



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2010

Présentation générale : Architecture et méthodologie du projet des Plus anciens documents linguistiques de la France, Édition électronique

Glessgen, Martin-Dietrich

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

ZORA URL: <https://doi.org/10.5167/uzh-60843>

Journal Article

Published Version

Originally published at:

Glessgen, Martin-Dietrich (2010). *Présentation générale : Architecture et méthodologie du projet des Plus anciens documents linguistiques de la France, Édition électronique*. Bibliothèque de l'École des Chartes, 168:7-24.

PRÉSENTATION GÉNÉRALE

ARCHITECTURE ET MÉTHODOLOGIE DU PROJET DES *PLUS ANCIENS DOCUMENTS LINGUISTIQUES DE LA FRANCE*, ÉDITION ÉLECTRONIQUE

par

MARTIN-D. GLESSGEN

L'idée des *Plus anciens documents linguistiques de la France*, développée d'abord par Paul Meyer depuis les années 1890¹, continuée par Clovis Brunel² et reprise, sous une forme encore plus ambitieuse, par Jacques Monfrin³, s'inscrit dans plus d'un siècle de recherche scriptologique gallo-romane, jalonnée par des noms comme ceux de Louis Remacle, Carl Theodor Gossen, Louis Carolus-Barré, Hans Goebel, Antonij Dees ou Max Pfister⁴.

1. Dès avant le volume des *Documents linguistiques du Midi de la France* (Paris, 1909), ont paru son étude sur « Le langage de Die au XIII^e siècle », dans *Romania*, t. 20, 1891, p. 70-85, et surtout la publication préliminaire des « Documents linguistiques des Basses-Alpes », dans *Romania*, t. 27, 1898, p. 337-441.

2. Voir Clovis Brunel, *Les plus anciennes chartes en langue provençale : recueil des pièces originales antérieures au XIII^e siècle*, Paris, 1926-1952, 2 vol., et son étude préliminaire « Les premiers exemples de l'emploi du provençal dans les chartes », dans *Romania*, t. 48, 1922, p. 335-364 ; voir aussi Jérôme Belmon et Françoise Vielliard, « Latin farci et occitan dans les actes du XI^e siècle », dans *Bibliothèque de l'École des chartes*, t. 155, 1997, p. 149-185.

3. Voir son premier article méthodologique, « Le mode de tradition des actes écrits et les études de dialectologie », dans *Revue de linguistique romane*, t. 32, 1968, p. 17-47, et son « Introduction » programmatique au premier volume de la collection dirigée par lui : Jean-Gabriel Gigot (éd.), *Chartes en langue française antérieures à 1271 conservées dans le département de la Haute-Marne*, Paris, 1974 (*Documents linguistiques de la France, série française*, 1), p. XI-LXXX.

4. Voir les excellents aperçus sur l'évolution et l'état de la recherche dans l'« Introduction » citée de J. Monfrin et dans Harald Völker, *Skripta und Variation : Untersuchungen zur Negation und zur Substantivflexion in altfranzösischen Urkunden der Grafschaft Luxemburg (1237-1281)*, Tübingen, 2003 (*Beihefte zur Zeitschrift für romanische Philologie*, 305), p. 9-79.

I. CONCEPTION DU PROJET ET STRUCTURE EXTERNE.

Le récent développement des moyens électroniques a introduit dans les années 1990 de nouvelles attentes et de nouveaux standards tant dans le domaine des éditions que dans celui des analyses scriptologiques. Le temps était alors aux grandes visions globalisantes, prévoyant de vastes ensembles de textes de l'ancienne langue, disponibles sous différentes formes de représentation, à commencer par de meilleures éditions critiques, enrichies de transcriptions « diplomatiques » (imitatives) et de reproductions photographiques des différents manuscrits. Inspiré par de telles perspectives, nous avons repris en 1999 l'idée initiale de Paul Meyer, soutenu, après le décès de Jacques Monfrin, par Françoise Vielliard et Olivier Guyotjeannin. L'objectif était alors celui de continuer, sous les auspices de l'ère informatique, l'entreprise de répertoire des documents français (et occitans) du XIII^e siècle conservés dans les différents départements de la France, pour les éditer *ex novo* ou, le cas échéant, pour reprendre d'anciennes éditions⁵.

Le projet concret s'est construit dans un monde en pleine effervescence puisque les transformations radicales dans les nouveaux médias ont amené l'historiographie linguistique et la philologie à une restructuration profonde de leurs outils disciplinaires. Si ces transformations médiales affectent l'intégralité des sciences et de la communication, et si chacune des disciplines doit forger une synthèse entre les nouveaux outils et ses propres traditions méthodologiques, les recherches philologiques et linguistiques en sont particulièrement touchées : elles ont comme objets immédiats des textes linéaires et s'inscrivent donc parfaitement dans la logique d'interprétation de l'ordinateur, elle aussi linéaire.

L'intégration des deux mondes implique, comme toute transformation radicale, une déperdition d'énergie notable puisqu'elle se heurte inévitablement à des difficultés tant pratiques que théoriques. Un premier problème concerne les compétences très diverses auxquelles les deux univers font appel : il existe, aujourd'hui encore, une séparation presque totale entre les personnes qui ont des connaissances et intérêts philologiques et historiques et celles qui ont de véritables capacités de programmation. Par ailleurs, non seulement l'informatique appliquée mais aussi la linguistique se trouvent

5. Une première phase a consisté en la prise en considération des matériaux légués par Jacques Monfrin à Françoise Vielliard ; cette étape a mené à des réflexions communes sur les opérations à suivre et les méthodes à appliquer. La deuxième phase, à savoir l'élaboration concrète des matériaux depuis 2005, est essentiellement réalisée et coordonnée par l'auteur de ces lignes, dans le cadre du projet « Écrit documentaire et élaboration linguistique du français au Moyen Âge tardif » du Fonds national suisse (Nationaler Forschungsschwerpunkt *Medienwandel–Medienwechsel–Medienwissen : historische Perspektiven* ; en ligne, www.mediality.ch).

actuellement aux prises avec des processus de transformation accélérés qui concernent aussi bien le domaine des interprétations que l'organisation concrète du milieu de la recherche⁶. Enfin, il subsiste deux problèmes épineux dans l'orientation de la recherche : (1) de très nombreux chercheurs travaillent, parallèlement, à l'interface entre la linguistique (historique) et l'informatique, ce qui brouille le regard ; (2) bien plus grave, la réflexion scientifique est perturbée par l'absence d'une doctrine formée concernant les outils informatiques et les méthodes qui en découlent ; les tentatives de hiérarchisation et d'évaluation de ces outils sont rares, ce qui rend leur accès plus difficile, notamment pour les non-spécialistes que sont les philologues.

