mPLM-Sim: Better Cross-Lingual Similarity and Transfer in Multilingual Pretrained Language Models

Anonymous ACL submission

Abstract

Recent multilingual pretrained language models (mPLMs) have been shown to encode strong language-specific signals, which are not explicitly provided during pretraining. It remains an open question whether it is feasible to employ mPLMs to measure language similarity, and subsequently use the similarity results to select source languages for boosting cross-lingual transfer. To investigate this, we propose mPLM-Sim, a language similarity measure that induces the similarities across languages from mPLMs using multi-parallel corpora. Our study shows that mPLM-Sim exhibits moderately high correlations with linguistic similarity measures, such as lexicostatistics, genealogical language family, and geographical sprachbund. We also conduct a case study on languages with low correlation and observe that mPLM-Sim yields more accurate similarity results. Additionally, we find that similarity results vary across different mPLMs and different layers within an mPLM. We further investigate whether mPLM-Sim is effective for zero-shot cross-lingual transfer by conducting experiments on both low-level syntactic tasks and high-level semantic tasks. The experimental results demonstrate that mPLM-Sim is capable of selecting better source languages than linguistic measures, resulting in a 1%-2% improvement in zero-shot cross-lingual transfer performance.[1]

1 Introduction

Recent multilingual pretrained language models (mPLMs) trained with massive data, e.g., mBERT (Devlin et al., 2019), XLM-R (Conneau et al., 2020) and BLOOM (Scao et al., 2022), have become a standard for multilingual representation learning. Follow-up works (Wu and Dredze, 2019; Pires et al., 2019; Libovický et al., 2020; Rama et al., 2020; Müller et al., 2021; Liang et al., 2021; Choenni and Shutova, 2022; Chang et al., 2022) show that these mPLMs encode strong language-specific signals which are not explicitly provided during pretraining. However, the possibility of using mPLMs to measure language similarity and utilizing the similarity results to pick source languages for enhancing cross-lingual transfer is not yet thoroughly investigated.

To investigate language similarity in mPLMs, we propose mPLM-Sim, a measure that leverages mPLMs and multi-parallel corpora to measure similarity between languages. Using mPLM-Sim, we intend to answer the following research questions.

(Q1) *What is the correlation between mPLM-Sim and linguistic similarity?*

We compute Pearson correlation between similarity results of mPLM-Sim and linguistic similarity measures. The results show that mPLM-Sim has a moderately high correlation with some linguistic measures, such as lexical-based and language-family-based measures. Additional case studies on languages with low correlation demonstrate that mPLMs can automatically acquire the patterns of similarity among languages through pretraining on massive data.

(Q2) *Do different layers of an mPLM produce different similarity results?*

Jawahar et al. (2019); Sabet et al. (2020); Choenni and Shutova (2022) have demonstrated that different linguistic information is encoded across different layers of an mPLM. We analyze the performance of mPLM-Sim across layers and show that mPLM-Sim results vary across layers, aligning with previous findings. Specifically, the embedding layer captures lexical information, whereas the middle layers reveal more intricate similarity patterns encompassing general, geographical, and syntactic aspects. However, in the high layers, the ability to distinguish between languages becomes less prominent. Furthermore, we observe that clustering of languages also varies by layer, shedding new light on how the representation of language-

[1] We will release our code upon publication.

specific information changes throughout layers.

(Q3) *Do different mPLMs produce different similarity results?*

We make a comprehensive comparison among a diverse set of 11 mPLMs in terms of architecture, modality, model size, and tokenizer. The experimental results show that input modality (text or speech), model size, and data used for pretraining have large effects on mPLM-Sim while tokenizers and training objectives have little effect.

(Q4) *Can mPLM-Sim choose better source languages for zero-shot cross-lingual transfer?*

Previous works (Lin et al., 2019; Pires et al., 2019; Lauscher et al., 2020; Nie et al., 2022; Wang et al., 2023) have shown that the performance of cross-lingual transfer positively correlates with linguistic similarity. However, we find that there can be a mismatch between mPLM subspaces and linguistic clusters, which may lead to a failure of zero-shot cross-lingual transfer for some low-resource languages. Intuitively, mPLM-Sim can select the source languages that boost cross-lingual transfer better than linguistic similarity since it captures the subspaces learned during pretraining (and which are the basis for successful transfer). To examine this, we conduct experiments on four datasets that require reasoning about different levels of syntax and semantics for a diverse set of low-resource languages. The results show that mPLM-Sim achieves 1%-2% improvement over linguistic similarity measures for cross-lingual transfer.

2 Related Work

2.1 Language Typology and Clustering

Similarity between languages can be due to common ancestry in the genealogical language tree, but also influenced by linguistic influence and borrowing (Aikhenvald and Dixon, 2001; Haspelmath, 2004). Linguists have conducted extensive relevant research by constructing high-quality typological, geographical, and phylogenetic databases, including WALS (Dryer and Haspelmath, 2013), Glottolog (Hammarström et al., 2017), Ethnologue (Saggion et al., 2023), and PHOIBLE (Moran et al., 2014; Moran and McCloy, 2019). The lang2vec tool (Littell et al., 2017) further integrates these datasets into multiple linguistic distances. Despite its integration of multiple linguistic measures, lang2vec weights each measure equally, and the quantification of these measures for language similarity computation remains a challenge.

In addition to linguistic measures, some non-lingustic measures are also proposed to measure similarity between languages. Specifically, Holman et al. (2011) use Levenshtein (edit) distance to compute the lexical similarity between languages. Lin et al. (2019) propose dataset-dependent features, which are statistical features specific to the corpus used, e.g., lexical overlap. However, these methods fail to capture deeper similarities beyond surface-level features.

Language representation is another important category of language similarity measures. Tan et al. (2019) represent each language by an embedding vector and cluster them in the embedding space. Fan et al. (2021b) find the representation sprachbund of mPLMs, and then train separate mPLMs for each sprachbund. However, these studies do not delve into the research questions mentioned in §1, and it motivates us to carry out a comprehensive investigation of language similarity using mPLMs.

2.2 Multilingual Pretrained Language Models

The advent of mPLMs, e.g., mBERT (Devlin et al., 2019), XLM (Conneau and Lample, 2019), and XLM-R (Conneau et al., 2020), have brought significant performance gains on numerous multilingual natural language understanding benchmarks (Hu et al., 2020).

Given their success, a variety of following mPLMs are proposed. Specifically, different architectures, including decoder-only, e.g., mGPT (Shliazhko et al., 2022) and BLOOM (Scao et al., 2022), and encoder-decoder, e.g., mT5 (Xue et al., 2021), are designed. Tokenizer-free models, including CANINE (Clark et al., 2022), ByT5 (Xue et al., 2022), and Charformer (Tay et al., 2022), are also proposed. Clark et al. (2022) introduce CANINE-S and CANINE-C. CANINE-S adopts a subword-based loss, while CANINE-C uses a character-based one. Glot500 (Imani et al., 2023) extends XLM-R to cover more than 500 languages using vocabulary extension and continued pretraining. Both InfoXLM (Chi et al., 2021a) and XLM-Align (Chi et al., 2021b) exploit parallel objectives to further improve mPLMs. Some mPLMs are specifically proposed for Machine Translation, e.g., M2M-100 (Fan et al., 2021a) and NLLB-200 (Costa-jussà et al., 2022). XLS-R-300M (Babu et al., 2021) is a speech (as opposed to text) model.

Follow-up works show that strong language-specific signals are encoded in mPLMs by means of probing tasks (Wu and Dredze, 2019; Rama et al.,

Model	Size	|Lang|	|Layer|	Tokenizer	Arch.	Objective	Modality	Data
mBERT (Devlin et al., 2019)	172M	104	13	Subword	Enc	MLM, NSP	Text	Wikipedia
XLM-R-Base (Conneau et al., 2020)	270M	100	13	Subword	Enc	MLM	Text	CC
XLM-R-Large (Conneau et al., 2020)	559M	100	25	Subword	Enc	MLM	Text	CC
Glot500 (Imani et al., 2023)	395M	515	13	Subword	Enc	MLM	Text	Glot500-c
mGPT (Shliazhko et al., 2022)	1.3B	60	25	Subword	Dec	CLM	Text	Wikipedia+mC4
mT5-Base (Xue et al., 2021)	580M	101	13	Subword	Enc-Dec	MLM	Text	mC4
CANINE-S (Clark et al., 2022)	127M	104	17	Char	Enc	MLM, NSP	Text	Wikipedia
CANINE-C (Clark et al., 2022)	127M	104	17	Char	Enc	MLM, NSP	Text	Wikipedia
XLM-Align (Chi et al., 2021b)	270M	94	13	Subword	Enc	MLM, TLM, DWA	Text	Wikipedia+CC
NLLB-200 (Costa-jussà et al., 2022)	1.3B	204	25	Subword	Enc-Dec	MT	Text	NLLB
XLS-R-300M (Babu et al., 2021)	300M	128	25	-	Enc	MSP	Speech	CommonVoice

Table 1: 11 mPLMs considered in the paper. |Layer| denotes the number of layers used for measuring similarity. Both the static embedding layer and all layers of the transformer are considered. For encoder-decoder architectures, we only consider the encoder. |Lang|: the number of languages covered. Arch.: Architecture. Enc: Encoder. Dec: Decoder. MLM: Masked Language Modeling. CLM: Causal Language Modeling. TLM: Translation Language Modeling. NSP: Next Sentence Prediction. DWA: Denoising Word Alignment. MT: Machine Translation. MSP: Masked Speech Prediction. CC: CommonCrawl.

2020; Pires et al., 2019; Müller et al., 2021; Liang et al., 2021; Choenni and Shutova, 2022) and investigating the geometry of mPLMs (Libovický et al., 2020; Chang et al., 2022). This inspires us to explore the potential of mPLMs for measuring language similarity, as well as to investigate how the resulting similarity measures can be used to analyze and exploit mPLMs in turn.

3 mPLM-Sim

Generally, a transformer-based mPLM consists of N layers: $N-1$ transformer layers plus the static embedding layer. Given a multi-parallel corpus[2], mPLM-Sim aims to provide the similarity results of N layers for an mPLM across L languages considered. In this context, we define languages using the ISO 639-3 code combined with the script, e.g., "eng_Latn" represents English written in Latin.

For each sentence x in the multi-parallel corpus, the mPLM computes its sentence embedding for the ith layer of the mPLM: $\boldsymbol{h}_i = E(x)$. For mPLMs with bidirectional encoders, including encoder architecture, e.g., XLM-R, and encoder-decoder architecture, e.g., mT5, $E(\cdot)$ is a mean pooling operation over hidden states, which performs better than [CLS] and MAX strategies (Reimers and Gurevych, 2019). For mPLMs with auto-regressive encoders, e.g., mGPT, $E(\cdot)$ is a position-weighted mean pooling method, which gives later tokens a higher weight (Muennighoff, 2022). Finally, sentence embeddings for all sentences of the L languages are obtained.

For ith layer, the similarity of each language pair is computed using the sentence embeddings of all multi-parallel sentences. Specifically, we get the cosine similarity of each parallel sentence of the language pair, and then average all similarity scores across sentences as the final score of the pair. Finally, we have a similarity matrix $\boldsymbol{S}_i \in \mathbb{R}^{L\times L}$ across L languages for the ith layer of the mPLM.

4 Setup

4.1 mPLMs, Corpora and Languages

We consider a varied set of 11 mPLMs for our investigation, differing in model size, number of covered languages, architecture, modality, and data used for pretraining. Full list and detailed information of the selected mPLMs are shown in Tab. 1.

We work with three multi-parallel corpora: the text corpora Flores (Costa-jussà et al., 2022) and Parallel Bible Corpus (PBC, (Mayer and Cysouw, 2014)) and the speech corpus Fleurs (Conneau et al., 2022). Flores covers more than 200 languages. Since both PBC and Fleurs are not fully multi-parallel, we reconstruct them to make them multi-parallel. After recostruction, PBC covers 379 languages, while Fleurs covers 67 languages. PBC consists of religious text, and both Flores and Fleurs are from web articles. The speech of Fleurs is aligned to the text of Flores, enabling us to compare text mPLMs with speech mPLMs. We use 500 multi-parallel sentences from each corpus. Languages covered by mPLMs and corpora are listed in §A.

[2] Monolingual corpora covering multiple languages can be also used to measure language similarity. Our initial experiments (§B.1) show that parallel corpora yield better results while using fewer sentences than monolingual corpora. Therefore, we use parallel corpora for our investigation.

Task	Corpus	|Train|	|Dev|	|Test|	|Lang|	Metric	Domain
Sequence	NER (Pan et al., 2017)	5,000	500	100-10,000	108	F1	Wikipedia
Labeling	POS (de Marneffe et al., 2021)	5,000	500	100-22,358	60	F1	Misc
Text	MASSIVE (FitzGerald et al., 2022)	11,514	2,033	2,974	44	Acc	Misc
Classification	Taxi1500 (Ma et al., 2023)	860	106	111	130	F1	Bible

Table 2: Evaluation dataset statistics. |Train|/|Dev|: train/dev set size (source language). |Test|: test set size (target language). |Lang|: number of target languages.

4.2 Evaluation

Pearson Correlation We compute Pearson correlation scores to measure how much mPLM-Sim correlates with seven linguistic similarity measures: LEX, GEN, GEO, SYN, INV, PHO and FEA. LEX is computed based on the edit distance of the two corpora. The six others are provided by lang2vec. GEN is based on language family. GEO is orthodromic distance, i.e., the shortest distance between two points on the surface of the earth. SYN is derived from the syntactic structures of the languages. Both INV and PHO are phonological features. INV is derived from PHOIBLE, while PHO is based on WALS and Ethnologue. FEA is computed by combining GEN, GEO, SYN, INV and PHO.

For each target language, we have the similarity scores between the target language and the other $L-1$ languages based on the similarity matrix $\boldsymbol{S}_i$ for layer i (see §3), and also the similarity scores based on the considered linguistic similarity measure j. Then we compute the Pearson correlation r_i^j between these two similarity score lists. We choose the highest correlation score across all layers as the result of each target language since the results for different languages vary across layers. Finally, we report MEAN (M) and MEDIAN (Mdn) of the correlation scores for all languages. Here, we consider 32 languages covered by all models and corpora.

Case Study In addition to the quantitative evaluation, we conduct manual analysis for languages that exhibit low correlation scores. We apply complete linkage hierarchical clustering to get the similar languages of the analyzed language for more detailed analysis. Specifically, the languages which have the most common shared path in the hierarchical tree with the target language are considered as similar languages. To analyze as many languages as possible, we consider the setting of Glot500 and PBC for manual analysis.

Cross-Lingual Transfer To compare mPLM-Sim with linguistic measures for zero-shot cross-lingual transfer, we run experiments for low-resource languages on four datasets, including two for sequence labeling, and two for text classification. Details of the four tasks are shown in Tab. 2.

We selected six high-resource and typologically diverse languages, namely Arabic (arb_Arab), Chinese (cmn_Hani), English (eng_Latn), Hindi (hin_Deva), Russian (rus_Cyrl), and Spanish (spa_Latn), as source languages. For a fair comparison, we use the same amount of source language data for fine-tuning and validation as shown in Tab. 2.

The evaluation targets all languages that are covered by both Glot500 and Flores and have at least 100 samples, excluding the six source languages. The language list for evaluation is provided in §A.

We obtain the most similar source language for each target language by applying each of the seven linguistic similarity measures (LEX, GEN, GEO, SYN, INV, PHO, FEA) and our mPLM-Sim. Here, we consider the setting of Glot500 and Flores for mPLM-Sim since extensive experiments (see §B.2) show that Flores provides slightly better similarity results than PBC. For the linguistic similarity measures, if the most similar source language is not available due to missing values in lang2vec, we use eng_Latn as the source language. We also compare mPLM-Sim with the ENG baseline defined as using eng_Latn as the source language for all target languages.

We use the same hyper-parameter settings as in (Hu et al., 2020; FitzGerald et al., 2022; Ma et al., 2023). Specifically, we set the batch size to 32 and the learning rate to 2e-5 for both NER and POS, and fine-tune Glot500 for 10 epochs. For MASSIVE, we use a batch size of 16, a learning rate of 4.7e-6, and train for 100 epochs. For Taxi1500, we use a batch size of 32, a learning rate of 2e-5, and train for 30 epochs. In all tasks, we select the model for evaluating target languages based on the performance of the source language validation set.

	XLM-R-Base		XLM-R-Large		mT5-Base		mGPT		mBERT		Glot500	
	M	Mdn	M	Mdn	M	Mdn	M	Mdn	M	Mdn	M	Mdn
LEX	0.740	0.859	0.684	0.862	0.628	0.796	0.646	0.848	0.684	0.882	0.741	0.864
GEN	0.489	0.563	0.570	0.609	0.577	0.635	0.415	0.446	0.513	0.593	0.527	0.600
GEO	0.560	0.656	0.587	0.684	0.528	0.586	0.348	0.362	0.458	0.535	0.608	0.674
SYN	0.637	0.662	0.709	0.738	0.594	0.612	0.548	0.591	0.611	0.632	0.577	0.607
INV	0.272	0.315	0.312	0.292	0.295	0.321	0.340	0.394	0.216	0.246	0.248	0.293
PHO	0.112	0.151	0.207	0.258	0.166	0.176	0.184	0.239	0.111	0.125	0.094	0.144
FEA	0.378	0.408	0.443	0.466	0.354	0.371	0.455	0.479	0.346	0.361	0.358	0.372
AVG	0.455	0.516	0.502	0.559	0.449	0.500	0.420	0.480	0.420	0.482	0.451	0.508

	CANINE-S		CANINE-C		NLLB-200		XLM-Align		XLS-R-300M		AVG	
	M	Mdn	M	Mdn	M	Mdn	M	Mdn	M	Mdn	M	Mdn
LEX	0.661	0.821	0.639	0.784	0.722	0.856	0.728	0.869	0.285	0.262	0.651	0.791
GEN	0.548	0.629	0.565	0.633	0.538	0.626	0.516	0.606	0.401	0.353	0.514	0.572
GEO	0.504	0.560	0.533	0.624	0.490	0.499	0.616	0.690	0.531	0.541	0.524	0.583
SYN	0.476	0.521	0.507	0.559	0.375	0.370	0.634	0.669	0.354	0.389	0.548	0.577
INV	0.329	0.390	0.369	0.406	0.337	0.373	0.252	0.315	0.191	0.180	0.287	0.321
PHO	0.112	0.137	0.117	0.173	0.101	0.108	0.105	0.143	0.124	0.115	0.130	0.161
FEA	0.317	0.297	0.367	0.360	0.311	0.326	0.368	0.399	0.203	0.175	0.355	0.365
AVG	0.421	0.479	0.442	0.506	0.411	0.451	0.460	0.527	0.298	0.288	0.430	0.481

Table 3: Comparison across mPLMs: Pearson correlation between mPLM-Sim and seven similarity measures for all mPLMs and Flores/Fleurs on 32 languages. mPLM-Sim strongly correlates with LEX, moderate strongly correlates with GEN, GEO, and SYN, and weakly correlates with INV, PHO, and FEA.

