Ancient Voices, Modern Technology: Low-Resource Neural Machine Translation for Coptic Texts *

Maxim Enis ME4@WILLIAMS.EDU
Andrew Megalaa ANM4@WILLIAMS.EDU
Williams College Computer Science

1. Introduction

Google Translate includes language translation models for a wide variety of language pairs, however many of the lesser-spoken languages in the world are not included. Some initiatives, such as Meta AI's No Language Left Behind program (NLLB) (Team et al., 2022), aim to address this concern. Even so, among the more than 31000 historically spoken languages, only a vanishingly small number have been successfully incorporated into translation models (Armstrong).

Coptic is a late-stage development of the Egyptian language that was primarily spoken in the period c. 325 – c. 800 AD (Layton, 2007). It is classified as extinct by the UNESCO Atlas of the World's Languages in Danger (UNESCO, 2010) and has no native speakers. However, many historical Egyptian texts are written in Coptic, and linguists and historians are interested in deciphering these texts. The Coptic Scriptorium is one such research program, with ongoing efforts to translate all texts available within the website (Schroeder and Zeldes, 2013-2023). However, due to the large quantity of texts and lack of qualified translators, many Coptic texts remain untranslated to this day.

In the present day, Coptic is primarily used in religious contexts within the Coptic Orthodox Church. However, there is community interest in reviving the language for everyday use. Despite past efforts that included developing resources on its vocabulary, grammar, and establishing teaching institutions, these initiatives did not lead to widespread adoption within the community (Atiya, 1991). The limited availability of online resources for learning Coptic and the absence of contemporary texts in the language pose significant challenge for future revival efforts, hindering its appeal to new learners.

To address these challenges, we created, to our knowledge, the first-ever contextual machine translator in the Coptic language. Using state-of-the-art techniques in low-resource machine translation, we demonstrate strong performance in translating unseen texts from a diverse range of religious sources. We also provide new translations for over 5700 previously untranslated sentences, completing translations for 109 Coptic texts. The model already demonstrates potential for utilization by historians, scholars of Early Christianity, and Egyptologists, as it automatically delivers the first-ever translations for a range of ancient religious texts. Ultimately, we aim to open-source the model code and host our translator for free use in the public domain[1]. In doing so, we hope to advance historic literature, assist in revival programs, and improve public fluency.

*. This title was inspired by GPT-4.
1. Website will be released at https://coptictranslator.com.

2. Preliminaries

Our approach uses deep neural networks and Transformers to leverage recent technological advancements in neural machine translation. On a high level, we fine-tune an existing Transformer-based multilingual model supplied on the Hugging Face libary on parallel Coptic data.

2.1 Neural Networks

Our approach uses neural networks to provide quality translations of Coptic sentences. For preliminary background describing neural networks, see Appendix 8.1.

2.2 NLP Essentials

In our project, we use state-of-the-art natural language processing architectures to design our machine translation model. In the following text, we provide a brief explanation of the technical details behind these architectures.

2.2.1 TOKENIZATION

In natural language tasks, neural networks usually cannot effectively take raw text as input. Rather, text should be *tokenized* into a finite set of categorical units, called tokens, which the model takes as input. The set of tokens which the model uses is called the *vocabulary*. We can represent tokens with *one-hot encoding vectors*, which are high-dimensional vectors (dimension equal to the vocabulary size) with zeroes in every entry except for one, uniquely representing the token.

2.2.2 ATTENTION

When humans read natural language sentences, we typically do not pay equal "attention" to each word within the sentence. In particular, different words carry varying importance and provide context to other words within the sentence. In machine learning, "attention" mathematically captures these relations by quantifying how much each pair of words contributes to the final sentence meaning.

2.2.3 THE TRANSFORMER

Prior to 2017, LSTMs were the state-of-the-art in attention-based deep learning architectures (Hochreiter and Schmidhuber, 1997). By using recursion to relate each following word in the sentence, they implemented a kind of "recurrent attention". However, since all sentences needed to be parsed sequentially, for long input sentences, LSTMs faced problems with efficiency. Therefore, in 2017, (Vaswani et al., 2017) described a new attention-based architecture: the Transformer. By using a learned, non-recursive attention mechanism, they solved the aforementioned problems with LSTMs. Due to their improved attention mechanism, they excel in tasks that require understanding the context and relationships within data, particularly in successful large language models like GPT-4.

The architecture consists of two primary components: an encoder, and a decoder. The encoder consists of a stack of N identical layers, each of which contains two sublayers: an attention head, and a feedforward neural network. The attention heads mathematically compute attention relations as vectors and augment the subsequent feedforward neural network with these resulting relations.

