UDPipe Frysk

version 20250407

Composition of corpus:

genre	#Tokens	#Words	#Sentences	%Tokens	%Words	%Sentences
historical	9200	8276	464	8	9	7
legal	1793	1596	78	2	2	1
narrative	49048	41158	3657	45	43	56
news	8737	7998	581	8	8	9
opinion	5993	5364	312	6	6	5
scientific	6163	5596	266	6	6	4
Wikipedia	27628	24653	1213	25	26	18
Total	108562	94641	6571	100	100	100

The training set includes 80% of the sentences, the validation (dev) set and the test set each 10%.

Results of k-fold cross-validation with $k = 10$ and using hyperparameter settings used for training based on LassySmall 2.5:

	Raw text		Gold tok		Gold tok + mor	
	mean	sd	mean	sd	mean	sd
f1 words	99.9	0.04				
f1 sents	98.6	0.73				
UPOS	95.8	0.17	95.8	0.20		
XPOS	93.1	0.34	93.2	0.36		
UFeats	94.6	0.27	94.7	0.29		
Lemma	96.6	0.11	97.0	0.11		
UAS	79.0	0.78	79.3	0.81	84.4	0.74
LAS	74.6	0.89	74.8	0.94	80.6	0.78