

In-Context Learning for Upper Sorbian to German Translation: Comparing Lexical and Morpho-Syntactic Information

Qimin Sheng¹ and Marion Di Marco^{1,2} and Alexander Fraser^{1,2}

¹School of Computation, Information and Technology,
Technical University of Munich (TUM)

²Munich Center for Machine Learning

{qimin.sheng|marion.dimarco|alexander.fraser}@tum.de

Abstract

LLM-based translation remains challenging for low-resource languages, as LLMs often lack linguistic knowledge due to their limited representation in pre-training data. This is particularly problematic for morphologically rich languages where a large vocabulary can lead to further data sparsity.

Previous work has shown that providing external guidance on the source sentence in the form of language-specific information can help LLMs with tasks such as machine translation. In this study, we investigate zero-shot context augmentation for Upper Sorbian-to-German translation by enriching prompts with different types of contextual information: translation candidates in the form of bilingual dictionary entries, and morpho-syntactic analysis provided through morphological tagging. We first compare the effectiveness of these types of information in a zero-shot setting and then evaluate whether they generalize across additional experimental settings. We find that context-augmented prompting can improve Upper Sorbian-to-German translation. Contexts only containing translation candidates perform overall best, whereas morpho-syntactic analysis offers limited benefits, both on its own and in combination with lexical cues.

1 Introduction

Large Language Models (LLMs) have shown strong capabilities in machine translation (MT), especially for English and other high-resource languages (HRLs) (Brown et al., 2020; Chowdhery et al., 2023). However, a performance gap remains between high- and low-resource languages (LRLs) (Robinson et al., 2023; Hendy et al., 2023; Nicholas and Bhatia, 2023), as LRLs are underrepresented with only limited data available (Ponti et al., 2019; Joshi et al., 2020).

The absence of sufficient training data is particularly problematic for languages with complex

morphology, because a large vocabulary can further increase data sparsity. When translating *out of* a morphologically rich language, words must be analyzed in the sentence context and their semantic content needs to be transported appropriately into the target language. With a large variety of inflected word forms, some forms might not be well attested in the LLM pre-training data, or are even entirely absent from it. At the same time, other word forms of the same lemma might be better attested, meaning that the LLM knows the words in principle but cannot necessarily access this knowledge optimally. For HRLs, large pre-training corpora can provide language competence. In contrast, for LRLs with only limited data available, focused *linguistic descriptions* can contribute structured and context-relevant information to compensate for data sparsity during pre-training, for example Tanzer et al. (2024) and Zhang et al. (2024b).

In-context learning (ICL) constitutes an interesting approach for LRL translation, as it reduces the reliance on large-scale training data (Cahyawijaya et al., 2024) by providing external linguistic information in prompts (Garg et al., 2023). Unlike fine-tuning, ICL does not require parameter updates and is therefore relatively cost-effective (Brown et al., 2020; Dong et al., 2024). Previous work has explored the use of additional linguistic resources in prompts, such as dictionaries, parallel examples, grammatical descriptions, morphological analysis, and glosses to support LLM-based MT for LRLs, with several studies reporting improvements on translation quality (Pei et al., 2025; Zhang et al., 2024b,a; Court and Elsner, 2024).

In this work, we explore context augmentation strategies for translation from Upper Sorbian, a low-resource morphologically rich Slavic language, into German, a high-resource language. We investigate two complementary sources of information: the morpho-syntactic information obtained from a morphological tagger, and the lexical information

from a translation dictionary derived from parallel data. They are intended (i) to help the model understand morphologically complex input and (ii) to facilitate the transfer to German. To evaluate their individual and combined effects, we compare four context strategies.

We address the following research questions:

- RQ1: *How do different context strategies perform, and what is the optimal configuration for each?*
- RQ2: *Do they perform effectively and stably across different settings?*

We first evaluate all context strategies in zero-shot setting using our primary model (llama3.3-70B-instruct), and then further evaluate whether each strategy’s optimal configuration in zero-shot generalizes to another model (gpt-4o) and also to few-shot prompting.

We find that providing contextual information in prompts can improve Upper Sorbian-German MT across different settings. Lexical context in the form of lemma-level dictionary entries performs best across most settings, whereas source-side morphology alone provides only limited benefit. Combining lexical and morphological information brings no additional stable gains, suggesting that direct source-to-target lexical mapping is the major contributor to translation performance.

2 Related Work

2.1 Low-Resource Scenarios in LLM Machine Translation

LLM-based MT formulates translation as a language modeling task (Ataman et al., 2025), as its multilingual pretraining (Lai et al., 2023) may enable cross-lingual transfer from HRLs to LRLs (Xu et al., 2025). LLMs achieve strong translation performance for HRLs such as English which are well represented in model training data (Brown et al., 2020; Chowdhery et al., 2023; Vilar et al., 2023). In contrast, LRLs with only limited available training data are underrepresented in multilingual models (Robinson et al., 2023; Hendy et al., 2023; Nicholas and Bhatia, 2023; Hasan et al., 2024). For morphologically rich languages, a large inventory of inflected forms further increases data sparsity (Samardzic et al., 2024). Consequently, LLMs lack sufficient linguistic knowledge of LRLs, making translating low-resource languages challenging.

2.2 In-Context Learning for Machine Translation

To improve source-side understanding for LRL translation, linguistic information can be incorporated into prompts via in-context learning (ICL) (Garg et al., 2023). Unlike fine-tuning, this approach gives LLMs the possibility to learn from externally provided knowledge without updating their parameters (Brown et al., 2020), while reducing reliance on large-scale annotated data (Cahyawijaya et al., 2024).

Dictionary. Previous work has explored dictionary-based prompting for translating low-resource languages. For example Ghazvininejad et al. (2023) and Zhang et al. (2024a) integrate bilingual dictionary information into translation prompts. Lu et al. (2024) explore Chain-of-Dictionary (CoD) Prompting, which uses multilingual dictionary chains. However, dictionary-only approaches can produce syntactically incoherent translation (Zhang et al., 2024b).

Grammar. Tanzer et al. (2024) showed that LLMs can learn to translate languages unseen during pre-training with in-context information obtained from grammar books and dictionaries. However, the effectiveness of grammatical information for LRL MT remains debated. Explicit grammar rules can help resolve lexical ambiguities and guide word-order transformations (Zhang et al., 2024b), while grammar explanations show limited or even negative effects (Court and Elsner, 2024; Pei et al., 2025). Zhang et al. (2025) argue that LLMs can apply explicitly provided grammar rules but may struggle to retrieve them on their own.

Morphology & Gloss. Morphological information can provide useful linguistic cues, for example number and case information (Zhang et al., 2024b), which is especially relevant for morphologically rich languages. Interlinear glosses, which combine lexical and morphological information, can improve prompting-based MT (Ramos et al., 2025).

