

Un Thesaurus informatisé de bibliographie gréco-latine

Rodrigue LARUE, Pierre SENAY et Florent TREMBLAY

Abstract. 1985: The Université du Québec à Trois-Rivières published the first edition of a standardized directory of name authority forms for Greek and Latin authors: *Clavis Scriptorum Graecorum et Latinorum*: It was initiated by Father Rodrigue LaRue, OFM, and included over 20,000 bibliographic records. It assembled in a single alphabetical index the authors, a Thesaurus of descriptors, the titles, and main subjects of the Classics.

Thesaurus Bibliographiae Graecae et Latinae (TBGL), Present and Future:

1994: The first CD-ROM in the *TBGL* series is published by the Université du Québec à Trois-Rivières, it marks an important event for researchers specializing in classical studies.

TBGL volume I contains all 9,000 bibliographic records for the letter 'A', as well as 33,000 digitized stills of original source material. At least ten more CD volumes are to appear over the next few years, completing the alphabet, and creating a new and definitive reference work.

The electronic *TBGL* comprises: the updated edition of the *Clavis* and the *Dossiers de bibliographie grecque et latine*.

- 1) The *Clavis-II*: over 70,000 bibliographic records; listings for authors, subjects, and disciplines from antiquity to the present; all preserved Greek and Latin writings from the tenth century B.C. to the present: 30 centuries of history; several million possible references; all subjects and areas of classical studies.
- 2) The *Dossiers*: 500,000 digitized stills of the original sources. Among the sources are: *L'Année philologique*, Bursian's *Jahresbericht Über Die Fortschritte der Klassischen Alterthumswissenschaft*, Fabricius's *Bibliotheca graeca et Bibliotheca latina*, and Wadding's *Scriptores Ordinis minorum*, etc.

TBGL, a User-Friendly Reference Tool:

- Easy to consult for anyone, from novice to expert.
- Search key based on familiar terms used in classical studies—ancient, medieval, as well as modern—and religious studies (subject, name authority forms, language, type of record,

✉ Rodrigue LARUE et Pierre SENAY; Études anciennes; C.P. 500; Université de Québec à Trois-Rivières; Trois-Rivières; Québec G9A 5H7 (Canada).

Fax : +1 819 376 5144

✉ Florent TREMBLAY, Collège militaire royal de Saint-Jean; Département de linguistique historique; Richelain; Québec J0J 1R0 (Canada).

Fax : +1 514 358 6799

century, descriptor, key word, bibliographical notes, principal works, list of still sources, publication years, bibliographies...).

- Records and *stills* connected by hypertext-type links, ensuring rapid and logical access to the data.
- Produced by specialists in classical studies, in documentation, and in computer science at the Université du Québec à Trois-Rivières.

Keywords: Bibliography, classical studies, **Mots-clés :** Bibliographie, études classiques, CD-ROM.

La réalisation de ce Thesaurus¹

Introduction

Il y a vingt ans, l'idée d'une banque bibliographique en études anciennes qui serait internationale, exhaustive et à la portée de tous les chercheurs, paraissait parfaitement utopique.

Progressivement, des chercheurs universitaires de tout le Québec, des membres de la Société des études anciennes du Québec, guidés dans leurs démarches par l'infatigable Professeur Rodrigue LaRue de l'Université du Québec à Trois-Rivières, (UQTR), ont dégagé une vision originale de la tâche qui s'imposait : connaissance générale d'un immense fonds documentaire et de son exploitation par des solutions informatisées définitives pour un *Thesaurus Bibliographiae* exhaustif et d'accès facile.

1. Les documents de base

Ce qui existe présentement, entreposé au sous-sol de la Bibliothèque de l'UQTR, peu exploité par les professeurs-chercheurs qui ignorent à peu près tout de ces fiches bibliographiques, est constitué d'une masse colossale de données accumulées sur une période de plus de 40 ans. C'est, *grosso modo*, un millier de boîtes d'archives contenant chacune quelque 3 000 fiches, ce

¹ Cet article fait suite à deux autres textes déjà parus dans *RISSH* : TREMBLAY (Florent) et LARUE (Rodrigue) : 1991, «Thesaurus bibliographiae græcæ et latinæ. An experiment in Computerizing Data with a Scanner and Storing them in CD-ROM'S», *RISSH* 27, pp. 213-231; TREMBLAY (Florent) : 1993, Thesaurus bibliographiae græcæ et latinæ. An experiment in Computerizing Data with a Scanner and Storing them in CD-ROM'S, Part 2 : Updating of a Bibliographical Material and Conclusion on the Technological Experimentation», *RISSH* 29, pp. 191-206.

qui, par simple comptabilité, signifie plus de 3 000 000 de pièces d'information bibliographique. Il y a les auteurs et il y a les disciplines.

