



**HAL**  
open science

# De la nécessité et de l'intérêt d'une mutualisation informatique des connaissances sur le lexique de notre langue

Jean-Marie Pierrel

► **To cite this version:**

Jean-Marie Pierrel. De la nécessité et de l'intérêt d'une mutualisation informatique des connaissances sur le lexique de notre langue. 1er Congrès Mondial de la Linguistique Française, 9-12 juillet, 2008, Paris, France. pp.15-30. halshs-00334787

**HAL Id: halshs-00334787**

**<https://shs.hal.science/halshs-00334787>**

Submitted on 28 Oct 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# De la nécessité et de l'intérêt d'une mutualisation informatique des connaissances sur le lexique de notre langue.

Jean-Marie Pierrel

ATILF, Nancy Université & CNRS

[Jean-Marie.Pierrel@atilf.fr](mailto:Jean-Marie.Pierrel@atilf.fr)

Une rapide analyse de l'évolution de la linguistique au cours du dernier demi-siècle montre que la confrontation avec l'informatique a permis à la linguistique de se définir de nouvelles approches. C'est ainsi qu'au-delà d'une simple linguistique descriptive s'est développée une *linguistique formelle*, couvrant aussi bien les aspects lexicaux que syntaxiques ou sémantiques, qui tend à proposer des modèles s'appuyant sur une double validation, *explicative* d'un point de vue linguistique, *opératoire* d'un point de vue informatique. C'est elle aussi qui a permis *l'émergence d'une véritable linguistique de corpus* (Habert et col. 1997) permettant au linguiste d'aller au-delà de l'accumulation de faits de langue et de confronter ses théories à l'usage effectif de la langue.

Cette évolution a provoqué au sein du champ des études lexicales une véritable révolution qui fit de l'informatique un outil indispensable pour :

- étudier le lexique et ses propriétés à travers l'exploitation intelligente de textes et de documents ;
- structurer et normaliser les connaissances lexicales et lexicographiques ;
- valoriser, partager et mutualiser les résultats de la recherche sur le lexique de notre langue.

En complément de la contribution de Robert Martin (2008) qui discute des points 1 et 2 et montre la nécessaire jonction de la lexicographie à destination humaine et des « dictionnaires-machine » indispensables pour le TAL, nous souhaitons, pour notre part, revenir sur le point 3.

En effet dans le cadre des travaux de recherche en linguistique, et plus particulièrement sur le lexique, les aspects de valorisation et mutualisation des résultats acquis sont particulièrement importants et stratégiques pour consolider la position du Français au sein de la communauté internationale. Au début de ce 21<sup>e</sup> siècle qui devrait être celui d'une société de la connaissance, au même titre que les 19<sup>e</sup> et 20<sup>e</sup> furent ceux d'une société industrielle, si l'on souhaite maintenir une société multilingue, un effort particulier doit être fait pour permettre au français de conforter ses positions. Nous pensons pour notre part que, dans cette société de la connaissance, seules les langues fortement étudiées et modélisées, permettant ainsi des traitements automatiques, subsisteront comme langues véhiculaires de travail et d'échange et que les autres se verront réduites à une dimension uniquement vernaculaire. Contrairement à ce que quelques esprits chagrins prétendent les jeux sont loin d'être faits comme le souligne une récente étude de l'Union Latine (Union Latine, 2008).

Nous rappellerons tout d'abord les principales conditions facilitant une meilleure mutualisation des résultats de recherche et le rôle important de l'informatique et du Web dans ce cadre. Puis nous montrerons quelques exemples d'une telle mutualisation et terminerons par la présentation d'un exemple d'intégration et de valorisation de données lexicographiques et lexicales au sein du portail lexical du Centre National de Ressources Textuelles et Lexicales (CNRTL, [www.cnrtl.fr](http://www.cnrtl.fr)) en nous appuyant, en

partie du moins, sur des bilans que nous avons faits par ailleurs dans (Pierrel & Petitjean 2007) et (Pierrel & Buchi à paraître).

## 1 Conditions pour une réelle mutualisation de nos résultats

### 1.1 Une volonté commune comme réponse aux défis du XXI<sup>ème</sup> siècle

Une des premières conditions à un vrai partage et une mutualisation de nos avancées en recherche réside sans aucun doute par un changement de mentalité. Trop souvent encore nous rencontrons des chercheurs faisant preuve d'une frilosité quasi malade pour partager leurs résultats de recherche et les ressources linguistiques qui en découlent. Les justifications qu'ils nous donnent sont très diverses mais toutes aussi peu convaincantes. On peut citer entre autres :

- La crainte d'être dépossédés de leurs productions. Dans le domaine de la constitution de ressources linguistiques informatisées (corpus, dictionnaires ou lexiques), sans aller jusqu'à refuser un tel partage sous prétexte qu'un autre chercheur pourrait, à partir de leurs résultats, développer des recherches nouvelles auxquelles ils n'ont pas encore pensé, beaucoup de chercheurs de notre discipline restent « assis » sur leur production, refusant de diffuser leurs résultats ou ne les diffusant que partiellement, rendant ainsi leur ré-exploitation pas d'autres quasi impossible.
- L'espoir, trop souvent vain, d'une valorisation industrielle de leurs résultats. C'est en particulier souvent le cas pour de vastes ressources lexicographiques ou lexicales et cela conduit à un gâchis énorme dans la mesure où souvent plusieurs équipes sont amenées à refaire un travail déjà fait par d'autres. On pourrait, à se propos, se poser la question de savoir s'il ne conviendrait pas d'opter pour notre langue sur une philosophie comparable à celle retenue en génomique qui conduisit à rendre accessible à tous le décodage du génome humain.
- La volonté d'une publication papier avant toute ouverture de ces ressources sur le Web. Si ce souhait est à première vue le plus défendable, il convient de noter que les clauses contractuelles d'édition empêchent souvent ensuite une valorisation et un partage informatique sous le Web.

Dans mes fonctions de directeur de l'ATILF j'ai été confronté à ces différents types de justifications : même au sein de l'ATILF, il ne fut pas toujours facile de défendre une ouverture de nos ressources. Pourtant aujourd'hui, alors que la plupart de nos productions sont facilement accessibles sur le Web sur le site du laboratoire ([www.atilf.fr](http://www.atilf.fr)) ou sur celui du CNRTL ([www.cnrtl.fr](http://www.cnrtl.fr)), tous nos collègues en sont fiers car, outre l'ouverture de ces ressources vers le monde de la recherche, cette valorisation de nos productions de recherche est une des fiertés et des motivations de travail de chacun d'entre nous.

Le coût de définition et de production de vastes ressources linguistiques de qualité (corpus, dictionnaires et lexiques) est important et c'est un gâchis énorme de vouloir, pour chaque projet de linguistique ou de TAL, redéfinir l'ensemble des ressources dont on a besoin. A titre d'exemple, la construction d'un dictionnaire de langue tel le Trésor de la Langue française a nécessité près de cent personnes durant trente ans, et l'établissement d'une base de données textuelles tel FRANTEXT ([www.atilf.fr/frantext](http://www.atilf.fr/frantext)) s'est chiffré aussi en dizaines d'hommes-an. Sans vouloir plaider ici pour une rentabilisation de la recherche, il convient de prendre conscience que sans une véritable mutualisation de telles ressources, dans le domaine des sciences du langage qui nécessite d'aborder des aspects aussi divers que le lexique, la syntaxe, la sémantique, la pragmatique, chaque équipe de recherche ou chaque chercheur se verrait dans l'obligation de tout réinventer, alors même que nul ne peut être spécialiste de chacun de ces sous-domaines.

Un second point plaidant pour la mutualisation de ressources concerne l'évaluation de nos productions de recherche (modèles, analyseurs, systèmes de traitement), qui nécessite, pour des besoins de comparaison, la disponibilité de ressources de référence (corpus textuels, lexiques, ressources dictionnaires) accessibles, partagées et clairement identifiables.

Enfin, il convient de noter qu'en termes de valorisation de la recherche et de partage de connaissances avec nos concitoyens, une disponibilité accrue, en particulier sur le web, de nos productions de recherche est indispensable. Outre le fait que cela peut permettre un meilleur partage entre le monde de la recherche et la société civile, cela répond aussi à un besoin de plus en plus grand de connaissance chez nos concitoyens.

Mais ne nous leurrions pas, la constitution et la valorisation de telles ressources de qualité nécessitent des investissements en temps importants. Si l'on souhaite que des chercheurs puissent consacrer une partie de leur temps à de telles tâches au service de l'ensemble de la communauté scientifique, il convient de mieux prendre en compte cette activité de production des documents et ressources numériques dans leur évaluation et de mettre en place une structure servant à la fois de validation et de diffusion de ces productions. C'est en partie du moins le rôle que le CNRS a confié aux Centres Nationaux de Ressources, dont le CNRTL (cf. paragraphe 3)

## **1.2 Un respect des normes et des recommandations internationales**

### **1.2.1 De l'importance de codages facilement exportables**

Rendre disponibles des ressources lexicales entraîne aujourd'hui un minimum d'exigences, en particulier une interface conviviale, pour faciliter des interrogations par l'humain, et un codage normalisé, pour permettre des traitements informatiques de ces données.

L'ergonomie des interfaces hommes machines est un champ de recherche en soi. Je ne prétendrai pas que nous avons révolutionné le monde sur ce point mais chacun conviendra que les interfaces proposées pour le TLFi (Trésor de la Langue Française informatisé : [www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) ou celles du portail lexical du CNRTL ([www.cnrtl.fr/portail/](http://www.cnrtl.fr/portail/)) permettent déjà un premier partage facilité. Au-delà de la simple consultation via le Web, il convient aussi de donner aux collègues accès à la ressource sous une forme informatique normalisée. A ce niveau si XML s'impose encore convient-il de pouvoir accompagner cette ressource d'un schéma ou DTD<sup>1</sup> et de prévoir une adresse Web normalisée, stabilisée et pérenne.

