

7

References

- [1] BLUMBERG, R.; ATRE, S.. **The problem with unstructured data, dm review magazine**. In: DM REVIEW MAGAZINE, p. –, 2003.
- [2] MANNING, C. D.; SCHUETZE, H.. **Foundations of Statistical Natural Language Processing**. The MIT Press, 1 edition, June 1999.
- [3] BRILL, E.. **Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging**. Comput. Linguistics, 21(4):543–565, 1995.
- [4] TJONG KIM SANG, E. F.; BUCHHOLZ, S.. **Introduction to the conll-2000 shared task: Chunking**. In: Cardie, C.; Daelemans, W.; Nedellec, C. ; Tjong Kim Sang, E., editors, PROCEEDINGS OF CONLL-2000 AND LLL-2000, p. 127–132. Lisbon, Portugal, 2000.
- [5] TJONG KIM SANG, E. F.; DÉJEAN, H.. **Introduction to the conll-2001 shared task: Clause identification**. In: Daelemans, W.; Zajac, R., editors, PROCEEDINGS OF CONLL-2001, p. 53–57. Toulouse, France, 2001.
- [6] TJONG KIM SANG, E. F.. **Introduction to the conll-2002 shared task: Language-independent named entity recognition**. In: PROCEEDINGS OF CONLL-2002, p. 155–158. Taipei, Taiwan, 2002.
- [7] TJONG KIM SANG, E. F.; DE MEULDER, F.. **Introduction to the conll-2003 shared task: Language-independent named entity recognition**. In: Daelemans, W.; Osborne, M., editors, PROCEEDINGS OF CONLL-2003, p. 142–147. Edmonton, Canada, 2003.
- [8] CARRERAS, X.; MÀRQUES, L.. **Introduction to the conll-2004 shared task: Semantic role labeling**. In: PROCEEDINGS OF CONLL-2004, p. 89–97. Boston, MA, USA, 2004.
- [9] CARRERAS, X.; MÀRQUEZ, L.. **Introduction to the CoNLL-2005 shared task: Semantic role labeling**. In: PROCEEDINGS OF THE

NINTH CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING (CONLL-2005), p. 152–164, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

- [10] HACIOGLU, K.. **Semantic role labeling using dependency trees.** In: COLING '04: PROCEEDINGS OF THE 20TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, p. 1273, Morristown, NJ, USA, 2004. Association for Computational Linguistics.
- [11] MCDONALD, R.; NIVRE, J.. **Characterizing the errors of data-driven dependency parsing models.** In: PROCEEDINGS OF THE 2007 JOINT CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING AND COMPUTATIONAL NATURAL LANGUAGE LEARNING (EMNLP-CONLL), p. 122–131, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [12] MILIDIU, R. L.; CRESTANA, C. E. M. ; DOS SANTOS, C. N.. **A token classification approach to dependency parsing.** In: PROCEEDINGS OF STIL 2009, 2009.
- [13] COVINGTON, M. A.. **A fundamental algorithm for dependency parsing.** In: IN PROCEEDINGS OF THE 39TH ANNUAL ACM SOUTHEAST CONFERENCE, p. 95–102, 2000.
- [14] JIJKOUN, V.; RIJKE, M. D. ; MUR, J.. **Information extraction for question answering: Improving recall through syntactic patterns.** In: IN COLING 2004, p. 1284–1290, 2004.
- [15] CUI, H.; LI, K.; SUN, R.; SENG CHUA, T. ; YEN KAN, M.. **National university of singapore at the trec 13 question answering main task.** In: IN PROCEEDINGS OF THE 13 TH TREC, 2005.
- [16] BOUMA, G.; MUR, J.; NOORD, G. V.; PLAS, L. V. D.; TIEDEMANN, J. ; GRONINGEN, R.. **Question answering for dutch using dependency relations.** In: IN: WORKING NOTES FOR THE CLEF 2005 WORKSHOP. Springer, 2005.
- [17] CUI, H.; SUN, R.; LI, K.; YEN KAN, M. ; SENG CHUA, T.. **Question answering passage retrieval using dependency relations.** In: IN SIGIR 2005, p. 400–407. ACM Press, 2005.
- [18] QUIRK, C.; ET AL.. **The impact of parse quality on syntactically-informed statistical machine translation,** 2006.

