Journal of Chinese Language Teaching in Europe 欧洲中文教育 2026/1, pp. 1–43

Zsuzsanna Karsai

(*Eötvös Loránd University, Budapest*)
(Hungarian–Chinese Bilingual Primary and Secondary School)

The vocabulary compatibility of the reading comprehension component of the intermediate-level written Chinese school leaving exam in Hungary with the Chinese Proficiency Grading Standards for International Chinese Language Education

Abstract

In recent decades, Chinese language education has gained increasing significance both globally and within Hungary's language education system. In Hungary, Chinese has been included among the optional foreign languages for the secondary school leaving examination since 2011. From 2015 onward, it has been possible to take the Chinese language exam at both intermediate and advanced levels. This study presents the characteristics of the intermediate-level Chinese school leaving exam based on publicly available official documents and exam descriptions found online. It then analyzes the vocabulary of the reading comprehension component of the intermediate-level written exam in terms of its compatibility with the Chinese Proficiency Grading Standards for International Chinese Language Education, published in 2021. The vocabulary of the tasks from the reading comprehension sections of the exams between 2011 and 2024 was analyzed using the Chinese Text Analyser software. The study presents the proportion of words corresponding to different language proficiency levels in percentage terms, thereby indicating the linguistic difficulty of the exam tasks. The practical value of the study lies in its potential to support not only students preparing for the exam but also future efforts in designing Chinese school leaving exam materials.

Keywords: Chinese language education, Chinese school leaving exam, vocabulary analysis

karsaizsuzsanna@gmail.com

https://doi.org/10.61637/cltcee.2026.1.1
https://orcid.org/0009-0005-8534-2257

如意

匈牙利中学毕业中文考试阅读理解试题的词汇与《国际中文教育中文水平等级标准》的匹配度研究

摘要

近年来，汉语教育在全球范围内以及匈牙利本国的语言教育中变得愈发重要。自2011年起，汉语被列为匈牙利中学毕业考试中可自由选择的现代外语科目之一。自2015年起，考生可以选择在中级和高级两个不同等级的汉语毕业考试。本研究基于公开获取的官方文件和考试说明，介绍匈牙利中级汉语毕业考试的主要特点，并重点分析了笔试中阅读理解部分的词汇，考察其与2021年发布的《国际中文教育中文水平等级标准》所设定各语言等级之间的匹配度。本研究借助 Chinese Text Analyser 软件，对2011年至2024年间的中级毕业考试阅读理解试题的词汇进行了分析。研究通过百分比的形式呈现了各语言等级词汇在文本中所占的比例，从而反映出考试材料的语言难度。本研究具有一定的实践意义，不仅能为备考汉语毕业考试的学生提供参考，同时也可为未来汉语毕业考试命题工作的改进与发展提供借鉴。

关键词：关键词：汉语教学，中学汉语毕业考试，词汇分析

Introduction

Chinese language learning is gaining increasing ground in Hungary. Initially, Chinese language education appeared in higher education, then gradually made its way into public education,[1] becoming one of the modern foreign languages taught in both primary and secondary schools—not only in Budapest, but also in several other cities across the country. The integration of Chinese language learning into public education can be attributed to the work of Confucius Institutes. It has been realized through the efforts of teachers working in both Confucius Institutes and Confucius Classrooms. Currently, there are seven Confucius Institutes operating in six Hungarian cities: Budapest, Szeged, Miskolc, Pécs,

[1] Ye 2021: 135-136.

Debrecen and Győr. Moreover, Chinese language education is also present in public education institutions of nearby towns, either as a second or third foreign language, or in the form of extracurricular classes.[2] In several secondary schools, students now also have the opportunity to take the Chinese school leaving exam.

According to data from the Hungarian Educational Authority, in May 2025, the intermediate-level Chinese school leaving exam was administered at four exam locations, while the advanced-level exam was held at two locations. In this examination period, 18 students took the intermediate-level Chinese exam, and 46 students participated in the advanced-level exam.[3] These numbers show an increase compared to publicly available data from two years earlier: in 2023, 11 students sat for the intermediate-level exam, and 35 for the advanced-level exam.[4] It is important to note that in Hungary, students whose native language is Chinese are also allowed to take the Chinese as a foreign language school leaving exam. This may explain the higher participation rate in the advanced-level exam.

Hungarian native speaker students learning Chinese often ask how they can effectively prepare for the Chinese school leaving exam, as there are currently no publications available on the Hungarian book market specifically designed for this purpose. Although it is commonly understood that the intermediate-level foreign language school leaving exam corresponds to Level B1 of the Common European Framework of Reference (CEFR), the question still frequently arises as to which level of the HSK (*Hanyu Shuiping Kaoshi* – Chinese Proficiency Test) best matches the vocabulary used in the exam. This question forms the basis for the relevance of the present study.

Purpose and Methodology of the Study

The aim of this study is to present the intermediate-level Chinese language school leaving exam in Hungary—its structure and exam requirements—based on publicly available information and documents accessible online. This may serve as a useful resource for students preparing for the exam. Another goal of the study is to provide an initial overview of where the vocabulary used in the reading comprehension tasks of the intermediate-level written Chinese exam generally falls within the levels defined by the Chinese Proficiency Grading

---

2 Karsai 2024: 92–93.

3 https://www.oktatas.hu/kozneveles/erettsegi/2025tavaszi_vizsgaidoszak/erettsegi_info_2025tavasz (last accessed: 17.05.2025)

4 *Ibid.*

Standards for International Chinese Language Education published in 2021, thereby offering an indication of the difficulty level of the exam tasks.

The sample examined in this study consists of every reading comprehension tasks from all intermediate-level written Chinese school leaving exams in Hungary, covering the years 2011 to 2024. These are available on the website of the Hungarian Educational Authority. These tasks include the reading texts and the related comprehension questions. The study focuses specifically on the vocabulary used in these tasks. The vocabulary is analyzed based on the proficiency levels defined in the Chinese Proficiency Grading Standards for International Chinese Language Education, published in 2021. The study presents the percentage distribution of the vocabulary according to these levels, aiming to map the compatibility between the vocabulary used in the reading comprehension tasks of the Chinese school leaving exam and the aforementioned Standards.

Although, due to the characteristics of the Chinese language and the difficulties in defining what constitutes a "word", the length of the reading tasks and the writing tasks in the exam is officially measured in number of characters. A character-based analysis would in fact be simpler to carry out, such an approach would not adequately indicate the approximate vocabulary level of the intermediate-level exam. Therefore, this study is an attempt to analyze the vocabulary, using the Chinese Text Analyser software to review and determine the proportional distribution of vocabulary levels.

The practical significance of this study lies in the method used to classify the vocabulary level of the exam tasks, which may contribute to future research—such as exam analyses, textbook development, or the item-writing process for the Chinese school leaving exam. The study also outlines the limitations of the current analysis, and thus the percentage-based data presented should be interpreted as indicative only, taking these limitations into account and ideally in conjunction with a close reading of the actual exam tasks. Due to these limitations, the present research primarily serves as a preliminary study that may support similar future investigations.

The Relationship Between Vocabulary Coverage and Reading Comprehension

Vocabulary coverage is a key factor in foreign language reading comprehension. Numerous studies have shown that the proportion of unknown words in a text significantly affects the success of comprehension. In this area, research has primarily focused on English-language vocabulary.

Laufer (1989) found that a 95% vocabulary coverage is necessary for adequate reading comprehension. In other words, if a learner knows less than this, their understanding of the text is likely to decline. However, it is not necessary to understand every single word when reading a text—some words can be ignored if they are not essential to grasping the overall meaning. Additionally, there are words whose meanings can be inferred from the context.[5] According to Hu and Nation (2000), a 98% vocabulary coverage is required for learners to understand a literary text without assistance. At 90–95% coverage, only some readers are able to achieve adequate comprehension, while the majority are not. When only 80% of the words in the text were familiar to readers, none achieved adequate comprehension, as the density of unknown words made reading very difficult.[6] Schmitt, Grabe, and Jiang (2011) also examined the relationship between the proportion of known vocabulary in a text and the level of comprehension of that same text. Their findings showed a linear relationship between vocabulary knowledge and reading comprehension, but their study did not identify a clear threshold at which a certain level of vocabulary knowledge leads to a sudden improvement in comprehension. According to their results, if 60% reading comprehension is considered sufficient, a 95% vocabulary coverage is required; if 70% comprehension is deemed necessary, then 98–99% vocabulary coverage is needed.[7] In relation to the Chinese language, Chang (2012) also starts from the widely accepted view that at least 95% vocabulary coverage is necessary for interpreting a written text, while 98% coverage is ideal for independent and fluent reading comprehension. Chang further concludes that knowledge of approximately 2300 words may be sufficient to reach the B1 Level.[8]

The required percentage of vocabulary coverage varies across studies, depending on what level of reading comprehension is considered sufficient. This research is merely an attempt to examine the compatibility between the vocabulary used in the reading comprehension section of the intermediate-level written Chinese school leaving exam and the vocabulary levels defined in the Chinese Proficiency Grading Standards for International Chinese Language Education published in 2021. It does not aim to make definitive claims about the difficulty level of the exam. To accurately determine the difficulty of an exam task set, several additional factors would need to be considered—such as the exam results of the participants (including their native languages, as native Chinese speakers are also allowed to take the exam[9]), the method of assessment,

---

5 Laufer 2013: 867–868.
6 Hu, Nation 2000: 403, 405.
7 Schmitt, Grabe, Jiang 2011: 35.
8 Zhang 2012: 88.
9 However, due to the protection of personal data, exam results are only available in anonymized

and the cut-off scores used in grading. Such an in-depth analysis is beyond the scope of the present study.

