Year: 2024

LLM-based Machine Translation and Summarization for Latin

Volk, Martin ; Fischer, Dominic P ; Fischer, Lukas ; Scheurer, Patricia ; Ströbel, Phillip

LLM-based Machine Translation and Summarization for Latin

Martin Volk, Dominic P. Fischer, Lukas Fischer, Patricia Scheurer, Phillip B. Ströbel

Department of Computational Linguistics, University of Zurich

volk@cl.uzh.ch

Abstract

This paper presents an evaluation of machine translation for Latin. We tested multilingual Large Language Models, in particular GPT-4, on letters from the 16th century that are in Latin and Early New High German. Our experiments include translation and cross-language summarization for the two historical languages into modern English and German. We show that LLM-based translation for Latin is clearly superior to previous approaches. We also show that LLM-based paraphrasing of Latin paragraphs from the historical letters produces English and German summaries that are close to human summaries published in the edition.

Keywords: Large Language Models, Machine Translation, Latin, Early New High German, GPT

1. Introduction

The advent and wide accessibility of large language models (LLMs) with their inherent multilingual abilities has founded a new paradigm for machine translation (MT). LLM-based MT is similar to neural MT but has advantages for low-resource languages because of cross-language knowledge transfer and the possibility of targeted translation suggestions. In this paper we explore GPT-4 (OpenAI, 2023) as MT system for Latin to English and to German. We tested GPT-4's MT performance on letters from the 16th century that are in Latin and Early New High German (ENH-German).

The MT community site[1] documents that MT for Latin is "supported by 10 APIs". We checked the corresponding websites and found that five of these allow for online testing: Google Translate, LingvaNex, ModernMT, Niutrans and Yandex, for all of which Latin is one among more than 100 supported languages. Our tests show that translation quality for Latin to English and German is low for most of these systems. For a first glimpse of the results see table 1. We will detail the figures in section 4.1.

Fischer et al. (2022) described a neural MT system for Latin to German translation that outperformed Google Translate on their test set. In the meantime the situation has changed. Recent multilingual LLMs show surprising performance for machine translation.

This paper proves that GPT-4 produces superior MT quality for Latin to German and Latin to English if prompted appropriately. We also show that the same technology is able to produce paraphrases of the historical letters which compare favorably with human-written summaries.

2. Previous Work on LLMs for Latin

Work on using language models for Latin started with Bamman and Burns (2020) who built Latin-BERT on more than 600 million words. This established a new state of the art for part-of-speech tagging for Latin and for predicting missing text. Following up, Nehrdich and Hellwig (2022) used the Latin BERT embeddings for PoS tagging and dependency parsing for Latin. Lendvai and Wick (2022) used Latin BERT for Word Sense Disambiguation. They confirm that the contextualized BERT representations finetuned on *the Thesaurus Linguae Latinae*[2] score better than static embeddings from a bidirectional LSTM classifier.

With the advent of ChatGPT the question arose: How good is the GPT technology for historical languages? And why is it so good? Burns (2023) addresses these questions in his blog post and estimates that GPT-3 has been trained on more than 300 million tokens of Latin text. This is only a small fraction of its total training corpus but enough to model the language for high-performance on tasks like part-of-speech tagging, spelling and grammar correction for Latin texts.

Riemenschneider and Frank (2023) investigated the use of LLMs for Latin and Ancient Greek. They focus on Greek, but also built a multilingual model with English and Latin (with roughly 200 million tokens in each language as training corpora). For Latin they evaluated their model against the EvaLatin 2022 dataset (Sprugnoli et al., 2022) and report superior performance for part-of-speech tagging and lemmatization.

