

# An Efficient Approach for Studying Cross-Lingual Transfer in Multilingual Language Models

Fahim Faisal, Antonios Anastasopoulos

Department of Computer Science, George Mason University  
{ffaisal, antonis}@gmu.edu

## Abstract

The capacity and effectiveness of pre-trained multilingual models (MLMs) for zero-shot cross-lingual transfer is well established. However, phenomena of positive or negative transfer, and the effect of language choice still need to be fully understood, especially in the complex setting of massively multilingual LMs. We propose an *efficient* method to study transfer language influence in zero-shot performance on another target language. Unlike previous work, our approach *disentangles downstream tasks from language*, using dedicated adapter units. Our findings suggest that some languages do not largely affect others, while some languages, especially ones unseen during pre-training, can be extremely beneficial or detrimental for different target languages. We find that no transfer language is beneficial for all target languages. We do, curiously, observe languages previously unseen by MLMs consistently benefit from transfer from *almost any* language. We additionally use our modular approach to quantify negative interference efficiently and categorize languages accordingly. Furthermore, we provide a list of promising transfer-target language configurations that consistently lead to target language performance improvements.<sup>1</sup>

## 1 Introduction

Pretrained Multilingual Models (MLMs) perform surprisingly well in terms of zero-shot cross-lingual transfer even though no explicit cross-lingual signal was present during pretraining. Subword fertility (Deshpande et al., 2022), token sharing (Dufter and Schütze, 2020), script (Muller et al., 2021), as well as balanced language representation (Rust et al., 2021) contribute to this effectiveness. But, by and large, the most important component seems to be the combination of languages the model is trained and evaluated on. It is important, hence, to

<sup>1</sup>Code and data are publicly available: [https://github.com/ffaisal93/neg\\_inf](https://github.com/ffaisal93/neg_inf)

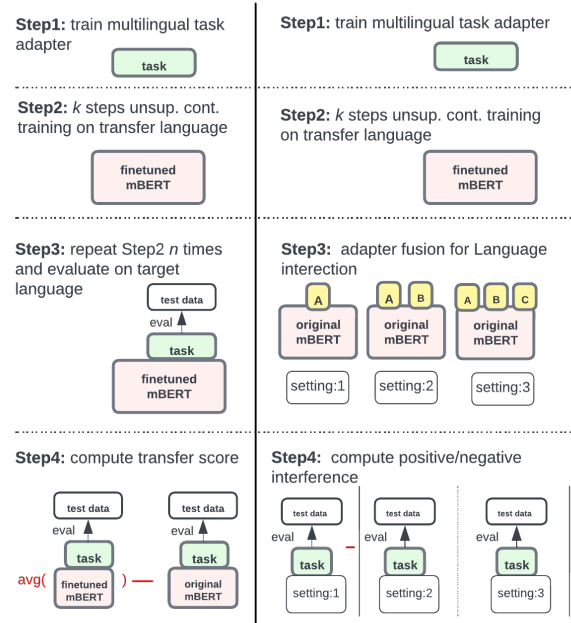


Figure 1: Our approach uses efficient few-step continued tuning (left) and adapter modules (right) to disentangle the effect of *task* and *language* to quantify the effect of a *transfer* language for a given task and model. The left panel depicts the framework for our cross-lingual transfer, while the right panel represents the scenario of multiple language interactions followed by quantifying negative interference.

understand why and when cross-lingual transfer is successful at the language level.

Previous attempts at studying cross-lingual transfer fall into two categories. First, the most popular approaches are those which, given a task and a MLM, task-tune the MLM on annotated data from a *transfer* language and then evaluate on a *target* language (e.g. Lin et al., 2019). The problem with such approaches is that (a) they do not disentangle the effect of task and language, since they train directly on the task using *annotated* data in the transfer language, and (b) it is expensive to task-tune the whole model for all possible transfer languages.

Second, other approaches tackle the inefficiency

problem by relying on bilingual approximations: Malkin et al. (2022) for instance train bi-lingual BERT (Devlin et al., 2019) models, task-tune them on the transfer language and then evaluate on the *target* one, and contrast this performance to a monolingual target-language BERT. While this approach ignores the fact that language interactions can be different in multilingual and bilingual models (Wang et al., 2020; Papadimitriou et al., 2022), it does correlate decently with transfer performance on multilingual models. However, it still does not disentangle task from language and is quite expensive, as studying  $n$  languages requires training  $n^2 + n$  BERT models.

In this work, we propose an *efficient* approach to study cross-lingual transfer, outlined in Figure 1, that also disentangles the effect of task-tuning and the effect of language, while operating within the framework of the same MLM. Our approach relies on learning a separate *task* adapter module to perform the downstream task, which needs to only be trained once (hence it is efficient). We then perform unsupervised finetuning on unannotated transfer language data for a minimal number of steps. Comparing the performance of the model on the target language with and without the previous step results in a direct assessment of the effect of the transfer language without changing the conditions under which the downstream task was learned. In addition, we extend this framework to quantify the negative interference resulted from the interaction of multiple languages (Figure 1(right)). With the aid of adapter-fusion tuning (Pfeiffer et al., 2021), we compare different combinations of language adapters and compute the interference occurring due to increased interactions.

We perform extensive analysis using this efficient approach on five downstream tasks using dozens of transfer and target languages (*184 in total*) and devise a metric (which we dub *transfer score*) to quantify which languages have/receive positive or adverse effects on/from others. Last, we focus our analysis on cross-lingual transfer for languages unseen during the pre-training of the MLM.

## 2 Methodology

Adapters (Pfeiffer et al., 2022) are light-weight parameter-efficient modules that can be injected between the layers of pretrained models. In their typical usecase, the rest of the model is frozen and only

the adapter modules are trained, to adapt a model to a new language, domain, or task. Importantly, for our goals, these adapters are also composable: one can stack an independently trained language adapter and task adapter to achieve decent performance for that language on that task. First we use an adapter-based setting to perform our analysis on cross-lingual transfer. Furthermore, we extend our study to negative interference and language interaction through another adapter-fusion-based setting.

**Cross-Lingual Transfer** The composable property of adapters allows us to disentangle learning a task from the language representations (the process is also outlined in Figure 1). In step 1, we first train a task-specific adapter [T] (e.g. named entity recognition), on data from as many languages as possible. This module will be responsible for performing the downstream task independently of input language. We then (step 2) finetune the [base] model (e.g. mBERT) on a transfer language  $\alpha$  with only a few steps (1, 10, or 100) using masked language modeling, obtaining [base $^\alpha$ ]. Now the language representations of this finetuned model will be (slightly) biased towards the transfer language.

Last, in step 3 we reinsert the task adapter in both the finetuned and the original pretrained model, and use both models to test and evaluate on target language data  $\beta$ . The difference in performance between these two models  $\text{score}(\beta; [\text{base}+\mathbf{t}]^\alpha) - \text{score}(\beta; [\text{base}+\mathbf{t}])$  will reveal whether transfer language  $\alpha$  benefits (if positive) or hurts (if negative) target language  $\beta$ .

An obvious caveat of our approach so far is that a single update (or 10 or 100) with a randomly sampled batch in any language does not allow for any robust conclusions. To avoid this issue, we repeat the above process  $n=10$  times for each transfer language with different data and aggregate these scores.

Our final transfer score  $\text{ts}(\alpha \rightarrow \beta; \text{base}, \mathbf{t})$  for a given model base and task  $\mathbf{t}$  turns the difference of the finetuned and original model into a percentage of the original baseline performance, for fairer comparisons at different levels of performance:

$$\text{ts}(\alpha \rightarrow \beta; \text{base}, \mathbf{t}) = \frac{\sum_1^{10} \text{score}(\beta; [\text{base}+\mathbf{t}]^\alpha) - \text{score}(\beta; [\text{base}+\mathbf{t}])}{n \cdot \text{score}(\beta; [\text{base}+\mathbf{t}])}$$

**Negative Interference** The typical definition of negative interference describes it as the phe-

nomenon when batches in different languages produce opposite gradients during training. We instead focus on downstream performance, in line with most studies focusing on cross-lingual transfer, assuming that a negative effect on performance implies negative interference. Another reason is that, in  $n$  dimensional spaces, there extremely high probability of two random vectors being orthogonal; hence any two gradient vectors could certainly be orthogonal without necessarily impacting downstream performance.

To quantify negative interference, we follow a modular-based approach depicted in Figure 1(right). Like before, we separate the task and language, followed by performing interaction among multiple languages. However, we use language adapters at this time instead of continuously finetuning the base model. This strategy allows us to efficiently train multiple language sub-parts only once (Step2) followed by mixing those modules through adapter fusion (Pfeiffer et al., 2021). In our experiments, we train a set of language adapters and make either monolingual settings or a combination of bilingual/trilingual interactions (Step3). Then we stack previously trained task adapter while only changing the underlying language combination. Finally, we extract the interference score from the difference between already computed multilingual and monolingual counterparts (Step4).

Having these interference scores at hand, we can tell whether a language actually gets benefits or not while influencing the associated languages in a positive/negative manner. For example, consider language A interacting with language B. We can easily quantify the interference of language A by calculating the loss/gain of this bilingual interaction [AB]: a score increase for A compared to its monolingual counterpart (i.e.  $+A = +_{[AB]-[A]}$ ) means positive interference for A in this particular setting. We can further extend this to a trilingual setting as well (i.e.  $+A = +_{[ABC]-[A]}$ ). Using these scores, we can get different combinations of interference scenarios by counting the co-occurred positive/negative interference. We use  $|+A, +B|$  to denote the number of cases where A benefits both itself and B, presenting all possible rules in Table 1. Utilizing these rules, we can identify how much language A actually gains or loses during its bilingual/trilingual interactions while providing substantial interference to other languages.

Moreover, we can use these interference combination counts to project languages in an interfer-

| Notations (+: win, -: loss) |            |   |
|-----------------------------|------------|---|
| 1.                          | $ +A $     | $= \text{count}(A \text{ gains in interaction } [AB] \text{ or } [ABC])$  |
| 2.                          | $ -A $     | $= \text{count}(A \text{ losses in interaction } [AB] \text{ or } [ABC])$   |
| 3.                          | $ +A, +B $ | $= \text{count}(\text{Both language gets benefit}).$ In other words, A gains. At the same time, B receives benefits while interacting with A. |

| Bilingual Interactions | Trilingual Interactions |                |
|------------------------|-------------------------|----------------|
| $ -A, -B $             | $ -A, -B, -C $          | $ -A, -B, +C $ |
| $ -A, +B $             | $ -A, +B, -C $          | $ -A, +B, +C $ |
| $ +A, -B $             | $ +A, -B, -C $          | $ +A, -B, +C $ |
| $ +A, +B $             | $ +A, +B, -C $          | $ +A, +B, +C $ |

Table 1: Interference calculation for language A.  $|+A|$  means the number of cases where A itself gets benefits. If the setting is bilingual, then  $|+A| = \text{count}(+_{[AB]-[A]})$  (i.e. if the evaluation score on task language A:  $[AB] - [A] > 0$  for the combination [AB], we get a  $+A$ .)

ence representation space. For example, consider a 2-D space of bilingual interaction where the X-axis represents the negative/positive interference a language receives from one such interaction and the Y-axis is for the interference it provides to other languages. We can project a language using the dot product of counts (eg.  $|+A, -B|$ ) with its corresponding quadrant identifier  $[1, -1]$ . As a result, the projection coordinates  $(x_A, y_A)$  for language A in a bilingual interaction could be obtained as follows:

$$\begin{aligned}
 C &= |-A, -B| + |-A, +B| + |+A, -B| \\
 &\quad + |+A, +B| \\
 (x_A, y_A) &= \frac{1}{C} \times (|-A, -B| \cdot [-1, -1] \\
 &\quad + |-A, +B| \cdot [-1, 1] + |+A, -B| \cdot [1, -1] \\
 &\quad + |+A, +B| \cdot [1, 1])
 \end{aligned}$$

Using the above-mentioned projections, we visualize a language in a way that represents how much interference it provides as well as receives (see example with each step of the calculation in Appendix §F). We can further extend this strategy to the trilingual setting, but now we have to deal with eight axes instead of four. In Figure 4 of the result section, we present the language interaction visualizations for bilingual and trilingual scenarios.

### 3 Experimental Setup

We conduct our experiments in two different settings targeted to perform two different analyses: first understanding the language effect on cross-lingual transfer and then, extending this to quantify language-language interaction.

Primarily, we use multilingual BERT as our base model and report XLM-R results for comparative model evaluation. We use a total of 38 transfer languages (11 unseen during pretraining) to finetune the MLM using masked language modeling with the process described above. Using these transfer languages, we do monolingual finetuning on mBERT for either 1, 10, 100, or 1000 steps and each experiment is repeated for 10 times. At the sametime, we trained multilingual task adapters followed by task evaluation on the following tasks:

- **Token-level:** Dependency Parsing (DEP), Part-of-Speech (POS) tagging and Named Entity Recognition (NER). Parsing and POS tagging are evaluated on a set of 114 languages from Universal Dependencies v2.11 (de Marneffe et al., 2021). For NER, we use 125 languages from the Wikiann (Pan et al., 2017) dataset.
- **Sentence-level:** Natural Language Inference (NLI) evaluated on XNLI (Conneau et al., 2018) and AmericasNLI (ANLI) (Ebrahimi et al., 2022) datasets.
- **Extractive Question Answering:** Evaluated on TyDiQA (Clark et al., 2020) gold task.

Additionally, we train 38 language adapters to perform the experiment on language-to-language interaction and negative interference. Here, we stack the previously trained task adapter on top of either one or a combination of double or triple language adapters (Figure 1(b)) and then perform the evaluation on the transfer languages having task data available. All training and evaluation datasets, implementation and hyper-parameter details are provided in Appendices C-E (Table 24-29).

## 4 Results and Discussion

First, in 4.1, we present a comparative scenario in between continuous training and language interaction in terms of performance improvement over the baseline model. Then in 4.2, we discuss the findings of continuous training in the context of cross-lingual transfer. After that, in 4.3, we present the representation of language interactions as well as interference following the strategy discussed in Section 2.

### 4.1 Continuous Training vs Language Interaction

Here we present 8 sets of scores for each token-level task. The baseline is where we stack the task adapter on the base pretrained mBERT (i.e. zero-shot

| Lang.              | Base        | Continuous Steps |             | Lang. Interaction |       |      |
|--------------------|-------------|------------------|-------------|-------------------|-------|------|
|                    |             | k=10             | k=1000      | [1A]              | [2A]  | [3A] |
| <b>Parsing</b>     |             |                  |             |                   |       |      |
| pcm                | <b>81.1</b> | 79.1             | 77.9        | 79.3              | 79.5  | 79.5 |
| wol                | <b>69.5</b> | 68.1             | 67.3        | 68.9              | 69.1  | 69.1 |
| kmr                | 31.9        | 31.7             | <b>45.3</b> | 32.6              | 32.1  | 32.0 |
| bam                | 29.9        | 30.9             | <b>38.1</b> | 30.8              | 30.8  | 30.8 |
| gub                | 21.7        | 20.9             | <b>34.5</b> | 23.8              | 23.73 | 23.5 |
| <b>POS Tagging</b> |             |                  |             |                   |       |      |
| pcm                | <b>92.9</b> | 92.2             | 91.2        | 92.3              | 92.5  | 92.6 |
| wol                | <b>85.6</b> | 84.2             | 82.1        | 84.1              | 84.7  | 84.8 |
| kmr                | 40.2        | 40.5             | <b>55.8</b> | 41.1              | 40.8  | 40.7 |
| bam                | 30.3        | 30.8             | <b>49.5</b> | 30.7              | 30.5  | 30.5 |
| gub                | 28.5        | 28.7             | <b>36.7</b> | 28.8              | 28.8  | 28.9 |
| <b>NER</b>         |             |                  |             |                   |       |      |
| ibo                | <b>61.1</b> | 57.2             | 55.4        | 57.5              | 57.8  | 57.7 |
| pms                | 88.2        | <b>88.9</b>      | 87.6        | 88.2              | 87.5  | 87.6 |
| kin                | <b>72.4</b> | 71.8             | 68.5        | 70.5              | 71.1  | 71.9 |

Table 2: Task results for transfer languages unseen by mBERT. **base:** zero-shot with task adapter [T]. **Continuous Steps:** do  $k$  steps of finetuning on that language plus [T]. **Lang. Interaction:** introducing language adapters; [1A]: just 1 adapter (in language) and evaluate on it; [2A]: 2 language adapters, the target lang. and one test (the result is averaged for all transfer langs.); [3A]: 3 lang. adapters (results are average again). The highest obtained score for each language is bolded.

task on pretrained mBERT+ [T]). Then for all the evaluation languages, we perform 4 sets of cross-lingual transfers (i.e. 1, 10, 100, and 1000 steps of continuous training). For the language-language interaction experiment, we only perform the evaluation on transfer languages where either 1, 2 or 3 language adapters are fused together before stacking the task adapter (i.e. [1A], [2A], [3A]).

**Only Unseen Transfers** In Table 2, we present our token-level evaluation report for transfer languages unseen during the pretraining phase. For the [2A] and [3A] language interaction results, we compute and report the average score where the evaluation language is also present in the [2A] or [3A] adapter fusion. For tasks where word-to-word relation plays a critical role (parsing and pos tagging), we observe similar patterns of improvement over baseline in both Cont. steps and lang. interaction settings. Whereas, for a task like NER, we do not observe any improvement over baseline both in sustained cont. (k=1000) and interaction settings. Even though we are evaluating the same language after continuous masked language modeling (mlm) or adapter fusion with another high-resource language, there is no clear winning formula that can

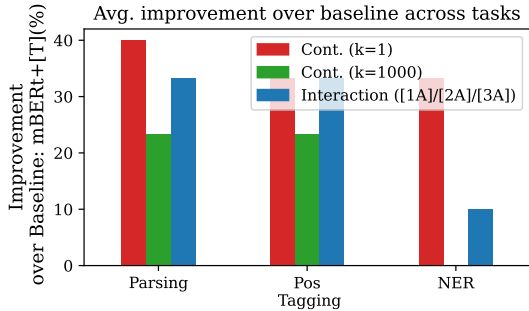


Figure 2: Average score improvement over baseline across tasks for the transfer languages (evaluated on itself). We observe a spike of over 33% positive score at continuous training step 1. Among these, only 23.3% cases result in sustained improvement after 1000 steps (0% in NER). On the contrary, standard language adapter interaction stays at 25% average improvement.

always serve the unseen low-resource languages.

**Unseen+Seen Transfers** On the other hand, when we consider the case of both unseen and seen languages together in token-level tasks, we see a spike of 33% average improvement over baseline with just 1 step of mlm training. However, this improvement percentage gets down to a sustained 23.3% (except task NER) when we evaluate again after having 1000 steps of training. Whereas, in language interaction settings where we fuse standard well-trained language adapters, we generally observe improvement for those languages which also get benefited from continuous training. The improvement percentage averaged over all 38 transfer languages is presented in Figure 2. In addition, we present all the scores for all 38 transfer languages and token-level tasks in App. Tables 7, 9, and 11.

## 4.2 Takeaways from Continuous Training

**No Universal Donor** First, we search for transfer languages that can be used for positive transfer for a large set of languages. However, we find no language out of 38 that can positively influence almost all languages using mBERT as base model. For this experiment, we rank the transfer languages based on their averaged transfer score (i.e. aggregated-transfer). In Table 3, we list the top 5 ranked transfer languages with their transfer score (base model: mBERT) and the percentage of target languages that do benefit from them (more details in Appendix H). We observe, most languages benefit within the range of 30-45% of target languages across tasks except NLI. However, we did not receive any positive transfer for

|   | Lang.          | ts    | +(%) | Lang.              | ts    | +(%) | Lang.         | ts   | +(%) |
|---|----------------|-------|------|--------------------|-------|------|---------------|------|------|
|   | <b>Parsing</b> |       |      | <b>POS Tagging</b> |       |      | <b>NER</b>    |      |      |
| 1 | mya            | 0.33  | 40.4 | <b>kin</b>         | 0.41  | 35.1 | zho           | 0.16 | 49.6 |
| 2 | ell            | 0.15  | 31.6 | <b>kmr</b>         | 0.36  | 36.9 | tel           | 0.08 | 32.8 |
| 3 | <b>kmr</b>     | 0.14  | 35.9 | <b>mos</b>         | 0.27  | 34.2 | hun           | 0.08 | 40.8 |
| 4 | yor            | 0.14  | 33.3 | hye                | 0.27  | 36.9 | heb           | 0.04 | 34.4 |
| 5 | <b>pcm</b>     | 0.13  | 31.6 | cym                | 0.22  | 37.7 | est           | 0.03 | 36.8 |
|   | <b>XNLI</b>    |       |      | <b>ANLI</b>        |       |      | <b>TyDiQA</b> |      |      |
| 1 | <b>hau</b>     | -34.4 | 0.0  | <b>bam</b>         | -15.0 | 0.0  | zho           | 0.7  | 77.8 |
| 2 | <b>bam</b>     | -34.9 | 0.0  | <b>hau</b>         | -17.8 | 0.0  | jpn           | 0.1  | 44.4 |
| 3 | <b>gub</b>     | -36.4 | 0.0  | <b>gub</b>         | -18.4 | 0.0  | gle           | -0.1 | 44.4 |
| 4 | <b>ewe</b>     | -36.7 | 0.0  | deu                | -19.8 | 0.0  | <b>wol</b>    | -0.1 | 44.4 |
| 5 | hin            | -37.1 | 0.0  | fin                | -19.9 | 0.0  | cym           | -0.1 | 33.3 |

Table 3: Top 5 transfer languages per task ranked using the aggregated transfer score (ts columns; see App. H for computation). Unseen ones are **bolded**. +(%) is the percentage of languages receiving positive transfer. No transfer language helps all target languages. (Complete rank with transfer scores: Table 15-18).

|       | Parsing | Pos Tagging | NER  | XNLI | ANLI | TyDiQA |
|-------|---------|-------------|------|------|------|--------|
| mBERT | 30.6    | 31.0        | 31.8 | 0    | 0    | 30.1   |
| xlmr  | 20.5    | 33.2        | 41.1 | 44.4 | 41.6 | 17.0   |

Table 4: Average percentage of languages receiving positive transfer (avg. +(%)) across models. Unlike mBERT, xlmr provides positive transferring in NLI.

both of the two different NLI task datasets (XNLI and ANLI). The maximum positive transfer percentage is from zho in both NER and TyDiQA. Interestingly, low-resourced unseen languages perform well in general as transfer languages: 31.7% (token-level) and 28.3% (sentence-level) of top 20 transfer languages are unseen languages.

**Base Model and Task Matters** To further investigate the discrepancy observed in NLI task, we replace the base model mBERT with XLM-R (Table 4). Unlike mBERT, XLM-R in NLI provides superior performance (XNLI: +44.4% and ANLI: +41.6%). This signifies how the choice of the base model in a setting with a disentangled language-task effect could drastically change the cross-lingual transfer performance of certain tasks.

Moreover, we observe the above-discussed rankings of transfer languages vary across tasks. To investigate the underlying similarity, we select a large subset of languages (the common 62 target languages across three token-level tasks) and rank the transfer languages as before. We then compute the Spearman rank correlation and statistical significance ( $p < 0.05$ ) of their transfer scores tasks (see Appendix Table 21). Only parsing and NER are positively correlated ( $\rho = 0.4$ ) whereas POS tag-

| Rank | Lang.<br># (max, min) | ts    | Var. | Type             |
|------|-----------------------|-------|------|------------------|
| 1    | <b>ibo</b> (10, 10)   | 0.05  | 23.5 | (+ <i>and</i> -) |
| 3    | <b>bam</b> (11, 15)   | 0.02  | 21.5 | (+ <i>and</i> -) |
| 6    | <b>mos</b> (13, 2)    | 0.09  | 16.1 | (+)              |
| 8    | <b>pcm</b> (1, 11)    | 0.13  | 13.4 | (-)              |
| 26   | eng (0, 0)            | -0.22 | 6.4  | neutral          |
| 36   | ara (0, 0)            | -0.12 | 5.1  | neutral          |

Table 5: Example of transfer languages ranked with their aggregated-transfer (ts) score variance (task: parsing). Unseen languages (**bold** font) exhibit high variance. # (max) represents the language count receiving maximum positive transfer. (see Appendix L)

ging is negatively correlated with the other two tasks. This is somewhat surprising, because we use the same underlying dataset for the parsing and POS tagging tasks. We find only a few transfer languages could effectively provide positive transfer simultaneously across tasks. The 5 common languages in the top 20 across tasks are: yor, mos, kin, hau, and tel. In sort, languages unseen by mBERT (in **boldface**), exhibit similar ranking across tasks (see Table 15-18), whereas others vary. For example, zho is the lowest-ranked one in parsing while being top-ranked in NER! Appendix Figure 6 shows the number of common languages across tasks.

### Unseen Languages Transfer with High Variance

We observe that transfer languages with high variance mainly fall into one of three categories:

1. (+ *and* -): boost performance for some languages while hurt significantly some others;
2. (+): mostly (small) positive transfer, significantly hurts only a few languages;
3. (-): mostly (small) negative transfer, significantly helps only a few languages.

See examples in Table 5 and Appendix L for details. Though unseen languages perform well as transfer languages, they usually exhibit the traits of high-variance transfer. Around 90% of unseen transfer languages are within top-20 languages sorted by variance (see Appendix Figure 7).

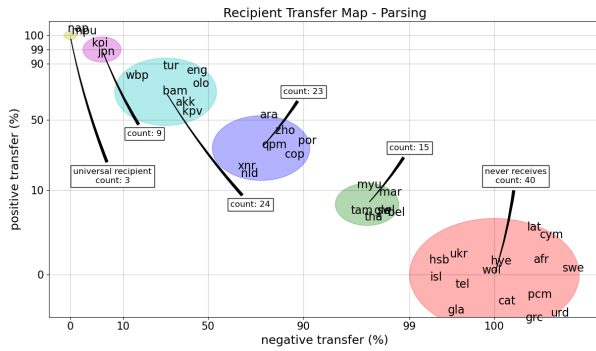
**Target Language Differences** Unlike transfer languages, we find target languages that are almost universal recipients of positive cross-lingual transfer, many of which are unseen by mBERT. On the other hand, some languages do not receive any benefit from the diverse set of transfer languages. In Figure 3(a), we plot the target languages based on the percentage of languages from which they receive positive or negative transfer (see additional

maps in Appendix Figure 5). We find around one-third of target languages across three token-level tasks never receive any positive transfer (parsing: 35.1%, POS: 28.1%, NER: 32.8%). Nevertheless, there are target languages (mostly unseen by mBERT) that benefit from all transfer languages (eg. nap, mpu in parsing). See Appendix I and Table 19 for additional results.

