

Yahoo!知恵袋における修辞ユニット分析の発話機能認定に関する諸問題

田中弥生[†] 浅原正幸^{††}

[†]東京大学大学院 総合文化研究科・人間文化研究機構 国立国語研究所 研究系音声言語研究領域

^{††}人間文化研究機構 国立国語研究所 コーパス開発センター

1. 背景と目的

修辞ユニット分析(Rhetorical Unit Analysis, 以下 RUA とする)は、英語の母子会話をもとに Cloran(1994, 1999)によって提案された選択体系機能言語理論の談話分析手法の一つで、テキストについてメッセージ(原則として節)単位で、「発話機能」(命題か提言か)、「中核要素」(主に主語で認定。空間的距離を示す。),「現象定位」(副詞や述部で認定。時間的距離を示す。)を認定して、その組み合わせから「修辞機能」と「脱文脈化指数」を特定するものである¹。筆者らはこれまでに、日本語のインターネット上の QA サイトやクチコミサイト、中学生の会話等の分析に使用し、修辞機能や脱文脈化程度の展開等に係るテキスト分析での有用性を示している(田中 2013 他)。従来にはなかった修辞機能や脱文脈化の観点からのテキスト分析手法であり、今後評判分析や作文指導など様々な学問領域での研究に広く利用されるためには、初学者によっても簡便な訓練により分析できるような認定基準の整備や、機械学習による自動認定を目指すことが必要になると考えられる。そこで、RUA の各種認定における問題点とその解決について検討を行うこととした。田中・浅原(2016)では、『現代日本語書き言葉均衡コーパス』(Mackawa et al. 2014) (以下 BCCWJ)に収録されているインターネット上の QA サイト、Yahoo!知恵袋への投稿を対象に、複数の作業員による作業結果から明らかになった中核要素の認定に

係る問題について報告した。本発表では、発話機能の認定について、報告する。2 節で RUA の手順について説明し、3 節で分析について報告し、4 節でまとめと今後の課題について述べる。

2. RUA の手順

2.1. RUA の概要

上述のとおり、RUA による修辞機能の特定と脱文脈化程度の確認の手順は、1. メッセージとその種類の認定、2. 発話機能・中核要素・現象定位の認定、3. 修辞機能の特定と脱文脈化指数の確認、である²。表 1 に示したように、発話機能・中核要素・現象定位の組合せから修辞機能を特定し、割り当てられた脱文脈化程度を特定することができる。脱文脈化程度は、図 1 に示したように、[1]が最も低く、より「今ここ」に近く、[14]が最も高く、より「今ここ」から遠い。

本稿では、1. メッセージの認定 (§ 2.2)と2. 発話機能の認定 (§ 2.3)のみを対象とする。

2.2. メッセージへの分割と種類の認定

まず、分析対象であるテキストをメッセージ単位に分割(segment)する。メッセージは原則として節を最小単位として表わされるものと捉える。主部や述部が省略されていると考えられる場合には補足してメッセージへの分割、統合を行う。図 2 にメッセージの分類を示す。

表 1 中核要素・発話機能と修辞機能

			発話機能						
			命題						
			現象定位						
			現在		過去	未来		仮定	
非習慣的・一時的	習慣的・恒久	意図	非意図						
中核要素	状況内	参加	[1]行動	[2]実況	[7]自己記述	[3]状況内回想	[4]計画	[5]	[6]
		非参加	n/a		[8]観測		状況内予想		状況内推測
	状況外			[9]報告	[13]説明	[10]状況外回想	[11]予測	[12]推量	
	定言		n/a	[14]一般化					

