Zeitschrift: Revue de linguistique romane
Herausgeber: Société de Linguistique Romane
Band: 64 (2000)
Heft: 253-254

Artikel: Ancien et moyen français sur le web : textes et bases de données
Autor: Kunstmann, Pierre
DOI: https://doi.org/10.5169/seals-400011

ANCIEN ET MOYEN FRANÇAIS SUR LE WEB: TEXTES ET BASES DE DONNÉES

Depuis plusieurs décennies, le texte dispose, avec l'ordinateur, d'un support nouveau; avec le web, il bénéficie, depuis quelques années, d'un moyen de diffusion radicalement différent, à la vitesse de la lumière, et couvrant virtuellement toute la planète. Cette situation n'est pas sans entraîner des conséquences importantes pour notre domaine d'études. Le présent article vise à signaler divers types de travaux consacrés à l'ancien et au moyen français (AF et MF), qu'on trouve maintenant sur la «Toile», le plus souvent en toute liberté d'accès. Nous nous intéresserons principalement aux textes anciens, qu'on peut lire en continu, et aux bases de données, qu'on peut interroger. Au préalable, cependant, nous mentionnerons des sites répertoires, panneaux indicateurs précieux pour nous mener sur les pistes du Web et nous apprendre à distinguer la pépite de la pacotille: effectivement, ce qui reluit n'est pas toujours de l'or!

1. SITES RÉPERTOIRES

Nous entendons par là des sites dont la fonction est de répertorier d'autres sites. Pour tout ce qui touche au français médiéval, le passage par *Ménestrel* (**Mé**diévistes sur l'Inter**Ne**t / **S**ources, **T**ravaux, **R**éférences **E**n **L**igne[1] s'impose. Ce groupe s'est formé à partir de l'équipe du *Médiéviste et l'Ordinateur*, en liaison avec l'École Nationale des Chartes et l'Institut de Recherche et d'Histoire des Textes, pour "favoriser le développement de ressources européennes et plus particulièrement francophones sur Internet"; il offre un répertoire intitulé *Ressources et outils documentaires sur Internet pour les médiévistes*, ordonné en 5 sections («Documents», «Ressources collégiales», «Thèmes», «Catalogue et index», «Instruments d'exploration sur internet»), qui ne prétend pas à l'exhaustivité, mais se veut "un choix raisonné et critique de sites". La première section comprend les rubriques suivantes: «Bibliographies», «Cartes et plans», «Dictionnaires», «Icônes, objets, peintures...», «Manuscrits enluminés», «Musique», «Ressour-

(1) www.ccr.jussieu.fr/urfist/omedirht.htm

ces pédagogiques», «Revues et articles», «Textes» et «Thèses et mémoires». L'avant-dernière rubrique (page réalisée et régulièrement mise à jour par R. Pellen) signale quelque 80 textes d'AF et de MF, chacun étant accompagné d'un bref paragraphe descriptif et critique; l'inventaire est bien fait, nous y renvoyons le lecteur. La plupart des documents ou fichiers dont nous parlerons dans cet article sont signalés, avec leur adresse, dans *Ménestrel*.

On aura avantage à consulter aussi la rubrique «Un choix de sites»(2) (elle aussi régulièrement actualisée) du **C**entre d'**É**tudes des **T**extes **M**édiévaux (CETM) de l'Université Rennes 2 Haute Bretagne. On y trouve des liens médiévaux, répartis en plusieurs sections; celle qui est consacrée aux «Textes médiévaux et approches critiques» contient les subdivisions suivantes: «Encyclopédies médiévales», «Matière de Bretagne», «Théâtre», «Fabliaux et Textes Moraux», «Textes épiques», «Chroniqueurs», «Textes tardifs», «Trouvères et Troubadours»; les commentaires descriptifs ou critiques y sont exceptionnels.

Dans le domaine français, le chercheur peut aisément se limiter à ces deux sites, par lesquels il peut rejoindre au besoin certains grands sites plus généraux (*Labyrinth* notamment, pour les études sur le moyen âge européen).

2. TEXTES EN CONTINU

Les textes ou images de textes d'AF et de MF commencent à s'accumuler sur le Web; aussi importe-t-il de savoir à qui on a affaire, d'où viennent les textes, comment ils sont présentés et quel degré de fiabilité on peut accorder à ces documents. Une première distinction s'impose, logique (parfois déontologique...): il faut effectivement faire le départ entre ce qui tient plutôt de la reproduction d'un document antérieur et ce qui est véritablement production d'un document nouveau.

2.1. *Reproductions de documents*

Les textes présentés sous format «image» relèvent, quasi par définition, de la première catégorie; c'est le cas de la grande majorité de ceux que, pour le moyen âge, Gallica Classique offre au lecteur sur le site Web de la BNF(3); le seul document en format «texte» est l'édition des *Tristan*

(2) www.uhb.fr/alc/medieval/liens_ma.htm#matière de Bretagne

(3) gallica.bnf.fr/classique/

en vers par J.-C. Payen (Garnier, 1974). Parcourir les images des autres ouvrages requiert des talents d'acrobate (en l'occurrence «Acrobat» de Adobe) et n'est certainement pas à la portée de l'usager moyen, même avec un bon appareil. Les images peuvent être lisibles, mais elles sont inutilisables pour toute recherche lexicale.

Il s'agit aussi de reproduction quand est transcrite sous format «texte» une édition déjà ancienne, entrée dans le domaine public. On peut ainsi lire, sur le site du CETM, *Erec et Enide* de Chrétien de Troyes dans l'édition qu'en a proposée W. Foerster(4). L'édition est composite, certes, mais sa reproduction soignée a demandé un important travail de révision et le format permet d'en explorer aisément le texte. Il est exceptionnel qu'on reproduise une édition récente, comme dans le cas du *Mystère de Saint Christofle* et du *Mystère de Sainte Venice*, pièces éditées par G. Runnalls, en 1973 et 1980 respectivement, dans les *Exeter French Texts Studies*; ces éditions étant épuisées, les responsables de la collection ont autorisé le CETM à les transposer électroniquement pour le Web. Ici encore les textes sont bien présentés et la lecture en est agréable.

Les textes offerts par les grandes bibliothèques électroniques doivent être considérés à part. Nous ne nous prononcerons pas sur la question des droits de reproduction, les limites semblant ici particulièrement floues. Mais il importe d'appeler un chat un chat et de préciser certaines choses. Le texte du *Tristan* de Béroul que propose ABU (Association des Bibliophiles Universels)(5) est une copie intégrale de l'édition des *Lettres gothiques* (1989)(6); pourquoi ne pas le mentionner? En Italie et en Amérique du Nord, il semble que la plupart des éditeurs acceptent que les textes d'éditions critiques soient reproduits intégralement (sans apparat); c'est une façon d'inciter le lecteur à se procurer ensuite le livre. D'ailleurs à la *Bibliotheca Augustana*, U. Harsch, qui reproduit à tour de bras des textes de littératures diverses, indique la source de ses textes; pour le *Tristan* de Thomas, il s'agirait d'une combinaison des éditions de J.-C. Payen et de C. Marchello-Nizia. En fait, le texte est bien emprunté au premier; c'est le fragment de Carlisle qui est tiré du volume de la Pléiade, morceau édité, comme on sait, par I. Short...

---

(4) Avec cependant, au vers 2268, une correction effectuée par M. Rousse...

(5) cedric.cnam.fr/cgi-bin/ABU/go?bertris1
«Texte produit par Christine Cloux» (apparaissant comme "copiste" dans la Notice); avec un moteur de recherche utile. Reconnaissons que, par ailleurs, pour la reproduction de la *Chanson de Roland* du ms. d'Oxford, le "copiste" est précisé (B. Woledge), et même l'intermédiaire...

(6) Cf. la leçon *voirre* au v. 2032...

Signalons au passage que l'*Oxford Text Archive* n'est pas riche en textes français du moyen âge. Il semble qu'il n'y en ait qu'un qui soit d'accès libre sur le Web; il est malheureusement bourré de fautes! Il s'agit de l'édition d'*Aliscans* par E. Wienbeck, W. Hartnacke et P. Rasch (Halle, 1903), «compiled by Steuart D. Hamilton»; le «compilateur» a réussi le tour de force de commettre 9 fautes – coquilles? – dans les 10 premiers vers (à vrai dire, 8 fois de la même erreur: confusion entre *n* et *u*).

