

PRÉSENTATION

par Robert Martin

Le *Dictionnaire du Moyen Français (DMF)*, qui s'élabore à Nancy, au Laboratoire ATILF ("Analyse et Traitement Informatique de la Langue Française", CNRS - Université de Lorraine), est un dictionnaire électronique librement accessible par Internet (www.atilf.fr/dmf). À l'occasion de la présente version, la sixième (DMF 2015), on rappelle ici :

1. les finalités qui orientent l'ouvrage ;
2. les options lexicographiques qui ont été prises.

1. Les finalités

1.1. Dans le sillage du TLF

L'ambition du DMF est de doter le moyen français d'un instrument comparable au TLF : dans l'idéal, il serait pour le français des XIV^e et XV^e siècles ce qu'est le TLF pour le français moderne. On sait que le TLF, dans sa conception initiale, devait, à l'image de l'*Oxford English Dictionary*, couvrir toute l'histoire de la langue, des plus anciens textes jusqu'au XX^e s. Il a vite fallu se rendre à l'évidence : le projet n'était pas réalisable dans les délais impartis. Dès lors, Paul Imbs, le fondateur et le premier directeur du TLF, a souhaité que d'autres ouvrages prennent le relais pour les périodes antérieures : le DMF réalise ce vœu pour le moyen français. Au reste, Paul Imbs a très chaleureusement soutenu le projet, notamment en 1980, lors du 3e Colloque international sur le moyen français, où l'idée a été évoquée pour la première fois (*Actes*, pp. 313-314).

La méthode lexicographique du DMF s'inspire largement de celle du TLF comme en témoignera plus bas l'exposé qui sera fait des "Options". Une des finalités essentielles du DMF est de mettre le dictionnaire de l'ancienne langue au niveau des dictionnaires modernes, tout particulièrement dans le traitement des mots sémantiquement complexes.

Ajoutons qu'une autre finalité rendait le projet très souhaitable : le Godefroy, qui fait une place importante à cette période, date, comme on sait, de la fin du XIX^e siècle. Depuis, la documentation s'est beaucoup enrichie, notamment grâce à de très nombreuses et souvent remarquables éditions critiques. Cette documentation élargie et plus sûre permet de donner du moyen français une image renouvelée. Naturellement, le Godefroy reste irremplaçable, ne serait-ce qu'en raison des innombrables documents d'archives qui s'y trouvent consignés : le DMF y renvoie constamment.

1.2. Un dictionnaire électronique

Depuis les origines, l'informatique se place au cœur du projet. Certes, initialement, l'intention était de doubler une version sur papier d'une autre, électronique. Le

développement spectaculaire de l'informatique et surtout de l'Internet a finalement fait opter, en 2001, pour une solution moins coûteuse et tout compte fait plus efficace : celle d'un dictionnaire exclusivement électronique.

Le DMF est un dictionnaire électronique à un double titre : sa consultation est électronique ; son élaboration s'effectue directement à l'écran.

- La *consultation électronique* présente des avantages désormais bien connus : à preuve le TLF informatisé. Certes, elle ne fait pas oublier l'attrait de la page imprimée, encore moins le plaisir de feuilleter le livre. L'artifice de l'écran, sa relative exigüité, les aléas qui demeurent de la connexion, tout cela peut faire naître la nostalgie de la lecture traditionnelle. Cela dit, un dictionnaire ne se lit pas, mais se consulte sur tel ou tel point particulier. Ce sont les ouvrages de consultation qui souffrent le moins des faiblesses de l'écran. En revanche, ils bénéficient de manière spectaculaire de l'accès électronique, délié de la linéarité et ouvert à l'hypertextualité, c'est-à-dire à la navigation à l'intérieur et en dehors de l'ouvrage.

Naturellement, la possibilité demeure de confectionner à tout moment, à partir du dictionnaire électronique, un dictionnaire sur papier. Un programme de conversion existe qui affiche immédiatement une forme imprimable avec toutes les exigences de l'imprimé traditionnel (colonnes, retraits, italiques, changements de corps...). En raison d'une certaine fragilité des données numériques, on a décidé de réaliser deux exemplaires imprimés du DMF 2015. L'un est en dépôt à Nancy. L'autre est abrité par la Bibliothèque de l'Institut de France : il comporte 28 volumes reliés d'environ 600 pages, soit près de 17.000 pages au total.

- Mais l'informatique pénètre le DMF d'une autre manière encore : elle se place au cœur de son *élaboration*. Désormais, le rédacteur construit son article à l'écran. La rédaction du DMF s'opère selon un "masque de saisie" ; par ailleurs le rédacteur dispose d'un "correcteur lexicographique". Tous les instruments nécessaires à la rédaction informatisée ont été développés par Gilles Souvay, Ingénieur de recherche à l'ATILF, responsable de l'élaboration informatique du projet.

Impossible de donner ici une idée précise du "masque de saisie". Disons que c'est un système qui, au fil de la rédaction, indique le type d'information qu'il convient de fournir ; il affiche au fur et à mesure des balises à remplir ou les choix à faire parmi les balises possibles à l'endroit où l'on est arrivé. Ainsi la première balise, obligatoire, est celle de la vedette : le curseur indique où il convient de l'écrire ; la forme s'enregistre automatiquement en gras ; vient ensuite un premier choix : on peut se borner à un renvoi (qui doit être précisé à l'endroit où le curseur s'est placé) ou bien on choisit de poursuivre, auquel cas, un exposant est possible (en cas d'homonymie), puis c'est le "code grammatical" qu'il convient d'indiquer ; le système l'écrit en bas-de-casse romain ; vient ensuite la balise "Dictionnaires" ; le système affiche les dictionnaires possibles ; le choix de l'un d'entre eux entraîne l'affichage du mode de référence que le DMF exige, par exemple le lemme pour le GD (Godefroy) ; le curseur demande à nouveau que le lemme soit précisé ; il l'écrit automatiquement en bas-de-casse italique ; et ainsi de proche en proche pour toutes les autres balises, jusqu'à la signature de l'article. Au fil du texte, les exemples sont directement transférables à partir des Bases textuelles, où l'occurrence demandée est automatiquement balisée, de même que la référence (elle-même balisée par auteur, par titre et par date dans la Bibliographie). Certains exemples

(surtout en provenance d'éditions récentes) doivent cependant être saisis par le rédacteur, qui enrichit ainsi quotidiennement les fonds du DMF de données supplémentaires.

Mais l'innovation la plus remarquable est la possibilité qui est offerte au rédacteur de soumettre son texte à un "correcteur lexicographique". Celui-ci, dans un temps extrêmement bref (de cinq à dix secondes selon la taille du fichier), signale quatre types d'erreurs possibles.

1. Erreurs de balisage (par exemple des valeurs incorrectes : sous la balise DOM[aine], on ne peut utiliser que des domaines reconnus, de même sous la balise CODE GR[ammatical] ; certaines balises ont pu rester incomplètement remplies : le système le décèle sans faille).

2. Erreurs matérielles : typographiques (espaces multiples ; absence d'espace après un signe de ponctuation double comme le point-virgule ; absence d'espace après la balise OCC[urrence] ; guillemets, parenthèses ou crochets qu'on a oublié de refermer ou qui ne sont pas ouverts...) ; ordre alphabétique défectueux des entrées ; ordre chronologique défectueux des exemples dans un même paragraphe ; numérotation incohérente...

3. Erreurs de référence : références bibliographiques inconnues (pour les exemples que le rédacteur saisit lui-même) ; présentation défectueuse d'un incunable (où il n'est pas possible de se contenter d'un numéro de page) ; mauvaise forme de l'étymon (le système contient une liste complète, avec leurs références, des étymons du FEW) ; mauvaise référence de l'étymon (qui par exemple se trouve dans un autre volume que celui que le rédacteur a erronément indiqué) ; mauvaise graphie du lemme de Tobler-Lommatzsch (là aussi le système contient toutes les graphies de la nomenclature).

