

# FRÉDÉRIC GODEFROY

*Actes  
du X<sup>e</sup> colloque international  
sur le moyen français*

organisé à Metz du 12 au 14 juin 2002  
par le centre « Michel Baude, littérature et spiritualité »  
et par l'ATILF (UMR 7118)

Textes réunis et présentés  
par  
FRÉDÉRIC DUVAL

PARIS  
ÉCOLE DES CHARTES  
2003

L'ÉLABORATION PHILOLOGIQUE  
ET L'ÉTUDE LEXICOLOGIQUE DES *PLUS ANCIENS*  
*DOCUMENTS LINGUISTIQUES DE LA FRANCE*  
À L'AIDE DE L'INFORMATIQUE

par

MARTIN-DIETRICH GLEßGEN

Ce projet a pour ambition l'élaboration électronique des sources romanes anciennes à des fins d'histoire linguistique. Lancé en 1998, il s'inscrit aujourd'hui dans l'entreprise des *Plus anciens documents linguistiques de la France*, inaugurée par Jacques Monfrin et poursuivie par Françoise Vielliard et Olivier Guyotjeannin<sup>1</sup>. La base de données repose actuellement sur un corpus de 290 chartes lorraines du XIII<sup>e</sup> siècle (1231-1265) qui avaient fait l'objet d'une transcription de qualité et d'une étude très solide par Michel Arnod en 1974<sup>2</sup>.

Le projet entend répondre à différents objectifs :

1. En premier lieu, il s'agit de préparer une édition fiable et moderne d'un corpus de textes non-littéraires en ancien français, inédits à ce jour. Cette édition alliera une version imprimée traditionnelle à une édition électronique qui permettra des choix entre une représentation plus diplomatique ou plus

1. Cf. la profession de foi de Jacques Monfrin, « Le mode de tradition des actes écrits et les études de dialectologie », dans *Revue de linguistique romane*, t. 32, 1968, p. 17-47 (rééd. dans *Les dialectes de France au Moyen Âge et aujourd'hui*, colloque organisé à Strasbourg du 22 au 25 mai 1967, éd. Georges Straka, Paris, 1972, p. 25-55) ainsi que son introduction au premier volume de la collection : « Les études sur les anciens textes gallo-romans non littéraires. Le recueil des documents linguistiques de la France », dans *Chartes en langue française antérieures à 1271 conservées dans le département de la Haute-Marne*, éd. Jean-Gabriel Gigot, Paris, 1974, p. XI-LXXIX.

2. Michel Arnold, *Publication des plus anciennes chartes en langue vulgaire antérieures à 1265 conservées dans le département de Meurthe-et-Moselle*, thèse dactylographiée, Nancy, 1974.

interprétative du texte transcrit. La version électronique intégrera aussi les photographies des documents.

2. Les textes édités devront être accompagnés d'une analyse grapho-phonétique, lexicologique et onomastique ainsi que morpho-syntaxique. Je souhaiterais donc développer un instrument informatique adapté à l'édition et l'analyse linguistique de textes romans anciens en général et de textes documentaires français en particulier. Les logiciels proposés actuellement sur le marché ne sont pas satisfaisants en raison de leur caractère commercial. Ce n'est certes pas là une tâche habituelle pour un historien de la langue, mais ces instruments informatiques nous sont devenus indispensables.

3. Au-delà de ces deux objectifs concrets, le projet devrait contribuer à la réflexion sur la formation des langues écrites dans la Romania. La nature particulière des sources que sont les chartes s'y prête particulièrement bien : leur caractère d'originaux les distingue des textes littéraires, techniques et scientifiques ; il permet un meilleur ancrage dans ce que la linguistique variationnelle appellerait le « diasystème de l'écrit » et un suivi plus fidèle des transformations linguistiques à travers le temps. La préstandardisation de la langue française entre le XIII<sup>e</sup> et le XV<sup>e</sup> siècle comporte notamment des phénomènes de normalisation graphématique et morphologique ainsi que d'élaboration lexicale et syntaxique qui pourront être décelés dans les textes étudiés.

4. Enfin, ce projet conduit à s'interroger sur la méthodologie à adopter en historiographie linguistique romane. La forte variation de la langue ancienne en fonction de la nature des sources écrites plaide pour l'utilité d'une « typologie linguistique des genres textuels » qui n'en est actuellement qu'à ses débuts. C'est seulement si l'on arrive à distinguer ce qu'est un changement linguistique lié à un genre textuel ou à un mode de transmission qu'il est possible d'établir, *ex negativo*, ce qu'est vraiment un changement dans le temps ou dans l'espace<sup>3</sup>.

3. Une première présentation du volet philologique du projet est fournie dans une publication en ligne, « L'écrit documentaire français dans la Lorraine médiévale : quelques réflexions méthodologiques », dans *Philology in the Internet Era / Philologie à l'ère de l'Internet*, qui réunit les interventions du Colloque de Budapest (7-11 juin 2000), organisé par Iván Horváth [<http://mayar-irodalom.elte.hu/colloquia/>]. — Le cadre historique et philologique général a été décrit dans « Das altfranzösische Geschäftsschrifttum in Oberlothringen : Quellenlage und Deutungsansätze », dans Kurt Gärtner *et al.* (éd.), *Skripta, Schreiblandschaften und Standardisierungstendenzen*, Beiträge zum Zweiten internationalen Urkundensprachen-Kolloquium vom 16.-18. September in Trier, Trier, 2001 (*Trierer Historische Forschungen*, 47), p. 257-294. Cet article de synthèse réunit l'essentiel de la bibliographie qui n'est donc pas reprise ici. — L'état actuel des principes philologiques ainsi que quelques éléments de l'analyse graphématique sont livrés – toujours en allemand – dans la brève contribution « Editorische, lexikologische und graphematische Erschliessung altfranzösischer Urkundentexte mit Hilfe von TUSTEP. Stand der Arbeiten », dans *Drittes Trierer Urkundensprachekolloquium* (20.-22. Juni 2001), éd. Kurt Gärtner et Günter Holtus, (sous presse).