Based on the findings of previous research, it can be considered valid for the present study as well that the higher the percentage of words in a text that belong to a specific proficiency level, the more successful reading comprehension is likely to be. As a result, the more accurately an exam can assess whether the examinee has reached the expected language proficiency level required by the exam.

The Chinese Language School Leaving Exam in Hungarian Public Education

In Hungary, after completing the 12th grade of secondary school, students are required to take school leaving exams in five subjects. These include mandatory exams in Hungarian language (for native Hungarian speakers this means Hungarian language and literature; foreign students may take the exam in Hungarian as a foreign language), mathematics, history, one foreign language, and a fifth subject chosen by the student, usually based on their interests or higher education plans. Since 2011, Chinese has been available as an optional subject among the modern foreign languages in the school leaving exam system. Between 2011 and 2014, the Chinese exam was only offered at the intermediate level. From 2015 onwards, it has been available at both intermediate and advanced levels.[10]

In Hungary, school leaving exams are held twice a year: during the autumn exam period (October–November) and the spring exam period (May–June). In addition to the regular school leaving exam, students also have the option to take advanced (early) school leaving exams in individual subjects. This allows them to ease their workload by reducing the number of subjects they need to prepare for at the end of 12th grade. For Chinese language, in order to be eligible for an advanced (early) school leaving exam, students must have final grades for Grades 9 through 12 in the subject. If the student has not yet completed the current school year, they are required to take a preliminary assessment exam covering the material of the ongoing academic year before sitting the school leaving exam.

As with other school leaving exam subjects, the Chinese language exam can be taken at two levels: intermediate and advanced. The school leaving exams in modern foreign languages correspond to levels of the Common European

form on the website of the Hungarian Educational Authority. In cases where the number of candidates is very low, the Authority does not publish the results at all to prevent identification.

[10] https://dari.oktatas.hu/erettsegi.utmutato.index (last accessed: 19.06.2025)

Framework of Reference for Languages (CEFR). According to the Hungarian Educational Authority, the intermediate-level exam corresponds to CEFR Level B1, while the advanced-level exam corresponds to CEFR Level B2.[11] At both levels, the exam consists of a written and an oral component. The date of the written exam is determined centrally and is held on the same day in all secondary schools. The date of the oral exam is set by the institution administering the exam.

In terms of content, the school leaving exam is organized around the following ten topics: personal matters and family; people and society; our environment; school; the world of work; lifestyle; leisure, culture, and entertainment; travel and tourism; science and technology; and the economy. These topics apply to both the written and the oral components of the exam.[12]

The Intermediate-Level Written Chinese School Leaving Exam

The written school leaving exam, at both intermediate and advanced levels, consists of four components in the following order: reading comprehension, grammar and language use, listening comprehension, and writing. Since the present study focuses on the intermediate-level exam, the following section will provide an overview of the parameters of the intermediate-level written Chinese school leaving exam.

Students have 180 minutes to complete the intermediate-level written Chinese school leaving exam. The reading comprehension section is allocated 60 minutes, the grammar and language use section 30 minutes, the listening comprehension section 30 minutes, and the writing section 60 minutes. The grammar and language use section is worth 18 points, while each of the other three components is worth 33 points, making a total of 117 points available for the written exam tasks.

The reading comprehension section consists of three to four tasks, containing a total of 25 to 30 items. The texts must cover different exam topics not only in terms of content but also in the types of items associated with them. These may include gap-filling, true/false, multiple choice, answering questions, etc. The combined length of the tasks must be between 1,000 and 1,200 words[13]. The

---

[11] https://www.oktatas.hu/pub_bin/dload/kozoktatas/erettsegi/vizsgakovetelmenyek2024/elo_id_nyelv_2024_e.pdf (lasd accessed: 13.06.2025)

[12] *Ibid.*

[13] The word count requirements apply uniformly to all intermediate-level matriculation exams in modern foreign languages. However, in the case of the Chinese language, determining the number of words is not straightforward. Therefore, for Chinese, the Hungarian Educational Authority specifies text lengths in terms of the number of characters rather than words.

grammar and language use section also consists of three to four tasks, with the most common task types being gap-filling and error recognition exercises. The tasks in the grammar and language use component must also relate to different exam topics and have a total length of 500–600 words[14], with the accompanying tasks consisting of 20–25 items. In the listening comprehension component, each of the two to three audio texts is followed by one task. The listening texts are approximately 6 to 9 minutes in total length. The tasks include 20–25 items, with true/false and multiple choice being the most common task types. In the writing component, students are generally required to produce two texts, typically in the form of a letter, email, or advertisement. The first text must be 120–150 characters long and is usually guided by three prompts that indicate the required content. The second text should be 150–180 characters long and must be written based on four given prompts.[15]

In the Chinese language, determining the exact number of words is not straightforward; therefore, for the entire exam, text lengths are specified in terms of the number of characters rather than the number of words.[16]

Exam component	Length	Time	Task specifications	Point
Reading comprehension	1000-1200 words	60 mins	25-30 items	33 points
Grammar and language use	500-600 words	30 mins	20-25 items	18 points
Listening comprehension	6-9 minutes	30 mins	20-25 items	33 points
Writing	120-150 characters, 150-180 characters	60 mins	3-4 given prompts	33 points

Table 1: Characteristics of the Components of the Intermediate-Level Written Chinese School Leaving Exam[17]

[14] The word count requirements apply uniformly to all intermediate-level matriculation exams in modern foreign languages. However, in the case of the Chinese language, determining the number of words is not straightforward. Therefore, for Chinese, the Hungarian Educational Authority specifies text lengths in terms of the number of characters rather than words.

[15] https://www.oktatas.hu/pub_bin/dload/kozoktatas/erettsegi/vizsgakovetelmenyek2024/elo_id_nyelv_2024_e.pdf (lasd accessed: 13.06.2025)

[16] https://dload-oktatas.educatio.hu/erettsegi/feladatok_2025tavasz_kozep/k_kinai_25maj_ut.pdf (last accessed: 21.06.2025)

[17] https://www.oktatas.hu/pub_bin/dload/kozoktatas/erettsegi/vizsgakovetelmenyek2024/elo_id_nyelv_2024_e.pdf (lasd accessed: 13.06.2025)

The Intermediate-Level Oral Chinese School Leaving Exam

The Chinese oral school leaving exam is not administered on a centrally determined date; the exam date is set independently by the educational institutions organizing the exam. The oral exam lasts 15 minutes, during which examinees are required to demonstrate their skills through the following three components: conversation, role-play between the examinee and the examiner, individual topic presentation, which includes a picture description. A total of 33 points can be earned in the oral exam.

The oral exam is conducted by drawing exam question, which are typically compiled by teachers at the given institution. Each set of exam questions must include both examiner and examinee versions. If there are fewer than 10 examinees, at least 10 questions must be prepared; if there are more than 10 examinees, at least 20 questions are required. Currently, Hungarian educational institutions are not obligated to publish or disclose the exam questions in advance, so examinees may only encounter the actual content of their selected question during the oral exam itself. According to the school leaving exam regulations effective from 2024, at least 10% of the questions in the oral exam set must be revised or replaced each year.[18]

Assessment of the Intermediate-Level Chinese School Leaving Exam

In the school leaving exam, examinees receive points for both the written and oral components. The total percentage score for the exam is calculated by adding the points from both parts. At the intermediate level, assessment is based on the following scale: Excellent (5): 80%–100%; Good (4): 60%–79%; Satisfactory (3): 40%–59%; Pass (2): 25%–39%; Fail (1): 0%–24%.[19]

Selection of the Standard Used for the Compatibility Analysis

Students learning Chinese often ask how they can effectively prepare for the school leaving exam without targeted preparatory materials available. As a starting point, guidance can primarily be gained from past exam papers, which are available on the Hungarian Educational Authority's website. Another use-

[18] https://www.oktatas.hu/pub_bin/dload/kozoktatas/erettsegi/vizsgakovetelmenyek2024/elo_id_nyelv_2024_e.pdf (lasd accessed: 13.06.2025)

[19] https://njt.hu/jogszabaly/1997-100-20-22 (last accessed: 22.06.2025)

ful indicator is that the intermediate-level school leaving exam corresponds to CEFR Level B1.

In Hungary, it is possible to take the Chinese bilingual language exam offered by the ELTE Origó Language Centre in Budapest, and a few sample tasks are available online.[20] However, this only provides limited practice material, with just a few sets of tasks and no accessible glossary. It is also possible to take the HSK exam in Chinese, which is the most popular Chinese language proficiency test in Hungary. The problem is that the task types in both of these language exam systems differ significantly from those of the Chinese school leaving exam. As a result, neither of them provides sufficient support for exam preparation on its own. Currently, no practice materials or vocabulary lists specifically designed for the Chinese school leaving exam are available.

Although there is a wealth of preparatory materials, textbooks, and vocabulary lists available for the HSK exam—including in online formats—it is more difficult to find clear information on which HSK level's vocabulary can adequately support preparation for the Chinese school leaving exam. Currently, students can primarily rely on the fact that HSK Level 3 was originally aligned with the CEFR B1 Level, which is the expected proficiency for the intermediate-level school leaving exam. One useful resource for exam preparation is the "Chinese Textbook for Hungarians" series published by the ELTE Confucius Institute, which follows the structure of the 2010 version of the HSK exam. The third volume of this series, published in 2013, was designed for the B1 Level, corresponding to HSK Level 3. However, the vocabulary used in the Chinese school leaving exam goes beyond the current HSK Level 3, which may be misleading for examinees.

