Simplified for Whom? On the Differences between Language Learner Texts and Simplified Language

Miriam Anschütz[1], Moritz Dittmeyer[2], Lisa Stengel[2], Georg Groh[1],
[1]Technical University of Munich, Germany, [2]Goethe-Institut, Germany,
Correspondence: miriam.anschuetz@tum.de

Abstract

Simplified language comes in various forms and targets diverse readers, including language learners and people with disabilities. These groups have different backgrounds and needs, reflected in language standards such as the Common European Framework of Reference (CEFR) and Easy Language (EL) rule sets. However, literature often uses these varieties interchangeably or claims that proposed approaches are useful to a broad range of readers. Therefore, this paper investigates the linguistic differences and overlaps between CEFR-level learner texts and simplified language varieties. We extract text samples from a German learning platform and conduct a corpus-based comparison of linguistic features across A1, A2, B1, EL, EL+, and Plain Language texts. We then test the readability of sample texts with members of the EL target group. Our findings show that Plain Language indeed overlaps with language learner texts, but that CEFR-level texts do not provide an adequate basis for Easy Language. We therefore caution researchers to be precise about which target groups their work claims to address.

1 Introduction

Accessible language comes in different forms. In the German-speaking context, two versions are particularly relevant: Plain Language (PL) and Easy Language (EL). PL targets a broad audience of non-expert readers, while EL is simpler, follows explicit rule sets, and is aimed specifically at people with cognitive disabilities or learning difficulties (Maaß, 2020; inclusion europe, 2021).

A third body of simplified texts comes from language teaching. Texts produced for learners at the A1 and A2 levels of the Common European Framework of Reference (CEFR; Council of Europe (2001)) use controlled vocabulary and short sentences to match beginner proficiency. While these texts are superficially similar to PL and EL, their consumers have different backgrounds and needs. An example is shown in Figure 1: English loanwords should not be used in simplified language because their pronunciation and meanings are often unfamiliar to inexperienced readers. In contrast, language learners may benefit as they are more proficient in English than in German.

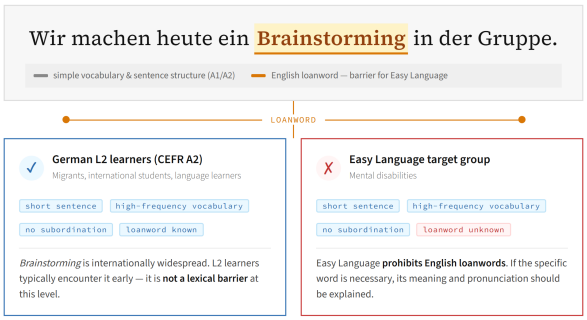


Figure 1: The example (EN: "We're doing a brainstorming session in the group today.") satisfies core simplification criteria (short sentence, high-frequency vocabulary, no subordination). Yet, the English loanword Brainstorming is not suitable for the EL target group.

Despite these distinct purposes, the registers are frequently conflated in NLP and computational linguistics research. A recurring pattern is to frame a system or dataset as targeting accessibility for people with cognitive disabilities, while using CEFR-level texts as the actual data (Battisti et al., 2020; Anschütz et al., 2023; Spring et al., 2021). This is problematic because treating texts that meet only CEFR standards as equivalent to EL can lead to materials that are not actually accessible to people with learning difficulties. Moreover, many resources for measuring the readability of a text or someone's language proficiency align with CEFR standards rather than targeting more general literacy scales, for example, by the OECD (2024).

Therefore, this paper sets out to investigate where A1, A2, B1, PL, and EL actually overlap and differ. For this, we performed two experiments:

First, a corpus-based analysis of German texts developed in collaboration with the Goethe-Institut, an institute of the Federal Republic of Germany with a global presence, facilitating international cultural exchange and promoting access to the German language. In a second step, people from the EL target group evaluated A1/A2 texts to determine whether they meet the accessibility standards required for readers with learning difficulties. Based on these results, we urge researchers and practitioners to be precise about the target groups their work addresses.

2 Background

2.1 Common European Framework of Reference (CEFR)

The CEFR (Council of Europe, 2001) targets non-native speakers and provides a reference for the learning, teaching, and assessing of European languages. Thus, language-learning curricula, textbooks, and examinations aligned with the CEFR ensure coherence and comparability. The different proficiency levels target various language activities like reception, interaction, or production, and cover modalities such as written text or spoken language. A detailed description for each activity, modality, and language level is provided in the CEFR Companion Volume. For example, language learners at a A2 level *"Can identify specific information in simpler material they encounter such as letters, brochures and short news articles describing events*" for the activity Reception → Reading comprehension → Reading for information and argument (Council of Europe, 2020, p. 57).

