

Martin-Dietrich Gleßgen (Zürich)

Le long chemin d'une charte vers l'ère multimédiale

1. L'écrit documentaire médiéval dans l'œuvre de Günter Holtus

La philologie éditoriale a toujours occupé une place primordiale dans l'œuvre de Günter Holtus.¹ En dehors de ses travaux fondamentaux sur le franco-italien, il a considérablement contribué, en synergie et symbiose avec notre ami Kurt Gärtner, à intensifier les études sur des textes documentaires médiévaux en langue vernaculaire. Les trois colloques thématiques orchestrés par eux à Trèves ont donné une impulsion décisive pour réveiller un ancien volcan que l'on croyait mort depuis longtemps. Mon initiative de la mise en valeur des *Plus anciens documents linguistiques de la France* a été présentée pour la première fois lors du premier colloque de Trèves (cf. Gärtner et al. 2001) et j'ai pu largement profiter des acquis méthodologiques et empiriques du groupe de travail trévérois et des autres collègues réunis lors de ces occasions.

Le projet des chartes françaises médiévales m'occupe maintenant depuis un peu plus de dix ans. Ce parcours a montré combien le chemin d'une charte médiévale dans notre ère moderne multimédiale peut être long et ardu mais également combien il peut réserver à tout moment des perspectives nouvelles et inattendues. Je suis aujourd'hui convaincu que le domaine des textes documentaires vernaculaires représente un univers à peine effleuré par la recherche et dont la mise à profit transformera en profondeur l'historiographie des langues modernes.

Sans vouloir reprendre ici les divers aspects de notre projet des *Plus anciens documents*,² je souhaiterais présenter par la suite les questions liées à l'édition des textes et, plus particulièrement, à leur présentation sous diverses formes électroniques. C'est un domaine pour lequel je partage une passion avec Günter Holtus, passion qui pour tous deux s'est nourrie de l'enseignement philologique toujours enthousiasmant de notre maître commun, Max Pfister.

2. Les principes fondamentaux de l'encodage

Les fondements de nos réflexions éditoriales ont été forgés dès les débuts de notre entreprise, aidés par de nombreuses discussions avec notre ami Harald Völker qui était tout particulièrement intéressé à ces questions. Le projet trévérois avait déjà mis en œuvre l'idée d'un encodage textuel neutre permettant des lectures et représenta-

¹ Cf. la présentation de son œuvre, au début de ce volume.

² Cf. la page internet Gleßgen 2009 et la présentation 2010.

tions multiples; nous avons suivi cette voie, riche en perspectives, en introduisant un jeu de balises définies en format XML; plus récemment, nous l'avons accompagné d'une description représentée plus loin sous la forme d'un schéma qui permet une vérification semi-automatique du balisage.³

Par ailleurs, nous avons développé l'idée d'une édition «mixte» ou «composite» qui tient compte, en même temps et dans la même mesure, des éléments caractéristiques de l'original (ponctuation, mise en page, majuscules) et des nécessités interprétatives d'une édition moderne. Cette combinaison entre les perspectives diplomatique et critique est enfin devenue possible, justement grâce à l'édition électronique qui permet différentes interprétations à partir d'une même base textuelle.

Résumons brièvement les éléments essentiels de ces deux principes de base, à savoir l'encodage textuel neutre et le double encodage diplomatique et interprétatif. Nous utilisons comme exemple la charte 2 du corpus de la Meurthe-et-Moselle, présentée plus en détail lors du troisième colloque de Trèves (Gleßgen 2005) puisque cela nous permet de renoncer ici à une description détaillée et de nous concentrer sur les éléments de structure les plus pertinents.

En réduisant les informations à l'essentiel, l'encodage XML de la charte 2 se présente sous la forme suivante:

```
<gl>
<id>555550002</id>
<zitif>chMM002</zitif>
<an>
<nom>002</nom>
<d>1234 (25 mars–31 décembre) ou 1235 (1er janvier– 24 mars)</d>
<type>charte: acensement de terres</type>
<r>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.</r>
<aut>non annoncé</aut>
<disp>abbaye de Salival</disp>
<s>disposant</s>
<b>disposant</b>
<act>Wirrion et Houillon, paysans de Juvelize</act>
<rd>scriptorium de l'abbaye de Salival</rd>
<f>Parchemin jadis scellé sur simple queue; 58x141</f>
<l>AD MM H 1244, fonds de l'abbaye de Salival</l>
</an>
<txt>
<div n="1"> <wn n= "1"><maj>C</maj>onue</wn> <wn n= "2">chose</wn> <wn n=
"3">soit</wn> <wn n= "4">à</wn>.<wn n= "5">toz</wn></div>
<div n="2"> q<abr>ue</abr> li abes <abr>et</abr> li chapitles de Salinvas /. at laissié a
Wirion
<zw/> <abr>et</abr> Huillon, les dous freres de Gev<abr>er</abr>lise, les anfanz Bertran
Bachelier,</div>
```

³ Pour cette opération, nous avons utilisé d'abord comme éditeur *xml-spy*, alors gratuit dans sa version *home*; depuis sa commercialisation intégrale, nous faisons appel au programme *Exchanger*, téléchargeable gratuitement et satisfaisant dans l'usage.

```

<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>
<zwt> /.II/. himas de blef,/. l'un d'avoine,/. l'autre de froment,/.</div> (...)
<par/>
<div n="6"> <maj>S</maj>i est ensi devisee /. q'au Tramble en
<zwt> at /.III/. jornas,/. un p<abr>ar</abr> lui<ful>Probablement <abr>Wirion</abr>, le
premier frère nommé dans le texte.</ful> /. <abr>et</abr> /.III/. ensemble,/.</div> (...)
<par/>
<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 verité,/. </div>
<div n="11"> l'an q<abr>ue</abr> li miliaires corroit p<abr>ar</abr> /.M/. <abr>et</abr>
CC/. <abr>et</abr> XXXIII/. anz,/.</div>
</txt>
</gl>

```

La charte est délimitée comme toute autre document du corpus par les balises <gl> </gl> qui indiquent une entité textuelle caractérisée par les mêmes paramètres diastématiques et de genèse: un même scribe, à un même moment, sur un même support. Pour des raisons de gestion, nous avons réuni toutes les chartes conservées dans les archives d'un même département dans un fichier informatique unique (reconnaisable ici par les chiffres « 55555 » dans la clé d'identification <id> </id>).

