TABLE OF CONTENTS 目录

FOREWORD 前言 v

RESEARCH ARTICLES 研究文章

MAXIMILIANO E. ORLANDO: Language Models Plus Specialized Documentation: Building Regular Expressions that Match Chinese Grammar Patterns
奥兰多·马西米利亚诺·埃德华多: 结合语言模型和专业文档 : 构建匹配汉语语法模式的正则表达式 1

庞泓雪 – 解植永: 匈牙利德布勒森市华裔儿童中文传承研究
HONGXUE PANG – ZHIYONG XIE: A Study on Chinese Heritage Language Transmission Among Ethnic Chinese Children in Debrecen, Hungary 25

李雪: 匈牙利高级阶段国际中文教育中古语词习得及教学研究
XUE LI: A Study on the Acquisition and Teaching of Archaic Words in Advanced International Chinese Education in Hungary 39

BALÁZS SZIGETHY: Diachronic Annotation and Network-Graph Structuring in Chinese Character Pedagogy: A Component-Based Framework
孔茗: 汉字教学中的历时注释与网络图结构建构：一种基于部件的教学框架 57

REPORTS 报告

ELTE Confucius Institute
匈牙利罗兰大学孔子学院 71

AUTHORS OF THE VOLUME 本期作者 77

Journal of Chinese Language Teaching in Europe 欧洲中文教育 2026/2, pp. 1–23

Maximiliano E. Orlando

(*Independent Researcher*)

Language Models Plus Specialized Documentation: Building Regular Expressions that Match Chinese Grammar Patterns

Abstract

Previous research had suggested using authentic teaching materials, corpora and regular expressions in Chinese as a Foreign Language (CFL) courses. Where regular expressions are concerned, however, it can be argued that language teachers are not expected to be experts on language programming or on authoring digital applications when it comes to build them. Given the importance of digital text in the study of additional languages nowadays, this case study discussed the use of language models in combination with specialized documentation to learn about the construction of regular expressions. It is hoped this discussion will be useful for both CFL teachers and learners who are interested in learning to use these expressions. The target language is the shì…de (*shì...de* (是…的)) grammar patterns. The existence of different *shì...de* (shì…de (是…的)) grammar patterns made it possible to illustrate the benefits of using language models to build regular expressions that match Chinese grammar patterns. Indeed, in enhancing the accuracy of the technical aspects of regular expressions, language models can also depict nuances in Chinese language text. Nevertheless, these benefits come with challenges, which together with the limitations of this study are tackled in this paper.

Keywords: Chinese language, regular expressions, grammar patterns, *Python*, digital text

maxorl@outlook.com

https://doi.org/10.61637/cltcee.2026.2.1
https://orcid.org/0009-0006-9351-2719

奥兰多·马西米利亚诺·埃德华多

结合语言模型和专业文档：构建匹配汉语语法模式的正则表达式

摘要

在汉语作为第二语言（CFL）教学中，既往研究建议使用真实的教学材料、语料库以及正则表达式。然而，就正则表达式的使用，语言教师无需成为编程或数字应用开发方面的专家。鉴于数字文本在当前的语言教学中的重要性，因此本文探讨了如何结合语言模型和专业文档以生成正则表达式的方法。希望本文的研究能对有意学习使用正则表达式的汉语教师和学习者提供参考。本文以"是…的"语法格式为研究对象。由于"是…的"存在不同格式，这为展示利用语言模型构建匹配汉语语法模式的正则表达式的优势提供了可能。实际上，语言模型不仅能提高正则表达式在技术层面的准确性，还能捕捉汉语文本中的细微差别。然而，这些优势也伴随着挑战。本文讨论了上述挑战与优势的同时，亦指出了本研究的局限性。

关键词： 中文，正则表达式，语法句式，*Python*，数字化课文

Introduction

In a discussion on the benefits of the application of task-based language teaching in the teaching of grammar in Chinese as a foreign language (CFL) contexts, Yuan (2025) argued that in many CFL courses taught in China there is an approach from which the following scenario is expected to result: Learners understand grammar rules, but they find it difficult to use them in real-world situations.[1] Therefore, Yuan (2025) highlighted the benefits of the use of authentic situations in foreign/second language teaching. These benefits have not only been emphasized in task-based language teaching, but also in the lexical approach.

In that approach, "noticing collocations and language patterns" and the use of language corpora[2] to that end play a fundamental role. Where electronic lan-

[1] Yuan 2025: 14.

[2] Torres Ramírez 2012: 249.

guage corpora are concerned, however, their use may require skills in handling concordancers. Indeed, the complexities of the query interface of an electronic corpus may be suspected to deter learners of an additional language who have little or no knowledge of corpus linguistics methodology from using it. In addition, the use of concordancers requires skills in the analyses of their outputs. Despite not making reference to these two situations specifically, Carter and Absalom (2023), for example, reported that some of the members of a group of tertiary students of Italian as an additional language participating in a study on the use of online resources did not employ corpora owing to the complexity of their use.[3] Nonetheless, Carter and Absalom (2023) mentioned other sources of naturally occurring language learners of additional languages can make use of. Two cases in point are online blogs and websites.[4]

Finding a specific character such as *shì* (是) in a corpus of digital text can be easily accomplished by using a find bar. However, a more sophisticated tool may come into play when looking for grammar patterns such as the *shì…de* (是…的) ones. In this paper, a *shì…de* (是…的) grammar pattern is defined as the order in which groups of words are used in conjunction with *shì* (是) and with *de* (的) to perform one of the six functions discussed below, and the tool in question is regular expressions. Nonetheless, teachers of additional languages in general and teachers of Chinese as an additional language in particular are not expected to be experts on language programming or on authoring electronic applications that can later be used by students in the courses they teach. For this reason, it was decided to share a report on an exploratory case study in which the combination of specialized documentation and artificial intelligence facilitated the creation of regular expressions. Indeed, the objective of this paper is to report on the use of artificial intelligence in combination with specialized documentation with a view to constructing regular expressions that match the *shì…de* (是…的) grammar patterns and that can be used in simplified present day Chinese text. Nonetheless, there is a disclaimer to be made. Because the purpose of the present study is to report on the process of designing the regular expressions in question, it is the responsibility of the users of the end result of this study, i.e. of these regular expressions, to verify their reliability and effectiveness and to improve them to cater for these users' specific needs.

---

[3] Carter-Absalom 2023: 59.

[4] Carter-Absalom 2023: 53.

Theoretical background

In a review of pre-electronic and computerized corpora, Xu (2019) observed that the quantitative study of sentence patterns was rarely tackled in pre-computer time, whereas it was underdeveloped in the latter.[5] Nonetheless, Xu (2019) underlined the importance of Shuhua Zhao's *Basic Sentence Patterns of Modern Chinese* project, where sentence types and grammatical categories were dealt with. A case in point was the "focus constructions 是…的 shi…de".[6] Xu (2019) also concluded that where the pedagogical implementation of corpora findings is concerned, there is still a lot of work to be done to adapt those findings "to classroom friendly teaching materials".[7] For example, in a discussion of the benefits of direct use and indirect use of Chinese corpora, Ištvánová (2021) noted "a limited number of teaching materials in the Slovak – Chinese language combination" [8], and went on to observe that direct and indirect use of corpora may come in useful to enhance language instruction. Nonetheless, where the direct use of a given corpus is concerned, Ištvánová (2021) explained that this use entails some training in the handling of the interface of the corpus with a view to facilitating the interpretation of queries.[9]

There is no denying that the use of electronic corpora and concomitantly of the interfaces that accompany them should be considered while teaching an additional language in contexts where the target learners can take advantage of this use.[10] Nevertheless, a traditional way of learning an additional language has been through printed materials, such as textbooks, literary works, and reference works among others. It is well known that nowadays some of these materials also come in electronic format. A case in point is electronic textbooks.

The search for grammar patterns in digital text can be achieved in different ways. One can read digital text with the express purpose of noticing the use of these patterns. A search bar can also come in useful in this case. For example, when studying the use of the *shì…de* (是…的) grammar patterns, a search bar can be employed to look for the occurrences of *shì* (是) or *de* (的). However, the problem that this tool may pose is that since *shì* (是) and *de* (的) are likely to appear in other grammar patterns, the user may be confronted by the time-consuming task of having to discard irrelevant ones. If the goal of a classroom

---

5 Xu 2019: 41.
6 Xu 2019: 41.
7 Xu 2019: 49.
8 Ištvánová 2021: 140.
9 Ištvánová 2021: 132.
10 Chen 2012.

task is strictly limited to the observation of the *shì…de* (是…的) grammar patterns in digital text, then the use of regular expressions is to come in useful.

A regular expression "is a sequence of characters that forms a search pattern"[11], and it allows the user to "specify the rules for the set of possible strings that [the user wants] to match".[12] As regards studies that relate to Chinese, regular expressions have proved to be useful in various contexts. Yuan et al. (2021), for example, used regular expressions to remove private information from the dataset used to build *WuDaoCorpora*. Regular expressions were used to match, for instance, phone numbers and identity numbers. Lau et al. (2024) used a "rule-based approach utilizing simple string matching"[13] instead of machine learning in a "tool for classifying Written Cantonese, Standard Written Chinese, and the intermediate varieties used by Cantonese-speaking users from Hong Kong".[14] In this approach, Cantonese and Standard Written Chinese features were defined in regular expressions, i.e. Cantonese and Standard Written Chinese features were identified by regular expression matches.[15] Fang (2017) employed regular expressions to normalize text in the "prototype of a text analyzer for a rule-based Mandarin Chinese [text-to-speech] voice".[16]

In a more pedagogical context, Gajdoš (2017) discussed the implications of using the Chinese monolingual corpus *Hanku*, and more specifically, the subcorpus of legal text *zh-law* for language pedagogy. In that discussion, Gajdoš (2017) suggested that students of Chinese practise writing regular expressions to look for patterns of sentences.[17] Owing to this suggestion, and to the notion that teachers of CFL can also take advantage of the use of regular expressions, the present study explores and reports on the use of artificial intelligence and specialized documentation to build regular expressions that match the Chinese *shì…de* (是…的) grammar patterns. The *shì…de* (是…的) grammar patterns have been chosen for three reasons.

The first reason is that these grammar patterns perform different grammatical functions. *Chinese Grammar Wiki* presents six uses. These will be introduced below in the order in which they are presented by *Chinese Grammar Wiki*. Examples that illustrate each use are present on *Chinese Grammar Wiki's* site and are not repeated in this paper owing to the length of the present study.[18]

---

[11] W3Schools n.d.: *Python RegEx* section.
[12] Kuchling n.d.: *Introduction* section.
[13] Lau et al. 2024: 28.
[14] Lau et al. 2024: 24.
[15] Lau et al. 2024: 27.
[16] Fang 2017: 2.
[17] Gajdoš 2017: 86.
[18] Chinese Grammar Wiki n.d.: *The "shi… de" patterns: an overview* section.

1. The first use consists in omitting a noun and in using *de* (的) to represent it.
2. In the second use, the *shì…de* (是…的) grammar pattern contains a distinguishing word.
3. The third case is the classic pattern. It is explained that it "is used to draw attention to certain information in a sentence", that it "is frequently used when asking or telling details about the past", and that "most commonly it emphasizes time, manner, or place".[19]
4. *Chinese Grammar Wiki* also describes the *shì…de* (是…的) pattern when it is used to explain "why you are doing something or what an object is used for".[20]
5. The fifth use is to talk about people's occupation.
6. The last use is to emphasize affirmations.

Nonetheless, for methodological purposes, in this paper the author has taken the view that there exists a generic, high-level, parent *shì…de* (是…的) pattern. This pattern can be filled up with different components, which gives birth to the six patterns described above. As a result, within the context of this paper each of the patterns presented above will be regarded as a child pattern of the generic *shì…de* (是…的) pattern. The author has adopted this view to use a generic term, namely, the pattern *shì…de* (是…的) or the *shì…de* (是…的) pattern, when prompting the models. This was done with a view to prompting the models from the general to the specific. Indeed, it will be seen how the author has moved from the general in Exact Prompt 1, where the models are asked to create an HSK 3 text containing about 100 words and the pattern *shì…de* (是…的), and in Exact Prompt 7, where the models are asked to try this pattern, to the specific in Exact Prompt 9, where it is requested that the models find the emphatic structure. It should be noted that in this instance the term *structure* was used because even though the author had used the term *pattern* in his conversation with the models, *Gemini* came up with the term "classic 'emphatic' structure", and *Qwen3.5-Plus* used the term "The *shì…de* (是…的) structure".

The second reason why the *shì…de* (是…的) grammar patterns have been selected for the purpose of this paper is that they are expected to occur in a wide range of learning materials. This is because they can be used to perform several communicative functions. Finally, it can be claimed that these grammar patterns are also expected to be tackled at more than one level of a CFL program. This is because the complexity of the elements that are within these patterns may vary.

The methodology used to conduct the present exploratory study will be described in the next section.

---

[19] Chinese Grammar Wiki n.d.: *The "shi… de" construction for emphasizing details* section.
[20] Chinese Grammar Wiki n.d.: *The "shi… de" construction for indicating purpose* section.

Methodology

This exploratory study began with a CFL learner's, namely, the author's, basic knowledge of regular expressions, and therefore with an elementary regular expression containing three groups and the non-greedy quantifier +?. A non-greedy quantifier matches "as little text as possible".[21] For example, in

Wǒ shì zuò wǒ de fēijī lái de. (我是坐我的飞机来的。)

Translation: I came on my plane.

(shì)(\w+)(de), namely, the exact expression (是)(\w+)(的) matches *shì zuò wǒ de fēijī lái de* (是坐我的飞机来的, to be precise), whereas *(shì)(\w+?)(de)*, namely, the exact expression (是)(\w+?)(的), matches *shì zuò wǒ de* (是坐我的, to be exact), i.e. it matches as little text as possible. More information on the latter expression will be given below.

Afterwards, artificial intelligence models were prompted through a series of conversation strings with a view to examining the possibility of enhancing the original expression and acquiring new knowledge on the part of the user. These models initially were a free online version of the *Gemini 3* family and the free online *Qwen3.5-Plus*. Since artificial intelligence can make mistakes, the information supplied by these models was contrasted with that provided by specialist online sites.

Even though it was a free online model of the *Gemini 3* family that was used, a description of *Gemini 3 Flash Preview (Reasoning)* will be provided. Two salient features of this model are that it is a leading model in intelligence and notably fast, but somewhat verbose.[22] It has a one million tokens context window.[23] It scores 51% on the *SciCode (Coding)*, whereas *Gemini 3.1 Pro Preview*, which leads the field, scores 59%.[24] The *Gemini 3* family was chosen owing to the following combination: It has a free online version; *Gemini 3 Flash Preview (Reasoning)* has a high score on the *Artificial Analysis Intelligence Index*. This score is 46. [25]

Since the online *Qwen3.5-Plus* was prompted, a description of *Qwen3.5-397B-A17B (Reasoning)* and of *Qwen3.5-Plus* will be presented. *Qwen3.5-397B-A17B (Reasoning)* scores high on the *Artificial Analysis Intelligence Index* (its score is 45), but it is somewhat verbose.[26] It is an open-weight model,

---

[21] Kuchling n.d.: *Greedy versus Non-Greedy* section.

[22] Artificial Analysis n.d.: *Gemini 3 Flash Preview (Reasoning) Intelligence, performance & price analysis* section.

[23] *Ibid*.

[24] *Ibid*.

[25] *Ibid*.

[26] Artificial Analysis n.d.: *Qwen3.5 397B A17B (Reasoning) Intelligence, performance & price analysis* section.

and it has been reported to score more highly than *Gemini-3 Pro* on three coding-related benchmarks.[27] *Qwen3.5-Plus* is the hosted model, has a one million tokens context window and features "official built-in tools and adaptive tool use".[28] *Qwen3.5-Plus* was selected because of the following features: It is free; it is online; *Qwen3.5-397B-A17B* has been reported to score well on coding-related benchmarks. However, when this paper was reviewed by the author, *Qwen3.6-Plus* had been launched and was used as well. Each model will be cited in the analyses.

Some of those regular expressions suggested by the above-mentioned models were tested on the online regular expression tester *regex101*. This tester has several interesting features, such as a section that explains each component of the target expression, a match information section that highlights each component of this expression as included in the target text, a quick reference section, and the possibility of selecting a programming language. As regards this language, *Python 3* was used. Indeed, the *Python* flavor was turned on on the *regex101* tester to validate the regular expressions. Chinese text produced by *Gemini* (*Appendix A*) and by *Qwen3.5-Plus* (*Appendix B*) was used during this process. The prompt used to generate this text was the following.

Exact Prompt 1:

"please give me an hsk 3 text containing about 100 words and the pattern 是…的"[29]

The prompt referred to the parent grammar pattern *shì…de* (是…的) to give the models ample opportunity to generate text concerning any of the six grammar patterns described above.

Four limitations of this study will be discussed.

First, this paper reports on a single case study. Indeed, its findings apply to the language models the researcher employed. This limitation has been minimized by bringing up concepts that it is hoped will come in useful in future research that examines learners' use of regular expressions as well as the application of corpora and regular expressions to teach and to learn Chinese. This leads to the next limitation.

In this study text generated by artificial intelligence has been used to ensure a high density of occurrence of *shì…de* (是…的) grammar patterns. It was thought that this density would facilitate testing regular expressions – in other words, that it would enable the realization of the exploratory nature of this

---

[27] Qwen Team 2026: *Qwen3.5: Towards native multimodal agents* section.

[28] *Ibid*.

[29] The author is citing the exact prompt used during the research. The pinyin of the characters used by the author is *shì…de* (是…的).

paper. However, the author is of the opinion that testing regular expressions on learning materials (either digitalized materials such as electronic textbooks or digital materials such as online articles) can be an interesting future endeavor.

A more technical limitation is that this study discusses the raw regular expressions that were pasted on *regex101's Regular Expression* bar. As a result, none of the following content will be included in the presentation of the target expressions: methods such as *match()* , module-level functions such as in *re.match()*, the use of the *r* prefix to avoid a *SyntaxError*[30] or the flags used such as the *global flag* and the *single line flag*. As a result, it is recommended that the content of this paper should also be checked on other platforms or on code editors.

Finally, *Chinese Grammar Wiki* describes situations in which *shì* (是) is optional. This is in the classic pattern and when *shì…de* (是…的) is used to emphasize affirmations. These cases have not been considered in this study since it has focused on the use of the full *shì…de* (是…的) grammar patterns.

Report and Discussion

The point of departure of this research is *(shì)(\w+?)(de)*, namely, the following exact regular expression: (是)(\w+?)(的). It is known that the first group matches the character *shì* (是), whereas \w in the second group "matches any alphanumeric character; this is equivalent to the class [a-zA-Z0-9_]".[31] Nonetheless, after testing the above-mentioned regular expression on the text from *Appendix A* and *Appendix B* on *regex101*, this tester explained that the second group matches "any word character in any script" and this is "equivalent to [\p{L}\p{N}_]". Not being familiar with the syntax of this character class, both models were consulted using the following prompt in *conversation string A*.

