



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2006

---

## **L'écrit documentaire dans l'histoire linguistique de la France**

Glessgen, M-D

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-34181>

Book Section

Originally published at:

Glessgen, M-D (2006). L'écrit documentaire dans l'histoire linguistique de la France. In: Guyotjeannin, O. La langue des actes. Actes du XIe congrès de la Commission internationale de diplomatique. Paris: Éditions en ligne de l'École des chartes (ELEC), 1-18.

# École nationale des chartes

Elec n° 7

## La langue des actes

Actes du XI<sup>e</sup> Congrès international de diplomatique  
(Troyes, jeudi 11-samedi 13 septembre 2003)

Sous la direction de Olivier Guyotjeannin

## L'écrit documentaire dans l'histoire linguistique de la France

*Martin-D. GLESSGEN*

professeur à l'université de Zurich, Romanisches Seminar, Zürichbergstr. 8, CH-8032 Zurich ;  
[glessgen@rom.unizh.ch](mailto:glessgen@rom.unizh.ch)

### Résumé

L'article met en relief l'apport de l'histoire linguistique à l'histoire et en même temps l'importance de l'écrit documentaire en tant qu'ensemble de genres textuels placé à l'interface entre la diplomatique et la linguistique. (I) Il identifie les domaines spécifiques dans lesquels l'écrit documentaire peut servir de source pour la linguistique aux différentes époques du Moyen Âge et (II) il introduit aux études graphématiques (la "scriptologie") et lexicologiques sur cet ensemble textuel. (III) Enfin, il souligne la nécessité d'intégrer la méthodologie informatique dans l'édition et l'analyse de l'écrit documentaire malgré les contraintes notables qu'elle implique.

## I. Le cadre terminologique et méthodologique

### I.1. L'histoire linguistique entre linguistique et histoire

L'histoire linguistique est une discipline ambivalente : elle appartient par ses interrogations et sa méthodologie autant à la linguistique qu'à l'histoire. Néanmoins, elle a toujours été liée, d'un point de vue institutionnel, à la linguistique puisque sa vocation première est de contribuer à mieux comprendre le fonctionnement de la langue. Cela se vérifie depuis sa genèse au XIX<sup>e</sup> siècle, advenue d'ailleurs largement après celle de l'histoire et de la linguistique en tant que telles<sup>1</sup> : l'histoire linguistique était à cette époque un facteur constitutif dans la formation de la linguistique moderne alors qu'elle n'a pas joué un rôle particulier dans les transformations, tout aussi importantes, de l'historiographie.

Si l'apport de la linguistique historique aux études historiques est plus restreint que sa contribution à l'analyse linguistique, il est néanmoins réel. De manière générale, les caractéristiques internes de la langue et leurs variations selon l'espace et le contexte d'usage reflètent des phénomènes externes à la langue comme la culture matérielle et intellectuelle, la cohésion sociale ou l'identité des hommes. Grâce à l'intensité des liens entre les phénomènes

internes et les facteurs externes, la langue possède la qualité de fournir des indices imparables pour des données historiques qui la dépassent. Il n'est pas facile de faire parler la langue comme témoin de l'histoire puisqu'elle possède une dynamique propre très affirmée ; de telles analyses supposent donc un fondement méthodologique adéquat et une certaine prudence<sup>2</sup>.

Il existe divers domaines concrets où les interactions entre linguistique et histoire se font particulièrement intenses. La formation et l'élaboration des langues écrites interfèrent immédiatement avec d'autres transformations socioculturelles et politiques ; elles sont liées aux centres de l'écrit et par là elles sont proches du pouvoir politique. L'analyse des centres du pouvoir gagne en profondeur si elle prend en considération les aspects de leur médialité, concentrée notamment dans l'écrit. Un autre exemple, plus circonscrit, est aussi éclairant : les noms de lieu permettent de déceler des mouvements de population ou l'intensité de l'habitat à certains moments historiques.

Si l'histoire linguistique est largement sous-employée par l'histoire, cela ne se justifie pas par des raisons de contenu ; il s'agit plutôt d'un effet secondaire des frontières institutionnelles posées entre les disciplines. Le cas est particulièrement flagrant dans le cas de la romanistique et des langues galloromanes, car la *Romania* fournit un cadre d'observation de choix pour les interrogations historiques : grâce à une documentation unique, elle permet de suivre en détail la trajectoire entre le latin et ses langues-filles à travers deux millénaires et demi. La *Romania* médiévale dispose de textes variés pour le français, l'italien, l'espagnol, le groupe occitano-catalan et le gallégo-portugais. La comparaison entre les différentes langues néolatines permet de mieux comprendre l'évolution de chacune d'entre elles, même quand l'étude se concentre sur une seule langue.

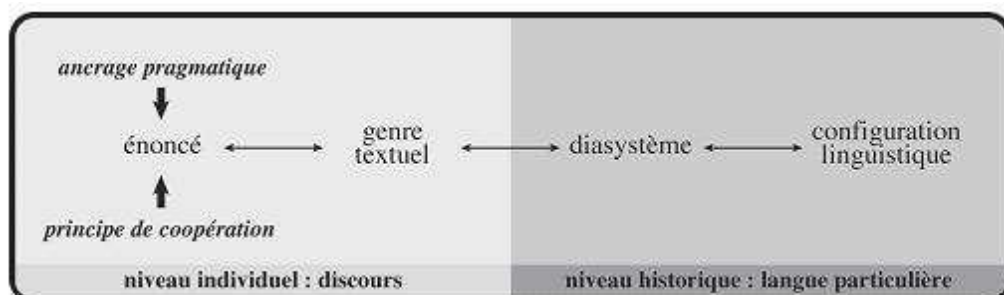
## **I.2. L'écrit documentaire parmi les genres textuels et dans le diasystème historique**

L'écrit documentaire représente l'un des trois grands ensembles textuels de l'Occident médiéval et moderne, avec les textes littéraires et les ouvrages d'un savoir spécialisé (médecine, droit, astronomie etc.). C'est l'ensemble le plus proche des préoccupations de l'historien, ce qui ressort immédiatement de sa solide présence dans la *Typologie des sources du Moyen Âge occidental* de Léopold Génicot. Mais il serait faux de l'étudier sans prendre en considération les autres genres textuels, liés à des domaines communicatifs et à des thématiques bien déterminés. Seule la comparaison des genres textuels peut déterminer des éléments de la langue qui reviennent dans toutes les situations et d'autres qui sont propres à un genre particulier. L'évolution générale du langage se produit par l'addition et l'interaction des

différentes traditions de l'écrit. L'écrit documentaire est partie prenante de ce processus d'élaboration linguistique<sup>3</sup> ; il reflète des évolutions enclenchées ailleurs et il introduit des innovations propres qui, partiellement, seront imitées dans d'autres genres.

Ces phénomènes ont une importance capitale pour l'histoire de la Romania et de l'Occident : entre la chute de l'Empire romain et l'époque actuelle, l'univers linguistique roman est déterminé par le passage du latin aux langues romanes, d'abord à l'oral, ensuite à l'écrit. La mise par écrit des langues romanes correspond en fait à un processus millénaire d'élaboration linguistique qui rythme toutes les transformations de détail dans les différents domaines de la langue. Dans ce processus, les éléments linguistiques qui sont propres à un ensemble de genres proches peuvent devenir caractéristiques pour certains contextes (= des caractéristiques 'diaphasiques' comme le langage juridique), pour le prestige des énoncés (= 'diastratique' — langage de haut ou de faible prestige), voire pour l'espace (= 'diatopique' — langage régional ou dialectal). L'ensemble de toutes les formes linguistiques existantes forment le 'diasystème' de la langue : la langue ne fonctionne pas comme une unité mais comme un ensemble de variétés parfois assez divergentes (= variétés 'diasystématiques'), réunies sous le 'toit' abstrait d'une langue historique<sup>4</sup>. Il est important de tenir compte de cette variabilité interne de la langue, parce qu'elle contient en même temps des informations précieuses pour l'histoire : tout usage définissable en termes diasystématiques peut être interprété à des fins socio-culturelles.

La place des genres textuels dans le fonctionnement et le changement linguistique est réelle. Le schéma suivant (tableau n° 1) met en relief les interactions qui existent entre (1) les énoncés concrets et individuels, (2) les genres, (3) le diasystème et (4) les règles linguistiques plus abstraites (= la 'configuration' ou le 'système' linguistique)<sup>5</sup> :



Les deux paramètres de l'élaboration linguistique et de la variation diasystématique permettent de donner une ligne interprétative générale à l'évolution de la langue des 'actes' en tant qu'ensemble de genres textuels. Les préoccupations de l'histoire linguistique et de la diplomatique se rejoignent dans l'attention qu'elles accordent toutes deux aux genres textuels

et à leurs caractéristiques. Leurs interrogations me semblent compatibles, même si elles ne sont pas identiques. Les interrogations de l'histoire linguistique se placent toujours dans l'optique de l'évolution générale de la langue. Il s'agit de savoir quelle a été la part de l'écrit documentaire dans l'élaboration linguistique d'une langue donnée à une époque donnée. Pour évaluer les qualités communicatives d'un texte documentaire donné, il s'agira en revanche de déterminer son degré d'élaboration dans un domaine de langue défini, et d'établir la distance linguistique qui le sépare d'autres textes et d'autres genres.

