

MCN コーパスにおける形式名詞「はず」「わけ」「つもり」のアノテーション

宇津木 舞香¹ 佐藤 未歩¹ 青木 花純¹ 田中 リベカ² 川添 愛³ 戸次 大介^{2,3,4}

¹ お茶の水女子大学理学部情報科学科

{g1120508,g1120522,g1120501}@is.ocha.ac.jp

² お茶の水女子大学大学院人間文化創成科学研究科

{tanaka.ribeka,bekki}@is.ocha.ac.jp

³ 国立情報学研究所

zoeai@nii.ac.jp

⁴ 独立行政法人科学技術振興機構, CREST

1 はじめに

自然言語で記述されたテキストには、事実だけでなく、推測や仮定などの様々な情報が含まれる。人間は、様相表現・条件表現・否定表現といった情報の確実性に関わる表現を手がかりとして、それらテキストに含まれる命題の真偽や信憑性を判断している。機械が自然言語を処理する際にも、これらの表現に着目することが重要となる。MCN コーパス(川添ら)[3]は、機械による確実性判断の基盤となるコーパスを構築するために作成されたものであり、情報の確実性に関わる表現にアノテーションを付与した言語データである。MCN コーパスのアノテーションでは、言語学的テストとして「ネガティブテスト」を採用したガイドラインを使用している。ここでの言語学的テストとは、文や文の一部の容認生や適切性を判定するものである。ネガティブテストは「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」(田中ら)[1, 2]に基づくものであり、「必要条件となる条件を満たさない場合、その分類に属さない」という形式のテストである。特に、表現の置き換えを利用したテストにおいては、「置き換え可能」という判断よりも「置き換え不可能」という判断の方がアノテータ間での一致度が高い傾向が見られることがわかっている事から、「置き換え不可能であれば、そのカテゴリに分類されない」という形式をとる。ネガティブテストを採用したガイドラインによるアノテーション作業では、消去法で分類先を一つに特定することになる。

本研究では、「はず」「わけ」「つもり」の三つの形

式名詞について、ネガティブテストを導入したガイドラインを作成し、実際にアノテーションを試みた。各表現の分類について『形式名詞がこれでわかる』(吉川ら)[4]を参考とした。本論文では、以下第3節で、「はず」「わけ」という二つの表現に共通する問題について考察し、その改善策を提案する。第4節では、「つもり」を通して慣用表現について考察する。第5節では実際のアノテーション作業に関する詳細と、その結果について述べる。

2 対象外とする表現について

MCN コーパスにおけるガイドラインは命題の確実性判断に関わる表現を対象としている。「わけ」を含む文は多々考えられるが、その中には確実性判断に関わる働きを持たない「わけ」も存在する。以下二つの例文はどちらも「わけもない」という表現を含むが、本ガイドラインでは例文2.は対象外としている。

1. 子供の私が親の気持ちなど理解出来るわけもない。
2. 算数の問題を解くことなどわけもない。

例文1.の「わけもない」は、論理的に考えて起こる可能性がないことを表し、筆者らのガイドラインでは[わけもない8]に分類される。例文2.の「わけもない」は「容易い、簡単である」ことを意味しており、命題の確実性を表す表現ではないのでアノテーションの対象外としている。また、他にも「わけがわからない」の「わけ」は理由などを表す名詞であり、形式名詞で

表 1: はずガイドライン抜粋

分類	特徴	例文	テスト
はず 2	認知的推量の「はず」。命題の真偽は話者にとって未知だが、真である確率が高い。	今日は負けたが、この次はきっと勝てるはずだ。	「違いない」置き換え不可であったらこのカテゴリでない。もしくは、「きっと」「おそらく」をつけて意味が変わる場合はこのカテゴリでない。
はず 2.5	認知的推量の「はず」。命題は真であるが、結果は偽である。	「見返りを求めない」はずのボランティア精神を損ないかねない。	形が「はずの」「はずが」「はずなのに」のいずれでもない場合はこのカテゴリでない。
はず 3	実現されていない予定や、事実と異なる想定を述べる「はず」。	新製品は昨年 5 月に発売されるはずだった。	「はず」の前の命題が偽であるとは言えない場合はこのカテゴリでない。

はない。よって本研究では、例文 2 の「わけ」や形式名詞でない「わけ」をアノテーション対象外とし、用言の連体形に接続する「わけ」のみを対象とした。

3 [はず・わけ] にみる曖昧な表現の分類について

「はず」「わけ」の二つの形式名詞についてのガイドラインを作成し、実際にアノテーション作業を行ったところ、この二つには本質的に曖昧な表現が存在することが分かった。以下では、「はず」「わけ」それぞれに存在する曖昧な表現の具体例を述べ、そのような表現を含む文章をアノテーションする際に生じる問題と、その解決策について考察する。

3.1 [はず]

本ガイドラインをもとに行ったアノテーション作業の結果、アノテータ間の答えの一致率は約 86 % となり、一定の成果は得られたものと思われるが、その一方で、一致しなかったカテゴリは [はず 2][はず 2.5][はず 3] の三つに集中していた。表 1 に「はずガイドライン」の一部抜粋を示し、各カテゴリの例文を以下に示す。

