



# Restoring Literary Style in Adaptations for EFL Learners

Iglika Nikolova-Stoupak<sup>1</sup>, Gaël Lejeune<sup>2</sup>, Eva Schaeffer-Lacroix<sup>3</sup>

<sup>1, 2, 3</sup> Sorbonne Université, France

## Abstract

*The proposed work continues a previous study in which we defined the distinctive characteristics of a literary work as its top deviations from a reference corpus in terms of a large selection of measurable textual characteristics, including readability features, grammatical constructions, and literary devices [1]. Our current step consists of 1) improving existing literary adaptations for EFL learners by reintroducing the encountered distinctive characteristics of the respective original work and 2) evaluating the intervention through a survey of reader experience that addresses text comprehension and appreciation. The conducted survey, answered by 45 participants, evaluated three types of literary adaptation: 1) an original extract from a professionally crafted graded reader ("baseline"); 2) the same extract as enriched through our proposed procedure ("enriched"); and 3) a version of the extract derived through automatic style transfer ("style transfer"). Three CEFR levels and two literary texts were involved. Based on participants' responses, we are inclined to think that the undergone enrichment has no significant effect on reader comprehension but might reduce appreciation of the text as defined by readers' judgement. Conversely, automatic style transfer offers a promising direction in associating a work's original style with positive reader experience.*

**Keywords:** computer-assisted literary adaptation, reader experience survey, literary style, corpus analysis

## 1. Introduction

Textual adaptation has been receiving growing attention in recent years as a means of making written content accessible to wider audiences, particularly through text simplification and Easy-Read. In particular, accessibility combined with natural exposure to language have been a focus in foreign-language learning [2-4]. In this context, graded readers, which are often adaptations of well-known literary works, provide texts that correspond to learners' proficiency levels. Such materials have been associated with vocabulary development as well as learner motivation and engagement [5-7]. However, questions are often raised about adapted literature's authenticity and literary qualities [8, 9], and graded readers have not been spared related criticism in their quest to meet linguistic and proficiency constraints [10, 11]. The present study investigates whether Natural Language Processing (NLP) methods can help mitigate the trade-off between accessibility and authenticity introduced by graded readers through restoring aspects of works' original style. We preliminarily defined a work's literary style as the deviations in a large number and variety of quantitative characteristics measured based on a literary corpus of a substantial size [1]. The current paper focuses on: 1) the use of the encountered distinguishing features of two literary texts (*The Picture of Dorian Gray* by Oscar Wilde and *Moby-Dick* by Herman Melville) to enrich existing adaptations for EFL learners [12-15] in a computer-assisted manner and 2) an evaluation of the derived texts based on a survey of reader experience. The survey features two graded-reader extracts per literary work at different CEFR levels. Each of these four texts is presented in three alternative versions: "baseline", "enriched" and "style transfer". The whole survey as well as other materials related to the present project can be accessed in the [associated repository](#).

## 2. Background and Related Work

### 2.1 Stylistics and Stylometry

Stylistics, the study of style, predates computational methods in its analysis of textual elements from vocabulary and syntax to semantics. Wellek and Warren [16] define style as "deviations and distortions from normal usage." While style has been traditionally associated with particular authors, alternative interpretations exist that take an isolated text as a stylistic unit [17]. Stylometry, in turn, which is also known as "computational stylistics", investigates full corpora in its quest to define and describe textual style, focusing on different patterns and frequencies. The quantitative textual features that stylometry measures may include those associated with readability, such as vocabulary richness and complexity



and different measures of length [18] and more strictly-defined linguistic phenomena, such as those defined in Biber's multi-dimensional analysis [19].

## 2.2 Automatic Adaptation

Although literary adaptation is not an established NLP task, simplification and summarisation have been part of the technological landscape since pre-neural times [20-22]. Currently, automated methods are increasingly employed in the generation and evaluation of adapted texts [23, 24]. Semantic aspects of text such as the use of metaphors are measurable through specially-trained models for the task [25]. While recent large language models (LLMs) are claimed to competitively engage in creative writing, they remain highly dependent on human prompts [26]. Moreover, their use for the purpose of writing literature is controversial on ethical grounds. Still, the application of LLMs and other contemporary NLP methods in narrowly defined aspects of the creative writing process has proven promising. Kreminski and Martens [27] discuss LLMs' potential to assist in cases of "writer's block", and tools like Story Centaur propose isolated sentences based on previous context and a "magic word" [28].

