

# 语料库驱动的捷克语汉源词研究

卢子玥

浙江外国语学院西方语言文化学院, 浙江 杭州

收稿日期: 2024年10月8日; 录用日期: 2024年11月12日; 发布日期: 2024年11月26日

## 摘要

汉源词是中国文化进入异域的重要载体, 同时是中国文化传播的重要体现。前期汉语词研究和中国文化研究多依托定性分析, 定量研究或两者结合的研究并不多见。本文以语料库驱动的研究方法, 拓展捷克语中的汉源词的研究, 考察中国文化在捷克的影响。首先分析捷克语中汉语借词的研究情况和基本特点, 进而利用语料库资源挖掘更多汉源词的存在。语言是文化的反映, 基于数据探索中国文化的异域呈现, 能够开拓当前中国文化传播研究新领域。

## 关键词

捷克语, 汉源词, 语料库

# A Corpus-Driven Study of Chinese-Origin Words in Czech Language

Ziyue Lu

School of European Languages and Cultures, Zhejiang International Studies University, Hangzhou Zhejiang

Received: Oct. 8<sup>th</sup>, 2024; accepted: Nov. 12<sup>th</sup>, 2024; published: Nov. 26<sup>th</sup>, 2024

## Abstract

Chinese-origin words act as conduits for the transmission of Chinese culture into foreign contexts, representing a crucial aspect of its dissemination. While existing research on these words and their connection to Chinese culture has predominantly relied on qualitative analysis, there is a significant lack of quantitative studies or those that integrate both qualitative and quantitative methodologies. This paper employs a corpus-driven approach to deepen the investigation of Chinese-origin words in the Czech language and to assess the influence of Chinese culture in the Czech Republic. The study first evaluates the current status and core characteristics of Chinese loanwords in Czech. It then utilizes corpus resources to identify additional Czech vocabulary derived from Chinese. As language

reflects culture, exploring the representation of Chinese culture in foreign countries through empirical data can open new avenues for research on the global spread of Chinese cultural elements.

## Keywords

Czech Language, Chinese-Origin Words, Corpus

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

当前的中国日益走近世界舞台中央，国际社会对中华文化的关注与日俱增。基于汉源词或汉语借词透视中国文化的研究多限于英语国家及中国周边国家，而中国与中东欧国家的文化关系研究尚处于起步阶段，开展广泛的中国与中东欧地区在文化、社会等领域的研究很有必要。语言是文化传播的媒介，基于语料库的研究可提供较为可靠的中国文化研究数据，助力中国文化软实力的理论研究。

## 2. 语料库驱动下的问题研究

语料库研究是当前中捷两国的热门研究领域之一，捷克国家语料库在捷克学界多应用于捷克语教学研究、捷克语历史研究和词汇研究等，而目前国内学界的捷克语语料库相关研究极为罕见。本课题的主要研究重点有二，一是介绍捷克语中汉源词典借词的基本情况及其与中国文化的关系；二是通过捷克国家语料库进一步检索捷克语中的汉源词，侧重于对汉源描写词及义借词的拓展探究。

### 2.1. 研究资源概述

笔者选取捷克国家语料库中的捷克语历时语料库版本 11 (SYN verze 11, 下文中简称为 SYN\_v11)作为语料来源，其语料丰富、规模庞大，是本研究启动前捷克国家语料库中捷克语历时语料库的最新版本。SYN\_v11 具有参考性，以捷克语新闻文本为主、科研文本及文学文本为辅，文本标记数为 6,067,313,960，词量为 50.32 亿。新闻基于现实事物创作，文本中的信息具有时效性，并以产生尽可能广的传播面而实现价值。该语料库涵盖 *Deník*, *Lidové noviny*, *Mladá fronta Dnes*, *Blesk* 等捷克主流报刊 1990~2021 年的文本数据，以 2004~2021 年的报刊文本为主[1]。

捷克国家语料库的功能完备，目前有 15 个语料研究工具，可以开展词汇变异、检索词分布、词汇对比等方面研究，也包含语料库的核心研究工具，如词频、词汇搭配、主题词研究等。笔者主要选用其中的一个工具 *Kontext*，并使用该工具的主功能 *Dotaz* (索引)、*Frekvence* (词频)等进行综合探索。