[はず 2]. 今日は負けたが、この次はきっと勝てるはずだ。

[はず 2.5]. 「見返りを求めない」はずのボランティア精神を損ないかねない。

[はず 3]. 新製品は昨年 5 月に発売されるはずだった。

三つの例文を見ると、それぞれの「はず」の意味や用法に違いがあるのは比較的明らかであり、異なる分類に分けられることも明らかである。それにも関わらず、これらの区別が難しくなっているのは、「はず」の中に、本質的に曖昧な表現が存在しているからである。次の例文 4 と例文 5 は、曖昧な表現の例である。

1. もっとゆるいカーブで丘裾の崖を降りていったはずだ。

上記の例文は、「はず」の前の命題が真であるとも偽であるとも捉える事が出来る。ここで、[はず 2] のテストを試みると「もっとゆるいカーブで丘裾の崖を降りていったに違いない」という文章が得られる。この文章は、「はず」の前の命題が真偽は未知だが、真である可能性が高いという、[はず 2] の特徴を満たしている。一方で、話者が勘違いして、「実際はカーブがゆるくなかった」と考える事もできる。このとき、例文 4 の「はず」の前の命題は偽であると考えられ、[はず 3] の特徴を満たす。このように聞き手の解釈により真とも偽ともとれる文については、本ガイドラインのテストはうまく機能しなくなってしまう。

テストがうまく機能していないのは、[はず 2.5] や [はず 3] に分類される「はず」に共通する、聞き手によって文の意味の捉え方が変化するという特徴をうまくとらえることができないからである。そのような特徴を捉え、テストとして使うことができれば、本ガイドラインはより信頼性の高いものになると思われるが、聞き手によって文の捉え方が変化してしまう以上、そのようなテストの作成は容易ではない。

3.2 [わけ]

「わけ」についても、同様の問題が生じる。「わけ」の否定表現である「わけではない」を例に考察する。

本ガイドラインにおいて「わけではない」は二つのカテゴリに分類できるとしている。[わけではない 5] は、命題の内容そのものを否定するという特徴をもつカテゴリであり、ネガティブテストは「「のではない」に置き換え不可能ならばこのカテゴリではない」とした。一方、[わけではない 6] は、状況や直前の発話から聞き手が推論を働かせて考えうる事柄を話し手が予め否定するという特徴をもつ。聞き手の信念についての仮定を含むことから、ネガティブテストは「「つも

りではない」に置き換え不可能ならばこのカテゴリではない」とした。以下の例文について考える。

1. 私と彼は付き合っているわけではない。

彼と私の二人は交際していないということは容易に読み取れるだろう。しかし、この文章に聞き手の信念についての仮定が含まれるかどうかは読み手の捉え方で異なり、アノテータ間でこの判断を一致させることは難しい。

例文1について、ネガティブテストによる分類判定を試みる。

1'. 私と彼は付き合っているのではない。

1''. 私と彼は付き合っているつもりはない。

文章1'は、[わけではない5]のネガティブテストを実施した場合である。これは文章として成立しており、例文1.は[わけではない5]に分類することが可能である。文章1''は[わけではない6]のネガティブテストを実施した場合であり、同様に、例文1.が[わけではない6]に分類することが可能であることが分かる。つまりネガティブテストによる分類を試みると、例文1.は[わけではない5][わけではない6]のどちらのカテゴリにも属する可能性を持ってしまう事が分かる。

ここで、二つの文章の特徴を見てみる。文章1'は、単純に「付き合っている」という内容そのものを否定していて、[わけではない5]の特徴を満たしている。同様に、文章1''は、「付き合っているように見えるだろうが、実際は付き合っていない」ことを意味していて、聞き手の信念についての仮定を含んでいるため[わけではない6]の特徴を満たしている。以上より、どちらの分類の特徴も満たしている事から、例文1.は、捉え方によってどちらのカテゴリにも属することが可能になってしまう。この事は、2.1節で述べた「はず」が抱える問題と良く似ていると言える。

3.3 まとめ

文章の捉え方の違いは読み手の感覚によるものであり、上で述べたような例文はカテゴリが一意に定まらず、本ガイドラインで分類する事は困難であると述べた。本質的に曖昧な例に関してはアノテータの解釈が分類を左右する。よって、偶然にもアノテータ間の感覚が似ていたために、アノテーション結果が一致したからといって、信頼性の高い結果が得られたと結論づけることは出来ない。