5 Results

5.1 Comparison Between mPLM-Sim and Linguistic Similarity

Tab. 3 shows the Pearson correlation between mPLM-Sim and linguistic similarity measures of 11 mPLMs, and also the average correlations of all 11 mPLMs. We observe that mPLM-Sim strongly correlates with LEX, which is expected since mPLMs learn language relationships from data and LEX similarity is the easiest pattern to learn. Besides, mPLM-Sim has moderately strong correlations with GEN, GEO, and SYN, which shows that mPLMs can learn high-level patterns for language similarity. mPLM-Sim also has a weak correlation with INV, and a very weak correlation with PHO, indicating mPLMs do not capture phonological similarity well. Finally, mPLM-Sim correlates with FEA weakly since FEA is the measure combining both high- and low-correlated linguistics features.

To further compare mPLM-Sim with linguistic similarity measures, we conduct a manual analysis on languages for which mPLM-Sim has weak correlations with LEX, GEN, and GEO. As mentioned in §4, with the setting of Glot500 and PBC, we apply hierarchical clustering and use similar results for analysis.

We find that mPLM-Sim can deal well with languages that are not covered by lang2vec. For example, Norwegian Nynorsk (nno_Latn) is not covered by lang2vec, and mPLM-Sim can correctly find its similar languages, i.e., Norwegian Bokmål (nob_Latn) and Norwegian (nor_Latn). Furthermore, mPLM-Sim can well capture the similarity between languages which cannot be well measured by either LEX, GEN, or GEO.

For LEX, mPLM-Sim can capture similar languages written in different scripts. A special case is the same languages in different scripts. Specifically, mPLM-Sim matches Uighur in Latin and Arabic (uig_Arab and uig_Latn), also Karakalpak in Latin and Cyrillic (kaa_Latn and kaa_Cyrl). In general, mPLM-Sim does a good job at clustering languages from the same language family but written in different scripts, e.g., Turkic (Latn, Cyrl, Arab) and Slavic (Latn, Cyrl).

For GEN, mPLM-Sim captures correct similar languages for isolates and constructed languages. Papantla Totonac (top_Latn) is a language of the Totonacan language family and spoken in Mexico. It shares areal features with the Nahuan languages (nch_Latn, ncj_Latn, and ngu_Latn) of the Uto-Aztecan family, which are all located in the Mesoamerican language area.[3] Esperanto (epo_Latn) is a constructed language whose vocabulary derives primarily from Romance languages, and mPLM-Sim correctly identifies Romance languages such as French (fra_Latn) and Italian (ita_Latn) as similar. The above two cases show the superiority of mPLM-Sim compared to GEN.

The GEO measure may not be suitable for certain language families, such as Austronesian languages and mixed languages. Austronesian languages have the largest geographical span among language families prior to the spread of Indo-European during the colonial period.[4] Moreover, for mixed languages, such as creole languages, their similar languages are often geographically distant due to colonial history. In contrast to GEO, mPLM-Sim can better cluster these languages.

The above analysis shows that it is non-trivial to use either LEX, GEN, or GEO for measuring language similarity. In contrast, mPLM-Sim directly captures similarity from mPLMs and can therefore produce better similarity results

However, we observe that obtaining accurate similarity results from mPLMs using mPLM-Sim can be challenging for certain languages. To gain further insights into this issue, we examine the correlation between performances, specifically the correlation between mPLM-Sim and GEN, and the sizes of the pretraining data. Surprisingly, we find a remarkably weak correlation (-0.008), suggesting that differences in pretraining data sizes do not significantly contribute to variations in performances.

Instead, our findings indicate a different key factor: the coverage of multiple languages within the same language family. This observation is substantiated by a strong correlation of 0.617 between the diversity of languages within a language family (measured by the number of languages included) and the performance of languages belonging to that particular language family.

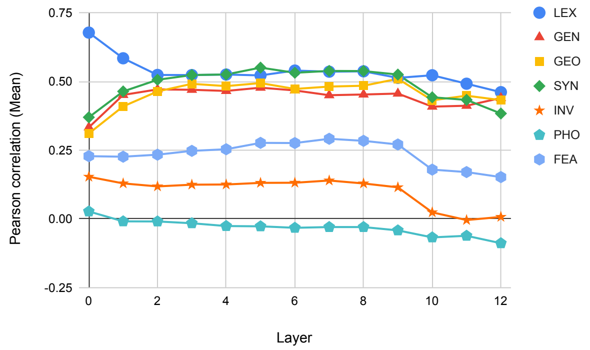


Figure 1: Comparison across layers: Pearson correlation (MEAN) between mPLM-Sim and linguistic similarity measures across layers for Glot500 and Flores on 32 languages. Correlation between mPLM-Sim and LEX peaks in the first layer and decreases, while the correlation with GEN, GEO, and SYN slightly increases in the low layers before reaching its peak.

5.2 Comparison Across Layers for mPLM-Sim

We analyze the correlation between mPLM-Sim and linguistic similarity measures across different layers of an mPLM, specifically for Glot500. The results, presented in Fig. 1, demonstrate the variation in mPLM-Sim results across layers. Notably, in the first layer, mPLM-Sim exhibits a high correlation with LEX, which gradually decreases as we move to higher layers. Conversely, the correlation between mPLM-Sim and GEN, GEO, and SYN shows a slight increase in the lower layers, reaching its peak in layer 1 or 2 of the mPLM. However, for the higher layers (layers 10-12), all correlations slightly decrease. We also performed further visualization and analysis across layers using the setting of Glot500 and Flores for mPLM-Sim (§C). The findings are consistent with our observations from Fig. 1.

Furthermore, our case study shows that the layers which have highest correlations between mPLM-Sim and LEX, GEN, or GEO vary across languages. For example, Atlantic–Congo languages achieve highest correlation with GEN at the 1st layer, while Mayan languages at the 6th layer. This finding demonstrates that language-specific information changes across layers.

[3] https://en.wikipedia.org/wiki/Mesoamerican_language_area

[4] https://en.wikipedia.org/wiki/Austronesian_languages

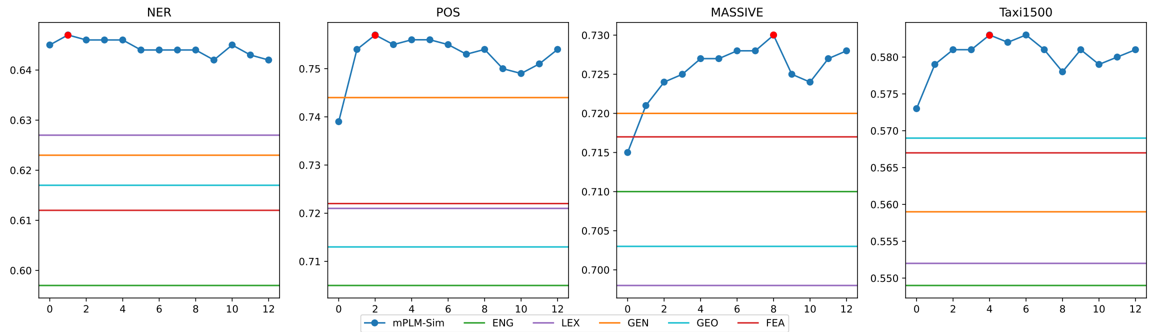


Figure 2: Macro average results (averaged over target languages) on cross-lingual transfer for baselines and for mPLM-Sim in all layers of Glot500. ENG represents using English as the source language. LEX, GEN, GEO, and FEA indicate using the most similar languages based on the corresponding similarity measures as the source language. The red dots of mPLM-Sim highlight the layer with the highest score.

5.3 Comparison Across Models for mPLM-Sim

Tab. 3 presents a broad comparison among 11 different mPLMs, revealing several key findings.

Firstly, the decoder architecture has a negative impact on performance due to the inherent difficulty in obtaining accurate sentence-level representations from the decoder. For example, the decoder-only mPLM mGPT performs worse than encoder-only mPLMs such as XLM-R and mBERT. This observation is reinforced by the comparison between XLM-R-Large and mT5-Base, which have nearly identical model sizes. Remarkably, XLM-R-Large outperforms mT5-Base on AVG by 5% for both Mean (M) and Median (Mdn) scores.

Additionally, tokenizer-free mPLMs achieve comparable performance to subword-tokenizer-based mPLMs. Notably, mPLMs such as mBERT, CANINE-S, and CANINE-C, which share pretraining settings, exhibit similar performances.

The size of mPLMs also influences mPLM-Sim in terms of LEX, GEN, and SYN. Comparing XLM-R-Base with XLM-R-Large, higher-level language similarity patterns are more evident in larger mPLMs. Specifically, XLM-R-Large shows a higher correlation with high-level patterns such as GEN and SYN, while having a lower correlation with low-level patterns like LEX, compared to XLM-R-Base.

The training objectives adopted in mPLMs also impact the performance of mPLM-Sim. Task-specific mPLMs, such as NLLB-200, perform slightly worse than general-purpose mPLMs. Besides, XLM-Align, which leverages parallel objectives to align representations across languages, achieves comparable results to XLM-R-Base. This highlights the importance of advancing methods to effectively leverage parallel corpora.

The choice of pretraining data is another important factor. For example, mBERT uses Wikipedia, while XLM-R-Base uses CommonCrawl, which contains more code-switching. As a result, XLM-R-Base has a higher correlation with GEO and achieves higher AVG compared to mBERT.

Speech mPLMs exhibit lower correlation than text mPLMs, consistent with findings from Abdullah et al. (2023). Speech mPLMs learn language similarity from speech data, which is biased towards the accents of speakers. Consequently, speech mPLMs have a higher correlation with GEO, which is more related to accents, than other similarity measures.

Factors such as the number of languages have minimal effects on mPLM-Sim. Glot500, covering over 500 languages, achieves comparable results with XLM-R-Base.

5.4 Effect for Cross-Lingual Transfer

The macro average results of cross-lingual transfer across target languages for both mPLM-Sim and baselines are presented in Fig. 2. Among the evaluated tasks, ENG exhibits the worst performance in three out of four tasks, emphasizing the importance of considering language similarity when selecting source languages for cross-lingual transfer. mPLM-Sim surpasses all linguistic similarity measures in every task, including both syntactic and semantic tasks, across all layers except layer 0. This indicates that mPLM-Sim is more effective in selecting source languages that enhance the performance of target languages compared to linguistic similarity measures.

		Language	GEN		mPLM-Sim		Δ		Language	GEN		mPLM-Sim		Δ
high end	NER	jpn_Jpan	0.177	eng_Latn	0.451	cmn_Hani	0.275	POS	jpn_Jpan	0.165	eng_Latn	0.534	cmn_Hani	0.369
		kir_Cyrl	0.391	eng_Latn	0.564	rus_Cyrl	0.173		mlt_Latn	0.603	arb_Arab	0.798	spa_Latn	0.196
		mya_Mymr	0.455	cmn_Hani	0.607	hin_Deva	0.153		wol_Latn	0.606	eng_Latn	0.679	spa_Latn	0.074
low end		pes_Arab	0.653	hin_Deva	0.606	arb_Arab	-0.047		ekk_Latn	0.815	eng_Latn	0.790	rus_Cyrl	-0.025
		tgl_Latn	0.745	eng_Latn	0.667	spa_Latn	-0.078		bam_Latn	0.451	eng_Latn	0.411	spa_Latn	-0.039
		sun_Latn	0.577	eng_Latn	0.490	spa_Latn	-0.087		gla_Latn	0.588	rus_Cyrl	0.548	spa_Latn	-0.040
high end	MASSIVE	mya_Mymr	0.616	cmn_Hani	0.707	hin_Deva	0.091	Taxi1500	tgk_Cyrl	0.493	hin_Deva	0.724	rus_Cyrl	0.231
		amh_Ethi	0.532	arb_Arab	0.611	hin_Deva	0.079		kin_Latn	0.431	eng_Latn	0.619	spa_Latn	0.188
		jpn_Jpan	0.384	eng_Latn	0.448	cmn_Hani	0.064		kik_Latn	0.384	eng_Latn	0.555	spa_Latn	0.172
low end		cym_Latn	0.495	rus_Cyrl	0.480	spa_Latn	-0.015		ckb_Arab	0.622	hin_Deva	0.539	arb_Arab	-0.083
		tgl_Latn	0.752	eng_Latn	0.723	spa_Latn	-0.028		nld_Latn	0.713	eng_Latn	0.628	spa_Latn	-0.085
		deu_Latn	0.759	eng_Latn	0.726	spa_Latn	-0.033		kac_Latn	0.580	cmn_Hani	0.483	hin_Deva	-0.097

Table 4: Results for three languages each with the largest (high end) and smallest (low end) gains from mPLM-Sim vs. GEN for four tasks. mPLM-Sim's gain over GEN is large at the high end and smaller negative at the low end. We report both the selected source languages and the results on the evaluated target languages. For mPLM-Sim, the results are derived from the layers exhibiting the best performances as shown in Fig. 2. See §E for detailed results for each task and each target language.

For low-level syntactic tasks, the lower layers (layer 1 or 2) exhibit superior performance compared to all other layers. Conversely, for high-level semantic tasks, it is the middle layer of the mPLM that consistently achieves the highest results across all layers. This can be attributed to its ability to capture intricate similarity patterns.

In Tab. 4, we further explore the benefits of mPLM-Sim in cross-lingual transfer. We present a comprehensive analysis of the top 3 performance improvements and declines across languages. We compare mPLM-Sim and GEN across four cross-lingual transfer tasks. By examining these results, we gain deeper insights into the advantages of mPLM-Sim in facilitating effective cross-lingual transfer.

The results clearly demonstrate that mPLM-Sim has a substantial performance advantage over GEN for certain target languages. On one hand, for languages without any source language in the same language family, such as Japanese (jpn_Jpan), mPLM-Sim successfully identifies its similar language, Chinese (cmn_Hani), whereas GEN fails to do so. Notably, in the case of Japanese, mPLM-Sim outperforms GEN by 27.5% for NER, 36.9% for POS, and 6.4% for MASSIVE.

On the other hand, for languages having source languages within the same language family, mPLM-Sim accurately detects the appropriate source language, leading to improved cross-lingual transfer performance. In the case of Burmese (mya_Mymr), mPLM-Sim accurately identifies Hindi (hin_Deva) as the source language, while GEN mistakenly selects Chinese (cmn_Hani). This distinction results in a significant performance improvement of 15.3% for NER and 9.1% for MASSIVE.

However, we also observe that mPLM-Sim falls short for certain languages when compared to GEN, although the losses are smaller in magnitude compared to the improvements. This finding suggests that achieving better performance in cross-lingual transfer is not solely dependent on language similarity. As mentioned in previous studies such as Lauscher et al. (2020) and Nie et al. (2022), the size of the pretraining data for the source languages also plays a crucial role in cross-lingual transfer.

6 Conclusion

In this paper, we introduce mPLM-Sim, a novel approach for measuring language similarities. Extensive experiments substantiate the superior performance of mPLM-Sim compared to linguistic similarity measures. Our study reveals variations in similarity results across different mPLMs and layers within an mPLM. Furthermore, our findings reveal that mPLM-Sim effectively identifies the source language to enhance cross-lingual transfer.

The results obtained from mPLM-Sim have significant implications for multilinguality. On the one hand, it can be further used in linguistic study and downstream applications, such as cross-lingual transfer, as elaborated in the paper. On the other hand, these findings provide valuable insights for improving mPLMs, offering opportunities for their further development and enhancement.

Limitations

(1) The performance of mPLM-Sim may be strongly influenced by the quality and quantity of data used for training mPLMs, as well as the degree to which the target language can be accurately represented. (2) The success of mPLM-Sim depends on the supporting languages of mPLMs. We conduct further experiment and analysis at §D. (3) As for §5.3, we are unable to conduct a strictly fair comparison due to the varying settings in which mPLMs are pretrained, including the use of different corpora and model sizes.

References

Badr M Abdullah, Mohammed Maqsood Shaik, and Dietrich Klakow. 2023. On the nature of discrete speech representations in multilingual self-supervised models. In *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 159–161.

Alexandra Y. Aikhenvald and R. M. W. Dixon. 2001. *Areal diffusion and genetic inheritance*. Oxford University Press, Oxford.

Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. 2021. XLS-R: self-supervised cross-lingual speech representation learning at scale. *CoRR*, abs/2111.09296.

Tyler A. Chang, Zhuowen Tu, and Benjamin K. Bergen. 2022. The geometry of multilingual language model representations. *CoRR*, abs/2205.10964.

Zewen Chi, Li Dong, Furu Wei, Nan Yang, Saksham Singhal, Wenhui Wang, Xia Song, Xian-Ling Mao, Heyan Huang, and Ming Zhou. 2021a. Infoxlm: An information-theoretic framework for cross-lingual language model pre-training. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 3576–3588. Association for Computational Linguistics.

Zewen Chi, Li Dong, Bo Zheng, Shaohan Huang, Xian-Ling Mao, Heyan Huang, and Furu Wei. 2021b. Improving pretrained cross-lingual language models via self-labeled word alignment. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 3418–3430. Association for Computational Linguistics.

Rochelle Choenni and Ekaterina Shutova. 2022. Investigating language relationships in multilingual sentence encoders through the lens of linguistic typology. *Comput. Linguistics*, 48(3):635–672.

Jonathan H. Clark, Dan Garrette, Iulia Turc, and John Wieting. 2022. Canine: Pre-training an efficient tokenization-free encoder for language representation. *Trans. Assoc. Comput. Linguistics*, 10:73–91.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 8440–8451. Association for Computational Linguistics.

Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 7057–7067.

Alexis Conneau, Min Ma, Simran Khanuja, Yu Zhang, Vera Axelrod, Siddharth Dalmia, Jason Riesa, Clara Rivera, and Ankur Bapna. 2022. FLEURS: few-shot learning evaluation of universal representations of speech. *CoRR*, abs/2205.12446.

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, Loïc Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation. *CoRR*, abs/2207.04672.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal dependencies. *Comput. Linguistics*, 47(2):255–308.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Matthew S Dryer and Martin Haspelmath. 2013. The world atlas of language structures online.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Michael Auli, and Armand Joulin. 2021a. Beyond english-centric multilingual machine translation. *J. Mach. Learn. Res.*, 22:107:1–107:48.

