Exploring Pre-Trained Transformers for Translating Portuguese Text to Brazilian Sign Language

José Mario De Martino
Universidade Estadual de Campinas
Fac. de Eng. Elétrica e de Computação
Depto. de Eng. de Computação e Automação
Avenida Albert Einstein, 400
13083-852 Campinas/SP, Brazil
martino@unicamp.br

Dener Stassun Christinele
ShowCase PRO
Avenida Antônio Artioli, 570
13049-253, Campinas/SP, Brazil
dstassun@showcasepro.com.br

Abstract

The paper focuses on machine translation from Portuguese text to Brazilian Sign Language (Libras) using Transformer-based models. In recent years, the Transformer architecture has established itself as a state-of-the-art approach for machine translation between written languages. To allow the use of the Transformer architecture for translating Portuguese into Brazilian Sign Language, we represent the sign language in a written form with glosses. As Brazilian Signing Language is a low-research language, the effective training of the Transformer model is challenging. The paper presents experimental results exploring transfer learning from pre-trained models of ten different language pairs: Portuguese-Galician, Galician-Portuguese, Portuguese-Catalan, Catalan-Portuguese, Portuguese-Ukrainian, Ukrainian-Portuguese, English-Spanish, English-French, German-Dutch, and German-Ukrainian. After transfer learning and considering the BLEU metric as the evaluation parameter, the experimental results show that the language pairs whose parent models had the biggest training datasets and vocabulary (English-Spanish, English-French, and German-Dutch) displayed the highest performances. The English-Spanish pair, the pair with the biggest training set, achieved the highest performance, followed by the English-French pair, the second biggest training set. The Galician-Portuguese pair, the pair with the smallest training set and vocabulary, presented the fourth-best BLEU score. One possible conjecture to explain this last result is the close relation between the languages.

1 Introduction

Sign Language Translation refers to the process of machine translating between spoken languages and sign languages, and also between sign languages, and presenting the result in a visual form using video or animation. The article focuses on machine translation from a spoken/written language, specifically Portuguese text, to a sign language, namely the Brazilian Sign Language (Libras). Our research tackles the sign translation task in two steps: 1) the machine translation from text to gloss using neural network architectures and 2) the animation of a 3D avatar controlled by the glosses generated by the translation step. In the paper, we present experiments using Transformers to perform the neural translation from text to gloss. The second step of our process is beyond the scope of the paper.

Sign languages are natural languages that convey meaning through manual and non-manual components. The manual elements include features like the configuration of the hand and its orientation and movement. Facial expressions, eye gaze, and upper body movement are examples of non-manual components. The visual-gestural modality of sign languages precludes the direct application of machine translation techniques devised for translating between spoken/written languages. To apply machine learning approaches for translation tasks involving sign language, glossing has been used to represent signs in a textual form and build parallel corpora (Zhu et al., 2023; De Martino et al., 2023; Ananthanarayana et al., 2021; Amin et al., 2021; McCleary et al., 2010). As a common practice, the written language used for glossing is the language of the speaking community in which the deaf community is immersed. For translation, in general, sign language glosses do not describe how signs are produced but are intended to label and encode the meaning of the signs. Typically, a gloss is a set of one or more words written in capital letters that labels a lexical item. In addition to the word(s) in a written language, glosses can be extended with special words and additional textual information in the form of prefixes, suffixes, and symbols such as the at sign (@), colon, and parentheses are used to identify partially-lexical signs like buoys and classifiers

Brazilian Portuguese	Gloss Representation
Dura em média 30 dias.	DURAR MAIS_OU_MENOS TRÊS_ZERO DIAS
Livrei-me de um bicho de pé	ALIVIO ANIMAL.2 PÉ
Leonardo faz homenagem a festeiros de São Benedito.	DAT:LEONARDO FAZER HOMENAGEM FESTA.2 SÃO_BENEDITO
De repente senti um leve toque de dedos em meu ombro.	DEPOIS EU SENTIR CL:TOQUE_OMBRO
Escreva uma palavra que também tenha esse som e compartilhe com a turma.	ESCREVER UM PALAVRA TAMBÉM TER PTF:EFI_CEN(SOM) SOM DEPOIS COMPARTILHAR TURMA

Table 1: Examples of Brazilian Portuguese sentences (translation source) and their respective gloss representations (translation target) .

and non-lexical signs like dactylology (Johnston, 2019, 2008; De Martino et al., 2023; McCleary et al., 2010). In this work, we adopt the glossing scheme described in De Martino et al. (2023) to build our text-to-text parallel corpus. This scheme is exemplified in Table 1 and commented in further detail in Section 3.2.

