Readability Criteria for Bulgarian: Lexical and Grammatical Dimensions

Ivelina Stoyanova
Institute for Bulgarian Language
Bulgarian Academy of Sciences
Sofia, Bulgaria
iva@dcl.bas.bg

This paper presents an empirical investigation into text readability for Bulgarian, with a focus on the lexical and grammatical features that predict word-level and sentence-level reading difficulty for primary-school children. We review established readability indices such as Flesch, LIX, Coleman-Liau, and discuss their limitations when applied to a morphologically rich Slavic language such as Bulgarian. We then propose a feature-based approach to readability assessment grounded in the actual reading behaviour of Bulgarian learners.

The empirical work is based on a corpus of finger-tracking reading data from 83 children across grades 2–5, reading age-appropriate narrative texts both aloud and silently. Each word is classified as easy, average, or difficult based on its letters-per-second speed distribution, and we examine the relationship between this difficulty classification and a comprehensive set of phonological, morphological, syntactic, and textual features.

Results identify a small set of strong and weak predictors of readability level. Word length, part-of-speech class (especially the content-vs-function-word distinction) are strong predictors, as well as the sentence-initial position and the position of clause and phrase boundaries, which are consistent across all grades. Repetition within the text reduces the processing cost of difficult content words with subsequent occurrences. Morphological complexity and stem alternation, as well as dependency relations have weak contribution to readability. We report specific findings on pronoun subtypes, on verb-initial and subject-initial sentences, and on POS-bigram syntactic structures.

The analysis also focuses on grade-by-grade comparison, as well as the differences between reading aloud and silent reading.

Keywords: readability, Bulgarian, lexical complexity, morphosyntax

1. Introduction

Readability is a multidimensional property of the written text that determines how easily a given reader can decode and comprehend the text. It depends jointly on properties of the text itself – its phonological, lexical, morphological, and syntactic complexity, and on the characteristics and abilities of the reader, including age, prior exposure to reading, vocabulary knowledge, and the degree to which decoding has become automatised. Good

https://doi.org/10.47810/JCLIB.2.2026.05

reading and text comprehension skills are key competences and an essential prerequisite for high-quality education. Reading fluency is a strong predictor of performance across all literacy-based subjects, with reading speed being one of the most robust indicators of overall reading proficiency. In the long term, students with early reading difficulties face serious challenges in learning, academic performance, and social integration.

Existing readability indices, in turn, were developed primarily for a particular language, and rely on surface measures such as average sentence length and average word length that are highly language-dependent and do not straightforwardly transfer to morphologically rich languages such as Bulgarian, thus the direct application of established readability formulae is not suitable.

Bulgarian has a relatively transparent orthography, with largely predictable grapheme–phoneme correspondences. Transparency facilitates early mastery of reading. However, Bulgarian exhibits a rich inflectional and derivational morphology, postpositional definite article realised as an enclitic, and relatively free word order – properties that influence readability in ways that surface formulae cannot capture.

The present work is part of the efforts to address this gap by developing a Bulgarian-specific, empirically grounded set of readability criteria, based on reading data collected from primary school children from grade 2 to grade 5.

Reading data were collected using the finger-tracking technique implemented in the *ReadLet* infrastructure (Taxitari et al. 2021; Crepaldi et al. 2022; Nadalini et al. 2023), a tablet-based platform that records the continuous movements of a reader's finger across a connected text, together with the voice signal in reading aloud sessions. Finger-tracking provides a fine-grained, temporally sensitive measure of reading dynamics at the word level, validated against eye-tracking in both adult and developing readers (Crepaldi et al. 2022; Nadalini et al. 2023), and is portable and classroom-friendly enough to be deployed at scale in primary schools (Lento et al. 2024; Marzi et al. 2026).

Crucially, by aligning per-word tracking times with linguistic annotations of the reading texts, we are in a position to relate the time a child takes to read each word to the phonological, lexical, morphological, and syntactic properties of that word, and to track how this relationship evolves across grades. The data analysed in the present study were collected within the bilateral project *Assessing reading literacy and comprehension of early graders in Bulgaria and Italy* (2023–2025),[1] jointly funded by the Bulgarian Academy of Sciences and the Italian National Research Council. Reading materials were originally authored in Italian and translated into Bulgarian, with sentence number and sentence type, age-appropriate vocabulary, and culturally neutral narrative content in similar proportions across the two languages (Pirrelli and Koeva 2024). Children read one text aloud and another text silently, on a tablet running the ReadLet application, in their regular classroom setting. The two texts are in the same age category. The present paper focuses exclusively on the Bulgarian data, with a view to identifying the text-internal features that determine word-, phrase-, and sentence-level reading difficulty and to laying the groundwork for a language-specific readability evaluation.

1 https://dcl.bas.bg/en/literacy/

The present study focuses on decoding speed as measured by per-word reading times, on the assumption that fluent decoding is a necessary component of, and a reliable proxy for, overall reading proficiency in early primary grades. Comprehension questions administered after each reading episode were recorded but fall outside the scope of the present paper. In Section 8 we briefly discuss the ways in which fast decoding can dissociate from full understanding of the text.

The objective of the study is to use reading speed data to derive a set of evidence-based readability criteria for Bulgarian texts. While established readability formulae such as Flesch, LIX, and Coleman–Liau, operate on a small number of surface variables and yield a single text-level score, the present approach grounds readability in observed word-level and sentence-level reading patterns and exploits a substantially richer feature inventory that includes phonological, morphological, syntactic, and textual properties. Identifying which of these features contribute robustly across all grades, and which are diagnostic only for beginner readers, is a necessary step towards a complex, developmentally sensitive readability index for Bulgarian. The design and validation of such an index is left to future work.

Building on the empirical data, we address the following research questions:

RQ1. Which phonological, lexical, morphological, and syntactic features of Bulgarian most strongly predict word-level and sentence-level reading speed in primary school children?

RQ2. Which features remain stable in importance as reading skills develop, and which lose (or gain) weight as decoding becomes increasingly automatised?

RQ3. Do the same features predict reading speed in reading aloud and in silent reading, or do the two modalities differ, and is that consistent across grades? Do phonological and articulatory features weigh more heavily in reading aloud, and are features equally pronounced?

The remainder of the paper is organised as follows. Section 2 reviews related work on readability assessment and on empirical finger-tracking studies of reading development, with Section 2.3 presenting the established readability assessment frameworks and their limitations for Bulgarian. Section 3 describes the dataset and the automatic linguistic analysis pipeline applied to the reading texts. Section 4 discusses the word-level and sentence-level feature inventory, including length of the word in number of letters and number of syllables, part of speech, with focus on function vs. content words, morphological complexity, syntactic structure, position of the word in the sentence. Section 5 presents the key findings that are directly relevant to the development of a readability index for Bulgarian. Section 6 draws practical recommendations for reading instruction and text design, Section 7 concludes and outlines directions for future work, and Section 8 briefly outlines the limitations of the work presented in the paper.

2. Related Works

Research relevant to a language-specific readability evaluation for Bulgarian spans across several directions. The first concerns the empirical methods used to observe reading be-

haviour, from eye-tracking studies to finger-tracking patterns among primary-school children. The literature also focuses on the linguistic features that most reliably modulate reading speed and comprehension – word length and frequency, morphological structure, part-of-speech class, and syntactic complexity. The developmental trajectory along which each of these features gains or loses weight with age and skills is relevant to establishing reliable criteria for readability. This section aims to outline these directions in the literature and to identify the limitations of established readability indices when applied to Bulgarian.

2.1 Works on Assessment of Reading Ability

Most work on the temporal dynamics of reading has relied on eye-tracking, which has established robust effects of word length, word frequency, and predictability on fixation durations and patterns (Rayner 1998, 2009). Eye-tracking, however, requires laboratory equipment and is difficult to deploy at the scale needed for developmental and educational research. Finger-tracking has recently emerged as a portable, classroom-friendly alternative that records the continuous movements of a reader's index finger across a connected text on a tablet screen. The technique has been validated against eye-tracking in both adult and developing readers, with finger movements showing strong correlations with fixation durations, saccade patterns, and articulation timing (Crepaldi et al. 2022; Nadalini et al. 2023). The data analysed in the present paper were collected using the *ReadLet* infrastructure (Taxitari et al. 2021; Ferro et al. 2018), a tablet-based platform purpose-built for large-scale, multimodal reading data collection in primary school settings.

Reading acquisition is a multidimensional developmental process that depends not only on the ability to decode written symbols into sounds, but also on efficient lexical access, working memory, attentional control, and the gradual automatisation of orthographic patterns (Ehri 2005; Share 2008; Perfetti 2007). Automatic word recognition is considered a prerequisite for fluent reading: as decoding becomes effortless, cognitive resources are freed for comprehension (Breznitz 2006).

One of the clearest behavioural markers of this developmental shift is considered to be the divergence between reading aloud and silent reading. In early grades, the two modalities are closely aligned, as silent reading still engages phonological processing and covert, subvocal articulation (LaBerge and Samuels 1974; Perfetti 1985; Wright, Sherman, and Jones 2004). With increasing proficiency, silent reading becomes faster than reading aloud and progressively less dependent on phonological decoding, reflecting the acquisition of larger orthographic units and more efficient lexical access (Brysbaert 2019; Rayner 2009). This divergence has been documented in English and Italian (Zoccolotti et al. 2009; Kim, Wagner, and Foster 2011) and in cross-linguistic finger-tracking work on Bulgarian and Italian primary school children (Lento et al. 2024; Marzi et al. 2026).

