

I. KALBOTYRA

Sabine Albert

Université de Cergy-Pontoise

CNRS UMR 7187 (Laboratoire LDI)

33, bd du Port

95011 Cergy-Pontoise cedex

Tel : 0134256758

Mel : sabine.albert@ac-versailles.fr

Champs de recherche : lexicographie, emprunts, constitution des lexiques français et anglais

POSSIBILITÉS OFFERTES PAR L'INFORMATISATION : LE CAS DU TRÉSOR DE LA LANGUE FRANÇAISE

En tant que reflets de la société, les dictionnaires se doivent de suivre les évolutions technologiques de leur temps. Le Trésor de la Langue Française ne fait pas exception : œuvre magistrale et imposante, après avoir bénéficié des découvertes techniques et technologiques les plus récentes lors de sa conception, puisque la constitution du corpus elle-même a été assistée par ordinateur afin de réunir un matériel de première importance, il est passé à l'ère de l'informatisation, répondant ainsi aux attentes nouvelles d'un public toujours plus exigeant. Dès lors que le support n'est plus celui du papier, les contraintes sont modifiées pour les lexicographes, et les attentes différentes pour les usagers, en droit d'attendre de nouveaux services. Les stratégies de consultation s'en trouvent transformées, et les recherches facilitées par des requêtes qui peuvent désormais combiner plusieurs critères en précisant, le cas échéant, les liens d'interdépendance. Il devient alors aisément possible de créer des listes de mots, de calculer des fréquences, autant de ressources de première importance pour quiconque s'intéresse de près au lexique. Dans le cadre de l'étude des emprunts aux langues étrangères, les possibilités démultipliées de consultation proposées par la version informatisée du TLF constituent une manne non négligeable pour les chercheurs et ouvrent la voie à une exploration analogique du traitement des mots venus d'ailleurs.

MOTS CLÉS : lexicographie, lexicologie, emprunts, dictionnaire informatisé, informatique.

Le TLF, déjà novateur dans sa conception et sa mise en place, a bénéficié, tout au long de son élaboration, des technologies les plus pointues, et en particulier des progrès en terme d'informatisation qui en font un véritable pionnier dans le domaine des dictionnaires électroniques. Il va sans dire que les recherches portant sur les emprunts et sur la terminologie employée pour les décrire sont grandement facilitées par la consultation informatisée.

Dès le début de son histoire, la fée informatique s'est penchée sur le berceau du TLF qui est né, s'est formé puis transformé en fonction des avancées technologiques de son temps. La jaquette originelle est d'ailleurs révélatrice de l'enthousiasme avec lequel les rédacteurs du TLF ont utilisé les toutes nouvelles possibilités offertes par l'informatique

naissante : « seul un ORDINATEUR pouvait dépouiller, classer et gérer, en moins de sept ans, 90 millions d'exemples extraits de bibliothèques de toutes disciplines. Employant pour la première fois cet instrument nouveau sur une aussi vaste échelle, 150 chercheurs et techniciens ont réalisé le DICTIONNAIRE NATIONAL qui manquait à la langue française contemporaine. Grâce à l'électronique, chaque mot est traité tant en fonction de données statistiques, que selon ses différentes significations et valeurs d'emploi. »

L'informatisation a donc assisté le travail des lexicographes dès la conception de l'ouvrage, mais son rôle en est arrivé à dépasser le simple moyen technique mis à la disposition des spécialistes : en informatisant le dictionnaire lui-même, ce sont des stratégies de consultation nouvelles qui sont mises en place, permettant ainsi au lecteur-consultant de plonger dans une nouvelle ère, l'ère analogique.

Nous nous proposons donc, dans cette étude, de repérer tout d'abord les différentes étapes qui ont conduit à l'informatisation du *TLF* et permis d'organiser la feuille de style nécessaire à la mise en place d'un document structuré. Nous nous efforcerons ensuite d'examiner les stratégies de consultation, forcément nouvelles sans être toutefois radicalement différentes, qui en découlent. Enfin, nous verrons dans quelle mesure la version informatisée du *TLF* ouvre de nouvelles perspectives pour la recherche, tout en essayant de mettre en lumière quelques améliorations qui pourraient être apportées dans le cadre particulier des emprunts.

