Assessing Reading Literacy of Bulgarian Pupils with Finger-tracking

Alessandro Lento[1], **Andrea Nadalini**[2], **Marcello Ferro**[2], **Claudia Marzi**[2], **Vito Pirrelli**[2], **Tsvetana Dimitrova**[3], **Hristina Kukova**[3], **Valentina Stefanova**[3], **Maria Todorova**[3], **Svetla Koeva**[3]

[1] Biomedical Campus University, Rome Italy

[2]Italian National Research Council, Institute for Computational Linguistics

[3] Institute for Bulgarian Language, Bulgarian Academy of Sciences

[1]alessandro.lento@unicampus.it

[2]{andrea.nadalini,marcello.ferro,claudia.marzi,vito.pirrelli}@cnr.it

[3]{cvetana,hristina,valentina,maria,svetla}@dcl.bas.bg

Abstract

The paper reports on the first steps in developing a time-stamped multimodal dataset of reading data by Bulgarian children. Data are being collected, structured and analysed by means of *ReadLet*, an innovative infrastructure for multimodal language data collection that uses a tablet as a reader's front-end. The overall goal of the project is to quantitatively analyse the reading skills of a sample of early Bulgarian readers collected over a two-year period, and compare them with the reading data of early readers of Italian, collected using the same protocol. We illustrate design issues of the experimental protocol, as well as the data acquisition process and the post-processing phase of data annotation/augmentation. To evaluate the potential and usefulness of the Bulgarian dataset for reading research, we present some preliminary statistical analyses of our recently collected data. They show robust convergence trends between Bulgarian and Italian early reading development stages.

Keywords: reading literacy assessment, finger-tracking, multimodal dataset.

1 Introduction

Good reading and text comprehension skills are key competences and essential prerequisites for high-quality education (Pikulski and David, 2011). Reading ability can predict performance in all subjects where reading and writing play a role, with reading speed being the most important indicator. In the long term, students with early reading difficulties have serious challenges with general learning, academic performance and social integration (Tichá et al., 2009; Nese et al., 2013). Hence, it is not surprising that educational systems across Europe have put considerable effort into teaching children core reading competencies. Literacy research has been at the forefront of this effort, supporting evidence-based practices for reading and language classes of schools everywhere. Ideally, education should be supported by continual observation of actual reading behaviour. However, major international organisations such as UNICEF (Chzhen et al., 2018) and OECD (OECD, 2023) have lamented a regrettable shortage of large scale reading data.

So far, two major sources of reading complexity have slowed down progress in collecting longitudinal reading data at scale. First, most recent and influential (eye-tracking) research on reading has typically focused on reading single words or sentences (Rayner, 1998, 2009). However, the need to monitor real reading data in real-life settings raises increasing concerns with the ecological validity of behavioural language data (Brennan, 2016; Demberg and Keller, 2019; Hasson et al., 2018; Willems, 2015), and requires shifting the research focus away from specific, highly controlled phenomena, to real-time processing issues (Jarodzka and Brand-Gruwel, 2017; Kaakinen and Hyönä, 2008; Verhoeven and Perfetti, 2008).

Secondly, the advent of eye-tracking technology at the services of eye movement research started a prolonged period of little interest in the vocal component of reading (with only few exceptions such as De Luca et al. (2013)) and a general neglect of the inherently multi-sensory nature of reading. In fact, a key cognitive insight in developing this ability occurs as learners are able to integrate three emerging sources of information about print and speech: i) the auditorily anchored understanding of syllables, ii) the linguistic-conceptual knowledge of words, and iii) the unfolding visuospatial understanding of printed words built upon the visual and (possibly) tactile exploration of the words' spatial dimension, as it occurs in finger-point reading (Mesmer and Lake, 2010; Mesmer and Williams, 2015). In attaining an efficient synchronisation between word pointing and the onset of word artic-

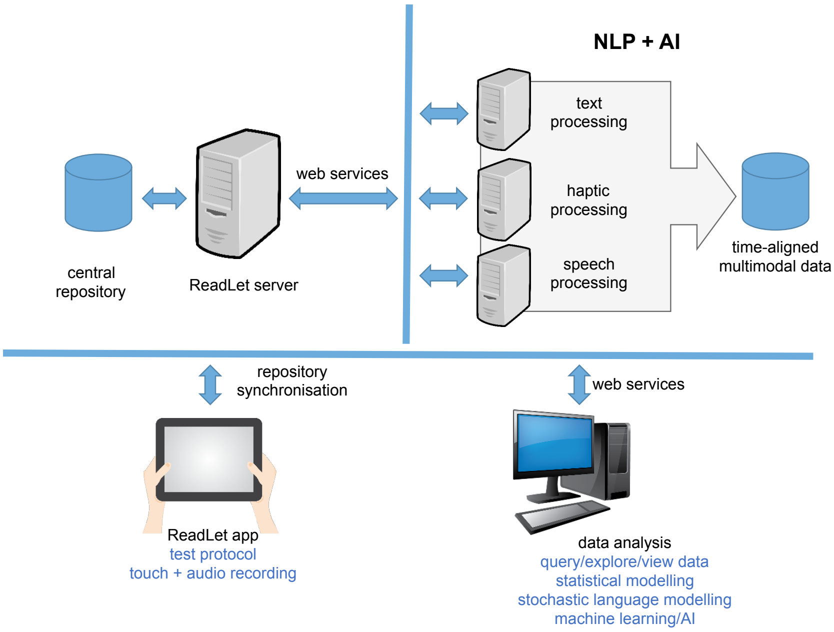

Figure 1: The physical architecture of the ReadLet infrastructure.

ulation, the learner must resolve the competing information between the multiple syllables that (s)he hears and feels and individual words that (s)he sees on a printed page (Mesmer and Lake, 2010; Uhry, 1999, 2002). While some of these reading aspects have been explored and investigated independently, much less work has been conducted so far to study their interaction, also because of the technical difficulty with concurrently recording asynchronous time-series of multimodal signals.