Exact Prompt 2:

"I am using your text to test regular expressions on regex 101. I use this regex: (是)(\w+?)(的). Regex 101 says that \w match any word character in any script (equivalent to [\p{L}\p{N}_] What is this?"[32]

As for *(shì)(\w+?)(de)*, namely, the author's exact regular expression (是)(\w+?)(的), *Qwen3.5-Plus* took a sentence from *Appendix B*, i.e. from the text it

---

30 Kuchling n.d.

31 Kuchling n.d.: *Matching Characters* section.

32 The author is citing the exact prompt used during the research. The pinyin of the characters used by the author is *shì* (是) and *de* (的). These characters are employed in the expression (是)(\w+?)(的).

had already generated, and explained why this expression should work. This is shown in Table One. Please note that the model did not use pinyin, and that its exact output was cited in this table.[33]

Aspects of regular expression	Model's exact output except for the translation
Model's sentence	我们是星期六早上出发的。
Translation of the model's sentence	We set off on Saturday morning.
Explanation of the first group	(是) matches: 是
Explanation of the second group	(\w+?) matches: 星期六早上出发 (Because these are all Unicode "letters")
Explanation of the third group	(的) matches: 的
Model's warning	While this works on regex101, be careful if you use this regex in actual programming code: Python 3: \w matches Unicode characters by default. (Your regex will work). JavaScript: \w does not match Chinese characters unless you use the Unicode flag (/u).

Table 1. Qwen3.5-Plus's explanation of the initial regular expression

Where *regex101*'s feedback regarding [\p{L}\p{N}_] is concerned, *Gemini* explained that \p{L} "matches any character in any language that is classified as a 'letter'", and that \p{N} "matches any kind of numeric character in any script". *Qwen3.5-Plus* explained that when *Unicode* is enabled, \p{L} "stands for Unicode Letter", and \p{N} "stands for Unicode Number". This content was validated on *The Modern JavaScript Tutorial* site, where it is explained that *Unicode* properties written as \p{…} can describe a character's category, that \p{L} "denotes a letter in any language", and that N is an alias for number.[34]

It is also interesting to point out that *Qwen3.5-Plus* issued a warning which concerns the use of flags: in *JavaScript* you need to use the /u flag to match Chinese characters, whereas in *Python 3* \w matches *Unicode* characters by default.

---

[33] The pinyin of the characters used by the model in Table One is *wǒmen shì xīngqíliù zǎoshang chūfā de* (我们是星期六早上出发的), *shì* (是), *xīngqíliù zǎoshang chūfā* (星期六早上出发), and *de* (的).

[34] Kantor 2022: *Unicode: flag "u" and class \p{...}* section.

However, in the original setting of the *Python* flavor on *regex101* that the author had used, the *Unicode* flag, namely, u, was used. Nonetheless, *regex101* informed that this was redundant in *Python 3*. Given the warning issued by *Qwen3.5-Plus* and the message given by *regex101*, the author still wanted to query *Gemini* on this issue through a prompt that belonged in *conversation string A1*. This prompt is cited below.

Exact Prompt 3:

"I am using Python on *regex101* ("flavor is set to Python"). I turned on the "u" (Unicode) flag, but *regex101* says: u modifier: unicode. Make \w, \W, \b, \B, \d, \D, \s and \S perform matching with Unicode semantic (redundant in Python 3). Do I really need this flag to match Chinese characters?"

When asked if the *Unicode* u flag was needed to match Chinese characters while using *Python 3* on *regex101*, *Gemini* explained that this was not so because in *Python 3* "all strings are native Unicode by default". This statement was validated by the *Python 3.14.3 documentation*, where it says that "in Python 3, Unicode characters are matched by default for str patterns", and that the *re.UNICODE* "flag is therefore redundant with no effect and is only kept for backward compatibility".[35] Where backward compatibility is concerned, Davies (2014), for example, tackled the use of the u prefix in *Python 2* while building a site for learning Chinese characters.[36]

Another topic was brought up by the models in *conversation string B*. This was matching only Chinese text. This happened when they were asked to build regular expressions for Chinese grammar patterns through the prompts that are shown below.

Exact Prompt 4:

"Do you know about regular expressions?"

Gemini answered "I certainly do!", provided a cheat sheet with some of the most common characters used in these expressions, supplied a real-world example, and offered help. *Qwen3.5-Plus* answered "Yes, I'm familiar with regular expressions (regex)!", provided three samples, and offered help. This was followed by the next prompt.

Exact Prompt 5:

"Ok. Can you help me build regex expressions to build Chinese patterns?"

At that point, the use of the script \p{Script=Han} was suggested by *Gemini*. However, this script failed when the *Python* flavor was selected on *regex101*.

---

[35] Python Software Foundation n.d.: *Module Contents* section.

[36] Giles 2014.

As a result, *Gemini* and *Qwen3.6-Plus* were prompted in *conversation string C*. This prompt is cited below.

Exact Prompt 6:

"Do we use this in Python 3: \p{Script=Han}?"

Both models agreed that using the third party *regex* module makes it possible to use this script in *Python 3*. When browsing the Web to get more details about this information, it was observed that several regular expressions had also been suggested to match Chinese characters.[37] Where the suggested script is concerned, Ayaka (n.d.) pointed out that \p{sc=Han} does not only match Chinese characters. Indeed, Goyvaerts (2025) explains that "Unicode offers the Hiragana, Katakana, Han, and Latin scripts that Japanese documents are usually composed of"[38], i.e. Han scripts that are used in Japanese. In addition, the *Unicode Consortium* (2024) explain that "the Unicode Standard leaves the issues of language tagging and word recognition up to a higher level of software and does not attempt to encode the language of the Han characters".[39]

Nonetheless, the objective of the present paper is not to delve into the limitations of the regular expressions or scripts that match Chinese characters. Nor is it to look into the historical analysis of rare or historic characters, or to discriminate between Chinese characters from the characters of other languages, but to report on the use of artificial intelligence in combination with specialized documentation with a view to constructing regular expressions that match the *shì…de* (是…的) grammar patterns and that can be used in simplified present day Chinese text.

Even though the script \p{Script=Han} failed when the *Python* flavor was selected on *regex101*, the author observed that this script worked on *regex101* when the *JavaScript* flavor was selected and when the *Unicode* flag was on. However, it will just be pointed out that a problem that was encountered while using \p{Script=Han} on *regex101* (more specifically, on *Java*, *PCRE2 (PHP)* or *JavaScript*) was naturally that it did not match numbers. However, the presence of numbers between *shì* (是) and *de* (的) in online Chinese newspaper articles should come as no surprise, and its study can be very interesting to further the knowledge of Chinese grammar. For example, the grammar pattern containing *shì* (是) and *de* (的) in the following sentence can be used to contrast this sentence's grammar pattern with the classic *shì…de* (是…的) grammar pattern.

---

[37] Ayaka, n.d.; Fu, 2021.

[38] Goyvaerts 2025: *Unicode Scripts and Script Extensions* section.

[39] Unicode Consortium 2024: Chapter 18, *Distinguishing Han Character Usage Between Languages* section.

Zhè shì 2025 nián 1 yuè 2 rì pāishè de mǎ lì zhèn "yún zhōng qiúchǎng"(wú rén jī zhàopiàn).

(这是2025年1月2日拍摄的马利镇"云中球场"（无人机照片）。)[40]

Translation: This is a January 2nd 2025 picture of Mali Town's "Field in the Clouds" (a drone photo).

Owing to the situation described above, it was decided to continue to use the \w character. Nevertheless, after *Exact Prompt 5* in *conversation string B*, the author sent the following prompt. The slip of the pen in the spelling of *Let* was kept for transparency.

Exact Prompt 7:

"LEt us try this pattern: 是…的"[41]

When asked to use regular expressions to match the *shì…de* (是…的) pattern, both models proposed *shì.+?de*, namely, the exact expression 是.+?的, where ". is often used where you want to match 'any character'", and "+ requires at least *one* occurrence"[42]. The non-greedy quantifier +? matches "as little text as possible".[43] *Qwen3.5-Plus* described the components of the suggested expression succinctly. This is shown in Table Two. Please note that the model did not use pinyin, and that the exact output was cited in this table.[44]

Aspects of regular expression	**Model's exact output except for the translation:**
Explanation of the first item: *shì* (是)	Literal character "是"
Explanation of the second item: .+?	Match any character, non-greedy (stops at the first 的)
Explanation of the third item: *de* (的)	Literal character "的"
Model's sentence	我是昨天来的，他是坐飞机来的
Translation of the model's sentence	I arrived yesterday, he came by plane.

[40] Xinhuanet 2026: 2025年新华社年度照片·体育新闻 (The author added the pinyin version).

[41] The author is citing the exact prompt used during the research. The pinyin of the characters used by the author is *shì…de* (是…的).

[42] Kuchling n.d.: *Matching characters* section and *Repeating things* section.

[43] Kuchling n.d: *Greedy versus Non-Greedy* section.

[44] The pinyin of the characters used by the model in Table Two is *shì* (是), *de* (的), *wǒ shì zuótiān lái de, tā shì zuò fēijī lái de* (我是昨天来的，他是坐飞机来的), *shì zuótiān lái de, tā shì zuò fēijī lái de* (是昨天来的，他是坐飞机来的), *shì zuótiān lái de* (是昨天来的), and *shì zuò fēijī lái de* (是坐飞机来的).

Explanation of the greedy component: *shì.+de*, namely, 是.+的	matches: 是昨天来的，他是坐飞机来的 ✗ (too long!)
Explanation of the non-greedy component: *shì.+?de*, namely, 是.+?的	matches: 是昨天来的 ✓ and 是坐飞机来的 ✓

Table 2. Qwen3.5-Plus's explanation of shì.+?de, namely, the exact expression是.+?的

As a result, *shìde* (是的) cannot be matched because at least one occurrence is required in between. The models also issued a series of warnings. *Qwen3.5-Plus* suggested adding the *re.DOTALL* flag in case content spanned lines, which made the author activate the *single line* (*Dot matches newline*) flag on *regex101*. In addition, if using *shì.+de*, namely, the exact expression是.+的 (without the non-greedy ?), *Gemini* recommended establishing the number of characters to be matched by replacing + with the {m,n} quantifier[45] "to limit the number of characters allowed in the middle". For example, *shì.{1,5}de*, namely, 是.{1,5}的 to be exact, matches *shì qí zìxíngchē qù de* (是骑自行车去的 to be precise (went by bike)) because there is at least one repetition and at most five.[46] However, it will not match *shì zuò dìtiě lái gōngsī de* (是坐地铁来公司的 to be exact (came to the company by subway)) because there are six repetitions in this chunk.

Furthermore, it was observed that in the sentence

shì de, tā shì zuò dìtiě lái de (是的，他是坐地铁来的to be exact),

Translation: Yes, he came by subway.

shì.+?de, namely, the exact expression是.+?的, would match the whole chunk because at least one occurrence is required. Nevertheless, *Qwen3.5-Plus* suggested different regular expressions in addition to *shì.+?de*, namely, the exact expression是.+?的. One of them was *shì.*?de*, namely, the exact expression是.*?的. This expression matched *shìde* (是的 to be precise), and *shì zuò dìtiě lái de* (是坐地铁来的 to be exact) separately because, as *Qwen3.5-Plus* put it, it allows "empty content". Indeed, "* allows zero or more characters" (*Qwen3.5-Plus*) as against +, where at least one occurrence is required. In other words, * "specifies that the previous character can be matched zero or more times, instead of exactly once".[47]

Another suggestion was made by *Gemini* in *conversation string D*, where the author prompted the following.

[45] Kuchling n.d.: *Repeating things* section.
[46] Kuchling n.d.: *Repeating things* section.
[47] Kuchling n.d.: *Repeating things* section.

Exact Prompt 8:

"I have the following sentence: 是的，他是坐地铁来的。The following regular expression 是.+?的 matched the segment that goes from the first 是 to the end. I want to match 是的 and 是坐地铁来的 separately in Python. Can you please give me a suitable regular expression?"[48]

The use of a character class starting with ^ was suggested in *shìde|shì[^de]+de*, namely, in the exact expression 是的|是[^的]+的. This matches "the characters not listed within the class by *complementing* the set".[49] *Gemini*'s explanation of this regular expression is provided in Table Three. Please note that the model did not use pinyin, and that the exact output was cited in this table.[50]

Aspects of regular expression	Model's exact output
Explanation of the first item: *shìde* (是的)	It looks for the exact character 是 immediately followed by the exact character 的
Metacharacter |[51]	| acts like an or operator in logic
Explanation of the second item: *shì* (是)	The engine looks for the character 是
Explanation of the third item: *[^de],* namely, [^的]	Any character that is not 的
Explanation of the fourth item: +	One or more times
Explanation of the third and the fourth items combined: *[^de]+*, namely, [^的]+	Tells the engine to match a continuous stream of text, but it must stop the moment it hits the character 的.
Explanation of the last item: *de*, namely, 的	Finally, the engine looks for the closing 的 character right after that stream.

Table 3. Gemini's explanation of shìde|shì[^de]+de, namely, the exact expression 是的|是[^的]+的

Afterwards, the author decided to move on to a specific *shì…de* (是…的) grammar pattern: the classic one or that which emphasizes details.[52] This was in *conversation string B*, i.e. after the models proposed *shì.+?de*, namely, the exact

[48] The author is citing the exact prompt used during the research. The pinyin of the characters used by the author is *shì* (是), *shì de, tā shì zuò dìtiě lái de* (是的，他是坐地铁来的), *shìde* (是的), *shì zuò dìtiě lái de* (是坐地铁来的), and *shì* (是) and *de* (的) in the expression是.+?的.

[49] Kuchling n.d.: *Matching Characters* section.

[50] The pinyin of the characters used by the model in Table Three is *shì* (是) and *de* (的).

[51] Kuchling n.d.: *More metacharacters* section.

[52] Chinese Grammar Wiki n.d.: *The "shi… de" construction for emphasizing details* section.

expression是.+?的. Therein, *Qwen3.5-Plus* offered to "extract the emphasized part (e.g., get '昨天来' from '我是昨天来的')"[53], and *Gemini* asked what the goal for the *shì…de* (是…的) pattern was. The answer is given below.

Exact Prompt 9:
"to find the emphatic structure."

Interesting warnings were made. Perhaps the most relevant one was that which pointed out that *shì…de* (是… 的) also has other uses. *Gemini* expressed "to find the emphatic structure (是...的) specifically, we need to be careful. In Chinese, 是 (is) and 的 (possessive/adjective marker) appear constantly in non-emphatic contexts".[54] *Qwen3.5-Plus*, for example, pointed out the importance of distinguishing the use of *shì* (是) and *de* (的) for possession and for emphasis.

Gemini offered several suggestions to match the classic grammar pattern specifically. It highlighted that "the emphatic structure usually highlights Time, Location, or Manner", and suggested using *shìzài* (是在 to be precise), at the beginning when the target is location, adding a verb such as *zuò* (做 to be precise) in *zuòde* (做的 to be exact) at the end when the target is location, and incorporating a verb such as *lái* (来to be precise) in *láide* (来的 to be exact) at the end when the target is manner. This is expressed in Table Four as quoted from *Gemini*. Indeed, please note that the model did not use pinyin, and that its exact output was cited in this table.[55] However, a column for translations was added by the author.

Target Emphasis	Regex Pattern	Example Match	Translation
Time (Years/Hours)	是(\d+年|\d+点)的	是2022年的	Belongs to the year 2022
Location (At/In)	是在([\u4e00-\u9fa5]+)做的	是在办公室做的	Was done in the office
Manner (How)	是([\u4e00-\u9fa5]{1,3})来的	是开车来的	It was by car that [someone] came

Table 4. Gemini's description of the emphatic structure

[53] The author is citing the exact model's output. The pinyin of the characters used by the model is *zuótiān lái* (昨天来) and *wǒ shì zuótiān lái de* (我是昨天来的).

[54] The author is citing the exact model's output. The pinyin of the characters used by the model is *shì* (是), *de* (的), and *shì…de* (是… 的).

[55] The pinyin of the characters used by the model in Table Four is *shì 2022 nián de* (是2022年的), *shì zài bàngōngshì zuò de* (是在办公室做的), and *shì kāichē lái de* (是开车来的). In addition, *shì* (是), *de* (的), *diǎn* (点) and *nián (*年*)* in是(\d+年|\d+点)的, and *shì* (是), *zài* (在), *zuò* (做), and *de* (的) in是在([\u4e00-\u9fa5]+)做的. Finally, *shì* (是), *lái* (来) and *de* (的) in 是([\u4e00-\u9fa5]{1,3})来的.

Gemini showed an inaccuracy: It is not clear whether *shì(\d+nián|\d+diǎn)de*, namely, 是(\d+年|\d+点)的 to be exact, where one or more digits (\d+) is followed by *nián* (年 to be exact) or *diǎn* (点 to be precise) is an emphatic pattern. In *conversation string D*, however, *Qwen3.6-Plus* expressed a few limitations concerning the manner pattern. This was done after exact prompt 10.

Exact prompt 10:

"please explain 是([\u4e00-\u9fa5]{1,3})来的".[56]

Qwen3.6-Plus explained that the manner pattern accepts up to three characters only ({1,3}), and [\u4e00-\u9fa5] excludes rare archaic characters amongst other items. In addition, it made a smart comment where it argued that the manner pattern "may match semantically odd phrases like 是苹果来的 if not filtered by a downstream NLP step (e.g., named entity recognition)".[57] Indeed, named-entity-recognition technology has the power of looking into semantic aspects of language.

Gemini also highlighted a *punctuation* problem. It explained that "in the emphatic structure, the 的 is often at the very end of the sentence" and suggested a strategy "if you want to be 100% sure you are catching the Grammatical Structure and not just a random '是...的' phrase".[58] This was *shì(.+?)de(?=[。！？]|\s|$)*, namely, the exact expression 是(.+?)的(?=[。！？]|\s|$), and it is described in Table Five.

Aspects of regular expression	**Model's exact output except where otherwise indicated**
Explanation of the first item: *shì(.+?)de*, namely, the exact expression是(.+?)的	Please, refer to Table Two (the author's note).
Positive lookahead assertion: (?=…)[59]	Only match this if it's followed by
Explanation of the second item: [。！？]	a Chinese period, an exclamation/question mark
Metacharacter |[60]	Please, refer to Table Three (the author's note).

[56] The author is citing the exact prompt used during the research. The pinyin of the characters used by the author is *shì* (是), *lái* (来) and *de* (的) in是([\u4e00-\u9fa5]{1,3})来的.

[57] The author is citing the exact model's output. The model used *shì píngguǒ lái de* (是苹果来的 to be precise).

[58] The author is citing the exact model's output. The pinyin of the characters used by the model is *de* (的) and *shì…de* (是...的).

[59] Kuchling n.d.: *Lookahead assertion*s section.

[60] Kuchling n.d.: *More metacharacter*s section.

Explanation of the third item: \s	a space
Metacharacter |[61]	Please, refer to Table Three (the author's note).
Explanation of the fourth item: $	the end of the string.

Table 5. Gemini's explanation of shì(.+?)de(?=[。！？]|\s|$), namely, the exact expression 是(.+?)的(?=[。！？]|\s|$)

One may have the impression that these analyses could go on ad infinitum. Indeed, making up regular expressions seems to be an endless task owing to the fact that catches may constantly appear. Some of them are spotted immediately; others are identified later during the testing process. As already explained, the purpose of this paper is not to present a perfect regular expression or systems of expressions to match the *shì…de* (是…的) grammar patterns. On the contrary, it is to report on the use of artificial intelligence and specialized documentation to construct these expressions.

In using artificial intelligence and specialized documentation to construct regular expressions, it can be observed that language models such as the ones used in this paper can perform several interesting functions where the study of the *shì…de* (是…的) grammar patterns is concerned.