## I. — L'ÉLABORATION PHILOLOGIQUE DES TEXTES

1. *Critères d'édition*

Les critères de l'édition électronique de notre corpus et – dans sa lignée – des *Plus anciens documents linguistiques de la France* entendent réaliser une synthèse entre l'ancienne et la nouvelle philologie : il faut rendre les textes anciens familiers et compréhensibles au lecteur moderne, ce qui oblige à des interventions qui en diminuent l'authenticité ; parallèlement, le linguiste doit pouvoir s'appuyer sur des reproductions très proches de l'original, si possible sous une forme informatique qui permet des quantifications.

De ce constat découlent les cinq plus importantes décisions éditoriales :

1. La ponctuation de l'original est maintenue ; elle est reproduite à mi-hauteur des lignes comme dans les documents authentiques. En complément, j'ai introduit une ponctuation moderne pour faciliter la lecture des textes ; celle-ci se trouve sur la ligne.

2. Les majuscules de l'original – dans la mesure où les majuscules peuvent être identifiées en tant que telles – sont en gras. Par ailleurs, j'ai introduit des majuscules selon l'usage moderne.

3. Des traits transversaux indiquent les sauts de ligne de l'original, des traits doubles chaque cinquième ligne ; celle-ci est numérotée automatiquement à côté du texte (lignes 5 et 10). Mais j'ai aussi introduit, par des chiffres arabes en gras, une structuration sémantique et syntaxique du contenu qui guide la compréhension du texte.

4. J'ai renoncé en revanche à la reproduction d'éléments calligraphiques : formes des lettres, abréviations, signes diacritiques, traits ou tildes à la fin de la ligne. Pour toutes les questions de paléographie ou de mise en page, le lecteur pourra se référer à la reproduction photographique de l'original dans l'édition électronique<sup>4</sup>.

5. Enfin, au-dessus du texte proprement dit apparaît un tableau analytique développé, comprenant l'indication de la date, le genre textuel, un résumé du contenu, les personnages concernés par le document, le rédacteur supposé, la description matérielle de la charte et le lieu de conservation. Ce tableau analytique replace le texte dans le diasystème en dégageant les paramètres diachroniques et diatopiques, mais aussi diastratiques (essai d'identification du rédacteur) et diaphasique (genre textuel). Le tableau est fondamental pour l'interprétation, puisque tout élément linguistique peut être recoupé à tout moment avec chacun des paramètres diasystématiques.

De tels principes éditoriaux permettent au lecteur d'avoir une image précise du document original sans renoncer pour autant à la facilité de lecture que

4. Cf. pour des questions paléographiques M.-D. Gleßgen, *Editorische, lexikologische und graphematische Erschließung...* [s.p.], n. 1 et 2.

procurent les interventions de l'éditeur moderne. Dans l'édition électronique le lecteur pourra choisir la représentation qu'il préfère :

- une reproduction mixte correspondant au modèle proposé par la suite ;
- une reproduction interprétative selon les principes des DocLingFr, ne contenant que les majuscules et la ponctuation modernes ;
- une reproduction diplomatique avec les majuscules et la ponctuation de l'original et structurée d'après les lignes de l'original, ce qui permet une comparaison immédiate avec la photographie ; dans ce cas, les apostrophes ou les séparations de mots introduites par l'éditeur peuvent être omises pour souligner les phénomènes d'agglutination.

À titre d'exemple nous reproduisons ci-contre l'interprétation mixte d'une des premières chartes de notre corpus. Elle nous semble réaliser un bon compromis pour une version imprimée.

Le contenu de cette charte peut être structuré de manière très cohérente d'après ses éléments de structure originaux. Chaque nouvelle entrée de la traduction en français moderne répond à un point dans l'original et/ou à une majuscule :

[majuscule = début] [point = résumé de l'acte juridique]	<b>1</b> Qu'il soit connu à tous <b>2</b> que l'abbé et le chapitre de Salival ont cédé à Wirrion et à Houillon, deux frères demeurant à Juvelize et fils de Bertran Bachelier, <b>3</b> treize journaux de terre en friche (= <i>treisse</i> ) <sup>5</sup>
[lieu] [durée]	<b>4</b> dans le territoire de Juvelize ; la cessation concerne aussi leurs héritiers ;
[cens] [cens en nature]	ceci pour la valeur de treize deniers de cens (annuel) et deux hémimes de grains, l'une d'avoine, l'autre de froment.
[condition]	<b>5</b> Au cas où ils ne payeraient pas le jour nommé à la fête de Saint Remy
[lieu de paiement] [conséquence] [spécification]	dans la maison (du chapitre) à Salival l'on prendrait comme gage la terre et ses fruits.
[point + majuscule : description du terrain]	<b>6</b> (Le terrain) est réparti de la manière suivante :
[première parcelle] [répartition entre les frères]	au lieu-dit Tramble se trouvent quatre journaux, dont un pour (Wirrion) seul,

5. Le terme – provenant du bas francique \**threosk* « Brachland / terre en friche » – appartient à l'ancien français nord-oriental (wallon, flandrin, champenois, lorrain, sous les formes *trie(z)* 1257, *trieu*, *trihe*, *triexhe*, *triot* et – seulement en Lorraine – *treixe* 1340 (FEW 17, 400b [définition peu heureuse de « terrain inculte »]). Notre charte fournit la première attestation de ce régionalisme. — Le syntagme *terre treisse* (qui n'a pas été retenu par le FEW) préfigure l'usage adjectival attesté uniquement en Lorraine et jusqu'ici exclusivement dans des dialectes modernes (« adj. Meuse *tríce*, Moselle *trix*, Metz *trex* », FEW, *ib.*).