The attention heads compute learned attention matrices describing how tokens in the sentences relate to one another using the following equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

for learned query matrix Q, key matrix K, and value matrix V. Meanwhile, the softmax function scales the matrix such that each row of the matrix can be interpreted as a probability vector, with all entries between 0 and 1 and summing to 1.

Ultimately, the "encoder" outputs a reduced-size vector encoding of the its input. The decoder takes the vector encoding of the last n tokens and produces a probability distribution across all possible tokens predicting the conditional probabilities of seeing the next token.

2.3 Neural Machine Translation

Machine translation is the task of translating a sentence from a source language (for example, Coptic) to a sentence with the same meaning in a target language (for example, English). Transformers are the state-of-the-art architecture in neural machine translation (NMT), or machine translation using deep learning (Vaswani et al., 2017). The powerful attention mechanism allows translation models to effectively learn two crucial components:

1. How to encode sentences from the source language into a k-dimensional vector, representing the *semantic meaning* of the source sentence,
2. How to decode sentences from a semantic meaning vector into a sentence in the target language.

2.3.1 TRANSFER LEARNING

By training transformer-based NMT architectures on extremely large parallel datasets between languages, translation models can achieve near human-level performance (Haddow et al., 2022). However, for low-resource languages like Coptic, traditional training techniques would result in extremely poor-performing models due to the lack of data. In 2016, (Zoph et al., 2016) found that by initializing the parameters of the NMT architecture to that of a "parent model" trained on high-resource language pair, and training the resulting model on the low-resource language pair, the performance of the translation model will substantially improve. In other words, Transformers can "transfer" knowledge about translation from one language to another by fine-tuning the high-resource model on the low-resource data.

2.3.2 BACKTRANSLATION

In some cases, a language may have a large corpus of *monolingual* sentences, most of which are unannotated. In these cases, translation quality can often be improved using a technique known as *backtranslation* (Sennrich et al., 2016). Often used alongside transfer learning, it involves using the parallel corpus to train a model that translates a low-resource language to a high-resource one, then using that model to translate the monolingual, low-resource corpus. These translations are then used as training data for the high-resource to low-resource model. This is generally the direction that backtranslation is applied in since it's easy to make a model that "understands" the high-resource language.

2.4 Evaluation Metrics

When creating a translation model, it is crucial to develop effective evaluation metrics to judge the quality of the model translations. For our project, we consider two popular evaluation metrics: BLEU, and chrF. The evaluation metrics are described in greater detail in Appendix 8.2.

3. Data

We obtained our data from the Coptic Scriptorium, which contains the largest parallel corpus for the Coptic language. All texts are translated by experts in the field. Original Coptic texts contain no spaces, but the Scriptorium normalizes the words and delimits them with spaces. The normalized-spaced sentences are labeled "norm_group". Additionally, the Scriptorium further tokenizes "norm_group" into more subwords, which they label "norm". This data was provided in a grid-formatted file that we parsed into a CSV file.

eng	She labored to stop him, but she was not able to.
norm	ⲛⲧⲟⲥ ⲇⲉ ⲁ ⲥ ⲭⲓⲥⲉ ⲉ ⲥ ϣⲱϣⲧ ⲙⲙⲟ ⲫ ⲙⲡⲉ ⲥ ϣ ⲃⲙⲃⲟⲙ ⲉⲣⲟ ⲫ
norm_group	ⲛⲧⲟⲥ ⲇⲉ ⲁⲥⲭⲓⲥⲉ ⲉⲥϣⲱϣⲧ ⲙⲙⲟⲫ ⲙⲡⲉⲥϣⲃⲙⲃⲟⲙ ⲉⲣⲟⲫ
unnormalized	ⲛⲧⲟⲥⲇⲉⲁⲥⲭⲓⲥⲉⲉⲥϣⲱϣⲧⲙⲙⲟⲫⲙⲡⲉⲥϣⲃⲙⲃⲟⲙⲉⲣⲟⲫ
norm_romanized	ntos deie a s hiauseie eie s shoousht mmo f mpeie s sh shmshom eiero f
norm_group_romanized	ntos deie ashiauseie eiesshoousht mmof mpeiesshshmshom eierof
unnormalized_romanized	ntosdeieashiauseieeiesshoushtmmofmpeiesshshmshomeierof
norm_greekified	ντοσ δε α σ ηισε ε σ σωστ μμο f μπε σ s ϲμϲομ ερο f
norm_group_greekified	ντοσ δε ασηισε εσσωστ μμοf μπεσϲμϲομ ερof
unnormalized_greekified	ντοσδεασηισεεσσωστμμοfμπεσϲμϲομερof

Table 1: Example row in dataset, after preprocessing and augmentation

3.1 Data Cleaning

Before creating any additional columns, we preprocessed all existing data in the dataset:

- We removed all rows where the Coptic field was empty.
- We identified all occurrences of accidental Latin characters in the Coptic field and replaced them with their associated Coptic letter.
- We removed all special characters from the Coptic text.
- We replaced instances of "&apos" to an apostrophe and the "·" character to regular periods.
- We filtered out all Coptic text with English interjections or with missing text denoted by "[...]".

