

従属節意味分類アノテーションに基づく修辞ユニット分析の分類単位認定の検討

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

筆者らは修辞ユニット分析 (Rhetorical Unit Analysis, 以下 RUA とする) の日本語における認定基準と自動化の検討を行っている。本発表では、松本(2018)が検討し『現代日本語書き言葉均衡コーパス』(Mackawa et.al 2014, 以下, BCCWJ) に付与を行っている従属節意味分類アノテーション情報を、RUA の分類単位であるメッセージ(おおむね節)の認定の際に利用することを検討し、報告する。

RUA は、Cloran(1994, 1995)によって提案された選択体系機能言語理論における英語の母子会話を対象とした談話分析手法の一つで、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」(主に主語から判断する。空間的距離を示す。), 「現象定位」(副詞や述部などから判断する。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」を特定し、修辞ユニット構造を確認するものである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、チラシやその作成の振り返り文、児童・生徒作文などの分析に使用し、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や可能性を示してきた(田中 2018a, 2018b 他)。修辞機能や脱文脈化の観点による、従来にはなかった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や教育現場などで広く利用されることが期待できる。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、機械処理による自動化を目指し、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っている。

RUA では、まずメッセージの種類を認定するが、従属節の種類を明確に分類することが必要である。現代日本語の従属節の意味分類に関連する研究には、丸山他(2004)、佐藤・丸山(2015)、佐藤他(2016)などがあり、松本(2018)は「鳥バンク」基準互換の日本語従属節の意味分類基準について再検討を行っている。本研究では、松本(2018)を基準として従属節の種類がアノテーションされているデータを用いて、RUA の認定に従属節意味分類アノテーションを活用することを検討する。以下、2 節で利用データについて説明し、3 節で RUA の手順とメッセージ認定について述べ、4 節で従属節意味分類アノテーション情報を利用したメッセージ分類について報告し、5 節でまとめと今後の課題について述べる。

2. 利用データ

本研究で使用したデータは BCCWJ に格納されている「特定目的サブコーパス「教科書」レジスター」コアデータの「国語科」「小学校」のうち 4 年生のデータである。サンプルは 3 件で、表 1 は形態素解析したデータの「文」項目で文頭を意味する「B」が付与されている数である。基本的には文の数に相当するが、箇条書きなどの場合にも 1 項目に 1 つ付与される。

表 1. 使用サンプルと文の数

	OT01_00006	OT01_00018	OT01_00030	合計
文頭	200	265	155	620

3. RUA の手順

3.1. RUA の概要

RUA 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、の順にすすめる。表 2 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確認することができる。

表 2. 発話機能・中核要素・現象定位からの修辞機能と脱文脈化指数

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

脱文脈化程度は指数で表されるが、図1に示したように、[1]が最も低くてより「今ここ」に近く、[14]が最も高く、より「今ここ」から遠い。RUA を用いたテキストの分類によって、修辞機

¹ 詳細は佐野(2010a,b)、佐野・小磯(2011)を参照のこと

能と脱文脈化程度の展開や様相など、そのテキストの特徴を知ることができる。

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]
行動	実況	状況内回想	計画	状況内予想	状況内推測	自己記述	観測	報告	状況外回想	予測	推量	説明	一般化

← 低 脱文脈化指数 高 →

図 1. 修辞機能と脱文脈化指数

3.2. メッセージの確認

上述のように、RUA では、まず分析対象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。図 2 にメッセージの分類を示す。