- [19] DING, Y.; PALMER, M.. **Machine translation using probabilistic synchronous dependency insertion grammars**, 2005.
- [20] SHEN, L.; XU, J. ; WEISCHEDEL, R.. **A new string-to-dependency machine translation algorithm with a target dependency language model**. In: PROCEEDINGS OF ACL-08: HLT, p. 577–585, Columbus, Ohio, June 2008. Association for Computational Linguistics.
- [21] CIRAVEGNA, F.; LAVELLI, A.. **Full text parsing using cascades of rules: an information extraction perspective**. In: IN PROC. OF THE 9TH CONF. OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 1999.
- [22] SÆTRE, R.; SAGAE, K.. **Syntactic features for protein-protein interaction extraction**.
- [23] ALPHONSE, E.; AUBIN, S.; BISSON, G.; HAMON, T.; LAGARRIGUE, R.; NAZARENKO, A.; PIERRE MANINE, A.; NÉDELLEC, C.; OULD, M.; VETAH, A.; POIBEAU, T. ; WEISSENBACHER, D.. **Event-based information extraction for the biomedical domain: The caderige project**. In: PROCEEDINGS OF THE INTERNATIONAL JOINT WORKSHOP ON NATURAL LANGUAGE PROCESSING IN BIOMEDICINE AND ITS APPLICATIONS. (2004), 2004.
- [24] KOLLER, A.; KOLLER, E. ; STRIEGNITZ, K.. **Generation as dependency parsing**. In: IN PROCEEDINGS OF THE 40TH ACL, p. 17–24, 2002.
- [25] DEBUSMANN, R.; DUCHIER, D. ; KUHLMANN, M.. **M.: Multi-dimensional graph configuration for natural language processing**. In: IN: CONSTRAINT SOLVING AND LANGUAGE PROCESSING. VOLUME 3438 OF LECTURE NOTES IN COMPUTER SCIENCE, p. 104–120. Springer, 2005.
- [26] COVINGTON, M. A.. **Syntactic Theory in the High Middle Ages**. Cambridge Studies in Linguistics 39. Cambridge University Press, New York, 2009.
- [27] JAN M. KRUIJFF, G.. **Formal and computational aspects of dependency grammar: History and development of dg**. Technical report, ESSLLI Course Notes, FoLLI, the Association of Logic, Language and Information, 2002.

- [28] MEL'CUK, I. A.. **Dependency Syntax: Theory and Practice**. State University of New York Press, New York, 1988.
- [29] TESNIÈRE, L.. **Eléments de Syntaxe Structurale**. Klincksieck, Paris, 1959.
- [30] NIVRE, J.. **Dependency grammar and dependency parsing**. Technical report, Växjö University: School of Mathematics and Systems Engineering, 2005.
- [31] NIVRE, J.; NILSSON, J.. **Pseudo-projective dependency parsing**. In: ACL '05: PROCEEDINGS OF THE 43RD ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, p. 99–106, Morristown, NJ, USA, 2005. Association for Computational Linguistics.
- [32] NIVRE, J.. **Constraints on non-projective dependency parsing**. In: IN PROC. EACL, p. 73–80, 2006.
- [33] NIVRE, J.; HALL, J.; KÜBLER, S.; MCDONALD, R.; NILSSON, J.; RIEDEL, S. ; YURET, D.. **The CoNLL 2007 shared task on dependency parsing**. In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 915–932, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [34] CARROLL, G.; CARROLL, G.; CHARNIAK, E. ; CHARNIAK, E.. **Two experiments on learning probabilistic dependency grammars from corpora**. In: WORKING NOTES OF THE WORKSHOP STATISTICALLY-BASED NLP TECHNIQUES, p. 1–13. AAAI, 1992.
- [35] EISNER, J.. **Bilexical grammars and their cubic-time parsing algorithms**, 2000.
- [36] COLLINS, M.; RAMSHAW, L. ; TILLMANN, C.. **A statistical parser for czech**. In: IN PROCEEDINGS OF THE 37TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, p. 505–512, 1999.
- [37] SAMUELSSON, C.. **A statistical theory of dependency syntax**, 2000.
- [38] KUDO, T.; MATSUMOTO, Y.. **Japanese dependency structure analysis based on support vector machines**, 2000.
- [39] KUDO, T.; MATSUMOTO, Y.. **Japanese dependency analysis using cascaded chunking**, 2002.

