



**Mots. Les langages du politique**

**84 | 2007**

**Politiquement sportif**

---

## Le TLFi ou les infortunes de la lexicographie électronique

**Charles Bernet**

---



### **Édition électronique**

URL : <http://journals.openedition.org/mots/1052>

DOI : 10.4000/mots.1052

ISSN : 1960-6001

### **Éditeur**

ENS Éditions

### **Édition imprimée**

Date de publication : 1 juillet 2007

Pagination : 85-100

ISBN : 978-2-84788-112-7

ISSN : 0243-6450

### **Référence électronique**

Charles Bernet, « Le TLFi ou les infortunes de la lexicographie électronique », *Mots. Les langages du politique* [En ligne], 84 | 2007, mis en ligne le 01 juillet 2009, consulté le 20 avril 2019. URL : <http://journals.openedition.org/mots/1052> ; DOI : 10.4000/mots.1052

---

## **Le TLFi ou les infortunes de la lexicographie électronique**

*Aquila non capit muscas*<sup>1</sup>.

Aujourd'hui, le *Trésor de la langue française informatisé* (TLFi) est, sans aucun doute, beaucoup plus consulté que l'ouvrage imprimé dont il est issu. Il est certain que l'informatisation du *Trésor de la langue française*<sup>2</sup> lui a apporté à la fois un public élargi et des modes de consultation beaucoup plus souples que ceux que permet le maniement de lourds volumes imprimés.

Pour tout utilisateur, la découverte du TLFi inspire de l'admiration, surtout en raison des larges possibilités d'accès à un contenu d'une grande ampleur. À l'usage, toutefois, si l'on ne se cantonne pas à la fonction d'affichage la plus sommaire, article par article, il apparaît que bien des expériences se soldent par des déconvenues. Vient alors le temps des questions. Quelques-unes de ces questions sont à l'origine des observations qui suivent.

On se propose d'explorer plusieurs zones d'ombre et d'élucider les causes de certains résultats déconcertants. L'accent sera mis sur la matière du dictionnaire elle-même, du point de vue de son intégrité, de sa présentation et de sa structure, telle qu'elle figure dans la version imprimée et telle qu'elle est restituée dans le TLFi.

Le TLFi est un outil auquel peuvent recourir tous ceux qui travaillent dans le domaine des sciences du langage ou dans des disciplines connexes. La suite de l'article est une invitation à découvrir son contenu réel et quelques-unes de ses limites ou de ses imperfections afin d'en faciliter une utilisation rationnelle et maîtrisée.

Plusieurs versions du TLFi coexistent aujourd'hui ; l'une est accessible sur la Toile, une autre est commercialisée sous forme de cédérom<sup>3</sup>. Les différences

1. Devise de sir Francis Drake.

2. *Trésor de la langue française. Dictionnaire de la langue du XIX<sup>e</sup> et du XX<sup>e</sup> siècle*, publié sous la direction de Paul Imbs (vol. 1-7), puis de Bernard Quemada (vol. 8-16), Paris, Gallimard, 1971-1994.

3. Centre national de la recherche scientifique, Université de Nancy 2, *Analyse et traitement informatique de la langue française, Trésor de la langue française informatisé Cédérom du texte intégral et son livre d'accompagnement*, Paris, CNRS, 2004 (version PC) ; une version pour Mac OS X est diffusée depuis 2005.

essentielles entre ces versions concernent des fonctionnalités annexes. Les requêtes porteront sur ces deux versions, en privilégiant le cédérom pour PC.

## Contenu textuel

Les présentations du TLF*i*, que ce soit sur le site qui le diffuse ou dans la documentation d'accompagnement du cédérom, apportent des informations et des précisions sur certains aspects techniques de l'opération et sur la valeur ajoutée à l'occasion de ce changement de support, notamment par le balisage de certains « objets textuels », repérables par leur typographie dans la version imprimée et par la prise en compte de l'interdépendance hiérarchique entre ces objets. En revanche, on ne trouve pas la mention explicite d'interventions sur le texte lexicographique en tant que tel. On sait simplement que les premiers travaux préparatoires à l'informatisation ont consisté à restituer une archive informatique, aussi proche que possible de l'état imprimé des articles du *Trésor*.

Les débuts de l'informatisation du dictionnaire se situent en 1990<sup>4</sup>. Dans un premier temps accessible sur la Toile, la commercialisation de cédéroms à la fin de l'année 2004 et en 2005, soit au bout d'une quinzaine d'années, devait marquer pour ses promoteurs une étape avancée de l'opération, sinon le couronnement d'un projet parachevé.

Dans les premiers temps, sur le site de l'INaLF-Nancy, puis de l'ATILF, un paragraphe de présentation précisait que le texte livré sur la Toile reproduisait sans changements – entendons sans corrections, modifications ni mise à jour – celui de l'ouvrage sur papier. En 2002 ou 2003, la présentation est devenue : « Le TLF est un dictionnaire des XIX<sup>e</sup> et XX<sup>e</sup> siècles en 16 volumes et 1 supplément. Le TLF*i* est la version informatisée du TLF. »<sup>5</sup> Cette formulation amène le lecteur à conclure que le TLF*i* reprend le TLF enrichi de son supplément, ce qui n'est nullement le cas. En dépit des méprises<sup>6</sup> provoquées par cette formule, le texte de présentation est toujours inchangé en 2007, alors que le *Trésor* reste constitué de seize volumes depuis 1994, et que le supplément n'est ni publié, ni diffusé sous quelque forme que ce soit. Dans le volume d'accompagnement

4. « 1990-1991 : Lancement de la première étude d'informatisation du TLF dans les locaux du laboratoire nancéien de l'Institut national de la langue française [...]. 1992 : Acquisition des bandes de photocomposition des tomes 9 à 16 et premiers essais d'informatisation du tome 14 au laboratoire. 1993 : Début des travaux au service informatique ; un prototype fonctionne sur le volume 14. » (Dossier de presse réalisé à l'occasion de la présentation du *Trésor de la langue française informatisé*, Paris, CNRS, 5 mars 2002).
5. *Trésor de la langue française informatisé En savoir plus Présentation*. Disponible sur <http://www.atilf.fr/> (informations accessibles aussi à partir de <http://www.cnrtl.fr/dictionnaires/modernes/>). Cette présentation fallacieuse figure dans les mêmes termes sous la rubrique *Ressources linguistiques publiées – TLF : Trésor de la langue française*.
6. Exemple : « [...] le projet TLF constitue une réelle avancée qui se manifeste par dix-sept volumes papier et par une version électronique désormais disponible sur le web [...]. » (Condamines, 2003, p. 21.)