À l'intérieur de toute entité, nous distinguons toujours une partie textuelle, balisée par <txt> </txt> et une partie analytique (<an> </an>) qui contient des informations matérielles (lieu de conservation, éléments de description externe), un résumé du contenu (<r[egeste]>) et les paramètres du diasystème (notamment la date <d>, le genre textuel <t> et le lieu d'écriture <rd>). De cette manière, tout mot grammatical (occurrence ou *token*) dans chaque charte est rattachée à son ancrage matériel et diastématique.

Le texte lui-même est structuré par les divisions sémantiques (<div n="1">) ainsi que par des paragraphes (<par/>). Chaque mot grammatical est cerné par une balise individuelle associée à un numéro (<wn n="2">chose</wn>), unique à l'intérieur du fichier (c55555).⁴ Toute occurrence du corpus global porte donc un numéro unique, composé du numéro du fichier et du numéro de l'occurrence (5555500002); ces numéros forment l'ancrage pour la gestion de notre base de données linguistique.

L'encodage XML du départ n'a pas connu de modifications structurelles depuis 2003, à l'exception du balisage systématique de toutes les occurrences, introduit seulement en 2009.⁵ Ce premier pas important a donc pu être réalisé relativement vite.

Le fichier qui contient les données textuelles et analytiques dans un encodage neutre représente la cheville ouvrière pour toute forme d'édition et pour toutes les analyses linguistiques mises en œuvre à travers une base de données. Quant à l'édition, le texte peut être présenté sous différentes vues, grâce à un programme d'interprétation: il peut apparaître sous une forme entièrement diplomatique ou entièrement

⁴ Dans l'exemple, nous n'avons reproduit que les numéros de la première division.

⁵ Notre ami Achim Stein a pris la même décision presque au même moment pour les textes du *Nouveau Corpus d'Amsterdam*, ce qui nous a réconforté dans notre choix. La numérotation systématique permet également une meilleure description des phraséologismes.

interprétative, ce qui permettra des possibilités de choix lors d'une présentation web. Voici une présentation mixte qui allie des éléments diplomatiques et interprétatifs:⁶

002

1234 (25 mars–31 décembre) ou 1235 (1er 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'acencement sont très contraignantes pour les paysans.*

Auteur: non annoncé

Disposant: abbaye de Salival

Sceau: disposant

Bénéficiaire: disposant

Autres acteurs: Wirrion et Houillon, paysans de Juvelize

Rédacteur: **scriptorium de l'abbaye de Salival**

Parchemin jadis scellé sur simple queue; 58 x 141

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

1 Conue chose soit à-toz 2 que li abes et li chapitres de Salinvas • at laissé à Wirion /
 et Huillon, les dous freres de Geverlise, les anfanz Bertran Bachelier, 3 XIII-jor/nas
 de terre treisse • en la fin de Geverlise • et à lor oirs 4 parmi XIII-deniers de cens • et
 5 / • II • himas de blef • l'un d'avoine • l'autre de froment • 5 et s'il ne paievent à jor //
 nomei à la feste sent Remi • à Giverlise, en la maison de Salinvas • que l'on se
 tan/roit à la terre • et ce que sus averoit •
 6 Si est ensi devisee • q'au Tramble en / at • III • jornas • un par lui 2 • et •
 III-ensemble • 7 et en la voie de Hignicort en at / V • jornas, • II • d'une part et • III •
 d'autre • 8 et en la voie de Marsal • II • jornas • / après la terre les Voves • 9 et en la
 10 voie de Donnereis • as Genoivres • 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 •

Les principes du double encodage, médiéval et moderne, qui régissent l'édition de notre corpus, ressortent déjà de ce bref exemple:

(1) Les mots grammaticaux sont séparés selon l'usage moderne, mais avec l'indication univoque de l'intervention éditoriale: à partir des formes *à-toz* ou *l'un*, un programme peut immédiatement reconstituer les formes précises du manuscrit (*atoz*, *lun*) ou proposer, si le lecteur le souhaite, une lecture très interprétative (*à toz*).

(2) L'ajout des apostrophes et des accents suit ce même modèle: elles appartiennent toujours à l'éditeur et peuvent être supprimées dans une perspective diplomatique.⁷

⁶ V. pour une version complète de l'édition, avec l'intégralité du tableau analytique et les notes éditoriales la publication électronique (Gleßgen 2009).

⁷ Un autre exemple: une forme manuscrite *alabe* sera transcrite dans le fichier XML par *à-l'abé*, une forme manuscrite *ala be* par *à-l'a_bé*; dans les deux cas, autant la lecture diplomatique que la lecture interprétative ou mixte peuvent être générées automatiquement par

(3) Les majuscules du manuscrit sont rendues (dans la mesure où elles sont identifiables) en gras, les abréviations en italique (dans le fichier, ce sont les balises <maj> </maj> et <abr> </abr>); toutes les autres majuscules, qui correspondent à l'usage moderne, sont introduites par l'éditeur, sans marquage aucun: là encore, le fichier informatique permet de reconnaître autant la forme manuscrite que les interventions éditoriales.

(4) Les ponctuations médiévale et moderne sont distinguées; dans la version mixte, nous renonçons à des signes de ponctuation modernes à des endroits où se trouvent des signes de ponctuation médiévale; mais le fichier informatique sépare de manière univoque les deux séries de marquage et permet de ne reproduire qu'un seul.

(5) Les sauts de ligne de l'original sont indiqués (par un trait transversal, dédoublé toutes les cinq lignes et accompagné alors d'une numération à côté du texte; balises <zw/> ou, pour les mots séparés par le saut de ligne, <zwt/>); parallèlement, le texte est segmenté d'après des critères sémantico-syntaxiques par des divisions qui correspondent donc à un choix d'éditeur (balises <div n="1"> </div>).

(6) Les *lapsus calami* sont corrigés (tout en les indiquant en note et en plaçant les formes restituées entre crochets dans le texte), ce qui permet à l'analyse linguistique de reposer sur des formes de langue et non pas sur des erreurs d'écriture (p.ex. *comte* et non pas, dans une copie, *conite*).

Ces principes d'éditions ont été affinés à travers les années et leur application concrète s'est précisée dans le travail sur plus d'un millier de chartes. La différenciation entre les lettres majuscules et minuscules a été améliorée grâce aux conseils de de Françoise Viellard.⁸ Nous avons introduit et continuons à introduire des balises complémentaires, par exemple, pour des fins d'interprétation historique. Mais là encore, les idées de base ont déjà pu être établies en 2003, donc relativement vite et tôt.

