

“Praktická korpusová lingvistika – čínsky jazyk”: Book review

Mateja PETROVČIČ

University of Ljubljana, Slovenia

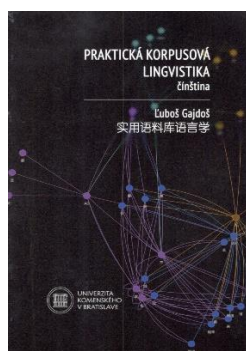
mateja.petrovcic@ff.uni-lj.si

Povzetek

Knjiga *Praktická korpusová lingvistika – čínština* (Praktično korpusno jezikoslovje – kitajščina) Ľuboša Gajdoša ponuja vpogled v uporabo kitajskih korpusov, pri čemer avtor k tematiki pristopi z vidika uporabnika. Monografija je tako neprecenljiv vir za vse, ki želijo bolje razumeti kitajski jezik. Delo najprej predstavi ključne izraze in kategorije v kitajski slovnici ter korpusnem jezikoslovju, pri čemer bralca sistematično popelje od osnovne do naprednejše rabe korpusov pri učenju jezika in nadaljnjem raziskovanju njegovih značilnosti. Kot sistematični vodnik, ki korak za korakom vodi od enostavnih do kompleksnejših vsebin, je ta knjiga zelo priporočljivo gradivo za bralce, ki že imajo določeno predznanje kitajskega jezika, vendar so na področju korpusnega jezikoslovja popolni začetniki ali srednje izkušeni uporabniki.

Summary

The book *Praktická korpusová lingvistika - čínština* (Practical Corpus Linguistics - Chinese) by Ľuboš Gajdoš provides insight into the use of Chinese corpora from a user-oriented viewpoint and is an invaluable resource for anyone who wants to gain a better understanding of the Chinese language. It introduces the key concepts and categories of Chinese grammar and corpus linguistics, and systematically guides the user from the basics to the intermediate use of corpora in both language learning and research. As a systematic step-by-step guide, it is highly recommended for readers with prior knowledge of Chinese, as well as for beginners and intermediate users in the field of corpus linguistics.



Author: Ľuboš Gajdoš

Title: Praktická korpusová lingvistika – čínština

Title in Chinese: 实用语料库语言学

Publisher: Univerzita Komenského Bratislava

Year: 2022

ISBN 978-80-223-5363-2

Language: Slovak

Pages: 164

Cover: Paperback

Size: 170 x 240 x 10 mm

Weight: 288 g

Acta Linguistica Asiatica, 13(1), 2023.

ISSN: 2232-3317, <http://revije.ff.uni-lj.si/ala/>

DOI: 10.4312/ala.13.1.119-122



Review

The book opens with a brief description of parts of speech in Chinese, especially from the perspective of morphological annotation and its features. Establishing a common understanding of word classes in Chinese is important and closely related to the issue of tagsets and tagging in corpora. From this point of view, Chapter 1 provides a bridge between the purely linguistic perspective and the corpus linguistic approach to a language. In the subchapters, the author briefly outlines the eleven categories of Chinese, points out selected details and relates them to the Slovak language where necessary. The eleven parts of speech are outlined quite briefly, which is fine for a beginner, but more demanding users might wish that more details were presented here.

The author assumes that a reader who picks up this book will already have some prior knowledge of Chinese language and grammar. Therefore, Chapter 2 covers the constituents of a sentence clause and the grammatical relationships between them. Following the previous chapter, the content is presented from the Chinese perspective, which is also used in Chinese didactics. This section concludes with a schematic representation of typical word order in Chinese.

Chapter 3 proceeds to the key concepts of corpus linguistics and thus provides essential information on the basic concepts such as *token*, *tag*, *concordance*, *collocation*, and *regular expression*, explaining each of these concepts with some examples from Chinese.

The following chapters are devoted to the description of the Chinese corpora, their access details, and the corresponding user interfaces. The parameters for each corpus are presented in the form of a table to provide easy insight into the characteristics of the selected corpora. To be more specific, Chapter 4 discusses the CCL corpus, Chapter 5 focuses on the CNC corpus, Chapter 6 explores the BCC corpus, Chapter 7 moves on to the Sihanku corpus, and Chapter 8 concentrates on the Hanku corpus.

Subsections that examine the characteristics of each corpus and their advantages and disadvantages are an important part of chapters 4-8. Here Gajdoš's book shows a wide range of possible applications. The author encourages a reader to include corpora in language teaching and consider them as tools for the in-depth study of language use.

Since the Hanku corpus uses the NoSketch Engine management system, i.e., a free version of its powerful commercial counterpart, Sketch Engine, the second part of Chapter 8 focuses on query options and available functions in the side menu. A reader who is not very familiar with corpora annotation might find Table 11 particularly useful, as it provides information about the Hanku tagset, the corresponding terms in Chinese and Slovak, and a lexical example in Chinese for each item.