002

1234 (25 mars-31 décembre) ou 1235 (1<sup>er</sup> janvier-24 mars)

Type de document : charte : acensement de terres

Objet : *L'abbé et le chapitre de Salival acensent à Wirrion et Houillon treize journaux de terre au finage de Juvelize contre un cens de treize deniers et deux hémines de grain ; les conditions de l'acensement sont très contraignantes pour les paysans.*

Auteur : non annoncé

Disposant : abbaye de Salival

Sceau : disposant

Bénéficiaire : disposant [la rédaction de la charte avantage surtout le chapitre]

Autres acteurs : Wirrion et Houillon, paysans de Juvelize

Rédacteur : scriptorium de l'abbaye de Salival [les paysans ne pouvaient pas disposer d'un scribe]

Parchemin jadis scellé sur simple queue ; 58 x 141

AD MM H 1244, fonds de l'abbaye de Salival

- 1 Conue chose soit a-toz 2 *que* li abes *et* li chapitles de Salinvas · at laissé a Wirion / *et* Huillon, les dous freres de Geverlise, les anfanz Bertran Bacheler, 3 ·XIII· jor/nas de terre treisse · en la fin de Geverlise · *et* a lor oirs · 4 *parmi* ·XIII· deniers de cens · *et* / ·II· himas de blef · l'un d'avoine ·
- 5 l'autre de froment · 5 *et* s'il ne paievent a jor // nomei a la feste sent Remi · a Giverlise, en la maison de Salinvas<sup>1</sup> · *que* l'on se tan/roit a la terre · *et* ce *que* sus averoit ·
- 6 Si est ensi devisee · q'au Tramble en / at ·III· jornas · un *par* lui<sup>2</sup> · *et* ·III· ensemble · 7 *et* en la voie de Hignicort en at / V· jornas, ·II· d'une part *et* ·III·
- 10 d'autre · 8 *et* en la voie de Marsal ·II· jornas · / après la terre les Vowes<sup>3</sup> · 9 *et* en la voie de Donneris · as Genovres · en at // ·II· jornas ·
- 10 Ci at mis li abes *et* li convenz de Salinvas son sael · en tesmoig/nage de verité · 11 l'an *que* li miliaires corroit *par* ·M· *et* CC· *et* XXXIII· anz ·

<sup>1</sup> L'abbaye possède donc une maison à Juvelize. <sup>2</sup> Probablement *Wirion*, le premier frère nommé dans le texte. <sup>3</sup> Les *Vowes* ou des *Vowes* ?

[id.]	les trois autres (pour les deux frères) en indivis ;
[deuxième parcelle]	7 et près du chemin qui mène à Hignycourt se trouvent cinq <sup>6</sup> journaux, deux pour un frère, trois pour l'autre ;
[troisième parcelle]	8 et près du chemin qui mène à Marsal deux journaux,
[localisation précise]	derrière le terrain dit « des Vowes » <sup>7</sup> ;
[quatrième parcelle]	9 et près du chemin qui mène à Donneroy
[localisation précise]	au lieu-dit Genouvres
[suite]	se trouvent deux journaux.
[point + majuscule : <i>corroboratio</i> ]	10 L'abbé et le couvent de Salival ont mis leur sceau,
[formule de <i>corroboratio</i> ]	en témoignage de vérité
[ <i>datatio</i> ]	11 en l'année 1234.

Les éléments de structure donnent des indications sur la manière dont les rédacteurs et scribes concevaient les unités de sens. Leur étude peut ainsi conduire à mieux connaître la conception et la conscience syntaxiques des professionnels de l'écrit médiévaux. C'est une particularité des chartes qui les distingue des textes littéraires ou scientifiques, normalement transmis sous forme de copies. Si la ponctuation suit une logique, il s'agit néanmoins d'une logique comportant de nombreux traits idiosyncrasiques qui ne sont reproduits que partiellement dans une copie. Ce phénomène empêche souvent d'interpréter de manière cohérente la ponctuation dans les manuscrits littéraires, où les systèmes d'écriture se chevauchent et sont difficiles à démêler.

## 2. Encodage des données textuelles

Les principes éditoriaux trouvent leur corollaire technique dans l'encodage informatique des données textuelles. Cet encodage est fondamental puisque nous ne possédons pas, à ce jour, de garantie sur la longévité des données informatiques. Pour minimiser les risques d'une perte des données, notre corpus suit les principes d'un encodage neutre de type XML, conforme aux propositions de la TEI (*Text Encoding Initiative*). Les différentes formules ou vues du texte présentées plus haut reposent toutes sur un seul fichier informatique et peuvent être facilement reproduites à partir de celui-ci sous un moteur de recherche. Par ailleurs, nos données sont saisies à l'aide de la plateforme extrêmement ouverte que constitue TUSTEP et qui sert aussi pour la programmation des interrogations linguistiques (= TUEbinger System von Textverarbeitungsprogrammen = système de logiciels textuels de Tübingen).

