

Lexicographie informatique et intelligence artificielle : analyse contrastive des évolutions en France et en Chine

Computer lexicography and artificial intelligence:
a contrastive analysis of developments in France and China

Lian Chen / 陈恋

LLL-CNRS, Université d'Orléans & CRLAO-CNRS-INALCO

lian.chen@univ-orleans.fr

<https://orcid.org/0000-0001-5609-7524>

Abstract

This article offers a comparative study of the development of computational lexicography in France and China, analyzing their historical trajectories, institutional frameworks, and technological orientations. In France, the discipline is rooted in an academic tradition structured around major infrastructures such as the ATILF, the TLFi, Frantext, and the CNRTL, with a strong emphasis on digital humanities, data standardization (XML/TEI, LLOD), and interoperability, particularly through tools and models such as NooJ and OntoLex-Lemon. In China, development, although more recent, has been rapid since the 1980s and is based on the creation of massive corpora, semantic databases such as HowNet, and the increasing integration of artificial intelligence and large language models into lexicographical production. The comparison highlights two complementary approaches: a heritage-based and methodologically rigorous tradition in France, and a pragmatic, technology-driven dynamic in China, paving the way for multilingual, and interconnected digital lexicography.

Keywords: Computational lexicography, Digital humanities, Artificial intelligence, French-Chinese

1. INTRODUCTION : TRAJECTOIRES SOCIO-LEXICOGRAPHIQUES EN CHINE ET EN FRANCE (XX^e–XXI^e SIÈCLES) ET GENÈSE DE LA LEXICOGRAPHIE INFORMATIQUE

1.1. CADRE THÉORIQUE

L'émergence de la lexicographie informatique s'inscrit dans un long processus d'intégration entre la lexicographie traditionnelle et les technologies informatiques. Depuis plus d'un demi-siècle, la lexicographie connaît une transformation profonde sous l'effet des évolutions technologiques. Dès le milieu du XX^e siècle, les progrès de la linguistique informatique et la création des premiers corpus numériques ont profondément modifié la manière de concevoir, de compiler et de consulter les dictionnaires. Ce mouvement, souvent désigné en anglais sous le nom de *computational lexicography* (Boguraev & Briscoe, 1989), ou en français *lexicographie informatique* (Pierrel, 2010) ou *lexicographie informatisée* (Béjoint, 2007 ; Martin, 2008 ; Aquino-Weber & Greub, 2020)¹, et en chinois 计算词典学 (Zhang, 2013), marque la rencontre entre la linguistique, l'informatique et le traitement automatique du langage (TAL). Ainsi, la lexicographie est passée d'un travail manuel sur fiches papier à une approche fondée sur la modélisation et la structuration numériques des données linguistiques. L'apparition des premiers corpus numériques, des bases de données lexicales et des outils de concordance a marqué le passage d'une description manuelle du lexique à une modélisation informatisée, appuyée sur l'extraction automatique d'informations linguistiques (p. ex. OntoLex-Lemon, Sens-Text).

Le paysage de l'automatisation de la lexicographie (Kilgariff & Pavel, 2010 ; Rundell & Kilgariff, 2011 ; Jakubíček & Rundell, 2023) a connu des avancées notables ces dernières années, en grande partie grâce à l'avènement des technologies d'intelligence artificielle (IA) telles que ChatGPT (Jakubíček & Rundell, 2023 ; De Schryver & Joffe, 2023 ; De Schryver, 2023a, 2023b ; De Schryver & Joffe, 2023 ; Lew, 2023 ; Merken, 2023). À l'ère du numérique et de l'IA, la lexicographie a connu d'importantes transformations, réinventant ses méthodes et ses outils (Pruvost, 2000 ; Zhang, 2013 ; Béjoint, 2007 ; Aquino-Weber & Greub, 2020 ; Molinari, 2021 ; Molinari & Vincent, 2021). Des projets monolingues et multilingues ont

¹ Le terme *lexicographie numérique* (Zaklani, 2020) est aujourd'hui largement employé dans la littérature francophone pour désigner les pratiques lexicographiques liées aux supports et outils numériques. Cependant, dans cet article, nous privilégions l'expression *lexicographie informatique*. En effet, lexicographie numérique constitue principalement une traduction de digital lexicography, laquelle insiste sur la dimension médiatique et technologique (supports digitaux, interfaces, diffusion en ligne) plus que sur les processus algorithmiques sous-jacents (Müller-Spitzer, 2014 ; Lew, 2015). À l'inverse, lexicographie informatique correspond davantage au champ de la *computational lexicography*, terme consacré dans la recherche internationale, qui renvoie explicitement à la modélisation, au traitement automatique et à l'exploitation computationnelle des données lexicales (Boguraev & Briscoe, 1989).

même expérimenté la « post-édition lexicographique » à grande échelle (Baisa *et al.*, 2019). Cette « métamorphose, de l'incunable à l'espace virtuel » (Pruvost, 2009), initialement déclenchée par l'introduction de l'outil informatique dans le processus lexicographique, s'est ensuite complexifiée avec l'émergence d'Internet et le développement de la lexicographie contributive. Elle illustre un ensemble de transformations successives et diverses qui ont redéfini la nature même de la lexicographie. Comme le soulignent divers (méta)lexicographes (Pruvost, 2000, 2004, [2006], 2021, 2014 ; Jacquet-Pfau, 2005 ; Béjoint, 2007 ; Polguère, 2014 ; Vincent, 2023), l'intégration du numérique dans le champ lexicographique a non seulement redéfini le rapport à l'espace des dictionnaires, la structure et l'usage des dictionnaires – en enrichissant leurs fonctionnalités – mais a également transformé leur conception même.

La France et la Chine ont su tirer parti de l'informatisation et de la numérisation pour améliorer la compilation, la diffusion et l'utilisation des dictionnaires. Néanmoins, ces mutations n'ont pas suivi un parcours homogène à l'échelle mondiale. Avant d'aborder le développement de la lexicographie informatique en France et en Chine aux XX^e et XXI^e siècles, il convient de ne pas négliger l'évolution de la socio-lexicographie (Béjoint, 1994 ; Gottlieb *et al.*, 2005) dans ces deux pays aux XIX^e et XX^e siècles. En effet, c'est dans un contexte profondément marqué par les transformations sociales, politiques, économiques et culturelles que se sont façonnées les trajectoires distinctes de la lexicographie française et chinoise. Ces contextes historiques ont jeté les bases des différences observables plus tard dans le développement des dictionnaires numériques.

1.2. LA SOCIO-LEXICOGRAPHIE EN CHINE ET EN FRANCE AUX XX^e ET XXI^e SIÈCLES

En France, avec la seconde moitié du XX^e siècle s'ouvre une nouvelle période, souvent qualifiée de « siècle d'or » (Pruvost, 2021), car elle s'avère révolutionnaire pour la lexicographie stimulée par l'informatique : on passe progressivement du support papier à l'ordinateur, d'abord au CD-ROM, puis directement à Internet, facilitant une diffusion plus large. Quatre grands dictionnaires de langues se distinguent : le *Dictionnaire alphabétique et analogique de Paul Robert*, en 6 volumes, achevé en 1964 ; le *Grand Larousse de la langue française*, achevé en 1978 ; le *Trésor de la langue française*, en 16 volumes, achevé en 1994 ; et le *Dictionnaire de l'Académie française*, publié au siècle suivant, en 2025. Il existe également la version multimédia Hachette, longtemps très présente sur CD-ROM. Cependant, une autre grande ère s'ouvre avec Internet, et le XXI^e siècle correspond à une révolution significative dans le monde des dictionnaires.

En Chine, le XX^e siècle est une période de transformations linguistiques. La lexicographie s'est développée dans un contexte de profonds bouleversements

socio-politiques : l'influence de la pensée occidentale, la fondation de la République populaire en 1949 et la volonté de modernisation linguistique ont conduit à la production de dictionnaires adaptés aux besoins éducatifs et culturels. Cette époque est marquée par des changements importants dus aux influences politiques et sociales, notamment avec la publication de 辭海 [Cíhǎi, *Mer des Lexie*] (1915), l'un des premiers dictionnaires complets. Entre 1949 et 1965, afin de répondre aux besoins urgents en matière d'alphabétisation, des dictionnaires pratiques, au format et au contenu nouveaux, apparurent. Cependant, le développement des dictionnaires connut une période difficile. La *Révolution culturelle*, de 1966 à 1977, marqua un recul : pendant dix ans, le seul dictionnaire disponible fut le 新华字典 [Xīnhuá zìdiǎn, *Dictionnaire Xinhua*], qui devint un ouvrage de référence essentiel. En juillet 1967, même *Mer des Lexie*, fut sévèrement critiqué par l'organisation centrale pour propagande et publication pendant la *Révolution culturelle*. Après la *Révolution culturelle*, les dictionnaires connurent une renaissance. Les innovations technologiques ont également permis la création du 现代汉语语法信息词典 [Xiàndài hànyǔ yǔfǎ xìnxi cídiǎn, *Dictionnaire d'information sur la grammaire chinoise moderne*] (1977), intégrant des éléments de grammaire moderne. Au XX^e siècle, le passage aux médias numériques a transformé la compilation et l'accès aux dictionnaires dans tous les pays. L'intégration de l'informatique et de la numérisation a joué un rôle crucial dans la transformation des processus lexicographiques traditionnels (compilation, organisation, recherche à l'aide d'outils numériques). La transition vers les supports numériques a profondément transformé la lexicographie française et anglaise, avec l'apparition des dictionnaires électroniques et des éditions sur CD-ROM. En Chine, cette évolution s'est amorcée avec des projets d'envergure tels que le 汉语大词典数据库 [Hànyǔ dà cídiǎn shùjùkù, *Base de données du Grand dictionnaire du chinois*] et le 国家语言资源监测与研究中心 [Guójiā yǔyán zīyuán jiāncè yǔ yánjiū zhōngxīn, *Centre national de surveillance et de recherche sur les ressources linguistiques*]. Bien que légèrement plus tardive, la transition numérique chinoise a également conduit au développement de dictionnaires numériques, de bases de données en ligne, ainsi que de dictionnaires bilingues et de projets de traduction automatique (Chen, 2021, pp. 298-304).

