HUMAN, TECHNOLOGIES AND QUALITY OF EDUCATION, 2026
https://doi.org/10.22364/htqe.2026.07

FREQUENCY OF CONSONANT COMBINATIONS IN THE LATVIAN LANGUAGE

Olga Lezina[1], Julija Svitaja[2], Antons Kutuzovs[3]

[1] Independent researcher, Latvia
[2] Riga Technical University, Latvia
[3] KU Leuven, Belgium

ABSTRACT

Text literacy is a fundamental element of quality education at all stages. However, text comprehension is impossible without well-developed reading technique, visual perception, and fluency. Today, children begin learning to read in the preschool period: they first get familiar with letters and then move on to combining letters into syllables. This process takes a considerable amount of time. This raises an important question: is it possible to reduce the time required for initial reading instruction and the development of reading technique in children? In addition to methodological approaches, linguistic strategies may also be applied. For example, it is possible to identify which letter combinations in the Latvian language are most frequently used and to focus on these combinations in early reading. When children encounter these frequently used letter combinations in words, they will be able to recognize them more easily. This can increase reading fluency and, consequently, improve overall text comprehension. The study was carried out using data from the dictionary tezaurs.lv. The research identified the frequency of consecutive consonant blendings consisting of 2, 3, 4, 5, 6, and 7 letters. The findings show a significant gap between highly frequent consonant combinations in Latvian and those that occur less frequently. The results present a basis for further work on developing more effective approaches to teaching children at the stage of reading skill formation.

Keywords: *consonant blendings, consonant combinations, consonant clusters, Latvian, reading fluency, sight comprehension*

Introduction

PIRLS, or Progress in International Reading Literacy Study, serves as the established benchmark for evaluating fourth-grade reading comprehension globally. PIRLS has been conducted every five years since 2001. In 2021, Latvia recorded its lowest result since joining the study (Ozola, 2025), showing a critical decline in early reading proficiency (Figure 1). Also, the Organisation for Economic Co-operation and Development (OECD, 2022), reviewing 2022 results in Latvia, proposes to employ effective learning strategies.

