

Cross-lingual Mediation: Readability Effects

Maria Kunilovskaya
Saarland University

maria.kunilovskaya@uni-saarland.de

Ruslan Mitkov
University of Lancaster

r.mitkov@lancaster.ac.uk

Eveline Wandl-Vogt

Ludwig Boltzmann Gesellschaft / Austrian Academy of Sciences

eveline.wandl-vogt@oeaw.ac.at

Abstract

This paper explores the readability of translated and interpreted texts compared to the original source texts and target language texts in the same domain. It was shown in the literature that translated and interpreted texts could exhibit lexical and syntactic properties that make them simpler, and hence, easier to process than their sources or comparable non-translations. In translation, this effect is attributed to the tendency to simplify and disambiguate the message. In interpreting, it can be enhanced by the temporal and cognitive constraints. We use readability annotations from the Newsela corpus to formulate a number of classification and regression tasks and fine-tune a multilingual pre-trained model on these tasks, obtaining models that can differentiate between complex and simple sentences. Then, the models are applied to predict the readability of sources, targets, and comparable target language originals in a zero-shot manner. Our test data – parallel and comparable – come from English-German bidirectional interpreting and translation subsets from the Europarl corpus. The results confirm the difference in readability between translated/interpreted targets against sentences in standard originally-authored source and target languages. Besides, we find consistent differences between the translation directions in the English-German language pair.

1 Introduction

Cross-lingual mediation is known to be a specific type of communication, where a message in one language is rendered into the other language either in spoken or written mode. The documents produced as a result of interpreting and translation were shown to have specific linguistic patterns, which makes them distinct from comparable originally-authored (non-mediated) documents. Distinctive features of translated language, usually

captured by statistical analyses, are traditionally referred to as *translationese*. Recent studies based on interpreting data demonstrated that the outcomes of cross-lingual mediation in the written and spoken modes are very dissimilar in their linguistic properties. The term *interpretese* was introduced to refer to the specificity of linguistic choices in interpreting.

In the literature, some trends in translational behaviour (in particular, *simplification*, *explicitation* and *normalisation*) are in part explained by a conscious strategy or subconscious tendency to clarify the communicative intent of the source text for the target audience (Olohan, 2001), and to improve document readability (Puurtinen, 2003). In simultaneous interpreting studies, simplification is viewed as a part of the coping strategy to mitigate the temporal and cognitive constraints imposed by the process. Shlesinger and Ordan (2012) found that simultaneous interpreting emphasises the spoken features of the language, which can contribute to simplification.

Theoretically, in terms of the readability outcomes of the mediation process, the effects of simplification, normalisation and explicitation together can be overcome by *interference*, a tendency which was more recently shown to have the stronger influence on the properties of translation (Evert and Neumann, 2017; Kunilovskaya and Lapshinova-Koltunski, 2020; Chowdhury et al., 2021).

This study aims to estimate the integral impact of cross-lingual mediation on the readability of translated and interpreted texts. For the purposes of this study, we do not make a distinction between readability and complexity/simplicity, assuming that the texts that are easier to read are also less complex, and vice versa, the texts that are simpler on any linguistic level are also easier to read. This assumption is often made in the related literature.

Unlike a lot of previous work, this study relies

on a computational, modelling-driven approach instead of corpus-based and statistical methods. The readability of translated and interpreted segments is captured as dependent on their respective sources and is contrasted with originally-authored data in the target language. The paper presents the results of two experiments comparing mediated segments (a) to their aligned sources and (b) to comparable originals (non-mediated segments) in the target language. Depending on the experimental setup, we train neural models that can distinguish original and simplified versions of the same segment or predict the readability level/score of unaligned segments with unrelated content, and apply these models to translated/interpreted segments aligned with their sources or to target language segments annotated as originals or mediated, respectively. If mediated language is simpler than sources, we expect the model to recognise the target segments in an aligned segment pair as simpler/more readable than the respective source. If mediated language is simpler than comparable non-mediated language in the target language, we expect mediated segments to get lower readability scores, signalling lower complexity.

The results are reported with regard to the mediation mode (written or spoken) and translation direction (German-English and English-German).

The rest of the paper is structured as follows. Section 2 provides a more in-depth analysis of the relations between readability and translational tendencies, especially simplification. In Section 3, we explain the rationale behind the proposed methodology and describe the setup of the two experiments. Subsection 3.1 contains the details on the textual data used to train and evaluate the models, and the testing data to for zero-shot transfer to translations and simultaneous interpreting. Subsection 3.2 offers a description of the modelling process and the measure used to estimate the readability of mediated segments. The results of the experiments are detailed and interpreted in Section 4. We conclude with Section 5, which summarises the study and highlights the findings.