For our experiments and comparisons, we consider only activities involving reading and ignore oral communication or text production, as these are not typically used in text simplification.

2.2 Simplified language

Simplified language is an accessibility-enhanced version of language that tries to facilitate the understanding of written information for inexperienced readers or people with disabilities. It comes in different versions with varying levels of simplification (see Figure 2) and rulesets. Plain Language (PL; DE: Einfache Sprache) targets lay people and, thus, avoids complex structures and expert terms. In practice, it is helpful for a broad group of people, such as people with low literacy or language learners. In contrast, Easy Language (EL; DE: Leichte Sprache) targets people with mental disabilities specifically.

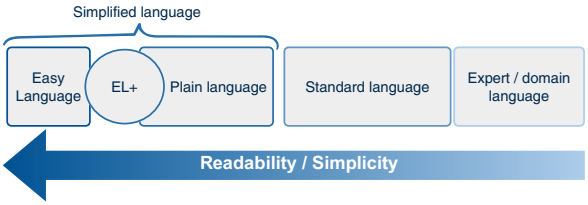


Figure 2: Different language levels in German. Figure adapted from Maaß (2020, Figure 1).

It follows strict guidelines outlined in the DIN SPEC 33429:2025-03, such as using only short, high-frequency words that align with its readers' everyday experiences. Sentences should be short, with only one statement per sentence and linebreaks after each sentence, and have a simple subject-verb-object order (Deutsches Institut für Normung e. V., 2025). As an intermediate between these two levels, Easy Language Plus (EL+) aims to balance the high comprehensibility of EL with the high acceptability and low stigmatization of PL. Therefore, it is closer to standard German and allows for a more natural sentence structure, yet remains easier to understand than PL (Maaß, 2020).

2.3 Comparison of rule sets

The CEFR descriptors explicitly reference "simple language" when describing skills of A1 or A2 level. However, simple language is never formally introduced and is also not a formalized version of simplified language, lacking a clear ruleset.

The CEFR often aligns the skill descriptions with common tourist activities, e.g. "Can understand information given in illustrated brochures and maps (e.g. the principal attractions of a city)" (Council of Europe, 2020, A2 "Reading for information and argument", p. 57) or "Can follow short, simple directions (e.g. to go from X to Y)" (Council of Europe, 2020, A1 "Reading instructions", p. 58). This is in contrast with Easy Language that aims to target all aspects of life (Deutsches Institut für Normung e. V., 2025, p. 11).

3 Experiment 1: Corpus analysis

The goal of our first experiment is to compare linguistic statistics across texts representative of the respective language level. For the CEFR texts, we extracted text samples from the Goethe-Institut's online learning platform, specifically from the A1, A2, and B1 courses. These texts are created by

Level	Goethe-Institut			German4All		
	A1	**A2**	**B1**	**EL**	**EL+**	**PL**
Num.	92	145	244	150	150	150

Table 1: Number of samples per language level.

language-learning experts at the Goethe-Institut and are specifically tailored to the course's language level and content. For each level, we used texts from the last three out of 18 chapters for the respective course level, and the first three chapters from the next course. Further details about which texts were selected are presented in Appendix A.

For the simplified texts, we took the *corrected* subset from the German4All corpus (Anschütz et al., 2025). This corpus contains aligned samples in EL, EL+, and PL, manually reviewed by EL experts. Overall, we compare 931 samples across six distinct language levels. The detailed sample counts are shown in Table 1.

We compare the average sentence length and the average word length, measured by the number of syllables. These features are most commonly used in readability formulae, such as FRE (Amstad, 1978) or the Wiener Sachtextformeln (Bamberger and Vanacek, 1984), and thus serve as good indicators of the samples' readability. The word length can also be seen as a proxy for word-level difficulty, as German compound words can be very long and thus harder to read. Similarly, we calculate the type-token ratio as the proportion of unique tokens among all tokens, and the distribution of word frequency bands.

