On the Use of Corpora in Second Language Acquisition – Chinese as an Example

Mária IŠTVÁNOVÁ
Comenius University in Bratislava, Slovakia
istvanova6@uniba.sk

Abstract

This paper aims to introduce the language corpora and the advantages of their use in the process of Chinese language acquisition. We provide practical examples of the corpora's direct and indirect use for teaching and learning Chinese as a second language. The exploratory approach towards Chinese by using various types of corpora is applicable for general language seminars as well as specialized translation seminars. The indirect use is mainly linked to the preparation of teaching materials and facilitates the curriculum design.

Keywords: Chinese; corpora; teaching methodology; second language acquisition; linguistic research

Povzetek

Članek predstavi jezikovne korpuse in prednosti njihove uporabe v procesu usvajanja kitajskega jezika ter ponudi številne praktične primere neposredne in posredne uporabe korpusov pri poučevanju in učenju kitajščine kot drugega jezika. Raziskovalni pristop do kitajščine s pomočjo različnih korpusov je primeren tako za splošne jezikovne vaje kitajščine kot drugega jezika kot za vaje prevajanja. Posredna uporaba korpusov pa je priporočljiva pri pripravi učnih gradiv in olajšuje oblikovanje učnega načrta.

Ključne besede: kitajščina; korpusi; metodologija poučevanja; učenje drugega jezika; jezikoslovno raziskovanje

1   Introduction

The corpora play an important role in the teaching and learning of a second language and although their use in language analysis is regarded as a relatively new approach, we already perceive a dynamic relationship established between corpora and linguistic research (Casas-Pedrosa et al., 2013, p. 2). The increasing number of studies embedded in the field of corpus linguistics and numerous published articles reflect the rising interest in the use of corpora for second language acquisition and the study of a learner's interlanguage. The emergence of corpora and their use in language learning transforms the traditional teaching approaches and methods used for the compilation of teaching materials.

2   Direct use of corpora in second language acquisition

The direct use of corpora enhances the process of second language acquisition, however, specialized training in corpus analysis for teachers and students is necessary to ensure a successful application of corpora in classroom instruction. Apart from that, the direct use of corpora requires certain adjustments in the curriculum design, therefore its incorporation in classroom activities is more common at higher educational institutions such as universities (McEnery & Xiao, 2011, p. 374). The introduction to corpus linguistics and familiarity with the user interface of the selected corpus facilitates the interpretation of the query, therefore in cases when students are not familiar with the use of corpora, the practical preparatory sessions would improve the efficiency of the corpora's use in the instruction.

Following the development of teaching methodology of a second language, the teacher's focus shifts from a linguistic theory to a student, and the student's ability to work independently is emphasized. We may therefore expect students to have mastered learning strategies, and find corpora a valuable source of information about the target language. At such a level, the students are able to easily compare various characteristics between their mother tongue and the studied language (Benko et al., 2019, p. 14). From the learners' perspective, the direct use of corpora in classroom instruction is beneficial because of their limited exposure to the target language, and the teacher's guidance is crucial especially for elementary or intermediate learners due to their limited knowledge of the target language (Tsui, 2004, p. 40). The exploration of a large amount of language material compiled in a corpus creates a suitable environment for the so-called *learning as discovery* approach. In this way, we enhance learner's individual research interests and improve their language awareness (Bernardini, 2004, p. 22). The use of corpora in instruction is closely linked to the data-driven learning method (DDL). This method is based on student's independent work with language data to explore the regularities and patterns of the target language (the corpus is perceived as the core teaching material). We shift the learner's passive

position of receiving information into an active position of a researcher of the second language (e.g., by using the concordance search feature or keyword in context – KWIC) (Johns, 1991, pp. 9-13). The integration of corpora and language instruction leads to the active participation of students, which further enables them to discuss the difficulties and learning strategies among each other. Apart from that, it strengthens the learner's focus on the form and meaning as well as the relation between the two (Bernardini, 2004, p. 28).

