



**HAL**  
open science

## Research and Resource Enhancement in French Lexicography: the ATILF Laboratory's computerized resources

Eva Buchi, Jean-Marie Pierrel

► **To cite this version:**

Eva Buchi, Jean-Marie Pierrel. Research and Resource Enhancement in French Lexicography: the ATILF Laboratory's computerized resources. *Lexicography in Italy and in Europe* / Silvia Bruti, Roberta Cella and Marina Foschi (eds), Cambridge Scholars Publishing, pp.39, 2008. halshs-00258126

**HAL Id: halshs-00258126**

**<https://shs.hal.science/halshs-00258126v1>**

Submitted on 21 Feb 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.

## CHAPITRE XXXX

# Research and Resource Enhancement in French Lexicography: the ATILF Laboratory's computerized resources

Pierrel, J.-M. & Buchi, E.

ATILF – Nancy-Université & CNRS  
Analyse et Traitement Informatique de la Langue Française  
44, avenue de la Libération BP 30687 F-54 063 Nancy cedex  
{Jean-Marie.Pierrel, Eva.Buchi}@atilf.fr

## 1 Introduction

During the second half of the 20th century, a number of major contributions to French lexicography have been developed in Nancy, first within the Centre de Recherche pour un Trésor de la Langue Française (CRTLF), then within the Institut National de la Langue Française (INaLF). The present research institute, Analyse et Traitement Informatique de la Langue Française (ATILF, [www.atilf.fr](http://www.atilf.fr)), aspires to become the latter's rightful successor.

Initiated as a result of the *Trésor de la Langue Française* project (TLF) – the main lines of which we return to in paragraph 2.1 – the Nancy work on French lexicography has now gone further. Beyond the compilation of the TLF, research has gone in two additional directions: historical lexicography and the computerized enhancement of existing resources. In the following pages, we assess the results of this work by presenting, in order:

- The *Trésor de la Langue Française Informatisé* (TLFi), which is the computerized enhancement (on a CD-Rom and on the Web) of the authoritative TLF dictionary (see paragraph 20). But it goes far beyond the simple computerization of the TLF's printed version. This project has laid the foundations for a new way of exploring lexicographical data, giving rise to genuinely innovative modes of interpretation. It has injected new life into this

great dictionary, now undoubtedly the most consulted institutional French dictionary on the Web.

- The *Dictionnaire du Moyen Français*, or *DMF*, decisively links computerized development with historical lexicography (see paragraph 3). It aims at filling the gap left between Tobler-Lommatzsch (for the earlier period) and Huguët (for classical French). It describes French vocabulary between 1330 and 1500, and is based on a concept of *evolving lexicography* (Martin 2007:1), made possible nowadays by computerization.
- Other recent developments in historical lexicography (see paragraph 4), more precisely the *TLF-Étym* programme, which aims at progressively revising the TLF’s “Etymology and history” sections; the *Base des Mots-fantômes*, a critical metalexicographical project seeking to list the *ghost words* or *fantasmas lexicográficos* (Pascual Rodríguez/García Pérez 2007:170), i.e. all the pseudo-lexemes with an erroneous lexicographical status (“the words that do not exist”), the ghost meanings and the wrong lemmatizations found in French historical and etymological reference dictionaries; finally, the *Bibliographie Godefroy*, which aims to identify the obscure sigla within the immense documentary bulk of the Godefroy dictionary. It does so either by cross-referring to more explicit abbreviations, or by philological work on the publications – and even the manuscripts, though more rarely – which include the different texts concerned.
- Finally, the *Portail Lexical* of the Centre National de Ressources Textuelles et Lexicales (CNRTL), created two years ago by the CNRS within our institute (see paragraph 5). This lexical portal seeks to enhance and share (in the first instance, with the scientific community) a collection of data gathered by researchers working on French lexis. As an ongoing project, this lexical database seeks to offer as much information as is available for each individual lexeme.

As will become clear through our presentation of each element, these varied contributions to French lexicography are based on common objectives, and depend centrally on our concern to enhance and to make available (thanks to computerization and the availability of information on the Web) the collection of lexicographical and historical data on French lexis which has been assembled and produced by our institute, and, more generally as far as the CNRTL is concerned, by the French language community.

## **2 From the *Trésor de la Langue Française* to the *Trésor de la Langue Française Informatisé***

### **2.1 The *Trésor de la Langue Française***

The *Trésor de la Langue Française*, a nineteenth- and twentieth-century language dictionary (Imbs et Quemada 1976-1994; cf. Radermacher 2004), is the result of a great lexicographical adventure which brought together more than one hundred contributors over thirty years, first under the direction of Paul Imbs, Recteur of the Académie of Nancy, then under that of Professeur Bernard Quemada.

Thus it was that by legislation enacted on 20th December, 1960, a research institute was created in Nancy with the aim of establishing the documentation, editorial process, and publication of a *Trésor de la Langue Française*. This project had been prepared and the decision had been taken, three years before, by an international conference on French and Romance lexicology and lexicography, organized by the head of the Centre de Philologie Romane de Strasbourg, Paul Imbs (cf. Imbs 1961: 285-289). Littré's dictionary (1863-1873) being then free of copyright restrictions, an interesting if controversial discussion ensued: one option was to republish that famous dictionary – because of its status of “monument” belonging to the science of its time – but the counterview prevailed, that the time had come to start building something entirely new. This new project would have to take into account the findings of 20th-century lexicology and lexicography, new possibilities regarding available documentation, and of course the changes which had arisen in the French language since the mid-19th century. The findings of the Strasbourg colloquium were very clear on the last point: “*Instrument de travail, le Trésor poursuivrait donc un double but: être le témoin objectif et impartial du vocabulaire français, mieux connu parce que mieux inventorié; être ce qu’avait été le Littré pour son temps: un exemple-type de lexicographie scientifique moderne*” [“The TLF, as a tool, has two aims: to offer an objective and impartial survey of the vocabulary of French, which will be better known if it is better indexed, and to be what Littré was in its era: an example of the best of modern scientific lexicography”]

If, today, one were to resituate the *Trésor* in relation to the demands of lexicography, it would still be appropriate to reiterate its initial aims as defined by Paul Imbs. Hence, the *Trésor* was to be:

- A dictionary of the French-speaking world. France had indeed to catch up with other countries in this matter. Twenty-five years earlier, England had compiled

its *New English Dictionary* (Oxford Dictionary), and other Latin, Germanic or Slavic countries had already been working on the publication of a national dictionary.

- A historical dictionary. The Trésor will not limit itself to only the current usage of words, but it will include, for each word, an “etymological and historical” section, fully reflecting current knowledge in that field.
- A linguistic dictionary, or dictionary of the language. As opposed to other dictionaries with an encyclopaedic perspective, the Trésor will try and define each word by its linguistic characteristics: its form, its meaning, and its stylistic and syntactic usages.
- A dictionary that has been the work of an entire generation. The creation of a research centre for a Trésor de la Langue Française coincided with the beginning of the use of mecanographic and computerized tools of documentation in the field of humanities. Hence, from 1964 onwards, thanks to powerful computer facilities, more than 1000 literary works were exhaustively combed. This enabled the editors to draw on an immensely rich collection of examples of usage (430 000 examples), which gave birth to what is undoubtedly the biggest textual database for a specific language: Frantext is regularly updated and nowadays covers more than 4000 works at the ATILF (around 80% French literary works and 20% technical works).

The *Trésor de la Langue Française* is the first dictionary to use a systematic methodology to analyze the real usages of our language’s words, through the use of a vast textual database. The compilation of that database had started in the 1960s, and its chief aim was to provide the TLF dictionary’s publishers with properly-organized data. Thus, when an editor had to write an article, he was provided with the systematic concordanced occurrences of that word, sorted by different criteria: the sources arranged in chronological order, equipped with left and right contexts in alphabetical order, in documents focusing on co-occurrences (“binary groups”, cf Gorcy et Col. 1970), but also syntactic constructions belonging to each discourse component in a specific order. Those concordances were used in order to effect a preliminary classification of the available documentation, and they then enabled the editors to obtain enlarged contexts, from which the examples which were to appear in the dictionary were selected.

## **2.2    *Le Trésor de la Langue Française Informatisé, a computerized version of the TLF***

### **2.2.1    What are the characteristics of the TLFi?**

The TLFi ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) remains a faithful image of the printed version of the TLF, even in its on-screen typography. Like the TLF, it is characterized by its rich and varied material, and by its complex structure:

- Importance of the nomenclature: 100,000 words with their etymologies and their history, along with 270,000 definitions.
- Variety of meta-textual objects included in each article (headwords, grammatical information, semantic or stylistic markers, field markers, definitions, documented examples...)
- Wealth of examples: 430,000 quotations from French (mainly literary) works from the last two centuries.
- Diversity of sub-sections: a section dedicated to synchronic semantic analysis (from 1789 onwards), a “pronunciation and spelling” section, an “etymology and history” section, a section devoted to lexical statistics and a bibliographical section for the main articles.

Moreover, the computerized version of the TLF (Dendien and Pierrel, 2003) includes access with a high level of sophistication (ability to disregard accents, tolerance of common spelling mistakes, phonetic and morphological treatment). Thus, spelling mistakes are automatically corrected, and access to articles is possible by typing a form and not necessarily a lemma or a headword; hence, a range of means of access is available.

We will not go back over the different phases of the computerization of the TLF, already discussed elsewhere (Dendien and Pierrel 2003), but will limit ourselves to a review (via examples) of different forms of access which the computerized version of the TLF provides.

### **2.2.2    Access to the TLFi**

The TLFi is a retro-conversion of the TLF’s printed version. Thanks to processes involving semi-automatic identification of the textual formats of the dictionary’s articles, detailed tagging has been introduced, both typographical (so we could keep a 100% faithful image of the TLF) and semantic (identification of the main textual forms within each article). Some figures regarding the precision of

the tagging process: after validation of all sixteen volumes, 36,613,712 XML tags were in place, i.e.: 17,364,854 typographical tags, 1,070,224 hierarchical tags, 18,178,634 tags tracking textual objects, among which were 92,997 entries and 64,346 phrases accompanied by their 271,166 definitions and illustrated by 427,493 examples.