Notre aperçu devra tenir compte des sérieuses lacunes de la recherche sur les genres textuels et, notamment, sur les genres documentaires. L'historiographie linguistique a en effet compris depuis longtemps l'importance des genres pour l'évolution de la langue, mais elle est loin d'avoir mené ce raisonnement à son terme. Nous ne disposons pas d'une typologie linguistique des genres textuels semblable à celle de Génicot. La *Typologie des sources*, quant à elle, ne peut nous servir de guide sûr, puisque les segmentations dans la langue sont plus larges que les genres textuels de l'histoire : une charte et une lettre se distinguent d'un point de vue diplomatique, mais elles peuvent appartenir, d'un point de vue linguistique, à un même ensemble variationnel, caractérisé par le même vocabulaire et une même syntaxe. À l'heure actuelle, nous n'avons pas d'idée précise sur les segmentations linguistiques entre les différents genres textuels, ce qui s'explique aussi par le fait que la linguistique s'est presque toujours basée sur les seules traditions littéraires pour étudier l'évolution des langues.

En effet, la tentative globalisante la plus ambitieuse menée pour structurer des textes anciens, le *Grundriß der romanischen Literaturen des Mittelalters* (GRLMA)<sup>6</sup>, prenait comme point de départ les traditions littéraires, même s'il les a dépassées par la suite<sup>7</sup>. Le plus récent *Inventaire systématique des premiers documents des langues romanes*<sup>8</sup> repose sur le *Grundriß* pour les genres traités mais il entend répertorier l'intégralité des documents originaux antérieurs à 1250 en les catégorisant d'après leurs genres textuels, puis d'après leur langue. Cet ouvrage bien conçu souffre malheureusement d'une réalisation philologiquement insatisfaisante<sup>9</sup> : la partie sur les sources documentaires<sup>10</sup> – complémentaire au *Grundriß* – est lacunaire et mal structurée<sup>11</sup>, la datation et la localisation n'ont pas fait l'objet d'une analyse critique, et l'identification des variétés linguistiques en question aurait mérité une réflexion plus approfondie<sup>12</sup>.

C'est donc sur une base encore partielle et d'un point de vue nécessairement personnel que je présenterai d'abord, d'un point de vue externe, le rôle de l'écrit documentaire parmi les différents genres textuels pendant les deux grandes périodes de l'histoire linguistique de la Romania (§ II : genèse des

langues romanes, passage du latin aux langues romanes à l'écrit). La réflexion consécutive sur les caractéristiques linguistiques internes de l'écrit documentaire repose — en l'absence d'études monographiques d'ensemble — sur un corpus empirique concret circonscrit, à savoir les documents français de Lorraine étudiés dans le cadre des *Plus anciens documents linguistiques de la France* (§ III : scriptologie, élaboration lexicale, syntaxe, onomastique). Je prendrai appui sur ce même projet, mené en collaboration avec l'École nationale des chartes (Françoise Vielliard, Olivier Guyotjeannin), pour étudier en conclusion les questions méthodologiques soulevées par l'informatique dans l'édition et l'analyse des textes anciens ; les possibilités informatiques, toutes récentes, ont en effet transformé en profondeur les perspectives et la pratique en philologie et nécessitent une attention particulière (§ IV : encodage textuel, analyse linguistique par l'informatique).

## **II. L'écrit documentaire dans les grandes époques de l'histoire linguistique de la Romania**

### **II.1. La genèse des langues romanes**

L'histoire linguistique de la Romania connaît deux grandes périodes qui correspondent approximativement au premier et au deuxième millénaires :

- (1) la formation des différentes langues romanes à l'oral, qui s'éloignent progressivement du latin parlé tardif, et
- (2) la formation des langues romanes écrites selon un processus d'élaboration plus ou moins développé suivant les pays.

La première transformation a déjà ses racines dans le latin parlé des I<sup>er</sup> ou II<sup>e</sup> siècles, mais s'intensifie après la chute de l'Empire d'Occident, notamment entre le milieu du VII<sup>e</sup> et le milieu du VIII<sup>e</sup> siècle, lorsque l'écrit latin est devenu trop faible pour contenir les divergences à l'oral<sup>13</sup>. Vers l'an 1000, tous les idiomes romans existent dans le sens où ils ont acquis une première autonomie interne dans leurs caractéristiques phonétiques, grammaticales et lexicales.

Mais il s'agit encore de langues purement orales : tous les témoignages écrits de type roman, de l'Andalousie à la Wallonie et à la Terre d'Otrante, n'atteignent pas, avant l'an 1000, le nombre de trente et sont souvent de la longueur d'une seule ligne. La transformation d'une langue de départ en une multitude de langues à l'arrivée n'est observable qu'à travers des sources écrites en latin.

Le latin des VI<sup>e</sup>-X<sup>e</sup> siècles reflète en effet certaines innovations romanes dans les graphies, dans le lexique et dans la grammaire, qui se concentrent

très curieusement dans les formes toponymiques ; en ligne générale, la *scripta latina rustica*, plus proche des langues romanes parlées<sup>14</sup>, apparaît notamment dans des énumérations ou dans des listes, formes écrites dont les fonctions communicatives se rapprochent de l'oral<sup>15</sup>. Cette variété intermédiaire et artificielle du latin crée un terrain propice à l'apparition de noms propres romans dont les liens avec le contexte extralinguistique des documents — et donc la charge pragmatique — sont particulièrement développés. Ils nous fournissent bien avant les autres éléments textuels des informations sur la différenciation géolinguistique des langues romanes, sur leur lexique courant et les types dérivationnels en usage, sur la chronologie des évolutions phonétiques de même que sur certaines marques et transformations grammaticales. Leur richesse est très loin d'avoir été épuisée, comme le montrent les études récentes de Jean-Pierre Chambon<sup>16</sup> et les publications du groupe d'onomastique sarrebruckois, instauré par le germaniste Wolfgang Haubrichs ; ce centre de recherches a mis en relief l'intérêt des noms de lieu dans la zone de frontière germano-romane pour l'histoire des langues et du peuplement<sup>17</sup> (cf. aussi *infra* § III.3.).

Étant donné le petit nombre de productions littéraires de cette époque, l'écrit documentaire représente à quelques rares exceptions près, le seul ensemble de sources disponible pour l'histoire linguistique<sup>18</sup>. Toute collection de documents ou tout relevé de données, notamment de noms propres, proposés par des historiens peuvent donc intéresser l'histoire linguistique. Malheureusement, la collection des *Chartae latinae antiquiores* s'est avérée jusqu'ici d'un intérêt plus faible pour la romanistique que, par exemple, le *Corpus Inscriptionum Latinarum* ; cela peut être dû à une exploitation linguistique insuffisante du corpus comme – jusqu'ici – pour les *Monumenta palaeographica Medii Aevi*. Souvent, nos données proviennent de cartulaires plus récents, exploités avec succès si l'on gère avec prudence les transformations potentiellement survenues dans les copies<sup>19</sup>. L'importante base de données textuelles des chartes originales antérieures à 1121 conservées en France, en cours de constitution à l'Artem, à Nancy, ou son équivalent pour la Belgique, publié sur cd-rom<sup>20</sup>, n'ont pas encore été étudiés d'un point de vue linguistique, exception faite du “vocabulaire de la souscription”, analysé par l'historien et diplomate Benoît-Michel Tock dans sa thèse d'habilitation<sup>21</sup>. Il en ressort notamment une grande variété dans ce vocabulaire spécialisé, ce qui laisse supposer des phénomènes de variation inattendus à cette époque en latin.

