A Continuous Improvement Framework of Machine Translation for Shipibo-Konibo

Erasmo Gómez Montoya✣   **Kervy Rivas-Rojas**✣   **Arturo Oncevay**✣♠

✣ Department of Engineering, Pontificia Universidad Católica del Perú
♠ School of Informatics, University of Edinburgh
{`hector.gomez,k.rivas`}`@pucp.pe`, `a.oncevay@ed.ac.uk`

Abstract

Shipibo-Konibo is a low-resource language from Peru with prior results in statistical machine translation; however, it is challenging to enhance them mainly due to the expensiveness of building more parallel corpora. Thus, we aim for a continuous improvement framework of the Spanish–Shipibo-Konibo language-pair by taking advantage of more advanced strategies and crowd-sourcing. Besides the introduction of a new domain for translation based on language learning flashcards, our main contributions are the extension of the machine translation experiments for Shipibo-Konibo to neural architectures with transfer and active learning; and the building of a conversational agent prototype to retrieve new translations through a social media platform.

1 Introduction

The focus on low-resource Machine Translation (MT) has driven further work with different machine learning settings to take advantage of Neural MT (NMT) methods, where the amount of training data is relevant to obtain quality results (Koehn and Knowles, 2017). For instance, with a Transfer Learning approach, we can learn specific components in a system from a resource-rich domain (e.g. a language-pair) and transfer the updated parameters to the real target (Zoph et al., 2016), usually in a resource-poor domain. Regarding the size of available corpora, with Active Learning meth-

ods, we can rank new samples to label (e.g. sentences to translate) to improve a learning system efficiently (Haffari et al., 2009). Besides, crowd-sourcing strategies and platforms, such as Amazon Mechanical Turk, have gained attention in translation studies and MT to retrieve less expensive corpora (Jiménez-Crespo, 2017).

Given the background, Peru offers a diversity-rich language context for MT research with more than 40 native languages (Simons and Fenning, 2019) that are typologically different from Castilian Spanish (spa), the primary official language in the country. Specifically, Shipibo-Konibo (shp) is an Amazonian language that has been addressed in Natural Language Processing (NLP) recently, including a statistical MT (SMT) study with religious and educational domain corpora (Galarreta et al., 2017). However, the language is far from being considered a rich-resource one with less than 20,000 sentences for the spa–shp language-pair. Thus, it is crucial to look for different approaches that could deliver better MT systems, and also, new parallel corpus.

Therefore, this study extends previous MT studies of Shipibo-Konibo by introducing a new domain for translation based on flashcards for language learning (see §4), experimenting with transfer and active learning strategies in neural architectures (see §5), and proposing a conversational agent prototype in social media to retrieve new translations from native speakers (see §6). Our main goal is to mount an initial framework able to continuously improve MT for Peruvian languages, with the potential to include further NMT features. To complement the article, §2 presents previous work on MT for Peruvian languages, §3 introduces more details about the target language, and finally, §7 concludes and proposes further steps.

	$\mathcal{S}$	$\overline{r}_{\text{shp–spa}}$	$\mathcal{T}$		$|\mathcal{V}|$		HLT	
			spa	shp	spa	shp	spa	shp
Religious	12,547	0.9476	195,887	185,638	13,620	19,091	6,426	11,115
Educational	5,982	0.9148	53,710	49,135	4,351	6,568	1,649	4,044
Flashcards	7,740	1.0966	20,858	22,874	6,382	5,133	4,234	3,312
Total	26,269	0.9526	270,455	257,647	21,710	28,024	10,954	16,875

Table 1: Details of the parallel corpora for spa–shp per domain and in total: $\mathcal{S}$ = number of sentences; $\overline{r}_{\text{shp–spa}}$ = average of the ratio$_{\text{shp–spa}}$ per sentence; $\mathcal{T}$ = number of tokens; $|\mathcal{V}|$ = vocabulary size; HLT = tokens with frequency equals to one.