To measure the texts' cohesion, i.e., logical flow, we also calculate the number of nouns that overlap between consecutive sentences (Lohmann et al., 2025). In addition, we measure how nested the texts are, i.e., how much cognitive load readers need to resolve sentence-level references or subclauses. For this, we measure the average number of tokens (distance) between dependent words according to the dependency tree (Liu, 2008), as well as the maximum dependency tree depth per sample (Anschütz and Groh, 2022). All annotations are obtained with SpaCy (Honnibal et al., 2020) and TextDescriptives (Hansen et al., 2023).

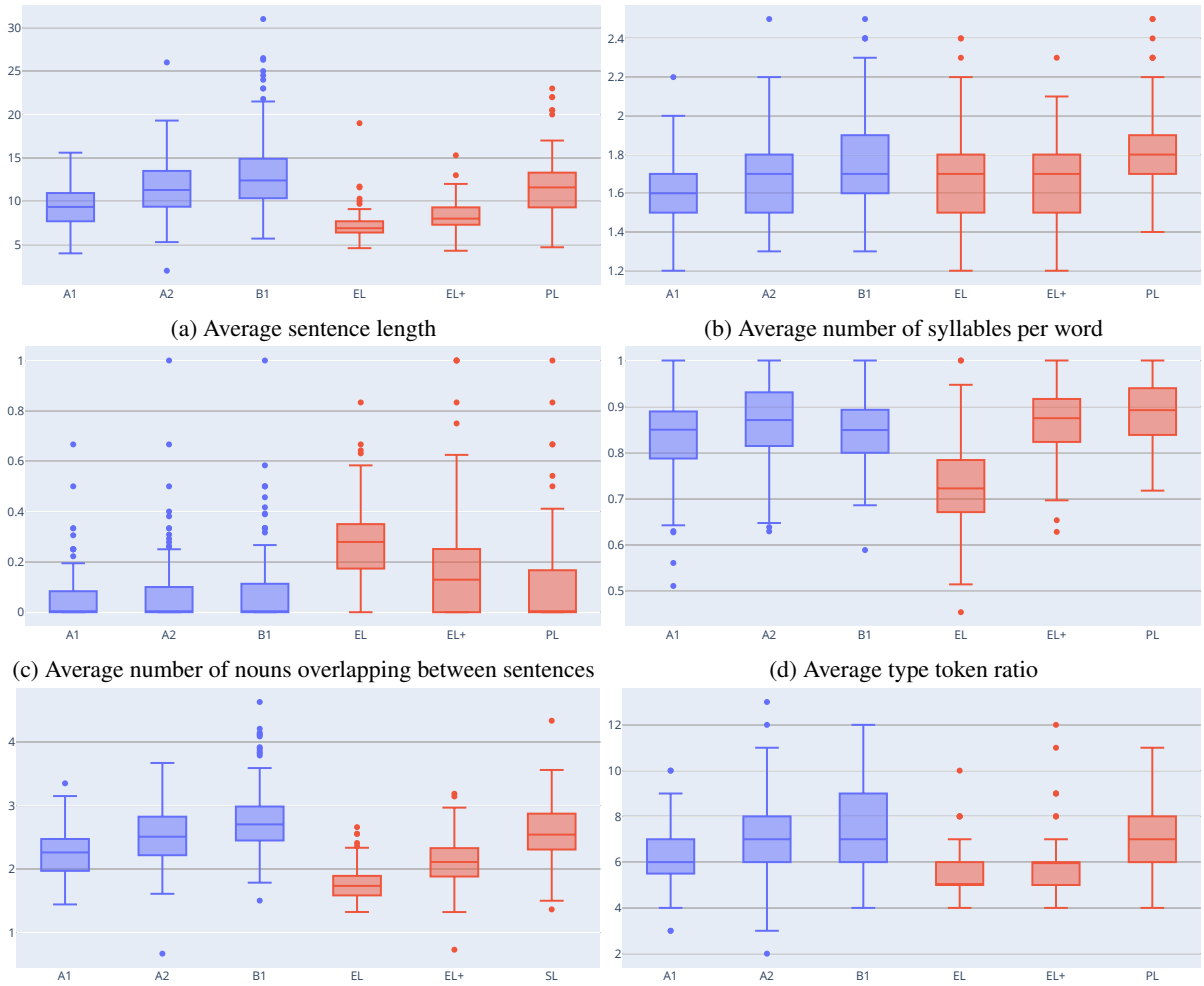
The results are presented in Figure 3. All features show continuous trends across the different readability levels, A1-B1 and EL-PL, indicating that they indeed measure simplicity-related aspects. The length-related features (Figures 3a and 3b) show an alignment of PL with CEFR levels, but words in simplified language tend to contain more syllables in general. For type-token-ratio (Figure 3d) and dependency structures (Figures 3e and 3f), PL samples exhibit a similar pattern as CEFR level, specifically A2 and B1. In contrast, the Easy Language samples are a clear outlier with shorter dependency paths, less nesting, and less variation in the type-token ratio. Moreover, the simplified language samples show a higher number of overlapping nouns between sentences (Figure 3c), indicating a higher cohesion. This is especially relevant for people with mental disabilities and inexperienced readers, who need to be guided through the text, and thus, implicit intermediate steps are made explicit in the reading flow.