「n/a」は該当なし。背景が灰色の部分に脱文脈化指数と修辞機能の種類

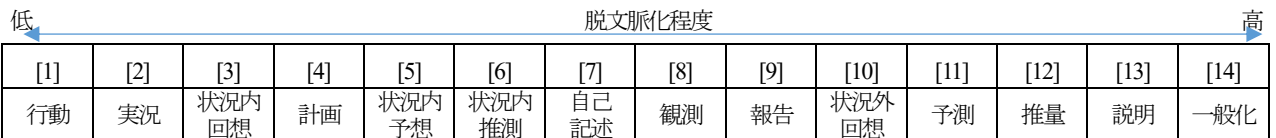


図 1 修辞機能と脱文脈化程度

¹ 詳細は佐野(2010a,b), 佐野・小磯(2011), 田中(2013)を参照のこと

依った。

² 各種認定および用語は原則として佐野(2010a), 佐野・小磯 (2011) に

がら、少数ではあるが「提言」が出現し、その「提言」の認定に問題があることがわかり、認定の不一致も多く見られた。本発表では、発話機能の認定における問題について報告する。

3. 分析

Yahoo!知恵袋の中カテゴリ「コスメ、美容」の質問と回答120セット(1480メッセージ)を対象に、2名の作業員(日本語学を勉強した修士修了生)によって、発話機能・中核要素・現象定位の認定を行い、現在、(1)認定の一致/不一致、(2)認定が不一致の場合の原因、について検証を行っている。以下に、発話機能の定量的分析と定性的分析を行う。

3.1. 定量的分析

発話機能の認定結果の分割表は表3のとおりであった。

2.3 で述べたように、Yahoo!知恵袋はインターネット上のコミュニケーションであり、情報のやり取りが行われる場で、基本的に品物や行為のやり取りはなく、発話機能は「命題」であると考えられる。1,480件のメッセージのうち、約75%が両者から「命題」と認定され、両者とも「提言」と認定したメッセージは1%以下であった。一方、認定の不一致も見られた。2名の作業員による発話機能の認定がどれだけ信頼性のあるものかを検討するためにカッパ係数を求めたところ、 $k = -.0005$ と信頼性は低い。この不一致の要因を明らかにする。

表3 作業員2名による発話機能認定結果

作業員Y \ 作業員X	提言	命題	認定なし	総計
提言	6 0.4%	71 4.8%	0 0.0%	77 5.2%
命題	2 0.1%	1,106 74.7%	258 17.4%	1,366 92.3%
認定なし	0 0.0%	29 2.0%	8 0.5%	37 2.5%
総計	8 0.5%	1,206 81.5%	266 18.0%	1,480 100.0%

3.2. 定性的分析

3.2.1. 不一致:「命題」と「提言」

(3)(4)は、作業員Xが「命題」と認定し作業員Yが「提言」と認定したものである。

(3) あと、韓国の女優さんとかは一重でも目がパッチリしているので一重のメイクは勧告の女優さんを見習って何かの雑誌で読んだことがあります。(OC09_02520(A))

(4) アイプチくらいでショック受けんなよ～！
(OC09_02604(A))

作業員Yは(3)(4)の発話機能を「提言」としている。(3)では「って何かの雑誌を読んだことがあります」を、前掲の(2)で述べた、話し手自身や他者による発話や考えが発話の中に引用されている投射ととらえて「一重のメイクは勧告の女優さんを見習え」の部分を対象に「提言」と認定し、(4)については質

問投稿者への行動の要求として「提言」と認定したが、作業員Xはいずれも情報の提供として「命題」と認定した。

一方、作業員Xが「提言」と認定し作業員Yが「命題」と認定したものは、(5)(6)(7)のように質問投稿で教示を依頼したり投稿を促したりするもの(38件)、(8)(9)のように回答投稿においてアドバイスを提供するもの(31件)、「がんばって」を含むもの(2件)であった。

- (5) 回答に心当たりのあるかた、教えてください。
(OC09_00237(Q))
- (6) 化粧を落とすクレンジングについてアドバイスお願いします。
(OC09_02508(Q))
- (7) こんな私に厳しい湯を入れて下さい。(OC09_02537(Q))
- (8) ミネラルウォーターをたくさん飲んでください！
(OC09_02469(A))
- (9) 手の甲などで余分な粉を落としてからつけて下さい。
(OC09_02492(A))

