

APPLICATION DES MÉTHODES MÉCANOGRAPHIQUES À UN AUTEUR MÉDIÉVAL

Les progrès de la technique moderne offrent à la philologie des possibilités nouvelles. Un domaine aussi peu exploré encore que le latin médiéval bénéficiera amplement de l'aide de la mécanographie et des ordinateurs électroniques. Bien peu de philologues ont osé, et à juste titre, entreprendre l'étude exhaustive de la langue d'un auteur. Si diligent qu'ait été Max Bonnet dans l'ouvrage classique qu'il a consacré à la langue de Grégoire de Tours, il reconnaît à plusieurs reprises n'avoir pu être exhaustif¹, et nous attendons toujours un *index verborum* de cet auteur. Remarquons d'ailleurs combien une telle étude est demeurée isolée. Les travaux portant sur la latinité d'auteurs médiévaux demeurent extrêmement rares ; bien souvent encore, leur caractère fragmentaire risque de déformer la réalité. La quasi totalité de la littérature médiolatine reste inexplorée du point de vue philologique. En 1929 déjà, dans son *Einführung in das Mittellatein*, Karl Streckler insistait sur la nécessité de soumettre chaque écrivain à une étude particulière et de consacrer à sa langue un examen spécial². Mais comment ne pas s'effrayer devant l'abondance de la production médiévale ? Aussi a-t-on trop souvent renoncé aux études exhaustives ; bien des phénomènes ont de la sorte échappé aux philologues.

La rapidité du traitement en machine permet ces études qui, jusqu'aujourd'hui, étaient pratiquement impossibles ou trop

1. Max BONNET, *Le latin de Grégoire de Tours*, Paris, 1890. Ainsi par exemple p. 495, après avoir souligné que l'étude de la syntaxe est le cœur même de son sujet, l'auteur écrit : « Il faut nécessairement nous borner à en tracer les grandes lignes, et faire un choix parmi les questions à traiter ».

2. Cf. Karl STRECKER, *Introduction à l'étude du latin médiéval*, trad. P. VAN DE WOESTIJNE, 3^e éd., Lille-Genève, 1948, p. 21.

fastidieuses. Comme l'écrivent MM. Pierre Demarne et Max Rouquerol, « l'ordinateur est un fantastique manœuvre intellectuelle »³. C'est dire par là que ces machines ne sont pas intelligentes ! Tout un travail préparatoire est à réaliser par le philologue avec une minutie aussi grande que possible. Les machines nous permettent à la fois de consacrer la plus grande partie de notre temps au travail proprement philologique et d'envisager de grands ensembles. Au point de départ elles nous obligent à poser les problèmes d'une façon extrêmement précise.

Un coup d'essai avait été tenté en 1951 déjà par le R. P. Busa avec les hymnes de saint Thomas d'Aquin⁴. Il s'est depuis attelé à la mise sur fiches de toute l'œuvre de saint Thomas⁵. Dom Jacques Froger a écrit en 1961 une étude sur l'*Emploi de la machine électronique dans les études médiévales*⁶. Il nous y offre un éventail des domaines d'application des machines électroniques en matière littéraire — l'édition critique des textes, l'étude de ceux-ci sous leurs divers aspects — non sans entrevoir des possibilités futures, comme la lecture automatique. Plus récemment, dans une communication faite à la Classe des Lettres de l'Académie royale de Belgique, M. Léopold Génicot a lancé un appel pour l'emploi des ordinateurs électroniques dans les études médiévales⁷. Si tous les philologues ne sont pas encore convaincus de l'apport que nous offre la technique d'aujourd'hui, et s'il est

3. Pierre DEMARNE et Max ROUQUEROL, *Les ordinateurs électroniques*, coll. *Que sais-je ?*, Paris, 1961, p. 61.

4. Roberto BUSA, *Sancti Thomae Aquinatis hymnorum ritualium varia specimina concordantiarum*, *Archivum Philosophicum Aloisianum*, série II, 7, Milan, 1951.

5. Le P. Busa a fait part aux médiévistes de l'avancement de son travail au premier congrès international de philosophie médiévale, qui s'est tenu à Louvain en 1958. Cf. R. BUSA, *L'automation appliquée à l'analyse linguistique des ouvrages de saint Thomas d'Aquin : programme, état actuel*, dans *L'homme et son destin d'après les penseurs du moyen âge, Actes du premier congrès international de philosophie médiévale*, Louvain-Paris, 1960, pp. 619-625. Depuis, le P. Busa a diffusé quelques *Indicis Thomistici specimen* qui portent la date du 31 mai 1963.

6. Jacques FROGER, *Emploi de la machine électronique dans les études médiévales*, dans *Bulletin de la Société internationale pour l'Étude de la philosophie médiévale*, 3 (1961), pp. 177-188.

7. Léopold GENICOT, *Ordinateurs électroniques et études médiévales*, dans *Bulletin de l'Académie royale de Belgique (Classe des Lettres)*, 5^e série, 49 (1963), pp. 66-76.

nécessaire dès lors que des savants expliquent et exhortent, nous sommes déjà passés, dans le domaine de l'étude des langues anciennes, au stade des réalisations. En 1961, en effet, M. Louis Delatte a créé à l'Université de Liège un Laboratoire d'Analyse statistique des Langues anciennes, et depuis, les diverses publications de l'équipe de ce Laboratoire permettent aux philologues de se faire une idée et des méthodes employées, et des premiers résultats obtenus⁸.

Étudiant l'œuvre de Raoul de Saint-Trond, nous avons nous-même bénéficié de l'équipement de ce Laboratoire et de l'expérience de son personnel⁹. Nous en avons suivi les techniques et les méthodes. Celles-ci ont été exposées de façon très claire par MM. Louis Delatte et Étienne Evrard. Nous ne saurions mieux faire que de renvoyer le lecteur à leurs exposés¹⁰, pour ne décrire ici que les diverses étapes du travail. Nous traiterons ensuite de diverses utilisations d'un fichier dûment établi, en matière lexicographique, grammaticale et stylistique, en fournissant quelques illustrations. Le plan de notre exposé sera donc le suivant :

1. Constitution du fichier

- a) Au point de vue lexicographique
- b) Au point de vue grammatical
- c) Au point de vue stylistique

8. Le Laboratoire d'Analyse statistique des Langues anciennes a entrepris l'étude de Sénèque le Philosophe ; trois *indices verborum* complets munis de relevés statistiques ont été publiés jusqu'à ce jour : *Sénèque, Consolation à Polybe : Index verborum, Relevés statistiques*, Liège, 1962 ; *Sénèque, Consolation à Helvia : Index verborum, Relevés statistiques*, La Haye, 1963 ; *Sénèque, Consolation à Marcia : Index verborum, Relevés statistiques*, La Haye, 1964.

9. Nous ne saurions trop souligner ici notre dette de reconnaissance à l'égard de M. le Professeur Louis Delatte, qui nous a offert de réaliser notre travail sur les machines et selon les méthodes de son Laboratoire, et n'a cessé de manifester pour nos recherches un intérêt constant. Quant à M. E. Evrard, il s'est bien souvent penché sur nos problèmes, et nous a apporté pour la réalisation de notre travail une aide irremplaçable.

10. Louis DELATTE et Étienne EVRARD, *Un Laboratoire d'Analyse statistique des Langues anciennes à l'Université de Liège*, dans *L'Antiquité Classique*, 30 (1961), fasc. 2, pp. 427-442.

Louis DELATTE, *Techniques et méthodes du Laboratoire d'Analyse statistique des Langues anciennes*, dans *Bulletin de l'Association des Classiques de l'Université de Liège*, 10 (1962), n° 2, pp. 33-54. *Un Laboratoire d'Analyse statistique des Langues anciennes à l'Université de Liège*, dans *Bulletin de l'Association des Amis de*

2. Diverses utilisations du fichier

a) Au point de vue lexicographique

- 1) *Index Verborum*
- 2) Listes de fréquence
- 3) Tableaux de distribution du vocabulaire
- 4) Mots significatifs les plus fréquents
- 5) Concordances

b) Au point de vue grammatical

c) Au point de vue stylistique

* * *

I. CONSTITUTION DU FICHIER.

La première étape du travail consiste à transcrire le texte sur cartes perforées. Toute l'œuvre en prose de Raoul de Saint-Trond a été ainsi mise sur cartes. Au préalable nous avons besoin d'une édition qui présentât le maximum de fidélité ; or ni l'édition de Kœpke dans les *Monumenta Germaniae Historica*, ni celle de de Borman ne remplissent cette condition ¹¹. Nous nous sommes donc efforcé d'établir une édition nouvelle aussi fidèle que possible

l'Université de Liège, 36 (1964), n° 1, pp. 5-34. On consultera également les introductions des indices publiés.

11. *Gesta abbatum Trudonensium*, éd. R. KOEPKE, dans *Monumenta Germaniae Historica, Scriptores*, X, Hanovre, 1852 ; *Chronique de l'abbaye de Saint-Trond*, I, éd. C. DE BORMAN, Liège, 1877.

La *Patrologie latine* reproduit le texte de Kœpke dans le tome 173, c. 659-708. La première édition, fragmentaire d'ailleurs, celle de Dom Luc d'Achery dans son *Spicilegium*, II, pp. 658-708, ne répond pas aux exigences d'une édition savante.

Cf. nos *Notes sur le texte de Raoul de Saint-Trond*, dans *ALMA*, 32 (1962), pp. 95-115. Citons deux exemples où se manifeste l'insuffisance des anciennes éditions pour l'étude du vocabulaire. M. Niermeyer a accueilli dans son *Mediae latinitatis lexicon minus un inclamosus*, qu'il a trouvé dans l'édition de de Borman (p. 59), qui lui-même a reproduit le texte de Kœpke (p. 249, l. 48) ; or si l'on se reporte aux manuscrits on lit, et très nettement, *in clamosa voce* ! *Reintroire*, d'autre part, qui n'est attesté ni chez Forcellini, ni chez Blaise, ni chez Ducange, ni chez Niermeyer, est devenu *introire* chez les éditeurs (p. 260, l. 37 éd. Kœpke ; p. 88 éd. de Borman).

aux manuscrits¹². Cette édition a été transcrite sur cartes perforées : un mot-forme par carte.

Après les vérifications nécessaires, un programme approprié a permis la perforation automatique sur chaque fiche de ses références propres : page, ligne, numéro d'ordre dans la phrase¹³, numéro d'ordre dans l'œuvre — pour rendre cette perforation automatique possible, il a fallu, lors de la transcription sur cartes perforées, munir de perforations spéciales les mots qui sont en fin de page, en fin de ligne ou en fin de phrase.