Le fichier informatique de base de la charte reproduite plus haut a la forme suivante :

6. Le point avant le chiffre a été omis à cause du saut de ligne.

7. Ou : le terrain appartenant à la famille Vowes.

```

<gl>
<t type ="123"/>
<id>555550002</id>
<zitf>002</zitf>
<an>
<nom>002</nom>
<d>1234 (25 mars-31 d%/ecembre) ou 1235 (1#'#'r janvier- 24
mars)</d>
<d0></d0>
<type>charte : acensement de terres</type>
<r>L'abb%/e et le chapitre de Salival acensent %\a Wirrion et
Houillon treize journaux de terre au finage de Juvelize
contre un cens de treize deniers et deux h%/emines de grain;
les conditions de l'acensement sont tr%\es contraignantes
pour les paysans.</r>
<aut>non annonc%/e</aut>
<disp>abbaye de Salival</disp>
<s>disposant</s>
<b>disposant [la r%/edaction de la charte avantage surtout le
chapitre]</b>
<act>Wirrion et Houillon, paysans de Juvelize</act>
<rd>scriptorium de l'abbaye de Salival [les paysans ne
pouvaient pas disposer d'un scribe]</rd>
<f>Parchemin jadis scell%/e sur simple queue; 58x141</f>
<l>AD MM H 1244, fonds de l'abbaye de Salival</l>
</an>
<txt>
<pub><div n=1> <maj>C</maj>onue chose soit a-toz</div> </pub>
<exp><div n=2> q<abr>ue</abr> li abes <abr>et</abr> li
chapitles de Salinvas /. at laissi%/e a Wirion <zw/>
<abr>et</abr> Huillon, les dous freres de
Gev<abr>er</abr>lise, les anfanz Bertran Bacheler,</div>
<div n=3> /.XIII/. jor<zwt/>nas de t<abr>er</abr>re treisse /.
en la fin de Gev<abr>er</abr>lise /. <abr>et</abr> a lor
oirs /.</div>
<div n=4> p<abr>ar</abr>mi /.XIII/. d<abr>eniers</abr> de cens
/. <abr>et</abr> <zw/> /.II/. himas de blef /. l'un d'avoine
/. l'autre de froment /.</div>
<div n=5> <abr>et</abr> s'il ne paievent a jor <zw/> nomei a
la feste sent Remi /. a Giv<abr>er</abr>lise, en la maison
de Salinvas <ful>L'abbaye poss%\ede donc une maison %\a
Juvelize.</ful> /. q<abr>u</abr>e l'on se tan<zwt/>roit a la
terre /. <abr>et</abr> ce q<abr>ue</abr> sus averoit
/.</div>
<par><div n=6> (...) </par> </exp>
<par><cor> <div n=10> <maj>C</maj>i at mis li abes
<abr>et</abr> li covenz de Salinvas son sael /. en
tesmoig<zwt/>nage de verit%/e /.</div></par> </cor>
<dat> <div n=11> l'an q<abr>ue</abr> li miliaires corroit
p<abr>ar </abr> /.M/. <abr>et</abr> CC/. <abr>et</abr>
XXXIIII/. anz /.</div></par> </dat>
</txt>
</gl>

```



La séparation du tableau analytique (codé par `<an> ~ </an>`) et du texte (codé par `<txt> ~ </txt>`) permet d'utiliser les données textuelles comme dans une base de données relationnelle et d'analyser tout élément linguistique à la lumière des quatre principaux paramètres du diasystème : le temps, le lieu, le rôle social du rédacteur et le genre textuel.

Plus en détail, les différentes balises, marquées par de petits crochets (`<>`) se divisent en trois ensembles :

1. Les éléments de structuration générale avec l'indication d'une entité textuelle `<gl> ~ </gl>`, le genre textuel `<t type = ...>` qui détermine la structure des balises, un numéro d'identification de cette même entité pour sa gestion informatique `<id>` et un sigle qui permet de citer brièvement le texte en question `<zitf>` :

<code>&lt;gl&gt;</code>	= début d'une entité
<code>&lt;t type = "123"/&gt;</code>	= genre textuel
<code>&lt;id&gt; &lt;/id&gt;</code>	= numéro d'identification
<code>&lt;zitf&gt; &lt;/zitf&gt;</code>	= sigle (« Zitierform »)
(...)	
<code>&lt;/gl&gt;</code>	= fin de l'entité

2. Le tableau analytique qui comprend les indications diasystématiques et la description philologique du texte :

<code>&lt;an&gt;</code>	= début du tableau analytique
<code>&lt;nom&gt; &lt;/nom&gt;</code>	= nom du texte (= ici : numéro de la charte)
<code>&lt;d&gt; &lt;/d&gt;</code>	= date
<code>&lt;type&gt; &lt;/type&gt;</code>	= type de document
<code>&lt;r&gt; &lt;/r&gt;</code>	= regeste
<code>&lt;aut&gt; &lt;/aut&gt;</code>	= auteur
<code>&lt;disp&gt; &lt;/disp&gt;</code>	= disposant (par exemple en cas de vidimus)
<code>&lt;s&gt; &lt;/s&gt;</code>	= sceau
<code>&lt;b&gt; &lt;/b&gt;</code>	= bénéficiaire
<code>&lt;act&gt; &lt;/act&gt;</code>	= autres protagonistes
<code>&lt;rd&gt; &lt;/rd&gt;</code>	= rédacteur
<code>&lt;sc&gt; &lt;/sc&gt;</code>	= scribe
<code>&lt;f&gt; &lt;/f&gt;</code>	= forme et description matérielle de la charte
<code>&lt;l&gt; &lt;/l&gt;</code>	= lieu de conservation
<code>&lt;ed&gt; &lt;/ed&gt;</code>	= édition éventuelle
<code>&lt;ana&gt; &lt;/ana&gt;</code>	= regeste ou analyse éventuels
<code>&lt;ec&gt; &lt;/ec&gt;</code>	= observations sur l'écriture
<code>&lt;met&gt; &lt;/met&gt;</code>	= observations sur la langue
<code>&lt;/an&gt;</code>	= fin du tableau analytique