## 3. Computer-Assisted Textual Adaptation

In the present section, we will present how we selected/produced the texts used in the reader experience survey in relation to the three investigated adaptation settings: baseline, enriched, and style transfer. Table 1 visualises all derived texts.

**Table 1.** Literary extracts and adaptation types included in the study.

| Original work                     | Level | Adaptation                             |
|-----------------------------------|-------|----------------------------------------|
| <i>Moby-Dick</i>                  | A2    | baseline                               |
|                                   |       | enriched                               |
|                                   |       | style transfer                         |
|                                   | B2    | baseline<br>enriched<br>style transfer |
| <i>The Picture of Dorian Gray</i> | B1    | baseline                               |
|                                   |       | enriched                               |
|                                   |       | style transfer                         |
|                                   | B2    | baseline<br>enriched<br>style transfer |

We used the CEFR-SP assessment model [29] to ensure the stability of the extracts' levels at different points of the pipeline. We averaged the sentence-level predictions that the model provides, using word counts as weights. The initial estimations differed from the professionally assigned ones - the B2 version of *Moby-Dick* was estimated as B1, and the rest of the extracts - as A2. We consider this mismatch not to constitute a significant limitation given our purpose to monitor level stability.

### 3.1 Baseline Adaptations

In order to ensure a representative number of survey participants, we focused on texts for learners of English. We selected two works that are canonically used in adaptation - *The Portrait of Dorian Gray* by Oscar Wilde and *Moby-Dick* by Herman Melville - and located two graded readers at different CEFR levels for each work. We manually selected extracts from each of the adaptations of between 500 and 700 words, ensuring that they comprise full episodes in the work's plot while not being longer than necessary for the evaluation of reader experience. In order to facilitate intertextual mapping, we attempted to select related episodes from the implied literary works; however, this was not achievable in the case of *Moby-Dick*'s adaptations due to the large differences in textual compression.

### 3.2 Enriched Adaptations



Based on the stylistic specificities of the two investigated original literary works as defined in Nikolova-Stoupak et al [1], we enriched the baseline adaptations through increasing the original works' prominent characteristics and overrepresented vocabulary. Additionally, we inserted a portion of original sentences. We laid out the following guidelines: to retain the extract's CEFR level, to base the number of interventions on the extract's length and the proportions observed in relation to the original text, and not to force modifications when they are not semantically suitable.

**Increase of prominent characteristics.** We first sought to reinforce five linguistic characteristics previously identified as particularly prominent in each original novel [1] (Table 2). For features expressed as average frequencies, the expected number of instances in an extract was estimated proportionally as  $E = NR/100$ , where  $N$  is the extract length and  $R$  - the feature's mean frequency per 100 tokens in the original novel. The extracts were then edited manually towards these values where this could be done naturally. For instance, the phrase "he put the statue" was altered as "he put *that* statue" in order to introduce a demonstrative determiner. Features characterised by high maximum values or standard deviations were treated as potentially clustered rather than uniformly frequent: existing clusters could therefore be reinforced, but new ones were not artificially introduced. Existing values above the target were also retained. Because the extracts are short, the aim was deliberately to emphasise rather than precisely reproduce the distributions found across the complete works.

**Table 2.** Prominent linguistic features reinforced in the adaptations. Avg: average; max: maximum; SD: standard deviation (aggregators of the feature in question)

| Work                              | Prominent characteristics                                                                                                       |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| <i>The Picture of Dorian Gray</i> | demonstrative pronouns (SD); past tense (SD); predictive modals (max); copular constructions (max); present tense (SD)          |
| <i>Moby-Dick</i>                  | seem/appear (max); demonstrative determiners (avg); by-passives (max); attributive adjectives (avg); non-finite fragments (max) |

**Increase of overrepresented vocabulary.** We similarly considered the ten lexical items previously found to be most overrepresented in each novel relative to a general literary frequency list [1] (Table 3). Where contextually appropriate, these words were introduced into the adaptations (e.g. "It was dreadful." became "It was hideous"). Closely related morphological forms were used where these produced more natural language (e.g. "mate" for "mates"). Words already sufficiently represented were left unchanged.

