

## Apports et usages de l'informatisation de connaissances sur le lexique français : l'exemple du TLFi (Trésor de la Langue Française Informatisée)

Jean-Marie Pierrel, ATILF UMR Nancy Université & CNRS  
[Jean-Marie.Pierrel@atilf.fr](mailto:Jean-Marie.Pierrel@atilf.fr)

Au cours de la seconde moitié du XXème siècle, de nombreuses contributions importantes à la lexicographie française se sont développées. A travers le Centre de Recherche pour un Trésor de la Langue Française (CRTL) puis de l'Institut National de la Langue Française (INaLF) dont notre laboratoire ATILF ([www.atilf.fr](http://www.atilf.fr)), Analyse et Traitement Informatique de la Langue Française, se veut être aujourd'hui le digne successeur, Nancy a su prendre une place privilégiée en ce domaine.

Initiées au départ autour du projet de Trésor de la Langue française (TLF : Imbs, Quamada 1971-1994) dont nous rappellerons les grandes lignes, les études nancéiennes en lexicographie française se sont poursuivies, au-delà de la rédaction du TLF, suivant deux orientations complémentaires : la lexicographie historique et la valorisation informatique des ressources produites dont un des premiers exemples fut la version informatique du TLF (Atilf 2004-2005), aujourd'hui disponible sur le Web ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)).

Cette seconde orientation a provoqué au sein du champ des études lexicales une véritable révolution qui fit de l'informatique un outil indispensable pour :

- étudier le lexique et ses propriétés à travers l'exploitation intelligente de textes et de documents ;
- structurer et normaliser les connaissances lexicales et lexicographiques ;
- valoriser, partager et mutualiser les résultats de la recherche sur le lexique de notre langue qui trop souvent encore reste dispersés.

Le partage et la mutualisation de résultats de recherche et de ressources informatisées sur le lexique français ouvrent en effet des perspectives intéressantes.

Ainsi le TLF, une référence en lexicographie française, a eu pendant longtemps la réputation tenace d'être un dictionnaire réservé à une élite et la diffusion de sa version papier s'est limitée à quelques milliers d'exemplaires au sein d'une intelligentsia somme toute limitée. Sa version informatique sous forme de Cédérom et de ressources librement accessibles sur le Web a rencontré un succès important tant auprès du grand public que des utilisateurs universitaires ou des professionnels de la langue. Sa version Web fait l'objet de plusieurs dizaines de milliers de connexions quotidiennes en provenance de tous les continents, et il est référencé par d'innombrables sources. La notoriété qu'il a acquise en fait aujourd'hui un dictionnaire incontournable et un outil de valorisation de la langue française et comme support pour l'enseignement du français.

L'intégration plus récente au sein du portail lexical du Centre National de Ressources Textuelles et Lexicales (CNRTL : [www.cnrtl.fr](http://www.cnrtl.fr)) de diverses ressources sur le lexique français permet encore une meilleure mutualisation des résultats de la recherche. L'outil informatique, à travers le Web, offre aujourd'hui des modes nouveaux de valorisation de ressources ou de résultats de recherche. Au-delà du seul monde universitaire, ces techniques permettent de mettre à disposition de l'ensemble de la société nos résultats de recherche. Aujourd'hui le portail lexical du CNRTL fait l'objet d'environ 300 000 requêtes par jour provenant d'horizons très divers (cf. : [www.cnrtl.fr/aide/stat/](http://www.cnrtl.fr/aide/stat/)), il est aujourd'hui l'un des sites Web sur le lexique français les plus utilisés.

Notons enfin que la généralisation de telles exploitations et valorisations de versions électroniques est ainsi en train de modifier notablement les modes de travail et d'échanges scientifiques au sein des communautés de recherche en sciences humaines, permettant un véritable travail en réseau comme c'est le cas en particulier aujourd'hui pour la communauté en lexicographie.