The *ReadLet* infrastructure (Crepaldi et al., 2022; Ferro et al., 2018; Taxitari et al., 2021) was designed and developed to meet most of these methodological and technical *desiderata*, with a view to collecting finely annotated, time-stamped, naturalistic reading sessions of connected texts, in both silent and oral reading modalities, by either child or adult readers. Here, we report on an ongoing project that uses *ReadLet* to collect sessions of reading data of Bulgarian primary school pupils, audio-recorded and “finger-tracked” in Sofia. The project aims to:

A) design and implement assessment strategies for monitoring and evaluating the reading and word comprehension skills of Bulgarian and Italian early graders;

B) collect reading and comprehension data from the two populations of children using the same protocols and comparable, rigorously controlled reading texts;

C) compare the data collected from children of different ages and languages (Bulgarian and Italian) and modelling the results.

Activities (B) and (C) are still underway and only some preliminary results are reported here. In what follows, we provide a broad outline of the *Readlet* architecture (section 2), a technical description of the materials and the experimental protocol adopted for data collection (section 3), an overview of the data collected so far (section 4), and a battery of preliminary analyses (section 5). Some conclusions and prospective directions of the current project are offered in the final section.

2 ReadLet

The *ReadLet* infrastructure supports a battery of specialised web services based on a cloud computing architecture (Figure 1). The user endpoint is a native app running on an ordinary tablet with an Internet connection. The app manages the administration of the reading protocol and the concurrent recording of finger touch events and the reader’s voice in oral reading sessions. Upon a reading session being completed, data are encrypted, pseudonymised and transmitted to the central repository through a secure Internet protocol. No session data are locally saved in the tablet’s internal storage after data transmission is successfully carried out.

The core of the architecture is a cloud server that exposes a set of functionalities interfacing the central repository with the app user. As new data are stored in the repository, cloud services are run to process text, audio and haptic data offline. Since all multi-modal data are eventually time-aligned, processes can cooperate to make the analysis more robust. Results are stored back to the central repository, where they remain available for post-processing.

A simple interface is provided for clinicians and operators to manage customised protocols and user profiles, configure a screening session, personalise the text files (e.g reading texts, questionnaires etc.) required by a specific screening protocol. Recorded data, as well as the output of offline data processing, are eventually made available through a set of web services provided by the cloud server. Web services are accessed through personal authentication by means of username and password, to allow for the rapid prototyping of third-party applications for data visualisation, analysis and modelling, while complying with requirements for personal data protection.

3 Data Acquisition

3.1 Participants and protocol

73 children were selected from single classes in a primary school in Sofia, from grade 2 to 5. All classes share the same teachers, and follow the Bulgarian curriculum and state education system. All children whose parents gave their consent took part in the study, with no selection bias for students with better or worse reading skills. The vast majority of participants had normal vision, with a very small percentage of them having their vision corrected to normal. None of them had hearing problems, attention deficit hyperactivity disorder, dyslexia, dysgraphia, dystrophy, aphasia, autism spectrum disorder, cognitive impairment or memory impairment.

During a reading session, participants were sitting at a school desk, in front of a tablet in portrait position. For the present collection campaign, we used a 10.5 inch Samsung Galaxy TAB A8 (2.0 GHz Dual+1.8 GHz Hexa-core, 3 GB RAM, 32 GB eMMC, Android 11), with a 246.8 x 161.9 x 6.9 mm screen and a 1920x1200 pixels resolution and a 1.4 inch Samsung Galaxy TAB S6 Lite (2.3 GHz, 1.8 GHz, 4 GB RAM, 64 GB eMMC, Android 12), with a 244.5 x 154.3 x 7 mm screen and a 2000x1200 pixels resolution. The reading text was displayed in Arial font (21.25pt).

Each child participated in two consecutive reading sessions, one silent and one oral. In each session, the child was asked to read one of four short children's stories originally created for this project, consisting of 5 self-contained episodes organised according to levels of increasing reading complexity (see section 3.2). The length of each episode varied between 120 and 155 words (see Table 1), with all episodes fitting a single tablet page. Second graders read only the first two episodes of each story, third graders read the first three episodes, fourth graders read the first four episodes, and fifth graders read all five episodes. The order of the two reading conditions (oral and silent) was counterbalanced across participants, with no child reading the same story in both conditions.

