

Extended Context at the Introduction of Complex Vocabulary in Abridged Literary Texts

Iglika Nikolova-Stoupak

Eva Schaeffer-Lacroix

Gaël Lejeune

Sens Texte Informatique Histoire, Sorbonne Université, Paris, France

iglika.nikolova-stoupak@etu.sorbonne-universite.fr,

{eva.lacroix, gael.lejeune}@sorbonne-universite.fr

Abstract

Psycholinguistics speaks of a fine-tuning process used by parents as they address children, in which complex vocabulary is introduced with additional context (Leung et al., 2021). This somewhat counterintuitive lengthening of text in order to aid one's interlocutor in the process of language acquisition also comes in accord with Harris (1988)'s notion that for every complex sentence, there is an equivalent longer (non-contracted) yet simpler one that contains the same amount of information. Within the proposed work, a corpus of eight renowned literary works (e.g. *Alice's Adventures in Wonderland*, *The Adventures of Tom Sawyer*, *Les Misérables*) in four distinct languages (English, French, Russian and Spanish) is gathered: both the original (or translated) versions and up to four abridged versions for various audiences (e.g. children of a defined age or foreign language learners of a defined level) are present. The contexts of the first appearance of complex words (as determined based on word frequency) in pairs of original and abridged works are compared, and the cases in which the abridged texts offer longer context are investigated. The discovered transformations are consequently classified into three separate categories: addition of vocabulary items from the same lexical field as the complex word, simplification of grammar and insertion of a definition. Context extensions are then statistically analysed as associated with different languages and reader audiences.

Keywords: corpus linguistics, literary abridgment, language acquisition, introduction of complex vocabulary

1 Introduction

The Cambridge Dictionary defines thus the verb 'to abridge': "to make a book, play, or piece

of writing shorter by removing details and information that is not important". Similarly, the Collins dictionary claims that 'to abridge' means "to reduce the length of (a written work) by condensing or rewriting". Despite this focus on textual length, however, it is logical to assume, in light of linguistic and psycholinguistic theory, that a mere reduction in size would be insufficient for said transformation to act as a 'bridge' (as its name implies) for an audience that would otherwise have difficulty in accessing a given text. This work, which builds upon Nikolova-Stoupak (2023)'s related findings, deliberately seeks and systematises instances where abridged literary works provide longer context than their original counterparts; specifically, when complex vocabulary is introduced. Particular contributions in the process include the composition of a larger corpus of full and abridged versions of famous literary works (including a significant variety of abridged versions in terms of audience, such as readers with comprehension difficulties and foreign language learners of several languages and levels), the exploration of both nouns and verbs as complex words on the basis of frequency lists, the inclusion of contexts that do not necessarily relate to identical parts of the examined texts and an elaborate pipeline that allows for the extraction and analysis of relevant textual samples.

The two main research questions investigated within the present work are: 1. Does complex vocabulary appear in a longer context (sentence) within an abridged rather than full literary text more frequently than random vocabulary? 2. When present, can the mentioned lengthening of context in an abridged text be further systematised with the help of specific examples and relevant linguistic theory?

2 Background

2.1 Literary Abridgement: Motivation

In a detailed study of 2006, the United Kingdom's *National Literacy Trust* notes that the practice of reading for pleasure has been diminishing over time, much to the detriment of children and adolescents' academic and even social performance (National Literacy Trust, 2006). A direct link between the reading of novels and academic results is established by Moje et al. (2008), who go on to note that motivation to read is triggered by the reading process itself. This vicious circle has also been referred to as the 'Matthew effect', according to which the gap between 'good' and 'bad' readers becomes larger as the latter's difficulty to comprehend reading materials leads to decreased interest in reading (National Literacy Trust, 2006). This is where literary abridgement comes into play and, although it has been claimed that the process compromises an original text's aesthetic or cultural value (Charyulu, 2018), there is significant evidence that the increased comprehensibility of abridged texts helps boost readers' confidence, motivation and ability to enjoy the process of reading (Arias Rodríguez and Roberto Flórez, 2018).

2.2 Linguistic Theory

There is evidence in support of the need and practice of expansion as a result of language simplification. The concept of 'baby talk' as well as the related one of 'foreigner talk' have an established presence in linguistic theory. Ferguson (1975) notes that these two modes of address to non-proficient interlocutors are characterised with an array of similarities and differences and that, having their own sets of 'rules', they are more complex than an intuitive imitation of the way the implied interlocutors (young children or foreign language learners) speak. After establishing that the mentioned linguistic varieties are language-specific, he goes on to gather a corpus of English foreigner talk as conceived by a group of students. The issuing language is derived through English based on a series of alterations, including replacements, omissions and, notably, expansion. The last includes, for instance, multiple negation and the inclusion of the subject 'you' in imperatives. Some reported lexical substitutions, such as 'yesterday'

becoming 'the day before', also lead to textual expansion.

A recent study by Leung et al. (2021) specifically addresses the process of introduction of new, complex vocabulary (such as by parents to their young children) as involving fine-tuning, in which the target words are presented along with complementary information that provides cues about their meaning (e.g. 'the dotted leopard' or 'the leopard ran like a cat' rather than the word 'leopard' in isolation). de Bot et al. (1997) seek to define a model for second language (L2) vocabulary acquisition. In particular, the ways in which learners make out the meaning of unknown words during the process of reading are discussed. In first place, the reader establishes the relative importance of the word. Then, they may directly use a dictionary to determine its meaning or rely on inference, which is most often based on grammatical knowledge, word morphology and world knowledge. The understanding and use of vocabulary are established to be inseparable from these of grammar. Ibrahim (2013) points out that lexical complexity may be voluntary or inherent to a language, whilst the complexity of accompanying syntax needs to be deliberately selected accordingly. Christophe et al. (2010) also imply the importance of balance between lexical and syntactic complexity for language acquisition at stating quasi-paradoxically that knowledge of different elements of language is both mutually helpful and inherently mandatory in order for additional knowledge to be achieved.