Yimin Fan, Yaobo Liang, Alexandre Muzio, Hany Hassan, Houqiang Li, Ming Zhou, and Nan Duan. 2021b. Discovering representation sprachbund for multilingual pre-training. In *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16-20 November, 2021*, pages 881–894. Association for Computational Linguistics.

Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, Swetha Ranganath, Laurie Crist, Misha Britan, Wouter Leeuwis, Gökhan Tür, and Prem Natarajan. 2022. MASSIVE: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages. *CoRR*, abs/2204.08582.

Harald Hammarström, Robert Forkel, and Martin Haspelmath. 2017. Glottolog 3.0. *Max Planck Institute for the Science of Human History*.

Martin Haspelmath. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? *Studies in Language*, 28(1):209–223.

Eric W Holman, Cecil H Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, et al. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52(6):841–875.

Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. 2020. XTREME: A massively multilingual multitask benchmark for evaluating cross-lingual generalisation. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 4411–4421. PMLR.

Ayyoob Imani, Peiqin Lin, Amir Hossein Kargaran, Silvia Severini, Masoud Jalili Sabet, Nora Kassner, Chunlan Ma, Helmut Schmid, André F. T. Martins, François Yvon, and Hinrich Schütze. 2023. Glot500: Scaling multilingual corpora and language models to 500 languages.

Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What does BERT learn about the structure of language? In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 3651–3657. Association for Computational Linguistics.

Anne Lauscher, Vinit Ravishankar, Ivan Vulic, and Goran Glavas. 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 2020, Online, November 16-20, 2020*, pages 4483–4499. Association for Computational Linguistics.

Sheng Liang, Philipp Dufter, and Hinrich Schütze. 2021. Locating language-specific information in contextualized embeddings. *CoRR*, abs/2109.08040.

Jindrich Libovický, Rudolf Rosa, and Alexander Fraser. 2020. On the language neutrality of pre-trained multilingual representations. In *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*, volume EMNLP 2020 of *Findings of ACL*, pages 1663–1674. Association for Computational Linguistics.

Yu-Hsiang Lin, Chian-Yu Chen, Jean Lee, Zirui Li, Yuyan Zhang, Mengzhou Xia, Shruti Rijhwani, Junxian He, Zhisong Zhang, Xuezhe Ma, Antonios Anastasopoulos, Patrick Littell, and Graham Neubig. 2019. Choosing transfer languages for cross-lingual learning. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 3125–3135. Association for Computational Linguistics.

Patrick Littell, David R. Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori S. 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, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 2: Short Papers*, pages 8–14. Association for Computational Linguistics.

Chunlan Ma, Ayyoob ImaniGooghari, Haotian Ye, Ehsaneddin Asgari, and Hinrich Schütze. 2023. Taxi1500: A multilingual dataset for text classification in 1500 languages.

Thomas Mayer and Michael Cysouw. 2014. Creating a massively parallel bible corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*, pages 3158–3163. European Language Resources Association (ELRA).

Steven Moran and Daniel McCloy, editors. 2019. *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.

Steven Moran, Daniel McCloy, and Richard Wright. 2014. Phoible online.

Niklas Muennighoff. 2022. SGPT: GPT sentence embeddings for semantic search. *CoRR*, abs/2202.08904.

Benjamin Müller, Yanai Elazar, Benoît Sagot, and Djamé Seddah. 2021. First align, then predict: Understanding the cross-lingual ability of multilingual BERT. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 2214–2231. Association for Computational Linguistics.

Ercong Nie, Sheng Liang, Helmut Schmid, and Hinrich Schütze. 2022. Cross-lingual retrieval augmented prompt for low-resource languages. *CoRR*, abs/2212.09651.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1946–1958. Association for Computational Linguistics.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4996–5001. Association for Computational Linguistics.

Taraka Rama, Lisa Beinborn, and Steffen Eger. 2020. Probing multilingual BERT for genetic and typological signals. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 1214–1228. International Committee on Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 3980–3990. Association for Computational Linguistics.

Masoud Jalili Sabet, Philipp Dufter, François Yvon, and Hinrich Schütze. 2020. Simalign: High quality word alignments without parallel training data using static and contextualized embeddings. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP 2020, Online Event, 16-20 November 2020*, volume EMNLP 2020 of *Findings of ACL*, pages 1627–1643. Association for Computational Linguistics.

Horacio Saggion, Sanja Štajner, Daniel Ferrés, Kim Cheng Sheang, Matthew Shardlow, Kai North, and Marcos Zampieri. 2023. Findings of the tsar-2022 shared task on multilingual lexical simplification. *arXiv preprint arXiv:2302.02888*.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilic, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, Jonathan Tow, Alexander M. Rush, Stella Biderman, Albert Webson, Pawan Sasanka Ammanamanchi, Thomas Wang, Benoît Sagot, Niklas Muennighoff, Albert Villanova del Moral, Olatunji Ruwase, Rachel Bawden, Stas Bekman, Angelina McMillan-Major, Iz Beltagy, Huu Nguyen, Lucile Saulnier, Samson Tan, Pedro Ortiz Suarez, Victor Sanh, Hugo Laurençon, Yacine Jernite, Julien Launay, Margaret Mitchell, Colin Raffel, Aaron Gokaslan, Adi Simhi, Aitor Soroa, Alham Fikri Aji, Amit Alfassy, Anna Rogers, Ariel Kreisberg Nitzav, Canwen Xu, Chenghao Mou, Chris Emezue, Christopher Klamm, Colin Leong, Daniel van Strien, David Ifeoluwa Adelani, and et al. 2022. BLOOM: A 176b-parameter open-access multilingual language model. *CoRR*, abs/2211.05100.

Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. 2022. mgpt: Few-shot learners go multilingual. *CoRR*, abs/2204.07580.

Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. 2019. Multilingual neural machine translation with language clustering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 963–973. Association for Computational Linguistics.

Yi Tay, Vinh Q. Tran, Sebastian Ruder, Jai Prakash Gupta, Hyung Won Chung, Dara Bahri, Zhen Qin, Simon Baumgartner, Cong Yu, and Donald Metzler. 2022. Charformer: Fast character transformers via gradient-based subword tokenization. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strötgen, and Hinrich Schütze. 2023. NLNDE at semeval-2023 task 12: Adaptive pretraining and source language selection for low-resource multilingual sentiment analysis. *CoRR*, abs/2305.00090.

Shijie Wu and Mark Dredze. 2019. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 833–844. Association for Computational Linguistics.

Linting Xue, Aditya Barua, Noah Constant, Rami Al-Rfou, Sharan Narang, Mihir Kale, Adam Roberts, and Colin Raffel. 2022. Byt5: Towards a token-free future with pre-trained byte-to-byte models. *Trans. Assoc. Comput. Linguistics*, 10:291–306.

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. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 483–498. Association for Computational Linguistics.

A Languages

Tab. 5-10 show the language list covered by mPLMs and corpora.

Tab. 11 provides the languages used for evaluating cross-lingual transfer.

	mBERT CANINE-S CANINE-C	XLM-R-Base XLM-R-Large	Glot500	mGPT	mT5-Base	XLM-Align	NLLB-200	XLS-R-300M	Flores	PBC	Fleurs
ace_Arab							✓		✓		
ace_Latn			✓				✓		✓	✓	
ach_Latn			✓							✓	
acm_Arab			✓				✓		✓		
acq_Arab							✓		✓		
acr_Latn			✓							✓	
aeb_Arab							✓		✓		
afr_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
agw_Latn			✓							✓	
ahk_Latn			✓							✓	
ajp_Arab			✓				✓		✓		
aka_Latn			✓				✓		✓	✓	
aln_Latn			✓							✓	
als_Latn			✓				✓		✓	✓	
alt_Cyrl			✓							✓	
alz_Latn			✓							✓	
amh_Ethi		✓	✓		✓	✓	✓	✓	✓	✓	✓
aoj_Latn			✓							✓	
apc_Arab			✓				✓		✓		
arb_Arab	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
arb_Latn							✓		✓		
arn_Latn			✓							✓	
ars_Arab							✓		✓		
ary_Arab			✓				✓		✓	✓	
arz_Arab			✓				✓		✓	✓	
asm_Beng		✓	✓			✓	✓	✓	✓	✓	✓
ast_Latn	✓		✓				✓		✓		✓
awa_Deva							✓		✓		
ayr_Latn			✓				✓		✓	✓	
azb_Arab	✓		✓				✓		✓	✓	
azj_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
bak_Cyrl	✓		✓	✓		✓	✓	✓	✓	✓	
bam_Latn			✓				✓		✓	✓	
ban_Latn			✓				✓		✓	✓	
bar_Latn	✓		✓							✓	
bba_Latn			✓							✓	
bbc_Latn			✓							✓	
bci_Latn			✓							✓	
bcl_Latn			✓							✓	
bel_Cyrl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
bem_Latn			✓				✓		✓	✓	
ben_Beng	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
bho_Deva			✓				✓		✓		
bhw_Latn			✓							✓	
bim_Latn			✓							✓	
bis_Latn			✓							✓	
bjn_Arab							✓		✓		
bjn_Latn			✓				✓		✓		
bod_Tibt			✓				✓	✓	✓	✓	
bos_Latn	✓	✓	✓				✓	✓	✓		✓
bqc_Latn			✓							✓	
bre_Latn	✓	✓	✓					✓		✓	
bts_Latn			✓							✓	
btx_Latn			✓							✓	
bug_Latn							✓		✓		
bul_Cyrl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
bum_Latn			✓							✓	
bzj_Latn			✓							✓	
cab_Latn			✓							✓	
cac_Latn			✓							✓	
cak_Latn			✓							✓	
caq_Latn			✓							✓	
cat_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	
cbk_Latn			✓							✓	
cce_Latn			✓							✓	
ceb_Latn	✓		✓		✓		✓	✓	✓	✓	✓
ces_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
cfm_Latn			✓							✓	
che_Cyrl	✓		✓							✓	
chk_Latn			✓							✓	
chv_Cyrl	✓		✓	✓				✓		✓	
cjk_Latn			✓				✓		✓		

Table 5: Languages covered by mPLMs and corpora.

	mBERT CANINE-S CANINE-C	XLM-R-Base XLM-R-Large	Glot500	mGPT	mT5-Base	XLM-Align	NLLB-200	XLS-R-300M	Flores	PBC	Fleurs
ckb_Arab		✓	✓		✓	✓	✓	✓	✓	✓	✓
ckb_Latn			✓							✓	
cmn_Hani	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cnh_Latn			✓					✓		✓	
crh_Cyrl			✓							✓	
crh_Latn			✓				✓		✓		
crs_Latn			✓							✓	
csy_Latn			✓							✓	
ctd_Latn			✓							✓	
ctu_Latn			✓							✓	
cuk_Latn			✓							✓	
cym_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
dan_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
deu_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
dik_Latn							✓		✓		
djk_Latn			✓							✓	
dln_Latn			✓							✓	
dtp_Latn			✓							✓	
dyu_Latn			✓				✓		✓	✓	
dzo_Tibt			✓				✓		✓	✓	
efi_Latn			✓							✓	
ekk_Latn	✓	✓	✓		✓	✓	✓	✓	✓		✓
ell_Grek	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
eng_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
enm_Latn			✓							✓	
epo_Latn		✓	✓		✓	✓	✓	✓	✓	✓	
eus_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
ewe_Latn			✓				✓		✓	✓	
fao_Latn			✓				✓	✓	✓	✓	
fij_Latn			✓				✓		✓	✓	
fil_Latn			✓		✓					✓	
fin_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
fon_Latn			✓				✓		✓	✓	
fra_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
fry_Latn	✓	✓	✓		✓			✓		✓	
fur_Latn			✓				✓		✓		
fuv_Latn							✓		✓		
gaa_Latn			✓							✓	
gaz_Latn		✓	✓				✓		✓		
gil_Latn			✓							✓	
giz_Latn			✓							✓	
gkn_Latn			✓							✓	
gkp_Latn			✓							✓	
gla_Latn		✓	✓		✓		✓		✓	✓	
gle_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
glg_Latn	✓	✓	✓		✓	✓	✓	✓	✓		
glv_Latn			✓					✓		✓	
gom_Latn			✓							✓	
gor_Latn			✓							✓	
grc_Grek			✓							✓	
guc_Latn			✓							✓	
gug_Latn			✓				✓	✓	✓	✓	
guj_Gujr	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
gur_Latn			✓							✓	
guw_Latn			✓							✓	
gya_Latn			✓							✓	
gym_Latn			✓							✓	
hat_Latn	✓		✓		✓		✓	✓	✓	✓	
hau_Latn		✓	✓		✓		✓	✓	✓	✓	✓
haw_Latn			✓		✓			✓		✓	
heb_Hebr	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
hif_Latn			✓							✓	
hil_Latn			✓							✓	
hin_Deva	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
hin_Latn		✓	✓		✓					✓	
hmo_Latn			✓							✓	
hne_Deva			✓				✓		✓	✓	
hnj_Latn			✓		✓					✓	
hra_Latn			✓							✓	
hrv_Latn	✓	✓	✓			✓	✓	✓	✓	✓	✓
hui_Latn			✓							✓	
hun_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 6: Languages covered by mPLMs and corpora.

	mBERT CANINE-S CANINE-C	XLM-R-Base XLM-R-Large	Glot500	mGPT	mT5-Base	XLM-Align	NLLB-200	XLS-R-300M	Flores	PBC	Fleurs
hus_Latn			✓							✓	
hye_Armn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
iba_Latn			✓							✓	
ibo_Latn			✓		✓		✓		✓	✓	✓
ifa_Latn			✓							✓	
ifb_Latn			✓							✓	
ikk_Latn			✓							✓	
ilo_Latn			✓				✓		✓	✓	
ind_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
isl_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	
ita_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ium_Latn			✓							✓	
ixl_Latn			✓							✓	
izz_Latn			✓							✓	
jam_Latn			✓							✓	
jav_Latn	✓	✓	✓		✓		✓	✓	✓	✓	✓
jpn_Jpan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
kaa_Cyrl			✓							✓	
kaa_Latn			✓							✓	
kab_Latn			✓				✓	✓	✓	✓	
kac_Latn			✓				✓		✓	✓	
kal_Latn			✓							✓	
kam_Latn			✓				✓		✓		✓
kan_Knda	✓	✓	✓		✓	✓	✓	✓	✓	✓	
kas_Arab							✓		✓		
kas_Deva							✓		✓		
kat_Geor	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
kaz_Cyrl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
kbp_Latn			✓				✓		✓	✓	
kea_Latn			✓				✓		✓		✓
kek_Latn			✓							✓	
khk_Cyrl							✓		✓		
khm_Khmr		✓	✓		✓	✓	✓	✓	✓	✓	
kia_Latn			✓							✓	
kik_Latn			✓				✓		✓	✓	
kin_Latn			✓				✓	✓	✓	✓	
kir_Cyrl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
kjb_Latn			✓							✓	
kjh_Cyrl			✓							✓	
kmb_Latn			✓				✓		✓		
kmm_Latn			✓							✓	
kmr_Cyrl			✓							✓	
kmr_Latn			✓				✓		✓	✓	
knc_Arab							✓		✓		
knc_Latn							✓		✓		
kng_Latn			✓				✓		✓		
knv_Latn			✓							✓	
kor_Hang	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
kpg_Latn			✓							✓	
krc_Cyrl			✓							✓	
kri_Latn			✓							✓	
ksd_Latn			✓							✓	
kss_Latn			✓							✓	
ksw_Mymr			✓							✓	
kua_Latn			✓							✓	
lam_Latn			✓							✓	
lao_Laoo		✓	✓		✓	✓	✓	✓	✓	✓	
lat_Latn	✓	✓	✓		✓	✓		✓		✓	
lav_Latn	✓	✓	✓	✓	✓	✓		✓		✓	
ldi_Latn			✓							✓	
leh_Latn			✓							✓	
lhu_Latn			✓							✓	
lij_Latn			✓				✓		✓		
lim_Latn			✓				✓		✓		
lin_Latn			✓				✓	✓	✓	✓	✓
lit_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
lmo_Latn	✓		✓				✓		✓		
loz_Latn			✓							✓	
ltg_Latn							✓		✓		
ltz_Latn	✓		✓		✓		✓	✓	✓	✓	✓
lua_Latn			✓				✓		✓		
lug_Latn			✓				✓	✓	✓	✓	

Table 7: Languages covered by mPLMs and corpora.

	mBERT CANINE-S CANINE-C	XLM-R-Base XLM-R-Large	Glot500	mGPT	mT5-Base	XLM-Align	NLLB-200	XLS-R-300M	Flores	PBC	Fleurs
luo_Latn			✓				✓		✓	✓	
lus_Latn			✓				✓		✓	✓	
lvs_Latn			✓				✓		✓		
lzh_Hani			✓							✓	
mad_Latn			✓							✓	
mag_Deva							✓		✓		
mah_Latn			✓							✓	
mai_Deva			✓				✓		✓	✓	
mal_Mlym	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
mam_Latn			✓							✓	
mar_Deva	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
mau_Latn			✓							✓	
mbb_Latn			✓							✓	
mck_Latn			✓							✓	
mcn_Latn			✓							✓	
mco_Latn			✓							✓	
mdy_Ethi			✓							✓	
meu_Latn			✓							✓	
mfe_Latn			✓							✓	
mgh_Latn			✓							✓	
mgr_Latn			✓							✓	
mhr_Cyrl			✓							✓	
min_Arab							✓		✓		
min_Latn	✓		✓				✓		✓	✓	
miq_Latn			✓							✓	
mkd_Cyrl	✓	✓	✓		✓	✓	✓	✓	✓	✓	
mlt_Latn			✓		✓	✓	✓	✓	✓	✓	✓
mni_Beng							✓		✓		
mon_Cyrl		✓	✓	✓	✓	✓		✓			✓
mos_Latn			✓				✓		✓	✓	
mps_Latn			✓							✓	
mri_Latn			✓		✓		✓	✓	✓	✓	✓
mrw_Latn			✓							✓	
mwm_Latn			✓							✓	
mxv_Latn			✓							✓	
mya_Mymr	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
myv_Cyrl			✓							✓	
mzh_Latn			✓							✓	
nan_Latn			✓							✓	
naq_Latn			✓							✓	
nav_Latn			✓							✓	
nbl_Latn			✓							✓	
nch_Latn			✓							✓	
ncj_Latn			✓							✓	
ndc_Latn			✓							✓	
nde_Latn			✓							✓	
ndo_Latn			✓							✓	
nds_Latn	✓		✓							✓	
nep_Deva	✓	✓	✓		✓	✓		✓		✓	✓
ngu_Latn			✓							✓	
nia_Latn			✓							✓	
nld_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
nmf_Latn			✓							✓	
nnb_Latn			✓							✓	
nno_Latn	✓		✓			✓	✓	✓	✓	✓	
nob_Latn	✓		✓				✓		✓	✓	
nor_Latn		✓	✓		✓	✓		✓		✓	
npi_Deva			✓				✓		✓	✓	
nse_Latn			✓							✓	
nso_Latn			✓				✓		✓	✓	
nus_Latn							✓		✓		
nya_Latn			✓		✓		✓		✓	✓	✓
nyn_Latn			✓							✓	
nyy_Latn			✓							✓	
nzi_Latn			✓							✓	
oci_Latn	✓		✓				✓	✓	✓		✓
ory_Orya		✓	✓			✓	✓	✓	✓	✓	
oss_Cyrl			✓	✓						✓	
ote_Latn			✓							✓	
pag_Latn			✓				✓		✓	✓	
pam_Latn			✓							✓	
pan_Guru	✓	✓	✓		✓	✓	✓	✓	✓	✓	

