HAL
open science

Les corpus numériques en didactique du FLE en Chine

Agnès Pernet-Liu, Jia Tan

To cite this version:

Agnès Pernet-Liu, Jia Tan. Les corpus numériques en didactique du FLE en Chine. Etudes en didactique des langues, 2018, Les corpus / Corpora, 31, pp.53-70. hal-04034780

HAL Id: hal-04034780

https://ut3-toulouseinp.hal.science/hal-04034780

Submitted on 17 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les corpus numériques en didactique du FLE en Chine

Agnès PERNET-LIU &Jia TAN
MCF et Lectrice
Université des Langues étrangères de Pékin (Chine)

Introduction

Le développement des corpus numériques en Chine et de leur lien avec la didactique des langues présente de nombreuses similitudes avec l'histoire des corpus et de l'apprentissage sur corpus dans les domaines anglo-saxon et francophone. C'est le signe d'une évolution de l'espace académique chinois en synergie avec la communauté universitaire internationale des linguistes et des didacticiens avec leurs usages propres et novateurs des nouvelles technologies et les problématiques inédites qui sont ainsi suscitées.

Nous faisons le point sur la situation des corpus numériques dans la didactique des langues en Chine, pour le chinois, pour l'anglais et pour le français, afin de faire apparaître des questionnements transversaux, communs aux différents contextes culturels et aux différentes langues enseignées, concernant l'utilisation didactique des corpus tout en prenant en compte des aspects spécifiques à la Chine, avec sa culture d'apprentissage et son environnement universitaire pour l'enseignement/apprentissage des langues.

Nous mettons l'accent sur les corpus écrits et sur le français langue étrangère et nous allons essayer de montrer que l'utilisation des corpus numériques en didactique du FLE dans le contexte universitaire chinois présente un certain intérêt à condition qu'une méthodologie adaptée au public des apprenants chinois de français en oriente l'utilisation pour que cet outil soit concrètement utilisable sur le terrain chinois de l'enseignement-apprentissage.

LES CORPUS NUMÉRIQUES ET LA DIDACTIQUE DES LANGUES EN CHINE

La place des corpus numériques en Chine, en chinois

On peut faire remonter la création des corpus numériques en chinois à la fin des années 80. Ils ont connu un développement rapide dans les années 90 en Chine. Les corpus en chinois créés depuis lors peuvent se diviser en deux

catégories: corpus en langue chinoise à usage courant (corpus généraux) et corpus en langue chinoise à usage spécialisé (corpus spécialisés).

.

Les corpus généraux

Les trois corpus en langue chinoise utilisés de la façon la plus fréquente (Tan, 2014: 27) sont les suivants.

Le corpus du CCL (Centre for Chinese Linguistics PKU), élaboré par le centre chinois de recherches en linguistique de l'Université de Pékin, est composé de trois sous-corpus: chinois contemporain, chinois classique et corpus bilingue chinois-anglais. Le sous-corpus chinois contemporain rassemble 1 696 ouvrages, permettant de formuler des requêtes en utilisant 9 711 caractères chinois, et le sous-corpus chinois classique comprend des textes datant de la dynastie Zhou (-1045 à -256) jusqu'à la République de Chine (1921-1949), permettant de formuler des requêtes en utilisant 22 580 caractères chinois. Ces deux sous-corpus contiennent en tout 477 millions de mots (Zheng, 2014: 163). N'ayant pas fait l'objet de découpage lexical, ces deux sous-corpus ne permettent que des types de requête très simples. Quant au troisième sous-corpus, il s'agit d'un corpus parallèle entre le chinois et l'anglais dont le nombre de couples phrastiques dépasse un million. Ce sous-corpus sert de base tant pour les analyses comparatives entre le chinois et l'anglais que pour la traduction automatique entre ces deux langues. L'Université de Pékin s'en est servie pour développer une plate-forme en vue de la rédaction d'un dictionnaire bilingue. Il faut noter que ce sous-corpus est réservé aux usagers au sein de l'Université.

Dans le corpus Cncorpus, créé par la Commission nationale des langues, les usagers peuvent trouver des textes authentiques du chinois contemporain (vingt millions de mots) et du chinois classique (cent millions de mots).

Quant au corpus PRF (corpus annoté du *Quotidien du Peuple*), c'est un corpus annoté en chinois rassemblant des articles de ce journal, créé conjointement par le Centre chinois de recherches en linguistique de l'Université de Pékin, l'entreprise Fujitsu et *Le Quotidien du Peuple*. Le corpus rassemble les articles publiés au cours de l'année 1998 et de l'année 2000. Ce corpus annoté a fait l'objet d'un découpage lexical et sert pour établir la liste des mots chinois à haute fréquence et étudier de façon quantitative les fonctions grammaticales des mots.

Les corpus spécialisés

Les corpus d'interlangue en chinois

Le corpus de production en chinois langue étrangère du test national HSK, développé par l'Université des langues et cultures de Pékin, comprend les productions écrites en chinois de 11 569 apprenants de chinois inscrits au test d'évaluation du chinois HSK, avec un nombre de caractères chinois qui atteint 4,24 millions.

Des corpus numériques d'interlangue de moindre envergure (Tan, 2014) ont aussi été créés, tels que le corpus de production orale en chinois des étudiants

étrangers et le corpus de production écrite en chinois des apprenants américains de chinois. Il s'agit de deux corpus d'interlangue longitudinaux, autrement dit, ils recueillent les différentes productions (orales ou écrites) de chaque apprenant à différentes étapes de son apprentissage. Ils ont tous les deux été créés par l'Université de Nankin. Ces corpus d'interlangue ont été créés pour fournir des données permettant des recherches sur la didactique du chinois (langue étrangère (LÉ) et sur l'interlangue des apprenants de chinois.

Les corpus en chinois classique

Depuis une dizaine d'années, bien des corpus en chinois classique ont été créés, comme le corpus de *Songci*, annoté, qui permet des études sur le style et le rythme des poèmes de la dynastie des Song, ou encore le corpus de documents de médecine traditionnelle en chinois classique qui, en numérisant les livres anciens, a favorisé l'informatisation dans le domaine de la médecine traditionnelle chinoise.

Les corpus phonétiques en chinois

Par rapport aux corpus de textes, les corpus phonétiques se développent très lentement vu les difficultés dans l'acquisition et l'annotation des données. Au cours des dix dernières années, seuls trois travaux de recherche ont abordé les corpus phonétiques en chinois (Tan, 2014). Il y a aussi des projets de créer des corpus phonétiques de langues régionales.

