INNOVATION IN LANGUAGE LEARNING International Conference

Automatic Generation of a Graded Reader in Old Church Slavonic

Iglika Nikolova-Stoupak[1], Gaël Lejeune[2], Eva Schaeffer-Lacroix[2], Aliona Shestakova-Stukun[3]

[1,2]Sorbonne Université, France
[3]Language School *Aspirantum*, Armenia
[1]iglika.nikolova-stoupak@etu.sorbonne-universite.fr
[2]{eva.lacroix, gael.lejeune}@sorbonne-universite.fr
[3]a.szestakowa@gmail.com

Abstract

Immersive reading, as well as graded readers as a case study of its application, have been highly valued within language education in the past few decades. Graded readers have so much as extended onto the so-called classical ('dead') languages, such as Latin and Greek. The reading of and listening to adapted texts in these languages has been shown to increase students' proficiency, independence and motivation when used either in isolation or combination with other teaching methods. However, as of now there is only a small number of related resources as well as of classical languages represented. The present study will investigate the current potential for (semi-)automatic generation of adapted classical-language readers while focusing on the Old Church Slavonic language. Old Church Slavonic is the earliest written Slavic language, which was to later branch into the modern-day West, East and South Slavic languages. It makes use of the Cyrillic and, more rarely and for earlier sources, the Glagolitic script. The language's interpretation is especially challenging due to the large geographical territory it encompassed, which was associated with diverse dialects. The following steps are taken in the framework of this project: 1) The linguistic characteristics of professional classical-language readers, such as the Latin Lingua latina per se illustrata *and the Greek* Athenaze*, are analysed (with a focus on representative readability-based characteristics). 2) Automatic generation of adapted Old Church Slavonic text is attempted through the use of GPT-5 (as per ChatGPT) in a one-shot setting. 3) The derived text's quality is assessed through both human evaluation and a comparison of its textual characteristics with those of professional texts as defined in point 1). The following Old Church Slavonic texts will be considered: the first chapter of the Biblical story of 'Genesis' and 'The Legend of Saint George and the Dragon'.*

Keywords: *classical languages, graded readers, large language models (LLMs)*

1. Introduction and Motivation

The immersive reading and listening of adapted texts in classical languages such as Latin and Ancient Greek (henceforth, Greek) has been shown to offer a number of benefits related to students' proficiency, independence and motivation when used either in isolation or combination with other methods, such as the question-answer method [3,11,16,18]. An increase in the number of relevant teaching materials that carry the established ones' qualities would therefore be of significant help in advancing the study of classical and other culturally-significant languages. In today's technological context, AI-based tools like ChatGPT are commonly used to facilitate the task of teaching professionals to create classroom and self-study materials, and this approach is to be applied in the current study. A discrete language will be experimented with, Old Church Slavonic (henceforth, OCS). On one hand, this language shares linguistic similarities with the classical languages in which established graded readers exist (in particular, Greek) as well as the common use for religious, notably Biblical, texts. On the other hand, OCS is significantly less resourced and less unified in terms of spelling and syntactic rules, thereby presenting a challenge. The deliverable OCS graded reader resources, consisting in two texts targeting two different proficiency levels, can be used by learners of the language, whether as part of a formal academic course or for independent study based on cultural, liturgical or academic interests.

2. Background

2.1 Graded Readers and Classical Languages

Related to the concept of comprehensible input as put forward by Krashen [8], graded readers of various modern languages have been in use for decades, typically associated with specific proficiency levels. The language within them is learner-friendly, containing accordingly simplified grammar and only a limited number of words that are complex for the level at hand. These reading materials have been particularly noted to help reaffirm vocabulary knowledge [19] and to increase students' motivation and sense of community [7]. Interestingly, albeit with certain modifications, the practice has extended to extinct, classical languages, such as Latin, whilst offering the same benefits as modern graded readers [3,11,16,18].