Au cours du XX^e siècle, la France et la Chine ont chacune produit d'importants ouvrages lexicographiques qui sont devenus des textes fondateurs pour leurs langues respectives (TLF, 汉字大词典 [Hànzì dà cídiǎn, *Grand dictionnaire chinois des sinogrammes*]), stimulés par des changements sociopolitiques et culturels. L'utilisation de corpus pour la compilation de dictionnaires est devenue une pratique courante dans ces pays, améliorant la richesse et l'accessibilité de leurs ressources linguistiques. Malgré ces similitudes, le contexte historique et les motivations des développements lexicographiques différaient considérablement entre les deux pays.

Les orientations typologiques et institutionnelles ont également divergé. En France, la tradition des grands dictionnaires monolingues, complets et réguliers, s'est maintenue, témoignant d'une approche visant à refléter l'état global de la langue et

des connaissances. En Chine, en revanche, l'accent a été mis sur la production de dictionnaires bilingues et pédagogiques, conçus pour favoriser la diffusion du mandarin standard et l'accès aux savoirs internationaux. Par ailleurs, si en France la lexicographie institutionnelle a longtemps reposé sur des initiatives privées, comme celles de *Larousse*, en Chine l'État a joué un rôle moteur, notamment après 1949, en soutenant la normalisation linguistique et la création de dictionnaires nationaux. Malgré des trajectoires historiques et politiques distinctes, la France et la Chine ont connu des dynamiques convergentes dans la modernisation de la lexicographie : utilisation croissante des corpus, informatisation des processus de compilation, et intégration progressive des technologies numériques dans la conception des ressources. Ces évolutions ont ouvert la voie, dans les deux pays, à la création de dictionnaires numériques, multimédias et en ligne, puis, plus récemment, à l'intégration de l'intelligence artificielle dans la production et la diffusion des ressources lexicographiques.

Dans ces deux pays, la lexicographie a intégré des outils informatiques et des corpus numériques, facilitant la création de dictionnaires enrichis par des modèles d'intelligence artificielle tels que ChatGPT. Ces corpus (*FRANTEXT* en France, le *corpus du Centre de Pékin* en Chine, *etc.*) assurent la représentativité des données lexicales et soutiennent le développement de divers dictionnaires spécialisés. La France privilégie les ouvrages détaillés tels que *Le Robert* et *Larousse*, ainsi que des projets comme le *TLFi* et le *DDF* (2018). La Chine se concentre sur les dictionnaires d'apprentissage et met l'accent sur les dictionnaires bilingues et multilingues, tels que le 汉外多语言词典数据库 [*Hàn wài duō yǔyán cídiǎn shùjùkù*, *Base de données de dictionnaire multilingue chinois-langues étrangères*] (2022), influencés par la *Nouvelle Route de la Soie*, pour intégrer les connaissances occidentales et promouvoir la mondialisation. L'État chinois a joué un rôle majeur dans la normalisation lexicale, contrairement à la France, où les institutions universitaires dominent et où les acteurs privés ont historiquement été plus influents. Les bouleversements sociopolitiques en Chine ont guidé ses orientations lexicographiques, tandis que la France a bénéficié de progrès scientifiques plus progressifs. Les deux pays ont adapté leurs traditions lexicographiques aux nouvelles technologies, mais des défis subsistent, notamment l'intégration éthique de l'intelligence artificielle : éviter les définitions biaisées, garantir la transparence, respecter la paternité et préserver la diversité linguistique.

L'arrivée récente de l'intelligence artificielle générative (IA générative) et des grands modèles de langue (LLM) ouvre aujourd'hui une nouvelle étape. Les dictionnaires ne sont plus seulement des produits finis, mais des systèmes dynamiques capables d'auto-apprentissage, de génération d'exemples bilingues, d'analyse sémantique ou d'interaction personnalisée. Cette transformation soulève à la fois des opportunités inédites (automatisation, accessibilité, enrichissement sémantique).

Ainsi, en partant des fondements de la lexicographie informatique jusqu'aux approches actuelles fondées sur les modèles de langage et la sémantique computationnelle, cette recherche vise à contribuer à une compréhension renouvelée de la

lexicographie comme interface entre les données, les savoirs et les cultures. Les connaissances relatives à la lexicographie informatique française présentées dans cet article proviennent principalement des ouvrages de Jean Pruvost, notamment *Les dictionnaires français : outils d'une langue et d'une culture* (Éditions Ophrys) et *Dictionnaires et nouvelles technologies* (PUF). En revanche, les connaissances concernant la lexicographie informatique et informatique en Chine s'appuient principalement sur l'ouvrage de Zhang Yihua, 计算词典学 [Jisuan cídiǎn xué, Computational Lexicography] (2013).

1.3. MÉTHODOLOGIE

Cette étude adopte une approche qualitative, comparative et historiographique de la lexicographie numérique en France et en Chine. Elle repose sur une analyse documentaire croisée de sources académiques, institutionnelles et techniques, comprenant des publications scientifiques, des rapports officiels, ainsi que l'examen de ressources lexicographiques numériques emblématiques (telles que le *TLFi*, le *CNRTL*, *HowNet* ou les grands dictionnaires chinois contemporains).

La comparaison s'articule autour de plusieurs dimensions : les cadres institutionnels, les modèles éditoriaux, les orientations méthodologiques, les choix technologiques et les logiques socio-économiques sous-jacentes. Cette démarche vise moins à établir une typologie exhaustive qu'à mettre en évidence des tendances structurantes et des différences systémiques entre deux écosystèmes lexicographiques distincts.

Enfin, l'analyse intègre les développements récents liés à l'IA générative afin d'évaluer leur impact sur la convergence ou la reconfiguration des modèles traditionnels de production lexicographique.

2. DÉVELOPPEMENT DE LA LEXICOGRAPHIE INFORMATIQUE : DE L'INFORMATISATION À L'INTERNET, PUIS À L'IA

2.1. ÉVOLUTION EN FRANCE

À partir des années 1950, sous l'impulsion de Quemada, les recherches sur le lexique s'appuient sur les premières technologies mécanographiques, notamment les cartes perforées utilisées dans les laboratoires de Besançon. Dès les années 1960, les lexicographes commencent à utiliser l'ordinateur pour assister la rédaction des dictionnaires papier. Ainsi, le premier dictionnaire ancien à être informatisé fut le *Thresor de la langue françoise* (1606) de Jean Nicot. Le texte a été saisi à Nancy et à Toronto entre 1979 et 1984, indexé à l'Université de Toronto sur grand ordinateur par le programme

de concordance COGS, publié sous forme d'une concordance sur microfiches en 1985, indexé sur un ordinateur personnel sous WordCruncher comme base de données interactive en 1988 et finalement mis en ligne dans la Toile en 1994.

Dans les années 1990, les dictionnaires numériques sur CD-ROM ont marqué une étape importante. Par exemple, la version numérique du *Petit Robert* (1996) permettait déjà des recherches complexes : anagrammes, homophones, étymologies ou citations. Des ressources comme le *Trésor de la langue française informatisé* (TLFi), le *Dictionnaire de l'Académie française* ou encore *Larousse.fr* deviennent les symboles de cette nouvelle ère numérique (Šonková, 2010). Dans ce contexte, la France a suivi l'exemple de l'*Oxford English Dictionary informatisé* en Angleterre. Fort de l'expérience du corpus *FRANTEXT*, l'INALF a entrepris la création d'une version numérique du *TLF*.

La lexicographie française a profondément évolué avec l'arrivée de l'informatique puis du Web, notamment grâce aux travaux de la *Text Encoding Initiative* (TEI). La mise en ligne du *TLFi* en 2002 a constitué une avancée majeure : accessible gratuitement, il offre des possibilités de recherche dans vingt-neuf champs différents (définitions, exemples, domaines, synonymes, etc.). Le *TLFi*, publié sur cédérom en 2004, représente l'aboutissement de ce travail. Il est aujourd'hui librement consultable en ligne sur le site de l'Analyse et Traitement Informatique de la Langue Française (ATILF). Cette mise à disposition marque une nouvelle étape : le passage à la lexicographie en ligne, ou lexicographie informatique. Le Web a également permis la création de dictionnaires interactifs et hypertextuels, tels que le *Dictionnaire universel francophone* (DUF), où chaque mot est relié à d'autres par des liens cliquables. D'autres grands dictionnaires français ont également été numérisés ou mis en ligne dans les années 1990 et 2000 : *Le Grand Robert* (1989), « premier grand dictionnaire de langue pouvant ainsi être consulté sur cédérom » (Pruvost, 2006, p. 93), le *Dictionnaire de l'Académie française*, *Le Petit Robert*, *Le Petit Larousse*, le *Dictionnaire encyclopédique multimédia Hachette*, ainsi que des plateformes telles que *Le Littré* (Reverso), *L'Internaute*, *Larousse.fr*, *TV5Monde Dictionnaire* ou encore le *Wiktionnaire*. Ces ressources illustrent la transition progressive de la lexicographie française vers le numérique, où les dictionnaires deviennent des bases de données consultables en ligne, interrogeables et souvent interactives. Ce mouvement s'inscrit dans une dynamique plus large de numérisation du patrimoine linguistique français, amorcée avec le *TLF* et poursuivie aujourd'hui par des projets ouverts et collaboratifs, tels que le *Wiktionnaire* et les bases lexicales de l'ATILF. En 2009, *Le Petit Robert*, *Le Grand Robert* et *Le Grand Robert & Collins* deviennent accessibles en ligne, depuis n'importe quel ordinateur, grâce à un identifiant ou à une clé d'accès liée à l'achat du dictionnaire papier. L'innovation se poursuit avec la première application mobile *Le Robert Dixel* en 2010, qui offre un accès rapide et interactif aux définitions sur iPhone et iPad. Depuis, les éditions *Le Robert* développent des outils numériques tels que *Le Robert Correcteur* (2014) et *La Certification Le Robert* (2016), afin d'aider chacun à améliorer sa maîtrise du français et à la faire valoir.

La lexicographie informatique s'enrichit également d'outils variés comme la traduction automatique (ex. Babelfish), conçue pour faciliter la compréhension rapide de textes en ligne.