### 2.2. 语料库研究设计

笔者前期的汉源词研究集中于对词典词的考察，但捷克语中源自汉语的不仅是词典词，还存在进入捷克文化圈而未收录的部分汉源词汇，这些词汇通常是义借词或仿译词，例如捷克语词 *hůlky* 被赋予“(中国)筷子”的含义[2]。

本研究借助捷克国家语料库确认捷克语中汉语义借词、仿译词或词典之外其他汉源词的存在。前期研究仅关注了“词典词”，即汉源词典借词，指的是收录于词典的汉语借词[3]。毋庸置疑，中国文化事物在捷克的命名也有短语形式的呈现，比如中医 *čínská medicína*、长城 *čínská zeď* 等。笔者以“中国的……”

“北京的……”为检索对象，使用语料库 SYN\_v11 中索引功能的子功能“高级索引”，筛选与中国密切相关的汉源词并进一步分析。

### 3. 捷克语词典中的汉源词

#### 3.1. 汉源词的界定

因语言无法自给自足而出现的外来词汇一般指借词，国内学界对借词的定义多有探讨，除了关注狭义范围内的音译借词[4]，对意译词和形译词的研究讨论是借词广义派学者的意愿，这样不仅能开拓当前借词研究的视野、丰富研究角度，还能捕捉到语言、文化和社会之间的深层次关系[5]。

捷克方面将借词看作两种或多种不同语言接触的结果，对借词的分类历来有不同的观点。捷克语术语有 *výpůjčka* 借出词、*přejímka* 借入词、*přejatá slova* 借词、*cizí slova* 外来词等，《(新)捷克语百科词典》以“*výpůjčka (přejímka)*”为搜索词条[6]，可见当前捷克学界对“借词”概念仍有较大探讨空间。捷克语借词在形式上主要有音借词、词态借词、形借词和构词型借词，捷克已有对义借词的考察，其被视作语义及命名两方面外语词汇介入的结果[7]。

萨丕尔将来自外语的借用称为“文化上的借用”，“用来称呼事物或行为的言语形式往往从一个民族传递给另一个民族”，反之，文化上的借用体现了民族间的文明互鉴[8]。详尽研究这些借词，便能追踪不同民族文化交融的历史痕迹，挖掘各个民族在世界思想文化传播中扮演的角色和所起到的作用。

笔者在本文中以更广阔的视野看待语言中出现的外来词汇。课题的主要研究对象“汉源词”是比借词更广泛的语言概念，“源”本义为水源、源泉，引申义是来源、根源。汉源词不仅包括异域语境中的汉语借词(单词)，同时还涉及来源于汉语、与中国文化事物相关的短语。借鉴相关研究中的概念[9]，对汉源词定义如下：

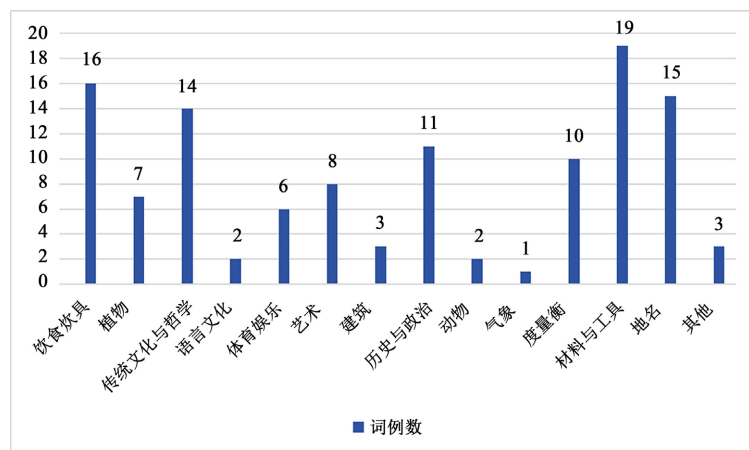
来源于汉语，在语音、语义及拼写方式等方面与汉语词有对应关系的词汇。

#### 3.2. 捷克语汉源词与中国文化

笔者前期研究着眼于词典借词，重点考察四部捷克语权威词典《(新版)外来词学术词典》《捷克语规范词典》《捷克语规范词典(学生及公众专用)》《捷克语便携辞典》中汉源词的呈现情况。这些词典借词融入捷克语的主要方式为音借和混合借入，音借词的主要音源为标准汉语读音，以汉语拼音和捷克式拼音为主要借入依据；混合借词则以词汇派生为主。本节重点从词义角度对捷克语中的汉语借词展开实例分析。

