

< 全文 > 言語資源ワークショップ2022発表論文集

著者	国立国語研究所言語資源開発センター
雑誌名	言語資源ワークショップ発表論文集
巻	1
ページ	1-416
発行年	2023
URL	http://doi.org/10.15084/00003716

Language Resources Workshop

言語資源ワークショップ2022

発表論文集

2022年8月30日(火)~31日(水)

大学共同利用機関法人 人間文化研究機構
国立国語研究所 言語資源開発センター

▼目次

o1-2s	「日本語日常会話コーパス」に見る ABAB 型基本オノマトベの音韻パターン： 日本語教育の視点から 3 廉 沢奇 (神戸大学国際文化学研究科)
	Phonological patterns of ABAB-Type onomatopoeia in CEJC corpus: from the viewpoint of Japanese Language Education LIAN Zeqi (Graduate School of Intercultural Studies, Kobe University)
o1-3	ユーザの特性情報付きチャットボットとの雑談対話コーパスの概要 17 伊藤 紀子 (同志社大学文化情報学部), 岩下 志乃 (東京工科大学コンピュータサイエンス学部), 杉本 徹 (芝浦工業大学工学部), 林 篤司 (東京工科大学コンピュータサイエンス学部非常勤講師), ト 秋予 (同志社大学大学院文化情報学研究科博士前期課程)
	Overview of Human-Chatbot Dialogue Corpus with Personality Information Noriko Ito (Doshisha University), Shino Iwashita (Tokyo University of Technology), Toru Sugimoto (Shibaura Institute of Technology), Atsushi Hayashi (Tokyo University of Technology), Bu Qiuyu (Doshisha University)
o1-4	疑問・非疑問発話の韻律典型性を捉える試み — 日本語母語話者と学習者の比較 — 29 波多野 博顕 (筑波大学), 王 可心, 陳 凱僑, 林 良子 (神戸大学)
	An Attempt to Understand Prosodic Typicality of Interrogative and Non-Interrogative Speech: Comparison of Native Speakers and Learners of Japanese Hiroaki Hatano (University of Tsukuba), Kexin Wang, Kaiqiao Chen, Ryoko Hayashi (Kobe University)
p1-1s	日本語学習者による用例を通時的観点で分析可能とするコーパス開発の意義とその可能性 39 上出 大河 (國學院大學 大学院 文学研究科)
	The Significance and Potential of Developing a Corpus to enable Diachronic Analysis of Usage by Learners of Japanese Taiga Kamide (Graduate School of Literature, Kokugakuin University)
p1-2	「急性」を含む病名の語構成 43 相良 かおる (西南女学院大学), 高崎 智子 (西南女学院大学/医師), 東条 佳奈 (大阪大学), 西嶋 佑太郎 (医師), 山崎 誠 (国立国語研究所)
	Word structure of disease names containing "acute" Kaoru Sagara (Seinan Jo Gakuin University), Satoko Takasaki (Seinan Jo Gakuin University, Physician), Kana Tojo (Osaka University), Yutaro Nishijima (Physician), Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

- p1-4** 『日本語話題別会話コーパス：J TOCC 語彙表』の公開と日本語教育むけ情報サイトにむけた
指標の検討 53
中俣 尚己 (大阪大学国際教育交流センター), 麻 子軒 (関西大学国際教育センター)
- ‘Vocabulary Table of Japanese Topic Oriented Conversation Corpus’ and a Consideration of Indices for Japanese Language Teaching**
NAKAMATA Naoki (Osaka University), MA Tzuhsuan (Kansai University)
- p1-6** 科学技術論文における「問題」の周辺文からの問題内容の抽出 59
平林 照雄 (東京農工大学 生物システム応用科学府), 古宮 嘉那子 (東京農工大学),
浅原 正幸 (国立国語研究所)
- Extracting Problem Contents from Surrounding Texts of a Word “Problem” in Science and Technical Papers**
Teruo Hirabayashi (Tokyo University of Agriculture and Technology),
Komiya Kanako (Tokyo University of Agriculture and Technology),
Masayuki Asahara (National Institute for Japanese Language and Linguistics)
- p1-7** 対のある自他動詞と文法を組み合わせた項目を用いたオンライン学習の解答傾向
— 5週間のオンライン学習で用いた項目の分析を中心に — 71
沖本 与子 (東京外国語大学)
- Student Tendencies of Five Week Online Study Materials for Paired Transitive/Intransitive Verb-Grammar Items**
Tomoko Okimoto (Tokyo University of Foreign Studies)
- p1-8s** NINJAL データベースを活用した言語研究の実施について 79
鈴木 成典 (国際基督教大学大学院), 五十嵐 陽介 (国立国語研究所), 李 勝勳 (国際基督教大学)
- Linguistic Research Using NINJAL Database**
Michinori Suzuki (Graduate School of International Christian University),
Yosuke Igarashi (National Institute for Japanese Language and Linguistics),
Seunghun J. Lee (International Christian University)
- p2-1** 文化的・言語的相違による「ハミング・鼻歌」の初発語彙使用
— 日・中・英・米語のコーパスによる分析 — 83
浅野 恵子 (順天堂大学医学部一般教育外国語研究室), 陳 森 (東京福祉大学 社会福祉学部)
- First lexical usage of the words "hum" across languages and cultures
- Analysis of a Japanese, Chinese, English, and American multilingual corpus -**
Keiko Asano (Juntendo University, Faculty of Medicine, Liberal Arts division),
Sen Chen (Tokyo University of Social Welfare, School of Social Welfare)
- p2-3** 作文評価における日本語教師の評価ポリシーの多様性 — ホリスティック評価の観点から — 91
安 芝恩
Diversity of Japanese Language Teachers' Assessment Policies in Writing Assessment: From the Viewpoint of Holistic Scoring
Jieun Ahn

p2-4 『子ども版日本語日常会話コーパス』の構築 103

小磯 花絵, 天谷 晴香, 居關 友里子, 白田 泰如, 柏野 和佳子, 川端 良子, 田中 弥生 (国立国語研究所), 藤 越 (東京大学大学院総合文化研究科/国立国語研究所), 西川 賢哉 (国立国語研究所)

Construction of CEJC-Child

Hanae Koiso, Haruka Amatani, Yuriko Iseki, Yasuyuki Usuda, Wakako Kashino, Yoshiko Kawabata, Yayoi Tanaka (NINJAL), Yue Teng (The University of Tokyo / NINJAL), Ken'ya Nishikawa (NINJAL)

p2-6 実践医療用語 語構成要素語彙試案表 Ver.2.0 の構築 109

東条 佳奈 (大阪大学), 黒田 航 (杏林大学), 相良 かおる (西南女学院大学), 高崎 智子 (西南女学院大学), 西嶋 佑太郎 (医師), 麻 子軒 (関西大学), 山崎 誠 (国立国語研究所)

Development of a Word Component Database for "Hands-On Medical Terms" (Version 2.0)

Kana Tojo (Osaka University), Kou Kuroda (Kyorin University), Kaoru Sagara (Seinan Jo Gakuin University), Satoko Takasaki (Seinan Jo Gakuin University, Physician), Yutaro Nishijima (Physician), Tzu-Hsuan Ma (Kansai University), Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

p2-9 テレビゲームコーパスの構築とその利活用 117

麻 子軒 (関西大学国際教育センター)

Construction and Utilization of Video Game Corpus

Ma Tzuhsuan (Kansai University)

o2-1 日本語の動詞とその結合価 127

青山 文啓 (桜美林大学大学院)

Valency and the Japanese Verbal

Fumihiro Aoyama (Graduate Division, J. F. Oberlin University)

o2-2s 英語学習用活用語彙リストの提案 133

— CEFR-J Wordlist のコロケーション・データセットの試み — 133
福田 航平 (東京外国語大学), 投野 由紀夫 (東京外国語大学)

Constructing a Collocation Database for the CEFR-J Wordlist

Kohei Fukuda (Tokyo University of Foreign Studies), Yukio Tono (Tokyo University of Foreign Studies)

o2-3s スロベニア人初級・中級・上級日本語学習者コーパスの構築と縦断的文法誤用分析 147

パウロヴィチ・ミハ (リュブリャナ大学文学部アジア研究学科)

The Construction and longitudinal analysis of a learner corpus of Japanese on the elementary, intermediate and advanced level

Miha Pavlovič (University of Ljubljana, Department of Asian Studies)

o3-1 現代日本語の名づけにおける「文の包摂」 155

泉 大輔 (東京外国語大学)

An Analysis of Sentential Compound in Japanese Naming

Daisuke Izumi (Tokyo University of Foreign Studies)

- o3-2 「のだ」の〈言い換え〉用法に接続詞が前置する条件の一考察 — 新書テキストを素材として — 187
石原 佳弥子

Examining the Conditions under which the Subjunctive Is Employed in the Paraphrase Usage of Noda: Using New Text Material

Kayako Ishihara

- o3-3s Twitter にみられる特徴的な慣用表現「名前をつけたい」に関する考察 195
星野 靖子 (放送大学文化科学研究科)

A Study on a Twitter-specific Idiomatic Expression ‘Need a Name for’

Yasuko Hoshino (School of Graduate Studies, the Open University of Japan)

- o4-1s 書き言葉・話し言葉コーパスデータに基づく高頻度漢語動名詞の品詞性の再考：
日本語教育の視点から 215
陳 迪 (神戸大学大学院国際文化科学研究科)

Reconsideration of Part-of-Speech of Sino-Japanese Verbal Nouns Based on Written and Spoken Corpus Data : A Perspective on Japanese Language Education

Di CHEN (Graduate School of Intercultural Studies, Kobe University)

- o4-2 上級日本語学習者の文体把握に関するケーススタディ
— 「BCCWJ 図書館サブコーパス文体情報」を用いた読解調査 — 233
小西 円 (東京学芸大学)

A Case Study of Advanced Japanese Language Learners' Understanding of Styles:

Reading Comprehension Survey Using the BCCWJ Library Subcorpus Stylistic Information

Madoka Konishi (Tokyo Gakugei University)

- p3-1 エンタメ小説における会話文の発話意図分析 — 発話文表現文型辞書の改良に向けて — 243
夏目 和子 (名古屋大学), 佐藤 理史 (名古屋大学)

Analyzing Utterance Intentions in Entertainment Novels

Kazuko Natsume (Nagoya University), Satoshi Sato (Nagoya University)

- p3-4 幼児と保護者によるごっこ遊びの相互行為 — 日常場面に関する知識の利用に着目して — 253
居關 友里子 (国立国語研究所), 小磯 花絵 (国立国語研究所)

Pretend Play Interactions Between Caregiver and Child

Yuriko Iseki (National Institute for Japanese Language and Linguistics),

Hanae Koiso (National Institute for Japanese Language and Linguistics)

- p3-6 「一周回って」の意味・用法をめぐって：Twitter の投稿データを言語資源として 263
朴 秀娟 (神戸大学), 于一楽 (滋賀大学)

On the Meaning and Usage of issyuu mawatte: using Twitter as a Language Resource

Sooyun Park (Kobe University), Yile Yu (Shiga University)

p3-9	近現代語における「もちろん」の用法 275 東泉 裕子 (東洋大学), 高橋 圭子 (東洋大学)	275
	Usage of <i>mochiron</i> in modern and contemporary Japanese Yuko Higashiizumi (Toyo University), Keiko Takahashi (Toyo University)	
p3-10	医学書テキストのたとえる表現 (2) — 接尾辞「状」の特徴 283 本多 由美子 (一橋大学・国立国語研究所), 三枝 令子 (元一橋大学)	283
	Illustrating Expressions in Medical Texts (2): Characteristics of the suffix “jou” Yumiko Honda (Hitotsubashi University / NINJAL), Reiko Saegusa (Hitotsubashi University)	
p4-1	打ち合わせにおける談話構造の修辞機能からの分析 297 田中 弥生 (国立国語研究所研究系)	297
	An Analysis of Discourse Structure of a Meeting from a Rhetorical Function Point of View. Yayoi TANAKA (National Institute for Japanese Language and Linguistics)	
p4-2	日常会話における「状況づけられた語り」 309 白田 泰如 (国立国語研究所研究系)	309
	“Situated Tellings” in Everyday Conversation Yasuyuki Usuda (National Institute for Japanese Language and Linguistics)	
p4-3	コーパスからの複合動詞の自動抽出の試み — 近現代作家の文学作品からの用例抽出を例に — 319 チャクマクピルギル・ニハル (アンカラ大学/麗澤大学), 千葉 庄寿 (麗澤大学)	319
	An Extraction of Japanese Compound Verbs from Corpus : An Example of Modern Literary Works Nihal Cakmak Bilgili (Ankara University/Reitaku University), Shoji Chiba (Reitaku University)	
p4-4	日本語学習者のコーパスツールの使用実態 — 作文での産出に着目して — 339 寺嶋 弘道 (立命館アジア太平洋大学), 板井 芳江 (立命館アジア太平洋大学)	339
	Use of Corpus Tool in Writing by Japanese Language Learners Hiromichi Terajima (Ritsumeikan Asia Pacific University), Yoshie Itai (Ritsumeikan Asia Pacific University)	
p4-6	児童作文における書き出しと結びの分析 347 加藤 恵梨 (愛知教育大学教育学部)	347
	A Study of the Opening and Closing Lines of Elementary Students' Compositions Eri Kato (Aichi University of Education)	
p4-7	「つまり」による換言が促す理解の範囲について 359 櫻井 芽衣子 (日本工業大学)	359
	The Scope of Influence that Promotes Understanding by Paraphrasing Tsumari Meiko Sakurai (Nippon Institute of Technology)	

p4-8s	自発対話音声に対する叫び声アノテーション	365
	白鳥 恵大, 大久保 港, 松田 匠翔, 有本 泰子 (千葉工業大学情報科学部)	

Scream and shout annotation for spontaneous dialog speech

Keita Shiratori, Minato Okubo, Takuto Matsuda, Yoshiko Arimoto
(Faculty of Information and Computer Science, Chiba Institute of Technology)

p4-9	学術論文と論説文における「だから」の使用の比較	375
	向坂 卓也 (中国・外交学院)	

Comparison of the Usage of "Dakara" in Academic Papers and Editorial Texts

Takuya Mukozaka (China Foreign Affairs University)

o5-1	少数言語のデジタルアーカイブ: PhoPhoNO と BantuDArc	387
	李 勝勳 (国際基督教大学), 倉部 慶太 (AA 研), 品川 大輔 (AA 研)	

Digital archives of understudied languages: PhoPhoNO and BantuDArc

Seunghun J. Lee (International Christian University), Keita Kurabe (ILCAA), Daisuke Shinagawa (ILCAA)

o5-2	「小中高大生による日本語絵描写ストーリーライティングコーパス」(JASWRIC) の構築: L1/L2 日本語研究の新しい資料として	393
	石川 慎一郎 (神戸大学), 友永 達也 (神戸大学附属小学校), 大西 遼平, 岡本 利昭, 勝部 尚樹, 川嶋 久予, 岸本達也, 村中 礼子 (神戸大学附属中等教育学校)	

Japanese Students' L1 Story Writing Corpus (JASWRIC): A New Dataset for Analysis of L1/L2 Japanese

Shin'ichiro ISHIKAWA (Kobe University), Tatsuya TOMONAGA (Kobe University, Elementary School),
Ryohei ONISHI, Toshiaki OKAMOTO, Naoki KATSUBE, Hisayo KAWASHIMA, Tatsuya KISHIMOTO,
Reiko MURANAKA (Kobe University, Secondary School)

▼プログラム

1日目：8月30日（火）9:30～17:20

9:30～9:50 オープニング

10:00～12:00 口頭発表セッション1

o1-1s 源氏物語における評価の形容詞

小原 みと希（中央大学文学研究科国文学専攻）

o1-2s 「日本語日常会話コーパス」に見る ABAB 型基本オノマトベの音韻パターン：日本語教育の視点から

廉沢 奇（神戸大学国際文化研究科）

o1-3 ユーザの特性情報付きチャットボットとの雑談対話コーパスの概要

伊藤 紀子（同志社大学），岩下 志乃（東京工科大学），杉本 徹（芝浦工業大学），
林 篤司（東京工科大学），ト 秋予（同志社大学大学院文化情報学研究科）

o1-4 疑問・非疑問発話の韻律典型性を捉える試み — 日本語母語話者と学習者の比較 —

波多野 博顕（筑波大学），王 可心（神戸大院），陳 凱僑（神戸大院），林 良子（神戸大）

13:00～14:00 招待講演1

▶開かれた音声情報処理のためのコーパス 高道 慎之介（東京大学）

14:10～15:40 ポスターセッション1（p1 コアタイム：14:10～14:55、p2 コアタイム：14:55～15:40）

p1-1s 日本語学習者による用例を通時的観点で分析可能とするコーパス開発の意義とその可能性

上出 大河（國學院大學大学院文学研究科）

p1-2 「急性」を含む病名の語構成

相良 かおる（西南女学院大学），西嶋 佑太郎，東条 佳奈，高崎 智子，山崎 誠

p1-3 史的文字連携システム API の利用：東京大学史料編纂所が公開する仮名漢字字形を検索するツールの開発

劉 冠偉（東京大学史料編纂所），中村 覚（東京大学史料編纂所），山田 太造（東京大学史料編纂所）

p1-4 『日本語話題別会話コーパス：J-TOCC 語彙表』の公開と日本語教育むけ情報サイトにむけた指標の検討

中俣 尚己（大阪大学），麻子 軒（関西大学）

p1-5 『現代日本語書き言葉均衡コーパス』書籍サブコーパスの小説サンプルに対するジャンル情報付与

加藤 祥（目白大学），浅原 正幸（国語研）

p1-6 科学技術論文における「問題」の周辺文の問題内容の抽出

平林 照雄（東京農工大学 生物システム応用科学府），古宮 嘉那子（東京農工大学），
浅原正幸（国立国語研究所）

p1-7 対のある自他動詞と文法を組み合わせた項目を用いたオンライン学習の解答傾向

— 5週間のオンライン学習で用いた項目の分析を中心に —
沖本 与子（東京外国語大学）

p1-8s NINJAL データベースを活用した言語研究の実施について

鈴木 成典 (国際基督教大学大学院アーツ・サイエンス研究科), 五十嵐 陽介 (国立国語研究所),
李 勝勳 (国際基督教大学, ヴェンダ大学)

p1-9 『名大会話コーパス』 中納言版・ひまわり版公開データの形態論情報の修正

柏野 和佳子 (国立国語研究所), 西川 賢哉 (国立国語研究所), 渡邊 友香, 小磯 花絵 (国立国語研究所)

p2-1 文化的・言語的相違による「ハミング・鼻歌」の初発語彙使用 一日・中・英・米語のコーパスによる分析一

浅野 恵子 (順天堂大学医学部), 陳 森 (東京福祉大学 社会福祉学部)

p2-2s 「女 (ひと)」のような二重表記が検索できる歌詞コーパスの設計と構築

胡 佳芮 (一橋大学大学院言語社会研究科)

p2-3 作文評価における日本語教師の評価ポリシーの多様性 — ホリスティック評価の観点から —

安 芝恩 (なし)

p2-4 『子ども版日本語日常会話コーパス』の構築

小磯 花絵 (国立国語研究所), 天谷 晴香 (国語研究所), 居關 友里子 (国語研究所),
白田 泰如 (国語研究所), 柏野 和佳子 (国語研究所), 川端 良子 (国語研究所), 田中 弥生 (国語研究所),
藤 越 (国語研究所), 西川 賢哉 (国語研究所)

p2-5 日本語教師養成のための音読観察実習における多段階の振り返りを考慮したビデオアノテーション共有手法

山口 昌也 (国立国語研究所), 森 篤嗣 (京都外国語大学外国語学部)

p2-6 実践医療用語_語構成要素語彙試案表 Ver.2.0の構築

東条 佳奈 (大阪大学), 黒田 航 (杏林大学), 相良 かおる (西南女学院大学), 高崎 智子 (西南女学院大学),
西嶋 佑太郎 (医師), 麻 子軒 (関西大学), 山崎 誠 (国立国語研究所)

p2-7s テ形従属節の用法分類に向けたアノテーションガイドラインの構築

野口 咲帆 (お茶の水女子大学人間文化創成科学研究科), 田中 リベカ (お茶大), 戸次 大介 (お茶大)

p2-8 TextGrid 不要の Praat アノテーション管理

西川 賢哉 (国立国語研究所)

p2-9 テレビゲームコーパスの構築とその利活用

麻 子軒 (関西大学)

15:50 ~ 17:20 □頭発表セッション 2**o2-1** 日本語の動詞とその結合価

青山 文啓 (桜美林大学大学院)

o2-2s 英語学習用活用語彙リストの提案 — CEFR-J Wordlist のコロケーション・データセットの試み —

福田 航平 (東京外国語大学 大学院総合国際学研究所 世界言語社会専攻), 投野 由紀夫 (東京外国語大学)

o2-3s スロベニア人初級・中級・上級日本語学習者コーパスの構築と文法的誤用分析

パウロ ヴイチ・ミハ (リュブリャナ大学文学部アジア研究学科)

2日目：8月31日(水) 9:20～17:30

9:20～10:50 □頭発表セッション3

o3-1 現代日本語の名づけにおける「文の包摂」

泉 大輔 (東京外国語大学)

o3-2 「のだ」の〈言い換え〉用法に接続詞が前置する条件の一考察 — 新書テキストを素材として —

石原 佳弥子 (なし)

o3-3s Twitter にみられる特徴的な慣用表現「名前をつけたい」に関する考察

星野 靖子 (放送大学文化科学研究科)

11:00～12:00 □頭発表セッション4

o4-1s 書き言葉・話し言葉コーパスデータに基づく高頻度漢語動名詞の品詞性の再考：日本語教育の視点から

陳 迪 (神戸大学)

o4-2 上級日本語学習者の文体把握に関するケーススタディ

— 「BCCWJ 図書館サブコーパス文体情報」を用いた読解調査 —

小西 円 (東京学芸大学)

13:00～14:00 招待講演2

▶ 芥川賞作品コーパスの構築のために 菅野 倫匡 (筑波大学)

14:10～15:40 ポスターセッション2 (p3 コアタイム：14:10～14:55、p4 コアタイム：14:55～15:40)

p3-1 エンタメ小説における会話文の発話意図分析

夏目 和子 (名古屋大学), 佐藤 理史 (名古屋大学)

p3-2 E テレの児童向け教育番組における単語出現頻度

北村 達也 (甲南大学知能情報学部), 川村 よし子 (東京国際大学)

p3-3s wav2vec モデルによる方言音声資料のテキスト化

峯尾 海成 (静岡理科大学大学院), LI XIAORAN (静岡理科大学大学院 理工学研究科),
谷口 ジョイ (静岡理科大学), 高野 敏明 (静岡理科大学)

p3-4 幼児と保護者によるごっこ遊びの相互行為：日常場面に関する知識の利用に着目して

居關 友里子 (国立国語研究所), 小磯 花絵 (国立国語研究所)

p3-5 唐話資料「三字話」系列の継承と展開

岩本 真理 (大阪公立大学非常勤講師 大阪市立大学名誉教授)

p3-6 「一周回って」の意味・用法をめぐって — Twitter の投稿データを言語資源として —

朴 秀娟 (神戸大学), 于一楽 (滋賀大学)

- p3-7** 医療記録における縮約表現の量的構造 — 医療用語との比較 —
山崎 誠 (国立国語研究所), 黒田 航 (杏林大学), 東条 佳奈 (大阪大学), 西嶋 佑太郎 (医師), 麻 子軒 (関西大学), 相良 かおる (西南女学院大学)
- p3-8** 様々な対話場面における speech-laugh の発生タイミングの分析
有本 泰子 (千葉工業大学), 真弓 花 (千葉工業大学)
- p3-9** 近現代語における「もちろん」の用法
東泉 裕子 (東洋大学), 高橋 圭子 (東洋大学)
- p3-10** 医学書テキストのたとえる表現 (2) — 接尾辞「状」の特徴
本多 由美子 (一橋大学・国立国語研究所), 三枝 令子 (元一橋大学)
- p3-11** 日本語日常会話における他称表現の使用傾向について
川端 良子 (国立国語研究所)
- p4-1** 打ち合わせにおける談話構造の修辞機能からの分析
田中 弥生 (国立国語研究所)
- p4-2** 日常会話における「状況づけられた語り」
白田 泰如 (国立国語研究所)
- p4-3** コーパスからの複合動詞の自動抽出の試み—近現代作家の文学作品からの用例抽出を例に—
チャクマクビルギル・ニハル (アンカラ大学 / 麗澤大学), 千葉 庄寿 (麗澤大学)
- p4-4** 日本語学習者のコーパスツールの使用実態 — 作文での産出に着目して—
寺嶋 弘道 (立命館アジア太平洋大学), 板井 芳江 (立命館アジア太平洋大学)
- p4-5** 医療記録における縮約表現の分析
東条 佳奈 (大阪大学), 黒田 航 (杏林大学), 相良 かおる (西南女学院大学), 西嶋 佑太郎 (医師), 麻 子軒 (関西大学), 山崎 誠 (国立国語研究所)
- p4-6** 児童作文における書き出しと結びの分析
加藤 恵梨 (愛知教育大学)
- p4-7** 「つまり」による換言が促す理解の範囲について
櫻井 芽衣子 (日本工業大学)
- p4-8s** 自発対話音声に対する叫び声アノテーション
白鳥 恵大 (千葉工業大学情報工学部), 大久保 港 (千葉工業大学), 松田 匠翔 (千葉工業大学), 有本 泰子 (千葉工業大学)
- p4-9** 学術論文と論説文における「だから」の使用の比較
向坂 卓也 (外交学院 (中国))
- p4-10** 関東・東北方言における動詞ラ行音節の撥音化と促音化 — COJADS データより —
佐藤 久美子 (国立国語研究所)

p4-11 リサーチデザインにおける言語資源の役割 — QA サイトコーパス（知恵袋データ）の場合 —
中渡瀬 秀一（国立情報学研究所）

15:50 ~ 16:50 □頭発表セッション 5

o5-1 少数言語のデジタルアーカイブ：PhoPhoNO と BantuDArc
李 勝勲（国際基督教大学），倉部 慶太（AA-研），品川 大輔（AA-研）

o5-2 「小中高大生による日本語絵描写ストーリーライティングコーパス」(JASWRIC) の構築：
L1/L2 日本語研究の新しい資料として
石川 慎一郎（神戸大学），友永 達也（神戸大学附属小学校），大西 遼平，岡本 利昭，勝部尚樹，川嶋 久予，
岸本 達也，村中 礼子（神戸大学附属中等教育学校）

16:50 ~ 17:30 クロージング



8月30日(水)

口頭発表セッション1, 2, ポスターセッション1

「日本語日常会話コーパス」に見る ABAB 型 基本オノマトペの音韻パターン：日本語教育の視点から

廉 沢奇 (神戸大学国際文化学研究所)

Phonological patterns of ABAB-Type onomatopoeia in CEJC corpus: from the viewpoint of Japanese Language Education

LIAN Zeqi (Graduate School of Intercultural Studies, Kobe University)

要旨

本研究は、日本語会話において頻出する基本的な ABAB 型オノマトペ (例：どんだん、そろそろ) の音象徴の解明を目指したものである。研究は母音と子音、そして併せた五十音、最後は清濁の差をめぐって検討する。手法については、まずは A・B 位置の音の要素を行・段で分解し、全体での行・段の傾向性を明らかにした。そして対応分析を用い、代表的な音象徴を持つ ABAB 型オノマトペを選定し、その意味傾向を探究するために 3:3 共起語を調べた。最後は清音語と濁音語の共起語の差を分析した。これより、ABAB 型オノマトペはその音で「動作系」と「変化系」に 2 つ分けられる。これ結果は、先行研究で示された印象に基づくイメージとは異なり、オノマトペの意味と音のつながりの新たな視角となっている。

1. はじめに

日本語の語彙体系において、オノマトペはきわめて重要な位置を占めており、小野 (2007) によると、その数は 4000 を超えるという。こうしたオノマトペの興味深い特性の 1 つは、特有の音韻傾向を有することである。オノマトペの多くは、聞いただけでオノマトペとわかる。

一方、日本語教育において、オノマトペは、語彙指導の直接の対象になりにくく、教科書の単元内でオノマトペが出現した場合も、語のおおよその意味が示されるだけで、その音韻特性について、詳しい説明がなされることは滅多にない。

しかし、オノマトペは、元来、人の感覚に訴えかける語であり、音が語の意味に強く影響している。このため、構成音素が少し変わるだけで全体としてはまったく異なる印象が生じる。下記の例を見てみよう。これらの用例は、日本語母語話者の自然発話を収集した『日本語日常会話コーパス』(CEJC) (詳細は以下に示す) に含まれる用例である。

- (1) 迫力がすごくて、ちょっとはらはら、どきどきもあってね。(K004_013)
- (2) 本とかもばらばらになってさ。(T016_002)
- (3) 雨が大して降ってないな、ばらばらだな。(T007_015)

(1)~(3)に含まれるのはいずれもほぼ同一のオノマトペであるが、(1)の「は」は清音で、

危惧を感じる心理的な状態を意味する。(2)の「ば」は濁音で、激しく散らかっている様子が浮かぶ。また、(3)の「ば」は半濁音で、量の少なさや、個々の雨粒が独立している様子が含意される。これらの例は、オノマトペの語構成における音韻要素の重要性を如実に示すものと言えよう。

以上の点をふまえると、日本語学習者に対する指導の現場においては、単に主要なオノマトペを抽出して提示するだけでなく、その音韻的傾向性と、音韻パタンごとの意味特性についても適切に指導することが重要となろう。筆者は、廉 (2022) において、オノマトペの典型的形態である ABAB 型に絞り、CEJC のジャンルごとに頻度調査を行うことで、高頻度かつ汎用的に用いられる「基本口語 ABAB 型オノマトペ」227 語を抽出した。本研究は、同リストに含まれる語を対象に、その音韻的傾向と、意味との関係を明らかにすることを目指す。

2. 先行研究

以下では、(1)オノマトペの全体的な音韻傾向に関する研究と、(2)オノマトペの持つ音韻パタンの意味特性に関する研究に分けて概観を行う。

まず、(1)については、各種の計量研究がなされている。たとえば、城岡 (1998) は、辞書の収録語を数えあげることで、一般語とオノマトペの語頭子音を比較した。この調査によれば、一般語彙の語頭音は、カ行 (19.1%) > サ行 (18.2%) > ハ行 (15.2%) > タ行 (14.0%) の順で多かったのに対し、オノマトペの語頭音は、ハ行 (33.9%) > カ行 (26.0%) > サ行 (14.6%) > タ行 (12.3%) の順となった。この研究は、オノマトペの語頭においてハ行音が多い可能性を示唆する。なお、この研究では、濁音・半濁音も含めた調査となっているため、より正確には、語頭でハ・バ・パ行音が多いということになる。一方、丹野 (2005) は、幼児 (2 歳 6 ヶ月～6 歳 6 ヶ月) 300 名が 1 週間に使用したすべてのオノマトペを保護者に記録させるという調査を行った。この調査により、各年齢段階で同様の音韻を有するオノマトペが使用されること、幼児は濁音を多用するほか、母音については、「い」が多く、「え」が少ないことなどが確認された。

次に、(2)について、生越 (2019) は、辞書に収録されたハ・バ・パ行で始まるオノマトペを対象に、辞書の解説文を比較することで、オノマトペの音による意味差を調査した。その結果、子音について、/h/音は擬音語では「空気の動き」(ひそひそ)、擬態語では「弱々しさ・柔らかさ」(ふわふわ)、/p/、/b/音は「表面への影響」「面の変化」(ぴちゃぴちゃ・ぼたぼた・ばらばら) というイメージがあると判断された。一方、母音について、/a/は「広がり」(ぱっ)、/i/は「直線」(ピン・ピタリ)、/u/は「つき出す」(ぶかりぶかり)、/e/は「不快」(べたっ)、/o/は「目立たなさ」(ぼっ) と「粒状性」(ぼちゃり) を表すと判断された。これにより、音のイメージが簡潔にまとめられたが、実際にはこの関係性に当てはまらない語も多く存在する。

このほか、実験調査に基づく研究もなされている。清音と濁音の印象の差について、丹野 (2006) は清音と濁音からなる 17 対の ABAB 型オノマトペ (はらはら／ばらばら、など) を 84 名の女子大学生に聞かせ、思い浮かぶ連想語を記録した。その結果、清音は軽快・小さい・愛らしいというイメージで、濁音は鈍重・大きい・荒々しいイメージであるという一般的な法則に合致するのは、17 対中 6 対のみであることが示された。篠原 (2010) は、実在語の意味が判断結果に影響しないよう、無意味の ABAB オノマトペ 40 種 (有声阻害音/b, d, g, z/, 無声阻害音/p, t, k, s/, 母音/a, i, u, e, o/) をローマ字で表記したものを 103 名の協力者 (日本語話者 42 名、英語話者 22 名、中国語 (北京語) 話者 20 名、韓国語話者 19 名) に聞か

せ、サイズについての印象を4段階で答えさせた。回答結果を分析することで、大きさを感じるのは、母音の開口度では低母音[a]>中母音[e,o]>高母音[i,u]の順、舌位置では後母音[u,o]>前母音[i,e]、子音タイプでは有声音>無声音であることが示された。被験者の国別で要因の効果を見ると、日本語母語話者は3つすべての要因が影響しているのに対し、英語・中国語・韓国語話者は母音の開口度に影響されず、韓国語話者は子音有声音にも影響されないことが分かった。この結果は、オノマトペの音表象に関して、母語話者が有している音韻感覚を学習者が共有していないことを示唆する。

上記で見たように、オノマトペの音に関する先行研究は、使用するデータの選定、試験の手法、最終的に得られた結論においてそれぞれ価値ある知見を示している。しかしながら、全体として小規模の研究が多く、系統的に母音・子音・清濁の意味差に関する調査はなされていない。

そこで本研究は、会話コーパスを調査資料とし、行段・清濁別に調査を行うことで、新しい観点から、重要な ABAB 型口語オノマトペの音象徴の解明を目指す。

3. リサーチデザインと手法

3.1 研究目的と RQ

すでに述べたように、本論文の目的は、4モーラの ABAB 型オノマトペ（「どんどん」「きよろきよろ」など）のうち、話し言葉における高頻度かつ汎用的な 227 語（廉、2022）をサンプルとして、A・B 要素別に音韻傾向を解明し、重要オノマトペの取りやすい典型的な音韻パターンを解明することである。この点に関しては、先行研究より、(1)語頭にはハ行音（濁音・半濁音含む）が多く、幼児のオノマトペ使用には濁音・「い」が多い、(2)ハ・バ・パ行に関して母音と子音の違いが意味の違いに関係する可能性がある、(3)清音＝小、濁音＝大というパターンにあてはまるものとあてはまらないものがあること、などが知られているが、A・B 要素を区別したコーパス調査は少ない。そこで、これらの点に関して新しい知見を得ることを目的として、以下の3つの研究設問を設定した。

RQ1 A・B 要素において典型的な音韻パターン（行・段・組み合わせ）はなにか？それらはどう整理できるか？（典型的な音韻パターン）

RQ2 音韻パターン別に意味の違いはあるか？（音韻パターン別意味特性）

RQ3 清音系と濁音系でサイズイメージの違いは確認できるか？（清濁別の意味特性）

3.2 調査対象語

本研究で調査対象にするのは、オノマトペの典型的形態である ABAB 型の中で、廉 (2022) で抽出された日本語重要会話オノマトペ 227 語である。これらは、『日本語日常会話コーパス』(Corpus of Everyday Japanese Conversation, 以下 CEJC) の 10 変種（雑談、用談・相談、会議・会合、授業・レッスンの 4 ジャンルに話者の年齢・性別・話者間関係の変数を組み込み、全体を 10 種に区分した）における平均頻度と、1 回以上出現している変種数（レンジ）の 2 つのデータを根拠として選定された。下記はその一部である。

表1 日本語重要会話オノマトペリスト(一部)

順位	オノマトペ	平均頻度	レンジ	重要度
1	どんどん	215.9	10	215.9
2	ぐるぐる	103.8	8	83.1
3	だんだん	80.4	9	72.4
4	そろそろ	49.5	9	44.6
5	くるくる	52.8	7	36.9
6	ふわふわ	42.0	8	33.6
7	ぎりぎり	39.3	8	31.5
8	しゅわしゅわ	43.9	7	30.7
9	めちやめちや	37.5	8	30.0
10	しゅっしゅっ	34.6	7	24.2

たとえば、「だんだん」の例で言うと、10変種での平均頻度（100万語あたりの調整頻度）は215.9である。また、10変種すべてにおいて「だんだん」は1回以上出現しており、レンジは10/10となる。このとき、当該語の重要度は $215.9 \times 10/10 = 215.9$ と計算される。また、「くるくる」の例で言うと、10変種での平均頻度（100万語あたりの調整頻度）は52.8である。また、10変種の中、7変種において「だんだん」は1回以上出現しており、レンジは7/10となる。このとき、当該語の重要度は $52.8 \times 7/10 = 36.9$ と計算される。

3.3 データ

本研究で利用したコーパスは前述のCEJCである。CEJCはさまざまな場面における自然な会話を収集することを目的としたコーパスである。性別・年齢などのバランスを考慮して抽出された40名の調査協力者を集め、機材機器を約3か月間貸与し、協力者の日常生活で自然に生じる会話を記録させた。発話者数は全体で約250名となる。今回利用したデータは「中納言」版の200時間分の会話である。

3.4 手法

RQ1では、はじめに、A要素とB要素を区別してそれぞれの特性を確認し、その後、統計手法を援用し、両者を組み合わせて特性を概観する。前者については、まず、A要素・B要素に出現する文字を、行（アイウエオの5行）と列（アカサタナハマヤラワの10段）からなる50音図上に記録していく。たとえば、「くるくる」を例にすると、A要素については「カ行・ウ段」の音で、B要素については「ラ行・ウ段」の音とみなす。227語のすべてについて同様の調査を行うことで、A要素に出やすい音韻、B要素に出やすい音韻を特定する。なお、B要素には50音図に含まれない促音「っ」と撥音「ん」が出現するところがあるが、これらについても、同時に記録する。続いて、後者については、97種の音（A要素で54種、B要素で43種）をケース、CEJCの4ジャンル（雑談、用談・相談、会議・会合、授業・レッスン）を変数とする頻度表に対して対応分析を実施する。対応分析とは、「データ表の行や列に含まれる情報を少数の成分に圧縮し、それらの関係を散布図に配置することで、視覚的なデータの俯瞰を可能にする」手法である（石川他、2010、p.245）。散布図上では、2つの次元（縦横軸）によって、データが4つの象限に分割される。

RQ2 では、RQ1 で得られた各象限に属するオノマトペの中から、パターンに完全に合致し、重要度の高い典型語を抽出し、それらの意味の違いを比較する。意味調査は、辞書調査・共起語調査の2段階で行う。まず、辞書調査については、上位5語の各々について、『日本国語大辞典 第二版』における副詞用法の語釈部分（用例や解説は含まない）を抽出して1つのテキストにまとめ、「Web 茶まめ」で形態素解析を行い、高頻度定義語を特定する。なお、『日本国語大辞典 第二版』は、総項目数50万、用例数100万を収録したもので、日本語辞書の最高権威とされるものである。次に、共起語調査として、同じく上位5語の各々についてCEJCで中心語の左右3語範囲に出現する共起語の頻度調査を行い、高頻度共起語を特定する。共起語分析には、コーパスコンコーダンサ AntConc (Version 4.0.8) の Collocate 機能を使用する。その後、これらの2つの調査結果を組み合わせることで、各象限に区分されたオノマトペの意味傾向の違いを探る。

最後に、RQ3 では、2段階で調査を行う。前述のように、篠原 (2010) の非実在語イメージ調査では、開口度では低母音[a] > 中母音[e,o] > 高母音[i,u]の順で、舌位置では後母音[u,o] > 前母音[i,e]の順で、子音タイプでは濁音・有声音 (/b, d, g, z/) > 清音・無声音 (/p, t, k, s/)の順で、イメージが大→小に転じていくことが報告されている。開口度・舌位置の両方の基準で、サイズイメージが大きくなるのは「オ」であり、小さくなるのは「イ」であることから、これらの中に最も明瞭なサイズイメージの差があるのではないかと推定される。また、有声音 (/b, d, g, z/) > 無声音 (/p, t, k, s/) の順位性を加味すると、有声音、つまりバ・ダ・ガ・ザ行の「オ」を含む語が最も大きなサイズイメージを持ち、無声音、つまりパ・タ・カ・サ行の「イ」を含む語が最も小さなサイズイメージを持つはずである。この点を検証するために、全227種の中から、A要素またはB要素においてバ・ダ・ガ・ザ行の母音「オ」を含む全9語（例：ごろごろ・おどおど）と、パ・タ・カ・サ行の母音「イ」を含む全19語（例：きらきら・ちくちく）を特定し、前述の辞書調査・共起語調査の手続きを経て、これら2種の間でのサイズイメージの違いの有無を検証する。続いて、分析対象語の範囲を広げ、全227語を純清音語（110語）と、そのほかの非清音系（117語）に二分し、辞書・共起語調査を行う。調査手法はRQ2に準じる。

4. 結果と考察

4.1 RQ1 典型的な音韻パターン

4.1.1 A要素・B要素の音韻傾向

A・B要素における主要な行・段・組み合わせ要素を調査したところ、以下の結果を得た。

表2 位置別の音要素

	A要素						B要素					
	ア段	イ段	ウ段	エ段	オ段	総計	ア段	イ段	ウ段	エ段	オ段	総計
ア行	1	1	4	0	1	7	3	11	4	0	0	18
カ行	9	2	6	3	5	25	7	8	11	0	5	31
ガ行	11	4	13	1	5	34	0	0	0	1	0	1
サ行	2	7	4	0	3	16	6	7	0	0	1	14
ザ行	2	7	3	1	0	13	1	5	2	0	0	8
タ行	1	15	4	1	6	27	10	21	3	1	1	36

ダ行	3	0	0	0	5	8	1	0	0	0	1	2
ナ行	0	2	1	1	0	4	0	3	0	1	0	4
ハ行	2	4	5	1	1	13	2	0	1	0	1	4
バ行	7	2	5	3	6	23	3	1	0	0	1	5
パ行	9	10	5	4	4	32	0	1	1	0	0	2
マ行	1	0	2	2	9	14	1	0	1	0	0	2
ヤ行	0	0	0	0	3	3	3	0	0	0	1	4
ラ行	2	0	0	0	0	2	21	16	7	2	18	64
ワ行	6	0	0	0	0	6	6	0	0	0	0	6
促音	/											19
撥音												7
総計	56	54	52	17	48	227	64	73	30	5	29	227

まず、A 要素について、行側はガ (34) > パ (32) > タ (27) > カ (25) > バ (23)・・・、段側はア (56) > イ (54) > ウ (52) > オ (48) > >エ (17)・・・、両者の組み合わせではチ (15) > グ (13) > ガ (11) > ピ (10)・・・の順になることが分かった。次に、B 要素について、行側はラ (64) > >タ (36) > カ (31) > >撥音 (19) > ア (18)・・・、段側はイ (73) > ア (64) > >ウ (30) > オ (29) > >エ (5)、組み合わせではチ (21) = ラ (21) > ロ (18) > リ (16)・・・の順になることが分かった。

つまり、ABAB 型オノマトペの初頭位置となる A 要素には音韻的に際立つ音(破裂音)が出現し、B 要素には、流音の「ラ」行音など、音韻的に目立たない音が出現するのが、現代日本語におけるオノマトペの基本的な音韻構成であるということになる。A 要素の破裂音は、初頭位置であることと相まって、オノマトペと一般語彙を音韻的に区別させる効果があると思われる。

ここで、今回の結果を先行研究における報告内容と比較してみよう。まず、城岡 (1998) は、辞書収録オノマトペの語頭位置、つまりは A 要素について、清音・濁音・半濁音を区別しない数え方で、ハ (33.9%) > カ (26.0%) > サ (14.6%) > タ (12.3%) …であると述べている (4 種の累計頻度は 86.8%)。今回の結果を城岡の基準に合わせると、A 要素については、ハ (32+23+13=68 : 30.0%) > カ (34+25=59 : 26.0%) > タ (27+8=35 : 15.4%) > サ (16+13=29 : 12.8%) (4 種の累計頻度は 84.2%) …となる。これより、辞書オノマトペと話し言葉オノマトペの音韻パタンの比率のずれは数パーセント以内で、はほとんど一致していることが確認された。また、城岡が調査対象としていない B 要素についても、同様に城岡の基準に合わせると、ラ (64 : 28.2%) > タ (36+2=38 : 16.7%) > カ (31+1=32 : 14.0%) > サ (14+8=22 : 9.7%) > 撥音 (19 : 8.4%) > ア (18 : 7.9%)・・・の順となった。B 要素は A 要素に比べて、ラが極端に多く (+27.3%)、カが少ない (-12%) ことがわかる。加えて、これらの比較により、オノマトペの音韻パターンを議論する際に、清音・濁音・半濁音を区別せずに計量することには問題がありうることも示唆された。たとえば、A 要素のハ系について言えば、清音のハの頻度は 23%、濁音のバの頻度は 13%、半濁音のパの頻度は 32% となり、必ずしも清音の頻度が多いわけではなく、3 種の間頻度の差も大きい。

次に、丹野 (2005) は、幼児の使用するオノマトペでは(1)濁音多用、(2)イ多用・エ少用という傾向が見られるとしている。幼児による無意識の選択は、オノマトペの音韻的典型性を反映している可能性がある。この点を検証すべく、今回の結果を整理すると、(1)については、A 要素では、濁音 (117 : 51.5%) と清音 (110 : 48.5%) の差は小さく、濁音多用ははっきりとは確認されなかった。また、B 要素では、濁音 (18 : 7.9%) は清音 (183 : 80.6%) より少なく、やはり濁音多用は確認されなかった。(2)については、A 要素では、ア (56 : 24.7%) > イ (54 : 23.8%) > ウ (52 : 23%) > オ (48 : 21.1%) > エ (17 : 7.5%)、B 要素では、イ (73 : 36.3%) > ア (64 : 32.0%) > ウ (30 : 15.0%) > オ (29 : 14.4%) > エ (5 : 2.5%) の順となる。イの多用は A 要素では確認できず、B 要素ではある程度示唆された。一方、エの少用は A、B 要素ともにはっきり確認された。つまり、幼児のオノマトペ使用の音韻特性のうち、エの少用のみが現代日本語のオノマトペ特性に合致していたことになる。

4.1.2 A 要素と B 要素の対応関係

以上で、A 要素、B 要素ごとに、典型的な音韻傾向を確認した。しかし、実際のオノマトペでは、両者が1つにまとまって語を形成している。そこで、コーパスジャンルを第1アイテム、A、Bそれぞれの音韻要素を第2アイテムとする頻度表に対して対応分析を実施したところ、以下の結果を得た。なお、図の視認性を高めるため、同一地点に重なっている語については実線の四角で、また、原点から遠くに布置されたデータについては破線の四角で、それぞれ別記した。

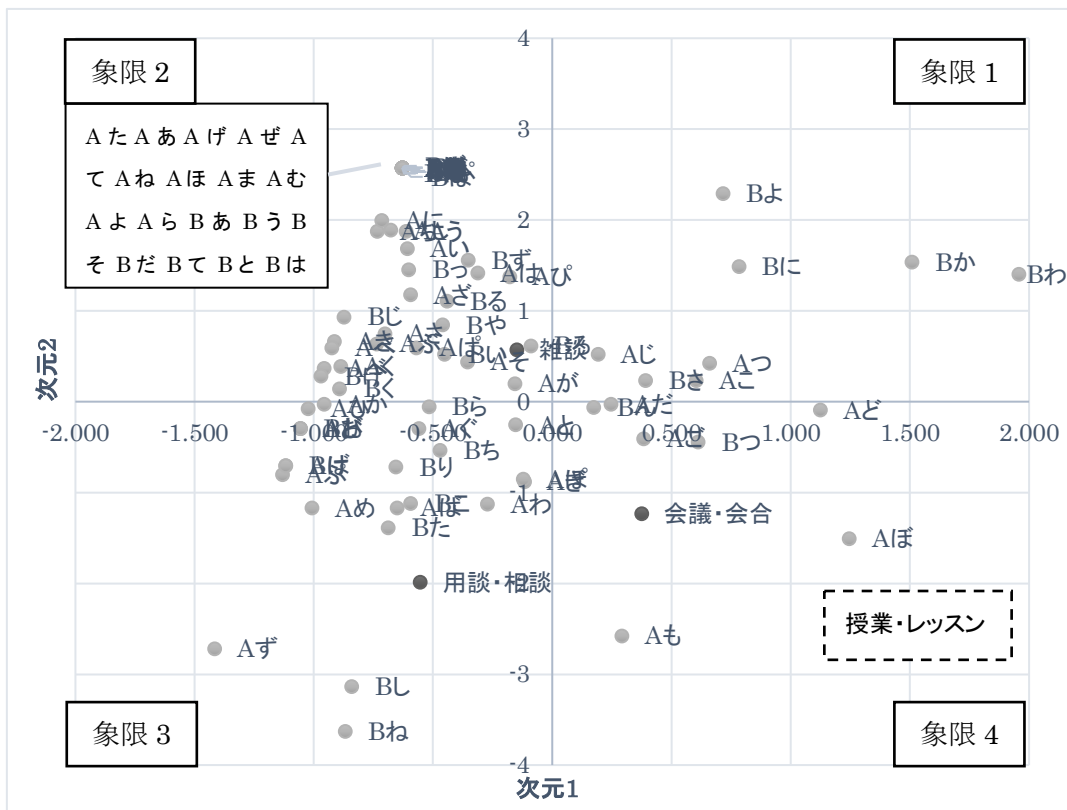


図1 対応分析による散布図

次元1（第1軸）の寄与率は44.2%であり、次元2（第2軸）の寄与率は39.5%である。2つの次元でデータの元の分散の約83.7%が説明されていることになる。

ここで、CEJCのジャンルを手掛かりとして、軸と象限の性質を考えてみたい。まず、第1軸の左側には「用談・相談」、「雑談」が、右側には「会議・会合」、「授業・レッスン」が含まれており、第1軸は私的（左側）と公的（右側）を区分する軸になっていると解釈できる。次に、第2軸の下部には「授業・レッスン」、「会議・会合」、「用談・相談」が、上部には「雑談」が含まれており、第2軸は、目的あり会話（下部）と目的なし会話（上部）を区分する軸であると解釈できる。これらの2つの軸により、データの全体は異なる性質を持つ4象限に区分されることになる。

表3 対応分析分類整理表

象限	A位置の音	B位置の音	音素数
1（公的・目的なし）	コ、ジ、ツ、フ	カ、サ、ニ、ヨ、ワ	9
2（私的・目的なし）	ア、イ、ウ、【ガ】、キ、ク、 ゲ、サ、ザ、シ、ゼ、ソ、 タ、【チ】、テ、ニ、ネ、ハ、 パ、【ピ】、ブ、ベ、ペ、ホ、 マ、ム、ヨ、ラ	ア、イ、ウ、ク、ゲ、ジ、 ズ、ソ、ダ、テ、ト、ハ、 ビ、フ、プ、ホ、ボ、マ、 ム、ヤ、ル、レ、【ロ】、 ツ	52
3（私的・目的あり）	オ、カ、ギ、【グ】、ケ、ズ、 ト、ヌ、バ、ヒ、ビ、プ、 ヘ、ポ、メ、ワ	コ、シ、タ、【チ】、ド、 ネ、バ、ピ、【ラ】、【リ】	26
4（公的・目的あり）	ゴ、ス、ダ、ド、ボ、モ	キ、ザ、ツ、ン	10

注：出現頻度において上位4位以内に入る音素は【】で示す。

ここからわかることを3点にまとめたい。まず、1点目は、A位置とB位置を合わせた延べの音素総数の70%が第2象限と第3象限に含まれることである。これらは、オノマトペが私的な言語環境で生じやすいことを示す。2点目は、同じ私的環境であっても、A位置の上位要素は第2象限に、B位置の上位要素は第3象限に含まれることである。高頻度音素に限って言えば、A要素は目的なし環境と、B要素は目的あり環境と親和性が高い。3点目は、各音素を清音・濁音にわけて濁音率を計算すると、第4象限（50%）>第3象限（38%）>第2象限（28%）>第1象限（11%）の順になるということである。一般の予想とは異なり、濁音は、公的・目的あり環境で相対的に出現しやすいと言える。

以上、RQ1に関しては、A位置、B位置に出現しやすい音素を特定するとともに、両者の組み合わせパターンを4分割した場合、公的・私的、目的あり・なしという2つの軸で区分されること、これらのタイプと特定の音素の間に結びつきが存在することが確認された。

4.2 RQ2 音韻パターン別意味特性

表4に示した各象限の典型的な音パターンに完全に合致する語を特定し、辞書におけるこれらの語の定義文中での高頻度出現語（上位3語）、また、コーパスにおける高頻度共起語

(上位 5 語) を調査したところ、以下の結果が得られた。

表 4 各象限の音パタンの合致語と辞書定義語とコーパス共起語

象限	合致語数と合致語例	辞書定義語	コーパス共起語	意味傾向
1	5 : ふわふわ、ふにゃふにゃ、つかつか、ふかふか、じわじわ	1. する (3) 2. 柔らかい (3) 3. 膨らむ (2)	1. 来る (12) 2. 香り (6) 3. じわじわ (6) 4. つかつか (6) 5. 溜まる (1)	状態および緩慢な変化
2	43 : そろそろ、くるくる、しゅっしゅっ、ちょろちょろ、にやにや…	1. する (5) 2. 行う (3) 3. 進行 (2)	1. する (110) 2. ている (63) 3. くるくる (52) 4. 回る (22) 5. 来る (10)	個別の変化・動作
3	36 : ぎりぎり、めちゃめちゃ、ばりばり、びしょびしょ、ばらばら…	1. する (8) 2. 勢い (4) 3. 降る (3)	1. 食べる (22) 2. まで (21) 3. 仕事 (11) 4. 働く (5) 5. 後ろ (4)	激的な変化・動作
4	5 : どンドン、だんだん、ぼんぼん、どきどき、ぼきぼき	1. する (3) 2. 次々 (2) 3. 続ける (2)	1. なる (84) 2. 変わる (34) 3. 上がる (20) 4. 進む (16) 5. 出す (15)	変化の安定的継続

注：合致オノマトペは重要度順に記載している。また、辞書定義語・コーパス共起語に添えた数値は頻度を示す。

以下、各象限に含まれるオノマトペの意味的性質について、オノマトペと関連語（辞書定義語とコーパス共起語を総称する）中の動詞タイプの組み合わせから解釈を試みる。まず、第1象限では、「(ふわふわ) している」や「(じわじわ) 来る・膨らむ」など、物事の固定的な状態や、段階的で緩慢な変化を含意する動詞が多い。用例を見ておこう。なお、用例中、[] で示したものは筆者が元文脈を確認して補った語である（以下同）。

(1) オムライスが空気を含み、柔らかくふわふわしている。(W005_001)

(2) [友人とパン作りについて] 発酵するのは 30 度とかじゃないですか、あの感じでじわじわと [膨らむ] ね。(K006_020)

(1) は、オムライスの柔らかい状態を「ふわふわ」というオノマトペで表現している。ここでは、「ふわふわ」でオムライスの固定的なありようを示している。(2) は生地がゆっくりとわずかずつ膨らむ状態を「じわじわ」というオノマトペで表現している。ここで「じわじわ」で段階的で緩慢なサイズ変化を示していることが分かる。

第2象限では、「(そろそろ) 行う」や「(くるくる) 回る」など、慎重な意図をもってなされる行為や、動きのはっきりした個別の動作を示す動詞が多い。

- (3) もうすぐ日が暮れちゃうから、そろそろ [家へ帰るために] 行こうかな (K006_013)
 (4) [カメラのレンズが] 回っている、あ、くるくる回っている。(K008_001a)

(3) は、「帰宅するためにその場に立ち去る」という個別の動作を慎重に時間をかけて行おうとしていることを含意する。また、(4) はカメラのレンズの回転がお刻みに連続的に行われていることを含意する。この2つの例は同じく「個別の動作」を強調していることが分かる。

第3象限では、「(めっちゃめっちゃ) 食べる・働く」や、「(ぎりぎりまで) やる・働く」など、極限まで進む勢いを持つ激烈な動作を含意する動詞が多い。

- (5) [小さいインコを見ながら] ちびちゃんがめっちゃめっちゃ食っていたような気がする。(T011_015)
 (6) [仕事後の飲み会で] 大泉さんはぎりぎりまで働いていたよね。(T011_017)

(5) は、小さいインコの食べ方は勢いがある、且つ普通食べると想定される量を超えて、その限界値まで食べていることを含意する。(6) は働き方が極限に至るまで猛烈であることを含意する。この2つの例は同じく「極限まで進む勢いを持つ激烈な動作」を含意していることが分かる。

第4象限では、「(どんどん) 続ける・変わる・上がる・進む」や「(だんだん) ~になる」、また、「(次々と) 変わる」のように、変化が連続的・継続的に起こることを含意する動詞が多い。

- (7) [老人が世の変化に感嘆して] やっぱり時代がどんどん変わっている。(K003_014)
 (8) [友達に料理を教えて] 出汁がだんだん上手になってきた。(T021_005)

(7) は、時代の変化が連続的に繰り返して起こっていることを含意する。(8) は出汁の取り方の上達の程度が段階的に継続的に上がっていることを含意する。この2つの例は同じく強烈で目立つ変化ではなく、時間が経つとともに起こる連続的变化を含意していることが分かる。

以上をふまえると、第1象限は「状態および緩慢な変化」型オノマトペ、第2象限は「個別の変化・動作」型オノマトペ、第3象限は「激烈な変化・動作」型オノマトペ、第4象限は「変化の安定的継続」型オノマトペと解釈できるだろう。これらは、変化の大きさ・顕著さの点で質的な差異を有する。以下は各象限の関係性をモデル化したものである。

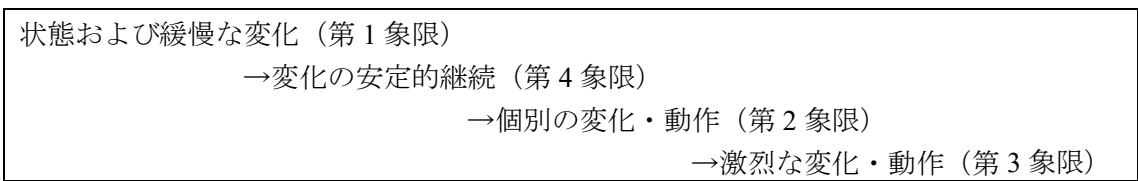


図2 ABAB 会話オノマトペの音パタンの連続モデル

第1象限と第4象限の順位性は、決定しがたいが、ここでは、そもそも変化しない状態性を包含する第1象限を第1層とし、ついで、変化が恒常的・安定的に起こることで結果的に変化が意識されにくくなる第4象限を第2層とした。これに対し、第2象限では小さいが具体的で個別的な変化性が出て来る。そして、第3象限はそうした動きの勢いが増し、極致までの激烈な動きが中心となる。

以上より、RQ2に関しては、4種のおノマトペの音韻パターンごとに異なる意味特性があること、また、それらの中に質的な段階性が存在することが確認された。そこで、本研究はABAB会話おノマトペの音パタンの連続モデルを提案した。

4.3 RQ3 清濁別の意味特性

まず、「清音・母音イ＝小」、「濁音・母音オ＝大」という先行研究の知見を検証するため、それらのパターンに合致する語を選び、RQ2と同様、辞書におけるそれらの語の定義文中での高頻度出現語（上位5語）とコーパスにおける高頻度共起語（上位5語）を調査したところ、以下の結果が得られた。

表5 清音+イ・濁音+オの合致語と辞書定義語とコーパス共起語

	合致語数と語例	辞書定義語	コーパス共起語	意味傾向
清音+イ	19： ちよきちよき、しくしく、かちかち、ちらちら、きらきら…	1. 音 (7) 2. する (4) 3. ない (3) 4. 刺激 (2) 5. 感じ (2)	1. する (67) 2. 座る (19) 3. 着る (16) 4. 見る (13) 5. 行く (6)	軽微で具体的な個人動作
濁音+オ	9： どんどん、どぼどぼ、ぼんぼん、ぼろぼろ、ごろごろ…	1. 音 (13) 2. する (5) 3. いる (4) 4. 激しい (4) 5. 物 (4)	1. なる (56) 2. 変わる (31) 3. 次 (27) 4. 入れる (22) 5. 進む (16)	抽象的な変化

上表に明らかなように、定義語・共起語からは、「清音・母音イ＝小」、「濁音・母音オ＝大」を支持するような明確な証拠は得られなかった。

もっとも、2つのグループ間では、何らかの意味の違いも示唆される。たとえば、清音イでは「する」が多く、「座る」「着る」「見る」など、軽微で具体的な個人的動作に関わる動詞が多い。一方、濁音オには「なる」が一番多く、「変わる」「入れる」「進む」など、抽象的な変化に関わる動詞が多い。ただ、これらは、母音イ、母音オに結び付く特性というよりは、清音・濁音の全体的特性ではないかと考えられる。

そこで、調査範囲を広げ、「基本口語 ABAB 型おノマトペ」227語を純清音語とそれ以外の濁音語に二分し、清音全体・濁音全体に対して同様の調査を行ったところ、以下の結果を得た。

表6 純清音語・濁音語とそれぞれの辞書定義語とコーパス共起語

	合致語数と語例	辞書定義語	コーパス共起語	意味傾向
清音	110： そろそろ、くるくる、 ふわふわ、ちょろち よる、にやにや…	1. する (5) 2. ある (3) 3. 行う (2) 4. 進行 (2) 5. 動作 (2)	1. する (368) 2. ている (175) 3. 回す (28) 4. 見える (27) 5. なる (24)	軽微で具体的個人動作の進行
濁音	117： どんどん、ぐるぐる、 だんだん、ぼろぼろ、 びしょびしょ…	1. する (4) 2. 物 (3) 3. 多い (3) 4. 強い (3) 5. 勢い (2)	1. なる (224) 2. 行く (144) 3. 見る (136) 4. 変わる (42) 5. 次 (34)	勢いのある抽象的变化

母音要件を外し、清濁系の語彙全体を調査した場合でも、前述のようなサイズの差ははっきりした形では確認できなかった。一方、「軽微で具体的な個人動作」と「抽象的变化」という意味の違いは、清濁系全体で見てもおよそ当てはまるように思われる。以下、それぞれの特徴を概観する。

まず、清音系の定義語・共起語には、「行う」「進行」「動作」や「する」「ている」などがあり、具体的な個人動作の中でも、とくに、進行中の動作を描写する傾向性が示唆される。この点に関して、以下のような用例が見られる。

(9) [綿あめづくりを体験している母親が子供に向かって] 一二三で回して一、さ、くるくるくるくるくるって回して (T003_012)

(10) [母親が知人に自分の子供を紹介して] 二歳なっているかな、もう歩ける子だけど、まだちょろちょろしている。(C001_001)

(11) [友人と就活について言って] もし将来自分が面接官やる側だと、なんか余裕でにやにやしながら、なんかやっちゃいそう。(T006_002)

(9) は、綿あめを作るために棒を機械に差し込んで、回転させるという動作が進行中であることを示している。(10) はうまく歩けない子供の動きという行為が今まさに進行中であることを示している。(11) は「笑う」という行為が面接の時間において進行していることを示している。この3つの例は、いずれも大きな作業負担を伴わない軽微な個人的動作の進行を描写していることが分かる。一方、濁音系の定義語・共起語には、「多い」「強い」「勢い」や、「なる」「変わる」「行く」などがあり、抽象的变化の中でも、とくに、勢いをもって変化していく様子を描写する傾向性が示唆される。この点に関して、以下のような用例が見られる。

(12) 肥料も効いているのかなと思って拝見して、すごくどんどん大きくなるのです。(T023_004)

(13) [母が雨の日に出かける子供に] 水溜まりに入らないようにね、びしょびしょになるから。(T003_012)

(14) [大事にしている本を見ながら] カバーもなんか、ちょっとぼろぼろになってきた。

(T006_005)

(12) は、肥料の効果で植物が勢いよく成長していくという変化を示している。(13) は水溜まりを踏むと水が一瞬にズボンを濡らすという強烈な変化が起こりうることを示している。(14) は本のカバーがもともと新しい状態から、今の破れている様子に至るまでの抽象的变化を示している。この 3 つの例は同じく元の状態から変化していくことを描写していることが分かる。

以上より、RQ3 の清濁音のイメージとしては、従来言われていたサイズ差は必ずしも明瞭ではなく、むしろ、動作と変化という意味の方向性の違いが存在する可能性が示唆された。

5. おわりに

本研究は日常会話で用いられた主要な ABAB 型オノマトペの音韻パターンを実証的に検討した。その結果、RQ1 (典型的な音韻パターン) については、A 要素の行側はガ・パ・タ、段側はア・イ・ウ、両者の組み合わせではチ・グ・ガなどが、B 要素の行側はラ・タ・カ、段側はイ・ア・ウ、組み合わせではチ・ラ・ロなどが多いことがわかった。また、対応分析により、個々の音素の持つ意味傾向が、公的・私的、目的あり・なしの 4 区分に分類できる可能性が示唆された。

RQ2 (音韻パターン別意味特性) では、辞書定義語とコーパス共起語に着目することで、前述の 4 区分が、状態および緩慢な変化 (第 1 象限) → 変化の安定的継続 (第 4 象限) → 個別の変化・動作 (第 2 象限) → 激的な変化・動作 (第 3 象限) という意味変化と合致している可能性が示された。

RQ3 (清濁別の意味特性) では、先行研究で言及された、「清音・母音イ＝小」、「濁音・母音オ＝大」という傾向性は確認されなかった。むしろ、清音・濁音は、動作の軽微性、抽象的な変化、といった意味と関連している可能性が示された。

以上で見たように、本研究は一定の成果を挙げたが、もちろん、課題も多い。オノマトペの音韻は多面的な特質を持ち、表面的な計量だけでその全貌をとらえることは困難である。この点をさらに深く議論するには、今後、コーパスに基づく計量調査と、心理学的な反応実験を組み合わせた分析アプローチが必要になってくるだろう。今後の研究において、これらの可能性を探っていきたいと考える。

引用文献

- 石川慎一郎・前田忠彦・山崎誠(2010)『言語研究のための統計入門』くろしお出版。
 生越真理(2019)「オノマトペ研究～ハ行の音象徴～」大阪教育大学国語教育専攻卒業論文。
 北原保雄ほか(2003)『日本国語大辞典 第 2 版』小学館。
 篠原和子(2010)「音象徴の言語間比較:大ききのイメージについて」『日本認知科学会第 27 回大会論文集』24-25。
 城岡啓二(1998)「日本語の擬音語・擬態語の語頭音の分布について」静岡大学人文学部論文集 48,169-195。
 丹野眞智俊(2005)「幼児の使用する日本語オノマトペの音韻分析」『児童教育学研究』24,1-6。
 丹野眞智俊・石橋尚子(2006)「オノマトペにおける清音と濁音の比較」『日本心理学会第 70 回大会』。

ユーザの特性情報付きチャットボットとの雑談対話コーパスの概要

伊藤 紀子 (同志社大学文化情報学部)
岩下 志乃 (東京工科大学コンピュータサイエンス学部)
杉本 徹 (芝浦工業大学工学部)
林 篤司 (東京工科大学コンピュータサイエンス学部非常勤講師)
ト 秋予 (同志社大学大学院文化情報学研究科博士前期課程)

Overview of Human-Chatbot Dialogue Corpus with Personality Information

Noriko Ito (Doshisha University)
Shino Iwashita (Tokyo University of Technology)
Toru Sugimoto (Shibaura Institute of Technology)
Atsushi Hayashi (Tokyo University of Technology)
Bu Qiuyu (Doshisha University)

要旨

本発表では、構築中のチャットボットとの雑談対話コーパスについて紹介する。本コーパスは、参加者がチャットアプリを通してチャットボットと雑談したデータに対して「対話行為」のタグ、チャットボットの発言に対して「対話破綻」のタグを付与したものである。また、各雑談データに参加者の対話に関する印象・満足度、性格特性・社会的スキルの数値を付与している点で特色のあるコーパスとなっている。

雑談データの概要を説明したのち、付与された2種類のタグと参加者の性別、性格特性、社会的スキルとの関連について予備的分析を行った結果を報告する。

1. コーパス構築の背景と目的

書き言葉であれ話し言葉であれ言葉づかいというのは言語使用者の性別、年齢、生育地などの人口統計学的属性と関連があることが、これまで数多くの研究で報告されてきた。またこのような言語行動と使用者の属性の関連を研究するためのコーパスやデータセットの整備、公開は国立国語研究所や国立情報学研究所を中心として数多く行われている。

近年ではこのような基本属性以外にも、人の言語行動に関連するものとして言語使用者の性格というものが注目されている。例えば、グループディスカッション時の非言語、パラ言語、言語的特徴と性格特性の関連を検討した林ら (2015) の研究では、参加者の外向性、協調性、勤勉性によって特徴量に有意な違いがあったことが報告されている。Twitter ユーザに24種類のパーソナリティテストへ回答してもらい、Twitterの情報から個人のパーソナリティをどの程度推定できるかを検討したMori, Haruno (2020) などもある。井・伊藤 (2020a, b) では、人とチャットボットとの雑談対話において、チャットボットの発言に違和感があった時にユーザが雑談の継続を維持するためにとる方略は性別や性格特性によって異なることが報告されている。これらの研究を踏まえると性格特性の情報は言語行動の解明に有用であると考えられるが、研究者向けにデータが公開されているものはほとんどない。この

ような状況の中、駒谷・岡田 (2022) が 2020 年より公開しているマルチモーダル対話に関するデータセット Hazumi は、雑談している様子を写した映像や Kinect データとともに、性格特性のデータも研究利用可能な形で公開されている点で非常に有益である。しかしながら、Hazumi のようにデータが公開されているものは稀であり、ほとんどの場合研究者が独自にその都度データを収集するのが現状である。

著者らの研究目的は、ユーザの対話継続欲求を維持、増進するために、「個性」を持った雑談対話システムを実現することである。イメージとしては、スマートフォンや服を選ぶ時のように、豊富なラインナップからユーザが自分にぴったりの（中長期的には性別、年齢、性格など、短期的にはその日の気分に合わせて）チャットボットを選べるようにしたいと考えている。対話システムにおける「個性」に関しては、先に挙げた性別や年齢などの人口統計学的属性、心理学的な性格、ステレオタイプ的な方言使用や特定のキャラクターに似せたものなどが考えられるが、本研究では性格に着目することにした。また、著者らが収集するデータを幅広く研究に利活用してもらえるように、コーパスとして公開することを視野に含めて、データ収集を行うことにした。

データ収集の目的は、単に人とチャットボットとの雑談対話のデータを集めることに留まらず、「どのようなタイプの人」がチャットボットとの雑談において「どのような振る舞い」をするのか、チャットボットの振る舞いに対して「どのような印象、評価」を抱くのかを解明するために必要な情報を集めることである。チャットボットとの雑談データについては、既存のチャットボットを用いてデータを収集した。参加者が「どのようなタイプか」については性格特性、社会的スキル等の心理尺度を用いてデータを収集した。雑談対話において「どのような振る舞い」をするのかについては、参加者とシステムの各発話に対して対話行為タグを付与した。「どのような印象、評価」を抱くのかについては、対話終了時に対話がうまくいったかどうかの評価やシステムに対する印象を尋ねてデータを収集した。さらに、システムの各発話に対して対話破綻タグを付与した。

著者らの目指すところは、これらの雑談対話データと対話参加者に関する様々なデータを合わせて機械学習を行うことで個性を持った雑談対話システムを実装することであるが、現時点ではそこまで到達していない¹。本発表は 2021 年秋に仮公開したコーパス (岩下 2021) に含まれるデータの解説を中心に行うものとし、言語行動と言語使用者の性格の関連並びに本コーパスの活用に関する活発な議論の呼び水としたい。

2. コーパスについて

コーパス構築のために、2020 年 5 月から 2021 年 6 月にかけて、著者らの所属する同志社大学、東京工科大学、芝浦工業大学でデータ収集実験を行った。本研究は、3 大学の倫理審査委員会それぞれ承認を受けた後、実験参加者より書面による同意を得て実施した。現在までの実験参加者数は 75 名（女性 18 名、男性 57 名）である。女子学生の参加が少ない点については、もうしばらくデータ収集を続けて均等になるように努めたい。

データ収集実験では、次の 3 つのテーマについて、参加者に LINE を介してチャットボット「かたらい」との雑談チャットをしてもらった。

¹ 本コーパスの対話システムへの応用に関しては杉本ら (2022)、仲村・杉本 (2022)を参照されたい。

- ・新しいクラスで初めて知り合った隣の席の学生と初めて会話する感じで
- ・好きな〇〇について (〇〇はアニメ、映画、食べ物など何でも構わない。途中で変えても大丈夫)
- ・小学校の時の思い出

雑談チャット終了後、参加者に、事後アンケートに回答してもらい、4種類のデータを得た。

- ・チャット全体に対する印象・満足度評価 (木村ら 2005)
- ・チャット相手 (システム) に対する印象評価 (金 1999)
- ・参加者自身の性格特性、社会的スキル等に関する質問 (菊池 1988; 相川 1991; 藤本・大坊 2007; 小塩ら 2012)
- ・性別、年齢

後日、収集したデータに対して、雑談チャットには関与していない第三者にアノテーションを行ってもらい、以下の2種類のデータを得た。

- ・対話行為タグ (北陸先端科学技術大学院大学白井研究室 2017)
- ・対話破綻タグ (東中ら 2016)

以下では、執筆時点でアノテーションが終了している74件のデータについて主に報告する。

2.1 雑談データとアノテーションについて

74名の参加者が3つのテーマで行った雑談チャットから実験開始や終了の合図となる発言を除くと、参加者の発話が7326、システムの発話が7322取得できた。

2.1.1 対話行為タグについて

対話行為のタグは、『JAIST タグ付き自由対話コーパス』(2017)に基づき、雑談チャットには関与していない第三者1名が、参加者とシステムの発話に対して<自己開示><質問(Yes/No)><質問(What)><応答(Yes/No)><応答(平叙)><あいづち><フィラー><確認><要求>の9種類を付与したものである。現在は74件のチャットデータに対するアノテーションが終わり、内訳は表1の通りである。各対話行為の発話例を(1)-(9)に挙げる。

表 1. 対話行為タグの度数 (74 件)

	自己開示	質問(Yes/No)	質問(What)	応答(Yes/No)	応答(平叙)
参加者	2419	963	882	85	1873
システム	2983	276	118	101	1048

	あいづち	フィラー	確認	要求	総計
参加者	268	3	719	114	7326
システム	362	1	2225	208	7322

(1) <自己開示>のシステム発話例 (DS01-013)

参加者 : ゲーム私も好きです!

システム : 世界で最初に作られたコンピューターゲームは、チェスのゲームなんだよ。

参加者：チェスってやった事ありますか？

(2) <質問(Yes/No)>のシステム発話例 (SIT01-011)

参加者：私はうどん派だなー

システム：そば湯はあるけど、うどん湯ってあるのかな？

参加者：うどん湯は聞いたことない笑

(3) <質問(What)>のシステム発話例 (DS01-014)

参加者：ゲーム通して付き合う人のことどう思う？

システム：どうだと思えますか？？

参加者：私はいんじゃないかなって思う

(4) <応答(Yes/No)>のシステム発話例 (SIT01-006)

参加者：そうなの？

システム：そうよ

参加者：物知りだね

(5) <応答(平叙)>のシステム発話例 (DS01-013)

参加者：好きな食べ物ってなんですか？

システム：好きな食べ物は、蕎麦かな。

参加者：お蕎麦いいですね～！

(6) <あいづち>のシステム発話例 (DS01-003)

参加者：へえ～

システム：うん

参加者：物知りだね

(7) <フィラー>のシステム発話例 (TEU02-019)

参加者：うーん

システム：んー

参加者：んー

(8) <確認>のシステム発話例 (DS01-005)

参加者：小学生の時に流行ったやつね

システム：今の小学生は、スマホに携帯ゲームに、遊び方が昔とは違うよね。

参加者：好きな遊び聞いても全然違うものになりそう

(9) <要求>のシステム発話例 (TEU01-006)

参加者：小学校の頃の思い出もうほとんど無いんだよね

システム：小学校の周りにはスクールゾーンがあるから、運転するときは気をつけて。

参加者：小学生コワイ

2.1.2 対話破綻タグについて

対話破綻のタグは、東中ら (2016) に基づき、システム発話に対して対話を継続することの困難さの度合いを3段階で判断したものである (表 2)。

表 2. 第三者による対話破綻タグの度数 (74 件)

	度数
○ 破綻ではない： 当該システム発話のあと対話を問題無く継続できる	3380 (46.16%)
△ 破綻と言いきれないが、違和感を感じる発話： 当該システム発話のあと対話をスムーズに継続することが困難	1374 (18.77%)
× あきらかにおかしいと思う発話、破綻： 当該システム発話のあと対話を継続することが困難	2568 (35.07%)
総計	7322

全体として、対話を問題なく継続できると評価されたシステム発話は約半数に留まっていることがわかる。前節で説明した対話行為と掛け合わせると、<自己開示><質問(What)><要求>で特に対話の継続に困難さが生じるようである (表 3)。発話例を見てみると、同じ語句を使っているにもかかわらず○と判断されることもあれば (1, 3, 9)、×と判断されることもある (10-12) ことがわかる。

表 3. 対話破綻×対話行為の度数 (74 件)

	自己開示	質問(YesNo)	質問(What)	応答(YesNo)	応答(平叙)
○	1147	114	46	96	597
△	598	56	15	2	180
×	1238	106	57	3	271
総計	2983	276	118	101	1048

	あいづち	フィラー	確認	要求	総計
○	298	1	1017	64	3380
△	13	0	474	36	1374
×	51	0	734	108	2568
総計	362	1	2225	208	7322

(10) <×><自己開示>のシステム発話例 (DS01-007)

システム：サッカー選手はゲームが好きですね

参加者：ウイイレは知ってる？

システム：世界で最初に作られたコンピューターゲームは、チェスのゲームなんだよ。

参加者：ウイニングイレブンのこと

(11) <×><質問(What)>のシステム発話例 (DS01-005)

参加者 : おお！僕もゲームが好きです どんなゲームをしますか？

システム : どうだと思えますか？？

参加者 : うーん...ツムツムとか？

(12) <×><要求>のシステム発話例 (DS01-011)

参加者 : 違う～！小学校の思い出！

システム : 小学校の周りにはスクールゾーンがあるから、運転するときは気をつけて。

参加者 : あ！この前免許とったよ～！

参加者のうち 12 名についてはチャットを行った本人にも破綻タグを付与してもらった。本人と第三者の破綻タグの一致をまとめたものが表 4 である。

表 4. 本人×第三者の対話破綻タグの度数 (12 件)

本人 \ 第三者	第三者			総計
	○	△	×	
○ 破綻ではない	315 (72.6%)	79 (18.2%)	40 (9.2%)	434
△ 破綻と言い切れないが、違和感を感じる発話	80 (23.1%)	101 (29.1%)	166 (47.8%)	347
× あきらかにおかしいと思う発話、破綻	27 (9.2%)	51 (17.5%)	214 (73.3%)	292
総計	422	231	420	1073

本人が○や×と分類した発話については第三者も 7 割程度で同じように分類している。その一方で、本人が△の「破綻とは言い切れないが、違和感を感じる」ものについては、第三者が×「破綻である」と分類しているものが 5 割程度あることから本人よりも第三者の方がやや厳しい判断をしていることが読み取れる。表全体の重み付きカッパ係数は、 $\kappa=0.489$ で中程度の一致であった。

今後は、追加で破綻アノテーション実験を実施し、東中ら (2016) や船越ら (2016) を参考に、破綻箇所 (△および×) について対話破綻類型に基づく分類を行い、分類毎に判断のばらつきがあるか調べる予定である。また、アノテーターの性別や性格特性・社会的スキルについても関連があるか明らかにしたい。

2.2 事後アンケートについて

2.2.1 チャット全体に対する印象・満足度評価

雑談チャット終了後、参加者にはチャット全体に対する質問として、木村ら (2005) の 18 項目と「また会話してみたいと思ったか」について 8 件法で回答してもらった。前者の 18 項目については、木村ら (2005, p.19) を参考に、<会話調整因子>として「会話を上手く調節していた」「協力的に会話が進んでいた」の 2 項目、<会話集中因子>として「好意的に会話していた」「相互に興味をもって会話していた」の 2 項目、<ぎこちない会話因子>として「会話にくそうだった」「会話は緊張感を伴うものであった」の 2 項目の平均値を算出した。後者については数値をそのまま用い、会話満足度と呼ぶことにした。アノテーショ

ンが終わっている 74 名分の印象・満足度評価の得点の基本統計量をまとめたものが表 5 である。

表 5. チャット全体に対する印象・満足度評価の得点

		会話調整	会話集中	ぎこちない	会話満足度
		因子	因子	会話因子	
全体 74 名	平均	4.53	4.59	4.95	4.60
	標準偏差	1.38	1.56	1.21	1.05
女性 17 名	平均	4.59	5.00	4.94	4.90
	標準偏差	1.50	1.56	1.56	1.03
男性 57 名	平均	4.51	4.46	4.96	4.52
	標準偏差	1.35	1.56	1.10	1.05

2.2.2 チャット相手（システム）に対する印象評価

参加者にはチャット相手であった雑談対話システムに対する質問として、金 (1999) の性格特性 5 因子に基づいて設定された相手のパーソナリティ特徴に関する 10 項目とコミュニケーターの信頼性に関する 7 項目について 5 件法で回答してもらった。前者の 10 項目については、金 (1999, p.128) を参考に、＜外向性＞＜友好性＞＜誠実性＞＜神経質性＞＜開放性＞の 5 つのパーソナリティ側面の 2 項目の平均値を取って、各性格特性の値とした。後者についても＜専門性＞と＜真実性＞を構成する項目の平均値を算出した (表 6)。

表 6. システムに対する印象評価の得点

		相手のパーソナリティ特徴				
		外向性 E	友好性 A	誠実性 C	神経質性 N	開放性 O
全体 74 名	平均	3.43	3.36	3.09	2.16	3.35
	標準偏差	0.92	0.93	0.99	0.97	0.97
女性 17 名	平均	3.82	3.56	3.53	2.12	3.47
	標準偏差	0.86	0.88	0.91	0.76	1.18
男性 57 名	平均	3.32	3.31	2.96	2.17	3.32
	標準偏差	0.91	0.94	0.99	1.03	0.91

		コミュニケーターの信頼性	
		専門性	真実性
全体 74 名	平均	4.28	3.27
	標準偏差	0.84	0.64
女性 17 名	平均	4.35	3.54
	標準偏差	0.77	0.56
男性 57 名	平均	4.26	3.19
	標準偏差	0.86	0.64

2.2.3 参加者自身の性格特性、社会的スキル等に関する質問

A. TIPI-J

参加者には、自分自身に関する複数の尺度からなるアンケートに回答してもらっている。その中の1つがTIPI-J(小塩ら,2012)で、性格特性5因子に関する2項目ずつの質問について7件法で回答してもらった。アノテーションが終わっている74名の性格特性5因子および先行研究の基本統計量をまとめたものが表7である。

表 7. TIPI-J の得点

		外向性 E	協調性 A	勤勉性 C	神経症傾向 N	開放性 O
全体 74 名	平均	7.07	9.73	6.20	8.70	8.84
	標準偏差	3.27	2.42	2.25	2.90	2.58
女性 17 名	平均	6.76	9.53	5.88	9.18	9.53
	標準偏差	3.61	2.23	1.97	3.22	1.97
男性 57 名	平均	7.16	9.79	6.30	8.56	8.63
	標準偏差	3.16	2.47	2.32	2.78	2.71

小塩ら (2012) が大学生約 900 名を対象に行った調査と比較すると、全体的に<外向性>と<神経症傾向>が低く、<開放性>が高い傾向にあることが確認できた。小塩 (2012) では<外向性>と<開放性>で性差が見られたとの報告があったが、本コーパスの参加者には性差は見られなかった。

B. KiSS-18

KiSS-18 (菊池 1988, 1994) を用いて、社会的スキルに関する 18 の質問に 5 件法で回答してもらった。74 名の社会的スキル得点の基本統計量をまとめたものが表 8 である。

表 8. KiSS-18 の得点

		基本統計量
全体 74 名	平均	57.51
	標準偏差	11.09
女性 17 名	平均	52.29
	標準偏差	9.13
男性 57 名	平均	<u>59.07</u>
	標準偏差	11.15

菊池 (1994) が大学生約 200 名を対象に行った調査と比較すると、本コーパスの女子学生は得点が低く、男子学生は得点が高い傾向にあることが確認できた。また、菊池 (1994) の調査では性差は見られなかったが、本コーパスの参加者では、Welch の t 検定の結果、男子学生の方が有意に得点が高かった。

C. ENDCORE

コミュニケーション・スキルを測る ENDCORES の短縮版である ENDCORE (藤本, 大坊, 2007) を用いて、<自己統制><表現力><解読力><自己主張><他者受容><関係調整>に関する質問 6 項目について 7 件法で回答してもらった。74 名の各得点の基本統計量をまとめたものが表 9 である。

表 9. ENDCORE の得点

		基本スキル			対人スキル		
		自己統制	表現力	解読力	自己主張	他者受容	関係調整
全体 74 名	平均	4.92	4.01	4.65	4.45	5.24	4.11
	標準偏差	1.61	1.69	1.55	1.55	1.12	1.72
女性 17 名	平均	4.18	3.41	4.06	3.71	5.00	3.12
	標準偏差	1.67	1.77	1.60	1.57	1.22	1.87
男性 57 名	平均	<u>5.14</u>	4.19	4.82	<u>4.67</u>	5.32	<u>4.04</u>
	標準偏差	1.54	1.64	1.50	1.48	1.09	1.57

藤本・大坊 (2007) が大学生約 230 名を対象に行った調査と比較すると、全体的に<自己主張>が高く、<関係調整>が低い傾向にあることが確認できた。藤本・大坊 (2007) では性差が見られたとの報告があった。本コーパスの参加者では、Welch の t 検定の結果、下線で示した 3 項目 (<自己統制><自己主張><関係調整>) において男子学生の方が有意に各得点が高かった。

D. 特性シャイネス尺度

特性シャイネス²の程度を測定するために、特性シャイネス尺度 (相川, 1991) の 16 の質問に 5 件法で回答してもらった。74 名の特性シャイネス得点の基本統計量をまとめたものが表 10 である。

表 10. 特性シャイネス尺度の得点

		基本統計量
全体 74 名	平均	49.92
	標準偏差	13.79
女性 17 名	平均	53.47
	標準偏差	14.81
男性 57 名	平均	48.83
	標準偏差	13.28

相川 (1994) が大学生を対象に行った調査と比較すると、男女ともに得点が高い傾向にあることが確認できた。相川 (1994) が性別×学部の分散分析を行ったところ両要因とも主効

² “特定の社会的状況を越えて個人内に存在し、社会的不安という情動状態と対人的抑制という行動特徴をもつ症候群” (相川 1994, p.150)

果も交互作用も有意ではなかったとの報告があった。本コーパスの参加者においても性差は見られなかった。

2.3 予備的分析

ユーザ毎に各対話行為タグの割合を算出し、男女別に参加者の性格特性および社会的スキル等 13 項目との関連を調べるために相関分析を行ったところ、女子学生については一部の項目で中程度以上の関連が見られた (表 11)。

表 11. 参加者の性格特性・社会的スキル等と対話行為タグの割合の相関 (女子学生)

	r	F
U 外向性×U 質問(What)%	0.586	7.86 *
U 勤勉性×U 応答(Yes/No)%	-0.534	5.99 *
U 神経症傾向×U あいづち%	-0.643	10.59 **
U 開放性×U 応答(平叙)%	-0.485	4.61 *
U 表現力×U 応答(Yes/No)%	-0.656	11.33 **
U 他者受容×U 質問(Yes/No)%	-0.584	7.77 *
U KiSS-18×U 応答(Yes/No)%	-0.662	11.72 **
U KiSS-18×U 質問(Yes/No)%	-0.545	6.35 *

* $p < 0.05$, ** $p < 0.01$

相関分析の結果より、＜外向性＞とユーザの＜質問 (What)＞の割合との間に正の相関、＜勤勉性＞と＜応答 (Yes/No)＞の割合との間に負の相関、＜神経症傾向＞と＜あいづち＞の割合との間に負の相関、＜開放性＞と＜応答 (平叙)＞の割合の間に負の相関が見られた。ENDCORE の＜表現力＞と＜応答 (Yes/No)＞の割合との間に負の相関、＜他者受容＞と＜質問 (Yes/No)＞の間に負の相関がみられた。KiSS-18 とユーザの＜応答 (Yes/No)＞およびユーザの＜質問 (Yes/No)＞の割合との間に負の相関が見られた。

次に、ユーザ毎に各対話破綻タグの割合を算出し、男女別に参加者の性格特性および社会的スキル等 13 項目との関連を調べるために相関分析を行ったところ、女子学生の＜外向性＞でのみ中程度以上の関連が見られた (表 12)。

表 12. 参加者の性格特性・社会的スキル等と対話破綻タグの割合の相関 (女子学生) z

	r	F
U 外向性×破綻△%	0.594	8.18 *
U 外向性×破綻○%	-0.529	5.84 *

相関分析の結果より、＜外向性＞と＜破綻△＞の割合との間に正の相関、＜破綻○＞との間に負の相関が見られた。

3. おわりに

本発表では、構築中の雑談対話コーパスに含まれるデータについて解説した。現在岩下 (2021) から入手できるのは上記データの一部であるため、整備が終わり次第、公開に努めたい。

予備的分析として、参加者の性格特性・社会的スキル等と対話行為、破綻との相関を調べた結果、女子学生にはいくつかの項目間で関連があることがわかったが、男子学生には関連が見られなかった。この違いの要因として、性格特性・社会的スキル等 13 項目のうち社会的スキルとコミュニケーション・スキルにおいて性差が見られたものについていずれも男子学生の方が得点が高かったことが考えられるが、現時点で男子学生と比較して女子学生のデータが少ないこと（約 3 分の 1）が関係している可能性もあることから、今後データ収集を行い、改めて検討する必要がある。

謝 辞

本研究は JSPS 科研費 JP19K12191 「雑談対話システムへの個性の付与とそのコミュニケーションへの影響について」の助成を受けて行われたものである。また、本研究で使用した雑談チャットおよび性格特性等のデータ収集に際して、同志社大学、東京工科大学、芝浦工業大学在学学生の方々にご協力いただいた。ここに記して感謝いたします。

文 献

- 相川充. (1991). 特性シャイネス尺度の作成および信頼性と妥当性の検討に関する研究. 心理学研究, 62:3, pp. 149-155.
- 井凌泓・伊藤紀子. (2020a). 雑談対話システムとの対話にみられる対話破綻とユーザのコミュニケーション方略. Proceedings of JASFL, 14, pp. 1-14.
- 井凌泓・伊藤紀子. (2020b). 雑談対話システムとの対話にみられる対話破綻とユーザのコミュニケーション方略—ユーザ特性による違い—. 日本機能言語学会第 28 回秋期大会.
- 小塩真司・阿部晋吾・カトローニ ピノ. (2012). 日本語 Ten Item Personality Inventory (TIPI-J) 作成の試み. パーソナリティ研究, 21:1, pp. 40-52.
- 菊池章夫. (1988). 思いやりを科学する: 向社会的行動の心理とスキル. 川島書店.
- 菊池章夫 (1994) 社会的スキルを測る KiSS-18 のこと. 菊池章夫, 堀毛一也, 編. 社会的スキルの心理学. 川島書店. pp. 177-183.
- 木村昌紀・余語真夫・大坊郁夫. (2005). 感情エピソードの会話場面における表出性ハロー効果の検討. 感情心理学研究, 12:1, pp. 12-23.
- 金官圭. (1999). CMC (computer-mediated communication) における印象形成に関する探索的研究. 社会心理学研究, 14:3, pp. 123-132.
- 沢田慶. (2021). 身近になった対話システム: 4. 一般ユーザとの雑談会話のための AI チャットボット. 情報処理, 62:10, pp. e19-e23.
- 杉本徹・岩下志乃・伊藤紀子・林篤司. (2022). 性格情報を持つ対話コーパスの構築と対話システムへの応用に向けた検討. 第 38 回ファジィシステムシンポジウム (FSS2022), FE1-1.
- 仲村詩穂・杉本徹. (2022). 性格情報を持つ対話コーパスに基づいた話者の性格特性と発言特徴の関係の分析. 第 38 回ファジィシステムシンポジウム (FSS2022), FE1-2.
- 林佑樹・二瓶英巳雄・中野有紀子・黄宏軒・岡田将吾. (2015). グループディスカッションコーパスの構築および性格特性との関連性の分析. 情報処理学会論文誌, 56:4, pp. 1217-1227.
- 東中竜一郎・船越孝太郎・荒木雅弘・塚原裕史・小林優佳・水上雅博. (2016). テキストチャットを用いた雑談対話コーパスの構築と対話破綻の分析. 自然言語処理, 23:1, pp. 59-86.