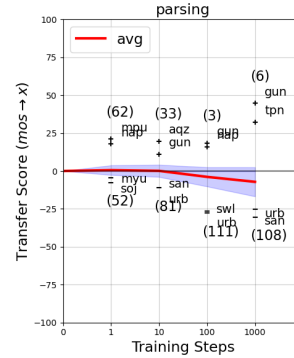
**Seen vs Unseen Languages** Transferring from either seen or unseen languages to unseen languages (i.e. transfer(seen/unseen→ unseen)) generally helps. For this experiment, we use the large set of token-level task evaluation and 11 transfer languages unseen during mBERT pertaining from diverse families including Indo-European, Afro-Asiatic, Mande, Niger-Congo and Tupian. We observe, that transferring to a large and diverse set of seen languages from unseen languages (i.e. transfer(unseen→ seen)) does not provide any substantial utility. Among the three tasks, we get the average transfer as positive for unseen transfer languages just once (dependency parsing, transfer(unseen→ unseen)). See Figure 8 for the difference of utility provided when the transfer-/target languages are seen vs unseen.

**Sustained Cross-Lingual Transfer** Our approach limits step 2 (continued training on the transfer language) to a minimal number of steps. For this section, we extend this to 1000 steps. In the vast majority of transfer-target language combinations, this leads to (small) negative transfer under our setting. We suspect this is due to the underlying model undergoing the first steps of catastrophic forgetting (McCloskey and Cohen, 1989).

There are some languages, though, mostly unseen ones (eg. nap, gun, tpn, aqz) that benefit more from this extended setting. See Appendix J Table 20, where we report the target language receiving the highest benefit from each transfer language for each setting (1,10,100,1000 steps). All the max-utility recipients aside from bar and nds are unseen languages. Figure 3(b) presents the training step progression of aggregated-transfer scores for Mossi, one of the most donating transfer languages, and Appendix N (Figures 9-18) shows the transfer progression graphs for all transfer languages. At the task level, POS tagging always ends up having comparatively higher target language performance variance with more training steps, while NER almost always ends up with negative results with longer training.



(a) Target languages mapped based on percentage of receiving positive/negative transfers.



(b) Aggregated-transfer score line with standard deviations through different training steps for Mossi (mos) as transfer language.

Figure 3: (a) Some languages exhibit universal recipient nature (yellow) while some never receive positive transfer (red). (b) Shown are the top and bottom two languages receiving maximum/minimum scores (eg. gun, tpn at 1000 steps) at each step, with total positive/negative transfers (in parenthesis) also shown. See Appendix N for other transfer language score lines.

### 4.3 Takeaway from Language Interactions

We plot all the transfer languages in a 2d axis for both two-language interactions and three-language interactions as shown in Figure 4.

**Bilingual Interactions** First of all, we observe most of the languages mainly fall into either one of the two categories: (1) A(+), B(+): getting benefits from interactions and helping others at the same time, (2) A(-), B(+): Helping other languages but do not get benefits from those languages. Secondly, there are resemblances in how certain languages from specific categories interfere across all 3 tasks. For example, consider the case of zho, swe, spa and fra. These languages fall to the lower right part of all three graphs. However, there are languages like ara that do not uniformly get benefits across three tasks while maintaining its positive interfering status. Although, there are debates whether English (eng) is an appropriate "hub" language or not (Anastasopoulos and Neubig, 2020), eng maintains its status in the upper right quarter making it a good transfer language in all Latin script majority settings.

**Trilingual Interactions** Now we increase the number of languages for a specific transfer language to influence. When we compare the bilingual settings with the trilingual ones (Figure 4 (2)), the left-right categorization remains the same. However, many languages receive an uplifting position meaning the strength of performing positive interference increases for those languages (eg. are in dependency parsing, zho in NER). Moreover, we observe an overall decrease in the lower-right cor-

ner for both dependency parsing and NER. However, there are languages like wol in POS tagging that goes from upper-left to lower-left. Nonetheless, very few different colored points (i.e. negative coordinate for 3rd language) signify the fact that a multilingual setting is beneficial towards a larger group of recipients.

## 5 Recommendations

Based on our above findings, we make a number of recommendations in choosing the appropriate transfer language and training scheme for a low-resource setting.

1. There is no universal donor but having multiple transfer languages in the training scheme helps in terms of language interference.
2. For universal recipient languages (eg. Typologically diverse unseen ones), including almost any language in the transfer scheme help.
3. Low resource unseen languages generally transfer with high variance. A good idea is to include them with other seen languages in the transfer scheme to stabilize the transfer output across a large number of target languages.
4. Only some of the unseen low-resource ones show sustained transfer toward other low-resource languages through continuous thousand-step training. Usually, the deviation happens during an early stage of training. So just continuing pretraining for longer is not optimal for a scenario with mixed-category languages.
5. The patterns of receiving positive transfer are similar when we use either one language small-

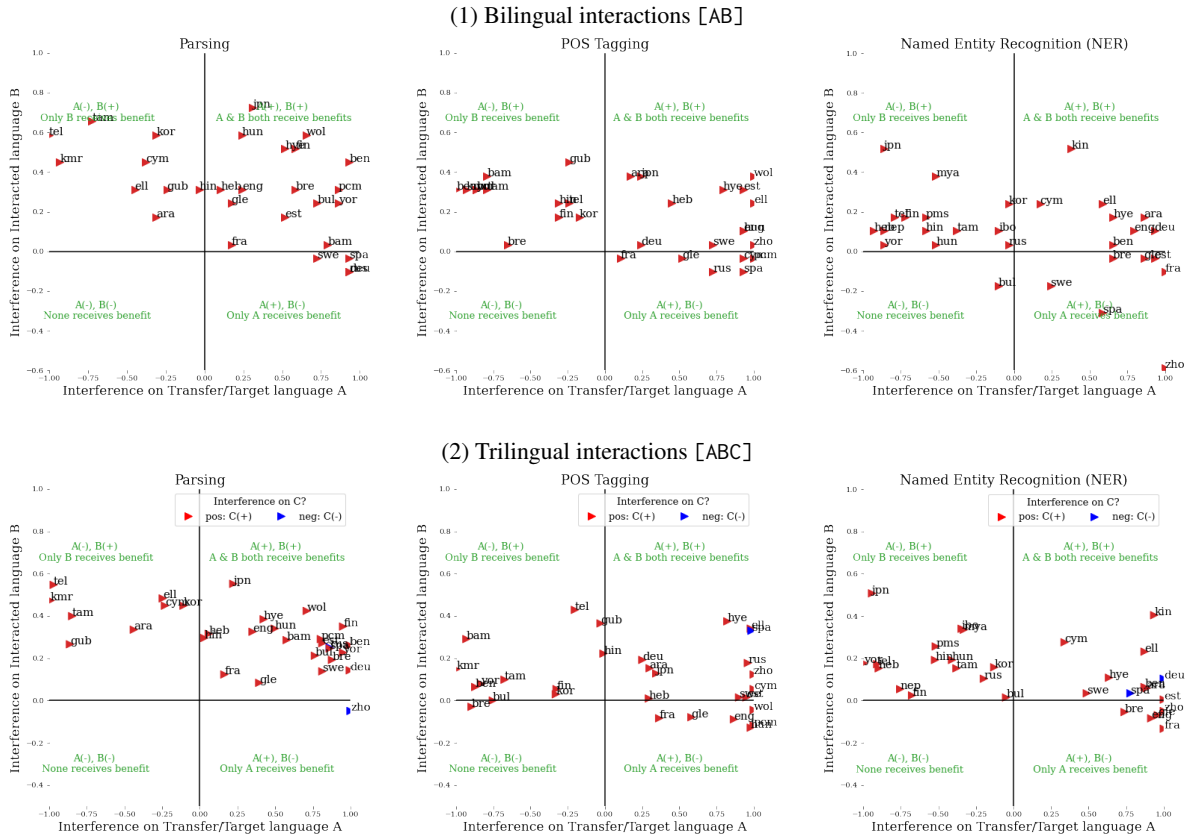


Figure 4: Language interaction representation for bilingual and trilingual settings. To identify the language coordinates, we use two and three adapters (i.e. [2A], [3A]) jointly fused. In [3A] plots, we show the position only for one interacted language B along with transfer/target language A. For the 3rd language C, we use color variation (red/blue) to depict whether C receives positive transfer or not.

step continuous training or 2/3 standard adapter fusion. So using a large set of trained language adapters fused together according to the need is a simpler way to deal with a large set of mixed-category target languages.

## 6 Conclusion and Future Work

We devise an efficient approach to study cross-lingual transfer in multilingual models for various tasks that disentangles task and language effects. We believe this disentanglement coupled with few-step fine-tuning has the potential to uncover currently uncharted model behaviors (eg. NLI evaluation). Our findings suggest languages unseen by MLMs clearly exhibit different behavioral pattern compared to other languages in general: they are universal as target, exhibit high variance as transfer language, and their behavior follows similar patterns across tasks. In addition, we do not find a universal donor (a language that benefits all others). Last, we find that some languages consistently benefit from settings that resemble "catastrophic

forgetting" for other languages, an observation we believe merits a dedicated follow-up study.

We hope that our approach will allow for further study of cross-lingual transfer for more languages and MLMs, and we plan to extend this in future work, as our findings suggest interesting differences in the behavior of languages used in pre-training and unused ones. Eventually, we hope that our study will also lead to guidelines for selecting appropriate transfer languages, as well as more informed methods for the adaptation of MLMs to new under-served languages. While our proposed approach being highly efficient to expand the paradigm of cross-lingual transfer evaluation, the findings shed light onto the easy adaptation of MLMs for new languages in a low-resource setting.

## Limitations

In this work, we primarily experiment with encoder models like mBERT and XLM-R, token-level syntactic tasks and two sentence-level tasks. In future, we would expand this work to recent large language



models and tasks involving natural language understanding. Moreover, our work only focus on low-resource setting with small-scale training data and parameter-efficient adapters. In future, instead of monolingual finetuning we will use this parameter efficient approach for multilingual finetuning thus unfolding effective multilingual pretraining configurations. As the base-language model choice, we only use mBERT. The evaluation of cross-lingual transfer needed to be expand to decoder based language models.

## Acknowledgements

This work has been generously supported by the National Science Foundation under grants IIS-2125466 and IIS-2327143. We are thankful to the anonymous reviewers and area chairs for their constructive feedback. This project was supported by resources provided by the Office of Research Computing at George Mason University (<https://orc.gmu.edu>) and funded in part by grants from the National Science Foundation (Awards Number 1625039 and 2018631).

## References

- Julien Abadji, Pedro Ortiz Suarez, Laurent Romary, and Benoît Sagot. 2022. [Towards a Cleaner Document-Oriented Multilingual Crawled Corpus](#). *arXiv e-prints*, page arXiv:2201.06642.
- David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruitter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajudeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. [A few thousand translations go a long way! leveraging pre-trained models for African news translation](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.
- Antonios Anastasopoulos and Graham Neubig. 2020. [Should all cross-lingual embeddings speak English?](#) In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8658–8679, Online. Association for Computational Linguistics.
- Jonathan H. Clark, Eunsol Choi, Michael Collins, Dan Garrette, Tom Kwiatkowski, Vitaly Nikolaev, and Jennimaria Palomaki. 2020. [TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages](#). *Transactions of the Association for Computational Linguistics*, 8:454–470.
- Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. [XNLI: Evaluating cross-lingual sentence representations](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2475–2485, Brussels, Belgium. Association for Computational Linguistics.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. [Universal Dependencies](#). *Computational Linguistics*, 47(2):255–308.
- Ameet Deshpande, Partha Talukdar, and Karthik Narasimhan. 2022. [When is BERT multilingual? isolating crucial ingredients for cross-lingual transfer](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3610–3623, Seattle, United States. Association for Computational Linguistics.
- 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Philipp Dufter and Hinrich Schütze. 2020. [Identifying elements essential for BERT’s multilinguality](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4423–4437, Online. Association for Computational Linguistics.
- Abteen Ebrahimi, Manuel Mager, Arturo Oncevay, Vishrav Chaudhary, Luis Chiruzzo, Angela Fan, John Ortega, Ricardo Ramos, Annette Rios, Ivan Vladimir Meza Ruiz, Gustavo Giménez-Lugo, Elisabeth Mager, Graham Neubig, Alexis Palmer, Rolando Coto-Solano, Thang Vu, and Katharina Kann. 2022. [AmericasNLI: Evaluating zero-shot natural language understanding of pretrained multilingual models in truly low-resource languages](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6279–6299, Dublin, Ireland. Association for Computational Linguistics.

- Fahim Faisal and Antonios Anastasopoulos. 2022. [Phylogeny-inspired adaptation of multilingual models to new languages](#). In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 434–452, Online only. Association for Computational Linguistics.
- Yoshinari Fujinuma, Jordan Boyd-Graber, and Katharina Kann. 2022. [Match the script, adapt if multilingual: Analyzing the effect of multilingual pretraining on cross-lingual transferability](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1500–1512, Dublin, Ireland. 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 Annual Meeting of the Association for Computational Linguistics*, pages 3125–3135, Florence, Italy. Association for Computational Linguistics.
- Dan Malkin, Tomasz Limisiewicz, and Gabriel Stanovsky. 2022. [A balanced data approach for evaluating cross-lingual transfer: Mapping the linguistic blood bank](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4903–4915, Seattle, United States. Association for Computational Linguistics.
- Michael McCloskey and Neal J Cohen. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier.
- Benjamin Muller, Antonios Anastasopoulos, Benoît Sagot, and Djamé Seddah. 2021. [When being unseen from mBERT is just the beginning: Handling new languages with multilingual language models](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 448–462, Online. Association for Computational Linguistics.
- 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 (Volume 1: Long Papers)*, pages 1946–1958, Vancouver, Canada. Association for Computational Linguistics.
- Isabel Papadimitriou, Kezia Lopez, and Dan Jurafsky. 2022. Multilingual bert has an accent: Evaluating english influences on fluency in multilingual models. arXiv:2210.05619.
- Jonas Pfeiffer, Naman Goyal, Xi Lin, Xian Li, James Cross, Sebastian Riedel, and Mikel Artetxe. 2022. [Lifting the curse of multilinguality by pre-training modular transformers](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3479–3495, Seattle, United States. Association for Computational Linguistics.
- Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. 2021. [AdapterFusion: Non-destructive task composition for transfer learning](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 487–503, Online. Association for Computational Linguistics.
- Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. 2020. [AdapterHub: A framework for adapting transformers](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 46–54, Online. Association for Computational Linguistics.
- Phillip Rust, Jonas Pfeiffer, Ivan Vulić, Sebastian Ruder, and Iryna Gurevych. 2021. [How good is your tokenizer? on the monolingual performance of multilingual language models](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3118–3135, Online. Association for Computational Linguistics.
- Zirui Wang, Zachary C. Lipton, and Yulia Tsvetkov. 2020. [On negative interference in multilingual models: Findings and a meta-learning treatment](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4438–4450, Online. Association for Computational Linguistics.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

## A Related Works

**Cross-Lingual Transfer** Studying cross-lingual transfer to prepare a better pretraining configuration is a well-explored topic. Malkin et al. (2022) propose a balanced-data approach to identify effective set of languages for model training through constructing bilingual language graph. They formulate the problem in terms of linguistic blood bank where language can either play the role of donor or receiver. This study comprises over a large set of languages while training a large number of bilingual models. However, how a large multilingual model (eg. mBERT) having a shared representation space larger than bilingual models perform in similar setting is not evaluated yet. Fujinuma et al. (2022) points out it is always better to have a diverse set of languages during pretraining for zero-shot adaptation. At the same-time, language relatedness in pretraining configuration always helps.

**Adaptation to Unseen Languages** The idea of performing effective zero-shot transfer is highly beneficial for model adaptation to new languages. According to Muller et al. (2021), transfer learning helps some new languages while some hard languages does not get the benefit mainly because of the difference in writing systems. Transliterating those languages to a more familiar form is a useful approach in this case.

**Parameter Efficiency** Recently parameter-efficient language modeling approaches are becoming more and more popular and capable. Adapter units (Pfeiffer et al., 2022) are such modular units containing small trainable set of parameters. Using adapters resolve the problem of model-capacity and training bottleneck. In addition, most of the parameters remain unchanged thus preventing the problem of negative interference. The most important benefit of adapter units are it’s modular design. It is also possible to train the adapters using language-phylogeny information (Faisal and Anastasopoulos, 2022) thus extending the base model capacity to unseen new language in an informed manner.

## B Terminologies

**Transfer Language:** The languages we use to perform monolingual finetuning of the base language model (mBERT) using masked language modeling.

**Target Language:** The languages we use to evaluate both the pretrained as well as finetuned mBERT on downstream tasks.

**Negative Transfer:** The scenario where language model performance drops because of finetuning it on a transfer language.

**Cross-lingual Transfer:** The established method of finetuning a language model on one transfer language and deploy it on another target language.

**Unseen Languages** Any language that were not part of the original pretraining step.

## C Dataset Details

### C.1 Transfer Languages

We perform mono-lingual finetuning as well as language adapter training on 38 transfer languages. Each language dataset contains 10k lines of text. We use texts from several corpus including OSCAR (Abadji et al., 2022) and African News Translation dataset (Adelani et al., 2022). 11 out of these 38 languages are unseen by mBERT during pretraining steps. The list is provided in Table 24.

### C.2 Adapter Training Dataset

**Dependency Parsing** We train a task adapter for performing dependency parsing task. For this step, we use Universal Dependency training dataset v2.11 (de Marneffe et al., 2021). To keep the data distribution balanced, we use not more than a thousand examples per language. Combining all these data together, we train a multilingual dependency tagging task adapter. The complete list of data-source languages for training this adapter is presented in Table 25.

**Parts-of-Speech Tagging** Here we also use the Universal Dependency training dataset v2.11 (de Marneffe et al., 2021). The languages are also the same ones used for dependency parsing previously.

**Named Entity Recognition** We use Wikiann (Pan et al., 2017) dataset for training a NER task adapter. The complete language lists are provided in Table 26.

**Natural Language Inference** We use XNLI (Conneau et al., 2018) dataset for training a NLI task adapter. The complete language lists are provided in Table 27.

**Extractive Question Answering** We use TyDiQA (Clark et al., 2020) dataset for training an Extractive Question Answering task adapter. The complete language lists are provided in Table 28.

### C.3 Evaluation Dataset

We use 125 languages for evaluating NER task from Wikiann. For udp and pos-tagging tasks we use 114 languages from Universal Dependency dataset. There are 62 languages which are common between these two sets of 125 and 114 languages. For NLI evaluation, we use 15 languages from XNLI (Conneau et al., 2018) dataset and 10 low-resource South American indigenous languages from Americas NLI (ANLI) (Ebrahimi et al., 2022) dataset. For the question answering task, we take 9 languages from TydiQA (Clark et al., 2020) to evaluate. The complete list of 184 evaluation languages are provided in Table 29.

## D Implementation Details

For all of our experiments, we use as well as modify the scripts from huggingface (Wolf et al., 2020) and adapterhub (Pfeiffer et al., 2020). For base language model, we use the model bert-base-multilingua-uncased from huggingface model repository.

## E Hyper-parameters

### Masked Language Modeling finetuning

- Train batch size: 8
- Evaluation batch size: 8
- Training Steps: 1, 10, 100 and 1000
- Learning Rate: 5e-5
- Maximum Sequence Length: 512

### Language Adapter Training: Language Interaction

- Train batch size: 8
- Evaluation batch size: 8
- Training Epochs: 3
- Learning Rate: 5e-4
- Maximum Sequence Length: 256
- Adapter Parameter Reduction Factor: 16

### Task Adapter Training: Dependency Parsing

- Train batch size: 36
- Evaluation batch size: 8
- Training Epochs: 5
- Learning Rate: 5e-4
- Maximum Sequence Length: 256
- Adapter Parameter Reduction Factor: 16

| Combination | Count |
|-------------|-------|
| $ -A, -B $  | 1     |
| $ -A, +B $  | 1     |
| $ +A, -B $  | 3     |
| $ +A, +B $  | 2     |

Table 6: Bilingual interaction counts

### Task Adapter Training: POS Tagging

- Train batch size: 36
- Evaluation batch size: 8
- Training Epochs: 5
- Learning Rate: 5e-4
- Maximum Sequence Length 256
- Adapter Parameter Reduction Factor: 16

### Task Adapter Training: NER

- Train batch size: 36
- Evaluation batch size: 8
- Training Epochs: 5
- Learning Rate: 5e-4
- Maximum Sequence Length: 256
- Adapter Parameter Reduction Factor: 16

### Task Adapter Training: NLI

- Train batch size: 32
- Evaluation batch size: 8
- Training Epochs: 5
- Learning Rate: 5e-5
- Maximum Sequence Length: 128
- Adapter Parameter Reduction Factor: 16

### Task Adapter Training: Extractive QA

- Train batch size: 32
- Evaluation batch size: 8
- Training Epochs: 5
- Learning Rate: 3e-5
- Maximum Sequence Length: 384
- Document Stride: 128
- Adapter Parameter Reduction Factor: 16

## F Language Interference Projection (an example)

For example, consider the case of Arabic [A] that interacts with Bengali [B] in a bilingual setting [AB]. The count from pair combinations of positive and negative interference counts are as follows:

So for language A we get,

$$\begin{aligned} C &= 1 + 1 + 3 + 2 \\ &= 7 \\ (\mathbf{x}_A, \mathbf{y}_A) &= \frac{1}{7} \times (1 \cdot [-1, -1] + 1 \cdot [-1, 1] \\ &\quad + 3 \cdot [1, -1] + 2 \cdot [1, 1]) \\ &= (0.43, -0.14) \end{aligned}$$

Here,  $|+A, +B| = 2$  means, in total two cases, Arabic gets positive interference score while the other associated language (Bengali) also gets positive interference. Similarly,  $|-A, -B| = 1$  means, for one language, both Arabic and Bengali get negative interference scores. Now  $(X_A, Y_A) = (0.43, -0.14)$ . So Arabic will be in the lower-right quartile of the graph (+x, -y) means, Arabic generally gets positive interference but it does not equally beneficial to other languages (gets penalized for cases  $|-A, -B|, |+A, -B|$ ). Here we consider only Bengali as a language to interact with. In practice, we use a set of other transfer languages to compute the total count of each combination for one specific language.

## G Comparison

### H Transfer Language Ranking

We rank the transfer languages by aggregating all the transfer scores. For example, consider getting transfer scores  $\{ts_1, ..ts_i, ..ts_n\}$  for a set of  $n$  target languages  $L_{tg}$  where  $i \in L_{tg}$  and the transfer language is  $tf$ . Then the aggregated transfer score for  $tf$  would be:

$$\text{aggregated-transfer}(tf) = \frac{\sum_{i=1}^n ts_i}{n}$$

The ranking of all transfer languages across three tasks are presented in Table ???. In addition, we report the percentage of positive transfers for each transfer language. Both in parsing and POS tagging, we observe significant presence of unseen languages in high ranked positions (percentage of unseen languages in top 10: parsing: 40%, POS tagging: 40%, NER: 20%). At the sametime, they provide positive scores similar to the cases of seen languages. On the contrary, in NER, we observe most of the unseen African languages are at the lower ranked positions.

### I Recipient Transfer Maps

In a similar manner of calculating the aggregated-transfer, we calculate

aggregated-target. For example, if a target language  $tg$  receives scores  $\{ts_1, ..ts_i, ..ts_m\}$  from a set of  $m$  transfer languages  $L_{tf}$  where  $i \in L_{tf}$ . Then the aggregated target score for  $tf$  would be:

$$\text{aggregated-target}(tg) = \frac{\sum_{i=1}^n ts_i}{n}$$

This way we identify how much a target language get benefited from all the transfer languages. In Figure 5, we present the Recipient Transfer Maps across tasks. We plot the percentage of positive/negative aggregated-target scores and corresponding target languages. Now looking at these maps, we observe the presence of universal target languages (2-5 %) which always receive positive transfer from all of the 38 source languages in two out of three tasks (exception: POS tagging). Wheres, around 28% languages in parsing and tagging, 32.8% in NER never receive any positive transfer. We observe out of 40 languages which receive positive transfer in more than 90% times, 25 languages are unseen low resourced languages. The complete list of target languages which never receive and which almost always receive positive transfer is presented in Table 19.

### J Maximum Score Recipients are low-resourced

In Table 20, we report all the recipients those receive maximum transfer scores at different steps of mlm fine-tuning. From the results, it is evident that, a multilingual model almost always benefits certain unseen, low-resource as well as endangered languages largely. We observe out of 19 max-recipients, 17 are mBERT-unseen languages. Moreover, the two other seen-languages: Bavarian German and Low German are also low-resourced languages.

### K Task Matters

In Figure 6, we present the commonality graph of transfer language ranking across all three tasks. Spearman rank correlation with p value is presented in Table 21.