Regarding the use of corpora with the objective to increase the learner's language proficiency, the study of collocations represents an efficient way for acquiring the prefabricated units that are necessary for achieving native-like language proficiency, because the number of acquired collocations facilitates language comprehension and determines the learner's language production (Cowie, 1994, p.3168). The perception of collocations differs among linguists and most linguists are convinced that the use of corpora to extract language data is the most reliable way of identifying collocations (Harris, 2006; Sinclair et al., 2004). The collocation boundaries are indistinct because collocations are either perceived as constructions composed of elements in direct proximity or in a broader sense, and the individual units are not necessarily followed one after another but they occur near the selected language unit (McEnery & Hardie, 2012, p. 123). Therefore, great importance is placed on the knowledge of lexical units with their co-occurring items, including their position in a sentence, instead of on individual words (McEnery & Xiao, 2011, p. 368).

3   Direct use of Chinese corpora

The use of Chinese corpora for teaching and learning purposes is favorable because of their quantitative approach in language study (Xu, 2019, p. 49). The corpora are suitable for displaying positive examples in a given context when focusing on the structures that need to be acquired by the learners (Chen & Tao, 2019, p. 75). The teaching approaches which base on information mined from the corpus concentrate on explanations of commonly used language forms in Chinese, and therefore the instruction time is focused mainly on the most relevant knowledge (Zhao & Kang, 2017, p. 86). An example of direct use of a corpus in the instruction has been introduced by Gajdoš in *Chinese Legal Texts – Quantitative Description* (Gajdoš, 2017). The subcorpus of Chinese legal texts as an integral part of the *Hanku* corpus enables teachers to concentrate on a specific field of language instruction. One of the prerequisites for the successful inclusion of corpora in the instruction is the learner's knowledge of CQL query and regular expressions, which is necessary for the retrieval of the language data (Gajdoš, 2017, p. 86). The query functionalities and search options vary based on the user interface of a chosen corpus. Not only the knowledge of the use of corpora but also a refined knowledge of grammatical rules of the target language is an important

prerequisite that allows a corpus to be successfully included in the process of language learning.

Incorporation of corpora in the translation seminars requires revisions of the existing curricula and training of teachers in corpus linguistics focused on the acquisition of necessary skills related to software use, statistical analysis, and data interpretation (Hu, 2011, p. 184). The main purpose of the translation teaching methodology is the development of the learner's translation competence and translation awareness. The translation competence represents a student's ability to transfer meaning of the source language into meaning in the target language by employing suitable translation strategies. The translation awareness reflects the student's understanding of the nature of the translation subject, object, or receptor and the relation between them (Liu, 2003, p. 1). Corpora is a useful training tool for future translators because it enhances understanding of typical translation problems solved by experienced translators when working for instance with parallel corpora (Bernardini, 2004, pp. 20-21). Figure 1 (Hu, 2011, p. 148) displays a variation in the translation of the word *shehui* 社会 into English as a target language. Based on the fixed terms in Chinese, it is necessary to select an appropriate translation and the query in the corpus facilitates the decision-making process, especially in the case of inexperienced learners and novice translators.

由"社会"组成的术语	英语译名
社会事业	social undertakings
小康社会	a well-off society
社会秩序	public order
社会主义和谐社会	a harmonious socialist society
社会监督	public scrutiny
社会体育组织	mass sports organization
社会保障体系	social security system

Figure 1: Translation of word *shehui* 社会 in different contexts

As mentioned in the previous section, the study of collocations improves learner's language production and comprehension skills. In this way, corpora represent a powerful tool when getting acquainted with linguistic conventions in Chinese (Jing-Schmidt, 2019, p. 23). Collocation functionalities enable learners to extend their knowledge of prefabricated units and increase their ability to identify collocations in the context or to employ them in language production. Figure 2 displays a collocation of the verb *zuzhi* 组织 followed by a noun by using Lancaster Corpus of Mandarin Chinese accessible on <corpus.leeds.ac.uk/query-zh.html>. The corpus functionality

among other statistical data (MI, Dice, and LL score) provides a T score and offers the possibility to display the chosen expression in a context with all corresponding entries.