Harris (1988)'s linguistic theory provides a detailed overview of the way language is composed, with a particular focus on its goal of transmitting information. Discussing grammar, he points out that every complex construction (typically, sentence) can also be expressed in simple terms whilst carrying the same amount of information. The two implied sentences are referred to as having 'descriptive equivalence'. Contractions, i.e. the skipping of inferrable parts of a sentence, cause ambiguity, whilst extension and redundancy facilitate comprehension. Harris (1988) assumes that different languages are associated with different types and frequencies of grammatical contractions.

3 Methods

3.1 Corpus Generation

The gathered corpus consists of full and abridged versions of eight famous literary texts with high involvement in school curricula: *Alice's Adventures in Wonderland* (Lewis Carroll), *A Christmas Carol* (Charles Dickens), *Don Quixote* (Miguel de Cervantes), *Les Misérables* (Victor Hugo), *Madame Bovary* (Gustave Flaubert), *Romeo and Juliet* (William Shakespeare), *Oliver Twist* (Charles Dickens) and *The Adventures of Tom Sawyer* (Mark Twain). The languages covered are English, French, Russian and Spanish. Up to four abridged versions per work and per language are included (based on budget and availability) as well as the original work or up to three translations in the cases of languages different from the original one of composition. Almost all texts are present in every language and in at least one full and one abridged version¹. The main source for full versions is Project Gutenberg², whilst abridged works come from a large variety of sources, mostly as distributed for free. Please refer to Appendix A for more detailed corpus characteristics.

All works were converted into plain text format. In cases where character recognition was necessary, it was run through the Adobe Acrobat software. Texts underwent minor pre-processing, including the removal of metatext (including prefaces), of non-alphanumeric symbols and of readily fixible mistakes resulting from character recognition (such as the presence of '_' instead of space between words).

The corpus of source texts gathered and used for the purpose of this study has been made publicly available by the authors.³

3.2 Context Extraction

The texts in the corpus were grouped into 72 full-abridged pairs, where each pair consisted of one full and one abridged version of the same text in the same language.

Due to the Russian language's strong morphology, Russian texts were firstly lemmatised.

¹exceptions: *A Christmas Carol*, *Les Misérables* and *Don Quixote* are not present in Russian; *Madame Bovary* is not present in English

²<https://www.gutenberg.org/>

³https://github.com/iglika88/corpus_original_and_abridged_texts/

All texts underwent part-of-speech (POS) tagging via Python's *spacy* library, following which lists of the nouns and verbs (selected due to carrying the key meaning within a sentence unit) in each text were composed. Then, lists of the intersecting nouns and verbs were extracted for each full-abridged textual pair and they were compared against frequency lists in the respective languages, each of which consisting of 5000 words. The English⁴ frequency list used is fittingly based on Project Gutenberg. Based on availability and size as well as the absence of a narrow domain focus, the French⁵ and Spanish⁶ frequency lists are based on Open Subtitles, and the Russian one⁷ – on the Russian National Corpus. All corpora were taken from the *Wiktionary* website. The Russian frequency list contains lemmas, whilst all others contain words as verbatim found in the text.

At this point, the lists of nouns and verbs associated with each full-abridged textual pair contained solely non-frequent (which for the purpose of the experiment is taken to be closely identical to 'complex') words found in both texts. The next step was to extract sentences in which the complex words appear for the first time in each of the two texts and to compare their length. Samples of ten parallel sentences per full-abridged pair were extracted and compared, and the pairs of sentences for which the one associated with the abridged text was longer were counted as well as set apart for further, qualitative analysis.

An additional experiment was carried out in order to test the hypothesis that complex words are more likely to appear within longer context in abridged texts than random words. For the purpose, before the derived lists of nouns and verbs that appear in each textual pair were further filtered based on frequency lists, an additional sample of parallel sentences was extracted from both texts, and the numbers of instances in which an abridged text provided a longer sentence were once again counted.

⁴https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/English/Project_Gutenberg

⁵https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/French_wordlist_opensubtitles_5000_

⁶https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/Spanish/Subtitles10K

⁷O. N. Lyashevskaya and S. A. Sharov

4 Results

A total of 720 sentence pairs per scenario (i.e. in relation to random versus complex nouns and verbs) were extracted from the corpus, ten per each full-abridged textual pair of the same literary work in the same language. There were 221 cases of longer sentences pertaining to the abridged text within the random sample (i.e. 30.69%) and 222 cases when only complex words were considered (30.83%). Due to the closeness of results, it is concluded that the null hypothesis, namely that abridged versions are as likely to involve longer context when they contain complex words, could not be rejected at this point.

When all sentence pairs related to complex words where the abridged sentence was longer were manually investigated, only 64 of them (28.83%) were discovered to factually conform to the presence of additional context associated with the vocabulary item in question. In the rest of the cases, the added length was attributable to textual processing (e.g. a punctuation sign being erroneously marked as a noun or verb during POS tagging or different punctuation leading to different textual units being extracted as sentences) or to the presence of additional information that, however, has no link to the complex word. In addition, some examples were disregarded where the examined words were not definable as complex despite their absence in the frequency lists (e.g. 'cake', 'pique-nique' ('picnic', Fr.)). Conversely, the present words that were conjugated forms of a common verb, a frequent phenomenon in French and Spanish text (e.g. 'pronunciaba' - 'to pronounce', Sp., 1st/3rd person singular, imperfect tense), were taken to correctly represent complex vocabulary due to their difficult comprehension and use by children and non-native speakers.

4.1 Types of Extended Context

In accord with the experiment carried out by Nikolova-Stoupak (2023), three main categories of extended context as used in abridged literary texts were detected: additional vocabulary items pertaining to the complex word's lexical field (59.38% of the discovered instances), grammatical transformations that render the sentence simpler (34.38%) and definitions of

the complex word (6.25%).

4.1.1 Additional vocabulary from the same lexical field

Full: "'Ha!' roared Charley Bates."^a

Abridged: "Oliver opened his mouth and roared *with all the force of his baby lungs*."^b

Full: "Nous devrions l'explorer encore."^c
("We should explore it further."; Fr.)

