

De la genèse à l'informatisation du *Trésor de la langue française* et du *Grand Robert électronique*

Jean-N. DE SURMONT
jdesurmont@yahoo.fr

RESUMEN

En un período de seis meses en 2004 y 2005, la publicación electrónica de diccionarios de lengua francesa ha conocido una renovación importante con la publicación del *Grand Robert électronique* en mayo de 2005 y la primera edición del *Trésor de la langue française informatisé* en disco compacto de la versión *on line* (aparecida en noviembre de 2004). Es recomendable examinar algunas características de estos dos diccionarios evocando la génesis de sus dos proyectos lexicográficos hasta las innovaciones electrónicas recientes.

Palabras clave: lexicografía (historia), edición electrónica, diccionarios franceses, *Trésor de la langue française*, *Grand Robert*.

RÉSUMÉ

En l'espace de six mois en 2004 et 2005, l'édition électronique des dictionnaires de langue française a connu un profond renouvellement avec la réédition du *Grand Robert électronique* (désormais *GRE*) en mai 2005 et la première édition du *Trésor de la langue française informatisé* en version cédérom (paru en novembre 2004). Il convient d'examiner quelques caractéristiques de ces deux dictionnaires en évoquant la genèse des deux projets dictionnaires jusqu'aux innovations électroniques récentes.

Mots-clé: lexicographie (histoire), édition électronique, dictionnaires français, *Trésor de la langue française*, *Grand Robert*.

En l'espace de six mois l'édition électronique des dictionnaires de langue française a connu de profonds changements avec la réédition du *Grand Robert électronique* (désormais *GRE*) en mai 2005 et la première édition du *Trésor de la langue française informatisé* (désormais *TLFI*) en version cédérom (paru en novembre 2004). Il convient d'examiner quelques caractéristiques de ces deux dictionnaires, sans aller dans les détails techniques des potentialités de consultations. Nous allons plutôt évoquer la genèse des deux projets dictionnaires jusqu'aux innovations électroniques récentes. Sans faire l'histoire des éditions du *Grand Robert* (désormais *GR*) et de l'édition du *Trésor de la langue française* (désormais *TLF*) nous expliquerons quelques faits saillants de leur développement tout en les rapprochant et en montrant leurs forces et leurs lacunes.

LE MOUVEMENT DE L'ÉDITION ÉLECTRONIQUE: LE TRÉSOR DE LA LANGUE FRANÇAISE

Le *TLF* constitue un des dictionnaires les plus ambitieux de la langue française. Plutôt que de décrire la langue depuis ses débuts comme le fait le *Oxford English Dictionary* pour l'anglais, les Français ont choisi de produire plusieurs dictionnaires synchroniques (bien qu'au début les objectifs furent les mêmes que celui d'Oxford)¹. Une vaste entreprise de description du français moderne a donc donné lieu à de nombreuses publications depuis une trentaine d'années. En novembre 2004, les éditions du Centre National de la Recherche Scientifique lançaient le *TLFI* en version cédérom, dix ans après la parution du dernier volume en 1994, le 16^e, du dictionnaire dont il est le produit dérivé le *TLF*. Ce produit cédérom s'ajoute ainsi à ceux que publient Hachette, Robert et Larousse (équivalents de leur produit papier) et à Champion et Redon qui lancent depuis une dizaine d'années des dictionnaires électroniques qui sont pour la plupart des rééditions de dictionnaires de langue relativement anciens.

Le projet de mise en ligne du *TLF* succède à celui de la base de données Frantext dont le corpus textuel a servi en partie de base linguistique aux dossiers de mots. En 1982, trois membres de l'Institut Nationale de la langue française (INaLF) arrivaient à Chicago afin d'élaborer une base de données textuelles. En 1989 la base de données textuelle Frantext était mise en ligne par les étatsuniens². Depuis 1998, le *TLF* est mis en ligne par les bons soins du Project for American and French Research on the Treasury of the French Language (ARTFL) stationné à Chicago et le supplément est progressivement ajouté à partir de 2000. N'est-il pas étonnant que ce projet d'envergure qui consistait à élaborer le plus grand dictionnaire de la langue française jamais paru (100 000 mots avec leur histoire et étymologie), ait d'abord été mis en ligne par les étatsuniens ? C'est que les dimensions du *TLF* n'ont rien de celles d'un dictionnaire conventionnel. Il contient 270 000 définitions, 430 000 exemples, 350 millions de caractères, 23 000 pages, 100 000 mots, 450 000 entrées et 430 000 citations précisément identifiées (à 80 % des citations littéraires mais peu extraites de recueils de poésies et la paralittérature est presque absente³), plus de trente millions d'occurrences scientifiques (le *TLF* comme le *GR* regorge de définitions terminologiques) ou techniques et soixante-dix millions d'occurrences littéraires. Jusqu'en 1985 l'INaLF avait relevé 900 000 néologismes au sein de sa base informatique Bornéo.