The classical language readers used as gold standard in the current project have been selected based on the presence of graded text (sometimes in addition to other learning materials), general quality/reputation and, ultimately, a quest for variety (in the face of different sizes, time frames of composition, and textual genres), so as to increase the robustness of the carried out analysis. The represented classical languages are Latin, Greek and Biblical Hebrew (henceforth, Hebrew). Each language is covered by two works: *Lingua latina per se illustrata* (henceforth, *LLPSI*) (Ørberg, 2011) and *Fabulae Faciles* (Ritchie, 1903); *Athenaze* (Balme and Lawall, 2003) and *Logos* (Martínez, 2023); *Graded Reader of Biblical Hebrew* (henceforth, *GRBH*) (Van Pelt and Pratico, 2006) and *Biblical Hebrew Easy Stories* (henceforth, *BHES*) (*Aleph with Beth*, 2020).[1] It is important to note that these languages differ significantly in relation to one another as well as to Old Church Slavonic. Hebrew comes as the clearest outlier, as it is the only language that does not issue from the Indo-European family. The other three investigated languages have closer alphabets as well as shared morphosyntactic features such as case systems.

2.2 The Old Church Slavonic Language

The name 'Old Church Slavonic is used to denote the language of the first Slavic manuscripts (10th-11th century AD) [10]. The language is characterised with around two centuries of use in a large geographical territory. The written system is credited almost exclusively to Constantine the Philosopher, a Thessaloniki-born scholar and monk. It initially made use of the Glagolitic alphabet, which then evolved into Cyrillic. The latter system is largely based on Greek letters as combined with additional symbols for typically Slavic sounds. There is only a limited number of established OCS manuscripts, which are considered to present the language's initial and unified characteristics. Extant OCS texts include Biblical translations, Saints' lives, prayers and sermons. OCS has strong word declension, which includes seven cases, three genders, three numbers, and three simple tenses. Typically, words come in sequences of open and closed syllables. The reduced vowels ѣ and ъ (respectively, strong and weak *yer*) are frequently used. Whilst the limited number of extant manuscripts clearly complicates the task of scholars of the OCS language, the most significant challenge comes in the presence of large variation within the language. Differences in dialects, encompassing vocabulary, spelling and grammar, progressively became significant since the establishment of the writing system, and no unified rules were ever explicitly laid out. Examples of dialect-based variation include the vocalisation of the open [e] sound and of nasal sounds and the use of uncontracted long adjectives [10]. Eventually, OCS gave place to what are now distinct 'Church Slavonic' languages, typical to the country or location in question.

The two texts that are going to be automatically adapted in the context of this project are the first chapter of the Biblical book of 'Genesis' (henceforth, 'Genesis: 1') and 'The Legend of Saint George and the Dragon' (henceforth, 'Saint George and the Dragon'). The former is the Biblical account of the creation of the world in six days. It is selected as the lower-level OCS text to be achieved due to its short length and relative simplicity. In contrast, 'Saint George and the Dragon', a hagiographic adventure story with a significant narrative line, is suitable to adapt into a higher-level text. The 'Genesis' text is reconstructed by Tomáš Spevák and follows the norms of the 9th-11th century OCS period. 'Saint George and the Dragon' is featured in a scholarly monograph by Alexander V. Rystenko, where multiple witnesses in both Greek and Slavic languages are featured and analysed. The language of writing is normalised based on later forms of Old Church Slavonic (as flowing into Russian Church Slavonic).

3. Methods

[1]For a detailed description of these resources, please refer to https://github.com/iglika88/automatically_generated_OCS_reader

3.1 Pre-processing

The primary sources were first preprocessed for use. All content apart from graded text in the target language (e.g. prefaces, vocabulary lists, grammatical sections) was discarded. Optical character recognition was applied for texts that were not already in a machine-readable format as per the proprietary tool 'Pen to Print' (selected due to its high-quality output in various languages and alphabets). To facilitate the forthcoming measurement of textual characteristics, each source was converted to full uninterrupted text, devoid of titles, tabs and new lines. Punctuation was standardised (e.g. the Greek interrogative ';' was replaced by '?'). Footnotes, notes within brackets and numerical annotations of lines and verses were also removed. The two volumes of *Athenaze* and *LLPSI* were merged into single texts.