### L Transfer Languages with High Variance

In Figure 7, we present the violin plots for all the transfer languages sorted by their

| Dependency Parsing |               |                  |       |       |       |                   |       |       |             |                |         |
|--------------------|---------------|------------------|-------|-------|-------|-------------------|-------|-------|-------------|----------------|---------|
| lang               | mBERT<br>base | Continious Steps |       |       |       | Lang. Interaction |       |       | Improvement |                |         |
|                    |               | 1                | 10    | 100   | 1000  | [1A]              | [2A]  | [3A]  | $Imp_{c:1}$ | $Imp_{c:1000}$ | $Imp_i$ |
| ell                | 92.82         | 92.73            | 92.39 | 92.09 | 91.46 | 91.97             | 91.91 | 91.98 | no          | no             | no      |
| tel                | 90.15         | 89.33            | 89.43 | 87.92 | 85.01 | 89.74             | 89.44 | 89.46 | no          | no             | no      |
| spa                | 90.02         | 89.87            | 89.42 | 89.00 | 88.16 | 89.09             | 89.20 | 89.24 | no          | no             | no      |
| hin                | 89.04         | 88.77            | 88.35 | 87.76 | 87.09 | 88.28             | 88.22 | 88.29 | no          | no             | no      |
| hun                | 87.20         | 87.14            | 86.49 | 85.65 | 85.06 | 86.25             | 86.29 | 86.36 | no          | no             | no      |
| heb                | 85.59         | 85.36            | 85.06 | 84.56 | 83.82 | 85.01             | 84.96 | 84.97 | no          | no             | no      |
| swe                | 85.45         | 85.32            | 85.03 | 84.88 | 84.30 | 84.76             | 84.81 | 84.88 | no          | no             | no      |
| tam                | 84.48         | 84.37            | 82.84 | 83.04 | 81.68 | 83.70             | 83.46 | 83.40 | no          | no             | no      |
| cym                | 83.26         | 83.15            | 82.58 | 82.45 | 81.95 | 83.11             | 83.06 | 83.10 | no          | no             | no      |
| hye                | 82.27         | 81.88            | 81.49 | 81.08 | 80.22 | 80.95             | 81.02 | 81.02 | no          | no             | no      |
| <b>pcm</b>         | 81.04         | 80.32            | 79.11 | 78.47 | 77.91 | 79.32             | 79.50 | 79.52 | no          | no             | no      |
| est                | 80.78         | 80.68            | 80.03 | 79.80 | 79.03 | 79.88             | 79.96 | 80.04 | no          | no             | no      |
| gle                | 79.65         | 79.21            | 79.03 | 78.66 | 78.99 | 79.09             | 79.09 | 79.14 | no          | no             | no      |
| zho                | 70.42         | 70.41            | 70.10 | 69.66 | 70.17 | 64.74             | 70.06 | 70.46 | no          | no             | yes     |
| <b>wol</b>         | 69.46         | 68.66            | 68.09 | 64.63 | 67.32 | 68.96             | 69.05 | 69.10 | no          | no             | no      |
| ara                | 31.25         | 31.15            | 31.29 | 30.99 | 29.97 | 30.45             | 30.38 | 30.37 | no          | no             | no      |
| fra                | 27.84         | 27.56            | 26.46 | 24.70 | 24.19 | 24.92             | 24.96 | 24.95 | no          | no             | no      |
| jpn                | 22.54         | 22.50            | 22.52 | 22.73 | 22.49 | 22.43             | 22.43 | 22.42 | no          | no             | no      |
| deu                | 89.37         | 89.40            | 88.67 | 88.39 | 87.43 | 89.10             | 89.28 | 89.35 | yes         | no             | no      |
| bul                | 89.32         | 89.32            | 89.01 | 89.01 | 89.09 | 89.12             | 89.18 | 89.21 | yes         | no             | no      |
| rus                | 88.09         | 88.14            | 87.96 | 87.53 | 87.03 | 87.82             | 87.88 | 87.94 | yes         | no             | no      |
| eng                | 79.62         | 79.76            | 79.56 | 79.39 | 78.86 | 80.18             | 80.18 | 80.24 | yes         | no             | yes     |
| ben                | 75.31         | 76.69            | 73.53 | 73.00 | 69.31 | 74.69             | 75.71 | 75.87 | yes         | no             | yes     |
| bre                | 70.79         | 71.90            | 69.82 | 71.04 | 72.12 | 73.58             | 73.76 | 73.86 | yes         | yes            | yes     |
| kor                | 64.02         | 64.33            | 64.31 | 63.76 | 65.02 | 64.33             | 64.35 | 64.43 | yes         | yes            | yes     |
| fin                | 63.54         | 64.58            | 63.45 | 63.26 | 63.84 | 64.22             | 64.37 | 64.45 | yes         | yes            | yes     |
| yor                | 40.92         | 42.80            | 40.40 | 42.17 | 47.59 | 42.55             | 42.84 | 42.93 | yes         | yes            | yes     |
| <b>kmr</b>         | 31.94         | 32.44            | 31.73 | 32.75 | 45.30 | 32.54             | 32.10 | 32.03 | yes         | yes            | yes     |
| <b>bam</b>         | 29.99         | 30.43            | 30.87 | 29.95 | 38.13 | 30.74             | 30.81 | 30.78 | yes         | yes            | yes     |
| <b>gub</b>         | 21.64         | 21.97            | 20.96 | 22.92 | 34.52 | 23.83             | 23.73 | 23.52 | yes         | yes            | yes     |

Table 7: Dependency Parsing results. Improvement:  $Imp_{c:1}$  cont. step 1-base.  $Imp_{c:1000}$ : cont. steps 1000-base.  $Imp_i$ : improvement in language interactions ([1A]/[2A]/[3A]) versus baseline. Languages unseen by mBERT are in **bold** font.

| Dependency Parsing (Model Comparison) |       |       |       |       |       |        |       |       |       |        |
|---------------------------------------|-------|-------|-------|-------|-------|--------|-------|-------|-------|--------|
| lang                                  | mBERT |       |       |       |       | XLM-R  |       |       |       |        |
|                                       | base  | 1     | 10    | 100   | 1000  | base_x | 1_x   | 10_x  | 100_x | 1000_x |
| ell                                   | 92.82 | 92.73 | 92.39 | 92.09 | 91.46 | 93.33  | 93.27 | 93.01 | 92.59 | 92.86  |
| tel                                   | 90.15 | 89.33 | 89.43 | 87.92 | 85.01 | 88.49  | 87.88 | 86.34 | 85.99 | 86.41  |
| spa                                   | 90.02 | 89.87 | 89.42 | 89.00 | 88.16 | 89.66  | 89.70 | 89.38 | 88.92 | 89.26  |
| deu                                   | 89.37 | 89.40 | 88.67 | 88.39 | 87.43 | 89.71  | 89.63 | 89.09 | 88.74 | 89.30  |
| bul                                   | 89.32 | 89.32 | 89.01 | 89.01 | 89.09 | 90.19  | 90.31 | 90.22 | 89.73 | 90.19  |
| hin                                   | 89.04 | 88.77 | 88.35 | 87.76 | 87.09 | 89.46  | 89.50 | 89.13 | 88.06 | 88.41  |
| rus                                   | 88.09 | 88.14 | 87.96 | 87.53 | 87.03 | 88.72  | 88.76 | 88.62 | 88.16 | 88.45  |
| hun                                   | 87.20 | 87.14 | 86.49 | 85.65 | 85.06 | 89.16  | 89.13 | 88.78 | 88.40 | 88.73  |
| heb                                   | 85.59 | 85.36 | 85.06 | 84.56 | 83.82 | 86.22  | 86.14 | 85.87 | 85.54 | 85.69  |
| swe                                   | 85.45 | 85.32 | 85.03 | 84.88 | 84.30 | 85.88  | 85.84 | 85.55 | 85.18 | 85.47  |
| tam                                   | 84.48 | 84.37 | 82.84 | 83.04 | 81.68 | 84.59  | 84.66 | 84.32 | 84.35 | 85.35  |
| cym                                   | 83.26 | 83.15 | 82.58 | 82.45 | 81.95 | 83.53  | 83.32 | 82.75 | 82.38 | 83.11  |
| hye                                   | 82.27 | 81.88 | 81.49 | 81.08 | 80.22 | 84.61  | 84.54 | 84.18 | 83.79 | 84.28  |
| pcm                                   | 81.04 | 80.32 | 79.11 | 78.47 | 77.91 | 79.58  | 79.38 | 78.48 | 77.87 | 79.37  |
| est                                   | 80.78 | 80.68 | 80.03 | 79.80 | 79.03 | 83.96  | 83.96 | 83.52 | 83.20 | 83.55  |
| gle                                   | 79.65 | 79.21 | 79.03 | 78.66 | 78.99 | 81.38  | 81.42 | 80.62 | 80.05 | 80.79  |
| eng                                   | 79.62 | 79.76 | 79.56 | 79.39 | 78.86 | 78.99  | 78.77 | 78.42 | 77.86 | 78.27  |
| ben                                   | 75.31 | 76.69 | 73.53 | 73.00 | 69.31 | 69.06  | 69.09 | 68.88 | 66.28 | 69.94  |
| bre                                   | 70.79 | 71.90 | 69.82 | 71.04 | 72.12 | 63.88  | 63.51 | 62.70 | 61.99 | 65.45  |
| zho                                   | 70.42 | 70.41 | 70.10 | 69.66 | 70.17 | 70.47  | 70.59 | 70.39 | 70.11 | 70.31  |
| wol                                   | 69.46 | 68.66 | 68.09 | 64.63 | 67.32 | 67.95  | 67.78 | 66.51 | 64.33 | 64.88  |
| kor                                   | 64.02 | 64.33 | 64.31 | 63.76 | 65.02 | 63.83  | 63.58 | 63.19 | 63.59 | 64.54  |
| fin                                   | 63.54 | 64.58 | 63.45 | 63.26 | 63.84 | 69.12  | 68.73 | 68.07 | 68.63 | 69.62  |
| yor                                   | 40.92 | 42.80 | 40.40 | 42.17 | 47.59 | 23.22  | 22.70 | 22.70 | 24.24 | 38.40  |
| kmr                                   | 31.94 | 32.44 | 31.73 | 32.75 | 45.30 | 64.53  | 64.08 | 62.50 | 64.94 | 66.41  |
| ara                                   | 31.25 | 31.15 | 31.29 | 30.99 | 29.97 | 9.42   | 9.54  | 10.04 | 9.84  | 8.82   |
| bam                                   | 29.99 | 30.43 | 30.87 | 29.95 | 38.13 | 29.68  | 29.70 | 29.28 | 29.16 | 34.20  |
| fra                                   | 27.84 | 27.56 | 26.46 | 24.70 | 24.19 | 19.59  | 19.99 | 18.49 | 16.18 | 19.34  |
| jpn                                   | 22.54 | 22.50 | 22.52 | 22.73 | 22.49 | 7.87   | 7.65  | 7.84  | 7.30  | 6.91   |
| gub                                   | 21.64 | 21.97 | 20.96 | 22.92 | 34.52 | 22.22  | 21.82 | 22.09 | 23.43 | 36.49  |
| Avg.                                  | 69.26 | 69.34 | 68.67 | 68.37 | 69.24 | 68.28  | 68.17 | 67.70 | 67.36 | 69.16  |

Table 8: Dependency Parsing results comparison using mBERT and XLM-R (for languages present in both transfer and target set.)

| POS Tagging |               |                  |       |       |       |                   |       |       |             |                |         |
|-------------|---------------|------------------|-------|-------|-------|-------------------|-------|-------|-------------|----------------|---------|
| lang        | mBERT<br>base | Continious Steps |       |       |       | Lang. Interaction |       |       | Improvement |                |         |
|             |               | 1                | 10    | 100   | 1000  | [1A]              | [2A]  | [3A]  | $Imp_{c:1}$ | $Imp_{c:1000}$ | $Imp_i$ |
| spa         | 97.84         | 97.80            | 97.66 | 97.60 | 97.41 | 97.67             | 97.71 | 97.72 | no          | no             | no      |
| ell         | 97.25         | 97.16            | 97.17 | 96.99 | 96.87 | 97.02             | 97.09 | 97.10 | no          | no             | no      |
| heb         | 95.22         | 95.07            | 94.91 | 94.72 | 94.52 | 94.86             | 94.89 | 94.89 | no          | no             | no      |
| swe         | 95.17         | 95.04            | 95.00 | 94.92 | 94.90 | 94.97             | 95.01 | 95.02 | no          | no             | no      |
| hun         | 94.28         | 94.17            | 94.03 | 93.94 | 93.82 | 93.91             | 94.01 | 94.03 | no          | no             | no      |
| rus         | 94.10         | 94.08            | 94.03 | 93.84 | 93.59 | 93.89             | 93.92 | 93.94 | no          | no             | no      |
| <b>pcm</b>  | 92.98         | 92.77            | 92.17 | 91.84 | 91.23 | 92.25             | 92.46 | 92.55 | no          | no             | no      |
| hin         | 92.23         | 92.00            | 91.76 | 91.61 | 91.39 | 91.98             | 92.00 | 92.01 | no          | no             | no      |
| est         | 91.12         | 90.90            | 90.65 | 90.48 | 90.65 | 90.68             | 90.77 | 90.80 | no          | no             | no      |
| hye         | 91.08         | 90.78            | 90.50 | 90.13 | 89.58 | 90.74             | 90.86 | 90.89 | no          | no             | no      |
| cym         | 89.69         | 89.36            | 89.03 | 88.83 | 88.60 | 88.96             | 89.15 | 89.22 | no          | no             | no      |
| tel         | 88.60         | 88.42            | 88.46 | 87.83 | 87.81 | 87.94             | 87.95 | 87.93 | no          | no             | no      |
| gle         | 88.29         | 88.08            | 87.62 | 87.13 | 87.49 | 87.72             | 87.81 | 87.83 | no          | no             | no      |
| <b>wol</b>  | 85.58         | 84.85            | 84.16 | 82.06 | 82.08 | 84.11             | 84.64 | 84.82 | no          | no             | no      |
| eng         | 84.65         | 84.64            | 84.63 | 84.62 | 84.58 | 84.62             | 84.74 | 84.77 | no          | no             | yes     |
| tam         | 83.10         | 82.74            | 82.19 | 82.39 | 82.38 | 82.87             | 82.72 | 82.70 | no          | no             | no      |
| ben         | 80.34         | 79.35            | 78.56 | 79.29 | 79.56 | 81.10             | 80.43 | 80.39 | no          | no             | yes     |
| <b>bam</b>  | 30.30         | 30.27            | 30.74 | 33.92 | 49.49 | 30.65             | 30.51 | 30.45 | no          | yes            | yes     |
| <b>gub</b>  | 28.49         | 28.11            | 28.66 | 30.02 | 36.64 | 28.82             | 28.77 | 28.85 | no          | yes            | yes     |
| jpn         | 7.85          | 7.73             | 7.80  | 7.91  | 7.84  | 7.58              | 7.67  | 7.68  | no          | no             | no      |
| bul         | 96.12         | 96.13            | 96.07 | 96.08 | 96.12 | 96.05             | 96.01 | 96.01 | yes         | no             | no      |
| deu         | 90.55         | 90.56            | 90.47 | 90.13 | 90.22 | 90.68             | 90.69 | 90.70 | yes         | no             | yes     |
| zho         | 80.45         | 80.50            | 80.54 | 80.55 | 79.72 | 79.34             | 79.71 | 79.91 | yes         | no             | no      |
| fin         | 77.98         | 78.30            | 77.83 | 77.78 | 78.36 | 77.82             | 77.83 | 77.83 | yes         | yes            | no      |
| bre         | 66.91         | 67.28            | 67.67 | 68.15 | 70.26 | 68.02             | 67.79 | 67.72 | yes         | yes            | yes     |
| kor         | 56.28         | 56.42            | 56.49 | 56.61 | 57.72 | 56.59             | 56.58 | 56.57 | yes         | yes            | yes     |
| yor         | 45.91         | 48.22            | 46.73 | 51.24 | 57.28 | 45.71             | 45.45 | 45.45 | yes         | yes            | no      |
| <b>kmr</b>  | 40.16         | 40.35            | 40.49 | 42.76 | 55.82 | 41.04             | 40.79 | 40.64 | yes         | yes            | yes     |
| fra         | 16.35         | 16.47            | 16.66 | 16.63 | 16.24 | 16.77             | 16.79 | 16.79 | yes         | no             | yes     |
| ara         | 8.61          | 8.70             | 8.76  | 8.53  | 5.17  | 8.74              | 8.86  | 8.88  | yes         | no             | yes     |

Table 9: POS Tagging results. Improvement:  $Imp_{c:1}$  cont. step 1-base.  $Imp_{c:1000}$ : cont. steps 1000-base.  $Imp_i$ : improvement in language interactions ([1A]/[2A]/[3A]) versus baseline. Languages unseen by mBERT are in **bold** font.



| POS Tagging (Model Comparison) |       |       |       |       |       |        |       |       |       |        |
|--------------------------------|-------|-------|-------|-------|-------|--------|-------|-------|-------|--------|
| lang                           | mBERT |       |       |       |       | XLM-R  |       |       |       |        |
|                                | base  | 1     | 10    | 100   | 1000  | base_x | 1_x   | 10_x  | 100_x | 1000_x |
| spa                            | 97.84 | 97.80 | 97.66 | 97.60 | 97.41 | 97.80  | 97.78 | 97.74 | 97.65 | 97.61  |
| ell                            | 97.25 | 97.16 | 97.17 | 96.99 | 96.87 | 97.51  | 97.50 | 97.47 | 97.43 | 97.45  |
| bul                            | 96.12 | 96.13 | 96.07 | 96.08 | 96.12 | 96.79  | 96.75 | 96.68 | 96.58 | 96.58  |
| heb                            | 95.22 | 95.07 | 94.91 | 94.72 | 94.52 | 96.41  | 96.30 | 96.16 | 96.16 | 96.24  |
| swe                            | 95.17 | 95.04 | 95.00 | 94.92 | 94.90 | 96.22  | 96.20 | 96.29 | 96.26 | 96.17  |
| hun                            | 94.28 | 94.17 | 94.03 | 93.94 | 93.82 | 95.53  | 95.57 | 95.51 | 95.24 | 95.17  |
| rus                            | 94.10 | 94.08 | 94.03 | 93.84 | 93.59 | 94.45  | 94.40 | 94.30 | 94.32 | 94.32  |
| pcm                            | 92.98 | 92.77 | 92.17 | 91.84 | 91.23 | 93.44  | 93.20 | 91.91 | 92.24 | 92.39  |
| hin                            | 92.23 | 92.00 | 91.76 | 91.61 | 91.39 | 93.47  | 93.42 | 93.31 | 92.96 | 93.12  |
| est                            | 91.12 | 90.90 | 90.65 | 90.48 | 90.65 | 93.46  | 93.42 | 93.30 | 93.16 | 93.35  |
| hye                            | 91.08 | 90.78 | 90.50 | 90.13 | 89.58 | 93.76  | 93.80 | 93.70 | 93.43 | 93.52  |
| deu                            | 90.55 | 90.56 | 90.47 | 90.13 | 90.22 | 90.04  | 90.03 | 90.01 | 90.00 | 90.00  |
| cym                            | 89.69 | 89.36 | 89.03 | 88.83 | 88.60 | 91.92  | 91.83 | 91.60 | 91.55 | 91.50  |
| tel                            | 88.60 | 88.42 | 88.46 | 87.83 | 87.81 | 91.58  | 91.78 | 91.17 | 90.91 | 91.38  |
| gle                            | 88.29 | 88.08 | 87.62 | 87.13 | 87.49 | 91.51  | 91.49 | 91.13 | 90.79 | 91.22  |
| wol                            | 85.58 | 84.85 | 84.16 | 82.06 | 82.08 | 84.14  | 83.84 | 83.21 | 81.84 | 81.99  |
| eng                            | 84.65 | 84.64 | 84.63 | 84.62 | 84.58 | 86.30  | 85.83 | 84.72 | 85.59 | 86.40  |
| tam                            | 83.10 | 82.74 | 82.19 | 82.39 | 82.38 | 85.55  | 85.71 | 85.79 | 85.76 | 85.73  |
| zho                            | 80.45 | 80.50 | 80.54 | 80.55 | 79.72 | 85.44  | 85.32 | 85.29 | 85.56 | 85.36  |
| ben                            | 80.34 | 79.35 | 78.56 | 79.29 | 79.56 | 83.65  | 83.36 | 83.37 | 83.35 | 84.36  |
| fin                            | 77.98 | 78.30 | 77.83 | 77.78 | 78.36 | 83.76  | 83.57 | 83.35 | 83.59 | 83.38  |
| bre                            | 66.91 | 67.28 | 67.67 | 68.15 | 70.26 | 61.11  | 60.97 | 61.72 | 61.73 | 64.70  |
| kor                            | 56.28 | 56.42 | 56.49 | 56.61 | 57.72 | 57.17  | 57.06 | 57.06 | 57.10 | 57.17  |
| yor                            | 45.91 | 48.22 | 46.73 | 51.24 | 57.28 | 26.88  | 26.41 | 26.37 | 27.63 | 45.90  |
| kmr                            | 40.16 | 40.35 | 40.49 | 42.76 | 55.82 | 74.85  | 74.96 | 75.78 | 76.26 | 76.95  |
| bam                            | 30.30 | 30.27 | 30.74 | 33.92 | 49.49 | 29.46  | 29.22 | 29.49 | 29.61 | 36.49  |
| gub                            | 28.49 | 28.11 | 28.66 | 30.02 | 36.64 | 29.97  | 30.17 | 31.05 | 31.48 | 41.21  |
| fra                            | 16.35 | 16.47 | 16.66 | 16.63 | 16.24 | 14.16  | 14.14 | 14.28 | 13.84 | 13.59  |
| ara                            | 8.61  | 8.70  | 8.76  | 8.53  | 5.17  | 8.15   | 8.27  | 8.36  | 8.38  | 7.03   |
| jpn                            | 7.85  | 7.73  | 7.80  | 7.91  | 7.84  | 7.61   | 7.60  | 7.44  | 7.46  | 7.34   |
| Avg.                           | 72.92 | 72.87 | 72.71 | 72.95 | 74.24 | 74.40  | 74.33 | 74.25 | 74.26 | 75.59  |

Table 10: POS Tagging results comparison using mBERT and XLM-R (for languages present in both transfer and target set.)

| NER        |               |                  |       |       |       |                   |       |       |             |                |         |
|------------|---------------|------------------|-------|-------|-------|-------------------|-------|-------|-------------|----------------|---------|
| lang       | mBERT<br>base | Continious Steps |       |       |       | Lang. Interaction |       |       | Improvement |                |         |
|            |               | 1                | 10    | 100   | 1000  | [1A]              | [2A]  | [3A]  | $Imp_{c:1}$ | $Imp_{c:1000}$ | $Imp_i$ |
| spa        | 90.66         | 90.56            | 90.18 | 89.41 | 87.74 | 90.20             | 90.26 | 90.26 | no          | no             | no      |
| bul        | 89.40         | 89.28            | 89.07 | 88.62 | 87.54 | 89.09             | 89.09 | 89.10 | no          | no             | no      |
| fra        | 88.14         | 87.95            | 87.56 | 86.53 | 85.19 | 87.63             | 87.77 | 87.80 | no          | no             | no      |
| fin        | 87.63         | 87.60            | 87.44 | 87.15 | 86.43 | 87.54             | 87.46 | 87.46 | no          | no             | no      |
| est        | 87.40         | 87.27            | 86.93 | 86.28 | 85.36 | 86.92             | 87.04 | 87.10 | no          | no             | no      |
| ell        | 85.81         | 85.71            | 84.95 | 84.36 | 83.02 | 85.44             | 85.50 | 85.54 | no          | no             | no      |
| gle        | 83.97         | 82.50            | 81.96 | 80.24 | 80.10 | 81.84             | 82.40 | 82.58 | no          | no             | no      |
| ara        | 83.69         | 83.50            | 82.73 | 80.85 | 79.29 | 83.04             | 83.18 | 83.23 | no          | no             | no      |
| bre        | 83.20         | 83.10            | 82.00 | 80.45 | 79.60 | 82.50             | 82.71 | 82.74 | no          | no             | no      |
| hin        | 82.53         | 82.31            | 81.87 | 80.21 | 76.62 | 82.56             | 82.40 | 82.41 | no          | no             | yes     |
| kor        | 81.67         | 81.66            | 81.27 | 79.75 | 78.22 | 81.43             | 81.41 | 81.42 | no          | no             | no      |
| eng        | 79.43         | 79.33            | 79.03 | 78.52 | 74.96 | 79.07             | 79.18 | 79.20 | no          | no             | no      |
| nep        | 78.33         | 77.32            | 77.97 | 75.63 | 69.80 | 78.15             | 77.71 | 77.62 | no          | no             | no      |
| tam        | 78.10         | 77.42            | 77.11 | 74.53 | 72.03 | 77.20             | 77.00 | 77.04 | no          | no             | no      |
| heb        | 76.53         | 76.41            | 75.92 | 75.09 | 73.56 | 76.19             | 76.04 | 76.07 | no          | no             | no      |
| tel        | 76.04         | 75.37            | 75.12 | 70.86 | 68.86 | 75.54             | 75.06 | 75.03 | no          | no             | no      |
| mya        | 73.15         | 72.92            | 70.71 | 69.51 | 63.59 | 71.88             | 71.54 | 71.70 | no          | no             | no      |
| zho        | 72.07         | 71.94            | 71.62 | 69.65 | 64.61 | 62.44             | 69.03 | 70.95 | no          | no             | no      |
| <b>ibo</b> | 61.06         | 57.30            | 57.21 | 53.72 | 55.40 | 57.52             | 57.76 | 57.70 | no          | no             | no      |
| jpn        | 59.85         | 59.59            | 58.38 | 56.79 | 53.38 | 58.93             | 58.65 | 58.68 | no          | no             | no      |
| swe        | 91.41         | 91.47            | 91.25 | 90.86 | 90.06 | 91.27             | 91.29 | 91.32 | yes         | no             | no      |
| hye        | 90.10         | 90.59            | 90.23 | 87.55 | 83.25 | 90.32             | 90.46 | 90.50 | yes         | no             | yes     |
| hun        | 88.42         | 88.49            | 88.30 | 87.49 | 86.78 | 88.40             | 88.35 | 88.37 | yes         | no             | no      |
| <b>pms</b> | 88.22         | 89.09            | 88.90 | 88.25 | 87.59 | 88.15             | 87.47 | 87.61 | yes         | no             | no      |
| cym        | 85.75         | 85.91            | 85.23 | 83.19 | 81.81 | 85.25             | 85.34 | 85.39 | yes         | no             | no      |
| deu        | 85.53         | 85.62            | 85.38 | 84.67 | 83.32 | 85.07             | 85.25 | 85.32 | yes         | no             | no      |
| rus        | 84.76         | 84.78            | 84.38 | 83.56 | 80.82 | 84.51             | 84.49 | 84.51 | yes         | no             | no      |
| ben        | 84.75         | 84.85            | 83.32 | 80.53 | 72.14 | 83.12             | 83.62 | 83.73 | yes         | no             | no      |
| <b>kin</b> | 72.38         | 72.74            | 71.76 | 68.79 | 68.50 | 70.48             | 71.11 | 71.85 | yes         | no             | no      |
| yor        | 67.53         | 70.33            | 72.11 | 69.40 | 51.34 | 79.11             | 77.58 | 76.04 | yes         | no             | yes     |

Table 11: NER results. Improvement:  $Imp_{c:1}$  cont. step 1-base.  $Imp_{c:1000}$ : cont. steps 1000-base.  $Imp_i$ : improvement in language interactions ([1A]/[2A]/[3A]) versus baseline. Languages unseen by mBERT are in **bold** font.