**Table 3.** Ten most overrepresented lexical items in the two original novels.

| Work                              | Most overrepresented lexical items                                                 |
|-----------------------------------|------------------------------------------------------------------------------------|
| <i>Moby-Dick</i>                  | whale; aye; mast; sir; mates; rigging; carpenter; jaw; aloft; aft                  |
| <i>The Picture of Dorian Gray</i> | duchess; sir; painter; portrait; coloured; gilt; studio; hideous; monstrous; valet |

**Insertion of original sentences.** To identify salient sentences from the original works to reintroduce, we used surprisal as a proxy for linguistic predictability. The surprisal of a word  $w_i$  is defined as  $-\log_2 P(w_i | w_1, \dots, w_{i-1})$ , higher values corresponding to less predictable language [30]. We initially hypothesised that memorable literary quotations would exhibit higher surprisal. To test this, we collected 100 popular one-sentence quotations for each novel from [www.goodreads.com](http://www.goodreads.com), as ordered by reader likes. Contrary to our hypothesis, the models GPT-2 Medium and BioGPT consistently assigned *lower* length-adjusted surprisal to famous quotations, with significant small-to-moderately large effects ( $|d| = 0.436 - 0.754$ ). A possible explanation for this correlation may be readers' preference for syntactically and lexically simple, logical sentences. TinyStories-33M, trained on child-oriented synthetic language, did not reproduce the pattern. Based on this investigation, we went on to use low surprisal as a feature for identifying potentially suitable sentences for enrichment. Using BioGPT, which showed the clearest and most consistent pattern, we calculated mean token surprisal for sentences from the sections of the original novels corresponding to the adapted extracts, shortlisting 20 lowest-surprisal sentences for each



episode. Candidates were manually inspected for contextual suitability and their estimated CEFR level was required to be no more than one level above that of the corresponding extract. Additions were to be stopped and the last one rejected upon alteration of the extract's CEFR level.

### 3.3 Style Transfer Adaptations

For style transfer, we used TinyStyler-Sim, a variant of TinyStyler that favours semantic preservation. The target style was derived from the unabridged literary works rather than from the graded adaptations. For each novel, eight random passages of approximately 80--160 words were encoded and combined into a target-style embedding using the authorship-style model AnnaWegmann/Style-Embedding. Style transfer was then performed sentence by sentence at a style strength of 0.25. Eight candidate paraphrases were generated per sentence. Candidates were discarded when: they were identical to the source, they were very short or very long compared to the source, or the cosine similarity with the source (calculated using the model all-mpnet-base-v2) was below 0.88. Three similarity levels were determined for the remaining candidates according to cosine similarity: over 0.95, 0.91-0.95, and 0.88-0.91. Table 4 lists examples of proposed sentences by similarity level. Additionally, only candidates whose cosine similarity to the target style exceeded that of the source sentence were retained. Candidate selection was probabilistic, balancing the three semantic-similarity levels with the magnitude of the stylistic gain. Replacements were executed at random. After each replacement, the CEFR level of the extract was recalculated, upon which replacements that altered the level were rejected, and the procedure continued from the last accepted version until at least six sentences had been altered.

**Table 4.** Examples of accepted style transfer outputs from the three defined semantic similarity levels.

| Band         | Original sentence                                                  | Transferred sentence                                               | Similarity |
|--------------|--------------------------------------------------------------------|--------------------------------------------------------------------|------------|
| Conservative | You came—oh, my beautiful love!—and you freed my soul from prison. | You came—oh, my beautiful love! And you freed my soul from prison. | 0.978      |
| Moderate     | Dorian Gray became pale as he watched her.                         | Dorian Gray paled as he watched her.                               | 0.945      |
| Exploratory  | What could they know of love such as ours?                         | What could they know of such love?                                 | 0.896      |

