

修辞ユニット分析からみた Q&A サイトの言語的特徴

田中 弥生 佐野 大樹

国立国語研究所 コーパス開発センター

{yayoi, toki.sano}@ninjal.ac.jp

1. はじめに

修辞ユニット分析 (Rhetorical Unit Analysis 以下, RUA) は, 元来話し言葉の談話分析手法であるが, 著者らはこの RUA を Q&A サイトの文章に適用し, 脱文脈化の視点から質問文の分類を試みる研究を進めている (田中・佐野 2011). Q&A サイトは知識伝達の場の一つであるが, 脱文脈化の視点で分析することによって, その知識の一般性・抽象性に焦点をあてた質問の分類が可能になる. 例えば, 同じ対象であっても一般的・抽象的な定義を聞いているのか, 具体的な使い方や個人的な感想を知りたいのか, などという観点から分類できるようになる.

Q&A サイトにおける質問の分類や分析についてはいくつかの先行研究があるが (三浦・川浦 2008, 栗山・神門 2009, 田村ほか 2008, 田中 2010 など), 質問の一般性・抽象性を観点とした分類は行われていなかった. 本研究は, 大規模データの分析を視野に入れ, Q&A サイトにおける質問文を対象にして, RUA の分類基準から見た語彙-文法的な特徴を示すものである.

以下, 2 で RUA を説明し, 3 で分析対象と手順を示した上で, 4 および 5 で具体的な基準を説明し, 6 でまとめと今後の課題を述べる.

2. RUA とは

RUA は, 発話機能, 中核要素, 現象定位の 3 つをメッセージ¹単位で認定して修辞機能を特定するもので, その結果として, 脱文脈化²の程度を知ることができる. RUA による修辞機能と脱文脈化指数を表 1 に示す. 脱文脈化指数の数値が大きいものほど脱文脈化の程度が高く, 小さいものほど脱文脈化の程度が低いことを示す³

表 1. 修辞機能の特定と脱文脈化指数

中核要素		発話機能								
		現象定位								
		現在		過去	未来		仮定			
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] 状況内予想	[6] 状況内推測		
	非参加	n/a	[9] 報告	[8] 観測						
	状況外	n/a	[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
背景が灰色の部分が修辞機能の種類
[]内は脱文脈化指数

¹ 選択体系機能言語理論の意味層 (semantics) における最小単位で, 基本的には節によって表される.
² Cloran(1999)に基づき脱文脈化言語を「一般化された要素の習慣的・恒久的な行動や状態について表現する言語」, 文脈化言語を「物質的状況に存在する要素の現在の行動や状況について表現する言語」とする.
³ 脱文脈化指数の詳細は佐野・小磯(2011)参照.

佐野(2010b)で述べたように, RUA はテキストの意味単位を特定するための手法だが, その過程においてメッセージの修辞機能の種類の特定期にも利用できる. 英語においては母子会話, 教師と生徒の会話等の分析に活用され, 知識伝達の分析に有用な枠組みと考えられているが, 日本語に適用した研究は佐野(2010b), 佐野・小磯(2011)などまだ少ない. また, Web 上のテキストについて, 田中・佐野(2011)で Q&A サイト Yahoo!知恵袋における, 中カテゴリ「コスメ, 美容」の「質問」へ RUA を適用し, 修辞機能と脱文脈化が図 1 のように分布し, 一般性, 具体性の異なる質問がみられることを示しているが, 語彙-文法的な特徴の記述は行っていない. 本研究は, RUA の中核要素と現象定位の種類ごとの語彙-文法的な特徴を示すものである.

