

文法論によるモダリティの細分類を判断、態度をアノテーションする基準に用いる問題点

小橋洋平*

東京工業大学大学院社会理工学研究科†

kobashi.y.aa@m.titech.ac.jp

1 はじめに

近年、書き手・話し手（以後、表現者）の命題に対する心的態度に着目して、テキスト、談話から重要度の高い情報を抽出する研究が行われている。表現者の事実認識から確実性の高い情報を特定する [1, 2, 3, 4, 5]、命題に対する表現者の評価から特定の主題に対する主張、意見を特定する [6, 7] など、対象とする心的態度を絞り込んでいる研究もあれば、体系的な心的態度のカテゴリを定め、網羅的に特定することを目的とした研究 [8] も存在する。

そして、これらの研究で提案されているアノテーションは、言語学のモダリティ論を参考にしているものが多い。モダリティ論とは、研究者によってその定義が異なる概念ではあるが、日本語学では、文は客観的な事柄を表す「命題」と表現者の心的態度を表すモダリティからなるという規定が多く、研究者に受け入れられてきた [12] 概念で、自然言語処理でも心的態度を指す概念として認識されていることが多い。

その一方で、モダリティでは心的態度を網羅的に捉えることができないという指摘もある。日本語学の文法論でモダリティとされる形式は陳述副詞、活用、助動詞、終助詞および助動詞、終助詞相当語句に限られる。対して、心的態度を表す表現には、用言、接続助詞といった他の形式も存在する。そのため、心的態度を網羅するにはモダリティの概念では十分ではなく、「拡張モダリティ」 [7] 「主観述語」 [9] といった新たな概念を導入すべきだという主張が展開されている。

ただし、モダリティの概念を拡張することによって、

アノテーションの難易度が上がることが予想される。原則、特定の表現が心的態度を表すかどうかは、文体や記述内容、書き手、読み手の社会的属性など、語用論的な要因によって変化すると考えられる。対象がモダリティだけであれば検討すべき対象は有限だが、文法形式以外に拡張すると「という見解が伝えられている」「ことが日本の未来に関わる」など、文脈次第で無数の表現が心的態度に関わる可能性がある。この結果、アノテーションする人の負担が膨大なものになる、または判断にぶれが生じると予想される。

これに対し、本発表では、既存の研究で引用されているモダリティのカテゴリは、文法形式を分類することが主旨のため、心的態度をアノテーションする主旨には向かないこと、および、意味論で議論されている「モダリティ」は、様相論理学の基本的枠組みと類似しつつ、言語学に合わせた直感的なカテゴリを提示しており、こちらに基づいてアノテーションすることが望ましいことを主張する。既存研究では、モダリティ表現への言い換え可能性によって心的態度の有無を評価する、または文法論によるモダリティのカテゴリをアノテーションの基本的枠組みに採用するといったものが主流となっている。しかし、文法形式は本質的に多義的であり、それらを機能によって分類すると、個別言語への対応は精緻なものとなる一方で、各カテゴリは複数の機能を内包するものになりやすい。一方、意味論に基礎をおくと、個別言語に捉われず、カテゴリが比較的、単純な概念となる。

以下、2節でモダリティ論のカテゴリと文法形式の多義性との関係から、カテゴリの具体性は文法形式によって与えられているのであって、概念の規定自体は抽象的であることを論じる。次に3節で、様相論理と意味論に基づいたアノテーションの枠組みと具体的な適用例を示す。そして4節でまとめを述べる。

* Kobashi, Yohei

† Tokyo Institute Technology

2 モダリティのカテゴリと心的態度

2.1 モダリティのカテゴリと文法形式

前述の通り、心的態度のアノテーションでは主に事実認識と評価を扱う研究が存在する。それぞれの心的態度は、一般言語学のモダリティ論で伝統的なカテゴリである認識的 (epistemic) モダリティと束縛的 (deontic) モダリティに対応している [10, 11]。この分類は、英語では法助動詞の多機能性を示すために引用される。例えば、must は、命題が事実である、または実現するということを推測していることを表す場合と、命題が実現することが価値観や規範上望ましいという為判断を表す場合とがある。この違いは、前者が認識的と後者が束縛的という形で解釈される。