Chaque fiche, et donc chaque mot du texte, a ensuite été analysé du point de vue lexicographique, grammatical et stylistique. Les résultats de cette analyse philologique ont été perforés selon un système de codification approprié. Donnons un exemple. Les cartes sont divisées en 80 colonnes : chaque colonne peut recevoir un certain nombre de perforations. L'analyse d'une forme comme *amabit* se perforera notamment dans les colonnes 71 à 75 ; on y notera 5113E, c'est-à-dire verbe (5 en c. 71) de la première conjugaison (1 en c. 72) indicatif (1 en c. 73) futur (3 en c. 74) troisième personne du singulier de la voix active (E en c. 75 — les lettres de l'alphabet correspondent à une combinaison de deux perforations numériques). Choisissons un autre exemple au point de vue de l'analyse stylistique : 2 en colonne 77 sur la fiche d'un verbe indique qu'il s'agit d'un verbe subordonné et que la proposition subordonnée dont il fait partie suit la principale. MM. Louis Delatte et Étienne Evrard ont publié le détail du système de codification qu'ils ont créé ; on voudra bien s'y reporter¹⁴.

a) Au point de vue lexicographique, il faut indiquer sur chaque fiche le lemme correspondant à la forme. Tâche moins facile qu'il n'y paraît ! Il faut procéder à partir d'un lexique de base, mais lequel ? Le *Thesaurus* est loin d'être achevé. A défaut, et à l'exemple du comité de rédaction du *Novum Glossarium*, le seul lexique de base actuellement utilisable, est le *Totius Latinitatis*

12. C'est celle que la Commission Royale d'Histoire a bien voulu nous charger de publier dans son *Recueil de textes pour servir à l'étude de l'histoire de Belgique*.

13. Le numéro d'ordre dans la phrase permettra de connaître de façon automatique toutes les longueurs de phrase. Cf. *infra* p. 152.

14. Cf. notes 8 et 10.

Lexicon de Forcellini, tel que l'a édité Corradini ¹⁵. Il faut distinguer dès lors les mots qui ne figurent pas chez Forcellini. Notons qu'il est parfois difficile de décider si telle forme doit être considérée comme une variante orthographique, ou s'il faut la considérer comme un mot nouveau.

Quelle attitude prendre à l'égard des évolutions sémantiques ? Le *Lexicon*, on le sait, contient parfois de fâcheuses inconséquences. Ainsi voit-on Forcellini tantôt citer les auteurs chrétiens et tantôt les ignorer. Pour tel mot rare, on ne trouve chez lui que la seule référence d'Isidore de Séville ; d'autres fois, un mot attesté chez un saint Augustin n'y figure pas. Faire intervenir le *Dictionnaire des auteurs chrétiens* de Blaise n'est guère possible, car comment présenter un relevé de vocabulaire objectif en usant dans une proportion inégale de deux lexiques conçus de façon différente ? Encore faut-il souligner le caractère fragmentaire et nécessairement provisoire de ce dictionnaire. Au point de départ nous avons voulu multiplier les distinctions, englober notamment dans la catégorie des mots qui ne figurent pas chez Forcellini, ceux qui ont subi une évolution sémantique profonde. Mais comment établir les limites ? Distinguer les mots qui ont trait aux institutions médiévales ? Comment cependant décider que tel mot est médiéval, alors qu'il pourrait n'être que la spécification du mot ancien ? Qu'on songe à des mots comme *claustrum* et *praepositus* tels qu'ils apparaissent le plus souvent dans les textes médiévaux ; les sens nouveaux restent dans la ligne de l'étymologie de ces mots. Le plus souvent, une distinction parfaitement claire pour un mot ne l'est plus du tout pour un autre. Aussi, après l'examen de tous les cas qui se présentent dans l'œuvre de Raoul de Saint-Trond, avons-nous décidé qu'actuellement, *au niveau des relevés de vocabulaire généraux*, pour que ceux-ci puissent vraiment servir, sans ambiguïté possible, il faut s'en tenir à une distinction matérielle : le mot figure ou ne figure pas chez Forcellini — ceci n'inclut pas, bien entendu, le cas de lemmes homographes. Il appartient à une phase ultérieure du travail de donner des précisions ; le caractère même d'un relevé général de vocabulaire ne permet pas au lecteur de

15. FORCELLINI, *Totius Latinitatis Lexicon*, éd. CORRADINI, 4 vol., Padoue, 1864.

contrôler la part de subjectivité dans l'interprétation des mots. Le but est de livrer aux chercheurs des instruments de travail qui ne permettent *aucune confusion*. Ce sont en effet de tels travaux qui nous font le plus cruellement défaut ; ceux-ci doivent être aussi objectifs que possible. Une même attitude nous a guidé en ce qui concerne les sous-lemmes, que ceux-ci figurent déjà chez Forcellini, ou que, pour plus de clarté, il ait fallu en créer de nouveaux. Ces sous-lemmes ne sont considérés comme créations nouvelles, que lorsque l'on aurait pu en faire un lemme nouveau. Ainsi on aurait pu créer le lemme *precaria* ; pour respecter la façon de procéder générale de Forcellini nous avons préféré adopter le lemme *precarius* et créer le sous-lemme *precaria* (que l'on considérera comme création nouvelle) ; c'est d'ailleurs la méthode que suit le *Mittelateinisches Wörterbuch*.

Un certain nombre de formes se trouvent toujours écrites dans les manuscrits en un seul mot ; ainsi *nullomodo*, *intantum*. Cette graphie se retrouve constamment ; on ne la considérera donc pas comme une simple habitude d'un copiste donné. Elle semble indiquer, tout comme le contexte d'ailleurs, dans un cas comme *intantum* par exemple, qu'il s'agit là d'un seul mot. En toute objectivité, ces mots sont également considérés comme ne figurant pas dans Forcellini ¹⁶.

Des formes orthographiques diverses d'un même mot ont été évidemment groupées sous le lemme tel qu'il est écrit chez Forcellini, puisque ce dernier est notre lexique de base. Il en va de même des noms propres, pour autant que ceux-ci figurent dans l'*Onomasticon* de Forcellini ¹⁷. Pour les formes différentes d'un même nom géographique nouveau, nous avons généralement retenu comme lemme la forme la plus fréquente ¹⁸.

b) Au point de vue grammatical, on fait figurer sur chaque fiche la catégorie grammaticale à laquelle le mot appartient.

16. Baxter et Johnson ont adopté une même attitude dans leur *Wordlist* : J. H. BAXTER, Ch. JOHNSON, Ph. ABRAHAMS, *Medieval latin wordlist from British and Irish sources*, Oxford, 1934. Cf. *quancilius*, *utquid*.

17. FORCELLINI, *Totius Latinitatis Onomasticon*, 2 vol., éd. PERIN, Padoue, 1913-1920.

18. Nous ne croyons pas nécessaire d'entrer ici dans tous les détails. Ceux-ci figureront dans l'introduction de notre *index verborum*.

En outre, pour un substantif : la déclinaison, le cas, le nombre ; on notera s'il est un attribut, s'il dépend d'une préposition, ou s'il s'agit d'un substantif au génitif dépendant d'un verbe. Pour les adjectifs : la classe, le cas, le nombre ; éventuellement le degré de comparaison, l'emploi substantivé ou adverbial. Pour les numéraux : le type, éventuellement des indications analogues aux substantifs ou aux adjectifs (cas, etc.). Pour les pronoms et les adjectifs-pronoms : le type, l'emploi pronominal, adjectif ou adverbial, et éventuellement des indications analogues aux deux premières catégories. Pour les verbes : la conjugaison, le mode, le temps, la voix, la personne, le nombre ; on notera s'il s'agit d'un verbe principal ou subordonné, et dans ce cas, on indiquera son mode de subordination ; pour les participes, éventuellement des indications analogues à celles des substantifs ou des adjectifs, selon leur emploi substantivé ou adjectivé ; pour les gérondifs et les adjectifs verbaux : le cas et le nombre ; des indications spéciales pour le verbe avec auxiliaire exprimé ou sous-entendu. Pour les adverbes : le type, éventuellement le degré de comparaison. Pour les prépositions : le cas régi ; on notera s'il s'agit d'une conjonction de coordination ou de subordination ; pour les conjonctions de subordination : le mode et le temps qu'elles entraînent (de même pour les adjectifs-pronoms et les adverbes qui introduisent une subordonnée).

c) Au point de vue stylistique, nous indiquons le nombre de syllabes de chaque forme, l'ordre des mots et des propositions ; l'ordre des mots dans le groupe verbe-sujet ; la place de l'attribut par rapport au verbe et au sujet ; la place du complément d'objet direct par rapport au verbe et au sujet ; la place du complément déterminatif, de l'épithète, du participe apposé, de l'élément significatif des formes verbales périphrastiques, de la préposition et de la conjonction ; la place des propositions subordonnées par rapport à la principale, en distinguant les divers types de subordonnées ¹⁹.

19. Pratiquement tous les cas sont envisagés. Ainsi pour l'analyse de l'ordre des mots dans le groupe verbe-sujet, distinguons-nous : le verbe précède le sujet et est — ou n'est pas — en tête de proposition ; le verbe suit le sujet et est — ou n'est pas — en fin de proposition ; le verbe n'a pas de sujet exprimé

Nous croyons pouvoir dire que l'on se trouve ainsi dans une perspective toute nouvelle : combien d'œuvres médiévales ont été soumises à un pareil traitement, où chaque mot du texte est l'objet d'une analyse aussi complète que possible ? Qui ne s'est jamais astreint à une telle discipline, peut difficilement se rendre compte de tous les problèmes insoupçonnés qui ne cessent de se poser, surtout pour cette langue médiévale constituée d'éléments divers.

Une perforation spéciale distingue les mots qui font partie d'une citation ; une codification propre fait état des questions relatives à la tradition manuscrite. Nous avons noté de même sur chaque fiche de quel manuscrit provient chacune des formes du texte. Ce détail apparaîtra dans notre *index verborum* ; nous rassemblons de la sorte pour le chercheur intéressé les éléments d'une étude de l'orthographe des manuscrits.

Nous accumulons ainsi sur nos cartes une foule de renseignements, dont l'intérêt ira croissant avec l'élaboration de travaux similaires portant sur d'autres auteurs médiévaux.

2. UTILISATIONS DU FICHER.

a) AU POINT DE VUE LEXICOGRAPHIQUE.

1) *Index verborum*.

Voilà donc notre fichier constitué : sur chaque fiche, suite aux opérations dont nous avons parlé, figure la forme, avec toutes ses références — rappelons-les : page, ligne, numéro d'ordre dans la phrase et dans l'œuvre —, le lemme correspondant, les résultats de l'analyse grammaticale et stylistique. Les *Gesta abbatum Trudonensium*, I-VII et les *Epistulae* de Raoul de Saint-Trond nous fournissent 35.752 fiches²⁰. L'analyse

est en tête de proposition ou est en fin de proposition ; le verbe constitue à lui seul la proposition. Pour la place de l'épithète, on distingue l'adjectif, le mot employé adjectivement, l'adjectif-pronom employé adjectivement, le numéral ; on note si l'épithète précède ou suit le mot auquel elle se rapporte, avec ou sans disjonction — disjonction d'un ou de plusieurs mots, de même nature ou non. C'est dire que finalement l'on obtient un dossier complet pour l'étude de l'ordre des mots.

20. Nous ne considérons pas ici la lettre de Raoul qui introduit le livre IX des *Gesta*. Cf. *infra* note 34.

philologique a naturellement demandé assez bien de temps ; mais une fois cette analyse transcrite sur les cartes, les résultats de nos recherches apparaissent avec une rapidité déconcertante.

