Family of Origin and Family of Choice:
Massively Parallel Lexiconized Iterative Pretraining for Severely Low Resource Text-based Translation

Zhong Zhou

Carnegie Mellon University

zhongzhou@cmu.edu

Alex Waibel

Carnegie Mellon University

alex@waibel.com

Abstract

We translate a closed text that is known in advance into a severely low resource language by leveraging massive source parallelism. In other words, given a text in 124 source languages, we translate it into a severely low resource language using only ~1,000 lines of low resource data without any external help. Firstly, we propose a systematic method to rank and choose source languages that are close to the low resource language. We call the linguistic definition of language family *Family of Origin* (FAMO), and we call the empirical definition of higher-ranked languages using our metrics *Family of Choice* (FAMC). Secondly, we build an *Iteratively Pretrained Multilingual Order-preserving Lexiconized Transformer* (IPML) to train on ~1,000 lines (~3.5%) of low resource data. To translate named entities correctly, we build a massive lexicon table for 2,939 Bible named entities in 124 source languages, and include many that occur once and covers more than 66 severely low resource languages. Moreover, we also build a novel method of combining translations from different source languages into one. Using English as a hypothetical low resource language, we get a +23.9 BLEU increase over a multilingual baseline, and a +10.3 BLEU increase over our asymmetric baseline in the Bible dataset. We get a 42.8 BLEU score for Portuguese-English translation on the medical EMEA dataset. We also have good results for a real severely low resource Mayan language, Eastern Pokomchi.

1 Introduction

We translate a closed text that is known in advance into a severely low resource language by leveraging massive source parallelism. In other words, we aim to translate well under three constraints: having severely small training data in the new target low resource language, having massive source language parallelism, having the same closed text across all languages. Generalization to other texts is prefer-

Eastern Pokomchi		English	
FAMD	*FAMP*	*FAMD*	*FAMP*
Chuj*	Dadibi	Danish*	Dutch*
Cakchiquel*	Thai	Norwegian*	Afrikaans*
Guajajara*	Gumatj	Italian	Norwegian*
Toba	Navajo	Afrikaans*	German*
Myanmar	Cakchiquel*	Dutch*	Danish*
Slovenský	Kanjobal	Portuguese	Spanish
Latin	Guajajara*	French	Frisian*
Ilokano	Mam*	German*	Italian
Norwegian	Kim	Marshallese	French
Russian	Chuj*	Frisian*	Portuguese

Table 1: Top ten languages closest to Eastern Pokomchi (left) and English (right) in ranking 124 source languages. *FAMD* and *FAMP* are two constructions of Family of Choice (*FAMC*) by distortion and performance metrics respectively. All are trained on ~1,000 lines. We star those in Family of Origin.

able but not necessary in the goal of producing high quality translation of the closed text.

2020 is the year that we started the life-saving hand washing practice globally. Applications like translating water, sanitation, and hygiene (WASH) guidelines into severely low resource languages are very impactful in tribes like those in Papua New Guinea with 839 living languages (Gordon Jr, 2005; Simons and Fennig, 2017). Translating humanitarian texts like WASH guidelines with scarce data and expert help is key (Bird, 2020).

We focus on five challenges that are not addressed previously. Most multilingual transformer works that translate into low resource language limit their training data to available data in the same or close-by language families or the researchers' intuitive discretion; and are mostly limited to less than 30 languages (Gu et al., 2018; Zhou et al., 2018a; Zhu et al., 2020). Instead, we examine ways to pick useful source languages from 124 source languages in a principled fashion. Secondly, most works require at least 4,000 lines of low resource data (Lin et al., 2020; Qi et al., 2018; Zhou et al., 2018a); we use only ~1,000 lines of low resource data to simulate real-life situation of having ex-

tremely small seed target translation. Thirdly, many works use rich resource languages as hypothetical low resource languages. Moreover, most works do not treat named entities separately; we add an order-preserving lexiconized component for more accurate translation of named entities. Finally, many multilingual works present final results as sets of translations from all source languages; we build a novel method to combine all translations into one.

We have five contributions. Firstly, we rank the 124 source languages to determine their closeness to the low resource language and choose the top few. We call the linguistic definition of language family *Family of Origin* (FAMO), and we call the empirical definition of higher-ranked languages using our metrics *Family of Choice* (FAMC). They often overlap, but may not coincide.

Secondly, we build an *Iteratively Pretrained Multilingual Order-preserving Lexiconized Transformer* (IPML) training on ~1,000 lines of low resource data. Using iterative pretraining, we get a +23.9 BLEU increase over a multilingual order-preserving lexiconized transformer baseline (MLc) using English as a hypothetical low resource language, and a +10.3 BLEU increase over our asymmetric baseline. Training with the low resource language on both the source and target sides boosts translation into the target side. Training on randomly sampled 1,093 lines of low resource data, we reach a 31.3 BLEU score testing on 30,022 lines of Bible. We have a 42.8 BLEU score for Portuguese-English translation on the medical EMEA dataset.

Thirdly, we use a real-life severely low resource Mayan language, Eastern Pokomchi, a Class 0 language (Joshi et al., 2020) as one of our experiment setups. In addition, we also use English as a hypothetical low resource language for easy evaluation.

We also add an order-preserving lexiconized component to translate named entities well. To solve the variable-binding problem to distinguish "Ian calls Yi" from "Yi calls Ian" (Fodor and Pylyshyn, 1988; Graves et al., 2014; Zhou et al., 2018a), we build a lexicon table for 2,939 Bible named entities in 124 source languages including more than 66 severely low resource languages.

Finally, we combine translations from all source languages by using a novel method. For every sentence, we find the translation that is closest to the translation cluster center. The expected BLEU score of our combined translation is higher than translation from any of the individual sources.

2 Related Works

2.1 Information Dissemination

Interactive Natural Language Processing (NLP) systems are classified into information assimilation, dissemination, and dialogue (Bird, 2020; Ranzato et al., 2015; Waibel and Fugen, 2008). *Information assimilation* involves information flow from low resource to rich resource language communities while *information dissemination* involves information flow from rich resource to low resource language communities. Taken together, they allow *dialogue* and interaction of different groups at eye level. Most work on information assimilation (Bérard et al., 2020; Earle et al., 2012; Brownstein et al., 2008). Few work on dissemination due to small data, less funding, few experts and limited writing system (Östling and Tiedemann, 2017; Zoph et al., 2016; Anastasopoulos et al., 2017; Adams et al., 2017; Bansal et al., 2017).

