Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022), pages 43–50
Language Resources and Evaluation Conference (LREC 2022), Marseille, 25 June 2022

Machine Translation of 16th Century Letters from Latin to German

Lukas Fischer, Patricia Scheurer, Raphael Schwitter, Martin Volk
Department of Computational Linguistics, University of Zurich
fischerl@cl.uzh.ch

Abstract

This paper outlines our work in collecting training data for and developing a Latin–German Neural Machine Translation (NMT) system, for translating 16th century letters. While Latin–German is a low-resource language pair in terms of NMT, the domain of 16th century epistolary Latin is even more limited in this regard. Through our efforts in data collection and data generation, we are able to train a NMT model that provides good translations for short to medium sentences, and outperforms GoogleTranslate overall. We focus on the correspondence of the Swiss reformer Heinrich Bullinger, but our parallel corpus and our NMT system will be of use for many other texts of the time.

Keywords: machine translation, low-resource language, medieval Latin

1. Introduction

Heinrich Bullinger (1504-1575) was a Swiss reformer with an extensive correspondence network across Switzerland and Europe. Roughly 10,000 handwritten letters addressed to Bullinger, and 2000 letters penned by himself have been preserved, but only a quarter of them have been edited. The Bullinger Digital[1] project aims to bring Bullinger's complete correspondence into digital form and make it accessible to the general public and to scholars. This includes scanning the original letters, recognising the hand-writings of the many writers (Ströbel et al., 2022), and making the letters available online.

The letters deal with politics, everyday life and religious questions, and discuss anything from the plague to thunderstorms. Since the addressees and writers range from relatives and close friends to the King of England, the style varies from colloquial to formal. The letters are predominantly written in Latin (LA), the second most frequent language is Early New High German (ENHG), and a significant number of letters contain code-switching between the two languages (Volk et al., 2022).

We will provide a translation of all Latin sentences within these letters into modern German (DE), which will be generated by a customised Machine Translation system. Therefore, one of the project's goals is the development of a Machine Translation model that is optimized for 16th century Latin. In this paper, we outline our approach in collecting training data for our Machine Translation models, and discuss the strategies that improved the performance of our translation systems.

2. Sentence Alignment

As we collect the majority of our training data ourselves, a crucial step in our pipeline is sentence alignment, to extract sentence-based parallel segments for training. We use two strategies for this, which are outlined below.

2.1. Bitext Miner

The LASER (Language-Agnostic SEntence Representations) library (Schwenk et al., 2017) provides an encoder to create sentence embeddings that was trained on 93 languages, including Latin and German. The library also includes a script that utilizes these sentence embeddings to find similar sentences across languages. For instance, Schwenk et al. (2019) use this method to compile parallel corpora from Wikipedia articles.

The algorithm assigns each found sentence pair a margin score: the higher the score, the more likely are the two sentences close translations. Thus, by discarding sentence pairs with a score below a certain threshold, the quality of the remaining dataset will increase, at the expense of its size. Schwenk et al. test for the optimal threshold by training multiple NMT systems for four different language pairs. They find that translation systems trained on datasets cut off using a margin threshold of 1.04 yield the best translations across the tested language pairs, therefore we also use this threshold in our pipeline.

2.2. Vecalign

Thompson and Koehn (2019) also utilize the LASER sentence embeddings in their Vecalign algorithm. In contrast to the Bitext Miner, which is built to find similar sentences in large, unordered datasets, Vecalign aims to create sentence alignments in parallel documents, which includes one-to-many and one-to-zero sentence alignments, similar to Hunalign (Varga et al., 2007) or Bleualign (Sennrich and Volk, 2011). Thompson and Koehn demonstrate that Vecalign outperforms the former two aligners, which is why we also adopt it in our pipeline.