3. Le texte lui-même est structuré selon ses entités sémantiques et/ou syntaxiques <div n = ...> et les sauts de lignes de l'original <zw/><sup>8</sup> (ou <zwt/> quand le saut de ligne implique une séparation du mot). D'autres balises marquent les abréviations <abr> et majuscules <maj> de l'original ou gèrent les deux séries de notes (<fue>, <ful>) :

<txt>	= début du texte
<div n = 1>	= chiffres gras pour la structuration sémantique du texte
<zw/> ~<zwt/>	= changement de ligne dans l'original (« Zeilenwechsel » + « Trennung »)
<par>	= nouveau paragraphe
. /, /; /.. /...	= signes de ponctuation originaux
. / ; / :	= ponctuation exclusivement moderne
<abr> </abr>	= abréviation dans l'original
<maj> </maj>	= majuscule dans l'original
<fue> </fue>	= notes éditoriales (série indiquée par a, b, c)
<ful> </ful>	= notes concernant le contenu de la charte (série indiquée par 1, 2, 3)
<v> </v>	= texte sur le verso
</txt>	= fin du texte

La transcription indique par ailleurs les séparations de mots (du type *a-savoir*) et les apostrophes introduites par l'éditeur ; ces éléments peuvent être omis dans une représentation diplomatique. S'ajoutent enfin des balises qui divisent les parties de la charte (*intitulatio* <int>, *publicatio* <pub>, *expositio* <exp>, *corroboratio* <cor>, *datatio* <dat>).

Cette structure informatique peut évoluer selon le texte étudié. Il est possible de déterminer pour tout genre textuel une série particulière de balises qui ne reprend que quelques éléments constitutifs du cadre général, comme les éléments de structure, la date ou les *divisiones*. À l'aide de l'indication <t type = ...>, le logiciel peut gérer alors les différentes séries de manière parallèle. Ce principe est plus flexible que le modèle de la TEI qui gère tous les textes avec une seule grande série de balises.

D'un point de vue philologique, cela veut dire qu'il est possible d'intégrer dans notre cadre des textes anciens très différents, des romans, de la poésie, des traités scientifiques ou religieux. À titre d'essai quelques textes ont déjà été adaptés à ce schéma, sur la base non pas d'éditions mais de transcriptions fidèles de manuscrits :

8. La forme de cette balise qui s'ouvre et se ferme en même temps est inhabituelle dans un schéma XML mais possible. Par ailleurs pour des raisons de convenance lors de la saisie, certaines balises ne correspondent pas, dans la version présentée, parfaitement aux règles de XML ni aux propositions de la TEI ; mais avec un petit programme déjà établi, la base peut être corrigée automatiquement.

- la traduction en moyen français du *De arte venandi cum avibus*, édité et étudié par Magali Grandval ;
- les huit manuscrits du *Chevalier de la Charrette* de Chrestien de Troyes (1177/81), transcrits et saisis en SGML par Karl Uitti (Princeton) [éd. Foulet-Uitti 1987] et réélaborés maintenant par Cinzia Pignatelli dans le cadre du Centre d'études médiévales de Poitiers ;
- les neuf manuscrits du *Charroi de Nîmes* (milieu du XII<sup>e</sup> siècle), transcrits très soigneusement par un groupe de travail sous la direction de Lene Schøsler (Copenhague) et Piet Van Reenen (Amsterdam).

De telles intégrations ouvrent de riches perspectives pour l'étude comparative des genres textuels en français médiéval. Avec Pierre Kunstmann et Achim Stein nous pourrions utiliser ce schéma pour une étude morphologique et lexicographique sur le « Corpus d'Amsterdam »<sup>9</sup>.

## II. — L'ANALYSE LEXICOLOGIQUE

Par la force des choses, mon projet connaît un glissement progressif de problématiques philologiques et informatiques, telles que l'édition, l'encodage ou le logiciel d'interrogation, vers des questions plus linguistiques, autant grapho-phonétiques que lexicologiques, morpho-syntaxiques et onomastiques. Celles-ci sont traitées au fur et à mesure. L'analyse grapho-phonétique constitue le premier volet de l'analyse linguistique puisqu'elle permet une identification des rédacteurs et donc l'ancrage diasystématique des chartes. Je n'insisterai pas ici sur cette partie de l'étude qui repose, d'un point de vue informatique, sur la même procédure que l'interrogation lexicologique<sup>10</sup>.

Le deuxième volet, lexicologique, connaîtra plusieurs étapes : il convient d'abord de regrouper les formes individuelles sous des lemmes cohérents. Pour distinguer des lemmes, j'ai adopté les mêmes principes étymologiques que ceux qui régissent la structure du PtRob ou le *Dictionnaire* de Godefroy. Les formes lemmatisées seront ensuite structurées, comme dans le DEAF, d'après leur forme et leur sens. Enfin, elles seront accompagnées par des commentaires lexicologiques et dictionnaires. L'ensemble de ces opérations permettra des interrogations sur le changement lexical qui prendront en compte la nature des sources et des données quantitatives.

Il a fallu développer des instruments qui facilitent la lemmatisation. Ainsi, notre logiciel fournit une concordance avec un contexte court pour chaque forme et la possibilité de consulter un contexte large, chaque fois que cela est nécessaire.