Table 8: Languages covered by mPLMs and corpora.

	mBERT CANINE-S CANINE-C	XLM-R-Base XLM-R-Large	Glot500	mGPT	mT5-Base	XLM-Align	NLLB-200	XLS-R-300M	Flores	PBC	Fleurs
pap_Latn			✓				✓		✓	✓	
pau_Latn			✓							✓	
pbt_Arab							✓		✓		
pcm_Latn			✓							✓	
pdt_Latn			✓							✓	
pes_Arab	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
pis_Latn			✓							✓	
pls_Latn			✓							✓	
plt_Latn	✓	✓	✓		✓		✓	✓	✓	✓	
poh_Latn			✓							✓	
pol_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
pon_Latn			✓							✓	
por_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
prk_Latn			✓							✓	
prs_Arab			✓				✓		✓	✓	
pxm_Latn			✓							✓	
qub_Latn			✓							✓	
quc_Latn			✓							✓	
qug_Latn			✓							✓	
quh_Latn			✓							✓	
quw_Latn			✓							✓	
quy_Latn			✓				✓		✓	✓	
quz_Latn			✓							✓	
qvi_Latn			✓							✓	
rap_Latn			✓							✓	
rar_Latn			✓							✓	
rmy_Latn			✓							✓	
ron_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
rop_Latn			✓							✓	
rug_Latn			✓							✓	
run_Latn			✓				✓		✓	✓	
rus_Cyrl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sag_Latn			✓				✓		✓	✓	
sah_Cyrl			✓	✓				✓		✓	
san_Deva		✓	✓			✓	✓	✓	✓	✓	
san_Latn			✓							✓	
sat_Olck			✓				✓		✓		
sba_Latn			✓							✓	
scn_Latn	✓		✓				✓		✓		
seh_Latn			✓							✓	
shn_Mymr							✓		✓		
sin_Sinh		✓	✓		✓	✓	✓	✓	✓	✓	
slk_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	
slv_Latn	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
sme_Latn			✓							✓	
smo_Latn			✓		✓		✓		✓	✓	
sna_Latn			✓		✓		✓	✓	✓	✓	✓
snd_Arab		✓	✓		✓	✓	✓	✓	✓	✓	✓
som_Latn		✓	✓		✓		✓	✓	✓	✓	✓
sop_Latn			✓							✓	
sot_Latn			✓		✓		✓		✓	✓	
spa_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sqi_Latn	✓	✓	✓		✓	✓		✓		✓	
srm_Latn			✓							✓	
srn_Latn			✓							✓	
sro_Latn			✓				✓		✓		
srp_Cyrl	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
srp_Latn			✓							✓	
ssw_Latn			✓				✓		✓	✓	
sun_Latn	✓	✓	✓		✓		✓	✓	✓	✓	
suz_Deva			✓							✓	
swe_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
swh_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sxn_Latn			✓							✓	
szl_Latn			✓				✓		✓		
tam_Latn		✓								✓	
tam_Taml	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
taq_Latn							✓		✓		
taq_Tfng							✓		✓		
tat_Cyrl	✓		✓	✓		✓	✓	✓	✓	✓	
tbz_Latn			✓							✓	
tca_Latn			✓							✓	

Table 9: Languages covered by mPLMs and corpora.

	mBERT CANINE-S CANINE-C	XLM-R-Base XLM-R-Large	Glot500	mGPT	mT5-Base	XLM-Align	NLLB-200	XLS-R-300M	Flores	PBC	Fleurs
tdt_Latn			✓							✓	
tel_Telu	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
teo_Latn			✓							✓	
tgk_Cyrl	✓		✓	✓	✓	✓	✓	✓	✓	✓	
tgl_Latn	✓	✓	✓	✓		✓	✓	✓	✓	✓	
tha_Thai		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
tih_Latn			✓							✓	
tir_Ethi			✓				✓		✓	✓	
tlh_Latn			✓							✓	
tob_Latn			✓							✓	
toh_Latn			✓							✓	
toi_Latn			✓							✓	
toj_Latn			✓							✓	
ton_Latn			✓							✓	
top_Latn			✓							✓	
tpi_Latn			✓				✓	✓	✓	✓	
tpm_Latn			✓							✓	
tsn_Latn			✓				✓		✓	✓	
tso_Latn			✓				✓		✓	✓	
tsz_Latn			✓							✓	
tuc_Latn			✓							✓	
tui_Latn			✓							✓	
tuk_Cyrl			✓							✓	
tuk_Latn			✓	✓			✓	✓	✓	✓	
tum_Latn			✓				✓		✓	✓	
tur_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
twi_Latn			✓				✓		✓	✓	
tyv_Cyrl			✓	✓						✓	
tzh_Latn			✓							✓	
tzm_Tfng							✓		✓		
tzo_Latn			✓							✓	
udm_Cyrl			✓							✓	
uig_Arab		✓	✓			✓	✓		✓	✓	
uig_Latn			✓							✓	
ukr_Cyrl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
umb_Latn			✓				✓		✓		
urd_Arab	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
urd_Latn		✓								✓	
uzn_Cyrl			✓							✓	
uzn_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
vec_Latn			✓				✓		✓		
ven_Latn			✓							✓	
vie_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
wal_Latn			✓							✓	
war_Latn	✓		✓				✓	✓	✓	✓	
wol_Latn			✓				✓		✓	✓	
xav_Latn			✓							✓	
xho_Latn		✓	✓		✓		✓		✓	✓	✓
yan_Latn			✓							✓	
yao_Latn			✓							✓	
yap_Latn			✓							✓	
ydd_Hebr		✓	✓		✓	✓	✓	✓	✓		
yom_Latn			✓							✓	
yor_Latn	✓		✓	✓	✓		✓	✓	✓	✓	
yua_Latn			✓							✓	
yue_Hani			✓				✓	✓	✓	✓	
zai_Latn			✓							✓	
zlm_Latn			✓							✓	
zom_Latn			✓							✓	
zsm_Latn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
zul_Latn			✓		✓		✓	✓	✓	✓	✓

Table 10: Languages covered by mPLMs and corpora.

Task	Language List
NER (108)	ace_Latn, afr_Latn, als_Latn, amh_Ethi, arz_Arab, asm_Beng, ast_Latn, azj_Latn, bak_Cyrl, bel_Cyrl, ben_Beng, bho_Deva, bod_Tibt, bos_Latn, bul_Cyrl, cat_Latn, ceb_Latn, ces_Latn, ckb_Arab, crh_Latn, cym_Latn, dan_Latn, deu_Latn, ekk_Latn, ell_Grek, epo_Latn, eus_Latn, fao_Latn, fin_Latn, fra_Latn, fur_Latn, gla_Latn, gle_Latn, glg_Latn, gug_Latn, guj_Gujr, heb_Hebr, hrv_Latn, hun_Latn, hye_Armn, ibo_Latn, ilo_Latn, ind_Latn, isl_Latn, ita_Latn, jav_Latn, jpn_Jpan, kan_Knda, kat_Geor, kaz_Cyrl, khm_Khmr, kin_Latn, kir_Cyrl, kor_Hang, lij_Latn, lim_Latn, lin_Latn, lit_Latn, lmo_Latn, ltz_Latn, mal_Mlym, mar_Deva, min_Latn, mkd_Cyrl, mlt_Latn, mri_Latn, mya_Mymr, nld_Latn, nno_Latn, oci_Latn, ory_Orya, pan_Guru, pes_Arab, plt_Latn, pol_Latn, por_Latn, ron_Latn, san_Deva, scn_Latn, sin_Sinh, slk_Latn, slv_Latn, snd_Arab, som_Latn, srp_Cyrl, sun_Latn, swe_Latn, swh_Latn, szl_Latn, tam_Taml, tat_Cyrl, tel_Telu, tgk_Cyrl, tgl_Latn, tha_Thai, tuk_Latn, tur_Latn, uig_Arab, ukr_Cyrl, urd_Arab, uzn_Latn, vec_Latn, vie_Latn, war_Latn, ydd_Hebr, yor_Latn, yue_Hani, zsm_Latn
POS (60)	afr_Latn, ajp_Arab, amh_Ethi, bam_Latn, bel_Cyrl, bho_Deva, bul_Cyrl, cat_Latn, ceb_Latn, ces_Latn, cym_Latn, dan_Latn, deu_Latn, ekk_Latn, ell_Grek, eus_Latn, fao_Latn, fin_Latn, fra_Latn, gla_Latn, gle_Latn, glg_Latn, heb_Hebr, hrv_Latn, hun_Latn, hye_Armn, ind_Latn, isl_Latn, ita_Latn, jav_Latn, jpn_Jpan, kaz_Cyrl, kmr_Latn, kor_Hang, lij_Latn, lit_Latn, mlt_Latn, nld_Latn, pes_Arab, pol_Latn, por_Latn, ron_Latn, san_Deva, sin_Sinh, slk_Latn, slv_Latn, swe_Latn, tam_Taml, tat_Cyrl, tel_Telu, tgl_Latn, tha_Thai, tur_Latn, uig_Arab, ukr_Cyrl, urd_Arab, vie_Latn, wol_Latn, yor_Latn, yue_Hani
Massive (44)	afr_Latn, als_Latn, amh_Ethi, azj_Latn, ben_Beng, cat_Latn, cym_Latn, dan_Latn, deu_Latn, ell_Grek, fin_Latn, fra_Latn, heb_Hebr, hun_Latn, hye_Armn, ind_Latn, isl_Latn, ita_Latn, jav_Latn, jpn_Jpan, kan_Knda, kat_Geor, khm_Khmr, kor_Hang, lvs_Latn, mal_Mlym, mya_Mymr, nld_Latn, nob_Latn, pes_Arab, pol_Latn, por_Latn, ron_Latn, slv_Latn, swe_Latn, swh_Latn, tam_Taml, tel_Telu, tgl_Latn, tha_Thai, tur_Latn, urd_Arab, vie_Latn, zsm_Latn
Taxi1500 (130)	ace_Latn, afr_Latn, aka_Latn, als_Latn, ary_Arab, arz_Arab, asm_Beng, ayr_Latn, azb_Arab, bak_Cyrl, bam_Latn, ban_Latn, bel_Cyrl, bem_Latn, ben_Beng, bul_Cyrl, cat_Latn, ceb_Latn, ces_Latn, ckb_Arab, cym_Latn, dan_Latn, deu_Latn, dyu_Latn, dzo_Tibt, ell_Grek, epo_Latn, eus_Latn, ewe_Latn, fao_Latn, fij_Latn, fin_Latn, fon_Latn, fra_Latn, gla_Latn, gle_Latn, gug_Latn, guj_Gujr, hat_Latn, hau_Latn, heb_Hebr, hne_Deva, hrv_Latn, hun_Latn, hye_Armn, ibo_Latn, ilo_Latn, ind_Latn, isl_Latn, ita_Latn, jav_Latn, kab_Latn, kac_Latn, kan_Knda, kat_Geor, kaz_Cyrl, kbp_Latn, khm_Khmr, kik_Latn, kin_Latn, kir_Cyrl, kng_Latn, kor_Hang, lao_Laoo, lin_Latn, lit_Latn, ltz_Latn, lug_Latn, luo_Latn, mai_Deva, mar_Deva, min_Latn, mkd_Cyrl, mlt_Latn, mos_Latn, mri_Latn, mya_Mymr, nld_Latn, nno_Latn, nob_Latn, npi_Deva, nso_Latn, nya_Latn, ory_Orya, pag_Latn, pan_Guru, pap_Latn, pes_Arab, plt_Latn, pol_Latn, por_Latn, prs_Arab, quy_Latn, ron_Latn, run_Latn, sag_Latn, sin_Sinh, slk_Latn, slv_Latn, smo_Latn, sna_Latn, snd_Arab, som_Latn, sot_Latn, ssw_Latn, sun_Latn, swe_Latn, swh_Latn, tam_Taml, tat_Cyrl, tel_Telu, tgk_Cyrl, tgl_Latn, tha_Thai, tir_Ethi, tpi_Latn, tsn_Latn, tuk_Latn, tum_Latn, tur_Latn, twi_Latn, ukr_Cyrl, vie_Latn, war_Latn, wol_Latn, xho_Latn, yor_Latn, yue_Hani, zsm_Latn, zul_Latn

Table 11: Languages for evaluating zero-shot cross-lingual transfer. The number in brackets is the number of the evaluated languages.

	mPLM-Sim	Mono	1	5	10
LEX	0.741	0.704	0.688	0.745	0.743
GEN	0.527	0.504	0.480	0.482	0.510
GEO	0.608	0.597	0.523	0.562	0.597
SYN	0.577	0.583	0.556	0.560	0.573
INV	0.248	0.245	0.226	0.265	0.260
PHO	0.094	0.109	0.114	0.118	0.102
FEA	0.358	0.369	0.347	0.371	0.360
AVG	**0.451**	0.444	0.419	0.444	0.449

Table 12: Comparison of pearson correlation result: Pearson correlation between seven similarity measurs and mPLM-Sim (500 multi-parallel sentences), Mono (Monolingual corpora) and the results of using different amounts (1, 5, 10) of multi-parallel sentences.

B Comparison Across Corpora for mPLM-Sim

B.1 Monolingual vs. Parallel

Both monolingual and parallel corpora can be exploited for obtaining sentence embeddings for measuring language similarity. We conduct experiments of exploiting monolingual corpora for measuring similarity across languages, and also provide the results of using different amounts (1, 5, 10, 500) of multi-parallel sentences.

For the experiment of pearson correlation in Sec. 5.1, the results (MEAN) are shown in Tab. 12. For the experiment of cross-lingual transfer in Sec. 5.4, the results are shown in Tab. 13. Based on these two experiments, we have the conclusions below:

- mPLM-Sim using multi-parallel corpora achieves slightly better results than using monolingual corpora.
- mPLM-Sim (500 sentences) requires less data than exploiting monolingual corpora. Besides, using mPLM-Sim (10 sentences) can achieve comparable results with mPLM-Sim (500 sentences). While including a truly low-resource language for similarity measurement, mPLM-Sim requires around 10 sentences parallel to one existing language, while monolingual corpora requires massive sentences.

In a word, exploiting parallel corpora is better for measuring language similarity than monolingual corpora.

B.2 Flores vs. PBC

To investigate the impact of multi-parallel corpora on the performance of mPLM-Sim, we compare

	mPLM-Sim	Mono	1	5	10
NER	0.647	0.644	0.644	0.646	0.647
POS	0.751	0.737	0.748	0.753	0.752
Massive	0.730	0.730	0.723	0.728	0.730
Taxi	0.583	0.585	0.580	0.582	0.582
AVG	**0.678**	0.674	0.674	0.677	**0.678**

Table 13: Comparison of cross-lingual transfer result: Cross-lingual transfer result for four tasks from mPLM-Sim (500 multi-parallel sentences), Mono (Monolingual corpora) and the results of using different amounts (1, 5, 10) of multi-parallel sentences.

	Flores		PBC	
	M	Mdn	M	Mdn
LEX	0.741	0.864	0.654	0.735
GEN	0.527	0.600	0.519	0.572
GEO	0.608	0.674	0.546	0.603
SYN	0.577	0.607	0.491	0.528
INV	0.248	0.293	0.254	0.276
PHO	0.094	0.144	0.103	0.098
FEA	0.358	0.372	0.333	0.357
AVG	0.451	0.508	0.414	0.453

Table 14: Comparison across corpora: Pearson correlation between mPLM-Sim and linguistic similarity measures for Glot500 and all corpora on 32 languages. Flores achieves higher correlations than PBC.

the results of Glot500 with Flores and PBC on 32 languages that are covered by both corpora.

Tab. 14 shows that Flores outperforms PBC across all similarity measures, except for PHO. To gain further insights, we conduct a case study focusing on languages that exhibit different performances between the two corpora.

In comparison to PBC, Flores consists of text that is closer to web content and spans a wider range of general domains. For example, a significant portion of Arabic script in Flores is written without short vowels, which are commonly used in texts requiring strict adherence to precise pronunciation, such as the Bible.[5] This discrepancy leads to challenges in tokenization and representation for languages written in Arabic, such as Moroccan Arabic (ary_Arab) and Egyptian Arabic (arz_Arab), resulting in poorer performance.

