

テ形従属節の用法分類に向けたアノテーションガイドラインの構築

概要

目標：テ形従属節の用法分類とアノテーションガイドラインを作成する

例文) 台風が来て、飛行機が遅れた

作成したガイドラインでアノテーションしたコーパスを教師データとして、テ形従属節の用法を判定するニューラル判定器を作成する予定

作成したガイドラインを用いて、試験的に小規模なアノテーションを行った

ガイドライン

信頼性の高いアノテーションを実現するために言語学的テスト^{[1][2]}を設計する

- ・対象表現を別の表現に置き換え可能かなどを作業者が言語直観に従って判定するもの

言語学的テスト欄の操作を上から順に適用。元の文と操作後の文の意味が同じであると判断したら、対応する用法をアノテーションし、以降の言語学的テストは行わない。

アノテーション

データ

BCCWJ^[3]のコアデータ（長単位）から

- ・品詞が助詞-接続助詞
- ・書字形出現形が「て」または「で」である要素を含む文を100文使用

※「において」などの複合辞は省かれる

ツール

brat^[4]で対象をハイライト
用法ラベルとUnlabeledラベルを用意

作業者

2名×4組が25文ずつ担当

結果

- 100文に含まれる124件中
- ・作業者間の判断一致…**42件**
- ・作業者間の判断不一致…**37件**
- ・アノテーション対象外…**45件**

準備段階で省けなかった

- ・テ形+助詞
「ても」、「ては」、「てから」など
- ・テ形+形式動詞
「ていく」、「てなる」など

[3]Maekawa et. al. (2014) 『現代日本語書き言葉均衡コーパス(BCCWJ)』
[4]Stenetorp et al. (2012) "brat: a Web-based Tool for NLP-Assisted Text Annotation", EACL 2012, https://github.com/nlplab/brat.

考察

- ・作業者間で判断が一致した対象には適切な用法がアノテーションされていた
- ・不一致となった原因としては、〈原因〉の言語学的テストである「ことで」への言い替えが〈手段〉のテ形従属節にもできたことなどが考えられる

判断が困難な例

用法はその用法の言語学的テストを適用したことを表す

◆針麻酔なので、本人は意識がとてまはっきりしていて、意外なほど元気に見えました^[5]

- ・意識がはっきりしていることを根拠にして元気に見えると述べているから〈原因〉だと作業者に判断してほしいが、「本人」の状態を書き並べているため〈並列〉に見える

並列 針麻酔なので、本人は意外なほど元気に見えて、意識がとてまはっきりしていました
原因 針麻酔なので、本人は意識がとてまはっきりしていたので、意外なほど元気に見えました

◆古い旅館に泊まって金縛りにかかったという話をよく聞きますが、おそらく成仏できない霊がいるのでしょう^[6]

- ・文全体を見れば古い旅館に泊まったことが金縛りの原因であるとわかるが、伝聞の段階で因果関係を認めて良いのか

原因 古い旅館に泊まったせいで金縛りにかかったという話をよく聞きますが、おそらく成仏できない霊がいるでしょう

- ・〈原因〉の言語学的テストを適用すると、この文の話者が伝聞した時点で古い旅館に泊まるのが金縛りの原因だとわかっている文になる

今後の目標

- ・ガイドラインを改良し、作業者間の判断一致率を高める
- ・大規模なコーパスアノテーションを行う

用法名	説明 言語学的テスト（テ形の置き換えなど）	例文
並列	前件・後件が並列関係にあり、独立である	おじいさんが山へ行って、おばあさんが川へ行ったのです ^[3]
	前件と後件を入れ替えられる	太郎は背が高く、力が強い ^[1]
逆接	前件から予想される事象に反する事象が後件にくる	彼はそのことを知っていて言わない ^[3]
	「のに」に置換可、かつ「ことに」「ために」に置換不可	花子は頼りなさそうに見えてけっこうしっかりしている ^[1]
仮定	前件が仮定条件を表す	彼が参加して5人になる ^[3]
	「たとしたら」または「たとしても」に置換可	そのブランドの財布は、安くて十万円ぐらいする ^[1]
原因	前件が後件の原因になる	雨が降って、涼くなった ^[2]
	「ので」または「せいで」または「ことで」に置換可	花子は勉強時間が足りなくて、いい成績が取れなかった ^[1]
手段	前件が後件の手段になる	眠って忘れよう ^[4]
	前件と後件を入れ替えた上で、「するために」に置換可	私は新幹線に乗って東京へ行った ^[1]
目的	前件が後件の目的になる	とにかく名前を言い当てたことに敬意を表して、鶏はそちらに進呈することにしたのである ^[4]
	「ために」に置換可	
状態	前件の状態において後件が成り立つ	歌を歌って帰る ^[2]
	「した状態で」または「しながら」に置換可	手品師は目隠しをして客の選んだカードを当てた ^[1]
継起	前件・後件に、時間的前後関係がある	学校へ行って、先生にあった ^[2]
	「した後」に置換可、または「それから」を前件の後に挿入可	花子は大学を卒業して、大手企業に就職した ^[1]

[1]戸次大介, 川添愛(JSeM), https://github.com/DaisukeBekki/JSeM.
[2]森田良行 (1975) 『複文の文型練習—「たら」「て」を含む文型を中心に—』, 講座日本語教育, (11), 1-15.
[3]森田良行 (1980) 『基礎日本語2』, 角川書店.
[4]仁田義雄 (1995) 『複文の研究(上)』, 87-127, くらしお出版.
[5]BCCWJ サンプルID: PB11_00013 藤原東演 『人生、考えすぎないほうがいい頭で動くなからだでつかめ』
[6]BCCWJ サンプルID: PB11_00021 天教院照玉 『それでも人は霊と共に生きている』

[1]田中リベカ, 小池恵里子, 戸次大介, 川添愛 (2012) 『言語学的テストに基づく意味アノテーションのガイドライン設計—確実性判断に関わる表現を中心に—』, 言語処理学会第18回年次大会発表論文集.
[2]戸次大介, 田中リベカ, 川添愛 (2012) 『MCNコーパス:モダリティ関連表現の曖昧性解消のためのアノテーションと言語学的テストの利用』, テキストアノテーションワークショップ・コンテスト, 国立情報学研究所.