Prior to preprocessing, our dataset included 33364 rows. After filtering, we retained 32535 rows.

3.2 Coptic Transcription

Because we fine-tuned an existing multilingual translation model, the model vocabulary was fixed. In other words, the tokens input into the model could not change. Since the model did not include Coptic characters in its vocabulary, we were unable to pass raw Coptic as input into the model.

3.2.1 ROMANIZATION

To resolve the above problem, we decided to take the approach of Amrhein and Sennrich to "romanize" the Coptic text into Latin characters (Amrhein and Sennrich, 2020). Using the `uroman` libary[2], we created additional columns "norm_romanized" and "norm_group_romanized" consisting of the romanized representations of the respective columns. The `uroman` algorithm uses a (potentially non-reversible) greedy dictionary mapping to map Coptic characters into associated strings in the Latin alphabet - see Table 1 for more information.

3.2.2 GREEKIFICATION

Additionally, since the closest related language to Coptic is Greek, we developed our own, reversible dictionary mapping to transcribe Coptic characters into their closely related Greek counterparts. Since some Coptic characters have no Greek equivalent, we transcribed these characters into English letters. Using Greek transcription, we generated the columns "norm_greekified" and "norm_group_greekified" - see Table 1.

Lastly, we removed spaces and punctuation from the norm_group to obtain the "unnormalized" column and applied romanization and greekification to this column as well. Table 1 shows an example row from the final data set.

4. Training And Validation Of Models

4.1 Pretrained Models

Since our approach relies on using transfer learning (see 2.3.1), we used pretrained massively multilingual translation models. For our Coptic-English translator, we used Helsinki-NLP/opus-mt-mul-en [3], and for English-Coptic, we used the Helsinki-NLP/opus-mt-en-mul[4]. Both models use encoder-decoder transformer architectures as described in 2.2.3, trained on translation for 119 different languages.

We separated our data into a training set with 32136 rows, a validation set with 145 rows, and a test set of 254 rows. We chose validation and test sets that were from different sources and translated by different people as shown in Table 5 and 6.

2. https://github.com/isi-nlp/uroman
3. https://huggingface.co/Helsinki-NLP/opus-mt-en-mul
4. https://huggingface.co/Helsinki-NLP/opus-mt-mul-en

4.2 Parameter Tuning

We ran extensive experiments to generate models using multiple techniques on each Coptic column of our dataset to see which one yielded the highest results. Many model parameters were fixed by the multilingual "parent" models, so we tuned the following train parameters:

1. `label_smoothing` $= 0$
2. `num_train_epochs` $= 24$
3. `base_learning_rate` $= 5e^{-5}$

5. Results

5.1 Baseline

Since there are no machine translation systems for Coptic, we implemented a dictionary baseline, using existing Coptic dictionaries to develop a word-by-word translation algorithm. Like all languages, Coptic words and meanings depend on context, so we expected our baseline to perform poorly and with high bias relative to an effective contextual model.

For our Coptic-English baseline, we split the Coptic into each `norm` unit and naively mapped each unit to its dictionary definition.

However, since the English definitions of words for our Coptic dictionary are phrases, we could not use a naive mapping-based approach like our `cop-eng` baseline translator. Instead, we emulated the existing Coptic dictionary[5] functionality by implementing semantic search. We encoded each definition into a vector using the `bert-base-cased`[6] model and normalized the embeddings using softmax. Then, given an English input word, we find the definition whose vector is closest to the input word embedding, and map it to Coptic word associated with that definition.

5.2 Model Results

We evaluated our `cop-eng` and `eng-cop` models on a variety of test texts, listed in Table 6. Each category of test data was translated by different authors with stylistically different translations, leading to significant variation within the results. Even so, the relative performance of all models on the test data was consistent with the model performance on validation data. All contextual models strongly outperform the dictionary baselines, demonstrating that even in an extremely low-resource scenario, NMT models can effectively incorporate contextual clues to improve translation quality.