| NER (Model Comparison) |       |       |       |       |       |        |       |       |       |        |
|------------------------|-------|-------|-------|-------|-------|--------|-------|-------|-------|--------|
| lang                   | mBERT |       |       |       |       | XLM-R  |       |       |       |        |
|                        | base  | 1     | 10    | 100   | 1000  | base_x | 1_x   | 10_x  | 100_x | 1000_x |
| swe                    | 91.41 | 91.47 | 91.25 | 90.86 | 90.06 | 89.83  | 89.91 | 89.96 | 89.71 | 89.61  |
| spa                    | 90.66 | 90.56 | 90.18 | 89.41 | 87.74 | 87.37  | 87.44 | 87.44 | 87.21 | 87.13  |
| hye                    | 90.10 | 90.59 | 90.23 | 87.55 | 83.25 | 89.85  | 89.77 | 89.56 | 89.98 | 89.88  |
| bul                    | 89.40 | 89.28 | 89.07 | 88.62 | 87.54 | 87.63  | 87.70 | 87.66 | 87.49 | 87.50  |
| hun                    | 88.42 | 88.49 | 88.30 | 87.49 | 86.78 | 86.59  | 86.58 | 86.26 | 86.29 | 86.07  |
| pms                    | 88.22 | 89.09 | 88.90 | 88.25 | 87.59 | 87.12  | 87.42 | 87.17 | 88.22 | 90.54  |
| fra                    | 88.14 | 87.95 | 87.56 | 86.53 | 85.19 | 84.54  | 84.53 | 84.28 | 84.18 | 84.20  |
| fin                    | 87.63 | 87.60 | 87.44 | 87.15 | 86.43 | 85.95  | 85.80 | 85.65 | 85.87 | 85.67  |
| est                    | 87.40 | 87.27 | 86.93 | 86.28 | 85.36 | 85.13  | 85.18 | 85.07 | 84.78 | 84.87  |
| ell                    | 85.81 | 85.71 | 84.95 | 84.36 | 83.02 | 84.16  | 84.25 | 84.09 | 84.07 | 83.96  |
| cym                    | 85.75 | 85.91 | 85.23 | 83.19 | 81.81 | 82.74  | 82.21 | 82.37 | 82.06 | 82.25  |
| deu                    | 85.53 | 85.62 | 85.38 | 84.67 | 83.32 | 83.08  | 83.17 | 83.33 | 82.81 | 82.78  |
| rus                    | 84.76 | 84.78 | 84.38 | 83.56 | 80.82 | 82.84  | 82.72 | 82.38 | 82.17 | 82.28  |
| ben                    | 84.75 | 84.85 | 83.32 | 80.53 | 72.14 | 81.42  | 81.82 | 81.52 | 80.14 | 80.20  |
| gle                    | 83.97 | 82.50 | 81.96 | 80.24 | 80.10 | 81.69  | 81.01 | 80.79 | 80.17 | 80.73  |
| ara                    | 83.69 | 83.50 | 82.73 | 80.85 | 79.29 | 80.97  | 80.91 | 80.56 | 80.23 | 80.24  |
| bre                    | 83.20 | 83.10 | 82.00 | 80.45 | 79.60 | 77.32  | 76.92 | 76.97 | 75.94 | 76.81  |
| hin                    | 82.53 | 82.31 | 81.87 | 80.21 | 76.62 | 80.92  | 80.76 | 81.66 | 81.33 | 80.86  |
| kor                    | 81.67 | 81.66 | 81.27 | 79.75 | 78.22 | 75.20  | 75.25 | 75.07 | 74.77 | 74.73  |
| eng                    | 79.43 | 79.33 | 79.03 | 78.52 | 74.96 | 76.56  | 76.56 | 76.82 | 76.21 | 75.35  |
| nep                    | 78.33 | 77.32 | 77.97 | 75.63 | 69.80 | 76.54  | 75.98 | 77.00 | 74.79 | 74.74  |
| tam                    | 78.10 | 77.42 | 77.11 | 74.53 | 72.03 | 76.35  | 76.24 | 76.25 | 75.92 | 75.79  |
| heb                    | 76.53 | 76.41 | 75.92 | 75.09 | 73.56 | 73.41  | 73.20 | 73.10 | 72.91 | 72.80  |
| tel                    | 76.04 | 75.37 | 75.12 | 70.86 | 68.86 | 76.07  | 76.27 | 75.78 | 74.77 | 74.09  |
| mya                    | 73.15 | 72.92 | 70.71 | 69.51 | 63.59 | 73.03  | 73.12 | 74.27 | 74.43 | 72.31  |
| kin                    | 72.38 | 72.74 | 71.76 | 68.79 | 68.50 | 71.23  | 72.72 | 72.00 | 67.94 | 63.44  |
| zho                    | 72.07 | 71.94 | 71.62 | 69.65 | 64.61 | 64.44  | 64.66 | 64.16 | 63.25 | 62.07  |
| yor                    | 67.53 | 70.33 | 72.11 | 69.40 | 51.34 | 72.10  | 72.07 | 74.73 | 68.77 | 76.73  |
| ibo                    | 61.06 | 57.30 | 57.21 | 53.72 | 55.40 | 63.68  | 63.15 | 60.89 | 57.25 | 61.05  |
| jpn                    | 59.85 | 59.59 | 58.38 | 56.79 | 53.38 | 54.92  | 54.80 | 54.31 | 53.07 | 52.81  |
| Avg.                   | 81.25 | 81.10 | 80.66 | 79.08 | 76.36 | 79.09  | 79.07 | 79.04 | 78.22 | 78.38  |

Table 12: NER results comparison using mBERT and XLM-R (for languages present in both transfer and target set.)

| XNLI  |       |                  |       |       |       |
|-------|-------|------------------|-------|-------|-------|
| lang  | base  | Continious Steps |       |       |       |
|       |       | 1                | 10    | 100   | 1000  |
| mBERT |       |                  |       |       |       |
| eng   | 81.02 | 45.42            | 45.01 | 42.86 | 43.18 |
| spa   | 77.33 | 46.59            | 47.52 | 44.87 | 40.96 |
| deu   | 76.27 | 46.79            | 47.71 | 46.63 | 39.57 |
| zho   | 75.43 | 46.45            | 44.74 | 43.23 | 40.07 |
| bul   | 75.13 | 46.57            | 46.70 | 44.85 | 36.58 |
| ell   | 73.89 | 44.51            | 44.38 | 44.21 | 36.28 |
| rus   | 73.59 | 45.78            | 46.28 | 44.24 | 38.53 |
| ara   | 71.36 | 42.82            | 41.44 | 41.37 | 39.34 |
| hin   | 67.68 | 41.74            | 42.61 | 40.19 | 35.02 |
| XLM-R |       |                  |       |       |       |
| eng   | 84.03 | 83.91            | 83.87 | 83.38 | 83.54 |
| spa   | 80.60 | 80.67            | 80.92 | 80.26 | 80.38 |
| bul   | 80.24 | 80.24            | 80.16 | 79.72 | 80.45 |
| deu   | 79.38 | 79.33            | 79.32 | 78.79 | 79.20 |
| rus   | 78.10 | 78.20            | 78.43 | 78.05 | 78.05 |
| ell   | 77.82 | 77.77            | 77.61 | 77.22 | 77.59 |
| zho   | 77.41 | 77.44            | 77.33 | 77.49 | 77.37 |
| ara   | 75.63 | 75.47            | 75.15 | 74.55 | 74.91 |
| hin   | 74.81 | 74.67            | 74.35 | 74.02 | 74.55 |

Table 13: XNLI results for (continuous training) languages present in both transfer and target set.

| TyDiQA |       |                  |       |       |       |
|--------|-------|------------------|-------|-------|-------|
| lang   | base  | Continious Steps |       |       |       |
|        |       | 1                | 10    | 100   | 1000  |
| mBERT  |       |                  |       |       |       |
| tel    | 58.45 | 58.19            | 57.53 | 56.55 | 56.50 |
| eng    | 56.14 | 55.89            | 55.91 | 53.05 | 53.89 |
| ara    | 54.83 | 54.73            | 54.40 | 49.28 | 50.99 |
| rus    | 50.37 | 49.93            | 49.15 | 36.16 | 43.74 |
| fin    | 50.13 | 50.09            | 50.45 | 44.16 | 46.85 |
| kor    | 47.83 | 46.59            | 46.74 | 44.09 | 44.38 |
| ben    | 45.13 | 46.11            | 48.05 | 45.13 | 44.78 |
| XLM-R  |       |                  |       |       |       |
| tel    | 56.20 | 56.46            | 55.65 | 55.72 | 56.35 |
| eng    | 52.50 | 52.82            | 53.18 | 52.89 | 52.70 |
| ara    | 51.57 | 51.69            | 49.16 | 48.02 | 51.13 |
| rus    | 47.41 | 47.32            | 45.04 | 44.19 | 46.26 |
| fin    | 45.65 | 46.24            | 45.88 | 45.10 | 45.19 |
| ben    | 44.25 | 42.39            | 43.27 | 40.97 | 43.45 |
| kor    | 42.03 | 42.32            | 42.90 | 43.01 | 43.22 |

Table 14: TyDiQA results (continuous training)for languages present in both transfer and target set.

| Transfer Languages Ranking using mBERT (Token Classification) |            |       |       |             |       |       |            |       |      |
|---|------------|-------|-------|-------------|-------|-------|------------|-------|------|
| Rank  | Parsing    |       |       | POS Tagging |       |       | NER        |       |      |
|   | Lang       | ts    | +(%)  | lang        | ts    | +(%)  | lang       | ts    | +(%) |
| 1   | mya        | 0.33  | 40.35 | <b>kin</b>  | 0.41  | 35.09 | zho        | 0.16  | 49.6 |
| 2   | ell        | 0.15  | 31.58 | <b>kmr</b>  | 0.36  | 36.84 | tel        | 0.08  | 32.8 |
| 3   | <b>kmr</b> | 0.14  | 35.96 | <b>mos</b>  | 0.27  | 34.21 | hun        | 0.08  | 40.8 |
| 4   | yor        | 0.14  | 33.33 | hye         | 0.27  | 36.84 | heb        | 0.04  | 34.4 |
| 5   | <b>pcm</b> | 0.13  | 31.58 | cym         | 0.22  | 37.72 | est        | 0.03  | 36.8 |
| 6   | nep        | 0.12  | 35.96 | jpn         | 0.18  | 37.72 | cym        | 0.03  | 40.0 |
| 7   | rus        | 0.11  | 32.46 | mya         | 0.17  | 39.47 | eng        | 0.02  | 38.4 |
| 8   | <b>mos</b> | 0.09  | 42.11 | nep         | 0.12  | 31.58 | <b>mos</b> | 0.00  | 32.0 |
| 9   | <b>pms</b> | 0.09  | 30.70 | <b>pms</b>  | 0.08  | 37.72 | tam        | 0.00  | 35.2 |
| 10  | heb        | 0.08  | 30.70 | zho         | -0.04 | 34.21 | <b>hau</b> | -0.07 | 35.2 |
| 11  | tel        | 0.05  | 26.32 | kor         | -0.07 | 31.58 | gle        | -0.07 | 32.8 |
| 12  | <b>ibo</b> | 0.05  | 31.58 | ben         | -0.08 | 32.46 | jpn        | -0.08 | 33.6 |
| 13  | <b>hau</b> | 0.04  | 37.72 | bul         | -0.08 | 24.56 | kor        | -0.09 | 35.2 |
| 14  | gle        | 0.03  | 28.07 | <b>bam</b>  | -0.12 | 30.70 | swe        | -0.11 | 29.6 |
| 15  | <b>wol</b> | 0.03  | 35.96 | ell         | -0.13 | 33.33 | nep        | -0.13 | 31.2 |
| 16  | <b>bam</b> | 0.02  | 32.46 | hin         | -0.13 | 32.46 | mya        | -0.17 | 35.2 |
| 17  | est        | 0.00  | 28.95 | tam         | -0.14 | 32.46 | hye        | -0.17 | 32.0 |
| 18  | hye        | -0.03 | 31.58 | <b>ibo</b>  | -0.14 | 32.46 | bul        | -0.18 | 27.2 |
| 19  | cym        | -0.03 | 28.95 | <b>wol</b>  | -0.14 | 27.19 | deu        | -0.20 | 32.8 |
| 20  | ben        | -0.06 | 31.58 | <b>pcm</b>  | -0.16 | 29.82 | bre        | -0.23 | 35.2 |
| 21  | <b>kin</b> | -0.06 | 29.82 | yor         | -0.16 | 36.84 | spa        | -0.28 | 28.0 |
| 22  | <b>ewe</b> | -0.08 | 38.60 | heb         | -0.26 | 30.70 | fin        | -0.29 | 27.2 |
| 23  | hin        | -0.10 | 32.46 | rus         | -0.28 | 28.07 | ell        | -0.29 | 34.4 |
| 24  | ara        | -0.12 | 31.58 | hun         | -0.29 | 25.44 | <b>pms</b> | -0.29 | 32.0 |
| 25  | deu        | -0.13 | 31.58 | ara         | -0.30 | 31.58 | ara        | -0.32 | 30.4 |
| 26  | <b>gub</b> | -0.14 | 34.21 | tel         | -0.31 | 28.95 | yor        | -0.32 | 27.2 |
| 27  | spa        | -0.18 | 30.70 | <b>hau</b>  | -0.34 | 29.82 | rus        | -0.34 | 26.4 |
| 28  | jpn        | -0.19 | 27.19 | gle         | -0.34 | 28.07 | <b>wol</b> | -0.41 | 32.8 |
| 29  | bul        | -0.21 | 25.44 | <b>gub</b>  | -0.34 | 23.68 | ben        | -0.54 | 27.2 |
| 30  | swe        | -0.21 | 28.95 | fin         | -0.35 | 25.44 | <b>ibo</b> | -0.54 | 29.6 |
| 31  | eng        | -0.22 | 27.19 | eng         | -0.35 | 26.32 | <b>kin</b> | -0.56 | 31.2 |
| 32  | bre        | -0.23 | 28.95 | est         | -0.35 | 29.82 | fra        | -0.59 | 29.6 |
| 33  | hun        | -0.23 | 21.93 | bre         | -0.39 | 31.58 | <b>kmr</b> | -0.61 | 25.6 |
| 34  | tam        | -0.24 | 21.05 | fra         | -0.41 | 28.95 | <b>pcm</b> | -0.69 | 25.6 |
| 35  | fin        | -0.26 | 26.32 | deu         | -0.44 | 26.32 | <b>bam</b> | -0.69 | 30.4 |
| 36  | fra        | -0.37 | 24.56 | spa         | -0.53 | 25.44 | hin        | -0.72 | 21.6 |
| 37  | kor        | -0.38 | 26.32 | swe         | -0.66 | 27.19 | <b>ewe</b> | -0.76 | 24.8 |
| 38  | zho        | -0.48 | 17.54 | <b>ewe</b>  | -0.79 | 23.68 | <b>gub</b> | -0.98 | 24.8 |

Table 15: Transfer Languages ranked by aggregated transfer scores (ts) overall target languages across token classification tasks using mBERT. Languages unseen by mBERT are in **bold** font.

| Transfer Languages Ranking using XLM-R (Token Classification) |            |       |       |             |       |       |            |       |      |
|---|------------|-------|-------|-------------|-------|-------|------------|-------|------|
| Rank  | Parsing    |       |       | POS Tagging |       |       | NER        |       |      |
|   | Lang       | ts    | +(%)  | lang        | ts    | +(%)  | lang       | ts    | +(%) |
| 1   | nep        | 0.02  | 23.68 | hin         | 0.18  | 39.47 | rus        | 1.15  | 40.0 |
| 2   | zho        | -0.00 | 34.21 | ben         | 0.10  | 36.84 | ell        | 1.04  | 41.6 |
| 3   | mya        | -0.01 | 21.93 | mya         | 0.04  | 34.21 | tel        | 0.76  | 41.6 |
| 4   | ben        | -0.08 | 26.32 | nep         | 0.01  | 34.21 | heb        | 0.70  | 46.4 |
| 5   | hin        | -0.09 | 29.82 | bul         | 0.01  | 40.35 | ben        | 0.64  | 36.0 |
| 6   | tam        | -0.09 | 27.19 | eng         | -0.00 | 35.96 | tam        | 0.48  | 44.0 |
| 7   | bre        | -0.11 | 23.68 | bre         | -0.01 | 42.11 | hin        | 0.47  | 44.0 |
| 8   | tel        | -0.13 | 20.18 | ara         | -0.05 | 37.72 | <b>pms</b> | 0.44  | 40.8 |
| 9   | deu        | -0.13 | 24.56 | gle         | -0.12 | 36.84 | bul        | 0.43  | 48.0 |
| 10  | kor        | -0.21 | 21.93 | cym         | -0.13 | 42.98 | hye        | 0.36  | 40.0 |
| 11  | est        | -0.25 | 24.56 | rus         | -0.13 | 31.58 | ara        | 0.33  | 43.2 |
| 12  | swe        | -0.25 | 22.81 | hye         | -0.16 | 36.84 | swe        | 0.32  | 48.8 |
| 13  | <b>pms</b> | -0.25 | 31.58 | tam         | -0.17 | 39.47 | <b>kmr</b> | 0.31  | 46.4 |
| 14  | hye        | -0.29 | 21.05 | heb         | -0.17 | 41.23 | fra        | 0.27  | 44.0 |
| 15  | jpn        | -0.34 | 21.93 | zho         | -0.24 | 25.44 | eng        | 0.25  | 51.2 |
| 16  | fin        | -0.35 | 18.42 | <b>hau</b>  | -0.24 | 38.60 | cym        | 0.22  | 48.8 |
| 17  | cym        | -0.36 | 18.42 | fra         | -0.25 | 36.84 | mya        | 0.22  | 45.6 |
| 18  | heb        | -0.37 | 22.81 | tel         | -0.26 | 40.35 | gle        | 0.21  | 46.4 |
| 19  | eng        | -0.39 | 23.68 | ell         | -0.30 | 39.47 | jpn        | 0.20  | 40.0 |
| 20  | bul        | -0.42 | 19.30 | deu         | -0.31 | 41.23 | fin        | 0.18  | 40.0 |
| 21  | rus        | -0.47 | 17.54 | kor         | -0.31 | 42.11 | hun        | 0.17  | 40.8 |
| 22  | <b>hau</b> | -0.50 | 16.67 | swe         | -0.32 | 41.23 | est        | 0.16  | 47.2 |
| 23  | yor        | -0.50 | 14.91 | spa         | -0.32 | 30.70 | spa        | 0.16  | 44.8 |
| 24  | ell        | -0.52 | 18.42 | est         | -0.34 | 38.60 | deu        | 0.15  | 42.4 |
| 25  | <b>kin</b> | -0.54 | 19.30 | <b>pms</b>  | -0.43 | 29.82 | nep        | 0.13  | 44.0 |
| 26  | gle        | -0.55 | 16.67 | fin         | -0.50 | 29.82 | bre        | 0.13  | 40.0 |
| 27  | fra        | -0.56 | 20.18 | hun         | -0.54 | 28.07 | <b>hau</b> | 0.12  | 42.4 |
| 28  | ara        | -0.57 | 19.30 | <b>kin</b>  | -0.56 | 28.95 | kor        | -0.02 | 44.0 |
| 29  | spa        | -0.66 | 17.54 | <b>kmr</b>  | -0.57 | 28.07 | <b>pcm</b> | -0.04 | 38.4 |
| 30  | hun        | -0.66 | 14.91 | <b>pcm</b>  | -0.70 | 16.67 | <b>wol</b> | -0.11 | 36.8 |
| 31  | <b>bam</b> | -0.74 | 15.79 | jpn         | -0.72 | 32.46 | <b>ibo</b> | -0.12 | 36.0 |
| 32  | <b>kmr</b> | -0.85 | 19.30 | <b>mos</b>  | -0.84 | 28.07 | <b>gub</b> | -0.17 | 36.0 |
| 33  | <b>mos</b> | -0.90 | 18.42 | <b>ewe</b>  | -0.88 | 21.93 | <b>mos</b> | -0.19 | 32.0 |
| 34  | <b>gub</b> | -0.95 | 15.79 | <b>bam</b>  | -0.89 | 20.18 | zho        | -0.20 | 30.4 |
| 35  | <b>ibo</b> | -1.22 | 14.91 | yor         | -0.90 | 26.32 | yor        | -0.20 | 37.6 |
| 36  | <b>wol</b> | -1.40 | 14.91 | <b>wol</b>  | -0.97 | 22.81 | <b>ewe</b> | -0.32 | 30.4 |
| 37  | <b>pcm</b> | -1.55 | 10.53 | <b>ibo</b>  | -1.00 | 23.68 | <b>kin</b> | -0.34 | 29.6 |
| 38  | <b>ewe</b> | -1.83 | 14.04 | <b>gub</b>  | -1.05 | 20.18 | <b>bam</b> | -0.41 | 32.0 |

Table 16: Transfer Languages ranked by aggregated transfer scores (ts) overall target languages across token classification tasks using XLM-R. Languages unseen by mBERT are in **bold** font.

| Transfer Languages Ranking using mBERT (Sentence Classification & QA) |            |        |      |            |        |      |            |       |       |
|---|------------|--------|------|------------|--------|------|------------|-------|-------|
| Rank  | XNLI       |        |      | ANLI       |        |      | TyDiQA     |       |       |
|   | Lang       | ts     | +(%) | lang       | ts     | +(%) | lang       | ts    | +(%)  |
| 1   | <b>hau</b> | -34.42 | 0.0  | <b>bam</b> | -14.97 | 0.0  | zho        | 0.67  | 77.78 |
| 2   | <b>bam</b> | -34.85 | 0.0  | <b>hau</b> | -17.82 | 0.0  | jpn        | 0.08  | 44.44 |
| 3   | <b>gub</b> | -36.40 | 0.0  | <b>gub</b> | -18.35 | 0.0  | gle        | -0.08 | 44.44 |
| 4   | <b>ewe</b> | -36.73 | 0.0  | deu        | -19.79 | 0.0  | <b>wol</b> | -0.12 | 44.44 |
| 5   | hin        | -37.08 | 0.0  | fin        | -19.93 | 0.0  | cym        | -0.14 | 33.33 |
| 6   | deu        | -37.33 | 0.0  | hun        | -20.01 | 0.0  | mya        | -0.15 | 22.22 |
| 7   | kor        | -37.86 | 0.0  | kor        | -20.15 | 0.0  | <b>mos</b> | -0.15 | 44.44 |
| 8   | spa        | -38.17 | 0.0  | zho        | -20.20 | 0.0  | hun        | -0.19 | 22.22 |
| 9   | <b>kmr</b> | -38.25 | 0.0  | hin        | -20.27 | 0.0  | fin        | -0.19 | 44.44 |
| 10  | fin        | -38.32 | 0.0  | <b>pms</b> | -20.28 | 0.0  | <b>kin</b> | -0.22 | 33.33 |
| 11  | rus        | -38.34 | 0.0  | yor        | -20.37 | 0.0  | est        | -0.25 | 33.33 |
| 12  | <b>pms</b> | -38.35 | 0.0  | <b>kin</b> | -20.63 | 0.0  | hye        | -0.25 | 33.33 |
| 13  | hun        | -38.57 | 0.0  | spa        | -20.66 | 0.0  | tel        | -0.26 | 33.33 |
| 14  | heb        | -38.72 | 0.0  | mya        | -20.68 | 0.0  | eng        | -0.28 | 33.33 |
| 15  | swe        | -38.86 | 0.0  | heb        | -20.68 | 0.0  | ell        | -0.29 | 22.22 |
| 16  | est        | -38.86 | 0.0  | <b>ewe</b> | -20.74 | 0.0  | <b>ewe</b> | -0.30 | 33.33 |
| 17  | gle        | -38.87 | 0.0  | rus        | -20.74 | 0.0  | yor        | -0.30 | 33.33 |
| 18  | bul        | -38.90 | 0.0  | est        | -20.91 | 0.0  | heb        | -0.33 | 22.22 |
| 19  | fra        | -38.91 | 0.0  | swe        | -20.93 | 0.0  | <b>pms</b> | -0.34 | 22.22 |
| 20  | yor        | -38.94 | 0.0  | gle        | -20.99 | 0.0  | tam        | -0.38 | 22.22 |
| 21  | ell        | -39.24 | 0.0  | bul        | -21.07 | 0.0  | ben        | -0.39 | 22.22 |
| 22  | <b>kin</b> | -39.37 | 0.0  | ell        | -21.10 | 0.0  | bul        | -0.41 | 33.33 |
| 23  | zho        | -39.44 | 0.0  | ara        | -21.11 | 0.0  | deu        | -0.41 | 22.22 |
| 24  | ara        | -39.50 | 0.0  | <b>kmr</b> | -21.11 | 0.0  | <b>gub</b> | -0.42 | 22.22 |
| 25  | mya        | -39.56 | 0.0  | nep        | -21.17 | 0.0  | nep        | -0.42 | 33.33 |
| 26  | eng        | -39.74 | 0.0  | fra        | -21.22 | 0.0  | swe        | -0.43 | 33.33 |
| 27  | hye        | -40.04 | 0.0  | eng        | -21.28 | 0.0  | kor        | -0.45 | 22.22 |
| 28  | bre        | -40.07 | 0.0  | cym        | -21.39 | 0.0  | hin        | -0.47 | 22.22 |
| 29  | cym        | -40.13 | 0.0  | jpn        | -21.43 | 0.0  | bre        | -0.48 | 11.11 |
| 30  | nep        | -40.25 | 0.0  | tam        | -21.45 | 0.0  | ara        | -0.51 | 33.33 |
| 31  | tel        | -40.31 | 0.0  | tel        | -21.51 | 0.0  | <b>ibo</b> | -0.57 | 33.33 |
| 32  | ben        | -40.31 | 0.0  | hye        | -21.62 | 0.0  | <b>bam</b> | -0.61 | 22.22 |
| 33  | jpn        | -40.53 | 0.0  | bre        | -21.65 | 0.0  | <b>kmr</b> | -0.62 | 33.33 |
| 34  | <b>mos</b> | -41.04 | 0.0  | <b>mos</b> | -21.65 | 0.0  | spa        | -0.66 | 22.22 |
| 35  | tam        | -41.04 | 0.0  | <b>wol</b> | -22.23 | 0.0  | rus        | -0.67 | 22.22 |
| 36  | <b>wol</b> | -42.67 | 0.0  | <b>pcm</b> | -22.24 | 0.0  | <b>hau</b> | -0.89 | 22.22 |
| 37  | <b>pcm</b> | -43.37 | 0.0  | <b>ibo</b> | -22.36 | 0.0  | fra        | -1.04 | 11.11 |
| 38  | <b>ibo</b> | -44.78 | 0.0  | ben        | -23.37 | 0.0  | <b>pcm</b> | -1.10 | 22.22 |

Table 17: Transfer Languages ranked by aggregated transfer scores (ts) overall target languages across Sentence Classification & QA tasks using mBERT. Languages unseen by mBERT are in **bold** font.

| Transfer Languages Ranking using mBERT (Sentence Classification & QA) |            |       |       |            |       |       |            |       |       |
|---|------------|-------|-------|------------|-------|-------|------------|-------|-------|
| Rank  | XNLI       |       |       | ANLI       |       |       | TyDiQA     |       |       |
|   | Lang       | ts    | +(%)  | lang       | ts    | +(%)  | lang       | ts    | +(%)  |
| 1   | <b>ewe</b> | 0.44  | 93.33 | hin        | 2.25  | 100.0 | <b>pcm</b> | -0.42 | 55.56 |
| 2   | bre        | 0.36  | 93.33 | ell        | 1.34  | 80.0  | ell        | -0.44 | 22.22 |
| 3   | <b>bam</b> | 0.30  | 93.33 | nep        | 1.33  | 100.0 | fin        | -0.46 | 44.44 |
| 4   | <b>pcm</b> | 0.30  | 93.33 | ara        | 1.31  | 80.0  | zho        | -0.47 | 44.44 |
| 5   | <b>ibo</b> | 0.28  | 80.00 | swe        | 1.31  | 100.0 | heb        | -0.47 | 11.11 |
| 6   | rus        | 0.22  | 73.33 | tam        | 1.03  | 70.0  | <b>ewe</b> | -0.51 | 55.56 |
| 7   | <b>wol</b> | 0.20  | 80.00 | bul        | 0.93  | 70.0  | tam        | -0.52 | 11.11 |
| 8   | <b>hau</b> | 0.13  | 73.33 | fra        | 0.73  | 70.0  | eng        | -0.53 | 33.33 |
| 9   | heb        | 0.08  | 66.67 | hun        | 0.39  | 60.0  | hin        | -0.55 | 22.22 |
| 10  | <b>kmr</b> | 0.08  | 66.67 | cym        | 0.36  | 60.0  | fra        | -0.56 | 33.33 |
| 11  | <b>pms</b> | 0.06  | 60.00 | deu        | 0.25  | 60.0  | tel        | -0.56 | 11.11 |
| 12  | jpn        | 0.05  | 60.00 | eng        | 0.17  | 70.0  | deu        | -0.64 | 11.11 |
| 13  | zho        | 0.05  | 60.00 | tel        | 0.13  | 60.0  | swe        | -0.67 | 11.11 |
| 14  | bul        | 0.03  | 53.33 | fin        | 0.10  | 40.0  | nep        | -0.67 | 11.11 |
| 15  | fra        | -0.00 | 53.33 | spa        | -0.08 | 60.0  | hun        | -0.69 | 11.11 |
| 16  | spa        | -0.01 | 46.67 | kor        | -0.10 | 50.0  | est        | -0.70 | 22.22 |
| 17  | mya        | -0.04 | 53.33 | rus        | -0.10 | 50.0  | <b>kmr</b> | -0.71 | 44.44 |
| 18  | <b>kin</b> | -0.04 | 40.00 | heb        | -0.10 | 50.0  | rus        | -0.72 | 0.00  |
| 19  | hye        | -0.05 | 53.33 | est        | -0.14 | 60.0  | gle        | -0.73 | 22.22 |
| 20  | deu        | -0.09 | 33.33 | mya        | -0.14 | 40.0  | <b>hau</b> | -0.77 | 22.22 |
| 21  | eng        | -0.11 | 33.33 | ben        | -0.19 | 50.0  | ben        | -0.77 | 11.11 |
| 22  | gle        | -0.11 | 26.67 | gle        | -0.27 | 30.0  | kor        | -0.78 | 0.00  |
| 23  | est        | -0.11 | 33.33 | <b>hau</b> | -0.48 | 60.0  | spa        | -0.79 | 0.00  |
| 24  | <b>mos</b> | -0.11 | 40.00 | zho        | -1.00 | 0.0   | bul        | -0.81 | 0.00  |
| 25  | swe        | -0.12 | 33.33 | <b>kmr</b> | -1.07 | 30.0  | hye        | -0.91 | 0.00  |
| 26  | tel        | -0.13 | 40.00 | hye        | -1.14 | 20.0  | cym        | -0.92 | 22.22 |
| 27  | cym        | -0.14 | 33.33 | jpn        | -1.32 | 10.0  | <b>gub</b> | -1.02 | 11.11 |
| 28  | ara        | -0.18 | 26.67 | <b>pcm</b> | -1.53 | 10.0  | <b>wol</b> | -1.02 | 11.11 |
| 29  | ben        | -0.20 | 20.00 | bre        | -2.10 | 0.0   | <b>ibo</b> | -1.03 | 33.33 |
| 30  | <b>gub</b> | -0.23 | 13.33 | <b>gub</b> | -2.19 | 20.0  | ara        | -1.08 | 0.00  |
| 31  | nep        | -0.23 | 20.00 | <b>pms</b> | -2.60 | 0.0   | <b>bam</b> | -1.12 | 11.11 |
| 32  | kor        | -0.36 | 6.67  | yor        | -2.88 | 0.0   | <b>mos</b> | -1.15 | 11.11 |
| 33  | hin        | -0.39 | 6.67  | <b>kin</b> | -4.28 | 0.0   | jpn        | -1.15 | 0.00  |
| 34  | ell        | -0.45 | 0.00  | <b>mos</b> | -4.54 | 10.0  | bre        | -1.19 | 11.11 |
| 35  | yor        | -0.45 | 13.33 | <b>bam</b> | -4.87 | 0.0   | <b>pms</b> | -1.22 | 0.00  |
| 36  | fin        | -0.46 | 6.67  | <b>wol</b> | -5.01 | 0.0   | <b>kin</b> | -1.30 | 11.11 |
| 37  | tam        | -0.47 | 6.67  | <b>ewe</b> | -5.02 | 10.0  | mya        | -1.49 | 0.00  |
| 38  | hun        | -0.70 | 0.00  | <b>ibo</b> | -5.95 | 0.0   | yor        | -1.57 | 11.11 |

Table 18: Transfer Languages ranked by aggregated transfer scores (ts) overall target languages across Sentence Classification & QA tasks using XLM-R. Languages unseen by mBERT are in **bold** font.