Le *TLF* fut la première entreprise dictionnaire française à employer, au début des années 1960 (au moment où le recteur Paul Imbs fondait le Centre de recherche pour un Trésor de la langue française qui deviendra à la fin des années 1970 sous la

¹ Le projet dictionnaire comprenait initialement 4 phases depuis les origines jusqu'au XX^e siècle, c'est la dernière phase qui a été retenue.

² Robert Morissey, 1993: p. 28.

³ En cela, la reconnaissance institutionnelle des auteurs de langue française relève surtout des auteurs de prose et néglige la poésie. En outre, étant donné le peu d'auteurs non-hexagonaux, on peut considérer qu'il s'agit surtout de décrire le français des auteurs connus et reconnus. (Voir H. Caputo, 1992: p. 78).

houlette de Bernard Quemada, l'INaLF), des techniques d'automatisation linguistique. Bien que sa confection ait surtout été réalisée à Nancy (44 avenue de la Libération, incidemment près de la Faculté de lettres de l'Université Nancy 2) les collaborateurs et les centres de recherches associés émergeaient tout aussi bien à Besançon qu'à Villetaneuse en banlieue parisienne pendant des périodes plus ou moins longues.

Paul Imbs avait reçu pour mission de réunir la documentation, d'assurer la rédaction et la publication du *TLF*. Le *TLF* se voulait, selon l'expression de Jean Pruvost, métalexigraphe laroussien réputé, une espèce de "décentralisation intellectuelle". *Dictionnaire non-conventionnel* il en est un aussi parce qu'il a cherché à rénover la science étymologique par un corpus d'œuvres littéraires qui dépasse largement celui des dictionnaires commerciaux comme le *GR* ou le *Grand Larousse de la langue française*. Cela n'empêche guère ces deux dictionnaires, pour ne nommer que ceux-là, de faire partie du corpus métalinguistique du *TLF* auquel s'ajoutent des centaines de dictionnaires et de grammaires de toute sorte dont une bonne partie sont localisés à Nancy. Dans une perspective *phylogénèse dictionnaire*, on peut ainsi constater les emprunts et plagiats entre dictionnaires de tradition différente.

GENÈSE DU TRÉSOR DE LA LANGUE FRANÇAISE

La naissance du *TLF*, dont le premier tome voit le jour en 1971, s'inscrit dans la mouvance d'une initiative étatique, représentant l'un des importants projets du président de la République Charles De Gaulle, tout comme la Bibliothèque Nationale le fut sous le septennat de François Mitterand. Le *TLF* a embauché dans les années 1960 une jeune génération d'une centaine de chercheurs rattachée à la composition, à la saisie (le dernier texte pour le XX^e siècle fut saisi en 1964) et à la correction des textes du corpus textuel et des articles de dictionnaires. Grâce à l'usage de l'ordinateur Gamma 60 Bull, une importante documentation fut traitée mécaniquement. Mais à l'époque, le dictionnaire électronique comme nous le connaissons sous sa forme actuelle n'était encore qu'un mirage dans le giron des spécialistes. En effet, Georges Matoré affirme en 1968: "Peut-être que les ouvrages lexicographiques du XXI^e siècle se présenteront-ils non sous la forme de livres, mais sous celle de fichiers minuscules, pourvus de dispositifs de lecture entièrement inédits. Le rêve n'est pas interdit au lexicographe qui, le soir venu, laisse sa pensée vagabonder au-delà de son travail quotidien"⁴. Dès 1970, quand débutait la rédaction, près de 80 millions d'occurrences étaient devenues lisibles par ordinateur⁵ auquel s'ajoutait les 6 millions de fiches de l'Inventaire générale de la langue française (IGLF) constitué à l'initiative de Mario Roques.

Si l'élaboration du *TLF* a méthodologiquement rénové la science étymologique en ce qu'il a travaillé avec un gros corpus; ses références sont souvent restées

⁴ Matoré, 1968: p. 246.

⁵ Robert Martin, 2000: p. 970.

imprécises faute de temps pour aller vérifier les datations. Le problème avait déjà été de longue date constaté par l'un des membres du comité de direction Matoré. En 1968, il écrit:

Le premier problème auquel ont à faire face les collaborateurs du Trésor est celui de la documentation: il s'agit de réunir au Centre un fonds complet de lexicographie et de lexicologie française ou romane, c'est-à-dire de grouper tous les fonds ou inventaires lexicologiques constitués par des savants décédés et non utilisés par eux (Godefroy, Delbrouille, Durafour, etc.), d'absorber les fiches de l'Inventaire général de la langue française, fondé par Mario Roques [dont le principe inspira l'équipe du Trésor de la langue française du Québec], et enfin de réunir des renseignements exhaustifs sur les différentes époques du vocabulaire français. Cette documentation "artisanale" sera complétée par une documentation obtenue mécaniquement grâce à un ordinateur qui dépouillera une masse d'environ 250 millions de mots tirés de textes littéraires (85 millions pour les XIX^e et XX^e siècles, 60 millions pour les XVII^e et XVIII^e siècles, etc.)⁶.