The TLF's detailed tagging and the using of the corresponding XML document has enabled us to provide access to the whole dictionary. It thus combines all the advantages of a dictionary with those of a textual resource and a proper lexical database:

- Word, phrase or lexical form search, more or less correctly spelled, along with the possibility, thanks to a “control panel”, of highlighting various fields in the search result (definitions, grammatical category, notional domain, example, author of an example, construction, marker, etc.).
- Possibility of hyper-navigation within the dictionary, enabling the user to move from one word of the meta-language to its lexicographical description (and more importantly its definition) in one click.
- Assisted searches or complex requests, exploiting the whole of the dictionary's structure through the combined application of various criteria.

### 2.2.3 Examples of searches in TLFi

The Web site [www.tlfi.fr](http://www.tlfi.fr) offers an introduction to the TLFi and displays its different search modes; yet, the best way to appreciate the effectiveness of the computerization of the TLF is either to use the TLFi CD-Rom (ATILF, 2004), or to go directly to the address : [www.atilf.fr/tlfi](http://www.atilf.fr/tlfi).

Three main types of access to articles are proposed: word search, assisted search and complex search.

#### Word search

This search enables the user to access a word through a system of correction and automatic lemmatization (artificial or not): hence, a search for the word *etique* (without the accent), leads to the two corresponding articles *étique* ou *éthique*; likewise, typing the form *sussiez* enables the user to be automatically redirected to the article *savoir*. It also offers the possibility of directly getting the definitions and the conditions of use:

- A lexical unit which is not the object of an independent lexicographical treatment (for instance, the masculine substantive *trompette* is found via the

- request “le trompette”, in a super-article *trompette* including both the masculine and the feminine).
- A phrase such as *battre la mesure*, focusing on the relevant element that was requested, and offering the possibility of highlighting a given textual object, thanks to a sort of electronic “coloured highlighter”. Thus, for instance, the following definition:

Objets de la recherche : 1 ¶ Paragraphe 1

**H** TROMPETTE, subst.

---

**II.** — *Subst. masc.* Personne qui joue de la trompette.

**1 A.** — Soldat chargé d'exécuter les sonneries. *Le trompette* de l'escadron, d'un régiment de cavalerie. Tu seras capitaine, avec une nuée de trompettes courant et sonnante devant toi (HUGO, *Légende*, t. 3, 1877, p. 390). 1

— *Loc. fam., vieilli.* Il est bon cheval de trompette. Il ne se laisse ni effrayer, ni intimider. Son air, un air de bon cheval de trompette qui ne craignait pas le bruit (A. DAUDET, *Tartarin de T.*, 1872, p. 13).

**B.** — Musicien jouant dans une fanfare, un orchestre. *Synon. trompettiste (infra dér.). Le trompette noir du dancing* (BEAUVOIR, *Mandarins*, 1954, p. 306).

### Assisted search

The second type of search offers, for instance, a means of establishing a list of all the compounds including a given element; thus, an enquiry which looks up compounds including the lexeme *queue*, generates 35 results, including:

COURTE-QUEUE, adj. et subst.
DEMI-QUEUE, subst. fém.
HOCHEQUEUE, HOCHE-QUEUE, subst. masc.
PAILLE-EN-CUL, PAILLE-EN-QUEUE, subst. masc.
PORTE-QUEUE, subst. masc.
QUEUE(-)D'ARONDE, voir ARONDE.
Etc.

The assisted search also enables the user to look up “verbs which in nautical usage, pertain to the handling of sails”. The user needs only to specify a search, in the category of verbs, for those verbs which belong to the notional domain of “nautical activity” and which correspond to a definition that includes an



inflectional form (singular or plural) of the word *voile*, i.e. more concisely formulated: [grammatical code: *verbe* ; domain: *sailing* ; object type: *definition*, content : &msail<sup>1</sup>]. Below are the 61 results thereby generated:

ABRIER, ABREYER, verbe trans.
3 Empêcher le vent, en l'interceptant, de passer jusqu'à (une autre voile) : 3
AGRÉER <sup>2</sup> , verbe trans.
3, Préparer ou travailler à la garniture, aux agrès d'un bâtiment, fourrer les dormans, estroper les poulies, garnir voiles, vergues, etc. : `` (WILL. 1831) : 3
AMURER, verbe.
3 Fixer l'amure d'une voile pour l'orienter selon le vent : 3
ETC.....

To take another example: for all the words whose definition includes the noun *liberté* [object type: *definition*; content: &mliberté ], there are 306 results, including:

Objets de la recherche : 1 Définition 1

ABUSER, verbe trans.
1 Exagérer dans l'usage d'une possibilité, d'une liberté : 1
AFFRANCHI, IE, part. passé, adj. et subst.
1 (Celui) à qui on a donné la liberté. : 1
AISE <sup>1</sup> , subst. fém.
1 Grande liberté. : 1
ALIÉNANT, ANTE, part. prés. et adj.
1 Qui prive l'homme de son humanité, de sa liberté : 1
Etc.....

### Complex search

Consulting the dictionary can entail even more complex processes. Thus, it is possible to launch the following request: “What nouns are borrowed from an

<sup>1</sup> &msubs enables us to try all the forms of a substantive, just as &cverbe enables us to try all the forms of a verb.

(undefined) foreign language and used in the terminology of cooking?”. This involves simply using the “complex search” rubric and specifying:

- Object 1 : type “*Entrée*”;
- Object 2 : type “*Code grammatical*”, content “*substantif*”, relationship “*inclus dans l'objet 1*”;
- Object 3 : type “*Domaine technique*”, content “*art culinaire*”, relationship “*dépendant de l'objet 1*”;
- Object 4 : type “*Langue empruntée*”, relationship “*dépendant de l'objet 1*”.

The relationship “inclus dans l’objet 1” of “Object 2” means that the entry is a substantive; the relationship “dépendant de l’objet 1” of “Object 3” suggests that the technical domain marker is within the scope of “Object 1”; and the relationship “dépendant de l’objet 1” of “Object 4” means that the object is in the article whose entry is “Object 1”.

Such a request gives us 42 results, including:

Objets de la recherche : 1. **Entrée** 2. **Code grammatical** 3. **Domaine technique** 4. **Langue empruntée**

	<b>BOR(T)SCH, subst. masc.</b>
<a href="#">1</a> <sup>4</sup> Empr. au russe <sup>4</sup>	
	<b>CARAMEL, subst. masc.</b>
<a href="#">2</a> <sup>4</sup> Empr. à l'esp. <sup>4</sup>	
	<b>CAVIAR, subst. masc.</b>
<a href="#">3</a> <sup>4</sup> Empr. au vénitien <sup>4</sup>	
	<b>CONDIMENT, subst. masc.</b>
<a href="#">4</a> <sup>4</sup> Empr. au lat. class. <sup>4</sup>	
	<b>ESSENCE<sup>3</sup>, subst. fém.</b>
<a href="#">5</a> <sup>4</sup> Empr. au lat. class. <sup>4</sup>	
	<b>ESTOUFFADE, subst. fém.</b>
<a href="#">6</a> <sup>4</sup> Empr. à l'ital. <sup>4</sup>	
	<b>GANACHE, subst. fém.</b>
<a href="#">7</a> <sup>4</sup> Empr. à l'ital. <sup>4</sup>	

### **3 Dictionnaire du Moyen Français (DMF)**

#### **3.1 Methodological characteristics**

##### **3.1.1 A new concept: evolving lexicography**

The idea of a *Dictionnaire du Moyen Français* (DMF) was first mooted in 1980 on the occasion of the 3<sup>rd</sup> International Conference on Middle French (cf. Wunderli 1982). It was developed by Robert Martin, who directed it from 1982 (the beginning of the preliminary work) until 2000. The DMF was then overseen by Bernard Combettes (2000-2002) and Hiltrud Gerner (2003-2007), and continues to develop under the direction of Sylvie Bazin-Tacchella (from 2008)<sup>2</sup>.

After the publication, in 1998, of an experimental “pre-publication” volume covering the alphabetical section A-AH (DMF0), the project took on another form, although it kept the same goal. The hardcopy version was – at least provisionally – dropped in favour of an electronic publication that would be produced in successive phases. It takes the form of an evolving lexical database encoded in XML format (cf. <http://www.atilf.fr/dmf>). The architect of the DMF has thus described the advantages of this solution:

“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. 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.” (Martin 2007 : 1). [“The central idea behind the DMF project is that computerization henceforth permits ‘evolving lexicography’: it is no longer necessary to edit a dictionary letter by letter, which means that it is not

---

<sup>2</sup> Greatly to the benefit of the DMF, Robert Martin and Hiltrud Gerner are still significantly involved in the editorial process.

finished until the end of the alphabet is reached, but it becomes possible to proceed instead in a series of phases, each of which can be completed whilst still allowing for further development. The ease with which, thanks to computerization, data can be expanded, corrected, and restructured is bound to impinge on lexicographical methodology. The system chosen for the DMF is based on the idea that modern scholarly (rather than commercially-produced) dictionaries, should not be finished products which can only be altered via putative revised editions which are inevitably expensive and which themselves are then fixed for a long period of time, but should instead be easily accessible computerized databases which can be subsequently enriched and improved as we wish, without the revisions costing very much”]

It is thus necessary to distinguish DMF1, available on the internet since 2003, from DMF2 (the current online version) and DMF3, still a work in progress (to say nothing of future versions and developments). DMF1 (cf. Gerner 2005) assembled thirteen lexica, each of them analyzing the vocabulary of a defined corpus. Those lexica cover either a representative text of the time (such as that which Jean-Loup Ringenbach produced for the *Passion d’Auvergne*), or again, the whole of a key Middle French author (for instance, Andrieu de la Vigne, available thanks to Annie Bertin); the lexica can, equally, cover the vocabulary pertaining to one specific genre (such as the “Lexicon of didactic texts” by Hiltrud Gerner). Preliminary lemmatization ensures that all thirteen lexica can be simultaneously consulted (32,779 lemmas are dealt with in 84,778 articles: one lemma is thus dealt with in an average of two to three different lexica).