1. The most obvious function is the presentation of regular expressions for the user to test them for their validity. Two cases in point were *shì.+?de*, namely, the exact expression是.+?的, and *shì(.+?)de(?=[。！？]|\s|$)*, namely, the exact expression 是(.+?)的(?=[。！？]|\s|$).
2. More interestingly, another function is the presentation of nuances not only of the components of regular expressions (*Qwen3.6-Plus*, for example, warned about the limitations of {1,3}), but also of the language these expressions match (*Gemini*, for example, suggested avoiding mixing up the emphatic and the possessive patterns containing *shì* (是) and *de* (的)). However, the following must be borne in mind. Even though language models should be expected to supply accurate information about regular expressions and about the language they match in response to users' prompts, they may make mistakes. In this regard documentation plays an essential role. This is to validate the content produced by language models. Indeed, the range of regular expressions that language models propose can be compared with information as presented by existing technical or linguistic specialized documentation. That is, specialized documentation

[61] Kuchling n.d.: *More metacharacters* section.

is expected to present accurate information about the components of regular expressions or about the language they match. Specialized documentation is also to illustrate this information with examples.

3. Language models can also help users advance their knowledge of regular expressions by presenting expressions that are more complex than the ones they are already familiar with. This can encourage users to move beyond their comfort zone. For example, *Gemini* introduced the concept of *lookahead*, which was later validated with specialized documentation.
4. In dealing with regular expressions, language models can also help users advance their knowledge of the Chinese language by means of the presentation of the nuances alluded to above. An example of these nuances is the different communicative functions that the *shì…de* (是…的) grammar patterns can perform. For example, in Table Four, *Gemini* explained the focus on manner, place or time where the emphatic pattern is concerned.
5. Language models can also make users aware of catches inside regular expressions. For example, *Qwen3.6-Plus* explained that *shì([\u4e00-\u9fa5]{1,3})láide*, namely,是([\u4e00-\u9fa5]{1,3})来的 to be precise, match odd phrases such as *shì píngguǒ lái de* (是苹果来的 to be exact (the author leaves the translation of this text to the readers' imagination)).

Conclusion

It can be concluded that the most complete function language models such as the ones discussed in this paper can perform while being used in conjunction with specialized documentation to construct regular expressions that match Chinese grammar patterns is the following: To enhance regular expressions while delving into the nuances that Chinese grammar patterns may convey. However, it must be said that the use of language models to make up regular expressions that match Chinese grammar patterns comes with challenges and requirements.

One of these challenges is the readiness to spot inaccurate information. For example, in Table Four, *Gemini* presented an example, *shì 2022 niánde* (是2022年的 to be exact), which was not clearly emphatic. Indeed, in the present context, knowledge of the nuances of the target grammar patterns or the ability to learn about these nuances from the text matched by the regular expressions provided by the models is required to analyze these models' information critically. For example, if the target grammar pattern is the classic one, the user can explore whether the text matched by the regular expressions provided by the models actually agrees with this pattern. As a result, in dealing with the nuances of the target grammar patterns, documentation that concerns Chinese grammar comes

into play. To illustrate this, one can cite *Chinese Grammar Wiki* to show that the use of the *shì…de* (是…的) grammar patterns transcends that of expressing emphasis on details concerning past events.

Another challenge is the complexity that is inherent in regular expressions. Indeed, one may argue that the use of regular expressions requires at least a basic understanding of the programming language that is selected to implement them. Information about this language can be provided by specialized documentation and should come in useful when regular expressions are tested. Specialized documentation also serves as a source that can be consulted to confirm the accuracy of the information provided by language models. As a result, reviewing specialized documentation may also be required with a view to succeeding in understanding regular expressions. These considerations lead to several pedagogical implications where the use of regular expressions to teach and learn Chinese is concerned.

The language models proved to be skillful in their presentation of regular expressions in spite of the above-mentioned challenges. They expressed nuances regarding not only the functions Chinese grammar patterns may perform, but also the text those expressions match. In so doing, they provided options, illustrated them, and explained them. As suggested above, users that employ these models to build regular expressions therefore should have at least some basic knowledge of the programming language selected to implement those expressions. This translates into learning about the components of these expressions, namely, of their code, and, as a result, into analyzing the range of data provided by models critically. Training that fosters this knowledge can make use of specialized documentation. Future studies could delve into the implications of providing this training. Indeed, reports on teachers' and learners' opinions of traversing Diagram A should provide interesting input in regard to the convenience and challenges of using language models and specialized documentation to build regular expressions.

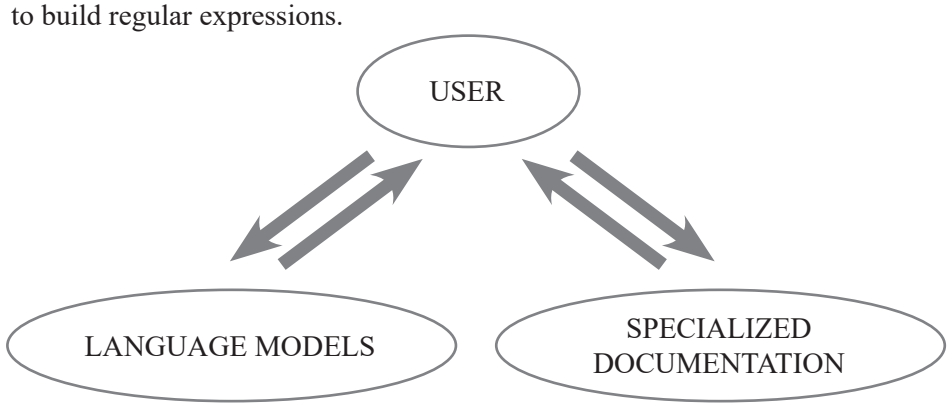


Diagram A. Regular expression learning graph

In Diagram A, the point of departure of the learning process can be a regular expression provided by a language model. The user consults specialized documentation that both explains and validates the features of this expression. Alternatively, a regular expression is suggested by specialized documentation. The user consults a language model to obtain further details about that expression, and then validates these details with specialized documentation. Another point of departure is the user, who comes up with a regular expression (possibly after consulting a language model or specialized documentation), and validates it by consulting a language model and specialized documentation. It must be highlighted that throughout these interactions the user tests regular expression candidates. Indeed, this paper closes by considering further research on the feasibility of using regular expressions to study Chinese grammar patterns in the classroom while navigating Diagram A, where the user reads specialized documentation, prompts language models, writes code and tests these expressions.

Bibliography

Artificial Analysis. (n.d.). *Gemini 3 Flash Preview (Reasoning) Intelligence, performance & price analysis*. https://artificialanalysis.ai/models/gemini-3-flash-reasoning (last accessed: 24.08.2026).

Artificial Analysis. (n.d.). *Qwen3.5 397B A17B (Reasoning) Intelligence, performance & price analysis*. https://artificialanalysis.ai/models/qwen3-5-397b-a17b (last accessed: 24.08.2026).

Ayaka (n.d.). *How to properly match Chinese characters with regular expression.* https://ayaka.shn.hk/hanregex/ (last accessed: 24.08.2026).

Carter, A. – Absalom, M. 2023. 'Giving Students the Tools: Looking at Teaching and Learning using Corpora', *The EuroCALL Review*, *30*(1): 52–62. https://doi.org/10.4995/eurocall.2023.14950 (last accessed: 24.08.2026).

Chen, H. 2012. *Using language corpora to enhance grammatical proficiency in Chinese* (Publication No. 3342). [Master's thesis, Brigham Young University]. https://scholarsarchive.byu.edu/etd/3342 (last accessed: 24.08.2026).

Chinese Grammar Wiki (n.d.). *The "shi… de" construction for emphasizing details*. https://resources.allsetlearning.com/chinese/grammar/The_%22shi..._de%22_construction_for_emphasizing_details (last accessed: 24.08.2026).

Chinese Grammar Wiki (n.d.). *The "shi… de" construction for indicating purpose*. https://resources.allsetlearning.com/chinese/grammar/The_%22shi..._de%22_construction_for_indicating_purpose (last accessed: 24.08.2026).

Chinese Grammar Wiki (n.d.). *The "shi… de" patterns: an overview*. https://resources.allsetlearning.com/chinese/grammar/The_%22shi..._de%22_patterns:_an_overview (last accessed: 24.08.2026).

Dib, F. (n.d.). *Regex101*. Regular Expressions 101. https://regex101.com/ (last accessed: 24.08.2026).

Fang, C. 2017. *A prototype text analyzer for Mandarin Chinese TTS system*. [Master's thesis in language technology, Uppsala University]. https://www.diva-portal.org/smash/get/diva2:1111826/FULLTEXT01.pdf (last accessed: 24.08.2026).

Fu, A. 2021. *Match Chinese characters*. Anthony Fu. https://antfu.me/posts/match-chinese-characters (last accessed: 24.08.2026).

Gajdoš, Ľuboš. 2017. 'Chinese legal texts – Quantitative Description', *Acta Linguistica Asiatica*, *7*(1): 77–87. https://doi.org/10.4312/ala.7.1.77-87 (last accessed: 24.08.2026).

Giles 2014. 'HSK东西 Scripts: A site for learning Chinese characters – or, "Handling Chinese characters with Python Unicode strings is less hassle than I thought it would be"', *PythonAnywhere*. https://blog.pythonanywhere.com/79/ (last accessed: 24.08.2026).

Google 2026. *Gemini* [Large language model]. https://gemini.google.com (last accessed: 24.08.2026).

Goyvaerts, J. 2025. *Unicode Scripts*. Regular-Expressions.info. https://www.regular-expressions.info/unicodescript.html (last accessed: 24.08.2026).

Ištvánová, M. 2021. 'On the Use of Corpora in Second Language Acquisition – Chinese as an Example', *Acta Linguistica Asiatica*, *11*(2): 131–142. https://doi.org/10.4312/ala.11.2.131-142 (last accessed: 24.08.2026).

Kantor, I. 2022. 'Unicode: flag "u" and class \p{...}', *The Modern JavaScript Tutorial*. https://javascript.info/regexp-unicode (last accessed: 24.08.2026).

Kuchling, A. M. (n.d.). 'Regular Expression HOWTO', *Python 3.14.3 documentation*. https://docs.python.org/3.14/howto/regex.html#regex-howto (last accessed: 24.08.2026).

Lau, C. – Lau, M. – To, A. 2024. 'The extraction and fine-grained classification of written Cantonese materials through linguistic feature detection', *Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI)*: 24–29. https://aclanthology.org/2024.eurali-1.4/ https://doi.org/10.63317/3obi2uptjo3x (last accessed: 24.08.2026).

Python Software Foundation (n.d.). 're – Regular expression operations', *Python 3.14.3 Documentation*. https://docs.python.org/3/library/re.html#id1 (last accessed: 24.08.2026).

Qwen Team 2026. 'Qwen3.5: Towards native multimodal agents', *Alibaba Cloud*. https://qwen.ai/blog?id=qwen3.5 (last accessed: 24.08.2026).

Torres Ramírez, A. 2012. 'The lexical approach: collocability, fluency and implications for teaching', *Revista De Lenguas Para Fines Específicos*, *18*: 237–254. https://ojsspdc.ulpgc.es/ojs/index.php/LFE/article/view/44 (last accessed: 24.08.2026).

Unicode Consortium 2024. *The Unicode® Standard Version 16.0 – Core Specification*. South San Francisco. https://www.unicode.org/versions/Unicode16.0.0/core-spec/ (last accessed: 24.08.2026).

Xinhuanet 2026. 2025年新华社年度照片, 体育新闻. Xinhua Agency's photos of the year 2025: Sports News. [Author's translation]. https://www.news.cn/photo/20260104/08216945fbbb4b10b38ec3f6f56a3623/c.html (last accessed: 24.08.2026).

Xu, J. 2019. 'The Corpus Approach to the Teaching and Learning of Chinese as an L1 and an L2 in Retrospect', In: Lu, X. – Chen, B. (eds). *Computational and Corpus Approaches to Chinese Language Learning. Chinese Language Learning Sciences*. Springer Nature Singapore: 33-53. https://corpus.bfsu.edu.cn/style/Xu2019_Chapter_TheCorpusApproachToTheTeaching.pdf https://doi.org/10.1007/978-981-13-3570-9_3 (last accessed: 24.08.2026).

W3Schools (n.d.). *Python RegEx*. https://www.w3schools.com/python/python_regex.asp (last accessed: 24.08.2026).

Yuan, S. – Zhao, H. – Du, Z. – Ding, M. – Liu, X. – Cen, Y. – Zou, X. – Yang, Z. – Tang, J. 2021. 'WuDaoCorpora: A super large-scale Chinese Corpora for pre-training language models', *AI Open*, *2*: 65–68. https://doi.org/10.1016/j.aiopen.2021.06.001 (last accessed: 24.08.2026).

Yuan, Y. 2025. 'Reflections of Task-Based Language Teaching on Grammar Teaching of Teaching Chinese as a Foreign Language', *GBP Proceedings Series*, *5*: 7–19. https://www.gbspress.com/index.php/GBPPS/article/view/326 https://doi.org/10.71222/tgvm3515 (last accessed: 24.08.2026).

Appendix A: Text generated by *Gemini*

我上个星期换了一份新工作。我的新公司就在火车站附近，环境非常好。我是在网上看到招聘信息的，然后发了我的简历。
周三下午，我去参加了面试。经理人很好，我们聊得很开心。他是坐地铁来公司的，而我是骑自行车去的。他问我什么时候可以开始上班，我说随时都可以。
我很喜欢这个新工作，因为同事们都很热情。我的这份工作是我努力争取来的，我一定会好好表现。

Translation:

Last week I changed jobs. My new company is right near the railway station, the atmosphere is very good. I saw the job advertisement online, I submitted my CV afterwards.

On Wednesday afternoon, I attended the interview. The manager was very nice, and the conversation went very well. He came to the company by subway, whereas I went by bike. He asked me when I could start work, and I said any time.

I really like this new job, because my colleagues are very welcoming. I earned this job of mine through hard work. I will certainly perform well.

Appendix B: Text generated by *Qwen3.5-Plus*

上个周末，我和朋友一起去北京旅游了。我们是星期六早上出发的。因为家离火车站很近，所以我们是走路去车站的。到了北京，我们是坐地铁去长城的。虽然有点累，但是我们玩得很开心。这张照片是我们在长城上拍的。这次旅游是我今年最高兴的事。

Translation:

Last weekend, my friends and I travelled to Beijing. We left on Saturday morning. We walked to the railway station because it is near our place. When we arrived in Beijing, we went to the Great Wall by subway. Even though we were a bit tired, we really had a lot of fun. We took this photo on the Great Wall. This trip is the happiest thing that has happened to me this year.

Statement about the use of artificial intelligence: Artificial intelligence tools were used to refine the Chinese text and the translated text.

Journal of Chinese Language Teaching in Europe 欧洲中文教育 2026/2, pp. 25–37

庞泓雪 – 解植永
(天津外国语大学国际教育学院)

匈牙利德布勒森市华裔儿童中文传承研究

摘要

本文以匈牙利德布勒森市新兴工业移民社区的华裔儿童为研究对象，基于问卷调查（儿童n=9，家长n=11）与3个典型家庭的深度访谈，系统考察该群体中文传承的实践状态与认同建构。研究发现：华裔家庭普遍形成功能性语言分区，中文使用域高度收窄，且儿童读写能力发展因缺乏体系化支持而明显滞后；家长期望整体偏高并带有较强的“母语者执念”，但理想语言身份与现实能力之间的落差构成了儿童内在的心理张力；值得注意的是，语言能力的退缩并未导致文化认同的同步松动，多数受访儿童通过节庆仪式等非语言路径独立维系了文化归属感。基于此，本文提出，中东欧华裔儿童传承语教学应从“母语者标准”向“情感与功能导向”转型，在教学中运用“跨语言实践”（Translanguaging）激活儿童内在动机；同时，呼吁依托当地孔子学院（如“OTT-HOME”社区课程）及在地中资企业资源，构建“家校社”协同的传承支持网络，以填补当地中文教育生态的体制性缺口。

关键词： 华裔儿童，传承语，文化认同，家庭语言政策，跨语言实践，匈牙利

13954871623@163.com
x-zy2000@163.com

https://doi.org/10.61637/cltcee.2026.2.2
https://orcid.org/0009-0007-5152-6626
https://orcid.org/0009-0008-9045-8985

Pang Hongxue – Xie Zhiyong

A Study on Chinese Heritage Language Transmission Among Ethnic Chinese Children in Debrecen, Hungary

Abstract

This study investigates the Chinese heritage language transmission practices and identity construction among ethnic Chinese children in Debrecen, Hungary, an emerging industrial immigrant community. Drawing on questionnaire surveys (children n=9, parents n=11) and in-depth interviews with three typical families, the research systematically examines their language reality. Findings reveal that Chinese families predominantly adopt functionally mixed language strategies, leading to a narrowed domain of Chinese language use and significantly lagging literacy skills due to the lack of systematic support. While parental expectations remain high and exhibit a strong 'native-speaker obsession,' the gap between idealized linguistic identity and actual proficiency creates internal psychological tension for the children. Crucially, however, the erosion of language proficiency does not translate into an equivalent loss of cultural identity; most children independently maintain a sense of cultural belonging through non-linguistic pathways such as family rituals and festivals. Based on these findings, this paper argues that heritage language education for Chinese children in Central and Eastern Europe should shift from a 'native-speaker standard' to an 'emotional and functional orientation,' utilizing 'Translanguaging' pedagogies to activate intrinsic motivation. Furthermore, it calls for the establishment of a collaborative 'family-school-community' support network—leveraging local Confucius Institute initiatives (e.g., the 'OTT-HOME' community program) and locally-based Chinese enterprises—to fill the institutional gaps in the regional Chinese language education ecosystem.

