Improving Spanish-Mapudungun Translation through Transfer Learning

Stanford CS224N Custom Project

Eban Ebssa
Department of Computer Science
Stanford University
eebssa@stanford.edu

Abstract

Neural Machine Translation (NMT) systems for low-resource languages like Mapudungun face significant challenges due to limited training data and linguistic complexities. This project aims to improve translation between Spanish and Mapudungun through transfer learning, leveraging pre-trained models on Spanish-English and Spanish-Finnish language pairs. Our contributions include demonstrating the effectiveness of transfer learning in this context and providing a comparative analysis of different parent models. Our main findings show that transfer learning significantly enhances translation performance, with not much of a difference between the Spanish-English and Spanish-Finnish pre-trained model performance. This suggests that factors beyond morphological similarity, such as data quality, play a crucial role in transfer learning success. These insights pave the way for future research into optimizing translation tools for low-resource languages and involving indigenous communities in the development process.

1 Key Information to include

- Mentor: Moussa Doumbouya (TA), Hernán Lira (External)
- External Collaborators (if you have any): N/A
- Sharing project: N/A

2 Introduction

Mapdungun (ISO 639-2 code: `arn`) is an indigenous language spoken in South America, of the Mapuche people originating from Chile and Argentina. As a result of colonization of the region, the Spanish language (ISO 639-1 code: `es`) has dominated systems of communication in South America, at the cost of the decline and extinction of many of these indigenous languages, leading to cultural, historical, and linguistic diversity loss [1]. Today, there are about 1.8 million Mapuche people, but only about 10% of them speak the language [2]. Even as the largest indigenous population in Chile, the Mapuche are not recognized by the law and their language is in danger of extinction [3]. Building a translator between Spanish and Mapudungun might lead to better incorporation of the Mapuche people into Chilean society, serve as an educational tool for the Mapuche, and serve as another resource to help prevent its extinction [4].

Given the scarcity of online parallel text data for Mapudungun, typical of low-resource languages, this project explores techniques tailored for such scenarios. While previous efforts for Mapudungun-Spanish translation have mainly focused on corpus development, statistical machine translation, or data augmentation, [5, 6, 7] the explicit application of transfer learning remains largely unexplored.

Specifically, our two research questions coming in to the project were:

1: Can transfer learning enhance existing neural machine translation systems between Spanish and Mapudungun?

2: Does higher morphological similarity between the parent and child model target languages lead to higher transfer learning performance?

3 Related Work

3.1 Computational Resources for Mapudungun

In 2006, Carnegie Mellon University and the Chilean Ministry of Education forged a collaboration to create the AVENUE project, a resource for Mapudungun-Spanish computational linguistics experiments [5]. This dataset of transcribed real conversations in Mapudungun is publicly available, consisting of approximately 260,000 parallel phrases and sentences in both languages [6].

Mingjund Duan and others (2020) [6] conducted some studies using this dataset, including Spanish-Mapudungun machine translation (in both directions). We use Duan's work as inspiration by drawing insight from one of their final conclusions. The authors suggest that the reason for poor NMT translation in the Spanish to Mapudungun direction compared to the opposite direction is because of the fact that Mapudugnun is a polysynthetic and agglutinative language, in which one word is often made up of many smaller morphemes, possibly making generation more complicated [7, 6]. This leads us to believe that transfer learning using a parent model that can already translate to a higher-resource agglutinative language (i.e. Finnish) might help with the generation.

3.2 Facilitating Low-Resource NMT

Machine translation (MT) converts text from a "source" language into output text in a "target" language, and consists of statistical (SMT), rule-based (RBMT), and neural (NMT) methods. Neural machine translation, which is the focus of this paper and the most current MT approach, uses deep learning to train a model on parallel text data in both the source and target languages.

While NMT systems excel with extensive training data (as in "high-resource" settings), the scarcity of such data for lower-resourced languages results in poorer NMT performance. Even when employing advanced NMT techniques like local attention, standard statistical machine translation (SMT) systems had often outperformed NMT in low-resource scenarios [8]. This disparity underscores the need for novel approaches tailored to the unique challenges of such languages in NMT.