The first Chinese-language version of the CEFR was published in Taiwan in 2007, followed by a different version released in mainland China at the end of 2008.[21] In 2009, the former Hanban (Office of Chinese Language Council International) published the correspondences between the 2010 version of the HSK exam and the CEFR levels, as follows:[22]

---

[20] https://www.onyc.hu/minta-es-gyakorlofeladatok/ (last accessed: 21.06.2025)

[21] Belassen 2012: 2.

[22] https://web.archive.org/web/20121026095039/http://www.chinesetesting.cn/gonewcontent.do?id=677487 (last accessed: 14.06.2025)

HSK	HSK1	HSK2	HSK3	HSK4	HSK5	HSK6
CEFR	A1	A2	B1	B2	C1	C2

Table 2: Correspondence Between HSK (Hanyu Shuiping Kaoshi) Levels and CEFR (Common European Framework of Reference) Levels According to the 2010 HSK Exam System[23]

However, this correspondence was widely criticized. For example, the Fachverband Chinesisch in Germany issued a statement in 2010 declaring that they did not accept this alignment. They argued that, while other foreign languages typically require 500 to 2000 words to reach CEFR Levels A1 to B1, the new HSK appeared to suggest that only about one-third of that was sufficient for Chinese. Furthermore, according to the former Hanban, the Chinese B2 Level (HSK Level 4) could be reached in 160–320 hours, whereas professional recommendations suggest that at least 1200 to 1600 hours are necessary to attain that level. The Fachverband Chinesisch proposed the following alternative alignment between the HSK and CEFR levels:[24]

HSK	KER
HSK1	---
HSK2	A1 without character knowledge
HSK3	A1-A2
HSK4	A2
HSK5	B1
HSK6	B2-C1

Table 3: Correspondence Between HSK Levels and CEFR Levels According to the 2010 Recommendation of the Fachverband Chinesisch [25]

According to this table, CEFR Level B1, which is the expected level for the intermediate-level school leaving exam, roughly corresponds to HSK Level 5. Therefore, HSK Level 5 preparatory materials and vocabulary lists may be more helpful for preparing for the intermediate-level Chinese school leaving exam. Due to this mismatch in level alignment, exam candidates may justifiably ques-

[23] *Ibid.*
[24] Fachverband Chinesisch 2010: 1-2.
[25] *Ibid.*

tion why the school leaving exam appears more difficult than HSK Level 3, and which HSK level actually provides effective support in preparing for the exam.

Based on this alignment, it would be worthwhile to use the HSK Level 5 vocabulary list for examining the compatibility of the vocabulary used in the school leaving exam. However, the HSK exam system is currently undergoing changes.[26] On July 1, 2021, the "Chinese Proficiency Grading Standards for International Chinese Language Education" (hereinafter referred to as "the Standards") came into effect. This document provides a comprehensive description and evaluation of Chinese language learners' skills and proficiency levels. The Standards divide Chinese language proficiency into three stages and nine levels, and for each level, they define the linguistic elements to be mastered—syllables, characters, vocabulary, and grammar. The Standards serve as an important reference document for institutions involved in Chinese language education. They provide a framework for curriculum development, textbook writing, and language testing across Chinese language education systems worldwide.[27]

Although the current form of the HSK exam has not yet fully aligned with the levels defined in the 2021 Standards, it is expected that the HSK will gradually be restructured according to this framework in the future. Therefore, this study has adopted the levels of the 2021 Standards as the basis for its vocabulary compatibility analysis. In the 2010 HSK system, learners were required to master 2500 words at HSK Level 5. According to the 2021 Standards, this corresponds roughly to a level between Level 3 (2245 words) and Level 4 (3245 words). The new levels outlined in the 2021 Standards are more accurately aligned with the CEFR levels, and as such, Level 3 of the 2021 Standards may serve as a more appropriate reference point in the future for the intermediate-level Chinese school leaving exam.

Selection of the Analysis Software

On the Chinese Test Service Website[28], an online query system linked to the Standards is available. This platform allows users to search for syllables, characters, vocabulary, and grammar to determine which proficiency level a given element belongs to. By clicking on the "Vocabulary" tab, users can access a

---

[26] https://www.abnewswire.com/pressreleases/the-first-exam-of-hsk-level-79-to-be-held-globally-soon_624081.html?utm_source=chatgpt.com (last accessed: 21.06.2025)

[27] http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202103/t20210329_523304.html (last accessed: 14.06.2025)

[28] https://old.chinesetest.cn/standardsAction.do?means=standardInfo (last accessed: 20.06.2025)

search interface where they can paste a text and check which proficiency levels the words in the text correspond to.[29] Each of the levels (1, 2, 3, 4, 5, 6, and 7–9) is represented by a different color, and the system highlights the words in the text using these colors. It then displays a table showing which proficiency levels the words and characters in the text correspond to. However, this interface does not provide percentage data, nor does it indicate the number of words or number of characters in the text. In the table, both characters and words are listed, which often results in duplication of entries. Therefore, while the color-coded text gives a rough visual impression of how much of the vocabulary belongs to a particular level, neither the colored text nor the table provides any percentage-based result.

For this reason, the present study did not use the query interface associated with the Standards to analyze the vocabulary compatibility of the school leaving exam tasks. Instead, it employed a different analysis software capable of providing a more in-depth examination. Numerous online vocabulary analysis tools are available—some even supported by artificial intelligence. In this study, the analysis was conducted using the software Chinese Text Analyser, which is available both online and as a downloadable desktop application, allowing it to be used offline as well. It is a free tool designed for analyzing Chinese texts. An additional advantage of this software for the purposes of the current research is that, even in its free downloadable version, it does not limit the length of the text to be analyzed, is capable of processing texts containing up to 200,000–300,000 characters.[30]

Using the Chinese Text Analyser Software – Procedure

The Chinese Text Analyser software allows users to import custom vocabulary lists for comparison. As a first step, four vocabulary lists were uploaded to the software, since the language level of the school leaving exam cannot be clearly classified under a single specific level of the new Standards. In this study, the four vocabulary lists correspond to the full word lists of Levels 3, 4, 5, and 6 of the Standards (that is the Level 3 list includes words from Levels 1 to 3, the Level 4 list includes words from Levels 1 to 4, the Level 5 list includes words from Levels 1 to 5, the Level 6 list includes words from Levels 1 to 6).

According to the Standards, the total number of words at each level is as follows: Level 3: 2245 words, Level 4: 3245 words, Level 5: 4316 words, Level 6: 5456 words. Some entries in the lists combine both single-character and two-character forms of the same word (e.g., 爸 | 爸爸), and some words appear in

[29] https://hyq.zzchinese.com/play/gjzwdj (last accessed: 20.06.2025)
[30] https://www.chinesetextanalyser.com/ (last accessed: 20.06.2025)

their erhua forms (e.g., 男孩儿). For the purposes of this analysis, the vocabulary lists were uploaded in an expanded form: single-character words also appear in their commonly used two-character versions, and both erhua and non-erhua forms were included (e.g., 聊天 and 聊天儿). As a result, the uploaded word lists contain more entries than the official numbers in the Standards. The total number of imported words was: 2273 words for Level 3, 3278 words for Level 4, 4354 words for Level 5, 5496 words for Level 6.

All reading comprehension tasks from the intermediate-level Chinese language written school-leaving examination papers between 2011 and 2024, were included in the analysis. These are available on the website of the Hungarian Educational Authority. This amounts to a total of 20 examination papers,[31] each containing three or four tasks, depending on the number of characters in the texts. The final sample thus comprised a total of 70 exam tasks, each including both the reading texts and the corresponding comprehension questions. After uploading the vocabulary lists, the reading comprehension texts from the school leaving exam papers were copied along with their related questions—but without numbering or letter labels such as A), B), C), etc. Each text was uploaded individually into the Chinese Text Analyser software, and for each one, the vocabulary lists corresponding to Levels 3, 4, 5, and 6 of the Standards were selected. The software then automatically processed the texts using these word lists.

Once the texts in the sample are opened, the software automatically identifies the boundaries of words and lexical units based on its algorithm—i.e., it segments the text. It then compares the segmented words to the user-uploaded vocabulary list, and highlights in red those words that are not included in the selected list, meaning words that exceed the given proficiency level. In addition, the software counts the number of words and characters, and displays, both in absolute numbers and percentages, how many of the words in the text are known (present in the list) and unknown (not present in the list) according to the selected vocabulary level. The software also has the ability to account for repeated words, and it provides a breakdown of how many of the repeated words are known or unknown within the text.

Manual correction is also possible: by double-clicking on a word in the text, users can manually mark additional words in red as unknown, or remove the red highlight, thereby marking words identified as unknown by the algorithm as known. To minimize possible segmentation errors, a second-round review was conducted. During this step, each word marked as unknown (in red) was individually checked using the query interface linked to the Standards on the

[31] Two shool leaving exams were administered in the following years: 2011 and 2020–2024. One school leaving exam was administered in each of the following years: 2012–2019. This results in a total of 20 exam papers included in the sample.

Chinese Test Service Website, by selecting the "Vocabulary" tab. This step made it possible to correct occasional segmentation mistakes by adjusting the status of certain words as either known or unknown, in order to achieve the most accurate results possible. All 70 tasks were processed using each of the four vocabulary lists (Levels 3, 4, 5, and 6). The data extracted from the Chinese Text Analyser software was then compiled into a summary table. The table includes: the total number of words in the texts (both with and without repetitions), and the number and percentage of words marked as known and unknown for each of the four levels (see Appendix).