Une autre forme de reproduction consiste à s'attacher, non pas au texte, mais au manuscrit qui l'a conservé. Les folios du manuscrit sont alors saisis sous format «image» et placés sur la Toile. Le meilleur exemple en est actuellement le «Projet *Charrette*», réalisé par l'équipe de K. Uitti à l'Université Princeton(7); on peut en consulter une version française sur le site du **C**entre d'**É**tudes **S**upérieures de **C**ivilisation **M**édiévale (CESCM, Université de Poitiers)(8). Les feuillets de 7 des 8 manuscrits du roman y sont reproduits avec une clarté et une lisibilité remarquables. L'un de ces manuscrits réapparaît sur le site du Laboratoire de Français Ancien (LFA)(9) de l'Université d'Ottawa: il s'agit des folios du ms. Garrett 125 où se trouve le *Chevalier au Lion*, reproduits avec la permission des "Princeton University Libraries". On annonce sur le Web deux autres projets de ce genre: la reproduction (sauf erreur, car le site est en cours d'élaboration et la présentation manque de précision) du manuscrit K (Lyon) de la *Quête du saint Graal*(10) à l'École Normale Supérieure de Fontenay / Saint-Cloud, et surtout «The *Lancelot-Graal* Project», qui se propose, dans une première phase, de photographier et de numériser le texte et les images de 5 manuscrits du roman en prose(11). Notons aussi dans ce domaine, sur le site du **C**entre d'**É**dition de **T**extes **É**lectroniques (CETE) de la Faculté des Lettres de Nantes(12), la reproduction d'une reproduction phototypique de manuscrit (le BNF fr. 19152), jadis publiée par E. Faral. Par ailleurs, la Bodleian Library d'Oxford a placé sur son site(13) d'excel-

(7) www.princeton.edu/~lancelot/

(8) www.mshs.univ-poitiers.fr/cescm/lancelot/

(9) aix1.uottawa.ca/academic/arts/lfa/

(10) www.lexico.ens-fcl.fr

(11) vrlab.fa.pitt.edu/STONES-WWW/VAlanc.html
Les mss sont les suivants: Amsterdam BPH1; Manchester Rylands Fr. 1; Oxford, Bodl. Douce 215; Londres, BL Add. 10292-4 et Royal 14.E.III. Collaborent au projet E. Kennedy, R. Middleton, K. Busby, M. Alison Stones, S. Blackman, M. Meuwese et K. Sochats.

(12) palissy.humana.univ-nantes.fr/CETE/CETE.html

(13) image.ox.ac.uk/pages/bodleian/medcanon.htm

lentes reproductions, très nettes, des feuillets de manuscrits importants pour la littérature médiévale française (notamment celui de la *Chanson de Roland* et du *Tristan* de Thomas). Il faut se féliciter d'une telle entreprise, souhaiter que la BNF lui emboîte le pas. Les grandes bibliothèques mondiales sont plus qualifiées et surtout mieux équipées que les universitaires pour procéder à la numérisation et assurer la diffusion et la conservation de tels documents.

2.2. *Production de documents nouveaux*

Passons à la production de documents nouveaux, en quelque sorte à valeur philologique ajoutée. On pourrait en distinguer 3 types:

- la transcription (diplomatique ou semi-diplomatique) de manuscrit,
- l'édition critique,
- le texte encodé.

Le gros travail effectué par K. Meyer (Université de Copenhague) sur le *Chevalier au Lion*(14) illustre bien le premier type. C'est une transcription quasi diplomatique (la segmentation des mots correspond à l'usage moderne) de tous les manuscrits et fragments qui ont conservé le roman, une extension, en quelque sorte, de l'étude qu'elle avait fait paraître en 1995 sur la copie de Guiot. Les textes se présentent en paquets de lignes; ainsi pour le premier vers de l'œuvre:

H 1 **Artus li boens rois de Bretaingne /79v°a/
P 1 **Li boins roys Artus de Bretaigne /61r°a/
V 1 **Li bons rois Artus de Bretaigne /34v°a/
F 1 **Li bons rois Artus de Bretaigne /207v°b/
G 1 **Artus li bons rois de Breteigne /1r°b/
A 1 **Artus li boins rois de Bretaingne /174r°a/
S 1 **Artus li boins rois de Bretagne /72r°a/
R 1 **Artus li boins rois de Bretagne /40r°a/
Ly 1 **Li bo..s rois Artus de Bretaigne /1r°/

En tête de ligne on trouve le sigle du manuscrit, ensuite le numéro du vers, puis le vers en question (les 2 astérisques signalent une initiale ornée), suivi de l'indication du folio. Les abréviations du copiste ont été développées et les lettres suppléées sont soulignées.

Par contre, les transcriptions des mêmes textes réalisées au LFA pour le *Dossier électronique du Chevalier au Lion*(15) peuvent être considérées comme semi-diplomatiques dans la mesure où la ponctuation y est utili-

(14) aix1.uottawa.ca/academic/arts/lfa/activites/textes/kmeyer/kpres.html
(15) aix1.uottawa.ca/academic/arts/lfa/activites/textes/chevalier-au-lion/index.html

sée, ce qui constitue déjà une première interprétation critique du document. Voici les premiers vers de la copie de Guiot:

79d.
1. Artus, li boens rois de Bretaingne,
2. La cui proesce nos enseigne
3. Que nos soiens preu et cortois,
4. Tint cort si riche come rois
5. A cele feste qui tant coste,
6. Qu'an doit clamer la Pantecoste.

Les transcriptions d'Ottawa ont été revues à Copenhague et vice versa(16). Le LFA présente d'autres transcriptions de textes: Chrétien de Troyes, *Le Conte du Graal*(17); Jean le Marchant, *Miracles de Notre-Dame de Chartres*(18); Alexandre du Pont, *Roman de Mahomet*(19); *Miracles de Notre-Dame tirés du «Rosarius»*(20); *Bestiaire marial tiré du «Rosarius»*(21).

Il faudrait mentionner à ce chapitre l'édition (travail en cours) du *Roman de Renart*(22) d'après les mss C et M, par une équipe japonaise de l'Université d'Hiroshima (N. Fukumoto, N. Harano et S. Suzuki). Ajoutons que les différents manuscrits du *Chevalier de la Charrette* ont fait l'objet de 2 séries de transcriptions: hyperdiplomatiques par le groupe de Princeton (mais les codes utilisés et le format SGML en rendent, pour l'instant, la lecture pratiquement impossible); diplomatiques par G. Jacquesson du CETE (cependant une brève vérification sur les images des premiers vers de la copie de Guiot fait apparaître trop d'erreurs; le travail devrait être revu soigneusement pour une prochaine mise à jour).

Dans le domaine du théâtre, le Web présente deux transcriptions (de type diplomatique) de pièces d'AF dues à O. Bettens (*Sponsus* et *Jeu de Robin et Marion*)(23). Le travail est original et la présentation a été conçue pour une utilisation en ligne.

---

(16) Si le lecteur remarque quelques différences, il s'agit des cas, rares, où les interprétations divergentes n'ont pu être accordées.
(17) Transcription du ms. Paris BNF fr. 794 par P. Kunstmann.
(18) Transcription par P. Kunstmann, suivant l'éd. P. A. Gratet Duplessis et la collation du ms. par C. Dunker.
(19) Transcription du ms. Paris BNF fr. 1553 par Y. Lepage.
(20) Transcription du ms. Paris BNF fr. 12483 par P. Kunstmann.
(21) Transcription du ms. Paris BNF fr. 12483 par A. Mattiacci.
(22) www.ipc.hiroshima-u.ac.jp/~france/RRenart.html#edition
Le site malheureusement ne donne pas de détail sur le caractère de l'édition.
(23) virga.org/robin
virga.org/cvf/exemples.html (texte et transcription phonétique des passages en langue vulgaire).