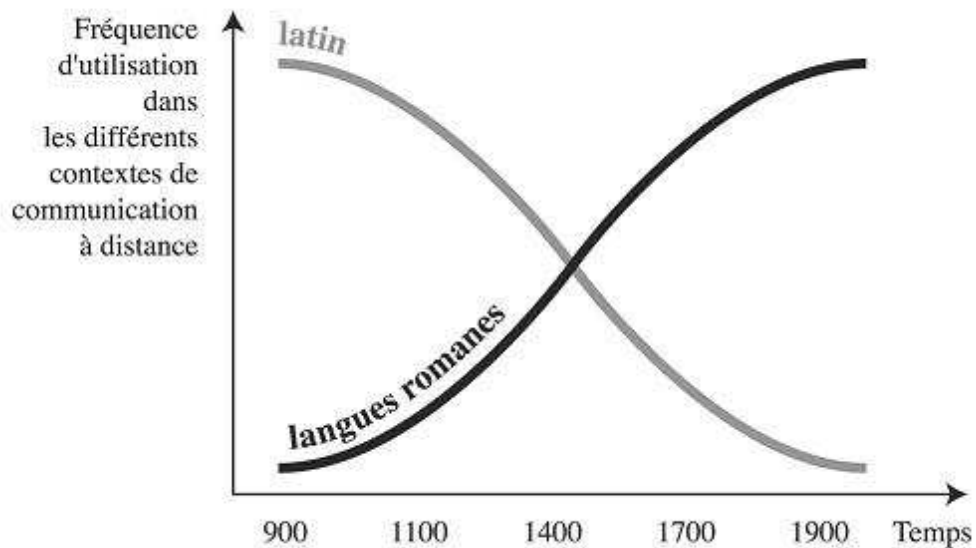
## II.2. Le passage du latin aux langues romanes à l'écrit

Le deuxième millénaire s'ouvre sur la mise à l'écrit et la standardisation des langues romanes, qui se substituent peu à peu au latin. Si le premier millénaire de l'histoire linguistique de la Romania est caractérisé par un oral

peu accessible, le second millénaire voit s'imposer l'écrit, ce qui donne plus d'aisance au romaniste.

Dans leur formation comme langues écrites, les idiomes romans ont pris modèle sur le latin qu'ils remplaçaient dans un processus d'acculturation. La substitution proprement dite s'est échelonnée à travers tout le millénaire : au début du XI<sup>e</sup> siècle, les variétés romanes étaient déjà le seul moyen de communication pour tout échange de proximité linguistique comme les conversations quotidiennes, amicales ou intimes ; le latin au contraire restait le seul outil de communication à distance, donc l'ensemble de l'écrit et, à l'oral, les registres formels et les énoncés de prestige. Le tableau suivant indique cette position du latin, en haut à gauche.

Au XIX<sup>e</sup> siècle, les langues vernaculaires ont pris la place des variétés de prestige et de distance, et reléguèrent le latin à une position marginale. La substitution du latin par les langues romanes est un processus linéaire qui traverse le deuxième millénaire, comme le souligne encore le tableau n° 2<sup>22</sup> :



Plus précisément, le processus de substitution s'est opéré en fonction des contextes communicatifs et des genres textuels. Selon les langues, les différents genres textuels témoignent pourtant d'évolutions décalées dans le temps : l'occitan connaît dès le XII<sup>e</sup> siècle des textes juridiques en grand nombre et une poésie profane, celle des troubadours, alors que le français fait encore appel, à la même époque, au latin pour les genres juridiques et administratifs, tout en développant le roman en vers ou en prose, rare en occitan.

D'autres langues montrent des retards plus importants : le francoprovençal et le romanche ou encore — dans l'Orient grecophone — le roumain n'ont



pratiquement pas d'existence à l'écrit avant le XVI<sup>e</sup> siècle. Le sarde reste pendant tout le Moyen Âge limité aux seules sources documentaires. Comme cela a été dit (§ I.1), la Romania médiévale se restreint donc dans les textes au français, à l'italien et à l'espagnol, au groupe occitano-catalan, encore peu différencié, et au gallégo-portugais, encore indistinct.

Mais malgré ces décalages, parfois importants, et malgré la variation dans les situations concrètes, le changement de codes s'est opéré dans toutes les langues et dans tous les cas de figure, à l'intérieur des différents genres textuels ; c'est l'addition des genres qui produit l'évolution générale.

D'un point de vue linguistique, le développement des genres textuels dépasse même les langues<sup>23</sup> : les formes de construction d'une charte, y inclus son vocabulaire et sa syntaxe, montrent des ressemblances fortes à travers toute l'Europe occidentale ; elles suivent partout le modèle du latin et elles répondent partout aux mêmes besoins de communication. Le processus de transformation appartient donc plus aux genres qu'aux langues : les études détaillées sur le passage du latin à une langue vernaculaire dans des contextes définis permettront ensuite de déterminer avec plus de précision la place de l'écrit documentaire dans une transformation générale des langues<sup>24</sup>.

Même si le passage du latin au français ou à l'occitan s'inscrit dans une logique bien déterminée, dans chaque situation historique concrète, le choix de l'une de ces trois langues à l'écrit peut être significatif. Faire appel au français à un moment où son usage n'était pas encore une nécessité implique toujours un sens pragmatique sous-jacent. Prenons un exemple : dans la deuxième moitié du XIII<sup>e</sup> siècle, l'on peut supposer que, très approximativement, un septième des chartes sur le territoire d'oïl était écrit en français, les autres en latin<sup>25</sup>. Or, un sondage que j'ai effectué sur les chartes françaises de l'évêque de Toul montre que celles-ci jouent un rôle politique primordial dans l'opposition avec le pouvoir ducal<sup>26</sup>. À l'intérieur de ce scriptorium, le choix du français doit être interprété dans l'optique de la gestion d'un pouvoir territorial et, plus généralement, dans celle de la mise en scène d'un pouvoir politique. Il est donc probable que l'étude des chartes françaises de cette époque revient à étudier un ensemble choisi de documents dotés d'une importance communicative particulière.

Pour les époques médiévale et — encore plus — moderne, l'écrit documentaire perd sa position de monopole pour l'histoire linguistique ; les genres littéraires et les textes d'un savoir spécialisé se développent de plus en plus, et le monde contemporain connaît une diversification extrême de l'écrit dans laquelle la notion même d'écrit documentaire se dissout. L'importance quantitative de l'écrit documentaire reste néanmoins écrasante jusqu'au

XVIII<sup>e</sup> siècle : pour la Lorraine médiévale — exemple parmi tant d'autres —, il existe peut-être cinquante fois plus de pages de textes documentaires que d'autres écrits, tous autres genres textuels confondus<sup>27</sup>. Il est toutefois difficile d'évaluer l'action des écrits documentaires en tant que tels sur d'autres genres textuels. L'immense majorité des textes administratifs ou juridiques n'ont jamais eu une réelle influence linguistique : les textes étaient enregistrés et archivés. Sans vouloir diminuer l'intérêt de la recherche historique sur la vie des documents après la promulgation, l'histoire linguistique considère actuellement que, en termes de communication, une partie significative de la production documentaire entrait dans une voie sans issue ; en échange, il est clair que ces textes reflètent des usages linguistiques bien ancrés et peuvent servir de témoins de leur temps.

Les interactions entre les différentes traditions de l'écrit sont plus immédiates dans les lieux de production : les chancelleries et scriptoria qui rédigeaient des actes produisaient aussi de nombreux textes d'autre nature et formaient des scribes qui se spécialisaient, par exemple, dans la copie d'ouvrages littéraires, religieux ou scientifiques. L'histoire linguistique devrait donc étudier plus qu'elle ne l'a fait dans le passé les processus de production des documents. L'histoire peut l'aider en identifiant plus précisément les lieux et conditions de production : en quels lieux combien de scribes produisaient quels types de textes en quel nombre et pour quel usage ? La linguistique pourra alors ajouter d'autres informations, par exemple : dans quelle mesure le langage écrit dans un scriptorium donné varie-t-il selon les genres textuels ? Nous avons des raisons de croire que l'écrit documentaire reflète plutôt la régionalité linguistique de son lieu de genèse — nous y reviendrons au prochain paragraphe — alors que les autres genres peuvent laisser paraître plus clairement l'origine personnelle des scribes.

En l'état actuel de nos connaissances sur les caractéristiques internes et externes de l'écrit documentaire, il est tout aussi impossible de proposer des segmentations linguistiques par sous-genres que d'évaluer leur impact réel sur la formation des langues standard modernes. Mais il est inconcevable que ces masses textuelles et leurs rédacteurs n'aient pas eu une importance déterminante dans l'élaboration et donc dans l'évolution linguistique générales.

### **III. La forme linguistique de l'écrit documentaire**

#### **III.1. L'étude de la variation graphématique et morphologique dans l'écrit documentaire**

Une langue définie dispose à un moment historique donné d'un stock important de formes invariables et identiques dans tout contexte et en tout

lieu. Le processus d'élaboration augmente considérablement ce stock et mène à une neutralisation notamment des variétés diatopiques. Pour les langues romanes comme pour les autres langues de l'Occident, la mise en place de telles 'formes standard' couvre tout le Moyen Âge jusqu'au XVI<sup>e</sup> siècle. D'innombrables pages écrites étaient nécessaires pour rapprocher les diverses formes diatopiques, pour harmoniser les variétés diastratiques et diaphasiques et pour produire cette homogénéité linguistique qui nous est, aujourd'hui, tout naturelle.

Seuls l'espagnol et le gallégo-portugais ont connu une standardisation précoce, dès le XIII<sup>e</sup> siècle, suite aux brassages de population lors de la *Reconquista* ; les neutralisations linguistiques fortes à l'oral qui en découlaient facilitaient la mise par écrit de ces langues. Notamment l'espagnol disposa très tôt, sous l'impulsion de la chancellerie royale d'Alfonse le Sage, d'une forme graphématique neutre sur tout le territoire ; par conséquent, si l'ancien espagnol connaît une forte variation dans les marques morphologiques, celle-ci ne peut pas être interrogée en terme de variation régionale.