3.2 One-Shot Prompting

OpenAI's popular chatbot ChatGPT was used for textual adaptation. Combined with relevant prompt engineering, the Large Language Model (LLM) is shown to provide promising results in the fields of textual simplification and summarisation, outperforming alternative models in terms of both automated scores and human preference [1,6,9], including in non-English settings [14, 17] and in relation to literary text [14]. For the purpose of this project, ChatGPT's most recent to-date version, GPT-5, is utilised through OpenAI's official interface. A new session is started between generation of the two texts in order to evaluate the model's performance in the OCS language based solely on its pretraining and a single provided text to adapt.

One-shot prompting is a setting in which, in addition to directions, the user provides the model with an example that illustrates the qualities that they would like its output to have. Previous research shows that in the presence of one-shot prompting, LLMs provide output sentences that resemble closer the various linguistic features of sentences that have been professionally crafted for the purpose of language teaching [15]. In relation to ChatGPT, a one-shot generation scenario has been shown to offer multilingual literary adaptations that share more closely the textual characteristics of human-made adaptations [14]. As no suitable original and learner-adapted textual pairs in OCS were discovered, the use of one-shot examples from another language was opted for. We resorted to the Latin language (as particularly high-resourced in the context of classical languages) as well as to *LLPSI* as a critically-acclaimed gold standard. Two stories at two different proficiency levels were taken to serve as examples for the two adapted OCS texts - the simpler one to be paired with 'Genesis: 1' and the more complicated one with 'Saint George and the Dragon'. *LLPSI* is a very exhaustive resource, starting from virtually no pre-required knowledge of the language and featuring mostly unadapted text in its second volume. The decision was therefore made to use one text from the middle and one from the end of volume 1 (respectively, 'Ch.18: Litterae latinea' and 'Ch.34: De arte poetica').

3.3 Evaluation

Firstly, the utilised professional classical-language graded readers were quantitatively analysed. For the purpose, shallow characteristics that have been established as highly relevant to readability were used [5]. They were selected to be highly language- and format-independent; to only rely on computational resources that are readily accessible; and to cover the general categories of 'length-based', 'vocabulary-related', 'syntax-related' and 'discourse-related' characteristics. Please see table 1 for an overview of the selected features. Content words were defined as those with part-of-speech (POS) tags NOUN, PROPN, VERB, ADJ and ADV; and function words - with AUX, ADP, DET, PRON, PART, CCONJ, SCONJ. 'Punctuation variety' was defined as the ratio of non-full-stop punctuation symbols over full stop symbols (or equivalent). This selection is not meant to be exhaustive in determining a text's characteristics; rather, it serves to provide a basis for analysis and comparison of various relevant aspects of the investigated texts. For lemmatisation and POS tagging, the open-source pipeline *UDPipe*[2] was employed via its web-based API.[3] The tool allows processing of all classical languages concerned by the current project (Latin, Hebrew, Greek and OCS), as models trained on the respective Universal Dependencies (UD) treebanks are available. Several problems arose with respect to a quantitative analysis of the primary texts' characteristics. Firstly, the texts come in significantly different lengths. Also, they do not all target the same level ranges. To make comparison between these readers and between them and the automatically generated OCS texts possible, an additional processing pipeline

[2] https://lindat.mff.cuni.cz/services/udpipe/

[3] the specific models used are: latin-perseus-ud-2.15-241121, ancient_greek-perseus-ud-2.15-241121, ancient_hebrew-ptnk-ud-2.15-241121 and later old_church_slavonic-proiel-2.15-241121

was elaborated, at the end of which relevant portions of the primary texts were extracted whose level is comparable to that of the derived OCS texts.

Type of textual feature	Selected features
Length-based	average number of letters per word average number of words per sentence
Vocabulary-related	word-based TTR lemma-based TTR
Syntax-related	average number of verbs per sentence average percentage of function words per sentence
Discourse-related	average number of pronouns per sentence punctuation variety

Table 1. Types of textual features selected for quantitative analysis.