For visually presenting sequences of glosses representing sign language sentences, in our approach, the articulation of the sign labeled by the gloss is registered with motion capture. The motion capture data drives the animation of our 3D avatar.

Due to its better performance over alternative machine learning models, such as convolutional and recurrent neural networks, the Transformer architecture introduced by Vaswani et al. (2017) has increasingly been used for machine translation. A transformer is a deep learning architecture based on an encoder-decoder model that relies on a parallel multi-head attention mechanism to handle language context dependencies. Currently, Transformer architectures produce state-of-the-art (SOTA) results. However, training SOTA Transformer models is challenging because of the requirement of vast volumes of parallel corpora. The challenge is even greater for low-resource languages, like the Brazilian Sign Language, that lack sufficient parallel corpora for building neural models.

To cope with the lack of data, transfer learning methods have successfully been applied in a diversity of natural language processing tasks. Typically, transfer learning methods reuse pre-trained models on high-resource language datasets to reduce the amount of training data required for low-resource languages (Zhuang et al., 2021; Torrey and Shavlik, 2009; Pan and Yang, 2010).

A relevant question associated with transfer learning concerns the choice of the base model for transfer learning. Seeking to cast some light on this issue, the paper presents experimental results exploring transfer learning from pre-trained models of ten different language pairs: Portuguese-Galician, Galician-Portuguese, Portuguese-Catalan, Catalan-Portuguese, Portuguese-Ukrainian, Ukrainian-Portuguese, English-Spanish, English-French, German-Dutch, and German-Ukrainian. Many research groups, institutions, and companies release models on large datasets that can be used as candidate models for transfer learning. This paper explores transformer models pre-trained and shared by the OPUS-MT project (Tiedemann and Thottingal, 2020). We test and evaluate transfer learning to tune the ten different models for translating from Portuguese text into a Brazilian Sign Language gloss representation.

Adhering to the terminology used in Zoph et al. (2016), we call the pre-trained models the *parent* models, and the models fine-tuned to translate from Portuguese to Brazilian Sign Language glosses the *child* models.

The remainder of this paper is organized as follows: We present an overview of related work in the field in Section 2. In Section 3, we describe our experiments, elaborating on the equipment and methods applied. Section 4 shares the results of our experiments. Finally, Section 5 concludes the paper.

2 Related Work

Machine translation (MT), the automatic translation of text from a source into a target natural language, has experienced major developments in the last decades. In recent years, Neural Machine

Translation (NMT) has established itself as a SOTA technique to overcome the deficiencies of translation strategies of the past, such as Rule-Based Machine Translation (RBMT) (Bhattacharyya, 2015) and Statistical Machine Translation (SMT) (Koehn, 2010). Unlike those strategies, the NMT approach seeks to define and train a neural network that can accommodate wider textual context windows in a flexible way (Bahdanau et al., 2015).

Sign Language Machine Translation (SLMT) cannot directly utilize MT approaches devised for translation between written languages. To overcome this barrier, written representations of sign languages have been tailored by different research groups. Despite its limitation as a linguistic representation (Pizzuto et al., 2006), glossing has been used to build parallel corpora to train machine learning translation approaches (Zhu et al., 2023). Previous research in SLMT-Text2Gloss includes Stoll et al. (2020); Saunders et al. (2020b). Stoll et al. (2020) apply a Recurrent Neural Network for Text2Gloss combined with Motion Graphs to estimate pose sequences. The pose sequences are fed to a Generative Adversarial Neural Network (GAN) to produce videorealistic animations. Saunders et al. (2020b) propose the Progressive Transformer model to translate from discrete text sentences to a skeleton representation of the sign language. Zhu et al. (2023) present experiments to improve the performance of Transformer models via data augmentation, semi-supervised technique, and transfer learning. All three works describe approaches to translate to the Deutsche Gebärdensprache (DGS – the German Sign Language) using the RWTH-PHOENIX14T dataset (Forster et al., 2014). Also, using the PHOENIX14T dataset, Egea Gómez et al. (2022) leverage Transformer models via (1) injecting linguistic features that can guide the learning process towards better translations and (2) applying a Transfer Learning strategy to reuse the knowledge of a pre-trained model. Differently, our experiments focus on Brazilian Sign Language as the target language and Brazilian Portuguese as the source language.