5.3 `cop-eng` model results

In the `cop-eng` direction, our translator obtains the highest BLEU on the New Testament data. We hypothesize that the New Testament contains the highest quality and most internally consistent data. It is also possible that the decoder of the parent model has already been trained on New Testament sentences and thus provides better translations.

5. https://coptic-dictionary.org
6. https://huggingface.co/bert-base-cased

	Old Testament	New Testament	Zeldes	Budge
dictionary_baseline	0.14, 14.49	0.23, 15.76	0.29, 10.29	0.10, 14.44
norm_romanized	28.25, **54.64**	**61.73, 73.63**	**19.46, 36.37**	**21.71, 43.42**
norm_greekified	26.97, 52.68	48.76, 64.20	15.30, 32.54	20.65, 41.44
norm_group_romanized	26.3, 51.49	43.11, 59.1	17.95, 33.47	18.49, 39.44
norm_group_greekified	**28.30**, 53.48	60.87, 72.19	14.34, 33.19	21.25, 41.48
unnormalized_romanized	23.97, 49.48	44.02, 61.03	14.12, 32.07	16.71, 36.60
unnormalized_greekified	21.69, 46.96	43.64, 60.27	12.06, 30.85	16.00, 35.85

Table 2: BLEU, chrF scores of Coptic-English models on each test set

	Old Testament	New Testament	Zeldes	Budge
dictionary_baseline	0.18, 20.12	0.19, 19.03	0.40, 15.66	0.20, 18.72
norm_romanized	24.77, 56.71	40.19, 66.19	**31.13**, **57.45**	30.24, 57.71
norm_romanized_bt	25.47, 56.72	41.23, **67.40**	21.71, 51.70	**31.29**, **59.13**
norm_greekified	23.24, 44.62	40.05, 58.33	23.83, 39.72	29.09, 48.65
norm_greekified_bt	**25.48**, 46.63	**41.33**, 59.20	27.11, 44.52	30.93, 49.82
norm_group_romanized	8.12, 56.30	14.42, 66.19	5.77, 52.89	9.94, 56.82
norm_group_romanized_bt	7.57, **57.05**	15.85, 66.34	9.35, 50.84	11.64, 57.38
norm_group_greekified	5.36, 45.80	16.12, 57.64	4.88, 43.89	11.22, 48.51
norm_group_greekified_bt	6.78, 44.99	15.95, 58.23	6.01, 44.49	10.10, 49.20

Table 3: BLEU, chrF scores of English-Coptic models on each test set

We consistently observe the best results from the norm_romanized model. The norm models perform best because norming provides a tokenization preprocessing that enforces more semantically meaningful word units. We also hypothesize that the romanized models perform better than the greekified models since the parent model is trained on more Latin alphabet data than Greek.

To qualitatively evaluate the norm_romanized model, we provide an example excerpt (part 3) of the first-ever English translation of the historic Coptic text "Life of Lucius and Longinus", by an unknown author. The translation is coherent and is about the life of a virtuous saint called Longinus who, having raised sick people from the dead, is glorified while visiting a monastery. We provide the translated text in Appendix 8.4.

5.4 eng-cop model results

The backtranslated models (xxx_bt) perform best, consistent with results in the literature (Sennrich et al., 2016). In addition, the norm based models achieve the highest performance, likely due to the tokenization advantage. We do observe a marked decrease in BLEU on norm_group models in comparison to the cop-eng norm_group models. This is however an artifact of the way BLEU pe-

nalizes by word as mentioned in Appendix 8.2.1. Since we have no way of merging `norm` units back to `norm_groups`, the `norm_group` models will be the most useful for providing full translations of English sentences.

6. Ablation Study

6.1 Dialect Differences

The Coptic language contains numerous dialects. Our translation model is trained chiefly on sentences from the Sahidic dialect since it had the most data available, however Bohairic is the primary dialect in use today among the Coptic community (Winand, 2022). To evaluate our model performance on Bohairic Coptic, we manually collected a small parallel dataset of 10 sentences (Isshak). We saw the best performance on our `norm_group_romanized` model with a BLEU score of 2.92, and a chrF score of 24.85. These results are expected given that Bohairic and Sahidic generally use different spellings of the same words and some times entirely different words for the same meaning.

6.2 Colloquial vs. Religious Contexts

Since the translator was trained entirely on a select few religious texts, the model may perform poorly on colloquial sentences. Unfortunately, due to absence of labeled data, we can not directly measure the performance of our models on non-religious sources. Thus, we ran an analysis using heuristic measures to quantify our model performance on out-of-domain data.