Grâce à une trieuse, on classe les fiches dans l'ordre alphabétique d'après le lemme, et pour chaque lemme, les diverses formes dans l'ordre grammatical adopté. A partir d'un fichier classé de la sorte, on imprimera l'*index verborum*²¹. On pourrait faire figurer dans l'*index verborum* tous les renseignements qui se trouvent sur les cartes, mais cela ne semble pas souhaitable ; aussi, pour la commodité du chercheur, faut-il choisir les données que l'on retiendra : le lemme (avec un signe spécial pour les lemmes non attestés dans Forcellini), les formes correspondantes avec leurs références, les indications relatives à la tradition manuscrite ; les mots faisant partie d'une citation ou servant de base à un enclitique, sont dotés d'un signe spécial. De plus nous avons noté le manuscrit dont provient chacune des formes. La reproduction par offset permettra la diffusion de cet *index*. Si le fichier au point de départ a été soigneusement contrôlé, aucune erreur ne se glissera dans les phases ultérieures du travail, soit dans la graphie du lemme ou celle des formes attestées, soit dans la référencement. La constitution d'*indices* est certes l'une des tâches les plus urgentes de la philologie médiolatine. Les moyens mécanographiques rendent ce travail réalisable pour un très grand nombre de textes et dans un court délai.

2) *Listes de fréquence.*

A partir du fichier complet d'une œuvre, on constitue, grâce à un programme approprié, un fichier de fréquence : ces nouvelles fiches comportent le lemme et sa fréquence d'emploi ; elles permettent la constitution de listes de fréquence, où, par exemple, les mots sont classés en ordre de fréquence décroissant — ce qui

21. Nos *indices* ont été réalisés sur l'organe d'impression de l'ordinateur Gamma-Tambour du Centre Interdisciplinaire de Calcul de l'Université de Liège, à raison de 140 lignes à la minute. Que M. le Professeur Linsman et les membres du personnel du Centre Interdisciplinaire de Calcul, auprès de qui nous avons toujours trouvé l'accueil le plus aimable, trouvent ici le témoignage de notre reconnaissance.

Actuellement, tous les travaux se font au siège même du Laboratoire d'Analyse statistique des Langues anciennes, qui dispose de toutes les machines mécanographiques nécessaires.

implique qu'au préalable on a trié ces fiches selon cet ordre ; comme la trieuse dont nous disposons travaille à la cadence de 1.000 cartes à la minute, toutes ces opérations sont extrêmement rapides. Nous avons dressé plusieurs listes de fréquence : le vocabulaire total de l'œuvre, le vocabulaire non attesté dans le *Lexicon* de Forcellini, le vocabulaire total des citations, le vocabulaire des citations non attesté chez Forcellini. Publier *in extenso* les listes des *Gesta* et des *Epistulae* nous est impossible dans le cadre du présent article. Ici, comme dans la suite, nous devons nous contenter de donner au lecteur en guise d'illustration, quelques *excerpta*. Voici, par exemple, les fréquences supérieures à vingt des *Gesta* et des *Epistulae* :

Gesta abbatum Trudonensium, I-VII ²²

<i>et</i>	971	<i>nos</i>	159	<i>magnum</i>	82
<i>qui</i> 1	581	<i>possum</i>	156	<i>quam</i>	82
<i>is</i>	530	<i>in</i>	acc 139	<i>et</i>	adv 79
<i>que</i>	382	<i>ecclesia</i>	138	<i>deus</i>	78
<i>noster</i>	358	<i>sed</i>	132	<i>tamen</i> 1	76
<i>de</i>	335	<i>per</i>	130	<i>totus</i> 2	75
<i>in</i>	abl 327	<i>Herimannus</i>	npr 128	<i>aut</i>	74
<i>ad</i>	318	<i>ut</i>	128	<i>dico</i> 2	73
<i>sum</i> 1	270	<i>habeo</i> 1	118	<i>tandem</i>	73
<i>sumus</i>	259	<i>facio</i>	114	<i>abbatia.</i>	72
<i>ille</i>	222	<i>cum</i> 1	112	<i>imperator</i>	70
<i>non</i>	222	<i>ipse</i>	106	<i>quis</i> 1	70
<i>abbas</i>	216	<i>sanctus</i>	105	<i>alius</i>	67
<i>sum</i> 1	aux 201	<i>omnis</i>	104	<i>si</i>	67
<i>hic</i> 1	198	<i>cum</i> 2	96	<i>noster</i>	66
<i>sui</i>	198	<i>dies</i>	96	<i>usque</i>	65
<i>a</i> 3	178	<i>ego</i>	88	<i>annus</i>	63
<i>episcopus</i> 2	169	<i>multus</i>	87	<i>locus</i>	62
<i>frater</i>	166	<i>video</i>	86	<i>monasterium</i>	61

22. Explication des codes et sigles qui apparaissent dans ces extraits de listes :

Indice numérique suivant le lemme ou le sous-lemme : correspond à l'ordre du *Lexicon* de Forcellini.

Point suivant le lemme ou le sous-lemme : signifie que ce lemme ou ce sous-lemme n'est pas attesté chez Forcellini.

Abl : ablatif ; acc : accusatif ; adj : sous-lemme adjectif ; adv : sous-lemme adverbe ; aux : auxiliaire ; c.c : sous-lemme conjonction de coordination ; npr : nom propre ; s : sous-lemme substantif.

Quant à la question des sous-lemmes, cf. *supra* p. 131.

<i>pro</i> 1		61	<i>Trudo</i>	npr	42	<i>nec</i>		26
<i>verus</i>	c.c	61	<i>tantus</i>	adv	41	<i>refero</i>		26
<i>volo</i> 2		61	<i>idem</i>		40	<i>religiosus</i>		26
<i>ex</i>		60	<i>aliquis</i>		39	<i>duo</i>		25
<i>apud</i>		59	<i>manus</i> I		39	<i>multum</i>		25
<i>comes</i>		59	<i>claustrum</i>		38	<i>quisquis</i>		25
<i>igitur</i>		59	<i>jam</i>		37	<i>sicut</i>		25
<i>nullus</i>		58	<i>Luiipo</i>	npr	37	<i>superus</i>	adv	25
<i>quidam</i>		58	<i>ordo</i>		37	<i>coenobium</i>		24
<i>tam</i>		58	<i>domnus</i>		36	<i>dum</i>		24
<i>aulem</i>		57	<i>nam</i>		36	<i>fio</i>		24
<i>tunc</i>		57	<i>Theodericus</i>	npr	36	<i>Gislebertus</i>	npr	24
<i>dominus</i> 2		55	<i>enim</i>		35	<i>invenio</i>		24
<i>etiam</i>		53	<i>nox</i>		35	<i>novus</i>		24
<i>inter</i>		53	<i>oppidum</i>		35	<i>parvus</i>		24
<i>mettensis</i>	npr	53	<i>sic</i>		35	<i>pono</i>		24
<i>quia</i>		53	<i>turris</i>		35	<i>primus</i>		24
<i>tempus</i>		53	<i>beatus</i>		34	<i>suus</i>	s. I	24
<i>quoque</i>		52	<i>malus</i> I	s. 2	34	<i>fero</i>		23
<i>ut</i>	adv	52	<i>super</i>	acc	34	<i>littera</i>		23
<i>do</i> 2		51	<i>bonus</i>		33	<i>modo</i>		23
<i>ibi</i>		51	<i>Lanzo</i>	npr	33	<i>opus</i> I		23
<i>quod</i> 1		51	<i>Mettis</i>	npr	33	<i>pax</i>		23
<i>Leodium</i>	npr	50	<i>coepio</i>		32	<i>quasi</i>		23
<i>Heinricus</i>	npr	49	<i>necnon</i>		32	<i>accipio</i>		22
<i>homo</i>		49	<i>ante</i>		31	<i>animus</i>		22
<i>neque</i>		49	<i>quidem</i>		31	<i>equus</i>		22
<i>res</i>		49	<i>secundus</i>		31	<i>paene</i>		22
<i>unus</i>		49	<i>consilium</i>		30	<i>propter</i>		22
<i>venio</i>		48	<i>cotidie</i>		30	<i>revertor</i>		22
<i>gravis</i>		47	<i>sine</i>		30	<i>suscipio</i>		22
<i>ita</i>		47	<i>sub</i>		30	<i>teneo</i>		22
<i>meus</i> I		46	<i>vita</i>		30	<i>vel</i>		22
<i>nilhil</i>		46	<i>audio</i>		29	<i>audeo</i>		21
<i>scilicet</i>		46	<i>post</i>		29	<i>clericus</i>		21
<i>ago</i>		45	<i>prius</i>		29	<i>contra</i>		21
<i>atque</i>		45	<i>vix</i> I		29	<i>honor</i>		21
<i>tantus</i>		45	<i>Adelardus</i>	npr	28	<i>multitudo</i>		21
<i>leodiensis</i>	npr	44	<i>dux</i>		28	<i>nimis</i>		21
<i>magis</i> I		44	<i>iste</i>		28	<i>nunc</i>		21
<i>monachus</i>		44	<i>judicium</i>		28	<i>praepositus</i> 2		21
<i>pars</i>		44	<i>domus</i>		27	<i>quantus</i>		21
<i>adhuc</i>		43	<i>interea</i>		27	<i>semper</i>		21
<i>nomen</i>		43	<i>necessitas</i>		27	<i>unde</i>		21
<i>vir</i>		43	<i>praesum</i>		27	<i>vivo</i>		21
<i>inde</i>		42	<i>debeo</i>		26			

Epistulae

<i>et</i>	434	<i>idem</i>	47	<i>frater</i>	26
<i>qui</i> 1	253	<i>omnis</i>	46	<i>pro</i> 1	26
<i>is</i>	186	<i>ego</i>	42	<i>coenobium</i>	25
<i>in</i>	abl 167	<i>cum</i> 1	41	<i>propter</i>	25
<i>de</i>	156	<i>pauper</i>	41	<i>que</i>	25
<i>sum</i> 1	136	<i>volo</i> 2	41	<i>vos</i>	25
<i>non</i>	109	<i>villa</i>	40	<i>pater</i>	24
<i>ad</i>	106	<i>dico</i> 2	39	<i>simonia.</i>	24
<i>noster</i>	98	<i>si</i>	39	<i>beatus</i>	23
<i>suus</i>	93	<i>tu</i>	37	<i>dies</i>	23
<i>hic</i> 1	89	<i>unus</i>	37	<i>annus</i>	22
<i>sum</i> 1	aux 78	<i>dominus</i> 2	36	<i>avaritia</i>	22
<i>deus</i>	73	<i>homo</i>	36	<i>do</i> 2	22
<i>ecclesia</i>	66	<i>quam</i>	36	<i>duo</i>	22
<i>sanctus</i>	66	<i>ipse</i>	35	<i>iste</i>	22
<i>a</i> 3	63	<i>possum</i>	35	<i>tuus</i>	22
<i>facio</i>	63	<i>multus</i>	32	<i>advocatus</i> 2	21
<i>sui</i>	63	<i>video</i>	32	<i>capio</i> 1	21
<i>sed</i>	60	<i>in</i>	acc 30	<i>cum</i> 2	21
<i>filius</i>	59	<i>magnus</i>	30	<i>quidam</i>	21
<i>ut</i>	56	<i>nos</i>	29	<i>quod</i> 1	21
<i>ille</i>	55	<i>quia</i>	29	<i>sicut</i>	21
<i>per</i>	54	<i>alius</i>	27	<i>totus</i> 2	21
<i>habeo</i> 1	51	<i>ex</i>	27		
<i>debeo</i>	48	<i>res</i>	27		

Compter le nombre de *et* — plus exactement, faire compter ce nombre par la machine ! — pourrait paraître à certains bien superflu. On remarquera avec intérêt que les mots-outils affectés des fréquences les plus élevées semblent nettement caractéristiques. Voyez dans les *Gesta* : *et* (971) qui précède de loin *qui* (581) ; *que* (382), *de* (335) ; *is* (530) devance de beaucoup *ille* (222), *hic* (198), et *iste* (28). De même dans les *Epistulae* : *et* (434) avant *qui* (253), *de* (156) ; *is* (186) avant *hic* (89), *ille* (55), *iste* (22).