Recent advances in realistic video generation guided by text prompts, such as seen in Ho et al. (2022) may eventually facilitate end-to-end models that perform translation from text to sign languages video without relying in a intermediate representation such as glosses. Some works already demonstrate translation pipelines that don't rely on glosses, such as Saunders et al. (2020a), where spoken language text is first fed to models that generate a sequence of poses which are then passed to a second model that attempts to generate realistic video from those poses.

Please refer to Kahlon and Singh (2023); Núñez-Marcos et al. (2023); Naert et al. (2020) for further surveys related to the main subject of the paper.

3 Materials and Methods

3.1 Parent Models

Ten different parent models were selected for fine-tuning. All chosen models are part of the OPUS-MT (Tiedemann and Thottingal, 2020) repository. Originally trained using MarianMT, a C++ machine translation framework (Junczys-Dowmunt et al., 2018), these models are available as PyTorch models on Hugging Face Hub and could be easily retrieved by code and fine-tuned using Hugging Face Transformers library[1].

Three of the ten models chosen were pre-trained to translate from Portuguese into a target language (Galician – pt-gl, Catalan – pt-ca, Ukrainian – pt-uk). The other three selected models involved the same language pairs but with reversed translation directions. Models involving Portuguese and somewhat related languages (Galician, Catalan) were chosen based on evidence that language relatedness between languages in parent and child models plays a role in transfer learning effectiveness (Dabre et al., 2017; Nguyen and Chiang, 2017; Zoph et al., 2016). The other four chosen models do not include Portuguese as the source or target language in their original task (English-Spanish – en-es, English-French – en-fr, German-Dutch – de-nl, German-Ukrainian – de-uk) and are included for comparison with those pre-trained on a translation task involving Portuguese. Furthermore, these models, with the exception of the German-Ukrainian model, were trained with a much larger dataset than the ones including Portuguese. There is evidence that the size of the training dataset plays a relevant role in the child model performance (Kocmi and Bojar, 2018).

To the best of our knowledge, all parent models were trained with the Tatoeba Challenge (Tiedemann, 2020) training datasets, subversion v2020-07-28[2].

[1] https://huggingface.co/docs/transformers/index

[2] https://github.com/Helsinki-NLP/Tatoeba-Challenge/tree/master/data/subsets/v2020-07-28

Model	BLEU	Vocab.	# Train
opus-mt-en-es	54.9	65001	952526014
opus-mt-en-fr	50.5	59514	180923860
opus-mt-de-nl	52.8	57567	38009174
opus-mt-pt-uk	39.8	62090	2350476
opus-mt-uk-pt	38.1	62090	2350476
opus-mt-de-uk	40.2	62523	1661237
opus-mt-pt-ca	45.7	20554	1164333
opus-mt-ca-pt	44.9	20554	1164333
opus-mt-pt-gl	55.8	5835	541122
opus-mt-gl-pt	57.9	5835	541122

Table 2: Summary of employed parent models with reported BLEU scores on their original test set, vocabulary size, and number of sentences in the original training dataset. Displayed BLEUs were reportedly measured against the Tatoeba Challenge test set for the language pair.

All models share the exact same architecture, with embedding output dimension of 512, 6 encoders, and 6 decoders, each with 8 attention heads and SiLU activation function. Each model has its own vocabulary and SentencePiece[3] pre-trained tokenizers. Further information on these models can be found on Helsinki-NLP Hugging Face Hub page[4].