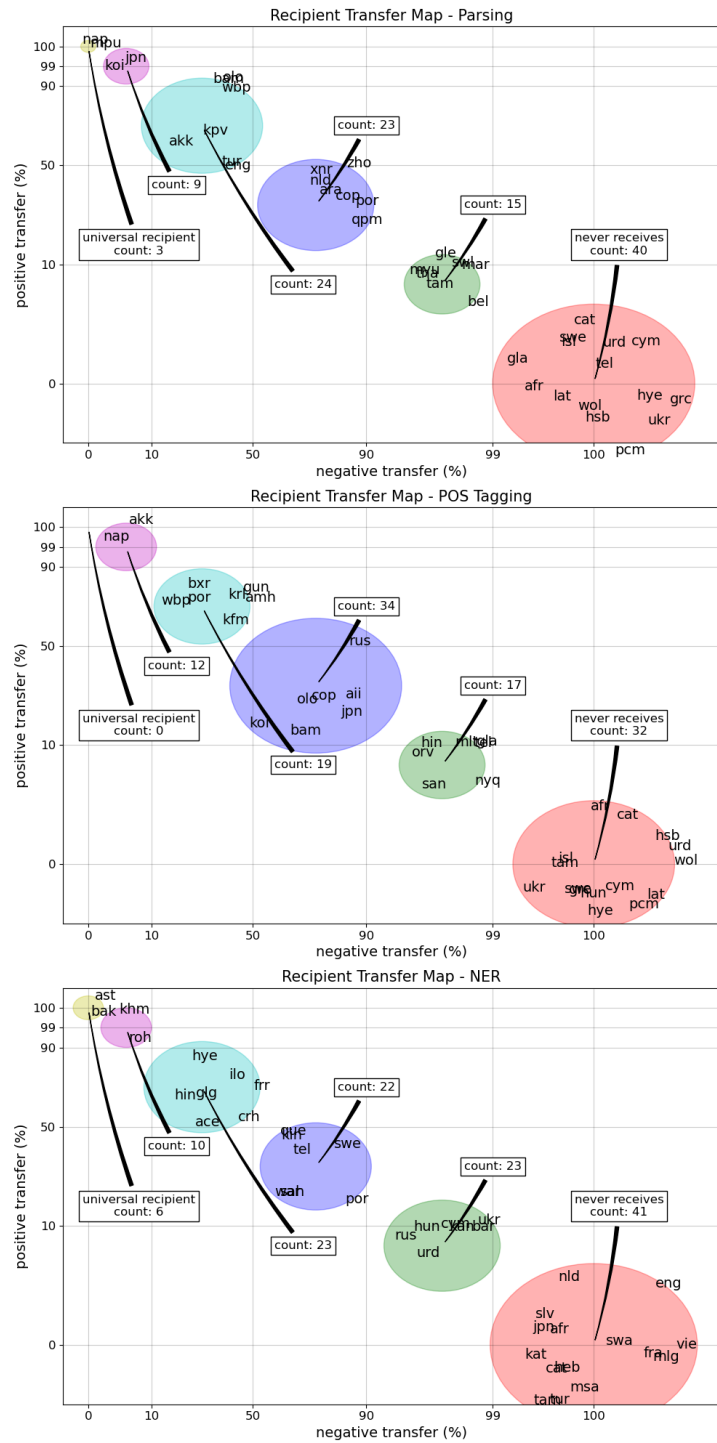


Figure 5: Recipient Transfer Map: we observe universal positive recipients as well as languages those never receive positive transfer across tasks. Circle size represents the percentage of languages fall to a transfer range.

| Tasks              | Never Receives  | Positive Transfer (%) (90-99]   | Universal Recipient                         |
|--------------------|---|---|---|
| Dependency Parsing | ell, isl, <b>grc</b> , <b>mlt</b> , fra, <b>qtd</b> , hrv, lav, <b>urb</b> , <b>fas</b> , ukr, spa, cym, tel, <b>pcm</b> , afr, swe, est, <b>nor</b> , <b>hsb</b> , <b>orv</b> , cat, slv, <b>chu</b> , <b>sme</b> , eus, slk, hye, <b>gla</b> , urd, hin, <b>fro</b> , lit, lat, <b>san</b> , <b>wol</b> , <b>lij</b> , <b>got</b> , srp, <b>lzh</b> | <b>tpn</b> , <b>koi</b> , <b>glv</b> , kor, <b>krl</b> , <b>mdf</b> , jpn, <b>amh</b> , sqi                     | <b>mpu</b> , <b>gun</b> , <b>nap</b>        |
| POS Tagging        | ell, nld, isl, <b>grc</b> , lav, <b>fas</b> , ukr, spa, cym, hun, <b>pcm</b> , afr, swe, est, <b>nor</b> , <b>hsb</b> , tam, cat, <b>chu</b> , <b>sme</b> , eus, hye, urd, ben, ita, ron, lit, lat, <b>wol</b> , <b>got</b> , srp, <b>lzh</b>   | fra, eng, <b>qhe</b> , dan, <b>myu</b> , <b>glv</b> , kor, tur, kaz, <b>akk</b> , <b>myv</b> , <b>nap</b>       |   |
| NER                | nld, ell, tgl, fra, ces, bul, zho, msa, sun, lav, gle, <b>fas</b> , kat, spa, heb, hbs, afr, est, <b>yid</b> , eng, tam, bre, vie, jpn, cat, <b>tha</b> , slv, ceb, tur, mlg, slk, swa, ben, uzb, ita, ron, tat, pol, <b>zea</b> , <b>lin</b> , <b>ibo</b>  | <b>ksh</b> , <b>pms</b> , aze, <b>mzn</b> , <b>oci</b> , tsk, <b>roh</b> , <b>khm</b> , <b>aym</b> , <b>csb</b> | bak, ast, <b>uig</b> , kaz, nds, <b>amh</b> |

Table 19: We find 25 languages out of 40 which receives positive transfer from almost any transfer languages (i.e. column 90-99% and 100%) are unseen by mbert. (language codes in **bold** font are the unseen ones)

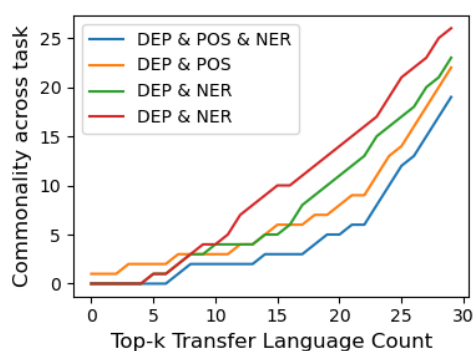


Figure 6: Extent of commonality of top-transfer languages across task. Unseen languages perform generally well while the other language rankings mostly vary across tasks.

aggregated-transfer score variance. We observe, the unseen languages (bold font) are the ones having large amount of variances across all three tasks. We find the languages with high variance can provide superior transfer for some languages but at the same time hurt significantly some other languages. For example, if we consider the case of dependency parsing, we find *ibo* (rank-1) and *bam* (rank-3) are two languages with high variance. They provide maximum amount of positive transfer some universal low-resourced target languages like *akk*, *koi*, *apu*, *tpn* from diverse families including afro-asiatic, uralic, tupian. At the same time, *ibo* also hurts a large number of languages (10) including *fra*, *nyq*, *sme*, *san* etc providing minimum amount of negative transfer. On the other hand, there can be languages with high variance providing either mostly positive aggregated-transfer scores like *mos* or mostly negative score like *pcm*. Interestingly, if we look at the aggregated-transfer score and variance of *pcm* in Table 22, we find the transfer is positive overall. Nevertheless it provides minimum negative scores to 11 languages thus making it a transfer language with high variance. On the other

| Transfer Language | Parsing |            |            |            | POS Tagging |            |            |            | NER |            |            |            |
|-------------------|---------|------------|------------|------------|-------------|------------|------------|------------|-----|------------|------------|------------|
|                   | 1       | 10         | 100        | 1000       | 1           | 10         | 100        | 1000       | 1   | 10         | 100        | 1000       |
| gub               | mpu     | mpu        | gun        | <b>gub</b> | aqz         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | bar        |
| est               | nap     | tpn        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| bre               | nap     | nap        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| eng               | nap     | mpu        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | sin        | <b>amh</b> | nds        |
| ben               | nap     | nap        | nap        | <b>gun</b> | cop         | nap        | nap        | <b>nap</b> | amh | sin        | amh        | <b>amh</b> |
| kmr               | nap     | mpu        | gun        | <b>kmr</b> | urb         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| spa               | nap     | mpu        | gun        | <b>gun</b> | nap         | nap        | <b>nap</b> | nap        | amh | amh        | <b>amh</b> | roh        |
| bul               | nap     | nap        | nap        | <b>gun</b> | nap         | nap        | nap        | <b>amh</b> | amh | amh        | <b>amh</b> | som        |
| pms               | nap     | <b>nap</b> | nap        | mpu        | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| gle               | nap     | nap        | mpu        | <b>gun</b> | aqz         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| nep               | nap     | tpn        | gun        | <b>gun</b> | aqz         | urb        | nap        | <b>nap</b> | amh | sin        | <b>amh</b> | roh        |
| cym               | nap     | nap        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>amh</b> | amh | <b>sin</b> | amh        | som        |
| fin               | nap     | nap        | tpn        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | sin        | <b>amh</b> | som        |
| hye               | nap     | nap        | nap        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | uig | sin        | <b>amh</b> | som        |
| mya               | nap     | nap        | wbp        | <b>gun</b> | aqz         | urb        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| hin               | nap     | tpn        | gun        | <b>gun</b> | aqz         | aqz        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| tel               | nap     | nap        | gun        | <b>gun</b> | aqz         | nap        | nap        | <b>amh</b> | amh | sin        | <b>amh</b> | roh        |
| tam               | nap     | nap        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | sin        | <b>amh</b> | som        |
| kor               | tpn     | mpu        | mpu        | <b>tpn</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | roh        |
| ell               | nap     | tpn        | nap        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| hun               | nap     | mpu        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| heb               | nap     | nap        | nap        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| zho               | tpn     | nap        | nap        | <b>mpu</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| ara               | nap     | nap        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | uig | amh        | <b>amh</b> | amh        |
| swe               | nap     | nap        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | sin        | <b>amh</b> | som        |
| jpn               | nap     | mpu        | mpu        | <b>tpn</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | amh        | <b>amh</b> |
| fra               | nap     | mpu        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| deu               | tpn     | mpu        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |
| rus               | nap     | nap        | gun        | <b>gun</b> | nap         | nap        | <b>nap</b> | nap        | amh | sin        | <b>amh</b> | roh        |
| bam               | mpu     | wbp        | <b>wbp</b> | gun        | nap         | <b>nap</b> | amh        | bam        | amh | uig        | amh        | <b>amh</b> |
| ewe               | nap     | gun        | tpn        | <b>gun</b> | nap         | <b>nap</b> | amh        | nap        | amh | sin        | <b>amh</b> | nds        |
| hau               | mpu     | nap        | gun        | <b>gun</b> | aqz         | <b>nap</b> | nap        | amh        | amh | sin        | amh        | <b>amh</b> |
| ibo               | tpn     | mpu        | gun        | <b>gun</b> | aqz         | <b>nap</b> | nap        | mpu        | sin | amh        | amh        | <b>amh</b> |
| kin               | mpu     | <b>nap</b> | nap        | tpn        | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| mos               | mpu     | aqz        | gun        | <b>gun</b> | nap         | <b>nap</b> | amh        | nap        | sin | sin        | <b>amh</b> | amh        |
| pcm               | nap     | nap        | wbp        | <b>gun</b> | kfm         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | amh        |
| wol               | nap     | aqz        | gun        | <b>gun</b> | nap         | nap        | nap        | <b>amh</b> | amh | sin        | <b>amh</b> | nds        |
| yor               | nap     | nap        | wbp        | <b>gun</b> | nap         | nap        | nap        | <b>nap</b> | amh | amh        | <b>amh</b> | som        |

Table 20: Only bar and nds are seen by mbert. All other languages receiving maximum benefits continuously are unseen by mbert (kfm, urb, gun, aqz, cop, roh, bam, tpn, som, kmr, uig, mpu, amh, sin, wbp, gub, nap). The maximum score across different steps of training are **bolded**.

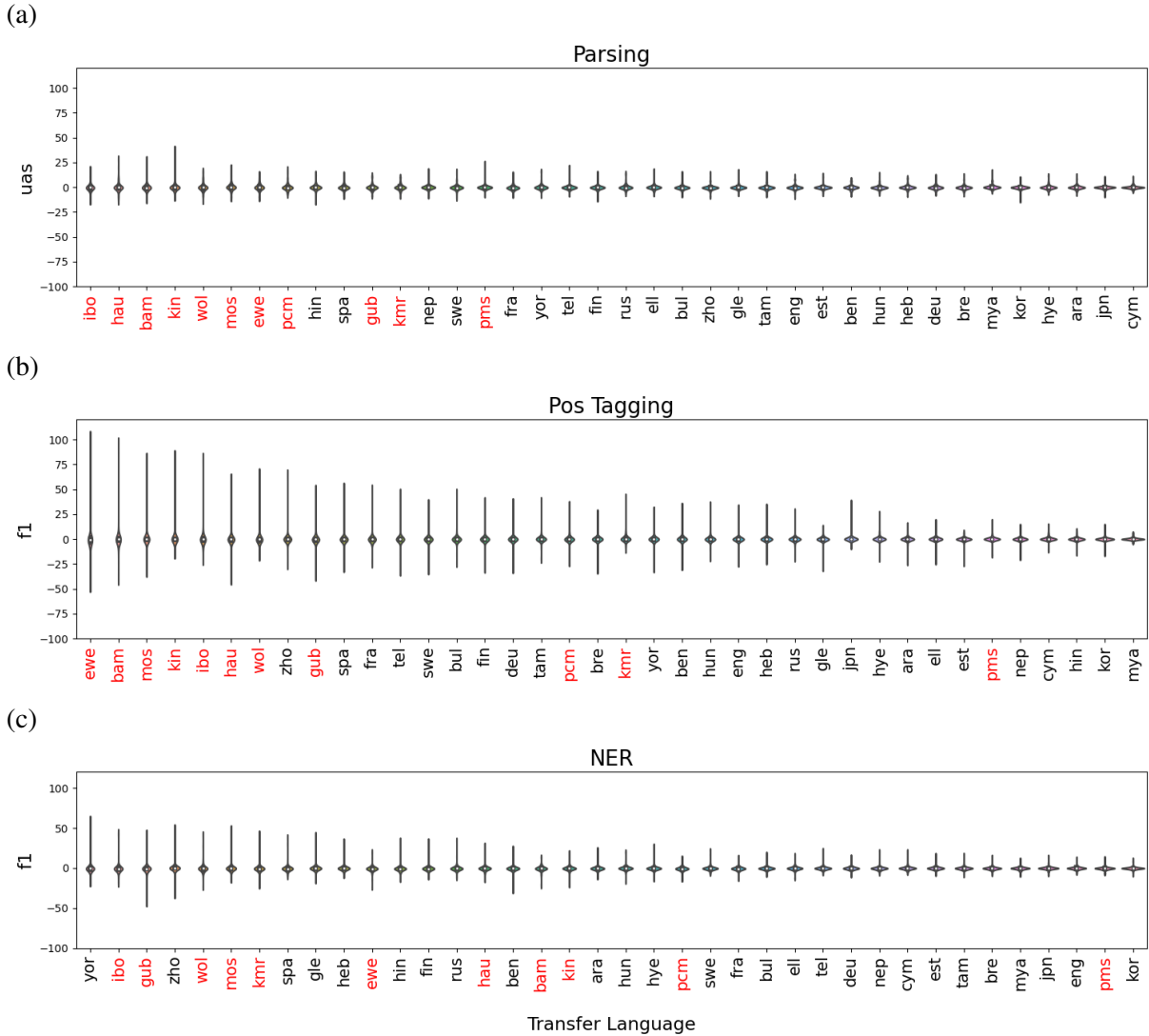


Figure 7: Violin plots of transfer languages sorted by transfer score variance. mBERT unseen languages are in red color font.

|             | DEP                  | POS                  | NER                 |
|-------------|----------------------|----------------------|---------------------|
| DEP         | -                    | <b>(-0.34, 0.04)</b> | <b>(0.40, 0.01)</b> |
| POS Tagging | <b>(-0.34, 0.04)</b> | -                    | (-0.15, 0.37)       |
| NER         | <b>(0.40, 0.01)</b>  | (-0.15, 0.37)        | -                   |

Table 21: (Spearman Rank correlation, p value) for correlation of transfer language ranking across token-classification tasks. Statistically significant relations are in **bold** font.

hand, low variance languages are the ones those do not significantly affect any transfer languages like arabic (rank 37). Though the overall transfer score is negative (-0.12) for arabic, it fails to provide maximum or minimum transfer score to any target language making it neutral. So, overall it is evident that, transfer languages with high variance are the

ones with either (i) mostly positive while significantly hurting a few, (ii) mostly negative while significantly boosting performance for a few, or (iii) Performing both (i) and (ii) concurrently being highly influential as well as detrimental at the same time. Languages unseen by mBERT during pretraining exhibit all three kinds of characteristics with high intensity (see Table 23 for examples). In Table 22, we report the transfer score with variance as well as the count of maximum/minimum transfer score recipients for all transfer languages across tasks.

## M Seen vs Unseen Languages

In Figure 8, we report the aggregated and averaged transfer scores we get for mBERT seen vs unseen languages.

| Rank | Parsing    |                 |              | POS Tagging |                 |              | NER        |                 |              |
|------|------------|-----------------|--------------|-------------|-----------------|--------------|------------|-----------------|--------------|
|      | Lang       | Transfer (Var.) | (Max, Min) # | Lang        | Transfer (Var.) | (Max, Min) # | Lang       | Transfer (Var.) | (Max, Min) # |
| 1    | <b>ibo</b> | 0.05 (23.5)     | (10, 10)     | <b>ewe</b>  | -0.79 (120.4)   | (5, 31)      | yor        | -0.32 (44.4)    | (2, 6)       |
| 2    | <b>hau</b> | 0.04 (22.7)     | (2, 2)       | <b>bam</b>  | -0.12 (101.2)   | (5, 0)       | <b>ibo</b> | -0.54 (42.0)    | (3, 18)      |
| 3    | <b>bam</b> | 0.02 (21.5)     | (11, 15)     | <b>mos</b>  | 0.27 (69.8)     | (4, 3)       | <b>gub</b> | -0.98 (41.2)    | (3, 14)      |
| 4    | <b>kin</b> | -0.06 (21.4)    | (3, 1)       | <b>kin</b>  | 0.41 (69.1)     | (3, 0)       | zho        | 0.16 (32.7)     | (46, 3)      |
| 5    | <b>wol</b> | 0.03 (17.6)     | (5, 6)       | <b>ibo</b>  | -0.14 (68.0)    | (5, 10)      | <b>wol</b> | -0.41 (32.4)    | (5, 6)       |
| 6    | <b>mos</b> | 0.09 (16.1)     | (13, 2)      | <b>hau</b>  | -0.34 (52.5)    | (1, 3)       | <b>mos</b> | 0.0 (29.5)      | (2, 1)       |
| 7    | <b>ewe</b> | -0.08 (14.5)    | (4, 5)       | <b>wol</b>  | -0.14 (51.9)    | (8, 10)      | <b>kmr</b> | -0.61 (25.6)    | (3, 8)       |
| 8    | <b>pcm</b> | 0.13 (13.4)     | (1, 11)      | zho         | -0.04 (45.6)    | (26, 7)      | spa        | -0.28 (24.5)    | (1, 1)       |
| 9    | hin        | -0.1 (13.1)     | (4, 3)       | <b>gub</b>  | -0.34 (41.3)    | (3, 4)       | gle        | -0.07 (21.1)    | (1, 1)       |
| 10   | spa        | -0.18 (11.5)    | (2, 3)       | spa         | -0.53 (39.6)    | (0, 2)       | heb        | 0.04 (21.1)     | (3, 0)       |
| 11   | <b>gub</b> | -0.14 (10.7)    | (5, 3)       | fra         | -0.41 (37.2)    | (2, 6)       | <b>ewe</b> | -0.76 (19.0)    | (0, 5)       |
| 12   | <b>kmr</b> | 0.14 (10.2)     | (0, 1)       | tel         | -0.31 (36.6)    | (0, 1)       | hin        | -0.72 (18.6)    | (1, 12)      |
| 13   | nep        | 0.12 (10.1)     | (8, 1)       | swe         | -0.66 (32.2)    | (1, 3)       | fin        | -0.29 (18.3)    | (0, 1)       |
| 14   | swe        | -0.21 (10.0)    | (0, 2)       | bul         | -0.08 (28.9)    | (0, 0)       | rus        | -0.34 (17.4)    | (0, 1)       |
| 15   | <b>pms</b> | 0.09 (9.7)      | (3, 1)       | fin         | -0.35 (28.9)    | (0, 1)       | <b>hau</b> | -0.07 (17.1)    | (7, 0)       |
| 16   | fra        | -0.37 (9.1)     | (1, 8)       | deu         | -0.44 (26.0)    | (0, 0)       | ben        | -0.54 (17.0)    | (4, 3)       |
| 17   | yor        | 0.14 (8.9)      | (3, 3)       | tam         | -0.14 (22.9)    | (0, 0)       | <b>bam</b> | -0.69 (16.1)    | (4, 1)       |
| 18   | tel        | 0.05 (8.7)      | (1, 2)       | <b>pcm</b>  | -0.16 (22.1)    | (3, 14)      | <b>kin</b> | -0.56 (13.2)    | (1, 3)       |
| 19   | fin        | -0.26 (8.6)     | (1, 1)       | bre         | -0.39 (21.4)    | (1, 0)       | ara        | -0.32 (13.1)    | (0, 1)       |
| 20   | rus        | 0.11 (8.5)      | (0, 1)       | <b>kmr</b>  | 0.36 (21.2)     | (11, 6)      | hun        | 0.08 (12.2)     | (2, 0)       |
| 21   | ell        | 0.15 (8.3)      | (1, 0)       | yor         | -0.16 (21.0)    | (1, 1)       | hye        | -0.17 (12.1)    | (0, 1)       |
| 22   | bul        | -0.21 (6.9)     | (1, 1)       | ben         | -0.08 (20.6)    | (3, 2)       | <b>pcm</b> | -0.69 (11.8)    | (6, 25)      |
| 23   | zho        | -0.48 (6.9)     | (5, 10)      | hun         | -0.29 (20.4)    | (3, 1)       | swe        | -0.11 (10.1)    | (0, 0)       |
| 24   | gle        | 0.03 (6.8)      | (1, 1)       | eng         | -0.35 (18.8)    | (2, 1)       | fra        | -0.59 (9.5)     | (0, 5)       |
| 25   | tam        | -0.24 (6.7)     | (1, 3)       | heb         | -0.26 (17.3)    | (0, 1)       | bul        | -0.18 (9.0)     | (0, 1)       |
| 26   | eng        | -0.22 (6.4)     | (0, 0)       | rus         | -0.28 (15.3)    | (0, 0)       | ell        | -0.29 (8.8)     | (1, 1)       |
| 27   | est        | 0.0 (6.1)       | (1, 1)       | gle         | -0.34 (13.7)    | (0, 2)       | tel        | 0.08 (8.6)      | (1, 0)       |
| 28   | ben        | -0.06 (6.0)     | (3, 1)       | jpn         | 0.18 (13.6)     | (11, 0)      | deu        | -0.2 (8.4)      | (0, 0)       |
| 29   | hun        | -0.23 (5.9)     | (0, 2)       | hye         | 0.27 (12.8)     | (2, 0)       | nep        | -0.13 (8.1)     | (1, 1)       |
| 30   | heb        | 0.08 (5.8)      | (0, 1)       | ara         | -0.3 (11.7)     | (0, 0)       | cym        | 0.03 (7.8)      | (2, 1)       |
| 31   | deu        | -0.13 (5.7)     | (1, 4)       | ell         | -0.13 (11.1)    | (1, 0)       | est        | 0.03 (7.8)      | (0, 0)       |
| 32   | bre        | -0.23 (5.7)     | (1, 1)       | est         | -0.35 (9.1)     | (1, 1)       | tam        | 0.0 (7.1)       | (3, 0)       |
| 33   | mya        | 0.33 (5.6)      | (9, 0)       | <b>pms</b>  | 0.08 (8.8)      | (2, 0)       | bre        | -0.23 (5.6)     | (1, 0)       |
| 34   | kor        | -0.38 (5.6)     | (0, 1)       | nep         | 0.12 (8.4)      | (6, 0)       | mya        | -0.17 (5.3)     | (5, 2)       |
| 35   | hye        | -0.03 (5.4)     | (0, 2)       | cym         | 0.22 (6.7)      | (0, 0)       | jpn        | -0.08 (5.3)     | (4, 1)       |
| 36   | ara        | -0.12 (5.1)     | (0, 0)       | hin         | -0.13 (6.4)     | (0, 3)       | eng        | 0.02 (5.1)      | (5, 1)       |
| 37   | jpn        | -0.19 (3.9)     | (14, 2)      | kor         | -0.07 (5.8)     | (1, 0)       | <b>pms</b> | -0.29 (5.0)     | (1, 0)       |
| 38   | cym        | -0.03 (2.9)     | (1, 3)       | mya         | 0.17 (2.7)      | (3, 1)       | kor        | -0.09 (4.4)     | (7, 1)       |

Table 22: Transfer languages are sorted by transfer score variance (mBERT unseen languages are in **bold** font). # Max Transfer and # Min Transfer denote the count of target languages which receive maximum and minimum transfer from this particular transfer language.