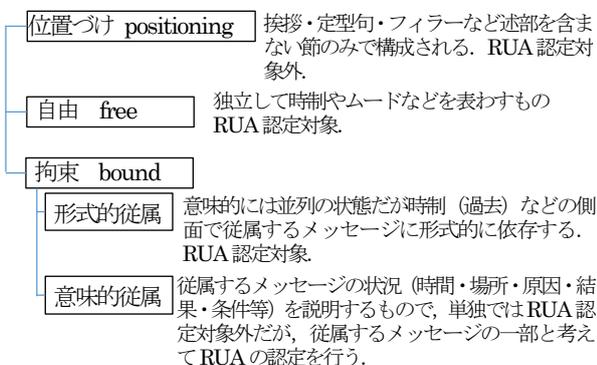


図 2. メッセージの分類

メッセージは、「位置づけ positioning」、「自由 free」、「拘束 bound」に分類する。(1)は「位置づけ」の例である。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、RUA の認定の対象とはしない。(2)は「自由」の例である。「自由」は独立して時制やムードなどを表わすもので認定対象となる。「拘束」は「拘束;形式的従属」と「拘束;意味的従属」に分けられる。(3)の下線部分は「拘束;形式的従属」の例である。意味的には並列の関係だが時制(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、RUA の認定の対象となる。(4)の下線部分は「拘束;意味的従属」の例である。従属するメッセージの状況(時間・場所・原因・結果・条件等)を説明するもので、単独では RUA 認定の対象とはせず、従属するメッセージの一部と考える。

- (1) ほう、ほう。(OT01_00018²)
- (2) ふつう、だえきは一日におおよそ一リットルほど出ます。(OT01_00030)

² カッコ内にサンプル ID を示す。

³ 「投射」については佐野(2010a)を参照のこと

- (3) 「物語」が多いのは予想どおりでしたが、他の種類の本を読む人が、思いの外たくさんいました。(OT01_00006)
- (4) 風がもうひとふきすれば、車がひいてしまうわい。(OT01_00018)

なお、「～と言った」「～と思う」「～と考える」「～だと聞いた」など、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られる場合は、引用された部分(被投射節 prefaced)を分析対象とする³。例(5)では、「なんだろう」を分析対象とする。

- (5) なんだろうって思いました。(OT01_00018)

3.3. 発話機能・中核要素・現象定位の認定及び修辞機能と脱文脈化指数の特定

メッセージの種類を確認した後、メッセージごとに発話機能、中核要素、現象定位を認定し、その組み合わせから、メッセージごとの修辞機能と脱文脈化指数を確認する。本発表では、メッセージの分類に関わる従属節分類について述べるため、これらの認定の説明は省略する⁴。

4. 従属節アノテーション情報の利用の検討

4.1. 従属節アノテーション情報とメッセージ分類

本節では、分析データに付与されている従属節意味分類アノテーション情報を用いたメッセージの分類の検討を行う。

「鳥バンク」の体系を再整理した松本(2018)の分類体系のうち、RUA で使用するメッセージ分類に関わるのは、連用節の副詞節及び並列節と、補足節の引用節であると考えられる。表 3 に、本発表の分析対象データにおける出現頻度を示す。

図 2 に示したように、メッセージの分類のうち、従属節に関わるのは「拘束;形式的従属」と「拘束;意味的従属」である。このどちらに分類されるかによって、その節を単独で RUA の分析対象とするか、単独では分析しないかが異なる。

「拘束;形式的従属」の例である(3)では、「「物語」が多いのは予想どおりでしたが」という従属節の中で「が」に「連用節-並列節-順接的並列」が付与されており、意味的には並列の関係だが時制(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、RUA の認定の対象となる。「連用節-並列節」の下位区分は「拘束;形式的従属」に分類されることが予想される。

⁴ RUA の詳しい手順については佐野・小磯(2011)、田中(2018a,b)などを参照のこと。

表 3. 連用節副詞節及び並列節と補足節引用節の内訳

			OT01_00006	OT01_00018	OT01_00030	計
連用節	並列節	順接的並列	9	24	8	41
		逆接的並列	0	0	0	0
	副詞節	時間	2	8	3	13
		原因理由手段	9	7	13	29
		条件・譲歩	7	15	17	39
		付帯状況	3	13	7	23
		様態	0	4	1	5
		逆接	0	2	0	2
		目的	3	3	1	7
		程度・比較	0	0	5	5
補足節	引用節	直接引用	3	1	2	6
		間接引用	3	4	1	8

一方、「拘束;意味的従属」の例である(4)では「風がもうひとふきすれば」という従属節の中で「ば」に「連用節-副詞節-条件・譲歩」が付与されている。従属する「車がひいてしまう」ことが発生する条件を表しているため、単独では RUA 認定の対象とはせず、従属するメッセージの一部と考える。「連用節-副詞節」の下位区分は「拘束;意味的従属」に分類されることが予想される。

- (3)(再掲) 「物語」が多いのは予想どおりでしたが、他の種類の本を読む人が、思いの外たくさんいました。
 (4)(再掲) 風がもうひとふきすれば、車がひいてしまうわい。