Before starting a reading session, participants were instructed to use the tip of the index finger of their dominant hand for finger-point reading. A short excerpt from a Bulgarian translation of the *Pinocchio* novel was used as practice session. The session was repeated if the child finger-tracked less than 60% of the practice text. After reading each text episode, the child was asked two reading comprehension questions consisting of a question stem (i.e. the actual question) and four randomly shuffled answers, only one of which was correct. Due to the different number of episodes read by children in different grades, the number of questions ranged from a minimum of 4 (2^{nd} graders) to a maximum of 10 questions (5^{th} graders).

3.2 Texts for the experiments

Five original Italian texts, created for the specific purposes of the *ReadLet* project (Taxitari et al., 2021), were translated into Bulgarian. In Italian, the readability of each text was automatically controlled according to a data-driven methodology that evaluates the reading difficulty of a text as a machine-learning binary classification problem (Dell'Orletta et al., 2011).

The linguistic features used to predict readability are categorised into four main groups: raw text, lexical, morpho-syntactic and syntactic features. *Raw text features* include sentence length, calculated as the average number of words per sentence, and word length, calculated as the average number of characters per word. *Lexical features* refer to the internal composition of the vocabulary of the text. For Italian, two different features were determined by comparing a text with a reference resource con-

BULGARIAN	episode 1		episode 2		episode 3		episode 4		episode 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
word length [letters]	4.57	2.28	4.78	2.68	4.88	2.82	4.80	2.80	4.82	2.94
text length [words]	119.75	7.14	133.25	14.13	140.75	6.85	151.00	12.99	155.50	9.33
type-token ratio	0.76	0.04	0.72	0.04	0.74	0.04	0.75	0.02	0.72	0.05
lexical density	0.61	0.07	0.61	0.05	0.59	0.04	0.58	0.02	0.56	0.02
PoS type	11.75	0.50	11.25	0.50	12.00	0.00	11.75	0.50	11.75	0.50
IPU length [words]	6.30	0.88	5.93	0.58	6.75	0.72	6.78	0.82	6.38	0.47
sentence length [words]	8.70	0.27	11.59	1.15	14.07	0.68	18.92	1.09	20.82	1.99
dependency length [words]	1.73	1.42	2.03	2.09	2.22	2.22	2.30	2.45	2.59	2.94
word frequency[1]	5.42	1.32	5.27	1.46	5.18	1.51	5.31	1.47	5.31	1.47

ITALIAN	episode 1		episode 2		episode 3		episode 4		episode 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
word length [letters]	4.21	2.17	4.66	2.46	4.71	2.53	4.87	2.76	5.07	3.01
text length [words]	141.00	0.82	152.00	0.82	166.00	4.83	169.75	3.2	178.00	7.96
type-token ratio	0.71	0.03	0.72	0.03	0.72	0.01	0.73	0.01	0.73	0.03
lexical density	0.59	0.03	0.61	0.02	0.60	0.01	0.60	0.03	0.60	0.01
PoS type	11.00	0.82	10.75	1.26	12.00	1.15	11.75	0.96	11.25	0.96
IPU length [words]	8.24	0.83	6.70	0.36	7.74	1.87	8.95	1.68	7.51	0.85
sentence length [words]	10.07	0.06	13.24	0.62	16.6	0.48	21.36	1.92	23.85	2.35
dependency length [words]	1.57	1.39	1.78	1.79	1.93	2.04	2.06	2.43	2.33	2.97
word frequency[1]	5.76	1.27	5.45	1.47	5.46	1.49	5.41	1.51	5.38	1.53

Table 1: Descriptive statistics for Bulgarian and Italian reading texts by text episodes (IPU = Implicit Prosodic Unit, SD = standard deviation).

taining a basic vocabulary: a) the percentage of unique words in the text that are also included in the reference list (calculated per lemma); and b) the internal distribution into usage classification classes as very frequent words, frequent words and words with relatively low frequency that refer to everyday objects or actions and are therefore well known to speakers. *Morpho-syntactic features* refer to lexical density, which refers to the ratio of content words (verbs, nouns, adjectives and adverbs) to the total number of lexical tokens in a text. *Syntactic features* are numerous, including the depth of the dependency tree, the relative order of the subordinate clauses in relation to the main clause and the length of the dependency.

Likewise, the Bulgarian translation of the Italian reading texts was preprocessed using the Bulgarian Natural Language Processing pipeline, which orchestrates several natural language processing tools, including the Bulgarian language processing chain (BGLPC) and the Universal Dependencies parser. The Bulgarian language processing chain consists of a sentence splitter, a tokeniser, a Part-of-speech (POS) tagger, a lemmatiser, a noun phrase (NP) extractor, a named-entity recogniser and a stop-word recogniser. All tools are self-contained and designed to work in a pipeline; i.e., the output of the previous component is the input for the next component, starting with the sentence splitter, and followed by the tokeniser, POS tagger and lemmatiser (Karagiozov et al., 2011).

The current version of the Bulgarian language processing chain uses an improved version of its components enabling simultaneous segmentation of texts into single words and multiword expressions (MWEs) as well as simultaneous POS tagging and lemmatisation of individual words and MWEs (Koeva et al., 2020). Although the accuracy of POS tagging was improved only marginally compared to the accuracy before retraining (0.033%), the most important result is the simultaneous processing of single words and MWEs, which is also reflected in the improvement of the existing lemmatiser. Universal dependency parsing is carried out with the NLP-Cube framework in API mode (Boroș et al., 2018). A Python script was created to enable access to the NLP-Cube functionality, and automate the processing of the Bulgarian texts. For each text, the NLP-Cube annotation and the BGLPC annotation are synchronised token by token, and a correspondence map is created between identical tokens in both documents. Based on this synchronisation, Universal Dependency relations are transferred to the BGLPC CoNLL-U Plus output and the relation index is recalculated.