2.2 Machine Polyglotism and Pretraining

Recent research on machine polyglotism involves training machines to be adept in many languages by adding language labels in the training data with a single attention (Johnson et al., 2017; Ha et al., 2016; Firat et al., 2016; Gillick et al., 2016; Zhou et al., 2018b). Some explores data symmetry (Freitag and Firat, 2020; Birch et al., 2008; Lin et al., 2019). Zero-shot translation in severely low resource settings exploits the massive multilinguality, cross-lingual transfer, pretraining, iterative back-translation and freezing subnetworks (Lauscher et al., 2020; Nooralahzadeh et al., 2020; Wang et al., 2020; Li et al., 2020; Pfeiffer et al., 2020; Baziotis et al., 2020; Chronopoulou et al., 2020; Lin et al., 2020; Thompson et al., 2018; Luong et al., 2014; Wei et al., 2020; Dou et al., 2020).

2.3 Linguistic Distance

To construct linguistic distances (Hajič, 2000; Oncevay et al., 2020), some explore typological distance (Chowdhury et al., 2020; Rama and Kolachina, 2012; Pienemann et al., 2005; Svalberg and Chuchu, 1998; Hansen et al., 2012; Comrie, 2005), lexical distance (Huang et al., 2007), Levenshtein distance and Jaccard distance (Serva and Petroni, 2008; Holman et al., 2008; Adebara et al., 2020), sonority distance (Parker, 2012) and spectral distance (Dubossarsky et al., 2020).

3 Methodology

3.1 Multilingual Order-preserving Lexiconized Transformer

3.1.1 Multilingual Transformer

In training, each sentence is labeled with the source and target language label. For example, if we translate from Chuj ("ca") to Cakchiquel ("ck"), each source sentence is tagged with `__opt_src_ca __opt_tgt_ck`. A sample source sentence is "`__opt_src_ca __opt_tgt_ck` Tec'b'ejec e b'a mach ex tzeyac'och Jehová yipoc e c'ool".

We train on Geforce RTX 2080 Ti using ~100 million parameters, a 6-layer encoder and a 6-layer decoder that are powered by 512 hidden states, 8 attention heads, 512 word vector size, a dropout of 0.1, an attention dropout of 0.1, 2,048 hidden transformer feed-forward units, a batch size of 6,000, "adam" optimizer, "noam" decay method, and a label smoothing of 0.1 and a learning rate of 2.5 on OpenNMT (Klein et al., 2017; Vaswani et al., 2017). After 190,000 steps, we validate based on BLEU score with early stopping patience of 5.

3.1.2 Star Versus Complete Configuration

We show two configurations of translation paths in Figure 1: *star* graph (multi-source single-target) configuration and *complete* graph (multi-source multi-target) configuration. The complete configuration data increases quadratically with the number of languages while the star configuration data increases linearly.

3.1.3 Order-preserving Lexiconized transformer

The variable binding problem issue is difficult in severely low resource scenario; most neural models cannot distinguish the subject and the object of a simple sentence like "Fatma asks her sister Wati to call Yi, the brother of Andika", especially when all named entities appear once or never appear in training (Fodor and Pylyshyn, 1988; Graves et al., 2014). Recently, researchers use order-preserving lexiconized Neural Machine Translation models where named entities are sequentially tagged in a sentence as `__NE`s (Zhou et al., 2018a). The previous example becomes "`__NE0` asks her sister `__NE1` to call `__NE2`, the brother of `__NE3`".

This method works under the assumption of translating a closed text known in advance. Its success relies on good coverage of named entities. To cover many named entities, we build on existing

(a) Complete graph configuration
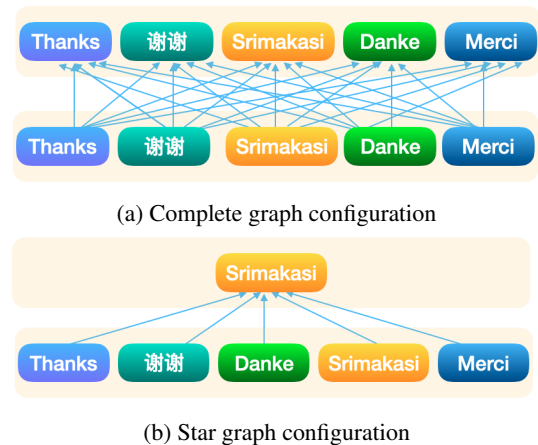

(b) Star graph configuration

Figure 1: (a) Complete graph configuration of translation paths (Many-to-many) in an example of multilingual translation. (b) Star configuration of translation paths (Many-to-one) using Indonesian as the low resource example.

research literature (Wu et al., 2018; Zhou et al., 2018a) to construct a massively parallel lexicon table that covers 2,939 named entities across 124 languages in our Bible database. Our lexicon table is an expansion of the existing literature that covers 1,129 named entities (Wu et al., 2018). We add in 1,810 named entities that are in the extreme end of the tail occurring only once. We also include 66 more real-life severely low resource languages.

For every sentence pair, we build a target named entity decoding dictionary by using all target lexicons from the lexicon table that match with those in the source sentence. In severely low resource setting, our sequence tagging is larged based on dictionary look-up; we also include lexicons that are not in the dictionary but have small edit distances with the source lexicons. In evaluation, we replace all the ordered `__NE`s using the target decoding dictionary to obtain our final translation.

Let us translate "Fatma asks her sister Wati to call Yi, the brother of Andika" to Chinese and German. Our tagged source sentence that translates to Chinese is "`__opt_src_en __opt_tgt_zh __NE0` asks her sister `__NE1` to call `__NE2`, the brother of `__NE3`"; and we use `__opt_tgt_de` for German. The source dictionary is "`__NE0`: Fatma, `__NE1`: Wati, `__NE2`: Yi, `__NE3`: Andika" and we create the target dictionaries. The Chinese output is "`__NE0`叫她的姐妹`__NE1`去打电话给`__NE3`的兄弟`__NE2`" and the German output is "`__NE0` bittet ihre Schwester `__NE1` darum, `__NE2`, den Bruder `__NE3`, anzurufen". We decode the named entities to get final translations.