Malgré ces solides obstacles, le nouveau projet des *Plus anciens documents* nous a séduits depuis le début par ses vastes perspectives, étant donné le nombre relativement faible de recherches linguistiques menées dans le passé sur des textes documentaires. Pour la linguistique, s'y ajoutent les qualités particulières des chartes ; elles sont souvent transmises sous leur forme originale, sont facilement datables et connaissent une nette variation géolinguistique aussi bien que diastatique. Les chartes se trouvent par ailleurs à la charnière entre l'écriture et son utilisation dans un contexte pragmatique défini ; la mise en scène dont elles forment le noyau cristallisateur poursuit des finalités politiques, économiques ou sociales diverses, tout en provoquant des effets ultérieurs multiples, d'ordre sociologique, cognitif

6. Ne serait-ce que dans le domaine de la scriptologie gallo-romane, ont paru, depuis la synthèse de H. Völker, *Skripta und Variation...*, toute une série de volumes thématiques pertinents ; voir les deux publications de l'École des chartes : *Frédéric Godefroy : actes du X^e colloque international sur le moyen français (Metz, 12-14 juin 2002)*, éd. Frédéric Duval, Paris, 2003, et *La langue des actes : actes du XI^e congrès de la Commission internationale de diplomatique (Troyes, 11-13 septembre 2003)*, éd. Olivier Guyotjeannin, Éditions en ligne de l'École des chartes, n° 7 [2006] (elec.enc.sorbonne.fr/sommaire163.html) ; les différents volumes d'actes des colloques de Trèves, en dernier lieu *Skripta, Schreiblandschaften und Standardisierungstendenzen : Urkundensprachen im Grenzbereich von Germania und Romania im 13. und 14. Jahrhundert, Beiträge zum Kolloquium (16.-18. September 1998)*, éd. Kurt Gärtner, Günter Holtus, Andrea Rapp et H. Völker, Trèves, 2001, et *Drittes Trierer Urkundensprachekolloquium (20.-22. Juni 2001)*, éd. K. Gärtner et G. Holtus, Trèves, 2005 ; *Ancien et moyen français sur le Web : enjeux méthodologiques et analyse du discours* (colloque, Ottawa, 4-5 octobre 2002), éd. Pierre Kunstmann, France Martineau et Danielle Forget, Ottawa, 2003 ; « De la philologie aux nouveaux médias : éditions de textes, linguistique de corpus, analyse informatique du langage », éd. H. Völker, Lene Schøsler, M.-D. Gleßgen et Costanzo Di Girolamo, section 2 des *Actes du XXIV^e Congrès international de linguistique et philologie romanes (septembre 2004, Aberystwyth)*, éd. David A. Trotter, Tübingen, 2007, t. I, p. 285-480 ; enfin les deux volumes diachroniques de linguistique de corpus : *Romanistische Korpuslinguistik II : Korpora und diachrone Sprachwissenschaft = Romance corpus linguistics II : Corpora and diachronic linguistics*, éd. Claus Pusch, Johannes Kabatek et Wolfgang Raible, Tübingen, 2005 (*ScriptOralia*, 130), et *Romanistische Korpuslinguistik III : Korpora und Pragmatik = Romance corpus linguistics III : Corpora and pragmatics*, éd. C. Pusch, Stefan Pfänder et W. Raible, Tübingen, à paraître en 2011 (*ScriptOralia*).

ou, justement, linguistique. Les chartes originales ont donc cette qualité particulière de préserver l'unité de genèse entre leurs caractéristiques externes (support, mise en page, écriture) et internes (contenu, état de langue) ; cette unité est rompue dans la plupart des textes anciens par l'acte de copie, puisque les caractéristiques externes d'une copie renvoient à un autre contexte de genèse que ses caractéristiques internes, liées à l'original. Les chartes originales permettent donc d'étudier l'utilisation de l'écrit dans un contexte donné et les effets produits par cette situation « médialisée », effets qui concernent autant le contexte en question que la forme et la langue du document même ⁷.

Un tel projet dépasse de loin les forces d'un seul chercheur. Ce problème était déjà apparu clairement à Jacques Monfrin, qui avait eu par conséquent l'idée de lancer un assez grand nombre de travaux de jeunes savants, chacun dans les limites d'un département actuel. Mais même sous cette forme, son projet n'a pas pu se réaliser comme il l'avait prévu, et bon nombre de volumes sont restés en suspens. Aujourd'hui, les exigences scientifiques sont encore plus rigoureuses puisque la méthodologie philologique s'est fortement développée, la linguistique est devenue plus performante et les outils informatiques permettent, certes, d'aller plus loin qu'avant mais demandent en revanche un effort supplémentaire.

En d'autres termes : la tentative d'introduire les genres textuels documentaires dans l'historiographie linguistique ne soulève pas seulement des réflexions disciplinaires, elle demande une réflexion de type épistémologique en vue d'établir les voies et les choix qui permettront sa réalisation. Nous sommes obligés, par conséquent, d'adopter dans la discussion disciplinaire des réflexions de type organisationnel, puisque nous sommes en face d'une recherche coordonnée et non d'un projet individuel.

Un premier choix a été celui de positionner le projet philologique des *Plus anciens documents* dans un cadre interdisciplinaire, à Strasbourg d'abord, puis à Zurich : cadre qui a permis de débattre avec des médiévistes littéraires (Christian Kiening) ou historiens (Benoît Tock, Simon Teuscher), des historiens du droit (Andreas Thier, Stefan Geyer), des historiens de l'art, ou encore des linguistes germanistes (Elvira Glaser). L'intégration dans le pôle de recherche national suisse (PRN) consacré à la « médialité » médiévale a introduit dans le projet ce nouveau concept (voir note 5). La médialité ne concerne pas seulement les supports (dans notre cas, la feuille de parchemin) et son exploitation en termes de mise en page (« Textgestalt als Zeichen », la forme textuelle en tant que signe) et de choix linguistiques (français/latin, syntaxe, lexique), mais aussi la situation dans laquelle un document est utilisé, les règles et contraintes du genre

7. Voir M.-D. Gleßgen, « L'écrit documentaire dans l'histoire linguistique de la France », dans *La langue des actes...*, (18 p.) et Paul Videsott, « À propos du plus ancien document en français de la chancellerie royale capétienne », dans le présent fascicule, p. 61-81 (pour la situation en Italie, voir la référence donnée par la note 12 du même article).

textuel en question ou l'état d'élaboration de la langue du document. Une telle conception a pour premier mérite de mettre l'accent sur les aspects matériels de la transmission linguistique (la médialité dans son acception traditionnelle, déjà trop peu prise en considération par l'historiographie) ; elle enrichit par ailleurs la théorie de la communication, en distinguant, d'une part, les contenus, formes et intervenants et, de l'autre, les modalités qui créent le lien entre eux (l'action médiale au sens large) ; elle permet enfin de cerner l'action et la puissance médiales de la langue en tant que telle, grâce aux multiples connotations (régionales, sociolectales, situationnelles et liées au genre textuel), que les termes et constructions véhiculent au-delà de leur sens dénotatif.