2 Cross-lingual Mediation and Readability

In the context of this research, the specificity of cross-lingual mediation as a type of communicative activity can be viewed as dominated by two opposite trends. On the one hand, mediated lan-

guage is expected to feature increased readability as an integral effect of simplification, explicitation and standardisation. On the other hand, translated and interpreted segments are likely to exhibit traits of interference from the source language, which might make them more difficult to read. Below, we give a brief overview of tendencies in translational behaviour that can be linked to readability.

One of the most discussed trends in both translators' and interpreters' behaviour is *simplification*. It can be described as "a subconscious tendency to simplify the language or message or both of the source" (Baker, 1996, p. 176). Evidence for various types of simplification in translation was reported in a number of corpus-based and computational studies, especially earlier ones (Puurinen, 2003; Corpas Pastor et al., 2008, to name just a few). Readability scores are used as an indicator of simplification on the assumption that easier-to-read texts must be less complex (Williams, 2005; Corpas Pastor et al., 2008; Redelinghuys, 2016). Importantly, most simplification-supporting evidence comes from lexical features, such as those used by (Volansky et al., 2015): TTR, mean word length, syllable ratio, lexical density, mean word rank, and mean sentence length. For example, Redelinghuys and Kruger (2015) reported some evidence in favour of a number of translationese trends, especially simplification, in translated English. However, if simplification is operationalised at the syntactic level, as in (Hu and Kübler, 2021), who looked at news articles in seven languages translated into Chinese, simplification hypothesis does not hold. Kunilovskaya (2023, p. 163, 222) also disproved simplification: sentences in English-to-Russian translations of mass-media texts had a strong tendency to be longer and more complex than in comparable non-translations.

Other tendencies that can be viewed as contributing to increased readability of mediated output are *normalisation* and *levelling-out*. They describe the trend in translation to prefer linguistic expressions that constitute prototypical features of the target language, which might lead to exaggerating these features (over-normalisation) and make translations more similar to each other than originals in either source or target language (Baker, 1996). The empirical support for this claim also varies depending on the language pair. For example, Hu and Kübler (2021) did not find evidence for normalisation in translated Chinese, while the research on translated

Russian suggests ample evidence that translators actually over-emphasise some of the features of the target language (Kunilovskaya, 2023). (Nikolaev et al., 2020) demonstrated that the relative contribution of normalisation depended on the distance between the source and target languages. Translations from structurally-similar languages were found to demonstrate greater conformity to the TL norms and were more predictable, while translations from structurally-divergent source languages contained more non-idiomatic features making them more entropic and unusual in terms of lexical density, mean sentence length, frequencies of conjunctions and passives, etc.

Finally, the tendency to make translations more explicit is usually linked to the potential readability gains. *Explicitation* is described as a trend of the target texts to spell out components that are implicit on the linguistic surface of the source text. The most studied explicitation phenomenon is the increased explicitness of cohesion in translations manifested by a greater number of connectives. Other explicitation phenomena include additional explanatory phrases and deciphered implicatures. Although the findings from the translationese studies are mixed, it is not uncommon to conclude that “compared to original texts, translations tend to be simpler, more standardised, and more explicit” (Toral, 2019).

From the readability perspective, a translationese trend that might work against increasing fluency of the text, is *shining-through* (Teich, 2003) or interference. This term is used to refer to a tendency in translation to follow the source language patterns where possible. Interference drives up the frequencies of the linguistic features shared by the source and target languages and results in unusual awkward wordings. Similarly, the potential effects of mediation (at least in the written mode) sometimes include *sentence lengthening* (Chesterman, 2010). Volansky et al. (2015) found that the mean sentence length in English translated from a number of languages (based on Europarl data) is higher than in comparable English originals, which contradicts the simplification hypothesis.

Investigations into simultaneous interpreting often compare the results to written translation. The findings usually align with (Shlesinger and Ordan, 2012), who concluded that interpreting is associated with a strong simplification effect. For example, (Kajzer-Wietrzny and Grabowski, 2021) estab-

lished reduced lexical variation in English-Polish interpreting based on Europarl speeches. (Dayter, 2018), using transcripts of speeches in the United Nations confirmed simplification (measured by lexical variety and density) for interpreting into English, but not into Russian. (Gast and Borges, 2023) found that there were fewer nouns and more pronouns in interpreted German than in comparable originals and translations, which was explained by the similarity of interpreting and unplanned spoken conversation.

To sum up, previous feature-based research on the properties of mediated language makes it difficult to judge about its comparative complexity/readability, mostly because of the atomistic nature of the features, pointing in opposite directions. Besides, the translational tendencies described above may overlap in terms of their operationalisation and interpretation with regard to readability. On the one hand, explicitation aims to make the text more accessible to the target audience and on the other hand, it increases the sentence length, and well-established readability formulae such as Flesh-Kincaid Reading Ease treat longer sentences as more difficult to read. Nonetheless, the previous research indicated that there might be good reasons to expect professional cross-lingual mediation, especially in interpreting, to increase text readability.