3. La préparation d'une édition électronique

La suite des opérations s'est avérée, comme attendu, plus laborieuse. En travaillant à différents chantiers en parallèle, nous avons entrepris avec toute une équipe de jeunes chercheurs et de collègues l'édition de presque 2.000 chartes du XIII^e siècle, ce qui a demandé des efforts considérables, efforts qui se sont échelonnés à travers

un programme de transformation (diplomatique: *alabe* et *ala be*, en omettant les tirets, apostrophes, accents et traits sur la ligne; interprétative: à *l'abé* dans les deux cas; mixte: à-*l'abé* et à-*l'a_bé*). Selon les finalités du lecteurs, les trois types de lecture ont leur raison d'être.

⁸ Notre amie et partenaire dans le projet des *Plus anciens documents linguistiques de la France* a proposé le principe de n'admettre une majuscule que dans les cas où le même manuscrit (ou au moins: le même scribe) distingue pour la lettre initiale en question au moins deux formes différentes dont l'une peut être interprétée (par sa taille, son ornement ou son ductus) comme majuscule. Une lettre initiale uniforme, même ornée et grande, ne pourra donc pas être considérée comme majuscule, en l'absence d'une opposition immédiate. Ce principe, surprenant au premier abord, s'est avéré dans la pratique très applicable et porteur.

huit années de travail.⁹ Toute édition d'un texte médiéval demande une attention particulière et un niveau certain de compétences paléographiques, linguistiques et lexicologiques, mais aussi diplomatiques et historiques. Nous avons eu la chance de rencontrer des collaborateurs qui se sont attelés à leur tâche avec un engagement remarquable. Cette tâche était diverse selon les différents cas de figure mais toujours centrée sur l'édition d'un corpus partiel: les éditions ont demandé à tous un investissement entier, dans la transcription, les multiples étapes de correction, l'édition et dans l'établissement des tableaux analytiques.

Au-delà des efforts individuels, une entreprise générale demande une homogénéisation des différents résultats à tous les niveaux. Indépendamment du niveau de difficulté des différentes chartes ou de la sécurité dans la transcription des individus, le résultat final doit connaître un degré relativement uniforme de fiabilité.

Les tableaux analytiques comportent notamment le relevé des acteurs principaux de la charte, ce qui demande une réflexion souvent poussée sur les contenus et enjeux de chaque acte, et l'identification des rédacteurs, qui est particulièrement épineuse. Reste enfin l'identification des toponymes et l'établissement des registes: ces deux opérations sont lourdes et demandent, pour différentes raisons, un travail systématique s'inscrivant dans sa logique propre que nous n'avons jusqu'ici entrepris que partiellement.

Les informations interprétatives ne sont pas toutes indispensables pour une (première) édition électronique et certains éléments gagneront, justement, à être traités de manière systématique à partir d'un corpus textuel plus large. Mais notre but reste un avancement maximal dans tous ces domaines. Notamment, nous souhaiterions pouvoir identifier l'essentiel des lieux d'écriture, ce qui a supposé le développement d'outils informatiques complémentaires et d'une méthodologie adéquate.¹⁰

À l'heure d'aujourd'hui (mai 2010), nous disposons d'un corpus d'environ 2000 chartes éditées¹¹ dont 1100 ont été pleinement corrigées dans leur partie textuelle; pour les autres 900, les opérations de correction et d'homogénéisation sont très avancées et devraient être terminées d'ici 2012. Les registes ont été établis pour la

⁹ V. l'aperçu des différents corpus Gleßgen 2010 [2011]. En dehors de Françoise Viellard et d'Olivier Guyotjeannin, avec qui nous avons lancé le projet, sont intervenus dans l'élaboration concrète des fichiers textuels depuis 2003 les élèves, collègues et amis suivants: Julia Alletsgruber, Thomas Brunner, Klaus Grübl, Dumitru Kihai, Anne-Christelle Matthey, Claire Muller, Martina Pitz, Cristina Solé, Sarah Tinner, Benoît Tock, David Trotter, Claire Vachon, Paul Videsott. Nous souhaiterions honorer en ce lieu la mémoire de Martina Pitz, décédée brutalement le 21 mai 2010; notre amie d'études de Sarrebruck, élève de Wolfgang Haubrichs et de Max Pfister, travaillait depuis plusieurs années sur le corpus de Metz (cf. Pitz 2001) et avait envisagé de suivre de plus près le volet onomastique du projet global.

¹⁰ Cf. Gleßgen 2008.

¹¹ Le tout est composite: certains corpus, publiés ou manuscrits, ont été repris de la première série des *Plus anciens documents* sous la direction de Jacques Monfrin, d'autres ont été nouvellement constitués à partir des documents d'archives; l'intégration d'un document correctement transcrit d'après les anciens critères de la collection dans le modèle actuel demande toutefois en moyenne une journée de travail par charte. C'est toujours un gain en termes de temps mais cela incite néanmoins à une certaine prudence dans les plannings.

grande majorité de ces 2000 documents mais sans l'homogénéisation qui nous semble nécessaire. Les lieux d'écriture ont été identifiés, jusqu'ici, pour une moitié des chartes; c'est le chantier que nous traitons actuellement en priorité puisque ces «lieux» permettent un ancrage sûr des documents dans le diasystème de leur époque; nous espérons atteindre un degré d'identification satisfaisant d'ici la fin 2012. Le corpus est naturellement en voie d'extension et les documents de plusieurs autres départements sont actuellement en suspens.

La préparation éditoriale des chartes a été accompagnée de diverses analyses linguistiques, notamment scriptologiques, lexicologiques et syntaxiques, du développement d'un logiciel utilisable pour ces fins et de réflexions théoriques sur le statut communicatif de ces documents dans l'optique de la médialité prémoderne. Nous passons ici sur tous ces aspects qui ne concernent que partiellement le projet éditorial.

4. La première présentation internet

Après plusieurs années de préparation, le corpus des chartes a ainsi atteint un volume et un degré d'élaboration qui nous ont semblé cohérents pour une première mise en ligne. Même si nos documents avaient été préparés dès les débuts pour une publication électronique et que nous avons fait des tests à plusieurs moments pour en évaluer la faisabilité, nous avons néanmoins rencontré le moment venu un obstacle majeur et nous avons dû résoudre toute une série de nouveaux problèmes de détail.