- [40] YAMADA, H.; MATSUMOTO, Y.. **Statistical dependency analysis with support vector machines.** In: IN PROCEEDINGS OF IWPT, p. 195–206, 2003.
- [41] VAPNIK, V. N.. **The nature of statistical learning theory.** Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [42] NIVRE, J.; HALL, J.; NILSSON, J. ; MARINOV, S.. **Labeled pseudo-projective dependency parsing with support vector machines.** In: IN PROCEEDINGS OF THE TENTH CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING (CONLL), p. 221–225, 2006.
- [43] NIVRE, J.. **An efficient algorithm for projective dependency parsing.** In: PROCEEDINGS OF THE 8TH INTERNATIONAL WORKSHOP ON PARSING TECHNOLOGIES (IWPT), p. 149–160, 2003.
- [44] JOHAN, J. N.. **Memory-based dependency parsing,** 2004.
- [45] NIVRE, J.; HALL, J.; NILSSON, J. ; MARINOV, S.. **Labeled pseudo-projective dependency parsing with support vector machines.** In: IN PROCEEDINGS OF THE TENTH CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING (CONLL), p. 221–225, 2006.
- [46] ATTARDI, G.; DELL’ORLETTA, F.; SIMI, M.; CHANEV, A. ; CIARAMITA, M.. **Multilingual dependency parsing and domain adaptation using DeSR.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 1112–1118, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [47] DUAN, X.; ZHAO, J. ; XU, B.. **Probabilistic parsing action models for multi-lingual dependency parsing.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 940–946, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [48] HALL, J.; NILSSON, J.; NIVRE, J.; ERYIGIT, G.; MEGYESI, B.; NILSSON, M. ; SAERS, M.. **Single malt or blended? a study in multilingual parser optimization.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 933–939, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

- [49] JOHANSSON, R.; NUGUES, P.. **Incremental dependency parsing using online learning.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 1134–1138, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [50] TITOV, I.; HENDERSON, J.. **Fast and robust multilingual dependency parsing with a generative latent variable model.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 947–951, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [51] SCHIEHLEN, M.; SPRANGER, K.. **Global learning of labeled dependency trees.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 1156–1160, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [52] NGUYEN, L.-M.; SHIMAZU, A.; NGUYEN, P.-T. ; PHAN, X.-H.. **A multilingual dependency analysis system using online passive-aggressive learning.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 1149–1155, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [53] SHIMIZU, N.; NAKAGAWA, H.. **Structural correspondence learning for dependency parsing.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 1166–1169, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [54] HALL, K.; HAVELKA, J. ; SMITH, D. A.. **Log-linear models of non-projective trees, k -best MST parsing and tree-ranking.** In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 962–966, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [55] MCDONALD, R.; PEREIRA, F.; RIBAROV, K. ; HAJIC, J.. **Non-projective dependency parsing using spanning tree algorithms.** In: PROCEEDINGS OF HUMAN LANGUAGE TECHNOLOGY CONFERENCE AND CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, p. 523–530, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