Les auteurs, il y en a 74 000 classés par ordre alphabétique et que nous entendons recenser sous le titre de *Clavis Scriptorum Graecorum et Latinorum* (la clé des auteurs grecs et latins). Cette *Clavis-II* (mise à jour de la *Clavis-I* de 20 000 auteurs qui fut publiée en 1985) deviendra le guide normatif des noms de personnes pour l'ensemble du *Thesaurus*. Quant aux disciplines, elles reprennent les dix divisions déjà officialisées par *L'Année philologique*.

L'inventaire de ces « dossiers » est relativement simple : la matière couvre trente (30) siècles. Sa variété et son immensité touchent à tous les domaines de la civilisation gréco-latine : Les dossiers des études grecques s'étendent de la préhistoire jusqu'à la prise de Constantinople (1453), mais les travaux, en cette matière, viennent jusqu'à nos jours. Les dossiers des études latines et les travaux qui s'y rattachent, dans toutes les langues européennes, s'étendent également de la préhistoire à nos jours.

Le nombre des ouvrages bibliographiques dépouillés pour constituer ces *Dossiers* est considérable (plus de 300 d'après une liste sommaire dressée par le prof. LaRue de l'UQTR). Il a fallu fixer un choix inspiré, à la base, par l'*Instrumentum* de J. Marouzeau (*Dix années*, p. VII) et par la *General Bibliography* du *Oxford Classical Dictionary* (2^e éd. 1970, p. 1151 ss.) pour les auteurs classiques.

Ce dépôt rassemble une masse imposante de quelque 25 000 chemises cartonnées qu'il faut absolument confier à la mémoire de l'informatique si on ne veut pas qu'elles soient bientôt perdues.

2. Les objectifs de travail

L'objectif principal de ce projet est d'informatiser cette masse bien précieuse de références bibliographiques et de les emmagasiner sur *CD-ROM*, en base de données indexées afin de les rendre disponibles aux autres chercheurs.

Les objectifs intermédiaires que nous avons prévus sont :

- a) Compléter au « scanner » la compilation statistique des 3 500 000 fiches bibliographiques (anciennes, médiévales et modernes; grecques, byzantines et latines; romanes, germaniques et anglo-saxonnes).
- b) Procéder à l'analyse et à la classification de tous ces dossiers par l'ajout de mots-clés.
- c) Organiser cette base de données selon une méthodologie rigoureusement définie et expliquée ci-après.

- d) Rendre cet *outil de recherche* accessible à tous les chercheurs grâce au disque compact vidéo (*CD-ROM* video).

3. La méthodologie suivie

Pour atteindre l'objectif établi, qui est d'informatiser par balayage optique et emmagasiner sur *CD-ROM*, la démarche suivante s'est imposée :

- a) La saisie des données se fait de façon « mécanique » en utilisant un procédé photographique par balayage optique. Chaque feuillet doit alors être « lu » par le *scanner*, nettoyé de ses impuretés et encodé selon les paramètres que nous avons établis.
- b) Une indexation de ces données est le procédé que nous choisissons afin de les rendre accessibles aux chercheurs : indexation des numéros de dossiers / auteur(s) / date(s), siècle, période / source du document concerné / descripteurs (mots-clés). Chacune de ces zones énumérées, (*i.e.* date, auteur, etc.) devient une porte d'accès à l'information. Certains ouvrages n'ont pas d'auteurs, certains ne sont pas datés; ils seront repérés soit par le titre, soit par d'autres mots-clés, car pour faciliter la consultation et pour obtenir un repérage plus étendu, on a multiplié les clés d'accès. L'ajout des mots-clés et le recours à la logique booléenne faciliteront la recherche des éléments et des sujets désirés.