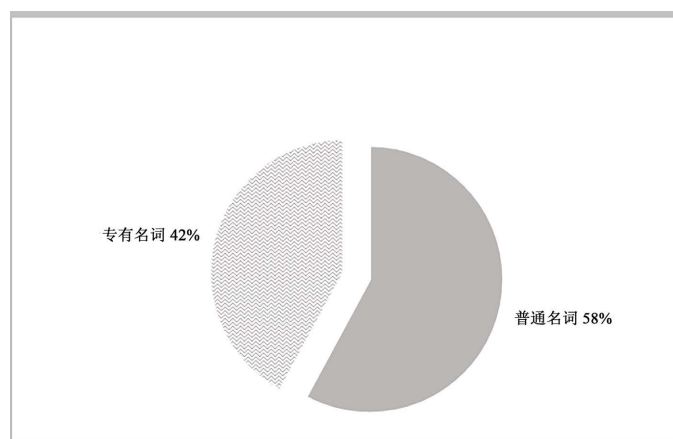
基于研究发现，按照语义，捷克语中的汉源词典词(114个)主要可以分为14类：饮食炊具、植物、传统文化与哲学、语言文化、体育娱乐、艺术、建筑、历史与政治、动物、气象、度量衡、材料与工具、地名、其他。这些词汇语义分布的具体情况(见图1)。

图中可见，材料与工具类借词(19个，占17%)最多，这类词所涉及内容包括产自中国的“高岭土 *kaolin*”“北京绸子 *peking*”“南京棉布 *nankin*”及其派生词等，还包含中国的古代水上交通工具“舢板 *sampan*”。其次便是饮食炊具类词汇(16个，占14%)居多，这类词汇以具有较强派生能力的“茶 *čaj*”为首，*čaj* 一词派生出 *čajovar* (泡茶机)、*čajník* (茶壶)、*čajovka* (茶花)、*čajový* (茶的)。值得注意的是，各类茶叶的名称并没有收录于捷克语词典中，茶叶的命名在捷克语中的形式也难以统一界定，比如龙井茶在捷克语中以 *Long Jing* 和 *Dračí studna* 的形式出现，前者是龙井的拼音式音译，后者是意译的“龙”和“井”，除此之外，还有英文名称 *Tiger Spring* 的使用。同时，饮食炊具类词汇含有中国炊具“镬、锅 *wok*”。除专有名词(即中国地名)之外，占据第三位的汉语借词便是传统文化与哲学类词汇(14个，占12%)，如“风水 *feng-šui*”“阴 *jin*”“阳 *jang*”“道 *tao*”等。



**Figure 1.** The semantic distribution of Chinese loanwords in Czech  
**图 1.** 捷克语中汉语借词的语义分布

从语义角度更进一步钻探，笔者留意到捷克语中的汉源词在很大程度上与中国专有名词有关，并存在专有名词普通化现象。笔者按照词汇含义与汉语普通名词、专有名词的相关度，将这些词汇分入“普通名词”和“专有名词”图层，据图 2 显示，共有 48 词与中国专有名词相关，占捷克语中汉源词典借词的 42%。



**Figure 2.** The distribution of Chinese loanwords in Czech related to common nouns and proper nouns  
**图 2.** 捷克语中汉语借词的普通名词与专有名词图层分布

捷克语中首字母大写的 Šantung 表示(中国)山东，为指称地名的专有名词，单词 šantung 指的是“山东绸”。与之类似的有 nankin(南京布)，据捷克词汇档案记载，该词最早于 19 世纪初(约 1826 年)出现在捷克语中，记载显示 nankin 又写作 nankýn，阳性名词，一种由南京传入捷克的轻薄面料，之后此布料较为普遍[10]。

汉源人名的普通化不似地名简单，而是借助捷克语语素生成混合式借词，比如捷克语中的 maoista 意为“毛泽东思想的追随者”，由 mao 加词缀-ista 生成；maoismus 意为“毛泽东思想”，由 mao 加国际化词缀-ismus 生成。这类现象的出现被看作“语言经济原则”的结果，以专有名词词义概念二元性为重要基础，在这一过程中转喻扮演着重要角色[11]。

捷克语中的汉语借词与中国文化事物息息相关，无论从语音还是语义上都有很强的“中华性”。因此，捷克语中的汉源词本身具有超强中国文化相关性，这类词汇在捷克语中的运用对中国文化的传播有一定的呈现度。