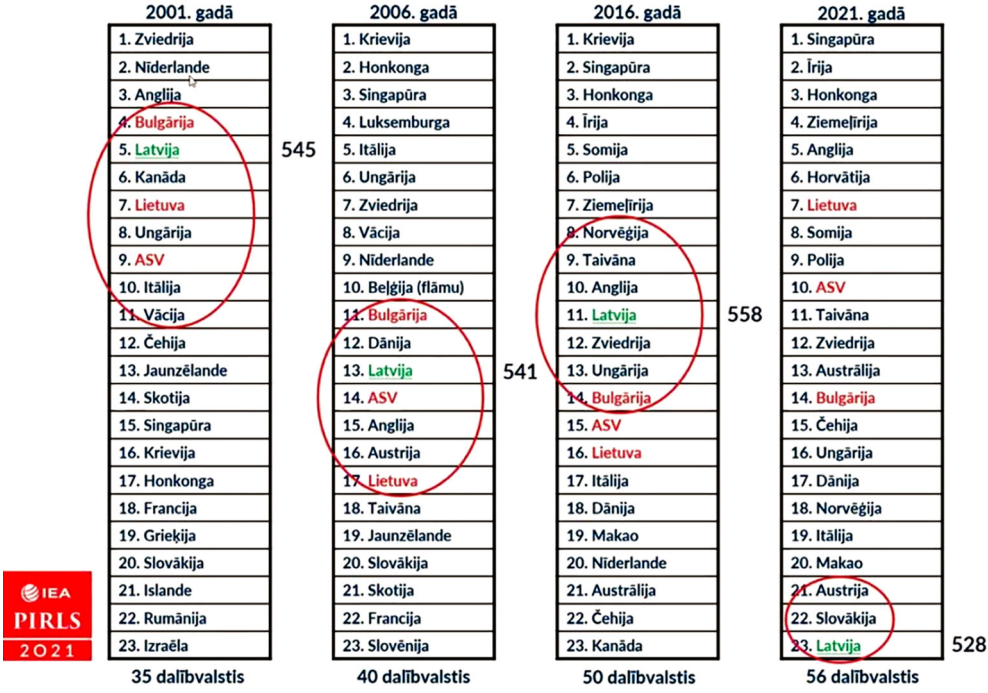


Figure 1. PIRLS assessment results 2025

The Reading Pyramid

The authors of the current study assume that, while reading literacy forms the foundation of academic success, the factors driving the decline, particularly regarding phonemic awareness, decoding speed, and reading fluency, remain insufficiently addressed within national education policy.

Reading instruction for beginner learners in Latvia lacks a systemic, clear, and age-appropriate framework. Regulations No. 747 un No. 716 (MK Noteikumi Nr. 747, 2018; MK Noteikumi Nr. 716, 2018) of the Cabinet of Ministers sets general outcomes, but does not indicate what characterises reading development; it does not also specify the particular methods required to achieve them. As the detailed step-by-step methodology does not exist, it seems to be assumed that once a student can recognize all the letters of the alphabet, they should automatically be capable of reading texts of any complexity level. (Birzniece, 2025)

Cognitive models of literacy development suggest that reading fluency development is based on a hierarchy of sub-skills, frequently visualized through the Reading Pyramid framework (National Reading Panel, 2000) (Figure 2).

1. **Phonemic Awareness:** Ability to identify and manipulate individual sounds in the words (phonemes).
2. **Phonics & Decoding:** Mapping phonemes to graphemes (letters) and letter combinations.

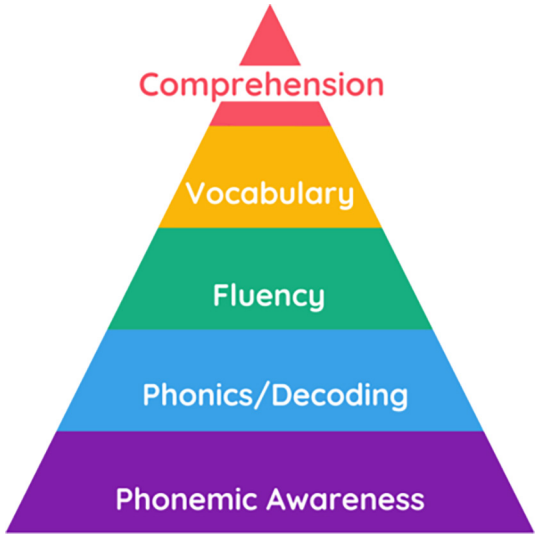


Figure 2. The reading pyramid

3. **Fluency:** Transition from effortful decoding to automatic word recognition.
4. **Vocabulary Acquisition:** Attaching semantic meaning to decoded words.
5. **Comprehension:** Deeply processing and understanding narrative and informational text.

According to Automaticity Theory (LaBerge & Samuels, 1974; Logan, 1997), reading fluency serves as the essential bridge between lower-level mechanical decoding and higher-level comprehension. When lower-level processing becomes automated, cognitive capacity is liberated for meaning understanding. When children remain trapped at letter-by-letter word decoding, their working memory becomes overloaded, which results in slow reading, diminished fluency, and insufficient comprehension (Boneva, 2020).

Syllable reading is an important step in early reading, helping children progress from blending individual sounds to decoding whole words, sentences, and texts. According to methodological guidelines for kindergarten teachers in Latvia, syllable decoding, letter recognition and learning to read whole words are considered essential during the second stage of preschool education (Zariņa, 2024). However, the existing standards and expectations overlook the core concept that children naturally build reading skills over time, first learning a word's meaning before learning to read it. It is assumed that when starting school, every child will, a priori, be able to read and understand any text, whether in a school textbook or in daily life, even though no one has intentionally taught them how to work through complex word structures. By expecting children to develop the whole range of reading sub-skills within preschool education, current educational demands risk fitting a complex, multi-stage developmental process into too short a timeframe, not giving students a chance to develop a great skill of text comprehension.