Tendances et problèmes

Avec le développement des corpus en langue chinoise, Tan (2014) voit dans les recherches en la matière trois tendances: primo, la priorité n'est plus au corpus général mais au corpus spécialisé; secundo, les corpus textuels tendent à la maturité alors que les corpus oraux avancent toujours avec tâtonnement; tertio, avec le développement de l'enseignement du chinois, les corpus d'interlangue suscitent de plus en plus d'intérêt.

Étant donné le développement de la linguistique de corpus à l'étranger et l'influence croissante de l'anglais, une grande importance est accordée en Chine aux corpus d'anglais alors qu'il y a beaucoup moins de corpus consacrés aux recherches sur la langue chinoise.

Faute de critères et de contrôle, il existe des problèmes dans l'élaboration des corpus en chinois: d'un côté, on crée souvent des corpus pour eux-mêmes et donc il y a des répétitions dans l'élaboration des corpus par manque de système de partage; de l'autre côté, peu de corpus ont fait l'objet de découpage et d'annotation, d'où un taux de consultation très faible.

L'exploitation des corpus numériques en langues étrangères

Les travaux sur les corpus numériques en langues étrangères débutent vers le début des années 80 en Chine, et peuvent se diviser en trois catégories qui correspondent à trois périodes: la rédaction de dictionnaires spécialisés, la

description et le diagnostic concernant l'interlangue, et les pratiques didactiques en cours de LÉ (Xu & Xu, 2017: 62-64).

Les corpus de sciences et technologies en LÉ donnant naissance à des dictionnaires techniques

JDEST (Jiao Da English for Science and Technology), créé par Yang Huizhong dans les années 80, est le premier corpus d'anglais de Chine et d'Asie et a été considéré par John Sinclair et Geoffrey Leech comme "le pionnier dans l'élaboration de corpus en Orient" (Zhen, 2010: 36).

Depuis, des établissements d'enseignement supérieur comme l'Université Tsinghua et l'Institut technique de Dalian ont élaboré des corpus de textes en sciences et technologies en LÉ, lesquels ont donné naissance à une série de dictionnaires de sciences et techniques en langues étrangères (anglais, russe, allemand et japonais) mais pas en français.

À partir des années 1990, beaucoup de disciplines (pétrole, navigation, agriculture, médecine, affaires commerciales) ont construit leurs propres corpus, essentiellement en anglais (Xu & Xu, 2017: 63). L'Université d'aérospatiale de Pékin a pris l'initiative de créer un corpus d'anglais académique dont le nom est BeijingCARE: Beijing Collection of Academic Research Essays. Des établissements d'enseignement supérieur ont élaboré aussi des corpus parallèles entre anglais et chinois, tels que le Corpus parallèle (anglais et chinois) des affaires militaires de l'Institut des langues étrangères de l'Armée populaire de libération, le Corpus parallèle chinois-anglais des lois et des règlements de Chine de l'Université de Shaoxing et le Nouveau corpus bilingue du tourisme de l'Université polytechnique de Hongkong.

Les recherches sur l'interlangue à partir des corpus de productions en anglais des apprenants chinois

À partir des années 90, la Chine s'est lancée dans l'exploitation et la création des corpus d'interlangue, en suivant le modèle de l'International Corpus of Learner English (ICLE), fondé par Sylviane Granger. Une série de corpus d'interlangue ont été élaborés, dont les plus importants sont présentés par Xu et Xu (2017: 63-64) de la manière suivante:

Sigle	**Nom complet**	**Contenu**
CLEC	Chinese Learner English Corpus	Productions écrites des tests d'anglais (niveaux 4 et 6) ainsi que celles d'étudiants de départements d'anglais, dont le nombre total de mots dépasse un million, avec annotation d'erreurs classées en onze catégories.
COLSEC	College Learners' Spoken English Corpus	Documents audio de productions orales d'étudiants dans les tests d'anglais (niveaux 4 et 6).
SWECCL	Spoken and Written English Corpus of Chinese Learners	Productions écrites et orales d'étudiants de départements d'anglais.
PACCEL	Parallel Corpus of Chinese EDL learners	Productions dans l'interprétation et la traduction d'étudiants de départements d'anglais.

CEM	Corpus for English Majors	Productions écrites et traductions d'étudiants de départements d'anglais.
SECOPETS	Spoken English Corpus of Public English Test System	Productions orales d'apprenants d'anglais de tous niveaux.
TECCL	Ten-thousand English Compositions of Chinese learners	10 000 compositions d'élèves de différents niveaux (école primaire, école secondaire et enseignement supérieur).

Tableau 1 – **Principaux corpus d'interlangue**

Il y a aussi des corpus d'anglais des apprenants chinois de l'école secondaire dont les plus connus sont MSEE, MSSW, MSSS et TSLC.

Ces différents corpus d'interlangue en anglais ont fait l'objet de nombreux travaux de recherches (Tang & He, 2004: 44) qui s'appuient essentiellement sur CEA (*Computer-aided error analysis*) et CIA (*Constrastive interlanguage analysis*).

L'utilisation de la linguistique de corpus en didactique des langues

À partir de 1995, des intellectuels ont proposé d'intégrer l'utilisation des corpus dans la pratique didactique. En Chine, Li & Pu (2001: 20) ont présenté en 2001 la méthode DDL (*data-driven learning*), proposée initialement par Johns (1991: 2) en 1991, laquelle consiste à inviter les apprenants à découvrir eux-mêmes les règles linguistiques à travers l'observation, la description et l'induction de quantité de données réelles issues de corpus. Voici un exemple d'application concrète: en 2009, Bai de l'École normale de Shenyang a développé, avec son équipe, l'apprentissage autonome des apprenants du département d'anglais de niveau avancé en se servant des corpus numériques et de la plateforme Wiki (Zhen, 2010: 37).

En 2017, Zhang a effectué une étude sur les travaux de recherche portant sur les corpus en Chine. Sur le site des publications académiques de Chine (CNKI), il a lancé une requête "corpus" et a ainsi recensé 3 128 articles publiés entre 1992 et 2015. À la suite d'une recherche plus détaillée en classant les articles en différentes catégories, on peut voir que sur les quelque 3 000 articles, seuls 581 concernent l'enseignement des langues, soit moins de 20%.

L'utilisation des corpus dans l'enseignement des LÉ se limite, dans la majorité des cas, à l'anglais et se concrétise dans trois domaines: dans les cours de production écrite, dans l'enseignement lexical (surtout pour maîtriser les collocations, la sémantique prosodique et les nuances entre synonymes) et aussi dans les cours de traduction, et ceci en s'appuyant sur les corpus parallèles et en promouvant surtout la méthode DDL.