Partout ailleurs dans la Romania, les variations graphématique et morphologique divergent d'un endroit à l'autre, ce qui permet d'étudier les divergences régionales et de les utiliser à d'autres fins, entre autres socio-historiques. Cette entreprise est suffisamment porteuse pour avoir mené à la formation d'une discipline à part entière, la scriptologie. Celle-ci relève et interprète les phénomènes graphiques, leur relation avec les données phonétiques sous-jacentes ainsi que des éléments morphologiques. Des scientifiques comme C. Th. Gossen (*Französische Skriptastudien*), Antonij Dees (*Atlas des chartes originales du XIII<sup>e</sup> siècle*), Hans Goebel (*Normandische Urkundensprache, Dialektometrische Darstellung von Dees' Sprachatlas*) ou Max Pfister ont contribué, en romanistique, à l'élaboration de cette discipline<sup>28</sup>.

Le rôle primordial de l'écrit documentaire pour la scriptologie réside dans deux qualités, absentes des autres traditions textuelles : la qualité d'original, fréquente dans l'écrit documentaire, et la présence d'éléments externes qui permettent une datation précise et donnent des indices sur le lieu de genèse des documents. H. Goebel a prouvé, il y a déjà vingt ans, que toute copie, même de textes documentaires, produit une neutralisation des caractéristiques linguistiques, un effet 'tâche d'huile'<sup>29</sup> : un texte original est toujours plus marqué régionalement qu'un texte copié, ce qui rend plus difficile l'étude des sources littéraires ou celles d'un savoir spécialisé, normalement transmises par des copies.

L'autre qualité est encore plus importante, parce qu'elle permet de

rapprocher les énoncés écrits de leurs lieux de production, des chancelleries et des scriptoria. Ceux-ci représentent les entités de base pour toute réflexion d'histoire linguistique médiévale : la langue écrite s'est formée et transformée très concrètement dans le réseau de ces 'lieux d'écriture' dont l'identification a par conséquent une importance capitale.

L'attribution d'un document à une chancellerie précise est néanmoins une opération délicate qui ne peut pas se valoir, chez les linguistes, d'une grande tradition d'études ; la scriptologie entendait plutôt identifier des régions d'origine que des 'lieux' précis.

L'identification d'un lieu d'écriture — entendu non pas dans l'espace mais dans la relation à un mandataire donné — suppose une combinaison de caractéristiques externes des documents (écriture, mise en page) et de caractéristiques linguistiques internes (graphies, morphèmes). Les premières sont difficiles à décrire de façon systématique et opératoire mais donnent des indices précieux, les secondes au contraire permettent des quantifications solides.

Des questions semblables ont été traitées récemment dans le cadre des projets de Trèves sur le langage des chartes en zone frontalière franco-allemande<sup>30</sup>. La thèse romaniste de Harald Völker sur *Skripta und Variation* a montré par exemple la force d'innovation de la chancellerie royale dans le choix des formes graphiques et morphologiques<sup>31</sup>. Pour un corpus lorrain des *Plus anciens documents linguistiques de la France*<sup>32</sup>, j'ai pu différencier tous les lieux d'écriture d'une réelle importance grâce au regroupement d'éléments externes et de plusieurs phénomènes graphiques et morphologiques<sup>33</sup> ; parallèlement, l'identification des lieux d'écriture permet de relever des caractéristiques linguistiques qui leur sont propres :

- la chancellerie des comtes de Bar, plus avancée que les autres, fait appel à des formes de prestige d'origine extrarégionale et parfois latinisantes, en accord avec le français central (comme *estable* contre la forme régionale lorraine *estaule*, l'article masculin régime *le* au lieu de *lo*, la forme *lettres* au lieu de *lattres*) ou le picard (comme *avera* contre la forme régionale *avra*) ; en outre, elle montre un plus haut degré d'homogénéité des graphies et des morphèmes ;

- les scribes indépendants forment un ensemble de 'lieux d'écriture' non constitués ; ils montrent un comportement linguistique très semblable à travers tout le territoire lorrain qui s'inscrit dans une norme de faible prestige dans le diasystème de l'écrit, une norme de 'proximité linguistique' ; ils forment une antipode polycéphale à la chancellerie des comtes de Bar par la présence notable de graphies

régionales (*estaule, avra, lattres*, l'article régime *lo*, le pronom démonstratif *ceas*) ;

- la chancellerie des ducs de Lorraine, occupe une position contradictoire, caractérisée par des alternances entre formes extrarégionales et régionales (*avera* et *avra*, *estable* et *estaule*), mais tendant, elle aussi, à l'uniformisation (*le, ceus*) et excluant des formes vraiment locales ;
- les *scriptoria* ecclésiastiques sont caractérisés par une attitude mélangée, en intégrant des formes d'origine et de prestige différents ; ils reflètent le mieux la variance de l'écrit lorrain.

Ces études montrent qu'il existe une dimension intermédiaire entre le scribe isolé et la *scripta* très abstraite d'une région. Les paramètres scriptologiques qu'avait appliqués Antonij Dees dans ses *Atlas* sur des bases parfois flottantes sont désormais intégrés à une méthodologie rigoureuse. Interroger la variation graphématique et morphologique permet en même temps de déceler les cheminements concrets d'élaboration et de neutralisation linguistiques et d'en identifier les acteurs et leur rôle. La simplicité de quantification des éléments graphématiques permet des constats qui sont aussi valables pour d'autres évolutions linguistiques ou historiques<sup>34</sup>. Ces recherches mettent en relief les 'lieux d'écriture' comme des entités à part entière ayant eu une action propre dans l'histoire de la langue écrite et dans tout ce qu'elle implique pour l'histoire politique et culturelle.

## III.2. L'étude de l'élaboration lexicale

### III.2.1. L'exemple du vocabulaire juridico-administratif : l'ancrage étymologique et diasystématique

L'élaboration d'un lexique spécialisé est sans doute l'élément linguistique le plus saillant de l'écrit documentaire. Les grands ensembles de termes juridiques et administratifs ainsi que de tournures figées ou semi-figées préfigurent les terminologies contemporaines et demandent en même temps un solide effort de formation pour la bonne compréhension des textes. L'histoire linguistique est tout autant concernée par la formation de ce vocabulaire que la diplomatique : le lexique véhicule des marques diaphasiques, liées à un domaine d'usage de la langue et reconnaissables comme telles par tous les locuteurs de l'époque.

La formation de ce vocabulaire fait appel aux mécanismes habituels de la formation des mots<sup>35</sup>, mais présente certaines particularités :

- les emprunts au latin sont très fréquents, surtout par emprunt direct, plus rarement par emprunt sémantique. Parallèlement aux emprunts véritables, le sens d'un terme latin peut être repris par un terme vernaculaire sans lien étymologique ;

- les changements sémantiques s'opèrent surtout à partir du vocabulaire déjà constitué : pour la plupart, il s'agit de spécialisations de sens par métonymie (*genus pro species*) ; l'intégration de termes provenant d'autres domaines sémantiques, par métaphore, est plutôt rare. Cette distribution entre métonymie, fréquente, et métaphore, épisodique, correspond aux habitudes générales de la langue, mais elle est encore accentuée ici ;
- les changements dérivationnels sont très fréquents, notamment entre noms et verbes ; l'ancien français fait souvent appel à la conversion, donc au changement de catégorie lexicale sans changement formel, ce qui complique parfois la compréhension des textes ;
- enfin, la formation de phraséologismes ou de formules figées est extrêmement fréquente dans l'écrit documentaire ; l'habitude des doublets synonymiques (*je donne et octroye, j'ai rendu et acquitté*) augmente encore la présence d'éléments préfabriqués dans ces textes<sup>36</sup>.

Mais s'interroger sur les mécanismes de formation du vocabulaire technique à partir d'un texte isolé est relativement décevant : une charte lorraine de 1427 ne contient normalement aucun terme spécialisé qui ne se trouve pas déjà dans des chartes de 1340 ou même dans les premières chartes du XIII<sup>e</sup> siècle, héritières directes des chartes latines. Les innovations lexicales individuelles sont bien plus fréquentes dans des textes littéraires, ce qui a sans doute contribué à favoriser l'analyse de ceux-ci dans l'histoire linguistique.