T score

Collocation	Joint	Freq1	Freq2	T score	Concordance
组织 结构	15		463	3.81	Examples
组织 机构	7		207	2.60	Examples
组织 形式	7		269	2.59	Examples
组织 理论	7		344	2.57	Examples
组织 群众	4		293	1.92	Examples
组织 领导	4		382	1.90	Examples
组织 程度	3		208	1.67	Examples
组织 学生	3		469	1.58	Examples
组织 蛋白	2		8	1.41	Examples
组织 民兵	2		9	1.41	Examples
组织 人力	2		17	1.41	Examples
组织 城镇	2		22	1.41	Examples

Figure 2: Collocation of *zuzhi* 组织 followed by a noun

Corpora are useful for improving learner's knowledge of the general vocabulary, but it is also possible to concentrate on a specific field based on a learner's future specialization. Such specializations are mainly required for academic or any other professional purposes with the objective to achieve native speaker-like language proficiency in a particular field (Zhao & Zhang, 2015, p. 38). Monolingual, as well as parallel corpora, are useful reference sources. Figure 3 (Guan & Tao, 2017, p. 178) displays different translations of the word *tiaojie* 调解 into English. The provided information is extracted from the specialized legal corpora *Beida fabao falv fagui shujuku* 北大法宝法律法规数据库 and *Wanshi falv xinxi shujuku* 万事法律信息数据库 (Guan & Tao, 2017, p. 178). Its translation into the target language varies based on the context and particular use of the expression. In the case of an existing specialized corpus in the language combination of the learner's mother tongue and Chinese, the corpora in the Chinese – English language combination might be employed as a supplementary reference. Translation of this term in a general web-based dictionary of Chinese is *to mediate* and *to bring parties to an agreement, bring together to an agreement* (retrieved from <mdbg.net> and <www.chinese-dictionary.org>). Specialized legal dictionary – *Chinese-English/English-Chinese Pocket Legal Dictionary* (Chen, 2008, p.19) also provides the translation of *tiaojie* 调解 as *mediation, to mediate*. In comparison to entries in the dictionaries, the query in the corpora provides another possibility of a translation in English – *to conciliate, conciliation*.

"调解"的用法	英 译	解 释	一般用法
作为名词	Mediation		
作为动词	Mediate	强调第三方的调节作用。	Mediate + xx disputes/cases/complaints
	Conciliate	强调审判或仲裁前的调解。	Conciliate + xx disputes; 作不及物动词
调解书	Conciliation statement	在相关机构主持下双方达成和解的法律协议。	
	Mediation document	民事诉讼中双方当事人向有关部门申请,调解书作为对双方均具有法律约束力的文件。	

Figure 3: Translation of *tiaojie* 调解

Apart from the use of parallel corpora in the field of translation didactics, we generally consider the parallel corpus useful for explaining studied expressions within the context, because it provides a conceptual framework for their understanding in the learner's native language as well as in their target language, Chinese in this case (Bluemel, 2019, pp. 85-86). Figure 4 displays query of *suzhi* 素质 extracted from *CCL Chinese-English parallel corpus* accessible on the website of Center for Chinese linguistics PKU (*Beijing daxue Zhongguo yuyanxue yanjiu zhongxin* 北京大学中国语言学研究中心, <ccl.pku.edu.cn:8080/ccl_corpus/index_bi.jsp>). We selected the first ten entries to display the variability in the English translation. The prevailing translation is *quality*, but there are also two entries translated into English as *competence* and one entry translated as *consciousness*. The chosen expression reflects a specific term in Chinese, and the context, as well as several translation options, enhance its correct understanding by learners. In comparison to query in the parallel corpus, the translation of *suzhi* 素质 in the web-based dictionaries (<mdbg.net> and <chinese-dictionary.org>) is *inner quality* and *basic essence*. In this case, the use of corpora provides valuable guidance for the selection of an appropriate translation in the target language.

