

Table S1. We draw data from three projects aimed at standardizing dependency annotation schemes and reformatting existing dependency corpora according to them: the Google Universal Dependency Treebank (23) (UDT), HamleDT 2.0 (24), and Universal Dependencies 1.0 (25) (UD)

Language	Family/subfamily	Standardization	Source
Arabic	Semitic	HamleDT	Prague Arabic Dependency Treebank (43)
Bulgarian	IE/Slavic	HamleDT	BulTreeBank (44)
Bengali	IE/Indo-Aryan	HamleDT	Hyderabad Dependency Treebank (45)
Catalan	IE/Romance	HamleDT	AnCorà-CA (46)
Chinese	Sino-Tibetan	—	Chinese Dependency Treebank[40]
Croatian	IE/Slavic	HamleDT+	Croatian Dependency Treebank (47)
Czech	IE/Slavic	UD	Prague Dependency Treebank (48)
Danish	IE/Germanic	HamleDT	Danish Dependency Treebank (49)
German	IE/Germanic	HamleDT	TIGER (50)
Modern Greek	IE/Greek	HamleDT	Greek Dependency Treebank (51)
English	IE/Germanic	UD	English Web Treebank (52)
Spanish	IE/Romance	HamleDT	AnCorà-ES (46)
Estonian	Uralic/Finnic	HamleDT	Eesti keele puudepank (53)
Basque	Isolate	HamleDT	Basque Dependency Treebank (54)
Persian	IE/Indo-Aryan	HamleDT	Persian Dependency Treebank (55)
Finnish	Uralic/Finnic	UD	Turku Dependency Treebank (56)
French	IE/Romance	UD	French Dependency Treebank (57)
Ancient Greek	IE/Greek	HamleDT	Perseus Project (58)
Hebrew	Semitic	HamleDT+	MILA HaAretz Treebank (59)
Hindi	IE/Indo-Aryan	HamleDT	Hyderabad Dependency Treebank (45)
Hungarian	Uralic/Ugric	HamleDT	Szeged Treebank (60)
Indonesian	Austronesian	UDT	Indonesian Dependency Treebank (61)
Irish	IE/Celtic	UD	Irish Dependency Treebank (62)
Italian	IE/Romance	UD	Italian Stanford Dependency Treebank (63)
Japanese	Isolate	HamleDT	Tüba-J/S (64)
Korean	Isolate	UDT	Universal Dependency Treebank 1.0
Latin	IE/Romance	HamleDT	Perseus Project (58)
Dutch	IE/Germanic	HamleDT	Alpino Treebank (65)
Portuguese	IE/Romance	HamleDT	Bosque/Floresta Sintá(c)tica (66)
Romanian	IE/Romance	HamleDT	Resurse pentru Gramaticile de Dependenta (67)
Russian	IE/Slavic	HamleDT	SynTagRus (68)
Slovak	IE/Slavic	HamleDT	Slovak Treebank (69)
Slovenian	IE/Slavic	HamleDT	Slovene Dependency Treebank (70)
Swedish	IE/Germanic	UD	Talbanken05 (71)
Tamil	Dravidian	HamleDT	TamilTB (72)
Telugu	Dravidian	HamleDT	Hyderabad Dependency Treebank (45)
Turkish	Turkic	HamleDT	METU-Sabancı Treebank (73)

The HamleDT project corpora are automatically converted from their original hand-parsed form to the Stanford Dependencies standard. We also used the scripts provided by the HamleDT project to convert some corpora not in the official HamleDT 2.0 release to HamleDT format (Hebrew and Croatian). These are labeled as HamleDT+ in the table.