En contrepartie, il est toujours utile d'appliquer aux termes ou phrasèmes d'un texte documentaire donné les trois interrogations suivantes qui mettent en relief leur ancrage diasystématique :

(1) Quelle est la trajectoire étymologique du terme : s'agit-il d'un terme nouveau ou traditionnel, bien attesté avant et après dans les textes ? Dans le premier cas, il peut s'agir d'un néologisme — donc une formation relativement récente au moment de la rédaction du texte — ou d'une lacune dans la documentation ; de telles lacunes de la recherche sont fréquentes malgré l'intensité des recherches sur les premières attestations des mots, toujours pour la même raison que les textes documentaires n'ont pas été étudiés aussi intensément que les autres genres textuels. Dans le cas où le terme ou le phrasème est bien attesté avant et après, son utilisation dans le texte est diachroniquement neutre. L'intérêt étymologique d'un tel terme est alors faible. Cette observation vaut également pour des termes d'une haute spécificité sémantique qui posent de grandes difficultés de compréhension à

un lecteur non initié, mais qui sont courants à leur époque et dans leur contexte.

L'occurrence concrète dans un texte donne toutefois des indications sur l'usage du mot et elle peut intervenir dans des calculs de fréquence d'usage pour des ensembles textuels plus importants. Par ailleurs, elle peut avoir un intérêt diaphasique comme je l'expliquerai par la suite.

(2) Parmi les termes bien attestés, diachroniquement neutres ou 'isochrones', il faut déterminer le degré de spécificité diaphasique : s'agit-il d'un mot du français général ou d'un terme spécialisé de l'administration ou du droit privé, donc d'un langage diaphasiquement marqué ? Cette question importante ne peut être traitée pour des raisons pratiques qu'après avoir considéré la trajectoire étymologique et les attestations du terme. L'on peut distinguer trois cas de figure, qui s'inscrivent dans un continuum qui a comme pôles extrêmes le langage général d'un côté et le langage diaphasiquement marqué de l'autre.

(i) Les mots du vocabulaire général peuvent tout à fait appartenir aux domaines sémantiques du monde juridico-administratif et être particulièrement fréquents dans les textes documentaires, mais ils sont en même temps intégrés dans le vocabulaire de la vie quotidienne et apparaissent tout aussi bien dans des textes non-documentaires. C'est le cas, pour notre corpus, des substantifs masculins *ban*, *gage* et *serment* ou de la locution adverbiale *estre en aage* "être majeur". L'usage fréquent de ces termes est une des contributions de l'écrit documentaire à l'ancrage d'un vocabulaire de base élargi dans une langue en voie d'élaboration.

(ii) Le vrai lexique spécialisé au contraire, placé de l'autre côté du continuum, supposait déjà à son époque des connaissances techniques ; son étrangeté pour les non-spécialistes justifie sa marque diaphasique comme 'langage juridique' ; cela vaut pour des termes comme les noms *entrepris(e)res* n.f.pl. "entreprises injustifiées, torts", *eschoite* n.f. "succession", *mortesmanie* n.f. "main morte" et *usuaire* n.m. "droit d'usage", le syntagme *letres pendans* pl. "acte scellé d'un sceau pendant" ou le verbe *aquiter* v.tr. "céder (une obligation, un cens)"<sup>37</sup>.

(iii) D'autres termes juridiques encore ont été forgés à l'intérieur de l'écrit documentaire, mais le dépassent dans l'usage, comme *bestens* n.m. "querelle, dispute", *concord* n.m. "accord", *coustange* n.f. "frais, dépenses", *creante* n.m./f. "approbation, consentement" ou *parrochage* n.m. "territoire d'une paroisse"<sup>38</sup>.

La distinction entre mots généraux et mots techniques permet de mesurer, pour l'ensemble d'un texte, son degré de technicité et d'évaluer par

conséquent sa valeur communicative : un texte très spécifique était autrement perçu et avait d'autres fonctions pragmatiques qu'un texte compréhensible par tous. Cette question implique aussi les sous-ensembles de l'écrit documentaire : un terme ou un phrasème peut caractériser plus particulièrement un genre donné. S'il y avait un vocabulaire récurrent, par exemple, dans des chartes d'arbitrage — indépendamment des dénominations de l'action —, ce genre textuel correspondrait à une segmentation linguistiquement pertinente.

(3) Pour les termes ou phrasèmes en question qui ne sont pas largement répandus dans le temps et l'espace, il faut spécifier en quoi la forme est peu courante : s'agit-il d'un mot régional, éventuellement lié à certaines chancelleries ? Cela est vrai pour *miliaires* "millésime", attesté surtout dans l'est du territoire d'oïl ou — comme l'a montré J. Monfrin<sup>39</sup> — pour *junet* "juin", non pas "juillet", sémantisme propre à la Lorraine et à la Franche-Comté. Est-ce un néologisme, un archaïsme ou un usage occasionnel ? Les marques diatopiques et diachroniques contribuent, elles aussi, à déterminer la valeur communicative du texte en question<sup>40</sup>.

En résumé, si l'on travaille sur des textes isolés, les interrogations seront moins étymologiques que diasystématiques ; si au contraire l'on travaille sur un corpus de textes plus important, les questions de genèse et de constitution du vocabulaire technique se posent constamment. Le lexique juridico-administratif a connu une transformation en profondeur par sa mise à l'écrit dans une langue vernaculaire : si certains termes-clés ont pu avoir une vie antérieure, le vocabulaire roman s'élargit et se spécifie par son usage à l'écrit. Le manque de stabilité d'un stock lexical encore récent se manifeste notamment par une grande variation dérivationnelle et syntagmatique des termes. L'idée, souvent évoquée, que tout ce langage existait déjà à l'oral au XI<sup>e</sup> voire au IX<sup>e</sup> siècle et ne fit qu'apparaître au XIII<sup>e</sup> sous une forme écrite est erronée ; elle ne tient pas suffisamment compte de l'impact de l'écrit sur l'évolution linguistique ni des changements provoqués par sa médialité.

### III.2.2. Le vocabulaire agricole et le vocabulaire général

L'écrit documentaire contient par ailleurs de nombreux termes du monde agricole ou urbain qui ne sont pas spécifiques du langage juridico-administratif et qui ouvrent des horizons sur d'autres domaines de l'histoire lexicale et culturelle. Pour ces termes, les actes ne font que reproduire un usage dont l'épicentre sociolinguistique se trouve ailleurs. Les lexèmes en question ont eu une existence largement antérieure aux textes du XIII<sup>e</sup> siècle ; ils sont nés à un moment indéterminé depuis le VI<sup>e</sup> siècle, s'ils ne continuent pas déjà un mot du latin parlé tardif, transformé phonétiquement. Leur évolution est peu affectée par leur usage dans les



chartes qui se font le miroir, dans un premier temps, de scènes de la vie agricole. Ce n'est que par les mécanismes plus généraux de la standardisation auxquels les chartes contribuent que l'histoire des termes agricoles peut être modifiée.

En même temps, les chartes représentent souvent les seules sources qui nous renseignent sur ce vocabulaire non-littéraire et sur les concepts qu'il véhicule. Elles fournissent par là des informations précieuses puisque tout terme mis à l'écrit contribue à la construction d'un sens culturel défini. Je dois avouer qu'après avoir étudié le vocabulaire juridique de sources documentaires dans ma thèse (*Lo Thesaur del hospital de Sant Sperit*, 1989), le jugeant particulièrement caractéristique pour ces genres, je m'intéresse bien plus maintenant aux autres domaines sémantiques concernés par ces documents, à commencer par le vocabulaire du monde agricole. Les questions qui se posent sont rigoureusement identiques à celles soulevées pour le vocabulaire juridico-administratif, étymologiques d'abord, diasystématiques ensuite (cf. n. 38).

### **III.2.3. Bases lexicographiques**

L'étude du lexique est certes lourde, mais elle est parfaitement accessible à des non-linguistes, bien plus que les études de grapho-phonétique ou de grammaire. Elle suppose essentiellement une utilisation intense des dictionnaires de référence, comme le *Französische Etymologisches Wörterbuch* (FEW) pour la Galloromania ou le *Lessico Etimologico Italiano* (LEI) pour l'Italoromania, et des bases de données lexicographiques, comme le *Tesoro della lingua italiana delle origini* (TLIO) pour l'italien ou le *Corpus diacrónico del español* (CORDE) pour l'espagnol. La lexicographie permet presque toujours de bien évaluer la trajectoire évolutive d'un terme et son ancrage diasystématique et d'en fournir une définition satisfaisante. Cela vaut en tout cas pour les variétés gallo- et italo-romanes ; les variétés ibéro-romanes et notamment le gallo-portugais sont moins bien décrites.

Grâce aux dictionnaires, il est possible d'identifier pour un texte vernaculaire concret la part de son vocabulaire qui est courant et généralement répandu et la part du vocabulaire spécialisé, régionalement ou diachroniquement marqué. L'on peut voir également si ce texte est ouvert vers d'autres domaines de la vie socio-culturelle. La réponse à ces questions permet de bien ancrer le texte dans son contexte d'usage et de déterminer sa valeur communicative et symbolique à son époque. Par ailleurs, bien plus que les phénomènes graphématiques ou morphologiques, le lexique permet d'identifier des sous-ensembles textuels, autant par la présence que par l'absence de certains ensembles de vocabulaire.