- [56] CARRERAS, X.. **Experiments with a higher-order projective dependency parser**. In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 957–961, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [57] MCDONALD, R.; PEREIRA, F.. **Online learning of approximate dependency parsing algorithms**. In: IN PROC. OF EACL, p. 81–88, 2006.
- [58] NAKAGAWA, T.. **Multilingual dependency parsing using global features**. In: PROCEEDINGS OF THE CONLL SHARED TASK SESSION OF EMNLP-CONLL 2007, p. 952–956, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [59] EISNER, J. M.. **Three new probabilistic models for dependency parsing: An exploration**. In: IN PROCEEDINGS OF THE 16TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS (COLING-96), p. 340–345, 1997).
- [60] CANISIUS, S.; BOGERS, T.; VAN DEN BOSCH, A.; GEERTZEN, J. ; TJONG KIM SANG, E.. **Dependency parsing by inference over high-recall dependency predictions**. 2006.
- [61] ATTARDI, G.. **Experiments with a multilanguage non-projective dependency parser**. 2006.
- [62] WU, Y.; LEE, Y. ; YANG, J.. **The exploration of deterministic and efficient dependency parsing**. 2006.
- [63] CARRERAS, X.; SURDEANU, M. ; MÀRQUEZ, L.. **Projective dependency parsing with perceptron**. 2006.
- [64] YURET, D.. **Dependency parsing as a classification problem**. 2006.
- [65] BICK, E.. **LingPars, a linguistically inspired, language-independent machine learner for dependency treebanks**. 2006.
- [66] NIVRE, J.; HALL, J.; NILSSON, J.; ERYİĞİT, G. ; MARINOV, S.. **Labeled pseudo-projective dependency parsing with support vector machines**. 2006.
- [67] SCHIEHLEN, M.; SPRANGER, K.. **Language independent probabilistic context-free parsing bolstered by machine learning**. 2006.

- [68] LIU, T.; MA, J.; ZHU, H. ; LI, S.. **Dependency parsing based on dynamic local optimization.** 2006.
- [69] DREYER, M.; SMITH, D. ; SMITH, N.. **Vine parsing and minimum risk reranking for speed and precision.** 2006.
- [70] JOHANSSON, R.; NUGUES, P.. **Investigating multilingual dependency parsing.** 2006.
- [71] MCDONALD, R.; LERMAN, K. ; PEREIRA, F.. **Multilingual dependency analysis with a two-stage discriminative parser.** 2006.
- [72] RIEDEL, S.; ÇAKICI, R. ; MEZA-RUIZ, I.. **Multi-lingual dependency parsing with incremental integer linear programming.** 2006.
- [73] SHIMIZU, N.. **Maximum spanning tree algorithm for non-projective labeled dependency parsing.** 2006.
- [74] CORSTON-OLIVER, S.; AUE, A.. **Dependency parsing with reference to Slovene, Spanish and Swedish.** 2006.
- [75] CHENG, Y.; ASAHARA, M. ; MATSUMOTO, Y.. **Multi-lingual dependency parsing at naist.** 2006.
- [76] RAMSHAW, L.; MARCUS, M.. **Text chunking using transformation-based learning.** In: Armstrong, S.; Church, K.; Isabelle, P.; Manzi, S.; Tzoukermann, E. ; Yarowsky, D., editors, NATURAL LANGUAGE PROCESSING USING VERY LARGE CORPORA. Kluwer, 1999.
- [77] DOS SANTOS, C. N.; MILIDIÚ, R. L.. **Foundations of Computational Intelligence, Volume 1: Learning and Approximation**, volumen 201 de **Studies in Computational Intelligence**, chapter Entropy Guided Transformation Learning, p. 159–184. Springer, 2009.
- [78] DOS SANTOS, C. N.; MILIDIÚ, R. L. ; RENTERÍA, R. P.. **Portuguese part-of-speech tagging using entropy guided transformation learning.** In: PROCEEDINGS OF PROPOR 2008, p. 143–152, Aveiro, Portugal, September 2008.
- [79] DOS SANTOS, C. N.; MILIDIÚ, R. L.; CRESTANA, C. E. M. ; FERNANDES, E. R.. **Etl ensembles for chunking, ner and srl.** In: PROCEEDINGS OF CICLING 2010, 2010.