Notons à ce propos que rien n'empêche de proposer divers niveaux de protection : ressources libres téléchargeables (c'est le choix que nous avons fait pour le lexique morphosyntaxique MORPHALOU), ressources accessibles uniquement en consultation via un logiciel spécifique (comme pour le TLFi) ou ressources sous droits (c'est le choix que nous avons fait pour le XML du TLFi qui est accessible dans le cadre de conventions de coopération avec le laboratoire). Dans tous les cas il convient que les métadonnées (description des ressources et des modes d'exploitation des ressources) soient largement accessibles. C'est un des objectifs que nous nous sommes donnés dans le cadre du CNRTL et c'est aussi ce que se propose de mettre en place le projet européen d'infrastructure de recherche CLARIN (Common Language Resources and Technology Infrastructure : [www.clarin.eu](http://www.clarin.eu)).

Notons à ce niveau que notre communauté francophone devrait s'investir un peu plus dans des initiatives internationales tel la TEI (Text Encoding Initiative : <http://www.tei-c.org/>) qui se propose de construire des recommandations internationales pour le codage de ressources linguistiques informatisées dont les dictionnaires et les lexiques.

### **1.2.2 La normalisation comme outil de partage de ressources**

Au-delà du suivi de recommandations, telle celles de la TEI, il convient aussi, dans la mesure du possible, de participer à l'émergence de normes internationales pour faciliter encore plus l'interopérabilité entre ressources. Les travaux autour de LMF (*Lexical Markup Framework*) sont de ce point de vue remarquables pour une double raison : d'une part c'est un des premiers exemples d'émergence d'une norme ISO dans le domaine des ressources lexicales, d'autre part, la communauté française s'est particulièrement investie pour faire émerger cette norme (Romary, 2004).

Deux exemples méritent d'être cités dans ce cadre :

**Le lexique MORPHALOU** des formes fléchies du français. Il s'agit d'un lexique extensionnel qui liste explicitement toutes les formes fléchies d'un lemme. Les données initiales (celles de la version 1.0) proviennent du *TLFNome*, la nomenclature du Trésor de la Langue Française qui a fourni 540.000 formes fléchies, réparties sur à peu près 68.000 lemmes. Dans la version 2.0, la couverture a été augmentée par environ 30.000 lemmes, extraits automatiquement du Trésor de la Langue Française informatisé. Il s'agit essentiellement de formes composés ou dérivées, ou de locutions n'ayant pas un statut d'entrée principale dans le TLFi. Actuellement, le lexique comporte 95.810 entrées lexicales. Le format de représentation suit les recommandations de normalisation pour les ressources lexicales de l'ISO (TC 37/SC 4) : il s'agit d'une instanciation actualisée du *Lexical Markup Framework*, version 9. *MORPHALOU 2.0* est en accès libre et téléchargeable sur le site du CNRTL à l'adresse [www.cnrtl.fr/lexiques/morphalou/](http://www.cnrtl.fr/lexiques/morphalou/).

Voici à titre d'exemple la structure XML d'une entrée dans MORPHALOU :

```
<lexicalEntry id="championne_1">
  <feminineVariantOf
target="champion_1">champion</feminineVariantOf>
  <formSet>
    <lemmatizedForm>
      <orthography>championne</orthography>

    <grammaticalCategory>commonNoun</grammaticalCategory>
      <grammaticalGender>feminine</grammaticalGender>
    </lemmatizedForm>
    <inflectedForm>
      <orthography>championne</orthography>
      <grammaticalNumber>singular</grammaticalNumber>
    </inflectedForm>
    <inflectedForm>
      <orthography>championnes</orthography>
      <grammaticalNumber>plural</grammaticalNumber>
    </inflectedForm>
  </formSet>
  <originatingEntry target="TLF">CHAMPION, ONNE,
subst.</originatingEntry>
</lexicalEntry>
```

**Le projet PROLEX** (Tran & Maurel 2006), piloté par le Laboratoire d'informatique (LI) de l'Université François-Rabelais de Tours, a pour but de fournir, à la communauté du traitement automatique des langues, des connaissances sur les noms propres, qui constituent, à eux seuls, 10% des textes journalistiques. La plate-forme PROLEX comprend un dictionnaire électronique relationnel multilingue de noms propres (Prolexbase), des systèmes d'identification des noms propres et de leurs dérivés, des grammaires locales, etc.

La modélisation du domaine des noms propres définie dans PROLEX repose sur deux concepts centraux : le nom propre conceptuel et le prolexème. Le nom propre conceptuel ne représente pas le référent, mais un point de vue sur ce référent. Il possède dans chaque langue un concept spécifique, le prolexème, qui est une famille structurée de lexèmes. Autour d'eux, sont définis d'autres concepts et des relations (synonymie, méronymie, accessibilité, éponymie, etc.). Chaque nom propre conceptuel est en relation d'hyponymie avec un type et une existence au sein d'une ontologie. Il n'est pas évident de définir la notion de nom propre. La plupart des définitions insistent sur le caractère unique de son référent et sur une sémantique et une syntaxe qui lui est propre. Le point de vue adopté ici est celui de (Jonasson, 1994) qui propose une définition plus large incluant ce qu'elle appelle les noms propres purs (noms de personne et noms de lieu) et les noms propres descriptifs qui résultent souvent de la composition d'un nom propre avec une expansion (Tour Eiffel, musée Rodin, etc.). Un nom propre descriptif peut être considéré comme

une expression définie figée ou en cours de figement (Jardin des Plantes, Médecins sans frontières, etc.). Cette définition est assez proche de celle utilisée dans le domaine du TAL depuis la conférence MUC6.

Une première version XML PROLEX est actuellement téléchargeable via le CNRTL ([www.cnrtl.fr/lexiques/prolex/](http://www.cnrtl.fr/lexiques/prolex/)) et une version au format XML est en cours de finalisation (Bouchou & Maurel à paraître)

## **2 L'informatique comme outil de mutualisation et de valorisation de productions lexicographiques**

### **2.1 Du Trésor de la Langue Française au Trésor de la Langue française informatisé**

Le *Trésor de la Langue Française* (TLF), Dictionnaire de la langue du 19<sup>e</sup> et 20<sup>e</sup> siècle (Imbs et Quemada 1976-1994) est le fruit d'une des dernières grandes aventures lexicographiques qui, sous les directions successives du Recteur Paul Imbs et du Professeur Bernard Quemada, regroupa à Nancy plus de cent collaborateurs durant 30 années. Reflet fidèle du TLF, jusque dans sa présentation, le TLFi ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) se caractérise, comme le TLF, par la richesse de son matériau et la complexité de sa structure :

- Importance de sa nomenclature : 100 000 mots avec leur étymologie et leur histoire, et 270 000 définitions.
- Richesse des objets méta-textuels inclus dans chaque article (vedettes, codes grammaticaux, indicateurs sémantiques ou stylistiques, indicateurs de domaines, définitions, exemples référencés...).
- Richesse des 430 000 exemples, tirés de plus de deux siècles de production française, essentiellement littéraire.
- Diversité des rubriques : une rubrique d'analyse sémantique synchronique (couvrant la période 1789 à nos jours), une rubrique « prononciation et orthographe », une rubrique étymologie et histoire, une rubrique de statistique lexicale et une rubrique bibliographie pour les principaux articles.

La version informatique du TLF (Dendien et Pierrel, 2003) intègre, de plus, des accès à très haut niveau de tolérance permettant une insensibilité aux accents, une tolérance aux fautes d'orthographe courantes, un traitement phonétique et un traitement morphologique. Ainsi, on peut offrir une correction automatique des fautes, permettre des accès à partir de formes et non plus uniquement de lemmes ou de vedettes et proposer des procédures d'accès diversifiées pour une consultation humaine.

Nous ne reviendrons pas ici sur les étapes d'informatisation du TLF traitées par ailleurs (Dendien et Pierrel 2003). Nous nous contenterons de rappeler les principaux accès que permet le TLFi qui correspond à une rétro-conversion de la version papier du TLF pour laquelle, par des procédures de repérage semi-automatique des objets textuels composant les articles du dictionnaire original, fut introduit un balisage fin, tant typographique (de manière à conserver une image fidèle du TLF) que sémantique (repérage des principaux objets textuels au sein de chaque article). Quelques chiffres peuvent donner un aperçu de la finesse de ce balisage : après validation sur l'ensemble des seize tomes, 36 613 712 balises XML ont été positionnées : 17 364 854 balises typographiques, 1 070 224 balises décrivant la hiérarchie, 18 178 634 balises repérant les objets textuels, dont 92 997 entrées et 64 346 locutions faisant l'objet de 271 166 définitions et illustrées par 427 493 exemples.

C'est ce balisage fin du TLF et l'exploitation du document XML correspondant qui nous permet de proposer des accès à l'ensemble du dictionnaire, cumulant les avantages d'un dictionnaire avec ceux d'une ressource textuelle et d'une véritable base de données lexicales :

- Recherche d'un mot, d'une expression ou d'une forme lexicale plus ou moins bien orthographiés, avec possibilité, via un « panneau de réglage », de mettre en évidence divers champs dans le résultat de la

recherche (définition, catégorie grammaticale, domaine notionnel, exemple, auteur d'exemple, construction, indicateur, etc.).