Vecalign is best used on documents that are close and complete translations. For example, a manual translation by a known expert would be a good fit, whereas web-crawled texts of different lengths will result in

[1] Bullinger Digital Project Website (German)

Testset	Segments	Token DE	Token LA
Bullinger	121	2,061	1,515
bible-uedin	200	5,571	3,575

Table 1: Number of segments, German and Latin tokens in the testsets.

poor alignments. In these situations, we use the Bitext Miner instead.

Vecalign reports an alignment cost for each found alignment, and we choose to drop alignments with an alignment cost above 1, to avoid including too much noise in our training data.

3. Collecting and Generating Parallel Data

Latin is a low resource language, especially in the context of Neural Machine Translation. While Latin has served as the language of Science and Church for centuries, only a limited number of texts are digitally available, and even fewer texts come with a close translation. For instance, the OPUS website (Tiedemann, 2016), which hosts large collections of parallel corpora, only includes 100,000 translated segments for the language pair LA–DE. This number does not change significantly for other language combinations with Latin. In contrast, the collection of English to German (EN–DE) training data that is available through OPUS includes 424 million segments.

Therefore, one of our main contributions is collecting and generating suitable training data for our NMT system. In this section, we describe our sources and techniques for the creation of our data set. We use two different approaches to automatically align translated sentences, as previously discussed in Section 2. Please refer to Table 2 for an overview of the training data collected, and Table 1 for the testsets that we set aside.

3.1. OPUS Corpora

As previously mentioned, a small number of parallel corpora is already available from the OPUS website (Tiedemann, 2016). We used the largest two of them in our training data.

Wikimatrix: This corpus was created by Facebook Research. It consists of automatically mined and aligned sentences from Wikipedia, using the Bitext Miner pipeline of Facebook's LASER framework (Schwenk et al., 2019). This way, 17,000 sentence pairs were automatically aligned. See section 2.1 for a detailed description of LASER. Latin Wikipedia articles are created by members of the Wikipedia community, and are available on a variety of predominantly modern topics. For example, the Latin and German sentences below are taken from an article on a video game.[2]

[2] https://la.wikipedia.org/wiki/Grand_Theft_Auto_III

la: *Unus enim ex anni 2001 venditissimis ludis factus est.*

de: *Ende 2001 war es das am zweithäufigsten verkaufte Spielzeug in den USA.*

en: *At the end of 2001, it was the second best-selling toy in the United States.*

Notably, the prepositional phrase *in den USA* (*in the United States*) is omitted in the Latin version. Indeed, as this is an automatically compiled corpus, sentence pairs are only quasi-parallel, therefore, some noise and rough translations are to be expected. Furthermore, sentences in this corpus tend to be short, with an average length of roughly 13 tokens per German sentence.

bible-uedin: As the most translated book in the world, the bible is an obvious choice for translations from Latin. We used the corpus collected by (Christodouloupoulos and Steedman, 2015). We shuffle the corpus and slice off 200 sentences to be used as a test set. This leaves 30,000 sentence pairs for training, with an average sentence length slightly above 20 German tokens.

3.2. Manual Translations

At the start of the project, a scholar of the Swiss Reformation Studies Institute manually translated a small number of the Bullinger collection. This serves as our primary test set (see Table 1).

In the meantime, we have periodically received additional manual translations by the Swiss Reformation Studies Institute, and we are adding them to our training data (Table 2). While 154 segments is a small number, these high quality translations of in-domain data are very valuable.

The example below (which stems from the test set) illustrates the epistolar language of our target domain:

la: *Diu nihil ad te scripsi, chare mi Myconi, sed modo copiosius tecum colloquar per librum, quem mitto.*

de: *Ich habe lange nicht an dich geschrieben mein lieber Myconius, aber nun möchte ich mich mit dir durch das Buch unterhalten, welches ich hier schicke.*

en: *I haven't written to you for a long time, my dear Myconius, but now I would like to talk to you through the book that I am sending here.*

3.3. Crawled Data

We collect a substantial part of our training data from the websites described in this section. For this we write customized scripts based on the Python library `scrapy`.