Les mots de fréquence très peu élevée caractérisent également — et parfois même davantage — la langue d'un auteur. Mais nous avons dû opérer un choix dans nos exemples, et les listes de mots à fréquence peu élevée sont évidemment beaucoup plus longues ²³.

23. Cf. pour les *Gesta*, I-VII : nous savons, grâce au tableau de distribution du vocabulaire, que 141 mots ont la fréquence 5, 207 la fréquence 4, 350 la fréquence 3, 687 la fréquence 2 et 1684 la fréquence 1.

Nous avons déjà noté qu'une des listes de fréquence groupe le vocabulaire non attesté dans le *Lexicon* de Forcellini. Étant donné que le nombre de mots nouveaux est, chez Raoul, remarquablement restreint, nous publions ces listes *in extenso*, tout en les accompagnant, au-dessous de la fréquence 10, des références au texte, d'après l'édition des *Monumenta Germaniae Historica, Scriptores*, X²⁴.

*Gesta abbatum Trudonensium, I-VII*²⁵,

<i>abbatia.</i>		72	
<i>marca.</i>		17	
<i>nullomodo.</i>		12	
<i>protomartyr.</i>		9	235, 37 ; 246, 14 ; 247, 44 ; 254, 17, 23 ; 255, 20 ; 256, 37 ; 266, 34 ; 268, 51 .
<i>episcopium.</i>		8	233, 44 ; 238, 31, 44 ; 245, 2 ; 246, 5 ; 248, 38, 44 ; 260, 6.
<i>auritrigium.</i>		7	256, 52 ; 257, 1, 3, 14, 15, 17, 17.
<i>matutinus</i>	s.	6	246, 23 ; 256, 2, 5, 6, 6, 10.
<i>alodium.</i>		5	264, 3, 7, 10, 10 ; 266, 34.
<i>bannus.</i>		5	238, 31, 44 ; 249, 1 ; 260, 46 ; 267, 50.
<i>beneficio.</i>		5	235, 35 ; 246, 40 ; 253, 25 ; 256, 49 ; 261, 23.
<i>paupost.</i>		5	239, 17, 37 ; 245, 15 ; 251, 8 ; 267, 51.
<i>refectorium.</i>		5	255, 43 ; 256, 13, 15, 24, 25.
<i>capella.</i>		4	255, 10, 22, 25 ; 271, 17.
<i>collateralis.</i>	s	4	265, 4 ; 268, 13, 30, 45.
<i>instantum.</i>		4	261, 17 ; 268, 14, 50 ; 269, 40.
<i>missaticus.</i>		4	246, 14 ; 254, 15 ; 262, 18 ; 263, 24.
<i>parefridus.</i>		4	234, 17 ; 250, 5 ; 268, 36, 40.
<i>vadium.</i>		4	235, 16 ; 249, 15 ; 256, 43 ; 268, 35.
<i>gemituosus.</i>		3	243, 45 ; 244, 20 ; 245, 7.

24. Cette liste a été évidemment constituée d'après notre propre édition, mais nous avons conservé les références à l'édition de Kœpke, considérée d'ailleurs comme édition *princeps*.

25. Dans cette liste de fréquence, figure, comme nous l'avons expliqué, le lemme suivi de sa fréquence. Dès lors les formes que ces mots médiévaux revêtent dans le texte, présentent parfois une orthographe différente. *Indices verborum* et listes de fréquence se complètent. D'autre part, nous ne croyons pas nécessaire d'exposer ici dans le détail les règles qui ont présidé à l'orthographe du lemme pour les mots non attestés chez Forcellini.

On constatera que pour certains mots, le sous-lemme seul est suivi d'un point ; lui seul doit être considéré en ce cas comme un mot nouveau. Le lecteur se rapportera aux explications données *supra* p. 131. L'avantage de cette façon de faire apparaît plus nettement dans le cadre général de l'*index verborum*.

<i>hominium.</i>	3	264, 47, 49; 269, 18.
<i>praelatus.</i>	3	232, 33; 236, 2; 249, 28.
<i>rehabeo.</i>	3	253, 23; 257, 22; 264, 13.
<i>ambulatorium.</i>	2	242, 33, 37.
<i>armarius.</i>	2	232, 17, 17.
<i>camerarius</i> s.	2	232, 24, 28.
<i>caminata.</i>	2	260, 18, 32.
<i>coanimatus.</i>	2	231, 30; 267, 44.
<i>comitissa.</i>	2	267, 45; 270, 1.
<i>crucifigo</i> s.	2	243, 53; 257, 11. *
<i>dominicalis.</i> s	2	246, 31, 37.
<i>feodum.</i>	2	253, 27; 268, 34.
<i>grisius.</i>	2	268, 36; 269, 47.
<i>inundantia.</i>	2	232, 27; 268, 24.
<i>laicalis.</i>	2	241, 32; 269, 15.
<i>multipost.</i>	2	260, 24, 26.
<i>pellicius</i> s.	2	268, 36; 269, 47.
<i>precarius</i> s.	2	235, 25; 251, 15.
<i>quamcito.</i>	2	254, 37; 256, 36.
<i>scutifer.</i>	2	265, 24, 45.
<i>secretarius.</i>	2	249, 11; 262, 41.
<i>simonia.</i>	2	269, 2; 271, 29.
<i>simoniacus.</i>	2	230, 10; 231, 16.
<i>spicarium.</i>	2	253, 6; 268, 45.
<i>tenuē.</i>	2	240, 16; 266, 17.
<i>tubeus.</i>	2	232, 19; 233, 29.
<i>ullomodo.</i>	2	265, 2; 272, 6.
<i>viscerabiliter.</i>	2	229, 28; 253, 39.
<i>advocatia.</i>	I	272, 6.
<i>anathematizatio.</i>	I	261, 1.
<i>auctorizo.</i>	I	238, 41.
<i>beneficio.</i> s	I	241, 29.
<i>bipennula.</i>	I	266, 11.
<i>capellula.</i>	I	245, 41.
<i>cappula.</i>	I	270, 44.
<i>carrata.</i>	I	261, 19.
<i>caudatus.</i>	I	257, 8.
<i>circumadjaceo.</i>	I	234, 33.
<i>circumfuro.</i>	I	244, 14.
<i>compatienter.</i>	I	228, 16.
<i>coopertura.</i>	I	255, 40.
<i>cortinatus.</i>	I	234, 13.
<i>decapito.</i>	I	258, 41.

* *crucifixus*, -i.

<i>delatrix.</i>	I	237, 45.
<i>densis.</i>	I	267, 14.
<i>diminutivo.</i>	I	233, 38.
<i>dominicalis.</i>	I	252, 41.
<i>dominicus</i> s.	I	271, 50.
<i>ecclesiastice.</i>	I	250, 16.
<i>exfestuco.</i>	I	248, 41.
<i>exlegalitas.</i>	I	265, 7.
<i>exsilio.</i>	I	236, 18.
<i>famelicitas.</i>	I	257, 37.
<i>firto.</i>	I	257, 7.
<i>flammivomans.</i>	I	240, 4.
<i>imaginiola.</i>	I	257, 9.
<i>inaestimatio.</i>	I	248, 7.
<i>inantea.</i>	I	268, 24.
<i>incomputabilis.</i>	I	234, 22.
<i>incorrigibilitas.</i>	I	264, 41.
<i>inedicibilis.</i>	I	261, 38.
<i>invadio.</i>	I	256, 49.
<i>invitatorius</i> s.	I	232, 14.
<i>irremeabiliter.</i>	I	248, 44.
<i>irrepressibilis.</i>	I	263, 53.
<i>laneus</i> s.	I	248, 27.
<i>leccatria.</i>	I	259, 47.
<i>leve.</i>	I	254, 1.
<i>lubricamentum.</i>	I	236, 53.
<i>macerialis.</i>	I	234, 47.
<i>medo.</i>	I	249, 35.
<i>monachilis.</i>	I	269, 28.
<i>mundalis.</i>	I	233, 41.
<i>nutrio</i> s.	I	255, 3*
<i>oblectabiliter.</i>	I	228, 3.
<i>obumbratorium.</i>	I	266, 14.
<i>odorificenter.</i>	I	233, 13.
<i>offirmatio.</i>	I	258, 34.
<i>papatus.</i>	I	271, 52.
<i>paterne.</i>	I	229, 36.
<i>pingue.</i>	I	234, 4.
<i>prasconor.</i>	I	243, 1.
<i>profanaticus.</i>	I	245, 32.
<i>prosperor.</i>	I	251, 17.
<i>quammultus.</i>	I	249, 42.
<i>radicabilis.</i>	I	236, 44.

* *nutritus*, -i.

<i>reacquirō.</i>	I	245, 26.
<i>readipiscor.</i>	I	249, 38.
<i>reascendo.</i>	I	261, 41.
<i>reingredior.</i>	I	271, 12.
<i>reintroduco.</i>	I	262, 15.
<i>reintroeo.</i>	I	260, 37.
<i>sacrista.</i>	I	271, 34.
<i>sæcularitas.</i>	I	237, 5.
<i>teutonice.</i>	I	229, 28.
<i>tuella.</i>	I	257, 10.
<i>tunicella.</i>	I	270, 44.
<i>turrensis.</i>	I	242, 46.
<i>tyrannizo.</i>	I	268, 50.
<i>utquid.</i>	I	228, 1.
<i>vesperus</i>	s. I	239, 26.
<i>violamentum.</i>	I	248, 39.
<i>werra.</i>	I	234, 31.

Epistulae.