[12] note that despite not being free of limitations, vocabulary size presents an efficient proxy for CEFR[4] level. In a later study, [13] go on to estimate the relative vocabulary knowledge (in terms of lemmas) of learners of different languages with previously determined proficiency levels. The ranges that they come up with, disregarding a clear outlier group of French learners in the UK, are the following for the first four levels: 894-1492 (A1), 1700-2237 (A2), 2194-3305 (B1), and 2450-4012 (B2). These conclusions were applied in the current project in order to approximate the level of each of the two *LLPSI* texts used as examples in the one-shot prompt to ChatGPT. The lemma-based vocabulary size (following rule-based preprocessing that eliminates punctuation, proper nouns, macron- or diacritic-based differences, and OCR-based errors) of the portion of the book all the way up to and including each text was calculated, and the resulting value is mapped to a proficiency level. In the case of the first extract, 1385 lemmas were determined, and of the second - 3212. These values lead to the unproblematic estimation of the respective CEFR levels as A1 and B1. Next, the portions of all readers pertaining to these levels' vocabulary ranges as determined by [13] were extracted. In cases where a reader's vocabulary does not reach the range associated with B1, only the A1 portion was extracted.[5] In cases where the whole range of a given level is not covered by the reader, an extract between the lower limit and the end of the reader was retained. [6] The situation is more complicated with respect to *GRBH*, as the reader contains only 847 unique words, which is below the lower limit for level A1. The reason is assumed to be the reliance of knowledge outside of the presented text, as admitted by the authors. Following the authors' estimation that the text commences at an established beginner level and eventually reaches an intermediate level, a decision was made for the text to be divided into three approximately equal parts, the first and last ones of which were labeled, respectively, as A1 and B1. The extracted portions of all texts, now labeled with an approximate CEFR level, were used in the quantitative textual analysis.

In turn, the output texts' qualitative evaluation is based on in-depth analysis, focused on the following textual aspects: understandability (level-appropriate vocabulary and grammar), correctness (lack of mistakes at the level of vocabulary, grammar, and punctuation), consistency (use of relevant verb tenses throughout the text as well as lack of large variation in terms of the language's features as bound by time period and geographical location), textual coherence (natural textual flow, easy anaphora resolution, lack of unnecessary redundancy) and aesthetic appeal (a more subjective measure involving the text's overall literary quality, length and register). On the basis of this analysis, the text was manually improved so as to correct errors, remove inconsistencies and increase understandability.

4. Results[7]

4.1 Quantitative

Please refer to table 2 for the detailed results of all texts' quantitative evaluation.

[4]Common European Framework of Reference for Languages
[5]This is the case with *Fabulae Facilis* (1739 words) and *BHES* (1423 words).
[6]*Athenaze*'s vocabulary comes at 2535, which is below the upper limit of level B1; it is therefore the portion of the reader between vocabulary size 2194 and its end that is taken.
[7]For the full OCS texts as output by ChatGPT as well as manually corrected, please refer to: https://github.com/iglika88/automatically_generated_OCS_reader

Text	*LLPSI*	*LLPSI*	*Fabulae Faciles*	*Athenaze*	*Athenaze*	*Logos*	*Logos*	*GRBH*	*GRBH*	*BHES*	"Genesis: 1"	"Saint George and the Dragon"
Level	A1	B1	A1	A1	B1	A1	B1	A1	B1	A1	A1	B1
Language	Latin	Latin	Latin	Greek	Greek	Greek	Greek	Hebrew	Hebrew	Hebrew	OCS	OCS
Avg letters/word	5.23	5.35	5.66	4.78	5.20	4.58	4.97	4.42	4.53	4.37	*3.70*	4.37
Avg words/s-ce	12.51	14.49	18.68	35.01	25.22	12.53	17.42	13.43	13.70	9.08	*7.54*	*9.00*
TTR (words)	0.36	0.42	0.40	0.40	0.52	0.33	0.40	0.50	0.57	0.28	0.42	*0.69*
TTR (lemmas)	0.21	0.25	0.27	0.30	0.40	0.25	0.30	0.43	0.50	0.20	0.37	*0.57*
Avg verbs/s-ce	1.43	1.84	3.19	4.08	4.20	1.33	2.29	2.48	2.10	2.46	*1.23*	1.65
Avg % funct. words/s-ce	27.89	30.38	34.54	33.11	33.11	31.31	33.78	30.86	22.34	32.50	33.51	33.49
Avg pronouns/s-ce	0.44	0.59	0.75	1.46	1.65	0.63	0.84	1.26	1.16	1.29	*0.35*	0.62
Punctuation variety	3.02	3.23	2.02	6.66	3.19	3.71	3.46	2.83	2.95	2.13	*0.97*	*1.65*
Total s-ces	905	1317	366	198	107	458	487	92	98	728	40	29