- 藤本学・大坊郁夫. (2007). コミュニケーション・スキルに関する諸因子の階層構造への統合の試み. *パーソナリティ研究*, 15:3, pp. 347-361.
- 船越孝太郎・東中竜一郎・稲葉通将・小林優佳・菅原朔・高梨克也・大塚裕子・小磯花絵・坊農真弓. (2016). 対話破綻検出チャレンジにおける対話破綻データと破綻検出結果の分析: 主観性の高い言語データにおける言語処理に関して. *言語処理学会第 22 回年次大会発表論文集*, pp. 433-436.
- Kazuma Mori, and Masahiko Haruno. (2020). Differential ability of network and natural language information on social media to predict interpersonal and mental health traits. *Journal of Personality*, 89:2, pp. 228-243.

関連 URL

- 岩下志乃. (2021). チャットボットとの対話コーパスと性格特性・社会的スキルのデータ公開について. https://www2.teu.ac.jp/iws/chat_corpus.pdf
- インターメディアプランニング・NTT ドコモ. (2022). かたらい. <https://www.katarai/>
- 北陸先端科学技術大学院大学 (JAIST) 白井研究室. (2017). JAIST タグ付き自由対話コーパス. 言語資源協会 (GSK): <https://www.gsk.or.jp/files/catalog/GSK2017-B/Readme.pdf>
- 駒谷和範・岡田将吾. (2022). 大阪大学マルチモーダル対話コーパス Hazumi 概要 (オンライン収録版). 情報学研究データリポジトリ (IDR): <https://www.nii.ac.jp/dsc/idr/rdata/Hazumi/documents/HazumiOverviewOnline.pdf>

疑問・非疑問発話の韻律典型性を捉える試み —日本語母語話者と学習者の比較—

波多野 博頭 (筑波大学) †

王 可心、陳 凱僑、林 良子 (神戸大学)

An Attempt to Understand Prosodic Typicality of Interrogative and Non-Interrogative Speech: Comparison of Native Speakers and Learners of Japanese

Hiroaki Hatano (University of Tsukuba) †

Kexin Wang, Kaiqiao Chen, Ryoko Hayashi (Kobe University)

要旨

日本語母語話者および学習者による疑問・非疑問発話を対象に、韻律の定量的な比較・検討を行なった。科研「三重データコーパスを用いた日本語韻律の習得・評価に関する多面的研究」によって構築中の音声コーパス「KANI-J (Kobe Archive of Nonnative Intonation in Japanese)」を用い、中国語・イタリア語・韓国語・ロシア語母語話者による動詞一語の発話を分析した。韻律データに階層的クラスタ分析を行なうことで典型性を捉えるとともに、学習者の日本語学習歴からその要因を検討した。また、疑問発話の韻律を日本語母語話者と学習者で比較し、学習者韻律の特徴を分析した。その結果、学習者では非疑問・疑問の別によってアクセント核の現れが異なることや、疑問上昇に至るまでの韻律動態に違いが見られた。

1. はじめに

自然な発音やイントネーションで話したいという日本語学習者のニーズは高く (佐藤 1998)、特にビジネス等の公的場面や国内の長期間生活では意思疎通以上の「外国人なまり」自体の克服が要求される (小河原 2001)。そのため、学習者の韻律動態や母語話者との異同を客観的に捉え、その典型性を定量的に理解することは、より実態に即した音声教育を考え上での基礎的な研究として重要であろう (前川 2009)。

日常会話では相手に何かを問いかける疑問発話が頻繁に現れ、その韻律は音声コミュニケーションの自然性に影響する (波多野・石井 2017)。日本語学習者の疑問発話に見られる韻律的特徴については、これまで文部省の重点領域研究「日本語音声における韻律的特徴の実態とその教育に関する総合的研究」のD1班「外国人を対象とする日本語教育における音声教育の方策に関する研究」での大規模なデータ収集および分析結果が報告されている (鮎澤・谷口 1991; 鮎澤 1992; 鮎澤 2003)。

本研究はそれらを継承・発展させたプロジェクト (林他 2018) のもと、日本語学習者による韻律の典型的な動態を定量的な音響分析に基づいて明らかにすることを目的とする。具体的には、音響分析から得た韻律データに階層的クラスタ分析を行なうことで大局的な観点から典型性を捉え、その要因を学習者の日本語学習歴から検討する。さらに、疑問発話における韻律を日本語母語話者と比較することによって、局所的な観点から学習者韻律の特徴を明らかにする。

† hatano.hiroaki.ge[at]u.tsukuba.ac.jp

2. データの概要

分析に用いたのは、科研「三重データコーパスを用いた日本語韻律の習得・評価に関する多面的研究」によって構築中の音声コーパス「KANI-J (Kobe Archive of Nonnative Intonation in Japanese)」である。KANI-J コーパスは日本語韻律教育の発展に寄与することを目的に、様々なタスクからなる日本語学習者および日本語母語話者の発話データが収集されている(林他 2018; Ueyama et al. 2019)。タスクは「疑問文」「曖昧文」「感情・態度」「母音無声化」「ダウンステップ」「フォーカス(強調)」の6つで構成されており、本研究はこれらのうち「疑問文」タスクを分析対象とした。

収録は主にオンライン音声収集システムである OVR (Online Voice Recorder、東京大学峯松研究室が開発) を用いて行なわれた。発話者はヘッドセットマイクロフォン等を装着し、インターネットに接続された PC 画面の指示に従い、表示されているテキストを発話した。音声はサンプリング周波数 48 kHz、量子化ビット数 18-bit で収録され、データは web 上に収集された。なお、本コーパスは科研プロジェクト終了後に公開される予定である。

本研究では、日本語母語話者(以下 JPN) 15 名、日本語学習者 73 名の計 88 名による発話を分析に用いた。JPN は男性 7 名・女性 8 名で、全て 20 代の首都圏(東京・埼玉・千葉・神奈川)出身者である。学習者は全て自国の大学で日本語を学んでいる学生で、母語は中国語(CHN)・イタリア語(ITA)・韓国語(KOR)・ロシア語(RUS)である。学習者の人数(男女)と平均日本語学習年数を表 1 に示す。

表 1 日本語学習者の人数(男女)と平均日本語学習年数

母語	人数(男性・女性)	平均日本語学習年数(SD)
CHN	20 (2・17)	2.0 (1.0)
ITA	16 (4・12)	2.9 (0.3)
KOR	23 (6・17)	3.5 (1.9)
RUS	14 (3・11)	2.4 (1.6)

分析した発話資料は、2 モーラ動詞「寝る」「飲む」と 3 モーラ動詞「終わる」「食べる」の各一語によるもので、アクセントがそれぞれ平板式と起伏式からなる。同じ語を非疑問・疑問の 2 種類で発話したため、1 人あたり 8 語、計 704 発話(8 語×88 名)を分析した。

発話者は PC 画面上に表示されたテキスト文末の「。」と「？」を手がかりに、非疑問と疑問を判断し(例:「寝る。」「寝る?」)、自分が適切だと思う音調で発話した。発話の際、モデル音声の聴取やアクセント・イントネーションの説明などは行なわれていない。

3. 分析方法

発話データに対し、音声認識エンジン Julius (Lee and Kawahara 2009) の音素セグメンテーションキット(segmentation-kit-4.3.1)を用いて音素境界の自動アライメントを行なった。結果は全て目視によって確認し、必要であれば境界位置の修正を行なった。その後、praat (version 6.1.40) で基本周波数(f_0)を抽出した。 f_0 抽出パラメータは、Time step: 0.001、Pitch floor: 50 Hz、Pitch ceiling: 600 Hz である。各話者のデータを比較可能にするため、話者ごとに f_0 を標準化(z-score normalization)、発話区間全体の時間を正規化(min-max normalization)した。なお、本研究では f_0 のみを分析対象とする。

マイクロプロソディの影響や抽出欠損に対処するため、以下の処理を行ない全体の f_0 をモデリングした(波多野他 2018; 波多野他 2021; 波多野他 2022)。まず、各母音区間においてフレーム全体の 50%以上から f_0 が抽出された箇所を対象に、 f_0 と時間情報から回帰直線を計算した。その際、区間内の f_0 から四分位数を求め、「第 1 四分位数 * 1.5 以上」「第 3 四分位数 * 1.5 以下」に該当した f_0 は外れ値として回帰直線の計算に含めなかった。また、疑問上昇の動態を捉えるため、語末母音は 2 等分して回帰直線を求めた。次に、回帰直線上

の時間方向で 25% と 75% の位置に制御点を設定した。最後に、これら制御点に基づいてスプライン補間を行なった。この処理により、 f_0 の局所的な不安定さに対処するとともに、発話全体の f_0 を欠損なく再構築した。なお、語頭・語末母音に f_0 の欠損が生じている場合は、直後・直前の母音の回帰直線における 0%・100% の位置にあたる f_0 を延伸して代用した。

図 1 に、JPN による疑問「終わる」(左図) および非疑問「食べる」(右図) の f_0 モデリング例を示す。図中、○印が f_0 の実測値、点線が回帰直線、●印が制御点、実線がそれらを通るスプライン曲線である。左図では/o/冒頭で外れ値となった f_0 を小さい●印で示しており、これらを回帰直線の計算に含めないことで妥当なモデリングとなっていることがわかる。また、右図では語末母音後半部/u2/の f_0 が欠損(区間内で 50% 以下の抽出)しているが、語末母音前半部/u1/の回帰直線で 100% の位置にある f_0 の値を延伸することで補完している。補完された f_0 から再度回帰直線を計算し、制御点を設定することで以降の分析に用いた。

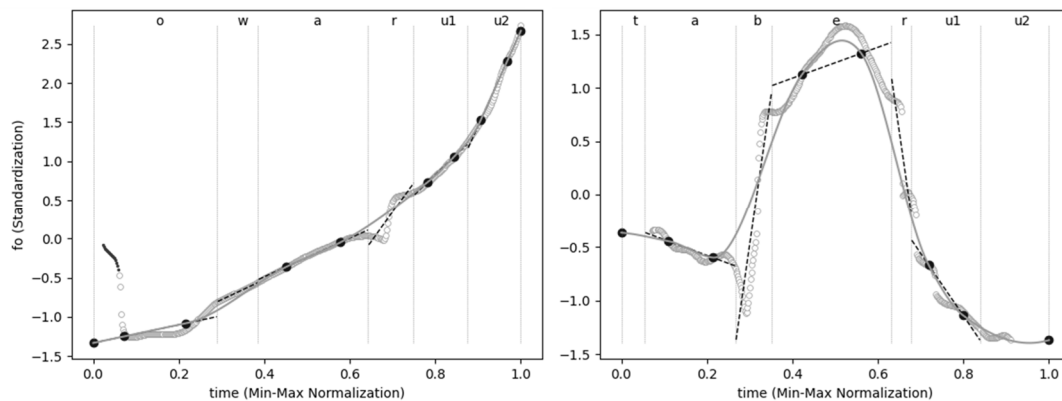


図 1 日本語母語話者による発話の f_0 のモデリング例 (左図:「終わる」右図:「食べる」)

動詞ごとに、非疑問・疑問を区別せず制御点の f_0 値を用いて階層的クラスター分析(ウォード法、ユークリッド距離を使用)を行ない、母語別で韻律の類型化を試みた。クラスター数の決定においては、デンドログラムおよび各ケースの所属数や韻律外形等を考慮した。また、各クラスターがどのような要因によって分類されるのかについて、学習者の日本語学習歴(年数)から検討した。さらに、疑問発話において学習者の f_0 動態が JPN と異なるかどうかをみるため、制御点ごとに f_0 の値をマン・ホイットニーの U 検定によって検証した。

4. 結果

4.1 2モーラ「寝る」「飲む」

4.1.1 クラスター分析の結果

2モーラ動詞について、図 2 に平板式「寝る」、図 3 に起伏式「飲む」の各制御点における f_0 分布を母語・クラスター別に示す(左図:非疑問、右図:疑問)。図中、V1・V2 はそれぞれ第一母音と第二母音を表し、a と b は当該母音における 25%・75% 位置にある制御点ということを示す(以下同様)。なお、末母音(ここでは V2) は二等分してそれぞれに制御点を置いているため、a, b, c, d の 4 つ存在している。非疑問・疑問別に各母語で最も所属数の多かったクラスターを実線で囲んで示している。

各クラスターの f_0 動態を観察すると、おおよそ「寝る」(図 2) では cluster 1 が平板式(ただし KOR の非疑問を除く)、cluster 2 が起伏式のアクセントを反映した韻律を示している。「飲む」(図 3) は逆で、cluster 1 が起伏式、cluster 2 が平板式を反映している。

図から、JPN では全ての話者で動詞のアクセント型通りの韻律(「寝る」は平板、「飲む」は起伏)に分類されている。疑問発話においても、当然ながら非疑問発話でのアクセント型が保たれたまま V2 で f_0 を上昇させている。

一方、学習者では非疑問・疑問でクラスターが異なる場合が多い。「寝る」では非疑問で多くのCHNがJPNと同じ平板式 (cluster 1) に分類されるものの、他では起伏式 (cluster 2) が多く、疑問では全ての母語で平板式の上昇 (cluster 1) が多い。「飲む」では、非疑問で全母語が起伏式 (cluster 1) 多勢であり、疑問ではCHNとKORの多くがJPNと同じ起伏式の上昇 (cluster 1) であるものの、他母語では平板式の上昇 (cluster 2) が多い (CHNとKORも平板式の上昇自体は多い)。以上から、学習者は「非疑問は有核・疑問は無核の上昇」で発話する傾向があると考えられる。RUSでは両語とも非疑問・疑問でクラスターの分類結果が特に偏っており、この傾向が顕著である。

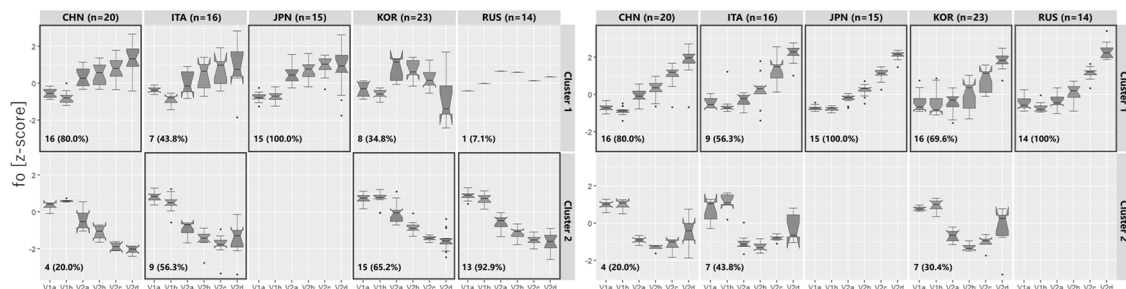


図 2 平板式「寝る」の母語別クラスター分類結果および各制御点における f_0 分布 (左図：非疑問、右図：疑問)

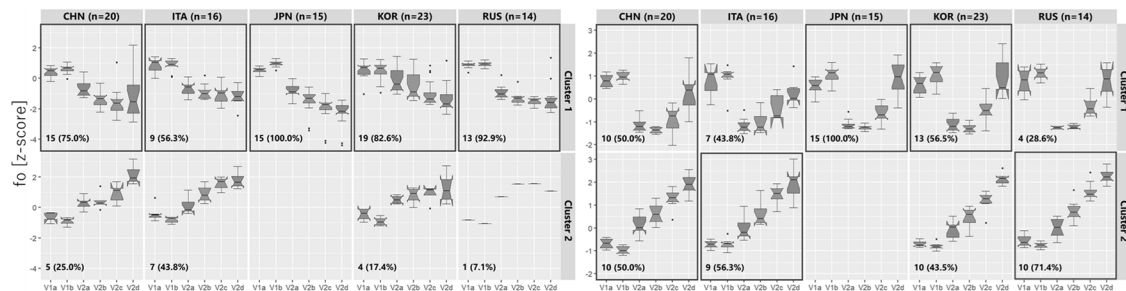


図 3 起伏式「飲む」の母語別クラスター分類結果および各制御点における f_0 分布 (左図：非疑問、右図：疑問)

4.1.2 クラスターと日本語学習年数の関係

学習者の母語別に、各クラスターにおける平均日本語学習年数を示す (表 2 「寝る」、表 3 「飲む」)。表では、非疑問・疑問ごとに各母語で最大のクラスターに網掛けをしている (以下同様)。cluster 1 と 2 の日本語学習年数に違いがあるかを見るため、非疑問・疑問で母語ごとにマン・ホイットニーの U 検定を行なった (RUS は各クラスターへの所属数の偏りが強いので除く)。その結果、全てのケースで日本語学習年数に有意差が見られなかった。

表 2 「寝る」の学習者母語別クラスター所属数および平均日本語学習年数 (SD)

Cluster	寝る (非疑問)				寝る (疑問)			
	CHN	ITA	KOR	RUS	CHN	ITA	KOR	RUS
1	16 名	7 名	8 名	1 名	16 名	9 名	16 名	14 名
	1.8 (0.9)	2.8 (0.4)	4.0 (2.4)	2.0	1.9 (1.0)	2.8 (0.4)	3.2 (1.4)	2.4 (1.6)
2	4 名	9 名	15 名	13 名	4 名	7 名	7 名	—
	2.5 (1.0)	3.0 (0.0)	3.3 (1.5)	2.4 (1.7)	2.0 (1.2)	3.0 (0.0)	4.3 (2.6)	—

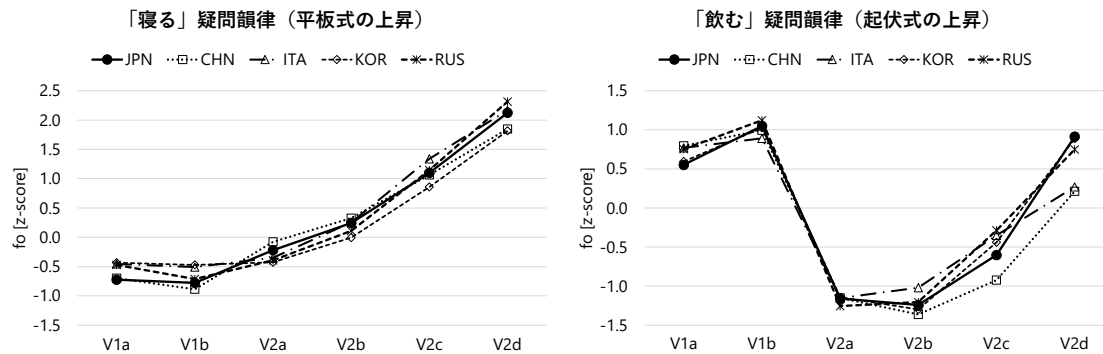
表 3 「飲む」の学習者母語別クラスター所属数および平均日本語学習年数 (SD)

Cluster	飲む (非疑問)				飲む (疑問)			
	CHN	ITA	KOR	RUS	CHN	ITA	KOR	RUS
1	15 名	9 名	19 名	13 名	10 名	7 名	13 名	4 名
	2.0 (1.0)	3.0 (0.0)	3.5 (1.3)	2.4 (1.7)	1.9 (0.9)	2.9 (0.2)	4.0 (2.1)	2.0 (1.6)
2	5 名	7 名	4 名	1 名	10 名	9 名	10 名	10 名
	1.7 (1.0)	2.8 (0.4)	3.5 (3.8)	1.8	2.1 (1.1)	2.9 (0.3)	2.9 (1.4)	2.6 (1.7)

4.1.3 疑問韻律の比較

疑問韻律における JPN と学習者各母語の異同を検討するため、「寝る」(平板式の上昇) と「飲む」(起伏式の上昇) の制御点 f_0 を用いてマン・ホイットニーの U 検定を行なった。

検定の結果、平板式上昇の「寝る」(図 4 左) では、V1b で CHN が JPN よりも有意に低かった (CHN: -0.89; JPN: -0.78; $U=68, p<.05$)。両者は V2a の f_0 に有意差がないことから、CHN では V1 から V2 にかけて疑問上昇が JPN よりも跳躍的に上昇することが示唆される。また、V2d では KOR の f_0 が JPN よりも有意に低かった (KOR: 1.82; JPN: 2.12; $U=54, p<.01$)。そのため、KOR による平板式の上昇では、JPN より末尾であまりピッチを上げない特徴がうかがえる。一方、起伏式上昇の「飲む」(図 4 右) では、いずれの制御点でも JPN と学習者の母語間に f_0 の有意差はみられなかった。

図 4 2 モーラ動詞の疑問韻律における各制御点の平均 f_0

4.2 3 モーラ「終わる」「食べる」

4.2.1 クラスター分析の結果

3 モーラ動詞について、図 5 に平板式「終わる」、図 6 に起伏式「食べる」の各制御点における f_0 分布を母語・クラスター別に示す (左図: 非疑問、右図: 疑問)。

各クラスターの f_0 動態を観察すると、例外があるものの大凡以下のようにまとめられる。「終わる」(図 5) では cluster 1 が平板式の非上昇、cluster 2 が平板式の上昇、cluster 3 が起伏式の韻律を反映している。「食べる」(図 6) では、cluster 1 が起伏式の非上昇、cluster 2 が平板式の上昇、cluster 3 が起伏式の上昇を反映している。

図から、JPN ではほぼアクセント型通りの韻律(「終わる」は平板式、「食べる」は起伏式)を示している。ただし、全員が同じクラスターに属しておらず、例外となったケースも存在する。これらは f_0 の抽出精度やモデリング結果の精査等、今後その原因について検討していく必要があるだろう。

学習者の韻律を見ると、「終わる」ではどの母語でも非疑問が起伏式（cluster 3）、疑問が平板式の上昇（cluster 2）に最も多く分類されている。「食べる」でも、非疑問は起伏式の非上昇（cluster 1）がいずれの母語でも最多である。疑問ではCHNとKORの多くがJPNと同じ起伏式の上昇（cluster 3）に分類されているものの、KORではcluster 2と3の差が僅かであることを考慮すると、他母語を含め全体的に平板式の上昇（cluster 2）が優勢であろう。以上から、2モーラと同じく、3モーラでも「非疑問は有核・疑問は無核の上昇」という学習者の韻律傾向が現れている。なお、ここでもRUSではこの傾向が特に顕著である。

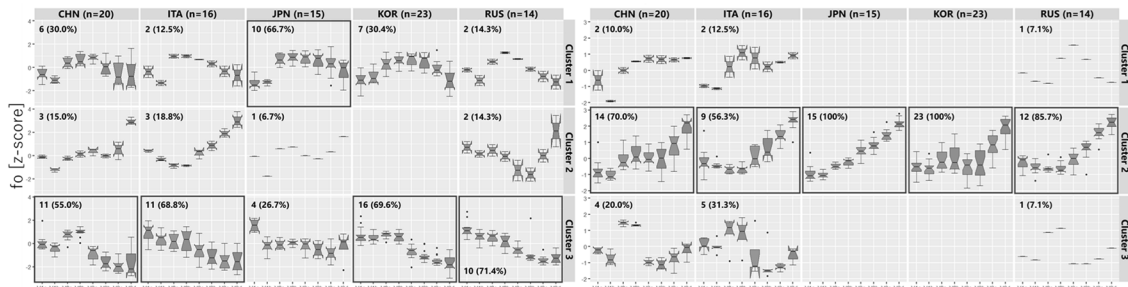


図 5 平板式「終わる」の母語別クラスター分類結果および各制御点における f_0 分布 (左図：非疑問、右図：疑問)

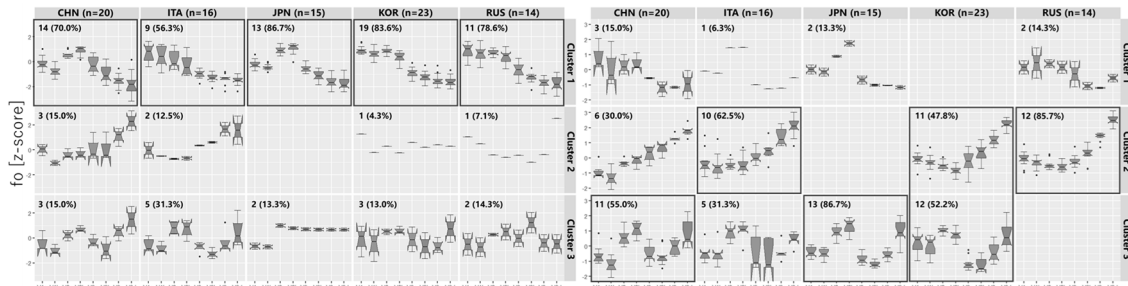


図 6 起伏式「食べる」の母語別クラスター分類結果および各制御点における f_0 分布 (左図：非疑問、右図：疑問)

4.2.2 クラスターと日本語学習年数の関係

学習者の母語別に、各クラスターにおける平均日本語学習年数を示す（表 4「終わる」、表 5「食べる」）。統計的検証を行なうには各クラスターに属するサンプルが少ないため定性的な観察に留まるが、日本語学習年数の高低と韻律分類の結果に明確な関係は見られない。

表 4 「終わる」の学習者母語別クラスター所属数および平均日本語学習年数 (SD)

Cluster	終わる（非疑問）				終わる（疑問）			
	CHN	ITA	KOR	RUS	CHN	ITA	KOR	RUS
1	6名 1.7 (0.8)	2名 3.0 (0.0)	7名 3.6 (2.8)	2名 1.9 (0.1)	2名 1.0 (0.0)	2名 3.0 (0.0)	—	1名 2.0
2	3名 1.7 (1.2)	3名 2.7 (0.6)	—	2名 3.0 (0.0)	14名 2.2 (1.0)	9名 2.9 (0.3)	23名 3.5 (1.9)	12名 2.6 (1.6)
3	11名 2.2 (1.0)	11名 3.0 (0.2)	16名 3.5 (1.4)	10名 2.4 (1.9)	4名 1.7 (1.0)	5名 2.9 (0.2)	—	1名 0.0

表 5 「食べる」の学習者母語別クラスター所属数および平均日本語学習年数 (SD)

Cluster	食べる (非疑問)				食べる (疑問)			
	CHN	ITA	KOR	RUS	CHN	ITA	KOR	RUS
1	14名 1.8 (0.9)	9名 3.0 (0.0)	19名 3.3 (1.3)	11名 2.6 (1.7)	3名 1.0 (0.1)	1名 3.0	—	2名 1.0 (1.4)
2	3名 1.6 (1.0)	2名 3.0 (0.0)	1名 9.0	1名 3.0	6名 2.3 (1.1)	10名 2.9 (0.3)	11名 3.4 (1.3)	12名 2.6 (1.6)
3	3名 3.2 (0.3)	5名 2.7 (0.4)	3名 3.3 (2.5)	2名 0.9 (1.3)	11名 2.1 (0.9)	5名 2.9 (0.2)	12名 3.7 (2.3)	—

4.2.3 疑問韻律の比較

疑問発話における韻律動態が JPN と学習者で異なるかどうかを検討するため、「終わる」(平板式の上昇)と「食べる」(起伏式の上昇)における各制御点の f_0 を用いてマン・ホイットニーのU検定を行なった。図7に各制御点における平均 f_0 に基づいた疑問韻律の動態を母語別に示し(左図:「終わる」、右図:「食べる」)、表6と表7に各制御点の平均 f_0 一覧を検定結果とともに示す。

検定の結果、「終わる」(図7左および表6)ではCHNとKORで最終母音V3のa,b,cの f_0 がJPNよりも有意に低かった。ただ、両母語ともV3dではJPNとの f_0 に有意差がみられないことから、CHNとKORでは最終母音の前半部でピッチを低く抑えて末尾で急激に上昇させる特徴があるといえる。また、ITAとRUSではV1a,bの f_0 がJPNより有意に高く、V2a,bは有意に低かった。JPNではV1からV3にかけて徐々に上昇していくが、ITAとRUSではV2で一旦ピッチが沈み込んでからV3で上昇するような動態が特徴的である。

「食べる」(図7右および表7)では、CHNでV1a,bの f_0 がJPNよりも有意に低く、V3bとV3cで有意に高い。このことから、CHNによる起伏式の疑問では、句頭上昇がやや誇張された形で実現された後、アクセント核によるピッチ下降の沈み込みが浅いまま上昇へ移行していく特徴がうかがえる。また、KORでは、V2bとV3aでJPNよりも有意に低い。そのため、KORでは起伏式の疑問においてアクセント核によるピッチ下降のタイミングがJPNよりも早いことが示唆される。

以上の結果が知覚上でどのような影響を与えているのかを明らかにするためには、今後更に検証が必要であろう。

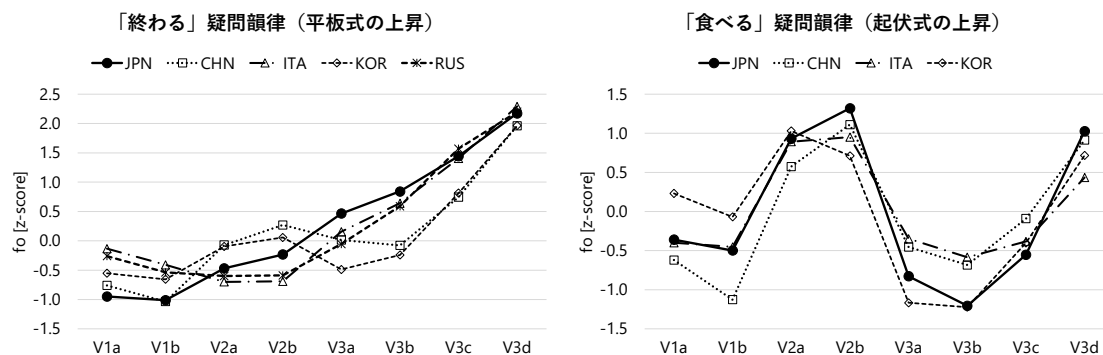
図 7 3モーラ動詞の疑問韻律における各制御点の平均 f_0

表 6 「終わる」疑問韻律における各制御点の平均 f_0

	n	V1a	V1b	V2a	V2b	V3a	V3b	V3c	V3d
JPN	15	-0.95	-1.01	-0.47	-0.23	0.47	0.84	1.44	2.17
CHN	14	-0.76	-1.04	-0.07	0.27*	0.02*	-0.08**	0.74*	1.96
ITA	9	-0.13**	-0.41**	-0.70*	-0.69**	0.16	0.64	1.41	2.29
KOR	23	-0.55*	-0.66**	-0.10	0.06	-0.49**	-0.24**	0.81*	1.97
RUS	12	-0.26**	-0.54**	-0.60**	-0.59**	-0.05*	0.59	1.57	2.20

(** = $p < 0.1$, * = $p < 0.5$ 。ただし、いずれも JPN との比較)表 7 「食べる」疑問韻律における各制御点の平均 f_0

	n	V1a	V1b	V2a	V2b	V3a	V3b	V3c	V3d
JPN	13	-0.36	-0.50	0.93	1.32	-0.83	-1.21	-0.55	1.03
CHN	11	-0.62*	-1.13**	0.57	1.11	-0.45	-0.68**	-0.09*	0.91
ITA	5	-0.40	-0.45	0.89	0.95	-0.35	-0.59	-0.38	0.44
KOR	12	0.23*	-0.07	1.03	0.71**	-1.17*	-1.22	-0.40	0.72

(** = $p < 0.1$, * = $p < 0.5$ 。ただし、いずれも JPN との比較)

5. おわりに

本研究では、日本語母語話者および学習者（中国語・イタリア語・韓国語・ロシア語の各母語話者）による動詞一語の非疑問・疑問発話を通して、その韻律の典型的な形を定量的に捉える試みを行なった。結果を以下にまとめる。

- 韻律クラスターの分類結果から、学習者は非疑問ではアクセント核を付けて（有核で）発話し、疑問では付けずに（無核で）発話する傾向が明らかとなった。この傾向はどの母語でも見られるが、特にロシア母語話者の発話に顕著であった。
- 韻律クラスターの分類結果と学習者の日本語学習年数に大きな違いは見られなかった。そのため、日本語母語話者と同じクラスターに分類される学習者たちが特別に長い学習歴を有しているわけではない。
- 疑問発話の韻律を日本語母語話者と学習者で比較した結果、母語ごとに特徴的な違いがいくつか明らかとなった。特に、3 モーラ動詞で末尾の上昇に至るまでの動態では、中国語と韓国語母語話者では末母音の前半部でピッチを低く抑えるが、イタリア語とロシア語母語話者では第 2 母音で低く抑える。

今後は学習者の母語における音声の特徴を整理し、今回の結果との対応を検討したい。また、クラスターの分類結果にアクセント型の評価を加えることで、より知覚に即した典型性の把握を行なう予定である。

謝 辞

本研究は JSPS 科研費 20K00610, 21H04348 の助成を受けたものです。

文 献

- Akinobu Lee, and Tatsuya Kawahara (2009). "Recent development of open-source speech recognition engine Julius", In proceedings of *Asia-Pacific Signal and Information Processing Association 2009 Annual Summit and Conference*, pp.131-137.

- Motoko Ueyama, Ryoko Hayashi, and Aaron Lee Albin (2019). “Development of an L2 Japanese speech corpus for the comparison of prosody across diverse L1 groups”, Presented at the 2nd Phonetics and Phonology in Europa 2019 (PaPE).
(https://www.researchgate.net/publication/333972766_Development_of_an_L2_Japanese_speech_corpus_for_the_comparison_of_prosody_across_diverse_L1_groups よりダウンロード可能)
- 鮎澤孝子 (1992). 「日本語の疑問文の韻律的特徴」『日本語の韻律に見られる母語の干渉 (2)ー音響音声学的対照研究ー』重点領域研究「日本語音声」研究報告書, pp.1-20.
- 鮎澤孝子 (2003) 「外国人学習者の日本語アクセント・イントネーション習得」『音声研究』7(2), pp.47-58.
- 鮎澤孝子・谷口聡人 (1991) 「日本語音声の韻律的特徴」『日本語の韻律に見られる母語の干渉ー音響音声学的対照研究ー』重点領域研究「日本語音声」研究報告書, pp.1-24.
- 小河原義朗 (2001). 「日本語非母語話者の話す日本語の発音に対する日本人の評価意識ー社会人の場合ー」『日本語教育方法研究会誌』8(2), pp.10-11.
- 佐藤友則 (1998). 「韓国および台湾の日本語学習者のニーズ調査」『言語科学論集』2, pp.49-60.
- 波多野博頭 (2021). 「機械学習による日本語アクセント型自動判定の検討：日本語教科書音声を用いた分析」日本語音声コミュニケーション学会 2021 年度秋季大会発表資料
(https://researchmap.jp/hhatano/presentations/36514692/attachment_file.pdf よりダウンロード可能)
- 波多野博頭・アルビン エレン・王睿来・石井カルロス寿憲 (2018). 「機械学習を用いた日本語アクセント型の分類ー母語話者と学習者による単語発話と朗読発話の比較ー」第 32 回日本音声学会全国大会予稿集, pp.48-53.
- 波多野博頭・石井カルロス寿憲 (2017). 「日本語自然対話に現れる質問発話の句末音調」『音声研究』21(1), pp.1-11.
- 波多野博頭・王可心・陳凱僑・林良子 (2022). 「日本語母語話者および日本語学習者による疑問・非疑問発話の韻律特徴：平均ピッチ曲線を用いた比較」日本音響学会 2022 年春季研究発表会講演論文集, pp.799-800.
- 林良子・磯村一弘・阿部新・上山素子・金村久美・柴田智子・中川千恵子・峯松信明・Aaron Lee Albin・波多野博頭・吉田夏也・松田真希子 (2018). 「日本語韻律学習のための音声アーカイブ構築」日本語教育国際研究大会 (ICJLE) 発表ポスター.
(https://www.researchgate.net/publication/326866490_ribenyuyunluxuexinotamenoyinshengakai_bugouzhu_Development_of_Sound_Archives_for_Learning_Japanese_Prosody よりダウンロード可能)
- 前川喜久雄 (2009). 「日本語学習者音声研究の課題」『日本語教育』142, pp.4-13.

関連 URL

- 大語彙連続音声認識エンジン Julius <https://julius.osdn.jp/>
- Praat: doing phonetics by computer [Computer program] <https://www.praat.org/>

日本語学習者による用例を通時的観点で分析可能とする

コーパス開発の意義とその可能性

上出 大河 (國學院大學大学院文学研究科) †

The Significance and Potential of Developing a Corpus to enable Diachronic

Analysis of Usage by Learners of Japanese

Taiga Kamide (Graduate School of Literature, Kokugakuin University)

要旨

日本語非母語話者(日本語学習者)による用例を収めたコーパスは現在「多言語母語の日本語学習者横断コーパス」(International Corpus of Japanese as a Second Language: I-JAS)等数種が存在している。それらはいずれも共時的な誤用分析、中間言語分析を行う際には有効である一方、通時的な観点から分析を加えようとする際にはその採録対象の時間的な偏りのためにある種の困難が付きまとうこととなる。そこで本発表では、このような問題を発展的に解消する手がかりとしてのコーパス開発の意義、可能性について少しく記述することとしたい。

1. はじめに

日本語母語話者の用例を採録対象としたコーパスと日本語非母語話者(日本語学習者)の用例を採録対象としたコーパスとの間に見られる最大の質的差異のひとつは、通時的な分析可能性の有無である。

日本語母語話者の用例を採録対象としたコーパスは「現代日本語書き言葉均衡コーパス」(Balanced Corpus of Contemporary Written Japanese: BCCWJ)のように用例の共時的検討を可能とするコーパスや「日本語歴史コーパス」(The Corpus of Historical Japanese: CHJ)のように用例の通時的検討を可能とするコーパス等が整備されている。しかし一方で、日本語非母語話者(日本語学習者)の用例を採録対象としたコーパスは、「多言語母語の日本語学習者横断コーパス」(International Corpus of Japanese as a Second Language: I-JAS)をはじめ「中国語・韓国語母語の日本語学習者縦断発話コーパス」(Corpus of Japanese as a Second Language :C-JAS)、「学習者作文コーパス なたね」等総じて通時的な検討を加える上では多大な制約を受ける性質のものが多数である。むしろこれらのコーパスは共時的な誤用分析、中間言語分析を行う上では有効であり、日本語非母語話者(日本語学習者)の犯しやすい言語的誤り等を考察しうるという点で高く評価しうるであろう。他方、日本語教育がその歴史的な営みを内省し今後の教授法を模索していくという上では、明治期に遡る日本語教育の萌芽的開始点に見られた用例も含めた通時的分析も重大な示唆を与えうるものではなかろうか。日本語非母語話者(日本語学習者)による用例を通時的観点

† y 224206@kokugakuin.ac.jp

で分析可能とするコーパスの開発は、現在のコーパスの部分的な問題点を発展的に解消し、このような方向付けでの研究を進展させる上で極めて重大な意味を有するというということをまず指摘しておきたい。

2. コーパス開発の意義

以下、日本語非母語話者（日本語学習者）による用例を通時的観点で分析可能とするコーパスが日本語教育の実践や言語研究を進める上でどのような意義を持ちうるのかという点について簡略的に記す。

2.1 学習者の言語習得上の通時的な難点の理解の促進

現在利用されている日本語非母語話者（日本語学習者）による用例を採録対象としたコーパスの整備が現代の誤用分析、中間言語分析に大きな研究上の進展をもたらしたのは周知の通りである。一方で、現在の分析の多くが共時的な分析にとどまっているために、それらの分析が総じて通時的視点を欠いたものになっていることは重く受け止められるべき事実である。ここに通時的視点を持ち込むことが容易になれば、学習者の音声的、文法的特性が通時的な連続性を有する特徴であるのか否か等が明らかになり、より重点を置くべき指導項目を絞り込むことが可能になることが期待される。また、通時的な分析で明らかに知る学習者の音声的、文法的特性の連続性及び非連続性については、2.2 に示す教材・教具の有効性の検証や 2.3 に示す日本語史的な分析可能性にもつながりを有するという点を申し添えておきたい。

2.2 教材・教具の有効性の検証可能性

日本語教育史領域での研究では、教授法、教材・教具の歴史的変遷等が研究上の関心の中心とされることが多く、資料に基づいた実証的な議論に耐えうる論が多数発表されている。これらの研究上の成果と、2.1 で言及したような学習者の音声的、文法的特性の連続性、非連続性とを併せて考察することで、その時期に中心的に採用されていた教授法、教材・教具がどのような言語上の性質を産み出したのか、そこに関連性が見出せるのか否か等を実証的に検討することが可能となり、日本語教育史領域に更なる知見の蓄積をもたらすものと期待される。ここに指摘するまでもなく、言語教育の最重要事項は学習者の言語習得のサポートであり、学習者の言語習得の実態を無視して教授法や教材を分析することはとりもなおさず言語教育の最重要事項を無視する態度に他ならない。しかし、歴史的な用例を採集し分析することは現状では多くの労量を費やす必要があり、また、網羅性や量的な妥当性を考えれば調査の困難であることは明白であり、このような現状の解消にはコーパス開発による用例調査の安定的な容易さの担保が必要であると思われる。

2.3 日本語史的な分析可能性

日本語非母語話者（日本語学習者）による用例を通時的観点で分析可能とするコーパスには、歴史的な用例として明治期、大正期の植民地政策の実践に伴う領有地での用例を採録する必要がある。この時期の台湾や朝鮮での用例は日本語非母語話者の日本語というだけではなく国家政策的道理上は皇民の国語という性質を有している。このような特殊事情に着目するにせよ無視するにせよ特定の言語は母語話者の専有物たりえないのであって、一定の歴史をもつ非母語話者の日本語も日本語史領域の一環として論じられる資格を有してい

るはずである。また、領有地にはそれぞれ内地方言の影響が見られるため、方言の展開の一側面としての分析も用例に研究用情報を付加すること等で進展していくものと期待したい。

その他にも日本語非母語話者(日本語学習者)による用例を通時的観点で分析可能とするコーパスはその史的立場づけのために歴史学や教科教育史領域等でも援用が可能となる可能性があり、その価値は極めて高いものと思われる。

3. 採録対象について

言語資料としてどのようなものを採録対象とするかというのは重大な問題である。現在利用されている日本語非母語話者(日本語学習者)による用例を採録対象としたコーパスとはその制約という面で隔たりが否定できない。最大の難点は録音資料の不足である。当時の内地人による音声的な報告の例を網羅的に採録することでこの問題を部分的に補うことができるかどうかは議論の余地があろう。

このような制約を考えると、当時の外地系児童の作文資料が採録対象としては大きな意味を持つ。作文資料としては以下のようなものが挙げられる。

<作文資料>

- ・新義州高等普通学校作文集『大正十二年伝説集』(1923年)
- ・『御大礼記念児童文集』(1929年)
- ・『全国小学児童綴方展覧会』(1936年)全6巻 等。

他にもガリ版刷りの作文集等を含めると、コーパスとして機能しうる最低限度の言語量は確保できるものとみてよいのではなかろうか。

作文資料も他者の手による規範化がなされている可能性があり、当時の外地人の言語状況が正確に反映されているのかという点で疑問が残る部分が存在する。この点については資料の質的な妥当性について検証した上で採録対象を選定する必要があるが、例えば『全国小学児童綴方展覧会』には「オカアサ」「オカアサン」、「イツツテ」「イッテ」というような誤記が分布の上で外地系児童に偏っており、その他方言要素の表出等を踏まえると、ここに見られる作文は過度に規範化されたものではないことが期待されるため、一定程度は言語状況を反映していたものと見られる。

音声に関しては、先述の事情で内地人の報告の観察に依存せざるを得ないが、これに関しては領有初期から比較的豊富な蓄積が存在し、量的な問題はクリアできそうである。

<内地人による音声上の報告に関する資料>

- ・山口喜一郎(1903) 「国語教授の際に気付きし児童発音の誤りに就きて」(『台湾教育学会雑誌』十七号)
- ・官立漢城外国語学校(1911) 『国語の発音及語法に関する調査』
- ・寺川喜四男(1939) 『北部台湾に於て福建系本島人の使用する国語のアクセント研究』 早稲田大学言語学会
- ・————— (1942) 『台湾に於ける国語音韻論』 台湾学芸社
- ・————— (1945) 『大東亜諸言語と日本語』 大雅堂 等。

その他、雑誌「日本語」、「音声学協会会報」等にも広く報告が見られる。

このような内地人の観察にどの程度の妥当性を認めるのかという問題は、コーパス採録の対象となりうるか否かという問題と地続きであり、多面的な角度からの検討を行う必要がある。

4. まとめ及び＜日本語非母語話者の日本語史＞研究について

以上、日本語非母語話者（日本語学習者）による用例を通時的観点で分析可能とするコーパスが整備されることで容易になると思われる研究の重要性やコーパス開発の可能性について簡略的に述べてきた。このようなコーパスを用いた分析の成果は、それを日本語教育学領域に位置づけるか日本語学領域に位置づけるかはさておき、いわば＜日本語非母語話者の日本語史＞研究とでも呼称すべきものであろう。この方向付けの研究の成果は純粋な言語学的分析の対象としても興味深く、また日本語教育がいつかきた道を繰り返さぬための道標としても重大な示唆を与えるものである。

現代の研究状況を眺めれば、母語話者の日本語のように非母語話者の日本語が通時的な分析を加えられることは僅少であり、日本語教育史領域ですら用例の通時的分析は充分行われていないと言わざるを得ない。少々事々しい例になるが、ソシュールは「記号学」の方向性を表明する際に、「それはまだ存在していないのであるから、どんなものになるかはわからない。しかし、それは存在すべき権利を有し、その位置はあらかじめ決定されている。¹⁾」と述べた。「記号学」の誕生ほどに重々しく華々しい事柄と位置付けるわけではないが、その重要性を踏まえれば、＜日本語非母語話者の日本語史＞研究も存在する権利を有しており、その地位はあらかじめ決定している、と主張しておきたい。

また、より多くの人々と問題意識を共有し、協力を得ながらそのような研究の進展をもたらさうとするコーパスの構築を目指すことを発表者の今後の課題とする。

¹⁾ Ferdinand de Saussure, *Cours de linguistique Générale*, 1916（小林英夫訳『一般言語学講義』1972年、岩波書店、p 29）

「急性」を含む病名の語構成

相良かおる(西南女学院大学)
高崎智子 (西南女学院大学、医師)
東条佳奈 (大阪大学)
西嶋佑太郎(医師)
山崎誠 (国立国語研究所)

Word structure of disease names containing "acute"

Kaoru Sagara (Seinan Jo Gakuin University)

Satoko Takasaki (Seinan Jo Gakuin University, Physician)

Kana Tojo (Osaka University)

Yutaro Nishijima (Physician)

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

「急性骨髄性白血病」のように「～性」が複数含まれ、かつ「急性」を含む病名において、「急性」の緊急度と語順の関係を調べるために、『現代日本語書き言葉均衡コーパス』(BCCWJ)に出現した当該の病名 28 語について、BCCWJ、医師経過記録、多職種共有の経過記録での出現頻度を求めた。

その結果、①「急性」は語頭に多く出現すること、②医療現場では使われない病名があること、③医療現場では「急性」無しの病名が多く使われていること、④「急性」とその他の「～性」の語順を変えた同義語が存在することが分かった。

1. はじめに

医療記録文の分かち書きを支援するために相良が 2008 年より無償で公開した ComeJisyoV1 (見出し語約 3 万語) は、随時更新を続け 2021 年 3 月公開の ComeJisyoUtf8-3 の見出し語数は 118,404 語となった。当初、医療記録に含まれる医療用語 (以下、「実践医療用語」という) の語単位の認定は困難だったことから、ComeJisyo には合成語が多く含まれている。

実践医療用語約 11 万語を見出し語とする ComeJisyo は、代表性は確保されていないものの、実践医療用語の言語資源と捉えることができることから、2018 年より日本語学、看護学、医学を専門とする研究者の協力を得て 7,192 語の合成語の語構成の解析と意味分類に着手した。そして、その成果物である「実践医療用語_語構成要素語彙試案表」を 2021 年 3 月に言語資源協会¹より公開している。

この共同研究において我々は、(1)一般的な日本語に比べて「～性」が複数つくものが多いこと、(2)「～性」を含む病名の中には、「接触皮膚炎」と「接触性皮膚炎」のように接尾辞「性」の無い同義語があることに気が付いた。

そこで(1)について、ICD10 (国際疾病分類第 10 版) に対応して開発された MEDIS-DC 標準病名マスター (以下、「標準病名マスター」という)²に登録されている病名 26,420 語を調べたところ、「～性」を含む病名は 8,172 語で 30.9%を占め、これらのうち、「～性」を

¹ <https://www.gsk.or.jp/>

² ICD10 対応標準病名マスターV5.04 2020 年 6 月 1 日改訂版 (byomei504)

2回含む病名は1,034語、3語含む病名は94語で、合わせると全体の3.93%を占めていることが分かった。

一方中納言 2.6.0 を用いて『現代日本語書き言葉均衡コーパス』（以下、「BCCWJ」という）の文字列検索を実施したところ、「男性」「女性」「性格」などの「性」を含めて「性」を二つ以上含む原文は21,478件であった³。その原文文字列の中には「一元性・階層性」のように「・」で複数の語が併記されているものがあるため、これらを削除したところ20,824語となり、これらの全検索対象語に占める割合は0.2%であった。そして20,824件のうち、「性」を一つ含む原文文字列は20,330語（97.63%）を占め、「性」を三つ含む5語は、「先天性腎性尿崩性」「良性発作性頭位変換性めまい」「慢性特異性肉芽性炎症」「特発性慢性仮性腸閉塞症」「薬剤誘発性慢性連日性頭痛」と全て病名に関する語であることから、(1)の一般的な日本語に比べて「～性」が複数つくものが多いことが示唆された。

現在我々は、複数の「～性」を持つ病名において、「～性」の語順や意味について何らかの関係性があるのではないかと考え、「～性」を複数持つ合成語の調査も行なっている。

本稿では、医療従事者ではない相良が、複数の「～性」を持つ病名の解析の中で、「症状が急に現れ、その進行が速いこと」を意味する「急性」とその他の「～性」の語順に意味的な関係があるのか否かに興味を持ったことが発端となり、調査したことについて報告する。

具体的には、BCCWJから「～性」を二つ以上含み、そのうちの一つが「急性」である病名28語を抽出し、ComeJisyoUtf8-3、医師経過記録、多職種共有の経過記録における出現頻度と語順について調べた結果について述べる。

2. 対象とする病名

医療従事者ではない者にも理解しやすい病名を対象とするために、中納言（2.6.0、データバージョン 2021.03）を用い、BCCWJの文字列検索により「～性」を複数含み、そのうちの一つが「急性」である病名28語（以下、「対象病名」という）を抽出した。また、これらに実践医療用語がどの程度含まれるのかの目安として、標準病名マスターおよびComeJisyoUtf8-3の見出し語と照合した。表1に対象病名の概要を示す。

BCCWJに出現した28語の病名の延べ語数は32語であり、そのうち22語（68.75%）のジャンルは自然科学であり、使用言語域（レジスター）では、25語（78.13%）が書籍であった。標準病名マスターとの照合で一致した病名は9語、ComeJisyoUtf8-3の見出し語との照合で一致したものは11語あった。

表1 対象病名28語とジャンル

No.	病名	ジャンル	言語使用域
1	急性アルコール性ショック症	4 自然科学	出版・書籍
2	急性リンパ球性白血病	エンターテインメントと趣味	特定目的・知恵袋
3	§* 急性リンパ性白血病	4 自然科学	図書館・書籍
4	急性化膿性炎症	4 自然科学	出版・書籍
5	§* 急性化膿性甲状腺炎	4 自然科学	出版・書籍
6	急性化膿性骨膜炎	4 自然科学	出版・書籍
7	急性化膿性疾患	健康、美容とファッション	特定目的・知恵袋

³ 【検索条件式】 キー: 全文検索 @@ "%性%性"

WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND keyDisplay="0" AND resultUnitWord="long" AND targetString="1" AND tglWords="20" AND unit="3" AND encoding="UTF-16LE" AND endOfLine="CRLF"

8	§ * 急性化膿性中耳炎	4 自然科学	出版・書籍
9	急性解離性大動脈瘤	4 自然科学	出版・書籍
10	急性壊死性歯肉炎	4 自然科学	出版・書籍
11	急性壊死性膵炎	4 自然科学	出版・書籍
12	§ * 急性間欠性ポルフィリン症	4 自然科学 5 技術・工学	出版・書籍 出版・書籍
13	急性間質性線維症	4 自然科学	出版・書籍
14	急性機能性便秘	4 自然科学	出版・書籍
15	§ * 急性骨髄性白血病	3 社会科学	出版・書籍
		4 自然科学	図書館・書籍
		4 自然科学	出版・書籍
		9 文学 9 文学 総合/一般/婦人誌	図書館・書籍 特定目的・ベストセラー 出版・雑誌
16	急性細菌性結膜炎	子育てと学校	特定目的・知恵袋
17	§ * 急性散在性脳脊髄炎	近畿地方	特定目的・広報紙
18	* 急性出血性胃炎	1 哲学	特定目的・ベストセラー
19	§ * 急性出血性結膜炎	4 自然科学	出版・書籍
20	急性腺窩性扁桃腺炎	4 自然科学	出版・書籍
21	§ * 急性前骨髄球性白血病	4 自然科学	図書館・書籍
22	* 急性前骨髄性白血病	4 自然科学	出版・書籍
23	急性熱性疾患	4 自然科学	出版・書籍
24	急性熱性病	4 自然科学	図書館・書籍
25	§ * 急性汎発性腹膜炎	9 文学	図書館・書籍
26	成人急性非リンパ性白血病	4 自然科学	出版・書籍
27	非細菌性急性胃腸炎	4 自然科学	出版・書籍
28	非乏尿性急性腎不全	4 自然科学	出版・書籍

注) § : 標準病名マスターに登録 * : ComeJisyoUtf8-3 の見出し語と一致

3. 調査方法

以下の手順で調査を行なった。

手順 1. 匿名加工済みの①医師経過記録と②多職種共有の経過記録、各 1 施設 1 年分のデータにおける対象病名の出現頻度を求める。

① 施設規模：1,267 床

データ量：3,255,371 行 596MB

② 施設規模：520 床

データ量：3,361,282 行 135MB

手順 2. 対象病名の「急性」を対義語の「慢性」で置き換えた語の出現頻度を求める。

手順 3. 対象病名から「急性」を除いた語の出現頻度を求める。

手順 4. 語構成および各語構成要素の意味分類を行う。

4. 結果

4.1 経過記録における出現頻度

表 2 は、匿名加工済みの医師経過記録と多職種共有の経過記録における対象病名の出現頻度をまとめたものである。

対象病名 28 語のうち、経過記録に出現した病名は 11 語あり、これらのうち標準病名マスターまたは ComeJisyoUtf8-3 に登録の病名は 9 語であった。経過記録に使われ、標準病名マスターまたは ComeJisyoUtf8-3 に登録されていない病名は、No.2「急性リンパ球性白血病」と No.23「急性熱性疾患」の 2 語であった。

なお、今回出現頻度を求めた経過記録データは、2 施設 2 種類であり、経過記録を代表するデータではない。

表 2 経過記録における出現頻度

No.	病名	BCCWJ	医師経過記録	多職種共有経過記録
1	急性アルコール性ショック症	1		
2	急性リンパ球性白血病	1	60	
3	§* 急性リンパ性白血病	2	508	19
4	急性化膿性炎症	1		
5	§* 急性化膿性甲状腺炎	1		1
6	急性化膿性骨膜炎	1		
7	急性化膿性疾患	1		
8	§* 急性化膿性中耳炎	2	6	
9	急性解離性大動脈瘤	1		
10	急性壊死性歯肉炎	1		
11	急性壊死性腭炎	1		
12	§* 急性間欠性ポルフィリン症	5	2	
13	急性間質性線維症	1		
14	急性機能的便秘	1		
15	§* 急性骨髄性白血病	12	1,326	18
16	急性細菌性結膜炎	1		
17	§* 急性散在性脳脊髄炎	1	12	
18	* 急性出血性胃炎	1		
19	§* 急性出血性結膜炎	3		
20	急性腺窩性扁桃腺炎	1		
21	§* 急性前骨髄球性白血病	1	58	5
22	* 急性前骨髄性白血病	1	27	4
23	急性熱性疾患	2		1
24	急性熱性病	2		
25	§* 急性汎発性腹膜炎	1	64	5

26	成人急性非リンパ性白血病	1
27	非細菌性急性胃腸炎	1
28	非乏尿性急性腎不全	1

4.2 「慢性」に置き換えた対象病名の出現頻度

表3は、対象病名の「急性」を対義語である「慢性」に置き換えた場合の出現頻度をまとめたものである。経過記録に出現した対象病名11語の対語と思われる病名はNo.2「慢性リンパ球性白血病」、No.8「慢性化膿性中耳炎」、No.15「慢性骨髄性白血病」の3語であり、No.9「慢性解離性大動脈瘤」は、表2のように「急性」のつく病名は経過記録には出現しなかったが、「慢性」に置き換えた病名のみが経過記録に出現した。

表3 「慢性」に置き換えた対象病名の出現頻度

No.	病名	医師経過記録	多職種共有経過記録
2	慢性リンパ球性白血病	42	
8 §*	慢性化膿性中耳炎	66	1
9	慢性解離性大動脈瘤	1	
15 §*	慢性骨髄性白血病	169	25

4.3 「急性」を除いた対象病名の出現頻度

表4は、対象病名のうち、「急性」を除いた病名（「～性」+「病名」）の出現頻度をまとめたものである。当該病名28語中20語に出現が認められ、そのうち、No.4「化膿性炎症」、No.7「化膿性疾患」、No.10「壊死性歯肉炎」、No.11「壊死性腭炎」、No.14「機能性便秘」、No.16「細菌性結膜炎」、No.18「出血性胃炎」、No.19「出血性結膜炎」の8語は、「急性」および「慢性」が付与された病名では出現がなかった。

標準病名マスターおよびComeJisyoUtf8-3の見出し語と照合したところ、「急性」のつく病名では登録されていなかったNo.9「解離性大動脈瘤」、No.11「壊死性腭炎」、No.16「細菌性結膜炎」の3語は両方に登録されていた。

表4 「急性」を除いた対象病名の出現頻度

No.	「～性」+「病名」	医師経過記録		多職種共有経過記録			
		全頻度	急性～	慢性～	全頻度	急性～	慢性～
1	アルコール性ショック症						
2	リンパ球性白血病	127	60	42	2		
3 §*†	リンパ性白血病	563	508		23	19	
4	化膿性炎症	5					
5 §*†	化膿性甲状腺炎				6	1	
6	化膿性骨膜炎						
7	化膿性疾患	1			1		
8 §*†	化膿性中耳炎	78	6	66			1
9 †	解離性大動脈瘤	201		1	42		
10	壊死性歯肉炎	2					

11 †	壊死性膵炎	93			1		
12 §*	間欠性ポルフィリン症	2	2				
13	間質性線維症						
14 †	機能性便秘	5					
15 §*†	骨髄性白血病	1,526	1,326	169	47	18	25
16 †	細菌性結膜炎	21			12		
17 §*	散在性脳脊髄炎	12	12				
18 *†	出血性胃炎	80			6		
19 §*	出血性結膜炎	1					
20	腺窩性扁桃腺炎						
21 §*	前骨髄球性白血病	58	58		5	5	
22 *	前骨髄性白血病	27	27		4	4	
23	熱性疾患	28			2	1	
24	熱性病						
25 §*†	汎発性腹膜炎	342	64		56	5	
26	成人非リンパ性白血病						
27	非細菌性胃腸炎						
28	非乏尿性腎不全						

注) “§”と“*”は、表1を再掲。“†”は、標準病名マスターに「～性+病名」が登録されていることを、“‡”は ComeJisyoUtf8-3 の見出し語にあることを示す。

4.4 対象病名の語構成と語構成要素の意味

表5は、共同研究者の医師2名により、対象病名を語構成要素に分割し、意味ラベルを付与したものである。

対象病名の語構成は、①「急性」+「～性」+「病名」と、②「～性」+「急性」+「病名」の2種類であり、①の対象病名は25語で、意味的な語順は、「経過/状態・原因/病名」となっており、②の経過が語中にある対象病名は、No.26「成人急性非リンパ性白血病」、No.27「非細菌性急性胃腸炎」、No.28「非乏尿性急性腎不全」の3語である。

No.26「成人急性非リンパ性白血病」は、「急性非リンパ性白血病 (ANLL: acute non-lymphocytic leukemia)」の前に「成人」がついたものと考えられる。今回、経過記録での出現はなかったため(表2)、CiNiiで検索したところ検索結果は44件あった。同様にCiNiiでNo.27「非細菌性急性胃腸炎」を検索した結果は6件、語順を換えた「急性非細菌性胃腸炎」では2件であり、どちらの語も使われていることが分かった。No.28「非乏尿性急性腎不全」の検索結果は11件で、英語表記は、“Non-Oliguric Acute Renal Failure”である。この英語表記をGoogle scholarで検索すると、検索結果は774件であった。また、急性腎不全の英語表記“Acute Renal Failure”はARFと略されることから、「急性」と「腎不全」の語間の結合が強いと考えられる。なお、語順を換えた「急性非乏尿性腎不全」の検索結果はCiNiiでは3件、Google scholarでは5件あったが、共にそのうち3件は著者が同じであった。

表 5 対象病名の語構成と語構成要素の意味

No.	病名	語構成と意味
1	急性アルコール性ショック症	急性 (経過) / アルコール性 (原因) / (病名)
2	急性リンパ球性白血病	急性 (経過) / リンパ球性 (状態) / (病名)
3	§ * 急性リンパ性白血病	急性 (経過) / リンパ性 (状態) / (病名)
4	急性化膿性炎症	急性 (経過) / 化膿性 (状態) / (病名)
5	§ * 急性化膿性甲状腺炎	急性 (経過) / 化膿性 (状態) / (病名)
6	急性化膿性骨膜炎	急性 (経過) / 化膿性 (状態) / (病名)
7	急性化膿性疾患	急性 (経過) / 化膿性 (状態) / (病名)
8	§ * 急性化膿性中耳炎	急性 (経過) / 化膿性 (状態) / (病名)
9	急性解離性大動脈瘤	急性 (経過) / 解離性 (状態) / (病名)
10	急性壊死性歯肉炎	急性 (経過) / 壊死性 (状態) / (病名)
11	急性壊死性腭炎	急性 (経過) / 壊死性 (状態) / (病名)
12	§ * 急性間欠性ポルフィリン症	急性 (経過) / 間欠性 (状態) / (病名)
13	急性間質性線維症	急性 (経過) / 間質性 (状態) / (病名)
14	急性機能的便秘	急性 (経過) / 機能的 (状態) / (病名)
15	§ * 急性骨髄性白血病	急性 (経過) / 骨髄性 (状態) / (病名)
16	急性細菌性結膜炎	急性 (経過) / 細菌性 (原因) / (病名)
17	§ * 急性散在性脳脊髄炎	急性 (経過) / 散在性 (状態) / (病名)
18	* 急性出血性胃炎	急性 (経過) / 出血性 (状態) / (病名)
19	§ * 急性出血性結膜炎	急性 (経過) / 出血性 (状態) / (病名)
20	急性腺窩性扁桃腺炎	急性 (経過) / 腺窩性 (状態) / (病名)
21	§ * 急性前骨髄球性白血病	急性 (経過) / 前骨髄球性 (状態) / (病名)
22	* 急性前骨髄性白血病	急性 (経過) / 前骨髄性 (状態) / (病名)
23	急性熱性疾患	急性 (経過) / 熱性 (状態) / (病名)
24	急性熱性病	急性 (経過) / 熱性病 (病名)
25	§ * 急性汎発性腹膜炎	急性 (経過) / 汎発性 (状態) / (病名)
26	成人急性非リンパ性白血病	成人 (属性) / 急性 (経過) / 非リンパ性 (状態) / (病名)
27	非細菌性急性胃腸炎	非細菌性 (原因) / 急性 (経過) / (病名)
28	非乏尿性急性腎不全	非乏尿性 (状態) / 急性 (経過) / (病名)

5. 考察

対象病名 28 語のうち、経過記録に出現するものは 11 語と半数以下であった (表 2)。今回、経過記録を参照したのが 2 施設のみであるという制約による可能性もあるものの、小説、雑誌記事などにのみ出現する対象病名もあり、対象病名には学術的なものと、医療施設で使われる実践的なもの以外の病名があることが示唆された (図 1)。

「急性」の対義語として「慢性」があるが、「急性」が付与される主要病名には、必ず「慢性」が付与される訳ではないことが分かる。

表 4 において、「急性～」と「慢性～」の出現頻度を足したものが全頻度と同程度である

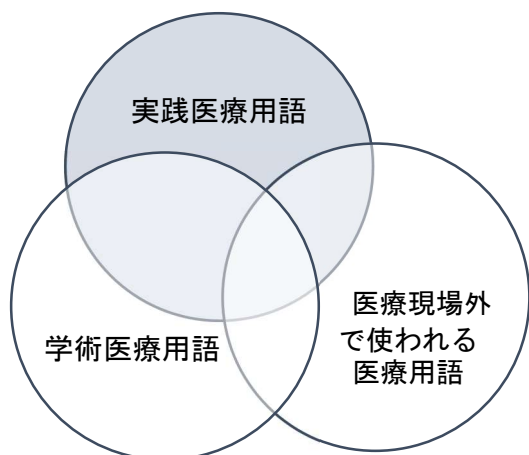


図 1 医療用語の使用域

主要病名 No.2「リンパ球性白血病」、No.8「化膿性中耳炎」、No.15「骨髄性白血病」において、「急性」と「慢性」は、意味のある語（有標的）であると考えられる。しかし、この場合の「急性」と「慢性」は、表面上は症状の進行の速さを表すものであるが、「急性」か「慢性」かによって病気の原因や機序の違いがあるため、単に症状の進行を表す以上に、病気の種類を区別するための重要な標識となっている。

一方、No.11「壊死性腭炎」とNo.16「細菌性結膜炎」の経過記録における出現頻度は、主要病名のみで「急性」および「慢性」が付与された対象病名では出現しない。このことから、これらの病名における「急性」は、重要な意味を持たないと推測される。

また、No.12「急性間欠性ポルフィリン症」とNo.17「急性散在性脳脊髄炎」は、経過記録中に「ポルフィリン症」や「脳脊髄炎」など「～性」を含まない病名での出現はなく、「急性」がつく対象病名でのみ出現した。『日本医学会医学用語辞典』WEB版を見ると、「ポルフィリン症」を含む用語には、「晩発性皮膚ポルフィリン症」「赤血球産生性プロトポルフィリン症」などがあるが、「急性」または「間欠性」いずれかを要素として持つのは「急性間欠性ポルフィリン症」のみである。また、「脳脊髄炎」を含む用語には、「壊死性脳脊髄炎」や「筋痛性脳脊髄炎」などがあるが、「急性」または「散在性」いずれかを要素として持つのは「急性散在性脳脊髄炎」のみである。「急性間欠性ポルフィリン症（acute intermittent porphyria）」「急性散在性脳脊髄炎（acute disseminated encephalomyelitis）」両病名ともに英語表記の逐語訳となっている。このことから、「急性間欠性」および「急性散在性」で病気の性質を表現する1つの要素であると考えられる。

6. おわりに

医療の知識を持たない筆者は、(1)「急性」の付く病名は、緊急度を表す重要な修飾語である、(2)病名において重要な意味を持つ修飾語の語順には何等かの規則がある、と考え、今回調査を行なった。

以下は「急性」の語義を手元にある国語辞典で調べた結果である。

- 急に症状を発して病気の進み方が速いこと [広辞苑 第七版]
- 症状が急に現れ、その進行が速いこと [明鏡国語辞典 第二版]
- 発病が急で病状の進行が早いこと [デジタル版 大辞泉]
- 急に起こって、症状の激しい病気 [新明解国語辞典 第七版]
- 病気が、急に症状を呈して激しく進行すること [精選版 日本国語大辞典]
- 病気が急に起こって、症状のはげしいこと [三省堂国語辞典 第七版]
- たちまち症状があらわれ、進み方の早い性質 [新選国語辞典 第九版]

医療の知識を持たない筆者は、「進行が早い」⇒「緊急度が高い」⇒「重要な語」⇒「語順に影響」と考え、今回調査を行なった。その結果、本対象病名における「急性」には、頻度調査の結果と医師らによるコメントから重要な意味を持たないものがあること、また、「急性」の語順は、英語表記の語順に影響を受けるものが少なくないことが示唆された。

合成語の語構成要素の意味分類において、非医療従事者である筆者が「状態」とするとこ

ろを、医療従事者である共同研究者は、外からみて分かることを「状態」、本人の訴えを「症状」、体の中で起きているメカニズム（機序）を「病態」というように細分化し、「症状」の進行の速さや激しさと「緊急度」には因果関係はないと認識している。

また、合成語の語構成要素の分割では、『急性出血性結膜炎』は、一語として認識しているので『急性』と『出血性』の重要度に違いはないなど、「これはひとまとまりの語として認識しているので語分割できない」との発言がある。

今回、医師を交えて分析する中で、病名は、医師の知識に付与された意味ラベルであり、「急性出血性結膜炎」から医師が得る情報は、「ウイルス性の結膜炎」という原因とその治療法であり、病名を構成する語構成要素の意味を集めたものが、医師にとって重要な情報ではないことに気付く。

機械学習の研究が進み、誤字・誤変換を含む表記の揺れのない膨大な学習データから医師の知識を含めた病名の意味を推測する精度は向上している。しかしながら現実には間違いない膨大な学習用データを集めることは困難である。我々は専門的な暗黙知を機械可読な表現で記述することで少ないデータでの推測が可能となると考えている。

謝 辞

本研究は JSPS 科研費 JP18H03499 ならびに JP21H03777 の助成を受けています。

関連 URL

国立国語研究所(編) (2022). 『現代日本語書き言葉均衡コーパス (BCCWJ)』,(データバージョン 2021.03,中納言バージョン 2.6.0) <https://chunagon.ninjal.ac.jp/> (2022年8月5日確認).
 CiNii Research <https://cir.nii.ac.jp/> (検索 : 2022年8月4日)
 ComeJisyo <https://ja.osdn.net/projects/comedic/>
 Google Scholar <https://scholar.google.com/schhp?hl=en-US&pli=1> (検索 : 2022年8月4日)
 ICD10 対応標準病名マスター <https://www2.medis.or.jp/stdcd/byomei/index.html>
 日本医学会 医学用語辞典 WEB 版 <https://jams.med.or.jp/dic/mdic.html>

『日本語話題別会話コーパス：J-TOCC 語彙表』の公開と 日本語教育むけ情報サイトにむけた指標の検討

中俣 尚己 (大阪大学 国際教育交流センター) †

麻 子軒 (関西大学 国際教育センター)

‘Vocabulary Table of Japanese Topic-Oriented Conversation Corpus’ and a Consideration of Indices for Japanese Language Teaching

NAKAMATA Naoki (Osaka University)

MA Tzuhsuan (Kansai University)

要旨

『日本語会話話題別コーパス:J-TOCC』の語彙表を公開する。表は2種類で、15ある話題間での特徴度を比較するための粗頻度ならびに LLR の表と、各話題ごとに、240名の調査協力者がそれぞれ何度その語を使用したかというデータを収めた表である。前者の表はどの話題に特徴的かという偏りを表し、後者の表は、ある話題を与えられた時に母語話者の何%がその語を使用するかという「使用者割合」を取り出せる。本プロジェクトの最終目標は日本語教育に役立つ「話題-語彙情報サイト」の構築であるが、現場に役立つ形で情報を整理するにはこの2種類の情報が必要であることを主張する。語の使用量の幅を見る指標としては tf-idf など存在するが、検討の結果、本データでは使用総頻度の影響が大きすぎるということがわかった。一方で、LLR は語の特徴語を効率よく抽出できるが、多義語など、他の話題の影響で値が低くなることもある。使用者割合はその点をカバーすることができる。

1. はじめに

筆者らは2021年3月に『日本語話題別会話コーパス：J-TOCC』(以下、「J-TOCC」)を公開した(中俣ほか2021)。J-TOCCは、日本語教育において話題を中心とした学習活動や教材開発を支援するため構築されたコーパスであり、筆頭著者のウェブサイトで公開されている(<http://nakamata.info/database/>)。J-TOCCの名称は Japanese Topic-Oriented Conversation Corpus の略で、「ジェイトック」と読む。

J-TOCCでは同じ会話協力者のペアが15の話題について、それぞれ5分間ずつ会話を行ってもらった。話題以外の条件が統制されていることが最大の特徴である。話題は表1に示した通り、身の回りの話題が11、社会に関わる話題が4である。1つの話題につき120ペア、合計10時間の会話が収められている。語数は記号を除くと延べ語数ではおよそ165万語、異なり語数ではおよそ4万語である。話題による語数の差は小さく、1話題あたりおよそ延べ語数は11万語、異なり語数は6千語程度となっている。ペアは様々な話題について話せるよう、仲の良い関係の大学生どうしに限定し、性別の組み合わせや録音地の東西でバランスをとった。

表1 J-TOCCの話題

身の回りの話題	社会にもかかわる話題
01.食べること 02.ファッション 03.旅行 04.スポーツ 05.マンガ・ゲーム 06.家事 07.学校 08.スマートフォン 09.アルバイト 10.動物 11.天気	12.夢・将来設計 13.マナー 14.住環境 15.日本の未来

† n.nakamata.ciee@osaka-u.ac.jp

また、話者がそれぞれの話題についてどのくらい詳しいかという「話題精通度」についても5段階で尋ね、付属データとして添付している。

今回は、話題ごとの語の使用実態を明らかにするために、基礎資料として『日本語話題別会話コーパス：J-TOCC 語彙表』を公開する。本発表ではその概要について述べるとともに、日本語教育への応用のため、いくつかの指標についても検討する。

2 公開データの概要と作成手順

2.1 公開データの概要

公開するデータは2種類ある。1つは、「話題別特徴語表」であり、縦に語が並び、横に15の話題が並ぶ。それぞれの語がどの話題で多く使われているかを見るための表である。語の切り方として、短単位と長単位があり、また指標として粗頻度と対数尤度比(LLR)を用意し、4枚のシートからなるxlsxファイルとして提供する。

もう1つは「話者別使用頻度表」であり、こちらも縦に語が並び、横にはJ-TOCCの240名の参加協力が並び、そして各セルには各話者がその話題で何度その語を使用したかの数字が並ぶ。ここから240人の話者のうち何人が特定の話題で語を使用したかという使用率(UR)を計算することができるほか、話者の個人差も考慮した統計モデリングを行うこともできる(中俣2023 予定)。話題ごとに表を作り、15個のcsvファイルで提供する。また、こちらも短単位と長単位の両方を用意した。

2.2 形態素解析の方法及びデータの加工

形態素解析を行うのに用いた解析器、辞書といった解析環境は下記の通りである。

短単位解析器 MeCab 0.996
 短単位解析辞書 UniDic-MeCab 2.1.2
 長単位解析器 Comainu 0.72
 長単位解析モデル CRF++-0.58

手順は、まずJ-TOCCの平文をMecabとUniDicで短単位に分割し、次にそれをComainuと長単位解析モデルCRFで長単位にまとめ上げる、という流れになるが、ComainuにはMecabを呼び出す機能が実装されているため、これら一連の処理は一度の実行で行えた。形態素解析の処理を行う前に、本文と見なさない話者記号及び丸括弧・山括弧で示されたメタタグとその内容(下記の例の下線部分)を削除した。

E-101-1M : お金かかるけど、(E-101-2M : うん) うまいんだよな。

E-101-2M : 好きだ(笑)。

W-109-2M : いや、結構<会話が重なり聞き取り不能>

語彙表を作成する際に、形態素解析を終えた文字列に対し、以下の加工を行った。

外来語はUniDicでは、「ファッション-fashion」のように語彙素の後ろに原語の綴りが併記されているが、検索の妨げとなるため、原語の綴りの部分を削除した。

また、LINEのような半角のアルファベットの文字列を短単位に分割すると、「LINE」のように1語になる場合と、「L」「I」「N」「E」のように4語になる場合がある。後者は長単位への結合処理で自動的にLINEの1語になるが、アルファベット1文字の語彙素じゃUniDicでは全角で登録されているため、結合後は全角「LINE」に変換される。一方、最初から短単位で「LINE」1語に切られた場合は半角であるため、作成した語彙表には同じ語にも関わらず、全角と半角が混在してしまう。これを解決するために、全角への一括置換を行った。