Ainsi, la DDL et les techniques multimédia offrent à la didactique des LÉ une nouvelle piste, surtout grâce à des logiciels de requête facilement maniables, y compris des logiciels professionnels tels que Wordsmith Tools, Concordance, et des logiciels téléchargeables gratuitement. Bien sûr, on en est toujours au début de

l'exploitation de la méthode DDL en Chine, et il y a un long chemin à parcourir en ce qui concerne l'application des corpus numériques en didactique des LÉ.

Le fait qu'en Chine la didactique de l'anglais recourt fréquemment aux corpus numériques s'explique, d'un côté, par le développement de la linguistique de corpus à l'étranger, surtout dans les pays anglo-saxons et, de l'autre, par la domination de l'anglais LÉ en Chine. À part des cas exceptionnels, comme le travail de Xu & Zhang (2016) sur les corpus oraux de français, la quasi absence de recherches sur les corpus en français s'explique, comparativement, par trois raisons: il y a moins d'apprenants de français en Chine; les corpus en français sont beaucoup moins nombreux par rapport aux corpus en anglais; pour les corpus en anglais, les logiciels de requête sont nombreux et faciles à manier, ce qui n'est pas le cas en français.

INTÉRÊT ET LIMITES DES CORPUS NUMÉRIQUES POUR LA DIDACTIQUE DU FLE EN CHINE

Les débats autour de l'utilisation des corpus en Chine

L'évolution des corpus numériques en Chine suit une chronologie analogue à celle des domaines anglophone et francophone et se caractérise aussi par un passage du domaine de la linguistique à celui de la didactique, avec ce phénomène d'emprunt par les didacticiens d'un outil conçu par les linguistes. Si l'anglais domine aussi en Chine, il faut noter cependant la place qu'occupent les corpus d'apprenants du chinois langue étrangère, un domaine académique dynamique, aussi bien pour l'enseignement que pour la recherche, du fait de l'essor de la mobilité d'étudiants étrangers en Chine et de la forte demande d'enseignants chinois formés pour enseigner dans les instituts Confucius qui se développent à une vitesse rapide dans de nombreux pays.

Pour les langues étrangères, alors que les corpus d'interlangue d'apprenants d'anglais sont de plus en plus nombreux, il n'y a pas de corpus numérisés de l'interlangue d'apprenants de français. Quant aux travaux concernant l'apprentissage sur corpus et la didactique à partir des corpus, ils restent limités, même pour l'anglais. Cela fait écho au constat de Boulton & Tyne (2014) du faible développement de l'apprentissage sur corpus pour le français par rapport à l'anglais, et, s'il l'est, c'est davantage pour l'oral que pour l'écrit. De manière générale, les enquêtes en milieux francophones ou anglophones montrent que les corpus sont finalement encore relativement peu utilisés dans l'enseignement des langues (Cavalla & Loiseau, 2014; Frankenberg-Garcia, 2014; Kübler, 2014).

Dans le domaine francophone, on s'est interrogé sur les raisons de ce faible développement de l'apprentissage sur corpus. Les critiques habituellement adressées à l'apprentissage sur corpus se retrouvent dans le contexte chinois. L'accès à des sites étrangers, qui peuvent être parfois payants, n'est pas toujours aisé compte tenu des conditions de navigation sur l'internet chinois. Le coût en temps est important (Chambers, 2010) pour se former à l'utilisation de l'outil et pour aboutir à des résultats significatifs dont la rentabilité est jugée faible. Cet

inconvénient généralement constaté (Ciekanski, 2014) est encore plus préjudiciable dans l'organisation des études supérieures chinoises où la tendance actuelle est aux emplois du temps de plus en plus chargés et où l'objectif prioritaire de la réussite des tests nationaux de niveau de français (TFS 4 et TFS 8) concentre une part importante du temps de l'enseignement et de l'apprentissage (Li *et al.*, 2012).

En Chine aussi, les corpus sont vus comme des outils conçus par et pour les linguistes et non pour une exploitation didactique. Or, la linguistique est relativement peu présente dans les cursus des départements chinois de français langue étrangère, traditionnellement orientés vers l'enseignement de la langue et de la littérature. L'enseignement de la langue donne la priorité à la grammaire mais la maîtrise du métalangage des linguistes n'est pas majoritairement répandue. Les constats faits sur le manque de formation des enseignants (Kübler, 2014) se retrouvent pour la Chine. Le renouvellement des pratiques didactiques se nourrit peu des travaux de recherche francophones parce qu'ils donnent beaucoup d'importance à l'exploration des fondements théoriques de l'apprentissage sur corpus et passent plus rapidement sur les descriptions concrètes des dispositifs.

Un aspect qui est peu visé hors de Chine par les critiques des corpus joue en Chine un rôle important, c'est la question de la pratique effective de la langue. Du fait de l'histoire des méthodologies d'enseignement des langues en Chine, l'observation de la langue, beaucoup pratiquée dans le cadre de la méthodologie grammaire-traduction dominante jusque dans les années 2000, y est vue actuellement comme faisant écran à l'acquisition et donc à la mise en œuvre de la compétence de communication, c'est-à-dire à l'usage réel de la langue dans la salle de classe et en dehors (Hu, 2011).

Enfin, plusieurs aspects qui ont été vus positivement en contexte francophone ne sont pas perçus ainsi en contexte chinois. La forte valorisation de l'apprentissage en autonomie dans les travaux anglophones et francophones (par exemple Ciekanski, 2014) est une tendance encore timide en Chine (Fu, 2014). De même, le développement d'un esprit critique remettant en question les savoirs magistraux ou la pratique d'une démarche métacognitive et constructiviste ne sont pas non plus des aspects mis en avant dans les conceptions et les pratiques didactiques. La question des limites de l'usage didactique des corpus numériques pour certains publics, en particulier les publics d'apprenants asiatiques, a été vue par Boulton & Tyne (2014: 156) quand ils s'interrogent sur son adaptation à des "cultures éducatives orientées vers la transmission des savoirs, laissant peu de place pour l'autonomie, la découverte, la gestion du flou".

L'intérêt des corpus numériques en contexte chinois

Faut-il pour autant renoncer à explorer davantage l'intérêt didactique des corpus numériques pour le FLE en contexte chinois d'enseignement-apprentissage?