Table 2. Statistics pertaining to the professional texts and the automatically generated OCS texts. The values for the OCS texts are given in *italics* when they fall outside of the range established by the professional texts for the variable.

Firstly, we may focus on the values exhibited per category for the professional texts, regardless of language and level, and note any deviations in ChatGPT's OCS output (i.e. any values that fall outside of the range observed for the variable in question). The 'number of letters per word' comes as too low for 'Genesis: 1' (3.7) when compared to the rest of the texts. In terms of the other length-based variable, 'number of words per sentence', both OCS texts' values are lower than the established standard. The type-to-token ratio (TRR) (for both words and lemmas) is higher than the established range for 'Saint George and the Dragon', speaking of high lexical variety within the text. The gap between the two types of TTR is also noted for each text, as it concerns the importance of inflection in lexical variety. In this aspect, the two OCS texts do not deviate from the standard. In terms of the 'number of verbs per sentence', 'Genesis: 1' once again demonstrates a reduced value, speaking of syntactic simplicity. The number of pronouns per sentence is also smaller than the established range in this text, thereby reducing the need for anaphora interpretation. Finally, punctuation variety comes as too small for both automatically generated texts, showing reduced variety in sentence types.

What follows are observations concerning the relationship between the different variables' values at the A1 vs. B1 proficiency level in the context of the same graded reader, where applicable. The average number of both 'letters per word' and 'words per sentence' tend to be higher for the higher proficiency level (with the exception of *Athenaze* in the case of the latter), and the two OCS texts fall neatly within this trend. The same can be said about the two types of TTR as well as the numbers of verbs and pronouns per sentence (in both cases, *GRBH* is an outlier at exhibiting smaller values in relation to B1 text). Concerning 'percentage of function words' and 'punctuation variety', no clear trends are established in relation to the professional texts.

Given the corpus size, only limited conclusions can be reached regarding tendencies related to the investigated variables per language. Hebrew is associated with the smallest 'number of letters per word,' which is natural due to the language's *abjad* (consonant-based) writing system. The gap between word- and lemma-based TTR is highest for Latin texts (the former value being larger by 0.13-0.17). As there is no gold standard in relation to OCS, any hypotheses concerning the language's specificity (such as a smaller number of letters per word compared to other classical languages) would need to be validated through statistical analysis based on professionally-crafted texts.

4.2 Qualitative

The language in 'Genesis: 1' strikes as simple and beginner-friendly. The sentences are short and the text makes effective use of the repetitive structure present in the original. The verses are numbered and, additionally, a line is skipped following each narrative unit. Still, there are some verses whose complexity is high due to a close reliance on the original language. For instance, in (21) 'И сътвори б҃ъ чловѣка: мѫжа и женѹ сътвори и')[8], the use of *и* as both a conjunction and a personal pronoun is likely to confuse a beginner learner. Similarly, anaphora resolution may be difficult in the following verb-less construction: (23) 'И виде б҃ъ вся, ѩже сътвори. И се, добро зѣло'[9]. The letters *оу* and *ѹ*, which are equivalent in representing the sound [u], are both present in the text, whilst in the original text only one

[8]'And God created the human: man and woman he created them'
[9]'And God saw everything that He created. And it [was] very good'

alternative is opted for. The same goes for the letters *ꙁ* and *ꙁ́* ([z]). Beginner students may get the wrong impression that there are specific reasons behind the choice of each of the letters in a pair.