Results and Conclusions

The data obtained through the above procedure were compiled into a table (see Appendix). Charts were also created based on the data from Level 3 and Level 4 of the Standards (approximately aligned with CEFR B1), which is considered relevant for the intermediate-level school leaving exam (see Chart 1, 2, 3). The results obtained and the conclusions that can be drawn from them are presented below.

I. The data show that in the first three years of the exam (2011–2014), the total number of words in the samples was often below 1000. From 2015 onward, the word count has consistently remained around 1100–1200, indicating that the length of the reading comprehension tasks (including both texts and related questions) has increased over the years. It is important to note, however, that determining the exact number of words in Chinese is not straightforward. Due to the segmentation settings of the Chinese Text Analyser software and potential errors in identifying word boundaries, the word count represents only an approximate figure.

II. From 2015 onward, not only did the length of the reading comprehension tasks change, but it can also be observed that the proportion of words marked as unknown (after processing the texts with the Level 3 vocabulary list of the Standards) consistently exceeds 10%. This may indicate that the difficulty level of the exam tasks has increased compared to those administered before 2015. Interestingly, it was also in 2015 that the advanced-level school leaving exam became available, which may have contributed to a refinement or adjustment in difficulty in the intermediate-level tasks as well.

III. The data also clearly show that the difficulty level of the exam texts is not uniform. Within a single reading comprehension section of a given exam, the percentage of vocabulary coverage at each level varies across

individual tasks. This suggests that exam papers typically include tasks of varying levels of difficulty. Tasks containing a higher proportion of unknown words are usually accompanied by others with a lower proportion of unfamiliar vocabulary, likely in an effort to balance the overall difficulty of the exam section.

IV. When the exam tasks are processed using the Level 3 vocabulary list of the Standards, the percentage of unknown words averages approximately between 10% and 20%. Using the Level 4 vocabulary list, the percentage of unknown words drops to around 10% on average. It is important to note that, due to the limitations of the study, the percentage values obtained should be interpreted as approximate indicators of text difficulty rather than precise measures of the actual proportion of unknown words. According to Hu and Nation (2000), when learners were familiar with only about 80% of the words in a text (i.e. when approximately 20% of the words were unknown), none of them achieved an adequate level of reading comprehension. Based on this, an unknown word ratio of around 10-20% may influence the difficulty of B1-level reading comprehension tasks and the success of text comprehension. The vocabulary of these exam tasks may be perceived as difficult by students on B1 level; however, the unknown word ratio is not the only factor that determines the level of difficulty of the task.

V. For each exam paper, the average proportion of unknown words was calculated across the tasks. Both the table in the Appendix and Chart 3 reveal a noticeable deviation in the October 2020 exam paper, where the proportion of unknown words is significantly lower, and the ratio of unknown words is closer to the proportions found in the test papers from the years 2011–2014. This difference may have several possible explanations, which could be the subject of future research. One possible explanation is that the vocabulary used in the reading comprehension section of this particular test paper aligns more closely with the vocabulary specified for the levels defined in the Standard. It may also be worthwhile to examine whether the differences in the proportion of unknown words across the test papers are related to the type or topic of the texts. However, such an investigation would require a separate, more detailed analysis and therefore falls beyond the scope of the present study.

It can be concluded that over the years, the length of the texts (within the permitted limits) has increased. The level of difficulty at the intermediate level has also risen following the introduction of the advanced-level school leaving exam. The difficulty of the individual texts within a single exam paper is not consistently homogeneous.

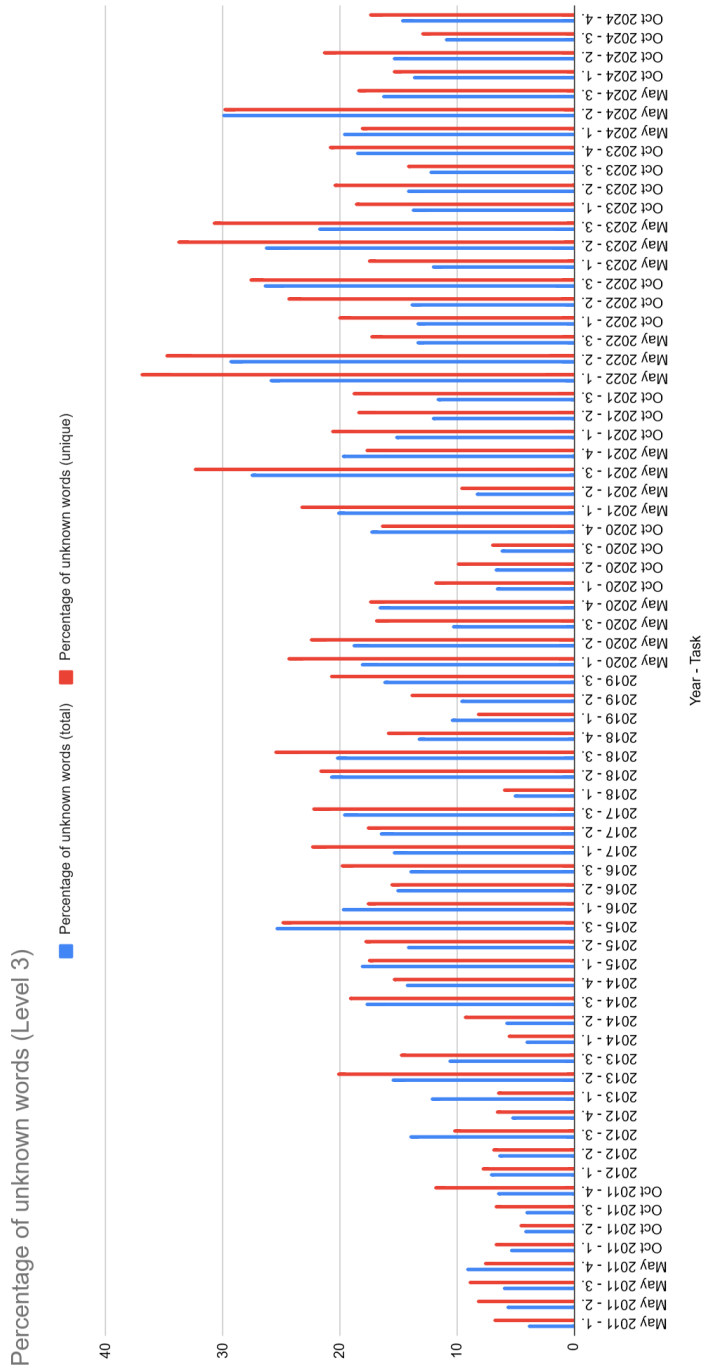

Chart 1. Proportion of Unknown Words in the Reading Comprehension Tasks of the School Leaving Exam at Level 3 of the Standards

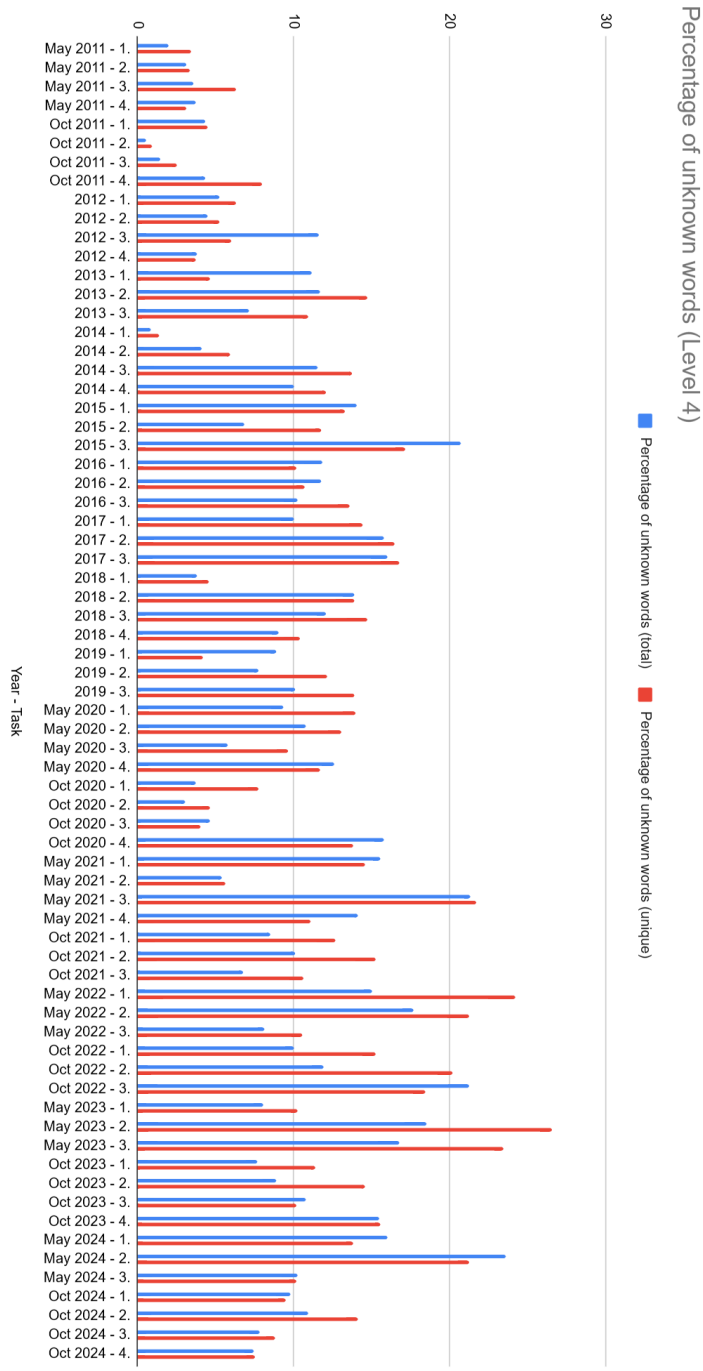

Chart 2. Proportion of Unknown Words in the Reading Comprehension Tasks of the School Leaving Exam at Level 4 of the Standards