En parallèle, à côté des grands projets institutionnels tels que le *TLFi* ou le *DUF*, on voit apparaître de nombreux glossaires et dictionnaires collaboratifs créés par des amateurs, qui participent à la diffusion du savoir linguistique sur le Web. La lexicographie informatique française s'appuie aujourd'hui sur les technologies du Web sémantique, l'IA et les bases de données linguistiques pour offrir des outils de recherche de plus en plus puissants, accessibles à la fois aux chercheurs et au grand public. Dans le laboratoire ATILF, et en s'appuyant sur la lexicologie explicative et combinatoire (Mel'čuk *et al.*, 1995) ainsi que sur les avancées récentes en matière de réseaux lexicaux permettant la génération automatique de dictionnaires numériques interconnectés, les travaux de Polguère (2000, 2009, 2014) sur la construction de lexiques en réseau démontrent comment les unités lexicales et leurs relations paradigmatisées (dérivations sémantiques) et syntagmatiques (collocationnelles) peuvent être modélisées sous forme de graphes navigables, conformément aux principes de la lexicologie explicative et combinatoire.

2.2. ÉVOLUTION EN CHINE

Le développement de la lexicographie computationnelle en Chine est étroitement lié à l'essor de la traduction automatique. Les dictionnaires numériques et bases lexicographiques automatisées constituent en effet l'une des ressources fondamentales des systèmes de traitement automatique du langage (TAL), tout en représentant à la fois leur socle de connaissance et leur principal goulot d'étranglement (Feng, 2001 ; Lin, 1999). En effet, le dictionnaire machine est considéré comme la base essentielle de tout système de TA : « Sans bon dictionnaire, la traduction automatique est impossible » (Feng, 2001, p. 99). La recherche sur la traduction automatique en Chine a débuté en 1956, avec un premier axe de travail centré sur la traduction du russe vers le chinois. Toutefois, en raison des limites technologiques et du contexte socio-économique de l'époque, les avancées ont rapidement marqué un ralentissement. À partir du milieu des années 1970, le pays a renoué avec ce champ de recherche, amorçant une phase de relance qui, entre le milieu des années 1980 et le début des années 1990, a constitué une étape charnière dans le développement du domaine. Depuis les années 1990, la Chine a vu apparaître plusieurs logiciels de traduction commerciale, tels que le 高立英汉系统 [*Gāo lì yīnghàn xìtǒng*, *Gao Li English-Chinese System*], 金山快译 [*Jīnshān kuài yì*, *Matrix et Kingsoft Express Translation*], marquant ainsi le passage de la recherche expérimentale à l'industrialisation de la linguistique informatique.

Le développement du *Dictionnaire automatique des verbes prédicats du chinois moderne* revêt une importance particulière, car les verbes prédicats

occupent une position centrale dans la structure syntaxique des phrases. Leur description formelle conditionne directement la qualité de l'analyse sémantique et constitue, selon Lin (1999, p. 158), le « maillon essentiel » de tout système lexicographique automatique. Parallèlement, le *Dictionnaire des concepts chinois*, élaboré par l'Université de Pékin, s'inspire du modèle *WordNet* pour construire un réseau sémantique centré sur les concepts du chinois. Ce projet permet à la fois la visualisation des relations conceptuelles et le développement d'interactions homme-machine, faisant de cette ressource une réalisation emblématique de la lexicographie informatique chinoise. Dans l'ensemble, le développement de la lexicographie informatique en Chine s'est opéré en plusieurs étapes : d'abord centrée sur la traduction automatique, elle s'est progressivement orientée vers la construction de dictionnaires automatisés, avant d'aboutir à l'intégration de ressources sémantiques interconnectées.

Depuis le début du XXI^e siècle, la lexicographie informatique tend à fusionner avec l'ontologie lexicographique, en mettant l'accent sur l'intégration entre dictionnaires automatiques et dictionnaires lisibles par l'humain. Dans cette perspective, *McNaught* (1988) avance que les dictionnaires numériques du futur devront réaliser une véritable « dualité homme-machine », conciliant à la fois la lisibilité pour les systèmes informatiques et la convivialité pour les utilisateurs humains. Ce concept offre ainsi une base théorique solide et une orientation pratique pour le développement des dictionnaires numériques et intelligents en Chine. Les dictionnaires machines chinois se sont diversifiés : 现代汉语语法信息词典 [Xiàndài hànyǔ yǔfǎ xìnxī cídiǎn, *Dictionnaire d'information grammaticale du chinois moderne*] (Université de Pékin) : riche en informations morphosyntaxiques ; 现代汉语语义词典 [Xiàndài yǔyì hànyǔ cídiǎn, *Dictionnaire sémantique du chinois moderne*] (Université de Pékin & Académie des Sciences) : orienté vers la sémantique computationnelle ; 现代汉语述语动词机器词典 [Xiàndài hànyǔ shù yǔ dòngcí jīqì cídiǎn, *Dictionnaire machine des verbes prédicatifs du chinois moderne*] (Tsinghua & RUC) : centré sur les verbes prédicatifs, considérés comme le cœur syntaxique de la phrase (Lin, 1999, p. 158) ; 中文概念辞书 [Zhōngwén gàiniàn cídiǎn, *Chinese Concept Dictionary*] : inspiré de *WordNet*, adapté à la structure sémantique du chinois.

3. ÉCOSYSTÈMES INSTITUTIONNELS ET ÉDITORIAUX

3.1. PROJETS LEXICOGRAPHIQUES PUBLICS FINANCÉS PAR L'ÉTAT

La lexicographie publique française s'est développée à travers une série de projets institutionnels soutenus et financés par l'État, visant à documenter, normaliser et valoriser la langue française dans toutes ses dimensions. Trois grandes étapes jalonnent cette évolution (voir Tableau 1) :

Tableau 1. Trois jalons institutionnels de la lexicographie française

Période	Projet	Institutions	Objectifs principaux	Innovations
XVII ^e siècle → aujourd’hui	<i>Dictionnaire de l’Académie française</i> (1 ^{re} éd. 1694)	Académie française	Normaliser la langue ; en garantir la pureté et l’usage	Démarche normative ; publication papier → mise en ligne ; outils de recherche
1960-1994 → version numérique (2002)	<i>Trésor de la langue française (TLF)</i> → <i>TLFi</i>	CNRS → ATILF	Documenter le français moderne sur base textuelle ; produire un dictionnaire scientifique historique	Corpus monumental (Frantext) ; balisage XML (~36 M balises) ; fonctions de recherche avancées ; navigation hypertextuelle
Depuis 2021	<i>Dictionnaire des francophones (DDF)</i>	Ministère de la Culture ; Huma-Num	Valoriser la diversité du français dans la francophonie ; intégrer données institutionnelles et contributions citoyennes	Modèle participatif encadré ; architecture RDF, OntoLex-Lemon ; LLOD ; SPARQL ; visualisation des variations diatopiques

La lexicographie chinoise contemporaine s’appuie sur plusieurs projets financés par l’État (voir Tableau 2), qui ont joué un rôle central dans la normalisation et la diffusion de la langue chinoise au XX^e siècle.

Tableau 2. Trois jalons institutionnels de la lexicographie française

Projet / Dictionnaire	Institution principale	Description
现代汉语词典 <i>Modern Chinese Dictionary</i>	Académie chinoise des sciences sociales (CASS) / Commercial Press	Dictionnaire normatif du mandarin moderne
汉语大词典 <i>Grand dictionnaire du chinois</i>	Maison lexicographique de Shanghai / Commission nationale de la langue	Dictionnaire historique du chinois, du classique au moderne
汉语方言大词典 <i>Grand dictionnaire des dialectes chinois modernes</i>	Université Fudan / Ministère de l’Éducation	Description et préservation des dialectes chinois
国家语言资源词典工程 <i>Projet national de ressources linguistiques</i>	Commission nationale de la langue	Ressources lexicographiques pour néologismes et langues minoritaires
中华通韵 <i>Compendium of Rhymes</i>	CASS	Dictionnaire des systèmes de rimes classiques
有声电子词典与数据库 <i>Dictionnaire audio et base phonétique</i>	CASS / Institutions éducatives	Données phonétiques et prononciation

3.2. MAISONS D'ÉDITION ET INDUSTRIE LEXICOGRAPHIQUE PRIVÉE

La lexicographie informatique s'est d'abord développée dans le cadre des maisons d'édition privées, qui ont joué un rôle central dans la transition du dictionnaire imprimé vers le dictionnaire numérique.

Deux acteurs majeurs dominent l'histoire de cette évolution en France : *Larousse* et *Le Robert*, qui exploitent aujourd'hui plusieurs modes de diffusion de leurs dictionnaires. 1) Fondée en 1856 par *Pierre Larousse*, la maison *Larousse* s'impose dès le XIX^e siècle comme un symbole de la diffusion du savoir encyclopédique. Avec l'essor de l'informatique, elle amorce une modernisation décisive : publication de l'*Almanach Hachette* en 1988, lancement du *Bibliorom Larousse 2.0* en 1996, puis du *Petit Larousse* sur CD-ROM en 2004. Ces versions numériques ont marqué le passage du dictionnaire papier à des formats interactifs, intégrant recherche rapide, hyperliens et contenus multimédias. Aujourd'hui, Larousse poursuit cette évolution avec une plateforme en ligne proposant dictionnaires, encyclopédies et ressources linguistiques accessibles gratuitement. 2) La maison *Le Robert*, fondée en 1951 par *Paul Robert*, a suivi une trajectoire parallèle. Dès 1989, elle propose une version informatisée du *Petit Robert*, avant de développer des éditions sur CD-ROM et en ligne. Sous l'impulsion de lexicographes tels qu'*Alain Rey*, *Le Robert* s'est imposé comme un acteur majeur de la lexicographie contemporaine, alliant rigueur scientifique et adaptation technologique. Son site web actuel illustre la continuité entre l'héritage lexicographique classique et les nouvelles pratiques numériques, à la croisée du savoir linguistique et de l'innovation numérique.

