



## 目 次

〈言語の科学〉へのいざない	v
学習の手引き	ix
<b>1 テキスト処理</b>	<b>1</b>
1.1 コンピュータ内部での文字の扱い	3
(a) 文字コードセット	4
(b) 文字コード体系	6
1.2 文字の表示	10
(a) 文字表示の仕組み	10
(b) フォント	11
1.3 テキストの表現形式	13
(a) SGML	14
(b) HTML	16
(c) T <sub>E</sub> X	17
(d) PostScript と PDF	19
1.4 文字列の整列	21
(a) 辞書式順序	21
(b) クイックソート	23
(c) 基底法	25
(d) $N$ グラム統計と KWIC	27
1.5 文字列の探索	29
(a) 2分探索	30
(b) ハッシュ法	32
(c) トライ法	34
1.6 文字列の照合	36
(a) 完全一致	36
(b) 正規表現の照合	39
(c) 近似照合	43
1.7 テキストデータベース	46

第1章のまとめ	50
2 情報検索	51
2.1 情報検索と情報検索支援	53
2.2 情報検索システム	55
(a) 情報検索システムの枠組	55
(b) 伝統的なキーワード検索	57
(c) 検索システムの評価	61
(d) 問題点	61
2.3 検索システムの高度化	65
(a) ベクトル空間法	65
(b) 関連フィードバック法	67
(c) 検索質問の自動合成	70
(d) パッセージ検索	71
2.4 全文検索	72
(a) 単語インデックスを用いた全文検索	72
(b) 文字列インデックスを用いた全文検索	74
2.5 情報の組織化	77
(a) 選 別	77
(b) 分 類	78
(c) 重要情報の抽出	81
(d) 要 約	83
2.6 インターネットと情報検索	85
(a) ワールドワイドウェブ	85
(b) ウェブページの検索	89
2.7 展 望	91
第2章のまとめ	93
3 機械翻訳	95
3.1 機械翻訳の原理	97
(a) 人間の翻訳と機械の翻訳	97
(b) 言語の約束とその解析	98

3.2	機械翻訳の方式	101
(a)	機械翻訳の基本方式	101
(b)	構文解析型トランスファー方式	102
(c)	意味解析型トランスファー方式	104
(d)	各種の翻訳方式	107
(e)	話し言葉の翻訳	111
3.3	機械翻訳システムの構成技術	112
(a)	日本語の文法解析	113
(b)	日本語の意味解析	117
(c)	日英変換技術	120
(d)	英語生成技術	124
3.4	翻訳辞書構成技術	127
(a)	翻訳知識ベースとしての辞書	127
(b)	機械辞書の役割	128
(c)	辞書の分類と参照順序	129
(d)	文法的解析のための辞書	130
(e)	意味解析のための辞書	133
(f)	機械翻訳用の辞書の構成	134
3.5	翻訳システムの現状と適用分野	137
(a)	翻訳能力の達成レベル	137
(b)	機械翻訳の適応分野	138
(c)	機械翻訳適用の方法	139
3.6	機械翻訳を取り巻く諸問題	141
(a)	機械翻訳の基本問題	141
(b)	当面する課題	146
	第3章のまとめ	148
4	言語の統計	149
4.1	ことばの数量化	151
(a)	言語について数を数えるということ	151
(b)	言語を測るものさし	154
(c)	ものさしの条件	155
4.2	統計的操作の基礎知識	157

(a)	言語の計量的研究の歴史	157
(b)	記述統計と推測統計	157
(c)	語彙分布の法則	159
<b>4.3</b>	<b>語彙調査の方法</b>	<b>163</b>
(a)	調査の企画	163
(b)	調査の目的	164
(c)	調査の方法	165
(d)	語彙調査の手順	166
(e)	調査単位	170
(f)	同語異語の判別	172
(g)	語彙の意味分布についての分析	176
<b>4.4</b>	<b>日本語の語彙調査例</b>	<b>180</b>
(a)	雑誌九十種の語彙調査	180
(b)	新聞の語彙調査	181
(c)	高校教科書の語彙調査	184
(d)	話しことばの語彙調査	188
(e)	テレビ放送の語彙調査	188
(f)	流行歌の語彙調査	190
<b>4.5</b>	<b>語彙の対照研究</b>	<b>194</b>
(a)	『星の王子さま』6か国語版の語彙の対照	194
(b)	中国流行歌と日本語の逐語訳の比較	196
	第4章のまとめ	198
	用語解説	201
	読書案内	203
	参考文献	207
	索引	215