Le cadre du pôle de recherche national apporte en même temps les financements nécessaires pour garantir l'infrastructure indispensable à la réalisation du projet. Celui-ci est intégré dans l'enseignement universitaire et donne lieu régulièrement à des travaux de jeunes chercheurs⁸. Il est impossible de réaliser un tel projet sans constituer une véritable équipe de travail, qui donne une certaine stabilité aux différentes opérations, bibliographiques, informatiques ou de numérisation, et garantisse un volume suffisant au répertoire et à l'édition des documents⁹.

D'autres partenariats et collaborations sont venus enrichir la mise en œuvre de ce projet tentaculaire (ci-dessous, II, 1) ; soulignons notamment l'inventaire des 3 600 chartes françaises et latines émanant de la royauté française et de l'entourage royal immédiat, établi par Paul Videsott¹⁰. À terme, les méthodes et outils élaborés dans le cadre des *Plus anciens documents* seront sans doute amenés à prendre leur indépendance et à dépasser ce contexte, puisque c'est un domaine de la recherche qui permet des segmentations cohérentes et opératoires. En dernière instance, il s'agit d'un nouvel univers à découvrir et à intégrer dans l'historiographie linguistique, en philologie et, par les aspects interdisciplinaires développés ici, également en histoire.

8. À Strasbourg, plusieurs étudiants ont rédigé, depuis 1999, leurs mémoires de maîtrise sur le corpus des chartes de Meurthe-et-Moselle (dans l'ordre chronologique : Jason D. Stein, Frédérique Gisquet, Delphine Harmand, Séverine Constans et Claire Muller) ; à Zurich, la thèse de doctorat d'Anne-Christelle Matthey (soutenue en 2007) a porté sur le corpus de la Meuse ; à l'École nationale des chartes, plusieurs thèses ont porté sur des documents d'oc (celle d'Aubin Leroy sur le cartulaire du consulat de Limoges, en 2005, et celle de Pascaline Todeschini sur les actes occitans dans le comté de Toulouse aux XI^e et XIII^e siècles, en 2005).

9. L'apport financier principal provient du Fonds national de la recherche scientifique suisse qui a financé les thèses de Julia Alletsgruber et de Dumitru Kihai, en cours d'achèvement, ainsi que celle de Claire Muller, soutenue en octobre 2010 ; depuis 2009, l'habilitation de Paul Gévaudan se place dans ce même cadre ; par ailleurs, à l'université de Zurich mes assistants Claire Vachon, Sarah Tinner et, précédemment, Harald Völker ont été pleinement impliqués dans le projet ; l'École des chartes garantit, de son côté, des vacances complémentaires indispensables.

10. Voir P. Videsott, « À propos du plus ancien document ... », ci-après, p. 64-65.

L'apport linguistique du matériau documentaire ressort d'autant plus dans la comparaison avec d'autres genres textuels, ce qui nous a amenés à explorer parallèlement les textes littéraires du XIII^e siècle, réunis dans le *Corpus d'Amsterdam*. Avec Achim Stein et Pierre Kunstmann, nous avons réélaboré cette base importante, qui avait été établie sous la direction d'Antonij Dees et de Pieter van Reenen. Nos interventions ont porté autant sur les aspects philologiques que sur le balisage linguistique¹¹. Il est dorénavant possible d'interroger les formes du *Nouveau Corpus d'Amsterdam* d'après les quatre paramètres des dates et lieux des œuvres comme des dates et lieux des manuscrits, ce qui confère une nouvelle sûreté de jugement à cet excellent outil linguistique¹².

L'idée directrice et fédératrice de ces différents projets est d'ordre philologique : la linguistique de corpus appliquée à des textes anciens doit se prévaloir d'un soin particulier accordé à la base philologique. Un enracinement (ancrage) diasystématique solide des formes linguistiques nécessite notamment des transcriptions (semi-)diplomatiques de manuscrits définis¹³.

II. LA STRUCTURE DES DONNÉES TEXTUELLES.

1. *Les données textuelles des « Plus anciens documents »*. – L'idée fondamentale de Paul Meyer, poursuivie par ses successeurs dans la chaire de philologie romane à l'École des chartes, était d'étudier, département par département, les collections d'anciens documents en langue vernaculaire. Par cette voie, à terme, l'intégralité des documents intéressants seront vus,

11. Voir M.-D. Gleßgen et Claire Vachon, « L'étude philologique et scriptologique du *Nouveau Corpus d'Amsterdam* », dans *Actes du XXVI^e Congrès international de linguistique et philologie romanes (septembre 2010, Valencia)*, éd. Emili Casanova et al., à paraître ; ainsi que M.-D. Gleßgen et Xavier Gouvert, « La base textuelle du *Nouveau Corpus d'Amsterdam* : ancrage diasystématique et évaluation philologique », dans *Le Nouveau Corpus d'Amsterdam, actes de l'atelier de Lauterbad, 23-26 février 2006*, éd. P. Kunstmann et Achim Stein, Stuttgart, 2007, p. 51-84. Nos deux projets interviennent aussi dans le cadre du Consortium international pour les corpus du français médiéval (CCFM, <http://ccfm.ens-lsh.fr/>). Le consortium a pour objectif de réunir différents corpus sur une même plateforme pour permettre, à terme, des interrogations communes. Il s'agit d'abord, outre les *Plus anciens documents* et le *Nouveau Corpus d'Amsterdam*, des textes réunis par le Laboratoire de français ancien d'Ottawa, de la Bibliothèque du français médiéval (École normale supérieure, Lyon) et de la base textuelle du *Dictionnaire du moyen français* (ATILF, Nancy).