We measured performance using an alternative metric: round-trip chrF (Zhuo et al., 2023). In essence, we measure translation competence by translating an input English sentence into Coptic, then translating back to English. We then compare the resulting, "round-trip" sentence with the original. Intuitively, the model will heavily distort sentences which contain unseen vocabulary or style, as observed by 7.77 chrF from gibberish sentences from Table 4. Technical texts and children's books also display moderate distortion. However, round-trip translations on religious texts are distorted much less. In conclusion, we heuristically demonstrate that despite having comparatively strong performance in the religious domain, our model may not yet be appropriate for use in translating technical or vernacular English sentences.

7. Discussion and Conclusion

Our models represent an exciting direction towards using AI in order to advance academic subfields such as history, religious studies, and linguistics. We hope to further inspire research directions in these areas.

7.1 Real-World Application

We recommend the use of our models in real-world contexts, for translating historic, religious Sahidic Coptic texts. However, once deployed, it is important for Coptic researchers to consider the following caveats:

	RTT chrF	Example Sentence	Example Round-Trip Sentence
Random Text	7.77	ihhe rcsnb a targwuw rnhosi zfwnk iegyk lltizbx.	eiereie p thua deie m p rhs shouao eievol tooun. eiereie p thua m p rhs sheieph tooun.
Green Eggs and Ham, Dr. Seuss	33.91	Would you like them in a house?	Thou wishest them in a house.
Coptic Wikipedia	24.66	legal documents and personal letters	and letters, and letters,
Tobit 1-3	46.72	The prayer of both was heard before the glory of the great God.	and they heard the two hymns before the glory of the great God.
Tobit Train Data	56.22	Tobias then answered and said, Father, I will do all things which thou hast commanded me:	Tobias then answered and said, father, I will do all that thou hast commanded me.

Table 4: Analysis of round-trip chrF on texts of various domains (children's books, technical documents, biblical books)

- Our best Coptic-English model, `norm_romanized`, requires preprocessing. Thus, rather than inputting into the model unnormalized Coptic text, it is necessary to tokenize the Coptic data into the `norm` format using the tokenization tools provided by the Coptic Scriptorium, then romanize the Coptic using the `uroman` package as described in Section 3.
- Like many NMT models, our model is subject to hallucination when encountering unknown vocabulary. Relatedly, although our models achieve high accuracy on some domains, they are subject to catastrophic failure outside of this domain or on more difficult sentences, and all outputs should be treated with skepticism.

7.2 Limitations

7.2.1 STYLISTIC CONSISTENCY

Substantial sections of our train set contain translations which differ drastically in style from other sections of our train data. Our models learns to emulate these stylistic differences and thus oftentimes outputs English which is not stylistically consistent. Furthermore, our models tend to output sentences extremely religious in nature, and oftentimes output sentences in the Early Modern English style.

7.2.2 REPETITIVE TRANSLATIONS

Qualitatively, we observe many repetitive translations outputted by the Coptic-English models. For example, the `norm_romanized` Coptic-English model outputs the following translation:

> **My Heart Is Crushed MONB.FL 105-106:** And it shall come to pass in the second month, on the fourth day of the month, on the four and twentieth day of the month, on

> the four and twentieth day of the month, on the four and twentieth day of the month, on the four and twentieth day of the month, ...

Repetitive translations are a common problem faced by low-resource machine translation systems. The problem can be partially alleviated by sampling translations by the best overall rather than greedy by the best token.

7.2.3 MODEL DOMAIN

We have demonstrated in subsection 6 that our model performs much better in historic and religious contexts. Therefore we provide this model as a useful tool for analyzing historic texts, but caution against its use in other contexts.

7.3 Future Research Directions

Low-resource neural machine translation is an active field of research. Our research problem opens unique and unsolved challenges in the following domains:

- extending translation models beyond their domains when monolingual and parallel data are extremely low-resource
- optimizing stylistic consistency in an extremely low-resource setting
- translating historic texts in extinct languages, and evaluating the confidence and quality of the resulting translations

We believe that dictionary-based data augmentation improve out-of-domain generalization of the model. In addition, using part-of-speech tagging provided by the Coptic Scriptorium could also improve the performance of our model. Furthermore, our model could become more consistent if translation data was preprocessed using a language model to convert sentences into more stylistically consistent formats. Lastly, OCR (optical character recognition) is actively being built for Coptic texts. This research holds promise for building a much larger monolingual Coptic dataset, which could lead to greatly improved performance of `eng-cop` models using backtranslation. Thus, as the monolingual Coptic corpus increases, we believe in the value of extending our model on this corpus.