- Possibilité d'hyper-navigation à l'intérieur du dictionnaire permettant en un clic-souris de passer d'un mot de la métalangue à sa description lexicographique et notamment à sa définition.
- Interrogations assistées ou requêtes complexes exploitant l'ensemble de la structure du dictionnaire à travers le croisement de multiples critères.

La mise à disposition sous le Web, parallèlement à la publication d'une version CDRom, a rencontré un vif succès tant auprès des chercheurs, des utilisateurs universitaires que du grand public : objet de plusieurs centaines de milliers d'interrogations quotidiennes en provenance de tous les continents, il est référencé par d'innombrables sources (collèges, lycées, universités, associations, publications à caractère culturel ou scientifique, etc.). La notoriété qu'il a acquise en fait un outil de promotion appréciable de la langue française.

Conscients que le TLFi constitue un patrimoine important, financé essentiellement par le CNRS, nous avons le souhait de le mettre le plus possible à la disposition de la communauté scientifique à travers des projets coopératifs de recherche. Et toute nouvelle proposition de coopération sera étudiée avec soin.

## 2.2 Le dictionnaire du Moyen Français (DMF)

### 2.2.1 Un nouveau concept : la lexicographie évolutive

C'est en 1980, lors du 3<sup>e</sup> Colloque international sur le moyen français, que naît l'idée d'un *Dictionnaire du Moyen Français* (DMF). Conçu par Robert Martin, qui l'a lui-même dirigé de 1982 à 2000, le DMF s'est poursuivi sous les directions successives de Bernard Combettes (2000–2002) et de Hiltrud Gerner (2003–2007), et continue à s'enrichir sous celle de Sylvie Bazin-Tacchella (depuis 2008)<sup>2</sup>.

Après la publication, en 1998, d'un volume expérimental regroupant la tranche alphabétique A-ah (DMF0), le projet a changé, non d'objectif, mais de forme. Le support papier a été – au moins provisoirement – abandonné au profit d'une publication électronique par étapes successives, sous forme de base de données lexicales évolutive encodée au format XML (cf. [www.atilf.fr/dmf](http://www.atilf.fr/dmf)). Voici comment le maître d'œuvre du DMF décrit les avantages de cette option : « *L'idée centrale qui guide le projet du DMF est 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 ouverte à tous les développements ultérieurs.* » (Martin 2007 : 1).

On distingue ainsi le DMF1, disponible depuis 2003 sur internet, le DMF2 (version actuellement en ligne) et le DMF3, en cours d'élaboration (sans parler des évolutions futures). La version DMF1 (cf. Gerner 2005) regroupe treize lexiques, qui analysent chacun le vocabulaire d'un corpus clos. Grâce à une lemmatisation préalable, les treize lexiques sont interrogeables ensemble (32.779 lemmes traités dans 84.778 articles : en moyenne, un lemme a donc été traité dans deux à trois lexiques différents).

Quant à la version actuellement disponible du DMF2 (cf. Martin/Gerner/Souvay à paraître), elle regroupe dix-sept lexiques, dont notamment le « Lexique complémentaire », lequel s'appuie sur un corpus très large, composé entre autres de textes saisis à l'ATILF, de dépouillements d'éditions scientifiques et de dictionnaires tant d'époque que rétrospectifs. Ce « Lexique complémentaire », qui compte à lui seul 35.000 lemmes, traite surtout des unités lexicales rares (car absentes des treize lexiques du DMF1) : hapax, créations d'auteur, latinismes sans grand lendemain, mots-fantômes dûment dénoncés en tant que tels. Mais on y relèvera aussi des lexèmes encore bien vivants à l'époque contemporaine, comme *cécité* ou *palissade* (Gerner 2007a : 71). La nomenclature du DMF2 se trouve ainsi presque doublée par rapport au DMF1, puisque le nombre de lemmes est porté à 60.241 (traités dans 117.723 articles, ce qui correspond à moins de deux articles par lemme en moyenne).

Pour ce qui est du DMF3, il se met entièrement au service de la cohérence interne : sa plus-value résidera essentiellement en un nombre substantiel de méta-articles synthétiques élaborés à partir de plusieurs articles consacrés à un seul lemme.

## 2.2.2 Une richesse d'exploitation inégalée grâce à sa valorisation informatique

Le DMF permet une variété de types de requêtes sans précédent pour un état ancien de la langue française (et peut-être sans précédent pour un état ancien de n'importe quelle langue). Ainsi, pour ne mentionner que quelques-uns des modes d'interrogation multidirectionnels prévus, l'ensemble d'une famille étymologique est accessible à travers un simple clic sur l'étymon du FEW ; l'exploitation informatique des définitions permet un accès onomasiologique aux données ; enfin un menu spécial est dévolu à la recherche de locutions, ce qui ravira les spécialistes de phraséologie. Et l'hypernavigation ne s'arrête pas aux frontières du dictionnaire stricto sensu, puisque le DMF se trouve au centre d'un réseau d'interconnexions de plus en plus dense (cf. Gerner 2007b).

Dans ce qui suit, nous illustrerons trois types de requêtes particulièrement utiles pour le chercheur en linguistique historique par des exemples issus des recherches réalisées à l'ATILF. Ces trois cas de figure relèvent de la lexicologie, de la pragmatique et de la morphologie constructionnelle. Inutile de dire que bien d'autres exploitations sont possibles.

### Lexicologie

La première vocation du DMF consistant à décrire les propriétés phonétiques, sémantiques, morphosyntaxiques et variationnelles du lexique du moyen français, l'apport du dictionnaire est particulièrement important dans le domaine des études lexicologiques synchroniques ou diachroniques. Un exemple rapide suffira pour illustrer cette utilisation classique du DMF.

Lors de la refonte des notices étymologiques des articles *adresse1*, *adresse2* et *adresse3* du TLFi réalisée dans le cadre du programme de recherche TLF-Étym, il s'agissait, entre autres, de tester la validité de la datation « depuis 1559 » que le TLFi propose pour le sens I. A. sous *adresse3* : « qualité d'une personne [...] parvenant aisément à atteindre un but ou à obtenir un résultat (le but visé est la réussite dans des actions nécessitant l'utilisation du corps et en particulier des mains) [synonymes : habileté, dextérité] ». Ni le FEW (von Wartburg 1928 in FEW 3, 84b, \**directiare* I), ni le Godefroy ou le Tobler-Lommatzsch (qui ne donnent pas ce sens s.v. *adrece*), ni l'AND (qui ne traite pas ce lexème) ne permettant d'aller au-delà de l'information disponible dans le TLFi, le regard des rédacteurs s'est rapidement tourné vers le DMF. Le mode « Recherche d'une entrée » du dictionnaire mène à neuf articles, issus de neuf lexiques différents, dont le suivant, tiré du lexique de Christine de Pizan par Joël Blanchard et Michel Quereuil :

#### Article 5/9

[ADRESSE](#)

FEW III [\\*DIRECTIARE](#)



**ADRESSE, subst. fém.**

- I. - "Chemin, direction" : Mais bien traihoit [Paris] en une *adresce* Fleches empaneas d'un arc ([CHR. PIZ., M.F., III, 1400-1403, 54](#)). ...tant veoir vous desiroye Que j'en ay empris longue voye, Par le renom de vo proece, Qui de ce me mit en l'*adrece* ([CHR. PIZ., M.F., III, 1400-1403, 143](#)).
- II. - "Art" : La roÿne d'Amasonie (...) D'armes savoit toute l'*adrece*. ([CHR. PIZ., M.F., III, 1400-1403, 141](#)).
- III. - "Celle qui montre le chemin, guide" : Philosophie y vi assise Moulte haultement, en tel devise Que bien semble haulte maistrresse Et des autres toutes l'*adresse* ([CHR. PIZ., M.F., II, 1400-1403, 104](#)).

*Pizan*

Joël Blanchard / Michel Quereuil

Le lecteur l'aura compris : l'attestation introduite sous II. avec la glose « art » antédater d'un siècle et demi (1559 > 1400/1403) l'apparition du sens « habileté manuelle » de *adresse*. Ajoutons que le DMF est aussi à l'origine de deux autres datations retenues pour ce groupe d'articles : avant 1343 (Guillaume de Machaut) pour « destination » (dans des locutions figées comme arriver à son adresse, cf. Petrequin/Buchi in TLF-Étym s.v. adresse1 I. A.) et à 1400 (Froissart) pour « moyen ou ensemble de moyens mis en œuvre par une personne pour atteindre un but (synonymes : diplomatie, doigté, finesse, manœuvre, procédé, ruse) » (cf. Petrequin/Buchi in TLF-Étym s.v. adresse3 II).

### Pragmatique

Contrairement à ce qu'un a priori réducteur pourrait faire penser, l'intérêt du DMF, même dans sa version actuellement disponible (DMF2), est loin de se cantonner au domaine lexical. Même si le traitement des unités grammaticales est réservé à une étape ultérieure, la consultation du dictionnaire s'impose dès à présent à tout chercheur intéressé par l'évolution des grammèmes, voire des marqueurs discursifs (ou pragmatèmes) du français.

Dans une publication récente, M.-B. Hansen (2005 : 47) date ainsi de la seconde moitié du 16<sup>e</sup> siècle (1587) l'apparition du marqueur *enfin* dans le sens « citons pour terminer (marque le dernier élément d'une énumération) » (actualisé par exemple dans Il y a à Paris trois polices : primo : la police du royaume [...] secundo : celle du régent [...] enfin celle de Dubois). Quiconque souhaite tester la solidité de cette datation de *enfin* « énumératif » se tournera vers le DMF, même si le dictionnaire ne contient pas, pour l'instant, d'entrée *enfin*. Par bonheur, en effet, le DMF connaît un mode « Recherche plein texte », qui propose 58 réponses à la requête *enfin*, dont la suivante :

#### Article 7/58 1 attestation

[BARATER](#)

FEW IX [prattein](#)



#### BARATER, verbe

[T-L : *barater* ; FEW IX, 330a : *prattein*]

Empl. trans.