LLMs are trained on large amounts of text, most of which is typically in English. But even small amounts of other languages in the training data enable the system to respond in multiple languages and to learn to translate. Briakou et al. (2023) find

---

[1] https://machinetranslate.org

[2] https://thesaurus.badw.de/

that only 1.4% of training instances for the PaLM are bilingual which still results in good translation performance for medium-resource languages like Bulgarian, Hebrew, and Greek (Latin not included in this study), especially for MT into English. Fine-tuning LLMs on translation tasks results in improved MT performance, as Xu et al. (2023) showed for LLaMA-2.

GPT-based MT has been evaluated by various researchers. Laskar et al. (2023) report that Chat-GPT scores slightly worse than the state-of-the-art for MT between high resource languages like English and French, but it is better than previous systems in translating Romanian into English and French into German.

We are the first to present a systematic evaluation of LLM-based MT and summarization for Latin and Early New High German.

3. The Corpus of Letters in Latin and ENH-German

We work with a large corpus of 16th-century letters (Volk et al., 2022; Ströbel et al., 2024). 3100 have been professionally edited and another 5400 have been manually transcribed. The letters include historical characters (like ę, ů, å, ỏ). Abbreviations have been spelled out by the transcribers (e.g. the greeting *S et p in domino Iesu* has been completed into *S[alutem] et p[acem] in domino Iesu*, EN: *Greetings and Peace in the Lord Jesus*). Paragraph boundaries are set by the transcribers, sentence boundaries have been automatically added. Three quarters of the letters are in Latin, the rest in ENH-German, many letters contain code-switching between the two languages. The letters contain occasional sentences in Greek. All sentences have been automatically assigned a language tag based on a self-trained language identifier that is able to distinguish between ENH-German and Latin with high accuracy (see (Volk et al., 2022)).

The letters are part of the correspondence to and from the Zurich reformer Heinrich Bullinger. They deal with politics, theological debates, regional and European news as well as education and family matters. The letters thus give a first-hand view into the life 500 years ago. The correspondence network extended from Zurich throughout Germany towards Denmark, England, and Poland. Some letters traveled more than 1000 km.

4. Experiments with LLM-based MT

4.1. Evaluation against a Test Set

We used the test set of Fischer et al. (2022) which consists of 8 letters which have been manually translated into German by a domain expert. This test set focuses on Latin letters, but contains one sentence that is code-switching from Latin into ENH-German *Indixit dry musterpletz: Fŭssen, Werdt und Nördlingen* (EN: *He designated three recruiting places: Füssen, Donauwörth and Nördlingen*).

These 8 letters sum up to a total of 121 Latin sentences, some of which are short greetings, others are as long as 47 words. The whole test set consists of 1240 words on the Latin side and 1768 words in the corresponding human-translated sentences in German.

In order to be able to re-use the test set for MT into English we automatically translated the human-translated German sentences into English with GPT-4.

We then translated the test set with Google Translate and the other online MT systems from Latin into German and into English in order to obtain the baseline scores. In a second step we fed the complete test set to GPT-4 with a single prompt: "Translate the following text from Latin into L" where L was first German and then English.

The resulting scores are in table 1: In translating Latin to German, GPT-4 outperforms Google Translate by close to 10 BLEU points on the test set. The other online MT systems score clearly worse than Google Translate both when measured with BLEU and with ChrF.[3]

Fischer et al. (2022) had reported a BLEU score of 19.5 for their own system and 17.07 for Google Translate. When testing Google Translate now, we obtain a score of 17.53, which is marginally higher. This means that Google Translate has not improved much for Latin MT in recent years. However, GPT-4 surpasses these results significantly, reaching a BLEU score of 27.07 for Latin to German MT on the test set (see table 1 for an overview).

We observe a similar quality increase in translating from Latin to English. Google Translate reached a BLEU score of 25.22 for this language direction, while GPT-4 again betters it considerably, reaching 34.50. This is an enormous improvement. Table 2 shows the differences in translation quality for an example sentence from our test set.

The discrepancy between English and German can be attributed to two major reasons:

1. English is by far the highest resource language on the internet, and many researchers reported better MT into English than into other languages (cf. section 2 above).

