

Gliederung

Vorwort		VIII
Einführung		1
1	Die Stellung der automatischen Lemmatisierung im Rahmen der Projekte des SFB 100	3
2	Allgemeine Probleme der Lemmatisierung	3
3	Wörterbücher als Grundlage der Lemmatisierung	5
4	Die Basisgrammatik	6
4.1	Das Regelsystem	6
4.2	Inhaltliche Beschreibung der Regeln	7
5	Einstweilige Grenzen des Lemmatisierungsverfahrens	13
I	Überblick über das Lemmatisierungsverfahren	15
I.1	Aufgabe	15
I.2	Definitionen	15
I.3	Probleme der automatischen Lemmatisierung	20
I.4	Grundlagen der automatischen Lemmatisierung	22
I.4.1	Das morpho-syntaktische Wörterbuch SADAW	23
I.4.2	Die Rolle der syntaktischen Analyse	25
I.5	Grenzen des Verfahrens	25
I.5.1	Technische Beschränkungen	25
I.5.2	Strukturelle Beschränkungen	26
I.6	Lösungsverfahren	27
I.6.1	Die Operatoren	28
I.6.2	Datenflußplan	32
I.6.3	Analysewortklassen	34
I.6.4	Inhalte der Analysespeicher	40
I.6.4.1	TX-Record	40
I.6.4.2	Segmentspeicher (AE-Speicher)	44
I.6.4.3	NOG-Record	45
I.6.4.4	Prädikatrecord	47
I.6.4.5	Der Subsatz-Speicher	48
I.7	Anwendungsfälle	49
I.8	Möglichkeiten der Weiterentwicklung	50
I.9	Auswertung eines Testlaufs	51
II	Das Wörterbuch SADAW (Saarbrücker Deutsches Analyse-Wörterbuch)	56
II.1	Aufgabe	56
II.2	Lexikoninhalt	56
II.2.1	Informationen zu Verbstämmen	57
II.2.2	Informationen zu Substantivstämmen	67
II.2.3	Informationen zu Adjektivstämmen	69

II.3	Liste zur Verschlüsselung bestimmter lexikalischer Informationen	72
II.3.1	Valenzliste	72
II.3.2	Liste der Verbzusatzklassen	73
II.3.3	Liste der Verbsimplexnummern	74
III	Texteingabe (Operator LESEN)	75
III.1	Aufgabe	75
III.2	Lösungsverfahren	75
IV	Wörterbuchsuche und morphologische Analyse (Operator WOBUDEU)	78
IV.1	Aufgabe	78
IV.1.1	Morphologische Analyse	78
IV.1.2	Nicht im Lexikon vorhandene Komposita und Ableitungen	79
IV.1.3	Feste Sequenzen	80
IV.2	Anwendungsgrenzen	80
IV.3	Lösungsverfahren	81
IV.3.1	Flexionsanalyse	81
IV.3.1.1	Wortformen	82
IV.3.1.2	Verbstämme	82
IV.3.1.3	Substantivstämme	83
IV.3.1.4	Behandlung der Adjektivstämme	83
IV.3.2	Zerlegung von Komposita und Ableitungen	83
IV.3.3	Erkennung unbekannter Wörter	85
IV.3.4	Erkennung fester Sequenzen	86
IV.4	Beispiele	87
V	Homographen-Analyse (Operator DIHOM)	89
V.1	Aufgabe	89
V.2	Lösungsverfahren	89
V.2.1	Untersuchungen zur Distribution einzelner Lesarten	90
V.2.2	Bildung von Ketten mit eindeutigen Lesarten	93
V.3	Anwendungsgrenzen	97
V.4	Beispiele	97
VI	Satzsegmentanalyse (Operator SEGMENT)	102
VI.1	Aufgabe	102
VI.2	Satzsegmentgrammatik	103
VI.3	Anwendungsgrenzen	108
VI.4	Lösungsverfahren	108
VI.4.1	Satzsegmentierung	108
VI.4.2	Segmentinventarisierung	109
VI.4.3	Überprüfung der Kompatibilität der Segmente	112
VI.4.4	Verbzusatzanalyse	112
VI.5	Beispiele	113
VII	Nominalgruppenanalyse (Operator NOVERA)	116
VII.1	Aufgabe	116
VII.2	Nominalgruppengrammatik	116
VII.2.1	Strukturierung der nominalen Gruppen (NOG)	116

VII.2.2	Darstellung der nominalen Gruppen	122
VII.3	Anwendungsgrenzen	124
VII.4	Lösungsverfahren	128
VII.5	Verwendete Begriffe	130
VII.6	Beispiele	130
VIII	Verbalgruppen- und Komplementanalyse (Operator SYNAN)	134
VIII.1	Aufgabe	
VIII.2.1	Grammatik der Verbalgruppen	134
VIII.2.2	Valenzgrammatik	136
VIII.2.2.1	Rekonstruktion getilgter Einheiten	136
VIII.2.2.2	Passivtransformationen	136
VIII.3	Anwendungsgrenzen	137
VIII.4	Lösungsverfahren	137
VIII.4.1	Verbalgruppenanalyse	137
VIII.4.2	Komplementanalyse	137
VIII.4.3	Zulässige Verbalgruppentypen und verbale Subklassen	138
VIII.5	Beispiele	144
IX	Semantische Disambiguierung (Operator SEDAM)	148
IX.1	Aufgabe	148
IX.1.1	Reduktion syntaktischer Mehrdeutigkeit durch Kontrolle von Valenzrahmen	148
IX.1.2	Ermittlung von Verblesarten durch Kontrolle von Valenzrahmen	149
IX.1.3	Disambiguierung durch selektionale Subkategorisierung	149
IX.1.4	Die Erkennung fester Syntagmen	150
IX.2	Lösungsverfahren	150
IX.2.1	Das Wörterbuch SESAM und der Operator SEDAM	150
IX.2.2	Die Disambiguierung der Präpositionen	151
IX.2.2.1	Nomenklassifikation	152
IX.3	Anwendungsgrenzen	153
X	Lemmatisierte Wörterbücher (Operator SALOP)	155
X.1	Aufgabe	155
X.2	Anwendungsgrenzen	156
X.3	Lösungsverfahren	156
X.3.1	Lemmatisierte Indices	156
X.3.2	KWASC-Index	156
X.4	Beispiele	159
X.4.1	Lemmatisierter Index	159
X.4.2	KWASC-Index (Indices auf der Basis vollständig durchgeführter Analysen)	162
X.4.2.1	Segmentindex	163
X.4.2.2	Nominalgruppenindex	165
X.4.2.3	Verbalgruppenindex	166
	Glossar	168
	Literatur	174