

日英機械翻訳のための言語知識の
構築と記述に関する研究

白 井 諭

目次

第1章 序論.....	1
1.1 本研究の背景.....	1
1.2 本研究の目的.....	2
1.3 本論文の構成.....	4
第2章 日本語の階層的認識構造と係り受け解析.....	5
2.1 緒言.....	5
2.2 日本語の階層的認識構造.....	6
2.3 日本語従属節の階層的分類と順序関係.....	7
2.3.1 問題となる日本語の従属節の種類.....	7
2.3.2 従属節の基本分類.....	7
2.3.3 従属節の派生的な分類.....	9
2.3.4 引用節と連体節の扱い.....	10
2.4 従属節依存関係の決定規則.....	11
2.4.1 述語種別判定.....	11
2.4.2 係り受け判定規則.....	12
2.5 従来方式との精度比較.....	13
2.5.1 評価対象試験文.....	13
2.5.2 無作為選択方式と直近係り先方式の場合.....	13
2.5.3 係り受け解析精度の比較.....	14
2.6 結言.....	16
第3章 係り受け制約を利用した日本文書き替え.....	18
3.1 緒言.....	18
3.2 書き替えの対象.....	20
3.2.1 自動書き替えの対象範囲.....	20
3.2.2 書き替え対象表現の分類.....	21
3.2.2.1 日本語内書き替え.....	22
3.2.2.2 疑似的日本語表現への書き替え.....	23
3.3 自動書き替え方式.....	25
3.4 実験と評価.....	29
3.4.1 実験と評価の条件.....	29
3.4.2 実験結果と考察.....	29
3.5 結言.....	32

第4章 日英機械翻訳に必要な結合価パターン対.....	35
4.1 緒言.....	35
4.2 前提条件.....	37
4.2.1 パターン対記述の枠組み.....	37
4.2.2 パターン対作成の方法.....	38
4.3 パターン対収集の方法.....	40
4.3.1 和英辞書の語義分類に基づく方法.....	40
4.3.2 日本語辞書の語義分類に基づく方法.....	41
4.3.3 人の知識を内省する方法.....	42
4.4 収集された用例とパターン対の数の比較推定.....	43
4.4.1 収集された用例の数とパターン対の数の比較.....	43
4.4.2 必要なパターン数と用例数の見積もり.....	45
4.5 用例の収集方法.....	46
4.5.1 用例の網羅的収集.....	46
4.5.2 多様な表現の収集.....	47
4.5.3 収集の経過と結果.....	48
4.6 結言.....	49
第5章 連鎖型および離散型共起表現の自動抽出.....	54
5.1 緒言.....	54
5.2 従来の方法とその問題点.....	56
5.3 連鎖型共起表現の抽出法.....	58
5.3.1 重複する文字列の扱い.....	58
5.3.2 文字列抽出アルゴリズム.....	60
5.4 離散型共起表現の抽出法.....	61
5.4.1 抽出する共起文字列の条件.....	61
5.4.2 文字列抽出アルゴリズム.....	64
5.5 共起表現の抽出実験.....	66
5.5.1 連鎖型共起表現の抽出.....	66
5.5.2 離散型共起文字列の抽出.....	68
5.5.3 今後の改良と応用について.....	71
5.6 結言.....	72
第6章 結論.....	73
謝辞.....	75

付録 A 不完全な対訳データを利用する用例利用型翻訳.....	76
A.1 緒言.....	76
A.1.1 ルール型翻訳方式.....	76
A.1.2 用例型翻訳方式.....	76
A.1.3 融合型翻訳方式.....	77
A.2 用例利用型翻訳方式の提案.....	77
A.3 用例利用型翻訳の適用例.....	78
A.3.1 対訳例文の収集.....	78
A.3.2 用例候補文の抽出.....	79
A.3.3 対訳用例の選択.....	79
A.3.4 対訳用例を用いた訳文生成.....	81
A.4 結言.....	84
参考文献.....	85
執筆論文リスト.....	91
[論文誌].....	91
[国際会議].....	92
[NTT 技術誌].....	95
[研究会・シンポジウム等].....	96
[特許].....	99
[著書].....	99

目次

図2.1	日本語述節間の階層的入れ子構造と係り受け関係.....	7
図2.2	係り受け関係に着目した日本語述節の分類.....	11
図2.3	係り受け派生規則の例.....	12
図2.4	述語ごとに見た係り先の数と正解率.....	13
付図	述語間係り受け解析の例.....	17
図3.1	書き替え規則が適用される場合と適用されない場合.....	26
図3.2	日本語書き替え型翻訳方式の構成.....	27
図3.3	書き替えによる多義削減の例.....	28
図3.4	書き替え方式による訳文品質向上効果.....	31
図4.1	ALT-J/Eにおけるパターン対記述の枠組み.....	37
図4.2	ALT-J/Eにおけるパターン対作成支援の仕組み.....	38
図4.3	人間用の辞書を使用したパターン対作成の例.....	40
図4.4	3種の方法で得られたパターン対の種類の関係.....	43
図4.5	日本語用言に対するパターン対の数の分布.....	45
図5.1	抽出対象文字列の例.....	56
図5.2	抽出文字列の相互関係.....	58
図5.3	文字列採否判定の方法.....	59
図5.4	連鎖型共起表現抽出アルゴリズム実施例.....	62
図5.5	オーバラップした文字列の扱い.....	63
図5.6	離散型共起表現抽出アルゴリズム実施例.....	65

表目次

表2.1	新聞記事標{に現れた従属節(述部)の分類	8
表2.2	日経産業新聞記事における述語数の分布	13
表2.3	述語間係り受けの精度比較(標本:日経産業新聞)	15
表3.1	日本語書き替えルールの構成	25
表3.2	書き替え前後の得点変化	30
表3.3	書き替えルールの適用箇所と訳文品質向上効果	30
付表	原文自動書き替えによる訳文変化の例(1/2)	33
付表	原文自動書き替えによる訳文変化の例(2/2)	34
表4.1	IPAL語義と文型の対応	42
表4.2	収集されたパターン対の数の比較(和語動詞の例)	44
表4.3	日英機械翻訳に必要なパタン対の数とその収集に必要な用例の数の見積もり	46
表4.4	用言の種類と例文数	48
付表4.1	最終的に得られたパターン対(「上がる」の場合)	50
付表4.2	第2の方法によるパターン対作成(IPAL用例「あがる」)	51
付表4.3	第3の方法によるパターン対作成(「あがる」の全90用例の一部)	52
付表4-A	和語動詞の対訳換言例文(一部)	52
付表4-B	複合和語動詞の対訳換言例文(一部)	52
付表4-C	イ型形容詞の対訳換言例文(一部)	53
付表4-D	ナ型形容詞の対訳換言例文(一部)	53
付表4-E	サ変動詞の対訳換言例文(一部)	53
表5.1	抽出された文字列の種類と延べ度数(その1)	66
表5.2	抽出された文字列の種類と延べ度数(その2)	66
表5.3	抽出された文字列の例	67
表5.4	出現頻度の高い文字列の例	67
表5.5	抽出された文字列の組の種類と延べ度数	68
表5.6	出現頻度の高い文字列の組の例	68
表5.7	合計文字数の大きい文字列の例	68
表5.8	抽出された離散型共起\現の例	69
表5.9	合計文字数の多い離散型共起表現の例	70