



HAL
open science

Les dictionnaires informatisés : un atout pour l’histoire du lexique

Agnès Steuckardt

► **To cite this version:**

Agnès Steuckardt. Les dictionnaires informatisés : un atout pour l’histoire du lexique. Les dictionnaires informatisés : un atout pour l’histoire du lexique, 2008, Paris, France. p. 1205-1215, 10.1051/cmlf08151 . hal-00782010

HAL Id: hal-00782010

<https://hal.science/hal-00782010>

Submitted on 28 Jan 2013

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.

Les dictionnaires informatisés : un atout pour l'histoire du lexique

Agnès Steuckardt

Université de Provence
Laboratoire Parole et Langage (UMR 6057)
Agnes.Steuckardt@univ-provence.fr

Les dictionnaires informatisés ouvrent un accès nouveau à la connaissance du lexique. Si l'apport de ces outils documentaires à la sémantique a été plus spécialement souligné (Martin, 2001), on aimerait ici en indiquer une autre exploitation possible, orientée vers une description diachronique du lexique¹. La première approche se focalise sur la rubrique « Définition » du dictionnaire ; celle que l'on propose s'appuie sur la date de publication du dictionnaire et sur la rubrique « Étymologie », quand elle existe. Idéalement, les informations contenues dans cette rubrique devraient permettre de constituer des ensembles lexicaux en fonction de la date de première attestation, de la langue d'origine, de l'étymon, des morphèmes dérivationnels. Dans la réalité actuelle, les dictionnaires informatisés n'offrent pas toujours la possibilité même d'une requête étymologique, parce qu'ils n'ont pas nécessairement jugé utile de fournir des informations de cette nature. Quand ils l'offrent, ils sont loin d'ouvrir les différents accès que constitue chaque caractéristique étymologique. Le *Trésor de la langue française informatisé (TLFi)*, le *Nouveau petit Robert électronique (PRE)* et le *Nouveau Littré électronique (NLE)* esquissent pourtant les premiers linéaments de ce type d'investigation. Les résultats de chacun à une même requête étymologique révèlent des disparités. Sans doute la qualité du balisage, et parfois de l'enquête étymologique, peut-elle être mise en cause, mais plus souvent ces disparités révèlent des décisions lexicographiques différentes, c'est-à-dire des questions linguistiques non résolues. On indiquera dans un premier temps les limites actuelles des recherches par critère étymologique dans les dictionnaires informatisés. À travers deux études de cas : la recherche par langue d'origine et la recherche par date, on montrera cependant comment le *TLFi*, le *PRE* et le *NLE* peuvent d'ores et déjà constituer non seulement un instrument, mais aussi un aiguillon pour l'histoire du lexique.

1 Les recherches étymologiques dans les dictionnaires informatisés

Le succès d'une recherche étymologique dans un dictionnaire informatisé dépend à la fois des informations contenues dans le texte des articles et du balisage que les informaticiens-linguistes y ont opéré. La première limite de cette recherche réside dans la présence même de l'information étymologique : de ce point de vue, lexicographie ancienne et lexicographie moderne ne présentent pas les mêmes normes. La seconde, qui ne concerne que les dictionnaires modernes, se trouve dans les choix de balisage et leur degré de précision.

1.1 L'absence de rubrique « Étymologie » dans les dictionnaires anciens

Le souci étymologique n'est apparu que progressivement chez les lexicographes. Les « Dictionnaires d'autrefois », pour reprendre la dénomination choisie par les concepteurs d'une précieuse base de données lexicographiques², ne comportent pas de rubrique « Étymologie ». Leur forme informatisée, aussi minutieuse soit-elle (Wionet, Tutin, 2001) ne peut donc pas directement permettre de navigation par ce type de critère. Deux biais toutefois rendent possibles leur exploitation dans une perspective d'histoire du lexique : l'exploitation des dates de publication des dictionnaires et les informations de nature étymologique dispersées dans le texte des articles.

À la faveur de leur réunion dans une même base de données, on peut comparer leur nomenclature : on saisira ainsi le moment de la lexicographisation d'un mot ou d'un sens nouveaux. Si la base *Dictionnaires*

d'autrefois permet ce type d'enquête, son maillage, avec sept dictionnaires pour parcourir cinq siècles, reste cependant un peu lâche. À cet égard, la base informatisée du *Petit Larousse* de 1905 à 2005, en cours d'élaboration au Laboratoire « Lexiques, Dictionnaires et Informatique » (Cergy-Pontoise), devrait procurer des résultats beaucoup plus fins, du moins pour le 20^e siècle. Valoriser, par le développement de telles bases de données, un patrimoine lexicographique qui constitue « une spécificité française au cœur de la galaxie Gutenberg » (Pruvost, 2000 : 10), c'est aussi donner un tremplin à la description diachronique du lexique français.