Abridged: "Ils découvrent une grotte et *pénètrent à l'intérieur pour explorer les galeries*."^d ("They discover a cave and enter inside it in order to explore the galleries."; Fr.)

^aDickens, Charles. *Oliver Twist*. Project Gutenberg, 1838.

^bDickens, Charles. *Oliver Twist*. In *Illustrated Stories by Dickens*, edited by M. Sebag-Montefiore, Abridged edition, Usborne, 2009.

^cTwain, Mark. *Les Aventures de Tom Sawyer*. Translated by P. F. Caillé and Y. Dubois-Mauvais, Ebooks libres et gratuits, 1917.

^dTwain, Mark. *Les Aventures de Tom Sawyer*. Abridged edition, edited by A. Culleton, Broché, 2020.

Figure 1: Additional related vocabulary: examples

In the examples in Figure 1, the added context in relation to the complex word aids the reader to both understand its meaning and associate it with common collocations (one roars with a force coming from the lungs; in order to explore a closed area, one needs to enter it).

4.1.2 Transformation into simpler grammar

A variety of grammatical transformations are encountered within the abridged texts, and they neatly point to Harris' claim that language is rendered more complex as a result of contraction. For instance, a clause such as "to ask that someone give you something" is typically contracted as "to ask for something"; anaphora (e.g. "like *this*", a single word in many languages) replaces a longer referent; and personal pronouns are often skipped in languages where they can be inferred from the verb's declension. Within the examined abridged works, one can notice a tendency for sentences to be

<p>Full: "Ningún mendigo le pedía limosna."^a ("No beggar asked him for alms."; Sp.)</p> <p>Abridged: "Ningún mendigo le suplicó que le diera una limosna."^b (lit. "No beggar asked him that he give him alms."; Sp.)</p> <p>Full: "так {реветь}"^c ("to howl like this"; Ru.)</p> <p>Abridged: "{реветь} в тпу ручей"^d ("to howl like a stream"; Ru.)</p> <p>Full: "corrí"^e ("ran"; 1st person singular, Sp.)</p> <p>Abridged: "yo corrí"^f ("I ran"; Sp.)</p> <hr/> <p>^aDickens, Charles. <i>Canción de Navidad</i>. Ediciones del Sur, 2004.</p> <p>^bDickens, Charles. <i>Canción de Navidad</i>. Abridged edition, Ediciones la Cueva, 1905.</p> <p>^cCarroll, Lewis. <i>Alisa v Strane chudes</i>. Translated by N. M. Demurova, Nauka, 1978.</p> <p>^dCarroll, Lewis. <i>Alisa v Strane chudes</i>. Abridged edition, edited by L. Yahnin, Eksmo, 1991.</p> <p>^eShakespeare, William. <i>Romeo y Julieta</i>. Biblioteca digital universal, 2003.</p> <p>^fShakespeare, William. <i>Romeo y Julieta</i>. Abridged edition, translated by A. L. Pujalte, educ.ar.</p>
--

Figure 2: Grammatical transformations: examples

re-extended to an associated fuller form (see Figure 2).

4.1.3 Definition

As illustrated in Figure 3, definitions within the abridged texts come in two distinct types: metatextual (akin to a glossary) and in-text (naturally embedded within the literary text).

4.2 Results by Language

As seen in Figure 4, the addition of vocabulary items of the same lexical field is the most common scenario of context extension within the abridged texts found in the corpus, and its prevalence is strongest in relation to English and French texts. Grammatical transformations are closely as common within Spanish and Russian abridged works and significantly less prominent in English and French. Def-

<p>Full: "délabré"^a ("run down", Fr.)</p> <p>Abridged: (in the text's margin) "délabré : en mauvais état."^b ("run down: in a bad state"; Fr.)</p> <p>Full: "Frère Jean, religieux du même ordre."^c ("Brother Jean, a monk of the same order"; Fr.)</p> <p>Abridged: "c'est un religieux, un homme d'église."^d ("He is a monk, a man of the church."; Fr.)</p> <hr/> <p>^aHugo, Victor. <i>Les Misérables</i>. Project Gutenberg, 1862.</p> <p>^bHugo, Victor. <i>Les Misérables</i>. Abridged edition, edited by B. Faucard-Martinez, CLE International, 2016.</p> <p>^cShakespeare, William. <i>Roméo et Juliette</i>. Translated by F.-V. Hugo, Kiléma, 1868.</p> <p>^dShakespeare, William. <i>Roméo et Juliette</i>. Translated by A. Hubert, edited by F.-V. Hugo, Kiléma.</p>

Figure 3: Definition of complex vocabulary: examples

initions are globally rarest, ranging from no instances in Russian to three in French texts.

4.3 Results by Target Audience

Figure 5 shows the presence of extended context in abridged texts by target audience. When it comes to an audience of (native) children, involvement of additional vocabulary is by far most commonly found. In contrast, although this type of extended context is also most common in texts meant for foreign language learners, the other two types follow closely, notably definitions taking the second place. No instances of the investigated context extension are found within the texts marked as being for readers with comprehension difficulties.

The current sample of sentence pairs is insufficient for tendencies to be established in relation to subgroups of readers, such as children of a particular age or language learners with a particular proficiency level.