Les deux éditeurs associent désormais leurs dictionnaires papier à un accès numérique sans publicité. Chez *Larousse*, l'achat du *Petit Larousse illustré* ouvre l'accès au *Larousse-Webdictionnaire* pour l'année correspondante, tandis que chez *Le Robert*, l'acquisition du *Robert illustré* permet de consulter en ligne, pendant quatre ans, une version numérique enrichie. Cette logique « bimédia » s'étend également aux dictionnaires scolaires (*Larousse junior*, *Robert collège*, etc.). Parallèlement, *Le Robert* est le seul à proposer un accès payant par abonnement à ses grands dictionnaires en ligne (*Le Grand Robert*, *Le Petit Robert*, *Le Robert collège* et *Le Robert junior*), un modèle qui constitue aujourd'hui le principal mode de diffusion du *Grand Robert*. Les deux maisons se positionnent aussi sur les applications mobiles : *Larousse* propose plusieurs titres pour iOS et Android, dont le Dictionnaire de français, *Larousse junior* et *Le Petit Larousse illustré 2019* pour iPhone, tandis que *Le Robert* privilégie iOS avec le *Petit Robert*, *Le Robert mobile* et le *Dictionnaire historique de la langue française*, seule version numérique de cet ouvrage. L'analyse de ces offres (Gasiglia, 2020, 2021) montre que la numérisation se concentre sur les dictionnaires de référence et les versions scolaires, que les applications iOS sont privilégiées, et que seul *Le Robert* maintient un modèle d'abonnement. Ce dernier illustre également la convergence entre correction linguistique et lexicographie informatique en intégrant une version

du *Robert illustré* dans son logiciel *Le Robert Correcteur*. Ainsi, la lexicographie informatique française s’organise autour d’une complémentarité entre accès gratuit et payant, papier et numérique, articulant objectifs commerciaux et pédagogiques : *Larousse* privilégie une diffusion large et gratuite soutenue par la publicité, tandis que *Le Robert* mise sur la valorisation et la monétisation de ses ressources.

Ainsi, la lexicographie informatique issue des maisons d’édition privées témoigne d’un mouvement progressif – du support imprimé au support numérique, puis au numérique en ligne – qui transforme profondément la production, la consultation et la diffusion des dictionnaires. Les entreprises privées comme Robert et Larousse ont abordé le marché d’Internet avec prudence, en pratiquant des tarifs élevés et en nouant des partenariats avec des écoles et des universités afin d’en proposer l’accès dans les bibliothèques. Dans les années 1990, le *Trésor de la langue française*, œuvre majeure du domaine public, a été numérisé et est devenu accessible gratuitement dès 1994.

En Chine, cinq grands éditeurs jouent un rôle de premier plan dans la publication de dictionnaires (voir Tableau 3).

Tableau 3. Cinq principaux éditeurs chinois spécialisés dans la lexicographie

Institution / Maison d’édition	Date de fondation	Jalons numériques	Innovations récentes
商务印书馆 The Commercial Press	1897	1998 : CD-ROM Xiandai Hanyu Cidian ; 2002 : premier dictionnaire en ligne	2013 : app mobile CP Dictionary ; 2021 : version interactive avec commande vocale, OCR et synchronisation cloud
外语教学与研究出版社 Foreign Language Teaching and Research Press	1979	1995 : CD-ROM English-Chinese Learner’s Dictionary ; 2005 : FLTRP Online	2012 : apps FLTRP Dic ; 2023 : assistant de traduction / rédaction basé sur GPT
上海外语教育出版社 Shanghai Foreign Language Education Press	1979	1998 : CD-ROM Oxford Advanced Learner’s Dictionary ; 2004 : portail en ligne	2011 : app DictBox ; 2022 : système de recommandation de collocations fondé sur corpus (API cloud)
上海译文出版社 Shanghai Translation Publishing House	1978	2000 : CD-ROM bilingue ; 2007 : site STPH Lexicon	2014 : app mobile ; 2023 : dictionnaires numériques avec OCR neuronal, compatibles liseuses (Kindle / Boox)
人民出版社 People’s Publishing House	1921	2003 : CD-ROM terminologie marxiste ; 2008 : portail terminologique juridique et politique	2018 : mini-app WeChat ; 2024 : diffusion de glossaires via API cloud dédiée

Dans l'ensemble, l'évolution lexicographique chinoise illustre une intégration progressive de l'intelligence artificielle, du cloud computing et du TAL dans la production et la diffusion des dictionnaires. Ce mouvement témoigne de la volonté des éditeurs chinois de faire de la lexicographie informatique un outil intelligent, interconnecté et participatif, au service de la recherche linguistique et de l'apprentissage des langues.

3.3. DICTIONNAIRES COLLABORATIFS

La lexicographie collaborative citoyenne représente une dimension essentielle de la lexicographie informatique contemporaine. Elle repose sur la participation active des internautes, qui contribuent collectivement à la création, à la mise à jour et à la diffusion des ressources lexicales. Ces initiatives, souvent financées par leurs communautés d'utilisateurs, reflètent une approche ouverte, participative et inclusive de la langue.

L'un des projets les plus emblématiques de la lexicographie informatique contemporaine est le *Wiktionnaire* francophone, lancé en 2004 dans le cadre du mouvement *Wikimedia*. Ce dictionnaire libre vise à décrire l'ensemble des langues du monde, dont le français, en proposant définitions, exemples, étymologies et traductions. Il incarne pleinement la lexicographie participative, où chaque usager peut devenir lexicographe en contribuant directement au contenu. Aujourd'hui, le *Wiktionnaire* francophone est l'une des éditions les plus dynamiques du projet international *Wiktionary*, avec plus de 20 000 nouvelles entrées françaises ajoutées chaque année. Fondé sur les principes du partage libre et de la participation collective, il s'est progressivement imposé comme une ressource de référence pour la langue française. Il repose sur une communauté très active qui enrichit constamment ses pages, notamment avec des définitions, exemples, illustrations, fichiers audio et données sémantiques. Le *Wiktionnaire* ne se limite pas à un usage en ligne : il a nourri la création du *Dictionnaire des francophones* et inspiré la publication du dictionnaire imprimé *Le Dico* (Éditions Garnier, 2019), qui rassemble 60 000 mots du français contemporain. Il s'inscrit également dans une dynamique d'innovation ouverte à travers des outils numériques gratuits, comme *Lingua Libre*, dédié à l'enregistrement de prononciations variées, et la *Dicothèque*, qui permet de consulter des définitions issues de dictionnaires historiques hébergés sur Wikisource, offrant une perspective lexicographique diachronique. Modèle du dictionnaire « ouvert et collaboratif » (Molinari, 2021, p. 28), le *Wiktionnaire* permet à tous les usagers de créer, modifier ou commenter des entrées, et de négocier les sens. Son fonctionnement s'appuie sur des mécanismes de « vigilance participative » (Pentzold, 2011) et d'« autorité dispositive » (Oger, 2021, p. 119), qui assurent une validation collective du contenu. Cette architecture favorise l'intégration d'usages variés – y compris familiers, néologiques ou régionaux – souvent absents

des dictionnaires traditionnels (Meyer, 2018, p. 306 ; Murano, 2019, p. 197). En 2015, le projet *Bob*, dictionnaire d'argot, est venu enrichir cette dynamique participative en se concentrant sur la langue familière et argotique. Ce dictionnaire en ligne, ouvert aux contributions du public, documente les usages populaires et non standard du français. Il permet ainsi d'intégrer dans la lexicographie informatique des variétés linguistiques souvent marginalisées dans les dictionnaires traditionnels. Enfin, créé en 2016, *Dicoado* incarne une nouvelle génération de lexicographie citoyenne, conçue par et pour les jeunes. Cette plateforme interactive offre aux adolescents la possibilité de proposer, définir et commenter des mots issus de leur vocabulaire quotidien. Dicoado favorise non seulement la créativité lexicale, mais aussi l'appropriation du français par ses jeunes locuteurs.

Ces trois projets – *Wiktionnaire*, *Bob* et *Dicoado* – témoignent de l'émergence d'une lexicographie ouverte, participative et décentralisée, où la communauté d'utilisateurs devient un véritable acteur de la description linguistique, en complément des institutions et des maisons d'édition traditionnelles.

En parallèle des grands projets lexicographiques institutionnels, la Chine a vu émerger depuis les années 2000 plusieurs dictionnaires collaboratifs et plateformes participatives, témoignant d'une évolution vers une lexicographie ouverte et numérique. Ces initiatives s'appuient sur la contribution d'internautes et de communautés linguistiques, tout en combinant modération humaine et technologies de gestion de contenu. Le site 汉典 [*Hàndiǎn*, *Zdic*] constitue une plateforme lexicale de référence, centrée sur la philologie et l'étymologie des sinogrammes. Bien qu'il ne soit pas ouvert à l'édition directe, il intègre un système de retours et de corrections communautaires, garantissant la fiabilité de ses sources tout en demeurant très riche en références classiques. *Baidu Baike*, l'encyclopédie du moteur de recherche Baidu, adopte un modèle similaire à celui de Wikipédia. Ouvert à l'édition publique et fortement modéré, il couvre un champ extrêmement vaste, allant des termes lexicaux aux notions culturelles et encyclopédiques. Le *Chinese Wiktionary* représente la version chinoise du dictionnaire collaboratif mondial de Wikipédia. Basé sur une modération communautaire, il reste en développement continu, notamment dans l'expansion de ses fonctionnalités multilingues et de ses ressources lexicales.

4. RESSOURCES ET TECHNOLOGIES

4.1. LINGUISTIQUE DE CORPUS ET RESSOURCES NUMÉRIQUES POUR LA LEXICOGRAPHIE

Avec l'essor des technologies informatiques, les données linguistiques ont pu être stockées, consultées et exploitées sous forme numérique, donnant naissance à la linguistique de corpus. Cette approche, devenue centrale pour l'étude empirique des

usages authentiques, a profondément transformé la recherche linguistique et joué un rôle clé dans le développement du TAL et de la lexicographie.

En France, cette dynamique s'est traduite par la constitution de grands corpus et d'outils numériques structurants, qui soutiennent aujourd'hui la recherche lexicographique. Plusieurs projets scientifiques visent à documenter, analyser et modéliser la langue à partir de ressources ouvertes à la communauté universitaire. Parmi ceux-ci, le Tableau 4 présente les principaux corpus et ressources numériques ayant marqué le développement de la lexicographie informatique en France.