In order to replicate the methodology used for

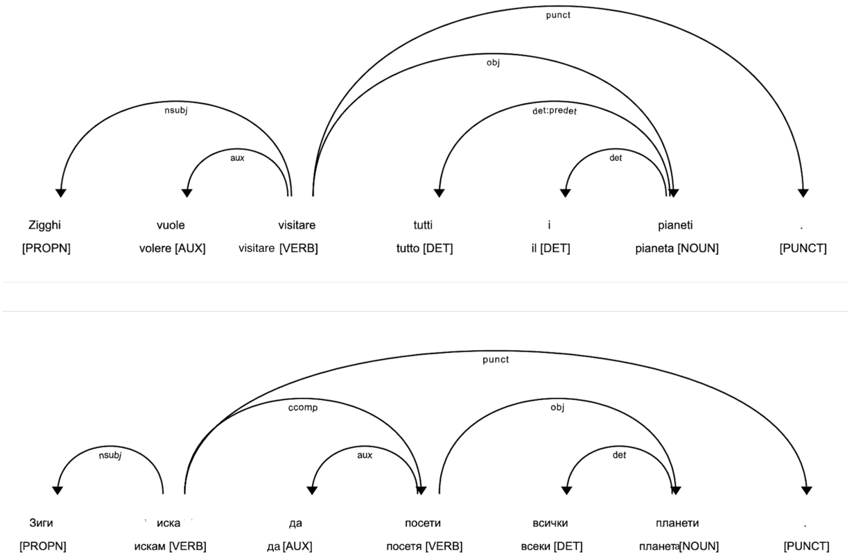

Figure 2: A Universal Dependency analysis of an Italian sentence (top) and its Bulgarian translation (bottom)

the Italian texts, the following principles and steps were applied to Bulgarian translations. First, lexical features in the Bulgarian texts were controlled in two ways: a) by comparing the vocabulary used in each translated text with the general Bulgarian lexis (Koeva and Doychev, 2022); and b) by calculating the *type/token ratio*, *i.e.* the ratio between the number of lexical *types* (i.e. lemmas) and the number of *tokens* (i.e. lexical forms) that actually occur in the texts. For example, the type/token ratio for the first episode of one of the Bulgarian texts is 0.67 (80 unique words and 119 tokens), while in Italian it is 0.61 (87 unique words and 141 tokens). Here, the general grammatical structure of the two languages (as well as the principles of tokenisation and lemmatisation) accounts for the difference between the ratios. Nevertheless, the overall lexical scores in the two languages are comparable.

To ensure a predominant usage of Bulgarian general lexis, we compared the vocabulary of each translated text with the vocabulary found in textbooks and dictionaries for the same age and educational level targeted by the reading texts. Some form of cultural adaptation of the target text was occasionally deemed necessary, as witnessed by some proper names in the Italian texts being replaced by more familiar Bulgarian names in the corresponding translation. For example, *Ivo* was used as a substitute for *Gianni*, and *Violetta* for *Viola*. To the extent possible, both number and type of sentences and clauses in the original Italian texts were kept in their Bulgarian translations. This is illustrated by the simple example in Figure 2, where two translationally-equivalent sentences are provided, together with their Universal Dependency annotations. The comparison of word and clause lengths in Bulgarian and Italian texts shows that the relatively small number of words in a sentence is maintained in Bulgarian, as is the tendency to use relatively short words with a relatively simple morphological structure. The slightly greater average word length in the Bulgarian texts is due to the morphological structure of nouns, adjectives and some pronouns and numerals, which incorporate definite articles. Likewise, the relatively lower average number of words in Bulgarian can generally be explained by Italian determiners being spelled as independent word tokens (Table 1).

Although the two languages present many morphological and syntactic similarities, there are also significant differences that manifest themselves in the grammatical structure. Overall, the Italian-Bulgarian translation required considerable effort to reproduce the meaning of the original text as faithfully as possible while maintaining the corresponding grammatical structure of the target language (Pirrelli and Koeva, 2024: 35-36).

The descriptive statistics of Table 1 show that all basic parameters of text complexity in the Bulgarian reading texts (letter length, morpho-syntactic

[1]Italian word frequency distributions are extracted from SUBTLEX-IT (Crepaldi et al., 2013). Bulgarian word frequency distributions are extracted from Bulgarian National Corpus, amounting in 1.2 Billion tokens (Koeva et al., 2012).

complexity, lexical density and syntactic complexity) increase with the text episodes. To illustrate: the average word length in letters increases from 4.57 to 4.82 in Bulgarian and from 4.21 to 5.07 in Italian, while the average syntactic dependency length (i.e. the number of tokens that can occur between a syntactic head and its dependent/modifying element in a reading text) increases from 1.73 to 2.59 in Bulgarian and from 1.57 to 2.33 in Italian. At the same time, as expected, the token frequency shows an inverse trend, ranging from a Zipf scale value of 5.42 to 5.31 for the Bulgarian episodes and from 5.76 to 5.38 for the Italian episodes.[2]

4 Data post-processing

The data collected with *ReadLet* include: speech recording, the tracking record of finger movements on the tablet touchscreen, the time taken to answer comprehension questions, and the answers provided by the reader. Data are then post-processed for them to be converted into time-series and then aligned with the text through the following steps.