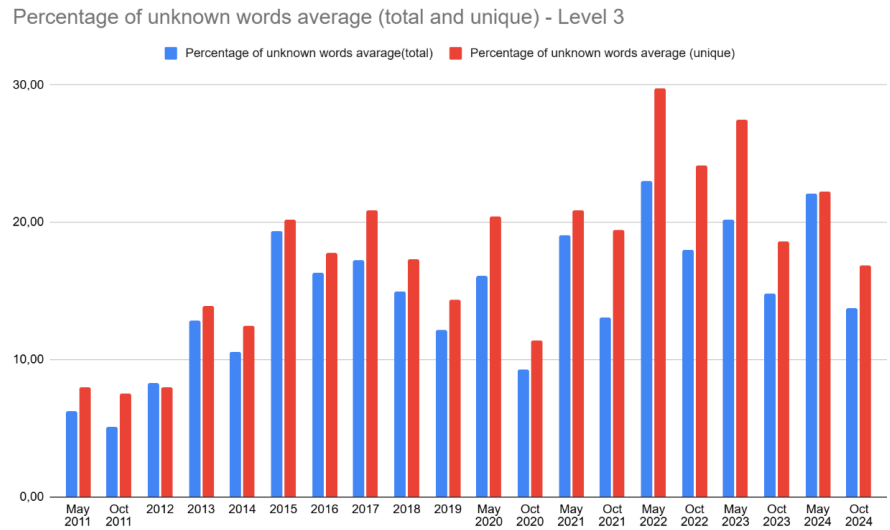

Chart 3. Average Proportion of Unknown Words in the Reading Comprehension Tasks of the School Leaving Exam at Level 3 of the Standards

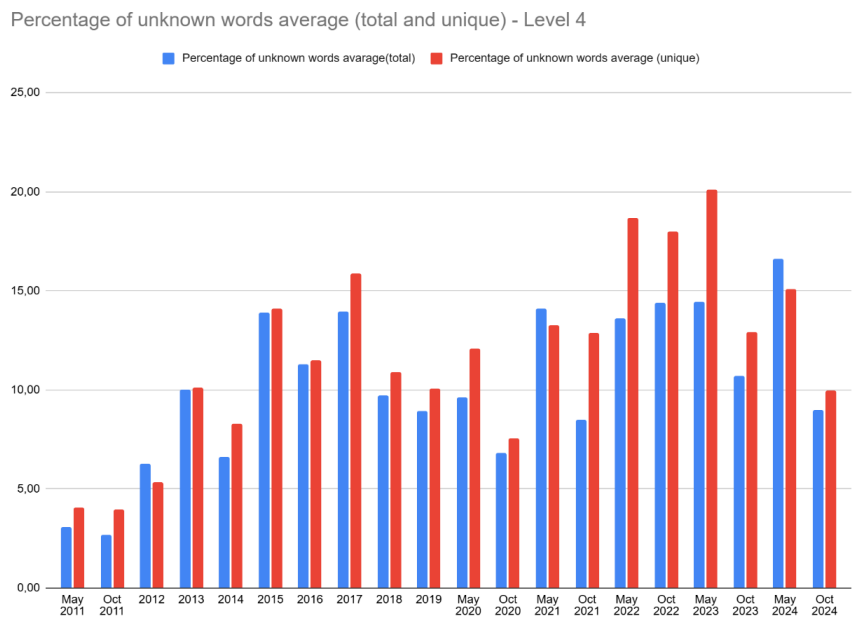

Chart 4. Average Proportion of Unknown Words in the Reading Comprehension Tasks of the School Leaving Exam at Level 4 of the Standards

Limitations of the Study

The present study has several limitations that affect the accuracy of the percentage results.

The vocabulary list of the Standard does not include proper nouns, such as the names of historical figures, nicknames, or geographical names (e.g., country names), whereas such terms do appear in the tasks of the graduation exam. Manually classifying these as known words in some tasks can reduce the percentage of unknown words by as much as 2–5%.

The software also marked certain compound words as unknown in the texts, such as names of Chinese historical periods (e.g., 春秋时期) and expressions related to Chinese culture (e.g., 春联), many of which can be inferred by learners at the B1 language level either from the context or from the composition of the characters. In this study, no character-based analysis of such words was carried out. The ability to infer the meaning of unknown words depends on the individual language skills of the test taker.

The vocabulary of the Standard published in 2021 was not only expanded at each level, but some words were also reassigned from lower to higher levels. For example, certain body parts (e.g., 耳朵 "ear", 鼻子 "nose") were moved up to Level 5, as were certain animals (e.g., 虎 "tiger", 大熊猫 "giant panda"). Some kinship terms only appear at Levels 7–9 (e.g., 外婆 "maternal grandmother", 外公 "maternal grandfather"). However, learners following the 2010 HSK system typically encounter these words relatively early in their studies, meaning they are likely familiar to many test-takers. Nevertheless, under the analysis based on the new 2021 Standard, these words are counted as unknown at Level 3. Therefore, a higher percentage of unknown words does not automatically indicate that the text is significantly more difficult than the expected level. Another example is how color terms are assigned to levels: while most colors (e.g., 红 "red", 红色 "the color red", 黑 "black", 黑色 "the color black") appear at Level 2, the word for "purple" (紫) only appears at Level 5—and only in its single-character form. Thus, when comparing a text with the Level 3 vocabulary list, this word will be marked as unknown, even though in practice, learners typically study color terms together at an early stage, and it would be reasonable to expect that a student at the B1 Level would know the word.

Rare segmentation errors resulting from the software's segmentation settings were manually corrected only in cases where the software marked as unknown certain words that were actually included in the vocabulary list. If the full word appeared in the Standard's vocabulary list at the given level, the word initially marked as unknown was manually reassigned as known after verifying it. However, the accuracy of segmentation for words already recognized by the software

was not reviewed. Such minor inaccuracies may have caused slight deviations in the percentage calculations.

A minor source of inaccuracy in the percentage results may arise from the fact that the software occasionally identifies certain expressions as four-character words that cannot be split. If each pair of characters within the four-character sequence belongs to different levels according to the Standard, it is not possible to label them separately as known or unknown. In this study, any such four-character expression, if partially marked as unknown by the system, was consistently classified as belonging to the higher level. It is important to emphasize that the percentage data obtained from the analysis can only be accurately interpreted in conjunction with reading the actual text.

Summary

This study attempted to review the vocabulary of the reading comprehension tasks (including the texts and corresponding questions) of the written Chinese language school leaving exam in Hungary between 2011 and 2024, and to examine how much the vocabulary overlaps with the levels defined in the Chinese Proficiency Grading Standards for International Chinese Language Education. Since the levels of the Standard are more closely aligned with the CEFR levels, Level 3-4 of the Standard can be regarded as corresponding to the B1 Level. In this study, the data obtained for Level 3 are considered the most relevant, though the analysis also includes the proportion of words appearing at Levels 4, 5, and 6. Words marked as "unknown" even at Level 6 either belong to the Level 7–9 category or are not included in the Standard's vocabulary list.

Although vocabulary analysis in Chinese is challenging due to difficulties in word segmentation (which was also evident during the use of the analysis software), examining vocabulary compatibility is of paramount importance for students, teachers preparing them, and exam designers alike. Due to the anticipated changes in the HSK examination system, the present study investigated the vocabulary compatibility of the tasks with the levels defined in the 2021 Chinese Proficiency Grading Standards for International Chinese Language Education. Running the same test materials through the vocabulary list of the 2010 HSK Level 5 (which, according to many experts, is closer to CEFR B1 than HSK Level 3) would yield different results. However, due to the implementation of the new Standard and the expected future revisions of the HSK system, this study chose the levels of the 2021 Standard as the basis for comparison. The comparison was conducted using the Chinese Text Analyser software.

The results show that, on average, the software identifies between 10% and 20% of words as unknown at Level 3 of the Standard, and using the Level 4 vocabulary list, the percentage of unknown words drops to around 10% on average. It is important to emphasize that the resulting percentages should not be interpreted as a direct indicator of how much more difficult a particular exam task is compared to the expected B1 Level, as the study has several limitations that affect the accuracy of the percentage outcomes.

The percentage results must always be interpreted in conjunction with a close reading of the texts and questions. Words marked as unknown at Level 3 should be examined to determine whether their meaning can be inferred by a test taker at the B1 Level—this includes personal names, culturally specific expressions, and other compound words not listed in the Standard vocabulary. It is also necessary to assess whether the presence of unknown words is relevant to task completion. However, percentages significantly exceeding the 10–20% range may indicate increased task difficulty. If a task contains a higher proportion of unknown vocabulary, it may suggest that test takers perceive the task as more difficult. Again, this must be evaluated in the context of the actual text, which can be the subject of future research.

The method applied in this study may be used in the future to fine-tune the difficulty level of exam texts and to help filter out texts that substantially exceed the expected B1 proficiency level. As such, this method could serve as a useful tool for the exam task development committees involved in designing the Chinese language school leaving exam.