En résumé, cette approche méthodologique nous mène aux étapes suivantes :

- i) la lecture optique (*scanning*) nous donne une image en mode point (*bitmap*);
- ii) ces images sont transformées en caractères ASCII;
- iii) l'image statique obtenue est alors compressée;
- iv) des mots-clés sont encodés; et enfin,
- v) l'information est emmagasinée sur *CD-ROM* (mémoire d'archive).

L'équipe du *TBGL* a opté pour l'option *CD-ROM* parce que consulter les données bibliographiques avec cette technologie ne génère pas de perte de temps comme dans le cas des services *on-line*. Mener une recherche à l'aide de mots-clés est chose facile. Bien plus, l'énorme quantité de données emmagasinées peut être transmise à distance rapidement et être imprimée partout. Le *CD-ROM* offre un outil de recherche unique parce qu'il n'y a actuellement sur le marché aucun produit tant soit peu comparable par sa capacité et par sa rapidité. À l'aide de cet outil, le chercheur pourra

constituer son propre fichier bibliographique et l'emmagasiner sur sa propre disquette. Au surplus, l'équipement et la configuration technique sont très simples d'utilisation; ils permettent une utilisation en *Windows* (PC 386, 2-4 Mo de mémoire vive, souris et lecteur CD) et en *Macintosh* (tout modèle équipé du système 6.07 ou mieux avec un minimum de 800 ko de mémoire vive).

Pour mener à terme ce projet, l'équipe a décidé de confier à la firme *Société nationale d'information Inc.*, filiale de *Periodica*, les étapes suivantes : la préparation et l'indexation des données à partir des fichiers de type .DBF et .DBT dont l'alimentation est assurée par une application développée avec la version 5.0 de *Clipper*; l'importation des images numérisées; l'élaboration d'un index à partir des informations contenues dans le nom attribué à chacune des images numérisées; le développement d'un logiciel de navigation convivial en langues française et anglaise dans l'environnement MS-DOS ou *Windows* et *Macintosh*, pour interroger la banque de données. Cet outil de navigation, en plus du repérage, permettra l'impression sur disquette des résultats de recherche; la fourniture d'une documentation technique en langues française et anglaise, le tout sous la forme de fichiers produits par une application du traitement de texte; le pré-matriçage; le pressage du *CD-ROM* et sa mise en marché.

L'équipe s'engage pour sa part à fournir sur disquettes, toutes les données des *Dossiers*, comprimées sous la forme de fichiers de type .DBF et .DBT et à fournir les images numérisées (images statiques des textes obtenues à partir de la lecture optique). À chaque image sera attribué un nom composé de quatre parties distinctes afin de permettre l'indexation et le repérage. Ce nom contiendra : le numéro de référence dans la *Clavis*; le code de la source bibliographique; l'année de la source et le numéro séquentiel; la liste des codes des sources bibliographiques avec leurs titres et les références.

La *Société nationale d'information (SNI-Montréal)* assure la commercialisation et la gestion des ventes.

4. Calendrier : étapes des opérations

Au printemps de 1994 fut achevée l'opération de numérisation (la lecture optique), de la lettre « A » qui comprend quelque 9 000 auteurs anciens et 33 000 « dossiers » des disciplines.

Vinrent ensuite l'indexation des textes, l'ajout des mots-clés et le matriçage sur disque compact.

En août 1994 eut lieu le lancement de cette première tranche (la lettre «A» au complet) lors du Congrès de la FIEC (Fédération Internationale des Études Classiques) qui, pour la première fois de son histoire, se tint en Amérique du Nord, à l'Université Laval, ville de Québec.

À partir de l'automne 1994, et dans les mois et les années qui suivront, nous espérons assurer la préparation du matériel restant, à raison de trois mois de travail d'indexation en moyenne par lettre alphabétique. Cela comprendra tous les auteurs, les matières et les disciplines. Resteront encore à établir les *CD-ROM* des commentaires et recensions de 30 siècles d'histoire!

Au rythme actuel, nos calculs nous mènent à l'année 2001. L'important, c'est de commencer et de poursuivre si on veut terminer!