vatican.va: The official website of the Vatican is accessible in 10 languages, among them German and Latin. We crawl all sites of the Latin version and check whether they contain a hyperlink to a German translation. If so, we save both documents and automatically align sentences. We find that Vecalign performs well

Corpus	Segments	Token DE	Token LA
WikiMatrix	17,847	225,673	174,303
bible-uedin	30,288	685,293	523,050
Bullinger Translations	154	4,021	2,994
vatican.va	60,589	805,508	598,877
Nuntii Latini	6,139	105,799	98,401
BKV	21,573	1,045,300	812,786
Vulgate	35,620	810,524	610,769
Perseus (DeepL)	14,870	287,960	190,957
Zurich Letters (DeepL)	1,825	47,672	34,385
Blarer (backtranslation)	2,868	54,415	43,679
Regests (backtranslation)	24,188	544,765	370,998
TOTAL	**215,961**	**4,616,930**	**3,461,199**

Table 2: Number of segments, as well as German and Latin tokens for each corpus included in our training data. The German segments of the Perseus and Zurich Letters data sets were translated from English with DeepL. The Latin segments in the Blarer and Regests data sets are backtranslations from German using GoogleTranslate.

for most documents, if they are structured identically and contain close and complete translations. For documents with incomplete translations, we use the Bitext Miner pipeline instead, since this algorithm ignores the document structure and excels at finding sentence pairs in large datasets. Please refer to section 2 for more information on the two sentence alignment algorithms. This way, we extract 60,589 quasi-parallel sentence pairs with an average length of 13 tokens.

The translations consist of the constitutions, declarations and decrees of the Second Vatican Council, as well as scriptures from the Apostolic Constitutions, the Catholic Catechisms and papal encyclicals. The following example is taken from an open letter by Pope Benedict XVI:

la: *Haec omnia divisiones genuerunt sive apud clerum sive apud fideles.*

de: *Das alles hat Spaltungen sowohl im Klerus als auch unter den Gläubigen verursacht.*

en: *All this has caused divisions both in the clergy and among the faithful.*

While these texts use a register that is different from 16th century letters, their ecclesiastical vocabulary is a good match for our target domain.

Nuntii Latini: Since 2004, *Vatican News*[3] has been publishing a weekly news summary in Latin and German. A typical entry consists of three short paragraphs per language. Therefore, the alignment of the paragraphs is straightforward, and we are able to add another 6,139 sentences to our training data.

While current news in modern Latin are not a perfect fit to our target domain, *Vatican News* is a source of high-quality close translations and therefore suitable as training data. The following example was published earlier this year:[4]

la: *Ad Diem universalem Pacis, qui I die mensis Ianuarii celebratur, quod attinet, Franciscus Papa enixe ad pacem fortius in mundo fovendam admonuit.*

de: *Zum Weltfriedenstag am 1. Januar hat Papst Franziskus eindringlich zu mehr Frieden in der Welt gemahnt.*

en: *On the occasion of the World Day of Peace on 1 January, Pope Francis made an urgent appeal for more peace in the world.*

BKV[5]: The Library of the Church Fathers (German: *Bibliothek der Kirchenväter*) is a collection of ancient Christian literature and corresponding (mostly German or French) translations. Notable authors in this corpus are Ambrosius, Hieronymus, Augustinus and Saint Gregory the Great. The excerpt below is taken of the book *Pastoral Care*:

la: *Quod Moyses utrumque miro opere explevit, qui praeesse tantae multitudini et noluit et obedivit.*

de: *Beides hat Moses in bewunderungswürdiger Weise beobachtet, als er nicht Führer eines so großen Volkes werden wollte und doch gehorchte.*

en: *Moses observed both in an admirable way when he did not want to become the leader of such a large people and yet obeyed.*

We crawl all Latin source texts with a German translation, and extract 21,573 parallel segments from this source. Notably, the average sentence length in this corpus is over 40 German tokens. Since some of these translations were incomplete, we use the Bitext Miner instead of Vecalign.