4. Erreurs de renvoi (renvoi à un lemme qui n'existe pas, à une subdivision incohérente...).

Tout cela facilite considérablement le travail du lexicographe et lui évite de menues erreurs souvent très difficiles à repérer. Une fois que l'habitude de l'assistance lexicographique est prise, il devient décidément impossible de s'en passer.

1.3. Une lexicographie évolutive

Le DMF s'élabore par une suite d'étapes dont chacune possède une cohérence suffisante pour constituer en soi un ouvrage clos, mais conçu de telle sorte qu'il reste ouvert aux extensions et aux améliorations. L'idée centrale qui guide le projet du DMF est en effet que l'informatique autorise désormais une *lexicographie évolutive* : il ne s'agit plus de rédiger le dictionnaire lettre par lettre, ce qui le laisserait dans l'inachèvement aussi longtemps que la lettre ultime n'est pas atteinte, mais plutôt de procéder par une suite d'étapes dont chacune possède sa propre clôture tout en restant disponible à tous les développements ultérieurs. La facilité avec laquelle les outils informatiques permettent d'augmenter, de corriger, de restructurer les données ne peut rester sans incidence sur la technique lexicographique. L'option choisie pour le DMF s'appuie fortement sur l'idée que les dictionnaires d'aujourd'hui, non pas commerciaux mais scientifiques, ne

devraient plus être des produits figés que seules peuvent modifier d'hypothétiques rééditions, inévitablement coûteuses et elles-mêmes figées pour longtemps, mais au contraire des bases informatisées, faciles d'accès et ouvertes à peu de frais à tous les enrichissements et à toutes les améliorations que l'on peut estimer souhaitables.

Le DMF en est donc actuellement à sa sixième version (DMF 2015) ; cette version a été précédée par le DMF1 (2002), le DMF2 (2007), le DMF2009, le DMF2010 et le DMF2012.

- Le **DMF1** (2002, sur Internet en 2003) rassemble sous des lemmes communs le contenu de 13 "Lexiques préalables", élaborés par des collaborateurs de l'ATILF (précédemment de l'INaLF) et par des collaborateurs universitaires qui ont bien voulu s'associer au projet. À ce stade, le DMF comporte une nomenclature de 26500 entrées, la matière des lexiques étant regroupée, mais pas encore synthétisée.

Certains de ces Lexiques avaient été précédemment publiés sur papier (ils sont en dépôt à la Librairie Champion) :

Denis Lalande (Univ. de Bordeaux 3), *Lexique de chroniqueurs français (XIVe s., début du XVe s.)*, 1995, 551 p.

Roger Dubuis (Univ. de Lyon 2), *Lexique des Cent nouvelles nouvelles*, 1996, 521 p.

Pierre Kunstmann (Univ. d'Ottawa), *Lexique des Miracles Notre Dame par personnages*, 1996, 528 p.

Danièle Jacquart (EPHE), Claude Thomasset (Univ. de Paris-Sorbonne), avec la collab. de Sylvie Bazin-Tacchella, Jean-Patrice Boudet, Thérèse Charmasson, Joëlle Ducos et Hervé L'Huillier, *Lexique de la langue scientifique (Astrologie, Mathématiques, Médecine...)*, 1997, 313 p.

Joël Blanchard (Univ. du Maine), Michel Quereuil (Univ. de Clermont 2), *Lexique de Christine de Pizan*, 1999, 402p.

D'autres étaient inédits :

Annie Bertin (Univ. de Paris 10), *Lexique d'Andrieu de la Vigne*

Charles Brucker (Univ. de Lorraine), *Lexique de Nicole Oresme*

Pierre Demarolle (Univ. de Lorraine), *Lexique d'Antoine de la Sale*

Hiltrud Gerner (ATILF), *Lexique de la littérature didactique* (Simon de Phares, Philippe de Mézières, Evrart de Conty, *Songe du vergier...*)

Martine Moulin (ATILF), *Lexique de Georges Chastellain, Fragments du Livre IV*

Noël Musso (ATILF), *Lexique de Guillaume de Machaut*

Edmonde Papin (ATILF), *Lexique d'archives : chartes, comptes, inventaires...*

Jean-Loup Ringenbach (ATILF), *Lexique de la Passion d'Auvergne*.

- Le **DMF2** (2007), ouvert à l'occasion du 25^e Congrès international de Linguistique et Philologie romanes qui s'est tenu à Innsbruck en Autriche, contient dans son entier la Base lexicale du DMF1, mais celle-ci est augmentée de trois lexiques nouveaux :

Jocelyne Bernardoff et Jean-Loup Ringenbach (ATILF), *Lexique de Jean d'Arras, Mélusine*

Jean-Loup Ringenbach (ATILF), *Lexique des Mystères*

Bernadette Suty (ATILF), *Lexique du Registre criminel du Châtelet* (lexique commencé par Monique Haas).

Le DMF2 traite par ailleurs plus de 30000 vocables qui ne sont pas déjà représentés dans le DMF1 : ils ont été inventoriés dans un document dénommé "Lexique complémentaire". La nomenclature comporte donc à ce stade plus de 60000 entrées. Les 30000 vocables supplémentaires sont des mots relativement rares puisqu'ils ne sont pas représentés dans les 13 lexiques du DMF1 : ils donnent lieu à des articles intégralement rédigés. En revanche la matière des lexiques, certes regroupée comme précédemment sous des lemmes communs, n'est toujours pas synthétisée à ce stade.

- Le travail de synthèse débute dans le **DMF 2009**. Les synthèses y sont de deux sortes: les unes (indicatif : "Synthèse") prennent en compte non seulement les données des lexiques, mais aussi l'entier de la documentation (électronique ou sur papier) dont le DMF dispose ; en d'autres termes, elles sont aussi complètes que possible ; dans le DMF 2009 ce type s'applique aux mots les plus fréquents (environ 1500, à travers tout l'alphabet). Les autres (indicatif : "Synthèse des Lexiques") ne réunissent que les lexiques et la documentation électronique (mais pas la documentation sur papier) ; elles doivent donc être complétées par la suite ; dans le DMF 2009, ce type est appliqué aux lettres A, B, C, H, I, J, K, L et P. Quant aux mots qui n'apparaissent que dans un seul Lexique, ils y restent provisoirement dans l'état du DMF2.

Le DMF 2009 comporte par ailleurs quatre lexiques nouveaux :

Lexique de Froissart, Chroniques par Jacqueline Picoche (Univ. d'Amiens)

Lexique de Griseldis par Jocelyne Bernardoff (ex-INaLF)

Lexique sélectif de Juvenal des Ursins par Martine Moulin (ATILF)

Lexique des Realia romana par Frédéric Duval (Univ. de Lorraine).

- Le **DMF 2010** poursuit le travail de synthèse. Quelque 500 mots supplémentaires sont intégralement synthétisés (c'est-à-dire à partir de la documentation électronique et de la documentation sur papier ; indicatif : "Synthèse"), soit désormais 2000 mots de grande fréquence. Le reste synthétise les lexiques et la documentation électronique (indicatif : "Synthèse des Lexiques") pour les lettres qui n'ont pas été traitées en 2009 (D, E, F, G, M, N, O, R, S, T, U, V, X, Y et Z).

Par ailleurs, le DMF 2010 ajoute le lexique suivant :

Lexique des Pèlerinages de Guillaume de Digulleville par Béatrice Stumpf (ATILF).