Keywords: ethnic Chinese children, heritage language, cultural identity, family language policy, Translanguaging, Hungary

一、引言

海外华裔儿童的中文传承是一个涉及语言能力、家庭政策与身份认同的复合议题。既有研究已确立若干基本命题：传承语学习者有其不同于外语学习者的独立习得逻辑（Valdés，2005）；家长的语言态度与日常实践对子女的中文保持程度具有决定性影响（Wang，2020）；在多语环境中成长的华裔儿童，往往呈现出听说能力相对稳定、读写能力明显滞后的“不对称双语”状态（Cummins，2005）。近年来，欧洲华裔社区的传承语研究已出现若干突破。Hu与Yagmur（2026）对荷兰华裔社区的个案研究表明，华裔儿童的语言保持受到家长、教育者与学生三方互动的共同塑造，族群归属感在传承语维系中扮演了不可忽视的角色。然而，中东欧语境下的案例研究仍几近空白，德布勒森这类新兴工业移民社区尤为如此。

与此同时，现有研究多将语言能力作为衡量传承语习得成效的核心甚至唯一指标，对文化认同与语言能力之间关系的讨论相对薄弱。Li Wei（2014）从跨语言实践的角度指出，双语者的身份建构并不简单等同于其语言能力的分布状态，语言使用的萎缩与文化归属感的维系可以并行共存。在语言能力与认同关系的研究上，Lee，Her与De Costa（2024）通过对英国华裔学习者的生命史叙事分析，揭示了传承语认同发展的动态性与非线性轨迹，指出认同建构并不随语言能力的高低而线性起伏，而是在不同生命阶段、不同交际对象与场所中持续协商。Shen与Jiang（2023）对澳大利亚三个华裔移民家庭的叙事研究同样发现，儿童在传承语习得中并非被动的政策执行者，其个人意愿与情感体验对实际语言保持程度具有独立的影响。在教学干预层面，Cun（2025）对美国华裔幼儿社区课堂的观察研究表明，跨语言实践（Translanguaging）能够有效激活儿童的多语资源，促进其语言身份的正向建构，为传承语课堂设计提供了可操作的实证依据。上述研究多集中于英语圈，这一论断在中东欧华裔社区是否同样成立，目前尚无实证研究加以检验。

匈牙利德布勒森市提供了一个颇具典型性的观察窗口。作为匈牙利第二大城市，德布勒森近年来因宁德时代、宝马等中资及跨国工业项目的大规模落地，吸引了相当数量的中国移民家庭定居。2019年成立的德布勒森大学孔子学院是本地中文教育的主要制度性支撑，但其课程以非华裔学习者为主要服务对象，针对华裔儿童传承语习得的专项支持几乎空白。在这一背景下，华裔家庭在中文传承上几乎处于缺乏系统性支持的状态，其语言实践与认同建构呈现出怎样的面貌，值得深入考察。

基于此，本研究通过问卷调查与深度访谈相结合的混合方法，重点考察以下三个问题：德布勒森华裔儿童的中文使用域呈现怎样的分布特征？家长的期望结构与语言政策如何影响儿童的学习实践？语言能力与文化身份认同之间是否存在非对应关系，其背后的维系机制是什么？

二、研究设计与方法

为全面、深入地探究中东欧新兴华裔社区儿童传承语习得的真实图景，本研究在方法论上采取了结构化问卷与半结构化深度访谈相结合的混合研究范式。以下为本研究的田野观测背景与受访者画像。

2.1 研究对象与田野背景

本研究的田野点选定为匈牙利德布勒森市。受近年来中资企业集中出海投资的影响，该市华人人口结构发生了显著变化，从早期的商贸群体逐渐过渡到以企业外派高管、技术人员及陪读家属为主的新型移民社区。本研究采用目的性抽样，选取了居住在德布勒森市的华裔儿童及其家长作为研究对象。共回收有效华裔儿童问卷9份，华裔家长问卷11份。9位华裔儿童平均年龄14岁左右，其中有5位是自出生起就生活在匈牙利，有4位是近5年内从中国来到匈牙利生活。为了更深入地探究语言实践背后的微观机制，本研究选取了三个具有典型代表性的华裔家庭进行深度访谈。这三个家庭在移民形态、家庭语言政策及未来规划上呈现出显著的差异性：

个案A（在地扎根型）：父母均为中国籍，在德布勒森从事个体商贸经营。两名子女均在匈牙利本地出生，受访的小儿子（6.5岁）目前正在德布勒森大学孔子学院的“OTT-HOME”社区项目接受系统的识字与阅读训练。

个案B（潜在回流型）：父母均为中国籍，在当地经营中餐饮业。受访的小儿子（6岁）于5岁时才移居德布勒森，具有一定的国内语言环境基础。

个案C（跨国婚姻型）：母亲为中国籍，父亲为匈牙利籍。受访女童（儿童A）自出生起生活在匈牙利。家庭内部呈现复杂的跨语言交际网络（父母间使用英语，父女间使用匈牙利语，母女间使用中文）。

访谈前均已征得受访者（及未成年人监护人）的书面知情同意。

2.2 数据收集与分析工具

本研究采用问卷调查与半结构化访谈相结合的方法，以实现数据的三角互证。家长问卷重点考察家庭语言政策的隐性与显性特征、期望结构与教育资源投入；儿童问卷涵盖日常语言使用场景的频率分布、中文学习困难自评与自我文化身份认知。量化数据主要运用描述性统计进行分布呈现。访谈提纲围绕“家庭日常语言互动”、“中文学习的阻力与动力”以及“文化节庆体验”三个维度展开。单次访谈时长约45-60分钟。访谈录音转写后，运用主题分析法（Thematic Analysis）对文本进行编码，提炼出关键的主题范畴。受限于样本规模（儿童n=9，家长n=11），本研究定位为探索性质性个案研究，量化数据不作统计推断，核心论点的支撑主要依赖对定性材料的“厚描述”。

三、 德布勒森华裔儿童中文传承状况调查

基于前期收集的问卷数据与访谈文本，本节将从家庭环境、家长期望、语域分布及文化认同四个维度，多角度剖析该群体中文传承的微观现实。

3.1 家庭语言政策与实践形态

访谈显示，德布勒森华裔家庭的语言环境并非铁板一块，而是沿着一个从严格单语到高度混用的连续统分布。坚持“在家只说中文”的家庭属少数，更普遍的形态是一种有意识或无意识形成的功能性分区：情感表达、家庭伦理、与祖辈沟通等场景倾向使用中文，学业讨论、本地社交及娱乐消费则自然切换至匈牙利语。

这种分区并非单纯的语言妥协，而是折射出中文在家庭生活中所承担的特定功能定位——它是代际情感的传递媒介，而非全功能的日常交际工具。一位受访家长如此描述：“我跟她说中文，她回答匈牙利语，但我们都听得懂，时间长了就这样了。”这种被动形成的混合模式，在许多华裔家庭中已成为常态，并非家长的主动政策选择。

尽管执行方式各异，家长的语言态度却呈现出高度的一致性。中文被普遍视为族裔身份的核心标志与代际情感的连接载体，其维护动力更多来自文化使命感与身份焦虑，而非对语言市场价值的理性计算。“不

学中文，他就不是中国人了”——这类表述在访谈中反复出现，折射出家长对语言与身份之间关系的深层预设，尽管这一预设在儿童身上并未完全得到印证。

3.2 家长期望的结构特征

11份家长问卷在“对子女中文水平的最终期望”这一问题上呈现出鲜明的分布特征（见图1）。

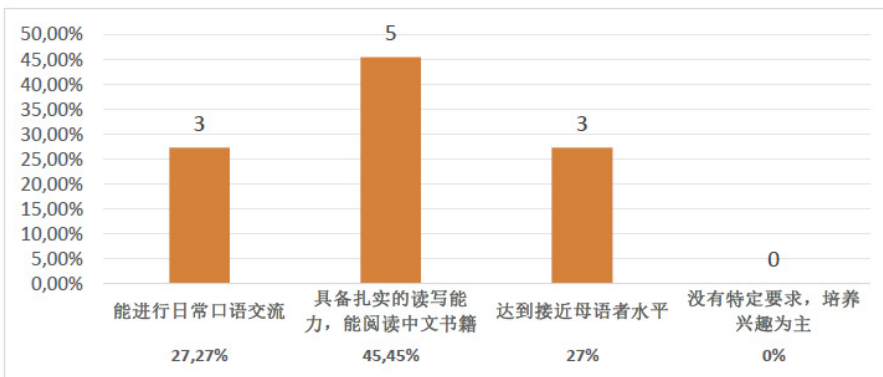


图1：华裔家长对子女中文水平的期望分布

最值得关注的数据是最后一项：无一位家长选择“以培养兴趣为主，没有特定要求”。这与公立学校非华裔少儿家长的普遍态度形成鲜明对比——后者多以兴趣培养为主要目标，对能力发展并无明确预期。华裔家长的目标设定是系统性的、能力取向的，且没有妥协余地。

这一高期望的形成有其内在逻辑。访谈揭示，家长的期望蓝图通常包含多个层次：基础层面是确保子女能与祖辈及中国亲属维持实质性沟通；中层是通过语言习得完成文化身份的代际传递；更深处则是对子女未来在双语劳动力市场中竞争优势的前瞻性布局。然而，这种接近“理想双语者”的培育愿景与儿童实际所处的语言环境之间存在相当距离，在体制内教育缺位的情况下，家庭亟需外部资源的介入，由于德布勒森本地缺乏专业的华裔儿童传承语教学机构，曾给华裔家庭带来巨大的时间与经济消耗。受访家长A指出，在德布勒森大学孔子学院“OTT-HOME”社区中文课开设之前，为了让子女接受正规的识字和阅读训练，他们不得不每周驱车前往首都布达佩斯上课，“时间成本很大，价格也很昂贵”。这一客观存在的资源瓶颈，导致许多家庭在中文传承上

长期处于“孤立作战”的困境。然而，“OTT-HOME”项目的在地化开展，显著改善了这一生态。在此说明，“OTT-HOME”项目是由德布勒森大学孔子学院主持开展的面向社会成员的中文听说读写课程项目。家长A明确表示，该课程不仅有效提升了儿童的读写水平，更重要的是通过集体的文化浸润，增强了儿童的文化归属感。

尽管问卷数据显示华裔家长普遍抱有高期望，但深度访谈揭示了这种期望背后的具体分化。对于潜在回流型的个案B而言，家长的首要焦虑是“语言磨蚀（Language Attrition）”。家长B坦言，虽然目前对孩子的整体水平要求不高，但核心底线是“怕他忘记该怎么写汉字”，因为家庭未来大概率会回流中国生活。这也印证了读写能力（尤其是汉字书写）是海外传承语保持中最脆弱、最易流失的一环。

3.3 中文使用的实际分布：语域收窄至家庭内部

9份华裔儿童数据呈现了一幅与家长期望形成明显落差的现实图景（见图2）。

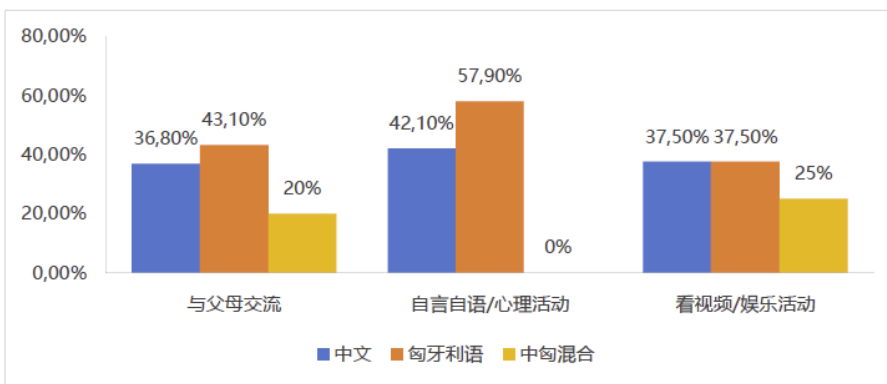


图2：华裔儿童在不同情境下的语言使用分布

数据显示，中文的实际使用场景高度集中于与父母的家庭内部对话；在自言自语、娱乐消费与同伴交流等情境中，匈牙利语均占据绝对主导地位。这是一种典型的“内在言语匈牙利化”[1]现象——当儿童独处

[1] 一个人的内在思考过程（内在言语），其运作的“语言系统”或“表达习惯”受到了匈牙利语的深刻影响，以至于匈牙利语成为了其思维的主要载体或“默认语言”。

时，其内部言语体系已完成向强势社会语言的切换，中文的输出主要由外部（家庭）要求触发，而非源于内在的自发需求。

语域收窄的成因是结构性的，而非个体态度问题。匈牙利语在学校、媒体与同伴网络中形成全方位的浸润优势，覆盖了儿童获取知识、建立社会关系、构建自我认同的绝大部分场景。中文在这一结构中能够维持的空间，只剩下家庭这一最后的“保留地”，且即便在家庭内部，受上文所描述的功能性混用模式影响，中文的实际使用也在持续萎缩。

读写能力的滞后是语域收窄最直接的后果之一，也是家长高期望与现实之间落差的核心体现。在缺乏系统性书面语输入环境的情况下，汉字认读与中文书面表达的发展几乎完全依赖家庭的主动干预，而这种干预在强度、系统性与专业性上均难以与正规教学机构相比。部分儿童的口语能够维持日常家庭对话的基本需求，却无法独立阅读一篇简单的中文文章，更遑论写作[2]。

在跨国婚姻家庭中，家庭语言政策（FLP）的执行面临更复杂的空间挤压。以个案C为例，受前文所述复杂家庭多语网络的影响，儿童呈现出典型的“不对称双语”状态：儿童能够进行日常中文对话，但不具备拼音基础，“不会认读，会写的汉字极少”。即便如此，中国籍母亲依然展现出强烈的中文维持意愿，将双语能力视为一种“竞争优势”，并期望借此维系对中国文化的感知。这表明，即便在混血家庭中，中文的工具性价值（语言优势）与整合性价值（文化认同）依然是家长坚持中文传承的双重驱动力。

3.4 文化认同的维系：语言之外的路径

上述语言能力图景与儿童的身份认同表达之间，存在一个引人关注的落差（见图3）。

即便在日常中文使用十分有限、读写能力明显滞后的情况下，本研究中9名受访德布勒森华裔儿童中，有5名认为自己“更像中国儿童”，另有4名儿童将自己定位为双重或混合文化身份。没有儿童将自己单一归类为“匈牙利人”。

---

[2] 以上判断来源于家长自我报告及访谈中儿童的自我评估，非标准化语言测试，结论应在此范围内理解。

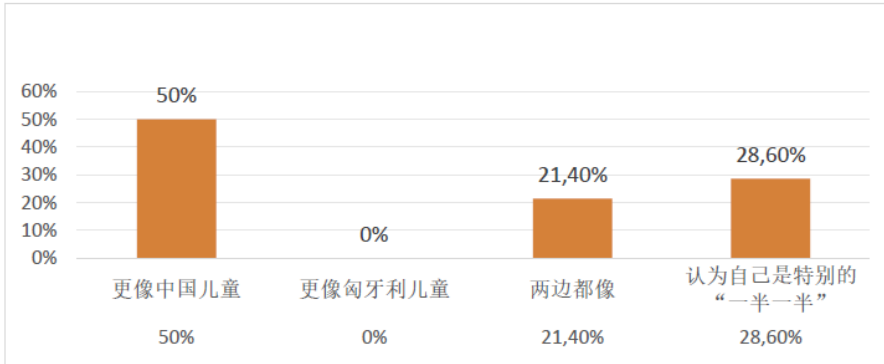


图3：华裔儿童自我文化身份认同分布

这一数据的意义在于：它表明文化认同的维系并不以语言能力为前提条件。进一步的访谈与开放题回答揭示了文化认同的实际维系载体——不是中文表达能力，而是一套由日常生活实践构成的情感经验：春节的红包与饺子、中秋的家庭聚会、饮食习惯中挥之不去的中国味道、祖辈定期的视频通话、家中悬挂的中国画与春联。这些仪式性的文化实践，在语言能力之外为儿童提供了持续的“中国性”体验，使其在认同层面保持与中国文化的连接。

值得关注的是，部分儿童在开放题中表达了一种特殊的心理张力：他们渴望能像“在中国长大的孩子”一样自如地使用中文，却清楚地意识到自己与这一理想形象之间存在距离。“我觉得我是中国人，但我的中文不够好”——这类表述折射出理想语言身份与现实能力之间的内在裂缝。这种裂缝既可能成为学习动力的来源，也可能在落差持续扩大时转化为自我效能感的损耗，进而加速学习动机的衰退。

四、华裔儿童中文传承的实践启示与对策建议

综合前文的实证调查结果，德布勒森华裔社区的语言实践为我们揭示了传承语习得中独特的运行逻辑与现实困境。本节将在此基础上提炼核心的理论与实践启示，并针对性地提出多维联动的干预对策。

4.1 实践启示：语言能力与文化认同的非线性演进与内在张力

本研究的核心发现可以概括为一个重要命题——在德布勒森华裔儿童群体中，语言能力的发展与文化认同的维系并非简单的线性正相关关系；文化归属感在很大程度上能够通过非语言路径得以独立维系，但这种维系并非没有隐患。

首先，这一发现与Valdés（2005）的论断形成了跨地域的实证呼应。传承语学习者的习得路径与认同逻辑有其自身的结构性特征。德布勒森的数据进一步证实，即便在语言能力发展受到强势社会语言（匈牙利语或英语）严重压制的情况下，本研究中多数受访华裔儿童依然可以通过家庭仪式、节庆参与等非语言渠道持续维系“中国性”体验。前文所述的日常跨国文化实践，实际上为儿童构建了一个物理空间之外的“文化飞地”，使得语言的功能性收缩并未自动导致文化认同的相应松动。

其次，这一结论对Wang（2020）关于家庭语言政策（FLP）的讨论既构成印证也形成了补充。德布勒森的数据证实了家长的高期望与高投入（如不辞辛劳跨城求学）对儿童保持中文具有正向促进作用。但更深层的启示在于，家长的日常语言管理与文化实践所塑造的不只是子女的语言能力，更是一个坚韧的“认同框架”。换言之，即便儿童“说”或“写”的能力存在短板，“是中国人”的内在感受依然能通过多重渠道得到确认。

然而，这种非线性关系并不意味着语言能力的停滞是毫无风险的。深度访谈揭示，理想语言身份与现实能力之间的落差，构成了儿童内在世界一种持续性的心理张力。这种张力在儿童尚小或处于跨国婚姻家庭（如个案C）时，可能表现为一种试图掌握双语优势的学习期待；但随着年龄增长、落差感加剧，尤其是在缺乏外部成就感正向激励的情况下，极易转化为对中文学习的自我效能感损耗乃至全面回避。如果不加干预，当语言能力长期停滞而家长的认同期待持续存在时，这种双向挤压的张力终将对儿童的母文化认同本身产生反噬与剥离。

4.2 对策建议：构建“家校社”协同的在地化传承支持网络

基于上述启示，德布勒森乃至中东欧新兴华裔社区的传承语教学亟须突破“家庭孤军奋战、家长充当草台班子”的单一模式。作为当地中文教育的核心制度性支撑，孔子学院应发挥枢纽作用，构建结构化的支持体系。具体建议如下：

第一，重塑家庭语言政策导向，从“母语者执念”向“功能与情感维系”转型。

在德布勒森缺乏全方位中文语境的现实下，家长单方面将传承语目标设定为“接近国内同龄人母语水平”是不切实际的，只会持续制造教育焦虑与代际冲突。鉴于潜在回流型家庭（如个案B）对汉字磨蚀的隐忧，教学机构的首要干预对象应当包含家长群体。建议教育机构或孔院定期开设“华裔家长语言规划工作坊”，协助家长建立科学的双语发展观。引导家长接受儿童的“不对称双语”现状，将务实目标降维至：首先维持中文作为家庭情感纽带的绝对地位，其次集中攻克读写能力的断层。通过循序渐进的家庭任务设计（如亲子共读中文绘本），使中文的使用场景逐步超越单调的“餐桌指令”边界。

第二，显性化身份认同建构，运用跨语言实践（Translanguaging）激活内在动机。

鉴于文化认同的维系具有相对独立性和抗性，这一优势应被教学设计有意识地加以利用。传统的对外汉语（L2）教学往往追求“全中文浸入”，但这在传承语课堂上容易让本就存在语言自卑感的华裔儿童感到挫败。针对个案C这类中外混血儿童或习惯用匈牙利语思维的华裔二代，课堂应积极引入跨语言实践理念（Translanguaging），即并非简单允许母语“夹杂”，而是将不同语言资源视为一个统一的语言库（unified linguistic repertoire），通过语言间的动态调用实现更深层的认知加工与身份认同建构（Li Wei，2014）。允许甚至鼓励他们在表达受阻时借用匈牙利语或英语作为支架（Scaffolding）。同时，将双语背景转化为课堂核心资源，例如开展“我的中匈双城记”主题演讲、中国节庆手工坊等活动。通过提供丰富的文化展示机会，将儿童对自身语言水平的“焦虑感”转化为拥有独特双重文化身份的“自豪感”，自内而外地激活其探索传承语的源动力。

第三，做深做实社区服务，依托孔子学院填补体制性与资源性缺口。

德布勒森大学孔子学院近期开展的“OTT-HOME”社区中文课提供了极佳的破局范本。该项目有效缓解了前文所述扎根型家庭（如个案A）跨城求学的资源瓶颈，及时填补了在地化传承语教学的体制性缺口。未来，孔院可从以下两个维度进一步深化在地化服务：一是实行精细化分层教学，针对不同移民形态实施定制化课程。例如，针对家长A与B的子女（有一定听说基础），开设以识字、阅读和中华文化体验为主的“传承语拔高班”；针对个案C（零起点混血儿童），则开设侧重基础交际与兴趣培养的“双语启蒙班”。二是首创“企院联动”机制，