2 Related work

In Peru, the Quechuan language family has been the primary target in MT. According to the survey of Mager et al. (2018), there are studies in rule-based MT (RBMT), based on the Apertium platform (Forcada et al., 2011), for Quechua Eastern Apurimac (qve) and Quechua Cuzco (quz) (Cavero and Madariaga, 2007). Another study in RBMT, from the project AVENUE, also targeted quz (Monson et al., 2006). Regarding topics closer to SMT, Ortega and Pillaipakkamnatt (2018) improved alignments for a Quechua variant by using an agglutinative language as Finnish as a pivot. The source for their parallel corpus is Rios et al. (2012), so we know that they worked with Quechua Cuzco (quz) too.

Apart from the Quechuan languages, just Aymara in RBMT (Coler and Homola, 2014), and Shipibo-Konibo in SMT (Galarreta et al., 2017) has been studied in MT, with Spanish as their paired language. Besides, the latter is the only Amazonian language in this scope. Moreover, to the best of our knowledge, there are not experiments with neural architectures or further learning strategies for languages of Peru or the Amazon.

Furthermore, and besides the Peruvian scope, there is a large body of knowledge on transfer learning for low-resource MT (Zoph et al., 2016), active learning for MT (Haffari et al., 2009), and collaborative translation (Jiménez-Crespo, 2017).

3 Language specifics

Shipibo-Konibo (shp) belongs to the Panoan language family, and there are more than 30,000 native speakers. It is a morphologically-rich language with agglutinative processes. Besides, there is a preponderance of suffixes in contrast to prefixes, and it includes some clitics particles. In contrast to Spanish, Shipibo-Konibo presents different word orders (e.g. predominance of SOV against SVO), which implies a more challenging scenario.

One of the reasons to target Shipibo-Konibo is the robust capabilities of the ethnic group to preserve its culture and language despite the several years of contact with Spanish speakers (Crevels, 2012). Moreover, they are one of the few native communities in Peru with a socio-political and cultural organisation[1].

Regarding the research and development in language technologies, there are outcomes in different levels, such as a spell-checker (Alva and Oncevay, 2017), a morphological analyser (Cardenas and Zeman, 2018), a lemmatiser with POS-tagger (Pereira-Noriega et al., 2017), a syntax dependency parser (Vásquez et al., 2018) and an SMT system paired with Spanish (Galarreta et al., 2017). Each study includes resources carefully crafted by bilingual speakers and linguists.

4 Dataset

A previous study of spa–shp introduced two corpora: religious and educational (Galarreta et al., 2017). The former is a compilation with post-processing steps of the Bible entries, whereas the latter contains translated sentences of bilingual educational texts from the Peruvian Government[2].

Besides those domains, we introduce a new parallel corpus that was built from a sample of sentences of the Tatoeba project, specifically, the Tab-delimited Bilingual Sentence Pairs in English–Spanish[3]. A few thousands of short sentences were translated from Spanish into Shipibo-Konibo for a certified bilingual translator. We named the new corpus Flashcards, as it is based on flashcards with bilingual sentences to easier memorisation in a language learning context[4].

Table 1 describes the corpus per domain and overall, including information about the number of

---

[1] *Coshikok*: http://www.coshikoxperu.org/
[2] We used an updated version: http://chana.inf.pucp.edu.pe/resources/parallel-corpus/
[3] http://www.manythings.org/anki/
[4] The new parallel corpus is going to be published

translated sentences, an average of the ratio of tokens per sentence between the Shipibo-Konibo and Spanish translations (Galarreta et al., 2017), the total number of tokens, the vocabulary size, and the amount of *hapax legomenon* tokens (HLT) or terms with frequency equals to one.

We observe that the Flashcards domain is proportionally bigger in vocabulary size and HLT regarding the other two, even when the amount of tokens per sentence in average is lower ($\mathcal{T}/\mathcal{S}$). Moreover, the averaged ratio of tokens ($\overline{r}_{\text{shp–spa}}$) has a particular value, as it is the only domain with more tokens per parallel sentence in the Shipibo-Konibo side than in the Spanish one. The following example illustrates a related case:

> **shp**: *Westiora kafe keniresa ea ike.*
> **spa**: *Sólo quería un café.*
> **eng**: *I just wanted a coffee.*

where there is a null subject in Spanish (*ea* or *I*), and Shipibo-Konibo merges *sólo quería* (*just wanted*) into *keniresa* and adds *ike* as an auxiliar.