- Le **DMF 2012** enrichit les articles qui, dans le DMF1, ne figuraient que dans un seul Lexique : les mots en cause (environ 10 000) n'ont pas été complétés dans les versions qui ont suivi ; chemin faisant, on a pris en compte les éditions critiques parues depuis 2000, ce qui a permis d'augmenter la nomenclature de quelques dizaines de vocables. Par ailleurs, les vocables signalés dans les versions 2009 et 2010 par la mention "Synthèse des Lexiques" ont également été retravaillés. Le DMF 2012 reprend, non pas l'entier du fonds sur papier, mais pour le moins ce qui a paru le plus important. Par la suite, ce fonds sera plus encore exploité, mais les versions ultérieures du DMF n'ont plus à reprendre systématiquement des tranches de nomenclature et des types d'articles : le DMF 2012 mettant tous les vocables à égalité, les interventions désormais seront ponctuelles, même si elles doivent être en très grand nombre.

Ajoutons que la version DMF 2012 comporte une rubrique nouvelle intitulée "Compléments au DMF 2012" : elle permet d'enregistrer des données qui trouveront leur place dans les versions futures. Ainsi les "Compléments" ajoutent des mots nouvellement rédigés, des mots à rédiger pour lesquels on fournit un exemple, des ajouts divers à des articles existants, des versions provisoires de lexiques en cours (p. ex. un Lexique des Préfixes et des Suffixes et un Lexique des mots grammaticaux). Par ailleurs, les corrections, voire les ajouts, à apporter au DMF sont consignés au fur et à mesure en "Annotation" sous les articles concernés.

- Le **DMF 2015** est une version corrigée, augmentée et enrichie de nombreux liens hypertextuels. Elle tient compte de toutes les "Annotations" qui ont été ajoutées au DMF 2012 ; elle accroît la nomenclature de près de 3000 entrées, fait une place aux mots grammaticaux, aux préfixes et aux suffixes, et propose une version revue de plusieurs milliers d'articles ; elle multiplie les liens hypertextuels qui permettent d'ouvrir, outre le Godefroy, le TLF et le FEW, tous les dictionnaires de l'ancienne langue accessibles sur la "toile" (AND2, DEAF, DECT). Le DMF 2015 ajoute aussi un Lexique nouveau, le *Lexique des Proverbes* rédigé par Pierre Cromer.

On observera que toutes les versions du DMF restent indéfiniment accessibles. Par défaut, c'est toujours la version la plus récente qui s'affiche ; mais il y suffit d'un clic pour ouvrir une version antérieure. C'est là un choix de grande importance. Sur l'Internet en effet, la tendance existe d'emprunts plus ou moins sauvages, sans renvoi explicite au document consulté, considéré comme plus ou moins anonyme. Alors qu'une publication sur papier donne lieu au dépôt légal et que l'on s'y réfère toujours à une édition précise, le risque se crée sur la "toile" qu'aucune date éditoriale ne s'attache plus à l'objet. Pour pallier ces inconvénients d'anonymat et d'instabilité, il faut veiller à garder disponibles, datées et signées, des versions successives correctement identifiables, et surtout ne pas livrer une production incessamment remodelée et qui, à la limite, varierait de jour en jour. Dénommées sans ambiguïté, les étapes doivent être stables comme le sont les éditions d'un livre. Les erreurs sont corrigées de version à version ; elles demeurent dans la version d'origine. En somme, l'ouvrage électronique s'améliore peu à peu ; mais c'est à la manière du livre et non comme une base en constante évolution. Cependant le système d' "Annotations" et de "Compléments" régulièrement augmentés permet de présenter avant même l'édition à venir les améliorations que celle-ci intégrera.

Il serait extrêmement souhaitable que d'autres Lexiques encore soient élaborés, surtout dans les domaines techniques, et que leur matière soit ensuite intégrée au Dictionnaire. Si le DMF reste ouvert aux connaissances qu'apporteront les travaux futurs sur le moyen français, il pourra indéfiniment s'améliorer, jusqu'au jour où la technique lexicographique qu'il met en œuvre sera à son tour dépassée.

1.4. Une lexicographie hypertextuelle

Informatisée et évolutive, la lexicographie que le DMF illustre est naturellement de nature hypertextuelle. Le Dictionnaire résulte d'une synthèse enrichie des Lexiques ; les Lexiques eux-mêmes sont construits à partir d'une Base textuelle, au reste indéfiniment extensible.

Ainsi se crée un système d'interrelations à trois niveaux :

- le Dictionnaire proprement dit
- les Lexiques
- les Bases textuelles.

La navigation entre ces trois niveaux peut être à bien des égards bénéfique. Ainsi la relation du Dictionnaire et des Lexiques : contrairement au Dictionnaire, les Lexiques procurent une vision particulière, par exemple en éclairant un texte, l'œuvre d'un auteur, un domaine. Par ailleurs, les Lexiques en attente d'intégration sont immédiatement accessibles.

Tout aussi importante est la relation du Dictionnaire avec les bases qui ont servi à le construire. Ainsi le consultant peut à tout moment, s'il juge l'article insuffisant sur le point qui l'intéresse, accéder à la documentation du DMF. Pour l'heure, seuls sont mis à sa disposition les textes intégralement saisis, et non pas le fonds des exemples relevés sélectivement par la lecture de milliers d'autres textes ; cette documentation, particulièrement riche, pourrait être elle aussi ouverte. Supposons que nous nous intéressions au syntagme *haute justice* ; le DMF fournit sous *justice* un certain nombre d'exemples ; mais le recours à la Base textuelle permet d'enrichir ces données d'une vingtaine d'exemples supplémentaires. Ainsi, les éventuelles faiblesses du Dictionnaire sont au moins partiellement palliées. Seule l'informatique procure cette possibilité remarquable : d'une certaine manière, elle associe le consultant à l'interprétation des données.

Au reste, un autre niveau encore s'est peu à peu instauré : celui d'outils, reliés au DMF, susceptibles de faciliter la construction de glossaires ou l'édition électronique de textes.

2. Les options lexicographiques

2.1. Les bornes chronologiques

Il a fallu tout d'abord fixer les limites chronologiques.

Le *terminus ad quem* s'est imposé pour des raisons pratiques : l'existence du Huguet conduit en effet à choisir, comme date ultime, les toutes dernières années du XV^e siècle (celles des guerres d'Italie). L'argument pratique emporte d'autant plus la conviction que la fin du XV^e siècle marque un incontestable clivage. Ce n'est qu'au XVI^e siècle que, par le contact direct avec le grec ancien, les hellénismes vont se répandre ; que l'italien, et, dans une moindre mesure, l'espagnol vont exercer leur influence ; que s'instaure une véritable réflexion grammaticale sur la langue vulgaire (avec des grammairiens comme Meigret, Palsgrave, Peletier du Mans ou encore Henri Estienne) ; que le développement de l'imprimerie jouera pleinement son rôle fixateur.

Plus délicat paraît la question du *terminus a quo*. Les opinions divergent très largement à ce propos. Il a semblé au total que les alentours de 1330 offraient le meilleur choix. L'avènement des Valois et la Guerre de Cent Ans qui en résulte marquent le début d'une période troublée : époque d'instabilité économique et de misère, encore aggravée par la Peste Noire (1347), sans parler des Grandes Compagnies qui dévastent le pays. La révolte avortée d'Etienne Marcel (1357) et la Jacquerie (1358) sont le signe

de bouleversements dont il serait étonnant qu'ils soient sans répercussion sur la langue. Le fait est qu'on assiste, au début du XIV^e siècle, à des phénomènes linguistiques de très grande portée : effondrement définitif de la déclinaison ; réduction analogique des radicaux verbaux ; substitution d'un ordre des mots fonctionnel à l'ancien ordre à fondement rythmique ; expression quasi-obligatoire du pronom sujet ; accord du participe passé quand le complément d'objet le précède ; véritable invasion de termes abstraits d'origine latine... Au reste, des données pratiques plaident aussi en faveur de cette date : le Tobler-Lommatzsch déborde légèrement sur le XIV^e siècle, mais fort peu ; on n'y trouve guère de textes de la fin du XIV^e siècle, et quasiment aucun du XV^e.