2.2 Studies on Aspects of Readability

The most extensively documented predictors of reading speed are word length and word frequency, both of which interact with orthographic transparency and with reading proficiency. Relatively transparent orthographies such as Bulgarian, with highly consistent

grapheme–phoneme correspondences and rules for exceptions, support early mastery of decoding (Schüppert et al. 2017; Seymour, Aro, and Erskine 2003; Share 2008), and have been argued to allow learners to rely on small, consistent grain sizes from the onset of literacy (Ziegler and Goswami 2005; Ziegler et al. 2010). Within this setting, the effect of word length is strongest in early grades and progressively diminishes with age, as serial decoding gives way to more holistic recognition (De Luca et al. 2008; Joseph et al. 2009; Zoccolotti et al. 2009; Marzi et al. 2020).

Word frequency shows the opposite trajectory: its facilitatory effect sharpens with experience, with high-frequency items accessed increasingly efficiently while low-frequency items remain costly (Lento et al. 2024; Marzi et al. 2026). These two effects are robust across measures of both reading time and reading speed and provide the empirical baseline against which any feature-based readability model must be evaluated.

Beyond length and frequency, morphological structure also plays a role in word recognition. Developing readers gradually learn to exploit morphological regularities to anticipate upcoming segments and reduce processing load (Beyersmann and Grainger 2018; Beyersmann et al. 2015; Carlisle 2000). Morphological decomposition effects have been observed across Spanish (Casalis et al. 2009), French (Beyersmann et al. 2015), German (Hasenäcker and Schroeder 2017), Italian (Burani 2010; Burani et al. 2018; Traficante et al. 2011), among others, suggesting a robust role for morphology in reading acquisition across languages. These effects become particularly visible in silent reading, where articulatory constraints no longer impose a strict left-to-right pace (Marzi et al. 2026). For Bulgarian, with its rich inflectional and derivational morphology and its postpositional definiteness marker, morphological cues are expected to be especially informative.

A further distinction that has emerged is the contrast between content and function words. Function words, being shorter, more frequent, and less semantically demanding, tend to reach automaticity earlier in development, whereas content words remain more sensitive to lexical variables such as length and frequency (Marzi et al. 2026). However, as function words play a distinct role in syntactic segmentation and express syntactic relations, their function also plays a role in reading speed.

These observations only emphasise the need to build a complex readability evaluation approach that incorporates all major features that determine the level of difficulty of a text relative to an age group or reading skills level.

2.3 Readability Assessment Frameworks

A number of indices for measuring text readability have been proposed in the literature, typically based on surface characteristics such as sentence and word length. Although they were developed independently and calibrated against different reference populations, these indices share a common assumption: that the difficulty of a text can be captured by a small number of easily computable variables – chiefly the average length of its sentences and the average length of its words. They differ mainly in how these variables are combined and in the scale on which the resulting score is reported.

The **Flesch Reading Ease Index** (FLI) (Flesch 1948), originally developed for English, assigns a score from 100 (very easy) to 0 (very difficult). It combines average sentence length, expressed as the number of words per sentence, with average word length, expressed as the number of syllables per word, and weights the latter heavily on the assumption that polysyllabic vocabulary is the dominant source of reading difficulty:

$$\text{FLI} = 206.835 - 1.015 \times (\text{avg. sentence length}) - 84.6 \times (\text{avg. word length}) \quad (1)$$

The two large multiplicative constants and the intercept of 206.835 were derived empirically by fitting the formula to reading comprehension data from school-age readers of English. A score above 90 corresponds to text accessible to a fifth-grade reader, whereas scores below 30 indicate text requiring college-level education.

The **LIX index** (Läsbarhetsindex) (Björnsson 1968) was developed for Swedish and is widely used across European languages. It ranges, in practice, from roughly 20 (easy) to 60 (difficult) and combines average sentence length with the percentage of long words, where long words are conventionally defined as those with more than six characters:

$$\text{LIX} = \frac{\#\text{ words}}{\#\text{ sentences}} + \frac{\#\text{ long words} \times 100}{\#\text{ words}} \quad (2)$$

By thresholding word length at six characters rather than averaging it, LIX is somewhat less sensitive to the presence of a few very long words and is in principle more portable across languages with different average word lengths than indices that rely on syllable counts.

The **Coleman–Liau Index** (CLI) (Coleman and Liau 1975) avoids syllable counting altogether and operates entirely on letter and sentence counts, which makes it particularly well-suited to automatic computation:

$$\text{CLI} = 0.0588 \times \frac{\#\text{ letters}}{\#\text{ words}} \times 100 - 0.296 \times \frac{\#\text{ sentences}}{\#\text{ words}} \times 100 - 15.8 \quad (3)$$

The resulting value is interpreted as the approximate number of years of formal education required to understand the text: a CLI of 8 corresponds to eighth-grade level, a CLI of 12 to the end of secondary school, and so on.

The dominant computational approach to text readability assessment treats reading difficulty as a classification problem driven by a feature inventory organised into four groups: text, lexical, morphological, and syntactic features (Dell'Orletta, Montemagni, and Venturi 2011). This framework was originally developed for Italian and has since served as a template for readability work in other languages, including Bulgarian (Pirrelli and Koeva 2024).

Text characteristics include sentence length, calculated as the average number of words per sentence, and word length, calculated as the average number of letters per word. Lexical characteristics comprise the percentage of unique words relative to an age-appropriate reference list, the distribution between basic vocabulary, high-frequency words, and relatively low-frequency words, and the number of occurrences of different words in the text. Morphosyntactic features refer to lexical density, defined as the ratio of content words

(verbs, nouns, adjectives, and adverbs) to the total number of words. Syntactic features include the depth of the syntactic trees, the overtness and transparency of the components of the clause, the average depth of the syntactic tree of subordinate clauses, the word order of subordinate clauses, and the length of dependent clauses.

2.4 Limitations for Bulgarian

Although there is general agreement between the assessments provided by different readability indices, there is no complete overlap. Critically, the indices were created and tested primarily for English and Swedish, and require adaptation for Bulgarian and other Slavic languages (for Russian, formulae with adjusted constants have been proposed).

More fundamentally, the assessment of texts using these measures is incomplete, as it does not account for the lexical, morphological, and syntactic features of Bulgarian words, nor for the semantic, selective, and grammatical restrictions arising from Bulgarian word order, lexical combinations, and syntactic constructions.

As an illustration, the excerpt from the short story *Lesson in Japanese* by Bulgarian author Viktoriya Beshliyska (Table 1) was assigned $FLI = 19.1$ (very difficult; college level), $LIX = 39.5$ (medium to difficult; approx. 9th grade), and $CLI = 10.1$ (difficult; approx. 10th/11th grade). While these results reflect the general difficulty of the text, they cannot precisely classify it into an age-related or difficulty-based category, and do not reflect the specific lexical, morphological, and syntactic sources of its difficulty. In the example, we notice a concentration of longer words such as *японското* 'Japanese', *семейство* 'family', *приближава* 'approaches' (verb), *вцепенявам* 'freeze' (verb), *качулката* 'hood', but they are part of the general vocabulary and do not pose a difficulty to fluent readers at the end of primary school.

Урок по японски Виктория Бешлийска	**Lesson in Japanese** Viktoriya Beshliyska
Сутринта отново виждаме японското семейство. Към нас приближава мъжът и аз се вцепенявам. Когато очите му срещат моите, той повдига любопитно вежди. Вероятно очаква да отвърна на поздрава му. Бързо нахлупвам качулката си. Тогава мъжът се усмихва и сякаш казва: „Не се притеснявай, разбирам те!“. В този миг се случва нещо странно: от устата ми започват да се нижат думи на японски.	In the morning, we see the Japanese family again. The man approaches us, and I freeze. When his eyes meet mine, he raises his eyebrows curiously. He's probably expecting me to return his greeting. I quickly pull my hood over my head. Then the man smiles, as if to say, "Don't worry, I understand!" At that moment, something strange happens: words in Japanese begin to flow out of my mouth.

Table 1
Short story excerpt demonstrating various readability indices.

3. Dataset and Automatic Preprocessing

Empirical reading data were collected with the *ReadLet* tablet-based platform (Taxitari et al. 2021; Ferro et al. 2018), which records the continuous movements of a reader's index finger across a connected text together with the voice signal in reading aloud sessions. Eighty-three Bulgarian children from grades 2 to 5 at a primary school in Sofia took part (Marzi et al. 2026), with no selection bias for reading skills, all native speakers of Bulgarian and with no documented learning-related special needs. Each child read two age-appropriate narrative texts on a tablet – one aloud and another one silently of similar length and complexity appropriate for the age, and answered two comprehension questions following each episode of the text. Texts ranged from approximately 280 words at grade 2 to approximately 700 words at grade 5, matched for narrative complexity to the age of the target readers (Pirrelli and Koeva 2024). Responses to the comprehension questions were recorded but are not analysed in the present paper, which focuses exclusively on decoding-speed measures; the analysis of comprehension data is left to future work.

All reading data are anonymised at the point of collection: each session is identified by a numeric code, with no personally identifying information retained beyond grade level, age in months, and gender. Analyses in the present study are conducted exclusively on aggregated data, grouped by grade and reading modality, with no individual reader being identifiable in any of the reported results. The dataset records the duration of the finger tracing within the token's bounding box, the reading modality (reading aloud or silent reading), and the resulting per-token reading speed, computed in both letters per second and syllables per second.