Même si les dictionnaires anciens ne consacrent pas systématiquement une rubrique spécifique à l'étymologie, certains, comme par exemple le *Dictionnaire critique de la langue française* de l'abbé Féraud (1787-1788), jugent utile de signaler la langue d'origine d'un mot emprunté. Les recherches en plein texte ouvrent dès lors une voie d'investigation. On pourra par exemple reprendre sur de nouveaux frais l'enquête menée par John Humbley, il y a une vingtaine d'années, sur les anglicismes dans le *Dictionnaire critique* de Féraud (Humbley, 1986). La recherche en plein texte du mot : *anglicisme* dans la base de données du *Dictionnaire critique* permet de repérer 10 mots que l'enquête traditionnelle n'avait pas relevés (*coalition, compassioner, dépravité, disgracieux, empiètement, inconsistance, inofficieux, non-sens, prescriptions, spontanément*) ; celle du mot : *anglais* en fournit encore 11 de plus (*bambou, importation, importer, interlope, lord, paquebot, ponche, redingote, tôte, toster, vote*). Une étude plus large sur le traitement des emprunts dans les dictionnaires d'autrefois pourra, grâce à leur informatisation, trouver un aboutissement rapide.

1.2 Le TLFi : un balisage limité à la « Langue empruntée »

Le TLFi offre une grande richesse dans le balisage général des articles : les critères de recherche, regroupés en 30 « Types d'objet » (par exemple : Synonymes, Domaine général, Construction, Indicateur) sont déterminés par l'utilisateur, qui pourra ainsi, dans le type *Domaine général* spécifier l'objet « vestimentaire », ou dans le type *Indicateur* spécifier « populaire ». Toutefois, pour la rubrique « Étymologie », un seul « Type d'objet » a été retenu : la *Langue empruntée*. Les autres informations, sur les dates et sources de première attestation, sur l'étymon, sur la formation du mot, n'ont pas été balisées et ne sont accessibles que par une recherche en plein texte, par le « Type d'objet » *Paragraphe*. Sans doute la rubrique « Étymologie » ne donnait-elle pas encore pleine satisfaction aux concepteurs du TLFi : une équipe de l'ATILF travaille actuellement à une révision des notices étymologiques, dont l'aboutissement est annoncé pour 2014 (Steinfeld, Petrequin, Evrard, 2005).

Les notices révisées sont publiées sur le web au fur et à mesure de leur rédaction : actuellement 150 notices sont consultables à l'adresse : www.atilf.fr/tlf-etym. Une interrogation commode par « classes étymologiques » est mise en place ; trois classes ont été prévues : « le lexique héréditaire », « les transferts linguistiques », « les formation française », avec, pour chacune, une sous-catégorisation rigoureuse. L'interrogation par date n'est cependant pas prévue, et l'information diachronique ne reste accessible que par la recherche en plein texte.

1.3 Le NLE : requêtes par langue d'origine et par « Ancienneté »

Le NLE a, du point de vue des requêtes étymologiques, suivi cet exemple de prudence. Si, avec 11 critères de recherche, il permet des requêtes diversifiées (notamment par Domaine, Usage, Code grammatical), la requête « Étymologie » n'informe que sur la langue d'origine. Quelque 150 langues ont été distinguées, avec un luxe de finesse dans la catégorisation ; pour le grec par exemple, le NLE ne définit pas moins de 14 sous-catégories. Quelques catégorisations prometteuses ont été, bien qu'externes à la notion de « Langue d'origine », insérées dans la liste : les catégories « Nom propre » et « Étymologie populaire ». Si elles ne fournissent pas pour l'instant de résultats considérables – trois résultats seulement pour la requête « Étymologie= Étymologie populaire » –, il suffirait de compléter les indications portées dans la rubrique « Étymologie » pour récolter de meilleures moissons.

Cette rubrique ne comporte pas, dans le *NLE*, d'indication sur la date de première attestation. Le *Littré* ne donnait pas de date de première attestation, mais seulement des indications d'usage, par siècle. Une approche diachronique est cependant possible par la requête « Ancienneté », qui exploite la présence de deux strates rédactionnelles dans le *NLE*. Elle met en évidence d'une part les mots qui se trouvent dans l'édition de 1874, mais sont sortis d'usage, et d'autre part ceux qui sont présents dans la nouvelle édition, mais ne l'étaient pas dans celle de 1874. Le *NLE* permet par exemple de lister 11 interjections sorties d'usage, de *ahi* à *tarare*, en passant par *diablezot* et *morguienne*. Menée à plus grande échelle, la requête par « Ancienneté » met à jour deux états synchroniques du lexique, celui de 1874 et celui de 2004/2006, tel que *Littré*, puis les rédacteurs du *NLE* les ont perçus.