3.2 Ranking Source Languages

Existing works on translation from multiple source languages into a single low resource language usually have at most 30 source languages (Gu et al., 2018; Zhou et al., 2018a; Zhu et al., 2020). They are limited within the same or close-by language families, or those with available data, or those chosen based on the researchers' intuitive discretion. Instead, we examine ways to pick useful source languages in a principled fashion motivated by cross-lingual impacts and similarities (Shoemark et al., 2016; Sapir, 1921; Odlin, 1989; Cenoz, 2001; Toral and Way, 2018; De Raad et al., 1997; Hermans, 2003; Specia et al., 2016). We find that using many languages that are distant to the target low resource language may produce marginal improvements, if not negative impact. Indeed, existing literature on zero-shot translation also suffers from the limitation of linguistic distance between the source languages and the target language (Lauscher et al., 2020; Lin et al., 2020; Pfeiffer et al., 2020). We therefore rank and select the top few source languages that are closer to the target low resource language using the two metrics below.

We rank source languages according to their closeness to the low resource language. We construct the Family of Choice (FAMC) by comparing different ways of ranking linguistic distances empirically based on the small low resource data.

Let S_s and S_t be the source and target sentences, let L_s be the source length, let $P(S_t = s_t|s_s, l_s)$ be the alignment probability, let F_s be the fertility of how many target words a source word is aligned to, let D_t be the distortion based on the fixed distance-based reordering model (Koehn, 2009).

We first construct a word-replacement model based on aligning the small amount of target low resource data with that of each source language using `fast_align` (Dyer et al., 2013). We replace every source word with the most probable target word according to the product of the alignment probability and the probability of fertility equalling one and distortion equalling zero $P(F_s = 1, D_t = 0|s_t, s_s, l_s)$. We choose a simple word-replacement model because we aim to work with around 1,000 lines of low resource data. For fast and efficient ranking on such small data, a word-replacement model suits our purpose.

We use two alternatives to create our FAMCs. Our distortion measure is the probability of distortion equalling zero, $P(D_t = 0|s_t, s_s, l_s)$, aggregated over all words in a source language. We use the distortion measure to rank the source languages and obtain the distortion-based FAMC (*FAMD*); we use the translation BLEU scores of the word-replacement model as another alternative to build the performance-based FAMC (*FAMP*). In Table 1, we list the top ten languages in FAMD and FAMP for Eastern Pokomchi and English. We use both alternatives to build FAMCs.

To prepare for transformer training, we choose the top ten languages neighboring our target low resource language in FAMD and FAMP. We choose ten because existing literature shows that training with ten languages from two neighboring language families is sufficient in producing quality translation through cross-lingual transfer (Zhou et al., 2018a). Since for some low resource languages, there may not be ten languages in FAMO in our database, we add languages from neighboring families to make an expanded list denoted by *FAMO*[+].

3.3 Iterative Pretraining

We have two stages of pretraining using multilingual order-preserving lexiconized transformer on the complete and the star configuration. We design iterative pretraining on symmetric data to address catastrophic forgetting that is common in training (French, 1999; Kirkpatrick et al., 2017).

3.3.1 Stage 1: Pretraining on Neighbors

Firstly, we pretrain on the complete graph configuration of translation paths using the top ten languages neighboring our target low resource language in FAMD, FAMP, and FAMO[+] respectively. Low resource data is excluded in training.

We use the multilingual order-preserving lexiconized transformer. Our vocabulary is the combination of the vocabulary for the top ten languages together with the low resource vocabulary built from the ~1,000 lines. The final model can translate from any of the ten languages to each other.

3.3.2 Stage 2: Adding Low Resource Data

We include the low resource data in the second stage of training. Since the low resource data covers $\sim 3.5\%$ of the text while all the source languages cover the whole text, the data is highly asymmetric. To create symmetric data, we align the low resource data with the subset of data from all source languages. As a result, all source languages in the second stage of training have $\sim 3.5\%$ of the text that is aligned with the low resource data.

Source Sentence	IPML Translation	Reference
En terwyl Hy langs die see van Galiléa loop, sien Hy Simon en Andréas, sy broer, besig om 'n net in die see uit te gooi; want hulle was vissers.	And as He drew near to the lake of Galilee, He Simon saw Andrew, and his brother, lying in the lake, for they were fishermen.	And walking along beside the Sea of Galilee, He saw Simon and his brother Andrew casting a small net in the sea; for they were fishers.
En toe Hy daarvandaan 'n bietjie verder gaan, sien Hy Jakobus, die seun van Sebedéüs, en Johannes, sy broer, wat besig was om die nette in die skuit heel te maak.	And being in a distance, He saw James, the son of Zebedee, and John, his brother. who kept the nets in the boat.	And going forward from there a little, He saw James the son of Zebedee, and his brother John. And they were in the boat mending the nets.
En verder Jakobus, die seun van Sebedéüs, en Johannes, die broer van Jakobus- aan hulle het Hy die bynaam Boanérges gegee, dit is, seuns van die donder-	And James the son of Zebedee, and John the brother of James; and He gave to them the name, which is called Boanerges, being of the voice.	And on James the son of Zebedee, and John the brother of James, He put on them the names Boanerges, which is, Sons of Thunder.

Table 2: Examples of Iteratively Pretrained Multilingual Order-preserving Lexiconized Transformer (IPML) translation from Afrikaans to English as a hypothetical low resource language using *FAMP*. We train on only 1,093 lines of English data.

We therefore create a complete graph configuration of training paths using all the eleven languages.

Using the pretrained model from the previous stage, we train on the complete graph configuration of translation paths from all eleven languages including our low resource language. The vocabulary used is the same as before. We employ the multilingual order-preserving lexiconized transformer for pretraining. The final model can translate from any of the eleven languages to each other.

