

## Appendix: Stack-Pointer Network for Dependency Parsing

### Appendix A: Hyper-Parameters

Table 5 summarizes the chosen hyper-parameters used for all the experiments in this paper. Some parameters are chosen directly or similarly from those reported in [Dozat and Manning \(2017\)](#). We use the same hyper-parameters across the models on different treebanks and languages, due to time constraints.

Layer	Hyper-parameter	Value
CNN	window size	3
	number of filters	50
LSTM	encoder layers	3
	encoder size	512
	decoder layers	1
	decoder size	512
MLP	arc MLP size	512
	label MLP size	128
Dropout	embeddings	0.33
	LSTM hidden states	0.33
	LSTM layers	0.33
Learning	optimizer	Adam
	initial learning rate	0.001
	$(\beta_1, \beta_2)$	(0.9, 0.9)
	decay rate	0.75
	gradient clipping	5.0

Table 5: Hyper-parameters for all experiments.

## Appendix B: UD Treebanks

Table 6 shows the corpora statistics of the treebanks for 12 languages. For evaluation, we report results excluding punctuation, which is any tokens with POS tags “PUNCT” or “SYM”.

	Corpora		<i>#Sent</i>	<i>#Token (w.o punct)</i>
Bulgarian	BTB	Training	8,907	124,336 (106,813)
		Dev	1,115	16,089 (13,822)
		Test	1,116	15,724 (13,456)
Catalan	AnCora	Training	13,123	417,587 (371,981)
		Dev	1,709	56,482 (50,452)
		Test	1,846	57,738 (51,324)
Czech	PDT, CAC CLTT FicTree	Training	102,993	1,806,230 (1,542,805)
		Dev	11,311	191,679 (163,387)
		Test	12,203	205,597 (174,771)
Dutch	Alpino LassySmall	Training	18,310	267,289 (234,104)
		Dev	1,518	22,091 (19,042)
		Test	1,396	21,126 (18,310)
English	EWT	Training	12,543	204,585 (180,308)
		Dev	2,002	25,148 (21,998)
		Test	2,077	25,096 (21,898)
French	GSD	Training	14,554	356,638 (316,780)
		Dev	1,478	35,768 (31,896)
		Test	416	10,020 (8,795)
German	GSD	Training	13,841	263,536 (229,204)
		Dev	799	12,348 (10,727)
		Test	977	16,268 (13,929)
Italian	ISDT	Training	12,838	270,703 (239,836)
		Dev	564	11,908 (10,490)
		Test	482	10,417 (9,237)
Norwegian	Bokmaal Nynorsk	Training	29,870	48,9217 (43,2597)
		Dev	4,300	67,619 (59,784)
		Test	3,450	54,739 (48,588)
Romanian	RRT	Training	8,043	185,113 (161,429)
		Dev	752	17,074 (14,851)
		Test	729	16,324 (14,241)
Russian	SynTagRus	Training	48,814	870,034 (711,184)
		Dev	6,584	118,426 (95,676)
		Test	6,491	117,276 (95,745)
Spanish	GSD AnCora	Training	28,492	827,053 (730,062)
		Dev	4,300	89,487 (78,951)
		Test	2,174	64,617 (56,973)

Table 6: Corpora statistics of UD Treebanks for 12 languages. *#Sent* and *#Token* refer to the number of sentences and the number of words (w./w.o punctuations) in each data set, respectively.

## Appendix C: Main Results

Table 7 illustrates the details of the experimental results. For each STACKPRT parsing model, we ran experiments with decoding beam size equals to 1, 5, and 10. For each experiment, we report the mean values with corresponding standard deviations over 5 runs.

