

# Contents

<b>Preface</b>	v
<b>List of figures</b>	ix
<b>1 Introduction</b>	1
<b>2 Motivation</b>	7
<b>3 Data</b>	17
3.1 Data creation . . . . .	18
3.1.1 Research domain . . . . .	18
3.1.2 Research question . . . . .	22
3.1.3 Variable selection . . . . .	24
3.1.4 Variable value assignment . . . . .	25
3.1.5 Data representation . . . . .	26
3.1.6 Data validation . . . . .	27
3.2 Data geometry . . . . .	28
3.2.1 Space . . . . .	28
3.2.2 Data transformation . . . . .	47
<b>4 Cluster</b>	153
4.1 Cluster definition . . . . .	153
4.2 Clustering methods . . . . .	156
4.2.1 Nonhierarchical clustering methods . . . . .	158
4.2.2 Hierarchical clustering . . . . .	201
4.3 Cluster validation . . . . .	224
4.3.1 Clustering tendency . . . . .	227
4.3.2 Validation . . . . .	232
<b>5 Hypothesis generation</b>	251
5.1 Cluster analysis of MDECTE . . . . .	251
5.1.1 Projection methods . . . . .	252
5.1.2 Nonhierarchical partitional methods . . . . .	255
5.1.3 Hierarchical partitional methods . . . . .	256

5.1.4	Comparison of results . . . . .	261
5.2	Hypothesis formulation . . . . .	270
<b>6</b>	<b>Literature Review</b>	<b>277</b>
6.1	Scope . . . . .	277
6.2	Context . . . . .	278
6.3	Cluster analysis in corpus linguistics . . . . .	285
6.3.1	Cluster analysis in grammatical research . . . . .	285
6.3.2	Cluster analysis in chronological variation research . . . . .	290
6.3.3	Cluster analysis in geographical variation research . . . . .	292
6.3.4	Cluster analysis in social variation research . . . . .	299
<b>7</b>	<b>Conclusion</b>	<b>301</b>
<b>8</b>	<b>Appendix</b>	<b>303</b>
8.1	Cluster analysis facilities in general-purpose statistical packages . . . . .	303
8.1.1	Commercial . . . . .	303
8.1.2	Freeware . . . . .	304
8.2	Cluster analysis-specific software . . . . .	304
8.2.1	Commercial . . . . .	304
8.2.2	Freeware . . . . .	305
8.3	Programming languages . . . . .	306
8.4	Corpus linguistic-specific applications using clustering . . . . .	309
	<b>References</b>	<b>310</b>
	<b>Subject index</b>	<b>359</b>