1. SUIVRE LE PROGRÈS

Les liens entre le *TLF* et l'informatique sont à considérer à deux niveaux : celui de l'élaboration de l'ouvrage, et celui de son informatisation dans sa version *Trésor de la Langue Française informatisé* (désormais *TLFi*). Le premier aspect porte sur la grande aventure lexicographique qui a permis au *TLF* d'être un pionnier dans l'utilisation des nouvelles technologies au service de la lexicographie. Dans son ouvrage *Dictionnaires et nouvelles technologies*, Jean Pruvost rappelle l'importance, dans ce domaine, de suivre le progrès : « de l'érudit-imprimeur, R. Estienne, auteur du premier dictionnaire de nomenclature française, au lexicologue-pionnier de la lexicographie mécanisée puis informatisée, B. Quemada, directeur du *TLF*, la démarche est la même : pour survivre et progresser, les lexicographes doivent être au plus près des technologies les plus performantes, et savoir les accompagner et les modifier à leurs fins. »¹ C'est ainsi qu'aidé par un gamma 60 bull, les lexicographes ont réuni un matériel impressionnant, de nature à donner une vision extrêmement large, à défaut d'être exhaustive, de la langue et de ses états. Des cartes perforées aux logiciels de plus en plus puissants, des millions d'informations ont pu être traitées de manière à faire naître le plus extensif des dictionnaires de langue française.

La seconde approche consiste à examiner comment on passe d'un dictionnaire élaboré à l'aide de l'informatique à un dictionnaire informatisé. Là encore, il s'agissait de suivre le progrès. Le développement de l'informatique et l'accès d'utilisateurs de plus en plus nombreux

¹ Jean Pruvost, *Dictionnaires et nouvelles technologies*, coll. « Écritures électroniques », Presses Universitaires de France, Paris, 2000, p. 5.

aux nouvelles technologies devaient incontestablement faire réfléchir les dictionnaristes, et les amener à inventer une nouvelle forme de dictionnaire. Le *TLF*, dans sa version papier, avait en effet traversé plusieurs époques lors de son élaboration, de ses très riches développements des premiers volumes à la plus juste mesure trouvée par la suite. L'objet n'était pas moins impressionnant : seize volumes, qui auraient pu, selon Jean Pruvost², atteindre les soixante si l'ouvrage n'avait trouvé de plus justes proportions à partir du quatrième et surtout du cinquième volume. L'avantage indéniable du modèle informatique est le gain de place, pour l'utilisateur — pour s'en convaincre, il suffit de comparer les seize volumes de la version papier et le compact-disc proposé dans la version informatisée, mais aussi pour le lexicographe, qui subit moins de pressions dictionnairiques. Il faut entendre par là que la quantité de signes ne joue plus dans l'univers du multimédia et que l'espace de texte ne connaît en conséquence plus de limitation. Il n'était cependant pas question de recommencer un *TLF*, ni de reprendre de bout en bout un ouvrage achevé : il fallait l'adapter à un nouveau modèle qui appelle d'autres formes de consultation. Pour cela, il était possible de s'inspirer du dictionnaire d'Oxford, l'*Oxford English Dictionary* ou *OED*, informatisé dès 1984. C'est en 1987 que naquit réellement l'idée d'un *TLFi*, qui exigeait, en raison de la logique informatique, une rigueur de rédaction encore accrue.