1.4 Le *PRE* : requêtes par « Langue d'origine », par « Date », par « Mots de l'étymologie »

Les concepteurs du *PRE* ont été, concernant les dates de première attestation, moins circonspects que ceux du *TLFi* : ils tirent parti, dès à présent, des informations contenues dans la rubrique « Étymologie » du *PRE*. Ont été balisées non seulement la langue empruntée, mais aussi la date de première attestation ; une recherche par intervalle entre deux dates a été rendue possible, ainsi qu'une recherche en plein texte limitée à la rubrique « Étymologie ». Ce dernier type de requête permet d'extraire du dictionnaire l'ensemble des mots issus du même étymon. Quelques interrogations par des métatermes, comme *sigle*, sont également possibles. Ainsi, une requête par « Mots de l'étymologie » avec le mot *sigle* fournit une liste des 145 articles ; en retranchant *siglé* et *siglaison*, dont l'étymologie mentionne *sigle* en tant qu'étymon et non en tant que descripteur, on obtient la liste des 143 sigles retenus comme entrées par le *PRE*.

Les requêtes par « Mots de l'étymologie » se révèlent cependant décevantes à cause de l'absence de systématisme dans l'emploi des métatermes : si un effort a été fait pour *sigle*, *acronyme* et *mot-valise*, on constatera que *suffixe*, *préfixe*, *composé*, *nom propre*, *attraction*, ou encore *métathèse*, ne sont utilisés que de façon épisodique ; *suffixe* par exemple ne semble employé que dans les étymologies non triviales. Ainsi lira-t-on dans l'étymologie de *chauffard* : « 1897 ; de *chauff(eur)* et suffixe péj. *-ard* », mais dans celle de *chauffeur* seulement : « 1680 ; de *chauffer* ». Une plus grande systématisme dans l'emploi des métatermes donnerait évidemment plus de valeur aux listes constituées à partir de ce type de requête. À l'heure actuelle, c'est donc seulement sur la langue d'origine que l'on peut mener l'enquête à la fois dans le *TLFi*, le *NLE* et le *PRE*, pour procéder à une analyse comparative des résultats.

2 Requête par la langue empruntée

On utilisera ici la comparaison des résultats pour mettre en évidence les problèmes certes informatiques, mais aussi linguistiques, qui fragilisent la fiabilité des requêtes par langue empruntée.

2.1 Un balisage à harmoniser

Le balisage de la langue empruntée présente, dans les trois dictionnaires contemporains, des anomalies. Par exemple, le mot *mah-jong*, pourtant signalé dans le *TLFi* et le *NLE* comme « mot chinois » dans la rubrique « Étymologie » de l'article qui lui est consacré, n'est trouvé ni dans le *TLFi*, par la requête « chinois » dans le type d'objet « Langue empruntée », ni, dans le *NLE*, par la requête « Étymologie=chinois » : les listes des résultats doivent être complétées par une fastidieuse recherche en plein texte. Dans le *TLFi*, le repérage de la langue-cible n'est pas toujours des plus aisés : ainsi pour trouver les mots issus de l'ancien scandinave, il ne convient pas de saisir « ancien scandinave », ni « norrois », ni « ancien nordique », mais « a. nord. ».

2.2 Balisage et identification des emprunts

Dans le *PRE*, le balisage des emprunts a été systématiquement revu. Toutefois, la mention de l'origine étrangère pour les locutions calquées, introduit dans les listes obtenues des mots que l'on peut percevoir comme des intrus. Apparaissent par exemple dans la liste des emprunts au chinois :

- *face*, parce que *perdre la face* serait un calque d'une locution chinoise,
- *révolution* à cause de *révolution culturelle*,
- *triade* à cause de l'emploi du mot au sens de « organisation secrète chinoise de type mafieux ».

Ce n'est pas ici la réalisation technique du balisage qu'il faut incriminer, mais son manque de finesse : en l'état actuel, il ne permet pas de distinguer les emprunts des calques, et renvoie aux contours donnés à la définition linguistique de l'emprunt (Tournier, 1985 ; Sablayrolles, 2000 ; Jacquet-Pfau, 2003).

Ces quelques imperfections des requêtes par langue empruntée étant prises en compte, expérimentons la recherche par langue empruntée sur un exemple : les emprunts au russe. Dans le *TLFi*, la requête par le Type d'objet « Langue empruntée » et l'objet : « Russe » ne fournit que 46 résultats. Elle doit être complétée par une recherche en plein texte, par le Type d'objet « Paragraphe » et l'objet « Russe ». Dans le *NLE*, la requête « Étymologie=Russe » ramène 47 résultats ; elle doit de même être complétée par la requête « Mot du texte=russe ». Dans le *PRE* en revanche, la requête étymologique choisissant pour critère la langue russe ramène 96 résultats, dont quelques-uns, comme *personnalité* ou *perspective*, surprennent. *Personnalité* figure en fait dans la liste seulement parce que le *PRE* signale que la collocation *culte de la personnalité* est une adaptation du russe ; et *perspective* est présent à cause de l'emploi de *perspective* au sens de « grande avenue rectiligne », qui est la traduction d'un mot russe. Ces mots seraient à prendre en considération dans une étude sur les calques, mais, comme les deux autres dictionnaires ne les ont pas intégrés, il est préférable dans cette étude de les laisser de côté. Les trois dictionnaires fournissent, après corrections, une liste d'environ 80 mots³. Si les trois ensembles comportent une intersection, chacun présente un sous-ensemble spécifique : comment expliquer ces différences ?