du cédérom, en revanche, on concède que le TLFi n'est que la réplique numérisée de son aîné<sup>7</sup>, mais, dans le même temps, la stratégie communicationnelle prend le relais en annonçant<sup>8</sup> : « *Le dictionnaire le plus complet par son nombre d'entrées et sa précision linguistique [...]* – l'intégralité de la version enrichie des 16 volumes contenus en un seul cédérom [...] – traitement exhaustif des mots du XIX<sup>e</sup> siècle et XX<sup>e</sup> siècle [...] ». Et ces propos sont repris, encore amplifiés, dans une publication scientifique : « L'intégralité des 16 volumes du Dictionnaire de référence de la langue française des XIX<sup>e</sup> et XX<sup>e</sup> siècles est réunie sur cédérom ! Enrichi d'une mise à jour complète, le TLFi constitue un instrument indispensable pour toutes les personnes qui pratiquent la langue française. »<sup>9</sup>

Les utilisateurs peuvent toutefois constater que de très nombreux mots et emplois qui se sont diffusés depuis le milieu du XX<sup>e</sup> siècle, et surtout après la rédaction du *Trésor*, ne sont pas traités dans le TLFi, tels *addiction*, *agro-alimentaire*, *communautarisme*, *entrepreneurial*, *flux tendu*, *illettrisme*, *langue de bois*, *nain de jardin*, *nettoyage ethnique*, *politiquement correct* ou *zéro défaut*. Les travaux en vue d'un supplément au *Trésor* ayant commencé en 1992 à l'INaLF-Nancy, il se pourrait que tout cela soit en attente d'être publié.

L'une des particularités de ce dictionnaire est de n'être, en apparence, l'œuvre d'aucun lexicographe. Le support numérique ne mentionne ni maîtres d'œuvre, ni équipe de rédaction, ni qui que ce soit qui ait apporté la plus modeste participation à son contenu. L'ouvrage d'accompagnement du cédérom, un peu moins discret, comporte une liste de noms, documentée et mise en forme avec dilettantisme et reléguée à la cent-cinquante-neuvième page.

Ce même ouvrage annonce que le TLFi comporte 54 280 articles et 92 997 entrées<sup>10</sup>. Le nombre des entrées est celui des unités lexicales traitées dans le dictionnaire. Les articles sont les « entrées principales » ou mots traités en « vedettes autonomes » du dictionnaire imprimé. Il y aurait donc 38 217 « entrées secondaires », entendons par là les mots figurant en sous-entrée<sup>11</sup>, c'est-à-dire sous des affixes, des éléments formants, dans des rubriques intitulées *Dér[ivé(s)]* ou dans des *REM[arques]*. Cette indication à l'unité près donne à entendre que ce dénombrement n'est pas une estimation mais le décompte précis d'entrées identifiées comme telles. Mais le nombre réel des vocables

7. « Si cette version informatisée correspond à un reflet fidèle de la version du TLF précédemment publiée, notre objectif est [...] d'y intégrer [...] des améliorations... » (Jean-Marie Pierrel, « Préface », p. 4).
8. « *Trésor de la langue française informatisé TLFi – version PC* ». Disponible sur <http://www.cnrsditions.fr/> ; repris dans « Après sa sortie sur PC, le TLFi version Mac OS X arrive ! » Disponible sur <http://www.tlfi.fr/>.
9. *Marges linguistiques. Revue semestrielle électronique en sciences du langage*, n° 9, mai 2005. Accessible à l'adresse <http://www.revue-texto.net/>.
10. Indications figurant p. 161.
11. L'aide en ligne qualifie ces entrées d'articles *cachés* : « [...] le TLF a la particularité de comporter un certain nombre d'articles *cachés* à l'intérieur d'autres articles. Par exemple, le mot MYXOMATOSE est défini dans l'article consacré à MYXOME. »

traités dans le *Trésor* est toujours inconnu car le repérage des entrées, effectué en vue de l'informatisation au moyen de critères élémentaires automatisables, n'a permis de reconnaître que les cas les plus simples. La plus grande partie des mots qui ne sont pas encore identifiés comme des sous-entrées figurent dans des *Remarques* dont l'exploration reste à faire car la typographie de ces rubriques n'a pris une forme identifiable sans ambiguïté par de simples automates qu'à partir du milieu du t. 8 de la version imprimée. En deçà de cette borne, il n'est pas possible de faire le tri entre les différents types de remarques en s'en tenant à des critères formels et sans analyse du contenu. Dans une fenêtre d'aide intitulée « Les types d'objets du TLF », il est indiqué : « Environ 50 % des vedettes du TLF sont secondaires », ce qui correspond à un nombre voisin de 54 000 ; si cette estimation était vérifiée, les entrées non encore identifiées comme telles seraient au nombre de 15 000, ce qui est considérable<sup>12</sup>.

Dans le premier menu du TLF*i*, intitulé « recherche d'un mot », une liste déroulante permettant d'accéder au contenu du dictionnaire est assortie d'une « info. importante » précisant à l'utilisateur que cet ensemble ne reprend « que les entrées principales ». Il est étonnant qu'après avoir fait le choix de présenter les articles du dictionnaire comme le ferait une banque de données constituée de mots, et après avoir mis au jour quelques milliers d'entrées secondaires « cachées », les listes déroulantes ne reprennent qu'une partie des entrées du dictionnaire.

L'évolution historique du *Trésor* a voulu que, pour économiser l'espace typographique, des mots, de plus en plus nombreux au fur et à mesure de l'avancement du dictionnaire, aient été traités dans des entrées secondaires. Il en résulte une nomenclature des entrées principales en accordéon, qui va d'une grande richesse dans les premières lettres de l'alphabet à des listes plus pauvres que certains dictionnaires courants vers la fin de l'alphabet. De nombreux mots fréquents et bien intégrés dans la langue ne sont pas en entrée principale (c'est le cas, en priorité, d'adverbes en *-ment*, mais aussi de substantifs et d'adjectifs en grand nombre comme *spéléologue*, *standardiste*, *stéréophonique*, *structuraliste*, *surpeuplement*, *tabagique*, *terrarium*, *totémique*, etc.). Tout cela n'a qu'une importance secondaire dans la version imprimée du dictionnaire, dans laquelle l'utilisateur repère sans peine ces mots sous une vedette morphologiquement apparentée. En revanche, l'utilisateur de la version informatisée, même prévenu, sera surpris de ne pas trouver ces vocables très courants dans les listes déroulantes.