[5] https://en.wikipedia.org/wiki/Arabic_diacritics

C Visualization and Analysis Across Layers

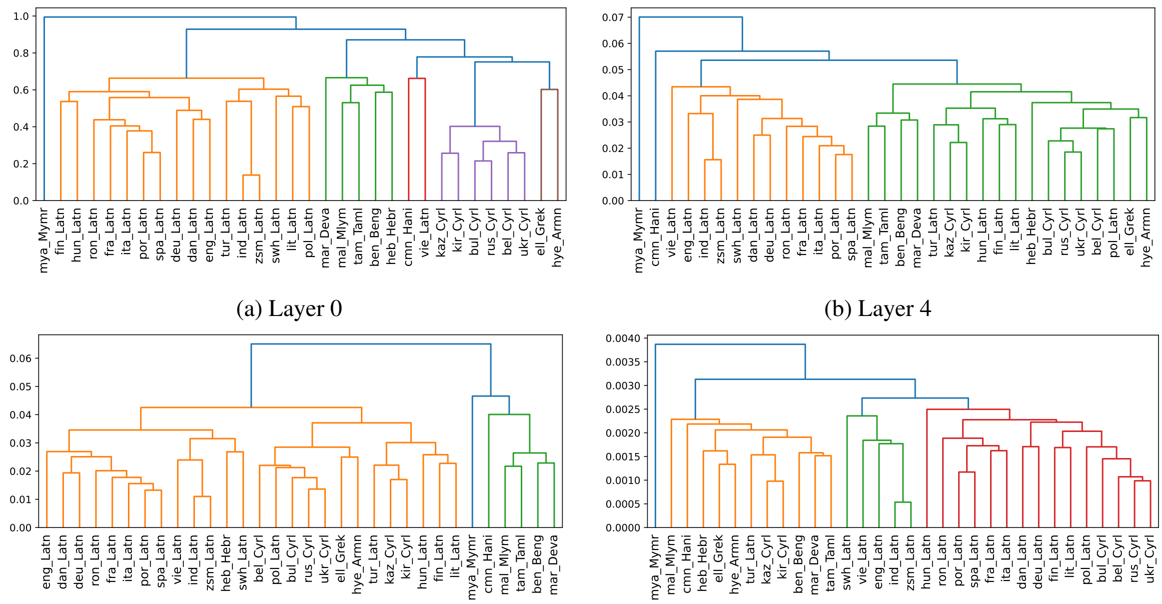
C.1 Hierarchical Clustering Analysis

We conducted hierarchical clustering analysis at different layers (0, 4, 8, and 12) using the setting of Glot500 and Flores for mPLM-Sim. The results, shown in Fig. 3, reveal distinct patterns of language clustering. In layer 0, the clustering primarily emphasizes lexical similarities, with languages sharing the same scripts being grouped together. As we progress to layers 4 and 8, more high-level similarity patterns beyond the surface-level are captured. For instance in these layers, Turkish (tur_Latn) and Polish (pol_Latn) are clustered with their Turkic and Slavic relatives although they use different writing systems. The similarity results of layer 12 are comparatively worse than those of the middle layers. For instance, English (eng_Latn) deviates from its Germanic and Indo-European relatives and instead clusters with Malay languages (ind_Latn, zsm_Latn). This phenomenon can be attributed to the higher layer exhibiting lower inter-cluster distances (comparison between the y-axis range across figures of different layers), which diminishes its ability to effectively discriminate between language clusters.

C.2 Similarity Heatmaps

Fig. 4-7 show the cosine simlarity values in heatmaps at layer 0, 4, 8 and 12, using the Glot500 and Flores settings for mPLM-Sim.

Generally, as the layer number increases, higher cosine similarity values are observed. Layer 0 exhibits a significant contrast in similarity values, whereas layer 12 demonstrates very low contrast. Notably, Burmese (mya_Mymr) consistently receives the lowest values across all layers, indicating the relationship between Burmese and other languages may be not well modeled.



(a) Layer 0 (b) Layer 4

(c) Layer 8 (d) Layer 12

Figure 3: Dendrograms illustrating hierarchical clustering results at layer 0, 4, 8, and 12 for Glot500 and Flores across 32 languages.

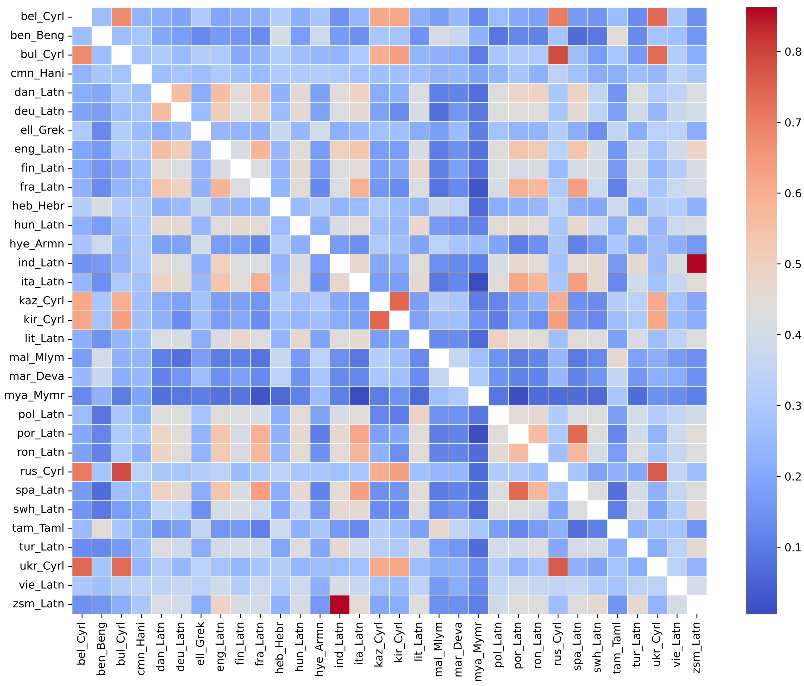


Figure 4: Heatmaps of cosine similarity results at layer 0 for Glot500 and Flores across 32 languages.

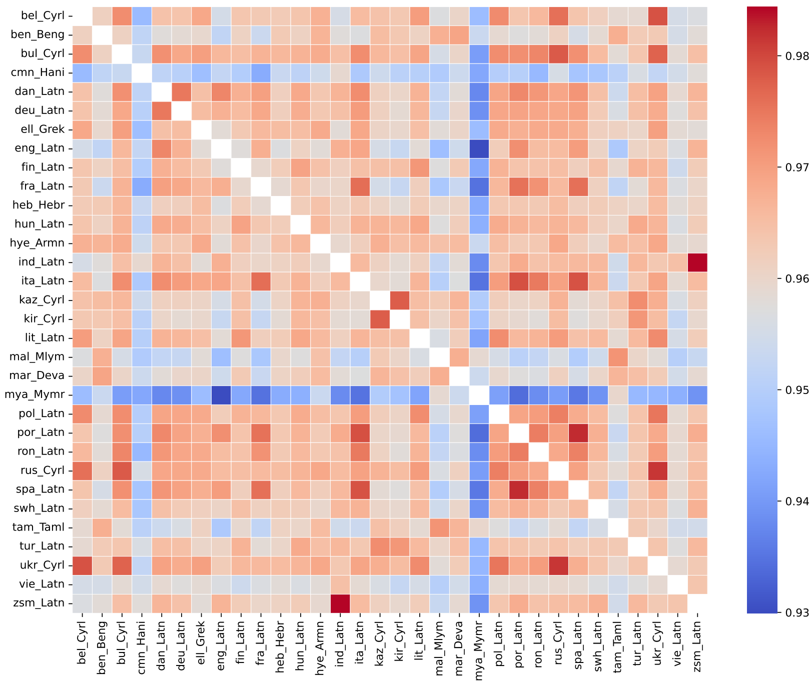


Figure 5: Heatmaps of cosine similarity results at layer 4 for Glot500 and Flores across 32 languages.

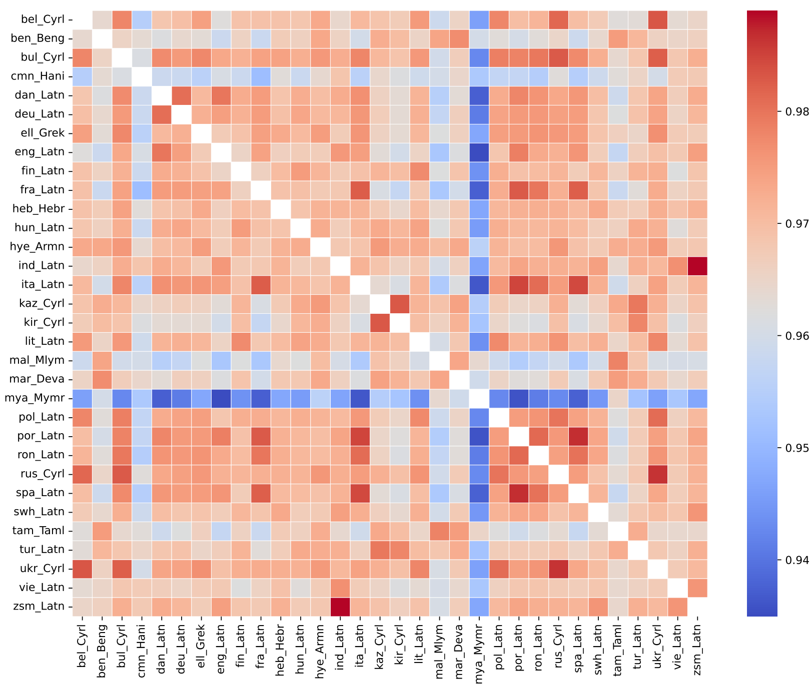


Figure 6: Heatmaps of cosine similarity results at layer 8 for Glot500 and Flores across 32 languages.

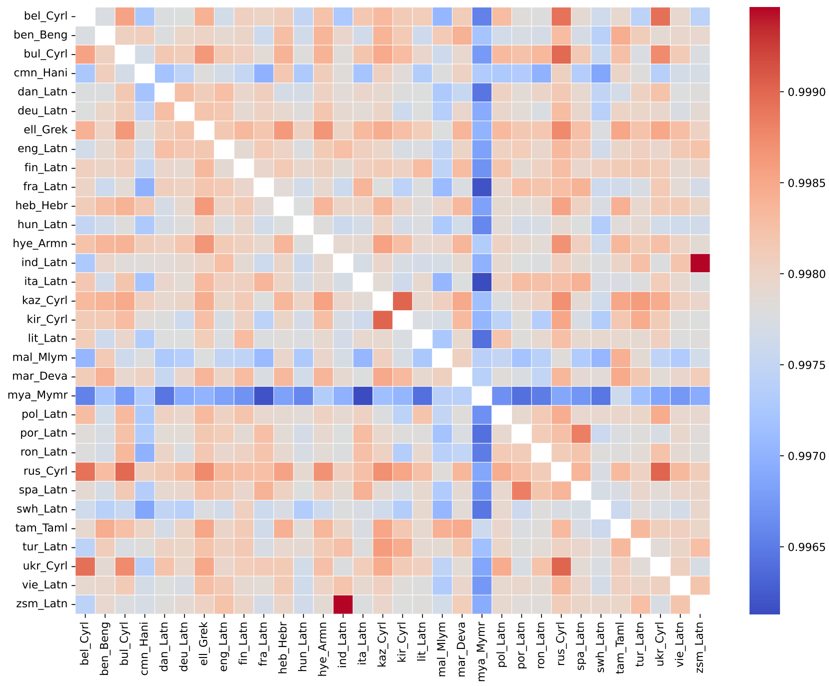


Figure 7: Heatmaps of cosine similarity results at layer 12 for Glot500 and Flores across 32 languages.

D Analysis on Unseen Languages of mPLMs

The success of mPLM-Sim depends on the supporting languages of mPLMs. To get more insights about languages which are this not supported by a specific mPLM, we conduct a new Pearson correlation experiment based on 94 languages unseen by XLM-R. Among 94 languages, there are 24 (25.5%) languages that achieve higher correlation than the average level of seen languages. These 24 languages usually have close languages seen by XLM-R, e.g, the unseen language, Cantonese (yue_Hani) is close to Mandarin (cmn_Hani). It shows that mPLM-Sim can be directly applied to some unseen languages which have close seen languages.

For the unseen languages which mPLM-Sim performs poorly, we can connect it to seen languages using traditional linguistic features, e.g., language family, and then use or weight the similarity results of seen languages as the results of the unseen languages. Since it is shown that mPLM-Sim provides better results than traditional linguistic features in our paper, connecting unseen languages to seen languages would be beneficial for unseen languages.

E Detailed Results of Cross-Lingual Transfer

We report the detailed results for all tasks and languages in Tab. 15-16 (NER), 17 (POS), 18 (MASSIVE), 19-21 (Taxi1500).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
ace_Latn	0.421	0.421	eng_Latn	0.421	eng_Latn	0.427	hin_Deva	0.421	eng_Latn	**0.439**	spa_Latn
afr_Latn	**0.739**	**0.739**	eng_Latn	**0.739**	eng_Latn	0.720	arb_Arab	0.707	rus_Cyrl	**0.739**	eng_Latn
als_Latn	0.767	0.767	eng_Latn	0.737	rus_Cyrl	**0.774**	spa_Latn	0.737	rus_Cyrl	**0.774**	spa_Latn
amh_Ethi	0.450	0.389	cmn_Hani	0.515	arb_Arab	0.515	arb_Arab	**0.554**	hin_Deva	**0.554**	hin_Deva
arz_Arab	0.491	**0.715**	arb_Arab	**0.715**	arb_Arab	**0.715**	arb_Arab	0.491	eng_Latn	**0.715**	arb_Arab
asm_Beng	0.661	0.603	arb_Arab	**0.720**	hin_Deva	**0.720**	hin_Deva	**0.720**	hin_Deva	**0.720**	hin_Deva
ast_Latn	0.813	**0.857**	spa_Latn	**0.857**	spa_Latn	**0.857**	spa_Latn	0.680	hin_Deva	**0.857**	spa_Latn
azj_Latn	0.625	0.625	eng_Latn	0.625	eng_Latn	**0.664**	arb_Arab	0.654	hin_Deva	0.648	spa_Latn
bak_Cyrl	0.558	0.675	rus_Cyrl	0.558	eng_Latn	0.675	rus_Cyrl	**0.681**	hin_Deva	0.675	rus_Cyrl
bel_Cyrl	0.728	**0.748**	rus_Cyrl	**0.748**	rus_Cyrl	0.728	eng_Latn	0.715	arb_Arab	**0.748**	rus_Cyrl
ben_Beng	0.670	0.647	arb_Arab	**0.692**	hin_Deva	**0.692**	hin_Deva	**0.692**	hin_Deva	**0.692**	hin_Deva
bho_Deva	0.544	**0.690**	hin_Deva	**0.690**	hin_Deva	**0.690**	hin_Deva	0.610	arb_Arab	**0.690**	hin_Deva
bod_Tibt	0.417	**0.544**	cmn_Hani	**0.544**	cmn_Hani	0.522	hin_Deva	**0.544**	cmn_Hani	**0.544**	cmn_Hani
bos_Latn	0.697	0.697	eng_Latn	**0.756**	rus_Cyrl	0.715	spa_Latn	0.702	arb_Arab	0.715	spa_Latn
bul_Cyrl	0.748	0.783	rus_Cyrl	0.783	rus_Cyrl	**0.787**	spa_Latn	0.783	rus_Cyrl	0.783	rus_Cyrl
cat_Latn	0.806	**0.808**	spa_Latn	**0.808**	spa_Latn	**0.808**	spa_Latn	0.806	eng_Latn	**0.808**	spa_Latn
ceb_Latn	**0.563**	**0.563**	eng_Latn	**0.563**	eng_Latn	0.211	cmn_Hani	0.530	spa_Latn	0.530	spa_Latn
ces_Latn	**0.760**	**0.760**	eng_Latn	0.741	rus_Cyrl	**0.760**	eng_Latn	0.741	rus_Cyrl	0.741	rus_Cyrl
ckb_Arab	0.707	**0.716**	arb_Arab	0.692	hin_Deva	**0.716**	arb_Arab	0.703	rus_Cyrl	**0.716**	arb_Arab
crh_Latn	0.521	0.521	eng_Latn	0.521	eng_Latn	0.472	arb_Arab	0.402	cmn_Hani	**0.551**	spa_Latn
cym_Latn	0.593	0.593	eng_Latn	0.617	rus_Cyrl	0.593	eng_Latn	0.542	arb_Arab	**0.636**	spa_Latn
dan_Latn	**0.792**	**0.792**	eng_Latn	**0.792**	eng_Latn	**0.792**	eng_Latn	0.747	arb_Arab	**0.792**	eng_Latn
deu_Latn	**0.714**	**0.714**	eng_Latn	**0.714**	eng_Latn	**0.714**	eng_Latn	**0.714**	eng_Latn	0.706	spa_Latn
ekk_Latn	0.713	0.713	eng_Latn	0.713	eng_Latn	0.713	eng_Latn	**0.729**	rus_Cyrl	0.729	spa_Latn
ell_Grek	0.686	0.686	eng_Latn	**0.733**	rus_Cyrl	0.729	spa_Latn	**0.733**	rus_Cyrl	**0.733**	rus_Cyrl
epo_Latn	0.639	0.639	eng_Latn	0.639	eng_Latn	0.639	eng_Latn	0.628	rus_Cyrl	**0.722**	spa_Latn
eus_Latn	0.516	0.516	eng_Latn	0.516	eng_Latn	0.552	spa_Latn	**0.588**	hin_Deva	0.552	spa_Latn
fao_Latn	0.706	0.706	eng_Latn	0.706	eng_Latn	0.706	eng_Latn	0.710	arb_Arab	**0.719**	spa_Latn
fin_Latn	0.728	0.728	eng_Latn	0.728	eng_Latn	0.728	eng_Latn	0.728	rus_Cyrl	**0.760**	spa_Latn
fra_Latn	0.730	0.730	eng_Latn	**0.805**	spa_Latn	0.730	eng_Latn	0.730	eng_Latn	**0.805**	spa_Latn
fur_Latn	0.567	0.567	eng_Latn	0.545	spa_Latn	0.567	eng_Latn	**0.605**	hin_Deva	0.545	spa_Latn
gla_Latn	0.571	0.571	eng_Latn	**0.612**	rus_Cyrl	0.571	eng_Latn	0.576	arb_Arab	0.582	spa_Latn
gle_Latn	0.670	0.670	eng_Latn	0.574	rus_Cyrl	0.670	eng_Latn	**0.688**	spa_Latn	**0.688**	spa_Latn
glg_Latn	0.768	**0.822**	spa_Latn	**0.822**	spa_Latn	**0.822**	spa_Latn	**0.822**	spa_Latn	**0.822**	spa_Latn
gug_Latn	0.552	0.552	eng_Latn	0.552	eng_Latn	**0.566**	spa_Latn	**0.566**	spa_Latn	**0.566**	spa_Latn
guj_Gujr	0.573	0.582	arb_Arab	**0.606**	hin_Deva	**0.606**	hin_Deva	**0.606**	hin_Deva	**0.606**	hin_Deva
heb_Hebr	0.458	0.300	cmn_Hani	**0.542**	arb_Arab	**0.542**	arb_Arab	0.463	rus_Cyrl	**0.542**	arb_Arab
hin_Deva	0.650	**0.697**	arb_Arab	**0.697**	arb_Arab	**0.697**	arb_Arab	**0.697**	arb_Arab	**0.697**	arb_Arab
hrv_Latn	0.738	0.738	eng_Latn	0.746	rus_Cyrl	0.738	eng_Latn	0.746	rus_Cyrl	**0.776**	spa_Latn
hun_Latn	0.727	0.727	eng_Latn	0.727	eng_Latn	0.727	eng_Latn	0.721	rus_Cyrl	**0.762**	spa_Latn
hye_Armn	0.518	**0.533**	arb_Arab	0.518	eng_Latn	**0.533**	arb_Arab	0.512	rus_Cyrl	0.531	hin_Deva
ibo_Latn	**0.574**	**0.574**	eng_Latn	**0.574**	eng_Latn	0.563	spa_Latn	**0.574**	eng_Latn	0.563	spa_Latn
ilo_Latn	0.673	0.673	eng_Latn	0.673	eng_Latn	0.577	cmn_Hani	0.673	eng_Latn	**0.716**	spa_Latn
ind_Latn	**0.594**	**0.594**	eng_Latn	**0.594**	eng_Latn	0.443	hin_Deva	**0.594**	eng_Latn	**0.594**	eng_Latn
isl_Latn	0.707	0.707	eng_Latn	0.707	eng_Latn	0.707	eng_Latn	0.707	eng_Latn	**0.726**	spa_Latn
ita_Latn	**0.764**	0.762	spa_Latn	0.762	spa_Latn	0.762	spa_Latn	0.762	spa_Latn	0.762	spa_Latn
jav_Latn	0.580	0.580	eng_Latn	0.580	eng_Latn	0.215	cmn_Hani	0.529	hin_Deva	**0.614**	spa_Latn
jpn_Jpan	0.177	**0.451**	cmn_Hani	0.177	eng_Latn	**0.451**	cmn_Hani	0.260	hin_Deva	**0.451**	cmn_Hani
kan_Knda	0.531	0.567	arb_Arab	0.531	eng_Latn	**0.638**	hin_Deva	**0.638**	hin_Deva	**0.638**	hin_Deva
kat_Geor	0.644	0.640	arb_Arab	0.644	eng_Latn	0.640	arb_Arab	**0.681**	hin_Deva	**0.681**	hin_Deva
kaz_Cyrl	0.416	**0.525**	rus_Cyrl	0.416	eng_Latn	**0.525**	rus_Cyrl	0.315	cmn_Hani	**0.525**	rus_Cyrl
khm_Khmr	0.404	0.404	eng_Latn	0.404	eng_Latn	0.467	hin_Deva	0.404	eng_Latn	**0.549**	arb_Arab
kin_Latn	0.626	0.626	eng_Latn	0.626	eng_Latn	0.672	arb_Arab	0.626	eng_Latn	**0.726**	spa_Latn
kir_Cyrl	0.391	**0.564**	rus_Cyrl	0.391	eng_Latn	**0.564**	rus_Cyrl	0.455	hin_Deva	**0.564**	rus_Cyrl
kor_Hang	0.470	0.445	cmn_Hani	0.470	eng_Latn	0.445	cmn_Hani	0.445	cmn_Hani	**0.551**	hin_Deva