本研究のアノテーションの方法では、ある表現が属するカテゴリは一つであると限定していたために、複数のカテゴリに属する可能性のある曖昧な表現の分類結果が一致しなかった。解決方法として、複数のカテゴリに属する可能性のある表現を新たなカテゴリとして確立させる方法が挙げられる。つまり、「分類aと分類bに属する可能性を持つ」分類cという新たなカテゴリを適宜増やしながらかつアノテーションを進めて行くとする。しかし、この方法では、カテゴリが膨大な数になってしまう可能性があり、実現することは困難である。よって筆者らは、本質的に曖昧な表現は一つのカテゴリに属することを正解とするのではなく、複数のカテゴリに同時に属することが正しい分類であると考え。つまり、本質的に曖昧な表現に対するアノテーションの結果が一致するとは、各アノテータが、ある曖昧な表現に対して分類可能であると判断したカテゴリの集合が、他のアノテータが考える集合と合致するときをいう。

4 [つもり]にみる慣用表現の分類について

MCNコーパスでは、慣用表現には別カテゴリを設けている。慣用表現は、構成的には決定できない固有の意味を持つ。そのような表現は独自のカテゴリとして分類すべきである。しかし、与えられた表現が慣用表現であるかどうかの判断は、常に自明とは限らない。一方で、すべての慣用表現をリストアップすることも、また現実的ではない。したがって、リストにない慣用表現を、アノテーション作業を通して発見する過程が求められる。

アノテーション作業において新たな表現が現れたとき、それが慣用表現であり、かつガイドラインが一定以上に網羅的である場合は、その表現は既存のどのカテゴリにも属さないことがテストを通して判断される。このように、ネガティブテストによって構成されたガイドラインは、新たな慣用表現の発見について一定の機能を果たすようにデザインされている。

しかしながら、慣用表現であることが疑われるような表現は、統語的に特異な構造を持つ場合が多く、テストの判断に揺れが生じることがある。よって判断の誤りを系統的に見出すことは難しく、慣用表現のカテゴリ作成には困難が伴う。

以下では、形式名詞「つもり」を例に、慣用的表現の分類について考察する。以下の例文を考える。

1. あなた、何様のつもり？

この表現の分類先を判断するために、ガイドラインの各カテゴリにおけるテストを適用する。たとえば、カテゴリ2のテスト(「つもりだ」を「と自負している」に置き換え可能であるかのテスト)を行うと、「あなた、何様と自負している？」となるが、これは置き換え不可能であり、条件を満たさないように思われる。その他のテストも同様に行った結果、上記の表現はどのカテゴリにも当てはまらず、したがって慣用表現として別カテゴリを設けた。

しかし、上述のテストが置き換え不可能である原因は、「何様」という疑問表現と、「のつもり」「と自負している」という二つの表現の統語的な違いの相互作用にある可能性もある。このように、その他の表現がもたらす統語環境がテストに影響してしまう場合があり、慣用表現のカテゴリを増やすという判断は慎重に行わなければならない。

5 評価

ここでは、三つの形式名詞「わけ」「はず」「つもり」それぞれについて行ったアノテーション活動の詳細を示すとともに、ガイドラインの有用性などについて述べる。

5.1 アノテーション活動の詳細

「わけ」「はず」「つもり」の三つの形式名詞のアノテーション作業は、「わけ」と「はず」についてはガイドライン設計者と文教育学部言語文化学科英語圏言語文化コースの学生の計2名で行い、「つもり」についてはガイドライン設計者と上記の学生、「わけ」「はず」「つもり」それぞれについていくつかのカテゴリを設定し、多くの文章の中から該当の表現が出現する部分を抜き出し、その用法がどのカテゴリに属するのかを、テスト等をもとに判断する、という形で行った。「わけ」「はず」「つもり」のガイドラインにおけるカテゴリ数は、それぞれ11、6、7である。

アノテーションの件数は「わけ」が約500件、「はず」が約50件、「つもり」が約80件であり、作業時間はおおよそ、「わけ」が12時間、「はず」が2時間、「つもり」が1.5時間である。

5.2 カップバ値

「わけ」「はず」「つもり」それぞれのアノテーション結果から、評価値としてカップバ値を計算したところ、「わけ」については0.88、「はず」については0.75、「つもり」については0.88という結果が得られた。全体的に高い値が得られており、信頼性の高いガイドラインが作成できたと判断できる。

6 おわりに

「はず」「わけ」「つもり」の三つの形式名詞について、ガイドラインを作成し、問題点を考察した。第5節で述べたように、本ガイドラインにおけるアノテーションの一致率は高いと言えるが、本質的に曖昧な表現の分類や慣用的表現についてはさらなる改良が必要である。

参考文献

- [1] 田中リベカ, 小池恵里子, 戸次大介, 川添愛 (2012a) 「言語学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.
- [2] 田中リベカ, 川添愛, 戸次大介 (2012b) 「MCN コーパス: 言語学的テストに基づくモダリティ・アノテーションの理論と実証」, 国立国語研究所第2回コーパス日本語ワークショップ予稿集, pp135-144.
- [3] 川添愛, 斎藤学, 片岡喜代子, 戸次大介 (2011) 「言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン Ver.2.4」, Technical Report of Department of Information Science, Ochanomizu University, OCHA-IS 10-4.
- [4] 吉川武時 編, 代表 小林幸江, 柏崎雅世 (2003) 『形式名詞がこれでわかる』, ひつじ書房.