12. Voir M.-D. Gleßgen et A. Stein, « Resources and tools for analyzing Old French texts », dans *Romanistische Korpuslinguistik II...*, p. 135-145.

13. J'ai présenté cette argumentation sous une forme embryonnaire et accompagnée de réflexions sur la faisabilité d'un tel projet dans l'article « Das altfranzösische Geschäftsschrifttum in Oberlothringen : Quellenlage und Deutungsansätze », dans *Skripta, Schreiblandschaften...*, p. 257-294.

surtout si l'on prend aussi en considération les documents conservés en Belgique, en Suisse romande et en Angleterre. Il sera alors possible de regrouper nouvellement les documents selon leur lieu de genèse médiéval, en vue de leur exploitation linguistique¹⁴.

Pour tous les documents, il s'agit désormais de les éditer d'après des critères définis et sous un format XML cohérent¹⁵, d'identifier les paramètres diasystématiques, de décrire les éléments de contenu de langue et de les photographier pour permettre une édition en ligne avec image à l'appui.

Le tableau suivant répertorie les quatorze ensembles de la langue d'oïl aujourd'hui en chantier. Il ne prend pas en considération les cinq ensembles en suspens pour lesquels existent des travaux préliminaires, ni les départements du Midi – dont certains sont pourtant à l'étude –, ni encore la Suisse romande¹⁶, le Luxembourg¹⁷ ou la Belgique¹⁸, qui sont ou ont été traités par d'autres collègues. Il indique par ailleurs les éditeurs et adaptateurs principaux et le nombre de documents en cours d'édition :

14. Il s'avère toutefois que de nombreuses chartes sont conservées encore aujourd'hui dans leur région d'origine, ce qui facilite les analyses dès les premières phases du travail.

15. Le standard XML (*Extensible Markup Language*) garantit la longévité des données textuelles et ouvre de bonnes perspectives pour leur interrogation. Il est cohérent d'un point de vue informatique et il répond, de par ses qualités, aux exigences tant de l'édition que de la linguistique. Voir M.-D. Gleßgen, « L'élaboration philologique et l'étude lexicologique des *Plus anciens documents linguistiques de la France* à l'aide de l'informatique », dans *Frédéric Godefroy...*, p. 371-386 (cet article, deuxième présentation du projet en français, indiquant les critères d'édition et le jeu des balises, a été repris, avec peu de variations, en allemand : « Editorische, lexikologische und graphematische Erschließung altfranzösischer Urkundentexte mit Hilfe von TUSTEP : Stand der Arbeiten », dans *Drittes Trierer Urkundensprachenkolloquium...*, p. 91-107) ; id., « Bases de données textuelles et lexicographie historique : l'exemple des *Plus anciens documents linguistiques de la France* », dans *De la philologie aux nouveaux médias...*, p. 373-380 (paru également dans *Nuove riflessioni sulla lessicografia : presente, futuro e dintorni del Lessico Etimologico Italiano, atti del Seminario di Lecce, 21-22 aprile 2005*, éd. Marcello Aprile, Galatina, 2007, p. 157-167 : cette troisième présentation explique les critères d'édition et les éléments du tableau analytique : les paramètres du diasystème et les genres textuels).

16. *Documents en langue française antérieurs à la fin du XIV^e siècle conservés dans les cantons du Jura et de Berne*, éd. Ernest Schüle, Rémy Scheurer et Zygmunt Marzys, avec le concours de Nicolas Barré, volume préparé par Bernadette Gavillet, Paris, 2002.

17. G. Holtus, Anja Overbeck et H. Völker, *Luxemburgische Skriptastudien : Edition und Untersuchung der altfranzösischen Urkunden Gräfin Ermesindes (1226-1247) und Graf Heinrichs V. (1247-1281) von Luxemburg*, Tübingen, 2003 (*Beihefte zur Zeitschrift für romanische Philologie*, 316) ; H. Völker, *Skripta und Variation...*

18. Pierre Ruelle, *Chartes en langue française antérieures à 1271 conservées dans la province de Hainaut*, Paris, 1984 (*Documents linguistiques de la Belgique romane*, 1) ; Reine Mantou, *Chartes en langue française antérieures à 1271 conservées en Flandre orientale et Flandre occidentale*, Paris 1987 (même collection, 2).

Départements	Éditeurs [adaptateurs]	Nb. de doc.
<i>1. Volumes publiés :</i>		
Oise	Carolus-Barré ¹⁹ [Tock, Grühl]	202
Haute-Marne	Gigot ²⁰ [Tock, Kihai]	142
Vosges	Lanher ²¹ [Trotter]	285
Aube, S.-et-M., Yonne	Coq ²²	103
<i>2. Volumes nouveaux en cours :</i>		
Jura	Muller	105
Marne	Kihai	230
Meuse	Matthey ²³	250
Moselle	Pitz [†]	180
Nièvre	Alletsgruber	30
Haute-Saône	Muller	135
Saône-et-Loire	Alletsgruber	95
Chancellerie royale	Videsott	150 [+ 350]
<i>3. Volumes manuscrits en cours de réélaboration :</i>		
Meurthe-et-Moselle	Arnod ²⁴ ; Gleßgen	290
Nord	Mestayer [Solé, Brunner]	350

Les quatre premiers départements (à savoir les départements de Meurthe-et-Moselle, de la Meuse, de la Marne et de la Haute-Marne, soit au total 1 100 chartes) ont été mis en ligne en janvier 2009²⁵. Dans les quatre années à venir suivront les neuf autres départements actuellement en chantier (sections 1, 2 et 3), à commencer par le Jura et la Haute-Saône, la Nièvre et la Saône-et-Loire. Une fois cet ensemble réuni, le total atteindra environ 2 500 chartes. Notons que le degré d'élaboration des volumes

19. Louis Carolus-Barré, *Les plus anciennes chartes en langue française*, t. I : *Problèmes généraux et recueil des pièces originales conservées aux archives de l'Oise*, Paris, 1964.

20. Volume cité note 3.

21. Jean Lanher (éd.), *Chartes en langue française antérieures à 1271 conservées dans le département des Vosges*, Paris, 1975 (*Documents linguistiques de la France, série française*, 2).

22. Dominique Coq (éd.), *Chartes en langue française antérieures à 1271 conservées dans les départements de l'Aube, de la Seine-et-Marne et de l'Yonne*, Paris, 1988 (*Documents linguistiques de la France, série française*, 3).