## 4. 捷克语中的汉源词新探

### 4.1. 语料库索引范围

接触语言学是借词研究的常见视角，从这一角度来看，借词研究实践集中在词典借词方面，对外来词借入的“动态过程”的关注则拓展了借词的研究范围，将“前借词”与“后造词”囊括在内[12]。借词的动态过程论亦为本课题的重要研究理论支撑，正如上文所说，本研究不仅聚焦于捷克语中的汉源词典词，还关注其他汉源词在捷克语中存在的可能性。与词典借词不同，其他汉源词汇的捷克语形式较为混杂，难以归总，或应实际命名需要而生，或为中捷文化交融的结果。笔者前期研究拘于词典，而忽略了另外一类汉源混合借词的存在，这类借词以描写中国事物为主，是语言应社会实际需求的发展结果，构成方式多为“来源地+事物”，被称为“描写词”。

此外，汉源词新探过程中义借词不可或缺，即仿译词，捷克语称之为 *kalky*。国内亦有将引申和仿译归为义借词的说法，捷克方面认为 *kalky* 是一类特殊的借词，主要有三类：

构词上的仿译：比如 *počít-ač (comput-er)*, *monitor-ování (monitor-ing)*；

多词仿译：比如 *horské kolo* 便是“模仿”了英文短语 *mountain bike*；

词意上的仿译：比如 *myš* 本意为“老鼠”，也指“鼠标”。

基于引申词与借出语的低关联性，捷克学者并不赞成义借词包括引申词的这一说法，比如捷克语中 *drak* 意为“龙”，*asijský drak* (本意“亚洲龙”)则引申为“发展迅速的亚洲地区或国家，香港、新加坡等”，这一说法源于 20 世纪通行于亚洲国家的“亚洲四小龙”概念，而与汉语本身的相关性不高。本课题的汉源词新探则主要讨论捷克语中的汉语仿译词，已有研究指出外语对汉语的仿译形式多为两词、三词短语，本文以“中国的……”“北京的……”系列短语为主要索引对象。

### 4.2. 语料库检索结果及分析

为清晰语料数据研究范围，笔者将词例按照词频由高到低排序，因不对词汇的词形变化展开分析，文本类型设置为“lemma (词元)”，并将词频最高的前 100 例划定为研究范围。由于语料库操作的局限性，研究数据中会出现不完整的短语，比如查询到的三词短语 *čínský císař Čchin*，完整短语应为 *čínský císař Čchin Š'chuan-ti* (中国皇帝秦始皇)，所以笔者借助语料库自带过滤设置“*Použít v pozitivním filtru*”进一步筛选汉源词汇，人工查看并过滤出汉源词汇在捷克语中出现的形式，因捷克语的屈折语特性，本研究中的筛选结果以“word”形式呈现。

按照前期研究设计，首先用语料库 SYN\_v11 中的子功能“高级索引”，对[*lemma="čínský"*][*tag="N.\*"*]进行用例查询，共得到 8794 例“中国的……”两词短语搭配。笔者发现以下词汇(按照相对词频从高到低排列)，见表 1：

**Table 1.** Search results for two-word phrases “Chinese...”  
**表 1.** “中国的……”两词短语搭配索引结果

汉源词	含义	相对词频(i.p.m)
<i>čínská medicína</i>	中医	1.61
<i>čínská zed'</i>	长城	0.62

续表

čínská provincie	中国省份	0.48
čínské zelí	小白菜、大白菜	0.36
čínská čtvrť	唐人街	0.31
čínská kuchyně	中餐	0.25
čínský znak	汉字	0.19
čínský drak	龙	0.18
čínský jüan	元(人民币单位)	0.12
čínský porcelán	中国瓷器	0.09
čínské nudle	中国面条	0.09
čínská tradice	中华传统	0.08

其次，笔者对[lemma="čínský"][tag="\*.\*\*"][tag="N.\*"]进行用例查询，“中国的……”三词搭配获得超过 10,000 例。如上文所说，从词频最高的前 100 例中筛选。这样的词汇有(按照相对词频从高到低排列)，见表 2:

**Table 2.** Search results for three-word phrases “Chinese...”

**表 2.** “中国的……”三词短语搭配索引结果

汉源词	含义	相对词频(i.p.m)
Čínská lidová republika	中华人民共和国	0.64
čínský nový rok	新春, 春节	0.16
čínská tušová malba	水墨画	0.03
čínská lunární kalendář	农历, 阴历	0.03
čínská tradiční medicína	中医, 中医学	0.02
Čínská lidová banka	中国人民银行	0.02
Čínská rozvojová banka	国家发展银行	0.01