2.3 LLR の計算方法

LLR の計算には対象コーパスと参照コーパスが必要となるが、今回は1つの話題を対象コーパスとし、残り14の話題を合わせたものを参照コーパスとした。計算方法は下記の通りである(田中・近藤 2011)。

$$2(\text{alna} + \text{blnb} + \text{clnc} + \text{dlnd} - (\text{a+b})\ln(\text{a+b}) - (\text{a+c})\ln(\text{a+c}) - (\text{b+d})\ln(\text{b+d}) - (\text{c+d})\ln(\text{c+d}) + (\text{a+b+c+d})\ln(\text{a+b+c+d}))$$

a : 当該資料での当該語の度数 b : 参照資料での当該語の度数

c : 当該資料の延べ語数 - a d : 参照資料の延べ語数 - b

ln は自然対数を表す。

a または b が 0 の場合、alna または blnb を 0 として計算する。

ad-bc < 0 の場合の場合、-1 を乗じる補正を行う

3. データの分析

3.1 LLR の概要

以降、「01.食べること」をケーススタディーとして取り上げる。短単位のデータを使用する。まず、単純に LLR が高い語を取り上げると、表 2 のようになる。

表 2 「01.食べること」で LLR が高い語

語	LLR	語	LLR	語	LLR
食べる	6555	ラーメン	1411	肉	593
美味しい	2665	寿司	1106	焼き肉	575
好き	2089	食う	917	料理	487
外食	1520	【店名】	817	カレー	477

表 2 はどれも納得のいく語ではあるが、『J-TOCC』語彙表では、特徴語の基準となる LLR10.83 以上の語が「01.食べること」だけでも 750 語以上存在する。この豊かな情報は、一方で具体的な教育場面では、一体どれを選べばよいのかというネックともなりうる。そこで、以下では代表的な品詞に分けて分析を行うとともに、もう一つの指標として 240 名のうち何%が使用したかという「使用率」(UR)にも着目する。

3.2 動詞の場合

ここでは、UniDic において品詞が「動詞—一般」であるものを対象にする。表 3 は左側に LLR が高い語、右側に UR が高い語をそれぞれ 10 語ずつ並べたものである。

表 3 「01.食べること」における高 LLR 動詞と高 UR 動詞

高 LLR 語	LLR	UR	高 UR 語	UR	LLR
食べる	6555	93%	言う	94%	-44
食う	917	36%	食べる	93%	6555
飲む	260	19%	思う	80%	-57
頼む	235	18%	分かる	70%	5
太る	207	14%	出る	40%	13
作る	90	38%	違う	38%	2
焼く	78	10%	作る	38%	90
好く	45	10%	食う	36%	917
並ぶ	38	7%	入る	30%	-4
炙る	33	1%	知る	28%	-4

まず、左側を見ると LLR が高い語であっても、「炙る」のように使用者割合が 1% というような場合も存在し、これはあまり重要な語とは認定できない。また、「食う」も教育という観点からはなかなか選びにくい語であろう。予想通り、これは男女によって使用割合に大きな差が見られ、男性の使用率は 63%、女性の使用率は 9% である。LLR の値は上位 10 語ほどではないが、使用率が比較的高い語としては売る (LLR20、UR20%) や「痩せる」 (LLR19、UR14%) がある。

次に、右側の UR に注目すると「言う」のように必ずしも話題とは関係のない語も選ばれてしまう。しかしながら、「言う」「思う」は一種の機能語として、「分かる」「違う」は情報処理のマーカーとして頻繁に使用されていると考えられる。また、「入る」は「店に入る」「砂糖が入る」のような「01. 食べること」と関係するコロケーションが多くあるものの、それ以上に「部活に入る」(04. スポーツ)、「バイトが入る」(09. アルバイト) など他の話題でより多くコロケーションが使用されたために、LLR が低く、特徴語とは判定されていない。このように、LLR は参照コーパスの性質によって、数値が変動する可能性がある。同様に、UR が 11 番目に高い語として「買う」があるが、この語もファッションでの頻度が非常に高かったため、UR27% に対し、LLR は -5 であった。

3.3 名詞の場合

ここでは、UniDic において品詞が「名詞—普通名詞—一般」であるものを対象にする。表 4 は左側に LLR が高い語、右側に UR が高い語をそれぞれ 10 語ずつ並べたものである。

表 4 「01. 食べること」における高 LLR 名詞と高 Ratio 名詞

高 LLR 語	LLR	UR	高 UR 語	UR	LLR
ラーメン	1411	39%	事	82%	27
寿司	1106	32%	奴	50%	0
肉	593	31%	人	46%	-106
焼き肉	575	23%	感じ	45%	-2
カレー	477	21%	方	40%	-1
御飯	452	36%	ラーメン	39%	1411
オムライス	391	14%	御飯	36%	452
チーズ	379	13%	家	36%	17
味	364	28%	本当	35%	1
食べ物	353	28%	物	33%	50

名詞の特徴として、LLR が高い語は UR も軒並み高いことが挙げられる。これ以降も、20 位までは UR は 10% を超え、38 位までは UR が 5% を超えている。一方、右側の UR に注目すると抽象的な語が多く抽出されてしまう¹。名詞については LLR のみで十分と言えるかもしれない。

3.4 形容詞の場合

ここでは、UniDic において品詞が「形容詞—一般」「形状詞—一般」であるものを対象にする。

表 5 「01. 食べること」における高 LLR 形容詞と高 UR 形容詞

¹ ただし、「もの」は LLR が 50 と高い。「もの」が食の特徴語であることは別コーパスを用いた研究(中俣 2015)からも示唆されている。また、副詞可能な名詞含めると、「時」「後」「最近」などの語が使用率が高い。

高 LLR 語	LLR	UR	高 UR 語	UR	LLR
美味しい	2665	78%	好き	90%	2089
好き	2089	90%	美味しい	78%	2665
旨い	274	29%	凄い	48%	-10
甘い	251	17%	多い	43%	29
大好き	150	16%	確か	42%	-2
辛い	147	15%	まじ	30%	2
脂っこい	65	3%	旨い	29%	274
安い	42	20%	やばい	28%	2
しょっぱい	41	2%	そんな	27%	1
濃い	21	5%	嫌	27%	-32

状況としては動詞に似ており、LLR が高くても UR が低い語が見られる。反対に UR が高いが LLR が低い語には「凄い」「まじ」のような程度表現や「やばい」「多い」「嫌」のように「01.食べること」に関する会話でよく使用される語が含まれており、使用率を LLR と併用することで、より効果的に語彙を選定できるようになると考えられる。

4. 指標の検討

ここまで述べてきたようなこと、例えば「炙る」は食の特徴語と言えるが重要度は低いということは経験のある教師ならば直観でわかることである。しかしながら、筆者らは話題を選ぶと必要な語彙を提供したり、語を選ぶと相性の良い話題を提供する「話題—語彙情報サイト」の構築を進めており、そのためには上記のような直観も何らかの形で指標化する必要がある。ある話題に特徴的かという指標としては LLR が頑健で一貫性のある指標であることがわかっている (内山ほか 2004)。一方で、重要度の指標としては今回用いた使用率(UR)のほか、使用頻度(frequency)や、その2つを組み合わせた tf-idf などの指標も用いられている。そこで、以上の4指標についての相関係数の計算を行った。記号を除いた全語、動詞、名詞、形容詞についてそれぞれ計算を行った結果が表6である。

表6 指標間の相関係数

全語	UR	freq.	tf-idf	LLR	動詞	UR	freq.	tf-idf	LLR
UR	1				UR	1			
freq.	0.79	1			freq.	0.85	1		
tf-idf	0.82	1.00	1		tf-idf	0.89	0.99	1	
LLR	<u>0.21</u>	0.18	0.20	1	LLR	<u>0.48</u>	0.81	0.79	1
名詞	UR	freq.	tf-idf	LLR	形容詞	UR	freq.	tf-idf	LLR
UR	1				UR	1			
freq.	0.96	1			freq.	0.91	1		
tf-idf	0.95	0.98	1		tf-idf	0.95	0.99	1	
LLR	<u>0.48</u>	0.61	0.67	1	LLR	<u>0.74</u>	0.90	0.87	1

まず、UR、frequency、tf-idfの3指標はどの品詞でも非常に相関が高い。このうち、教育目的としては使用率が有効な指標と言える。頻度では258回という数字の意味は単独ではわからないが、使用率は「その話題においては、母語話者の80%が使用する」というように単独で意味を解釈できる指標だからである。また、森(2017、2019)におけるコーパスの単位を文書(人)にすべきであるという主張からも正当性があると言える。

次に、LLRとURの相関に着目すると、全語では0.21で低く、動詞と名詞では0.48と

中程度の相関を示し、形容詞では 0.74 と高くなった。特に形容詞の相関がこれだけ強くなったのは不明である。LLR は他の 3 つとは異なるタイプの指標であるとは言える。

また、3. における高 LLR 語と高 UR 語の比較から、指標の選定については以下のように言えそうである。

- 名詞の選定には LLR のみで十分である。
- 動詞・形容詞の選定には LLR に加えて UR を加味し、UR が低い語の重要度を下げ、UR が高い語を加えるべきである。

上述の提言は他の話題のデータにも成り立つであろうか。「06.家事」と「15.日本の未来」について調査を行ったところ、同様の傾向を示すことが確認された。

5. おわりに

本研究では『日本語話題別会話コーパス：J-TOCC 語彙表』を作成・公開した。語彙表は特徴度を示す LLR の表と、重要度を示す「話者別使用頻度表」からなる。名詞においては特徴度が高い語はたいてい重要度が高かったが、動詞や形容詞においては両者は必ずしも一致しないことがわかった。

謝 辞

本研究は JSPS 科研費 18H00676、22H00668 の助成を受けた。

文 献

- 内山将夫・中條清美・山本英子・井佐原均(2004)「英語教育のための分野特徴単語の選定尺度の比較」『自然言語処理』11:3, 165-197.
- 田中牧郎・近藤明日子(2011)「教科書コーパス語彙表」『言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用』pp.55-63, 2011 文部科学省科学研究費特定領域研究「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」言語政策班。
- 中俣尚己(2015)「「日中 Skype 会話コーパス」を用いた話題別語彙の抽出—「食」の場合—」『第8回コーパス日本語学ワークショップ予稿集』11-18.
<http://doi.org/10.15084/00003427>
- 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣(2021)「「日本語話題別会話コーパス：J-TOCC」」『計量国語学』33:1, 205-213.
- 中俣尚己(2023 予定)「間投助詞「さ」の使用に話題が与える影響」中俣尚己(編)『話題別コーパスが拓く日本語教育と日本語学』ひつじ書房。
- 森秀明(2017)「コーパス間における単語使用率の比較—観察単位(ケース)は単語か文書か—」『計量国語学』31:3, 205-221.
- 森秀明(2019)『コーパスの計量的分析法再考』東北大学博士論文。

関連 URL

データベース | 中俣尚己ウェブサイト <http://nakamata.info/database/>

科学技術論文における「問題」の周辺文からの問題内容の抽出

平林 照雄 (東京農工大学 生物システム応用科学府) *

古宮 嘉那子 (東京農工大学)

浅原 正幸 (国立国語研究所)

Extracting Problem Contents from Surrounding Texts of a Word “Problem” in Science and Technical Papers

Teruo Hirabayashi (Tokyo University of Agriculture and Technology)

Komiya Kanako (Tokyo University of Agriculture and Technology)

Masayuki Asahara (National Institute for Japanese Language and Linguistics)

要旨

科学論文では、しばしば「問題内容」と「解決法」が主題となる。我々は、日本語科学論文において「問題内容」と「解決法」のペアを効率的に取得するシステムの作成を目指している。Heffernan ら (2018) は、“problem” 及びその類義語に注目し、パターン抽出を用いて英語科学論文における「問題内容」提起箇所の抽出を行った。しかし、日本語は英語とは異なり、決まったパターンがないことから、単純なパターン抽出では「問題内容」を抽出することができない。そのため、本論文では、「問題」という語を含む文とそれらの前後文を少量用意し、それらに「問題内容」が含まれているかどうかのアノテーションを行う。また、それらのアノテーションデータを用いて機械学習を行い、大量のデータに適用することで、自己学習により「問題内容」の有無のタグを付与し、付与されたタグの精度を調査した。

1. はじめに

増え続ける科学技術論文の全てを読解することは非常に困難であるから、その読解を助ける目的で、自動知識抽出システムの研究は盛んである。我々は、主題を自動で抽出するシステムについて研究している。学術論文において、主題は問題解決プロセスであることが多い。一般に、問題を提起し、その解決法として手法を提案し、問題の解決を目指すという流れをとる。そこで、我々はこれらの問題内容と解決法のペアの知識を効率的に取得することで、主題読解を機械的に行うシステムを目指している。以降、取得すべき問題内容および解決法を記述した部分をそれぞれ「問題内容」と「解決法」と表記する。

Heffernan and Teufel (2018) は英語科学論文からパターン抽出を利用して“problem” に注目した「問題内容」と「解決法」のペアの情報抽出を行った。日本語科学論文でも同様に、「問題内容」を明示するため、「問題」（以降、「問題」という語を「問題」と表記する）をはじめとする特徴語が、同一文または、周辺文に存在すると予想される。しかし、日本語においては、「問題は～である。」と明記されていることは稀であり、パターン抽出による情報抽出手法を用い

* s213645z@st.go.tuat.ac.jp

ることが難しい。そこで本論文では、「問題」に注目し、「問題」が含まれる文とその前後文に「問題内容」が含まれるか否かを判定する分類器を作成し、作成した分類器で自己学習することで精度の向上を目指す。具体的には、はじめに、日本語科学技術論文中の「問題」を含む文及び、その前後一文を対象に、文に「問題内容」が含まれるか否かを人手でタグ付けしたコーパスを作成し、それらから文の「問題内容」の有無を判定する分類器を作成し、この精度を調べる。次に作成した分類器を、まだアノテーションがされていない、「問題」を含む文及びその前後一文を対象に適用することで、機械的にタグ付けを行い、コーパスの増補を行う。最後に、増補されたコーパスから新しく分類器を作成し、この精度を調べる。

2. 関連研究

科学技術論文を対象にした自動知識抽出システムの研究は多い。例えば、Peng and McCallum (2006) は、科学技術論文を対象に、論文のヘッダーと呼ばれる冒頭部分から、CRF を用いて、タイトル、著者、所属などの 15 種類の情報を抽出している。

Heffernan and Teufel (2018) は、“problem” とその類義語を利用した情報抽出により、英語科学論文から問題提起箇所を含む「問題内容」と「解決法」コーパスを作成し、さらに、文中に「問題内容」または「解決法」を含むか否かという分類を行った。

また、「問題内容」を抽出するために、「問題」に注目している研究は他にもあり、Scott (2001) は、「問題」や解決法という語に類似した単語を特定し、それらの単語が問題解決プロセスを示す際にどのように用いられるかを明らかにしていた。

本論文で用いている自己学習手法は、分類器の精度の向上に貢献する。例えば、Yarowsky (1995) や鈴木ほか (2019) は、ブートストラップ法による自己学習を行い、教師なし学習による語義曖昧性解消の精度が向上している。

3. 提案手法

3.1 「問題」の語義曖昧性解消

まずはじめに、平文のコーパスから、「問題」を抽出し、その語義曖昧性解消を行う。先行研究として、平林ほか (2021) で、「問題」には、困っていること、解決したいことを意味する “problematic”、クイズなどのお題という意味の “task” 等の意味があるが、「問題内容」および「解決法」が存在する「問題」は “problematic” の意味のみであるから、「問題」の語義曖昧性解消を必要とすることを述べ、「問題」の語義をタグ付けしたコーパスの作成を行った。先行研究では、コーパス中の「問題」の一部の用例について人手で語義曖昧性解消を行い、残りを機械的に付与したコーパスを作成したが、本研究で用いられる「問題」を含む文は全て人手で語義曖昧性解消を行った文である。

3.2 「問題内容」のアノテーション

本研究では、「問題」の意味をタグ付けしたコーパスから、本研究では、「問題」の意味をタグ付けしたコーパスから、「問題」を含む文及び、その前後一文を対象に、文に「問題内容」が含まれるか否かのタグを付与する。具体的には、対象三文に、人手で一文ずつ、後述のタグ

付けガイドラインに基づき、文に、III(a)~III(c) のタグを付与する。ここで対象とする「問題」は、「問題」の語義を特定済みで、なおかつ「問題内容」とその「解決法」が存在し得る ‘‘problematic’’ の意味のものに限られる。さらに本研究では、「問題」の指し示す「問題内容」の抽出範囲を、その「問題」を含む文か、その周辺文のみとした。タグ付けのガイドラインは以下である。

- I 対象とする三文内で「問題内容」であると判断できない場合、タグを付与しない。
- II 「問題内容」は文中に叙述されている場合のみとし、具体的には「X は問題だ」と言い換えられる場合のみ、タグを付与する。
- III 「問題」との意味関係から「問題内容」を下記の3種類に分け、それぞれタグを付与する。

(a) 「問題」が直接指す「問題内容」

[例文] 従って、原則をルールとして生成しても、有効に機能しない場合があるという問題がある。

【出典】『言語処理学会論文誌 LaTeX コーパス』 V12/V12N02-01.tex

(b) 「問題」が直接指す「問題内容」と同じ内容で言い換え

下記例文の二文目に該当する「問題内容」を含む。

[例文] この比喩的な表現の問題を解決するには、比喩に関する人間の常識的な推論が必要である。例えば、「頭が痛い」「寒気がする」「発熱がある」など、疾患・症状が比喩的に使用される例は多くある。

【出典】『言語処理学会論文誌 LaTeX コーパス』 V22/V22N05-02.tex

(c) 「問題」が直接指す「問題内容」とは異なる内容で補足・展開

下記例文の二文目に該当する「問題内容」を含む。

[例文] この選別における問題は、選別の妥当性を確保することである。さらに、選別の対象であるがん用語の候補集合が、なるべく多くのがん用語を網羅していることを保証する必要もある。

【出典】『言語処理学会論文誌 LaTeX コーパス』 V16/V16N02-01.tex

3.3 問題内容を含む文か否かを分類する分類器の作成

作成した「問題内容」のタグを手で付与したコーパスを用いて、「問題」を含む文とその前後一文の計三文を対象に分類器を学習する。この時、対象三文に別の分類器を作成するか否か、使用するモデル、使用する「問題内容」のタグ付け範囲をそれぞれ変化させ、以下の8種類の実験を行った。対象三文に共通の分類器を作成するほかに、対象三文に別の分類器を作成する理由としては、「問題」の前後文の分類器に、それぞれ「問題」の前後文であるという情報を分類器に入力として加えることによる精度の変化を調査するためである。

(1) 対象とする三文内で共通の分類器を学習する。

この時、各文はそれぞれ独立して分類器に入力する。

また、3.2 節の III で定義した「問題内容」の分類のうち、(a) と (b) と (c) を正例とする。

- i モデルは、線形の SVM を用いる。
- ii モデルは、BERT(Devlin et al. 2018) を fine-tuning したものをを用いる。

(2) 対象とする三文内で異なる分類器を学習する。

「問題」を含む文の前文の分類器を学習する際には、「問題」を含む文の前文と「問題」を含む文の 2 文を分類器の素性として用いる。

「問題」を含む文の後文の分類器を学習する際には、「問題」を含む文と「問題」を含む文の後文の 2 文を分類器の素性として用いる。

「問題」を含む文の分類器を学習する際には、「問題」を含む文の 1 文のみを分類器の素性として用いる。

- i モデルは、線形の SVM を用いる。
 - A 3.2 節の III で定義した「問題内容」の分類のうち、(a) のみを正例とする。
 - B 3.2 節の III で定義した「問題内容」の分類のうち、(a) と (b) を正例とする。
 - C 3.2 節の III で定義した「問題内容」の分類のうち、(a) と (b) と (c) を正例とする。

- ii モデルは、BERT を fine-tuning したものをを用いる。
 - A 3.2 節の III で定義した「問題内容」の分類のうち、(a) のみを正例とする。
 - B 3.2 節の III で定義した「問題内容」の分類のうち、(a) と (b) を正例とする。
 - C 3.2 節の III で定義した「問題内容」の分類のうち、(a) と (b) と (c) を正例とする。

線形の SVM による分類器を学習する際には、対象コーパスをランダムに学習データとテストデータに分け、BERT の fine-tuning による分類器を学習する際には、テストデータを線形の SVM と同じ範囲になるように、対象コーパスを学習データと開発データとテストデータに分け、学習データでモデルを学習し、開発データでパラメータ調整を行った後、学習データと開発データを合わせて学習した（図 1）。

その後学習した分類器を用いて、テストデータにより精度の評価を行った。

3.4 自己学習による分類器の追加学習

最後に、作成した分類器を、「問題内容」のタグが付与されていないコーパスにそれぞれ適用することで、「問題内容」のタグを機械的に付与し、コーパスの増補を行う。また、増補したコーパスから、それぞれ分類器を再学習し、その精度を評価する。

この時、線形の SVM による分類器を学習する際には、増補分のコーパスを全て、学習データに加え、コーパス増補前のテストデータで精度を求めた（図 2）。

BERT による分類器を学習する際には、増補分のコーパスをランダムに学習データと開発データに分け、増補前のコーパスの学習データに増補分の学習データを加え、モデルを学習し、増補前のコーパスの開発データと増補分のコーパスの開発データでパラメータ調整を行った後、増補前と増補分の学習データと開発データを合わせて学習し、増補前のテストデータで精度を求めた（図 3）。

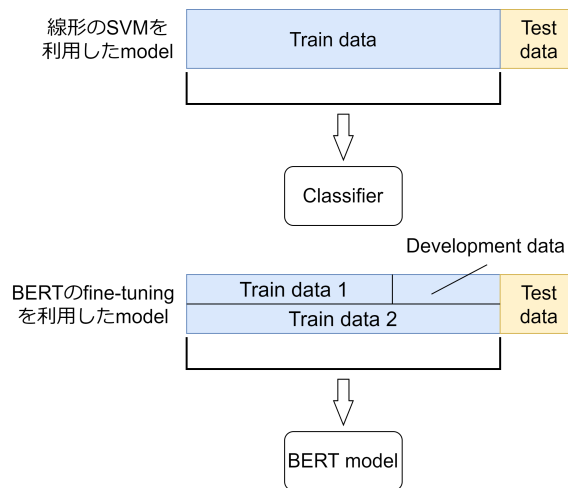


図1 線形のSVMを利用したmodelとBERTのfine-tuningを利用したmodel(ただし、二つのmodel間でTest dataの範囲が同一になるようにとる)

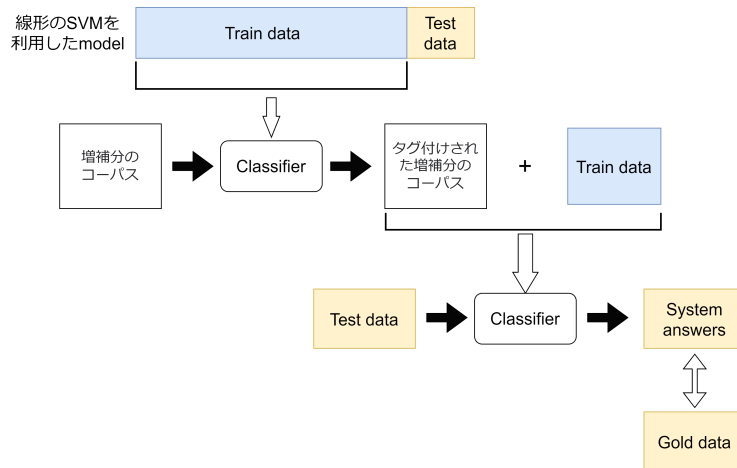


図2 線形のSVMを利用したmodelの再学習とその評価

4. 実験設定

4.1 コーパス情報

本論文ではコーパスとして『言語処理学会論文誌 LaTeX コーパス』⁽¹⁾と、『現代日本語書き言葉均衡コーパス (BCCWJ)』(Maekawa et al. 2014)を利用した。人手で「問題内容」のタグを付与した「問題」を含む文、「問題」を含む文の前文、「問題」を含む文の後文それぞれに対してアノテーションを行い、(a)のみを正例としたときの内訳を表1、(a)と(b)を正例としたときの内訳を表2、(a)と(b)と(c)を正例としたときの内訳を表3に示す。この時、人手でつけた対象コーパスはすべて『言語処理学会論文誌 LaTeX コーパス』内の文である。

⁽¹⁾ https://www.anlp.jp/resource/journal_latex/index.html

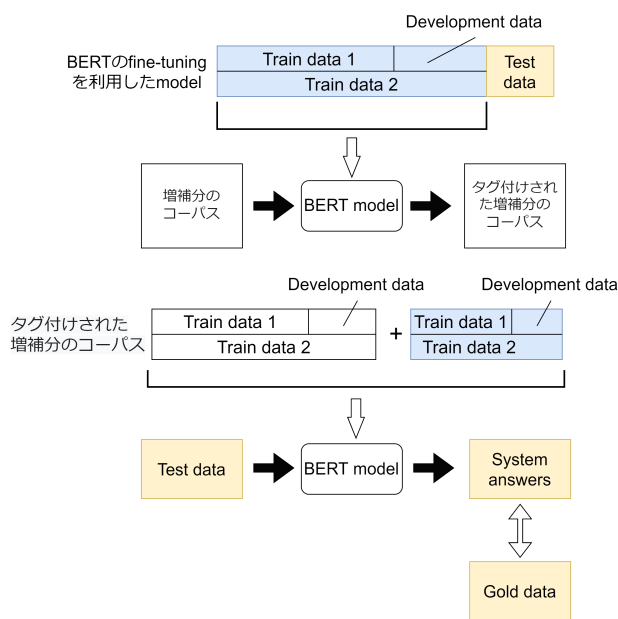


図3 BERTのfine-tuningを利用したmodelの再学習とその評価

表1 (a)のみを正例とするときの人手でタグを付与した対象三文の内訳

	「問題」を含む文		前文		後文		総計
	正例	負例	正例	負例	正例	負例	
人手によるアノテーション総数	133		133		133		399
学習データ	48	31	10	69	1	78	237
開発データ	16	11	3	24	0	27	81
テストデータ	16	11	2	25	0	27	81
総数	80	53	15	118	1	132	399

表2 (a)と(b)を正例とするときの人手でタグを付与した対象三文の内訳

	「問題」を含む文		前文		後文		総計
	正例	負例	正例	負例	正例	負例	
人手によるアノテーション総数	133		133		133		399
学習データ	48	31	11	68	7	72	237
開発データ	17	10	5	22	2	25	81
テストデータ	15	12	3	24	2	25	81
総数	80	53	19	114	11	122	399

表3 (a)と(b)と(c)を正例とするときの人手でタグを付与した対象三文の内訳

	「問題」を含む文		前文		後文		総計
	正例	負例	正例	負例	正例	負例	
人手によるアノテーション総数	133		133		133		399
学習データ	45	34	15	64	10	69	237
開発データ	20	7	4	23	6	21	81
テストデータ	19	8	7	20	5	22	81
総数	84	49	26	107	21	112	399

コーパスの増補時に用いた対象文は、2021年9月6日-2021年9月8日に実施した yahoo! クラウドソーシングにより整備した。実施時の作業員への質問例を図4に示す。

設定した設問ID: 6222ZV16N03Z01

以下の【問題】という単語の使い方について当てはまるものすべてにチェックをいれてください

また、精度を低下させる原因として、属性・属性値とfacetを含む関係を上位下位関係と誤判定する【問題】が多いことも分かった。

- 「--は【問題】でない」「--は【問題】にならない」と同じ意味である
- 「○○【問題】」など複合語の一部である
- 「【問題】点」に言い換えられる
- 「タスク」に言い換えられる
- 「問い」に言い換えられる
- 「疑問点」に言い換えられる
- 「困難」に言い換えられる
- 今の技術では簡単には解けそうにはない【問題】である
- 「対処する(される)」「解消する(される)」【問題】である

[6222ZV16N03Z01]

図4 クラウドソーシング作業員への質問例

コーパスの増補分の、「問題」を含む文数を表4に示す。

4.2 作成した分類器

SVMはScikit-learnライブラリのlinearSVCモデルをデフォルト値で使用した。また、事前学習済みBERTモデルは、東北大学乾研究室で公開されている“cl-tohoku/bert-base-

表4 クラウドソーシングによる対象文数と、コーパスの増補に用いた文数の内訳

クラウドソーシング実施対象文数	10,000
内、「問題」を含む文数 (コーパスの増補分の文数)	1,353
内、『言語処理学会論文誌 LaTeX コーパス』内の文数	450
内、BCCWJ 内の文数	903
学習データ	1,082
開発データ	271

japanese-v2⁽²⁾を使用した。また、いずれの分類器も入力として、事前学習済み BERT モデルから出力された分散表現を用いた。対象三文内でそれぞれ異なる分類器を作成する時、「問題」を含む文から作成される分類器を以降の表内で「抽」と表記し、同様に「問題」を含む文の前文から作成される分類器を「前」、「問題」を含む文の後文から作成される分類器を「後」と表記する。

人手でタグを付与した「問題内容」コーパスに対して fine-tuning により学習した BERT モデルの学習率と epoch 数を表5に示す。

表5 人手でタグを付与した「問題内容」コーパスに対して学習した BERT モデルの各種パラメータ

分類器の番号	(1)-BERT		(2)-BERT
学習率	0.001		
epoch 数	3	A	抽 5
			前 0
			後 0
		B	抽 3
			前 4
			後 0
		C	抽 7
			前 6
			後 7

コーパス増補後の「問題内容」コーパスに対して fine-tuning により学習した BERT モデルの学習率と epoch 数を表6に示す。

⁽²⁾ <https://huggingface.co/cl-tohoku/bert-base-japanese-v2>

表 6 コーパス増補後の「問題内容」コーパスに対して学習した BERT モデルの各種パラメータ

分類器の番号	(1)-BERT		(2)-BERT
学習率	0.001		
epoch 数	26	A	抽 29
			前 0
			後 0
		B	抽 22
			前 20
			後 0
		C	抽 24
			前 1
			後 3

5. 実験結果

対象三文内でそれぞれ異なる分類器を作成する時、対象三文の分類器による判定が全て正しい時の精度を表中で「計」と表記する。コーパス増補前の分類器の正解率を表 7 に示す。

表 7 コーパス増補前の分類器の正解率

分類器の番号	(1)-SVM	(1)-BERT		(2)-SVM	(2)-BERT
正解率	0.70	0.79	A	抽 0.56	抽 0.70
				前 0.89	前 0.85
				後 1.0	後 1.0
				計 0.48	計 0.63
			B	抽 0.63	抽 0.89
				前 0.89	前 0.85
				後 0.93	後 0.89
				計 0.59	計 0.67
			C	抽 0.70	抽 0.78
				前 0.85	前 0.85
				後 0.78	後 0.78
				計 0.48	計 0.48

コーパス増補後の分類器の正解率を表 8 に示す。

表8 コーパス増補後の分類器の正解率

分類器の番号	(1)-SVM	(1)-BERT		(2)-SVM	(2)-BERT
正解率	0.72	0.69	A	抽 0.78	抽 0.67
				前 0.93	前 0.85
				後 1.0	後 1.0
				計 0.74	計 0.52
			B	抽 0.67	抽 0.56
				前 0.85	前 0.89
				後 0.85	後 0.89
				計 0.52	計 0.41
			C	抽 0.85	抽 0.70
				前 0.74	前 0.74
				後 0.85	後 0.81
				計 0.52	計 0.30

6. 考察・展望

表7と表8を見比べると、(2)-SVM-Bの分類器を除いて、SVMを用いた分類器では自己学習による精度の向上がみられる。一方、BERTを用いた分類器では、コーパス増補後は一律精度が下がっている。これには二つの可能性が考えられる。一つは、学習データに開発データを加えて再学習する手法がBERTのfine-tuningとは相性が悪い可能性が考えられる。学習データに開発データを加えて再学習する手法は、今回のように、学習データが非常に少ない時に、学習データを増やす目的で利用され、平林ほか(2021)でも同様の手法をとり、今回の実験結果と同様、線形のSVMでは効果があることが確認されている。しかし、BERTでのfine-tuningでは初めての試みであったため、これらの手法が有効であるか確認できていない。もう一つは、異なる分野での学習の影響をBERTでのfine-tuningでは強く受けてしまう可能性が考えられる。本論文では日本語の科学技術論文における「問題内容」のアノテーションを目指しているが、日本語の科学技術論文が十分に集まらなかったため、他分野コーパスを用いて精度向上を狙った。しかし、他分野コーパスの文が多い学習データで、日本語の科学技術論文のテストデータをあてるのがBERTでのfine-tuningによる分類器には難しいタスクであった可能性がある。

これらの可能性を踏まえ、増補分のコーパスの精度を人手でタグをつけることで、それぞれの分類器のタグ付け精度を求め、追加のコーパスを精度良く機械的にタグ付けを行えるシステムを探していく。

また、複数回実験を行うことによる正解率の差が大きいため、十分な実験数をとって追報告したい。

7. おわりに

本論文では、日本語科学技術論文中の「問題」を含む文及び、その前後一文を対象に、「問題内容」が含まれるか否かを人手でタグ付けしたコーパスを作成し、そのコーパスから作成される分類器の精度と、それをを用いてコーパスを増補し、自己学習を行い分類器の精度の向上を目指した。その結果、線形の SVM からなる分類器の精度の向上は概ね見られたが、BERT の fine-tuning からなる分類器の精度の向上をすることが出来なかった。自己学習の手法を見直し、さらなる分類器の精度の向上により、機械的に文に「問題内容」を含むかをタグ付けし、最終目標である日本語論文において「問題内容」と「解決法」のペアを効率的に取得するシステムの作成を目指したい。

謝辞

本研究は、国立国語研究所共同研究プロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」により作成されたコーパスを利用したものです。また、本研究は JSPS 科研費 18K11421、17KK0002 の助成を受けています。深く感謝いたします。

文献

- Kevin Heffernan, and Simone Teufel (2018). “Identifying problems and solutions in scientific text.” *Scientometrics*, 116:2, pp. 1367–1382.
- Fuchun Peng, and Andrew McCallum (2006). “Information extraction from research papers using conditional random fields.” *Information Processing & Management*, 42:4, pp. 963–979.
- Mike Scott (2001). “Mapping key words to problem and solution.” *Patterns of Text: in Honour of Michael Hoey. Benjamins, Amsterdam*, pp. 109–127.
- David Yarowsky (1995). “Unsupervised word sense disambiguation rivaling supervised methods.” *33rd annual meeting of the association for computational linguistics*, pp. 189–196.
- 鈴木類・古宮嘉那子・浅原正幸・佐々木稔・新納浩幸 (2019). 「概念辞書の類義語と分散表現を利用した教師なし all-words WSD」 *自然言語処理*, 26:2, pp. 361–379.
- 平林照雄・河野慎司・古宮嘉那子・新納浩幸 (2021). 「日本語の論文コーパスにおける「問題」の語義アノテーション」 *言語処理学会第 27 回年次大会*, pp. 1151–1155.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). “Bert: Pre-training of deep bidirectional transformers for language understanding.” *arXiv preprint arXiv:1810.04805*.
- Kikuo Maekawa, 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*, 48:2, pp. 345–371.

対のある自他動詞と文法を組み合わせた項目を用いた オンライン学習の解答傾向 —5週間のオンライン学習で用いた項目の分析を中心に—

沖本 与子 (東京外国語大学) †

Student Tendencies of Five Week Online Study Materials for Paired Transitive/Intransitive Verb-Grammar Items

Tomoko Okimoto (Tokyo University of Foreign Studies)

要旨

本稿は2021年秋学期に実施したオンライン学習システム調査で得た解答データの中から、特に対のある自他動詞と文法を組み合わせた項目を選出しその分析を行うことを目的とした。その結果、動詞と文法を組み合わせた項目の基本統計量は、動詞のみの項目と、平均・標準偏差 (*SD*) とともに大きな違いがないことが確認された。また文法を用いた項目の解答 (2,000件) は、グラフ上、全解答とも動詞項目とも異なる推移を見せ、正答率は第4週・第5週ともに月曜日から金曜日の間に下降・上昇を示し、標準偏差は日を追うごとに広がった。最後に、インタビュー分析から文法を用いた項目に対し、まず選択肢の助詞または動詞を確認する学習者が大半であることが確認された。

1. はじめに

本研究は数年に渡り実施しているオンライン学習で得た解答傾向を分析し、オンライン学習システムの開発に繋げることを目的とする。本稿では用意した項目の中から、特に対のある自他動詞と文法を組み合わせた項目 (以下、文法項目) を選出しその分析を行う。

今まで筆者は対のある自他動詞のみを研究対象とし、オンライン学習システムの構築を行ってきたが、対のある自他動詞を用いた項目 (以下、動詞項目) における学習者の解答傾向や反応についてある程度の蓄積ができ、また、くり返し学習による対のある自他動詞の知識定着が確認できた。そのため、重複しない項目を用いた場合の学習者の解答傾向と、過去の調査に参加した日本語学習者から、対のある自他動詞以外の様々な日本語学習項目について同様にオンライン学習ができるよう要望が多く出たことから、新たな試みとして文法項目を作成し、調査から取得したデータを分析することとした。本稿では、以下の研究課題を明らかにすることを目的とする。

研究課題1: 文法項目の基本統計量は、動詞項目の基本統計量と異なるのか

研究課題2: 文法項目は、どのような解答傾向を示すのか

研究課題3: 文法項目に対し、学習者は解答する際にどのような反応を示すのか

2. 先行研究

対のある自他動詞を教育分野で応用した研究は複数あるが (守屋 1994・小林 1996・中石 2003・伊藤 2014・沖本 2020)、オンライン学習に応用し学習者の知識定着を確認した研究

† tokimoto@tufs.ac.jp

は沖本（2020）から始まっている。一方、オンラインで文法を指導する研究またはオンライン学習教材の開発は前原他（2007）以降始まっている。実際に授業で運用しつつ教材開発を行っている研究が多く、自立的なオンライン学習は「まるごと e ラーニング」や「NIHONGO e な」などが挙げられる。様々な教材や研究の中で、対のある自他動詞と初中級文法を組み合わせた項目を用いた自立的なオンライン学習の提供及び研究は管見の限り見当たらない。

3. 調査概要・項目概要・取得データ

3.1 調査概要と調査参加者

調査は、2021年11月15日から12月17日までの5週間、オンライン上で実施した。また、5週間のオンライン学習前後に筑波 SPOT・プレテスト・ポストテスト・作文産出・インタビューを実施した。調査参加者（以下、学習者）は、都内の高等教育機関に所属する日本語学習者20名であり、JLPT取得と母語の内訳は表1と表2の通りである。

表1 JLPT取得級と人数

N1	10
N2	3
N3	4
N4	1
N5	0
JLPTなし	2
合計	20

表2 母語の内訳

中国語母語話者	13
韓国語母語話者	1
英語母語話者	1
アジア諸言語	2
ヨーロッパ諸言語	3
合計	20

3.2 項目概要と取得データ

学習者は、月曜日～金曜日の5日間の学習を1週とし、5週間（5回）のオンライン学習を受けた。学習は1回につき50項目提示され、学習者は約30分で解答していることが確認された。また、学習者は日本国内外にあり、時差の関係もあったことから、オンライン学習は学習者の任意の時間帯に自主的に行うようにしていた。調査では、以下の条件を基に沖本（2021）を改善し、1,250項目用意した。

表3 項目作成の条件

- | |
|---|
| <p>1) 初級から上級の対のある自他動詞を用いること</p> <p>2) 項目文は、1つの動詞につき3～6文作成し、文の重複がないように作成すること
 なお、漢字には後ろに（ ）で読み仮名を付与すること
 例：上がる（あがる）
 成績(せいせき)が【 】。
 （2文比較で出題：ペアは「成績(せいせき)を【 】。」
 雨(あめ)が【 】。
 階段(かいだん)【 】【 】。
 打(う)ち上(あ)げ花火(はなび)【 】【 】。
 驚(おどろ)きの声(こえ)が【 】。
 （2文比較で出題：ペアは「驚(おどろ)きの声(こえ)を【 】。」</p> <p>3) 対のある自他動詞と初中級文法を組み合わせた項目を作成すること
 例：あのビルの前（まえ）で車（くるま）【 】【 】ください。(ID979 : N5)
 この天気（てんき）なら飛行機（ひこうき）【 】【 】はずだ。(ID1080 : N4)</p> |
|---|

また、項目は日ごと、週ごとに旧 JLPT の難易度を基にレベルを上げている。その中で文法項目は第 4 週と第 5 週に 1 日 10 項目ずつ出題した。第 1 週から第 3 週は動詞項目を 50 項目提示し、第 4 週と第 5 週は動詞項目を 40 項目、文法項目を 10 項目の計 50 項目を出題した。本システム開発に当たり、対象者を中級レベルの日本語学習者としていることを鑑み、文法項目は、文法は旧 JLPTN2～N5 レベルを、動詞は旧 JLPT の N2～N4 レベルを使用し作成した。図 1 は項目の内、設問 1 での出題設計である。例えば、左端の W1-1 は第 1 週の 1 日目を表しており、設問 1 では b (beginner) の旧 JLPTN4 レベルである初級 5 項目と、i (intermediate) の旧 JLPTN2 レベルである中級 5 項目の計 10 項目を出題している。週を追うごとに移動動詞を用いた項目・上級レベルの項目・文法項目を追加した。また新しい項目を追加する前日までに、解説をまとめた 5 分程度のビデオ視聴を依頼していた。これは今までの調査に参加した学習者から「最初に、対のある自他動詞について説明が欲しい」「予め移動動詞の説明が欲しい」などの要望が出たことにより、学習者の情意フィルターを低くするために用意した。

	設問 1								
	初級		中級		移動動詞		上級		文法
	項目数	レベル	項目数	レベル	項目数		項目数	レベル	項目数
W1-1	5	b4	5	i2					
W1-2	5	b4	5	i2					
W1-3	5	b4	5	i2					
W1-4	5	b3	5	i2					
W1-5	5	b3	5	i2					
W2-1	4	b4	3	i2	3	m			
W2-2	4	b4	3	i2	3	m			
W2-3	4	b3	3	i2	3	m			
W2-4	4	b3	3	i2	3	m			
W2-5	4	b3	3	i2	3	m			
W3-1	2	b4	2	i2	3	m	3	i0, i1, eとする	
W3-2	2	b4	2	i2	3	m	3	i0, i1, eとする	
W3-3	2	b3	2	i2	3	m	3	i0, i1, eとする	
W3-4	2	b3	2	i2	3	m	3	i0, i1, eとする	
W3-5	2	b3	2	i2	3	m	3	i0, i1, eとする	
W4-1	2	b4	2	i2	2	m	2	a2	2
W4-2	2	b4	2	i2	2	m	2	a2	2
W4-3	2	b3	2	i2	2	m	2	a2	2
W4-4	2	b3	2	i1	2	m	2	a2	2
W4-5	2	b3	2	i1	2	m	2	a2	2
W5-1	2	b4	2	i1	2	m	2	a2	2
W5-2	2	b4	2	i1	2	m	2	a2	2
W5-3	2	b3	2	i1	2	m	2	a2	2
W5-4	2	b3	2	i1	2	m	2	a2	2
W5-5	2	b3	2	i1	2	m	2	a2	2

図 1 出題項目数とレベル (設問 1)

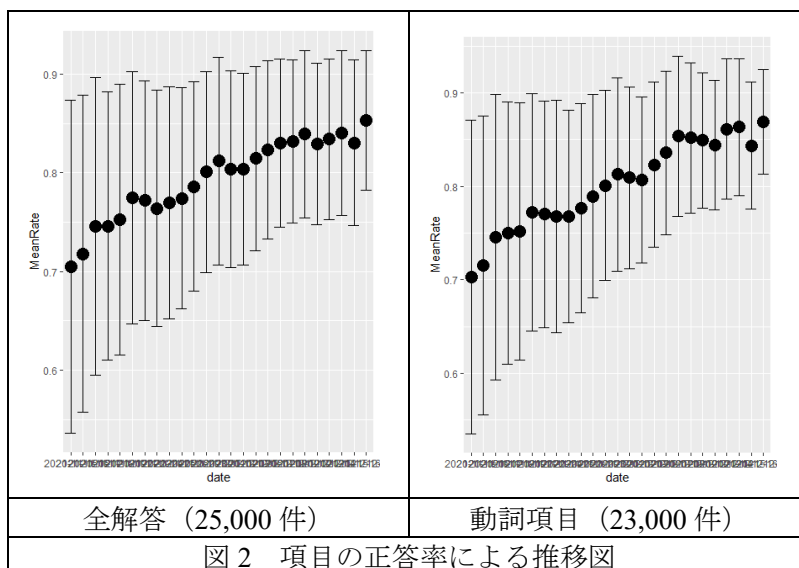
前述のように 1 日 50 項目、1 週間につき 5 日間を 5 週間実施し、参加者が 20 名であったため、25,000 件の解答を得た (50 項目×5 日×5 週間×20 名)。欠損値はなかったため、全データを使用する。文法項目は 100 項目用意しており、取得データは 25,000 件中 2,000 件に当たる (100 項目×20 名)。こちらも欠損値がないため、全データを使用する。

4. 分析と考察

4.1 全体の解答傾向

まず、取得したデータ 25,000 件を、動詞項目の 23,000 件と文法項目の 2,000 件に分けて正答率の推移を確認した。図 2 は黒丸が正答率の平均であり、上下の線が標準偏差を表す。全解答と動詞項目の推移図を確認すると、どちらも日を追うごとに正答率が上昇し、標準偏

差が狭まり，ほぼ同じ傾向を示すことが分かる。また，この推移図から学習者が5週間のオンライン学習で，重複のない項目であっても，正答率を上昇させることが確認された。



4.2 文法項目の解答傾向

続いて，動詞項目と文法項目の比較を行う。表4は動詞項目と文法項目の基本統計量である。調査で用いた項目は，正答は1，誤答は0としている。

表4 動詞項目と文法項目の基本統計量（20人）

	動詞項目	文法項目
受験者数（人）	20	20
解答数（件）	23,000	2,000
平均	0.79	0.76
SD	0.40	0.43
中央値	1	1
最小点	0	0
最大点	1	1

研究課題1「文法項目の基本統計量は，動詞項目の基本統計量と異なるのか」については，動詞項目の解答数は文法項目の解答数の10倍以上になるが，平均・標準偏差（SD）ともに大きな違いがないことが確認された。これは動詞項目と文法項目に対する学習者の解答傾向が数値上の差異がないことを示している。

続いて，全体の解答傾向と文法項目の解答傾向を確認する。文法項目は，5週間の内第4週と第5週に出題した。そのため図3にあるように，左図「全解答」の四角枠に当たる箇所が文法項目を含んでおり，この中から2,000件を抜粋し推移を確認した。その結果グラフ上では文法項目は，全解答とも動詞項目とも異なる推移を見せることが分かった。文法項目の正答率は第4週・第5週ともに月曜日から金曜日の間に下降・上昇を示し，標準偏差は日を追うごとに広がった。

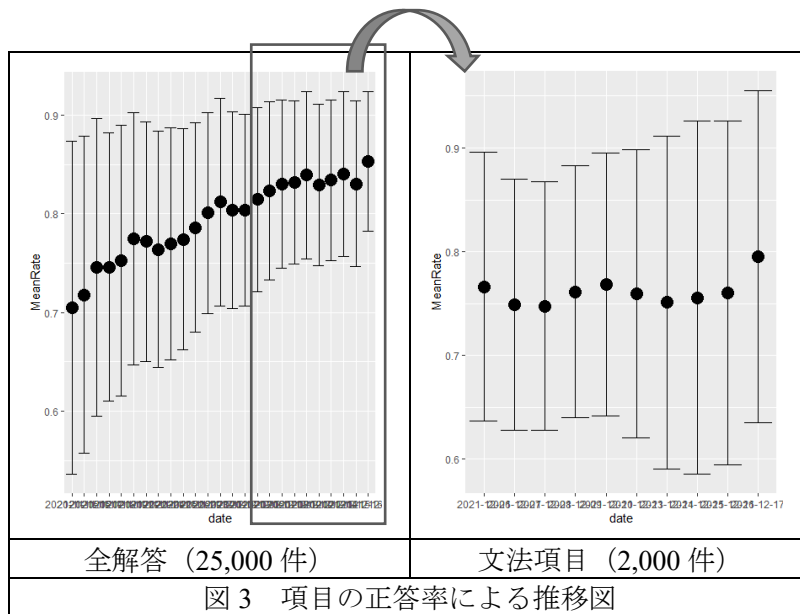


図3 項目の正答率による推移図

研究課題2「文法項目は、どのような解答傾向を示すのか」については、前述のように全体の推移とはことなる動きを見せることが確認された。文法項目は、文法は旧 JLPT の N2 から N5 レベルを、動詞は N2~N4 レベルを用いて作成し、できるだけ難易度が均一になるように目視で文法のレベルと項目数を確認しながら第4週から第5週に項目を配置した。図3の分析を受け、1週間の内、水曜日に難易度の高い項目があるかを確認するために、便宜上旧 JLPT のレベルを数値としてとらえ、平均値を算出しグラフ化した(図4)。Y軸の1~5は旧 JLPT のレベルを表し、X軸は日付を表す。例えば、W4-1は第4週の1日目(=月曜日)を表す。なお、Y軸は1から積みあがるため、グラフが上昇するほど、より難易度が低く学習者にとって既習に当たるであろう項目が出題されていることを示す。

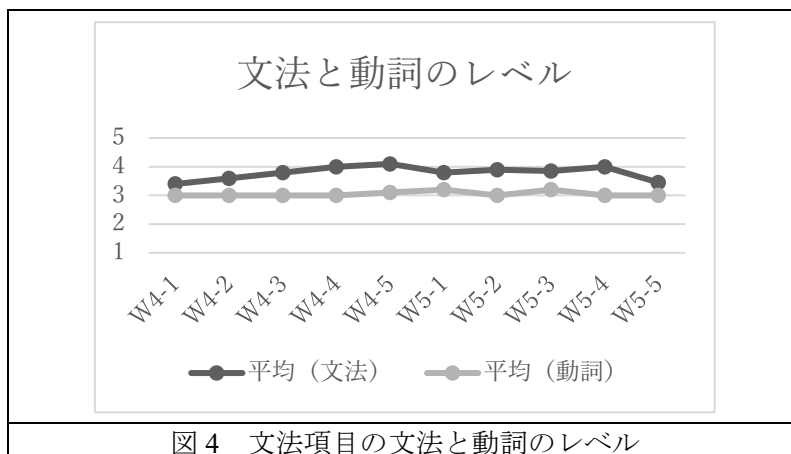


図4 文法項目の文法と動詞のレベル

図3と図4を確認すると、週半ばの水曜日に向けて項目の難易度が上がっているわけではない。そこで水曜日にどのような誤答が生じているのかを確認することにする。水曜日の項目は第4週が項目番号851-900であり、第5週が項目番号1101-1150を使用している。その中で、文法項目かつ正答率が50%以下(25%~40%)の項目を選出した。

表5 第4週と第5週の3日目における低正答率の項目
（ ）内の数字は旧 JLPT レベルを示す

出題日	項目番号	項目 (学習時には漢字に読み仮名が付与してある)	正答率%	正答	最も多い誤答
W4-3	869	テーブルに, 8人分のお皿が【 】あった。(4)	25	並べて(4)	並んで(4)
W4-3	870	遅い時間には, 駅前の道を【 】ようにしている。(3)	20	通る(3)	通す(3)
W4-3	880	客を乗せた船が川【 】【 】いく。(4)	25	を下って(2)	が下って(2) に下って(2)
W4-3	889	私の部屋は壁に時計【 】【 】ある。(4)	40	がかけて(4)	がかかって(4) にかけて(4)
W5-3	1139	友達に貸した漫画【 】【 】こない。(4)	35	が返って(3)	を返して(4)
W5-3	1140	日本語がわからないときはいつも姉【 】【 】もらう。(4)	30	に助けて(2)	が助かって(2)
W5-3	1149	①アルバイトの面接と試験の日が【 】しまった。(4) ②大きい皿から順に【 】もらいたい。(4)&(5)	25	重なって(3)・ 重ねて(2)	重ねて・(2) 重なって(3)

正答率の低い項目は動詞に旧 JLPT の N2 レベルを使用しているが(例: 1149), 初中級で学習する N3・N4 レベルの動詞も使用しており(例: 869, 870), 組み合わせた文法も N3・N4・N5 レベルを使用している。特に得点が下がる要素のある項目でもなければ, 前後に影響を受けるような, 例えばより難易度の高い項目も配置されていない。

第4週と第5週の週半ばに向けて下降し, その後上昇した平均点については, 明確な数値上の根拠がないため, 学習者が回答したインタビューを確認することとした。

4.3 インタビュー分析

インタビューは20名の内8名に, 2021年12月22日~2022年1月初旬の間に実施した。対象者としてプレテスト・ポストテストを比較し, 伸び率の高い学習者, 伸び率の低い学習者, ポストテストの得点が高い学習者(100点~125点) ポストテストの得点が高い学習者(89点~99点)を基準に12名ずつ選び依頼をした。12名中8名と各自約1時間のインタビューを実施した。インタビュー回答から, 文法項目の解答方法に関する学習者の回答を以下にまとめる。なお, 第4週開始前に文法項目が加わる旨を連絡していたため, 学習者は文法項目に気づいており, 動詞項目と比較したコメントを述べた。

なお, 学習者はA~Hとし, 「」内が学習者の回答であり()は筆者の加筆である。筆者の加筆は学習者に確認し加えている。

① 動詞項目と文法項目では, 動詞項目の方が簡易だと見なしている場合

「動詞だけの方が簡単でした。短いから, どんどん答える。でも文法の問題が出ると時間がかかる。(中略) 短い文と比べると, (文法項目は) 長いです。」[学習者C]

「(文法項目は) 難しかった。もともと文法がきらい, あー, 苦手。だから, あ, これ文法だって分かっていたら, いやだった。(中略) 動詞の方が簡単でした。文が短い, 短いし, 動詞が分かるとき, 意味が分かる。」[学習者E]

② 文法項目を解答する際、選択肢を確認しその後文全体を確認する傾向

「文法は最後にあって、他の文より長いから、あー、「あ、来た」と思って、止まって。一度で分からないときもあるから、何度も読みました。最初にチョイス（選択肢）を探します。知っている文法は見ます、でも最初に particle（助詞） とか、動詞だけを見ました。」〔学習者 B〕

「（文法項目は）実は読みませんでした。あー読まない文もありました。全部じゃないです。でも文法（の）文は長いから、多分これって答えを選んでました。全部じゃないです。読んだ文はちょっと短い、答えの動詞を知っている文。答え（選択肢）を見て意味が分かったら、文を読む、読みました。」〔学習者 G〕

もちろん、上記②とは異なり、学習者の中には文法項目は文が長いから、文中からヒントを得て解答に繋げると回答した学習者もいる。

「動詞だけ（の）文は動詞が分からないと答えられない。でも文法（の）文は、何となく文から意味が分かった。（中略）文が長いから、分からないことばがあった、あっても、分かる時があった。」〔学習者 A〕

「たくさん読んで、そのうち分かってきます。（同じ文を繰り返し読むという意味）それで、答え（選択肢）を読んで、選ぶ、そんな順番でした。」〔学習者 D〕

週内で正答率が下降・上昇する理由と、研究課題3「文法項目に対し、学習者は解答する際にどのような反応を示すのか」を明らかにするために、インタビュー回答をまとめたことで、学習者ごとに、文法項目への解答の仕方が異なることが分かった。動詞項目に比べ、文法項目は複数要素が組み合わされるため、まず選択肢の助詞または動詞を確認する学生が多くいた。文を最初に読み、文法を確認しそれに合う動詞の活用を選ぶ学生は2名で、6名は選択肢を確認しその後文を読む解答方法を取っていた。

文法項目は対のある自他動詞と文法を組み合わせることにより、動詞項目より長く、文全体の意味確認、動作主（主語）の確認、動詞の活用（または助詞の選択）、など一つの文に複数要素が存在する。そのため、動詞項目のようなほぼ単文を読む場合と異なり、Cognitive overload（認知負荷：Fayol, Lamy, and Lemaire, 1994）がかかった可能性がある。

文法項目に対し苦手意識を持っている学習者がいたこと、中級レベルの日本語学習者であり、人数も20名と多くないことから、学習レベルが十分でない場合、様々な影響要因が介在したと考えられる。これらは参加者の学習レベルや、参加者人数の増加により更に変化する可能性がある。

5. まとめ

本稿は2021年秋学期に実施したオンライン学習システム調査で得た解答の中から、特に文法項目（＝対のある自他動詞と文法を組み合わせた項目）を選出しその分析を行うことで、三つの研究課題を明らかにすることを目的とした。

まず、研究課題1「文法項目の基本統計量は、動詞項目の基本統計量と異なるのか」については、動詞項目の解答数は文法項目の解答数の10倍以上になるが、平均・標準偏差（SD）ともに大きな違いがないことが確認された。これは動詞項目と文法項目に対する学習者の解答傾向が数値上の差異がないことを示している。

次に、研究課題2「文法項目は、どのような解答傾向を示すのか」については、2,000件の取得データの推移を確認した。その結果グラフ上では文法項目は、全解答とも動詞項目とも異なる推移を見せることが分かった。正答率は第4週・第5週ともに月曜日から金曜日の間に下降・上昇を示し、標準偏差は日を追うごとに広がった。

最後に、研究課題3「文法項目に対し、学習者は解答する際にどのような反応を示すか」

については、学習者ごとに、文法項目への解答の仕方が異なることが分かった。動詞項目に比べ、文法項目は複数要素が組み合わせられるため、まず選択肢の助詞または動詞を確認する学生が多くいた。

6. 今後の課題

研究課題については一定の解答を得たが、分析の途中で出現した、文法項目の正答率が週内で下降・上昇する課題については、明確な理由が判明しなかった。この課題については、次回以降の調査で項目の順番を入れ替える、調査人数を増やすなどの変更を行った上で、取得データを比較する必要があると考えられる。

謝 辞

本研究は JSPS 科研費 21K00633 「日本語学習者と日本語教師のためのオンライン自他動詞習得支援システムの開発」(2021-2023) の助成を受けて行われたものである。

文 献

- 伊藤秀明 (2014) 「日本語教育における対のある自動詞・他動詞の研究—文脈重視の文法教育論の構築に向けて—」筑波大学博士 (国際日本研究) 学位請求論文 筑波大学
- 沖本与子 (2020) 「日本語学習者の助詞・動詞選択における解答時間と誤答率の傾向—5週間のオンライン学習項目の分析を中心に—」『言語資源活用ワークショップ 2020 発表論文集』 pp.43-59. 国立国語研究所
- 沖本与子 (2021) 「「対のある自他動詞」の練習・学習を支援するシステムの開発—5週間のオンライン学習における効果—」『2021年度日本語教育学会春季大会予稿集』 pp.153-158. 日本語教育学会
- 小林典子 (1996) 「相対自動詞による結果・状態の表現—日本語学習者の習得状況—」『文芸言語研究言語篇』 29号 pp.41-56. 筑波大学文藝・言語学系
- 中石ゆうこ (2003) 「対のある自動詞・他動詞の習得研究の動向と今後の課題」『広島大学大学院教育学研究科紀要第二部 (文化教育開発関連領域)』 52号 pp.167-174. 広島大学大学院教育学研究科
- 前原かおる・増田真理子・李相穆・菊地康人 (2007) 「日本語の意味的・構造的理解を強化するための聴解型オンライン学習教材の開発」『日本語教育方法研究会誌』 14 (2), pp.54-55
- 守屋三千代 (1994) 「日本語の自動詞・他動詞の選択条件—習得状況の分析を参考に—」『講座日本語教育』 第29分冊 pp.151-165. 早稲田大学日本語研究教育センター
- Fayol, M., Largy, P., and Lemaire, P. (1994) Cognitive Overload and Orthographic Errors: When Cognitive Overload Enhances Subject-Verb Agreement Errors. A Study in French Written Language. *The Quarterly Journal of Experimental Psychology Section A*, 47, 2, pp. 437-464.

関連 URL

- 「まるごと eラーニング」 <<https://www.marugoto.org/e-learning/>>, 2022年7月12日閲覧
- 「NIHONGO e な」 <<https://nihongo-e-na.com/jpn/>>, 2022年7月12日閲覧

NINJAL データベースを活用した言語研究の実施について

鈴木 成典 (国際基督教大学大学院) †

五十嵐 陽介 (国立国語研究所)

李 勝勲 (国際基督教大学)

Linguistic Research Using NINJAL Database

Michinori Suzuki (Graduate School of International Christian University)

Yosuke Igarashi (National Institute for Japanese Language and Linguistics)

Seunghun J. Lee (International Christian University)

要旨

本稿では、国立国語研究所の共同利用型共同研究（登録型）で利用可能な豊富なデータベース（研究資料室収蔵資料：fo0245）を用いたデータ処理方法について紹介する。今回利用したデータベースでは多くのデータへのアクセスが可能な反面、実験全体の録音ファイルのみが利用可能であったため、初めに実験の録音音声をもとに刺激のメタデータを構築し、単語や実験内でのセクション、話者、繰り返しの有無をもとに研究後にも識別可能な刺激のアーカイブ ID を作成した。次に、Praat スクリプトを用いて録音全体における刺激間の境界の配置や各刺激への ID の付与、録音音声全体の個別刺激ファイルへの分割を行うことで、分析対象に対するアノテーションの保存を半自動的に可能とした。本研究手順により、録音データベースを用いたより効率的な研究が可能となるだろう。また、処理したデータをアーカイブすることで将来の様々な研究に役立てることができるだろう。

1. はじめに

本稿では、国立国語研究所の共同利用型共同研究（登録型）で利用可能な豊富なデータベースを用いたデータ処理方法について紹介する。共同利用型共同研究（登録型）は、国立国語研究所が所有する多種多様な研究資料・言語資源・分析装置を活用して研究を行うことができる制度である。著者らの行なった研究では、「国立国語研究所研究資料室収蔵資料」（主に1950年代から2000年代初頭にかけて行われた250以上の研究プロジェクトの研究資料）にある、「fo0245：日本語音声における韻律的特徴の実態とその教育に関する総合的研究」において収集された録音音声データを用いて分析を行った。データ処理の過程で録音データに含まれる各刺激のアーカイブ ID を作成し、この ID をもとに各刺激を個別の音声及びテキストグリッドファイルへの分割を行ったため、本研究以降に同一データを活用して研究を実施する際に役立て、より効率的に研究を行うことが可能である。2022年8月現在においてもコロナ禍は依然として続いており、新たに実験を行うことが簡単ではないため、共同利用型共同研究（登録型）の利用や識別可能な ID とともに処理したデータのアーカイブを行うことで、このような社会情勢化においても研究を促進することが可能となると考えている。

† Email address: g239704k@iccu.ac.jp

2. 使用したデータベースについて

著者は、1990年代初頭に実施された「fo0245：日本語音声における韻律的特徴の実態とその教育に関する総合的研究」の実験録音データを使用して日本語の有声性の対立に関する研究を行なった。本節ではこのデータ全体に関する説明を行う。

2.1 被験者

被験者は日本の各地方（北海道・東北・関東・中部・近畿・中国・四国・九州・沖縄）に在住の日本語方言話者であり、年齢も小学生、中学生、若年層、壮年層、中年層、高年層（老年層）と多様な被験者による録音が存在する。

2.2 刺激

録音実験を行なったオリジナルの研究プロジェクトは日本語の韻律をトピックとしており、実験録音データは（1）に示したような様々な種類の刺激を含んでいた。

- (1) 刺激の種類
 - a. 五十音
 - b. 名詞
 - c. 動詞とその活用
 - d. 形容詞とその活用
 - e. 文章
 - f. 童話（桃太郎）の朗読
 - g. 数字（1から9までと四桁の数字）

これらの刺激は地域や被験者に応じて刺激やセクションの順番が異なっている場合があり、その差異は名詞の種類や名詞に格助詞「が」を付けた繰り返しの有無などに見られた。例を挙げると、後述するように東北方言と東京方言の録音実験は異なる刺激リストを用いて行われており、東北方言の録音データにおいては繰り返しがあったものの、東京方言の録音データには繰り返しはなかった。

2.3 実験手順

録音実験では、まず実験者が被験者に名前や生年月日の確認などを行い、例文を読み上げてもらい練習をして実験の手順について説明した後に、被験者が紙に印刷されている刺激を一つずつ読み上げていくという形で行われた。実験の途中で外を走る車の音などの外部音が入った場合や、録音が不明瞭な場合などは実験者の判断により被験者に再度の読み上げをお願いしていた。

ここまで第二節では利用した「fo0245 日本語音声における韻律的特徴の実態とその教育に関する総合的研究」のデータベース全体の説明を行なったが、次節ではこの中から著者の行った研究で焦点を当てたデータの処理方法について詳説する。

3. データ処理

本節ではデータの処理方法について順を追って詳しく説明する。著者は日本語方言間での有声性の対立に焦点を当てており、方言間での音声的な差異を研究するため、東北方言と東京方言を分析した。分析対象とした話者はデータベース上の「高齢層」であり、それぞれ

の方言から 10 名のデータに対して分析を行なった。被験者の男女比は、東北方言では女性のデータが存在しなかったため全員男性であったのに対し、東京方言では 5:5 であった。該当するデータの被験者全員分の年齢は正確には判明しなかったが、分析した録音実験は 1991 年に実施されており、少なくとも 20 名中 12 名が 1910 年代から 1930 年生まれのため、「高齢層」に分類されている被験者は実験実施時点で 60 歳以上であったと推察される。

また、本データベース上には録音リストが存在するものの、刺激のリストを見つけることができなかつたため、まず初めに実験音声をもとに刺激リストを作成した。刺激は前節で述べたように東北方言と東京方言間で名詞や文章が異なっており、東北方言の録音実験では、ある名詞を発話した後に格助詞「が」を付けた繰り返しが存在した。刺激のリストを作った後に、各刺激に対して「単語 ID」-「繰り返し番号」-「実験内のセクション ID」-「被験者 ID」となるように、本研究が終わった後においても識別可能なアーカイブ ID を作成した。被験者 ID はデータベース上で各音声ファイルに割り振られていた番号を使用した。例えば、「リスト上の一つ目の単語(W001)」で「セクション番号 S1」、「被験者 ID が 001(JPD001)」だった場合は「W001-S1-JPD001」となり、繰り返しがある場合には最後に「W001-1-S1-JPD001」のように繰り返し番号を付けた。単語 ID と繰り返し ID を頭の部分に配置した理由は、のちの処理で個別の音声ファイルへと分割される際に、フォルダ内で同じ刺激のファイルをまとめて表示させるためである。

録音音声データの処理は、Praat (Boersma and Weenink, 2022) を用いて行なった。本研究では名詞に焦点を当て、まず録音音声ファイルの中から(1b)や(1e)のセクションを切り取った。その後 Praat スクリプトを用いてセクション内の各刺激の境界を配置し、さらにその境界をもとにして各刺激に対して上で説明したアーカイブ ID を与えた。また、各刺激をアーカイブ ID がファイル名となる個別の音声ファイルへと分割し、保存した。ここまでの処理が終わった後、個別の音声ファイルに対して手動でアノテーションを行なった。これらの Praat 上での処理は、Praat スクリプトを用いることでアノテーションを除き半自動的に行なった。

アノテーションが完了した後は、個別の音声ファイルとアノテーションを行なったテキストグリッドファイルの入っているフォルダに対し Praat スクリプトを使用してアノテーションを行なった箇所の情報を抽出した。その後、抽出されてできたテキストファイルをエクセルから開き、統計分析ソフトウェアの R (R Core Team, 2021) を用いて分析できるように加工した後に csv ファイルとして保存した。

4. おわりに

本稿では国立国語研究所の「共同利用型共同研究（登録型）」を用いて実施した研究手法を説明した。今回の研究で実施した処理過程の中でも、アーカイブ ID の使用が重要な点だと考える。一つの研究の後でも識別可能なアーカイブ ID をそれぞれの刺激に対し与えることで、将来同じデータベースを用いて研究を行う際に役立てることができ、また先行研究の結果の再現性の確保にもつながるだろう。特に、現在の社会情勢を念頭に置くと、まだ新たな実験を対面で実施することは必ずしも簡単ではないため、データベースを活用した研究の重要性がさらに高まると考えられる。その際にオリジナルのデータベースに加え、処理されたデータが識別可能な ID とともにアーカイブされていたとしたら、仮に研究トピックが異なっていたとしても、将来行われる研究の大きな促進につながると考えている。

謝 辞

本稿は、国立国語研究所の共同利用型共同研究プロジェクト「日本語の有声性の対立への複数の音響指標の影響」（研究代表者：李勝勳）の研究成果である。また、第一著者は国際基督教大学より「国際基督教大学博士研究員（A種）」の助成を受けている。

参 考 文 献

- Boersma, P. and Weenink, D. (2022). “Praat: doing phonetics by computer [computer program] (version 6.2.14)”.
- R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

文化的・言語的相違による「ハミング・鼻歌」の初発語彙使用 — 日・中・英・米語のコーパスによる分析 —

浅野 恵子 (順天堂大学医学部一般教育外国語研究室) †

陳 森 (東京福祉大学 社会福祉学部)

First lexical usage of the words "hum" across languages and cultures — Analysis of a Japanese, Chinese, English, and American multilingual corpus —

Keiko Asano (Juntendo University, Faculty of Medicine, Liberal Arts division)

Sen Chen (Tokyo University of Social Welfare, School of Social Welfare)

要旨

同じ音声的及び音響的特徴をもちながら、文化や気候風土によって変化する音声行動があり、無意識に行われているものが少なくない。その一つとして、/m,n/などの有声鼻音の音声特徴は自然発話としては一般的であり、それをさらに上咽頭に響かせる音の「ハミング」がある。日本語では「鼻歌」と呼ばれている。他言語が理解できなくても音声行動としては個別言語の域を超えて普遍的に発せられる声音である。日常の発声時行動様式が文化的・言語別にどのように呼ばれているか、またいつから使われているかを日・中・英・米語の各言語のコーパスを比較し、初めて使用された時期や当時の意味などから推移を分析する。

1. はじめに

例えば、次の質問、「ハミングとは何ですか」や、「ハミングをしてください」といわれた人は誰でも、一度は聞いたか自分自身で行ったことのある行為で、意味や行動が不明であることは恐らくないであろう。それはどの言語においても、誰に聞いてもその語の意味がわからないという返答はないぐらい、「ハミング」は対話には使用しないが、自然発話では一般的な音声行動と言える。

しかしながら、この「ハミング」という行為はいつから行われているものであろうか。1888年以降、現時点まで版を重ねて音声言語は国際音声表記 (IPA) で表記することが可能である。しかしながら、この「ハミング」が実際の音声として記録する方法が確立していない時代からの行為であると仮定すると、歴史語彙コーパスを用いて、その語彙の使用された状況を分析する手法が効果的だと示唆される。

音声表記において、同じ音声的及び音響的特徴をもちながら、文化や気候風土によって変化するものがあり、音声行動としては無意識に行われているものが少なくない。その一つとして、/m,n/などの有声鼻音の音声特徴は自然発話としては一般的であり、それをさらに上咽頭に響かせる音の一つに「ハミング音」がある。日本語では「鼻歌」と呼ばれている。「ハミング」とは何か、なぜ、どのような状況で行うのか、人にとってハミングとは何かについての分析は進んでいないのが現状である。このハミング音は他言語が理解できなくても音声行動としては個別言語の域を超えて普遍的に発せられる発声音でもある。また、特定の人種や言語に関わらず、ポジティブな感情を伴う時に発することが多いとも思われる

† keasano@juntendo.ac.jp

Dunbar(2006)。この万国共通の音声言語であるハミング音及び行動環境がコーパスからの初発語の記載を調査することにより、音声としての記録に相当すると推測される。また、いつからこの行為が行われていたかを観察し、言語間の比較をすることを目的とする。

2. 各言語間における語彙「ハミング」について

2.1 各言語における使用コーパス種類

対象語彙である「ハミング」(各言語により名称が異なる)が、各言語コーパスによって、その語彙がいつ初めて使われたか(初発語)を調べるためのコーパスについて明記する。表1に各言語とその対象初発語彙及び検索コーパスサイト名を示す。初発語彙がいつであるかを調べるのが目的ではあるが、出現する度数調査についても、表記することとする。

まずは日本語の語彙「鼻歌」と「ハミング」については、初発語という観点からは、外来語としてカタカナ表記ではない行為との連動を調べたいため、前者「鼻歌」に対しての初発語として『日本語歴史コーパス(CHJ)』(国立国語研究所 2022)から調べる。しかしながら、外来語の借用からのカタカナ表記で使用された時期としての初発語にも興味を持ったためこちらも『日本語話し言葉コーパス (CSJ)』と『現代日本語書き言葉均衡コーパス(BCCWJ)』から調べた。

中国語の語彙については、浅野(2022)において、中国語で「ハミング」をどう呼ぶかというアンケート調査で、多くの呼び名が方言を含めて存在することが判明した。ここでは一般的に多く「ハミング」を示す内容としての語彙、「哼哼」に絞って調べることにした。その際に、『北京大学 CCL 语料库检索系统(网络版)』と『台湾 中央研究院の Sinica コーパス』から、初発語を検索した。

英語コーパスは English Copora.org 内の Online Etymology Dictionary、Early English Books Online(EEBO)版を、米語コーパスでは、English Copora.org 内の Corpus of Historical American English (COHA) を検索に使用した。英語の場合、同じ言語ではあるが、文化的、あるいは建国の年代が異なり、使用された書物や環境も異なるため英語と米語のそれぞれのコーパスから初発語彙使用がいつかを調べることにした。各関連 URL は最後に記載する。

表1 各言語と対象初発語彙及び検索コーパスサイト名

対象言語	言語別の初発対象語彙	検索コーパスサイト名
日本語	鼻歌 ハミング	国語研日本語ウェブ NWJC 『中納言』 『日本語歴史コーパス(CHJ)』 『日本語話し言葉コーパス (CSJ)』 『現代日本語書き言葉均衡コーパス(BCCWJ)』
中国語	哼哼	北京大学 CCL 语料库检索系统(网络版) 台湾 中央研究院の Sinica コーパス
米語	hum	English Copora.org Corpus of Historical American English (COHA)
英語	hum	English Copora.org Online Etymology Dictionary Early English Books Online(EEBO)

2.2 日本語の語彙、「鼻歌」、「ハミング」についての初発語

日本語の語彙として「ハミング」を「鼻歌」という場合も多い。「鼻歌」の初発語彙については、『日本語歴史コーパス(CHJ)』を中心に検索した。特に、外来語からの借用語であるカタカナ表記の「ハミング」については『日本語話し言葉コーパス (CSJ)』および、『現代日本語書き言葉均衡コーパス(BCCWJ)』を用いた。表2に「鼻歌」と「ハミング」の度

数を示す。

表2 日本語語彙「鼻歌」と「ハミング」の度数

	「鼻歌」	「ハミング」
CHJ	15	0
BCCWJ	115	60
CSJ	0	17

まずは、CHJによる実際に得られた用例の「鼻歌」の初発語彙について示す。

- (1) と言うたら、こなさんも尋常に死んでくださせと戸を引き立ててよりかかり、口に鼻歌、心には弥陀の名号一筋の、紙鳶の糸よりなほ細く切れかかりたる玉の
【出典】CHJ サンプルID：51-近松 1709_17002 近松門左衛門『心中刃は水の朔日』

初発語彙以外に、近松の作品には、再度この「鼻歌」という語彙が登場する作品がある。

- (2) 腹這ひにはつたとのめらせ腰骨を、七つ、八つ、うんといふほど踏みつけて鼻歌に懐手、吾妻、つきど、をかしさ堪へ笑ひを殺す笑止顔彦介やう / \ 起き上がり聞えた
【出典】CHJ サンプルID：51-近松 1718_10001 近松門左衛門『山崎与次兵衛寿の門松』

以下、代表的な文学作品においても出現している代表例を示す。

- (3) 廻りと生意気は七つ八つよりつので、やがては肩に置手ぬぐひ、鼻歌のそそり節、十五の少年がませかた恐ろし、学校の唱歌にもぎつちよんちよんと拍子
【出典】CHJ サンプルID：60N たけ 1895_11001 樋口一葉『たけくらべ』
- (4) は無いさ、ただ暫く塾するといふだけのことだらう須らく笑つて酒を汲んで鼻歌でも迂鳴るべしだ、さあ飲み玉へ、君ほどにも無い、其様な顔つきは止し
【出典】CHJ サンプルID：60M 太陽 1901_02022 幸田露伴『緑の糸』

次に、語彙「鼻歌」及び「ハミング」について、『現代日本語書き言葉均衡コーパス(BCCWJ)』の出典の初発語彙を見ていく。

- (5) それからせわしない手つきでアイスペールの氷をガチャガチャとグラスに放りこみ、何か鼻歌のようなものを口ずさみながら酒を注いだ。何か言いたいことがある
【出典】BCCWJ サンプルID：OB1X_00173 三田 誠広 『僕って何』1977年
- (6) 方面にバンを走らせながら、彼は『皇太子の初恋』の酒場の歌をハミングで歌った。「チャイニーズ？」 「うん？」 「ねえ、胡椒の
【出典】BCCWJ サンプルID：LBa9_00088 トマス・ペリー(著)飯島 宏(訳)『メッツガーの犬』1986年

「ハミング」という語彙は、多くは文学作品の翻訳として出現する機会が多いが、日本語では恐らく村上春樹の小説に初発が出現している。

- (7) か気づきもしないようだった。僕はトレーンのソロにあわせて小さな声でハミングしな

がら車を走らせた。湘南から夜に東京に帰ってくる道は

【出典】BCCWJ サンプルID: OB3X_00282 村上 春樹(著)『ダンス・ダンス・ダンス』
1988年

語彙「鼻歌」及び「ハミング」について、『日本語話し言葉コーパス (CSJ)』での初発も検索を試みた。

(8) 音で鼻音つまり唇で作った音で鼻から出す音とえば先、先程のハミングのように口を
で閉じて「ん」という風に鼻から出すというのが

【出典】CHJ 講演ID: M03F0019

(8) は、まさにハミング音の方法を説明している発話となっている点が興味深い。他のCHJに見られる発話は学会等での「ハミング音、音響的なハミング窓」の説明などが多く検索された。CHJでは、「鼻歌」についての度数は検出されなかった。

2.3 中国語の語彙、「哼哼」についての初発語

次に中国語のハミングに相当する語彙「哼哼」についての初発を検索した。上記で述べたように、まずは北京大学 CCL 語料庫検索システム(ネットワーク版)を使用した。「哼哼」以外に「ハミング」を表す語彙が存在し、それらの語彙も含めて度数は165であった(哼哼:145、哼哼唧唧:15、哼哼叭叭:2)。「哼哼」に似た語彙に「气哼哼」(度数:35)があるが、ぷんぷん怒ると解釈されるので、今回は省いてある。

(9) 王神姑走近前来听上一听, 只见人果真有个人在里头, 一会儿哼哼的哭, 一会儿又不哭, 一会儿骨弄的响, 一会儿又不响。

【出典】『三宝太监西洋记(二)』、罗懋登(著) 作品年 1597年

もう一つの中国語コーパス、『台湾 中央研究院の Sinica コーパス』において、検索を試みた。度数は2であったが、出典情報に不明確なものが多く提示不可能である。

2.3 英語の語彙、“hum”についての初発語

英語の場合、文化間の違いなども考慮に入れ、英語と米語の各コーパスから検索することにした(表1を参照)。両言語を調べるコーパスは English Corpora.org に集約されているものを用いた。両英語を検索したコーパス内での該当語彙の度数を表3に示す。

表3 英・米語語彙 “hum” の度数

	コーパス	“hum”
英国英語	EEBO	6(1480s)/1902(1470s~1902s)
米国英語	COHA	43(1820s)/ 4355(1820s~2010s)

2.3.1 英語(英国)の語彙、“hum”についての初発語

英語(英国)の語彙で使用したコーパスは Early English Books Online(EEBO)であり、語彙の意味語源を知るために Online Etymology Dictionary(Online)を用いた。このコーパスにおいて、初発に出現しているのは以下のとおりである。

(10) and his men were alle slayne that none of hem escaped and therfor is that water callyd hum bar and euermore shal be callyd thus for enchesen that kynge humbar was thereyn drenchyd .

【出典】 Title: The cronycles of Englonde, Author: Caxton, William, ca. 1422-1491., 作品 1482 年

上記の出現が初発ではあるが、Online Etymology Dictionary(OED) によると今回調査したい意味として出現している作品は以下の(11)の方が初発だと想定できる。以下 OED の hum についての定義である。

hum (v.):late 14c., hommen "make a murmuring sound to cover embarrassment," later hummen "to buzz, drone" (early 15c.), probably of imitative origin. Sense of "sing with closed lips" is first attested late 15c.; that of "be busy and active" is 1884, perhaps on analogy of a beehive. Related: Hummed.

hum (n.):mid-15c., "a murmuring sound made with the voice," from hum (v.).