- [80] MILIDIÚ, R. L.; DOS SANTOS, C. N. ; DUARTE, J. C.. **Portuguese corpus-based learning using ETL**. Journal of the Brazilian Computer Society, 14(4), 2008.
- [81] FERNANDES, E. R.; MILIDIU, R. L. ; SANTOS, C. N.. **Portuguese language processing service**. In: 18TH INTERNATIONAL WORLD WIDE WEB CONFERENCE, April 2009.
- [82] FINGER, M.. **Técnicas de otimização da precisão empregadas no etiquetador tycho brahe**. In: PROCEEDINGS OF PROPOR, p. 141–154, São Paulo, November 2000.
- [83] MEGYESI, B.. **Shallow parsing with pos taggers and linguistic features**. Journal of Machine Learning Research, 2:639–668, 2002.
- [84] DOS SANTOS, C. N.; OLIVEIRA, C.. **Constrained atomic term: Widening the reach of rule templates in transformation based learning**. In: EPIA, p. 622–633, 2005.
- [85] MILIDIÚ, R. L.; DOS SANTOS, C. N.; DUARTE, J. C. ; RENTERIA, R. P.. **Semi-supervised learning for portuguese noun phrase extraction**. In: PROCEEDINGS OF 7TH WORKSHOP ON COMPUTATIONAL PROCESSING OF WRITTEN AND SPOKEN PORTUGUESE, p. 200–203, Itatiaia, Brazil, May 2006.
- [86] MANGU, L.; BRILL, E.. **Automatic rule acquisition for spelling correction**. In: PROCEEDINGS OF THE FOURTEENTH ICML. Morgan Kaufmann, 1997.
- [87] FREITAS, M. C.; DUARTE, J. C.; DOS SANTOS, C. N.; MILIDIÚ, R. L.; RENTERIA, R. P. ; QUENTAL, V.. **A machine learning approach to the identification of appositives**. In: PROCEEDINGS OF IBERO-AMERICAN AI CONFERENCE, Ribeirão Preto, Brazil, October 2006.
- [88] FLORIAN, R.. **Named entity recognition as a house of cards: Classifier stacking**. In: PROCEEDINGS OF CONLL-2002, p. 175–178. Taipei, Taiwan, 2002.
- [89] MILIDIÚ, R. L.; DUARTE, J. C. ; CAVALCANTE, R.. **Machine learning algorithms for portuguese named entity recognition**. In: PROCEEDINGS OF FOURTH WORKSHOP IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, Ribeirão Preto, Brazil, 2006.

- [90] HIGGINS, D.. **A transformation-based approach to argument labeling.** In: Ng, H. T.; Riloff, E., editors, HLT-NAACL 2004 WORKSHOP: EIGHTH CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING (CONLL-2004), p. 114–117, Boston, Massachusetts, USA, May 6 - May 7 2004. Association for Computational Linguistics.
- [91] DOS SANTOS, C. N.. **Entropy Guided Transformation Learning.** PhD thesis, Pontificia Universidade Católica do Rio de Janeiro, 2008.
- [92] QUINLAN, J. R.. **C4.5: programs for machine learning.** Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [93] SU, J.; ZHANG, H.. **A fast decision tree learning algorithm.** In: PROCEEDINGS OF THE TWENTY-FIRST AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2006.
- [94] CURRAN, J. R.; WONG, R. K.. **Formalisation of transformation-based learning.** In: PROCEEDINGS OF THE ACSC, p. 51–57, Canberra, Australia, 2000.
- [95] BUCHHOLZ, S.; MARSI, E.. **Conll-x shared task on multilingual dependency parsing.** In: IN PROC. OF CONLL, p. 149–164, 2006.
- [96] KROMANN, M. T.. **The Danish dependency treebank and the underlying linguistic theory.** Växjö, Sweden, 2003.
- [97] NILSSON, J.; HALL, J. ; NIVRE, J.. **MAMBA meets TIGER: Reconstructing a Swedish treebank from antiquity.** In: PROC. OF THE NODALIDA SPECIAL SESSION ON TREEBANKS, 2005.
- [98] EINARSSON, J.. **Talbankens skriftspråkskonkordans,** 1976.
- [99] TELEMAN, U.. **Manual för grammatisk beskrivning av talad och skriven svenska (MAMBA),** 1974.
- [100] VAN DER BEEK, L.; BOUMA, G.; MALOUF, R. ; VAN NOORD, G.. **The Alpino dependency treebank.** In: COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS (CLIN). Rodopi, 2002.
- [101] VAN DER BEEK, L.; BOUMA, G.; DACIUK, J.; GAUSTAD, T.; MALOUF, R.; VAN NOORD, G.; PRINS, R. ; VILLADA, B.. **The Alpino dependency treebank.** In: ALGORITHMS FOR LINGUISTIC PROCESSING, NWO PIONIER progress report 5. 2002.