The position of the text on the tablet touchscreen is encoded with the pixel coordinates (`x_min`, `y_min`, `x_max` `y_max`) of the *bounding box* of each text character on the screen (including blank spaces and punctuation marks).[3] In turn, continuous finger movements are discretized into *touchmove* events on the tablet touchscreen, with each *touchmove* event being associated with its time onset and its pixel coordinates on the screen. Text-coordinates and finger-coordinates are then aligned using a custom convolutional algorithm that finds the largest match between text and finger coordinates (Ferro et al., 2024). Finally, after finger-coordinates are matched with letter-coordinates, we compute the finger-tracking time of each letter in the text as the difference between the last time tick and the first time tick in the time series of touch events falling within a letter's bounding box. The finger-tracking time for a text unit containing more letters is computed as a summation of the tracking times of the letters the unit spans over.

At the moment of writing the paper, the audio-recordings of oral reading sessions are in the process of being automatically converted into text using Whisperx (Bain et al., 2023), a free open-source toolkit built on top of Whisper (Radford et al., 2022). For each spoken word, the toolkit outputs an alphabetic transcription and the associated confidence level, together with onset and offset time-points of the word's articulation. After this first processing step, a procedure aligning word transcriptions with the original text is executed using an alignment model (downloadable here), based on a version of `Wav2Vec2 XLS-R` (Babu et al., 2021) fine-tuned on Bulgarian speech data (downloadable here). This second step is taken to provide more reliable timestamps associated with the actual word in the original text. At the time of writing the paper, Bulgarian children's reading data are being post-processed for speech-to-text conversion. Thus, the present preliminary analyses are exclusively focused on finger-tracking data.

Original audio-recordings of reading sessions will not be made openly available. Nonetheless, we provide open-access information about the onset and offset time-points of a word's articulation, as computed by the speech-to-text conversion tool.

5 Data analysis

The original dataset was trimmed by excluding individual data points (word tokens) whose finger-tracking time was lower than 0.01 seconds or higher than 3.5 seconds. This procedure resulted in 4.4% of the original data being removed, corresponding to 2 subjects of the original set of participants. The resulting dataset was analysed with *R* using Generalised Additive Models (*GAMs*), using the package `gamm4`, version 0.2-6 (Wood, 2017).

To understand the factors affecting the pace of finger-point reading, we entered *token tracking time* (i.e. the time taken by the finger to underline an individual word token) as the dependent variable of two *GAM* models with the independent variables *grade level* (from 2^{nd} to 5^{th}) and *reading type* (aloud vs. silent) as categorical factors, and *word length* or *word frequency* as numeric predictors. Finally, to take into account the inter-individual variability in our sample and control for effects of lexical variability in our texts, *subjects* and *word tokens* were entered as random effects.

Results are plotted in Figure 3. Here, the box plot in the top panel shows that finger-tracking times are significantly shorter in silent reading than in the aloud reading condition for grades 3, 4 and 5

[2] As frequency measures for the two languages came from corpora of different size, raw counts were transformed using the *Zipf scale*: log_{10}(frequency per million words) + 3 (Van Heuven et al., 2014).

[3] `x_min` and `y_min` are coordinates of the top-left corner of the bounding box; `x_max` and `y_max` are coordinates of the bottom-right corner of the bounding box.