Souvent les datations qui se sont avérées imprécises, notamment, parce que l'on ne citait pas l'édition d'origine. Depuis le 1^{er} janvier 2005, l'Atilf procède ainsi à une révision progressive des rubriques "Étymologie et histoire" de l'article lexicographique du *TLF* en y réinjectant le résultat des recherches récentes et des sources nouvellement publiées. Comme signale Laurent Bray, on retrouve aussi dans le *GR* des datations inexactes aux entrées *arbre*, *arrêt*, *assemblage*, etc. qui présentent Richelet 1680 comme la première attestation alors qu'il s'agit de l'édition lyonnaise de 1728⁷.

La source la plus importante qui a servi à retourner aux racines de notre langue fut incontestablement le *Französische etymologische Wörterbuch* élaboré jadis à Bâle sous la tutelle de W. von Wartburg⁸. Le *TLF* est indubitablement une entreprise lexicographique de grande ampleur, héritière des entreprises lexicographiques du XIX^e siècle, dans la mesure où l'importance en volume de la documentation traitée au XX^e siècle est inégalée. En plus, son influence sur la rédaction d'articles lexicographiques tant au Québec (*Dictionnaire historique du français québécois* (1998), qu'en Suisse (*Dictionnaire suisse romand* 1997) est capitale. De l'ampleur de cette œuvre résulte une absence de réédition et donc de mots récents ce qui rend forcément une partie de la description de la nomenclature désuète. En définitive, la question se pose de savoir si le *TLF* répond d'avantage aux objectifs d'un dictionnaire d'usage qu'à celui d'un dictionnaire philologique.

Si les dirigeants se targuent de ne pas avoir de concurrents, ce n'est pas sans que se tissent des relations houleuses dans le milieu. Nul ne peut cacher que le succès des uns fait la jalousie des autres. Il arrive que chercheurs dénigrent les talents intellectuels pour les subordonner aux traits de personnalité, encore faut il que cette

⁶ Georges Matoré, 1968: p. 187.

⁷ 1995: p. 98.

⁸ Son *ex libris* est d'ailleurs hébergé dans les bâtiments de l'ATILF.

réussite soit le fruits des mêmes conditions de travail. Le *TLF* ne répond pas aux mêmes objectifs que le dictionnaire de l'Académie française qui est de nature plus prescriptive et ne donne pas le même éventail de définitions. Malgré ses grandes qualités, le *TLF* est un dictionnaire produit par l'élite et pour l'élite et soutenu par une politique pourtant démocratique. N'est-il pas, comme le dit Alain Rey, "le plus important modèle du lexique dans l'histoire des dictionnaires de langue française"⁹, tout en ajoutant que "les nouveautés technoscientifiques, familières ou argotiques sont moins bien représentés"¹⁰. Rey soulève un problème qui n'est pas étranger à celui de l'étendue de la dite description synchronique (1789 à nos jours) occasionnant des états d'incompatibilité fonctionnelle.

S'il n'est pas un dictionnaire commercial comme le *GR*, il n'est pas étonnant dans ce contexte que le *TLF* et les publications de l'ATILF soient essentiellement destinés aux chercheurs. Mais depuis peu l'ATILF s'est rapprochée du milieu universitaire; elle est devenue une unité rattachée à l'Université Nancy 2 et a ainsi étendu son réseau d'influence. Même si ce dictionnaire est une œuvre de scientifiques "chevronnés", encore faudrait-il rendre à César ce qui est à César en permettant aux chercheurs de toutes les avenues de bénéficier de ces ressources uniques. La politique d'accès à ces grands trésors de l'Etat, une documentation unique en France réunit en des mêmes locaux, répond elle vraiment à la politique démocratique dans le cadre de laquelle ils ont été rassemblés? En effet, l'accès aux locaux très restreint, faute d'espace, valorise parfois le séjour de chercheurs étrangers (Bulgare, Russe, etc.) au détriment des Français vu les volontés de diffusion de la langue française à l'étranger. Même à Nancy, où le dictionnaire et la base de données textuelle Frantext de 4000 œuvres ont été élaborés, le base Frantext restait encore inexploitée par les étudiants de l'université en 2001 (un cédérom, Discotext 1, corpus de trois cents œuvres littéraires (1827-1923) en est issu). Le *TLFI* était jusque tout récemment consulté par un tiers d'enseignants et de chercheurs. Il faudrait aussi que les chercheurs locaux y aient accès ou en connaissent l'existence. L'une des critiques de la communauté de recherche en France était que la véritable richesse de la France n'était pas tant le dictionnaire que le corpus de textes.

