Report embeddings

For this assignment I chose Russian language as a source and Czech language as a target and got the embeddings from the fastText website.

I used the VecMap tool in unsupervised way to create a cross-lingual mapping using only first 100000 words

```
python3 map_embeddings.py --unsupervised cut.cc.ru.300.vec cut.cc.cs.300.vec ru_mapped.emb cs_mapped.emb
```

and then just concatenated resulting mappings:

```
cat ru_mapped.emb cs_mapped.emb > mapped.emb
```

Then I chose to do the parsing. I trained the parser with bilingual embeddings on Russian training data from

```
./udpipe --train --tokenizer=none --tagger=none --parser='mapped.emb' ru_parser.model < UD_Russian-GSD/ru_gsd-ud-train.conllu
```

I compared the resulting parser on target Czech data and compared it with delexicalized Czech parser from tree_translation assignment:

```
./udpipe --parse --accuracy ru_parser.model < UD_Czech-PUD/cs_pud-ud-test.conllu
```

```
./udpipe --parse --accuracy cs-delex.udpipe < UD_Czech-PUD/cs_pud-ud-test.conllu
```

	LAS	UAS
Delex	61.81%	48.69%
Bilingual embedding	62.27%	52.53%

I also trained a delexicalized parser for Russian language:

```
./udpipe --train --tokenizer=none --tagger=none --parser='embedding_form=0;embedding_feats=0' ru.delex.parser.udpipe < UD_Russian-GSD/ru_gsd-ud-train.conllu
```

and compared the accuracy of both parsers on source language data:

```
./udpipe --parse --accuracy ru_parser.model < UD_Russian-GSD/ru_gsd-ud-test.conllu
```

```
./udpipe --parse --accuracy ru.delex.parser.udpipe < UD_Russian-GSD/ru_gsd-ud-test.conllu
```

	LAS	UAS
Delex	76.61%	71.20%
Bilingual embedding	85.25%	82.12%

In both cases, accuracy is better for the bilingual embedding parser.