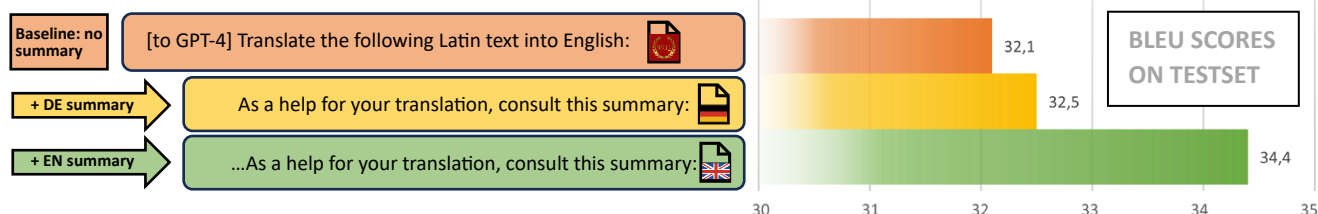


LLM-based Translation for Latin: Summaries Improve Machine Translation

Dominic P. Fischer, Martin Volk

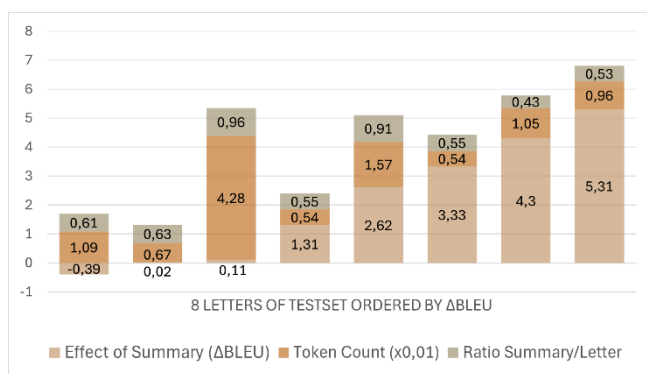
Could we use summaries that typically accompany historical text editions to boost translation?

In a Nutshell



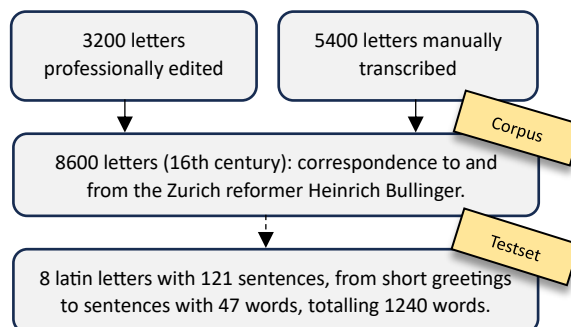
Providing the summary in the target language English improves translation by 2,3 BLEU points.

NB: The original summary is in German, while its English translation was created by GPT-4. Translation into German with the German summary yielded a comparable improvement of 2,0 BLEU points. ChrF scores exhibited the same trend: Improvements of around 2 points with the target language summary, and only marginal improvements with the non-target language summary.



Neither summary length nor token ratio appear correlated to summary effect.

The Corpus and Testset



A Look beyond

LATIN ORIGINAL	COMMISSUM HABEAS ADOLESCENTULUM; POLICEOR ET EGO ME NON INGRATUM FORE.
EN REFERENCE	I recommend the young boy to you; I assure you that I too will not be ungrateful.
W/O SUMMARY	You may have committed the young man; I also promise that I will not be ungrateful.
SUMMARY DE	You have the young man in your care; I promise that I will not be ungrateful.
SUMMARY EN	You may have the young man in your care; I promise that I will not be ungrateful.
ALIGNED SUMMARY EN	You have a committed young man; I promise that I will not be ungrateful.

Only correctly translated when summary is EN and not aligned!

- Experiments with aligned Summaries in the target language were conducted in both DE and EN: with a simple (P1) and a more verbose prompt (P2).
- Brittle approach: P2 led to the best results yet in English (BLEU 35,3), yet the worst results in German, worse than no summary at all.

	amount	percentage
the same	51	42
different	70	58
... of which		
neut. (≈)	36	30
pos. (+)	23	19
neg. (-)	4	3
major pos. (++)	7	6

On the left: Effects of summary inclusion as found through manual comparison of the 121 test set sentences translated with and without summary.

- Minor positive changes contained predominantly name corrections/normalizations (*Marcus* → *Markus*, *Caesar* → *Kaiser KarlV.*), clarifications of pronouns (*these* → *these news*, *he* → *it*)
- Negatives included wrongful reversals of such things, like *these questions* → *these*.
- Major positive changes greatly affected the understanding of the sentence, including changes of modus (imperative → conjunctive) or of an entire (part of a) sentence (cf. above).
- Major negative or neutral changes were not present.

Conclusion

Human-made summaries in the target language significantly enhance LLM-based translation quality.

LLMs are sensitive to prompt formatting: Aligning the summary and letter texts led to inconsistent results.

Rather than summary length, be it absolute or relative to the letter, the contents matter for translation quality.

Outlook

- Could summaries function as baseline for translation tasks?
- How does replacing the summary with unrelated text affect the translation?
- Would an LLM-generated summary have the same positive effect?