In Section 6, we take for granted that round-trip translation is an appropriate measure of model quality in a particular domain. Although (Zhuo et al., 2023) has supported the use of round-trip translation for model evaluation, further research needs to be done to better understand its ability to delineate stylistic and domain limitations of a model.

References

Chantal Amrhein and Rico Sennrich. On Romanization for model transfer between scripts in neural machine translation. In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2461–2469, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.223. URL https://aclanthology.org/2020.findings-emnlp.223.

Richard Armstrong. Language death. URL https://engines.egr.uh.edu/episode/2723#:~:text=Historically%20speaking%2C%20the%20vast%20majority,languages%20spoken%20in%20the%20world.

Aziz Suryal Atiya. The coptic encyclopedia. https://ccdl.claremont.edu/digital/collection/cce/id/1039, 1991.

Barry Haddow, Rachel Bawden, Antonio Valerio Miceli Barone, Jindřich Helcl, and Alexandra Birch. Survey of low-resource machine translation. *Computational Linguistics*, 48(3):673–732, September 2022. doi: 10.1162/coli_a_00446. URL https://aclanthology.org/2022.cl-3.6.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, nov 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

Nabil Isshak, February 2006. URL http://www.coptic.net/copticweb/contributions/copticlanguagelessons.pdf. A course of lessons in Coptic language.

Bentley Layton. *Coptic in 20 lessons: Introduction to sahidic coptic: With exercises amp; vocabularies*. Peeters, 2007.

Maja Popović. chrF: character n-gram F-score for automatic MT evaluation. In Ondřej Bojar, Rajan Chatterjee, Christian Federmann, Barry Haddow, Chris Hokamp, Matthias Huck, Varvara Logacheva, and Pavel Pecina, editors, *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-3049. URL https://aclanthology.org/W15-3049.

Caroline T. Schroeder and Amir Zeldes, 2013-2023. URL http://copticscriptorium.org. Coptic SCRIPTORIUM.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1009. URL https://aclanthology.org/P16-1009.

Aditya Siddhant, Ankur Bapna, Orhan Firat, Yuan Cao, Mia Xu Chen, Isaac Caswell, and Xavier Garcia. Towards the next 1000 languages in multilingual machine translation: Exploring the synergy between supervised and self-supervised learning, 2022.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. No language left behind: Scaling human-centered machine translation, 2022.

UNESCO, editor. *Atlas of the world's languages in danger*. Memory of Peoples Series. UNESCO, 3 edition, February 2010.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Jean Winand. Ucla encyclopedia of egyptology. *UCLA Encyclopedia of Egyptology*, June 2022. URL https://escholarship.org/uc/item/8tr5w9nc.

Terry Yue Zhuo, Qiongkai Xu, Xuanli He, and Trevor Cohn. Rethinking round-trip translation for machine translation evaluation. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 319–337, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.22. URL https://aclanthology.org/2023.findings-acl.22.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer learning for low-resource neural machine translation. In Jian Su, Kevin Duh, and Xavier Carreras, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, Texas, November 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1163. URL https://aclanthology.org/D16-1163.

8. Appendix

8.1 Neural Network Preliminaries

Deep neural networks are a technique in machine learning, inspired by the human brain, used to approximate arbitrary difficult mathematical functions using a computation graph with learned weights. In essence, they are directed acyclic graphs where vertices can be partitioned into distinct sets $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_k$, often referred to as layers, such that vertices in $\mathcal{L}_i$ can only have vertices in $\mathcal{L}_{i-1}$ as parents. The first layer is referred to as the input layer, the last layer as the output layer while the layers in between are called hidden layers. Each edge between a vertex in layer $\mathcal{L}_i$ and a vertex in $\mathcal{L}_{i+1}$ represents a learnable weight. These weights are learned through a process known as stochastic gradient descent. The weights are intially set to a random guess then a batch of the training data is fed through the neural net to come up with a prediction. This prediction is used to compute the value of a loss function. This function describes how far off the prediction is, and the goal of the neural net is to minimize it's output. As the data traverses the network, gradients are computed, and through a mechanism known as backpropagation, the neural network's weights are adjusted iteratively. Backpropagation calculates the gradient of the loss function with respect to each weight in the network, employing the chain rule from calculus. This gradient signifies the direction and magnitude of the adjustment required for each weight to minimize the error.

8.1.1 TRAIN EPOCHS

Neural networks are trained on a number of train epochs, or iterations of the train data. It is crucial to optimize the number of epochs to increase performance without overfitting on the train data. When trained on too few epochs, the model doesn't yet converge to optimal performance, and when trained on too many, the model overfits on train data. Typically, the optimal number of epochs is found by training the model until validation loss increases.