2.2. Les sources textuelles.

Les textes du DMF appartiennent donc à la période qui va de 1330 à 1500. Les écarts sont en très petit nombre : quand l'essentiel de l'œuvre d'un auteur entre dans les limites chronologiques que l'on s'est données, on a tout de même admis, pour la complétude, les textes légèrement postérieurs ; c'est le cas pour Jean Molinet ou pour Andrieu de la Vigne. Même chose pour les genres littéraires : certaines farces, dont la datation est imprécise, peuvent appartenir au début du XVI^e siècle. Il va sans dire que le corpus constitué est loin d'être exhaustif : c'est proprement impossible. Mais on a cherché à constituer une sélection suffisamment vaste pour être représentative de la langue de l'époque dans ses différents aspects. Si une place particulière est faite aux grands auteurs, tous les genres et tous les types de textes ont été pris en compte : la littérature narrative, les chroniques, les récits de voyage ; la poésie ; le théâtre, religieux et profane ; les traités, philosophiques, scientifiques, religieux ; les sermons ; mais aussi des chartes, des comptes, des inventaires, les ordonnances ou les lettres royales. À quelques exceptions près, il s'agit toujours de textes imprimés – de préférence dans des éditions critiques – et non pas de manuscrits. L'exploitation systématique des comptes rendus auxquels les éditions donnent lieu permet de tenir compte des suggestions lexicologiques qui y sont faites.

Une Bibliographie détaillée, établie par Willy Stumpf et complétée pour les localisations régionales par Yan Greub et Pascale Baudinot, fournit la liste des textes exploités dans le DMF. Ils se répartissent en plusieurs classes : les uns appartiennent à une Base textuelle qui les recueille dans leur intégralité ; cette Base fait partie de FRANTEXT ; d'autres entrent dans une Base dite de « partiels » ; pour d'autres enfin, on s'est borné à exploiter le glossaire de l'édition critique, ou bien on les a lus en constituant des dossiers traditionnels sur papier.

1. Base textuelle FRANTEXT.

Des textes dits « intégraux » (220 au total) forment une sous-base de FRANTEXT. La plupart de ces textes ont été saisis au laboratoire. Certains cependant viennent d'autres Équipes qui les ont généreusement communiqués au DMF. Le DMF est particulièrement redevable :

- à l'Institut d'études médiévales de l'Université de Montréal (Serge Lusignan) ;
- à l'Université du Québec à Montréal (Monique Lemieux) ;
- à l'Institut des langues romanes et classiques de l'Université de Jyväskylä (Ulla Jokinen, Ellen Sakari) ;
- au Département de français de l'Université d'Aberdeen (James C. Laidlaw) ;
- à la Faculté des Lettres de l'Université de Lausanne (Eric Hicks) ;
- à l'ancienne Équipe de Linguistique informatique de l'École Normale Supérieure de Fontenay–Saint Cloud (Christiane Marchello-Nizia).

Une mention à part doit être faite du Laboratoire de français ancien de l'Université d'Ottawa ; un fructueux accord de collaboration permet de conduire divers projets en étroite collaboration ; en particulier, l'enregistrement des *Miracles de Notre Dame par personnages* du Lexique de Pierre Kunstmann a été réalisé à Ottawa.

Par ailleurs, plusieurs chercheurs ont mis gracieusement à la disposition du DMF des enregistrements textuels :

- Rae S. Baudouin (FROISSART, *Ballades et Rondeaux*) ;
- Dominique Billy (MESCHINOT, *Les Lunettes des princes*) ;
- Jean-Patrice Boudet (SIMON DE PHARES, *Astrologie*, version utilisée dans la thèse qu'il a soutenue) ;
- Frédérique Hamm (DAUDIN, *De la erudition*) ;
- Helena Häyrynen (FILLASTRE, *Le Traité de Conseil*) ;
- Madeleine Jeay (*Les Evangiles des Quenouilles*) ;
- Hans-Erich Keller (BAGNYON, *Histoire de Charlemagne*) ;
- Carole Lambert (*La maniere de henter soutillement* ; LA VIGNE, *La Ressource de la Chrestienté*) ;
- François Rouy (CHARTIER, *Le Livre de l'Espérance*) ;
- Alain Surdel (FLAMANG, *La Vie et Passion de monseigneur saint Didier*) ;
- Marie-Hélène Tesnière (BERSUIRE I, 1 ; I, 9) ;
- Florence Waskiewicz (LA HAYE, *Poème sur la grande peste de 1348*).

Cette Base continue de s'enrichir. Une dizaine de textes se trouve en attente d'intégration (notamment, grâce à la générosité de George T. Diller, l'édition qu'il a consacrée au ms. d'Amiens de la Chronique de Froissart).

2. Base des « partiels ».

Un grand nombre de textes (environ 500), dits « partiels », ont fait l'objet d'une informatisation sélective. Les extraits retenus constituent une Base particulière ; ils sont interrogeables, au delà des faits qui ont motivé leur sélection, sur l'ensemble des

vocables qu'ils comportent. Cette Base est provisoirement réservée aux Rédacteurs du DMF : elle n'est pas accessible par FRANTEXT.

3. *Documentation sur papier.*

Les éditions critiques dont on a exploité le glossaire ont donné lieu à une table lemmatisée qui rassemble toutes les entrées ; ce document, appelé *Glossaire des glossaires*, a été réalisé grâce au concours d'une quarantaine de médiévistes, français et étrangers ; Yorio Otaka, Professeur à l'Université d'Osaka, a pris dans cette opération une part considérable.

Grâce à Yan Greub (et à Pascale Baudinot), la Bibliographie mentionne désormais l'appartenance régionale des textes (ces données s'affichent article par article par le bouton *Sources*, "Source des exemples").

Par ailleurs, chaque texte est pourvu d'une date, au moins d'une fourchette plus ou moins approchée. La date indiquée se rapporte à la composition du texte :

- 1400 : «texte rédigé en 1400»
- a.1400 : «texte rédigé peu avant 1400»
- p.1400 : «texte rédigé peu après 1400»
- c.1400 : «texte rédigé aux alentours de 1400»
- 1400-1410 : «texte dont la rédaction s'est étendue de 1400 à 1410»
- c.1400-1410 : «texte dont la rédaction peut être située approximativement entre 1400 et 1410 – que la date soit impossible à préciser davantage ou que la rédaction se soit approximativement étendue entre ces deux dates».

Une date entre crochets signale :

- l'existence d'une version antérieure, qui n'est pas celle de l'édition, mais dont on peut supposer qu'elle n'est pas sans laisser de traces (*Percef. I*, T., c.1450 [c.1340]) ;
- l'époque tardive du manuscrit ou de l'imprimé à travers lequel le texte est connu (*All. foi C.*, c.1350-1400 [p.1478]).

Dans les articles du DMF, les exemples sont classés, à l'intérieur de chaque rubrique, par ordre chronologique. Dans le cas d'une fourchette, l'ordre est déterminé par la date initiale. Lorsque plusieurs dates sont mentionnées (*Jug. maître-échev. Metz S.M.S.*, t.1, a.1494 [1363]), c'est la première dans l'ordre séquentiel qui l'emporte (ici : a.1494). Un des objectifs du DMF est évidemment de relever les "premières" et les "dernières" attestations, non seulement de vocables, mais aussi de sens et même de constructions. Cependant, ces attestations ne sont pas distinguées des autres : d'une part certaines ont pu passer à travers les mailles du filet et d'autre part les informations de cette nature se périment très vite, puisqu'il suffit d'étendre le corpus pour qu'elles se modifient.