<i>simonia.</i>	24	
<i>marca.</i>	20	
<i>allodium.</i>	7	324, 47 ; 325, 1 ; 327, 3, 9, 9, 11 ; 328, 41.
<i>precarius</i>	s. 7	327, 17, 41, 45, 50 ; 328, 1, 10, 11.
<i>mansus.</i>	5	324, 49 ; 327, 9, 11, 12, 31.
<i>parefridus.</i>	4	326, 34 ; 327, 43 ; 329, 6, 13.
<i>simoniacus.</i>	4	323, 46, 52, 52 ; 324, 9.
<i>scabinus.</i>	3	326, 26, 29, 31.
<i>stipendio.</i>	3	320, 21 ; 321, 13 ; 322, 52.
<i>annono.</i>	2	328, 29, 49.
<i>aurifrigium.</i>	2	330, 30, 31.
<i>bonuarium.</i>	2	328, 49 ; 329, 49.
<i>censionarius.</i>	2	329, 44, 46.
<i>coenobialis.</i>	2	324, 17, 22.
<i>curtis.</i>	2	327, 21, 329, 22.
<i>dominicalis.</i>	2	327, 47 ; 329, 22.
<i>jeodium.</i>	2	324, 49 ; 327, 12.
<i>firto.</i>	2	327, 19, 26.
<i>nullomodo.</i>	2	329, 37, 38.
<i>proculdubio.</i>	2	319, 3 ; 324, 9.
<i>protomartyr.</i>	2	324, 41, 44.
<i>subtalaris</i>	s. 2	330, 34 ; 331, 35.
<i>abbatia.</i>	I	325, 1.
<i>advocatia.</i>	I	324, 49.
<i>brascena.</i>	I	326, 40.
<i>camba.</i>	I	326, 44.

<i>clamitosus.</i>	I	318, 52.
<i>curtilis.</i>	s I	327, 1.
<i>deplanctio.</i>	I	326, 35.
<i>devastatio.</i>	I	326, 23.
<i>dominicalis.</i>	s I	324, 44.
<i>dominicus</i>	s. I	327, 54.
<i>exemplifico.</i>	I	323, 32.
<i>flandrigena.</i>	I	328, 31.
<i>garba.</i>	I	326, 27.
<i>incommotus.</i>	I	330, 39.
<i>marescalcus.</i>	I	328, 30.
<i>multopost.</i>	I	318, 28.
<i>oblatratio.</i>	I	617, 10*
<i>rejudico.</i>	I	326, 43.
<i>solidarius.</i>	I	326, 34.
<i>spicarium.</i>	I	327, 49.
<i>stipendiator.</i>	I	322, 27.
<i>subadvocatus.</i>	I	327, 54.
<i>super globo.</i>	I	331, 27.
<i>teutonice.</i>	I	331, 48.
<i>tyrannizo.</i>	I	327, 35.
<i>valeo</i>	s. I	329, 48**
<i>werra.</i>	I	326, 35.

Chacun de ces mots mériterait une étude particulière, mais le cadre du présent article ne nous le permet pas ; nous désirons seulement présenter ici de façon globale l'application des méthodes mécanographiques aux auteurs médiévaux.

On peut constituer de même des relevés de mots anciens à sens nouveau.

Ainsi que nous l'avons déjà noté, nous avons distingué le vocabulaire proprement dit de Raoul du vocabulaire de ses citations. Bien que cette distinction semble aller de soi, on néglige souvent de l'établir. Ce vice de méthode a pu se glisser jusque dans le *Mittelateinisches Wörterbuch*, que l'on pourrait croire à l'abri d'une telle erreur : s.v. *alternitas*, qui est un mot relativement rare, on trouve seulement quelques références, et notamment les *Gesta* de Raoul ; or le texte de Raoul est une citation : il

* Cette référence renvoie à l'édition de la lettre de Raoul à Rupert de Deutz publiée par F. ROHN, *Ein Brief des Chronisten von St. Trond an Rupert von Deutz*, dans *Neues Archiv*, 17 (1892), pp. 617-618.

** *valens*, *-entis*.

fallait donc le signaler. Bien sûr, il ne faut pas non plus négliger les citations, car l'écrivain a choisi telle citation plutôt que telle autre ; elles peuvent révéler ses lectures, ses auteurs préférés. Tout le vocabulaire des citations des *Epistulae* est attesté chez Forcellini. Il n'en va pas de même pour les *Gesta* : un certain nombre de mots nouveaux y apparaissent, mais tous dans une même citation, celle de l'inventaire de 870, qui figure au début de premier livre. Voici la liste de fréquence de ces mots nouveaux, accompagnés de leur référence :

<i>corbus.</i>	3	231, 7, 8, 9.
<i>buxta.</i>	2	230, 33 ; 231, 1.
<i>cussinus.</i>	2	230, 38 ; 231, 2.
<i>repa.</i>	2	230, 23, 24.
<i>abbatia.</i>	1	230, 19.
<i>aquamabile.</i>	1	231, 2.
<i>cambuta.</i>	1	230, 37.
<i>clavicella.</i>	1	230, 37.
<i>crucicula.</i>	1	230, 31.
<i>dalmaticus</i> s.	1	230, 38.
<i>facistergius.</i>	1	230, 40.
<i>jaldo.</i>	1	231, 2.
<i>jano.</i>	1	230, 40.
<i>frisum.</i>	1	230, 39.
<i>hanapus.</i>	1	231, 1.
<i>mancosus.</i>	1	231, 6.
<i>sutilis.</i>	1	230, 34*.

3) Tableaux de distribution du vocabulaire.

Pour évaluer statistiquement le vocabulaire d'un auteur, il importe d'établir des tableaux de distribution. Cette opération s'effectue au moyen d'un programme approprié. Ces tableaux nous montrent la répartition générale du vocabulaire, depuis 1 mot répété x fois jusqu'à y mots employés 1 fois. Nous en avons constitué plusieurs :

1. le vocabulaire de Raoul
 - a) en y comprenant les noms propres

* *Sutilis* est la forme qui apparaît dans les manuscrits. Kœpke a corrigé *sutilis* en *buxtilis*, qui n'est cependant pas attesté par ailleurs. Nous croyons dès lors que jusqu'à preuve du contraire, il faut conserver *sutilis*, même si, actuellement, on n'en perçoit pas nettement le sens.

Quant à *aquamabile*, l'*index* signale qu'il s'agit d'une correction de J. Gessler.

b) en les excluant — vu leur nombre, ils pourraient en effet défigurer certains résultats.

2. le vocabulaire des noms propres
3. le vocabulaire non attesté chez Forcellini
4. le vocabulaire des citations
5. le vocabulaire des citations non attesté chez Forcellini.

En ce qui concerne le vocabulaire de Raoul lui-même, noms propres exclus, on voit que sa répartition va, pour les *Gesta*, du même mot employé 971 fois jusqu'aux 1565 mots employés 1 fois — les mots employés 1 fois représentent donc 43,9 % du vocabulaire —; et pour les *Epistulae*, du même mot employé 434 fois jusqu'aux 879 mots employés 1 fois — soit 50,5 % du vocabulaire —. Ces tableaux de distribution nous fournissent des indications extrêmement précieuses pour juger la richesse de vocabulaire d'un auteur ou d'une œuvre.

Voici les chiffres obtenus en fin de tableau, qui totalisent les occurrences et le vocabulaire ²⁶ :

Pour les *Gesta* :

	Occurrences	Vocabulaire
Total	26.270	4214
Citations exclues	25.440	3791
Citations et noms propres exclus	24.367	3566
Mots non attestés chez Forcellini	317	122
Mots faisant partie d'une citation non attestés chez Forcellini	22	17

Pour les *Epistulae* :

	Occurrences	Vocabulaire
Total	9482	2130
Citations exclues	8926	1888
Citations et noms propres exclus	8640	1741
Mots non attestés chez Forcellini	130	49

On ne manquera pas d'être frappé par le nombre minime de mots employés par Raoul, qui ne sont pas attestés dans le *Lexicon* de Forcellini. Si on se rapporte aux listes de fréquence ²⁷, plusieurs

26. Par occurrences, on entend le nombre de mots d'un texte ; par vocabulaire le nombre de mots différents.

27. Cf. *supra* pp. 138-142.

constatations s'imposent. Raoul n'emploie généralement des mots nouveaux que par nécessité. Il ne peut évidemment échapper à l'emploi de *abbatia* pour les *Gesta*²⁸, de *simonia* et de *marca* pour les *Epistulae* — la première lettre est en somme un bref traité sur la simonie, et la troisième, une longue réclamation à l'évêque de Metz. L'emploi répété de ces mots — ainsi qu'il ressort du seul examen des références — se situe très souvent dans un même passage, voire dans une même phrase. Nombreux sont les mots de cette catégorie qui ont trait aux institutions médiévales : *episcopium*, *alodium*, *bannus*, *beneficio*, etc. Dans les *Epistulae*, sur un vocabulaire de 49 mots nouveaux, 36 apparaissent dans la seconde et la troisième lettres, lesquelles précisément touchent à des questions d'organisation économique.

A côté de cette attitude générale, on constate pour les *Gesta* un phénomène particulier, dont, ici encore, le lecteur s'apercevra en se reportant aux listes de fréquence des mots non attestés chez Forcellini²⁹. Dans les *Gesta*, en effet, Raoul affectionne certaines créations nouvelles, en raison sans doute de leur caractère expressif. Il s'agit, en général, de mots notablement longs : *viscerabiliter*, *incomputabilis*, *incorrigibilitas*, *irremeabiliter*, *irrepressibilis*, *lubricamentum*, etc. On y remarquera aussi la fréquence des verbes formés à l'aide du préfixe *re-* : *rehabeo*, *reacquirō*, *readipiscor*, *reascendo*, etc. Dans les *Epistulae*, au contraire, aucun mot nouveau n'est particulièrement long ; pour le préfixe *re-*, la seule attestation est *rejudico*. Cette constatation que nous faisons à propos des *Gesta*, doit être mise en parallèle avec l'abondance des superlatifs attestés dans cette œuvre, ainsi que le fait apparaître le relevé statistique correspondant : ils y constituent 13,03 % du nombre d'adjectifs, alors que dans les *Epistulae* ils n'en représentent que 3,53 %.

4) Mots significatifs les plus fréquents.