## N Transfer Progression Graphs

From Figure 9 to 18, we present the transfer progression graphs for all 38 transfer languages. We observe POS tagging always have comparatively larger deviation which increase with the progression of training steps. In addition, for different time steps in each graph, we provide percentage of positive/negative transfers and the top performing target languages. This way, we observe top target languages that can get continuous improvement for each transfer language even after thousands of steps.

| Type      | Transfer Language | Variance | max(+) $\rightarrow$  | min(-) $\rightarrow$  |
|-----------|-------------------|----------|---|---|
| (+ and -) | ibo               | high     | aii, ajp, apu, arr, ces, gle, gub, koi, krl, yor                | grc, hsb, hye, kfm, otk, san, sme, sqi, srp, urb                          |
| (+ and -) | bam               | high     | bho, bam, bre, bxr, kfm, kmr, kpj, mpu, rus, soj, wbp           | ajp, ara, chu, gla, got, krl, lzh, nld, orv, qpm, qtd, swl, tha, tgl, zho |
| (+)       | mos               | high     | aqz, bel, bul, eng, ind, ita, kaz, lit, myv, pol, tam, tgl, ukr | arr, wbp  |
| (-)       | pcm               | high     | tha   | aii, bho, ell, eng, eus, hrv, isl, lat, lit, nor, qhe                     |
| neutral   | eng               | low      | -   | -   |
| neutral   | ara               | low      | -   | -   |

Table 23: Characteristics of example transfer languages with different intensity of variance derived from dependency parsing task results. max(+) $\rightarrow$  represents set of target language which receive maximum score for the specific transfer language whereas, min(-) $\rightarrow$  represents the complete opposite.

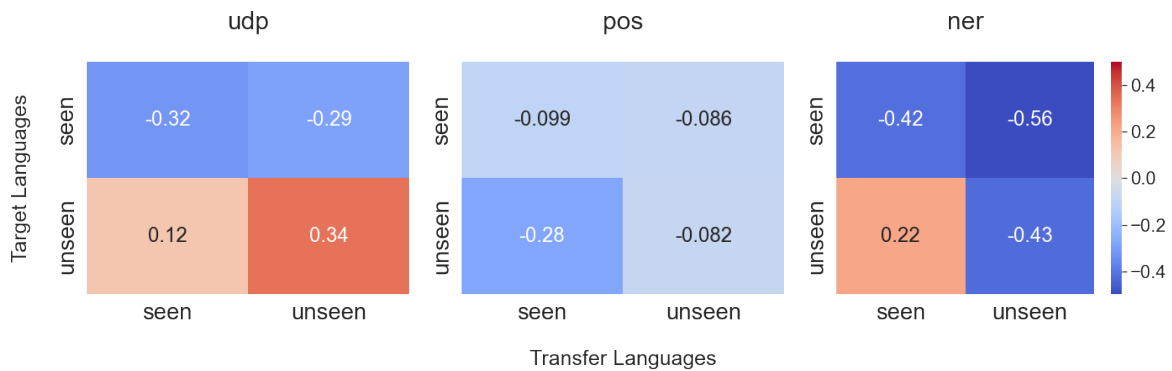


Figure 8: Transfer-Target Heatmap for mbert seen and unseen languages

| Language               | iso-639 | Family        | Genus               | Script      | mBERT-seen? |
|------------------------|---------|---------------|---------------------|-------------|-------------|
| Hebrew                 | heb     | Afro-Asiatic  | Semitic             | Hebr        |             |
| Arabic                 | ara     | Afro-Asiatic  | Semitic             | Arab        |             |
| Hausa                  | hau     | Afro-Asiatic  | West Chadic         | Latn        | ×           |
| Telugu                 | tel     | Dravidian     | Dravidian           | Telu        |             |
| Tamil                  | tam     | Dravidian     | Dravidian           | Taml        |             |
| Armenian               | hye     | Indo-European | Armenian            | Armn        |             |
| Breton                 | bre     | Indo-European | Celtic              | Latn        |             |
| Irish                  | gle     | Indo-European | Celtic              | Latn        |             |
| Welsh                  | cym     | Indo-European | Celtic              | Latn        |             |
| English                | eng     | Indo-European | Germanic            | Latn        |             |
| Swedish                | swe     | Indo-European | Germanic            | Latn        |             |
| German                 | deu     | Indo-European | Germanic            | Latn        |             |
| Modern Greek (1453-)   | ell     | Indo-European | Greek               | Grek        |             |
| Bengali                | ben     | Indo-European | Indic               | Beng        |             |
| Nepali (macrolanguage) | nep     | Indo-European | Indic               | Deva        |             |
| Hindi                  | hin     | Indo-European | Indic               | Deva        |             |
| Northern Kurdish       | kmr     | Indo-European | Iranian             | Arab        | ×           |
| French                 | fra     | Indo-European | Romance             | Latn        |             |
| Spanish                | spa     | Indo-European | Romance             | Latn        |             |
| Piemontese             | pms     | Indo-European | Romance             | Latn        | ×           |
| Bulgarian              | bul     | Indo-European | Slavic              | Cyrl        |             |
| Russian                | rus     | Indo-European | Slavic              | Cyrl        |             |
| Japanese               | jpn     | Japanese      | Japanese            | Jpan        |             |
| Korean                 | kor     | Korean        | Korean              | Kore        |             |
| Bambara                | bam     | Mande         | Western Mande       | Latn        | ×           |
| Kinyarwanda            | kin     | Niger-Congo   | Bantu               | Latn        | ×           |
| Yoruba                 | yor     | Niger-Congo   | Defoid              | Latn        |             |
| Ewe                    | ewe     | Niger-Congo   | Gbe                 | Latn        | ×           |
| Igbo                   | ibo     | Niger-Congo   | Igboid              | Latn        | ×           |
| Mossi                  | mos     | Niger-Congo   | Oti-Volta           | Latn        | ×           |
| Wolof                  | wol     | Niger-Congo   | Wolof               | Latn        | ×           |
| Burmese                | mya     | Sino-Tibetan  | Burmese-Lolo        | Mon-Burmese |             |
| Chinese                | zho     | Sino-Tibetan  | Chinese             | Chinese     |             |
| Guajajára              | gub     | Tupian        | Maweti-Guarani      | Latn        | ×           |
| Estonian               | est     | Uralic        | Finnic              | Latn        |             |
| Finnish                | fin     | Uralic        | Finnic              | Latn        |             |
| Hungarian              | hun     | Uralic        | Ugric               | Latn        |             |
| Nigerian Pidgin        | pcm     | other         | Creoles and Pidgins | Latn        | ×           |

Table 24: Transfer Languages we use in our study for mBERT fine-tuning

| UDP and POS Tagging Task Adapter Training Dataset |         |                 |            |                |                     |           |
|---|---------|-----------------|------------|----------------|---------------------|-----------|
| Language  | iso-639 | UD Identifier   | # Examples | Family         | Genus               | Script    |
| Coptic  | cop     | cop_scriptorium | 403        | Afro-Asiatic   | Egyptian-Coptic     | Coptic    |
| Arabic  | ara     | ar_nyuad        | 1963       | Afro-Asiatic   | Semitic             | Arab      |
| Hebrew  | heb     | he_hhb          | 491        | Afro-Asiatic   | Semitic             | Hebr      |
| Maltese   | mlt     | mt_mudt         | 518        | Afro-Asiatic   | Semitic             | Latn      |
| Kazakh  | kaz     | kk_ktb          | 1047       | Altaic         | Turkic              | Cyrl      |
| Turkish   | tur     | tr_gb           | 2880       | Altaic         | Turkic              | Latn      |
| Uighur  | uig     | ug_udt          | 900        | Altaic         | Turkic              | Uighur    |
| Vietnamese  | vie     | vi_vtb          | 800        | Austro-Asiatic | Vietic              | Latn      |
| Indonesian  | ind     | id_pud          | 1000       | Austronesian   | Malayo-Sumbawan     | Latn      |
| Basque  | eus     | eu_bdt          | 1799       | Basque         | Basque              | Latn      |
| Turkish German                                    | qtd     | qtd_sagt        | 805        | Code switching | Code switching      | Latn      |
| Tamil   | tam     | ta_mwtt         | 534        | Dravidian      | Dravidian           | Taml      |
| Telugu  | tel     | te_mtg          | 146        | Dravidian      | Dravidian           | Telu      |
| Armenian  | hye     | hy_armtdp       | 278        | Indo-European  | Armenian            | Armn      |
| Latvian   | lav     | lv_lvth         | 1823       | Indo-European  | Baltic              | Latn      |
| Lithuanian  | lit     | lt_alksnis      | 684        | Indo-European  | Baltic              | Latn      |
| Welsh   | cym     | cy_ccg          | 953        | Indo-European  | Celtic              | Latn      |
| Scottish Gaelic                                   | gla     | gd_arcosg       | 538        | Indo-European  | Celtic              | Latn      |
| Irish   | gle     | ga_idt          | 454        | Indo-European  | Celtic              | Latn      |
| Gothic  | got     | got_proiel      | 1029       | Indo-European  | Germanic            | Gothic    |
| Afrikaans   | afr     | af_afribooms    | 425        | Indo-European  | Germanic            | Latn      |
| Danish  | dan     | da_ddt          | 565        | Indo-European  | Germanic            | Latn      |
| German  | deu     | de_hdt          | 18459      | Indo-European  | Germanic            | Latn      |
| English   | eng     | en_ewt          | 2077       | Indo-European  | Germanic            | Latn      |
| Faroese   | fao     | fo_of           | 1208       | Indo-European  | Germanic            | Latn      |
| Icelandic   | isl     | is_icepahc      | 5157       | Indo-European  | Germanic            | Latn      |
| Dutch   | nld     | nl_lassysmall   | 875        | Indo-European  | Germanic            | Latn      |
| Norwegian   | nor     | no_bokmaal      | 1939       | Indo-European  | Germanic            | Latn      |
| Swedish   | swe     | sv_talbanken    | 1219       | Indo-European  | Germanic            | Latn      |
| Modern Greek (1453-)                              | ell     | eL_gdt          | 456        | Indo-European  | Greek               | Grek      |
| Ancient Greek (to 1453)                           | grc     | grc_perseus     | 1306       | Indo-European  | Greek               | Grek      |
| Urdu  | urd     | ur_udtb         | 535        | Indo-European  | Indic               | Arab      |
| Sanskrit  | san     | sa_vedic        | 1473       | Indo-European  | Indic               | Brahmi    |
| Hindi   | hin     | hi_hdtb         | 1684       | Indo-European  | Indic               | Deva      |
| Marathi   | mar     | mr_ufal         | 47         | Indo-European  | Indic               | Deva      |
| Persian   | fas     | fa_perdt        | 1455       | Indo-European  | Iranian             | Arab      |
| Northern Kurdish                                  | kmr     | kmr_mg          | 734        | Indo-European  | Iranian             | Arab      |
| Latin   | lat     | la_ittb         | 2101       | Indo-European  | Italic              | Latn      |
| Catalan   | cat     | ca_ancora       | 1846       | Indo-European  | Romance             | Latn      |
| French  | fra     | fr_ftb          | 2541       | Indo-European  | Romance             | Latn      |
| Old French (842-ca. 1400)                         | fro     | fro_srcmf       | 1927       | Indo-European  | Romance             | Latn      |
| Galician  | glg     | gl_ctg          | 861        | Indo-European  | Romance             | Latn      |
| Italian   | ita     | it_vit          | 1067       | Indo-European  | Romance             | Latn      |
| Portuguese  | por     | pt_gsd          | 1204       | Indo-European  | Romance             | Latn      |
| Romanian  | ron     | ro_nonstandard  | 1052       | Indo-European  | Romance             | Latn      |
| Spanish   | spa     | es_ancora       | 1721       | Indo-European  | Romance             | Latn      |
| Belarusian  | bel     | be_hse          | 889        | Indo-European  | Slavic              | Cyrl      |
| Bulgarian   | bul     | bg_btb          | 1116       | Indo-European  | Slavic              | Cyrl      |
| Old Russian                                       | orv     | orv_torot       | 1756       | Indo-European  | Slavic              | Cyrl      |
| Russian   | rus     | ru_syntagrus    | 6491       | Indo-European  | Slavic              | Cyrl      |
| Serbian   | srp     | sr_set          | 520        | Indo-European  | Slavic              | Cyrl      |
| Ukrainian   | ukr     | uk_iu           | 892        | Indo-European  | Slavic              | Cyrl      |
| Church Slavic                                     | chu     | cu_proiel       | 1141       | Indo-European  | Slavic              | Glag+Latn |
| Czech   | ces     | cs_pdt          | 10148      | Indo-European  | Slavic              | Latn      |
| Croatian  | hrv     | hr_set          | 1136       | Indo-European  | Slavic              | Latn      |
| Upper Sorbian                                     | hsb     | hsb_ufal        | 623        | Indo-European  | Slavic              | Latn      |
| Polish  | pol     | pl_pdb          | 2215       | Indo-European  | Slavic              | Latn      |
| Pomak   | qpm     | qpm_philotis    | 635        | Indo-European  | Slavic              | Latn      |
| Slovak  | slk     | sk_snk          | 1061       | Indo-European  | Slavic              | Latn      |
| Slovenian   | slv     | sl_sst          | 1110       | Indo-European  | Slavic              | Latn      |
| Japanese  | jpn     | ja_bccwj        | 7871       | Japanese       | Japanese            | Jpan      |
| Korean  | kor     | ko_kaist        | 2287       | Korean         | Korean              | Kore      |
| Russia Buriat                                     | bxr     | bxr_bdt         | 908        | Mongolic       | Altic               | Cyrl      |
| Wolof   | wol     | wo_wtb          | 470        | Niger-Congo    | Wolof               | Latn      |
| Cusco Quechua                                     | qhe     | qhe_hienccs     | 225        | Quechuan       | Quechuan            | Latn      |
| Literary Chinese                                  | lzh     | lzh_kyoto       | 4469       | Sino-Tibetan   | Chinese             | Chinese   |
| Chinese   | zho     | zh_hk           | 1004       | Sino-Tibetan   | Chinese             | Chinese   |
| Estonian  | est     | et_edt          | 3214       | Uralic         | Finnic              | Latn      |
| Finnish   | fin     | fi_ood          | 2122       | Uralic         | Finnic              | Latn      |
| Livvi   | olo     | olo_kkpp        | 106        | Uralic         | Finnic              | Latn      |
| Northern Sami                                     | sme     | sme_giella      | 865        | Uralic         | Saami               | Latn      |
| Hungarian   | hun     | hu_szeged       | 449        | Uralic         | Ugric               | Latn      |
| Nigerian Pidgin                                   | pcm     | pcm_nsc         | 972        | other          | Creoles and Pidgins | Latn      |
| Swedish Sign Language                             | swl     | swl_sslc        | 34         | other          | Sign Languages      | Latn      |

Table 25: Task Adapter training dataset details (taken from Universal Dependency v2.11 (de Marneffe et al., 2021)) for dependency parsing and pos tagging.



NER Task Adapter Training Dataset

| Language               | iso-639 | Family         | Genus                      | Script         |
|------------------------|---------|----------------|----------------------------|----------------|
| Somali                 | som     | Afro-Asiatic   | Lowland East Cushitic      | Latn           |
| Arabic                 | ara     | Afro-Asiatic   | Semitic                    | Arab           |
| Amharic                | amh     | Afro-Asiatic   | Semitic                    | Ethi           |
| Hebrew                 | heb     | Afro-Asiatic   | Semitic                    | Hebr           |
| Maltese                | mlt     | Afro-Asiatic   | Semitic                    | Latn           |
| Mongolian              | mon     | Altaic         | Mongolic                   | Mongolian      |
| Bashkir                | bak     | Altaic         | Turkic                     | Cyrl           |
| Chuvash                | chv     | Altaic         | Turkic                     | Cyrl           |
| Kazakh                 | kaz     | Altaic         | Turkic                     | Cyrl           |
| Yakut                  | sah     | Altaic         | Turkic                     | Cyrl           |
| Crimean Tatar          | crh     | Altaic         | Turkic                     | Cyrl+Latn+Arab |
| Kirghiz                | kir     | Altaic         | Turkic                     | Kyrgyz+Cyrl    |
| Azerbaijani            | aze     | Altaic         | Turkic                     | Latn           |
| Tatar                  | tat     | Altaic         | Turkic                     | Latn           |
| Turkmen                | tuk     | Altaic         | Turkic                     | Latn           |
| Turkish                | tur     | Altaic         | Turkic                     | Latn           |
| Uzbek                  | uzb     | Altaic         | Turkic                     | Latn           |
| Uighur                 | uig     | Altaic         | Turkic                     | Uighur         |
| Khmer                  | khm     | Austro-Asiatic | Khmer                      | Khmer          |
| Vietnamese             | vie     | Austro-Asiatic | Vietic                     | Latn           |
| Malagasy               | mlg     | Austronesian   | Barito                     | Latn           |
| Cebuano                | ceb     | Austronesian   | Greater Central Philippine | Latn           |
| Tagalog                | tgl     | Austronesian   | Greater Central Philippine | Latn           |
| Waray (Philippines)    | war     | Austronesian   | Greater Central Philippine | Latn           |
| Javanese               | jav     | Austronesian   | Javanese                   | Latn+Javanese  |
| Achinese               | ace     | Austronesian   | Malayo-Sumbawan            | Latn           |
| Malay (macrolanguage)  | msa     | Austronesian   | Malayo-Sumbawan            | Latn           |
| Sundanese              | sun     | Austronesian   | Malayo-Sumbawan            | Latn           |
| Iloko                  | ilo     | Austronesian   | Northern Luzon             | Latn           |
| Maori                  | mri     | Austronesian   | Oceanic                    | Latn           |
| Aymara                 | aym     | Aymaran        | Aymaran                    | Latn           |
| Basque                 | eus     | Basque         | Basque                     | Latn           |
| Kannada                | kan     | Dravidian      | Dravidian                  | Kannada        |
| Malayalam              | mal     | Dravidian      | Dravidian                  | Malayalam      |
| Tamil                  | tam     | Dravidian      | Dravidian                  | TamI           |
| Telugu                 | tel     | Dravidian      | Dravidian                  | TelU           |
| Albanian               | sqi     | Indo-European  | Albanian                   | Latn           |
| Armenian               | hye     | Indo-European  | Armenian                   | Armn           |
| Latvian                | lav     | Indo-European  | Baltic                     | Latn           |
| Lithuanian             | lit     | Indo-European  | Baltic                     | Latn           |
| Breton                 | bre     | Indo-European  | Celtic                     | Latn           |
| Welsh                  | cym     | Indo-European  | Celtic                     | Latn           |
| Scottish Gaelic        | gla     | Indo-European  | Celtic                     | Latn           |
| Irish                  | gle     | Indo-European  | Celtic                     | Latn           |
| Western Frisian        | fry     | Indo-European  | Germanic                   | West Frisian   |
| Afrikaans              | afz     | Indo-European  | Germanic                   | Latn           |
| Bavarian               | bar     | Indo-European  | Germanic                   | Latn           |
| Danish                 | dan     | Indo-European  | Germanic                   | Latn           |
| German                 | deu     | Indo-European  | Germanic                   | Latn           |
| English                | eng     | Indo-European  | Germanic                   | Latn           |
| Faroese                | fao     | Indo-European  | Germanic                   | Latn           |
| Northern Frisian       | frr     | Indo-European  | Germanic                   | Latn           |
| Icelandic              | isl     | Indo-European  | Germanic                   | Latn           |
| Kölsch                 | ksh     | Indo-European  | Germanic                   | Latn           |
| Luxembourgish          | ltz     | Indo-European  | Germanic                   | Latn           |
| Low German             | nds     | Indo-European  | Germanic                   | Latn           |
| Dutch                  | nld     | Indo-European  | Germanic                   | Latn           |
| Norwegian              | nor     | Indo-European  | Germanic                   | Latn           |
| Swedish                | swe     | Indo-European  | Germanic                   | Latn           |
| Yiddish                | yid     | Indo-European  | Germanic                   | Latn           |
| Zeeuws                 | zea     | Indo-European  | Germanic                   | Latn           |
| Modern Greek (1453-)   | ell     | Indo-European  | Greek                      | Grek           |
| Sindhi                 | snd     | Indo-European  | Indic                      | Arab           |
| Urdu                   | urd     | Indo-European  | Indic                      | Arab           |
| Assamese               | asm     | Indo-European  | Indic                      | Assamese       |
| Bengali                | ben     | Indo-European  | Indic                      | Beng           |
| Hindi                  | hin     | Indo-European  | Indic                      | Deva           |
| Marathi                | mar     | Indo-European  | Indic                      | Deva           |
| Nepali (macrolanguage) | nep     | Indo-European  | Indic                      | Deva           |
| Gujarati               | guj     | Indo-European  | Indic                      | Gujarati       |
| Oriya (macrolanguage)  | ori     | Indo-European  | Indic                      | Odia           |
| Punjabi                | pan     | Indo-European  | Indic                      | Shahmukh       |
| Sinhala                | sin     | Indo-European  | Indic                      | Sinhala        |
| Dhivehi                | div     | Indo-European  | Indic                      | Thaana         |
| Persian                | fas     | Indo-European  | Iranian                    | Arab           |
| Ossetian               | oss     | Indo-European  | Iranian                    | Cyrl           |
| Tajik                  | tgk     | Indo-European  | Iranian                    | Cyrl+Latn      |
| Kurdish                | kur     | Indo-European  | Iranian                    | Latn+Sorani    |
| Mazanderani            | mzn     | Indo-European  | Iranian                    | Persian        |
| Pushto                 | pus     | Indo-European  | Iranian                    | Pushto         |
| Asturian               | ast     | Indo-European  | Romance                    | Latn           |
| Catalan                | cat     | Indo-European  | Romance                    | Latn           |
| French                 | fra     | Indo-European  | Romance                    | Latn           |
| Galician               | glg     | Indo-European  | Romance                    | Latn           |
| Italian                | ita     | Indo-European  | Romance                    | Latn           |
| Ligurian               | lij     | Indo-European  | Romance                    | Latn           |
| Neapolitan             | nap     | Indo-European  | Romance                    | Latn           |
| Occitan (post 1500)    | oci     | Indo-European  | Romance                    | Latn           |
| Piemontese             | pms     | Indo-European  | Romance                    | Latn           |
| Portuguese             | por     | Indo-European  | Romance                    | Latn           |
| Romansh                | roh     | Indo-European  | Romance                    | Latn           |
| Romanian               | ron     | Indo-European  | Romance                    | Latn           |
| Spanish                | spa     | Indo-European  | Romance                    | Latn           |
| Belarusian             | bel     | Indo-European  | Slavic                     | Cyrl           |
| Bulgarian              | bul     | Indo-European  | Slavic                     | Cyrl           |

|                         |     |                   |                |             |
|-------------------------|-----|-------------------|----------------|-------------|
| Macedonian              | mkd | Indo-European     | Slavic         | Cyrl        |
| Russian                 | rus | Indo-European     | Slavic         | Cyrl        |
| Ukrainian               | ukr | Indo-European     | Slavic         | Cyrl        |
| Czech                   | ces | Indo-European     | Slavic         | Latn        |
| Kashubian               | csb | Indo-European     | Slavic         | Latn        |
| Serbo-Croatian          | hbs | Indo-European     | Slavic         | Latn        |
| Upper Sorbian           | hsb | Indo-European     | Slavic         | Latn        |
| Polish                  | pol | Indo-European     | Slavic         | Latn        |
| Slovak                  | slk | Indo-European     | Slavic         | Latn        |
| Slovenian               | slv | Indo-European     | Slavic         | Latn        |
| Japanese                | jpn | Japanese          | Japanese       | Jpan        |
| Georgian                | kat | Kartvelian        | Kartvelian     | Georgian    |
| Mingrelian              | xmf | Kartvelian        | Kartvelian     | Latn        |
| Korean                  | kor | Korean            | Korean         | Kore        |
| Chechen                 | che | Nakh-Daghestanian | Nakh           | Cyrl        |
| Kinyarwanda             | kin | Niger-Congo       | Bantu          | Latn        |
| Lingala                 | lin | Niger-Congo       | Bantu          | Latn        |
| Swahili (macrolanguage) | swa | Niger-Congo       | Bantu          | Latn        |
| Yoruba                  | yor | Niger-Congo       | Defoid         | Latn        |
| Igbo                    | ibo | Niger-Congo       | Igboid         | Latn        |
| Quechua                 | que | Quechuan          | Quechuan       | Latn        |
| Tibetan                 | bod | Sino-Tibetan      | Bodic          | Tibetan     |
| Burmese                 | mya | Sino-Tibetan      | Burmese-Lolo   | Mon-Burmese |
| Chinese                 | zho | Sino-Tibetan      | Chinese        | Chinese     |
| Thai                    | tha | Tai-Kadai         | Kam-Tai        | Thai        |
| Guarani                 | grn | Tupian            | Maweti-Guarani | Latn        |
| Estonian                | est | Uralic            | Finnic         | Latn        |
| Finnish                 | fin | Uralic            | Finnic         | Latn        |
| Veps                    | vep | Uralic            | Finnic         | Latn        |
| Hungarian               | hun | Uralic            | Ugric          | Latn        |

Table 26: Task Adapter training dataset details (taken from Wikiann (Pan et al., 2017)) for Named Entity Recognition.