Bibliography

Belassen, Joël 2012. 'Is Chinese Eurocompatible? Is the Common Framework Common? The Common European Framework of Reference for Languages Facing Distant Languages. 'https://www.joelbellassen.com/wp-content/uploads/2019/07/Is-Chinese-eurocompatible-The-CEFR-facing-distant-languages_TOKYO-WolSec-2011.pdf (last accessed: 16.06.2025).

Chang Li-ping 2012. 'The study of the vocabulary size at the CEFR levels for CFL/CSL learners,' *Journal of Chinese Language Teaching*, 9(2): 77–96. [張莉萍2012. '對應於歐洲共同架構的華語詞彙量', 華語文教學研究，9(2): 77–96.]

Chinese Text Analyser 2025. https://www.chinesetextanalyser.com/ (last accessed: 20.06.2025).

Educational Authority 2025. 'Chinese Language. Intermediate-Level Written Examination. Marking and Assessment Guide.' [Oktatási Hivatal 2025. 'Kínai nyelv. Középszintű írásbeli vizsga. Javítási-értékelési útmutató.'] https://dload-oktatas.educatio.hu/erettsegi/feladatok_2025tavasz_kozep/k_kinai_25maj_ut.pdf (last accessed: 21.06.2025).

Educational Authority 2023. 'Information about the School-Leaving Examination.' [Oktatási Hivatal 2023. 'Információk az érettségiről.'] https://www.oktatas.hu/kozneveles/erettsegi/2024tavaszi_vizsgaidoszak0/erettsegi_info_2023tavasz?itemNo=6 (last accessed: 17.05.2025).

Educational Authority 2025. ‘Information about the School-Leaving Examination.’ [Oktatási Hivatal 2025. ‘Információk az érettségiről.’] https://www.oktatas.hu/kozneveles/erettsegi/2025tavaszi_vizsgaidoszak/erettsegi_info_2025tavasz (last accessed: 17.05.2025).

Educational Authority 2024. ‘Modern Foreign Language.’ [Oktatási Hivatal 2024. ‘Élő idegen nyelv.’] https://www.oktatas.hu/pub_bin/dload/kozoktatas/erettsegi/vizsgakovetelmenyek2024/elo_id_nyelv_2024_e.pdf (lasd accessed: 13.06.2025).

Educational Authority 2025. ‘Search among the examination papers and marking guides from previous school-leaving examination periods.’ [Oktatási Hivatal 2025. ‘Keresés a korábbi érettségi vizsgaidőszakok érettségi feladatlapjai, javítás-értékelési útmutatói között.’] https://dari.oktatas.hu/erettsegi.utmutato.index (last accessed: 19.06.2025).

ELTE Origó Language Centre. ‘Sample and Practice Tasks.’ [ELTE Origó Nyelvi Centrum. ‘Minta és gyakorlófeladatok.’] https://www.onyc.hu/minta-es-gyakorlofeladatok/ (last accessed: 21.06.2025).

Fachverband Chinesisch 2010. ‘Statement of the Fachverband Chinesisch e.V. on the New Chinese Proficiency Test (HSK).’ [Fachverband Chinesisch 2010. ‘Erklärung des Fachverbands Chinesisch e.V. zur neuen Chinesischprüfung HSK.’] https://www.fachverband-chinesisch.de/fileadmin/user_upload/Der_Verband/Offizielle_Statements/FaCh_2010_HSK-Erklaerung_de.pdf (last accessed: 15.06.2025).

Hanban 2009. ‘Introduction on New HSK Test.’ https://web.archive.org/web/20121026095039/http://www.chinesetesting.cn/gonewcontent.do?id=677487 (last accessed: 14.06.2025).

Hu, M. – Nation, P. 2000. ‘*Unknown Vocabulary Density and Reading Comprehension,’ Reading in a Foreign Language* 13(1): 403–430. https://doi.org/10.26686/wgtn.12560354

Karsai, Zsuzsanna 2024. ‘An Overview of the Milestones and Achievements in Chinese Language Education of the 20-Year-Old Hungarian–Chinese Bilingual Primary and Secondary School,’ *Journal of Chinese Language Teaching in Europe* 欧洲中文教育 2024(2): 91–115. https://doi.org/10.61637/cltcee.2024.2.6

Laufer, Batia 2013. ‘Lexical Thresholds for Reading Comprehension: What They Are and How They Can Be Used for Teaching Purposes’, *Tesol Quarterly* 47(4): 867–872. https://doi.org/10.1002/tesq.140

Ministry of Education of the People’s Republic of China 2021. ‘Release of the International Chinese Language Education Chinese Proficiency Grading Standards.’ [中华人民共和国教育部 2021. ‘国际中文教育中文水平等级标准》发布.’] http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202103/t20210329_523304.html (last accessed: 14.06.2025).

National Legislation Database 2025. ‘Government Decree 100/1997 (VI. 13.) on the Issuance of the Regulations of the School-Leaving Examination.’ [Nemzeti Jogszabálytár 2025. ‘100/1997. (VI. 13.) Korm. rendelet az érettségi vizsga vizsgaszabályzatának kiadásáról.’] https://njt.hu/jogszabaly/1997-100-20-22 (last accessed: 22.06.2025).

Schmitt, N. – Grabe, W. – Jiang, X. 2011. ‘The percentage of words known in a text and reading comprehension’, *The Modern Language Journal* 95: 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Ye Qiuyue 2021. ‘The History and Current Situation of Chinese Language Education in Hungary’ *Far Eastern Studies* 12 (2020/2): 135–149. [Ye Qiuyue 2021. ‘A magyarországi kínai nyelvoktatás története és Jelenlegi Helyzete’, *Távol-Keleti Tanulmányok* 12 (2020/2): 135–149.] https://doi.org/10.38144/TKT.2020.2.6

Appendix

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
May 2011	1	3	153	147	96,08	6	3,92	87	81	93,1	6	6,9
		4	153	150	98,04	3	1,96	87	84	96,55	3	3,45
		5	153	152	99,35	1	0,65	87	86	98,85	1	1,15
		6	153	152	99,35	1	0,65	87	86	98,85	1	1,15
	2	3	258	243	94,19	15	5,81	120	110	91,67	10	8,33
		4	258	250	96,9	8	3,1	120	116	96,67	4	3,33
		5	258	251	97,29	7	2,71	120	117	97,5	3	2,5
		6	258	254	98,45	4	1,55	120	118	98,33	2	1,67
	3	3	280	263	93,93	17	6,07	111	101	90,99	10	9,01
		4	280	270	96,43	10	3,57	111	104	93,69	7	6,31
		5	280	270	96,43	10	3,57	111	104	93,69	7	6,31
		6	280	271	96,79	9	3,21	111	105	94,59	6	5,41
	4	3	293	266	90,78	27	9,22	129	119	92,25	10	7,75
		4	293	282	96,25	11	3,75	129	125	96,9	4	3,1
		5	293	286	97,61	7	2,39	129	127	98,45	2	1,55
		6	293	287	97,95	6	2,05	129	128	99,22	1	0,78

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
Oct 2011	1	3	183	173	94,54	10	5,46	89	83	93,26	6	6,74
		4	183	175	95,63	8	4,37	89	85	95,51	4	4,49
		5	183	181	98,91	2	1,09	89	88	98,88	1	1,12
		6	183	183	100	0	0	89	89	100	0	0
	2	3	190	182	95,79	8	4,21	107	102	95,33	5	4,67
		4	190	189	99,47	1	0,53	107	106	99,07	1	0,93
		5	190	189	99,47	1	0,53	107	106	99,07	1	0,93
		6	190	190	100	0	0	107	107	100	0	0
	3	3	405	388	95,8	17	4,2	161	150	93,17	11	6,83
		4	405	399	98,52	6	1,48	161	157	97,52	4	2,48
		5	405	400	98,77	5	1,23	161	158	98,14	3	1,86
		6	405	401	99,01	4	0,99	161	160	99,38	1	0,62
	4	3	349	326	93,41	23	6,59	176	155	88,07	21	11,93
		4	349	334	95,7	15	4,3	176	162	92,05	14	7,95
		5	349	345	98,85	4	1,15	176	172	97,73	4	2,27
		6	349	348	99,71	1	0,29	176	175	99,43	1	0,57

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2012	1	3	209	194	92,82	15	7,18	127	117	92,13	10	7,87
		4	209	198	94,74	11	5,26	127	119	93,7	8	6,3
		5	209	202	96,65	7	3,35	127	122	96,06	5	3,94
		6	209	205	98,09	4	1,91	127	125	98,43	2	1,57
	2	3	245	229	93,47	16	6,53	114	106	92,98	8	7,02
		4	245	234	95,51	11	4,49	114	111	97,37	6	5,26
		5	245	234	95,51	11	4,49	114	111	97,37	6	5,26
		6	245	235	95,92	10	4,08	114	112	98,25	2	1,75
	3	3	249	214	85,94	35	14,06	116	104	89,66	12	10,34
		4	249	220	88,35	29	11,65	116	109	93,97	7	6,03
		5	249	226	90,76	23	9,24	116	114	98,28	2	1,72
		6	249	228	91,57	21	8,43	116	115	99,14	1	0,86
	4	3	318	301	94,65	17	5,35	134	125	93,28	9	6,72
		4	318	306	96,23	12	3,77	134	129	96,27	5	3,73
		5	318	307	96,54	11	3,46	134	130	97,01	4	2,99
		6	318	309	97,17	9	2,83	134	131	97,76	3	2,24