5 Neural Machine Translation for Shipibo-Konibo

The NMT paradigm has achieved state-of-the-art results mostly with large-resource settings. The training of NMT systems is an open challenge for low-resource language-pairs (Koehn and Knowles, 2017), but we consider a must the alignment to this paradigm, as it is going to be the main focus of the MT research for the following years.

NMT is based on an encoder-decoder framework to perform an end-to-end translation using sequence-to-sequence neural networks (Sutskever et al., 2014). For the encoder, we have a recurrent neural network that receives a source sentence and outputs a dense encoded vector. Similarly, the decoder is another recurrent network that transforms the vector into a target sentence.

In this paper, we use a two-layer encoder-decoder LSTM network. Additionally, we use teacher forcing with 0.5 in the encoder and an attention mechanism in the decoder to look back at the source (Luong et al., 2015). Besides, the number of units of the hidden layer is 1024, the embedding size is 128, and the batch size is 64. We use Adam optimiser and train for ten epochs.

Given the baseline settings, we performed the first experiments at word- and subword-level. For the latter, we use Byte Pair Encoding (BPE) (Sennrich et al., 2016) with different merge operations.

| | BLEU$_{\text{w}}$ | BLEU$_{\text{BPE}}$ | |
		5k	15k
Religious	1.29	**2.08**	1.33
Educational	4.10	**4.91**	3.21
Flashcards	**11.95**	11.15	11.11
Total	3.76	**3.94**	2.46

Table 2: BLEU scores with the NMT baseline settings at word- (w) and subword-level using joint BPE with 5,000 (5k) and 15,000 (15k) merge operations for the latter.

Whereas for evaluation, we take 10% of the corpus as development and other 10% as testing sets per each domain and overall.

As we can see in Table 2, the initial results were meagre as expected, with an exception in the new Flashcards domain, where the BLEU score might be higher due to the short length of the sentences. Regarding the subword evaluation with BPE, there are slightly better values in some cases (with the lower amount of merge operations), which is an anticipated trend for the agglomerative nature of the language. Nevertheless, the scores in both religious and educational domains are lower than the SMT system of Galarreta et al. (2017), and the overall result confirms the neediness of using additional strategies for improving the low-resource NMT setting. We examine the next steps only at word-level to control the variables.

5.1 Transfer learning

Following the study of Zoph et al. (2016), we defined a parent language-pair (Spanish to L or spa–L) to benefit a child language-pair (Spanish to Shipibo-Konibo) by pre-initializing parameters of the child using the updated values at the end of the parent training in the encoder-decoder. For exploration purposes, we use a short but diverse set of L languages regarding their potential closeness to Shipibo-Konibo in typological properties.

Table 3 presents the set of languages analysed[5]. The parallel corpora aligned with Spanish is retrieved from several sources: Turkish from OPUS (Tiedemann, 2012), German and Hebrew from the TED Multilingual Parallel Corpus[6], and English from the same source as the new Flashcards corpus. In the case of Hebrew, we transliterated the corpus to the Latin alphabet.

---

[5] We choose four languages to make the experiments: English, German, Turkish and Hebrew. The four languages were chosen due to the availability of the datasets

[6] `https://github.com/ajinkyakulkarni14/TED-Multilingual-Parallel-Corpus`

L (lang.)	$\mathcal{S}_{\text{spa}-L}$	BLEU$_{\text{spa}-\text{shp}}$	$D_{(\text{shp},L)}$
English	120,566	6.34	0.2822
German	452,661	4.45	0.3382
Turkish	7,177	9.22	**0.1764**
Hebrew	486,466	**12.34**	0.4264

Table 3: Transfer learning experiments using spa–L as a parent language-pair. $\mathcal{S}$ indicates the size of the corpus, BLEU the score of translation in the child language-pair spa–shp, and D is the Hamming distance between L and shp.

	Initial	+ Rand	+ AL
Religious	4.12	4.70	**5.78**
Educational	5.65	5.89	**6.30**
Flashcards	10.20	12.30	**14.71**
Total	9.12	9.75	**10.43**

Table 4: BLEU scores for the 40% incremental step over the initial 50% in the Active Learning experimental setting.

