

書籍の文体情報と脱文脈度の関連 —客観度との対応のパイロットスタディ—

田中弥生(国立国語研究所) 柏野和佳子(国立国語研究所) 加藤祥(目白大学)

1. はじめに

文章の文体についての研究は、品詞の分布、語彙、語種、読点など、さまざまな観点から行われている(金2009)。柏野(2013)は文体を分類するための指標として、専門度、客観度、硬度、くだけ度、語りかけ性度の分類指標を提案し、柏野(2015)にて現代日本語書き言葉均衡コーパス(The Balanced Corpus of Contemporary Written Japanese: 以下、BCCWJ)の図書館サブコーパス(LB)のサンプルに、人手による印象評定から文体情報を付与している。テキストが専門的か子供向けか、客観的か主観的か、硬いか軟らかいか、など感じるのは、どのような言語的特徴と関わりがあるのだろうか。柏野(2015)によって付与された文体情報について、例えば浅原ほか(2014)は語彙の観点から文体との対応を確認している。また、馬場(2019)は図書館サブコーパスに含まれる語(語彙素)の文体差を数値化し、品詞別特徴、語種別特徴の全体的傾向を示すとともに、内省判断との相関を明らかにしている。文章の印象を決める要因を明らかにすることは、作文や文章表現指導などに寄与できると考えられる。

本発表は書籍に人手で付与された文体情報を修辞機能と脱文脈度の観点から検討する研究の一環である。修辞機能・脱文脈度を測るために、修辞機能分析の分類法(田中2022)を用いる。本分類法では、「修辞機能」を「話し手書き手が発信する際に、言及する対象である事態や事物や人物等を捉え表現する様態を分類し概念化したもの」と定義する。また文脈とは話者の「いま・ここ・わたし」とし、脱文脈度を「発話がコミュニケーションの場「いま・ここ・わたし」にどの程度依存しているか」の程度を表す概念とする。研究の目的は、1) 人手で付与された専門度、客観度、くだけ度などが文の要素の分類から特定される脱文脈度とどのような関連があるかを明らかにすること、及び、2) 修辞機能分析の分類法によって述部の時制と主語・主題の交差から特定される修辞機能・脱文脈度が何を測ることができるのかを明らかにすることである。これまで修辞機能分析によって、家庭での児童の会話や高齢者の談話などの話し言葉(田中・江口・小磯2023, 田中・小磯・大武2022)、児童作文、高齢者の小作文などの書き言葉の分析(田中・佐尾・宮城2023, 田中・小磯・大武2023)などを行った。BCCWJの図書館サブコーパスのサンプルでのパイロットスタディ(田中・柏野・加藤2023)では、分析した4つのサンプルには共通する修辞機能があることと、サンプルごとに特徴的な修辞機能があることが確認されている。本発表では文体情報と脱文脈度との関連を明らかにするパイロットスタディとして、客観度が異なるサンプルについて客観度の違いと脱文脈度の関連を確認する。

2. 分析

2.1. 分析対象

分析対象データは、文体情報が付与されているBCCWJの図書館サブコーパスのサンプルで、本発表では、ランダムに抽出した中から、客観度との関連を確認するパイロットスタディとして、客観度が異なり他の文体情報が一致する4サンプルを対象にした。表1に分析対象サンプルの情報を示す。

表 1. 分析対象サンプル

サンプルID	NDC ¹	書籍タイトル	専門度	客観度	硬度	くだけ度	語りかけ性度
LB17_00057	7 芸術・美術	ウディ・アレンのすべて	3 一般向き	1 とても客観的	3 どちらかといえば軟らかい	3 くだけていない	3 特に語りかけ性は低い
LBo2_00014	2 歴史	貴婦人たちの華麗なる犯罪	3 一般向き	1 とても客観的	3 どちらかといえば軟らかい	3 くだけていない	3 特に語りかけ性は低い
LBm1_00032	1 哲学	しあわせ眼鏡	3 一般向き	4 とても主観的	3 どちらかといえば軟らかい	3 くだけていない	3 特に語りかけ性は低い
LBp9_00125	9 文学	父への恋文	3 一般向き	4 とても主観的	3 どちらかといえば軟らかい	3 くだけていない	3 特に語りかけ性は低い

2.2. 分析方法

修辞機能分析は、選択体系機能言語理論(Systemic Functional Linguistics)の枠組みの英語談話分析法Rhetorical Unit Analysis(Cloran1994, 1999)を日本語に適用した修辞ユニット分析(佐野2010, 佐野・小磯