The motivation behind this study is to leverage the power of modern language models and estimate the readability of translated and interpreted texts in a holistic manner, refraining from designing features and detecting specific trends. Translation scholars convincingly hypothesise that clarification of the original communicative intent and disambiguation of the original message are integral parts of cross-lingual mediation, which should improve the comprehensibility of mediated texts for the target audience. If the effects of simplification, explicitation and normalisation were not counteracted by shining-through, it is not unreasonable to suggest that mediated subcorpora should have higher readability scores than non-mediated subcorpora in the source or target language. This claim is stronger for high-quality professional translation/interpreting (used in this study) because professional translators and interpreters (unlike amateurs or students) can be expected to effectively counteract interference and follow the best practices disseminated and established through professional training (Redel-

inghuys, 2016).

3 Methodology

Our methodology is based on fine-tuning a multilingual neural model on English readability-annotated data and applying the resulting models to English-German translational data in a zero-shot transfer scenario. This approach is inspired by (Artetxe and Schwenk, 2019) who reported remarkably strong zero-shot performance for large multilingual models, fine-tuned on English and evaluated on cross-lingual inference and classification tasks. The assumed reliability of zero-shot transfer helps us circumvent the lack of readability data in German.

3.1 Data

This subsection gives a general description of the readability corpus used to train models and translation/interpreting subcorpora used to obtain readability estimates and address our research question of the impact of cross-lingual mediation on the readability/complexity of language.

The parameters of the datasets in Tables 1 and 2 are reported after filtering and preprocessing, including sentence and word tokenisation. Segments shorter than 5 words were filtered out from all datasets.

Readability corpus The fine-tuning tasks are formulated based on the annotated data from the Newsela corpus¹ which was officially obtained for this study under an academic licensing agreement. The corpus contains 1130 news articles simplified by professional editors several times to fit the reading proficiency of children at different grade levels. Each text comes with extensive metadata, including annotations for grade level and Lexile level². This corpus is distributed with a segment-aligned Newsela-based dataset created by (Xu et al., 2015) to facilitate research on text simplification. The dataset maps grade and readability levels of Newsela to 5 versions of the same text (or simplification levels, ranging from the original V0 version to the most simple V4 version). The sentences from all versions of the same text were automatically aligned pairwise (using Jaccard similarity on overlapping word lemmas) resulting in pairs of sentences like V0-V1, V3-V4, V0-V3, etc, where the

¹<https://newsela.com/>

²a quantitative readability metric based on individual words and sentence lengths (see <https://en.wikipedia.org/wiki/Lexile>)

first member of the segment pair is from the more complex document.

For our purposes, we filtered out all segment pairs that did not contain V0 as the more complex version.

A closer inspection of the filtered dataset revealed that the aligned version of the Newsela corpus does not respect sentence splitting and explicitation as simplification strategies. The same original sentence (V0) can be multiply aligned with various parts of the simplified version at the same level. For example, according to the full-text V4 Newsela version (see corporal-punishment.en.4.txt) ORIGINAL (V0) in (1) was rendered as two sentences (given in SIMPLIFIED (V4)). Each of these sentences is paired to the original sentence in the aligned dataset.

- (1) ORIGINAL (V0): *“All studies point to the fact that corporal punishment does not make for a more peaceful, happier child,” she said at the Capitol on Wednesday.*
SIMPLIFIED (V4): *Corporal punishment does not work, she said. It “does not make for a more peaceful, happier child.”*

Non-unique originals within the same simplification level were grouped together, and split bits of their simplified versions were concatenated. The number of segment pairs that were affected by multiple alignments varied across levels from 1370 (V1) to 2930 (V3) (1-2% of the input number of segment pairs for each readability level). This preprocessing step reduced the repetitiveness of V0 and de-noised the association between V0 and the simplified versions.

Table 1 displays the quantitative parameters of the resulting dataset for each type of alignment. It can be seen that the number of V0 segments aligned to VN versions varies across simplified versions. It means that in many cases V0 segments do not have corresponding versions at all simplification levels. In fact, we detected only 290 V0 segments that were aligned to all four VN.

The average segment length for V0 (original document) is about 28 words. The segment length for the simplified versions ranges between 25.1 (V1) and 16.8 (V4).