Tableau 4. Principales ressources et plateformes lexicographiques et linguistiques numériques en France

Ressource / Projet	Institution	Type	Caractéristiques principales
<i>CNRTL (Centre National de Ressources Textuelles et Lexicales)</i>	ATILF, Nancy	Plateforme textuelle et lexicale	Regroupe dictionnaires (TLF, DMF, Académie), corpus et outils d'analyse morpho-syntaxique et sémantique
<i>DÉS (Dictionnaire Électronique des Synonymes)</i>	Université de Caen	Dictionnaire électronique	Recherche interactive de synonymes / antonymes ; mises à jour régulières
<i>Frantext</i>	ATILF, Nancy	Corpus textuel (accès contrôlé)	Plus de 270 millions de mots ; textes du XVI ^e s. à aujourd'hui ; recherche complexe (formes, lemmes, catégories)
<i>ESLO (Enquêtes Sociolinguistiques d'Orléans)</i>	Université d'Orléans	Corpus oral	Enregistrements et transcriptions du français parlé depuis les années 1960 ; étude de la variation sociolinguistique
<i>DiCo / GLAWI (via plateforme REDAC)</i>	Laboratoire CLLE, Toulouse	Ressources lexicales	Dictionnaire de cooccurrences (DiCo) ; base lexicale GLAWI ; informations morphologiques, syntaxiques, sémantiques
<i>Phraséotext / Lexicoscope</i>	LIDILEM, Grenoble	Outils d'analyse phraséologique	Outils dédiés à la phraséologie et à la lexicométrie ; exploration et visualisation des relations lexicales

Plusieurs projets ont également émergé pour valoriser le patrimoine lexicographique et en faciliter l'accès grâce aux technologies numériques. Le premier projet emblématique est *Le Musée Virtuel des Dictionnaires*, développé à Cergy. Ce musée en ligne offre une exploration historique et critique des dictionnaires français depuis leurs origines. Il permet d'accéder à des notices détaillées sur les grandes œuvres lexicographiques, leurs auteurs et leurs contextes de production. L'utilisateur peut effectuer des recherches chronologiques ou thématiques, consulter des bibliographies spécialisées et suivre l'évolution de la lexicographie à travers les siècles. Ce projet illustre la volonté de préserver et de diffuser la mémoire lexicographique française dans un cadre numérique et interactif. Un autre projet majeur est *Nénufar*, porté par l'Université de Montpellier. Il s'agit d'une édition numérique critique des anciens

dictionnaires Larousse, couvrant la période de 1906 à 1948. Ce projet vise à redonner vie aux éditions historiques du *Petit Larousse illustré*, en permettant la consultation en ligne des articles, des formes lexicales et des illustrations originales. Chaque entrée, comme l'exemple de *Nénufar*, donne accès à des informations sur les différentes éditions, les évolutions orthographiques et les contextes d'usage, tout en intégrant des fonctionnalités de recherche avancée (formes, images, XML, *etc.*). Ainsi, ces deux initiatives – *Le Musée Virtuel des Dictionnaires* et *Nénufar* – témoignent de l'importance croissante de la métalxicographie numérique en France. Elles conjuguent recherche historique, valorisation patrimoniale et innovation technologique, en transformant les anciens dictionnaires en véritables objets d'étude interactifs et accessibles au grand public.

Enfin, la plate-forme *FranceTerme*, hébergée par le ministère de la Culture français, offre une base terminologique destinée à enrichir la langue française via la recommandation de termes en français afin de remplacer les anglicismes ou les emprunts non francisés. Elle couvre de nombreux domaines (informatique, économie, transport, environnement, *etc.*) et propose non seulement une interface de recherche de termes (par exemple « actif numérique », « réseau métalloorganique ») mais aussi des « recommandations d'usage » publiées au Journal officiel.

En Chine, la période allant de 1979 à 1989 marque une étape fondatrice dans le développement des corpus linguistiques chinois, posant les bases de la lexicographie informatique et des études statistiques sur le chinois moderne. Plusieurs universités ont alors joué un rôle pionnier dans la constitution de ressources essentielles à la recherche linguistique et à l'enseignement du chinois. Le Tableau 5 présente les principaux corpus constitués durant cette période.

Tableau 5. Corpus chinois (1979-1989)

Corpus / Projet	Institution	Année	Type	Taille	Contribution principale
<i>Corpus of Modern Chinese Literature</i>	Université de Wuhan	1979	Corpus littéraire moderne	5,72 M mots	Premier corpus chinois d'envergure ; fondation de l'analyse fréquentielle moderne
<i>JDEST Academic English Corpus</i>	Université Jiaotong de Shanghai	1982	Corpus d'anglais académique	1 M → 3,6 M mots	Premier corpus multilingue académique en Chine
<i>Corpus of Middle School Chinese Textbooks</i>	Université normale de Pékin (BNU)	1983	Corpus pédagogique	1,068 M mots	Analyse de la langue scolaire et standardisation pédagogique
<i>Corpus of Modern Chinese Word Frequency Statistics</i>	Université des langues et cultures de Pékin (BLCU)	1983	Corpus de fréquence lexicale	1,82 M mots	Base statistique lexicale moderne pour l'enseignement et la recherche

<i>Corpus of Modern Chinese</i>	Université Beihang (BUAA)	1983	Corpus général moderne	20 M mots	Premiers outils statistiques à grande échelle sur le chinois
<i>Extended Corpus of Modern Chinese</i>	Université Beihang (BUAA)	1985	Corpus général étendu	20,24 M mots	Extension et consolidation du corpus général
<i>Modern Chinese Word Frequency Statistics Diachronic Corpus</i>	Université du Shanxi	1989	Corpus diachronique	2,5 M mots	Introduction de la dimension historique dans l'analyse du chinois moderne

Ces initiatives inaugurales, enrichies et diversifiées à partir des années 1990, ont jeté les fondements de la linguistique de corpus et de la lexicographie computationnelle en Chine, contribuant à faire du pays l'un des acteurs majeurs de la recherche linguistique appuyée sur les données (voir Tableau 6).

Tableau 6. Corpus depuis les années 1990 en Chine

Corpus / Projet	Institution	Type	Contribution
<i>Modern Chinese Corpus (NLC)</i>	National Language Committee	Corpus général	Référence nationale pour le chinois standard
<i>Modern Chinese Corpus</i>	Université de Pékin	Corpus général	Analyse de fréquence et usages lexicaux
<i>HuaYu Corpus</i>	Université Tsinghua	Corpus général	Corpus large pour études lexicales contemporaines
<i>HSK Dynamic Composition Corpus</i>	BLCU	Corpus d'apprenants (écrit)	Étude de l'acquisition du chinois langue étrangère
<i>Contemporary Beijing Spoken Corpus</i>	BLCU	Corpus oral	Documentation du chinois parlé contemporain
<i>Modern Chinese Spoken Corpus</i>	Académie chinoise des sciences sociales	Corpus oral	Étude systématique de l'oralité moderne
<i>CCL Chinese-English Parallel Corpus</i>	Université de Pékin	Corpus parallèle chinois-anglais	Recherche contrastive et traduction automatique
<i>Ancient Chinese Corpus</i>	National Language Committee	Corpus historique	Ressource clé pour le chinois classique
<i>Chinese Learner English Corpus (CLEC)</i>	GDUFS & SJTU	Corpus d'apprenants (anglais)	Analyse des erreurs des apprenants chinois d'anglais

Ces corpus, produits initialement entre 1979 et 1989 puis enrichis à partir des années 1990, constituent la base de la linguistique de corpus en Chine et ont permis

le développement de la lexicographie computationnelle, de la pédagogie du chinois et des études contrastives. Ils ont progressivement positionné la Chine comme un acteur majeur de la recherche linguistique outillée au niveau international.

Enfin, *HowNet* (知网) est une ressource lexicographique et ontologique chinoise pionnière, créée à la fin des années 1980 et mise en ligne en 1999, qui combine lexique, encyclopédie et ontologie linguistique en s'appuyant sur des *sememes* pour modéliser hiérarchiquement les relations sémantiques entre concepts (synonymie, hyperonymie, partie-tout, rôles agent-patient, *etc.*), intégrant un système de gestion et un dictionnaire sino-anglais de plus de 180 000 entrées, et jouant un rôle majeur dans les applications TALN (annotation sémantique, désambiguïsation, modélisation conceptuelle) comme dans la transition vers une lexicographie computationnelle et ontologique.

4.2. PLATES-FORMES ET OUTILS DE COMPILATION LEXICOGRAPHIQUE ASSISTÉ PAR ORDINATEUR

Une plateforme de compilation de dictionnaires, ou système de compilation de dictionnaires assisté par ordinateur, est désignée à l'international sous le nom de *Dictionary Writing System (DWS)*, parfois aussi appelée *Dictionary Production System (DPS)*. Il s'agit d'un logiciel informatique spécialisé – un éditeur conçu spécifiquement pour la compilation lexicographique. Ce système intègre un corpus lexicographique ainsi que des interfaces permettant l'importation du corpus et l'exportation des données du dictionnaire. Il offre ainsi un ensemble intégré de fonctions couvrant toutes les étapes de la compilation lexicographique : gestion du corpus, compilation, édition, stockage des données et production finale du dictionnaire.

En Occident, grâce à une mise en place plus précoce des corpus linguistiques, plusieurs grandes maisons d'édition de dictionnaires ont développé, dès les années 1970-1980, leurs propres plateformes de compilation et d'édition. Parmi les plus représentatives figurent : le *DPS* développé par la société IDM (France), et le projet Papillon, issu d'une collaboration franco-japonaise. Le *DPS*, développé par la société française IDM (Ingénierie Diffusion Multimédia) fondée en 1994, est une plateforme modulaire et collaborative de compilation lexicographique basée sur le format XML. Conçu initialement en partenariat avec *Longman* et *Oxford University Press*, ce système intègre plusieurs modules – édition, recherche, gestion du flux de travail et relecture – permettant une chaîne complète et centralisée pour la compilation, l'édition et la publication de dictionnaires, qu'ils soient bilingues, thématiques ou encyclopédiques. Compatible avec Unicode et divers formats DTD, il facilite l'importation, l'exportation et la diffusion des données lexicographiques. Son interface ergonomique, fondée sur une structure en quatre zones éditables, offre une navigation fluide et une forte interactivité. Adopté par plus de 500 institutions dans le monde, dont *Longman*, le *DPS* constitue aujourd'hui une référence

internationale en matière de numérisation et d'automatisation de la lexicographie. Le projet *Papillon*, initié par Mathieu Mangeot (NII, Tokyo) et Gilles Sérasset (GETA-CLIPS, Grenoble) parmi d'autres, est une initiative franco-japonaise pionnière visant à construire un dictionnaire numérique multilingue à architecture interlingue (Mangeot, Sérasset, Lafourcade, 2003). Enfin, la plateforme *NooJ* (<https://nooj.univ-fcomte.fr/>, par Silberztein) occupe une place particulière : à la fois environnement logiciel et communauté structurée, elle permet de concevoir des dictionnaires électroniques formalisés (morphologie, syntaxe, sémantique) et d'articuler lexicographie, linguistique de corpus et TAL.