¹ 日本十進分類法

2011) に、日本語文法の枠組みで修正を加えたものである。以下に概要を示す。

2.2.1. 分析単位の分割と分析対象の特定

分析単位であるメッセージは概ね節に相当し、「定型句類」(相槌, 挨拶, 定型句, 節の形でないものなど), 「主節」(単文, 及び主節), 「並列」(従属度の低い従属節), 「従属」(従属度の高い従属節), 「引用」(“と思う”などで引用されている部分)に分類する。「主節」「並列」「引用」についてこのあとの分類を行う。

2.2.2. 発話機能・時間要素・空間要素

発話機能・時間要素・空間要素を分類し、表2に示したように、これらの組み合わせから修辞機能と脱文脈指数が特定される。【行動】[1]がもっとも文脈に依存した表現で、【一般化】[14]がもっとも脱文脈度の高い表現である²。

発話機能は「提言」か「命題」に分類する。「提言」は、品物・行為の交換に関する提供・命令で、基本的には同じ時空に存在する会話者同士の行為にかかわる「そのハサミを取って」「お醤油どうぞ」などが該当し、【行動】[1]と特定される。「命題」は、情報を交換する陳述・質問で、「私は桜が大好き」「この桜はピンクが濃いね」「桜はバラ科の植物だ」などが該当する。発話機能が「命題」のメッセージについて、このあと時間要素と空間要素を認定する。時間要素は、話者のいる時間「いま」を基準として、メッセージで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。基本的にテンスや時間を表す副詞などによって表現され、「習慣・恒久」「現在」「過去」「未来意志的」「未来非意志的」「仮定」に分類する。「いまお花見してる」は「現在」、「去年見た青森の桜はすごかった」は「過去」、「来週、お花見に行こう」は「未来意志的」、「私は桜が大好き」のような嗜好は「習慣・恒久」、「桜はバラ科の植物だ」は恒久的と判断し「習慣・恒久」に分類する。書籍サンプルでは著者が執筆した時間を「いま」と考える。例えば、「今、ここにりんごが一つある。」という文は、読者が読んでいる時間を基準にすると「過去」とも考えられるが、本研究では著者が執筆した時間を基準として「現在」に分類する。空間要素は、話者や著者のいる場所「ここ・わたし」を基準として、主語、主題、述部の主体との空間的距離から判断するもので、「参加」「状況内」「状況外」「定義」に分類する。「私」「あなた」が主語であれば「参加」、「太郎が花びらを拾っている」は太郎が話者と同じ時空にいると考えられれば「状況内」、「去年見た青森の桜はすごかった」「桜は日本中で楽しめる」の桜は話者のいる時空には存在していないと考えられるので「状況外」、「桜はバラ科の植物だ」は桜一般について述べているので「定義」に分類する。

表 2. 修辞機能と脱文脈化指数の特定

定義	高 ↑ 空間的距離のレベル ↓ 低						一般化 [14]
		報告 [9]	状況外回想 [10]	予測 [11]	推量 [12]	説明 [13]	
状況外							
状況内		実況 [2]	状況内回想 [3]	状況内予想 [5]	状況内推測 [6]	観測 [8]	
参加	行動 [1]			計画 [4]		自己記述 [7]	
空間要素	今 ここ わたし	← 時間的距離のレベル →					
時間要素		現在	過去	未来意志的	未来非意志的	仮定	習慣・恒久
発話機能		提言		命題			

時間要素は、話者のいる時間「いま」を基準として、メッセージで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。基本的にテンスや時間を表す副詞などによって表現され、「習慣・恒久」「現在」「過去」「未来意志的」「未来非意志的」「仮定」に分類する。「いまお花見してる」は「現在」、「去年見た青森の桜はすごかった」は「過去」、「来週、お花見に行こう」は「未来意志的」、「私は桜が大好き」のような嗜好は「習慣・恒久」、「桜はバラ科の植物だ」は恒久的と判断し「習慣・恒久」に分類する。書籍サンプルでは著者が執筆した時間を「いま」と考える。例えば、「今、ここにりんごが一つある。」という文は、読者が読んでいる時間を基準にすると「過去」とも考えられるが、本研究では著者が執筆した時間を基準として「現在」に分類する。空間要素は、話者や著者のいる場所「ここ・わたし」を基準として、主語、主題、述部の主体との空間的距離から判断するもので、「参加」「状況内」「状況外」「定義」に分類する。「私」「あなた」が主語であれば「参加」、「太郎が花びらを拾っている」は太郎が話者と同じ時空にいると考えられれば「状況内」、「去年見た青森の桜はすごかった」「桜は日本中で楽しめる」の桜は話者のいる時空には存在していないと考えられるので「状況外」、「桜はバラ科の植物だ」は桜一般について述べているので「定義」に分類する。