(11) and thadmyrall commanded that he shold hastily be taken but whan the hors saw that they wold haue taken hum he ranne and smote out ; cessed not tyl he cam to the yate of the castel in whych the barons of frauce were enclosed

【出典】 Title: Thystorye and lyf of the noble and crysten prynce Charles the grete kynge of Frauce [sic], Author: Caxton, William, ca. 1422-1491., 作品 1485 年

2.3.2 英語（米国）の語彙、“hum” についての初発語

次に米語の初発についての検索結果を示す。米国における作品中の初発は比較的年代が浅いが、建国された時代との関連が想定される。

(12) Made their disapprobation known By many a murmur, hum and groan, That to his speech supplied the place Of counterpart in thorough bass. Thus bagpipes, while the tune they breathe, Still drone and grumble underneath;

【出典】 Title: The poetical works, Author: Trumbull, John, 1750-1831., 作品 1820 年

3. 考察

表 4 に本稿における各言語間の初発出現年と出現作品名をまとめる。

表 4 各言語間の初発出現年と出現作品名

言語	語彙	検索コーパス名	初発年	著者・作品名
日本語	鼻歌 ハミング	CHJ (鼻歌)	1709	近松門左衛門『心中刃は氷の朔日』
		BCCWJ (鼻歌)	1977	三田 誠広 『僕って何』
		BCCWJ(ハミング)	1986	トマス・ペリー(著)/飯島 宏(訳)『メツツガーの犬』
		CSJ(ハミング)		M03F0019 40 前半女性 (独話)
中国語	哼哼	北京大学 CCL 台湾 Sinica	1597 N/A	罗懋登 (著)『三宝太监西洋记 (二)』 N/A
米語	hum	COHA	1820	John Trumbull, <i>The poetical works</i>
英語	hum	EEBO	1485	Caxton, William, <i>Thystorye and lyf of the noble and crysten prynce Charles the grete kynge of Frauce</i>

4. おわりに

本稿では音声を録音する技術がない時代の発話行為である「ハミング」がいつから行われる自然発話であるかを知るために、歴史的足跡を追えるコーパスを中心に初発現語彙を調べた。「ハミング」という行動はどの言語話者にも特に説明をしなくても理解できる対話を伴わない音声言語である点から、日本語、中国語、英語、米語のコーパスから初発年代とその作品を調べてみた。今回は、初発を中心に焦点を当てて、その語彙を調べたため、品詞などの文法関連や語彙の意味合いの変化推移など、現代で使用されている意味との相互性についてはさらに精査は必要であろう。

コーパスに記述がみられる語彙以外に、どの言語でも地域性や方言で使われている呼称があると思われるため、今後は代替語彙としてどのような言い方があるかも、内容との関連からも調査を続けていく予定である。

また、無論、作品中に初発で出現することと、そのハミングを行うこと自体が同等にその行動の初めを意味するわけではない。また、初発語彙として作品をコーパスからは検索はできたものの、現在の意味合いと同じ内容であるか否かはさらに精査する必要がある。また、コーパスからの検索以外に「ハミング」という行動を文化や言語を超えて、人として、いつからどのような目的で行うようになったかを知る術を今後も模索していくための一歩として、今回の研究の目的はあるとも示唆される。

謝 辞

本研究において、中国語の知識提供者として、専修大学大学院文学研究科、大学院生・李丹さんのご助言を受けた。また、この研究は 2020, 2021 年度の順天堂大学医学部一般教育共同研究で採択され助成金を受けた。

文 献

- 浅野恵子(2021) 人はなぜハミングを行うのか ―日本語話者における使用状況のアンケート調査―, 日本音響学会秋季研究論文集 (DC-ROM) pp.823-824.
- 浅野恵子(2022) 文化的・言語的相違によるハミング行動の分析 ―日・中・英・米母語話者へのアンケート調査― 日本音響学会 2022 年度秋季研究論文集 (DC-ROM) To appear.
- 周 振, 吉本 啓一(2020) 統語・意味情報付き中国語コーパスの構築 ―統語解析の詳細について― 国際文化研究(オンライン版) 26, pp.89-104, 2020-03-31 URL <http://hdl.handle.net/10097/00127390>
- 国立国語研究所(編)(2022). 『日本語歴史コーパス』,(バージョン 2022.3,中納言バージョン 2.6.1) <https://chunagon.ninjal.ac.jp/> (2022 年 8 月 14 日確認).
- Dunbar, L. (2004) The Human Story. London: Faber and Faber. Humming makes you happy. London Zoo electronic Newsletter from 14th March.
- Jordania, J. (2006) Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech. Tbilisi State University, Logos.
- 「西遊記」オンライン版 https://www.jy135.com/guwen/183792/18.html?ivk_sa=1024320u(2022 年 8 月 16 日確認)

関連 URL

コーパス検索アプリケーション 『中納言』 <https://chunagon.ninjal.ac.jp/>

北京大学 CCL 语料库检索系统 (网络版) http://ccl.pku.edu.cn:8080/ccl_corpus/

台湾 中央研究院の Sinica コーパス <http://asbc.iis.sinica.edu.tw/>

English Corpora. org <https://www.english-corpora.org/>

Corpus of Historical American English (COHA) <https://www.english-corpora.org/coha/>

Online Etymology Dictionary <https://www.etymonline.com/>

Early English Books Online(EEBO) <https://www.english-corpora.org/eebo/>

作文評価における日本語教師の評価ポリシーの多様性 —ホリスティック評価の観点から—

安芝恩 (なし) †

Diversity of Japanese Language Teachers' Assessment Policies in Writing Assessment: From the Viewpoint of Holistic Scoring

Jieun Ahn (Not affiliated)

要旨

本研究は、日本語教師が日本語学習者の作文をどのように評価しているのかという評価者の評価ポリシーを明らかにすることを目的とするものである。特に、評価者一人ひとりにおける評価ポリシーは存在するのか、作文評価において評価ポリシーは一貫性が保たれているのかに注目した。日本語学習者 13 名による作文 (4 コマイラストの描写ストーリー文) を対象とし、日本語教師 12 名に Google フォームを用いた作文評価をしてもらった。評価方法については、全体的・総合的に評価してもらうため、ホリスティック評価を採用し調査を行った。その結果、日本語教師は、明示的な認識があるとは言えないものの日本語学習者の作文に対して各自の評価ポリシーを持っており日本語学習者の作文によってその評価ポリシーは個人の中でも変動することが明らかになった。評価者個人が持っている評価ポリシーの多様性を認めつつ、安定した評価への必要性があることが示唆された。

1. はじめに

1.1 研究背景

評価には、「評価のばらつき・ゆらぎ」(宇佐美 2014)がある,または、「評価の不一致をもたらす要因」(田中ほか 2009, 田中 2016)があるという「評価の多様性」が報告されている。さらに、「評価価値観の変容や自己調整,といったことも研究の対象として取り上げられなければならない」(宇佐美 2016)という指摘があるように,評価は、「変動するもの」としての評価価値観として捉えられているために,複数の評価者による評価には,評価の不一致が生まれるものである。

最近,複数の評価者による評価を一致するための試み(阿部ほか 2016)や日本語の学習者と教師のための Web システムである jWriter, 読み手と構成を意識した日本語ライティングである GoodWriting Rater といった評価ツールが開発されている。

しかし,日本語教育現場では,いまだ評価者による作文評価が行われており,日本語の添削に焦点が置かれている現状にある。そのため,作文の構成や内容よりは文法の正確性が問われる傾向がある。おそらく作文評価は添削という認識が評価ポリシーの根底に位置付けられているだろう。

そこで,本研究では,日本語教師は,日本語学習者の作文をどのように評価しているのかという評価者の評価ポリシーを明らかにすることを目的としている。本研究の調査結果から,

† parkhc2015@gmail.com

評価者個人が持っている評価ポリシーの多様性を認めつつ、その評価ポリシーにおける「安定した評価」への必要性が示唆されることに意義があると思われる。

1.2 用語定義

では、本研究で用いる「評価ポリシー」および「作文」の定義について説明する。「評価ポリシー」は、宇佐美(2014)での「評価」の用語定義を援用し、「主体がもつ内的・暗黙的な価値観に基づいて、対象についての情報を収集し、主体なりの解釈を行ったうえで、価値判断を行うまでの一連の認知プロセス。またその結果として得られる判断」と定義する。さらに、宇佐美(2016)で示されているように、「価値観に基づく判断」であり、価値観の相違によって判断も、それに先立つ情報収集や解釈のありようも変わってくる」と捉える。「作文」については、日本語学習者が4コマイラスト(安 2017)を見て各コマの内容について描写したストーリー文を「作文」と定義する。日本語学習者から収集された作文のスタイルは、登場人物の会話文に注目した「会話型」と叙述形式の「説明型」が混在している。

2. 先行研究

2.1 先行研究の概観

では、「評価者による評価」および「評価ツールによる評価」についての従来の研究について述べる。

評価ポリシー(例:ループリック評価)における理解が、評価者それぞれ異なる解釈をしている(藤浦ほか 2018)ことが報告されている。田中(2016)によると、各評価者におけるポリシーについては、「評価基準やガイドライン等に柔軟に対応する人となかなか自身の評価ポリシーを譲らない人がいる」と指摘されている。一方、宇佐美(2016)では、「評価」に対する明示的な認識の不在が指摘されている。これは、評価基準が設けられていても各評価者における評価ポリシーが存在している(田中ほか 2009)ことが影響されていると考えられる。研究は背景でも述べたように、「評価のばらつき・ゆらぎ」(宇佐美 2014)は、「評価の不一致をもたらす要因」(田中ほか 2009, 田中 2016)である、評価ポリシーに対する明示的な不在から起因している可能性がある。

近年は、「評価者による評価」から発生する「ばらつき」「ゆらぎ」といったものを減らすため、「GoodWriting Rater」や「jWriter」のような評価ツールが開発されている。これらの評価ツールは、それぞれの評価ポリシーに基づき評価していることが特徴である。

「GoodWriting Rater」は、ホリスティック評価とマルチプルトレイト評価によって評価されるツールであり、矢部ほか(2019)によると、「人間」と「機械」による評価の統合的活用」の一部として開発された評価ツールであると言及されている。文字数400~1600字の作文が評価の対象となっている。また、「jWriter」は、総合的評価によって評価されるツールであり、「日本語学習者の作文テキストを入力すると、推測される作文力の到達レベルを5段階で判定」できる。最小入力文字数は300文字で、600字前後でもっとも良い判定ができると言及されている。このような総合的評価について、占部(2007)では、妥当性があるとは言えないものの実用性は高いという利点があると指摘されている。一方、評価者による評価のばらつき・ゆらぎを防げるものの、より良い判定のための文字数制限があるため、全ての日本語教育現場での作文評価に活用できるとは限らない。

以上を踏まえ、日本語教育現場では、「評価者による評価」が求められている現状で、評価者一人ひとりにおける評価ポリシーがあるのか、その評価ポリシーは作文評価において評価の

対象が変わっても一貫性が保たれているのかといったことについてはまだ明らかにされていない。

2.2 研究課題

そのため、以下の3つの研究課題を設定し調査を行った。

【課題 1】日本語学習者の作文に対する日本語教師の評価には、評価者それぞれの間にどのような違いが見られるのか

【課題 2】日本語教師は、日本語学習者の作文に対する評価において、どのような評価ポリシーを持っているのか

【課題 3】評価ポリシーは、評価者個人の中で一貫性が保たれているのか

課題 1 は、評価者による評価の不一致といった評価ポリシーの多様性を見るために、課題 2 は、評価者個人における明示的な評価ポリシーの存在有無を確認するために、課題 3 は、評価者個人において評価ポリシーが存在する場合、評価対象の作文すべてにおいて、その評価ポリシーは一貫性のある働きをするのかを確認するために設定したものである。

3. 研究方法

3.1 コーパス

本研究に用いられたコーパスは、現在構築中の「日本語学習者 13 名による作文 (4 コマイラストの描写ストーリー文)」である。4 コマイラスト(安 2017)を用いて、日本語学習者に描写ストーリー文を作成してもらったものである。次の図 1 のように、日本語学習者から収集された作文のスタイルは、登場人物の会話文に注目した「会話型」と叙述形式の「説明型」が混在している。

A さん：お腹がすいたなあ。あ、前のコンビニへ買いに行きます。
 スタッフ：いらっしゃいませ。
 A さん：食べ物を買いました。これください
 スタッフ：はい、わかりました
 A さん：すみません。現金がないので、カードではらえますか。
 スタッフ：すみませんが、ここで現金だけです。
 A さん：わかりました、待ってください、ATMで現金を出してはらいます。
 スタッフ：はい、わかりました。

・「会話型」作文の例

会社から家に帰る途中で、急に何か
 甘い物を食べたくなった私は汐井公園の近くにあるコンビニに寄りました。コンビニに入って、色々なお菓子があって、みんなは美味しそうだった。どれにしようかなあと迷っちゃた。
 やっぱ私の大好物のどら焼きが一番いいと思って、三つのどら焼きにしました。
 会計するとき、財布が家に忘れてしまったことを気がついた。
 どうしようかなあ、とても恥ずかしくて、顔が真っ赤になった
 仕方がない、ATMからお金を引き出そうと思った私は引き出して、払いました。

・「説明型」作文の例

図 1 日本語学習者 13 名による作文の例

また、コーパスの構築は、表 1 の日本語学習者 13 名の協力を得て行われた。協力した日本語学習者の国籍は、表 1 に示した通りでベトナム・中国・マレーシア・ミャンマー・インドネシアである。日本語の関連資格 (JLPT, J-TEST)については、いずれかの資格を保持してい

る日本語学習者が10名、いずれの資格も保持していない日本語学習者が3人である。

表 1 日本語学習者の属性

No.	性別	国籍	年齢	来日時期	日本語学習機関	JLPT	J.TEST
作文 1	男	ベトナム	20	2020年3月	日本語学校と独学	N4	ない
作文 2	男	中国	24	2019年1月	独学	ない	A-C レベル
作文 3	女	中国	33	2020年11月	日本語学校	ない	D-E レベル
作文 4	女	中国	33	2020年12月	日本語学校	ない	ない
作文 5	男	ベトナム	30	2013年7月	日本語センター	N1	A-C レベル
作文 6	男	ベトナム	20	2020年11月	日本語学校	ない	A-C レベル
作文 7	女	中国	25	2020年11月	日本語学校	N5	ない
作文 8	男	マレーシア	24	2020年10月	日本語学校	N4	ない
作文 9	未回答	ミャンマー	24	2020年10月	日本語学校	ない	ない
作文 10	男	インドネシア	18	2020年11月	日本語学校	N5	ない
作文 11	女	ベトナム	26	2019年	大学	N1	ない
作文 12	男	ベトナム	20	2021年1月	日本語学校	N5	ない
作文 13	女	ベトナム	29	2021年1月	日本語教室	ない	ない

3.2 調査対象および調査期間

調査対象者は、日本語教師12名であり、属性については表2に示した通りである。調査は、2021年3月～4月に日本語教師12名に、3.1で紹介した「日本語学習者13名による作文（4コマイラストの描写ストーリー文）」の13作文の評価を依頼し行われた。日本語教師（日本語教師4を除く）は、日本語専攻/副専攻・420時間養成課程修了、日本語教育能力検定試験合格等の日本語教育関連資格を保持している。

日本語学習者のレベルが作文評価に影響されることを排除するため、3.1の表1日本語学習者の属性については、伏せた状態で日本語学習者の作文のみで評価してもらった。

表 2 調査対象者の属性

No.	指導経験	指導期間	指導機関	日本語教育関連資格	性別	年齢
教師 1	あり	4年～5年	日本語学校	大学副専攻	女	20代後半
教師 2	あり	11年	日本語学校	420時間養成課程修了、日本語教育能力検定試験合格	女	50代半ば
教師 3	あり	3年	高校、日本語教室	中等学校の教員資格（日本語、英語）	女	40代前半
教師 4	あり	10年	日本語学校	大学院専攻、日本語教育能力検定試験合格	男	30代後半
教師 5	あり	1年	大学	なし	男	30代後半
教師 6	あり	6～7年	日本語学校	420時間養成課程修了	女	30代後半
教師 7	あり	1年6ヶ月	日本語学校	420時間養成課程修了	男	40代半ば
教師 8	あり	3年	日本語学習塾、大学（ア	420時間養成課程修了、	女	20代後半

			シスタント),ボランティアセンター	大学院専攻		
教師 9	あり	18年	日本語学校,短期大学	420時間養成課程修了,日本語教育能力検定試験合格,大学院専攻	女	50代
教師 10	あり	7年	日本語学校	420時間養成課程修了	女	30代後半
教師 11	あり	10年以上	日本語学校,技能実習,専門学校	大学副専攻,日本語教育能力検定試験合格	女	30代後半
教師 12	あり	29年	日本語学校,大学,企業	420時間養成課程,日本語教育能力検定試験合格	女	50代後半

3.3 調査および分析方法

本研究の調査は,日本語教師 12 名に「日本語学習者 13 名による作文 (4 コマイラストの描写ストーリー文)」の 13 作文を評価してもらい,その結果は Google フォームを用いて収集した。ホリスティック評価の観点から,図 2 のように総合的に評価してもらい,気づいた点についてのコメントの記入に加え,5 段階評価 (わるい 1 ~ よい 5) のスコアをつけてもらった。作文に対するコメントについては, KH-Coder3(樋口 2020)を用いた計量テキスト分析を,評価のスコアについては,Excel のピボットテーブルを用いて分析を行った。図 2 に,作文 6 に対する日本語教師 4 のコメントを例で示してある。

本研究で採用しているホリスティック評価は,一般的に全体的・総合的評価 (holistic scoring) と知られており,占部 (2007)によると,「評価者個人の判断をもとにして全体評価として 1 つのスコアをつける方法」であると定義されている。また,「代表的なものとして, TOEFL の writing section での評価があり, 6 段階の全体的評価になっている」ことが紹介されている。

文のスタイルがバラバラ、長さは十分、語彙力は初級後半位、状況の理解はできていると思う。イラストの状況描写よりも、自分の日本語力を最大限に表出したい学生ではないだろうか。評価を4にしたいと思ったが、イラストの状況描写と関係のない部分を除外して考えてみると作文のレベルが高いとは思えない。

上記の作文の評価をお願いします。*

わるい 1 2 3 4 5 いい

○ ○ ● ○ ○

図 2 作文の 5 段階評価

4. 調査結果および考察

4.1 評価ポリシーのカテゴリー

まず,日本語学習者の作文に対する日本語教師 12 名のコメントの傾向を見るために, KH-Coder3 で分析を行った。その結果,図 3 のように,【作文,状況,描写,ストーリー,印象,コマ,話】という 7 つのカテゴリーが抽出された。日本語教師による作文評価において,本研究での作文が,4 コマイラストの描写ストーリー文であるという共通認識が働いて図 3 のようなカテゴリーが抽出されたと考えられる。

樋口(2020)によると,出現頻度の多い語は,大きく描画されている語であり,共起の程度は,太い線で描画されており共起の程度が強いことを意味している。

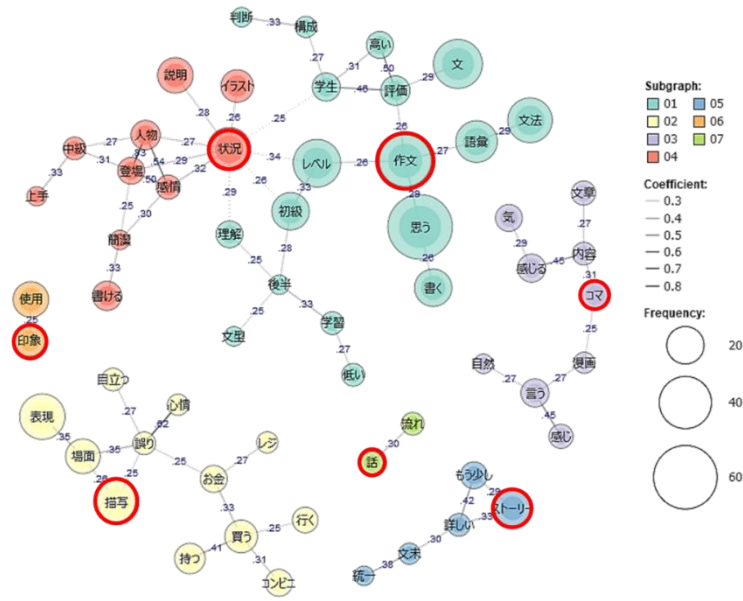


図 3 作文評価コメントの共起ネットワーク図

作文ごとのコメントの抽出語を表 3 に示した。作文それぞれに対する評価ポリシーは異なっているものの、日本語教師による評価コメント全体の結果から抽出された【作文,状況,描写,ストーリー,印象,コマ,話】の7つのカテゴリーがメインとなっていることが見て取れる。また,紙面の都合上,各作文に対するコメントの共起ネットワーク図は割愛するが,作文それぞれにおいて日本語教師の評価ポリシーは一貫性が保たれているとは言えない結果となった。

表 3 作文に対するコメントの抽出語

No.	抽出語
作文 1	レベル、作文、表現、ストーリー、過去、書く等
作文 2	語彙、初級、コマ、文法、簡潔等
作文 3	状況、描写、日本語等
作文 4	会話、勘違い、普段、ストーリー等
作文 5	作文、レベル、初級、状況等
作文 6	状況、イラスト、統一、ストーリー、使い方等
作文 7	文、会話、ストーリー、イラスト
作文 8	作文、文法、文型、描写、場面、使い方等
作文 9	会話、文、コマ、説明等
作文 10	間違い、文法、状況、レベル、描写等
作文 11	会話、文法、描写、展開、場面、使い方等
作文 12	会話、状況、敬語、描写、文等
作文 13	作文、文法、説明、状況、話等

ここで、注目したい箇所は、赤い字で示した「会話」であるが、日本語教師のコメントのうち、「会話」についてのコメントが24件見られた。「会話」につながるコメントは、ポジティブな評価とネガティブな評価に分かれていることが分かった。おそらく「作文」に対する日本語教師の用語定義がそれぞれ異なっており、「会話型」の作文を柔軟に受け入れないことが影響されていると言える。日本語学習者の作文が、「会話型」か「説明型」かが、評価ポリシーに影響され、スコア（5段階評価）の判定に影響を与えており、評価者の間に評価ポリシーのズレが存在（藤浦ほか 2018）していると考えられる。おそらく、「会話型」の作文スタイルをどのくらい容認してくれるかがスコア判定の判断の材料となっていると窺える。一方、イラストを、描写前提での作文という共通認識から、【状況、描写、ストーリー、コマ】のような語が抽出されたと考えられる。

4.2 評価ポリシーの多様性

では、4.2では、【課題1】日本語学習者の作文に対する日本語教師の評価には、評価者それぞれの間にどのような違いが見られるのかについて見ていきたい。

評価者による評価の不一致といった評価ポリシーの多様性を見るために、日本語教師による作文評価のスコアをまとめた。日本語教師1~12による「日本語学習者13名による作文（4コマイラストの描写ストーリー文）」の13作文に対するスコア（5段階評価）を表4に示した。同じ作文に対して日本語教師の評価スコアは、同じ場合もあればそうでない場合もあることが確認できた。

表4 日本語教師による作文評価のスコア

No.	教師1	教師2	教師3	教師4	教師5	教師6	教師7	教師8	教師9	教師10	教師11	教師12
作文1	3	3	2	3	2	3	3	4	3	3	3	4
作文2	4	2	2	2	3	2	3	4	3	3	2	3
作文3	1	2	2	2	2	2	4	4	3	3	2	3
作文4	4	1	2	2	2	1	3	2	4	3	3	3
作文5	2	2	1	1	3	2	3	2	3	3	1	2
作文6	4	5	2	3	4	4	2	4	5	4	4	4
作文7	1	1	1	1	1	1	2	1	2	3	1	1
作文8	4	3	2	4	2	4	3	3	3	3	1	3
作文9	5	2	2	2	3	1	4	4	5	3	3	3
作文10	2	3	2	3	2	3	3	3	4	3	3	3
作文11	4	3	3	2	3	1	3	3	5	4	4	5
作文12	4	2	2	3	4	1	3	4	4	3	3	3

また、同じ作文を提示して評価を依頼したものの、評価ポリシーにおける理解が評価者それぞれ異なる解釈をしている（藤浦ほか 2018）可能性が窺えた。図4は、表4に基づき作文の評価スコア（5段階評価）の揺れ幅をグラフで示したものである。作文10に対しては、スコア（5段階評価）2~4に分布しており揺れ幅3を見せているものの、スコア（5段階評価）3をつけた日本語教師が8名ということからある程度日本語教師による評価ポリシーが一致している可能性が見受けられる結果となった。

一方,作文 8 に対しては,スコア (5 段階評価) 1~4 に分布しており揺れ幅 4 を見せている。さらに,作文 9 に対しては,スコア (5 段階評価) 1~5 に分布しており,評価ポリシーの不一致が最も目立つ揺れ幅 5 を見せている。

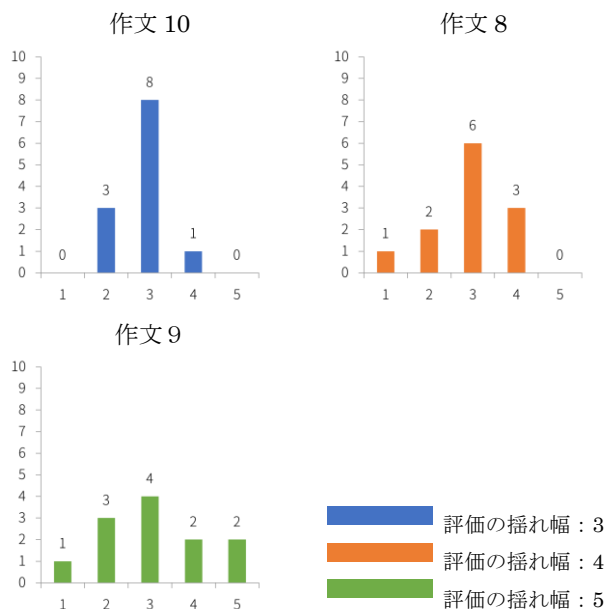


図 4 作文評価スコアの揺れ幅

特に,4.1 と 4.2 の調査結果の評価スコアとコメントを比較した結果,スコアの判定が高評価の場合は, [ストーリー,説明,状況] 等のストーリーの拡張に焦点が置かれている一方,低評価の場合は, [文法,語彙,初級,レベル] 等のような日本語の間違いに焦点が置かれており正確性が問われていることが窺えた。日本語学習者の日本語レベル等の属性については,日本語教師 12 名に伏せた状態で調査を行ったが,作文を評価する際は,日本語学習者の作文から改めてレベルを類推し作文の評価を行っていることが特徴であると言える。

4.3 評価ポリシーの明示的な認識

続いて,4.3 では【課題 2】日本語教師は,日本語学習者の作文に対する評価において,どのような評価ポリシーを持っているのかについて見ていきたい。評価スコア (5 段階評価) と評価コメントとの比較から分析を行った。

評価ポリシーがあると考えられる調査対象者は,日本語教師 4・10・11 であり,日本語教師 4 の場合は, [日本語学習者のレベルの想定,状況説明,文法] に焦点が置かれていることが分かった。一方,日本語教師 10 の場合は [ストーリー,使い方] に,日本語教師 11 の場合は [文法,話の流れ,日本語学習者の書いた文の意図] に注目していることが確認された。

詳しくは,表 5 の日本語教師 4 のコメントを参照されたいが,日本語教師 4 のコメントから,作文評価において事前に評価ポリシーが提示されていなくても作文評価に対する個人の評価ポリシーを持っていることが分かった。また, [日本語学習者のレベルの想定,状況説明,文法] のように具体的な言及をしていることから,ある程度安定した個人の評価ポリシーに基づき評価が行われたと考えられる。

表 5 日本語教師 4 のコメント

No.	コメント (※原文ママ)
作文 1	レベルは初級後半、作文の構成は文同士のつながりがよくはない、文法の間違いや語彙の特徴から中国の学生なのかなと思う、状況理解して説明していると思う。
作文 2	初級であって作文の構成を判断できるほどの長さではなく、状況理解しているのかどうか判断できるほどの語彙が見られない、4 コマ漫画の各コマを説明すると言うより 4 コマ漫画全体の流れを要約したような感じ。使える語彙だけを使ってなるべく安全に（間違いを避けて）作文を済ませた感じ。
作文 3	文法や状況描写においては初級後半だろうと思うが、三つの文のスタイルがそれぞれで一致しないのは、文型や語彙に対する理解度の低さが表れているのだろうと思う
作文 4	4 コマ漫画と言うこともあり、セリフを書いている感じ。状況を文章で説明することの難しさから、会話文の羅列に逃げている可能性もあると考える。作文と言うより普通の会話活動の文字化に近い。正確でない表記は普通の会話の中での耳コピーに原因があるのではないかと思う。
作文 5	状況理解よりも自分が正確に使える自信がある語彙の使用を優先した作文だと思う。手書きの作文ではない分、漢字の変換ミスをどのように捉えるべきなのかは迷う（作文のレベルの判断基準にすることを迷う）。
作文 6	文のスタイルがバラバラ、長さは十分、語彙力は初級後半位、状況の理解はできていると思う。イラストの状況描写よりも、自分の日本語力を最大限に表出したい学生ではないだろうか。評価を 4 にしたいと思ったが、イラストの状況描写と関係のない部分を除外して考えてみると作文のレベルが高いとは思えない。
作文 7	初級前半の学生と会話をしているような作文。文の長さが足りないため作文の構成、文法、語彙等については判断できる素材がない。ただこれが全力の作文であると仮定して判断すると評価は 1。
作文 8	初級後半以降の学習者が、新しく習得した文型を無理に使った感じがする。しかしそこを間違っているのが残念。流れが自然ではないが「そして」を使って文を繋げようとしているところは、これが作文であると言うことを意識している表れだと思う。文法的な間違いさえなければより評価は高かったと思う。
作文 9	4 コマ漫画を見て会話文を書くのは自然な流れかもしれないが、やはり作文の評価としては、叙述文を書くことの難解さから学生が逃げているのではないかと言う疑念を持ってしまふ。文法や語彙だけを考えると初級前半の会話文レベルだと思う。
作文 10	状況描写・作文と言う意識がしっかりある学生だと思う。漢字の変換については手書きではないので考慮しないことにしたい。学生が自分が使える語彙や文法を優先していると言うよりはイラストの内容やその状況描写を優先して、それらを説明するために頑張っている様子がうかがえる。文法的な間違いや語彙のレベルの問題から評価はそこまで高くはならない。
作文 11	かっこつけている。バイト先・日本での生活で耳コピーした日本語を自慢している。その割に文法的正確さや語彙のレベルはそこまで高くない。会話文であることを除いても、作文としての評価は 1・2 レベル。
作文 12	状況描写は少なく、イラストの状況でよく使われる会話文をつないでいる感じ。ただ他の学生より会話文の正確性は高い。敬語もちゃんと使っている（状況にあってはいるが）。語彙や文法の理解度は高いが、その運用能力がまだ備わっていない学生ではないかと思う。初歩的な間違いが少ない分評価を高くしたいとも思ったが、状況を描写した作文になっているのかと思うとそうではないので評価を低くした。
作文 13	状況描写ではなく日記のように作文をしているが、スタイルが首尾一貫しているなのでそこを評価したい。作文の構成としては文全体がまとまりがあって、ちゃんと作文として成立し

ていると思う。完璧な作文ではないが他の学生と相対的に高いレベルにあると思う。

一方、上述した日本語教師 4・10・11 以外の日本語教師のコメントは、気づいた点に焦点が置かれているコメントとなっており、個人における評価ポリシーが存在するかどうかも確認ができなかった。そのため、個人における評価ポリシーがあるとは言えない結果となっている。

4.4 一貫性が問われる評価ポリシー

4.3 では、個人における評価ポリシーがあるケースとあるとは言えないケースが確認されたが、4.4 ではより詳しく、【課題3】評価ポリシーは、評価者個人の中で一貫性が保たれているのかについて考察する。

表6は、日本語教師1がスコア（5段階評価）4判定をした作文に対する評価コメントである。同じ4というスコアではあるが、「描写できていないコマがある。文の作り方は良いが、レベルはそれほど高くない。助詞の使い方が間違っているのが残念。」というネガティブな評価をしている。一方、「ことばは間違えているが、きちんと描写できている。耳で聞いたことばをそのまま覚えているため、正しい表記ができていない。レベルは高め。」とポジティブな評価をしている。

表6 同じスコアに対する日本語教師1の評価コメント

スコア	コメント（※原文ママ）
4	描写できていないコマがある。文の作り方は良いが、レベルはそれほど高くない。助詞の使い方が間違っているのが残念。
4	ことばは間違えているが、きちんと描写できている。耳で聞いたことばをそのまま覚えているため、正しい表記ができていない。レベルは高め。
4	多くのことばを知っていて、ある程度使えるのでレベルは高め。会話が得意な印象。主語が私になっているのが気になる！
4	きちんと描写できている。不自然な部分はあるが、少し練習すればもっと良い文が書けそう。
4	日本での生活が1年以上？ 日常良く使う言い回しを知っている。
4	敬語がきちんと入っている。 アルバイトでレジをしたことがあるかも。きちんと描写できている。
4	文をきちんと組み立てることができている。 分かりやすい。

このことから、日本語教師1は、個人における評価ポリシーがあるとは言えないと考えられる。つまり、個人における評価ポリシーが確立されていないため、スコアの判定においても揺れが生じていると判断できる。おそらく同じ4というスコアではあるものの、その都度気づいた点に焦点を当て判断したため、評価コメントにも揺れが見られたと思われる。ほかの日本語教師の評価においても、このような傾向が見られたことから、評価ポリシーは個人の中でも評価対象の作文が変わることによって、個人における評価ポリシーも変動していく（一貫性が保たれているとは言えない）ものであることが分かった。

5. おわりに

以上の調査結果から、評価者である日本語教師に作文評価における明示的な評価ポリシーへの認識があるとは言えないものの、日本語学習者の作文に対する認識していない何らかの各自の評価ポリシーを持っており(宇佐美 2016)、日本語学習者の作文によってその評価ポリシーは個人の中でも変動していく可能性があることが明らかになった。本研究は、日本語教育に携わっている者に日本語学習者の作文評価に対する評価において、評価者個人が持っている評価ポリシーの多様性を認めつつ、その評価ポリシーにおける「安定した評価」(宇佐美 2016)への認識を提示し、その必要性が示唆されたと考える。評価ポリシーが提示されない場合のある日本語教育現場でも、評価者による「安定した評価ポリシー」に基づき「安定した評価」が行われることにより、妥当性のある作文評価につながるものと期待される。

今後は、日本語教師からの意見を踏まえ Google フォームの評価シートを改善していきたい。特に、調査対象者からの意見の中で「普段は日本語学習者の作文用紙をめくりながら評価(添削)していたが、それができなかったため、評価に一貫性が保たれていない」や「評価基準がないから評価しにくい」といった意見があったため、より明確な指示の必要性が確認できた。また、Google フォームによる作文評価に加え、フォローアップ調査を実施し、評価ポリシーの変動についての実証研究を行いたい。

文 献

- 阿部新・田中真理(2016). 「グループによるライティング評価における個人評価点の統一パターン」『「評価」を持って街に出よう―「教えたこと・学んだことの評価」という発想を超えて―』宇佐美洋(編), くろしお出版.
- 安芝恩(2017). 「第二言語環境における中級及び上級日本語学習者の言い換えストラテジーについて」九州大学大学院地球社会統合科学府地球社会統合科学専攻修士学位論文.
- 宇佐美洋(2014). 『「非母語話者の日本語」は、どのように評価されているか―評価プロセスの多様性をとらえることの慈義―』ココ出版.
- 宇佐美洋(2016). 「人間探求のための「評価」、という新しい視点」『「評価」を持って街に出よう―「教えたこと・学んだことの評価」という発想を超えて―』宇佐美洋(編), くろしお出版.
- 占部昌蔵(2007). 「項目応答理論を応用した英作文評価者トレーニングの有効性について」『STEP BULLETIN』19回, pp.14-22.
- 田中真理・長阪朱美(2009). 「ライティング評価の一致はなぜ難しいか:人間の介在するアセスメント(<特集>言語・コミュニケーションの学習・教育と社会言語科学-人間・文化・社会をキーワードとして-)」『社会言語科学』12巻1号, pp.108-121.
- 田中真理(2016). 「パフォーマンス評価はなぜばらつくのか?―アカデミック・ライティング評価における評価者の「型」―」『「評価」を持って街に出よう―「教えたこと・学んだことの評価」という発想を超えて―』宇佐美洋(編), くろしお出版.
- 樋口耕一(2020). 『社会調査のための計量テキスト分析―内容分析の継承と発展を目指して―第2版』ナカニシヤ出版.
- 藤浦五月・宇野聖子・小針奈津美・坂井菜緒・柴田幸子・服部真子・中川純子・長松谷有紀(2018). 「初年次レポート・ライティング教育における共通ルーブリック使用時の評価のズレに関する研究」『武蔵野大学しあわせ研究所紀要』1号, pp.3-24.
- 矢部玲子・桐山聡(2019). 「「人間」と「機械」による評価の統合的活用に関する検証―「GoodWriting Rater」を用いた留学生対象の日本語文章作成指導に基づいて―」『情報教育』1巻, pp.1-7.

関連 URL

jReadability PORTAL 日本語の学習者と教師のための Web システム <https://jreadability.net/>
GoodWriting.jp 読み手と構成を意識した日本語ライティング <https://goodwriting.jp/wp/>

『子ども版日本語日常会話コーパス』の構築

小磯花絵*・天谷晴香・居關友里子・白田泰如・柏野和佳子

川端良子・田中弥生 (国立国語研究所)

藤越 (東京大学大学院総合文化研究科/国立国語研究所)・西川賢哉 (国立国語研究所)

Construction of CEJC-Child

Hanae Koiso, Haruka Amatani, Yuriko Iseki, Yasuyuki Usuda, Wakako Kashino

Yoshiko Kawabata, Yayoi Tanaka (NINJAL)

Yue Teng (The University of Tokyo / NINJAL), Ken'ya Nishikawa (NINJAL)

要旨

2022年3月に公開した『日本語日常会話コーパス』(CEJC)は、成人中心のコーパスであり、未成年者、とくに10歳未満の子どもの会話はあまり含まれていないという問題がある。そこで国立国語研究所共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」(2022~2027年度)では、子どもを中心とする多様な場面・相手との会話を含む映像付きコーパスを新たに開発し、成人中心のCEJCと接続させることにより、コミュニケーションを含む言語の発達・変化の過程を、子どもから高齢者まで多世代に渡り実証的に研究できる基盤を構築することを目指している。発表では、新たに構築する子ども版の日常会話コーパスの設計や収録状況について報告する。

1. はじめに

これまで乳幼児を含む子どもの言語発達に関する研究が数多く行われてきたが、発達研究は乳幼児に限られるものではなく、学童期、青年期、成人初期、壮年期、老年期など、多世代に渡り見ていく視点も不可欠である。国立国語研究所共同研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」(2016~2021年度)では、多様な話者による多様な場面の日常会話200時間をバランスよく集めた『日本語日常会話コーパス』(CEJC)を開発し2022年3月に一般公開したが、このコーパスは成人の調査協力者を中心に会話を収集したため、未成年者、特に10歳未満の子どもの数がかなり少ないという問題がある。乳幼児から高齢者までの多世代を対象とするコーパス言語学的手法に基づく話し言葉研究を広く推進するには、子どもを対象とする会話コーパスが不可欠である。

子どものデータの記録と共有化については、これまでCHILDES⁽¹⁾などを中心に積極的に進められてきた(宮田2004)。また幼児とその両親の自然発話500時間以上を集めた大規模な『NTT乳幼児音声データベース』も2008年に公開された⁽²⁾。

乳幼児の言語発達において養育者、特に母親の子どもに対する影響が強いことから、既存の

* koiso@ninjal.ac.jp

(1) <https://childes.talkbank.org/>

(2) <http://research.nii.ac.jp/src/INFANT.html>

コーパス・データベースでは家庭での母子間会話が対象とされることが多かった。しかし成長するにつれ、友達や幼稚園の先生との会話といったように、多様な場面・相手との会話が言語の発達に深く影響するようになる。そのため、子どもの成長とともに広がる多様な場面・相手との会話を含むコーパスの構築が求められている。

こうした状況を受け、国立国語研究所共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」（2022～2027年度）では、子どもを中心とする映像付き会話コーパスを新たに開発し、成人中心のCEJCと接続させることにより、コミュニケーションを含む言語の発達・変化の過程を、子どもから高齢者まで多世代に渡り実証的に研究できる基盤を構築することを目指して活動している。

本プロジェクトで構築するコーパスは幾つかあるが、その中で最も力を入れているのは、子どもを中心とする多様な場面・相手との日常会話を経年的に収録した『子ども版日本語日常会話コーパス（CEJC-Child）』（仮）である。そのために、2019年度より収録の準備を進めてきた。本発表では、コーパスの設計や収録状況について報告する。

2. コーパスの基本設計

2.1 設計方針

成人中心のCEJCと合わせて研究できるようにするために、設計方針もCEJCを踏襲し、次の方針のもとでコーパスを構築することとした。

1. 収録のために集められた状況での会話ではなく、日常場面の中で自然に生じる会話を対象とする。
2. 子どもの成長とともに広がる多様な場面における多様な話者との会話をできるだけ収録する。
3. 音声データだけでなく映像データも記録・公開する。

このような方針のもと、2019年度より順次、会話収録を進めてきたが、新型コロナウイルス感染拡大の影響により、2020年以降、対象とする場面や話者に偏りが見られるようになった。これについては3節で状況を報告する。

2.2 会話の収録

■収録法 日常場面で自然に生じる会話を収録するために、CEJCの個人密着法に準拠した収録法を採用した。具体的には、3節で述べる調査協力世帯に収録機材等一式を貸し出し、調査対象とする子どもを中心とする多様な場面・話者との会話を収録してもらうこととした。自然な会話を記録するため、調査者が収録に介入しないという点もCEJCと同じである。協力世帯には、できるだけ毎月1時間程度、1～4年程度の中長期に渡り収録してもらうよう依頼した。

■収録の流れ 収録調査の流れについても、原則としてCEJCに準ずることとした（田中ほか2018）。協力世帯に依頼することの概要は次の通りである。

1. 参加者に対して収録調査の趣旨やデータ公開の方法などについて説明
2. 参加者に対してデータ収録・公開に関する同意書への署名を依頼
3. 参加者に対してフェイスシート（話者の性別や出身地など）への記入を依頼

4. 収録の日時や使用機材、参加者等の情報を記録
5. カメラ・ICレコーダーを用いた会話の収録（毎月1時間程度）
6. 収録データの調査者への提出（3～6ヶ月に1回程度）

先述の通り、調査者は収録に介在しないため、親戚や友人、知人など、会話に新たに参加する人（参加者）に対して収録調査の趣旨を説明し、データ収録・公開に関する同意をとってもらうなど、その内容は多岐に渡る。

新たな参加者がいる場合には、上記の1～3を行う必要がある。収録の都度、4と5を行う。収録時間は、毎月1時間程度としたが、子育て世帯の多忙さを考慮し、厳密に徹底するよう依頼することはしなかった。6については、録音・録画データのほか、4の収録の記録、及び、2の同意書、3のフェイスシート（新規参加者がいる場合）をまとめて提出してもらう。

■収録機材 収録には、原則として表1に示すカメラ2台、ICレコーダー1台を用いた（基本収録）⁽³⁾。

表1 基本収録で用いた機材と設定

	品名	設定	使用台数
映像	ZOOM Q2n-4K	1920 × 1080, 30fps	2
音声	SONY ICD-SX1000	リニア PCM 44.1kHz, 16bit	1

CEJCでは、映像収録用にカメラを3台、また音声収録用に会話全体を録るためのICレコーダー1台と各話者の音声を録るための人数分のICレコーダーを使用するといったように、多くの機材を用いる収録方法を採用したが、子育て世帯の負担を軽減するため、収録機材は上記の通り最小限に留めた。

このようにICレコーダーを1台に減らしたため、カメラについては、音声を高精度に収録できる小型の機器を選択した。採用したZOOM Q2n-4Kは、非圧縮で高いサンプリング周波数の音声を収録することができる（設定：リニアPCM 48kHz, 24bit）。1台のICレコーダーと2台のカメラを分散して配置することによって、例えば部屋の中で遊びながら移動するような場合などであっても、各話者の音声をある程度収録することが可能となる。

また、表1の機材を基本としつつも、機材設置の準備が間に合わない場合や、屋外での収録のためにこれらの機材を持ち出すことが難しい場合には、必要に応じてスマートフォンなど容易に利用できる機材を用いて収録してもよいこととした。

収録の様子を図1に示す。

2.3 コーパスの規模

コーパスの規模（目標）は100時間である。収録したデータの中から、会話や話者のバランス、データの質や倫理的・法的な問題などを考慮し、コーパスに格納するデータを選別する。

⁽³⁾ 収録を進めながら、協力世帯から意見をもらい、機材を確定させた。そのため、GoProなど異なる機材で収録したものが一部含まれている。また家の間取りなどの都合により、カメラを3台用いる世帯もある。



図1 収録した会話の映像の例（左は基本収録の機材で、右は iPhone で収録した映像）

これまでに収録したデータから、コーパスの規模は、230 セッション、340 会話⁽⁴⁾、延べ話者数 900 人、総語数 120 万語程度になると予想される。

2.4 転記テキスト・アノテーション

コーパスに格納する 100 時間の会話について、原則 CEJC と同様の基準で転記テキストを作成する（白田ほか 2018）。転記テキスト作成の際、発話単位（JDRI 2017）も認定する。その上で、2 種類の形態論情報（短単位・長単位）を自動で付与する。また 1 割に相当する 10 時間を「コア」データセットと定め、形態論情報を人手で修正するほか、人手で談話行為情報（Iseki, et al. 2019）も付与する計画である。

3. 収録状況

表 2 に示す 8 世帯 14 名の子どもを対象に収録を進めている。子ども会話の収録では、特に子どもが小さいうちは家庭での収録が中心となる。そのため、兄弟のいる世帯や家庭がバイリンガル環境の場合など、できるだけ家族構成などに多様性を持たせるようにしている。また、1~4 年程度にわたり経年的に調査する予定であるが、その中で対象とする子どもの年齢の多様

⁽⁴⁾ 協力が 1 回に収録したものを「セッション」と、セッションからある程度のまとまりをもった範囲を切り出したものを「会話」と称す。倫理的・法的な問題や協力者の希望などを考慮し、問題のある部分をカットした結果、一つのセッションが複数の会話に分かれることもある。

性を確保するため、収録開始時期の月齢についても幅を持たせている。

表2 対象児・協力世帯の情報 (2022年8月現在)

世帯 ID	性別	収録開始時点の月齢	収録開始時期	収録期間 (予定)	同居家族	備考
Y001	女	2歳 6ヶ月	2019年6月	48ヶ月	父・母	
	女	0歳 1ヶ月	2021年8月	22ヶ月		
Y002	女	5歳 7ヶ月	2019年6月	36ヶ月	父・母	日韓バイリンガル
Y005	女	1歳 7ヶ月	2020年1月	36ヶ月	父・母	日中バイリンガル
	男	0歳 2ヶ月	2021年5月	20ヶ月		
Y006	男	6歳 6ヶ月	2020年1月	36ヶ月	父・母・ 姉(小学生)	
	女	1歳 6ヶ月				
Y008*	女	0歳 9ヶ月	2020年6月	18ヶ月	父・母	
Y009*	男	4歳 2ヶ月	2020年6月	25ヶ月	父・母	
	女	1歳 2ヶ月				
Y010	男	3歳 8ヶ月	2020年9月	30ヶ月	父・母	
	男	0歳 10ヶ月				
Y011*	男	4歳 11ヶ月	2021年2月	14ヶ月	父・母	
	男	1歳 6ヶ月				

* 世帯 ID に “*” を付けた 3 世帯は 2022 年 8 月現在で収録を終えている

2.1 節で言及したように、多様な場面における多様な話者との会話を収録するという方針のもと、2019年6月から会話収録を進めてきたが、新型コロナウイルス感染拡大の影響により、2020年以降、対象とする場面や話者に偏りが見られるようになった。

図2は、2019年6月から2021年12月末までに収録された会話を対象に、会話の収録場所および会話相手との関係性の割合を年毎に示したものである⁽⁵⁾。図から、2019年は自宅での収録だけでなく、親類宅など自宅以外の室内や飲食店などの公共商業施設での収録も多く、会話の相手も家族だけでなく親戚や友人知人なども一定数見られたのに対し、2020年、2021年は自宅での収録が多くを占め、会話の相手も家族の割合が高くなっていることが分かる。

こうした偏りはあるものの、調査者が収録に介入せず協力世帯に収録を依頼したことが幸いし、収録自体は停滞せずに進めることができた。

なお、プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」では、幼稚園や小学校での会話を対象とするコーパスも別に構築する予定である。CEJC-Childで不足する友達や先生との会話は、これらのコーパスで補填することができると考えている。

4. おわりに

本発表では、現在構築中の『子ども版日本語日常会話コーパス』の設計と収録状況について報告した。成人話者を中心とする CEJC と接続させ、子どもから高齢者まで長期に渡り実証的に研究できる基盤を構築することを目指し、設計や収録法、アノテーション等はできるだけ CEJC を踏襲することとした。収録状況としては、新型コロナウイルスの影響で、2020年以

⁽⁵⁾ あくまで収録されたデータの分布であり、そこから選別してコーパスに格納するデータの分布ではないことに注意する必要がある。

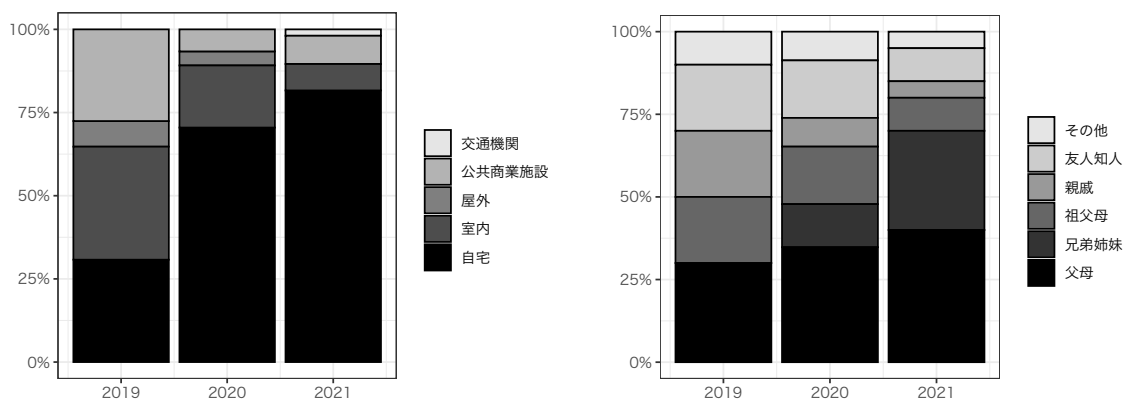


図2 収録された会話の収録場所・相手との関係性

降、自宅での家族との会話の収録の割合が高くなっているが、協力世帯に収録を依頼していることから、収録自体は順調に進んでいる。2025年頃の公開を目指して開発を進める。

謝 辞

本研究は、国語研究所共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」および科研費20H01264による成果である。

参考文献

- Yuriko Iseki, Keisuke Kadota, and Yasuharu Den(2019) “Characteristics of everyday conversation derived from the analysis of dialog act annotation”, *Proceedings of the 22nd Oriental COCOSA*, pp.16.
- JDRI (2017) 『発話単位ラベリングマニュアル』 version 2.1, 2017.
- 白田泰如・川端良子・西川賢哉・石本祐一・小磯花絵 (2020) 「『日本語日常会話コーパス』における転記の基準と作成手法」『国立国語研究所論集』15号, pp.177-193.
- 小磯花絵・天谷晴香・石本祐一・居關友里子・白田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉・渡邊友香 (2022) 「『日本語日常会話コーパス』の設計と特徴」『言語処理学会第28回年次大会発表論文集』pp.2008-2012.
- 田中弥生・柏野和佳子・角田ゆかり・伝康晴・小磯花絵 (2018) 「『日本語日常会話コーパス』の構築—会話収録法に着目して—」『国立国語研究所論集』14, 275-292.
- 宮田 Susanne 編 (2004) 『今日から使える発話データベース CHILDES 入門』ひつじ書房.

実践医療用語_語構成要素語彙試案表 Ver. 2.0 の構築

東条佳奈 (大阪大学)
黒田航 (杏林大学)
相良かおる (西南女学院大学)
高崎智子 (西南女学院大学)
西嶋佑太郎 (医師)
麻子軒 (関西大学)
山崎誠 (国立国語研究所)

Development of a Word Component Database for "Hands-On Medical Terms" (Version 2.0)

Kana Tojo (Osaka University)
Kou Kuroda (Kyorin University)
Kaoru Sagara (Seinan Jo Gakuin University)
Satoko Takasaki (Seinan Jo Gakuin University, Physician)
Yutaro Nishijima (Physician)
Tzu-Hsuan Ma (Kansai University)
Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

医療記録データには、複数の単語が連結された合成語が多く存在する。そのため、自然言語処理を効率的に行うためには、合成語の語構成や、それらの構成要素の意味に着目し、合成語の構造を明らかにする必要がある。しかし、医療記録は非公開という資料的特質のため、言語学的な調査があまり行われてこなかった。また、医療関係者における意味のある言語単位も定まっておらず、整理の必要があった。こうした背景に基づいて作成した言語資源が『実践医療用語_語構成要素語彙試案表 Ver.2.0』である。本試案表は、『実践医療用語辞書 ComeJisyoSjis-1』より抽出した合成語より作成した『実践医療用語_語構成要素語彙試案表 Ver.1.0』を更新したもので、7,087 語の合成語について、それぞれを構成する語構成要素 6,633 種と、語構成要素に付与した意味ラベル 41 種を収録している。本発表では、Ver.1.0 からの変更点と、本言語資源の特徴、意味ラベルに注目した語構成要素について概観を行った。

1. はじめに

本発表は、2022 年に言語資源協会より無料で配布を開始した電子データである『実践医療用語_語構成要素語彙試案表 Ver.2.0』¹ (以下、「本試案表」とする) について紹介するものである。

電子カルテシステムの普及により医療記録データが蓄積されていくにつれ、利用者が求

¹ <https://www.gsk.or.jp/catalog/gsk2020-g>

める情報の適切な検索や、二次利用のための技術の重要性が高まっている。自然言語処理では、処理を行うための最初のステップとして、テキストを意味のある最小の単位（一般には単語）に分割する前処理が必要となる。英語のテキストの場合、単語間は基本的にスペースで区切られているため、コンピュータでの単語に分割する処理も容易である。一方、分かち書きのされない日本語のテキストは、単語分割の処理が容易ではない。

特に、医療記録データには「腹腔鏡補助下低位前方切除 D2 郭清施行」の様に複数の単語が連結された合成語が多く、医療関係者にとって意味のある最小の単位がどのようなものかという統一的な見解もないため、語分割のための情報が重要となる。加えて、医療記録は各々の医療現場で作成されるため、表記の揺れ、誤字、標準的でない用語などを含む専門的かつ実践的な医療用語が多用されており、効率的な自然言語処理のためには、それらの語構成や語種構成の実態の調査も必要である。

しかし、医療記録データは個人情報豊富に含まれているために非公開とされており、こうした資料的特質から、実践医療用語の言語学的な調査は、十分に行われてはこなかった。

こうした背景のもと、発表者らは、医療記録文の自然言語処理支援のために相良が作成・公開した形態素解析器 MeCab 用のユーザー辞書『ComeJisyo』の登録語²より、医学的な知識が不足していても語の境界がわかりやすいように、『『分類語彙表—増補改訂版—』に収録されている語を含む合成語』を抽出し、語構成要素とそれぞれの意味ラベルを付与した『実践医療用語_語構成要素語彙試案表 Ver.1.0』(以下、「試案表 Ver.1.0」とする)を作成した(詳細な手順については相良他 2019、相良他 2020、相良 2021 を参照)。試案表 Ver.1.0 において見直しの必要があった意味ラベルの妥当性(相良 2021)を踏まえて、合成語や意味ラベルを見直し、新規の情報を追加するなどのアップデートを施したものが、Ver.2.0 にあたる本試案表である。

具体的には、試案表 Ver.1.0 より不要な語を除いた 7,087 語を対象合成語とし、これらの合成語を構成する語構成要素 6,633 要素、およびそれぞれの要素に付与した 41 種類の意味ラベルを収録したものとなっている。

ファイル種別は Excel ファイル(拡張子 xlsx)で、「はじめに」(用語の定義及び各シートの解説)、「対象語」(対象語一覧)、語構成要素(語構成要素一覧)、意味ラベル(意味ラベル一覧)の4つのシートから構成されており、それぞれのシートには、以下の情報が収録されている。

(1) 「対象語」シート (7,087 語) :

通し番号 (ID) ・ 対象合成語 (見出し) ・ 対象合成語の読み ・ 語構造 ・ 文字長情報
「語構成要素」シート (6,633 種) :

通し番号 (ID) ・ 語構成要素 ・ 語構成要素の読み ・ 意味ラベル ・ 対象語における語構成要素の出現頻度 ・ 当該語構成要素が語頭に出現した頻度 ・ 当該語構成要素が語末に出現した頻度 ・ 語構成要素が対象語と一致する場合の対象語 ID

「意味ラベル」シート (41 種) :

通し番号 (ID) ・ 意味ラベル ・ 当該意味ラベルに分類された語構成要素の個数 ・ 意味ラベルの属性 (具体か抽象か)

² 分かち書き用実践医療用語辞書 ComeJisyo
<https://ja.osdn.net/projects/comedic/>
試案表作成の際には ComeJisyoSjis-1 の登録語 (111,664 語) を使用した。

以下、本試案表における用語と、試案表 Ver.1.0 からの変更点について触れながら、本言語資源の特徴を述べる。

2. 本試案表における用語について

本試案表が対象とする「実践医療用語」は、医療施設で使われる医療記録に含まれる「学術上の専門用語」と「それ以外の専門用語」を指す(相良 2021)。

また、本試案表で扱う「語構成要素」は、相良(2021:559)に従い、「合成語を構成する要素で、合成語を医療の観点から意味的にまたは統語的に分割可能なすべての部分文字列」と定義し、分割できない合成語については元の合成語を語構成要素ととらえる。例えば(2)のように、「脳幹多発性硬化症」という合成語は、「脳幹」「多発性」「硬化症」という短い語構成要素のほか、「多発性硬化症」という語構成要素同士が結合した形も一つの語構成要素として捉えて集計している。

(2) 合成語：脳幹多発性硬化症

語構成要素：脳幹，多発性，硬化症，多発性硬化症

また、本試案表においては、それぞれの語構成要素に、医療の観点での意味を表すラベルとなる「意味ラベル」を付与している。意味ラベルは語構成要素と必ずしも1:1の関係ではなく、語構成要素が多義の場合は複数の意味ラベルを「，」で区切り、列挙して示している。例えば語構成要素「流動食」の意味ラベルは《医療行為，食品，治療食》を付与している³。

「流動食」は、食品という点では一般の食品が含まれており、同時に《治療食》という《医療行為》の一つとも考えられる。そこで《医療行為》と《食品》と《治療食》の意味ラベルを併記して示している。なお、ここでの「意味」は、用語の意味を表すもの(例：《衛生物品》)と、医療者の使用面での心的な捉え方を表すもの(例：《病因》)が含まれている。

3. 『実践医療用語_語構成要素語彙試案表 Ver. 1.0』からの主な変更点

1節で述べたように、本試案表は、2021年に公開した試案表 Ver.1.0を見直し、変更を加えたものである。主な変更点は、①不要な対象合成語の削除、②合成語の語構造情報の付与、③意味ラベルの見直し・統合である。そのほか、対象合成語・語構成要素の見直しに伴い変更した情報についても更新を行っている。

①については、異体字による重複・誤字・現在では使われない語等を削除した。②の語構造情報は、(3)のように対象合成語の分割位置を医師が確認し、[]で区切りを入れたものである。これにより、(4)のような長大な合成語についても意味理解がしやすくなったといえる。また、合成語がどのような結合パターンになっているかについての検討なども可能になった。

(3) 対象合成語：悪性脳腫瘍

語構造：悪性[脳[腫瘍]]

語構成要素：悪性，脳，腫瘍，脳腫瘍

³ 本発表では便宜上、意味ラベルは《》に入れて示す。

- (4) 対象合成語：脳淡蒼球内オイルプロカイン注入療法
語構造：[脳[淡蒼球内]][[オイル[プロカイン]][注入[療法]]]

次に③についてだが、本試案表では、80種類とかなり細分化されていた試案表 Ver.1.0 の意味ラベルを見直して統合を行い、41種類にまとめた。(5)に試案表 Ver.1.0より削除したラベル46種類を、(6)には本試案表のラベル41種類を示す。なお、今回新たに追加したラベル7種類には下線を付している。

- (5) 「試案表 Ver.1.0」より削除したラベル46種：
意向，動き，音，化学現象，課題，感覚，関係，基準，軌跡，基礎，規則，機能，教育，距離，傾向，行為，作用，時間，色彩，社会，手技，主体，順序，数量，成育，性質，制度，増減，属性，体外物質，体内物質，知的産物，調節，程度，認識，熱，能力，波動，光，文法用語，変化，方向，保健衛生，保留，様相，例示
- (6) 「試案表 Ver.2.0」のラベル41種：
維持行為，位置，医薬品，医療行為，衛生物品，化学物質，患者属性，患部，機器，経過，形状，検査，サービス，施設，指標，種類，症状，状態，食品，身体機能，身体部位，精神，生体物質，生理，組織，治療食，動植物，排泄物，場所，ヒト，費用，病因，病原体，病態，病名，物品，部分，法規，方法，予防行為，#未定

試案表 Ver.1.0 から削除した意味ラベルの中で、最も語構成要素が多かったのは《体内物質》、次いで《体外物質》であった。これらはより分類しやすいラベルとして、《生体物質》《化学物質》のラベルに変更した。また、(5)には、《感覚》(例：「感覚」「知覚」「聴覚」など)《動き》(例：「感覚運動」「眼球運動」「関節運動」など)《機能》(例：「肝機能」「呼吸機能」など)のように、どのように異なるのか区別が判然としない類義の意味ラベルも含まれていた。これらのラベルについては上位ラベルとして《身体機能》を想定し、《身体機能，精神》《身体機能，生理》のように複数のラベルを付与することで整理を行った。結果として、《感覚》《動き》《機能》などのラベルは、《身体機能，生理》という2種のラベルを付与することで、「自我意識」「欲求」などの精神的な身体機能と区別できるようになった。

加えて、試案表 Ver.1.0 に付与されていた意味ラベルのうち、該当する語構成要素がごくわずかしかなかったものは、合成語の中でどのような意味合いかを再度検討し、より上位の意味ラベルに変更した(例：「認知」《認識》→《身体機能，生理》、「白色」《色彩》→《状態》など)。削除した意味ラベルの中には、過分割していた語構成要素の見直しを行った結果、不要になったものもある。例を挙げると、「任意」《意向》→「任意入院」《医療行為》のような変更である。一方で、本来は分割すべき語構成要素を短くすることも行っている(例：「ぶどう膜炎」《病名》より、「ぶどう膜」《身体部位》を語構成要素として追加)。このように本試案表では、より医療の観点で有用になるように語構成要素の分割位置に関する調整も行った。

そのほか、命名が困難なものや、医療用語特有のものではない一般的な語構成要素⁴に「#未定」のラベルを付与したことも大きな変更点である。

⁴ 『岩波国語辞典第五版タグ付きコーパス 2004』に立項されているかどうかを判断の基準とした。

4. 意味ラベルからみる語構成要素の概観

本試算表では、新たに意味ラベルごとに、分類された語構成要素の個数の情報を追加した。語構成要素数の降順に 20 位まで示したものが表 1 である。

最も多いものが《病名》(1918 種)で、《状態》(1669 種)、《身体部位》(1639 種)、《病態》(1202 種)と続く。この 4 種の意味ラベルで全体の約 70%という割合となる。本試算表に収録されている対象語のうち、大部分が病名である⁵ため、《病名》ラベルが最も多くなる。同様に、病名にはその症状が発生している患部が同時に含まれることが多いため、《身体部位》ラベルが付与される語構成要素も高頻度となった。

《状態》ラベルが付与された語構成要素のほとんどは、《病態》ラベルと併記しているものになった。《病態》は、病的な状態や、単独では病名を構成しない語構成要素に付与しており、《状態》の下位にあたる意味ラベルである。さらに、《病態》の中には、「細菌性食中毒」における「細菌性」のように、《病因》であり、かつ《病原体》であるというものもある。この場合、(7)に示すように、意味ラベル同士の関係は等位ではなく、上位下位関係にあると解釈される。

表 1 語構成要素数上位 20 の意味ラベル

意味ラベル	語構成要素数
病名	1,918
状態	1,669
身体部位	1,639
病態	1,202
医療行為	319
症状	249
化学物質	241
部分	239
位置	235
経過	176
病因	145
# 未定	141
病原体	123
身体機能	96
生体物質	89
種類	83
生理	77
医薬品	72
形状	60
ヒト	54

⁵ 《病名》もしくは《病態》ラベルが付与されたものが対象語全体のうち 73.6% (5,218 語)であった。

(7) 対象合成語：細菌性食中毒

語構成要素	意味ラベル
「細菌性」	《状態》 > 《病態》 > 《病因》 > 《病原体》
「食中毒」	《病名》

つまり、本試案表に付与されている意味ラベルのうち、「,」で列挙しているものの中には、包含関係で表されるものもあるということである。《状態》ラベルは意味ラベルの中では上位概念にあたるラベルとなっており、他にも、《患者属性》（例：「老人性貧血」の「老人性」）・《形状》（例：「線状網膜炎」の「線状」）・《経過》（例：「進行性筋萎縮」の「進行性」）などが《状態》の下位概念にあてはまる。

試案表 Ver.1.0 では、「～性」となる語構成要素をはじめ、あらゆるものに《状態》ラベルが付与されており、「意味」による分類が出来ているとは言い難いものになっていたが、本試案表で意味ラベルを複数付与・併記したことで、《状態》ラベルの細分化を行うことができたといえる。例えば、同じ「～性」でも、「サルコイドーシス性」は《病態》で、「コクシジオイデス性」は《病因》かつ《病原体》である、というように、「性」の前要素に関する知識がなくとも意味判断ができるようになった。

しかし、本試案表ではこのような上下関係または包含関係にあたるものも、そうではなく、多義のために併記しているもの（例：《位置》であり《経過》を示す「前」など）もいずれも「,」で区切って列挙をしており、見かけ上区別がつかず混在しているため、意味ラベルの関係性の整理が必要と思われる。

また、表 1 で示したように、《#未定》ラベルを付与している語構成要素が少なからずある。これらは前述の通り、命名が困難なものや、医療用語特有のものではない語構成要素があてはまるが、ラベルを作るほど合成語や語構成要素の数が少ない、というものも含まれる。本試案表の作成の目的の一つである、自然言語処理の効率化等のためには、未定ラベルの整理も課題であるといえよう。

5. おわりに

本発表では、『実践医療用語_語構成要素語彙試案表 Ver.2.0』の概要について、主に Ver1.0 との変更点と意味ラベルを中心に説明した。ComeJisyo より抽出した実践医療用語の合成語について、語構成要素と、それぞれに対応する意味ラベル、合成語の語構造情報を付与した本試案表を利活用することで、医療用語の語彙的な特徴の分析のほか、将来的に新たな医療記録データの解析に応用していくことが可能であると思われる。

意味ラベル間関係の整理や、未定ラベルの検討についても引き続き行いながら、本試案表の精緻化を今後も行っていく予定である。また、これまでは、合成語の抽出の際に『分類語彙表一増補改訂版』に収録されている「一般語を含む語」であることを条件にしていたため、語構成要素がすべて専門用語からなる合成語についても対象にしていく必要があるだろう。いずれも今後の課題としたい。

謝 辞

本研究は JSPS 科研費 JP18H03499 ならびに JP21H03777 の助成を受けている。

文 献

- 国立国語研究所 (2004) 『分類語彙表—増補改訂版—』 大日本図書.
- 相良かおる・山崎誠・麻子軒・東条佳奈・小野正子・内山清子 (2019) 「実践医療用語の語構成要素意味を基準とした分割」『人文科学とコンピュータシンポジウム論文集』pp.57-64.
- 相良かおる・小野正子・高崎智子・東条佳奈・麻子軒・山崎誠 (2020) 「実践医療用語の語構成と意味—語構成要素語彙試案表の作成にむけて—」『人文科学とコンピュータシンポジウム論文集』 pp.289-296.
- 相良かおる (2021) 「実践医療用語における語構成要素の意味ラベルについて」『言語処理学会第 27 回年次大会発表論文集』 pp.559-562.
https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/P3-10.pdf

関連 URL

- 『岩波国語辞典第五版タグ付きコーパス 2004』 <https://www.gsk.or.jp/catalog/gsk2010-a/>
- 『分かち書き用実践医療用語辞書 ComeJisyo』 <https://ja.osdn.net/projects/comedic/>
- 『実践医療用語_語構成要素語彙試案表 Ver.2.0』 <https://www.gsk.or.jp/catalog/gsk2020-g>

テレビゲームコーパスの構築とその利活用

麻 子軒 (関西大学国際教育センター) †

Construction and Utilization of Video Game Corpus

Ma Tzuhsuan (Kansai University)

要旨

コーパス言語学の普及に伴い、これまでは書籍・雑誌・新聞・小説・漫画など、様々な媒体によるコーパスが構築されてきたが、テレビゲームを対象に作成されたものはまだない。本稿では、キャラクターのセリフが多く含まれるロールプレイングゲームを例に取り、テレビゲームコーパスを構築する方法論を述べた上で、その利活用の例を示し、同コーパスに言語資源としての価値があることを主張する。具体的に、役割語や日本語教育といった分野における研究例が挙げられる。また、ロールプレイングゲームには、ナレーションによる状況描写が少ないことや、一部会話のやり取りがプレイヤーの想像で補完しなければならないことなどの特徴がある。このことから、本稿で作成したテレビゲームコーパスを用いれば、同じ「作られた話し言葉」として代表される小説・漫画では観察できない言語的特徴を見出すことができると考えられる。

1. はじめに

コンピュータの計算能力が飛躍的に向上したことで、大規模な言語データを高速に処理することが可能となり、コーパス言語学が大いに発展を遂げた。2011年に、国立国語研究所によって日本語に関する初めての大規模コーパス『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)が公開された(国立国語研究所 2015)。それ以来、ウェブコーパス、歴史コーパス、学習者コーパスなど、異なる目的で開発されたコーパスが次々と世に出た。また、安価かつ高性能なコンピュータの発売により、研究者も各自の研究目的に応じてコーパスを自作することが可能となった。

こうした背景の下、これまで書籍・雑誌・新聞・小説・漫画¹など、様々な媒体によるコーパスが構築されてきた。しかし、テレビゲーム(以下、ゲーム)を対象に作成されたコーパスはまだない。その理由として、ゲームの文字情報を表示させるのにプレイヤーの操作が必要であることや、テレビ画面上に表示された文字情報を電子化する作業が煩雑であることが考えられる。媒体が紙である小説や新聞の場合は、紙に印刷された文字を画像にスキャンし、OCRソフトウェアを用いれば電子化作業がある程度自動化できるが、ゲームはその媒体の性質上同じ処理ができない。また、ゲームは子どもの遊びである固定観念がまだ強く、研究の対象として扱われていないことが多い。

そこで、本稿ではゲームコーパスを構築する方法論を述べた上で、その利活用の例を示し、同コーパスに言語資源としての価値があることを主張する。今回は言語情報が比較的多く含むと思われるロールプレイングゲームを例に取るが、本稿で提案した方法論が今後ほかのジャンルのゲームコーパス作成の参考になればと思う。

† kenji.ma@kansai-u.ac.jp

¹ 公開されていないが、研究のために各研究者によって作成された漫画コーパスはある。自作の漫画コーパスで漫画の言語的特徴について考察したウンサーシュッツ (2013) がその一例である。

2. コーパスの構築方法

構築方法は、まず、ゲームを一つ選定し、そのゲーム内の言語情報をすべて電子化し、通し番号や文字区分などの情報を付与した後、プログラミング言語で処理しやすいように CSV 形式の構造化データとして保存する（以下、本文データ）。次に、本文データの言語情報のみ形態素解析器で処理を行い、語彙素や品詞などのアノテーションが付与されたデータを別個の CSV 形式で保存する（以下、形態素解析データ）。以下、手順を紹介する。

2.1 対象ゲーム

ゲームにもアクションゲームやパズルゲームなどの様々なジャンルがある。本稿では言語学として研究する可能性を考慮し、ロールプレイングゲーム（以下、RPG）というジャンルをコーパス化の対象とする。RPG とはプレイヤーがゲームの主人公になり、架空の世界でほかのキャラクターに話しかける、情報を集めるなどして、物語を展開させていくものである。そのため、キャラクター間の会話が大量含まれており、言語情報がほかのゲームジャンルと比べて比較的多いと思われる。

具体的に選定したのは、1988年にエニックス社（現在はスクウェア・エニックス）によって発売された『ドラゴンクエスト3』である。同ゲームが日本で最も代表的なRPGの一つであり、後に複数のリメイク版が制作された。本稿ではコーパス化する際に、1996年のリメイク版のものを使用する。昔のゲームを選んだ理由は、画面構成とシステムが比較的単純であり、方法論を確立するために適しているからである。

2.2 電子化

コーパスは、機械可読性を持たせるために、電子化する必要がある。電子化の対象は、ゲーム内におけるすべての言語情報である。RPGのゲーム画面に現れる言語情報は、主にメニュー、メッセージ、セリフの3種類に分けられる。「メニュー」は図1の「ヒール」「HPを30～回復する」のようなゲーム内の道具・魔法名やそれに対する説明文（HPの数字などの変数は含まない）で、「メッセージ」は図2の「主人公はタンスを調べた！300ゴールドを手に入れた！」のようなゲーム内の進行状況を語るナレーション（小説での地の文に該当する）で、「セリフ」は図3の「こんにちは。ここは国立国語研究所ですわ。」のような物語の展開と関わるキャラクターによる発話（小説での会話文に該当する）である²。電子化する際、一つの枠に表示されたすべてのテキストを一行の文字列として記録する。

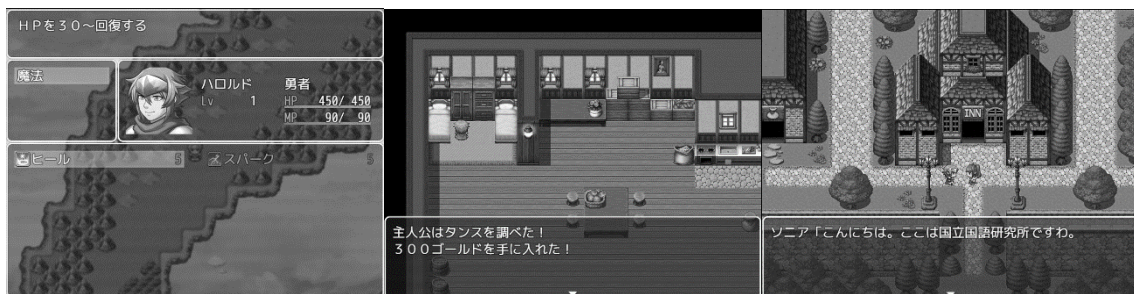


図1 メニュー

図2 メッセージ

図3 セリフ

² 著作権の関係で、実際に発売されたゲームの画面を掲載することができないため、図1～3は筆者が「RPG ツクールMV」というソフトウェアで作成したイメージである。

電子化する最も効率的な方法は、ゲーム媒体に記録された情報から直接文字を読み出すことであるが、昔のゲームの文字コードは PC 用の Shift-JIS や UTF-8 といった標準化されたものでなく、各ゲームメーカーの独自の方法でコーディングされたものが多かった。そのコード表が既知の場合は、理論的にそれを利用して文字情報を自動的に取り出せるが、直接ゲームの記録媒体にアクセスし、情報を抽出することは「解析」に該当する行為で、ゲームの利用規約に違反するおそれがある。

そこで、本稿では代わりに下記の方法を提案する。まず、実際にゲームを一通りプレイし、言語情報が画面に表示されるたびにスクリーンショット画像に保存した後、OCR ソフトウェアで文字を認識させる³。次に、一行の文字列ごとに番号や文字区分などの属性を付与し、後の整形や形態素解析に備え、プレーンテキスト（文字コード UTF-8）の CSV 形式で「本文データ」と名付けて保存する。設定した属性は表 1 の通りである⁴。

表 1 「本文データ」の構成

属性	説明
番号	すべての行の文字列に対し、重複しないように一意に付与された通し番号
区分	言語情報の種類を表す属性で、「メニュー」「メッセージ」「セリフ」の3種類がある
画像	キャラクターの外見を表す画像を名付けた名称（区分がセリフの場合のみ）
性別	キャラクターの性別（区分がセリフの場合のみ。判断不能の場合は「*」と記す）
名前	キャラクターの名前（区分がセリフの場合のみ。設定されていない場合は「*」と記す）
場所	キャラクターの所在（判断不能の場合は「*」と記す）
内容	ゲーム画面に表示された言語情報 ⁵ （区分がセリフの場合は発話内容）

実際に作成した「本文データ」の例は表 2 である。

表 2 「本文データ」の作成例

番号	区分	画像	性別	名前	場所	内容
932	セリフ	王	男	*	アリアハン	もはや 魔王に いどめるような者は おらぬ。 われわれには もう 希望はないのか...
933	セリフ	主人公の母	女	*	アリアハン	いいえ 王さま。 この子が います。 オルテガの 血をひく [主人公の名前] が。
934	セリフ	主人公の母	女	*	アリアハン	夫の 遺志は きっと この子が ついでみせますわ！
935	メッセージ				オリビアの岬	どこからともなく 悲しげな 歌声が 聞こえる.....。
936	セリフ	オリビア	女	オリビア	オリビアの岬	ああ エリック！ 私の 愛しきひと。 あなたをずっと まっていたわ。
937	セリフ	エリック	男	エリック	オリビアの岬	オリビア ぼくのオリビア もう君を はなさない！
938	セリフ	オリビア	女	オリビア	オリビアの岬	エリックーっ！

³ 同一対象に何度話しかけても同様の内容が表示される場合、電子化するの是最初の一回のみとする。ただ、ゲームの進行によってセリフが変わるキャラクターの場合は異なる発話をすべて収録する。

⁴ 今回のような小規模コーパスは必要ないが、複数のゲームを含むコーパスを作成する場合は、ゲームタイトル、ゲームジャンル、制作会社などの属性も必要になってくると思われる。なお、画像や名前を入れたのは、後に述べる役割語研究において必要な情報だからである。

⁵ 主人公の名前のような、プレイヤーの入力によって変わる変数は適当な文字列を付けて [] で示す。

2.3 形態素解析

検索を効率的に行うためには、文字列を形態素に分割し、品詞などのアノテーション情報を付与する必要がある。ただ、半角文字が混在すると、後のプログラミングによる処理でトラブルに繋がることが多いため、形態素解析を施す前に、前作業として半角を全角に置換する。また、ゲームに出現するキャラクター名や地名などの固有名詞を事前にユーザー辞書に登録しておくことで解析精度が上がる。

実際に形態素解析を行う対象は「本文データ」の「内容」欄の情報である。用いる解析器は MeCab0.996⁶、辞書は UniDic1.3.12⁷である。解析結果をプレーンテキスト（文字コード UTF-8）の CSV 形式で「形態素解析データ」と名付けて保存する。ここで、「本文データ」で付与された「番号」を残し、両データが互いに紐付けられるようにしておけば、区分や名前などの情報が必要な際に、プログラミング言語を用いて抽出することができる。MeCab の処理で得られるのは短単位データであるが、Comainu⁸を用いれば、長単位データを得ることもできる。表 3 は、番号 934 の内容に対応する形態素解析の結果である。

表 3 「形態素解析データ」の作成例

番号	原文文字列	品詞(大)	品詞(中)	品詞(小)	活用型	活用形	語彙素	語彙素読み	語種
934	夫	名詞	普通名詞	一般	*	*	夫	オット	和
934	の	助詞	格助詞	*	*	*	の	ノ	和
934	遺志	名詞	普通名詞	一般	*	*	遺志	イシ	漢
934	は	助詞	係助詞	*	*	*	は	ワ	和
934	きつと	副詞	*	*	*	*	急度	キット	和
934	この	連体詞	*	*	*	*	此の	コノ	和
934	子	名詞	普通名詞	一般	*	*	子	コ	和
934	が	助詞	格助詞	*	*	*	が	ガ	和
934	つい	動詞	一般	*	五段-ガ行	連用形-イ音便	繼ぐ	ツグ	和
934	で	助詞	接続助詞	*	*	*	て	デ	和
934	みせ	動詞	非自立可能	*	下一段-サ行	連用形-一般	見せる	ミセル	和
934	ます	助動詞	*	*	助動詞-マス	終止形-一般	ます	マス	和
934	わ	助詞	終助詞	*	*	*	わ	ワ	和
934	!	補助記号	句点	*	*	*	!		記号

上記の手順を経て得られた「本文データ」の文字列は 5,900 行で、内訳はメニューが 72、メッセージが 1,162、セリフが 4,666 である。キャラクターは全部で 79 種類である。

「形態素解析データ」の総語数は延べ 67,715 語、異なり 3,164 語で、記号類を除けば延べ 58,407 語、異なり 3,150 語になる。今回用いたのは 1996 年発売のゲームであるが、記録媒体の大容量化が進み、近年制作されたゲームはさらに規模が増えると予想される。

⁶ <https://taku910.github.io/mecab/>

⁷ 最新版は 3.1.10 である。 <https://clrd.ninjal.ac.jp/unidic/>

⁸ <http://comainu.org/>

2.4 注意点

ゲームは、書籍や新聞などの媒体とは異なる性質を持つため、コーパス化する際にいくつかの注意点がある。以下にそれを記しておく。

一つ目、ストーリーの時系列が一定でないゲームがある。書籍の場合、最初のページから最後のページまで読む順番が決まっており、ストーリーの進行が一本道である。それに対し、自由度が高いゲームの場合、先にAに行ってからBに行くこともできれば、Bに行ってからAに行くこともできる。文脈を問題にする研究の場合は要注意である。

二つ目、ストーリーの本筋とは別に、サブイベントが設定されているゲームがある。これらのイベントは見なくてもゲームをクリアできる。書籍の場合、書かれている内容は意図的に飛ばさない限り、すべて目にするのが原則であるが、ゲームの場合はプレイヤーのよって目に入る言語情報の量が異なる。また、いわゆる隠しイベントもあるため、熟練したプレイヤーでないと、すべての言語情報を抽出できない可能性がある。

三つ目、プレイヤーの行動によってストーリーが変わるゲームがある。マルチエンディングを採用しているゲームがその典型例である。具体的に、Aルートに進めば、Bルートのストーリーが見られないため、一周回のプレイですべての言語情報を収集できない。研究目的によって、作品総語数の認定が難しくなる場合がある。

四つ目、ゲームの歴史は書籍に比べてまだ浅く、著作権はほとんどまだゲーム会社が持っており、著作権処理は個人研究者として交渉することが困難であろう。著作権が切れた書籍のように自由に公開できるゲームがほぼない現状では、ゲームコーパスは個人研究用として構築することが可能であっても、公開が難しいことが予想される。

五つ目、ゲームは最初に発売されたオリジナル版と、後に異なるプラットフォームに移植するため新たに作成されたりリメイク版とが存在する場合がある。後者には内容の追加と削除が行われることがあり、収録された言語情報も変化する可能性がある。書籍に例えるならば異なる「版」に類似した概念である。コーパス化する際に、どのプラットフォームのバージョンを用いたかを明記する必要がある。

本節では、ゲームコーパスを構築する一例として、『ドラゴンクエスト3』というRPGを取り上げた。RPGに特化した方法ではあるが、ほかのゲームジャンルにも応用できると思われる。ただ、以上見てきたように、ゲームはその媒体の特殊性によって、書籍や新聞コーパスと同列に扱えない点にも注意が必要である。次節では、ゲームコーパスが実際にどの分野の研究に役立つかについて述べる。

3. ゲームコーパスの応用例

3.1 役割語に関する研究

金水(2003)は、特定の人物像を思い浮かべるある特定の言葉づかいを役割語と定義した。また、金水(2014)は、「役割語は、言葉のステレオタイプ」とした上で、「ステレオタイプが最も効力を発揮するのは、フィクションにおいて」と指摘した。そのフィクションには、具体的に漫画、アニメ、小説などが含まれ、これまで多くの研究が行われたが、ゲームに関しては麻(2019a)以外あまり見られず、以下にその例を挙げる。

麻(2019a)は、『ドラゴンクエスト3』を対象に、役割語の調査を行った。具体的に、作中の登場キャラクターのセリフに出現した人称代名詞・助動詞・助詞・感動詞の4つの品詞を持つ表現の量的分布を統計量とし、クラスター分析を適用した。その結果、キャラクターをその話し方によって、「異人ことば」「女ことば」「中性ことば」「老人ことば」「男ことば」

「少年ことば」の6グループに分類することができた。さらに、この6グループの分類結果を利用して、特化係数で各グループのキャラクターの特徴的な言語表現を表4のように抽出した(麻2019aより一部改変)。

表4 各グループのキャラクターの特徴語

	代名詞	助動詞	格助	副助	接助	終助	準体	感動詞
異人	わし, おまえ, だれ, わたし	ていく, てる, ない		か				おお
女	あたし	ざます, てらっしゃる, てあげる, ちゃう, ておく		ったら, なんて, って	けれど, けど	かしら, もの, わ, の, ね (他1語)		きゃー, まあ, おほほほほ, あら, はあ (他8語)
中性	オイラ, あなた, 彼女	ます, ベシ						やあ
老人	おぬし, そなた, わし, おのれ	じゃ, とる, である, まい		ぞ, なん, なぞ		のう, ぞい, ぞ		はくしょん, ふーむ, ふあ, やれやれ, ふむ (他3語)
男	あたい, おめえ, オレ, あんた, われ (他1語)	やがる, ちまう, てやる, らしい, てる		なんか, って, しか	けれど, ちゃ	ぜ, い, さ, や, な	ん	へい, おい, へー, おっと, ひゃー (他4語)
少年	ボク	ちゃう, てる, ていく		って, なんて	けれど, から	よう, さ, や, なあ, の (他2語)	ん	わーい, ねえ, ん, え

麻(2019a)は、ゲームコーパスにも言語資源としての価値があることを示唆した。また、上記の研究では言及されていないが、(1)～(3)のような侍ことば、幼児ことば、ロボットことばもゲームで観察できる。(3)は表記の手段で表現された役割語である。

- (1) わしもいく**でござる**！世界をこのままほうっておくわけにはい**か**ない**でござる**！
(『ファイナルファンタジー6』カイエン)
- (2) ええっ！？フェアリーちゃんが？た**た**、たいへ**ん**で**ち**！どうすればたすけられ**ま**ちか？めがみさま！(『聖剣伝説3』シャルロット)
- (3) ケガのこうみょうというやつ**デショウカ**？ぐうぜんにも、飛行機能が手に入り**マシタネ**！(『クロノ・トリガー』ロボ)

さらに、定延(2007)などで述べられたような、人工的な表現または方言を用いるキャラクターもゲームに存在する。下記の3例がこれに該当する。(4)が人工的な表現で、(5)(6)が関西弁と名古屋弁をベースにした表現だと思われる。

- (4) みんな！みんな死んじ**ゃ**ったと思**っ**た**クポ**！生きてた**クポ**ー！よ**か**った**クポ**ー！ボクもいっしょに戦**うクポ**！！(『ファイナルファンタジー6』モーグリ)

- (5) ちゃうねん！ええ話やわあ。あんさん達の話に、めっちゃ感動しとったんよ！マナが少なくなったせいか、最近どうも涙もろうて... (『聖剣伝説3』 ウンディーネ)
- (6) でも、度重なるじしんで、洞窟の入り口がふさがってしまったぎゃー。山の行き止まりの大岩の近くだぎゃー。(『聖剣伝説3』 ダークプリースト)

以上見てきたように、ゲームコーパスは役割語の分析に役立つことが分かった。この点は次に述べる日本語教育にも関連している。

3.2 日本語教育に関する研究

国際交流基金が2018年に実施した「海外日本語教育機関調査」によると、世界の日本語教育機関に在籍する学習者の学習目的・理由として挙げた項目のうち最も回答が多かったのは「マンガ・アニメ・J-POP・ファッション等への興味」(66.0%)であるという(国際交流基金2020)。このなかにゲームも含まれており、具体的な内訳の割合が提示されていないが、ゲームも日本語学習者にとって重要な動機づけになっていると思われる。

ゲームと日本語教育の関係に焦点を当てた研究に麻(2008, 2019b)がある。麻(2008)は3種のゲームからそれぞれ延べ1,000語まで抽出する方法で語彙調査を行い、ゲームに出現した語彙の旧日本語能力試験における難易度を明らかにしようとしたが、調査単位が統一されていないなどの問題点がある。麻(2019b)はその改良版として『ドラゴンクエスト3』を対象に行った調査である。その結果の一部を表5に示す(麻2019bより引用)。

表5 旧日本語能力試験レベルの分布状況

	級外	1級	2級	3級	4級
異なり語数	1,003 (33.19%)	358 (11.85%)	889 (29.42%)	355 (11.75%)	417 (13.80%)
延べ語数	3,333 (11.23%)	1,205 (4.06%)	5,526 (18.62%)	5,485 (18.48%)	14,131 (47.61%)

異なり語数で見れば最も多く占めたのは級外の33.19%で、その次は2級の29.42%である。一方、延べ語数で見れば最も多いのは4級の47.61%であることが分かる。なお、最も難しい1級では異なり語数でも延べ語数でも低い割合を占めていることも読み取れた。また、名詞に関しては、「魔王」「勇者」「魔物」「魔法」「剣」のような、ゲームの世界観を反映させたものが多く、形容詞(イ形容詞)には「愛しい」「麗しい」「劳しい」のような文語の形容詞が多いという結論も同調査で得られた。

このように、ゲームには初中級の言葉が多い反面、出現した名詞と形容詞の頻出語が日常生活のそれとは異なる側面をもっており、特に教師のいない独学の学習者の場合は、目にした表現をそのまま覚えることがあるため、注意が必要である。この点は前述した役割語に関しても同じことが言える。これもゲームコーパスを調査してはじめて分かったことである。

3.3 その他の分析例

ほかにゲームコーパスに見られる言語表現に関する分析例として、二例挙げておく。

一つ目の例は「って」という表現の使用である。この表現は一般的に「引用する内容」を表す格助詞として用いられているが、ゲームの中には(7)(8)のような、日常生活であまり耳にしない使い方が存在する。

- (7) 戦争が、始まるかも知れないんだ**って**!? ホントウなのかい? あんた傭兵なんだから、何とかしておくれよ! (『聖剣伝説3』老婆)
- (8) ちょっと待って...いっしょに、ここから出よう。どうやってだ**って**? ふん、オイラに、まかせろ! (『聖剣伝説3』ケヴィン)

上記の例はいずれも主人公が別のキャラクターに話しかけた後に返ってきた返事である。「って」で表す引用の内容は、主人公が話しかけた内容の繰り返しだと思われるが、日常生活でわざわざ相手の発話内容を「って」で繰り返す人はあまりいない。それにもかかわらず、ゲームの世界ではこのような発話が多く見られる理由は何だろう。一つ考えられるのは、ゲーム内の発話は登場キャラクター間のコミュニケーションだけではなく、ゲーム制作者が登場キャラクターを通してプレイヤーに伝えるメッセージも含まれる、という解釈である。一般的に、プレイヤーに何らかの状況を説明するのに最も直接的な手段はナレーションだが、RPGは「プレイヤー自身が主人公である」という一人称視点の雰囲気を作るため、ナレーションと相性が悪い。残された手段は発話になるが、RPGは上記の一人称視点という理由で、主人公の発話を表示せず、代わりにプレイヤー自身にその内容を脳内補完させることが多い。つまり、プレイヤーは「話しかける」という行為を実行した後に、相手の返事内容を見て主人公の発話を想像するわけである。制作者がこのゲーム内特有の仕組みを利用して、プレイヤーに伝達したい情報を登場キャラクター経由で言わせている可能性がある。

これは、同じ「作られた話し言葉」である漫画や小説では観察しにくい現象である。漫画では、長方形の吹き出しで「ナレーション」、角を泡にした吹き出しで「心的描写」を表現することができる。小説でも「地の文」で類似した効果が得られる。漫画や小説では、作者がこれらの手段を用いて読者にメッセージを伝えることができる一方、ゲームの場合はそれがしにくいいため、「って」という表現を用いて同様の効果を狙うのである。

二つ目の例は、命令形の多用である。命令形は「命令」や「指示」を表す一種の待遇表現である。人間関係などの制限により、現実世界では命令形は限られた場面でしか用いられない。しかし、ゲームでは(9)～(11)のように、比較的多く現れている。

- (9) さあ! この水筒をもってとつと**いきやがれっ!** うぐっ! (『ドラゴンクエスト3』あらくれ)
- (10) よし、じゃあふたてに分かれよう。おまえらは、港に**急げ**。この町から脱出するんだ! (『聖剣伝説3』デュラン)
- (11) セリス将軍!! さあ、もう芝居はよい。そいつらの魔石を持ってこっちへ**来い** (『ファイナルファンタジー6』ケフカ)

ゲームはほとんど架空の世界観を持っており、シチュエーションや登場キャラクター間の人間関係も現実世界では実在しにくいものが多い。それが結果的に「命令形」を使いやすい環境に仕立て上げたのだと思われる。

このように、ゲームというフィクションの世界での言語表現は、リアルの世界のそれとは異なることがある。本節で取り上げた「って」、命令形、及び役割語は現実世界で多用すると、場合によって幼稚に見られ、不利益を被るおそれがある。特に日本語を独学している学習者は、教師によるフィードバックがないため、自分で判断することが困難である。こうした注意が必要な表現も、ゲームコーパスの調査を通して発見することが可能である。

4. ゲームコーパスの位置づけ

ここまで、ゲームコーパスが言語資源としての活用例を見てきたが、本節では、ゲームコーパスがコーパス言語学での位置づけを考える。結論を先に述べると、均衡コーパスの一部、対訳コーパス、マルチメディア・コーパスとして活用する可能性があると思われる。以下、このことについて少し触れておく。

4.1 均衡コーパスの一部として活用する可能性

最も均衡性が確保されている BCCWJ であっても、作業上などの理由でゲームが収録されていない。すでに1節で述べたように、確かにゲームの言語情報を電子化するのに手間がかかるが、ゲームに現れた言語表現は日本語の中に実在している。また、3節で見たように、ゲームはその特殊な性質で、ほかの言語使用域とは異なる用法も複数存在する。これらの理由を考慮し、今後、ゲームを均衡コーパスの母集団の一部として取り入れる検討も必要になってくると考えられる。無論、どのジャンルのどのゲームをどれほど対象に入れるかは、日本全国の制作本数を総合的に考慮し、慎重にサンプリングする必要があるが、日本語における言語使用実態を解明する一環としてやる価値はあると思われる。

4.2 対訳コーパスとして活用する可能性

ゲームコーパスは均衡コーパスとしてではなく、特別の目的を果たすために構築された簡易コーパスとしての役割もある。対訳コーパスがその一例である。ゲームは、ほかの言語作品と比べて、対訳コーパスの構築が比較的容易である。その理由は、人気なゲームは多言語版も発売されていることが多く、そして、ゲームは一度画面の枠に表示できる文字情報の量が限られているため、ページごとに保存されることが一般的である。そのページの情報を利用すれば、両言語のほぼ文単位での対応関係を得ることができる。

4.3 マルチメディア・コーパスとして活用する可能性

石井・孫 (2013) はテレビ放送のデータを対象に、マルチメディア・コーパスを作成し、指示詞と指差しのような言語表現と非言語表現との関係を調べた。これが現実世界を反映させたマルチメディア・コーパスであれば、ゲームがフィクションの世界を反映させたマルチメディア・コーパスという位置づけになる。前述したように、フィクションでは言語表現としてのステレオタイプが観察されやすいが、同じことは非言語表現に対しても言える。近年のゲームは、キャラクターの身振りや表情なども工夫を凝らして作られることが多く、非言語表現のステレオタイプを観察するのに利用することもできよう。

5. おわりに

本稿では、まず『ドラゴンクエスト3』というRPGを例にして、ゲームコーパスの構築方法を説明した。次に、ゲームコーパスを言語資源として活用する事例をいくつか紹介した。具体的に、役割語研究や日本語教育研究などの領域での利用例を見た。また、「って」のようなゲーム特有の言語的特徴も確認した。最後に、ゲームコーパスがコーパス言語学における位置づけを述べた。本稿が今後ゲームコーパス作成の参考、またはゲーム関連研究のきっかけになればと思う。

ただし、ゲームジャンルはRPGのみでなく、アクションゲームやパズルゲームなども存在する。本稿で提案した方法論で異なるジャンルのゲームコーパスを作成する際に、さらな

る調整が必要となるだろう。また、本格的にゲームコーパスを作成することになると、先行研究、例えば石川 (2012) でまとめられた、(1) 書き言葉や話し言葉などの現実の言語、(2) 大規模、(3) 基準に沿って網羅的・代表的に収集するもの、(4) コンピュータ上で処理できるデータとして保存するもの、(5) 言語研究に使用するもの、といったコーパスの観点を満たすように構築しなければならない。いずれも今後の課題とする。

文 献

- 石井正彦・孫栄爽 (2013) 『マルチメディア・コーパス言語学—テレビ放送の計量的表現行動研究—』大阪大学出版会
- 石川慎一郎 (2012) 『ベーシック コーパス言語学』ひつじ書房
- ウンサーシュッツ ジャンカーラ (2013) 『マンガにおける言語の役割—構造・語彙・登場人物という三つの観点から』一橋大学博士論文
- 金水敏 (2003) 『ヴァーチャル日本語 役割語の謎』岩波書店
- 金水敏 (編) (2014) 『<役割語>小辞典』研究社
- 国際交流基金 (2020) 『海外の日本語教育の現状 2018 年度日本語教育機関調査より』 <<https://www.jpff.go.jp/j/project/japanese/survey/result/dl/survey2018/all.pdf>>最終アクセス：2022 年 8 月 17 日
- 国立国語研究所 (2015) 『現代日本語書き言葉均衡コーパス利用の手引 第 1.1 版』 <<https://clrd.ninjal.ac.jp/bccwj/doc.html>>最終アクセス：2022 年 8 月 16 日
- 定延利之 (2007) 「キャラ助詞が現れる環境」金水敏 (編) 『役割語研究の地平』くろしお出版
- 麻子軒 (2008) 「テレビゲームが日本語教育の課外教材としての位置づけ—RPG を中心とした語彙調査の中間発表—」『大学院教育改革支援プログラム 平成 19 年度活動報告書 海外研修事業編』お茶の水女子大学, pp.245-249.
- 麻子軒 (2019a) 「計量的アプローチによる役割語の分類と抽出の試み—テレビゲーム『ドラゴンクエスト 3』を例に—」『計量国語学』32 巻 2 号, pp.103-116.
- 麻子軒 (2019b) 「テレビゲーム『ドラゴンクエスト 3』の語彙調査—日本語教育への応用可能性に向けて—」『日本語教育学会予稿集』2019 年度秋季大会, pp.131-136.