Thus, the authors emphasize the necessity, expediency, and relevance of a program that incorporates scientifically researched features of the Latvian language in terms of syllables and words (high-frequency words, multisyllable words, compounds, derivatives), as well as sentences of various syntactic structures and texts retaining all units of language thoroughly studied and selected.

The teacher plays a significant role in building reading motivation of their students. This involves the dynamic selection and design of reading materials that reflect the pupils' different developmental levels, sociocultural backgrounds, and perceptual types (auditory, visual, kinesthetic, and tactile). As Zariņa (2024) explains, the role of a teacher is to regularly update the classroom environment with authentic materials and motivate their students to co-create reading materials, elevating child-authored texts to equal status, facilitating students' interest. Therefore, the teacher serves as the link between the institution and the home sphere of their students, actively validating family reading experiences; they empower parents through dedicated workshops and collaborative development of materials. This combined framework of classroom differentiation and parental partnership transforms the educator into a vital architect of early childhood literacy. (Zariņa, 2024)

Acknowledging teachers' crucial role, to save their time and boost teaching productivity, researchers and methodologists must equip them with scientific evidence on the unique characteristics of the Latvian language, including data on the most high-frequency letters and syllable combinations.

The authors of the current study assume that consonant combinations in the Latvian language can be particularly challenging for beginner readers, including children and adults from diverse linguistic backgrounds, such as linguistic minorities, the Latvian diaspora, immigrants, and dyslexic persons.

Consonant blends (e.g., *bl*, *st*, *skr*) can be explicitly taught to help readers rapidly blend two or three distinct consonant sounds while maintaining the individuality of each sound. This skill is essential for decoding unfamiliar words, enhancing reading fluency, and strengthening spelling accuracy by teaching students to perceive and accurately reflect all the sounds present within a word (Reading Universe)

While syllables like 'vowel+consonant' or 'consonant+vowel' are generally easy to select, combinations of two or more consonants represent statistically more variants and frequently cause selection difficulties. Yet, such combinations are quite common in Latvian words.

Decoding relies on recognizing the word's actual sounds separately, but in everyday speech, sounds are articulated together, which is called coarticulation. They merge in speech so quickly that pronouncing them separately and decoding the actual letter combination can be quite a challenge. This often depends on the surrounding consonants or the word's position in the sentence.

In linguistics, this involves processes like assimilation, voicing shifts, sound blending, and elision (Skujiņa et al., 2011). For example, in words such as *nav*, *pats*, *mežs*, and *galds*, the consonants are pronounced very differently from their standard alphabetical sounds. (Ozola, 2021).

All the above mentioned pose problems for primary school students learning to read and write, as it complicates word splitting into individual phonemes. Likewise, it sets an obstacle for students with weak phonemic awareness.

Thus, the current study suggests that the systematic identification and methodological integration of high-frequency consonant combinations and the syllables they form into everyday teaching practices can reduce young learners' cognitive load during learning to read, thereby enhancing reading fluency among early readers.

The present research conducts an applied quantitative analysis of consonant combinations in the Latvian language.

The study aims to identify and analyze the frequency of consonant combinations in the Latvian language to provide empirical data for the development of evidence-based strategies for effective early reading programs.

Methodology: Computational Extraction and Quantification of Consonant Combinations

Given these circumstances, it is clear that since a sequential, structured reading curriculum is lacking, we must equip educators with scientifically grounded data. This allows them to efficiently use their time and resources to prepare preschoolers for fluent reading, deep comprehension, and future academic success.

To design effective learning materials for beginner readers, the next step is to identify which specific consonant combinations are most common in the Latvian language, which ones feature in the largest vocabulary pool, and which syllable patterns can best help children quickly learn to recognize these consonant blends in everyday words.

Blends are consonants that stand next to each other in a word or syllable, and each consonant makes its own sound. The word *griež*, for example, begins with the blend gr. Both the /g/ sound and the /r/ sound are pronounced.