D'éditions de type critique préparées directement pour le Web, nous n'en connaissons pour l'instant que deux: celle des *Miracles de Notre-Dame* de Jean le Conte (travail en cours: le texte est édité, la présentation reste à paraître) et celle du *Miracle de l'enfant donné au diable*, publiées toutes deux dans les *Archives Miracles de Notre-Dame*, au LFA(24). Ce dernier travail comporte sept sections: Introduction, Texte, Apparat critique, Notes, Concordance, Index, Lexique, Analyse des formes verbales. L'introduction présente l'intérêt et l'importance de ce miracle. Une série de liens hypertextuels relient cette œuvre à des versions plus anciennes – ainsi en cliquant on peut faire venir une page où se trouve transcrit le vingt-deuxième miracle de Gautier de Coinci ainsi que sa probable source latine. D'autres liens mènent à des textes postérieurs: la compilation du franciscain Jean le Conte, écrite vers la fin du XIVe siècle, ou celle de Jean Miélot, secrétaire de Philippe le Bon. Ces liens sont établis entre éléments contenus dans le site (fichiers); d'autres sont faits avec l'extérieur (URL) et mettent tel point de l'introduction en relation avec tel site qui présente un point commun. Le texte du miracle peut se lire de façon continue en deux cellules juxtaposées: à gauche le texte de MF, à droite sa traduction en langue moderne. Lors de la lecture du texte, si une forme fait difficulté, le lecteur peut la noter et cliquer sur INDEX, dans le sommaire. En s'aidant de la fonction «recherche» de son navigateur, il pourra localiser la forme, qui apparaîtra classée sous le lemme correspondant. Cliquant alors sur ce lemme, il verra s'ouvrir le Lexique à l'article ou à la section d'article où la forme reçoit sa définition. Toutes les occurrences et toutes les acceptions des mots, lexicaux et grammaticaux, sont indiquées pour le premier miracle; les références comportent deux parties: l'une en chiffres romains pour indiquer le numéro du miracle dans le recueil; l'autre en chiffres arabes renvoie au vers. Quant à la dernière section, Analyse des formes verbales, il en sera question plus loin (3.2.).

Une autre édition critique fondée sur l'hypertexte est annoncée sur le Web, avec une belle maquette de présentation: il s'agit de l'édition de l'ensemble des poèmes à listes du XIIIe au XVIe siècle, dans le cadre du projet *Hyperlistes* de M. Jeay à l'Université McMaster(25).

Le corpus de textes encodés le plus important est celui de l'Université Libre d'Amsterdam: c'est le *Corpus des textes littéraires* constitué sous la responsabilité d'A. Dees et de P. van Reenen et qu'il est question de

(24) aix1.uottawa.ca/academic/arts/lfa/activities/textes/archives_miracles_ND.html

(25) www.humanities.mcmaster.ca/~hyplist/

rendre maintenant accessible sur le Web. Chaque mot du texte est accompagné d'un indice d'analyse morphologique, dans la perspective de la confection de l'*Atlas des formes linguistiques des textes littéraires de l'ancien français* d'A. Dees. Instrument très utile pour l'étudiant qui déchiffre un texte ou le chercheur en quête de relevés rapides. Mais l'indexation empêche la lecture de ces œuvres en continu. En fait, un tel corpus est déjà pratiquement une base de données.

3. BASES DE DONNÉES

3.1. *Bases textuelles*

La concordance en constitue la forme la plus simple. Elle peut être de deux sortes: statique ou dynamique. Le premier type est le plus ancien, bien connu dans notre domaine. Le degré zéro en est la concordance dite "brute", qui résulte d'une simple opération de machine, donc sans travail philologique ajouté, ce qui la rend d'ailleurs précieuse, garantie contre toute dérive interprétative. Citons l'exemple des concordanciers complets des formes graphiques occurrentes publiés par le CREL/CUERMA à Aix-en-Provence. Au degré immédiatement supérieur, la concordance brute est accompagnée d'un répertoire où les graphies sont regroupées par vedettes; c'est le cas du travail effectué jadis sur le *Charroi de Nîmes*, premier titre de la collection *Textes et Traitement Automatique* publiée sous la direction de G. de Poerck. Gravissant un échelon, on trouve la concordance dite "lemmatisée", type maintenant courant, où les formes occurrentes sont regroupées avec leur contexte sous le lemme approprié (le choix du système de lemmes et la question du regroupement / dégroupement des entrées restant toujours des points délicats); c'est le genre de documents produits par l'équipe de Liège. Montant encore d'un degré, on obtient une concordance où les graphies sont classées sous la rubrique du lemme selon des critères morphologiques (nombre, genre, cas; modes, temps, personnes); c'est ainsi que se présente la *Concordance analytique de La Mort le Roi Artu* (P. Kunstmann et M. Dubé). Le sommet de l'échelle est certainement tenu par le travail de M.-L. Ollier, *Lexique et Concordance de Chrétien de Troyes*, où l'analyse est finement menée dans le menu détail; la classification des mots grammaticaux y atteint des limites qui ne sont plus dépassables...

Malgré toute son utilité, la concordance statique souffre d'un gros désavantage: sa taille! Comme l'a calculé D. Megginson (*Old-Fashioned Concording*[26]), une concordance d'un texte de 200 pages, à raison de

(26) www.chass.utoronto.ca/epc/chwp/merrily1/dlv_3.html

40 mots par citation, représenterait, avec la même taille de caractères, un texte de 8000 pages... La concordance de *La Mort le Roi Artu* frise les 2000 pages; quant à celle de Chrétien de Troyes, il a fallu la publier de façon compacte et sur microfiches, ce qui la rend d'ailleurs actuellement pénible à lire. D'où l'intérêt des concordances de type dynamique: à partir d'un texte numérisé, un logiciel permet à l'utilisateur de bâtir une concordance partielle ou complète; il peut alors généralement jouer sur deux tableaux: l'un présentant le mot recherché avec un contexte d'une page, l'autre en donnant les occurrences sous forme KWIC («Key word in context»: courtes citations avec le mot centré). Ce type de concordance est idéal pour un chercheur travaillant individuellement sur un texte spécifique, à son ordinateur. Par contre, il est difficile de l'offrir tel quel sur le Web – ne serait-ce que pour des raisons de droits de reproduction.

En fait, la concordance dynamique mène naturellement à un autre genre: la base de données interactive, qui, elle, est bien représentée sur la Toile (voir, par ex., *Rabelais et son temps* et *Recherche hypertextuelle dans la Comédie Humaine de Balzac* sur le site de l'INaLF). Pour l'AF et le MF, on peut depuis deux ans interroger ainsi la base *Textes de Français Ancien* (TFA)[27]. Établie par le LFA en collaboration avec l'American Research on the Treasury of the French Language (ARTFL) de l'Université de Chicago[28], cette banque textuelle est un produit dérivé de deux types de travaux effectués au LFA: d'une part les opérations de saisie, de correction et de formatage d'œuvres de l'ancienne langue (à partir des éditions modernes) pour la constitution d'une Base lemmatisée (voir infra); d'autre part, les transcriptions semi-diplomatiques mentionnées plus haut. Le fonds principal est constitué de textes des 12e et 13e siècles; à ce fonds se sont ajoutées des œuvres de MF. La banque fait l'objet de mises à jour régulières et s'enrichit à chaque fois par l'apport de nouveaux documents.

---

(27) http://humanities.uchicago.edu/ARTFL/projects/LFA/

(28) L'ARTFL, mis sur pied en 1981 à partir d'une collaboration entre l'Institut National de la Langue Française (INaLF) et la Division des Humanités et Sciences Sociales de l'Université de Chicago, a évolué cependant de façon autonome, sinon indépendante. La base de données principale ne correspond plus exactement à la base FRANTEXT de l'INaLF; s'y sont ajoutés d'autres projets, notamment une base de données sur les textes poétiques en ancien provençal et, en collaboration avec le CNR de Florence (consortium ItalNet), l'*Opera del Vocabolario Italiano* (1410 textes antérieurs à 1375, année de la mort de Boccace; quelque 17 millions de mots). L'entreprise américaine est actuellement dirigée par Robert Morrissey; Mark Olsen en est la cheville ouvrière.

Les textes sont saisis (par scan ou copie manuelle) le plus fidèlement possible. Aucune correction n'y est apportée; certaines corrections sont cependant suggérées entre crochets droits («l» après un crochet ouvrant signifie «lire plutôt»). On trouvera, sur le site du LFA, des notices détaillées que M. Plouzeau (Université de Provence) a consacrées à plusieurs de ces œuvres (*Ipomédon*, *Miracles de Saint Louis*, *Prise d'Orange*), y proposant un certain nombre de corrections. Pour une recherche fine sur ces textes, on aura donc intérêt à en tenir compte.