DMF2, the currently-available online version (cf. Gerner 2007a; Martin/Gerner/Souvay, forthcoming) gathers together seventeen lexica. Among them, of particular importance is the “Lexique complémentaire”, which is based on a very large corpus consisting – among others – of texts digitized at the ATILF, of gleanings from published editions, and of dictionaries of medieval French as well as historical dictionaries. This “Lexique complémentaire” (in itself containing 35,000 lemmas), deals mainly with rare lexical units (missing from the thirteen lexica of the DMF1): hapaxes, authorial inventions, one-time Latinisms, ghost-words (see below, 5.2) that are identified as such. But it also contains lexemes that are still in use in contemporary French, such as *cécité* or *palissade* (Gerner 2007a: 71). The nomenclature of DMF2 is thus almost twice the size of the DMF1, with the number of lemmas increasing to 60,241 (discussed in 117,723 articles, which corresponds to an average of fewer than two articles per lemma).

DMF3 is deliberately intended to fit in with the project’s own internal coherence and planning: its additional contribution will chiefly be in the provision

of a significant number of synthetic meta-articles, established on the basis of several articles devoted to one single lemma.

### 3.1.2 Larger-scale objectives

Scholars are unable to agree on the chronological boundaries of Middle French. Those that were fixed for the DMF (1330-1500) are justified by reasons which are both historical (the accession of the Valois to the throne in 1328; the beginning of the Italian wars in 1497) and linguistic (such as the appearance of a proclitic subject pronoun, compulsory in the main clause, cf. Posner 1997: 378-388). But they are also justified by the complementary nature of the DMF and Tobler-Lommatzsch for the earlier period, and Huguet in the aftermath Middle French. Although a variationist perspective is taken into account by the DMF (see for instance its careful approach to diatopia), the period under scrutiny in the dictionary is treated as a broad synchronic sweep, similar to that of the TLFi (and not as a narrow diachronic selection).

The nomenclature of the DMF takes into account all the lexical units (nouns, adjectives, verbs, adverbs, interjections and numerals) of the corpus, including the hapaxes. For the time being, it leaves out grammatical units (DMF2). In order to maintain some semblance of control over the immense formal variety presented by the vocabulary of a non-standardized period of a language such as that of Middle French, Gilles Souvay devised for the DMF the LGeRM lemmatizer (Lemmes, Graphies lemmatisées et Règles Morphologiques; cf. Souvay 2007). This instrument<sup>3</sup> connects each word (or noun-form) from a Middle French text (and specifically, from a quotation in the DMF) to the lexeme of which it is a graphical or morphological variant. This enables the DMF user to access the relevant lemma from any inflectional or graphical variant, and to navigate (by double-clicking) inside the dictionary, including between the Frantext database and the DMF. The LGeRM lemmatizer “*opère avec une probabilité de réussite qui dépasse à présent les 90%*” (“is probably 90% accurate”; Martin/Gerner/Souvay, forthcoming: 1); this level of accuracy is impressive.

As well as the central position occupied by the DMF in research into Middle French lexicology and synchronic lexicography, it is impossible to overstate its enormous heuristic potential for diachronic lexicology and etymology. Indeed, the vocabulary dealt with in the DMF refers to etymology – briefly but

---

<sup>3</sup> For it is an instrument, not a simple tool (for further reading on the distinction of those two notions, cf. Habert 2005).

systematically – by reference to the FEW (*Französisches Etymologisches Wörterbuch*). This is not only the case where the FEW deals with the lexical unit under scrutiny, but also when the DMF entry is an addition to the FEW's nomenclature, in which case this is indicated by “\*FEW”. Cases of this are far from rare: a quick survey reveals that the DMF1 includes 97 etymological references to the alphabetical section B- in the first volume of the FEW. Among those, 43 can be linked to articles that are due to appear in the new version of the FEW (see above, 2)<sup>4</sup>. The DMF thus emerges as *the* Middle French etymological dictionary, a feature which has tended to be overlooked.

A further example might be the lexicon of scientific language compiled by Danièle Jacquart and Claude Thomasset; in comparison with the FEW, this lexicon offers 339 new lexemes, 74 new etyma, along with hundreds of antedatings of individual words (Gerner/Martin 2007). The lexicon of the *Miracles de Nostre Dame par personnages*, compiled by Pierre Kunstmann, included 29 antedatings and one postdating (Gerner, to be published). These figures suggest that the future use of the DMF for historical lexicology is very promising.

### 3.1.3 International strategic position

It is not particularly surprising that the analysis offered by the DMF should overwhelmingly supersede Godefroy (1881-1902). Godefroy is the only dictionary to cover (albeit only partly) the same period of the history of French. During the last century, lexicographical techniques and, above all, much more sharply-focused lexicological analysis (notably in the field of semantic description) have advanced very significantly. In comparison with its illustrious ancestor – which is nonetheless still very useful (see below 5.3) – the DMF's contribution is particularly striking in terms of both quality and quantity. Hence, for the alphabetical range I-, R. Martin and G. Souvay (2003:398) provide 300 additional lexemes in comparison with Godefroy.

---

<sup>4</sup>The additions deal with 22 FEW articles: \*BABA (DMF s.v. *débaver, débaver [soi]*), BAJULARE (s.v. *baillable, baillette, baillie<sup>2</sup>*), BALBUS (s.v. *bauberie, baubie, bauboiment*), BALNEARE (s.v. *baigneresse, baigneur, baigneux*), BARBA (s.v. *barbelé, barbière, barbillonner*), BASSUS (s.v. *basser, subaste*), BASTUM (s.v. *bastorné*), BATARE (s.v. *baie<sup>2</sup>*), BLASPHEMARE (s.v. *blasphemeur*), BOS (s.v. *bouvatier*), BRACHIUM (s.v. *bracelette, bracelot, brasset, brassière*), BRANCA (s.v. *brancherie, branchier*), BREVIS (s.v. *desbreveter*), BRITTUS (s.v. *breton, bretonnant*), BROCCUS (s.v. *brochardre*), \*BRUSCIA (*brousser*), BÜCCA (s.v. *bellebouche, embouchoir*), BULLA (s.v. *billette, billeter, boulonnet*), BÜRRA (s.v. *bourrée, embourroumer [s']*), BÜTYRUM (s.v. *butirosité, butyreux*), BŪXUS (s.v. *buissière*), BYRSA (s.v. *boursal, boursière*).

Twenty years after its launch, the DMF is clearly *the* authoritative reference-work in the field of Middle French lexicology. Each new finding which emerges in the area: discovery of lexemes thus far undescribed; antedatings and postdatings; improved semantic analysis; attribution of a lexeme to a language variety – refers of necessity to the information available in the DMF's columns. Another invaluable asset is its productive relationship with the other two giants of medieval French lexicography, the DEAF (*Dictionnaire Etymologique de l'Ancien Français*) and the AND (*Anglo-Norman Dictionary*).

In addition, , at the beginning of the 21<sup>st</sup> century, the DMF is a natural focal point for the best research in lexicology, concentrating on the 14<sup>th</sup> and the 15<sup>th</sup> centuries. Mention should be made in this context of the enrichment of the DMF by Frédéric Duval's study of classical Latin and Greek terminology, or the contribution made by Yan Greub's 2003 thesis on the localisation of Middle French farces.

## **3.2 Computerized resource enhancement**

### **3.2.1 An unrivalled variety of use**

The DMF allows for a range of types of interrogation (cf. Gerner 2005: 159-161; Martin/Gerner/Souvay, forthcoming) which are unprecedented for a historical period of French (and perhaps for a historical period of any language). In what follows, we discuss only a few of the multidirectional modes of interrogation which are available: a complete etymological family is available via a single click on a FEW etymon; the computerized interrogation of definitions allows onomasiological access to the underlying data; and a specific menu is dedicated to the retrieval of locutions, much to the delight of phraseology scholars. Moreover, hyper-navigation does not stop where the dictionary *stricto sensu* ends, since the DMF is now at the centre of an increasingly dense interconnected network of research (cf. Gerner 2007b).

We illustrate below three types of interrogation that are of considerable value to scholars in historical linguistics. The examples derive from research undertaken at the ATILF. Our three examples deal with lexicology (3.2.2), pragmatics (3.2.3) and constructional morphology (3.2.4). Suffice it to say that any number of other types of enquiry are possible.

### **3.2.2 Lexicology**

The DMF was first and foremost designed to produce a description of the phonetic, semantic, morphosyntactic and variational aspects of the vocabulary of

Middle French. It goes without saying that the dictionary's contribution is particularly significant in the field of lexicology, whether synchronic or diachronic. One brief example demonstrates this obvious use of the DMF.

When the TLFi etymological notices of the articles *adresse*<sup>1</sup>, *adresse*<sup>2</sup>, and *adresse*<sup>3</sup> were revised as part of the TLF-Étym research programme (see below 4.1), one of the aims was to check the accuracy of the date “since 1559” suggested by the TLFi for the meaning I.A. under *adresse*<sup>3</sup>: “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*]”. The FEW (von Wartburg 1928 in FEW 3, 84b, \*DIRECTIARE I), Godefroy and Tobler-Lommatzsch (neither gives this meaning s.v. *adrece*), and the AND (which has no entry for this lexeme) yielded no information beyond that already available in the TLFi. Hence the TLF-Étym editors turned to the DMF: the choice of “Recherche d’une entrée” in the dictionary leads to nine articles *adresse*, taken from nine different lexica. Among them is found the following, an extract from the Christine de Pizan lexicon, compiled by Joel Blanchard and Michel Quereuil:

ADRESSE	FEW III *DIRECTIARE
<b>ADRESSE, subst. fém.</b>	
<p><b>I.</b> - "Chemin, direction" : Mais bien traihoit [Paris] en une <i>adresce</i> Fleches empanees d'un arc (CHR. PIZ., <i>M.F.</i>, 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'<i>adrece</i> (CHR. PIZ., <i>M.F.</i>, III, 1400-1403, 143).</p>	
<p><b>II.</b> - "Art" : La roïne d'Amasonie (...) D'armes savoit toute l'<i>adrece</i>. (CHR. PIZ., <i>M.F.</i>, III, 1400-1403, 141).</p>	
<p><b>III.</b> - "Celle qui montre le chemin, guide" : Philosophie y vi assise Moul haultement, en tel devise Que bien semble haulte maistresse Et des autres toutes l'<i>adresse</i> (CHR. PIZ., <i>M.F.</i>, II, 1400-1403, 104).</p>	
<i>Pizan</i>	Joël Blanchard / Michel Quereuil

The reader will immediately spot the fact that the attestation given under II. with the gloss “art” antedates the occurrence of the meaning “habileté manuelle” of *adresse* by one and a half centuries (1559 → 1400-1403), which is itself worthy of attention. Two other datings emerging from this collection of articles also originate in the DMF: before 1343 (Guillaume de Machaut) for “destination” (in fixed



locutions such as *arriver à son adresse*, cf. Petrequin/Buchi in TLF-Étym s.v. *adresse*<sup>1</sup> I. A.) and ca. 1400 (Froissart) for “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. *adresse*<sup>3</sup> II).