Table 2 summarises the distribution of the Bulgarian finger-tracking dataset across the eight groups categorised by grade (2 – 5) and modality (aloud / silent), comprising a total of 248 reading sessions from 83 children and a total of 128,386 tracked words.

Grade	Modality	Sessions	Children	Words read	Words/session
2	aloud	23	21	6,675	290
2	silent	21	21	6,088	290
3	aloud	38	38	16,970	447
3	silent	38	38	16,698	439
4	aloud	32	32	18,928	592
4	silent	32	32	18,980	593
5	aloud	32	31	21,750	680
5	silent	32	31	22,297	697
Total	**aloud**	**125**	–	**64,323**	–
Total	**silent**	**123**	–	**64,063**	–
Total		**248**	**83**	**128,386**	–

Table 2
Distribution of the Bulgarian finger-tracking dataset across the eight groups (grade × modality).

Preliminary automatic processing of the texts involves annotation at the following linguistic levels: sentence segmentation; tokenisation; part-of-speech (POS) tagging; lemmatisation; and sentence structure identification using a Universal Dependencies (UD) model. The analysis was performed using the Bulgarian multi-component system for processing and linguistic annotation (Koeva, Obreshkov, and Yalamov 2020).

The annotation supports statistical investigation of readability features across different linguistic levels:

- phonological and general lexical complexity (word length and lexical diversity);
- coverage of different parts of speech and their word forms;
- lexical typicality (statistical frequency of usage and distribution of words);
- morphological complexity (composition of words);
- syntactic functions of words in the sentence;
- word order positions and word combinations.

4. Analysis of Readability Features

The objective of this study is to investigate the features at phonological, lexical, morphological and syntactic level, that contribute to the classification of words and larger language units (phrases, sentences) according to their reading difficulty.

Our current specific tasks are related to the research questions set in the Introduction in Section 1. Our future work aims at developing a comprehensive system of criteria for assessing reading difficulty which can be formalised and implemented to perform automatic text classification based on readability.

Difficulty measure. Reading difficulty was operationalised as the average reading speed for a word, calculated as the number of letters read per second. Words are classified as: *difficult*/slow to read (read at a slow speed), *average* (read at the most common speed), and *easy*/fast to read (read at a high speed). The class boundaries were determined from the distribution of reading speeds: the speed recorded for the majority of words (60%) is considered normal; the slowest 20% constitute the difficult class, and the fastest 20% the easy class.

The following features were defined at the word-level:

1. length in number of letters and number of syllables;
2. ratio of consonants to total number of letters and number of clusters of 3 or more consonants;
3. part of speech;
4. number of prefixes and suffixes, as well as the definite article;
5. phonological irregularities in wordforms;

6. number of occurrences in the text so far;
7. syntactic function;
8. position of the word in the sentence.

4.1 Average Reading Speed Thresholds by Grade and Modality

For each grade × modality, the 20th and 80th percentiles of per-token reading speed (in letters per second) were taken as the boundaries between the three difficulty classes: tokens read at a speed below the 20th percentile are labelled *difficult*, those between the 20th and 80th percentiles – *average*, and those above the 80th percentile – *easy*. The resulting thresholds are reported in Table 3 and Figure 1.

Grade	Modality	Difficult (slow)	Average	Easy (fast)
2	aloud	< 5.49	5.49–17.09	≥ 17.09
2	silent	< 5.84	5.84–17.61	≥ 17.61
3	aloud	< 6.54	6.54–20.16	≥ 20.16
3	silent	< 7.76	7.76–23.92	≥ 23.92
4	aloud	< 7.75	7.75–22.45	≥ 22.45
4	silent	< 9.22	9.22–26.61	≥ 26.61
5	aloud	< 9.98	9.98–23.95	≥ 23.95
5	silent	< 10.93	10.93–29.76	≥ 29.76

Table 3
Reading speed thresholds (letters per second) for the three difficulty classes.

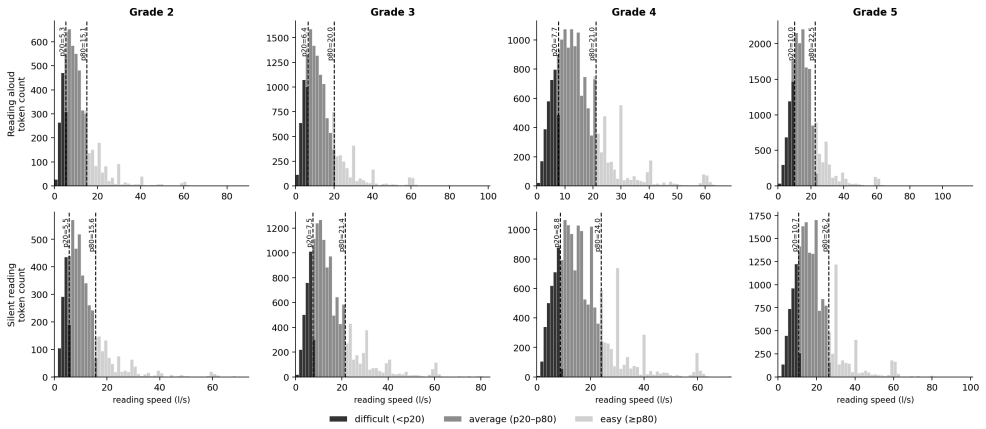

Figure 1
Per-token reading speed distributions (letters per second) for each grade × modality. Dashed lines mark the 20th and 80th percentiles, which define the boundaries between the three difficulty classes (dark – *difficult*; gray – *average*; light gray – *easy*).

Two patterns are immediately visible: distributions shift progressively rightward across grades, reflecting developmental gains in reading speed, and silent thresholds sit systematically above aloud thresholds within each grade, i.e. silent reading is faster, with the gap widening with grade progress.

4.2 Phonological Features

Word Length in Letters and syllables. Table 4 shows that word length is significantly associated with reading difficulty in every grade and modality as the average length (both in terms of number of letters and in terms of number of syllables) of easier/faster to read words is shorter than for more difficult words. However, the difficult and average classes are not separable by length alone.

		Length in letters				Length in syllables			
Gr	**Mod**	H	p	D / A	E	H	p	D / A	E
2	aloud	292	$<10^{-60}$	4.63 / 4.60	3.24	287	$<10^{-62}$	1.98 / 1.92	1.33
2	silent	340	$<10^{-70}$	4.79 / 4.45	3.11	346	$<10^{-75}$	2.04 / 1.87	1.25
3	aloud	411	$<10^{-89}$	4.61 / 4.62	3.60	399	$<10^{-86}$	1.95 / 1.94	1.49
3	silent	685	$<10^{-148}$	4.67 / 4.55	3.31	682	$<10^{-148}$	1.97 / 1.90	1.35
4	aloud	835	$<10^{-181}$	4.51 / 4.76	3.30	774	$<10^{-168}$	1.90 / 2.01	1.39
4	silent	1200	$<10^{-260}$	4.69 / 4.65	3.06	1114	$<10^{-241}$	1.99 / 1.96	1.25
5	aloud	1027	$<10^{-223}$	4.51 / 4.80	3.30	986	$<10^{-214}$	1.92 / 2.03	1.38
5	silent	1520	$<10^{-300}$	4.75 / 4.68	3.00	1454	$<10^{-300}$	2.02 / 1.98	1.23

Table 4
Word length in letters (left) and in syllables (right) across the three difficulty classes, by grade and modality. H is the Kruskal-Wallis statistic on 2 degrees of freedom. Mean length is reported as D / A (the *difficult* and *average* classes do not differ significantly) and E (*easy*).

Consonant Contents. Table 5 shows that both the consonant-to-letter ratio and the presence of a 3+ consonant clusters are significantly associated with reading difficulty for all grades and both modalities, but the differences are small between the difficult/average and the easy category. Only 2.5% of tokens overall contain a 3+ consonant cluster, so the results reported on this feature are not reliable.

4.3 Part of speech

Results show that part of speech is significantly associated with reading difficulty (Figure 2): function words (coordinating and subordinating conjunctions, particles, prepositions, pronouns) are over-represented in the easy class, while content words (nouns, verbs, adjectives, numerals, proper nouns) are over-represented in the average and difficult classes. This function/content asymmetry replicates the one reported by Marzi et al. (2026) for Bulgarian and Italian children and is consistent with the broader claim that function words automatise earlier in development than content words.

		Consonant-to-letter ratio				3+ consonant cluster			
Gr	**Mod**	H	p	D / A	E	χ^2	p	D / A	E
2	aloud	11.1	.004	0.538 / 0.537	0.527	10.6	.005	3.65% / 3.57%	1.74%
2	silent	3.1	.209	0.545 / 0.532	0.530	6.2	.045	4.09% / 3.12%	2.30%
3	aloud	10.4	.006	0.538 / 0.538	0.529	13.3	.001	3.36% / 2.71%	1.89%
3	silent	8.6	.013	0.541 / 0.537	0.532	10.6	.005	2.90% / 2.74%	1.78%
4	aloud	17.1	$<10^{-3}$	0.531 / 0.535	0.523	24.7	$<10^{-5}$	2.71% / 2.70%	1.27%
4	silent	30.3	$<10^{-6}$	0.531 / 0.536	0.512	29.7	$<10^{-6}$	2.78% / 2.68%	1.17%
5	aloud	27.1	$<10^{-5}$	0.532 / 0.534	0.524	28.0	$<10^{-6}$	2.14% / 2.67%	1.26%
5	silent	52.8	$<10^{-11}$	0.537 / 0.535	0.514	35.2	$<10^{-7}$	2.42% / 2.53%	1.02%

Table 5
Consonant-to-letter ratio (left, Kruskal-Wallis) and presence of at least one cluster of three or more consonants (right, chi-square on the binary feature), across the three difficulty classes by grade and modality. Mean ratio and % of tokens containing a 3+ cluster are reported as D / A (the *difficult* and *average* classes do not differ meaningfully) and E (*easy*).