Le TLF étant transformé en une manière de base de données, dans laquelle la notion de page n'a plus de pertinence, se pose le problème du libellé de

12. Dans une récente version simplifiée du TLF*i* (<http://www.cnrtl.fr/lexicographie/>), les requêtes ne portent que sur les entrées identifiées comme telles. Cet outil ne donne donc accès qu'à une partie de la nomenclature du dictionnaire. Noter aussi que, dans cette version, certains vocables figurant en « entrées principales » sont inaccessibles.

certains renvois internes du *Trésor*. Dans la version électronique, les renvois à des tomes et à des pages sont sans objet (ex. s.v. *-odonte* : « v. *anodonte* ds TLF t. 3, p. 82 » ou *flancon(n)ade* : « suff. *-ade*\* v. TLF t. 1, p. 648a »). On trouve, de même, des renvois à la *Liste des textes dépouillés du Trésor*, qui n'est pas reprise dans le TLFi (p. ex. s.v. *poissard*<sup>13</sup>).

Il y a aussi, dans le texte lexicographique proprement dit, des citations extraites de la préface du t. 1 (ex. 2 s.v. *colloque*, ex. 10 s.v. *condition*, ex. s.v. *diachronie* B2, ex. s.v. *entrée* II B2, etc.), ce qui met en évidence l'absence, dans le TLFi, d'éléments du paratexte qui, du coup, deviennent des données externes au dictionnaire informatisé. En effet, la version électronique ne restitue pas les préfaces ou textes liminaires de Paul Imbs (t. 1, 2, 3, 4, 5 et 7), ni un texte liminaire (t. 8) et la postface (t. 16) de Bernard Quemada. Tous ces documents sont pourtant riches d'informations sur la mise en place et sur l'évolution des pratiques rédactionnelles du *Trésor*, ils appartiennent à son histoire et contribuent à en expliquer les modifications. On voit par là que l'informatisation du *Trésor* n'a pas consisté, comme cela se fait pour les dictionnaires anciens, à restituer l'intégralité de l'œuvre originale. Ce choix montre que le TLFi n'est pas destiné à répondre à des objectifs strictement patrimoniaux.

Il faut mentionner aussi le maintien de renvois à des index non repris dans le TLFi, par exemple s.v. *-opsie* : « v. *ampélopsis* (TLF t. 7, p. 1324) ». Il se trouve que ces index, placés à la fin de plusieurs volumes du *Trésor*, comportent des corrections et des ajouts que la version informatisée avait précisément pour vocation de réintroduire dans le dictionnaire proprement dit. Le mot *ampélopsis* (nom de la « vigne vierge »), qui est enregistré dans tous les dictionnaires courants et attesté notamment chez Marcel Proust<sup>13</sup>, est bien traité dans le *Trésor* mais ne figure donc pas dans le TLFi. Ne figurent pas non plus d'autres ajouts ni des entrées réellement cachées, et en particulier la locution *faire chanter qqn* « exercer un chantage sur qqn », qui est traitée dans l'index du t. 6 ; il en résulte que cette expression bien connue figure dans tous les dictionnaires du français contemporain, à l'exception du TLFi.

Par son contenu, le TLFi n'est donc pas à proprement parler la restitution fidèle et intégrale du *Trésor* imprimé.

## Restitution du texte

La réalisation d'une version informatisée sous la forme d'une simple réplique de l'imprimé est, toutes questions techniques mises à part, une opération sans difficultés apparentes, mais qui n'est pas sans risques. En effet, un transfert de support, surtout lorsqu'il s'accompagne d'interventions manuelles, est la source d'inévitables erreurs.

13. Dans *Du côté de chez Swann* (avec la graphie *empelopsis*).

On ne s'attardera pas sur des fautes ponctuelles telles que *température dispensé* (s.v. *chaleur*) ou *opératiof* (s.v. *opération*), ni sur des défauts de mise en forme provoqués par la substitution d'une référence explicite aux indications<sup>14</sup> *op. cit.*, *loc. cit.*, *id.* ou *ibid.* (espaces manquants, successions de virgules, doublons lexicaux, etc.).

Parmi les coquilles sérielles, l'une des plus curieuses est le remplacement de la lettre *x* par un point : *.IX<sup>e</sup>* pour *XIX<sup>e</sup>* (s.v. *loup* et *onglette*), *.X<sup>e</sup>* pour *XX<sup>e</sup>* (s.v. *poulbot*), *di.ain* (s.v. *majeur*) ou *e.près* (s.v. *ordre*) ; ce syndrome affecte aussi une série de formes en *-eux* comme *musculeu.*, *officieu.*, *vieu.* ou *yeu.* Un autre cas remarquable est celui de *Æ* et *œ*. Dans une tranche alphabétique qui s'étend de la fin de la lettre *l* à la fin de la lettre *p*, les digrammes ligaturés *Æ* et *œ* ont été dissociés en *OE* et *oe*. La substitution n'a pas de répercussions sur le résultat des requêtes qui portent sur le texte en caractères latins. En revanche, le remplacement de *œ* par *oe* dans des transcriptions phonétiques en code API est particulièrement fâcheux. Cette anomalie prend un tour pleinement matériel dans la version du TLFi diffusée sur la Toile, qui propose une rubrique intitulée « Prononcer ». Dans son fonctionnement normal cette rubrique, présentée comme expérimentale, donne une illustration sonore de la vedette au moyen d'un système de synthèse de la parole. La restitution de mots appartenant à la tranche alphabétique concernée donne des résultats inouïs. Tous pourraient figurer dans des tests auditifs de discrimination des voyelles et plusieurs rappellent les mégapneumies de la poésie lettriste ; entre autres : [loen(t)ʃ] (*lunch*), [moe:Rs] (*mœurs*), [œvRet] (*œuvrette*) ou [œjdøboef] (*œil-de-bœuf*), séquence dans laquelle la même erreur figure deux fois.

Le recours à des outils informatiques adaptés aurait permis de corriger sans peine des fautes de la version imprimée<sup>15</sup> comme *assujeti* pour *assujetti* (s.v. *impôt*), *ommission* pour *omission* (s.v. *être<sup>1</sup>* et *fait*), *sensibilité* pour *sensibilité* (s.v. *lysotype*, *phénomène*, etc.), *sprit* pour *esprit* (s.v. *être<sup>2</sup>*), ou même des noms propres comme *Mitterand* pour *Mitterrand* (s.v. *durcissement*, *nous*, *nucléarisation*, *patrie*, *politique-fiction*, *préfectoral*, etc.) ou *Wilcok* pour *Hitchcock* (s.v. *policier*).