23. A.-C. Matthey, *Les plus anciens documents linguistiques de la France : le cas du département de la Meuse*, thèse de doctorat, Université de Zurich, septembre 2007, dactyl., 1 vol. d'étude et 2 vol. d'édition ; présentations dans *Romanistische Korpuslinguistik II...*, p. 123-134, et dans « De la philologie aux nouveaux médias... », p. 421-433.

24. Michel Arnod, *Publication des plus anciennes chartes en langue vulgaire antérieures à 1265 conservées dans le département de Meurthe-et-Moselle*, thèse de 3^e cycle, Université Nancy-II, 1974, dactyl.

25. En ligne : www.mediaevistik.uzh.ch/docling.

déjà publiés sous forme imprimée est moins poussé que celui des nouveaux volumes, puisque les principes éditoriaux et de description ont considérablement évolué depuis les années 1980²⁶.

Le tableau fait ressortir clairement la difficulté inhérente à l'harmonisation et à l'élaboration cohérente de ces collections. De nombreux auteurs et adaptateurs, aux statuts très divers, interviennent sur des éditions qui se trouvent, elles, dans un état d'élaboration et d'avancement très variable. Même si l'adaptation d'un ancien volume de la série offre des garanties complémentaires dans la qualité de l'édition, il s'est avéré que l'élaboration d'une charte d'après nos critères et à partir d'une transcription dactylographiée des années 1980 coûte presque autant de temps que la nouvelle édition d'une charte non publiée, qui peut s'inscrire immédiatement dans la logique actuelle du projet (voir l'étude de cas ci-après, « La terre lorraine... »). L'importance des efforts nécessaires pour intégrer dans l'édition électronique des transcriptions anciennes, pourtant de très bonne qualité, nous a amenés dans certains cas à renoncer à des textes déjà transcrits ; pour cette même raison, l'avenir des volumes manuscrits en suspens n'est pas mieux établi, à l'heure actuelle, que celui des départements de l'Ouest et du Midi.

Pour ne pas surcharger la lecture, le tableau ne comporte pas les dates extrêmes des différents corpus, pourtant significatives puisque les documents ne couvrent pas partout la même période. Pour la Meurthe-et-Moselle, très riche en documents français anciens, Michel Arnod avait arrêté les transcriptions, à juste titre, en 1265, comptabilisant déjà jusque-là 290 chartes. En revanche, pour les départements franc-comtois et bourguignons, les dates extrêmes ont dû être repoussées jusqu'en 1290, voire 1330, pour pouvoir documenter les débuts de l'écrit français dans la région²⁷. Les collègues travaillant sur les documents de la Suisse romande ont avancé jusqu'au xiv^e siècle (voir note 16).

Ces décalages créent des difficultés d'interprétation, mais ils correspondent à une réalité historique multiforme : on ne trouve pas, dans toutes les régions de la Romania, à une époque donnée, les mêmes genres textuels, ce qui complique la constitution de corpus équilibrés et l'étude des genres dans une optique comparatiste. Mais nous pourrions étudier malgré tout les débuts de l'écrit documentaire dans les régions en question et identifier les voies de diffusion des modèles linguistiques.

2. *L'élaboration informatique des données textuelles et linguistiques.* – L'édition des *Plus anciens documents* est accompagnée, depuis ses débuts,

26. École nationale des chartes, *Conseils pour l'édition des textes médiévaux*, t. I : *Conseils généraux*, Paris, 2005 ; t. II : *Actes et documents d'archives*, Paris, 2001.

27. Voir Julia Alletsgruber, « Approche du lexique de l'agriculture dans des chartes françaises du xiii^e siècle », dans *Actes du XXV^e Congrès international de linguistique et philologie romanes*, Tübingen, 2010, t. VI, p. 3-10.

d'analyses lexicologiques, onomastiques et graphématiques qui enrichissent les données textuelles. L'étude des lexèmes et des toponymes est indispensable, ne serait-ce que pour la bonne compréhension et l'établissement du texte. Les éléments graphématiques et, dans une certaine mesure, morphologiques contribuent avec l'analyse paléographique à l'identification du lieu de genèse des documents (« lieux d'écriture » ou « rédacteurs »). Ce sont donc des analyses primaires, indispensables dans cette entreprise.

Les données lexicologiques, onomastiques, graphématiques et morphologiques sont répertoriées dans une base de données interprétative, interdépendante avec la base de données textuelles²⁸. La configuration informatique des différentes bases repose sur de nombreuses listes établies, en dernière instance, à partir des mots graphiques numérotés dans la base textuelle encodée en XML.

Les données textuelles sont organisées par fichiers individuels regroupant des ensembles thématiques (par exemple un fichier pour les 289 chartes de Meurthe-et-Moselle, un autre pour les 250 chartes de la Meuse, etc.). L'édition de chaque charte (cernée de balises <txt>...</txt>) est accompagnée par un tableau analytique individuel (<an>...</an>), regroupant toutes les informations pertinentes pour la description diasystématique et les premiers éléments de la description historique²⁹.

(1) fichier textuel :

```
<gl>
<an> [date, rédacteur, etc.]</an>
<txt> <div n="1"> Le chapitre et li abbes de Salival (...) </div>
</gl>
```

Les formes graphiques du fichier textuel sont toutes entourées d'une balise portant un numéro univoque (<wn n="1">); à partir de ce numéro, elles sont enrichies, au fur et à mesure, par des informations linguistiques diverses placées dans la base de données interprétative : le lemme (lexical ou onomastique) auquel une forme peut être rattachée, sa catégorisation et description morphologique, les caractéristiques graphématiques ou, éventuellement, des variables dans le marquage morphologique (forme de l'article, etc.) :

(2) fichier textuel numéroté :

```
<gl>
<an>[date, rédacteur, etc.]</an>
```

28. Sur les problèmes soulevés par la lemmatisation, voir M.-D. Gleßgen, « La lemmatisation de textes d'ancien français : méthodes et recherches », dans *Ancien et moyen français sur le Web...*, p. 55-75. Le lemmatiseur est présenté dans M.-D. Gleßgen et Matthias Kopp, « Linguistic annotation of texts in non-standardized languages : the program procedures of the tool PHOENIX », dans *Romanistische Korpuslinguistik II...*, p. 147-154.