Parallel Examples. Few-shot prompting can provide a model with a small number of source-target translation examples (Brown et al., 2020). Increasing the number of demonstrations generally improves translation performance, although gains tend to diminish after only a few examples (Vilar et al., 2023). Using similar and higher-quality

demonstrations also benefits MT (Sia and Duh, 2023; Agrawal et al., 2022; Vilar et al., 2023).

2.3 Machine Translation between Upper Sorbian and German

Upper Sorbian (HSB) is a West Slavic minority language spoken in eastern Germany, with an estimated fewer than 15,000 speakers (Howson, 2017). Its rich morphology (e.g. the seven-case nominal system and uncommon archaic dual number) is particularly challenging for MT.

Sorbian $\leftrightarrow$ German MT has been studied in a series of WMT Shared Tasks since WMT20 (Fraser, 2020; Libovický and Fraser, 2021; Weller-Di Marco and Fraser, 2022; Okabe et al., 2025), for both Upper and Lower Sorbian.

Summary. This study explores the effectiveness of in-context learning (ICL) in this low-resource translation task. Previous studies on low-resource MT have shown that dictionaries, morphological analysis and parallel examples are generally beneficial, whereas grammatical information provides rather limited benefits (Pei et al., 2025; Zhang et al., 2024b; Court and Elsnér, 2024). Motivated by these findings, this work mainly investigates the effects of dictionary and morphological contextual information for Sorbian–German MT.

3 Tagging and Dictionary Creation

Machine translation remains challenging for Upper Sorbian due to its low-resource nature and rich morphology. We explore context augmentation strategies to better support the model in translating a Sorbian input sentence by providing two complementary types of information: *morphological tags* and bilingual *dictionary entries*. The morphological analysis provides an abstract description of the input sentence in which inflectional features such as case, number and tense are directly accessible, as well as the respective lemma. Relying on a bilingual dictionary, translation options support the model with lexical mappings from the source to the target language.

3.1 Parallel Data

All datasets, a parallel HSB-DE training corpus, as well as our test set, are sourced from previous WMT Shared Tasks on the Sorbian languages,

WMT 2022¹ and WMT2025².

The parallel training corpus used to create a HSB-DE dictionary contains 510,077 unique sentence pairs after deduplication. As test set, we use the TUM-NLP 2025 MT development set. After removing sentence pairs that overlap with the training corpus, we randomly sampled 2000 sentences for testing. The remaining 902 sentence pairs formed the candidate pool for few-shot example selection.

3.2 Pre-Processing: Tagging

We used RNNTagger³ (Schmid, 2019) to obtain morphological information. The tagger is applied to both the training corpus used for dictionary construction and the test set. Each tagged token is represented by its surface form, morphological full tag (POS tag and inflectional features) and lemma. Table 1 shows an example of tagging output.

3.3 Dictionary Creation

The HSB-DE dictionary is derived from word-aligned parallel data, from which we extract 1-1 and 1-*n* translation pairs.

Alignment. Word alignment was obtained with Eflomal^{4,5} (Östling and Tiedemann, 2016)). For the alignment, each token is represented by its lemma and morphological full tag. We follow the standard procedure of computing alignments for both the forward and the reverse direction, and then apply symmetrization using *atools* from FastAlign⁶.

Dictionary at the Lemma|CoreTag Level. To mitigate the data sparsity issue exacerbated by the rich morphology of Upper Sorbian, the HSB-DE dictionary is constructed at the Lemma|CoreTag-level according to the tagged training corpus and corresponding alignments. Each dictionary entry maps a source word to a frequency-ranked list of target-language translation candidates. The construction process consists of five steps:

¹https://statmt.org/wmt22/unsup_and_very_low_res.html

²<https://github.com/TUM-NLP/llms-limited-resources2025/tree/main/Sorbian/hsb/MT>

³<https://www.cis.uni-muenchen.de/~schmid/tools/RNNTagger/>

⁴<https://github.com/robertostling/eflomal>

⁵We opted for statistical word alignment as opposed to a neural alignment tool like SimAlign (Jalili Sabet et al., 2020) where the alignment quality depends on the coverage of the language in the underlying model; for example, mBERT (Devlin et al., 2019) does not support Sorbian.

⁶https://github.com/clab/fast_align

1. Extracting source-target token pairs according to alignment indices.
2. Filtering noisy token types, e.g. PROP, NUM.
3. Aggregating occurrence frequencies for associated target candidates.
4. Representing one-to-many alignments as a multi-word expression with a PHRASE tag.
5. Ranking target candidates by descending frequency.

The resulting HSB-DE dictionary contains 177,436 unique source entries with an average of 3.77 translation candidates per entry. Using the same token filtering criteria as in dictionary construction, the dictionary reaches 96.21% coverage at the Lemma level and 95.68% at the Lemma|CoreTag level on the test set. Most uncovered entries belong to the noun (NOUN/NN/NE) and adjective (ADJ/ADJA/ADJD) word types, and the majority of their lemmas are tagged as <unknown>.

Dictionary Lookup. When applying the dictionary to retrieve translation candidates for context construction, each token in the HSB test set is looked up using the following fallback order: (1) exact matching at the Lemma|CoreTag level, (2) matching at the Lemma level, and (3) matching using lowercased or title-cased lemma forms. The first successful match returns the ranked list of target translation candidates at the Lemma|CoreTag level used for subsequent context construction. Filtered or unmatched tokens receive no translation candidates.

4 Context Strategies

To compare the individual and combined effects of **lexical information** from the generated Lemma|CoreTag dictionary and **morphological information** from the source-side tagged output, we explore four context construction strategies: **Strategy A** combines the most frequent translation candidate with optional morphological information in a pseudo-gloss format; **Strategy B** provides multiple translation candidates; **Strategy C** uses morphological information only; **Strategy D** combines multiple translation candidates with complete morphological information. Throughout this section, we use the example in Table 1 to illustrate different context construction strategies.

Source: Zo bychu prózdnjeńcu před přistupom kunow abo wrónow škitali, zamurjuja pikotače zachod.

Reference: Um die Höhle vor dem Zugriff von Mardern oder Krähen zu schützen, mauern die Kleiber den Eingang zu.

English for readability: To protect the nest cavity from access by martens or crows, nuthatches wall up the entrance.