2.3 La documentation

Une particularité bien connue du *TLFi* tient à sa documentation d'orientation principalement littéraire. Elle explique notamment la présence des emprunts introduits au 19^e siècle par la vogue des romans russes traduits en français : pour cette période, les rédacteurs du *TLFi* ont, contrairement à ceux du *NLE* et du *PRE*, choisi de retenir les mots : *tcherkess* attesté dans Frantext chez Alain Fournier et Romain Gary, *barine* (Gaston Leroux, Paul Bourget, Camus), *sotnia* (Leroux, Kessel) et *nagaïka* (Kessel, Cavanna).

La documentation de Littré était, elle aussi, à dominante littéraire. Le *NLE*, comme le *TLFi*, et contrairement au *PRE*, retient *kirghiz*, introduit en français par la traduction d'un roman de Pouchkine. Il est le seul à mentionner *byline*, « épopée populaire russe ».

La documentation du *TLFi* diffère de celle du *NLE* et du *PRE* par sa date. Le dernier russisme enregistré par le *TLFi* est *samizdat* (1971) ; le *PRE* et le *NLE* ont pu ajouter cinq emprunts postérieurs : *tokamak*, *nomenklatura*, *refuznik*, *glasnost*, *perestroïka*. La postériorité de publication n'explique cependant pas tout. Enregistré par le *PRE* et le *NLE*, le mot *goulag*, par sa date de première attestation, 1938, aurait semblé susceptible de constituer une entrée du *TLFi* ; il ne figure pourtant pas dans la nomenclature. C'est sans doute ici l'usage qui a guidé les rédacteurs du *TLF*. Jusqu'en 1981 en effet, *goulag* n'est mentionné dans Frantext qu'en citation du titre de l'ouvrage de Soljenitsyne : il n'a pas encore acquis le statut de nom commun autonome, ce qui justifie l'abstention des rédacteurs du *TLF*. Dans un autre domaine, on pourrait de même s'étonner que le *TLFi*, contrairement au *PRE* et au *NLE*, ne comporte pas d'entrée *daïcha* (1843), ni *blini* (1883) : là encore, l'usage au moment de la publication du dictionnaire peut expliquer ces choix ; les modes immobilières et alimentaires ont déterminé l'enregistrement de ces emprunts dont la première attestation est ancienne, mais l'entrée dans l'usage plus récente.

2.4 Le traitement des ethnonymes

Le *TLFi* et le *NLE* se montrent plus accueillants à l'égard des ethnonymes que le *PRE* : ils enregistrent *tchérémissé*, peuple d'origine finnoise et *kirghiz*, peuple nomade d'Asie centrale ; le *TLFi* retient en outre : *ossète*, peuple indo-européen du Caucase central, et *tcherkessé*, peuple du nord-ouest du Caucase. Alain Rey et Josette Rey-Debove justifient la mise à l'écart des ethnonymes de la façon suivante :

Les noms propres de personne (anthroponymes) ou de lieux (toponymes) produisent de nombreux dérivés, adjectifs ou noms ordinaires. Ces mots sont embarrassants pour le lexicographe dans la mesure où les noms propres sont internationaux et donc translinguistiques. Ce n'est qu'une question d'usage, rien n'empêche de produire des adjectifs français à partir de tous les noms allemands ou arabes d'une encyclopédie, par exemple. D'autre part la plupart de ces dérivés n'ont pas véritablement de sens linguistique : ils désignent par leur radical et signifient seulement par leur suffixe. (Rey, Rey-Debove, 1993 : 13)

Deux arguments sont avancés : premièrement, la production des ethnonymes est virtuellement non limitée aux lexèmes effectivement attestés, « ce n'est qu'une question d'usage » ; deuxièmement, ces dérivés n'ont pas « véritablement de sens » car « leur radical », c'est-à-dire le nom propre, « désigne », et ne signifie pas. Ces deux arguments peuvent être contestés. On peut s'étonner de voir balayée d'un revers de main la question de l'usage dans un dictionnaire de langue : le *TLFi* et le *NLE* ont préféré mentionner, non pas tous les ethnonymes virtuellement possibles, mais ceux dont l'emploi est attesté.

Le second argument repose sur une conception classique du nom propre, développée par Stuart Mill et reprise par Kripke : cette conception, selon laquelle le nom propre serait dépourvu de sens, est aujourd'hui mise en débat (Siblot, 1995 ; Leroy, 2004). Le *TLFi* et le *NLE* manifestent une plus grande ouverture à ce que Paul Siblot appelle la « signifiante » des noms propres : ce choix linguistique les amène à inclure dans leur nomenclature non seulement les ethnonymes cités ci-dessus, mais aussi des mots à référent unique comme *kremlin*, pour lequel le *TLFi* donne la définition : « Résidence des tsars à Moscou », tandis que le *PRE* l'écarte.

