Appendix

Table 3 gives the percentage of head-initial syntagmatic dependencies and mean distances of syntagmatic dependencies for all the languages in the UD inventory of treebanks (as of November 2018), on both UD annotation and purely syntactic annotation.

Table 1: Percentage of head-initial syntagmatic dependencies and mean of syntagmatic distance (MDD) on both annotation schemes, UD and purely syntactic, for all the languages currently in the UD inventory of treebanks.

Language	% head initial UD	% head initial pur. synt.		MDD UD	MDD pur. synt.
Afrikaans	30.4	46.1		2.21	1.96
Amharic	47.1	48.9		1.00	1.16
Ancient Greek	37.6	43.8		1.81	1.81
Arabic	68.1	87.9		1.76	1.72
Armenian	27.3	29.0		1.49	1.42
Basque	37.2	28.0		1.31	1.32
Belarusian	38.2	60.5		1.39	1.23
Breton	43.4	62.9		1.16	0.92
Bulgarian	35.1	67.0		1.26	0.90
Buryat	15.5	9.0		1.32	1.32
Cantonese	44.9	48.5		1.32	1.30
Catalan	34.2	64.3		1.70	1.32
Chinese	26.4	32.5		2.17	2.11
Coptic	34.5	71.4		1.50	1.11
Croatian	35.4	61.4		1.44	1.10
Czech	35.5	59.9		1.41	1.18
Danish	33.5	61.1		1.73	1.31
Dutch	24.1	49.7		1.98	1.63
English	29.5	59.3		1.55	1.09
Estonian	31.0	44.4		1.35	1.17
Faroese	24.7	63.3		0.96	0.54
Finnish	33.4	48.9		1.15	0.95
French	33.7	62.0		1.48	1.07
Galician	44.0	63.6		1.68	1.47
German	23.3	47.1		2.13	1.91
Gothic	46.3	59.7		1.29	1.26
Greek	33.3	51.1		1.51	1.22
Hebrew	38.5	67.1		1.50	1.17

Hindi	41.2	11.8		2.48	2.13
Hungarian	18.7	21.6		1.89	1.77
Indonesian	45.2	69.6		1.40	1.20
Irish	50.2	78.8		1.90	1.64
Italian	33.9	62.0		1.43	1.08
Japanese	50.7	13.3		2.25	1.89
Kazakh	11.5	5.4		1.46	1.48
Komi	32.8	37.6		1.19	1.13
Korean	8.6	5.3		1.81	1.79
Kurmanji	35.3	53.7		1.72	1.89
Latin	36.9	55.3		1.69	1.59
Latvian	27.2	46.0		1.29	1.09
Lithuanian	25.2	46.3		1.46	1.29
Marathi	26.5	9.7		0.96	0.90
Naija	30.7	58.3		1.63	1.24
NorthSami	42.1	50.2		0.97	0.86
Norwegian	31.4	64.2		1.39	0.95
Old Church Slavonic	52.3	61.0		1.22	1.20
OldFrench	24.9	50.8		1.37	1.15
Persian	41.0	61.7		2.77	2.74
Polish	43.0	61.9		0.85	0.71
Portuguese	35.1	62.4		1.60	1.23
Romanian	41.5	71.3		1.36	1.09
Russian	36.3	58.5		1.30	1.15
Sanskrit	18.3	17.4		1.34	1.35
Serbian	38.0	65.2		1.43	1.02
Slovak	34.7	56.7		1.03	0.86
Slovenian	29.7	60.7		1.49	1.14
Spanish	35.1	64.4		1.58	1.20
Swedish	31.5	60.4		1.43	1.06
Swedish Sign	46.0	52.9		1.02	0.98
Tagalog	53.4	66.5		0.85	0.74
Tamil	19.8	7.7		1.46	1.40
Telugu	5.9	2.9		0.71	0.68
Thai	51.6	75.9		1.50	1.32
Turkish	15.2	8.6		1.68	1.61
Ukrainian	36.2	59.3		1.35	1.17
Upper Sorbian	27.4	53.8		1.86	1.61
Urdu	41.6	14.1		2.67	2.29

Uyghur	9.5	5.1		1.76	1.79
Vietnamese	51.2	64.6		1.17	1.07
Warlpiri	21.7	41.9		0.97	0.59
Yoruba	34.4	64.8		1.79	1.36

The scatter plot visualizes the MDD numbers in a form that achieves a comprehensive overview. All languages above the diagonal have lower MDD numbers on the purely syntactic analysis as compared to on the UD analysis. One can see the 6 languages that are anomalous (the ones below the line) as well as the minor extent to which they are anomalous.