Finally, we looked at the distribution of the word frequency bands of words and lemmata in the texts, obtained as zipf_frequency by the Python wordfreq package (Speer, 2022). We define the bucket of very rare words with $z < 2$, the very common words as $z \geq 5$, and the remaining buckets continuously in between. Figure 3g shows the distribution of frequency bands per language level. All three CEFR levels are very similar with very few low-frequency words overall, while Easy and Plain language seem to have more low-frequency words. A more detailed analysis of the observation is presented in Appendix B. We find that most of the low-frequency words in simplified language are proper names that were repeated multiple times in one sample (e.g., "Evoque", a car model), while the low-frequency words in the Goethe-Institut data are indeed long, complex words like "Lieblingsjahreszeit" (EN: favorite time of the year).

Overall, this corpus analysis paints a mixed picture: Plain language shows a high overlap with CEFR levels, but simpler language versions in the form of Easy Language (plus) are more distinct.

4 Experiment 2: Target group evaluation

One of the core features of Easy Language texts is that they should be examined by members of the target group themselves to ensure they are easy to read. Therefore, we worked together with an EL test group and two professional translators to test text samples from the Goethe-Institut courses. The testing was conducted with the help of the FÜP - FortSchritt Übersetzungs- & Prüfbüro für Leichte Sprache. In total, we had two testers who were compensated for their efforts at an hourly rate



(a) Average sentence length

(b) Average number of syllables per word

(c) Average number of nouns overlapping between sentences

(d) Average type token ratio

(e) Average token distance of dependent tokens

(f) Maximum dependency tree depth

(g) Word frequency bands

Figure 3: Linguistic comparison between language learner texts from the Goethe-Institut (blue) and texts in simplified language from the German4All corpus (red).

of 14€ per hour and two professional translators as supervisors. The test group usually consists of about 10 people, and we would have liked to include more testers. However, many of the testers have a very low literacy or cannot read at all, which was a strong mismatch with the sample texts. Thus, the supervising translators advised us to include only testers with higher literacy to avoid frustration among the test group.

We selected four texts from the Goethe-Institut's course platform, two from the first chapter of an A2 course (=A1 level), and two from the first chapter of a B1 course (=A2 level). We excluded B1 texts here because the previous experiments showed they were too complex. The texts come from typical reading comprehension exercises, which present

a text and ask students to answer questions about it. These texts have at least multiple coherent sentences and are in a communication format, e.g. an email or a brochure. Typical for language learner texts, all texts address a trip or introduce the reader to popular places in Germany.

For the test session, we printed the exercises and sat together with the testers and two translators who supported them. The two testers took turns reading the texts out loud so we could hear when they struggled with certain words or misread them. We explained the exercise tasks orally, even though the exercises had written instructions, and guided the testers through the questions, e.g by asking follow-up questions why they thought a statement was correct or where they found this information. Our qualitative observations can be summarized as follows:

- The testers immediately stated that these texts were not Easy Language and identified many aspects they would correct if they had to test them as usual, such as the lack of line breaks in the layout. After reminding them that this was no standard test session and these texts don't claim to be Easy Language, they could work with them.
- Both testers were able to read the texts quite fluently, but stumbled over particular words. These include gendered forms ("Mitarbeiterinnen" *EN female employees*) that are not commonly used in EL, specific places ("Norderney", "Mosel"), or loanwords ("Kitesurfen", "Opening-Party").
- Two exercises asked the reader to answer true/false questions about the text, while the other two exercises required content linking, e.g. assigning a headline to a text snippet. These linking exercises were answered with a very high accuracy. For the other exercise, 7/10 questions were answered correctly. However, due to our interactive setup, it is hard to quantify performance exactly because sometimes one tester would give a wrong answer, which the other tester would then correct.
- One exercise required the testers to combine given sentence beginnings and endings. Most sentences were in SVO order and posed no challenge to the testers. However, one sentence used a compound clause ("K believes, A likes parties.") that confused the testers a lot. Similar complex structures like M-dashes or information in parentheses were present in other texts as well.
- The two supervising translators gave us feedback that they perceived the texts as very complex, but were proven wrong and impressed by the performance of the two testers. However, all four testers and translators acknowledged that these texts were far from Easy Language and they would not let them pass in a real EL testing.