Les textes, une fois balisés (jalons indiquant diverses sections, les pages de l'édition, les folios du manuscrit ou les vers du poème), sont enregistrés en HTML et expédiés à Chicago, où ils sont alors placés dans la base TFA. Celle-ci fonctionne à l'aide du logiciel *Philologic*, sur système UNIX. La base, encore embryonnaire par rapport à l'OVI, son équivalent italien, comprend néanmoins actuellement quelque 2 millions de mots pour un total de 79 documents. Ceux-ci sont regroupés, par ordre alphabétique des titres, dans une Bibliographie, qui précise, pour chacun d'entre eux, l'édition (ou le manuscrit) utilisée, donne, le cas échéant, le nom de l'auteur et la date de rédaction, et spécifie le genre de composition (narratif, dramatique, didactique ou lyrique) ainsi que la forme (vers ou prose). Certaines indications de date ont subi une transformation qui pourrait surprendre un lecteur non averti. Pour la recherche par année ou par tranche d'années, le logiciel utilisé ne reconnaît que les chiffres. D'où, par exemple, les distorsions suivantes: «1250» pour «c. 1250»; «1250» pour «13e siècle»; «1200» pour «Début du 13e siècle»; «1299» pour «Fin du 13e siècle». Les œuvres sous droits sont signalées, dans la bibliographie, par un astérisque; les mots sont alors présentés avec un contexte réduit.

Voici la liste des œuvres figurant actuellement dans la base:

- *Aiol*, éd. J. Normand et G. Raynaud.
- *Ami et Amile*, éd. P. Dembowski.
- *Le Bestiaire Marial tiré du Rosarius (Paris, ms. BN fr. 12483)*, éd. A. Mattiacci.
- Chrétien de Troyes, *Le Chevalier au Lion (Yvain)*, transcription du ms. Paris, BN fr. 794 par P. Kunstmann.
- Pierre Sala, *Le Chevalier au Lion*, transcription du ms. Paris, BN fr. 1638 par P. Kunstmann.
- Chrétien de Troyes, *Le Conte du Graal (Perceval)*, transcription du ms. Paris, BN fr. 794 par P. Kunstmann.
- *Le Couronnement de Louis*, éd. Y. Lepage.
- *La Deuxième Continuation de Perceval*, éd. W. Roach.
- Gautier d'Arras, *Eracle*, éd. G. Raynaud de Lage.

- *La Folie Tristan de Berne*, éd. J. Bédier.
- *La Folie Tristan d'Oxford*, éd. J. Bédier.
- Hue de Rotelande, *Ipomédon*, éd. A. J. Holden.
- Marie de France, *Lais*, éd. K. Warnke.
- *Miracles de Nostre Dame par personnages*, éd. G. Paris et U. Robert (40 documents).
- Gautier de Coinci, *Miracles de Notre-Dame*, éd. V. F. Koenig (4 documents).
- Jean le Marchant, *Miracles de Notre-Dame de Chartres*, transcription par P. Kunstmann, suivant l'éd. P. A. Gratet Duplessis et la collation du ms. par C. Dunker.
- Guillaume de Saint-Pathus, *Les Miracles de saint Louis*, éd. P. B. Fay.
- *La Mort le roi Artu*, éd J. Frappier.
- *La Prise d'Orange*, éd. Cl. Régnier.
- *La Queste del Saint Graal*, éd. A. Pauphilet.
- *Le Roman d'Alexandre*, éd. E. C. Armstrong, D. I. Buffum, Bateman Edwards, L. F. H. Lowe, 1937 (4 branches).
- Jean Renart, *Le Roman de la Rose ou de Guillaume de Dole*, éd. F. Lecoy.
- *Le Roman de l'Estoire dou Graal*, éd. W. A. Nitze.
- *Le Roman de Renart*, éd. M. Roques (branches I, VII, VIII, IX, X, XI).
- *Le Roman de Thèbes*, éd. G. Raynaud de Lage (2 documents).
- Béroul, *Le Roman de Tristan*, éd. E. Muret, révisée par L. M. Defourques.
- Thomas, *Le Roman de Tristan*, éd. F. Lecoy.
- Guernes de Pont-Sainte-Maxence, *La Vie de saint Thomas Becket*, éd. E. Walberg.
- Benedeit, *Le Voyage de saint Brandan*, éd. I. Short et B. Merrilees.

L'accès est direct pour les abonnés à l'ARTFL; les autres utilisateurs doivent s'adresser au LFA pour obtenir un «user name» et un mot de passe; ceux-ci sont accordés aux étudiants et professeurs d'université ainsi qu'aux personnes reliées à un centre de recherche. Le mot de passe une fois donné, on obtient le «Search Form» (à partir d'ici les instructions sont en anglais seulement), grille d'interrogation.

La grille elle-même comporte deux parties. La première est consacrée à la définition du corpus: si l'on veut consulter la base entière, aucune sélection n'est requise; mais qui désire restreindre le corpus a plusieurs possibilités. On peut, en effet, préciser un auteur ou un groupe d'auteurs: chaque nom d'auteur, dans ce cas, doit être séparé par une virgule; ainsi quand on tape dans la case rectangulaire *chretien, renart, gautier de coinci*, un groupe de 7 textes est retenu: 2 de Chrétien de Troyes, 1 de Jean

Renart et 4 de Gautier de Coinci. La sélection peut se faire aussi par titre: si l'on cherche un document particulier, on peut se contenter de taper, sans accent, un mot caractéristique du titre (*thebes* par exemple, pour *Le Roman de Thèbes*); mais si le mot se retrouve dans un autre titre de la bibliographie, on risque d'avoir plusieurs documents. Le plus sûr, dans ce cas, est d'écrire le titre entier: par exemple *Miracle de saint Jehan Crisothomes*. Le mot *miracle* donnerait 46 documents: les 40 des miracles par personnages, 4 de Gautier de Coinci et les textes de Jean le Marchant et de Guillaume de Saint-Pathus. En tapant *miracle d*, on obtient tous les miracles dramatiques. La sélection se fait donc automatiquement par ordre alphabétique des lettres, de gauche à droite. Pour qui travaille sur une période particulière, il suffit d'indiquer une tranche chronologique dans la case «Dates»; ainsi quand on choisit *1300-1350*, le corpus se restreint, dans l'état actuel de la base, aux 12 premiers miracles par personnages. Il reste deux autres possibilités de définition:

– par genre: narration et théâtre sont bien représentés; le lyrisme se réduit aux chansons de Gautier de Coinci; il n'y a, pour l'instant, pas d'œuvre didactique;

– par mode: les textes sont en vers, à l'exception de deux romans en prose (*La Mort le roi Artu* et *La Queste del Saint Graal*).

La seconde partie de la grille d'interrogation est consacrée à la recherche lexicale (case «Word, words or phrases»). Celle-ci peut porter sur un mot ou plusieurs. Pour les troncatures et tous les cas de «pattern matching» on peut utiliser les symboles habituels de UNIX:

– le point remplace un caractère quelconque. Si l'on tape *am.r*, on obtient en réponse les occurrences de *amer*, *amor*, *amur*.

– le point suivi d'un astérisque remplace toute suite de caractères. Tapant *am.*r*, on trouvera *amour*, mais aussi, le cas échéant, des indésirables: *amender*, *amonester*, etc. L'utilisation de l'astérisque est naturellement indispensable pour repérer les formes fléchies ou dérivées. Ainsi à partir d'*amor.** sont repérables les formes suffixées *amoreux*, *amorous*, *amoreusement*, *amorete*; pour les préfixes, *.*mordre* fera apparaître *amordre*, *desamordre*, *remordre*, etc.

– le dièse précédant une lettre indique qu'il s'agit d'une majuscule.

Quand la recherche porte sur plusieurs mots, on peut avoir recours aux opérateurs booléens. Si l'on s'intéresse à un thème ou un concept, l'opérateur «ou» (marqué par une ligne verticale) s'avérera utile. On peut, par exemple, facilement repérer chez Gautier de Coinci les verbes à l'infinitif exprimant le concept «frapper», récemment étudié ici même par

G. Lavis; il suffit d'écrire la suite *ferir|batre|boter|brochier|hurter|poindre|empeindre* et après validation on peut lire 41 occurrences dans les textes du prieur de Saint-Médard. L'opérateur «et» sert à chercher deux ou plusieurs mots dans la même phrase, quel que soit l'ordre dans lequel ils sont placés. A la demande *larges cortois* correspond ainsi dans *Le Roman d'Alexandre*:

> Biaus iert et avenans si tint son cors molt chier
> Et *larges* et *cortois*, n'i ot que ensegnier; [p. 100]

aussi bien que

> Se il fust crestïens, ainc tels rois ne fu nes,
> Si *cortois* ne si *larges*, si sages, si menbrés, [p. 355]

Remarquons cependant que le logiciel ne permet pas de préciser la distance (le nombre de mots) entre les deux unités choisies.