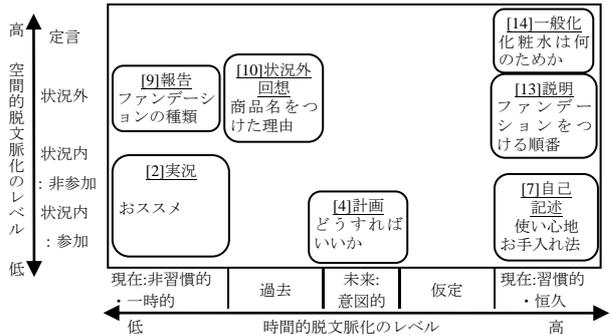


図 1. Yahoo!知恵袋「コスメ, 美容」の修辞機能と脱文脈化の分布

3. Q&A サイトへの RUA の適用基準

3.1. 分析対象

分析には, 『現代日本語書き言葉均衡コーパス』領域内公開データ (2009 年度版) に収録されている「Yahoo!知恵袋」データを利用した. 中カテゴリ「コスメ, 美容」の質問文 1 つから成る「質問」90 投稿のうち, 国立国語研究所(1960,1963)で示されている「要求表現」の「質問的表現」に含まれる表現が使用されている 59 投稿を分析対象とした. RUA では, 当該のメッセージだけではなく他のメッセージの照応なども考慮して中核要素と現象定位を認定するが, 本研究では, 分類基準となる語彙-文法的な特徴を示すパイロットスタディとして, 照応を考慮する必要のない単一の質問文から成る「質問」投稿を対象とした.

3.2. 分析手順

本研究で用いる RUA の詳細は, 佐野(2010b), 佐野・小磯(2011)で述べた. 分析では, 1.分析対象のメッセージと種類の認定, 2.発話機能の認定, 3.中核要素の認定, 4.現象定位の認定, 5.修辞機能の特定と脱文脈化指数の確認, を行う. 1.から 4.によって発話機能, 中核要素, 現象定位が認定されれば, 5 の修辞機能と脱文脈化指数が決まる. ここでは, 1.分析対象のメッセージと種類の認定と 2.発話機能の認定について述べる.

3.2.1. 分析対象のメッセージと種類の認定

RUAでは、テキストをメッセージという単位に分割し、メッセージの種類を認定する⁴。メッセージは、「位置付け positioning」「拘束 bound」「自由 free」のいずれかに分類し、さらに「拘束」は「拘束：意味的従属」と「拘束：形式的従属」に分類する。「位置付け」は、挨拶・定型句・フィラーなど述部を含まない節のみによって構成されるものである。「拘束：意味的従属」は、従属するメッセージの状況（時間・場所・原因・結果等）を説明するもので、従属しているメッセージの一部と考える。「拘束：形式的従属」は、意味的には並列の関係であるが、時制（過去）などの側面で、従属するメッセージに形式的に依存するものである。「自由」は独立して時制やムードなどを表わすものである。RUAでは、これらのうちの「自由」と「拘束・形式的従属」について、修辞機能の認定を行う。(1)は「自由」の例である。(2)は、【】⁵内の「T字型カミソリで剃っていると」が原因を示しているため、「拘束：意味的従属」で、それ以外の部分が「自由」である。また、(3)は、「こういう長い連休でずっと家にいる方」は「髭そったり」「髪をセットしたり」に関わるが便宜上 a に含め、a. と b. は並列となる。(3)a. は「拘束：形式的従属」、(3)b. が「自由」である。

- (1) ヒリヒリしみる化粧水って、自分に合わないってことでしょうか？
- (2) 男性のヒゲは、【T字型カミソリで剃っていると】濃くなるんですか？
- (3) a. こういう長い連休でずっと家にいる方、髭そったり、
b. 髪をセットしたりしますか？

3.2.2. 発話機能の認定

メッセージの種類を認定した後、発話機能を認定する。発話機能は、「提言」か「命題」のどちらかに分類する。「提言」は品物・行為の交換に関するメッセージ、「命題」は情報の交換に関するメッセージが該当する。例えば、(4)は「お塩を取る」という行為を要求する「提言」で、(5)は「おいしいお塩」についての情報を要求する「命題」である⁶。

- (4) お塩を取っていただけますか？
- (5) おいしいお塩を教えてくださいませんか？

上述の(1)(2)(3)の発話機能は、いずれも「命題」である。以降で、中核要素と現象定位の認定基準となる言語的特徴について説明する。

4. 中核要素の認定の基準となる言語的特徴

中核要素はメッセージの中心となる要素で、基本的には主語によって表現される。田中・佐野(2011)で図2に示したように、中核要素はまず「状況内要素」「状況外要素」「定言要素」のいずれかに分類し、「状況内要素」はさらに「参加要素」「非参加要素」に分類する。

⁴ メッセージの種類認定については、佐野(2010a)参照。
⁵ 【】部分は、メッセージの種類が「拘束：意味的従属」であることを示す。
⁶ (4)(5)は筆者の作例である。