Commençons avec l'obstacle majeur: l'essentiel des programmations pour notre logiciel *Phoenix*, destiné à l'édition et l'analyse linguistique de textes médiévaux, a été programmé à l'aide du langage script *Tustep*, d'abord par Matthias Kopp, ensuite par Matthias Osthof. Les programmes sont pour l'essentiel limpides et bien commentés; le logiciel remplit pleinement les fonctionnalités demandées. Par ailleurs, *Tustep* peut être utilisé pour une installation de serveur, ce qui permet de transposer immédiatement sur internet les fonctions d'interrogation de *Phoenix*. Or, au moment de la mise en ligne, nous nous sommes rendus compte que les serveurs gérés par *Tustep* sont extrêmement rares; en 2008, il y en avait précisément trois (à Tübingue même, à Wurzburg et à Duisbourg). Le responsable de l'installation dans le centre d'études hébraïques, à Duisbourg, est un collègue philologue avec des compétences informatiques remarquables et il a très aimablement proposé à M. Osthof de placer notre corpus sur ce serveur. Mais la faible diffusion des serveurs *Tustep* nous a semblé un obstacle important, d'autant plus que nos programmes s'inscrivent dans la logique d'un accès libre et sont destinés, à terme, à une diffusion libre dans la communauté scientifique. Nous avons donc décidé de revenir sur le choix initial de *Tustep*: un tel changement d'orientation n'est pas très surprenant, étant donné la grande mouvance qui caractérise l'univers des outils de programmation. Il nous a toutefois obligé à une restructuration globale de nos logiciels et de nos données, ce qui a ouvert, du même coup, de nouvelles perspectives.

Avant d'entreprendre cette restructuration importante, nous tenions à réaliser une version internet, ne serait-ce que restreinte, pour permettre à la communauté scientifique un premier accès à nos documents. Ce test en dimensions réelles nous a

également permis de rendre apparents tous les problèmes de détail qu'une telle opération pouvait soulever.

Pour éviter des programmations lourdes qui seraient devenues obsolètes avec la restructuration globale, nous avons opté pour une représentation statique à l'aide d'un encodage PHP. Le format PHP est proche de HTML (et donc de XML), et la transformation de nos données de XML en PHP était par conséquent relativement simple. Cette solution a permis la réalisation d'une mise en page internet qui intègre les photographies des documents; en revanche, nous avons dû renoncer dans ce premier cadre à toute possibilité d'interrogation linguistique ou historique: chaque charte représente un document à part entière et, en l'absence d'une programmation supplémentaire, il n'était pas possible de réaliser des interrogations sur plusieurs chartes voire sur le corpus complet. Il restait la possibilité de dédoubler le corpus et de placer nos textes sous un autre format dans le cadre des interrogations, très performantes, du logiciel TWIC qu'Achim Stein a créé pour le *Nouveau Corpus d'Amsterdam*.¹² Par ailleurs, pour permettre des vues alternatives sur nos chartes, diplomatique, interprétative ou mixte, nous aurions pu générer trois versions, toujours statiques, de ces possibles vues. Mais cette démultiplication de versions intermédiaires de notre corpus nous a semblé trop lourde, et nous avons préféré postposer les deux volets à la restructuration des données.

Malgré notre prudence, la «simple» mise en ligne statique de notre corpus en format PHP, avec une représentation d'édition de type mixte et l'accès parallèle aux photographies, a occupé notre programmeur Matthias Osthof, mon assistante Claire Vachon, notre réviseur-paléographe Dumitru Kihai et nous-même pendant trois mois à temps plein. Le nombre de révisions des données textuelles à l'aide de notre schéma pour homogénéiser en détail les quatre premiers corpus mis en ligne était considérable (concernant par exemple la position précise des balises et les espaces qui les entouraient ou encore les séparations des mots qui impliquaient l'introduction ou non de tirets ou de traits sous la ligne mais également le problème des chiffres romains à mettre ou non en majuscules et à entourer ou non de points, ou encore la numérotation des séries .jpg et .tiff des photographies recto et verso pour permettre un accès automatique). À ces questions souvent sans implication linguistique s'ajoutaient la révision globale des noms propres dans les tableaux analytiques et celle des rédacteurs déjà identifiés. Tous ces éléments avaient déjà été traités à de multiples reprises auparavant mais l'intervention de plusieurs transcrip-teurs et la présence de plusieurs séries de photos avaient malgré tout généré des incohérences à niveau microscopique qu'il a fallu redresser dans leur intégralité pour la mise en ligne.

Le résultat, publié en janvier 2009 sur la page internet de la médiévistique de Zurich, allie une bonne qualité textuelle avec une mise en page élégante et un accès rapide aux photographies (par simple click). En voici l'exemple de la charte 2:

¹² Cf. Stein 2010 et Kunstmann/Stein 2007.

Zürcher Mediävistik Les corpus - Charte

Home

Les plus anciens documents linguistiques de la France
Corpus : chartes de la Meurthe-et-Moselle (ChMM)
Directeur du volume : Martin-D. Gledgen
Transcription et édition électronique : Martin-D. Gledgen / Michel Arnold
Relecture de la transcription : Dumitru Chihai

Les corpus textuels

Descriptif du projet

Contact

ChMM 002

1234 (25 mars-31 décembre) ou 1235 (1^{er} janvier- 24 mars)

Type de document: *charte: accensement de terres*

Objet: L'abbé et le chapitre de Salival accensent à Wirrion et Houillon treize journaux de terre au finage de Juvelize contre un cens de treize deniers et deux hémines de grain.