Culture d'apprentissage chinoise et hypothèse théorique du corpus mental

Certes, des aspects de la culture d'apprentissage chinoise peuvent, dans une certaine mesure, représenter un obstacle à l'apprentissage sur corpus. Cependant

cette culture d'apprentissage se trouve être particulièrement en cohérence avec des hypothèses sur l'acquisition langagière qui servent de fondement théorique à l'apprentissage sur corpus.

L'hypothèse d'un corpus mental, posée par Taylor (2012), est tout à fait intéressante si on la lit à partir du terrain chinois. Taylor entend par corpus mental la modalité de stockage en mémoire des connaissances langagières. La mémoire répertorie, sous toutes ses dimensions, l'expérience langagière vécue, y compris le contexte. Le cerveau n'apprendrait pas *a priori* des règles d'usage ou des normes pour ensuite les appliquer, mais la mémoire accumulerait un stock important d'exemples de fragments langagiers que le cerveau traiterait, avec de plus en plus de certitude et de précision, en faisant émerger des tendances et des régularités. C'est la mémoire de l'expérience linguistique qui, selon Taylor, permet l'apprentissage et cette mémoire fonctionne grâce à la très grande sensibilité des fonctions cognitives à la langue rencontrée. Ces fonctions se mobilisent constamment dans un processus continu d'innovation permanente qui dure toute la vie (*ibid.*: 285). Plus il y a exposition à la langue, plus cette expérience s'enrichit et s'affine et ainsi le cerveau généralise à partir de ces exemples avec de plus en plus de finesse. Pour Taylor, l'acquisition se fait donc à partir d'exemples en série.

Cette conception correspond à la représentation implicite qui est au fondement des méthodes d'acquisition de l'écriture dans la culture d'apprentissage chinoise. L'apprentissage de l'écriture en chinois, dès l'école primaire, repose sur l'habitude de copier des modèles de phrases ou de textes et d'emprunter aux textes lus. L'apprentissage de l'écriture n'est pas d'abord une expérience de création langagière à partir du donné intérieur de sa propre pensée mais un rapport à établir et à cultiver avec les textes d'autrui, par la copie et l'imitation. L'observation attentive des textes concerne d'abord la graphie de chaque sinogramme puis s'étend à la phrase et ensuite au texte. La compétence d'écriture s'acquiert par l'observation de l'écriture d'autrui, par le prélèvement de phrases et leur réemploi, d'abord littéral puis en effectuant progressivement des interventions qui sont la commutation et la substitution d'éléments équivalents, de manière simple puis de plus en plus complexe. L'élève chinois est donc habitué à construire, en quelque sorte, des lignes de concordance intériorisées qu'il enrichit en s'essayant à la variation. On pourrait dire, à la suite de Goody (1979), que l'apprentissage graphique forge des dispositions et habiletés cognitives qui peuvent être mises à profit pour l'apprentissage sur corpus. Le corpus représente alors un moyen pratique et efficace d'intensifier l'exposition à des séries langagières d'exemples dont l'usage est attesté, ce qui en fait une approche moins artificielle que la fabrication hors contexte d'exemples isolés.

Les corpus et l'évolution de la représentation de la langue chez les enseignants chinois

Le corpus, en tant que dictionnaire intelligent de la langue en usage qui permet de "démontrer l'existence de faits langagiers avérés" (Boulton *et al.*, 2013), présente un intérêt évident pour des enseignants non natifs de FLE qui disposent ainsi d'un outil plus fiable qu'un simple moteur de recherche. La formation des enseignants reste à développer en Chine. Elle pourrait s'inspirer des modèles théoriques et des

expériences menées ailleurs (Kübler, 2014). Mais il faut tenir compte des représentations des enseignants qui, dans la pratique, placent les outils renseignant sur le système de la langue avant ceux renseignant sur la langue en usage (en particulier les moteurs de recherche). Il est vrai que l'accès aux moteurs de recherche francophones n'est pas toujours facilité et les moteurs de recherche chinois classent en premier, dans les réponses, les sites officiels chinois francophones (par exemple les pages francophones des médias officiels ou de l'agence de presse Xinhua) dont les traducteurs souvent ne sont pas natifs.

Or, l'utilisation des corpus remet en question certaines manières de présenter le système de la langue. L'enseignant doit apprendre à gérer la tension entre l'usage et le discours normatif. L'observation des usages offre une palette beaucoup plus large que ce que les présentations systématiques, mais lacunaires, de la langue (les grammaires, les dictionnaires, etc.) parviennent à enseigner (Bilger & Cappeau, 2013). Ces présentations forgent même parfois des représentations de l'usage qui sont véhiculées dans le discours d'enseignement. Par exemple, en Chine, on trouve dans le discours d'enseignement la distinction entre "riche en" et "riche de", "riche en" étant présenté comme suivi d'un nom exprimant une réalité concrète. Or une requête comparative (l'expérience a été faite sur le corpus de *L'Est républicain* sur la plate-forme ScienQuest) fait apparaître facilement que "riche de" peut aussi être suivi d'un nom exprimant une réalité concrète. Le recours aux corpus conduit à un renouvellement du discours didactique de présentation théorique de la langue et met en lumière des idées reçues sur les usages. Le dictionnaire ou un certain discours établi de présentation systématique de la langue n'ont pas le dernier mot sur la langue étrangère. C'est une représentation traditionnelle en Chine que le corpus fait évoluer.

Développer l'intuition langagière de l'écart et de la convention dans la formation des futurs traducteurs chinois

L'intérêt des corpus pour la formation des futurs traducteurs a été analysé dans le domaine francophone (Kübler, 2014). Pour eux, les corpus représentent un dictionnaire des collocations habituelles très utiles pour le choix des termes en traduction. C'est un aspect à renforcer dans les formations en Chine où l'accent est mis sur l'acquisition du lexique et sur les glossaires. Un autre aspect à travailler serait la capacité à évaluer le degré du caractère conventionnel de la langue utilisée dans un certain genre. Les requêtes sur corpus peuvent être ciblées en fonction des genres et permettre l'acquisition des caractéristiques des genres et du langage attendu pour chacun. Ceci suppose une approche du langage qui prenne en compte l'écart qui travaille plus ou moins l'écriture dans les choix faits pour s'exprimer et donc pour traduire: le futur traducteur doit apprendre quels genres recourent à un langage à fortes contraintes, par exemple les genres juridico-administratifs, et quels genres se dispensent de contraintes ou s'en affranchissent ou en jouent, et c'est la liberté de l'écrivain dans les genres littéraires.