Une informatisation moins asservie au texte lexicographique imprimé aurait pu réorganiser partiellement le contenu de certains articles en mettant à profit la possibilité d'un agencement non linéaire et non séquentiel, en établissant des liens entre des informations dispersées et en substituant des liens à des données répétées.

En effet, certaines lexies complexes sont enregistrées à plusieurs reprises ; c'est par exemple le cas de *vin délimité de qualité supérieure*, pour lequel un

14. Ces notations abrégées ont été remplacées par des indications en clair dans une partie de la copie.

15. Il n'aurait pas été superflu de tirer parti des comptes rendus publiés au fur et à mesure de la sortie des volumes du *Trésor* tels que ceux de Robert-Léon Wagner et de Henri Cottez parus, de 1973 à 1980, dans le *Bulletin de la Société linguistique de Paris*.

renvoi à l'article *vin*, sous chaque entrée consacrée aux autres vocables, aurait été préférable au maintien du traitement actuel sous deux entrées distinctes (s.v. *supérieur* et *vin*; le sigle *VDQS* étant traité dans l'entrée consacrée à la lettre *V*).

C'est aussi le cas d'unités lexicales comme le verbe *saccageoter* :

s.v. -OTER, -OTTER : *saccageoter*, verbe trans. *Empl. abs.* Se livrer à de petits saccages. *Nous vivotions du côté de Chalon-sur-Saône, pillotant, saccageotant* (ARNOUX, *Rhône*, 1944, p. 305).

s.v. SACCAGER : **Saccageoter**, verbe trans., hapax. *Nous vivotions du côté de Chalons-sur-Saône [sic], pillotant, saccageotant* (ARNOUX [...]).

La reproduction du même exemple sous des vedettes différentes est plus comode pour le lecteur que des renvois dans un dictionnaire en plusieurs volumes. Dans un dictionnaire informatisé, en revanche, il serait de règle de débusquer ces répétitions inutiles et de gérer des renvois.

Les exemples sont parfois reproduits à l'identique, mais il arrive souvent que la taille ou le découpage du contexte varie. Ainsi, une citation de Vallès (*L'Enfant*, p. 18), est mise à contribution à trois reprises, d'abord dans une version longue, s.v. *bachelier*, et à deux reprises sous une forme plus courte, s.v. *compagnon* (ex. 10) et *devoir*<sup>3</sup> (ex. 2).

Il est inutile de développer ce point davantage. Des liens vers des exemples bien documentés et d'une longueur optimale auraient été préférables au maintien de contextes répétés, lorsqu'ils sont brefs et opaques.

Plusieurs autres anomalies, aisément repérables, auraient justifié une mise à niveau des citations, facilitée par l'existence de liens hypertextuels entre le TLFi et Frantext. Sachant que plusieurs œuvres avaient été référencées de façon fautive dans les documents mis à la disposition de la rédaction du *Trésor*, et qu'une liste des références suspectes donne toutes les indications utiles, il aurait suffi, pour éliminer les erreurs résiduelles, de rechercher dans le TLFi les citations tirées de ces œuvres et de les confronter à la base Frantext, dans laquelle les indications correctes ont été rétablies. Les deux exemples suivants, dont le lecteur peut percevoir sans analyse le caractère insolite, appartiennent à cette série :

[...] je ne t'ai pas écrit hier au soir parce que nous avons eu une algarade terrible à propos de ma mère. Elle a été empoisonnée par des huîtres [...] (V. HUGO, *Les Contemplations*, t. 3, 1856, p. 235); ex. 6 s.v. *algarade*;

J'envisage avec effroi les ennuis qu'entraînera pour moi la publication de cette ode (LEGOUVÉ, *Mort Henri IV*, acte V, scène 5, 1806, p. 177); ex. s.v. *effroi*.

En fait, la première citation est tirée de la *Correspondance* de Gustave Flaubert et la seconde des *Lettres à la fiancée* de Victor Hugo.



L'utilisateur peut être amené à constater que le contenu actuel de Frantext ne restitue pas toujours les mêmes contextes que le dictionnaire. En effet, il est arrivé, pour certains textes, qu'une édition soit substituée à une autre ; c'est le cas, par exemple, pour l'œuvre de Balzac<sup>16</sup>. Par ailleurs, la base s'est enrichie d'ouvrages qui, dans le *Trésor*, ont été cités à partir d'autres sources, et particulièrement à partir de *l'Inventaire général de la langue française*, avec quelques différences. On peut découvrir des divergences de toute nature : absence du mot traité, du passage cité, graphies différentes, modifications du contexte, etc. Ainsi :

J'ai ôté d'instinct mon chapeau – pour saluer le drapeau noir [...] étendard des canuts, bannière de la Guillotière! (VALLÈS, *J. Vingtras*, Bachel., 1881, p. 342). [s.v. *drapeau*].

Ce texte correspond à l'édition Charpentier de 1881 revue par Jules Vallès ; dans Frantext, on trouve une leçon différente<sup>17</sup> : en fin de phrase *la Croix-Rousse* remplace *la Guillotière*.

## Maintien de la typographie de l'imprimé

Le dictionnaire électronique reproduit, pour les vedettes, les caractéristiques typographiques du dictionnaire imprimé. Il présente donc certaines variations morphologiques ou des variantes formelles de manière synthétique, sans reproduire chaque forme en toutes lettres : *fédéral*, *ale*, *aux* ; *constitution(n)alisme* ; *auto(-)détermination*. Les formes ainsi obtenues peuvent conduire, lors des recherches, à des résultats aberrants.

Si nous cherchons le mot *acte*<sup>18</sup>, les résultats de la requête sont les entrées : ACTE<sup>1</sup>, subst. masc. ; ACTE<sup>2</sup>, subst. masc. ; ACTE<sup>3</sup>, subst. masc. ; COMPACT, ACTE, adj. ; INEXACT, -ACTE, adj. et INTACT, -ACTE, adj. C'est-à-dire que toute séquence de caractères correspondant à la forme demandée, qu'il s'agisse d'un mot autonome ou d'une finale, est considérée comme une réponse pertinente. Il aurait suffi, pour ces vedettes, de développer intégralement la forme féminine pour éviter cet inconvénient<sup>19</sup>.

Lorsque les variations graphiques des entrées sont notées entre parenthèses, les formes identifiées ont bien fait l'objet, dans le TLFi, d'une réécriture

16. Il y a aussi le cas, rarissime, du maintien dans Frantext d'une édition utilisée dans le *Trésor* malgré l'ajout d'une édition plus récente (p. ex. *La Légende des siècles* de Victor Hugo).