2.3. Les principes de macrostructure

2.3.1. Principes d'admission

Le DMF est fondamentalement un dictionnaire de corpus. Le principe du corpus suppose que l'on enregistre dans le dictionnaire tous les faits repérés, sans aucune exception. Cela signifie que les hapax, les fantaisies verbales, les créations sans lendemain, voire les mots fantômes que les éditions peuvent comporter entrent dans le dictionnaire, où naturellement ils sont signalés pour ce qu'ils sont. Aucune sélection n'est opérée sur des critères de fréquence. Seuls sont exclus, en principe, les mots «étrangers» (y compris occitans, franco-provençaux, franco-italiens et même les mots proprement anglo-normands pour lesquels on renvoie à l'AND2). Les mots latins ne sont admis que si leur syntaxe est française (*un ave, un vidimus...*).

Il faut ajouter cependant que des faits complémentaires, plus ou moins aléatoirement détectés, trouvent place dans le DMF. Ils viennent du dépouillement de revues, d'études de toutes sortes, du fonds Barbier ou bien de l'IGLF (Inventaire général de la langue française, initié par Mario Roques).

2.3.2. Règles de dégroupement homonymique

En principe, le dégroupement homonymique se fonde sur l'étymologie. Sont dégroupés en entrées multiples tous les vocables de même forme mais d'étymologie distincte. Dans la très grande majorité des cas, ce principe est d'une application relativement simple : sous chaque vocable traité, le DMF renvoie au FEW ; dès lors que le FEW révèle une pluralité d'étymons (non pas d'hypothèses étymologiques, mais de sources étymologiques distinctes), il s'ensuit, dans le DMF, comme dans le Tobler-Lommatzsch, un dégroupement des entrées. Donnent donc lieu à des entrées séparées des vocables comme *biere*¹ (germ. *bier*) et *biere*² (franc. **bëra*), *delit*¹ (*delictum*) et *delit*² (*delectum* de *delectare*), *louer*¹ (*laudare*) et *louer*² (*locare*), *affoler*¹ (sur *fou*) et *affoler*² (sur *fouler* «blesser»), *son*¹ («ce qui frappe l'ouïe», *sonus*), *son*² («résidu de la mouture du grain», étymologie discutée) et *son*³ (*suus*).

Ailleurs on adopte une seule entrée, même s'il n'est pas sûr (mais comment en juger ?) que dans la conscience linguistique des locuteurs de l'époque un lien existait, si ténu soit-il, entre les diverses significations. Ainsi *robe* au sens de «pillage» et *robe* au sens de «vêtement» se trouveront regroupés sous une entrée commune, articulés seulement par des chiffres romains (ceux-ci étant réservés aux distinctions de *sens* et non pas de simple *acception*, c'est-à-dire non reliables par des indicateurs comme *p. ext.*, *p. restr.*, *p. méton.*, *p. anal.* ou *au fig.*). Cependant l'éloignement peut être tel que le dégroupement paraît s'imposer même en moyen français (par exemple *savoir* «avoir le goût de» et *savoir* verbe épistémique ; ou bien *joncher* «couvrir le sol de» et *joncher* «tromper»). Mais ce cas est rarissime.

L'unité morphologique et historique conduisant à une seule entrée, les simples variations grammaticales sont traitées, non pas sous forme d'entrées dégroupées, mais au moyen de distinctions en chiffres romains. Ainsi l'adjectif substantivé masculin trouve place sous l'adjectif ; le participe présent ou passé, adjectivé ou substantivé au masculin, ainsi que l'infinitif substantivé vont normalement sous le verbe. En revanche, les substantifs que distingue un écart formel (*boulangier/boulangere* ; *batailleur/batailleresse...*) sont toujours enregistrés séparément, d'autant plus qu'il s'y ajoute fréquemment des variations sémantiques (*lionne* ne conduit pas aux mêmes emplois métaphoriques que *lion*). Les locutions conjonctives ou prépositives formées sur un substantif sont mentionnées sous le substantif concerné (ainsi à *cause de* sous *cause*).

Ces règles appellent cependant certains traitements particuliers :

— ainsi, quand une opposition grammaticale produit des sens que la base ne connaît pas, le dégroupement peut se justifier ; *abondant*, en raison de la locution *d'abondant* «en outre, de surcroît», qui ne peut se rattacher directement au verbe, bénéficiera d'une entrée séparée ;

— quand l'opposition grammaticale ne vaut que pour un effet de sens particulier, au lieu que soit ouverte une rubrique en chiffres romains, elle est signalée à la seule place où elle est pertinente (ainsi *bon* en emploi substantif correspond au sens de «brave, valeureux» ; c'est donc à cet endroit que la substantivation gagne à être mentionnée).

Les mots composés, dont les limites, comme on sait, sont bien difficiles à tracer, trouvent place sous le composant qui en constitue le noyau syntaxique (*main mise* sous *main* ; *œuvre naturelle* sous *œuvre...*). Cependant, un principe formel d'analyse syntaxique conduira à des entrées distinctes chaque fois que la syntaxe du composé n'est pas celle des éléments qui le composent : dans *un porte-faix* (*déterminant + verbe + substantif), *une alarme* (*déterminant + préposition + déterminant + substantif : **une à l'arme*), *la bienvenue* (*déterminant + adverbe + participe passé), les suites catégorielles ne sont pas normalement des suites acceptables.

2.4. Les données de microstructure et le balisage

L'examen des balises permet de préciser les options lexicographiques choisies pour la microstructure du DMF. Elles peuvent être groupées en plusieurs espèces :

— balises de lemmes, d'entrées et de codes grammaticaux ;

— balises de renvoi aux dictionnaires (selon la présence du vocable dans le Tobler-Lommatzsch, le Godefroy, le DEAF, l'AND2, le DECT le FEW, le TLF) ;

— balises de structuration de l'article (balises de rubriques) ;

— balises métalinguistiques ;

— balises d'exemples.

2.4.1. Balises de lemmes, d'entrées et de codes grammaticaux

Le DMF distingue des "entrées" et des "lemmes". À l'origine, cette opposition est due au fait que dans les "Lexiques préalables" les rédacteurs ont appliqué des règles qui n'étaient pas a priori unifiées mais qui valaient pour le corpus en cause. Le regroupement a donc imposé une lemmatisation.

À présent que le Dictionnaire est intégralement synthétisé, on pourrait juger cette distinction superflue. Elle est maintenue pour deux raisons :

- le lemme, intangible, est une étiquette commode pour passer du Dictionnaire aux Lexiques et inversement ;
- le choix du lemme se fonde sur un principe qui n'est pas exactement celui qui gouverne les entrées : le lemme se rapproche le plus possible des graphies du français moderne ; les entrées reflètent plutôt l'état du moyen français. Ainsi, à côté du lemme *château* figure l'entrée *chastel* ; à côté du lemme *âme*, l'entrée *ame* ; à côté du lemme *connaître*, l'entrée *connoistre*.

Le lemme est associé à son code grammatical. Quand le genre du substantif est incertain, on se limite à <subst.> ; quand il est variable, on mentionne <subst. masc. et fém.>. Aucune information n'est donnée à ce niveau sur la transitivité, l'intransitivité ou l'emploi pronominal des verbes ; ces précisions figurent dans le corps des articles. Dans le cas de mots douteux (de mots qui ont tout l'air d'être des fantômes), le code grammatical est remplacé par la marque (?), ce qui est une manière de les signaler d'emblée.