8.1.2 DROPOUT

Dropout is a regularization technique during neural network training. The basic idea is to "drop" some edges in the network with probability p during train time, enforcing a kind of bagging + L1 regularization on the network parameters. During test time, the network then scales parameter values by $1 - p$ to maintain consistency.

8.2 Evaluation Metric Preliminaries

8.2.1 BLEU

BLEU is a popular metric that measures the similarity between the machine-translated text and one or more reference translations. It operates by comparing n-grams (sequences of n consecutive words) in the machine translation to n-grams in the reference translation. BLEU calculates a score based on the precision of matching n-grams, adjusting for the length of these n-grams to ensure translations don't overly favor longer phrases. Higher BLEU scores, closer to 1, indicate a better alignment between the machine translation and the reference(s). BLEU scores are generally reported out of 100 rather than 1. In practice, sentences between languages can be translated in different ways using different synonyms.

However, BLEU only compares the words between the translated sentence and the reference, potentially resulting in low scores even when the two sentences capture the same meaning. Even so, BLEU has been shown to correlate well with subjective human perception of translation quality, particularly when multiple reference translations are available.

8.2.2 CHRF

Although BLEU works well in languages such as English, some languages, such as Coptic, have different syntactic structure incompatible with BLEU. Coptic is *agglutinative*, meaning that words in the language can be composed of connected strings of arbitrarily long "lemmas", or units of meaning. Researchers at Google Research found that agglutinative languages are unfairly penalized by BLEU (Siddhant et al., 2022). A single incorrect character or lemma within a word results in a completely incorrect n-gram evaluation by BLEU, even if most lemmas within the word are correct.

chrF, a popular character-level adaptation of BLEU, provides more representative scoring translation quality in the agglutinative case (Popović, 2015). The chrF metric operates on a character level, considering character n-grams instead of word n-grams like BLEU. The upshot is that chrF more closely measures similarity between translations and references in agglutinative languages since it appropriately considers the closeness of spelling (and therefore, lemmatization) between words. The metric has been shown to correlate with human perception of translation quality better than BLEU (Popović, 2015). Since the NMT literature often records BLEU within their results, for comparison purposes, we typically provide both BLEU and chrF scores when evaluating translation quality.

8.3 Validation and Test Data Sources

Category	Source
Old Testament	Deuteronomy_1
New Testament	Mark Chapter 1
Budge	Martyrdom of Victor part 8

Table 5: Validation set sources

Category	Source
Old Testament	09_I Samuel_17
New Testament	42_Luke_1
Budge	Letter of pseudo-Ephrem MERC.AT 126-134
Amir Zeldes	Acephalous 22: YA 517-518

Table 6: Test set sources

8.4 Life of Longinus and Lucius Part 3

Using `norm_romanized_cop_eng`, we provide a translation for Part 3 of the previously untranslated Coptic text "Life of Longinus" by an unknown author[7]. The text appears coherent, narrating the tale of a virtuous saint called Apa Logos who flees a monastery after he becomes glorified by the other monks.

> The country of Apa Logos, which he knew well, was a merchant in the province of Lacia, and he used to ask questions in his monastery. And it came to pass that when he had come to Rakote because of the need for his pragmatia, he went to the shrine of Saint Victor, and he prayed, and he also desired to go to the monastery of the blessed Gaius, the elder according to his custom. having great faith in the place where he lived, When he had come, and had received the blessing, he asked him to pray with him. When he was sitting at the door of the assembly, the holy man Gaius related many words of instruction from the Holy Scriptures to the salvation of his soul. And the Pragmatist looked, and saw Apa Mothius, who had raised the dead, and had performed many miracles in the country of Lycia, and he also confirmed that he was not himself, but that he was like unto him. For the hair had grown old on him, and he was in great pain, therefore he knew him not: and a sign had been on his face from his youth, and when the Pharao had looked intently at the sign which was on his face, he said to the holy man Gaius, 'Do you have such a great man in this Monastery that he was filled with strength?'
>
> The prophet said to him, 'Who among the brothers is this that you say about him? The gatekeeper said to him, 'I speak as to this man who knocks at the door.' For this Apa Logos, who lived in the country of Lycia, he and his father, Apa Lucius, who had raised the dead, and had cast out demons from among the people and multitudes of the healed ones, God worked through him in that country. When his fame went out into that land, he was glorified by all. And he forsook his father, and fled secretly to this place, and fled from the vain glory of men. And when he had said these things concerning the holy man Apa Logos, he was in the place of the gate which was sealed, and the governor cried out to him, saying, Why didst thou hide thyself from us, and didst not tell us that thou wast a monk? And straightway the humble man in his heart cast himself down under the feet of the Gaiusite, and repented, saying, 'Forgive me, O my father.' And straightway the Prophet commanded the hair of his head to be shaved, and he prayed over him, and he gave over him the crown and the Devil, according to the garb of Rakote, and when the brethren who were in the monastery heard these words concerning Apa Logos, they set him before them as a great perfect man, and they began to ascribe glory unto him exceedingly. And the holy man Apa reasoningos, when he saw that the men were praising him, grieved greatly. But the younger did not remember the word which his father, the Spirit, gave him, saying, 'Let us flee from the glory of vain men.'
>
> Then he took counsel in himself, saying, I have forsaken my country for the sake of the glory of men, and have departed to my father, and how can I remain in this place, whereas this great multitude has magnified themselves against me? For this reason he came forth