使用ゲーム

- スクウェア (開発元) (1994) 『ファイナルファンタジー6』スクウェア
- スクウェア (開発元) (1995) 『聖剣伝説3』スクウェア
- ハートビート・アルテピアッツァ (開発元) (1996) 『ドラゴンクエスト3』エニックス

日本語の動詞とその結合価

青山 文啓 (桜美林大学大学院)

Valency and the Japanese Verbal

Fumihiko Aoyama (Graduate Division, J. F. Oberlin University)

要旨

単語はその用法しか記述対象にできない。英語の五文型は構成要素単位で動詞の記述を試みるが、ここでは単語を記述の中心にすえ、日本語の動詞を例に取りあげる。用言は語幹と語尾に分割され、語幹の結合価(文型)と、語尾の担う三種の従属関係(連体、連用、終止)とは相互に関連し、従属関係は結合価の増減にもかかわる。結合価の記述例として、ペアをなす動詞のうち他動詞/自動詞に焦点をあてる。ヲ格/ガ格に体言が共有される動詞ペアに着目すれば、体言が結合価の決定に大きくあずかることが分かる。結合価は(a)用言がいくつ体言を取り、(b)どのような体言が、(c)どのような助詞に仲介されるかを記述する。どの言語も体言は最大のメンバーを誇り、(b)の様な記述は望めない。すべての単語が一度は体言であり、固有名詞、外来語、複合名詞、専門用語、短縮語、略語、メタファーすべてを扱える記述モデルは虹の向こうにある。用言のなかに体言を囲いこむ対処法について検討する。

1. はじめに

小論では、結合価(文型)を想定しなければ説明できない、傍証をいくつか提示することを目的とする。前提として、日本語で単語が音形とアクセント型で外形を定義されるという基本的なことがらを確認する。言うまでもなく、印刷物としての辞書(dictionary)では単語をどのような見出し語形として登録するかを決めない限り、その見出し語に品詞情報さえ振ることはできない。

2. 結合価を想定すべきか？

ここでは結合価を想定しなければ説明のできない問題を二つ取りあげよう。例えば〈ココガイタイ〉と〈ココニイタイ〉とは助詞が異なるだけである。しかし、この助詞の違いをもとに私たちは頭のなかの辞書(mental lexicon)を引き、結合価を割り出し〈ココ／ガ／イタイ〉〈ココ／ニ／イ／タイ〉のようにべつべつの語形に分割する。イタイもイルも助詞ガを取りうるが、助詞ニを取るのはイルだけである。さらに結合価は、二つの文に現れるココの解釈にもおおきな差異を強制する。つまり、最初のココは身体部位だが、二つめの例では話し手の存在する空間を指す。こうした解釈の違いも、イタイとイルの語幹がそれぞれ有する結合価によりもたらされるものである。

当面、ある用言の漢字仮名表記を無視して、音形とアクセント型が活用を通じておなじであれば、同一の単語と考える——例えば「書く」「搔く」「欠く」はおなじ単語の異表記である可能性がある。しかし「欠く」にはペア「欠かす／欠ける」があり、前二者とは異なる。自他ペアがあるかないかは動詞の用法を分類する場合、有力な目安である。

本題にもどろう。結合価を想定すべき傍証としてつぎに昔話「一寸法師」をあげる。

(1) むかし、むかし、あるところに、おじいさんとおばあさんがありました。

【X】子どもがなかったものですから、【X】子どもがほしくて、ほしくて、明けても暮れても、このことばかり、【X】神さまにおねがいしておりました。[『日本むかしばなし集 1』坪田譲治 1975]

第一文には「おじいさんとおばあさん(と)」という並列された体言が現れるだけであり、第二文の三カ所のスロット【X】に入るのはこの体言でしかありえない。しかし「おじいさんとおばあさんニ 子どもガ ない」という結合価は「子ども」を【実子】と解釈させ、【おじいさんとおばあさんが 親ではない】という昔話特有のナンセンスな推論を呼び出す(例えば「公園ニ 子どもガ いる」と比較すると分かりやすい)。つまり用言に結合価を想定することは、無数にある体言と、その組み合わせから生まれる際限のない推論を、用言のもとに囲いこむための有効な方法である(さらに踏みこんで言えば、奥田靖雄の一連の仕事をそれに近づくための方法と見ても間違いとはいえないだろう：言語学研究会[編]1983)。

3. 用言から眺めた体言

どの言語でも体言はほかの品詞を圧倒しておおい。これは、無限の単語を有限個の類に収め、品詞表示を与えるだけではすまない問題である。『日本国語大辞典』(初版1976年完結)は44万語の見出し語総数を誇るが、編集に深くかかわった松井栄一(1979)によれば見出し語を主要品詞別に見れば以下のようなになるという(以下の要約版から主要品詞だけを抜粋する：『図説日本語』1982)。

『日本国語大辞典』	見出し語総数	438,357
	名詞【73.19%】	320,824
	動詞【5.72%】	25,087
	形容詞【1.00%】	4,381
	形容動詞【1.09%】	4,760
	副詞【1.25%】	5,483

つまり、体言(名詞)はつぎつぎ造られるが、体言が造語されるたびにあらたに用言が造られることはない。辞典とよく対比される百科事典には、体言以外に収録される単語はなく、大量に排出された複合名詞の集積場所とも見られる。

一面で、複合名詞の形成は、既存の用言に対応させるための補填手段でもある。例えばコロナが世界に蔓延して、大学の授業も会議もオンラインが一般化した。〈教室〉から複合名詞〈バーチャル教室〉を造り、〈教室〉が〈行く〉を始めとする用言のあいだに築いた資産を有効利用するための方策だとは考えられないだろうか。〈教室で転ぶ〉ことはありそうだが、〈バーチャル教室で転ぶ〉ことはコミカルな推論しか生まないはずである。

逆方向の例もあげておきたい。〈教室に行く〉とはいえても〈先生に行く〉とはいえず〈先生のところに行く〉という(田窪1984)。これは用言〈行く〉が人を指す体言〈先生〉に対して、場所と理解されたければ形式名詞〈ところ〉に従属するよう強制する例ともいえる。類例に〈先生の話を話す／君のことが好きだ〉〈バラのにおいをかぐ〉などがあるが、強制的度合いは用言による。主要部〈ところ〉ではなく、従属部〈先生〉と用言〈行く〉との関係に記述の焦点を絞るほうが、実りある成果が得られそうである(宮島2005)。

4. テキストから単語へ

ところで、辞書を編むまえに見出し語形を確定しなければ、品詞情報は与えられない(水谷1951, 1952)。(2a)は一語だが(2b)が二語であることは、(2c)とこれらのアクセント型まで比較すれば納得がいくところである。

(2a)【とても】大きな男だった。燃えるような緑の目をしていた。(『TVピープル』村上春樹1993)

(2b)エスパニャと争い、イタリアを狙うヴァロア朝のフランス王【とても】同様であった。(『略奪の海カリブ』増田義郎1989)

(2c)先生は気長に治すことが大事ですと笑っているが、彼【とて】治療に自信があるようには見えない。(『マリコ／マリキータ』池澤夏樹 1994)

もう一つ類似の例をあげよう。(.....新興勢力である HIS が【大手】に価格競争を挑み台頭した)では〈HIS ガ 大手ニ 価格競争ヲ 挑む〉という三つの体言を取る(この例文はつぎの論考による:藤原 2016)。この場合の〈大手〉を〈大手企業／大手航空会社〉の略と理解し、助詞ニは用言〈挑む〉の必須とする体言を仲介すると、解釈する非母語話者はかなり読める人だろう。しかし〈新型コロナの感染が再拡大してきましたが、【大手】を振ってマスクを外せる日がやってくるのでしょうか? :『東京新聞』朝刊 31JUL2022〉では「おおで」であって先の「お]おて」とはべつの単語である(閉じのブラケットでアクセント核を示した)。しかし〈大手を振って中小に入る〉ではどちらの単語にも対応する。

ついでに助詞ニのむずかしさをべつの観点から二つ指摘しておきたい。まず奥田(1956)のあげる以下の二例を引用しよう(表記の一部を漢字に改めた):

(3a) 親切に感謝する

(3b) 親切に世話する

つまり述語〈感謝する〉は必須の体言として〈親切〉を要求するが、〈世話する〉にとつて〈親切に〉は副詞としてしか解釈されない(接続助詞ノニに関して同様の指摘が可能である:青山 2013)。助詞ニが体言を必須とするか否かは個々の用言が決める専権事項である。もう一つ、三つの助詞ヲ、ガ、ニのうち、ニにおおきな負荷がかかることを見るために架空の箱根駅伝から例文をあげよう:

(4a) 東洋が青山学院に勝つ

(4b) 東洋に青山学院に勝ってほしい

〈ほしい〉が助動詞(補助用言)として使われれば〈東洋が〉を〈東洋に〉と助詞を入れ換え、自身の結合価に納めこの語順を維持するほかない。類例については次節でもふれる。

5. 自他ペアと“能格”

ここではカケ]ル/カカ]ルのように他動詞/自動詞の順に表示し、/kakéru/と/kakáru/を単純にペアと呼ぶ。しかし、ある動詞が「他動詞」か「自動詞」かは、その動詞がどのような助詞を介して体言を取るかどうかにかかっている、語形の問題とはいえない。このように、動詞を含む用言がどのような助詞を介して、どのような体言をいくつ取るかを記述目標に置く立場、あるいはそのような観点からなされた記述結果を「結合価」と呼ぶ。言い換えれば、自他の区別はもっとも原初的な結合価への視点と見ていいだろう。

具体例があると話が進めやすい。ヲ格にもガ格にも「時間をカケル・時間がカカル」でおなじ体言が立つが、「眼鏡をカケル・?眼鏡がカカル」「?霏をカケル・霏がカカル」のように体言をおなじくしないペアはめずらしくない。体言が共通する場合を自他ペアと呼ぶことには問題はない(英語圏では70年代前後に有力な研究者が、共有される体言を“能格”と呼んだが最近この用語法はほとんど顧みられることはない:Lyons 1968, Halliday 1970;青山 1983はこの用語法にならったものである)。しかし共有される体言を欠いたペアにまで、この用語を拡張することは混乱を増すばかりだろう。

この点を念頭においたうえで、それぞれの動詞をハイフンで語幹と語尾に分離し(5a)(5b)に示す。ハイフンの左側が語幹で、右側が語尾である。(5a)(5b)のように二つの語幹に共有される/kak/を語根と呼び、それぞれの語幹から共有部分を抜き去ったあとに‘é/ár’という形態ペアが残る。この組み合わせから他動か自動かが判別できるとされてきた。

(5a) /kaké-ru/

(6a) /sazuké-ru/

(5b) /kakár-u/

(6b) /sazukár-u/

しかし、「子どもをサズケル・子どもをサズカル」は(6a)(6b)に示すように、おなじ形態ペアでありながら両者ともヲ格体言を取る他動詞である。つまり形態情報より結合価が優先されなければならないことになる。

6. 結合価：単文と複文を区別するために

最初に単純な定義から出発しよう。一つの用言とそれが要求する一つ以上の体言からなる文を「単文」と呼ぶ。これに対して、二つ以上の用言とそれぞれが要求する一つ以上の体言からなる文を「複文」と呼ぶ(そして二つの用言は従属か、並列かの関係に立つ)。複文では、複数の用言が体言を共有する場合がある。このように考えたとしても、単文と複文のあいだに一線を引くのは、なかなかむずかしい。

そのむずしさは用言が活用することに由来する。例えば「一体に監察医の鑑定は、死後の時間をなるべく内輪に見がちだということだ」(『黒い手帖』松本清張、中公文庫 1974)では「見る」の連用形が命題を形成するが、主語の「監察医の鑑定」はこの動詞にも文末に現われる助動詞ダにも従属する。例えば、ある単語が用言であるか否かは結合価と活用という二つの性質によって支えられているが、日本語では活用が用言の結合価を変容させて助詞や助動詞への道を歩ませる。実際、私たちがある文を単文と認定するか、複文と認定するかは結合価によって支えられる命題の数が決め手になる。大切なことは、結合価とはこの“命題”を単語の連鎖として捉えたものと読み換えられることである。問題点をすべてリストするには程遠いが、これまで思いついた単文と複文との境界例を以下にあげる。それぞれ(7)は動詞-助動詞/動詞-動詞(つまり複文)、(8)は複合動詞/動詞-動詞、(9)は複合助詞/動詞-動詞、(10)(11)(12)は註釈句/動詞(形容詞)-動詞の境界例である。

■簡単に解説を加える。日本語の文法は動詞と助動詞(補助動詞)の一線をあいまいなままにしてきた。(7a)のクレルは助動詞と考えるしかない。この場合の〈紙に〉をクレルの直前に移動することはできないからだ。しかし〈私に〉は(7a)(7b)両様の解釈を許す。(7b)のように〈私の体に〉という解釈と、あくまでも受益者という(7c)の解釈である。

(7a) [彼は[[紙に地図をかいて]くれ]た](『深夜特急 2』沢木耕太郎 1994)

(7b) [彼は[[私に地図をかいて]くれ]た]

(7c) [彼は[私に[地図をかいて]くれ]た]

(7d) [彼は[[紙に地図をかいて]私に]くれ]た。

しかし(7c)のクレルは(7d)のように〈私に〉を移動させられるため、はっきりと動詞は二つあることになり(7c)(7d)はともに複文である。つまり語幹の持つ結合価情報は命題の認定に不可欠である。

■複合動詞のなかにはそれぞれが述語として、目的語を共用しうる場合がある。べつべつに自立すれば、当然アクセント型は単一語のそれからは変更される(cf. 長嶋 1976)。

(8a) ラスコリニコフは斧で老婆をなぐり殺した。

(8b) ラスコリニコフは老婆を斧でなぐり殺した。

(8c) ラスコリニコフは斧でなぐり老婆を殺した。

(8d) ラスコリニコフは老婆をなぐり斧で殺した。

最後の(8d)はそのほかの例文が共有する同義性からは遠く離れてしまう。複合動詞は一語であり複文とは違って、時間的な前後関係を表示できないからである。

■ニョッテとニョルトにはいろいろな漢字の書き分けが考えられるが、主語を取り得る(9c)(9e)だけが動詞である。そのほかは複合助詞だが受動と呼応するのは(9b)だけである。

(9a) 喫茶店によって、コーヒーの味はまちまちだ。(青山 2000)

(9b) 消息筋によって、この情報は伝えられた。

(9c) 【私は】喫茶店によって、トルコで地震が起きたことを知った。

(9d) 消息筋によると、地震は午前一時ごろ起こった。

(9e) 【私が】喫茶店によると、テレビはちょうどそのニュースを報じていた。

■ 註釈句(10a)が特異なところは、主語は話し手以外に考えられないのに、それが明示できないところである。

(10a) はっきりいって、ぼくは大学の将来が不安だ。(cf. 杉戸 1983)

(10b) 【彼女に】はっきりいって、自分の誠意を示せよ！

形容詞からつくられる註釈句もおなじく主語は表面には現れない。

(11a) 悪いけど、そろそろ行かなくっちゃ。

(11b) 【天気は】悪いけど、そろそろ行かなくっちゃ。

(12) くどいようだけれど、あたしは全く時間には遅れていなかったのだ。(『星影のステラ』林真理子 1986)

■ つぎの(13)は見逃されやすい表現型だが、問題の句〈私を相手に(して)〉は、どちらの場合も独自の主語を取らない(cf. 村木 1983, 1991, 3部2章; 寺村 1983; さらに、村木ほか 1984)。両方とも単文内部の問題と捉えるべきである。

(13a) その男は私を相手にして冗談ばかり言っていた。

(13b) その男は私を相手に冗談ばかり言っていた。

7. まとめ

結合価は用言を中心に体言の連鎖を見たものであり、命題を単語の連鎖として見なすことに近似する。つまり複文から単文を区別して命題を取り出すには結合価を想定する以外にない。小論では結合価を想定しなければ説明のできない用例を列挙したが、先行研究を咀嚼し改訂の機会があればもう少し読みやすいかたちに整理できればと思う。

謝 辞

本研究の一部は、国立国語研究所のプロジェクト「学習者辞書用語彙資源の構築」による。小論をまとめるにあたり、いろいろなかたにお世話になった。現状は未完成なままでありここに名前をあげるのはもうすこし論述を整理してからにしたい。

文 献

- 青山文啓(1986) 体言と用言の結ぶ二つの関係：能格性と対称性、『ソフトウェア文書のための日本語処理の研究』7 (情報処理振興事業協会)
- 青山文啓(1987) 料理の文章における提題化の役割、『計量国語学と日本語処理』水谷静夫教授還暦記念会[編], 秋山書店
- 青山文啓(1998) 二重主語構文と辞書、『言語』27(3)
- 青山文啓(2000) 活用と統語—日本語とスペイン語との比較から、『日本語と外国語との対照研究VI：日本語とスペイン語(3)』くろしお出版
- 青山文啓(2000) 統語論—単語の二重分節を中心として、『一橋論叢』124(4).
- 青山文啓(2004a) ことばの研究と辞書に記載される情報、『桜美林論叢』31: 37-45.
- 青山文啓(2004b) 日本語の正書法へのささやかな願い、『日本言語政策学会会報』4.
- 青山文啓(2013) 二つの四階層モデル—『文型』と『構造』のための読書ノート、『基本文型の研究』林四郎, 明治図書 1960, 復刊: ひつじ書房
- 青山文啓(2016) ハルキのレトリック, ルービンのツボ, 『ことばと文字』6
- 青山文啓(2019) 鈴木孝夫とその言語圏, 『ことばと文字』11

- 青山文啓(2020) 『雪国』冒頭部の翻訳をめぐって—中島文雄『日本語の構造』を機縁として(1987年), 『ことばと文字』13
- 奥田靖雄(1956) ことばの組み立て, 『講座日本語I: 民族とことば』大島義夫[編], 大月書店
- 亀井孝/河野六郎/千野栄一[編](1996) 『言語学大辞典 6: 術語編』三省堂
- 言語学研究会[編]1983 『日本語文法・連語論(資料編)』むぎ書房
- 情報処理振興事業協会(1990) 『計算機用日本語基本形容詞辞書 IPAL(Basic Adjectives)』同技術センター
- 須賀一好/早津恵美子[編](1995) 『動詞の自他』ひつじ書房
- 杉戸清樹(1983) 待遇表現としての言語行動—「注釈」という視点, 『日本語学』2(7)
- 田窪行則(1984) 現代日本語の「場所」を表す名詞類について, 『日本語・日本文化』12
- 寺村秀夫(1983) 「付帯状況」表現の成立の条件—XをYに……する—という表現をめぐって, 『日本語学』2(10): 38-46.
- 長嶋善郎(1976)複合動詞の構造, 『日本語講座 4: 日本語の語彙と表現』鈴木孝夫[編], 大修館書店, 再収: 須賀/早津[編](1995)
- 橋本三奈子/青山文啓(1992) 形容詞の三つの用法: 終止, 連体, 連用, 『計量国語学』18
- 林大/宮島達夫/野村雅昭/江川清/中野洋/真田信治/佐竹秀雄[編](1982) 『図説日本語』角川書店
- 藤原未雪(2016) 中国語を母語とする上級日本語学習者が学術論文を読むときの困難点, 『日本語/日本語教育研究』7
- 水谷静夫(1951) 形容動詞辨, 『国語と国文学』28(5); 『日本の言語学 4: 文法II』服部四郎ほか[編], 大修館書店 1979.
- 水谷静夫(1952) 形容動詞と謂ふもの, 『国文学 解釈と鑑賞』17(12)
- 宮島達夫(1972) 『動詞の意味・用法の記述的研究』(国語研報告#43)秀英出版
- 宮島達夫(1996) カテゴリー的多義性, 『日本語文法の諸問題—高橋太郎先生古希記念論文集』鈴木泰/角田太作[編], ひつじ書房
- 宮島達夫(2005) 連語論の位置づけ 『国文学解釈と鑑賞』70(7)
- 村木新次郎(1983) 「地図をたよりに, 人をたずねる」という言いかた, 『副用語の研究』渡辺実[編], 明治書院
- 村木新次郎/青山文啓/六条範俊/村田賢一(1984) 辞書における格情報の記述, 『情報処理学会自然言語処理研究会資料』46-3
- 村木新次郎(1991) 『日本語動詞の諸相』ひつじ書房
- Halliday, M. A. K. (1970) Language structure and language function. In: *New Horizons in Linguistics* ed. by John Lyons, London: Penguin Books. pp.140-165. ([『現代の言語学(上)』田中春美ほか[訳], 大修館書店, 1973)
- Lyons, J (1968) *Introduction to Theoretical Linguistics*. ([『理論言語学』國廣哲弥ほか[訳], 大修館書店, 1973)

英語学習用活用語彙リストの提案 — CEFR-J Wordlist のコロケーション・データセットの試み —

福田 航平 (東京外国語大学) †
投野 由紀夫 (東京外国語大学) ††

Constructing a Collocation Database for the CEFR-J Wordlist

Kohei Fukuda (Tokyo University of Foreign Studies)
Yukio Tono (Tokyo University of Foreign Studies)

要旨

外国語学習において、語彙を運用する力を学習者が身につけるためには語彙リストで単語を個別に記憶するだけでなく、使用頻度の高いフレーズで提示するのが重要である。近年、コーパス準拠英語語彙・フレーズリストが発表されているが、対象学習者レベル、受容・産出語彙の区別は明確ではない。本研究では、CEFR 準拠英語学習語彙表 CEFR-J Wordlist の活用度を上げ、学習に資するコロケーション選択するための資料となるコロケーション・データセットを整備する。1億語のイギリス英語均衡コーパスである British National Corpus から Universal Dependencies に基づいた構文解析を行い、共起フレーム別 (例: 動詞+名詞、形容詞+名詞) に共起語セットを抽出した。教育的に有用なコロケーションを教師や学習者が選定するための情報源として、単純共起頻度以外に検索語-共起語ペアについて、各語の CEFR-J Wordlist に基づいた CEFR レベル情報、共起統計 (MI, MI2, MI3, t-score, z-score, logDice, log-likelihood, chi-squared) と散布度指標 (DP) の情報を付与した。

1. はじめに

コロケーションは外国語学習において重要な側面であることは広く認識されているが、学習者にとって習得が困難な領域である (Martinez and Murphy, 2011)。コロケーション学習における問題の一つは、学習者が習得すべきコロケーションを判断する基準の合意がないこと (Rogersa, Brizzarda, et. al., 2015)、また個々のコロケーションの学習レベルに関する研究が少ないことである。本研究では、コロケーションの学習レベルを CEFR 準拠英語学習語彙表を活用して付与すると同時に多様な共起統計を付属したコロケーション・データセットを整備することで学習すべきコロケーション特定のための言語教育資料を提供することを目的とする。

2. コロケーション

2.1 コロケーションとは

コロケーションは、最も簡単には「語の自然な組み合わせ」¹と定義される (McCarthy & O'Dell, 2017, p.4)。コロケーションは研究によって様々な定義がなされているが、大きく2つの方向性に分類できる。構成性に焦点を当てたフレイジオロジー的アプローチと、コーパス情報に基づく頻度情報に焦点を当てた統計的アプローチである (Evert, 2008; Granger &

† fukuda.kohei.r0[at]tufs.ac.jp ※[at]は@に置き換え

†† y.tono[at]tufs.ac.jp

¹ 原文は“a natural combination of words”

Paquot, 2008; Gablasova, Brezina & McEnery, 2017)²。前者のアプローチでは語結合の構成性 (compositionality) に着目し、自由結合句 (free combination) とイディオムの中に位置するものとしてコロケーションを捉える。この見方では自由結合句とコロケーションとイディオムは連続体として考えられ、客観的にコロケーションとそうでないものを線引きして区別することはできない。後者のアプローチでは、コーパス内での共起頻度や、後述する共起統計指標を用い、数値化してコロケーション性を判断する。これにより一定の数値基準を設け、その基準を満たしているかどうかでコロケーションとそうでないものに区別することができる。設定する基準はその目的に応じて変わり、ある程度恣意的なものとならざるを得ないが、妥当な基準を設定できれば、比較的客観的かつ効率的なコロケーション選択が可能となる。本研究では教育的に有用なコロケーションをその用途に応じて客観的に判断するための統計情報を提供するのが主目的であるため、後者の統計的アプローチでコロケーションを捉える。

2.2 フレーズリスト研究

英語教育分野では語彙表の研究は長い歴史を持ち、多くの教育用語彙表が提案されてきた (中條, 2015)。その後は1つ1つの単語の振る舞いを学習辞典の形式でまとめる工夫が多く提案され、その中でコロケーションもコーパス研究の成果を活かして積極的に辞書記述に取り入れられていった。日本では戦前にすでに勝俣銓吉郎による『英和活用大辞典』(1939)があり、欧米では BBI Dictionary (revised edition) of English Word Combinations (1997)、the LTP Dictionary of Selected Collocations (1999)、Oxford Collocations Dictionary for Students of English (2009)、Longman Dictionary of Collocations & Thesaurus (2013) などがコロケーションの辞書としては知られているが、一方で Academic Formulas List (Simpson-Vlach & Ellis, 2010)、Phrasal Expression List (Martinez & Schmitt, 2012) のような語彙表と同じようなコロケーションやフレーズのリストの提案は比較的最近のことである。特にこれらの資料はコーパスからの機械的な n-gram 情報とそれらをフレーズとしての機能やまとまりなどの観点から人手で評価する情報を合成している。こういった合成の判断基準も研究によってまちまちであり、ある意味で異なる情報の重み付けは恣意的に行われているのが現状である。

3. データ整備

3.1 共起語抽出

コーパスからコロケーションを抽出する方法には大きく4つの方法がある (Bhalla & Klimcikova, 2019)。(1) ウィンドウスパンを用いた方法、(2) n-gram、(3) 正規表現と POS タグを用いた方法 (regex-over-pos)、(4) 構文解析による統語依存関係 (syntactic dependency) を利用した方法である。(1) のウィンドウスパンを用いた手法は中心語 (node) の前後数語に共起する語を抽出する手法で、英語コーパス研究で最もよく用いられる手法である。Sinclair (1991) によるコロケーションの定義もこの手法を前提としている³。(2) の n-gram は任意の語数の隣接する句を機械的に抽出する手法である。(3) は品詞タグと正規表現によるパターン抽出によって特定の文法関係にある語を抽出する手法で、コーパス検索 Web

² Evert (2008) では、後者によるコロケーションを“empirical collocations”と呼んでいる。

³ Sinclair (1991) ではコロケーションを“the occurrence of two or more words within a short span of each other in a text” (p.170) と定義している。

ンターフェースである Sketch Engine の Word Sketch 機能で利用されている (Kilgarriff, Bida et al., 2014)。

コロケーション抽出方法を比較した研究では、(4) の手法が最も正確にコロケーション抽出をすることができるという結果が報告されている (Bartsch, Evert et al., 2014; Bhalla & Klimcikova, 2019; Evert, Uhrig et al., 2017)。従って、本研究では大規模コーパスに構文解析を行い、統語依存関係に基づき共起フレーム別に共起ペアを抽出した。今回は Python の自然言語処理ライブラリである Stanza (Qi, Zhang et al., 2020) を利用し構文解析を行った。Stanza の構文解析器は Universal Dependencies treebanks を用いて訓練されており、統語依存関係の解析結果は Universal Dependencies に基づいた表示がなされる (de Marneffe, Manning et al., 2021)。

今回のデータ整備に当たっては、British National Corpus XML edition (BNC Consortium, 2007) のテキストデータを用いた。テキスト全体に対して構文解析を行い、解析結果を用いて (a) 動詞+目的語、(b) 形容詞+名詞、(c) 名詞+名詞、(d) 副詞+動詞、(e) 副詞+形容詞の共起ペアを抽出し、全ての共起ペアの頻度を算出した⁴。なお全てレマ単位で抽出・集計している。

3.1.1 動詞+目的語の抽出

Stanza での構文解析結果を用い obj の係り受け関係 (dependency relation) で結ばれている主辞 (head) と依存部を「動詞+目的語」のペアとして抽出した。また、受動態の動詞とその主語、名詞を修飾する過去分詞とその被修飾語も動詞+目的語のペアとして抽出するために、nsubj:pass の係り受け関係で結ばれている主辞と依存部、acl の係り受け関係で結ばれている主辞と依存部 (依存部の語の品詞が xpos で VBN のものに限定) も動詞+目的語の共起ペアとして抽出した。また、conj の依存関係タグを用いて、1つの動詞に対して複数目的語が並列されている場合も考慮した。目的語となる名詞は upos のタグが NOUN のものに限定し、代名詞や固有名詞は含んでいない。データセットでは obj のシートにこの共起フレームのデータを収録している。

以下に BNC 中の実際の文から例を示す。1つ目の括弧は BNC のジャンルとファイル名である。2つ目以降の括弧には該当文から抽出される共起ペアとその際の係り受け関係タグである。

- (1) The pictures look like war photos and the **costs** will not easily be **assessed** but the first estimates to **repair** the **damages** run between \$2.5 and \$4 million. (W misc; J1B) (nsubj:pass; assess + cost) (obj; repair damage)
- (2) The firm that **makes** the **ties sold** in the country's department stores, is optimistic about prospects at home. (S brdcast news; K6H) (obj; make + tie) (acl; sell + tie)
- (3) If anyone could **post** the half-time **score** and final **result** when it comes through I'd appreciate it. (W email; J1G) (obj; post + score) (conj; post + result)

⁴ 詳しい処理は Python コード参照のこと。

- (4) They therefore **put** forward a wide **spectrum** of policies to **cover** all politically significant **aspects** of national life, as well as foreign **affairs**. (W ac polit law edu; GV5) (obj; put + spectrum) (obj; cover + aspect) (conj; cover + affair)

上記の(1)～(4)は今回の抽出方法で抽出可能な動詞+目的語のペアの一例を示している。

(3)のように1つの動詞に対して複数目的語が付いていて、目的語となる名詞が比較的離れている場合でも抽出可能である。(4)のように *as well as* で結ばれた **aspects** と **affairs** を両方とも **cover** の目的語として抽出することができている。もちろんこの抽出方法でも完璧ではなく、(4)の例では **put** と **spectrum** が動詞+目的語ペアと判定されている。これは実際には **put forward** という句動詞の目的語であると判定すべきあり、さらには意味内容を考慮すると、**spectrum** よりも **policy** を目的語とする方が正確である。しかしこれはウインドウスパンを用いた抽出方法でも同様であり、例えば名詞を中心語 (node) として左5右5のスパンで共起する動詞を抽出した場合、**spectrum** の共起語として **put** と **cover** が抽出されることになる。以上のことを考えると統語依存関係に基づいた共起語抽出の方が比較的過不足なく目的となる共起ペアを抽出することが可能であると考えられる。

3.1.2 形容詞+名詞の抽出

Stanza での構文解析結果を用い *amod* の係り受け関係で結ばれている主辞と依存部を「形容詞+名詞」のペアとして抽出した。名詞は *upos* が *NOUN* のものに限定し、代名詞や固有名詞は含んでいない。*conj* のタグを用い1つの名詞に複数の形容詞が並列されている場合も考慮している。データセットでは *amod* のシートにこの共起フレームのデータを収録している。

- (5) A **fate worse** than a cookbook. (W commerce; CBU) (*amod*; bad + fate)
- (6) The **thin, pale sunshine** gave **little warmth**, but it was a **welcome extension** of the autumn, and unusually fine for the **last day** of October. (W fict prose; GVT) (*amod*; thin + sunshine) (*amod*; pale + sunshine) (*amod*; little + warmth) (*amod*; welcome + extension) (*amod*; last + day)
- (7) It wasn't simply because he was ugly — he knew **men** equally **ugly** and ten times more **lovable**. (W fict prose; GVT) (*amod*; ugly + man) (*conj*; ugly + lovable)
- (8) **Individual, social** and **historical contexts** are also key. (W misc; G34) (*amod*; individual + context) (*conj*; social + context) (*conj*; historical + context)

(5)の例のように形容詞が後ろから前の名詞を修飾する場合も *amod* の係り受け関係で抽出可能である。(6)の **thin, pale** のように2つの名詞が連続して1つの名詞を修飾している場合も *amod* の係り受け関係で両方抽出可能である。(7)の例では *amod* と *conj* のタグを用いることで **ugly** と **lovable** を **men** に対する形容詞修飾として抽出に成功している。(8)のように1つの名詞に対して3つの形容詞が修飾している場合も3つも抽出することができる。

3.1.3 名詞+名詞の抽出

Stanza での構文解析結果を用い compound の係り受け関係で結ばれている主辞と依存部を抽出した。主辞・依存部ともに upos で NOUN のタグが付いているものに限定している。これによって下記の (9) のように名詞+名詞で 1 つの表現を成しているペアを抽出可能である。データセットでは nounmod のシートにこのデータを収録している。

- (9) One arises out of the need for **state provision** of education to be efficient and economical, which means that **policy considerations** might outweigh individual parental preference. (W ac polit law edu; AN5) (compound; state + provision) (compound; policy consideration)

3.1.4 副詞+動詞の抽出

Stanza での構文解析結果を用い advmod の係り受け関係で結ばれている主辞とその依存部を抽出した。advmod のタグは副詞が動詞を修飾する場合だけでなく形容詞や他の副詞を修飾する場合も含むため、主辞の upos タグが VERB のものに限っている。依存部の語の品詞は where などのような疑問副詞が入らないように、xpos のタグが RB, RBR, RBS のものに限っている。conj の依存関係を用いて 1 つの動詞に対して複数の副詞が修飾する場合も考慮している。データセットでは advmod_verb のシートにこの共起フレームのデータを収録している。

- (10) Saturday, 13th: **Arrived** at Delhi airport over two hours **late**, at 06.30. (W misc; KAN) (advmod; late + arrive)

- (11) If she opts for cardiocography it must be **done properly** and not **haphazardly**. (W ac medicine; EA2) (advmod; do + properly) (conj; do + haphazardly)

3.1.5 副詞+形容詞の抽出

Stanza での構文解析結果を用い advmod の係り受け関係で結ばれている主辞のうち upos が ADJ のものと、その依存部を副詞+形容詞の共起ペアとして抽出した。副詞+動詞の場合と同様に副詞の品詞は RB, RBR, RBS に絞っている。conj の係り受け関係を利用して 1 つの形容詞に対して複数の副詞が修飾する場合も考慮している。データセットでは advmod_adj のシートにこの共起フレームのデータを収録している。

- (12) The action of the sea, however, is concentrated within **comparatively narrow** limits. (W ac nat science; GV0) (advmod; comparatively narrow)

- (13) **Initially** the field-worker's relations with respondents in the field were **especially** and **unusually warm**, which runs counter to the norm in ethnographic research. (W ac polit law edu; A5Y) (advmod; initially + warm) (advmod; especially + warm) (conj; unusually + warm)

(13) の例では initially と warm を副詞+形容詞のペアとして抽出している。それと同時に especially と unusually も warm を修飾する副詞として抽出することに成功している。

3.2 共起統計指標

コーパスを用いた統計的アプローチでのコロケーション抽出では広く共起統計指標 (association measures; AM) が用いられる⁵。これまで様々な AM が提案されており、どの AM がコロケーション抽出に適しているか研究がなされている (e.g. Evert, 2004)。教育目的のコロケーションリスト作成においても、コロケーション候補抽出の際に AM を用いて基準設定をすることが多い。例えば、Ackermann & Chen (2013) はアカデミック英語用 (EAP) のコロケーションリスト作成において、MI と t-score を用いて基準を設定し、一定の値を超えたものをリストに入れる候補にしている。

AM にはそれぞれ特徴があるため目的に応じて使い分けるべきである (Brezina, 2018)。例えば MI や z-score は低頻度語を構成素に含む共起ペアを評価しやすい一方で、t-score は高頻度語を構成素に含む共起ペアに対して高い値を返す (Evert, 2008; 石川, 2008; Brezina, 2018)。本研究ではデータセットの利用者がその目的に応じて必要な AM を選択できるように全ての共起ペアに対して MI, MI2, MI3, t-score, z-score, logDice, log-likelihood, chi-squared の 8 つの AM を算出し情報を付した⁶。

AM の算出には共起頻度、各構成素の頻度、コーパス総語数が必要である。本研究での共起語抽出は統語依存関係による共起語抽出⁷であるため、Evert (2008) に習い、各構成素の頻度はコーパス全体における当該語の総頻度ではなく、共起フレームごとに見て、その共起フレーム内に出現する頻度になっている。データセットでは、w1_in_rel と w2_in_rel の列が各共起フレーム内での構成素頻度を表している。コーパス総語数は、統語依存関係による共起語抽出では共起フレームの出現回数である。表 1 は共起フレーム別の出現総頻度である。

表 1 共起フレーム別総頻度

共起フレーム	総頻度
動詞+目的語 (obj)	4,633,147
形容詞+名詞 (amod)	5,188,122
名詞+名詞 (nounmod)	2,295,774
副詞+動詞 (advmod_verb)	2,948,140
副詞+形容詞 (advmod_adj)	890,696

3.3 散布度指標

教育的に有用な語句を抽出するにあたっては頻度のみではなく散布度 (dispersion) も考慮することが重要である。これは単語だけでなくコロケーションにも同様に言えることである (Rogersa, Brizzarda, et al., 2015)。共起統計指標は全て頻度に基づいて算出したものであるため、頻度とは別の軸で教育有用度を測る 1 つの指標として散布度の指標を用いることは重要である。散布度の指標を用いることで、限定的な場面でしか使用されない表現ではなく、様々な場面で使用されるコロケーションを選択することができる。しかしながら、著者の知る限りでは散布度指標を明示したコロケーション・データセットはほとんどない。従っ

⁵ association measures はより一般的に「相関度指標」「関連度指標」とも訳される。

⁶ 各指標の詳しい計算方法については Evert (2008) を参照。logDice については Rychlý (2008) または、恒川 (2020) を参照。

⁷ Evert (2008) ではこれを 'syntactic cooccurrence' と呼んでいる。

て、本研究のデータセット整備では、教育的に有用なコロケーションを選択するために重要な情報として、散布度指標を算出し明示した。

本研究では散布度指標として DP (Deviation of Proportions) を用いた⁸。DP はコーパス内のサブコーパスやファイルごとの語数を考慮に入れた上で、各サブコーパスやファイル間で、特定の語句がどれだけ均等に用いられているかを数値化した指標である。0 から 1 の値を取り、0 に近いほど均等に用いられていることを示す。BNC は各ファイルや各ジャンルの語数が均等ではないことと、膨大なデータを処理しなければならないことを考慮し、算出が比較的簡素である DP を本研究では取り入れた。

DP を算出するにあたっては、コーパス分割の基準を定める必要があるが、今回は BNC のジャンル分けに従った。British National Corpus XML edition では classCode のタグ内に各ファイルのジャンルが示されている。タグを参照しジャンルごとに処理することで、各ジャンルの総語数と、各ジャンルにおける共起ペアの共起頻度データを取り出し DP を算出した。

3.4 CEFR-J Wordlist の利用

今回のコロケーション・リストを作成する際の学習レベル情報は、ヨーロッパ言語共通参照枠 (CEFR) を日本の英語教育に適用した CEFR-J プロジェクトで構築された CEFR レベル別語彙表 CEFR-J Wordlist Version 1.6⁹を用いた (投野, 2013)。これはアジア圏の英語教科書分析および海外の CEFR 語彙表データを比較して学習語彙表として整備されたもので、A1 から B2 まで 7801 語 (品詞別の語数、表層形では 6868 語) を選定している。CEFR-J Wordlist を用いることで、共起ペア内の単語の CEFR レベルを明示することができ、有益なコロケーションが 7801 語の語彙表内の単語でどの程度カバーできるか、語彙表の単語を身につけることの生産性と、コロケーションによる活用度・有用度の向上を目指した。

4. データセットの概要

本節では今回整備したコロケーション・データセットの概要を説明する。本データセットの中に含まれる情報は、図 1 の列名で示されるように各共起ペアの構成素 (w_1, w_2)、構成素の CEFR-J レベル (w_1_CEFR, w_2_CEFR)、共起フレーム (relation)、共起頻度 (cooccurrence)、共起フレーム内での各構成素の頻度 ($w_1_in_rel, w_2_in_rel$)、散布度指標 (DP)、期待値頻度 (expected_freq)、共起統計指標 (MI 以降) である。全部で 5 通りの共起フレーム (amod, obj, nounmod, advmod_verb, advmod_adj) を各シートに分けている。扱いやすいデータサイズにするために、収録しているコロケーションデータは共起頻度 5 以上かつ、共起頻度が期待値頻度以上 ($cooccurrence \geq expected_freq$) のものに限定している¹⁰。本データセットに収録されている共起ペア総数は表 2 の通りである。

⁸ DP についての詳しい説明や計算方法は Gries (2008) を参照。

⁹ http://www.cefr-j.org/download.html#cefrj_wordlist

¹⁰ これに限定しない全データを研究用に公開する。付録参照。

表 2 共起フレーム別 共起ペア数

共起フレーム	共起ペア数
動詞＋目的語 (obj)	114,582
形容詞＋名詞 (amod)	135,940
名詞＋名詞 (nounmod)	72,340
副詞＋動詞 (advmod_verb)	43,992
副詞＋形容詞 (advmod_adj)	16,180
合計	383,034

	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	w1	w2	w1.CEFR	w2.CEFR	relation	cooccurrence	freq.w1	freq.w2	w1_in_rel	w2_in_rel	DP	expected freq	MI	
2	last	year	A2	A1	amod	14280	78446	176485	59008	52433	0.388692426	596.355765	4.58167892	!!
3	first	time	A1	A1	amod	8622	124338	189496	78511	56321	0.205860426	852.2964632	3.338595312	!!
4	last	night	A2	A1	amod	8575	78446	38745	59008	13826	0.608845761	157.2523946	5.76898278	!!
5	same	time	A1	A1	amod	7641	61855	189496	48995	56321	0.162705428	531.8778924	3.844594483	!!
6	local	authority	A2	B1	amod	7152	47279	28451	43662	14385	0.525139906	121.0607364	5.884543799	!!
7	long	term	A1	B1	amod	6467	82846	46578	38307	19778	0.329249197	146.0327737	5.46873252	!!
8	next	year	A2	A1	amod	6455	45646	176485	35459	52433	0.352138152	358.3612234	4.170930652	!!
9	other	hand	A1	A1	amod	5578	182918	59458	128190	14167	0.279434234	350.0433741	3.994142331	!!
10	last	week	A2	A1	amod	5503	78446	48163	59008	14998	0.513578273	170.5823387	5.011678139	!!
11	many	people	A1	A1	amod	4966	89532	117876	64619	40160	0.247660507	500.2000801	3.311507068	!!
12	long	time	A1	A1	amod	4652	82846	189496	38307	56321	0.244187168	415.8515445	3.483710601	!!
13	other	people	A1	A1	amod	4535	182918	117876	128190	40160	0.200474189	992.2878452	2.192271964	!!
14	young	man	A1	A1	amod	4156	35531	93946	26878	31477	0.373086437	163.0722651	4.671612316	!!
15	great	deal	A1	A2	amod	4061	66218	28841	52014	7528	0.18531049	75.47266468	5.749737005	!!
16	nineteenth	century	-	A2	amod	3959	4207	27308	4058	24393	0.472953046	19.07950391	7.696968597	!!
17	young	people	A1	A1	amod	3762	35531	117876	26878	40160	0.369457095	208.0561097	4.176455289	!!
18	many	year	A1	A1	amod	3614	89532	176485	64619	52433	0.219059021	653.0625199	2.468303489	!!
19	high	level	A1	A2	amod	3592	64571	41241	44473	18363	0.345744296	157.4091162	4.512196349	!!
20	past	year	B1	A1	amod	3521	26267	176485	9267	52433	0.364867557	93.65558693	5.232476357	!!
21	large	number	A1	A1	amod	3500	49196	61044	42006	14692	0.345426329	118.9548264	4.878869209	!!
22	few	year	A2	A1	amod	3453	46810	176485	38943	52433	0.198886735	393.5717624	3.133151714	!!
23	other	word	A1	A1	amod	3411	182918	42876	128190	12459	0.35514389	307.8414906	3.46993516	!!
24	old	man	A1	A1	amod	3333	69388	93946	48501	31477	0.411698942	294.2617728	3.501649277	!!
25	next	week	A2	A1	amod	3328	45646	48163	35459	14998	0.488149559	102.5060864	5.020873955	!!

図 1 コロケーション・データセットのサンプル画像

5. データセットを用いた調査例

本研究で整備したデータセットは、学習者や教師が学ぶべきコロケーションを選択するための情報として役立つことを想定しているが、研究目的も含め、その目的に応じて様々な利用が可能であろう。本節では学習に有用なコロケーション選択のために使用することを念頭に置いたうえで、教育的に有用なコロケーションを抽出するために本データセットをどのように用いるのがよいか考えるための、簡易的な調査を行う。

コロケーションを選択するためには共起頻度や AM や DP の値を用いてソートした上で、その中から使用者が必要であると思われるものを選択するのが基本であるが、その際にどの AM を使用するのか、適切な選択が必要である。また、DP を用いた適切な基準値で境界線を設けることで、共起頻度や AM の値は高いが、特定のジャンルに偏っているものを除外することができる。つまり適切な AM の選択と DP の値の基準値を考える必要がある。従って以下の 2 点を調査する。

1. 本データセットを用いて教育的に有用なコロケーションを抽出するにはどの AM を用いるのが最適か。

2. 本データセットを用いた教育的に有用なコロケーションの抽出に当たって、基準値となる適切な DP の値はいくつか。

5.1 ゴールドスタンダード

コロケーション抽出の評価にはゴールドスタンダードを使った手法がよく用いられる (e.g. Evert, 2008; Bhalla & Klimcikova, 2019; Evert, Uhrig et al., 2017)。本研究は CEFR-J Wordlist 内の語の組み合わせでできるコロケーションを教育的に有用なコロケーションと定義する。今回は Oxford Collocations Dictionary を用いて (McIntosh, Francis & Poole 2009)、この辞書に記載があり、かつ CEFR-J Wordlist に掲載されている語同士の組み合わせのコロケーションをゴールドスタンダードと仮定する。

今回はデータセットの形容詞修飾 (amod) と名詞修飾 (nounmod) を対象とする。データセット中の amod と nounmod の共起ペア総数は 205,310 である。Oxford Collocations Dictionary では名詞+名詞と形容詞+名詞を区別していないため、この二つの共起フレームを同時に扱う。具体的には Oxford Collocations Dictionary の名詞を見出し語とするエントリーの中の ADJECTIVE に分類されている共起ペアを対象とする。これに該当するコロケーションを Oxford Collocations Dictionary から全て取り出したうえで、構成単語の片方もしくは両方が CEFR-J Wordlist にないものを削除し、残ったものをゴールドスタンダードとした。ゴールドスタンダードとなったコロケーション数は 38,362 であった。

5.2 AM と DP の評価

上記の手順で作成したゴールドスタンダードと作成したデータセットを用いて、n-best リスト (共起頻度や特定の AM の値順に並べ一定の数で区切ったもの) 中の適合率 (precision; n-best リスト中のゴールドスタンダードの割合)、再現率 (recall; n-best リスト中のゴールドスタンダードの数の、全ゴールドスタンダードに占める割合) を算出した。AM の中から共起頻度 (cooccurrence)、MI、MI3、t-score、z-score、logDice、log-likelihood をそれぞれ使い、precision と recall を算出した。precision と recall のグラフを DP による基準値別に表したものが図 2-1 から図 2-6 である。

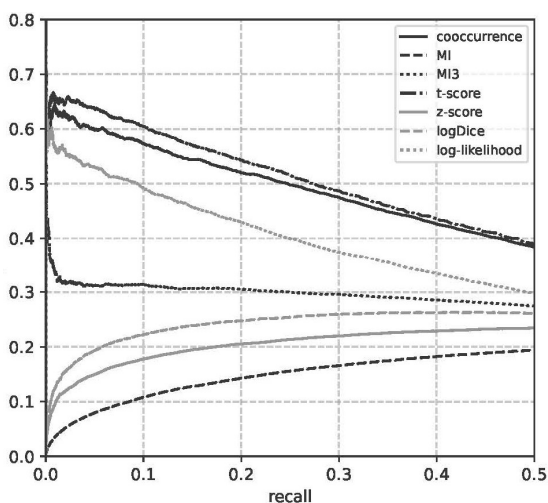


図 2-1 precision/recall (DP < 1)

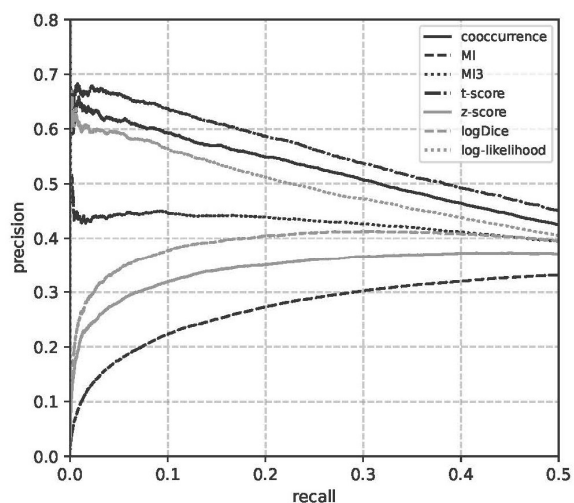


図 2-2 precision/recall (DP < 0.8)

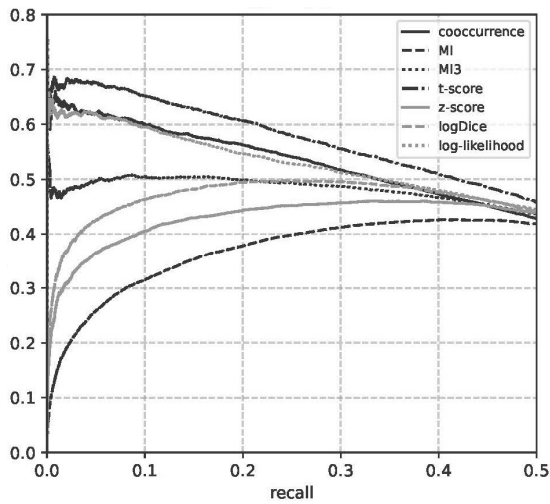


図 2-3 precision/recall (DP < 0.7)

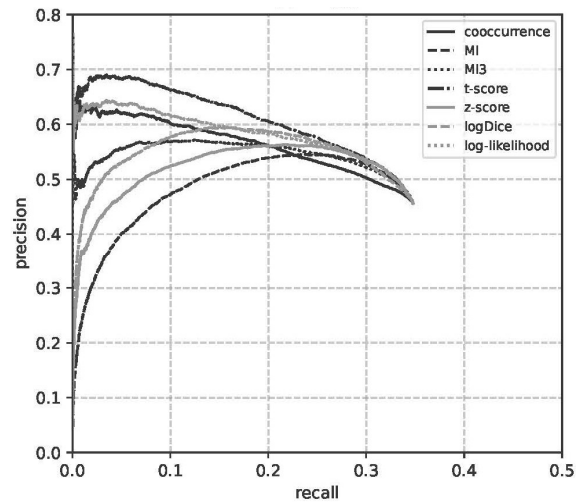


図 2-4 precision/recall (DP < 0.6)

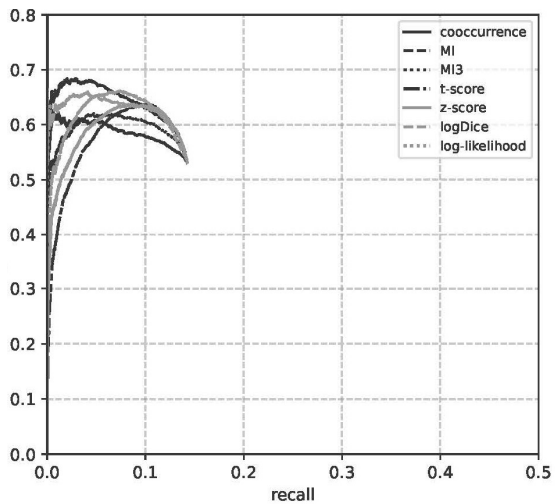


図 2-5 precision/recall (DP < 0.5)

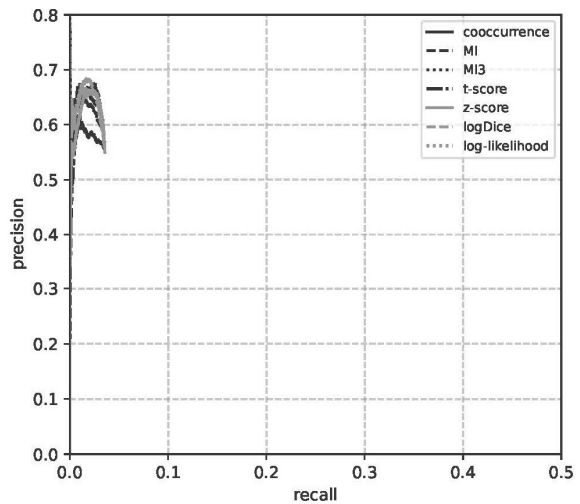


図 2-6 precision/recall (DP < 0.4)

図 2-1 から図 2-6 を見ると、常に t-score がよい成績であることが分かる。従って教育的に有用なコロケーションを選択するという点では t-score を用いるのが良いであろうということがグラフから推察される。DP < 0.6、DP < 0.5、DP < 0.4 のグラフは途中で切れているが、これは DP の値を低く設定すると、多くのコロケーションが除外され、一定以上 recall が上がらないことを意味する。例えば、DP を 0.4 以下に設定すると、2,487 の共起ペアしか残らないため、これ以上 n-best リストを長くすることはできない。この中に含まれるゴールドスタンダードの数は 1,366 であるため、DP < 0.4 の場合の recall は約 3.6% を超えることができない。従って、DP < 0.4 や DP < 0.5 は学ぶべきコロケーションのうちの多くを除外することになる。

グラフだけでは具体的な数値で AM・DP の成績を評価することができないため、precision と recall の観点から AM・DP の成績を数値化して評価したものが表 3 である。表 3 の AP30 は recall = 0.3 までの precision の平均値である。Evert, Uhrig et al. (2017) は、recall = 0.5 ま

での precision の平均値である AP50 によって各種のコロケーション抽出法の評価を行っているが、教育目的でのコロケーション抽出では数を絞る必要があることを勘案し、recall = 0.3 に設定した。DP の値を 0.5 未満に設定すると recall が 0.3 には達さないため AP30 は DP < 1.0、DP < 0.8、DP < 0.7、DP < 0.6 の 4 つの基準値別に算出した。

表 3 DP の基準値別 AP30

DP	AM	AP30
DP < 1.0	共起頻度(cooccurrence)	54.2%
	MI	9.7%
	MI3	30.9%
	t-score	56.5%
	z-score	17.3%
	logDice	21.6%
	log-likelihood	45.2%
DP < 0.8	共起頻度(cooccurrence)	56.7%
	MI	21.0%
	MI3	43.9%
	t-score	60.5%
	z-score	31.3%
	logDice	36.8%
	log-likelihood	53.3%
DP < 0.7	共起頻度(cooccurrence)	57.6%
	MI	30.4%
	MI3	49.7%
	t-score	62.1%
	z-score	39.9%
	logDice	45.3%
	log-likelihood	56.7%
DP < 0.6	共起頻度(cooccurrence)	57.2%
	MI	45.8%
	MI3	55.1%
	t-score	62.0%
	z-score	51.1%
	logDice	54.9%
	log-likelihood	59.3%

DP の基準値別に見てみると、0.7 に基準を設定するのが適切であると思われる。AP30 で最も高い値が出るのは DP < 0.7 の t-score の場合である。DP < 0.6 にすると、わずかであるが t-score の AP30 の値が下がる。他の AM は、DP の値を低く設定するほど成績が良くなっていることが分かるが、t-score の値を超えることはない。

以上のことを踏まえると基本的には DP < 0.7 辺りに設定するのが妥当であると考えられ

るが、実際には対象となる学習者のレベルに応じて、選択するコロケーションの数は絞ることになる。その場合 $DP < 0.6$ や $DP < 0.5$ に設定することも考えられるが、 t -score を使用する場合は、 DP を低く設定し必要なものを除外してしまうことを考えると、やはり $DP < 0.7$ あたりが妥当な目安になるであろう。

もちろん対象となる学習者やデータセットの使用の目的によって、 AM や DP やその他の情報の使用法は様々である。対象となる学習者レベルが高い場合は、 MI や \logDice の値を用いることも十分考えられる。これらの評価例を参考に、目的に応じて有用なコロケーションを $CEFR$ レベルを勘案して選定できるのがこのデータベースの新規性であるといえよう。

6. おわりに

本研究では、 $CEFR$ 準拠英語学習語彙表と多様な統計指標を活用して、学習すべきコロケーション特定のためのコロケーション・データセット作成の報告を行い、データセット内での DP と AM の評価を行った。データセットの使用及び DP ・ AM の教育的な観点での評価は暫定的なものであり、対象となる学習者レベルを考慮した調査を行い、構成素のレベルも踏まえたうえで、どのようなコロケーションが抽出されるのかより子細に調査する必要があるが、一応の目安として t -score を用い、 $DP < 0.7$ に設定すると、 $CEFR$ の $A1$ ~ $B2$ レベルの語彙で構成された学習レベル的に有益なコロケーションが選定可能になると考えられる。

付 録

本研究で作成したデータセットとプログラムコードは以下のリンクから利用可能である。全データについては、共起フレーム別に CSV 形式で保存してあるものと、 $Python$ プログラム上で呼び出し可能な $pandas DataFrame$ の形で保存してある。公開しているデータは本稿を適切に引用することで研究教育目的にのみ無償で利用可能である。

(<https://drive.google.com/drive/folders/1CxLqPpAL9UGTYd234siJgCOJNGv85Duw?usp=sharing>)

参考文献

- Kirsten Ackermann and Yu-Hua Chen (2013). “Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach”, *Journal of English for Academic Purposes*, 12:4, pp.235-247.
- Sabine Bartsch and Stefan Evert (2014). “Towards a Firthian notion of collocation”, In A. Abel and L. Lemnitzer (eds.), *Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern*, number 2/2014 in OPAL - Online publizierte Arbeiten zur Linguistik, pp.48-61. Institut für Deutsche Sprache, Mannheim.
(<https://www.stephanie-evert.de/PUB/BartschEvert2014.pdf> よりダウンロード可能)
- Vishal Bhalla and Klara Klimcikova (2019). “Evaluation of automatic collocation extraction methods for language learning”, In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pp.264-274, Florence, Italy. Association for Computational Linguistics. (<https://aclanthology.org/W19-4428/>よりダウンロード可能)
- BNC Consortium (2007). *British National Corpus, XML edition*, Oxford Text Archive, <http://hdl.handle.net/20.500.12024/2554>.
- Vaclav Brezina (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
- Stefan Evert (2004). “The Statistics of Word Cooccurrences: Word Pairs and Collocations”, Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart.
(<https://www.stephanie-evert.de/PUB/Evert2004phd.pdf> よりダウンロード可能)
- Stefan Evert (2008). “Corpora and collocations”, In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, article 58, pp.1212-1248. Mouton de Gruyter, Berlin.
(http://purl.org/stefan.evert/PUB/Evert2007HSK_extended_manuscript.pdf よりダウンロード可能)
- Stefan Evert, Peter Uhrig, Sabine Bartsch, and Thomas Proisl (2017). “E-VIEW-alation – a large-scale evaluation study of association measures for collocation identification”, In *Electronic lexicography in the 21st century. Proceedings of the eLex 2017 conference*, pp.531-549.
(<https://www.stephanie-evert.de/PUB/EvertUhrigEtc2017.pdf> よりダウンロード可能)
- Dana Gablasova, Vaclav Brezina, and Tony McEnery (2017). “Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence”, *Language Learning*, 67:S1, pp.155-79. (<https://onlinelibrary.wiley.com/doi/full/10.1111/lang.12225> よりダウンロード可能)
- Sylviane Granger and Magali Paquot (2008). “Disentangling the phraseological web”, In Sylvia Granger and Fanny Meunier (eds.) *Phraseology: An Interdisciplinary Perspectives*, pp.27-49, Benjamins.
- Stefan Th Gries (2008). “Dispersions and adjusted frequencies in corpora”, *International Journal of Corpus Linguistics*, 13:4, pp.403-437.
(https://stgries.info/research/2008_STG_Dispersion_IJCL.pdf よりダウンロード可能)
- 石川慎一郎 (2008). 『英語コーパスと言語教育 —データとしてのテキスト』大修館書店.
- Adam Kilgarriff, Vít Baisa, Jan Bušta et al. (2014). “The Sketch Engine: ten years on”, *Lexicography ASI/ALE*, 1, pp.7-36. (<https://link.springer.com/article/10.1007/s40607-014-0009-9> よりダウンロード可能)

- Ron Martinez and Norbert Schmitt (2012). A Phrasal Expressions List. *Applied Linguistics*, 33:3, pp.299-320.
- Ron Martinez and Victoria A. Murphy (2011). Effect of Frequency and Idiomaticity on Second Language Reading Comprehension. *TESOL Quarterly*, 45:2, pp.267-290.
- Michael McCarthy and Felicity O'Dell (2017) *English Collocations in Use Intermediate* [second edition] Cambridge University Press.
- Colin McIntosh, Ben Francis and Richard Poole (2009). *Oxford collocations dictionary: for students of English (Second)*, Oxford University Press.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman (2021). “Universal Dependencies”, *Computational Linguistics*, 47:2, pp.255-308.
(<https://direct.mit.edu/coli/article/47/2/255/98516/Universal-Dependencies> よりダウンロード可能)
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. (2020). “Stanza: A Python Natural Language Processing Toolkit for Many Human Languages”, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp.101-108, Association for Computational Linguistics. (<https://aclanthology.org/2020.acl-demos.14/>よりダウンロード可能)
- James Rogersa, Chris Brizzarda, Frank Daultonb et al. (2015). “On using corpus frequency, dispersion, and chronological data to help identify useful collocations”, *Vocabulary Learning and Instruction*, 4:2, pp.21-37.
- Pavel Rychlý (2008) A Lexicographer-Friendly Association Score, In Proceedings of Recent Advances in *Slavonic Natural Language Processing*, RASLAN, pp.6-9.
(<https://nlp.fi.muni.cz/raslan/2008/papers/13.pdf>よりダウンロード可能)
- Rita Simpson-Vlach and Nick C. Ellis (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31:4, pp.487-512.
- John Sinclair (1991). *Corpus, Concordance, Collocation*, Oxford University Press.
- 投野由紀夫 (編) (2013). 『CAN - DO リスト作成・活用 英語到達度指標 CEFR - J ガイドブック』大修館書店.
- 恒川元 (2020). 「logDice 係数はどのような共起指標か」『言語文化論究』45, pp.35-44.
(https://catalog.lib.kyushu-u.ac.jp/opac_detail_md/?lang=0&amode=MD100000&bibid=4104141よりダウンロード可能)

スロベニア人初級・中級・上級日本語学習者コーパスの構築と 縦断的文法誤用分析

パウロヴィチ・ミハ (リュブリャーナ大学文学部アジア研究学科)

The Construction and longitudinal analysis of a learner corpus of Japanese on the elementary, intermediate and advanced level

Miha Pavlovič (University of Ljubljana, Department of Asian Studies)

要旨

初級、中級、上級レベルのスロベニア人日本語学習者コーパスを構築した。その中の文法誤用の分析を行った。各レベルで最も頻繁な文法誤用を判別し、レベル間の結果を比較することにより、さまざまなタイプの誤用の発生傾向の変化を分析した。その結果、237 のテキストと、1778 のマークおよび分類された文法誤用を含むコーパスが作成された。習熟度のレベルごとに、誤用の数が最も多いカテゴリーを決定した。コーパスには縦断的なデータも含まれているため、そのデータも分析し、両方の分析の結果を比較した。最後に、誤用が最も多い文法をより詳細に分析した。分析の結果は、スロベニア人日本語学習者のデータを含む初めての公的コーパスが公開された。

1. はじめに

言語的誤用は言語習得過程の重要な一部である。本研究では日本語学習者の文章に見られる誤用を扱っている。外国語学習者の言語使用で発生する誤用に関するデータは、言語習得過程のより詳しい洞察を得るのに不可欠な情報源である。学習者にとって取得しにくい言語要素についての洞察を可能とし、このデータに基づいて、教師はカリキュラムを調整し、指導を改善することができ、学習者はより効率よく言語を習得できる。また、研究者、教科書や言語教材の作成に関わる者、そして学習者自身にも役立つと言える。

そのため、言語学習者コーパスというツールが情報源として非常に有利である。最も理想的な場合に、これらのコーパスは簡単にアクセスできるように公開されており、そしてデータの作成した学習者に関するさまざまな情報 (メタデータ) を備えており、さまざまな言語研究を可能にする。

本研究を行っている現時点では、日本語学習者コーパスは多数存在するが、複数のスロベニア語母語話者から取得したデータが含まれているものはまだない。従って、本研究の主な目標は、スロベニア語母語話者のデータを含む包括的な日本語学習者コーパスを構築することにより、以上述べた問題の解決策を提供することであった。もう一つの目標は、作成したコーパスの実用の試みとしてスロベニア人日本語学習者のデータを分析することであった。具体的には、個々の学習段階で見られる文法誤用を対象とする 3 つの分析のことである。

本論文では、構築されたコーパスを使用して行った 3 つの研究について紹介する。これは、三つの学習段階での最もよく見られる文法誤用の分析と、個々の段階間の比較である。分析の目的は、スロベニア人日本語学習者にとって、言語学習過程のさまざまな段階 (初級・中級・上級) にて最も困難な文法要素を特定し、それらの誤用をより詳細に分析することであった。

2. 研究背景

2. 1. データ収集とコーパス構築について

本研究の一環として、三つの学習段階でスロベニア語母語話者のデータを収集した。下記学年の学生より、それぞれデータを提供いただいた結果、合計で42人の学習者から237の作文を収集した。

- ・ 初級 → 137文 → 平均280字
2016/2017年に在籍していたリュブリャーナ大学日本研究コースの学部生の1年生
- ・ 中級 → 38文 → 平均500字
2017/2018年に在籍していたリュブリャーナ大学日本研究コースの学部生の2年生
- ・ 上級 → 62文 → 平均600字
2020/2021・2021/2022年に在籍していたリュブリャーナ大学日本研究コースの大学院生の2年生

集めたデータを上記のように学習能力の段階に応じて分類し、デジタル形式に変換した。データにて見られる誤用をアノテーションして、その分類をした。データは、簡単に使用および共有できるような形式で保存されている。詳細に説明すると最初はマイクロソフトのExcelを使用していたが、より共有しやすくするため、データをGoogle Sheetsに移動させた。また、できるだけ多くの研究を可能にすることを目的として、文章やその筆者である学習者に関する詳細な情報もアンケートで集めて、個々の文章に加えている。本当はちゃんとしたユーザーインターフェースやコンコダンサーがついているコーパスフレームワークを利用したかったが、一人でこのプロジェクトを行っていたため、バジェット（資金/使えるお金）が非常に限られていた。

2.2 仮説

調査を実施する前に、次の仮説が立てられた。

- ・ 最も頻繁な誤用は、最も頻繁に使用される文法要素と一致するだろう。
- ・ 学習者が日本語を長く学ぶほど、誤用は減少するだろう。
- ・ 初級と上級という2つの段階の文章とは異なり、中級段階の文章はテスト場面で作成されたため、誤用が最も多く見られるのは中級段階であろう。
- ・ 上級段階では、語の活用などの形態的誤用が減少し、学習者が学んだばかりの要素でより頻繁に誤用が発生するだろう。
- ・ 縦断調査の結果は、コーパス全体の調査で最も頻繁な誤用カテゴリーの結果と一致するが、学習者間の個人差による偏差がある。

3. 分析

3.1 第1分析

まずは、最も頻繁な文法誤用タイプを、初級、中級、上級のそれぞれの段階において分析した。各段階で誤用が最も多いカテゴリーを分析し、誤用を言語的特徴（品詞等）によって15のカテゴリーに分類した。

今回、上記の分類方法の参考として、市川（1995）で使われている分類化に近い海野（2012）の『オンライン日本語誤用辞典』に使われるものをもとにして、本コーパスのニーズに応じて幾らか修正したものを使用している。

また、Google Sheets を利用して各段階で最も頻繁な誤用の結果に基づいて、最も学生に困難な誤用のカテゴリーを特定した。段階間のより信頼性の高い比較ができるように、誤用の絶対数ではなく、文章の 1000 字にあたる相対頻度という値を使用した。各段階での誤用の頻度を比較することにより、初級と中級、および中級と上級の段階に共通する誤用のタイプが特定された。結果は次の通りであった。

表 1: 初級・中級・上級段階で最も誤用の多いカテゴリー

初級段階				中級段階				上級段階			
カテゴリー	誤用数	割合	頻度	カテゴリー	誤用数	割合	頻度	カテゴリー	誤用数	割合	頻度
格助詞	129	24.86%	3.21	格助詞	115	21.78%	5.47	格助詞	170	23.16%	6.66
基本文型	110	21.19%	2.73	2文の接続	107	20.27%	5.09	2文の接続	104	14.17%	4.08
修飾	82	15.80%	2.04	取り立て助詞	81	15.34%	3.85	取り立て助詞	101	13.76%	3.96
テンス・アスペクト	54	10.40%	1.34	テンス・アスペクト	39	7.39%	1.85	修飾	92	12.53%	3.61
2文の接続	50	9.63%	1.24	表現文型・接辞	39	7.39%	1.85	表現文型・接辞	57	7.77%	2.23
取り立て助詞	39	7.51%	0.97	修飾	37	7.01%	1.76	テンス・アスペクト	46	6.27%	1.80
形式名詞	14	2.70%	0.35	基本文型	29	5.49%	1.38	ヴォイス	44	5.99%	1.72
複合辞	12	2.31%	0.30	ヴォイス	24	4.55%	1.14	形式名詞	44	5.99%	1.72
ヴォイス	8	1.54%	0.20	形式名詞	23	4.36%	1.09	基本文型	30	4.09%	1.18
表現文型・接辞	8	1.54%	0.20	指示詞	14	2.65%	0.67	指示詞	28	3.81%	1.10
2語の接続	5	0.96%	0.12	複合辞	11	2.08%	0.52	複合辞	13	1.77%	0.51
指示詞	4	0.77%	0.10	2語の接続	9	1.70%	0.43	2語の接続	4	0.54%	0.16
終助詞	3	0.58%	0.07	終助詞	0	0.00%	0.00	疑問詞	1	0.14%	0.04
疑問詞	1	0.19%	0.02	疑問詞	0	0.00%	0.00	終助詞	0	0.00%	0.00
待遇表現	0	0.00%	0.00	待遇表現	0	0.00%	0.00	待遇表現	0	0.00%	0.00

3.2. 第2分析

個々の段階で文章を執筆した学習者が異なっているため、異なる学習者集団が結果に影響を与える可能性もある。より正確な結果を得るには、各段階で同時に収集された同じ学習者の文章が必要になる。同じ学習者から数年連続でデータを取得する必要があるため、このようなデータを取得することは非常に困難である。本研究で構築された SLJ コーパスでは、同じ学習者から 3 つの異なる時点に取得された縦断的データのサブセットも含まれているため、縦断的誤用分析が可能であった。このタイプのデータは 5 人の女子学習者から取得されており、縦断的サブコーパスを構成している。

このデータに基づいて、各段階での個々の学習者の文章の最も頻繁な誤用と、段階間の誤用の変化の傾向を分析した。次に、これらの傾向について、個々の学習者間と以上に述べたコーパス全体の分析の結果を比較した。比較を行った誤用カテゴリーは、最初の調査で最も頻繁で、SLJ コーパスのすべての誤用の 70%以上を占める 5 つの誤用カテゴリーである。そのカテゴリーは：

- ・ 取り立て助詞（「は」、「も」、「しか」等）
- ・ 格助詞（「を」、「に」、「で」、「が」等）
- ・ 基本文型（肯定文や否定文の構造、コンピュータの使用等）
- ・ 修飾（連体修飾、連用修飾、連体助詞「の」等）
- ・ 2文の接続（条件節、理由節など、文の接続に関する文法）

その他のカテゴリーは、誤用の数が少ないため比較を行わなかった。以下は格助詞に関する誤用の縦断的データをまとめた表である。左側の数値は 1000 字にあたる相対頻度であり、グラフの折れ線の色はそれぞれの学習者のデータを表している。

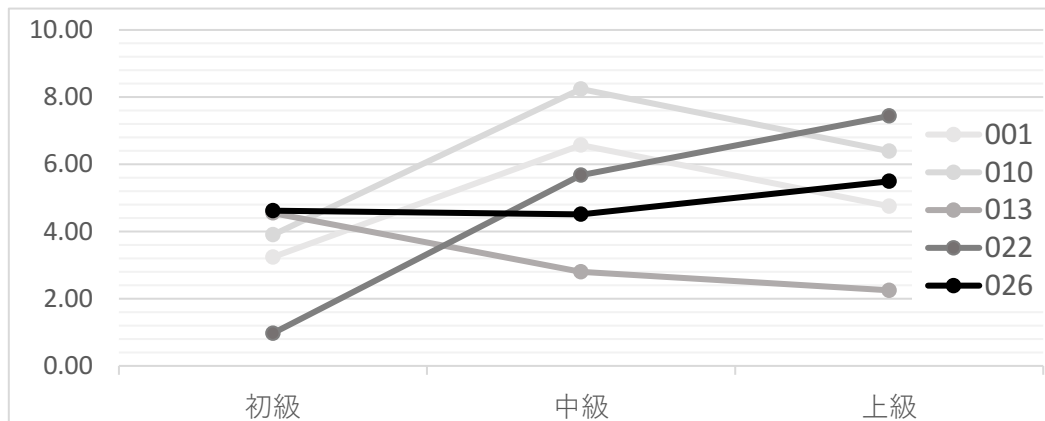


図 1：格助詞の誤用頻度の縦断的分析

3.3. 第 3 分析

第 1 調査では、助詞に関する誤用が最も頻繁であることが判明し、第 2 調査では、5 人の個々の学習者の誤用間で最大の差が見られるのも助詞の誤用であった。したがって、コーパスの誤用の最大の割合を占める 6 つの助詞をより詳細に分析することにした。その 6 つの助詞は「は」、「が」、「を」、「で」、「に」と「の」であった。これらの 6 つの助詞について、段階間の遷移での誤用の発生頻度の変化を比較し、誤用の最も頻繁な原因と、これらの助詞が最も頻繁に混同される文法要素を特定した。

誤用分析に対する批判の一つは、研究者が誤用の絶対数だけに基づいて結論を出すことが多く、ある要素が他の要素よりも頻繁に使用されて誤用が発生しやすいことを考慮に入れていないことである。これは、本研究の SLJ コーパスでも観察できた。初級段階では、接続詞の誤用は頻繁ではないが、中級段階では、学習者が条件節、原因節等の接続詞を知っていてこれらの誤用の割合ははるかに高くなる。ある要素が実際に学習者にとって困難であるか、または単にその要素の使用数が多いため誤用も多いかを確認するには、誤用の数だけでなく、誤用率も決定する必要がある。それは具体的に、標準的な文章において予想される使用量と、実際の使用量の比率のことである。つまり使用されるべきであった回数（教師によって採点された文章の回数）と実際に学習者が使用した回数の比率である。以上述べた六つの助詞のそれぞれの誤用率を上記のように計算し、これらの結果を比較した。

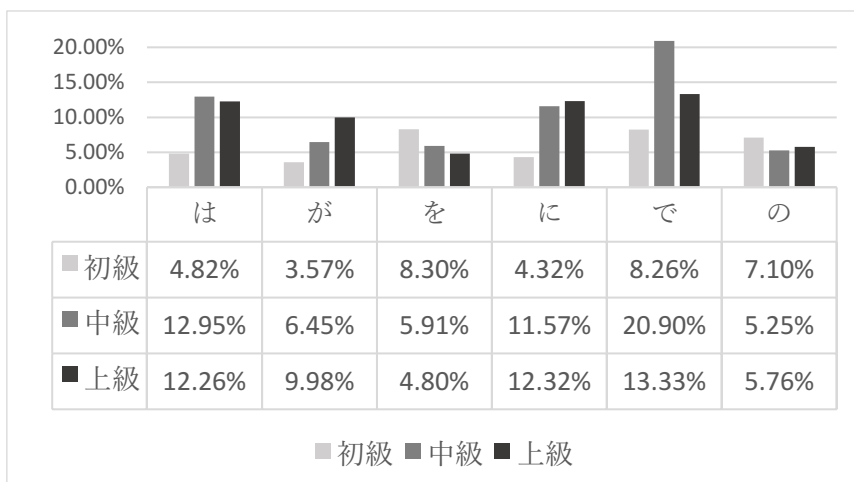


図 2：六つの最も誤用の多かった助詞の誤用率

4. 結果

研究結果からは次のことが分かった。

すべての段階で最も頻繁な誤用が助詞に関する誤用である。格助詞の「が」、「で」、「に」と「を」、取り立て助詞の「は」と連体助詞「の」の誤用が多いことが判明した。

初期段階では、基本文型、特にコピュラ「です」の使用、動詞と形容詞の活用などに関連する誤用も多く見られるが、学習時間が長いほどこれらの誤用は非常にまれになる。助詞の場合は逆である。すなわち、上記の六つのカテゴリーの誤用はある段階から次の段階に移行するにつれて誤用率が増えるのである。したがって、最も頻繁に使用される文法要素の誤用が最も頻繁であるという仮説が確認された。

それぞれの学習者で最も頻繁に発生する誤用の傾向に共通点が見られるという仮説も確認されたが、学習者間の個人差により、偏差もある。中級段階の作文状況が他の二つの段階とは異なるため、誤用の頻度が中級段階のデータで最も高く見られるという仮説は、部分的に確認された。その理由は、中級段階の誤用頻度は初期段階よりも高いものの、上級段階では、カテゴリーの半分以上の誤用頻度が中級段階よりもさらに高いためである。

より詳細に分析された六つの助詞のうち、四つの場合に初期段階での誤用割合は、上級段階での誤用割合よりも低くなっている。したがって、言語学習時間が長くなると誤用の頻度が減少するという仮説は確認されなかった。

中級以上の段階で最も頻繁な誤用は、学習者が学んだばかりの文法要素であるという仮説は、部分的に確認された。中級段階では、初級段階の文章の執筆時に学習者がまだ知らなかった接続詞に関する誤用が中級段階では増加すると同時に、かなり早い時点から学ぶ格助詞の誤用がすべての段階で最も頻繁で、上級段階では、学んだばかりの文法要素の誤用割合に増加は見られないからである。もちろん、学習者がそれらの要素を避けているために誤用が発生しない可能性もあるが、その仮説を確認するにはさらなる調査が必要である。

5. おわりに

このように、本研究では、スロベニア語母語話者の言語使用データから成る日本語学習者コーパスを作成し、コーパスのデータを元にして三つの言語分析を行い、スロベニア人日本語学習者の文章における文法誤用の発生傾向を個々の段階において分析した。作成したコーパスはオンライン公開されているため、それに基づいて他の一連のコーパス分析を実行することも可能としている。学習者の文章内の語彙やスタイルの誤用の比較、または、他の学習者コーパスとの比較、他の母語話者のコーパスとの比較などが可能である。本コーパスは、今後も発展させ、できるだけ多くのユーザーに利用できるよう継続していく予定である。

文 献

Ellis, Rod (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Granger, Sylviane in Magali Paquot (2017). »Towards standardization of metadata for L2 corpora.«

Presentation at the CLARIN workshop on Interoperability of Second Language Resources and

Tools, 2018. <https://sweclarin.se/swe/workshop-interoperability-l2-resources-and-tools>

Meunier, Fanny (2016). »Introduction to the LONGDALE Project«. In *Studies in learner corpus linguistics. Research and applications for foreign language teaching and assessment*, eds. Erik Castello, Katherine Ackerley in Francesca Coccetta, pp. 123-126. Berlin: Peter Lang.