図3は、中核要素を認定する際の指標となる語彙-文法的特徴のフローである。このフローにそって、以下に中核要素認定にかかわる語彙-文法的特徴を示す。

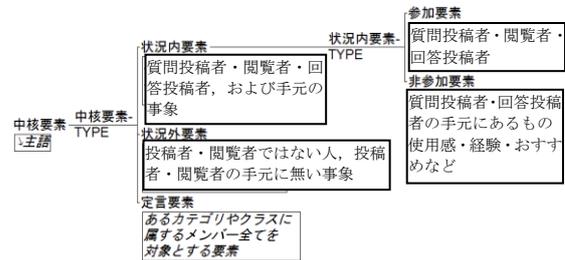


図2. Yahoo!知恵袋「コスメ、美容」における中核要素 (田中・佐野 2011)

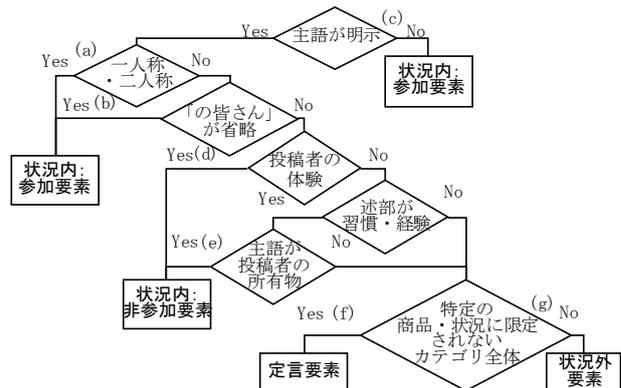


図3. 中核要素認定のためのフロー

4.1. 状況内要素

4.1.1. 状況内：参加要素

「状況内：参加要素」は、基本的には主語が一人称・二人称の場合ある。まず、主語が明示されているか否かを確認する。

- (a) 主語として一人称・二人称（「わたしは」「あなたは」「皆さんは」）が明示されている場合。
- (6) 朝のぼさぼさの髪をみなさん⁷どうやってセットしてますか？
- (b) 主語に一人称・二人称以外が明示されていて、「の皆さん」が省略されていると考えられる場合。
- (7) 30代前半の女性は若い女性が羨ましいですか？⁸
- (8) 最近の若い男性は、脇の毛を処理したりするんですか？⁹

なお、(7)では述部が感情に関わる表現であるため、主語は二人称であると判断できるが、(8)のように述部が感情に関わる表現以外の場合には、該当する人以外でも回答することが考えられ、主語が二人称相当であるとは必ずしも言い切れない。この認定については、質問投稿のみからでは困難である。本研究では、統一して、質問投稿者は、該当する人が回答することを想定していると考え、(8)のような場合にも二人称相当と判断することとした。

⁷ 中核要素を太字ゴシック体で示す。
⁸ 「30代前半の女性のみなさん」と考えられる。
⁹ 「最近の若い男性のみなさん」と考えられる。

(c) 一文のみで構成されている質問で主語が明示されていない場合。

(9) ϕ クリスマス限定キット何を買う予定ですか？

(10) ϕ 髭剃りって毎日やっていますか？

(11) 【肌をキレイにするには】 ϕ どうすればいいですか？

(9)の「～する予定ですか」のように、今後の意図できる行動を聞く場合、(10)のように「使っていますか」「やっていますか」など、習慣的な行動を聞く場合は、二人称が省略されていると考える。(11)の「どうすればいいですか？」のように今後のとるべき行動を聞いている場合には、一人称が省略されていると考える。

4.1.2. 状況内：非参加要素

(d) 主語が一人称・二人称以外で「おすすめ」「使用感」など体験が含まれている場合。

(12) **リキッドアイライナーでお勧めって何ですか**？

(e) 述部が習慣・経験で主語が投稿者の所有物等の場合。

(13) **ファンデーションなどどのくらいの期間でなくなりませんか**？¹⁰

4.2. 定言要素

(f) 主語が一人称・二人称以外で特定の商品や特定の状況に限定されていない場合で、尚且つ述部が「どのようなもの」「何のこと」のように性質を問うものが使用されている場合。

(14) **収れん化粧水とはどのようなものですか**？

4.3. 状況外要素

(g) 主語が一人称・二人称以外で特定の商品や特定の状況に限定されなくて述部が性質を問うもの以外の場合、あるいは、主語が一人称・二人称以外で特定の商品や特定の状況に限定されている場合。