2. We translated the German reference translation into English using GPT-4, which may in-

---

[3] BLEU (Papineni et al., 2002) is a precision-oriented word n-gram overlap metric which is often used in MT evaluation. ChrF (Popović, 2015) is a character n-gram metric which uses precision and recall.

MT System	Languages	MT Latin into German		MT Latin into English	
		BLEU	ChrF	BLEU	ChrF
GPT-4	unknown	**27.07**	**50.55**	**34.50**	**54.6**
Google Translate	134	17.53	43.23	25.22	47.48
LingvaNex	109	12.08	37.54	17.72	39.76
Yandex Translate	102	11.36	35.35	12.64	35.43
ModernMT	200	9.78	32.42	13.56	34.2
Niutrans	449	4.45	26.8	5.52	26.9

Table 1: BLEU and ChrF scores when translating the Latin test set (121 sentences) into German and English. The first column has the number of supported languages per system.

Original Latin	Quid sibi hęc societas velit, facile divinari potero.
Human Reference German	Was dieses Bündnis bedeutet, kann ich mir leicht vorstellen.
Human Reference English (transl. from DE by GPT-4)	What this alliance means, I can easily imagine.
MT System	**Translation**
GPT-4	What this alliance means, I can easily guess.
GoogleTranslate	What this company wants for itself, I can easily guess.
LingvaNex	What society wants for itself here is that I will be able to be divined easily.
Niutrans	I'm afraid it's hard to predict why Szczesny himself chose to participate in the league.

Table 2: A Latin sentence taken from a letter of Johannes Gast to Heinrich Bullinger, 1. April 1544 (see https://www.bullinger-digital.ch/letter/11930), translated by different translation systems, ordered by their automatic evaluation scores; with GPT-4 performing best and Niutrans worst.

troduce a bias, as the English translation may now be skewed towards a GPT-4 style of writing. When used as a reference for the evaluation of the Latin-English translations, that bias might lead to higher BLEU scores for GPT-4. As the BLEU score increase between Google Translate and GPT-4 remains approximately the same for both language pairs Latin-German and Latin-English, we conclude that this bias cannot be the decisive factor.

4.2. Evaluation against Paragraph Summaries

Reference translations are tedious and costly to create. With GPT having proven its quality in translation from Latin to both German and English, we investigated whether we can use letter summaries to evaluate GPT-4 translations.

For each of the 3100 edited letters we have a summary in German which was written by experts of the Institute for Swiss Reformation Studies. For the initial volumes of the edition, which date back to the 1970s, the summaries consisted of a few sentences or paragraphs. Over time the summaries increased in length. The three most recent volumes of the letter edition (published in the years 2017 to 2022, cf. Gäbler et al. (1973–2022)) contain paragraph-by-paragraph summaries that can be seen as shortened paraphrases. The alignment between the Latin paragraph in the letter text and the German summary is given. For an example letter with summaries see appendix A.

We used 10 medium-sized letters (5-7 paragraphs each) in Latin from the volume 18 of the edition, where the human-written summaries are paragraph-by-paragraph. Since the human summaries in this volume are close to the letter text we hypothesized that the summaries could be used as reference translations.

With this setup GPT-4 achieved a low 4.93 BLEU points when we compare the automatic translation to the human summary in German. In analogy to our test set evaluation we also translated the human summaries from German into English with GPT-4. GPT-4 MT from Latin to English then results in 6.80 BLEU. Google Translate resulted in 3.43 BLEU for German and 5.84 for English. Interestingly, the MT scores are slightly higher when we translate the summaries from German into English with DeepL, which proves that GPT-4 translation DE-EN of the reference texts does not favor the MT results LA-EN towards GPT-4. See table 3 for the results.

Evaluating GPT-4 MT against the human summaries shows again that GPT-4 clearly outperforms Google Translate. But the scores differ by few BLEU points only and do not show the GPT-4 advantage as clear as with the test set.