2.5 La description de l'emprunt

La réflexion linguistique sur l'emprunt a souffert, dans la perspective synchronique longtemps privilégiée par la linguistique, d'un certain désintérêt ; pour Saussure en effet, « le mot emprunté ne compte plus comme tel, dès qu'il est étudié au sein du système » (Saussure, 1915 : 42). On perçoit dans la présentation lexicographique des emprunts une hésitation entre description diachronique et synchronique.

2.5.1 Les russismes à étymon latin

Se voulant outil de transmission d'un patrimoine historique, les trois dictionnaires choisissent en principe une description diachronique ; pourtant, des flottements apparaissent dans la description des emprunts dont la morphologie est accessible en synchronie. Pour le *TLFi*, *décembriste* et *planifier* sont des russismes, alors que le *PRE* y voit des dérivés de *décembre* et de *planifier* ; inversement, pour le *PRE*, *défaitisme* et *vernalisations* sont des russismes, tandis que, pour le *TLFi*, il s'agit de dérivés de *défaite* et de *vernal*. Selon le *NLE*, aucun de ces mots ne serait emprunté au russe. Tous, de fait, sont formés sur des étymons latins : même si ce sont des locuteurs russes qui ont créé le dérivé, le rédacteur a pu être tenté de négliger une étymologie quelque peu contre-intuitive. L'intérêt récent des linguistes pour la notion d'événement linguistique (Guilhaumou, 2006) amènera peut-être à ne pas sous-estimer l'importance de l'acte de création, et à trancher plus nettement en faveur d'une description diachronique.

2.5.2 Les mots-voyageurs

Une hésitation de même nature apparaît dans la description des mots-voyageurs. Pour le *TLFi*, *cosaque* et *polatouche* sont des emprunts au polonais, qui les emprunte lui-même au russe ; le *PRE* signale leur

origine russe, sans mentionner leur passage par le polonais. Inversement, le *PRE* ne mentionne pour *tsarine* que l'emprunt à l'allemand, tandis que le *TLFi* n'indique que l'origine russe. Il est évidemment un peu surprenant de voir le mot *tsarine* décrit comme un emprunt à l'allemand, mais si l'on veut serrer au plus près les étapes du transfert lexical, il importe de connaître la langue de transmission.

La comparaison des listes fournies par les trois dictionnaires informatisés révèle donc une double instabilité. La première est inhérente à la nature de l'outil de recherche : le dictionnaire ne donne pas un recensement exhaustif des emprunts qu'a pu faire, au cours de son histoire, le français, mais retient seulement ceux qu'il juge utiles à l'utilisateur contemporain : de là, par exemple, la présence de *datcha* et *blini* dans le *PRE* et le *NLE*, mais pas dans le *TLFi*. Une seconde instabilité est due à des incertitudes dans l'analyse linguistique : les ethnonymes ont-ils leur place dans un dictionnaire de langue ? Les mots-voyageurs doivent-ils être présentés comme des emprunts à la langue qui les a créés ou à la langue qui les a transmis ? Si le dictionnaire tranche clairement en faveur d'une lexicologie diachronique, il devra tenir indiquer les différentes étapes du parcours et articuler le balisage des langues à celui des dates de première attestation.

3 Requête par date de première attestation

La contribution la plus évidente du dictionnaire à la connaissance de l'histoire du lexique est la datation de la première attestation d'une forme lexicale ou de ses changements de sens. La requête par date de première attestation n'est actuellement possible que dans le *PRE*.

3.1 Quelle date choisir ?

Si, généralement, la rubrique « Étymologie » indique une seule date de première attestation, il est deux situations où elle en mentionne deux, voire davantage.

Lorsque la forme graphique du mot a évolué, le *PRE* donne la date de la forme stabilisée en français moderne, et indique la date de première attestation de la graphie initiale : par exemple, pour *kopeck*, il donne comme date de première attestation : 1806, et signale : « *copec* ; 1607 ».

Une autre situation de pluri-datation se présente lorsqu'il y a décalage entre la première attestation du mot et son entrée dans l'usage. Ainsi, pour *aristocrate*, le *PRE* donne : « 1550 ; répandu en 1778 » et, avec une autre présentation, pour *patriotique* : « 1750 ; hapax 1532 ». Or les deux datations ont été balisées : ainsi, *aristocrate* et *patriotique* seront ramenés à la fois par une requête sur le 16^e siècle, et par une requête sur le 18^e ; de même, *kopeck* apparaîtra à la fois dans les listes du 17^e et du 19^e siècle. Dès lors, une hiérarchisation dans le balisage des dates semble souhaitable ; elle suppose qu'ait été menée en amont une réflexion linguistique sur ce qu'il convient de considérer comme l'entrée du mot dans la langue.

Dans une perspective d'histoire du lexique, le développement de la notion d'événement linguistique, avec la dimension sociale et institutionnelle qu'il comporte, semble devoir conduire à privilégier la date où le mot est « répandu » dans l'usage, donc à préférer pour *aristocrate* et *patriote* la seconde date. En revanche *kopeck* peut très bien avoir été répandu dans l'usage avec une graphie variable, et la première datation peut sembler plus pertinente. La réponse du lexicographe à la question de la datation requiert à la fois une réflexion théorique sur la constitution de la norme lexicale, et un travail d'analyse de discours sur l'archive.