Auteur: non annoncé
Disposant: abbaye de Salival
Bénéficiaire: abbaye Notre-Dame de Salival
Autres acteurs: Wirrion et Houillon, paysans de Juvelize
Sceau: non annoncé

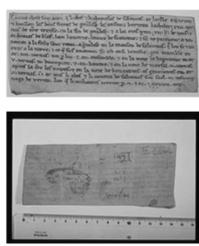
Rédacteur: *scriptorium de l'abbaye Notre-Dame de Salival* [pal]
Support: parchemin judis scellé sur simple queue; 58x141
Verso: néant

Lieu de conservation: AD MM H 1244, fonds de l'abbaye de Salival
Edition: -

¶1 Conue chose soit à-toz ¶2 que li abes et li chapitres de Salivas - at laissié à Wirion ¶3 et Houillon, les dous freres de Geverlise, les anzanz Bertran Bacheler, ¶4 XIII-jor ¶5 nas de terre treisse - en la fin de Geverlise - et à lor oirs - ¶6 pormi XIII-deniers de cens - et ¶7 ¶8 himas de blef - l'un d'avoine - l'autre de froment - ¶9 et s'il ne paievent à jor ¶10 nomei à la feste sent Remi - à Giverlise, en la maison de Salivas - que l'on se tan ¶11 roit à la terre - et ce que sus averoit -

¶12 Si est ensi devisee - q'au Tramble en ¶13 at III-jornas - un par lui - et III-ensemble - ¶14 et en la voie de Higlicort en at ¶15 V-jornas, II-d'une part et III-d'autre - ¶16 et en la voie de Marsal-II-jornas - ¶17 après la terre les Voves - ¶18 et en la voie de Donnerreis - as Genovieres - en at ¶19 II-jornas -

¶20 Et Ci at mis li abes et li covenz de Salivas son sael - en tesmoig ¶21 nage de verité - ¶22 l'an que li milliaires corroit par-M- et CC-et XXXIII- anz -



Kompetenzzentrum Zürcher Mediävistik

Ecole Nationale des Chartes

Cette version internet souffre, comme nous l'avons dit, de l'absence de possibilités d'interrogation et de l'absence d'accès aux informations linguistiques (graphématiques, morphologiques et lexicologiques) dont nous avons pourtant longuement enrichi notre base de données textuelles. Elle a également l'inconvénient de toute publication électronique, de garder malgré tout un statut inférieur à celui d'une publication papier, ne serait-ce que par l'absence de comptes rendus qui permettent normalement aux monographies d'être perçues et évaluées par la communauté scientifique.

Ce dernier problème se résoudra au fur et à mesure dans les années à venir. Quant au premier point, nous avons entrepris dès 2009 la réorganisation de nos outils informatiques, toutefois sur la base désormais fiable d'un corpus des 1.100 documents édités en ligne.

5. Le développement de la deuxième présentation internet

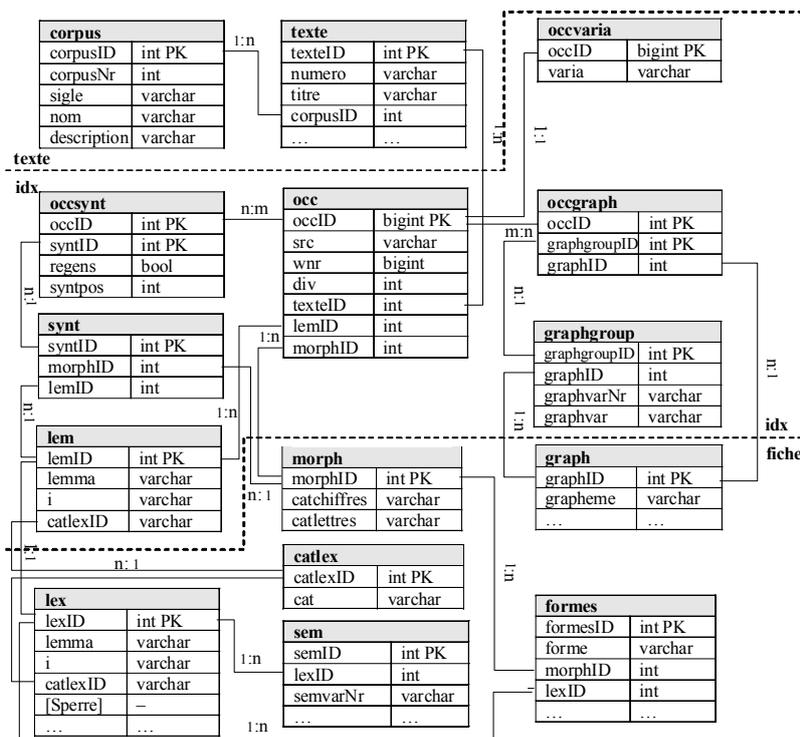
Pour pouvoir réaliser une version internet pleinement interrogeable et permettant des vues dynamiques sur les textes, nous avons dû réorganiser les données non pas textuelles mais interprétatives. La gestion des informations linguistiques qui accompagnent les données textuelles repose actuellement sur un cheminement à trois paliers:

- (1) Le niveau textuel où chaque occurrence individuelle est accompagnée d'un numéro unique (dans le fichier c55555: <wn n= "2">chose</wn>) et est liée aux informations du tableau analytique (du texte 0002 dans ce même fichier).

(2) Le premier niveau interprétatif qui correspond à un fichier-index et qui contient des informations essentielles pour chaque numéro d'occurrence, la forme (<s[our]ce>), le lemme (<lex>), la catégorie de nom (f="c" pour un nom commun, f="t" pour un toponyme), un renvoi alpha-numérique à une description graphématique (<graph f="ou1">) ainsi qu'à la définition morphologique et sémantique (<wn>2</wn> <src>chose</src> <lex f="c">chose</lex> <graph f="(...)"> <morph f="(...)"> <sem f="(...)">).

(3) Le second niveau interprétatif qui correspond à un fichier de données interprétatif assez complexe réunissant toutes les informations linguistiques générées dans notre logiciel *Phoenix* à partir des occurrences individuelles.

Paul Gévaudan, qui a rejoint en 2009 notre projet, nous a convaincu de l'utilité d'une segmentation radicalement différente de nos données interprétatives. L'idée est celle d'une segmentation maximale de toutes les informations linguistiquement pertinentes (= une «normalisation» en théorie des structure de données). Les liens entre les occurrences, les données diasystématiques et les différents éléments linguistiques sont alors établies par de nombreuses séries chiffrées qui servent de renvois d'une entité à une autre. Un aperçu (simplifié) de la structure des données établie par P. Gévaudan donne une notion de la nature de ce concept, même si les détails échappent facilement au lecteur linguiste; chaque « ID » dans le tableau correspond à une série chiffrée, « idx » représente le premier niveau interprétatif, « fiche » le second:



Cette structure de données, éclatée uniquement en apparence, permet un accès extrêmement rapide à toute information, ce qui est la condition essentielle pour la gestion dans le cadre d'un moteur de recherche (navigateur, *browser*). Une fois les informations actuelles exportées dans le nouveau système, elles sont gérées à tout moment à partir d'un serveur.