Importantly, this dataset ignores text-level simplification strategies such as reordering or deleting entire sentences containing unimportant de-

aligned	docs	segs	wc V0	wc VN
V0-V1	1,130	15,1K	440 K	382 K
V0-V2	1,130	17,4K	497 K	389 K
V0-V3	1,129	16,6K	466 K	322 K
V0-V4	1,125	13,2K	365 K	221 K

Table 1: Description of the aligned Newsela for English

tails. It only reflects sentence-level simplification transformations and seems to omit sentence pairs where sentences from two simplification versions coincide. Only 37% of the original V0 sentences present in the full-text documents are found in the aligned version of Newsela (20621 out of 55946 V0 segments). The aligned version is thus a focused sentence-level simplification dataset, which is not diluted by sentence pairs without simplification transformations. It is particularly suited for training models that can distinguish complex and simple sentences.

The aligned dataset was used to construct training data in Experiments 1 and 2, as detailed in Section 3.2.

Translation/interpreting data The results from the translation data in all settings are reported for each translation direction in the English-German language pair and for each mode (written and spoken). EPIC-UdS (Przybyl et al., 2022) and Europarl-UdS (Karakanta et al., 2018) were used as the sources of document- and segment-aligned parallel data, representing *spoken* and *written* mode of cross-lingual mediation, respectively. EPIC-UdS was built from transcribed speeches by MEPs and their transcribed simultaneous interpretation, whereas Europarl-UdS includes officially published speeches and their written translations. Throughout this paper, we will refer to interpreted or translated text as *targets* (tgt) or *mediated*, to the source language segments aligned with mediated text as *sources* (src), and to comparable material in the target language as *originals* (org).

3.2 Experimental setup

We are interested in comparing targets to their sources in the other language (cross-lingual comparison) as well as to domain-comparable originals in the same target language (monolingual comparison). To account for the specificity of each task, we designed two experiments. In Experiment 1 *paired* Newsela segments were used to obtain fine-tuned

		docs	segs	src_wc	tgt_wc
DE-EN	sp	165	2,748	56,720	57,880
	wr	170	2,896	68,358	77,721
EN-DE	sp	137	2,965	66,146	57,020
	wr	170	2,930	72,296	70,327

Table 2: Balanced subsets of bidirectional English(EN)-German(DE) corpus representing spoken (sp) and written (wr) mediation modes by translation direction

models that could be applied to aligned sources and targets and establish which of the two was estimated as more complex. Experiment 2 had a different type of training data, namely single segments that did not share semantic content. The models obtained in this setup were applied to compare targets to the originals in the target language. The paragraphs below provide more details on how we approached each task.

The models in both experiments were trained in the same neural networks framework, simple transformers library (Rajapakse, 2019) build on top of Hugging Face’s Transformers (Wolf et al., 2020). As a starting point in all experiments, we used pre-trained *xlm-roberta-base* model (Conneau et al., 2019) available from the Hugging Face repository. The initial training hyper-parameters were set to the following values: batch 32, epochs 10, learning rate 2e-6, warmup_ratio 0.05. We trained models with an adaptive learning rate, using the AdamW optimizer with the weight_decay 1e-6 to improve regularization and to avoid overfitting. The training process is also equipped with the early_stopping rule (delta: 1e-5, patience: 3). Thirty percent (30%) of the data available for each training process was reserved for validation during training (10%) and for measuring the models’ skill on Newsela-based readability tasks (20%).

Experiment 1: Paired segments The comparison of aligned segments was cast as a binary classification task, based on paired segments as input. We trained four models: one for each set of V0 segments aligned with V1, V2, V3, V4 versions. Fifty percent (50%) of paired instances in each set had V0-VN order and were assigned label 0, while the other 50% had the order of segments swapped (VN-V0) and were labelled 1. We expect that the accuracy of these classifiers would increase for subsets using simpler versions. Additionally, we experimented with a model trained on all V0

paired with a simpler version regardless of the level of simplification (see ‘V0-any’ in Tables 4 and 5). Arguably, it is a more challenging task, with more heterogeneous and noisy instances, where the same V0 can be aligned to several dissimilar versions. These binary classifiers were perfectly balanced across 0 and 1 classes, and we evaluate them on accuracy only, reporting the size of support for each classifier.

These models are applied to predict source-target pairs in a zero-shot manner. Unlike the training data, translations are full-text documents aligned at the segment level. The explored readability relation between sources and targets in the individual segment pairs can vary. To obtain an overall estimate of targets’ simplicity in comparison to their sources, we calculate the ratio of 0 returned for source-target pairs in each mode subcorpus and translation direction. Recall that 0 is used to label pairs where the first member is more complex than the second one (e.g. V0-V4). The higher the ratio of 0, the more target sentences in a document were predicted as simpler by our simplification-aware models.