Prenons l'exemple d'une classe d'étudiants d'un cours de master de traduction chinois-français devant traduire des consignes pour le retrait des médicaments dans

un hôpital de médecine chinoise de Pékin. Dans le dispositif, l'enseignant collecte les traductions individuelles et réalise le concordancier de passages significatifs du texte pour faire observer comparativement aux étudiants les choix faits par chacun dans sa traduction. Par exemple, *shaodeng* se traduit par: veuillez patienter, merci d'attendre, attendez un peu, attendez, attendez quelques instants, attendez un moment, vous devez attendre, il faut attendre avec de la patience, attendez pour un moment, patientez, attendez avec patience, veuillez attendre un instant, veillez (*sic*) attendre, patientez pour un moment, etc. Ensuite la recherche ciblée sur Frantext dans des genres analogues et dans des genres différents permet d'apprécier, en fonction des attentes liées au genre de la consigne administrative, les choix heureux (au sens de la linguistique pragmatique) que l'on pourrait appeler avec Taylor (2012: 100-119) des choix "idiomatiques".

Cette formation est d'autant plus nécessaire que la question des genres est peu travaillée dans l'écriture en chinois du fait que l'enseignement de l'écrit au lycée est centré sur les types de textes qui sont très familiers aux étudiants et qui fonctionnent davantage chez eux pour catégoriser les manières d'écrire.

L'importance du français sur objectifs universitaires en Chine

La question du genre est centrale dans le domaine du français sur objectifs universitaires (FOU). En Chine, ce domaine connaît un développement impressionnant du fait de l'essor de la mobilité internationale. L'expérience de l'initiation à Scientext (Falaise *et al.*, 2011) dans un cours sur l'écriture du mémoire universitaire destiné à des étudiants de quatrième année de licence avec un projet de mobilité pour un master dans une université francophone, leur permet de vérifier des usages d'écriture de la langue académique et d'observer les "routines phraséologiques" (Tutin, 2014: 30). Tutin propose une typologie des séquences polylexicales de ce "langage préconstruit" de l'écriture académique (*ibid.*: 35) en distinguant les séquences à fonction référentielle, à fonction discursive, à fonction interpersonnelle et les routines sémantico-rhétoriques.

Avec des apprenants chinois, il est intéressant de travailler aussi avec une perspective plus large qui touche à la question de l'éthos discursif du scripteur (Maingueneau, 2014). Par exemple, l'enseignant demande aux étudiants de chercher sur Scientext "gentil, cœur, ami, s'amuser", des mots qu'ils utilisent habituellement sans tenir compte des genres. Ils prennent alors conscience de la non pertinence du registre affectif dans un écrit de genre académique en français alors qu'en chinois une démonstration intellectuelle tient aussi sa force de l'émotion qu'elle suscite et de la sensibilité qu'elle exprime.

Une autre activité consiste à faire entrer sur Scientext des proverbes qu'ils connaissent tous et utilisent fréquemment ("petit à petit l'oiseau fait son nid", "revenons à nos moutons", "c'est le premier pas qui coûte", "chaque médaille a son revers", etc.) pour faire découvrir qu'ils ne sont pas d'usage. Puis on les invite à entrer le mot "proverbe" ou des formules d'intégration de proverbes ("comme dit le proverbe"). Le peu de résultats obtenus les amène à émettre l'hypothèse que la sagesse proverbiale sert peu d'argument dans un écrit académique en français à la

différence de l'écriture des intellectuels en chinois où les proverbes sont valorisés pour leur valeur spirituelle et parce qu'ils sont l'un des indices du caractère littéraire que recherche traditionnellement l'écriture lettrée. Prenons alors l'exemple de "non… point" qui, dans les représentations des apprenants, correspond à une manière soutenue d'exprimer la négation, d'où son emploi fréquent dans les mémoires universitaires. Les résultats sur Scientext permettent d'observer l'effet littéraire de l'emploi. Laissé à lui-même, l'apprenant risque de tirer comme conclusion de ce constat que son choix d'employer fréquemment "non... point" dans les genres académiques est justifié.

Ces exemples montrent bien que le "guidage des apprenants" (Cavalla & Loiseau, 2014) dans le travail sur corpus est particulièrement nécessaire en contexte chinois, avec une méthode appropriée qui repère les besoins, dirige la formulation des requêtes et encadre l'interprétation des résultats.

UNE MÉTHODE D'UTILISATION DES CORPUS NUMÉRIQUES ADAPTÉE AU PUBLIC ET AU CONTEXTE CHINOIS

Limites et acquis de l'analyse des erreurs pour une didactique des langues à partir des corpus

Cette méthode, mise au point dans le cadre d'une expérimentation réalisée à l'Université des langues étrangères de Pékin, repère préalablement les besoins grâce à une liste d'une centaine de difficultés les plus fréquentes à l'écrit en français chez les apprenants chinois de cette université, liste mise au point à partir de l'observation empirique, depuis plusieurs années, d'un corpus non numérique (en l'absence jusqu'à aujourd'hui de corpus numérisés pour le français) de plusieurs centaines de leurs productions écrites. Ce corpus empirique permet de cibler un groupe d'apprenants passés par le même parcours de formation en français (manuels, enseignants, méthodes, cours, etc.).

Avec cette liste, il ne s'agit pas de refaire les tentatives d'analyse des erreurs en proposant des inventaires et typologies théoriques des erreurs commises et en comparant les langues pour diagnostiquer des transferts de la langue maternelle. La critique de l'analyse des erreurs (Besse & Porquier, 1991: 204-209; Vogel, 1995: 29-57) a montré que l'erreur n'est jamais prévisible et reste foncièrement individuelle. L'analyse des erreurs est une impasse si c'est la tentative de généraliser et de parler de l'interlangue commune à tous les étudiants chinois de français avec ses fossilisations typiques. La généralisation qui est ici pratiquée à l'échelle du groupe d'étudiants impliqués est une forme d'intuition temporaire qui sert seulement à repérer des phénomènes récurrents qui travaillent la langue de ces apprenants et qui appellent des techniques pour être améliorés. Cette généralisation, il faut ensuite progressivement y renoncer. Dans notre démarche, la perspective est de permettre, à terme, à chaque apprenant de se constituer sa propre liste de points à travailler à partir de ses travaux corrigés par les enseignants et de tenter personnellement des requêtes dont il présente ensuite la conception et les résultats.

Il y a, cependant, un acquis positif de l'histoire des recherches en analyse contrastive et en analyse des erreurs qui est d'avoir attiré l'attention sur les stratégies subjectives d'apprentissage et de construction de sens de l'apprenant et sur sa créativité (Corder, 1967), de poser la question du degré de distance entre les deux langues selon l'angle de perception de l'apprenant et d'envisager autrement la question de l'erreur (Marquillo Larruy, 2003) qui n'est plus vue comme inévitable ou nuisible, comme une faiblesse ou une hantise, mais comme une caractéristique de la progression. Les erreurs sont le "signe que l'apprenant est en train d'explorer le système d'une langue nouvelle" (Corder, 1967: 169).

