UCSP Submission to the AmericasNLP 2025 Shared Task

Jorge Asillo Congora **Julio Santisteban** **Ricardo Lazo Vasquez**
Department of Computer Science, Universidad Católica San Pablo
Arequipa - Peru
{jorge.asillo, jsantisteban, ricardo.lazo}@ucsp.edu.pe

Abstract

Quechua is a low-resource language spoken by more than 7 million people in South America. While Quechua is primarily an oral language, several orthographic standards do exist. There is no universally adopted writing standard for Quechua, and variations exist across dialects and regions; its current writing is based on how it is uttered and how the sound is written. Quechua is a family of languages with similarities among the seven variants. The lack of a parallel dataset has reduced the opportunities for developing machine translation. We investigated whether increasing the current Quechua Parallel dataset with synthetic sentences and using a pre-trained large language model improves the performance of a Quechua machine translation. A Large language model has been used to generate synthetic sentences to extend the current parallel dataset. We use the mt5 model to fine-tune it to develop a machine translation for Quechua to Spanish and vice versa. Our survey identified the gaps in the state of the art of Quechua machine translation, and our BLEU/Chrf++ results show an improvement over the state of the art.

1 Introduction

In this paper we present the submission of the Universidad Católica San Pablo to the Workshop on Natural Language Processing (NLP) for Indigenous Languages of the Americas (AmericasNLP) 2025 Shared Task on machine translation systems for Indigenous languages. We participated in two directions: Spanish to Quechua and Quechua to Spanish.

Quechua is an indigenous language from the south of Peru that has expanded to Bolivia, Chile, and Ecuador. It is an indigenous language family with 7 variations and almost 8 to 10 million speakers. Quechua is actively used in Peru and Bolivia and is the official language of the Peruvian, Bolivian, and Ecuadorian governments.

Quechua is a phonetic language where each letter represents a specific sound. Quechua is well-studied linguistically and does have defined grammatical rules. Each Quechua dialect has its own semantics and vocabulary. Quechua is an agglutinative language where a prefix or suffix is added to the root of a word to create a new word with a different meaning. Quechua writing is as it sounds and according to the utterance and listener.

A parallel dataset restricts machine translation (MT). In the case of Quechua, the most used resource is the JW300 (Agić and Vulić, 2019), which presents 2 Quechua variants: Ayacucho Quechua (quy), Cuzco Quechua (quz), and the Bolivian variety of Quechua (que). There are also scarce resources with few parallel sentences.

There is a clear need to develop a machine translation and other tools to support Quechua speakers, and current proposals do not achieve an appropriate machine translation. The current research and development of a Quechua MT lacks of an appropriate parallel dataset, making it more challenging to develop an Quechua MT.

The AmericasNLP Shared Task on Machine Translation into Indigenous Languages has been promoting the research of 11 indigenous languages, including Quechua, from 2021 to 2024. The AmericasNLP Shared Task is a competition for research on machine translation. The AmericasNLP Shared Task is based mostly on the Quechua Ayacucho (quy) variant. The Shared Task is framed on a given dataset and open resources, including pre-trained models. The focus has been to translate Quechua–Spanish; to our knowledge, no other research has translated English into Quechua and Quechua into English.

The benchmark for a MT of Spanish (es) to Quechua (quy) has been set on The AmericansNLP 2024 as follows: chrF of 28.81 developed by Helsinki (Vázquez et al., 2021) and ChrF of 34.01 developed by Sheffield (Gow-Smith and Villegas,

Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP), pages 84–91
May 4, 2025 ©2025 Association for Computational Linguistics

2023) for the test set and 28.78, 30.22 respectively for the development set.

We aim to identify if extending the JW300 (Agić and Vulić, 2019) Parallel dataset by generating synthetic sentences in English would improve the machine translation performance. In addition, we want to identify if using a Large Language Model would improve the machine translation performance.

The following sections present a review of the state of the art, our method, and our results and Conclusion section.

2 Related Work

2.1 Early antecedents

Rios (2015) developed a hybrid machine translation for Spanish to Cuzco Quechua. The MT is a classical rule-based supported by statistical modules. Rios also developed a Quechua text normalisation to rewrite Quechua texts in different orthographies or dialects to standard orthography. Rios developed a Quechua dependency treebank and spell checker. Achieving a BLEU score of 57.98 for words and 63.13 for morphemes. Rios' work includes the use of verb morphology (Rios and Göhring, 2013) and rule-based(Rios and Göhring, 2016).