Table 15: Cross-Lingual Transfer Results of NER (Part 1): The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 1).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
lij_Latn	**0.431**	**0.431**	eng_Latn	0.413	spa_Latn	0.413	spa_Latn	0.395	hin_Deva	0.413	spa_Latn
lim_Latn	**0.646**	**0.646**	eng_Latn	**0.646**	eng_Latn	**0.646**	eng_Latn	0.605	hin_Deva	0.621	spa_Latn
lin_Latn	0.486	0.486	eng_Latn	0.486	eng_Latn	**0.555**	arb_Arab	0.486	eng_Latn	0.519	spa_Latn
lit_Latn	**0.707**	**0.707**	eng_Latn	0.699	rus_Cyrl	**0.707**	eng_Latn	0.699	rus_Cyrl	0.699	rus_Cyrl
lmo_Latn	**0.712**	**0.712**	eng_Latn	0.706	spa_Latn	0.706	spa_Latn	0.559	hin_Deva	0.706	spa_Latn
ltz_Latn	0.646	0.646	eng_Latn	0.646	eng_Latn	0.646	eng_Latn	**0.663**	spa_Latn	**0.663**	spa_Latn
mal_Mlym	0.591	0.642	arb_Arab	0.591	eng_Latn	**0.709**	hin_Deva	**0.709**	hin_Deva	**0.709**	hin_Deva
mar_Deva	0.583	**0.725**	hin_Deva	**0.725**	hin_Deva	**0.725**	hin_Deva	**0.725**	hin_Deva	**0.725**	hin_Deva
min_Latn	0.405	0.405	eng_Latn	0.405	eng_Latn	0.363	hin_Deva	0.405	eng_Latn	**0.423**	spa_Latn
mkd_Cyrl	0.696	**0.767**	rus_Cyrl	**0.767**	rus_Cyrl	0.730	spa_Latn	**0.767**	rus_Cyrl	**0.767**	rus_Cyrl
mlt_Latn	0.667	0.667	eng_Latn	0.597	arb_Arab	**0.732**	spa_Latn	0.641	rus_Cyrl	**0.732**	spa_Latn
mri_Latn	0.531	0.531	eng_Latn	0.531	eng_Latn	0.433	cmn_Hani	0.531	eng_Latn	**0.572**	spa_Latn
mya_Mymr	0.493	**0.612**	arb_Arab	0.455	cmn_Hani	0.607	hin_Deva	0.493	eng_Latn	0.607	hin_Deva
nld_Latn	0.779	0.779	eng_Latn	0.779	eng_Latn	0.779	eng_Latn	0.779	eng_Latn	**0.781**	spa_Latn
nno_Latn	**0.762**	**0.762**	eng_Latn	**0.762**	eng_Latn	**0.762**	eng_Latn	0.686	hin_Deva	**0.762**	eng_Latn
oci_Latn	0.678	**0.802**	spa_Latn	**0.802**	spa_Latn	**0.802**	spa_Latn	**0.802**	spa_Latn	**0.802**	spa_Latn
ory_Orya	0.230	0.262	arb_Arab	**0.300**	hin_Deva	0.230	hin_Deva	**0.300**	hin_Deva	**0.300**	hin_Deva
pan_Guru	0.464	**0.470**	hin_Deva	**0.470**	hin_Deva	**0.470**	hin_Deva	**0.470**	hin_Deva	**0.470**	hin_Deva
pes_Arab	0.386	0.606	arb_Arab	**0.653**	hin_Deva	0.606	arb_Arab	**0.653**	hin_Deva	0.606	arb_Arab
plt_Latn	**0.533**	**0.533**	eng_Latn	**0.533**	eng_Latn	0.424	arb_Arab	0.510	rus_Cyrl	0.507	spa_Latn
pol_Latn	**0.754**	**0.754**	eng_Latn	0.719	rus_Cyrl	**0.754**	eng_Latn	0.719	rus_Cyrl	0.719	rus_Cyrl
por_Latn	0.745	**0.803**	spa_Latn	**0.803**	spa_Latn	**0.803**	spa_Latn	0.745	eng_Latn	**0.803**	spa_Latn
ron_Latn	0.632	0.632	eng_Latn	**0.746**	spa_Latn	0.632	eng_Latn	0.614	rus_Cyrl	**0.746**	spa_Latn
san_Deva	0.306	**0.523**	hin_Deva	**0.523**	hin_Deva	**0.523**	hin_Deva	**0.523**	hin_Deva	**0.523**	hin_Deva
scn_Latn	0.676	0.676	eng_Latn	**0.750**	spa_Latn	**0.750**	spa_Latn	0.623	arb_Arab	**0.750**	spa_Latn
sin_Sinh	0.536	0.560	arb_Arab	**0.727**	hin_Deva	**0.727**	hin_Deva	**0.727**	hin_Deva	**0.727**	hin_Deva
slk_Latn	**0.745**	**0.745**	eng_Latn	0.721	rus_Cyrl	**0.745**	eng_Latn	0.659	hin_Deva	0.721	rus_Cyrl
slv_Latn	**0.766**	**0.766**	eng_Latn	0.724	rus_Cyrl	**0.766**	eng_Latn	0.724	rus_Cyrl	0.724	rus_Cyrl
snd_Arab	0.374	0.441	arb_Arab	**0.530**	hin_Deva	**0.530**	hin_Deva	**0.530**	hin_Deva	0.441	arb_Arab
som_Latn	0.598	0.598	eng_Latn	0.562	arb_Arab	0.562	arb_Arab	0.579	hin_Deva	**0.605**	spa_Latn
srp_Cyrl	**0.627**	0.586	rus_Cyrl	0.586	rus_Cyrl	**0.627**	eng_Latn	0.586	rus_Cyrl	0.586	rus_Cyrl
sun_Latn	**0.577**	**0.577**	eng_Latn	**0.577**	eng_Latn	0.492	hin_Deva	**0.577**	eng_Latn	0.490	spa_Latn
swe_Latn	**0.632**	**0.632**	eng_Latn	**0.632**	eng_Latn	**0.632**	eng_Latn	**0.632**	eng_Latn	**0.632**	eng_Latn
swh_Latn	**0.687**	**0.687**	eng_Latn	**0.687**	eng_Latn	0.503	arb_Arab	0.662	spa_Latn	0.662	spa_Latn
szl_Latn	**0.670**	**0.670**	eng_Latn	0.655	rus_Cyrl	**0.670**	eng_Latn	0.631	hin_Deva	0.655	rus_Cyrl
tam_Taml	0.498	0.597	arb_Arab	0.498	eng_Latn	**0.626**	hin_Deva	**0.626**	hin_Deva	**0.626**	hin_Deva
tat_Cyrl	0.630	**0.715**	rus_Cyrl	0.630	eng_Latn	**0.715**	rus_Cyrl	0.672	arb_Arab	**0.715**	rus_Cyrl
tel_Telu	0.420	0.516	arb_Arab	0.420	eng_Latn	**0.539**	hin_Deva	**0.539**	hin_Deva	**0.539**	hin_Deva
tgk_Cyrl	0.588	**0.652**	rus_Cyrl	0.598	hin_Deva	**0.652**	rus_Cyrl	0.629	arb_Arab	**0.652**	rus_Cyrl
tgl_Latn	**0.745**	**0.745**	eng_Latn	**0.745**	eng_Latn	0.466	cmn_Hani	0.667	spa_Latn	0.667	spa_Latn
tha_Thai	0.049	**0.074**	cmn_Hani	0.049	eng_Latn	0.014	hin_Deva	0.049	eng_Latn	**0.074**	cmn_Hani
tuk_Latn	0.577	0.577	eng_Latn	0.577	eng_Latn	0.579	arb_Arab	0.553	cmn_Hani	**0.615**	spa_Latn
tur_Latn	0.712	0.712	eng_Latn	0.712	eng_Latn	0.707	arb_Arab	0.707	rus_Cyrl	**0.758**	spa_Latn
uig_Arab	0.460	**0.547**	arb_Arab	0.460	eng_Latn	0.525	rus_Cyrl	0.485	cmn_Hani	**0.547**	arb_Arab
ukr_Cyrl	0.695	**0.802**	rus_Cyrl	**0.802**	rus_Cyrl	0.695	eng_Latn	**0.802**	rus_Cyrl	**0.802**	rus_Cyrl
urd_Arab	0.596	0.689	arb_Arab	**0.743**	hin_Deva	**0.743**	hin_Deva	**0.743**	hin_Deva	**0.743**	hin_Deva
uzn_Latn	0.713	0.713	eng_Latn	0.713	eng_Latn	0.716	rus_Cyrl	0.479	hin_Deva	**0.792**	spa_Latn
vec_Latn	0.624	0.624	eng_Latn	**0.680**	spa_Latn	**0.680**	spa_Latn	0.549	hin_Deva	**0.680**	spa_Latn
vie_Latn	**0.654**	**0.654**	eng_Latn	**0.654**	eng_Latn	0.406	cmn_Hani	**0.654**	eng_Latn	0.546	rus_Cyrl
war_Latn	0.554	0.554	eng_Latn	0.554	eng_Latn	0.425	cmn_Hani	0.425	cmn_Hani	**0.585**	spa_Latn
ydd_Hebr	0.496	0.496	eng_Latn	0.496	eng_Latn	0.496	eng_Latn	**0.609**	hin_Deva	0.569	arb_Arab
yor_Latn	**0.614**	**0.614**	eng_Latn	**0.614**	eng_Latn	0.612	spa_Latn	0.532	rus_Cyrl	0.612	spa_Latn
yue_Hani	0.261	**0.635**	cmn_Hani	**0.635**	cmn_Hani	**0.635**	cmn_Hani	**0.635**	cmn_Hani	**0.635**	cmn_Hani
zsm_Latn	**0.654**	**0.654**	eng_Latn	**0.654**	eng_Latn	0.522	hin_Deva	**0.654**	eng_Latn	**0.654**	eng_Latn

Table 16: Cross-Lingual Transfer Results of NER (Part 2): The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 1).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
afr_Latn	0.850	0.850	eng_Latn	0.850	eng_Latn	0.599	arb_Arab	0.809	rus_Cyrl	**0.854**	spa_Latn
ajp_Arab	**0.671**	0.648	arb_Arab	0.648	arb_Arab	0.648	arb_Arab	0.651	hin_Deva	0.648	arb_Arab
amh_Ethi	0.648	0.645	cmn_Hani	0.670	arb_Arab	0.670	arb_Arab	**0.704**	hin_Deva	**0.704**	hin_Deva
bam_Latn	0.451	0.451	eng_Latn	0.451	eng_Latn	0.411	spa_Latn	**0.484**	hin_Deva	0.411	spa_Latn
bel_Cyrl	0.824	**0.934**	rus_Cyrl	**0.934**	rus_Cyrl	0.824	eng_Latn	0.719	arb_Arab	**0.934**	rus_Cyrl
ben_Beng	0.767	0.583	arb_Arab	**0.803**	hin_Deva	**0.803**	hin_Deva	**0.803**	hin_Deva	**0.803**	hin_Deva
bho_Deva	0.520	**0.682**	hin_Deva	**0.682**	hin_Deva	**0.682**	hin_Deva	0.536	arb_Arab	**0.682**	hin_Deva
bul_Cyrl	0.871	**0.899**	rus_Cyrl	**0.899**	rus_Cyrl	0.882	spa_Latn	**0.899**	rus_Cyrl	**0.899**	rus_Cyrl
cat_Latn	0.860	**0.962**	spa_Latn	**0.962**	spa_Latn	**0.962**	spa_Latn	0.860	eng_Latn	**0.962**	spa_Latn
ceb_Latn	0.605	0.605	eng_Latn	0.605	eng_Latn	0.481	cmn_Hani	**0.634**	spa_Latn	**0.634**	spa_Latn
ces_Latn	0.826	0.826	eng_Latn	**0.874**	rus_Cyrl	0.826	eng_Latn	**0.874**	rus_Cyrl	**0.874**	rus_Cyrl
cym_Latn	**0.621**	**0.621**	eng_Latn	0.612	rus_Cyrl	**0.621**	eng_Latn	0.602	arb_Arab	0.618	spa_Latn
dan_Latn	**0.873**	**0.873**	eng_Latn	**0.873**	eng_Latn	**0.873**	eng_Latn	0.640	arb_Arab	**0.873**	eng_Latn
deu_Latn	**0.850**	**0.850**	eng_Latn	**0.850**	eng_Latn	**0.850**	eng_Latn	**0.850**	eng_Latn	0.784	spa_Latn
ekk_Latn	**0.815**	**0.815**	eng_Latn	**0.815**	eng_Latn	**0.815**	eng_Latn	0.790	rus_Cyrl	0.790	rus_Cyrl
ell_Grek	0.822	0.822	eng_Latn	**0.871**	rus_Cyrl	0.834	spa_Latn	**0.871**	rus_Cyrl	**0.871**	rus_Cyrl
eus_Latn	0.625	0.625	eng_Latn	0.625	eng_Latn	0.681	spa_Latn	**0.702**	hin_Deva	0.681	spa_Latn
fao_Latn	0.869	0.869	eng_Latn	0.869	eng_Latn	0.869	eng_Latn	0.701	arb_Arab	**0.876**	spa_Latn
fin_Latn	0.771	0.771	eng_Latn	0.771	eng_Latn	0.771	eng_Latn	**0.773**	rus_Cyrl	**0.773**	rus_Cyrl
fra_Latn	0.838	0.838	eng_Latn	**0.885**	spa_Latn	0.838	eng_Latn	0.838	eng_Latn	**0.885**	spa_Latn
gla_Latn	0.571	0.571	eng_Latn	**0.588**	rus_Cyrl	0.571	eng_Latn	0.498	arb_Arab	0.548	spa_Latn
gle_Latn	0.578	0.578	eng_Latn	**0.624**	rus_Cyrl	0.578	eng_Latn	0.624	spa_Latn	0.624	spa_Latn
glg_Latn	0.796	**0.864**	spa_Latn	**0.864**	spa_Latn	**0.864**	spa_Latn	**0.864**	spa_Latn	**0.864**	spa_Latn
gug_Latn	0.213	0.213	eng_Latn	0.213	eng_Latn	**0.256**	spa_Latn	**0.256**	spa_Latn	**0.256**	spa_Latn
heb_Hebr	0.636	0.560	cmn_Hani	0.696	arb_Arab	0.696	arb_Arab	**0.704**	rus_Cyrl	0.696	arb_Arab
hin_Deva	**0.665**	0.612	arb_Arab	0.612	arb_Arab	0.612	arb_Arab	0.612	arb_Arab	0.612	arb_Arab
hrv_Latn	0.829	0.829	eng_Latn	**0.899**	rus_Cyrl	0.829	eng_Latn	**0.899**	rus_Cyrl	**0.899**	rus_Cyrl
hun_Latn	0.801	0.801	eng_Latn	0.801	eng_Latn	0.801	eng_Latn	0.740	rus_Cyrl	**0.811**	spa_Latn
hye_Armn	0.817	0.595	arb_Arab	0.817	eng_Latn	0.595	arb_Arab	**0.846**	rus_Cyrl	**0.846**	rus_Cyrl
ind_Latn	**0.814**	**0.814**	eng_Latn	**0.814**	eng_Latn	0.695	hin_Deva	**0.814**	eng_Latn	**0.814**	eng_Latn
isl_Latn	**0.805**	**0.805**	eng_Latn	**0.805**	eng_Latn	**0.805**	eng_Latn	**0.805**	eng_Latn	0.802	spa_Latn
ita_Latn	0.852	**0.906**	spa_Latn	**0.906**	spa_Latn	**0.906**	spa_Latn	**0.906**	spa_Latn	**0.906**	spa_Latn
jav_Latn	**0.742**	**0.742**	eng_Latn	**0.742**	eng_Latn	0.543	cmn_Hani	0.645	hin_Deva	0.731	spa_Latn
jpn_Jpan	0.165	**0.534**	cmn_Hani	0.165	eng_Latn	**0.534**	cmn_Hani	0.402	hin_Deva	**0.534**	cmn_Hani
kaz_Cyrl	0.724	**0.739**	rus_Cyrl	0.724	eng_Latn	**0.739**	rus_Cyrl	0.545	cmn_Hani	**0.739**	rus_Cyrl
kmr_Latn	0.748	0.748	eng_Latn	0.719	hin_Deva	0.646	arb_Arab	0.748	eng_Latn	**0.777**	spa_Latn
kor_Hang	**0.497**	0.447	cmn_Hani	**0.497**	eng_Latn	0.447	cmn_Hani	0.447	cmn_Hani	0.491	hin_Deva
lij_Latn	0.739	0.739	eng_Latn	**0.819**	spa_Latn	**0.819**	spa_Latn	0.685	hin_Deva	**0.819**	spa_Latn
lit_Latn	0.787	0.787	eng_Latn	**0.840**	rus_Cyrl	0.787	eng_Latn	**0.840**	rus_Cyrl	**0.840**	rus_Cyrl
mal_Mlym	**0.847**	0.680	arb_Arab	**0.847**	eng_Latn	0.804	hin_Deva	0.804	hin_Deva	0.804	hin_Deva
mar_Deva	0.813	**0.830**	hin_Deva	**0.830**	hin_Deva	**0.830**	hin_Deva	**0.830**	hin_Deva	**0.830**	hin_Deva
mlt_Latn	0.776	0.776	eng_Latn	0.603	arb_Arab	**0.798**	spa_Latn	0.787	rus_Cyrl	**0.798**	spa_Latn
nld_Latn	**0.874**	**0.874**	eng_Latn	**0.874**	eng_Latn	**0.874**	eng_Latn	**0.874**	eng_Latn	0.855	spa_Latn
pes_Arab	0.675	0.690	arb_Arab	**0.709**	hin_Deva	0.690	arb_Arab	**0.709**	hin_Deva	0.690	arb_Arab
pol_Latn	0.791	0.791	eng_Latn	**0.881**	rus_Cyrl	0.791	eng_Latn	**0.881**	rus_Cyrl	**0.881**	rus_Cyrl
por_Latn	0.857	**0.910**	spa_Latn	**0.910**	spa_Latn	**0.910**	spa_Latn	0.857	eng_Latn	**0.910**	spa_Latn
ron_Latn	0.747	0.747	eng_Latn	**0.816**	spa_Latn	0.747	eng_Latn	0.794	rus_Cyrl	**0.816**	spa_Latn
san_Deva	0.217	**0.319**	hin_Deva	**0.319**	hin_Deva	**0.319**	hin_Deva	**0.319**	hin_Deva	**0.319**	hin_Deva
sin_Sinh	0.546	0.520	arb_Arab	**0.652**	hin_Deva	**0.652**	hin_Deva	**0.652**	hin_Deva	**0.652**	hin_Deva
slk_Latn	0.820	0.820	eng_Latn	**0.865**	rus_Cyrl	0.820	eng_Latn	0.743	hin_Deva	**0.865**	rus_Cyrl
slv_Latn	0.743	0.743	eng_Latn	**0.805**	rus_Cyrl	0.743	eng_Latn	**0.805**	rus_Cyrl	**0.805**	rus_Cyrl
swe_Latn	**0.891**	**0.891**	eng_Latn	**0.891**	eng_Latn	**0.891**	eng_Latn	**0.891**	eng_Latn	**0.891**	eng_Latn
tam_Taml	0.733	0.586	arb_Arab	0.733	eng_Latn	**0.771**	hin_Deva	**0.771**	hin_Deva	**0.771**	hin_Deva
tat_Cyrl	0.675	**0.692**	rus_Cyrl	0.675	eng_Latn	**0.692**	rus_Cyrl	0.587	arb_Arab	**0.692**	rus_Cyrl
tel_Telu	**0.791**	0.653	arb_Arab	**0.791**	eng_Latn	0.781	hin_Deva	0.781	hin_Deva	0.781	hin_Deva
tgl_Latn	0.695	0.695	eng_Latn	0.695	eng_Latn	0.416	cmn_Hani	**0.719**	spa_Latn	**0.719**	spa_Latn
tha_Thai	**0.502**	0.499	cmn_Hani	**0.502**	eng_Latn	0.453	hin_Deva	**0.502**	eng_Latn	0.499	cmn_Hani
tur_Latn	0.671	0.671	eng_Latn	0.671	eng_Latn	0.522	arb_Arab	0.671	rus_Cyrl	**0.697**	spa_Latn
uig_Arab	0.660	0.536	arb_Arab	0.660	eng_Latn	0.670	rus_Cyrl	0.525	cmn_Hani	**0.687**	hin_Deva
ukr_Cyrl	0.821	**0.918**	rus_Cyrl	**0.918**	rus_Cyrl	0.821	eng_Latn	**0.918**	rus_Cyrl	**0.918**	rus_Cyrl
urd_Arab	0.589	0.580	arb_Arab	**0.889**	hin_Deva	**0.889**	hin_Deva	**0.889**	hin_Deva	**0.889**	hin_Deva
vie_Latn	0.648	0.648	eng_Latn	0.648	eng_Latn	0.442	cmn_Hani	0.648	eng_Latn	**0.658**	rus_Cyrl
wol_Latn	0.606	0.606	eng_Latn	0.606	eng_Latn	**0.679**	spa_Latn	0.606	eng_Latn	**0.679**	spa_Latn
yor_Latn	0.644	0.644	eng_Latn	0.644	eng_Latn	0.651	spa_Latn	**0.658**	rus_Cyrl	0.651	spa_Latn
yue_Hani	0.196	**0.787**	cmn_Hani	**0.787**	cmn_Hani	**0.787**	cmn_Hani	**0.787**	cmn_Hani	**0.787**	cmn_Hani