En Chine, la recherche et le développement de plateformes de compilation lexicographique ont commencé dans les années 1990. De nombreux travaux et réalisations techniques ont depuis vu le jour. Parmi les projets les plus représentatifs, on peut citer : 1) CONULEXID, un système bilingue de compilation lexicographique basé sur un corpus, développé conjointement par le Centre de recherche sur les dictionnaires bilingues de l'*Université de Nanjing et la Commercial Press*, dans le cadre de la compilation du *Dictionnaire anglais-chinois de la Nouvelle ère* ; 2) le Système web de compilation lexicographique et de génération automatique basé sur un corpus, fruit de la collaboration entre l'*Université des langues étrangères du Guangdong* et la *Maison d'édition de l'éducation du Guangdong* ; 3) et la Plateforme de compilation lexicographique bilingue basée sur un corpus, développée conjointement par la *Foreign Language Teaching and Research Press* et l'*Institut de linguistique computationnelle de l'Université de Pékin*.

5. ÉCOSYSTÈME ACTUEL DE LA MÉTALEXICOGRAPHIE NUMÉRIQUE

5.1. CENTRES DE RECHERCHE ET LABORATOIRES

La métalexicographie numérique correspond à un domaine de recherche qui étudie les dictionnaires eux-mêmes, leur histoire, leur structure et leur évolution à l'ère du numérique (Chen, 2026, à paraître). La recherche en lexicographie et métalexicographie en France s'appuie sur plusieurs centres et laboratoires de référence, qui jouent un rôle clé dans l'étude, la conception et la valorisation des dictionnaires.

L'ATILF (Analyse et Traitement Informatique de la Langue Française), rattaché à l'Université de Lorraine et au CNRS, est l'un des pôles les plus importants pour la recherche lexicographique et lexicologique numérique. ATILF est l'héritier direct des travaux lexicographiques menés à Nancy depuis la seconde moitié du XX^e siècle au sein du CRTLF puis de l'INaLF.

Parmi les figures majeures du domaine, Jean Pruvost occupe une place centrale. Professeur émérite à CY Cergy Paris Université, il a dirigé le Laboratoire Lexiques,

Textes, Discours, Dictionnaires (LTD2 – Centre Jean Pruvost), spécialisé dans l’histoire des dictionnaires, l’analyse lexicologique et la culture du mot. Il est également à l’origine du groupe Métadif et du Laboratoire LDI (Lexiques, Dictionnaires, Informatique), qui relie la lexicographie traditionnelle à l’analyse informatique des corpus linguistiques.

Le laboratoire *Praxiling* de l’Université Paul-Valéry Montpellier 3 se consacre à l’analyse du discours, à la sémantique et à la linguistique de corpus. Le CRIS-CO (Centre de Recherches Interdisciplinaires sur les Significations et les Contextes) de l’Université de Caen Normandie explore les liens entre lexique, signification et contexte d’usage, en articulant linguistique et traitement des données.

Au-delà de ces pôles, d’autres traditions ont marqué la lexicographie computationnelle en France. D’une part, le LADL (Laboratoire d’Automatique Documentaire et Linguistique), fondé autour des travaux de Maurice Gross, a joué un rôle pionnier dans la formalisation des descriptions lexicales et syntaxiques à grande échelle, notamment à travers les lexiques-grammaires, qui constituent des ressources structurées et directement exploitables pour l’analyse automatique. D’autre part, le LLI (Laboratoire de Linguistique Informatique) s’inscrit dans une dynamique de linguistique outillée et de modélisation lexicale, en particulier dans des perspectives multilingues, utiles à la lexicographie numérique et au traitement automatique des langues.

Ces institutions forment un réseau scientifique solide, reliant tradition lexicographique, humanités numériques et IA, et contribuent à faire de la France un acteur majeur de la recherche internationale sur les dictionnaires et le lexique.

En Chine, la recherche lexicographique s’est structurée autour de nombreux centres spécialisés, souvent en collaboration étroite avec des maisons d’édition universitaires et des institutions linguistiques. Parmi ces structures, plusieurs jouent aujourd’hui un rôle particulièrement actif dans l’innovation lexicographique, la numérisation des ressources linguistiques et le développement de dictionnaires bilingues ou spécialisés : 1) 广东外贸大学词典学研究中心 [Guǎngdōng wàimào dàxué cídiǎn xué yánjiū zhōngxīn, Centre de recherche en lexicographie de l’Université de commerce extérieur du Guangdong]. 2) 南京大学双语词典研究中心 [Nánjīng dàxué shuāngyǔ cídiǎn yánjiū zhōngxīn, Centre de recherche sur les dictionnaires bilingues de l’Université de Nanjing]. 3) 厦门大学双语词典中心 [Xiàmén dàxué shuāngyǔ sōusuǒ zhōngxīn, Centre des dictionnaires bilingues de l’Université de Xiamen]. 4) 四川外国语大学词典学研究所 [Sīchuān wàiguóyǔ dàxué cídiǎn xué yánjiū suǒ, Institut de recherche en lexicographie de l’Université des langues étrangères du Sichuan]. 5) 复旦大学 [Fùdàn dàxué, Université Fudan]. 6) 黑龙江大学俄罗斯语言文学与文化研究中心辞书研究所 [Hēilóngjiāng dàxué èluósī yǔyán wénxué yǔ wénhuà yánjiū zhōngxīn císhū yánjiū suǒ, Institut de recherche lexicographique du Centre d’études russes de l’Université du Heilongjiang].

Ces centres constituent aujourd’hui le moteur de la lexicographie numérique chinoise, en combinant recherche universitaire, développement de corpus, intelligence

artificielle et innovations éditoriales. Ils contribuent activement à la modernisation des pratiques lexicographiques et au rayonnement international de la recherche linguistique en Chine.

5.2. ORGANISMES ET ASSOCIATIONS PROFESSIONNELLES

Les activités scientifiques liées à la lexicographie et à la métalexigraphie s'organisent autour de plusieurs événements et réseaux internationaux majeurs. Depuis 1994, la *Journée des dictionnaires* constitue un rendez-vous incontournable pour les chercheurs, lexicographes, enseignants et amateurs de la langue française. Fondée par Jean Pruvost au sein du laboratoire LTD2 à CY Cergy Paris Université, cette journée offre un espace de réflexion et d'échanges sur les dictionnaires d'hier et d'aujourd'hui, en explorant leurs évolutions, leurs usages et leurs enjeux contemporains. En 2021, la création du réseau international de métalexigraphie (METALEX) a marqué une nouvelle étape dans la structuration de la recherche lexicographique au niveau mondial. Ce réseau, présidé par Christophe Rey et vice-présidé par Jean Pruvost, favorise la collaboration entre chercheurs issus de différents horizons disciplinaires et linguistiques, tout en promouvant le développement de projets communs autour de la lexicographie numérique. La recherche lexicographique française est caractérisée par une forte dimension internationale : des lexicographes de Grande-Bretagne, d'Allemagne, d'Espagne, et d'autres pays publient régulièrement dans des revues françaises. Les collaborations internationales sont fréquentes, comme le montrent divers colloques, tels que le colloque de 2019 organisé par Do-Hurinvillle *et al.*, ou celui de 2024 organisé par Pruvost et C. Rey, réunissant des experts de toute l'Europe et au-delà. Enfin, la communauté scientifique participe activement aux conférences internationales TOTH (Terminology & Ontology : Theories and Applications), organisées chaque année à l'Université Savoie Mont-Blanc et à l'Université de Crète. Ces colloques rassemblent des spécialistes de la terminologie, de l'ontologie et du traitement automatique du langage, et constituent un lieu privilégié pour les échanges entre linguistique, IA et humanités numériques. En somme, ces initiatives traduisent une dynamique scientifique forte, ancrée à la fois dans la tradition lexicographique et dans les innovations du web sémantique et des données ouvertes.

En Chine, la tradition lexicographique moderne s'est institutionnalisée dès les années 1980 avec la création de plusieurs associations régionales et nationales dédiées à l'étude et à la compilation des dictionnaires. Parmi les premières figurent l'*Association de Shanghai pour la lexicographie* (1982), l'*Association lexicographique de la province du Shanxi* (1985), l'*Association lexicographique de la province du Fujian* (1989) et l'*Association lexicographique de la province d'Anhui* (1991). À l'échelle nationale, la fondation de la *China Association for Lexicography* en 1992 a marqué une étape décisive dans la coordination et la modernisation des travaux

lexicographiques. Ces institutions ont contribué à la standardisation des méthodes, à la création de dictionnaires régionaux et spécialisés, ainsi qu'à la transition vers la lexicographie numérique, intégrant progressivement les technologies informatiques et les bases de données dans la recherche lexicale et la production de ressources linguistiques.

La *China Association for Lexicography* a été fondée en septembre 1992 à l'Université du Hubei, avant de tenir son assemblée inaugurale à Pékin en octobre de la même année. Son secrétariat général est établi à la Commercial Press, la plus ancienne maison d'édition chinoise spécialisée dans la lexicographie. L'association comprend un comité académique et huit comités spécialisés, couvrant les principaux domaines de la discipline : dictionnaires de langue, dictionnaires spécialisés, dictionnaires bilingues, encyclopédies, dictionnaires des minorités, théorie et histoire de la lexicographie, technologies modernes de compilation, ainsi que l'édition et la publication des dictionnaires. Depuis sa création, l'Association s'attache à fédérer les chercheurs et les professionnels de la lexicographie à l'échelle nationale, à promouvoir les échanges scientifiques et à encourager l'innovation dans la compilation et la recherche sur les dictionnaires. Elle organise régulièrement des colloques nationaux et internationaux, facilite la coopération entre la recherche lexicographique et l'édition, et soutient la formation de jeunes chercheurs à travers des séminaires et des programmes de mentorat. Elle joue également un rôle actif dans la normalisation de la production lexicographique, la mise en place de standards éditoriaux et la sélection des ouvrages d'excellence. Grâce à ses activités de recherche, de formation et de diffusion, la China Association for Lexicography occupe aujourd'hui une position centrale dans le développement de la lexicographie contemporaine chinoise, alliant héritage scientifique, innovation technologique et ouverture internationale.

5.3. REVUES SCIENTIFIQUES ET PUBLICATIONS SPÉCIALISÉES

La recherche en lexicographie s'appuie sur plusieurs revues scientifiques spécialisées, qui jouent un rôle essentiel dans la diffusion des travaux théoriques et appliqués du domaine.