Experiment 2: Content-unique segments This experimental setup aims to facilitate the comparison between targets and similar originals in the target language. Unlike the source-target case, segments do not share semantic content and cannot be reasonably paired. For this experiment, we constructed balanced subsets of content-unique Newsela segments annotated with V0, V1, V2, V3, V4 simplification version labels. To obtain these subsets, each original segment was aligned with all simplification versions available for it. Each set of aligned versions was represented in the new dataset only once: one item from a set of versions for each original segment was selected in accordance with its version. This ensured that there were no segments with very similar content across the simplification levels. Besides, each segment was matched in the alternative Newsela format to access grade and Lexile readability scores, available for it. The segments that did not match were skipped. The number of skipped segments varied from 2% in V3 to 29% in V1. Table 3 presents a quantitative description of the resulting dataset.

As can be seen from Table 3, the five version-based categories in the unique-content segments datasets are reasonably balanced in size. The dataset spans 11 grade values and 125 Lexile level

	docs	segs	wc	grade	lexile
V0	1,066	3,948	121 K	12	1288
V1	1,061	4,182	101 K	8	1112
V2	1,101	4,369	94 K	6	972
V3	1,095	4,462	83 K	5	834
V4	1,025	3,630	60 K	4	710

Table 3: Parameters of the dataset based on unique-content segments annotated for various complexity levels, including mean scores for grade and lexile level

values, with their averages consistent with the expected decrease in text complexity from V0 to V4.

This dataset was used to train five classifiers, similar to Experiment 1, except the input was single segments annotated for complexity level: four classifiers use [V0, V1], [V0, V2] etc. as categories plus a multiclass classifier on the five labels [V0, V1, V2, V3, V4]. Additionally, the entire dataset was used to train two regressors using grade level and Lexile level as training targets.

4 Results and Discussion

The results are reported by experiment, starting with the evaluation study on Newsela and then, focusing on the outcomes of the zero-shot transfer to translated/interpreted data.

4.1 Experiment 1: Are Targets Simpler than their Sources?

In this experiment, we fine-tuned *xlmr-roberta-base* to recognise the order of more complex and less complex versions of the same segment in a pair with a view to apply trained models to predict source-target pairs. Five models were produced: four models based on the alignment of the original V0 segments with each of the four simplified versions plus a model on the entire dataset, where the simplified member of the pair was not differentiated by simplification level. The evaluation results for the five models on Newsela are listed in Table 4.

model	acc	train	test
V0-V1	0.79	12 K	3,040
V0-V2	0.88	14 K	3,479
V0-V3	0.92	13 K	3,325
V0-V4	0.95	11 K	2,633
V0-any	0.91	50 K	12,477

Table 4: Binary classifiers results on aligned Newsela

The results are intuitively expected: the greater the difference in complexity between the aligned segments, the higher the performance of the classifier. The lowest accuracy of 79% was seen for V0-V1 pairs, which was still higher than the random baseline of 50%. The best results of 95% accuracy were seen on V0-V4, where the original was aligned to its simplest version. A generic model, which was supposed to capture the different complexity of the segments regardless the annotated readability level, returned a high score of 91%.

The preliminary fine-tuning experiments with other values for batch and starting learning rate returned some fluctuation but the overall ranking of models’ performance on Newsela and the relation between mediation modes and translation directions predicted by the models were the same.

The results of the zero-shot transfer of the English Newsela models to the Europarl spoken and written mediation data for German-English and English-German are displayed in Table 5.

direction	model	spoken	written
DE-EN	V0-V1	0.96	0.93
	V0-V2	0.87	0.78
	V0-V3	0.80	0.71
	V0-V4	0.89	0.86
	V0-any	0.61	0.49
EN-DE	V0-V1	0.81	0.59
	V0-V2	0.64	0.45
	V0-V3	0.56	0.34
	V0-V4	0.65	0.50
	V0-any	0.42	0.28

Table 5: Ratio of segment pairs where the target was estimated as more readable than its source by model, translation direction and mediation mode

Table 5 invites a few observations. First, most models predicted targets as easier to read than their sources. The ratio of source-target segment pairs predicted as 0 was over 50%. This ratio was higher for the V0-V1 model, which was fine-tuned on pairs with small complexity contrast between the aligned segments. Interestingly, the V0-V4 model, which was trained on the segment pairs with the greatest complexity contrast, also had a tendency to predict targets as easier than their sources in both modes and translation directions. This might mean that the nature of transformations performed

in translation/interpreting is more similar to simplification transformations typical for V1 and V4. Also, recall that the V0-V1 model had a relatively low accuracy on Newsela (79% vs 95% for the V0-V4 model). Therefore, the predictions by this model might be less reliable.

Second, the ratios of segment pairs where targets were predicted as simpler than their sources were consistently higher for interpreting (spoken) than for translation (written). For spoken production, the ratios of cross-lingual pairs with the simpler target were over 50% for all testing conditions, except the V0-any model. For written production, these ratios were not only consistently lower, but in English-to-German direction some models predicted the prevalence of segment pairs where targets were more complex than their sources.