### **III.3. D'autres domaines de la langue : grammaire, onomastique**

La syntaxe de la phrase simple fournit un grand nombre de renseignements sur les processus de neutralisation et normalisation linguistiques ; la syntaxe de la phrase complexe donne au contraire la mesure peut-être la plus fiable pour évaluer le degré d'élaboration d'un texte ou d'un état de langue.

Je n'ai pas encore développé dans mon corpus d'ancien français des règles paramétrables pour une évaluation syntaxique. Nous y viendrons sans doute, mais, plus encore que dans le domaine de la graphématique, je suis sceptique sur leur utilisation par des non-linguistes : la construction d'une proposition par le groupe verbal et les groupes nominaux, le système de marquage verbal et nominal et la détermination interne à ces groupes, les fonctionnalités du système TMA (temps-mode-aspect), l'importance des périphrases, les mécanismes de cohésion textuelle, tous ces sujets sont porteurs mais demandent une formation poussée qui dépasse la consultation de dictionnaires.

Cet exemple me semble par ailleurs caractéristique pour toute expérience transdisciplinaire : les représentants de deux disciplines peuvent toujours discuter les bases méthodologiques et les intérêts abstraits de leur recherche ; ils peuvent aussi apprendre sans trop de mal certaines techniques de travail de l'autre discipline qui leur sont particulièrement utiles ; mais ils rencontrent toujours des domaines définis qui demandent un tel effort d'apprentissage que celui-ci ne se justifie pas en termes de temps humain et de pratique d'un métier défini. C'est là où la transdisciplinarité doit se transformer en une véritable collaboration, si toutefois il y a un intérêt de recherche partagé.

Un dernier domaine de la langue, décisif pour l'écrit documentaire et accessible aux historiens, est l'onomastique. Après la mise par écrit des langues romanes, l'onomastique n'a plus la même valeur de témoignage unique qu'elle possède pendant l'époque précédant l'an mille. Notamment, elle ne joue plus le même rôle dans l'émergence des formes romanes qu'auparavant ; par ailleurs, la formation des noms de lieu est nettement antérieure à leur utilisation dans les chartes ; même si les chartes fournissent des attestations très précieuses de noms de lieu, elles présentent un état lexical et dérivationnel révolu depuis des siècles.

Cela n'enlève rien à l'intérêt historique de ces formes, et la présence massive des noms propres demande une réelle attention, notamment d'ordre lexicologique. Les règles de traitement linguistique des noms de personne et des noms de lieu ressemblent à ceux du lexique, à la différence près que les répertoires existants sont nettement moins fiables. Les identifications étymologiques des meilleurs ouvrages, comme ceux de A. Vincent, A. Dauzat, M.-Th. Morlet et E. Nègre, manquent de précision et leur utilisation demande une attention constante<sup>41</sup>.

Dans une logique d'évolution de la langue, les deuxièmes et troisièmes noms de personne (surnoms, noms de famille) sont donc plus intéressants, puisqu'ils se forment au Moyen Âge tardif. Les sources documentaires fournissent ici des renseignements sur certaines évolutions socio-culturelles en cours.

## IV. Exigences philologiques et linguistique informatique

### IV.1. L'encodage textuel

Le grand nombre des sources documentaires est à la fois une force et un réel handicap pour l'histoire linguistique. La masse des textes sériels dépasse les capacités d'analyse des linguistes, tenus à une lecture minutieuse des sources. À l'avenir, le développement des techniques informatiques permettra sans doute d'approcher ces grands ensembles textuels en les quantifiant. Ce n'est pas une entreprise facile, mais l'informatique semble le seul outil capable de tirer de l'écrit documentaire des informations linguistiques qui dépassent l'anecdotique.

L'intégration de l'informatique dans les études philologiques demande une attention méthodologique et empirique particulière, à commencer par l'encodage des textes : à terme, il faudra renoncer pour des éditions de textes anciens à la facilité apparente de Word et travailler avec des instruments comme OpenOffice sous Linux, LaTeX ou Tustep, qui font appel à un encodage transparent, pour la plupart de type *xml*. Un encodage de type *xml* demande une certaine discipline, mais transforme très favorablement la nature des données textuelles. Il permet notamment des interrogations multiples sur les données textuelles et peut garantir une longévité des bases de données inconnue jusqu'ici. Il ouvre la perspective que les transcriptions faites par des historiens pourront être utilisées plus facilement par des linguistes et vice versa. Tout historien pourra, après publication, mettre ses sources à la disposition d'une des grandes entreprises de corpus anciens, ce qui permettra des interrogations nouvelles sur ces mêmes textes.

Un encodage informatique neutre désamorce aussi la problématique séculaire des critères d'édition qui divergent selon les disciplines, les écoles et les époques. Prenons un exemple : une charte lorraine contient la séquence *alabe* ; pour une édition papier, il faudra trancher entre une transcription à *l'abé*, en transformant la donnée de départ, mais en rendant le texte compréhensible, une forme *alabe*, en respectant l'original, mais en compliquant la lecture, ou enfin une voie de compromis, *à-l'abé*, compréhensible et en même temps respectueuse de l'original. Notre corpus informatique permet d'obtenir les trois interprétations divergentes à partir d'une même transcription : la forme saisie *à-l'abé* (ou précisément *%a-l'ab%/e*) peut être rendue à l'impression ou dans un *browser* aussi sous les deux formes, *à-l'abé*

ou *alabe* ; la représentation imitative suppose la suppression automatique, programmée dans le logiciel, de tous les accents, tirets et apostrophes.

Lors de la saisie, il suffit d'introduire quelques signes interprétatifs supplémentaires et le texte gagne en informations sans aucune perte d'authenticité. La segmentation des mots est presque indispensable pour une interrogation informatique cohérente : reconnaître dans la suite *alabe* une préposition, un article et un nom demande un effort considérable à un programme informatique alors que la segmentation par le transcritteur coule de source.

Naturellement, il existe des cas de figure plus délicats, mais nous avons toujours trouvé des solutions non contradictoires de double encodage – interprétatif et imitatif – pour toutes les questions, notamment les majuscules, les abréviations, la ponctuation et la séparation des mots<sup>42</sup>. Il serait également possible d'intégrer des informations sur la forme différente des lettres, mais sur ce point je n'arrive pas à suivre mes collègues germanistes et m'en tiens très traditionnellement à une forme unique par lettre.

Le texte saisi doit ensuite être accompagné d'indications précises sur son ancrage dans le diasystème : la date du texte et, le cas échéant, de la copie, l'identification du lieu d'origine, idéalement de la chancellerie ou du scriptorium où le texte a été rédigé, enfin la détermination du genre textuel, au moins selon des critères diplomatiques.

## **IV.2. Les interrogations linguistiques par informatique**

Si l'encodage informatique d'un texte documentaire ancien est une question résolue au moins intellectuellement, il n'en va pas de même pour les instruments d'interrogation. Il existe différents programmes de lemmatisation comme *gatto*, le moteur utilisé par le *TLIO*, ou comme *phœnix*, le logiciel que nous développons actuellement et qui est utilisé pour les *Plus anciens documents linguistiques de la France*. Mais la recherche sur des outils qui permettront de dominer le problème des masses documentaires en est encore à sa phase initiale.

Prenons la distance lexicale existant entre deux textes ou ensembles textuels que plusieurs logiciels permettent de mesurer. D'un point de vue informatique, il n'est pas difficile de relever les formes qui apparaissent avec une fréquence particulière ou exclusive dans une série de chartes, comparée à une autre série, d'une autre région ou époque. À travers ces interrogations, il sera possible d'identifier sur une base objective des genres textuels d'après une définition linguistique et de suivre leur évolution à travers les siècles.

Je suis cependant encore à la recherche d'une voie pour poser ces dernières

questions de manière adéquate en utilisant les données textuelles complexes de mon corpus ; en dernière instance, le linguiste se heurte ici aux mêmes problèmes que peut rencontrer un non-linguiste avec la grammaire : faire appel à des outils informatiques pour des questions de ce genre revient encore, actuellement, à entreprendre auparavant une formation très intense à une nouvelle discipline, l'informatique, qui peut dépasser les forces disponibles dans une vie de chercheur, sans parler du fait que la programmation n'est pas nécessairement une passion pour un historien de la langue.

Mes tentatives personnelles ne me semblent pas encore pleinement satisfaisantes même si je crois qu'elles s'inscrivent dans une voie viable : le logiciel *phoenix* développé en collaboration avec deux amis programmeurs<sup>43</sup> répond à des exigences de base et pourra être utilisé par d'autres chercheurs avec un investissement raisonnable en termes de temps ; actuellement, je fais des sondages avec X-Query pour des interrogations ultérieures et je suis certain que les modèles de programmes seront utiles ; devant les difficultés de dialogue avec des collègues informaticiens, il me semble nécessaire de développer une culture informatique pour les historiens et les linguistes, même si celle-ci restera toujours partielle.