Jacques Dendien fut le maître d'œuvre du projet : il était en effet indispensable de créer un nouveau type de document structuré en fonction d'une feuille de style rigoureuse et bien définie, en d'autres termes, repenser intégralement la forme du texte, sans en changer le contenu. Comme l'explique Jacques Dendien, « informatiser un dictionnaire ne consiste pas à écrire des logiciels plus ou moins géniaux, mais consiste beaucoup plus prosaïquement à **le transformer en document structuré** »³. Le concept SGML (Standard Generalized Markup Language, un langage de balisage) fut utilisé pour décrire la structure des articles. Cela revient à définir les éléments constituant les articles et à les encadrer de balises textuelles afin d'indiquer, pour chacun d'eux, non seulement le type d'information contenue (définition, exemple, étymologie, bibliographie...), mais encore le début et la fin de cette dernière : une véritable syntaxe de l'article se dessine alors. Encore fallait-il que l'ensemble du texte fût disponible sur un support informatique, alors que les huit premiers tomes avaient été composés « au plomb », et que les tomes 11 à 13 présentaient un texte parfois lacunaire. Ce fut la Bibliothèque nationale de France qui en finança la saisie ; en 1994, grâce au soutien du CNRS et à l'énorme travail mené par le laboratoire dirigé par Jacques Dendien, lorsque fut achevée la rédaction du dernier volume, l'informatisation à proprement parler pouvait débuter, pour aboutir à l'« un des documents structurés les plus volumineux et complexe réalisé à ce jour »⁴.

Il ne faudrait pas s'imaginer, cependant, qu'informatiser le *TLF* revienne simplement à transposer un texte déjà lisible sur format papier sur un format numérique. L'utilisateur

² *Id.*, p. 95.

³ Jacques Dendien, « Histoire de l'informatisation du *TLF* » dans le livret d'accompagnement au *TLFi*, CNRS éditions, Paris, 2004, p. 14-15.

⁴ *Id.*, p. 19.

du *TLFi* est en droit d'attendre de nouvelles possibilités, et de nouveaux axes de recherche et de consultation, que nous allons à présent détailler.

2. STRATÉGIES DE CONSULTATION

L'informatisation représente un moyen très puissant de consulter le dictionnaire et elle touche un large public. La première version du *TLFi* a été mise en ligne sur Internet ; il s'agissait en quelque sorte d'un test grandeur nature qui permettait non seulement de vérifier le bon fonctionnement du nouveau système, mais aussi de mesurer son impact auprès du public. La présentation officielle se déroula le 5 mars 2002 : alors qu'au mois de février de la même année, moins de cinquante pages étaient consultées quotidiennement, fin mars, on en comptait plusieurs milliers, et on en a dénombré plus de 160 000 par jour en 2004⁵.

Avant de montrer quelle peut être l'étendue des possibilités de consultation du *TLFi*, il semble nécessaire de préciser les raisons pour lesquelles un usager de la langue est amené à consulter un dictionnaire : c'est, en effet, en fonction de ces raisons que peuvent s'expliquer les choix mis en œuvre dans la version informatisée du *TLF*.

Le recours le plus courant au dictionnaire s'explique sans conteste par la vérification orthographique. En consultant un dictionnaire alphabétique traditionnel dans ce but, le lecteur est souvent amené à tâtonner et à essayer toutes les orthographes possibles, parfois inattendues dans le cas de mots empruntés. Avec le *TLFi*, trois possibilités s'offrent au consultant dans l'onglet « Recherche d'un mot » : saisir le mot, dans ces conditions entre en jeu un correcteur orthographique intégré capable, entre autres, de rétablir les cédilles ou accents oubliés, mais aussi de proposer plusieurs mots d'orthographe approchante, utiliser les listes défilantes, dont la consultation rappelle celle d'un dictionnaire alphabétique traditionnel, ou bien effectuer une saisie phonétique — sans recourir à l'alphabet phonétique international, que tout lecteur n'est pas censé connaître, mais à l'aide d'un clavier phonétique dans lequel chaque son est clairement expliqué — qui dirige directement vers l'article recherché. Si l'on veut vérifier l'orthographe du mot *shampooing*, on cliquera ainsi sur les touches *CH-AN-P-OU-IN*, qui amèneront à une fenêtre proposant de se rendre à l'article *shampooing*, *shampoing*. On pourra aussi, comme dans de nombreux dictionnaires, s'assurer de la prononciation d'un mot, parfois périlleuse quand il s'agit d'emprunts tels que le *krish* malais, le *quattrocentiste* italien ou le *chiasme* grec.