3.4 Final Training

Finally, we focus on translating into the low resource language. We use the symmetric data built from the second stage of pretraining. However, instead of using the complete configuration, we use the star configuration of translation paths from the all source languages to the low resource language. All languages have $\sim 3.5\%$ of the text.

Using the pretrained model from the second stage, we employ the multilingual order-preserving lexiconized transformer on the star graph configuration. We use the same vocabulary as before. The final trained model can translate from any of the ten source languages to the low resource language. Using the lexicon dictionaries, we decode the named entities and obtain our final translations.

3.5 Combination of Translations

We have multiple translations, one per each source language. Combining all translations is useful for both potential post-editing works and systematic comparison of different experiments especially when the sets of the source languages differ.

Our combination method assumes that we have the same text in all source languages. For each sentence, we form a cluster of translations from all source languages into the low resource language. Our goal is to find the translation that is closest to the center of the cluster. We rank all translations according to how centered this translation is with respect to other sentences by summing all its similarities to the rest. The top is closest to the center of the translation cluster. We take the most centered translation for every sentence to build the combined translation output. The expected BLEU score of our combined translation is higher than translation from any of the individual source languages.

4 Data

We use the Bible dataset and the medical EMEA dataset (Mayer and Cysouw, 2014; Tiedemann, 2012). EMEA dataset is from the European Medicines Agency and contains a lot of medical information that may be beneficial to the low resource communities. Our method can be applied to other datasets like WASH guidelines.

For the Bible dataset, we use 124 source languages with 31,103 lines of data and a target low resource language with ~1,000 lines (~3.5%) of data. We have two setups for the target low resource language. One uses Eastern Pokomchi, a Mayan language; the other uses English as a hypothetical low resource language. We train on only ~1,000 lines of low resource data from the book of Luke and test on the 678 lines from the book of Mark. Mark is topically similar to Luke, but is written by a different author. For the first stage of pretraining, we use 80%, 10%, 10% split for training, validation and testing. For the second stage onwards, we use

95%, 5% split of Luke for training and validation, and 100% of Mark for testing.

Eastern Pokomchi is Mayan, and English is Germanic. Since our database does not have ten members of each family, we use FAMO$^+$, the expanded version of FAMO. For English, we include five Germanic languages and five Romance languages in FAMO$^+$; for Eastern Pokomchi, we include five Mayan languages and five Amerindian languages in FAMO$^+$. The Amerindian family is broadly believed to be close to the Mayan family by the linguistic community.

We construct FAMCs by comparing different ways of ranking linguistic distances empirically based on ~1,000 lines of training data. In Table 1, we list the top ten languages for Eastern Pokomchi and English in FAMD and FAMP respectively.

To imitate the real-life situation of having small seed target translation data, we choose to use ~1,000 lines (~3.5%) of low resource data. We also include Eastern Pokomchi in addition to using English as a hypothetical low resource language. Though data size can be constrained to mimic severely low resource scenarios, much implicit information is still used for the hypothetical low resource language that is actually rich resource. For example, implicit information like English is Germanic is often used. For real low resource scenarios, the family information may have yet to be determined; the neighboring languages may be unknown, and if they are known, they are highly likely to be low resource too. We thus use Eastern Pokomchi as our real-life severely low resouce language.

In addition to the Bible dataset, we work with the medical EMEA dataset (Tiedemann, 2012). Using English as a hypothetical language, we train on randomly sampled 1,093 lines of English data, and test on 678 lines of data. Since there are only 9 languages in Germanic and Romance families in EMEA dataset, we include a slavic language Polish in our FAMO$^+$ for experiments.

The EMEA dataset is less than ideal comparing with the Bible dataset. The Bible dataset contains the same text for all source languages; however, the EMEA dataset does not contain the same text. It is built from similar documents but has different parallel data for each language pair. Therefore, during test time, we do not combine the translations from various source languages in the EMEA dataset.

Experiments	IPML	MLc	MLs	PMLc	PMLs	AML
Pretrained	✓			✓	✓	
Iterative	✓					
Lexiconized	✓	✓	✓	✓	✓	✓
Symmetrical	✓	✓	✓	✓	✓	
Star	✓		✓		✓	
Complete	✓	✓		✓		✓
Combined	37.3	13.4	14.7	34.7	35.7	27.0
German	35.0	11.6	12.3	33.3	34.5	25.4
Danish	36.0	12.5	12.4	33.3	34.2	26.2
Dutch	35.6	11.5	11.1	32.3	33.7	25.0
Norwegian	35.7	12.3	12.0	33.2	34.1	25.8
Swedish	34.5	11.8	12.4	32.3	33.4	24.9
Spanish	36.4	11.7	11.8	34.1	35.0	26.2
French	35.3	10.8	10.8	33.1	34.0	25.8
Italian	35.9	11.7	11.7	34.3	34.5	26.1
Portuguese	31.5	9.6	10.1	30.0	30.4	23.1
Romanian	34.6	11.3	12.1	32.3	33.2	25.0

Table 3: Comparing our iteratively pretrained multilingual order-preserving lexiconized transformer (IPML) with the baselines training on 1,093 lines of English data in *FAMO$^+$*. We checkmark the key components used in each experiments and explain all the baselines in details in Section 5.

5 Results

We compare our iteratively pretrained multilingual order-preserving lexiconized transformer (IPML) with five baselines in Table 3. *MLc* is a baseline model of multilingual order-preserving lexiconized transformer training on complete configuration; in other words, we skip the first stage of pretraining and train on the second stage in Chapter 3.3.2 only. *MLs* is a baseline model of multilingual order-preserving lexiconized transformer training on star configuration; in other words, we skip both steps of pretraining and train on the final stage in Chapter 3.4 only. *PMLc* is a baseline model of pretrained multilingual order-preserving lexiconized transformer training on complete configuration; in other words, we skip the final stage of training after completing both stages of pretraining. *PMLs* is a baseline model of pretrained multilingual order-preserving lexiconized transformer training on star configuration; in other words, after the first stage of pretraining, we skip the second stage of pretraining and proceed to the final training directly. Finally, *AML* is a baseline model of multilingual order-preserving lexiconized transformer on asymmetric data. We replicate the ~1,000 lines of the low resource data till it matches the training size of other source languages; we train on the complete graph configuration using eleven languages. Though the number of low resource training lines is the same as others, information is highly asym-