5 Conclusion

The corpus analysis and experiment results together paint a nuanced picture. Plain Language shows considerable overlap with CEFR-level texts, but Easy Language is more distinct. Testers read the learner texts largely fluently, but immediately identified them as not being Easy Language and knew exactly what they would change. The recurring problems, such as loanwords and syntactically complex structures, are precisely the features that Easy Language rules prohibit. Hence, we conclude that CEFR-level texts are indeed close to PL. However, EL remains a standalone, accessibility-enhanced version that must not be confused with other simplified or language-learner-friendly texts.

Limitations

We only study the overlaps between simplified language and language learner texts in German. There are two reasons: Only a few languages have strict rule sets for Easy Language, a language version specifically targeted at people with disabilities. English, for example, does not strictly distinguish the different language forms, and thus, would give less insight into the suitability of language learner texts. Moreover, we prioritized human supervision and expert support from the Goethe-Institut and EL testers. This established system of EL test groups is less common in other countries than in Germany. Nevertheless, since the CEFR applies to all European languages, we believe our findings transfer to other languages as well.

Our experiments rely entirely on data from a single language-learning institute (Goethe-Institut). While we worked together with experts in language teaching, acquisition, and the CEFR, this may introduce a bias through their course content. Moreover, the data from Goethe-Institut is from their commercial courses, and thus, cannot be made publicly available. To account for this, we conducted an ablation in Appendix C and observed similar trends in a publicly available CEFR dataset. Similar to the language selection, we focused on expert support

and qualitative findings rather than the quantity of experiments.

Acknowledgements

Thanks to the Goethe-Institut for providing access to their online courses and for supporting this research with additional insights.
This research was partially funded by the German Federal Ministry of Research, Technology and Space (BMFTR) through grant 01IS23069 Software Campus 3.0 (Technical University of Munich) as part of the Software Campus project "LIANA".
The authors used an AI writing assistant in the form of Claude Chat to improve the paper's formulations and phrasing. All suggestions were thoroughly revised by the authors.

References

Toni Amstad. 1978. *Wie verständlich sind unsere Zeitungen?* Ph.D. thesis, Universität Zürich.

Miriam Anschütz and Georg Groh. 2022. TUM social computing at GermEval 2022: Towards the significance of text statistics and neural embeddings in text complexity prediction. In *Proceedings of the GermEval 2022 Workshop on Text Complexity Assessment of German Text*, pages 21–26, Potsdam, Germany. Association for Computational Linguistics.

Miriam Anschütz, Joshua Oehms, Thomas Wimmer, Bartłomiej Jezierski, and Georg Groh. 2023. Language models for German text simplification: Overcoming parallel data scarcity through style-specific pre-training. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1147–1158, Toronto, Canada. Association for Computational Linguistics.

Miriam Anschütz, Thanh Mai Pham, Eslam Nasrallah, Maximilian Müller, Cristian-George Craciun, and Georg Groh. 2025. German4All – a dataset and model for readability-controlled paraphrasing in German. In *Proceedings of the 18th International Natural Language Generation Conference*, pages 390–407, Hanoi, Vietnam. Association for Computational Linguistics.

Richard Bamberger and Erich Vanacek. 1984. *Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Diesterweg.

Alessia Battisti, Dominik Pfütze, Andreas Säuberli, Marek Kostrzewa, and Sarah Ebling. 2020. A corpus for automatic readability assessment and text simplification of German. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3302–3311, Marseille, France. European Language Resources Association.

Adriane Boyd, Jirka Hana, Lionel Nicolas, Detmar Meurers, Katrin Wisniewski, Andrea Abel, Karin Schöne, Barbora Štindlová, and Chiara Vettori. 2014. The MERLIN corpus: Learner language and the CEFR. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1281–1288, Reykjavik, Iceland. European Language Resources Association (ELRA).