Si, par contre, on recherche une suite syntagmatique, il faut, à la rubrique «Search Options», cocher la case «Phrase Search». Un utilisateur qui voudrait obtenir les occurrences des formules assertives du type *se Dieu m'aist* dans le corpus des miracles dramatiques, pourrait taper *se Dieu*, cliquer sur «Phrase Search» et obtiendrait 226 occurrences, dont voici les 10 premières (en présentation KWIC):

Searching 40 documents for se + Dieu.* Total Occurrences = 226

MirPer1, 4: moustier? LE SEIGNEUR. Dame,	**se**	Dieu me vueille aidier, Je y alay
MirPer1, 25: la doulce vierge puissant,	**Se**	Dieu plaist, vous avez voir dit. J
MirPer2, 75:. Mes suers, il m'est bien,	**se**	Dieu plaist; Mais je me doubt d'es
MirPer2, 76: si est, dame, sanz raison;	**Se**	Dieu plaist, n'arez se bien non; N
MirPer2, 78: ER CLERC A L'EVESQUE. Sire,	**se**	Dieu me doint leesse, Je croy qu'i
MirPer2, 91: ien fait? L'ABBEESSE. Sire,	**se**	Dieu plaist, tel meffait Ne trouve
MirPer3, 119: Sainte Marie! BELOT. Oil,	**se**	Dieu me beneie, Sire, c'est il, n'
MirPer3, 124: ous avez dit bonne raison,	**Se**	Dieu m'aist. SECOND CHANOINE. Ja p
MirPer3, 124: desdit De riens qui soit,	**se**	Dieu me voie. Je croy que c'est la
MirPer4, 156: on seigneur. LE ROY. Amis,	**se**	Dieu te doint honneur, Va, si li d

Lui resterait alors à trier les exemples pertinents.

Le résultat d'une recherche lexicale est présenté par défaut sous forme de concordance, le mot sélectionné apparaissant en gras au centre d'un contexte d'une centaine de mots. S'il s'agit d'œuvres qui ne sont pas sous droits, on peut cliquer sur le numéro de la page et obtenir un contexte plus large, celui de la page entière; des indications «page précédente»/«page suivante» permettent même de remonter ou de descendre dans le texte. Le programme offre aussi un choix entre deux présentations particulières des résultats: le «rapport KWIC» (c'est l'exemple du para-

graphe précédent) et le «rapport de fréquence par titre». Dans le premier cas, le mot clé est centré, en gras, dans une ligne de texte; ici également, pour les œuvres du domaine public, un clic sur la référence élargit le contexte à toute la page. Dans le second cas, on n'obtient pas de texte, mais une liste (par ordre de fréquence décroissant) de titres d'œuvres où figure l'élément recherché. Voici, pour la séquence *se Dieu* retenue précédemment, les 10 premiers titres:

> Searching **40** documents for **se+Dieu.*** Total Occurrences = **226**
>
> **Frequency by title in descending numeric order:**
>
> 15 *Miracle du roy Thierry* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 5, 1880.) 1374
>
> 15 *Miracle de la fille du roy de Hongrie* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 5, 1880.) 1371
>
> 13 *Miracle de la fille d'un roy* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 7, 1883.) 1379
>
> 13 *Miracle de l'empereris de Romme* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 4, 1879.) 1369
>
> 10 *Miracle de un enfant que Nostre Dame resucita* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 2, 1877.) 1353
>
> 10 *Miracle de sainte Bautheuch* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 6, 1881.) 1376
>
> 10 *Miracle de la femme du roy de Portigal* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 1, 1876.) 1342
>
> 10 *Miracle de Robert le dyable* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 6, 1881.) 1375
>
> 8 *Miracle de saint Jehan Crisothomes* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 1, 1876.) 1344
>
> 8 *Miracle de Theodore* (Miracles de Nostre Dame par personnages, éd. par Gaston Paris et Ulysse Robert, Paris, SATF, tome 3, 1878.) 1357

Chaque type de rapport est suivi d'indications bibliographiques pour les textes cités.

La base TFA présente le grand avantage d'être accessible gratuitement à tous les chercheurs œuvrant dans le domaine du français ancien.

Les responsables comptent l'enrichir régulièrement, en accordant la priorité à l'AF et particulièrement aux textes du XII^e siècle. Il est question, à moyen terme, d'utiliser le dictionnaire des formes en voie d'établissement au LFA (voir infra) pour une semi-lemmatisation de la base. Ainsi en tapant *avoir*, par exemple, l'automate rechercherait les occurrences de toutes les formes graphiques de ce verbe. Il serait souhaitable que les TFA deviennent, pour les chercheurs travaillant sur l'AF, l'équivalent, en quelque sorte, des bases ARTFL et FRANTEXT pour les spécialistes du français classique et moderne.

Par contre, si les recherches portent sur la période 1330-1500, mieux vaudrait, bien sûr, rendre accessible sur le Web le trésor textuel accumulé par l'INaLF, au fil des années, en vue de la rédaction du *DMF*. Les Bases textuelles de moyen français (BTMF) comprennent principalement une banque d'œuvres saisies intégralement (220 au total, ce qui représente quelque 6 millions d'occurrences); on peut y accéder sur le Web, mais de façon limitée: le mot de passe est réservé aux membres de l'équipe de rédaction du dictionnaire. Il serait souhaitable que ces documents soient réellement accessibles "par les modalités de Frantext" (comme l'indique, par anticipation, R. Martin dans sa présentation du premier fascicule, à tirage limité, du *DMF*). Un récent sondage effectué par l'INaLF indique d'ailleurs qu'un bon nombre d'abonnés aimeraient que FRANTEXT soit consultable gratuitement sur l'internet - ce qui, d'après les responsables de la base, ne serait possible que pour les œuvres du domaine public.

Pour le MF, on peut également consulter la banque de textes entreprise vers la fin des années 80 par U. Jokinen à l'Université de Jyväskylä. Couvrant la période 1300-1550, elle comporte aujourd'hui plus d'un million d'occurrences; y figurent surtout des textes en saisie intégrale[29]. Ce corpus n'est cependant pas accessible sur le Web.

Ajoutons qu'il existe une autre base de données textuelles, celle constituée naguère par l'**É**quipe **L**inguistique et **I**nformatique (ELI), autour de C. Marchello-Nizia, à l'École Normale Supérieure Fontenay / St-Cloud. La base est plus restreinte quantitativement, mais s'étend sur une période plus longue, du XI^e au XVI^e siècle. Les textes saisis et vérifiés sont accompagnés d'une concordance et peuvent être consultés sur place avec le logiciel *Analyser*, conçu par P. Bonnefois. L'ELI a fusionné avec l'UMR 9952 (Lexicométrie et textes politiques) pour constituer

(29) S'adresser, pour la consultation, à O. Merisalo (merisalo@tukki.jyu.fi), responsable du Corpus.

l'UMR 8503 (Laboratoire d'analyse de corpus linguistiques, usages et traitements – dont les directeurs sont P. Fiala et B. Habert). Cette unité, membre de l'INaLF, possède un site Web; en voici l'adresse:

www.lexico.ens-fcl.fr

Ce qui intéresse les linguistes travaillant sur le moyen âge se trouve malheureusement sous une grosse rubrique *Intranet des projets en cours (Accès réservé)*. Comme l'on sait, l'intranet étant l'inverse de l'internet, on ne peut accéder aux sous-rubriques, la principale d'entre elles étant justement la *Base de Français Médiéval* (BMF). En cliquant, par exemple, sur la subdivision *Textothèque Médiévale*, on est immédiatement accueilli par un message où l'on se voit réclamer un «User Name» et un «Password», sans qu'on puisse savoir d'ailleurs où solliciter ces mots de code. Le chercheur qui n'est pas de la maison peut cependant cliquer sur une autre subdivision, *Catégorisation de la BFM*, et trouver une page où il peut tout apprendre sur ce chapitre (l'intranet n'est donc pas tout à fait étanche). Il apparaît que les textes (en totalité ou en partie?) sont actuellement l'objet d'une procédure d'étiquetage, morphologique principalement, mais avec parfois des retombées syntaxiques; le programme utilisé, pour ce faire, est SATO («Système d'analyse de textes par ordinateur»), conçu il y a plusieurs années par F. Daoust, à l'Université du Québec à Montréal.

