

INTRODUCTION OF COMPLEX VOCABULARY IN LITERATURE THROUGH FINE-TUNING

A CORPUS-BASED STUDY

IGLIKA NIKOLOVA-STOUPAK

SORBONNE UNIVERSITY

Innovation in Language Learning
2023



Abridgement?

Quantitative vs Qualitative Abridgement?

Summary?

Adaptation?

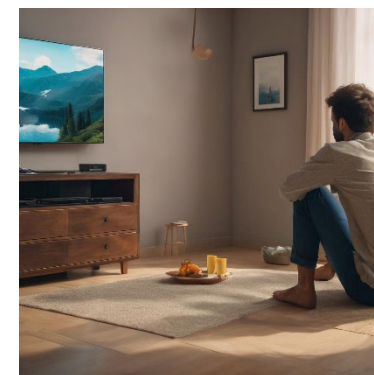
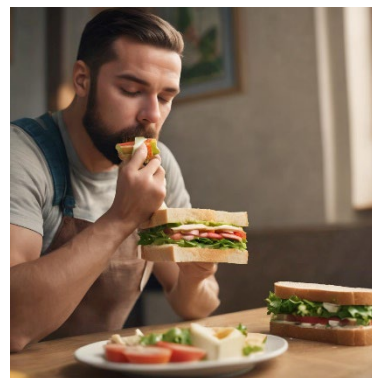
Can a version of a literary work specially created for children or foreign language learners involve addition rather than subtraction?

YES! EXAMPLES

- Illustrations
- Glossaries
- Shorter versions of a sentence may carry the same information while being more complex (Harris) => making text longer can mean making it simpler

1) John ate a sandwich. After that, John sat to watch the news on TV.

2) Having eaten a sandwich, John sat to watch the news.



RESEARCH QUESTION

- A particular case of abridgement through addition will be examined
- Leung *et al.* 2021: Parents introduce complex vocabulary through fine-tuning, providing additional explanatory context

The **dotted** leopard ran after the rabbit.

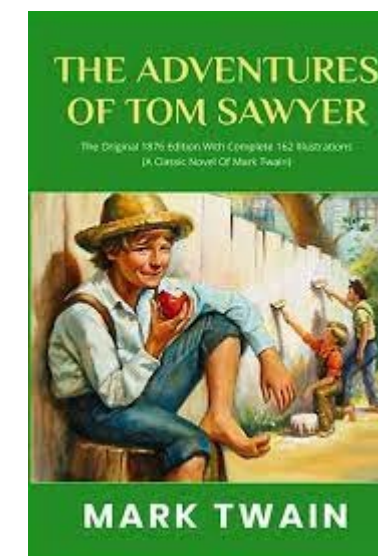
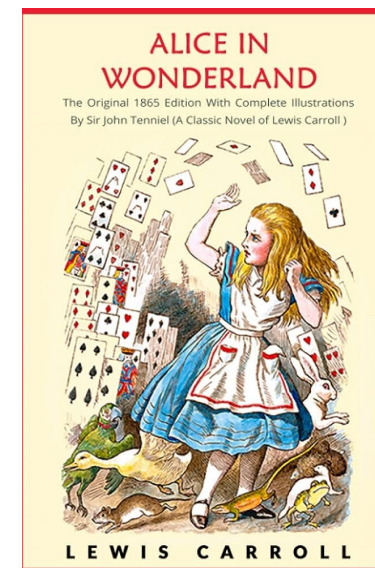
The leopard ran after the rabbit **like a cat**.



- Do literary works adapted for children and foreign language students also feature additional context at the introduction of complex vocabulary?
- **Additional questions:** To what extent? Are there differences in relation to different audiences? Are there differences by language?

EXPERIMENTAL SETUP: CORPUS

- Famous literary works (same time period)
 - Charles Dickens, *A Christmas Carol*
 - Lewis Carroll, *Alice's Adventures in Wonderland*
 - Mark Twain, *The Adventures of Tom Sawyer*
- Languages
 - English
 - French
 - Russian
 - Spanish
- Both full and abridged versions



EXPERIMENTAL SETUP: TEXTUAL PROCESSING

- Traditional preprocessing (e.g. removal of capitalisation and punctuation)
- POS tagging applied and a list of nouns taken from full texts
- Only longest words in the list retained (Zipf's Law of Least Effort)
- Mistakes analysed and, if possible, removed:
 - non-words (“adventuresbeginning”)
 - non-nouns (“aficionada”, adj.)
 - common words (“Christmas,” “cumpleaños”)
 - phrases in FR and RU that are hyphenated based on grammatical rules (“commença-t-elle”, “как-нибудь”)

EXPERIMENTAL SETUP: COMPARISON

- The list is limited to the nouns also present in an abridged version (for every full-abridged pair)
- The contexts of first appearance of the nouns are extracted and juxtaposed
- The examined feature (i.e. introduction of a word with additional context in an abridged version) is sought and analysed