The two main current approaches for facilitating low-resource NMT include exploiting monolingual data and utilizing auxiliary languages [9]. The use of monolingual data includes strategies like back translation, parallel corpus mining, and word replacement through a bilingual dictionary. Auxiliary language strategies include the use of multilingual models, transfer learning, and pivot translation. This paper focuses on the use of transfer learning as an approach for low-resource NMT.

3.3 Transfer Learning Considerations for Mapudungun

Transfer learning is a technique that applies knowledge gained from one task to enhance performance on a related task, thereby reducing the amount of training data needed. While transfer learning had previously demonstrated success in other natural language processing (NLP) tasks, its application to NMT had been previously unexplored. Barret Zoph and others (2016) [8] provide a straightforward yet effective transfer learning framework for NMT. First, train an NMT model, referred to as the "parent" model, on a large parallel text corpus (e.g., French to English, or in one of our cases, Spanish to Finnish). Next, initialize another NMT model, known as the "child" model, by further training the parent model on the small parallel corpus (e.g., Uzbek to English, or in one of our cases, Spanish to Mapudungun). This strategy improves NMT performance despite the parallel text resource constraints.

In addition to choosing a fine-tuning method and vocabulary sharing, choosing an appropriate parent model is one of the design decisions in order to apply transfer learning for low-resource NMT. Most research suggests that selecting a parent model based on language family is important [9]. This means utilizing a parent model trained on a high-resource language within the same family as the target low-resource language. Furthermore, the conclusion from the original paper [8] by Zoph and others that introduced transfer learning for low-resource machine translation is that the more similar the parent model is to the child language, in terms of distributional characteristics or vocabulary, the better the expected performance.

Linguists classify Mapudungun as an Araucarian language, along with other languages like Huilliche spoken in the same region [3]. Unfortunately, the other languages in this family are also low-resource in terms of their online data presence, limiting the common strategy of using a high-resource language from the same family. So, we aimed to explore the grammar aspect of language families as a similarity criterion. Mapudugnun is a polysynthetic agglutinative language, in which words are often composed of many smaller morphemes [7]. For example, according to our dataset, the word "tünkülepürakatun" in Mapudungun means "ahora estoy calmada" in Spanish ("now I am calm" in English), effectively functioning as a whole sentence [10].

Finnish, another agglutinative language, shares these properties and may thus be more related to Mapudungun than other non-agglutinative languages [11]. By choosing Finnish as the parent language, we hoped to exploit these grammatical similarities to improve the translation performance of the Mapudungun-Spanish NMT model.

4 Approach

4.1 Question 1: Exploring the Effect of Transfer Learning

To investigate the effect of transfer learning on translation results in the Spanish to Mapudungun (es→arn) direction, we consider the Spanish to Mapudungun model published by Duan et al. in 2020 [6], which was trained from scratch without transfer learning. We refer to this model as the "es→arn baseline training" model. The diagram below conceptually illustrates the baseline training process for the es→arn direction:

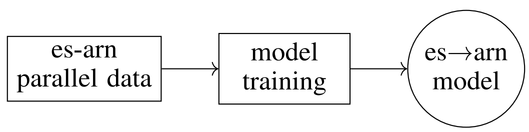


Figure 1: Baseline Training process for es→arn translation.

For comparison, the corresponding model to this baseline is the Spanish to Mapudungun child model we developed by fine-tuning a parent model pre-trained on Spanish to Finnish translation. This model is referred to as the "es→arn transfer learning" condition. Below is a conceptual illustration of the transfer learning process in this direction:

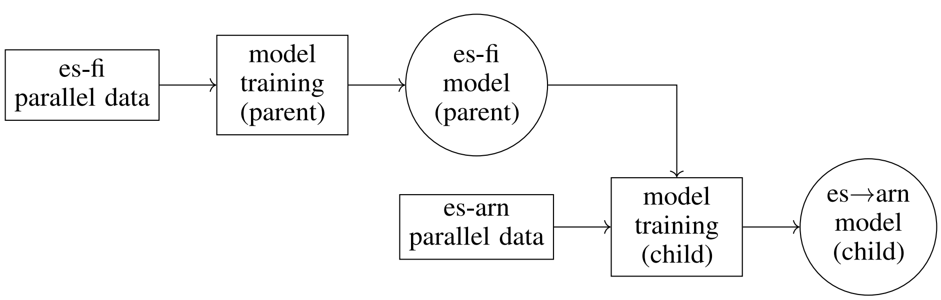


Figure 2: Transfer Learning process for es→arn translation. In this case, the parent model is trained on Spanish-Finnish data, and the child model is the result of fine-tuning parameters of the parent model on Spanish-Mapudungun data. Note that the source languages in both models are the same – the arrow between the parent model and the child training is to demonstrate the transfer of learned knowledge.

Similarly, to explore the effect of transfer learning on translation results in the opposite direction (arn→es), we refer to Duan's Mapudungun to Spanish model as the "arn→es baseline training" model. The corresponding model for comparison is "arn→es transfer learning", the Mapudungun to Spanish child model we developed by fine-tuning a parent model that was pre-trained on Finnish to Spanish translation.

We compare the results of our transfer learning models against the baseline models trained from scratch using the same data sizes and corpus source. The baseline training models, as described in the paper by Duan and others, each utilized a sequence-to-sequence Transformer architecture [12], with 5 encoder and 5 decoder layers, an embedding size of 512, a feed-forward transformation size of 2048, 8 encoder and 8 decoder heads, a 0.4 dropout probability, label smoothing set to 0.1, and the Adam optimizer. The transfer learning child models we developed by fine-tuning pre-trained Spanish to Finnish and Finnish to Spanish models are Opus-MT models developed by the Helsinki NLP group, available on the Huggingface Transformers hub[1] [2], which we treat as the parent models. The transfer learning models are also sequence-to-sequence Transformers, with 6 encoder and 6 decoder layers, an embedding size of 512, a feed-forward transformation size of 2048, 8 encoder and 8 decoder heads, a 0.1 dropout probability, and the Adam optimizer. Due to these slight architectural differences between the baseline training and transfer learning models, we consider our comparison a preliminary one.

4.2 Question 2: Exploring the Effect of Morphological Similarity in the Parent Language

Next, we investigate the effect of grammatical linguistic similarity in the parent model language on the performance of the child model. We explore this by comparing transfer learning for Spanish to Mapudungun translation when fine-tuned from two different parent models: one translating Spanish to Finnish (an agglutinative language like Mapudungun) and the other translating Spanish to English (a mostly non-agglutinative, fusional language). The Spanish to Finnish parent model is the same pre-trained model accessed on Huggingface as the one used in the es→arn transfer learning condition in the first question. The Spanish-English parent model is also Opus-MT models developed by the Helsinki NLP group and available on the Huggingface Transformers hub [3], and it has the same model architecture as the Spanish to Finnish model in the above section.

5 Experiments

5.1 Data

Corpus: We construct a dataset using the Mapudungun-Spanish parallel corpus, sourced from the AVENUE project—a collaboration between Carnegie Mellon University and the Chilean Ministry of Education [5]. This corpus consists of about 260,000 transcribed phrases from Mapudungun conversations and their corresponding Spanish translations [10].

Example Pair (Raw Data):

```
M: fey manchana ka fey ka ti <uh>.
C: la manzana tambièn dicen.
```

Formatting: The files in the corpus were not always uniformly formatted. Typically, each entry consists of a human-generated transcription phrase in Mapudungun, prefaced by "M:", followed by the human-generated Spanish translation prefaced by "C:". However, variations in formatting were present due to different translators' practices. For instance, some entries had both "M" and "C" phrases on the same line. Others had variations in prefixes such as "C." or "C" instead of "C:". To standardize the format, we developed a custom regex expression to consistently extract and pair the phrases.

Example Pair (Extracted):

```
arn_phrase = "fey manchana ka fey ka ti <uh>."
es_phrase = "la manzana tambièn dicen."
```

Meaning: "The apple, too, they say."