Nous avons choisi comme langage de programmation *PHP*, plus abstrait et par conséquent dans certains domaines moins élaboré que *Tustep* mais dans la plupart des constellations plus performant que celui-ci et en même temps infiniment plus répandu et accessible; pour la gestion de la base de données nous avons retenu *MySQL* comme le DEAF électronique, ce qui facilitera le rapprochement souhaitable des données lexicologiques.¹³ La représentation des informations se construira à l'aide d'interfaces CGI par lesquels le serveur relie les moteurs de recherche des utilisateurs aux programmes qui, à leur tour, ont accès à la base de données et assurent par là un assemblage dynamique en fonction des requêtes formulées.

La réorganisation conceptionnelle a occupé pendant presque six mois Paul Gévaudan, Matthias Osthof et nous-même. La reprogrammation, qui sera réalisée pour la plus grande partie par notre jeune collaborateur informaticien et linguiste Samuel Läubli, demandera probablement autant de temps.

Si notre planning se réalise correctement, au moment de la parution des mélanges en l'honneur de Günter Holtus, la deuxième version internet des *Plus anciens documents linguistiques de la France* sera accessible. Elle permettra alors les interrogations linguistiques pertinentes à l'aide d'une série d'algorithmes appropriés et elle donnera ainsi accès aux bases de données graphématique, morphologique et lexicologique que nous construisons en parallèle. Nous espérons qu'elle offrira également des vues alternatives sur les documents, de la version la plus diplomatique à la plus interprétative. Cela prendra éventuellement un peu plus de temps mais nous sommes confiants que cette idée, toute simple et bien en place dès les débuts de notre projet, finira par trouver une réalisation pleinement satisfaisante.

6. Une ouverture pédagogique: la plateforme *Ad fontes*

Parallèlement à la mise en ligne interactive de nos chartes, nous avons été incité par notre collègue zurichois en histoire médiévale, Simon Teuscher, à préparer une version commentée d'un choix de nos documents pour la plateforme paléographique *Ad fontes*, jusqu'ici exclusivement dédiée aux textes latins et allemands. Nous avons pu gagner pour cette réalisation pédagogique assez prenante le secours de notre amie Hélène Carles habituée au travail avec des sources documentaires et passionnée par l'enseignement de la paléographie et de la philologie médiévales.

Le nouveau volet (gallo-)romaniste de la plateforme *Ad fontes* comporte actuellement une introduction générale qui réunit notamment les critères d'édition à suivre

¹³ Cf. la page internet du DEAF (www.deaf-page.de/deaff.htm) et Tittel 2010.

ainsi qu'une série, réduite mais significative, de chartes commentées.¹⁴ En voici la page d'accueil:

TRAINING >

Exercices de transcription
Vous avez ici la possibilité d'exercer et améliorer votre capacité à déchiffrer et transcrire les écritures anciennes.
Veillez choisir un exercice de transcription dans la liste suivante:

Intitulé	Institution	An	Langue	Niveau	Bilan
Charte de l'abbaye de Salival	Archives départementales de la Meurthe-et-Moselle	1234	Français	facile	0%
Charte de l'épiscopat de Toul	Archives Départementales de la Meurthe-et-Moselle	1237	Français	facile	0%
Charte de Metz, transaction de rente	Archives départementales de la Meurthe-et-Moselle	1239	Français	facile	0%
Charte de l'abbaye de Fontenay	Archives Départementales de la Côte-d'Or	1253	Français	facile	0%

Le programme est conçu pour suivre votre transcription au fur et à mesure. Afin de vérifier si ce que vous transcrivez est correct, vous pouvez cliquer sur la touche 'vérifier les données'. Les mots qui contiennent des erreurs seront alors marqués en rouge. Le programme ignore les questions de ponctuation mais tient compte en revanche de l'identification des majuscules et des minuscules. Il indique aussi à quel endroit vous avez omis de transcrire un mot (...) et là où il devrait y avoir un saut de ligne (§).

Lorsque vous ne parvenez plus à avancer dans votre transcription, le programme peut, si vous le souhaitez, vous donner des indices (de type paléographique et/ou éditorial) ou la solution pour un mot donné. Pour avoir accès à ces conseils, nous vous invitons à cliquer sur l'onglet 'Conseils' et ensuite à positionner le curseur sur le mot qui fait difficulté. Ces conseils permettent également d'attirer votre attention sur des éléments dignes de notes et de vous sensibiliser ainsi à des questions diverses pour mieux vous faire progresser dans l'édition des documents médiévaux.

TUTORIUM
[Transkriptionsregeln](#)

RESSOURCEN
[Critères d'édition \(PDF\)](#)

Les exercices de transcription reposent sur les critères d'édition accessibles à partir du TUTORIUM. Dans les RESSOURCEN, se trouve une version PDF imprimable de ces critères d'édition. Pour les particularités propres à une écriture donnée, on se reportera aux conseils relatifs aux exercices d'entraînement concernés.

ÜBERSICHT

La qualité particulière d'*Ad fontes* réside dans le fait que les étudiants peuvent s'exercer à la transcription d'un texte à l'intérieur de la plateforme et que le logiciel leur indique, au moment choisi par l'étudiant (pas à pas ou bien tout à la fin de l'exercice), si la transcription et l'édition accomplies sont erronée. La séquence apparaît alors en rouge. À ce moment-là, il leur est possible d'accéder à des informations complémentaires concernant le mot faisant difficulté, informations visant à leur permettre de transcrire et d'éditer correctement la forme. Si les étudiants n'y parviennent pas, ils peuvent accéder à la solution pour la forme concernée.

