

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

田中弥生[†] 浅原正幸^{††}

[†]神奈川大学 外国語学部・人間文化研究機構 国立国語研究所 理論・構造研究系

^{††}人間文化研究機構 国立国語研究所 言語資源研究系

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1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、修辞ユニット分析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.

本発表では、『現代日本語書き言葉均衡コーパス』(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!知恵袋を対象とする、対面等のコミュニケーションとは異なる空間的距離を表す「中核要素」認定における問題点について、複数の作業員による作業結果から明らかになったことを報告する.

2. 中核要素認定の概要

中核要素はメッセージの中心になるものがコミュニケーションの場面に存在するか否かによって特定する要素で、基本的には主語によって表現される. 中核要素の分類を図1に示す. 中核要素はまず「状況内要素 co-present entity」「状況外要素 absent entity」「定言要素 generalised entity」のいずれかに分類し、「状況内要素」はさらに「参加要素 interactants」「非参加要素 non-interacting entity」に分類する.

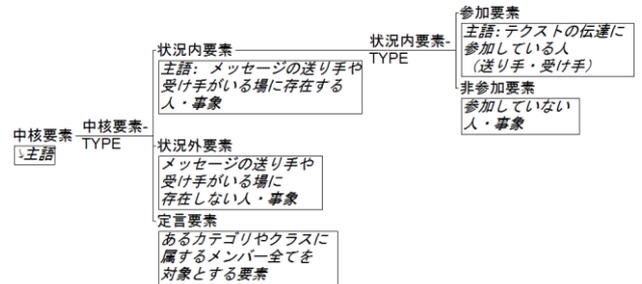


図 1: 中核要素の分類(佐野・小磯 2011)

RUA はもともと英語の会話、すなわち話し言葉を分析するために用いられたもので、日英語の違いや、書き言葉における「発話機能」、「中核要素」、「現象定位」の定義における難しさに直面することもあった. 本研究の分析対象のインターネット上のコミュニケーションである Yahoo!知恵袋への投稿は、物理的な空間は共有しておらず、時間も共有していないため時間差(タイムラグ)があり、書き言葉が用いられる、というように、対面コミュニケーションとは異なる部分がある. 図2に、田中・佐野(2011)にて示した Yahoo!知恵袋「コスメ, 美容」における分類を示して、以下で各要素について説明する.

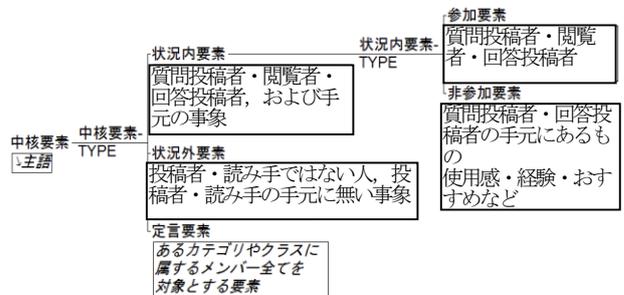


図 2. Yahoo!知恵袋「コスメ, 美容」における中核要素

2.1. 状況内要素

主語が「メッセージの送り手や受け手がいる場に存在する人・事象」である場合に「状況内要素」と認定され、さらに「参加要素」と「非参加要素」に分類される.

2.1.1. 状況内;参加要素

「状況内;参加要素」は主語が「メッセージの送り手や受け手がいる場に存在する人・事象」でなお且つ「テキストの伝達に参加している人」で、基本的には一人称と二人称が該当す

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

る。典型的なものは(1)のような「私は」である。(2)のように省略されている主語は復元して認定する。また、対面等では直接のコミュニケーションの相手である「あなたは」や娘から父親への質問「お父さんビール飲む？」の「お父さん」が「状況内;参加要素」である。QA サイトは通常特定の相手に向けられてはいないが、QA サイトを閲覧したり投稿したりする不特定の相手をさして、相手に向けて用いられる(3)のような「みなさんは」などが使用されることがある。