Blends can be made up of two consonants, like gr, or three consonants, like 'str', or more.

Consonant blends, unlike digraphs, do not make a new sound, each consonant retains its usual sound, which students already know. The focus here should be on blending these known sounds together as students read, and separating them as they spell (Reading Universe).

To analyze Latvian consonant combinations, a custom Python script was developed to process the lexical data from the Tezaurs.lv database (tezaurs.lv), which currently contains 413,577 entries. Tēzaurs is a rich and detailed Latvian dictionary that has been growing at the Artificial Intelligence Laboratory (AiLab) at the University of Latvia's Institute of Mathematics and Computer Science since the 1990s. The project is still very much alive today. The entries in Tēzaurs include words, set phrases, and their meanings, which have been carefully collected from dictionaries, encyclopedias, and many other sources – over 340 in total.

The research analyzes the data from this dictionary using a computer programming approach to isolate all consonant combinations in sequences of 2, 3, 4, and even 5–7 consecutive consonants.

In the entries that form the lexical network of the Latvian language, the division of meanings has been significantly restructured based on data from Korpuss.lv.

Definition of 'word'

For each Tezaurs entry, the lemma form was taken as the word: specifically, the text content of the < orth type = "lemma" > element within the entry's < form > element. The entries were not filtered to remove multi-word entries, entries with numbers or abbreviations. The code is written so as to ignore all symbols except the specified consonant letters. That means vowels, numbers, hyphens and other symbols are treated in the same way and are ignored by the code. Hence, entries that contain no letters or only vowels (such as "80286" or "Ā") are ignored completely. Entries with at least one combination of consonants, regardless of whether it contains numbers or other symbols (such as "Abgunst-Gruenfeldt") is used in the count.

Frequency metric

The code measures number of times a specific combination of consonants is met in the Tezaurs database. It didn't count the number of entries where the combination is met, but rather the number of combinations themselves. If a word has two identical combinations, the code counts it two times (e.g. "papildsods" will result in a combination "ds" counted 2 times). It is similar to type frequency, but not exactly it since a single word can be counted several times for the same combination.

The algorithm operated through the following steps.

1. **Consonant Filtering:** A predefined array containing all 24 Latvian consonants (b, c, č, d, f, g, ģ, h, j, k, ķ, l, ļ, m, n, ņ, p, r, s, š, t, v, z, ž) was loaded into the program to serve as the filtering criterion.
2. **Character Scanning:** The script performed a character-by-character scan of each word. When a consonant was detected, it was appended to a temporary variable. This accumulation continued until a vowel or a word boundary was encountered.
3. **Length Verification and Segmentation:** The extracted consonant sequence was then checked against a target length (L). If the sequence was shorter than L, it was deleted. If it was longer, the script automatically split it into all possible sub-combinations of the required length (e.g., if $L = 2$, the cluster *"pkr"* was broken down into *"pk"* and *"kr"*).
4. **Data Aggregation:** Finally, the program updated a global frequency counter for each unique combination and exported the final statistical results into an Excel spreadsheet for further analysis.

For computational study the Python script is provided at Github/Antons-Kutuzovs.

Results

This section (Figure 3) identifies the rarely used consonant blends, which mostly appear in compounds or borrowed words. There is no need to specifically teach this