The AVENUE project at the Language Technologies Institute (Llitjós, 2005) had developed an MT which would be used to translate Quechua if a Parallel dataset exist. The AVENUE is a statistical machine translation. One extension of AVENUE had developed a Quechua Parallel dataset, which reached 1,700 sentences. As a result, a Quechua Morphology Analyzer to assist the MT was developed by Llitjós et al. (2005).

Vilca also developed a morphological analyzer (Vilca et al., 2009), Huarcaya Taquiri (2020) developed the first transformer model for an MT Spanish to Quechua Chanka with an outstanding BLEU score of 39.5 using the JW300 Parallel dataset (Agic and Vulic, 2019). Quechua Chanka is also know as Quechua Ayacucho (quy).

2.2 Quechua's resources

There are few resources of a Parallel dataset of Quechua, and the following parallel dataset is well established: the most used is the JW300 (Agic and Vulic, 2019), which presents 3 Quechua variants: Ayacucho Quechua (quy), Cuzco Quechua (quz).

The following parallel dataset are small repositories in which the validity of the Quechua variant is not clear: Sentences extracted from the official dictionary of the Minister of Education (MINEDU)(AmericasNLP, 2021), Huarcaya(Moreno, 2021), Oncevay(Arturo and Diego, 2021), the Peruvian(Congreso de la República del Perú, 2008) and Bolivian(Ministerio de la Presidencia de Bolivia, 2012), constitutions (Tiedemann, 2012), Wikipedia crawls(Tiedemann, 2020) and The JHU Bible parallel dataset (McCarthy et al., 2020).

Well-know Quechua dictionaries, Quechua Spanish and Spanish Quechua produced by Calvo Pérez (2007), Calvo works for the recognition and normalization of the Quechua language and its harmonization with the Spanish language. Calvos's dictionary holds 51233 Quechua and 74395 Spanish words. The website Runasimi.de (2006) provides a dictionary of several Quechua variants to German, English, Spanish, Italian and French.

2.3 State of the art of Quechua machine translation

Table 1 shows the best MT score for es->quy held by BSC (Garcia Gilabert et al., 2024) in the AmericasNLP 2024 Shared Task. For quy->es the score is held by Chen and Fazio (2021) focusing on a morphologically guided segmentation.

The state of the art concerning Quechua machine translation has its own limitations. The pertinent literature does not show a clear development and presents outlier results that are not viable to achieve like Huarcaya Taquiri (2020) reports a 39.50 BLEU score in the JW300 dataset(Agić and Vulić, 2019). Similarly, Ebrahimi and et. al. (2022) report 68.00 BLEU score for en -> quy using the same dataset. There are two logical conclusions: the results are inconclusive or use an incorrect interpretation of the BLEU score.

The BSC team (Garcia Gilabert et al., 2024) achieved the highest performance in the Quechua language. Their approach focused on fine-tuning the NLLB-200 for Quechua and Guarani, inparallel datasetting data from multiple sources and applying a rigorous cleaning process. They experimented with two model sizes, 3.3B and 1.3B, finding that the larger model only improved Quechua results. In particular, fine-tuning NLLB 1.3B with LoRA yielded a new benchmark score of 38.21 ChrF++ for Quechua, the highest among all submissions.

Other teams also contributed innovative approaches to the AmericasNLP 2024 Shared Task. The NordicAlps team (Attieh et al., 2024), based on