Source-side tagged representation		
Word	Morph	Lemma
Zo	ADP	zo
bychu	AUX.Cnd.Plur.3.Fin	być
prózdńjeńcu	NOUN.Acc.Fem.Sing	prózdńjeńca
před	ADP	před
přistupom	NOUN.Ins.Neut.Sing	přistupo
kunow	NOUN.Gen.Fem.Plur	kuna
abo	CCONJ	abo
wróńow	NOUN.Gen.Fem.Plur	wróńa
škitali	VERB.Inan.Masc.Plur.Past.Part.Act	škitać
,	PUNCT	,
zamurjuja	VERB.Ind.Sing.3.Pres.Fin	zamurjować
pikotače	ADJ.Acc.Neut.Sing.Part.Pass	pikotači
zachod	NOUN.Inan.Acc.Masc.Sing	zachod
.	PUNCT	.

Table 1: Example sentence with morphological analysis.

4.1 Context Strategy A: Pseudo-Gloss

We consider two variants of Context Strategy A, illustrated in Table 2.

Context Strategy A1 constructs a pseudo-gloss by replacing each matched source token with its most frequent target translation candidate, while unmatched tokens keep their original form. The resulting lexical sequence is interleaved with Upper Sorbian morphological tags to provide inflectional information. This context design provides an abstract target-language representation while preserving the original source token order.

While some Upper Sorbian-specific morphological features (e.g. dual number and grammatical gender and some cases) have no direct mapping in German, we assume that they still provide relevant linguistic signals to help form a complete translation into German. To assess the impact of the morphological information, we also evaluate a variant without source-side morphological tags (A2), resulting in a lexical-only sequence containing German lemmas for matched tokens and original forms for unmatched tokens, in Upper Sorbian word order.

4.2 Context Strategy B: Dictionary Entries

An obvious issue with Context A is that it relies solely on the highest-ranked translation option. Context B provides multiple translation candidates to better address lexical ambiguity and to offer a larger range of options. This is relevant not

A1: Translation candidates with source morpho-tags.

<gloss> dass ADP sein AUX.Cnd.Plur.3.Fin Höhle
NOUN.Acc.Fem.Sing vor ADP Beitritt NOUN.Ins.Neut.Sing Marder
NOUN.Gen.Fem.Plur oder CCONJ Krähe NOUN.Gen.Fem.Plur schützen
VERB.Inan.Masc.Plur.Past.Part.Act , PUNCT mauern
VERB.Ind.Sing.3.Pres.Fin Kleiber ADJ.Acc.Neut.Sing.Part.Pass
Eingang NOUN.Inan.Acc.Masc.Sing . PUNCT </gloss>

A2: Translation candidates only.

<pseudo_translation> dass sein Höhle vor Beitritt Marder oder
Krähe schützen , mauern Kleiber Eingang . </pseudo_translation>

Note: The representation uses a single line, wrapped here only because of space constraints.

Table 2: Example for Context Strategy A

only for words with clearly different meanings, but also for function words like conjunctions or prepositions that have a large variety of possible translations which strongly depend on the context. To select the translation candidates, we consider the following factors:

Representation: Keeping vs. not keeping the main tag of the translation candidates, as illustrated in the dictionary entry for *zamurjować* (EN: 'to wall up'; DE: 'einmauern') below:

- **With Tag & Raw** keeps Lemma|CoreTag format to provide shallow syntactic information.
zamurjować (VERB): mauern (VFIN), einmauern (VINF), einmauern (VPP)
- **No Tag & Unique** provides unique lemma-level target candidates ranked by aggregated frequency. zamurjować: einmauern, mauern

Probability Threshold: Only translation candidates with relative frequency above the threshold are kept. Table 9 summarizes the statistics for Context B across thresholds.

4.3 Context Strategy C: Morphological Analysis

To help the model capture the rich morphology of Upper Sorbian, Context Strategy C provides morphological information for each source token, including the surface form, morphological tags (POS tag and inflectional features) and lemma, cf. <morph>...</morph> in Table 3. This representation links each inflected form to its lemma to address data sparsity at the lexical level, and also provides a morpho-syntactic analysis in the sentence context.

```
<dict>
zo: dass, um, damit
być: sein, haben
prózdnjeća: Höhle
před: vor
přistupo: Beitritt, Zugang
kuna: Marder, Marderin, Baumarder
abo: oder
wróna: Krähe, Nebelkrähe
škitać: schützen
zamurjować: einmauern, mauern
pikotači: Kleiber
zachod: Eingang
</dict>
<morph>
Word=Zo Morph=ADP Lemma=zo
Word=bychu Morph=AUX.Cnd.Plur.3.Fin Lemma=być
Word=prózdnjeću Morph=NOUN.Acc.Fem.Sing Lemma=prózdnjeća
Word=před Morph=ADP Lemma=před
Word=přistupom Morph=NOUN.Ins.Neut.Sing Lemma=přistupo
Word=kunow Morph=NOUN.Gen.Fem.Plur Lemma=kuna
Word=abo Morph=CCONJ Lemma=abo
Word=wrónow Morph=NOUN.Gen.Fem.Plur Lemma=wróna
Word=škitali Morph=VERB.Inan.Masc.Plur.Past.Part.Act
Lemma=škitać
Word=, Morph=PUNCT Lemma=,
Word=zamurjuja Morph=VERB.Ind.Sing.3.Pres.Fin
Lemma=zamurjować
Word=pikotače Morph=ADJ.Acc.Neut.Sing.Part.Pass
Lemma=pikotači
Word=zachod Morph=NOUN.Inan.Acc.Masc.Sing Lemma=zachod
Word=. Morph=PUNCT Lemma=.
</morph>
```

Note: The representation uses one token per line. Long lines wrap here only because of space constraints, with continuation lines indented to distinguish them from new token entries.

Table 3: Context Strategy D: lexical information (<dict>) and morphological information (<morph>).

4.4 Context Strategy D: Combined Dictionary Entries and Morphological Analysis

Context Strategy D combines lexical information from the best-performing configuration of Context Strategy B with morphological information from Context Strategy C. The combined representation varies along the following dimensions:

- **Structure:** *All-in-One*, where all information for each source token appears on a single line, or *Block-by-Block*, where the two information types are grouped into separate blocks.
- **Order:** *OrderBC*, where lexical information is presented first, or *OrderCB*, where morphological information is presented first.

5 Experiments

To address RQ1, we compare the context-free baseline (baseline prompt) with the four context-augmented strategies (extended prompts with described context) in a zero-shot setting using

Translate the following Upper Sorbian sentence into German: <hsb> {sentence} </hsb>
The following context is provided to assist with the translation:
[Context block]
Provide exactly one translation (no alternatives, no explanations) in this format: <de> {German translation} </de>

Table 4: Zero-shot Generic-Intro prompt template.

Llama3.3-70B-Instruct with the prompt template shown in Table 4⁷. For RQ2, we further investigate whether the observed effects in the LLaMA zero-shot setting generalize to another model (GPT-4o) as well as to the LLaMA few-shot setting.

5.1 Experimental Setup

Experiments were conducted on our test set of 2000 HSB source sentences; few-shot examples were drawn from an extra pool of 902 sentences.