## 4. Reader Experience Survey

The survey was conducted using Google Forms. Respondents were firstly asked to estimate their level of English as A2, B1 or B2. Then, a single text was allocated based on the indicated level. Each of the four adaptation extracts was presented in three versions (baseline, enriched and style transfer), giving a total of 12 variants of the survey. As level B2 was associated with both of the original literary works, respondents were assigned one of six variants for this level, while there were three variants each for the other two levels. Texts were followed by three text-specific multiple choice comprehension questions. The next two questions addressed reader appreciation, asking the participants to evaluate the read extract's interest and style on a scale of 1 to 5. Participants were also asked whether they recognised the work in question and prompted to write its title in case that they did. Finally, participants were asked to answer in free text what they think will happen next in the story. Respondents were not provided information about the nature of the adaptation version they were assigned. Versions were rotated manually in an attempt to provide an equal number of replies per adaptation version. However, balance was not achieved in relation to other variables, such as the number of respondents per CEFR level. There were no other preconditions for participants safe for their status as learners of English at one of the indicated levels. The survey was anonymous, but participants were asked to provide their email address in case they would like to be included in a draw for five 5 € Amazon vouchers.

## 5. Results

A total of 45 English learners responded to the survey: the majority were at level B2 (31), followed by B1 (8) and A2 (6). 15 learners responded to each of the three investigated adaptation types. See Table 5 for detailed participant distribution. Table 6 reports participant-level mean scores for the three adaptation types. "Comprehension" represents the percentage of correct answers across the



comprehension questions. The scalar ratings associated with general interest and style evaluation were converted to percentages of the maximum score, and the two values were averaged into a single variable referred to as “appreciation”. Baseline texts are associated with the highest comprehension value, while enriched texts and style transfer demonstrate an equal, slightly lower value. Appreciation is highest for style transfer and lowest for enriched adaptations.

**Table 5.** Distribution of participants across texts, CEFR levels, and adaptation types.

| Book                              | Level | Adaptation type | N         |
|-----------------------------------|-------|-----------------|-----------|
| <i>Moby-Dick</i>                  | A2    | baseline        | 1         |
|                                   | A2    | enriched        | 1         |
|                                   | A2    | style transfer  | 4         |
|                                   | B2    | baseline        | 5         |
|                                   | B2    | enriched        | 5         |
|                                   | B2    | style transfer  | 5         |
| <i>The Picture of Dorian Gray</i> | B1    | baseline        | 3         |
|                                   | B1    | enriched        | 3         |
|                                   | B1    | style transfer  | 2         |
|                                   | B2    | baseline        | 6         |
|                                   | B2    | enriched        | 6         |
|                                   | B2    | style transfer  | 4         |
| <b>Total</b>                      |       |                 | <b>45</b> |

**Table 6.** Mean reader-evaluation scores by adaptation type. Compr: comprehension; Appr: appreciation.

| Adaptation     | N  | Compr. | Appr. |
|----------------|----|--------|-------|
| baseline       | 15 | 86.7   | 73.3  |
| enriched       | 15 | 84.4   | 81.3  |
| style transfer | 15 | 84.4   | 64.0  |

To gain further insight, we went on to investigate the simultaneous relationships between adaptation types and 1) CEFR level and 2) literary text. Style transfer adaptations demonstrated the highest uniformity in terms of comprehension between levels (83.3%-85.2%) and were on par with enriched adaptations for literary texts (81.5%-83.3%). In contrast, variation was strikingly high among levels for enriched texts (33.3%-90.9%). For level A2 and *Moby-Dick*, appreciation was highest for baseline adaptations, whilst the measure was highest for style transfer adaptations in the rest of cases. At interpreting these results, one should keep in mind that CEFR levels are unevenly represented within the survey data. Moreover, there is an inherent link between CEFR level and literary work as levels A2 and B1 are represented by a single work each.

In order to examine the association between experimental and participant characteristics, we calculated correlation ratios ( $\eta$ ) for the nominal variables (adaptation type and literary work) and Spearman's rank correlations ( $\rho$ ) for the ordinal CEFR variable. Table 7 reports these associations at the individual-participant level. The largest association observed was between adaptation type and appreciation ( $\eta = 0.372$ ), whereas the association between adaptation type and comprehension was very small ( $\eta = 0.044$ ). CEFR level showed a positive association with comprehension ( $\rho = 0.365$ ), whereas the association between CEFR level and appreciation was small and negative ( $\rho = -0.138$ ). The positive association of CEFR level with comprehension was statistically significant ( $p=0.014$ ), whereas its association with appreciation was not ( $p=0.366$ ). Comparatively small associations were observed between literary work and both comprehension and appreciation ( $\eta = 0.083$  and  $\eta = 0.115$ , respectively).