[1]https://huggingface.co/Helsinki-NLP/opus-mt-es-fi
[2]https://huggingface.co/Helsinki-NLP/opus-mt-fi-es
[3]https://huggingface.co/Helsinki-NLP/opus-mt-es-en

Text Cleaning: First, we removed the notes taken by the data transcribers. These included notes like "<*SPA>" to indicate that the Mapudungun speaker used the Spanish word instead of the Mapudungun word, or "<Noise>" for when the audio was not clear. Additionally, we removed duplicate punctuation, any remaining unclosed notes, diacritic typos, and capitalization. We also performed normalization, but only experiment 1, in which all Unicode diacritics were also removed as part of text cleaning and replaced with their ASCII counterparts. Spacing was adjusted post-cleaning, and the cleaned dataset was saved in a TSV file with appropriate splits for ease of subsequent loading. Our complete dataset consists 268,265 parallel phrases. However, due to limited computational resources, we use a subset of 10,000 pairs for training and 1,250 pairs each for validation and testing.

Example Pair (Cleaned):

```
arn_phrase = "fey manchana ka fey ka ti."
es_phrase = "la manzana también dicen."
```

Tokenization: We used the extracted cleaned monolingual Mapudungun data to train a new Byte-Pair Encoding (BPE) subword tokenizer [13] implemented via SentencePiece. Subword tokenization is particularly effective for highly agglutinative languages like Mapudungun, where words are typically composed of subwords. New tokens discovered in the Mapudungun tokenizer were added to the shared vocabulary of the respective parent model, which is the common approach for low-resource NMT [9].

5.2 Evaluation method

We use two quantitative evaluation metrics: BLEU and chrF. BLEU, a widely-used metric [14], relies on n-grams, and is computed using the Sacre BLEU toolkit [15]. The chrF metric, on the other hand, measures overlap on the character-level, and so it may be more suitable for evaluating text in languages with complex morphologies, such as the agglutinative Mapudungun [16].

5.3 Experimental details

Effect of Transfer Learning: We fine-tuned both the arn→es transfer learning model and the es→arn transfer learning model from their respective parent model on the Mapudungun-Spanish parallel corpus training subset size of 10,000, and with a learning rate of 3e-4, a weight decay of 0.01, and a batch size of 8. We trained for three epochs each on a single GPU on the Google Cloud Platform, which took about 21 minutes for the es→arn child model and 16 minutes for the arn→es child model. The metric for what the best model to save during training was validation set BLEU score.

Effect of Morphological Similarity in the Parent Language: We fine-tuned the two es→arn models from the Spanish-Finnish pre-trained model and the Spanish-English pre-trained model each on the Mapudungun-Spanish parallel corpus training subset size of 10,000, and with a learning rate of 3e-4, a weight decay of 0.01, and a batch size of 8. We trained for up to 15 epochs each on a single Google Cloud Platform GPU, which took about 1 hour and 45 minutes for the Spanish-English pre-trained model and 1 hour and 53 minutes for the Spanish-Finnish pre-trained model. The metric for what the best model to save during training was validation set BLEU score.

5.4 Results

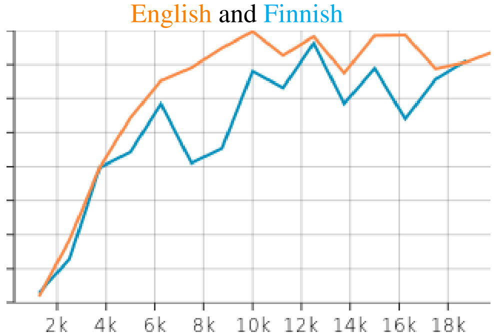
Effect of Transfer Learning:

The following tables report the test BLEU and chrF scores for 10,000 training examples, comparing the Duan baseline model (trained from scratch for up to 200 epochs) with the transfer learning child model (fine-tuned from the Finnish-Spanish parent model for 3 epochs) for translation in the Mapudungun to Spanish (arn→es) direction and the Spanish to Mapudungun (es→arn) direction.