- (1) **わたしは**²けっこうなくせ毛です。 OC09_00175_Q³
- (2) **φ (=私は)** 普通にヴァセリン塗ってます。
OC09_02724_A
- (3) **みなさんは**ホントにお湯だけで落としていますか？
OC09_02515_Q

2.1.2. 状況内;非参加要素

「状況内;非参加要素」は、「メッセージの送り手や受け手がいる場に存在する人・事象」で、なお且つ「テキストの伝達に参加していない人・事象」である。対面では、例えば両親と娘の3人で会話をしている場合、母親から娘に向けて送られた「昨日お父さんが花束をくれたの」というメッセージの主語「お父さんが」は、その場にいるがそのメッセージの伝達には参加していないので、中核要素は「状況内;非参加要素」となる。また、「このカレーはとても辛い」の「このカレーは」は、その場にあるがテキストの伝達には参加していないため、「状況内;非参加要素」である。QA サイトにおいては、例えば(4)の投稿者の「顔のしみが」がメッセージの送り手がいる場に存在する事象として「状況内;非参加要素」と認定される。また使用感やおススメについても、その場に存在するものと考え、「状況内;非参加要素」と認定した。

- (4) 最近**顔のしみが**どんどん薄くなってきました。
OC09_00109_A
- (5) **わたしのお勧めは**資生堂モアリップです。
OC09_00058_A

2.2. 状況外要素

「状況外要素」は「メッセージの送り手や受け手がいる場に存在しない人・事象」である。QA サイトにおける(6)の「何もしないのに肌きれいな子が」や(7)の「使い捨てコンタクトの使用期限って」が「状況外要素」である。

- (6) 友達にも、**何もしないのに肌きれいな子が**います

² 例文中、中核要素を太字で示す。

³ BCCWJのサンプルID及び質問 (Q) 回答 (A) の別を示す。

が、〈略〉 OC09_02701_Q

- (7) **使い捨てコンタクトの使用期限って**ありますか？
OC09_00184_Q

2.3. 定言要素

「定言要素」は、「あるカテゴリやクラスに属するメンバー全てを対象とする要素」で、例えば「醤油は大豆からできている」の「醤油は」が「定言要素」となる。主語のみでなく述部もあわせて判断する。(8)の「**油分は**」は、油分というカテゴリ全体を指しているため、「定言要素」と認定される。

- (8) 時間が経つにつれ**油分は**酸化します。OC09_02500_A

3. 分析

3.1. 定量的分析

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(1221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及び学部卒業生)によって認定を行い、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。中核要素の認定結果の分割表は表1のとおりであった。

表 1. 作業員2名による中核要素の認定結果

作業員 X \ 作業員 Y	主語不明	状況外	状況内; 参加	状況内; 非参加	定言	提言 ⁴	総計
主語不明	0 0.0%	1 0.1%	5 0.4%	1 0.1%	2 0.2%	0 0.0%	9 0.7%
状況外	4 0.3%	209 17.1%	229 18.8%	211 17.3%	60 4.9%	11 0.9%	724 59.3%
状況内; 参加	0 0.0%	9 0.7%	355 29.1%	38 3.1%	15 1.2%	11 0.9%	428 35.1%
状況内; 非参加	0 0.0%	0 0.0%	2 0.2%	3 0.2%	0 0.0%	0 0.0%	5 0.4%
定言	0 0.0%	3 0.2%	2 0.2%	3 0.2%	10 0.8%	3 0.2%	21 1.7%
提言 ⁴	0 0.0%	1 0.1%	31 2.5%	1 0.1%	1 0.1%	0 0.0%	34 2.8%
総計	4 0.3%	223 18.3%	624 51.1%	257 21.0%	88 7.2%	25 2.0%	1,221 100.0%

2名の作業員による中核要素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、 $k=0.47$ で、中程度の一致とみなされる結果となっている。