Author	BLEU	ChrF	Direction of Translation
AmericasNLP 2024 BSC (Task)	4.85	38.21	es ->quy
AmericasNLP 2024 BSC (NLLB-3.3B)	4.07	36.39	es ->quy
AmericasNLP 2024 Baseline dev.	-	30.22	es ->quy
AmericasNLP 2024 Baseline test	-	34.01	es ->quy
Gow-Smith and Villegas (2023)	4.61	39.52	es ->quy
Vázquez et al. (2021)	5.38	39.40	es ->quy
NLLB Team et al. (2022) 1.3B parameter	-	29.2	es ->quy
Thesis: (Huarcaya Taquiri, 2020)	39.50	0.24	es ->quy
Ebrahimi and et. al. (2022) Baseline	1.58	0.33	es ->quy
Ebrahimi and et. al. (2022) XLM-R Large +MLM	68.00	-	es ->quy
Chen and Fazio (2021)	23.70	-	quz ->es
Ortega et al. (2020) Morfessor	20.30	-	qu ->es
Ortega et al. (2020) BPE-Sennrich	22.90	-	qu ->es
Oncevay (2021) Pairwise	8.20	30.90	quy ->es
Oncevay (2021) Multiling.	4.23	37.80	es ->quy
Ortega et al. (2021) es,qu,fi	22.60	-	quz ->es
Ortega et al. (2021) es,qu,fi,cni	17.00	-	quz ->es
Ortega et al. (2021) es,qu,cni	20.10	-	quz ->es

Table 1: State of the art of Quechua machine translation

the Helsinki system (De Gibert et al., 2023), used various tokenization strategies, with their BPE-MR model ranking first in five languages. The DC_DMV team (Degenaro and Lupicki, 2024) worked with two approaches using the NLLB-200 and the Mamba-based model, obtaining the second-best result for Quechua with the NLLB model. Meanwhile, the University of Edinburgh (Iyer et al., 2024) fine-tuned Llama-2 7B, Mistral 7B, and MaLA-500 using LoRA but did not achieve outstanding performance.

Due to the nature of the Quechua and its lack of writing rules, there are attempts to use morphological tools to normalise the Quechua (Ebrahimi and et. al., 2022) (Chen and Fazio, 2021) (Ortega et al., 2020) (Ortega et al., 2021); prefixes and suffixes are used to normalise (Ortega et al., 2020), and text normalization to keep under control the text pass to a Neural Network (Vázquez et al., 2021). There are interesting approaches, but those rules are like if someone is building the grammar and syntaxes of the Quechua. Reported results range from 17 to 24 BLEU scores; most proposals do not use the ChrF, which might help corroborate the results. Some proposals use variations of the JW300 (Agić and Vulić, 2019) and in most cases, the dataset used is small and domain-constrained.

There are clear limitations to the development of Quechua machine translation. The first is the variety of Quechua dialects or variations. The second is the lack of writing rules, which causes the same pronounced word to be written differently. The last limitation is the lack of a Parallel dataset; all research is based on the JW300 parallel dataset, and no efforts are made to develop a new dataset even though there are 11 million Quechua speakers. Most of the testing is based on Opus biblical, a Peruvian magazine article, testing in a close domain. (Mager and et. al, 2021).

The present work tries to develop a machine translation based and extending the Parallel dataset with syntactic sentences. Tens of Indigenous languages exist in Western South America, some of which are in the process of extinction, and others have disappeared. We aim to preserve the Quechua and make it available to Quechua speakers.

3 Method

3.1 Data sources

Our sources of parallel dataset are shown in Table 5. Most of the data are based on the JW300 parallel dataset (Agić and Vulić, 2019). (Calvo Pérez, 2007) is a dictionary, and the sentences have been extracted almost manually. All our data has been cleaned up by removing irrelevant text, extracting only sentences in lowercase, and keeping only characters a-z and ñ. Data has been shuffle, and we reserve 85% for training and 15% for testing. The

JW300 was only used for que <-> en MT.

3.2 Parallel dataset Expansion

parallel dataset expansion is primarily based on the generation of synthetic sentences. This method consists of taking a sentence from a high-resource language such as Spanish or English, applying a **POS** and replacing words in the original sentence. We will use two approaches: Wordnet and based on LLM.

Based on WordNet, each sentence will be scanned to identify the parts of the speech. The subject and verb of the sentence will be selected. Using WordNet, similar words will be identified based on the four types of similarity defined by WordNet: synonyms, similar, hypernyms, and hyponyms. The new words, subject and verb, will be identified. Synthetic sentences are generated by combining the new words. The combination will be progressive, changing one word, then two, and then three. Several subsets of synthetic sentences are generated depending on the degree of combinatorics.

Based on LLM, each sentence will be parsed (POS) to identify the parts of speech. The subject and/or verb of the sentence will be selected. Using an LLM, the word (subject or verb) will be replaced with another semantically similar word in the context of the sentence. The answer sentence within the LLM answer will be extracted (clean the answer).