Table 8 reports the answers to the survey questions pertaining to participants' recognition of the underlying literary work. Close to a quarter of participants recognised the work, the rate being higher for *The Picture of Dorian Gray*: 25.0% versus 19.0% for *Moby-Dick*. Only one participant claimed to recognise the work but went on to name it incorrectly. Recognition was highest for baseline adaptations, followed by style transfer and enriched.



**Table 7.** Associations between adaptation type, CEFR level, literary work, and reader evaluations. Bold values highlight the two largest associations observed in the sample.

| Predictor       | Outcome       | Measure | Association  |
|-----------------|---------------|---------|--------------|
| Adaptation type | Comprehension | $\eta$  | 0.044        |
| Adaptation type | Appreciation  | $\eta$  | <b>0.372</b> |
| CEFR level      | Comprehension | $\rho$  | <b>0.365</b> |
| CEFR level      | Appreciation  | $\rho$  | -0.138       |
| Book            | Comprehension | $\eta$  | 0.083        |
| Book            | Appreciation  | $\eta$  | 0.115        |

**Table 8.** Recognition of the source literary work. “Yes/ No/ Maybe” denote the respective participant answers. “Correct” indicates participants who correctly named the work. Values are percentages (number of participants).

| Group                             | N  | Yes        | Maybe     | No         | Correct    |
|-----------------------------------|----|------------|-----------|------------|------------|
| Overall                           | 45 | 24.4% (11) | 20.0% (9) | 55.6% (25) | 22.2% (10) |
| Baseline                          | 15 | 33.3% (5)  | 13.3% (2) | 53.3% (8)  | 33.3% (5)  |
| Style transfer                    | 15 | 26.7% (4)  | 20.0% (3) | 53.3% (8)  | 20.0% (3)  |
| Enriched                          | 15 | 13.3% (2)  | 26.7% (4) | 60.0% (9)  | 13.3% (2)  |
| A2                                | 6  | 33.3% (2)  | 16.7% (1) | 50.0% (3)  | 33.3% (2)  |
| B1                                | 8  | 25.0% (2)  | 37.5% (3) | 37.5% (3)  | 12.5% (1)  |
| B2                                | 31 | 22.6% (7)  | 16.1% (5) | 61.3% (19) | 22.6% (7)  |
| <i>Moby-Dick</i>                  | 21 | 19.0% (4)  | 19.0% (4) | 61.9% (13) | 19.0% (4)  |
| <i>The Picture of Dorian Gray</i> | 24 | 29.2% (7)  | 20.8% (5) | 50.0% (12) | 25.0% (6)  |

## 6. Discussion

Given the proximity of comprehension results per adaptation type and the attested small association between the two variables, it is safe to assume that the presented enrichment procedure does not have a significant effect on reading comprehension in any direction. Conversely, keeping in mind the low associated result combined with the highest measured correlation, appreciation may be affected negatively - observation that comes contrary to our hypothesis. In this context, it is important to note that readers expressed highest appreciation for style transfer adaptations, whose production was associated with higher automatisisation.

Style transfer adaptations were also associated with highest stability between CEFR levels in relation to text comprehension. Comprehension and CEFR level demonstrated the second-largest association in our data in absolute terms, comprehension increasing with proficiency level. Ideally, comprehension should remain relatively stable across CEFR levels when texts are appropriately calibrated to their intended readers. The small differences may therefore suggest that style transfer did not disadvantage particular reader profiles, a desirable property of proficiency-adapted texts [31, 32]. We additionally examined this association, separately by adaptation type. It was strongest for baseline adaptations ( $p = 0.614$ ,  $p = 0.015$ ), followed by enriched adaptations ( $p = 0.381$ ,  $p = 0.161$ ), and was considerably weaker for style transfer ( $p = 0.165$ ,  $p = 0.557$ ). These within-group results should nevertheless be interpreted cautiously given the small sample sizes and particularly small numbers of A2 and B1 participants.