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2013	1	3	188	165	87,77	23	12,23	107	100	93,46	7	6,54
		4	188	167	88,83	21	11,17	107	102	95,33	5	4,67
		5	188	169	89,89	19	10,11	107	103	96,26	4	3,74
		6	188	170	90,43	18	9,57	107	103	96,26	4	3,74
	2	3	360	304	84,44	56	15,56	163	130	79,75	33	20,25
		4	360	318	88,33	42	11,67	163	139	85,28	24	14,72
		5	360	331	91,94	29	8,06	163	146	89,57	17	10,43
		6	360	338	93,89	22	6,11	163	151	92,64	12	7,36
	3	3	335	299	89,25	36	10,75	155	132	85,16	23	14,84
		4	335	311	92,84	24	7,16	155	138	89,03	17	10,97
		5	335	318	94,93	17	5,07	155	142	91,61	13	8,39
		6	335	319	95,22	16	4,78	155	143	92,26	12	7,74

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2014	1	3	121	116	95,87	5	4,13	71	67	94,37	4	5,63
		4	121	120	99,17	1	0,83	71	70	98,59	1	1,41
		5	121	121	100	0	0	71	71	100	0	0
		6	121	121	100	0	0	71	71	100	0	0
	2	3	343	323	94,17	20	5,83	169	153	90,53	16	9,47
		4	343	329	95,92	14	4,08	169	159	94,08	10	5,92
		5	343	335	97,67	8	2,33	169	164	97,04	5	2,96
		6	343	339	98,83	4	1,17	169	167	98,82	2	1,18
	3	3	208	171	82,21	37	17,79	109	88	80,73	21	19,27
		4	208	184	88,46	24	11,54	109	94	86,24	15	13,76
		5	208	193	92,79	15	7,21	109	100	91,74	9	8,26
		6	208	194	93,27	14	6,73	109	101	92,66	8	7,34
	4	3	160	137	85,63	23	14,38	58	49	84,48	9	15,52
		4	160	144	90	16	10	58	51	87,93	7	12,07
		5	160	149	93,13	11	6,88	58	52	89,66	6	10,34
		6	160	151	94,38	9	5,63	58	54	93,1	4	6,9

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2015	1	3	504	412	81,75	92	18,25	210	173	82,38	37	17,62
		4	504	433	85,91	71	14,09	210	182	86,67	28	13,33
		5	504	442	87,7	62	12,3	210	190	90,48	20	9,52
		6	504	449	89,09	55	10,91	210	196	93,33	14	6,67
	2	3	468	401	85,68	67	14,32	195	160	82,05	35	17,95
		4	468	436	93,16	32	6,84	195	172	88,21	23	11,79
		5	468	454	97,01	14	2,99	195	183	93,85	12	6,15
		6	468	462	98,72	6	1,28	195	189	96,92	6	3,08
	3	3	251	187	74,5	64	25,5	128	96	75	32	25
		4	251	199	79,28	52	20,72	128	106	82,81	22	17,19
		5	251	206	82,07	45	17,93	128	111	86,72	17	13,28
		6	251	235	93,63	16	6,37	128	117	91,41	11	8,59

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2016	1	3	303	243	80,2	60	19,8	147	121	82,31	26	17,69
		4	303	267	88,12	36	11,88	147	132	89,8	15	10,2
		5	303	280	92,41	23	7,59	147	138	93,88	9	6,12
		6	303	298	98,35	5	1,65	147	143	97,28	4	2,72
	2	3	178	151	84,83	27	15,17	121	102	84,3	19	15,7
		4	178	157	88,2	21	11,8	121	121	100	13	10,74
		5	178	164	92,13	14	7,87	121	111	91,74	10	8,26
		6	178	168	94,38	10	5,62	121	115	95,04	6	4,96
	3	3	575	494	85,91	81	14,09	280	224	80	56	20
		4	575	516	89,74	59	10,26	280	242	86,43	38	13,57
		5	575	526	91,48	49	8,52	280	251	89,64	29	10,36
		6	575	558	97,04	17	2,96	280	265	94,64	15	5,36

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2017	1	3	548	463	84,49	85	15,51	222	172	77,48	50	22,52
		4	548	493	89,96	55	10,04	222	190	85,59	32	14,41
		5	548	508	92,7	40	7,3	222	198	89,19	24	10,81
		6	548	519	94,71	29	5,29	222	204	91,89	18	8,11
	2	3	374	312	83,42	62	16,58	164	135	82,32	29	17,68
		4	374	315	84,22	59	15,78	164	137	83,54	27	16,46
		5	374	324	86,63	50	13,37	164	141	85,98	23	14,02
		6	374	327	87,43	47	12,57	164	144	87,8	20	12,2
	3	3	350	281	80,29	69	19,71	197	153	77,66	44	22,34
		4	350	294	84	56	16	197	164	83,25	33	16,75
		5	350	309	88,29	41	11,71	197	173	87,82	24	12,18
		6	350	314	89,71	36	10,29	197	178	90,36	19	9,64

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2018	1	3	367	348	94,82	19	5,18	198	186	93,94	12	6,06
		4	367	353	96,19	14	3,81	198	189	95,45	9	4,55
		5	367	360	98,09	7	1,91	198	193	97,47	5	2,53
		6	367	361	98,37	6	1,63	198	194	97,98	4	2,02
	2	3	201	159	79,1	42	20,9	115	90	78,26	25	21,74
		4	201	173	86,07	28	13,93	115	99	86,09	16	13,91
		5	201	184	91,54	17	8,46	115	104	90,43	11	9,57
		6	201	189	94,03	12	5,97	115	108	93,91	7	6,09
	3	3	314	250	79,62	64	20,38	156	116	74,36	40	25,64
		4	314	276	87,9	38	12,1	156	133	85,26	23	14,74
		5	314	289	92,04	25	7,96	156	142	91,03	14	8,97
		6	314	293	93,31	21	6,69	156	146	93,59	10	6,41
	4	3	277	240	86,64	37	13,36	144	121	84,03	23	15,97
		4	277	252	90,97	25	9,03	144	129	89,58	15	10,42
		5	277	256	92,42	21	7,58	144	133	92,36	11	7,64
		6	277	258	93,14	19	6,86	144	134	93,06	10	6,94

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
2019	1	3	293	262	89,42	31	10,58	120	110	91,67	10	8,33
		4	293	267	91,13	26	8,87	120	115	95,83	5	4,17
		5	293	267	91,13	26	8,87	120	115	95,83	5	4,17
		6	293	271	92,49	22	7,51	120	116	96,67	4	3,33
	2	3	309	279	90,29	30	9,71	165	142	86,06	23	13,94
		4	309	285	92,23	24	7,77	165	145	87,88	20	12,12
		5	309	290	93,85	19	6,15	165	149	90,3	16	9,7
		6	309	292	94,5	17	5,5	165	151	91,52	14	8,48
	3	3	633	530	83,73	103	16,27	245	194	79,18	51	20,82
		4	633	569	89,89	64	10,11	245	211	86,12	34	13,88
		5	633	590	93,21	43	6,79	245	224	91,43	21	8,57
		6	633	596	94,15	37	5,85	245	227	92,65	18	7,35

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
May 2020	1	3	494	404	81,78	90	18,22	208	157	75,48	51	24,52
		4	494	448	90,69	46	9,31	208	179	86,06	29	13,94
		5	494	455	92,11	39	7,89	208	185	88,94	23	11,06
		6	494	473	95,75	21	4,25	208	195	93,75	13	6,25
	2	3	296	240	81,08	56	18,92	168	130	77,38	38	22,62
		4	296	264	89,19	32	10,81	168	146	86,9	22	13,1
		5	296	273	92,23	23	7,77	168	153	91,07	15	8,93
		6	296	278	93,92	18	6,08	168	156	92,86	12	7,14
	3	3	258	231	89,53	27	10,47	135	112	82,96	23	17,04
		4	258	243	94,19	15	5,81	135	122	90,37	13	9,63
		5	258	248	96,12	10	3,88	135	126	93,33	9	6,67
		6	258	253	98,06	5	1,94	135	130	96,3	5	3,7
	4	3	318	265	83,33	53	16,67	171	141	82,46	30	17,54
		4	318	278	87,42	40	12,58	171	151	88,3	20	11,7
		5	318	284	89,31	34	10,69	171	155	90,64	16	9,36
		6	318	287	90,25	31	9,75	171	157	91,81	14	8,19

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
Oct 2020	1	3	403	376	93,3	27	6,7	142	125	88,03	17	11,97
		4	403	388	96,28	15	3,72	142	131	92,25	11	7,75
		5	403	396	98,26	7	1,74	142	137	96,48	5	3,52
		6	403	399	99,01	4	0,99	142	140	98,59	2	1,41
	2	3	264	246	93,18	18	6,82	130	117	90	13	10
		4	264	256	96,97	8	3,03	130	124	95,38	6	4,62
		5	264	257	97,35	7	2,65	130	125	96,15	5	3,85
		6	264	261	98,86	3	1,14	130	128	98,46	2	1,54
	3	3	174	163	93,68	11	6,32	99	92	92,93	7	7,07
		4	174	166	95,4	8	4,6	99	95	95,96	4	4,04
		5	174	167	95,98	7	4,02	99	96	96,97	3	3,03
		6	174	167	95,98	7	4,02	99	96	96,97	3	3,03
	4	3	259	214	82,63	45	17,37	145	121	83,45	24	16,55
		4	259	218	84,17	41	15,83	145	125	86,21	20	13,79
		5	259	228	88,03	31	11,97	145	133	91,72	12	8,28
		6	259	233	89,96	26	10,04	145	135	93,1	10	6,9