De Cock & Tyne (2014: 2) font remarquer que les phénomènes observés sur les corpus d'interlangue sont rarement expliqués. D'un point de vue didactique, en tenant compte de l'impasse de l'analyse des erreurs, ce qui est plus intéressant c'est d'amener l'apprenant à formuler lui-même un diagnostic sur les difficultés constatées, diagnostic qui peut être varié (calque du chinois, calque de l'anglais, etc.) et qui peut d'ailleurs le conduire à se démarquer du groupe ("moi, je ne fais pas cette erreur-là"), une position qui ouvre à la correction entre pairs. De Cock & Tyne (*ibid.*: 12) insistent sur l'utilisation des "preuves positives" (usages corrects) et sur les exercices de consolidation. Il est intéressant aussi d'avoir une approche dynamique et souple de l'erreur et de l'usage. Par exemple, travailler sur la confusion fréquente entre "soutenir" et "supporter" avec un concordancier fait prendre conscience que l'interférence de l'anglais gagne les usages en français où "supporter quelqu'un", au sens de soutenir, se dit dans le monde du spectacle ou du sport.

Le guidage indispensable

C'est sur la question du sous-emploi ou du suremploi de mots ou d'expressions que le recours aux corpus est particulièrement intéressant à travailler avec des apprenants chinois, en s'appuyant notamment sur l'utilisation parallèle d'un corpus chinois et d'un corpus français, à savoir la partie contemporaine de Cncorpus et Scientext ou bien le corpus de BLCU ou celui de l'Université de Pékin (CCL) et Frantext. Les caractéristiques de ces corpus chinois (BLCU, 2015; Zhan *et al.*, 2003; Xiao, 2009; 2012) correspondent bien à Scientext et à Frantext, avec des genres académiques, scientifiques ou universitaires pour Cncorpus et une grande variété de genres dans une langue contemporaine pour les corpus de BLCU et CCL.

La liste de constats empiriques fait ressortir des phénomènes de suremploi, par exemple:

- *guocheng* / processus: avec le processus de mondialisation (à comparer à: avec la mondialisation);
- *nuli* / faire tous ses efforts: le gouvernement fait tous ses efforts pour résoudre ce problème (à comparer à: s'efforce de, essaie de);
- *qifa* / inspiration: cet article m'a donné des inspirations (à comparer à: … m'a fait réfléchir, m'a donné des idées).

Ceci concerne des mots mais aussi des usages syntaxiques, par exemple: commencer ses phrases par "et" ou par "parce que". À l'inverse, il y a des constats

de sous-emploi, par exemple le recours au gérondif est sous-employé, il lui est préféré des formules comme: *il présente cette idée par expliquer[1] (à comparer avec le chinois *yong*); au mot "évolution" est préféré le mot "développement" (à comparer avec le chinois *fazhan*), etc. Le but est didactique, c'est-à-dire qu'il ne s'agit pas d'établir un diagnostic contrastif, mais de stimuler chez l'apprenant des intuitions langagières qu'il n'a pas, ou pas de manière active, c'est-à-dire d'avoir dans la pratique l'idée même d'employer un mot ou une structure. Donc l'exercice n'a pas pour finalité l'observation (construire un modèle théorique d'explication des sous-emplois ou des suremplois) mais l'exploration.

Une autre activité consiste à repérer des expressions ou mots n'existant pas en français mais fréquents dans le discours écrit des étudiants chinois. Par exemple,
- "jeter/établir des bases": pendant ces trois années, j'ai *établi/jeté des bases en français (…j'ai acquis une connaissance générale du…);
- "sous la circonstance de": *sous la circonstance de la mondialisation, cette question est devenue importante (dans le contexte de la mondialisation…);
- "faire une partie importante": le thé *fait une partie importante de la culture chinoise (le thé est un élément important de…);
- "attitude sur": l'auteur présente son *attitude sur cette question (opinion).

Dans ces cas-là, il s'agit de faire vérifier si l'expression existe (par exemple, au début de l'expérimentation, la presque totalité des étudiants sont convaincus que l'expression "l'attitude sur" existe en français), de faire des hypothèses sur des expressions alternatives puis de les tester et ensuite de faire des requêtes pour observer les collocations des mots-clés aussi bien pour les mots-clés des expressions fautives ou non usuelles (ici: base, circonstance, partie, attitude) que pour les mots-clés des expressions attestées (ici: connaissance, contexte, élément, opinion). Ceci permet de les fixer en mémoire et de consolider les acquis.

Comme souvent dans les expérimentations didactiques, les requêtes concernant la syntaxe tiennent une place assez importante, mais elles donnent lieu à des observations relativement faciles à faire et c'est un bon entraînement pour l'acquisition des procédés de formulation des requêtes. La liste oriente le travail sur les collocations usuelles en français, mal connues des étudiants chinois qui leur préfèrent souvent d'autres collocations pas forcément fautives mais peu ou pas usuelles. Par exemple: profiter du soleil/bénéficier du soleil; c'est la même chose pour/c'est le même cas pour; l'enseignement supérieur/l'éducation supérieure; donner son avis/donner son idée; acquérir des connaissances/prendre des connaissances; le style d'un écrivain/le style d'écrire d'un écrivain, etc.

Prenons l'exemple d'une activité pour des étudiants débutants dans l'utilisation des corpus, invités à travailler sur les collocations de "habitude" (la liste recense l'emploi systématique et fautif de "former l'habitude" par les apprenants). L'activité est guidée par étape. La première étape consiste d'abord à compléter librement un texte à trous tiré d'un concordancier élaboré à partir des résultats de requêtes sur le corpus de *L'Est républicain* (sur ScienQuest).

[1] L'astérisque signale les formes erronées.

1. Les élèves de l'école et du collège St-Joseph ont depuis quelques années _____ l'habitude de jouer ou d'aller au théâtre accompagnés de leurs enseignants, parfois de leurs parents. Ce week-end, le théâtre était dans l'établissement et les adultes sur les planches.

2. En ce moment, on le voit plusieurs fois par jour, perché sur une échelle en train de donner à manger à deux petits cigogneaux nés ce printemps. Il en _____ l'habitude puisque ce n'est pas la première couvée.