**A.** - "Tromper, frauder" : ...qui tousjours estoient ententis et occupez en saintes meditations et par leurs bonnes euvres edifioient eulx et leur proesme. Ilz ne baretoient nullui ne bleçoient mais deduisoient leur vie en purté et en simplese, et crucifioient et tourmentoient leurs corpz sans pitié. ([Horloge de sapience S., c.1389, 81](#)). ...luxure donc y est pour deliter, et avarice y est pour profiter, et trayson **enfin** pour l'amant bareter. (EVR. CONTY, *Eschez amour. mor. G.-T.R., c.1400, 245*).

**B.** - "Agiter, battre"

- *Lait baraté*. V. [lait](#)

*Littérature didactique*

Hiltrud Gerner

L'article *barater* du DMF2 contient en effet, sous la rubrique A. « tromper, frauder », une attestation d'Evrart de Conty qui invite à antédater *enfin* « énumératif » de presque deux siècles (1587 > 1400)<sup>3</sup>, ce qui a des incidences certaines sur la modélisation de l'agencement historique des différentes valeurs, tant aspectuelles qu'énonciatives, de ce marqueur (cf. Buchi/Städtler 2008).

### Morphologie constructionnelle

Il est peut-être plus étonnant encore de constater que l'exploitation du DMF est d'un rendement tout aussi évident pour une autre sous-discipline linguistique, à savoir la morphologie constructionnelle.

Une étude stimulante de notre collègue Denis Apothéloz (2003) met en évidence le caractère double du préfixe IN- du français contemporain. Il distingue d'une part un préfixe in-1 « de base », qui est caractérisé par une certaine allomorphie : /in-/ devant une base en voyelle (inutile), /ɛ-/ devant une base en consonne bruyante (imbattable) et /i-/ devant une base en consonne sonante (illégal). Son sémantisme est soit négatif (inutile « non utile »), soit de type superlatif (inqualifiable « d'une extrême bassesse »).

Pour ce qui est du second terme du couple, in-2, il ne frappe que des bases en consonne sonante, et il est toujours réalisé /ɛ̃-/ ; son sémantisme est exclusivement négatif (inréparable « qui ne peut pas être réparé »). D. Apothéloz (2003 : 43) pose la question de l'époque à laquelle remontent les doublons en /ɛ̃-/ des dérivés en /i-/ formés sur une base en consonne sonante. Le DMF est-il à même de contribuer à répondre à cette question : comporte-t-il des lexèmes présentant le préfixe in- (/ɛ̃-/) devant des bases en r- ou l- ? Pour ce genre de recherches transversales, on dispose, sous « Recherche sur les entrées », d'un bouton « + options » qui permet de sélectionner, pour le filtre sur les entrées, non pas le mode par défaut « texte exact », mais une option de recherche d'une chaîne de caractères, en l'occurrence à l'initiale. Voici, à titre d'exemple, les résultats à la requête inr- :

The screenshot shows a search interface with a list of words on the left and a sidebar on the right. The list of words is as follows:

inrabilité	1
inracontable	1
inraisonnable	2
inrégi	1
inrégulé	1
<b>inrégularité</b>	<b>1</b>
inrégulièrement	1
inréméable	1
inrémédiable	1
inrémédiablement	1
inremuable	1
inrémunéré	1
inrenommé	1
inréparable	2

The sidebar on the right contains the following options:

- Structure de l'article
- Article sans exemples
- Article complet
- Formes de l'entrée
- Exemples de l'entrée
- Rechercher dans les textes

At the bottom of the interface, it says: "Nombre d'entrées : 28"

Le résultat de cette requête est d'une richesse insoupçonnée, puisqu'elle met en évidence 28 formations en inr-, dont la majorité non répertoriées par d'autres sources lexicographiques. Le témoignage du DMF aura donc été déterminant pour répondre à la question de l'âge du système morpho-sémantique complexe mis en évidence par D. Apothéloz (cf. Buchi à paraître).

## 2.3 Autres développements récents en lexicographie historique

### 2.3.1 TLF-Étym

Le programme de recherche TLF-Étym (cf. [www.atilf.fr/tlf-etym](http://www.atilf.fr/tlf-etym) et Buchi 2005), dirigé par Gilles Petrequin, se propose de réviser progressivement les notices intitulées « Étymologie et histoire » du TLFi. Il ne s'agit pas de réélaborer l'ensemble des quelque 54.000 notices historiques du TLF(i), mais d'une part de réinjecter dans cet ouvrage de référence les trouvailles faites par les différents chercheurs dans leur activité propre, d'autre part de réviser plus systématiquement des ensembles de notices particulièrement déficientes (par exemple celles consacrées aux anglicismes ou aux déonomastiques). Cette initiative fédère les forces vives de l'étymologie française à travers le monde (parmi les chercheurs étrangers impliqués dans le projet, citons par exemple Thomas Städtler et l'équipe du DEAF [Heidelberg], Franz Rainer [Vienne] ou Takeshi Matsumura [Tokyo]) et à revigorer ce secteur de la linguistique historique française.

Les notices révisées, téléchargeables sur le site TLF-Étym, contiennent des étymologies inédites mais aussi des antédatations (Stumpf/Evrard s.v. fabulateur : 1541 → ca 1360/1380 ; Städtler s.v. laconique : 1529 → ca 1372/1374 ; Petrequin s.v. ostensor : 1771 → 1673 ; etc.), et des rétrodatations (Koehl s.v. iota : ca 1240 → ca 1300 ; Steinfeld s.v. féodalement ; 1483 → 1514 ; Robin/Buchi s.v. véloce : 1804

→ 1818 ; etc.). Par ailleurs, TLF-Étym représente la structure d'accueil naturelle pour les notices étymologiques refondues dans le cadre du projet DETCOL (Développement et Exploitation Textuelle d'un Corpus d'Œuvres Linguistiques), dirigé par Bernard Colombat (cf. [http://ctlf.ens-lsh.fr/documents/ct\\_projet\\_detcol.pdf](http://ctlf.ens-lsh.fr/documents/ct_projet_detcol.pdf)). Le site propose ainsi une vue renouvelée de l'étymologie et de l'histoire de termes techniques de la grammaire, comme antécédent, conjonction, gémation, gérondif, négation, parfait, pluriel, pronom ou encore temporel. À titre d'exemple, nous invitons le lecteur à comparer la notice étymologique de l'article parfait<sup>2</sup> du TLFi avec sa version refondue sous TLF-Étym :

#### TLFi :

**Étymol. et Hist.** XIV<sup>e</sup>s. gramm. adj. *prétérit parfait* (*Ms. Fonds St Victor 867. Doctrinal avec glose, Anonyme du XIV<sup>e</sup>s.* d'apr. Ch. THUROT ds *Notices et Extraits des mss de la bibl. impériale et autres bibl.*, XXII, 2, p.184) ; 1596 gramm. subst. « temps qui marque un passé accompli » (HULSIUS, introd.). Représente la trad., par la forme *parfait*, du lat. *perfectum tempus* ou, plus brièvement, *perfectum* p. subst. de l'adj. *perfectus*, terme de gramm. du lat.

#### TLF-Étym :

*parfait*<sup>2</sup>, subst. masc.

#### ÉTYMOLOGIE

##### Histoire :

C. *prétérit parfait* loc. nom. masc. « temps verbal présentant le procès comme accompli et l'envisageant dans son résultat actuel ». Attesté de la fin 12<sup>e</sup> siècle [par référence à la grammaire latine] ([AelfricFH](#), page 102, in [STÄDTLER, TraLiPhi 37](#), page 128 : tempore preterito perfecto : *par le preterit parfet*, et plusquamperfecto : *et plusqueparfet*) à 1878 ([Ac<sup>7</sup>](#) s.v. *parfait* : En Grammaire, **Prétérit parfait**, ou substantivement, *Parfait*, Le prétérit qui marque une chose parfaite, une chose arrivée dans un temps qui n'est ni précis ni déterminé, comme *J'ai aimé, j'ai dit* [...]. L'emploi de *Parfait* [...] comme substantif[s] est le plus ordinaire). Cf. [STÄDTLER, Grammatiksprache](#), page 270 pour des attestations des 13<sup>e</sup>—15<sup>e</sup> siècles. Dernière attestation textuelle : 1775 ([CONDILLAC, Cours d'étude](#), volume 1, partie 2, chapitre 10, page 196, in [Gallica = Frantext](#) : On appelle *je ferois, prétérit imparfait ; je fis & j'ai [fait], prétérit parfait ; & j'avois fait, plusque parfait*). -

A./B. *parfait* subst. masc. « temps verbal présentant le procès comme accompli et l'envisageant dans son résultat actuel ». Attesté depuis 1690 ([FURETIERE<sup>1</sup>](#) s.v. *Parfait* : En termes de Grammaire on appelle le preterit *parfait*, un temps passé certain & défini, *J'ay aimé, J'ai leu*. Un **parfait** indéfini, *J'aymay, Je leus*). Si le rapport entre *prétérit parfait* et *parfait* est encore allusif dans cette citation, il deviendra explicite (“ou substantivement”) à partir de 1835 ([Ac<sup>6</sup>](#) [texte identique à celui de [Ac<sup>7</sup>](#), cf. ci-dessus C.]). — Une attestation isolée de 1596 ([HULSIUS<sup>1</sup>](#), introduction grammaticale : Des Verbes [...] Present. Die zeit die jetze da ist / als : Je mange, ich eß. / Imparfait. Zeit die halb fürüber ist / Je mangeoye, ich aß / Indiffinit. Zeit so auch fast fürüber / Je mangeay, ich aß / **Parfait**. Zeit die gar fürüber ist / J'ai mägé, ich hab gessen. / Advenir. Zeit so noch zukünftig ist / Je mangeray, ich werde essen) est à mettre à part ; il s'agit peut-être d'un emprunt à l'allemand *Perfekt* subst. neutre « temps verbal présentant le procès comme accompli et l'envisageant dans son résultat actuel », même si ce terme est attesté dans cette langue depuis le 17<sup>e</sup>/18<sup>e</sup> siècle seulement, [SCHULZ, Fremdwörterbuch<sup>1</sup>](#). -

##### Origine :

C. Transfert linguistique : calque du latin *praeteritum perfectum* loc. nom. neutre « temps verbal présentant le procès comme accompli et l'envisageant dans son résultat actuel (terme de grammaire) » (attesté depuis Quintilien, [TLL](#) 10/1, 1378, s.v. *perficio*), cf. *prétérit\** et *parfait<sup>1\*</sup>*. À ajouter [FEW](#) 9, 322b, PRAETERIRE 1 a β ; cf. [STÄDTLER, Grammatiksprache](#) 270.