4.2. 連用節-並列節

上で(3)を例に述べたように、「連用節-並列節」の下位区分は、RUA のメッセージの種類としては「拘束;形式的従属」に分類されることが予想された。分析データについてメッセージの種類認定を行った結果を表 4 に示す。「その他」は、慣用的表現の一部や連体節の一部であると考えられるため、メッセージと認定しないものなどである。

表 4. 連用節-並列節のメッセージの種類

メッセージの種類	件数
拘束;形式的従属	29
拘束;意味的従属	5
その他	7
計	41

「拘束;形式的従属」に分類された例を、(6)(7)に示す。(6)では、「クローバーが青々と広がる」と、「わた毛と黄色の花の混ざったたんぽぽが点々のもようになってきている」ことは、並列のことと考えられる。(7)では、「仲間たちに出会う」と「次から次へとおかしな事件にまきこまれる」ことは時間の経過は考えられるものの、因果関係はない。

⁵当該のタグが付与されている形態素を下線で示す。

- (6) クローバーが青々と広がる、わた毛と黄色の花の混ざったたんぽぽが、点々のもようになってきています。(OT01_00018)
 (7) 森の中でトンカチはきみょうな仲間たちに出会う、次から次へとおかしな事件にまきこまれます。(OT01_00018)

「拘束;意味的従属」に分類された例を、(8)にあげる。「言葉や表現に注意」しながら「いろいろな物語を楽しむ」と考えられるため、「言葉や表現に注意すること」だけを RUA の分析対象とはしない。

なお、(9)は、下線部分の「て」は「お母さんが虫とりあみをかまえる」と「あの子がぼうしをそうっと開ける」という 2 つの行動を並列しているが、その両方の節が「とき」という名詞にかかる副詞節の構成要素であるため、下線部分については、メッセージの認定を行わないものである。(10)でも、下線部分の「たり」は「表す」と「想像させる」が並列となっているが、「言葉」を修飾する節を構成しており、「においや色を表したり」はメッセージと認定されない。

- (8) 言葉や表現に注意して、いろいろな物語を楽しみましょう。(OT01_00006)
 (9) 「お母さんが、虫とりあみをかまえて、あの子がぼうしをそうっと開けたときー。」(OT01_00018)
 (10) 「白いぼうし」には、においや色を表したり想像させたりする言葉がたくさん出てきます。(OT01_00018)

4.3. 連用節-副詞節

先に(4)を例に述べたように、「連用節-副詞節」のタグが付与されている場合、その従属節は RUA で用いるメッセージの種類としては「拘束;意味的従属」に該当することが予想された。表 5 にメッセージの種類認定結果を示す。ほとんどが「拘束;意味的従属」に分類された。(11)は「時間」の例、(12)(13)は「原因理由手段」の例である。「拘束;形式的従属」に分類されたのは、(14)の下線部分「が」である。従属節アンテーションでは、「連用節-副詞節-逆接」が付与されている。

表 5. 連用節-副詞節のメッセージの種類

メッセージの種類	連用節-副詞節								計
	時間	原因理由手段	条件・譲歩	付帯状況	様態	逆接	目的	程度・比較	
拘束;形式的従属	0	1	0	0	0	1	0	0	2
拘束;意味的従属	12	22	29	21	5	0	6	5	101
その他	1	6	10	2	0	1	1	0	20
計	13	29	39	23	5	2	7	5	123

- (11) 項目がはっきりしている場合、結果が数字で出ている場合は、調べたことを表やグラフの形に整理すると分かりやすくなります。(OT01_00006)
- (12) 感想や意見を交流して、生活について、さらに考えを深めましょう。(OT01_00006)
- (13) 信号が赤なので、ブレーキをかけてから、運転手の松井さんは、ここにこして答えました。(OT01_00018)
- (14) 意地悪なことを言うこともありますが、本当はやさしい心のもち主です。(OT01_00018)

「その他」に分類されたのは、(15)の下線部「ば」で、「～ほど」のような慣用的表現の一部分であり、「かめば」のみで一つのメッセージとは認められない。(16)の下線部分「の」は「連用節-副詞節-逆接」が付与されているが、「だれも見たはずはないのに、本当にそうだったような気持ちになる」がその後の助詞の「の」(波線部分)にかかってこの文の主語(主部)となっている。