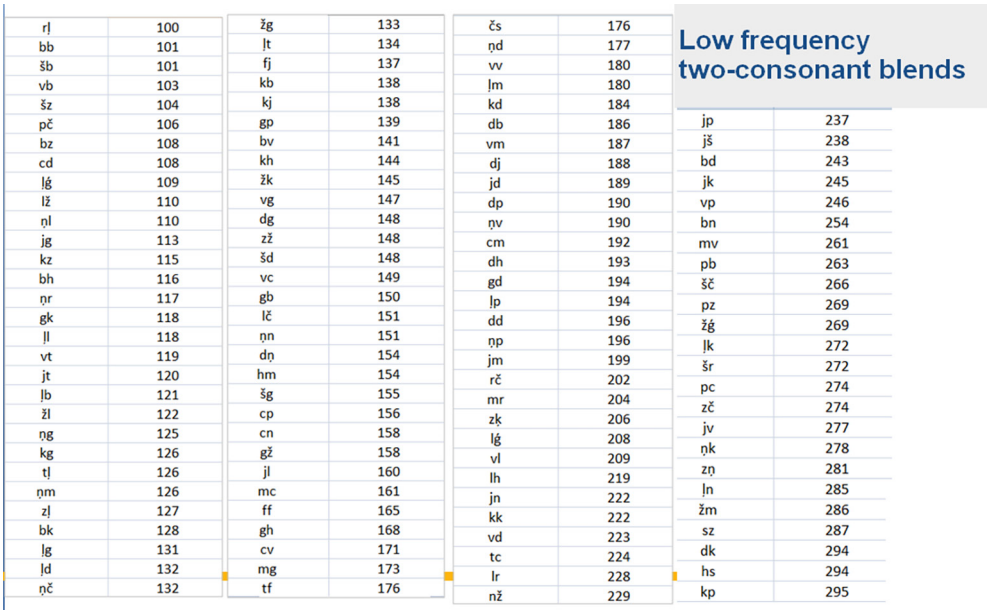


Figure 3. Combinations of two consonants with low frequency of use

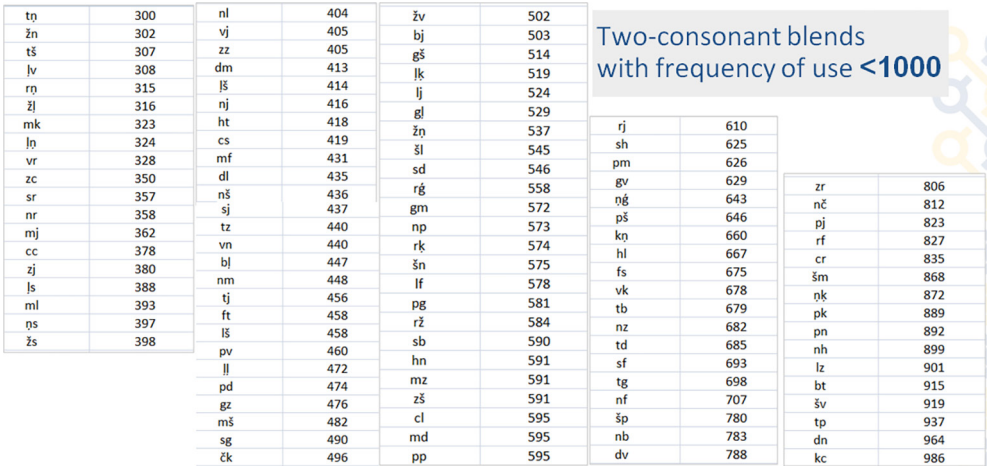


Figure 4. Combinations of two consonants with a frequency of use < 1000

group to children, and words with these consonant combinations should be avoided when choosing text for beginner readers. The script allows exploring the complete list of the corresponding words.

This group (Figure 4**)** comprises the most high-frequency consonant blends in the Latvian language. It is worth noting how sharply their frequency of use increases.

For instance, pairs such as *bļ* (447) and *bj* (503) look and sound similar to a child. A beginner reader may confuse these consonant combinations with one another—both during reading and writing.