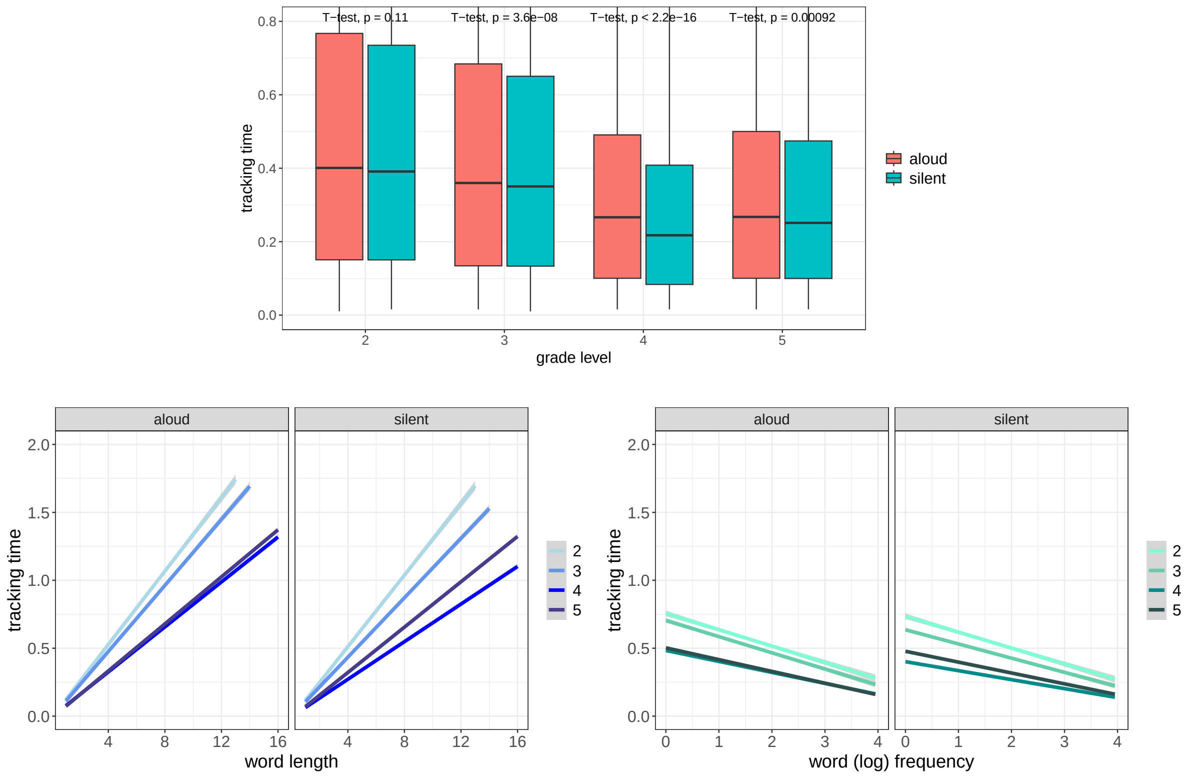

Figure 3: Upper panel: boxplot distributions of tracking time across grades and reading types. Lower panel: linear regression of word length (left) and frequency (right) on tracking time across grades and reading types. Shaded areas refer to 95% confidence intervals.

(p-value < 0.001). In addition, one can observe a decreasing trend of finger-tracking times for increasing grade levels (p-value $< 2e-16$), with a plateau between grade 4 and 5 for both reading conditions. We believe this levelling effect to be due to the greater complexity of the additional episodes read by 5th graders. Effects of word length (left bottom panel) and word frequency (right bottom panel) on tracking times exhibit a significant interaction with the grade level, with early graders (2nd and 3rd) being more affected by both word length and word frequency than late graders (4th and 5th) (p-values $< 2e-16$). The effect, also observed in Italian child's reading data collected with the same finger-tracking technology (Marzi et al., 2020; Ferro et al., 2024), shows that the orthographic lexicon of Bulgarian 4 and 5 graders contains not only more word tokens than the orthographic lexicon of 2 and 3 graders, but also distinctively longer and less frequent ones. As a result, the finger-tracking pace of Bulgarian 4 and 5 graders is less prominently slowed down by longer words (left panel) and less frequent words (right panel), than the pace of Bulgarian 2 and 3 graders is. The evidence confirms that even very young readers of a script such as the Bulgarian Cyrillic alphabet[4] tend to opt for a lexical reading strategy as early as possible for the sake of reading efficiency.

Finally, the two plots also show a significant interaction between word length or frequency and reading conditions, with finger-point reading being significantly more affected by both word length and word frequency in the aloud reading condition than in the silent one (p-values $< 2e-16$). Not only does this evidence suggest that longer and rarer words are more difficult to process and access. Also their articulation take longer to be planned and executed. Incidentally, this provides further, indirect evidence of the strong correlation between finger-tracking times and articulation times in finger-point reading. We expect this evidence to be confirmed by text-aligned and time-aligned speech-recognition data.

[4] Although there is no one-to-one correspondence between letters and sounds in either the Bulgarian or Italian alphabet, both scripts are much closer to this ideal condition than – say – in English or French. Accordingly, Bulgarian and Italian are classified as orthographically *transparent* languages, i.e. languages where a word pronunciation can largely be predicted from its spelling.

6 Discussion and outlook

In this study, we capitalised on the huge potential of mobile information technology, cloud computing and NLP for behavioural data collection and analysis, to investigate developmental trends in the reading data of Bulgarian early graders. Preliminary results significantly replicate benchmark effects attested in the reading literature on transparent scripts, and are in line with the finger-tracking data collected from Italian children with the same protocol (Ferro et al., 2024; Marzi et al., 2020). Overall, the data confirm that children confronted with a transparent script tend to resort to a lexical reading strategy as early as possible, since direct access to orthographic lexical information allows for a more fluent and efficient reading performance than a sublexical reading strategy.

There are several reasons to recommend an extensive usage of finger-tracking and NLP technologies for literacy research and education. First, the use of a simple tablet supports unobtrusive collection of multimodal reading data in ecological contexts. Following the Italian experience with the *Readlet* infrastructure, we were able to replicate, on a smaller scale, a relatively straightforward and efficient protocol for Bulgarian child's reading data collection at school. Pupils were extremely responsive to the task, and showed a great familiarity with using the tablet for reading. This made the process of data collection surprisingly quick and most effective, suggesting that finger-tracking can aptly be used for extensive reading assessment in primary schools. In particular, a tablet can be used as a reading book for large-scale studies, paving the way to generalisable results. In addition, the possibility to take repeated single-subject measurements makes finger-tracking evidence suitable not only for group assessment, but also for individual developmental studies.