9. Cf. la contribution de P. Kunstmann et A. Stein dans ce même volume, p. 387-392.

10. Cf. l'exemple du -d- épenthétique (*semondre versus lorrain semondre*) présenté dans M.-D. Glebgen, *Editorische, lexikologische und graphematische Erschließung...* [s.p.], chap. 5.

En guise de prélemmatisation, le logiciel sépare d'emblée les noms propres des noms communs (sur simple repérage des majuscules de l'édition). Par ailleurs, il réduit la variance des formes en introduisant des équivalences graphiques que l'on peut choisir librement.

Dans notre corpus, *bb* double et *b* simple sont presque toujours interchangeables ; ils ne distinguent presque jamais deux mots différents et ne constituent donc pas une paire minimale. Cela vaut aussi pour les autres consonnes ou voyelles doubles : *bb-b*, *cc-c*, *dd-d*, *ff-f*, *ll-l*, *mm-m*, *rr-r*, *ss-s* ; *aa-a*, *ee-e*, *ee/-é/* [= *ee* ou *é* en fin de mot], *oo-o*.

Cette variance est d'ailleurs une caractéristique générale d'une écriture non standardisée ou dominée. D'autres « équivalences » graphiques sont plus subtiles et cachent différents types de variation linguistique. Dans notre corpus, nous pouvons relever quelques groupes consonantiques latinisants (*cq-q*, *ct-t*), quelques homophonies (latentes) (*en-an*, *z/-s/*, (*n*)*gni-ngn-gn*) et certains graphèmes redondants (*y-i*, *k-qu*, *x/-us/*). D'autres variantes sont dues à la présence simultanée d'évolutions phonétiques ou graphématiques régionales différentes dans le même corpus<sup>11</sup> :

*w-g(u)* (*wage-gage*), *ei/-é/* (*abbei-abbé*), *eir/-er/* (*torneir-torner*), *np-mp*  
(*Champaigne-Chanpaigne*)  
*au(v)le-auble-able* (*estauvle-astable-estable*), *atr-etr* (*lattes-lettres*)

Cette opération qui bien entendu, n'a dans un premier temps qu'une fonction pratique, réduit la variance graphique de notre corpus de moitié environ. L'entrée *abé* par exemple (« abbé ») rassemble les formes *abé*, *abbé*, *abei*, *abbei* et *abbey*.

Grâce à cette réduction des formes il devient relativement aisé de regrouper les mots sous un lemme, quitte à corriger l'un ou l'autre regroupement non étymologique. Après marquage avec la souris, le programme crée automatiquement des balises dans le texte de départ. La principale utilité de ce marquage est de maintenir une relation active entre le texte et la base lexicale (ou base grapho-phonétique) et de permettre ainsi des corrections manuelles dans l'édition même après avoir commencé l'analyse linguistique. Il est important de rappeler qu'un glossaire doit précéder l'établissement définitif d'un texte<sup>12</sup>.

11. La première orientation pour la variation régionale des graphies est toujours fournie par Carl Theodor Gossen, *Französische Skriptastudien. Untersuchungen zu den nordfranzösischen Urkundensprachen des Mittelalters*, Vienne, 1967 (*Sitzungsberichte der Österr. Akad. der Wissenschaften. Phil.-hist. Kl.*, 253), à compléter par Antonij Dees (et al.), *Atlas des formes et des constructions des chartes françaises du 13<sup>e</sup> siècle*, Tübingen, 1980 ainsi que *Atlas des formes linguistiques des textes littéraires de l'ancien français*, Tübingen, 1987. Les particularités lorraines ont été étudiées plus en détail par Max Pfister, « Nordöstliche Skripten im Grenzgebiet Germania-Romania (13./14. Jahrhundert) », dans K. Gärtner et al. (éd.), *Skripta, Schreiblandschaften und Standardisierungstendenzen...*, p. 223-244.

12. C'est l'enseignement de Frankwalt Möhren, *Édition et lexicographie*, dans *Alte und neue*

Cette procédure basique de lemmatisation peut être appliquée à tout texte qui respecte le balisage présenté ou qui est adapté à celui-ci. Une procédure plus élaborée de lemmatisation supposera au contraire la comparaison des formes textuelles avec un inventaire de formes lemmatisées ou à la morphologie définie. Pierre Kunstmann et Achim Stein ont présenté en ce même lieu un tel inventaire sur la base des formes du Godefroy, du Tobler-Lommatzsch et du « Corpus d'Amsterdam »<sup>13</sup>. Mais il faut savoir que même les formes regroupées de façon automatique doivent être ensuite contrôlées pour une lemmatisation donnée. C'est là où intervient l'interface que je viens de présenter.

Actuellement, Matthias Kopp, un collègue de Tübingen qui a rédigé l'essentiel du logiciel d'interprétation, et moi-même, préparons la prochaine étape de l'analyse linguistique : la structuration des formes retenues sous un lemme d'après des critères de forme et de sens. D'un point de vue informatique, cela n'est pas chose banale, car il faut garder à tout moment un lien vivant avec le texte de départ. Mais la conception de ce passage décisif est bien avancée. D'un point de vue linguistique au contraire, l'historien de la langue retrouve ses habitudes. Il s'agit là de préparer un glossaire amélioré en utilisant la structure exemplaire du DEAF qui sépare le volet des formes de celui des sens.

La préparation informatique de cette glossographie permettra ainsi de gérer dans une seule base tous les volumes des *Plus anciens documents linguistiques de la France*. Une fois réalisé un premier volume, les formes des volumes suivants seront intégrées dans le cadre créé, ce qui réduira considérablement le travail de définition et évitera des contradictions éventuelles.