	Chinese	English
1	现在飞行事故多，当然这里面有部队训练不好、飞行员素质差、地面指挥水平低等问题，但是也要研究飞机质量方面的问题。	Airplane accidents are numerous nowadays, and of course some are attributable to the poor training of the men, the incompetence of the pilots or the deficiencies of the ground-control crews. But sometimes the fault may lie in the quality of the aircraft.
2	...在不打仗，你根据什么来考验干部，用什么来提高干部，提高军队的素质，提高军队的战斗力？	But now that we are not at war, how are we to test our cadres, raise their level, and improve the quality and combat effectiveness of our troops?
3	通过学习、拉练、演习，可以提高各级干部的素质。	The quality of cadres at different levels can be enhanced through study, camp and field training, and military exercises.
4	...要承认我们军队的人数虽然多，但是素质比较差。	We must also admit that our army is not sufficiently capable of conducting modern warfare, and that although it is numerically strong, it is of relatively poor quality.
5	我们军队的素质在战争年代是很好的，在抗美援朝时也是很好的。	It was of very good quality during the war years and the War to Resist U.S. Aggression and Aid Korea.
6	老实说，有够样的素质，就是用现有的武器，也可以打仗，也可以学会适应现代化战争的条件...	Indeed, given quality of that kind, our army could fight even with its present weapons and could learn to adapt itself to the conditions of modern warfare and defeat the powerful imperialists.
7	现在的问题是，由于遭受林彪、"四人帮"的干扰破坏，我们军的素质确实不如以前了。	The point is that because of the interference and sabotage by Lin Biao and the Gang of Four, the quality of our army just isn't as good as before.
8	...七五年我们就提出要把训练提高到战略地位，因为不打仗，部队军事素质的提高就得靠训练。	In 1975 we proposed that training should be considered of strategic importance because, in the absence of war, it is the only way to improve the army's quality.
9	要大力加强政治、公安部门的建设和工作，提高这些部门1人队的政治素质和业务素质。	Great effort should be made to strengthen the public security, procuratorial and judicial departments, improve their work and enhance the political quality and professional competence of their personnel.
10	...特点，摸索了现代条件下诸军兵种协同作战的经验，提高了部队军队素质和实战水平。	These exercises have given us an opportunity to assess our achievements in building modern, regularized armed forces, and have simulated modern warfare fairly well. They represent part of our effort to explore combined operations by the various services and arms under modern conditions, and have enhanced the political consciousness and military capability of our army, especially its capacity to fight actual battles.

Figure 4: translation of *suzhi* 素质

The incorporation of corpora in classroom instruction is not a trouble-free process. It is on a teacher to decide whether it is necessary to compile a new corpus corresponding to the requirements of the teaching objectives, or whether the existing one would fulfilling such prerequisites. Despite the fact that most of the commonly taught second languages already have several existing corpora, there are nevertheless cases when creating a new corpus is necessary. In such a case, a teacher has to deal with curriculum adjustments and would face general issues related to the creation of a corpus including the compilation procedure, representativeness, or source of language material (Casas-Pedrosa et al., 2013, p. 2). It is therefore necessary to create more corpora suitable for didactical use in the frame of their content and design to enhance their use in language instruction (McEnery & Xiao, 2011, pp. 374-375). The direct use of corpora in the instruction is still less frequent in comparison to its indirect implications, therefore it is necessary to increase the extent of the employed data-driven learning approaches in the process of Chinese language acquisition and start using more computer-assisted software tools that facilitate the learning process (Xu, 2019, p. 47). When working with corpora, it is important to keep in mind that the search results are restricted to the findings in that particular corpus, which is processed by different tools (e.g., the text is tokenized and annotated). These factors also have an impact on the results extracted from the corpus (Gajdoš, 2020, p. 122).

Based on the variety of introduced approaches aimed at a direct use of corpora, we expect that the use of different Chinese corpora becomes an inseparable part of a teaching and learning practice in the near future, especially with successive attempts to simplify procedures related to the creation of corpora thanks to fully automated or semi-automated processes.

4   Indirect use of Chinese corpora

The indirect use of corpora is often linked to the revision of an existing curriculum or to designing a new one, as well as the creation of teaching materials. The advantage of corpus-based teaching materials is that the examples reflect real language utterances.

We expect that learners improve their communication skills when using this kind of teaching materials because they get used to the language utterances of the native speakers and the distinction of specific nuances in Chinese becomes more evident (McEnery & Xiao, 2011, pp. 367-368). The traditional compilation of teaching materials used to be based mainly on the individual preference, language intuition, and teaching experience of the author. In comparison to the teaching materials created with the use of corpora, the research aimed at the content of the existing teaching materials compiled with the traditional method showed that in many cases, descriptions of words or expression utterances are inconsistent with the natural language usage (Luo, 2008, p. 48). The frequency lists of morphemes, expressions, and grammar structures are one of the main determining features of corpus-aided teaching materials. The textbook's difficulty level is easily controllable using corpus functionalities, and the arrangement of the content is considered more reasonable. We determine the textbook's difficulty level by the text's length, the core grammar structures to be acquired as well as by the ratio of the new vocabulary and the total word count. In addition, it is possible to adjust the occurrence rate of more common words or increase the reoccurrence rate of new words (Zhao & Kang, 2017, p. 85).