We went on to analyse the open question in which participants were asked to give their opinion regarding what happens next in the story. This question was included in order to evaluate readers' engagement, and the correctness of answers is not of direct relevance. Average answer length, which could be regarded as a proxy to engagement, was highest for style transfer (17.7 words), followed by enriched (12.3) and baseline adaptations (11.1). Further analysis shows that answer length was also highest for level B2, for *The Picture of Dorian Gray*, as well as in cases where readers correctly identified the literary work. 40 out of the 45 answers to the question were meaningful. Non-meaningful answers were: empty, a variation of “I don't know” or unrelated to the question. Answers were still considered meaningful in the following cases: when there were linguistic errors; when the expressed thoughts did not unambiguously include predictions or were vague (e.g. “He left for ever [sic]”; “They will come to a solution”); when an expression such as “I don't know” was followed by meaningful content (e.g. “I don't know, maybe they will both fish together”) and when an answer was clearly far-fetched or humorous



(e.g. “may be they are all going to eat a dish with some human meat ? [sic]”). There were cases in which respondents clearly demonstrated their prior knowledge of the work (e.g. “I think the narrator and Queequeg will become friends and may have more adventures together, possibly involving a whaling voyage.”). One answer, associated with an enriched adaptation, included personal feeling through the expression “I hope”.

## 7. Conclusion

The current study illustrated a computer-assisted procedure for enriching existing literary adaptations for EFL learners that partly restores the original work’s literary style as expressed by prominent quantitative characteristics. Based on evaluation through a reader experience survey, we showed that the achieved enriched adaptations closely retained comprehension levels but reduced reading appreciation as expressed by participants. In turn, the style transfer adaptations we utilised for comparison exhibited positive characteristics linked to high appreciation, more constant comprehension among reader types and weaker positive correlation between reader comprehension and CEFR level.

Our goal was to investigate the potential of the proposed enrichment method in aiding aspects of the reading process. We compared adaptations derived using the method with a baseline as well as an alternative achieved through automatic style transfer. The results confirm potential benefits of original style retention on the reading process and lead us to focus further on the style transfer alternative, possibly combining it with elements of computer-assisted enrichment. In the future, an interface may be proposed to writing and teaching professionals that implements style transfer at different settings, automatically calculates a literary work’s distinguishing features, proposes candidate modifications and, with the help of Large Language Models, assists users with implementation and final editing.

## Limitations

The computer-assisted text enrichment procedure has been handled only in relation to the few relevant characteristics for the texts at hand, and it is currently difficult to apply by non-NLP specialists. A clearer protocol or defined pipeline can be established only upon careful consideration and relevant grouping of a large number of characteristics. Also, the procedure for adapting a complete text will differ from the one for adapting an extract of limited size, where characteristics had to be concentrated in a small space. It is also important to note the limited participant pool and the very limited representation of the A2 and B1 CEFR levels. Also, participants subjectively self-assigned to a CEFR level based on the provided level descriptions and their own experience.

## Ethical Considerations

The survey responses were collected anonymously. Participants could optionally provide an email address to enter a draw for five 5 € Amazon vouchers. No email addresses or other personally identifiable information was made publicly available. The utilised adapted texts are copyright-protected and were therefore not shared in their entirety.

## REFERENCES

- [1] Nikolova-Stoupak, I., Schaeffer-Lacroix, E., & Lejeune, G. (2026). Quantifying Literary Style: A Corpus-Based Study. In *Proceedings of the Seventh International Conference on Computational Linguistics in Bulgaria*. Sofia, Bulgaria.
- [2] Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.
- [3] Godwin-Jones, R. (2018). Contextualized Vocabulary Learning. *Language Learning & Technology*, 22(3), 1–19.
- [4] Restrepo Ramos, F. D. (2015). Incidental Vocabulary Learning in Second Language Acquisition: A Literature Review. *PROFILE Issues in Teachers’ Professional Development*, 17(1), 157–166.
- [5] Wan-a-rom, U. (2008). Comparing the Vocabulary of Different Graded-Reading Schemes. *Reading in a Foreign Language*, 20(1), 43–69.
- [6] Hill, D. R. (2013). Graded Readers. *ELT Journal*, 67(1), 85–125.