La version cédérom rendra donc plus accessible une analyse détaillée du français des XIX^e et XX^e siècle. Un balisage raffiné de plus de 36 millions de balises XML, 17 364 854 balises typographiques, 1 070 224 balises décrivant la hiérarchie, 18 178 634 balises repérant les objets textuels, permet une convivialité accrue. L'informatisation du *TLF* avait suscité de nombreux débats et réflexions donnant lieu à un colloque au milieu des années 1990 deux ans après la sortie de la première édition du *GRE*. C'était alors la phase de traitement informatique (traitement des archives de photocomposition à partir de 1992), assurée par le CNRS au sein de l'INaLF. Une partie des chercheurs continuait néanmoins les recherches portant sur

⁹ 1977 : p. 150. Au moment où Rey rédigeait ces lignes il était conseiller de rédaction au TLF, collaboration qui s'est terminé environ en 1977. Le pouvoir politique de l'époque influençait semble-t-il de manière importante le choix des directeurs de laboratoire. Source: "Pierre Lerat, conversation personnelle, 8 juin 2005.

¹⁰ *Ibid.*

d'autres synchronies comme l'Ancien ou le Moyen Français, la rédaction des articles du complément, ou du *Französische Etymologische Wörterbuch* abrité dans les locaux de l'ATILF et dirigé par Jean-Pierre Chauveau.

La création de l'ATILF en 2001 n'est pas qu'une simple reconfiguration administrative de l'INALF. Elle correspond au fait que devenue depuis plusieurs années la langue française est une véritable industrie nouée par des réseaux internationaux. L'institut nancéien devait se mettre au goût du jour et c'est ce qu'il a fait en diversifiant aussi les partenaires de son réseau scientifique (Université Nancy 2, Université Nancy-Metz, etc.).

Même si le *TLFI* et par conséquent le cédérom n'ont pas pu tenir compte de la réforme de l'orthographe, l'ATILF a néanmoins publié en collaboration avec le premier ministre d'alors Lionel Jospin, un guide de féminisation des grades et fonctions. Celui-ci arrivait sur le marché près de dix ans après les publications du Québec et de la Belgique dont les dénominations marquent parfois un écart avec l'usage réel. La parité des salaires est sûrement un maigre bond par rapport à l'intégration dans la langue de la reconnaissance de la femme. Encore faut-il que les femmes elles-mêmes acceptent le compromis en reniant une partie des us et coutumes de la tradition linguistique.

Bref, la mise sur cédérom n'est ni une pure stratégie de marketing, ni un pur rejeton de la République. Il répond au fait que la demande en ligne ne cesse d'augmenter et que les utilisateurs avaient de plus en plus besoin d'insérer le dictionnaire dans d'autres applications comme Explorer et Word. La convivialité des rapports entre Word et le *TLF* en ligne (depuis mars 2002) laissant parfois à désirer, on a décidé d'améliorer l'accès aux ressources informatiques. L'approche du dictionnaire a évolué, est devenue plus structurée, logique et permet de stocker une masse d'informations, de faire des recoupements entre les critères de recherches. L'interface permet trois possibilités graduées de recherche: la recherche simple: consistant à rechercher les articles concernant le mot tapé par l'utilisateur, la recherche assistée proposant un formulaire de recherche permettant à l'utilisateur d'imposer des contenus textuels à des types d'objets données et enfin la recherche complexe requête plus complexe que l'on retrouve aussi dans le *TLFI*. Dans ce cas, l'utilisateur peut spécifier les types des objets recherchés, leurs contenus textuels éventuels et leurs relations hiérarchiques¹¹. Mais l'examen du cas du *TLFI* est généralisable à l'analyse des structures de beaucoup d'autres dictionnaires électroniques. Ainsi, le transfert "du dictionnaire 'texte-papier' sur support électronique, en somme son inscription dans les schémas fonctionnels du dispositif informatique, affecte son potentiel significatif (limitation des processus de construction de sens). En effet l'intelligibilité du texte lexicographique prend appui sur des caractéristiques typographiques et topographiques (positionnements relatifs des composantes textuelles, et [Ide, Le Maître & Véronis] montrent clairement comment certaines segmentations opérées dans le texte d'un article en vue de l'établir sous forme de base de données relationnelle ruinent son unité sémantique et, corollairement, sa portée informative¹².

¹¹ Voir Jean-Marie Pierrel, Jacques Dendien et Pascale Bernard, 2004: p. 168.

¹² David Piotrowski, 2004: p. 118.

“Par contre, de dire les concepteurs mêmes du dictionnaire, chercheurs et administrateurs à l’ATILF, Jean-Marie Pierrel, Jacques Dendien et Pascale Bernard, le TLF se distingue des autres dictionnaires électroniques par la finesse de la structuration des données en ‘objets’ interrogeables selon divers critères, et par une interface simple et conviviale [...]”¹³.