一方、言語によっては認識的と束縛的という違いが文法上の違いとして現れる。その例として挙げられるのが日本語である。日本語学のモダリティでも、一般言語学と同様に認識のモダリティと評価のモダリティという認識的、束縛的に対応するカテゴリが存在するが [12, 13]、こちらは、それぞれのカテゴリに対して異なる文法形式が属している。認識のモダリティには「方がいい」「べきだ」「ざるを得ない」など、評価のモダリティには「だろう」「かもしれない」「らしい」などの文法形式が属する。

モダリティ論によって与えられるカテゴリの重要な特徴として、カテゴリの規定自体は言語普遍的である一方で、カテゴリが各言語で文法形式とどのように対応づけられているのか、詳細な言語観察によって体系化されていることが挙げられる。

2.2 モダリティのカテゴリの規定としての抽象度

モダリティのカテゴリが言語普遍的で、かつ文法形式との関係の記述にも適応できる背景には、カテゴリが抽象的な概念として規定されていることがある。ここで [13] による日本語学における認識のモダリティと評価のモダリティの規定を表 1 に示す。それぞれ、規定自体には「認識」「評価」というカテゴリの名前以上の情報は含まれていない。その一方でカテゴリの細分類が与えられることで具体性が高いものになっている。

このとき、各細分類を理解する上で重要なのは各細分類に属する文法形式である。例えば、推量の「だろう」が蓋然性や証拠性を表し得ないのかということを考えると、推量、蓋然性、証拠性の境界は明確ではないように思われる。にもかかわらず、文法形式が与えられること

で、母語話者はその文法形式から細分類の特徴を感覚的に理解することができる。そして、母語話者による観察を通して統語上の共通点という客観的な特徴も明らかにされている。すなわち、各カテゴリに、具体性を与えているのは文法形式であるといえよう。

2.3 モダリティの文法形式を目安とすることの問題点

心的態度の各カテゴリがモダリティ論に基づいて与えられることになると、自然と各カテゴリの典型的な表現はモダリティの文法形式なると予想される。しかし、ここで注意しなければならないのは、文法形式自体は多義的であるという点である。例えば、評価のモダリティで必要に属する「なくてはならない」だが、これは表現者の価値観を表明している場合もあれば、外部の規則を伝えている場合もあり、前者と後者をそれぞれ用言を用いて書き直すと、「ことが望ましい」「と定められている」などと全く異なるものとなる。さらに他の命題との因果関係を表す場合もあり、その場合は心的態度としての側面が極めて弱まる。

このような傾向は程度の差こそあれ、どの文法形式にも存在する。よって文法形式との機能的な類似性に基づいて心的態度のアノテーションを行うと、全く異質な表現が同じカテゴリに存在することになりかねない。文法形式に基づくアノテーションでは、あくまでモダリティと類似した表現のアノテーションになってしまい、心的態度を分類するには適していないといえよう。

3 心的態度の意味論によるカテゴリ化

3.1 意味論に基づく心的態度のカテゴリ

前節では文法論のモダリティのカテゴリに基づくアノテーションでは、心的態度の分類としては適切ではないと述べた。その代わりとして挙げられるのが意味論による「モダリティ」のカテゴリである。意味論では「モダリティ」という用語を特定の言語の文法形式によらず規定している。ただし、特定の言語表現との結びつきがないため、カテゴリが細かく規定されているとアノテーションの作業者が混乱する恐れが高い。そこで、本発表では、意味論の規定を用いつつ、細分類にまでは踏み込まず、日本語に適していると思われる独自のカテゴリを提示する。