- [7] Dickinson, P. (2017). Effects of Extensive Reading on EFL Learner Reading Attitudes. In *Proceedings of the 21st Conference of the Pan-Pacific Association of Applied Linguistics* (pp. 28–35). Tamkang University, Taiwan.
- [8] Hutcheon, L. (2006). *A Theory of Adaptation*. New York: Routledge.
- [9] Thiel, E. (2013). Downsizing Dickens: Adaptations of *Oliver Twist* for the Child Reader. In A. Müller (Ed.), *Adapting Canonical Texts in Children's Literature* (pp. 143–162). London: Bloomsbury Academic.
- [10] Claridge, G. (2005). Simplification in Graded Readers: Measuring the Authenticity of Graded Texts. *Reading in a Foreign Language*, 17(2), 144–158.
- [11] Canning, N. A., & Nelson, M. E. (2018). Unraveling the Affordances of *Silas Marner* in a Japanese University EFL Context. *L2 Journal*, 10(2).
- [12] Wilde, O. (1994). *The Picture of Dorian Gray* (L. A. Hill, Adapt.; B2). Oxford Progressive English Readers. Oxford UP.
- [13] Wilde, O. (2020). *The Picture of Dorian Gray* (V. Heward, Adapt.; B1). Black Cat. CIDEB.
- [14] Melville, H. (2006). *Moby-Dick* (K. Burke, Adapt.; A2). Penguin Readers. Pearson.
- [15] Melville, H. (2020). *Moby-Dick* (F. Mackenzie, Adapt.; B2). Penguin Readers. Penguin Books.
- [16] Wellek, R., & Warren, A. (1949). *Theory of Literature*. New York: Harcourt, Brace and Company.
- [17] Hatzfeld, H. A. (1953). *A Critical Bibliography of the New Stylistics Applied to the Romance Literatures, 1900–1952*. University of North Carolina Studies in Comparative Literature, No. 5. Chapel Hill: University of North Carolina Press.
- [18] DuBay, W. H. (2004). *The Principles of Readability*. Costa Mesa, CA: Impact Information.
- [19] Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge UP.
- [20] Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159–165.
- [21] Devlin, S., & Tait, J. (1998). The Use of a Psycholinguistic Database in the Simplification of Text for Aphasic Readers. In *Proceedings of the International Conference on Artificial Intelligence (IC-AI'98)* (pp. 161–167). Las Vegas, NV, USA: CSREA Press.
- [22] Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)* (pp. 168–177). New York, NY, USA: Association for Computing Machinery.
- [23] Freyer, N., Kempt, H., & Klöser, L. (2024). Easy-Read and Large Language Models: On the Ethical Dimensions of LLM-Based Text Simplification. *Ethics and Information Technology*, 26(3).
- [24] Kew, T., Chi, A., Vásquez-Rodríguez, L., Agrawal, S., Aumiller, D., Alva-Manchego, F., & Shardlow, M. (2023). BLESS: Benchmarking Large Language Models on Sentence Simplification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 13291–13309). Singapore: ACL.
- [25] Wachowiak, L., Troiano, E., & Klinger, R. (2022). Drum Up Support: Systematic Analysis of Image-Schematic Conceptual Metaphors. In *Proceedings of the 3rd Workshop on Figurative Language Processing*. Abu Dhabi: ACL.
- [26] Shanahan, M., & Clarke, C. A. M. (2023). Evaluating Large Language Model Creativity from a Literary Perspective. *arXiv preprint arXiv:2312.03746*
- [27] Kreminski, M., & Martens, C. (2022). Unmet Creativity Support Needs in Computationally Supported Creative Writing. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)* (pp. 74–82). Dublin, Ireland: ACL
- [28] Swanson, B., Mathewson, K., Pietrzak, B., Chen, S., & Dinalescu, M. (2021). Story Centaur: Large Language Model Few Shot Learning as a Creative Writing Tool. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 244–256). Online: Association for Computational Linguistics.
- [29] Arase, Y., Uchida, S., & Kajiwara, T. (2022). CEFR-Based Sentence Difficulty Annotation and Assessment. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 6206–6219). Abu Dhabi, United Arab Emirates: ACL.
- [30] Hale, J. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*. ACL.
- [31] Stenner, A. J., & Stone, M. H. (2022). Does the Reader Comprehend the Text Because the Reader Is Able or Because the Text Is Easy? In *Explanatory Models, Unit Standards, and Personalized Learning in Educational Measurement* (pp. 133–152). Springer.
- [32] Sheehan, K. M. (2017). Helping Students Select Appropriately Challenging Text: Application to a Test of Second Language Reading Ability. *ETS Research Report Series*.