On peut aussi souhaiter vérifier le sens d'un mot, auquel cas la question de l'orthographe ne se pose pas. Non seulement le *TLFi* propose une ou plusieurs définitions, mais on a aussi la possibilité de les rendre plus lisibles encore en peignant l'objet définition dans une couleur appropriée. De la même manière, il est possible de faire apparaître très nettement à l'écran les domaines techniques, exemples, auteurs d'exemples, codes grammaticaux... en somme tous les éléments constitutifs de l'article, ce qui permet, par exemple, de s'assurer que tel emploi est correct. La recherche d'expressions est d'ailleurs permise dans le cadre

⁵ Ces chiffres sont donnés par Jean-Marie Pierrel dans la Préface du livret d'accompagnement au *TLFi*, *op. cit.*, p. 4.

« tapez le mot » et l'utilisateur pourra aussi bien s'assurer de la validité d'une métaphore que se renseigner sur la correction de la syntaxe de tel membre de phrase.

Un écueil est parfois rencontré par les usagers de dictionnaires papier : celui de la lemmatisation. Un usager non averti peut ignorer que le terme *targui* correspond à la forme savante du singulier de *touareg*. Le *TLFi* conduit automatiquement à cette forme, alors qu'une consultation sur support papier aurait entraîné le lecteur à fouiller dans toutes les entrées de la lettre T, tâche décourageante s'il en est.

Toutes ces utilisations relèvent encore des méthodes les plus traditionnelles de consultation. Mais un dictionnaire informatisé ouvre d'autres perspectives bien plus novatrices, celles de la recherche transversale. Comme le signale Jacques Dendien, « il était nécessaire d'offrir à l'utilisateur des **services supplémentaires**, inédits dans le monde des dictionnaires électroniques, qui lui facilitent au maximum l'accès au *TLF*. »⁶ Si le dictionnaire papier voit souvent son utilisation limitée aux seules entrées, le dictionnaire informatisé permet d'ouvrir la recherche à l'ensemble du texte. Pour cela, il fallait mettre en œuvre un moyen de communication entre le lecteur et le *TLF* : le logiciel STELLA (Système de Traitement et d'Exploitation en Ligne de la Langue) réalisé par le service Développements informatiques de l'ATILE. À l'aide d'un langage spécialement conçu pour le *TLF*, STELLA permet de formuler des requêtes et d'extraire des résultats dans le corps-même de l'œuvre. C'est, dans le fond, un retour à l'analogie que permet l'informatique

3. L'ANALOGIE RETROUVÉE

« En simplifiant, on peut dire qu'informatiser le dictionnaire c'est le rendre lisible après coup sur ordinateur, non pas seulement en substituant à sa forme éditoriale classique celle d'écrans-pages, mais en transformant son contenu en base de données relationnelles qui permettra d'afficher des données que la consultation manuelle ne pourrait parfaitement rassembler — la consultation, alors, n'étant plus seulement une démarche de reconnaissance et de lecture, mais une procédure de découverte, c'est-à-dire une exploration. Informatiser le dictionnaire revient à en faire un dictionnaire électronique, destiné à la consultation humaine. »⁷ déclarent Gérard Gorcy et Françoise Henry. Il faut entendre par là que plus qu'un dictionnaire, le dictionnaire informatisé joue le rôle d'une véritable base relationnelle avec des affichages « intelligents » qui permettent au lecteur de formuler des requêtes précises.

En effet, contrairement à son frère de papier qui ne connaît qu'une méthode de consultation, formelle et alphabétique, le dictionnaire informatisé libère l'information de son carcan linéaire en lui rendant sa dimension analogique. C'est ce que relève Jean Pruvost

⁶ Jacques Dendien, « Histoire de l'informatisation du *TLF* » dans le livret d'accompagnement au *TLFi*, *op. cit.*, p. 20.