3.2 Un exemple d'exploitation

Dans l'état actuel de son élaboration, le *PRE* fournit dès à présent un outil utile à l'historien du lexique, et permet de mener des investigations d'une amplitude qu'il lui aurait été difficile d'atteindre sans le secours du dictionnaire informatisé. Pour en donner une illustration, procédons à une recherche comparée des emprunts du français au néerlandais et à l'allemand du 16^e au 20^e siècle.

3.2.1 Approche quantitative

Une interrogation par période de 50 ans permet d'évaluer rapidement l'évolution de ces emprunts. Le travail le plus long réside dans la correction des réponses contestables, nécessaire à cause des hésitations que nous avons notées plus haut :

- enregistrement des calques, comme *surhomme* (1892), pour traduire *Übermensch*,
- enregistrement de la langue d'origine alors que ce n'est pas la langue par laquelle le mot est entré en français : par exemple, *commodore* apparaît dans les résultats de la requête des mots issus du néerlandais au 18^e siècle, alors qu'il s'agit d'un emprunt certes du 18^e siècle, mais à l'anglais (qui l'avait emprunté au néerlandais) ;
- double enregistrement des mots dont la graphie a évolué, comme *asticoter*, qui figure à la fois dans les résultats du 18^e et du 17^e siècle, parce que la rubrique étymologie signale la graphie *dasticoter*, attestée dès 1642 ;
- double enregistrement des mots dont la date de première attestation est décalée par rapport à la date d'entrée dans l'usage, comme *L.S.D.*, sigle formé en allemand à partir du mot *Lysergsäure-diäthylamid*, attesté dès 1948, mais répandu seulement en 1966.

Après ces corrections, on obtient les résultats suivants :

Tableau 1

Emprunts du français au néerlandais et à l'allemand depuis 1500, par tranches de 50 ans

Période	Néerlandais	Allemand
1500-49	3	4
1550-99	11	12
1600-49	18	11
1650-99	22	13
1700-49	9	10
1750-99	13	35
1800-49	5	52
1850-99	7	114
1900-49	4	110
1950-99	1	23

La grande époque des emprunts au néerlandais correspond au « siècle d'or » des Pays-Bas. En 1681, l'Acte de La Haye proclame l'indépendance des Provinces unies, qui se libèrent du joug espagnol. Les Hollandais développent leur puissance maritime et disputent, souvent avec succès, aux Espagnols et aux Portugais leurs possessions coloniales. Le pays est prospère et ouvert aux réfugiés politiques ou religieux.

Les emprunts à l'allemand se situent, pendant la période du 16^e au 18^e siècle, à un niveau comparable à ceux du néerlandais. C'est à partir du début du 19^e siècle qu'ils franchissent un seuil quantitatif. Du 19^e siècle jusqu'au début du 20^e siècle, au moment du plus grand rayonnement scientifique et philosophique

de l'Allemagne, les emprunts à l'allemand restent à un niveau élevé, et amorcent leur déclin dans la première moitié du 20^e siècle.

3.2.2 L'âge pré-industriel : la mer et la terre

Les listes obtenues permettent de préciser la nature des emprunts. Sur la période 16^e-18^e siècle, emprunts au néerlandais et emprunts à l'allemand s'opposent nettement par leur domaine : alors que les premiers sont liés à la puissance maritime des Provinces unies, les seconds ressortissent du vocabulaire de la mine.

Les emprunts au néerlandais manifestent en effet de deux manières le rayonnement des Provinces unies sur les mers :

- d'une part, ils relèvent du vocabulaire spécialisé de la mer : dès le 16^e, on importe des noms de poisson (*flétan, églefin*) ; puis au 17^e siècle, ce sont plutôt des mots qui nomment les parties du bateau (*foc, rouf*),
- le néerlandais, d'autre part, sert d'intermédiaire entre le français et les mots venus des comptoirs : comme *thé*, venu du chinois (ou du malais), et *palissandre* venu d'un créole de Guyane. De la même manière, il sert d'intermédiaire pour diffuser les dénominations portugaises de relais lointaines, comme dans le cas de *sargasse*, ou de *pamplemousse*.

À la même époque, c'est pour nommer les ressources de la terre que le français recourt à l'allemand, avec *cobalt, gangue* au 16^e, *castine, zinc, bismuth* au 17^e, *quartz, blende, hornblende, pechblende, spath, feldspath, mispickel, gneiss* au 18^e. L'exploitation des ressources minières des montagnes hercyniennes, de part et d'autre de la forêt noire, s'est inspirée du modèle allemand.

Comme le néerlandais, l'allemand peut être une langue de transmission. Le facteur historique déterminant n'est pas alors la puissance maritime, mais la guerre : les transferts lexicaux sont passés par les mercenaires allemands, qui servent sous le commandement français depuis le 15^e siècle, auxquels ce sont ajoutés, après leur défaite à Marignan, les Suisses. Sont ainsi empruntés à l'allemand : *obus* (du tchèque *haufnice*), *sabre* et *hussard* (du hongrois *szablya* et *huszar*, « le vingtième », parce qu'en 1458 le gouvernement hongrois ordonna la levée d'un homme sur huit pour former une cavalerie), *uhlan* (du polonais *oglan*).