Le problème des masses documentaires n'est pourtant pas encore résolu par une utilisation standard de l'informatique. Les prochaines années verront la constitution de bases textuelles dans une mesure encore difficilement imaginable, sans doute avec le concours de toutes les disciplines historiques, comme par exemple l'histoire littéraire ou l'histoire de la médecine. La question de la sélection des données restera entière : comment identifier, dans un corpus qui contient 1500 occurrences du mot *loyer*, éventuellement déjà lemmatisées, les cinq formes qui ont un sens particulier et trahissent l'usage d'une chancellerie donnée ? Pour ne pas se noyer dans la masse des occurrences, il faudra tenir compte de la fréquence d'un terme dans les sous-ensembles textuels et de son contexte syntagmatique : un mot dont la fréquence et l'entourage changent d'un genre textuel à un autre ou d'une époque à une autre, risque d'avoir aussi connu des changements sémantiques. Là encore, un grand effort de dialogue dans la recherche transdisciplinaire nous attend.

## **Conclusion : l'apport de l'histoire linguistique à l'histoire**

L'étude de l'écrit documentaire apparaît comme un domaine de recherche à part entière qui jusqu'ici n'a été entamé que très partiellement par les historiens linguistes. Ceux-ci restent tributaires des avancées des historiens et diplomatistes qui utilisent et interprètent systématiquement leurs sources. En échange, l'histoire linguistique peut enrichir la réflexion historique

proprement dite. Elle ajoute de nouvelles facettes à la description de l'écrit documentaire et à l'interprétation de ses conditions de production et de réception. Notamment dans des domaines bien délimités comme ceux des genres textuels, du changement latin-roman, des formes graphiques régionales, du vocabulaire spécialisé, de l'onomastique ou de l'élaboration philologique des textes, elle apporte des outils et des approches complémentaires, pleinement utilisables par des historiens. Si les linguistes et les historiens n'ont pas l'habitude de collaborer, ils ont toutefois beaucoup de choses à se dire.

---

## Notes

---

1. Cf. Gerhard Ernst, Martin-D. Gleßgen, Christian Schmitt, Wolfgang Schweickard (éd.), *Romanische Sprachgeschichte / L'histoire linguistique de la Romania* (HSK 23.1-3) [désormais *RSG*, suivi de l'indication du numéro d'article (le premier volume correspond aux articles 1-100, le deuxième aux articles 101-200, le troisième, en cours d'achèvement, aux articles 201-263), du titre et du nom de l'auteur de l'article], 3 vol., Berlin, New York, 2003-[2007], art. 1, *Romanistik und Sprachgeschichtsschreibung* (iid.).

2. Les différents facteurs sont étudiés de manière plus approfondie dans *RSG* ; synthèse plus succincte dans mon introduction *Linguistique romane, domaines et méthodes en linguistique française et romane* (A. Colin, 2007 [désormais *DomMéth*]), chap. 3 (« Histoire externe des langues et variétés romanes »). La *RSG* a été conçue pour un usage en parallèle avec le *Lexikon der romanistischen Linguistik [LRL]*, éd. Günter Holtus, Michael Metzeltin, Christian Schmitt, 12 vol., Tübingen, 1988-2003 ; cf. pour nos interrogations notamment les volumes I/1, I/2 [*Méthodologie*, 2001] et II/1, II/2 [*Latin / Roman, Grammaire historique comparée des langues romanes*, 1996, *Les différentes langues romanes du Moyen Âge et à la Renaissance*, 1995].

3. L'élaboration linguistique présente deux mouvements opposés : d'un côté, une neutralisation et normalisation qui affectent les domaines de la phonétique et de la graphématique, le lexique de base et la syntaxe de la phrase simple ; de l'autre côté, l'élargissement d'un lexique spécialisé dans les différents domaines du savoir et le développement de la syntaxe complexe permettant un encodage plus dense et décontextualisé.

4. La terminologie de la linguistique variationnelle (diasystème, variétés diatopiques, diastratiques, diaphasiques, langage de proximité et de distance, 'langues-toit') semble déroutante à première vue, mais elle est bien établie et très utile ; cf. par ex. *RSG*, art. 220, *L'architecture comparée des langues romanes* (P. Koch et W. Oesterreicher), ou *DomMéth*, chap. 1.5. (« L'étude variationnelle des langues romanes »).

5. Cf. *DomMéth*, chap. 1.5.6., n° 3. Le schéma est développé dans M.-D. Gleßgen, « Diskurstraditionen zwischen pragmatischen Regeln und sprachlichen Varietäten », *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*, éd. Angela Schrott et Harald Völker, Göttingen, 2005, p. 207-228, à la p. 217.

6. Éd. Hans Robert Jauss et al., Heidelberg, Winter, 1968-.

7. P. ex. le dernier volume (vol. 11, 1987) totalise 1150 pages en trois tomes sur les seuls textes historiographiques.

8. Éd. Barbara Frank et Jörg Hartmann avec la collaboration de Heike Kürschner, 5 vol., Tübingen, 1997 ; la réflexion sur laquelle repose l'ouvrage est expliquée par Peter Koch, « Pour une typologie conceptionnelle et médiale des plus anciens documents / monuments des langues romanes », *Le passage à l'écrit des langues romanes*, éd. Maria Selig, Barbara Frank et Jörg Hartmann, Tübingen, 1993, p. 39-81.

9. Cf. le compte-rendu de Françoise Viellard dans *Cahiers de civilisation médiévale*, 43 (2000), p. 294-298 ; à la p. 298 : « Cet inventaire repose donc sur une documentation incomplète et peu sûre qui donne une fausse impression de précision en jouant d'artifices typographiques ».

10. C'est une importante partie qui couvre la fin du troisième (*Législation*) et tout le quatrième volume (*Chartes, Lettres, Documents administratifs*).

11. Cf. M.-D. Gleßgen, « Das altfranzösische Geschäftsschrifttum in Oberlothringen : Quellenlage und Deutungsansätze », *Skripta, Schreiblandschaften und*

*Standardisierungstendenzen, Urkundensprachen im Grenzbereich von Germania und Romania im 13. und 14. Jahrhundert*, éd. Kurt Gärtner, Günter Holtus, Andrea Rapp et Harald Völker, Trier, 2001, p. 257-294, à la p. 265.

12. Un exemple : si une charte nomme comme auteur un personnage résidant à Metz, ce n'est pas pour autant que ce texte pourra être qualifié linguistiquement comme "français - lorrain", en tout cas pas avant une étude préalable ; les rédacteurs des chartes ne coïncident souvent pas avec les auteurs de celles-ci ; par ailleurs, une entité géolinguistique comme "lorrain" est très vague dans ce contexte. La même critique est faite par H. Völker, *Skripta und Variation, Untersuchungen zur Negation und zur Substantivflexion in altfranzösischen Urkunden der Grafschaft Luxemburg (1237-1281)*, Tübingen, 2003, à la p. 68, n. 292.

13. Cf. *DomMéth*, chap. 3.3. (« La genèse de la Romania [V<sup>e</sup>/X<sup>e</sup> siècles] »).

14. Cf. en premier Francesco Sabatini, « Dalla 'scripta latina rustica' alle 'scriptae' romanze », *Studi medievali*, 9 (1968), p. 320-358 ; plus récemment Jean-Pierre Chambon, « L'identité langagière des élites cultivées d'Arvernie autour de l'an Mil et la *scripta latina rustica* : réflexions à propos du *Breve de libros* du chapitre cathédral de Clermont (984-1010) », *Revue de linguistique romane*, 62 (1998), p. 381-408.

15. Cf. Peter Koch, « Graphé, ihre Entwicklung zur Schrift, zum Kalkül und zur Liste », *Schrift, Medien, Kognition*, éd. Peter Koch et Sybille Krämer, Tübingen, 1997, p. 43-81, aux p. 69 et suiv.

16. P.ex. « L'agencement spatial et fonctionnel des vicairies carolingiennes dans le midi de la Gaule : une approche linguistique », *Revue de linguistique romane*, 63 (1999), p. 55-174 ; « L'onomastique du censier interpolé (ca 946) dans la charte de fondation du monastère auvergnat de Sauxillanges », *ibid.*, 68 (2004), p. 105-180.

17. Cf. le résumé donné par Wolfgang Haubrichs, *RSG*, art. 64 (*Die verlorene Romanität im deutschen Sprachraum*) ainsi que les thèses germano-romanistes de Monika Buchmüller-Pfaff (*Siedlungsnamen zwischen Spätantike und frühem Mittelalter, die -(i)acum-Namen der römischen Provinz Belgica Prima*, Tübingen, 1990), Martina Pitz (*Siedlungsnamen auf -villare [-weiler, -villers] zwischen Mosel, Hunsrück und Vogesen, Untersuchungen zu einem germanisch-romanischen Mischtyps der jüngeren Merowinger- und der Karolingerzeit*, 2 vol., Saarbrücken, 1997) ou Maria Besse (*Namenpaare an der Sprachgrenze, eine lautchronologische Untersuchung zu zweisprachigen Ortsnamen im Norden und Süden der deutsch-französischen Sprachgrenze*, Tübingen, 1997) ; pour la région alpine, cf. aussi Max Pfister, « La popolazione del Trentino-AltoAdige verso l'anno 600 », *Italia longobarda*, éd. Gian Carlo Menis, *Italia longobarda*, Venise, 1991, p. 175-225.