A./B. Formation française : ellipse de *prétérit parfait* (cf. ci-dessus C.). Cette analyse se recommande en raison des témoignages explicites des sources lexicographiques ainsi que du parallélisme d'*imparfait* (<

*prétérit imparfait*, cf. *imparfait\** et *prétérit\**). Cf. VON WARTBURG in [FEW](#) 8, 237b, PERFECTUS I 1 b α, qui avance à tort l'hypothèse d'un calque du latin *perfectum* subst. neutre « temps verbal présentant le procès comme accompli et l'envisageant dans son résultat actuel (terme de grammaire) » (attesté depuis Varron, [TLL](#) 10/1, 1377, s.v. *perficio*) sur *parfait*<sup>1\*</sup>.

Dès les premiers témoignages d'un discours grammatical français, on relève le latinisme *prétérit parfait* (ci-dessus C.). C'est seulement à l'époque classique (en 1550, MEIGRET, [Traité](#), page 70 emploie encore *prétérit parfait*) que *parfait* (ci-dessus A./B.) vient concurrencer la locution nominale, avant de l'évincer complètement au 19<sup>e</sup> siècle. — L'adjectif *parfait* à sens grammatical (cf. STÄDTLER, [Grammatiksprache](#) 250-251) n'a pas de rapport étymologique direct avec le lexème traité ici ; il serait à classer s.v. *parfait*<sup>1\*</sup>.

Rédaction TLF 1986 : Étienne Ammann. - Mise à jour 2007 : Éva Buchi. - Relecture mise à jour 2007 : Jean-Paul Chauveau ; Thomas Städtler ; Jean-Pierre Chambon ; Melanie Lang ; Gilles Roques ; Takeshi Matsumura ; Gilles Petrequin ; Nadine Steinfeld.

### 2.3.2 Base des Mots-fantômes

La Base des Mots-fantômes (cf. [www.atilf.fr/MotsFantomes](http://www.atilf.fr/MotsFantomes) et Steinfeld à paraître) est une ressource de métalexigraphie critique qui se propose, sous la direction de Nadine Steinfeld, de recenser les *ghost words* ou *fantasmas lexicográficos* (Pascual Rodríguez/García Pérez 2007 : 170), c'est-à-dire les pseudo-lexèmes disposant à tort d'un statut lexicographique (« ces mots qui n'existent pas »), les sens fantômes et les lemmatisations erronées qui se trouvent dans les dictionnaires historiques et étymologiques français de référence.

Le dictionnaire de Godefroy contient une entrée labaille qui se lit comme suit :

LABAILLE, s. f., syn. d'*escoppe*, mod. *écoppe*, sorte de pelle creuse qui sert à vider l'eau entrée dans une embarcation :  
Et en doivent les vaisseaulx qui viennent esditz havres chargez de blez en grenier chascun vaissel plaine une escoppe ou *labaille* ou l'en puche l'eau. (1413, *Denombr du baill. de Constantin*, Arch. P 304, f<sup>o</sup> 116 v<sup>o</sup>.)

Von Wartburg n'a pas réussi à étymologiser cette donnée et l'a reléguée, en 1968, dans les matériaux d'origine inconnue ou incertaine (FEW 23, 108b, s.v. « *écoppe* »). Ce classement dans le « purgatoire » du FEW équivalait implicitement à un appel à la communauté scientifique à formuler des hypothèses étymologiques pour clarifier d'origine de labaille.

Ce n'est qu'en 2006 que le mystère a été élucidé : dans le cadre de la refonte de l'article *bājula* « bonne d'enfants » du FEW, Jean-Paul Chauveau (2006 : 2 : n. 4) identifie le vocable \*labaille comme une mélecture, par agglutination fautive de l'article défini, à rattacher au lexème *baille* n.f. « baquet de bois en forme de demi-tonneau ou de cône tronqué, spécialement utilisé sur les bateaux », attesté depuis 1340.

Or, étant donné la distance sémantique qui sépare « *écoppe* » de « *bonne* », il est extrêmement peu probable que le lecteur du Godefroy intéressé par l'étymologie de \*labaille consulte spontanément le FEW s.v. *bājula*. C'est précisément là que réside le principal intérêt de la Base des Mots-fantômes : elle centralise de façon commodément accessible les identifications de mots-fantômes dispersées dans des articles de revues, des communications de colloque, des comptes rendus et des articles lexicographiques, sans parler des rattachements inédits.

Nous sommes en présence ici d'un bel exemple de valorisation réciproque entre le FEW et la Base des Mots-fantômes.

## LABAILLE

1331-1500

lexie

### Source première :

Gdf 4, 686a

labaille, subst. fém.

« syn. d'*escope*, mod. *cope*, sorte de pelle creuse qui sert à vider l'eau entrée dans une embarcation »

Et en doivent les vaisseaulx qui viennent esditz havres chargez de blez en grenier chascun vaissel plaine une *escope* ou *labaille* ou l'en puche l'eaue 1413 (*Denombr. du baill. de Constentin*, Arch. P 304, f° 116v°)

### Source(s) secondaire(s) :

#### Source secondaire :

FEW 23, 108b

Concept : ÉCOPE

Anorm. *labaille* f. « sorte de pelle creuse qui sert à vider l'eau entrée dans une embarcation » (Coutances 1413)

### Justification :

mélecture portant sur un phénomène d'agglutination liée à l'article *la* 1413

Il convient de lire *la baille* « le baquet »

### Solution :

#### LA BAILLE

FEW 1, 206a BAJULA 2

### Correction des sources :

#### Source à corriger :

correction, déplacement-S

GdfC 8, 273b

baille, subst. fém.

« baquet »

Rem. : *Baille*, dans l'attestation de J. Du Clercq, *Mém.*, l. V, ch. XVIII, III, 375, citée ici, a été interprété à tort par Gay comme signifiant « palissade ».

#### Source à corriger :

correction, déplacement

FEW 1, 206a

BAJULA

2005 : J.-P. Chauveau ; T. Städtler, T. Matsumura

Entrée « labaille » dans la base des Mots-fantômes

### 2.3.3 Bibliographie Godefroy

« Les utilisateurs du dictionnaire de Godefroy sont partagés entre l'intérêt que présente l'immense masse documentaire contenue dans les huit mille pages de ses dix volumes et la frustration de ne pouvoir disposer d'une bibliographie des sources exploitées » (Ringebach 2003 : 191). Ce constat est à l'origine de la Bibliographie Godefroy en cours de construction par Jean-Loup Ringebach (cf. <http://www.atilf.fr/BbgGdf/> ; Ringebach à paraître). Le système de sigles utilisé par Godefroy est en

effet erratique. Pour ne citer qu'un exemple, le *Dit de buffet* (texte du milieu du 13<sup>e</sup> siècle) est tantôt cité sous la forme « Dit de buffet » ou « Dit du buffet », tantôt sous celle de « Du Vilain au buffet », sans que ces sigles soient accompagnés d'une datation. De ce fait, la Bibliographie Godefroy est en passe de devenir aussi incontournable que le complément bibliographique du DEAF.

La méthodologie utilisée pour l'élaboration de la Bibliographie Godefroy consiste à identifier les sigles opaques soit par recoupement avec des références plus explicites, soit par un travail philologique sur les éditions, plus rarement aussi les manuscrits, qui contiennent les différents textes.

Dans le cadre de notre recherche sur l'histoire du préfixe IN- du français contemporain (cf. ci-dessus 3.2.4), l'attestation suivante tirée du Godefroy nous laissait perplexe : Quant est de toy, tu mors et pince Par ton envye inraisonnable Plus c'un serpent (Envye, Estat et Simplese, p. 6, ap. Ler. de Lincy et Michel, Farces, Moral. et Serm. joy., t. I). De quelle source s'agissait-il ? Comment fallait-il la dater ? Grâce à la Bibliographie Godefroy, le texte en question a pu être identifié :

<b>Envye, Estat et Simplese</b>
<i>Envye, Estat et Simplese</i>
1) Exemples :
- Gdf : <i>irraisonnable</i> : <i>Envye, Estat et Simplese</i> , p. 6, ap. Ler. de 24341 et Michel, <i>Farces, Moral. et Serm. joy.</i> , t. I
- autres citations : Gdf : <i>grogneu, oppresse, ragace</i>
2) Bibliographie :
<i>Moralité à troys personnages, c'est assavoir Envye, Estat et Simplese</i>
- <b>Texte : milieu du 16e s.</b>
- Manuscrit : PARIS BN fr 24341 (Recueil La Vallière), f° 50a-53b, ca 1575
Édition(s) : <i>Recueil de farces, moralités et sermons joyeux</i> , éd. Leroux de Lincy et Francisque Michel, Paris, t. 1, 1837, n° 10.
Renvois :
- Bossuat : 5844
- Louis Petit de Julleville, <i>Répertoire du théâtre comique en France au moyen âge : histoire du théâtre en France</i> , Paris, 1886, 62-63.