29. M.-D. Gleßgen, « L'élaboration philologique ... », p. 371-386 ; id., « Bases de données... ».

```
<txt> <div n="1"> <wn n="1">Le</wn> <wn n="2">chapitre</wn> <wn
n="3">et</wn> <wn n="4">li</wn> <wn n="5">abes</wn> <wn n="6">de</wn>
<wn n="7">Salival</wn> (...) </txt>
</gl>
```

(3) informations dans la base de données interprétative :

```
<wn>6</wn>: <src>abes</src> <lex f="c">abbé</lex> <graph f="(.)"> <morph
f="(.)"> <sem f="(.)">
```

Il est possible, grâce à un programme spécifique, d'importer les informations linguistiques du fichier-index dans le fichier textuel pour permettre des interrogations en texte plein :

(2 + 3) fichier textuel avec les informations du fichier-index :

```
<gl>
<an>[date, rédacteur, etc.]</an>
<txt> <div n="1"> Le chapitre et li <wn n="5"> <idx><lex f="c">abbé</lex>
<graph f="(.)"> <morph f="(.)"> <sem f="(.)"> </idx> abes</wn> de Salival (...)
</txt>
</gl>
```

Enfin, dans la base de données interprétative (3), les éléments regroupés auparavant sont réunis (par exemple toutes les formes appartenant à un lemme ou toutes les formes pour une variable graphématique) et peuvent être classés, décrits et commentés; voici quelques éléments d'entrée (simplifiés) pour *abbé* dans la version primitive de la base de données :

```
<entry s="abbe" t="00" i="c">
<name> abbe </name>
<name_typ> n. m. </name_typ>
<descr f="short"> abbé </descr>
<etymon> ABBAS </etymon>
</entry>
<group s="abbe" t="01" r="001 001 001 002" g="0" i="c">
<name> abbe(i) </name>
<ruler> n. m. sg. reg. </ruler>
<occur>
<occ id="555550005" div="3" wnr="2942"></occ>
<occ id="555550013" div="3" wnr="3012"></occ>
(...)
</occur>
</group>
```

L'architecture de la base de données a connu une transformation intégrale, après le premier achèvement, en 2009. Nous avons abandonné le

logiciel initial *Tustep*³⁰, malgré ses excellentes fonctionnalités, au profit d'une base de données MySQL, plus répandue et plus facile à rapprocher d'autres projets en philologie informatique³¹. La structure de cette base n'est pas banale mais elle est très clairement dessinée. Son architecture prévoit notamment des élargissements et même des transformations de la base de données textuelles, ce qui permet des corrections dans le texte au cours de l'analyse, opération impossible dans bon nombre de bases relationnelles actuelles.

3. *Formes de publication prévues.* – La publication des bases textuelles des *Plus anciens documents* est prévue, essentiellement, sous une forme informatique, tout comme les textes du *Nouveau Corpus d'Amsterdam*, déjà disponibles sur le web (www.uni-stuttgart.de/lingrom/stein/corpus). Nous avons renoncé, pour la collection complète, à une publication systématique sur papier qui se serait inscrite dans l'ancienne collection des *Plus anciens documents linguistiques de la France*. Le fait de travailler sur plusieurs départements voisins à la fois, dont les rédacteurs (lieux d'écriture) se recoupent, aurait rendu incohérente la publication traditionnelle par volumes, dédiés chacun à un département. La publication en ligne permettra néanmoins une impression de qualité de toutes les chartes, organisées d'après des critères divers.

La publication en ligne est évolutive. Chaque nouveau sous-ensemble départemental augmentera la base textuelle et, par là, les témoignages pour les différents lieux d'écriture. L'édition électronique permettra bientôt aux utilisateurs de choisir entre une présentation diplomatique ou interprétative des textes ; la version qui est proposée pour l'impression prévoit un système mixte d'encodage double (voir ci-dessous, III, 2, 1). La forme informatique intègre les photographies des documents, ce qui en augmente l'utilité. Enfin, la version en ligne des textes sera liée, à partir de 2010, aux bases de données linguistiques ; celles-ci resteront modestes au début mais évolueront au fur et à mesure, sous forme de répertoires, voire de dictionnaires et d'une base de données graphématiques.

III. LES DONNÉES ET INTERPRÉTATIONS LINGUISTIQUES.

1. *Les réalisations linguistiques et philologiques ciblées.* – L'élaboration des données textuelles s'accompagne d'emblée d'éléments d'interprétation linguistique. La réalisation de ces derniers représente l'investissement

30. En ligne : www.uni-tuebingen.de/zdv/zrinfo/tustep-des.html ; le programme Tustep, moins répandu que Perl ou LaTeX, est pourtant utilisé par de très nombreux projets d'édition.

31. Voir l'histoire de ces transformations dans M.-D. Gleßgen, « Le long chemin d'une charte vers l'ère multimédiale », dans *Festschrift für Günter Holtus*, éd. Anja Overbeck, Wolfgang Schweickard et Harald Völker (sous presse).

majeur à l'intérieur de nos projets, autant en temps qu'en termes de réflexion. Les quatre bases de données interprétatives en construction (lexicologique, onomastique, graphématique et morphologique) sont réunies dans une même architecture informatique (voir ci-dessus, II, 2) ; elles correspondent cependant à différentes interrogations et font appel à différentes méthodologies :

(1) Les variations graphématiques et morphologiques sont interprétées dans une optique scriptologique, dont l'analyse est indispensable pour l'identification des lieux d'écriture. Les éléments d'analyse graphématique entreront dans différents travaux en chantier et déboucheront aussi sur des études monographiques d'interprétation scriptologique.

(2) Le vocabulaire des documents est lemmatisé (pour les mots lexicaux) et certains segments sont étudiés de façon monographique. La première étude, par Julia Alletsgruber, concerne le vocabulaire du monde agricole, particulièrement présent dans les chartes ; les fiches lexicologiques, très développées, seront une partie intégrante de sa thèse en cours³². Sur la base de certaines de ces entrées, nous avons étudié le marquage diasystématique dans le vocabulaire médiéval³³. Par ailleurs, nous envisageons d'intégrer les formes du corpus lemmatisés dans la nouvelle version électronique du DEAF³⁴ qui servirait alors de plateforme d'accès et formerait un supplément substantiel au *Dictionnaire* de Godefroy.

(3) Les noms de personne et, surtout, les toponymes sont traités d'après des méthodes lexicologiques, en vue d'un premier dictionnaire onomastique informatisé pour la *Galloromania*, petit mais évolutif. Parallèlement au volume lexicographique, un deuxième volume imprimé pourrait réunir les entrées onomastiques contenues dans le même corpus textuel.