- [102] AFONSO, S.; BICK, E.; HABER, R. ; SANTOS, D.. **Floresta sintá(c)tica: a treebank for portuguese**. In: Rodríguez, M. G.; Araujo, C. P. S., editors, LREC2002, p. 1698–1703. ELRA, 2002.
- [103] ERALDO R. FERNANDES, C. N. D. S.; MILIDIÚ, R. L.. **A machine learning approach to portuguese clause identification**. In: ., editor, PROCEEDINGS OF PROPOR 2010. Porto Alegre, Brazil, 2010.

A

Corpora Description

A.1

Danish Corpus

The Danish *corpus* provides no lemma annotation.

Table A.1: Danish Part-of-Speech Tags.

Coarse-grained Part-of-speech	Description	Fine-grained Part-of-speech	Description
A	adjectives	AN	normal adjective
C	conjunction	AC	cardinals
I	interjection	AO	ordinals
N	noun	CC	coordinating conjunction
P	pronoun	CS	subordinating conjunction
RG	adverb	I	interjection
SP	preposition	NP	proper noun
U	unique	NC	common noun
V	verb	PP	personal pronoun
X	extra-linguistic units	PD	demonstrative pronoun
		PI	indefinite pronoun
		PT	interrogative/relative pronoun
		PC	reciprocal pronoun
		PO	possessive pronoun
		RG	adverb
		SP	preposition
		U	unique
		VA	main verb
		VE	medial verb
		XA	abbreviation
		XF	foreign word
		XP	punctuation
		XR	formulae
		XS	symbol
		XX	other

A.2 Dutch Corpus

In the Dutch *corpus* the fine-grained part-of-speech tag is identical to the coarse-grained part-of-speech except for multi-word units, where it is the concatenation of the coarse-grained part-of-speech tags of the words.

Table A.2: Dutch Part-of-Speech Tags.

Coarse-grained Part-of-speech	Description
Adj	adjective
Adv	adverb
Art	article
Conj	conjunctive
Int	interjection
N	noun
Num	numeral
Misc	miscellaneous
Prep	preposition
Pron	pronoun
Punc	punctuation
V	verb

A.3 Portuguese Corpus

Table A.3: Portuguese Part-of-Speech Tags.

Coarse-grained Part-of-speech	Description	Fine-Grained Part-of-speech	Description
adj	adjective	adj	adjective
adv	adverb	adv	adverb
art	article	art	article
conj	conjunction	conj-c	coordinating conjunction
in	interjection	conj-s	subordinating conjunction
n	noun	in	interjection
num	numeral	n	noun
pron	pronoun	num	numeral
prop	proper noun	pron-det	determiner pronoun
prp	preposition	pron-indp	independent pronoun
v	verb	pron-pers	personal pronoun
		prop	proper noun
		prp	preposition
		v-fin	finite verb
		v-ger	gerund verb
		v-inf	infinitive verb
		v-pcp	participle verb

B
Baseline Systems

B.1
Danish

Table B.1: Danish Baseline System for One ETL Model.

Part-of-Speech	Most Common Head
A	1_P_L
C	1_V_L
I	root
SP	1_V_L
N	1_SP_L
P	1_V_L
RG	1_V_L
U	1_V_R
V	root
X	1_V_L

Table B.2: Danish Baseline Systems for 1st Subtask.

Part-of-speech	Head Side
A	L
C	L
I	root
N	L
P	L
RG	L
SP	L
U	L
V	L
X	L

Table B.3: Danish Baseline Systems for 2nd Subtask.

