

Inhaltsverzeichnis

1	Computerlinguistik – Was ist das?	1
1.1	Aspekte der Computerlinguistik	1
1.1.1	Computerlinguistik: Die Wissenschaft	2
1.1.2	Computerlinguistik und ihre Nachbardisziplinen	3
1.1.3	Teilbereiche der Computerlinguistik	6
1.1.4	Theoretische Computerlinguistik	8
1.1.5	Wissensbereiche	11
1.1.6	Industrielle Anwendungen	14
1.1.7	Berufsfelder für Computerlinguisten	16
1.1.8	Literaturhinweise	17
1.2	Zur Geschichte der Computerlinguistik	18
1.2.1	Die Ursprünge	18
1.2.2	Symbolische Sprachverarbeitung	19
1.2.3	Korpusstatistische Verfahren	21
1.2.4	Anwendungen der Computerlinguistik	23
2	Formale Grundlagen	27
2.1	Mengenlehre und Logik	28
2.1.1	Mengenlehre	28
2.1.2	Aussagenlogik	33
2.1.3	Prädikatenlogik	45
2.1.4	Typenlogik	53
2.1.5	Der Lambda-Kalkül	60
2.1.6	Literaturhinweise	65
2.2	Automatentheorie und Formale Sprachen	66
2.2.1	Grundlegende Definitionen	66
2.2.2	Grammatiken	67
2.2.3	Endliche Automaten, einseitig-lineare Grammatiken und reguläre Sprachen	70
2.2.4	Kontextfreie Sprachen und Grammatiken	79
2.2.5	Nicht-kontextfreie Sprachen und Grammatiken	84
2.2.6	Komplexitäts- und Entscheidbarkeitseigenschaften	90
2.2.7	Zusammenfassung	92
2.2.8	Literaturhinweise	93
2.3	Graphentheorie und Merkmalsstrukturen	94
2.3.1	Graphen und Bäume	94
2.3.2	Merkmalsstrukturen	97
2.3.3	Unifikation	103

2.3.4	Generalisierung	106
2.3.5	Typisierte Merkmalsstrukturen	108
2.3.6	Literaturhinweise	113
2.4	Statistische Grundlagen	114
2.4.1	Wahrscheinlichkeitstheoretische Grundlagen	114
2.4.2	Hidden-Markov-Modelle	130
2.4.3	Evaluation und Optimierung statistischer Modelle	147
2.4.4	Literaturhinweise	157
2.5	Texttechnologische Grundlagen	159
2.5.1	HTML – Hypertext Markup Language	160
2.5.2	XML – Extensible Markup Language	161
2.5.3	Verarbeitung XML-annotierter Daten	163
2.5.4	Texttechnologie und Computerlinguistik	167
2.5.5	Literaturhinweise	168
3	Methoden	169
3.1	Phonetik und Phonologie	170
3.1.1	Grundlagen der Computerphonologie	172
3.1.2	Empirische Methoden	190
3.1.3	Formale Methoden	197
3.1.4	Zusammenfassung und weitergehende Lektüre	211
3.2	Verarbeitung gesprochener Sprache	214
3.2.1	Spracherkennung	215
3.2.2	Sprachsynthese	223
3.2.3	Gemeinsamkeiten und Unterschiede	229
3.2.4	Literaturhinweise	235
3.3	Morphologie	236
3.3.1	Überblick	236
3.3.2	Grundbegriffe und -probleme	236
3.3.3	Modelle aus der Generativen Linguistik	240
3.3.4	Morphologie mit endlichen Automaten	244
3.3.5	Default-Vererbungsnetze: DATR	251
3.3.6	Erweiterte Finite-State-Ansätze	257
3.3.7	Morphologie und generative Kapazität	262
3.3.8	Zusammenfassung und Ausblick	263
3.3.9	Literaturhinweise	263
3.4	Flache Satzverarbeitung	264
3.4.1	Tokenisierung	264
3.4.2	Wortart-Tagging	271
3.4.3	Chunk-Parsing	275
3.4.4	Literaturhinweise	278
3.5	Syntax und Parsing	280
3.5.1	Syntax	281
3.5.2	Parsing	303
3.5.3	Literaturhinweise	328

3.6	Semantik	330
3.6.1	Grundlagen der natürlichsprachlichen Semantik	332
3.6.2	Formale Semantik	339
3.6.3	Diskursrepräsentationstheorie	359
3.6.4	Ansätze zur Unterspezifikation	371
3.6.5	Lexikalische Semantik	377
3.6.6	Literaturhinweise	393
3.7	Pragmatik	394
3.7.1	Text, Diskurs und Dialog	395
3.7.2	Anaphernresolution	399
3.7.3	Implikaturen und Präsuppositionen	410
3.7.4	Benutzermodellierung	422
3.8	Textgenerierung	436
3.8.1	Aufgaben der Planung und Umsetzung	437
3.8.2	Funktionalität des Planungsprozesses	439
3.8.3	Methoden zur Diskursplanung	446
3.8.4	Satzplanungsverfahren	453
3.8.5	Verfahren zur Oberflächenrealisierung	458
3.8.6	Linguistische Theorien zur Generierung	461
3.8.7	Ausblick	464
3.8.8	Literaturhinweise	465
3.9	Programmiersprachen in der Computerlinguistik	466
3.9.1	Die Anfänge: Hochsprachen und symbolische Sprachverarbeitung	466
3.9.2	C/C++	469
3.9.3	Programmierarchitekturen: Java und .Net	472
3.9.4	Dynamische Sprachen: Perl und Python	475
3.9.5	Von der Desktop- zur Web-Applikation	479
4	Ressourcen	481
4.1	Korpora	482
4.1.1	Aufbau eines Korpus	483
4.1.2	Typologie	486
4.1.3	Anwendungen	489
4.1.4	Weiterführende Informationen	491
4.2	Baumbanken	492
4.2.1	Zentrale Eigenschaften	492
4.2.2	Die wichtigsten Baumbanken	496
4.2.3	Suche in Baumbanken	502
4.2.4	Literaturhinweise	503
4.3	Lexikalisch-semantische Ressourcen	504
4.3.1	Lexikalisch-semantische Wortnetze	504
4.3.2	FrameNet	511
4.3.3	Literaturhinweise	514
4.4	Lexika für multimodale Systeme	515
4.4.1	Grundlagen	515