5. LLM-based Summarization

In the previous section we tested whether the German summaries in the letter edition may serve as reference translations. Here we extend this idea to check whether GPT-4 can produce English or

Model	MT LA into DE	MT LA into EN (GPT-4)	MT LA into EN (DeepL)
Google Translate	3.43	5.84	6.59
GPT-4	**4.93**	**6.80**	**7.47**

Table 3: BLEU scores when translating 10 Latin letters into German and English, evaluated against the human summary in German, and a machine-translated summary (DE-EN) in English

German summaries for Latin and ENH-German letters.

In this experiment, we used the same 10 Latin letters as above, as well as 10 ENH-German letters from the edition. We prompted GPT-4 to produce a paragraph-by-paragraph summary of the given letter in the following way: "I have this letter by {sender} to {addressee} with {nr} paragraphs: {original_letter}. For each paragraph, write a summary in English from a third-person perspective."

We evaluated again by comparing the GPT-4 output with the human summary in German and the machine-translated summary (DE-EN) in English. When summarizing in German, GPT-4 achieves a BLEU score of 6.23 for the Latin letters and 5.45 for the letters in ENH-German.

In order to evaluate the summarization into English, we used both GPT-4 and DeepL to translate the human summaries from modern German into English and used these translations as reference. For Latin-English, GPT-4 now scores 9.98 on the DeepL reference translation and 10.40 on the GPT-4 reference translation. For ENH-German to English, the scores are 7.75 on the DeepL translation and 8.48 on the GPT-4 translation.

The BLEU scores for the automatic summaries are low, but confirm that GPT's output in English is of slightly higher quality than in German. A comparison of the scores for the ENH-German letters with the Latin letters is not possible. These are different letters.

Even though the summarization scores are low, the summaries look very good. In order to check the quality and assess their usefulness, we conducted a manual evaluation of GPT's automatically produced German summaries, using the following criteria. We checked for each paragraph whether

- the names (persons, locations) that are mentioned in the human summary are also contained in the generated GPT summary

- the events and times of the human summary are included in the generated GPT summary

- the information from the human summary is contained **completely** in the generated GPT summary

- the information of the human summary is **correctly** contained in the generated GPT summary

Three annotators compared and judged the human-written summaries to the GPT-produced German summaries paragraph-by-paragraph.

This evaluation yielded the results in table 5. Names are well represented in the generated GPT summaries, in particular person names. GPT-4 shows some issues with consistency: "Schweiz" (EN: *Switzerland*) is repeatedly used synonymously to "Eidgenossenschaft" (EN: *confederation*), which historically does not make sense. The average human evaluation score with regards to the names is 47.2 out of 58.

Times and Events were best captured by GPT with a score of 54.5 out of 58. Dates and temporal expression were accurately transferred into the summary. With regards to completeness, human evaluation yields a score of 48.2 out of 58, showing slight differences between the generated and the reference summary. It is to be noted, however, that completeness is sometimes subjective, since the expert editors weigh events by importance and thus decide whether or not to include them in the summaries. In a few cases, GPT-4 provided additional information that was pertinent, yet not contained in the human summary. Correctness was the lowest of the 4 metrics, with 43.5 out of 58 points. In some cases, potentially sensitive or offensive information was not correctly rendered, possibly due to censoring by GPT. Moreover, mistranslation of a few words or phrases led to opposite interpretation (e.g. "mirari" as "admire" instead of "be astounded").

We also noted positively that GPT-4's summaries of our test letters are free of any hallucinations: all information that is found in GPT's summaries is derived from the original Latin letter.

6. Advantages of LLM-based MT

Our results show clear advantages of LLM-based MT quality for Latin and ENH-German over the previous generation of neural MT systems. In addition, there are some technical aspects that speak in favor of LLM-based MT.

