

03
1552
1999
4G

日本語処理における頑健な文節区切り手法とその応用

1999年7月

鈴木 恵美子

寄贈
鈴木恵美子氏

00003644

目次

1	序論	1
1.1	研究の背景	1
1.2	研究の目的	4
1.3	本研究の工学的意義	5
1.4	研究の概要	6
2	計算機による日本語処理の諸問題	9
2.1	はじめに	9
2.2	日本語形態素解析の諸問題	9
2.2.1	形態素解析の定義	9
2.2.2	形態素解析技術	10
2.2.3	日本語形態素解析の新たなアプローチ	12
2.3	まとめ	14
3	文字列パターンを用いた日本語文自動分割	15
3.1	はじめに	15
3.2	文字列パターンを用いた日本語文節区切りアルゴリズム	16
3.3	「漢字でない文字 - 漢字 - ひらがな - ひらがな」に現れるひらがな2文字の特徴	18
3.4	文節区切り記号挿入方法について	29
3.5	文節区切り記号挿入方法の改良	33
3.6	評価および考察	35
3.7	まとめ	36
4	構造化文書を用いた日本語文書校正支援	37
4.1	はじめに	37
4.2	ワードプロセッサで作成された文書の誤り調査	38
4.3	日本語文書校正支援の必要性	39

4.4	日本語文書校正支援システムの試作	40
4.4.1	試作システムのシステム構成	40
4.4.2	校正用知識ベース	53
4.5	試作システムの校正知識	53
4.5.1	ソース表現上の校正知識	53
4.5.2	KWIC表現上の校正知識	56
4.6	校正知識の開発環境	59
4.7	日本語文書校正支援における拡張機能	62
4.7.1	重要語検出方法	62
4.7.2	拡張機能の動作	63
4.7.3	拡張機能の評価	64
4.8	試作システムの機能と評価	65
4.8.1	誤りの分類と調査	65
4.8.2	実験結果と市販のワープロ機能との評価	66
4.9	まとめ	69
5	統計的手法を用いた日本語形態素解析の曖昧さ解消方式	70
5.1	はじめに	70
5.2	形態素解析における文節の曖昧さを含む語句の分類	70
5.3	提案する曖昧さ解消方式	71
5.4	形態素解析における品詞多義の問題	73
5.5	統計的手法を用いた日本語形態素解析の曖昧さ解消方式の評価	76
5.6	形容詞の多義を減らすために提案する曖昧さ解消方式	78
5.7	まとめ	80
6	点字翻訳ボランティアのための対話型分かち書き支援手法	83
6.1	はじめに	83
6.2	分かち書きの規則と問題点	84
6.2.1	点字翻訳のための分かち書き	84
6.2.2	従来方式とその問題点	85
6.2.3	提案する手法の基本方針	86
6.3	実験システムの構成と分かち書き手法	86
6.3.1	実験システムの構成	86
6.3.2	自動分割部	87

6.3.3	対話処理部	91
6.4	評価および考察	93
6.4.1	実験方法	93
6.4.2	不正解部分に関する考察	94
6.4.3	市販の点字翻訳プログラムとの比較結果	95
6.4.4	他の対話システムとの比較	95
6.5	まとめ	96
7	結論	98
	謝辞	101
A.1	統計情報に基づく形態素解析の曖昧さ解消方式	110
A.1.1	「で」に関する調査結果	110
A.1.2	格助詞相当連語に関する調査結果	115
A.1.3	「と」に関する調査結果	119
A.1.4	「と」を受ける動詞および形容詞の引用性と随伴性調査結果	120
A.1.5	「と」に伴う各種調査結果	127
A.2	実験に使用した天声人語	131