3.2.3 La langue populaire, du 16^e au 20^e siècle

Matelots et soldats, mais aussi main d'œuvre en quête de travail, ont au fil des siècles laissé des traces dans la langue populaire. Les apports du néerlandais sont quantitativement modestes, mais réguliers :

- au 16^e siècle : *ripaille*⁴, *drôle*,
- au 17^e, *gribouiller, gredin, micmac*,
- au 18^e, *campuse, bastringue*,
- au 19^e, *gribiche* (« mégère »), *vaser* (« pleuvoir »),
- au 20^e, *margaille* (« bagarre, désordre »).

Ceux de l'allemand, plus nombreux, témoignent d'une circulation linguistique constante, particulièrement riche au 19^e siècle, lorsque, de part et d'autre du Rhin, la Révolution industrielle jette sur les routes des ouvriers itinérants à la recherche d'embauche. On relève :

- au 16^e siècle : *chenapan, brinde* (devenu *bringue* au 19^e siècle)
- au 17^e, *glass* (« verre d'une boisson alcoolisée »), *nouille*,
- au 18^e, *asticoter, drille, loustic, schnaps*,
- au 19^e, *chnouf* (« drogue »), *turne* (« chambre »), *estourbir, schlague* (« coup »), *nase, chlinguer, arpète* (« jeune couturière apprentie »), *flingue, clamser, schproum* (« dispute violente »), *mouise*.

Dans un contexte historique d'affrontement, le 20^e siècle multiplie les emprunts servant de dénominations familières et péjoratives des Allemands, avec *fritz*, *fridolin*, *frisé* ; de l'argot des camps restent *Kapo*, *Stalag*, *Oflag*.

3.2.4 Les emprunts savants à l'allemand

À partir de la fin du 18^e siècle, un nouveau type d'emprunt apparaît : les emprunts formés par les savants allemands sur des bases latines ou grecques. Des domaines scientifiques très variés sont concernés ; on en donnera ici seulement quelques illustrations, choisies parmi les emprunts du 19^e siècle :

- la chimie : *benzine*, *émulsine*, *barbiturique*,
- la médecine : *leucémie*, *embolie*, *homéopathie*,
- la philosophie : *noème*, *déterminisme*, *pragmatisme*,
- et, bien sûr, la linguistique : *morphologie* (créé par Goethe d'abord pour la botanique), *orthoépie*, *stylistique*, *syntactique*, *sémasiologie*.

Certes, du point de vue de la morphologie lexicale, ces mots ne présentent pas de caractère germanique (Jacquet-Pfau, Moreaux, 1998), mais, du point de vue de l'histoire de la langue, l'interrogation par le dictionnaire informatisé met en évidence un moment de forte influence du discours savant allemand sur le lexique français.

Les dictionnaires informatisés facilitent, et donc favorisent une approche historique du lexique français. Adossés à une riche documentation étymologique, ils sont susceptibles de donner à voir sa structuration diachronique. Certes, les balisages actuels, encore timides ou imparfaits, pourront être améliorés ; plus fondamentalement, la représentation de la structure diachronique dessinée par les dictionnaires ne coïncide pas exactement avec sa réalité historique : les dictionnaires ne retiennent dans leur nomenclature que les mots qu'ils jugent pertinents dans le discours contemporains, et, du fait de cette sélection, renvoient de l'histoire du lexique une image déformée. Mais cette image elle-même, variable d'un dictionnaire à l'autre, évolutive, constitue la trace mémorielle active du passé linguistique (Paveau, 2006) et fait elle-même partie de la description du français. Aux diachroniciens de comparer les images successives données par les dictionnaires, et de confronter ces représentations à l'histoire du lexique telle qu'ils peuvent la reconstituer, notamment grâce à l'informatisation des dictionnaires anciens. Forte d'une tradition lexicographique exceptionnelle (Pruvost, 2000 : 9-11), la linguistique française a aujourd'hui, avec les dictionnaires informatisés, l'opportunité de développer un secteur des sciences du langage longtemps relégué au second plan par le primat d'un structuralisme synchroniste.

Références bibliographiques

- Blum, C., Pruvost, J. (2007). *Le Nouveau Littré*. Version électronique. Paris : Éditions Garnier.
- Bouverot, D., Steuckardt, A. (à paraître en 2008). À travers le *TLFi* : les emprunts au russe. *Neologica*, n° 2.
- Caron, P., Dajenais, L., Gonfroy, G. (1992). Le programme d'informatisation du *Dictionnaire critique de la langue française* de l'abbé Féraud (1787). In Wooldridge, T.R. (éd.), *Historical Dictionary Databases*. Édition en ligne : *Computing in the Humanities Working Papers*, 1996 (<http://www.chass.utoronto.ca/epc/chwp/>).
- Dendien, J. (Conception et réalisation informatiques) (2002). *Le Trésor de la Langue française informatisé*. Version informatique du *TLF* (1971-1994). ATILF et CNRS Éditions.
- Guilhaumou, J. (2006). *Discours et événement. L'histoire langagière des concepts*. Besançon : Presses Universitaires de Franche Comté.