6.1. Steering the Translation

One striking advantage of LLM-based MT is the possibility for the user to suggest the translation of specific terminology to the LLM. For example, we observed that GPT-4 translates the Latin word *caesar* with the same word in English. However, in our

Model	ENH-German into German	Latin	ENH-German into English (GPT-4 / DeepL)	Latin
GPT-4	5.45	6.23	8.48 / 7.75	10.40 / 9.98

Table 4: BLEU scores when summarising 10 ENH-German and 10 Latin letters paragraph-wise into German and into English

	Names	Times & Events	Complete	Correct
Judge 1	44	56	49	43.5
Judge 2	45	52	46.5	36.5
Judge 3	52.5	55.5	49	50.5
Average	47.2	54.5	48.2	43.5

Table 5: Evaluator scores for the four evaluation categories on automatic summarization. The maximum points per category is 58, which means 1 point each for the 58 paragraphs in the test letters.

context *caesar* refers to the German emperor (Karl V. until 1556, and Ferdinand I. afterwards). Adding the instruction "Translate 'caesar' with 'emperor' " to the GPT-4 prompt is enough to steer the translation of *caesar* with its inflected forms *caesarem, caesare* etc. to be translated in the desired way. If needed, the translation instruction can be enriched with world knowledge, e.g. by specifying the name of the respective emperor.

We observed such rare mistranslations not only with nouns but also with names. GPT-4 knows a surprising number of Latin city names and translates them correctly into modern day equivalents (e.g. Basilea → Basel, Lutetia → Paris, Tigurinę → Zurich). Still it gets confused when old names are homographs to modern names. In our case of 16th century Latin *Argentina* refers to the city of *Strasbourg* but is often mis-translated as the country name. The simple instruction "Translate 'Argentina' with 'Strasbourg' " solves this problem for us, since the country name does not occur in our texts.

In the experiments reported in this paper we did not use the option of steering the translation.

6.2. Preserving XML Tags

Our corpus is annotated in XML for sentence boundaries, person and place names, for footnotes and page breaks. In order to use this valuable information after translation, the XML tags need to preserved in the target language. On a side project we experimented with MT for Latin to English with XML tags for sentence boundaries and names. We find that they are well-preserved when we translate with GPT-4. This requires specific prompting to inform the system about the XML in the input and the request for preserving the tags in the output.

7. Conclusion

GPT-4's performance on Machine Translation for historical languages is impressive. We experi-mented with letters from the 16th century that are partly in Latin and partly in Early New High German. The quality for translating both languages to modern English and German is high, much higher than with previous neural MT technology. We measure an improved score of plus 10 BLEU points for both Latin to English and Latin to German LLM-based MT over Google Translate on a test set of 121 sentences. This is a huge improvement of the state-of-the-art.

In a second round of experiments we evaluated LLM-based paragraph-wise summarization against expert-written summaries. Our manual evaluation showed that the automatically generated summaries capture names, events and other pieces of information accurately.

We deal with letters that contain a lot of code-switching between Latin and ENH-German. Unlike previous MT the new generation of LLM-based MT is robust against language mix, which is a big advantage. We will investigate this aspect in more detail in future work.

This paper focused on GPT-4 as a prominent LLM example. Future work should compare GPT's performance to other LLMs like LLama or Google Gemini. There, it will also be interesting to check how translation quality can be improved further by fine-tuning the LLMs to the Latin and ENH-German translation task.

We argued that it is easy to feed special terminology to the system in order to influence the translation (see (Bogoychev and Chen, 2023) for a systematic study). We plan to investigate the steering of the translation as a finetuning step by automatically identifying terms that require special translation instructions.

9. Bibliographical References

David Bamman and Patrick J. Burns. 2020. Latin BERT: A contextual language model for classical philology. In *ArXiv*.