Finally, as prompted above, the results from this experiment are suggestive of some asymmetry between translation directions. Any mediation into English leads to a greater simplification effect (against the aligned sources) than mediation into German. The asymmetry in translational properties of translation into English and into other languages, including German, attracted some attention from the research community before. In particular, results reported by [Kunilovskaya et al. \(2023\)](#) obtained using other methods, confirm our current observation that written translation into German seems to increase text complexity (unlike all other mediation settings). However, this can also be an effect of a zero-shot setup: the models were fine-tuned on the English data only.

4.2 Experiment 2: Are Targets Simpler than Comparable Originals?

The models trained in this experiment were designed to compare the complexity of mediated text vs. originals in the target language. Table 6 reports the performance of the four binary classifiers and a multiclass classifier, described in Section 3.2, on the Newsela corpus.

As was the case with the classification of the paired segments, the performance of the classifier followed the increase in the contrast between the classes, achieving the best accuracy score of 0.88 for V0-vs-V4 classifier. The confusion matrix for the multiclass indicates that the classifier struggles most with V2 and returns the highest accuracy for the extreme classes: V0 and V4.

A regressor fine-tuned on grade level values re-

model classes	acc	train	test
V0-vs-V1	0.66	6,504	1,626
V0-vs-V2	0.74	6,652	1,665
V0-vs-V3	0.84	6,728	1,682
V0-vs-V4	0.88	6,061	1,517
V0,V1,V2,V3,V4	0.41	16,472	4,119

Table 6: Newsela evaluation results for classifiers on content-unique segments

turned a Pearson correlation of 0.664 and root mean square error (RMSE) of 2.12 (grades spanned 11 values from 2 to 12). The results on 125 types of Lexile scores were very similar: Pearson 0.669 and RMSE 165.59 (Lexile score range from 320 to 1640).

Models’ transfer to the translation/interpreting data yielded the following results. In this classification setup, targets and target language originals were predicted independently of each other. Table 7 reports the aggregated predictions on the translational data from the binary classifiers. For considerations of space, it omits the outcomes for the intermediate simplification levels (V0 vs V2 and V0 vs V3).

	type	spoken	written
V0 vs V1			
DE	org	0.27	0.35
	tgt	0.15	0.44
EN	org	0.28	0.39
	tgt	0.17	0.41
V0 vs V4			
DE	org	0.4	0.51
	tgt	0.24	0.62
EN	org	0.46	0.65
	tgt	0.34	0.67

Table 7: Ratios of instances predicted as the more complex class (V0) by the classifiers fine-tuned on the least and most contrasting classes for originals and targets in each language by mode

The results from the omitted models were consistent with the trend established by the reported simplification models: in spoken production, targets had a twice lower ratio of complex sentences than comparable target language originals, while

in written mode, targets were a bit more difficult than the originals. Unlike the cross-lingual experiments, this observation holds for both translation directions. For example, in German, the V0-vs-V4 model predicted 40% of spoken originals and 24% of spoken targets as complex, while in written production 51% of originals and 62% of targets were complex. This finding is corroborated by the results from the multiclass model given in Table 8. The ratios of segments predicted as the complex V0 class were twice lower for spoken targets than for spoken originals in both languages and were higher for written targets than for written originals. While in the multiclass model, the V0 option was one of the five categories, the absolute values were lower compared to predictions of binary classifiers, as expected.

	type	spoken	written
DE	org	0.15	0.20
	tgt	0.06	0.30
EN	org	0.19	0.30
	tgt	0.09	0.33

Table 8: Ratios of instances predicted as the more complex class (V0) by the multiclass classifier for originals and targets in each language by mode

The predictions from zero-shot transferred regressors were well-aligned with the observation from classifiers in this experiment. They confirmed that cross-lingual mediation in spoken mode comes with a considerable simplification effect against similar target language originals. Tables 9 and 10 present average predicted grade and Lexile levels, respectively, for originals and targets in each language. For both metrics the lower the score, the lower the text complexity.

	type	spoken	written
DE	org	6.7 (+/-1.7)	7.2 (+/-1.6)
	tgt	6.0 (+/-1.4)	7.5 (+/-1.6)
EN	org	7.0 (+/-1.7)	7.9 (+/-1.6)
	tgt	6.6 (+/-1.5)	7.9 (+/-1.6)

Table 9: Mean predicted *grade levels* for originals and targets in each language by mode

It can be seen that the scores predicted for the targets tend to be lower than for the originals in the spoken mode, but higher in the written mode,

	type	spoken	written
DE	org	914.0 (+/-136)	951.8 (+/-133)
	tgt	844.9 (+/-121)	978.3 (+/-132)
EN	org	972.6 (+/-160)	1047.1 (+/-156)
	tgt	927.6 (+/-138)	1056.4 (+/-156)

Table 10: Mean predicted *Lexile levels* for originals and targets in each language by mode

except for written English where the difference was small and outside of the number of decimal places reported in Table 9.