Part-of-Speech	<u>Predicted Head-Side</u>		
	Root	L	R
A	root	P	N
C	root	V	V
I	root	V	V
N	root	SP	V
P	root	V	V
RG	root	V	V
SP	root	V	V
U	root	SP	V
V	root	V	V
X	root	V	V

Table B.4: Danish Baseline Systems for 3rd Subtask.

Part-of-Speech	<u>Predicted Head-Side</u>		
	Root	L	R
A	root	1	1
C	root	1	2
I	root	1	1
N	root	1	1
P	root	1	1
RG	root	1	1
SP	root	1	1
U	root	1	1
V	root	1	1
X	root	1	1

B.2 Dutch

Table B.5: Dutch Baseline System for One ETL Model.

Part-of-Speech	Most Common Head
Adv	1_V_R
Punc	1_N_L
MWU	1_Prep_L
Int	root
Misc	1_N_L
N	1_Prep_L
Pron	1_V_R
Num	1_N_R
V	root
Art	1_N_R
Conj	1_V_L
Adj	1_N_R
Prep	1_N_L

Table B.6: Dutch Baseline System for 1st Subtask.

Part-of-speech	Head Side
Adv	R
Adj	R
Art	R
Conj	L
Int	root
Misc	L
MWU	L
N	L
Num	R
Prep	L
Pron	R
Punc	L
V	L

Table B.7: Dutch Baseline System for 2nd Subtask.

Part-of-Speech	<u>Predicted Head-Side</u>		
	Root	L	R
Adv	root	V	N
Adj	root	V	V
Art	root	V	N
Conj	root	V	V
Int	root	Prep	V
Misc	root	N	V
MWU	root	Prep	V
N	root	Prep	V
Num	root	Prep	N
Prep	root	N	V
Pron	root	V	V
Punc	root	N	V
V	root	V	Conj

Table B.8: Dutch Baseline System for 3rd Subtask.

Part-of-Speech	<u>Predicted Head-Side</u>		
	Root	L	R
Adv	root	1	1
Adj	root	1	1
Art	root	1	1
Conj	root	1	1
Int	root	1	1
Misc	root	1	1
MWU	root	1	1
N	root	1	1
Num	root	1	1
Prep	root	1	1
Pron	root	1	1
Punc	root	1	1
V	root	1	1

B.3 Portuguese

Table B.9: Portuguese Baseline System for One ETL Model.

Part-of-Speech	Most Common Head
adv	1_v_R
pp	1_v_L
art	1_n_R
in	1_v_R
ec	1_n_R
prop	1_prp_L
vp	1_n_L
pron	1_v_R
num	1_n_R
prp	1_n_L
v	root
punc	1_v_L
conj	1_v_R
n	1_prp_L
adj	1_n_L

Table B.10: Portuguese Baseline Systems for 1st Subtask.

Part-of-speech	Head Side
adv	R
pp	L
art	R
in	L
ec	R
prop	L
vp	L
pron	R
num	R
prp	L
v	L
punc	L
conj	L
n	L
adj	L

Table B.11: Portuguese Baseline Systems for 2nd Subtask.

Part-of-Speech	<u>Predicted Head-Side¹</u>		
	Root	L	R
adj	root	n	n
adv	root	v	v
art	-	prp	n
conj	root	v	v
ec	-	prp	n
in	root	v	v
n	root	prp	v
num	root	prp	n
pp	root	v	v
pron	root	v	v
prop	root	prp	v
prp	root	n	v
punc	-	v	v
v	root	v	v
vp	-	n	-

Table B.12: Portuguese Baseline Systems for 3rd Subtask.

Part-of-Speech	<u>Predicted Head-Side²</u>		
	Root	L	R
adj	root	1	1
adv	root	1	1
art	-	1	1
conj	root	1	1
ec	-	1	1
in	root	1	1
n	root	1	1
num	root	1	1
pp	root	1	1
pron	root	1	1
prop	root	1	1
prp	root	1	1
punc	-	1	1
v	root	1	1
vp	-	1	1

The values marked with a “-” are the ones that do not appear in the training set, therefore their values are not automatically predictable. Since they do not appear in the development or test set either, no default value was needed. However, in those cases it is easy to infer that *root* would make sense to choose as default value.