### 3.2.3 Pragmatics

Despite what readers might naïvely expect, the interest of the DMF is not restricted to the purely lexical, even in its currently available version (DMF2). Even though the treatment of grammatical units has been set aside for later, consultation of the dictionary has become a requirement for any scholar with an interest in the evolution of grammatical units and for that matter of discourse markers (or pragmatic units) in the French language.

In a recent publication, M.-B. Hansen (2005: 47) dates the first occurrence of the marker *enfin* with the meaning “marking the final element of a series of items quoted” (for instance in *Il y a à Paris trois polices: primo: la police du royaume [...] secundo: celle du régent [...] enfin celle de Dubois*) to the second half of the 16<sup>th</sup> century. Anyone who wants to check the accuracy of the dating of *enfin* will turn to the DMF, even though the dictionary does not include – for the time being – the entry *enfin*. But as luck would have it, the dictionary has a means of interrogation “Recherche plein texte”, which offers 58 results for *enfin*. Among them, we find the following: **Article 7/58 1 attestation**

<b>BARATER</b>	<b>FEW IX <i>prattein</i></b>	
<b>BARATER, verbe</b>		
[T-L : <i>barater</i> ; FEW IX, 330a : <i>prattein</i> ]		
Empl. trans.		
<p><b>A.</b> - "Tromper, frauder" : ...qui tousjours estoient ententis et occupez en saintes meditacions 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. (<i>Horloge de sapience</i> S., c.1389, 81). ...luxure donc y est pour deliter, et avarice y est pour profiter, et trayson enfin pour l'amant bareter. (<i>EVR. CONTY, Eschez amour. mor. G.-T.R.</i>, c.1400, 245).</p>		
<p><b>B.</b> - "Agiter, battre"</p> <p>- <i>Lait baraté</i>. V. <i>lait</i></p>		
Littérature didactique		Hiltrud Gerner

DMF2's article *barater* includes, under A. "tromper, frauder", an attestation from Evrart de Conty antedating *enfin* "énumératif" by nearly two centuries (1587 → ca. 1400)<sup>5</sup>. This undoubtedly has implications for our understanding of the historical development of this marker, whether aspectual or enunciative (cf. Buchi/Städtler, in progress).

### 3.2.4 Constructional morphology

It may be even more surprising to discover that the DMF is of considerable use for another linguistic sub-discipline, i.e. constructional morphology.

A stimulating study produced by our colleague Denis Apothéloz (2003) highlights the dual nature of the prefix *IN-* in contemporary French. On the one hand, Apothéloz distinguishes a "basic" prefix *in*<sup>-1</sup>, characterized by allomorphy: /in-/ before a basic form starting with a vowel (*inutile*), /ẽ-/ before a basic form starting with an obstruent consonant (*imbattable*), and /i-/ before a basic form starting with a resonant consonant (*illégal*). *In-* has either a negative meaning (*inutile* "not useful") or a superlative meaning (*inqualifiable* "extremely bad or poor"). Regarding the second term of the pair, *in*<sup>-2</sup>, it only arises with basic forms starting with an obstruent consonant, and it is always (in Apothéloz's findings) realized as /ẽ-/ ([ẽnɔmabl]); its meaning is exclusively negative (*irréparable* "irreparable, unable to be repaired"). D. Apothéloz (2003: 43) asks the question as to when the pairs in /ẽ-/ of the derivatives starting with /i-, and formed on a basic form with an obstruent consonant, evolved. Can the DMF answer this question? Does it include lexemes with the prefix *in-* (/ẽ-/) before roots in *r-* or *l-*? For this kind of oblique interrogation, the DMF offers, under the heading "Recherche sur les entrées", a "+ options" button which enables the user to select, in order to filter the entries, an option to look up a character string – via the initial letters of a series of words. This is different from the default mode "texte exact". Here is a sample of the results for *inr-*:

---

<sup>5</sup> The continuity of the tradition is guaranteed by an intermediary attestation, in 1452, proposed by the DMF2 s.v. *tare*.

inrabilité	1
inracontable	1
inraisonnable	2
inrégi	1
inréglé	1
inrégularité	1
inrégulièrement	1
inréméable	1
inrémédiable	1
inrémédiablement	1
inremuable	1
inrémunéré	1
inrenommé	1
inréparable	2

Structure de l'article

Article sans exemples

Article complet

Formes de l'entrée

Exemples de l'entrée

Rechercher dans les textes

Nombre d'entrées : 28

The result of such a request is surprisingly rich: it highlights 28 forms starting with *inr-*, and the majority of them are not listed in other lexicographical sources. The DMF data will thus prove to be crucial in answering D. Apothéloz's question concerning the chronology of this complex morpho-semantic system (cf. Buchi, forthcoming).

## 4 Recent developments in historical lexicography

### 4.1 TLF-Étym

#### 4.1.1 Methodology of TLF-Étym

The *TLF-Étym* research project (cf. <http://www.atilf.fr/tlf-etym>, and Buchi, 2005) is directed by Gilles Petrequin. It aims at progressively revising the sections entitled “Étymologie et Histoire” in the TLF*i*. There can be no question of revising all the 54,000-odd historical notices of the TLF*i*. On the one hand, the project endeavours to incorporate work by various researchers in their different fields; on the other hand, it will attempt a systematic revision of etymological notices in certain categories which are recognizably deficient (for instance, those which deal with Anglicisms and deonomastic words).

The methodological framework of the TLF-Étym project is based on the history and etymology of each word: this conception stems from the idea that etymology is not simply the identification of an etymon. It aims at clearly stating the formal and semantic link between the etymon and the lexeme under examination; word-history

thus plays an integral part in etymology. This conception derives ultimately from Schuchardt, Meillet, Ernout and above all von Wartburg. But Baldinger (1959: 239) probably put it best: “l’étymologie, [...] c’est [...] la biographie du mot” (“etymology is the biography of a word”).

Indeed, a proper methodology lies at the core of the elaboration of this programme: “TLF-Étym a la chance de pouvoir bénéficier des réflexions méthodologiques des meilleurs spécialistes qui se sont exprimés lors des séminaires de méthodologie en étymologie et lexicologie historique qui se sont déroulés de septembre 2005 à juin 2006” (Chambon/Carles 2007: 315: note 2). The proceedings of those seminars can be downloaded from the TLF-Étym website (under the heading “En savoir plus”). Such an initiative from the ATILF has thus made it possible to associate with our research institute, the main contemporary experts in French etymology from all around the world (foreign researchers such as Thomas Städtler and the DEAF team [Heidelberg], Franz Rainer [Vienna] or Takeshi Matsumura [Tokyo], for instance, are closely involved in the project). It has also helped to revitalize this specific field of French historical linguistics. It is also important to note that the elaboration of etymological notices is often accompanied by explicit, independently-published theoretical commentary:

- *anil et indigo* → Benarroch, forthcoming;
- *casuel<sup>2</sup> et défectif* → Andronache, in progress;
- *cheire* → Chambon/Grémois 2007 (*cf.* also Gouvert 2007);
- *claire* → Chambon forthcoming and Petrequin/Andronache in progress;
- *défectif, différencier, obole* → Steinfeld/Andronache forthcoming;
- *lucarne* → Pitz 2006.

The revised notices can be downloaded from the TLF-Étym website; they include unpublished etymologies (Turcan s.v. *bienfaisance*: calqued on Latin, rather than generated within French; Chauveau s.v. *bigler*: continuation of proto-Romance, rather than a borrowing from Latin; Buron/Baudinot s.v. *fare*: borrowing from Breton instead of “unknown origin” etc.), antedatings (Stumpf/Evrard s.v. *fabulateur*: 1541 → *ca.* 1360/1380; Städtler s.v. *laconique*: 1529 → *ca.* 1372/1374; Petrequin s.v. *ostensoir*: 1771 → 1673; etc.), and ultimately, postdatings (Koehl s.v. *iota*: *ca.* 1240 → *ca.* 1300; Steinfeld s.v. *féodalement*: 1483 → 1514; Robin/Buchi s.v. *vélocipède*: 1804 → 1818; etc.). Nevertheless, a reasonable conclusion is that the greatest value of the TLF-Étym project will be the precision and coherence of the models of analysis developed for the twenty-two etymological sub-classes, whether with regard to inherited vocabulary (Steinfeld s.v. *claire*: “Continueur régulier du protoroman régional \*/kleta/” [TLFi: “du gaul. *cleta*”]), lexical borrowings (Navrátilová s.v. *riesling*: “emprunt à l’allemand” [TLFi: “mot all.”]) or French developments (Leroy s.v. *cerbère*:

“Formation française: translation déonomastique du nom propre de créature mythologique *Cerbère*” [TLFi: “Empr. au lat. *Cerberus*”]).