Model	beam	English							
		Dev				Test			
		UAS	LAS	UCM	LCM	UAS	LAS	UCM	LCM
BiAF	–	95.73±0.04	<b>93.97±0.06</b>	60.58±0.77	47.47±0.63	95.84±0.06	<b>94.21±0.04</b>	59.49±0.23	49.07±0.34
Basic	1	95.71±0.02	93.88±0.03	62.33±0.33	47.75±0.32	95.71±0.06	94.07±0.06	60.91±0.35	49.54±0.48
	5	95.71±0.04	93.88±0.05	62.40±0.45	47.80±0.44	95.76±0.11	94.12±0.11	61.09±0.43	49.67±0.41
	10	95.72±0.03	93.89±0.04	<b>62.40±0.45</b>	<b>47.80±0.44</b>	95.77±0.11	94.12±0.11	61.09±0.43	49.67±0.41
+gpar	1	95.68±0.04	93.82±0.02	61.82±0.36	47.32±0.14	95.73±0.04	94.07±0.05	60.99±0.46	49.83±0.59
	5	95.67±0.01	93.83±0.02	61.93±0.32	47.44±0.20	95.76±0.06	94.11±0.06	61.23±0.47	50.07±0.59
	10	95.69±0.02	93.83±0.02	61.95±0.32	47.44±0.20	95.78±0.05	94.12±0.06	61.24±0.46	<b>50.07±0.59</b>
+sib	1	95.75±0.03	93.93±0.04	61.93±0.49	47.66±0.48	95.77±0.15	94.11±0.06	61.32±0.37	49.75±0.29
	5	95.74±0.02	93.93±0.05	62.16±0.22	47.68±0.54	95.84±0.09	94.17±0.09	61.52±0.57	49.91±0.76
	10	95.75±0.02	93.94±0.06	62.17±0.20	47.68±0.54	95.85±0.10	94.18±0.09	<b>61.52±0.57</b>	49.91±0.76
Full	1	95.63±0.08	93.78±0.08	61.56±0.63	47.12±0.36	95.79±0.06	94.11±0.06	61.02±0.31	49.45±0.23
	5	95.75±0.06	93.90±0.08	62.06±0.42	47.43±0.36	95.87±0.04	94.20±0.03	61.43±0.49	49.68±0.47
	10	<b>95.75±0.06</b>	93.90±0.08	62.08±0.39	47.43±0.36	<b>95.87±0.04</b>	94.19±0.04	61.43±0.49	49.68±0.47
Model	beam	Chinese							
		Dev				Test			
		UAS	LAS	UCM	LCM	UAS	LAS	UCM	LCM
BiAF	–	90.20±0.17	88.94±0.13	43.41±0.83	38.42±0.79	90.43±0.08	89.14±0.09	42.92±0.29	38.68±0.25
Basic	1	89.76±0.32	88.44±0.28	45.18±0.80	40.13±0.63	90.04±0.32	88.74±0.40	45.00±0.47	40.12±0.42
	5	89.97±0.13	88.67±0.14	45.33±0.58	40.25±0.65	90.46±0.15	89.17±0.18	45.41±0.48	40.53±0.48
	10	89.97±0.14	88.68±0.14	45.33±0.58	40.25±0.65	90.48±0.11	89.19±0.15	45.44±0.44	40.56±0.43
+gpar	1	90.05±0.14	88.71±0.16	45.63±0.52	40.45±0.61	90.28±0.10	88.96±0.10	45.26±0.59	40.38±0.43
	5	90.17±0.14	88.85±0.13	46.03±0.53	40.69±0.55	90.45±0.15	89.14±0.14	45.71±0.46	40.80±0.26
	10	90.18±0.16	88.87±0.14	<b>46.05±0.58</b>	40.69±0.55	90.46±0.16	89.16±0.15	45.71±0.46	40.80±0.26
+sib	1	89.91±0.07	88.59±0.10	45.50±0.50	40.40±0.48	90.25±0.10	88.94±0.12	45.42±0.52	40.54±0.69
	5	89.99±0.05	88.70±0.09	45.55±0.36	40.37±0.14	90.41±0.07	89.12±0.07	45.76±0.46	40.69±0.52
	10	90.00±0.04	88.72±0.09	45.58±0.32	40.37±0.14	90.43±0.09	89.15±0.10	45.75±0.44	40.68±0.50
Full	1	90.21±0.15	88.85±0.15	45.83±0.52	40.54±0.60	90.36±0.16	89.05±0.15	45.60±0.33	40.73±0.23
	5	90.23±0.13	88.89±0.14	46.00±0.54	<b>40.75±0.64</b>	90.58±0.12	89.27±0.11	<b>46.20±0.26</b>	<b>41.25±0.22</b>
	10	<b>90.29±0.13</b>	<b>88.95±0.13</b>	46.03±0.54	<b>40.75±0.64</b>	<b>90.59±0.12</b>	<b>89.29±0.11</b>	<b>46.20±0.26</b>	<b>41.25±0.22</b>
Model	beam	German							
		Dev				Test			
		UAS	LAS	UCM	LCM	UAS	LAS	UCM	LCM
BiAF	–	<b>93.60±0.13</b>	<b>91.96±0.13</b>	58.79±0.25	49.59±0.19	<b>93.85±0.07</b>	<b>92.32±0.06</b>	60.60±0.38	52.46±0.24
Basic	1	93.35±0.14	91.58±0.17	59.64±0.78	49.75±0.64	93.39±0.09	91.85±0.09	61.08±0.31	52.21±0.53
	5	93.49±0.14	91.72±0.16	59.99±0.69	49.82±0.54	93.61±0.09	92.07±0.08	61.38±0.30	52.51±0.43
	10	93.48±0.14	91.71±0.17	<b>60.02±0.69</b>	49.84±0.54	93.59±0.09	92.06±0.08	61.38±0.30	52.51±0.43
+gpar	1	93.39±0.07	91.66±0.13	59.59±0.54	49.81±0.42	93.44±0.07	91.91±0.11	61.73±0.47	52.84±0.48
	5	93.47±0.09	91.75±0.10	59.81±0.55	50.05±0.39	93.68±0.04	92.16±0.04	62.09±0.44	53.13±0.42
	10	93.48±0.08	91.76±0.09	59.89±0.59	50.09±0.40	93.68±0.05	92.16±0.03	62.10±0.42	<b>53.14±0.4</b>
+sib	1	93.43±0.07	91.73±0.08	59.68±0.25	49.93±0.30	93.55±0.07	92.00±0.08	61.90±0.50	52.79±0.22
	5	93.53±0.05	91.83±0.07	59.95±0.23	50.14±0.39	93.75±0.09	92.20±0.08	62.21±0.38	53.03±0.18
	10	93.55±0.06	91.84±0.07	59.96±0.24	<b>50.15±0.40</b>	93.76±0.09	92.21±0.08	<b>62.21±0.38</b>	53.03±0.18
Full	1	93.33±0.13	91.60±0.16	59.78±0.32	49.78±0.29	93.50±0.04	91.91±0.11	61.80±0.28	52.95±0.37
	5	93.42±0.11	91.69±0.12	59.90±0.27	49.94±0.35	93.64±0.03	92.10±0.06	61.89±0.21	53.06±0.36
	10	93.40±0.11	91.67±0.12	59.90±0.27	49.94±0.35	93.64±0.03	92.11±0.05	61.89±0.21	53.06±0.36

Table 7: Parsing performance of different variations of our model on both the development and test sets for three languages, together with the BiAF parser as the baseline. Best results are highlighted with bold print.