Une fois la datation du texte *Envye, Estat et Simplese* arrêtée au milieu du 16<sup>e</sup> siècle, on s'aperçoit que nous sommes en présence du dernier témoignage de l'adjectif *irraisonnable*, au moment où le français est en train d'opter de façon définitive pour *irraisonnable* (attesté depuis le 14<sup>e</sup> siècle). Inversement, l'intérêt de cette attestation pour la lexicologie historique (et, au-delà, pour la morphologie constructionnelle historique) demeurerait obscur sans l'aide de la Bibliographie Godefroy.

### 3 Un exemple d'intégration de données lexicographiques : le portail lexical du CNRTL

Le portail lexical du CNRTL a pour vocation de valoriser et de partager, en priorité avec la communauté scientifique, un ensemble de données issues des travaux de recherche sur le lexique français. Projet évolutif, cette base de connaissances lexicales propose, à partir d'une unité lexicale, d'intégrer un maximum de connaissances disponibles.

#### 3.1 Informations lexicographiques

Au premier rang de ces connaissances se placent les données lexicographiques. A ce jour nous avons intégré dans ce portail les informations issues du TLF ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) qui apparaissent par défaut lorsqu'on demande des informations lexicographiques. Elles sont complétées par des informations facilement accessibles, via un menu, et issues :

- du dictionnaire de l'Académie Française (4ème, 8ème et 9ème éditions) ([www.atilf.fr/academie](http://www.atilf.fr/academie)) dont l'informatisation a été réalisée au sein du laboratoire dans le cadre d'un partenariat avec l'Académie,
- de la Base de Données Lexicographiques Panfrancophone (BDLP : <http://www.tlfq.ulaval.ca/bdlp/>),
- De la Base Historique du Vocabulaire Français (Datations et Documents Lexicographiques : [www.atilf.fr/ddl](http://www.atilf.fr/ddl)) constituée de datations du vocabulaire français, s'appuyant sur des données des 48 volumes de la collection Datations et Documents Lexicographiques (Quemada 1970-1998)

L'ensemble de ces informations lexicographiques sont également accessibles directement, pour une forme donnée, par <http://www.cnrtl.fr/lexicographie/suivi> de la forme que l'on souhaite interroger. Ainsi : <http://www.cnrtl.fr/lexicographie/aguerrit> permet d'accéder aux informations lexicographiques disponibles pour le verbe *aguerrir*.

The screenshot shows the CNRTL website interface. At the top, there's a navigation bar with 'Accueil', 'Portail lexical', 'Corpus', 'Lexiques', 'Dictionnaires', 'Outils', and 'Contact'. Below that, a secondary navigation bar includes 'Morphologie', 'Lexicographie', 'Etymologie', 'Synonymie', 'Antonymie', 'Proxémie', 'Concordance', and 'Aide'. The main content area features a search bar with 'aguerrit' entered and a 'Chercher' button. Below the search bar, the entry for 'AGUERRIR, verbe trans.' is displayed. It includes a definition: 'Habituer à la guerre.' and a note: 'A.- Emploi trans. [En parlant de pers.]'. The entry is followed by three numbered points: 1. 'Domaine milit. Aguerrir qqn. L'accoutumer à mener une vie de combattant, avec les fatigues et les dangers qu'elle entraîne.'; 2. 'P. ext., au propre et au fig. Habituer à vivre de façon plus combative, plus dure, plus apte à supporter des épreuves.'; 3. '- Prenez garde, monsieur, au luxe accoutumé, Contre la pauvreté vous êtes désarmé, Et l'assaut des besoins vous sera bien plus rude Qu'aux hommes aguerris par la vieille habitude.'

### 3.2 Informations morphosyntaxiques

Ces informations morphosyntaxiques sont issues de la base Morphalou ([www.atilf.fr/morphalou](http://www.atilf.fr/morphalou)), construite au départ à partir de la nomenclature du TLFi. Elles sont aussi accessibles directement à partir de n'importe quelle forme du paradigme, telle *aguerrit*, par : <http://www.cnrtl.fr/morphologie/aguerrit>

The screenshot shows the Morphalou website interface. At the top, there's a navigation bar with 'Morphologie', 'Lexicographie', 'Etymologie', 'Synonymie', 'Antonymie', 'Proxémie', 'Concordance', and 'Aide'. The main content area features a search bar with 'aguerrit' entered and a 'Chercher' button. Below the search bar, the entry for 'Morphologie du verbe "aguerrir"' is displayed. It includes a table with the following columns: Orthographe, Mode, Temps, Nombre, Personne, and Genre. The table lists the following forms:

Orthographe	Mode	Temps	Nombre	Personne	Genre
aguerrir	infinitif				
aguerris	indicatif	présent	singulier	1 <sup>ère</sup> personne	
aguerris	indicatif	présent	singulier	2 <sup>ème</sup> personne	
<b>aguerrit</b>	<b>indicatif</b>	<b>présent</b>	<b>singulier</b>	<b>3<sup>ème</sup> personne</b>	
aguerrissent	indicatif	présent	pluriel	3 <sup>ème</sup> personne	

### 3.3 Informations étymologiques

Ces informations étymologiques sont issues du TLF ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) et du projet TLF-Etym de mise à jour des rubriques étymologiques du TLF (cf. ci-dessus 2.3.1). Elles sont accessibles directement, pour une forme donnée, par : [www.cnrtl.fr/etymologie/aguerrir](http://www.cnrtl.fr/etymologie/aguerrir)

The screenshot shows the TLF website interface. The 'Etymologie' tab is selected. The search bar contains 'aguerrir'. The results show the verb 'AGUERRIR' with its etymology and history. A pop-up window titled 'Informations' provides details about the source 'De Selve Vies Plutarque'.

**AGUERRIR**, verbe trans.

**Étymol. ET HIST.** – 1535 actif, sens propre « habituer aux périls de la guerre » (G. DE SELVE, *Vies de Plutarque*, 104 v<sup>o</sup>, éd. 1547 ds *R. Hist. litt. Fr.* t. 1, pp. 493-94 : Bonnes gens et bien **aguerriz**); d'où 1665 *id.*, fig. « accoutumer aux choses pénibles » (GRAINDORGE ds *Fr. mod.*, 14, 289 : je prends un plaisir indicible à vous voir **aguerrir** aux pauvres chiens); 1694 pronom., au propre et au fig. « s'endurcir, se faire à » (*Ac.* : ces troupes **se sont aguerries**. Il n'est pas fait au grand monde, il **s'aguerrira**).  
Dér. de *guerre*\*; préf. *a*<sup>1</sup>\*, dés. *-ir*.

**MISE À JOUR DE LA NOTICE ÉTYMOLOGIQUE PAR LE PROGRAMME DE RECHERCHE TLF-ÉTYM :**

**Histoire :**

**A. 1.** « entraîner aux exercices de la guerre et aux exigences du métier des armes ». Attesté depuis 1543 (DE SELVE, *Vies de Plutarque*, *Paul Émile*, page 109 : ilz [les Romains] n'entendoient point que la perte que fait Philippe, avoit fait d'**aguerrir** davantage la puissance [« armée »] des Macedoniens). Pour ce qui est de TLF (datée improprement de 1535), elle concerne l'adjectif *aguerr\** (cf. A. 1.). - Attesté depuis av. 1592 (MONTAIGNE, *Essais* 1595, tome 1, livre I, chapitre XXV, page 143 = nces amollit et effemine les courages, plus qu'il ne les fermit et **aguerrit**). - s'entraîner aux exercices de la guerre et aux exigences du métier des armes ». Attesté *alogue*, page 192 = *Frantext* : ilz **s'aguerrissent**, ilz deviennent meilleurs pour le mestier des s'accoutumer aux difficultés de toute nature, s'endurcir ». Attesté depuis 1580 (MONTAIGNE, *Essais*<sup>1</sup>, tome 1, livre II, chapitre XXVII, pages 693-694 = *Frantext* : Les meurtres des victoires s'exercent ordinairement par le peuple et par les officiers du bagage : et ce qui fait voir tant de cruautéz inouies aux guerres populaires, c'est que cette canaille de vulgaire **s'aguerrit** et se gendarme à s'ensanglanter jusques aux coudes et à deschiqueter un corps à ses pieds,

**Informations**

**De Selve Vies Plutarque**  
DE SELVE, *Vies de Plutarque* = Selve (George de), 1543.  
*Thémistocle/Camille/Périclès/Fabius Maximus/Alcibiade/Coriolan/Timoléon/Paul Émile*, in : *Vies de Plutarque*, édité par Michel de Vascozan, Paris, J. de Tournes.