Table 1: Effect of Transfer Learning on arn→es Translation

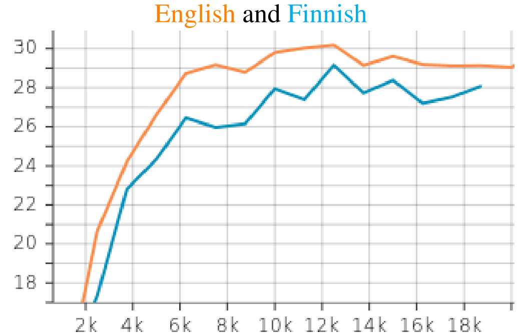
Model	**Test BLEU Score**	**Test chrF Score**
Baseline Training	6.26	10
Transfer Learning	10.03	30.3

Figure 3: Validation BLEU over Time for English and Finnish



Training Step (18,750 total)

Figure 4: Validation chrF over Time for English and Finnish



Training Step (18,750 total)

We observe that, with the same dataset size and original source, the application of transfer learning via a pre-trained parent model leads to higher BLEU (+3.77) and chrF (+20.3) scores compared to baseline training from scratch in the arn→es direction (Table 1). This aligns with our hypothesis, as we expect some aspects of translating into Spanish to be transferable from the Finnish-Spanish parent model to the child task of translating arn→es.

Similarly, in the es→arn direction (Table 2), transfer learning results in significantly higher BLEU (+8.68) and chrF (+21.2) scores. This supports our hypothesis that aspects of translating into Mapudungun are transferable from the Spanish-Finnish parent model to the child task of translating es→arn. This finding also aligns with Duan's study, which attributes the lower baseline BLEU score in the Spanish to Mapudungun direction to the difficulty of generating text in a polysynthetic, agglutinative language with limited training data [6].

Table 2: Effect of Transfer Learning on es→arn Translation

Model	**Test BLEU Score**	**Test chrF Score**
Baseline Training	1.09	10
Transfer Learning	9.77	31.2

Effect of Morphological Similarity in the Parent Language:

Table 3 shows the test BLEU and chrF scores for the es→arn translation task using transfer learning from two different parent models: Spanish-English and Spanish-Finnish.

Table 3: Effect of Parent Morphological Similarity on es→arn Translation

Parent Target	**BLEU Score**	**chrF Score**
English	8.07	0.29
Finnish	8.11	0.28

There does not seem to be a significant difference in BLEU or chrF scores between using the Spanish-English pre-trained model and the Spanish-Finnish pre-trained model as the parent. In fact, the Spanish-English parent model performs slightly better in terms of chrF score.

Figures 3 and 4 show the trajectory of validation BLEU and chrF scores during training (with 1,250 steps equivalent to one epoch). The validation metrics indicate that the Spanish-English pre-trained model might be performing better than the Finnish model, with both models plateauing around 15 epochs.

6 Analysis

In this section, we qualitatively evaluate our system by inspecting key characteristics and outputs of our model. We analyze our model's performance through various methods including example commentary, error analysis, and comparing behaviors across different systems.

6.1 Transfer Learning Evaluation

We observe that the BLEU scores for the models using transfer learning via a pre-trained parent model are higher than their corresponding baseline models in both translation directions. Due to slight architectural differences between the baseline and transfer learning models, our comparison is preliminary. The baseline model and tokenizer developed by Duan et al. are not publicly available. Nevertheless, prior studies support the benefits of transfer learning [8, 9]. Additionally, the baseline paper claims that the lower BLEU score in the Spanish to Mapudungun direction is due to the challenges of generating text in a polysynthetic, agglutinative language with limited training data [6]. We believe there might be a transfer of knowledge when translating into agglutinative languages.

6.2 Tokenization and Vocabulary Analysis

The method of tokenization and the corresponding vocabulary significantly impact translation quality. The text cleaning steps, such as removing diacritics, also play a crucial role. For instance, in experiment 1, removing diacritics from Spanish and Mapudungun texts and replacing them with normalized ASCII characters may have simplified translation leading to slightly higher scores. This is particularly relevant given the inconsistent use of diacritics in the original dataset (e.g., not using the acute "á" in "más").