43. Smrz O, et al. (2008) Prague Arabic dependency treebank: A word on the million words. *Proceedings of the Workshop on Arabic and Local Languages (LREC 2008)*, eds Choukri K, Diab M, Maegaard B, Rosso P, Soudi A, Farghaly A (European Language Resources Association, Marrakech, Morocco), pp 16–23.
44. Simov K, Osenova P (2005) Extending the annotation of BulTreeBank: Phase 2. *The Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*, eds Civit M, Kuebler S, Martí MA (Universitat de Barcelona, Barcelona), pp 173–184.
45. Husain S, Mannem P, Ambati B, Gadde P (2010) The ICON-2010 tools contest on Indian language dependency parsing. *Proceedings of ICON-2010 Tools Contest on Indian Language Dependency Parsing*, eds Husain S, Mannem P, Ambati B, Gadde P (IIIT-Hyderabad, Kharagpur, India).
46. Taulé M, Martí MA, Recasens M (2008) AnCorà: Multilevel annotated corpora for Catalan and Spanish. *Proceedings of the 6th International Conference on Language Resources and Evaluation*, eds Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Tapias D (European Language Resources Association, Marrakech, Morocco), pp 96–101.
47. Berović D, Agić Ž, Tadić M (2012) Croatian dependency treebank: Recent development and initial experiments. *Proceedings of the 8th International Conference on Language Resources and Evaluation*, eds Calzolari N, Choukri K, Declerck T, Uğur Doğan M, Maegaard B, Mariani J, Moreno A, Odijk J, Piperidis S (European Language Resources Association, Istanbul, Turkey), pp 1902–1906.
48. Hajič J, et al. (2006) *Prague Dependency Treebank 2.0* (Linguistic Data Consortium, Philadelphia).
49. Kromann M (2003) The Danish dependency treebank and the DTAG treebank tool. *Proceedings of the Second Workshop on Treebanks and Linguistic Theories*, eds Nivre J, Hinrichs E (Växjö University Press, Växjö, Sweden), pp 217–220.
50. Brants S, Dipper S, Hansen S, Lezius W, Smith G (2002) The TIGER treebank. *Proceedings of the Workshop on Treebanks and Linguistic Theories*, eds Hinrichs E and Simov K (Sopotol, Bulgaria), pp 24–42.
51. Prokopiadis P, Desipri E, Koutsombogera M, Papageorgiou H, Piperidis S (2005) Theoretical and practical issues in the construction of a Greek dependency treebank. *Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT)*, eds Hajič J, Nivre J (Charles University in Prague, Department of Formal and Applied Linguistics, Prague), pp 149–160.
52. Silveira N, et al. (2014) A gold standard dependency corpus for English. *Proceedings of the 9th International Conference on Language Resources and Evaluation*, eds Calzolari N, Choukri K, Declerck T, Loftsson H, Maegaard B, Mariani J, Moreno A, Odijk J, Piperidis S (European Language Resources Association, Reykjavik, Iceland), pp 2897–2904.
53. Bick E, Uibo H, Müürisep K (2004) Arborest: A VISL-style treebank derived from an Estonian constraint grammar corpus. *Proceedings of Treebanks and Linguistic Theories*, eds Kuebler S, Nivre J, Hinrichs E, Wunsch H (Eberhard-Karls-Universität Tuebingen, Tuebingen, Germany), pp 9–20.
54. Aduriz I, et al. (2003) Construction of a Basque dependency treebank. *Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories*, eds Nivre J, Hinrichs E (Växjö University Press, Växjö, Sweden), pp 201–204.