MT like mt5-small are sensitive to the direction of the translation. Asymmetric model supports this assumption (Santisteban and Tejada-Cárcamo, 2015). We will train the model in both directions.

The objective is to evaluate the machine translation for Quechua based on the expanded parallel dataset. Two *transformer* models will be used, the base Transformer model by (Vaswani et al., 2023) and a pre-trained multilingual MT5-small (Xue et al., 2021).

3.3 Generation of synthetics sentences

Two different approaches were used for synthetic sentence generation. Initially, an English dataset was processed using WordNet, where part-of-speech (POS) tagging identified the first noun. This noun was then replaced using WordNet and Phi-3 (Abdin et al., 2024), resulting in two synthetic sentences. For example, given the sentence "pay attention to how you listen", the POS tagging selected the word "pay". The synthetic sentence generated with WordNet was "wage attention to how you listen", while Phi-3 produced "focus on how you listen". The prompt used is as follows: ”Replace '$word$' in '$sentence$' with another word while maintaining the semantic meaning”.

Quechua	Original	Clean	Synthetic
que	135,068	131,430	*
quy	114,408	111,655	111655
quz	128,252	125,341	121,480

Table 2: English synthetic sentences generated

In the second approach, a Spanish dataset was used without prior POS tagging. Instead, Phi-3.5 (Abdin et al., 2024) was prompted to replace either a verb or a noun in the given sentence while preserving its semantic meaning. For instance, starting with "aproveche momentos en que estén relajados.", Phi-3.5 generated "aproveche momentos de calma" This adjustment improved the quality of the generated sentences while maintaining coherence. The prompt used is as follows: ”Reemplaza un '$sustantivo$' o '$verbo$' por otro semanticamente similar en la oracion: "{oracion}". dame la primera oracion alternativa. respuesta corta. sin explicacion”.

4 Tests and Results

4.1 English-Quechua

For en->qu and vice versa, we only used the JW300 parallel dataset in English (Agić and Vulić, 2019). We used Phi3-mini due to its compact size and average performance compared to other larger models.

4.1.1 Synthetic Generation Results

Generation with Wordnet lacks of quality. It is unable to find a suitable synonym; it also fails to take the word's context into account, rendering the new sentence meaningless. The evaluation was empirical, based on a review of sentences.

A more satisfactory result was obtained regarding the synthetic sentences generated with Phi3-mini. It takes the word's context into account and can replace the verb, connectives, etc., associated with some nouns, resulting in synthetic sentences with better semantic meaning. Some sentences did not generate any results due to the absence of a noun in the sentence. The original parallel dataset increases with the synthetic sentences by 96%, almost doubling the size of the original parallel dataset.

Dataset	Sense	BLEU	ChrF
JW300	en quy	3.64	32.92
	quy en	5.70	23.43
	en quz	3.82	31.03
	quz en	5.49	22.54
JW300 Clean	en quy	2.68	33.87
	quy en	4.98	23.60
	en quz	3.17	31.70
	quz en	5.42	23.76
JW300 Extended	en quy	*	*
	quy en	5.22	23.35
	en quz	5.67	29.24
	quz en	3.08	31.51

Table 3: MT5-small trained in English

	Sense	BLEU	ChrF
JW300 Clean	en quz	1.89	28.88
	en quy	1.92	29.40
JW300 Expanded	en quz	1.83	28.46
	en quy	1.83	28.52

Table 4: Basic Transformers with synthetic data

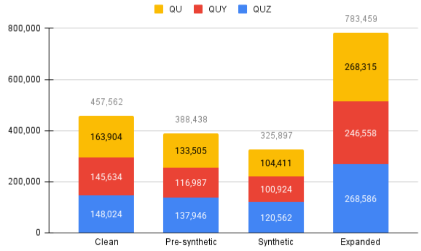

Figure 1: Numbers of synthetic sentences generated in Spanish from the original Spanish parallel dataset

4.1.2 Training the Transformer Models

Two models were used for training: the basic (untrained) transformer model by (Vaswani et al., 2023) and the MT5-small model by (Xue et al., 2021), which is a large, pretrained multilingual text-to-text transformer.