5 Discussion

Going back to research question 1 as stated in Section 1, the null hypothesis was not rejected in an initial experiment that compared

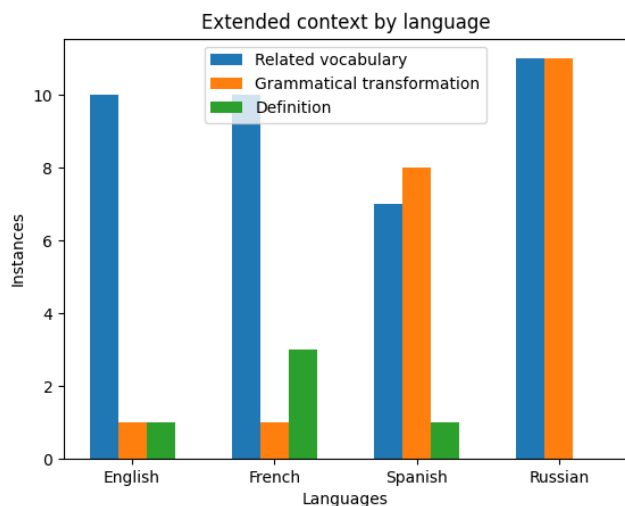


Figure 4: Instances of extended context by language

the number of instances where abridged texts contained longer sentences than their full counterparts in a random sample and in a sample associated with complex vocabulary. However, it was discovered that, in fact, less than half of the cases where the abridged sentences were longer were actually attributable to extended context related to the examined word (rather than to specificities of textual processing, for example). Therefore, results would be more accurate if the experiment is repeated with a larger sample and/or includes manual filtering of the sentences. Moreover, additional factors may exert an influence on the length of sentences within a literary text. For instance, translators have an established tendency to 'simplify' and/or 'explicate' a work's original text, thus respectively reducing and extending the size of context (Volansky et al., 2015).

As regards research question 2, when present, extended context within abridged texts was indeed systematisable as belonging to three main types: addition of vocabulary items of the same lexical field, transformations into simpler grammar (such as anaphora resolution and addition of implied personal pronouns) and definitions of the investigated complex vocabulary item. The first and third types are relatable to Harris (1988)'s notion of 'report' (description, definition), while the second type resembles the linguist's notion of 'paraphrase' (morphological and syntactic modification).

Instances of added vocabulary were most nu-

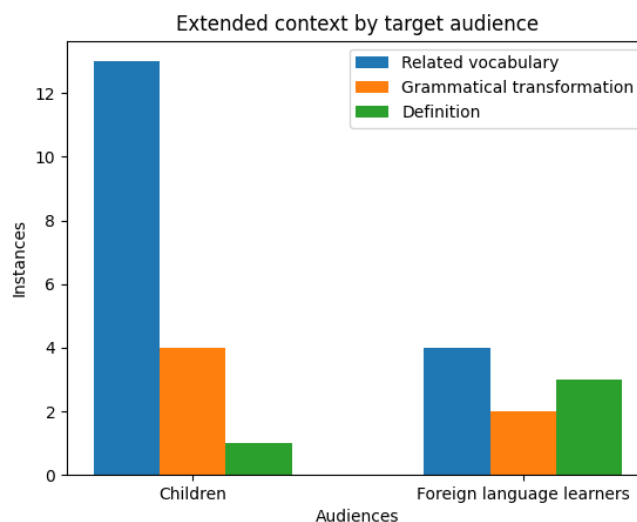


Figure 5: Instances of extended context by target audience

merous, potentially speaking of the importance for learners to acquire knowledge of a new word along with its common collocations and in a common and natural usage scenario (rather than through an explicit definition, for example). In addition, lexical fields are of key importance in the understanding of discourse. The words belonging to one field are used in a systematised pattern throughout a text and may as much as appear synonymously as anaphora referents (e.g. the phrase "I climbed (...)", as used without a direct object being followed by the definitive "The peak(...)" in the next sentence) (Christophe et al., 2010). Also, as demonstrated in de Bot et al. (1997)'s experiment, readers make especially high use of discourse content in order to determine the meaning of an unknown lexical item.

Evidence was found that grammatical transformations are significantly more common in morphologically rich languages like Russian and Spanish while, as de Bot et al. (1997) point out, languages like English and French, which have fewer word categories and argument structures, are associated with easier syntactic inference of a new word's meaning. It also appears that a larger variety of extended context is offered to foreign language learners. When it comes to the lack of detected instances of the investigated feature in relation to texts targeted for readers with comprehension difficulties, it is at least partly explainable through the fact that

most of these texts were by a single publisher, Almadraba (Spain) and that they contained designated glossaries, which were not processed as part of the texts.

Finally, the study revealed additional cases of simplified vocabulary and grammar in sentences containing complex vocabulary that are worthy of notice despite not directly leading to an increase in length, as they demonstrate a curious tendency for balance to be sought within the overall complexity present in an abridged textual unit. For instance, an abridged text thus uses a simpler verb as associated with the complex noun 'fragancia': "cuya fragancia hacía recordar" ⁸ ("whose fragrance made remember"; Sp.) in place of "fragancia que evocaba" ⁹ ("a fragrance that evoked"; Sp.). Likewise, a negative sentence that contains a relatively complex conjugated verb, like "vous ne me reconnaissez pas ?" ¹⁰ ("Don't you recognise me?"; Fr.) is made positive in an abridged text: "vous me reconnaissez ?" ¹¹ ("Do you recognise me?"; Fr.).

6 Conclusion and Future Directions

The current study shows that there are cases when abridged literary works make use of extended rather than reduced context, in particular when it comes to the introduction of complex vocabulary. Several distinct types of context extension are detected, and there is evidence that their application and frequency are language-specific. However, further work is required in order to determine the particular cases in which the practice is applied as well as its efficiency. The authors are planning for the latter to be evaluated through a reading comprehension survey, in which pairs of short texts will be presented, one of which will introduce complex vocabulary through extended context in scenarios resembling the ones found within the examined corpus.

To go further, a tendency has been detected for complex vocabulary to influence its context

⁸Dickens, Charles. Canción de Navidad. Abridged edition, edited by S. R. Santerbás, Anaya, 1986.

⁹Dickens, Charles. Canción de Navidad. Ediciones del Sur, 2004.

¹⁰Hugo, Victor. Les Misérables. Project Gutenberg, 1862.

¹¹Hugo, Victor. Les Misérables. Abridged edition, edited by B. Faucard-Martinez, CLE International, 2016.

in ways that are not related to length but to overall complexity in terms of syntax and vocabulary. This claim deserves to be further investigated in order for optimal abridgement practices to be defined that coincidentally render a text simpler and encourage vocabulary acquisition.