We used Meta:Llama 3.3 70B Instruct (hereafter LLaMA) (Meta AI, 2024) as primary model, while OpenAI: GPT-4o (hereafter GPT) (OpenAI et al., 2024) was used only for the cross-model comparison. All requests were sent through OpenRouter API with the temperature set to 0.

5.2 Prompt Settings

The zero-shot baseline prompt contains only the translation instruction and the output format, whereas the context-augmented prompt additionally includes a context introduction and a strategy-specific context block.

In the few-shot setting, all prompts additionally include the same three demonstration pairs with source sentences of 10–15 tokens, randomly sampled with a fixed seed of 42 from a separate parallel HSB-DE sentence pool disjoint from the test set. The source and target sentences serve as demonstration inputs and outputs, respectively, while demonstration contexts for context-augmented variants are constructed according to the strategy being evaluated.

5.3 Evaluation Metrics

We evaluate translation quality using BLEU (Papineni et al., 2002) for lexical overlap via the im-

⁷In preliminary experiments, we compared a *Generic-Intro* prompt with a *Described-Intro* variant that replaces the generic term context with a strategy-specific description (e.g. *token-aligned German gloss with morpho-syntactic features*), but found the generic prompt more effective.)

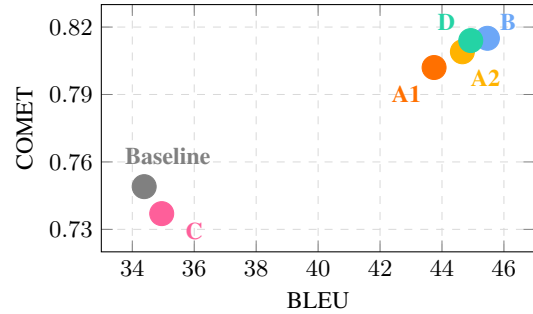


Figure 1: Cross-Strategy Performance Comparison measured by BLEU and COMET. See numerical results in Table 10 in Appendix B.

plementation of SACREBLEU⁸ (Post, 2018) and COMET (Rei et al., 2020) for semantic similarity. Model outputs were post-processed by removing configuration headers and formatting tags to ensure sentence alignment, without modifying the translation content.

5.4 RQ1 Experimental Results

Figure 1 shows an overview of all strategies, displaying each in its best-performing configuration.

Lexical information is the primary contributor to translation performance. All strategies providing source-to-target lexical mappings (Contexts A1, A2, B, and D) substantially outperform the baseline in both BLEU and COMET (Figure 1), with lexical-only Context B performing best overall. Strategies providing multiple translation candidates (Contexts B and D) consistently outperform those using only the top-ranked candidate (Contexts A1 and A2). In contrast, morphology-only Context C shows a marginal BLEU improvement and a slight COMET decline. Adding source-side morphology to lexical information also offers limited benefit, as Context D and A1 does not outperform the corresponding lexical-only strategies, B and A2, respectively.

Compact sets of unique, tag-free translation candidates perform best. The *No Tag & Unique* variant consistently outperforms *With Tag & Raw* across all thresholds in both BLEU and COMET (Figure 2), with a particularly large BLEU gain at the 20% threshold, indicating that removing tags and deduplicating translation candidates provides more effective lexical context. As thresholds increase, both variants improve and then decline,

⁸BLEU signature:nrefs:1|case:mixed|eff:no|tok:13a|smooth:exp|version:2.6.0

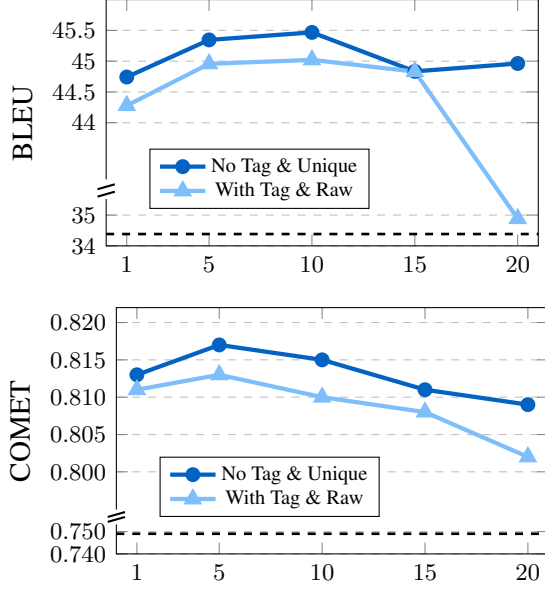


Figure 2: Comparison of Context B Variants across different frequency-filtering thresholds measured by BLEU and COMET. The dashed lines indicate context-free baseline. See numerical results in Table 11 in Appendix B.

with BLEU and COMET for *No Tag & Unique* peaking at 10% and 5%, respectively. The corresponding averages of 1.6307 and 2.3691 candidates per source token (Table 9 in Appendix A) suggest BLEU favors a more focused candidate set than COMET. We therefore select *No Tag & Unique* variant with a 10% threshold for subsequent Context B experiments, as it achieves the highest BLEU and near-peak COMET.

A lexical block followed by a morphological block performs best in combining information. *OrderBC* (lexical information first) generally outperforms *OrderCB* (morphological information first) in both BLEU and COMET (Figure 3), except for a slight COMET advantage for *OrderCB* in *All-In-One*. This suggests that presenting lexical information before morphology is generally more beneficial. While *All-In-One* remains relatively stable across orderings, *Block-By-Block* is more sensitive to information order, achieving the highest BLEU and COMET scores with its *OrderBC* variant. We therefore select *Block-By-Block OrderBC* for subsequent Context D experiments.

5.5 RQ2: Evaluating Additional Settings

For RQ2, we first evaluate whether the observed improvements also hold with another model (GPT). Then, we evaluate the four context strategies in a

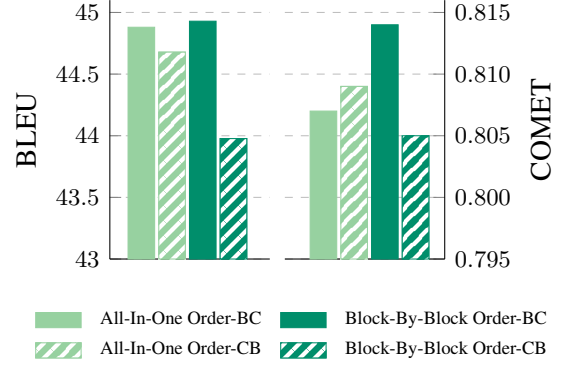


Figure 3: Comparison of Context D Variants across different representations measured by BLEU and COMET. See numerical results in Table 12 in Appendix B.

few-shot setting with our primary LLaMA model.