(4) La syntaxe est analysée tout d'abord dans une optique de structuration textuelle à partir de l'utilisation des temps dans les phrases principales

32. Voir J. Alletsgruber, « Approche du lexique de l'agriculture... », et ci-après, Annexe, p. 83-94.

33. Voir M.-D. Gleßgen, « Realia und Urkunden : die Teilung eines lothringischen Stadthauses kurz nach 1400 », dans *Romanische Sprachwissenschaft, Zeugnisse für Vielfalt und Profil eines Faches : Festschrift für Christian Schmitt zum 60. Geburtstag*, éd. Alberto Gil et al., Francfort (etc.), 2004, p. 423-447. En outre, id., « Vergleichende oder einzelsprachliche historische Textwissenschaft », dans *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten ?*, éd. Wolfgang Dahmen et al., Tübingen, 2006 (*Romanistisches Kolloquium*, 20), p. 319-340 ; id., « Diskurstitionen zwischen pragmatischen Regeln und sprachlichen Varietäten », dans *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*, éd. Angela Schrott et H. Völker, Göttingen, 2005, p. 207-228 (les deux articles précédents sont synthétisés en français dans « Corpus historiques et pragmatique : genres textuels et variétés linguistiques », dans *Romanistische Korpuslinguistik III...*). Voir enfin ci-après mon article « La terre lorraine, les puissants et les hommes : étude de la plus ancienne charte française du *scriptorium* épiscopal de Toul (1237) », p. 25-46.

34. Voir Sabine Tittel, « DEAF électronique », un avenir pour la lexicographie », dans *Revue de linguistique romane*, t. 74, 2010, p. 301-311.

et subordonnées. C'est le sujet de la thèse récemment soutenue par Claire Muller³⁵. Les recherches actuelles portent sur la formation de systèmes syntaxiques complexes, observable dans les chartes et qui se trouve en interaction avec l'élaboration des normes grammaticales dans l'ancienne langue ; l'étude de Paul Gévaudan sur la phrase complétive prend en considération notamment la structure de dialogue sous-jacente des chartes, liée aux constellations communicatives et donc médiales dans lesquelles elles ont été conçues.

Les analyses linguistiques sont accompagnées d'une tentative de description paléographique du support des chartes, dans l'optique de leur médialité. La thèse de Dumitru Kihai, notamment, proposera des éléments de structure pour l'analyse paléographique, indispensable comme l'étude graphématique pour l'identification des différents *scriptoria* et chancelleries³⁶.

2. *Les objectifs méthodologiques en linguistique historique.* – Les éditions et analyses présentées ici poursuivent différents objectifs d'ordre méthodologique, qui font et feront l'objet de publications thématiques. Voici les objectifs les plus saillants :

(1) Contribuer à une norme satisfaisante dans les principes d'édition. La tradition romaniste nous livre les meilleurs modèles pour les éditions de texte mais, malheureusement, nous sommes loin de voir ces modèles appliqués dans chacune des nouvelles éditions qui voient le jour. Un élément novateur de nos projets est le principe de l'encodage double, intégrant dans une même édition (papier ou web) des éléments de structure (majuscules, ponctuation) médiévaux et modernes³⁷. Le jeu de balises que nous avons mis au point pour la saisie d'un texte pourra être utile à d'autres éditeurs qui voudront s'inspirer de ces mêmes normes.

Jusqu'ici, nous n'avons travaillé que sur les transcriptions d'un manuscrit défini. La prise en considération conjointe de plusieurs témoins textuels d'une même œuvre complique considérablement les choses, notamment pour créer des liens opératoires entre les différentes versions. Dans ce domaine, la réflexion reste ouverte.

La variation des genres textuels introduit d'autres impératifs difficiles à concilier. Nous considérons que la TEI (*Text Encoding Initiative*) est trop

35. Voir Claire Muller, « Analyse textuelle et analyse phrastique du dispositif dans les plus anciennes chartes françaises : les relations entre structure interne et genre textuel », dans *Actes du XXV^e Congrès...*, t. VI, p. 249-258.

36. D. Kihai, « Les lieux d'écriture dans le domaine oriental d'oïl : les paramètres extralinguistiques de leur positionnement dans le réseau scriptologique », dans ce fascicule, p. 47-60.

37. Voir les exemples illustrés dans ce fascicule, « La terre lorraine... », II, et P. Videsott, « À propos du plus ancien document... », IV ; voir aussi M.-D. Gleßgen, « L'élaboration philologique... », p. 371-386 ; id., « Editorische, lexikologische und graphematische Erschließung... », et id., « Bases de données... ».

lourde pour pouvoir être utilisée par la majorité des philologues élaborant des éditions de texte. Il faudra trouver des compromis viables, ce qui sera néanmoins nettement plus facile que la gestion des différentes versions d'un texte.

(2) Préciser et rendre opératoire la notion de « lieux d'écriture » dans l'historiographie linguistique. L'analyse des chartes lorraines nous a amené, depuis 2002, à prendre en considération les « lieux d'écriture », c'est-à-dire les *scriptoria* et les chancelleries, comme entités à part entière dans le processus de l'élaboration de l'écrit, au moins médiéval³⁸. Sur la base des résultats obtenus, nous supposons que la dimension du lieu d'écriture doit être placée entre l'individualité des scribes et la norme abstraite d'une variété régionale ou sociale de la langue. Dans une assez forte mesure, les scribes s'adaptent à des normes reconnaissables quand ils travaillent dans un lieu d'écriture défini. Cela concerne autant la mise en page que le choix des graphèmes. Le comportement des scribes semble alterner, par ailleurs, selon le genre textuel. Mais sans entrer dès maintenant dans la description de ces micro-normes, il semble important d'identifier le lieu et l'institution concrets auxquels se rattache la rédaction d'un texte.

Le lieu d'écriture contient en même temps une dimension dans l'espace et dans l'univers sociologique ; le prestige linguistique d'une charte épiscopale de Toul est radicalement différent de celui d'une charte rédigée par un scribe libre travaillant dans la même ville. Cela complique d'ailleurs considérablement la représentation cartographique des formes de *scriptae* médiévales, qui devra intégrer par conséquent une dimension diastatique.

(3) Établir une structuration linguistique des genres textuels. Les chartes constituent un ensemble relativement homogène parmi les genres textuels. Il est naturellement possible et utile de distinguer des sous-genres, mais la variation linguistique concerne plutôt les différentes parties des chartes (*intitulatio*, *publicatio*, *corroboratio*, *datatio* / *expositio*, *dispositio*, clauses) que les différents types de contenu (vente, arbitrage, contrat féodal)³⁹. Le *Nouveau Corpus d'Amsterdam* introduit plus clairement la nécessité de réfléchir sur une structuration linguistique possible pour les genres textuels qui dépasse les catégories traditionnelles, d'origine littéraire. Les genres n'existent pas de manière absolue mais dépendent de l'optique de l'observateur ; d'un point de vue linguistique, il est légitime d'identifier comme genre un ensemble textuel qui se caractérise par un nombre défini de variables linguistiques. Cette recherche en est encore à ses débuts mais elle est prometteuse.