7 Limitations

Although the utilised corpus has been specifically compiled in a way as to provide a variety of abridged versions of famous literary works, limitations in terms of budget and availability have influenced its extent. In addition, some texts (particularly, Russian-language ones) are not clearly denoted as abridged or full. Finally, the quality of texts is imperfect due to the processes of character recognition and conversion between textual formats; a limitation that has, however, not proven significant.

References

- Gladis Leonor Arias Rodríguez and Eliana Edith Roberto Flórez. 2018. Using the abridged version of some novels as a way to encourage students' written and oral production. *GiST Education and Learning Research Journal*, (16):6–32.
- Kees de Bot, T. Sima Paribakht, and Marjorie Birmingham Wesche. 1997. Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from esl reading. *Studies in Second Language Acquisition*, 19(3):309–329.
- Gomatam Mohana Charyulu. 2018. Complications in reading abridged texts: a study on cultural destruction by ellis in meaning-making process. *Review of Research*, 7(9):1–5.
- Anne Christophe, Séverine Millotte, Perrine Brusini, and Elodie Cauvet. 2010. Early bootstrapping of syntactic acquisition. In *Language Acquisition Across Linguistic and Cognitive Systems*, pages 53–66.
- Charles A. Ferguson. 1975. Toward a characterization of english foreigner talk. *Anthropological Linguistics*, 17(1):1–14.
- Zellig Harris. 1988. *Language and Information*. Columbia University Press.
- Amr Helmy Ibrahim. 2013. Une mesure unifiée de la complexité linguistique : l'analyse matricielle définitoire. *Nouvelles perspectives en sciences sociales*, 9:17–80.

- A. Leung, A. Tunkel, and D. Yurovsky. 2021. Parents fine-tune their speech to children’s vocabulary knowledge. *Psychological Science*, 32:975–984.
- Elizabeth Birr Moje, Melanie Overby, Nicole Tysvaer, and Karen Morris. 2008. The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard Educational Review*, 78(1):107–154.
- National Literacy Trust. 2006. [Reading for pleasure: A research overview](#). Accessed April 6, 2023.
- Iglika Nikolova-Stoupak. 2023. Introduction of complex vocabulary in literature through fine-tuning: A corpus-based study. *Innovation in Language Learning*, 16.
- Vered Volansky, Noam Ordan, and Shuly Wintner. 2015. [On the features of translationese](#). *Digital Scholarship in the Humanities*, 30(1):98–118.

Appendix A Corpus Characteristics

Language	Text	Version	Length (char.)	Audience
English	<i>Alice in Wonderland</i> ¹²	Full	144018	
English	<i>Alice in Wonderland</i> ¹³	Abridged 1	34836	FL ¹⁴ A2 ¹⁵
English	<i>Alice in Wonderland</i> ¹⁶	Abridged 2	53003	(not indicated)
English	<i>A Christmas Carol</i> ¹⁷	Full	158394	
English	<i>A Christmas Carol</i> ¹⁸	Abridged	39314	FL B1 ¹⁹
English	<i>Don Quixote</i> ²⁰	Full	2167742	
English	<i>Don Quixote</i> ²¹	Abridged 1	130587	children
English	<i>Don Quixote</i> ²²	Abridged 2	81001	age 8+ ²³
English	<i>Les Miserables</i> ²⁴	Full	3199531	
English	<i>Les Miserables</i> ²⁵	Abridged 1	1493465	(not indicated)
English	<i>Les Miserables</i> ²⁶	Abridged 2	166826	FL C1
English	<i>Les Miserables</i> ²⁷	Abridged 3	1790853	(not indicated)
English	<i>Oliver Twist</i> ²⁸	Full	887739	
English	<i>Oliver Twist</i> ²⁹	Abridged 1	21041	age 9+
English	<i>Oliver Twist</i> ³⁰	Abridged 2	147207	FL B2/C1
English	<i>Oliver Twist</i> ³¹	Abridged 3	22948	age 7+
English	<i>Romeo and Juliet</i> ³²	Full	140481	
English	<i>Romeo and Juliet</i> ³³	Abridged 1	32375	children
English	<i>Romeo and Juliet</i> ³⁴	Abridged 2	316247	(not indicated)
English	<i>Romeo and Juliet</i> ³⁵	Abridged 3	74939	(not indicated)
English	<i>Romeo and Juliet</i> ³⁶	Abridged 4	19758	children

¹²Carroll, Lewis. *Alice's Adventures in Wonderland*. Project Gutenberg, 1865.

¹³Carroll, Lewis. *Alice's Adventures in Wonderland*. Abridged edition, edited by J. Bassett, Oxford University Press, 2000.

¹⁴foreign language

¹⁵indicated as '700 headwords'

¹⁶Carroll, Lewis. *Alice's Adventures in Wonderland*. Abridged edition, Sam'l Gabriel Sons and Company, 1916.

¹⁷Dickens, Charles. *A Christmas Carol*. The Baker & Taylor Company, 1843.

¹⁸Dickens, Charles. *A Christmas Carol*. Abridged edition, edited by P. Lagendijk, Mediasat Poland Bis, 2004.

¹⁹1000 headwords

²⁰Cervantes, Miguel de. *Don Quixote*. Translated by J. Ormsby, Project Gutenberg, 2004. (Original work published 1605)

²¹Cervantes, Miguel de. *Don Quixote*. Abridged edition, edited by E. Underdown, Thomas Nelson and Sons, 1910.

²²Cervantes, Miguel de. *Don Quixote*. Abridged edition, edited by K. Casey, Dover Publications, 1979.

²³where relevant, it is the lowest age for which a text is recommended that is considered

²⁴Hugo, Victor. *Les Miserables*. Translated by I. F. Hapgood, Project Gutenberg, 1887.

²⁵Hugo, Victor. *Les Misérables*. <https://lesmiserablesabridged.blogspot.com/>.

²⁶Hugo, Victor. *Les Miserables*. Abridged edition, easybusyomsk.ru, 2017.

²⁷Hugo, Victor. *Les Miserables*. Translated by C. E. Wilbour, edited by F. M. Cooper, Abridged edition, Pranava Books, 2022.

²⁸Dickens, Charles. *Oliver Twist*. Project Gutenberg, 1838.

