
Contents

Introduction: The Handbook of Linguistic Annotation	1
Nancy Ide	
Part I The Science of Annotation	
Designing Annotation Schemes: From Theory to Model	21
James Pustejovsky, Harry Bunt and Annie Zaenen	
Designing Annotation Schemes: From Model to Representation	73
Nancy Ide, Christian Chiarcos, Manfred Stede and Steve Cassidy	
Community Standards for Linguistically-Annotated Resources	113
Nancy Ide, Nicoletta Calzolari, Judith Eckle-Kohler, Dafydd Gibbon, Sebastian Hellmann, Kiyong Lee, Joakim Nivre and Laurent Romary	
Overview of Annotation Creation: Processes and Tools	167
Mark A. Finlayson and Tomaž Erjavec	
The Evolution of Text Annotation Frameworks	193
Graham Wilcock	
Tools for Multimodal Annotation	209
Steve Cassidy and Thomas Schmidt	
Collaborative Web-Based Tools for Multi-layer Text Annotation	229
Chris Biemann, Kalina Bontcheva, Richard Eckart de Castilho, Iryna Gurevych and Seid Muhie Yimam	
Iterative Enhancement	257
Markus Dickinson and Dan Tufiş	
Crowdsourcing	277
Massimo Poesio, Jon Chamberlain and Udo Kruschwitz	

Inter-annotator Agreement	297
Ron Artstein	
Ongoing Efforts: Toward Behaviour-Based Corpus Evaluation	315
Takenobu Tokunaga	
Machine Learning for Higher-Level Linguistic Tasks	333
Anna Rumshisky and Amber Stubbs	
Sustainable Development and Refinement of Complex Linguistic Annotations at Scale	353
Dan Flickinger, Stephan Oepen and Emily M. Bender	
Linguistic Annotation in/for Corpus Linguistics	379
Stefan Th. Gries and Andrea L. Berez	
Developing Linguistic Theories Using Annotated Corpora	411
Marie-Catherine de Marneffe and Christopher Potts	
 Part II Case Studies	
MULTEXT-East	441
Tomaž Erjavec	
The Groningen Meaning Bank	463
Johan Bos, Valerio Basile, Kilian Evang, Noortje J. Venhuizen and Johannes Bjerva	
Case Study: The Manually Annotated Sub-Corpus	497
Nancy Ide	
OntoNotes: Large Scale Multi-layer, Multi-lingual, Distributed Annotation	521
Sameer Pradhan and Lance Ramshaw	
Prague Dependency Treebank	555
Jan Hajič, Eva Hajičová, Marie Mikulová and Jiří Mírovský	
German Treebanks: TIGER and TüBa-D/Z	595
Stefanie Dipper and Sandra Kübler	
Sinica Treebank	641
Chu-Ren Huang and Keh-Jiann Chen	
The Hindi/Urdu Treebank Project	659
Riyaz Ahmad Bhat, Rajesh Bhatt, Annahita Farudi, Prescott Klassen, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, Ashwini Vaidya, Sri Ramagurumurthy Vishnu and Fei Xia	

Semantic Annotation of MASC	699
Collin Baker, Christiane Fellbaum and Rebecca J. Passonneau	
VerbNet/OntoNotes-Based Sense Annotation	719
Meredith Green, Orin Hargraves, Claire Bonial, Jinying Chen, Lindsay Clark and Martha Palmer	
Current Directions in English and Arabic PropBank	737
Claire Bonial, Kathryn Conger, Jena D. Hwang, Aous Mansouri, Yahya Aseri, Julia Bonn, Timothy O’Gorman and Martha Palmer	
FrameNet: Frame Semantic Annotation in Practice	771
Collin F. Baker	
MPQA Opinion Corpus	813
Theresa Wilson, Janyce Wiebe and Claire Cardie	
The JDPa Sentiment Corpus for the Automotive Domain	833
Jason S. Kessler and Nicolas Nicolov	
Czech Named Entity Corpus	855
Jana Straková, Milan Straka, Magda Ševčíková and Zdeněk Žabokrtský	
Crowdsourcing Named Entity Recognition and Entity Linking Corpora	875
Kalina Bontcheva, Leon Derczynski and Ian Roberts	
Case Study: Chemistry	893
Colin Batchelor, Peter Corbett and Simone Teufel	
Building FactBank or How to Annotate Event Factuality One Step at a Time	905
Roser Saurí	
ISO-TimeML and the Annotation of Temporal Information	941
James Pustejovsky	
It-TimeML and the Ita-TimeBank: Language Specific Adaptations for Temporal Annotation	969
Tommaso Caselli and Rachele Sprugnoli	
ISO-Space: Annotating Static and Dynamic Spatial Information	989
James Pustejovsky	
Spatial Role Labeling Annotation Scheme	1025
Parisa Kordjamshidi, Martijn van Otterlo and Marie-Francine Moens	

VU Amsterdam Metaphor Corpus	1053
Tina Krennmayr and Gerard Steen	
Annotation of Linguistic and Conceptual Metaphor	1073
Ekaterina Shutova	
FATE: Annotating a Textual Entailment Corpus with FrameNet	1101
Aljoscha Burchardt and Marco Pennacchiotti	
The Recognizing Textual Entailment Challenges:	
Datasets and Methodologies	1119
Luisa Bentivogli, Ido Dagan and Bernardo Magnini	
Phrase Detectives	1149
Massimo Poesio, Jon Chamberlain and Udo Kruschwitz	
NAIST Text Corpus: Annotating Predicate-Argument	
and Coreference Relations in Japanese	1177
Ryu Iida, Mamoru Komachi, Naoya Inoue, Kentaro Inui and Yuji Matsumoto	
The Penn Discourse Treebank: An Annotated Corpus of	
Discourse Relations	1197
Rashmi Prasad, Bonnie Webber and Aravind Joshi	
Pair Annotation as a Novel Annotation Procedure:	
The Case of Turkish Discourse Bank	1219
İşin Demirşahin and Deniz Zeyrek	
ANNODIS and Related Projects: Case Studies	
on the Annotation of Discourse Structure	1241
Nicholas Asher, Philippe Muller, Myriam Bras, Lydia Mai Ho-Dac, Farah Benamara, Stergos Afantenos and Laure Vieu	
NICT Kyoto Dialogue Corpus	1265
Kiyonori Ohtake and Etsuo Mizukami	
Case Study: The AusTalk Corpus	1287
Steve Cassidy, Dominique Estival and Felicity Cox	
Annotations in the Nordic Dialect Corpus	1303
Janne Bondi Johannessen	
The Corpus of Interactional Data: A Large Multimodal	
Annotated Resource	1323
Philippe Blache, Roxane Bertrand, Gaëlle Ferré, Berthille Pallaud, Laurent Prévot and Stéphane Rauzy	

**Annotating the Clinical Text – MiPACQ, ShARe, SHARPN
and THYME Corpora** 1357
Guergana Savova, Sameer Pradhan, Martha Palmer, Will Styler,
Wendy Chapman and Noémie Elhadad

**The Colorado Richly Annotated Full Text (CRAFT) Corpus:
Multi-Model Annotation in the Biomedical Domain** 1379
K. Bretonnel Cohen, Karin Verspoor, Karën Fort,
Christopher Funk, Michael Bada, Martha Palmer
and Lawrence E. Hunter

The GENIA Corpus: Annotation Levels and Applications 1395
Paul Thompson, Sophia Ananiadou and Jun’ichi Tsujii

De-identification of Medical Records Through Annotation 1433
Amber Stubbs and Özlem Uzuner