The importance of the tokenizer vocabulary is evident when comparing the pre-trained tokenizer to a newly trained tokenizer on Mapudungun text. For example, consider the Mapudungun sentence "Tünkülepürakatun" ("Ahora estoy más calmada" in Spanish, "Now I am more calm" in English). The pre-trained Spanish-Finnish tokenizer produces the tokens `["_T", "ü", "nk", "ü", "le", "p", "ü", "ra", "katu", "n"]`, which include many single-character tokens. In contrast, the new tokenizer trained on monolingual Mapudungun data generates `["_T", "ün", "küle", "püra", "katun"]`, capturing more meaningful morphemes. According to an online Spanish-Mapudungun dictionary [4], "püra" approximates "más" ("more"), and "küle" might relate to an emotional state.

6.3 Error Analysis

The pretrained Spanish to Finnish model's performance was unexpectedly lower compared to the Spanish to English model. For instance, consider the following test dataset example translation:

Spanish Source Text: `Toda la gente lo ve.` *English: All the people see it.*

Mapudungun Target Translation: `Kom che pey.`

Predicted Target (Spanish-English Model): `Kom che pe ngeafuy.`

Predicted Target (Spanish-Finnish Model): `Che pe pape feyu.`

Using online Mapudungun resources [5] [6] and patterns from the parallel corpus, we infer that "Kom" translates to "Toda" (meaning "all") and "che" to "la gente" (meaning "the people"). The word "pey" is a third-person singular intransitive form of "pen" (meaning "to see"). The Spanish-English model uses the first-person singular intransitive form "pen" with the suffix "geafuy," which might relate to an impersonal pronoun or passive voice, as in "pigeafuy" ("it is said"). The Spanish-Finnish model may have produced the "pepapeyu" which we infer relates to "coming to see [something or someone]" ("venir" is "to come" and "ver" is "to see").

[4]https://www.conadi.gob.cl/storage/docs/Diccionario_mapudungun.pdf

[5]https://www.conadi.gob.cl/storage/docs/Diccionario_mapudungun.pdf

[6]https://en.wikipedia.org/wiki/Mapuche_language

6.4 Qualitative Comparison

In some cases, the source Spanish text and target Mapudungun translation are similar due to the presence of Spanish words in Mapudungun utterances. For example:

Spanish Source Text: `y así nunca más` *English: and so never again*

Mapudungun Target Translation: `y femuechi nunca ma`

Predicted Target (English Parent): `fey tu rpu müten`

Predicted Target (Finnish Parent): `y asi nunca no ma`

In this instance, the child model from the Finnish parent produced a qualitatively similar translation to the reference compared to the English parent model.

7 Conclusion

Transfer learning on a parent model seems to improve on non-transfer learning for Spanish-Mapudungun translation, even with very limited epochs. There does not seem to be much of a difference in BLEU or chrF scores between English and Finnish as target languages for the parent model. If anything, the Spanish-English model seems to be a better parent. Primary limitations were the difference between the model architecture of the baseline training and transfer learning models in experiment 1, and the difference in tokenization and vocabulary of the two trials in the Finnish vs. English experiment. Perhaps it is the similarity in vocabulary and tokenization that has a bigger influence than morphological similarity in grammar. Future work would investigate the effect of tokenization method and extend training epochs. Future work might also compare transfer learning results using parent models trained on a variety of languages, including those with different scripts and morphological structures, such as Arabic, which is also agglutinative but uses a different script.

8 Ethics Statement

One of the ethical challenges in developing a Neural Machine Translation (NMT) system between Spanish and Mapudungun is the risk of misrepresentation. As the developer is not a native speaker of either language, there is a possibility of inaccuracies in translation that may not be evident through quantitative metrics like BLEU scores alone. These metrics do not always capture semantic accuracy, leading to potential mistranslations or cultural misrepresentations. To mitigate this risk, the project includes a handover of the model and related work to Inria Chile, an organization based in Santiago, which will engage native speakers in the evaluation and further development process. This ensures that qualitative assessments are incorporated and that the translations are culturally and linguistically accurate.