结合德布勒森近年来宁德时代、宝马等大型跨国及中资企业入驻的宏观背景，孔子学院可联合这些在地企业，开展“华裔青少年中企参访日”或“未来双语人才夏令营”。这一举措能够将中文的价值从单一的“爷爷奶奶的语言”（整合性动机）拓展为极具市场竞争力的“未来职业语言”（工具性动机），用现实的经济与职业前景重塑高年级华裔儿童的中文学习意愿。

五、结语

本研究以德布勒森新兴工业移民社区为窗口，揭示了华裔儿童“中文功能性退缩”与读写滞后的“不对称双语”格局。研究核心发现提示，语言能力的弱化并未自动导致文化认同的剥离，多数儿童通过非语言路径维系了坚韧的“中国性”归属感。这一发现深化了对家庭语言政策隐性认同形塑功能的认知（Wang, 2020），但也警示了理想身份与现实能力落差所带来的潜在心理张力。

在实践层面，本文强烈呼吁，面对中东欧华裔儿童这一特殊群体，传承语教育必须走出“孤军奋战”与“唯母语者标准”的困局。德布勒森大学孔子学院“OTT-HOME”项目的在地化实践表明，体制性力量的介入至关重要。未来，依托孔子学院构建“家校社”协同网络，在教学中引入跨语言实践（Translanguaging）理念，并积极联动在地中资企业（如宁德时代、宝马等），将是破局中东欧传承语教育生态困境的必由之路。

同时，本研究也存在较大局限，尚属探索性质的个案考察。样本规模（儿童n=9，家长n=11）的局限性使得结论暂不具备广泛的统计推断效力；单一城市的横截面数据亦无法完整呈现语言与认同随年龄增长的动态演变轨迹。未来研究可从三个方向推进：其一，在中东欧多国（如捷克、波兰、罗马尼亚）华裔社区扩展样本，开展跨国家比较研究，以检验德布勒森模式的区域普适性；其二，引入纵向追踪设计，考察同一批儿童在青春期前后语言能力与文化认同之间关系的动态演变；其三，将跨语言实践（Translanguaging）教学干预作为准实验变量，检验其对传承语儿童语言自我效能感与学习动机的实际影响效果。

参考文献

康晓娟 2015. 海外华裔儿童华语学习、使用及其家庭语言规划调查研究——以马来西亚3～6岁华裔儿童家庭为例[J]. 语言文字应用, 2015(2): 26–34. [A Survey on Overseas Chinese Children's Chinese Learning, Use and Family Language Planning: A Case Study of Chinese Families with Children Aged 3–6 in Malaysia]

黎敏 2020."一带一路"沿线国家匈牙利的汉语教学现状及发展策略研究[J].国际汉语教育（中英文）, 2020 (02): 100–110. [Research on the Current Status and Development Strategies of Chinese Language Teaching in Hungary, a Country along the Belt and Road]

马燕华 2003. 初级华裔留学生汉语家庭语境调查与分析[J].华文教学与研究, 2003 (1): 47–53. [Investigation and Analysis of Chinese Family Context for Beginner Heritage Chinese Students]

牟芳芳, 郑咏滟 2022. 跨国家庭的家庭语言政策与儿童双语发展——以在沪中英/中法跨国家庭为例[J]. 外语界, 2022(3)：56–64. [Family Language Policy and Children's Bilingual Development in Transnational Families: A Case Study of Chinese–English and Chinese–French Families in Shanghai]

钟曼妮 2025. 匈牙利线上中文教学现状调查与分析[J]. 欧洲中文教育, 2025(2): 53–70. [Investigation and Analysis of the Current Status of Online Chinese Language Teaching in Hungary] https://doi.org/10.61637/cltcee.2025.2.3

Cummins, Jim. 2005. 'A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom,' The Modern Language Journal, 2005, 89(4): 585–592.

Cun, Aijuan 2025. 'Translanguaging practices in a community-based Chinese heritage early childhood classroom,' International Journal of Multilingualism, 2025, 29(3): 667–683. https://doi.org/10.1177/13670069241243204

Hu, Hong – Yagmur, Kutlay 2026. '"Jij bent Chinese! Ik niet! (You are Chinese! I am not!):" a case study of Chinese heritage language maintenance in the Netherlands,' Journal of Multilingual and Multicultural Development, 2026, 47(2): 1236–1250. https://doi.org/10.1080/01434632.2024.2418463

Lee, Her – De Costa 2024. 'Heritage language identity matters: tracing the trajectory of a Chinese heritage mother and contested Chinese dual language bilingual education,' Studies in Second Language Learning and Teaching, 2024, 14(1): 75–96. https://doi.org/10.14746/ssllt.42372

Li, Wei 2014. 'Translanguaging knowledge and identity in complementary classrooms for multilingual minority ethnic children,' Translanguaging in Education, 2014, 1(1): 11–26.

Shen, Chunxuan – Jiang, Wenying 2023. 'Parents' planning, children's agency and heritage language education: re-storying the language experiences of three Chinese immigrant families in Australia,' Frontiers in Psychology, 2023, 13:1083813. https://doi.org/10.3389/fpsyg.2022.1083813

Valdés, Guadalupe 2005. 'Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized?,' The Modern Language Journal, 2005, 89(3): 410–426. https://doi.org/10.1111/j.1540-4781.2005.00314.x

Wang, Yining 2020. Heritage language maintenance attitudes of Chinese migrant parents and their children (Ph.D. dissertation Melbourne Monash University)

Journal of Chinese Language Teaching in Europe 欧洲中文教育 2026/2, pp. 39–55

李雪

(布达佩斯罗兰大学)

匈牙利高级阶段国际中文教育中古语词习得及教学研究

摘要

古语词作为现代汉语词汇中的特殊词汇类型，在高级汉语学习中不可忽视。本文以匈牙利罗兰大学中文系的12名研究生为研究对象，采取课堂观察、问卷调查和访谈相结合的方法，对匈牙利高级阶段汉语学习者古语词习得情况进行实证分析。研究结果显示，学习者整体掌握水平偏低，古语词习得存在滞后性，受教学系统性欠缺、输入不足、学习策略单一及使用机会有限等因素的影响。基于研究结果，本文从教材、教师和学生三个维度提出针对性的教学建议，为匈牙利地区的国际中文教育及高级阶段的古语词教学提供参考。

关键词： 古语词，高级汉语教学，国际中文教育，匈牙利学习者

lixue2024918@gmail.com

https://doi.org/10.61637/cltcee.2026.2.3

https://orcid.org/0009-0002-1138-7303

Li Xue

(*Eötvös Loránd University, Budapest*)

A Study on the Acquisition and Teaching of Archaic Words in Advanced International Chinese Education in Hungary

Abstract

Archaic words constitute a distinctive category within the modern Chinese lexicon and should not be overlooked in advanced-level Chinese language learning. This study takes twelve master's students from the Department of Chinese Studies at Eötvös Loránd University in Hungary as research participants. Through a combination of classroom observation, questionnaires and interviews, the study conducts an empirical analysis of the acquisition of archaic words among advanced-level Chinese learners in Hungary. The findings reveal that learners' overall mastery is relatively low and that the acquisition of archaic words demonstrates a lagging effect. The learning outcomes are influenced by factors such as insufficient systematic instruction, limited input, reliance on single learning strategies, and restricted opportunities for use. Based on these findings, the study proposes targeted pedagogical suggestions from three dimensions: textbook, teachers, and learners, providing references for international Chinese education in Hungary and for the teaching of archaic words at the advanced level.

Keywords: archaic words, advanced Chinese language teaching, international Chinese education, Hungarian learners

一、引言

语音、词汇和语法共同构成语言三要素，词汇是语言要素中发展最快也是最复杂的要素，任何语言的学习都离不开词汇的学习。在高级汉语教学阶段，不仅对教师的词汇教学能力，也对学习者的词汇学习提出了更高要求。汉语学习者仅掌握基础词汇已难以满足高级阶段语言学习的需要。高级阶段的汉语教材文本多选材于中国经典散文、小说、诗歌及权威期刊，这就要求学习者不仅能够掌握现代汉语常用词汇，还需要理解并掌握具有特殊语言特征的词汇类型，如古语词。

古语词是现代汉语词汇体系的重要组成部分，同时也是具有特殊来源和功能特征的词汇群体，它来源于古代汉语，在词汇发展中被现代汉语吸收，区别于古代汉语词汇，成为现代汉语词汇中的成员。古语词常用于文学作品和历史著作，在公文和政论文中起到特殊的表达效果，很多古语词还反映着中华民族特定的风俗习惯和价值观念，是中国文化的重要表现形式，但古语词在语义表达和使用方式上保留了一定的古代汉语特征，语言形式相对凝练晦涩，这决定学习者在理解和习得上存在一些困难。

在匈牙利罗兰大学高级阶段的汉语教学实践中，所使用的教材内容涉及到较多的古语词，在教学过程中笔者发现，涉及古语词的教学内容，学生习得速度相对较慢，且就算理解掌握后在口头或书面语中也较少主动使用。而对希望全面提升汉语能力并达到较高语言水平的学习者而言，古语词是不可忽视的部分。因此对古语词教学开展系统研究，对于促进学习者现代汉语词汇习得、提升整体汉语水平以及深化中国文化理解具有重要意义。

对此，本文以匈牙利罗兰大学中文系研究生一年级的12名学生为研究对象，采用课堂观察、问卷调查和访谈相结合的研究方法，考察学生古语词习得情况，并对调查结果进行系统分析，探讨当前匈牙利高级阶段古语词教学过程中存在的困难和问题。并结合研究结果提出相应地教学建议，以期为匈牙利地区现代汉语词汇教学研究及汉语教学实践提供一些参考。

二、研究概述

2.1 古语词本体研究

古语词作为现代汉语词汇系统的重要组成部分，长期以来一直是现代汉语词汇研究的重要对象，现代汉语词汇在长期历史发展过程中逐渐形成，在此过程中部分以旧质要素为主的古代汉语词汇得以保留，并在现代汉语中继续使用，主要用于书面语表达或特定交际语境。古语词作为现代汉语词汇研究的重要内容，其本体研究长期受到学界关注，并取得了较为丰富的研究成果。现有研究主要围绕古语词的概念界定、范围划分、性质特征、分类标准及其与古代汉语词汇的关系等问题展开讨论，不同学者从不同角度对古语词的内涵和外延进行了阐释。

张世禄（1957）较早地系统论述了“古语词”这一词汇类别，明确区分古语词与历史词，并对其性质、范围及功能进行了详细分析。刘世俊（1980）进一步从概念和范围上对古语词、文言词及历史词进行阐述，就古语词和现代词的区分问题进行讨论。

刘叔新（1990）认为，古语词是产生于古代而在现代书面语中仍在沿用的词语，强调其历时延续性。葛本仪（2001）也从历时与共时两个维度探讨古语词问题，认为判断某一词是否为古语词，应结合在共时层面是否能够被重新启用加以考察。

孙常叙（2006）从内容与形式的关系以及语法功能出发，对汉语中古今词的区分标准进行了分析，指出现代汉语中仅能充当词素、不能独立成词，或在现代汉语中同音同义但词义发生变化的词，应归入古语词范畴。

黄伯荣、廖序东（2007）将古语词[1]界定为来源于古代文言著作，在现代汉语中仍在使用，可在特定语境中出现或用于构词的词语。根据其所指对象的不同，古语词可分为文言词和历史词两类，文言词所指事物或现象至今仍然存在，但具有庄重、典雅、简洁等书面语色彩；历史词所指事物或现象已经消失，主要用于历史叙述或特定场合。该界定在学界具有较高认可度，本文关于古语词的界定亦采用黄、廖的观点。

---

[1] 本文对古语词的界定采用的是黄伯荣、廖序东《现代汉语》（增订四版），北京：高等教育出版社，2007年中的观点，该观点在学界被广泛认可，即古语词是现代汉语中仍在使用、来源于古代汉语的词语，是现代汉语词汇系统的重要组成部分。需要说明的是，古语词并非泛指古代汉语中的全部词汇，而是指经过历史演变被现代汉语吸收并沿用至今的那部分词汇，已经完全消亡的古代词语不属于古语词的范畴。

邱峰（2010）从历时角度系统梳理了学界对古语词认识的发展过程，总结当前较为通行的观点，并对争议问题进行讨论，提出现代汉语中古语词的研究对象应聚焦于从古代汉语沿用下来的，既可独立使用又可充当语素的部分。

除概念界定外，古语词本体的少数研究聚焦在方言中的古语词及古语词在现代汉语中的应用情况。例如，肖九根（2007）、褚红&古敬恒（2010）从历史语言学视角考察方言中古语词的特点及其文化价值。在现代汉语应用层面，古语词研究多集中于公文写作、新闻语体及修辞表达等方面。如张燕峰（2009）系统分析了古语词在新闻报道中的使用情况；林万菁（1998）探讨了古语词所蕴含的文化意义与情感色彩在现代汉语修辞中的功能。

综上可知，古语词本体研究已形成较为清晰的理论框架，但相关研究主要集中于词汇学层面的概念界定与语言应用分析，对于其在第二语言教学中的实践转化问题关注不足。

2.2 国际中文教育中古语词研究

随着全球汉语学习者数量的增加及汉语学习需求的多元化发展，古代汉语的相关内容在国际中文教育中的作用逐渐得到大家的关注。虽然讨论古语词在国际中文教育中作用的研究相对有限，但仍有些研究者注意到了古语词的重要地位。

有些研究者在探讨古代汉语对国际中文教学的作用时强调了古语词的重要性。高惠敏（2001）认为古语词学习可以有效增加学习者的词汇量，并从教学目标、原则、作用及教学方法等多方面对其合理性和实施方案进行论述。朱炎炜（2007）提出，通过押韵规律的讲解可帮助学习者理解汉语韵母系统及同音现象，从而加深对同音字及一词多义古语词的掌握。

少数研究者以具体古语词为重点，分析该古语词在国际中文教学中的偏误情况及教学策略。郭翠翠（2015）基于HSK动态语料库，对外国学习者使用“其”时产生的偏误类型进行系统分析，并提出针对性的教学建议。杨荣华（2015）亦从多角度探讨学习者在使用“之”时的偏误特征，揭示古语词在理解与运用方面的难点。

有些研究以不同水平阶段的汉语学习者为对象，探讨古语词在教学实践中的功能及存在的问题。冯雨晴（2015）通过课堂观察与案例分析，研究高级阶段留学生古语词学习情况，并从教材设置、学习者特点

及课堂教学实施等方面分析问题成因。毛璟玲（2017）则以中级学习者为研究对象，对国际汉语教材中古语词的分布情况进行统计分析，探讨其教学意义及习得偏误特征。

现有研究虽在个案分析与教学策略探讨方面取得一定成果，但研究对象多以在华留学生为主，且多集中于个别词项或单一教学情境。以具体国家为单位，对特定国别学习者古语词习得状况进行系统实证分析的研究仍较为缺乏。

2.3 当前研究的不足与本文研究价值

目前古语词的本体研究已相对成熟，相关成果较为丰富。然而在国际中文教育领域围绕古语词教学的研究数量仍旧有限，研究视角与覆盖范围亦有待拓展，整体上尚缺乏系统而深入的探讨。现有研究多集中于古语词偏误分析及词汇教学策略等方面，且研究对象主要以在华留学生为主。以具体国家为研究单位，系统考察特定国别汉语学习者古语词的习得状况，并在此基础上探讨具有国别针对性的教学路径的研究尚较为匮乏。因此从宏观理论建构与微观教学实践两个层面来看，古语词在国际中文教育中的研究仍有进一步深化的必要。

基于目前研究现状，本文以匈牙利汉语学习者为研究对象，通过课堂观察、问卷调查与访谈相结合的方法，系统考察匈牙利高级阶段汉语学习者古语词的习得情况及其在教学过程中存在的问题。在实证分析的基础上，结合匈牙利学习者的语言背景与学习特点，提出具有针对性的教学策略与改进建议。以期在一定程度上弥补古语词教学国别研究的不足，为在匈牙利开展国际中文教育实践的教师及高级阶段的古语词教学提供可资参考的教学思路与经验。

三、研究方法

本文以匈牙利罗兰大学中文系研究生一年级的12名学生为研究对象，综合运用课堂观察、问卷调查与访谈等研究方法，对匈牙利高级汉语学习者的古语词习得情况进行实证分析。从理论上，本文通过梳理国内外古语词教学的研究成果来确定本文的研究方向与分析角度。在实证上，通过观察课堂、设计调查问卷来收集研究数据，结合卷后访谈获取

补充性研究资料。调查问卷以教材[2]中选取的15个高频古语词为测试项目，测试词语的选取主要遵循以下原则：一是词语在教材中出现频率较高，具有一定的教学代表性；二是属于高级阶段学习者应掌握的常见古语词，在现代汉语书面语中具有一定的使用频率；三是兼顾实词和虚词，以保证测试内容的代表性和涵盖面。通过测试考察匈牙利高级阶段汉语学习者对古语词的掌握情况，并结合卷后访谈进一步了解并分析学习者在古语词习得、理解和识别过程中存在的主要问题，从而为后续教学实践提出针对性的教学策略与建议。

四、匈牙利高级阶段汉语学习者古语词习得情况

4.1 问卷调查结果分析

本次调查问卷发放共12份，实收12份，问卷具体数据如下：

正确率水平	人数（n）	比例（%）
≥ 50%	2	16.7
< 50%	10	83.3
其中完全未掌握（0%）	3	25.0

表1[3] 匈牙利高级汉语学习者古语词测试成绩分布情况

词类类型	题数（n）	平均正确率
实词	10	46.3%
虚词	5	21.5%

表2[4] 测试古语词实词与虚词正确率占比

据表1的调查数据可知，在12名受试者中，仅有2名学习者测试正确率达到50%以上，占总人数的16.7%，正确率低于50%的学习者为10人占

[2] 学生现代汉语课堂所采用的教材是《博雅汉语·冲刺篇Ⅱ》
[3] 百分比基于总样本量N＝12计算。
[4] 表中样本量N=15，平均正确率按不同词类答对情况计算。

83.3%。其中有3名学习者未能正确识别任何古语词项目占总人数的25.0%。综合数据来看，匈牙利高级汉语学习者在古语词测试中的正确率水平相对较低。

从词类角度分析，根据表2测试结果显示，学习者对实词的掌握明显优于虚词。实词的平均正确率为46.3%，而虚词仅为21.5%。这一差异表明学习者在古语词的词类习得上存在明显的不均衡现象。实词由于语义具体，在语境中使用频率较高，学习者比较容易掌握，而虚词语义抽象、功能性较强，且在日常交流和文本中出现频率低，因此学习者对虚词的掌握低于实词。

从整体正确率分布来看，多数学习者掌握水平偏低，这说明古语词尚未成为稳定掌握的词汇内容，这一结果表明即使在汉语水平达到高级阶段后，学习者在古语词识别与理解方面仍旧存在困难，古语词知识的发展相对滞后于现代汉语词汇的发展。

测试结果显示学习者之间也存在一定的个体差异,部分学习者能够正确识别一定数量的古语词，而有些学习者则完全未能识别相关词汇。这种差异可能与学习者自身的语言接触环境和阅读经验等有关，具有较多书面语阅读经验的学习者可能更容易接触古语词，从而形成一定程度的词汇积累，而接触较少的学习者则难以获得相关语言经验，这种差异反映出学习者自身古语词掌握情况存在不均衡现象。

从词汇习得角度来看，相较于现代汉语常用词汇，古语词在当代语言交际中的使用频率相对较低，学习者在自然语言环境中接触相关词汇的机会有限，另外古语词虚词的功能性和抽象性又增加了学习难度。因此在缺乏系统输入和明确教学支持的情况下，学习者较难通过自然习得掌握古语词，这一特点可能是导致本研究中多数学习者整体正确率较低的重要原因。