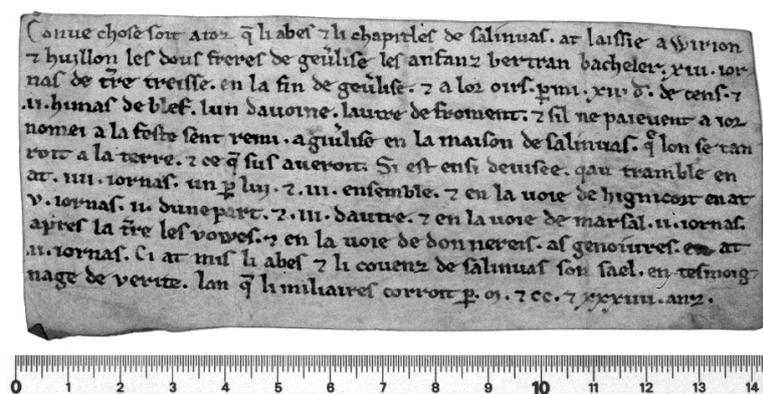
D'un point de vue pédagogique, cette méthode est d'une utilité réelle puisque l'apprentissage paléographique contient un potentiel de frustration considérable, frustration induite d'une part par la non-reconnaissance d'une forme mais aussi plus généralement par l'insécurité des résultats obtenus. Les commentaires d'*Ad fontes* introduisent ainsi un niveau intermédiaire entre l'incompréhension totale et la solution parfaite; par ailleurs, le logiciel résout le problème de l'insécurité dans les choix des transcribers dans des conditions optimales. Les premières applications avec des étudiants ont donné, en effet, d'excellents résultats dans l'apprentissage

¹⁴ Cf. *Ad fontes* <www.adfontes.uzh.ch/1000.php>: l'inscription obligatoire est gratuite et rapide.

philologique des textes documentaires. Les avancées ont été très satisfaisantes et la possibilité d'établir un bilan à la fin de la transcription a permis à chacun une auto-évaluation synthétique.

L'inconvénient de cette procédure est que toute occurrence individuelle doit être préalablement accompagnée d'un commentaire, ne serait-ce que bref, apte à aider la transcription ou la compréhension du mot en question. Notre charte 2, longue d'une demie page dans son édition internet, a ainsi généré une liste de douze pages de commentaires individuels portant principalement sur des questions paléographiques et ecclotiques.

Nous reproduisons par la suite en guise d'exemple les commentaires pour le début du document dont voici d'abord la photographie:



mot	commentaire
<i>Conue</i>	Conservez la majuscule car c'est le premier mot du texte. Observez ici la différence de sens des jambages pour <i>u</i> et <i>n</i> . Observez la petite barre montante caractéristique du ductus de <i>e</i> dans ce manuscrit.
<i>chose</i>	La quatrième lettre est <i>s</i> long.
<i>soit</i>	La première lettre est <i>s</i> long et la troisième est <i>i</i> . Ne la confondez pas avec <i>r</i> .
<i>à-toz</i>	Éditez la première lettre avec un accent grave afin de distinguer la préposition <i>à</i> de la troisième personne du singulier du verbe <i>avoir</i> . Introduisez un tiret entre la première et la deuxième lettre car il s'agit du début d'un mot distinct. On opère ici une segmentation graphique conforme à celle du français moderne. La dernière lettre est <i>z</i> . Ne la confondez pas avec <i>r</i> .
<i>que</i>	Il s'agit d'une abréviation qui consiste en la lettre <i>q</i> surmontée d'un signe conventionnel appelé 'tilde'. Elle doit être résolue en <i>que</i> , conjonction de subordination.

mot	commentaire
<i>li</i>	Il s'agit de deux lettres dont la première est <i>l</i> . C'est un article défini masculin singulier au cas sujet.
<i>abes</i>	La dernière lettre est <i>s</i> long. Elle note le morphème flexionnel du cas sujet singulier.
<i>et</i>	Résolvez l' 'éperluette', qui est une abréviation par signe conventionnel, en <i>et</i> .
<i>li</i>	Ce mot est un article défini masculin singulier au cas sujet.
<i>chapitles</i>	Il s'agit d'un mot de neuf lettres. <i>l</i> au lieu de <i>r</i> attendu. <i>chapitre</i> est un latinisme (CAPITULUS).
<i>de</i>	Les deux lettres <i>d</i> et <i>e</i> se chevauchent partiellement (phénomène de 'ligature').
<i>Salinvas</i>	Introduisez une majuscule car il s'agit d'un nom de lieu. La première et la dernière lettre est <i>s</i> long. La sixième lettre, qui consiste en deux jambages doit être éditée <i>v</i> car elle note un phonème consonantique. Omettez le point dans l'original.

L'exercice de transcription est accompagné d'une brève introduction pour chaque document remplaçant le texte dans son contexte, ainsi que de quelques remarques historiques; voici les textes correspondant à notre charte:

Charte de l'abbaye de Salival

Il s'agit d'une charte émanant du scriptorium de l'abbaye de Salival, datant de mars 1234 et étant conservée aux Archives Départementales de la Meurthe-et-Moselle. Cette charte est l'un des premiers textes documentaires de Lorraine à avoir été rédigé en langue vernaculaire, c'est-à-dire en français et non pas en latin. Le plus ancien document, quant à lui, a été émis par le scriptorium du chapitre cathédral de Metz (à une cinquantaine de kilomètres au nord-ouest de Salival) en 1215.

Ce document a été choisi comme second exercice paléographique par contraste avec la charte du premier exercice. L'écriture est dominée mais elle est beaucoup moins empreinte d'ornementation et très statique. Le ductus des lettres est minimaliste et l'usage des majuscules extrêmement restreint. Des lettres majuscules, toujours sans ornement, apparaissent exclusivement en début des trois phrases qui constituent la charte. Dans un même ordre d'idées, le scribe de cette charte a utilisé davantage d'abréviations que dans le document précédent. Sont illustrés la plupart des types d'abréviations en vigueur dans les chartes médiévales françaises, à savoir les abréviations par signe conventionnel, par lettre suscrite ainsi que les abréviations finales par suspension. La charte est émise par une abbaye qui a également produit des documents plus élaborés; la simplicité formelle du présent manuscrit et l'absence de →l'intitulatio s'expliquent, entre autres, par la condition roturière des personnages avec lesquels l'abbaye conclut l'acte juridique.

La ponctuation originale de ce document est particulièrement signifiante puisqu'elle structure de manière très cohérente le contenu de la charte. On se reportera aux conseils disponibles dans l'exercice de transcription pour des remarques relatives.

Remarques historiques

Les contenus de cette charte contribuent aux stratégies mises en œuvre par le clergé régulier afin d'intensifier la gestion de son temporel. La charte réalise deux opérations complémentaires, une juridique et une administrative. D'un point de vue juridique, le chapitre fait appel à son monopole de l'écrit pour notifier un contrat dont les dispositions favorisent clairement l'abbaye au détriment des paysans: ces derniers obtiennent certes l'usufruit d'un ensemble de terres appartenant au clergé mais ceci contre paiement d'un impôt conséquent dans des conditions très contraignantes. La clause de garantie (qui justifie à elle-seule la rédaction de l'acte) spécifie en effet qu'à défaut de paiement le jour dit, les paysans se verraient retirer jusqu'à l'intégralité de leur récolte et ils perdraient par ailleurs l'usufruit des terres. D'un point de vue administratif, la charte décrit minutieusement les terrains en question (par exemple *en la voie de Donneris as Genoivres*). En l'absence de cadastre, une telle description qui combine la microtoponymie et l'infrastructure routière correspond à une mainmise sur le territoire, mainmise qui va de pair avec celle de l'abbaye sur les hommes de sa contrée.