3.2. *Fichiers d'analyse morphologique et syntaxique*

Deux fichiers de ce type, en cours d'élaboration, sont actuellement ouverts et consultables sur le Web. Le premier est une *Base grammaticale du Chevalier au Lion* (sur le site du LFA), où F. Martineau analyse en priorité les formes verbales du ms. H (copie de Guiot). Les fiches, pour le verbe, sont structurées selon les critères morphologiques et syntaxiques suivants: forme occurrente; nombre d'occurrences de cette graphie; lemme; source du lemme (le Tobler-Lommatzsch; sinon, l'édition); catégorie grammaticale; texte (le manuscrit retenu); référence (numéro de vers); lien (voir infra); personne; temps (simple ou composé); mode; voix; construction (les constructions à possession inaliénable sont distinguées des constructions personnelles); citation (quelques lignes du texte); en cliquant sur la rubrique «Lien», on obtient, pour l'occurrence sélectionnée, un très large contexte: celui d'une des 10 sections de la transcription placée sur le site du LFA. Les données, gérées par le logiciel FileMaker de Claris, sont accessibles suivant deux modes: visualisation et recherche. Le premier permet de parcourir le fichier, de la première à la dernière fiche; le second offre de multiples possibilités d'interrogation, de la simple recherche d'une forme à la demande combinant plusieurs rubriques. Si la question porte sur la forme occurrente

ou sur le lemme, on peut même la moduler de cinq façons, en choisissant entre «contient», «égale», «n'égale pas», «commence par» et «finit par».

Le deuxième fichier, qui utilise aussi FileMaker, est également consacré au verbe et placé sur le site d'Ottawa. Il s'agit de l'*Analyse des formes verbales du Miracle de l'enfant donné au diable*, document réalisé par Lene Schøsler (Institut d'Études Romanes, Université de Copenhague). L'auteur présente, dans une première section, des précisions définitoires concernant la grille établie pour l'examen de ces formes; ses considérations portent sur les divers types d'emploi (auxiliaire, verbe support, verbe à sens plein...), sur les diverses acceptions des verbes polysémiques, sur les distinctions de voix, de mode et de temps, ainsi que sur la valence (réalisée ou maximale). La deuxième section contient les fiches d'analyse; voici, en exemple, la première, qui correspond à la forme *abregeras*:

forme graph. occ.	abregeras
lemme	abréger
source du lemme	○ TL ○ Gdf ○ SGdf ○ Ed
texte	Mirl
référence	1452
forme: simple, composée	● simple ○ 1 auxil. ○ 2 auxil.
voix	● actif ○ passif ○ pronom. ○ pronom à sens passif ○ pronom. réfl. ○ pronom. réciproque
mode	○ inf. ○ part. «prés.» ○ part. «pas.» ○ subj. ○ impér. ● ind.
temps	○ prés. ○ pas. simple ○ imp. ● fut. ○ cond.
personne	○ 1 ● 2 ○ 3 ○ 4 ○ 5 ○ 6
valence réalisée	○ 0 actant ● 1 actant ○ 2 actants ○ 3 actants ○ 4 actants
autres compléments	● oui ○ non
valence maximale	○ 1 actant ● 2 actants ○ 3 actants ○ 4 actants
commentaire	S implicite; COD: «tes voies»

Quand on affiche le mode «recherche», le logiciel permet, comme pour le fichier précédent, d'interroger la base à partir d'une seule rubrique ou de plusieurs à la fois. Les références correspondent à l'édition mentionnée dans la première partie de cet article.

Il convient sans doute de signaler ici que commencent à être placés sur le Web (même site) des fichiers, par ordre alphabétique, correspondant

aux données qui avaient été réunies par R. Martin à l'INaLF, au début des années soixante. Ce dernier avait, en effet, constitué un *Fonds de formes flexionnelles* consacré au verbe, d'une taille considérable (entre 16000 et 20000 fiches manuscrites de formes verbales analysées, pour la période allant de l'AF au français de la Renaissance). Les graphies étaient tirées des grands dictionnaires (*Gdf*, *TL* et *HUGUET*), qui avaient été dépouillés de façon exhaustive, des manuels de morphologie ainsi que d'une trentaine d'éditions critiques d'AF. Ces fiches sont actuellement en dépôt à Ottawa, dans le cadre de la convention qui unit l'INaLF et le LFA. Elles ont fait l'objet d'une révision, d'une réorganisation du point de vue du choix des lemmes et d'une saisie électronique. Les données se trouvent maintenant dans un fichier FileMaker (intitulé *Base de graphies verbales*), à raison d'une fiche par forme graphique. A titre d'exemple, la graphie *ennuiera* y est analysée ainsi:

forme	ennuiera
lemme	enoiier
dictionnaires	○ TL ◉ Gdf ○ GdfS ○ Hug.
autres sources	
mode	○ inf. ○ part. ○ subj. ○ impér. ◉ ind.
temps	○ prés. ○ pas. (ps/pp) ○ imp. ◉ fut. ○ cond.
personne	○ 1 ○ 2 ◉ 3 ○ 4 ○ 5 ○ 6
remarques	

Ici encore, le logiciel permet toutes les possibilités d'interrogation. Les fiches sont classées dans l'ordre alphabétique des lemmes; ceux-ci correspondent aux vedettes du *TL*, à défaut à celles du *Gdf* et sinon à celles de *HUGUET*.

3.3. *Index lexicaux*

Un texte une fois numérisé, il est facile d'en tirer des listes de mots lexicaux et grammaticaux, ainsi que de noms propres, de façon à embrasser le vocabulaire de l'œuvre dans son intégralité. Un programme d'indexation peut livrer, en quelques secondes, un index brut, c'est-à-dire une liste alphabétique des formes occurrentes, avec indication de fréquence et de référence; ce résultat peut ensuite être l'objet de différents traitements: minimum, quand on veut en tirer séparément une liste de noms propres; plus important, délicat et plus long quand on procède à une lemmatisation des formes occurrentes (regroupées alors par mots-vedettes). Le traitement serait maximal si la lemmatisation s'accompagnait d'une analyse morphologique détaillée.

Des index de ce genre sont souvent diffusés sur support papier. A notre connaissance, le site du LFA est le seul actuellement à en présenter en libre accès sur le Web. Certaines des œuvres qu'on peut y lire en transcription semi-diplomatique (*Bestiaire marial*, *Miracles de Notre-Dame de Chartres*, *Roman de Mahomet*) sont suivies d'un index complet; la liste des mots lexicaux et gramaticaux y est divisée en tranches alphabétiques; la liste des noms propres se consulte séparément. Le LFA a entrepris également d'établir et d'afficher sur son site des index bruts pour tous les textes de sa base TFA, ce qui donnera de ces œuvres une vue différente de la présentation en contexte sur le serveur de Chicago. Pour ce qui est des index lemmatisés, on en compte pour l'instant 3 au LFA:

– l'index du *Miracle de l'enfant donné au diable*, qui ouvre une série de 40, pour l'ensemble du recueil

– les index du *Couronnement de Louis* (éd. Y. Lepage) et du *Conte du Graal* (copie de Guiot, transcription de P. Kunstmann), qui constituent le début de la *Base lemmatisée d'ancien français* dont il sera question plus bas.

Signalons ici une initiative conjointe avec le CESCM. Le LFA, dans le cadre de son projet *Chevalier au Lion*, procède actuellement à la lemmatisation des textes de tous les manuscrits et fragments de manuscrits où l'on trouve ce roman; une équipe du CESCM, sous la responsabilité de R. Pellen, lemmatise le texte critique du *Chevalier de la Charrette (Lancelot)*, édité par A. Foulet et K. Uitti, et envisage de passer ensuite à la lemmatisation des textes des manuscrits, tels qu'ils ont été transcrits par le groupe de Princeton. Étant donné les liens qui unissent les deux romans (en particulier les allusions dans l'un à l'action se déroulant dans l'autre), il serait intéressant d'en rassembler le vocabulaire dans une base qui serait une étape vers la réalisation d'un index lemmatisé des textes des manuscrits de l'œuvre de Chrétien.

Ces index lemmatisés, s'ajoutant à la saisie en cours (effectuée au LFA) des graphies du *Gdf* et aux formes occurrentes de la *Base de graphies verbales*, serviront à l'établissement d'un dictionnaire des formes de l'AF. On prévoit d'ailleurs, dans le cadre d'un accord entre l'Institut d'Études Romanes de l'Université de Copenhague et le LFA, l'exploitation à cet effet du *Corpus des textes littéraires* d'Amsterdam (voir supra); il serait intéressant d'enrichir ce dictionnaire par l'apport des données de la BFM. Ce dictionnaire pourra servir à une meilleure consultation de la base TFA à Chicago; augmenté régulièrement, il pourrait être mis sur le Web et consulté à distance ou téléchargé.

