Model / Task	**Average**	**Tokens**	**Sentences**	**Words**	**UPOS**	**XPOS**	**UFeats**	**AllTags**	**Lemmas**	**Time CPU**	**Time GPU**
concraft	91.61	98.56	71.33	99.64	95.88	90.04	90.59	90.04	96.79	3504.6	–
udpipe + fastText	94.43	99.75	90.51	**99.73**	97.36	90.64	90.97	90.64	95.86	**43.6**	45.2
combo + fastText	95.75	99.12	**93.33**	99.04	97.25	93.82	93.61	92.98	96.90	468.9	145.8
combo + HerBERT	**96.67**	99.12	**93.33**	99.04	**97.80**	**95.66**	**95.75**	**95.20**	**97.42**	827.8	166.6
stanza + fastText	95.89	**99.76**	92.70	99.45	97.43	93.57	93.90	93.36	96.94	163.1	52.1
spacy + pl-core-news-lg	75.38	99.56	61.85	98.46	96.30	90.97	31.03	30.14	94.77	62.6	**26.9**
spacy + fastText	75.15	99.56	61.85	98.46	95.89	89.93	31.03	30.08	94.43	61.9	27.1
spacy + BERT-pl	76.12	99.56	61.85	98.46	97.02	94.60	31.03	30.46	95.98	160.9	42.3
trankit + xlm-Roberta-base	92.59	98.37	89.39	97.84	95.36	89.74	90.05	88.73	91.19	480.3	192.3

Table 2: Results (F1 scores) and inference time (in seconds) of benchmarking morphosyntactic analyzers on the Morfeusz tagset averaged by the datasets (*byName* and *byType*). The evaluated systems are divided into two categories: non-neural and systems based on neural networks.

Model / Task	**Average**	**Tokens**	**Sentences**	**Words**	**UPOS**	**XPOS**	**UFeats**	**AllTags**	**Lemmas**	**Time CPU**	**Time GPU**
udpipe + fastText	92.30	**99.79**	92.44	**99.78**	97.33	89.97	90.37	89.35	95.23	51.0	51.5
combo + fastText	94.04	99.18	**94.29**	98.77	96.64	93.30	93.48	91.97	96.53	415.7	135.2
combo + HerBert	**95.51**	99.21	**94.29**	98.77	**97.57**	**95.33**	**95.61**	**94.54**	**97.13**	826.8	152.6
stanza + fastText	94.25	99.77	93.92	99.43	97.33	92.88	92.90	91.63	96.60	178.6	48.2
spacy + pl-core-news-lg	88.39	99.58	65.05	98.47	96.36	90.95	91.22	89.65	93.62	40.1	39.9
spacy + fastText	87.68	99.58	65.05	98.47	95.79	89.77	90.05	88.37	93.37	**40.0**	**29.6**
spacy + BERT-pl	90.70	99.58	65.05	98.47	97.26	94.68	94.84	94.09	94.89	173.5	48.4
trankit + xlm-Roberta-base	92.91	98.88	92.44	98.52	96.50	91.74	91.91	90.21	90.47	439.8	172.6

Table 3: Results (F1 scores) and inference time (in seconds) of benchmarking morphosyntactic analyzers on UD tagset averaged by the datasets (*byName*, *byType*, and *PDB-UD*). Similarly as in Table 2 the evaluated systems are divided into two categories. (Concraft is not included because it does not allow data in the UD tagset.)