Chapter 9 forms the core of the book and is therefore the longest section at nearly forty pages. Its focus is placed on the corpus query language (CQL), which offers a variety of search options. The author asserts that although the notation may seem complicated and abstract at first glance, it is very logical and easy to learn, even for those who do not work with corpora daily.

The reader who is learning the basics of CQL will find Table 12 as one of the spots where to add a bookmark for the future. Namely, this is the place where all metacharacters are explained and briefly illustrated with examples from Chinese. These characters have a special meaning and use within the Sketch Engine management system. To be able to take advantage of CQL, one must first learn what the different types of parentheses, special characters, and letters mean. After clarifying what the *word*, *lemma*, and *tag* attributes mean, Table 13 shows examples of the use of wildcard characters in CQL queries. If a desired expression can be formulated in more than one way, all possibilities are listed there for consideration.

At this point, the author presents some exercises, gives the solutions to the tasks set, and discusses the results of a query. Throughout the book, 34 tasks are presented, ranging from simple to more complex ones.

This chapter also explains how to search the left and right context, which is possible only with the operators *meet*, *union*, *containing*, and *within*. More complex and advanced examples are additionally provided with graphical representations to ensure that a reader understands the 'technical mindset' required to formulate such query strings. At this stage, for example, the user should be able to explore complex postverbal structures or verb complements known in Chinese as *buyu* 补语, questions related to prepositional phrases, *ba*-constructions (*baziju* 把字句), *bei*-constructions (*beiziju* 被字句), and others. Exact explanations of the operators *meet* and *union* can already be found in the Sketch Engine documentation, but since the examples are given for English language, it might be challenging to apply them in Chinese. This is one of the places where the book *Praktická korpusová lingvistika - čínština* (Practical Corpus Linguistics - Chinese) by Ľuboš Gajdoš comes to the rescue.

The last section of Chapter 9 introduces token comparison in a query, whereby global conditions are set for the individual tokens. Chapter 10 brings queries to the next level, where the author shows how to take advantage of advanced search options that combine multiple conditions into a complex query string.

Chapter 11 is aimed at a reader who already has some understanding of Chinese grammar and would like to deepen their knowledge of grammatical relationships between the constituents of a clause. This chapter explores the possibilities and limitations of regular expressions in unveiling the syntactic features of Chinese. The author demonstrates how to define and extract a sentence object (*binyu* 宾语), adverbial adjuncts (*zhuangyu* 状语), six kinds of verbal complements (*buyu* 补语), a

subject (*zhuyu* 主语), and attributives (*dingyu* 定语). This is the background knowledge of what we call *word sketches* in the commercial version of Sketch Engine, a one-page summary of the grammatical and collocational behavior of a word. It is very convenient and easy to get an overview of the grammatical relationships of a desired word with a few clicks, however, getting there from scratch is another thing.

The book closes with a glossary of selected terms in Chinese corpus linguistics, references, an excerpt from the Slovak-Chinese parallel corpus Sihanku, an ordered list of the 34 tasks, and an index. The author's concluding remarks indicate that "Part 2" may be compiled in the future:

Although in this publication I have focused mainly on didactics and linguistic research of the Chinese language, I believe that the acquired knowledge about the use of language corpora can also be used for translation purposes from and into Chinese (Gajdoš, 2022, p. 135).

Praktická korpusová lingvistika – čínština (Practical Corpus Linguistics – Chinese) by Ľuboš Gajdoš is an invaluable resource for anyone who wants to improve their understanding of Chinese through the corpus-linguistic perspective. It provides an illustrative step-by-step explanation of how Chinese corpora are used and applied. Although this book is written in Slovak, it is easy to understand for speakers of other languages as well. The book contains chapters on topics such as Chinese grammar from a Chinese perspective, Chinese language corpora, basics of corpus linguistics, and researching the language using a corpus query language. Each chapter provides brief but comprehensive information and examples that illustrate the value of these corpora in actual use. After reading in depth and testing the queries on one's own, this book offers a reader the opportunity to explore the nuances of the Chinese language and its usage based on real language data. The trump card of this publication is the rich collection of examples that encourage users to formulate queries according to their own preferences and interests. Overall, this book is a must-have for anyone studying or researching the Chinese language and striving to test the information on large language corpora. The book is also an excellent reference for teachers of Chinese to prepare representative language examples.

References

- Gajdoš, L. (2022). *Praktická korpusová lingvistika – čínsky jazyk*. Bratislava: Univerzita Komenského Bratislava.
- CQL – meet & union | Sketch Engine. (2017, February 9).
<https://www.sketchengine.eu/documentation/cql-meet-union/>