17. L'édition saisie reproduit le texte corrigé et enrichi des *uvres de Jules Vallès* sous la direction de Lucien Scheler (les Éditeurs français réunis, impr. 1955).

18. Par l'un des deux procédés suivants : « recherche d'un mot » ou « recherche assistée d'une «vedette égale ou contenant un mot donné» ».

19. Avec le mot *acte*, le nombre de réponses incorrectes est limité ; si le choix se porte sur un mot comme *elle*, ce nombre est supérieur à quatre cents (v. *actuel*, *elle* ; *constitutionnel*, *elle* ; *oppositional*, *-elle* ; *présidentiel*, *-elle* ; etc.).

développée<sup>20</sup>. Il aurait été nécessaire, dans le même temps, de neutraliser les possibilités de recherche sur les formes comportant des parenthèses, car dans ce cas, le logiciel traite chaque séquence située hors des parenthèses, ou dans les parenthèses, comme un mot autonome. Ainsi, la vedette CHEF(F)ESSE constituera l'une des réponses aux requêtes portant sur *chef*, sur *fou* sur *esse*, et PAS(S)ION(N)ARIA sur cinq entrées du dictionnaire : *pas*, *s*, *ion*, *n* et *aria*.

On retrouve ce phénomène de « scission » du mot en plusieurs séquences lorsqu'il est le lieu de changements d'attributs de caractères, ainsi l'inaccessible *abribus*, imprimé partiellement en italique et partiellement en gras, que l'utilisateur ne trouvera pas s'il formule sa requête en écrivant le mot entier. S'il lui vient à l'idée de chercher, avec le menu « recherche assistée », les deux formes autonomes *abri* et *bus* dans un « paragraphe quelconque », il finira par le localiser sous l'entrée *-bus*. Il semble y avoir normalement deux traitements différents selon que le mot est, ou n'est pas, une unité de traitement lexicographique. On observe ainsi, sous l'entrée *-able*, *-ible*, *-uble*, les deux cas pour *ministrable*. Le logiciel l'identifie bien comme tel dans la partie « synchronique » de l'article : *ministrable* « qui peut devenir ministre », mais dans la rubrique *Étymol. et Hist.* : « *ministrable* (1936) »<sup>21</sup>, le logiciel n'accède au mot que par des requêtes portant soit sur *ministr* soit sur *able*.

Les caractères spéciaux, ou rares en français, comme ñ ou Ñ ne sont pas reconnus par le logiciel et sont toujours traités comme des séparateurs. Une requête portant sur *do*, avec la fonction « recherche d'un mot » aura pour résultats « *do*, subst. masc. » et « *doña*, subst. fém. » ; une requête portant sur l'entrée *or* aura, parmi ses résultats « *or*, subst. masc. » et « *señor*, subst. masc. ».

On rencontre les mêmes difficultés dans la recherche de lexies complexes ou de locutions (« syntagmes définis » ou « syntagmes » dans la terminologie du TLFi). Le logiciel impose que la séquence recherchée soit saisie exactement telle qu'elle figure dans le dictionnaire. Ainsi, les requêtes portant sur *accord amiable*, *calamités agricoles*, *persona non grata*, *privilège royal*, *programme de recherche* ou *violation de territoire* aboutissent toutes à une impasse car leur présentation typographique comporte des parenthèses ou une barre oblique : *accord (à l') amiable*, *calamité(s) agricole(s)*, *persona grata/non grata*, *privilège (du roi/royal)*, *programme (de recherche(s))* ou *violation (de territoire)*. Pour contourner ces désagréments, l'utilisateur devra chercher non pas des séquences entières, mais des « syntagmes définis » comportant à la fois *accord* et *amiable*, *calamité* et *agricole*, etc., au risque d'ajouter des réponses parasites.

L'usage massif d'abréviations est caractéristique des ouvrages de lexicographie. Cela est justifié, dans une publication, pour des raisons d'économie.

20. Ex. : EST(H)ONIEN, IENNE, (estonien, esthonien) ; les vedettes en italique ont été ajoutées dans la version informatisée ; seules les formes masculines ont été développées.

21. Ce mot est traité par ailleurs en vedette autonome, avec une première attestation moins tardive (1885).

Dans un dictionnaire informatisé, il était envisageable de développer ces abréviations sélectivement ou dans leur ensemble. Le TLFi a préféré opter pour une gestion par le biais du logiciel. La version du dictionnaire sur cédérom, et elle seule, propose à l'utilisateur une table à laquelle on accède par un bouton intitulé «Abrév. et bibliog.» [sic]. Celle-ci regroupe des abréviations (usuelles ou propres au métalangage des lexicographes) et des références bibliographiques abrégées. La « bibliog. » correspond non pas à l'ensemble des ouvrages cités mais à ceux que la version sur papier du *Trésor* réunit sous l'étiquette *Études fréquemment citées*. Les rubriques bibliographiques sont généralement d'une incontestable précision. Leur mise en forme est plus décevante. Pour ne prendre que l'exemple du *Littré*, les mentions *Littré*, *Littré Add.*, *Littré Suppl.* et *Littré Suppl. Add.* renvoient à une fenêtre unique dans laquelle sont groupées quatre éditions fort dissemblables de ce dictionnaire parmi lesquelles l'utilisateur doit faire lui-même son choix.

Le décodage des abréviations et de la bibliographie ne se fait pas directement (par un clic droit de la souris, par exemple), mais par l'ouverture laborieuse d'une série de fenêtres successives.

Le contenu de cette liste laisse beaucoup à désirer : coquilles<sup>22</sup>, absence de majuscules à l'initiale des noms propres et des substantifs en langue allemande<sup>23</sup>, lacunes multiples – certains noms d'auteurs abrégés sont retenus alors que d'autres ne le sont pas<sup>24</sup>; de très nombreuses abréviations ne sont pas mentionnées : **a. fr.** « ancien français », **c.-à-d.** « c'est-à-dire », **ch. de fer** « chemin de fer », etc. L'équivalence donnée pour certaines abréviations est problématique, c'est le cas de « **cr.** croate » qui n'apparaît en tant que telle que sous quelques entrées (« serbo-cr. », s.v. *charivari*, *guzla*, etc.), alors que **cr** est aussi l'abréviation, bien plus fréquente, de *compte rendu*.

