

係り受け解析による動詞の人称不一致に関する調査

松本 大輝[†] 奥村 紀之[†][†]香川高等専門学校 情報工学科

1 はじめに

非母語話者が書く英文には誤りが含まれることが多く、中でも動詞の人称不一致は英語を使いこなした人であっても起こしやすい誤り^[1]である。動詞の人称はそれに係る主語で決定されるため、動詞の人称不一致の検出には動詞と主語の係り受け関係を把握する必要がある。

係り受け関係は既存の構文解析器によって解析することができるが、誤りを含む文の係り受け関係を正しく解析できるかは自明ではない。本研究では、既存の構文解析器の解析結果から誤りが含まれる文の主語を特定することで動詞の人称不一致を解消可能か検証している。

2 関連研究

関連研究として、水本らは統計的機械翻訳 (Statistical Machine Translation, SMT) を用いた誤りの種類を限定しない英文の誤り訂正^[2]を試みている。SMTはコーパスベースの手法であるため、一般的にコーパスサイズを大きくすると誤りの検出や訂正の精度が向上することが知られている。しかし、動詞の人称不一致はコーパスサイズを大きくしても検出や訂正精度があまり向上しないことが示されており、SMTを用いて動詞の人称不一致の検出や訂正を行うことは困難であると考えられる。

また、Dongfengらは主格の関係代名詞の直後に続く動詞と主語の係り受け関係を同定することで動詞の人称の正誤を自動で判断する手法^[3]を提案している。提案手法では、対応テーブルと構文解析器によって得られた品詞タグを元に文中の修飾語等を削除することで文を簡素化し、加えて関係代名詞をそれが指し示す名詞に置換する処理を入力文に行っている。文の簡素化と関係代名詞の置換を個別に入力文へ適用した場合と、これらの処理を組み合わせて適用した場合のいずれも正誤判断の精度が向上している。このことから文の簡素化は動詞の人称不一致の検出に対して有効な処理であると考えられる。

3 誤りを含む文の入力時における構文解析器の解析結果の調査実験

既存の構文解析器の解析結果から動詞と主語の正しい係り受け関係が読み取れるかを検証するために、動詞の人称不一致を含む文(以下、誤り文)とその人称が正しい文(以下、正しい文)を既存の構文解析器に入力し、その解析結果の傾向を手動で検証する。

3.1 評価基準

正しい係り受け関係を読み取れるかの判断基準を2つ設ける。下記(1)、(2)の両方を満たす場合に、動詞の人称不一致を含む文において正しい係り受け関係を読み取れると判断する。判断基準を以下に示す。

(1) 正しい文と誤り文の両文において正しい係り受け関係を読み取れる

(2) 誤り文において人称不一致を起こしている動詞の品詞タグが動詞である

(1)は、3章の内容を踏まえて設定した。(2)は、品詞タグが誤っていると、仮に係り受け関係にある動詞と主語を同定できたとしても2単語間の関係が動詞と主語の係り受けであるかを判断できなくなることを踏まえて設定した。

3.2 実験に使用したツールとデータ

構文解析器はStanford-Parser¹を使用する。構文解析器への入力文は、誤り文と正しい文を1組として合計436組を使用する。誤り文は、The NICT JLE Corpus²から動詞の人称不一致タグが付与されている文を抽出し、正しい文は抽出した誤り文に付与されているタグに従って訂正したものを使用する。入力文の例を表1に示す。

表 1: 実験に使用した入力文の例

誤り文	But my daughter like animals very much .
正しい文	But my daughter likes animals very much .

An Investigation of person disagreement of verb using Semantic Analysis

[†] Daiki Matsumoto (acselroad@gmail.com)

[†] Noriyuki Okumura (okumura@di.kagawa-nct.ac.jp)

Department of Information Technology, National Institute of Technology Kagawa college (†)

¹ <http://nlp.stanford.edu/software/lex-parser.shtml>

² <https://alaginrc.nict.go.jp/nict-jle/>

誤り文については、文中に含まれる誤りの種類は動詞の人称不一致のみとし、人称が誤っている動詞が1つとなるように整形している。また、The NICT JLE Corpus は日本人の英語発話記録から構成されているため、言い淀みや言い直しが含まれている。本実験で使用した入力文は言い淀みなどを削除してある。

