



HAL
open science

Thesaulangue

Viviane Berthelie, Isabelle Turcan

► **To cite this version:**

| Viviane Berthelie, Isabelle Turcan. Thesaulangue. 2006. halshs-00096877

HAL Id: halshs-00096877

<https://shs.hal.science/halshs-00096877>

Submitted on 25 Sep 2006

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.

NOTES ET COMPTES RENDUS

THESAULANGUE

Le thesaurus de linguistique française du laboratoire ATILF

Thesaulangue est un recueil qui se veut exhaustif de tous les termes de linguistique française figurant dans la littérature scientifique internationale (revues et monographies en particulier), analysée par les documentalistes du laboratoire INaLF (Institut National de la Langue Française), sous la responsabilité d'Evelyne Martin, à partir de principes définis en 1971¹, et repris par ceux du laboratoire ATILF (Analyse et traitement informatique de la langue française, ATILF-CNRS-Universités de Nancy) qui lui a succédé en 2001. Sa fonction principale est de servir à exprimer toutes les notions en usage dans ce domaine.

Ce thesaurus de linguistique française, nommé *Thesaulangue* dans la dernière phase de son élaboration, a été repris à partir du thesaurus existant à usage interne, dans l'état où il se trouvait en 2000 : une liste de termes collectés pendant deux décennies à l'INaLF, au gré de la réalisation des travaux scientifiques du laboratoire et de l'entrée des ouvrages à la bibliothèque. Initialement conçu pour être une des composantes de la *Banque de données Lexicologiques et lexicographiques*, indissociable du *Trésor de la Langue Française*, le thesaurus de linguistique française a d'emblée visé à la maîtrise de l'exploitation automatisée de vastes ressources documentaires, par le recensement systématique des mots-clés métalinguistiques utilisés pour la description linguistique en langage naturel.

Thesaulangue est avant tout un outil documentaire pour une indexation scientifique, une homogénéisation de l'indexation des notions linguistiques sur le plan terminologique, destiné à faciliter toutes les recherches bibliographiques et favoriser la pertinence des résultats d'une requête automatisée sur la base bibliographique du centre de documentation et à terme sur HAL (hyper article en ligne) qui fournit une interface permettant aux auteurs de déposer sur la base du CCSD (Centre pour la Communication Scientifique Directe –UPS 2275) des manuscrits d'articles scientifiques dans toutes les disciplines.

C'est un outil dynamique, ouvert à l'enrichissement progressif de son contenu en fonction de l'émergence de nouvelles notions ou termes scientifiques. De ce fait il offre un suivi historique retraçant l'évolution de la linguistique, de ses concepts et de sa terminologie au travers des différentes publications scientifiques prises en compte.

La dynamique de ce projet a permis de poursuivre la valorisation des ressources du laboratoire et leur exploitation bibliographique automatisée.

Ont été effectués :

¹ Martin, R., Vienney, R., Martin, E. (1971) : La Mécanisation de la documentation bibliographique au Centre de recherche pour un Trésor de la langue française. *Le français moderne*, 39-2, 181-190.
Martin, E. (1975) : le thesaurus de linguistique française. *Le français moderne*, fasc. 2, supplément.

- Un toilettage terminologique des termes obsolètes ou leur remplacement par un synonyme dont la fréquence d'apparition dans les textes est significative ;
- Une actualisation terminologique correspondant aux nouvelles notions propres à chacun des objets de recherche scientifique en linguistique ;
- Un enrichissement du corpus au prorata des travaux effectués depuis 2000 et dont l'analyse scientifique a été réalisée – et continuera de l'être- au Centre de documentation
- Une harmonisation des descripteurs tant sur le plan des entrées préférentielles que celui des renvois et des « employés pour » ;
- Une traduction en langage RAMEAU, dans la mesure du possible : la passerelle ainsi réalisée facilite les échanges de notices avec le Sudoc (Système Universitaire de documentation), base nationale gérée par l'ABES (agence bibliographique de l'enseignement supérieur) ;
- Une consultation et une relecture d'experts en linguistique et en terminologie afin de confirmer la pertinence et la validité scientifique du thesaurus.

Sont actuellement en projet :

- Pour répondre à l'attente de la communauté linguistique internationale : une traduction en anglais et en allemand ;
- Pour une mutualisation des ressources terminologiques à l'échelle nationale : un reformatage en XLM de la version actuelle en Word, en vue de son intégration dans *TermSciences*. Ce portail terminologique, initié par la Direction de l'Information Scientifique du CNRS et développé par l'INIST en association avec le LORIA et l'ATILF a pour but de valoriser et de mutualiser l'ensemble des ressources terminologiques (lexiques, dictionnaires, thesaurus) des organismes publics de recherche et d'enseignement supérieur pour aboutir à la constitution d'un référentiel terminologique commun ;
- Pour donner une valeur ajoutée aux notices bibliographiques ATILF qui enrichissent la base documentaire du SCD de l'Université de Nancy 2 : une intégration de *Thesaulangue* au logiciel de gestion documentaire afin de créer des contrôles automatiques des termes d'indexation et de générer une indexation automatique en langage RAMEAU, adopté par le Sudoc.

La réactualisation du thesaurus de l'ATILF a permis aux documentalistes du Centre de documentation de poursuivre l'indexation et le dépouillement de la littérature scientifique et en particulier des articles parus dans les périodiques auxquels le Centre de documentation est abonné.

Nous pouvons ainsi enrichir notre base bibliographique d'une valeur ajoutée reconnue par la communauté scientifique comme indispensable pour réunir avec la plus grande pertinence une bibliographie sur une notion donnée alors que cette même notion n'est pas forcément mentionnée dans le titre des articles et ouvrages.

Ainsi *Thesaulangue* propose aux chercheurs une arborescence de descripteurs pour l'indexation de la production scientifique du laboratoire ATILF, dans le cadre de son dépôt dans la base d'évaluation du CNRS (HAL), à travers *TermSciences* auquel il sera bientôt intégré.

Thesaulangue représente désormais le résultat d'un travail qui s'imposait véritablement pour un laboratoire tel que l'ATILF dont la mission est à la fois ancrée dans l'histoire de la langue

française enregistrée sur le plan documentaire, et ouverte à sa régénération par les renouvellements de sa matière vivante en tant que langue naturelle, un laboratoire dont la mission lexicographique, descriptive et analytique, ne saurait se passer ni d'une réflexion terminologique ni d'outils novateurs assortis en matière de transmission des savoirs et des savoirs-faire.

Mais surtout, et par delà l'utilité première d'un inventaire normalisé, le Centre de documentation, répondant à la vocation d'observatoire de la langue du laboratoire, a créé pour la future histoire du métalangage linguistique, un outil et une nouvelle ressource documentaire. *Thesaulangue* doit être compris par la communauté linguistique comme une réserve terminologique de l'histoire de démarches épistémologiques, comme le conservatoire de données précieuses accumulées au gré de l'évolution de la recherche, comme un lieu de mémoire vivante offert aux futurs chercheurs.

Viviane Bertheliet et Isabelle Turcan.
Université de Nancy 2 / ATILF.