### 3.4 Synonymies et antonymies

Ces informations de synonymie et d'antonymie proviennent du dictionnaire de synonymes de Caen (<http://www.crisco.unicaen.fr/>), construit à partir de données issues de l'INaLF. Ces informations sont aussi directement accessibles par : [www.cnrtl.fr/synonymie/aguerrir](http://www.cnrtl.fr/synonymie/aguerrir)

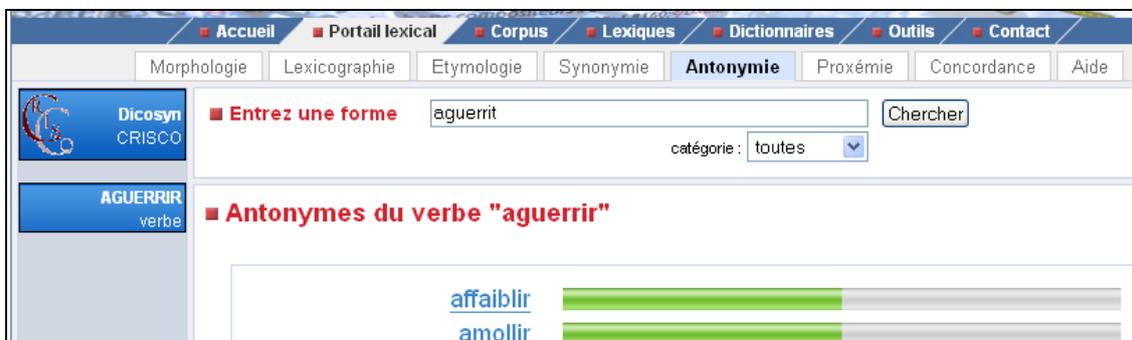
The screenshot shows the Dicosyn CRISCO website interface. The 'Synonymie' tab is selected. The search bar contains 'aguerrir'. The results show a list of synonyms for the verb 'aguerrir'.

**AGUERRI** substantif  
**AGUERRI** adjectif  
**AGUERRIR** verbe

**Synonymes du verbe "aguerrir"**

- endurcir
- accoutumer
- entraîner
- exercer
- rompre
- tremper
- préparer
- affermir
- cuirasser

ou [www.cnrtl.fr/antonymie/aguerrit](http://www.cnrtl.fr/antonymie/aguerrit)



Accueil Portail lexical Corpus Lexiques Dictionnaires Outils Contact

Morphologie Lexicographie Etymologie Synonymie **Antonymie** Proxémie Concordance Aide

■ Entrez une forme  Chercher

catégorie : toutes

■ Antonymes du verbe "aguerrir"

[affaiblir](#) 

[amollir](#) 

### 3.5 Concordance

Cette concordance utilise le corpus des textes libres de droit de la base Frantext ([www.atilf.fr/frantext](http://www.atilf.fr/frantext)) ; elle offre aussi la possibilité d'exporter les résultats du concordancier au format XML/TEI. C'est à notre connaissance le seul site permettant à un utilisateur d'exporter dans un format normalisé un concordancier français d'une telle importance. Ces concordances sont aussi directement accessibles par : <http://www.cnrtl.fr/concordance/aguerrir>



Accueil Portail lexical Corpus Lexiques Dictionnaires Outils Contact

Morphologie Lexicographie Etymologie Synonymie Antonymie Proxémie **Concordance** Aide

■ Entrez une forme  Chercher

■ Concordances de "aguerrir"

Résultats de 0 à 14 sur 14

jours comme très prochaine. Mais je m'y suis [aguerrir](#), et je ne souhaite aux orthodoxes qu'une pai  
s galeries et n'[apercevait](#) pas "Bineau. Tout [aguerrir](#) qu'il était, le chef d'orchestre se troubla,  
ni co [chercher 'apercevait' en](#) regard [aguerrir](#) dont l'orgueil est coulé dans un moule de fo  
tes e [aguerrir](#) moins que la première. J'en prends mon parti : ava  
t ap [aguerrir](#) un regard à la fois digne et caressant, il p  
con [aguerrir](#) avaient [aguerrir](#) contre toute espèce de dangers, et le rendai  
, sobre sa [aguerrir](#) nement [aguerrir](#), doué de clartés et d'accents inconnus, maît  
aillis [aguerrir](#) un bras [aguerrir](#). Et l'arc ploie et se brise avec un bruit te  
ail [aguerrir](#) s'était [aguerrir](#) vaillamment au métier des armes, et que ses  
mérite. un jeune missionnaire, non encore [aguerrir](#) [aguerrir](#) contre les dangers, comme ces vieux chefs to  
st un jeune avocat d'un tempérament robuste, [aguerrir](#) par la lecture des pièces dans les procédu  
ficile d'avoir tout à la fois un peuple très-[aguerrir](#) dans les combats, et très-soumis dans l'ence  
languissante pour exprimer ses maux ! L'oeil [aguerrir](#) des chirurgiens se ferme d'horreur ; leur ma  
'avoient agité la veille. N'étant pas encore [aguerrir](#) contre les incertitudes du savoir, ma peur a

Navigation

Export XML-TEI

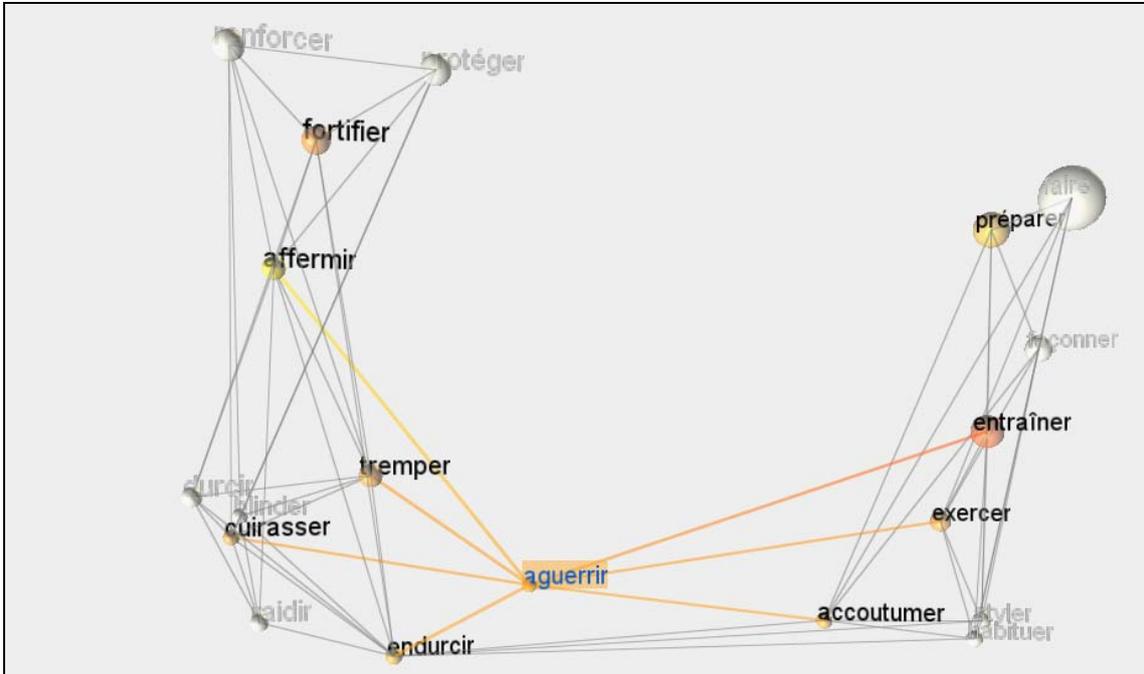
- [morphologie](#)
- [lexicographie](#)
- [etymologie](#)
- [synonymie](#)
- [antonymie](#)
- [proxémie](#)
- [concordance](#)

Le portail lexical permet également, à partir d'un double clic sur un mot-forme, une hyper-navigation vers toutes les informations lexicales disponibles pour l'unité lexicale concernée. Par exemple, si l'on veut obtenir des informations sur la forme *apercevait* du deuxième exemple de concordance, un double clic sur le mot-forme affiche un menu qui permet d'hyper-naviguer vers l'ensemble des informations disponibles pour cette forme et pour le verbe *apercevoir* :

De plus, un simple clic droit sur un des exemples permet d'obtenir la référence complète de l'exemple sélectionné.

### 3.6 Proxémie

Une représentation en trois dimensions de la proxémie (représentation des proximités sémantiques) des mots de la langue réalisée en coopération entre l'IRIT et l'ERSS (<http://Prox.irit.fr>) (Gaume 2006) est accessible elle aussi directement par : <http://www.cnrtl.fr/proxemie/aguerrir>.



## 4 Conclusion

Le partage et la mutualisation de résultats de recherche et de ressources informatisées sur le lexique de notre langue ouvrent des perspectives intéressantes. Ainsi le TLF, une référence en lexicographie française, a eu pendant longtemps la réputation tenace d'être un dictionnaire réservé à une élite et la diffusion de sa version papier s'est limitée à quelques milliers d'exemplaires au sein d'une intelligentsia somme toute limitée. Sa version informatique sous forme de Cédérom et de ressources librement accessibles sur le Web a rencontré un succès important tant auprès du grand public que des utilisateurs universitaires ou des professionnels de la langue. Sa version Web fait l'objet de plus de 250 000 connexions quotidiennes en provenance de tous les continents, et il est référencé par d'innombrables sources. La notoriété qu'il a acquise en fait aujourd'hui un dictionnaire incontournable et un outil de promotion appréciable de la langue française.

L'intégration plus récente au sein du portail lexical du CNRTL de diverses ressources sur le lexique français permet une meilleure mutualisation des résultats de la recherche. De plus l'outil informatique, à travers le Web, offre aujourd'hui des modes nouveaux de valorisation de ressources ou de résultats de recherche. Au-delà du seul monde universitaire, ces techniques permettent de mettre à disposition de l'ensemble de la société nos résultats de recherche. Aujourd'hui le portail lexical du CNRTL fait l'objet d'environ 300 000 requêtes par jour provenant d'horizons très divers, il est aujourd'hui l'un des sites Web sur le lexique français les plus utilisés (Cf. ci-après un extrait des statistiques d'utilisation du portail lexical).