Additionally, we include the transfer learning results for translation, in terms of BLEU score, using the entire spa–shp corpus[7]. We observe that English and German slightly overcome the NMT baseline results; however, Turkish and Hebrew show a significant improvement. The case of Turkish is even more promising, as its corpus size is the smallest among the four languages.

We also present a language similarity score with Shipibo-Konibo. Alike Agić (2017), we compute a language distance using the Hamming distance function with language vectors extracted from the WALS (World Atlas Language Structure) knowledge base of linguistic typology (Dryer and Haspelmath, 2013). We only considered syntactic properties (e.g. word order), as we were performing translation at word-level, and we took advantage of the 103 binary features processed in `lang2vec` (Littell et al., 2017). However, it is worth noting that there are several missing values, specially for Shipibo-Konibo, due to the sparsity of WALS. Thus, we solely preserved the categories with completed entries across the five languages involved, lowering the dimensionality to 68.

A recent study in transfer learning for MT (Kocmi and Bojar, 2018) argued that it might be more important the size of the corpus of the parent language-pair rather than the similarity of the languages concerned. Our results are partially aligned with their claim, but we observe that English and German cannot overcome Turkish despite the large difference in corpus size. However, we cannot derive further conclusions about language distance as a proper measurement for improving transfer learning results, as Hebrew was the most different language, in terms of syntax, and obtained the best translation score in the transfer setup. Nevertheless, we think the metric should be reviewed

carefully, as there are several missing records in WALS. Moreover, the Spanish–Turkish parallel corpus is composed only by GNOME and Ubuntu localisation files, a scanty and limited domain for translation.

A more objective analysis could be performed using similar size and domain corpora, although those requisites are tough to satisfy in MT. Furthermore, if we want to evaluate a subword-level transfer context, we should include morphological features to the language similarity measure as well. Nonetheless, for the next experimental setting, we take as a basis the parameter values learned in the Spanish–Hebrew language-pair.

5.2 Active learning

In this part of the study, we emulate a pool-based active learning setting, where we need to select iterations of sample batches to incrementally improve the MT system. For the sampling query, we partially adapt elemental heuristics proposed for SMT (Haffari et al., 2009). Specifically, we focus on n-gram heuristics (1-gram) to select new sentences based on out-of-vocabulary (OOV) words and term frequency. Due to the high presence of unique and HLT in the corpora, it is relevant to deal with OOV terms, and even more when the target language is an agglutinative one. Besides, this heuristic could be insightful for further subword-level experimentations using BPE.

The evaluation of the active learning approach is performed per domain and altogether. We separate 20% of the parallel corpora as the validation and test sub-sets with 10% each, and the rest of the corpus is used for the pool-based evaluation. We take half of the sentences available as the baseline subset (Initial), and we perform a one-step increment (+40%).

Table 4 presents the BLEU scores, and we observe that the Active Learning criterion achieved better results than random in all the experiments. Although, it is worth noting that the overall results are very low, mainly due to the amount of data

---

[7]We decided not to divide the corpus per domain due to the amount of data, and because we only need the parent language-pair parameters to pre-initialize the next experiment in Active Learning

available. We expect to integrate novel queries and active learning settings proposed directly in the NMT paradigm (Liu et al., 2018). Nevertheless, as the primary goal of the study is the development of a continuous improvement framework, we consider that different AL strategies could offer a proper foundation to incrementally enhance the MT systems for Shipibo-Konibo.

6 Conversational agent prototype for crowd-sourcing

We take inspiration on the actions taken in an humanitarian emergency (Munro, 2010), where there was a need to solve translation queries on-the-fly. In our context, we consider that the endangerment of a native language is an emergency for the community as well, and we would like to reach the speakers to involved them in this revitalisation effort from a computational perspective. Thus, to support the continuous improvement of the MT system developed so far, we expect to retrieve collaborative and crowd-sourced translations from native speakers through social media, which provides extended channels with few access constraints or limitations. For this reason, we decided to build a conversational agent, and we describe our work in progress of the current prototype.

6.1 Interaction strategies

To apply the collaborative learning for translation, we designed a persistent model to support the interaction between the user and the application. The model includes features such as: the storage of potential translations from users and non-translated sentences (in Spanish), the selection of non-translated sentences to be presented to the users, and the integration of the new translations in following training iterations.