Mais si l'utilité d'un logiciel qui permet l'édition et l'analyse linguistique de textes anciens se défend aisément, il reste néanmoins à estimer dans quelle mesure le vocabulaire des plus anciennes chartes françaises apportera des données nouvelles à l'histoire de la langue. Les sources documentaires sont riches pour le vocabulaire juridique et féodal, celui de la hiérarchie seigneuriale et ecclésiastique ou pour la langue administrative et diplomatique ; les chartes contiennent également de nombreux termes liés à l'exploitation du sol et au commerce médiéval (mesures et monnaies). Mais ce sont des domaines sémantiques restreints sur lesquels nous sommes assez bien renseignés pour le français, grâce justement au Gdf et au FEW.

L'apport des chartes à notre connaissance du vocabulaire de l'ancienne langue en termes chronologique et diatopique ressort de nos brèves observations sur le syntagme *terre treisse* « terre en friche »<sup>14</sup> :

Le terme – provenant du bas francique... – appartient à l'ancien français nord-oriental (...), sous les formes *trie(z)* 1257, *trieu*, *trihe*, *triexhe*, *triot* et – seulement en

*Philologie*, éd. M.-D. Gleßgen et Franz Lebsanft, Tübingen, 1997 (*Beihefte zu editio*, 8), p. 153-166.

13. Cf. de nouveau la contribution de P. Kunstmann et A. Stein dans ce volume, p. 387-392.

14. Cf. *supra* note 5.

Lorraine – *treixe* 1340 (FEW 17, 400b). Notre charte fournit la première attestation de ce régionalisme.

Le syntagme *terre treisse* (qui n'a pas été retenu par le FEW) préfigure l'usage adjectival attesté uniquement en Lorraine et jusqu'ici exclusivement dans des dialectes modernes (« adj. Meuse *trîce*, Moselle *trix*, Metz *trex* », FEW *ibid.*).

Le FEW – qui s'appuie sur le Godefroy – connaît le type *trieu(s)* pour le domaine nord-oriental et la variante grapho-phonétique de notre charte, *treisse*, pour la Lorraine. Par ailleurs, il relève sur la base de la documentation dialectale l'usage adjectival du terme comme lorrain. Notre charte précise alors :

- la première attestation, avancée de 25 ans pour le lexème (1232 *versus* 1257), d'un siècle pour le type grapho-phonétique lorrain (1340) ;
- une préfiguration par l'apposition *terre treisse* de l'usage adjectival dans les dialectes modernes, ce qui permet de lier plus étroitement la documentation dialectale de cet usage à la langue médiévale.

Des chartes plus longues aux riches descriptions contiennent souvent beaucoup plus de données lexicales nouvelles, mais il ne s'agit néanmoins pas ici d'apports révolutionnaires.

### III. — LE GODEFROY ET L'ANALYSE LEXICOLOGIQUE DES CHARTES

L'analyse systématique du vocabulaire des chartes n'en est qu'à ses débuts. Mais dans le cadre de trois mémoires de maîtrise, tous les mots spécialisés du volume sur la Meurthe-et-Moselle ont été classés dans des catégories sémantiques et une bonne centaine d'entre eux ont été analysés en détail<sup>15</sup>. Sur cette base, il est possible de formuler quelques observations et hypothèses sur le traitement du vocabulaire des chartes dans le Godefroy et le FEW, ce dernier apparaissant bien comme une réinterprétation du Godefroy : le FEW fournit pour les formes de Gdf une synthèse en termes graphématique, géolinguistique, chronologique et sémantique ; par ailleurs, il les place dans un cadre plus large, sans toutefois modifier considérablement le nombre de sources documentaires<sup>16</sup>.

15. Il s'agit des mémoires suivants : Jason M. Dawdy Stein, *Les chartes lorraines du XIII<sup>e</sup> siècle*, Strasbourg, 1999 ; Frédérique Gisquet, *Étude philologique et lexicologique d'un corpus de chartes lorraines du XIII<sup>e</sup> siècle*, Strasbourg, 2000 ; Delphine Harmand, *Id. Approche méthodologique*, Strasbourg, 2000. Trois autres mémoires sur des aspects linguistiques ponctuels de notre corpus sont actuellement en cours.

16. Cf. la contribution de J.-P. Chauveau dans ce volume, p. 323-344.

## 1. *Nomenclature*

Sur l'ensemble du corpus lexical étudié, je n'ai pas relevé un seul cas où le mot des chartes lorraines n'a pas pu être lié à une entrée du Godefroy et du FEW. L'utilisation du dictionnaire ne soulève donc dans ce contexte que des observations de détail :

Parfois, les données du dictionnaire conduisaient à une erreur d'interprétation comme pour la forme *croeire(s)* que j'ai relevée lors du traitement de *croeies* (1228). Ce dernier lexème est une forme métathétique de *corvée* (comme *crowee*, *croee*)<sup>17</sup> et Gdf 2, 377a contient le renvoi correct de *croee* à *corvee*, mais dans la même colonne apparaît l'entrée *croiere* s. f. avec la définition « lieu rempli de craie » et le contexte : « *contenant les diz heritages vint acres qui sont en croieres* (1379, Arch. MM [= Archives départementales de Meurthe-et-Moselle] 30, f° 116 r°.) ».

La forme du document lorrain se rattache sans doute comme dérivé à *croeie* « corvée ». Le rattachement étymologique à CRÊTA a mené Gdf à une mauvaise définition qui se retrouve telle quelle dans le FEW sous cet étymon :

« Mfr. *croiere* "lieu où il y a beaucoup de craie" (1379), *croiere* (1547) ; mfr. nfr. *crayère* "lieu d'où l'on extrait de la craie" (seit 1408, Runk) » (FEW 2<sup>2</sup>, 1330b).

