

テ形従属節の意味論的分析に向けた用法分類とそのアノテーション

野口 咲帆 (指導教員：戸次大介)

1 はじめに

本研究では、テ形従属節が用いられた複文¹の意味を適切に表す論理式を自動的に導出することを目指す。日本語表現に頻出するテ形従属節（以下、テ形節）には、様々な意味用法がある。「空がにわかにか曇って、すぐに雨が降ってきた」(JSeM:Conj3463)²という文の前件と後件には、空が曇った後に雨が降ってきたという順序関係が成り立つ。一方、「花子は仕事ができる、性格がいい」(JSeM:Conj3631)という文の前件と後件には、花子が仕事ができることと性格がいいことの並列関係が成り立つ。

本研究の目標を達成するために、入力テキストに現れるテ形節の用法を曖昧性なく判定するニューラル判定器を作成し、その判定結果を用いて意味表示を導出するというタスクを考える。ニューラル判定器の教師データとしてアノテーション済みコーパスを構築する必要があるため、本研究では、テ形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。

ガイドラインの作成には田中ら[1]、川添ら[3] が用いた MCN コーパスへのアノテーション手法を採用し、言語学的テストを設計する。言語学的テストとは、言語学の理論構築及び検証に用いられるテストで、作業者が母語話者としての言語直観に照らして答えられるように作問される。作業者はテキスト中の表現が置き換え可能かなどを判定する。定義や例文と照らし合わせて判定する従来のアノテーション手法よりも、作業者の判断が揺れにくいという特長がある。

2 先行研究

日本語学の先行研究では様々なテ形節の分類が考えられているが、著者が定めた分類とそれに属する文の集合に対しての特徴が説明されたものであり、分類の判断基準は必ずしも明確に示されていない。本研究が先行研究と異なる点は、分類の判断基準を明確に提示し、アノテーションを行う点である。

先行研究として森田[6] が挙げられる。テ形節の意味が前件・後件の意味関係によって8種に分類されている。森田[6] が提案した8分類を表1に示す。

先行研究を基に本研究のアノテーションガイドラインを作成する際に生じた課題を3点挙げる。1点目は、同じ意味の表現へ言い換えができるかを、そのまま言語学的テストとして用いることはできない点である。森田[7] では、〈同時進行〉のテ形節を用いた複文は、前件・後件に「～ナガラ」の関係が成り立つとしている。その上で、「～ナガラ」は状態性を表す前件には接続しないため、テ形節とは性格の違いがあると述べた。そのため、〈同時進行〉のテ形節であっても、「～ナガラ」への言い換えができない場合がある。例として「立っ

¹複数の節を含む文を複文と呼ぶ。

²日本語意味論テストセット[2]

(<https://github.com/DaisukeBekki/JSeM>)

以下、(JSeM:XX) は日本語意味論テストセット[2](<https://github.com/DaisukeBekki/JSeM>) 中の JSeM id が XX であることを示す。

表 1: 森田[6] の用法分類と同じ意味の表現

用法	同じ意味の表現
並列	ソシテ
同時進行	～ナガラ
順序	ソレカラ、～テカラ
原因・理由	～カラ、～ノデ
手段・方法	～ニヨッテ、～デ
逆接	～ノニ、～ガ、～ニモカカワラズ
仮定の結果	～ト、～バ

て話す」を「立ちながら話す」に言い換えると不自然になることを挙げた。

2点目は、1つのテ形節に、複数の用法がラベル付けされてしまう点である。たとえば、「～後」「～ノデ」への言い換えができるかどうかでテ形節を異なる用法へ分類するならば、「台風が来て、東京行きの飛行機が欠航になった」(JSeM:Conj4346)という文のテ形節は2つの分類にまたがってしまう。このような場合、用法を1つに特定できないため、アノテーション結果に揺れが生じる恐れがある。

3点目は、複数の用法のテ形節に使用できる言い換え表現が存在する点である。例として、森田[6] から「歌を歌って帰る」と「彼はそのことを知っていて言わない」の2文を取り上げる。前者が〈同時進行〉、後者が〈逆接〉に分類されているが、両者とも「～ナガラ」への言い換えが可能である。このような表現を用いて言語学的テストを設計してしまうと、分類に曖昧性が残ってしまう。

3 テ形の用法分類

表2に本研究で提案するテ形節の用法分類と言語学的テストを示す。

3.1 アノテーション手法

アノテーション対象のテキストに対して、表2の言語学的テスト欄の操作を上から順に行う。元のテキストと操作後のテキストの意味が同じであると判断した場合、その言語学的テストに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。

3.2 言語学的テスト操作の具体例

テ形節の各分類について、テ形節を含む複文と言語学的テストが定める言い換えを施した文の例を以下に示す。

言語学的テストのうち、「～ノニ」「～ノデ」「～セイデ」は、「～スル」または「～シタ」に接続する。どちらかを用いて言い換えができた場合、言い換え可能であると判断する。前件が「～ナイデ」の形である場合は、「～シタ状態デ」の代わりに「～ナイ状態デ」への言い換えを行う。

並列 1. おじいさんが山へ行行って、おばあさんが川へ行行ったのです[7]