Input Language Family					
By Linguistics		By Distortion		By Performance	
FAMO[+]		*FAMD*		*FAMP*	
Source	BLEU	Source	BLEU	Source	BLEU
Combined	37.3	Combined	38.3	Combined	39.4
German	35.0	German	36.7	German	37.6
Danish	36.0	Danish	37.1	Danish	37.5
Dutch	35.6	Dutch	35.6	Dutch	36.7
Norwegian	35.7	Norwegian	36.9	Norwegian	37.1
Swedish	34.5	Afrikaans	38.3	Afrikaans	39.3
Spanish	36.4	Marshallese	34.7	Spanish	38.4
French	35.3	French	36.0	French	36.6
Italian	35.9	Italian	36.9	Italian	37.7
Portuguese	31.5	Portuguese	32.9	Portuguese	33.1
Romanian	34.6	Frisian	36.1	Frisian	36.9

Table 4: Performance of Iteratively Pretrained Multilingual Order-preserving Lexiconized Transformer (IPML) training for English on *FAMO*[+], *FAMD* and *FAMP*. We train on only 1,093 lines of English data.

Input Language Family					
By Linguistics		By Distortion		By Performance	
FAMO[+]		*FAMD*		*FAMP*	
Source	BLEU	Source	BLEU	Source	BLEU
Combined	23.0	Combined	23.1	Combined	22.2
Chuj	21.8	Chuj	21.9	Chuj	21.6
Cakchiquel	22.2	Cakchiquel	22.1	Cakchiquel	21.3
Guajajara	19.7	Guajajara	19.1	Guajajara	18.8
Mam	22.2	Russian	22.2	Mam	21.7
Kanjobal	21.9	Toba	21.9	Kanjobal	21.4
Cuzco	22.3	Myanmar	19.1	Thai	21.8
Ayacucho	21.6	Slovenský	22.1	Dadibi	19.8
Bolivian	22.2	Latin	21.9	Gumatj	19.1
Huallaga	22.2	Ilokano	22.5	Navajo	21.3
Aymara	21.5	Norwegian	22.6	Kim	21.5

Table 5: Performance of Iteratively Pretrained Multilingual Order-preserving Lexiconized Transformer (IPML) training for Eastern Pokomchi on *FAMO*[+], *FAMD* and *FAMP*. We train on only 1,086 lines of Eastern Pokomchi data.

metric.

Pretraining is key as IPML beats the two baselines that skip pretraining in Table 3. Using English as a hypothetical low resource language training on FAMO[+], combined translation improves from 13.4 (MLc) and 14.7 (MLs) to 37.3 (IPML) with iterative pretraining. Training with the low resource language on both the source and the target sides boosts translation into the target side. Star configuration has a slight advantage over complete configuration as it gives priority to translation into the low resource language. Iterative pretraining with BLEU score 37.3 has an edge over one stage of pretraining with scores 34.7 (PMLc) and 35.7 (PMLs).

All three pretrained models on symmetric data, IPML, PMLc and PMLs, beat asymmetric baseline AML. In Table 3, IPML has a +10.3 BLEU increase over our asymmetric baseline on combined translation using English as a hypothetical low resource language training on FAMO[+]. All four use the same amount of data, but differ in training strategies and data configuration. In severely low resource scenarios, effective training strategies on symmetric data improve translation greatly.

We compare IPML results training on different sets of source languages in FAMO[+], FAMD, and FAMP, for English and Eastern Pokomchi in Table 4 and 5. FAMP performs the best for translation into English while both FAMP and FAMD outperforms FAMO[+] as shown in Table 4. FAMD performs best for translation into Eastern Pokomchi as shown in Table 5. Afrikaans has the highest score for English's FAMD and FAMP, outperform-

Source	BLEU
Combined	N.A.
German	34.8
Danish	37.7
Dutch	39.7
Swedish	37.7
Spanish	42.8
French	41.6
Italian	39.2
Portuguese	42.8
Romanian	40.0
Polish	34.1

Table 6: IPML Performance on the EMEA dataset trained on only 1,093 lines of English data.

Source	BLEU
Combined	31.3
German	29.4
Danish	28.8
Dutch	29.9
Norwegian	29.7
Swedish	29.0
Spanish	30.3
French	28.9
Italian	29.7
Portuguese	24.4
Romanian	28.8

Table 7: IPML Performance on the entire Bible excluding ~1k lines of training and validation data.

ing Dutch, German or French. A reason may be that Afrikaans is the youngest language in the Germanic family with many lexical and syntactic borrowings from English and multiple close neighbors of English (Gordon Jr, 2005). When language family information is limited, constructing FAMC to determine neighbors is very useful in translation.

Comparing Eastern Pokomchi results with English results, we see that translation into real-life severely low resource languages is more difficult than translation into hypothetical ones. The combined score is 38.3 for English in Table 4 and 23.1 for Eastern Pokomchi on FAMD in Table 5. Eastern Pokomchi has ejective consonants which makes tokenization process difficult. It is agglutinative, morphologically rich and ergative just like Basque (Aissen et al., 2017; Clemens et al., 2015). It is complex, unique and nontransparent to the out-

Source Sentence	IPML Translation	Reference
Caso detecte efeitos graves ou outros efeitos não mencionados neste folheto, informe o médico veterinário.	If you notice any side effects or other side effects not mentioned in this leaflet, please inform the vétérinaire.	If you notice any serious effects or other effects not mentioned in this leaflet, please inform your veterinarian.
No tratamento de Bovinos com mais de 250 Kg de peso vivo, dividir a dose de forma a não administrar mais de 10 ml por local de injecção.	In the treatment of infants with more than 250 kg in vivo body weight, a the dose to not exceed 10 ml per injection.	For treatment of cattle over 250 kg body weight, divide the dose so that no more than 10 ml are injected at one site.
No entanto, uma vez que é possível a ocorrência de efeitos secundários, qualquer tratamento que exceda as 1-2 semanas deve ser administrado sob supervisão veterinária regular.	However, because any of side effects is possible, any treatment that 1-5 weeks should be administered under regular supraveghere.	However, since side effects might occur, any treatment exceeding 1–2 weeks should be under regular veterinary supervision.