2.4.2. Balises de renvoi aux dictionnaires

À l'intérieur de la balise <DICT>, le renvoi aux dictionnaires se fait dans l'ordre suivant :

- Tobler-Lommatzsch : le sigle (<T-L>) est suivi du lemme de ce dictionnaire, éventuellement pourvu d'un exposant (ajouté par nos soins, les renvois de T-L entrant dans le décompte) ;
- Godefroy et le Complément de Godefroy : le sigle (<GD> ou <GDC>) est également suivi du lemme, pourvu de l'exposant que Godefroy lui affecte ; les références au tome et à la page (ou à la colonne de T-L) sont mentionnées, si c'est nécessaire, dans le corps de l'article ;
- DEAF : le sigle (<DEAF>) est suivi de la référence (lettre et colonne) et du lemme ; l'AND et le DECT sont suivis du lemme ;
- FEW : le sigle (<FEW>) est suivi de la référence au tome, à la page et à la colonne, ainsi que de l'étymon (<ETYM : ...>) ; cette information est capitale ; elle permet en effet les interrogations à partir de l'étymon ; on attire tout particulièrement l'attention de l'utilisateur sur la possibilité, dans toute interrogation, de «cliquer» sur l'étymon et d'obtenir ainsi l'affichage de tous les vocables de même étymologie ;

c'est assurément une commodité très appréciable ; lorsque le mot ne figure pas dans le FEW, on indique le lieu où il devrait être, le sigle FEW étant alors précédé d'un astérisque (<*FEW>) ; cette mention est suivie d'un point d'interrogation en cas de doute sur le rattachement proposé ; lorsque l'étymon lui-même ne figure pas dans le FEW, c'est mentionné ainsi : <FEW Ø, nom de la langue source : <ETYM>>, par exemple, pour *bigame* :

<FEW Ø, lat. : <ETYM> : *bigamus*> ;

— TLF : le sigle (<TLF>) est suivi de la référence au tome, à la page et à la colonne, ainsi qu'au lemme du TLF, pourvu de l'exposant que le TLF lui affecte ; cette information rend possible l'interrogation par les formes de français moderne, même si le lemme du DMF s'en écarte.

2.4.3. Balises de structuration de l'article

Les balises de structuration font apparaître les ordinations sémantiques du DMF, dont on voudrait exposer au moins sommairement les principes. Indiquons cependant au préalable la hiérarchie pratiquée : la plus élevée est marquée par les chiffres romains ; en dessous figurent les lettres majuscules ; en dessous les chiffres arabes ; en dessous les lettres latines minuscules ; soit **I. – A. – 1. – a**) ; chacune de ces indications figure dans une balise <NUM>. À chaque niveau, sans qu'il en découle aucune distinction hiérarchique, il est possible de subdiviser au moyen d'un tiret (–), et, en dessous du tiret, par un point (.), par exemple pour introduire une nuance sémantique ou un emploi locutionnel qui ne mettent pas en cause la structuration elle-même.

Le fondement de l'analyse est en principe de nature synchronique. Le DMF couvre certes une période de quelque 170 années. Au cours d'une durée aussi longue, des évolutions ont forcément eu lieu : beaucoup de vocables sont tombés en désuétude, d'autres ont vu le jour, parfois éphémères et parfois plus durables ; en 170 ans, la succession des générations et l'évolution des mentalités n'ont pas pu rester sans conséquences sur les usages linguistiques. Cependant, *grosso modo*, la période en cause constitue une unité synchronique – une synchronie large, comparable en étendue à celle du TLF, et dont la stabilité est suffisante pour que les techniques de l'analyse synchronique puissent y être valablement appliquées. Dans son principe donc, le DMF est un dictionnaire de synchronie ; il n'est historique que par la synchronie qu'il traite, c'est-à-dire une synchronie ancienne.

Cela signifie que l'organisation des sens est largement comparable à celle du TLF. Dans le cas où les sens ont au plus un sème commun, sans qu'on puisse les relier par un indicateur métalinguistique du type *p. ext.*, *p. restr.*, *p. méton.* ou *au fig.*, la numérotation adoptée est **I., II., III.** ..., en chiffres romains. Cette numérotation est également celle des distinctions grammaticales ; celles-ci sont largement privilégiées dans le DMF, en particulier pour les verbes, où l'on distingue, chaque fois que c'est

justifié, l'emploi transitif et l'emploi intransitif ou pronominal ; ces distinctions s'accompagnent presque toujours, en effet, d'une différence sémantique et elles ont l'avantage de procurer des structures lisibles et faciles de consultation. Une *règle d'ordination* donne la priorité au sens propre (A) sur les sens figurés ou bien sur les emplois par extension, par restriction ou par métonymie (B, C, D...).

Mais fréquemment l'ordre est synchroniquement indécidable, et la seule issue est alors de recourir à l'histoire. Le choix entre l'extension ou la restriction ne peut avoir de fondement logique : *homme* au sens de «être humain» conduit par restriction à *homme* «être humain de sexe masculin» ; mais l'inverse, c'est-à-dire l'extension de sens, est, logiquement, tout aussi imaginable. Ce qui fait choisir la restriction de sens, c'est la connaissance que nous avons du latin *homo* (extensif par rapport à *vir*). La même constatation vaut pour certaines métonymies : on va indifféremment du tout à la partie ou de la partie au tout. Mais nous savons, par l'histoire, qu'il faut par exemple partir de *bureau* «pièce de toile (de *bure*) posée sur la table à écrire» pour arriver à *bureau* «table à écrire». Logiquement, l'inverse serait à nouveau tout aussi acceptable. Il n'est pas jusqu'aux analogies, quand on reste dans le concret, où l'on ne puisse aller dans un sens ou dans l'autre. Pour *calice*, l'analogie de forme associe le calice d'une fleur et le calice à boire. Là encore, ce sont les connaissances historiques qui seules font interpréter comme secondaire la désignation de l'objet botanique. Quant à la polysémie de sens, rien ne peut jamais faire préférer un ordre plutôt qu'un autre si ce n'est l'histoire (les considérations de fréquence parfois évoquées ne sauraient être qu'arbitraires).

Une des préoccupations du lexicographe est donc de retrouver, même quand la visée est fondamentalement synchronique, les traces du sens étymologique. Prenons pour exemple le mot *avis*. La grande masse des occurrences se rapporte au sens, relativement abstrait, de «ce qui semble à qqn» : *à mon avis, à son avis, de mon avis, de son avis, selon mon avis, selon son avis* ; *par avis* «comme il apparaîtrait, comme il semble» ; *par droit avis...*, *il est à qqn avis que...*, *il semble avis à qqn que...*, *avoir avis que...* ; ou bien au sens de «opinion concertée», de «réflexion qui conduit à une opinion, une décision» : *avoir avis avec qqn/ensemble/entre soi* «se concerter» ; *avoir avis si...* «avoir une réflexion commune pour savoir si...» ; *prendre avis* «se décider» ; ou encore au sens de «capacité de jugement, bon sens, intelligence, sagesse» : *avoir petit d'avis* «avoir peu de sagesse, d'intelligence», *perdre (son) avis* «perdre le jugement», *pourvu d'avis* «avisé», *de grant/haut avis* «intelligent, sage», *par bon/grant avis* «avec intelligence, avec sagesse ou prudence», *par pauvre avis* «sottement». Cependant un certain nombre d'exemples renvoient à des significations beaucoup plus concrètes et plus proches de *visum* «chose vue, objet vu, vision». Ainsi la locution *d'un avis* signifie «d'une seule visée», c'est-à-dire «d'un seul coup» ; ailleurs *avis* prend le sens de «apparence» (en particulier de «apparence gracieuse») ; ailleurs encore, il est synonyme d'*avision* et désigne une vision, un songe, ce que l'on croit voir. Une grande attention doit être portée à de tels effets qui rappellent les origines. Dans l'article lexicographique, il y a tout à gagner, pour la cohérence historique, à les placer en tête.

