

## 文章難易度に寄与する構文の自動検出システムの開発

内田聖也(甲南大学学生) 北村達也(甲南大学) 川村よし子(東京国際大学)

### 1. 研究の目的

本研究では日本語文章の難易度判定システム開発の一環として日本語学習者にとって難解な構文とされる主格省略[1]を，自然言語処理技術を使って検出するシステムを開発した。主格省略とは，文中の述語に対応する主格が省略されていることである。例えば「うどんが食べたい。」は動作主である「私は」が省略されている。本研究では文章を入力した際に，主格の有無を判定するシステムを開発した。

### 2. 主格省略の検出

本研究では構文解析システム KNP [2] の出力を基に主格省略の有無を判定する。主格省略か否かの判定は，(1) 述語に係っている形態素の列挙，(2) 主格を取りうる助詞の有無の判定，(3) 述語に係っている名詞のカテゴリに基づく主格省略の判定，という流れで行われる。以下では各ステップについて説明する。

まず，KNP の出力に基づき，入力文中の述語に係る形態素を列挙する。

次に，述語と名詞を接続し，主格を取りうる助詞を検出する。本研究では，主格を取りうる助詞 12 種類「が」，「は」，「しか」，「だけ」，「ばかり」，「くらい(ぐらい)」，「ほど」，「でも」，「まで」，「さえ」，「こそ」，「も」，を選定した。ただし，これらの助詞は主格を取る可能性があるだけで，述語と名詞がこれらの助詞で接続された場合，「うどんが食べたい」の文のように，必ずしも主格を取るわけではない。

主格を取りうる助詞が検出された場合，その直前の名詞のカテゴリを利用して主格省略の有無を判定する。カテゴリは KNP の解析結果から得られる。例えば，「うどん」なら「人工物」，「犬」なら「動物」とカテゴリ分けされる。

本研究では旧日本語能力検定試験出題基準 4 級の動詞それぞれに対し，主格になりうるカテゴリのデータベース

	人	組織 団体	動物	植物	人工物	自然物	場所	抽象物	形 複 種	色	数量	時間
知る	1	1	1	0	0	0	0	0	0	0	0	0
吸う	1	1	1	1	1	0	0	0	0	0	0	0
住む	1	1	1	1	0	0	0	0	0	0	0	0
する	1	1	1	1	1	1	0	0	0	0	0	0
座る	1	1	1	0	0	0	0	0	0	0	0	0
出す	1	1	1	1	1	1	0	0	0	0	0	0
立つ	1	1	1	1	1	1	0	0	0	0	0	0
建つ	0	0	0	0	1	0	0	0	0	0	0	0
頼む	1	1	0	0	0	0	0	0	0	0	0	0
食べる	1	1	1	1	0	0	0	0	0	0	0	0

図 2: データベースの一部

を作成した。データベースは1か0で表され、動詞に係っている名詞のカテゴリを主格として取る場合は1となっている。

主格省略の検出例を以下に示す。

入力文：あなたが貸してくれた本をなくしてしまいました。

出力結果：

述語 貸してくれた：主格あり(「あなた(人)」は「貸してくれた」の主格になる)

述語 なくしてしまいました：主格省略(主格に接続する助詞がない)

この例では述語「貸してくれた」の主格である「あなた」を正しく認識できている。また、「なくしてしまいました」は接続している助詞「を」が主格を取らないため主格省略となる。

もう1つの例を示す。

入力文：うどんが食べたい。

出力結果：

述語 食べたい：主格省略(「人工物(うどん)」は「食べる」の主格にならない)

この例では名詞が主格を取りうる助詞「が」に接続されており、述語である「食べる」に係っている。しかしこれはデータベース上では述語「食べる」はカテゴリ「人工物」を主格として取らないため、主格省略となっている。

### 3. 評価実験と考察

朝日新聞の社説より抽出した100文を使用して本システムの評価実験を行った。その結果、主格省略と判断された文章は83文あり、そのうち実際に主格が省略されていた文章は64文あった。つまり正解率は77%であった。現在は動詞データベースが旧日本語能力検定試験4級の動詞にしか対応していないため、いくつかの判別不能な文章が現れた。また、本システムは並列文章(例「社会保障と税の一体改革を議論する」)には対応していない。そのため、並列文章が多く出てくる新聞記事では正解率が低くなったと考えられる。

今後はこれらの問題点に対応していく予定である。

#### 参考文献

[1] 川村よし子, 前田ジョイス, 保原麗, 川村ヒサオ, “文章の難易度判定システム構築のための基礎調査,” ヨーロッパ日本語教育 15(印刷中).

[2] 黒橋禎夫, “結構やるな, KNP,” 情報処理 41(11)1215-1220(2000).