Du vocabulaire de Raoul, extrayons la liste des cinquante mots significatifs les plus fréquents :

28. On remarquera en outre, que si notre lexique de base pouvait être le *Thesaurus*, *abbatia* y serait attesté.

29. Cf. *supra* pp. 138-141.

Gesta abbatum Trudonensium, I-VII :

Mots	Fréquence	Mots	Fréquence	Mots	Fréquence
1. <i>abbas</i>	216	18. <i>locus</i>	62	35. <i>claustrum</i>	38
2. <i>episcopus</i>	169	19. <i>monasterium</i>	61	36. <i>ordo</i>	37
3. <i>frater</i>	166	20. <i>volo (velle)</i>	61	37. <i>dommus</i>	36
4. <i>possum</i>	156	21. <i>comes</i>	59	38. <i>nox</i>	35
5. <i>ecclesia</i>	138	22. <i>dominus</i>	55	39. <i>oppidum</i>	35
6. <i>habeo</i>	118	23. <i>tempus</i>	53	40. <i>turris</i>	35
7. <i>facio</i>	114	24. <i>do</i>	51	41. <i>beatus</i>	34
8. <i>sanctus</i>	105	25. <i>homo</i>	49	42. <i>malum</i>	34
9. <i>dies</i>	96	26. <i>res</i>	49	43. <i>bonus</i>	33
10. <i>multus</i>	87	27. <i>venio</i>	48	44. <i>coepio</i>	32
11. <i>video</i>	86	28. <i>gravis</i>	47	45. <i>consilium</i>	30
12. <i>magnus</i>	82	29. <i>ago</i>	45	46. <i>vita</i>	30
13. <i>deus</i>	78	30. <i>monachus</i>	44	47. <i>audio</i>	29
14. <i>dico (dicere)</i>	73	31. <i>pars</i>	44	48. <i>dux</i>	28
15. <i>abbatia.</i>	72	32. <i>nomen</i>	43	49. <i>judicium</i>	28
16. <i>imperator</i>	70	33. <i>vir</i>	43	50. <i>domus*</i>	27
17. <i>annus</i>	63	34. <i>manus</i>	39		

Epistulae :

Mots	Fréquence	Mots	Fréquence	Mots	Fréquence
1. <i>deus</i>	73	18. <i>res</i>	27	35. <i>manus</i>	19
2. <i>ecclesia</i>	66	19. <i>frater</i>	26	36. <i>pars</i>	19
3. <i>sanctus</i>	66	20. <i>coenobium</i>	25	37. <i>redimo</i>	19
4. <i>facio</i>	63	21. <i>pater</i>	24	38. <i>vestis</i>	19
5. <i>filius</i>	59	22. <i>simonia.</i>	24	39. <i>suscipio</i>	18
6. <i>habeo</i>	51	23. <i>beatus</i>	23	40. <i>teneo</i>	17
7. <i>debeo</i>	48	24. <i>dies</i>	23	41. <i>christus</i>	16
8. <i>pauper</i>	41	25. <i>annus</i>	22	42. <i>invenio</i>	16
9. <i>volo (velle)</i>	41	26. <i>avaritia</i>	22	43. <i>justitia</i>	17
10. <i>villa</i>	40	27. <i>do</i>	22	44. <i>audio</i>	15
11. <i>dico (dicere)</i>	39	28. <i>advocatus</i>	21	45. <i>spiritus</i>	14
12. <i>dominus</i>	36	29. <i>capio</i>	21	46. <i>bonus</i>	14
13. <i>homo</i>	36	30. <i>aufero</i>	20	47. <i>martyr</i>	14
14. <i>possum</i>	35	31. <i>cogo</i>	20	48. <i>nox</i>	14
15. <i>multus</i>	32	32. <i>corpus</i>	20	49. <i>offero</i>	14
16. <i>video</i>	32	33. <i>marca.</i>	20	50. <i>tollo</i>	14
17. <i>magnus</i>	30	34. <i>solidus (-i)</i>	20		

Ces listes également mériteraient tout un commentaire, que nous ne pourrions faire figurer ici.

* Ont encore la fréquence 27 : *necessitas* et *praesum*.

A partir de ces mots significatifs les plus fréquents, il sera particulièrement intéressant de dégager la liste des mots-clés. Ceux-ci se distinguent par une fréquence qui s'écarte notablement de la fréquence normale pour une époque ou un genre littéraire donné. C'est dire qu'il faudrait disposer d'un *index* de fréquence général pour un ensemble d'œuvres contemporaines. Ces mots-clés caractériseront nettement une œuvre ou un auteur.

5) *Concordances.*

A partir d'un tel fichier, on peut constituer diverses concordances selon la nécessité des recherches — il suffit d'introduire ce fichier en mémoire ; ainsi une concordance qui livre au chercheur la liste de tous les mots nouveaux avec leur contexte, ce dernier étant, par exemple, toute la phrase ; ou encore une concordance de tel mot institutionnel, ou de telle notion philosophique ³⁰.

Faut-il souligner que les pages qui précèdent, n'ont évidemment pas épuisé toutes les formes de recherches lexicologiques possibles au moyen des méthodes mécanographiques. Notre exposé permet, croyons-nous, de se rendre compte de toutes les possibilités offertes.

b) AU POINT DE VUE GRAMMATICAL.

Si l'analyse philologique constitue un gros travail qui, ultérieurement, sera en grande partie allégé par une analyse automatique, nous ne pourrions assez insister sur l'extraordinaire rapidité que présente la suite des opérations.

Tous les renseignements qui figurent sur les cartes peuvent apparaître dans des relevés statistiques. A partir du fichier des *Epistulae*, dûment classé au préalable au point de vue grammatical — rappelons que ces tris se font à la cadence de 1.000 cartes à la minute — la trieuse statistique IBM 108, à laquelle est combinée une machine à écrire automatique 866, nous a permis d'obtenir, en une heure seulement, le comptage complet de

30. Le Laboratoire d'Analyse statistique des Langues anciennes a effectivement mis au point un programme d'ordinateur IBM 1620, qui permet l'établissement d'une concordance donnant pour chaque mot toute la phrase qui le renferme.

Nous discutons à la fin du présent article le problème des concordances automatiques réalisées au moyen d'un programme «KWIC». Cf. pp. 155-160.

toutes les particularités morphologiques ; pour les verbes, par exemple, l'on connaîtra pour l'indicatif présent de la première conjugaison, le nombre de premières personnes du singulier, de deuxièmes, etc., de la voix active, puis la somme du nombre d'indicatifs présents actifs ; puis de même pour la voix passive, la somme du nombre d'indicatifs présents passifs, puis la somme d'indicatifs présents, et ainsi de suite, pour toutes les conjugaisons pour tous les modes et tous les temps. Imagine-t-on le temps qu'il faudrait pour constituer, de façon traditionnelle, de pareils relevés ? De plus, il faut noter qu'à partir d'un fichier dûment vérifié, aucune erreur ne peut se glisser dans les relevés au cours de ces opérations ! Une erreur de manipulation apparaîtrait instantanément.

Nous nous proposons d'offrir au lecteur deux exemples de tableaux grammaticaux : les tableaux généraux de distribution des occurrences et du vocabulaire selon les catégories grammaticales, dont les citations et les noms propres sont exclus. Ces derniers, en effet, doivent être distingués dans l'étude grammaticale d'une œuvre médiévale ; sinon ils risquent de fausser certains résultats — ainsi le nombre de substantifs indéclinables paraîtrait particulièrement élevé, alors que des 90 substantifs indéclinables attestés dans les *Gesta*, 84 sont des noms propres (une telle constatation ne se vérifiera évidemment pas pour des œuvres antiques). Les noms propres font donc l'objet de tableaux particuliers. Dans la dernière colonne de droite de ces tableaux, nous avons noté les moyennes d'emploi. On calcule celles-ci en divisant le nombre d'occurrences par le nombre de mots différents, c'est-à-dire par le vocabulaire³¹. Si chaque occurrence représentait un mot différent, le quotient de cette division serait 1. Le texte est donc d'autant plus riche que sa moyenne d'emploi se rapproche de 1. Il faut évidemment être particulièrement vigilant quand on compare les moyennes d'emploi de textes de longueur différente ; dans un texte plus long, l'auteur sera en effet amené à se répéter plus souvent.

31. Cf. note 26 p. 144.

Gesta abbatum Trudonensium, I-VII

Distribution des occurrences et du vocabulaire selon les catégories grammaticales. Noms propres exclus.

Catégories	Nombre d'occurrences		Nombre de mots		Moyenne d'emploi
	Val. absol.	%	Val. absol.	%	
Substantifs	6 250	25,85	1 316	36,97	4,75
Adjectifs	1 819	7,52	581	16,32	3,13
Numéraux	384	1,59	59	1,66	6,51
Adjectifs-pronoms	3 734	15,45	60	1,69	62,23
Verbes	4 424	18,30	1 079	30,31	4,10
Adverbes	2 845	11,77	373	10,48	7,63
Prépositions	2 116	8,75	38	1,07	55,68
Conjonct. de coordination	2 046	8,46	23	0,65	88,96
Conjonct. de subordination	525	2,17	24	0,67	21,88
Interjections	32	0,13	7	0,20	4,57
Totaux	24 175	100,—	3 560	100,—	6,79
<i>Esse</i> employé comme auxiliaire	201		1		
	24 376*		3561**		

* Le tableau de distribution du vocabulaire donne un total de 24.367 occurrences. (cf. *supra* p. 144). La différence de 9 s'explique par *nequis* (1 occurrence) et *siquis* (8 occurrences) qui dans ce tableau grammatical, figurent dans deux catégories distinctes. *Nequis* et *siquis* apparaissent dans nos textes écrits en un seul mot, et Forcellini leur consacre un lemme spécial.

** Ce nombre concorde avec celui obtenu dans le tableau de distribution du vocabulaire (3566). En effet, dans le tableau des catégories grammaticales *nequis* et *siquis* ne sont pas comptés comme mots différents, étant donné que *ne*, *si* et *quis* indéfini sont attestés par ailleurs. De plus, dans le tableau de distribution du vocabulaire figurent 7 cas de sous-lemmes adjectifs-pronoms qui, d'un point de vue grammatical, ne doivent pas être distingués comme mots différents (donc 3561 + 2 + 7 = 3.570). D'autre part dans le tableau de distribution du vocabulaire *qualis*, *quantus*, *quisque* et *quot* ne sont comptés qu'une seule fois, alors qu'ils entrent dans deux catégories différentes — relatif et interrogatif ou relatif et indéfini — (donc 3566 + 4 = 3.570).

Epistulae

Distribution des occurrences et du vocabulaire selon les catégories grammaticales. Noms propres exclus.

Catégories	Nombre d'occurrences		Nombre de mots		Moyenne d'emploi
	Val. absol.	%	Val. absol.	%	
Substantifs	2 142	25,01	648	37,35	3,31
Adjectifs	622	7,26	218	12,56	2,85
Numéraux	216	2,52	45	2,59	4,80
Adjectifs-pronoms	1 398	16,32	46	2,65	30,39
Verbes	1 704	19,89	512	29,51	3,38
Adverbes	786	9,18	196	11,30	4,01
Prépositions	809	9,45	31	1,79	26,10
Conjunct. de coordination	671	7,83	18	1,04	37,28
Conjunct. de subordination	207	2,42	17	0,98	12,18
Interjections	10	0,12	4	0,23	2,50
Totaux	8 565	100,—	1 735	100,—	4,94
<i>Esse</i> employé comme auxiliaire	78		1		
	8 643*		1 736**		

L'examen de la distribution des occurrences selon les catégories grammaticales nous révèle que la mieux représentée est celle des substantifs. Elle fournit, dans les deux œuvres, un quart des occurrences. Les verbes, par contre, n'en représentent que 18,30 % dans les *Gesta* et 19,89 % dans les *Epistulae*. Le

* La différence de 3 avec le nombre obtenu au tableau de distribution du vocabulaire (cf. *supra* p. 144) s'explique par le fait que *nequis* (1 occurrence) et *siquis* (2 occurrences) apparaissent dans le tableau grammatical en deux catégories distinctes.

** Le tableau de distribution du vocabulaire donne un total de 1741 mots. A cause de l'absence d'autres attestations de *quis* indéfini, seul *siquis* a été retenu dans le tableau grammatical en vue du comptage des mots différents. Reste donc *nequis*. En outre 6 sous-lemmes adjectifs-pronoms figurent dans le tableau de distribution du vocabulaire (donc 1736 + 1 + 6 = 1743). D'autre part dans le tableau de distribution du vocabulaire, *quantus* et *quot* ne sont comptés qu'une seule fois, alors qu'ils entrent dans deux catégories différentes : les relatifs et les interrogatifs (donc 1741 + 2 = 1743).

pourcentage des adjectifs est 7 %. Le pourcentage des adverbes est élevé surtout dans les *Gesta*. Raoul a donc un style nominal.