Parmi elles, en France, la revue ÉLA – Études de Linguistique Appliquée, dirigée actuellement par Jean Pruvost, occupe une place importante dans le paysage francophone. Elle couvre un large éventail de thématiques allant de la didactologie des langues-cultures à la lexiculture, tout en intégrant des études sur la lexicographie et la sémantique. Les *Cahiers de lexicologie*, créés en 1959 par Bernard Quemada, constituent une revue de référence dans le domaine. Publiée aujourd'hui par Classiques Garnier, elle explore les dimensions linguistiques, historiques et théoriques du lexique et de la lexicographie. Sous la direction de Christine Jacquet-Pfau et Alain Polguère, cette revue contribue à la réflexion scientifique sur la structuration du vocabulaire,

les dictionnaires et la description des langues. Enfin, la revue *Les Cahiers du dictionnaire*, également publiée par Classiques Garnier, s'intéresse plus particulièrement aux rapports entre dictionnaire et patrimoine linguistique, ainsi qu'aux discours du dictionnaire et des dictionnaires du discours. Dirigée par Giovanni Dotoli, elle réunit des chercheurs issus de plusieurs traditions linguistiques et culturelles, et constitue un espace de dialogue entre lexicographie, lexicologie et analyse du discours.

Ces revues témoignent du dynamisme scientifique de la lexicographie francophone, en reliant tradition académique et ouverture internationale, et en favorisant la circulation des savoirs entre chercheurs, enseignants et praticiens du langage.

En Chine, la lexicographie numérique connaît un essor remarquable, soutenu par des institutions académiques, des maisons d'édition spécialisées et des revues de référence. Parmi celles-ci, la revue 辞书研究 [*Císhū yánjiū*, *Lexicographical Studies*] occupe une place centrale dans la recherche théorique et appliquée sur la lexicographie contemporaine. D'autres périodiques influents tels que le 中国辞书学报 [*Zhōngguó císhū xuébào*, *Chinese Journal of Dictionary*], 中国语文 [*Zhōngguó yǔwén*, *Langue chinoise*] et le 语言战略研究 [*Yǔyán zhànlüè yánjiū*, *Chinese Journal of Language Policy and Planning*] publient régulièrement des travaux sur la numérisation des dictionnaires, les nouvelles méthodes de compilation et l'intégration de l'intelligence artificielle dans la production lexicographique. En Chine, la lexicographie numérique s'appuie sur un réseau de revues académiques et de maisons d'édition spécialisées qui jouent un rôle essentiel dans la modernisation de la discipline.

5.4. INTELLIGENCE ARTIFICIELLE, TRANSFORMATION DES PRATIQUES ET COOPÉRATIONS INTERNATIONALES

L'évolution de la lexicographie s'inscrit désormais à la croisée des humanités numériques et du traitement automatique du langage naturel (NLP), deux domaines complémentaires qui transforment en profondeur la manière de concevoir, produire et diffuser les dictionnaires.

Du côté des Humanités numériques, les efforts se concentrent sur la numérisation et l'archivage à long terme des ressources lexicographiques, la curation de contenus et la promotion du mouvement Open Science, garantissant un accès libre et durable aux données. Les projets s'appuient sur des corpus linguistiques normalisés (UNICODE), des outils collaboratifs et des démarches participatives de documentation des langues. De nouveaux instruments sont également développés pour assister les lexicographes dans leurs tâches d'analyse, de gestion et de mise à jour des bases de données lexicales. Ces initiatives impliquent non seulement des chercheurs, mais aussi des linguistes et amateurs de langue, favorisant une dynamique collective et ouverte. Du côté du Traitement automatique du langage (NLP), la lexicographie bénéficie de technologies avancées telles que la reconnaissance optique de caractères (OCR), la

reconnaissance d'écriture manuscrite (HTR) et la production automatisée de définitions avec post-édition. L'intégration du Big Data, des systèmes de gestion de données et des Linguistic Linked Open Data (LLOD) permet d'interconnecter les ressources linguistiques à une échelle mondiale. Ces avancées sont menées par des data scientists, des informaticiens et des spécialistes du langage, qui développent de nouvelles technologies pour relier et enrichir les données lexicales.

Ainsi, la lexicographie académique en France s'oriente principalement vers les Humanités numériques, tout en intégrant progressivement les apports du NLP pour favoriser une approche hybride, à la fois collaborative, technologique et interdisciplinaire.

La Chine modernise rapidement sa lexicographie grâce aux technologies d'intelligence artificielle. Deux projets phares illustrent cette évolution : 1) le « Dictionnaire multilingue et corpus de langues étrangères chinoises » de l'Université des études étrangères de Pékin, un projet quinquennal réunissant 150 chercheurs pour élaborer un dictionnaire multilingue (anglais, arabe, russe, français, coréen, *etc.*) ; et 2) la « Plateforme d'emprunts européens et américains » de l'Université des études étrangères du Guangdong, qui utilise de grands modèles linguistiques pour générer 30 000 entrées d'emprunts – chacune vérifiée par des éditeurs – démontrant ainsi que l'IA peut considérablement accélérer le travail lexicographique tout en maintenant la nécessité d'une validation humaine. Ces initiatives montrent que les grands modèles linguistiques vont profondément remodeler la lexicographie chinoise (et au-delà), tout en soulevant de nouvelles questions liées au contrôle qualité et à la propriété des données, nécessitant une collaboration étroite entre lexicographes et spécialistes de l'IA.

La Chine accorde une attention particulière au développement des dictionnaires à l'ère de l'IA. En 2024, les universités en Chine et en France, en collaboration avec *China Association for Lexicography*, a organisé le colloque international *Lexicographie et services de connaissances à l'ère du CGIA*, réunissant des experts internationaux provenant de plus de dix pays d'Europe et d'Asie. Cet événement s'inscrivait dans le cadre du projet international financé par le Ministère chinois des Sciences et des Technologies, visant à analyser les trajectoires d'innovation en lexicographie numérique en Chine et en Europe au cours des vingt dernières années et à renforcer les échanges scientifiques en invitant des spécialistes européens. Les éditeurs commerciaux chinois s'adaptent également activement à l'ère de l'IA. Lors de ce colloque en 2024, un forum spécial leur a été consacré. Des représentants de Shanghai Century Publishing (Group) Co., Ltd., de The Commercial Press Ltd., de Foreign Language Teaching & Research Press, de Shanghai Foreign Language Education Press et d'autres maisons d'édition ont présenté leurs stratégies d'intégration des dictionnaires aux technologies d'intelligence artificielle et aux grands modèles linguistiques.

Cette initiative s'articule également avec la série internationale de séminaires *International in Computational Lexicology and Lexicography*, favorisant un dialogue régulier entre chercheurs européens et chinois autour de la lexicologie et lexicographie

numériques. Dans la continuité de cette dynamique, un second projet international, financé par le Programme national pour les experts en innovation scientifique et technologique de Chine, porte sur les innovations locales et les avancées internationales en lexicographie intelligente. Dans ce cadre, la conférence internationale NPCL² 2026 – New Perspectives in Computational Lexicology and Lexicography : From Data to Dictionaries est en préparation et réunira des experts de Chine, de France, d'Italie, du Portugal, du Canada, entre autres, afin d'instaurer un dialogue direct et productif entre informaticiens et spécialistes en lexicographie pour répondre aux défis scientifiques et technologiques émergents. Ces initiatives témoignent du rôle croissant de la Chine dans la recherche mondiale en lexicographie numérique et dans le développement des dictionnaires intelligents, ainsi que du dynamisme des coopérations sino-européennes dans ce domaine stratégique.

6. CONCLUSION

La comparaison entre la France et la Chine montre que, malgré des trajectoires historiques et institutionnelles distinctes, les deux pays convergent aujourd'hui vers une même dynamique : la transformation profonde de la lexicographie à l'ère du numérique et de l'intelligence artificielle. En France, l'évolution s'est appuyée sur un héritage lexicographique ancien, le développement précoce des corpus, la structuration de ressources ouvertes et l'intégration progressive des normes du web sémantique, dans une logique scientifique centrée sur la qualité, la modélisation et l'interopérabilité. En Chine, la modernisation lexicographique s'est accélérée depuis les années 1980 grâce à la constitution de grands corpus, à la création de bases sémantiques innovantes telles que *HowNet*, et, plus récemment, à l'impulsion décisive de l'IA générative et des grands modèles de langage dans la production de ressources bilingues et multilingues.

Ces évolutions mettent en lumière des complémentarités : une tradition française fondée sur la rigueur philologique et l'ouverture scientifique, face à une innovation chinoise marquée par l'intégration rapide des technologies avancées et par l'orientation vers des usages éducatifs, multilingues et industriels. Dans les deux contextes, la lexicographie se repositionne au croisement des humanités numériques, des sciences du langage et de l'intelligence artificielle, posant des enjeux nouveaux : qualité et contrôle humain, standards de structuration des données, droit d'auteur, diffusion ouverte, et formation de nouvelles compétences hybrides.

Ainsi, la lexicographie contemporaine n'est plus seulement un travail d'inventaire et de description du lexique : elle devient un domaine interdisciplinaire stratégique, où se rencontrent savoir linguistique, technologies de traitement automatique du langage, ingénierie des connaissances et humanités numériques. Les perspectives futures reposent sur la collaboration internationale, l'interopérabilité des données lexicales, et l'articulation entre intelligence humaine et intelligence artificielle, conditions

indispensables pour construire des dictionnaires vivants, connectés, multilingues et adaptés aux besoins des sociétés du XXI^e siècle.

La transition vers la lexicographie numérique bouleverse profondément le rôle des dictionnaires à l'ère de l'IA. Comme le souligne de Schryver, l'IA exploite massivement ces ressources pour l'apprentissage automatique, la traduction ou la recherche, obligeant les lexicographes à collaborer avec ces entreprises sous peine de devenir « irrelevants » (Lew, 2023).

L'arrivée des grands modèles de langage transforme en profondeur les pratiques : alors que les travaux fondateurs défendaient un modèle de post-édition, dans lequel « la machine produit un brouillon et l'humain le révise » (Rundell & Kilgariff, 2011), les LLM semblent désormais capables de « contourner les étapes techniques traditionnelles » et d'approcher une forme de génération lexicographique quasi directe (Rundell, 2023). Les perspectives discutées dans les colloques *ELEX 2025* et *ASIALEX 2025* confirment cette transition vers une lexicographie automatisée-assistée, combinant intelligence computationnelle, supervision humaine et modélisation explicable.