²⁹Dickens, Charles. *Oliver Twist*. Abridged edition, translated by G. Tavner, Real Reads, 2009.

³⁰Dickens, Charles. *Oliver Twist*. Abridged edition, edited by R. Rogers, Oxford Bookworms, 2009.

³¹Dickens, Charles. *Oliver Twist*. In *Illustrated Stories by Dickens*, edited by M. Sebag-Montefiore, Abridged edition, Usborne, 2009.

³²Shakespeare, William. *Romeo and Juliet*. Project Gutenberg, 1597.

³³Shakespeare, William. *Romeo and Juliet*. Abridged edition, edited by M. Lamb and P. Middelboe, Shakespeare Schools Festival.

³⁴Shakespeare, William. *Romeo and Juliet*. Abridged edition, Spark Publishing, 2003.

³⁵Shakespeare, William. *Romeo and Juliet*. Abridged edition, edited by J. Tanner, The Wichita Shakespeare Co., 2017.

³⁶Shakespeare, William. *Romeo and Juliet*. Abridged edition, edited by M. Heyes, Eleanor Palmer Primary

English	<i>The Adventures of Tom Sawyer</i> ³⁷	Full	386043	
English	<i>The Adventures of Tom Sawyer</i> ³⁸	Abridged 1	141205	FL
English	<i>The Adventures of Tom Sawyer</i> ³⁹	Abridged 2	21864	FL A1 ⁴⁰
English	<i>The Adventures of Tom Sawyer</i> ⁴¹	Abridged 3	32220	FL A1 ⁴²
French	<i>Alice in Wonderland</i> ⁴³	Full 1	183660	
French	<i>Alice in Wonderland</i> ⁴⁴	Full 2	187474	
French	<i>Alice in Wonderland</i> ⁴⁵	Abridged 1	12235	age 5+
French	<i>Alice in Wonderland</i> ⁴⁶	Abridged 2	73638	children
French	<i>Alice in Wonderland</i> ⁴⁷	Abridged 3	64692	age 9+
French	<i>A Christmas Carol</i> ⁴⁸	Full 1	194592	
French	<i>A Christmas Carol</i> ⁴⁹	Full 2	188831	
French	<i>A Christmas Carol</i> ⁵⁰	Abridged	6550	children
French	<i>Don Quixote</i> ⁵¹	Full	2391260	
French	<i>Don Quixote</i> ⁵²	Abridged 1	93066	(not indicated)
French	<i>Don Quixote</i> ⁵³	Abridged 2	606051	(not indicated)
French	<i>Les Misérables</i> ⁵⁴	Full	3674712	
French	<i>Les Misérables</i> ⁵⁵	Abridged 1	59606	FL A2
French	<i>Les Misérables</i> ⁵⁶	Abridged 2	670357	(not indicated)
French	<i>Madame Bovary</i> ⁵⁷	Full	683958	
French	<i>Madame Bovary</i> ⁵⁸	Abridged	60448	FL B2
French	<i>Oliver Twist</i> ⁵⁹	Full 1	467498	
French	<i>Oliver Twist</i> ⁶⁰	Full 2	660534	
French	<i>Oliver Twist</i> ⁶¹	Full 3	932256	

School, London, 2016.

³⁷Twain, Mark. *The Adventures of Tom Sawyer*. Project Gutenberg, 1876.

³⁸Twain, Mark. *The Adventures of Tom Sawyer*. Abridged edition, Global Publishing Solutions.

³⁹Twain, Mark. *The Adventures of Tom Sawyer*. Translated by J. Kehl, Pearson Education, 2000.

⁴⁰300 headwords

⁴¹Twain, Mark. *The Adventures of Tom Sawyer*. Abridged edition, edited by N. Bullard, Oxford University Press, 2000.

⁴²400 headwords

⁴³Carroll, Lewis. *Les Aventures d’Alice au pays des merveilles*. Translated by J. Papy, Gallimard Jeunesse, 2001.

⁴⁴Carroll, Lewis. *Les Aventures d’Alice au pays des merveilles*. Translated by H. Bué, Hachette, 1908.

⁴⁵Carroll, Lewis. *Les Aventures d’Alice au pays des merveilles*. Abridged edition, edited by H. Parisot, Editions Corentin, 1975.

⁴⁶Carroll, Lewis. *Les Aventures d’Alice au pays des merveilles*. Abridged edition, edited by P. Protet, Auzou, 2012.

⁴⁷Carroll, Lewis. *Les Aventures d’Alice au pays des merveilles*. Abridged edition, edited by P. Rouard, Bayard Jeunesse, 1992.

⁴⁸Dickens, Charles. *Conte de Noël*. Translated by A. De Goy and De Saint-Romain, La Bibliothèque électronique du Québec, 1890.

⁴⁹Dickens, Charles. *Conte de Noël*. pitbook.com, 2002.

⁵⁰Dickens, Charles. *Un Chant de Noël*. Abridged edition, edited by L. Papineau, Héritage, 2004.

⁵¹Cervantes, Miguel de. *Don Quichotte*. Translated by L. Viardot, La Bibliothèque électronique du Québec, 1936.

⁵²Cervantes, Miguel de. *Don Quichotte*. Abridged edition, edited by D. Guerin de Bouscal, Theatre classique, 2015.

⁵³Cervantes, Miguel de. *Don Quichotte*. Translated by J.-P. Florian, Hachette, 1882.

⁵⁴Hugo, Victor. *Les Misérables*. Project Gutenberg, 1862.

⁵⁵Hugo, Victor. *Les Misérables*. Abridged edition, edited by B. Faucard-Martinez, CLE International, 2016.

⁵⁶Hugo, Victor. *Les Misérables*. La Bibliothèque électronique du Québec.

⁵⁷Flaubert, Gustave. *Madame Bovary: Mœurs de province*. Project Gutenberg, 1857.

⁵⁸Flaubert, Gustave. *Madame Bovary*. Abridged edition, edited by B. Faucard-Martinez, CLE International, 2019.