Nikolay Bogoychev and Pinzhen Chen. 2023. Terminology-aware translation with constrained decoding and large language model prompting. In *Proceedings of the Eighth Conference on Machine Translation (WMT)*, pages 890–896, Singapore. Association for Computational Linguistics.

Eleftheria Briakou, Colin Cherry, and George Foster. 2023. Searching for needles in a haystack: On the role of incidental bilingualism in PaLM's translation capability. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 9432–9452, Toronto, Canada.

Patrick J. Burns. 2023. Research Recap: How much Latin does ChatGPT 'know'? Online blog post of May 19th, 2023.

Lukas Fischer, Patricia Scheurer, Raphael Schwitter, and Martin Volk. 2022. Machine translation of 16th century letters from Latin to German. In *Proceedings of 2nd Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) at LREC-2022*, pages 43–50, Marseille.

Ulrich Gäbler, Endre Zsindley, Kurt Maeder, Matthias Senn, Kurt Jakob Rüetschi, Hans Ulrich Bächtold, Rainer Heinrich, Alexandra Kess, Christian Moser, Reinhard Bodenmann, Judith Steiniger, and Yvonne Häfner, editors. 1973–2022. *Heinrich Bullinger Briefwechsel*. Heinrich Bullinger Werke. Theologischer Verlag Zürich.

Md Tahmid Rahman Laskar, M Saiful Bari, Mizanur Rahman, Md Amran Hossen Bhuiyan, Shafiq Joty, and Jimmy Xiangji Huang. 2023. A systematic study and comprehensive evaluation of ChatGPT on benchmark datasets. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 431–469.

Piroska Lendvai and Claudia Wick. 2022. Finetuning Latin BERT for Word Sense Disambiguation on the Thesaurus Linguae Latinae. In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon*, pages 37–41.

Sebastian Nehrdich and Oliver Hellwig. 2022. Accurate dependency parsing and tagging of Latin. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022) at LREC*, pages 20–25, Marseille.

OpenAI. 2023. GPT-4 technical report.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of ACL*, pages 311–318, Philadelphia, PA, USA.

Maja Popović. 2015. chrf: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Frederick Riemenschneider and Anette Frank. 2023. Exploring large language models for classical philology. In *Proceedings of the 61st Annual Meeting of the ACL*, pages 15181–15199.

Rachele Sprugnoli, Marco Passarotti, Flavio Massimiliano Cecchini, Margherita Fantoli, and Giovanni Moretti. 2022. Overview of the EvaLatin 2022 evaluation campaign. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages (LREC)*, pages 183–188.

Phillip Benjamin Ströbel, Lukas Fischer, Raphael Müller, Patricia Scheurer, Bernard Schroffenegger, and Martin Volk. 2024. Multilingual workflows in Bullinger Digital: Data curation for Latin and Early New High German. *Journal of Open Humanities Data*, 10.

Martin Volk, Lukas Fischer, Patricia Scheurer, Raphael Schwitter, Phillip Ströbel, and Benjamin Suter. 2022. Nunc profana tractemus. Detecting code-switching in a large corpus of 16th century letters. In *Proceedings of LREC-2022*, pages 2901–2908, Marseille.

Haoran Xu, Young Jin Kim, Amr Sharaf, and Hany Hassan Awadalla. 2023. A paradigm shift in machine translation: Boosting translation performance of large language models. In *ArXiv*.