Chi-Editor (*Hanyu yuedu fenji zhinanzhen* 汉语阅读分级指难针) accessible on <langaugedata.net/editor> represents a useful tool for assessing the difficulty of the textual material chosen for the instruction practice. Its functionality enables a user to retrieve language material that is sorted out based on individual proficiency levels together with the vocabulary lists corresponding to the requirements of HSK, as well as Chinese syntactic and lexical structures (Jin et al., 2018). Figure 3 (Jin et al., 2018, p. 7) displays a sample text with different colors according to varying difficulty of the vocabulary. As the authors of *Chi-Editor* state, the proficiency levels are in accordance with *International Curriculum for Chinese Language Education* published by Confucius Institute Headquarters in 2015 (Jin et al., 2018). As revisions of the HSK exams have been announced, we expect certain adjustments and updates to follow. At this point, we consider *Chi-Editor* as a suitable tool for the regular preparation of teaching materials, and its use is not restricted by the potential absence of an update in the future.

☑超纲词　☑更高级词　☑高级词　☑中级词　☑初级词　☑专有名词　☑HSK词汇　　　　显示标注

原文

四合院 是 老 北京 人 居住 的 主要 建筑 形式 。四合院 是 由 东、南、西、北 四面 房子 围起来 形成 的 内院式 住宅 。它 是 封闭 的,对外 只有 一个 大门,非常 适合 一家人 居住 。四合院 里 四周 的 房屋 各自 独立,房子 的 门 都 向 院内 开 。四合院 中,北面 的 房子 最好,叫做 正房,是 家中 的 长辈 居住 的 地方 。东西 两边 的 房子 比 北房 差 一些,是 晚辈们 居住 的 地方 。四合院 大小 不一样,有 的 相差 很大 。由 四面 房子 围起 一个 院子 是 四合院 的 一个 单元,有 的 四合院 有 四个 或者 五个 这样的 单元,多 的 还有 九个 单元 的 大 。四合院 北京 的 四合院 很 讲究 绿化。院内 的 空地 上 常常 种树、种花、种草 。一般 在 正房 前 左右,两边 各种 两棵树,有 海棠 树、石榴树、枣树 等 。春天 这些 树 会 开出 漂亮 的 花,夏天 人们 可以 在 树下 乘凉,秋天 树上 结 出 很多 水果 。四合院 中 花草 很多,有 的 大 四合院 还 另有 花园。每年 春、夏、秋 三季,园中 的 鲜花 五颜六色,十分 漂亮。北京 的 四合院 已经 有 几百年 的 历史 了,它 留给 人们 的 不 只是 一座座 建筑,而且 还是 宝贵 的 文化遗产。

Figure 3: Text sample processed by *Chi-Editor*

When focusing on advanced learners of Chinese, there is still a rather limited number of studies devoted to them and their academic Chinese as a target language. Learners' corpora composed of texts written by students at the advanced level depicts learners' needs in the frame of creating new teaching material (Chen & Tao, 2019, p. 58).

Not only the Chinese learner corpus but also the corpus used for the analysis of existing teaching materials such as *Duiwai Hanyu jiaocai yuliaoku* 对外汉语教材语料库 created by National Language Resources Monitoring and Research Center at the Xiamen University represents a helpful tool. It contains texts taken from widely used teaching materials for international students of Chinese that were published between 1992 and 2006. Based on the above-mentioned corpus, researchers or teachers are able to study and assess the content of the existing teaching materials (Su et al., 2010) and conclude whether rearrangements are necessary to increase teaching efficiency. The corpus is accessible on the website of the research center and provides a complete list of included teaching materials. The study of the existing materials is necessary for the verification of the teaching materials' compatibility with the curriculum requirements. The corpus composed of texts from existing teaching materials is also useful for the retrieval of studied vocabulary or structures to display all entries in the corpus with corresponding examples in the context to strengthen the learner's knowledge (Luo, 2008, p. 48). In the more recent corpus linguistics research, there are increasing tendencies aimed at the automation of processes facilitating the use of corpora. It is possible to sort out the language material in the corpus based on the proficiency level. An evident benefit of corpus-aided teaching materials is the natural language providing examples for vocabulary lists enhancing the acquisition of authentic language by the learners (Zhao & Kang, 2017, pp. 85-86). In addition to traditional methods, it is possible to supplement the exercise design based on the findings from

the corpora which focus on the use of collocations and lexical bundles (Chen & Tao, 2019, p. 64).