2.2.3. 修辞機能と脱文脈度の特定

表2を参照し、発話機能、時間要素、空間要素の組み合わせから、修辞機能と脱文脈度を特定する。

2.2.4. アノテーション例

表3、表4に、本発表の分析対象のサンプルの冒頭部分のアノテーション結果を示す³。本研究の分析対象データの書籍サンプルに含まれる章や節の見出しは節の形式でないことが多く、分析対象外とした。

表 3. LB17_00057 (客観度: 1 ととても客観的) 冒頭

メッセージ	発話機能	時間要素	空間要素	修辞機能[指数]
a 第五章 マンハッタンへの道	(見出しのため対象外)			
b 憧れのマンハッタン	(見出しのため対象外)			
c これまで書いてきたように、	(従属節)			
d ウディ・アレンはブロンクスで生まれ、	命題	過去	状況外	状況外回想[10]
e ブルックリンで少年時代を過ごしたわけだが、	命題	過去	状況外	状況外回想[10]
f その頃から彼には常に、きらびやかなマンハッタンが憧れの対象として	(節の一部)			
g 存在していた。	命題	過去	状況外	状況外回想[10]
h いや、これはアレンだけではなく、多くのニュー Yorker に	命題	習慣・恒久	状況外	説明[13]
i ニューヨークでは、大勢の人間がマンハッタンに憧れ、	命題	習慣・恒久	状況外	説明[13]
j ニューヨーク以外に住むアメリカ人や外国人にとっても、ニュー	命題	習慣・恒久	定義	一般化[14]
ヨークとはすなわちマンハッタンのことなのである。				

² 以下、本文中では、修辞機能を【】で、脱文脈化指数を[]で示す。

³ アノテーションは2名で行い、信頼性検討のためカッパ係数を求めたところ $k = .727$ という実質的に一致しているとみなされるカッパ係数が確認された。

表 4. LBm1_00032 (客観度：4 ととても主観的) 冒頭

メッセージ	発話機能	時間要素	空間要素	修辞機能[指数]
a 感謝の言葉	(見出しのため対象外)			
b 何か親切なことをしたときに	(従属節)			
c 「ありがとう」と言われると	(従属節)			
d 嬉しい.	命題	習慣・恒久	参加	自己記述[07]
e 礼を言ってほしくてしているわけでもないが,	命題	習慣・恒久	参加	自己記述[07]
f 相手から何の反応も返ってこないと,	(従属節)			
g 拍子抜けがしてしまう.	命題	習慣・恒久	参加	自己記述[07]
h ところで, それが肉親の間になると	(従属節)			
i どうだろう.	命題	仮定	状況外	推量[12]
j 親子の間で, どの程度感謝の言葉を言うだろう.	命題	仮定	参加	状況内推測[06]

3. 分析結果と考察

サンプルごとの修辞機能の出現頻度と割合を表5および図1に示す。4サンプルではほぼ出現しない修辞機能があること、客観的とされた2サンプルでは【状況外回想】[10]と次いで【説明】[13]が多いこと、主観的とされた2サンプルでも【状況外回想】[10]と【説明】[13]は用いられているが、客観的なサンプルより多様な修辞機能が用いられていることがわかる。

表 5. サンプルごとの修辞機能の出現頻度

	客観 / 主観	行動 [01]	実況 [02]	状況内 回想 [03]	計画 [04]	状況内 予想 [05]	状況内 推測 [06]	自己 記述 [07]	観測 [08]	報告 [09]	状況外 回想 [10]	予測 [11]	推量 [12]	説明 [13]	一般化 [14]	合計
LB17_00057	客観	0	1	12	1	0	0	5	9	1	99	0	0	44	1	173
LBo2_00014	客観	0	0	0	1	0	0	2	1	11	139	3	1	35	0	193
LBm1_00032	主観	1	1	7	0	0	2	5	2	6	16	0	1	20	0	61
LBp9_00125	主観	7	11	67	6	3	4	29	13	5	73	0	0	36	0	254