Table 8: Examples of IPML translation on medical EMEA dataset from Portuguese to English using $FAMO^+$.

sider (England, 2011). Indeed, translation into real severely low resource languages is difficult.

We are curious of how our model trained on ∼1,000 lines of data performs on the rest of the Bible. In other words, we would like to know how IPML performs if we train on ∼3.5% of the Bible and test on ∼96.5% of the Bible. In Table 7, we achieve a BLEU score of 31.3 training IPML on randomly sampled 1,093 lines of data for English on FAMO$^+$. Note that the training data is randomly sampled in Table 7 comparing to training on Luke in Table 4 and Table 5. We use this experiment to show that we have good results not only with specific book, but also with randomly sampled data.

We show qualitative examples in Table 2 and 9. The source content is translated well overall and there are a few places for improvement in Table 2. The words "fishermen" and "fishers" are paraphrases of the same concept. IPML predicts the correct concept though it is penalized by BLEU.

Infusing the order-preserving lexiconized component to our training greatly improves qualitative evaluation. But it does not affect BLEU much as BLEU has its limitations in severely low resource scenarios. This is why all experiments include the lexiconized component in training. The BLEU comparison in our paper also applies to the comparison of all experiments without the order-preserving lexiconized component. This is important in real-life situations when a low resource lexicon list is not available, or has to be invented. For example, a person growing up in a local village in Papua New Guinea may have met many people named "Bosai" or "Kaura", but may have never met a person named "Matthew", and we may need to create a lexicon word in the low resource language for "Matthew" possibly through phonetics.

We also see good results with the medical EMEA dataset. Treating English as a hypothetical low resource language, we train on only 1,093 lines of English data. For Portuguese-English translation, we obtain a BLEU score of 42.8 while the rest of languages all obtain BLEU scores above 34 in Table 6 and Table 8. In Table 8, we see that our translation is very good, though a few words are carried from the source language including "vétérinaire". This is mainly because our ∼1,000 lines contain very small vocabulary; however, by carrying the source word over, key information is preserved.

6 Conclusion

We use ∼1,000 lines of low resource data to translate a closed text that is known in advance to a severely low resource language by leveraging massive source parallelism. We present two metrics to rank the 124 source languages and construct FAMCs. We build an iteratively pretrained multilingual order-preserving lexiconized transformer and combine translations from all source languages into one by using our centric measure. Moreover, we add a multilingual order-preserving lexiconized component to translate the named entities accurately. We build a massively parallel lexicon table for 2,939 Bible named entities in 124 source languages, covering more than 66 severely low resource languages. Our good result for the medical EMEA dataset shows that our method is useful for other datasets and applications.

Our final result can also serve as a ranking measure for linguistic distances though it is much more expensive in terms of time and resources. In the future, we would like to explore more metrics that are fast and efficient in ranking linguistic distances to the severely low resource language.

References

Oliver Adams, Adam Makarucha, Graham Neubig, Steven Bird, and Trevor Cohn. 2017. Cross-lingual word embeddings for low-resource language modeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 937–947.

Ife Adebara, El Moatez Billah Nagoudi, and Muhammad Abdul Mageed. 2020. Translating similar languages: Role of mutual intelligibility in multilingual transformers. *arXiv preprint arXiv:2011.05037*.

Judith Aissen, Nora C England, and Roberto Zavala Maldonado. 2017. *The Mayan languages*. Taylor & Francis.

Antonios Anastasopoulos, Sameer Bansal, David Chiang, Sharon Goldwater, and Adam Lopez. 2017. Spoken term discovery for language documentation using translations. In *Proceedings of the Workshop on Speech-Centric Natural Language Processing*, pages 53–58.

Sameer Bansal, Herman Kamper, Sharon Goldwater, and Adam Lopez. 2017. Weakly supervised spoken term discovery using cross-lingual side information. In *Acoustics, Speech and Signal Processing*, pages 5760–5764. IEEE.

Christos Baziotis, Barry Haddow, and Alexandra Birch. 2020. Language model prior for low-resource neural machine translation. *arXiv preprint arXiv:2004.14928*.

Alexandre Bérard, Zae Myung Kim, Vassilina Nikoulina, Eunjeong Lucy Park, and Matthias Gallé. 2020. A multilingual neural machine translation model for biomedical data. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.

Alexandra Birch, Miles Osborne, and Philipp Koehn. 2008. Predicting success in machine translation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 745–754.

Steven Bird. 2020. Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519.

John S Brownstein, Clark C Freifeld, Ben Y Reis, and Kenneth D Mandl. 2008. Surveillance sans frontieres: Internet-based emerging infectious disease intelligence and the healthmap project. *PLoS medicine*, 5(7):e151.

Jasone Cenoz. 2001. The effect of linguistic distance, l2 status and age on cross-linguistic influence in third language acquisition. *Cross-linguistic influence in 2nd language acquisition: Psycholinguistic perspectives*, 111(45):8–20.

Koel Dutta Chowdhury, Cristina España-Bonet, and Josef van Genabith. 2020. Understanding translationese in multi-view embedding spaces. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6056–6062.

Alexandra Chronopoulou, Dario Stojanovski, and Alexander Fraser. 2020. Reusing a pretrained language model on languages with limited corpora for unsupervised nmt. *arXiv preprint arXiv:2009.07610*.

Lauren Eby Clemens, Jessica Coon, Pedro Mateo Pedro, Adam Milton Morgan, Maria Polinsky, Gabrielle Tandet, and Matthew Wagers. 2015. Ergativity and the complexity of extraction: A view from mayan. *Natural Language & Linguistic Theory*, 33(2):417–467.

Bernard Comrie. 2005. *The world atlas of language structures*. Oxford University Press.

Boele De Raad, Marco Perugini, and Zsófia Szirmák. 1997. In pursuit of a cross-lingual reference structure of personality traits: Comparisons among five languages. *European Journal of Personality*, 11(3):167–185.

Zi-Yi Dou, Antonios Anastasopoulos, and Graham Neubig. 2020. Dynamic data selection and weighting for iterative back-translation. *arXiv preprint arXiv:2004.03672*.