本次调查问卷结果表明，匈牙利高级阶段汉语学习者古语词掌握程度整体有限，多数学习者尚未有效掌握相关词汇，同时存在明显个体差异和词类掌握不均衡现象。实词掌握相对较好，而虚词习得不足。

4.2 访谈分析

为进一步了解匈牙利高级阶段汉语学习者古语词习得情况及其相关影响因素，本研究在测试结束后对受试者进行了访谈。访谈内容主要涉及学习者汉语学习背景、古语词学习情况、学习经历、教学情况、学习

态度及学习策略等方面。并结合问卷调查结果对学习者古语词情况进行综合分析。

访谈结果显示大多数受试者汉语学习时间已超过四年，且学习目的主要为留学和就业，说明学习者具有较强的学习动机和较长的汉语学习经历。同时学习者通过课堂学习、阅读、影视观看以及日常交流等多种途径接触汉语，具备一定的语言输入基础。但结合测试结果可以发现，多数学习者古语词测试正确率仍处于较低水平，这表明古语词习得并未随着整体汉语水平的提高而同步发展，呈现出一定的滞后性特征，学习者难以通过一般语言输入掌握相关词汇。

在古语词接触与教学情况方面，虽然约70%的学习者表示曾在课堂上接触过古语词，但超过一半的学习者指出相关词汇未被作为重点内容进行系统讲解，即使进行讲解也主要通过现代汉语释义进行简单解释，这说明古语词在课堂教学中尚未形成系统的教学内容。从词汇习得角度来看，词汇的有效习得通常需要明确输入及语境支持，而古语词在教学过程中缺乏系统呈现，在这种教学输入不足的情况下，一定程度上影响了学习者古语词的习得效果。

在学习态度方面，超过80%的学习者对古语词学习兴趣不高，这反映出学习者对该类词汇的学习动机较弱。然而超过90%的学习者同时认为古语词学习对汉语学习具有积极作用，这表明尽管学习者对古语词学习兴趣有限，但仍能够认识到在汉语学习中的重要价值，只是在学习过程中缺乏内在动机。学习动机和学习兴趣是影响学习效果的重要因素，当学习者缺乏兴趣时，对相关语言项目的注意和投入程度就相对较低从而影响学习效果。

在学习策略与使用频率方面，大多学习者主要通过书写和机械练习等方式学习古语词，学习策略相对单一，同时学习者普遍反映古语词难以记忆，且由于实际使用机会较少容易遗忘，这说明学习者在古语词学习过程中缺乏有效的学习策略支持和充分的使用机会。多数学习者认为古语词在日常交际中使用频率较低，主要出现在书面语如文章和学术论文中，由于语言使用频率和词汇保持密切相关，较低的使用频率可能在一定程度上影响学习者对古语词的习得。

在教师教学上，学习者对多样化教学方式具有明确需求，学习者普遍倾向于教师采用多样化的教学方式，如通过举例说明、故事讲解及多媒体运用等方式进行古语词教学，这反映出学习者更容易接受情境化的教学方式。情境化教学有助于学习者将词汇与具体语境相结合，从而促进词汇的理解与记忆。因此多样化地教学方式可能有助于提高学习者对古语词的理解程度和学习效果。

4.3 分析与总结

综合问卷调查与访谈结果可以看出，匈牙利高级阶段汉语学习者在古语词习得方面整体水平有限。量化数据表明，多数学习者整体掌握率偏低，仅16.7%的学习者测试正确率达到50%以上，而有25.0%的学习者未能掌握任何古语词。从词类分析来看，学习者对实词的掌握明显优于虚词。访谈结果进一步揭示，古语词习得受多重因素影响，包括学习输入量不足、课堂教学系统性欠缺、学习策略单一、学习兴趣和动机差异以及实际使用机会有限等。

总体来看，高级阶段汉语学习者的古语词习得呈现滞后性，并伴随显著的个体差异。这些结果不仅验证了量化数据的发现，也为了解解古语词学习过程提供了行为与认知层面的实证解释，同时为今后古语词教学策略的优化提供了参考依据。

五、匈牙利高级阶段古语词教学的建议与启示

针对匈牙利高级阶段汉语学习者在古语词习得方面存在的问题，本文从教材、教师教学与学生学习三个层面提出具体建议，力求在教学实践中形成“教材系统呈现—教师策略优化—学生主动建构”的联动机制，促进古语词由被动识别向主动运用转化，从而提升匈牙利高级阶段国际中文教育中古语词教学的整体质量。

5.1 对教材编写的启示

5.1.1 建立古语词系统板块，增强教学系统性

近年来，面对国际中文教育的古代汉语教材建设虽取得一定进展，但针对不同国家学习者语言背景与学习特点，按国别进行差异化编写的古代汉语教材仍较为匮乏，这一问题亟待长期关注与系统解决。古语词教学不仅存在于专门的古代汉语课程中，也广泛渗透于中高级汉语综合课、阅读课等其他课程类型。然而当前各类教材对古语词的处理普遍缺乏系统性，因此如何在更广泛的教材体系中合理融入古语词内容，同样值得重视。当前匈牙利本土汉语教材的建设尚处于发展初期，在编写相关教材时，应充分关注古语词在语言教学中的特殊地位。尤其在高阶汉

语教材中，古语词分布零散，缺乏系统归类和重点呈现，会导致学习者难以对古语词建立完整的认知结构。因此在进行高级汉语教材编写时可在教材中设立古语词专栏，将教材中出现频率较高的古语词进行集中归类，并明确标注其来源、词义演变及适用语体。通过这种系统呈现让学习者在整体框架下理解古语词，避免词汇学习的碎片化输入，形成结构化的知识网络，为词汇的长期记忆和实际运用奠定基础。

5.1.2 区分词类，实施分层设计

本研究调查结果显示，虚词正确率显著低于实词，这表明在古语词教学过程中，不同词类在认知加工方式和习得难度方面存在差异。实词通常具有较为具体的语义内容，能够通过语境推断其基本意义，并在书面语中仍有一定使用频率，因而更易被学习者理解记忆。而虚词主要承担语法功能或逻辑关系表达，语义抽象、语法位置灵活，且往往缺乏明确的现代汉语对应形式，学习者难以通过直观理解掌握其功能，这在一定程度上解释了虚词正确率偏低的现象。对此教材编写应当实施词类分层设计，针对不同词类，采取不同的呈现方式与训练路径。如在实词教学方面，可侧重词义理解与语境运用，教材可通过提供多样化例句、固定搭配及语篇片段，引导学习者在语境中理解词义，对于与现代汉语意义存在差异或发生语义演变的实词，应通过对比说明其古今意义变化，使学习者意识到词义发展的轨迹，避免误用。教材在呈现虚词时，可单独设立专题讲解板块，对其语法功能和句法位置进行系统分析，避免简单释义处理。通过分层设计不仅可以降低学生的认知负担，还能提高教学针对性和效率。

5.1.3 增强文化与语境相结合

古语词不仅是语言系统中的词汇单位，更是历史文化的重要载体，其语义的构成与文化语境密不可分。如果教材仅以现代汉语释义的方式呈现，往往会削弱古语词的文化含义，使学习者停留在表层理解，难以真正掌握语用价值。因此高级阶段教材在呈现古语词时，应强化文化与语境的结合，适度增加古语词的历史背景说明，对于具有明显历史来源的词汇，可简要介绍其历史背景，学习者从而更容易建立联系，形成深层理解。还可以结合文学作品、历史典故或现实语境提供具体使用实例，文学语篇中常保留较多的古语词，且语境相对完整，有利于学习者

观察词汇在语篇中的实际功能，教材可选取适度简化的文学片段或现实文本材料，如新闻评论、公文语段等，在语境中展示古语词的搭配方式和表达效果。

5.2 对教师教学的启示

5.2.1提升教师专业素养与教学能力

教学效果很大程度上取决于教师的专业素养与教学能力，无论教师教何种内容，采用何种教学方法，教师专业素养与教学能力都不可或缺。在古语词教学中，教师应加强自身的古语词和词汇学知识储备，古语词因涉及词义演变、语法功能及历史文化背景，若教师对其来源和发展缺乏深入了解，教学讲解容易停留在表层释义层面。教师学习系统的古语词基础知识才能在课堂上进行清晰、准确的讲解，才能针对学习者疑问做出专业回应。此外教师应重视专业发展与教学反思提升自己的教学能力，可通过教学培训、学术研讨等方式，不断更新教学理念，并在实践中反思总结，如分析学生作业和测试结果来识别共同难点，调整教学策略，使古语词教学更贴合学生实际需要。

5.2.2增加输入与重复呈现

多数匈牙利汉语学习者在课堂中接触古语词的机会有限，缺乏持续性和系统性的输入支持，这种低频、零散的接触方式很难形成稳定记忆，也不利于学习者习得词汇。根据词汇习得规律，词汇掌握需要反复接触，才能实现由表层识记向深层记忆转化。因此在课堂教学中，教师应有意识地增加古语词的输入频率，强化学习效果。如可以在不同课文讲解、阅读训练、听力材料及课堂讨论活动中反复加入已学习的古语词，使其在多样化语境中重复出现，或将古语词与现代汉语词汇进行对比讲解也是提升输入的重要途径。通过提高课堂呈现频率、强化词汇输入和系统复习，可以为古语词的有效习得提供坚实基础。

5.2.3采取多样化教学策略

在匈牙利高级阶段古语词教学中，若教学方式单一很容易导致课堂形式固化，学习者参与度下降进而影响学习效果。访谈显示部分学习者对古语词学习兴趣不足，主要依赖教师讲解与机械记忆。对此教师应采

取多样化地教学策略，如通过情境创设、多媒体辅助及互动活动等方式，提高课堂吸引力和学习投入度。情景教学有助于增强古语词的语用体验，教师可围绕具体主题创设模拟情境，使古语词在特定语境中自然呈现，如在模拟正式会议或公文写作场景中引导学生使用带有书面色彩的古语词，使其体会词汇的语体特征与表达效果。而多媒体教学手段的引入能够增强课堂表现力，教师可利用电子课件、视频片段、影视作品等展示古语词在真实语境中的使用情况，让学生在视听结合的环境中感受古语词的表达氛围。互动式教学活动能够提高学习者的主动性，教师可组织小组讨论、角色扮演或合作任务，让学生在交流过程中有意识地使用所学词语，促进学习者在理解基础上实现灵活运用。

5.2.4激发学生学习兴趣，增强主动性

若学习者对古语词学习存在畏难心理将会影响学习的投入与保持。学习兴趣与内在动机是影响词汇习得的重要因素，若缺乏主动性，即使增加输入，也难以达到理想效果。在教学过程中，教师不仅应关注知识讲解，还要注重学生学习兴趣的激发，增强学习者的主动性。如可以通过任务教学法增强学习者的参与度，教师可以设计与学生兴趣相关的主题任务。还可营造积极的课堂氛围，教师在讲解古语词时可结合故事、历史趣闻等使教学更具有趣味性。此外，还应鼓励学习者进行自主探索，教师可推荐适合高级阶段阅读的材料，引导学生在课外阅读中主动学习古语词，从而增强持续学习的动力。激发学习兴趣与增强主动性是提高古语词教学效果的重要保障，只有学习者愿意主动学习运用古语词，才能真正实现词汇内化。

5.3 对学生学习的启示

5.3.1转变观念，发挥课堂主体作用

传统教学教师是主体，学生被动吸收教师所教授的知识，但在现代教学中学生已从课堂的参与者转变为课堂的主体，学生应认识到自身在课堂中的主体地位，充分发挥主观能动性。课堂教学并非单向知识传递的过程，而是教师引导与学生主动构建相结合的过程。在古语词学习中学习者应积极参与课堂，主动提出疑问，并在教师的引导下尝试归纳词义规律和语用特点，通过参与式学习加深对词汇的理解，而不是单纯停留在机械记忆层面。

5.3.2主动扩大输入与增加语言输出

古语词多用于书面语和正式语体，日常口语使用较少，学习者如果只依赖课堂学习，接触的机会有限，难以真正掌握，因此应主动拓展语言输入渠道，接触多样化的书面语材料，如散文、诗歌、小说、新闻报道及学术文章等，通过主动阅读，学习者在真实语境中理解掌握古语词的词义及用法。语言输出也是巩固和内化古语词的重要环节，学习者应有意识的在口语表达、写作和课堂活动中使用古语词，将词汇知识转化为实际运用的能力，如在文章写作或课堂展示中运用。通过多样化地输出，学习者不仅能巩固词汇，还能发现自身掌握的不足，并在实际运用当中提升熟练度。

六、结语

古语词作为现代汉语词汇体系中的特殊词汇类型，在高级阶段汉语教学中具有不可忽视的地位，对于高级汉语学习者而言，古语词的掌握一定程度上反映了其语言能力和书面语表达水平。本文根据笔者自身在匈牙利罗兰大学中文系的教学实践经历，针对匈牙利高级阶段的汉语学习者，来探讨古语词的习得与教学，基于问卷调查与访谈资料的实证分析发现，匈牙利高级阶段的汉语学习者在古语词习得方面整体水平偏低，存在词类掌握不均衡、输入不足、教学系统性欠缺及运用能力薄弱等问题，从影响因素看，古语词习得滞后并非由单一因素所导致，而是受教学方式、输入频率、学习策略、学习动机等多重因素影响。因此基于研究结果，本文从教材编写、教师教学与学生学习三个方面提出针对性建议，一定程度上为匈牙利高级阶段的古语词教学提供了教学实践思路。

本文以匈牙利汉语学习者为具体研究对象，拓展了古语词教学研究的国别视角，对国际中文教育情境下的高级阶段词汇教学研究具有一定的参考价值，但受研究对象规模与研究条件的限制，结论具有一定的局限性，未来可在更大研究范围内进行验证，进一步探讨古语词习得情况与其教学干预效果。加强古语词教学的理论与实践探索，对于提升国际中文教育高级阶段词汇教学的质量具有现实意义与长远价值。

参考文献

李晓琪，赵延风2015. 博雅汉语·中级冲刺篇Ⅱ[M]. 北京：北极大学出版社，2015.1. [Li Xiaoqi – Zhao Yanfeng 2015. *Boya Chinese: Intermediate Sprint II*. Beijing: Peking University Press.]

符淮青2004. 现代汉语词汇[M]. 北京：北京大学出版社，2004.4: 169– 172. [Fu Huaiqing. 2004. *Modern Chinese Vocabulary.* Beijing: Peking University Press: 169–172.]

黄伯荣, 廖序东2007. 现代汉语（增订四版）[M]. 北京：高等教育出版社，2007.6. [Huang Borong – Liao Xudong 2007. *Modern Chinese*. Beijing: Higher Education Press.]

刘叔新2005. 汉语描写词汇学[M]. 北京：商务印书馆. [Liu Shuxin 2005. *Descriptive Lexicology of Chinese*. Beijing: The Commercial Press.]

孙常叙2006. 汉语词汇（重排版）[M]. 北京：商务印书馆. [Sun Changxu 2006. *Chinese Vocabulary (Revised Edition)*. Beijing: The Commercial Press.]

刘叔新1984. 词汇学和词典学问题研究[M]. 天津：天津人民出版社，1984：42–68. [Liu Shuxin 1984. *Studies on Lexicology and Lexicography*. Tianjin: Tianjin People's Publishing House: 42–68.]

叶秋月2021. 匈牙利汉语教学的历史与现状——以罗兰大学中文系和罗兰大学孔子学院为例. 汉语国际传播研究, (01), 177–185. [Ye Qiuyue 2021. 'The history and current situation of Chinese language teaching in Hungary: A case study of the Chinese Department and Confucius Institute at Eötvös Loránd University,' *Chinese Language International Communication Research, (01)*: 177–185.]

叶秋月2023. 朗读在对外汉语语音教学中的应用——以匈牙利罗兰大学中文系学生为例. 国际中文教育(中英文), 8(01): 66–72. [Ye Qiuyue 2023. 'The application of reading aloud in teaching Chinese pronunciation to foreign learners: A case study of students at the Chinese Department of Eötvös Loránd University in Hungary,' *International Chinese Education (Bilingual), 8(01)*: 66–72.]

冯雨晴2015. 高级阶段留学生的古语词教学研究[D].华中师范大学. [Feng Yuqing 2015. *A Study on the Teaching of Archaic Words to Advanced International Students.* Central China Normal University.]

郭翠翠2015. 外国留学生文言词“其”的偏误情况分析[J]. 语文期刊, 2015(05): 29–30+37. [Guo Cuicui 2015. 'An analysis of errors in the use of the classical function word “qi” by international students,' *Language Periodical*, (05): 29–30, 37.]

刘世俊1980. 现代汉语中的古语词[J]. 宁夏大学学报(哲学社会科学版), 1980(03): 23–27. [Liu Shijun 1980. 'Archaic words in modern Chinese,' *Journal of Ningxia University (Philosophy and Social Sciences Edition)*, (03): 23–27.]

邱峰2010. 略论古语词. 商丘师范学院学报[J], 26(11): 120–122. [Qiu Feng 2010. 'A Brief Study on Classical Chinese Vocabulary,' *Journal of Shangqiu Normal University*, 26(11): 120–122.]

王勤1981. 略论现代汉语中的古语词[J]. 湘潭大学(社会科学学报), 1981(01): 94–97+87. [Wang Qin 1981. 'On archaic words in modern Chinese,' *Journal of Xiangtan University (Social Sciences Edition)*, (01), 94–97, 87.]

毛璟玲2017. 中高级阶段对外汉语教材古语词研究[D]. 湖南师范大学. [Mao Jingling 2017. *A Study of Archaic Words in TCFL Textbooks at the Intermediate and Advanced Levels*. PhD dissertation. Hunan Normal University.]

褚红, 古敬恒2010. 江苏方言古语词的历史积淀及若干特点[J]. 扬州大学学报(人文社会科学版), 14(02): 117–122. [Chu Hong-Gu Jingheng 2010. 'The historical accumulation and characteristics of archaic words in Jiangsu dialects,' *Journal of Yangzhou University (Humanities and Social Sciences Edition)*, 14(02): 117–122.]

肖九根2007. 赣方言古语词及其在汉语史研究中的实践价值. 江西社会科学, (06): 220－223. [Xiao Jiugen 2007. 'Archaic words in Gan dialect and their practical value in Chinese historical studies,' *Jiangxi Social Sciences*, (06): 220－223.]
刘延新1999. 古语词分布状况和使用频率考察——兼评《现代汉语词典》和《现代汉语频率词典》的收词[J]. 辽宁大学学报(哲学社会科学版), (02): 32－35. [Liu Yanxin 1999. 'A survey of the distribution and frequency of archaic words: With comments on "*Modern Chinese Dictionary*" and "*Modern Chinese Frequency Dictionary*",' *Journal of Liaoning University (Philosophy and Social Sciences Edition)*, (02): 32－35.]
张燕峰2009. 试论古语词在报纸新闻报道中的运用[M]. 浙江大学. [Zhang Yanfeng 2009. *On the use of archaic words in newspaper reporting*. MA thesis Zhejiang University.]
林万菁1998. 以借用古语词为例论修辞与文化的关系[J]. 平顶山师专学报, (05): 67－69+80. [Lin Wanjing 1998. 'On the relationship between rhetoric and culture from the perspective of borrowing archaic words,' *Journal of Pingdingshan Teachers College*, (05): 67－69, 80.]
李如龙, 吴茗2005. 略论对外汉语词汇教学的两个原则[J]. 语言教学与研究, (02): 41－47. [Li Rulong － Wu Ming 2005. 'Two principles of vocabulary teaching in Teaching Chinese as a Foreign Language,' *Language Teaching and Research*, (02): 41－47.]
张和生2005. 对外汉语词汇教学研究述评[J]. 语言文字应用, (S1): 6－8. [Zhang Hesheng 2005. 'A review of research on vocabulary teaching in Teaching Chinese as a Foreign Language,' *Applied Linguistics*, (S1): 6－8.]