3. Pourquoi renoncer à la direction de l'école? "Le plus important à l'école, c'est l'enseignement", confie-t-il. "Il faut un enseignement moderne, _____ les habitudes. Il faut des idées novatrices, un œil nouveau…" Et d'ajouter: "C'est une décision personnelle. Je m'investirai afin que le passage se fasse positivement, pour le bénéfice de l'école".

La deuxième étape consiste à envoyer les étudiants sur le même corpus avec les formules écrites très précises des requêtes (en haut de la page à gauche vous lisez "mot 1", écrivez... puis cliquez sur "mot 2" et écrivez, etc.) afin qu'ils cherchent les verbes en usage avec "habitude". Certes les étudiants chinois font preuve d'une habileté dans l'usage des nouvelles technologies; cependant une enquête informelle auprès de plusieurs classes d'étudiants montre qu'ils ne se lancent pratiquement pas seuls sur les plateformes en français. Ceci explique le caractère très dirigé des activités proposées.

La troisième étape consiste à vérifier l'exercice à trous en retrouvant les exemples dans les résultats qui s'affichent. L'étape suivante consiste alors à chercher d'autres verbes pouvant s'employer avec "habitude". Le principe est que les apprenants, de manière individuelle ou collective, se constituent eux-mêmes un exemplier en ne retenant dans les réponses obtenues que les exemples qu'ils comprennent. La dernière étape, c'est l'exploration libre du mot "habitude" afin de récolter des exemples d'autres collocations déjà connues de manière active ou passive. Dans cette dernière étape, les étudiants s'essaient à la manipulation de l'outil en testant de nouvelles formules pour leurs requêtes, l'enseignant jouant le rôle d'assistant technique.

Les requêtes concernant la syntaxe sont non seulement les plus faciles à formuler mais aussi celles qui donnent des résultats interprétables en observant un voisinage restreint des occurrences. Par exemple, les prépositions après les verbes (aider *de/à faire; espérer *de/Ø faire; il vaut mieux de/Ø; traduire *comme/par), la forme syntaxique de structures (comme *ce que/le dit cet auteur; il n'est pas *si grand comme/aussi grand qu'elle; en tant qu'*un/Ø média; *n'importe/peu importe s'il est; *tous ces trois/tous ces/ces trois arguments), la syntaxe à l'échelle de la phrase (bien que… *mais; même si.. *mais; *selon moi, je; *il y a un auteur s'appelle…), les calques du chinois (déjà, tous, un peu / *yijing*, *dou*, *yixia*). Ce n'est pas le cas des activités de comparaison du sens de mots proches et confondus (par exemple, situation/ condition; mou/doux/tendre; connaître/savoir; découvrir/trouver; prolonger/ allonger; trop/très; visiter/rendre visite; vouloir bien/vouloir; gouvernement/État) qui demandent d'explorer plus longuement le mot ou l'expression dans son contexte.

Conclusion

Sur le terrain chinois de la didactique du français, les perspectives d'avenir pour les corpus numériques sont liées à l'essor des formations. Certaines sont organisées sur la base d'un manuel en chinois pour l'exploitation des corpus (Liang, Li & Xu, 2010), qui en est à sa sixième édition et qui figure au classement officiel chinois des ouvrages pour les enseignants des langues étrangères dans le supérieur, dans la série "méthodologie de la recherche" (*quanguo gaodeng xuexiao waiyu jiaoshi congshu, keyan fangfa xilie*), ce qui est le signe de sa valorisation officielle, importante en Chine.

Par ailleurs, il faut souhaiter qu'en Chine la numérisation de corpus d'apprenants chinois de FLE se développe, afin de ne pas travailler seulement à partir de corpus empiriques non numérisés. Mais la limite même des corpus empiriques permet d'avoir une approche prudente de l'interlangue avec la conscience de la tentation intellectuelle des généralisations dans l'analyse de productions d'apprenants alors que les généralisations ne sont qu'un relais provisoire dans une démarche qui, pour éveiller à l'autonomie des sujets apprenants, doit avoir pour horizon la singularité de leurs situations particulières.

Une autre perspective pour la Chine est la création d'une plate-forme facile à utiliser, et si possible unique, pour interroger en ligne tous les corpus numériques en français avec des requêtes très simples. La simplification dans l'usage des technologies est une cause et une conséquence de leur diffusion de plus en plus large. En Chine, unifier les fonctionnalités diverses offertes par une technologie nouvelle pour la rendre praticable par le plus grand nombre est une piste déjà explorée de manière performante dans d'autres domaines de la vie sociale. Cependant vivre l'étape actuelle de la complexité d'utilisation des outils existants fait aussi réfléchir à la relation entre l'homme et la machine, un domaine avec d'immenses perspectives en linguistique et en didactique mais aussi avec des enjeux éthiques fondamentaux, une réflexion nécessaire en contexte chinois.

Références bibliographiques

BESSE, HENRI & RÉMI PORQUIER. 1991. *Grammaire et didactique des langues*. Paris: Didier.

BLCU. 2015. *Beijing yuyan daxue BCC yuliaoku* (*Le corpus numérique BCC de l'Université des langues et cultures de Pékin*). URL: http://yyjx.blcu.edu.cn/art/2015/6/17/art_9112_1098463.html.

BILGER, MIREILLE & PAUL CAPPEAU. 2013. Comment les données de corpus pourraient renouveler les manuels de grammaire? Illustrations à partir de quelques pronoms et de la forme même. *Linx* 68-69, 177-199.

BOULTON, ALEX & HENRY TYNE. 2014. *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. Paris: Didier.

BOULTON ALEX, EMMANUELLE CANUT, EMMANUELLE GUÉRIN, CHRISTOPHE PARISSE & HENRY TYNE. 2013. Corpus et appropriation de L1 et L2. *Linx* 68-69, 9-32.

CAVALLA, CRISTELLE & MATHIEU LOISEAU. 2014. Scientext comme corpus pour l'enseignement. AGNÈS TUTIN & FRANCIS GROSSMANN (dir.). *L'écrit scientifique: du lexique au discours*. Rennes: Presses universitaires de Rennes, 163-180.

CHAMBERS, ANGELA. 2010. L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé. *Revue française de linguistique appliquée* 15: 2, 9-20.

CIEKANSKI, MAUD. 2014. Les corpus: de nouvelles perspectives pour l'apprentissage des langues en autonomie? *Recherches en didactique des langues et des cultures. Les Cahiers de l'Acedle* 11: 1, 1-19.

CORDER, PIT. 1967. The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching* 5: 4, 161-170.