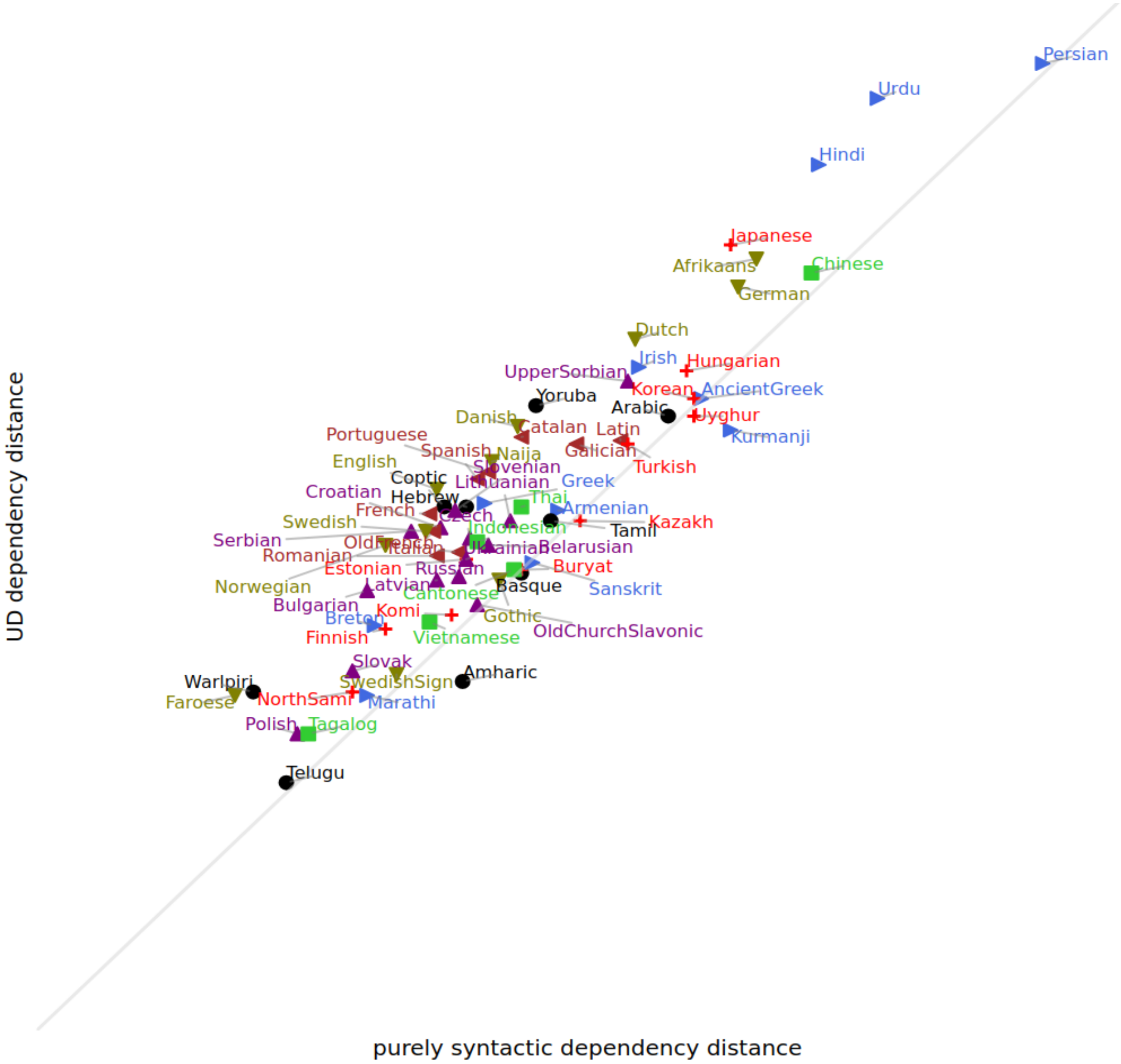

Figure 1: Scatter plot showing mean dependency distance (MDD) for each of the 71 languages currently in the UD inventory of treebanks.