附录

高级阶段汉语学习者古语词习得情况调查问卷

各位同学大家好！感谢您参与调查，本问卷旨在了解高水平汉语学习者古语词习得情况，所有问卷均用于学术研究，不作他用，问卷采取匿名方式，请放心填写！本问卷共计15题。

请解释下列划线词语

1、他总是一副斯文的模样。
2、要是生气，也不妨迁怒一下。
3、他上面还有高堂老母。
4、他擅长书法，尤擅草书。
5、霍光是镇守疆土的将军。
6、不要小看市井之地孕育出来的文化。
7、人有不同个性，猫，又何尝不是？
8、读了这篇文章，他感触颇深。

9、干坐着岂不着急？
10、凭你怎么斯文，那朵颐的样子也遮掩不住。
11、斗争之激烈，难以形容。
12、跟父母谈心之后，他不再彷徨了。
13、我驰骋在草原上。
14、学问是决不会辜负人的。
15、他们的水平参差。

访谈问题

1、你学习汉语多长时间？
2、你学习汉语的目的是什么？
3、主要通过什么途径学习汉语？
4、你在课上学习、接触过这些古语词吗？
5、你对学习这些古语词感兴趣吗？
6、教材中出现这些古语词时会被当作生词详细讲解吗？
7、教师在课堂上是采取什么方法讲解古语词？
8、在教学过程中你更倾向教师使用什么方法教授古语词？
9、你是采用什么方法来学习古语词？
10、你在学古语词过程中的困难是什么？
11、你学习过的古语词在日常生活中是否被应用，应用的多吗？
12、你认为什么场合会使用古语词？
13、你认为学习古语词对你学习汉语有帮助吗？

Journal of Chinese Language Teaching in Europe 欧洲中文教育 2026/2, pp. 57–69

Balázs Szigethy

(*Eötvös Loránd University Doctoral School of Linguistics Sinology Programme*)

Diachronic Annotation and Network-Graph Structuring in Chinese Character Pedagogy: A Component-Based Framework

Abstract

This article proposes a standards-aligned approach to Chinese character literacy that integrates two complementary components. First, a diachronic annotation file is prepared for each character, grounded in palaeographic and historical lexicographic evidence, and aligned with *bùshǒu* (部首) indexing standards (GF 0011-2009; GB 13000.1). Second, a network-graph approach represents *bùjiàn* (部件) as role-labelled components (semantic, phonetic, or neutral) within a dependency structure. The framework specifies a reproducible decomposition policy, a sequencing logic combining usage frequency with network topology (following Yan et al. 2013), and teacher-facing governance to ensure consistency and prevent folk etymology. The approach is intended to support recognition, pronunciation inference, and transfer, without presupposing measurable gains over existing approaches.

The proposal is intended as a character literacy module that complements, rather than replaces, communicative four-skill CSL instruction, and is most appropriate once learners have established a basic lexical and phonological foundation. An empirical study is outlined to evaluate learning outcomes relative to list-based instruction; claims regarding efficiency remain subject to future validation.

Keywords: Chinese character pedagogy, component-based instruction, network-graph approach, diachronic etymology, standardisation.

szigethybalazs@gmail.com

https://doi.org/10.61637/cltcee.2026.2.4

https://orcid.org/0009-0006-2188-1195

孔茗

汉字教学中的历时注释与网络图结构建构：一种基于部件的教学框架

摘要

本文提出了一种与国家标准对齐的汉字识读教学方法，由两个相互补充的组成部分构成。第一，为每个汉字建立一份历时性注释档案，以古文字学与历史词汇学研究为依据，并与部首索引体系（GF 0011-2009；GB　13000.1）保持一致。第二，构建网络图方法，将部件作为带有语义、语音或中性功能标注的组成单位，置于依存结构中进行呈现。该框架明确了一套可复现的分解策略，以及一种结合使用频率与网络拓扑的排序逻辑（参照闫小勇等，2013），并配套面向教师的治理机制，以确保分析一致性并避免民间字源学的偏差。该方法旨在通过结构化的部件意识，促进汉字识别、读音推断与迁移能力的发展。

本研究所提出的框架定位为一种汉字识读模块，用于补充而非取代以听、说、读、写为核心的综合性对外汉语教学体系，适用于学习者已具备基本词汇与语音基础之后的阶段。文中同时设计了一项实证研究方案，用以比较该方法与传统字表式教学在学习效果上的差异；关于教学效率的相关判断仍有待进一步实证检验。

关键词： 汉字教学法, 部件化教学, 网络图方法, 历时词源学, 标准化

Introduction

Although component-based approaches have been explored in CSL pedagogy (e.g., Taft – Chung 1999; Shen – Ke 2007), these have not resulted in a standardised, interoperable model aligned with both historical scholarship and modern indexing systems.[1] Across contexts, L2 Chinese instruction still leans on list-based rote memorization that treats characters as isolated units. This obscures the system's compositional nature: complex graphs are systematically built

[1] For example: Shen – Ke 2007: 106–109; Tong – Yip 2015: 173–177; Wong 2017: 211–225; Xu – Chang – Perfetti 2014: 789; Takashima – Verhoeven 2019: 77–79.

from recurrent components with relatively stable functional roles. A pedagogy that foregrounds internal structure—components, dependencies, and reusability—may support more efficient learning, better transfer across families, and more transparent assessment.

Prior work has represented characters as nodes in a Directed Acyclic Graph (DAG) where edges encode "is-a-subcomponent-of" relations (e.g., *mù* (木) → *lín* (林) → *sēn* (森)). Using 3500 high-frequency characters and principled decompositions, Yan, Fan, Di, Havlin, and Wu (2013) built a network with a giant component (n = 3687; 7024 links), revealing a hierarchical structure and Zipf-like offspring distributions. They defined learning cost as a function of unlearned subcomponents and proposed *Distributed Node Weight* (DNW) centrality—hybridizing usage frequency with topology (optimal b ≈ 0.35)—which outperformed frequency-only and naïve bottom-up baselines on efficiency: earlier coverage with lower cumulative cost, front-loading reusable components, and surpassing textbook orders.

Yet there is no shared pedagogical standard that reconciles historical analysis with modern orthographic conventions, prescribes a transparent sequencing policy that can be used in curricula and examinations, and provides teacher-facing tooling and quality governance to curb folk etymologies and inconsistent decomposition. The present article proposes a "quasi-Kodály" model (after Zoltán Kodály, 1882–1967, a Hungarian composer–educator whose method builds literacy from systematically sequenced elementary units and graded pattern acquisition) for Chinese characters: a standardised, component-first approach that aims to make aspects of the writing system's generative structure more explicitly teachable and analysable within a pedagogical setting.

The contribution is threefold. Conceptually, it sets out a standards-aligned way to teach components that respects historical scholarship whilst remaining interoperable with contemporary indexing and IT standards. Theoretically, it shows how a network-graph approach operationalises well-established principles in cognitive science and usage-based linguistics. Practically, it details governance, accessibility, and assessment practices that make the approach potentially adoptable at scale.

This proposal does not aim to replace comprehensive CSL curricula that develop speaking, listening, reading, and writing in parallel. Instead, it defines a modular orthographic strand focusing specifically on character literacy. In typical instructional settings, communicative competence—particularly speaking and listening—remains primary at the beginner stage, while structured character analysis is introduced progressively once learners have acquired a minimal working vocabulary. The present proposal is therefore intended as a complemen-

tary system for reading and writing development, not as a standalone approach governing all aspects of language acquisition.

The present article is conceptual in nature and proposes a testable framework rather than reporting empirical results.

Standardised Component Taxonomy

1. *Bùshǒu* (部首) vs. *bùjiàn* (部件): Division of Labour

The proposed standard begins with a clear division of labour between *bùshǒu* (部首) and *bùjiàn* (部件). *Bùshǒu* developed historically as a lexicographic device for finding characters. Xu Shen's *Shuōwén Jiězì* (说文解字) grouped roughly ten thousand graphs under 540 head graphs (indexing key components), a system later consolidated as 214 head graphs in the *Kāngxī Zìdiǎn* (康熙字典) and legally normalised in the PRC by GF 0011-2009 and GB 13000.1 for education, publishing, and information processing. *Bùshǒu* therefore answer the question "where do we look?", and they should remain the basis for interoperable lookup and assessment.

Learning, however, benefits most when instruction targets *bùjiàn* (部件), the synchronically functional subgraphs that serve as semantic or phonetic units, sometimes as both, and sometimes as neutral structural units (as in simplified or abbreviated/omitted forms). The terminological catch-all *piānpáng* (偏旁) is reserved here for historical discussion, since it blurs functional distinctions in the modern script. In teaching practice, each component instance in a character receives a single primary role—semantic, phonetic, phono-semantic or neutral/structural—determined by synchronic evidence and cross-family productivity. Thus, *qīng* (清) 'clear' is analysed as *shuǐ* (氵, the 'three-dots water' form of *shuǐ* 水), which conveys the semantic field of water/cleanliness and at the same time acts as the *bùshǒu* (部首), the indexing component, for lookup and retrievability purposes, together with *qīng* (青) serving as the phonetic component; *qíng* (情) 'feelings, emotions' as *xīn* (忄, the 'standing heart' form of *xīn* 心) plus *qīng* (青); *qǐng* (请) 'to ask, to invite' as *yán* (讠, the 'speech' form of *yán* 言) plus *qīng* (青). By contrast, differentiators without stable contribution (such as the dot distinguishing *yù* (玉) 'jade' from *wáng* (王) 'king') are treated as neutral. This neutral class becomes especially important in simplified forms where graphic residues persist without current semantic or phonetic function, and earlier character forms have to be analysed for semantic clarification. From

a pedagogical stance, it can be said that *bùshǒu* is for interoperable lookup, whilst *bùjiàn* is for decoding and learning.[2]

2. Allographs and position

In order to maintain interoperability with indexing and digital infrastructure whilst teaching the forms learners actually see, surface allographs — *shuǐ* (氵), *shǒu* (扌), *xīn* (忄), *yán* (讠), *jīn* (钅), *mì* (纟), *cǎo* (艹) — are mapped to the canonical lemmas *shuǐ* (水), *shǒu* (手), *xīn* (心), *yán* (言), *jīn* (金), *mì* (糸) and *cǎo* (艸) respectively for metadata (GF 0011-2009; GB 13000.1), while instruction uses surface forms.[3] Positional alternations such as *fù* (阝, from *fù* 阜 'mound') when located on the left side and *yì* (阝, from *yì* 邑 'city') when located on the right side of a character are distinguished with typical semantic tendencies.[4] Further terminological confusions arise with expressions such as *zuǒ'ěrpáng* (左耳旁), *yòu'ěrpáng* (右耳旁) and *shuāng'ěrpáng* (双耳旁), 'left ear' key, 'right ear' key and 'double ear' key respectively, which have nothing to do with the character for ear, nor do they have ear-related meanings; the resemblance is merely visual. This approach preserves dictionary skills and functional clarity. Characters are parsed by layout schema — *zuǒyòu* (左右), *shàngxià* (上下), *bāowéi* (包围), *zuǒzhōngyòu* (左中右), *shàngzhōngxià* (上中下), *pǐnzì* (品字), *chuānchā* (穿插) — and split along joints that maximise reuse of attested *bùjiàn* while avoiding ad-hoc fragments. This yields reproducible analyses and stable graph neighbourhoods. Each analysis stores: the *bùshǒu* (部首) in GF 0011-2009, *guībù* (归部)/stroke residuals under GF 0012-2009 for the GB 13000.1 character set, Unicode, and major variants of *Kāngxī* (康熙) characters. It has to be acknowledged that some heads (e.g., *piě* 丿) are convenience categories; indexing convenience should not leak into functional claims.[5]

The cognitive and linguistic rationale for a component-first approach can be stated without resorting to lists. Working memory is tightly constrained (≈4±1 units).[6] Treating characters as unrelated wholes imposes a heavy burden on novices, whereas teaching reusable subparts promotes chunking[7] and allows the same knowledge to be encoded verbally (labels, pinyin, glosses) and visually (layout and edge semantics), which improves retrieval.[8] Moreover, many

[2] Chen 2015: 44–50; Packard 2000: 310–317.
[3] Galambos 2006: 33–61.
[4] Taft – Zhu – Peng 1999: 498–502.
[5] Qiu 2000: 26–58.
[6] Cowan 2001: 88.
[7] Sweller 1988: 257–285.
[8] Clark – Paivio 1991: 166, 170–177.

character families align naturally with embodied image-schemas—verticality, path/goal, container—which help link orthographic form, morphemic meaning, and usage.[9] Category learning itself tends to organise around prototypes with graded extensions[10], a fact reflected in the behaviour of radical families such as *shuǐ* (氵), which prototypically signals liquids yet extends to flow or cleansing/cleanliness. Visual "radial maps" around prototypical cores help learners avoid over-generalisation. At the same time, the modern script is dominated by phono-semantic compounds. Treating shared phonetic elements as families reframes reading as hypothesis-testing: learners form pronunciation expectations for unseen series members and then adjust when they meet exceptions.[11] Finally, frequent form-meaning pairings in the language can be taught as constructions; modelling such patterns as higher-order nodes connected to exemplars supports both comprehension and production.[12]

These theoretical considerations motivate a specific stance on sequencing in the proposed approach. A sequencing policy should reflect both how often learners will encounter an item and how much it unlocks by way of dependents.[13] DNW-type centrality captures that logic by combining text frequency with network topology—reusability and dependency depth. Rather than treating DNW as a one-off optimisation that merely produces a better list, the present proposal treats the network itself as the curricular substrate. Syllabi, pacing guides, assessments, and analytics are all derived from the same graph. In practical terms, prerequisite components precede their dependents; characters that contribute many reusable subparts are introduced early; lesson design respects learnability by limiting the number of similar shapes introduced together; and coverage targets (e.g., HSK/CEFR thresholds) are reached with transparent trade-offs that can be tuned for particular cohorts. The analogy with Kodály is not ornamental: just as musical literacy is built from elemental patterns in graded repertoire, component literacy is built from a small set of high-utility forms whose reuse is then expanded along dependency paths.

While DNW centrality has demonstrated efficiency in network-based simulations, its pedagogical applicability requires empirical validation in classroom settings.

There is, nonetheless, a perennial tension between the mnemonic appeal of historical etymology and the demands of synchronic accuracy. Historical notes can vividly motivate forms such as *běn* (本), *xiū* (休), or *ān* (安), but not all

---

[9] Johnson 1987: 90–100; Lakoff – Johnson 1980: 14–21.
[10] Rosch 1975: 192–233; Lakoff 1987: 58–67.
[11] Perfetti – Tan 1998: 114; Taft – Chung 1999: 249–250.
[12] Goldberg 2009: 201–224; Ellis 2003: 82–95.
[13] Bybee 2006: 724.

historical motivations remain productive for modern learners. For this reason, every etymological comment in the Diachronic Annotation File is paired with a plain "Today it functions as…" line that states the current contribution of each component—semantic, phonetic, or ornamental—with no ambiguity. The result is a pedagogy that uses history responsibly whilst keeping present-day inference on firm ground. It also acknowledges that diagrams function as external memory: when relations are made persistent in the environment, novices can off-load some of the burden and rely on perception to guide retrieval.[14] Under realistic classroom time budgets this typically yields better near-transfer and delayed retention than linear lists, a pattern anticipated by the theory and borne out in prior studies grouping characters by component families.[15]

These requirements imply that the model is not immediately applicable in all instructional contexts and may be more suitable for digitally supported or research-oriented programmes.

From Visual Networks to Semantic–Phonetic Truth

A network-graph is only as good as the decompositions that license its edges. If the parsing is driven solely by modern regular-script silhouettes (or even graphic variants), the resulting graph inherits late graphic biases and the logic of traditional indexing rather than the historically correct semantic–phonetic relations that motivated character formation.[16] This produces precisely the sort of false generalisations that a principled component method is meant to avoid.

The case of *qiáng* (强 / 強 / 彊)[17] 'strong' is instructive. Diachronically, *qiáng* (彊) is analysable as *gōng* (弓) 'bow' contributing the semantic field of 'tension/force' (or the force needed to draw a bowstring) and *jiāng* (畺) functioning as the phonetic component. Later simplifications compact the right-side complex into an opaque block.[18] A visual parser that relies only on modern contours may mislabel the right side as a free-standing semantic element or dismiss it as an ornament, erasing the phonetic cue learners most need. The present account centres on visual reanalysis rather than phonological development; the phonetic trajectory of that component in *qiáng* (彊 / 強) as reconstructed for Old Chinese[19] is acknowledged but not explored here. Modern forms also contain many

[14] Risko – Gilbert 2016: 681–685.
[15] Xu – Chang – Perfetti 2014: 789; Takashima – Verhoeven 2019: 77–79.
[16] Qiu 2000: 208; Boltz 1994: 158–167.
[17] Schuessler 2007: 427.
[18] Su 2009: 34–35.
[19] cf. Baxter – Sagart 2014. Their work systematically reconstructs Old Chinese phonology and examines phonetic series, sound change, and related semantic shifts—demonstrating how pho-

box-like components that resemble *kǒu* (口) 'mouth/opening' but are unrelated historically. Examples such as *wéi* (韦 / 韋), *chū* (出), *gè* (各), or *zhèng* (正) easily tempt learners (and automated tools) to create spurious "mouth" connections, when in fact the shapes descend from enclosures, frames, places, cities or other complexes. Without appropriate diachronic anchoring, such mislabelling proliferates.

For an ideal standardised syllabus, the registry behind the graph must therefore carry diachronic metadata for each component (oracle/bronze → small seal → clerical → regular) with role tags at each stage; inheritance rules that preserve phonetic tagging when simplification compresses a complex into an opaque block; disambiguation of look-alikes that only mimic *kǒu* (口); and explicit alignment with contemporary indexing standards that also notes where indexing convenience diverges from historical function. In classroom tools, a guiding principle is that visual edges are licensed by etymology. The interface may show modern components for recognition and lookup, but every node must be able to reveal its period-specific role and series membership on demand. In practice, a learner hovering over the right side of *qiáng* (强) would see that the element corresponds to *jiāng* (畺) with phonetic value in earlier stages, and the network would connect *qiáng* (强) to its phonetic series siblings rather than to arbitrary rectangles. Where a phonetic element also carries meaning, the interface should render both roles explicitly.

A standards-aligned approach requires infrastructure. On the authoring side, teachers need assisted decomposition checks, graph builders that enforce role labelling, and exportable network sheets for classroom use. On the programme side, governance must include versioning, audit trails, peer review of proposed changes, and explicit citation of palaeographic claims, so that the standard remains both stable and improvable. Accessibility is designed in from the start: colour-independent encodings, screen-reader semantics, keyboard navigation, and functional rather than decorative visuals. These choices are not cosmetic. They determine whether institutions can adopt the approach at scale without producing new inequities.

Position within Four-Skill CSL Instruction

A central concern in CSL pedagogy is the balance between communicative competence and orthographic knowledge. The present framework is designed to operate within this balance by targeting character literacy as a distinct, but

netic evolution can reshape meanings, relations, and category boundaries over time.

integrated, component of instruction; furthermore, it does not assume that character structure determines communicative competence, nor that orthographic structure is a primary driver of overall language acquisition; rather, it targets specific aspects of literacy development.

At the beginner level, learning priorities centre on pronunciation, basic vocabulary, and functional communication. For initial, high-frequency expressions such as *nǐ hǎo* (你好), learning is optimally achieved through communicative practice; structural decomposition is introduced only once learners possess sufficient familiarity with the writing system to benefit from it.