5 Conclusion

This project adopts a modelling-driven approach to the study of readability/complexity of translated and interpreted texts against their sources and comparable originals in the target language. It is designed to test a theoretical claim, often made in translation studies and supported by some empirical evidence, that professional translation, and especially interpreting, entails a considerable simplification effect. In this study, the properties of translations and interpreting are contrasted with the sources (based on parallel segment-aligned data) and with comparable originals in the target language. Our method consists in fine-tuning a pre-trained multilingual model in a number of settings (required to respect the specificity of the two types of comparisons – targets to sources and targets to originals) and applying the fine-tuned models to texts produced in various cross-lingual mediation conditions in a zero-shot transfer scenario. We use the annotations in the Newsela readability corpus to create computational models of linguistic complexity and then transfer them to the bidirectional English-German translational data from Europarl reporting the results for spoken and written mediation mode in each translation direction separately.

Our findings from several experimental setups reveal a certain pattern of readability/simplification effects in cross-lingual mediation. When compared to their sources, targets tend to be easier to read, especially in interpreting and in German-to-English direction. Written translation into German might be an exception to this trend: German written targets were more often predicted as more complex than their sources. When compared to the originals in the target language, targets are simpler in interpreting, but not in translation. Written translations

were found more difficult to read than originally-authored tests in the same language, possibly with some exceptions for English, where the difference between the categories was small.

It is important to bare in mind that the nature and strength of any translationese effects are register- and language-pair-dependent. The claims made in this study are only applicable to the specific register and domain of the underlying translational data. We plan to extend this study to Spanish Newsela data and Spanish segment of Europarl to explore the properties of zero-shot predictions. Also, we leave the qualitative analysis of simplification transformations in the Newsela simplification versions and in translation data as well as the analysis of the models’ training process for future work.

Acknowledgments

This research was funded within the CHIST-ERA programme under the following national grant agreement: FWF-I 3441 (FWF Austrian Science Fund, Austria)

Lay Summary

This paper explores the impact of translation and interpreting on the readability of texts comparing the outcomes to originally-authored texts in the source and target languages. Previous studies of translated language demonstrate that translators have a tendency to make texts simpler, more readable and less ambiguous than originals in either source or target language. We expect that this general trend is stronger in simultaneous interpreting. Interpreters experience additional difficulties because they have to deliver their interpretations in the other language as they process the incoming original speech in the source language. This general expected trend towards simpler output in translation/interpreting can be counteracted by the tendency to render the original word-for-word, where possible. This tendency is known as shining-through and leads to a familiar Master Yoda talk in translations. So, we are interested in whether overall translated/interpreted messages are more readable than their sources and comparable texts in the target language. Our translated/interpreted materials come from a collection of speeches delivered in the European Parliament. The written edited versions of the original speeches and their written translations are available on their website, while the spoken versions and their simultaneous interpretations were transcribed from

the video recordings. Our approach is based on training computational models that can distinguish between sentences with higher and lower readability scores. The sentences for training were obtained from the Newsela corpus, which contains news articles in English manually simplified by experts in creating reading materials for schoolchildren of various ages. The trained models demonstrated a good ability to tell apart (1) more readable and less readable versions of the same sentence and (2) more readable and less readable sentences of unrelated content. These models were applied to translational data in two conditions corresponding to the two training setups: classification of sentence pairs into those where the source is more complex than the target or where it is not, and classification of sentences into marked as translated/interpreted or originally-authored in the target language. Our experiments yielded evidence that simultaneous interpreting comes with a strong simplification effect for both translation directions and both types of comparison (vs sources in the other language and vs non-translations in the same language). However, in written translation, the results are more varied. The simplification effect was only seen for the comparison against sources in the German-to-English direction.