Parallèlement au dictionnaire des formes, il importe d'indiquer la parution d'un dictionnaire des lemmes: le *Lexique d'ancien français* de D. Walker (Université de Calgary). On peut le trouver en cliquant sur le site du LFA ou en passant directement à l'adresse suivante:

www.acs.ucalgary.ca/~dcwalker/Dictionary/dict.html

Cette base contient 48000 mots tirés du *TL*; il s'agit des lemmes à proprement parler (les vedettes retenues, dans le dictionnaire, pour les différents articles; elles apparaissent en caractères gras) et de certaines graphies particulièrement fréquentes; celles-ci sont de deux sortes: en caractères gras il s'agit de renvois aux vedettes qui y correspondent; en caractères ordinaires, ce sont des variantes graphiques qui suivent immédiatement le mot-vedette, et sur la même ligne, au début de l'article. L'auteur du lexique appelle la vedette «forme primaire» et réserve au reste le nom de «variante» (y incluant même les vedettes d'articles différents mais reliés l'un à l'autre par l'abréviation *vgl.* «confer»). Il reprend en gros les principes qui l'ont guidé pour la confection du *Dictionnaire inverse de l'ancien français* (Ottawa, 1982). La base, comme le dictionnaire, a bénéficié de la collaboration de feu H. H. Christmann, qui a communiqué à D. Walker les lemmes des fascicules non encore terminés. L'outil présenté est donc particulièrement précieux.

Le moteur de recherche utilisé permet une recherche fine et rapide. L'interrogation peut se faire par catégorie grammaticale (nous en avons compté pas moins de 58!) ou par composition graphique. Quand on choisit le deuxième type, on peut remplir jusqu'à 4 paires de cases pour préciser si le mot «commence par / se termine en», «comprend / contient» (et leur négation) telle ou telle suite graphique particulière. Ainsi, sélectionnant les formes commençant par *re* et se terminant en *aille*, on obtient les 9 entrées suivantes: *remembraille*, *repentaille*, *reponaille*, *reposaille*, *repostaille*, *resaille*, *retaille*, *retenaille*, *revisdaille*; si l'on ajoute la catégorie «sf» à la sélection, *resaille* («sm/adj») disparaît de la liste. Autre option pour les résultats de la recherche: on peut les obtenir dans l'ordre alphabétique normal (de gauche à droite) ou de façon inversée (de droite à gauche). Ainsi, dans l'ordre inverse, pour la terminaison *-aille*, on reçoit une liste de 207 lemmes (ou de 238 formes primaires et variantes, si l'on retient cette option) commençant par *aille*, *bäaille*, *bläaille*, *mäaille*, *gäaignemäaille*, *espargnemäaille*...

Il s'agit donc d'un magnifique instrument pour explorer, sinon directement le lexique de l'ancien français, du moins les vedettes conventionnelles du *TL* et indirectement les formes qu'on trouve réellement dans les

textes (par le biais des exemples cités dans le dictionnaire). Par contre, il semble que l'outil ne soit pas encore suffisamment au point en ce qui concerne les «variantes». Outre le choix de ce terme, qui paraît malheureux (*desfroi* et *frait*, par ex., ne peuvent évidemment pas être considérés comme des «variantes» des «formes primaires» *desfraitier* et *fraitier*) et un problème de machine qui fait disparaître les signes diacritiques dès qu'un mot ne figure plus dans la colonne de gauche («entrée») mais dans une des colonnes à la droite de celle-ci, l'apparition des variantes semble quelque peu aléatoire. Ainsi, pour reprendre les exemples proposés dans l'introduction du dictionnaire de 1982, si à ***erbee*** (entrée du *TL*), placé dans la colonne de gauche, correspond (*herbee*) [autre réalisation de l'entrée principale] dans la colonne de droite, à *herbee* placé à gauche ne correspond pas ***erbee** à droite (à vrai dire, dans ce cas, le logiciel semble ne pas même reconnaître l'entrée *herbee*). Certaines inconséquences peuvent être dues à des coquilles ou à des oublis: à *niele* à gauche correspond ***neele** à droite (alors que c'est *nielle* et non *niele* qui figure à l'entrée ***nëele*** dans le *TL*), mais pas ***nieule*** sf (c'en est pourtant une variante); quand il s'agit de deux homonymes avec même indice grammatical mais deux entrées différentes dans le *TL* (par ex., ***nieule***1 «nuage», ***nieule***2 «gâteau»), la distinction est oblitérée quand la forme apparaît à droite: *niule* sf à gauche renvoie uniquement à ***nieule*** à droite. En revanche, quand la forme est dans la colonne de gauche, la distinction s'effectue bien; ainsi, quand on tape *noël*, on obtient logiquement le résultat suivant:

Entrée	**Catégorie**	**Forme primaire**	**Variante(s)**
noël	sm		(Noel, Nael)
nöel	sm	***neel**	
nöel	sm	***noiel**	
nöel	adj.	***novel**	

Ce sont là de petits défauts qui devraient être corrigés quand la base fera l'objet d'une mise à jour; en attendant, cependant, il convient d'observer une certaine prudence si l'on travaille sur ces «variantes».

Index lemmatisés, dictionnaires de formes et de lemmes, voilà des outils bien commodes pour l'élaboration progressive d'une banque lexicale qui serait un répertoire, avec indice grammatical et références aux œuvres, des mots que présentent les textes d'ancien français. Rêve naguère chimérique, le projet est réalisable de nos jours avec l'appui de l'informatique, si l'on est prêt à y consacrer les moyens et le temps nécessaires. C'est en somme le but que se propose l'équipe du LFA, qui, à la sugges-

tion de G. Roques et dans le cadre de son accord de collaboration avec l'INaLF, a entrepris d'organiser et de construire peu à peu une *Base lemmatisée d'AF*, fichier (FileMaker) des formes graphiques occurrentes regroupées par vedettes pour faciliter la recherche sur le lexique (notamment sur l'évolution du vocabulaire par genre et par région), la graphématique et la morphologie du français des 12e et 13e siècles. Pour ce faire, le LFA a adopté une politique de lemmatisation des principaux textes d'AF, à commencer par ceux du 12e siècle. Dans une première phase, ont été retenus, à cet effet, une quinzaine de textes de genres littéraires différents. Dans la base, chaque lemme est accompagné de ses graphies, d'un indice grammatical et de la référence au texte; la base est mise en relation avec l'ensemble de sous-bases que constituent les index lemmatisés d'œuvres particulières, où l'on peut trouver des informations plus détaillées. Ainsi, pour le fichier du *Couronnement de Louis*:

Lemme	abaissier
Indice grammatical	v
Formes et nombre d'occurrences	abessa (1) 1109 abesse (1) 954 abessier (3) 81, 587, 1294
Source	Couronnement de Louis

La *Base lemmatisée d'AF*, "dictionnaire" sans définition, sera placée sur le Web en accès libre.

3.4. *Lexiques et dictionnaires*

Nous entrons ici dans le domaine des listes de mots à définition ajoutée! On connaît les activités de l'INaLF sur le moyen français. La préparation du *DMF* a entraîné la confection d'une série de lexiques préalables (lexique d'une œuvre, lexique d'auteur, lexique de genre). On peut en voir la liste aux pages V et VI du premier fascicule. Quatre de ces lexiques ont été publiés chez Klincksieck, dans la collection «Matériaux pour le Dictionnaire du Moyen Français»:

– Roger Dubuis, *Lexique des Cent nouvelles nouvelles*, 1996.

– Danièle Jacquart, Claude Thomasset (avec la collaboration de S. Bazin-Tacchella, J.-P. Boudet, Th. Charmasson, J. Ducos, H. L'Huillier), *Lexique de la langue scientifique (Astrologie, Mathématiques, Médecine...)*, 1997.

– Pierre Kunstmann, *Lexique des Miracles Nostre Dame par personnages*, 1996.

– Denis Lalande, *Lexique de chroniqueurs français (XIVe s., début du XVe s.)*, 1995.