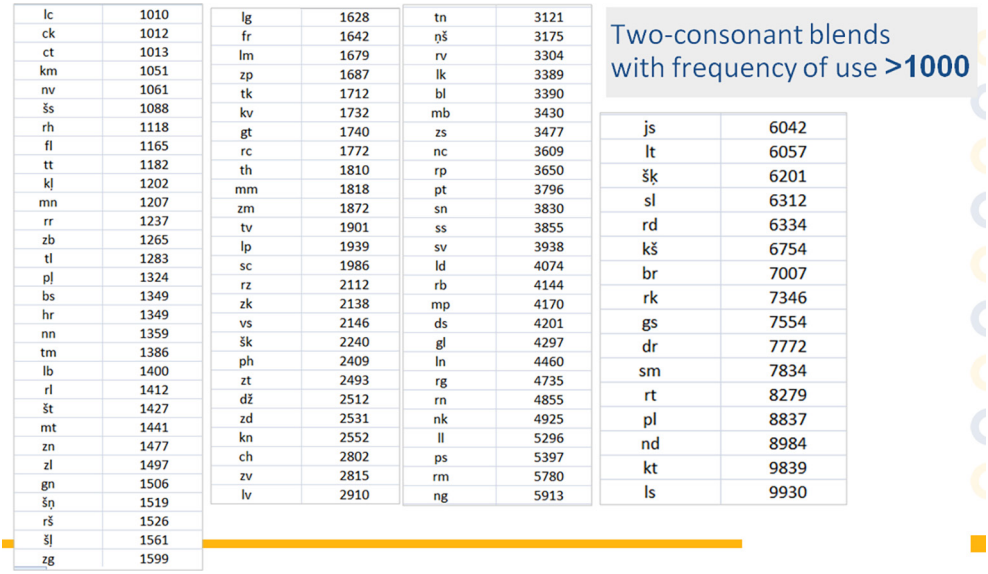


Figure 5. Combinations of two consonants with a frequency of use > 1000

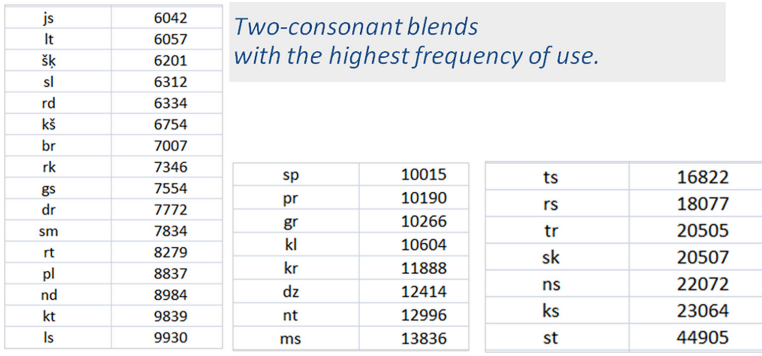


Figure 6. Combinations of two consonants with the highest frequency of use

A remarkably high frequency of use of particular consonant combinations is found (See Figure 5). The difference in frequency between the top and bottom positions in this list is tenfold. This group also includes the combination bl. If bļ and bj are added together to this figure, it totals more than four thousand words in the language that contain these combinations.

This section (Figure 6) aggregates the most frequently used two-consonant combinations in the Latvian language.

These clusters include combinations found in more than 6,000 words in the Latvian language, followed by a group appearing in over 7,000 to 9,000 words, and a separate category spanning 10,000 to 13,000 words. Beyond these, separated by a wide margin, are five consonant blends that are highly prevalent in Latvian vocabulary, occurring in more than 20,000 and one nearly 45,000 instances, respectively.

Three-consonant blends with the highest frequency of use

rkl	514	nsk	1008	skr	2047
ndz	526	lst	1020	sms	2174
lsk	545	trs	1047	sks	3496
gst	570	ndr	1058	sts	3592
ngs	600	pst	1073	str	5526
rts	653	rkš	1163	kst	5667
lks	658	lns	1165		
kls	711	ntr	1388		
rdz	769	ksn	1476		
nst	829	rks	1557		
rds	855	nts	1569		
ktr	917	rst	1752		
kts	994	spr	1820		
		kšķ	1844		

Figure 7. Combinations of three consonants with the highest frequency of use

rksn	104
rkls	110
rsts	120
lksn	124
strs	124
nkts	127
zstr	144
zspr	156
pstr	172
kstr	192
nkšķ	200
nkst	213
nskr	220
lkst	243
lsts	255
nstr	273
rkšķ	461
ksts	531
rkst	1023

The four-consonant blends with the highest frequency of use

Figure 8. Combinations of four consonants with the highest frequency of use

This section (Figure 7**)** features a selection of three-consonant combinations. As can be observed, some combinations occur approximately 500 times, while others hover around the one-thousand mark. However, there are also 'champions' that lead by a wide margin, reaching two thousand and even five thousand instances. These blends are found in high-frequency, authentic Latvian words, like *str*azds, *str*aujs, *str*ādāt, ra*kst*īt, rie*kst*i, tu*kst*ošs, *skri*et *skr*ūve.