3.3 実験結果

表2に436組の文の解析結果を評価基準に従って集計した結果を示す。Correctは正しい係り受け関係が読み取れた文数、Incorrectは正しい係り受け関係が読み取れなかった文数、Totalは総文数を表す。動詞の種類は動詞の人称が誤っている動詞の種類を表す。

表2: 誤り文を入力した場合に構文解析器で正しい係り受けが読み取れた文数

動詞の種類	Correct	Incorrect	Total
一般動詞	221	8	229
be 動詞	207	0	207
合計	428	8	436

表1より、be動詞の人称が誤っている文が207文あり、その全てについて正しい係り受け関係が読み取れた。また、解析結果から正しい係り受け関係が読み取れないものが8文あり、そのうち3文は”like”の人称が誤っているものであった。

3.4 考察

正しい係り受け関係を読み取れなかった文の解析結果の一例を図1に示す。図1より、”like”には”IN”の品詞タグが付与されていることから、”like”が前置詞であると判断されたことがわかる。正しい係り受け関係を読み取れない8文のうち5文は、図1のように動詞の品詞判定が誤っている。これは、使用した構文解析器が確率文脈自由文法(Probabilistic Context Free Grammar, PCFG)を利用していることに起因する。PCFGは与えられた文の品詞列を利用するため、図1のように品詞タグに誤りが含まれる場合には正しい構文木を出力をできなくなると考えられる。このことから、”like”のように様々な品詞になり得る可能性があることで品詞の判定を誤りやすい動詞の人称不一致の検出に対しては、既存の構文解析器の解析結果を利用することが困難であると考えられる。

また、be動詞においては人称が誤っていても正しい係り受け関係を207文全てで読み取れた。これは、be動詞が一般動詞のように様々な品詞になり得る可能性がなく、品詞の判定を正しく行えたからだ考える。

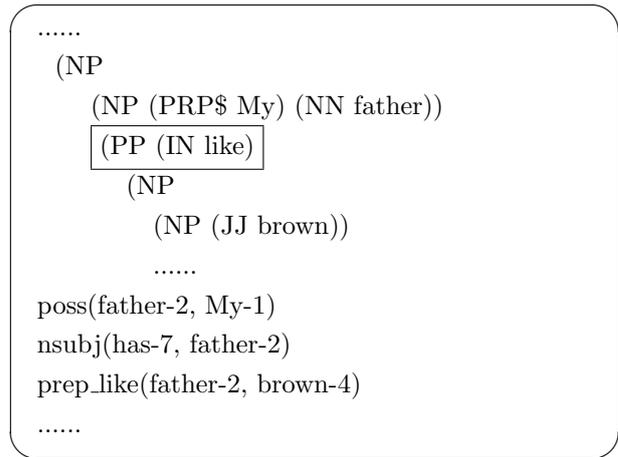


図1: 正しい係り受け関係を読み取ることができない文の解析結果の例

4 まとめと今後の課題

既存の構文解析器によって得られる係り受け関係を動詞の人称不一致の検出に利用できるかを検証した。be動詞の人称が誤っている文においては、207文全てで解析結果から正しい係り受け関係を読み取れた。一般動詞においては、”like”のような品詞の判定を誤りやすい動詞の人称が誤っている場合には正しい係り受け関係を読み取ることができない事例があった。”like”のような動詞の人称不一致の検出には、既存の構文解析器の結果を利用することは困難である。誤り文の正しい係り受け関係を把握するには、誤りが含まれる例文を学習させることで誤り文に特化した解析器を構築する必要があると考える。また、人称誤りの有無で主語と修飾語の係り受けが異なるなど、評価基準外の部分で解析結果に差異が見られる事例が存在した。このような事例の解析結果の傾向から人称不一致を検出可能か検証する。

参考文献

- [1] 伊東美津, 英語学習者のエラーについて. 九州国際大学 教養研究 第18号, pp.23-24, 2012.
- [2] 水本智也, 小町守, 林部裕太, 永田智明, 松本裕治, 大規模英語学習者コーパスを用いた英作文の文法誤り訂正の課題分析. 研究報告自然言語処理, 2012-NL-209(5), pp.1-8, 2012.
- [3] Dongfeng Cai, Yonghua Hu, Xuelei Miao, Yan Song, Dependency Grammar Based English Subject-Verb Agreement Evaluation, Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 1, pp.63-71, 2009.