18. Il y a naturellement des exceptions, quelques textes littéraires ou encore des textes de pharmacognosie comme le *Dioscoride* latin ou les réceptaires édités par Sigerist ; cf. aussi les études récentes sur les monnaies mérovingiennes : Jean-Pierre Chambon et Yan Greub, « Données nouvelles pour la linguistique gallo-romane : les légendes monétaires mérovingiennes », *Bulletin de la Société de linguistique de Paris*, 95 (2000), p. 147-182 ; J.-P. Chambon, « L'identification des noms d'ateliers monétaires mérovingiens (Arverne et environs) : point de vue de linguiste », *Revue numismatique*, 2001, p. 347-405.

19. Sur ce point, la position de Jacques Monfrin, en lançant la série des *Plus anciens documents linguistiques de la France*, était trop rigoriste ; cf. « Le mode de tradition des actes écrits et les études de dialectologie », *Revue de linguistique romane*, 32 (1968), p. 19-47, repr. dans Id., *Études de philologie romane*, Genève, 2001 (*Publications romanes et françaises*, 230), p. 145-173, à la p. 169 : « En revanche, les copies tardives (...) n'offrent pas la moindre garantie ».

20. Cf. Benoît-Michel Tock (dir.), *La diplomatie française du Haut Moyen Âge vue à travers les originaux, Inventaire des chartes originales antérieures à 1121 conservées en France*, 2 vol., Turnhout, 2001 ; CETDOC, *Thesaurus Diplomaticus (de la Belgique des origines à 1200)*, cd-rom, Turnhout, 1997.

21. Cf. B.-M. Tock, *Scribes, souscripteurs et témoins dans les actes privés en France (VII<sup>e</sup>-début XII<sup>e</sup> siècle)*, Turnhout, 2005, p. 109-143.

22. Cf. *DomMéth*, chap. 3.4.3, n° 2.

23. Cf. de nouveau P. Koch, « Pour une typologie conceptionnelle et médiale... ».

24. Le changement de code suppose une coexistence des deux langues pendant une époque de transition qui peut englober plusieurs siècles ; il implique des interférences linguistiques et mène parfois à des textes mixtes, donc une alternance de codes (angl. *code switching*) ; cf. de nouveau J.-P. Chambon, « L'identité langagière des élites cultivées... ».

25. D'après une évaluation de Paul Videsott, fondée sur sa base de données d'environ 4 000 chartes latines et françaises rédigées dans l'entourage royal.
26. Cf. « Église et puissance temporelle en terre lorraine : une charte épiscopale de Toul de 1237 en l'honneur de Gilles Roques », *Mélanges de linguistique et de philologie romanes*, éd. J.-P. Chambon, J.-P. Chauveau et P. Swiggers (sous presse).
27. Pour l'écrit documentaire un ensemble très approximatif de 400 000 pages françaises manuscrites (Gleißgen, « Das altfranzösische Geschäftsschrifttum... », p. 268), pour les autres genres peut-être 150 textes d'une taille supérieure à une dizaine de pages.
28. Cf. Carl Th. Gossen, *Französische Skriptastudien, Untersuchungen zu den nordfranzösischen Urkundensprachen des Mittelalters*, Wien, 1967 ; Hans Goebel, *Die normandische Urkundensprache, ein Beitrag zur Kenntnis der nordfranzösischen Urkundensprachen des Mittelalters*, Wien, 1970 ; Anthonij Dees, *Atlas des formes et des constructions des chartes françaises du 13<sup>e</sup> siècle*, Tübingen, 1980 ; Max Pfister, « Scripta et koinè en ancien français aux XII<sup>e</sup> et XIII<sup>e</sup> siècles ? », *Écriture, langues communes et normes*, éd. Pierre Knecht et Zygmunt Marzys, Genève, 1993 ; Hans Goebel et Guillaume Schiltz, « Der 'Atlas...' von Anthonij Dees - dialektometrisch betrachtet », *Skripta, Schreiblandschaften...*, p. 169-221 ; cf. l'aperçu sur l'histoire de la scriptologie donné par H. Völker, *Skripta und Variation...*, p. 9-79.
29. Cf. Hans Goebel : « *Le Rey est mort, vive le Roy* : nouveaux regards sur la scriptologie », *Travaux de linguistique et de littérature*, 13 (1975), p. 145-209, à la p. 167.
30. Les avancées ressortent bien d'une série de volumes d'actes publiés après trois colloques qui se sont tenus à Trèves : *Urkundensprachen im romanisch-germanischen Grenzgebiet*, éd. K. Gärtner et G. Holtus (éd.), Mainz, 1997 ; *Skripta, Schreiblandschaften...* ; *Überlieferungs- und Aneignungsprozesse im 13. und 14. Jahrhundert auf dem Gebiet der westmitteldeutschen und ostfranzösischen Urkunden- und Literatursprachen*, éd. K. Gärtner et G. Holtus, Trier, 2005.
31. Cf. H. Völker, *Skripta und Variation...*, p. 194 et suiv.
32. Il s'agit des 290 documents conservés en Meurthe-et-Moselle et produits entre 1232 et 1265 ; cf. Michel Arnot et M.-D. Gleißgen, *Les plus anciens documents linguistiques de la France : Meurthe-et-Moselle*, éd. électronique en cours.
33. Cf. M.-D. Gleißgen, « Les 'lieux d'écriture' et leur identification dans les documents lorrains du XIII<sup>e</sup> siècle », *Revue de linguistique romane*, 70 (2006), [sous presse] ; cf. aussi Dumitru Chihaï, *Les chartes de Toul : caractéristiques graphématiques*, mémoire de maîtrise, Strasbourg, 2003, et Anne-Christelle Matthey, *La préstandardisation dans la scripta lorraine du XIII<sup>e</sup> siècle*, thèse, Zürich, 2006 (à paraître).
34. Cf. néanmoins l'étude de H. Goebel, « La dialectométrie corrélative, un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme », *Revue de linguistique romane*, 69 (2005), p. 321-368, qui prouve qu'il peut exister un sérieux décalage entre différents domaines de la langue.
35. Cf. l'aperçu dans *DomMéth*, chap. 2.4 (« La lexicologie »).
36. Cf. Claire Muller, *Marquages phraséologiques et formules dans le dispositif*, mémoire de maîtrise, Strasbourg, 2003.
37. Plus précisément, la marque diaphasique dépend de la fréquence avec laquelle un terme apparaît dans des contextes spécialisés, de sa faible fréquence voire absence dans d'autres contextes et du degré de spécificité de son sens.
38. Les exemples proviennent actuellement des mémoires de maîtrise élaborés dans le cadre de ce projet à Strasbourg : Jason S&n, *Les chartes lorraines du XIII<sup>e</sup> siècle*, 1999 ; Frédérique Gisquet, *Étude philologique et lexicologique d'un corpus de chartes lorraines du XIII<sup>e</sup> siècle*, 2000 ; Delphine Harmand, *Id., approche méthodologique*, 2000 ; les résultats seront intégrés dans les articles de la base de données lexicologique en cours d'élaboration.
39. Cf. J. Monfrin, « Notes lexicographiques (...) II. Après la feste sain Johan, en juignet (Villehardouin 490) » [1973], repr. dans *Études de philologie romanes...*, p. 197-210.
40. J'ai essayé d'analyser plus en détail les qualités diaphasiques dans le vocabulaire des chartes dans les deux études suivantes : « Realia und Urkunden, die Teilung eines lothringischen Stadthauses kurz nach 1400 », *Romanische Sprachwissenschaft. Festschrift für Christian Schmitt zum 60. Geburtstag*, éd. A. Gil et al., Frankfurt a.M., 2004, p. 423-447, et « Vergleichende oder einzelsprachliche historische Textwissenschaft », *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten?*,



*Romanistisches Kolloquium XX*, éd. W. Dahmen et al., Tübingen, 2006, p. 319-340.

41. Cf. encore les articles de Jean-Pierre Chambon cités ci-dessus, n. 16.

42. Cf. la présentation des critères d'édition dans « L'élaboration philologique et l'étude lexicologique des *Plus anciens documents linguistiques de la France* à l'aide de l'informatique », Frédéric Godefroy. *Actes du X<sup>e</sup> colloque international sur le moyen français*, éd. Frédéric Duval, Paris, 2003 (*Mémoires et documents de l'École des chartes*, 71), p. 371-386.

43. Matthias Kopp (Tübingen) et Matthias Osthof (Tübingen/Zürich).