Du point de vue du vocabulaire, le pourcentage des substantifs est aussi nettement supérieur à celui des verbes. Les adjectifs, qui ne représentent que relativement peu d'occurrences, atteignent un pourcentage fort élevé par rapport au vocabulaire total. Ils forment la catégorie grammaticale pour laquelle Raoul dispose proportionnellement du nombre le plus élevé de mots différents. Leur moyenne d'emploi est d'ailleurs de 3,73 dans les *Gesta*, où celle des substantifs est de 4,75 et celle des verbes 4,10 ; dans les *Epistulae*, elle est de 2,85, alors qu'elle est de 3,31 pour les substantifs et de 3,38 pour les verbes. Bien qu'on ne puisse guère comparer les moyennes d'emploi quand les textes ne sont pas de même étendue, nous en faisons néanmoins état afin de mieux mettre en lumière les divers rapports à l'intérieur de chaque œuvre.

L'importance des mots-outils est sensiblement la même dans les deux œuvres. Leur vocabulaire est proportionnellement plus riche dans les *Epistulae* que dans les *Gesta*. Il ne représente cependant qu'un très faible pourcentage par rapport au vocabulaire total.

Nos constatations font abstraction des noms propres. En les faisant intervenir, nous ne ferions qu'accentuer les caractéristiques que nous venons de dégager. Le pourcentage des substantifs est majoré, et le caractère nominal de la langue de Raoul s'en trouve accusé davantage encore. Grâce à cette distinction des noms propres, nous savons que nos pourcentages sont sûrs. Les noms propres, qui représentent dans les *Gesta* 4,25 % du nombre total d'occurrences et 5,92 % du vocabulaire total, et dans les *Epistulae*, respectivement 3,23 % et 7,81 %, n'influencent donc en rien nos conclusions.

Chacune des catégories grammaticales fait l'objet de tableaux détaillés. Ceux des verbes sont particulièrement développés. Les tableaux de distribution des verbes subordonnés selon le type de subordination, le mode et le temps, sont une véritable synthèse de la syntaxe de l'auteur.

Les tableaux des *Gesta* et ceux des *Epistulae* nous donnent fréquemment des pourcentages différents. Les tableaux des modes

impersonnels, notamment, font apparaître une différence marquée entre les deux œuvres : 16,36 % d'ablatifs absolus dans la première contre 8,8 % dans la seconde. Cette fréquence de l'ablatif absolu, est-elle la caractéristique d'une époque ou d'un genre littéraire donné, ou faut-il l'attribuer à la personnalité de l'auteur ? L'évolution de la langue est-elle plus marquée dans l'une des deux œuvres ? est-ce, au contraire, le genre littéraire, qui est à la source d'une telle différence ? Il faudra attendre d'autres études pour répondre avec certitude à ces questions. Actuellement, nous ne concluons qu'à une différence de style entre les *Gesta* et les *Epistulae*.

A partir des fiches des verbes classées dans l'ordre souhaité, l'on imprimera la liste de tous les verbes de propositions subordonnées selon le mode, le type de subordination et le temps, en les accompagnant de leurs références au texte. Toutes ces opérations demandent un minimum de temps.

c) AU POINT DE VUE STYLISTIQUE.

Nous ne nous attarderons pas sur les tableaux concernant l'ordre des mots et des propositions. Les principes de travail sont toujours les mêmes. Ces tableaux seront d'autant plus précis et détaillés que l'aura été l'analyse préalable ³². Pour les tableaux concernant l'ordre des propositions, on distingue soigneusement les divers types de propositions subordonnées.

Examinons quelque peu les longueurs de phrases. Elles ont été l'objet d'une attention toute particulière. Un tri sur la colonne, où est perforée l'indication de fin de phrase, rassemble tous les mots qui terminent une phrase ; comme chaque fiche a reçu automatiquement son numéro d'ordre dans la phrase, nous groupons de la sorte toutes les longueurs de phrases. On établira notamment un tableau de distribution des phrases selon leur longueur — tel nombre de phrases comportant tel nombre de mots. Le nombre moyen de mots par phrase est de 24 pour les *Gesta*, et de 18 pour les *Epistulae*. Si l'on élimine les citations, nous obtenons respectivement les longueurs moyennes de 25 et de 19 mots par phrase.

32. Cf. *supra* p. 132.

Au moyen de ces fiches, on constitue un « listing » comportant à chaque ligne le mot qui est en fin de phrase, la référence du texte, la longueur de la phrase que ce mot termine. Ces données peuvent naturellement être transposées sur des graphiques, notamment celui où les longueurs de phrases figurent suivant l'ordre du texte ; on y indiquera également la moyenne mobile c'est-à-dire la longueur moyenne calculée pour chaque phrase, en tenant compte du groupe qui précède et de celui qui suit chaque point considéré. On observe de la sorte qu'en général, chez Raoul, une phrase longue est entourée de phrases courtes — ainsi les deux phrases les plus longues des *Gesta*, comptant chacune 106 mots, ne font nullement monter la moyenne mobile. Le graphique des *Gesta* et celui des *Epistulae* ont un aspect très différent. Celui des *Gesta* a une physionomie en dents de scie qui s'accroît de plus en plus nettement à mesure que l'œuvre avance. La fin du livre V et le livre VI, qui raconte l'abbatiate tourmenté de l'abbé Thierry, marquent l'apogée de cette tendance. Au livre VII, dans le pittoresque récit du voyage de Raoul à Metz, on retrouve des différences de longueur généralement moins marquées.

Quant au graphique des *Epistulae*, on n'y rencontre nulle part des variations de longueur comparables à celles des *Gesta*. On distingue nettement par ailleurs les diverses lettres. Dans la première, se dessinent deux groupes de phrases particulièrement courtes : au début, Raoul cite saint Augustin et, vers la fin, il a recours à de brèves citations des Pères de l'Église. Au début, les moyennes mobiles sont inférieures à 10. Un autre endroit où apparaissent beaucoup de phrases courtes correspond précisément à un passage, où l'auteur paraphrase saint Augustin. Dans la troisième lettre, les moyennes mobiles ne subissent pas de variations notables. De fait, un grand nombre des griefs que Raoul y accumule, sont énoncés d'une façon identique. Une certaine monotonie dans l'énoncé de griefs répétés peut d'ailleurs être voulue, en vue de donner l'impression d'un fardeau sous lequel l'adversaire sera accablé. Dans les lettres courtes, la deuxième et la cinquième — celle adressée à Rupert de Deutz — on observe une longueur de phrase moyenne fort élevée. Dans la deuxième, qui ne compte en tout que cinq phrases, on en trouve

deux qui ont 70 et 87 mots ; la cinquième lettre s'achève par une phrase de 70 mots. La comparaison des deux graphiques montre combien le style des deux œuvres est différent.

Soulignons, une fois encore, que nous nous limitons ici à quelques exemples destinés à illustrer ces méthodes pour les lecteurs de l'*ALMA* ; les possibilités sont en effet proportionnelles à la multitude de renseignements qui figurent sur les cartes. En outre chaque type de renseignement est extrait du fichier en un minimum de temps.

* * *

Application des méthodes mécanographiques à un auteur médiéval : on voit que ce dernier y est « cerné » de toutes parts ! Rappelons-nous ici le jugement porté par Manitius sur la langue de Raoul³³ ? Il suffit de prêter attention au tableau général de la distribution des occurrences et du vocabulaire selon les catégories grammaticales, pour se rendre compte déjà que cette langue est loin d'être *ohne Wortschwall*. Sur quoi Manitius fondait-il son jugement ? Sur quelques impressions de lecture, probablement.

Si nous avons fourni au lecteur tous les tableaux concernant les *Gesta* et les *Epistulae*, il aurait constaté, ainsi que les exemples le suggèrent déjà, des différences considérables entre ces deux œuvres. Peut-être, au moyen âge, époque des arts poétiques et des *artes dictaminis*, la différence entre le style des divers genres littéraires se marque-t-elle souvent davantage qu'entre le style des écrivains. On entrevoit ce que de telles méthodes d'analyse apporteront à l'étude des genres littéraires et aux critiques d'authenticité.

On se plaît à rêver au moment où plusieurs œuvres ou auteurs médiévaux seront « cernés » de la sorte ! A vrai dire, il ne s'agit plus d'un rêve, mais d'un plan de travail, dont on peut prévoir le temps de réalisation³⁴.

33. Max MANITIUS, *Geschichte der lateinischen Literatur des Mittelalters*, dans *Handbuch der Altertumswissenschaft*, IX, 2, 3, München, 1931, p. 560.

34. Dès à présent, après avoir établi une édition nouvelle de la *continuatio prima* des *Gesta abbatum Trudonensium*, nous avons terminé l'*index verborum*, ed même que les diverses listes de fréquence et les relevés statistiques du livre

A la fin de cet aperçu des méthodes et des travaux réalisés à l'heure actuelle au Laboratoire d'Analyse statistique des Langues anciennes en matière médiévale, une question doit être clairement posée : n'est-il pas préférable de constituer par les méthodes mécanographiques des concordances plutôt que des *indices verborum* ?

Certes, il n'est guère concevable aujourd'hui d'envisager la constitution d'*indices* ou de concordances sans recourir aux moyens mécanographiques, au moment où les divers comités nationaux du *Novum Glossarium* désirent constituer des *thesauri* des textes médiolatins de leur pays respectif, pour élaborer ensuite un *thesaurus* général du latin médiéval.

Une discussion pour l'*index* contre la concordance, ou inversement, pour la concordance contre l'*index*, nous semble tout à fait vaine, parce qu'à partir des méthodes que nous avons employées, les possibilités sont extrêmement vastes.

La technique moderne peut nous fournir des *indices* ou des concordances, et de plus toutes sortes d'*indices* ou de concordances. Tout dépend du travail préalable ; tout dépend des méthodes suivies.

Confiera-t-on à la machine des formes nues ou des formes analysées ? Telle est la question primordiale. C'est entamer la discussion sur les concordances automatiques qui procèdent par formes nues, c'est-à-dire sans analyse philologique préalable.

Nos machines peuvent facilement et rapidement nous fournir de telles concordances, où chaque forme du texte sera encadrée de son contexte, à concurrence d'un certain nombre de signes alphabétiques prédéterminés par les capacités d'impression de l'ordinateur. Il suffit de perforer tous les mots d'un texte ; un programme « KWIC » (*keyword in context*) réalisera automati-

IX. Les livres VIII et X à XIII sont pour l'instant transcrits sur cartes perforées. On sait que cette *continuatio* est d'un auteur anonyme, qui a voulu insérer dans son œuvre un texte de Raoul, le livre IX. Nous ne pouvions en cet article introduire des exemples empruntés au livre IX, ni en publier la liste de fréquence des mots nouveaux ; il faut au préalable en faire la critique d'authenticité. Le *Mittelateinisches Wörterbuch*, dans son relevé des sources, a, parmi les œuvres de Raoul, maintenu un point d'interrogation à côté de la mention du livre IX. On ne saurait douter des éclaircissements nouveaux que nous apporteront de tels travaux pour l'ensemble de la *continuatio*. Nous comptons entreprendre de même la constitution d'*indices* et de listes de fréquence des grandes chroniques contemporaines.

quement la concordance. On obtiendra en fin de compte un « listing », où chaque forme figurera entourée d'un nombre fixe de signes alphabétiques.