Biblia Vulgata: The Vulgate is a Latin translation of the Bible, dating back to the 4th century. It has been translated into German by Joseph Franz von Allioli in the 1830s. Since the Vulgate is structured into numbered verses, aligning the translations was straightfor-

[3] `www.vaticannews.va`

[4] Nuntii Latini – Die IV mensis ianuarii MMXXII

[5] `https://bkv.unifr.ch/`

ward. We collected 35,620 parallel segments this way, providing some alternate translations to the verses contained in the **bible-uedin** corpus.

3.4. Perseus

While there is no large digital collection of German translations to classical Latin texts, the Perseus Digital Library (Clérice et al., 2022) includes a large number of English translations of canonical Latin literature. All texts are available for download via their git repository[6]. We use LASER to mine English-Latin sentence pairs, and then translate the English sentences into German using DeepL[7], to create a synthetic parallel corpus. As can be seen in table 2, this method yields 14,870 sentence pairs, with an average sentence length of 19 tokens in German. Below is a sample sentence, which originally stems from *The Epistles of Ovid*. The Latin and English sentences are collected from Perseus, the German is a translation by DeepL.

la: *Increpet usque licet - tua sum, tua dicar oportet; Penelope coniunx semper Ulixis ero.*

de: *Lass ihn schimpfen; ich bin dein und muss dein genannt werden; Penelope wird immer die Frau des Odysseus bleiben.*

en: *Let him chide on; I am yours, and must be called yours; Penelope will ever remain the wife of Ulysses.*

3.5. Scans and Backtranslations

Some letters of the Bullinger correspondence have already been transcribed and translated in other edition projects. Since they are available in print only, we scan these letters and use an OCR software[8] to extract the text.

Zurich Letters (Hastings, 1968): This edition consists of the correspondence between Bullinger and other Swiss reformers with English Bishops. The letters are available in Latin and English, which allows us to automatically align sentences using Vecalign. This way, we collect 1825 English-Latin sentence pairs, and we use DeepL once again to translate the English sentences into German. The average sentence length of 16 tokens is identical to the one found in the Bullinger translations described in section 3.2, which is unsurprising, as they are from the same domain.

Below is a sample sentence from the Zurich letters data set. The German translation is by DeepL, while the Latin and English sentences are taken from the edition.

la: *Superiori die accepimus literas ex Anglia, quibus mors Mariae, inauguratio Elisabeth, et obitus cardinalis Poli confirmatur.*

de: *Wir haben gestern einen Brief aus England erhalten, in dem der Tod von Maria, die Thronbesteigung von Elisabeth und das Ableben von Kardinal Pole bestätigt wird.*

en: *We yesterday received a letter from England, in which the death of Mary, the accession of Elizabeth, and the decease of cardinal Pole is confirmed.*

Blarer Correspondence (Schieß, 1908): This edition contains the correspondence of the Blarer brothers (Ambrosius and Thomas), who were both in frequent contact with Heinrich Bullinger. Unfortunately, the German translations that come with these letters are summaries. Therefore it is not possible to create a sentence alignment suitable for training an NMT model. However, as GoogleTranslate vastly improved the quality for translations from and to Latin (see Section 4.1), we exploit this to translate the German sentences of the Blarer letters into Latin, following the idea of back-translation proposed by Sennrich et al. (2016b).