Volodina, Elena, Lena Granstedt, Arild Matsson, Beá a Megyesi, Ildikó P lán, ulia Prentice, Rosén

- Dan, et al. (2019) "The SweLL Language Learner Corpus: From Design to Annotation". *Northern European Journal of Language Technology (NEJLT)*. pp. 67-104.
- Pavlovič, Miha (2018) *Izgradnja in analiza korpusa usvajanja japonščine kot tujega jezika na osnovni in srednji ravni (bachelor's thesis)*, Ljubljana: University of Ljubljana.
- 市川保子 (1997) 『日本語誤用例文小辞典』 凡人社.
- 海野多恵・他 (2012) 「オンライン日本語誤用辞典」 東京外国語大学
<http://cblle.tufs.ac.jp/llc/ja/index.php?menulang=en>
- 川口良 (1995) 「中上級日本語学習者の作文にみる誤用の一例」 *言語文化と日本語教育*, pp. 178-188.
- 佐治圭三 (1992) 『外国人が間違えやすい日本語の表現お』 ひつじ書房.
- Serrano, Nestor・入野みはる (2018) 『「誤用」から日本語学習の「成果」をはかる試み――中級学習者による「旅」の作文を用いて』 *JLTANE 2018*, Dartmouth, 2018年8月18日.
- 寺村秀夫 (1990) 『外国人日本語学習者の日本語誤用例集』 寺村誤用例集データベース.
<http://teramuradb.ninjal.ac.jp/teramura.goyoureishu.pdf>.
- 原沢伊都夫. (2012) 「日本語初中級学習者の作文指導：学習者の誤用分析をもとに」 *静岡大学国際交流センター紀要*, 6, pp. 79-92.
- 丸山岳彦・田野村忠温 (2007) 「コーパス日本語言語学の射程」 *日本語科学*. 22, pp. 5-12. 国立国語研究所
- 望月通子 (2012) 「日本語教育における学習者コーパスの構築と ICLEAJ」 *関西大学外国語学部紀要* 7, pp. 111-119. 関西大学外国語学部.
- 野田尚史・迫田久美子 (2019) 『学習者コーパスと日本語教育研究』 くろしお出版.
- 吉田桃子 (2016) 「イタリア語話者の日本語作文の誤用分析から分かること」 *日本語教育シンポジウム報告・発表論文集*.



8月31日（水）

口頭発表セッション3, 4, ポスターセッション2

現代日本語の名づけにおける「文の包摂」

泉 大輔 (東京外国語大学) †

An Analysis of Sentential Compound in Japanese Naming

Daisuke Izumi (Tokyo University of Foreign Studies)

要旨

本稿で取り上げるのは、「振り込め詐欺」「早く帰れオーラ」「幻のポケモンをもらおう!キャンペーン」「いいねボタン」「かまってちゃん」「朝はパンだ派」など、合成語の前項に「文相当の要素」が生起する言語現象(以下、「文の包摂」)である。一般に日本語の語形成規則では、語(小さい言語単位)の中に文(大きい言語単位)は入り得ない(*明日行こう店)。しかし、「振り込め詐欺」という表現は、「〇〇詐欺」という合成語の中に「振り込め」という命令文相当の要素が含まれている点で逸脱的な表現と言える。本研究では主にコーパスを用いて「文の包摂」の実例を収集し、その使用実態を記述した。その上で、「文の包摂」は個人が臨時的に名づけやネーミングに用いられ、「新奇性」という表現効果とその使用の動機づけになっていると考察した。

1. 研究の目的と研究の背景

本研究で研究の対象としているのは、「振り込め詐欺」「幻のポケモンをもらおう!キャンペーン」「いいねボタン¹」「朝はパンだ派」「かまってちゃん²」「なんでも言うこと聞きます券」など、「文相当の要素」に名詞や接尾辞が直接後続している言語形式である³。本研究では、このような言語形式を、合成語のその内部に文相当の要素が包み込まれた言語現象と捉え、便宜的に「文を包摂する合成名詞⁴」(文包摂名詞)と呼ぶこととする⁵。

通常、動詞の命令形(「振り込め」)や意志形(「幻のポケモンをもらおう」)、テ形(「か

† daisuke.izumi7@gmail.com

¹ 「いいねボタン」とは、種々の SNS (ソーシャルネットワークサービス) の中で、誰かが投稿した文章や画像などに対して「好き」「おもしろい」「支持できる」といった共感などの気持ちを示すための機能(クリックするボタン)の総称である。SNS の 1 つである Facebook の「いいねボタン」はこぶしを握り、親指を立てて上に向けた形をしている。

² 「かまってちゃん」とは、「俗に、他人にほめられたい、親切にされたいなどの気持ちが強く、周囲の人の気を引くような言動を繰り返す人」のことである(『デジタル大辞泉』2021年6月28日閲覧)。

³ 本稿における下線はすべて筆者による。

⁴ 接尾辞「感」の用法について考察した曾(2017:152)では、「やっちゃまったなあ感」「バブル後に全ていなくなりました感」などを例に挙げ、文相当の要素と「感」が結びつく表現を「文を包摂する「感」と呼んでいる。本研究ではこれに倣い、便宜的に「文を包摂する合成名詞」という用語を用いるが、実際には文ではなく、談話相当の要素と結びつく例も見られる(例:「行った先でどうにかなるさ〜。買えばいいじゃん」主義)。「めんどくせー出かけたくないよー」病)。「熱?平熱じゃん」クラス(の超微熱)」「緊急事態宣言!不要不急の外出は控えてください」要請)など。そのため、当該の言語現象を指し示す用語については今後も検討する必要がある。

⁵ 類似する言語現象には、「句の包摂」(影山 1993)がある。これは「中世のフランス風」「懐かしの名器展示会」「24日に開催された美人コンテスト会場」(下線部が句の部分)など(影山 1993:326-327)、合成語の内部に「句」が包み込まれていると捉えられる現象のことである。語形成規則上、通常は合成語の構成要素にはなり得ないものが前項になっている点で、「句の包摂」と「文の包摂」は共通している。しかし、「文の包摂」は、「中世のフランス」や「懐かしの名器」などの名詞句ではなく、「やっちゃまったなあ」や「行くぞ!」などの文相当の要素が前項となる点で「句の包摂」とは異なる。

まって」）、名詞に後接する「だ」の終止形（「朝はパンだ」）、終助詞（「いいね」）、丁寧形（「なんでも言うこと聞きます」）などで終了する文は、合成語の前項になり得ない。前項語基が名詞や動詞の連用形、形容詞語幹である合成語（「経済問題」「上り坂」「嬉し涙」など）であれば通常の語形成の範囲で扱えるが、文包摂名詞は本来なら合成語の前項にはなれないはずの要素を持ち、旧来の語形成論ではとらえきれない現象である。この点で文包摂名詞は、従来の文法規則には当てはまらない特異な言語現象であると言える。

このように、文包摂名詞は逸脱的な表現であるにもかかわらず、近年、その使用が広く観察される。例えば、ブログなどのウェブ上のテキスト、SNS上のテキスト、広告表現やキャッチフレーズ、日常会話など、話し言葉や打ち言葉⁶で特にその実例が多数見られる。しかし、従来の日本語研究では、文包摂名詞は研究の対象として詳細に記述・考察されてこなかった。そのため、このような特異な言語現象について、その実態は明らかにされておらず、本質的な理解は得られていない。

そこで、本研究では、現代日本語における文包摂名詞の実例を収集してその出現状況を調査し、文包摂名詞の使用実態について記述する。特に、名づけの観点からその使用実態について考察することで、文包摂名詞が使用される動機付けの解明に迫る。

2. 先行研究の検討

文包摂名詞に関連する先行研究を見ると、次の2つの問題点があると考えられる。

問題点①：文包摂名詞の全体像が明らかになっていない

影山（1993）では、「[夏目漱石と正岡子規]展」「[カラオケとゲーム]大会」といった表現を取り上げ、このように語の内部に句が包み込まれる現象を「句の包摂」と呼んでいる。なお、通常、語の内部には語より大きい単位が入れないため、語が句を包摂する表現は語形成規則を逸脱する例外的な現象であるとされている。このような現象については、林（1982）や石井（2007）でも取り上げられており、前項を句とする臨時的な合成語として位置づけられている。しかし、これらの先行研究では語の内部に文相当の要素が包み込まれる表現については記述されていない。

個別の接尾辞を扱った先行研究の中では、「的」「系」「風」「式」「感」などの接尾辞に前接する語基が、語から句、さらに文にまで拡大し、派生語の内部に文が包み込まれる用法があると述べられている（山下 2005、金田 2014、曾 2017 など）。先行研究では文を前項とする派生語の用法について、それを形成できる種々の接尾辞が取り上げられているが、一方、文包摂名詞を形成する後項名詞については、名詞「状態」しか取り上げられていない。

名詞「状態」について考察した新屋（2014）では、「「何だコリヤ??」状態」「「こんなに難しかったっけ」状態」といった表現について、文を前項、「状態」を後項とする複合名詞と位置付け、その意味・表現機能などを記述している。しかし、実例を観察すると、「系」「式」「風」「感」といった接尾辞や、「状態」という名詞の他にも種々の名詞が文包摂名詞の後項になる場合が見られ、その全体像は明らかになっていないと言える。

問題点②：文包摂名詞を分析する上で適切なジャンルのデータを十分に確保できていない

⁶ 「打ち言葉」とは、SNSなどのインターネットを介したサービス上のテキストで用いられる言葉のことである（文化庁 2018）。

文包摂名詞は従来の語形成規則を逸脱する例外的な表現であるため、規範的な書き言葉においては出現することがなく、周辺的な問題として位置付けられてきた。しかし、当該の形式はブログ、SNS などの口語体のテキストを見ると、その実例の数は必ずしも少なくはない。

文包摂名詞について取り上げた先行研究（山下 2005、金田 2014、新屋 2014 など）では、口語体のテキスト（雑誌の記事、エッセー、『現代日本語書き言葉均衡コーパス』における「Yahoo!知恵袋」および「Yahoo!ブログ」、ウェブ上のテキスト）から得られた用例を提示しているものの、その用例は少数にとどまり、十分な量のデータに基づく検討は行われていない。

また、これらの先行研究の用例は主に 2010 年代以前のもものが中心で、2010 年代以降の比較的新しいデータを対象とした定量的な調査は行われていない。文包摂名詞のデータを収集するうえでは、出現しやすいウェブ上のテキスト（ブログなど）のデータが収録されたコーパスを使用することが必要である。

以上の 2 つの問題点をふまえて、本稿では、文包摂名詞について、その全体像を明らかにするために、まずはその出現状況を調査する。その際、文包摂名詞が出現しやすいウェブ上のテキストデータが収録された大規模コーパスを用いて用例の収集を行う。

3. 調査方法

3.1. コーパスの選定

文包摂名詞は、新聞や論文などの規範的な書き言葉ではその使用がほとんど見られない。しかし、近年では CM やポスターなどの広告、ネットニュースの見出し、雑誌などの本文、テレビ番組のタイトルなどで広くその実例が見られる。特に、2000 年以降のインターネットの普及に伴い、ブログや SNS などのウェブ上のテキストで文包摂名詞の実例が数多く観察される。

しかし、前節で指摘した通り、先行研究では書籍（雑誌やエッセーなど）から採取された少数の用例や、『現代日本語書き言葉均衡コーパス』の中の「Yahoo!ブログ」および「Yahoo!知恵袋」といった一部のウェブ上のテキストで見られた用例のみが取り上げられている。そのため、文包摂名詞の用例を大規模に収集し、どのような出現の分布を見せるのか、どのような諸特徴を有するのかについては明らかになっていない。そこで、本研究では、文包摂名詞の諸特徴を詳細に記述するために、その使用が広く観察され、より多くの実例が収集可能なウェブ上のテキストを調査資料として用いることとする。

ウェブ上のテキストから実例を収集するにあたって、本研究では『国語研日本語ウェブコーパス』を使用する。このコーパスは国立国語研究所がウェブ上の日本語テキストから収集したデータをもとに開発した日本語コーパスである。ウェブ上のデータの収集期間は 2014 年 10～12 月とされており、約 100 億語が収録されている。なお、『国語研日本語ウェブコーパス』では検索ツールに「梵天」と呼ばれるアプリケーションが用いられる。

このコーパスを使用する理由には次の 3 つがある。第一に、本コーパスはウェブ上のテキストを収録したコーパスとしては 2022 年現在で収録語数が最大（約 100 億語）であり、文包摂名詞の実例を採取できる可能性が高いと考えられるためである。第二に、先行研究で取り上げられている文包摂名詞の実例は、そのほとんどが最新のものでも 2008 年のものではあったが、本コーパスを用いることで 2014 年時点での実例が収集でき、2008 年よりも新しい

データが大量に観察できるためである。第三に、Google などの検索エンジンでは通常は文字列検索しかできないが、本コーパスの検索ツールである『梵天』を用いることによって、品詞を指定して検索ができ、収集したデータの加工や集計も可能になるためである。これにより、用例の収集および集計を効率的に行えるようになる。以上の理由から本研究では『国語研日本語ウェブコーパス』を用いて実例を収集する。

3.2 コーパスでの検索方法の検討

続いて、『国語研日本語ウェブコーパス』から用例を収集するにあたって、検索系『梵天』でどのような検索条件を指定するのかについて検討を行う。本コーパスの検索ツールである『梵天』では「品詞」や「活用形」などの条件を指定することで用例の検索が可能である。文包摂名詞は、「文相当の要素」の直後に名詞または接尾辞が後続する形式であるため、このような形式を抽出するには、文の直後に名詞または接尾辞が共起するという条件を指定できればよい。しかし、コーパスでは検索条件に「名詞」や「接尾辞」を指定して検索することはできても、「文」という単位を検索する機能はない。そのため、文を検索するための別の条件を設定しなければならない。なお、「文」の単位を検索するために、句点(「。」、感嘆符(「!」)、疑問符(「?」)を検索条件に指定するという方法も想定されるが、『国語研ウェブコーパス』ではその仕様上、これらの記号を検索条件に指定して検索することはできない。

そこで本研究では、文包摂名詞の多くの用例において、その前項の内部に種々の文末形式が現れることに着目する。そのような文末形式のうちのいくつかをコーパスの検索条件に指定することで、文包摂名詞の用例の抽出を試みる。例えば、コーパスの検索条件で、「命令形」の直後に「名詞」が後続するように指定すれば、命令形「辞めろ」に名詞「発言」が後続した「今すぐ辞めろ発言」といった用例を抽出することができる。

本研究では、次の4つの検索条件によってデータを収集する。すなわち、(a)「活用形」に「命令形」を指定するという条件(「早く帰れオーラ」など)、(b)「活用形」に「意志推量形」⁷を指定するという条件(「幻のポケモンをもらおうキャンペーン」など)、(c)「語彙素」に「だ」、「活用形」に「終止形」を指定するという条件(「犯人はお前だ宣言」など)、(d)「語彙素」に「な」を指定するという条件(「こんなもんだな程度」など)である。これらの形式が文末に現れる文は、後続する名詞を「という」などの形式を介在せずに直接修飾することはできない。そのため、通常の変体修飾構造ではなく、文包摂名詞だと判断できる。また、接尾辞についても、このような文末形式に直接後接することはできないため、文包摂名詞と判断可能である。

(a)「命令形」および(b)「意志推量形」を選定したのは、筆者が観察した限り、文包摂名詞の前項の述語が「命令形」や「意志形」である用例が多く見られたためである。(c)「だ」かつ「終止形」という条件を選定した理由は、この検索条件であれば、次のような種々の「文相当の要素」を前項とする文包摂名詞の実例が収集できるためである。例えば、名詞に「だ」の終止形が後接する名詞述語文(「これは夢だ」状態)など、な形容詞の終止形で終了する述語文(「どうせダメだ」発言)など、説明のモダリティを表す「のだ」で終了する文(「どうしたらいいんだ状態」)などがある。(d)「な」を選定する理由は、筆者が行

⁷ 『国語研ウェブコーパス』の検索条件として指定できる「活用形」の1つである。この「意志推量形」を検索条件に指定すると、意志形(「読もう」「食べよう」など)および推量形(「だろう」「でしょう」など)が抽出される。

った予備的な調査において、文包摂名詞の前項の文末に現れる終助詞のうち、終助詞「な」の用例が最も多く見られたからである⁸。

3.3. コーパスでの検索および用例の収集・集計の手順

用例の検索・収集および集計には、『国語研日本語ウェブコーパス』(検索系『梵天』)、『茶器』(Chaki.NET)⁹、『Excel』を用いる。以下、表1の流れにしたがってその手順を述べる。

表1 検索および用例の集計の手順

操作	使用ツール	手順
検索	『国語研ウェブコーパス』 検索系『梵天』	①「品詞列検索」を用いて検索条件の指定および検索を行う。
		②検索結果を「CaboCha形式」でダウンロードする。
変換	『茶器』	③「CaboCha形式」を『茶器』にインポートする。
		④「SQLite コーパス」を作成する。
		⑤「Tag 検索」を用いて検索条件の指定および検索を行う。
		⑥検索結果を「Excel」にエクスポートする。
集計	『Excel』	⑦「ピボットテーブル」を用いて用例の集計を行う。

手順①および手順②：『国語研ウェブコーパス』での用例の検索およびダウンロード

まず、表1中の手順①および手順②について説明する。検索をする際には『国語研ウェブコーパス』の「品詞列検索」を用いて、以下の表2における(a)～(d)のように検索条件を指定する。また、検索条件を指定した検索画面を以下の図1に示す。

表2 コーパスにおける検索条件

検索条件	ボックス0	ボックス1	ボックス2
(a)「命令形」	<活用形1> 命令形	<語彙素> 」	<品詞1> 名詞
(b)「意志推量形」	<活用形1> 意志推量形		<品詞2> 普通名詞
(c)「だ」かつ「終止形」	<活用形1> 終止形 <語彙素> だ		または
(d)「な」	<語彙素> な		<品詞1> 接尾辞 <品詞2> 名詞的

⁸ 『国語研ウェブコーパス』を用いて「終助詞」に「名詞」が直接後続する形式および「終助詞」と「名詞」の間に「鉤括弧閉」(J)を介する形式を検索した。検索の結果として得られた2万件のデータを集計したところ、文包摂名詞の前項の文末に生起する終助詞には、「な」(303件)、「よ」(291件)、「ね」(171件)、「か」(165件)、「ぞ」(57件)、「の」(39件)、「ぜ」(33件)、「わ」(29件)などがあった(泉2018)。

⁹ 『茶器』とは文字列、単語列および係り受け関係による検索機能を備えたコーパス管理ツールである。

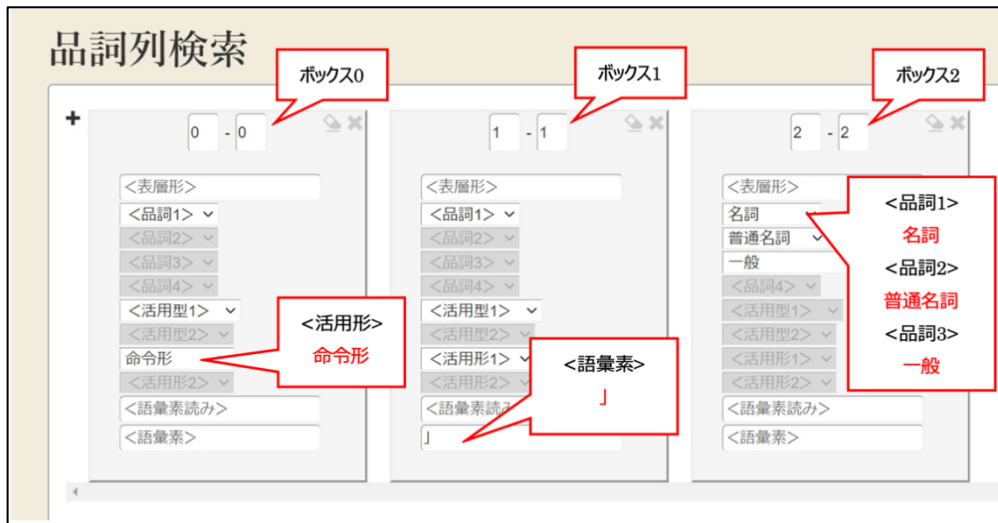


図1 コーパスの検索画面（(a)「命令形」の場合）

図1のように、まず、「ボックス0」（図1左側）では各文末形式の条件を指定する。例えば、(a)「命令形」の場合では、図1中の<活用形>に「命令形」を指定する。続いて、「ボックス1」（図1中央）および「ボックス2」（図1右側）では次のような検索条件を指定する。「ボックス1」では前項の文末形式と後項となる形式との間に「鉤括弧閉」（「」）が介在するように条件を指定する。具体的には、<語彙素>に鉤括弧閉「」を指定する。実例を観察すると鉤括弧に囲まれていないものも一定数見られる。しかし、その場合、書かれたテキストからだけでは前項の「文相当の要素」がどこから始まるのか、前項の範囲が判断しにくい。本章における調査では、前項の範囲を明確にするため、前項が鉤括弧で囲まれたもののみを検索対象とする。

「ボックス2」では、後項が名詞の場合は、<品詞1>に「名詞」を、<品詞2>に「普通名詞」を指定し、後項が接尾辞の場合は、<品詞1>に「接尾辞」、<品詞2>に「名詞的」を指定する。その後、検索結果は「CaboCha形式」という形式のファイルでダウンロードする。「名詞的」とは、名詞的な接尾辞のことで、派生名詞を作る「系」「派」「ちゃん」「さん」などがある。この「CaboCha形式」でダウンロードするのは、後述の『茶器』というツールで形態素解析をする際に必要となるためである。

「CaboCha形式」でダウンロードするうえで問題となるのが、「CaboCha形式」は『国語研ウェブコーパス』の仕様上、1万件を超える用例をダウンロードできないことである。そこで、検索結果をできるだけダウンロード可能数の上限である1万件以下になるようにする。その手順としては、「普通名詞」の下位分類5種¹⁰を順に検索条件に指定し（図1では<品詞3>を「一般」に指定）、5回に分けて検索を行う。このようにして得られたデータを「CaboCha形式」でダウンロードする。ただし、上記の下位分類に分けて検索を行っても1

¹⁰ 「普通名詞」の下位分類の5種とは、「サ変可能」（「勉強」など）、「サ変形状詞可能」（「心配」など）、「一般」（「山」「犬」など）、「形状詞可能」（「危険」など）、「助数詞可能」（「点」「組」など）である。

万件以上の用例がある場合は、無作為に選ばれた1万件の用例のみをダウンロードする。

得られたデータの件数は、合計 80,453 件である (表 3)。その内訳は、「命令形」の場合が 28,043 件 (後項が名詞のものが 24,817 件、後項が接尾辞のものが 3,226 件)、「意志推量形」の場合が 12,465 件 (後項が名詞のものが 11,135 件、後項が接尾辞のものが 1,330 件)、「だ」かつ「終止形」が 19,551 件 (後項が名詞のものが 15,641 件、後項が接尾辞のものが 3,910 件)、「な」が 20,394 件 (後項が名詞のものが 17,920 件、後項が接尾辞のものが 2,474 件) である。

表 3 データの件数の内訳

検索条件	データの件数		
	名詞	接尾辞	合計
(a) 「命令形」	24,817	3,226	28,043
(b) 意志推量形	11,135	1,330	12,465
(c) 「だ」かつ「終止形」	15,641	3,910	19,551
(d) 「な」	17,920	2,474	20,394
合計	69,513	10,940	80,453

手順③~⑥：『茶器』(Chaki.NET)におけるファイル形式の変換

次の工程では、コーパス管理ツール『茶器』(Chaki.NET)を用いる。ダウンロードした「CaboCha形式」のファイルはそのままでは検索や形態素解析、集計などを行うことができない。そのため、「CaboCha形式」のファイルを『茶器』に取り込み、「SQLiteコーパス」を作成する。この「SQLiteコーパス」というのは、「CaboCha形式」のファイルを検索が可能な形式に変換したものである。その後、「SQLiteコーパス」の中から「Tag検索」の機能を用いて、検索を行う。例えば、(a)「命令形」の場合は、一番左側の検索条件の「活用形」に「命令形」を、その後方の検索条件の「語彙素」に鉤括弧閉「」を、さらにその後方の検索条件に「名詞」「普通名詞」「一般」を指定して検索を行う(図2)。これによって得られた検索結果は、Excelファイル形式でエクスポートできるため、用例の集計が可能となる。

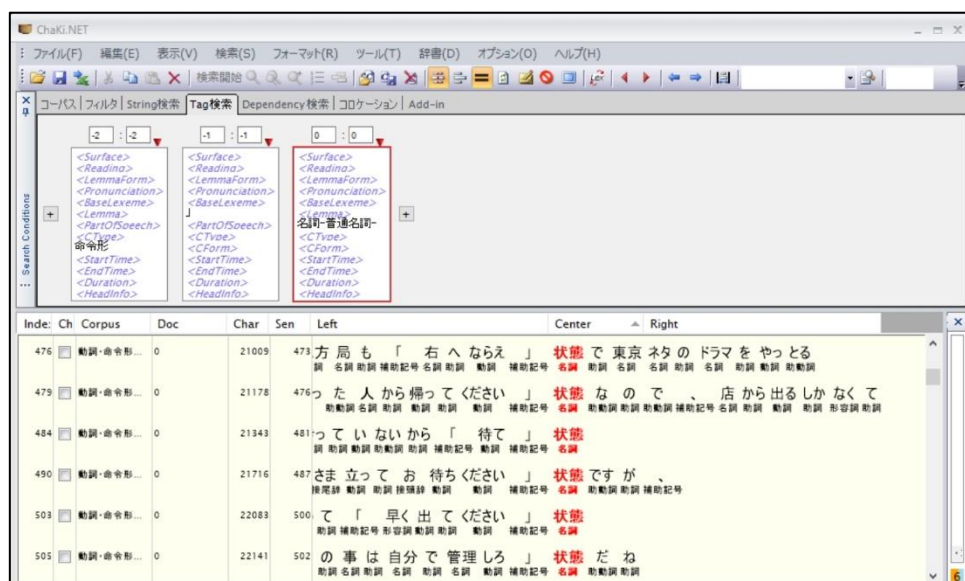


図2 『茶器』に取り込んだ「SQLite コーパス」における「Tag 検索」画面

手順⑦: Excel ファイルでの集計

Excel ファイルにエクスポートし、「ピボットテーブル」の機能を用いて集計を行う。集計作業では、用例数が 50 件以上見られるものを抽出する¹¹。重複する用例を Excel の機能を用いて削除したうえで、分析の対象外となるデータ¹²を目視で確認し、除外する作業を行う。最終的に得られた用例数を次節で示す。

4. 調査結果

前節で述べた手順で用例数を集計した結果、最終的には有効なデータとして 15,763 件の用例が得られた。収集期間および集計期間は 2019 年 4 月～2021 年 8 月である。得られた用例を見ると、文包摂名詞を形成する後項になり得るものには少なくとも 50 種類の名詞や接尾辞があることが明らかとなった (表 4)。

表 4 用例数の内訳

検索条件	後項の種類数	度数
(a) 「命令形」	38	5,932
(b) 意志推量形	28	4,745
(c) 「だ」かつ「終止形」	10	1,782
(d) 「な」	8	3,304
合計	50 (異なり語数)	15,763

用例数は「命令形」(5,932 件)、「意志推量形」(4,745 件)、「な」(3,304 件)、「だ」の「終止形」(1,782 件)の順に多い。用例数の多い「命令形」および「意志推量形」に関しては、その後項の種類も多く、それぞれ 38 種類、28 種類の形式が後項になり得る。「だ」かつ「終止形」と「な」は、後項がそれぞれ 10 種類、8 種類と比較的少ない。さらに、「な」に関しては、「だ」かつ「終止形」の約 2 倍の用例数が見られるにもかかわらず、後項となり得る形式の種類が少ない点が特徴的である。このように、前項の文末形式によって、後項に共起しやすい形式の種類数および用例数には差が見られる。文包摂名詞を形成する後項になり得る形式の一覧とその出現数は、次の表 5 の通りである。

表 5 文包摂名詞の後項となり得る形式とその出現数

順位	後項	前項の文末形式				合計
		「命令形」	「意志推量形」	「だ」かつ「終止形」	「な」	
1	程度	226	372	513	1,936	3,047
2	発言	717	257	550	371	1,895

¹¹ 用例数が 50 件未満の場合、コーパスの誤解析によって抽出された用例が多いため、用例数が 50 件以上見られたもののみを本研究では取り上げる。

¹² 分析の対象外となるデータは基本的にコーパスの誤解析によるものである。例えば、名詞「ひっかけ」(引っ掛け)が動詞「引っ掻く」の命令形として誤解析された例、名詞「おけ」(桶)が動詞「置く」の命令形として誤解析された例などがある。

3	状態	700	184	234	256	1,374
4	キャンペーン	114	773			887
5	コール	722				722
6	シリーズ	153	474			627
7	感		173	124	316	613
8	攻撃	361	119			480
9	メール	300	74			374
10	レベル	96	75	50	151	372
11	オーラ	160	50		146	356
12	企画		295			295
13	さん	131	160			291
14	コーナー	102	188			290
15	式	144	54	59		257
16	運動	94	160			254
17	作戦	89	161			250
18	精神	123	100			223
19	イベント	54	167			221
20	プロジェクト		216			216
21	アピール	111	57			168
21	派	109			59	168
23	論	94		66		160
24	モード	81	77			158
25	宣言	68		73		141
26	系			60	69	129
27	計画		124			124
28	タイプ	114				114
29	遊び	107				107
30	指令	105				105
31	ツアー		96			96
32	講座		91			91
33	命令	85				85
34	メッセージ	83				83
35	ポーズ	80				80
36	画面	78				78
37	方式	76				76
38	篇		67			67
39	編		66			66
40	風 (ふう)	63				63
41	ゲーム	60				60
41	表示	60				60
43	様		59			59
44	スレ	58				58
45	コメント	56				56
45	特集		56			56
47	展	54				54
48	エンド			53		53

48	連呼	53				53
50	デモ	51				51
	合計	5,932	4,745	1,782	3,304	15,763

上の表 5 から読み取れる点について述べる。まず、用例数の多さで上位 3 位に入る「程度」「発言」「状態」はいずれも用例数が 1,000 件を超え、どの文末形式も前項に生起する用例が見られる。ただし、4 種の文末形式と共起可能であるといっても、文末形式ごとに用例数には差が見られる。

例えば、「程度」は「な」と共起する用例数が圧倒的に多く、「発言」は「命令形」と「だ」かつ「終止形」と共起する用例が比較的多く、「状態」は「命令形」と共起する用例が多い。また、用例数の多さが上位 4 位および 5 位である「キャンペーン」と「コール」は、それぞれ「意志推量形」と共起する場合と、「命令形」と共起する場合にその出現が偏っている。さらに、出現する用例数が少ない形式になればなるほど、共起する文末形式には偏りが見られる。

このことから、用例数が多く観察され、文包摂名詞として用いられやすい形式の中にも、種々の文末形式と共起可能なものと、一部の文末形式との共起に偏るものがあると言える。このような偏りが見られるのには、その前項の「文相当の要素」が表すモダリティと、後項の意味的な特徴との親和性が影響していると考えられる。

さらに、上の表 6 中の後項となり得る形式以外にも、筆者が Google 上のテキストやテレビ番組などから採取した文包摂名詞の用例には、次のような名詞や接尾辞が後項となるものも見られる（182 種類）。したがって、上述のコーパスで得られた後項になり得る形式 50 種類と合わせると、少なくとも 232 種類の形式が文包摂名詞の後項になり得るということになる。

以下の表 6 はおよび表 7 は、文包摂名詞の後項別に出現数とその用例を示したものである。表 6 中の用例はウェブコーパスから得られた中核的な資料中のものであり、表 7 の用例は筆者が個別に採取した補助的な資料中のものである。

表 6 文包摂名詞の後項となり得る形式とその出現数および用例一覧（50 位まで）

順位	後項	合計	用例
1	程度	3,047	「頑張れ」程度、「注意して {くれ/ください}」程度、「気を付けろ」程度、「とりあえずがんばろう」程度、「これでいいだろう」程度、「気をつけよう」程度、「様子をみましょう」程度、「へえ、こんながあるんだ」程度、「ああ、そういう人がいるんだ」程度、「あったらいいな」程度、「ないよりマシかな」程度、「あーそういうことあったなあ」程度
2	発言	1,895	「戦ってください」発言、「かかってこい」発言、「早く座れ」発言、「辞めろ」発言、「空気読め」発言、「黙れ」発言、「ナチスに学べ」発言、「飽きた、もう帰ろう」発言、「結婚しよう」発言、「どうしよう」発言、「最後は金目でしょう」発言、「おまえのためだ」発言、「謝罪すべきだ」発言、「女は子を産む機械だ」発言、「内政干渉するな」発言、「勘違いするな」発言、「金ないのに結婚するな」発言
3	状態	1,374	「ご自由に {お持ち帰り/お取り/お使い} ください」状態、「しばらくお待ちください」状態、「勘弁してくれ」状態、「やめてくれ」状態、「行って来い」状態、「果報は寝て待て」状態、「どうにでもなれ」状態、「どうしよう」状態、「こんなの無理だろ」状態、「ま〜何とかなるでしょう」状態、「早くおうちに帰ろう」状態、「もういやだ」状態、「俺はダメだ」状態、「お前は何を言っているんだ」状態、「お前が言うな」状態、「なんだかなあ」状態、「これでいいのかな」状態

4	キャンペーン	887	「{子どもたち/地球}を救え」キャンペーン、「(キャラクター名)を探せ」キャンペーン、「ミッフィーシールをあつめよう」キャンペーン、「そうだ京都へいこう」キャンペーン、「東北を応援しよう」キャンペーン
5	コール	722	「がんばれ」コール、「帰ってこい」コール、「帰れ」コール、「出て行け」コール、「辞めろ」コール、「シュート打て」コール、「金返せ」コール
6	シリーズ	627	「ウォーリーをさがせ」シリーズ、「よいこになあれ」シリーズ、「宇宙へ飛び出せ」シリーズ、「とにかく文章を書こう」シリーズ、「夏までに腹筋を割ろう」シリーズ
7	感	613	「どうしよう」感、「こんなもんだろ」感、「頑張ろう」感、「オレ様は客だ」感、「こりゃダメだ」感、「これどうすればいいんだ」感、「もういいかな」感、「やっちゃったな」感、「なんだかなー」感
8	攻撃	480	「{ごはん/おやつ/エサ}くれ」攻撃、「買ってくれ」攻撃、「{遊び/旅行}に連れて行け」攻撃、「食べる」攻撃、「遊ぼう遊ぼう」攻撃、「公園行こう」攻撃、「早く帰ろう」攻撃
9	メール	374	「{教えて/助けて/譲って}ください」メール、「(早く)帰って来い」メール、「ごめんなさい」メール、「頑張れ」メール、「デートしよう」メール、「ごはん行きましょう」メール、「一緒に帰ろう」メール
10	レベル	372	「注意してください」レベル、「しばらくお待ちください」レベル、「ご遠慮ください」レベル、「ま、こんなもんだろう」レベル、「これからがんばりましょう」レベル、「このまま様子見ましょう」レベル、「そんなことあったんだー」レベル、「こりゃ、どー見ても本物だ」レベル、「あったらいいな」レベル、「できたらいいな」レベル、「聞いたことがあるかなー」レベル
11	オーラ	356	「話しかけないでください」オーラ、「早く帰れ」オーラ、「俺ってかっこいいでしょ」オーラ、「早く帰ろう」オーラ、「近寄るな」オーラ、「声をかけるな」オーラ、「入ってくんな」オーラ
12	企画	295	「レビューを書いてプレゼントをもらおう」企画、「サプライズで誕生日を祝おう」企画、「お花を贈ろう」企画
13	さん	291	こまったさん、わかったさん、有りがたうさん
14	コーナー	290	「{もらって/教えて/譲って}ください」コーナー、「ご自由に{お持ち帰り/お取り}ください」コーナー、「{アトム/仲間はずれ}を探せ」コーナー、「選手に聞いてみよう」コーナー、「みんなで短歌・俳句を詠もう」コーナー、「世界の国のあそびを楽しもう」コーナー
15	式	257	「ご自由にお取りください」式、「習うより慣れよ」式、「俺についてこい」式、「紙に目標を書いてがんばろう」式、「作者の意図を読みとりましょう」式、「何でも自分でやろう」式、「これが真相だ」式、「何をしても心配だ」式、「考えるな、感じるんだ」式
16	運動	254	「{ウォール街/東京}を占拠せよ」運動、「出て行け」運動、「東京にオリンピックを招致しよう」運動、「元気に挨拶しましょう」運動、「毎日豆乳でも飲もう」運動
17	作戦	250	「押してダメなら引いてみろ」作戦、「急がば回れ」作戦、「買い物は歩いていこう」作戦、「飛行機乗るまで寝ないでいこう」作戦、「数を撃てばそのうち当たるだろ」作戦
18	精神	223	「やってみなはれ」精神、「急がば回れ」精神、「習うより慣れる」精神、「当たって砕けろ」精神、「郷に入っては郷に従え」精神、「無いなら自分達で作ろう」精神、「できることからはじめよう」精神、「とりあえずやってみよう」精神
19	イベント	221	「スティッチを探せ」イベント、「{娘さん/お嬢さん}をください」イベント、「東北の酒をのんで応援しよう」イベント、「政治家と話そう」イベント、「本を読んで寄付をしよう」イベント
20	プロジェクト	216	「手作りトートバッグで被災地を応援しよう」プロジェクト、「カンボジアに学校を建てよう」プロジェクト
21	アピール	168	「{エサ/おやつ/ごはん/めし}くれ」アピール、「{遊んで/買って}くれ」アピール、「俺カッコイイだろ」アピール、「早く帰ろう」アピール、「遊ぼう」アピール、「私って面白いでしょ」アピール
21	派	168	「習うより慣れる」派、「あたってください」派、「オレについてこい」派、「好きにきなさい」派、「ラーメンに並びたくないよなあ」派、「今日出来ることは明日に延ばすな」派
23	論	160	「ドイツを見習え」論、「専門家に任せておけ」論、「文句があるなら自分で作れ」論、「果報は寝て待て」論、「本当に怖いのは人間だ」論、「このままでは日本はダメだ」論、「迷惑かけてないからいいんだ」論

24	モード	158	「おやすみなさい」モード、「勘弁してください」モード、「どうしよう」モード、「帰りましょう」モード、「みんなに任せよう」モード、「遊ぼう」モード
25	宣言	141	「娘さんをください」宣言、「結婚してください」宣言、「付き合ってください」宣言、「ごめんなさい」宣言、「絶交だ」宣言、「俺がルールだ」宣言、「お前には無理だ」宣言
26	系	129	「いい話だなー」系、「お前が言うな」系
27	計画	124	「お家を建てよう」計画、「秋あたりに温泉に行こう」計画、「今から東京行こう」計画
28	タイプ	114	「俺に {ついてこい/任せろ} 」タイプ、「習うより慣れろ」タイプ、「果報は寝て待て」タイプ
29	遊び	107	「 {取って/持って} 来い」遊び
30	指令	105	「 {牛乳/餃子} 買って来い」指令、「 (早く) 帰って来い」指令
31	ツアー	96	「そうだ、京都へ行こう」ツアー、「アンコールワットを見つけに行こう」ツアー、「そらジローに会いにいこう」ツアー
32	講座	91	「コンピュータで音楽をつくろう」講座、「親子でリースを作ろう」講座、「世界遺産を学ぼう」講座
33	命令	85	「帰れ」命令、「来い」命令、「降りろ」命令
34	メッセージ	83	「頑張ってください」メッセージ、「しばらくお待ちください」メッセージ
35	ポーズ	80	「ごめんなさい」ポーズ、「くれくれ」ポーズ、聞こえませんポーズ
36	画面	78	「ログインしてください」画面、「しばらくお待ちください」画面、「パスワードを入力してください」画面
37	方式	76	「ご自由にお持ちください」方式、「急がば回れ」方式、「習うより慣れろ」方式、「損して得とれ」方式
38	篇	67	「手をつなごう」篇、「大人の旅を楽しもう」篇、「ハッピーになろう」篇
39	編	66	「だまされないよう頑張ろう」編、「かわいく着よう」編、「オープンキャンパスで会いましょう」編
40	風 (ふう)	63	「自由にお持ちください」風、「何でもいから書いてくれ」風、「同情するなら金をくれ」風
41	ゲーム	60	どこ行くんですかゲーム、「箱の中身はなんだろう」ゲーム、「スクリーンショットだけでゲームタイトルを当てられるかな」ゲーム、「待て」ゲーム、「 {ミッキー/スパイ} を探せ」ゲーム
41	表示	60	「しばらくお待ちください」表示、「充電してください」表示
43	様	59	「小説家になろう」様、「新時代ではこうしよう」様
44	スレ	58	「この本を読め」スレ、「作曲できる奴ちょっと来い」スレ
45	コメント	56	「おかえりなさい」コメント、「食べてみてください」コメント
45	特集	56	「できるかな」特集、「海女になろうよ」特集、「設計ミスをなくそう」特集、「下町散歩に出かけよう」特集、「ニッポンを食べよう」特集
47	展	54	書いたよ作ったよ展、あっ、これいいね展
48	エンド	53	「俺達の戦いはこれからだ」エンド
48	連呼	53	「ごめんなさい」連呼、「出て行け」連呼、「帰れ」連呼

50	デモ	51	「ウォール街を占拠せよ」デモ、「尖閣諸島を守れ」デモ
	合計	15,763	

表7 文包撰名詞の後項となり得る形式とその出現数および用例一覧 (51位以下)

順位	後項	合計	用例
51	祭り	7	起きよ祭り、「寿々ちゃん明日が1歳だね」祭り、「母の作った稲荷寿司はなぜあんなにうまいのか」祭り、「イチローすごいなあ」祭り、「すごいじゃん」祭り、「なんでやねん」祭り、「春ですね俳句選手権ですよ」祭り
52	合戦	6	「誰がやるか」合戦、「俺は知ってるぜ」合戦、「何でもいいからいかに試合を早く進めるか」合戦、オヤツよこせよこせ合戦、待て合戦、ここ掘れ合戦
53	おぼけ	5	もったいないおぼけ、つられたらたべちゃうぞおぼけ、くれくれおぼけ、やだやだおぼけ、どっか連れてけオバケ
53	顔	5	「やったぜ」顔、「知ってますよ」顔、「もういいや」顔、「ボク知らないよ」顔、「わかってるよ」顔
53	詐欺	5	今年こそ痩せるよ詐欺、「やばいよやばいよ」詐欺、「すぐ買わないと売れちゃうよ」詐欺、「アイドルになりませんか」詐欺、出すよ出すよ詐欺
53	欲求	5	「こうするとどうなるか」欲求、「投げるとどうなるのか」欲求、「どうしても投げてみたい」欲求、「ピザ食いたーい」欲求、「勉強やり直したいぞ」欲求
57	仮説	4	地理要因が成長に「直接」効くよ仮説、政治制度が重要だよ仮説、人が重要だよ仮説、夏のほうが好きな人のほうが多いんじゃないかな仮説
57	級	4	「あんた死ぬよ」級 (の脅し文句)、「おまえさんはサイボーグか」級 (の美しさ)、「異国に来たなあ」級 (の新鮮な驚き)、「こいつ、勝つためなら人間捨てるのか」級 (の超衝撃シーン)
57	教	4	「晴れろ〜！」教、頑張れ教、何でもござれ教、「ま、いっか」教
57	クラス	4	「走りたいぞ」クラス、「熱?平熱じゃん」クラス (の超微熱)、「(ツッコミが)「なんでやねん」クラス (の、何のヒネリもないもの)、「よくこんな守備でレギュラー取れたな」クラス (の守備)
57	コース	4	「黒目川の下流を目指してどこまでも」コース、「あの時の黒目川のお花見提灯を目指すぜ」コース、編集したら5分ないよ、基礎からみっちり教えるわよコース
57	思考	4	「まあいっか」思考、「俺の酒が飲めないのか」思考、「なんくるないさー」思考、「今だけ楽しければいいじゃん」思考
57	時代	4	「産めよ・殖やせよ」時代、「そうですね」時代、「ガンガンいこうぜ」時代、「あったかななかったかな」時代
57	主義	4	事なかれ主義、「行った先でどうにかなるさ〜。買えばいいじゃん」主義、「どうにかなるさ」主義、「どーでもいいよ」主義
57	賞	4	頑張ったね賞、あいつ、やるな賞、来年があるさ賞
57	星人	4	やだやだ星人、何だっけ星人、構ってよ星人、待ってればいいよ星人
57	調	4	効くんですよ調、線路は続くよ調、待っていたわ調、やめて、そんなこと言わないで調
57	プラン	4	「昼食を兼ねて復活した高橋果実店のものを食べよう」プラン、「甲子園へ行こう」プラン、「ハワイ旅行 or ダイソンのファンヒーターで2013年の冬をあたたく過ごそう」プラン、「ワンコと同室お泊り・草原を駆け回ろう」プラン
57	ブーム	4	人は褒めて伸ばすよブーム、関西魂とにかつつこめブーム、真水を飲むと病気をしなくなるよブーム、出せ出せブーム
57	要求	4	「なんかくれ」要求、「空気を読め」要求、「トイレに連れて行け」要求、「いつまでもなでていなさい」要求

57	理論	4	「社会じゃあ通用しないぞ」理論、「鶏が先か卵が先か」理論、「かわいい子には旅をさせよ」理論、「押すなよ絶対押すなよ」理論
72	映画	3	「恋人が死んで悲しいな」映画、「僕は夢を信じるよ」映画、「岩に挟まっちゃったよ〜ん」映画
72	おじさん	3	早く帰れおじさん、結果出せおじさん、なんちゃっておじさん
72	型	3	「俺は嫌だ」型、「助けてくれ」型、「インフルエンザかもね」型
72	願望	3	「つくりたいーい」願望、「私は他人と違うのよ」願望、「欧米人になりたいーい」願望
72	機能	3	「いいね」機能、「あったらいいな」機能、「なんだかなあ」機能
72	疑惑	3	この家傾いてるぞ疑惑、水樹って薫子さんが好きなんじゃね疑惑、課題終わらないよ疑惑
72	グループ	3	「1200mまでなら、僕たちの世界だよ」グループ、「天使を倒そうぜ」グループ
72	警告	3	「インクが無いよ」警告、「死ぬぞ」警告、「すぐなくなっちゃいますよ」警告
72	劇場	3	NHK たぶんこうだったんじゃないか劇場、NHK きっとこんな感じなんじゃないか劇場、NHK つまりこういうことなんだ劇場
72	月間	3	「とりあえず毎日ブログつけるぜ」月間、「腹を引っ込ませるぞ」月間、「相続登記はお済ですか」月間
72	券	3	なんでも言うこと聞きます券、誕生日だから半額にしますよ券、スタバーのコーヒーおごるよ券
72	現象	3	「以前お会いしましたよね〜」現象、「ジブリだからとりあえず見とくか」現象、「コレ誰だっけ」現象
72	広告	3	「今日のお昼、牛丼食べたいよね」広告、「あなたのPCにはウイルスがいますよ」広告
72	婚	3	できちゃった婚、こんなもんかな婚、リアルね婚
72	週間	3	「お前ら自分で勉強しとけよー」週間、「社会人、楽しいことばかりじゃないね」週間
72	症候群	3	「死んでもいいや」症候群、「ま、いっか」症候群
72	情報	3	「こっちの方がすいているよ」情報、「レースがありますよ」情報、「レースが終わりました」情報
72	証明書	3	「投票しましたよ」証明書、「お仕事休んで下さいね」証明書、「ちゃんと受け取りましたよ」証明書
72	スイッチ	3	頑張ろうスイッチ、遊ぼうスイッチ、「行くぞ」スイッチ
72	スタンス	3	「負けないからな」スタンス、「まあ聞けるならいいや」スタンス、「自分が子供を守るの」スタンス
72	セット	3	「甘いもの、これでもか」セット、「何にしようかな」セット、「越後屋、お主もなかなか悪よの〜」セット
72	ソング	3	頑張れソング、「いっといで」ソング、「原発なくてもええじゃないか」ソング
72	チーム	3	「そこであきらめたら試合終了よ」チーム、「ガンガンいこうぜ」チーム、「いのちだいじに」チーム

72	伝説	3	「両親ともに栗毛でないのに、栗毛の仔は走るぞ」伝説、「ラブは3歳になったら落ち着くよ」伝説、「シゲってそういうところあるよね」伝説
72	度	3	「明日会社行きたくねえ」度、「いい湯だな」度、「まあ、いっか」度
72	ドリンク	3	はじめの地球を守るッス！ドリンク、パイマンの宇宙人なめんなよドリンク、O・Dのやりすぎー！ドリンク
72	パターン	3	「お前が言うな」パターン、何となく見てたら、結局最後まで見ちゃったよパターン、勘弁してよパターン
72	バージョン	3	「遅れてごめんね」バージョン、「泳ぐな」バージョン、「もしよかったら、電話かメールくださいね」バージョン
72	病	3	イタイイタイ病、「めんどくせー出かけたくないよー」病、アップルパイが食べたいよ病
72	風味	3	「おれに聞くな」風味、「ええ仕事してますねえ」風味、「おうおう〜この桜吹雪が目に入られえかー」風味
72	部分	3	「この屏風何も描いてないじゃん」部分、「好きだよ」部分、「おーなかが減ったよー」部分
72	ボタン	3	いいねボタン、読んだよ！ボタン、「予約開始時点でお知らせメール送るよ」ボタン
72	マン	3	どこでも開けちゃうマン、絶対〇〇マン、お金貸してマン
72	ムード	3	「北京五輪楽しみですね」ムード、「俺（私）たちはラブラブだぜ」ムード、「どうせ選ばれるわけないよ」ムード
72	ランキング	3	「こんなのがあったらイヤだな」ランキング、「何が一番美味しかったか」ランキング、「うまいねん」ランキング
72	リスト	3	「あんなこといいな、できたらいいな」リスト、「死ぬ前に絶対行くもんね」リスト、「観ようかな」リスト
72	論争	3	「ニワトリが先か、卵が先か」論争、「一生借家でよいのか持ち家か」論争、「癌と闘うな」論争
109	アドバイス	2	「こうしたらいいよ」アドバイス、「こうすれば直るよ」アドバイス
109	アラート	2	もうすぐ地震が来るよアラート、「少ないよ」アラート
109	アラーム	2	帰れアラーム、「鼻水たれてますよ」アラーム
109	祝い	2	産休入るよ祝い、引越し終わったよ祝い
109	女	2	保育園作ります女、土地買ってくれませんか女
109	化	2	「孫見に行かねーとな」化、「俺が法律だ」化
109	会	2	「産休頑張ってるね」会、「打ち上げと称してしゃぶしゃぶ食おうぜ」会
109	係	2	「元気出していこうぜ」係、「探したんだけど見つからないよ」係
109	活動	2	「みんなで募金をよびかけよう！」活動、「見てね」活動
109	期	2	構って期、ほっといて期
109	競争	2	「自分のほうが苦勞しているんだぞ」競争、「誰が給食を早く食べ終わるか」競争
109	禁止令	2	「フィギュア買うな」禁止令、「でもさあ」禁止令
109	グッズ	2	疲れてちゃもったいないよグッズ、あるといいなグッズ

109	組	2	「独りでそんなに困らないよ」組、「ここでいいや」組
109	軍団	2	みんな額が広いぞ軍団、寝ないぞ軍団
109	芸人	2	「嫌なら見るな」芸人、相方どうかしてるぜ芸人
109	研究所	2	あったらいいな研究所、アクティブラーニングこんなのどうだろう研究所
109	公演	2	歌舞伎座さよなら公演、「Party が始まるよ」公演
109	効果	2	「いい子にしないとサンタさん来ないよ」効果、「おごりますよ」効果
109	コンテスト	2	「誰が一番安全運転か」コンテスト、日本語話そうコンテスト
109	実験	2	「満員電車には最大何%乗れるのか」実験、「ハロウィンまでにどれだけ作れるか」実験
109	質問	2	「なんでそっち選んだのよ」質問、「どうしたの」質問
109	証明	2	「夫婦仲良くやってるのよ」証明、「受診しましたよ」証明
109	人生	2	「まいったか」人生、「音楽だけが友達さ」人生
109	勢	2	やるならやるぞー勢、遊びじゃねんだよ勢
109	制度	2	「日頃の苦勞忘れるために皆で打ち上げ行こうぜ」制度、「チケットを持っていない方もグッズを買えますよ」制度
109	セール	2	「来年こそは絶対優勝するぞ」セール、「来年こそは絶対勝つぞ」セール
109	センサー	2	ご飯もらえるぞセンサー、やばいぞセンサー
109	宣伝	2	「これで学ぶとこんなに英語ペラペラですよ」宣伝、「共産市政許すな」宣伝
109	戦略	2	「一気にいくぜ」戦略、「押してもダメなら引いてみな」戦略
109	騒動	2	「肖像権の侵害で訴えるぞ」騒動、「2 番じゃダメなんですか」騒動
109	隊	2	地球を守るぞ隊、人間ドック一緒に行くぞ隊
109	大使	2	ふくしま知らなかった大使、浜松市やらまいか大使
109	大賞	2	うわ！ダマされた大賞、「お前が言うな」大賞
109	大臣	2	産めよ増やせよ大臣、「注射怖いだもん」大臣
109	魂	2	負けじ魂、「俺達銀魂愛してるぜー」魂
109	ちゃん	2	構ってちゃん、困ったちゃん
109	党	2	「お前が言うな」党、「もうどうでもいいや」党
109	同盟	2	「子供は遊ぶぞ」同盟、もう何もかもどうでもいいよ同盟
109	トーク	2	「ばれたら大変だよな」トーク、「最近私身体壊して病院行ってるんだよな」トーク
109	能力	2	「なんでやねん！」能力、使わない知識なんて忘れてしまえ能力
109	パーティ	2	またねパーティ、しばらくお別れだねパーティ
109	ハラスメント	2	お前無理だろハラスメント、年収いくら？ハラスメント

109	版	2	みんなで歌おう版、お小遣い稼ぎ頑張るぞ版
109	評価	2	「いいね」評価、「まあいいや」評価
109	プログラム	2	虫歯の一本もない子になろうプログラム、福島の子どもを守ろうプログラム
109	報告	2	「彼氏できたよ」報告、「結婚するよ」報告
109	物語	2	日本やるじゃん物語、次男頑張れ物語
109	モンスター	2	くれくれモンスター、やだやだモンスター
109	問題	2	「いい加減にしろよ」問題、北千住と南千住どっちなんだ問題
109	要請	2	「緊急事態宣言！不要不急の外出は控えてください」要請、ビールもこっち入れる要請
109	力	2	「うまいこと言うな」力、晴れる力
109	ルール	2	「黙って借金なんかしたら許さんぞ」ルール、「3日空けちゃイカンぜよ」ルール
109	レター	2	「相談してね」レター、「今なら電池パック交換サービスするわよ」レター
109	WiFi	2	どんなときもWiFi、どこよりもWiFi
109	割引	2	「いいね」割引、「ブログ見たよ」割引
165	アイドル	1	もっと小顔にしなきゃいけないわよアイドル
165	アンケート	1	あったらいいアンケート
165	ウィーク	1	頑張らなくてもいいんだよウィーク
165	おばさん	1	引っ越せおばさん
165	カード	1	4枚引いてくださいカード
165	会議	1	経済ってそういうことだったのか会議
165	会見	1	「笑っちゃうよ」会見
165	解散	1	バカヤロー解散
165	勧告	1	「遠くから来てる人は早く帰っていいよ」勧告
165	関西人	1	なんやて関西人
165	記念	1	夏休み無事乗り越えたよ記念
165	記念日	1	「親になったぞ」記念日
165	教室	1	経済ってこうなってるんだ教室
165	教祖	1	晴れる～教祖
165	くん	1	「オレがルールだ」君
165	警察	1	マスクつける警察
165	結婚	1	できちゃった結婚

165	研究会	1	「彼はなぜに譜面をみるのか」研究会
165	検証	1	「2 クロンの餌 300 個の釣果はどうなの」検証
165	講演会	1	みやざきの木をたっぷり使って良い家を建てよう講演会
165	構文	1	「カキ料理は広島が本場だ」構文
165	裁判	1	スマホ持つか持たないか裁判
165	作業	1	「あーでもねーこーでもねー」作業
165	試合	1	「取れるときに点が取れないとこういうことになるんだぞ」試合
165	じいさん	1	マスクしろ爺さん
165	シール	1	晴れるシール
165	事件	1	ミッドに巨大ロボが攻めて来たぜ事件
165	自慢	1	「マンション買ったわよ」自慢
165	集団	1	「ええじゃないか」集団
165	需要	1	やっぱり要るな需要
165	状況	1	お父さん頑張ってるよ状況
165	症	1	偏頭痛と一緒にときどき突発的に目まいが来ますよ症
165	図鑑	1	「香川照之の昆虫すごいぜ!」図鑑
165	ストーリー	1	「良かったね」ストーリー
165	制	1	いつでも好きなときに来て下さいね制
165	生活	1	「ま、なんとかなるさ」生活
165	説	1	「64GB モデルもあるよ」説
165	然	1	さあ、いらっしやい然
165	選手権	1	思い出せるか選手権
165	戦法	1	「余ったからあげるよ」戦法
165	族	1	「俺、サッカー好きやねん」族
165	大作戦	1	カップルにマナーを教えてあげよう大作戦
165	態度	1	「そんなことも知らないのか」態度
165	チケット	1	本社まで来てくれたらドリンクおごるよチケット
165	チャンネル	1	山形を熱くしよう!チャンネル
165	調査	1	「日本人は幸福な生活を送る上で、何が必要か」調査
165	ついで	1	「住所録リストにあるから」ついで

165	電車	1	「きっとサクラサクよ」電車
165	展覧会	1	みんなでつくろう展覧会
165	トレーニング	1	「勝手にドアの外に出るんちゃねーぞ」トレーニング
165	並み	1	こっちだって寝てないんだよ並み (の言い訳)
165	ニーズ	1	「メイクを教えて」ニーズ
165	日記	1	「なんでやねん」日記
165	忍者	1	カメラに映っちゃダメ忍者
165	俳句	1	頭悪そう俳句
165	話	1	きつかった話
165	パフェ	1	いちごにキュンですパフェ
165	弁当	1	バカが食うやつだな！弁当
165	待ち	1	そんなことないよ待ち
165	メロディ	1	はよ帰れメロディ
165	モデル	1	「こっち見んな」モデル
165	用	1	後で見よう用 (のメモ)
165	流	1	まっいいか流
165	旅行	1	「ゆっくりでいいさ」旅行
165	令	1	帰りなさい令
165	レポート	1	「こんな人たちがいたよ」レポート
165	練習	1	「調子にのっちゃっあいけないけど、まだ負けないぞ」練習
165	枠	1	特待生になりきれない枠
	合計	384	

上記の調査資料における文包摂名詞の出現状況を見ると、文包摂名詞の後項の名詞・接尾辞の意味的な特徴としては、抽象的な概念を表すものが多いという傾向が見られる。例えば、言語活動を表す形式（「発言」「トーク」「コール」「コメント」「宣言」など）、状態・様相を表す形式（「状態」「状況」「感」「オーラ」「ムード」「風」「調」など）、類型を表す形式（「型」「系」「タイプ」「バージョン」など）、集団を表す形式（「グループ」「組」「クラス」「チーム」など）、活動・イベントを表す形式（「イベント」「祭り」「パーティ」「プロジェクト」「会」「展」など）、働きかけを表す形式（「詐欺」「攻撃」「ハラスメント」「アピール」など）などがある。また、用例数は多くないものの、物体や人物を表す形式（「パフェ」「ドリンク」「カード」「おじさん」「女」など）も見られる。

また、前項の形式的な特徴として、前項は独立語文・述語文、単文・複文、言いさし文、

談話相当の要素など、どのような形式をとる要素であつてもなり得るという点が見られる。前項の表記的な特徴としては、前項の文相当の要素には、感嘆符や疑問符、顔文字や絵文字、鉤括弧などが付された実例が多数見られ、話し言葉や打ち言葉に特有の記号が用いられやすいという点が挙げられる。

ただし、このような形式的な自由度の高さがあるものの、どのような形式の文であっても前項になるわけではない。既述のとおり、後項の名詞や接尾辞が表す意味と、意味的に親和性の高いモダリティを表す文が前項になりやすい。例えば、「キャンペーン」という名詞は、何らかの目的を持って多くの人に呼びかける活動を表すことから、勧誘のモダリティを表す文相当の要素と親和性が高い。そのため、「キャンペーン」と結びつく前項の文末は、勧誘のモダリティを表す動詞の意志形が出現しやすいという形式的な特徴が見られる。

5. 考察

5.1 個人による臨時的な名づけとしての文包摂名詞

実例を観察すると、文包摂名詞は一般的な語と同様に次のような性質が見られる。すなわち、語彙的な側面から見ると、「語」には命名的な機能があり、人・物・事、動き・変化、状態・性質といった現実の個々の断片を一般的に名づける（奥田 1985、村木 2002、石井 2014 など）という性質である。例えば、現実世界にある「平地よりも高く隆起した地塊」を「山」と名づけて一般化しているのと同様に、「振り込め詐欺」という文包摂名詞は「電話などで相手を騙し、指定の口座に現金を振り込ませる犯罪行為」を名づけたものである。

そこで以下では、文包摂名詞が名づけとして用いられていることに着目し、個人が臨時的な名づけを行っている場合を例に挙げ、その使用実態を記述する。以下の例（1）（2）の文包摂名詞は、それぞれ個人が世界の中から切り取った断片を名づけている。

(1) <次回のトライアスロンのレースに向けた水泳の練習にて>

そんな泳ぎに変えるのじゃー。名付けて「みんな来年の俺を見たらビックリするぞ作戦」。ということで今日は体の動きと使うべき筋肉を意識した練習。

（ブログ『ホノルルマラソンに出てから』¹³）

(2) <他の選択肢とは種類の異なる選択肢を選ぶ形式の問題について>

テキストの各章から満遍なく3～5問ずつ出題されており、中にはサービス問題 → 「仲間はずれを探せ」コーナー（勝手に命名）もあります。（ブログ『kimidoring』¹⁴）

例（1）では、「（自身の泳ぐフォームを変えて練習を行うことにより、）来年、自分を見た人たちを驚かせようとする企て」を「みんな来年の俺を見たらビックリするぞ作戦」と名づけている。例（2）では、話者が受験した試験問題のうち、「他の選択肢とは種類の異なる選択肢を選ぶ形式の問題」を「「仲間はずれを探せ」コーナー」と名づけている。命名者が例

（2）のように命名したのは、1つだけ種類の異なる選択肢を選ぶ出題形式を、「仲間はずれを探せ」という言葉で表現し、さらに、試験問題の中でそのような形式の問題が複数出題されている箇所を1つの区画（「コーナー」）のように捉えたからであると推測される。例（1）

（2）に共通しているのは、前項語基が名詞である一般的な複合語の構造（例：「ダイエット

¹³ <http://macoto1127.blog45.fc2.com/blog-date-201308.html>（2017年11月3日閲覧）

¹⁴ <http://xxxkimidorinxxx.blog115.fc2.com/?m&no=179>（2018年11月13日閲覧）

作戦」「間違い探しコーナー」など)では、命名者が表現対象を的確に表し切れないという点であると考えられる。なお、例(1)(2)のそれぞれの用例中には「名付けて」「勝手に命名」(用例中の波線部)という共起表現があり、話者(書き手)自身が名づけ(あるいは造語)を行っていると感じていることが窺える。

また、以下の例(3)の下線部は、既存の病名があるものの、あえて文包摂名詞を用いてわかりやすく再命名している例である。

(3) <病院で診断された病名とその症状について>

病名は『良性発作性頭位目まい症』という何だか長くて、難しそうな名前

まあ、早く言えば『偏頭痛と一緒にときどき突発的に目まいが来ますよ症』だそうです
(ブログ『MASCARADE ~仮面舞踏会~』¹⁵)

上の例(3)の「偏頭痛と一緒にときどき突発的に目まいが来ますよ症」とは、「良性発作性頭位目まい症」(波線部)を正式名称とする病名を指し、突発的にめまいが生じる疾患を表している。このブログの筆者は、「良性発作性頭位目まい症」という難解な病名に対して、それを平易に言い換える形で「偏頭痛と一緒にときどき突発的に目まいが来ますよ症」と換言している。このように表現することで、ブログの読み手に対して診断された病名をわかりやすく伝えることが可能となる。

他にも、文包摂名詞を用いてわかりやすく換言する例には、「4枚引いてくださいカード」という表現がある。これは「UNO」というカードゲームの「Wild Draw Four」というカードを指している。あるプレイヤーが「Wild Draw Four」というカードを出すと、その次のプレイヤーは山札から4枚のカードを引かなければならない。「Wild Draw Four」というカードの名前を「4枚引いてくださいカード」と表現することで、UNOをしたことがない聞き手や英語が分からない聞き手にも、そのカードについてわかりやすく伝えられる。

文包摂名詞の前項が文相当の要素であることにより、命名者は自分が表現したいと思う対象を聞き手にもイメージが喚起されやすい形で提示できるのではないかと考えられる。これは一般的な構造の複合語では得られない「文包摂名詞ならではの表現効果」と言え、文包摂名詞が使用される動機付けの一つではないかと推察される。

人が新語を形成する動機として、ある新しい事態に遭遇した場合や、今まで感じたことのない感情を抱いた場合に、手持ちの言葉ではそれを表現しきれず、それまではなかったような語を作り出して表現しようとする(小野 2018:691-692)。換言すると、表現の対象を的確に表す語が心的語彙の中に見当たらず、類義関係にある語では適切に表現しきれない場合、その表現対象を新たに名づけることとなる。上の例(1)~(3)の用例かのように、文包摂名詞は、既知の語では表現の対象を的確に言い表せない場合や、説明の難解なものを端的に伝達したい場合に有効な表現手段であるということがうかがえる。

文包摂名詞の構造を利用したこのような命名法は、現代日本語において若者を中心に比較的定着しているものではないかと思われる。実際に大学生がそのような型を利用して新しい名づけをしていることが窺える研究には、新語の形成過程について考察した黒崎(2018)がある。そこでは、大学生がグループの中で、「前から歩いてくる人にぶつかりそうになり、左右に避けたところまたぶつかりそうになる現象」に対してどのような命名を行うのかと

¹⁵ <http://requiemwokanadeyou.blog53.fc2.com/blog-category-3.html> (2021年7月21日閲覧)

いう実験を行っている¹⁶。

実験結果を見ると、まず、A グループでは、27 の名前の候補が出されているが、文包撰名詞は用いられていない。B グループでは、14 の候補が出され、そのうちの3つ (全体の約 21.4%) が文包撰名詞を用いた命名である (「お前もか現象」「こっち来るの現象」「こっち来んな現象」)。C グループでは、76 の候補が出され、そのうちの24 の候補 (全体の約 31.6%) が文包撰名詞を用いた命名である (「ぶつかっちゃうよ現象」「気が合うね現象」「どうも、ごめんなさいね現象」「道譲り合っちゃったよ現象」など)。なお、文包撰名詞以外の候補には、「譲り合い現象」「お見合い現象」「ミラーリング現象」「相互回避現象」といった典型的な複合名詞が挙げられている。

通常、多くの人には当該の現象を経験したことがあっても、それを一般に名づける語を心的語彙の中から見つけられるとは限らない¹⁷。黒崎 (2018) の実験結果を見ると、そのような言い表しにくい表現対象に対する臨時的な命名として、文包撰名詞の使用は、少なくとも大学生の間では比較的定着している造語法なのではないかと考えられる¹⁸。

さらに、このように臨時的な名づけによって誕生した文包撰名詞は、徐々にその使用が定着するものも見られる。例えば、地名として用いられる文包撰名詞の例には「おいてけ堀」がある。「おいてけ堀」は、『日本国語大辞典』(第二版)によると、江戸本所 (現在の東京都墨田区南部地域) にあった池の名称であり、本所七不思議¹⁹という伝承の1つでもある。この池で釣りをすると、水中から「置いてけ、置いてけ」と呼ぶ声がし、釣った魚をすべて返すまではこの声が止まなかったという。この「置いてけ、置いてけ」という声を文包撰名詞の前項に引用し、その声のする場所を表す「堀」という名詞を後項に用いることで、当該の池を「おいてけ堀」と名づけているのだと考えられる。「おいてけ堀」が江戸時代における池の名称であることから、文包撰名詞を用いた命名法は、江戸時代には既に存在していたと言える²⁰。

祭りの名として定着している文包撰名詞の例としては、宮崎県日向市美々津町で伝統的に催されている「おきよ祭り」が挙げられる。「おきよ祭り」は、古語「起く」の命令形「起きよ」と、「祭り」とが結びついた文包撰名詞である。宮崎県日向市の公式ホームページ²¹によると、この祭りは、八朔 (旧暦の8月1日) の夜に、子どもたちが「起きよ、起きよ」と言いながら、短冊飾りのついた笹で町の家々の戸を叩いて回る祭りであるという。この祭りの起源は、神武天皇が東征した際の説話であるとされる。神武天皇はこの地から船出したと伝えられているが、風向きの変化によって急遽出発することとなったため、町の人に「起き

¹⁶ なお、実験は、4人ずつ3グループ (A、B、Cグループとする) に分かれて行われている。

¹⁷ このような現象の発生率とその環境要因について考察した堤他 (2020) では、当該の現象は「連続回避本能」と呼ばれていると述べられている。この「連続回避本能」というのが当該の現象を表す専門用語ではないかと思われる。

¹⁸ なお、黒崎 (2018) では、当該の現象に対してどのような名づけを行うかという実験を小学生のグループにも行っている。データを見ると、3つの小学生のグループでは文包撰名詞を使用した命名が1例も観察されないため、小学生には文包撰名詞を用いた造語法が定着していないのではないかとと思われる。

¹⁹ 本所七不思議とは、本所地域 (現在の東京都墨田区南部地域) にまつわる自然や動物の異変、伝説や妖怪に関する伝承のことである (江東区深川江戸資料館『資料館ノート第92号』https://www.kcf.or.jp/cms/files/pdf/original/7905_%E8%B3%87%E6%96%99%E9%A4%A8%E3%83%8E%E3%83%BC%E3%83%8892.pdf 2021年6月22日閲覧)。

²⁰ なお、現代日本語において「置き去りにすること」の意味を表す「おいてきぼり」という語は、「おいてけぼり」が変化した形であり、その語源はこの本所七不思議の故事に由来するとされる (『暮らしのこぼ語源辞典』)。

²¹ <https://www.hyugacity.jp/sp/display.php?cont=140317132835> (2021年6月23日閲覧)

よ、起きよ」と声を掛けて回り、船出の準備をしたという。この由来をふまえると、「おきよ祭り」の前項「おきよ」は、町を回りながら呼びかける「起きよ、起きよ」という声を引用したものであると考えられる。なお、昭和15年4月には、神武天皇の東征2600年記念事業として、「おきよ丸」という船が造られ、進水したという²²。この「おきよ丸」は、「おきよ祭り」の「おきよ」と、船の名前を表す造語成分である「丸」が結びついた文包撰名詞であり、船の名前として用いられている²³。

歴史上の事実の俗称として定着している文包撰名詞の例には、「バカヤロー解散」がある。「バカヤロー解散」は1953年に吉田茂首相が行った衆議院解散の名称である。野党からの追及に対して吉田首相が「バカヤロー」と発言したことが問題となり、内閣不信任決議案が可決され、解散に至ったとされる(『デジタル大辞泉²⁴』)。解散の引き金となった吉田茂首相の「バカヤロー」という発言は、当該の衆議院解散を特徴づけるものであり、それがそのまま文包撰名詞の前項に引用されている。

以上、本節では文包撰名詞が臨時的な名づけとして用いられ、既存の語では表現しきれない対象を名づけ、聞き手にわかりやすくイメージを喚起させられることを述べた。また、文包撰名詞が現代日本語における造語法として比較的定着しているものであり、歴史を遡ると、文包撰名詞を用いて名づけが行われた実例があることも確認した。

5.2 商業・広報などを目的としたネーミングとしての文包撰名詞

文包撰名詞は個人による臨時的な名づけ以外にも、主に商業・広報を目的としたネーミングとして用いられる場合が見られる。「ネーミング」とは、『新明解国語辞典』(第七版)によると、「消費者などに印象づけることを目的とする、商品などの効果的な命名」のことである。本研究では、商品名に限らず、商業や広報を主な目的として、聞き手に印象付けるために戦略的に考えられた名づけのことを「ネーミング」と呼ぶこととする。以下、順に文包撰名詞がネーミングに用いられている例を見ていく。

商品名として用いられている文包撰名詞の例には、「いちごにキュンですパフェ²⁵」という苺が使われたパフェの名称がある。「キュンです²⁶」は2020年頃から若者の間で流行しているフレーズであり、このような一種の定型的な表現を取り入れた文包撰名詞が商品名となっている。このパフェを食べた人物の感動を、流行語を用いて言語化するならば、「いちごにキュンです」という言葉であると想定し、それを文包撰名詞の前項に引用したのでは

²² 宮崎県季刊誌『Jaja』(https://www.pref.miyazaki.lg.jp/contents/org/honbu/hisho/jaja/10_tosen.html 2021年6月23日閲覧)

²³ 「おきよ祭り」と「おきよ丸」では命名の過程が異なる。「おきよ祭り」の前項は「起きよ、起きよ」という呼び声を引用したものであると考えられる。それに対して、「おきよ丸」の前項は「起きよ、起きよ」という呼び声を引用したのではなく、既に祭りの名称として定着した「おきよ祭り」の前項を利用して作り出した二次的なものである。

²⁴ 2021年6月23日閲覧

²⁵ 回転寿司チェーン店「スシロー」が2021年春頃に販売を開始したパフェである。公式ホームページの商品説明には、「旬のフレッシュいちごを贅沢に使用したスイーツ♪チーズをふんだんに使用した濃厚なストロベリーチーズケーキアイスと、甘酸っぱいいちごアイス、いちごのムースやソースと、いちごをたっぷり組み合わせ、いちご好き必見の、心がときめいて“キュン”とするスイーツに仕上げました。」と書かれている(<https://www.akindo-sushiro.co.jp/cafebu/> 2021年6月22日閲覧)。

²⁶ 2020年頃から若者の間で流行しているフレーズで、「胸がきゅんとときめいた」ことを意味する。また、片手の親指と人差し指を交差させて小さなハートマークを作る「きゅんですポーズ」を撮った動画の投稿がTikTokなどのSNSでブームとなった(女性向けメディア『Oggi.jp』<https://oggi.jp/6325547> 2021年6月22日閲覧)。

ないかと考えられる。

番組名に用いられている文包撰名詞の例には、「うわ！ダマされた大賞」がある。「うわっ！ダマされた大賞」（通称「ダマされた大賞」）はいわゆるドッキリ²⁷を仕掛ける番組の名称である²⁸。「うわっ！ダマされた」はまさにドッキリを仕掛けられた人物が、ドッキリであることを明かされた際に発話すると想定される言葉である。番組内で紹介された「ドッキリ」のうち、特に面白かった企画に贈られる賞もまた「ダマされた大賞」と呼ばれ、これも文包撰名詞の一例である。

キャンペーンの名称として用いられている文包撰名詞の例には、「ウルトラ！ゼンリョク！幻のポケモンをもらおうキャンペーン」がある。これは、株式会社ポケモンが販売する関連商品を購入して応募すると、『ポケットモンスター』というゲームの中に登場する珍しいポケモンが入手できるというキャンペーンである。「幻のポケモンをもらおう」は、幻のポケモンを手に入れたらというゲームプレイヤーたちに対して、キャンペーンの実施者が呼びかけ・誘いかけとして発話すると想定されるものであり、それを文包撰名詞の前項に引用したものではないかと思われる。

キャンペーンの名称から派生して、SNS 上のハッシュタグの名称に用いられる例もある（例（4））。

（4）＜森永乳業が製造・販売するアイスクリーム「パルム」のCMについて＞

「#今日のアイスどうする問題²⁹」とは、今日食べるアイスは濃厚なものにするか、さっぱりとしたものを選ぶか、定番商品にするか、新商品にするか・・・など、多くのアイス好きが抱えている、日々の悩める問題です。本CMでは、どんな時もスマートに決断をする竹野内豊さん演じる敏腕部長が、コンビニのアイスケースの前で「#今日のアイスどうする問題」について真剣に悩むお茶目な姿が描かれています。

（森永乳業『パルム』公式ホームページ³⁰）

上の例（4）では、命名者（森永乳業）は用例中の波線部「今日食べるアイスは濃厚なものにするか、さっぱりとしたものを選ぶか、定番商品にするか、新商品にするか・・・など、多くのアイス好きが抱えている、日々の悩める問題」を「今日のアイスどうする問題」と名づけている。アイスの購入に際してあれこれと悩む人々の心内発話として想定されるもの（「今日のアイス、どれにしよう」「今日ほどのアイスを食べようかな」など）を「今日のアイスどうする」という形式で代表させて前項に配置し、そのような悩ましい事柄を「問題」と捉えたことで、「今日のアイスどうする問題」という文包撰名詞が作り出されたのだと推察される。

「振り込め詐欺」「母さん助けて詐欺」（例（5））といった特殊詐欺の名称も、一種の広報目的の表現であると言える。

²⁷ 「人にいたずらを仕掛けて驚かすこと」（『広辞苑』第七版）のことである。

²⁸ なお、この番組は不定期に放送され、2020年に放送されたスペシャル番組では、「うわっ！ダマされた大賞2020」という西暦が付された名称であった（『日本テレビ』公式ホームページ <https://www.ntv.co.jp/topics/articles/19jwcdpkdawkqtxybt.html> 2020年12月28日閲覧）。

²⁹ 「#今日のアイスどうする問題」の「#」という記号は「ハッシュタグ」である。「ハッシュタグ」とは、「ツイッターなどのソーシャルメディアにおいて、特定のテーマについての投稿を検索して一覧表示するための機能。」のことである（『デジタル大辞泉』2021年6月1日閲覧）。

