

Appel à contributions pour un projet international



中外辞书数字图书馆

Digital Bibliography and Library of Lexicography: Chinese-Foreign Languages

Bibliographie et Librairie Digitales de la Lexicographie : chinois-langues étrangères

À la suite du projet international « *Digital Lexicography in China and Europe: Status Quo and Future Directions* » et du 9ème *Colloque International PL<L*, nous lançons un projet visant à collecter et intégrer une bibliographie et un corpus sur la (méta)lexicographie numérique chinois-langues étrangères : Bibliographie & Librairie Digitales de la Lexicographie.

Calendrier :

- **Lancement du projet** : Juin 2024
- **Financement** : Demande en cours
- **Appel à contribution des experts lexicographes de différentes langues** : 30 août 2024

Description du projet

Contexte

Bien que les experts en (méta)lexicographie aient depuis longtemps entrepris des projets visant à collecter des références sur la recherche sur les dictionnaires, tant en Chine qu'à l'étranger, ces efforts sont souvent marqués par plusieurs limitations. Qu'il s'agisse de ressources imprimées ou en ligne, les tentatives de recensement et d'enregistrement de la production scientifique dans ce domaine sont souvent incomplètes, obsolètes (plus mises à jour), ou se concentrent uniquement sur des données bibliographiques. La littérature scientifique est généralement dispersée, publiée dans divers revues et actes par différentes associations, avec des niveaux d'accessibilité variés. Pour compliquer les choses, de nombreuses publications anciennes ne sont disponibles qu'en version imprimée (*An Index of Chinese Metalexical papers 1911-1989* ; Xu & Ding 2003 ; Wang D. & Wang L. 2013 ; Wang 2019 ; Wei 2014, Wei & Geng & Wang 2011, etc.), ce qui rend leur consultation difficile. Par ailleurs, certaines ressources en ligne (*Euralex Bibliography of Lexicography*, *Kleines Wörterbuch der Verlaufsformen im Deutschen*, *GLex*, *IHMLL*, *LexBib*, *Musée Virtuel des Dictionnaires* en France, etc.)¹ restent souvent monolingues, et les barrières linguistiques entravent l'accès à la littérature dans d'autres langues, isolant ainsi certains chercheurs de la communauté lexicographique.

Les résultats de recherche dans les portails bibliographiques dépendent de la précision et de l'exhaustivité des métadonnées, soulignant ainsi la nécessité d'une bibliographie bien conçue, qui répond à des critères stricts pour être véritablement utile. Cela engendre un besoin pressant d'un effort concerté pour développer des outils adaptés qui répondent aux exigences spécifiques de la recherche dans cette discipline, en intégrant des corpus variés tels que des articles et des œuvres complètes.

Objectifs

Le projet *Digital Bibliography and Library of Lexicography (DB2L) : Chinese-Foreign Languages* vise à résoudre ces problèmes en développant bibliographie en ligne spécifique au domaine de la lexicographie et de la recherche sur les dictionnaires (c'est-à-dire la métalexicographie) contemporaines (à partir des années 2000 avec le numérique, soit environ 30 ans après la naissance de la métalexicographie au niveau international). Ce projet fournira des métadonnées de publication validées manuellement au fur et à mesure qu'elles seront nécessaires pour les citations, tout en étant complété par le développement d'une chaîne d'outils en traitement du langage naturel (NLP). Cela permettra d'offrir une bibliographie structurée et de tirer parti des avancées en NLP pour améliorer l'accès à l'information. En plus d'assurer une couverture complète de la littérature lexicographique, l'importance de cette proposition réside dans la possibilité d'une extraction facile des citations et l'introduction de l'indexation automatique des mots-clés (et de leur évaluation).

¹ Exemples de projets multilingues : *Elexifinder*, *Bibliography of Lexicography* de R.R.K. Hartmann (2007) ; *DBLP*.