Occasionally, the adapted text's spelling speaks of a later variety of OCS than can be assumed from the original text. For instance, the letter *е* sometimes replaces *ѥ* and *ь* (e.g. *единъ* instead of *ѥдинъ*; *посредь* instead of *по срьдь*). We consider that it would be more fitting to make use of the language's earlier spelling conventions as laid out in established textbooks and to seek uniformity where possible. Also, prepositions and pronouns are occasionally not separated from the word they modify (e.g. *съберетъсѧ*; the original reads *съберетъ сѧ*). Whilst such spelling is not impossible in the language, we judge that consistent separation of these words would increase readability and help students achieve notions of word formation.

Gradually moving from the voicing of preferences to the detection of actual mistakes, we note the use of the letter *я* (e.g. *вся*), which is not typical to OCS but to later Slavic languages. Also, accents are sometimes placed above vowels (e.g. *что́*) in an unpredictable way, possibly as a result of interference from the Greek language. The Russian word *меншее* is used in place of the appropriate *мьнѥѥ*, as found in the original. Another possible interference from Russian is the word *летаютъ*, whose spelling should be *летаѭтъ*. Mistakes in word declension have also been noticed: (15) the verb *изведѫтъ* should be *изведѫтъ*.[10] There are several nouns whose case has been mistaken (or, alternatively, their declension type), e.g. (9) *воды* (correct: *водъ*); (9) *морѧ* (correct: *морꙗ*); (11) *землѧ* (correct: *землꙗ*); (13) *днѫ* (correct: *дни*); (15) *птици* (correct: *пътицѧ*). Finally, there is a spelling mistake in the word *четвѫртыи* (correct: *четврьтыи*). Please see Fig. 1 for an extract of the output text as juxtaposed to its manually corrected version.

Original	12 И рече ѣъ: да бѫджтъ свѣтила на небеси, свѣтити *землѫ*. 13 И сътвори ѣъ два свѣтила: свѣтило *вєлико* дн8, и свѣтило меншее нощи, и *звѣзды*. 14 И виде ѣъ, ꙗко добро. И бѣ вечерь и бѣ *оутро*, дьнь четв8рты́и.
Corrected	12 И рече ѣъ: да бѫджтъ свѣтила на небеси, свѣтити ꙁемлѫ. 13 И сътвори ѣъ два свѣтила: свѣтило велико дни, и свѣтило мьнѥѥ нощи, и ꙁвѣзды. 14 И виде ѣъ, ꙗко добро. И бѣ вечерь и бѣ 8тро, дьнь четвертыи.

Figure 1. 'Genesis: 1': an extract of original vs. corrected output. Legend: underline = mistake; *italics* = stylistic choice (the two categories are not always unequivocally distinguishable).

The adapted story 'Saint George and the Dragon' is given the title 'Чюдо сѣаго Гєwргіѧ' ('Saint George's Miracle') and it, too, clearly shows qualities of text suitable for learners, including short and simple sentences and clearly distinguishable dialogue. In our opinion, the rendition may be a little too short, thereby limiting the story's action and losing some of its aesthetic appeal. Influences of later Slavic languages are more frequent than in 'Genesis: 1'. Apart from the letter *я* (*змия*), the letters *ё* and *й*, non-existent in traditional OCS, appear (*нёмъ*; *твой*). Once again, *е* is found to occasionally replace the traditional *ь* (*наконецъ*). The letters *ъ* and *ь* are sometimes confused, such as in the word *царь*. Shorter verb conjugations of a more recent nature are also opted for, particularly in the case of imperfect forms: *живьше*; *метаху* (in place of *живьаше*; *метааху*). We consider that a learner would benefit from use of the established imperfect forms, both because they are laid out as correct in contemporary OCS study materials and because they are more easily recognisable due to the distinctive presence of two adjacent vowels.