This results in 2,868 German segments with Latin translations as additional training data. Since back-translations are synthetically generated and likely erroneous, we follow Caswell et al. (2019) and tag these segments with a special symbol (<*bt*>), before adding them to our training data:

la: *<bt> Mantuae concilium in Septembris sive proximo anno dilatum erit.*

de: *Das Konzil von Mantua soll auf September oder nächstes Jahr verschoben sein.*

en: *The Council of Mantua is said to be postponed to September or next year.*

Regests: The already edited Bullinger letters are prefaced by a regest, a German summary of the letter's content. We also use GoogleTranslate to create back-translations of these letter summaries. Unfortunately, typical characteristics of letters, such as direct speech and the use of second person singular, are exchanged with third person statements in the summaries. Nevertheless, using these texts as training data ensures that the model encounters the names of most of Bullinger's correspondents, as well as other named entities and specific vocabulary. The summaries consist of 24,188 segments, with an average length of 23 German tokens. The example below highlights characteristics of the summaries, as well as the sometimes erroneous nature of the back-translations in the duplication of *officorum* in Latin.

la: *<bt> Bullinger, qui officiorum officiorum assidue intermittitur, iustam causam habet ut alios petat ut hoc exemplum faciant.*

de: *Bullinger, der andauernd von Amtspflichten unterbrochen wird, hätte guten Grund, andere mit dieser Abschrift zu beauftragen.*

en: *Bullinger, who is constantly interrupted by official duties, would have good reason to hire others to do this transcript.*

[6] https://github.com/PerseusDL/canonical-latinLit

[7] https://www.deepl.com/

[8] ABBYY FineReader

4. Machine Translation

We have continually added more data to our training set and thus gradually improved our NMT system. Therefore, we are going to present eight models to highlight the impact of additional training data, as well as crucial strategies that improved the translation quality. In addition, we also describe our baseline models in this Section.

4.1. GoogleTranslate Baseline

We choose GoogleTranslate as a baseline to compare our NMT systems against. We use the BLEU metric (Papineni et al., 2002) to compare the performance of our systems to the baseline.

When we started the project in early 2021, we created a set of translations of the testsets using GoogleTranslate. At this time, GoogleTranslate still used Statistical Machine Translation (SMT) for language pairs that include Latin (GoogleTranslate, 2021) and while the translation of the bible testset was of an acceptable quality, the translation of the Bullinger testset was mostly unintelligible, which was also reflected in the low BLEU score of only 7.36 (see Table 3, B1, GoogleTranslate SMT).

However, GoogleTranslate implemented an NMT model for translations from Latin (to all of the 100+ available languages of the online system) over the course of the last year, achieving a much higher BLEU score of 17.07 for German when we translated the test sets again in the fall 2021 (B2, GoogleTranslate NMT).

4.2. Transformer Architecture

In all experiments, we use the transformer architecture with the base configuration by Vaswani et al. (2017). More specifically, we use the sockeye framework (Hieber et al., 2017). While Araabi and Monz (2020) have shown that optimizing hyperparameters for low-resource NMT greatly improves translation quality, we plan to test this once we have completed data collection, as the optimal settings change with increasing training corpus size.

Table 3 shows the results of our experiments E1–E8. Each subsequent experiment incorporates all training data and adopted strategies of the previous experiments.

4.3. Impact of Training Corpora, E1–E3

In our first experiment (E1), we use the **wikimatrix**, **bible-uedin**, **vatican.va**, **Nuntii Latini** and **Vulgate** corpora as training data, which amounts to 150,000 sentence pairs, or roughly 2 million Latin tokens. With this setup, we achieve a BLEU score of 11.14 on our indomain testset, which is comparable to the results reported by Martínez Garcia and García Tejedor (2020). They compile a Latin–Spanish training corpus from Saint Augustine translations with a similar size (2,2 million Latin tokens) to train a NMT model, which reaches a BLEU score of 10.01 on their indomain testset.

The E1 model already outperforms the SMT baseline for both testsets by a great margin. However, the GoogleTranslate NMT baseline still has a lead of 6 BLEU points on the Bullinger testset. Our training data has significant overlaps with the bible testset, due to the inclusion of the **Vulgate** translations and the fact that the bible is often quoted on **vatican.va**. Therefore, a direct comparison with the baseline is not possible for this testset.