For the choice of tokenizers in the case of the MT5-small transformer, the model was trained using a word tokenizer for both the source and target languages. Retraining the model requires using the same tokenizers. In the case of the base transformer, since this model is trained from scratch, we chose a word tokenizer for English and a BPE tokenizer for Quechua.

Hyperparameters for MT5-small are as follows: batch size 8, learning rate 2e-5, seq_len 512, epoches 30, d_{model} 512. For base Transformer are batch size 32, learning rate 1e-4, seq_len 128, epoches 30, d_{model} 512.

4.1.3 Transformer Model Training Results

The fine-tuning of the MT5-small was tested as shown in Table 3 and 4. For the base transformer, we can see the output of both the model trained with the original parallel dataset and the model trained with the expanded parallel dataset.

Table 3 shows The training of MT5-small with different datasets. JW300 is the basic one (no data processing). JW300 Clean, without punctuation marks, verses, and others. JW300 Extended, the clean parallel dataset plus the synthetic parallel dataset. Trained in both directions, from the source language to the target language. BLEU (sacreBLEU) and ChrF metrics. Using two Quechua languages: Ayacucho Quechua (**quy**) and Cuzco Quechua (**quz**), and English (en). Synthetic parallel dataset generated with Phi3-mini.

As show in 4 the training the Base Transformer with different datasets. JW300 Clean, without punctuation marks, verses, and others. JW300 Extended, the clean parallel dataset plus the synthetic parallel dataset. Trained in both directions, from the source language to the target language. BLEU (sacreBLEU) and CharF metrics. Using Cuzco Quechua (**quz**) and Ayacucho Quechua (**quy**), and English (en). Synthetic parallel dataset generated with Phi3-mini.

The base Transformer and the MT5-small obtained lower scores in both metrics when training with the expanded parallel dataset than with the original. This drop in metrics may indicate that the generated synthetic sentences are not of good quality.

4.2 Spanish-Quechua

The Quechua-Spanish parallel dataset is from 7 sources. Those that could not be identified by the Quechua used were marked as Southern Quechua. A total of 457,562 entries were obtained for the new original parallel dataset, divided into three groups, “quz", “quy", and “qu", as shown in figure 1

4.2.1 Training the Transformer Models

The base transformer model (Vaswani et al., 2023) and the MT5-small (Xue et al., 2021) were used, with the same hyperparameters and tokenizers as in the english-Quechua phase. In the case of the MT5-small, the model was fine-tuned using a word

Author	File	Quechua	quantity
REPU-CS-2021	Constitution (REPU-CS-2021)	quz	812
	Handbook	quy	2,297
	Lexicon	quy	6,154
	Regulation	quz	217
	Webmics	quy	980
Portocarrero	Emotion analysis	-	1,722
AmericasNLP 2024	Dict_misc	quy	8,955
	Minedu	quy	643
	JW300	quy	115,620
	JW300	quz	124,833
Julio Calvo Perez	Spanish Quechua Dictionary Vol. 2	sur	20,606
JRXYZ	Various books	-	140,878
Llamacha	audio transcription	sur	698
Runasimi	dictionary	quy	10,986
	dictionary	quz	22,162

Table 5: Spanish-Quechua parallel dataset.

tokenizer. In the case of the base transformer we chose a BPE tokenizer.

4.2.2 Transformer Model Training Results

Two different sets were used: a validation set and a testing set. The validation set comes from the same original and expanded parallel dataset. The testing set is a parallel dataset provided by AmericasNLP 2024 to compare models.

Table 6 shows the model results for the original parallel dataset, and table 7 shows the expanded parallel dataset. A clear improvement was observed with the expanded parallel dataset over the original in both BLEU and ChrF. Although the scores are low compared to the best scores from AmericasNLP 2024. Considering resource constraints like vanilla transformer without pre-training and MT5-small fine-tuned on a domestic GPU (NVIDIA GeForce GTX 1070), results highlight opportunities for further progress.

5 Conclusion

Synthetic generation of sentences in English did not improve the machine translation. This is because the WordNet technique to generate synthetic sentences was not reliable. On the other hand, using Phi3.5 to generate synthetic sentences improves the MT, particularly in Spanish-Quechua.