An inconsistency comes in the face of the spelling of *гражане*, which is neither South nor East Slavic in nature, as the former would call for the presence of *жд* instead of *ж*, and in the latter the root would be *город* rather than *град*. A declension mistake is found in the adjective *лютъ*, which should be *люто* in agreement with the neuter noun, *змиище*. We also found two cases of wrong word choice. Saint

[10] 3rd person plural, aorist; correct in the original

George addresses the princess as *господже* ('Madam'), whilst appropriate words for a young unmarried woman would be *дѣвица* and *отроковица*. Also, after the hero has slayed the dragon, he is said to be 'leading' (*веди*) it to the city. A more suitable word with similar spelling would be *влеци*, 'to drag'. Please see Fig. 2 for an extract of the output text as juxtaposed to its manually corrected version.

Original	Бысть градъ Ласиа, и в нёмъ царъ именемъ Соломонь. Близъ града *бѣше* озеро велико, и в томъ озерѣ *живѣше* змиище лютъ.
Corrected	Бысть градъ Ласиа, и в нѣмъ царь именемъ Соломонь. Близъ града бѣаше озеро велико, и в томъ озерѣ живѣаше змиище люто.

Figure 2. 'Saint George and the Dragon': an extract of original vs. corrected output.

5. Discussion

The output texts showcase GPT-5's language-independent ability to generate textual adaptations for learners in a one-shot setting: they feature short sentences, simple grammar and, where applicable, dialogue. Importantly, the LLM also demonstrates proficiency in relation to the OCS language, proving its already existent notions in the language. Words and declensions not present in the provided unadapted texts are used. Correct links are made between alternative symbols, such as *ѣ* and *ѣ*. Still, the output text is not mistake-free, in particular in relation to grammatical cases and word choice when it comes to more complex vocabulary. There is a perceptible tendency for later versions of OCS or even modern Slavic languages to interfere with the output, which can be seen as a natural consequence of these languages' (in particular, Russian's) significantly higher resourcedness.

On the other hand, we also noted a degree of dependence of the output on the original unadapted text provided. A text that uses later language conventions resulted in higher interference with modern languages. Also, when a rarely used letter, such as *оу*, was present in the original text, it also appeared in the generated one. It is due to this perceived reliance on the associated source text that we deem that it would be non-optimal to develop rules for automatic correction and standardisation of the issuing text. Instead, we manually revised the two stories.

The undergone quantitative analysis showed that the difference between the generated texts in terms of level matches all trends established in relation to professional texts. The perceived 'lack of action' in the more complex story, 'Saint George and the Dragon', can be confirmed by the fact that, although the 'number of verbs' does not fall outside of the overall range established by the original texts, the feature's value is lower than the lowest value observed at level B1. Several additional deviations from the ranges observed in the gold standard texts may speak of excessive simplicity, especially in relation to 'Genesis: 1'. Conversely, characteristics such as the lack of variety in terms of sentence mood and construction as well as high lexical variety (in the case of 'Saint George and the Dragon') may increase the texts' difficulty.

6. Conclusion and Future Work

The two adapted OCS texts generated by ChatGPT exhibit minor drawbacks, and the work of language professionals (mostly related to language standardisation) is indispensable in order for them to be made suitable for learners. Still, the economy in time and effort granted by the automatic intervention is considerable. Future plans related to the project include experimentation with other models and generation methods. In particular, an LLM may be finetuned with a large amount of OCS text [2,4], and the results of quantitative and qualitative analysis of the issuing text may be compared to the current ones. Given the scarcity of original OCS texts, it would be highly beneficial to also experiment with a setting, in which text in another language is used as source.

7. Limitations

When dealing with human languages, especially those that are no longer spoken, it is important to acknowledge their role as cultural heritage, as well as to note that LLM-issuing text has no 'authenticity' and, therefore, not the same value as an original one. Only two primary OCS texts are used in the project, and they are reconstructed. Therefore, whilst they are a suitable choice in view of the generation of a language reader, they do not represent the original language in its fullness. Also, as the diversity of the classical language resources used in the project demonstrates, there is no clear formula of what

makes up a good graded reader. For instance, it is subject to one's opinion whether it is best to present students with adapted text in isolation or to add explanatory notes, translations and exercises. The limited number of classical readers available further complicates the tasks of defining their features and evaluating their relevance.