Mais ces avancées révèlent aussi des défis critiques : dépendance au *prompt engineering*, risques d'erreurs factuelles, hallucinations sémantiques, biais linguistiques, gestion des contenus sensibles, et nécessité d'un contrôle éditorial rigoureux. Les orientations identifiées incluent : (1) modèles spécialisés (« LexGPT »), (2) *fine-tuning* sur des corpus lexicographiques, (3) traçabilité et explicabilité, (4) systèmes FAIR et interopérables, (5) capitalisation du retour utilisateur. Comme le soulignent De Schryver & Joffe (2023), l'IA peut déjà « take over some of the most laborious aspects » du travail lexicographique, mais doit encore progresser dans la détection automatique des sens, la gestion des exemples et la validation par les usagers. La tendance observée est donc celle d'un flux de travail hybride, où les modèles produisent des esquisses d'articles – un processus que Rundell (2023) qualifie de « post-édition lexicographique » désormais réalisable – et où l'expert redirige ses compétences vers la validation, la normalisation et le contrôle de la qualité.

Enfin, ces mutations invitent à reconsidérer le statut théorique du domaine. Si Landau (2001) soutient que « lexicography is first and foremost an Art & Craft » – et qu'il « n'existe aucun besoin » d'une théorie unifiée – l'avènement de l'IA ne réduit pas cette dimension artisanale ; il la reconfigure. Le lexicographe devient chef d'orchestre, garant de la qualité, de la cohérence et de l'éthique, face à des systèmes capables de dévorer les données lexicales et de reformuler les savoirs linguistiques à grande vitesse. L'avenir se joue donc dans la coopération avancée homme-machine, l'interconnexion des ressources, la transparence méthodologique et l'adaptation à des usages multimodaux, pédagogiques, industriels et mobiles. La lexicographie se positionne durablement à l'intersection des humanités numériques, du *Big Data*, de la linguistique computationnelle et de l'ingénierie cognitive – au service de dictionnaires intelligents, évolutifs et scientifiquement responsables.

BIBLIOGRAPHIE

- Aquino-Weber, D. & Greub, Y. (éds.). (2020). *La lexicographie informatisée : les vocabulaires nationaux dans un contexte européen*. Berne : Académie suisse des sciences humaines et sociales. <https://doi.org/10.5281/zenodo.3550494>
- Baisa, V. *et al.* (2019). Automating dictionary production: a Tagalog – English – Korean dictionary from scratch. In *Electronic Lexicography in the 21st Century. eLex Proceedings* (pp. 805-818). Brno : Lexical Computing CZ.
- Béjoint, H. (1994). Langue populaire américaine et dictionnaires : quelques remarques sociolexicographiques. *Revue française d'études américaines*, 60, 117-124.
- Béjoint, H. (2007). Informatique et lexicographie de corpus : les nouveaux dictionnaires. *Revue française de linguistique appliquée*, XII (1), 7-23.
- Boguraev, B. & Briscoe, T. (1989). *Computational lexicography for natural language processing*. London : Longman.
- Bosque-Gil, J. & Gracia, J. (2019). The OntoLex-Lemon lexicography module. *W3C Community Group Final Report*. <https://www.w3.org/2019/09/lexicog/>
- Chen, L. (2021). *Analyse comparative des expressions idiomatiques en chinois et en français (relatives au corps humain et aux animaux)* [thèse de doctorat]. Cergy Paris Université.
- Chen, L. (2026, à paraître). *Nouvelles réflexions et perspectives sur la métalexigraphie contemporaine : une analyse contrastive entre le chinois, le français et l'anglais*. Paris : Honoré Champion, collection Lexica.
- De Schryver, G.-M. (2023a). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, 36 (4), 1-33. <https://doi.org/10.1093/ijl/ecad021>
- De Schryver, G.-M. (2023b). Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography*, 16 (2), 143-199. https://www.researchgate.net/publication/31424125_Lexicographers%27_Dreams_in_the_Electronic-Dictionary_Age
- De Schryver, G.-M. & Joffe, D. (2023). The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks. *20th CODH Seminar*, Tokyo. <http://codh.rois.ac.jp/seminar/lexicography-chatgpt-20230227/>
- Feng, Z.W. 冯志伟. (2001). 计算语言学基础 [Jisuan yǔyán xué jīchǔ, *Fondements de la linguistique informatique*]. Pékin : Commercial Press.
- Feng, Z.W. 冯志伟. (2011). 计算语言学的历史回顾与现状分析 [Jisuan yǔyán xué de lìshǐ huíguò yǔ xiànzhuàng fēnxī, *Rétrospective historique et état actuel de la linguistique computationnelle*]. *Lan-gues étrangères (Journal of Shanghai International Studies University)*, 1, 91-97.
- Gasiglia, N. (2020). Étude généalogique d'articles dictionnaires affichés par Google. *Academic Journal of Modern Philology*, 9, 91-105. https://ajmp.uwr.edu.pl/wp-content/uploads/sites/39/2020/12/J_8_Gasiglia.pdf
- Gasiglia, N. (2021). Identification des articles dictionnaires non référencés fournis par google.fr. *Éla. Études de linguistique appliquée*, 201, 63-89. <https://doi.org/10.3917/ela.201.0064>
- Gottlieb, H., Mogensen, J.E. & Zettersten, A. (2005). *Symposium on lexicography XI: Proceedings of the eleventh international symposium on lexicography, May 2-4, 2002, University of Copenhagen*. Lexicographica Series Maior 115. Berlin : De Gruyter. <https://doi.org/10.1515/9783110928310>
- Jacquet-Pfau, J. (2005). Pour un nouveau dictionnaire informatisé. *Éla. Études de linguistique appliquée*, 137 (1), 51-71.
- Jakubiček, M. & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? In *Electronic Lexicography in the 21st Century (eLex 2023) Proceedings* (pp. 508-523). Brno : Lexical Computing CZ.
- Kilgariff, A. & Pavel, R. (2010). Semi-automatic dictionary drafting. In G.-M. De Schryver (éd.), *A way with words: Recent advances in lexical theory and analysis. A festschrift for Patrick Hanks* (pp. 299-312). Kampala : Menha Publishers.

- Landau, S.I. (2001). *Dictionaries: The art and craft of lexicography*. Cambridge : Cambridge University Press.
- Lew, R. (2015). Research into the use of online dictionaries. *International Journal of Lexicography*, 28 (2), 232-253. <https://doi.org/10.1093/ijl/ecv010>
- Lew, R. (2023). ChatGPT as a COBUILD lexicographer. *Humanities and Social Sciences Communications*, 10, 1-10. <https://www.nature.com/articles/s41599-023-02119-6>
- Lin, X. 林杏光. (1999). 词汇语义和计算语言学 [Cíhuì yǔyì hé jìsuàn yǔyán xué, Sémantique lexicale et linguistique computationnelle]. Pékin : Éditions Yuwen.
- Mangeot, M., Sèrasset, G. & Lafourcade, M. (2003). Construction collaborative d'une base lexicale multilingue : le projet Papillon. *Traitement automatique des langues*. <https://hal.science/hal-00340723v1>
- Martin, R. (2008). Perspectives de la lexicographie informatisée. In J. Durand, B. Habert & B. Laks (éds.), *Congrès Mondial de Linguistique Française – CMLF'08* (pp. 1251-1256). Paris : Institut de Linguistique Française. <https://doi.org/10.1051/cmlf08332>
- Merken, S. (2023). New York lawyers sanctioned for using fake ChatGPT cases in legal brief. *Reuters*. <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/>
- Meyer, C.M. (2018). Kollaborative Lexikografie: Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze. In S. Engelberg *et al.* (éds.), *Wortschätze: Dynamik, Muster, Komplexität* (pp. 293-310). Berlin/Boston : De Gruyter. <http://doi.org/10.1515/9783110579963-016>
- Molinari, C. (2021). La lexicographie numérique : un nouvel outil pour l'enseignement du FLE ? *Éla. Études de linguistique appliquée*, 201 (1), 27-47. <https://doi.org/10.3917/ela.201.0028>
- Molinari, C. & Vincent, N. (éds.). (2021). *Dictionnaires et culture numérique dans l'espace francophone 2. Perspectives pragmatiques de la description numérique du français*. Paris : Didier-Érudition / Klincksieck.
- Murano, M. (2019). Les dictionnaires collaboratifs et la phraséologie : de la description du patrimoine existant à l'invention de nouvelles séquences figées. *Éla. Études de linguistique appliquée*, 194, 193-209.
- Müller-Spitzer, C. (éd.) (2014). *Using online dictionaries*. Berlin & Boston : De Gruyter.
- Oger, C. (2021). *Faire référence. La construction de l'autorité dans le discours des institutions*. Paris : EHESS.
- Pentzold, C. (2011). Imagining the Wikipedia community: What do Wikipedia authors mean when they write about their 'community'? *New Media & Society*, 13 (5), 704–721. <https://doi.org/10.1177/1461444810378364>
- Pierrel, J.-M. (2010). À la découverte d'un trésor : le mariage de l'informatique et de la lexicographie au service de la valorisation de la langue française. *Revue française de linguistique appliquée*, XV (1), 103-118.
- Polguère, A. (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. In U. Heid *et al.* (éds.), *Proceedings of the 9th EURALEX International Congress* (pp. 517-527). Stuttgart : Institut für Maschinelle Sprachverarbeitung.
- Polguère, A. (2009). Lexical systems: graph models of natural language lexicons. *Language Resources and Evaluation*, 43 (1), 41-55. <https://link.springer.com/article/10.1007/s10579-008-9078-4>
- Polguère, A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, 27 (4), 396-418. <https://doi.org/10.1093/ijl/ecu017>
- Pruvost, J. (2000). *Dictionnaires et nouvelles technologies*. Paris : PUF.
- Pruvost, J. (2004). Des dictionnaires de langue française et des nouvelles technologies d'hier à aujourd'hui. In F. Argod-Dutard (dir.), *Quelles perspectives pour la langue française ?* (pp. 137-172). Bordeaux : Presses universitaires de Bordeaux.
- Rundell, M. & Kilgariff, A. (2011). Automating the creation of dictionaries: where will it all end? In F. Meunier *et al.* (éds.), *A taste for corpora. A tribute to Sylviane Granger* (pp. 257-281). Amsterdam & Philadelphia : John Benjamins.
- Zhang, Y. 章宜华. (2013). 计算词典学 [Jìsuàn cídiǎn xué, Computational Lexicography]. Pékin : Commercial Press, coll. « Bibliothèque des études lexicographiques ».