7. https://data.copticscriptorium.org/texts/lifelonginuslucius/life-of-longinus-and-lucius-part-3/norm

from that monastery, where very few God-loving men built for him, or a monastery by the sea. He made himself worthy to work the school, and to suffer for the work of his hands, so that it might be found by those who labor in the work of his hands, which is like the work of the hands of the Apostle Paul, saying, 'These hands have ministered to my necessities and to those who are with me, and that I may find more happiness in giving than in receiving.' Many of the captain of the ship went up to him, having taken his armour, and had made them with his holy hands, and fastened them in the midst of their sails, as far as bows, and believed. He prayed for them, and the Lord delivered their ships into the sea. After a little while the three who were monks came with him, and made disciples of him.

Sitting with his disciples, and meditating on the word of God, and working with his hands, he sat down on his seat in the place where he was working, in the memory of the prophet David, saying, Let me not sleep with mine eyes, I pray thee, with my tears. While he was sitting by night, sleep fell upon him, and he was exhausted: and if a man should stand before him, saying, Rise up quickly, and go to the other side of the sea, and thou shalt find thy father Apa Lucius, and he shall come forth to thee from the land. And he rose up straightway, and did not shew any one of his disciples, and he came to the place which he had been told in the revelation, and he found the holy man Apa Lucius who had come down from the boat immediately after they had saluted each other, and the old man said unto him, 'I did not tell thee that I was going to the Christ, and that I should come to thee, and that the glory of God might bring them to the light, and that the waters might be divided into two parts, and that the wall might be divided into two parts, and that the gates might be For the great and righteous trumpet of the prophets have called these men men Joel, because they are lovers of money, and they persecute the race of serpents like enemies: so also are the holy ones, fleeing from the confusion of cities, dwelling in the wilderness, and they continue in the evil places of the good things of the world at all times, which are the spirits of wickedness.

And they went on, and entered together. And his disciples were sitting and pondering the work, and they saw the holy man Apa reasonos walking with his father, and they marvelled, and they rose up with great joy, and saluted one another, and they prayed and sat down. And Apa reasonings said to them, This is your father from this day forward: then he told them what he had done to them in the vision concerning him. And the holy man Apa Lucius made Apa reasoning men to teach him how to do the work of the instructors, and they persevered zealously in the service of God, and they lived together in great peace, each one keeping watch over his rank in every coast, so that the fame of their city, and their life, was filled with the great city of Rakote and all the country of Egypt.

For a woman, whose little daughter had an issue of blood, having heard of the resurrection of the dead, has been cast forth.' And when she had come to the other side of the sea, she met the holy man Apa Logos, who was lying on the bank of the sea, and she asked him, saying, 'In what place is the servant of God, Apa Logos?' He said, 'I want them to come with me.' And she informed him of the disease which was in it. He said to her, 'What

do you want?' For neither was this woman independent of the man, nor independent of the woman, but, 'My Lord Jesus Christ, be with you.' She believed, and returned to her house. But when she had washed your feet, she went away naked. She went into the city, and told all the things which had happened to her by the angel of the Lord, saying, 'What kind of man is this?' And she shewed them the token of his blessed countenance, and they told her that Apa reasoningos had healed her. And the woman glorified God.

O what man is there that can speak of the healings and miracles which God hath performed? If I wish to speak to each one of them, he will not do so; for the time will come when I will let me speak, but I will declare of them a few of them, and I will declare to thee the completion of the word to this place, and I will spare some of the men of truth because of the multitude of the miracles which they perform. For many came to him out of the city of Rakote, and out of her coasts, desiring to know the place where he was, and brought to him all who were sick with various diseases, and he healed them all.