FULL: “and dashed the wild year through there stood a **solitary lighthouse** great heaps of seaweed clung to its base”

ABRIDGED: “crashed violently underneath them the spirit took scrooge to a **lighthouse built on a lonely rock** several miles from land”

RESULTS: TYPES OF ADDITION

- A total of 62 instances of the feature were found in 377 parallel introductions of complex vocabulary
- They were categorised into 3 types:
 - addition of related vocabulary (46.8 %)
 - “brouillard” vs “épais brouillard”
 - “гостеприимство” vs “щедрое гостеприимство”
 - grammatical transformation (into a simpler structure) (41.9 %)
 - “tejer una guirnalda de margaritas” vs “juntar margaritas para trenzar una guirnalda”
 - explanation or definition (11.3 %)
 - “в суде заседают, потому и называются ‘присяжные заседатели’”

RESULTS: BY LANGUAGE

Language	Prominence of the Feature (proportion)	Prominence of the Feature (%)	Breakdown by Type of Additional Context
English	4/69	12.2%	50% gram. trans.
			50% related voc.
French	14/67	20.9%	28.6% gram. trans.
			42.9% related voc.
			28.6% explanation
Russian	15/85	17.6%	46.7% gram. trans.
			33.3% related voc.
			20% explanation
Spanish	29/156	18.6%	44.8% gram. trans.
			55.2% related voc.

- The number of instances is lowest for English texts and highest for French
- The results for French and Spanish are comparable
- Explanations are very limited in number and appear only in French and Russian texts

RESULTS: BY AUDIENCE

Audience	Prominence of the Feature (proportion)	Prominence of the Feature (%)	Breakdown by Type of Additional Context
General/Undefined	24/139	17.3%	50% gram. trans.
			41.7% related voc.
			8.3% explanation
FL Students	4/15	26.7%	50% gram. trans.
			50% related voc.
0-500 words	N/A	N/A	N/A
500-1000 words	2/5	40%	50% gram. trans.
			50% related voc.
Children	28/169	16.6%	35.7% gram. trans.
			44.4% related voc.
			17.9% explanation
Age 5-8	2/10	20%	100% explanation
Age 9-11	14/95	14.7%	50% gram. trans.
			50% related voc.

- Clearer tendencies as compared with language
- Very few of the complex words are present in abridged versions for young children
- The tendency is identical with lower levels of foreign language learners
- The feature is most prominent with higher-level foreign language learners (40% of considered words)

CONCLUSION AND FUTURE WORK

- The addition of context at the introduction of complex vocabulary can be used as a readability feature
- Conclusions would benefit from a larger corpus and as well as additional languages
- Better NLP tools should be sought (e.g. word lists instead of longest words)
- The study can be placed within a relevant context (e.g. Harris's transformations by type; influence of complex vocabulary on the text surrounding it)

BIBLIOGRAPHY

- Charyulu, G. M. (2018). Complications in Reading Abridged Texts: A Study on Cultural Destruction by ELLs in Meaning-Making Process, *Review of Research* 7, 9: 1-5.
- DuBay, W. H. (2007), *The Classic Readability Studies*, Clearinghouse.
- Gala, N., Todirascu, A., Bernhard, D., Wilkens, R. and Meyer, J.-P. (2020). Transformations syntaxiques pour une aide à l'apprentissage de la lecture : typologie, adéquation et corpus adaptés, *Congrès Mondial de Linguistique Française*, Montpellier, France.
- Harris, Z. (1988). *Language and Information*, Columbia University Press.
- Krashen, S. (1982). *Principles and Practices of Second Language Acquisition*, Pergamon.
- Leung, A., Tunkel, A. and Yurovsky, D. (2021). Parents Fine-Tune Their Speech to Children's Vocabulary Knowledge, *Psychological Science*, 32: 975–984
- Moje, E. B., Overby, M., Tysvaer, N. and Morris, K. (2008). The Complex World of Adolescent Literacy: Myths, Motivations, and Mysteries, *Harvard Educational Review* 78, 1: 107–54.
- Reading for Pleasure: A Research Overview. (2006). *National Literacy Trust*. <https://literacytrust.org.uk/research-services/research-reports/reading-pleasure-research-overview/>.
- Rodriguez, A., Leonor, G. and Flórez, E. E. R. (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.
- Wilkens, R., Alfter, D., Wang, X., Pintard, A., Tack, A., Yancey, K., and François, T. (2022) FABRA: French Aggregator-Based Readability Assessment toolkit, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 1217–1233.
- Xia, M., Kochmar, E. and Briscoe, T. (2016). Text Readability Assessment for Second Language Learners, *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, 12-22.

THANK YOU FOR YOUR KIND ATTENTION!

QUESTIONS ARE WELCOME