
Contents

1	Introduction	1
1.1	Aims of our Work	4
1.2	Structure of the Book	4
2	Theoretical Fundamentals	5
2.1	Anaphora and Coreference	5
2.2	Prague Tectogrammatics	7
2.3	Coreference in Tectogrammatics	9
2.4	Types of Expressions	12
2.4.1	Central Pronouns	12
2.4.2	Relative Pronouns	15
2.4.3	Zeros	16
2.4.4	Other Expressions	18
2.4.5	Delimiting the Mentions	18
3	Related Work	23
3.1	Monolingual Coreference Resolution	23
3.1.1	A Historical Overview of Supervised Coreference Resolution	23
3.1.2	Mention-pair and Mention-ranking Models	24
3.1.3	Treatment of Non-anaphoric Mentions	26
3.1.4	Specialized Models	26
3.1.5	Coreference Resolution in Czech	27
3.2	Cross-lingual Approaches to Coreference	
	Resolution	28
3.2.1	Coreference Projection	29
3.2.2	Delexicalized Approaches	32

3.2.3	Multilingually Informed and Joint Multilingual Resolution	32
4	Data Sources, Tools and Evaluation	35
4.1	Data Resources	35
4.1.1	Prague Dependency Treebank	38
4.1.2	Prague Czech-English Dependency Treebank	38
4.1.3	CzEng	41
4.1.4	CoNLL 2012 Test Set	42
4.2	Treex Pre-processing Pipeline	42
4.2.1	Czech and English Analysis	43
4.2.2	Monolingual Alignment	47
4.2.3	Original Cross-lingual Alignment	50
4.3	Coreference Systems to Compare	51
4.3.1	CzEng CR	52
4.3.2	Stanford CR	53
4.4	Coreference Evaluation Measures	54
4.4.1	Standard Measures	55
4.4.2	Addressing the Issues of Standard Measures	55
4.4.3	Prague Anaphora Score	57
5	Analysis of the Parallel Data	61
5.1	English Central Pronouns	62
5.2	Czech Central Pronouns	66
5.3	Czech Relative Pronouns	68
5.4	English Relative Pronouns	71
5.5	English Anaphoric Zeros	73
5.6	Czech Anaphoric Zeros	74
5.7	Summary	75
6	Cross-lingual Alignment of Coreferential Expressions	77
6.1	Manual Alignment	78
6.2	Supervised Alignment	79
6.2.1	Design of the Aligner	79

6.2.2	Evaluation	81
6.3	Summary	85
7	Adding Cross-lingual Features to Coreference Resolution	87
7.1	Treex Coreference Resolver	88
7.1.1	Tectogrammatical Analysis	90
7.1.2	System Design	90
7.1.3	Feature Sets	91
7.1.4	Cross-lingual Extension	93
7.2	Monolingual Resolution	95
7.2.1	Overall Evaluation Results	96
7.2.2	Fine-grained Evaluation Results on Czech	97
7.2.3	Fine-grained Evaluation Results on English	98
7.2.4	Learning Curves	100
7.3	Bilingually Informed Resolution	101
7.3.1	Bilingually Informed vs. Monolingual	101
7.3.2	Contribution of Cross-lingual Feature Sets	103
7.3.3	Alignment and Aligned Coreference Oracles	104
7.4	Comparative Analysis of Mono CR and BI CR	105
7.4.1	Quantitative Analysis	105
7.4.2	Qualitative Analysis	108
7.5	Summary	111
8	Coreference Projection	113
8.1	Projection Mechanism	114
8.2	Gold Projections	115
8.2.1	Error Analysis	118
8.2.2	Effect of Alignment Quality	123
8.3	Resolver Trained on Projected Gold Coreference	124
8.3.1	Projected vs. Monolingual Coreference	125
8.4	Summary	126
9	Conclusion	129
	Summary	131

List of Figures	133
List of Tables	135
A Diagnostics in Prague Anaphora Score	137
B Distributions of Coreferential Expressions and Their Counterparts	139
B.1 Distributions of Coreferential Expressions	139
B.2 Distributions of Expressions' Counterparts	142
C Learning Curves	145
Bibliography	149