Le projet pédagogique d'*Ad fontes* a fourni également l'impulsion à la rédaction d'une synthèse des critères d'édition en vigueur dans l'édition des *Plus anciens documents*. Celle-ci présente intégralement les éléments de l'encodage double, même si les exercices concrets d'*Ad fontes* ne peuvent prendre en compte certains éléments particulièrement épineux en raison de l'intégration du volet galloromaniste dans le projet préexistant. Grâce à la souplesse de notre collègue médiéviste responsable du projet, Simon Teuscher, les critères appliqués sont toutefois nettement plus linguistiques que ceux que l'on rencontre habituellement en histoire médiévale.

Connaissant l'importance que Günter Holtus attache aux outils et à la doctrine de l'enseignement, nous sommes certain que cette dernière étape sur le chemin de notre charte lorraine trouvera son plein accord, accord dont nous savons apprécier la valeur. L'élève aîné de Max Pfister est donc tout autant présent à la fin de ce long cheminement d'un document vers l'ère multimédiale qu'il l'a été à ses débuts.

Bibliographie

- «Ad fontes» - *Eine Einführung in den Umgang mit Quellen im Archiv*, <www.adfontes.uzh.ch/1000.php>.
- Gärtner, Kurt, et al. (edd.), *Skripta, Schreiblandschaften und Standardisierungstendenzen* (Beiträge zum Kolloquium vom 16. bis 18. September 1998 in Trier), Trier, THF, 2001.
- Gärtner, Kurt/Holtus, Günter (edd.), *Überlieferungs- und Aneignungsprozesse im 13. und 14. Jahrhundert auf dem Gebiet der westmitteldeutschen und ostfranzösischen Urkunden- und Literatursprachen. Beiträge zum Kolloquium vom 20. bis 22. Juni 2001 in Trier*, Trier, Kliomedia, 2005.
- Gleßgen, Martin-Dietrich, *Das altfranzösische Geschäftsschrifttum in Oberlothringen: Quellenlage und Deutungsansätze*, in: Gärtner et al. 2001, 257–294.
- Gleßgen, Martin-Dietrich, *Editorische, lexikologische und graphematische Erschließung altfranzösischer Urkundentexte mit Hilfe von TUSTEP. Stand der Arbeiten*, in: Gärtner/Holtus 2005, 91–107.

- Gleßgen, Martin-Dietrich, *Bases de données textuelles et lexicographie historique: l'exemple des Plus anciens documents linguistiques de la France*, in: Trotter 2007, 373–380.
- Gleßgen, Martin-Dietrich, *Les «lieux d'écriture» dans les chartes lorraines du XIII^e siècle*, *Revue de linguistique romane* 72 (2008), 413–540.
- Gleßgen, Martin-Dietrich, *Les plus anciens documents linguistiques de la France. Édition électronique*, Collection fondée par Jacques Monfrin, poursuivie par M.-D. G., en collaboration avec Françoise Viellard et Olivier Guyotjeannin, en partenariat avec Paul Videsott, <www.mediaevistik.uzh.ch/docling>, 2009 [1133 documents (1224–1275), Marne, Haute Marne, Meurthe-et-Moselle, Meuse].
- Gleßgen, Martin-Dietrich, *Élaborations philologiques et linguistiques sur la base de corpus textuels en français ancien – architecture du projet*, in: Maria Iliescu/Heidi Siller-Runggaldier/Paul Danler (edd.), *Actes du XXV^e Congrès International de Linguistique et Philologie Romanes* (sept. 2007, Innsbruck), Berlin, De Gruyter, 2010, tome 6 (section XIII – *Philologie et linguistique des textes anciens; lexicographie diachronique*), 135–149.
- Gleßgen, Martin-Dietrich/Kihai, Dumitru/Videsott, Paul, *L'élaboration philologique et linguistique des «Plus anciens documents linguistiques de la France. Édition électronique»*, *Trois études réunies par M.-D. G., D. K. et P. V.*, *Bibliothèque de l'École des Chartes* 168 (2010) [2011], 5–94.
- Holtus, Günter/Völker, Harald, *Diskussion aktueller Probleme: Editionsriterien in der Romanischen Philologie*, *Zeitschrift für romanische Philologie* 115 (1999), 397–409.
- Kunstmann, Pierre/Stein, Achim (edd.), *Le Nouveau Corpus d'Amsterdam* (Actes de l'atelier de Lauterbad, 23–26 février 2006), Stuttgart, Steiner, 2007.
- Overbeck, Anja, *Literarische Skripta in Ostfrankreich. Edition und sprachliche Analyse einer französischen Handschrift des Reiseberichts von Marco Polo (Stockholm, Kungliga Biblioteket, Cod. Holm. M 304)*, Trier, Kliomedia, 2003.
- Pitz, Martina, *Volkssprachige Originalurkunden aus Metzger Archiven bis zum Jahr 1270*, in: Gärtner et al. 2001, 295–392.
- Stein, Achim, *Le Nouveau Corpus d'Amsterdam*, <www.uni-stuttgart.de/lingrom/stein/corpus>, 2010.
- Tittel, Sabine, *Le «DEAF électronique» – un avenir pour la lexicographie*, *Revue de linguistique romane* 74 (2010), 301–311.
- Trotter, David A. (ed.), *Actes du XXIV^e Congrès International de Linguistique et Philologie Romanes* (sept. 2004, Aberystwyth), vol. 1, section 2: *De la philologie aux nouveaux médias: éditions de textes – linguistique de corpus – analyse informatique du langage*, Tübingen, Niemeyer, 2007, 285–480.
- Videsott, Paul, *À propos du plus ancien document en français de la chancellerie royale capétienne*, *Bibliothèque de l'École des Chartes*, 168 (2010) [2011], 61–81.
- Völker, Harald, *Skripta und Variation. Untersuchungen zur Negation und zur Substantivflexion in altfranzösischen Urkunden der Grafschaft Luxemburg (1237–1281)*, Tübingen, Niemeyer, 2003.