(15) 【ニキビ跡のシミをなくすには、】**何が**一番良いですか？

(16) **一番くずれない、取れないファン**で¹¹はなんですか。

(17) **鼻の頭の黒いぶつぶつを上手に取る方法**はありませんか。

5. 現象定位の認定の基準となる言語的特徴

次に現象定位の認定について説明する。現象定位はメッセージが伝達されている時 (Time of speaking 以下、Ts) を基準として、メッセージによって表現されている出来事がいつ起こったかを示す要素である。現象定位はまず「現在」「過去」「未来」「仮定」に分類し、「現在」はさらに「非習慣的・一時的」「習慣的・恒久」に、「未来」は「意図的」「非意図的」に分類する。田中・佐野 (2011) で提示した

¹⁰ 「あなたのファンデーションなどは」と考えられる。

¹¹ 原文のまま

Yahoo!知恵袋における分類の基準を図4に示す。

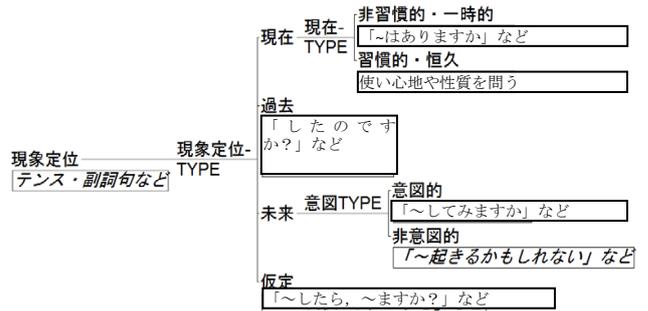


図4. Yahoo!知恵袋「コスメ、美容」における現象定位 (田中・佐野 2011)

図5は、現象定位を認定する際の指標となる言語的特徴のフローである。このフローにそって、以下に現象定位認定にかかわる言語的特徴を示す。

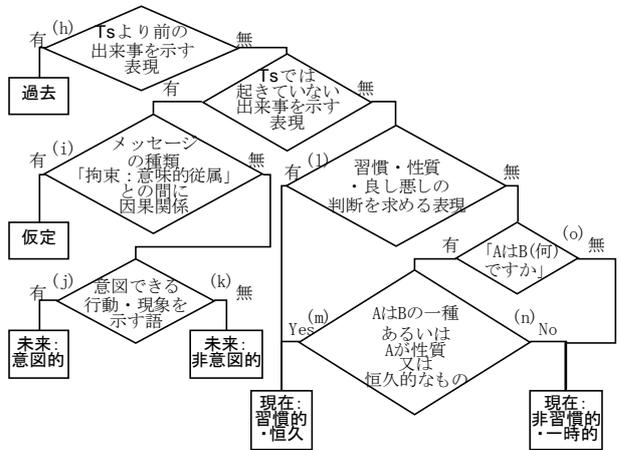


図5. 現象定位認定のためのフロー

5.1. 過去

(h) Tsより前に起こった出来事である表現がある場合。

(18) **ジョンソンから新発売のボディソープ「ソフトウォッシュ」を使ってみた方、どんな感じでしたか**？¹²

5.2. 仮定

Tsでは起きていない出来事を示している場合、現象定位は「仮定」か「未来」である。

(i) メッセージの種類「拘束：意味的従属」との間に因果関係が見られる場合。

(19) 【**まめに塗り直せば、**】**【SPFの低いもので海辺にいても】** ϕ 大丈夫でしょうか

5.3. 未来

Tsでは起きていない出来事で「仮定」でなければ、「未来」になり、意図できるか否かで「意図的」と「非意図的」に分かれる。

¹² 現象定位をイタリック体で示す。

5.3.1. 未来：意図的

- (j) 意図できる行動・現象を示す語がある場合。
- (20) ϕ クリスマス限定キット何を買う予定ですか？
(9)再掲
- (21) 【肌をキレイにするには】 ϕ どうすればいいですか？ (11)再掲

5.3.2. 未来：非意図的

- (k) 意図できる行動・現象を示す語がない場合。
- (22) ちふれ化粧品の福袋は出るのでしょうか？

5.4. 現在

Ts の時点に起きているかどうか問うものは「現在」である。また、過去から未来へと続いていくものもここに含まれる。習慣性・恒久性があるかないか、一時的かによって「習慣的・恒久」と「非習慣的・一時的」に分かれる。