Another significant ethical concern is the assumption that a digital translation tool is desired by the Mapuche people. Mapudungun, traditionally an oral language, only adopted the Latin alphabet post-colonial contact [4]. Imposing a digital translation tool without proper consultation could undermine the linguistic sovereignty and cultural heritage of the Mapuche community. To address this, we propose setting up recurring focus groups with Mapuche people to involve them actively in the development process. This approach will ensure that their needs and perspectives are respected and that they are treated as collaborators rather than passive subjects [17]. By engaging with the community, we can ensure that the tool aligns with their linguistic and cultural values and supports their goals for language preservation and revitalization.

References

[1] Javiera Moya-Santiagos, Paulina; Quiroga-Curín. An overview of indigenous peoples in chile and their struggle to revitalise their native languages: The case of Mapudungun. In *London Review of Education*, Santiago, Chile, May 2022. Association for Computational Linguistics.

[2] Begoña Pendas, Andres Carvallo, and Carlos Aspillaga. Neural machine translation through active learning on low-resource languages: The case of Spanish to Mapudungun. In Manuel Mager, Abteen Ebrahimi, Arturo Oncevay, Enora Rice, Shruti Rijhwani, Alexis Palmer, and

Katharina Kann, editors, *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 6–11, Toronto, Canada, July 2023. Association for Computational Linguistics.

[3] Scott Sadowsky, Héctor Painequeo, Gastón Salamanca, and Heriberto Avelino. Mapudungun. *Journal of the International Phonetic Association*, 43(1):87–96, 2013.

[4] Elizabeth Jacqueline Quintrileo Llancao. Technology's role in mapudungun language teaching and revitalization. *AlterNative: An International Journal of Indigenous Peoples*, 15(3):205–216, 2019.

[5] Lorraine Levin, Rodolfo M Vega, Jaime G. Carbonell, Ralf D Brown, Alon Lavie, Eliseo Canulef, and Carolina Huenchullan. Transfer learning for low-resource neural machine translation. In *Language Resources and Evaluation (LREC 2002)*, 2000.

[6] Mingjun Duan, Carlos Fasola, Sai Krishna Rallabandi, Rodolfo Vega, Antonios Anastasopoulos, Lori Levin, and Alan W Black. A resource for computational experiments on mapudungun. In *Association for Computational Linguistics (ACL)*, 2020.

[7] Christian Monson, Ariadna Font Llitjós, Roberto Aranovich, Lori Levin, Ralf Brown, Eric Peterson, Jaime Carbonell, and Alon Lavie. Building nlp systems for two resource-scarce indigenous languages: Mapudungun and quechua. *Strategies for developing machine translation for minority languages*, page 15, 2006.

[8] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer learning for low-resource neural machine translation. In *Association for Computational Linguistics (ACL)*, 2016.

[9] Rui Wang, Xu Tan, Renqian Luo, Tao Qin, and Tie-Yan Liu. A survey on low-resource neural machine translation. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4636–4643. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Survey Track.

[10] Flor Caniupil; Héctor Painequeo; Rosendo Huisca; Hugo Carrasco; Rodolfo M Vega; Lori Levin; Jaime Carbonell Luis Caniupil. Mapudungun speech corpus.

[11] Dr. Alfréd Tóth. Are all agglutinative languages related to one another? 2007.

[12] 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.

[13] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Association for Computational Linguistics (ACL)*, 2016.

[14] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, Philadelphia, PA, 2002. Association for Computational Linguistics (ACL).

[15] Matt Post. A call for clarity in reporting bleu scores. In *Proceedings of the Third Conference on Machine Translation (WMT18)*, page 186–191, Brussels, Belgium, 2018. Association for Computational Linguistics.

[16] Maja Popović. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, page 392–395, Lisbon, Portugal, 2015. Association for Computational Linguistics.

[17] Manuel Mager, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. Ethical considerations for machine translation of indigenous languages: Giving a voice to the speakers, 2023.