Les forces et lacunes se côtoient et la plus grande qualité que l’on peut affubler à ce produit dictionnaire est le fait de rendre les données plus accessibles. La nouveauté du cédérom est aussi de permettre d’utiliser le dictionnaire, même si l’on n’est pas connecté à un haut débit sur internet. En somme, si l’on a démocratisé l’accès aux ressources déjà publiées de l’ATILF, la démocratisation de la langue n’est pas acquise dans tous les secteurs et on gagnerait à y faire voir davantage ses *trésors* cachés pour qu’en définitive, les chercheurs puissent se faire les ambassadeurs de cette langue si *magnifique*. Ce dictionnaire cédérom sera apprécié d’autant plus que l’édition sur support papier est difficilement accessible. Il constitue un *lieu de mémoire* au sens où l’entend l’historien Pierre Nora.

LE GRAND ROBERT ÉLECTRONIQUE: COMBLER LE “VIDE LEXICOGRAPHIQUE”

Bien que le *TLF* ait été le premier à utiliser des moyens mécaniques pour l’élaboration du corpus linguistique¹⁴, le *GRE*, édité la première fois en 1994, constitue le point de départ de l’informatisation des dictionnaires imprimés quatre ans avant que s’engage, en 1998, un véritable mouvement d’informatisation du corpus dictionnaire par les maisons Redon¹⁵ et Champion électronique¹⁶. En 1998 le *Robert Junior* (fonctionnant avec 16 Mo de Ram en version PC) est lancé et, deux ans plus tard, le *Petit Larousse* et le *Petit Robert* (32 Mo de Ram en version PC).

L’AVENTURE ÉDITORIALE DES DICTIONNAIRES ROBERT

L’aventure éditoriale des éditions Robert, anciennement la Société du Nouveau Littré (cette dénomination cherchait à marquer l’esprit de continuation entre le lexicographe du XIX^e siècle et le dictionnaire rédigé sous la direction de Paul Robert),

¹³ Jean-Marie Pierrel, Jacques Dendien et Pascale Bernard, 2004: p. 165.

¹⁴ L’apport de l’informatique du TLF a été brièvement et clairement analysé par Robert Martin, 2000.

¹⁵ Redon se lança notamment dans l’informatisation des dictionnaires de l’Académie française qui sortent en novembre 2000 et le Littré qui sort en novembre 1997, l’*Atelier historique de la langue française* qui comprend sept dictionnaires sort en avril 1999, Le *Grand Atelier de la langue française* comprenant sept dictionnaires supplémentaires en novembre 2001, etc.). Les Dictionnaires Robert acquièrent Redon en 2001.

¹⁶ Champion pratique des prix plus élevés offrant des produits un peu différents de ceux de Redon comme le *Dictionnaire de l’ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle* Frédéric Godefroy mais aussi, comme Redon, les dictionnaires de l’Académie Française et le *Grand dictionnaire universel* de Larousse. En revanche l’espace disque dur disponible nécessaire pour supporter les produits Champion est moins important en plus que le processeur suffisant pour alimenter n’est que le PC 486.

commence à Casablanca dans les années 1950. En 1955, une partie de la rédaction déménage à Paris évitant la mobilisation de certains rédacteurs vu les difficultés au Maroc. En août 1964 le dernier fascicule du *Dictionnaire alphabétique et analogique de la langue française* est publié (paru en six volumes entre 1953 et 1966. Dans sa deuxième édition en 1985, il deviendra le *Grand Robert de la langue française*. On a depuis oublié que Paul Robert avait été approché par Paul Imbs pour la direction du *Trésor de la langue française*. En effet, Paul Robert écrit dans son récit autobiographique *Aventures et mésaventures d'un dictionnaire*:

Paul Imbs me reçut, au Centre de philologie romane, d'une façon extrêmement cordiale et de plus en plus chaleureuse, à mesure que se révélait entre nous une parfaite identité de vues sur les questions que nous envisagions, notamment sur l'élaboration d'un "Trésor de la langue française" qu'il m'engageait à promouvoir. Il me confirmait par lettre, quelques jours plus tard, son souhait de me voir prendre la tête du mouvement et diriger l'exécution du projet avec mon équipe de collaborateurs, sous le patronage de hauts comités. Je répondis que j'étais, en principe, d'accord avec lui, mais que je ne pourrais rien entreprendre avant l'achèvement du dictionnaire, c'est-à-dire avant plusieurs années, quatre au moins, puisque je pensais, alors terminer mon ouvrage en cinq volumes¹⁷.