Mark Breuker. 2023. *CEFR Labelling and Assessment Services*, pages 277–282. Springer International Publishing, Cham.

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing, Strasbourg.

Council for Cultural Co-operation. Education Committee. Modern Languages Division Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Deutsches Institut für Normung e. V. 2025. DIN spec 33429:2025-03, Empfehlungen für Deutsche Leichte Sprache (Guidance for German Easy Language).

Lasse Hansen, Ludvig Renbo Olsen, and Kenneth Enevoldsen. 2023. Textdescriptives: A python package for calculating a large variety of metrics from text. *Journal of Open Source Software*, 8(84):5153.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Joseph Marvin Imperial, Abdullah Barayan, Regina Stodden, Rodrigo Wilkens, Ricardo Muñoz Sánchez, Lingyun Gao, Melissa Torgbi, Dawn Knight, Gail Forey, Reka R. Jablonkai, Ekaterina Kochmar, Robert Joshua Reynolds, Eugénio Ribeiro, Horacio Saggion, Elena Volodina, Sowmya Vajjala, Thomas François, Fernando Alva-Manchego, and Harish Tayyar Madabushi. 2025. UniversalCEFR: Enabling open multilingual research on language proficiency assessment. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 9703–9755, Suzhou, China. Association for Computational Linguistics.

inclusion europe. 2021. Information for all: European standards for making information easy to read and understand.

Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159–191.

Julian F Lohmann, Fynn Junge, Jens Möller, Johanna Fleckenstein, Ruth Trüb, Stefan Keller, Thorben Jansen, and Andrea Horbach. 2025. Neural networks or linguistic features?-comparing different machine-learning approaches for automated assessment of text quality traits among l1-and l2-learners' argumentative essays. *International Journal of Artificial Intelligence in Education*, 35:1178–1217.

Christiane Maaß. 2020. *Easy language–plain language–easy language plus: Balancing comprehensibility and acceptability*. Frank & Timme, Berlin.

OECD. 2024. *Do Adults Have the Skills They Need to Thrive in a Changing World?: Survey of Adult Skills 2023*. OECD Skills Studies. OECD Publishing, Paris.

Robyn Speer. 2022. rspeer/wordfreq: v3.0.

Nicolas Spring, Annette Rios, and Sarah Ebling. 2021. Exploring German Multi-level Text Simplification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1339–1349, Held Online. INCOMA Ltd.

A Goethe-Institut online course selection criteria

The online courses cover a range of exercises, from gap-fill exercises to practice grammar and vocabulary to text production and listening. The texts selected for our corpus analysis had to fulfill these criteria:

- At least two consecutive sentences. Sentences must not be incomplete
- No exercise texts requiring user inputs (no gap texts or shuffled sentences that should be sorted by the user)
- No special texts like poems
- No chat or other informal communication
- No enumerations/lists

To keep the total lengths consistent, we removed contents in parentheses and split texts with multiple paragraphs into separate samples, each containing a single paragraph.

B Detailed anaysis of frequency band words

Figure 4 shows the top three most frequent words per frequency word band and language level. The low-frequency Goethe samples are most often long German compounds like "Lieblingsjahreszeit" (EN: favorite time of the year), "Überstunden" (EN: overtime hours), or "Tagesmutter" (EN: baby minder). In addition, we observe common names (Mónica or Anabel), company names (Chancenwerk), or places in Germany (Rothenburg). Many of these words like "Lieblingsfächer" (EN: favorite subject) or "Schüleraustausch" (EN: student exchange) are explicitly introduced in the vocabulary of the respective chapters and give an overview of the topics covered.

In contrast, the simplified texts show almost exclusively proper names like "Evoque" (a car model) and "Stilgar" (a fictional character), or ways to introduce them ("heißt", EN: means, or "namens", EN: called). These names are potentially an artifact of the Wikipedia origin of the German4All dataset.

The most common word in the language learner texts is the word "und" (EN: and), consistent across all language levels. For simplified language, this only appears at the EL+ and PL levels, indicating again that EL is restricted to only one statement per sentence and, thus, does not need conjunctions.

C Ablation: publicly available CEFR data

Multiple CEFR-labeled datasets are publicly available (Imperial et al., 2025). However, many of them consist of language learner texts, i.e., texts that were created by language learners themselves, such as the German Merlin corpus (Boyd et al., 2014). These datasets focus on the production activity of the CEFR, whereas we focus on reception and literacy. Thus, we reproduced our linguistic experiments on the ELG corpus by Breuker (2023), which was created to train CEFR-level prediction models.