3.2 Parallel Corpus

The corpus employed for model training and testing is composed of sentences from two elementary school textbooks chosen from the National Program of Books and Teaching Materials, a program of the Brazilian federal government. The translation was carried out sentence by sentence, first registering the translation in a reference video and then annotated with glosses. Along with the gloss translation, each sentence was also recorded with a Vicon Motion Capture System[5] and annotated on Elan[6]. The motion capture data and the Elan annotation are not used in the present work. The translation team was composed of four bilingual members fluent both in Brazilian Portuguese and Brazilian Sign Language and four deaf researchers who are native speakers of Libras.

Glosses were annotated using the formalism described in De Martino et al. (2023). The scheme is an adaptation of the concepts presented by Johnston (2019). In our project, a gloss represents a simplified "translation" of a sign expressed by Brazilian Portuguese words and is uniquely associated with the realization of the sign. The annotation follows the general form [PREFIX:]ID-GLOSS[.n], where elements in the square bracket are optional. The ID-GLOSS element is composed of one or more Brazilian Portuguese words in capital letters. If the ID-GLOSS is formed by several words, they are separated by underscores. The optional numeric value "n" is included in the case of sign variation, that is, if the sign associated with ID-GLOSS can be articulated in more than one manner. The numeric index allows the correct identification of the associated articulation. PREFIX supplements the information expressed by ID-GLOSS. Although other prefixes are specified in the glossing scheme, our dataset, beyond glosses with no prefix, contains only glosses with the prefixes DAT: for dactylology (fingerspelling), CL: for classifiers, and PTF: for pointing signs where a fixed referent is pointed in the signing space.

Examples of the used glossing schema are shown in Table 1. Further details on the glossing scheme can be found in De Martino et al. (2023).

Before use, all Brazilian Portuguese sentences were spelled, checked, and corrected if needed. All Brazilian Sign Language glossed sentences were checked for typos and to see if they were all conforming to the glossing scheme.

Selected Transformer-based models were fine-tuned for Text2Gloss translation using a parallel corpus of 4553 Portuguese - Brazilian Sign Language gloss sentence pairs. 4096 (90%) were used in training, while the remaining 457 were used for testing. When splitting in train/test, the dataset was stratified so that splits contained a balanced number of sentences from each of the two selected textbooks. The glossed sentences contain 5109 unique glosses and a total of 31284 glosses. Out of this total, 1909 (6.1%) glosses accommodate prefixes that convey additional meaning for that gloss (i.e. DAT:, CL:, PTF:)

3.3 Experiments

Experiments were performed using two different pre/post-processing pipelines over the dataset described in Section 3.2.

In the first one, named "lower", glosses are just lower-cased before being passed to the tokenizer. Due to our usage of pre-trained tokenizers from the selected models, passing the glosses in their

[3] https://github.com/google/sentencepiece
[4] https://huggingface.co/Helsinki-NLP
[5] https://www.vicon.com/
[6] https://archive.mpi.nl/tla/elan

Original Glossing	DAT:LEONARDO FAZER HOMENAGEM FESTA.2 SÃO_BENEDITO
Variant "lower"	dat:leonardo fazer homenagem festa.2 são_benedito
Variant "tags"	[DAT_BEG] leonardo [DAT_END] [GLOSS_BEG] fazer [GLOSS_END] [GLOSS_BEG] homenagem [GLOSS_END] [GLOSS_BEG] festa [VAR_2] [GLOSS_END] [GLOSS_BEG] são benedito [GLOSS_END]

Table 3: Example of a glossed sentence, in original form and as it is passed to model on "lower" and "tags" experiments.

original upper-case format would likely negatively impact tokenization and model performance.

In the second one, named "tags", we additionally wrap each gloss inside tags to cue the start/end of the gloss, the gloss prefixes, and the optional information associated with it. After wrapping, glosses are stripped of the special symbols used by the annotation scheme (prefixes, underscore, parenthesis, colon, etc.), as the tags already unambiguously denote what was implied by the original annotation. The employed tags are added as additional tokens on the pre-trained tokenizers so that each tag is tokenized as a single unique token. We enlarge the pre-trained models' token embedding layer input dimension to accommodate the new tokens.