The indirect use of corpora is also applicable in the field of lexicography, and numerous existing dictionaries compiled by the use of corpora continue to grow. One of the examples is *A frequency dictionary of Mandarin Chinese: Core vocabulary for learners* compiled by Xiao et al. (2009), *Frequency-based HSK vocabulary* by Yang Ying (2016), covering all six proficiency levels in the individual volumes, and the vocabulary is supplemented by example sentences.

The indirect use of corpora for the creation of teaching materials or frequency dictionaries not only spares the time needed for their compilation but also increases the relevance of the selected content. Corpora composed of teaching materials such as the one created at the Xiamen University, or the tool *Chi-Editor*, facilitate the selection of appropriate teaching materials. In case that the employed teaching materials are not identical with the processed textbooks in *Duiwai Hanyu jiaocai yuliaoku* 对外汉语教材语料库, it might serve as a suitable model for teachers who create corpora composed of their own teaching materials to facilitate the analysis.

5   Conclusion

The direct use of corpora in the instruction is beneficial to teachers and students because they are transformed into active participants of the instruction practice. Students learn how to work with corpora and its features and gain experience in interpreting the query results. As experienced corpus users, learners obtain the opportunity to improve their proficiency level based on the examples of native speakers reflecting the natural language production of Chinese. The corpora as a part of translation seminars enhance the students' translation skills and quality of translation. Corpora besides enables learners to concentrate on a specific field of Chinese and teach how to explore the regularities of the target language. The indirect use of corpora is mainly employed for the creation of teaching materials, dictionaries, and curricula, providing the students with a reliable source of native speaker-like language utterances. From the perspective of the authors of teaching materials, the processing time is relatively short and the contents are easily controlled using corpus functionalities with the objective to increase the efficiency of the acquisition process. Taking into consideration a limited number of teaching materials in the Slovak – Chinese language combination, the continuous development of the use of corpora (both direct and indirect) not only facilitates linguistic research but improves the instruction practice, and as such represents an important source for reference.

Acknowledgments

Abbreviations

DDL   Data Driven Learning
KWIC   Key Word in Context
CQL   Corpus Query Language
HSK   Hanyu shuiping kaoshi

References

Benko, V., Butašová, A., Lalinská, M., Paľová, M., Puchovská Z., Segretain, A. & Zeleňáková, M. (2019). *Webové Korpusy Aranea : Učebnica pre Učiteľov Cudzích Jazykov, Prekladateľov, Tlmočníkov, Filológov a Študentov Filologických Odborov*. Bratislava: Univerzita Komenského.

Bernardini, S. (2004). Corpora in the Classroom. An Overview and Some Reflections on Future Developments. In J. McHardy Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp.15-38). Amsterdam: Benjamins.

Bluemel, B. (2019). Pedagogical Applications of Chinese Parallel Corpora. In X. Lu & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp.81-98). Singapore: Springer.

Casas-Pedrosa, A.V., Fernández-Domínguez, J. & Alcaraz-Sintes , A. (2013). Introduction: the Use of Corpora for Language Teaching and Learning. *Research in Corpus Linguistics*, *1*, 1–5.

CCL Han-Ying shuangyu yuliaoku CCL 汉英双语语料库 [CCL Chinese-English Parallel Corpus of Center for Chinese linguistics PKU]. Retrieved from http://ccl.pku.edu.cn:8080/ccl_corpus/index_bi.jsp.

Chen, H.-J. & Tao, H.-Y. (2019). Academic Chinese: From Corpora to Language Teaching. In X. Lu & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp. 57-79). Singapore: Springer.