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
May 2021	1	3	296	236	79,73	60	20,27	158	121	76,58	37	23,42
		4	296	250	84,46	46	15,54	158	135	85,44	23	14,56
		5	296	253	85,47	43	14,53	158	138	87,34	20	12,66
		6	296	261	88,18	35	11,82	158	146	92,41	12	7,59
	2	3	392	359	91,58	33	8,42	195	176	90,26	19	9,74
		4	392	371	94,64	21	5,36	195	184	94,36	11	5,64
		5	392	373	95,15	19	4,85	195	186	95,38	9	4,62
		6	392	385	98,21	7	1,79	195	190	97,44	5	2,56
	3	3	300	217	72,33	83	27,67	166	112	67,47	54	32,53
		4	300	236	78,67	64	21,33	166	130	78,31	36	21,69
		5	300	249	83	51	17	166	139	83,73	27	16,27
		6	300	264	88	36	12	166	147	88,55	19	11,45
	4	3	227	182	80,18	45	19,82	135	111	82,22	24	17,78
		4	227	195	85,9	32	14,1	135	120	88,89	15	11,11
		5	227	197	86,78	30	13,22	135	122	90,37	13	9,63
		6	227	202	88,99	25	11,01	135	126	93,33	9	6,67

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
Oct 2021	1	3	588	498	84,69	90	15,31	236	187	79,24	49	20,76
		4	588	538	91,5	50	8,5	236	208	88,14	30	12,71
		5	588	553	94,05	35	5,95	236	216	91,53	20	8,47
		6	588	565	96,09	23	3,91	236	223	94,49	13	5,51
	2	3	247	217	87,85	30	12,15	151	123	81,46	28	18,54
		4	247	222	89,88	25	10,12	151	128	84,77	23	15,23
		5	247	230	93,12	17	6,88	151	135	89,4	16	10,6
		6	247	232	93,93	15	6,07	151	138	91,39	13	8,61
	3	3	486	429	88,27	57	11,73	216	175	81,02	41	18,98
		4	486	453	93,21	33	6,79	216	193	89,35	23	10,65
		5	486	465	95,68	21	4,32	216	202	93,52	14	6,48
		6	486	476	97,94	10	2,06	216	209	96,76	7	3,24

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
May 2022	1	3	399	295	73,93	104	26,07	186	117	62,9	69	37,1
		4	399	339	84,96	60	15,04	186	141	75,81	45	24,19
		5	399	357	89,47	42	10,53	186	153	82,26	33	17,74
		6	399	369	92,48	30	7,52	186	162	87,1	24	12,9
	2	3	435	307	70,57	128	29,43	235	153	65,11	82	34,89
		4	435	358	82,3	77	17,7	235	185	78,72	50	21,28
		5	435	384	88,28	51	11,72	235	199	84,68	36	15,32
		6	435	402	92,41	33	7,59	235	213	90,64	22	9,36
	3	3	393	340	86,51	53	13,49	190	157	82,63	33	17,37
		4	393	361	91,86	32	8,14	190	170	89,47	20	10,53
		5	393	367	93,38	26	6,62	190	175	92,11	15	7,89
		6	393	376	95,67	17	4,33	190	181	95,26	9	4,74

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
Oct 2022	1	3	318	275	86,48	43	13,52	203	162	79,8	41	20,2
		4	318	286	89,94	32	10,06	203	172	84,73	31	15,27
		5	318	293	92,14	25	7,86	203	178	87,68	25	12,32
		6	318	302	94,97	16	5,03	203	186	91,63	17	8,37
	2	3	530	456	86,04	74	13,96	257	194	75,49	63	24,51
		4	530	467	88,11	63	11,89	257	205	79,77	52	20,23
		5	530	485	91,51	45	8,49	257	221	85,99	36	14,01
		6	530	496	93,58	34	6,42	257	229	89,11	28	10,89
	3	3	404	297	73,51	107	26,49	184	133	72,28	51	27,72
		4	404	318	78,71	86	21,29	184	150	81,52	34	18,48
		5	404	341	84,41	63	15,59	184	164	89,13	20	10,87
		6	404	358	88,61	46	11,39	184	170	92,39	14	7,61

			Total					Unique				
				Known words		Unknown words			Known words		Unknown words	
Year	Task number	Levels of the Standard	Total words (100%)	Number (pcs)	Ratio (%)	Number (pcs)	Ratio (%)	Total words (100%)	Number (pcs)	Ratio (%)	Number (pcs)	Ratio (%)
May 2023	1	3	411	361	87,83	50	12,17	215	177	82,33	38	17,67
		4	411	378	91,97	33	8,03	215	193	89,77	22	10,23
		5	411	390	94,89	21	5,11	215	200	93,02	15	6,98
		6	411	394	95,86	17	4,14	215	204	94,88	11	5,12
	2	3	291	214	73,54	77	26,46	177	117	66,1	60	33,9
		4	291	237	81,44	54	18,56	177	130	73,45	47	26,55
		5	291	243	83,51	48	16,49	177	136	76,84	41	23,16
		6	291	264	90,72	27	9,28	177	153	86,44	24	13,56
	3	3	429	335	78,09	94	21,91	230	159	69,13	71	30,87
		4	429	357	83,22	72	16,78	230	176	76,52	54	23,48
		5	429	383	89,28	46	10,72	230	193	83,91	37	16,09
		6	429	393	91,61	36	8,39	230	201	87,39	29	12,61

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
Oct 2023	1	3	353	304	86,12	49	13,88	176	143	81,25	33	18,75
		4	353	326	92,35	27	7,65	176	156	88,64	20	11,36
		5	353	335	94,9	18	5,1	176	161	91,48	15	8,52
		6	353	343	97,17	10	2,83	176	166	94,32	10	5,68
	2	3	337	289	85,76	48	14,24	185	147	79,46	38	20,54
		4	337	307	91,1	30	8,9	185	158	85,41	27	14,59
		5	337	317	94,07	20	5,93	185	168	90,81	17	9,19
		6	337	326	96,74	11	3,26	185	176	95,14	9	4,86
	3	3	194	170	87,63	24	12,37	98	84	85,71	14	14,29
		4	194	173	89,18	21	10,82	98	88	89,8	10	10,2
		5	194	181	93,3	13	6,7	98	94	95,92	4	4,08
		6	194	182	93,81	12	6,19	98	95	96,94	3	3,06
	4	3	284	231	81,34	53	18,66	148	117	79,05	31	20,95
		4	284	240	84,51	44	15,49	148	125	84,46	23	15,54
		5	284	248	87,32	36	12,68	148	132	89,19	16	10,81
		6	284	252	88,73	32	11,27	148	135	91,22	13	8,78

			Total					Unique				
				Known words		Unknown words			Known words		Unknown words	
Year	Task number	Levels of the Standard	Total words (100%)	Number (pcs)	Ratio (%)	Number (pcs)	Ratio (%)	Total words (100%)	Number (pcs)	Ratio (%)	Number (pcs)	Ratio (%)
May 2024	1	3	344	276	80,23	68	19,77	159	130	81,76	29	18,24
		4	344	289	84,01	55	15,99	159	137	86,16	22	13,84
		5	344	299	86,92	45	13,08	159	142	89,31	17	10,69
		6	344	309	89,83	35	10,17	159	147	92,45	12	7,55
	2	3	339	237	69,91	102	30,09	207	145	70,05	62	29,95
		4	339	259	76,4	80	23,6	207	163	78,74	44	21,26
		5	339	276	81,42	63	18,58	207	176	85,02	31	14,98
		6	339	281	82,89	58	17,11	207	179	86,47	28	13,53
	3	3	391	327	83,63	64	16,37	167	136	81,44	31	18,56
		4	391	351	89,77	40	10,23	167	150	89,82	17	10,18
		5	391	357	91,3	34	8,7	167	156	93,41	11	6,59
		6	391	380	97,19	11	2,81	167	160	95,81	7	4,19

			Total					Unique				
				Known words		Unknown words			**Known words**		**Unknown words**	
Year	**Task number**	**Levels of the Standard**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**	**Total words (100%)**	**Number (pcs)**	**Ratio (%)**	**Number (pcs)**	**Ratio (%)**
Oct 2024	1	3	306	264	86,27	42	13,73	148	125	84,46	23	15,54
		4	306	276	90,2	30	9,8	148	134	90,54	14	9,46
		5	306	282	92,16	24	7,84	148	138	93,24	10	6,76
		6	306	292	95,42	14	4,58	148	140	94,59	8	5,41
	2	3	310	262	84,52	48	15,48	177	139	78,53	38	21,47
		4	310	276	89,03	34	10,97	177	152	85,88	25	14,12
		5	310	285	91,94	25	8,06	177	161	90,96	16	9,04
		6	310	293	94,52	17	5,48	177	166	93,79	11	6,21
	3	3	487	433	88,91	54	11,09	192	167	86,98	25	13,02
		4	487	449	92,2	38	7,8	192	175	91,15	17	8,85
		5	487	457	93,84	30	6,16	192	181	94,27	11	5,73
		6	487	458	94,05	29	5,95	192	182	94,79	10	5,21
	4	3	54	46	85,19	8	14,81	40	33	82,5	7	17,5
		4	54	50	92,59	4	7,41	40	37	92,5	3	7,5
		5	54	51	94,44	3	5,56	40	38	95	2	5
		6	54	52	96,3	2	3,7	40	39	97,5	1	2,5