55. Rasooli MS, Moloodi A, Kouhestani M, Minaei-Bidgoli B (2011) A syntactic valency lexicon for Persian verbs: The first steps towards Persian dependency treebank. *5th Language & Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics*, ed Vetulani Z (Adam Mickiewicz University, Poznań, Poland), pp 227–231.
56. Haverinen K, et al. (2010) Treebanking Finnish. *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories (TLT9)*, eds Dickinson M, Müürisep K, Passarotti M (Northern European Association for Language Technology, Tartu, Estonia), pp 79–90.
57. Candito M, Crabbé B, Denis P (2010) Statistical French dependency parsing: Treebank conversion and first results. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, eds Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (European Language Resources Association, Valletta, Malta), pp 1840–1847.
58. Bahman D, Crane G (2011) The Ancient Greek and Latin dependency treebanks. *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series, Theory and Applications of Natural Language Processing*, eds Sporleder C, von den Bosch A, Zervanou K (Springer, Berlin), pp 79–98.
59. Sima'an K, Itai A, Winter Y, Altman A, Nativ N (2001) Building a tree-bank of Modern Hebrew text. *Traitement Automatique des Langues* 42(2):347–380.
60. Csendes D, Csirik J, Gyimóthy T, Kocsor A (2005) The Szeged Treebank (TSD). *Text, Speech and Dialogue, 8th International Conference, Lecture Notes in Computer Science*, eds Matousek V, Mautner P, Pavelka T (Springer, Berlin), Vol 3658, pp 123–131.
61. Green N, Larasati SD, Žabokrtský Z (2012) Indonesian dependency treebank: Annotation and parsing. *Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation*, eds Manurung R, Bond F (Faculty of Computer Science, Universitas Indonesia, Bali, Indonesia), pp 137–145.
62. Lynn T, et al. (2012) Irish treebanking and parsing: A preliminary evaluation. *Proceedings of the 8th International Conference on Linguistic Resources and Evaluation*, eds Calzolari N, Choukri K, Declerck T, Uğur Doğan M, Maegaard B, Mariani J, Moreno A, Odijk J, Piperidis S (European Language Resources Association, Istanbul, Turkey), pp 1939–1946.
63. Bosco C, Montemagni S, Simi M (2013) Converting Italian treebanks: Towards an Italian Stanford dependency treebank. *The 7th Linguistic Annotation Workshop Interoperability with Discourse*, eds Pareja-Lora A, Liakata M, Dipper S (Association for Computational Linguistics, Sofia, Bulgaria), pp 61–69.
64. Kawata Y, Bartels J (2000) *Stylebook for the Japanese Treebank in Verbmobil*. Report 240 (Eberhard-Karls-Universitaet Tuebingen, Tuebingen, Germany).
65. van der Beek L, et al. (2002) The Alpino dependency treebank. *Algorithms for Linguistic Processing NWO PIONIER Progress Report*, eds van der Beek L, Bouma G, Daciuk J, Gaustad T, Malouf R, van Noord G, Prins R, Villada V (Graduate School for Behavioral and Cognitive Neurosciences Alfa-informatica, Groningen, The Netherlands), Chap 5.
66. Afonso S, Bick E, Haber R, Santos D (2002) "Floresta sintá(ç)tica": A treebank for Portuguese. *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, eds Zampolli A, (European Language Resources Association, Las Palmas, Spain), pp 1968–1703.
67. Căliăcean M (2008) Data-driven dependency parsing for Romanian. MS thesis (Uppsala Univ, Uppsala).
68. Boguslavsky I, Grigorieva S, Grigoriev N, Kreidlin L, Frid N (2000) Dependency treebank for Russian: Concept, tools, types of information. *Proceedings of the 18th Conference on Computational Linguistics*, eds Kay M, (Association for Computational Linguistics, Saarbruecken, Germany), Vol 2, pp 987–991.
69. Šimková M, Garabik R (2006) Sintaksičeskaja razmetka v slovackom nacional'nom korpusse. Trudy mezhdunarodnoj konferencii Korpusnaja lingvistika, (University of St. Petersburg, St. Petersburg, Russia), pp 389–394.
70. Dzeroski S, et al. (2006) Towards a Slovene dependency treebank. *Proceedings of the Fifth International Language Resources and Evaluation Conference*, eds Calzolari N, Maegaard B, Choukri K, Marconi L (European Language Resources Association, Genoa, Italy), pp 1388–1391.
71. Nivre J, Nilsson J, Hall J (2006) Talbanken05: A Swedish treebank with phrase structure and dependency annotation. *Proceedings of the 5th International Conference on Language Resources and Evaluation*, eds Calzolari N, Maegaard B, Choukri K, Marconi L (European Language Resources Association, Genoa, Italy), pp 1392–1395.
72. Ramasamy L, Žabokrtský Z (2012) Prague dependency style treebank for Tamil. *Proceedings of the 8th International Conference on Language Resources and Evaluation*, eds Calzolari N, Choukri K, Declerck T, Uğur Doğan M, Maegaard B, Mariani J, Moreno A, Odijk J, Piperidis S (European Language Resources Association, Istanbul, Turkey), pp 1888–1894.
73. Atalay NB, Oflazer K, Say B (2003) The annotation process in the Turkish treebank. *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC)*, eds Abeillé A, Hansen-Schirra S, Uszkoreit H (Association for Computational Linguistics, Budapest), pp 33–38.