Same trend for GPT. Although GPT achieves higher translation quality overall, the effects of context augmentation follow the same general pattern as with LLaMA (cf. Figure 4): improvements over the baseline for context strategies providing source-target mappings, and the worst performance for the morphology-only context. Notably, the GPT baseline reaches a BLEU score comparable to several context-augmented LLaMA results.

Few-shot consistently outperforms zero-shot for LLaMA. Across all strategies, including the baseline, few-shot prompting improves both BLEU and COMET (Figure 5). Notably, Context A2 achieves the highest BLEU score in the few-shot setting, with Contexts B and D close behind.

Summary. The context strategies with their selected configurations show broadly similar performance patterns across LLaMA zero-shot, GPT zero-shot and LLaMA few-shot settings. Con-

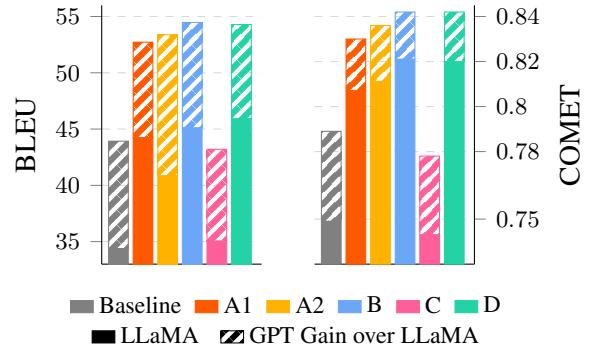


Figure 4: Comparison of LLaMA and GPT under zero-shot prompting measured by BLEU and COMET. See numerical results in Table 13 in Appendix B.

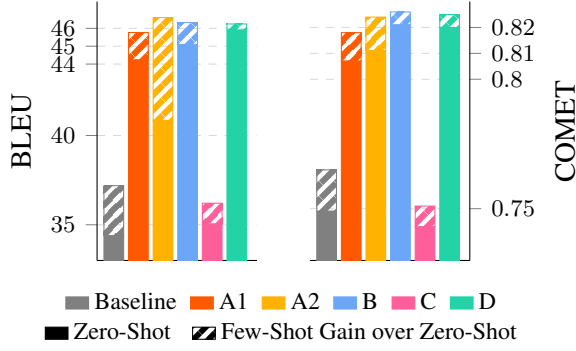


Figure 5: Zero-shot and few-shot performance of LLaMA measured by BLEU and COMET. See numerical results in Table 14 in Appendix B.

texts B and D consistently form the top-performing group, while A1 and A2 tend to rank in the middle. Context C remains the weakest context strategy in almost all evaluated settings, often performing below the baseline.

6 Discussion

The experimental findings of RQ1 (Section 5.4) are examined through sentence-level case studies⁹, focusing on the benefits of multiple translation candidates and the conditional usefulness of source-side morphology. The full corresponding Context D examples are provided in Appendix C.

Multiple translation candidates help context-dependent lexical selection. This example (Table 5) contains two source words with multiple plausible translations: *wopismo*, with candidates *Urkunde* (*document*) and *Bescheinigung* (*certificate*), and *pućowanje*, with candidates *Wanderung* (*hike*) and *Reise* (*trip*) (see Table 15 in Appendix C). The reference translation uses *Bescheinigung* and *Reise*. Given only the most frequent candidate for each word in Context A1, the model produces *Urkunde* and *Wanderung*, which are plausible dictionary translations in isolation but less suitable in this context. By contrast, Contexts B and D provide multiple candidates, allowing the model to choose the contextually appropriate alternatives *Bescheinigung* and *Reise*.

Reliable morphological tags can be beneficial. In this example (Table 6), the morphological tags are largely accurate (see Table 16 in Appendix C). Contexts A1, B, and D correctly recover key lexical

⁹With the context configurations selected in Section 5.4; using *Zero-Shot Baseline* prompt for the baseline and *Zero-Shot Generic-Intro* prompt for context-augmented examples.

Source Mějće tute wopismo na pućowanju při sebi.
Reference Führen Sie diese **Bescheinigung** auf der **Reise** mit.

Strategy	Translation	COMET
Baseline	Haben Sie dieses <i>Schreiben</i> auf der Reise bei sich.	0.634
Context A1	Habt diese <i>Urkunde</i> auf eurer <i>Wanderung</i> bei euch.	0.596
Context B	Habt diese Bescheinigung auf eurer Reise bei euch.	0.818
Context D	Habt diese Bescheinigung auf eurer Reise bei euch.	0.818

Note: English for readability: “Carry this **certificate** with you during your **trip**.” See Context D in Table 15 in Appendix C.

Table 5: Selection of translation candidates.

Source Spisany wot mnicha ze srjedźowěka, rozprawješe list wo pokładze wjercha.
Reference Verfasst von einem Mönch aus dem Mittelalter, **berichtete** der Brief über den **Schatz** des **Fürsten**.

Strategy	Translation	COMET
Baseline	Der von einem Mönch im Mittelalter geschriebene Brief handelt von der Haltung des Herrn.	0.667
Context A1	Von Mönchen im Mittelalter verfasst , berichtet der Brief über den Schatz des Fürsten .	0.871
Context B	Der von einem Mönch im Mittelalter verfasste Brief berichtet über den Schatz des Fürsten .	0.869
Context C	Geschrieben von einem Mönch des Mittelalters, handelte der Brief über die Haltung des Herrschers.	0.770
Context D	Verfasst von einem Mönch des Mittelalters, berichtet der Brief über den Schatz des Fürsten .	0.904

Note: English for readability: “**Written** by a medieval monk, the letter **recounted** the **prince’s treasure**.” See Context D in Table 16 in Appendix C.

Table 6: Reliable morphological tags

choices consistent with the reference, such as *verfasst* (*written*) and *Schatz* (*treasure*), while the baseline and Context C produce less appropriate lexical choices. The baseline correctly reproduces the first part of the sentence, but fails to translate the second part. Nevertheless, Context C outperforms the baseline and is the only variant that correctly reproduces the source tense, likely due to the explicit past-tense tag (VERB.Ind.Sing.3.Past.Fin). Context D achieves the highest COMET score, suggesting that reliable morphological tags can further improve translation quality when combined with lexical candidates.

Incorrect morphological tags can cause error propagation. This example (Table 7) illustrates how morphological ambiguity in Upper Sorbian can lead to contextually inappropriate tags, as illustrated in Table 8¹⁰.