The tagging scheme is an attempt to improve tokenization of glosses. After sentences are tagged, they become a sequence of special tokens (i.e. the tags) and plain text Portuguese words without underscores and other notation-specific characters and constructions that do not occur in parent languages.

In both schemes, when decoding results to compute metrics, we post-process the generated text to revert to the original annotation scheme. Table 3 shows an example of the schemes.

For each of the two pre/post-processing pipelines, we executed 3 fine-tuning runs on each selected parent model. Additionally, each experiment variation was also trained once with randomized weights instead of the pre-trained weights to verify whether knowledge transfer was actually occurring. In total, 80 training runs were executed. Each run was comprised of 6 training epochs with a constant learning rate of 1e-4, batch size 8, adamW optimizer, and cross-entropy loss. The training phase was conducted with the aid of the Hugging Face Transformers, Accelerate, and Tokenizers libraries. We employed an NVIDIA GeForce RTX2080 Ti card to execute training and testing. Each training and testing run took an average of 8 minutes.

3.4 Metrics

We used SacreBLEU v2.2.1 (Post, 2018) library to compute BLEU scores for our test set. When configuring SacreBLEU parameters, we explicitly direct the library not to perform any additional tokenization since glosses should not be additionally broken down (e.g. "DAT:BORGES" would be split to "DAT: BORGES") and skew the metric. All other configurable parameters were left with their standard value provided by the library.

Furthermore, we compute two additional metrics. The first one, called "Vocabulary Score", is the ratio of glosses generated by the model that are present in the training dataset. An ideal Vocabulary Score of "100" means that all glosses generated were previously seen on the training dataset. Since leveraging the parent models' weights meant using their pre-trained SentencePiece tokenizers, we expected our child models to generate glosses that were not originally seen in the training dataset. This effect is troubling because, since glosses are linked to their unique realization in Brazilian Sign Language, we wouldn't want the model to generate glosses that don't necessarily have a realization associated with them. Therefore, we compute this metric to quantify this effect.

The second one, called "Syntax Score," is the ratio of glosses generated by the model that correctly follows the annotation scheme syntax mentioned in Section 3.2. An ideal Syntax Score of "100" means that all glosses generated by the model conform to the annotation scheme. For instance, if the model generated the gloss "CAT:FESTA", the Syntax Score would decrease since "CAT" is not a valid prefix in our notation. In the same manner, if it generated the gloss "GATO_", the Syntax Score would decrease since glosses never end with an underscore. This way, this metric tries to quantify how well the child model is capable of correctly reproducing our glossing scheme.

Experiment	Randomized BLEU	BLEU	Vocab. Score	Syntax Score
en-es-lower	1.32	**24.06**	92.56	99.13
en-es-tags	0.19	22.21	91.10	99.48
en-fr-lower	1.70	22.44	90.76	99.38
en-fr-tags	0.30	22.09	90.36	99.61
de-nl-lower	1.69	21.62	90.29	99.11
de-nl-tags	0.32	20.40	89.23	99.67
pt-uk-lower	1.71	16.79	93.16	98.90
pt-uk-tags	0.09	15.64	94.15	99.38
uk-pt-lower	1.22	16.68	90.83	99.16
uk-pt-tags	0.17	16.13	**94.75**	**99.75**
de-uk-lower	1.48	18.31	87.23	99.43
de-uk-tags	0.14	16.81	86.80	99.60
pt-ca-lower	1.31	19.16	90.81	98.89
pt-ca-tags	0.19	18.33	90.52	99.58
ca-pt-lower	1.10	18.83	90.59	98.92
ca-pt-tags	0.21	19.44	91.68	99.41
pt-gl-lower	1.85	19.76	89.52	98.97
pt-gl-tags	0.21	18.55	88.40	99.68
gl-pt-lower	1.44	19.68	89.16	98.92
gl-pt-tags	0.41	20.50	91.68	99.45

Table 4: Measured Randomized BLEU, BLEU, Vocabulary, and Syntax Score for each experimental setup. Randomized BLEU was obtained in 1 run where parent model weights were discarded before the training procedure. BLEU, Vocabulary, and Syntax Score are mean values for the 3 runs of each setup. The table is ordered by parents' training dataset size (see Table 2) and grouped by language pairs.