We investigated the same linguistic features as with the Goethe-Institut texts. The results are presented in Figure 5. The public dataset exhibits similar overall trends and overlaps with simplified language. Only for the type-token-ration, the ELG corpus has lower average values, which are closer to EL.

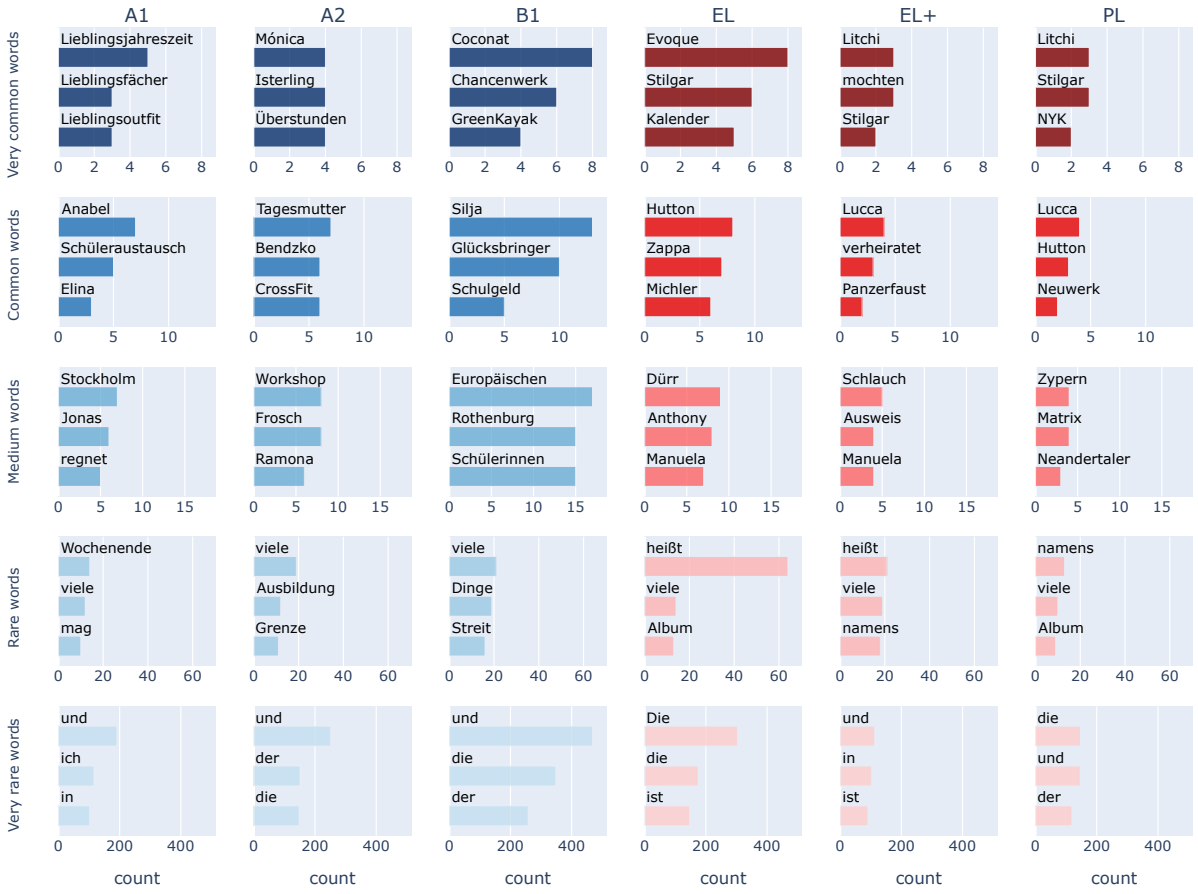
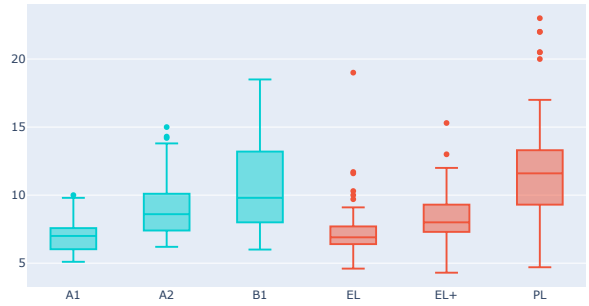
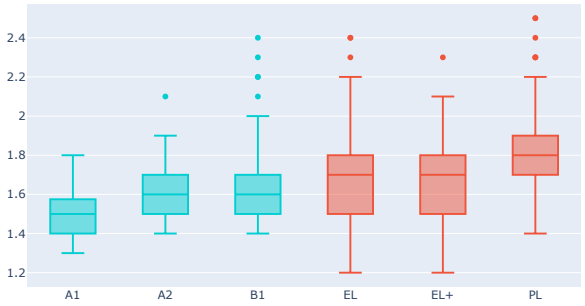