En 1967, le dictionnaire le *Petit Robert* est publié. Il est extrait du GR achevé en 1964. Depuis 1971, il comporte un deuxième tome consacré aux noms propres dont certaines biographies, notamment celles des auteurs cités, sont reprises dans le *GRE*. L'aventure du *Robert électronique* commence en 1987 par la collaboration entre les équipes rédactionnelles du Robert et les informaticiens du bureau van Dijk aux Pays-Bas. Deux ans de travail préparatoire vont permettre la création d'outils d'aide à la rédaction sur disque optique compact. Si l'on se fie au propos d'Alain Duval, on pourrait évaluer le corpus citationnel du GR de 1985 à 158 636, celui de son corpus à 10 000 ouvrages écrits par 1250 auteurs¹⁸. La version électronique reproduit la dernière édition du GR, c'est-à-dire 86 000 entrées, 13 500 pages et 120 millions de signes, 200 000 définitions et 350 000 sens. Le *GRE* est une version enrichie de la deuxième édition remaniée et comporte 410 000 exemples et 175 000 citations (par rapport à 158 636 de l'édition de 1985) ainsi que de nouveaux mots et de nouvelles acceptions par rapport à l'édition de 2001. La nomenclature et le corpus citationnel se sont donc enrichis de plusieurs milliers de mots depuis l'édition de 1985. Le GR a été élaboré dans l'esprit de la description alphabétique illustrée par des exemples littéraires, analogue au Littré mais s'intéressant davantage à l'usage de ses contemporains?. Il constitue un dictionnaire "analogique" qui regroupe les expressions diverses d'une même "idée" et s'est inspiré au départ de *La pensée et la langue* de Ferdinand Brunot, comme l'affirme Alain Rey dans la préface de la deuxième édition du *Grand Robert de la langue française*¹⁹.

¹⁷ 1966: p. 116.

¹⁸ 1995: p. 469.

¹⁹ Préfaces, 2005: p. XIX.

Les différentes statistiques mentionnées ci-dessus peuvent servir d'éléments de comparaison entre le *TFLI* et le *GRE*. En outre, ce qui distingue le *GR* du *TLF* est que ce dernier n'a pas été remanié au moment de son édition électronique. Le *GRE* offert au public constitue la deuxième édition du *GR* et la dernière édition électronique parue en mai 2005 intègre les modifications apportées par l'équipe rédactionnelle en 2001. Certes le *TLFI* a sorti sa version cédérom en novembre 2004 six mois avant celle du cédérom du *GR* mais ce dernier en était à sa deuxième édition.

LES AMÉLIORATIONS DE L'ARTICLE LEXICOGRAPHIQUE

Il est nullement notre intention de présenter l'ensemble des modifications apportées au *GR* entre sa première mouture et sa deuxième édition en 1985 (ce que révèle ce que je nomme l'*ontogenèse dictionnaire*). En revanche, on peut tout de même évoquer le fait que dans la première édition les variations patronymiques des 120 000 citations (approximativement.) étaient plus importantes que dans la deuxième édition. Laurent Bray relève que le titre *Le Petit chose* de Daudet est référencé de quatre façons DAUDET, s. v. *aussi*, cit. 25, DAUD. s. v. *aussitôt* cit. 4, A. Daudet, avec cit. 31 et A. Daud., s. v. *aviser* cit. 5²⁰. Cela s'ajoute aux nombreuses références bibliographiques tout simplement incomplètes et donc non opératoires. Les rédacteurs de l'édition de 1985 ont uniformisé les éléments identificateurs des références mais d'une manière systématique. Pourtant, la politique du Robert semble être claire à ce sujet. Ainsi A. Rey affirme: "[c]es datations sont présentées dans le *Grand Robert* quelquefois avec une référence complète, parfois avec la mention d'un auteur ou d'un texte. Il nous a en effet semblé que ces références, indispensables pour le spécialiste, n'offraient qu'un intérêt limité si elles n'étaient pas très précises, ce qui auraient occupé une place trop importante —et déjà autrement utilisée— dans l'ouvrage; l'utilisateur sera plus éclairé par une date ou une époque que par la mention d'auteurs et de textes souvent peu connus"²¹.

Il en va autant des marques diatopiques relatives aux belgicisms dont on retrouve plusieurs formules syntaxiques. Ce manque d'homogénéité dans la rédaction de l'article lexicographique et des formules métalinguistiques peuvent témoigner de l'hétérogénéité des rédacteurs et d'un manque d'indications dans les guides de rédaction destinés aux auteurs des articles. Le *GR* propose deux catégories de citations généralement placées l'une après l'autre sans pratique d'homogénéité contrairement à la pratique du *PR*:

- les citations proprement dites, entre guillemets et référencées, nettement identifiables (corps typographique distinct, alinéa), numérotées et souvent fort longues;
- les autres extraits, en italiques, en général sans guillemets, suivis d'un renvoi à l'entrée sous laquelle se trouve la citation numérotée dont est tiré le fragment"²².

²⁰ Laurent Bray, 1995: p. 93.