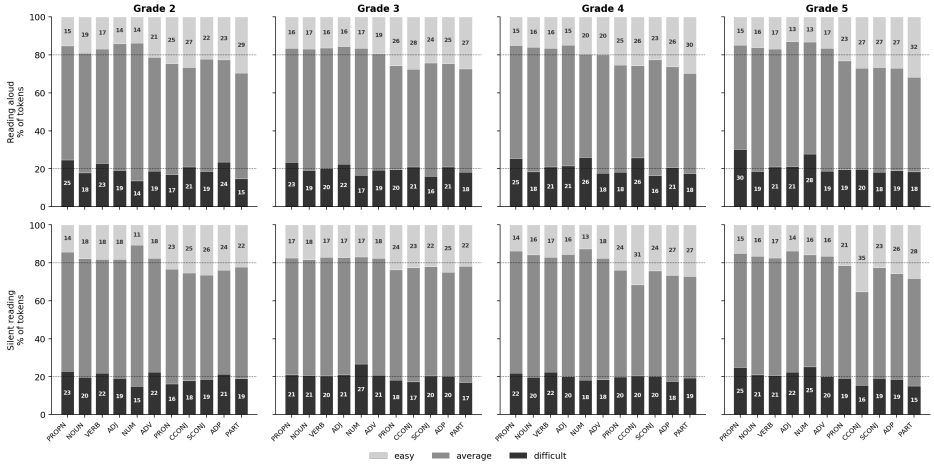

Figure 2
Classification of each POS (grade × modality) into the three difficulty classes. Colours: *difficult* (dark gray, bottom), *average* (gray, middle), and *easy* (light gray, top). The dashed lines mark the 20% and 80% boundaries that would obtain under independence.

However, the pattern is not uniform across function words. Despite being short and high-frequency, pronouns and conjunctions do not share the same level of automatisation as the other function-word categories (Figure 3).

Pronouns require additional processing time to resolve their referential meaning to an antecedent – a cognitive cost that reflects in the reading speed.

Similarly, conjunctions require processing time to uncover the structure of the sentence and the relation between the phrases or clauses they connect.

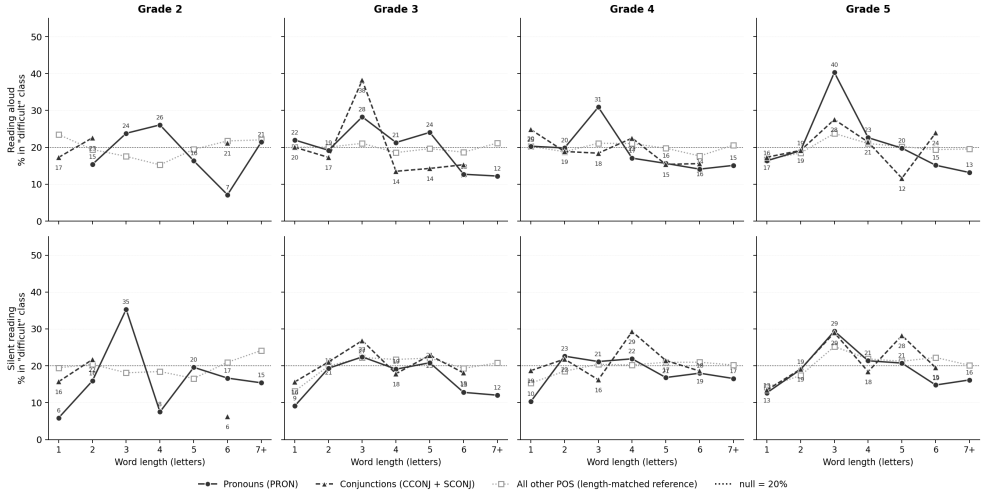

Figure 3
Comparison on pronouns and conjunctions classified into the difficult category, for aloud and silent reading, by grade.

Another diversion is that of proper nouns – they show the largest share of difficult tokens of any category (24.6%), reflecting the cost of reading unfamiliar names (*Зиги*, *Андрей*, etc.) for which children have no pre-existing lexical knowledge.

A closer look at pronouns and conjunction reveals that in some cases they are classified as difficult (read slowly by children) much more often than compared to function words in general. Two important observations can be made about the results on Figure 3. First, we are interested in the peaks for short pronouns and conjunctions which are clear outliers, signifying particular lexical elements that cause problems and slow reading. Secondly, we notice that these differences are dropped for silent reading, which can have a significance for the quality of reading and comprehension in silent mode if the reading speed of these items is in fact linguistically motivated.

Figure 4 compares the different pronoun subtypes, separately for each grade and modality. Two patterns dominate. First, articulation while reading aloud exposes systematic differences between pronoun subtypes that silent reading largely neutralises. Second, the indefinite and negative pronouns are read substantially more slowly than the others. Developmental change across grades 2–5 is small for all subtypes, suggesting that the general cost reflects the inherent referential demands of these pronouns rather than a transient difficulty that primary-school readers outgrow.

Figure 5 of the most difficult pronoun forms show two different trends across grades. For some items difficulty declines monotonically from grade 2 to grade 5, indicating genuine automatisation. However, for most of the pronouns difficulty plateaus or rises, reflecting the increasing referential complexity of the constructions in which these pronouns appear in higher-grade texts.

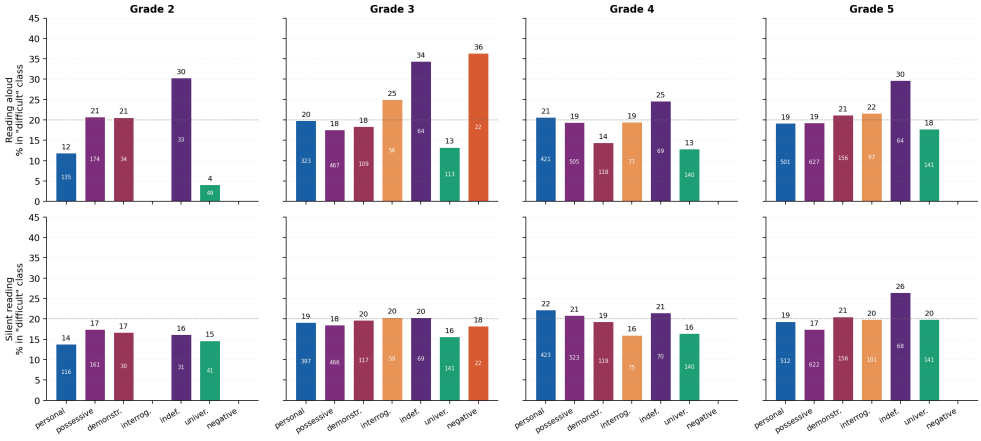

Figure 4
Reading difficulty by pronoun subtype, by grade and modality.

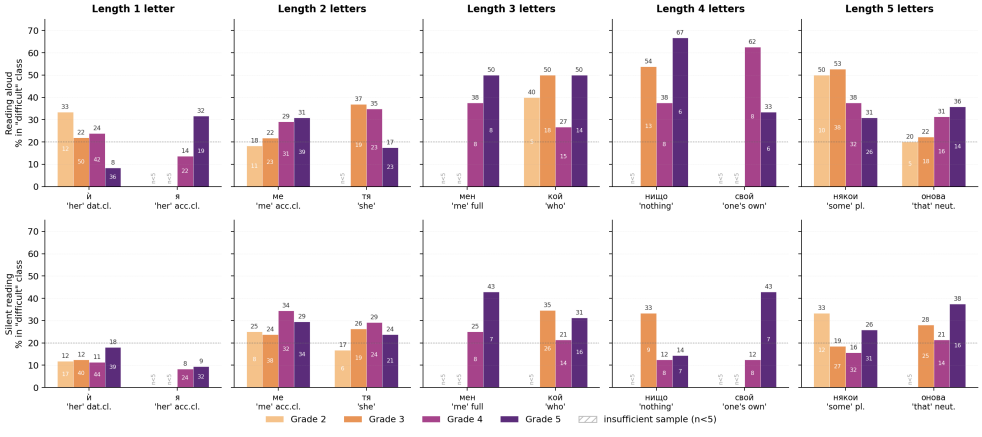

Figure 5
Most difficult pronoun forms of lengths 1–5 letters, by grade and modality.

Figure 5 also reveals a consistent modality asymmetry at individual pronoun level as seen in generalised terms before (Figure 3): reading aloud sits above silent reading in almost all cases. As discusses, this may be a sign of better automatisation while reading silently, but does not eliminate the risk of mechanical reading with less comprehension depth. These observations show the need to perform a more focused analysis of reading patterns with view to function words, and in particular pronouns.

The difficult nouns in grade 2 are mostly definite plural or definite singular forms of moderate length (5–7 letters), e.g. *вълните* ‘wave’ (fem., pl., def.), *балкона* ‘balcony’ (masc., sg., def.), *стаята* ‘room’ (fem., sg., def.). From grade 3 onwards, the difficult nouns are

longer and not part of everyday vocabulary, e.g. *небосвода* ‘sky dome’ (figurative, masc., sg., def.), *аквариум* ‘aquarium’ (masc., sg.), *въртележки* ‘carousel’ (fem., pl.), *лунапаркът* ‘amusement park’ (masc., sg., def.).