⁷ Gérard Gorcy et Françoise Henry, « Informatisation du *Trésor de la Langue Française (TLF)* : problèmes et perspectives » cités par Gérard Gorcy dans « Le *Trésor de la Langue française (TLF)* : son originalité et les voies ouvertes par son informatisation » in *Autour d'un dictionnaire: Le Trésor de la Langue Française, témoignages d'atelier et voies nouvelles*, Dictionnaire et lexicographie 1, publications de l'Institut National de la Langue Française, Centre National de la Recherche Scientifique, Didier Érudition, Paris, 1990., p. 195.

quand il écrit que « l'ordre analogique reprend tous ses droits à travers les recherches par critères, par auteurs de citation, par renvois et contraires, à travers les recherches croisées, les accès thématiques, etc. »⁸ Comment l'analogie est-elle mise en œuvre ? De quels biais l'utilisateur dispose-t-il ?

Nous avons déjà vu les possibilités proposées dans l'onglet « Recherche d'un mot », mais elles restent encore proches, bien qu'indéniablement plus puissantes, des méthodes de consultation traditionnelles. Avec les onglets « Recherche assistée » et plus encore « Recherche complexe » le *TLFi* ouvre les portes de la transversalité, puisqu'il est désormais possible de formuler des requêtes portant sur l'examen sélectif de certains éléments des articles d'un type donné, ce qui permet d'effectuer en un temps record des recherches qui auraient nécessité au moins une, sinon plusieurs, lectures intégrales du *TLF*. Il semble utile d'examiner un peu plus en profondeur ces deux méthodes. La recherche assistée permet de combiner cinq critères de sélection dont le code grammatical, l'indicateur d'emploi ou encore le type d'objet textuel recherché lié à un contenu spécifié. Ce dernier critère est précieux au plus haut point lorsqu'on s'intéresse de près aux emprunts. En effet, on peut sélectionner le type « langue empruntée », compléter le formulaire en spécifiant, par exemple, le contenu « amharique » et l'on obtient la liste des mots dont la notice étymologique signale qu'ils ont été empruntés à cette langue. Mais évidemment, il faut pour cela que l'emprunt ait été direct, car le *TLFi* lie obligatoirement le type et le contenu, c'est-à-dire que, dans la notice étymologique, on trouvera la mention « empr. à l'amharique ». Les fonctionnalités de la recherche complexe sont encore plus puissantes, puisqu'elle propose une transposition de la démarche humaine à l'aide du formulaire où l'on peut définir non seulement le type et le contenu textuel, mais encore la relation d'interdépendance entre eux sur six niveaux, ce qui représente un outil hors norme, d'une puissance inégalée.

Enfin, le *TLFi* propose de créer des listes de mots, soit manuellement, si l'utilisateur saisit lui-même une liste, soit automatiquement, ou par flexion d'un lemme, ou par extraction à partir de l'ensemble du texte du *TLF* et non plus seulement à partir des entrées. Dès lors, ce sont de véritables corpus de travail qui peuvent être constitués.

D'infinies possibilités d'accès aux données, de consultation, d'analyse, d'extraction d'informations, d'interrogations sont désormais ouvertes, ce qui représente sans doute une manne pour les chercheurs. Jacques Dendien écrit d'ailleurs à propos de l'informatisation du *TLF* qu'« elle offre à la communauté scientifique travaillant dans le domaine des industries de la langue un matériau d'une richesse incomparable permettant l'extraction de connaissances portant sur le français et la réalisation de moyens nouveaux pour en assurer le traitement informatique. »⁹ Constituer un corpus de travail, des exempliers, des anthologies, spécifier des requêtes, calculer des fréquences, dater la forme d'un syntagme sont autant de nouveaux horizons ouverts et à parcourir avec la facilité et la rapidité propres à l'informatique, de quoi s'exclamer qu'« il y a de la boîte de Pandore dans un dictionnaire informatisé, celle correspondant à la tradition de la jarre d'Héphaïstos contenant non pas

⁸ Jean Pruvost, *Dictionnaires et nouvelles technologies*, op. cit., p. 21.