Consequently, if a child learns to automatically decode and, conversely, write down these common consonant blends by ear through targeted instruction, their reading speed and writing accuracy will improve significantly.

It might also be of interest to examine the existing four-consonant combinations (Figure 8**)**. Once again, a significant predominance over the other variations is clearly observable.



293	stscr	5
294	stskr	5
295	gstsp	6
296	ktssk	6
297	lkŗsp	6
298	ltspr	6
299	nsstr	6
300	rpsts	6
301	stspr	6
302	nksts	8
303	rkstg	8
304	rkstn	8
305	gstsk	9
306	kstsp	9
307	lgtsp	9
308	rkstm	9
309	gstfr	11
310	dzstr	13
311	lksts	13
312	ktssv	15
313	ksngr	17
314	tzsch	19
315	rksts	48

chtsch	1
gzdstr	1
ksťstr	1
kšķstr	1
kšņstr	1
lkšņst	1
lstsgr	1
lstspr	1
lstssv	1
ltkrps	1
mdcccl	1
mpjstr	1
rdzmpk	1
rftsch	1
rkmpsr	1
rkstsp	1
rkstst	1
schtsc	1
stschr	1
tkrpsr	1
ltschm	2
schsch	2
gstspr	4
ntzsch	4

Blends of 5-6-7 consonants

lkšņstr	1
ltkrpsr	1
schtsch	1

Figure 9. Combinations of 5-6-7- consonants with the highest frequency of use

The number of words containing 5–7 consecutive consonants is negligible and these are most frequently loanwords or proper names: Scilla mi*schtsch*en*k*oana, A*lkšņstr*auts, ha*ntzsch*ia au*gstspr*aigums and others.

The actual count is following: there were found 315 combinations (694 words) of 5 consecutive consonants, 25 combinations (32 words) of 6 consecutive consonants and 3 combinations (3 words) of 7consecutive consonants.

Conclusions

This study addresses the declining reading literacy scores among primary school students in Latvia by offering an evidence-based linguistic framework to optimize early decoding instruction. By performing a computational analysis of lexical data from the Tezaurs.lv database (413,577 entries), the research quantified the distribution and occurrence frequency of consecutive consonant combinations ranging from 2 to 7 letters in the Latvian language.

The conclusions of the study are drawn as follows.

1. The computational analysis reveals a vast quantitative gap between rare consonant combinations and high-frequency clusters. While a large number of multi-consonant clusters occur infrequently (primarily in compound or loan words), a select group of 2-, 3-, and 4-consonant blends occurs in thousands of words across the core Latvian lexicon.
2. Current national teaching methods seem to expect that basic alphabet recognition naturally leads into fluent reading. However, meeting unautomated, complex consonant clusters creates a significant cognitive problem for early readers. Explicitly teaching students to recognize and blend high-frequency consonant combinations

is able to liberate working memory, accelerating the transition from mechanical letter-by-letter decoding to instant sight recognition and reading fluency.

3. As consonant blends undergo natural coarticulation (e.g., assimilation, voicing shifts), the authors assume that targeted training in cluster recognition may sharpen phonemic awareness, improve spelling accuracy, and reduce letter omissions in writing (e.g., distinguishing sounds in words such as *pats*).
4. This research leaves significant room for investigation and the selection of lexical material with regard to its frequency, collocations, morphological composition, and grammatical variations, appropriate for learners of different age and level of proficiency on reading.
5. A more systematic and detailed approach to reading instruction, including teaching high-frequency consonant clusters, will offer a structured tool for students from uneven linguistic backgrounds, including students with dyslexia, children with weak phonemic awareness, non-native speakers, diaspora populations, and adult language learners.