The dual verb *namotwitej* (‘call upon’) identifies *Koaliciskej frakciji* (‘two coalition parliamentary groups’) as the dual nominative subject and *knježerstwo* (‘government’) as the accusative ob-

¹⁰Morphological analyses follow Upper Sorbian paradigms: <https://baltoslav.eu/hsb/>

Source	Koaliciskej frakciji namolwitej knjezerstwo namjety MSPI, dalokož Europska komisija je sčěhuje, bórže pruwować.	
Reference	Die beiden Koalitionsfraktionen forderten die Regierung dazu auf, die Vorschläge von MSPI, soweit ihnen die Europäische Kommission folgt, bald zu prüfen.	
Strategy	Translation	COMET
Baseline	Die Regierung der Koalitionsfraktion hat den Vorschlag der MSPI angenommen, der von der Europäischen Kommission verfolgt wird, um ihn zu überprüfen.	0.660
Context B	Die Koalitionsfraktion hat die Regierung aufgefordert, den Vorschlag des MSPI zu prüfen, soweit die Europäische Kommission ihm folgt, und zwar zunächst.	0.728
Context C	Der Koalitionsfraktion , die die Regierung gebildet hat, schlägt die MSPI vor, da die Europäische Kommission sie verfolgt, um sie besser zu prüfen.	0.541

Note: English for readability: “**The two coalition parliamentary groups** called upon the government to promptly review MSPI’s proposals, insofar as the European Commission adopts them.” See Context D in Table 17 in Appendix C.

Table 7: Incorrect morphological tags

HSB	Possible morphological analyses	Role
ADJ <i>Koaliciskej</i>	Sing.Dat/Ins/Loc or Dual.Nom/Acc	subject (Dual.Nom)
NOUN <i>frakciji</i>	Sing.Dat/Loc or Dual.Nom/Acc/Voc	
VERB <i>namolwitej</i>	Ind.Dual.3.Pres.Fin	verb (Dual)
NOUN <i>knjezerstwo</i>	Sing.Nom/Acc/Voc	object (Acc)

Table 8: Relevant Morphological Information.

ject. The baseline already misinterprets the subject-object relation, translating the first part into “*The government of the coalition parliamentary group ...*”. While this translation is somewhat ambiguous due to case ambiguity in German, it does not correctly reproduce the relation between the two phrases.

Context C further amplifies this error by explicitly providing contextually inappropriate tags, where the actual subject *Koaliciskej frakciji* is labeled as locative and the non-subject element *MSPI* as nominative (see Table 17 in Appendix C). This appears to lead the model to treat *die MSPI* as the subject. In contrast, Context B avoids this subject-object confusion, suggesting that translation candidates presented in source-token order provide useful anchors for generation. All strategies fail to recover the dual feature from Upper Sorbian, translating *Koaliciskej frakciji* as singular rather than *Die beiden Koalitionsfraktionen* (*the two coalition parliamentary groups*) as shown in the reference. This suggests that the source-side dual feature is not reliably captured by neither the tagger nor the model.

Taken together, these case studies support the overall experimental findings. Providing multiple translation alternatives enables the model to select contextually appropriate lexical choices. In

contrast, source-side morphology is conditionally useful: reliable tags can provide complementary structural cues, while incorrect tags can introduce noise and lead to error propagation.

7 Conclusion

This study investigates how lexical and morphological information provided in prompts affects LLM-based Upper Sorbian-to-German translation via ICL, using four representations: pseudo-gloss, lexical-only, morphology-only, and combined contexts. Context augmentation improves translation quality, with lexical-only context performing best overall. Morphological information shows limited benefit on its own and no consistent gains when combined with lexical guidance (RQ1). The relative ranking among context strategies remains largely stable across experimental settings (RQ2). These findings suggest that source-to-target lexical mapping is the major contributor to translation improvements. Case studies further show that the usefulness of source-side morphology can be limited by morphological ambiguity. Overall, the effectiveness of context augmentation depends on the type and reliability of the information provided.

8 Limitations

We briefly discuss some limitations of our study.

Representation and reliability of source-side tags. We used the morphological tags produced by RNNTagger without modification. However, the rich inflectional system of Upper Sorbian makes morphological analysis challenging, as reflected in the erroneous tags identified in our case studies which raises questions of the overall tagging quality. Furthermore, not all of the Sorbian features have a direct equivalent in German inflection, and thus might not provide a useful signal.

Translation candidate selection and word sense disambiguation. The most frequent translation candidates are not always contextually appropriate, but might put less frequent word senses at a disadvantage. Future work could consider the sentence-level context in candidate selection. Similarly, we only consider 1-1 and 1-many translation options, excluding longer $n-m$ phrase translations. Investigating more flexible translation units might help to preserve more contextual information, but is not trivial due to potential phrase overlap.

Translation direction. This study explores the translation from Upper Sorbian to German. Future work can investigate the reverse direction, shifting the focus from understanding low-resource input to generating low-resource text.

9 Acknowledgements

This work was co-funded by the European Union (ERC, EPICAL, 101141712). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

- Sweta Agrawal, Chunting Zhou, Mike Lewis, Luke Zettlemoyer, and Marjan Ghazvininejad. 2022. [In-context examples selection for machine translation](#). *Preprint*, arXiv:2212.02437.
- Duygu Ataman, Alexandra Birch, Nizar Habash, Marcello Federico, Philipp Koehn, and Kyunghyun Cho. 2025. [Machine translation in the era of large language models: a survey of historical and emerging problems](#). *Information*, 16(9).
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. [Language models are few-shot learners](#). *Preprint*, arXiv:2005.14165.
- Samuel Cahyawijaya, Holy Lovenia, and Pascale Fung. 2024. [LLMs are few-shot in-context low-resource language learners](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 405–433, Mexico City, Mexico. Association for Computational Linguistics.
- Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, and 48 others. 2023. Palm: scaling language modeling with pathways. *J. Mach. Learn. Res.*, 24(1).
- Sara Court and Micha Elsner. 2024. [Shortcomings of llms for low-resource translation: Retrieval and understanding are both the problem](#). *Preprint*, arXiv:2406.15625.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Jingyuan Ma, Rui Li, Heming Xia, Jingjing Xu, Zhiyong Wu, Baobao Chang, Xu Sun, Lei Li, and Zhifang Sui. 2024. [A survey on in-context learning](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1107–1128, Miami, Florida, USA. Association for Computational Linguistics.
- Alexander Fraser. 2020. [Findings of the WMT 2020 shared tasks in unsupervised MT and very low resource supervised MT](#). In *Proceedings of the Fifth Conference on Machine Translation*, pages 765–771, Online. Association for Computational Linguistics.
- Shivam Garg, Dimitris Tsipras, Percy Liang, and Gregory Valiant. 2023. [What can transformers learn in-context? a case study of simple function classes](#). *Preprint*, arXiv:2208.01066.
- Marjan Ghazvininejad, Hila Gonen, and Luke Zettlemoyer. 2023. [Dictionary-based phrase-level prompting of large language models for machine translation](#). *Preprint*, arXiv:2302.07856.
- Md. Arif Hasan, Prerona Tarannum, Krishno Dey, Imran Razzak, and Usman Naseem. 2024. [Do large language models speak all languages equally? a comparative study in low-resource settings](#). *Preprint*, arXiv:2408.02237.
- Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. [How good are gpt models at machine translation? a comprehensive evaluation](#). *Preprint*, arXiv:2302.09210.
- Phil Howson. 2017. [Upper sorbian](#). *Journal of the International Phonetic Association*, 47(3):359–367.
- Masoud Jalili Sabet, Philipp Dufter, François Yvon, and Hinrich Schütze. 2020. [SimAlign: High quality word alignments without parallel training data using static and contextualized embeddings](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 1627–1643, Online. Association for Computational Linguistics.
- Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. [The state and fate of linguistic diversity and inclusion in the NLP world](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

- Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. 2023. [Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning](#). *Preprint*, arXiv:2304.05613.
- Jindřich Libovický and Alexander Fraser. 2021. [Findings of the WMT 2021 shared tasks in unsupervised MT and very low resource supervised MT](#). In *Proceedings of the Sixth Conference on Machine Translation*, pages 726–732, Online. Association for Computational Linguistics.
- Hongyuan Lu, Haoran Yang, Haoyang Huang, Dongdong Zhang, Wai Lam, and Furu Wei. 2024. [Chain-of-dictionary prompting elicits translation in large language models](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 958–976, Miami, Florida, USA. Association for Computational Linguistics.
- Meta AI. 2024. [Llama 3.3 70b instruct model card](#). GitHub Repository. Accessed: 2026-05-29.
- Gabriel Nicholas and Aliya Bhatia. 2023. [Lost in translation: Large language models in non-english content analysis](#). *Preprint*, arXiv:2306.07377.
- Shu Okabe, Daryna Dementieva, Marion Di Marco, Lukas Edman, Katharina Haemmerl, Marko Měškank, Anita Hendrichowa, and Alexander Fraser. 2025. [Findings of the WMT 2025 shared task LLMs with limited resources for Slavic languages: MT and QA](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 503–519, Suzhou, China. Association for Computational Linguistics.
- OpenAI, :, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, and 401 others. 2024. [Gpt-4o system card](#). *Preprint*, arXiv:2410.21276.
- Robert Östling and Jörg Tiedemann. 2016. [Efficient word alignment with Markov Chain Monte Carlo](#). *Prague Bulletin of Mathematical Linguistics*, 106:125–146.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Renhao Pei, Yihong Liu, Peiqin Lin, François Yvon, and Hinrich Schuetze. 2025. [Understanding in-context machine translation for low-resource languages: A case study on Manchu](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8767–8788, Vienna, Austria. Association for Computational Linguistics.
- Edoardo Maria Ponti, Helen O’Horan, Yevgeni Berzak, Ivan Vulić, Roi Reichart, Thierry Poibeau, Ekaterina Shutova, and Anna Korhonen. 2019. [Modeling language variation and universals: A survey on typological linguistics for natural language processing](#). *Computational Linguistics*, 45(3).
- Matt Post. 2018. [A call for clarity in reporting BLEU scores](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.
- Rita Ramos, Everlyn Asiko Chimoto, Maartje ter Hove, and Natalie Schluter. 2025. [Grammamt: Improving machine translation with grammar-informed in-context learning](#). *Preprint*, arXiv:2410.18702.
- Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. [COMET: A neural framework for MT evaluation](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.
- Nathaniel Robinson, Perez Ogayo, David R. Mortensen, and Graham Neubig. 2023. [ChatGPT MT: Competitive for high- \(but not low-\) resource languages](#). In *Proceedings of the Eighth Conference on Machine Translation*, pages 392–418, Singapore. Association for Computational Linguistics.
- Tanja Samardzic, Ximena Gutierrez, Christian Bentz, Steven Moran, and Olga Pelloni. 2024. [A measure for transparent comparison of linguistic diversity in multilingual NLP data sets](#). In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3367–3382, Mexico City, Mexico. Association for Computational Linguistics.
- Helmut Schmid. 2019. [Deep learning-based morphological taggers and lemmatizers for annotating historical texts](#). In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, DATeCH2019*, page 133–137, New York, NY, USA. Association for Computing Machinery.
- Suzanna Sia and Kevin Duh. 2023. [In-context learning as maintaining coherency: A study of on-the-fly machine translation using large language models](#). In *Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track*, pages 173–185, Macau SAR, China. Asia-Pacific Association for Machine Translation.
- Garrett Tanzer, Mirac Suzgun, Eline Visser, Dan Jurafsky, and Luke Melas-Kyriazi. 2024. [A benchmark for learning to translate a new language from one grammar book](#). In *The Twelfth International Conference on Learning Representations*.
- David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2023. [Prompting palm for translation: Assessing strategies and performance](#). *Preprint*, arXiv:2211.09102.

Marion Weller-Di Marco and Alexander Fraser. 2022. Findings of the WMT 2022 shared tasks in unsupervised MT and very low resource supervised MT. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 801–805, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Yuemei Xu, Ling Hu, Jiayi Zhao, Zihan Qiu, Kexin Xu, Yuqi Ye, and Hanwen Gu. 2025. A survey on multi-lingual large language models: corpora, alignment, and bias. *Frontiers of Computer Science*, 19(11).

Chen Zhang, Jiuheg Lin, Xiao Liu, Zekai Zhang, and Yansong Feng. 2025. Read it in two steps: Translating extremely low-resource languages with code-augmented grammar books. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3977–3997, Vienna, Austria. Association for Computational Linguistics.

Chen Zhang, Xiao Liu, Jiuheg Lin, and Yansong Feng. 2024a. Teaching large language models an unseen language on the fly. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8783–8800, Bangkok, Thailand. Association for Computational Linguistics.

Kexun Zhang, Yee Choi, Zhenqiao Song, Taiqi He, William Yang Wang, and Lei Li. 2024b. Hire a linguist!: Learning endangered languages in LLMs with in-context linguistic descriptions. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 15654–15669, Bangkok, Thailand. Association for Computational Linguistics.

A Statistics

Threshold	Variant	#Entries	#Avg. Cand.
1%	Raw	25063	6.9048
	Unique	25063	6.2589
5%	Raw	25062	2.6179
	Unique	25062	2.3691
10%	Raw	25000	1.7735
	Unique	25012	1.6307
15%	Raw	24781	1.4482
	Unique	24834	1.3533
20%	Raw	24266	1.2714
	Unique	24520	1.2279

Table 9: Context B Statistics by Threshold.