⁹ Jacques Dendien, « Histoire de l'informatisation du *TLF* » dans le livret d'accompagnement au *TLFi*, op. cit., p. 7.

tous les maux mais tous les bienfaits, avec l'émergence imparable de nouvelles urgences à satisfaire. »¹⁰

Dans l'harmonie de ces louanges, on pourrait cependant relever un bémol. Certes, le *TLFi* n'est pas simplement un dictionnaire électronique proposant sur un écran les mêmes services que ceux dont dispose le lecteur sur la version papier : les moyens de consultation sont indubitablement plus variés, plus puissants, et ouvrent le champ à une véritable plongée au sein du texte, une exploration en profondeur du dictionnaire. Or la fidélité au texte du *TLF* interdit des mises à jour qui pourraient pourtant sembler nécessaires. En effet, si chaque article recense précisément toutes les connaissances à la pointe du savoir scientifique au moment de sa rédaction, ce savoir est susceptible de vieillir et de mériter parfois des compléments. Ainsi, on pourrait déplorer l'absence de mots, dans la nomenclature, pourtant d'emploi courant aujourd'hui. C'est le cas du *cookie*, que connaissent aussi bien les gourmands que les utilisateurs malchanceux d'Internet, malheureusement absent, puisque trop récent sans doute, des textes dépouillés. Il serait aussi possible de songer à compléter l'information bibliographique, d'enregistrer des sens plus récents, comme celui de *kif*, par exemple, et de nouvelles dérivations (*kiffer*), ou encore d'élargir la part faite aux sens inventoriés dans la francophonie. Gérard Gorcy cite l'exemple québécois d'*avoir les yeux comme des pistolets*, « avoir les yeux animés d'un grand sentiment de colère »¹¹, ignoré en France, mais pourtant existant et qui mérite donc sa place dans ce grand dictionnaire de la langue française. De tels écueils sont inévitables dans la mesure où la version informatisée du *TLF* se veut absolument fidèle à la version originale, mais l'informatisation ouvre des pistes non négligeables pour un apport ultérieur de compléments.

CONCLUSION

Né sous le signe de la modernité, bercé par les progrès informatiques, nourri des travaux les plus novateurs, le *TLF* s'est inscrit, dès sa conception, dans une perspective d'avenir tant lexicographique que dictionnaire. La parution de son dernier volume, en 1994, ne pouvait pas signer la fin de l'aventure. Des bases étaient posées, qui devaient tracer un nouveau chemin dans la forêt dictionnaire, pour reprendre l'image de Franz-Josef Hausmann. L'informatisation de l'œuvre n'était qu'une première étape. Comment résister au plaisir de plonger dans le fonds extraordinaire réuni pour prolonger l'ouvrage ? À peine la version informatisée était-elle achevée que foisonnaient de nouveaux projets : « nos recherches en lexicologie et lexicographie ne vont pas s'arrêter là. Si cette version informatisée correspond à un reflet fidèle de la version du *TLF* précédemment publiée, notre objectif est de la faire vivre et d'y intégrer, dans des versions ultérieures, des améliorations de contenu liées à l'évolution de nos connaissances sur la langue, son histoire et son usage, mais aussi des ajouts prenant en compte des changements de la langue au cours des périodes les plus récentes afin de permettre au *TLFi* de demeurer le dictionnaire de référence de la

¹⁰ Jean Pruvost, *Dictionnaires et nouvelles technologies*, op. cit., p. 80.

¹¹ Gérard Gorcy, « Le Trésor de la Langue Française (*TLF*) : son originalité et les voies ouvertes par son informatisation » in *Autour d'un dictionnaire : Le Trésor de la Langue Française, témoignages d'atelier et voies nouvelles*, op. cit., p. 206.

langue française. »¹² Par ces mots, Jean-Marie Pierrel montre une volonté intacte de servir la langue, d'en suivre les mouvements, d'en étudier les évolutions, de faire en somme un travail inédit en traquant l'inouï.