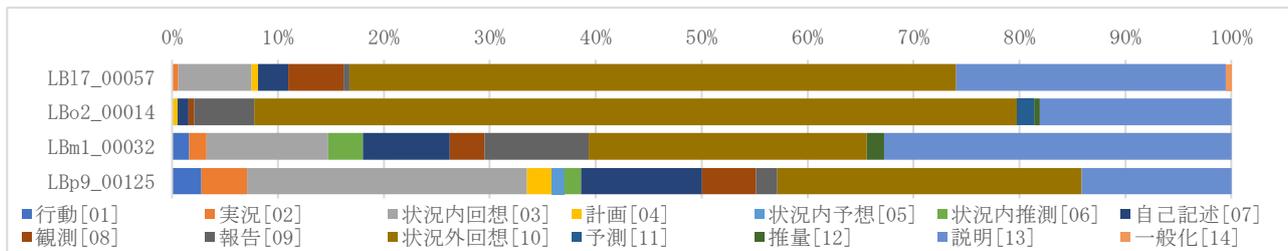


図 1. サンプルごとの修辞機能の割合

また、サンプルと修辞機能との対応関係を調べるために対応分析を行った⁴。結果を図2に示す⁵。客観的であるとされた2サンプル(客LB17, 客LBo2)は第1軸(寄与率82.3%)で右側に配置され、主観的であるとされた2サンプル(主LBm1, 主LBp9)は左側に配置された。右には脱文脈化指数の高い修辞機能、左は脱文脈化指数の低い修辞機能が配置されている。対応分析の結果から次のようなサンプルと修辞機能との関連が見られた。

- 4つのサンプルの中央に【説明】[13]が位置しており、いずれのサンプルにも共通する修辞機能と考えられる。このことは田中・柏野・加藤(2023)でも確認されている。
- 客観的なサンプルは、【状況外回想】[10]、【説明】[13]、【報告】[09]が特徴的と考えられる。
- 主観的なサンプルは、【観測】[8]、【自己記述】[07]、【状況内回想】[03]、【実況】[02]、が特徴的と考えられる。ただし客LBm1については、第1軸の中央寄りに位置し、【説明】[13]、【報告】[09]がより近い。

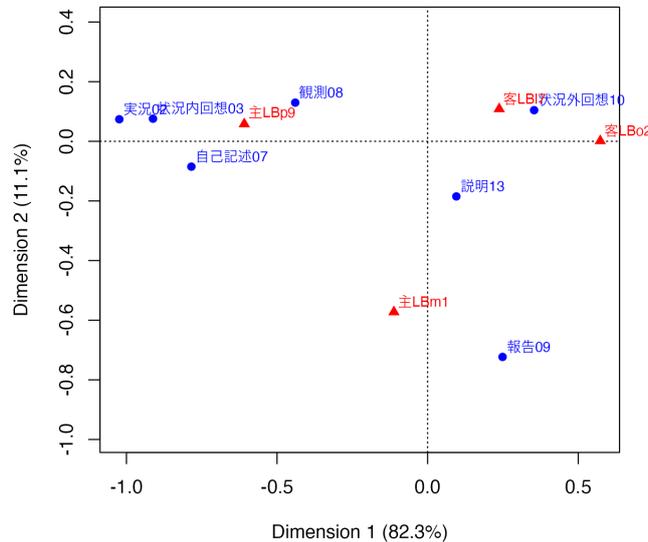


図 2. サンプルと修辞機能の対応分析結果

⁴ 分析には R の ca 関数を用いた。出現頻度が 10 件以下の修辞機能は除外した。

⁵ 図内ではサンプル ID を省略して LBxx の部分のみ表示し、サンプル ID の前に客観主観の区別(客/主)を示している。

これらのことから、書籍サンプルにおいて基本的には修辞機能及び脱文脈度と客観度には関連がある可能性があることがうかがえる。客観的なサンプルでは主観的なサンプルより脱文脈度の高い修辞機能の割合が高い。一方、主観的なサンプルでもそれら2つの修辞機能が多く用いられているが、客観的なサンプルより多様な修辞機能が用いられていることが特徴と言える。図2で中央寄りに位置した主LBm1は、図1からわかるように、同じ主観的なサンプルである主LBp9に比べ脱文脈度の高い修辞機能が多く用いられている。しかし、客観的なサンプルと比較すると脱文脈度の低い修辞機能の割合が高いことから、印象評定では主観的だと判断された可能性がうかがえる。修辞機能分析で特定される脱文脈化指数は順序尺度であり、脱文脈度は相対的なものであるが、表2に示したように、【報告】[09]より脱文脈化指数が大きいものは、空間要素(主語や主題)が「状況外」か「定義」であり、【観測】[08]より小さいものは、空間要素が「状況内」か「参加」である。主語が「私」や「私たち」であれば「参加」であるため、脱文脈度は【観測】[08]より小さくなる。人手で印象評定をアノテーションする際に、それらの言語表現が手がかりとなっていることが考えられるだろう。