DE COCK, SYLVIE & HENRY TYNE. 2014. Corpus d'apprenants et acquisition des langues. *Recherches en didactique des langues et des cultures. Les Cahiers de l'Acedle* 11: 1, 1-23.

FALAISE, ACHILLE, AGNÈS TUTIN & OLIVIER KRAIF. 2011. Une interface pour l'exploitation de corpus arborés par des non-informaticiens: la plate-forme ScienQuest du projet Scientext. *TAL* 52: 3, 103-128.

FRANKENBERG-GARCIA, ANA. 2014. How language learners can benefit from corpora, or not. *Recherches en didactique des langues et des cultures. Les Cahiers de l'Acedle* 11: 1, 1-15.

FU, RONG & PHILIPPE MOGENTALE (dir.). 2014. *Autonomie dans l'apprentissage des langues et apprentissage de l'autonomie. Synergies Chine* 9.

GOODY, JACK. 1979. *La raison graphique. La domestication de la pensée sauvage.* Paris: Éditions de Minuit.

HU, YU. 2011. Place du communicatif et didactique dans les pratiques et dans les représentations d'enseignants chinois de français. *Synergies Chine* 11, 117-126.

JOHNS, TIM. 1991. Should you be persuaded: two samples of data-driven learning materials. TIM JOHNS & PHILIP KING. *Classroom concordancing. ELR Journal* 4, 1-16.

KÜBLER, NATHALIE. 2014. Mettre en œuvre la linguistique de corpus à l'université. *Recherches en didactique des langues et des cultures. Les Cahiers de l'Acedle* 11: 1, 1-34.

LI, KEYONG, ZHIHONG PU, RONG FU & JEAN-JACQUES RICHER (dir.). 2012. *Curriculum et évaluation. Synergies Chine*7.

LI, WENZHONG & JIANZHONG PU. 2001. Yuliaoku suoyinzai waiyu jiaoxue zhong de yingyong (L'usage de l'indexation des corpus de recherche dans la didactique des langues étrangères). *Journal académique de l'Institut des langues étrangères de l'APL* 24: 2: 20-25.

LIANG, MAOCHENG, WENZHONG LI & JIAJIN XU. 2010. *Yuliaoku yingyong jiaocheng* (*Manuel d'utilisation des corpus numériques*). Pékin: Éditions de l'enseignement et de la recherche en langues étrangères.

MAINGUENEAU, DOMINIQUE. 2014. Retour critique sur l'éthos. *Langage et société* 149, 31-48.

MARQUILLO LARRUY, MARTINE. 2003. *L'interprétation de l'erreur*. Paris: CLE International.

TAN, XIAOPING. 2014. Jinshinian hanyu yuliaoku jianshe yanjiu zongshu (Une décennie de recherches sur l'évolution des corpus en chinois). *Actes des travaux du 7e Forum de l'enseignement de Chinois langue étrangère de Pékin*. Pékin: Éditions de l'Institut de didactique du chinois langue étrangère de l'Université de Pékin, 26-31.

TANG, JIEYI & ANPING HE. 2004. Yuliaoku zai waiyu jiaoxue zhong de yingyong (L'usage des corpus dans la didactique des langues étrangères). *Computer-Assisted Foreign Language Education* 99, 42-45.

TAYLOR, JOHN. 2012. *The mental corpus: how language is represented in the mind*. New York: Oxford University Press.

TUTIN, AGNÈS. 2014. La phraséologie transdisciplinaire des écrits scientifiques: des collocations aux routines sémantico-rhétoriques. AGNÈS TUTIN & FRANCIS GROSSMANN (dir.). *L'écrit scientifique: du lexique au discours*. Rennes: Presses Universitaires de Rennes, 27-43.

VOGEL, KLAUS. 1995. *L'interlangue*. Toulouse: Presses universitaires du Mirail.

XIAO, HANG. 2009. *Guojia yuwei xiandai hanyu tongyong pingheng yuliaoku biaozhu yuliaoku shuju ji shiyong shuoming* (*Le corpus numérique, général et équilibré, de chinois contemporain de la Commission linguistique nationale. Les données du corpus annoté et leur utilisation*). Publication de l'Institut de linguistique appliquée du ministère de l'Éducation sur le portail chinois des sciences sociales. URL: http://www.cssn.cn/yyx/ yyxcyzy/201509/W020150922441291714295.pdf.

XIAO, HANG. 2012. *Guojia yuwei xiandai hanyu yuliaoku jieshao* (*Présentation du corpus numérique de chinois contemporain de la Commission linguistique nationale*). Publication de l'Institut de linguistique appliquée du ministère de l'Éducation sur le portail chinois des sciences sociales. URL: http://www.cssn.cn/yyx/yyxcyzy/ 201509/W020150922441286080921.pdf.

XU, RIXUAN & CHANG ZHANG. 2016. Enjeux des corpus oraux pour l'enseignement universitaire du français langue étrangère en Chine. L'exemple de la base de données CLAPI. *Synergies Chine* 11, 177-188.

XU, XUILING & JIAJIN XU. 2017. Woguo waiyu jiaoxue zhong de yuliaoku yingyong 40 nian (L'utilisation des corpus en Chine pour l'enseignement des langues étrangères depuis 40 ans). *Didactique des langues étrangères en Chine* 4, 62-68.

ZHAN, WEIDONG, RUI GUO & YIRONG CHEN. 2003. *Beijing daxue zhongguo yuyanxue yanjiu zhongxin CCL yuliaoku (guimo: 7 yizi; shijian: gongyuanqian 11 shiji – dangdai)* (*Le corpus numérique CCL du Centre chinois de recherches en linguistique de l'Université de Pékin, rassemblant 700 millions de sinogrammes, du 11e siècle avant notre ère à aujourd'hui*). URL: http://ccl.pku.edu.cn:8080/ccl_corpus/CCLCorpus_Read me.html.

ZHANG, XINJIE. 2017. Guonei yuliaoku yuyanxue yanjiu: huigu yu zhanwang – jiu hexin qikan 24 nian wenxian de tongji fenxi (Les recherches en linguistique de corpus en Chine: histoire et perspectives. Analyse statistique de 24 publications clés). *Journal académique de l'Université des langues étrangères de Xian* 25, 36-41.

ZHEN, FENGCHAO. 2010. Yuliaoku yuyanxue zai zhongguo de chengzhang yu fazhan (La maturation et le développement de la linguistique de corpus en Chine). *Recherches contemporaines en langues étrangères* 3, 36-41.

ZHENG, MENG. 2014. Guonei hanyu yuliaoku gaikuang (Les corpus en chinois en Chine). *Perspectives technologiques* 27, 163-164.