#### 4.1.2 Example

Moreover, the TLF-Étym forms a natural home for the etymological notices currently being revised as part of the DETCOL project (Développement et Exploitation Textuelle d’un Corpus d’Œuvres Linguistiques), directed by Bernard Colombat (cf. [http://ctlf.ens-lsh.fr/documents/ct\\_projet\\_detcol.pdf](http://ctlf.ens-lsh.fr/documents/ct_projet_detcol.pdf)). The website thus offers a renewed panorama on etymology and the history of technical grammatical terms, such as *antécédent*, *conjonction*, *gémiation*, *gérondif*, *négation*, *parfait*, *pluriel*, *pronom* or *temporel*. The reader can compare, below, the etymological notice for the article *parfait*<sup>2</sup> in the TLFi with its revised version in 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 perfectio : *par le preterit parfet*, et plusquamparfetio : *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 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 / l'ai mägé, ich hab gessen. / Advenir. Zeit so noch zukünftig ist / Je mangeray, ich werde essen). Première attestation dans une grammaire française : 1606 (**MASSET**, *Acheminement*, page 11 : Nos verbes ont cinq modes. Indicatif, impératif, optatif, subiunctif, & infinitif : Trois temps principaux, present, passé & futur. Le passé se diuise en imparfait, aoriste simple, **parfait**, plus que parfait, aoriste composé, & parfait, tres-parfait ; cf. aussi **MAUPAS**, *Grammaire*<sup>2</sup>, page 92 v<sup>o</sup>/93 v<sup>o</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*\*. 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 (cf. “ou substantivement” dans la citation ci-dessus C.) 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>. On écarte de même l'hypothèse d'un emprunt à l'allemand *Perfekt* subst. neutre « temps verbal présentant le procès comme accompli et l'envisageant dans son résultat actuel » (attesté depuis le 17<sup>e</sup>/18<sup>e</sup> siècle seulement, **SCHULZ**, *Fremdwörterbuch*<sup>1</sup>), même si la première attestation absolue se trouve dans un dictionnaire allemand-français.

Dès les premiers témoignages d'un discours grammatical français, on relève le latinisme *prétérit parfait* (ci-dessus C.). C'est seulement à la toute fin du 16<sup>e</sup> siècle (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 : E. Ammann. – Mise à jour 2007 : E. Buchi. - Relecture mise à jour 2007 : J-P. Chauveau ; T. Städtler ; J-P. Chambon ; M. Lang ; G. Roques ; Y. Matsumura ; G. Petrequin ; N. Steinfeld.

## 4.2 *Base des Mots-fantômes* (“Database of ghost-words”)

### 4.2.1 Methodological characteristics

The *Base des Mots-Fantômes* is a resource in critical metalexigraphy, directed by Nadine Steinfeld. It aims to list the *ghost words* or *fantasmas lexicográficos* (Pascual Rodríguez/García Pérez 2007:170), i.e. all the pseudo-lexemes of erroneous lexicographical status (“those words that do not exist”), the ghost meanings and the wrong lemmatizations that can be found in major French historical and etymological dictionaries.

### 4.2.2 Example

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

LABAILLE, s. f., syn. d'*escope*, mod. *écope*, sorte de pelle creuse qui sert à vider l'eau entrée dans une embarcation :  
Et en doivent les vaisseaux 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 Constantin*, Arch. P 304, f<sup>o</sup> 116 v<sup>o</sup>.)

Von Wartburg failed to supply an etymology for the word and relegated it (in 1968) to the “materials of unknown or uncertain origin” (FEW 23, 108b, s.v. “écope”). This incarceration in the FEW’s “purgatory” constituted an implicit appeal to the scientific community to come up with etymological hypotheses to clarify the origin of *labaille*.

The mystery was eventually solved in 2006: as part of the revision of the article BAJULA “bonne d’enfants” of the FEW, Jean-Paul Chauveau (2006: 2: n.4) identified the word *\*labaille* as a misreading, by erroneous agglutination of the definite article, of a form which should be attached to the lexeme *baille* n.f. “baquet de bois en forme de demi-tonneau ou de cône tronqué, spécialement utilisé sur les bateaux”, attested since 1340.

Now, because of the semantic gap between “écope” and “bonne”, it is very unlikely that any reader of Godefroy who was interested in the etymology of *\*labaille* would automatically think of consulting the FEW s.v. BAJULA. This is precisely where the *Base des Mots-Fantômes* comes into play: it centralizes in a very convenient way the identifiable ghost-words scattered across articles in scholarly journals, conference proceedings, notices and lexicographical studies, as well as establishing new hypotheses regarding connections between forms. This is how the entry *labaille* appears in this data base:

<b>LABAILLE</b>	
1331-1500	lexie
<b>Source première :</b>	
Gdf 4, 686a labaille, subst. fém. « syn. d' <i>escope</i> , mod. <i>cope</i> , 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 <i>escope</i> ou <i>labaille</i> ou l'en puche l'eaue 1413 ( <i>Denombr. du baill. de Constantin</i> , Arch. P 304, f° 116v°)	
<b>Source(s) secondaire(s) :</b>	
<b>Source secondaire :</b>	
FEW 23, 108b Concept : ÉCOPE Anorm. <i>labaille</i> f. « sorte de pelle creuse qui sert à vider l'eau entrée dans une embarcation » (Coutances 1413)	
<b>Justification :</b>	
mélecture portant sur un phénomène d'agglutination liée à l'article <i>la</i> 1413 Il convient de lire <i>la baille</i> « le baquet »	
<b>Solution :</b>	
<b>LA BAILLE</b> FEW 1, 206a BAJULA 2	
<b>Correction des sources :</b>	
<b>Source à corriger :</b>	
correction, déplacement-S GdfC 8, 273b baille, subst. fém. « baquet » Rem. : <i>Baille</i> , dans l'attestation de J. Du Clercq, <i>Mém.</i> , l. V, ch. XVIII, III, 375, citée ici, a été interprété à tort par Gay comme signifiant « palissade ».	
<b>Source à corriger :</b>	
correction, déplacement FEW 1, 206a BAJULA	
2005 : J.-P. Chauveau ; T. Städtler, T. Matsumura	



This is, then, a nice example of reciprocal enrichment operating between the FEW and the Base des Mots-Fantômes.

### 4.3 *Bibliographie Godefroy*

#### 4.3.1 Methodological characteristics

“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” (“The users of Godefroy are torn between being fascinated by the vast amounts of information contained in the 8,000 pages of the ten volumes of the dictionary, and frustration at the absence of a proper bibliography of sources used”; Ringenbach 2003: 191). This observation was the starting-point of the Bibliographie Godefroy project, in progress, thanks to Jean-Loup Ringenbach (cf. <http://www.atilf.fr/BbgGdf>; Ringenbach 2003; forthcoming). The sigla system used by Godefroy is, it has to be said, unreliable. One example: the *Dit de buffet* (mid 13<sup>th</sup> century) is sometimes referred to as “Dit de buffet” or “Dit du buffet”, and sometimes as “Du Vilain au buffet”, without a date being supplied. The unreliability of the dictionary’s own reference-system means that the Bibliographie Godefroy will become every bit as indispensable as the bibliography to the DEAF.

The methodology used in the production of the Bibliographie Godefroy consists in identifying obscure sigla, either by cross-reference to more explicit abbreviations, or by philological investigation of editions of texts – and even of manuscripts – which contain the texts concerned.

#### 4.3.2 Example

As part of our research on the history of the prefix IN- in contemporary French (see above 3.2.4), the following attestation from Godefroy remained a puzzle: *Quant est de toy, tu mors et pinces Par ton envye irraisonnable Plus c’un serpent* (*Envye, Estat et Simplese*, p. 6, ap. Ler. de Lincy et Michel, *Farces, Moral. et Serm. joy.*, t. I). What was the source? How were we to date it? Thanks to the Bibliographie Godefroy, this text can now be identified:

<p><b>Envye, Estat et Simplese</b></p> <p><i>Envye, Estat et Simplese</i></p> <p>1) Exemples :</p> <p>- Gdf : <i>irraisonnable</i> : <i>Envye, Estat et Simplese</i>, p. 6, ap. Ler. de 24341 et</p>
--

Michel, *Farces, Moral. et Serm. joy.*, t. I  
- autres citations : Gdf : *grogneu, oppresse, ragace*

2) Bibliographie :

*Moralité à trois personnages, c'est assavoir Envye, Estat et Simplese*

- Texte : milieu du 16e s.
- Manuscrit : PARIS BN fr 24341 (Recueil La Vallière), f° 50a-53b, ca 1575
- Édition(s) : *Recueil de farces, moralités et sermons joyeux*, éd. Leroux de Lincy et Francisque Michel, Paris, t. 1, 1837, n° 10.

Renvois :

- Bossuat : 5844
- Louis Petit de Julleville, *Répertoire du théâtre comique en France au moyen âge : histoire du théâtre en France*, Paris, 1886, 62-63.

Once the text *Envye, Estat et Simplese* had been dated to the mid-16<sup>th</sup> century, we realized that this was the last occurrence of the adjective *inraisonable*, at a time when French was opting decisively for *irraisonnable* (attested from the 14<sup>th</sup> century). Conversely, the value of such an attestation for historical lexicography (and over and above that, for historical constructional morphology) would have gone unnoticed without the help of the Bibliographie Godefroy.

## 5 The CNRTL Lexical Portal

### 5.1 Objectives of the CNRTL

The Centre National de Ressources Textuelles et Lexicales (CNRTL: [www.cnrtl.fr](http://www.cnrtl.fr)) was created in 2005 by the CNRS and is directly associated with our institute, ATILF. It aims to gather together within one single portal as many computerized resources and consulting tools as possible for the study, understanding, and diffusion of the French language.

Because it results from the bringing-together of studies undertaken in various research institutes, the CNRTL is able to maximize the production, verification, harmonisation, diffusion and sharing of resources – whether computerized lexical or textual data, or tools and instruments designed to facilitate informed access to their content.

The decision to create the CNRTL is part of CNRS policy, which aims to establish new infrastructures, which will be vital for research work involving the whole scientific community. It is also the consequence of a joint programme by the “Direction de l’Information Scientifique” and the “Département Homme et

Société” within the CNRS. The acknowledged scientific expertise of the CNRTL, along with the numerous national and international collaborative projects run within the various research institutes of which it is part, have enabled the project to establish itself within European research, thanks to:

- direct collaborations with partner institutes, in Great Britain (University of Oxford), Germany (Computing centres in Trier and Würzburg, DFKI in Saarbrücken, MPI), and in the Netherlands (University of Nijmegen).
- participation in CLARIN (<http://www.mpi.nl/clarin/>) the European network of resource management and centres of linguistic technology.

