

## A Detailed Hyperparameters

### A.1 Training Hyperparameters

We use exactly the same hyperparameters as the public multilingual BERT for finetuning our models. We train the part-of-speech tagging task for 10 epochs and the morphology task for 50 epochs.

For distillation, we use the following hyperparameters for all tasks.

- learning rate: 1e-4
- temperature: 3
- batch size: 256
- num epochs: 24

We take the Wikipedia pretraining data as is and drop sentences with fewer than 10 characters.

### A.2 Small BERT structure

We use the vocab and wordpiece model included with the cased public multilingual model on GitHub.

We use the BERT configuration of the public multilingual BERT with the following modifications for *mMiniBERT*.

- Hidden size = 256
- Intermediate layer size = 1024
- Num attention heads = 4
- Layers = 3

Model	Distilled	Part-of-Speech F1	Morphology F1
<i>mBERT</i>	No	94.5	91.0
<i>mMiniBERT</i>	Yes	93.7	88.6
<i>mMiniBERT</i>	No	90.2	85.5

Table 7: Ablation study to show the effect of distillation. The model without distillation has 3.5 points lower macro-averaged F1 on the part-of-speech task and 3.1 lower F1 on the morphology task.

## B Detailed Results

### B.1 The Importance of Distillation

To understand the importance of distillation in training *mMiniBERT*, we compare it to a model with the MiniBERT structure trained from scratch using only labeled multilingual data the teacher is trained on. Table 7 shows that distillation plays an important role in closing the accuracy gap between teacher and student.

### B.2 Per-Language Results

We show per-language F1 results of each model in Table 8 and Table 9. For per-language models, no models are trained for treebanks without tuning data, and metrics of those languages are not reported. All macro-averaged results reported exclude those languages.

treebank	BERT	Meta-LSTM	mBERT	mMeta-LSTM	mMiniBERT
af_afribooms	97.62	97.63	97.49	93.16	96.08
am_att			3.28	5.6	3.16
ar_padt	90.46	90.55	90.32	89	90.06
ar_pud			71.59	68.96	71.06
be_hse	94.81	91.05	95.02	87.59	94.95
bg_btb	99.01	98.77	98.72	96.43	98.19
ca_ancora	98.84	98.62	98.77	97.57	98.45
cs_cac	99.17	99.43	99.3	98.46	98.48
cs_cltt	87.48	87.25	87.67	87.62	87.53
cs_fictree	98.62	98.63	98.25	97.2	97.18
cs_pdt	99.06	99.07	98.99	98.22	98.61
cs_pud			97.13	96.53	97
da_ddt	97.59	97.47	97.18	92.36	95.93
de_gsd	94.81	94.17	94.53	91.94	93.82
de_pud			88.76	87.42	88.7
el_gdt	97.97	97.4	97.91	94.87	97.16
en_ewt	95.82	95.45	95.2	92.24	94.19
en_gum	96.22	95.02	94.79	92.33	94.24
en_lines	97.22	96.81	95.79	93.96	95.25
en_partut	96.11	95.9	95.02	93.29	94.61
es_ancora	98.87	98.78	98.17	96.27	97.8
es_gsd	93.7	93.9	89.65	90.61	89.58
es_pud			85.87	86.1	85.71
et_edt	97.27	97.17	97.02	94.32	95.64
eu_bdt	96.2	96.1	95.51	91.53	94.15
fa_seraji	97.57	97.17	97.17	95.29	96.92
fi_ftb	96.26	96.12	93.15	87.23	89.79
fi_pud			95.55	93.23	95.01
fi_tdt	96.81	97.02	93.9	91.58	92.6
fr_gsd	96.62	96.45	96.23	95.37	96.05
fr_partut	96.18	96	95.43	94.35	94.93
fr_pud			90.77	90.1	90.64
fr_sequoia	96.77	97.59	97.07	95.91	96.75
fr_spoken	97.55	95.78	96.1	90.07	93.25
ga_idt	91.92	91.55	90.83	84.16	85.72
gl_ctg	96.99	97.21	96.5	92.87	95.84
gl_treegal			93.4	91.28	91.9
he_htb	82.76	82.49	82.69	80.93	81.93
hi_hdtb	97.31	97.39	97.1	96.2	96.43
hi_pud			86.48	85.33	85.68
hr_set	97.79	97.94	97.47	96.24	97.2
hu_szeged	96.51	94.71	95.99	85.5	95.47
hy_armtdp	84.42		86.62	63.82	86.98
id_gsd	93.06	93.37	93.3	90.81	93.35
id_pud			63.52	63.5	63.33
it_isdt	98.33	98.06	98.27	96.7	97.8
it_partut	98.12	98.17	98.09	96.99	98.06
it_postwita	95.66	95.86	95.6	94.17	93.2
it_pud			93.84	92.72	93.67