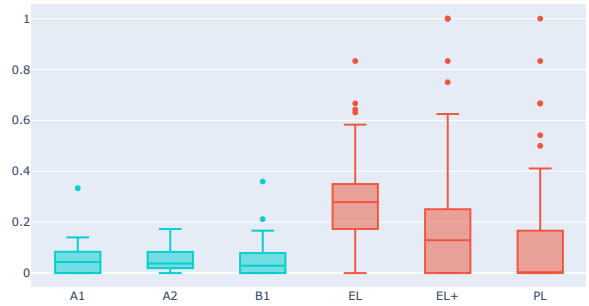
Figure 4: Top 3 words per word frequency band. Blue items are from the Goethe-Institut courses, red items from German4All.



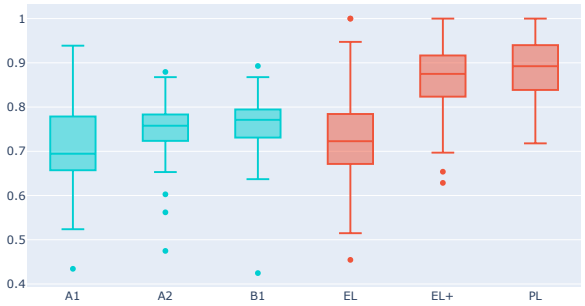
(a) Average sentence length



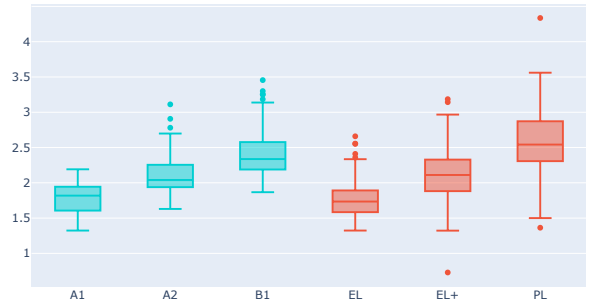
(b) Average number of syllables per word



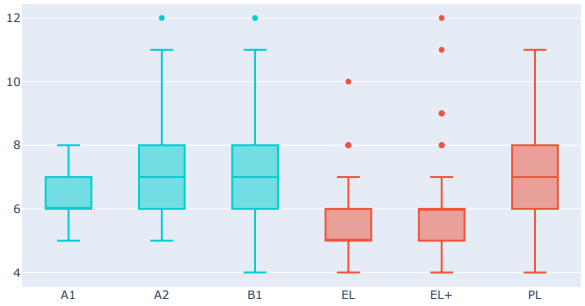
(c) Average number of nouns overlapping between sentences



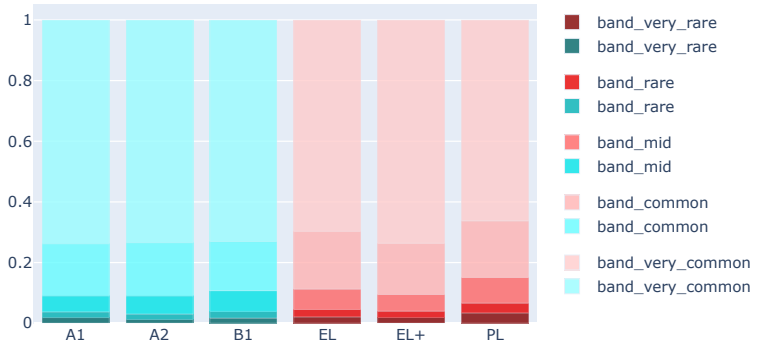
(d) Average type token ratio



(e) Average token distance of dependent tokens



(f) Maximum dependency tree depth



(g) Word frequency bands

Figure 5: Linguistic comparison between language learner texts from the ELG corpus (turquoise) and German4All (red).