REFERENCES

[1] Bogireddy, S. R., & Dasari, N. (2024). Comparative analysis of ChatGPT-4 and LLaMA: Performance evaluation on text summarization, data analysis, and question answering. In *Proceedings of the 15th International Conference on Computing Communication and Networking Technologies (ICCCNT 2024)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICCCNT61001.2024.10725662
[2] Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Fiedel, N., *et al.* (2022). PaLM: Scaling language modeling with Pathways. *arXiv*. https://arxiv.org/abs/2204.02311
[3] Diller, K. C., & Walsh, T. M. (1978). "Living" and "dead" languages: A neurolinguistic distinction. In J.-G. Savard & L. Laforge (Eds.), *Actes du 5e congrès de l'Association Internationale de Linguistique Appliquée*. Montréal: Les Presses de l'Université Laval.
[4] Dodge, J., Ilharco, G., Schwartz, R., Farhadi, A., Hajishirzi, H., & Smith, N. A. (2020). Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. *arXiv*. https://arxiv.org/abs/2002.06305
[5] DuBay, W. H. (2007). *The classic readability studies*. ERIC Clearinghouse.
[6] Engelmann, B., Haak, F., Kreutz, C. K., Nikzad Khasmakhi, N., & Schaer, P. (2023). Text simplification of scientific texts for non-expert readers. *arXiv*. https://arxiv.org/abs/2307.03569
[7] Hill, D. R. (2013, January). Graded readers. *ELT Journal*, 67(1), 85–125. https://doi.org/10.1093/elt/ccs067
[8] Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press. http://www.sdkrashen.com/content/books/principles_and_practice.pdf
[9] Leroy, G., Kauchak, D., Harber, P., Pal, A., & Shukla, A. (2024, May). Text and audio simplification: Human vs. ChatGPT. *AMIA Joint Summits on Translational Science Proceedings*, 2024, 295–304.
[10] Lunt, H. G. (2001). *Old Church Slavonic grammar* (7th ed.). Berlin: Mouton de Gruyter.
[11] McMenamin, C. (2022). Greek club: Resurrecting dead languages in secondary schools. *Journal of Classics Teaching*, 23(46), 121–123. https://doi.org/10.1017/S2058631022000058
[12] Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters. https://doi.org/10.21832/9781847692900-007
[13] Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for Languages. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 194–211). London: Palgrave Macmillan. https://link.springer.com/chapter/10.1057/9780230242258_12
[14] Nikolova-Stoupak, I., Lejeune, G., & Schaeffer-Lacroix, E. (2024). Contemporary LLMs and literary abridgement: An analytical inquiry. In *Proceedings of the Sixth International Conference on Computational Linguistics in Bulgaria (CLIB 2024)* (pp. 39–57). Sofia: Institute for Bulgarian Language, Bulgarian Academy of Sciences.
[15] Nikolova-Stoupak, I., Bibauw, S., Dumont, A., Stas, F., Watrin, P., & François, T. (2024). LLM-generated contexts to practice specialised vocabulary: Corpus presentation and comparison. In *Actes de la 31e conférence sur le Traitement Automatique des Langues Naturelles* (pp. 472–498). Toulouse: ATALA & AFPC. https://aclanthology.org/2024.jeptalnrecital-taln.33/
[16] Philips, F. C. (1988). The language laboratory and the teaching of "dead" languages. *The Classical World*, 82(2), 105–108. https://doi.org/10.2307/4350305
[17] Pu, X., Gao, M., & Wan, X. (2023). Summarization is (almost) dead. *arXiv*, abs/2309.09558. https://api.semanticscholar.org/CorpusID:262044218
[18] Venditti, E. (2021). Using comprehensible input in the Latin classroom to enhance language proficiency. *Journal of Classics Teaching*, 22(43), 22–28. https://doi.org/10.1017/S2058631021000039
[19] Wan-a-rom, U. (2008). Comparing the vocabulary of different graded-reading schemes. *Reading in a Foreign Language*, 20(1), 43–69.