Les responsables de l'INaLF ont eu l'excellente idée de les placer également sur leur site Web en accès libre(30). L'utilisateur a le choix entre la «consultation d'une entrée particulière» et la «consultation d'un lexique». Dans le premier cas, on l'invite à choisir une tranche alphabétique, puis à sélectionner une entrée et à valider son choix; s'il retient par ex. la lettre «c» et choisit l'entrée *corps*, il obtiendra 4 articles, classés dans l'ordre alphabétique des noms de leurs auteurs. Dans le deuxième cas, il lui faudra cliquer sur le lexique qu'il veut consulter, puis choisir une tranche alphabétique, ce qui fera apparaître, pour le lexique sélectionné, tous les articles sur les mots commençant par la lettre choisie; il pourra les faire défiler et parcourir cette section du lexique en continu. Quant aux lexiques en cours de rédaction, l'accès en est réservé aux rédacteurs du *DMF*.

Pour le MF, signalons aussi le lexique correspondant à la nouvelle édition critique des *Miracles de Nostre Dame par personnages* en cours de parution sur le site du LFA. Ce lexique n'est pas une simple reprise de celui de 1996; il est construit sur des principes nouveaux. Toutes les occurrences et toutes les acceptions des mots, lexicaux et grammaticaux, y sont indiquées pour le premier miracle; les références comportent deux parties: l'une en chiffres romains pour indiquer le numéro du miracle dans le recueil; l'autre en chiffres arabes renvoie au vers.

L'AF est pour l'instant moins représenté sur la Toile. Est annoncé cependant (LFA) un lexique du *Chevalier au Lion*, exhaustif puisqu'y seront examinés tous les mots et toutes les acceptions de tous les manuscrits. Si le travail mené à Poitiers aboutit aussi à la confection d'un lexique de ce genre, on peut espérer parvenir un jour à remplacer le *Wörterbuch zu Kristian von Troyes' Sämtlichen Werken* de W. Foerster par un ouvrage qui tienne compte systématiquement de toute la tradition manuscrite et qui réponde aux normes actuelles de la lexicographie. Il serait hautement souhaitable de présenter ces données dans un fichier FileMaker ou Access, ce qui permettrait alors de fructueuses recherches de type onomasiologique, sémasiologique et thématique.

Abordant maintenant les dictionnaires, nous devons distinguer ceux de l'ancienne langue de ceux qui portent sur la langue ancienne. Pour les premiers, il est un site dont la fréquentation s'impose: c'est celui de Brian Merrilees(31) et de son projet REFLEX (**R**esearch in **E**arly **F**rench

(30) www.ciril.fr/~hamon/paghtm/wwwlex.html

(31) Sur la tradition de la lexicographie médiévale et l'article dictionnairique, on peut lire, avec profit, du même auteur, «The Shape of the Medieval Dictionary Entry» à l'adresse:
www.chass.utoronto.ca/epc/chwp/merrily2/merr_res.htm#res

Lexicography)[32]. Outre l'annonce des éditions (chez Brépols) du *Catholicon* des manuscrits de Montpellier (Fac. de Médecine, H110) et de Stockholm (Bibl. Klung. N78) [dict. latin-français, fin XIV^e^ s.], du *Glossarium Gallico-latinum* (Paris BNF lat. 7684) [dict. français-latin du XV^e^ s.] et du *Vocabularius familiaris et compendiosus* (Rouen, c. 1490) [dict. latin-français], on y trouve une base, très bien présentée, sur l'*Aalma* [dict. latin-français du XIV^e^ s.]; elle est établie d'après le manuscrit BNF. Lat. 13032 (XIV^e^ s.) et l'édition de Mario Roques (*Recueil général des lexiques latins-français du moyen âge*, t. 2, Paris, Champion, 1938), et contient tous les termes latins et français du manuscrit. Un moteur de recherche permet d'interroger le texte de quatre façons: par mot entier, par série de caractères (minimum de 3 lettres) au début ou à la fin d'un mot, ou bien par simple série de caractères. Si l'on s'intéresse, par ex., à la série *-aille* en fin de mot, on obtient 70 réponses, dont nous donnons les 5 dernières:

66: Tenuis et hoc tenue - *tenve, subtil, graille* - o

67: Testa, teste - *ecruche, esquaille ou truche* - f

68: Texera .xere - *vaissel ou mesure ou dez ou signe de bataille ou d'ostelerie ou jeu de tables ou quarreau de pavement ou maniere de blé* - f

69: Trituro .ras .ravi .rare - *batre, attraire le grain de la paille* - a

70: Vectura .ture - *porture, veiture* - f - *ou le pris que l'en baille pour porter*

Pour les dictionnaires portant sur le français médiéval, l'unique ouvrage à mentionner (mais il est de taille!) est, bien sûr, le *DMF*. Comme l'indique son maître d'œuvre, "le *DMF* est d'emblée *un dictionnaire informatisé*" (fasc. 1, p. IX). Il sera donc consultable pratiquement sous tous les angles. Reste à savoir si la version électronique, dont il ne nous appartient pas de faire ici la présentation, sera diffusée sur CD-ROM ou disponible sur le Web. Notons que le site de l'INaLF annonce que la rédaction de la lettre «D» est en cours.

En ce qui concerne l'AF, une équipe de chercheurs sous la responsabilité de P. Blumenthal à l'Université de Cologne mûrit un projet de numérisation du *TL* et peut-être du *Gdf*; il ne s'agirait pas de réviser ces dictionnaires, mais de les mettre facilement à la disposition du public, probablement sous forme d'un CD-ROM[33]. Remarquons en passant qu'il

(32) www.chass.utoronto.ca/epc/chwp/merrily2/merr_res.htm#res

(33) Pour l'instant, les amateurs ne peuvent s'offrir que le *Dictionnaire historique de l'ancien français* de La Curne de Sainte-Palaye (texte intégral de l'édition Favre de 1876), sur le CD-ROM *L'atelier historique de la langue française* édité par

serait temps, pour les textes de cette période, de songer à donner une suite à ces monuments de la lexicographie française. On pourrait justement, à cet effet, rassembler sur le Web et classer différents types de matériaux, constituer progressivement une sorte de "glossaire des glossaires" comme pour le *DMF*, mais avec définition des termes; pour ce faire, il faudrait commencer à partir des éditions récentes, en examiner soigneusement les glossaires et les diverses critiques dans les revues savantes, et remonter jusqu'à la date d'achèvement des travaux du *Gdf* et des différents fascicules du *TL*. Les lexiques complets et la *Base lemmatisée d'AF*, une fois développée, constitueraient des aides précieuses. Une telle entreprise devrait être multipolaire et l'internet en serait le moyen de communication tout indiqué.

*
* *

Perspectives intéressantes donc, prometteuses probablement; mais cela ne va pas sans risques. En effet, les matériaux qui permettent de stocker les données numérisées sont dégradables; il semble qu'une disquette ou un CD-ROM devienne inutilisable au bout d'une douzaine d'années. La technologie, appareils et logiciels, change rapidement – ce qui peut rendre impossible la lecture de documents enregistrés avec des modèles anciens, à moins d'assurer un transcodage régulier et systématique, processus difficile et coûteux. Aussi vaudrait-il mieux, sur les sites consacrés à la recherche, résister à la tentation de l'image envahissante, des effets visuels, viser plutôt la clarté de la présentation, la lisibilité des textes. Un texte conservé en caractères d'usage courant est facilement téléchargeable, utilisable dans le cadre d'autres programmes et toujours transcodable; la simplicité garantit la liberté et la flexibilité.

Il conviendrait de prévoir plusieurs lieux (au sens propre) de conservation des mêmes données; celles-ci peuvent circuler à une vitesse fulgurante, mais elles peuvent s'effacer (accident ou malveillance) aussi rapidement. Les bibliothèques universitaires auront certainement là un rôle à jouer.

Il faudrait enfin que la qualité et la fiabilité des documents soient assurées par l'observation des mêmes normes et critères que pour les tra-

---

«REDON» (voir: www.dictionnaires-france.com); un vrai régal d'ailleurs puisque pour une même vedette on peut passer, par un simple clic de la souris, de l'article de ce dictionnaire à ceux de 5 autres, dont le *Furetière* et le *Littré*.

vaux imprimés. D'où la nécessité d'instances critiques pour l'évaluation des sites et de leur contenu. Certains «travaux en cours» n'aboutissent jamais, des sites «en construction» peuvent rester éternellement à l'état de projets, d'autres bien établis peuvent disparaître. Menace de l'éphémère; le virtuel ne laisse pas de trace. Il importerait que les revues savantes consacrent une rubrique régulière (quelques pages par numéro) à la critique des parutions sur le Web. C'est à ces conditions seulement que la *translatio* d'un support à l'autre pourra s'effectuer sans dommage, voire même avec un gain appréciable.

Ottawa. Pierre KUNSTMANN