RÉFÉRENCES BIBLIOGRAPHIQUES

DENDIEN, J. 2004. « Histoire de l'informatisation du *TLF* » dans le livret d'accompagnement au *TLFi*, CNRS éditions, Paris, p. 7–24.

GORCY, G. 1990. « Le *Trésor de la Langue Française (TLF)* : son originalité et les voies ouvertes par son informatisation » in *Autour d'un dictionnaire : Le Trésor de la Langue Française, témoignages d'atelier et voies nouvelles*, Dictionnaire et lexicographie 1, publications de l'Institut National de la Langue Française, Centre National de la Recherche Scientifique, Didier Érudition, Paris.

HAUSMANN, F.-J. 1988. « Le *T.L.F.* prix d'excellence ? » in *Zeitschrift für französische Sprache und Literatur*, XCVII/2.

IMBS, P., QUEMADA, B. (dir.) 1971–1994. *Trésor de la Langue Française : Dictionnaire de la langue du xix^e et du xx^e siècle (1789-1960)*, 16 volumes, éditions du CNRS/Gallimard, Paris.

PIERREL, J.-M. 2004. Préface du *Trésor de la langue française informatisé*, Centre National de la Recherche Scientifique, Université de Nancy-2, Analyse et Traitement Informatique de la Langue Française, CNRS éditions, Paris, p. 1–5.

PRUVOST, J. 2000. *Dictionnaires et nouvelles technologies*, coll. « Écritures électroniques », Presses Universitaires de France, Paris.

QUÉMADA, B. (dir.) 1990. *Autour d'un dictionnaire : le Trésor de la Langue Française, témoignages d'atelier et voies nouvelles*, Dictionnaire et lexicographie 1, publications de l'Institut National de la Langue Française, Centre National de la Recherche Scientifique, Didier Erudition, Paris.

CÉDÉROM

TLFi, Trésor de la Langue Française informatisé, CNRS/Université Nancy 2/ATILF, CNRS éditions, Paris, 2004.

(version Internet : <http://stella.atilf.fr/>)

Sabine Albert

The University of Cergy-Pontoise

Research interest: lexicography, loan-words, constitution of French and English lexicons

COMPUTERIZED DICTIONARIES, NEW POSSIBILITIES : THE CASE OF THE TRÉSOR DE LA LANGUE FRANÇAISE

Summary

Since the beginning, when it was decided on the project of the *Trésor de la Langue Française*, the most advanced technologies were used to assist the lexicographers. The use of the computer helped to gather important material, giving access to an enlarged vision of language. But if informatics was an essential point from the conception of the dictionary, through the computerization of the *TLF* itself,

¹² Jean-Marie Pierrel, Préface du *Trésor de la langue française informatisé*, *op. cit.*, p. 4-5.

another step was taken, since new strategies of consultation appeared for the user. There are a lot of advantages in setting forth a computerized version of a dictionary which necessarily takes a reduced space, while permitting to abolish any limitation for the text space. The *Oxford English Dictionary*, computerized since 1984, furnished the model on which the lexicographers founded their work. The text had to be changed into a structured document in order to define the constitutive elements of each article and to organize them with textual tags. Thanks to the extreme precision in the structuring, it is possible to search for a word through three different ways: by entering a word, using scrolling lists, or leaning upon phonetics. In the text of the article, different zones can be painted to highlight them (definitions, synonyms, etc) and there is no more problems of lemmatization. Moreover, the user can formulate very precise requests: the assisted research permits to combine five selection criteria while the complex research transposes the human process in a form where textual type and content are defined, specifying on six levels of interdependence links. Other applications can also be found, as the creation of words lists, the calculation of frequencies is a godsend for a researcher. Unfortunately, one problem remains: the impossibility of updating since the computerized version of the *TLF* is still the reflection of the dictionary, definitively finished in 1994.

KEY WORDS: lexicography, lexicology, loan-words, computerized dictionary, informatics.

Jteikta 2014 metų liepos 15 d.