## 5.2 *Resources located within CNRTL*

The CNRTL revolves around five activities: a lexical portal for French; text and data corpora, annotated or not; encyclopaedic and linguistic dictionaries (old and modern); phonetic, morphological, syntactical, semantic lexica; linguistic tools (labellers, analysers, aligners, concordancers, annotation tools). Among the resources that are already part of the CNRTL – the lexical portal is further discussed in the following paragraph – the following should be noted:

- corpora of texts free of writers’ and publishers’ copyright (in the first instance, 500 texts taken from Frantext): selected according to their authors, titles, dates and genres, with the possibility of downloading selected texts in XML format – with a TEI-compliant DTD ([www.tei-c.org](http://www.tei-c.org))<sup>6</sup>; the user thus has access to files including the DTD and the XML/TEI encoding of the texts. To the best of our knowledge, the CNRTL is the first website to offer a collection of French corpora in XML/TEI format with around 150 million characters.
- the Morphalou lexicon, derived from the TLF nomenclature, whose access is free for consultation and downloading: an open lexicon of French inflected forms, containing 524,725 inflected forms belonging to 95,810 lemmas. These have been linguistically validated (approved by an editorial committee) and conform to the proposal for normalization of ISO lexical data (TC37/SC4).
- Computerized versions of dictionaries, whether modern (TLFi; Dictionnaire de l’Académie française: 8<sup>th</sup> (1932-1935) and 9<sup>th</sup> (1992-) editions) or older (Dictionaries of R. Estienne (1552), of Jean Nicot (1606), of Bayle (1740), of Féraud (1787-1788), of the Académie (1<sup>st</sup> edition, 1694; 4<sup>th</sup> edition, 1762; 5<sup>th</sup>

---

<sup>6</sup> Let us here remind our reader that Nancy, thanks to the association of the ATILF with the INIST and the LORIA, is now the European support centre for the TEI.

edition, 1798; 6<sup>th</sup> edition, 1835), along with the *Encyclopédie* by Diderot and d’Alembert<sup>7</sup>.

The CNRTL also offers to the scientific community linguistic tools and instruments that can be used directly on the website, from a simple internet browser. Among the different projects in progress or planned, we will be offering simple and user-friendly access to tools such as:

- FLEMM: a flexional analysis tool for French texts that have been previously labelled, via one of the two categorizers: Brill or TreeTagger.
- POMPAMO: a detection tool for potential cases of formal and categorical neologism, based on the use of restricted lexica. This project makes use of lexical resources such as Morphalou and offers the possibility of creating new resources as well.

### **5.3    *The Lexical Portal, an example of lexical data integration***

The CNRTL lexical portal aims to enhance and share – mainly with the scientific community – various types of data taken from research studies on the vocabulary of French. An ongoing project, this lexical data-base endeavours to provide users with as much available information as possible, the starting-point of each enquiry being the individual lexical unit.

#### **5.3.1    Lexicographical information**

Lexicographical data are the priority in our work. Thus far, we have managed to include in this portal all the information provided by the TLF ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) which appears by default when lexicographical information is requested. This information is supplemented by easily available information via a menu. This is taken from:

- the Académie Française dictionary (4<sup>th</sup>, 8<sup>th</sup> and 9<sup>th</sup> editions) ([www.atilf.fr/academie](http://www.atilf.fr/academie)), which has been computerized within our institute as part of a partnership programme with the Académie.
- the Base de Données Lexicographiques Panfrancophones (BDLP: <http://www.tlfq.ulaval.ca/bdlp/>). This international-scale project is part of the “Trésor des vocabulaires français” enterprise originally initiated by Bernard

---

<sup>7</sup> Most of the computerized versions of old dictionaries, along with that of the *Encyclopédie* of Diderot and d’Alembert are the result of a partnership with the ARTLF: (<http://humanities.uchicago.edu/orgs/ARTFL/>).

Quemada in the 1980s. The aim of the BDLP is to construct and to assemble representative databases of French as spoken in each French-speaking area. The databases are designed in order that they can be consulted separately or at the same time, and so that they constitute a supplement to the *Trésor de la Langue Française informatisé*. With an international perspective, the BDLP project is sponsored by the Agence Universitaire de la Francophonie, which supports it via its network for the study of French in the French-speaking world (<http://www.eff.auf.org/>).

- the Base Historique du Vocabulaire Français (*Datations et Documents Lexicographiques*: [www.atilf.fr/ddl](http://www.atilf.fr/ddl)), made up of datings for French vocabulary, and based on the data found in the 48 volumes of the *Datations et Documents Lexicographiques* collection (Quemada 1970-1998).

All of these lexicographical data are also directly available, for any given form, at the address <http://www.cnrtl.fr/lexicographie/>, followed by the form which the user wishes to consult. Thus:

<http://www.cnrtl.fr/lexicographie/aguerrir> enables the user to access all available lexicographical information for the verb *aguerrir*.

The screenshot displays the CNRTL website interface. At the top, the header reads 'Centre National de Ressources Textuelles et Lexicales'. Below the header is a navigation menu with tabs for 'Morphologie', 'Lexicographie', 'Etymologie', 'Synonymie', 'Antonymie', 'Proxémie', 'Concordance', and 'Aide'. The 'Lexicographie' tab is active. A search bar contains the word 'aguerrir' and a 'Chercher' button. Below the search bar, the entry for 'AGUERRIR, verbe trans.' is shown. The entry includes the following information:

- Habituer à la guerre**
- A. – Emploi trans.** [En parlant de pers.]
- 1. Domaine milit.** *Aguerrir* qqn. **L'accoutumer à mener une vie de combattant, avec les fatigues et les dangers qu'elle entraîne.**
  - 1. Il fallait remettre l'ordre en Italie. Il fallait substituer peu à peu aux légions indociles qui avaient vaincu à Philippes, une armée qui valait celle d'Antoine; la discipliner, **l'aguerrir**. Il fallait, sous les yeux de Sextus, maître de la mer, construire des vaisseaux, exercer des matelots. J. MICHELET, *Histoire romaine*, t. 2, 1831, p. 308.
  - 2. Ce n'est point par la force qu'il convient de résister à la force, car la lutte **aguerrit** les combattants et le sort des batailles est douteux. A. FRANCE, *Le Puits de Sainte-Claire*, 1895, p. 202.
- 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. F. PONSARD, *L'Honneur et l'argent*, 1853, II, 6, p. 40.
  - 4. J'aurais de la peine à vous expliquer comment une fantaisie aussi hardie pouvait naître dans un esprit que je vous ai montré d'abord si pusillanime; mais bien des épreuves m'**avaient aguerris**. E. FROMENTIN, *Dominique*, 1863, p. 194.
  - 5. La semaine suivante, Charles fut emporté par une fièvre typhoïde. Un soir, sa grand'mère lui avait relu le combat du *Vengeur* pour **l'aguerrir**; et le délire l'avait pris dans la nuit. É. ZOLA, *Le Capitaine Burle*, 1883, p. 60.
  - 6. – Vous devrez obligatoirement jouer dehors, sauf s'il pleut. – Mais s'il fait très froid? hasarda mademoiselle.

### 5.3.2 Morphosyntactical Information

Morphosyntactical data are taken from the Morphalou base ([www.atilf.fr/morphalou](http://www.atilf.fr/morphalou)), which was set up along with the TLFi nomenclature in the first place. This information is also directly available from any form of the paradigm, as in *aguerrit*, via: <http://www.cnrtl.fr/morphologie/aguerrit>

The screenshot shows the 'Morphologie' section of the Morphalou ATILF website. The search bar contains 'aguerrit' and the category is set to 'verbe'. Below the search bar, a table lists the morphological forms of the verb 'aguerrir'.

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>	
aguerrissons	indicatif	présent	pluriel	1 <sup>ère</sup> personne	

### 5.3.3 Etymological Information

Etymological data are taken from the TLF ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) and the TLF-Étym, which updates the etymological notices of the TLF (see above 4.1). They are directly available, for a given form, via: <http://www.cnrtl.fr/etymologie/aguerrit>

The screenshot shows the 'Etymologie' section of the TLFi website. The search bar contains 'aguerrit' and the category is set to 'toutes'. Below the search bar, the etymological information for the verb 'AGUERRIR' is displayed.

**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 t d'aguerrir davantaige la puissance [« armée »] des Macedoniens). Pour ce qui est de TLF (datée improprement de 1535), elle concerne l'adjectif *aguerrit* (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 devient 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).