5.4.1. 現在：習慣的・恒久

- (l) 習慣、性質、よし悪しの判断を求める表現がある場合。
- (23) ϕ 髭剃りって毎日やってますか？ (10)再掲
- (24) ϕ シャンプー・トリートメント、なに使ってますか？
- (25) おしろいって ϕ 【下地とかしないで】そのままつけていいんですか？
- (26) 取れん化粧水とはどのようなものですか？ (14)再掲
- (m) 「AはBですか」の形式において、AがBの一種(スクラブはあるカテゴリの中の一つ)の関係にあるか、あるいはAが性質や本質的なもので、Bにおいて性質が述べられる(スクラブ=何)場合。
- (27) スクラブってなんですか？

5.4.2. 現在：非習慣的・一時的

- (n) 「AはBですか」の形式において、Aが行為や出来事に関わることで、Bにおいて性質以外が述べられる(リキッドアイライナーでお勧め=何)場合。
- (28) リキッドアイライナーでお勧めって何ですか？ (12)再掲
- (o) 存在の有無を聞くものなど
- (29) 顔の毛穴を目立たなくするお勧めの化粧品、コスメありませんか？

6. まとめと今後の課題

本研究では、Q&Aサイトの質問文にRUAを適用するための分類基準となる言語的特徴を、Yahoo!知恵袋の中カテゴリ「コスメ、美容」内の1文で構成されている「質問」投稿を利用して記述した。日本語の質問文やWeb上のテキストへRUAを適用するためのこのような記述はこれまで行われておらず、これによって、RUAの適用の自動化など、将来的な大規模データの分類も検討でき、質問の一般性・抽象性を観点とした分類が可能になると考えられる。

本研究の分析対象カテゴリが「コスメ、美容」であったための基準も含まれていることから、今後はコスメ以外の異なる話題への適用の可否を確認し、Q&Aサイトの質問に汎用的な基準と言語的特徴を検討していく予定である。また、回答についても言語的特徴を記述していくことによって、Q&Aサイト全体へRUAを適用できるようになれば、質問と回答の修辞機能や脱文脈度の程度の対応関係を知ることができ、また、修辞機能や脱文脈化指数を投稿に付与すれば、サイト利用者が検索する際に一般性や個別性を指定して、知りたい情報を抽出するのに役立つと考えられる。さらに今後は、カテゴリ間の差異や、Q&Aサイトの違いによる異同などを分析して、話題やコミュニティの違いと、一般性・抽象性との関連性なども、明らかにしていきたいと考えている。

謝辞

本研究は、文部科学省研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築:21世紀の日本語研究の基盤整備」(平成18~22年度、領域代表者:前川喜久雄)、及び、科学研究費補助金(基盤C)「書き言葉コーパスに基づくテキスト分類尺度の探索的研究」(平成21年度~23年度、研究課題番号:21520493、研究代表者:小磯花絵)による補助を得ています。本研究では、『現代日本語書き言葉均衡コーパス』領域内公開データ(2009年度版)に含まれる「Yahoo!知恵袋」のデータを利用させていただきました。記して感謝の意を表します。

文献

- Cloran,C.(1999)Contexts for learning. In Christie, F. (ed.) Pedagogy and the Shaping of consciousness,London: Cassell, 31-65
- 国立国語研究所(1960).話しことばの文型 1 一対話資料による研究 秀英出版
- (1963).話しことばの文型 2 一独和資料による研究 秀英出版
- 栗山和子・神門典子(2009).「Q&Aサイトにおける質問と回答の分析」. 情報処理学会研究報告第95回情報学基礎研究会. 2009-FI-95.
- 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 / 閲覧日 2011.1.19
- 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. 掲載頁未定
- 田中弥生(2010).「Q&Aサイトの「質問-回答」における結束性-省略の特徴分析-」『信学技報』109(390), NLC2009-34, pp.7-12.
- 田中弥生・佐野大樹(2011).「Yahoo!知恵袋における質問の修辞ユニット分析-脱文脈化-文脈化の程度による分類-」『信学技報』, 110(400), NLC2010-32, pp. 13-18
- 田村元秀・村上仁一・徳久雅人・池原悟(2008).「Web検索エンジンを用いたWhy型質問応答システムに関する研究(特許分類)」『情報処理学会研究報告. 自然言語処理研究会報告』2008(4), pp.15-21.
- 三浦麻子・川浦康至(2008).「人はなぜ知識共有コミュニティに参加するのか:質問行動と回答行動の分析」『社会心理学研究』23(3), pp.233-245.