Certaines abréviations sont inconnues ou fantaisistes (« **phm.** philémon »), d'autres ne sont jamais attestées dans le dictionnaire (« **dépl.** dépliant », « **erw.** erweitert », etc.) ; on trouve enfin dans la liste des caractères incorrects (« **esp.** espa(ol) », « **fø.** folio », « **nø.** numéro », etc.).

Dans le fonctionnement du logiciel (sur cédérom ou sur la Toile), les abréviations et les termes abrégés correspondants sont dans une relation symétrique. En d'autres termes, toute requête, qu'elle soit formulée sous l'une ou l'autre forme, donne pour résultat l'ensemble des attestations des deux formes<sup>25</sup>. En pratique, elle donne des excédents pour plusieurs raisons<sup>26</sup>. La première est

22. « **bibliohécon.** bibliothéconomie », « **mozar.** mozarable », etc.

23. « **exup.** exupéry », « **flaub.** flaubert, flaube [sic] », « **wiss.** wissenschaft ».

24. Parmi les nombreux absents : **Fustel de Coul.** (Fustel de Coulanges), **Lautréam.** (Lautréamont), **Martin du G.** (Martin du Gard) ou **Robesp.** (Robespierre).

25. Pour la forme abrégée, les effectifs sont évidemment différents selon que la requête est formulée avec ou sans le point abrégatif.

26. Ne nous attardons pas sur les abréviations qui figurent bien dans la liste mais ne sont pas traitées comme telles par le logiciel. C'est, entre autres, le cas de *souv.* (*souvent*), dont l'omission

que les abréviations n'étant pas identifiées (ou balisées) en tant que telles, toute forme homographe est considérée comme une attestation de cette abréviation. En cherchant les définitions comportant *fin*. (pour *finances* ou *financier*), près de deux résultats sur trois sont des définitions se terminant par *fin* (adj. ou subst.) ; en cherchant *luxembourgeois* (abrégé en *lux.*) dans des exemples, le logiciel propose « fiat lux » (s.v. *abîmer*, *audace*, etc.), le « certificat de lavabilité "Lux" » (entrée *-té*, *-eté*, *-ité*) ou « Guy Lux » (s.v. *guili-guili*). La seconde raison est qu'il peut exister, dans un autre environnement, des formes abrégées homographes non répertoriées. En cherchant le mot *marine*, dont l'abréviation *mar.* est surtout employée dans la désignation d'un « domaine technique », le logiciel donne des résultats inattendus dans des titres tels que : BALZAC, *Mém. jeunes mar.*, TOULET, *Mar. Don Quichotte* ou MUSSET, *Caprices Mar.* où l'abréviation est mise pour *mariés*, *mariage* et *Marianne*.

Ici, le meilleur remède est probablement de s'abstenir de formuler des requêtes sur les abréviations reconnues par le logiciel. L'utilisateur du cédérom peut en consulter la liste et tenter d'agir en conséquence ; en revanche, l'utilisateur de la version disponible sur la Toile, ne disposant pas d'indications qui lui permettraient de se tirer d'affaire, en est réduit à pester contre les mystères de l'informatique.

## Deux « objets textuels »

L'aide du TLFi indique : « Les articles du TLF, comme ceux de tout dictionnaire, sont constitués d'un certain nombre d'"objets textuels" dont la nature et l'organisation sont familières au lecteur. » Ces objets élémentaires vont de l'entrée aux exemples en passant par la structure de l'article, les constructions grammaticales, plusieurs types d'indicateurs, les définitions, etc. Certains ont été évoqués plus haut, comme les « entrées principales », les « syntagmes » ou les « syntagmes définis ». On s'interroge sur leur nombre exact<sup>27</sup>. La principale rubrique du dictionnaire, qui décrit les différents sens des mots attestés aux XIX<sup>e</sup> et XX<sup>e</sup> siècles, a fait, seule, l'objet d'un balisage réputé complet. Nous examinerons successivement le traitement de deux objets textuels qui, à première vue, n'ont rien de problématique : le « titre d'exemple », entendons par là le titre d'un ouvrage duquel est extrait un exemple, et la « langue empruntée », ou, en clair, la langue d'origine d'un emprunt.

La plupart du temps, lorsque l'exemple est tiré du corpus textuel réuni pour le dictionnaire, le « titre d'exemple » correspond au titre de l'ouvrage. Dans

---

évit en l'occurrence des confusions avec la même forme employée pour *souvenir(s)* (p. ex. Gide, *Souv. Cour d'ass.*, Martin du G., *Souv. autobiogr.*, Renan, *Souv. enf.* ou Stendhal, *Souv. égotisme*).

27. « En tout, plus d'une quarantaine d'objets différents » (Bernard *et al.*, 2002, p. 139) ; dans l'aide du TLFi, une fenêtre intitulée « Les types d'objets du TLF » n'en mentionne que trente.

le *Trésor*, la formulation des titres varie selon le statut de l'exemple (placé à la suite de la définition ou dans les pavés réservés aux citations), et selon la tranche alphabétique du dictionnaire. Pour économiser l'espace typographique de l'ouvrage imprimé, les références bibliographiques ont pris, avec le temps, des formes de plus en plus synthétiques. Ainsi, pour les *Considérations sur les principaux événements de la Révolution française* de Madame de Staël, on rencontre, outre la forme intégrale : *Considérations sur les princ. événements de la Révolution fr.*, *Consid. Révol. fr.*, *Consid. Révol.*, *Consid. Rév. fr.* et *Consid.* ; pour *L'Ancien régime et la Révolution*, de Tocqueville, on trouve, en plus du titre *in extenso*, *Anc. Régime et Révol.*, *Ancien Rég. et Révol.*, et *Anc. Rég. Révol.* Avec de telles variations de forme, il devient quasiment impossible de sélectionner, dans une même requête, l'ensemble des exemples tirés d'une œuvre. Lorsque ces variations se combinent avec des abréviations mal décodées, les résultats peuvent être plus déroutants : dans le titre précédent, l'abréviation adoptée pour *régime*, *rég.*, est, de fait, reconnue comme l'équivalent de *régulier* ; une requête sur les « titres d'exemples » comportant le mot *champ*, dont l'abréviation est *ch.*, donne une liste dans laquelle on trouve des œuvres aussi inattendues que : Bailleul, *Notions de matériel roulant des ch. [chemins] de fer* ; Goncourt, *Ch. [Charles] Demailly* ; Labiche, *Club champ. [champenois]* ; Mirbeau, *Journal femme ch. [chambre]* ou J. Monnerot, *Les Faits sociaux ne sont pas des ch. [choses]*. On comprend mal que, dégagée des contraintes d'espace typographique de la publication sur papier, la version informatisée ait conservé, dans leur diversité, toutes ces abréviations chargées d'ambiguïté.