³⁰ <https://parm-ice.jp/cm/>（2021年6月1日閲覧）

- (5) 「振り込め詐欺³¹」に代わる新たな名称を募集していた警視庁は12日、東京都中央区の歌舞伎座で開いたイベントで、応募のあった約1万4千点から最優秀作品として「母さん助けて詐欺」を選んだと発表した。

(『日本経済新聞』(電子版) 2013年5月12日³²)

これらの新名称は、もともとは「オレオレ詐欺」などの名称で呼ばれていたものが、詐欺の手口の巧妙化に伴い、再命名されたものである。

以上、本節では、文包摂名詞が個人による臨時的な名づけ以外にも、主に法人が商業や広報を目的としてネーミングを行う際にも用いられるということを見た。ネーミングではキャッチーな表現が求められると考えられるが、次節では文包摂名詞が持つ性質がそのようなネーミングに用いられやすい基盤になっていることを述べていく。

5.3 文包摂名詞の表現効果

本研究では文包摂名詞が広く使用される動機付けの1つに「新奇性」という表現効果があると考えられる。以下では、文包摂名詞ならではの表現効果である「新奇性」について考察を行う。「新奇性」に関する明確な定義はないが、鈴木(2020)を参考に、本研究では「新奇性」を「受け手側に目新しさや特異さを感じさせ、ユニークでユーモアのある印象やインパクトを与えるという伝達上の表現効果³³」とする。

本研究で文包摂名詞と位置づけられる形式について、上述の「新奇性」という表現効果が見られることを指摘した先行研究には、新屋(2014)および鈴木(2020)がある。新屋(2014)は、後項が名詞「状態」である文包摂名詞(「交換システム作って～状態」「枯れ木?!状態」など)について考察を行っている。そのうえで、後項が名詞「状態」である文包摂名詞には、若者語のコミュニケーション機能の1つである「娯楽機能」(米川1996:16-27)があると述べている³⁴。

また、新しい表現が創発される過程について論じた鈴木(2020:202-204)では、種々の新奇な表現を扱っているが、その中で「めっちゃコンサート行きましたアピール」という文包摂名詞を取り上げている。そこでは、このようなユニークな表現が発話の場で臨時的に創造された際に、受け手側には相応のインパクトがあるものとして受け取られると述べられている。それは、表現者が文包摂名詞のような特異な表現を用いることで、「新しいことをした」「うまいことを言った」というのが受け手側に伝わるからだという。

以下の例(6)および例(7)は、「母さん助けて詐欺」という表現について、ウェブ上の

³¹ 「振り込め詐欺」とは、「電話・インターネットなどで虚偽の請求や懇願をし、指定口座に金を振り込ませる詐欺。」(『広辞苑』(第七版))のことである。また、『デジタル大辞泉』には「オレオレ詐欺」「架空請求詐欺」「融資保証金詐欺」「還付金詐欺」など、銀行口座に現金を振り込ませる手口の詐欺犯罪を総称したもの。平成16年(2004)12月、警察庁が命名。なりすまし詐欺。[補説]近年、現金を手渡しさせる手口が増えたことから、警視庁が実態に見合う新名称を公募。平成25年(2013)5月、「母さん助けて詐欺」「ニセ電話詐欺」「親心利用詐欺」の3点を発表した。」と記述されている(『デジタル大辞泉』<https://dictionary.goo.ne.jp/word/%E6%8C%AF%E8%BE%BC%E3%82%81%E8%A9%90%E6%AC%BA/#jn-195641> 2020年2月2日閲覧)。

³² https://www.nikkei.com/article/DGXNASFK12008_S3A510C1000000/ (2018年7月3日閲覧)

³³ 受け手側が面白さを感じるという点では「娯楽性」、受け手側に何らかの衝撃や影響を与えうるという点では「インパクト性」とも言えるような表現効果である。

³⁴ 「娯楽機能」とは、「会話に笑いを生じさせ、楽しむ機能」(米川1996:16-27)のことである。

質問サイトで投稿された質問とその回答である。例（7）の回答者は、「母さん助けて詐欺」という文包撰名詞をインパクトのある表現であると捉えていることが窺える。

（6）＜「振り込め詐欺」に代わる新名称の「母さん助けて詐欺」に関する質問＞

警視庁が、「振り込め詐欺」に代わる新名称を一般から募集していましたが、このたび、「母さん助けて詐欺」という名称が最優秀作品に選ばれたそうです。優秀作は、「ニセ電話詐欺」「親心利用詐欺」ということですが、この、「母さん助けて詐欺」という名前、どう思いますか？私はむしろ、優秀作だった「ニセ電話詐欺」の方がしっくりきます。

（7）＜上記の質問に対する回答＞

母さん助けて詐欺はいい名称だと思いました。ちょっとどうかな～と言う人も出て来ると思いますが、インパクトを持たせるにはこれくらい印象的なネーミングでないといけません。[……]³⁵

＞私はむしろ、優秀作だった「ニセ電話詐欺」の方がしっくりきます。

これは無難だけど無難ではインパクトがなく効果もないです。

（質問&回答コミュニティ『教えて goo』³⁶）

例（6）では質問者が「母さん助けて詐欺」という名称に対する感想を求めており、それに対して、例（7）では回答者が「インパクトを持たせるにはこれくらい印象的なネーミングでないといけません」（用例中の波線部）と述べている。このことから、受け手側が「母さん助けて詐欺」をインパクトのある印象的な表現だと捉えていることが窺え、文包撰名詞の使用による表現効果として「新奇性」が生じていると言える。

上記の例（6）と同様に、文包撰名詞によって新奇性が発現していることが窺える用例は他にも見られる。新語の形成過程について考察した黒崎（2018）では、大学生のグループの中である現象に対してどのような名づけを行うのかという実験を行っている。グループ内での議論では、名づけの際に文包撰名詞が用いられ、それに対する周囲の反応として、ユーモアやインパクトを感じている様子が見受けられる（以下の例（8））。なお、グループで新たに名前を付ける対象は「前から歩いてくる人にぶつかりそうになり、左右に避けたところまたぶつかりそうになる現象」である。

（8）＜ある現象に名前を付けるというグループワークにて³⁷＞

A 「バシッと決めたいね。」（030364A³⁸）

D 「お前もか現象。」（030365D）

全員（笑う）（030366ABCD）

A 「いや…（笑う）」（030367A）

C 「なんか、違くない？（笑う）」（030368C）

B 「でもそれ、すごい好き！（笑う）」（030369B）

[……]

B 「はあ。」（030389B）

³⁵ [……] は筆者が引用の際に省略したことを示す。

³⁶ <https://oshiete.goo.ne.jp/qa/8084403.html>（2021年5月15日閲覧）

³⁷ グループは4名（A～D）で構成され、全員大学生である（黒崎 2018:11）。

³⁸ 括弧内は黒崎（2018）がグループワーク内で学生の発話や行動に付与した ID である。

A 「お前もか現象。」 (030390A)

B 「うん。面白い。それはすごい好きかもしれん。」 (030391B)

A 「なるほど。」 (030392A)

[……]

C 「なんかもう、お前もか現象って流行語にしか聞こえん。」 (030706C)

全員 (笑う) (030707ABCD)

D 「インパクト強いよね。」 (030708D)

A 「強い。だって忘れないもん。」 (030709A)

B 「うん、絶対忘れない。多分あたしこれからも言う。」 (030710B)

(黒崎 (2018:164-168) 実験データ A-2 より一部記号などを省略して抜粋)

上の例 (8) において、まず大学生 D が当該の現象を「(すれ違いざまに自分と同じ方向に避けようとする相手に対して)『お前も (同じ方向に避けるの) か』と言いたくなるような現象」と捉え、「お前もか現象」と名づけている (030365D)。この命名の直後に、大学生 D と聞き手であった大学生 A~C の全員が笑っていることから (030366ABCD)、この表現形式に面白さを感じていることが窺える。その後のディスカッションでも、大学生 B が「お前もか現象」について「面白い」「すごい好き」(030391B の波線) と述べていることから、このネーミングにユーモアを感じ、好意的に捉えていると推察される。

さらに、命名者である大学生 D 自身も「お前もか現象」について「インパクト強い」(030708D の波線) と評価している。これに反応する形で大学生 A が「強い。だって忘れないもん」(030709A の波線) と述べていることから、忘れられないほどのインパクトがある表現だと感じていることがわかる。続けて大学生 B は「絶対忘れない」「これからも言う」(030710B の波線) と大学生 A に同調していることから、今後も使用したくなるほどインパクトのある表現だと評価していたことが窺える。上の例 (8) の会話参加者の反応をふまえると、表現者自身を含め、その場にいた受け手全員が「お前もか現象」という文包摂名詞をユーモアとインパクトのある表現と評価していることが見て取れる。

このような「新奇性」という表現効果の発現基盤には「逸脱性」があると考えられる。鈴木 (2020:188-192) は新奇性を持った表現形式³⁹のデータを収集し、その新奇性が発現する基盤として最も多いのは、「合成⁴⁰」という造語法における「逸脱性⁴¹」であるとしている⁴²。具体的に言うと、一般に合成語は語基と語基、あるいは語基と接辞の組み合わせによって形成されるが、本来は形態的・意味的に組み合わせざるはずのない要素同士が結びつくことによって生じる「違和感」が新奇性の基盤になっているのだと推察される。

例えば、鈴木 (2020:191) で挙げている「逸脱性」由来の新奇な表現の例には、「めっちゃ

³⁹ 鈴木 (2020) では、新奇性を持った表現形式であると認定するにあたって、「聞いたことがない」「変わっている・面白いという印象を与える」「辞書には掲載されていない表現」などの複数の基準を設けている (p.185)。

⁴⁰ 鈴木 (2020:188) では新奇な表現を形成する方法の呼称として、自立形式同士が組み合わせる場合は「合成」、自立形式に接辞的な形式が結びつく場合は「要素付加」という用語を用いて区別している。一般的には前者は「複合」、後者は「派生」と呼ばれ、2つを合わせて「合成」とすることが多い。そのため、本研究においては鈴木 (2020) の「合成」と「要素付加」の両者を合わせて「合成」と呼ぶことにする。

⁴¹ 鈴木 (2020) では「選択制限の逸脱」という用語を用いている。

⁴² 鈴木 (2020:189) では、「逸脱性」以外に新奇性を発現させる基盤としていくつか提示している。例えば、「特殊な発声」「生起文脈の意外性」「先行発話の奇抜な言い換え」「意味的特殊性のある縮約」などがある。

コンサート行きましたアピール」という文包撰名詞のほかに、「おにぎり中」「小さぶる」などの合成語がある。前者の「おにぎり中」は「おにぎりを食べている最中」という意味である。「最中」の意味を持つ接尾辞「中」は、「授業」「工事」などの活動を表す名詞と結びつくのが一般的である。しかし、「おにぎり中」は物を表す名詞「おにぎり」が「中」と結びついており、その制約を逸脱しているという。

一方、後者の「小さぶる」は「小さく見せようとする⁴³」という意味である。鈴木（2020:191）によると、接尾辞「ぶる」は通常、「偉（ぶる）」「学者（ぶる）」「上品（ぶる）」などの自分を尊大に見せようとするさまを示す語基と結びつくのが一般的だが、「小さぶる」は尊大さを連想させる語基と結びつくという制約を逸脱しているとされる⁴⁴。「おにぎり中」も「小さぶる」も本来は結びつかないはずの要素同士が結合している点で「逸脱性」があり、逸脱していることによる「違和感」が基盤となって受け手側に新奇性を感じさせるのだという。

文包撰名詞が持つ「逸脱性」によって受け手側に違和感が生じていることと、その違和感によって「新奇性」が発現するということは、前掲した「母さん助けて詐欺」をめぐる問答からも窺える（例（9））。

(9) = (7) 母さん助けて詐欺はいい名称だと思いました。ちょっとどうかな～と言う人も出て来ると思いますが、インパクトを持たせるにはこれくらい印象的なネーミングでないといけません。[……]

>私はむしろ、優秀作だった「ニセ電話詐欺」の方がしっくりきます。

これは無難だけど無難ではインパクトがなく効果もないです。

上の例（9）では、回答者（受け手側）は「母さん助けて詐欺」という名称に対して「ちょっとどうかな～と言う人も出て来ると思いますが」（用例中の1つ目の波線部）と述べており、「母さん助けて詐欺」が持つ「違和感」によって、この新奇表現を許容しがたい人が一定数いると推測している。同時に、回答者は「インパクトを持たせるにはこれくらい印象的なネーミングでないといけません」（用例中の2つ目の波線部）と主張しており、「新奇性」を発現させるにはそのような違和感が必要であると考えていることが推察される。「母さん助けて詐欺」は広報を目的に名づけられたものであるため、一般市民への注意喚起を効果的に行ううえでは、インパクトのある印象的な表現でなければならないのであろう。

また、「ニセ電話詐欺」という名称のほうが良いという質問者の意見に対しては、回答者は「これは無難だけど無難ではインパクトがなく効果もない」（用例中の3つ目の波線部）と指摘している。「ニセ電話詐欺」という表現を回答者が「無難」だと感じてしまうのは、これが複合名詞として適格な内部構造を持ち、「母さん助けて詐欺」という逸脱した表現と

⁴³ ただし、「小さぶる」が、「小さ」の意味を消極的に捉え、「自分を過小評価したり卑下したりすることで、自分を取るに足りない存在であるように見せる」という意図で用いられているのか、「小さ」の意味を肯定的に捉え、「（顔や体などを）小さく見せることで、（スタイルなどを）よく見せようとする」という意味で用いられているのか、鈴木（2020）では文脈情報が提示されていないため判断できない。

⁴⁴ 鈴木（2020:191）は、「ぶる」が「偉」「学者」「上品」のような「自分を尊大に見せようとするさま」を示す語基と結びつくとしているが、「悪者ぶる」「被害者ぶる」のように、必ずしも立派だと一般に見なされるものやさまを表す語基だけに後接するわけではない。「ぶる」が「偉」「学者」「上品」などの尊大さを連想させる語基と結びつくために「自分を尊大に見せようとするさま」を表すのであって、「ぶる」自体には尊大さを示す語基と結びつくという意味的な制約はないと考えられる。ここでは単に、前項語基の意味的な制約に逸脱しているのではなく、「小さ」と「ぶる」という定着していない組み合わせ自体が逸脱しているように感じさせるのだと思われる。

比較すると形態的に安定した造語だからではないかと推測される。形態的に安定した表現であればそれだけ「新奇性」が感じられにくくなるのは自然なことである。したがって、文包摂名詞における「新奇性」の発現基盤となるのは、やはり「逸脱性」なのだと考えられる。

広告で用いるといった使用場面の要請を受けて「新奇性」の高い表現を求めるとなれば、表現者は「新奇性」の発現基盤を担保しなければならなくなる。「新奇性」の発現基盤として最も多いのが「逸脱性」であることをふまえると（鈴木 2020:189-190）、文包摂名詞の形成を可能とする1つの要因として考えられるのは次のようなものである。すなわち、「新奇性」を期待する表現者の表現意図に込めるうえで、「逸脱性」こそがその実現に不可欠な要素であるため、文包摂名詞の形成が可能になっているのではないかということである。換言すれば、文包摂名詞は一般的な語形成規則を「逸脱しているにもかかわらず」用いられているのではなく、「逸脱しているからこそ」狙い通りの表現効果が得られ、多少違和感のある表現であったとしても広く使用されているのだと推察される。なお、広告媒体などで文包摂名詞が用いられやすいのは、「新奇性」が宣伝効果に直結するものであり、両者の親和性が高いからだと考えられる。

以上をまとめると次のようになる。「新奇性」とは表現形式の目新しさから、受け手側にユーモアやインパクトを感じさせるといった表現効果のことである。文包摂名詞は語構成上の逸脱性によって生じる違和感があり、この違和感が「新奇性」を発現させる基盤となるのである。

6 本研究で得られた示唆と今後の課題

本研究では、文包摂名詞の考察を通していくつかの示唆が得られた。まず、引用研究の射程が広がった点が本研究の意義の1つであると考えられる。文包摂名詞の前項は文相当の要素であるが、これはほとんどがどこかから引用された言葉であると考えられる。従来、引用研究は構文論や談話研究において文の単位や談話の単位が考察の対象とされてきた（藤田 2000、加藤 2010 など）。しかし、本研究は、文包摂名詞の形成を支える基盤について考察する過程で、引用された言葉が語構成要素として機能することを示した。これにより、引用研究は語形成論において語の単位も考察の対象となり得ることが明らかとなった。

次に、本研究では現代日本語における文包摂名詞の記述および考察を射程としているが、本研究の成果は日本語の歴史的な研究の足掛かりにもなると考えられる。例えば、江戸時代には既に「おいてけぼり」という本研究の研究対象とする形式と同様の形式的特徴を備えた表現が観察される。「おいてけぼり」（「置き去りにすること」の意味）の語源とされる「置行堀（おいてけぼり）」は、「置行（おいてけ）」という「命令文相当の要素」と名詞「堀」が結びついた形式であり、本研究の研究対象とする形式と同様の形式的特徴を備えた表現である。

また、後項「堀」に着目すると、「ほり」が「ぼり」に連濁している。連濁は複合語に生じる現象であることから、「おいてけぼり」は複合名詞相当の形式であると考えられる。なお、「置行堀」は江戸本所にあった池の名称であり、釣った魚を持ち帰ろうとすると「置いてけ、置いてけ」という声が聞こえるという本所七不思議の1つでもある（『日本国語大辞典』（第二版））。

さらに遡ると、平安時代にも本研究の研究対象とする形式と同様の形式的特徴を備えた表現が観察される。以下の例（10）～（15）はいずれも『日本語歴史コーパス』から採取した用例である。

- (10) よくなしごといと多かりや (源氏物語・行幸)
- (11) 事なし草 (ことなしぐさ) は、思ふ事をなすにやと思ふもをかし (枕草子)
- (12) 一日はつれなし顔 (つれなしがお) をなむ (源氏物語・若菜上)
- (13) われもかうに劣らじ顔 (おとらじがお) にぞおはします (堤中納言物語)
- (14) 負けじ心 (まけじごころ) にて、また、散らさじと惜しみおきけることのはをきながらだにぞ今朝はとはまし (蜻蛉日記)
- (15) 負けじ魂 (まけじだまし) に、怒りなば、せぬことどももしてん (源氏物語・玉鬘)

上の用例 (10) ~ (12) はいずれも形容詞の連体形ではなく終止形の形で名詞に接続し、連濁が生じている。連濁は複合語に生じる現象であることから、上述の「おいてけぼり」の場合と同様に、例 (10) ~ (12) の「よくなしごと」「事なし草」「つれなし顔」はいずれも複合名詞相当の形式であると考えられる。

また、例 (13) ~ (15) の「劣らじ顔」「負けじ心」「負けじ魂」も例 (10) ~ (12) と同様に連濁が生じていることから複合名詞相当の形式であると考えられる。助動詞「じ」は未然形・連用形・命令形を持たず、終止形・連体形・已然形がいずれも「じ」の形になる。「顔」「心」「魂」といった名詞が後続しているため、ここでの「じ」は本来ならば連体形であるはずだが、通常の連体修飾であれば後続する名詞に連濁は生じない。したがって、例 (13) ~ (15) の助動詞「じ」は例 (10) ~ (12) と同様に、終止形で名詞に接続していると解釈することも可能ではないかと考えられる⁴⁵。

終止形は文末で言い切る際に用いられる活用形であるため、本来は複合名詞の前項にはなり得ない形式である。本来は文末に来る形式が複合名詞の前項として用いられている点では、例 (10) ~ (15) のような用例も文包摂名詞と類似する現象が生じているのではないかと考えられる。

本研究の研究対象とする言語現象は、先行研究では日本語の一般的な語形成規則に逸脱する例外的な現象とされ、十分に記述や考察がなされてこなかったものである。しかし、本研究の成果は、例 (10) ~ (15) のような平安時代の言語形式にも通ずるものであり、日本語の造語法の通時的な研究にも貢献できる可能性があると思われる。今後はこのような歴史的な観点から文包摂名詞について考察を進めていくことで、日本語における造語の性質の一側面を明らかにし、語形成論に資する研究を目指す。

さらに、通言語的な観点も視野に入れながら、文包摂名詞について考察を進めることも重要であると考えられる。英語には、「Save-the-WHALES campaign」(鯨を救おう運動) や「We love Anne DIAMOND' party」(アン・ダイヤモンド愛してます党) のように、日本語の文包摂名詞に類似する合成名詞が見られる⁴⁶。前者は「Save the WHALES」という命令「文相当の要素」がハイフンを伴って合成名詞の前項になっている。後者は「We love Anne DIAMOND」という「文相当の要素」が引用符を伴って合成語の前項になっている。他にも、「I don't know face」(わからん顔) や、「"Don't Say Gay" bill」⁴⁷ (ゲイと言ってはいけない法案) といった

⁴⁵ ただし、濁音での表記は校訂者によるものであり、この時代のこのような表現に連濁が本当に起きていたのかどうかについては別の傍証を見つける必要がある。

⁴⁶ 英語の用例はそれぞれ、窪菌 (1995:136) の例 (160) a および b である。和訳は筆者によるものである。

⁴⁷ <https://www.harpersbazaar.com/jp/lifestyle/daily-life/a39822575/dont-say-gay-bill-florida-220508-lift3/> (2022年7月20日閲覧)

用例も見られる。英語の例だけにはなるが、文包摂名詞は日本語のみならず、通言語的にも普遍的な現象ではないかという示唆が得られた。

本研究では、文包摂名詞という逸脱表現を取り上げ、その出現状況の調査と使用実態を記述し、名づけの観点から考察を行ってきた。文包摂名詞以外にも、実際の言語使用では、規則を逸脱した言語表現(ら抜き言葉・二重敬語など)が数多く観察される。金澤他編(2021)でも指摘されている通り、このような逸脱表現の産出は、言葉が新しい意味や機能を獲得する過程の一端を示すため、言語変化の要因を読み解く知見になり得る。今後も文包摂名詞の記述および理論的な考察を継続し、現代日本語の言語変化を動的に捉え、そのプロセスの解明に資する研究を目指す。

参 考 文 献

- 石井正彦(2007)『現代日本語の複合語形成論』, ひつじ書房.
- 石井正彦(2014)「語」, 『日本語大事典(上)』, 朝倉書店, pp.725-726.
- 泉大輔(2018)「文が名詞に直接接続する構文の形式的特徴に関する考察」, 『第16回日本語教育研究集会予稿集』, 日本語教育研究集会, pp.26-29.
- 奥田靖雄(1985)『ことばの研究・序説』, むぎ書房.
- 小野正弘(2018)「名前・命名」, 日本語学会編『日本語学大辞典』, 東京堂出版, pp.691-692.
- 影山太郎(1993)『文法と語形成』, ひつじ書房.
- 加藤陽子(2010)『話し言葉における引用表現—引用標識に注目して—』, 日本語研究叢書(25), くろしお出版.
- 金澤裕之・川端元子・森篤嗣編(2021)『日本語の乱れか変化か—これまでの日本語、これからの日本語』, ひつじ書房.
- 金田英里(2014)「接尾辞としての『-感』に前接する要素の拡大」, 『日本語教育論集』第23号, 姫路獨協大学大学院, pp.1-8.
- 窪菌晴夫(1995)『語形成と音韻構造』, くろしお出版.
- 黒崎貴史(2018)「熟議を利用した新語形成プロセスに関する研究」, 山口大学大学院東アジア研究科博士論文.
- 新屋映子(2014)『日本語の名詞指向性の研究』, ひつじ書房.
- 鈴木亮子(2020)「新表現の創発—新しくない中にめっちゃ新しさ見えてるアピール—」, 中山俊秀・大谷直輝編『認知言語学と談話機能言語学の有機的接点—用法基盤モデルに基づく新展開—』, ひつじ書房, pp.183-207.
- 曾睿(2017)「語構成から文構成へ—形態素『-感』と自立語『感』との関わりから」, 『国語学研究』第56巻, 「国語学研究」刊行会, pp.142-155.
- 堤琴里・田村裕一・佐藤克成(2020)「環境要因による連続回避本能の発生率」, 『第25回日本バーチャルリアリティ学会大会論文集(CD-ROM)』, ROMBUNNO.3C1-7. (http://conference.rvs.j.org/ac2020/program/doc/3C1-7_PR0055.pdf 2021年7月1日閲覧)
- 中島晶子(2011)「『系』と『風』の新奇用法について」, 『お茶の水女子大学比較日本学教育研究センター研究年報』第7号, pp.203-207.
- 林四郎(1982)「臨時一語の構造」, 『国語学』第131号, 国語学会, pp.15-26.
- 藤田保幸(2000)『国語引用構文の研究』, 和泉書院.
- 文化庁(2018)『分かり合うための言語コミュニケーション(報告)』 <http://www.bunka.go.jp/k>

oho_hodo_oshirase/hodohappyo/_icsFiles/afieldfile/2018/04/09/a1401904_01.pdf (2019年2月8日閲覧)

村木新次郎 (2002) 「第3章 意味の体系」, 北原保雄監修・斎藤倫明編『【新装版】朝倉日本語講座4 語彙・意味』, 朝倉書店, pp.54-78.

メイナード・K・泉子 (2004) 『談話言語学—日本語のディスコースを創造する談話・レトリック・ストラテジーの研究—』, くろしお出版.

メイナード・K・泉子 (2008) 『マルチジャンル談話論—間ジャンル性と意味の創造』くろしお出版.

山下喜代 (2005) 「漢語と文体—漢語接尾辞を含む合成語と引用表現を中心にして—」, 中村明・野村雅昭・佐久間まゆみ・小宮千鶴子編『表現と文体』, 明治書院, pp.87-97.

米川明彦 (1996) 『現代若者ことば考』, 丸善.

辞書・事典・データベース・関連 URL

『CaboCha』 <http://code.google.com/p/cabocha/> (2019年7月1日閲覧)

『暮らしのことば語源辞典』 (1998), 講談社.

『現代日本語書き言葉均衡コーパス』, 国立国語研究所 (<https://chunagon.ninjal.ac.jp/bccwj-research> 2021年3月11日閲覧).

『広辞苑』 (第七版) (2018), 岩波書店.

『国語研日本語ウェブコーパス』, 国立国語研究所 (http://pj.ninjal.ac.jp/corpus_center/nwjc 2021年3月11日閲覧).

『新明解国語辞典』 (第七版) (2012), 三省堂.

『精選版 日本国語大辞典』 (2005-2006), 1~3巻, 小学館.

『茶器』 <http://sourceforge.jp/projects/chaki/> (2019年7月1日閲覧)

『デジタル大辞泉』 (<https://dictionary.goo.ne.jp/jn/> 2021年9月28日閲覧)

『日本国語大辞典』 (第二版) (2000-2002), 小学館.

『日本語歴史コーパス』, 国立国語研究所 (<https://chunagon.ninjal.ac.jp/chj/search> 2021年3月11日閲覧).

『MeCab』 <http://mecab.sourceforge.net/feature.html> (2019年7月1日閲覧)

『例解新国語辞典』 (第七版) (2010), 三省堂.

「のだ」の〈言い換え〉用法に接続詞が前置する条件の一考察

—新書テキストを素材として—

石原 佳弥子

Examining the Conditions under which the Subjunctive Is Employed in the Paraphrase Usage of Noda: Using New Text Material

Kayako Ishihara

要旨

庵 (2018 : 67) は日本語教育の観点から平叙文の「のだ」の主な用法として〈言い換え〉〈理由〉〈状況に対する解釈〉の3つを挙げている。この中の〈言い換え〉用法は、前の文(脈)の内容を繰り返し、別の表現で言い換えた文に後置する「のだ」の用法を指すものだが、この用法の文頭には「つまり」「すなわ(即)ち」「よう(要)するに」といった言い換えの接続詞が前置する文と、前置しない文とが見られる。そこで本稿では新書作品のコーパスを作成して、接続詞が前置する「のだ」文を抽出し、接続詞が前置する条件を考察した。その結果、同一の主題(topic)の元でひとまとまりとなった複数の文の中で、〈言い換え〉用法の「のだ」文の主題が、言い換えの文の主題と同一であることがわかりにくいテキスト的条件にあると接続詞が前置する傾向にあることがわかった。

1. はじめに

〈言い換え〉用法の「のだ」は、テキストの前方の文(脈)と同じ内容を、異なる表現で繰り返し述べている文に後置するマーカーだが、同じように「言い換え」を意味する「つまり」「すなわ(即)ち」「よう(要)するに」といった接続詞が文頭に前置する文と、前置しない文が見られる。本稿では、新書を調査の対象として、〈言い換え〉用法の「のだ」文の文頭に接続詞が前置する条件について主題(topic)に焦点を当てて考察する。

2. 先行研究

2.1 〈言い換え〉用法の「のだ」

庵 (2018 : 68) は「のだ」の〈言い換え〉用法の例として(1)を挙げ、この「のだ」は「(彼は)16歳から18歳までカナダで過ごした」と「カナダの高校で勉強した」が意味的に等価な関係にあることを示し、前の文(脈)を言い換えているというマーカーであり、テキスト的に動機づけられたものだという。また、〈言い換え〉用法の「のだ」は「わけだ」に置き換えることができるものが多いという。

(1) 彼は16歳から18歳までカナダで過ごした。カナダの高校で勉強したのだ。

庵 (2018)

清水 (2012 : 84) は学術論文を分析し、「のだ」文の助詞「は」に率いられる名詞句が段落内の前文にある場合、その「のだ」文は前文までの内容とつながりを持ちながら前の内容をまとめており、前文にない場合にはその段落の内容だけでなく、さらにその前の段落に出

現した名詞句を「のだ」文の「は」と共に出し、「のだ」文の所属する段落やそれ以前の段落をまとめあげていると述べている。本稿では、清水(2012)が指摘する前者の「は」、つまり、段落内の前文にある「は」に率いられる名詞句を〈言い換え〉用法の「のだ」文の主題(topic)と呼ぶことにする。この「主題」は、「のだ」文の「前提」と呼ばれることもある。

奥田(1990:186)は「のだ」と「のだ」の前方の文との関係を〈説明の文〉と〈説明される文〉の対立とし、〈説明の文〉は〈説明される文〉に後続する「つけたし」な運びが原則的だという。特に小説では〈説明される文〉が心理的な活動を差し出し、〈説明の文〉がその心理的な活動を方向づけている出来事を差し出すことがあり、その場合、〈説明の文〉に差し出される出来事はそんな風に思い、感じることの〈理由〉となることがあるとして(2)の例を挙げている。

- (2) まだ子どものくせに胸ばかりおおきく発達している。ちぐはぐの感じの、生意気そうな娘だった。マダム蘭子はこういう娘には本能的な反感をもっていた。つまり、女としてのかたちがくずれているのだ!。奥田(1990)

確かに(2)は「マダム蘭子がこういう娘に反感を持つ」理由が「女としてのかたちがくずれている」と「のだ」文で述べられている解釈が可能だ。しかし、(3)のように、この例文よりさらに前の文までさかのぼって見ると、「のだ」文の主題を「マダム蘭子」とすることは難しいように感じられる。

- (3) やがて三人の一组が来た。すぐそのあとから六人の一组がやって来て、スタンドはほとんど満員になった。六人組の方には若い娘が二人まざっていた。ひとりは少しずんぐりした、顔色の青い娘で、髪をばさばさに大きくひろげている。オーヴァーの下に赤いスーターを着て、グレイのスラックスをはいていた。まだ子どものくせに胸ばかりおおきく発達している。ちぐはぐの感じの、生意気そうな娘だった。マダム蘭子はこういう娘には本能的な反感をもっていた。つまり、女としてのかたちが崩れているのだ。……それが影山径子だった。(洒落た関係)

(3)は、「ひとりは」から「影山径子」についての叙述が始まり、「影山径子」について述べるひとつのまとまったテキストを成している。つまり、「のだ」文の主題は「マダム蘭子」ではなく、「生意気そうな娘＝影山径子」と考えられるので、本稿では、この「のだ」を〈言い換え〉用法と解釈する。このように言い換えの接続詞「つまり」「すなわ(即)ち」「よう(要)するに」が前置した「のだ」文はすべて〈言い換え〉用法とみなして考察を進めることにする²。

また、石原(2021:28)は、〈言い換え〉用法の「のだ」は、「のだ」文と言い換えられた文とが同一の主題を持ち、「のだ」文の主題が非顕現となるスタイルが新書の中で一番多く見られることを指摘し、このようなテキスト的条件にあるものが「のだ」の〈言い換え〉用法の典型だと述べている。言い換えられる文が複数文あっても、文脈から「のだ」文の主題

¹ 『洒落た関係』石川達三(1966)文藝春秋

² 庵・三枝(2013:55)は、「前の文の内容を別の文で言い換えたときには普通「のだ」を使い、「つまり、すなわち、要するに」とともに使うことがある」と述べている。

が容易に判断できれば、「のだ」文の主題は非顕現になることが多いという。例えば、(4)は「脳は」という主題 (topic) の下にテキストがひとつにまとめあげられており、「のだ」の言い換えの始点は主題「脳は」が顕現する文になる。

ちなみに、(4)の「のだ」は、1つの話題の下にまとまったテキストの最後に現われていて、「のだ」を省略して読んでもテキストの結束性が維持されるので、石原(2021)は、〈言い換え〉用法の下位に位置する〈結論〉用法³と考えている。

- (4) 脳は退屈が嫌いだ。「何も新しいことを考えるな」と命じられると、手持ち無沙汰のあまり、思い出を材料に「不安」「焦り」「嫉妬」といったゴミのような感情ばかり作り出す。逆に、考えるネタをふんだんに与えれば、「楽しい、もっとやりたい」という感情を放出する。子どものように単純なのだ。⁴ 石原(2021)

2.2 〈理由〉用法の「のだ」

「のだ」の〈理由〉用法についても概観しておく。庵(2012:246)は「P。Qのだ。」という文でQがPの理由を説明するとき「のだ」は〈理由〉用法になるという。この用法の例として庵(2018:67)は(5)を挙げている。

- (5) 昨日はいつもより早く家に帰った。結婚記念日だったのだ。 庵(2018)

石原(2021)は「のだ」の〈理由〉用法は「S2だ。S1のだ⁵。」という文を「S1ので、S2だ。」に置き換えられるものだという。つまり、もともと「S1ので、S2だ。」という1つの文が「S2だ。S1のだ。」と2つの文に分かれたものが〈理由〉用法であり、したがって、本来、主節である「S2だ。」に対して「S1のだ。」が従属する度合は高く、ノデ節に相当する「S1のだ。」には南(1974)の階層構造のB類(「が」の文)までしか現われないと石原(2021)は述べている。本稿でも〈理由〉用法の定義を石原(2021)と同じように捉えることとし、元々1つの文であった「S2だ。S1のだ。」という2つの文では、述べられている事態(命題)は1つと考える。

3. 調査方法

新書作品10冊のコーパスを作成し、言い換えの接続詞「つまり」「すなわ(即)ち」「よう(要)するに」が前置する「のだ」文を抽出して考察を行う。コーパス化した作品は用例出典に記したものとする。

³ 石原(2021)は、前方の文(脈)を言い換えるだけでなく、話題の境目の最後に置かれ、結論的な叙述をする「のだ」を〈結論〉用法と呼んでいる。〈結論〉用法の「のだ」は、後続する文との関係性を強く表わす必要がないことから、「のだ」を省略して読んでもテキストの結束性は維持され则认为している。

⁴ 『すべての教育は「洗脳」である』堀江貴文(2017)

⁵ 石原(2021)では「S1だ。S2のだ。」(「S2ので、S1だ。」)と表記しているが、文の流れからすると「S2だ。S1のだ。」(「S1ので、S2だ。」)と表記するほうがふさわしいため、本稿では後者を用いることにする。

4. 考察

4.1 「のだ」の〈言い換え〉用法と接続詞

新書では、(6) (7) のように「つまり」「すなわ（即）ち」「よう（要）するに」という接続詞が前置するのではなく、「一言でいえば」「何が言いたいか」というとなどが「のだ」文に前置して言い換えを表わす例が見られたが、調査の対象は接続詞「つまり」「すなわ（即）ち」「よう（要）するに」の3つに限ることとする。接続詞の前置する文と接続詞以外が前置する文とは意味的にも、文の構造的にも異なるものとなるからだ。

(6) しかしこの大災害は、リスボン以降続いてきた大災害とは質が違う災害であったように、僕には感じられた。一言でいえば、建築をいくら「強く合理的で大きく」しても、この大災害には対抗できないと感じられたのである。（建築）

(7) だから、少額で投資している人にはやっていたのだが、これを買うこと自体はそれほど「おめでたい」わけではない。

何が言いたいかという、その状況判断をするために「おめでたさ」が有効になるのである。大暴落が買い場になるからだ。（おめでたい）

「のだ」文の文頭に前置した「つまり」「すなわ（即）ち」「よう（要）するに」の数を作品ごとにまとめたものが表1である。

表1 〈言い換え〉用法の「のだ」に前置する接続詞数

作品名	文数	つまり	すなわち	要するに	接続詞 合計	「のだ」 言い換え用法
おめでたい人の思考は現実化する	1,442	6	2	5	13	71
すべての教育は「洗脳」である	2,501	27	0	0	27	159
食べもの探訪記	1,966	1	0	0	1	30
小さな建築	1,829	0	1	0	1	114
テツに学ぶ楽しい鉄道旅行入門	2,202	0	0	1	1	76
二軍監督の仕事	2,142	6	0	0	6	57
農業新時代	2,256	0	0	1	1	52
武士の家計簿	2,683	4	0	0	4	96
料理は女の義務ですか	2,380	2	0	0	2	46
和食の知らされる世界	2,271	11	0	1	12	85
総計	21,672	57	3	8	68	786

表1からは、接続詞の出現数が作品ごとに大きく異なり、「のだ」の〈言い換え〉用法の数と接続詞の数が比例するわけではないことがわかる。つまり、言い換えの接続詞が〈言い換え〉用法の「のだ」に前置するのは、何らかの理由があると考えられる。そこで、どのような条件の下で接続詞が用いられるかを「のだ」文の主題に焦点を当てて考察を進める。

4.2 〈言い換え〉用法に接続詞が現れる条件

表2は、接続詞が前置する〈言い換え〉用法の「のだ」文を、述語に係る主語が「は」の

文、「が」の文、そして、「その他」の文の3つに分類し、その数をまとめたものである。「は」の文には「とは」も含む。

表2 〈言い換え〉用法の「のだ」文の文ごとの接続詞数

	「は」の文	「が」の文	その他	合計
つまり	23	10	24	57
すなわ(即)ち	0	0	3	3
よう(要)するに	3	0	5	8
合計	26	10	32	68

表2からは「は」の文、「が」の文、「その他」の文のすべてで接続詞の前置が見られ、特に「その他」の文で接続詞の前置が一番多いことがわかった。

そこで、まず、「その他の文」で接続詞が前置した例を挙げる。

- (8) しかし、この手の詐欺師は、それにもめげずに毎日100回でも200回でもかけ続けることができるのである。

犯罪者たちのやっていることだから、電話をかける役にはブラック企業以上の締め付けがあるのだろうが、それにしても一日中かけ続けるのは並大抵のことではない。

どれだけ断られても、怒鳴られても、ひたすら電話をかけ続けられる理由のひとつは、彼らは確率論で考えないからだろう。つまり「ひとり‘カモ’をひっかければ300万円だ」と思えば、打率は関係ない。要するに1日に1件、それどころか、1週間に1件カモに出会うことができれば、相当の売り上げを確保することができるのである。だからこそ、どんなに失敗を重ねても、チャレンジの回数を増やすことが目的にかなっている。(おめでたい)

- (9) ネットネイティブ世代は、「画像がすぐ消える世界」の価値を直感的に見抜く。数秒で消えるなら、どんな変顔写真だろうがお互い瞬間的にネタにして終わりにできるし、データが残って「黒歴史」になることも避けられる。毎日「いいね!」を稼がなければ、フォロワーを増やさなければというストレスもない。つまり、どこまでも気楽なのだ。彼らにとって、余計な情報やモノは、むしろウザいゴミにすらなる。(教育)

(8)の「のだ」文には「要するに」と言い換えの接続詞が前置している。(8)の「のだ」文の主題(topic)は「彼らは」と考えられるが、「彼ら」はさらに前の文脈にある「この手の詐欺師」を言い換えたものである。また、「のだ」文には主題が顕現していないので、言い換えの文と同一主題であることがわかりにくくなっている。

一方、(9)は「のだ」文の主題(topic)が「ネットネイティブ世代は」とわかりやすく、前掲の(4)とテキスト的条件が似ているように感じられる。しかし、(9)は「のだ」文とその直前の文を「どこまでも気楽なので、毎日「いいね!」を稼がなければ、フォロワーを増やさなければというストレスもない。」と「S1なので、S2だ。」に置き換えられ、1つの事態を述べる文が成立する。つまり、接続詞の前置がなければ〈理由〉用法の解釈のほうが妥当となる。

本稿では、〈理由〉用法を野田 (1997) の「スコープの「の (だ)」」⁶に相当するものと捉え、〈言い換え〉用法を野田 (1997) の「モダリティの「のだ」」⁷に相当するものと捉えている。一般に、「のだ」の用法は語用論的な問題と考えられているように、「スコープの「の (だ)」」と「モダリティの「のだ」」は連続性を持つものだが、テキストを正確に理解するためには、〈理由〉用法と〈言い換え〉用法の「のだ」を正しく解釈する必要がある。そのため、〈言い換え〉用法の「のだ」の主題がわかりにくくなっているときに言い換えの接続詞が前置され、「のだ」という文末の情報からだけでは解釈が難しい〈言い換え〉用法を文頭で明示するものと考えられる。

次に「のだ」文が「が」の文となった例を (10) に挙げる。

- (10) たとえば、こんな二人を考えてみよう。
一人は、自動車免許を持っている、早大卒の商社マン。
もう一人は、ムエタイの達人で、歌手活動もしている女子高生。
あなたは、どちらに価値を感じるだろうか。
まず、前者のような会社員はどこにでもいる。しかし、後者のような高校生を探そうと思ったら骨が折れるに違いない。つまり、後者の方がより“レアキャラ”なのだ。(教育)

(10) のテキストは、主題 (topic) を「自動車免許を持っている、早大卒の商社マン」と「ムエタイの達人で、歌手活動もしている女子高生」の双方と解釈できるが、テキスト中ではこれらの主題となる名詞句に「は」が後置して、明示的に導入にされておらず、このテキスト部分での主題がわかりにくくなっている。

さらに、(10) では、「のだ」文が前の文脈の「あなたは、どちらに価値を感じるだろうか」という問いの答えとも解釈でき、その場合には「が」を菊地 (1997) の「解答提示⁸」と読むことができ、文末にモダリティ形式を持たない「が」の文の解釈が妥当となる。しかし、言い換えの接続詞が前置することで、書き手 (1人称) の存在が示唆され、文末にモダリティ形式を持つことが暗示されるため、「が」の文ではなく「は」の文、つまり独立性が高く、前の文を言い換えている文という解釈が可能になると本稿では考える。接続詞は書き手の存在を暗示し、文末にモダリティがあることを示唆し、意味的側面から独立性の高い文であることを示すものと本稿では捉える。このように考えると、「のだ」の主題の存在を暗示する接続詞が「その他」の文で一番多く見られる理由の説明がつかうのではないかと。

⁶ 野田 (1997) は、「のだ」の用法を大きく 2 つに分け、述語以外の対象語が否定などの範囲に入るものを「スコープの「の (だ)」」と呼んでいる。

⁷ 野田 (1997) は、話し手の心的態度を表すものを「ムードの「のだ」」と呼び、「スコープの「の (だ)」」は「準体助詞の「の」 + 「だ」と組成のままに近いものだが、「ムードの「のだ」」は話し手が事態をどう捉え、聞き手にどう伝えるかという心的態度を表し、「のだ」と一語化したものだという。ただし、野田 (1997 : 34) は、「スコープの「の (だ)」」もガノ可変が成立しないことから助動詞として一語化したものと認めなければならず、一語化はしているが「の」 + 「だ」という組成に近いものとして位置づけるとも述べている。

⁸ 菊地 (1997 : 105) は「が」の〈総記〉の用法を〈解答提示〉と呼んでいる。〈解答提示〉は「関心の対象」であればその読みが可能になるという。(10) では前の文脈で「あなたは、どちらに価値を感じるだろうか。」と質問が提示されており、その答えが「のだ」文において「ooが〈プラス評価〉」と解答が提示されていると解釈することが可能だ。

それでは、「は」の文ではどのような条件において接続詞が前置するのか。その例を(11)(12)に挙げる。(11)は「のだ」文に「6番は」と主題(topic)が顕現している。言い換えの始発となる文の主題も同じはずだが、テキスト中には「は」でなく「(僕は)6番打者こそ」と「こそ」で導入されている。そのため、「のだ」文には主題が顕現しているわけだが、言い換えの始発文の主題と「のだ」文の主題が異主題と解釈される恐れがあり、接続詞の前置が見られる。

他方、(12)は言い換えの始点が「日本の企業は」の文で、「のだ」文の主題も同じと考えられるが、「どんなに企業の業績が上がっても給料は」と対比用法の「は」が現われているため、言い換えの文と「のだ」文が同一主題であることがわかりにくくなっており接続詞が前置している。

(11) 僕は6番打者こそ「ポイントゲッター」になれると思っている。なぜなら、クリーンナップは「OPS」の高い選手が並んでいるから(OPSとは、出塁率と長打率の総和で、アメリカでは打者を評価するにあたって重要な指標とされ、スポーツ専門のサイトでは、OPSがボックススコアに記されるほど)、6番に打順が回る時は、ランナーがいる可能性が高い。つまり、6番はかなり高い確率でチャンスが巡ってくる打順なのである。(二軍)

(12) あるいは「給料なんて上げたら会社がつぶれる」「ストなんかするやつは自分勝手なサヨクだ」と言われて、頭から信じている人。アベノミクスで景気がよくなる、幸せになれると思っている人はおそらくそれに近い。

アベノミクス以前、超不景気といわれていた時代でさえ、日本の企業は内部留保を260兆円も貯め込んでいた。それが今350兆円を超えている。つまり、どんなに企業の業績が上がっても給料は上がらないのである。(おめでたい)

このように言い換えの文の主題が「は」で導入されていない場合、または「のだ」文に対比用法の「は」が現われて、言い換えの文と「のだ」文の主題が表層的に同一ではないように見える場合に接続詞の前置が見られた。

本稿の調査では、前の主題が「のだ」文にまで継続し、非顕現となるといった、石原(2021)が〈言い換え〉用法の典型とするテキスト的条件において言い換えの接続詞が前置していたのは68例中2例であった。

4.3 考察のまとめ

「のだ」文を述語に係る主語という観点から「は」の文、「が」の文、「その他」の文と3分類し、それぞれの文で言い換えの接続詞が前置する例を記述し、その前置の条件を考察した。まず、接続詞が前置するのは、「その他」の文が一番多いことが明らかとなった。そして、接続詞の前置は、1つの主題の下でひとまとまりとなったテキストにおいて、「のだ」の「は」に前置する名詞句が言い換えの始まりの文の主題と同一ではない、または、同じ主題となっていることがわかりにくいといったテキスト的条件にあるときに見られた。

5. 結語

言い換えの接続詞は、その前置によって書き手(1人称の主題)の存在が暗示され、「のだ」文がモダリティを持つものであることが示唆される。そのため、独立性が高く、1つの

事態を述べる文、つまり〈言い換え〉用法の「のだ」の解釈が可能、もしくは容易になると本稿では結論づける。

言い換えの接続詞は、あくまで文末の「のだ」で明確に示しきれないテキストの関係性を文頭で補足するものである。

参考文献

- 庵功雄（2012）『新しい日本語学入門第2版』スリーエーネットワーク（2001年を改訂）。
——（2018）『一歩進んだ日本語文法の教え方2』くろしお出版。
- 庵功雄・三枝令子（2013）『日本語文法演習 まとまりを作る表現：指示詞，接続詞，のだ・わけだ・からだ』スリーエーネットワーク。
- 石原佳弥子（2021）「テキストにおける「のだ」の〈言い換え用法〉と〈理由用法〉：「は」と「が」の情報構造の焦点から」『一橋大学国際協力交流センター紀要』3，pp.27-38。
- 奥田靖雄（1990）「説明（その1）：のだ，のである，のです」『ことばの科学4』，pp.173-216，むぎ書房。
- 菊地康人（1997）「「が」の用法の概観」，川端善明・仁田義雄編『日本語文法 体系と方法』，pp.101-123，ひつじ書房。
- 清水まさ子（2012）「学术论文でノダ文はどのように用いられているのか」『日本語／日本語教育研究』3，pp.73-89，ココ出版。
- 野田春美（1967）『「の（だ）」の機能』くろしお出版。
- 南不二男（1974）『現代日本語の構造』大修館書店。

用例出典

『「おめでたい人」の思考は現実化する』和田秀樹（2016）小学館，『すべての教育は「洗脳」である』堀江貴文（2017）光文社，『食べもの探訪記』吉本隆明（2001）光芒社，『小さな建築』隈研吾（2013）岩波書店，『テツに学ぶ楽しい鉄道旅入門』野田隆（2016）ポプラ社，『二軍監督の仕事』高津臣吾（2018）光文社，『農業新時代』川内イオ（2019）文藝春秋，『武士の家計簿』磯田道史（2003）新潮社，『料理は女の義務ですか』阿古真理（2017）新潮社，『和食の知られざる世界』辻芳樹（2013）新潮社

Twitter にみられる特徴的な慣用表現「名前をつけたい」に関する考察

星野靖子(放送大学文化科学研究科)†

A Study on a Twitter-specific Idiomatic Expression ‘Need a Name for’

Yasuko Hoshino (School of Graduate Studies, the Open University of Japan)

要旨

本稿では、Twitter にみられる一種の慣用表現「名前をつけたい」のコミュニケーション論的特徴を明らかにすることを目的とする。従来は「子供には季節を感じる名前をつけたい」「ファイルに別の名前をつけたい」等の人やモノを対象とする命名表現だが、Twitter では「試験前に部屋を片付けたくなる現象に名前をつけたい」などの個人的な出来事や感情を述べる特徴がみられ、抽象的な名詞とは対照的に名詞修飾節の内容が個別具体的である点から、「名前」から想起される一般性の高さに反して認知意味論的なミスマッチが生じている。そこで、Twitter の用例を収集・分類し、国語研現代日本語書き言葉均衡コーパスの用例を比較した結果、①当該表現は2007年以來 Twitter で頻出し、その大半は命名を意図しないこと、②後続の抽象名詞に通時的変化がみられ、共起語「現象」とあわせて慣用表現化していることが明らかになり、③「名前をつけたい人生だった」等のメディア特有の変異形が確認された。

1. 研究の目的と背景

本稿の目的は、Twitter(ツイッター)にみられる特徴的な慣用表現「名前をつけたい」について実例に基づき検討を行い、その形式的な特徴およびコミュニケーション論的な特徴を明らかにすることである。Twitter をはじめとするソーシャル・ネットワーキング・サービス(SNS)やブログ、動画サイトでは「Nに名前をつけたい」形式の文がよくみられる。

(1) エコバッグをうまく取り出せなくて、レジ前でなんともいえない空気が流れる現象に名前をつけたい。

(2020年7月15日 キングジム公式)

(2) うろおぼえの情報が検索して見つからないとすごく不安になる。この不安に名前をつけたい。

(2008年8月20日 平川哲生)

上の(1)(2)に共通する特徴は、①必須要素の対象を示す二格+願望を表す「二格+つけたい」の構文であり、②前部要素の後続部には抽象名詞「現象」や気持ちを表す名詞「不安」が用いられる¹。③前部要素のテンスは現在形(「流れる」「不安になる」=スル、ナル)であ

† yasukohosn[at]gmail.com

¹ Twitter 用例は原文のまま掲載し、投稿年月日を付記する。2022年8月現在アクティブな企業や著名人の公式アカウントのツイートにはユーザー(投稿者)名を併記する。用例中の

る。③前部要素の内容は、投稿者の日常体験(「エコバッグをうまく取り出せない」)や心情(「情報が見つからず不安になる」)である。③のような個人的な体験や心情に名前をつけるのは、事物を一般化するための名称付与行為とは異なる行為であり、内容が個別具体的であるにもかかわらず、その属性である後続名詞が抽象概念を表す「現象」である点にコミュニケーション論上のミスマッチが生じている。

「名前をつける」は一般的な動詞述語文である。しかし、「つける」に願望を表す助動詞「たい」が付いた「名前をつけたい」は、従来の用法においてきわめて少ない。「国語研現代日本語書き言葉均衡コーパス(以下、BCCWJ)」を「中納言」で検索したところ、「名前をつ(付)け～」の検索結果は、名前を"つける"、"つけて"、"つけない"、"つけました"等の 393 例あったが、"名前をつけたい"のフレーズを検索した結果は 14 例のみだった。

表 1 BCCWJ 「(～に)名前(を)(が)つ(付)けたい」 検索結果

BCCWJ	出現数
名前(を)(が)つ(付)けたい	14
名前をつ(付)け[たい]	393

この BCCWJ にみられた「名前をつけたい」の希少な 14 例は、以下の(3)～(5)のような用例である。

- (3) 将来、お子さんが仕事などで海外で活躍するようになったときにも、外国の人たちに親しんでもらえるような名前をつけたい。そう思う方もいらっしゃるはず。

PB51_00109 秋月智朱(監修)『しあわせ赤ちゃんのすこやか名前事典』

- (4) 大きな白い羽を広げて、海洋を滑空するグライダーに似た優雅な鳥を人類はどうしても残さなければなりません。そのためにも、アホウドリを改名して、もっと美しく賢そうな名前をつけたいものです。

PB24_00012 中村幸昭(著)『鳥羽水族館館長のジョーク箱』

- (5) 「保存ファイル一覧」ダイアログボックスの一覧でファイルを 1 つ選択し、そのファイルに別の名前を付けたい場合には、「ファイル名」に新しいファイル名を入力する。

PB15_00178 岡村友之(著)『しっかりわかるホームページ・ビルダー-version 6』

(3)と(4)の命名対象は、それぞれ人、鳥である。人や鳥にはその上位概念として「生物」>「動物」という抽象体系がすでに存在する(森岡・山口 1985, pp.30-60)。(3)と(4)に修飾語「～な」(「親しんでもらえるような」、「賢そうな」)が付いているのは、人や鳥の属性や固有性、個性を表す意図である。(5)は、名付けの対象となるデジタルファイルが多数存在するなかで、他との違いを名に表すことで対象を識別する(=ラベリング)意図である。(3)～(5)のいずれも命名対象は生物やモノであり、Twitter の用例(1)「現象」(2)「不安」のような、抽象概念や感情を対象とした例は BCCWJ の検索結果にはみられない。

下線は筆者によるものである。

さらに、Twitterの「～に名前をつけたい」には、下記(6)(7)のような「あるある」「わかる」といった共起語を伴う例が多い。

(6) 中学まで一応軟式だけはやきうやってたから大丈夫👉w

って言ってもこの前行った時、たった3ゲームでまめできて終わった……泣
やきう見てると打ちたくなる現象に名前付けたいってくらいあるあるな事

(2022年7月17日)

(7) A: 1時間作業しようと思って結局5時間ぐらい作業してしまいなんだこれとなる現象に名前をつけたい

B: わかります!いつも決めた予定の倍以上の時間は必ずかかる www

(2015年1月28日くまみき)

(6)(7)には、①二格+願望構文、②前部要素のテンスはスル、ナルなどの現在形、③投稿者の日常体験や気持ちを述べる内容、という(1)(2)と共通の特徴がある。(6)では、野球(原文「やきう」)の試合を見て、自ら(バットでボールを)打ちたくなかったという投稿者自身の体験を「あるあるな事」と述べていることから、以前にも投稿者が(おそらく野球観戦をするなかで)同じ経験をしていると推察される。また、「名前付けたいってくらい」という過剰な量や程度を表す副助詞「くらい(位)」からは、それほど打ちたい気持ちが高まっていることがうかがえる。つまり、「名前付けたいってくらいあるあるな事」とは、野球観戦をするうちに自分もボールを打ちたくなる気持ちの追体験を述べたものであり、「名前をつける」意図はないとみられる。

(7)は、Aの「名前をつけたい」というツイートに対し、Bが引用RT(リツイート)²をする形で応じている。Aは、一つの個人体験(「1時間作業しようと思って結局5時間ぐらい作業してしまいなんだこれとなる」)について「名前をつけたい」と述べている。これに対しBは、Aの前部要素「1時間作業しようと思って結局5時間ぐらい作業してしまいなんだこれとなる現象」の内容に共感を示したうえで、Bも同様の経験(「決めた予定の倍以上の時間がかかる」)と述べている。文末に感情記号(「www」)を付けていることから、同様の体験の習慣性(「いつも」「必ず」)に対し、ある種の自嘲的なユーモアを表している。なおBはAの「名前をつけたい」には反応していない。AのA「名前をつけたい」→B「わかります」という応答は、一見するとAとBの会話が成立していないように見える。しかし、「名前をつけたい」という願望表現に対し、「わかる」「あるある」と反応する形式の談話はTwitterによくみられる。これらの従来の用法と大きく異なる用法を見る限り、Twitterにみられる「Nに名前をつけたい」文は、語義通りの命名目的ではない、慣用的な表現になっていると考えられる。

本研究では、上記のような異なる観察結果をふまえて、Twitterにみられる慣用表現「名前をつけたい」の①命名を意図しない用例の使用状況、②名前と内容の認知意味論的關係、

² 他のユーザーのツイートを引用する形でコメントを付けてツイートする機能。元ツイートのユーザーだけでなく、引用したユーザーのフォロワーにもツイートが公開される。

の2点を主な研究目的とする。具体的には、対象となる表現を含む用例を Twitter で収集し、BCCWJ コーパスとの比較を通じて形式的な特徴と意味的な特徴の検討を行う。

なお、本稿では Twitter にみられる「名前をつけたい」文を、慣用表現の一種と定義する。土屋(2011)はことわざや慣用句などの定型表現を形式的に分類し、慣用句は動詞句が中心であることを明らかにしている。本稿の「名前をつけたい」は「AをB+たい」形式であり、同形式の慣用句には「猫の手も借りたい」「穴があったら入りたい」などがある。これらの形式的共通性や使用実態を鑑みて、慣用句の一様相として慣用表現と位置づける。慣用表現について、本稿では、中村(2007, p.87)の「世間で、あるいは、ある社会で、そのように表現することが慣用となっている用語や言い回しや決まり文句など。「慣用句」ほど厳密な定義で用いることのない、緩やかな一般語。」と定義する。慣用句については先行研究にさまざまな定義がみられるが、呉(2017)による詳細な分類を元に、①統語論・意味論的に連結した2語以上からなる表現とし、さらに中村(2007, pp.85-86)による②2語以上が連結したことばの意味の総和として論理的に導けないような意味を全体としてあらわす句(腕が鳴る、水に流すなど)とする。

2. 先行研究と本研究の位置づけ

かつて若者中心のメディアといわれたソーシャルメディアの利用層が年々拡大し、全年代に利用されるようになった。2019年時点で40代以降の利用者は10代から30代の利用者に迫る³。ソーシャルメディアの利用目的にも変化がみられ、官公庁の情報提供や企業・マスコミの公式アカウントによる情報発信の拡大⁴に伴って、従来の個人交流や日常的な投稿とあわせて情報収集や閲覧の目的で使われるようになったといわれる⁵。このような現況において、ソーシャルメディアにおけることばの様相も、世代的な特徴という枠組みにとらわれず、多様な側面から検討する必要があると考える。

Twitter やソーシャルメディアを対象とした先行研究はすでに多数あり、中でもソーシャルメディアのテキストデータをコーパスとして活用し、そこにみられる特徴的な語句や表現を取り上げたさまざまな研究例がある。五味・辰巳ほか(2011)では動詞「違う」が「チガクナイ」「チガクッタ」などの形容詞型活用をする現象について Twitter の用例を収集し、分析している。野田 (2014)は話し言葉と書き言葉の中間的な打ち言葉＝擬似独話が Twitter に多くみられるとし、情報受発信の様態の多様化に伴い、投稿文の読み手の範囲が曖昧になっている状況のなかで、独話の形をとりながらも読み手を意識する「擬似独話の時代」と位置付けている(野田 2014, pp.64-65)。神澤 (2019)はブログや SNS などに多くみられる特徴的な連体修飾節を取り上げている。泉(2019)は複合名詞の前項が文単位であ

³ 総務省「令和2年度 情報通信白書」図表 5-2-1-9「年齢階層別ソーシャルネットワーキングサービスの利用状況」によると、2019年のソーシャルメディア利用者率は10代から30代が80%以上と最も多いが、40代から50代は70%以上、60代も50%以上である。70代、80代は40%以上で2018年からの伸び率が最も大きい。

<https://tinyurl.com/35b7erwb> (閲覧：2022年7月20日)

⁴ 総務省「平成30年通信利用動向調査」3 ソーシャルネットワーキングサービス等の利用動向(PDF) <https://tinyurl.com/bddpvvdq> (閲覧：2022年7月20日)

⁵ マイボイスコム株式会社「第12回 Twitterに関する調査」プレスリリース 2021年1月14日 <https://tinyurl.com/2uebcxa5>(閲覧：2022年7月14日)

り、語の内部に包摂した形式がブログや SNS などに多くみられると指摘し、その形式的な特徴を分析している。Twitter にみられる特有の語句用法については岡田(2013)、宇野(2015)、林・松浦(2019)らが Twitter やインターネット特有の新しい用法や若者ことば、ネットスラングを分析している。

本稿で取り上げる Twitter における慣用表現の先行研究例では、岡田(2014)が「爪痕を残す」を対象に、従来にはない新しい用法を分析した。しかし、Twitter で初出が確認された慣用表現の研究例は管見の限り見当たらない。また、本研究の対象になる「名前をつけたい」は、個人ユーザーの投稿のみならず企業公式アカウントの PR 文や、中高年著名人の投稿からも用例が確認されている。また、本稿で取り上げる「名前をつけたい」文の検討対象にユーザー個々の体験や感情を述べた前部要素があり、その表現は多様である。このことから、本稿では Twitter にみられることばの様相を若者ことばに限定することなく、メディアの特徴的な表現として調査・分析を進めていきたい。

3. 研究方法と集計結果

Twitter API を利用して、「(～に)名前をつけたい」を含むツイートを表記や時期を指定して用例を収集し、分類項目やアノテーションを付与して分析を行った。検索対象としたのは、「名前をつけたい」、「名前つけたい」、「名前がつけたい」、「名前を付けたい」、「名前付けたい」、「名前が付けたい」の、漢字/かな表記、助詞の異なる 6 種類である。以降、本稿では表記を区別せず「(～に)名前をつけたい」と記す。用例には元の表記をそのまま記載する。

用例収集の対象期間は、Twitter サービス開始日である 2006 年 3 月 21 日を起点とし、①2006 年 3 月 21 日から 2010 年 12 月 31 日、②2016 年 3 月 1 日から 3 月 31 日、③2022 年 3 月 1 日から 3 月 31 日の 3 期間とした。用例数が非常に多く、2006 年から 2022 年までの全期間にわたる用例を網羅的に収集することは困難だったが、当該表現の出現状況を通時変化から観察したいと考えたため、期間を 3 期間に分割して 6 年間隔で一定数のサンプルを抽出することにした。この期間内における bot ツイート⁶とリプライツイート⁷を除く全ツイートを収集し、引用 RT は元のツイートを 1 件と数えた。

この結果、最も古い用例として 2007 年 5 月 31 日の(8)が確認された。

(8) 元ネタより先にボーガスニュース⁸を讀んじやったときの変なトリップ感に名前をつけたい

(2007 年 5 月 31 日)

Twitter では自分のツイートやアカウントを削除することができるため、過去のツイートが残っていない場合がある。そのため、これ以前にも使用例があった可能性が考えられるが、本調査時において確認する限りでは、(8)が Twitter サービス開始以降「名前をつけたい」用例が確認された最初の例である。そこで、以後の集計値を(8)の投稿日 2007 年 5 月 31 日以降と記載する。これをふまえて、各期間における 6 表記の「名前をつけたい」を合わせた用例数は 21406 件になった。

⁶ Twitter の機能を利用して機械によってツイートするアカウント。

⁷ 元ツイートに対する返信ツイート

⁸ さまざまな時事情報をパロディ化した記事を掲載するサイト(<http://bogusne.ws/>)。2004 年 1 月開設、最終更新日は 2013 年 5 月 16 日(閲覧日：2022 年 7 月 30 日)。

<1>2007年5月31日～2010年12月31日	6712件
<2>2016年3月1日～2016年3月31日	5068件
<3>2022年3月1日～2022年3月31日	9626件
計	21406件

対象3期間のうち<2>と<3>を1ヶ月間、<1>のみを4年以上の期間に設定した。<1>で初出から2010年末までの用例をすべて収集したのは、残り2期間との用例数のバランスを取るためと、初期出現内容を見ることで用例の質的な分析を行うねらいからである。期間<1>のツイートは6712件で日平均では3.84件、<2>と<3>の日平均は237件で用例数に差がある。この要因には、2010年以前時点でこの表現がまだあまり使われていなかったことが考えられるが、Twitter日本語ユーザー数や日本語のツイート数が相対的に少なかったことも要因としてみられる。また、Twitterの投稿者自身がツイートやアカウントを削除できる機能によって古いログが残っていない可能性も考えられる。このため通時的量的な単純比較はできないが、「名前をつけたい」文が2007年頃からTwitterで使われはじめ、2022年8月現在も使用され続けていることは確認できた。⁹

4. 考察

4.1 本当に「名前をつけたい」のか

全21406件の用例を観察したところ、従来型からTwitterの特徴的な型までさまざまなパターンがみられた。それらを①従来型②非命名抽象型③命名抽象型④非命名具体型の4分類に区分した。区分別の特徴と代表的な用例を以下に挙げる。

【①従来型】従来の動植物、モノなどを対象とした命名目的文

(9) 将来結婚して子供ができれば好きなアニメキャラの名前を付けたい、という気持ち分かるかな？

(2016年3月7日)

【②抽象非命名型】前部要素の後続部が抽象名詞の非命名文

(10) はじめきいたときはなんとも思わなかった曲なのに間をおいて「あれめっちゃ良い曲やん！」てなる現象に名前をつけたい。

私あるある。

(2016年3月5日)

(9)と(10)は目的・用法の異なる対照的な用例であり、(9)は、1. (3)～(5)のBCCWJ、(10)は(1)(2)のTwitterの特徴と一致している。(9)は従来の命名目的の文で、「子供」という具体的な名詞を対象とし、対象の属性や個性などを表す連体修飾語を伴う。これに対して、(10)

⁹ 過去のデータ収集についてTwitter APIを利用した手動収集を行ったほか、「名前をつけたい」ツイートのリアルタイム出現状況をあわせて観察する目的で、2022年7月20日から8月10日にかけてGoogleスプレッドシートのアドオン「Tweet Archiver」<https://tinyurl.com/2kedybw3>を利用してツイート収集を行った。この結果、「名前をつけたい」の非命名型ツイートはリツイートを含め1日あたり200～400件程度みられた。

は抽象概念的な名詞(「現象」)を対象に、必須要素の二格+願望「～たい」が結びついたパターンであり、(6)と同様に、体験の再現性や反復性を示す「あるある」が共起語に用いられていることから、命名の意図がないとみられる。(9)のような動植物、モノなど具体的な対象への命名目的文を①従来型とし、(10)のような抽象名詞を対象とした命名を意図しない文を②抽象非命名型と規定する。

次に、上記①②に当てはまらない用例を確認した。

【③抽象命名型】 前部要素の後続部が抽象名詞の命名目的文

- (11) あ、今頃になって今日のノルマやってないことに気付いた。まさかの三日坊主にさえ届かないこの状況。私はこの状態に「一日坊主」という名前をつけたいと思う
(2010年8月29日)

(11)の例は、抽象概念的な名詞を対象(「状態」)とし、必須要素の二格と願望を表す「～たい」のパターンである点は(10)と同様である。しかし、二格に続き「一日坊主」という具体的な名称候補を挙げている点が異なる。前部要素の後続部が従来のではない抽象名詞でありながら「～という名前をつけたい」と名称候補を挙げるパターンを命名目的とみなし、③抽象命名型と分類する。この分類には、「急募」「緩募」¹⁰「なにかいい名前ありませんか」などの命名を意図した共起語が含まれる場合が多い。

次に挙げる(12)は「風邪」という具体的な名詞を対象としており、一見すると従来の命名型に見える。

【④具体非命名型】 個人的・具体的な動植物やモノなどを対象とした非命名文

- (12) 何故か空いているのにあまり人が通らない改札に名前つけたい。
(2022年3月1日)

(12)の「改札」は具象名詞であり、従来型と同様に名づけの対象になりうる。しかし、前部の修飾節「何故か空いているのにあまり人が通らない」は、投稿者の個人的な体験である。また、名称候補を挙げておらず、命名意図がみられない。前部要素の後続部が具象的でありながら内容が限定的で、命名意図がみられないパターンを【④具体非命名型】に分類する。区分①～④のうち、【①従来型】と【③命名抽象型】を<命名系>、【②抽象非命名型】と【④具体非命名型】を<非命名系>とした。そのうえで、全件を①～④のいずれかに分類し、命名系、非命名系それぞれの集計を行ったところ、表2の結果になった。

¹⁰ 急ぎではないが情報やメンバーを求める意味。参考:デジタル大辞泉「ネット上で【緩募】という言葉を見かけたことはありませんか?」2017年7月20日(2022年8月15日閲覧)tinyurl.com/msbdp3wj

表 2 Twitter「名前をつけたい」出現数と命名/非命名分類結果

分類 対象期間	命名系		計	非命名系		計	合計
	① 従来	② 抽象		③ 抽 象	④ 具体		
	<1>20070531- 20101231	2859	51	2910	3562	240	3802
<2>20160301- 20160331	961	64	1025	3949	94	4043	5068
<3>20220301- 20220331	1825	43	1868	7617	141	7758	9626
計	5645	158	5803	15128	475	15603	21406

表 2 をみると、全期間において非命名型の用例が従来型を上回り、出現時期が新しくなるほど比率が高まっていることがわかる。時期<1>の非命名型使用率は 54.6%、時期<3>は 80.59%である。全期間の平均では非命名型が占める割合は 72.89%になった(図 1)¹¹。

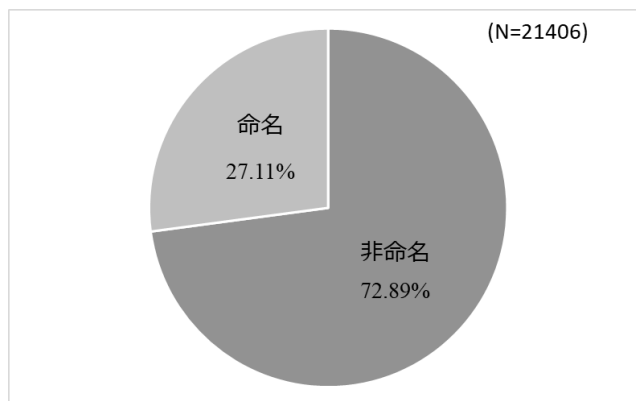


図 1 Twitter「名前をつけたい」ツイートの非命名／命名区分比率

このことから、Twitter の「名前をつけたい」文は、語義とは異なり命名目的でない用例数が従来型を上回る特徴がみられる。これらの非命名型の特徴①②について、認知命名論の観点から詳しくみていきたい。

森岡・山口(1985)は、日常語の名における概念に大小の違いがあるとし、その上位には抽象度が高い概念があるとする(図 2 の<抽象 1>、<抽象 2>)。この体系のうち、具体的な属性や個性を表す命名(ネーミング)の対象になるのは<一次>から<四次>までであり、それぞれの名にはその属性を象徴し抽象化・一般化に寄与する「表示性」と、属性内における個性を表し具体化・差別化に寄与する「表現性」を内包する。森(2019)は、森岡・山口(1985)が示したこれらの「表示性」と「表現性」を細分化し、上下の階層による属性概念に加えて、それぞれの属性や個性の区分を具体的に示した。図 2 のように、従来例では名称の表示性や

¹¹ 命名系と非命名系の各期間の合計値で χ^2 二乗検定にかけたところ、時期<1>の命名系が多く非命名系が少なく、時期<2>および<3>の命名系が少なく非命名系が多い、との結果を得た($\chi^2 = 1307.144$, $p < .01$, Cramer's $V = 0.247$)。期間の長さが異なるなど条件面に問題はあがるが、相対的な比率としての結果なので参考までに記す。

表現性は下位に属するほど細分化される。一方で<抽象1>、<抽象2>の「気象」や「自然」は、名というよりも「語」であり、命名対象にならない。

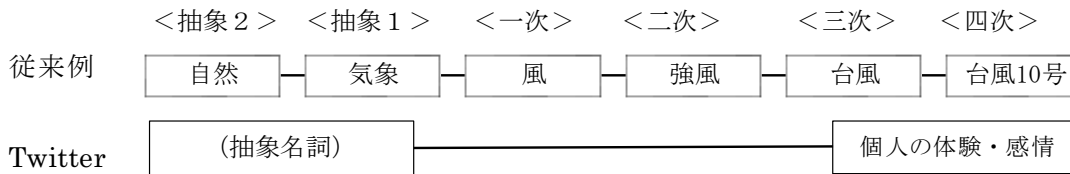


図2 名の概念体系とTwitter「名前をつけたい」文構造の比較
(森岡・山口 1985, p.32 の図を基に筆者作成)

図2の下段は、Twitterにみられる「名前をつけたい」の②非命名型である。この類型に後続する現象、状況などの抽象名詞は<抽象1>、<抽象2>に該当する。それら抽象名詞は属性の区分性が乏しく命名対象にならないため、前部要素である個の体験・感情という極端な表現性のみを示す構図になる。つまり、Twitterの用例にみられる個人的・限定的な体験や感情に名前をつけたいという行為は、属性のないものを属性化して表示性を与えたいといった相矛盾する関係になり、その不自然な構図からも従来とは異なる意味用法であることがわかる。

このようにTwitterの「Nに名前をつけたい」文には、形式的な特徴および意味的な特徴から、従来型とは明らかに異なる命名目的のない用例がみられ、その数が従来型を上回っていることがわかった。その特異性は、森岡・山口(1985)および森(2019)の命名体系と認知意味論的な表示性・表現性の観点からみても明らかである。また、用例の返信や引用RTを観察するなかで、①～④の複合的な例もみられた。その例を、本節の最後に紹介したい。

(13) A: 年をとるにつれて、読了した本におけるノンフィクションの比率が高くなっていく現象に名前をつけたい。

B: マジメ化 ww 僕もその現象です

A2: しかし年々真面目になっているかと言われると、笑ってごまかすしか(笑)。

(2010年1月2日)

(13)は、Aの投稿に対しBが引用RTの形で返信している。Aの元ツイートを見る限り、命名目的があるかどうかは判別できない。BはAの投稿を受けて一応の名称案を挙げている(「マジメ化」)が、それと同時に「僕もその現象です」と共感を示している。これに対してA2は、「マジメ化」というBの名称案を元に話題を展開するが、真剣に名称案を募集する意図はないことが感情記号(笑)から推察される。Bの反応にも笑いを示す感情記号「ww」があり、Aに命名意図がないことをBが理解したうえでユーモアを示しているとみることができる。名称と共感の両方を軽い形で示すBを受けて、A2は他愛のない雑談の形でゆるやかに反応している。

(13)のように、「名前をつけたい」の元投稿に対し、フォロワーからの返信や引用RTに名称候補が挙がる例はみられたが、「あるある」「私も同じ」と共感する例はそれ以上に多く、さらに別の体験談を挙げるなどして話題を発展させる態様もみられた。つまり、Twitterに

おける「名前をつけたい」は命名か否かという単純な分類ではなく、その表現自体が慣用表現として浸透し、話題の発展・深化のきっかけになっているとみることができる。さらに、次のような「名前をつけたい」表現そのものに言及するツイートが一定数¹²みられることから、この表現の定着化がうかがえる。

(14) あるあるネタを見つけたら「～という現象に名前をつけたい」と言ってしまう現象に名前をつけたい

(2009年12月23日)

4.2 慣用表現としての意味・用法

4.1 で示した通り、Twitter にみられる「名前をつけたい」文の多くが従来型とは異なる抽象概念を対象とし、命名を目的としないことがわかった。では、語義通りの意味でない慣用表現であるとするならば、どのような場面において用いられるのだろうか。その慣用表現としての意味や語義を探るため、前部要素、つまり「名前をつけたい」とする出来事や感情を内容の面から見ていく必要があると考えた。これまでに述べた通り、前部要素の内容は後続の抽象名詞に対して、個人的な体験や感情が中心である。それらの体験や感情を非命名型のうち無作為抽出した500例を分析したところ、下記A～Dの主に4種類に分類された。

【非命名型の内容区分と定義・用例】 (カッコはN=500の出現率)

<A 反復・再現> 以前に体験した出来事が再来する。何度も同様の経験をする (37.10%)

(15) 埃を払ったのに数分後にはまた積もってるこの現象に名前をつけたいね

(2016年3月15日)

<B 落胆> 出来事や体験に対し落胆・失望する(24.19%)

(16) 会社のボールペンを胸ポケットに入れたまま帰宅してガッカリする現象に名前付けたい

(2016年3月2日)

<C 予想外> 突然起きる予想外の出来事や限定的な出来事(16.13%)

(17) 弊社の名前がニュースに踊る日にかぎって、フォロワー数がキリ番になる現象に名前をつけたい。

(2016年2月25日 シャープ公式)

<D 興奮・驚き> 出来事や体験に気持ちが昂る(14.52%)

(18) 独り言をつぶやき続けていると自分のテンションがだんだん上がってくる現象に名前をつけたいと思う土曜日の昼下がり

(2010年9月4日)

4分類のうち37.10%と最も多い「反復・再現」は、過去に投稿者が体験した出来事や感情が日常生活において再現した際に、その反復性自体を話題にしている。また、「落胆」「予想外」「興奮・驚き」は、日常の規則的なルーティンから逸脱した意外性のある状況について述べている。いずれの分類にも共通しているのは、日常生活において、ふだんとは少し異なる場面に遭遇したときに、投稿者がその出来事への反復性または意外性を独白しつつも読み手に共感を求めたいという意図ではないだろうか。

¹² 「名前をつけたい」という表現自体に言及したツイートは21406中45件、全体の0.39%みられた。

次に、非命名型「Nに名前をつけたい」のN部分の名詞について、出現例と使用度を集計した。非命名型用例15603件のうち100回以上の出現例がみられた名詞は、順に現象、気持ち、感情、法則、症状、状態、衝動の7種類であり(付録表6)、このうち「現象」が全体の78.58%と大半を占めた。時期別では、<1>2007年5月31日～2010年12月31日は60.70%、<2>2016年3月1日～3月31日は81.82%、<3>2022年3月1日～3月31日には85.66%と、その比率が年々高まっていることが明らかになった。また、「現象」の使用度率が高まるに伴って、現象以外の名詞や名詞句の使用種類が少なくなっている傾向もみられた(図3)。

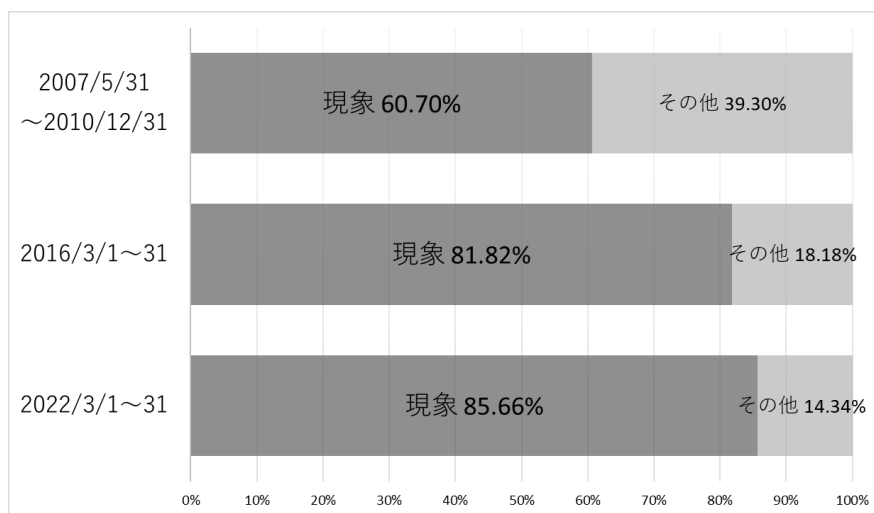


図3 非命名型「名前をつけたい」前部要素の後続部「現象」使用度推移

なぜ前部要素の後続部が「現象」一語に集約され、他の名詞が使われなくなっているのか。その要因を探るため、Twitter「名前をつけたい」における「現象」の意味用法を、従来型用例との比較において調べた。まず、前部要素の被修飾部を内容別に分類するため、a～fの6項目を設定した。

- a 記憶や感情 主体の記憶や感情を表す内容
- b 個人体験 主体の実体験を叙述する内容
- c 対人・対社会 主体の見聞や経験に基づく人や社会に関する内容
- d 生理的欲求 睡眠、食事など生理的欲求にかかわる内容
- e 娯楽・機器 ゲームやアイドルなどの内容とPCや端末等の内容
- f 科学・社会 科学的な事象、社会・歴史的事実などの客観的な内容、学術用語

分類用サンプルは、Twitterは「名前をつけたい」のうち前部要素の後続部が「現象」である非命名型用例700件を無作為抽出した。BCCWJは「現象」のみで検索すると5,754件と大量になったため、「現象+ニ格」の用例に絞り込んで398件の用例をサンプルにした。Twitter、BCCWJ各サンプルの意味内容を分類したところ、表3の結果になった。

表 3 Twitter「Nに名前をつけたい」の「現象」とBCCWJ「現象」の意味別分類

	Twitter	BCCWJ
記憶や感情	29	0
個人体験	357	17
対人・対社会経験	102	25
生理的欲求	82	0
娯楽・機器	130	0
科学・社会	0	356
計	700	398

Twitterの「Nに名前をつけたい」とBCCWJにおける「現象」の前部要素には対照的な違いがみられる。Twitterでは(18)にみられるような投稿者個人の具体的かつ主観的な出来事に基づく内容のみであり、科学・社会に言及した用例は1例もみられなかった。一方、BCCWJの用例では、科学・社会の一般的かつ客観的な事実を述べた内容が398件中356件と全体の89%以上を占めた。個人的な出来事や体験を述べた用例は17例みられるが、感情や記憶にかかわる用例は1例もなかった(付録表7)。

【Twitter「個人体験」の例】

(19) 忙しい時期が終わってふっと気を抜いた瞬間に風邪をひいてしまう現象に名前をつけたい

(2022年3月18日 ベンザブロック公式)

【BCCWJ「科学・社会」の例】

(20) いまでこそ、呼吸は穏やかな燃焼に外ならないことが知られているが、両者の共通性が明らかにされたのは、十八世紀も末になってからであった。そもそも燃焼という身近な現象について、曲がりなりにも化学らしい議論が行われるようになったのが、十七世紀後半からなのである。

LB14_00052 小山慶太『神さまはサイコロ遊びをしたか—宇宙論の歴史』

(19)の前部要素が投稿者個人の体験であるのに対し、(20)は「燃焼」という科学的事象を示す。「現象」と類似した抽象名詞「状態」について、新屋(2008)は語彙的、統語的、語用論的な態様を、年代の異なるコーパスの用法比較を通じて通時的に比較・考察している。このなかで新屋は、「状態」が明治以降の近代語形成過程において定着していった時代的背景も反映され「名詞本来の機能である格成分よりも述語、特に文末名詞として機能するものが増えている」と指摘する(新屋2008, pp.303)。現象についても同様に、西洋からの翻訳語を漢語的に表した術語であり、語義のあいまいさを内包していたがゆえに用法が広がったものと思われる¹³。

TwitterとBCCWJの「現象」にみられる最大の違いは複合語の出現性である。BCCWJの「二格現象」398例中197例が複合語だったのに対し、Twitter「名前をつけたい」文の現象が複合語として使われている例は、全11888例中わずか7種類13例だった(表4)。

¹³ 「現象」は西周が明治時代に哲学用語 *phenomenon* の訳語として用い、それが「哲学字彙」に採用されてから一般化したとされる。出典:日本国語大辞典, JapanKnowledge, <https://japanknowledge-com.libproxy.ouj.ac.jp>, (参照 2022年8月20日)

表 4 Twitter「名前をつけたい」の「現象」と BCCWJ「現象+に格」の複合語比較

Twitter “Nに名前をつけたい” の現象を後項とする複合語	怪奇現象、怪現象、逆転現象、謎現象、不思議現象、侵略現象、不思議現象 全7種類 13例
BCCWJ「現象+に」の現象を 後項とする複合語	学術、専門用語(フェーン現象、毛細管現象、光行差現象など) 一般用語(怪奇現象、超常現象、文化現象など) 造語(よろず社会現象、アタマでっかち現象、不登校現象など) 全 112種類 197例