In experiment E2, we add the **Bullinger translations** and the **BKV** corpus to the previous training data, which increases the training data size by 21,000 segments and bumps the BLEU score up to 12.15.

In in E3, we further add the **regest** backtranslations, which is another 24,000 segments, and an increase of BLEU points to 13.72.

While adding training data gradually increases BLEU, we observe that all previous NMT models particularly struggle with longer sentences, and they often fail at easy tasks such as correctly copying digits or translating dates.

4.4. Pretraining, E4–E7

Following Zoph et al. (2016), we add a pretraining step to our pipeline to make our model more robust and especially to improve its ability to preserve numbers (e.g. denoting years or measurements). The idea is that the model learns fluent German from a larger training corpus. The source language of this pretraining model should be closely related to Latin (Zoph et al., 2016), which is why we use Italian. However, we expect that using a German–Spanish or German–French corpus for pretraining will yield a similar result.

In E4, we download the German–Italian parallel segments of the Europarl corpus (Koehn, 2005) and train an Italian to German NMT system on this data. We use the same parameters as for the previous experiment. After training converges, we replace the IT–DE training data with our Latin–German corpora (all corpora listed in experiments **E1–E3** in Table 3) and continue training until the model converges again. This results in an increase of over 1.2 BLEU. Moreover, pretraining has the desired impact of improving the model's capability of correctly translating dates and numbers.

In E5 and E6 we add the **perseus**, **Zurich letters** and **Blarer** datasets to the training data, while retaining the pretraining in the pipeline. Overall, we add 16,000 segments, which improves the BLEU score by 1.5.

In E7, we replace the Europarl corpus with the Paracrawl corpus (Bañón et al., 2020) for pretraining. Thus, we increase the size of the pretraining data from 1,2 million to 6 million segments, which results in a stronger IT–DE pretrain model. In addition, this also improves the quality of the LA–DE model by another 0.6 BLEU points, which means our NMT performs as well as the GoogleTranslate-NMT baseline.

Exp	Description	N Segments	Bullinger	bible-uedin
B1	GoogleTranslate SMT (Feb. 2021)	-	7.36	9.67
B2	GoogleTranslate NMT (Oct. 2021)	-	17.07	15.89
E1	vulgate, vatican.va, nuntii latini, wikimatrix, bible	150,483	11.14	27.89
E2	+ Bullinger translations, BKV	172,210	12.15	28.10
E3	+ regests	196,398	13.72	27.65
E4	+ pretraining (Europarl IT–DE)	196,398	14.92	28.00
E5	+ perseus, Zurich letters	213,093	16.05	28.47
E6	+ blarer	215,961	16.57	26.97
E7	+ pretraining (Paracrawl IT–DE)	215,961	17.15	28.74
E8	+ normalize	215,961	19.50	28.63

Table 3: BLEU values on the Bullinger and the Bible testset achieved by the two GoogleTranslate baselines (**B1** and **B2**) and our own experiments **E1–8**. Our experiments all build on each other, thus, **E8** incorporates all the data sets of the previous seven experiments. **N Segments** gives the total number of segments in the training set for each experiment.

4.5. Normalization, E8

We use the CLTK (Classical Language Toolkit) normalizer (Johnson et al., 2021) to preprocess the Latin segments of our training data. In particular, CLTK automatically replaces any letters that have accents or macrons with their base form (e.g. *à* is replaced with *a*) and splits ligatures into their base characters (e.g. *æ* to *ae*).

CLTK also includes the option to replace all instances of *j* with *i* and *v* with *u*, as Latin often does not distinguish between these letters. However, we find that this has a negative effect on the BLEU value and therefore do not implement this option.

In addition, we also remove the diacritics from all *e caudatae* (ę and Ę), which frequently occur in the Bullinger correspondence, but are used inconsistently. Similarly, Sennrich et al. (2016a) also remove diacritics in the source language for Romanian–English translation to great effect, improving their BLEU score by 1.4.