The model can be adjusted with different parameters, such as a limited sentence length or term frequency for the selection process, or the number of references translations required from different users given a non-translated sentence. The latter is a significant feature in crowd-sourcing settings, as we cannot assume that a professional translator is going to provide all the feedback, thus, we need to take many references from the crowd to reduce potential noise.
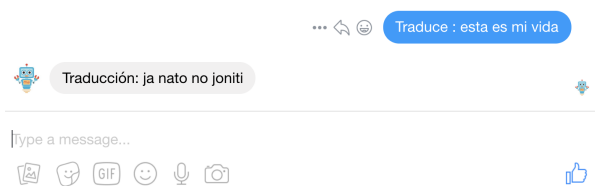

Figure 1: First story: a user requests an automatic translation from Spanish into Shipibo-Konibo. User says: "Translate: This is my life", and the conversational agent answers with the translation ("Traducción")
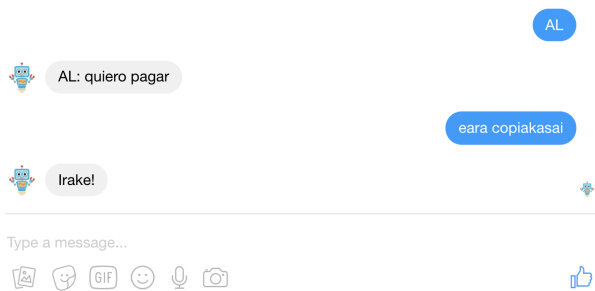

Figure 2: Second story: a user offers its help and the system requests for a new translation. First, the user writes "AL" to start the interaction (more natural expressions are going to be integrated). Then, the system requests for a translation ("I want to pay") and the user answers. Finally, the system ends the interaction with thanks in Shipibo-Konibo.

6.2 Design and implementation

We built a framework for developing a webhook that supports the interaction with the Facebook Messenger API[8] (version 3.2). The webhook supports two types of interactions, known as stories.

The first story refers to requests from users to the conversational agent for translating a phrase or sentence. The translation request must be from Spanish to Shipibo-Konibo, as we can see in Figure 1. The aim of this first story is to engage potential learners interested in the language, or professional translators that want to analyse and post-edit the output of MT systems.

The second story, in contrast with the first one, takes advantage of crowd-sourcing, as it involves a translation requirement from the system to the user after receiving a manifest of support. In Figure 2 we observe the conversational agent asking for the translation of a sentence. The text has been extracted from the pool of non-translated flashcards by using the active learning criterion.

---

[8] We chose this platform because it has been our main communication channel with the certified translators during the corpus development. Official site: `https://developers.facebook.com/docs/graph-api`

6.3 Further development

Apart from technical details to support the model persistence in a large-scale number of interactions, there must be a focus in building a robust communication flow in the stories. For instance, the fist story could be extended to accept feedback of the users in a post-edition setting, although there should be a mechanism to distinguish professional translators from other speakers. In case of the second story, the system could ask to retrieve more translations instead of ending the interaction immediately. Moreover, there should be a usability test for the conversational agent to identify the best interaction flow for the users (native speakers).

7 Conclusion

We presented additional MT results for Shipibo-Konibo using sequence-to-sequence neural networks, altogether with transfer learning and active learning strategies. We also introduced a new parallel corpus domain which texts are used in a language learning context. Overall results are aligned to the amount of data available; however, we observed a promising upward trend in the performance, even more when the new domain is involved. Thus, we integrated an NMT model within a conversational agent prototype to retrieve crowd-sourcing and collaborative translations through social media. These have been the initial steps to set up a continuous improvement framework for MT in Shipibo-Konibo.

Furthermore, as we built the current system in the NMT paradigm, we could integrate novel features to steadily improve the performance. Also, we plan to complete the pairwise-system with the translation direction from Shipibo-Konibo into Spanish, and take advantage of monolingual data in Shipibo-Konibo to enhance the encoder-decoder components at subword-level.

Acknowledgments

References

Agić, Željko. 2017. Cross-lingual parser selection for low-resource languages. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 1–10.