- Humbley, J. (1986). Les anglicismes dans le *Dictionnaire critique* et dans le *Supplément. Autour de Féraud. La lexicographie en France de 1762 à 1835*. Paris : École Normale Supérieure de Jeunes Filles, 147-155.
- Jacquet-Pfau, C., Moreaux, M.-A. (1998). Motivation et transparence des emprunts gréco-latins en français et en allemand. In Clas, A., Mejri, S., Baccouche, T. (éds), *La mémoire des mots*, Tunis : AUPELF-UREF et Serviced, 587-600.
- Jacquet-Pfau C. (2003). Du statut de l'emprunt en traitement automatique des langues. In Sablayrolles, J.-F. (éd.), *L'innovation lexicale*. Paris : Honoré Champion, 79-97.
- Leroy, S. (2004). *De l'identification à la catégorisation. L'antonomase du nom propre en français*. Louvain-Paris : Éditions Peeters.
- Martin, R. (2001). *Sémantique et automate : l'apport du dictionnaire informatisé*, Paris : Presses Universitaires de France.
- Paveau, M.-A. (2006). *Les prédiscours. Sens, mémoire, cognition*, Paris : Presses de la Sorbonne Nouvelle.
- Pierrel, J.-M. (2003). Un ensemble de ressources de référence pour l'étude du français : *TLFi, Frantext* et le logiciel *Stella*. *Revue québécoise de linguistique*, volume 32, n° 1, 155-176.
- Pruvost, J. (2000). *Dictionnaires et nouvelles technologies*. Paris : Presses Universitaires de France.
- Rey, A., Rey-Debove, J. (1993). *Préface du Nouveau Petit Robert*. Paris : Le Robert, 9-23.
- Rey, A., Rey-Debove, J. (dir.) (2006). *Le Nouveau Petit Robert 2007*. CD-ROM. Nouvelle édition du *Petit Robert* de Paul Robert, remaniée. Le Robert Éditions électroniques.
- Saussure, F. (1916). *Cours de linguistique générale*, Paris : Payot.
- Sablayrolles J.-F. (2000), *La néologie en français contemporain*, Paris : Honoré Champion.
- Siblot, P. (1995). Noms propres et image de marque. De la construction du sens dans les noms propres. In Noailly, M. (éd.), *Nom propre et nomination*. Paris : Klincksieck, 147-186.
- Steinfeld, N., Petrequin, G., Evrard, A. (2005). *Projet TLF-Étym : mise à jour des notices étymologiques du Trésor de la langue française informatisé*. Dossier de présentation, Nancy : Laboratoire ATILF.
- Steuckardt, A. (2008). Les emprunts aux langues germaniques : parcours diachronique. Site web du Cercle des Linguistes d'Aix-en-Provence. <http://sites.univ-provence.fr/wclaix/>
- Tournier, J. (1985). *Introduction descriptive à la lexicogénétiq ue de l'anglais contemporain*. Paris : Honoré Champion.
- Wionet, C., Tutin, A. (2001). *Pour informatiser le dictionnaire de Basnage 1702 et de Trévoux 1704*, Paris : Honoré Champion.

¹ Jean-Marie Pierrel décrivait ainsi en 2003, outre « la possibilité d'écriture et d'exploitation de grammaire », sept pistes d'exploitation des dictionnaires informatisés : la « recherche des cooccurrences et collocations », « l'extraction des sous-lexiques », les « études morphologiques », les « études de syntaxe locale », les « études de sémantique », les « études de stylistique », et la piste très ouverte : « Autres possibilités ». C'est sous la rubrique « Autres possibilités » qu'il évoquait, très brièvement, la piste diachronique : « Dans le *TLFi*, on peut trouver les origines et premières attestations des mots dans les domaines tant médical que philosophique, techniques ou autres » (Pierrel, 2003 : 175). Les origines et premières attestations ne sont pas seulement, selon nous, des données propres à satisfaire la curiosité des étymologistes sur les vocabulaires spécialisés : elles constituent des informations susceptibles d'aboutir à des recherches d'envergure sur l'histoire du lexique français.

² L'informatisation de ce dictionnaire a été dirigée par Philippe Caron et Louise Dagenais ; il est, depuis 2004, consultable sur le site des laboratoires ATILF (Analyse et Traitement Informatique de la langue Française, Nancy) et ARTLF (*American Research on the Treasury of French Language*, Chicago).

³ Pour une comparaison plus complète des listes données par le *TLFi* et par le *Petit Robert électronique*, voir Danielle Bouverot et Agnès Steuckardt, 2008, à paraître.

⁴ Les emprunts sont présentés dans l'ordre chronologique, en fonction de leur date de première attestation.