Grade 2 difficult verbs are short (3–7 letters) and high-frequency but either impersonal (*има* ‘(there) is/are’) or in past tense (*имаше* ‘was having’), and in some cases there is grammatical ambiguity of the verb form (*видя* can be 1st person singular present tense ‘see’, or 2nd/3rd person singular aorist ‘saw’). From grade 3 onwards, the difficult category is dominated by past participles used adjectivally (*набраздена* ‘furrowed’, *обрасла* ‘overgrown’, *обзаведена* ‘furnished’, *вкаменен* ‘petrified’), which are formally verbs but functionally adjective-like and morphologically complex.

Grade 2 difficult adjectives are usually phonologically or morphoogically challenging, e.g. *кръгъл* ‘round’ (contains two letters ‘ъ’), *внушителни* ‘impressive’ (pl.), *сияещо* ‘glowing’ (neut.), etc. From grade 3 onwards difficult adjectives are long, low-frequency, semantically rich and inflectionally demanding, e.g. , *благородната* ‘noble’ (fem., sg., def.), *теракотена* ‘terracotta’ (fem.), *причудливи* ‘whimsical’ (pl.).

The auxiliary verb *съм* ‘to be’ appears to be predominantly easy to read (Figure 6), which is not surprising since its forms are also short. More thorough analysis is required to determine what causes the difficulty and the differences across modality. Across grades, while general verb difficulty follows consistent distribution, the difficulty of the auxiliary verb shows irregular patterns. The fluctuations in difficulty are likely to be caused by the more complex verb forms, e.g. perfect, passive, etc., and constructions that involve the auxiliary (*трябва да беше дошъл* ‘should have probably arrived’; *са били категорично против* ‘have been firmly against’).

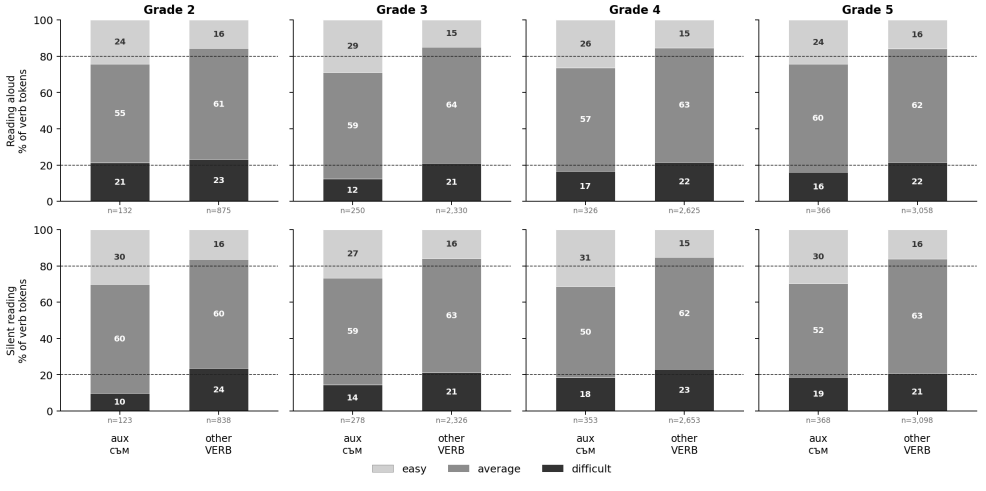

Figure 6
Difficulty level of the auxiliary verb, by grade and modality, compared to general verb reading difficulty.

4.4 Morphological Complexity

As already seen, morphologically complex words usually take longer time to read, but since they are also longer in length, it is difficult to judge these two features separately. Here we examine which particular morphemes may pose more challenges and whether the difficulty fades with grade progress.

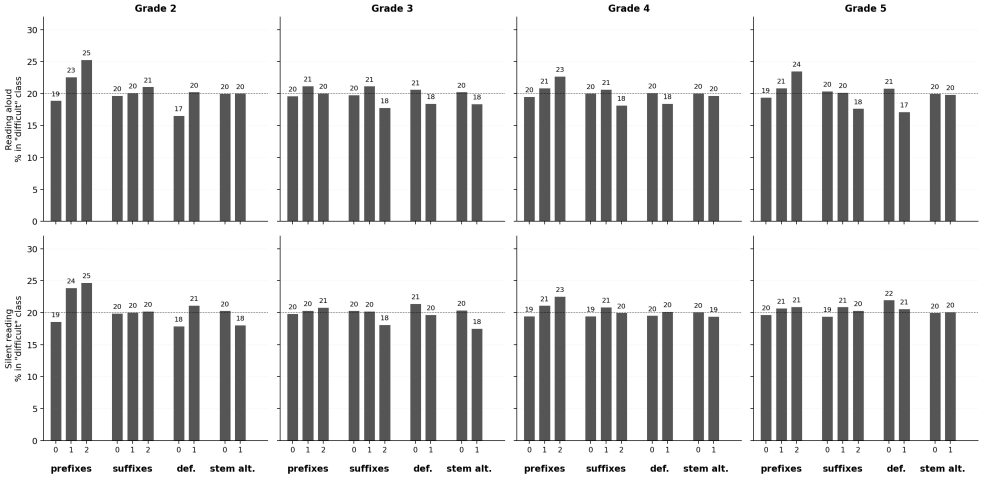

Figure 7
Difficulty of morphologically complex words, by grade and modality.

Figure 7 shows the influence of four morphological and phonological features – number of prefixes, number of suffixes, presence of a definite article, and stem alternation in wordforms. The results show that these are statistically significant but weak associations with reading difficulty (much smaller than the effect of word length).

Among the four, prefixation impedes reading the most. Multisuffix words are generally no slower to read and at higher grades are even slightly faster. This confirms the common belief that when a multi-morphemic ending is a familiar pattern, the reader processes it as a unit and the complexity of the form does not require longer processing time.

This is further confirmed by the fact that the definite article shows almost no effect across grades, indicating that this overt Bulgarian-specific morphological marker is largely automatised very early, even in grade 2.

Stem alternations also show no consistent effect, but since the lexemes that undergo such changes in the texts are high-frequency items (*ръка – ръце*, *око – очи*, *голям – големи*) whose forms are learned early and retrieved holistically, no reliable conclusions can be drawn.

The fact that none of the four features shows a clear developmental trajectory across grades 2–5 reinforces the overall picture that morphological properties are not slowly automatised across primary school, but are either trivial from the start (definite article, regular suffixation) or remain a relatively trivial challenge (prefixation). For a readability index, this

means that morphological complexity should have a lower weight alongside other more pronounced features.

4.5 Frequency Properties

In general, across content categories, the slowest words are low-frequency words that likely the readers encounter for the first time. Figure 8 shows a robust trend stable across grade and modality, that subsequent occurrences are read faster than the first one, and most often the difficulty level moves out of the *difficult* into the *average* or *easy* category.

Silent reading produces slightly larger differences with repetition than reading aloud at grades 2–3, and the two modalities converge at grades 4–5. Some lemmas, particularly *вещица* 'witch' and *кралица* 'queen' – remain in the *difficult* class even after multiple encounters, indicating that within-session repetition does not always influence lexically demanding items; for these cases, the contextual complexity of the constructions in which they appear continues to impose a processing cost that repetition cannot fully overcome.

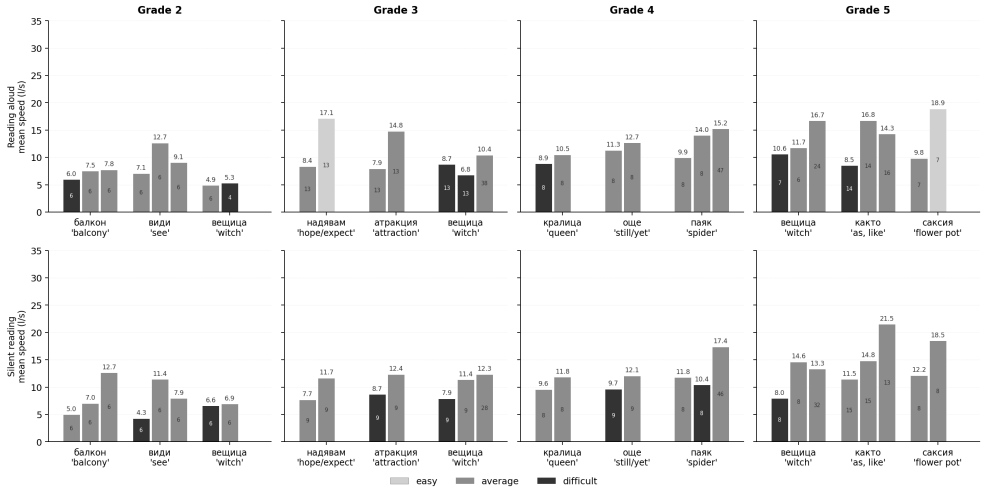

Figure 8
Progression of reading speed for content words at first, second and any further occurrences of the words within a reading session, by grade and modality.

The detailed analysis of word frequency in the language and their reading difficulty is beyond the scope of the current study. Here we only acknowledge that with each subsequent occurrence of a word, in general its reading difficulty decreases, at least for the span of the text. This is valid for both silent reading and reading aloud.

Overall, feature-based readability index should incorporate token-recurrence count alongside the static lexical features of word length, frequency, and part of speech. Moreover, this can be implemented as a strategy for learning new and more complex words by repetition in the learner texts.