2章で述べたように、RUAでは中核要素の分類には「その場に存在するか否か」がかわる。Yahoo!知恵袋は物理的

⁴ 中核要素認定の前段階「発語機能」の認定による。

な空間ではないため、「その場」をどうとらえるかによって「状況内」と「状況外」の認定が分かれる可能性があると考えられる。表1では、2名の作業者が共に「状況内;参加」と認定したメッセージが29.1%と最も多く、共に「状況外」と認定したのも17.1%と多い。しかし、作業者Yが「状況外」と認定しているメッセージを作業者Xが「状況内;参加」(18.8%)、「状況内;非参加」(17.3%)としたものも同程度あり、ゆれが生じている。

このようなゆれは、Yahoo!知恵袋が、物理的空間を持たないインターネット上の空間であるために「メッセージの送り手や受け手がいる場」の認識が人によって異なることに起因する可能性がある。

本発表では、中核要素における不一致の中で、作業者Xが「状況内;参加」と認定し作業者Yが「状況外」と認定した229件の不一致の原因について分析を行った結果を報告する。表2に作業者Xが認定した中核要素の内訳を示す。

表 2. 作業者 X の中核要素内訳

(作業者 X 「状況内;参加」, 作業者 Y 「状況外」と認定分)

		中核要素	件数
明示		私も	1
		みなさんが	1
		誰か	1
非明示(復元)	私	私は	80
		私の場合は	1
		自分	2
	あなた	あなた	4
		あなたは	118
		あなたが	2
		あなたに	1
	皆さん	皆さん	4
		皆さんは	10
		皆さんも	1
		(特定できず) ⁵	3
			計

3.2. 定性的分析

(9)について、作業者Xは「皆さんは」が省略されているとして中核要素を「状況内;参加」と認定しているのに対し、作業者Yは「アイラインとマスカラとビューラーは」を主語として中核要素を「状況外」と認定している。(10)について作業者Xは「あなたは」が省略されているとして「状況内;参加」と認定したのに対して、作業者Yは「香り強めの香水は」を主語として「状況外」と認定している。(11)について、作業者Xはその主語「私の場合は」が省略されているとして中核要素に認定しているのに対し、作業者Bは明示されている「その3点は」を主語として中核要素と認定している。(12)は、作業者Xは「私は」が省略されている主語であると考え、作業者Yは「ソ

イエイは」が省略されていると判断した。これらのように作業者Yでは「マスカラとビューラーは」を中核要素の認定対象としているものが多い見られた。これらは、述語項構造解析を使用することによって主語の認定が行える。

- (9) **φ(=皆さんは)** アイメイクをする時にアイラインとマスカラとビューラーはどの順番でやっていますか? OC09_02496_Q
- (10) そうそう、**φ(=あなたは)** 香り強めの香水は耳たぶの後ろには付けちゃダメだよ。 OC09_02672
- (11) **φ(=私の場合は)** ファンデーションがルースパウダーだけのお休みバージョンのお化粧のときもその3点は欠かさないです。 OC09_02710
- (12) 剃るよりも「ゴマ状態」になるまでの時間が長いので**φ(=私は)** 重宝してます。 OC09_00083

また、助言の中でのやり方やその効果として(13)や(14)のような列挙について作業者Xは「あなたは」が省略されていると判断し、作業者Yは「素肌をきれいにする方法は」や「ミネラルウォーターは」が省略されているとしている。(15)について作業者Xは「あなたは」が省略されていると判断し、作業者Yは「脂取り紙だと」を「脂取り紙は」と同等だと認定している。(16)の回答(A)の下線のメッセージ「目の周りがかすんでいるのかも」について、作業者Xは「あなたは」を、作業者Yは「アイシャドウがきれいにならぬ事は」が省略されているとして中核要素を認定している。これらのように助言の受け手を主語とするか話題になっている事象を主語とするかの違いが見られた。(13)は「目的」(~するには)、(14)(15)は「手段/対策」(~をすると)、(16)は「原因」(~のは)と分類できる。