La datation de Runk[ewitz] qui fournit l'attestation de 1408 est fortement sujette à caution : l'ouvrage date souvent des formes françaises contenues dans un cartulaire d'après la date du document latin original. Les deux dates de 1379 (notre charte) et de 1408 sont donc à retirer de l'article, ce qui rend probable une formation plus tardive de *crayère* / *croiere* « lieu d'où l'on extrait de la craie » au XVI<sup>e</sup> siècle.

Parfois, Godefroy a cité précisément, à l'aide de son « informateur » Bonnardot, une des chartes de notre corpus. Il se peut donc que la collection des *Plus anciens documents* repose sur le même ensemble de sources que les lectures de Godefroy, ce qui expliquerait le grand nombre de recoupements. En tout cas, sur la base de mon échantillon, il faut considérer la nomenclature du *Dictionnaire* comme très complète. Il serait inutile de continuer notre travail si notre seul but était un élargissement de cette nomenclature ou l'amélioration des seules premières attestations.

## 2. *Contexte linguistique et pragmatique*

Le Godefroy est beaucoup moins convaincant dans l'identification de la forme lexicale et de son contexte. Si une identification sémantique et une défi-

17. Cf. M.-D. Glebgen, *Das altfranzösische Geschäftsschrifttum in Oberlothringen...*, p. 282 ; FEW 2<sup>2</sup>, 1226b sq.

nition peu fiables peuvent parfois être corrigées – et ont souvent été réinterprétées dans le FEW –, l'exercice est plus difficile pour l'identification des sources concrètes. Dans le cas des chartes, le Godefroy nous fournit une date et un lieu de conservation ou de provenance. Mais nous ne connaissons ni la nature de la source (vente, acte d'inféodation, compromis etc.) ni le lieu ni le contexte sociologique auxquels il faut rattacher le texte<sup>18</sup>. L'ancrage diasystématique des formes reste donc très partiel. Enfin, si le contexte donné pour le *Dictionnaire* permet souvent l'identification de syntagmes, le contexte plus large de la forme est irrécupérable, ce qui empêche toute réflexion sur le rôle et la place d'un mot dans un texte donné. L'étude du vocabulaire des chartes à partir d'une transcription complète place les formes dans leurs contextes syntagmatique et pragmatique et ouvre par-là de nouvelles perspectives d'interprétation.

### 3. Usage des mots

Dans le même ordre d'idées, le Godefroy ne permet pas d'évaluer la représentativité des mots de la langue médiévale écrite. Il ne fournit que des informations très approximatives sur l'usage des lexèmes, leur fréquence en comparaison avec d'autres termes, sur leurs liens avec des genres textuels spécifiques ou sur leurs réseaux synonymiques. La méthode de lecture non systématique qui caractérise le *Dictionnaire de l'ancienne langue* induit une décontextualisation des mots. Même si cette méthode a mené, grâce à la force de travail et à la clarté d'esprit de son auteur, à une nomenclature très complète, elle ne permet pas à l'historien de la langue de mener à leur terme toutes les interrogations pertinentes. Notamment, beaucoup de questions globalisantes et quantificatives qui sont devenues possibles grâce à l'informatique supposent comme base de départ un corpus textuel large plutôt que des glanures lexicales, même importantes. À l'aide d'un réseau plus dense d'attestations, il sera notamment possible de caractériser les liens qui se tissent entre les lieux d'écriture et de déterminer le cheminement et la rapidité de processus d'innovation ou d'expansion lexicale voire de neutralisation et de standardisation du vocabulaire.

### 4. Onomastique

Enfin, le domaine des noms propres, dont les chartes fournissent un nombre infini d'exemples, reste absent du Godefroy. Il s'agit de premiers noms de personne, d'origine onomastique gréco-latine ou germanique – qui le cas

18. Cf. Harald Völker, *Skripta und Variation. Untersuchungen zur Negation und zur Substantivflexion in altfranzösischen Urkunden der Grafschaft Luxemburg (1237-1281)*, Tübingen, 2002.



échéant deviennent aussi des deuxièmes noms –, de deuxièmes ou troisièmes noms de personne d'origine lexicale (les surnoms), et enfin de toponymes, indépendants ou formant une partie de noms de personnes. L'origine des toponymes renvoie à son tour aux noms de personnes ou à des lexèmes, notamment dans la microtoponymie. Les noms d'origine lexicale, autant anthroponymiques que toponymiques, viennent donc enrichir le stock de mots de l'ancienne langue. Puisque les noms propres contenus dans notre corpus de chartes seront organisés dans une base parallèle à la base lexicale, il sera facile de les intégrer à la réflexion lexicologique.

\*

Si notre projet continue à son rythme actuel, lent mais régulier, nous disposerons d'ici une dizaine d'années de quelques milliers d'entrées lexicologiques, élaborées à partir d'une base textuelle qui couvrira un bon tiers du domaine d'oïl. Un grand nombre de formes seront lemmatisées, mais seule une partie restreinte des lexèmes, jugée pertinente pour le vocabulaire spécifique des chartes, connaîtra des analyses lexicales détaillées. Cet ensemble-là constituera alors un supplément au Godefroy. Il n'en élargira pas fortement la nomenclature, sauf pour le volet onomastique ; mais il enrichira l'utilisation du *Dictionnaire de l'ancienne langue* et permettra une véritable contextualisation des mots dans leur usage, dans leur entourage syntagmatique et dans leur trajectoire pragmatique.