まず、カテゴリの基本的な枠組みを様相論理に基づいて示す。様相論理では命題 p の属性として、真理値の他に可能性と必然性という概念を導入し、かつフレームと

表1 認識のモダリティと評価のモダリティの規定 [13]

カテゴリ	規定	カテゴリの細分類
認識	事態に対する話し手の認識的な捉え方を表す	断定と推量 (断定形, だろう etc)
		蓋然性 (かもしれない, にちがいない etc)
		証拠性 (ようだ, らしい etc)
評価	話し手が何らかの事態を述べ伝えるときに, その事態に対する話し手の評価的なとらえ方を表す	必要 (方がいい, べきだ, ざるを得ない etc)
		許可・許容 (てもいい etc)
		不必要 (なくてもいい, ことはない etc)
		不許可・非許容 (てはいけない etc)

いう概念によって認識と束縛や仮想という視点の違いも記述できるようになっている。この様相論理を用いて自然言語のモダリティを記述する試みも数多く行われているが、現状では欠点が完全に解消されたものが提案されているとは言い難い [15]。

ここでは厳密な推論ができる枠組みをつくることは放棄し、様相論理の

- 可能性と必然性
- フレーム

の概念を、意味論の「モダリティ」で提示されている力動的 (dynamic), 存在的 (数量化 quantificational と対応), および仮想世界の概念 [11] と置き換えることでアノテーションの枠組みを提案する。

1. 最初にどのフレームか決める
 - 現実
 - 表現者の価値観または規則
 - 仮想 (その他)
2. 可能性、必然性の代わりに、次の4つを考える
 - 任意の1時点 (時間的幅があってもよい。時間幅の最大は無限)
 - 断続的, 定期的に起こること
 - 潜在的な能力、可能性 (典型的には can や「できる」)¹
 - 決められない (その他)
3. 上の2つが定まった上で真偽について話し手が判断しているかをアノテーションする
 - 真と思っている

¹ これはアノテーションの目的によっては不要

- 偽と思っている
- わからない、真偽を考えていない (その他。目的によっては疑問を分けてもよい)

という基準で評価する。このモデルは原則、真偽と束縛、仮想の違いをフレームの違いと捉えて、どちらにせよ心的態度としては { 真, 偽, わからない } という3択を示しているだけとみなしている。言語や文脈によっては1や2の特定が困難な場合もあり得るが、その場合は無理な絞り込みはせず、現実以外のものはアノテーションから外すという形で対応すればよい。枠組み事態は、様相論理の比喩的なものに過ぎないが、将来的に厳密な様相論理のモデルが提案された際に対応させやすいという点でも有効だと思われる。

3.2 確実性判断への意味論カテゴリの適用

近年、医学・生物学テキストから確実性の高い情報を特定するための研究が盛んに行われているが、この分野では、アノテーションの対象がモダリティに近いものに限定されている研究がある [4]。(1)は「だろう」、(2)は断定形としてそれぞれ事実認識の情報を読み取ることができる。

(1) たしかに、米国の大手金融機関が破綻したり、ドルが暴落したりする可能性はごく小さいだろう。(朝日新聞 08/01/17)

(2) 最近ではインターネットの発達で、誰でも簡単に情報を交換したり、発信したりできる。(日本経済新聞 08/01/18)

この(1)と(2)の主節に対し上記のアノテーションの手続きを適用すると1では両方とも事実、2では前者は1時点、後者は潜在的な能力、3は両方とも「真」とな

る。また、「かもしれない」は、「かもしれないし違うかもしれない」と言えることから、手続きの3で「わからない」となる。

また、この手続きでは、モダリティ以外の心的態度もアノテーションの対象となりうる。[5]は、推測と否定をアノテーションするだけでは確実性判断を全て取り扱うことができないと主張している。例えば、次の(3)や(4)のように仮定表現が含まれる文の場合、その命題は事実とは言えない。