4 Results and Discussion

The experimental results are presented in Table 4.

In the cases where the parent models' weights were discarded before the training procedure (Randomized BLEU column), all models performed poorly (below 1.85 BLEU for the "lower" variant and below 0.41 BLEU for the "tags" variant) indicating that the parent model's pre-trained weights were beneficial for the child's translation task.

The best-performing experiment, BLEU-wise, was the "en-es-lower" variant. The English-Spanish parent model was trained with the most sentences on their original translation task, compared to all other parent models. It was trained on 1760 times more sentences than the Portuguese-Galician model, which is the parent pair with fewer training sentences. This way, its superior performance is consistent with findings that report that parent training set size may play a significant role in child model performance, such as seen in Kocmi and Bojar (2018). Nevertheless, language relatedness may also have played a role in the result since Spanish and Portuguese are closely related romance languages. The same may be said of the second-best performing model, trained with the English-French parent.

Between experiments where Portuguese was part of the parent models, the Portuguese-Galician and Galician-Portuguese models achieved the best results in general, with the experiment "gl-pt-tags" achieving the best BLEU among these. Portuguese-Catalan and Catalan-Portuguese models followed closely. Interestingly, Portuguese-Galician was the parent model with fewer sentences in its original training set. Therefore, the model's performance may be related to the fact that, in addition to the presence of Portuguese in the parent pair, the second language of the pair is also closely related to Portuguese. This is consistent with reports of more efficient transfer learning in cases of closely related languages, such as seen in Dabre et al. (2017); Nguyen and Chiang (2017).

In general, experiments using the "tags" scheme had slightly lower BLEU than their "lower" counterparts, except in the "gl-pt" and "ca-pt" experiments, where a small increase in BLEU was observed. Syntax scores for the "tags" variant were, for all models, slightly better than their counterparts. Nevertheless, all experiments resulted in a

Brazilian Portuguese	Nome dado a determinado tipo de história.
Reference Translation	NOME PRÓPRI@ TIPO HISTÓRIA
Best en-es-plain	NOME PRÓPRI@ HISTÓRIA TIPO
Best en-es-tags	NOME PONTO DETALHE TIPO HISTÓRIA
Brazilian Portuguese	Releia o que o Sapo gritou. O que significa o sinal de pontuação !?
Reference Translation	LER NOVAMENTE O_QUE SAPO GRITAR.2 O_QUE SIGNIFICA SINAL DAT:PONTUAÇÃO PONTO_EXCLAMAÇÃO
Best en-es-plain	RELER O_QUE SAPO GRITAR O_QUE SIGNIFICAR SINAL DAT:PONTUAÇÃO
Best en-es-tags	RELER O_QUE SAPO GRITAR O_QUE SIGNIFICAR SINAL DAT:PONTUAÇÃO
Brazilian Portuguese	Assinale a alternativa correta.
Reference Translation	MARCAR RESPOSTA CORRET@
Best en-es-plain	MARCAR RESPOSTA CORRETA
Best en-es-tags	ASSINALAR ALTERNATIVA CERT@
Brazilian Portuguese	Qual é a relação entre essa placa e o quadro?
Reference Translation	PLACA PTF:ESI_CEN(PLACA) QUADRADO OS_DOIS RELAÇÃO O_QUE
Best en-es-plain	QUAL RELAÇÃO PLACA TAMBÉM QUADRO
Best en-es-tags	QUAL RELAÇÃO ENTRE ESSA PLACA TAMBÉM QUADRO

Table 5: Examples of translations produced by the fine-tuned pre-trained English-Spanish model.

Syntax Score of over 98.89, and the improvement brought by the “tags” scheme was marginal and, in this case, possibly not worth the decrease in other metrics. Additionally, inspecting the generated translation, we found translations made by models trained with the "tags" scheme to be more conservative on the generation of glosses containing special annotation prefixes, producing roughly half as much prefixed glosses as their "lower" counterparts over the test set.