Le volume de ces concordances sera matériellement énorme, puisque toutes les formes du texte y sont notées à raison d'une ligne par forme. Pour quantité de mots-outils, cette ligne sera complètement inutile ³⁵.

Dans une telle concordance, les formes seules sont classées dans l'ordre alphabétique : les différentes formes d'un même lemme ne sont donc pas nécessairement groupées ; cela constitue un grave inconvénient. Dans le cas de *tollo*, on trouvera certaines formes à la lettre *t*, et d'autres à la lettre *s*, et encore à des endroits différents ; dans le cas de *esse*, il faudra chercher aux lettres *e*, *f* et *s*. Inconvénient plus redoutable encore : les homographes ne sont pas distingués. Cette distinction demande en effet une analyse préalable.

D'autre part, le contexte sera, on le devine, très souvent insuffisant. Ce contexte insuffisant, on peut évidemment le doubler, mais cela correspond à doubler le volume, déjà tellement considérable. Un contexte même doublé ne lèvera pas toutes les équivoques.

Dans certains cas, il sera manifestement impossible de comprendre le texte ; dans d'autres — et ceci est particulièrement grave — on aura l'illusion de le comprendre. Le philologue se refusera à deviner le sens, et il sera fatalement amené à retourner au texte.

Prenons quelques exemples dans les textes de Raoul de Saint-Trond, en les citant selon l'édition de Koepke. Nous tenons compte du maximum de signes possible. La plupart des ordinateurs nous offrent un maximum de 132 signes par ligne ; une machine du type IBM 1443 peut en avoir 144 ³⁶. La référence aux *Monumenta Germaniae Historica* exige 7 signes (page et ligne). En

35. Les *Gesta abbatum Trudonensium, I-VII* comptent 971 et pour 26.270 occurrences. Pour les six millions de mots que comportent approximativement les textes médiolatins belges, il faudrait compter à proportion égale, 10 volumes de 400 pages, uniquement pour les *et*. Une étude lexicographique ne peut cependant éliminer les mots-outils.

36. N'insistons pas sur le peu de commodité d'un ouvrage d'une largeur telle qu'il compte 132 ou 144 signes par ligne ; un tel ouvrage n'est guère publiable cette constatation est également importante.

comptant d'après le maximum, c'est-à-dire 144 signes, il nous reste donc 137 signes pour chaque forme. Nos exemples en contiennent dès lors 137³⁷.

1) *dis promptuariis toto anni circulo liquata autumnii tempore ad tempus SEPONANTUR et nova musta naturali motu adhuc calentia et necdum pror* (228,36).

Il est impossible de s'assurer du sens de *seponere*, sans en connaître ni le sujet, ni le complément. Quant à *dis promptuariis*, le recours au texte nous apprend que *dis* est la finale de *gelidis*. Cet exemple illustre la difficulté que l'on rencontrera souvent, de s'assurer du sens d'un verbe, parce qu'on en ignore le sujet, ou le complément, ou les deux à la fois. On ne sera jamais sûr qu'il n'y a pas, dans le reste de la phrase, un élément qui modifie le sens.

2) *irmorum neque die habebat ignem neque nocte lumen nisi fratres in ea IACENTES undecumque sibi interdum etiam ex victu et vestitu suo ea* (255,50).

Quel est le sens de *iacentes*? Niermeyer signale pour *iacere* un sens spécial de « gîter, exercer le droit de gîte ». Ce sens convient-il ici? On ne peut le dire. Cet exemple montre que, normalement, il est impossible d'être certain du sens d'un verbe, quand on ignore le substantif, que remplace l'adjectif-pronom, qui en dépend.

3) *pecuniosus avarus ut aequanimius ferat cum audierit quod non vult. SUSCAEPTIONES igitur de aecclesiis et in aecclesiis maturarum et imm* (319,23).

Le contexte donné ne nous permet pas de connaître le sens de *susceptio*. Il en ira de même de beaucoup de mots qui commencent une phrase : le contexte qui précède, généralement, ne sert à rien ; celui qui suit, est nettement insuffisant³⁸.

4) Comment procédera-t-on dans le cas de formes périphrastiques aux membres disjoints, ou dont l'un des membres est sous-entendu ?

37. En réalité, d'autres indications, également nécessaires, restreindront le nombre de signes disponibles pour le contexte.

38. On pourrait faire figurer en début de ligne, les formes qui commencent une phrase. Mais à défaut d'une analyse philologique préalable du texte, l'on n'aura pas éliminé les cas de mauvaise ponctuation. Pour le moyen âge, nous ne disposons guère encore de bonnes éditions, qui évitent ces risques d'erreur.

5) Prenons le simple cas de *cum* : nous avons examiné les *cum* qui figurent dans les *Gesta, I-VII* : dans 19 % des emplois de *cum*, conjonction de subordination, le verbe n'apparaîtra pas dans le contexte cité ; on supposera, plus ou moins valablement, qu'il s'agit d'une conjonction de subordination. Dans 5 autres %, le *cum*, conjonction de subordination, est suivi directement d'un ablatif, et le contexte cité se comprend à merveille avec un *cum* préposition.

On voit par les exemples qui précèdent, qu'il faudrait procéder par phrase : le contexte devrait chaque fois être la phrase. C'est parfaitement possible. Mais examinons à titre d'exemple, les longueurs de phrases de la *Chronique* de Raoul. Les 137 signes correspondent à une vingtaine de mots ; or sur un total de 1118 phrases, 125 comptent de 26 à 30 mots ; 142, de 31 à 40 mots ; 93 de 41 à 50 mots ; 41 de 51 à 60 mots ; 12 de 61 à 70 mots, etc. Il y en a même 2 de 106 mots. Pour tous les *et*, et pour tous les mots-outils contenus dans ces phrases, toute la phrase sera chaque fois reproduite ! D'ailleurs, dans certains cas, la phrase n'y suffit pas.

Un travail lexicographique ne peut donc se faire sur la base d'une telle concordance. La probabilité d'erreur est trop élevée. Il serait extrêmement dangereux de faire un choix à partir d'une concordance, parce que celle-ci masque une série de faits que l'on ne peut découvrir que par un examen approfondi du texte.

Les exemples cités illustrent déjà l'impossibilité d'une analyse grammaticale correcte. Nous voudrions ajouter un exemple spécifiquement grammatical :

rat hac necessitate irrecuperabiliter venditus est silva in kircheym EXCISA et vendita substantia alia multum attenuata quod que magis do (261,52).

S'agit-il d'un indicatif parfait, ou d'un ablatif absolu ? Sans connaître l'analyse grammaticale correcte ³⁹, il est par ailleurs impossible d'analyser le style. D'ailleurs la seule compréhension d'un texte exige auparavant une analyse grammaticale correcte,

39. Au point de vue grammatical, on peut sans doute, dans certains cas, distinguer de quelle forme il s'agit ; mais on ne disposera pas de relevés numériques automatiques, que nous procure un fichier analysé.

car on ne consentira ni à deviner le sens, ni à le connaître plus ou moins vaguement.

Grâce aux concordances automatiques, on espère échapper à une analyse philologique préalable du texte, mais ces concordances ne pourraient servir de base à une étude lexicographique et grammaticale.

Jusqu'à présent, nous avons étudié les possibilités de compréhension d'un contexte tronqué, et nous avons nié cette possibilité comme base d'un travail scientifique. Il faut en outre attirer particulièrement l'attention sur de graves problèmes généraux, que ces concordances ne permettent pas de résoudre, ni même dans certains cas, de soupçonner. Ainsi en ira-t-il du problème des variantes, du problème des citations, qui sont souvent nombreuses dans les textes médiévaux. Il faut mentionner également les questions de ponctuation, qui exigent un examen attentif du texte. Pour combien de textes médiévaux, en effet, disposons-nous de bonnes éditions ?

La rapidité avec laquelle les machines mécanographiques constituent des concordances, est peut-être séduisante au premier abord ; en réalité, on ne peut échapper à une étude approfondie du texte. Nous écartons dès lors ces concordances, parce qu'elles sont fallacieuses, inutilisables dans certains cas, inutiles dans beaucoup d'autres. D'autre part, quand on a constitué une telle concordance, on peut sans doute entreprendre un autre travail, mais alors il faut tout recommencer, car nous n'avons pas de fichier de base, qui permette autre chose que le premier travail effectué.

Les méthodes qui consistent à confier à la machine non des formes nues, mais des formes analysées, évitent seules ces graves inconvénients. Elles requièrent comme base une analyse philologique approfondie du texte. Une fois cette analyse philologique terminée ⁴⁰ et transcrite sur des cartes perforées, on peut réaliser ce que l'on veut : *indices* de lemmes ou de formes, complets ou partiels, généraux ou spéciaux, *glossaires médiévaux*, concordances de tel mot institutionnel ou de telle notion

40. Rappelons que prochainement nous bénéficierons de l'analyse automatique. On peut considérer que 60 % environ des formes d'un texte seront analysées de la sorte.

philosophique, ou tout autre type de concordance — il suffit de mettre le texte en mémoire, et on le fera au moyen du fichier préconisé. Tout travail ultérieur reposera sur des bases scientifiques sûres. De plus, à côté des études lexicographiques, on disposera de relevés grammaticaux et stylistiques exhaustifs !

Ce qu'il faut mettre en vive lumière, est l'extraordinaire souplesse d'un tel fichier ; il se prête à toutes sortes de recherches ; il reste le document de base. On pourra constituer peu à peu une bibliothèque de fiches, une cartothèque, dont le spécialiste tirera n'importe quel renseignement ⁴¹.

Nous citions au début de notre article cette réflexion de MM. Pierre Demarne et Max Rouquerol, que « l'ordinateur est un fantastique manœuvre intellectuel ». Nous adjoignons au philologue ce manœuvre intellectuel ; ne croyons pas qu'on puisse sans grave dommage remplacer le philologue par ce manœuvre.

Le médiéviste certes ne consentira pas à limiter son champ de recherches par les concordances automatiques dont nous avons parlé, et à fournir de la sorte des instruments de travail peu maniables, insuffisants et fallacieux ; il optera au contraire pour un système aux possibilités multiples, qui permette l'analyse des textes sous tous leurs aspects.

L'absence d'*indices*, de relevés lexicographiques, grammaticaux et stylistiques englobant l'ensemble d'une œuvre, se fait cruellement sentir dans l'étude de la philologie et de la culture médiolatine. Notre désir est de remédier pour notre modeste part à ce lourd handicap, grâce aux moyens nouveaux que nous offre la technique moderne.

Liège.

Paul TOMBEUR

Chargé de Recherches du F.N.R.S.

41. Le philologue qui désire obtenir rapidement un relevé lexicographique d'un texte non dépouillé, pourra s'en tenir à une simple lemmatisation, et réaliser ultérieurement l'analyse grammaticale et stylistique approfondie. Il est toujours possible de compléter le fichier par après.