本発表の分析でも田中・柏野・加藤(2023)でも【説明】[13]が書籍サンプルに共通して用いられる修辞機能であることが明らかになっている。基本的には書き手の時空に読み手が存在することのない書籍というテキストにおいては、脱文脈度の高い修辞機能が用いられるのが自然だと考えると、脱文脈度の低い修辞機能の使用割合が高い場合に、より主観的だと判断される可能性があると言えるのではないだろうか。

4. まとめ

本発表では、文体情報と脱文脈度との関連を明らかにするパイロットスタディとして、客観度が異なるサンプルについて、客観度の違いと脱文脈度の関連を確認した。その結果、客観的とされたサンプルでは、主観的とされたサンプルより脱文脈度の高い修辞機能の占める割合が高いこと、主観的なサンプルでは客観的なサンプルより多様な修辞機能が用いられていることが明らかになった。書籍サンプルという性質上、脱文脈度の高い修辞機能の使用が基本であって、書籍のテキストは客観的なものであるが、脱文脈度の低い修辞機能の使用割合が高い場合にはより主観的だと判断される可能性がうかがえた。

本発表は、5つの文体情報のうち4つが同じで客観性のみ異なる、「とても客観的」とされたサンプル2件と「とても主観的」とされたサンプル2件、計4件のみでの分析であった。今後該当するサンプルすべてにアノテーションを行い、このパイロットスタディの結果を確認する予定である。また、客観性以外の文体情報についても脱文脈度との関連を確認していく。なお、本発表では、あらかじめランダムに抽出された中から条件にあう4サンプルを分析対象としたため、NDCの統制は取れておらず、今後の課題としたい。

謝辞 本研究は JSPS 科研費 JP19K00588, JP23K00569 によるものです。

参考文献

- 浅原正幸・加藤祥・立花幸子・柏野和佳子(2014). 文体指標と語彙の対応分析 第6回コーパス日本語学ワークショップ予稿集, 11-20.
- 馬場俊臣(2019). BCCWJ 文体情報の各文体指標の特徴語: 『BCCWJ 図書館サブコーパスの文体情報』を用いて 北海道教育大学紀要. 人文科学・社会科学編 69(2), 1-14.
- C. Cloran(1994). *Rhetorical units and decontextualisation: An enquiry into some relations of context, meaning and grammar*. Nottingham: University of Nottingham.
- C. Cloran(1999). Contexts for learning. Frances C (Ed.), *Pedagogy and the Shaping of Consciousness: Linguistic and Social Processes*. London: Continuum International Publishing. 31-65.
- 柏野和佳子(2013). 「〈共同研究プロジェクト紹介〉萌芽・発掘型: テキストの多様性を捉える 分類指標の策定 書籍サンプルの文体を分類する 国語研プロジェクトレビュー4:1, 43-53.
- 柏野和佳子(2015). BCCWJ 図書館サブコーパスの文体情報(2015年公開第1版). <https://doi.org/10.15084/00003109>
- 金明哲(2009). 計量文献学 計量国語学会(編)計量国語学事典 朝倉書店 pp. 238-252.
- 佐野大樹(2010). 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), (RUAの方法と手順 ver. 0. 1. 1). https://researchmap.jp/multidatabases/multidatabase_contents/detail/228720/34ec2728989b398c8dd30e659251320e?frame_id=486573
- 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-「書き言葉らしき話し言葉らしさ」と脱文脈化言語・文脈化言語の関係- 機能言語学研究, 6, 59-81.
- 田中弥生(2022). 修辞機能と脱文脈化の観点からの日本語談話分析 博士論文, 東京大学大学院総合文化研究科. (未公開)
- 田中弥生・江口典子・小磯花絵(2023). 家庭での食事場面における親子会話の脱文脈度の観点からの分析 言語資源ワークショップ発表論文集, 1, 329-339.
- 田中弥生・柏野和佳子・加藤祥(2023). 書籍の文体と修辞機能の分析のパイロットスタディ 言語資源ワークショップ発表論文集, 1, 142-150.
- 田中弥生・小磯花絵・大武美保子(2022). 共想法談話のテーマと修辞機能の関連についての分析 言語処理学会第28回年次大会発表論文集, 1439-1443.
- 田中弥生・小磯花絵・大武美保子(2023). 共想法による話し言葉・書き言葉における修辞機能の特徴-テーマとの関係に着目して- 言語処理学会第29回年次大会発表論文集, 1356-1360.
- 田中弥生・佐尾ちとせ・宮城信(2023). 児童の作文における表現の脱文脈化観点による可視化 日本語文法学会第24回大会発表予稿集, 197-202.