Chen, Y. (2008). *Chinese-English/English-Chinese Pocket Legal Dictionary*. New York: Hippocrene Books.

Cowie, A. (1994). Phraselogy. *The Encyclopaedia of Language and Linguistics*, *6*, 3168–3171.

Gajdoš, Ľ. (2017). Chinese Legal Texts – Quantitative Description. *Acta Linguistica Asiatica*, 7(1), 77-87. https://doi.org/10.4312/ala.7.1.77-87.

Gajdoš, Ľ. (2020). Verb Collocations in Chinese- Retrieving, Visualization and Analysis of Corpus Data. *Studia Orientalia Slovaca*, 19(1), 121-138.

Guan, X.-C. 管新潮 & Tao, Y.-L. 陶友兰. (2017). *Yuliaoku yu fanyixue 语料库语翻译学* [Corpus and Translation]. Shanghai: Fudan daxue chubanshe.

Harris, A. (2006). Revisiting Anaphoric Islands. *Language, 82*(1), 114–30.

Hu, K.-B. 胡开宝. (2011). *Yuliaoku fanyi jiaoxue gailun 语料库翻译教学概论* [An Introduction to Corpus Translation Teaching]. Shanghai: Shanghai Jiaotong Daxue chubanshe.

Jin, T. 金檀, Lu, X. 陆小飞, Lin, Y. 林筠, & Li, B. 李百川. (2018). Hanyu yuedu fenji zhinanzhen：shiyong shouce 汉语阅读分级指难针：使用手册. Retrieved from https://www.languagedata.net/editor/manual.pdf.

Jing-Schmidt, Z. (2019). Corpus and Computational Methods for Usage-Based Chinese Language Learning: Toward a Professional Multilingualism. In X. Lu & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp. 13-31). Singapore: Springer.

Johns, T. (1991). Should You Be Persuaded: Two Examples of Data Driven Learning. In T. Johns & P. King (Eds.), *Classroom Concordancing* (pp. 1-13). Birmingham: ELR.

Lancaster Corpus of Mandarin Chinese. Retrieved from http://corpus.leeds.ac.uk/query-zh.html.

Luo, L. 骆琳. (2008). Guanyu jianli Hanyu xuexizhe yuliaoku de sikao 关于建立汉语学习者语料库的思考. *Gaodeng gongcheng yanjiu 高等工程教育研究*, 47-49.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. New York: Cambridge University Press.

McEnery, T. & Xiao, R. (2011). What Corpora Can Offer in Language Teaching and Learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 364-380). London: Routledge.

Sinclair, J., Jones, S., Daley, R., & Krishnamurthy, R. (2004). *English Collocational Studies: The OSTI Report*. London: Continuum.

Su, X.-C. 苏新春, Tang, S.-Y. 唐师瑶, Luo, C.-Y. 罗春英 & Hong, G.-Z. 洪桂治. (2010). Jiaocai yuyan tongji yanjiu de duoweidu gongneng 教材语言统计研究的多维度多功能. *Proceedings of the Innovation of International Chinese Teaching Theories and Models Conference*.

Tsui, A. (2004). What Teachers Have Always Wanted to Know – and How Corpora Can Help. In J. McHardy Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 39-61). Amsterdam: Benjamins.

Xiao, R., Rayson, P. & Mcenery, T. (2009). *A frequency dictionary of Mandarin Chinese: Core vocabulary for learners*. London: Routledge.

Xu, J. (2019). The Corpus Approach to the Teaching and Learning of Chinese as an L1 and an L2 in Retrospect. In X. Lu & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp. 33-53). Singapore: Springer.

Yang, Y. (2016). *Frequency-based HSK vocabulary*. Beijing: Sinolingua.

Zhao, L.-Z. 赵连振 & Zhang, G.-J. 张桂军. (2015). Guonei yuliaoku fuzhu waiyucihui jiaoxue wenxian zongshu 国内语料库辅助外语词汇教学文献综述. *Yaoxue jiaoyu 药学教育*, 31(2), 34-39.

Zhao, X. 赵星 & Kang, D.-M. 康冬梅. (2017). Yuliaoku zai Hanyu jiaoxue zhong de yingyong tanxi 语料库在汉语教学中的应用探析. *Xiandai yuwen 现代语文*, 7, 85-87.