Conclusion

Quand il sera terminé, le *Thesaurus* représentera le fruit de toutes les données que le Père LaRue a accumulées en plus de 40 ans de diligentes compilations. Ce sera un *outil de recherche* unique en son genre, éminemment utile et non-existant. C'est un projet technologique ambitieux, partiellement réalisé parce que toutes les fiches sont déjà classées. Reste à les informatiser! Le *TBGL* contribuera d'une façon très significative à mettre à la portée des chercheurs les données qui permettront d'entreprendre de nombreux autres projets de recherche. Lorsqu'il sera terminé, ce *Thesaurus* informatisé et disponible sur *CD-ROM* deviendra, à n'en pas douter, un outil essentiel à tous les «historiens» de la grande culture gréco-romaine; son importance pour l'étude des textes grecs, latins et byzantins sera incontestable.

En résumé

1985 : L'Université du Québec à Trois-Rivières publie la première édition d'un répertoire de normalisation des formes de noms d'auteurs grecs et latins : la *Clavis Scriptorum Graecorum et Latinorum* : Fruit de la vision du Père Rodrigue LaRue, OFM, elle regroupait plus de 20 000 notices. Sous un seul ordre alphabétique, on y retrouvait des auteurs, un thésaurus de descripteurs, des titres d'œuvres, des sujets.

Thesaurus Bibliographiae Graecae et Latinae (TBGL), aujourd'hui et demain :

1994 : la publication du premier disque optique numérique (*CD-ROM*) de la série *TBGL*, par l'Université du Québec à Trois-Rivières, marque une date importante pour les chercheurs spécialisés en études anciennes.

Le volume I du *TBGL* comprend les 9 000 notices d'auteurs débutant par la lettre «A», ainsi que 33 000 *photographies* numérisées. Au cours des prochaines années,

une dizaine de nouveaux disques seront édités, par lettre ou groupe de lettres. Et un nouveau corpus virtuellement exhaustif verra le jour.

Le *TBGL* électronique comprend deux types de contenus : la *Clavis* enrichie et les *Dossiers de bibliographie grecque et latine*.

- 1) La *Clavis-II* : c'est plus de 70 000 notices comprenant l'ensemble des auteurs, matières et disciplines, depuis l'Antiquité jusqu'à nos jours; c'est tout ce qui s'est écrit, en grec ou en latin, entre le dixième siècle avant Jésus-Christ et aujourd'hui : 30 siècles d'histoire; c'est un potentiel de quelques millions de références.
- 2) Les *Dossiers* : 500 000 *photographies* numérisées des sources elles-mêmes, parmi lesquelles : *L'Année philologique*, *Jahresbericht Über Die Fortschritte der Klassischen Altertumswissenschaft* de Bursian, *Bibliotheca graeca et Bibliotheca latina* de Fabricius, *Scriptores Ordinis minorum...* de Wadding, etc.

TBGL, un produit convivial :

- Accès intuitif pour l'expert comme pour le novice.
- Clés de recherche qui s'inspirent des termes familiers employés en études classiques, médiévales et modernes, ainsi qu'en études religieuses (sujets, formes de nom d'auteur, langue, type de notice, siècle, descripteurs, mots-clés, notes bibliographiques, œuvres caractéristiques, sources bibliographiques, année d'édition, bibliographie...).
- Notices et *photographies* reliées entre elles par des liens de type hypertexte assurant un accès rapide et logique à l'information.
- Une réalisation des spécialistes en études anciennes, en documentation et en information de l'Université du Québec à Trois-Rivières.

Bibliographie

Cedrom Technologies inc. *Création d'une nouvelle entreprise d'information électronique au Québec*. Communiqué de presse, *PERIODICA*, 1^{er} octobre 1992, 7 pages.

GOULET : 1989, *Dictionnaire des Philosophes antiques*, pp. 18-23.

LARUE (Rodrigue) : 1992, «Thesaurus international de bibliographie gréco-latine : Problèmes et solutions», *Cahier des Études anciennes*, n° 27, pp. 133-178.

Pontifical Institute of Mediaeval Studies : 1989, *The Greek Index Project Series*, Toronto, p. 9.

SENAY (P.) et TREMBLAY (F.) : 1992, Eds. «Séminaire international sur la faisabilité du *Thesaurus*, 2-3 oct. 92, à l'UQTR», *Cahier des Études anciennes* XXVII, 176 pages. (Douze articles par autant d'auteurs sur la faisabilité de ce même projet).

Société nationale d'information inc. 1993, *Repérage de toute information sur CD-ROM grâce à 30 clés de recherche*. (Outremont, Montréal : SNI).