In relation to obtained Syntax Scores, results show that the child models successfully learned to reproduce the gloss annotation schema when generating text, regardless of their BLEU scores. Vocabulary Scores show that, in all models, roughly 10% of produced glosses were not previously present on the training set. Although this is not ideal, post-processing pipelines that deal with out-of-vocabulary glosses by removing or replacing them with similar known ones could be sufficient to mitigate this effect.

Some examples of glossed text generated by the best "en-es-plain" and "en-es-tags" models can be seen in Table 5.

5 Conclusion and Future Work

In this work, we presented experiments conducted to explore the possibility of leveraging pre-trained translation models to perform Brazilian Portuguese to glossed Brazilian Sign Language translation. The observed results lead us to believe that the parent model’s previous competence in processing Portuguese is not a necessary factor for reaching relatively good performance in our translation task, seeing that the best-performing model was pre-trained to translate English to Spanish. The English-Spanish parent model was also the model with the most sentences in its original training dataset, with up to 1760 times more sentences than the parent model with the least sentences (Galician-Portuguese). This suggests that the size of the parent’s original training dataset plays a significant role in the child model performance, consistent with what is reported in Kocmi and Bojar (2018). Nevertheless, the fourth best-performing language pair parent, Galician-Portuguese, yielded better results than other models despite having the smallest training dataset among all models. In this case, we believe language relatedness may have played a part and mitigated the effects of the small training set.

Experiments were also conducted utilizing a tag-

ging scheme devised to facilitate glossed text tokenization and also force the model to correctly produce glosses that comply with the annotation scheme syntax. In general, the tagging scheme produced marginal improvements in compliance with the glossing scheme but reduced measured BLEU in most cases.

In our experiments, we repeated a simplistic fine-tuning scheme for all experiments, with a fixed number of epochs and a constant learning rate. It is likely that refining the training procedure with techniques such as learning rate scheduling or early stopping could improve model performance. Data augmentation through back-translation or other techniques could also be employed to tackle data scarcity, such as those described by Zhu et al. (2023). Techniques that would allow us to more efficiently use pre-trained model tokenizers and enable us to increase its vocabulary could also be applied, like seen in Lakew et al. (2018).

If the presented models were used to drive sign language video generation or drive a 3D avatar, further post-processing measures would have to be conceived to deal with out-of-vocabulary or incorrect syntax glosses, which we believe are bound to be generated (even if seldom) in the present case where we leverage pre-trained models and their SentecePiece tokenizers.

We intend to conduct further investigations using a larger Portuguese-Libras dataset in the future. Further expansion of the used corpus is expected, increasing its size and vocabulary variety.

Acknowledgements

This study was partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and grant nº 88887.091672/2014-01, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant nº 458691/2013-5, and Financiadora de Estudos e Projetos (Finep) grant nº 2778/20.

References

Mohamed Amin, Hesahm Hefny, and Ammar Mohammed. 2021. Sign language gloss translation using deep learning models. *International Journal of Advanced Computer Science and Applications*, 12(11).

Tejaswini Ananthanarayana, Priyanshu Srivastava, Akash Chintha, Akhil Santha, Brian Landy, Joseph Panaro, Andre Webster, Nikunj Kotecha, Shagan Sah, Thomastine Sarchet, and Raymond Ptuchaand Ifeoma Nwogu. 2021. Deep learning methods for sign language translation. *ACM Transactions on Accessible Computing*, 14(4):22.1–22.30.

Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations – ICRL 2015*, San Diego, USA.

Pushpak Bhattacharyya. 2015. *Machine Translation*. CRC Press.

Raj Dabre, Tetsuji Nakagawa, and Hideto Kazawa. 2017. An empirical study of language relatedness for transfer learning in neural machine translation. In *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation*, pages 282–286. The National University (Phillippines).

José Mario De Martino, Ivani Rodrigues Silva, Janice Gonçalves Temoteo Marques, Antonielle Cantarelli Martins, Enzo Telles Poeta, Dener Stassun Christinele, and João Pedro Araújo Ferreira Campos. 2023. Neural machine translation from text to sign language. *Universal Access in the Information Society*, pages 1615–5297.