B Performance Score Tables

Strategy	BLEU / COMET
Baseline	34.384/0.749
Context A1	43.745/0.802
Context A2	44.659/0.809
Context B	45.467/0.815
Context C	34.950/0.737
Context D	44.929/0.814

Note: All performance results are reported as BLEU / COMET, with the best result shown in bold.

Context B configuration: NoTag & Unique & 10% threshold.
Context D configuration: Block-By-Block OrderBC.

Table 10: Zero-Shot LLaMA: Baseline vs. Contexts.

Threshold	With Tag & Raw	No Tag & Unique
1%	44.278 / 0.811	44.742 / 0.813
5%	44.957 / 0.813	45.345 / 0.817
10%	45.021 / 0.810	45.467 / 0.815
15%	44.830 / 0.808	44.834 / 0.811
20%	34.884 / 0.802	44.962 / 0.809

Note: All performance results are reported as BLEU / COMET, with the best result shown in bold.

Table 11: Performance of Context Strategy B Variants.

Ordering	All-In-One	Block-By-Block
Order-BC	44.880/0.807	44.929/0.814
Order-CB	44.678/0.809	43.977/0.805

Note: All performance results are reported as BLEU / COMET, with the best result shown in bold.

Lexical configuration: No Tag & Unique & 10% threshold.

Table 12: Performance of Context Strategy D variants.

Strategy	LLaMA Zero-Shot	GPT Zero-Shot
Baseline	34.384/0.749	43.925/0.789
Context A1	44.230/0.807	52.720/0.830
Context A2	40.845/0.811	53.373/0.836
Context B	45.099/0.821	54.469/0.842
Context C	35.041/0.743	43.206/0.778
Context D	45.907/0.820	54.287/0.842

Note: All performance results are reported as BLEU / COMET, with the best result shown in bold.

Table 13: Zero-Shot: LLaMA vs. GPT.

Strategy	LLaMA Zero-Shot	LLaMA Few-Shot
Baseline	34.384/0.749	37.187/0.765
Context A1	44.230/0.807	45.766/0.818
Context A2	40.845/0.811	46.606/0.824
Context B	45.099/0.821	46.325/0.826
Context C	35.041/0.743	36.207/0.751
Context D	45.907/0.820	46.246/0.825

Note: All performance results are reported as **BLEU / COMET**, with the best result shown in bold.

Table 14: LLaMA: Zero-Shot vs. Few-Shot.

C Context D for Case Study Examples

```

<dict>
mějčí: Macht
tutón: dieser
wopismo: Urkunde, Bescheinigung
na: auf, an
pučowanje: Wanderung, Reise
při: bei, dabei, an
seba: sich
</dict>
<morph>
Word=Mějće Morph=ADJ.Nom.Neut.Sing.Ord Lemma=mějčí
Word=tute Morph=DET.Dem.Nom.Neut.Sing Lemma=tutón
Word=wopismo Morph=NOUN.Nom.Neut.Sing Lemma=wopismo
Word=na Morph=ADP Lemma=na
Word=pučowanju Morph=NOUN.Loc.Neut.Sing Lemma=pučowanje
Word=při Morph=ADP Lemma=při
Word=sebi Morph=NOUN.Loc.Fem.Sing Lemma=seba
Word=. Morph=PUNCT Lemma=.
</morph>

```

Table 15: Selection of translation candidates. (Test 842)

```

<dict>
spisany: verfassen, schreiben, verfasst, geschrieben
wot: von, seit
mnich: Mönch
z: mit, aus
srjedžowěk: Mittelalter
rozprawjeć: berichten
list: Brief
wo: über, um, von
pokład: Schatz, Bodenschatz
wjerch: Fürst, Decke
</dict>
<morph>
Word=Spisany Morph=ADJ.Nom.Masc.Sing.Part.Pass
Lemma=spisany
Word=wot Morph=ADP Lemma=wot
Word=mnicha Morph=NOUN.Inan.Gen.Masc.Sing Lemma=mnich
Word=ze Morph=ADP Lemma=z
Word=srjedžowěka Morph=NOUN.Inan.Gen.Masc.Sing
Lemma=srjedžowěk
Word=, Morph=PUNCT Lemma=,
Word=rozprawješe Morph=VERB.Ind.Sing.3.Past.Fin
Lemma=rozprawjeć
Word=list Morph=NOUN.Inan.Nom.Masc.Sing Lemma=list
Word=wo Morph=ADP Lemma=wo
Word=pokładže Morph=NOUN.Inan.Loc.Masc.Sing Lemma=pokład
Word=wjercha Morph=NOUN.Inan.Gen.Masc.Sing Lemma=wjerch
Word=. Morph=PUNCT Lemma=.
</morph>

```

Note: The representation uses one token per line. Long lines wrap here only because of space constraints, with continuation lines indented to distinguish them from new token entries.

Table 16: Reliable morphological tags (Test 1145)

```

<dict>
koaliciski: Koalitionsvertrag, Koalitionsverhandlung,
Koalitionspartner
frakcija: Fraktion, der
namoŭić: auffordern, fordern, appellieren, aufrufen
knježerstwo: Regierung, Bundesregierung
namjety: Vorschlag, Änderungsantrag
dalokož: soweit, sofern
europski: europäisch
komisija: Kommission
być: sein, haben
ščěhować: folgen, verfolgen
bórže: zunächst, zuerst, erst, bald
pruwować: prüfen
</dict>
<morph>
Word=Koaliciskej Morph=ADJ.Loc.Fem.Sing Lemma=koaliciski
Word=frakciji Morph=NOUN.Loc.Fem.Sing Lemma=frakcija
Word=namoŭitej Morph=VERB.Ind.Dual.3.Pres.Fin
Lemma=namoŭić
Word=knježerstwo Morph=NOUN.Nom.Neut.Sing Lemma=knježerstwo
Word=namjety Morph=ADV.Sup Lemma=namjety
Word=MSPI Morph=PROP.N.Inan.Nom.Masc.Sing Lemma=MSPI
Word=, Morph=PUNCT Lemma=,
Word=dalokož Morph=CONJ Lemma=dalokož
Word=Europska Morph=ADJ.Nom.Fem.Sing Lemma=europski
Word=komisija Morph=NOUN.Nom.Fem.Sing Lemma=komisija
Word=je Morph=AUX.Ind.Sing.3.Pres.Fin Lemma=być
Word=ščěhuje Morph=VERB.Ind.Sing.3.Pres.Fin Lemma=ščěhować
Word=, Morph=PUNCT Lemma=,
Word=bórže Morph=ADV.Cmp Lemma=bórže
Word=pruwować Morph=VERB.Inf Lemma=pruwować
Word=. Morph=PUNCT Lemma=.
</morph>

```

Note: The representation uses one token per line. Long lines wrap here only because of space constraints, with continuation lines indented to distinguish them from new token entries.

Table 17: Incorrect morphological tags (Test 1897)