Une *règle d'explicitation* consiste à spécifier partout les motifs de l'organisation lexicographique. Le DMF se fonde sur des structures arborescentes sans vide. Ainsi on s'interdit d'écrire :

A. – 1. «...»

...

B. – Au fig.

en préférant :

A. – Au propre

1. «...»

...

B. – Au fig.

Une balise <EMPL> («conditions d'emploi») permet de noter toutes les données qui justifient les distinctions sémantiques que l'on adopte, en particulier les données du contexte qui accompagnent l'émergence du sens en cause (notamment la sous-catégorisation : *bon* a le sens de «courageux, valeureux» quand il s'applique à un substantif désignant un chevalier, un soldat, un homme d'armes).

On ajoutera enfin une *règle de «cohérence structurelle à l'intérieur d'une famille étymologique»* : elle conduit à aligner le plus possible le plan d'un dérivé sur celui du mot de base.

2.4.4. Balises métalinguistiques

Les indicateurs métalinguistiques figurent généralement sous une balise <IND>. On peut en distinguer de plusieurs sortes :

– Indicateurs grammaticaux

À l'accompli	Au compar.	Constr. dir.
À l'actif	Au fém.	Double constr.
À l'impér.	Au masc.	Élém. de compos.
Adj.	Au passif	Empl. abs.
Adj. et subst. masc.	Au plur.	Empl. adj.
Adj. indéf.	Au sing.	Empl. adv.
Adj. num.	Au sing. à valeur coll.	Empl. exclam.
Adj. num. cardinal	Au subj.	Empl. factitif
Adj. num. ordinal	Au superl.	Empl. impers.
Adj. poss.	Au vocatif	Empl. interj.
Adv.	Constr.	

Empl. intrans.	Interj.	Part. prés. en empl. adj. ou subst.
Empl. intrans. ou pronom.	Loc.	Part. prés. en empl. adj. ou subst. fém.
Empl. intrans. (ou pronom.)	Loc. adj.	Part. prés. en empl. adj. ou subst. masc.
Empl. massif	Loc. adv.	Part. prés. en empl. subst.
Empl. prép.	Loc. conj.	Part. prés. en empl. subst. fém.
Empl. pronom.	Loc. exclam.	Part. prés. en empl. subst. fém.
Empl. pronom. (ou intrans.)	Loc. fig.	Part. prés. en empl. subst. masc.
Empl. pronom. à sens passif	Loc. interj.	Part. prés. en empl. subst. masc.
Empl. pronom. réciproque	Loc. nom.	Prép.
Empl. pronom. réfl.	Loc. prép.	Pron.
Empl. subst.	Loc. prov.	Pron. indéf.
Empl. subst. fém.	Loc. verb.	Sens actif
Empl. subst. masc.	Part. passé	Sens passif
Empl. trans.	Part. passé en empl. adj.	Subst.
Empl. trans. dir.	Part. passé en empl. adj. ou subst.	Subst. à valeur de neutre
Empl. trans. indir.	Part. passé en empl. adj. ou subst. fém.	Subst. fém.
En appellatif	Part. passé en empl. adj. ou subst. masc.	Subst. fém. plur.
En appos.	Part. passé en empl. subst.	Subst. masc.
En compos.	Part. passé en empl. subst. fém.	Subst. masc. et fém.
En incise	Part. passé en empl. subst. masc.	Subst. masc. plur.
Inf. subst.	Part. prés.	Subst. masc. à valeur de neutre
	Part. prés. en empl. adj.	

— Indicateurs sémantiques

Au fig.	P. ext.
Au propre	P. métaph.
Au propre et au fig.	P. méton.
Dimin. de *	P. restr.
En partic.	
P. anal.	

— Indicateurs stylistiques

Arg.	P. allég.
Empl. pléonastique	P. antiphr.
En apostrophe	P. compar.
Fam.	P. ell.
Fréq.	P. euphém.
Latinisme	P. exagér.
Littér.	P. hyperb.

P. iron.
P. personnif.
P. plaisant.
Par dérision
Péj.
Poét.
Pop.

Prov.
Rare
Sav.
Vulg.
Vx

Les indicateurs de domaine figurent sous une balise spécifique (<DOM>); les versions ultérieures du DMF devraient en augmenter et en améliorer sensiblement l'application :

ADMIN.
ADMIN. ECCL.
AGRIC.
ALCHIM.
ALG.
ANAT.
ANTIQ.
ARCHIT.
ARITHM.
ARM.
ART MILIT.
ASTR.
BOT.
CARR.
CHARPENT.
CHASSE
CHIR.
COMM.
CONSTR.
COST.
COST. LITURG.
COST. MILIT.
COUT.
CUIS.
DÉMON.
DIPLOM.
DIVIN.
DR.
DR. CANON
DR. COUTUM.
DR. FÉOD.
DR. MATRIMONIAL
DR. ROMAIN
DRAP.

EAUX ET FORÊTS
ESCR.
FAUCONN.
FIN.
FORTIF.
GÉOGR.
GÉOL.
GÉOM.
GRAMM.
HÉRALD.
HIPPI.
HIST.
HIST. GR.
HIST. GR. ET
ROMAINE
HIST. JUIVE
HIST. ROMAINE
HORLOG.
ICONOGR.
INSTIT.
INSTIT. ROMAINE
JEUX
JOAILL.
LING.
LITURG.
LITURG. CATH.
LITURG. CHRÉT.
LOG.
MAGIE
MAR.
MATH.
MÉD.
MÉD. VÉTÉR.
MES.

MÉTIER
MINES
MOBIL.
MONN.
MOR.
MOULINS
MUS.
MUS. GRECQUE
MYTH.
NUMISM.
ORFÈVR.
ORNITH.
PEAUSS.
PÊCHE
PHILOS.
PHYS.
PHYSIOL.
REDEV.
RELIG.
RELIG. CHRÉT.
RELIG. JUDÉO-CHRÉT.
RELIG. JUIVE
RELIG. PERSE
RELIG. ROMAINE
RHÉT.
SC.
SCOLAST.
SPIRITUALITÉ
TECHNOL.
THÉOL.
THÉÂTRE
TOURN.
VERSIF.

Les régionalismes sont repérés par la balise <RÉGION>. Sont actuellement utilisés les indicateurs suivants : anglo-normand, Auvergne, Bourgogne, Bretagne, Champagne, Dauphiné, Est, Flandres, Franche-Comté, francoprovençal, Hainaut, Lorraine, Nord, Nord-Est, Normandie, Ouest, Picardie, Poitou, Provence, Savoie, Suisse romande, Sud (c'est-à-dire Sud du domaine d'oïl), Sud-Est (du domaine d'oïl), Wallonie. En raison des difficultés inhérentes à la notion de "régionalisme", le DMF fait certes de cette balise un usage restreint, la préférence étant donnée à l'affichage, par le biais de la Bibliographie, de la localisation régionale des textes où les exemples sont pris. Mais dans les éditions à venir, l'usage de la balise <RÉGION> devrait s'étendre et s'appliquer notamment aux "Remarques", où le caractère régional est fréquemment mentionné.

Outre ces balises proprement métalinguistiques, le DMF présente des balises diverses qui structurent son métatexte.

