

システム概要

誤り検出対象: 主語・動詞の人称・数の一致に関する誤り

Flowers is very beautiful. Reading books are interesting.
 検討したい英文を入力

まず, 品詞タグを使って誤りを検出する
 品詞タグでは判断が困難な場合は検索結果数を使う

品詞タグによる誤り検出

主語の単複・人称

[NP Flowers/NNS]

→ 複数形

動詞の活用形

[VP is/VBZ]

→ 三人称単数現在

「主語の単複・人称」

「動詞の活用形」

が一致しなければ動詞が誤り

検索結果数による誤り検出

検索クエリ生成

“Reading books are”

“Reading books is”

検索結果数の比較

変更

“Reading books are”

少 15,400件

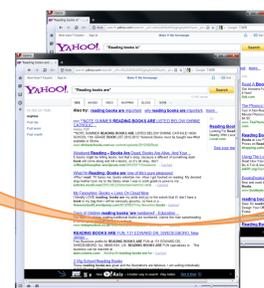
“Reading books is”

多 34,400件

検索結果数が多い方が正しい

検索結果数の取得

Web検索



(※ http://www.yahoo.com)

エラータグ付きの英文を出力

Flowers <v_agr>is</v_agr> very beautiful. Reading books <v_age>are</v_agr> interesting.

品詞タグによる誤り検出

1. 検出対象の動詞の探索

次のチャンクパターンの出現箇所を探索

① NP + VP

例: [NP Flowers/NNS] [VP is/VBZ] ...

② NP + ADVP + VP

例: [NP I/PRP] [ADVP often/RB] [VP go/VBP] ...

③ WHNP + VP

例: ... [WHNP who/WP] [VP is/VBZ] ...

- これらのパターンが出現した時, NP, VPは主語と対応する述語になる可能性がある

- 各パターンのVPの先頭の単語がVB, VBP, VBZなら誤り検出の対象とする

VB, VBP

am, are, go, live, like, listen, study
read, play, were(※) など

VBZ

is, goes, lives, likes, listens, studies
reads, plays, was(※) など

(※ was, wereはVBDであるが例外として扱う)

➤ 品詞などはStanford Parserにより取得

2. 対応する主語の探索

1. のNPは主語とは限らない → 主語を探索

① There is 構文の場合

例: There are many books.

[NP There/EX] [VP are/VBP] [NP many/JJ books/NNS]

形式上の主語

変更

[NP There/EX] [VP are/VBP] [NP many/JJ books/NNS]

② NP + PP + NP + VP の場合

例: The streets of Nagasaki were very beautiful.

[NP The/DT streets/NNS] [PP of/IN] [NP Nagasaki/NNP]

主語

主語ではない

[NP The/DT streets/NNS] [PP of/IN] [NP Nagasaki/NNP]

③ WHNP + VP の場合

例: ... someone who is ...

... [NP someone/NN] [WHNP who/WP] [VP is/VBZ] ...

主語

関係代名詞

... [NP someone/NN] [WHNP who/WP] [VP is/VBZ] ...

主語の単複・人称と

動詞の活用形が一致しない場合

動詞にエラータグを付与

3. 主語の単複・人称

主語を三人称単数とそれ以外に分類

一人称・二人称・複数

慣用句: a lot of が前後にある

名詞の列挙:

NP + , + NP + , + NP の一部である

主語に名詞 + and + 名詞 が含まれる

代名詞: he, she, it, one, 以外

限定詞: these, those

NPの最後の品詞: NNS, NNPS

三人称単数

代名詞: he, she, it, one

NPの最後の品詞: NN, NNP

➤ どちらにも当てはまらない場合は無視する

検索エンジンの検索結果数による誤り検出

主語の単複・人称の判断が困難な場合, 検索結果数を用いて検出

検索エンジンを用いる条件

動名詞の可能性がある場合, 検索する

① NPにingで終わる名詞が含まれている場合

[NP Reading/NNP books/NNS] [VP is/VBZ] ...

[NP Writing/JJ novels/NNS] [VP is/VBZ] ...

② NPの直前にVBGのみからなるVPがある場合

[VP Playing/VBG] [NP tennis/NN] [VP is/VBZ] ...

➤ 検索結果数はYahoo!検索APIを使用

http://developer.yahoo.co.jp/

検索クエリ生成

I. NPと動詞の連結

“Reading books is”

II. 動詞の活用形の変更

“Reading books are”

I と II でフレーズ検索し,
その検索結果数を比較

結果

	再現率	適合率	F値
ドライラン(提出時)	0.614	0.429	0.505
ドライラン(※)	0.855	0.542	0.664
フォーマルラン	1.000	0.590	0.724

(※ フォーマルランのシステムで評価)

今後の課題

- 検索エンジンを使用する条件の検討
- 主語が誤っている場合の対応