Haim Dubossarsky, Ivan Vulić, Roi Reichart, and Anna Korhonen. 2020. The secret is in the spectra: Predicting cross-lingual task performance with spectral similarity measures. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2377–2390.

Chris Dyer, Victor Chahuneau, and Noah A Smith. 2013. A simple, fast, and effective reparameterization of ibm model 2. In *Proceedings of the 12th Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technologies*, pages 644–648.

Paul S Earle, Daniel C Bowden, and Michelle Guy. 2012. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 54(6).

Nora C England. 2011. *A grammar of Mam, a Mayan language*. University of Texas Press.

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. Multi-way, multilingual neural machine translation with a shared attention mechanism. In *Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technologies*, pages 866–875.

Jerry A Fodor and Zenon W Pylyshyn. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1):3–71.

Markus Freitag and Orhan Firat. 2020. Complete multilingual neural machine translation. *arXiv preprint arXiv:2010.10239*.

Robert M French. 1999. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135.

Dan Gillick, Cliff Brunk, Oriol Vinyals, and Amarnag Subramanya. 2016. Multilingual language processing from bytes. In *Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technologies*, pages 1296–1306.

Raymond G Gordon Jr. 2005. Ethnologue, languages of the world. *http://www. ethnologue. com/*.

Alex Graves, Greg Wayne, and Ivo Danihelka. 2014. Neural turing machines. *arXiv preprint arXiv:1410.5401*.

Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor OK Li. 2018. Universal neural machine translation for extremely low resource languages. *arXiv preprint arXiv:1802.05368*.

Thanh-Le Ha, Jan Niehues, and Alexander Waibel. 2016. Toward multilingual neural machine translation with universal encoder and decoder. *arXiv preprint arXiv:1611.04798*.

Jan Hajič. 2000. Machine translation of very close languages. In *Sixth Applied Natural Language Processing Conference*.

Lynne Hansen, Karri Lam, Livia Orikasa, Paul Rama, Geraldine Schwaller, and Ronald Mellado Miller. 2012. In the beginning was the word. *Second Language Acquisition Abroad: The LDS Missionary Experience*, 45:89.

Theo Hermans. 2003. Cross-cultural translation studies as thick translation. *Bulletin of the School of Oriental and African Studies*, 66(3):380–389.

Eric W Holman, Søren Wichmann, Cecil H Brown, Viveka Velupillai, André Müller, Dik Bakker, et al. 2008. Advances in automated language classification. *Quantitative investigations in theoretical linguistics*, pages 40–43.

Chu-Ren Huang, Laurent Prévot, I-Li Su, and Jia-Fei Hong. 2007. Towards a conceptual core for multicultural processing: A multilingual ontology based on the swadesh list. In *International Workshop on Intercultural Collaboration*, pages 17–30. Springer.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. *arXiv preprint arXiv:2004.09095*.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. Opennmt: Opensource toolkit for neural machine translation. *Proceedings of the 55th annual meeting of the Association for Computational Linguistics, System Demonstrations*, pages 67–72.

Philipp Koehn. 2009. *Statistical machine translation*. Cambridge University Press.

Anne Lauscher, Vinit Ravishankar, Ivan Vulić, and Goran Glavaš. 2020. From zero to hero: On the limitations of zero-shot language transfer with multilingual transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4483–4499.

Zheng Li, Mukul Kumar, William Headden, Bing Yin, Ying Wei, Yu Zhang, and Qiang Yang. 2020. Learn to cross-lingual transfer with meta graph learning across heterogeneous languages. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2290–2301.

Yu-Hsiang Lin, Chian-Yu Chen, Jean Lee, Zirui Li, Yuyan Zhang, Mengzhou Xia, Shruti Rijhwani, Junxian He, Zhisong Zhang, Xuezhe Ma, et al. 2019. Choosing transfer languages for cross-lingual learning. *arXiv preprint arXiv:1905.12688*.

Zehui Lin, Xiao Pan, Mingxuan Wang, Xipeng Qiu, Jiangtao Feng, Hao Zhou, and Lei Li. 2020. Pretraining multilingual neural machine translation by leveraging alignment information. *arXiv preprint arXiv:2010.03142*.

Minh-Thang Luong, Ilya Sutskever, Quoc V Le, Oriol Vinyals, and Wojciech Zaremba. 2014. Addressing the rare word problem in neural machine translation. *arXiv preprint arXiv:1410.8206*.

Thomas Mayer and Michael Cysouw. 2014. Creating a massively parallel bible corpus. *Oceania*, 135(273):40.

Farhad Nooralahzadeh, Giannis Bekoulis, Johannes Bjerva, and Isabelle Augenstein. 2020. Zero-shot cross-lingual transfer with meta learning. *arXiv preprint arXiv:2003.02739*.

Terence Odlin. 1989. *Language transfer: Crosslinguistic influence in language learning*. Cambridge University Press.

Arturo Oncevay, Barry Haddow, and Alexandra Birch. 2020. Bridging linguistic typology and multilingual machine translation with multi-view language representations. *arXiv preprint arXiv:2004.14923*.

Robert Östling and Jörg Tiedemann. 2017. Neural machine translation for low-resource languages. *arXiv preprint arXiv:1708.05729*.

Steve Parker. 2012. Sonority distance vs. sonority dispersion–a typological survey. *The sonority controversy*, 18:101–165.

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. Mad-x: An adapter-based framework for multi-task cross-lingual transfer. *arXiv preprint arXiv:2005.00052*.

Manfred Pienemann, Bruno Di Biase, Satomi Kawaguchi, and Gisela Håkansson. 2005. Processability, typological distance and l1 transfer. *Cross-linguistic aspects of Processability Theory*, pages 85–116.

Ye Qi, Devendra Singh Sachan, Matthieu Felix, Sarguna Janani Padmanabhan, and Graham Neubig. 2018. When and why are pre-trained word embeddings useful for neural machine translation? *arXiv preprint arXiv:1804.06323*.

Taraka Rama and Prasanth Kolachina. 2012. How good are typological distances for determining genealogical relationships among languages? In *Proceedings of COLING 2012: Posters*, pages 975–984.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2015. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*.

