

## INHALT

0.	Einleitung	9
1.	Kommunikationstechnologie und Sprache	13
1.1	Die Rolle der natürlichen Sprache in der vermittelten Kommunikation	13
1.2	Darstellungsformen von Sprache und Text	17
1.3	Informationserschließung	19
1.4	Maschinelle Sprachübersetzung (MÜ)	22
1.4.1	Die heutige Bedeutung der MÜ	22
1.4.2	Hauptprobleme der MÜ	22
1.4.3	Ablauf maschineller Übersetzungsprozesse	23
1.5	Natürlichsprachliche Kommunikation mit Computern	28
2.	Linguistische Grundlagen	31
2.1	Linguistik – eine empirische Wissenschaft	31
2.1.1	Der Korpusgedanke	31
2.1.2	Prozeduren	33
2.2	LDV als explizite Beschreibung von Sprache	35
2.2.1	Das Problem der sprachlichen Einheit	36
2.2.2	Strukturbeschreibung	43
2.3	Sprache und Wissen	47
3.	Verarbeitungstechniken	51
3.1	Algorithmen	51
3.1.1	Begriffsbestimmung	51
3.1.2	Darstellungsformen von Algorithmen	52
3.2	Sprachliche Zeichen und Codes	57
3.2.1	Erkennen von Sprache allgemein	57
3.2.2	Erkennen von Sprache durch Computer	58
3.2.3	Schriftsysteme	59
3.2.4	Binärcodes	59
3.2.5	Optische Datenerfassung durch Computer	62
3.3	Sortieren	63
3.4	Suchen und Abbilden	64

<b>3.5</b>	<b>Isolierung von sprachlichen Einheiten</b>	<b>66</b>
3.5.1	Ebene der Grapheme	66
3.5.2	Ebene der Morpheme	67
3.5.3	Ebene der Sätze und Satzteile	70
3.5.4	Ebene der Silben	71
<b>3.6</b>	<b>Darstellung von Information: Listen, Bäume, Netze</b>	<b>73</b>
3.6.1	Einfache Listen	73
3.6.2	Strukturierte Listen und Bäume	74
3.6.3	Netze	76
3.6.4	Darstellung und Verarbeitung von Listen	78
<b>4.</b>	<b>Typische Algorithmen</b>	<b>81</b>
4.1	Text erfassung und -segmentierung	81
4.1.1	Zum Begriff der Kodierung	81
4.1.2	Kodierungssysteme	82
4.1.3	Verarbeitung von Kodierungen	85
4.1.4	Segmentierung	86
4.2	Register, Index, Konkordanz	88
4.2.1	Wortformenregister	90
4.2.2	Index und Konkordanz	90
4.2.3	Problematik ausschließlich wortformenbasierter Register und Indizes	93
4.3	Morphologische Analyse	94
4.3.1	Kombination und Dekomposition	95
4.3.2	Lemmatisierung	97
4.3.2.1	Voraussetzungen	97
4.3.2.2	Verfahren	98
4.4	Deskriptive Grammatiken	100
4.4.1	Allgemeine Grundlagen	100
4.4.2	Fragment einer deskriptiven Grammatik	101
4.4.3	Nutzen deskriptiver Grammatiken	105
4.5	Parsing-Techniken	106
4.5.1	Begriffsbestimmung und allgemeine Voraussetzungen	106
4.5.2	Parser und Parsing-Strategien	107
<b>5.</b>	<b>Praktische Umsetzung von Algorithmen in Programme</b>	<b>113</b>

<b>5.1</b>	<b>Möglichkeiten der Programmiersprache PL/1</b>	<b>113</b>
<b>5.2</b>	<b>Zeichenkettenverarbeitung</b>	<b>115</b>
<b>5.2.1</b>	<b>Datenein- und -ausgabe</b>	<b>115</b>
<b>5.2.1.1</b>	<b>Zeichenorientierter Modus</b>	<b>115</b>
<b>5.2.1.2</b>	<b>Satzorientierter Modus</b>	<b>117</b>
<b>5.2.2</b>	<b>Druckgestaltung</b>	<b>119</b>
<b>5.2.3</b>	<b>Manipulation von Zeichenketten</b>	<b>123</b>
<b>5.2.3.1</b>	<b>Von PL/1 bereitgestellte Möglichkeiten</b>	<b>123</b>
<b>5.2.3.2</b>	<b>Erweiterungsmöglichkeiten</b>	<b>128</b>
<b>5.3</b>	<b>Verarbeitung von Texten</b>	<b>130</b>
<b>5.3.1</b>	<b>Segmentierung</b>	<b>131</b>
<b>5.3.2</b>	<b>Kodierung und Umkodierung</b>	<b>134</b>
<b>5.3.3</b>	<b>Register</b>	<b>138</b>
<b>5.3.4</b>	<b>Lexikonvergleich</b>	<b>142</b>
<b>5.3.5</b>	<b>Index und Konkordanz</b>	<b>146</b>
<b>5.3.5.1</b>	<b>KWIC</b>	<b>146</b>
<b>5.3.5.2</b>	<b>Wortstellenindex</b>	<b>150</b>
<b>5.3.6</b>	<b>Suche nach Zeichenketten</b>	<b>152</b>
<b>5.3.7</b>	<b>Statistische Untersuchungen</b>	<b>156</b>
<b>5.4</b>	<b>Textanalyse</b>	<b>162</b>
<b>5.4.1</b>	<b>Morphologische Analysen</b>	<b>163</b>
<b>5.4.1.1</b>	<b>Verbformenbestimmung</b>	<b>163</b>
<b>5.4.1.2</b>	<b>Substantivsegmentierung</b>	<b>168</b>
<b>5.4.1.3</b>	<b>Flexionsformengenerierung</b>	<b>173</b>
<b>5.4.2</b>	<b>Parser</b>	<b>176</b>
<b>5.5</b>	<b>Erläuterungen zu den Aufgaben aus 5.2 – 5.4</b>	<b>180</b>
<b>5.6</b>	<b>Unterschiede beim Einsatz von Microcomputer-PL/1</b>	<b>182</b>
<b>Literaturverzeichnis</b>		<b>189</b>
<b>Sachregister</b>		<b>197</b>