²¹ *Préfaces*, 2005: p. XXXIX.

²² Alise Lehmann, 1995: p. 108.

Le *GR* n'a certes pas bénéficié de l'informatique dès le début de sa confection, mais il a inauguré le mouvement des dictionnaires électroniques de langue française utilisant les technologies d'encodage XML et des programmes de contrôles éditoriaux. Les changements qui lui ont été apportés, du moins ceux que nous signalent les préfaces et la lecture de quelques travaux relatifs au contenu lexicographique, sont surtout relatifs à la précision des référenciations, à l'amplification et au remaniement de la nomenclature. Mais l'originalité du *GRE* est notamment de mêler les informations encyclopédiques du *Petit Robert des noms propres* (désormais *PR 2*), dont la première édition paraît au début des années 1970, à la nomenclature des mots de la langue. Ainsi, 2000 notices issues du *PR2* constitue une espèce de dictionnaires des auteurs intégrés faisant référence aux grands écrivains de la langue française (y compris certains linguistes) ainsi qu'aux grandes œuvres citées dans le *GR*. Les références sont accessibles par hyperliens depuis la citation du dictionnaire. Une bibliographie générale de 3400 entrées est également accessible par le nom des auteurs à partir de n'importe quelle référence de citation. De plus, on peut accéder à toutes les citations (environ 200 000 liens) issues des renvois provenant d'autres articles. Par conséquent, En ce sens le *GRE* constitue un outil doté de fonctionnalités conviviales (améliorant le logiciel de consultation précédent) et permet la construction d'un véritable réseau de liens. Le cédérom peut être appelé via n'importe quelle autre application comme le traitement de texte notamment. Une telle manipulation posait problème quant à la version en ligne du TLF.

CONCLUSION

C'est dans la lexicographie française du XVII^e siècle qu'on pourrait retrouver les sources de la lexicographie de ces deux dictionnaires. Laurent Bray soutient que "[t]ous les deux innovent. Richelet livre avec une honnêteté quasi scientifique les références des attestations, établit partiellement son dictionnaire à partir d'un corpus de données et préfigure la méthodologie du Trésor de la Langue Française qui ne décrit rien qui ne soit attesté. Pour sa part, Furetière se plaît à émailler la description lexicographique de citations littéraires et, en dépit de ses choix arbitraires, il annonce à sa façon Littré et son goût pour l'énoncé littéraire"²³. Dans la préface de la deuxième édition du *GR*, Alain Rey est lucide sur la dimension novatrice du *GR* par rapport au *TLF*:

Entre le gigantisme 'philologique' du Trésor de la langue française — qui ne porte que sur le français de 1790 à 1960 —, et l'information trop partielle ou rendue archaïque des autres dictionnaires de langue, nous avons visé la richesse d'information dans l'économie de présentation, la modernité sur fond historique, la simplicité de l'exposé pour maîtriser la complexité des faits décrits; en un mot, la communication d'une image sociale: celle de la culture francophone classique et contemporaine à travers le kaléidoscope des mots²⁴.

²³ Laurent Bray, 1995: p. 52.

²⁴ *Préfaces*, 2005: p. XX.

Chronologiquement c'est au *GR* que nous devons d'avoir comblé le "vide lexicographique" de la première moitié du XX^e siècle que Rey a de maintes fois dénoncé. Il couvre une période historique qui va au-delà du projet initial du *TLF* (1950-1960) et donne lieu à des produits dérivés (*Petit Robert*, *Robert Junior*, etc.) en plus des ajouts et remaniements annuels. Le *TLF* ne possède qu'une édition unique (sauf un supplément) qui le rend plutôt similaire aux dictionnaires anciens lesquels établissaient plutôt une tradition de réimpression valorisant les amplifications. Les remaniements à chaque millésime ne pouvaient s'en tenir, vu le marché dictionnaire et les obstacles de la technique lexicographique d'alors, qu'à de rares rééditions (parfois très éloignées dans le temps) ou des suppléments. Aussi les citations du *TLF* ne commencent qu'avec la Révolution tandis que les citations du *GR* couvrent plus de cinq siècles.

La confection du *GR* semble aussi avoir été mue par un investissement très différent si l'on considère le fait que la description des premiers tomes est beaucoup plus sommaire par rapport aux autres et devient de plus en plus extensive alors que le *TLF*, au contraire, est plus développé dans les premières lettres que dans la suite. Plus de vingt ans séparent l'édition du premier tome du *TLF* et son dernier tome, contrairement au *GR* qui est achevé en moins de quinze ans même, s'il est vrai, ses articles sont beaucoup plus réduits. Cela donne au *GR* un aspect commercial que le *TLF*, dictionnaire avant tout destiné à l'élite. A cet égard, le *GR* est tributaire des recherches et des "excellentes synthèses procurées par les articles diachroniques du *TLF*" selon Alain Rey dans la préface de la deuxième édition²⁵.