Our finding shows that expanding the parallel dataset with synthetic sentences improves the performance of the MT, even if we use a pre-trained transformer (MT5-small) or base transformer model and even though we run our model on a domestic GPU (NVIDIA GeForce GTX 1070).

Identification of the Quechua varieties is still an open problem. It is natural for Quechua speakers, but to our understanding, there are no steps for language identification.

Fluency in the sentences is absent in all current proposals, which needs to be addressed. Fluency would be evaluated by its readability, rhythm, pacing, and the way the sentence structure mirrors natural speech patterns.

Limitations

The parallel dataset is small and domain-constrained, expanding it with synthetic sentences does not guarantee the expansion of the MT in other domains. Despite the existence of millions of Quechua speakers.

References

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, and 110 others. 2024. Phi-3 technical report: A highly capable language model locally on your phone. *Preprint*, arXiv:2404.14219.

Željko Agić and Ivan Vulić. 2019. JW300: A wide-coverage parallel corpus for low-resource languages. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3204–

Model	Dataset	Validation		Testing	
		Bleu	ChrF++	Bleu	ChrF++
MT5	quz	9.97	29.94	1.06	19.85
	quy	11.42	32.24	1.23	21.13
	quz + quz + qu	19.40	*	9.96	29.94
Transformer Base	quz	4.04	35.85	0.02	22.35
	quy	5.26	40.27	0.02	23.90
	quz + quz + qu	*	*	*	*

Table 6: Results of the transformers trained with the original Spanish Quechua corpus.

Model	Dataset	Validation		Testing	
		Bleu	ChrF++	Bleu	ChrF++
MT5	quz	8.56	28.65	1.06	20.38
	quy	9.81	30.50	2.00	22.73
	quz + quz + qu	*	*	*	*
Transformer Base	quz	9.14	41.39	0.04	24.70
	quy	12.38	45.64	0.10	27.43
	quz + quz + qu	*	*	*	*

Table 7: Results of the transformers trained with the expanded Quechua Spanish parallel dataset

3210, Florence, Italy. Association for Computational Linguistics.

Željko Agic and Ivan Vulic. 2019. Jw300: A wide-coverage parallel corpus for low-resource languages. Association for Computational Linguistics.

AmericasNLP. 2021. Mt for spanish (es) - quechua ayacucho (quy). Accessed: 2025-03-20.

Oncevay Arturo and Huarcaya Diego. 2021. Mt-es-quy: Machine translation for spanish-quechua. Accessed: 2025-03-20.

Joseph Attieh, Zachary Hopton, Yves Scherrer, and Tanja Samardžić. 2024. System description of the NordicsAlps submission to the AmericasNLP 2024 machine translation shared task. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 150–158, Mexico City, Mexico. Association for Computational Linguistics.

Julio Calvo Pérez. 2007. Estrategias lexicológicas sobre terminología (en el nuevo diccionario español-quechua/quechua-español). *Estrategias lexicológicas sobre terminología (en el Nuevo Diccionario español-quechua/quechua-español)*, pages 737–757.

William Chen and Brett Fazio. 2021. Morphologically-guided segmentation for translation of agglutinative low-resource languages. In *Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021)*, pages 20–31.

Congreso de la República del Perú. 2008. *Perú Suyu Hatun Kamay Pirwa 1993: Constitución Política del Perú*. Congreso de la República del Perú. Accessed: 2025-03-20.

Ona De Gibert, Raúl Vázquez, Mikko Aulamo, Yves Scherrer, Sami Virpioja, and Jörg Tiedemann. 2023. Four approaches to low-resource multilingual NMT: The Helsinki submission to the AmericasNLP 2023 shared task. In *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 177–191, Toronto, Canada. Association for Computational Linguistics.

Dan Degenaro and Tom Lupicki. 2024. Experiments in mamba sequence modeling and NLLB-200 fine-tuning for low resource multilingual machine translation. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 188–194, Mexico City, Mexico. Association for Computational Linguistics.

Abteen Ebrahimi and et. al. 2022. AmericasNLI: Evaluating zero-shot natural language understanding of pretrained multilingual models in truly low-resource languages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6279–6299, Dublin, Ireland. Association for Computational Linguistics.

Javier Garcia Gilabert, Aleix Sant, Carlos Escolano, Francesca De Luca Fornaciari, Audrey Mash, and Maite Melero. 2024. BSC submission to the AmericasNLP 2024 shared task. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 143–149, Mexico City, Mexico. Association for Computational Linguistics.