⁵⁹Dickens, Charles. *Olivier Twist*. Translated by A. Gérardin, Librairie Hachette et Cie., 1893.

⁶⁰Dickens, Charles. *Oliver Twist*. Translated by E. De La Bédolère, Culturea, 2022.

⁶¹Dickens, Charles. *Oliver Twist*. Translated by A. Girardin, Ink Books, 2017.

French	<i>Oliver Twist</i> ⁶²	Abridged	22047	age 10+
French	<i>Romeo and Juliet</i> ⁶³	Full 1	166742	
French	<i>Romeo and Juliet</i> ⁶⁴	Abridged 1	71651	(not indicated)
French	<i>Romeo and Juliet</i> ⁶⁵	Abridged 2	19449	CD ⁶⁶
French	<i>The Adventures of Tom Sawyer</i> ⁶⁷	Full 1	404458	
French	<i>The Adventures of Tom Sawyer</i> ⁶⁸	Full 2	417108	
French	<i>The Adventures of Tom Sawyer</i> ⁶⁹	Abridged	7191	age 7+
Russian	<i>Alice in Wonderland</i> ⁷⁰	Full 1	173205	
Russian	<i>Alice in Wonderland</i> ⁷¹	Full 2	140446	
Russian	<i>Alice in Wonderland</i> ⁷²	Abridged 1	127433	(not indicated)
Russian	<i>Alice in Wonderland</i> ⁷³	Abridged 2	31018	age 6+
Russian	<i>Madame Bovary</i> ⁷⁴	Full	619263	
Russian	<i>Madame Bovary</i> ⁷⁵	Abridged	610777	(not indicated)
Russian	<i>Oliver Twist</i> ⁷⁶	Full	854364	
Russian	<i>Oliver Twist</i> ⁷⁷	Abridged	74449	children
Russian	<i>Romeo and Juliet</i> ⁷⁸	Full 1	154285	
Russian	<i>Romeo and Juliet</i> ⁷⁹	Full 2	120633	
Russian	<i>Romeo and Juliet</i> ⁸⁰	Abridged	33182	(not indicated)
Russian	<i>The Adventures of Tom Sawyer</i> ⁸¹	Full	954134	
Russian	<i>The Adventures of Tom Sawyer</i> ⁸²	Abridged 1	218344	(not indicated)
Russian	<i>The Adventures of Tom Sawyer</i> ⁸³	Abridged 2	374766	age 11+
Spanish	<i>Alice in Wonderland</i> ⁸⁴	Full 1	156666	
Spanish	<i>Alice in Wonderland</i> ⁸⁵	Full 2	142262	
Spanish	<i>Alice in Wonderland</i> ⁸⁶	Abridged	66023	age 10+

⁶²Dickens, Charles. *Oliver Twist*. Abridged edition, translated by F. Ledoux, edited by P. Arrou-Vignod, Gallimard Jeunesse, 2018.

⁶³Shakespeare, William. *Roméo et Juliette*. Translated by F.-V. Hugo, Kiléma, 1868.

⁶⁴Shakespeare, William. *Roméo et Juliette*. Abridged edition, edited by E. Gwénola and P. Fièvre, Theatre classique, 2014.

⁶⁵Shakespeare, William. *Roméo et Juliette*. Translated by A. Hubert, edited by F.-V. Hugo, Kiléma.

⁶⁶readers with comprehension difficulties

⁶⁷Twain, Mark. *Les Aventures de Tom Sawyer*. Translated by P. F. Caillé and Y. Dubois-Mauvais, Ebooks libres et gratuits, 1917. (Original work published 1876)

⁶⁸Twain, Mark. *Les Aventures de Tom Sawyer*. Translated by F. De Gail, Flammarion, 1996.

⁶⁹Twain, Mark. *Les Aventures de Tom Sawyer*. Abridged edition, edited by A. Culleton, Broché, 2020.

⁷⁰Carroll, Lewis. *Alisa v Strane chudes*. Translated by N. M. Demurova, Nauka, 1978.

⁷¹Carroll, Lewis. *Alisa v Strane chudes*. Translated by Y. Nesterenko, Biblioteka Maksima Moshkova, 2000.

⁷²Carroll, Lewis. *Alisa v Strane chudes*. Abridged edition, edited by L. Yahnin, Eksmo, 1991.

⁷³Carroll, Lewis. *Alisa v Strane chudes*. Abridged edition, Eksmo, 2018.

⁷⁴Flaubert, Gustave. *Gospozha Bovari*. Translated by A. Chebotarevski, Klub semeynovo dosuga, 2011.

⁷⁵Flaubert, Gustave. *Gospozha Bovari*. Translated by N. Lyubimov. <https://www.100bestbooks.ru/>.

⁷⁶Dickens, Charles. *Priklyucheniya Olivera Tvista*. Translated by A. Krivtsov, Eksmo, 2023.

⁷⁷Dickens, Charles. *Priklyucheniya Olivera Tvista*. Translated by I. O. Rodin, Biblioteka shkol'nika.

⁷⁸Shakespeare, William. *Romeo i Zhuletta*. Translated by D. Mihailovski, educ.ar.

⁷⁹Shakespeare, William. *Romeo i Zhuletta*. Translated by B. Pasternak, Biblioteka dramaturgii agenstva.

⁸⁰Shakespeare, William. *Romeo i Zhuletta*. Abridged edition, edited by Z. Barbe and M. Karre, Orchestre et Choeurs du Théâtre National de l'Opéra de Paris, 1968.

⁸¹Twain, Mark. *Priklyucheniya Toma Soyera*. Translated by K. Chukovskiy, Kaliningradskoe Knizhnoe Izdatelstvo, 1972.

⁸²Twain, Mark. *Priklyucheniya Toma Soyera*. Abridged edition, edited by I. O. Rodin, Biblioteka Shkolnika, 2014.

⁸³Twain, Mark. *Priklyucheniya Toma Soyera*. Abridged edition, edited by N. L. Daruzes, Vita Nova, 2014.