38. Voir mon étude monographique à partir du corpus de Meurthe-et-Moselle : M.-D. Gleßgen, « Les "lieux d'écriture" dans les chartes lorraines du XIII^e siècle », dans *Revue de linguistique romane*, t. 72, 2008, p. 413-540.

39. C. Muller, « Analyse textuelle... ».

(4) Garantir un « ancrage » diasystématique détaillé des formes linguistiques. Nos éditions fournissent pour tout document une description diasystématique détaillée. Celle-ci comporte le temps (éventuellement dédoublé entre l'époque de la genèse d'une œuvre et la rédaction d'une copie peu postérieure), l'espace (lui aussi dédoublé si nécessaire), le prestige social, le lieu d'écriture (qui synthétise en un certain sens l'espace et le prestige social), éventuellement le scribe ou l'auteur, de même que le genre textuel (qui traduit en même temps l'ancrage pragmatique et qui intervient dans le marquage diaphasique du texte).

Chaque forme linguistique individuelle est donc liée à ces paramètres qui permettent une description (un ancrage) diasystématique exhaustive (voir ci-dessus, II, 2). C'est seulement sur une telle base qu'il est possible de concevoir une historiographie linguistique qui prenne en compte le diasystème de la langue ou, du moins, de l'écrit.

Notre jeu de balises prévoit une entrée pour chacun des paramètres diasystématiques, et celles-ci sont pleinement prises en compte dans nos bases de données interprétatives. L'introduction de ces paramètres dans le *Nouveau Corpus d'Amsterdam*, jusqu'ici partielle, s'est révélée très utile pour les interrogations lexicologiques.

3. *Les développements informatiques*⁴⁰. – Un volet à part entière du projet est le développement du logiciel *Phoenix*, conçu en collaboration avec Matthias Kopp (Tübingen), Matthias Osthof (Tübingen et Zurich) et Paul Gévaudan qui en ont réalisé la programmation (d'abord avec le programme *Tustep*, ensuite avec *Perl*). Ce logiciel réunit un programme d'édition, un lemmatiseur (qui permet aussi le groupement de variables graphématiques et morphologiques), l'outil de saisie des bases de données interprétatives et un programme de représentation de ces bases.

Parallèlement, nous utilisons un éditeur XML pour la vérification du jeu de balises (par un schéma) et pour des interrogations à l'aide du langage *X-Query*⁴¹. Sont prévues en outre des interfaces avec le *TreeTagger*, baliseur morphologique très performant, programmé par Heinrich Schmid et Achim Stein.

Les outils développés sont maintenant disponibles, et seront téléchargeables à partir de 2011. Ce sont des réalisations utiles mais néanmoins

40. Voir aussi M.-D. Gleßgen, « Esigenze della tecnologia informatica nella filologia e lessicografia storica », dans *Nuovi media e lessicografia storica, atti del colloquio in occasione del settantesimo compleanno di Max Pfister*, éd. Wolfgang Schweickard, Tübingen, 2006, p. 15-24, et id., « Philologie und Sprachgeschichtsschreibung in der Romanistik : die "informatische Wende" », dans *Edition und Sprachgeschichte : Baseler Fachtagung, 2.-4. März 2005*, éd. Michael Stolz, Tübingen, 2007 (*Beihefte zur Editio*, 26), p. 201-212.

41. Malheureusement, la plupart des éditeurs XML sont de type commercial. Actuellement, nous travaillons avec le logiciel *Exchanger*, très satisfaisant et gratuit pour des usages universitaires.

exigeantes, dont l'emploi demande un effort indéniable. Un éditeur de texte et linguiste qui voudra utiliser ces outils devra :

- saisir ses données textuelles sous une forme XML ou les transformer en une forme XML ;
- adapter cette forme XML à notre jeu de balises ⁴² ;
- installer une petite série de programmes sur son ordinateur et apprendre un certain nombre de commandes pour pouvoir lancer les programmes ;
- comprendre le fonctionnement des différents volets qui constituent *Phoenix*, notamment le lemmatiseur et le programme des bases de données ;
- éventuellement, se familiariser avec *X-Query*.

Pour faciliter l'accès à ces outils, nous sommes en train de rédiger des documents descriptifs. Dans cette même perspective, les notes du présent article réunissent tous les textes rédigés jusqu'ici dans le cadre des projets ; les publications sont complémentaires et permettent d'accéder à la logique du projet des *Plus anciens documents*, y compris à celle des programmes informatiques. Malgré tous les efforts en faveur d'une meilleure accessibilité, il est probable qu'il ne sera pas possible de maintenir en deçà d'un certain degré la complexité d'utilisation de tels outils informatiques pour l'histoire linguistique.

4. *Finalités interprétatives ultérieures*. – Les finalités interprétatives se placent essentiellement dans l'optique du changement linguistique, entendu comme un outil pour comprendre le fonctionnement de la langue. Nos projets impliquent notamment le rôle de la variation diasystématique et celui des genres textuels dans le fonctionnement et le changement linguistiques. Ce sont ces mêmes paramètres qui permettent de mettre en relief le rôle de la langue en tant que média. Les chartes créent par ailleurs un lien particulièrement étroit avec l'infrastructure politique dans laquelle elles évoluent, ce qui permet de mieux cerner l'impact de l'ancrage pragmatique et des phénomènes d'une médialisation complexe sur le développement des genres textuels et sur l'élaboration linguistique.

Il est vrai que le chemin à parcourir entre les recherches empiriques et leur interprétation finale est très long. Dans ce sens, le projet des *Plus anciens documents* a l'inconvénient de demander, dans l'élaboration primaire des données linguistiques, des efforts qui pourront paraître démesurés à certains, mais ils ont l'avantage unique de fournir des données suffisamment complexes pour permettre de véritables conclusions. Les lourdeurs inévitables de la matière garantissent en même temps des résul-

42. M.-D. Gleßgen, « L'élaboration philologique... », p. 371-386 ; id., « Bases de données... ».

tats novateurs et solides dans les différents domaines interprétatifs, au profit de l'histoire de la langue, de l'histoire des textes et de l'histoire tout court.

Martin-D. GLESSGEN.