Méthodes / Techniques

Le projet DB2L utilise diverses méthodes de NLP pour optimiser la collecte et l'organisation des métadonnées. Tout d'abord, des outils comme EndNote, Mendeley et Zotero seront employés pour recueillir des titres, auteurs, mots-clés, sources, langues, URL de publications, dates, éditeurs, ORCID et DOI/ISBN, avec une validation manuelle des métadonnées, qui pourront également inclure les affiliations des auteurs. Ces informations seront ensuite fusionnées avec des données structurées (TEI-XML ou BibTeX) à l'aide de Python, R, BBEdit, etc. pour enrichir DB2L et faciliter le référencement croisé. Les publications, notamment celles au format PDF ou HTML, seront nettoyées et converties en textes bruts. Les corps de texte seront isolés et lemmatisés à l'aide d'outils comme TreeTagger dans TXM et des outils développés par l'IMS, permettant ainsi de réaliser une extraction de termes thématiques et de références bibliographiques. Des outils NLP spécifiques au traitement du chinois, tels que Thulac pour la segmentation et HanLP pour des tâches avancées comme l'analyse syntaxique et l'extraction d'entités nommées, seront également intégrés. D'autres outils, tels que SpaCy, NLTK ou Jieba, pourraient également être utilisés pour le traitement linguistique et l'analyse. Par ailleurs, GROBID sera mis à profit pour extraire et structurer les références, construisant ainsi un réseau de citations pour visualiser les relations entre publications. Enfin, le projet aspire à devenir une plateforme collaborative (eg. PHP+ MySQL) pour enrichir la bibliographie et évaluer les métadonnées. Toutes les notices bibliographiques seront accessibles via une API, facilitant le téléchargement des métadonnées par des gestionnaires de références comme EndNote, Mendeley et Zotero, ce qui procurera une grande commodité pour les chercheurs dans ce domaine (méta)lexicographique.

Collaboration

Nous invitons les experts en lexicographie, les Talistes, les informaticiens, etc. à participer à la création de ce projet interdisciplinaire. L'objectif principal est d'enrichir les théories fondamentales de la métalexicographie et de développer des projets et outils reposant sur la riche littérature chinoise et étrangères. Parallèlement, nous souhaitons favoriser les échanges culturels et le développement des dictionnaires en chinois et dans d'autres langues.

Les références employées seront principalement axées sur les théories et pratiques liées aux dictionnaires, couvrant tous les sujets depuis les années 2000, y compris des ouvrages spécialisés, des articles de revues, et des dictionnaires influents tels que le 汉语辞海, le 现代汉语词典, le 新华字典, des références françaises comme le *Robert*, le *Larousse*, et le *Littré*, ainsi que les dictionnaires anglais comme l'*Oxford English Dictionary*, *Oxford Advanced Learner's Dictionary* et le *Collins COBUILD Advanced Learner's Dictionary*, etc.

1. (Méta)Lexicographie monolingue : monographies, articles, lexicographie informatique, encyclopédies, dictionnaires spécialisés, utilisation du dictionnaire, dictionnaires d'apprentissage, etc.
2. (Méta)Lexicographie bilingue : monographies théoriques sur les dictionnaires bilingues en lien avec le chinois et d'autres langues.

Pour encourager une collaboration internationale, nous nous intéressons également aux aspects suivants :

3. Principales associations en lexicographie
4. Revues de premier plan dans chaque pays (tel que 辞书研究, *IJL*, *Lexikos*, *Lexicography*, *ELA*, *Cahier de dictionnaires*, etc.)
5. Laboratoires de dictionnaires
6. Principaux événements liés aux dictionnaires (colloques, journées du dictionnaire, etc.)
7. Biographies des grands (méta)lexicographes tels que HUANG Jianhua en Chine, PRUVOST Jean en France, TARP Sven du Danemark, GOUWS Rufus H. en Afrique du Sud, etc.

Activités scientifiques :

- Colloques, journées d'études, publications

Coordinateurs :

CHEN Lian 陈恋 : lian.chen@univ-orleans.fr & loselychen@gmail.com
(Cergy Paris University - LT2D - Centre Jean Pruvost & LLL - CNRS, Orléans University)

GENG Yundong 耿云冬 : benedict_geng@mail.hzau.edu.cn
(Département des langues étrangères - Université agricole de Huazhong, Chine)