- (15) そうして、かめばかむほど、食べ物は口の中でだえきとまじり、だんだん小さく、やわらかくなっていきます。(OT01_00030)
- (16) だれも見たはずはないのに、本当にそうだったような気持ちになるのは、あざやかで力強い絵と、えらびぬかれた短い言葉のためでしょうか。(OT01_00018)

4.4. 補足節の引用節

補足節の引用節は、例(5)で示した、投射の判断の際に利用できる。(5)は間接引用で「って」にタグが付与され、(16)は直接引用で「と」に付与されている。一方、例(17)の下線部分「と」にも「間接引用」のタグが付与されているが、「見たいと思う」という節が「番組」を修飾している連体節のため、RUAではこの部分を分析の対象とはしない。

- (17) 「今の子は、昔ほど外遊びをしなくなった。」と、母が言っていました。(OT01_00006)
- (18) 兄弟で、見たいと思う番組がちがうことがよくあります。(OT01_00006)

5. まとめと今後の課題

本発表では、松本(2018)が検討し BCCWJ に付与を行っている従属節意味分類アノテーション情報を、RUA の分類単位であるメッセージ(おおむね節)の認定の際に利用することを検討し、報告した。

RUA ではまずメッセージへの分割と種類の認定を行い、分類対象を確定するが、従属節に関して、作業による「拘

束;形式的従属」と「拘束;意味的従属」の認定のゆれが生じることがある。本発表で検討したように、松本(2018)による従属節意味分類のアノテーションを活用することによって、ゆれの無い安定した認定が可能になることがうかがえた。

今後の課題として、メッセージ認定の後に行う「現象定位」の認定における従属節意味分類アノテーションの利用の検討を進める。「現象定位」は述部の時制や副詞などから「過去」「現在」「未来」「仮定」のいずれかに分類するものだが、「仮定」は、「もし～であれば」のような仮定の条件が従属節に現れている場合に該当する。松本(2018)の意味分類の「連用節-副詞節-条件・譲歩」のタグが付与されている従属節をさらに検討することによって、利用の可否が確認できると考えている。

これらのアノテーションを活用して、RUA を一部でも自動化していくよう、さらに検討を進めていく。

謝辞

本研究の一部は JSPS 科研費 JP15K02535 の助成を受けたものです。

参考文献

- Cloran, C. (1994) *Rhetorical Units and Decontextualisation: an Enquiry into some Relations of Context, Meaning and Grammar*. Monographs in Systemic Linguistic Linguistics, No.6. Nottingham: Department of English Studies, University of Nottingham.
- (1995) Defining and Relating Text Segments: Subject and Theme in Discourse, In Hasan R. & Fries P.H. (eds.), *On Subject and Theme: A Discourse Functional Perspective*, 361-403, Amsterdam: John Benjamins.
- Maekawa, Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den (2014) Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation* 48: 345-371
- 佐野大樹(2010a)日本語における修辭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—(修辭機能編) <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 —修辭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辭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(2018a)「日本語非母語話者向け自治会加入勧誘チラシとその作成振り返りコメントの分析—修辭機能と脱文脈化程度の観点から—」『言語情報科学』16, pp.73-88. 東京大学大学院総合文化研究科
- 田中弥生(2018b)「児童・生徒作文の日本語修辭ユニット分析と教員評価の検討」『言語資源活用ワークショップ 2018 発表論文集』, pp.91-104.
- 松本理美(2018)「日本語従属節の意味分類基準策定について: 「鳥ノク」節間意味分類体系再構築の提案」『国立国語研究所論集』15, pp.107-133.
- 丸山岳彦・柏岡秀紀・熊野正・田中英輝(2004)「日本語節境界検出プログラム CBAP の開発と評価」『自然言語処理』11(3)pp.39-68.
- 佐藤理史・丸山岳彦(2015)「節境界認定に関する諸問題」『第 8 回コーパス日本語学ワークショップ予稿集』pp.225-232.
- 佐藤理史・丸山岳彦・夏目和子(2016)「現代日本語書き言葉均衡コーパスに対する節境界付与」『言語処理学会第 22 回年次大会発表論文集』pp.409-412.