(3) この改革がきちんと実現すれば、霞が関の透明度はぐっと高まる。(朝日新聞 09/03/31)

(4) 人材源の拡大が急務なのに、世襲候補を放置したら、多様な人材の政治への進出は困難になるばかりだ。(毎日新聞 09/03/07)

ここで上記の手続きを(3)の「a. この改革がきちんと実現する」「b. 霞が関の透明度はぐっと高まる」「c. この改革がきちんと実現すれば、霞が関の透明度はぐっと高まる」の3つの命題に適用する。まずaとbは「仮想(その他)」だが、cについては現実世界の属性を記述している。次に2だが、この文の情報だけでは仮想世界の時間軸が分からずa, bの時間を定めることはできない。cは潜在的な可能性なので潜在的な能力に含まれる。そして3の真偽は、aは仮想世界で仮定しているという意味で「真」、cは現実世界の潜在的な可能性として「真」だが、bは仮定(その他)に分類され、仮定されているわけではないので、「わからない」ということになる。

4 おわりに

本発表では、モダリティのカテゴリが文法形式によって具体性を持つこと、およびモダリティの文法形式が多岐であることから、心的態度のカテゴリとしては適当ではないと主張した。その上で、様相意味論と意味論に沿ったカテゴリを提案し、具体的な適用例を示した。

参考文献

- [1] Light, M., Qiu, YX. and Srinivasan, P.: The Language of Bioscience: Facts, Speculations, and Statements In Between, Proceedings of BioLink 2004 Workshop on Linking Biological Literature, Ontologies and Databases: Tools for Users, pp.17-24 (2004).
- [2] Medlock, B. and Briscoe, T.: Weakly Supervised Learning for Hedge Classification in Scientific Literature, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp.992-999 (2007).
- [3] Szarvas, G., Vincze, V., Farkas, R. and Csirik, J.: The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts, BioNLP 2008: Current Trends in Biomedical Natural Language Processing, pp.38-45 (2008).
- [4] Thompson, P., Venturi, G., McNaught J., Montemagni, S. and Ananiadou, S. Categorising Modality in Biomedical Texts, LREC 2008 workshop Building and Evaluating resources for biomedical text mining (2008).
- [5] 川添愛, 齊藤学, 片岡喜代子, 戸次大介: 確実性判断に関わる意味的文脈アノテーションの試み, 情報処理学会研究報告, 189, pp.77-84 (2009).
- [6] 永野圭一郎, 辻井潤一, 鳥澤健太郎: 談話文からの命題-様相の抽出システム, 言語処理学会第7回年次大会論文集, pp.38-41 (2001).
- [7] 乾裕子, 井佐原均: 拡張モダリティの提案: 自由回答から回答者の意図を判定するために, 電子情報通信学会技術研究報告 言語理解とコミュニケーション, 102(414), pp.31-36 (2002).
- [8] 江口 萌, 松吉 俊, 佐尾 ちとせ, 乾 健太郎, 松本 裕治: モダリティ、真偽情報、価値情報を統合した拡張モダリティ解析, 言語処理学会第16回年次大会論文集, pp.37-48 (2010).
- [9] 牧野武則: 日本語の主観表現の機能的構造—主観述語—, 情報処理学会研究報告, 2009-NL-191, 2 (2009).
- [10] Palmer, F.R.: Mood and Modality Second edition, Cambridge University Press (2001).
- [11] 澤田治美: モダリティ, 開拓社 (2006).
- [12] 宮崎和人, 安達太郎, 野田晴美, 高梨信乃: 新日本語文法選書4 モダリティ, くろしお出版 (2002).
- [13] 日本語記述文法研究会(編): 現代日本語文法4 第8部 モダリティ, くろしお出版 (2003).
- [14] Palmer, F.R.: Mood and Modality, Cambridge University Press (1986).
- [15] Portner, P.: Modality, Oxford Univ Pr (Txt) (2009).