4.6 Syntactic Features

Syntactic function correlates with reading difficulty (Figure 9), but the effect is derived in combination with the lexical content typically appearing in each role. It is not entirely clear what the relation contributes once those lexical correlates are accounted for, and further analysis is needed in order to determine the weight of generalised dependency features in a feature-based readability index.

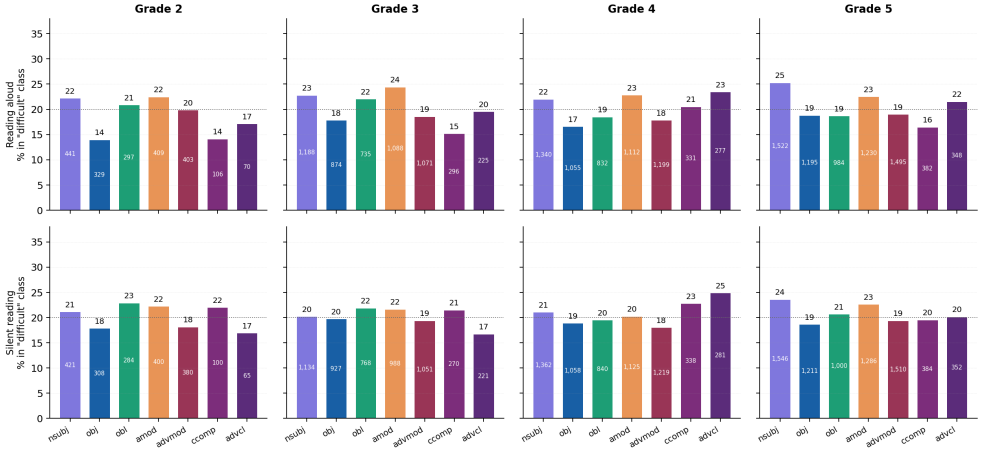

Figure 9
Reading difficulty for seven core dependency relations (UD schema), by grade and modality.

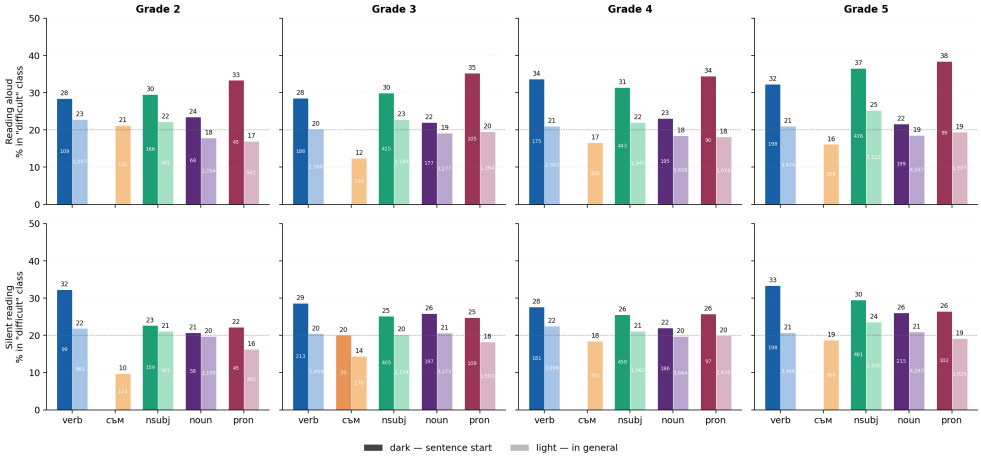

Figure 10
Comparing reading difficulty of verbs and nouns at the start of the sentence and in general, by grade and modality.

However, a more focused investigation of syntactic properties involving other aspects of the syntactic structure and in particular, features specific for Bulgarian, reveal more relevance.

Figure 10 shows that the position in the sentence is correlated with difficulty, and this is consistent across categories, grades, and modalities. Sentence-initial position shows more difficulty compared to the corresponding overall rate. The effect is more pronounced for verbs as an initial verb requires the reader to recover the non-overt subject, or process an impersonal sentence.

Subjects in initial position also require more processing time, most notably if it is expressed by a pronoun, which also requires anaphora resolution. Nouns are the least affected category.

Across modalities, reading aloud and silent reading produce comparable positional effects, but in general, for silent reading the difficulty gap is less pronounced.

For a feature-based readability index, sentence-initial position deserves an independent entry – particularly when scoring sentences whose first word is a verb or a pronoun subject, because the cost it imposes is not predicted by length, frequency, or POS alone.

Further, our observations show that reading speed of word bigrams can also serve as a predictor of grade-level fluency (Figure 11). Speed of main syntactic structures rises monotonically with grade, the slope is steeper than any single-word feature. For a feature-based readability index, the type of POS sequence a sentence is composed of carries diagnostic value beyond its individual word features. Figure 11 also shows that the most frequent syntactic structures are read with speed close to the average speed for the age group. This means that typical structures are generally not problematic for the children, and more focus is needed on the use of structures of lower frequency.

One final observation concerns phrase and clause boundaries within the sentence marked by punctuation (commas) or conjunctions. Figure 12 show that the average reading speed of bigrams that include the boundaries falls at the lower end of the *average* category. Phrase and clause boundaries cause slight slowing and this effect is consistent through grades and modalities, which shows that this is a regular reading feature rather than a learner's deficiency.

5. Discussion of Results

The results reveal several key findings that are directly relevant with a view to designing a readability index in Bulgarian.

1. **Word length is by far the strongest single predictor of reading difficulty.** Every other lexical, morphological, or syntactic feature we tested shows a lesser effect. Length emerges as the dominant predictor consistently across grade and modality. This in fact reaffirms the previously discussed indices such as FLI, LIX, CLI, etc. (see Section 2.3).

2. **Part-of-speech class is also a strong predictor and the results reveal a clear function-vs.-content words division.** The split is symmetric across modal-

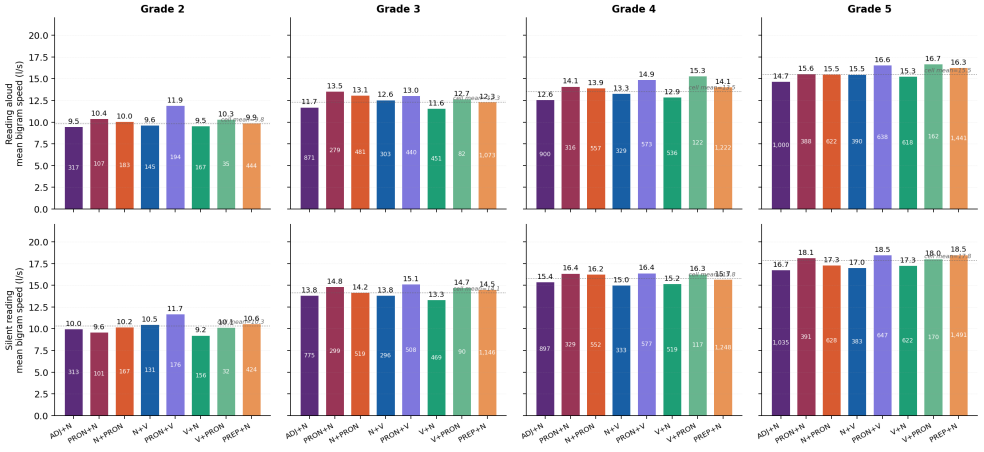

Figure 11
Reading speed of selected syntactic structures (bigrams), by grade and modality.

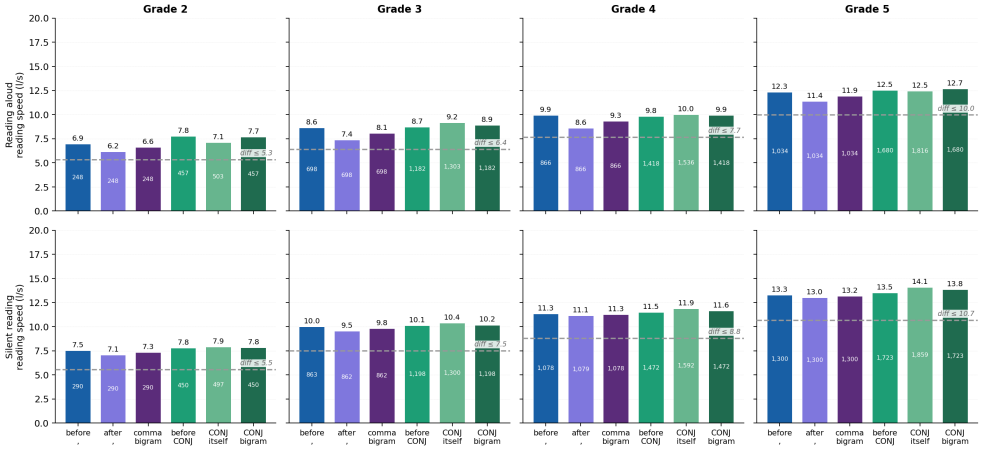

Figure 12
Reading speed at mid-sentence punctuation and conjunction boundaries, by grade and modality.

ities and stable across grades. In general, the auxiliary verb *съм* does not pose any significant difficulties, except in certain more complex structures in higher grade texts.

3. **Pronouns are associated with increased processing cost consistent across grades.** Among pronoun subtypes, indefinite (*някой, нещо*) and negative pronouns (*никой, нищо*) bear the heaviest cost because they introduce or exclude referents rather than retrieve them from prior discourse.