Table 17: Cross-Lingual Transfer Results of POS: The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 2).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
afr_Latn	**0.732**	**0.732**	eng_Latn	**0.732**	eng_Latn	0.589	arb_Arab	0.701	rus_Cyrl	**0.732**	eng_Latn
als_Latn	0.708	0.708	eng_Latn	0.721	rus_Cyrl	**0.727**	spa_Latn	**0.727**	spa_Latn	**0.727**	spa_Latn
amh_Ethi	0.557	0.470	cmn_Hani	0.532	arb_Arab	0.532	arb_Arab	**0.611**	hin_Deva	**0.611**	hin_Deva
azj_Latn	0.773	0.773	eng_Latn	0.773	eng_Latn	0.705	arb_Arab	**0.793**	hin_Deva	**0.793**	hin_Deva
ben_Beng	0.676	0.625	arb_Arab	**0.768**	hin_Deva	**0.768**	hin_Deva	**0.768**	hin_Deva	**0.768**	hin_Deva
cat_Latn	0.731	**0.833**	spa_Latn	**0.833**	spa_Latn	**0.833**	spa_Latn	0.731	eng_Latn	**0.833**	spa_Latn
cym_Latn	0.492	0.492	eng_Latn	**0.495**	rus_Cyrl	0.492	eng_Latn	0.433	arb_Arab	0.480	spa_Latn
dan_Latn	**0.838**	**0.838**	eng_Latn	**0.838**	eng_Latn	**0.838**	eng_Latn	0.720	arb_Arab	**0.838**	eng_Latn
deu_Latn	**0.759**	**0.759**	eng_Latn	**0.759**	eng_Latn	**0.759**	eng_Latn	**0.759**	eng_Latn	0.726	spa_Latn
ell_Grek	0.715	0.715	eng_Latn	**0.729**	rus_Cyrl	0.717	spa_Latn	**0.729**	rus_Cyrl	**0.729**	rus_Cyrl
fin_Latn	0.677	0.677	eng_Latn	0.677	eng_Latn	0.677	eng_Latn	**0.701**	rus_Cyrl	**0.701**	rus_Cyrl
fra_Latn	0.812	0.812	eng_Latn	**0.816**	spa_Latn	0.812	eng_Latn	0.812	eng_Latn	**0.816**	spa_Latn
heb_Hebr	0.697	0.576	cmn_Hani	0.691	arb_Arab	0.691	arb_Arab	**0.714**	rus_Cyrl	0.691	arb_Arab
hun_Latn	0.673	0.673	eng_Latn	0.673	eng_Latn	0.673	eng_Latn	**0.698**	rus_Cyrl	**0.698**	rus_Cyrl
hye_Armn	**0.781**	0.729	arb_Arab	**0.781**	eng_Latn	0.729	arb_Arab	0.780	rus_Cyrl	0.780	rus_Cyrl
ind_Latn	**0.819**	**0.819**	eng_Latn	**0.819**	eng_Latn	0.779	hin_Deva	**0.819**	eng_Latn	**0.819**	eng_Latn
isl_Latn	0.658	0.658	eng_Latn	0.658	eng_Latn	0.658	eng_Latn	0.658	eng_Latn	**0.664**	rus_Cyrl
ita_Latn	0.772	**0.817**	spa_Latn	**0.817**	spa_Latn	**0.817**	spa_Latn	**0.817**	spa_Latn	**0.817**	spa_Latn
jav_Latn	**0.507**	**0.507**	eng_Latn	**0.507**	eng_Latn	0.416	cmn_Hani	0.504	hin_Deva	0.495	spa_Latn
jpn_Jpan	0.384	**0.448**	cmn_Hani	0.384	eng_Latn	**0.448**	cmn_Hani	0.363	hin_Deva	**0.448**	cmn_Hani
kan_Knda	0.682	0.628	arb_Arab	0.682	eng_Latn	**0.729**	hin_Deva	**0.729**	hin_Deva	**0.729**	hin_Deva
kat_Geor	0.618	0.605	arb_Arab	0.618	eng_Latn	0.605	arb_Arab	**0.620**	hin_Deva	**0.620**	hin_Deva
khm_Khmr	**0.655**	**0.655**	eng_Latn	**0.655**	eng_Latn	0.636	hin_Deva	**0.655**	eng_Latn	0.611	arb_Arab
kor_Hang	0.758	0.643	cmn_Hani	0.758	eng_Latn	0.643	cmn_Hani	0.643	cmn_Hani	**0.768**	hin_Deva
lvs_Latn	0.661	0.661	eng_Latn	0.661	eng_Latn	0.661	eng_Latn	0.651	hin_Deva	**0.722**	rus_Cyrl
mal_Mlym	0.717	0.678	arb_Arab	0.717	eng_Latn	**0.764**	hin_Deva	**0.764**	hin_Deva	**0.764**	hin_Deva
mya_Mymr	0.688	0.656	arb_Arab	0.616	cmn_Hani	**0.707**	hin_Deva	0.688	eng_Latn	**0.707**	hin_Deva
nld_Latn	**0.813**	**0.813**	eng_Latn	**0.813**	eng_Latn	**0.813**	eng_Latn	**0.813**	eng_Latn	**0.813**	eng_Latn
nob_Latn	**0.847**	**0.847**	eng_Latn	**0.847**	eng_Latn	**0.847**	eng_Latn	**0.847**	eng_Latn	**0.847**	eng_Latn
pes_Arab	**0.831**	0.780	arb_Arab	0.817	hin_Deva	0.780	arb_Arab	0.817	hin_Deva	0.817	hin_Deva
pol_Latn	0.768	0.768	eng_Latn	**0.788**	rus_Cyrl	0.768	eng_Latn	**0.788**	rus_Cyrl	**0.788**	rus_Cyrl
por_Latn	0.793	**0.839**	spa_Latn	**0.839**	spa_Latn	**0.839**	spa_Latn	0.793	eng_Latn	**0.839**	spa_Latn
ron_Latn	0.791	0.791	eng_Latn	**0.814**	spa_Latn	0.791	eng_Latn	0.790	rus_Cyrl	**0.814**	spa_Latn
slv_Latn	0.643	0.643	eng_Latn	**0.720**	rus_Cyrl	0.643	eng_Latn	**0.720**	rus_Cyrl	**0.720**	rus_Cyrl
swe_Latn	**0.834**	**0.834**	eng_Latn	**0.834**	eng_Latn	**0.834**	eng_Latn	**0.834**	eng_Latn	**0.834**	eng_Latn
swh_Latn	0.465	0.465	eng_Latn	0.465	eng_Latn	0.468	arb_Arab	**0.499**	spa_Latn	**0.499**	spa_Latn
tam_Taml	0.698	0.657	arb_Arab	0.698	eng_Latn	**0.737**	hin_Deva	**0.737**	hin_Deva	**0.737**	hin_Deva
tel_Telu	0.695	0.657	arb_Arab	0.695	eng_Latn	**0.756**	hin_Deva	**0.756**	hin_Deva	**0.756**	hin_Deva
tgl_Latn	**0.752**	**0.752**	eng_Latn	**0.752**	eng_Latn	0.648	cmn_Hani	0.723	spa_Latn	0.723	spa_Latn
tha_Thai	**0.791**	0.714	cmn_Hani	**0.791**	eng_Latn	0.752	hin_Deva	**0.791**	eng_Latn	0.714	cmn_Hani
tur_Latn	0.747	0.747	eng_Latn	0.747	eng_Latn	0.650	arb_Arab	0.731	rus_Cyrl	**0.786**	hin_Deva
urd_Arab	0.716	0.686	arb_Arab	**0.806**	hin_Deva	**0.806**	hin_Deva	**0.806**	hin_Deva	**0.806**	hin_Deva
vie_Latn	**0.771**	**0.771**	eng_Latn	**0.771**	eng_Latn	0.680	cmn_Hani	**0.771**	eng_Latn	**0.771**	eng_Latn
zsm_Latn	**0.754**	**0.754**	eng_Latn	**0.754**	eng_Latn	0.731	hin_Deva	**0.754**	eng_Latn	**0.754**	eng_Latn

Table 18: Cross-Lingual Transfer Result of MASSIVE: The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 8).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
ace_Latn	0.624	0.624	eng_Latn	0.624	eng_Latn	**0.726**	hin_Deva	0.624	eng_Latn	0.654	spa_Latn
afr_Latn	0.600	0.600	eng_Latn	0.600	eng_Latn	0.455	arb_Arab	0.522	rus_Cyrl	**0.604**	spa_Latn
aka_Latn	**0.518**	**0.518**	eng_Latn	**0.518**	eng_Latn	0.471	spa_Latn	0.469	hin_Deva	0.471	spa_Latn
als_Latn	**0.575**	**0.575**	eng_Latn	0.557	rus_Cyrl	0.536	spa_Latn	0.557	rus_Cyrl	0.536	spa_Latn
ary_Arab	0.421	**0.484**	arb_Arab	**0.484**	arb_Arab	0.465	spa_Latn	0.421	eng_Latn	**0.484**	arb_Arab
arz_Arab	0.325	**0.430**	arb_Arab	**0.430**	arb_Arab	**0.430**	arb_Arab	0.325	eng_Latn	**0.430**	arb_Arab
asm_Beng	0.574	0.548	arb_Arab	**0.600**	hin_Deva	**0.600**	hin_Deva	**0.600**	hin_Deva	**0.600**	hin_Deva
ayr_Latn	**0.694**	**0.694**	eng_Latn	**0.694**	eng_Latn	0.645	spa_Latn	0.564	cmn_Hani	0.685	hin_Deva
azb_Arab	0.527	0.585	arb_Arab	0.527	eng_Latn	0.585	arb_Arab	**0.639**	hin_Deva	**0.639**	hin_Deva
bak_Cyrl	0.632	**0.667**	rus_Cyrl	0.632	eng_Latn	**0.667**	rus_Cyrl	0.635	hin_Deva	**0.667**	rus_Cyrl
bam_Latn	0.487	0.487	eng_Latn	0.487	eng_Latn	**0.617**	spa_Latn	0.531	hin_Deva	**0.617**	spa_Latn
ban_Latn	0.446	0.446	eng_Latn	0.446	eng_Latn	0.483	cmn_Hani	**0.497**	hin_Deva	0.489	spa_Latn
bel_Cyrl	**0.622**	0.571	rus_Cyrl	0.571	rus_Cyrl	**0.622**	eng_Latn	0.530	arb_Arab	0.571	rus_Cyrl
bem_Latn	0.418	0.418	eng_Latn	0.418	eng_Latn	0.477	arb_Arab	**0.517**	spa_Latn	**0.517**	spa_Latn
ben_Beng	**0.667**	0.568	arb_Arab	0.634	hin_Deva	0.634	hin_Deva	0.634	hin_Deva	0.634	hin_Deva
bul_Cyrl	0.612	**0.618**	rus_Cyrl	**0.618**	rus_Cyrl	0.574	spa_Latn	**0.618**	rus_Cyrl	**0.618**	rus_Cyrl
cat_Latn	0.496	**0.614**	spa_Latn	**0.614**	spa_Latn	**0.614**	spa_Latn	0.496	eng_Latn	**0.614**	spa_Latn
ceb_Latn	0.565	0.565	eng_Latn	0.565	eng_Latn	**0.565**	cmn_Hani	0.456	spa_Latn	0.456	spa_Latn
ces_Latn	**0.620**	**0.620**	eng_Latn	0.577	rus_Cyrl	**0.620**	eng_Latn	0.577	rus_Cyrl	0.577	rus_Cyrl
ckb_Arab	0.544	0.539	arb_Arab	**0.622**	hin_Deva	0.539	arb_Arab	0.589	rus_Cyrl	0.539	arb_Arab
cym_Latn	0.488	0.488	eng_Latn	0.435	rus_Cyrl	0.488	eng_Latn	0.469	arb_Arab	**0.501**	spa_Latn
dan_Latn	**0.556**	**0.556**	eng_Latn	**0.556**	eng_Latn	**0.556**	eng_Latn	0.401	arb_Arab	**0.556**	eng_Latn
deu_Latn	0.559	0.559	eng_Latn	0.559	eng_Latn	0.559	eng_Latn	0.559	eng_Latn	**0.561**	spa_Latn
dyu_Latn	0.520	0.520	eng_Latn	0.520	eng_Latn	**0.587**	spa_Latn	0.568	hin_Deva	**0.587**	spa_Latn
dzo_Tibt	0.495	0.612	arb_Arab	**0.682**	cmn_Hani	0.681	hin_Deva	0.681	hin_Deva	0.681	hin_Deva
ell_Grek	0.532	0.532	eng_Latn	**0.547**	rus_Cyrl	0.485	spa_Latn	**0.547**	rus_Cyrl	**0.547**	rus_Cyrl
epo_Latn	**0.548**	**0.548**	eng_Latn	**0.548**	eng_Latn	**0.548**	eng_Latn	0.511	rus_Cyrl	0.530	spa_Latn
eus_Latn	0.196	0.196	eng_Latn	0.196	eng_Latn	**0.299**	spa_Latn	0.268	hin_Deva	**0.299**	spa_Latn
ewe_Latn	0.480	0.480	eng_Latn	0.480	eng_Latn	**0.589**	spa_Latn	0.530	hin_Deva	**0.589**	spa_Latn
fao_Latn	**0.658**	**0.658**	eng_Latn	**0.658**	eng_Latn	**0.658**	eng_Latn	0.591	arb_Arab	0.526	spa_Latn
fij_Latn	0.512	0.512	eng_Latn	0.512	eng_Latn	0.525	cmn_Hani	**0.576**	spa_Latn	**0.576**	spa_Latn
fin_Latn	0.465	0.465	eng_Latn	0.465	eng_Latn	0.465	eng_Latn	**0.518**	rus_Cyrl	**0.518**	rus_Cyrl
fon_Latn	0.462	0.462	eng_Latn	0.462	eng_Latn	**0.562**	spa_Latn	0.462	eng_Latn	**0.562**	spa_Latn
fra_Latn	0.566	0.566	eng_Latn	**0.627**	spa_Latn	0.566	eng_Latn	0.566	eng_Latn	**0.627**	spa_Latn
gla_Latn	0.489	0.489	eng_Latn	0.476	rus_Cyrl	0.489	eng_Latn	0.464	arb_Arab	**0.503**	spa_Latn
gle_Latn	0.375	0.375	eng_Latn	0.387	rus_Cyrl	0.375	eng_Latn	**0.502**	spa_Latn	**0.502**	spa_Latn
gug_Latn	0.396	0.396	eng_Latn	0.396	eng_Latn	**0.561**	spa_Latn	**0.561**	spa_Latn	**0.561**	spa_Latn
guj_Gujr	**0.717**	0.646	arb_Arab	0.680	hin_Deva	0.680	hin_Deva	0.680	hin_Deva	0.680	hin_Deva
hat_Latn	0.571	0.571	eng_Latn	**0.644**	spa_Latn	0.571	eng_Latn	0.584	arb_Arab	**0.644**	spa_Latn
hau_Latn	0.486	0.486	eng_Latn	**0.560**	arb_Arab	0.550	spa_Latn	0.486	eng_Latn	0.550	spa_Latn
heb_Hebr	**0.398**	0.391	cmn_Hani	0.359	arb_Arab	0.359	arb_Arab	0.373	rus_Cyrl	0.359	arb_Arab
hin_Deva	**0.705**	0.618	arb_Arab	0.618	arb_Arab	0.618	arb_Arab	0.618	arb_Arab	0.618	arb_Arab
hne_Deva	0.708	**0.711**	hin_Deva	**0.711**	hin_Deva	**0.711**	hin_Deva	**0.711**	hin_Deva	**0.711**	hin_Deva
hrv_Latn	0.569	0.569	eng_Latn	**0.680**	rus_Cyrl	0.569	eng_Latn	**0.680**	rus_Cyrl	**0.680**	rus_Cyrl
hun_Latn	0.540	0.540	eng_Latn	0.540	eng_Latn	0.540	eng_Latn	**0.609**	rus_Cyrl	**0.609**	rus_Cyrl