Edward Gow-Smith and Danae Sánchez Villegas. 2023. Sheffield's submission to the americasnlp shared task on machine translation into indigenous languages. *arXiv preprint arXiv:2306.09830*.

Diego Huarcaya Taquiri. 2020. Traducción automática neuronal para lengua nativa peruana.

Vivek Iyer, Bhavitvya Malik, Wenhao Zhu, Pavel Stepachev, Pinzhen Chen, Barry Haddow, and Alexandra Birch. 2024. Exploring very low-resource translation with LLMs: The University of Edinburgh‘s submission to AmericasNLP 2024 translation task. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 209–220, Mexico City, Mexico. Association for Computational Linguistics.

Ariadna Font Llitjós. 2005. Developing a quechua-spanish machine translation system.

Ariadna Font Llitjós, Roberto Aranovich, and Lori Levin. 2005. Building machine translation systems for indigenous languages. In *Second Conference on the Indigenous Languages of Latin America (CILLA II), Texas, USA*.

Manuel Mager and et. al. 2021. Findings of the americasnlp 2021 shared task on open machine translation for indigenous languages of the americas. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 202–217.

Arya D McCarthy, Rachel Wicks, Dylan Lewis, Aaron Mueller, Winston Wu, Oliver Adams, Garrett Nicolai, Matt Post, and David Yarowsky. 2020. The johns hopkins university bible corpus: 1600+ tongues for typological exploration. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2884–2892.

Ministerio de la Presidencia de Bolivia. 2012. *Estadoq Kuraq Kamachiynin: Constitución Política del Estado*. Ministerio de la Presidencia y Fundación Konrad Adenauer (KAS). Accessed: 2025-03-20.

Oscar Moreno. 2021. The REPU CS' Spanish–Quechua submission to the AmericasNLP 2021 shared task on open machine translation. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 241–247, Online. Association for Computational Linguistics.

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, and 20 others. 2022. No language left behind: Scaling human-centered machine translation.

Arturo Oncevay. 2021. Peru is multilingual, its machine translation should be too? In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 194–201.

John E Ortega, Richard Castro Mamani, and Kyunghyun Cho. 2020. Neural machine translation with a polysynthetic low resource language. *Machine Translation*, 34(4):325–346.

John E Ortega, Richard Alexander Castro Mamani, and Jaime Rafael Montoya Samame. 2021. Love thy neighbor: combining two neighboring low-resource languages for translation. In *Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021)*, pages 44–51.

Annette Rios. 2015. *A basic language technology toolkit for quechua*. Ph.D. thesis, University of Zurich.

Annette Rios and Anne Göhring. 2013. Machine learning disambiguation of quechua verb morphology. Association for Computational Linguistics.

Annette Rios and Anne Göhring. 2016. Machine learning applied to rule-based machine translation. *Hybrid approaches to machine translation*, pages 111–129.

Runasimi.de. 2006. Runasimi - quechua language resources. Accessed: 2025-03-20.

Julio Santisteban and Javier Tejada-Cárcamo. 2015. Unilateral jaccard similarity coefficient. In *GSB@ SIGIR*, pages 23–27.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *Lrec*, volume 2012, pages 2214–2218. Citeseer.

Jörg Tiedemann. 2020. The tatoeba translation challenge–realistic data sets for low resource and multilingual mt. *arXiv preprint arXiv:2010.06354*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention is all you need. *Preprint*, arXiv:1706.03762.

Raúl Vázquez, Yves Scherrer, Sami Virpioja, and Jörg Tiedemann. 2021. The helsinki submission to the americasnlp shared task. In *Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 255–264. The Association for Computational Linguistics.

Hugo David Calderon Vilca, Vilca César David Mamani Calderón, Flor Cagniy Cárdenas Mariño, and Edwin Fredy Mamani Calderón. 2009. Traductor automático en linea del español a quechua, basado en la plataforma libre y código abierto apertium. *Revista de Investigaciones de la Escuela de Posgrado de la UNA PUNO*, 5(3):81–99.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mt5: A massively multilingual pre-trained text-to-text transformer. *Preprint*, arXiv:2010.11934.