Après avoir présenté les principales conditions facilitant une meilleure mutualisation des résultats de recherche et le rôle important de l'informatique et du Web dans ce cadre, nous montrerons un exemple d'une telle valorisation de ressources lexicographiques à travers la version informatisée du TLF ([www.atilf.fr/tlfi](http://www.atilf.fr/tlfi)) et son intégration au sein du portail lexical du CNRTL ([www.cnrtl.fr](http://www.cnrtl.fr)), abrité au sein

de notre laboratoire et qui regroupe, dans le domaine de la lexicographie et des études lexicales, des contributions d'horizons divers :

- fonds de ressources issues de notre laboratoire ATILF : *Trésor de la Langue Française Informatisée*, *Dictionnaire du Moyen Français ou DMF*, développements récents en lexicographie historique avec en particulier le programme *TLF-Etym*, etc. ;
- mais aussi contributions de nombreux partenaires : Académie Française ([www.academie-francaise.fr/](http://www.academie-francaise.fr/)) à travers l'informatisation des principales éditions de ce dictionnaire de la 1ère datant de 1694 à la 9ème édition en cours de finition ; ARTLF de Chicago (<http://humanities.uchicago.edu/orgs/ARTFL/>) pour les principaux dictionnaires informatisés présent dans notre rayon des dictionnaires anciens ; l'université de l'Etat de São Paulo – *UNESP* à travers un dictionnaire électronique d'expressions idiomatiques - français-portugais / portugais-français ; l'université Laval de Québec, en association avec de nombreux contributeurs du monde francophone, à travers la base de données lexicographiques panfrancophones (<http://www.tlfq.ulaval.ca/bdlp/>) ; Le laboratoire CRISCO de Caen à travers son dictionnaire des synonymes ([www.crisco.unicaen.fr](http://www.crisco.unicaen.fr)), le laboratoire CLEE-ERSS de Toulouse à travers la présentation de la *proxémie* entre les mots (<http://erss.irit.fr/>) ; etc.

Cette présentation sera l'occasion d'une réflexion sur les possibilités, d'une part, d'usage de telles ressources dans le cadre de l'enseignement et de la recherche et, d'autre part, de coopération et de mutualisation impliquant la communauté des enseignants-chercheurs et doctorants de linguistique française qui, depuis de nombreuses années, mènent des études et recherches sur le lexique français sans que l'ensemble de la communauté universitaire en soit toujours informée et qui pourraient trouver, au sein du CNRTL, un vecteur de valorisation permettant une plus grande visibilité de leurs résultats de recherche.

### Références bibliographiques

- ATILF (2004-2005) *Trésor de la Langue Française informatisé*, CNRS Editions, 591 p. et CD du texte intégral, Version PC, ISBN 2-271-06273-X, 2004, Version Mac OS X, ISBN 2-271-06365-5, 2005.
- Imbs P., Quemada B. (dir.). *Trésor de la Langue Française. Dictionnaire de la langue du XIXe et du XXe siècle (1789–1960)*, 16 vol., Paris, Éditions du CNRS/Gallimard, 1971–1994.
- Dendien J., Pierrel J.M. (2003) Le Trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence, *Traitement Automatique des Langues*, Vol 44 – n° 2/2003, Hermes Sciences Edition, p. 11-37
- Pierrel, J.M. et Petitjean, E. (2007) Valorisation et exploitation scientifiques de documents numériques pour la recherche en linguistique : l'exemple du CNRTL, *Actes de CIDE 2007 Congrès International sur le Document Numérique*, Nancy, 2-4 juillet 2007, p13-24, Europa, ISBN 978-2-909285-38-2
- Pierrel, J.M. et Buchi, E. (2009) Research and Resource Enhancement in French Lexicography: the ATILF Laboratory's computerized resources, in "Lexicography in Italy and in Europe", Silvia Bruti, Roberta Cella and Marina Foschi, editors, Cambridge Scholars Publishing, pp. 79-118