Table 19: Cross-Lingual Transfer Results of Taxi1500 (Part 1): The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 4).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
hye_Armn	0.650	**0.678**	arb_Arab	0.650	eng_Latn	**0.678**	arb_Arab	0.654	rus_Cyrl	0.654	rus_Cyrl
ibo_Latn	0.544	0.544	eng_Latn	0.544	eng_Latn	**0.566**	spa_Latn	0.544	eng_Latn	**0.566**	spa_Latn
ilo_Latn	0.511	0.511	eng_Latn	0.511	eng_Latn	0.463	cmn_Hani	0.511	eng_Latn	**0.591**	spa_Latn
ind_Latn	0.720	0.720	eng_Latn	0.720	eng_Latn	**0.795**	hin_Deva	0.720	eng_Latn	0.720	eng_Latn
isl_Latn	0.497	0.497	eng_Latn	0.497	eng_Latn	0.497	eng_Latn	0.497	eng_Latn	**0.602**	spa_Latn
ita_Latn	**0.608**	0.593	spa_Latn	0.593	spa_Latn	0.593	spa_Latn	0.593	spa_Latn	0.593	spa_Latn
jav_Latn	0.445	0.445	eng_Latn	0.445	eng_Latn	0.428	cmn_Hani	0.441	hin_Deva	**0.516**	spa_Latn
kab_Latn	0.259	0.259	eng_Latn	0.368	arb_Arab	**0.396**	spa_Latn	0.259	eng_Latn	**0.396**	spa_Latn
kac_Latn	0.451	0.451	eng_Latn	**0.580**	cmn_Hani	0.483	hin_Deva	**0.580**	cmn_Hani	0.483	hin_Deva
kan_Knda	**0.673**	0.637	arb_Arab	**0.673**	eng_Latn	0.640	hin_Deva	0.640	hin_Deva	0.640	hin_Deva
kat_Geor	0.558	0.464	arb_Arab	0.558	eng_Latn	0.464	arb_Arab	**0.672**	hin_Deva	**0.672**	hin_Deva
kaz_Cyrl	0.587	**0.636**	rus_Cyrl	0.587	eng_Latn	**0.636**	rus_Cyrl	0.629	hin_Deva	**0.636**	rus_Cyrl
kbp_Latn	0.357	0.357	eng_Latn	0.357	eng_Latn	0.361	spa_Latn	0.357	eng_Latn	**0.378**	hin_Deva
khm_Khmr	0.653	0.653	eng_Latn	0.653	eng_Latn	**0.679**	hin_Deva	0.653	eng_Latn	**0.679**	hin_Deva
kik_Latn	0.384	0.384	eng_Latn	0.384	eng_Latn	0.456	arb_Arab	**0.555**	spa_Latn	**0.555**	spa_Latn
kin_Latn	0.431	0.431	eng_Latn	0.431	eng_Latn	0.530	arb_Arab	0.431	eng_Latn	**0.619**	spa_Latn
kir_Cyrl	0.623	0.601	rus_Cyrl	0.623	eng_Latn	0.601	rus_Cyrl	**0.750**	hin_Deva	0.601	rus_Cyrl
kng_Latn	0.353	0.353	eng_Latn	0.353	eng_Latn	**0.455**	arb_Arab	**0.455**	arb_Arab	0.381	spa_Latn
kor_Hang	0.614	0.602	cmn_Hani	0.614	eng_Latn	0.602	cmn_Hani	0.602	cmn_Hani	**0.686**	hin_Deva
lao_Laoo	0.689	0.689	eng_Latn	0.689	eng_Latn	**0.711**	cmn_Hani	0.689	eng_Latn	**0.711**	cmn_Hani
lin_Latn	0.504	0.504	eng_Latn	0.504	eng_Latn	**0.541**	arb_Arab	0.504	eng_Latn	0.450	spa_Latn
lit_Latn	0.566	0.566	eng_Latn	**0.594**	rus_Cyrl	0.566	eng_Latn	**0.594**	rus_Cyrl	**0.594**	rus_Cyrl
ltz_Latn	0.546	0.546	eng_Latn	0.546	eng_Latn	0.546	eng_Latn	**0.547**	spa_Latn	**0.547**	spa_Latn
lug_Latn	0.474	0.474	eng_Latn	0.474	eng_Latn	**0.564**	arb_Arab	0.510	spa_Latn	0.510	spa_Latn
luo_Latn	0.394	0.394	eng_Latn	0.394	eng_Latn	**0.435**	arb_Arab	0.394	eng_Latn	0.427	spa_Latn
mai_Deva	0.698	**0.724**	hin_Deva	**0.724**	hin_Deva	**0.724**	hin_Deva	**0.724**	hin_Deva	**0.724**	hin_Deva
mar_Deva	**0.720**	0.665	hin_Deva	0.665	hin_Deva	0.665	hin_Deva	0.665	hin_Deva	0.665	hin_Deva
min_Latn	0.482	0.482	eng_Latn	0.482	eng_Latn	0.464	hin_Deva	0.482	eng_Latn	**0.552**	spa_Latn
mkd_Cyrl	**0.701**	0.648	rus_Cyrl	0.648	rus_Cyrl	0.629	spa_Latn	0.648	rus_Cyrl	0.648	rus_Cyrl
mlt_Latn	0.503	0.503	eng_Latn	0.519	arb_Arab	0.527	spa_Latn	**0.556**	rus_Cyrl	0.527	spa_Latn
mos_Latn	0.360	0.360	eng_Latn	0.360	eng_Latn	**0.506**	spa_Latn	0.360	eng_Latn	**0.506**	spa_Latn
mri_Latn	**0.522**	**0.522**	eng_Latn	**0.522**	eng_Latn	0.391	cmn_Hani	**0.522**	eng_Latn	0.484	spa_Latn
mya_Mymr	0.581	0.574	arb_Arab	0.537	cmn_Hani	**0.674**	hin_Deva	0.581	eng_Latn	**0.674**	hin_Deva
nld_Latn	**0.713**	**0.713**	eng_Latn	**0.713**	eng_Latn	**0.713**	eng_Latn	**0.713**	eng_Latn	0.628	spa_Latn
nno_Latn	**0.704**	**0.704**	eng_Latn	**0.704**	eng_Latn	**0.704**	eng_Latn	0.691	hin_Deva	**0.704**	eng_Latn
nob_Latn	**0.656**	**0.656**	eng_Latn	**0.656**	eng_Latn	**0.656**	eng_Latn	**0.656**	eng_Latn	**0.656**	eng_Latn
npi_Deva	0.694	**0.712**	hin_Deva	**0.712**	hin_Deva	0.694	eng_Latn	**0.712**	hin_Deva	**0.712**	hin_Deva
nso_Latn	0.514	0.514	eng_Latn	0.514	eng_Latn	0.519	arb_Arab	0.519	arb_Arab	**0.564**	spa_Latn
nya_Latn	0.560	0.560	eng_Latn	0.560	eng_Latn	0.584	arb_Arab	0.584	arb_Arab	**0.624**	spa_Latn
ory_Orya	**0.698**	0.635	arb_Arab	0.683	hin_Deva	**0.698**	eng_Latn	0.683	hin_Deva	0.683	hin_Deva
pag_Latn	**0.618**	**0.618**	eng_Latn	**0.618**	eng_Latn	0.572	cmn_Hani	0.610	spa_Latn	0.610	spa_Latn
pan_Guru	**0.709**	0.675	hin_Deva	0.675	hin_Deva	0.675	hin_Deva	0.675	hin_Deva	0.675	hin_Deva
pap_Latn	0.572	0.572	eng_Latn	0.538	spa_Latn	0.538	spa_Latn	**0.607**	arb_Arab	0.538	spa_Latn
pes_Arab	0.624	0.619	arb_Arab	**0.668**	hin_Deva	0.619	arb_Arab	**0.668**	hin_Deva	**0.668**	hin_Deva

Table 20: Cross-Lingual Transfer Results of Taxi1500 (Part 2): The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 4).

	ENG	LEX		GEN		GEO		FEA		mPLM-Sim	
plt_Latn	0.503	0.503	eng_Latn	0.503	eng_Latn	0.495	arb_Arab	**0.627**	rus_Cyrl	0.562	spa_Latn
pol_Latn	**0.690**	**0.690**	eng_Latn	0.690	rus_Cyrl	**0.690**	eng_Latn	0.690	rus_Cyrl	0.690	rus_Cyrl
por_Latn	**0.615**	0.605	spa_Latn	0.605	spa_Latn	0.605	spa_Latn	**0.615**	eng_Latn	0.605	spa_Latn
prs_Arab	0.677	0.653	arb_Arab	0.665	hin_Deva	0.665	hin_Deva	**0.691**	cmn_Hani	0.665	hin_Deva
quy_Latn	0.696	0.696	eng_Latn	0.696	eng_Latn	0.693	spa_Latn	**0.718**	hin_Deva	0.693	spa_Latn
ron_Latn	0.582	0.582	eng_Latn	**0.617**	spa_Latn	0.582	eng_Latn	0.589	rus_Cyrl	**0.617**	spa_Latn
run_Latn	0.470	0.470	eng_Latn	0.470	eng_Latn	0.508	arb_Arab	**0.546**	hin_Deva	0.504	spa_Latn
sag_Latn	0.476	0.476	eng_Latn	0.476	eng_Latn	**0.491**	arb_Arab	0.476	eng_Latn	0.442	spa_Latn
sin_Sinh	0.582	0.652	arb_Arab	**0.663**	hin_Deva	**0.663**	hin_Deva	**0.663**	hin_Deva	**0.663**	hin_Deva
slk_Latn	0.568	0.568	eng_Latn	0.592	rus_Cyrl	0.568	eng_Latn	**0.635**	hin_Deva	0.592	rus_Cyrl
slv_Latn	0.635	0.635	eng_Latn	**0.718**	rus_Cyrl	0.635	eng_Latn	**0.718**	rus_Cyrl	**0.718**	rus_Cyrl
smo_Latn	0.600	0.600	eng_Latn	0.600	eng_Latn	**0.630**	cmn_Hani	0.549	arb_Arab	0.625	spa_Latn
sna_Latn	0.443	0.443	eng_Latn	0.443	eng_Latn	0.444	arb_Arab	**0.555**	spa_Latn	**0.555**	spa_Latn
snd_Arab	0.694	0.621	arb_Arab	**0.726**	hin_Deva	**0.726**	hin_Deva	**0.726**	hin_Deva	**0.726**	hin_Deva
som_Latn	0.355	0.355	eng_Latn	0.454	arb_Arab	0.454	arb_Arab	0.424	hin_Deva	**0.485**	spa_Latn
sot_Latn	0.441	0.441	eng_Latn	0.441	eng_Latn	**0.537**	arb_Arab	**0.537**	arb_Arab	0.516	spa_Latn
ssw_Latn	0.437	0.437	eng_Latn	0.437	eng_Latn	0.424	arb_Arab	0.424	arb_Arab	**0.497**	spa_Latn
sun_Latn	0.493	0.493	eng_Latn	0.493	eng_Latn	**0.548**	hin_Deva	0.493	eng_Latn	0.514	spa_Latn
swe_Latn	**0.665**	**0.665**	eng_Latn	**0.665**	eng_Latn	**0.665**	eng_Latn	**0.665**	eng_Latn	**0.665**	eng_Latn
swh_Latn	**0.642**	**0.642**	eng_Latn	**0.642**	eng_Latn	0.558	arb_Arab	0.574	spa_Latn	0.574	spa_Latn
tam_Taml	0.684	0.643	arb_Arab	0.684	eng_Latn	**0.695**	hin_Deva	**0.695**	hin_Deva	**0.695**	hin_Deva
tat_Cyrl	**0.670**	0.664	rus_Cyrl	**0.670**	eng_Latn	0.664	rus_Cyrl	0.648	arb_Arab	0.664	rus_Cyrl
tel_Telu	0.557	0.594	arb_Arab	0.557	eng_Latn	**0.684**	hin_Deva	**0.684**	hin_Deva	**0.684**	hin_Deva
tgk_Cyrl	0.490	**0.724**	rus_Cyrl	0.493	hin_Deva	**0.724**	rus_Cyrl	0.426	arb_Arab	**0.724**	rus_Cyrl
tgl_Latn	**0.628**	**0.628**	eng_Latn	**0.628**	eng_Latn	0.563	cmn_Hani	0.567	spa_Latn	0.567	spa_Latn
tha_Thai	0.600	**0.669**	cmn_Hani	0.600	eng_Latn	0.651	hin_Deva	0.600	eng_Latn	**0.669**	cmn_Hani
tir_Ethi	0.487	0.497	cmn_Hani	0.531	arb_Arab	0.531	arb_Arab	**0.601**	hin_Deva	**0.601**	hin_Deva
tpi_Latn	**0.621**	**0.621**	eng_Latn	**0.621**	eng_Latn	0.579	cmn_Hani	**0.621**	eng_Latn	0.609	spa_Latn
tsn_Latn	0.397	0.397	eng_Latn	0.397	eng_Latn	0.447	arb_Arab	0.413	cmn_Hani	**0.495**	spa_Latn
tuk_Latn	0.537	0.537	eng_Latn	0.537	eng_Latn	**0.649**	arb_Arab	0.592	cmn_Hani	0.604	hin_Deva
tum_Latn	0.559	0.559	eng_Latn	0.559	eng_Latn	0.528	arb_Arab	**0.642**	hin_Deva	0.533	spa_Latn
tur_Latn	0.609	0.609	eng_Latn	0.609	eng_Latn	0.602	arb_Arab	0.615	rus_Cyrl	**0.640**	hin_Deva
twi_Latn	**0.532**	**0.532**	eng_Latn	**0.532**	eng_Latn	0.507	spa_Latn	**0.532**	eng_Latn	0.507	spa_Latn
ukr_Cyrl	0.506	**0.558**	rus_Cyrl	**0.558**	rus_Cyrl	0.506	eng_Latn	**0.558**	rus_Cyrl	**0.558**	rus_Cyrl
vie_Latn	0.642	0.642	eng_Latn	0.642	eng_Latn	**0.656**	cmn_Hani	0.642	eng_Latn	0.614	rus_Cyrl
war_Latn	0.449	0.449	eng_Latn	0.449	eng_Latn	0.472	cmn_Hani	0.472	cmn_Hani	**0.505**	spa_Latn
wol_Latn	0.396	0.396	eng_Latn	0.396	eng_Latn	**0.400**	spa_Latn	0.396	eng_Latn	**0.400**	spa_Latn
xho_Latn	0.486	0.486	eng_Latn	0.486	eng_Latn	**0.507**	arb_Arab	0.486	eng_Latn	0.422	spa_Latn
yor_Latn	0.542	0.542	eng_Latn	0.542	eng_Latn	0.556	spa_Latn	**0.584**	rus_Cyrl	0.556	spa_Latn
yue_Hani	0.577	**0.718**	cmn_Hani	**0.718**	cmn_Hani	**0.718**	cmn_Hani	**0.718**	cmn_Hani	**0.718**	cmn_Hani
zsm_Latn	0.658	0.658	eng_Latn	0.658	eng_Latn	**0.694**	hin_Deva	0.658	eng_Latn	0.658	eng_Latn
zul_Latn	0.504	0.504	eng_Latn	0.504	eng_Latn	0.527	arb_Arab	0.526	rus_Cyrl	**0.529**	spa_Latn

Table 21: Cross-Lingual Transfer Results of Taxi1500 (Part 3). The first column is the target language. For each language similarity measure, we report both the source language selected based on similarity and also the evaluation results on target language using the source language. For mPLM-Sim, we report the layer achieving best performance (layer 4).