据语料库数据显示，“中国的……”两词短语的相对词频(i.p.m)整体比三词短语高。短语 *čínská medicína* 的相对词频最高，为 1.61，表示“中医、中医学”的还有 *čínská tradiční medicína*，相比前者该短语多了对“传统(tradiční)”的表达，但是词频却与前者相差甚远，为 0.02，由此可见 *čínská medicína* 在捷克语中更为常见、使用频率更高。

短语 *čínské zelí* (中国白菜)在捷克文化中指小白菜和北京白菜，北京白菜 *pekingské zelí* 即中国通常所说的“大白菜”，*čínské zelí* 亦被称为 *bok čoj* 或 *pak čoj zelí*，后两个词均为音借词，借汉语中“白菜”的发音，也有借粤语中“白菜”读音的说法。在捷克商店卖 *čínské zelí* 的货架上，我们既能看到小白菜，也可能发现卖的是大白菜。在捷克语非正式用语中，有用 *Číňanka* (本意“中国女人”)指称中国大(小)白菜的用法。另外一种表示汉源食物的短语 *čínské nudle* 本意为中国面条，在语料库文本中这一短语所指称的食物各有不同。

下面出现在捷克杂志《面点》(Těstoviny)的这段文本将“粉条”称为“中国面条”：



Čínské nudle “fen-tchiao” se vyrábějí nejčastěji z rýžového škrobu. (中国面条“粉条”通常由米淀粉制成。)

也有直接将“炒面”称为“中国面条”的说法：

Čínským nudlím se tady říká “Chowmein”. (中国面条，这里指“炒面”。)

同理，笔者对[lemma=”pekingský”][tag=”N.\*”]和[lemma=”pekingský”][tag=”.\*”][tag=”N.\*”]进行用例查询，搜索到 1169 例“北京的……”两词搭配及 1486 例三词短语。按照上述步骤，笔者筛选到以下词汇(按照相对词频从高到低排列)，见表 3：

**Table 3.** Search results for two-word and three-word phrases “Peking...”

**表 3.** “北京的……”两词、三词短语搭配索引结果

汉源词	含义	相对词频(i.p.m)
pekingské zelí	大白菜	0.12
Pekingská univerzita	北京大学	0.12
pekingská opera	京剧	0.11
pekingská kachna	北京烤鸭	0.07
pekingská kuchyně	北京菜	0.003
Pekingské náměstí Tchien-an-men	北京天安门广场	0.09
pekingský palácový psík	北京狗、狮子狗	0.02
pekingské Zakázané město	紫禁城	0.02
pekingské Ptačí hnízdo	北京鸟巢	0.01
pekingský Palác lidu	北京人民大会堂	0.01
Pekingská univerzita Čching-chua	清华大学	0.004
pekingský císařský palác	故宫	0.002
pekingská Lidová univerzita	中国人民大学	0.002
Pekingská sportovní univerzita	北京体育大学	0.002
Pekingská pedagogická univerzita	北京师范大学	0.001
Pekingská technická univerzita	北京科技大学	0.001
pekingský Letní palác	圆明园	0.001
pekingský Chrám nebes	天坛	0.001

“北京的……”系列词汇中真正的汉源词不多，带有北京地域特色的专有名词反而占比更多，其中包括北京历史古迹和北京高校的名称，语料库中具体搜寻到七所北京高校的名称。有汉语词源、带中华文化特点的普通名词主要涉及食物、戏曲和动植物三类。

笔者对语料库中所提取到的语料数据展开分析，属于描写词的有 čínská medicína 中医，čínské zelí 小白菜、大白菜，čínský porcelán 中国瓷器，čínský nový rok 新春、春节，čínská lunární kalendář 农历、阴历，pekingská kachna 北京烤鸭等，属于义借词的有 Čínská lidová republika 中华人民共和国，pekingské Zakázané město 紫禁城，pekingský císařský palác 故宫等。本研究以“中国的……”“北京的……”系列短语为研究实例，其中描写词占很大一部分，这类词语是捷克语对汉源新生事物的描述，通过表明其元文化而对这一事物定性，同时这些词语又是中国文化在异域渗透的有力工具。汉语中也常有“胡、番、