References

- Mikel Artetxe and Holger Schwenk. 2019. [Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond](#). *Transactions of the Association for Computational Linguistics*, 7:597–610.
- Mona Baker. 1996. [Corpus-based translation studies: the challenges that lie ahead](#). *Terminology, LSP and translation: Studies in language engineering, in honour of Juan C. Sager.*, pages 175–186.
- Andrew Chesterman. 2010. [Why study translation universals?](#) *Kiasm. Acta Translatologica Helsingiensia*, 1:38–48.
- Koel Dutta Chowdhury, Cristina España i Bonet, and Josef van Genabith. 2021. [Tracing source language interference in translation with graph-isomorphism measures](#). In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 380–390. INCOMA Ltd.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Unsupervised cross-lingual representation learning at scale](#). *CoRR*, abs/1911.02116.
- Gloria Corpas Pastor, Ruslan Mitkov, Naveed Afzal, and Viktor Pekar. 2008. [A corpus-based NLP study of convergence and simplification](#). In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas (AMTA'08)*, pages 75–81. Association for Computational Linguistics.
- Daria Dayter. 2018. [Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of russian-english interpreting \(siren\)](#). *FORUM: International Journal of Interpretation and Translation*, 16.
- Stefan Evert and Stella Neumann. 2017. [The impact of translation direction on characteristics of translated texts: A multivariate analysis for English and German](#). *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 300:47–80.
- Volker Gast and Robert Borges. 2023. [Nouns, verbs and other parts of speech in translation and interpreting: Evidence from english speeches made in the european parliament and their german translations and interpretations](#). *Languages*, 8(1):39.
- Hai Hu and Sandra Kübler. 2021. [Investigating Translated Chinese and Its Variants Using Machine Learning](#). *Natural Language Engineering*, 27(3):339–372.
- Marta Kajzer-Wietrzny and Lukasz Grabowski. 2021. [Formulaicity in constrained communication: An intermodal approach](#). *MonTi Monografías de Traducción e Interpretación*, 13:148–183.
- Alina Karakanta, Mihaela Vela, and Elke Teich. 2018. [Europarl-UdS: Preserving metadata from parliamentary debates](#). In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).
- Maria Kunilovskaya. 2023. [Translationese indicators for human translation quality estimation \(based on English-to-Russian translation of mass-media texts\)](#). Ph.D. thesis, University of Wolverhampton.
- Maria Kunilovskaya and Ekaterina Lapshinova-Koltunski. 2020. [Lexicogrammatic translationese across two targets and competence levels](#). In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 4102–4112. The European Language Resources Association (ELRA).
- Maria Kunilovskaya, Heike Przybyl, Elke Teich, and Ekaterina Lapshinova-Koltunski. 2023. [Simultaneous Interpreting as a Noisy Channel: How Much Information Gets Through](#). In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2023)*, page in print. INCOMA Ltd.

- Dmitry Nikolaev, Taelin Karidi, Neta Kenneth, Veronika Mitnik, Lilja Saeboe, and Omri Abend. 2020. **Morphosyntactic predictability of translationese**. *Linguistics Vanguard*, 6(1):1–12.
- Maeve Olohan. 2001. **Spelling out the optionals in translation: a corpus study**. In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie, and Shereen Khoja, editors, *Proceedings of the Corpus Linguistics*, pages 423–432. Lancaster University.
- Heike Przybyl, Ekaterina Lapshinova-Koltunski, Katrin Menzel, Stefan Fischer, and Elke Teich. 2022. **Epic uds - creation and applications of a simultaneous interpreting corpus**. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 1193–1200, Marseille, France. ELDA.
- Tiina Puurtinen. 2003. **Genre-specific features of translationese? Linguistic differences between translated and non-translated Finnish children’s literature**. *Literary and Linguistic Computing*, 18(4):389–406.
- Thilina Rajapakse. 2019. **Simple transformers**. <https://www.simpletransformers.ai/>. [Online; accessed Feb. 12, 2021].
- Karien Redelinghuys. 2016. **Levelling-out and register variation in the translations of experienced and inexperienced translators: A corpus-based study**. *Stellenbosch Papers in Linguistics*, 45(1):189–220.
- Karien Redelinghuys and Haidee Kruger. 2015. **Using the features of translated language to investigate translation expertise: A corpus-based study**. *International Journal of Corpus Linguistics*, 20(3):293–325.
- Miriam Shlesinger and Noam Ordan. 2012. **More spoken or more translated?: Exploring a known unknown of simultaneous interpreting**. *Target. International Journal of Translation Studies*, 24(1):43–60.
- Elke Teich. 2003. *Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts*. Mouton de Gruyter, Berlin.
- Antonio Toral. 2019. **Post-editeese: an Exacerbated Translationese**. In *Proceedings of MT Summit XVII*, pages 273–281, Dublin.
- Vered Volansky, Noam Ordan, and Shuly Wintner. 2015. **On the features of translationese**. *Digital Scholarship in the Humanities*, 30(1):98–118.
- Donna A Williams. 2005. *Recurrent features of translation in Canada: A corpus-based study*. Ph.D. thesis, University of Ottawa (Canada).
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. **Transformers: State-of-the-art natural language processing**. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.
- Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. **Problems in Current Text Simplification Research: New Data Can Help**. *Transactions of the Association for Computational Linguistics*, 3:283–297.