

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 (甲南大学知能情報学部, t-kitamu@konan-u.ac.jp)
 川村よし子 (東京国際大学言語コミュニケーション学部, kawamura@tiu.ac.jp)

1. 研究の背景と目的

言語教育においては、**学習者の学習段階に応じた**教材を用いることが重要

我々のレベル判定システム (川村ら, 2009) では単語レベルでの難易度判定を行っていた

↓ 単語レベルだけでの判定でいいの?

単語だけでなく、構文の難しさも文章難易度に関係している (川村ら, 2011)

→ **文章難易度に寄与する構文 (本研究では主格省略) の自動検出システムの開発**

2. 主格省略とは

文中の述語に対する主格が省略されていること

具体例

- うどんが食べたい。
- 貸してもらった本をなくしてしまいました。
- 先生がどこにいらっしゃるかごぞんじですか。

例では動作主である「私」などが省略されている
 このような文章の主格の有無を判定する

3. 主格省略の検出の流れ

以下の3つの文から主格省略を検出する

- 彼が食べた。
- うどんを食べる。
- うどんが食べたい。

2. 主格を取りうる助詞の有無の判定

表1: 主格を取りうる助詞

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

これらの助詞があると、主格を取りうる判定

逆にこれらの助詞が無い場合は主格を取らないと判定する

彼が食べた 彼…名詞 が…助詞 食べた…動詞	うどんが食べたい うどん…名詞 が…助詞 食べたい…動詞	うどんを食べる うどん…名詞 を…助詞 食べる…動詞
---------------------------------	---------------------------------------	-------------------------------------

主格をとりうる助詞がないので主格省略

4. 主格省略判定の結果

彼が食べた

主格をとりうる助詞あり
述語に掛かっている名詞カテゴリ動詞に対応

主格省略無し

うどんが食べた

主格をとりうる助詞あり
述語に掛かっている名詞カテゴリ動詞に非対応

主格省略あり

うどんを食べる

主格をとりうる助詞なし

主格省略あり

1. 文をKNPで解析

それぞれの文をKNP(京都大学)で構文解析を行う
品詞情報や係り受け情報が得られる

彼が 食べた	うどんが 食べたい	うどんを 食べる
彼が食べた 彼…名詞 が…助詞 食べた…動詞	うどんが食べたい うどん…名詞 が…助詞 食べたい…動詞	うどんを食べる うどん…名詞 を…助詞 食べる…動詞

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

うどんが食べたい
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の判定をする

彼が食べた 彼…名詞 ←カテゴリ-人 が…助詞 食べた…動詞	うどんが食べたい うどん…名詞 ←カテゴリ-人工物 が…助詞 食べたい…動詞
---	---

	人	組織団体	動物	植物	人工物	自然物	場所	抽象物	形模様	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: 主格カテゴリデータベース

動詞「食べる」はカテゴリ-人工物の場合、主格として取らないので主格省略
 食べるという動詞の場合、主格となりうる名詞は人、組織・団体、動物、植物である

5. 評価実験

3月1日から3月5日朝日新聞の社説より無作為に抽出した100文を使用して本システムの評価実験を行った

正解率…77%

表2: 検出結果

検出方法	主格省略	主格あり	判別不能
自動検出	83	7	10
手動	64	36	0

6. まとめと課題

まとめ

日本語学習者が学ぶ上で難しいとされる、主格省略の自動検出システムを開発した

- 主格を取りうる助詞での判定
- 述語に係っている名詞のカテゴリによる判定

課題

- 並列文に未対応
- サ変動詞に未対応
- 対応動詞の範囲の拡大

謝辞 本研究の一部は平成22年度甲南大学総合研究所および平成23年度科学研究費補助金基盤研究(B)(21320095)の支援を得て行われた