Yahoo!知恵袋というインターネット上のコミュニケーションは、基本的には情報のやり取りが行われる場面であるが、雑誌からの引用である「見習え」や、質問者に対する「ショック受けるな」「飲んでください」のようなアドバイスや、質問者からの「教えてください」という定型表現などを、行為の要求とみるか、情報提供・情報要求とみるか、判断が分かれていることが不一致の要因の一つであることがわかった。

3.2.2. 不一致:「命題」と「認定なし」

次に、どちらか一方の作業員が発話機能を認定していないメッセージについて確認を行う。表4、表5に、作業員Xと作業員Yの一方が「命題」と認定し、もう一方が認定しなかったメッセージについて、そのメッセージの種類や性質を確認し、本来発話機能をどう認定すべきであったかをまとめた。図2に示したように、RUAの認定は、メッセージの種類によって単独で行う場合、行わない場合、従属するメッセージとあわせて行う場合がある。(10)(11)において(a)(b)は従属節である。作業員Xは、(a)(b)(c)それぞれを「命題」と認定し、作業員Yは一つにまとめて「命題」と認定している。これは、(a)(b)のメッセージの種類を作業員Xは、「拘束;形式的従属」と判断し、作業員Yは「拘束;意味的従属」と判断したためで、その結果、作業員Yでは(a)(b)それぞれ単独には認定されず、作業員Xの認定とは不一致となった。

- (10)
 - (a) 顔全体のニキビがひどく、
 - (b) その上とても顔が乾燥していて
 - (c) ボロボロです。(OC09_00289(Q))

6 原文ママ

- (11)
- (a) 化粧品や乳液などではなくて、
- (b) 下地、ファンデーションなどで、皮脂を押さえたり、
- (c) 化粧品ずれの少ないものを教えてください。
- (OC09_02469(Q))

表 4 発話機能不一致要因(作業員 X「命題」、作業員 Y 認定せず)

発話機能	メッセージの種類や問題となる性質	出現数	割合
「命題」ではなく認定不要	「拘束;意味的従属」	148	57.4%
	「拘束;意味的従属」の中での列挙の判断	22	8.5%
	投射の扱い	2	0.8%
「命題」でなく「位置づけ」とすべきもの	署名	8	3.1%
	箇条書き	6	2.3%
「命題」であるもの	「拘束;形式的従属」の判断	60	23.3%
	「拘束;意味的従属」+句点(〜から。)	7	2.7%
	「自由」の認定	2	0.8%
検討が必要なもの	カッコ内や、前の文の補足の扱い	2	0.8%
	体言止めの主語の扱い	1	0.4%
	計	258	100.0%

表 5 発話機能不一致要因(作業員 X 認定せず、作業員 Y「命題」)

発話機能	メッセージの種類や問題となる性質	出現数	割合
「命題」ではなく認定不要	「拘束;意味的従属」	2	6.9%
	投射	5	17.2%
	文内	2	6.9%
	修飾句内	1	3.4%
	埋め込み節	1	3.4%
	列挙の1つ	1	3.4%
「命題」であるもの	「拘束;形式的従属」	10	34.5%
検討が必要なもの	カッコ内の扱い	6	20.7%
	文末の「…」の判断	1	3.4%
	計	29	100%

2.2 で述べたように、メッセージに分割する際、あるメッセージが状況(時間・場所・原因・結果・条件等)を説明している場合には、「拘束;意味的従属」として従属するメッセージの一部と考え、単独ではその後の認定を行わない。一方、意味的には並列で、時制表現などのために形式的に従属しているものは「拘束;形式的従属」で、単独でその後の認定を行う。発話機能不一致における要因の内訳をみると、表 4 では 80%以上が、表 5 では 40%が、この「拘束」の判断に係るものであることがわかる。