On a more technical note, the use of a tablet as a reading book allowed us to collect finger-tracking and audio-recording streams concurrently, and take advantage of their being both text-aligned and time-aligned. In spite of recent progress in the accuracy of NLP technologies, they can occasionally be brittle and error-prone, particularly when confronted with real language data, which are collected in inherently noisy, ecological communication contexts.

The bonus of having multiply time-aligned multimodal data streams is that, in processing raw data, noise in one channel can be filtered out by integrating synchronous information coming from a less noisy channel. For example, the drift of a finger-tracking signal in a particular time window, can be corrected using the voice signal sampled and text-aligned in the same time window (Ferro et al., 2024). In fact, the latter can provide reliable information about which text line the reader is currently processing. This is expected to offer better finger-tracking data but also better transcribed spoken data, which, in turn, can be aligned more reliably both individually with the text being read, and with each other.

In the near future, we consider exploring several lines of research. First, we intend to investigate the complementary role of n-gram frequencies in affecting Bulgarian finger-point reading, which is likely to attest to a dynamic interaction of sublexical and lexical reading strategies through literacy development (Orsolini et al., 2006). In addition, based on speech-recognition data, we intend to compute the correlation between finger-tracking times and articulation timing in aloud reading, and assess the Implicit Prosody Hypothesis (Breen, 2014), i.e. the idea that, in silent reading, readers activate prosodic representations that are similar to those they would produce when reading the text aloud.

Another aspect of our present study worth exploring in the near future is the interactive dynamic between voice recording and finger-tracking data. A recent analysis of the correlation between voice articulation, finger-tracking patterns and eye-tracking patterns in adult reading (Nadalini et al., 2024) showed that, in finger-point reading, the finger is most often located few characters ahead of the voice, and that the finger's pace is a rhythmic proxy for a reader's articulation rate. In particular, the distance of the finger from the voice correlates with the distance of the eye from the voice (or eye-voice span: Inhoff et al., 2011), a measure of the capacity of a reader's phonological buffer and reading fluency (Laubrock and Kliegl, 2015; Silva et al., 2016).

Thus, one can reasonably expect that monitoring the development of the finger-voice span in early readers can provide evidence of more and less typical developmental patterns of reading skills in the first years of primary school, when reading difficulties are more critical but manifest less clearly. This will offer a suitable benchmark for continual assessment of reading proficiency.

Acknowledgments

The present study is carried out within the bilateral cooperation project *Assessing reading literacy and comprehension of early graders in Bulgaria and Italy* (2023 – 2025), supported by the National Research Council of Italy (CNR) and the Bulgarian Academy of Sciences. The paper's first author, Alessandro Lento, is a PhD student enrolled in the *National PhD in Artificial Intelligence*, XXXVII cycle, course on Health and Life sciences, organized by Università Campus Bio-Medico of Rome. All other authors contributed equally to the study.

References

Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Wu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. 2021. Xls-r: Self-supervised cross-lingual speech representation learning at scale.

Max Bain, Jaesung Huh, Tengda Han, and Andrew Zisserman. 2023. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. In *Proc. INTERSPEECH 2023*, pages 4489–4493.

Tiberiu Boroș, Stefan Daniel Dumitrescu, and Ruxandra Burtica. 2018. NLP-cube: End-to-end raw text processing with neural networks. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 171–179, Brussels, Belgium. Association for Computational Linguistics.

Mara Breen. 2014. Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2):37–50.

Jonathan Brennan. 2016. Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass*, 10(7):299–313.

Yekaterina Chzhen, Anna Gromada, Gwyther Rees, Jose Cuesta, and Zlata Bruckauf. 2018. An unfair start: Inequality in children's education in rich countries. Technical Report 15, UNICEF Office of Research, Innocenti, Florence.

Davide Crepaldi, Marcello Ferro, Claudia Marzi, Andrea Nadalini, Vito Pirrelli, and Loukia Taxitari. 2022. Finger movements and eye movements during adults' silent and oral reading. In *Developing Language and Literacy: Studies in Honor of Dorit Diskin Ravid*, pages 443–471. Springer International Publishing.

Davide Crepaldi, Emmanuel Keuleers, Pavel Mandera, and Michael Brysbaert. 2013. SUBTLEX-IT: A frequency list based on movie subtitles. (unpublished manuscript).

Maria De Luca, Maria Pontillo, Silvia Primativo, Donatella Spinelli, and Pierluigi Zoccolotti. 2013. The eye-voice lead during oral reading in developmental dyslexia. *Frontiers in human neuroscience*, 7:696.

Felice Dell'Orletta, 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*, pages 73–83, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Vera Demberg and Frank Keller. 2019. Cognitive models of syntax and sentence processing. *Human language: From genes and brains to behavior*, pages 293–312.

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

Marcello Ferro, Claudia Marzi, Andrea Nadalini, Loukia Taxitari, Alessandro Lento, and Vito Pirrelli. 2024. Readlet: a dataset for oral, visual and tactile text reading data of early and mature readers. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, Turin, Italy.

Uri Hasson, Giovanna Egidi, Marco Marelli, and Roel M Willems. 2018. Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, 180:135–157.

Albrecht W Inhoff, Matthew Solomon, Ralph Radach, and Bradley A Seymour. 2011. Temporal dynamics of the eye–voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*, 23(5):543–558.