表 2: テ形節の用法分類と言語学的テスト

適用順	用法	特徴	言語学的テスト
1	並列	前件・後件が並列関係にあり、独立である	前件・後件の入れ替え
2	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	「～ノニ」
3	仮定	前件が仮定条件を表す	「～タトシタラ」または「～タトシテモ」
4	原因	前件が後件の原因になる	「～ノデ」または「～セイデ」
5	手段	前件が後件の手段になる	前件・後件の入れ替え後「～スルタメニ」
6	状態	前件の状態において後件が成り立つ	「～シタ状態デ」または「～シナガラ」
7	継起	前件・後件に、時間的前後関係がある	「～シタ後」または「ソレカラ」

2. おばあさんが川へ行って、おじいさんが山へ行ったのです
- 逆説**
1. 彼はそのことを知っていて言わない[7]
 2. 彼はそのことを知っているのに言わない
- 仮定**
1. 彼が参加して5人になる[7]
 2. 彼が参加したとしたら5人になる
- 原因**
1. 500m 走って、疲れた[6]
 2. 500m 走ったので（せいで）、疲れた
- 手段**
1. 眠って忘れよう[4]
 2. 忘れるために眠ろう
- 状態**
1. 歌を歌って帰る[6]
 2. 歌を歌いながら帰る
- 継起**
1. 学校へ行って、先生にあった[7]
 2. 学校へ行って、それから先生にあった

形+ナガラ」という言い換えは、〈逆接〉の文にも適用可能な場合がある。このことを踏まえ、状態の言語学的テストとして「連用形+ナガラ」を用いるため、〈逆接〉には別の言い換えを言語学的テストとして用い、〈逆接〉の言語学的テストを〈状態〉の言語学的テストよりも先に行うようにした。

4 現状と今後の課題

テ形節を対象とした先行研究から例文を270件抜き出し、1名の作業者がアノテーションを行った。現時点では248件の分類ができた。分類できていない例文として、「事務所の奥へ来てもらい、犯人の素性を確かめて驚いた」[4]などが挙げられる。提案手法では〈継起〉に分類されるが、これは〈原因〉に分類されるべきである。そのため、言語学的テストの改良をする必要があると考えている。

また、〈並列〉に分類されるが、同時に「～ノニ」「～ノデ」への言い換えが可能であるテ形節についても考慮したい。これには「普通の人がほしがめるものはほしくなくて、ほしがらないものがほしい」[5]「この山道は険しくて、歩きにくい」(JSeM:Conj3627)が該当する。

5 おわりに

本研究では、テ形従属節の用法分類の提案を行い、それに基づいたアノテーションガイドラインを作成した。今後は、用法分類と言語学的テストの妥当性を確認するため、100件程度の試験的なアノテーションを実施する。その後、改良したガイドラインを用いて、現代日本語書き言葉均衡コーパス(BCCWJ)のうち、テ形節を含む複文をアノテーションする予定である。

参考文献

- [1] 田中リベカ, 小池恵里子, 戸次大介, 川添愛. 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, pp. 401-404, 2012.
- [2] 川添愛, 田中リベカ, 峯島宏次, 戸次大介. 日本語意味論テストセットの構築. 言語処理学会第21回年次大会, 2015.
- [3] 川添愛, 田中リベカ, 戸次大介. 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, 2012.
- [4] 仁田義雄. シテ形接続をめぐって. 複文の研究, pp. 87-126, 1995.
- [5] 益岡隆志, 田窪行則. 基礎日本語文法-改訂版, 1992.
- [6] 森田良行. 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, No. 11, pp. 1-15, 1975.
- [7] 森田良行. 基礎日本語 2. 角川小辞典, 1980.

3.3 〈手段〉の言語学的テスト詳細

文末が「ウ接続形+ウ」という形の場合、一旦「～ウ」以前の部分のテキストに言語学的テスト操作を行い、最後に文末を「ウ接続形+ウ」の形にする。「眠って忘れよう」[4]の場合は以下の通りになる。

1. 「眠って忘れよ」+「ウ」
2. 「忘れるために眠ろ」+「ウ」
3. 「忘れるために眠ろう」

〈手段〉の言語学的テスト操作を施すと、後件の事象が過去に起こった事実であっても、変換後には既に実現したかどうか不明な出来事になる。たとえば、「大いにがんばって、仕事を全部すませた」[6]を「仕事を全部すませるために、大いにがんばった」と変換すると、仕事が全部済んだのかが曖昧になる。〈手段〉の言語学的テスト操作後は、上記の意味の変化は看過して判断する。

3.4 言語学的テストの順序

言語学的テストの順序を表2の通り提案した理由を述べる。まず、2つの事象が独立か従属かは一番大きな分類であるため、〈並列〉の言語学的テストを最初に行う。次に、〈逆接〉〈仮定〉の言語学的テストは、〈原因〉以下の4分類の文に対して適用すると不自然であると判断を下しやすい場合が多いため、〈逆接〉〈仮定〉を2、3番目に置いた。〈原因〉以下の4分類については、〈原因〉〈手段〉のテ形節の複文が〈状態〉〈継起〉の言語学的テストにも通る場合が多いため、表2の順序にした。

〈状態〉の言語学的テストとして用いている「連用