| NLI Task Adapter Training Dataset |         |                |          |         |
|-----------------------------------|---------|----------------|----------|---------|
| Language                          | iso-639 | Family         | Genus    | Script  |
| Arabic                            | ara     | Afro-Asiatic   | Semitic  | Arab    |
| Turkish                           | tur     | Altaic         | Turkic   | Latn    |
| Vietnamese                        | vie     | Austro-Asiatic | Vietic   | Latn    |
| German                            | deu     | Indo-European  | Germanic | Latn    |
| English                           | eng     | Indo-European  | Germanic | Latn    |
| Modern Greek (1453-)              | ell     | Indo-European  | Greek    | Grek    |
| Urdu                              | urd     | Indo-European  | Indic    | Arab    |
| Hindi                             | hin     | Indo-European  | Indic    | Deva    |
| French                            | fre     | Indo-European  | Romance  | Latn    |
| Spanish                           | spa     | Indo-European  | Romance  | Latn    |
| Bulgarian                         | bul     | Indo-European  | Slavic   | Cyrl    |
| Russian                           | rus     | Indo-European  | Slavic   | Cyrl    |
| Swahili (macrolanguage)           | swa     | Niger-Congo    | Bantu    | Latn    |
| Chinese                           | zho     | Sino-Tibetan   | Chinese  | Chinese |
| Thai                              | tha     | Tai-Kadai      | Kam-Tai  | Thai    |

Table 27: Task Adapter training dataset details (taken from XNLI (Conneau et al., 2018)) for Natural Language Inference.

| Extractive Question Answering Task Adapter Training Dataset |         |               |           |        |
|---|---------|---------------|-----------|--------|
| Language  | iso-639 | Family        | Genus     | Script |
| Arabic  | ara     | Afro-Asiatic  | Semitic   | Arab   |
| Indonesian  | idn     | Astronesian   | Malay     | Latn   |
| Telugu  | tel     | Dravidian     | Dravidian | Telu   |
| English   | eng     | Indo-European | Germanic  | Latn   |
| Bengali   | ben     | Indo-European | Indic     | Beng   |
| Russian   | rus     | Indo-European | Slavic    | Cyrl   |
| Korean  | kor     | Korean        | Korean    | Kore   |
| Swahili (macrolanguage)                                     | swa     | Niger-Congo   | Bantu     | Latn   |
| Finnish   | fin     | Uralic        | Finnic    | Latn   |

Table 28: Task Adapter training dataset details (taken from TyDiQA (Clark et al., 2020)) for Extractive Question Answering.

| Evaluation Languages |                           |         |     |             |      |      |        |
|----------------------|---------------------------|---------|-----|-------------|------|------|--------|
|                      | Language                  | iso-639 | NER | UDP and POS | XNLI | ANLI | TyDiQA |
| 1                    | Achinese                  | ace     |     | X           | X    | X    | X      |
| 2                    | Afrikaans                 | afr     |     |             | X    | X    | X      |
| 3                    | Assyrian Neo-Aramaic      | aii     | X   |             | X    | X    | X      |
| 4                    | South Levantine Arabic    | ajp     | X   |             | X    | X    | X      |
| 5                    | Akkadian                  | akk     | X   |             | X    | X    | X      |
| 6                    | Amharic                   | amh     |     |             | X    | X    | X      |
| 7                    | Apurinã                   | apu     | X   |             | X    | X    | X      |
| 8                    | Akuntsu                   | aqz     | X   |             | X    | X    | X      |
| 9                    | Arabic                    | ara     |     |             | X    | X    |        |
| 10                   | Karo (Brazil)             | arr     | X   |             |      | X    | X      |
| 11                   | Assamese                  | asm     |     | X           | X    | X    | X      |
| 12                   | Asturian                  | ast     |     | X           | X    | X    | X      |
| 13                   | Aymara                    | aym     |     | X           | X    |      | X      |
| 14                   | Azerbaijani               | aze     |     | X           | X    | X    | X      |
| 15                   | Bashkir                   | bak     |     | X           | X    | X    | X      |
| 16                   | Bambara                   | bam     | X   |             | X    | X    | X      |
| 17                   | Bavarian                  | bar     |     | X           | X    | X    | X      |
| 18                   | Belarusian                | bel     |     |             | X    | X    | X      |
| 19                   | Bengali                   | ben     |     |             | X    | X    |        |
| 20                   | Bhojpuri                  | bho     | X   |             | X    | X    | X      |
| 21                   | Tibetan                   | bod     |     | X           | X    | X    | X      |
| 22                   | Breton                    | bre     |     |             | X    | X    | X      |
| 23                   | Bulgarian                 | bul     |     |             | X    | X    | X      |
| 24                   | Russia Buriat             | bxr     | X   |             |      | X    | X      |
| 25                   | Catalan                   | cat     |     |             | X    | X    | X      |
| 26                   | Cebuano                   | ceb     |     | X           | X    | X    | X      |
| 27                   | Czech                     | ces     |     |             | X    | X    | X      |
| 28                   | Chechen                   | che     |     | X           | X    | X    | X      |
| 29                   | Church Slavic             | chu     | X   |             | X    | X    | X      |
| 30                   | Chuvash                   | chv     |     | X           | X    | X    | X      |
| 31                   | Chukot                    | ckt     | X   |             | X    | X    | X      |
| 32                   | Coptic                    | cop     | X   |             | X    | X    | X      |
| 33                   | Crimean Tatar             | crh     |     | X           | X    | X    | X      |
| 34                   | Kashubian                 | csb     |     | X           | X    | X    | X      |
| 35                   | Welsh                     | cym     |     |             | X    | X    | X      |
| 36                   | Danish                    | dan     |     |             | X    | X    | X      |
| 37                   | German                    | deu     |     |             |      | X    | X      |
| 38                   | Dhivehi                   | div     |     | X           | X    | X    | X      |
| 39                   | Modern Greek (1453-)      | ell     |     |             |      | X    | X      |
| 40                   | English                   | eng     |     |             |      | X    |        |
| 41                   | Estonian                  | est     |     |             | X    | X    | X      |
| 42                   | Basque                    | eus     |     |             | X    | X    | X      |
| 43                   | Faroese                   | fao     |     |             | X    | X    | X      |
| 44                   | Persian                   | fas     |     |             | X    | X    | X      |
| 45                   | Finnish                   | fin     |     |             | X    | X    |        |
| 46                   | French                    | fra     |     |             |      | X    | X      |
| 47                   | Old French (842-ca. 1400) | fro     | X   |             | X    | X    | X      |
| 48                   | Northern Frisian          | frf     |     | X           | X    | X    | X      |
| 49                   | Western Frisian           | fry     |     | X           | X    | X    | X      |
| 50                   | Scottish Gaelic           | gla     |     |             | X    | X    | X      |
| 51                   | Irish                     | gle     |     |             | X    | X    | X      |
| 52                   | Galician                  | glg     |     |             | X    | X    | X      |
| 53                   | Manx                      | glv     | X   |             | X    | X    | X      |
| 54                   | Gothic                    | got     | X   |             | X    | X    | X      |
| 55                   | Ancient Greek (to 1453)   | grc     | X   |             | X    | X    | X      |
| 56                   | Guarani                   | grn     |     | X           | X    | X    | X      |
| 57                   | Swiss German              | gsw     | X   |             | X    | X    | X      |
| 58                   | Guajará                   | gub     | X   |             | X    | X    | X      |
| 59                   | Gujarati                  | guj     |     | X           | X    | X    | X      |
| 60                   | Mbyá Guaraní              | gun     | X   |             | X    | X    | X      |
| 61                   | Serbo-Croatian            | hbs     |     | X           | X    | X    | X      |
| 62                   | Hebrew                    | heb     |     |             | X    | X    | X      |
| 63                   | Hindi                     | hin     |     |             |      | X    | X      |
| 64                   | Croatian                  | hrv     | X   |             | X    | X    | X      |

|     |                        |     |   |   |   |   |   |
|-----|------------------------|-----|---|---|---|---|---|
| 65  | Upper Sorbian          | hsb |   |   | X | X | X |
| 66  | Hungarian              | hun |   |   | X | X | X |
| 67  | Armenian               | hye |   |   | X | X | X |
| 68  | Igbo                   | ibo |   | X | X | X | X |
| 69  | Iloko                  | ilo |   | X | X | X | X |
| 70  | Indonesian             | ind | X |   | X | X |   |
| 71  | Icelandic              | isl |   |   | X | X | X |
| 72  | Italian                | ita |   |   | X | X | X |
| 73  | Javanese               | jav |   | X | X | X | X |
| 74  | Japanese               | jpn |   |   | X | X | X |
| 75  | Kannada                | kan |   | X | X | X | X |
| 76  | Georgian               | kat |   | X | X | X | X |
| 77  | Kazakh                 | kaz |   |   | X | X | X |
| 78  | Khunsari               | kfm | X |   | X | X | X |
| 79  | Khmer                  | khm |   | X | X | X | X |
| 80  | Kinyarwanda            | kin |   | X | X | X | X |
| 81  | Kirghiz                | kir |   | X | X | X | X |
| 82  | Northern Kurdish       | kmr | X |   | X | X | X |
| 83  | Komi-Permyak           | koi | X |   | X | X | X |
| 84  | Korean                 | kor |   |   | X | X |   |
| 85  | Komi-Zyrian            | kpv | X |   | X | X | X |
| 86  | Karelian               | krl | X |   | X | X | X |
| 87  | Kölsch                 | ksh |   | X | X | X | X |
| 88  | Kurdish                | kur |   | X | X | X | X |
| 89  | Latin                  | lat | X |   | X | X | X |
| 90  | Latvian                | lav |   |   | X | X | X |
| 91  | Ligurian               | lij |   |   | X | X | X |
| 92  | Lingala                | lin |   | X | X | X | X |
| 93  | Lithuanian             | lit |   |   | X | X | X |
| 94  | Luxembourgish          | ltz |   | X | X | X | X |
| 95  | Literary Chinese       | lzh | X |   | X | X | X |
| 96  | Malayalam              | mal |   | X | X | X | X |
| 97  | Marathi                | mar |   |   | X | X | X |
| 98  | Moksha                 | mdf | X |   | X | X | X |
| 99  | Macedonian             | mkd |   | X | X | X | X |
| 100 | Malagasy               | mlg |   | X | X | X | X |
| 101 | Maltese                | mlt |   |   | X | X | X |
| 102 | Mongolian              | mon |   | X | X | X | X |
| 103 | Makuráp                | mpu | X |   | X | X | X |
| 104 | Maori                  | mri |   | X | X | X | X |
| 105 | Malay (macrolanguage)  | msa |   | X | X | X | X |
| 106 | Burmese                | mya |   | X | X | X | X |
| 107 | Mundurukú              | myu | X |   | X | X | X |
| 108 | Erzya                  | myv | X |   | X | X | X |
| 109 | Mazanderani            | mzn |   | X | X | X | X |
| 110 | Neapolitan             | nap |   |   | X | X | X |
| 111 | Low German             | nds |   |   | X | X | X |
| 112 | Nepali (macrolanguage) | nep |   | X | X | X | X |
| 113 | Dutch                  | nld |   |   | X | X | X |
| 114 | Norwegian              | nor |   |   | X | X | X |
| 115 | Nayini                 | nyq | X |   | X | X | X |
| 116 | Occitan (post 1500)    | oci |   | X | X | X | X |
| 117 | Livvi                  | olo | X |   | X | X | X |
| 118 | Oriya (macrolanguage)  | ori |   | X | X | X | X |
| 119 | Old Russian            | orv | X |   | X | X | X |
| 120 | Ossetian               | oss |   | X | X | X | X |
| 121 | Old Turkish            | otk | X |   | X | X | X |
| 122 | Panjabi                | pan |   | X | X | X | X |
| 123 | Nigerian Pidgin        | pcm | X |   | X | X | X |
| 124 | Piemontese             | pms |   | X | X | X | X |
| 125 | Polish                 | pol |   |   | X | X | X |
| 126 | Portuguese             | por |   |   | X | X | X |
| 127 | Pushto                 | pus |   | X | X | X | X |
| 128 | Cusco Quechua          | qhe | X |   | X | X | X |
| 129 | Pomak                  | qpm | X |   | X | X | X |
| 130 | Turkish German         | qtd | X |   | X | X | X |
| 131 | Quechua                | que |   | X | X |   | X |
| 132 | Romansh                | roh |   | X | X | X | X |

|     |                         |     |   |   |  |   |   |   |
|-----|-------------------------|-----|---|---|--|---|---|---|
| 133 | Romanian                | ron |   |   |  | × | × | × |
| 134 | Russian                 | rus |   |   |  |   | × |   |
| 135 | Yakut                   | sah |   | × |  | × | × | × |
| 136 | Sanskrit                | san | × |   |  | × | × | × |
| 137 | Sinhala                 | sin |   | × |  | × | × | × |
| 138 | Slovak                  | slk |   |   |  | × | × | × |
| 139 | Slovenian               | slv |   |   |  | × | × | × |
| 140 | Northern Sami           | sme | × |   |  | × | × | × |
| 141 | Skolt Sami              | sms | × |   |  | × | × | × |
| 142 | Sindhi                  | snd |   | × |  | × | × | × |
| 143 | Soi                     | soj | × |   |  | × | × | × |
| 144 | Somali                  | som |   | × |  | × | × | × |
| 145 | Spanish                 | spa |   |   |  |   | × | × |
| 146 | Albanian                | sqi |   |   |  | × | × | × |
| 147 | Serbian                 | srp | × |   |  | × | × | × |
| 148 | Sundanese               | sun |   | × |  | × | × | × |
| 149 | Swahili (macrolanguage) | swa |   | × |  |   | × |   |
| 150 | Swedish                 | swe |   |   |  | × | × | × |
| 151 | Swedish Sign Language   | swl | × |   |  | × | × | × |
| 152 | Tamil                   | tam |   |   |  | × | × | × |
| 153 | Tatar                   | tat |   | × |  | × | × | × |
| 154 | Telugu                  | tel |   |   |  | × | × |   |
| 155 | Tajik                   | tgk |   | × |  | × | × | × |
| 156 | Tagalog                 | tgl |   |   |  | × | × | × |
| 157 | Thai                    | tha |   |   |  |   | × | × |
| 158 | Tupinambá               | tpn | × |   |  | × | × | × |
| 159 | Turkmen                 | tuk |   | × |  | × | × | × |
| 160 | Turkish                 | tur |   |   |  |   | × | × |
| 161 | Uighur                  | uig |   |   |  | × | × | × |
| 162 | Ukrainian               | ukr |   |   |  | × | × | × |
| 163 | Urubú-Kaapor            | urb | × |   |  | × | × | × |
| 164 | Urdu                    | urd |   |   |  |   | × | × |
| 165 | Uzbek                   | uzb |   | × |  | × | × | × |
| 166 | Veps                    | vep |   | × |  | × | × | × |
| 167 | Vietnamese              | vie |   |   |  |   | × | × |
| 168 | Waray (Philippines)     | war |   | × |  | × | × | × |
| 169 | Warlpiri                | wbp | × |   |  | × | × | × |
| 170 | Wolof                   | wol | × |   |  | × | × | × |
| 171 | Mingrelian              | xmf |   | × |  | × | × | × |
| 172 | Kangri                  | xnr | × |   |  | × | × | × |
| 173 | Yiddish                 | yid |   | × |  | × | × | × |
| 174 | Yoruba                  | yor |   |   |  | × | × | × |
| 175 | Cantonese               | yue | × |   |  | × | × | × |
| 176 | Zeeuws                  | zea |   | × |  | × | × | × |
| 177 | Chinese                 | zho |   |   |  |   | × | × |
| 178 | Bribri                  | bzc | × | × |  | × |   | × |
| 179 | Asháninka               | cni | × | × |  | × |   | × |
| 180 | Huichol                 | hch | × | × |  | × |   | × |
| 181 | Nahuatl                 | nah | × | × |  | × |   | × |
| 182 | Otomí                   | oto | × | × |  | × |   | × |
| 183 | Shipibo-Konibo          | shp | × | × |  | × |   | × |
| 184 | Rarámuri                | tar | × | × |  | × |   | × |

Table 29: Evaluation languages for all six tasks.

## O FAQ

1. What are the main contributions of this study and the difference of our approach with other methods?
  - First, note that our paper introduces a method for studying cross-lingual transfer, not necessarily a method for improving cross-lingual transfer. We deviate from this “standard” way of using adapters for two reasons:
    - (a) Training a task adapter on many languages, as a preliminary step, allows this component to learning the task, regardless of language. This is necessary for disentangling the effect of task and language in our analysis.
    - (b) We then finetune the whole model (and not introduce a new adapter) exactly because we now want to study the effect of the language. While introducing a new language adapter might have a similar effect, there’s additional hurdles to do so: the language adapter would need more data to be trained, as it would be randomly initialized; our approach instead can work even with a single batch/update, so it is applicable even for very, very low-resource scenarios.
  - Secondly, we propose a strategy to visually represent the language-language interaction utilizing the adapter-based fusion method. In general, training fully bilingual or trilingual for a different combination of languages are very expensive. This is why, we opt to have trained language adapter modules and then fuse together according to the need in an efficient manner.
2. What is the reason for selecting the 38 transfer languages, including the 11 unseen languages? Why include the 11 unseen languages from pre-training?
  - **Language selection:** No other particular reasons except selecting a broader range of transfer languages covering language families and typological diversity. These 38 languages in total cover 10 language families, 26 genus and 14 script variations.