Alva, Carlo and Arturo Oncevay. 2017. Spell-checking based on syllabification and character-level graphs for a Peruvian agglutinative language. In *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, pages 109–116. Association for Computational Linguistics.

Cardenas, Ronald and Daniel Zeman. 2018. A morphological analyzer for Shipibo-konibo. In *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 131–139, Brussels, Belgium, October. Association for Computational Linguistics.

Cavero, Indhira Castro and Jaime Farfán Madariaga. 2007. Traductor morfológico del castellano y quechua (Morphological translator of Castilian Spanish and Quechua). *Revista I+ i*, 1(1).

Coler, Matthew and Petr Homola, 2014. *Rule-based machine translation for Aymara*, pages 67–80. Cambridge University Press, 10.

Crevels, Mily, 2012. *Language endangerment in South America: The clock is ticking*, volume 2, pages 176–234. Walter de Gruyter.

Dryer, Matthew S. and Martin Haspelmath, editors. 2013. *WALS Online - World Atlas of Language Structures*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Forcada, Mikel L, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim ORegan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, and Francis M Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine translation*, 25(2):127–144.

Galarreta, Ana Paula, Andrés Melgar, and Arturo Oncevay-Marcos. 2017. Corpus creation and initial SMT experiments between Spanish and Shipibo-konibo. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, Varna, Bulgaria, September 2 - 8, 2017*, pages 238–244.

Haffari, Gholamreza, Maxim Roy, and Anoop Sarkar. 2009. Active learning for statistical phrase-based machine translation. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 415–423, Boulder, Colorado, June. Association for Computational Linguistics.

Jiménez-Crespo, Miguel A. 2017. *Crowdsourcing and online collaborative translations: Expanding*

the limits of translation studies, volume 131. John Benjamins Publishing Company.

Kocmi, Tom 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, Belgium, Brussels, October. Association for Computational Linguistics.

Koehn, Philipp and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39. Association for Computational Linguistics.

Littell, Patrick, David R Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori Levin. 2017. Uriel and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, volume 2, pages 8–14.

Liu, Ming, Wray Buntine, and Gholamreza Haffari. 2018. Learning to actively learn neural machine translation. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 334–344, Brussels, Belgium, October. Association for Computational Linguistics.

Luong, Minh-Thang, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.

Mager, Manuel, Ximena Gutierrez-Vasques, Gerardo Sierra, and Ivan Meza-Ruiz. 2018. Challenges of language technologies for the indigenous languages of the Americas. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 55–69. Association for Computational Linguistics.

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

Munro, Robert. 2010. Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge. In *AMTA 2010 Workshop on Collaborative Crowdsourcing for Translation*, pages 1–4.

Ortega, John and Krishnan Pillaipakkamnatt. 2018. Using morphemes from agglutinative languages like Quechua and Finnish to aid in low-resource translation. In *Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018)*, pages 1–11.

Pereira-Noriega, José, Rodolfo Mercado, Andrés Melgar, Marco Sobrevilla-Cabezudo, and Arturo Oncevay Marcos. 2017. Ship-LemmaTagger: Building an NLP toolkit for a Peruvian native language. In *International Conference on Text, Speech, and Dialogue*, Lecture Notes in Computer Science, pages 473–481. Springer, July.

Rios, Annette, Anne Ghring, and Martin Volk. 2012. Parallel Treebanking Spanish-Quechua: how and how well do they align? *Linguistic Issues in Language Technology*, 7(1).

Sennrich, Rico, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, August. Association for Computational Linguistics.

Simons, Gary F. and Charles D. Fenning, editors. 2019. *Ethnologue: Languages of the World. Twenty-second edition*. Dallas Texas: SIL international. Online version: http://www.ethnologue.com.

Sutskever, Ilya, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in OPUS. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Mehmet Uur Doan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Vásquez, Alonso, Renzo Ego Aguirre, Candy Angulo, John Miller, Claudia Villanueva, Željko Agić, Roberto Zariquiey, and Arturo Oncevay. 2018. Toward universal dependencies for Shipibo-konibo. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 151–161, Brussels, Belgium, November. Association for Computational Linguistics.

Zoph, Barret, 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, November. Association for Computational Linguistics.