The network-graph approach becomes relevant at a later stage, when learners encounter a growing number of characters and require strategies for managing orthographic complexity. At this point, component awareness supports recognition, generalisation across character families, and informed pronunciation inference.

The diachronic annotation system is primarily teacher-facing in early stages, with only selectively disclosed information presented to learners. This principle of progressive disclosure ensures that historical and structural knowledge supports learning without introducing unnecessary cognitive load.

In this sense, the framework aligns with communicative CSL instruction by strengthening literacy outcomes while remaining subordinate to broader language-learning objectives.

The model does not claim that orthographic structure is a primary driver of overall language acquisition, but rather that it can provide targeted support within the domain of literacy.

Methodology

The study evaluates character-level outcomes (recognition, retention, and inference) as one component of overall language development, without claiming to measure full four-skill proficiency. The empirical study is designed to evaluate the core claims of the proposal under realistic classroom conditions. It asks the following questions:

RQ1: *Does network-graph instruction improve immediate and delayed recall and recognition compared to traditional list-based teaching?*

RQ2: *Do the effects of network-graph instruction generalise across different L1 backgrounds and levels of prior exposure?*

RQ3: *What cognitive-load and usability trade-offs arise in network-graph instruction, and does progressive disclosure help to reduce overload?*

RQ4: *Does network-graph instruction enhance learners' ability to infer the pronunciation of previously unseen characters that share phonetic components with known families?*

RQ5: *Does network-graph instruction enhance learners' ability to infer semantic categories or approximate meanings of unfamiliar characters by recognising shared semantic components?*

Participants include L2 learners with varied L1s, stratified by prior exposure. Materials consist of matched character sets balanced for frequency, stroke count, and family membership; all characters in the network condition are supported by brief Diachronic Annotation Files pairing historical notes with a clear statement of the current function of each component. Instruction in the network condition proceeds by decomposition and reassembly, practice over phonetic and semantic families, and the use of construction nodes, with ordering guided by DNW-style centrality and lessons staged to avoid visual interference. The comparison group studies frequency-ordered lists with the same examples but without graphs. Measures include a brief cognitive-load instrument, immediate and delayed recognition tests, pronunciation prediction for unseen series members, sense classification tasks, short construction tasks, and near/far transfer probes in which students place unseen items into an existing network. Analysis uses mixed-effects models with participant and item as random factors, preregistered contrasts, and time-on-task controls.

The study is designed to test whether network-graph instruction leads to differences in subjective load, retention, pronunciation inference, and sense classification relative to list-based approaches.

For each taught character the programme produces two artefacts. The *Diachronic Annotation File* is a concise, referenced, standards-aligned note that links the historical dossier to modern function. The *Network Sheet* presents the character's local neighbourhood with role-labelled nodes and family links; it is designed for teaching, practice, and assessment. Roll-out aligns with HSK/CEFR levels; teacher training focuses on reproducible decomposition and on how to read and build network sheets; and a public repository tracks issues and proposed changes. Known limits are acknowledged: phonetic families contain irregulars; some simplified forms obscure earlier cues; and diachronic certainty varies. The standard addresses these with family-level exception tags, explicit neutral labels, and transparent confidence notes.

Conclusion

This article proposes a two-part, standards-aligned approach to Chinese character literacy that integrates diachronic annotation with network-graph instruction. The framework is designed to support structured inference and component awareness while remaining compatible with existing CSL curricula. Part I—*diachronic annotation*—preserves *bùshǒu* (部首) indexing for interoperability with dictionaries and examinations while correcting common misanalyses through palaeographic evidence. Part II—*network-graph instruction*—teaches *bùjiàn* (部件) as role-labelled components within a dependency graph, operationalising findings on chunking, dual coding, and phono-semantic regularities.

The proposed framework offers three properties that make it adoptable at scale:

1. Strict alignment with national standards;
2. Reproducible decomposition under peer-reviewed governance;
3. Graph-aware sequencing that integrates usage frequency with reusability.

An empirical study has been outlined to test recall, retention, cognitive load, and transfer, including learners' ability to infer pronunciation and meaning of unfamiliar characters.

In sum, the approach reframes character learning from rote memorisation to structured inference, while preserving the librarian's path of traditional lookup. Implemented as paired *Diachronic Annotation Files* and *Network Sheets*, it offers a potential basis for a more systematic descriptive framework—historically accurate, pedagogically motivated, and accessible across contexts.

The proposal is deliberately bounded. It is not intended as a comprehensive CSL approach, nor is it designed for absolute beginners, and its effectiveness will depend on appropriate sequencing, teacher training, and integration with communicative instruction. It is therefore best understood as a theoretically motivated and operationalised proposal, whose pedagogical value remains to be established through empirical study under classroom conditions.

Bibliography

Baxter, William H. – Sagart, Laurent. 2014. *Old Chinese: A New Reconstruction*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199945375.001.0001

Boltz, William G. 1994. *The Origin and Early Development of the Chinese Writing System*. American Oriental Series, vol. 78. New Haven, CT: American Oriental Society.

Bybee, Joan. 2006. 'From Usage to Grammar: The Mind's Response to Repetition,' *Language* 82 (4): 711–733. https://doi.org/10.1353/lan.2006.0186

Chen, Chunlei. 2015. '现代汉语教材中偏旁、部首和部件的定义问题 [Definitional Issues of *Piānpáng*, *Bùshǒu*, and *Bùjiàn* in Modern Chinese Textbooks],' *Academic research on language and writing* (5): 44–50. https://doi.org/10.14014/j.cnki.cn11-2597/g2.2015.05.013.

Clark, James M. – Paivio, Allan 1991. 'Dual Coding Theory and Education,' *Educational Psychology Review* 3: 149–210. https://doi.org/10.1007/BF01320076

Cowan, Nelson. 2001. 'The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity,' *Behavioral and Brain Sciences* 24 (1): 87–114. https://doi.org/10.1017/S0140525X01003922

Ellis, Nick C. 2003. 'Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure', in Catherine J. Doughty – Michael H. Long (eds.). *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell, 63–103. Reprinted 2008. https://doi.org/10.1002/9780470756492.ch4

Galambos, Imre. 2006. *Orthography of Early Chinese Writing: Evidence from Newly Excavated Manuscripts*. Budapest: Department of East Asian Studies, Eötvös Loránd University, (Budapest Monographs in East Asian Studies 1): 33–61.

Goldberg, Adele E. 2009. 'Constructions work,' *Cognitive Linguistics*, vol. 20 (1): 201–224. https://doi.org/10.1515/COGL.2009.013

Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226177847.001.0001

Lakoff, George – Mark Johnson. 1980. *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lakoff, George. 1987. *Women, Fire and Dangerous Things*: *What Categories Reveal about the Mind.* Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226471013.001.0001

Ministry of Education of the People's Republic of China [中华人民共和国教育部] and National Language Commission [国家语言文字工作委员会] 2009. *Specification for Identifying Indexing Components of the GB 13000.1 Chinese Character Set* [*GB13000.1*字符集汉字部首归部规范]. Beijing: Chinese Language and Literature Publishing House [语文出版社]. Standard GF 0012–2009. http://www.moe.gov.cn/jyb_sjzl/ziliao/A19/200901/t20090102_186104.html (last accessed: 02.10.2025.)

Ministry of Education of the People's Republic of China [中华人民共和国教育部] and National Language Commission [国家语言文字工作委员会] 2009. *The Table of Indexing Chinese Character Components* [汉字部首表]. Beijing: Chinese Language and Literature Publishing House [语文出版社]. Standard GF 0011–2009. Released 2009-01-12; implemented 2009-05-01. https://archive.org/details/GF0011-2009/page/8/mode/2up (last accessed: 02.10.2025.)

Packard, Jerome. L. 2000. *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486821

Perfetti, Charles A. – Li-Hai Tan. 1998. 'The Time Course of Graphic, Phonological, and Semantic Activation in Chinese Character Identification,' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24(1): 101–118. https://doi.org/10.1037/0278-7393.24.1.101

Qiu Xigui [Mattos, Gilbert L. – Norman, Jerry (trans.)] 2000. *Chinese Writing. (Early China Special Monograph Series No. 4)*. Berkeley: The Society for the Study of Early China – The Institute of East Asian Studies, University of California.

Risko, Evan F. – Sam J. Gilbert. 2016. 'Cognitive Offloading'. *Trends in Cognitive Sciences* 20 (9): 676–88. https://doi.org/10.1016/j.tics.2016.07.002

Rosch, Eleanor. 1975. 'Cognitive Representations of Semantic Categories.' *Journal of Experimental Psychology: General, 104(3):* 192–233. *https://doi.org/10.1037/0096-3445.104.3.192*

Schuessler, Axel. 2007. *ABC Etymological Dictionary of Old Chinese*. Honolulu: University of Hawai'i Press.

Shen, Helen H. – Cheng-Chao Ke. 2007. 'Radical Awareness and Character Acquisition,' *The Modern Language Journal* 91 (1): 97–111. https://doi.org/10.1111/j.1540-4781.2007.00511.x

Su, Peicheng. 2009. '简化汉字60年 [Sixty Years of Simplified Chinese Characters]'. 语言文字应用 [*Applied Linguistics*], 4: 26–35. https://doi.org/10.16499/j.cnki.1003-5397.2009.04.004

Sweller, John. 1988. 'Cognitive load during problem solving: effects on learning,' *Cognitive Science, 12(2):* 257–285. https://doi.org/10.1207/s15516709cog1202_4

Taft, Marcus – Kevin Chung. 1999. 'Using Radicals in Teaching Chinese Characters to Second Language Learners,' *Psychologia* 42 (4): 243–251.

Taft, Marcus – Xiaoping Zhu – Danling Peng. 1999. 'Positional Specificity of Radicals in Chinese Character Recognition,' *Journal of Memory and Language* 40 (4): 498–519. https://doi.org/10.1006/jmla.1998.2625

Takashima, Atsuko – Ludo Verhoeven. 2019. 'Radical Repetition Effects in Beginning Learners of Chinese as a Foreign Language Reading,' *Journal of Neurolinguistics* 50: 71–81. https://doi.org/10.1016/j.jneuroling.2018.03.001

Tong, X. – Yip, J.H.Y. 2015. 'Cracking the Chinese character: radical sensitivity in learners of Chinese as a foreign language and its relationship to Chinese word reading,' *Reading and Writing* 28: 159–181. https://doi.org/10.1007/s11145-014-9519-y

Wong, Yu Ka. 2017. 'The Role of Radical Awareness in Chinese-as-a-Second-Language Learners' Chinese Character Reading Development,' *Language Awareness* 26 (3): 211–225. https://doi.org/10.1080/09658416.2017.1400039

Xu, Y. – Chang, L. – Perfetti, C. A. 2014. 'The effect of radical-based grouping in character learning in Chinese as a foreign language,' *The Modern Language Journal, 98(3):* 773–793. https://doi.org/10.1111/modl.12122

Yan, Xiaoyong – Ying Fan – Zengru Di – Shlomo Havlin – Jinshan Wu. 2013. 'Efficient Learning Strategy of Chinese Characters Based on Network Approach.' *PLoS ONE* 8 (8): e69745. https://doi.org/10.1371/journal.pone.0069745

Journal of Chinese Language Teaching in Europe 欧洲中文教育 2026/2, pp. 71–75

Anna Csikó

Looking Back on the Past 20 Years of the ELTE Confucius Institute

On December 7, 2006, Hungary's first Confucius Institute opened its doors at Eötvös Loránd University. Over the past twenty years, the Institute has accumulated not only extensive experience and many cherished memories, but also numerous achievements. Already in its second year of operation, in 2007, it received the Hanban Excellence Award, which it was honored with again in 2013.

Professor Ferenc Hudecz, former rector of ELTE and a dedicated supporter of the Institute, was presented with an Excellence Award in 2009 in recognition of his work, and in 2011 he was appointed honorary member of the Central Governing Board of the Confucius Institutes.

In the same year, Imre Hamar received the 'Outstanding Confucius Institute Director' award in recognition of his leadership. The distinction was awarded to only 20 directors out of the 350 Confucius Institutes and 500 Confucius Classrooms worldwide.

The important role played by the ELTE Confucius Institute in Hungarian – Chinese relations is clearly demonstrated by the fact that in 2011, Wen Jiabao visited the Institute.



Fig. 1. Chinese Premier Wen Jiabao Visits the Confucius Institute

Another significant achievement was the establishment of the Eötvös Loránd University Beijing Center at Beijing Foreign Studies University in 2015, making ELTE the first Hungarian university to open a center in China. In the same year, the Institute was awarded the title of 'Model Confucius Institute,' a distinction granted for the first time to only 15 out of 500 Confucius Institutes worldwide.

In December 2020, the Institute received the 'Outstanding Chinese Test Center 2020–2021' award, while Imre Hamar was honored with one of the most prestigious distinctions awarded by the Chinese government, the 'Special Book Award of China.' The Institute also received the 'Outstanding Chinese Test Center' award again in 2021, 2022, 2023, and 2025.

In July 2021, our partner university, Beijing Foreign Studies University, recognized the outstanding work carried out during the 2020/2021 academic year by presenting our Chinese director, Li Denggui, with the 'Outstanding Confucius Institute Director' award. He has served as the Institute's Chinese director since September 2019.

In September 2021, Imre Hamar was awarded the 'Confucius Institute Directors' Commemorative Medal' in recognition of his fifteen years of service as director.



Fig. 2. Professor Imre Hamar Receives the "Confucius Institute Directors ' Commemorative Medal"

In 2023, we reached a major milestone: the 100th anniversary of the founding of the Department of Chinese Studies within the Institute of East Asian Studies, Faculty of Humanities, Eötvös Loránd University. To commemorate this distinguished occasion, in June 2023 we organized an international conference entitled ‘ELTE – Cambridge Doctoral Conference to Commemorate the 100th Anniversary of the Establishment of the ELTE Chinese Department’, while in November 2023 we held a grand anniversary gala in the Aula Magna.

In 2024, our Institute received numerous prestigious awards: both Imre Hamar and Li Denggui were honored with the title ‘Expert of the Confucius Institutes.’ In addition, Imre Hamar received the Chinese Embassy’s ‘Outstanding Sinologist’ award, while Ferenc Hudecz was presented with the ‘Excellence of the Confucius Institutes’ distinction. Our examination center was also awarded the title ‘Model Chinese Test Center’ in recognition of its outstanding achievements in language testing. This distinction was granted for the first time to only 19 centers out of 1,404 Chinese language test centers worldwide.



Fig. 3. Professor Ferenc Hudecz with the “Excellence of the Confucius Institutes” Award

One of the ELTE Confucius Institute's largest language-education projects, jointly supported by the Hungarian government and Hanban, was the compilation and publication of the Chinese – Hungarian Dictionary and the Hungarian – Chinese Dictionary. The ceremonial presentation of the Chinese – Hungarian Dictionary took place at the closing event of the 16+1 Summit held in Budapest in 2017, at the Hungarian Parliament. Imre Hamar presented a deluxe copy of the newly published, more than 2,000-page, two-volume Chinese – Hungarian Dictionary to Prime Minister Viktor Orbán, while Melinda Pap, lecturer at the ELTE Chinese Department, presented another deluxe copy to Prime Minister Li Keqiang. Since then, the Hungarian – Chinese Dictionary has also been completed, and its revised online edition is scheduled for publication this year.

Over the past twenty years, the Institute has also achieved significant results in teacher training and research. In November 2014, the Central and Eastern European Regional Teacher Training Center of ELTE officially opened its doors. Since then, it has organized twelve teacher-training programs and one webinar.



Fig. 4. Closing Ceremony of the Central and Eastern European Regional Teacher Training Center at ELTE, 2025

In September 2016, the Senate of Eötvös Loránd University approved the establishment of the 'One Belt, One Road' Research Center, whose mission was twofold: on the one hand, to study the cultures along the ancient Silk Road, and on the other, to analyze the economic and political dimensions of the modern Silk Road initiative. The center was officially inaugurated during the celebrations marking the tenth anniversary of the ELTE Confucius Institute. Closely con-

nected to this initiative was the MTA–ELTE–SZTE Silk Road Research Group, founded by Imre Hamar in 2017 and active until 2022.

In 2022, a new research group was established within the ELTE Confucius Institute: the ELTE Research Group on the Roman World and the Far East, led by archaeologist Krisztina Kinga Hoppál. The group focuses on the study of long-distance connections in antiquity, particularly in relation to the Roman world, Southeast Asia, and China.

In June 2024, we celebrated the opening of another joint Chinese – Hungarian research center, established together with Renmin University of China under the name 'China – Hungary Joint Research Centre for Exchanges and Mutual Learning between Civilizations.'



Fig. 5. 20th Anniversary Celebration of the ELTE Confucius Institute

AUTHORS OF THE VOLUME

Balázs Szigethy

Eötvös Loránd University, Doctoral School of Linguistics, Sinology Programme
Email: szigethybalazs@gmail.com
ORCID ID: 0009-0006-2188-1195

Balázs Szigethy is a doctoral candidate in the Sinology Programme of the Doctoral School of Linguistics at Eötvös Loránd University, Budapest. His research concerns the teaching of Chinese characters to adult learners of Chinese as a foreign language, with a focus on component-based methodology, the pedagogical use of palaeographic evidence, and the alignment of teaching materials with PRC language standards.

Hongxue Pang

Master's student, School of International Education, Tianjin Foreign Studies University
Email: 13954871623@163.com
ORCID ID: 0009-0007-5152-6626

Research interests: International Chinese Language Education.

Maximiliano E. Orlando

Independent Researcher
Email: maxorl@outlook.com
ORCID ID: 0009-0006-9351-2719

Maximiliano E. Orlando teaches English and French in Canada and studies Chinese. His research interests include Additional Language Teaching, particularly the application of computer programming in this field.

Xue Li

Eötvös Loránd University, Chinese Department
Email: lixue2024918@gmail.com
ORCID ID: 0009-0002-1138-7303

Xue Li is a PhD candidate in Linguistics at Eötvös Loránd University, Hungary. Her research interests focus on Chinese phonological acquisition and International Chinese Language Education. She has taught Chinese at Cairo University in Egypt and Eötvös Loránd University in Hungary, gaining extensive experience in teaching Chinese as a foreign language. Her current research focuses on phonological acquisition errors among Hungarian learners of Chinese and related pedagogical strategies.

Zhiyong Xie

Professor, School of International Education, Tianjin Foreign Studies University
Email: x-zy2000@163.com
ORCID ID: 0009-0008-9045-8985

Research interests: History of the Chinese language and international Chinese language education.

本期作者

奥兰多·马西米利亚诺·埃德华多 (Orlando Maximiliano Eduardo)，目前在加拿大教授英文和法语，同时也在学习中文。他对语言教学与习得，尤其是编程在语言教学中的应用有浓厚兴趣。

孔茗 (**巴拉日·西吉特希**Balázs Szigethy)，匈牙利罗兰大学语言学博士学院汉学专业博士研究生。主要研究方向为面向成人学习者的对外汉字教学，关注部件教学法、古文字学证据在教学中的应用，以及教学材料与国家语言文字规范的对接。

李雪，匈牙利罗兰大学语言学博士研究生，主要研究方向为汉语语音习得与国际中文教育。曾在埃及开罗大学和匈牙利罗兰大学从事中文教学**工作**，具有较丰富的国际中文教学与实践经验。目前主要关注匈牙利汉语学习者的语音习得偏误及其教学策略。

庞泓雪，天津外国语大学国际教育学院硕士研究生，研究方向为国际中文教育。

解植永，天津外国语大学国际教育学院教授，研究方向为汉语史及汉语国际教育。