FICHE TECHNIQUE TLF1

Configuration matérielle requise: Processeur Intel Pentium II 350 MégaHertz 64 Mo Ram, 450 Mo d'espace disque disponible Lecteur de cédérom. Carte graphique 256 couleurs. Résolution d'écran de 800 par 600. Configuration logicielle requise: Microsoft Windows 98-Me NT (SP6)=2000-XP Navigateur Internet (non fourni): Internet explorer version 5 ou ultérieure Mozilla version 1.6 ou ultérieure Netscape version 7.1 ou ultérieure. Machine virtuelle JAVA de Sun Microsystems (fournie sur ce cédérom).

FICHE TECHNIQUE DU GRE

Configuration requise: Processeur Intel Pentium II ou supérieur, Méga Hertz 32 Mo Ram, 15 Mo de disque dur disponible pour l'installation minimum, 250 Mo pour l'installation complète. Microsoft Windows 98/2000/Me/NT4 ou XP. Internet explorer version 5. Disponible en version réseau (le cédérom est installé sur un serveur), en licence sur site (permet d'équiper des postes nomades) et pour certains titres en version HTML (permet l'accès simultané de plusieurs utilisateurs par Intranet).

²⁵ *Op. Cit.*, p. XXXVII.

BIBLIOGRAPHIE

- BRAY, Laurent, "Notes sur la référence bibliographique", *Lexique*, 12-13, PUS, 1995, p. 91-103.
- CAPUTO, Hélène, "Norme(s), Langue(s) et Dictionnaires Monolingues", *Cahier du C.I.E.L., Facteurs d'hétérogénéité dans le lexique*, Paris, Centre Interlangue d'Etudes en Lexicologie, Université Paris 7, pp. 73-92.
- CORBIN, Pierre, "Le monde étranger des dictionnaires (8). Du Petit Robert (1967) au Micro Robert (1971): le recyclage de citations", *Lexique*, n° 12-13, 1995, pp. 125-144.
- DUVAL, Alain, "Clefs d'exploitation pour les citations du *Robert Electronique*", *Lexique*, n° 12-13, 1995, pp. 469-477.
- HENRY, Françoise, "Les écrivains et le dictionnaire, d'après le corpus du Trésor de la langue française", *Lexique*, 12/13, 1995, 389-402.
- JACQUET-PFAU, Christine, "Dictionnaires et encyclopédies du français: la version électronique", dans *Proceedings of the Eleventh Euralex International Congress Euralex 2004*, Lorient, France, July 6-10 2004, vol. 1, Université de Bretagne Sud, pp. 53-62.
- LEHMANN, Alise, "La citation d'auteurs dans les dictionnaire de la fin du XVIIIe siècle (Richelet et Furetière)", *Langue française*, n° 106, mai 1995, p. 35-54.
- "Du Grand Robert au Petit Robert: les manipulations de la citation littéraire", *Lexique*, n° 12-13, 1995, pp. 105-124.
- MARTIN, Robert, "Le Trésor de la langue française", dans Gérard Antoine et Bernard Cerquiglini (sous la dir. de), *Histoire de la langue française*, Paris, CNRS éditions, 2000, pp. 969-977.
- MORISSEY, Robert, "Text and Contexts: The ARTFL Database in French Studies", *Profession 93, The Modern Language Association of America*, Phyllis Franklin (ed), 1993, pp. 27-31.
- NORA, Pierre, "Le dictionnaire de pédagogie de Ferdinand Buisson, cathédrale de l'école primaire", *Les lieux de mémoire. 1, La République*, 1984, pp. 354-378.
- PIERREL, Jean-Marie, Jacques Dendien, Pascale Bernard, "Le TLFi ou Trésor de la langue française informatisé", dans Geoffrey Williams et Sandra Vessier (eds.), *Proceedings of the Eleventh Euralex International Congress Euralex 2004*, Lorient, France 6-10, 2004, vol. 1, Université de Bretagne-Sud, pp. 165-170.
- PIOTROWSKI, David, *L'hypertextualité ou la pratique formelle du sens*, Paris, Honoré Champion, 2004, 226 p.
- Préfaces, extrait des préfaces de la dernière édition*, Paris, Dictionnaire Le Robert, 2005, LVIII p. [document joint au *Grand Robert électronique*].
- PRUVOST, Jean, *Dictionnaires et nouvelles technologies*, Paris, Presses universitaires de France, 2000, (coll. "Ecritures électroniques").
- REY, Alain, *Le lexique, images et modèles, du dictionnaire à la lexicologie*, Paris, Librairie Armand Colin, 1977, 307 p.
- ROBERT, Paul, *Aventures et mésaventures d'un dictionnaire*, Paris, Société du Nouveau Littre, le Robert, 1966, [149] p.