ja_gsd	88.63	88.73	88.54	87.03	88.43
ja_modern			41.55	51.26	21.61
ja_pud			89.15	87.96	89.3
kk_ktb	75.93	61.7	81.36	52.91	80.06
ko_gsd	95.92	95.64	90.3	86.39	88.62
ko_kaist	95.56	95.42	93.86	87.46	93.43
ko_pud			41.93	46.11	31.96
la_ittb	98.34	98.42	98.3	97.18	97.65
la_perseus			89.91	83.85	85.23
la_proiel	96.34	96.37	95.97	92.02	93.78
lt_hse	88.88	81.43	90.01	65.6	86.9
lv_lvttb	94.79	94.47	93.71	88.25	91.3
mr_ufal	77.45	72.1	75.92	65.48	75.41
nl_alpino	97.1	96.16	97.33	93.78	96.19
nl_lassysmall	95.54	95.92	95.72	94.4	95.47
no_bokmaal	98	98	97.95	95.27	97.04
no_nynorskliia			94.08	88.27	92.55
no_nynorsk	97.94	97.92	97.69	94.91	96.59
pl_lfg	98.7	98.5	98.39	95.21	97.48
pl_sz	98.56	97.91	98.05	94.73	97.29
pt_bosque	96.74	96.73	96.16	95.53	95.85
pt_gsd	95.83	95.44	93.84	93.07	94.44
pt_pud			89.48	89.66	89.29
ro_nonstandard	94.67	94.48	94	92.05	91.9
ro_rrt	97.63	97.52	97.47	95.78	96.71
ru_gsd	92.23	91.39	90.84	88.13	90.14
ru_pud			89.7	88.92	89.52
ru_syntagrus	98.3	98.65	98.32	97.13	98.03
ru_taiga			93.62	92.75	93.18
sa_ufal			32.47	29.58	27.11
sk_snk	97.08	96.32	96.98	93.61	96.35
sl_ssja	97.07	96.68	96.89	94.24	95.58
sl_sst			94.51	90.34	91.79
sr_set	98.63	98.33	98.31	94.79	97.36
sv_lines	97.21	96.59	96.99	93.64	95.57
sv_pud			94.52	92.06	94.32
sv_talbanken	98.03	97.34	97.77	94.91	96.76
ta_ttb	75.71	72.7	74.28	61.51	74.6
te_mtg	94.25	92.72	93.42	87.32	93.42
th_pud			2.37	2.73	1.54
tl_trg			70.69	28.62	68.28
tr_imst	93.96	94.03	93.1	84.64	91.8
tr_pud			73.1	68.36	72.47
uk_iu	97.29	96.6	97.28	93	96.88
ur_udtb	93.83	93.87	93.69	93	93.05
vi_vtb	77.67	76.42	77.44	72.01	77.06
yo_ytb			43.48	30.85	34.59
zh_cfl			49.83	39.77	49.42
zh_gsd	87.6	85.7	85.96	82.76	86.08
zh_hk			66.29	57.88	65.86
zh_pud			83.3	73.3	82.95

---

Table 8: POS tagging F1 of all models.