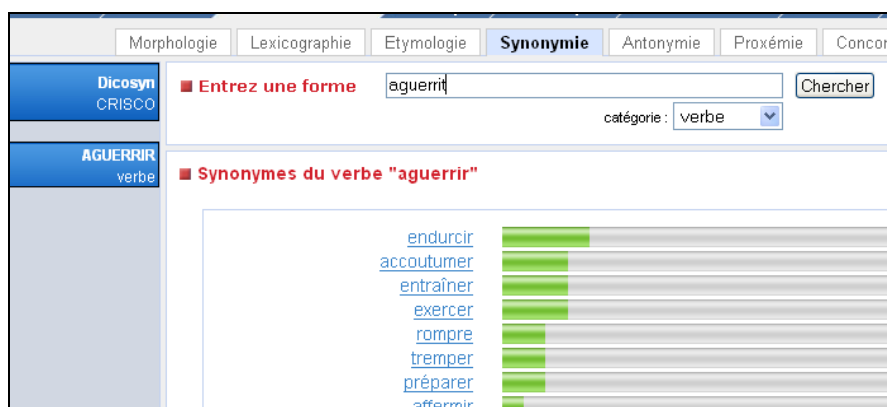
**Informations**

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

### 5.3.4 Synonyms and antonyms

Data on synonyms and antonyms come from the Caen dictionary of synonyms (<http://www.crisco.unicaen.fr/>), which was set up on the basis of data from the INaLF. These data are also directly available via:

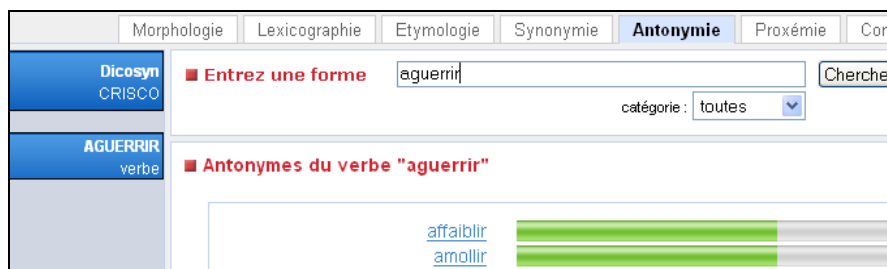
<http://www.cnrtl.fr/synonymie/aguerrir>



The screenshot shows the 'Synonymie' tab of the Dicosyn CRISCO interface. The search bar contains 'aguerrir' and the category is set to 'verbe'. Below the search bar, the title 'Synonymes du verbe "aguerrir"' is displayed. A list of synonyms is shown with green progress bars indicating their frequency or relevance:

Synonym	Progress Bar
endurcir	Full green bar
accoutumer	Medium green bar
entraîner	Medium green bar
exercer	Medium green bar
rompre	Short green bar
tremper	Short green bar
préparer	Short green bar
affermir	Short green bar

or <http://www.cnrtl.fr/antonymie/aguerrir>

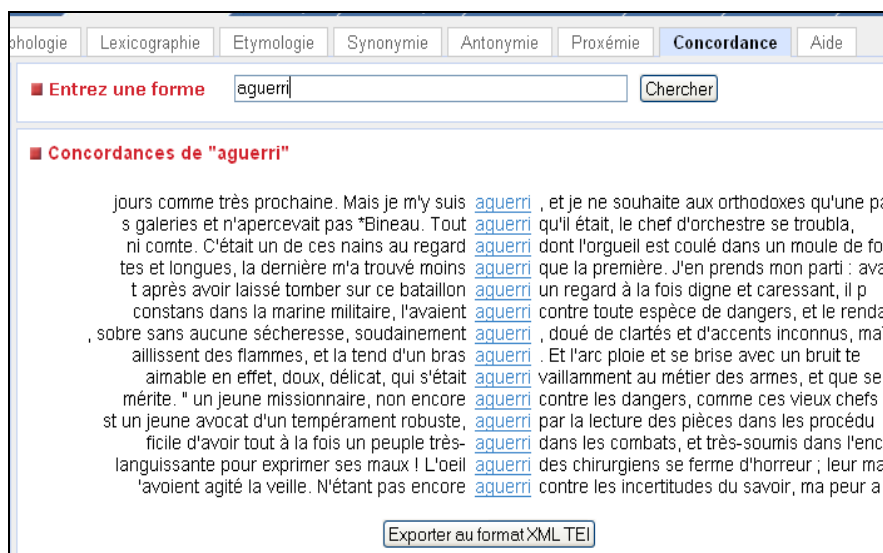


The screenshot shows the 'Antonymie' tab of the Dicosyn CRISCO interface. The search bar contains 'aguerrir' and the category is set to 'toutes'. Below the search bar, the title 'Antonymes du verbe "aguerrir"' is displayed. A list of antonyms is shown with green progress bars indicating their frequency or relevance:

Antonym	Progress Bar
affaiblir	Full green bar
amollir	Full green bar

### 5.3.5 Concordance

This concordance makes use of the corpus of copyright-free texts on the Frantext database ([www.atilf.fr/frantext](http://www.atilf.fr/frantext)); it also provides the option of converting the results of the concordancer into XML/TEI format. As far as we are aware, it is the only website which enables a user to convert a French concordancer of this size into a normalized format. The concordances are also directly available via: <http://www.cnrtl.fr/concordance/aguerrir>



Moreover, the user only has to right-click on one of the examples to get its complete reference. Hence, for the first example:



The lexical portal also permits double-clicking on a word-form and hypernavigation to all the lexical data available regarding the given lexical unit. For instance, in order to get information concerning the form *apercevait*, from the second example in the concordance, a double-click on the word-form produces a new menu, directing the user towards all available data on this form and on the verb *apercevoir*:





## **6 Conclusion**

### **6.1 Computerization as a means to enhance lexicographical research.**

As far as French is concerned, the TLFi has opened up ground-breaking perspectives, thanks to its rich content, entirely XML-encoded. For a long time, the TLF had been said to be a dictionary designed for an elite. Such a view of the TLF could be justified by at least three characteristics of the dictionary's printed version:

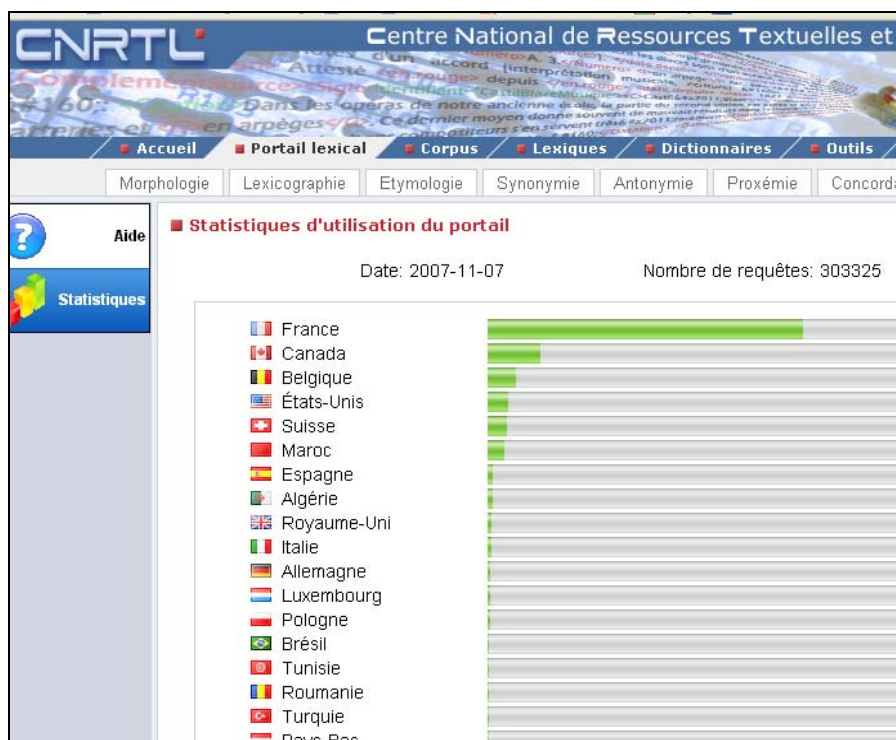
- its length, 16 volumes, each of them above 1000 pages.
- its descriptive richness, which could sometimes be viewed as an obstacle to casual reading, at least for the more complex articles; the article *aimer*, for instance, extends over 12 pages, or 24 columns, and it was not always easy for non-specialists to cope with such rich information.
- its cost, about 1500 euros, which prevented it being widely accessible.

Even though it soon became an authoritative reference work in the field of French lexicography, its publication was limited to a thousand or so copies, distributed to a limited public mainly of the intelligentsia.

Its computerized version in the form of a CD-Rom (around 15,000 copies sold in less than four years) or as a freely-accessible, web-based database achieved great success, among people at large, but also among scholars and language professionals. There are now 300,000 daily connections to its Web version from all around the world, and it is referenced in an innumerable range of sources. Its notoriety makes it a great means of promoting French.

Its more recent integration within the CNRTL lexical portal and its interconnections with other types of resources for French vocabulary studies place it at the core of a collection of resources on French in which it retains pride of place. It thus shows that its elitist reputation is clearly unjustified, and its diffusion on the CNRTL lexical portal (the target of around 300,000 daily hits from all around the world) makes it one of the most consulted dictionaries on the internet to date.

Finally, a comment on the fact that the diffusion of a computerized version of an authoritative scientific reference-work now offers new scope for the production and dissemination of research findings. Beyond the purely scholarly world, new technology offers the opportunity to make our research findings available to society at large. The value of this is demonstrated by comments on the Web, on various institutional or professional websites.



The increase in the use of enhanced, computerized versions is thus dramatically changing working methods, and the extent of collaboration, within the humanities and social sciences research communities.

## 6.2 From French to Romance languages

The strategic position of the ATILF in the field of computerized lexicographical resources is, to date, chiefly of value for French alone. Nonetheless, this restriction is far from being set in stone: on the one hand, the CNRTL positively welcomes resources and instruments dedicated to languages other than French, and on the other hand, the ATILF is soon to move from French to Romance languages, at least in the field of historical lexicology.

This diversification mainly concerns Gallo-Romance languages in the first place: two types of computerized lexicography are under progress at the moment within the FEW (cf. <http://atilf.fr/few>). The first part of this global enterprise is intended to lay the foundations of the future retroconversion of the dictionary (cf.

Renders to be published; in progress); the second type of work concerns the computerization of the editorial procedure (cf. Matthey/Nissille to be published).

A development encompassing the whole of the Romance language family is already under way thanks to the launching of the DÉRom project (*Dictionnaire Étymologique Roman*, cf. <http://www.atilf.fr/derom> and Buchi/Schweickard, forthcoming) in January 2008. The ideal might be, in a near future, the possibility of hyper-navigation between the FEW, the LEI and the DÉRom; only time will tell if such a dream will be fulfilled...