A. Example of a Mixed Language Letter

Letter in ENH-German and Latin	Human Summary (German) from the Heinrich Bullinger edition	GPT-4 Summary (English)
S. Gratulor tibi, honorande mi Myconi, ob recuperatam sanitatem tuam, quam dominus velit esse diuturnam, ut diu utilis esse pergas ecclesiae suae; quo etiam omnia tua studia convertas!	[1] Gut, dass es Myconius wieder besser geht! Der Herr möge ihn noch lange seiner Kirche erhalten. Ihr soll er sich völlig widmen!	Heinrich Bullinger expresses his congratulations to Oswald Myconius on recovering his health, hoping it endures so Myconius can continue serving the church and focusing his efforts on it.
A comitiis Badensibus ne hac quidem hora quicquam certi habemus. Legati enim nostri nec redierunt nec literam ulla de re scripserunt. Nihil ergo vel de literis caesaris, quibus de tu scribis, vel de ullis aliis rebus scribere possum.	[2] Die Zürcher Gesandten [Johannes Haab und Itelhans Thumysen], die noch auf der Badener Tagsatzung sind, haben keine Nachricht übermittelt. Bullinger hat also nichts über den von Myconius erwähnten Brief des Kaisers Karl V. gehört.	Bullinger informs Myconius that there has been no certain news from the Baden assembly, as their delegates have neither returned nor written back, leaving him unable to discuss the matters Myconius mentioned or any other issues.
Argentinam vix aliis recipiet in gratiam conditionibus caesar quam alias urbes, quas mellitis verbis et nimbo quodam promissionum occaecat; in fine videbitur cuius.	[3] Straßburg wird kaum unter anderen Bedingungen begnadigt werden als die anderen Städte, die vom Kaiser mit schönen Versprechungen benebelt wurden. Am Ende wird man sehen, von wem [diese Versprechungen ausgingen]!	He speculates that Strasbourg will not easily reconcile with the emperor under conditions different from other cities, which have been blinded by sweet words and promises, indicating a future revelation of true intentions.
...		
Ulma et reliquae miserę urbes exemplum praebent. **Was dörffend sich die barmhertzigen lüt also uffzegåben? Was nåmend sy das tüfelvolck hinyn? Was behaltend sy ir stett nitt unbefleckt und unbeherrschet?** Sed haec est manus domini percutientis nos propter ingratitudinem nostram et vitam poenitere nesciam.	[5] Ulm und die anderen armseligen Städte sind schon ein gutes Beispiel dafür. Warum mussten diese sich ergeben, fremde Besatzungen aufnehmen und ihre Freiheit preisgeben? Bestimmt bestraft Gott so die nicht bußfähigen Menschen für ihre Undankbarkeit.	He discusses the example set by Ulm and other unfortunate cities, questioning the pride of those who accept the deceitful, and lamenting the divine punishment reflected in their plight due to ingratitude and unrepentant living.
...		
Recte iudicas de concordia et discordia caesaris et papae. Ego arbitror saepiae esse atrorem. **Es ist bůbery, damitt die lüt umbgand. Trüw inen ja frylich der tüfel!** Quin potius annunciemus nos regnum dei et salutiferum evangelium filii dei, pontificis et regis nostri, qui syncere agit, neminem fallit et mox venturus est in iustitia, iudicaturus vivos et mortuos. Huic placere in omnibus satagamus! In illo vale cum omnibus bonis. Tiguri, 4. martii circa 9 antemeridianam 1547. Saluta fratres. Bullingerus tuus.	[8] Myconius' Beurteilung von Kaiser und Papst [Paul III.] ist zutreffend. Beide vertuschen nur ihre Kungeleien, um die Menschen besser an der Nase herumzuführen. Umso wichtiger ist es, das Evangelium Christi, des wahren Priesters und treuen Königs, der bald alle richten wird, weiter zu verkündigen! Gruß, auch an die Kollegen. Geschrieben gegen neun Uhr vormittags.	Bullinger critiques the relationship between the emperor and the pope, suggesting it's often more tumultuous than it appears, and advocates for the preaching of God's kingdom and the true gospel, urging to please God in all things. Bullinger concludes with greetings and a personal sign-off, noting the letter's time of writing in Zurich and asking Myconius to greet other brothers.

Excerpt of the human summary (German) vs. the automatic summary (English) of a letter from Heinrich Bullinger to Oswald Myconius, 4. March 1547. Sentences in Early New High German are in bold. See https://www.bullinger-digital.ch/letter/12884.