treebank	BERT F1	Meta-LSTM F1	mBERT F1	mMeta-LSTM F1	mMiniBERT F1
af_afribooms	97.11	97.36	96.53	88.98	93.75
am_att				32.36	32.36
ar_padt	88.26	88.24	87.76	83.14	85.34
ar_pud			36.33	34.28	36.08
be_hse	82.83	74.03	87.52	59.16	81.82
bg_btb	97.54	97.58	97.47	91.41	95.4
ca_ancora	98.37	98.21	98.28	96.04	97.67
cs_cac	96.33	96.49	96.54	88.11	93.47
cs_cltt	81.61	79.89	83.86	78.82	80.61
cs_fictree	96.39	96.4	94.09	83.37	87.59
cs_pdt	97.18	96.91	97.15	89.77	94.63
cs_pud			93.88	87.44	91.81
da_ddt	97.22	97.08	95.62	89.82	94.08
de_gsd	90.84	90.58	90.4	80.69	88.99
de_pud			30.41	30.55	30.4
el_gdt	94.57	93.95	94.83	87.6	92.07
en_gum	96.87	96	93.79	90.11	93.71
en_lines	97.32	96.68	93.11	87.49	92.07
en_partut	94.88	95.38	90.76	79.99	90.18
en_pud			93.25	91.23	93.1
es_ancora	98.45	98.42	97.6	95.17	97
es_gsd	93.52	93.72	88.72	89.26	88.78
es_pud			52.7	52.8	52.73
et_edt	96.14	96.11	95.78	90.51	92.14
eu_bdt	93.27	92.56	92.67	76.72	84.53
fa_seraji	97.35	97.25	96.91	93.82	96.28
fi_ftb	96.34	96.48	92.32	77.89	86.47
fi_pud			93.58	91.12	91.65
fi_tdt	95.03	95.58	90.96	88.44	87.48
fr_gsd	96.05	96.11	94.67	86.97	94.51
fr_partut	93.32	92.93	88.9	87.48	87.05
fr_pud			59.15	57.5	58.94
fr_sequoia	97.09	97.13	91.54	85.23	90.74
fr_spoken	100	100	98.62	80.67	96.67
ga_idt	82.2	81.78	81.2	63.44	66.82
gl_ctg	98.98	98.95	95.27	89.98	95.1
gl_treegal			80.05	68.73	75.97
he_htb	81.27	80.85	80.79	76.89	78.74
hi_hdtb	93.32	93.85	92.91	89.09	90.65
hi_pud			22.1	22.37	22.03
hr_set	91.99	91.85	91.24	81.62	87.81
hu_szeged	93.65	91.28	92.93	71.25	87.36
hy_armtdp	41.13	54.45	51.08	36.59	46.43
id_gsd	94.84	96	94.85	91.62	94.39
id_pud			39.83	42.79	39.79
it_isdt	97.7	97.82	97.87	95.47	97.37
it_partut	97.35	97.73	98.01	96.33	97.9

it_postwita	95.62	96.05	95.03	91.52	93.17
it_pud			57.82	57.41	57.6
ja_gsd	90.29	90.45	90.29	90.39	90.41
ja_modern			63.9	61.17	63.99
ja_pud			57.4	57.26	57.27
kk_ktb			64.6	25.55	59.49
ko_gsd	99.62	99.55	99.4	98.99	99.37
ko_kaist	100	100	99.94	99.24	99.93
ko_pud			38.33	38.66	38.27
la_ittb	96.7	96.94	97.15	90.78	93.91
la_perseus			82.09	64.73	72.24
la_proiel	90.82	91.01	91.51	79.08	83.99
lt_hse	75.21	69.65	73.61	42.51	65.22
lv_lvtb	88.61	91.34	88.1	79.11	81.91
mr_ufal	63.95	59.11	64.2	33.63	54.01
nl_alpino	96.22	96.13	96.53	91.9	95.67
nl_lassysmall	96.46	96.02	95.55	92.16	95.28
no_bokmaal	96.85	97.13	96.48	91.17	95.31
no_nynorskliia			94.22	89.56	91.08
no_nynorsk	96.7	97.04	96.49	92.12	94.79
pl_lfg	95.85	94.68	84.96	47.99	84.56
pl_sz	93.9	91.93	71.4	73.02	65.36
pt_bosque	96.27	96.16	87.04	83.13	85.72
pt_gsd	97.2	95.33	67.72	76.01	71.88
pt_pud			52.06	49.79	50.95
ro_nonstandard	88.52	88.91	86.89	82.1	82.14
ro_rrt	97.02	97.23	96.58	93.2	94.85
ru_gsd	88.83	86.73	81.44	64.2	78.93
ru_pud			37.97	35.26	37.49
ru_syntagrus	97.02	96.9	95.99	91.96	94.33
ru_taiga			88.56	84.02	86.01
sa_ufal			15.9	16.14	16.33
sk_snk	92.06	89.63	91.58	68.25	85.29
sl_ssj	94.39	93.78	94.41	82.69	89.23
sl_sst	88.46		91.89	78.22	85.59
sr_set	94.83	94.71	92.79	73.51	90.48
sv_lines	89.54	89.55	88.66	83.27	86.4
sv_pud			77.39	73.94	76.79
sv_talbanken	96.92	96.56	96.13	90.23	94.49
ta_ttb	72.91	71.01	73.75	46.9	70.22
te_mtg	98.96	98.96	98.54	98.68	98.54
th_pud			8.27	0	8.43
tl_trg			29.31	28.62	25.17
tr_imst	89.5	91	88.63	73.23	81.99
tr_pud			23.72	23.84	23.46
uk_iu	92.4	90.98	92.64	79.49	88.79
ur_udtb	82.24	83.72	82.64	81.89	82.48
vi_vtb	83.74	84	83.93	83.58	83.94
yo_ytb			58.78	86.82	61.88
zh_cfl			46.55	43.55	45.73
zh_gsd	87.64	88.38	88.31	87.05	88.5

zh_hk	66.33	64.97	66.23
zh_pud	86.35	83.6	86.14

---

Table 9: Morphology F1 of all models.