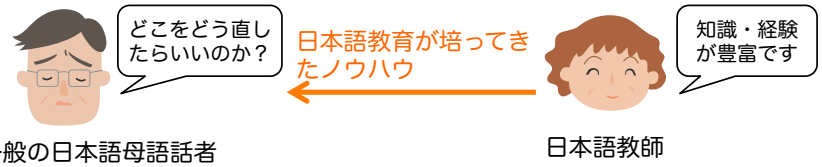
甲南大学 KONAN UNIVERSITY　　東京国際大学 TOKYO INTERNATIONAL UNIVERSITY

検索エンジンを用いた主格省略文の自動判定

中村慶太，北村達也（甲南大学知能情報学部）
川村よし子（東京国際大学言語コミュニケーション学部）

1 研究の背景と目的

日本語非母語話者への情報伝達においてそれぞれの母国語で伝えるのはコスト大
→ 彼らにも理解しやすい平易な日本語で情報伝達



文の難易度を高める要因 (川村ら2011)
→ 単語の難易度と構文の複雑さ，ゼロ格 (特に主格省略)，モダリティやアスペクトに関する補助動詞，視点の移動，慣用表現

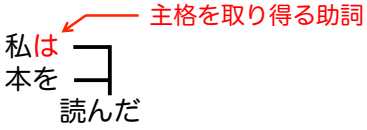
本研究：日本語教育が培ってきたノウハウを活かし主格省略を検出するシステムを開発
検索エンジンを利用した用例検索により主格の有無を判定

2 処理の流れ（一般の動態述語の場合）

Step 1 入力文を構文解析する

「私は本を読んだ」 → 構文解析システム KNP ver. 3.01 → 私は 本を 読んだ

Step 2 述語に係っている文節の助詞が主格を取り得るか調べる



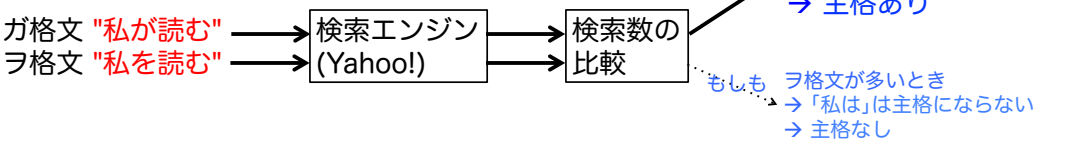
表：主格を取り得る助詞のリスト

が	しか	ばかり
ぐらい	だけ	ほど
こそ	でも	まで
さえ	は	も

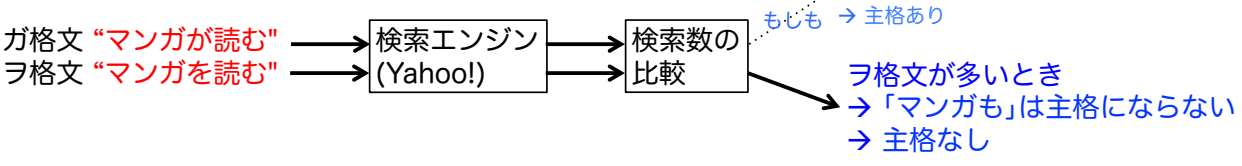
主格を取り得る助詞がない
→ 主格なし

Step 3 検索文を作り検索エンジンで検索し，主格の有無を判定する

「私は」は「読んだ」の主格になるか調べる



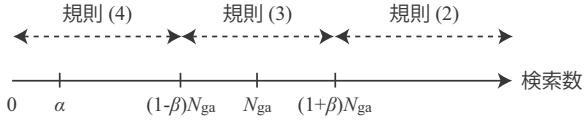
「マンガも読んだ」の場合



この比較で判断できない場合，助詞と述語の間に任意の形態素が入ることを考慮して再検索する

ガ格文 "私が*読む"，"私が**読む"
ヲ格文 "私を*読む"，"私が**読む"　※アスタリスク1個が形態素1個に対応

3 判定の詳細



図：判定基準

ガ格文の検索数をN_{ga}，ヲ格文の検索数をN_{wo}とする

規則(1) $N_{ga}<\alpha$かつ$N_{wo}<\alpha$ならば，判定不能である

規則(2) $(1+\beta)N_{ga}<N_{wo}$ならば，主格にならない

規則(3) $(1-\beta)N_{ga}\leqq N_{wo}$かつ $(1+\beta)N_{ga}\geqq N_{wo}$ならば，主格になる可能性がある

規則(4) $(1-\beta)N_{ga}>N_{wo}$ならば，主格になる

一言で言えば，検索数が少なすぎる場合や僅差の場合に対処

4 処理例

α=100，β=0.2，γ=10000で処理
(γは可能を表す動態述語などに対して使う変数)

1.「彼は東京に行った。」
→ 主格あり：［彼は］は［行った］の主格になる

2.「彼と美しい彼女は東京に行った。」
→ 主格あり：［彼と彼女は］は［行った］の主格になる

3.「英語が話せる。」
→［話せる］の主格は省略されている

4.「英語が学びたい。」
→［学びたい］の主格は省略されている

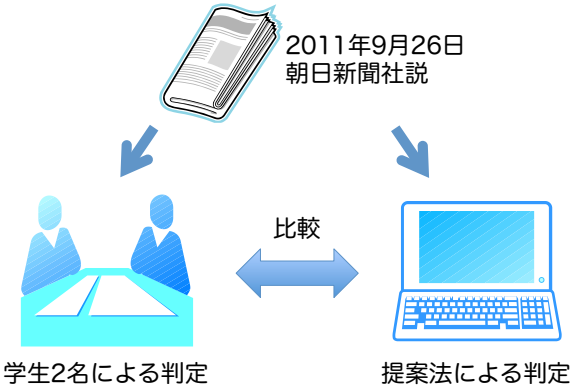
5.「英語ができる。」
→［できる］の主格は省略されている

6.「彼は英語が好きだ。」
→ 主格あり：［彼は］は［好きだ］の主格になる

7.「本が欲しい。」
→［欲しい］の主格または目的格が省略されている

8.「彼は話しながら歩いた。」
→［話しながら］の主格は［歩いた］の主格と同じ
→ 主格あり：［彼は］は［歩いた］の主格になる

5 評価実験



評価対象となった33個の述語に対して，
・人の判定と一致：28個 (85 %)
・人の判定と不一致：3個 (9 %)
・判定不能：2個 (6 %)

6 今後の課題

- より多くの文章を対象にした評価実験の実施
- 他の構文にも対応させ，構文チェッカーとして公開
- 語彙チェッカーと統合

謝辞　本研究の一部は，平成23年度科研費基盤研究(B) (21320095) および平成23年度私立大学等経常費補助金の支援を得て行われた．開発にご協力いただいた，甲南大学知能情報学部3回生小林謙太郎君に感謝します．