As shown in Table 3, E8, adding normalization to Latin source sentences in addition to pretraining greatly improves the translation quality, as we achieve a BLEU score of 19.5. This system produces good translations for shorter and mid-length sentences. For instance, find below this model's translation of the example sentence from section 3.2:

la: *Diu nihil ad te scripsi, chare mi Myconi, sed modo copiosius tecum colloquar per librum, quem mitto.*

de: *Lange Zeit habe ich dir nichts geschrieben, mein lieber Myconius, aber jetzt werde ich mich ausführlicher mit dir unterhalten durch das Buch, das ich sende.*

en: *For a long time I have written nothing to you, my dear Myconius, but now I shall converse with you more fully through the book I am sending.*

This is a good translation, and even the name *Myconius* is translated correctly. In the following example translation, however, the idiom *die Spreu vom Weizen trennen (separate the wheat from the chaff)* is translated too literally:

la: *Spero tamen dominum tanto magis nos liberaturum, quanto magis paleae hae a tritico segregantur.*

de: *Ich hoffe jedoch, dass der Herr uns um so mehr befreien wird, je mehr diese Stroh von dem Weizen getrennt werden.*

en: *However, I hope that the more these straws are separated from the wheat, the more the Lord will set us free.*

For longer, more complicated sentences, our best system still struggles to produce accurate translations. However, with this setup, we outperform GoogleTranslate by two BLEU points.

5. Conclusion and Future Work

We have shown that by using a combination of existing parallel corpora, manual translations, web crawls, digitized texts and synthetic training data, we surpass the translation quality of our baseline.

Additionally, we have shown that pretraining on a similar language pair and normalizing Latin diacritics greatly enhances translation quality. Since there is limited previous research on Machine Translation from or into Latin, our research fills in an important gap in Digital Humanities and will hopefully inspire similar projects in the future.

While we are currently ahead of the GoogleTranslate baseline by two BLEU points, we have yet to evaluate whether this difference is apparent to human evaluators, and we plan to carry out such a qualitative comparison.

Furthermore, we are still collecting more training data, for example the translations by Schwitter (2018). In addition, we plan to greatly increase the amount of back-translations in our training corpus, and test different methods of data augmentation, for example the tasks

proposed by Sánchez- Cartagena et al. (2021).
Finally, once we are happy with the size of our training corpus, we will optimize the hyperparameters of the transformer architecture.

6. Bibliographical References

Araabi, A. and Monz, C. (2020). Optimizing Transformer for Low-Resource Neural Machine Translation. *arXiv:2011.02266 [cs]*, November. arXiv: 2011.02266.

Bañón, M., Chen, P., Haddow, B., Heafield, K., Hoang, H., Esplà-Gomis, M., Forcada, M. L., Kamran, A., Kirefu, F., Koehn, P., Ortiz Rojas, S., Pla Sempere, L., Ramírez-Sánchez, G., Sarrías, E., Strelec, M., Thompson, B., Waites, W., Wiggins, D., and Zaragoza, J. (2020). ParaCrawl: Web-scale acquisition of parallel corpora. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4555–4567, Online, July.

Caswell, I., Chelba, C., and Grangier, D. (2019). Tagged Back-Translation. *arXiv:1906.06442 [cs]*, June. arXiv: 1906.06442.

Christodouloupoulos, C. and Steedman, M. (2015). A massively parallel corpus: the Bible in 100 languages. *Language resources and evaluation*, 49(2):375–395.

Clérice, T., Cerrato, L., Babeu, A., Almas, B., Jovanović, N., annettegessner, Burns, P. J., Stella, mkonieczny9805, Munson, M., Jøhndal, M., maryam foradi, zachhimes, MMernitz, Seydi, M., Celano, G. G. A., Scott, S., Huskey, S. J., and TDBuck. (2022). Perseusdl/canonical-latinlit. `https://doi.org/10.5281/zenodo.6418631`, April.