西、洋”等语素表示“舶来品”，比如胡麻、胡萝卜、番茄、洋葱、西蓝花等。

捷克语中为描写中华文化事物而生的词汇愈来愈多，将类似的词汇尽数列出毫无可能，也非本研究的目的，本文主要从宏观角度探析汉源词在捷克语中存在的可行性。当前捷克语中也出现很多来源于中国的新词汇、新名称，比如 Pás a stezka 一带一路、Hedvábná stezka 丝绸之路、perlový čaj 珍珠奶茶、horký kotlík/horký vývar hot pot 火锅等，这些词汇尚未收录于捷克语词典，但却是中国文化、中国故事渗透捷克的体现。

## 5. 结语

语言是反映国家文化的一面镜子，词汇之间所体现的文化借用正是民族文明之互鉴。本文基于捷克语中汉源词典的前期研究，首先剖析了捷克语中汉源词典借词的特点及其对中国文化的呈现。相比于其他欧洲国家语言和斯拉夫语言对捷克语的影响，汉语借入捷克语的词汇不多，以 čaj (茶) 为典型词汇，其词汇衍生力也是最强的。捷克语中的汉源词典词整体对中国文化有较高呈现度，从语义角度可以看出这些借词与中国特色文化有很强的联系。词典词中不乏有专有名词的出现，如中国地名，也存在不少专有名词普通化的现象，这种现象或会直接影响捷克人对中国某位名人或某个地方的第一认识。

本文与学界已有的汉语借词研究有所不同，将定量分析与定性分析有效结合，并选用语料库驱动的研究方法。基于捷克国家语料库的检索，笔者筛选出捷克语中更多的汉源词汇，形式上看这些词主要以短语为主，语义上来说，本研究的语料索引格式多为“中国的……”“北京的……”。最终的检索结果显示该类词汇中描写词占多数，同时存在一部分的义借词。这些词汇在捷克语中的使用频率均较低，但对比发现，捷克社会对中国传统文化、中国美食和中国经济等方面都多有关注。随着中国文化的境外传播，许多中华文化事物进入捷克等欧洲国家，从不认识到认识再到熟知，中国文化走向全世界还有很长的路。

## 基金项目

本文系杭州市外文学会 2023 年研究课题“基于语料库驱动的捷克语汉源词研究及中国文化透视”(立项编号: HWKT2023015)的阶段性成果。

## 参考文献

- [1] Petkevič, J., Procházka, V., Skoumalová, P., Šindlerová, H.J. and Škrabal, M. (2022) Korpus SYN, verze 11 ze 14. 12. <https://www.korpus.cz>
- [2] 北京外国语大学《新捷汉词典》编写组, 编. 新捷汉词典[M]. 第 2 版. 北京: 商务印书馆, 2018.
- [3] Lu, Z. (2020) Přejatá slova z čínštiny v češtině. Master's Thesis, Ostravská Univerzita.
- [4] 史有为. 异文化的使者[M]. 上海: 上海辞书出版社, 2004.
- [5] 徐通锵. 基础语言学教程[M]. 北京: 北京大学出版社, 2001.
- [6] Nekula, M. (2017) Výpůjčka. In: Karlík, P., Nekula, M. and Pleskalová, J., Eds., *CzechEncy-Nový encyklopedický slovník češtiny*. <https://www.czechency.org/slovník/VÝPŮJČKA>
- [7] Svobodová, D. (2015) Současný stav internacionalizace lexika. In: Martinková, M. and Uličný, O., Eds., *Studie k moderní mluvnici češtiny*. 4, *Synamika českého lexika a lexikologie*, Univerzita Palackého v Olomouci, 131-151.
- [8] Bloomfield, L. (1933) *Language*. Holt, Rinehart and Winston.
- [9] 冯雪峰, 吕菲. 汉源词偏误及韩国学生的心理词典[J]. 长江学术, 2008(4): 143-149.
- [10] Hujer, O., Smetánka, E., Weingart, M., Havránek, B., Šmilauer, V. and Ziskal, A. (1911) Dodnes. Novočeský lexikální archiv. <https://bara.ujc.cas.cz/psjc/search.php>
- [11] 雷玉兰. 英语专有名词普通化的认知阐释[J]. 湖南农业大学学报(社会科学版), 2023, 24(1): 94-100.
- [12] 陈胜利. 英语中的汉语借词研究——接触语言学视角[D]: [博士学位论文]. 苏州: 苏州大学, 2014.