DECLARATION OF CONFLICTING INTERESTS

The authors have no conflicts of interest to declare that are relevant to the content of this article and confirm full responsibility for the article's content.

STATEMENT ON AI TOOL USE

While writing this work, the authors used Grammarly.com to improve readability and polish the English phrasing. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

Birzniece, E. (2025) Lasīšanas grūtību identificēšana un mazināšana 4. klasē – Lasīšanas raitums kā lasītprasmes novērtēšanas kritērijs un tās uzlabošanas instruments. *Latvijas Zinātņu Akadēmijas Vēstis,* 103–117 https://www.lasproceedings.lv/wp-content/uploads/2025/08/7-Eva-Birzniece-1.pdf

Boneva, D. (2020), Lasīšanas un rakstīšanas grūtību identificēšana un atbalsts. *Rēzekne, Latvija* https://ec.europa.eu/programmes/erasmus-plus/project-result-content/968667ce-eb44-4266-8eaa-827b1c1d8f12/Good_Practice_Book_-_LV.pdf

Github/Antons-Kutuzovs: The code used to count combinations of consonants in Latvian language https://github.com/Antons-Kutuzovs/Tezaurs

LaBerge, D. and Samuels, S. J. (1974) Toward a Theory of Automatic Information Processing in Reading. Cognitive Psychology, 6, 293–323. http://dx.doi.org/10.1016/0010-0285(74)90015-2

Logan, G. D. (1997). Automaticity and Reading: Perspectives from the Instance Theory of Automatization. Reading & Writing Quarterly, 13, 123–147. https://doi.org/10.1080/1057356970130203

Ministru kabineta noteikumi Nr. 747 (2018) Noteikumi par valsts pamatizglītības standartu un pamatizglītības programmu paraugiem. https://likumi.lv/ta/id/303768-noteikumi-par-valsts-pamatizglitibas-standartu-un-pamatizglitibas-programmu-paraugiem

Ministru kabineta noteikumi Nr. 716 (2018) Noteikumi par valsts pirmsskolas izglītības vadlīnijām un pirmsskolas izglītības programmu paraugiem. https://likumi.lv/ta/id/303371-noteikumi-par-valsts-pirmsskolas-izglitibasvadlinijam-un-pirmsskolas-izglitibas-programmu-paraugiem

National Reading Panel (2000). Teaching Children to Read. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/documents/report.pdf

OECD PISA 2022 datu bāze https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/latvia_f58bca29-en.html

Ozola, M. (2021). "Protu, protu rakstīt" Vardu pareizrakstība. *Avots*

Ozola, A. (2025). Lasītprasme un tekstpratība sākumskolās Latvijā: situācija un metodiskie ieteikumi. *Latvijas Universitāte/ Izglītības pētniecības institūts.* https://etwinning.lv/wp-content/uploads/2025/09/PIRLS_2021_Twinning.pdf

Reading Universe. (Date or publication is not available). What are consonant blends? https://readinguniverse.org/skill-explainer/phonics-patterns/blends-skill-explainer/overview-what-are-blends

Skujiņa, V., Anspoka, Z., Kalnbērziņa, V., Šalme, A. (2011). Lingvodidaktikas terminu skaidrojošā vārdnīca Latviešu valodas aģentūra, Latviešu valodas institūts. https://valoda.lv/wp-content/uploads/docs/E-Gramatas/08-Vardnica.pdf

Tezaurs.lv https://tezaurs.lv/_/anotacija

Zariņa, S. (2024), Metodiskie ieteikumi pirmsskolas pedagogiem bērnu lasītprasmes veicināšanai https://du.lv/wp-content/uploads/2025/01/Skolotaja_gramata_metodiskie_ieteikumi_05_12_2024.pdf