— Balise de définition (<DEF>) : elle permet des recherches sur la métalangue définitoire (par exemple le repérage de toutes les définitions qui comportent le mot *cheval*). Dans le DMF, la définition est préférentiellement de type hyperonymique. Dans le cas où la définition du moyen français et celle du français moderne sont strictement identiques et s'il n'y a pas de risque de confusion polysémique, on peut se passer de définition et se contenter de traduire : *bleu* «bleu» (et non «qui est de la couleur du ciel quand il est sans nuages»). Pour les termes techniques, il peut être utile de définir sommairement et en même temps de traduire : *farcin* «maladie contagieuse du cheval, farcin». Une grande attention est portée aux données contextuelles, nécessaires au décodage polysémique mais qui n'entrent pas dans le contenu même de la définition ; elles figurent dans les conditions d'emploi (<EMPL>) ou bien, entre parenthèses, dans la définition elle-même : *abandonner* a le sens de «quitter» quand l'objet désigne un lieu ; soit : «quitter (un lieu)», ou bien : [Un lieu] «(le) quitter». Lorsqu'une définition est empruntée, la source est marquée par une balise <EMPR>.

— Une autre balise signale les constructions, les syntagmes plus ou moins figés ou bien les locutions (<LOC>). Les locutions sont suivies d'une définition, contrairement aux syntagmes, relevés en raison de leur fréquence, mais qui sont compositionnels, c'est-à-dire compréhensibles par le sens des parties qui les composent. Les locutions prennent place sous le premier substantif qu'elles comportent (de la gauche vers la droite), à défaut sous le premier adjectif, à défaut le premier adverbe, à défaut le premier verbe (on ne tient donc pas compte des mots grammaticaux). Par exemple :

<i>gens de labour</i> «paysans»	est traité sous	<i>gens</i>
<i>mere au lait</i> «nourrice»	est traité sous	<i>mere</i>
<i>à cheval</i>	est traité sous	<i>cheval</i>
<i>donner paix</i>	est traité sous	<i>paix</i>
<i>prendre en gré</i>	est traité sous	<i>gré</i>

<i>se jeter hors de blâme</i> «se justifier»	est traité sous	<i>blâme</i>
<i>mettre la main à l'œuvre</i> «travailler»	est traité sous	<i>main</i>
<i>estre en mauvaise voie</i>	est traité sous	<i>voie</i>
<i>avoir fol parler</i>	est traité sous	<i>parler</i>
<i>mettre sa vielle sous le banc</i> «renoncer»	est traité sous	<i>vielle</i>
<i>gracieuse maladie</i> «mal d'amour»	est traité sous	<i>maladie</i>
<i>cent fois pour une</i>	est traité sous	<i>fois</i>
<i>avoir froid</i>	est traité sous	<i>froid</i>
<i>aller ensemble</i>	est traité sous	<i>ensemble</i>
<i>par ouïr dire</i>	est traité sous	<i>ouïr</i>

De tels principes, d'application mécanique, ont l'avantage de la commodité. Mais il va sans dire qu'ils peuvent souffrir des exceptions, l'interrogation électronique permettant en tout état de cause de s'y retrouver. En particulier, la locution gagne à être traitée sous le vocable qui s'éloigne le plus de son sens habituel ; ainsi *baptiser le vin* «le couper d'eau» sera mentionné de préférence sous *baptiser* (d'autant plus que *vin* est commutable avec des mots qui désignent des sortes de vins). De même, on choisira comme lieu de traitement le vocable rare que la locution peut comporter ; *vin claré* «vin clairé» sera mis préférentiellement sous *claré*.

Une autre balise encore permet de renvoyer aux synonymes. Deux cas de figure se présentent : la balise <SYNON> complète la définition, la balise <SYNON DE> en tient lieu. Celle-ci ne peut être utilisée que dans le cas d'une synonymie apparemment absolue (notamment quand une variation suffixale, par exemple *-age /-ment*, n'induit aucune différence sémantique explicitable) ; l'autre est préférée quand la synonymie n'est qu'approchée et que les définitions respectives divergent tant soit peu. En cas de synonymie multiple, l'ensemble des synonymes peut n'être explicité que sous un seul des vocables, celui-ci étant mentionné comme synonyme sous tous les autres.

La balise «remarque» se présente elle aussi sous deux formes :

- <REM> est utilisé si l'article tout entier est touché ;
- <Rem> l'est si la portée est locale et si la remarque ne touche qu'une rubrique particulière.

Les renvois se font au moyen d'une balise <CIBLE>, la cible étant indifféremment l'entrée ou le lemme.

2.4.5. Balises d'exemples

Les exemples sont tous balisés en tant que tels. Ils comportent aussi des balises internes qui en structurent les données :

- balise d'occurrence ;
- balise de «commentaire» (<COMM>) : on y enregistre notamment les variantes ; le commentaire sert aussi à noter les anaphores et, plus généralement, toute indication, notamment de traduction, susceptible d'éclairer l'exemple cité ; placé à la fin de l'exemple, après la référence, le commentaire autorise toutes les remarques que l'on souhaite y apporter ; la mention «seul ex.» signifie que les dictionnaires régulièrement consultés ne connaissent pas d'autres exemples ;
- balise de «référence» : elle entoure la référence bibliographique abrégée ; on y distingue un champ «auteur», un autre de «titre», un autre encore de «date» ; la référence se fait toujours, sauf indication explicite, à la page de l'édition citée.

Dans tous les cas, le texte reproduit est celui de l'édition. Les seuls changements, tout à fait mineurs, concernent les noms de personnages dans les textes de théâtre : on les note toujours en majuscules ; sont également mises systématiquement en majuscules les initiales des vers. Une coupure pratiquée par le rédacteur est signalée par (...); une lacune du texte par [...]; la présentation (.), (,)... marque un signe de ponctuation que l'on propose de supprimer ; [.,], [,]... un signe de ponctuation que l'on propose d'ajouter ou qui modifie un signe de l'édition. Les corrections d'éditeur sont reproduites à l'identique de l'édition. Cependant on dissocie, mais seulement pour la forme occurrente, les éléments que l'éditeur a éventuellement agglutinés (sous l'entrée *entendement*, *lentendement* est noté *l'entendement*).

* *
*

On voit la perspective qui est celle du DMF. Les choix sont résolument orientés du côté de la lexicographie informatisée. Dans l'histoire récente de la lexicographie, l'informatisation connaît au moins trois degrés de perfectionnement :

- à un premier niveau, seule est informatisée la documentation sur laquelle le projet lexicographique se fonde ; c'était le cas du TLF dès les années 1960 ; on a vu, chemin faisant, la part très importante des données informatisées dans la documentation du DMF ;

- à un niveau plus élaboré, le dictionnaire lui-même est informatisé ; toutes les informations qu'il véhicule sont alors strictement balisées et se prêtent du fait même à toutes les interrogations croisées que l'on souhaite, hors des contraintes de la linéarité ; il en est ainsi du TLF informatisé (TLFi) et aussi du DMF, à cette différence près cependant que le balisage du TLF a été pratiqué *a posteriori*, à partir

de la version imprimée, moyennant des procédures complexes de «rétroconversion», alors que dans le DMF il s'opère au moment même de la rédaction ;

— à un niveau plus élevé encore, la rédaction elle-même est informatisée : elle est alors conduite au moyen d'instruments qui assistent le rédacteur dans sa tâche. On a vu que le DMF bénéficie d'outils qui permettent au rédacteur d'introduire directement les données et de les organiser à l'écran et que différents modules de contrôle vérifient l'adéquation au modèle et la conformité aux normes qui ont été fixées. C'est grâce à ces instruments que le travail lexicographique peut s'engager pleinement dans la voie d'avenir d'une *lexicographie évolutive*.

L'avantage le plus appréciable d'une *lexicographie évolutive* est qu'à tout moment le dictionnaire reste ouvert aux améliorations souhaitables. C'est dire que les observations éventuelles que l'on voudra bien nous faire seront accueillies avec reconnaissance. Il est réconfortant de penser que les erreurs et les insuffisances ne sont plus désormais irréversibles.

(Dernière mise à jour de ce texte : le 31.10.2017)