- (13) Q: 〈略〉 素肌をきれいにするにはどうすればいいんでしょうか? 〈略〉
A: 〈略〉 **φ(=あなたは)** 2時から2時までなるべく熟睡しているようにする。 **φ(=あなたは)** 生ではなく火を通した緑黄色野菜をたっぷり摂る。 **φ(=あなたは)** 体を冷やさないようにする。 〈略〉 OC09_00289
- (14) 上の方の書いている「お手入れにプラスして、ミネラルウォーターをたくさん飲んでください!」
〈略〉 **φ(=あなたは)** 肌の調子もよくなるし、 **φ(=あなたは)** 太らないし、 **φ(=あなたは)** お腹の調子もいいし、 **φ(=あなたは)** 汗の匂いも消える! OC09_02469

⁵ 省略されているものは特定できないが中核要素を「状況内;参加」と認定したもの。

⁶ 脱毛処理器具の名称

⁷ 原文ママ

(15) 脂取り紙だと、取れ過ぎちゃう場合があるのでφ
(=あなたは)要注意です。OC09_00256

(16) Q: アイシャドウがきれいに発色しない事ってありませんか?例えばターコイズっぽい色をまぶたにのせるとグリーンっぽくなってしまったり。色が変わってしまうことを踏まえて使っているの今のところ問題はないのですが、何か解決策はないでしょうか?

A: φ(=あなたは)目の周りがくすんでいるのかも。目の周りはくすみが出やすい部分です。OC09_02730

また、(17)では作業員Xは「あなたは」が省略されていると判断したのに対して、作業員Yは埋め込み節内で省略されている「エステ券は」をこのメッセージ全体の主語と判断している。(18)については作業員Yは「みなさんがイイと思うマスカラは」を主語と判断したのに対して、作業員Xは埋め込み節内の「みなさんが」をこのメッセージ全体の主語と判断している。(19)では、作業員Xは「あなたは」が省略されていると判断したのに対し、作業員Yは「お化粧を落とさないで寝るのは」を主語として中核要素を認定している。これらのような埋め込み節が主語となるメッセージでの認定の違いも見られた。

(17) ちっともお得でない事をφ(=あなたは)知っておきましょう! OC09_02427

(18) みなさんがイイと思うマスカラはなんですか? OC09_02536

(19) でも、φ(=あなたは)お化粧を落とさないで寝るのは危険です! OC09_02701

(20)の下線部は後続の主節について説明する従属節で、「意味的従属」と考える。独立したメッセージとは見なさず、主節に付随するものと考え、主節部分の主語が中核要素となる。また主節の「~と思います」は選択体系機能理論で「投射projection」⁹と呼ばれ、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と考える。「思います」(投射節)は分析対象とせず、「粉をはたく前にティッシュで余分な油を抑えた方がいい」という部分(被投射節)を分析する。作業員Xは「あなたは」が省略されていると判断し、作業員Yは従属節内の「リキッドは」を本メッセージ全体の主語としてとらえた。このように「意味的従属」にかかわる中核要素認定の違いも確認された。

(20) あと、リキッドはカバー力がある分油分もあるので、φ(=あなたは)粉をはたく前にティッシュで余分な油を抑えた方がいいと思います。OC09_02492

4. まとめ

本発表では、2名の作業員による認定作業結果のうち、メッセージの中心になるものがコミュニケーションの場面に存在するか否かによって特定する要素である「中核要素」について、一方の作業員がそのメッセージのやりとりに参加している「状況内;参加」と認定し、もう一方の作業員がその場には存在しない「状況外」と認定したメッセージの不一致の要因について確認を行った。

その結果、「ワ格」を中核要素の認定対象としているもの、助言の受け手を主語とするか話題になっている事象を主語とするかの違い、埋め込み節が主語となるメッセージでの認定の違い、「意味的従属」にかかわる中核要素認定の違い、などがあることが明らかになった。述語項構造解析の利用や、従属節に関わる基準の徹底などによって認定の一致度の向上の可能性がうかがえた。その他の不一致についても検討を続けていく。

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)、国語研基幹型共同研究プロジェクト「コーパスアノテーションの基礎研究」および国語研「超大規模コーパス構築プロジェクト」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.
- 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(システミック理論)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・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.

⁸ RUA では埋め込み節は単独では認定対象としない

⁹ 「投射」については佐野(2010a)を参照のこと