Edward Sapir. 1921. How languages influence each other. *Language: an Introduction to the Study of Speech*.

Maurizio Serva and Filippo Petroni. 2008. Indoeuropean languages tree by levenshtein distance. *EPL (Europhysics Letters)*, 81(6):68005.

Philippa Shoemark, Sharon Goldwater, James Kirby, and Rik Sarkar. 2016. Towards robust crosslinguistic comparisons of phonological networks. In *Proceedings of the 14th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 110–120.

Gary F Simons and Charles D Fennig. 2017. *Ethnologue: languages of Asia*. sil International Dallas.

Lucia Specia, Stella Frank, Khalil Sima'an, and Desmond Elliott. 2016. A shared task on multimodal machine translation and crosslingual image description. In *Proceedings of the 1st Conference on Machine Translation*, volume 2, pages 543–553.

Agneta M-L Svalberg and Hjh Fatimah Bte Hj Awg Chuchu. 1998. Are english and malay worlds apart? typological distance and the learning of tense and aspect concepts. *International Journal of Applied Linguistics*, 8(1):27–60.

Brian Thompson, Huda Khayrallah, Antonios Anastasopoulos, Arya D McCarthy, Kevin Duh, Rebecca Marvin, Paul McNamee, Jeremy Gwinnup, Tim Anderson, and Philipp Koehn. 2018. Freezing subnetworks to analyze domain adaptation in neural machine translation. *arXiv preprint arXiv:1809.05218*.

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

Antonio Toral and Andy Way. 2018. What level of quality can neural machine translation attain on literary text? *arXiv preprint arXiv:1801.04962*.

USAID. 2009. How to Wash Hands. https://www.pseau.org/outils/biblio/resume.php?d=3319&l=en. [Online; accessed 23-Nov-2020].

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Alex Waibel and Christian Fugen. 2008. Spoken language translation. *IEEE Signal Processing Magazine*, 25(3):70–79.

Zirui Wang, Zachary C Lipton, and Yulia Tsvetkov. 2020. On negative interference in multilingual models: Findings and a meta-learning treatment. *arXiv preprint arXiv:2010.03017*.

Hao-Ran Wei, Zhirui Zhang, Boxing Chen, and Weihua Luo. 2020. Iterative domain-repaired backtranslation. *arXiv preprint arXiv:2010.02473*.

Winston Wu, Nidhi Vyas, and David Yarowsky. 2018. Creating a translation matrix of the bible's names across 591 languages. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*.

Zhong Zhou, Matthias Sperber, and Alex Waibel. 2018a. Massively parallel cross-lingual learning in low-resource target language translation. In *Proceedings of the 3rd conference on Machine Translation Worshop of the 23rd Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Zhong Zhou, Matthias Sperber, and Alex Waibel. 2018b. Paraphrases as foreign languages in multilingual neural machine translation. *Proceedings of the Student Research Workshop at the 56th Annual Meeting of the Association for Computational Linguistics*.

Changfeng Zhu, Heng Yu, Shanbo Cheng, and Weihua Luo. 2020. Language-aware interlingua for multilingual neural machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1650–1655.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 21st Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575.

7 Appendix

In Table 1 and Table 5, Kanjobal is Eastern Kanjobal, Mam is Northern Mam, Cuzco is Cuzco Quechua, Ayacucho is Ayacucho Quechua, Bolivian is South Bolivian Quechua, and Huallaga is Huallaga Quechua.

We show an illustration of WASH guidelines in Figure 2. We also show IPML translations into Eastern Pokomchi (Mayan) in Table 9.

Figure 2: An Amharic illustration of translation of water, sanitation, and hygiene (WASH) guidelines in Ethiopia (USAID, 2009).

Source Sentence	IPML Translation	Reference
Ket idi limmabas iti dinna ti baybay ti Galilea, nakitana ni Simon ken ni Andres a cabsatna, nga iwaywayatda ti iket iti baybay; ta dumadaclisda idi.	Eh noq ojik i rub'an i Jesús juntar i k'isa palaw i Galilea, xrilow reje i Simón ruch'ihil i Andres, re' i rutuut i k'isa palaw, ruum jinaj i k'isa palaw barco.	Noq k'ahchi' rik'iik i Jesús chi chii' i k'isa palaw ar Galilea, xrilow reje wach i Simón ruch'ihil i ruchaaq', Andres rub'ihnaal. Re' keh aj karineel taqe, k'ahchi' kikutum qohoq i kiya'l pan palaw.
Ket idi nagna pay bassit nakitana ni Santiago nga anac ni Zebedeo ken ni Juan a cabsatna, nga addada idi iti barangayda, a tartarimaanenda dag-iti iketda.	Eh noq ojik i rub'an i Jesús, xrilow i Jacobo, re' i Jacobo rak'uun i Zebedeo, re' Juan rub'ihnaal, ruch'ihil taqe i raj tahqaneel. eh xkikoj wo' wach chinaah i k'isa palaw.	Eh junk'aam-oq chik i xb'ehik reje i Jesús, xrilow kiwach i ki'ib' chi winaq kichaaq' kiib', re' Jacobo, re' Juan, rak'uun taqe i Zebedeo. Eh wilkeeb' chupaam jinaj i barco, k'ahchi' kik'ojem wach i kiya'l b'amb'al kar.
Ket immasideg ni Jesus ket inigga-manna iti imana ket pinatacderna; ket pinanawan ti gorigor , ket nagservi cadacuada.	Eh re' Jesús xujil i koq riib', xutz'a'j i koq chinaah i q'ab'. eh re' i kaq tz'a' chi riij. eh jumehq'iil xwuktik johtoq, re' chik i reh xutoq'aa' cho yej-anik kiwa'.	Eh re' i Jesús xujil i koq riib' ruuk' i yowaab', xuchop chi q'ab', xruksaj johtoq, eh jumehq'iil xik'ik i tz'a' chi riij. Eh re' chik i reh xutoq'aa' cho yej-anik kiwa'.

Table 9: Examples of Iteratively Pretrained Multilingual Order-preserving Lexiconized Transformer (IPML) translation from Ilokano to Eastern Pokomchi using *FAMD*. We train on only 1,086 lines of Eastern Pokomchi data.

80