4. **Sentence-initial position imposes an additional processing cost that does not fade across grades.** Verbs and pronoun subjects at sentence-initial position show more challenges relative to the general baseline. The cost is consistent across grades, indicating a structural rather than developmental effect.

5. **Clause and phrase boundaries within the sentence require an additional processing cost consistent across grades.** Words before and after punctuation, as well as conjunctions, which mark phrase and clause boundaries, show lower reading speed. The cost is consistent across all four grades, indicating a structural property.

6. **Within-session word repetition is one of the most consistent accelerators of reading fluency.** The share of tokens classified as *easy* grows with every subsequent occurrence of a word in the text. The effect is strongest at grade 2 and gradually attenuates, but does not disappear at higher grades.

7. **Morphological complexity features show small influence.** The number of prefixes, suffixes, and the presence of a definite article, as well any stem irregularity that appear in a wordform, correlate with difficulty in the predicted direction but are not able to predict difficulty on their own. Notably, the effect disappears at higher grades, indicating that compound morphology is chunked and automated efficiently with progress of reading skills. This is a feature that can be used to evaluate readability only for beginner learners and to identify problems early on.

8. **Syntactic function carries a small but reliable signal that overlaps with POS.** Across the core dependency relations (`nsubj`, `obj`, `obl`, etc.), the slowest is the subject, which also is associated often with a sentence-initial position, and is a content word. The effect is largely a consequence of the lexical content typical of each role rather than the role itself.

9. **Reading speed of bigram POS structures is a developmental signal.** Mean reading speed for bigrams (ADJ+N, N+V, ADP+N, *etc.*) rises monotonically from grade 2 to grade 5, both in silent reading and reading aloud.

10. **Silent reading is consistently faster than reading aloud across all grades and all features, and the gap widens with age and development of reading skills.** Silent reading compresses the variability across some features, e.g. the difference between the slowest and fastest pronoun subtypes. On one hand, articulation exposes processing differences that silent reading partly absorbs. However, on the other hand, irregular patterns in speed in silent reading may suggest distraction and lower quality of reading comprehension.

Concerning the features most strongly predicting reading speed (**RQ1**, see Section 1), the empirical data provides clear insights. Word length is the dominant lexical predictor and outweighs every other tested feature in isolation. Other important features are POS, primarily the content-vs-function words distinction that drives much of the variance in

Feature	Strong	Weak	Stable across grades	Stable across modality
Word length	✓	×	✓	partially
POS class (content vs. function)	✓	×	✓	✓
Auxiliary *съм* vs. lexical verbs	✓	×	✓	✓
Sentence-initial position	✓	×	✓	partially
Clause / phrase boundary	✓	×	✓	×
Within-text repetition	✓	×	×	partially
POS bigram structures	✓	×	×	✓
Number of prefixes	×	✓	✓	✓
Suffixal complexity	×	✓	partially	✓
Definite article	×	✓	✓	✓
Stem alternations	×	✓	✓	✓
Syntactic function (UD dependency)	×	✓	✓	✓

Table 6
Summary of the features and their classification according to their contribution to readability. **Strong** = strong predictor of difficulty in the empirical data; **Weak** = statistically significant but weak predictor; **Stable across grades** = the effect magnitude does not change substantially between grades 2 and 5; **Stable across modality** = the effect magnitude is approximately the same in reading aloud and silent reading. The label 'partially' indicates that the effect varies systematically or shows only a partial dependency.

reading speed, sentence-initial position with a focus on sentences starting with a verb (impersonal or pro-drop), clause and phrase boundary positions. Morphological features – prefixes, suffixes, definiteness, stem alternation, carry only small independent weight. Syntactic function contributes only marginally to what POS already captures. Within-text repetition modulate the cost of individual challenging words by 30–70 % in each subsequent mention which suggests this as a possible strategy to improve readability.

For a feature-based readability index, these observations give a minimal effective feature set consisting of strong basic features and weaker secondary features to evaluate the readability of a text.

Concerning the features that remain stable across grades and the ones which gain or lose weight as decoding becomes more automatised (**RQ2**, see Section 1), the analysis points to two qualitatively different developmental patterns. The first applies to the lexical predictors (length, POS, function-vs-content split), which attain stable importance although with decreasing magnitude – the effects remain detectable at grade 5 although the absolute reading speed at every feature rises. The second pattern is characteristic of within-session repetition and the processing cost of sentence-level components, which show stable magnitude – the repetition gain and the cost of processing pronouns, sentence-initial words and clause and phrase boundaries are present at both grade 2 and grade 5, even as children

become more fluent in reading. The morphological features (prefixes, suffixes, definiteness, stem alternations) show only small magnitude and do not exhibit a clear developmental trajectory, suggesting that these features represent stable, low-weight processing costs that are neither automatised away nor accentuated as proficiency grows.

The most significant developmental feature is the bigram-speed trajectory of typical and high frequency syntactic sequences (ADJ+N, N+V, etc.), which doubles between grade 2 and grade 5 in roughly parallel fashion, indicating that what improves with development is the reader's general fluency on multi-word chunks of particular syntactic relations. More thorough analysis of syntactic features is necessary in order to integrate syntactic information into a readability index.

Concerning the features that are consistent or differ in reading aloud and in silent reading (**RQ3**, see Section 1), we make two main observations. First, silent reading consistently raises absolute reading speed across all features and grades. Second, and more interesting, in some cases silent reading drops some of the relevant features. One possible explanation is that silent reading neutralises features tied to articulatory or phonological complexity. However, smaller difficulty gaps under silent reading in structural positions requiring extra processing cost (e.g., anaphora resolution, sentence structure, etc.) may suggest loss of attention and deteriorated comprehension. While in general, the hypothesis articulated in RQ3 — that phonological and articulatory features weigh more heavily in reading aloud, is supported by the data, any further conclusions need to be based on a detailed examination of reading comprehension in reading aloud as compared to reading silently, especially in lower grades.

Several findings of the present analysis converge with those reported by Lento et al. (2024) and Marzi et al. (2026) on the same finger-tracking dataset. In particular, the studies confirm a robust developmental trajectory of reading speed from grade 2 to grade 5 and a stable content-vs-function-word contrast in which function words are read faster and automatised earlier than content words in both modalities. The consistency of these observations supports their status as developmental phenomena in Bulgarian primary-school reading, and contribute to the efforts to distinguish reading developmental features (which evolve across grades) from stable decoding and comprehension features (which are related to text processing and information comprehension).

6. Notes on Reading Development Strategies

Based on the findings presented in the paper, we can compile a list of practical recommendations to support reading instruction and development of reading skills and text design for Bulgarian learners.

Vocabulary and Morphology. Before or after reading, attention should be paid to unfamiliar or more difficult words and those that require more processing time. In learners' texts, new, unfamiliar or long words should appear more than once, in different morphological forms and in different syntactic positions.

Pronoun Use. Reading should be combined with exercises on the use and comprehension of pronouns, including the correct identification of antecedents in more complex cases.

Particular attention should be paid to possessive and personal reflexive pronouns, indefinite and negative pronouns, and to the correct resolution of antecedents from the context.

Verb-Initial Constructions and Sentence Segmentation. Sentences that begin with a verb deserve special attention, especially those with complex verb forms where an auxiliary verb appears in sentence-initial position. Text and sentence segmentation during reading should also be explicitly taught, with particular emphasis on pausing correctly at the boundaries of phrases and sentences. Syntactic constructions that appear frequently in the language are more suitable for beginner readers, and gradual introduction of more complex syntactic structures could facilitate their acquisition without impeding reading speed significantly.

7. Conclusion

This paper has presented an analysis of reading data from children from grades 2 to 5. The main objective was to identify features measuring and predicting text readability for Bulgarian, with a focus on lexical and grammatical features. Our analysis confirms that established readability indices developed for English and Swedish are applicable but insufficient for Bulgarian, as they do not provide in-depth information to support any teaching or text design approaches to improve reading skills of learners.

The empirical findings yield a small but consistent set of strong predictors for text readability that need to be investigated further in order to be implemented into a readability index for Bulgarian.

The observations also show grade-by-grade developmental tendencies which reflect the gradual improvement of reading skills.

The aloud-vs-silent reading comparison reveals a systematic but bounded asymmetry. While silent reading is faster than reading aloud across all grades as it generally removes the cost of articulatory processing, it also raises the important concern of the quality of comprehension in younger readers.

Beyond the immediate application to readability scoring, the findings reported here suggest several directions for future work. First, the within-session repetition effect deserves dedicated study with materials designed specifically to test priming gradients across longer time intervals and across different text positions. Second, the modality discrepancies could be tested experimentally with stimuli that controllably vary articulatory load, sentence structure, and referent accessibility. Third, the lexicon-specific findings (the difficult words at each grade, the bigram patterns) could inform the construction of graded reading materials in which the introduction of demanding vocabulary is paced to take advantage of the priming gains demonstrated in the study.

The finger-tracking paradigm proved well-suited for collecting naturalistic reading-time data from young children at scale, and the resulting dataset provides an empirical foundation that text-corpus statistics alone cannot offer. The continued development of such corpora, in Bulgarian and in other under-resourced languages, is a productive route toward creating readability frameworks that are grounded in the actual reading behaviour of the children and the language specifics for which they are designed.

8. Limitations

Several caveats should be taken into account when interpreting the findings reported in this paper.