⁸⁴Carroll, Lewis. *Las aventuras de Alicia en el país de las maravillas*. Translated by M. Aguirre, Ediciones del Sur, 2003.

⁸⁵Carroll, Lewis. *Las aventuras de Alicia en el país de las maravillas*. Translated by L. Maristany, Titivillus, 1996.

⁸⁶Carroll, Lewis. *Las aventuras de Alicia en el país de las maravillas*. Abridged edition, edited by F. Díez de

Spanish	<i>A Christmas Carol</i> ⁸⁷	Full	170634	
Spanish	<i>A Christmas Carol</i> ⁸⁸	Abridged 1	164266	(not indicated)
Spanish	<i>A Christmas Carol</i> ⁸⁹	Abridged 2	40993	(not indicated)
Spanish	<i>Don Quixote</i> ⁹⁰	Full	2093496	
Spanish	<i>Don Quixote</i> ⁹¹	Abridged 1	212235	FL B1 ⁹²
Spanish	<i>Don Quixote</i> ⁹³	Abridged 2	302115	(not indicated)
Spanish	<i>Don Quixote</i> ⁹⁴	Abridged 3	1022357	children
Spanish	<i>Les Miserables</i> ⁹⁵	Full	851969	
Spanish	<i>Les Miserables</i> ⁹⁶	Abridged	15518	age 12+
Spanish	<i>Madame Bovary</i> ⁹⁷	Full 1	688237	
Spanish	<i>Madame Bovary</i> ⁹⁸	Full 2	695263	
Spanish	<i>Madame Bovary</i> ⁹⁹	Abridged 1	89352	CD
Spanish	<i>Madame Bovary</i> ¹⁰⁰	Abridged 2	661392	(not indicated)
Spanish	<i>Oliver Twist</i> ¹⁰¹	Full	994225	
Spanish	<i>Oliver Twist</i> ¹⁰²	Abridged 1	112781	CD
Spanish	<i>Oliver Twist</i> ¹⁰³	Abridged 2	95452	(not indicated)
Spanish	<i>Romeo and Juliet</i> ¹⁰⁴	Full	240967	
Spanish	<i>Romeo and Juliet</i> ¹⁰⁵	Abridged 1	78402	CD
Spanish	<i>Romeo and Juliet</i> ¹⁰⁶	Abridged 2	127332	(not indicated)
Spanish	<i>The Adventures of Tom Sawyer</i> ¹⁰⁷	Full	415424	
Spanish	<i>The Adventures of Tom Sawyer</i> ¹⁰⁸	Abridged 1	90334	(not indicated)
Spanish	<i>The Adventures of Tom Sawyer</i> ¹⁰⁹	Abridged 2	135697	age 11+
Spanish	<i>The Adventures of Tom Sawyer</i> ¹¹⁰	Abridged 3	36930	CD
Spanish	<i>The Adventures of Tom Sawyer</i> ¹¹¹	Abridged 4	95491	CD

Miranda, Zig Zag, 2018.

⁸⁷Dickens, Charles. Canción de Navidad. Ediciones del Sur, 2004.

⁸⁸Dickens, Charles. Canción de Navidad. Abridged edition, Ediciones la Cueva, 1905.

⁸⁹Dickens, Charles. Canción de Navidad. Abridged edition, edited by S. R. Santerbás, Anaya, 1986.

⁹⁰Cervantes, Miguel de. Don Quijote de la Mancha. Editorial Porrúa S.A. de C.V., 1605.

⁹¹Cervantes, Miguel de. Don Quijote de la Mancha. Abridged edition, edited by U. Malmlose, Egmont, 1972.

⁹²2500 headwords

⁹³Cervantes, Miguel de. Don Quijote de la Mancha. Abridged edition, edited by D. Rieu, Librairie des Éditions Espagnoles, 1955.

⁹⁴Cervantes, Miguel de. Don Quijote de la Mancha. IES Maese Rodrigo, 2012.

⁹⁵Hugo, Victor. Les Misérables. Livros grátis, 2013.

⁹⁶Hugo, Victor. Los Miserables. Translated by M. Pujadó, Anaya, 2020.

⁹⁷Flaubert, Gustave. Madame Bovary. Translated by M. Armiño, Siruela, 2014.

⁹⁸Flaubert, Gustave. Madame Bovary. Translated by M. T. G. Urrutia, Hofmiller, 2012.

⁹⁹Flaubert, Gustave. Madame Bovary. Abridged edition, edited by C. S. Baudet, La mar de fácil, 2018.

¹⁰⁰Flaubert, Gustave. Madame Bovary. Abridged edition, Culturea, 2010.

¹⁰¹Dickens, Charles. Las Aventuras de Oliver Twist. Culturea, 2003.

¹⁰²Dickens, Charles. Oliver Twist. Abridged edition, edited by N. Martí Constans, Almadra, 2016.

¹⁰³Dickens, Charles. Oliver Twist. Abridged edition, Guagua Luna, 2020.

¹⁰⁴Shakespeare, William. Romeo y Julieta. Biblioteca digital universal, 2003.

¹⁰⁵Shakespeare, William. Romeo y Julieta. Abridged edition, edited by E. O'Callaghan, Almadra, 2009.

¹⁰⁶Shakespeare, William. Romeo y Julieta. Abridged edition, translated by A. L. Pujalte, educ.ar.

¹⁰⁷Twain, Mark. Las Aventuras de Tom Sawyer. Translated by J. Torroba, Biblioteca Virtual Universal, 2003.

¹⁰⁸Twain, Mark. Las Aventuras de Tom Sawyer. Abridged edition, edited by B. Palacios, Dirección General de Bibliotecas, 2010.

¹⁰⁹Twain, Mark. Las Aventuras de Tom Sawyer. Abridged edition, edited by L. I. Barrena, Anaya, 2007.

¹¹⁰Twain, Mark. Las Aventuras de Tom Sawyer. Abridged edition, edited by M. A. Rubio and M. Sintes, Almadra, 2006.

¹¹¹Twain, Mark. Las Aventuras de Tom Sawyer. Abridged edition, edited by M. A. Rubio, Almadra, 2005.