「拘束;意味的従属」と「拘束;形式的従属」の判断は、その連用の意味が分かりにくい時に難しさが生じるものと考えられる。例えば、(10)(b)「乾燥して」と(11)(a)「ではなくて」はともに「テ形」だが、前者は「拘束;形式的従属」、後者は「拘束;意味的従属」と考えられる。益岡(2013:178)は、連用複文構文にかかわる中立型接続とテ形接続の関係について、「連用複文構文における中立型接続は単純列挙(並列)、及び、そこから派生する部分的な連用関係(継起/因果の関係)を表す。他方、テ形接続は「テ」の不可(組み込み)により、広義の連用関係(「並列・時間・論理(広義因果)・様態」の関係)を

全面的・明示的に表す。」としている。いずれの型にも「因果」が含まれていることから、判断の難しさがうかがえる。

また、表 4、表 5 ともに、カッコ内の扱いや、前の文の補足の扱いなど、どこまでを1つのメッセージとするか、基準の明確化が求められるものがあることがわかった。

4. まとめと今後の課題

本発表では、2名の作業員による RUA 認定作業結果のうち、そのメッセージが品物や行為の交換か、情報の交換かを示す「発話機能」の結果を確認し、不一致のものについてその要因を検討した。その結果、インターネット上の Q&A サイトにおける情報要求表現や行為要求表現の「発話機能」との捉え方に違いがあることが明らかになった。また、「発話機能」認定以前のメッセージの種類認定における、「拘束;意味的従属」と「拘束;形式的従属」の認定の違いや、カッコ内のとらえ方の違いなどが、その後の「発話機能」認定の不一致に影響を与えることが確認された。

今後の課題として、インターネットという空間における品物・行為の交換と情報の交換の解釈の再検討と、メッセージの種類の実証的な認定への基準策定を進める。また、節認定ツールの活用も検討していく。

謝辞

本研究の一部は科研費基盤(C)「修辞機能」と「脱文脈化程度」の観点からのテキスト分析手法確立と自動化の検討(15K02535)によるものです。

参考文献

- Cloran, C. (1994) *Rhetorical Units and Decontextualisation: an Enquiry into some Relations of Context, Meaning and grammar*. Nottingham: University of Nottingham.
- Cloran, C. (1999) 'Contexts for learning.' In F. Christie (ed.) *Pedagogy and the Shaping of Consciousness*, London: Cassell, 31-65.
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48(2), 345-371.
- 佐野大樹(2010a)日本語における修辞ユニット分析の方法と手順 ver.0.1.1—選択体系機能言語理論(システム理論)における談話分析—(修辞機能編) <http://researchmap.jp/systemists/>資料公開 (RUA の方法と手順 ver.0.1.1) 2016/1/11 閲覧
- 佐野大樹(2010b)「選択体系機能言語理論を基底とする 特定目的のための作文指導方法について—修辞ユニットの概念から見たテキストの専門性—」『専門日本語教育研究』12, pp.19-26.
- 佐野大樹・小磯花絵(2011)「現代日本語書き言葉における修辞ユニット分析の適用性の検証—「書き言葉らしさ・話し言葉らしさ」と脱文脈化言語・文脈化言語の関係—」『機能言語学研究』6, pp.59-81.
- 田中弥生(2013)「評価の高低によるクチコミサイト「アットコスメ」における談話構造の特徴—修辞ユニット分析を用いて—」『神奈川大学 言語研究』35, pp.1-23.
- 田中弥生・浅原正幸(2016)「Yahoo!知恵袋における修辞ユニット分析の中核要素認定に関する諸問題」『言語処理学会第22回年次大会(NLP2016)』
- 田中弥生・佐野大樹(2011)「Yahoo!知恵袋における質問と回答の分類—修辞ユニット分析を用いた脱文脈化・文脈化の程度による検討」『社会言語科学会第27回大会発表論文集』, pp.208-211.
- 益岡隆志(2013)『日本語構文意味論』くろしお出版