Notons enfin que la généralisation de telles exploitations et valorisations de versions électroniques est ainsi en train de modifier notablement les modes de travail et d'échanges scientifiques au sein des communautés de recherche SHS, permettant un véritable travail en réseau comme s'est le cas en particulier aujourd'hui pour la communauté en lexicographie historique.



## Références bibliographiques

- ATILF (2004-2005) *Trésor de la Langue Française informatisé*, CNRS Editions, 591 p. et CD du texte intégral, Version PC, ISBN 2-271-06273-X, 2004, Version Mac OX X, ISBN 2-271- 06365-5, 2005.
- Apothéloz, D. (2003) Le rôle de l'iconicité constructionnelle dans le fonctionnement du préfixe négatif *in-*, *Cahiers de linguistique analogique* 1, p. 35-63.
- Bouchou, B. et Maurel, D. (à paraître) Prolexbase et LMF: vers un standard pour les ressources lexicales sur les noms propres, à paraître dans *Traitement Automatique des Langues*, Vol 49 : 1, Varia
- Buchi, É. (2005) Le projet TLF-Étym (projet de révision sélective des notices étymologiques du *Trésor de la langue française informatisé*), *Estudis romànics* 27, p. 569-571.
- Buchi, É. (à paraître) *Réel, irréel, inréel* : depuis quand le français connaît-il deux préfixes négatifs *IN-*?, in : Brun-Trigaud G. et al. (éd.), *Actes du colloque GalRom07. Diachronie du gallo-roman. Évolution de la phonologie et de la morphologie du français, du francoprovençal et de l'occitan (Nice, 15/16 janvier 2007)*.
- Buchi, É. et Städtler, Th. (2008) La pragmatization de l'adverbe *enfin* du point de vue des romanistes (« Enfin, de celui des francisants qui conçoivent leur recherche dans le cadre de la linguistique romane »), Congrès Mondial de Linguistique Française 2008.
- Chauveau, J.-P. (2006) FEW : article BAJULA, Nancy/ATILF, (<http://www.atilf.fr/few/bajula.pdf>), 7 p., 2006.
- DEAF = Baldinger, K. et al., (1974- ) *Dictionnaire Étymologique de l'Ancien Français*, Québec/Tübingen/Paris, Presses de l'Université Laval/Niemeyer/Klincksieck.
- Dendien, J. et Pierrel, J.M. (2003) Le Trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence, *TAL* Vol 44 – n° 2, Hermès Sciences Edition, p. 11-37.

- DMF<sup>0</sup> = Martin R. (dir.), *Dictionnaire du moyen français (DMF) : 1330–1500 : A-AH*, Nancy, INaLF, 1998.
- DMF = *Dictionnaire du Moyen Français*. Nancy, ATILF/Nancy-Université & CNRS, <http://www.atilf.fr/dmf>.
- FEW = Wartburg W. von *et al.*, *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes* (25 vol.), Bonn/Heidelberg/Leipzig-Berlin/Bâle, Klopp/Winter/Teubner/Zbinden, 1922–2002.
- Gaume, B. (2006) Cartographier la forme du sens dans les petits mondes lexicaux, in *JADT 2006*, p 541-465.
- Gerner, H. (2005) *La Base des Lexiques du Moyen Français (BLMF) : Première étape vers le DMF électronique*, in : Kabatek J./Pusch Claus D./Raible W. (éd.), *Romanistische Korpuslinguistik II : Korpora und diachrone Sprachwissenschaft*, Tübingen, Narr, 2005, p. 155-162.
- Gerner, H. (2007a) *Le Lexique Complémentaire : deuxième étape vers le DMF électronique*, in : Trotter D. (éd.), *Actes du XXIV<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes*, Tübingen, Niemeyer, vol. IV, p. 69-77.
- Gerner, H. (2007b) Constitution et évolution des corpus textuels et lexicaux à l'ATILF. Interconnexion des ressources, in : Kunstmann P./Stein A. (éd.), *Le Nouveau Corpus d'Amsterdam. Actes de l'atelier de Lauterbad, 23-26 février 2006*, Stuttgart, Steiner, p. 101-109.
- Godefroy, F. (1881–1902) *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX<sup>e</sup> au XV<sup>e</sup> siècle* (10 vol.), Paris, Vieweg.
- Habert, B., Nazarenko, A. et Salem, A. (1997) *Les linguistiques de corpus*, Paris, Armand Colin.
- Hansen, M.-B. (2005) Mosegaard, From prepositional phrase to hesitation marker. The semantic and pragmatic evolution of French *enfin*, *Journal of Historical Pragmatics* 6, 37-68.
- Huguet, E. (1925–1967) *Dictionnaire de la langue française du seizième siècle* (7 vol.), Paris, Champion/Didier.
- Imbs, P., Quemada, B. (dir.) (1971–1994) *Trésor de la Langue Française. Dictionnaire de la langue du XIXe et du XXe siècle (1789–1960)*, 16 vol., Paris, Éditions du CNRS/Gallimard.
- Jonasson, K. (1994) *Le nom propre. Constructions et interprétations*, Duculot, Louvain-La-Neuve
- LEI = Pfister M./Schweickard W. (dir.), *Lessico etimologico italiano*, Wiesbaden, Reichert, 1979–.
- Martin, R. (2008) Perspectives de la lexicographie informatisée, Contribution à la table ronde « Lexique », 1<sup>er</sup> Congrès Mondial de Linguistique Française, Paris.
- Martin R., *Dictionnaire du Moyen Français (DMF) (1330–1500). Seconde version : DMF2. Présentation in : DMF* (<http://atilf.atilf.fr/gsovay/dmf2/PrésentationDMF2.pdf>), 2007.
- Martin, R., Gerner, H., Souvay (à paraître) G., Présentation de la seconde version du DMF (*Dictionnaire du Moyen Français*), in : Iliescu M. *et al.* (éd.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique Romane (Innsbruck, 3-8 septembre 2007)*, Tübingen, Niemeyer.
- Pascual Rodríguez, J. A. et García Pérez, R. (2007) *Límites y horizontes en un diccionario histórico*, Salamanca, Diputación de Salamanca.
- Pierrel, J.M. et Petitjean, E. (2007) Valorisation et exploitation scientifiques de documents numériques pour la recherche en linguistique : l'exemple du CNRTL, *Actes de CIDE 2007 Congrès International sur le Document Numérique*, Nancy, 2-4 juillet 2007, p13-24, Europa, ISBN 978-2-909285-38-2
- Pierrel, J.M. et Buchi, E. (à paraître) Research and Resource Enhancement in French Lexicography: the ATILF Laboratory's computerized resources, à paraître in "Lexicography in Italy and in Europe", Silvia Bruti, Roberta Cella and Marina Foschi, editors, Cambridge Scholars Publishing (39 pp.)
- Quemada, B. (dir.) (1970-1998) *Matériaux pour l'histoire du vocabulaire français. Datations et documents lexicographiques*, 2<sup>e</sup> série, 48 vol., Besançon/Paris, Centre d'Étude du Vocabulaire Français/Didier/Klincksieck.
- Ringebach, J.-L. (2003) Bibliographie des sources de Frédéric Godefroy, in : Duval F. (éd.), *Frédéric Godefroy. Actes du X<sup>e</sup> colloque international sur le moyen français organisé à Metz du 12 au 14 juin 2002 par le Centre « Michel Baude, littérature et spiritualité » et par l'ATILF (UMR 7118)*, Paris, École des Chartes, 191-206.
- Ringebach, J.-L. (à paraître) La Bibliographie du Dictionnaire de Godefroy, in : Iliescu M. *et al.* (éd.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique Romane (Innsbruck, 3-8 septembre 2007)*, Tübingen, Niemeyer.

Romary, L., Salmon-Alt, S., Francopoulo, G. (2004) Standard going concrete: from LMF to Morphalou. Actes de la conférence COLING-2004/Genève.

Steinfeld, N. (à paraître) Étymologie-origine, étymologie-histoire et déonomastiques : le cas de *poubelle* (avec une annonce de la création de la Base des Mots-fantômes), in : Iliescu M. *et al.* (éd.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique Romane (Innsbruck, 3-8 septembre 2007)*, Tübingen, Niemeyer.

Tran, M., Maurel, D. (2006) « Prolexbase : Un dictionnaire relationnel multilingue de noms propres », *TAL*, vol. 47, n° 3, p. 115-139.

Tobler-Lommatzsch = Tobler, A./Lommatzsch, E./Christmann, H. H. (1925–2002) *Altfranzösisches Wörterbuch* (11 vol.), Berlin/Wiesbaden/Stuttgart, Weidmann/Steiner.

Union Latine (2008), Langue et cultures sur la toile : enquête 2007, [http://dtil.unilat.org/LI/2007/index\\_fr.htm](http://dtil.unilat.org/LI/2007/index_fr.htm)

---

<sup>1</sup> DTD : La Document Type Definition (DTD), ou Définition de Type de Document, est un document permettant de décrire un modèle de document XML.

<sup>2</sup> Pour le plus grand bonheur de l'œuvre, Robert Martin et Hiltrud Gerner continuent à s'impliquer de façon notable dans le travail de rédaction.

<sup>3</sup> La continuité de la tradition est assurée par une attestation intermédiaire, de 1452, que le DMF2 propose s.v. tare.