Santiago Egea Gómez, Luis Chiruzzo, Euan McGill, and Horacio Saggion. 2022. Linguistically enhanced text to sign gloss machine translation. In *Natural Language Processing and Information Systems*, pages 172–183, Cham. Springer International Publishing.

Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-weather. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1911–1916, Reykjavik, Iceland. European Language Resources Association (ELRA).

Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. 2022. Imagen video: High definition video generation with diffusion models.

Trevor Johnston. 2008. From archive to corpus: transcription and annotation in the creation of signed language corpora. In *22nd Pacific Asian Conference on Language, Information, and Computation*, pages 16–29.

Trevor Johnston. 2019. Auslan corpus annotation guidelines. Technical report, Macquarie University (Sydney) - La Trobe University (Melbourne).

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch. 2018. Marian: Fast neural machine translation in C++. In *Proceedings of ACL 2018, System Demonstrations*, pages 116–121,

Melbourne, Australia. Association for Computational Linguistics.

Nevroz Kaur Kahlon and Williamjeet Singh. 2023. Machine translation from text to sign language: a systematic review. *Universal Access in the Information Society*, 22:1–35.

Tom Kocmi and Ondřej Bojar. 2018. Trivial transfer learning for low-resource neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 244–252, Brussels, Belgium. Association for Computational Linguistics.

Philipp Koehn. 2010. *Statistical Machine Translation*. Cambridge University Pres.

Surafel M. Lakew, Aliia Erofeeva, Matteo Negri, Marcello Federico, and Marco Turchi. 2018. Transfer learning in multilingual neural machine translation with dynamic vocabulary. In *Proceedings of the 15th International Conference on Spoken Language Translation*, pages 54–61, Brussels. International Conference on Spoken Language Translation.

Leland McCleary, Evani Viotti, and Tarcísio Arantes Leite. 2010. Descrição das línguas sinalizadas: a questão da transcrição dos dados. *Alfa: Revista de Linguística*, 54(1):265–289.

Lucie Naert, Caroline Larboulette, and Sylvie Gibet. 2020. A survey on the animation of signing avatars: From sign representation to utterance synthesis. *Computers & Graphics*, 92:76–98.

Toan Q. Nguyen and David Chiang. 2017. Transfer learning across low-resource, related languages for neural machine translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 296–301, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Adrián Núñez-Marcos, Olatz Perez de Viñaspre, and Gorka Labaka. 2023. A survey on sign language machine translation. *Expert Systems with Applications*, 213:118993.

Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.

Elena Pizzuto, Rossini Paolo, and Russo Tommaso. 2006. Representing signed languages in written form: questions that need to be posed. In *2nd Workshop on the Representation and Processing of Sign Languages "Lexicografic matters and didactic scenarios"*, pages 1–6, Genoa, Italy.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020a. Everybody sign now: Translating spoken language to photo realistic sign language video.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020b. Progressive transformers for end-to-end sign language production. In *16th European Conference on Computer Vision - ECCV 2020*, pages 687–705, Glasgow, UK.

Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden. 2020. Text2sign: Towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, 14:891–908.

Jörg Tiedemann. 2020. The tatoeba translation challenge – realistic data sets for low resource and multilingual MT. In *Proceedings of the Fifth Conference on Machine Translation*, pages 1174–1182, Online. Association for Computational Linguistics.

Jörg Tiedemann and Santhosh Thottingal. 2020. OPUS-MT – building open translation services for the world. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal. European Association for Machine Translation.

Lisa Torrey and Jude Shavlik. 2009. Transfer learning. In E. Soria, J. Martin, R. Magdalena, M. Martinez, and A. Serrano, editors, *Handbook of Research on Machine Learning Applications*. IGI Global.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008. Curran Associates, Inc.

Dele Zhu, Vera Czehmann, and Eleftherios Avramidis. 2023. Neural machine translation methods for translating text to sign language glosses. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12523–12541, Toronto, Canada. Association for Computational Linguistics.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. 2021. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, Texas. Association for Computational Linguistics.