Lorsque l'exemple est tiré d'une revue, d'une publication collective, d'un dictionnaire, d'une compilation ou d'ouvrages étrangers au corpus du *Trésor*, on trouve dans l'objet textuel « titre d'exemple » des données très diverses. Ci-dessous, le titre d'exemple est entre accolades et l'auteur entre crochets :

[A.-G. CIMARELLI,] {*Armes blanches et armes défensives*, adapté par C. Bosson, Paris, Grange Batelière}

[W. FREYHAN,] {*Retour à la Tradition*, trad. de l'all. par A. Dreyfus}

[DE GAULLE] {ds *Paris-Match*, 1<sup>er</sup> mai 1965, n° 838 ds RHEIMS}

[V. GISCARD D'ESTAING,] {*Lettre du 20 déc. 1976* ds S. NORA, A. MINC, *L'Informatisation de la société*, Paris, éd. du Seuil}

[NICOLAS ds] {G.-H. Roger, F. Widal, J. Teissier, *Nouveau traité de médecine*, fasc. 4}

[TAINÉ,] {*De l'Intelligence*, livre II, ch. 1, § 1 ds LAL. 1968}

Le « titre d'exemple » peut comporter des noms d'éditeurs, de traducteurs, d'auteurs, des noms d'ouvrages, de revues, de lieu d'édition, des mentions de langue d'origine, de chapitres ou des dates, toutes indications qui n'ont

rien à voir avec ce que l'on entend par *titre*. Cela a des incidences sur certaines requêtes. En cherchant les exemples tirés d'œuvres dont le titre comporte le mot *minuit*, on identifiera bien des attestations de titres comme *La Confession de minuit* de Georges Duhamel ou *Le Démon de minuit* de Hervé Bazin, mais s'ajouteront plus d'une quarantaine d'exemples tirés d'ouvrages publiés par les Éditions de Minuit. La même expérience donne des résultats encore plus massifs avec un nom de ville comme *Paris* (6863 ex.). On voit que le contenu peu homogène du « titre d'exemple » peut, dans bien des cas, décourager toute recherche.

On observe aussi que le balisage des différents objets textuels entrant dans les références bibliographiques n'est pas toujours cohérent.

Dans l'article *kitsch*, deux citations sont reprises au *Dictionnaire des mots contemporains* (Gilbert, 1980) ; la fonction du TLFi « peindre les objets »<sup>28</sup> permet d'en identifier les objets « publication » et « date d'exemple » dont les contenus diffèrent dans les deux cas :

- {*Elle*, 1<sup>er</sup> mars 1971 ds GILB.} {*Mots contemp.*} {1980} ; les accolades encadrent ici {la « publication »}, {le titre} et {la date}.
- {*Paris-Match*,} {13 nov. 1971} ds GILB. *Mots contemp.* 1980 ; les accolades encadrent {la « publication »} et {la date}.

Le plus souvent, le balisage de la « date d'exemple » reproduit la date du document initial, mais les exceptions sont nombreuses.

Une autre question importante, que nous ne ferons qu'évoquer ici, est celle de la datation des textes cités dans le *Trésor*. Il n'est pas rare qu'un même texte soit affecté, selon les endroits, de datations différentes. Dans de nombreux cas, il s'agit de corrections effectuées par l'équipe de rédaction du *Trésor* au fil du temps. À défaut d'explications, certains cas restent énigmatiques pour les utilisateurs du TLFi. Celui de *L'Itinéraire espagnol* d'Albert T'Serstevens, cité plus de huit cents fois dans le *Trésor*, est de ceux-là. Ce texte est daté de 1933 au début du dictionnaire jusqu'à la vedette *arbre* (proche de la fin du t. 3), et de 1933 ou 1963 à partir d'*architrave*, la seconde date étant attribuée un peu plus fréquemment que la première. Les cas de double datation de cette œuvre à l'intérieur d'un même article ne sont pas rares (v. *distribuer*, *motif*, *nomade*, *plateresque*, *polychromie*, *sur*, *vigne*<sup>2</sup> et *volant*<sup>3</sup>).

Le *Trésor* étant un dictionnaire de culture largement fondé sur un corpus, sa version informatisée aurait dû comporter une liste complète des œuvres citées. Grâce à l'informatique, il aurait été possible, de surcroît, de déceler des incohérences, de repérer des coquilles<sup>29</sup> ou des erreurs et de gérer ces informations de façon souple et dynamique.

28. Une bizarrerie de cette fonction veut que la liste des objets comporte 16 items avant d'activer la touche « valider » et 17 après validation. L'objet additionnel est précisément la « publication ».

29. Les références bibliographiques des œuvres étrangères au corpus initial, n'ayant pas fait l'objet des mêmes vérifications que les autres, sont plus exposées aux erreurs : HARDELET pour HARDELLET

L'objet textuel « langue empruntée » apparaît dans la rubrique *Étymol. et Hist.* et permet des requêtes sur la langue d'origine des emprunts (ou des xénismes) figurant à la nomenclature du TLFi. Un moyen simple d'en souligner l'intérêt, la pertinence et les limites sans longue démonstration est d'examiner les résultats obtenus pour une langue à laquelle le français n'a fait des emprunts qu'en nombre limité.

À l'aide de la fonction « recherche assistée », on active l'objet textuel « langue empruntée » appliqué au japonais et l'on obtient une liste très succincte, qui comporte en tout et pour tout trois entrées : *netské*, *netsuké*, *netzké*, subst. masc. ; *roténone*, subst. fém. et *saké*, *sakî*<sup>1</sup>, subst. masc. Même en misant sur un nombre de réponses limité, un résultat aussi modeste est bien en deçà de ce que l'on pouvait attendre. On observe que le point commun à ces trois articles est de comporter, dans la rubrique *Étymol. et Hist.*, la formule « Empr. au japonais » et il semble bien que, pour cette fonction, le logiciel n'identifie que les termes mentionnés dans une phrase commençant par « Empr. » Une analyse des rubriques *Étymol. et Hist.* montre que ce critère unique n'est pas la formule la mieux adaptée. En effet, la mention « empr. au ... » n'est pas toujours en tête de phrase et les emprunts sont très souvent mentionnés d'une autre manière. Pour la langue choisie, en effectuant une recherche sur la séquence « mot japonais », on obtient une liste de trente-deux vocables dans laquelle on a la satisfaction de retrouver plusieurs mots d'origine japonaise qui viennent spontanément à l'esprit. Ces entrées sont : *daïmio*, *daïri*, *dan*, *go*, *hara-kiri*, *jiu-jitsu*, *judo*, *kabuki*, *kakémono*, *kakî*<sup>2</sup>, *kami*, *kamikaze*, *kana*, *karaté*, *kimono*, *koto*, *makemono*, *mikado*, *moxa*, *nippon*, *nô*, *norimon*, *obi*<sup>3</sup>, *samouraï*, *sen*, *shamisen*, *tatami*, *torii*, *tsuga*, *tsunami*, *yen* et *zen*. Avec la séquence « terme japonais » s'ajoute le mot *abrasin*. En poursuivant la recherche, on relève encore *geisha* « Mot empr. à la langue japonaise [...] » ou *moussmé* « Transcr. du japonais [...] ». Finalement, dans cette fonction, le rendement du logiciel est dérisoire car il est inférieur à 8 % des réponses attendues. Et il convient d'ajouter que, si le TLFi avait fait l'objet d'une mise à jour, cette liste serait plus longue et comprendrait par exemple *aïkido*, *bonsaï*, *kanban*, *kata*, *kendo*, *sashimi*, *surimi*, *sushi*, *tempura* ou *yakitori*.

