Song, Jae Jung. *Linguistic typology: Morphology and syntax*. Routledge, 2014.

Zeman, Daniel. "Reusable Tagset Conversion Using Tagset Drivers." *LREC*. 2008.

Appendix A. *Selected Language Data*

Our study is based on the UD 2.0 treebanks of 43 languages combining 67 corpora.

As an example, we provide a table with the (alphabetically) first functions of rounded DDD data per language:

name	acl	advcl	advmod	amod	appos	aux
Arabic	3,37	9,87	3,42	1,39	3,43	-1,05
Bulgarian	5,07	2,73	-1,33	-1,09	2,58	-1,32
Catalan	5,51	7,41	-1,24	0,89	5,26	-1,45
Czech	5,58	1,72	-1,22	-0,97	4,83	-2,14
Old Church Slavonic	2,37	0,02	-0,97	0,66	1,63	0,79
Danish	5,42	5,15	-0,24	-0,63	2,59	-2,31
German	9,9	7,47	-1,84	-1,17	2,29	-4,54
Greek	4,25	4,01	-1,04	-1,08	5,67	-1,14
English	3,48	2,4	-0,93	-1,16	4,07	-1,58
Spanish	4,94	6,11	-1,16	0,7	3,45	-1,5
Estonian	2,07	3,39	-0,63	-1,04	2,84	-1,98
Basque	-1,83	-0,03	-1,93	0,43	4	0,78
Persian	7,81	-4,98	-5,66	0,95	2,81	-1,64
Finnish	1,4	2,24	-0,56	-1,19	2,96	-1,66
French	3,72	4,59	-1,17	0,65	3,2	-1,46
Irish	3,13	8,37	1,88	1,3	4,59	0
Galician	4,33	5,07	-1,06	0,78	5,14	-1,31
Gothic	3,35	1,04	-1,09	0,17	2,34	0,96
Ancient Greek	4,6	-0,52	-1,91	0,37	3,66	-1,73
Hebrew	4,53	2,83	-0,33	1,8	4,15	-1,96
Hindi	3,73	-5,67	-2,35	-1,32	0	1
Croatian	4,55	2,99	-1,48	-1,2	2,34	-1,54
Hungarian	8,67	4,22	-2,26	-1,39	3,67	0
Indonesian	3,81	4,65	-1,15	1,25	3,7	-1,33
Italian	3,84	2,46	-1,51	0,53	4,98	-1,32
Japanese	-6,35	0	-8,99	-1,43	0	1,76
Korean	-1,55	-5,22	-3,26	-1,08	-6,52	0
Latin	3,55	0,85	-2,33	0,1	3,5	0,55
Latvian	3,41	1,52	-1,5	-1,42	5,67	-1,11
Dutch	5	4,39	-1,67	-1,07	2,27	-2,62
Norwegian	3,77	3,71	-0,67	-0,94	4,79	-1,77
Polish	4,7	1,85	-1,13	-0,34	1,7	0,05
Portuguese	4,37	3,76	-1,29	0,46	3,68	-1,43
Romanian	4,13	3,37	-1,21	1	4,95	-1,21
Russian	4,19	3,07	-1,17	-1,05	2,31	-0,89
Slovak	4,57	1,73	-1,14	-1,06	3,68	-0,64
Slovenian	5,77	1,04	-1,28	-1,17	3,35	-2,35
Swedish	3,66	3,06	-0,64	-1,07	5,6	-1,95
Turkish	-2,46	0	-1,05	-1,9	2,11	1,35
Ukrainian	4,06	2,15	-1,28	-1,19	2,22	-0,65
Urdu	5,84	-3,73	-6,4	-1,43	0	1
Vietnamese	0	-3,61	-0,66	1,18	3,83	-0,77
Chinese	-4,88	-8,17	-2,5	-2,18	1,5	-2,67

The unabridged data used in this paper is available on https://gerdes.fr/papiers/2017/dependencyTypology/

code	Language	tokens
ar	Arabic	233, 712
ar_nyuad	Arabic	670, 612
bg	Bulgarian	123, 178
ca	Catalan	417, 453
cs	Czech	1, 174, 076
cs_cac	Czech	426, 274
cs_cltt	Czech	22, 000
cu	Old Church Slavonic	39, 394
da	Danish	80, 351
de	German	245, 524
el	Greek	47, 343
en	English	194, 428
en_lines	English	58, 223
en_partut	English	34, 195
es	Spanish	377, 020
es_ancora	Spanish	443, 951
et	Estonian	29, 051
eu	Basque	82, 516
fa	Persian	113, 699
fi	Finnish	152, 583
fi_ftb	Finnish	118, 747
fr	French	349, 973
fr_partut	French	16, 328
fr_sequoia	French	53, 635
ga	Irish	11, 627
gl	Galician	105, 844
gl_treegal	Galician	13, 819
got	Gothic	37, 931
grc	Ancient Greek	161, 184
grc_proiel	Ancient Greek	171, 524
he	Hebrew	127, 018
hi	Hindi	262, 007
hr	Croatian	161, 533
hu	Hungarian	27, 607
id	Indonesian	82, 588
it	Italian	254, 058
it_partut	Italian	38, 768
ja	Japanese	149, 147
ko	Korean	43, 921
la	Latin	15, 978
la_ittb	Latin	254, 683
la_proiel	Latin	134, 030
lv	Latvian	38, 476
code	Language	tokens
nl	Dutch	170, 665
nl_lassysmall	Dutch	73, 373
no_bokmaal	Norwegian	243, 529
no_nynorsk	Norwegian	240, 917
pl	Polish	63, 236
pt	Portuguese	196, 032
pt_br	Portuguese	260, 983
ro	Romanian	177, 755
ru	Russian	78, 025
ru_syntagrus	Russian	872, 362
sk	Slovak	79, 704
sl	Slovenian	113, 498
sl_sst	Slovenian	16, 389
sv	Swedish	65, 954
sv_lines	Swedish	56, 661
tr	Turkish	37, 167
uk	Ukrainian	11, 312
ur	Urdu	99, 024
vi	Vietnamese	25, 979
zh	Chinese	103, 614