## 7 Bibliography

- Andronache M., Le problème de la continuité en lexicologie historique. Réflexions à partir de la pratique lexicographique dans le cadre du projet DETCOL, proposition de communication soumise au Congrès Mondial de Linguistique Française 2008, in progress.
- ATILF *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 OS X, ISBN 2-271-06365-5, 2005.
- Apothéloz D., Le rôle de l'iconicité constructionnelle dans le fonctionnement du préfixe négatif *in-*, *Cahiers de linguistique analogique* 1, 2003, p. 35-63.
- Baldinger K., L'étymologie hier et aujourd'hui, *Cahiers de l'Association internationale des Études françaises* 11, 1959, p. 233-264.
- Benarroch M., Le traitement des emprunts au portugais dans le TLFi, in: Pierrel J.-M. et al. (éd.), *Actes du Colloque international "Lexicographie et informatique: bilan et perspectives" (Nancy, 23-25 janvier 2008)*, forthcoming.
- Buchi É., Le projet TLF-Étym (projet de révision sélective des notices étymologiques du *Trésor de la langue française informatisé*), *Estudis romànics* 27, 2005, p. 569-571.
- Buchi É., *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)*, forthcoming.
- Buchi É./Schweickard W., À la recherche du protoroman: objectifs et méthodes du futur *Dictionnaire Étymologique Roman* (DÉRom), in: Iliescu M. et al. (éd.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique Romane (Innsbruck, 3-8 septembre 2007)*, Tübingen, Niemeyer, forthcoming.
- Buchi É./Städler Th., 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”), proposition de communication soumise au Congrès Mondial de Linguistique Française 2008, in progress.
- Chambon J.-P., Pratique étymologique en domaine (gallo-)roman et grammaire comparée-reconstruction. À propos du traitement des mots héréditaires dans le TLF et le FEW, forthcoming (Festschrift for a colleague).
- Chambon J.-P./Carles H., À propos du traitement des emprunts à l’occitan dans le *Trésor de la langue française*, in: Rézeau P. (éd.), *Richesses du français et géographie linguistique*, vol. 1, Bruxelles, de Boeck, 2007, p. 313-325.
- Chambon J.-P./Grélois E., Pour la révision de l’article *cheire* du *Trésor de la langue française*, in: Rézeau P. (éd.), *Richesses du français et géographie linguistique*, vol. 1, Bruxelles, de Boeck, 2007, p. 327-360.
- Chauveau J.-P., FEW: article BAJULA, Nancy/ATILF, publication électronique (<http://www.atilf.fr/few/bajula.pdf>), 7 p., 2006.
- DEAF= Baldinger K. et al., *Dictionnaire Étymologique de l’Ancien Français*, Québec/Tübingen/Paris, Presses de l’Université Laval/Niemeyer/Klincksieck, 1974-.
- Dendien J., Pierrel J.M. Le Trésor de la Langue Française informatisé: un exemple d’informatisation d’un dictionnaire de langue de référence, *TAL* Vol 44 – n° 2/2003, 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, site internet: <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.
- Frantext=INaLF, puis ATILF. *Frantext* (outil de consultation de ressources informatisées sur la langue française), site internet: <http://www.frantext.fr>, Nancy, ATILF (CNRS & Nancy-Université), 1992-.
- Gaume B., Cartographier la forme du sens dans les petits mondes lexicaux, in *JADT* 2006, p. 541-465.
- Gerner H., *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., *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, 2007a, vol. IV, p. 69-77.
- Gerner H., 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, 2007b, p. 101-109.
- Gerner H., Cueillette de quelques attestations intéressantes dans le Lexique des *Miracles Notre Dame par personnages*, forthcoming (Festschrift for a colleague).
- Gerner H./Martin R., Le lexique de la langue scientifique à l'aune du FEW, in: James-Raoul D./Soutet O. (éd.), *Par les mots et les textes. Mélanges de langue, de littérature et d'histoire des sciences médiévales offerts à Claude Thomasset*. Paris, Presses de l'Université de Paris-Sorbonne, 2005, p. 357-370.
- Godefroy F., *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, 1881–1902.
- Gorcy G., Martin R., Maucourt J., et Vienney R., “Le traitement des groupes binaires”, *Cahiers de Lexicologie*, n° 2, 1970, p. 15-46.
- Gouvert X., Le traitement étymologique des “francoprovençalismes” dans le *Trésor de la langue française*. Problèmes méthodologiques et étude de cas, in: Rézeau P. (éd.), *Richesses du français et géographie linguistique*, vol. 1, Bruxelles, de Boeck, 2007, p. 361-413.
- Greub Y., *Les Mots régionaux dans les farces françaises. Étude lexicologique sur le Recueil Tissier (1450–1550)*, Strasbourg, Bibliothèque de linguistique romane, 2003.
- Habert B., Traitements probabilistes et Corpus, *TAL* Vol 36- N°1-2, Paris, Hermès, 1995.
- Habert B., Portrait de linguiste(s) à l'instrument. *Texto!*, vol. X, n° 4, disponible sur: [http://www.revue-texto.net/Corpus/Publications/Habert/Habert\\_Portrait.html](http://www.revue-texto.net/Corpus/Publications/Habert/Habert_Portrait.html), 2005.
- Habert B., Nazarenko A. et Salem A. *Les linguistiques de corpus*, Paris, Armand Colin, 1997.
- Hansen M.-B. Mosegaard, From prepositional phrase to hesitation marker. The semantic and pragmatic evolution of French *enfin*, *Journal of Historical Pragmatics* 6, 2005, 37-68.
- Huguet=Huguet, E., *Dictionnaire de la langue française du seizième siècle* (7 vol.), Paris, Champion/Didier, 1925–1967.
- Imbs P. (éd.). *Lexicologie et lexicographie françaises et romanes. Orientations et exigences actuelles (Strasbourg, 12-16 Novembre 1957)*, Paris, Éditions du CNRS, 1961.

- Imbs P., Quemada B. (dir.). Trésor de la Langue Française. Dictionnaire de la langue du XIXe et du XXe siècle (1789—1960), 16 vol., Paris, Éditions du CNRS/Gallimard, 1971–1994.
- Laporte E., Mots et niveau lexical, TAL, vol. 38, n°2, 1997.
- LEI=Pfister M./Schweickard W. (dir.), *Lessico etimologico italiano*, Wiesbaden, Reichert, 1979–.
- Martin R., *Sémantique et automate*, PUF, Paris 2001.
- Martin R., Dictionnaire du Moyen Français (DMF) (1330–1500). Seconde version: DMF2. Présentation in: DMF (<http://atilf.atilf.fr/gsouvey/dmf2/PresentationDMF2.pdf>), 2007.
- Martin R./Gerner H./Souvay 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, forthcoming.
- Martin R./Souvay G., Quelle postérité électronique pour le *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, 2003, 393-403.
- Matthey A.-C./Nissille C., L'irruption de l'informatique dans la rédaction d'un dictionnaire historique: l'exemple du FEW (*Französisches Etymologisches Wörterbuch*), in: Iliescu M. et al. (éd.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique Romane (Innsbruck, 3-8 septembre 2007)*, Tübingen, Niemeyer, forthcoming.
- Pascual Rodríguez J. A./García Pérez R., *Límites y horizontes en un diccionario histórico*, Salamanca, Diputación de Salamanca, 2007.
- Petrequin G./Andronache M., Le projet TLF-Étym: apports récents de l'étymologie comparée-reconstruction, proposition de communication soumise au XIII<sup>e</sup> Congrès International d'Euralex (Barcelone, 15-19 juillet 2008), in progress.
- Pierrel J.M., *Ingénierie des langues*, Hermès Editions, 2000, 354 p.
- Pierrel J.M., Un ensemble de ressources de référence pour l'étude du français: TLFi, Frantext et le logiciel Stella, *Revue Québécoise de Linguistique*, volume 32/1, TAL Web et Corpus, p. 155-176, 2005.
- Pitz M., Pour une mise à jour des notices historiques consacrées aux emprunts à l'ancien francique dans le *Trésor de la langue française informatisé*, in: Buchi, É. (éd.), *Actes du Séminaire de méthodologie en étymologie et histoire du lexique (Nancy/ATILF, année universitaire 2005/2006)*, Nancy, ATILF, publication électronique ([http://www.atilf.fr/atilf/seminaires/Seminaire\\_Pitz\\_2006-11.pdf](http://www.atilf.fr/atilf/seminaires/Seminaire_Pitz_2006-11.pdf)), 28 p., 2006.
- Posner R., *Linguistic Change in French*, Oxford, Clarendon, 1997.

- Quemada B. (dir.), *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, 1970-1998.
- Radermacher R., *Le Trésor de la langue française. Une étude historique et lexicographique*, thèse Université de Strasbourg, 2004.
- Renders P., L'informatisation du FEW: quels objectifs, quelles possibilités?, in: Iliescu M. et al. (éd.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique Romane (Innsbruck, 3-8 septembre 2007)*, Tübingen, Niemeyer, forthcoming.
- Renders P., Modélisation d'un discours étymologique. Prolégomènes à l'informatisation du FEW (*Französisches Etymologisches Wörterbuch*), thèse en préparation à l'Université de Liège et à l'Université Nancy 2.
- Ringebach J.-L., 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, 2003, 191-206.
- Ringebach J.-L., 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, forthcoming.
- Rothwell W. et al., *Anglo-Norman Dictionary*, Londres, Maney Publishing/Modern Humanities Research Association (site internet: <http://www.anglo-norman.net>), 2005–<sup>2</sup>[1977–1992<sup>1</sup>].
- Souvay G., LGeRM: un outil d'aide à la lemmatisation du moyen français, in: Trotter D. (éd.), *Actes du XXIV<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes*, Tübingen, Niemeyer, 2007, vol. I, p. 457-466.
- Steinfeld N., 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, forthcoming.
- Steinfeld N./Andronache M., Quoi de neuf du côté de la lexicographie étymologique? La méthode utilisée dans le cadre du projet TLF-Étym pour distinguer les emprunts au latin de l'Antiquité de ceux faits au latin médiéval, in: Pierrel J.-M. et al. (éd.), *Actes du Colloque international "Lexicographie et informatique: bilan et perspectives" (Nancy, 23-25 janvier 2008)*, forthcoming.
- Tobler-Lommatzsch=Tobler A./Lommatzsch E./Christmann H. H., *Altfranzösisches Wörterbuch* (11 vol.), Berlin/Wiesbaden/Stuttgart, Weidmann/Steiner, 1925–2002.
- Wunderli P., *Du mot au texte: Actes du III<sup>e</sup> Colloque international sur le moyen français (Dusseldorf, 17-19 septembre 1980)*, Tübingen, Narr, 1982.