Annotation/corpus projects with or for external partners 1999-2009

External partners Contact & institution	**lang-uage**	**Corpus / project**	**size**	**annotation**	**date**
Helena Britto, Marcelo Finger IME USP	pt	Tycho Brahe corpus historical Port. compatible annot.	60 m	morf/syn	Sept. '99
Sandra Aluísio ICMC, São Carlos, USP	pt	Tycho-Brahe 'CLAWS-sytle' annotation	100 m	morf/syn	Dec. '99
Ana Martins (IP), Ernestina Carrilho (CLUL)	pt	CORDIAL-SYN dialect speech	30 m	treebank	Mar '00
Roseli Imbernom do Nascimento	pt		117 m	pos/syn	May '00
Mark Davies ILSTU	pt	interviews pt vs. br	1 M	pos/syn	Nov '00
Diana Santos (SINTEF, Oslo) ACDC --> Linguateca	pt	Público, Folha, ... PALAVRAS installed	300 M	pos/syn	2001- ... ongoing
Diana Santos	pt	polylexicals/MWE dictionary	9000		Mar '01
Marcelo Modolo USP	pt	cartas, anuncios, diarios (historical texts)	80 m	pos/syn	2001 ...
M. Finatto	pt	frequency lists	2 M		Aug '01
Lene Petersen, Markus Dahl SDU/University of Bristol	en	KEMPE Shakespeare	9 M	syntax	2001-2003
Jørg Asmussen DSL	da	Corpus 90/2000	50 M	pos/syn	2002
Iris Kurz Gatti	pt, de	pt-de comparison of subjunctives	300 m	pos	Jan '02 Mar '04
Caroline Varaschin Gasperin Unisinos (Renata Vieira)	pt		10 m	treebank	Sep '02
Dorte Haltrup Hansen CST (via Nomen Nescio)	da	NER training corpus from Korpus2000	100 m	cg, NER	Apr. '02 ... Mar. '03
Christina Mota LABL, IST, UTL	pt	test texts	3 m	NER	Mar '03
Diana Santos Linguateca	pt	Morpholympics (winning system)		morph, pos	2003 Jan '04
Andreia Gentil Bonfante	pt		online	cg->trees	May '03
Paulo Quaresma University of Évora	pt	PALAVRAS installed	offline	sem	2003 ...

Susanne Salmon-Alt Ane Dybro Johansen LORIA, Nancy	fr	Ananas corpus (new text, le Monde) Goriot corpus	80 m 95 m	treebank treebank	Sep '03 Nov '03
Heli Uibo Tartu University	est	ArborEst SophieEst CG cooperation	2 m 200 s	cg treebank	2003-2004
Nordic treebank network coord. Joakim Nivre	da, se, est, pt	Sofies Verden parllel corpus	200 s / lang	revised treebank	Feb. '04
Susanne Salmon Alt LORIA, France	fr	talana	860 m	treebank	June '04
DEJO/VISL	eo	Esperanta Lingvo-Banko	18 M	pos/syn	2004
Aminova (Martin Carlsen, Moren Baun Møller, Mikael Colding-Jørgensen)	da	QA-corpus based on encyclopedic data	offline	pos/syn	2004
Paula Chesley, Susanne Salmon Alt LORIA, France	fr	verb complement subcategorisation lexicography	600 m 7 m	constituent treebank	Aug '04 June '05
Nuno Cardoso Linguateca	pt	HAREM joint evaluation (winning system)	500 m	NER	Feb '05 (preparations '04, eval '05)
Anna Eldman Ohio State University	es	Europarl sample	117 m	pos/syn	Oct '05
Sabine Buchholz Toshiba UK	da	CoNNL tests		dep	Nov '05
Tyler Ulrich	es	teaching SOAP access	online	syntax	2006-2009
Heini Justinussen	fo	Sosialurin Wikipedia	206 m	pos/syn	May '06
Line Jacobsen	da	comparative studies	100 s	dep treebank	Oct. '06
Stefan Evert Universität Osnabrück	de	Wikipedia sample	7 m	treebank	Oct. '06
Renata Vieira Unisinos, RG	pt	PALAVRAS installed dep-extraction (extract_deprel)	offline	sem	May '06 Nov '06
Graça Nunes NILC, São Carlos	pt	PLN-BR project PALAVRAS installed	offline	sem	2006
Nevena Tinkova Dept. Linguistica General Universitat de Barcelona	es	for comparative studies	100 s	treebank	Oct. '06 Mar/Apr. '07
Arina Greavu	ro	Business texts --> CorpusEye	21.4 M	pos/syn	2007

Leif Grönqvist Göteborg University / hum	se	Göteborgsposten --> CorpusEye	63 M	pos/syn	May '07
SDI (Johan Norberg, Martin Volk, Christian)	da, se	subtitles	10 M	pos (syn)	Dec. '07 Jan. '08
Violeta Quental	pt	Prepoly MWE lexicography		lex	Sep '06 - '07
Teresa Cadierno	en	helpdesk dialogue		pos/syn	2006/2007
Chris Callison-Burch John Hopkins University, US	es	ACL workshop data	100 m	treebank	March '08
Eloize ICMC, USP (NILC)	pt	dep-alignment	offline	dep-trees	March '08
Haitao Liu	eo	News, Mixed	70 m	dep	Jan. '08
Linguateca. polo Coimbra Cláudia Freitas, Paulo Rocha		Floresta Selva	1.5 M		2008
Dagbladet Information (Johannes Wehner)	da	OpenCalais corpus --> CorpusEye	80 M	pos/syn inline NER	2008
Paulo Quaresma University of Évora	en	Reuters corpus eurlex, for text classification	10 M	treebank	Jan '09
Francisco Pereira	pt	test interest in WBR-99	small	treebank	feb. '09
Paula Carvalho, Mario Silva di.fc.ul (plan only)	pt	SOAP access NER for opinion mining	online	NER	feb. '09
Vládia Pinheiro Fundação Cearense de Pesquisa e Cultura	pt	SOAP access	online		June '09 ...
Helmara F. Real de Moraes USP, Brazil	pt	upload	online		2. aug. '09
Heliana Mello, Tommaso Raso UFMG, Brazil	pt	C-ORAL (speech data)		pos/syn	2009-2010
Jorge Baptista (Faro), Nuno Mamede (Lissabon)	pt	Semantic prototypes lexicon		sem	2009-2010
Martin Volk, Claire Lavanchy Universität Zürich	fr	Sofies verden	250 s	treebank	4.-13. aug. '09
Aline Villavicencio, Adriano Zanette, Leonardo Zilio UFRGS, Brazil	pt	PALAVRAS installed	offline		112. jan. '09
Elma Kerz, Institut für Anglistik RWTH Aachen University	en	advertisements	300 m	pos/syn	13.-15. dec. 2009