Halszka Jarodzka and Saskia Brand-Gruwel. 2017. Tracking the reading eye: Towards a model of real-world reading.

Johanna K Kaakinen and Jukka Hyönä. 2008. Perspective-driven text comprehension. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(3):319–334.

Diman Karagiozov, Svetla Koeva, Maciej Ogrodniczuk, and Cristina Vertan. 2011. ATLAS – A Robust Multilingual Platform for the Web. In *Proceedings of the German Society for Computational Linguistics and Language Technology Conference (GSCL 2011)*, page 223–226, Hamburg, Germany. European Language Resources Association.

Svetla Koeva and Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the 5th*

International Conference on Computational Linguistics in Bulgaria (CLIB 2022), pages 203–213, Sofia, Bulgaria. Department of Computational Linguistics, IBL – BAS.

Svetla Koeva, 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*, pages 6988–6994, Marseille, France. European Language Resources Association.

Svetla Koeva, Ivelina Stoyanova, Svetlozara Leseva, Tsvetana Dimitrova, Rositsa Dekova, and Ekaterina Tarpomanova. 2012. The Bulgarian National Corpus: Theory and Practice in Corpus Design. *Journal of Language Modelling, 0, 1*, page 65 — 110.

Jochen Laubrock and Reinhold Kliegl. 2015. The eye-voice span during reading aloud. *Frontiers in psychology*, 6:1432.

Claudia Marzi, Anna Rodella, Andrea Nadalini, Loukia Taxitari, and Vito Pirrelli. 2020. Does finger-tracking point to child reading strategies? In *Proceedings of 7th Italian Conference on Computational Linguistics*, volume 2769, Bologna.

Heidi Anne E. Mesmer and Karen Lake. 2010. The role of syllable awareness and syllable-controlled text in the development of finger-point reading. *Reading Psychology*, 31(2):176–201.

Heidi Anne E Mesmer and Thomas O Williams. 2015. Examining the role of syllable awareness in a model of concept of word: Findings from preschoolers. *Reading Research Quarterly*, 50(4):483–497.

Andrea Nadalini, Claudia Marzi, Marcello Ferro, Loukia Taxitari, Alessandro Lento, Davide Crepaldi, and Vito Pirrelli. 2024. Eye-voice and finger-voice spans in adults' oral reading of connected texts. Implications for reading research and assessment. *The Mental Lexicon*.

Joseph F. T. Nese, Gina Biancarosa, Kelli D. Cummings, Patrick C. Kennedy, Julie Alonzo, and Gerald A. Tindal. 2013. In search of average growth: describing within-year oral reading fluency growth across grades 1-8. *Journal of school psychology*, 51 5:625–42.

OECD. 2023. Equity and inclusion in education: Finding strength through diversity. `https://doi.org/10.1787/e9072e21-en`, OECD Publishing, Paris.

Margherita Orsolini, Rachele Fanari, Valeria Tosi, Barbara De Nigris, and Roberto Carrieri. 2006. From phonological recoding to lexical reading: A longitudinal study on reading development in italian. *Language and cognitive processes*, 21(5):576–607.

John J. Pikulski and J. Chard David. 2011. Fluency: Bridge between decoding and reading comprehension. *The Reading Theacher*, 58:510–519.

Vito Pirrelli and Svetla Koeva. 2024. Developing Materials for Assessing Reading Literacy and Comprehension of early Graders in Bulgaria and Italy. *Foreign Language Teaching*, 51(1):29–37.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.

Keith Rayner. 2009. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology*, 62(8):1457–1506.

Susana Silva, Alexandra Reis, Luís Casaca, Karl M. Petersson, and Luís Faísca. 2016. When the Eyes no longer lead: Familiarity and Length Effects on Eye-Voice Span. *Frontiers in Psychology*, 7.

Loukia Taxitari, Claudia Cappa, Marcello Ferro, Claudia Marzi, Andrea Nadalini, and Vito Pirrelli. 2021. Using mobile technology for reading assessment. In *Proceedings of 6th IEEE Congress on Information Science & Technology (IEEE CiST'20)*, Agadir, Morocco.

Renáta Tichá, Christine A. Espin, and Miya Miura Wayman. 2009. Reading progress monitoring for secondary–school students: Reliability, validity, and sensitivity to growth of reading–aloud and maze–selection measures. *Learning Disabilities Research & Practice*, 24(3):132–142.

Joanna K Uhry. 1999. Invented spelling in kindergarten: The relationship with finger-point reading. *Reading and Writing*, 11:441–464.

Joanna K. Uhry. 2002. Finger-point reading in kindergarten: The role of phonemic awareness, one-to-one correspondence, and rapid serial naming. *Scientific Studies of Reading*, 6(4):319–342.

Walter JB Van Heuven, Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. Subtlex-uk: A new and improved word frequency database for british english. *Quarterly journal of experimental psychology*, 67(6):1176–1190.

Ludo Verhoeven and Charles Perfetti. 2008. Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(3):293–301.

Roel M Willems, editor. 2015. *Cognitive neuroscience of natural language use*. Cambridge University Press.

Simon N Wood. 2017. *Generalized additive models: an introduction with R*. Chapman and Hall.