4.4.2	Die Lexikographie	517
4.4.3	Lexikalische Struktur- und Informationstypen	520
4.4.4	Literaturhinweise	523
4.5	Sprachdatenbanken	524
4.5.1	Definition	524
4.5.2	Primärdaten	524
4.5.3	Sekundärdaten	526
4.5.4	Tertiärdaten	528
4.5.5	Software	529
4.5.6	Anwendungsgebiete	530
4.5.7	Literaturhinweise	531
4.6	Nicht-sprachliches Wissen	532
4.6.1	Die Relevanz nicht-sprachlichen Wissens für die CL	532
4.6.2	Was ist „Wissen“ (nicht)?	533
4.6.3	Wissen und Wissensrepräsentation	533
4.6.4	Aspekte der Wissensrepräsentation	534
4.6.5	Wissensrepräsentation für die CL	541
4.6.6	Literaturhinweise	543
4.7	Das World Wide Web als computerlinguistische Ressource	544
4.7.1	Einleitung	544
4.7.2	Aspekte des Web als Korpus	544
4.7.3	Sozio-Semantisches Web	546
4.7.4	Sprachverarbeitungsanwendungen mit Nutzung des World Wide Web als Ressource	550
4.7.5	Computerlinguistik und Sprachtechnologie für das Web	550
4.7.6	Literaturhinweise	551
5	Anwendungen	553
5.1	Korrektursysteme	555
5.1.1	Korrektur von Nichtwörtern	556
5.1.2	Kontextabhängige Korrektur	559
5.1.3	Rechtschreibkorrektur für Suchmaschinen	561
5.1.4	Grammatikkorrektur	562
5.1.5	Perspektiven	563
5.1.6	Literaturhinweise	564
5.2	Computergestützte Lexikographie und Terminologie	566
5.2.1	Lexikographie und Terminologie	566
5.2.2	Die Teilbereiche im Überblick	567
5.2.3	Akquisition von lexikalischem Wissen	568
5.2.4	Verwaltung und Repräsentation lexikalischen Wissens	571
5.2.5	Nutzung von lexikalischem Wissen	573
5.2.6	Computerlinguistische Unterstützung lexikographischer Arbeit	574
5.2.7	Literaturhinweise	575
5.3	Text-basiertes Informationsmanagement	576
5.3.1	Überblick	576

5.3.2	Information Retrieval	587
5.3.3	Informationsextraktion	594
5.3.4	Domänenoffene Fragebeantwortung	606
5.3.5	Textzusammenfassung	611
5.3.6	Multilinguales und sprachübergreifendes TIM	613
5.3.7	Perspektiven	614
5.3.8	Literaturhinweise	615
5.4	Sprachein- und -ausgabe	616
5.4.1	Spracheingabe	616
5.4.2	Sprachausgabe	621
5.4.3	Literaturhinweise	623
5.5	(Multimodale) Dialogsysteme	624
5.5.1	Multimodale Kommunikation	624
5.5.2	Sprachdialogsysteme	626
5.5.3	Struktur eines multimodalen Dialogsystems	628
5.5.4	Modellierung und Repräsentation	631
5.5.5	Literaturhinweise	632
5.6	Angewandte natürlichsprachliche Generierungs- und Auskunftssysteme	633
5.6.1	Was ist angewandte NLG?	634
5.6.2	Beispiele für angewandte NLG-Systeme	635
5.6.3	Mechanismen und Methoden	637
5.6.4	Perspektiven	640
5.6.5	Literaturhinweise	641
5.7	Maschinelle und computergestützte Übersetzung	642
5.7.1	Einleitung	642
5.7.2	MÜ-Ansätze	644
5.7.3	Regel-basierte Systeme	645
5.7.4	Statistische Maschinelle Übersetzung	647
5.7.5	Evaluation von MÜ-Systemen	653
5.7.6	Computergestützte Übersetzung – CAT	654
5.7.7	Aktueller Stand und Perspektiven	656
5.7.8	Literaturhinweise	657
6	Evaluation von sprachverstehenden und -generierenden Systemen	659
6.1	Einführung	659
6.1.1	Warum wird evaluiert?	659
6.1.2	Wann und wie wird evaluiert?	660
6.1.3	Was wird evaluiert?	662
6.2	Evaluationskriterien für sprachverarbeitende Systeme	664
6.2.1	Spracherkennungssysteme	664
6.2.2	Evaluation von Dialogsystemen	669
6.2.3	Informationssuchsysteme	674
6.2.4	Sprachsynthesysteme	674
6.2.5	Maschinelle Übersetzung	675
6.2.6	Fazit	677

6.2.7 Literaturhinweise	678
Literaturverzeichnis	679
Index	717
Die Autorinnen und Autoren	734