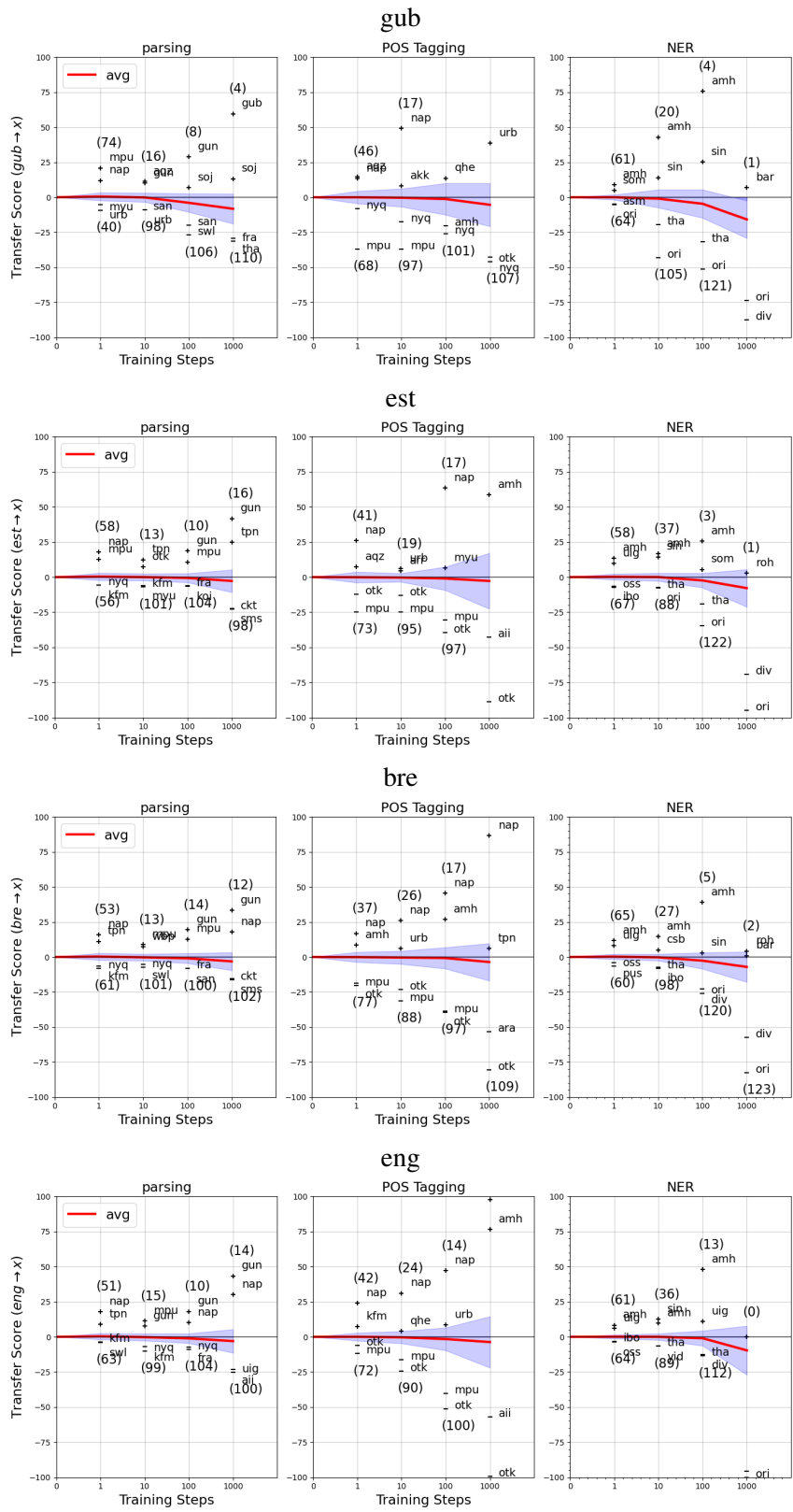


Figure 9: Aggregated Transfer Progression through training steps

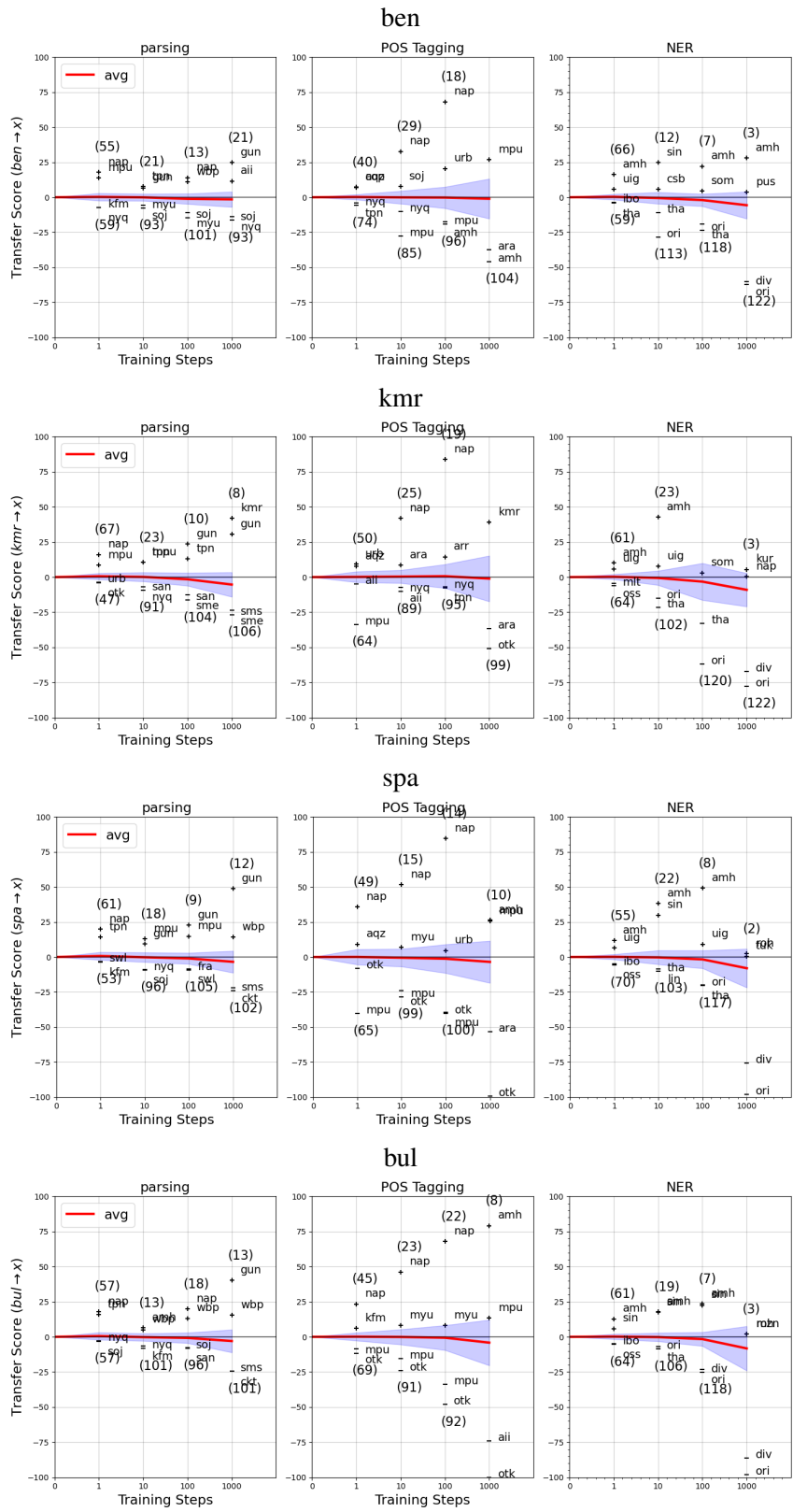


Figure 10: Aggregated Transfer Progression through training steps



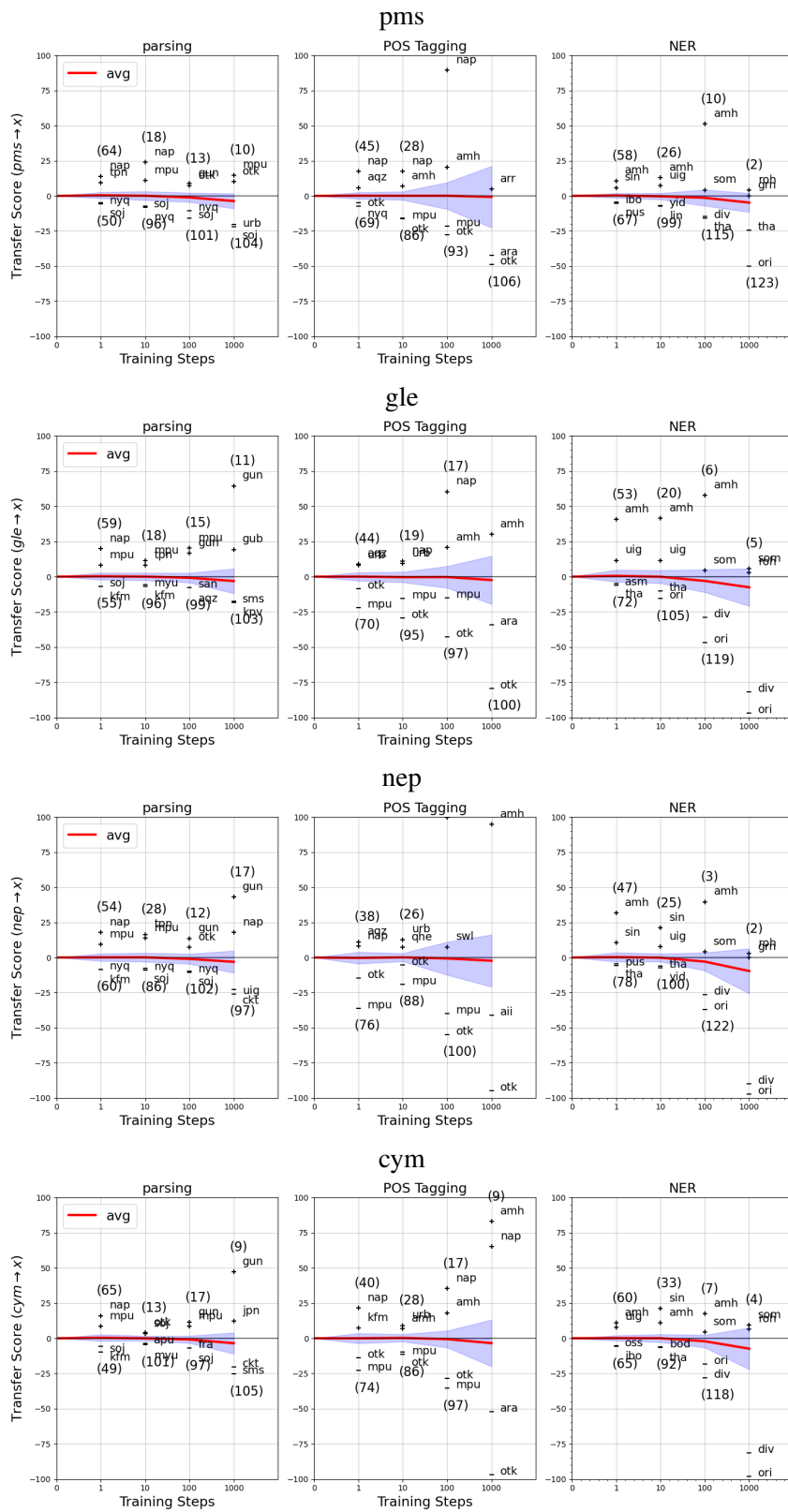


Figure 11: Aggregated Transfer Progression through training steps

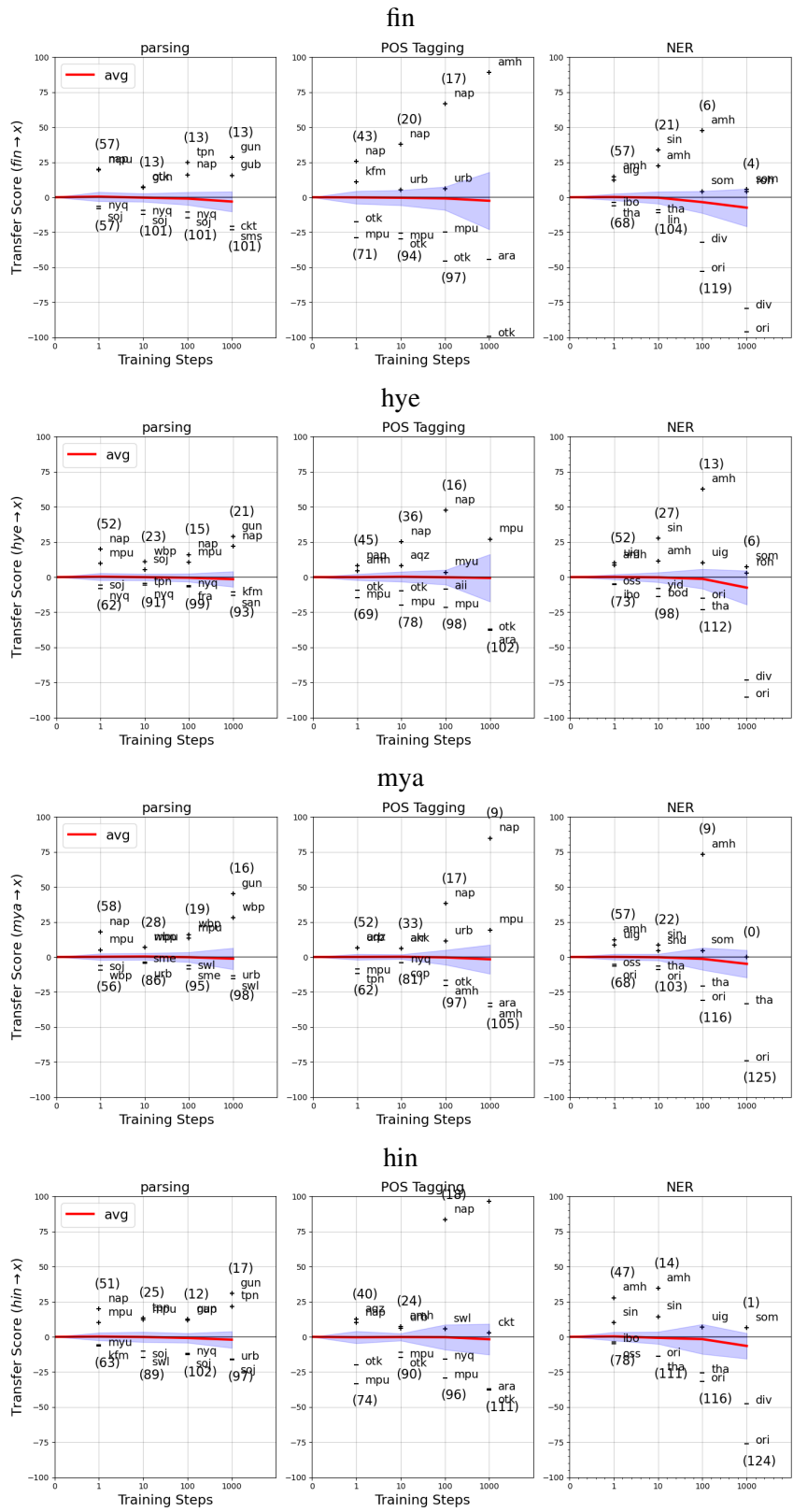


Figure 12: Aggregated Transfer Progression through training steps

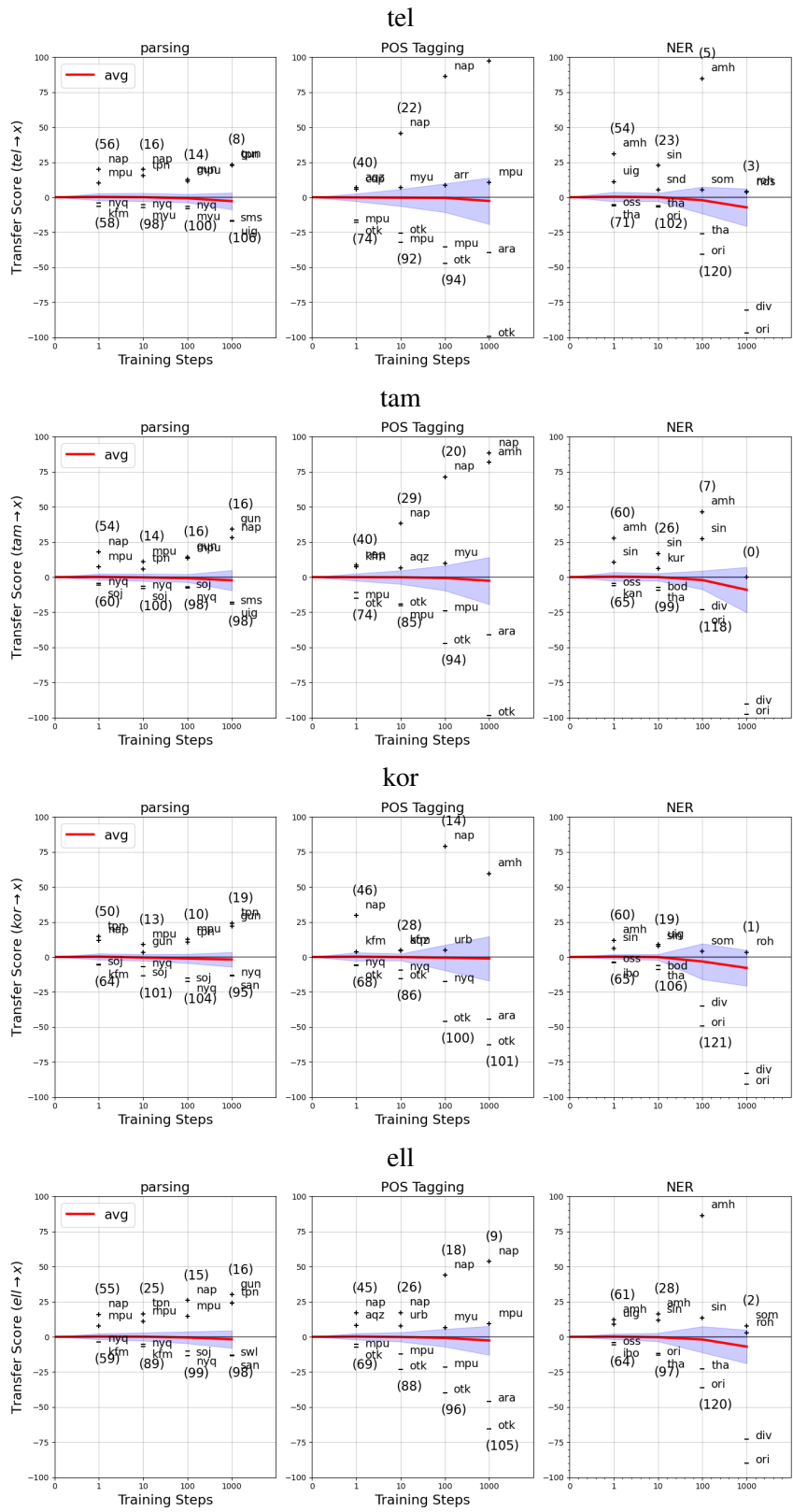


Figure 13: Aggregated Transfer Progression through training steps

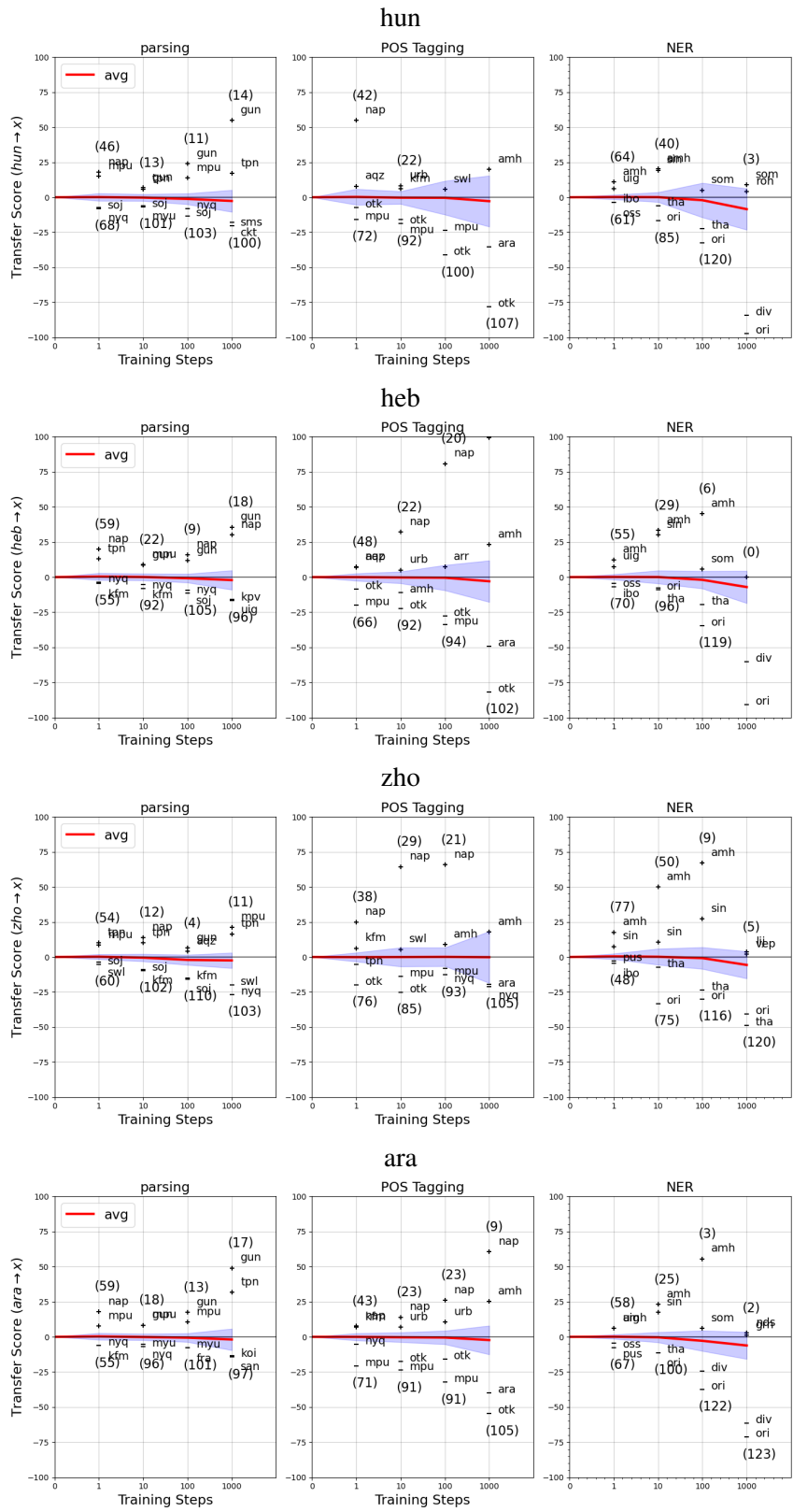


Figure 14: Aggregated Transfer Progression through training steps

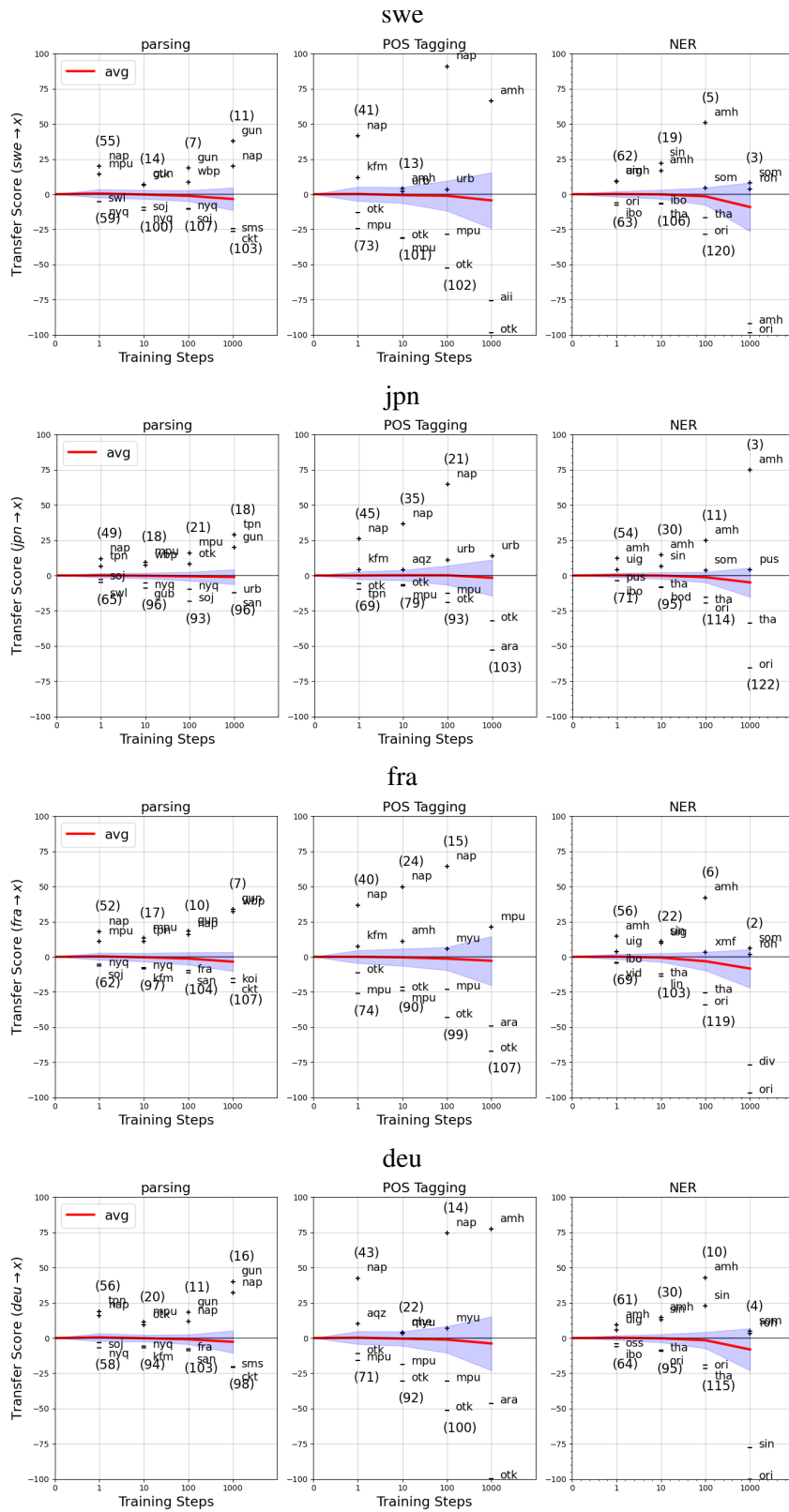


Figure 15: Aggregated Transfer Progression through training steps

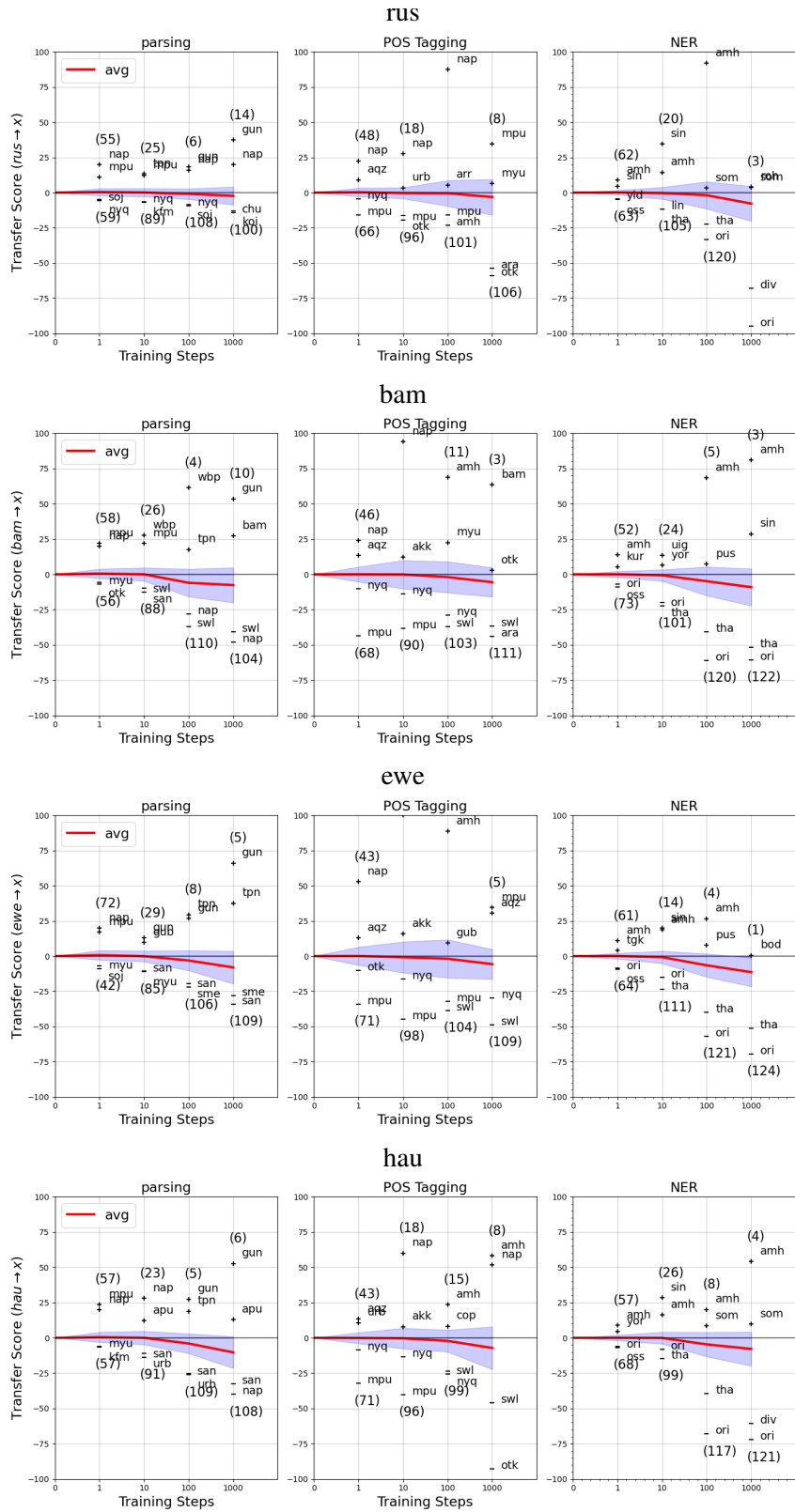


Figure 16: Aggregated Transfer Progression through training steps

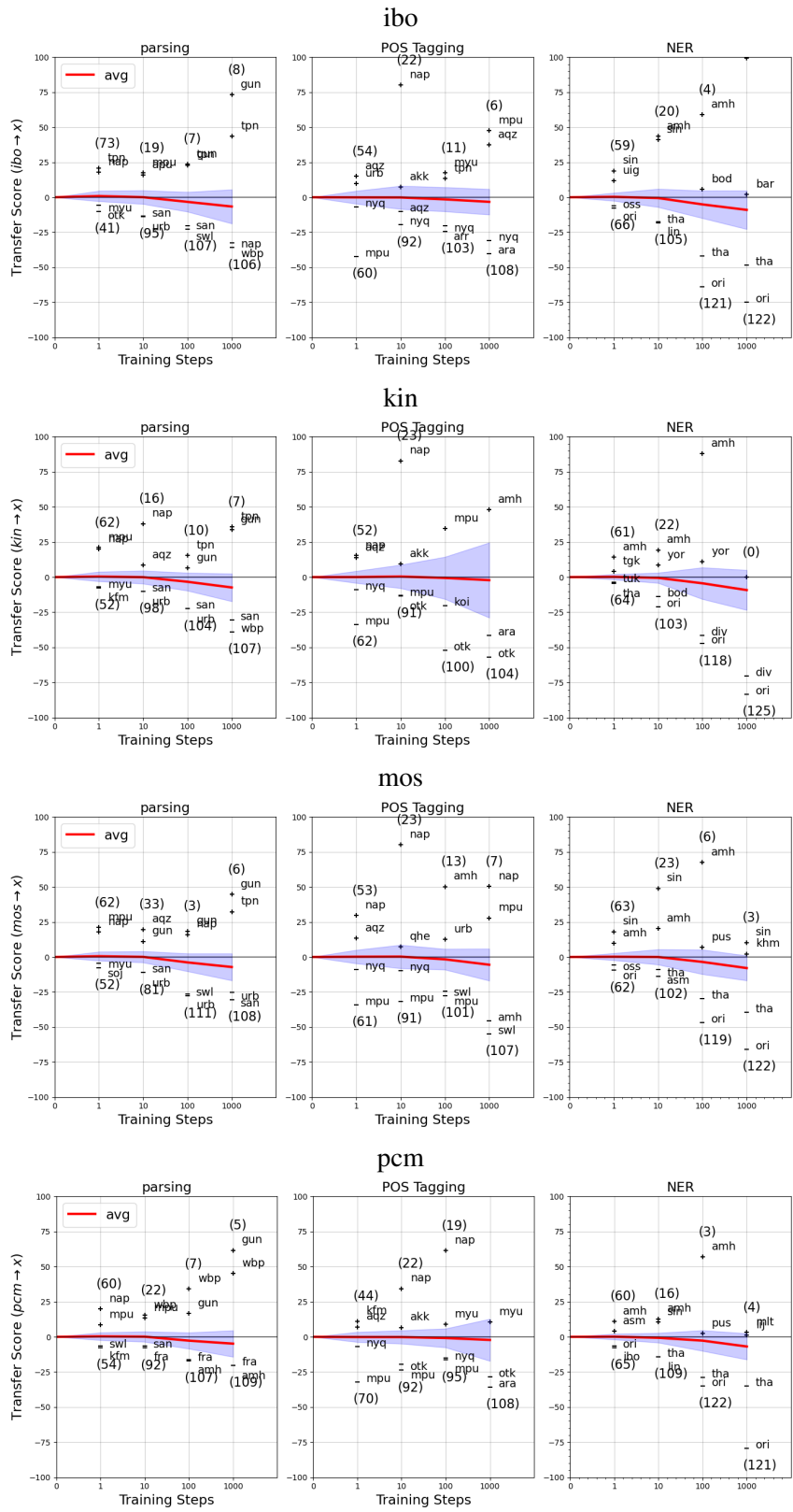


Figure 17: Aggregated Transfer Progression through training steps

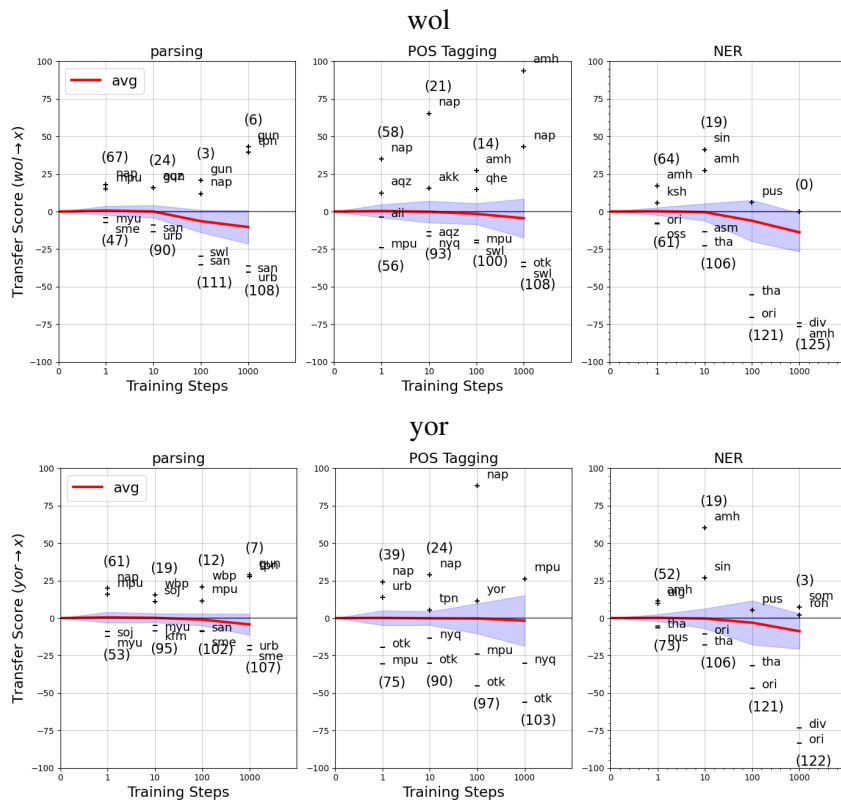


Figure 18: Aggregated Transfer Progression through training steps