GoogleTranslate. (2021). Google Languages. `https://translate.google.com/intl/en/about/languages/#!/la`. (Accessed on 02/11/2021).

Hastings, R. (1968). *The Zurich Letters: Comprising the Correspondence of Several English Bishops and Others, with Some of the Helvetian Reformers, During the Early Part of the Reign of Queen Elizabeth*. Cambridge. Original Publication: 1842-1845.

Hieber, F., Domhan, T., Denkowski, M. J., Vilar, D., Sokolov, A., Clifton, A., and Post, M. (2017). Sockeye: A toolkit for neural machine translation. *CoRR*, abs/1712.05690.

Johnson, K. P., Burns, P. J., Stewart, J., Cook, T., Besnier, C., and Mattingly, W. J. B. (2021). The Classical Language Toolkit: An NLP framework for pre-modern languages. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 20–29, Online, August. Association for Computational Linguistics.

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand, September.

Martínez Garcia, E. and García Tejedor, Á. (2020). Latin-Spanish Neural Machine Translation: From the Bible to Saint Augustine. In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 94–99, Marseille, France, May. European Language Resources Association (ELRA).

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of ACL*, pages 311–318, Philadelphia, PA, USA.

Sánchez-Cartagena, V. M., Esplà-Gomis, M., Pérez-Ortiz, J. A., and Sánchez-Martínez, F. (2021). Rethinking Data Augmentation for Low-Resource Neural Machine Translation: A Multi-Task Learning Approach. *CoRR*, abs/2109.03645.

Schieß, T. (1908). *Briefwechsel der Brüder Ambrosius und Thomas Blaurer*. Freiburg.

Schwenk, H., Tran, K. M., Firat, O., and Douze, M. (2017). Learning joint multilingual sentence representations with neural machine translation. *CoRR*, abs/1704.04154.

Schwenk, H., Chaudhary, V., Sun, S., Gong, H., and Guzmán, F. (2019). Wikimatrix: Mining 135m parallel sentences in 1620 language pairs from wikipedia.

Raphael Schwitter, editor. (2018). *Martin Le Franc, Agreste Otium – De Bono Mortis*. Wiesbaden.

Sennrich, R. and Volk, M. (2011). Iterative, MT-based Sentence Alignment of Parallel Texts. In *Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)*, pages 175–182, Riga, Latvia, May. Northern European Association for Language Technology (NEALT).

Sennrich, R., Haddow, B., and Birch, A. (2016a). Edinburgh Neural Machine Translation Systems for WMT 16. *arXiv:1606.02891 [cs]*, June.

Sennrich, R., Haddow, B., and Birch, A. (2016b). Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany, August. Association for Computational Linguistics.

Ströbel, P. B., Clematide, S., Volk, M., Schwitter, R., Hodel, T., and Schoch, D. (2022). Evaluation of HTR models without ground truth material. *CoRR*, abs/2201.06170.

Thompson, B. and Koehn, P. (2019). Vecalign: Improved sentence alignment in linear time and space. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1342–1348, Hong Kong, China, November. Association for Computational Linguistics.

Tiedemann, J. (2016). OPUS – parallel corpora for everyone. In *Proceedings of EAMT*, page 384, Riga, Latvia.

Varga, D., Halacsy, P., Kornai, A., Nagy, V., Nemeth, L., and Tron, V. (2007). Parallel corpora for medium density languages. page 7.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of NIPS*, June.

Volk, M., Fischer, L., Scheurer, P., Schwitter, R., Ströbel, P., and Suter, B. (2022). Nunc profana tractemus. detecting code-switching in a large corpus of 16th century letters. In *Proceedings of LREC-2022*, Marseille.

Zoph, B., Yuret, D., May, J., and Knight, K. (2016). Transfer Learning for Low-Resource Neural Machine Translation. *arXiv:1604.02201 [cs]*, April.