Sampling and population coverage. The data were collected at a single school across grades 2–5. The sample is therefore not statistically representative of Bulgarian primary-school readers as a whole, and the absolute reading-speed values, the thresholds, and the magnitudes of the effects should be interpreted as characterising this cohort rather than the wider population. The analyses report on correlations between linguistic features and reading difficulty and point to candidate predictors worth further investigation on larger and more diverse samples.

Difficulty thresholds. Reading difficulty is operationalised throughout a 20/60/20 split of the letters-per-second speed distribution, and words are classified as *difficult*, *easy*, and *average*. This threshold choice was made for analytic convenience, but this is a coarse discretisation of a continuous and skewed distribution of word reading speeds which may be approximated more precisely with other thresholds (e.g., a quintile-based segmentation). The analyses should be read as illustrative of tendencies rather than as precise quantifications, and any operational readability scoring system would need to justify any threshold choice in light of its intended end use.

Feature representation sets are mostly small. Some of the observations are made on a limited number of occurrences, which is insufficient to draw reliable conclusions. For some features a more targeted experiment would be needed in order to determine its significance for the readability of the text (e.g., pronouns in sentence-initial position).

Confounds between features. The features we examined are not statistically independent. Word length correlates with morphological complexity, frequency with the division between content and function words, part of speech with syntactic function, etc. A more complex analysis should be performed, e.g. multivariate regression, to disentangle the contributions of correlated features.

Reading comprehension was not measured. Reading speed is a measure of fluency, but does not reflect reading comprehension and information extraction. The difficulty classification applied here captures the *decoding* difficulty of words and sentences, and not necessarily the content difficulty of the text and the ability to process the information structure of the text. A future study pairing the finger-tracking data and decoding with post-reading comprehension and information extraction questions would be necessary to validate that the features identified as difficulty-correlated are also features that impede understanding, not just features that slow down decoding.

Acknowledgments

The work presented here is carried out with the support of the National Scientific Program "Development of Scientific Research and Innovation in Bulgarian Preschool and School Education", Institute of Education, Ministry of Education and Science, Agreement SP-13/04.11.2025.

References

Beyersmann, Elisabeth and Jonathan Grainger. 2018. Support from the morphological family when unembedding the stem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(1):135–142.

Beyersmann, Elisabeth, Jonathan Grainger, Séverine Casalis, and Johannes C. Ziegler. 2015. Effects of reading proficiency on embedded stem priming in primary school children. *Journal of Experimental Child Psychology*, 139:115–126.

Björnsson, Carl Hugo. 1968. *Läsbarhet*. Liber, Stockholm.

Breznitz, Zvia. 2006. *Fluency in Reading: Synchronization of Processes*. Routledge.

Brysbaert, Marc. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, 109.

Burani, Cristina. 2010. Word morphology enhances reading fluency in children with developmental dyslexia. *Lingue e Linguaggio*, 9(2):177–198.

Burani, Cristina, Stefania Marcolini, Daniela Traficante, and Pierluigi Zoccolotti. 2018. Reading derived words by Italian children with and without dyslexia: The effect of root length. *Frontiers in Psychology*, 9:647.

Carlisle, Joanne F. 2000. Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12(3):169–190.

Casalis, Séverine, Mathilde Dusautoir, Pascale Colé, and Stéphanie Ducrot. 2009. Morphological effects in children word reading: A priming study in fourth graders. *British Journal of Developmental Psychology*, 27(3):761–766.

Coleman, Meri and T. L. Liau. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2):283–284.

Crepaldi, Davide, Marcello Ferro, Claudia Marzi, Andrea Nadalini, Vito Pirrelli, and Loukia Taxitari. 2022. Finger movements and eye movements during adults' silent and oral reading. In Ronit Levie, Amalia Bar-On, Orit Ashkenazi, Elitzur Dattner, and Gilad Brandes, editors, *Developing Language and Literacy: Studies in Honor of Dorit Diskin Ravid*. Springer International Publishing, pages 443–471.

De Luca, Maria, Laura Barca, Cristina Burani, and Pierluigi Zoccolotti. 2008. The effect of word length and other sublexical, lexical, and semantic variables on developmental reading deficits. *Cognitive and Behavioral Neurology*, 21(4):227–235.

Dell'Orletta, Felice, Simonetta Montemagni, and Giulia Venturi. 2011. READ-IT: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies, Edinburgh, Scotland, UK*, pages 73–83, Association for Computational Linguistics.

Ehri, Linnea C. 2005. Learning to read words: Theory, findings, and issues. In *Scientific Studies of Reading*, volume 9. Taylor & Francis, pages 167–188.

Ferro, Marcello, Claudia Cappa, Sara Giulivi, Claudia Marzi, Ouafae Nahli, Franco Alberto Cardillo, and Vito Pirrelli. 2018. ReadLet: Reading for understanding. In *Proceedings of the 5th IEEE Congress on Information Science and Technology (IEEE CiSt'18), Marrakech, Morocco*, pages 1–6.

Flesch, Rudolf. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233.

Hasenäcker, Jana and Sascha Schroeder. 2017. Syllables and morphemes in German reading development: Evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, 38(3):733–753.

Joseph, Holly S. S. L., Simon P. Liversedge, Hazel I. Blythe, Sarah J. White, and Keith Rayner. 2009. Word length and landing position effects during reading in children and adults. *Vision Research*, 49(16):2078–2086.

Kim, Young-Suk, Richard K. Wagner, and Elizabeth Foster. 2011. Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading*, 15(4):338–362.

Koeva, Svetla, Nikola Obreshkov, and Martin Yalamov. 2020. Natural language processing pipeline to annotate Bulgarian legislative documents. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020), Marseille, France*, pages 6988–6994, European Language

Resources Association.
LaBerge, David and S. Jay Samuels. 1974. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2):293–323.
Lento, Alessandro, Andrea Nadalini, Marcello Ferro, Claudia Marzi, Vito Pirrelli, Tsvetana Dimitrova, Hristina Kukova, Valentina Stefanova, Maria Todorova, and Svetla Koeva. 2024. Assessing reading literacy of Bulgarian pupils with finger-tracking. In *Proceedings of the Sixth International Conference on Computational Linguistics in Bulgaria (CLIB 2024), Sofia, Bulgaria*, pages 140–149.
Marzi, Claudia, Marcello Ferro, Andrea Nadalini, Vito Pirrelli, Maria Todorova, Tsvetana Dimitrova, Valentina Stefanova, Hristina Kukova, and Svetla Koeva. 2026. Comparable reading development in Bulgarian and Italian: Cross-linguistic insights from a finger-tracking study. *Languages*, 11(4):70.
Marzi, Claudia, Anna Rodella, Andrea Nadalini, Loukia Taxitari, and Vito Pirrelli. 2020. Does finger-tracking point to child reading strategies? In *Proceedings of the 7th Italian Conference on Computational Linguistics (CLiC-it 2020), Bologna, Italy*, volume 2769.
Nadalini, Andrea, Claudia Marzi, Marcello Ferro, Loukia Taxitari, Alessandro Lento, Davide Crepaldi, and Vito Pirrelli. 2023. Eye-voice and finger-voice spans in adults' oral reading of connected texts: Implications for reading research and assessment. *The Mental Lexicon*, 18(3):366–400.
Perfetti, Charles. 1985. *Reading Ability*. Oxford University Press, New York.
Perfetti, Charles. 2007. Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4):357–383.
Pirrelli, Vito and Svetla Koeva. 2024. Developing materials for assessing reading literacy and comprehension of early graders in Bulgaria and Italy. *Foreign Language Teaching (Chuzhdoezikovo Obuchenie)*, 51(1):29–37.
Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422.
Rayner, Keith. 2009. Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8):1457–1506.
Schüppert, Anja, Wilbert Heeringa, Jelena Golubovic, and Charlotte Gooskens. 2017. Write as you speak? A cross-linguistic investigation of orthographic transparency in Germanic, Romance and Slavic languages. In Martijn Wieling, Martin Kroon, Gertjan van Noord, and Gosse Bouma, editors, *From Semantics to Dialectometry: Festschrift in Honour of John Nerbonne*. College Publications, pages 312–427.
Seymour, Philip H. K., Mikko Aro, and Jane M. Erskine. 2003. Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2):143–174.
Share, David L. 2008. On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, 134(4):584–615.
Taxitari, Loukia, Claudia Cappa, Marcello Ferro, Claudia Marzi, Andrea Nadalini, and Vito Pirrelli. 2021. Using mobile technology for reading assessment. In *Proceedings of the 6th IEEE Congress on Information Science and Technology (CiSt), Agadir, Morocco*, pages 302–307.
Traficante, Daniela, Stefania Marcolini, Alessandra Luci, Pierluigi Zoccolotti, and Cristina Burani. 2011. How do roots and suffixes influence reading of pseudowords: A study of young Italian readers with and without dyslexia. *Language and Cognitive Processes*, 26(4–6):777–793.
Wright, Gary, Ross Sherman, and Timothy B. Jones. 2004. Are silent reading behaviors of first graders really silent? *The Reading Teacher*, 57(6):546–553.
Ziegler, Johannes C., Daisy Bertrand, Dénes Tóth, Valéria Csépe, Alexandra Reis, Luís Faísca, Nina Saine, Heikki Lyytinen, Anniek Vaessen, and Leo Blomert. 2010. Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, 21(4):551–559.
Ziegler, Johannes C. and Usha Goswami. 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131(1):3–29.
Zoccolotti, Pierluigi, Maria De Luca, Gloria Di Filippo, Anna Judica, and Marialuisa Martelli. 2009. Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing*, 22(9):965–992.