Observons aussi que les résultats sont le plus souvent sans nuances, même lorsque les emprunts ont pris des voies complexes ou détournées. Pour le mot *soja*, la « langue empruntée » indiquée par le logiciel est le néerlandais, soit la première langue mentionnée dans un commentaire qui laisse la place au doute : « Empr., prob. par l'intermédiaire du néerl. *soja*, au japonais *shōyu* "sauce de soja" et celui-ci au chinois [...] ». Dans le cas du mot *kapok*, la « langue empruntée » est le malais, soit la première langue mentionnée dans un

---

(s.v. *solitaire*), PINGUET pour PINGET (s.v. *mon*, *ma*, *mes*), MOUNIER pour MONNIER (s.v. *char*<sup>1</sup> : *Grandeur et décadence de M. Joseph Prudhomme*) ; mais les auteurs du corpus n'ont pas toujours été épargnés ; ainsi, justement, MONNIER pour MOUNIER s.v. *gravité* : « [Emmanuel] MONNIER, *Traité caract.* ».

commentaire construit à l'inverse du précédent : « Empr. au malais *kāpuq*, de même sens, prob. par l'intermédiaire du néerl. ». La place d'une incise suffit à provoquer une erreur, comme pour *embarrasser* : « Empr., de même que l'ital. *imbarazzare*, à l'esp. *embarazar* », qui est donné par le logiciel comme un emprunt à l'italien.

Pour couronner le tout, une autre anomalie veut que, quelle que soit la langue recherchée, cette fonction ne repère aucun vocable du début de l'alphabet (comme *agio*, *apache*, *armada*, *arquebuse*, *atoll*, *autarcie*, *babouche* ou *bacon*). Le balisage de l'objet textuel « langue empruntée » a-t-il fait l'impasse sur une partie de la nomenclature du *Trésor* ?

L'étude d'autres objets textuels, comme les « définitions » dans leurs différentes formes, les indications de « domaines » ou les « dates d'exemples », demanderait des développements plus longs et plus techniques mais aboutirait à un constat similaire.

Au terme de cet examen, qui a simplement consisté à soulever le coin du voile pour tenter d'aller au-delà des premières impressions, il faut convenir que l'informatisation du *Trésor*, limitée aujourd'hui à la reprise d'articles du dictionnaire imprimé, demandera encore de nombreux perfectionnements pour parvenir à un *Trésor informatisé* cohérent et sûr.

Au cours du colloque fondateur de 1995, la nécessité de retoucher la copie pour lui donner une plus grande unité avait été soulignée par plusieurs intervenants. Robert Martin (1996, p. 14) indiquait : « Certains articles atypiques exigent une refonte [...]. Les modifications indispensables touchent essentiellement les premiers volumes. » Il est vrai que les deux premiers volumes ont été rédigés alors que plusieurs sources de la documentation d'exemples n'étaient pas encore intégralement disponibles. Les rubriques initiales ont été redéfinies et leur contenu s'est modifié ; la rubrique *Styl[istique]* a été abandonnée, la rubrique *Dér[ivé(s)]*, qui était constituée de listes de mots, est devenue une sous-entrée, de même que l'une des formes particulières de la *Rem[arque]*. Enfin, le discours lexicographique ne s'est stabilisé que de manière progressive et le protocole typographique a subi des modifications à plusieurs reprises.

Après un examen approfondi du *Trésor*, Françoise Henry (1996, p. 79) indiquait, au cours du même colloque, quels seraient les travaux préalables indispensables pour donner une assise solide à l'informatisation du dictionnaire en tant que tel et pour en faire l'élément central d'une base lexicale, l'objectif étant de répondre à « une exigence fondamentale, celle d'une présentation formalisée rigoureuse, analytique et univoque des données lexicographiques préalablement nettement définies et circonscrites ».

Ces recommandations ont conservé toute leur pertinence. En attendant une version revue du TLFi, le recours à l'édition imprimée n'est pas à négliger pour



une appréhension du dictionnaire dans son intégralité. La version électronique, animée par un moteur de recherche puissant et rapide, a surtout les avantages liés à son support. L'utilisateur aura intérêt à s'en tenir aux fonctions les plus simples et les plus basiques, comme l'affichage mot par mot et, pour éviter les inconvénients d'un balisage qui n'est pas un modèle de perfection, il aura avantage, chaque fois que le nombre des réponses n'atteint pas une taille rédhibitoire, à effectuer des requêtes en plein texte<sup>30</sup>.

## Références

- BERNARD Pascale *et alii*, 2002, « Les ressources de l'ATILF pour l'analyse lexicale et textuelle : TLFi, Frantext et logiciel Stella », *JADT 2002, 6<sup>es</sup> Journées internationales d'analyse statistique des données textuelles*, 13-15 mars 2002, Rennes.
- CONDAMINES Anne, 2003, *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*, Toulouse, Université Toulouse-Le Mirail.
- GILBERT Pierre, 1980, *Dictionnaire des mots contemporains*, Paris, Le Robert.
- HENRY Françoise, 1996, « Pour une informatisation du TLF », *Lexicographie et informatique. Actes du Colloque international de Nancy (29, 30, 31 mai 1995) publiés par David Piotrowski*, Paris, Didier Érudition.
- MARTIN Robert, 1996, « Introduction », *ibid.*

30. En activant, dans les « recherches assistées », le type d'objet « paragraphe quelconque ».