複合語の内容と出現頻度についても、新屋(2008)による「状態」の例が当てはまる。BCCWJの「現象」は複合語の後項要素となる率が高いのと同時に、「〇〇現象」という前項要素がきわめて多様化している。これに対して Twitter の「名前をつけたい」文では、文の前部要素である投稿者の個人体験・感情が「現象」に連結し、単純語としての「現象」が個々の体験や感情の表現対象にされるため、複合語として意味を付加する必要がない。また、この単純語「現象」には、「意味的な抽象性、上位性のゆえに、実質的な意味を持つものであればそのような語句とも結合することが可能」(新屋 2010, p.316)になっていることから、後項とする語は必ずしも「現象」でなくてもよい。実際に、Twitter「名前をつけたい」文の「現象」は、その類概念にあまり厳格な意味がなく、前部要素の意味内容を見ても、多くの名詞が他に置き換え可能であることがわかる。

(21) a. めっちゃ元気な曲聴いてると逆に泣きたくなる現象に名前つけたい
(2016/3/21)

(21) b. めっちゃ元気な曲聴いてると逆に泣きたくなる気持ちに名前つけたい

(21)をみると、前部要素の後続部を現象から気持ちに置き換えても違和感なく文が成立することがわかる。他にも感情、法則、状態など、出現例のあるほとんどの名詞に置き換え可能である。また、次のような例もある。

(22) a. 家に帰ったらこたつがつけっぱなしで、へこむと同時にホカホカ幸せという微妙な気持ちの現象に名前を付けたい。
(2010/12/28)

(22) b. 家に帰ったらこたつがつけっぱなしで、へこむと同時にホカホカ幸せという微妙な気持ちに名前を付けたい。

(22)は、(へこむと同時にホカホカ幸せという)「微妙な気持ち」が主題である。ここに、後続となる「現象」を付ける必然性はあまりない。これらの例から、「現象」や「気持ち」にはそれ自体の意味が薄く、前部要素である連体修飾節を含む主語節の文末に付けられた機能語化、補語標識化しているとみることができる。そのような機能において「Nに名前をつけたい」のN名詞が淘汰されて「現象」の一語に収束され、「(前部要素)+現象に名前をつけたい」という慣用表現が定着したと考えられる。

以上の分析から、Twitter「名前をつけたい」文には前部要素を後項とする名詞に「現象」が多く、そのほとんどは単独語として用いられるが、意味は厳密ではなく、他にも置き換えということがわかった。また、これまでにみえた態様から、Twitterにみられる非命名型「名前をつけたい」の慣用表現には、①個人の日常的な体験や感情のうち、反復や再現、予想外の事態、興奮、躊躇などの非日常的な出来事を述べる際に使われる②命名を目的としない独話形式であり、ユーザー同士の共感反応がみられる場合が多いことがわかった。これらのうち「名前をつけたい」の前部要素と後続名詞の文節には、泉(2019)が文が包摂した複合

名詞形式(「母さん助けて詐欺」「早く帰れオーラ」など)に挙げた特徴の、①後部要素が抽象的な概念を表す名詞であることと、②前部要素の文に心情や印象を述べた内容が多いことが共通している。これらは、神澤 (2019)が取り上げた用例(「すまふおの容量足りない問題」「親父の大河ドラマの音デカすぎ問題」など Twitter にみられる特徴的な連体修飾節)にも共通する特徴である。神澤 (2019)は特徴のひとつに、連体修飾節でピッチの下降が生じないことを挙げている。本稿で取り上げた文節(「エコバッグをうまく取り出せなくて、レジ前でなんともいえない空気が流れる現象」)は直後に「に名前をつけたい」が続く文節であり、それ自体が独立した名詞句として用いられる可能性は低く、この点は泉(2019)、神澤 (2019)の事例と異なる。しかし、対象となる名詞が抽象概念的であることや、前部要素に投稿者個人の出来事や感情を自由に綴る点が共通していることは確かであり、ソーシャルメディアにおける文体の特徴的な様相のひとつとみることができるだろう。

4.3 変異形の出現

「名前をつけたい」文は通常は必須要素の二格「～に」+願望の「たい」型で用いられるが、これらの出現状況を通時的に追うなかで、いくつかの変形例を確認した。表 5 は、Twitter 「名前をつけたい」変形例をまとめたものである。それぞれ末尾変化やモダリティ(「～つけようぜ」「～つけてほしい」「なんか名前はあるのかな」)、程度表現(「名前つけてもいいくらい」「名前つけてもいいレベル」、あいさつ表現(「名前を付けたい今日この頃」「おはようございます(連結)」)などの出現と一定数の使用例が確認された。また、2013 年以降は「名前をつけたい人生だった」という新しい慣用表現が一定数確認されている。「願望～たい」+「人生だった」形式は Twitter で 2012 年頃から使われている一種の慣用表現であり¹⁴、これの前部要素に「名前をつけたい」が連結した形と思われる。

表 5 「名前をつけたい」変異形の初出日と用例

初出年月日	用例(下線太字は変異形)
2006年3月21日	(Twitter サービス開始)
2007年5月31日	元ネタより先にボーガスニュースを読んじゃったときの変なトリップ感に <u>名前をつけたい</u>
2007年8月17日	1人お題:「『発言が反映されない』って書くと反映される法則に <u>名前つけようぜ!!</u> 」
2007年10月14日	市場規模が10倍になると価格が50%になるってのに、 <u>なんか名前はあるのかな</u>
2008年4月17日	絵描きのひとが上達してデッサン力が高まった結果、絵が怖くなって可愛さが失われる現象、に <u>だれか名前をつけてほしい</u> 。がっかりした経験多数
2008年10月17日	肩をポンポンして振り向くホッペに指を、ってのに <u>名前を付けたい気分</u>
2009年6月15日	そろそろ誤って洗顔フォームで歯磨きする行為に <u>名前を付けたい今日この頃</u> 。

¹⁴ 初出例は「家でのおんびり漫画読んでたい人生だった(2012年9月25日)」(閲覧:2022年6月4日)。意味用法はいくつかみられるが、主に投稿者の願望(漫画読んでたい)が実際にはかなわなかった現状を嘆く例が多い。

2009年10月24日	「徹夜しようかなあ、と思った途端、お腹が空く」現象は <u>名前つけてもいいくらいに</u> 一般的なことだろうか
2009年11月10日	そろそろかばんの中でソフトパックのタバコが漏れて折れる現象に <u>名前つけようか</u> ・・・
2009年12月21日	画像を使わずに CSS だけでナントカしたいがために、マーク付けが冗長になって得意げにしてるのを、 <u>なんか名前をつけて呼びたい</u> 、という「 <u>名前つけたい症候群</u> 」に冒されてしまいました。
2010年8月19日	トイカに慣れ過ぎて、地下鉄で手に定期入れ持ってるのに直前まで切符出すの忘れて、改札にタッチしようとしてないのに気付いて慌て切符を出す行為にそろそろ <u>名前をつけてもいいレベル</u>
2010年9月28日	<u>おはようございます！朝からミスドでコーヒー飲みながら脳内整理タイム♪最近はこの時間がすごく大切☆脳内整理って可愛くないから、なんか名前をつけたいなあ。このいとおいしい時間に。そんなわけで、今日も1日がんばりましょう☆</u>
2011年10月23日	書店で購入した文庫本と同じ本が以前注文した amazon の箱から出てくる現象に、 <u>名前をつけたいようなつけないような</u> 。
2013年10月25日	伸びたら椅子が後ろに倒れそうになってビビる症状にかっこいい <u>名前を付けたい人生だった</u>
2016年1月2日	Naked Romance を踊ってもらうとめちゃくちゃかわいい現象に <u>名前を付けたくありませんか？</u>

土屋(2011)は定型表現の拡張用法について、「コミュニティのメンバーとしての同一性」や、「背景的知識」の共有を「相互に認証し合う作業を伴う」とし、「定型表現を用いた創造的言語使用による修辞性は文脈への適応のみならず、発話者と聞き手間のこの相互認証を求める談話マネジメント的機能を有している」と述べる。表5の各変異形は、検索結果を見る限り、いずれも「名前をつけたい」以降に出現し、各例の使用率がそれぞれ数十件から数百件程度と、非命名型「名前をつけたい」用例の15603件よりも少ないことから、「名前をつけたい」を原型とした変形例と位置付ける。石田(2015)は、コーパスを用いた慣用表現の「変異形」研究について、従来式の内省や少数の用例のみの研究手法では把握しきれなかった変異形やその対応関係が体系的に抽出できる可能性が高いとする。本稿では詳しい分析には至らなかったが、今後は個々の意味や形態、「名前をつけたい」との対応関係、Twitterやソーシャルメディアにみられる慣用表現と個々の意味用法や修辞性について詳しく調べていきたい。

5. まとめと今後の課題

本稿では、Twitterにみられる慣用表現「名前をつけたい」の従来的な用法との違いを明らかにすることを目的とし、国語研現代日本語書き言葉均衡(BCCWJ)コーパス用例の比較調査を通じて、その特徴を検討した。この結果、①Twitterの「名前をつけたい」表現の8割近くが非命名目的であり、名前と内容の認知意味論的關係からも従来型との違いがみられること②後続の抽象名詞と対照的に前部要素部分の大半は個人体験や心情に基づく内容であること③後続の抽象名詞に通時的変化がみられ、共起語「現象」とあわせて慣用表現化していることが明らかになった。

これまでにみてきた態様から、Twitter にみられる慣用表現「名前をつけたい」には下記のような語義が考えられる。

- ①個人的な出来事や感情を中心に反復や再現、予想外の事態、驚きや興奮などの非日常体験に直面し、
- ②一語では書き表せないようなもどかしい状況やその心情におかれたときに用いる。
- ③大半が命名を目的としない独話形式であるが共感を求める願望形を伴い、ユーザー同士のテキストコミュニケーションにおいて共感の反応が多く、命名反応や話題の発展もある。
- ④通常は必須要素の二格「～に」+願望の「たい」型で用いられるが、変則表現も多数出現している。

一言で表せない混沌とした状況や感情をことばで表そうとする行為はインターネット以前から存在し、人の言語表現活動そのものにかかわるともいえる。個人的・限定的な出来事や感情を書き綴る行為についても、古くから日記や手紙の形式で存在した。Twitter では、その行為がソーシャルメディアという不特定多数を対象とした非対面式のゆるいつながりのなかで広がったもので、そこに「名前をつけたい」という共通表現が加わったことで慣用表現化したと考えられるだろう。

北村・佐々木・河井(2016)によると、Twitter の個人ユーザーがツイートする内容は主に自分に起きた出来事や気持ちを書くことであり、その意図の多くは「ツイートすること自体が楽しい」、「ツイートすることで気分が楽になる」という独話目的である。ただし、共起語として「共感して欲しい」「伝えたい」などの願望表現がみられることから、読み手に対し共感や承認を求める意識も表れているという(北村・佐々木・河井 2016, pp.69-109)。独話の形をとりながらも「欲しい」「したい」という願望を表現することは、「押しつけがましくならず自分の心情や本音を人に知らせる」擬似独話形式と共通し(野田 2014, p.65)、Twitter における「名前をつけたい」文はそれらの特徴に合致する。個人的な出来事や気持ちを独話形式で綴りながらも「名前をつけたい」という願望に乗せることで、読み手からのゆるい反応を求めているといえる。

このように独話と願望を織り交ぜた形式は、相手への配慮を示すと同時にゆるい共感を求める近接的・共感的なポライトネスのストラテジーであり、Brown and Levinson (1987) の枠組みでいえば「ポジティブ・ポライトネス」のストラテジーとみることができる。Twitter のユーザー同士の反応や談話には、このようなメディア特有のポライトネス・ストラテジーが働いているといえるだろう。今後は、本稿で収集した「名前をつけたい」の用例を元に、ユーザー同士で用いられる配慮やポライトネス・ストラテジーを詳細に分析していきたいと考える。具体的には、本稿で取り上げた独話式のツイートのみならず、ツイートのリプライ数やリツイート数、「お気に入り」登録数などの反応状況を類型化し、個々のリプライやリツイートの談話を質的に分析することで、Twitter の慣用表現をポライトネス・ストラテジーの観点から考察する方向性が考えられる。

また、本稿では Twitter における「名前をつけたい」の変異形出現状況から元の表現の広がりを確認したが、この表現は、Twitter での使用にとどまらず、2014 年頃から複数のラジオ番組におけるレギュラー企画としても定番化している。このことから「名前をつけたい」をはじめとする Twitter やソーシャルメディアで用いられることばが、他メディアへと使用範囲を拡大する可能性も考えられる¹⁵。このようなメディア別の出現状況をふまえ、Twitter

¹⁵ tinyurl.com/4ab369fm, <https://www.joqr.co.jp/ag/article/46581/>, <https://www.tatsuyanomura.com/radio> 等のラジオ番組企画を確認した(2022年8月20日閲覧)。企画内容はリスナーの体験談をもとにパーソナリティが名称を付けるもの、リスナーから名称を募集するものまでさまざまである。

にみられる他の慣用表現事例を対象とした形式的、認知意味論的な研究、ポライトネス・ストラテジーの観点からリプライやリツイートも含めた談話分析等を進めることで、ソーシャルメディアにおけることばの様相の一端をとらえていくことを今後の課題としたい。

付録

表 6 非命名型「Nに名前をつけたい」対象 N 名詞の使用度

度数 順位	前部 要素 の後 続名 詞	使用度 数	使用度率	前部要素の例
1	現象	11888	78.58%	新幹線とか、ちょっと長距離移動する時に、崎陽軒のシュウマイを食べたくなる (2016/3/28)
2	気持ち	305	2.02%	PC 落として寝ようとしたときに、更新して再起動ボタンを押してしまったときのこのなんとも言えない (2016/3/12)
3	感情	205	1.36%	弾き語り路上ライブを見かけた時に沸き起こるあの (2010/11/24)
4	法則	177	1.17%	昨日さんざん悩んだバグが翌日あっさりと見つかる (2016/3/4)
5	症状	174	1.15%	旅行に出るたびになんか忘れ物している気持ちになるこの (2010/11/10)
6	状態	116	0.77%	フレームの小さい眼鏡で下を見ると半分見えてない変な状態になるけど、この (2010/12/27)
7	衝動	101	0.67%	何の用事もないのに冷蔵庫を開けてしまう (2010/12/15)

表 7 Twitter と BCCWJ 「現象」の意味別分類

分類	Twitter, BCCWJ の用例 []は出現数
① 記憶や感情	<p>【Twitter】 [29] 深夜に過去の思い出に浸って戻りたくなる現象に名前付けたい (2022年3月5日)</p> <p>【BCCWJ】 [0]</p>
② 個人体験	<p>【Twitter】 [357] 片付けしてて懐かしいものを見つけると片付けが進まずそれに夢中になる現象に名前を付けたい (2016年3月17日)</p> <p>【BCCWJ】 [17] ドイツだけのことかと思いきや、先日パリ市内を走っていたらまったく同じことだったし、海峡を渡ったロンドンやバーミンガムでも、同様の現象に出くわした。パッサとヘンデル記念の年をひめやかに送り、突如モーツァルト・ブームにわきたった(日本ほどではないが)ヨーロッパ中がまたしてもマーラーに戻ってきた。 LBg2_00043 小塩節『ザルツブルクの小径 音楽と食彩の旅』</p>

<p>③ 対人、対社会</p>	<p>【Twitter】 [102] ピュアなクリエイターの人がネットにはびこる陰謀説に傾きがちになる現象に名前を付けたい。 (2008年9月17日 津田大介)</p> <p>【BCCWJ】 [25] 自分さえよければ後で利用する人のことはまったく考えずに、利用した判例集を元の棚に返さずに帰るという現象に、これぞアメリカの公共心の退化だ、今にアメリカはつぶれるぞ、と一人悪態をつきながら目指す判例集を探し回ったものだった。 LBa3_00005 阿川尚之『アメリカン・ロイヤーの誕生』</p>
<p>④ 生理的欲求</p>	<p>【Twitter】 [82] 眠い眠いって布団に入ったら目が覚める現象に名前をつけたい (2022年6月21日)</p> <p>【BCCWJ】 [0]</p>
<p>⑤ 娯楽・機器</p>	<p>【Twitter】 [91]特定のポケモンを対策していると全く出会わないのに対策を切った途端に出会ってしまう現象に名前をつけたい。 (2016年3月18日)</p> <p>【BCCWJ】 [0]</p>
<p>⑥ 科学、社会</p>	<p>【Twitter】 [0] 【BCCWJ】 [356] 臨死体験の解釈として、それを現実体験と解釈する立場と、脳内で起きた現象にすぎないとする立場と、基本的に二つの立場があるということとは前に述べた。そして、普通、脳の話を持ち出す人は、だいたい後者なのだが、アトウォーターさんはちがうのである。 LBi1_00015 立花 隆『臨死体験』</p>

謝 辞

本稿の執筆に際して、指導教官である放送大学 滝浦真人教授にはきめ細やかなご指導、ご助言を賜りました。心より感謝申し上げます。

文 献

- 森雄一(2019)「命名論と認知言語学」辻 幸夫・楠見 孝・菅井 三実他『認知言語学大事典』4C.9, pp.609-615,朝倉書店
- 森岡健二,山口仲美(1985)『命名の言語学』東海大学出版会
- 野田 春美(2014)「疑似独話と読み手意識」石黒圭・橋本行洋(編)『話し言葉と書き言葉の接点』, pp.57-74,ひつじ書房
- 泉大輔(2019)「文を包摂する名詞の形式的な特徴に関する考察」国立国語研究所編『言語資源活用ワークショップ 2019 発表論文集』
- 神澤克徳(2019)「メディアなどで見られる拡張的な連体修飾表現の分布：主観性と間主観性の観点から」日本語用論学会第 21 回大会発表論文集第 14 号, pp.25-32

- 五味伸之・辰巳暢・新田優喜(2011)「Twitter を利用した言語形態の変化についての研究」『福井工業高等専門学校研究紀要 人文・社会科学』第 45 号
- 岡田祥平(2013)「Twitter を利用した新語・流行語研究の可能性—アイドルグループ「Sexy Zone」の略語を例に一」新潟大学教育学部研究紀要第 6 巻第 1 号, pp.49-74
- 岡田祥平(2014)「「爪痕を残す」の「新用法」について」『日本語学会 2014 年度秋季大会予稿集』 pp.185-192
- 宇野和(2015)「Twitter における『新しいミ形』」『国文』123, pp.106-94. 掲載誌 国文 / お茶の水女子大学国語国文学会 編, pp.106-94
- 林智昭・松浦光(2019)「インターネットスラングにおける意味変化—新規表現「耐え」を中心に—」日本語用論学会第 21 回大会発表論文集第 14 号, pp.89-96
- 石田プリシラ(2015)「コーパスと辞書に見られる慣用句の「変異形」」砂川有里子・他『汎用的日本語学習辞書開発データベース構築とその基盤形成のための研究』研究成果報告書 筑波大学
- 呉 琳(2017)「コーパスに基づく日本語慣用句の研究」北海道大学文学研究科博士論文
- 新屋映子(2008)「総合雑誌に見る名詞「状態」の用法」『日本語の名詞指向性の研究』ひつじ書房
- 新屋映子(2010)「類義語「状況」「状態」の統語的分析」『日本語の名詞指向性の研究』ひつじ書房
- 土屋智行(2011)「言語の創造性の基盤としての定型表現 —慣用句およびことわざの拡張用法の調査—」『認知科学』18(2), pp.370-374
- 中村明(2007)『日本語の文体・レトリック辞典』東京堂出版
- 滝浦真人(2008)『ポライトネス入門』研究社
- Penelope Brown and Stephen C. Levinson (1987) “Politeness: Some Universals In Language Usage”, Cambridge University Press
- 北村智・佐々木裕一・河井大介(2016)『ツイッターの心理学：情報環境と利用者行動』誠信書房

関連 URL

- Twitter <https://twitter.com/> (閲覧：2021 年 7 月 17 日～2022 年 8 月 15 日)
- Twitter ヘルプセンター「高度な検索の使い方」<https://help.twitter.com/ja/using-twitter/twitter-advanced-search> (2022 年 7 月 31 日閲覧)
- 国立国語研究所『現代書き言葉均衡コーパス(BCCWJ)』中納言 2.6.0 データバージョン 2021.03. <https://chunagon.ninjal.ac.jp>
- 国立国語研究所『分類語彙表—増補改訂版データベース』<https://clrd.ninjal.ac.jp/goihyo.html> (閲覧：2022 年 8 月 15 日)
- 総務省情報通信政策研究所「令和 2 年度 情報通信メディアの利用時間と情報行動に関する調査」2021 年 8 月 25 日掲載 https://www.soumu.go.jp/iicp/research/results/media_usage-time.html (閲覧：2022 年 7 月 20 日)

書き言葉・話し言葉コーパスデータに基づく 高頻度漢語動名詞の品詞性の再考：日本語教育の視点から

陳迪 (神戸大学大学院国際文化学研究科)

Reconsideration of Part-of-Speech of Sino-Japanese Verbal Nouns Based on Written and Spoken Corpus Data : A Perspective on Japanese Language Education

Di CHEN (Graduate School of Intercultural Studies, Kobe University)

要旨

漢語動名詞は、単独で名詞として使用されるほか(例：勉強は大事だ)、「する」を付加することでサ変動詞の一部として使用されることもある(例：勉強する)。このとき、個々の漢語動名詞が主としていずれの品詞で多く使われるのかは必ずしも明確でない。そこで、本研究では、「現代日本語書き言葉均衡コーパス」と「日本語日常会話コーパス」のデータを用い、動詞性判定テストと名詞性判定テストによって、重要漢語動名詞 200 種の品詞用法別頻度調査を行った。その結果、漢語動名詞は、動詞的動詞・名詞的動詞・動詞的名詞・名詞的名詞の 4 種に区分でき、それぞれ、異なる統語特性と意味特性を有することが示された。得られた知見は日本語教育にも応用が可能であると思われる。

1. はじめに

日本語には漢語動名詞 (Verbal Noun, VN) が多く存在するが、その品詞特性は極めて複雑で、単語間に名詞性や動詞性の濃淡の差が存在する。漢語動名詞は、一般に、名詞と動詞の両方の性質を備えた語群とされる(小林, 2004)。名詞であれば、格助詞を介在させて「漢語 VN+を/が+V」と言え、動詞であれば「漢語 VN+する」と言えるはずだが、実際には、語によって、許容度の差がある。以下の例を見てみよう(すべて筆者による作例)。

- (1) *能力の【発揮】をしている [名詞的用法]
- (2) 能力を【発揮】している [動詞的用法]
- (3) 心の【準備】が整いました [名詞的用法]
- (4) *心を【準備】しました [動詞的用法]
- (5) 日本語の【学習】を頑張ってください [名詞的用法]
- (6) 日本語を【学習】したことがありますか [動詞的用法]

このことは、語によって、名詞的用法と動詞的用法の両方を許容する場合、名詞的用法のみを許容する場合、動詞的用法のみを許容する場合とがあることを示す。

日本語教育の観点から見た場合、中国人学習者は漢語の知識を有するが、それでも、漢語動名詞の習得は難しい。以下は「日本語学習者作文コーパス」(林・李・宮岡・柴崎・趙, 2012)における中国人学習者の作文実例である。なお、修正例はコーパス開発者が付与したものである。

- (7) 自分の会話力 (→力) やイヤリング (→ヒアリング) が【上達】になるだ (→上達す

る)と思う。(CG038)

(8) いま、ぼくは日本語 N1 試験を【準備】しています(→の準備をしています)。(CG050)

学習者が、漢語動名詞の使用上の問題に直面した場合、教科書や辞書に頼ることとなるが、個々の漢語動名詞の品詞特性については、十分な情報が示されていないのが現状である。こうしたそこで、本研究では、主要な漢語動名詞について、その品詞特性を調査し、学習者向けの指導モデルを考えることを目指したい。

2. 先行研究における漢語動名詞の品詞性の判定

漢語動名詞の品詞用法については様々な研究が行われてきた。とくに重要な研究として、小林(2004)がある。同研究は、漢語動名詞が動詞的性質と名詞的性質の両方を有することを指摘する。時間表現内で使用可能で(「目的語+[を/に]+VN+[中/後](に)」-資料を入力中に、群馬方言を調査後)、尊敬語を付加できる(「お/ご~になる」「お/ご~ください」-ご訪問になる、ご検討ください)点では動詞的である。また、格助詞を付けて単独で主語や目的語として使用可能で(勉強が~、勉強を~)、用途を表す「~用」を付加できる(料理用の酒、受験用の本)点では名詞的である。なお、漢語動名詞の名詞用法は、一般の名詞と完全には同じではない。第1に、普通の名詞は「~する」の形で使えないが、動名詞は「する」をつけてサ変として使用できる(*騎手する/騎乗する)。第2に、普通の名詞には「人(ジン)」が付加されるが、動名詞には「人(ニン)」が付加される(外国人(ジン)/管理人(ニン))。

このほか、多くの研究が、近代日本語や現代日本語における漢語動名詞の使用実態を調査してきた。永澤(2010)は、『太陽コーパス』や新聞用例を使い、二字漢語の品詞用法を調査した。間淵(2015)は、『太陽コーパス』と『現代日本語書き言葉均衡コーパス(BCCWJ)』を用い、サ変用法を調査した。間淵(2016)は、『明六雑誌コーパス』、『国民之友コーパス』、『太陽コーパス』、『近代女性雑誌コーパス』、BCCWJ 出版サブコーパスを使い、従前の分析を拡張した。田辺・中條・船戸(2012)は、独自に作成した日英対訳の新聞コーパスを用い、漢語動名詞のコロケーションを調査した。田辺・中條(2014)は、同じコーパスとBCCWJの図書館・書籍サブコーパスを比較して、漢語動名詞を調査した。李(2014)は、BCCWJのジャンル別データを用いて、主な漢語動名詞の出現パターンや言語特性を調査した。

各研究は、漢語動名詞の品詞性の議論にあたり、小林(2004)の動詞・名詞の判定基準を基本にしながら、それぞれの研究目的に応じて、判定基準の工夫を試みている。以下は、各研究で採用された判定基準をまとめたものである。

表1 先行研究における動詞用法・名詞用法の判定基準

先行研究	動詞用法判定基準	名詞用法判定基準
小林(2004)	~する・~中/後(に)・ご~になる/くださる	~〈格助詞〉・~用
永澤(2010)	~す・~する	~〈格助詞〉・~〈係助詞〉・~〈副助詞〉
間淵(2015)	~する・~出来る	
間淵(2016)	~為る・~出来る・~致す・~奉る・~仕る・~遊ばす・~下さる・~なさる・	~〈格助詞〉・~〈係助詞〉・~〈副助詞〉・~ \emptyset

	～申す	
田辺他(2012)	高頻度コロケーションでサ変動詞用法が多いもの	高頻度コロケーションで名詞用法が多いもの
田辺他(2014)	～する・～いたす・～なさる・～できる	動詞用法以外のすべての品詞用法
李(2014)	～する	～をする

こうした先行研究の判定基準にはいくつかの課題もある。1 点目は、各研究が採用した動詞用法の判定基準にずれが見られることである。たとえば、「～する」(活用形含む)を動詞用法の判定基準とすることはすべての研究に共通しているが、間淵(2015)は「～出来る」を、間淵(2016)と田辺・中條(2014)は「～出来る」と「～する」の敬語形をそれぞれ基準に追加している。2 点目は、名詞性を直接的に判断した研究が極めて少ないことである。たとえば、田辺・中條(2014)では、動詞用法以外のすべての品詞用法を名詞用法と仮認定している。これに対し、間淵(2016)は、サ変率を根拠として動詞性の強さと名詞性の強さ(あるいは他の品詞性)を議論する工夫を行っているが、名詞性を直接に判定したわけではない。3 点目は、言語学で提唱されてきた各種の動詞用法・名詞用法の判定基準を網羅的に調査した研究が少ないことである。今回の先行研究の概観の限りにおいては、小林(2004)で報告された判定基準をすべて用いて分類を行った研究はなかった。

3. リサーチデザインと手法

3.1 研究目的と RQ

本研究では、重要漢語動名詞を対象として、その品詞性を調査・分類し、分類群ごとの統語的・意味的特性の調査を行うことで、学習者用の指導モデル開発の手掛かりを得ることを目指す。品詞性の判定にあたっては、先行研究で言及された品詞判定基準を網羅的に取り込んだ新しい動詞性判定テストと名詞性判定テストを作り、動詞性・名詞性をそれぞれ独立して評価していく。上記をふまえ、以下の3つのリサーチクエスチョン(RQ)を設定した。

- RQ1 動詞性・名詞性が高い漢語動名詞はなにか。
- RQ2 動詞性・名詞性を統語的に評価した場合、漢語動名詞はどのように区分され、各区分にどのような統語的特性があるか。
- RQ3 各区分にどのような意味特性があるか。

3.2 データ

本研究では、「現代日本語書き言葉均衡コーパス」(Balanced Corpus of Contemporary Written Japanese, BCCWJ)と「日本語日常会話コーパス」(Corpus of Everyday Japanese Conversation, CEJC)を利用する。BCCWJは、国立国語研究所が中心となって開発した現代日本語の書き言葉の大規模コーパスである。BCCWJは「新聞」「雑誌」「書籍」「白書」「教科書」など、13種の大ジャンルに区分されており、ジャンルによってはさらに詳細な下位区分がある。CEJCは、国立国語研究所が開発した話し言葉コーパスで、40名の調査協力者の3か月の自然会話200時間分の会話データを収録している。CEJCの会話データは、形式別に「雑談」「用談・相談」「会議・会合」「授業・レッスン」の4ジャンルに区分されている。本研究で使用しているのは2022年の本公開版である。両コーパスとも、国立国語研究所が開発・運営するコーパス検索インタフェースである「中納言」上でアクセスする。

2つのコーパスには、様々なジャンルデータが含まれるが、本研究では、BCCWJの13ジャンルのうち、体言止めの多用など、特殊な品詞使用がなされている可能性のある「韻文」を除く12ジャンルと、CEJCの4ジャンルを組み合わせた16種の日本語ジャンルデータを資料として、重要漢語動名詞200種の動詞的用法と名詞的用法の頻度を調査する。

3.3 手法

3.3.1 調査対象語

調査する対象語の範囲は、BCCWJの65変種およびCEJCの場面別7変種（「室内」「職場」「自宅」「施設」「交通機関」「学校」「屋外」）をあわせた72変種から均衡的に抽出された重要漢語動名詞200種である。

表2 重要漢語動名詞（重要度順）

利用	実施	使用	対応	紹介	確認	増加	発生	表示	作成	説明	参加	推進
発表	存在	理解	表現	期待	活用	注意	設定	設置	開催	減少	変化	購入
提供	比較	検討	結婚	選択	指定	展開	判断	入力	用意	検索	質問	確保
工夫	実現	評価	計算	開始	変更	発見	登録	注目	拡大	規定	導入	決定
相談	完成	指摘	勉強	構成	提出	移動	改善	保存	整備	解決	開発	整理
考慮	低下	安定	促進	維持	把握	登場	上昇	予想	成立	終了	反映	成功
形成	発展	販売	記載	報告	採用	記録	重視	発揮	意識	安心	認識	処理
撮影	配慮	練習	意味	削除	実行	生活	該当	集中	主張	達成	心配	無視
活躍	交換	連絡	観察	成長	分析	充実	強化	対処	継続	掲載	設立	適用
強調	経験	活動	失敗	支援	想像	努力	発達	提案	追加	発行	取得	学習
独立	調整	電話	更新	増大	卒業	加入	推移	吸収	代表	貢献	回答	我慢
調査	発売	一致	接続	派遣	希望	分解	保護	出席	分類	指導	反応	納得
悪化	解消	運転	印刷	進行	相当	準備	経過	違反	出品	反対	記入	感染
提示	証明	落札	到着	防止	放送	否定	演奏	改正	請求	応援	入院	出発
苦勞	訪問	通過	死亡	収集	表明	注文	策定	支給	制定	就職	批判	感謝
施行	所属	支配	共有	合格								

なお、これは、先に陳（2022）で提案した200種を改訂したものとなっている。以下に、両研究における抽出過程の違いを記載する。

表3 重要漢語動名詞の抽出方法

	陳（2022）	今回
対象ジャンル	BCCWJの65変種+CEJCの7変種=72変種	
CEJCデータ	モニター公開版（50時間）	本公開版（200時間）
抽出条件	語種：「漢」 品詞大分類：「名詞」 品詞小分類：「普通名詞-サ変可能」 後方共起：「語彙素読み」が「スル」	語種：「漢」 品詞大分類：「名詞」 品詞小分類：設定なし 後方共起：「語彙素読み」が「スル」
抽出過程	(1) 該当漢語数が10未満の4変	(1) 該当漢語数が20未満の1変

種を除く (2) 68 変種の上位 10 語、延べ 680 語を抽出 (3) 異なり 214 語を抽出 (4) 重要度 (平均頻度×レンジ比率) で並べ替え、上位 200 語を決定	種を除く (2) 71 変種の上位 20 語、延べ 1420 語を抽出 (3) 異なり 363 語を抽出 (4) 重要度 (平均頻度×レンジ比率) で並べ替え、上位 200 語を決定
---	--

2 種の処理の違いについて概略する。まず、CEJC データについては、陳 (2022) の作業段階ではモニター公開版を使っていたが、今回の処理では本公開版を使用した。

次に、抽出条件についてであるが、陳 (2022) では「サ変可能」とあるもののみを対象とした。しかし、検証の結果、形態素解析で「サ変可能」と分類されていない語の中にも、サ変化している例が確認された (例: 「それはお母さんを安心させるためでもあるんだよ」 (PB59_00148)、「こうした雰囲気的な情感に浸ることが哲学することなのではない。」 (PM21_00626))。そこで、今回の処理では、漢語動名詞をより網羅的に抽出するため、名詞の小分類をあえて設定せずに単語抽出を行った。なお、範囲を広げた新しい処理では、漢語動名詞でありうるものを広く拾える一方、研究対象である二字漢語動名詞以外のものまで拾ってしまうという問題もある。たとえば、副詞と「する」が結びついた「(騒音を) 結構する」、助数詞とする「(二) 時間する (と)」、二字漢語以外を含む「(一) 段落する」など、である。そこで、抽出されたすべての連鎖について、再度、「Web 茶まめ」で形態素解析を行い、漢語成分の品詞を確認した上で、対象以外ものは除去する処理を加えた。

以上の処理により、陳 (2022) で提案した重要漢語 200 語のうち、約 62 語が入れ替わった。具体的には、「出土」や「定植」などがはずれ、代わりに「安心」や「調査」などが加わった。

3.3.2 動詞性・名詞性判定テスト

漢語動名詞の動詞性・名詞性は、何らかのテストによって決定される必要がある。先行研究では、様々なテストが考案されてきたわけであるが、本研究では、過去の研究で採用された項目を幅広く取り込み、以下のテスト内容を決定した。両テストはそれぞれ 3 つの判定基準を内包する。

表 4 本研究の動詞性・名詞性判定テストの項目

	動詞性判定テスト	名詞性判定テスト
基準 1	「スル」付加 (～する (敬語形含む)、～出来る)	格助詞付加 (格助詞「が・を・に…」の後続)
基準 2	時間表現付加 (～中、～後)	係助詞付加 (係助詞「は・も…」の後続)
基準 3	尊敬表現化 (ご～になる/くださる)	「用」付加 (～用)

なお、本研究では、過去の研究で扱われた基準を広く取り込むことを目指したが、動詞性判定テストにおける「奉る・仕る・遊ばす・申す」後続基準と、名詞性判定テストにおける副助詞後続基準は採用しなかった。この点について説明したい。

まず、間淵 (2016) が動詞性判定で採用した「奉る・仕る・遊ばす・申す」後続基準につ

いてであるが、BCCWJを検索したところ、「漢語 VN+奉る」が4例、「漢語 VN+仕る」が63例、「漢語 VN+遊ばす」が27例、「漢語 VN+申す」が28例、確認されたものの、多くは現代日本語としては違和感の残るものであった。下記は用例の一部である。

- (9) 一万石ほど、ご返納【奉る】つもりはないか（出版・書籍、PB29_00071、『黒船擾乱』、笹沢左保（著））
- (10) 慶喜が、薩土越宇の四藩をふくむ諸藩の意見を「取捨【仕り】候は、私の任に御座候」と言い放ったことである（図書館・書籍、LBs2_00093、『徳川慶喜』、家近良樹（著））
- (11) このたびは上公御隠居【遊ばされ】、誠に恐れ入った次第。（図書館・書籍、LBf2_00008、『幕末の水戸藩』、山川菊栄（著））
- (12) ほんとうの親でもない私が、これほど一生懸命にお世話【申して】、あなたを大切な人と思っているのが（特定目的・ベストセラー、OB2X_00146、『円地文子の源氏物語』、紫式部（著）/円地文子（訳））

(9)～(12)は、いずれも、歴史小説や、歴史をテーマにした文献内での用例であり、狭義の現代日本語の用例としては特殊なものだと思われる。この基準を加えた場合、他の基準のウェイトが低減すると思われることから、今回は採用しないこととした。

次に、永澤（2010）および間淵（2016）が名詞性判定で採用した、（漢語が補語となる場合の）副助詞後続基準についてであるが、同じく、BCCWJで検索したところ、以下のように、副助詞が名詞ではなく動詞に後続する事例がいくつか確認された。

- (13) 形を工夫し【さえ】すれば、肉まんだって、ほら、このとおり。（出版・書籍、PB55_00045）
- (14) 民訴法百六十九条（新百三条）などを準用する【のみ】で（図書館・書籍、LBm3_00048）
- (15) ただ、それがこの世のどこかに存在する【だけ】でいい。（図書館・書籍、LBr7_00065）

この基準を加えた場合、上記のような用例が分析対象から外れてしまうため、これも今回は採用しないこととした。

3.3.3 手順

まず、RQ1では、本研究で決定した判定基準に基づき、「中納言」上で、以下の検索条件で検索を実施した。

表5 各品詞判定基準の検索条件

	動詞性判定テスト	名詞性判定テスト
基準1	「スル」付加（～する（敬語形含む）、～出来る）→ キー：語彙素が漢語 VN、後方共起：語彙素が“ <u>為る 出来る 致す 為さる</u> ”	格助詞付加（格助詞「が・を・に…」の後続）→ キー：語彙素が漢語 VN、後方共起：品詞が“助詞-格助詞”
基準2	時間表現付加（～中、～後）→ キー：語彙素が漢語 VN、後方共起：語彙素が“ <u>中 後</u> ”	係助詞付加（係助詞「は・も…」の後続）→ キー：語彙素が漢語 VN、後方共起：品詞が“助詞-係助詞”

基準3	尊敬表現化(ご~になる/くださる) → キー: 語彙素が漢語 VN、前方共起: 語彙素が“御”、(2)後方共起 1:語彙素が“に”、後方共起 2: 語彙素が“成る” (2)後方共起:“下さる”	「用」付加(~用)→ キー: 語彙素が漢語 VN、後方共起: 語彙素が“用”
-----	--	--

順位化の方法として、まず、第1基準に関して、個々の語について、全用例数と、そのうち、当該基準を満たす用例数を求め、基準合致率を計算する。そして、基準合致率の高いものから、対象語全体を並べ替え、順位を決定する。その後、第2基準、第3基準について、同じ作業を繰り返す。最後に、個々の語について、3つの基準に基づく合致率順位の和を計算する。そして、順位和の低いものから全体を並べ替え、最終的な順位を決定する。ここでは、「掲載」を例に説明したい。この語は、動詞性判定テストにおいて、第1基準の合致率は65.99%で全200語中の順位は16位であった。また、第2基準の合致率は1.4%で順位は28位であった。第3基準の合致率は0%で順位は65位であった(※0%の語は多数あるため、すべて65位となる)。これらの順位和は16+28+68=109となる。最後に、順位和で200語すべてを並べ替え、最終順位(この場合は6位となる)が決定される。

RQ2では、重要漢語動名詞の品詞用法を区分するため、200種の漢語動名詞を第1アイテム、各合致率を第2アイテムとする対応分析を実施する。対応分析とは、頻度表における行・列の関係を組み替え、頻度表に含まれる情報を少数の成分(次元)にまとめることで、行・列を整理する多変量解析手法である。対応分析によって、データが視覚的に散布図上に布置することができ、直観的に関係を解釈することが可能になる。また、対応分析において、変数とケースの区別は行わず、列方向に入るデータの集合を第1アイテム、行方向に入るデータを第2アイテムと呼ぶ(石川他、2010)。

RQ3では、各区分に含まれる各漢語動名詞の意味特性を『分類語彙表』(増補改訂版)に基づき調査する。『分類語彙表』では、体・用・相の各類の中に、大きな意味的まとまりとして「抽象的關係」、「人間活動の主体」、「人間活動・精神および行為」、「生産物及び用具」、「自然物及び自然現象」といった5つの部門を設け、各部門の内部は「事柄」、「類」、「存在」などのような中項目に分かれた。さらに同じ中項目に属する単語はいくつかのグループ(段落)に細分された。本研究では、表6に示したように、部門を大分類、段落を小分類として採択する。

表6 『分類語彙表』の枠組み

大分類	小分類
抽象的關係	理由・目的・証拠、異同・類似...
人間活動の主体	男女、夫婦、相手・仲間...
人間活動・精神および行為	感覚、表情・態度、学習・習慣・記憶...
生産物及び用具	産物、資材・ごみ、飾り...
自然物及び自然現象	光、気象、物質の変化...

ここで注意すべきは、漢語動名詞が多義である場合の扱いである。本研究では、『デジタル大辞泉』(以下、『大辞泉』)(ジャパンナレッジ版、2022年7月に検索)における当該語の第1語義の意味分類を選ぶことにする。たとえば、「反応」には、(1)ある働きかけに応じて起こる物事の動き、(2)

生体が刺激に応じて活動を起こすこと、(3)物質の相互作用によって別の物質を生じること、の3つの語義がある。また、『分類語彙表』における「反応」の意味特性は、「抽象的關係－作用・変化」、「人間活動・精神および行為－感覚」、「自然物及び自然現象－生理」、「自然物及び自然現象－物質の変化」の4種である。この場合、第1語義に合致するものとして、「抽象的關係－作用・変化」を選ぶ。

4. 結果と考察

4.1 RQ1 動詞性・名詞性が高い漢語動名詞

2種の品詞判定テストを行った結果、各品詞判定基準の合致率の順位和の上位20位の漢語動名詞、すなわち動詞性が高い漢語動名詞と名詞性が高い漢語動名詞がそれぞれ抽出された。

表7 動詞性が高い漢語動名詞

動詞性 順位	高動詞 性漢語	合致率			順位			
		スル	時間	尊敬	スル	時間	尊敬	順位和
1	確認	0.6568	0.0059	0.0128	18	54	8	80
2	活躍	0.5704	0.0434	0.0008	40	7	33	80
3	到着	0.5729	0.0265	0.0002	39	10	46	95
4	検討	0.5058	0.0245	0.0014	66	11	24	101
5	記入	0.6177	0.0035	0.0066	27	71	10	108
6	掲載	0.6599	0.0140	0.0000	16	28	65	109
7	開催	0.5979	0.0236	0.0000	32	13	65	110
8	提出	0.6438	0.0031	0.0020	20	73	19	112
9	成立	0.6038	0.0198	0.0000	30	17	65	112
10	通過	0.6550	0.0133	0.0000	19	30	65	114
11	購入	0.5423	0.0077	0.0013	46	46	25	117
12	出席	0.5892	0.0027	0.0061	35	76	11	122
13	入力	0.5952	0.0064	0.0003	33	52	42	127
14	使用	0.5060	0.0105	0.0012	65	38	26	129
15	実施	0.6000	0.0108	0.0000	31	34	65	130
16	参加	0.4452	0.0090	0.0132	90	43	7	140
17	考慮	0.8336	0.0035	0.0000	6	70	65	141
18	完成	0.5223	0.0150	0.0000	56	26	65	147
19	表明	0.6893	0.0014	0.0005	13	94	41	148
20	採用	0.5888	0.0047	0.0002	36	60	54	150

表8 名詞性が高い漢語動名詞

名詞性 順位	高名詞 性漢語	合致率			順位			
		格助詞	係助詞	「用」	格助詞	係助詞	「用」	順位和
1	練習	0.3707	0.0522	0.0099	19	23	2	44
2	記録	0.3693	0.0615	0.0038	21	16	12	49
3	活動	0.5506	0.0581	0.0011	3	21	29	53
4	生活	0.4924	0.0436	0.0059	7	41	6	54
5	連絡	0.3645	0.0548	0.0025	24	22	15	61
6	電話	0.4965	0.0366	0.0019	6	52	18	76
7	回答	0.3781	0.0748	0.0004	16	8	55	79
8	準備	0.3979	0.0604	0.0003	14	18	64	96
9	勉強	0.3369	0.0604	0.0006	32	19	47	98
10	規定	0.5557	0.0889	0.0000	2	3	94	99
11	分析	0.3287	0.0434	0.0014	35	42	24	101
12	放送	0.2858	0.0441	0.0055	53	39	9	101
13	反応	0.4590	0.1122	0.0000	10	1	94	105
14	努力	0.5427	0.0704	0.0000	4	10	94	108
15	意味	0.5695	0.0657	0.0000	1	14	94	109
16	変化	0.4871	0.0801	0.0000	9	7	94	110
17	交換	0.2849	0.0403	0.0032	54	45	13	112
18	意識	0.5160	0.0657	0.0000	5	13	94	112
19	評価	0.3772	0.0728	0.0001	17	9	88	114
20	経験	0.4916	0.0674	0.0000	8	12	94	114

では、本研究の枠組みによって、先行研究に対してどの程度異なる結果が得られたのであろうか。下記は、本研究で得られた上位5語をサンプル語として、それらの順位が、先行研究（田辺他、2011）の動詞性判定（サ変率のみに基づく判定）および名詞性判定（動詞比率から間接的に推定した名詞比率に基づく判定）基準あればどのようになっていたかを示した結果である。

表9 本研究と先行研究の動詞性・名詞性判定基準の比較

動詞順位の変動例			名詞順位の変動例		
語	本研究の順位	サ変率に基づく 5語の順位	語	本研究の順位	推定名詞比率に基 づく5語の順位
確認	1	18	練習	1	28
活躍	2	40	記録	2	12
到着	3	39	活動	3	3
検討	4	66	生活	4	1

記入	5	27	連絡	5	29
----	---	----	----	---	----

上記に示すように、従来の基準と今回の基準で、「活動」以外の個々の語の順位が大きく変動することが示された。動詞性判定については、間淵（2015）も指摘するように、サ変率だけで、動詞性（あるいは名詞性）の強さを計ることは問題が多いと言えるだろう。また、名詞性判定については、動詞性比率のみを直接に求め、名詞性比率はその残余とみなす間接推定法にも問題があると言えるだろう。

4.2 RQ2 漢語動名詞の分類と区分ごとの統語特性

200種の漢語動名詞を第1アイテム、6つの品詞用法判定基準の合致率を第2アイテムとする対応分析を実施した結果、以下のような散布図が得られた。次元1と次元2の寄与率はそれぞれ57.8%と25.2%となり、全体の約83%が説明できる。なお、時間表現付加は第3象限において他と離れた位置に布置されているため、別途記載している。

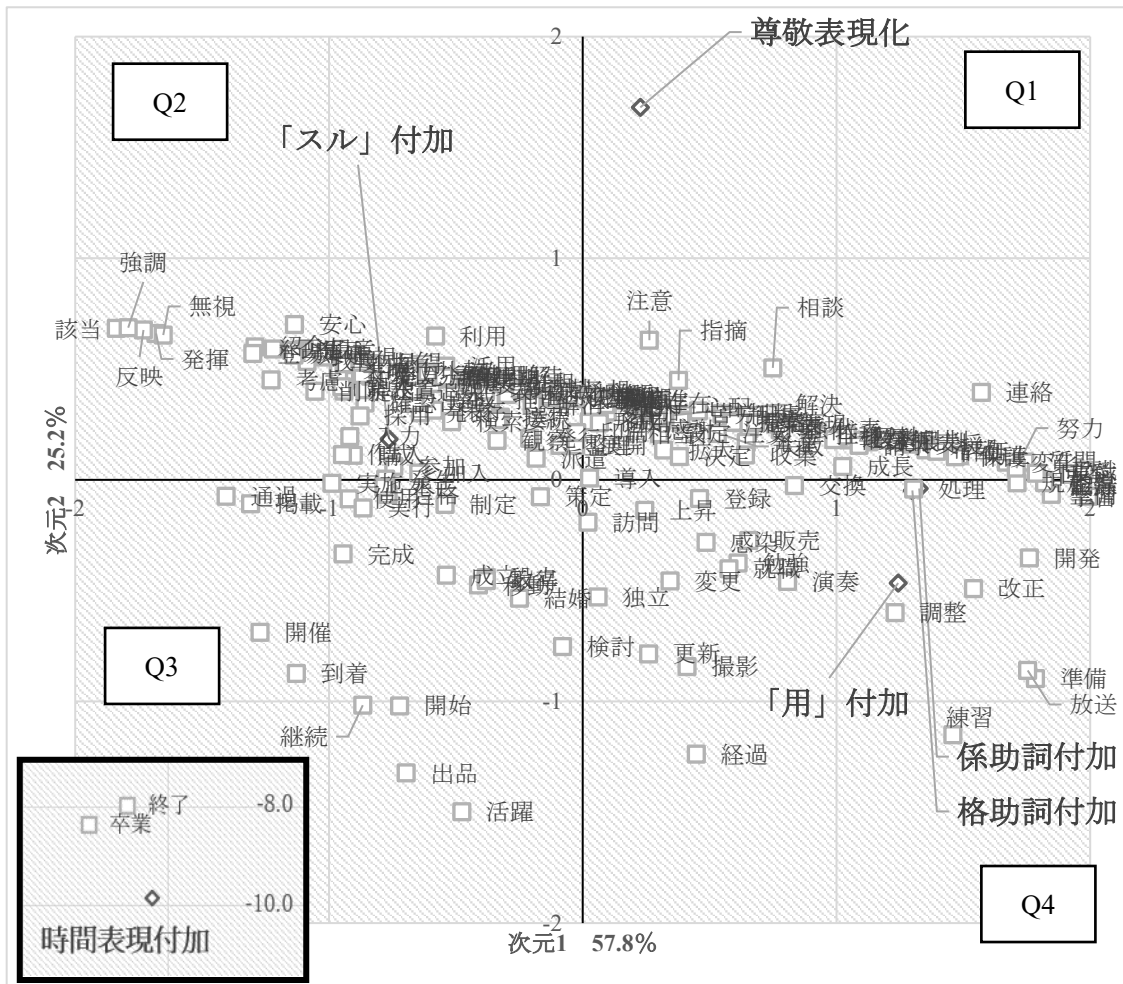


図1 対応分析の結果

全体は、2つの次元（軸）によって4つの象限にわかれる。判定基準を手掛かりに考えると、第1象限（Q1）は尊敬表現化と関連が深い。尊敬表現は一般には動詞標識とされるが、

「ご指摘になる」という用例に見られるように、漢語その自体を考えれば本質的には名詞的な語群であるため、動詞的性質を包含した名詞群（動詞的名詞）と言える。第2象限（Q2）は、「強調する」「無視する」など、「スル」付加と関連が深い。これはもっとも典型的な動詞性を持つ動詞群（動詞的動詞）と言えよう。第3象限（Q3）は、「進行中」「卒業後」など、時間表現付加と関連する。時間表現は一般には動詞標識とされるが、実際には、「進行」や「卒業」という内容は、話者が当該の行為の外の視点に立ち、行為を客体化して呼称しているもので、名詞的な動詞群（名詞的動詞）と言える。第4象限（Q4）は、「運転が」「準備は」「練習用」など、格助詞・係助詞・「用」付加と密接に関連しており、典型的な名詞群（名詞的名詞）と言える。これらを本研究では、新しい動名詞の区分として提案したい。

表10 動詞用法判定テストと名詞用法判定テストを統合した新しい品詞分類

品詞区分	象限	主な統語的特性	漢語動名詞
I 名詞的名詞	Q4	格助詞・係助詞付加、一部は「用」後続できる	生活、調査、電話、運転、経過、練習、準備、放送、撮影、更新、調整、独立、改正、演奏、変更、活動、就職、勉強、開発、感染、販売、訪問、上昇、登録、学習、整備、意味、処理、交換、記録、規定、指導（32語）
II 動詞的名詞	Q1	尊敬表現付加	注意、相談、指摘、連絡、確保、構成、対応、説明、悪化、分類、促進、支配、感謝、存在、増大、心配、安定、相当、苦勞、増加、判断、解決、配慮、失敗、形成、提案、認識、適用、充実、表現、代表、計算、強化、注文、設定、推移、反対、違反、報告、改善、工夫、批判、分析、発展、発達、支援、請求、希望、拡大、評価、収集、防止、決定、保護、質問、変化、成長、努力、意識、回答、反応、経験、導入（63語）
III 名詞的動詞	Q3	時間表現付加	卒業、終了、発売、入院、進行、施行、落札、活躍、出品、開始、継続、到着、検討、開催、結婚、移動、取得、設立、成立、完成、実行、制定、掲載、使用、策定、通過、合格、死亡、実施（29語）
IV 動詞的動詞	Q2	「スル」付加(敬語形・可能形含む)	安心、強調、該当、反映、發揮、無視、利用、紹介、否定、用意、提示、集中、登場、対処、重視、我慢、共有、納得、把握、活用、比較、一致、表明、吸収、記載、理解、分解、発見、期待、維持、注目、提供、考慮、実現、出席、証明、所属、支給、記入、貢献、減少、設置、予想、表示、提出、削除、追加、想像、主張、指定、応援、成功、達成、推進、低下、出発、確認、解消、保存、発表、採用、選択、接続、検索、印刷、入力、発行、観察、整理、展開、作成、購入、派遣、参加、加入、発生（76語）

以上の4つのタイプは、動詞性の強弱の点で、段階的な位置づけにある。4つのタイプの

関係性と、全 200 種に占める各々の比率をまとめると、以下のようになる。

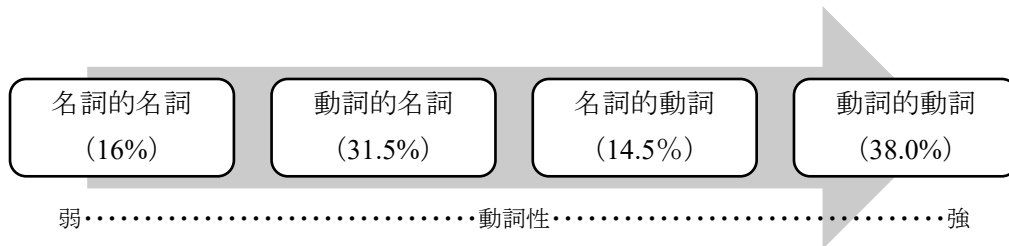


図 2 動詞性・名詞性の連続体モデル

上記より、動名詞が名詞と動詞にまたがる性質を有すること、いわゆる名詞と認定されるものは 2 割未満にとどまることが確認された。以下、BCCWJ と CEJC における具体的用例を見る。(16) ～ (17) は名詞的名詞、(18) ～ (19) は動詞的名詞、(20) ～ (21) は名詞的動詞、(22) ～ (23) は動詞的動詞の実例で、用例中、当該の漢語動名詞は【】記号で示す。

- (16) 私たちの日常生活でも、先にふきんや台所スポンジ、野菜類の細菌汚染について見たように、無数ともいえる細菌類に囲まれた【生活】をしています。(BCCWJ、図書館・書籍、LB14_00044)
- (17) と相澤さん座談会でイイッティ最初の大島の【調査】を挙げていたんですけど(CEJC、会議・会合、W001_001b)
- (18) お忘れ物なさいませぬようにご【注意】ください(CEJC、雑談、K012_007a)
- (19) この辺、何か郵政大臣は、文部省なり科学技術庁あたりと御【相談】になって、やるべき方向を見出すべきだと思うんですね。(BCCWJ、特定目的・国会会議録、OM21_00010)
- (20) その【卒業】後どうゆう活動されてるか(CEJC、会議・会合、T003_017)
- (21) ☆チケットは、市民会館・ぴあ・ローソン・クラモトレコードで【発売】中！(BCCWJ、特定目的・広報紙、OP77_00002)
- (22) 発生予防対策を【強調】したが、実際問題として発生を完全に予防することが困難である(BCCWJ、特定目的・白書、OW1X_00575)
- (23) ゴールデンウィークぐらいまではさ【安心】できない(CEJC、雑談、C002_008)

まず、(16) と (17) における「生活」と「調査」は、格助詞「を」をつけて、動詞「する」「挙げる」の目的語として機能しているので、もっとも典型的な名詞用法であると言える。次に、(18) と (19) における「注意」と「相談」は、いずれも尊敬表現に入ったことで、動詞的性質を備わっている名詞であると考えられる。(20) と (21) における「卒業」と「発売」は、時間表現である「中／後」が付加できることで、動詞性を有するが、「卒業」は「その」という連体修飾語で修飾することができることで、このような漢語動名詞は名詞性も備わっていると考えられる。(22) と (23) における「強調」と「安心」は、「する」あるいは可能形である「できる」をつけて述語として機能していることで、もっとも典型的動詞用法であると言える。

こうした用例からも明らかなように、同じ漢語動名詞であっても、その用法には多様性が

ある。品詞特性の観点から漢語動名詞を分類することで、そうした用法理解の手掛かりが得られるであろう。

4.3 RQ3 漢語動名詞の区分ごとの意味特性

前節で、漢語動名詞を品詞性の観点から4区分に分けた。では、各グループには意味的な違いはないのであろうか。この点を確認するため、『分類語彙表』に基づき、意味の大分類・小分類で整理したところ、以下の結果を得た。なお、大分類に関して、「人間」は「人間活動・精神および行為」を、「抽象」は「抽象的關係」を、「自然」は「自然物および自然現象」を示す。小分類については上位3種を選び、筆者の判断で、変化・認知・活動という3つのレベルに区分して整理する。

表 11 品詞区分別漢語動名詞の意味分類

名詞的 <----->		動詞的	
I 名詞的名詞	II 動詞的名詞	III 名詞的動詞	IV 動詞的動詞
大分類			
人間 (72%)	人間 (60%)	抽象 (55%)	人間 (66%)
抽象 (25%)	抽象 (38%)	人間 (41%)	抽象 (32%)
自然 (3%)	自然 (2%)	自然 (3%)	自然 (2%)
小分類 (上位3種、数字は順位を示す)			
レベル1: 変化			
1. 作用・変化 (例: 働く、反応、変える、反発、転移)【 抽象的变化 】	1. 作用・変化 (例: 働く、反応、変える、反発、転移)【 抽象的变化 】 3. 増減・補充 (例: 増大、減少、加熱、加える、付ける)【 増減的变化 】	1. 成立 (例: 完成、熟成、仕上げ、出来上がり、独立)【 変化の完遂 】	φ
レベル2: 認知			
1. 学習・習慣・記憶 (例: 習う、練習、磨く、学習、独学…)【 学習的認知 】	2. 注意・認知・了解 (例: 注意、用心、配慮、気づき、注目)【 一般的認知 】	φ	1. 注意・認知・了解 (例: 注意、用心、配慮、気づき、注目)【 一般的認知 】
レベル3: 活動			
2. 書き (例: 書写、執筆、表記、登録、記載)【 記録的活動 】 2. 技術・設備・修理 (例: 設ける、準備、飾る、処理、片付ける)【 整備的活動 】	φ	2. 動き (例: 始動、主動、駆動、運動、波動)【 一般的活動 】 2. 行為・活動 (例: する、受ける、取り組む、振る舞う、行動)【 一般的活動 】	2. 見る (例: 発見、拝見、閲覧、観察、見回す)【 視覚的活動 】

注：例は『分類語彙表』から引用したものである。

大分類について言うと、一部のずれはあるものの、人間>抽象>自然というパターンは共通であるように考えられる。むしろ、重要なのは小分類である。表 11 からわかるように、上位 3 種の小分類に注目すると、I からIVへ、変化、認知、活動という 3 つのレベルにおいて意味が変化している可能性が認められる。

まず、第 1 レベルについて、グループ I では、「働く」や「変える」のような抽象的变化の含意を持つ語が多い。グループ II では、抽象的变化を含意する語のほか、「増大」や「減少」のような増減を含む変化を表す語も多い。グループ III では、「完成」や「熟成」のような変化の完遂を表す語が多い。また、グループ IV の上位 3 種の小分類には変化を表す語がないが、すでに述べたように、4 つの区分は動詞性・名詞性の強弱によって連続体モデルになることで、φ は一つ前のグループの状態が継続しているものと考えられ、グループ IV は完遂した状態の継続を含意するものと言える。実際、グループ IV では変化の完遂を含意する漢語動名詞 2 語が出現している（「作成」「設置」）。これより、I からIVへ、「一般的で抽象的变化→増減を含む抽象的变化→変化の完遂→（変化の完遂）」のような意味が変化している可能性があると思われる。BCCWJ におけるそれぞれの区分での漢語動名詞の具体的用例を見てみよう。なお、当該語は【】記号で示し、当該語が所属する品詞区分は[]で示す（以下同）。

- (24) 結婚で改姓のある場合は、できるだけ早く知らせてもらえるよう、社員に提示しています。ウチの場合は、社内の書類の改姓手続きだけなら、大した作業ではありませんが、給料の振込の際、本店への電算処理の依頼と、本人の口座の名義の【変更】との兼ね合いがあるので、そこだけは変更時期に注意を要します。（特定目的・知恵袋、OC11_01452）[名詞的名詞]
- (25) 業種別では、デジタル家電販売が好調を継続しているほか、アジア向け輸出の【増加】などにより自動車も堅調で、大企業の製造業を中心に収益が改善。（出版・新聞、PN4d_00029）[動詞的名詞]
- (26) このダムの工事は、地元の農夫が土を運び、あるいは鳶の仕事まで身につけて【完成】させた。大正十五（千九百二十六）年三月に起工した工事は、昭和五（千九百三十）年に【完成】した。（図書館・書籍、LBe9_00024）[名詞的動詞]
- (27) 確認事務を受託した法人は、警察署長の指示に基づき、巡回計画書を【作成】し、警察署長の認可を受けた上で、計画に従い、駐車監視員を巡回させます。（出版・書籍、PB46_00114）[動詞的動詞]

以上の用例の下線部はどれも物事の変化（またその完遂）を表している。(24) は結婚によって口座の名義を変え改めること、(25) は自動車をアジア向けの輸出の量が増えること、(26) はダムの工事の継続状態が変化し、昭和五年にそれが完了したこと、(27) は巡回計画書の作成という状況が変化し、それが完成したことをそれぞれ意味する。名詞的名詞は一般的で抽象的变化、動詞的名詞は増減を含む抽象的变化、名詞的動詞と動詞的動詞は変化の完遂を含意することがわかる。

次に、第 2 レベルについて、グループ I では、「習う」や「練習」のような学習的認知の含意を持つ語が多い。グループ II とグループ IV では、「注意」や「用心」のような一般的認

知行為を表す語が多い。また、グループⅢはグループⅡの意味の継続であり、一般的認知行為を含意すると考えられる。実際、グループⅢでは一般的認知行為にまとめられる[決心・解決・決定・迷い]（「制定」）や[思考・意見・疑い]（「検討」）に属する漢語動名詞も出現している。これより、ⅠからⅣへ、「学習的認知行為→一般的認知行為→（一般的認知行為）→一般的認知行為」のような意味が変化している可能性があると考えられる。

- (28) 前田は家庭教師二人に来てもらって、明けても暮れても熱心にロシア語の【学習】を始めた。(特定目的・ベストセラー、OB1X_00077) [名詞的名詞]
- (29) これは、談合行為を反復して行っていた事業者に対し、従業員などに、当該行為が違法であることを明白に【認識】させ、違反行為を繰り返させないようにすることを狙った措置です。(図書館・書籍、LBj3_00128) [動詞的名詞]
- (30) こういような場合には、その段階での金利や物価や交通量等の状況を適切に把握して、機構の償還状況や、国、地方の支援措置の可能性などを総合的に【検討】した上で適切に判断する、こういうことになろうかと思えます。「名詞的動詞」
- (31) 引き落とし金融機関によって違うようです。URLより 【確認】 してみてください。(特定目的・知恵袋、OC03_00008) [動詞的動詞]

以上の用例の下線部はいずれも人間の認知行為を表している。(28) はロシア語という外国語を身に付けるための学習的認知行為、(29) は違法行為をよく知る・理解するという認知行為、(30) は様々な状況をよく調べて、それ（償還計画）がいいかどうかを考える認知行為、(31) は「引き落とし金融機関によって違う」という事実をはっきり認める行為である。これより、名詞的名詞は学習的認知行為、動詞的名詞と名詞的動詞、動詞的動詞は一般的認知行為を含意することがわかる。

最後に、第3レベルについて、グループⅠでは、「書写」や「執筆」のような記録の活動と「設ける」や「準備」のような整備的活動、すなわち、具体的で細かい作業と言える活動の含意を持つ語が多い。グループⅢでは、「する」や「取り組む」、「振る舞い」のようなより日常的で頻繁に行っている一般的活動を表す語が多い。グループⅣでは、「発見」や「観察」のような視覚的活動、つまり、人間の感覚器官が外界の物事を捉え、それが何であるかを区別する知覚的活動を表す語が多い。また、グループⅡはグループⅠの意味の継続であり、細かい作業的活動を含意すると考えられる。実際、グループⅡでは細かい作業的活動にまとめられる[扱い・操作・使用]（「適用」）や[測定・計算]（「計算」）に属する漢語動名詞も出現している。これより、ⅠからⅣへ、「具体的で細かい作業的活動→（具体的で細かい作業的活動）→抽象的で一般的活動→知覚的活動」のような意味が変化している可能性があると考えられる。

- (32) 平成十三年度の典型7公害の苦情のうち、直接処理した苦情件数について、苦情の【処理】のための防止対策の有無、防止対策を講じたものについてはその内容及び対策の実施までに要した期間をみると、次のとおりとなっている。(特定目的・白書、OW6X_00357) [名詞的名詞]
- (33) 帯分数のかけ算やわり算は、帯分数を仮分数になおして、真分数と同じように【計算】します。(特定目的・教科書、OT11_00013) [動詞的名詞]
- (34) 円滑な労働移動とは、失業を経ないということだけでなく、個々の能力を十分発揮

できるような転職先への【移動】をも意味する。（特定目的・白書、OW5X_00176）〔名詞的動詞〕

- (35) 同見学会では、河内浦城跡公園などを見学した後、恐竜の足跡の化石を【観察】。参加した子どもたちは、市の学芸員の説明を聞きながら、熱心に【観察】していました。（特定目的・広報紙、OP95_00001）〔動詞的動詞〕

以上の用例の下線部はいずれも人間の活動を含意するが、(32) はどのような対策を講じるか、どのように対策を実施するかなどの具体的な作業を通して苦情を処理する活動、(33) は帯分数を仮分数になおすという細かい作業を通して帯分数のかけ算やわり算をする活動、(34) は転職という抽象的で一般的な活動、(35) は子供たちが学芸員の説明を聞きながら、化石を理解しようと自分の目でよく見るといった知覚的活動である。これより、名詞的名詞や動詞的名詞は具体的で細かい作業的活動、名詞的動詞は抽象的で一般的活動、動詞的動詞は知覚的活動を含意することがわかる。

以上より、I からIVへ、つまりは名詞性の強い漢語動名詞から動詞性の強い漢語動名詞にかけて、以下のような意味拡張のパターンが存在している可能性がある。

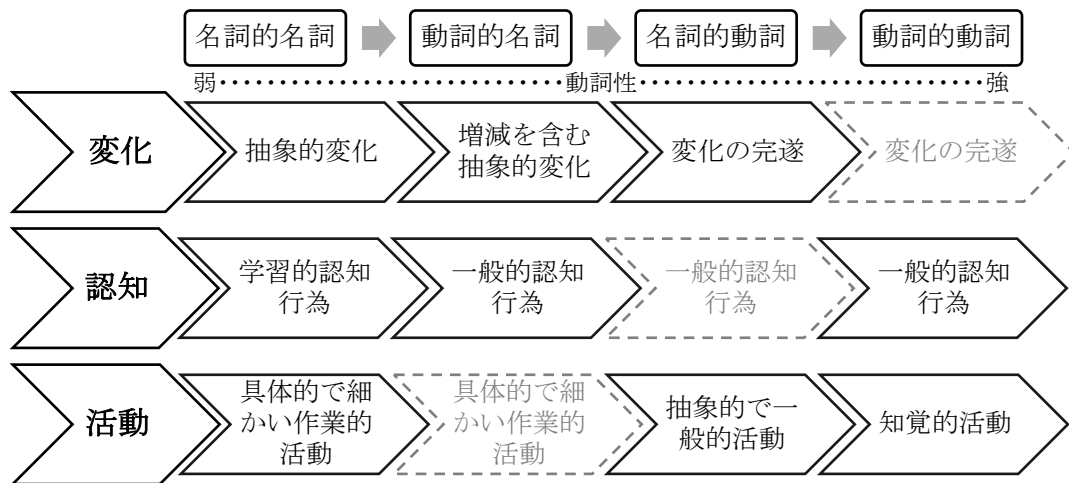


図3 品詞区分別漢語動名詞の意味拡張パターン

RQ3 からわかるように、各タイプの意味特性については重なりも多いものの、全体的に見れば、漢語動名詞の動詞性が高まるにつれて、典型的な意味が、変化、認知、活動の3つのレベルで、図3のような段階的に変化している可能性が示唆される。

5. まとめと今後の課題

以上、本研究では、書き言葉・話し言葉のコーパスデータを用い、動詞性判定テストと名詞性判定テストによって、重要漢語動名詞 200 種の品詞用法別の頻度を調査した。3つのRQを通して、漢語動名詞が4群に区別されること、それらが異なる統語的・意味的特性を有する可能性が示唆された。以下は知見をまとめた図である。

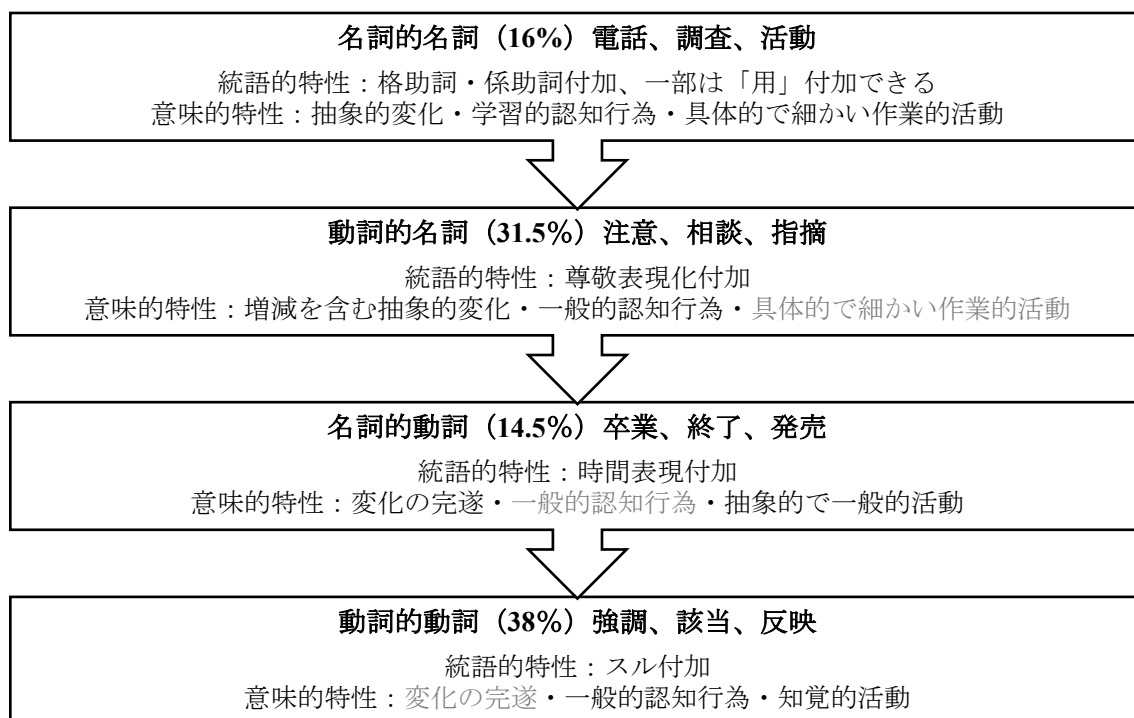


図4 名詞性・動詞性テストを統合した漢語動名詞の品詞分類の枠組み

日本語教育の観点で言えば、漢語動名詞は数も多く、体系的に教えることが難しい。しかし、本研究で得られたような分類枠を用い、グループ別の特性を理解させ、グループを代表する漢語動名詞とその用例を示すようにすれば、学習者が理解しにくい漢語動名詞の品詞用法をある程度効率的に指導できるのではないかと考えられる。

教育的援助の一例としては、本研究で得られた知見を身に着けるための学習タスクの開発が考えられる。例えば、図2や図3を提示したうえで、表10からいくつかの漢語動名詞を選択し、学習者個人で、あるいはペアによる相談で、個々の漢語動名詞がそれぞれどのグループに入るかを推測させるようなタスクを行うこともできる。こうしたタスクを通して、学習者は、漢語動名詞が一つのかたまりではないこと、品詞性と意味が混ざり合う形で、いくつかの内部的なグループに分かれることを、体験的に理解することができるだろう。

しかしながら、今回の調査には課題も多く残されている。たとえば、意味タイプの判別は筆者が単独で行ったため、母語話者の協力を得て妥当性の検証を行う必要がある。また、語構成や、語の造語年などを含めた調査も必要であろう。今後、調査の範囲を広げ、漢語動名詞の品詞性をさらに多角的に検討していく予定である。

文 献

- 石川慎一郎・前田忠彦・山崎誠 (2010) 『言語研究のための統計入門』くろしお出版。
 国立国語研究所 (2004) 『分類語彙表：増補改訂版』国立国語研究所資料集 (14)。
 小林英樹 (2004) 『現代日本語の漢語動名詞の研究』ひつじ書房。
 田辺和子・中條清美・船戸はるな (2011) 「新聞コーパスにおける二字漢語動名詞の動詞的・

- 名詞的ふるまいについて」『日本女子大学紀要文学部』 61, pp.19-32.
- 田辺和子・中條清美 (2014) 「日英新聞コーパス及び BCCWJ 比較に基づく二字漢語動名詞使用の分析」『日本女子大学紀要文学部』 63, pp.1-11.
- 陳迪 (2022) 「書き言葉・話し言葉を含む 72 変種のコーパスデータの調査に基づく「重要漢語動名詞 200」の抽出の試み」『統計数理研究所共同研究レポート』 456, pp. 21-41.
- 永澤済 (2010) 「変化パターンからみる近現代漢語の品詞用法」『東京大学言語学論集』 30, pp.115-168.
- 間淵洋子 (2015) 「二字漢語名詞サ変用法の変化—『太陽コーパス』『BCCWJ』を用いて—」『第7回コーパス日本語学ワークショップ予稿集』 pp.275-284.
- 間淵洋子 (2016) 「近現代漢語におけるサ変動詞用法の変化—形態論情報付きコーパスを用いて—」『国際日本学研究論集』 4, pp.17-36.
- 李楓 (2014) 「漢語サ変動詞の卓立性の再考—動詞形・構文形比率を手掛かりとして—」『第6回コーパス日本語学ワークショップ予稿集』 pp.63-72.
- 林炫情・李在鎬・宮岡弥生・柴崎秀子・趙垺熙 (2012) 「言語処理技術を利用した日本語学習者作文コーパスの開発」『日本文化学報』 56, pp.129-142.

関連 URL

- | | |
|---------------------|---|
| コーパス検索アプリケーション『中納言』 | https://chunagon.ninjal.ac.jp/ |
| 形態素解析ツール『Web 茶まめ』 | https://chamame.ninjal.ac.jp/ |
| 日本語学習者作文コーパス | http://sakubun.jpn.org/ |
| 『デジタル大辞泉』 ジャパンナレッジ版 | https://japanknowledge.com/library/ |

上級日本語学習者の文体把握に関するケーススタディ —「BCCWJ 図書館サブコーパス文体情報」を用いた読解調査—

小西 円 (東京学芸大学) †

A Case Study of Advanced Japanese Language Learners' Understanding of Styles: Reading Comprehension Survey Using the BCCWJ Library Subcorpus Stylistic Information

Madoka Konishi (Tokyo Gakugei University)

要旨

本研究では、日本語学習者のための類義表現の記述方法に関する研究の一環として、上級日本語学習者が日本語の文章を読む際に文体をどのように把握しているかを探るために調査を行った。本研究は、その調査をケーススタディとして、分析を行うものである。調査には「BCCWJ 図書館サブコーパス文体情報」を用いて、そこで示された指標と本研究の調査との差を分析した。その結果、文体的な指標では特に「硬度」「くだけ度」の理解に難しさが見られたが、その理解を生む過程が、学習者によって異なっていることがわかった。丁寧体・普通体の理解が学習者によって異なり、それが影響している側面があった。また、自分が知らない語や漢字語を硬く感じる、オノマトペを軟らかく感じるなどの特徴が見られた。また、「客観的で硬い」のに「くだけている」文章の文体的な理解が難しいこともわかった。

1. はじめに

日本語学習者が多様な類義表現を使い分けるためには、それぞれの表現をいつ使うかに関する情報が必要である。その違いは、「話しことば」「書きことば」といった用語で説明されたり、「レポートや論文で使う」というように文章・談話のジャンルを用いて記述されることが多い。しかし、「話しことば」「書きことば」という二項対立はあいまいさを含んでいる。

通常、「話しことば」「書きことば」は媒体が音声か文字かという点によって区別される(日本語記述文法研究会編 2009)。それと同時に「話しことば」は、場面依存性、縮約性、即興性、冗長性、対面性を持つ(日本語記述文法研究会編 2009)。「なければ」が「なきゃ」になるなどの音声化の負担を軽減する縮約性や、その場で考えたことを瞬時に発話するために起こる倒置や主述の不一致(即興性)、また、発話したその瞬間から消えていく音声的な情報の伝達性を高めるために同じことを何度も繰り返す等することによる文の整合性の欠如(冗長性)等によって、「話しことば」は必然的に「くだけ度」の高さを兼ね備えてしまうと言えるだろう。「くだけている」「改まっている」というような文体的な特徴は、本来は「媒体が音声であること」と区別されるべきであろうが、「話しことば」という用語ではそれらを区別することが難しい。

また、「レポートや論文」というジャンルは、そのようなジャンルの文章を産出する必要の学習者にとっては必須で有用な提示方法であるが、それ以外のジャンルでの対応が難し

† kmadoka@u-gakugei.ac.jp

い。

そのため、類義表現の使い分けは「話しことば」「書きことば」といった二項対立にとどまらない多様な情報によって示される必要がある。その一つとして、「硬い／軟らかい」「くだけている」などの文体に関する情報がある。文体情報を用いて類義表現を記述するためには、日本語学習者が文体をどのように把握しているかを知る必要があるが、現状では十分に明らかになっているとは言えない。そのため、本発表では、日本語学習者が日本語の文章の文体をどのように把握しているかに関する調査をケーススタディとして行い、その結果と考察を行う。このケーススタディは、12名の日本語学習者を対象とした読解調査を通して彼らの文体把握の様子を分析するものである。以下、詳細を示す。

2. 調査について

2.1 「BCCWJ 図書館サブコーパス文体情報」

本調査では使用データとして「BCCWJ 図書館サブコーパス文体情報」（以下、「図書館サブコーパス文体情報」）を用いる。「図書館サブコーパス文体情報」は、『現代日本語書き言葉均衡コーパス』（以下、BCCWJ）の「図書館サブコーパス」に収録されている書籍サンプルに対して、人手で文体情報を付与したデータである（柏野 2013、国立国語研究所 2015）。付与されている文体情報は、(A) 対象読者に想定される読解レベル、(B) テキストの作成意図、(C) さまざまな文体情報の3種である。具体的には、(A) として「専門度」、(B) として「客観度」、(C) として「硬度」「くだけ度」「語りかけ性度」の5つの指標がある。それぞれの指標は表1のような段階があり、各書籍サンプルに以下のような数値が付与されている。

表1 「図書館サブコーパス文体情報」の各指標

専門度	1 専門家向き, 2 やや専門的な一般向き, 3 一般向き, 4 中高生向き, 5 小学生・幼児向き
客観度	1 とても客観的, 2 どちらかといえば客観的, 3 どちらかといえば主観的, 4 とても主観的
硬度	1 とても硬い, 2 どちらかといえば硬い, 3 どちらかといえば軟らかい, 4 とても軟らかい
くだけ度	1 とてもくだけている, 2 どちらかといえばくだけている, 3 くだけていない
語りかけ性度	1 とても語りかけ性がある, 2 どちらかといえば語りかけ性がある, 3 特に語りかけ性はない

「図書館サブコーパス文体情報」の文体研究における有用性は馬場（2018）によっても示されている。馬場（2018）は BCCWJ の図書館サブコーパス内の語（語彙素）に対して「図書館サブコーパス文体情報」の5つの指標別の平均値を算出し、それらの平均値が語の文体差に関する内省判断と強い相関があることや、和語・漢語・外来語の特徴の違いが平均値の違いに現れていることなどを示した。そのため、「図書館サブコーパス文体情報」のアノテーションデータは文体情報に関する一定の信頼性があると判断し、本調査のデータとして使用する。

2.2 使用データ

本調査では、「図書館サブコーパス文体情報」の5指標を参考にして、さまざまな文体的特徴を持った以下の5つのサンプルを選出し、これを調査使用データとした。各サンプル

とタイトルと 5 指標の詳細を表 2 に示す。各サンプルは「j-Readability」¹によって難易度を調査し、調査協力者となる日本語学習者にとって難易度が高すぎないように考慮した。表 2 に示した難易度は、「j-Readability」のサイトで判定されたものである。

サンプルの具体的な文章は表 3 に示す。実際の調査では、表 3 の文章を含む 500 文字前後を使用した。また、中級以上の漢字語にはフリガナをつけた。

表 2 調査使用データ

ID	タイトル (略称)	文体情報の指標	難易度	
LBi4_00028	岩波講座現代の物理学 (物理)	1 専門家向き 1 とても硬い 3 特に語りかけ性はない	1 とても客観的 3 くだけていない	中級後半
LBt1_00037	松岡修造のカッコいい大人になるための 7 つの約束 (約束)	4 中高生向き 4 とても軟らかい 1 とても語りかけ性がある	4 とても主観的 1 とてもくくくくくく	中級前半
LBon_00004	キラキラ恋うらない (占い)	5 小学生・幼児向き とても軟らかい とても語りかけ性がある	4 とても主観的 1 とてもくくくくくく 1	中級前半
LBf9_00067	男はオイ！女はハイ… (男女)	3 一般向き 4 とても軟らかい 2 どちらかといえば語りかけ性がある	4 とても主観的 1 とてもくくくくくく	中級前半
LBd4_00005	新史・動物行動記 (動物)	3 一般向き 2 どちらかといえば硬い くだけていない 3 特に語りかけ性はない	2 どちらかといえば客観的 1 とても	上級前半

表 3 調査使用データの文章 (一部)

物理	われわれは、なにゆえに、自由な Dirac 場が Fermi 統計をみだし、自由な Klein-Gordon 場が Bose 統計に従うかをみてきた。相対論的な自由場はこれ以外にもさまざまなスピンを記述するものが知られている。これらを逐一述べるのは本書の目的ではない。ただスピンと統計の関係として次の定理 (W.Pauli, 1940) はよく知られている。 * 証明法はいろいろ考えられるが、前節の方法の一般化としては、例えば文献 [8] を参照。
約束	僕はスポーツが大好きです。自分の頭、身体、そして心、持てる力のすべてをありったけぶつけることができますからです。 結果がはっきりと出るのもいいですね。記録や勝ち負けという形で、そのときの能力やそれまで積み重ねてきた毎日の努力が、具体的に表れてきます。「目標を立てる」→「努力する」→「(残念な) 結果が出る」→「目標を立て直す」→「また努力する」→「(いい) 結果が出る！」という流れが、スポーツをやっていると自然に身についてくるのです。 これは、将来スポーツと関係のない仕事に就いたとしても役立つことだと思うんです。だから、みんなにも、何かひとつはスポーツを経験してもらえたら、と思います。
占い	コインやトランプ…etc。さあ、今度はいろ～んなタイプのうらないで、恋のことさぐっちゃお。 1 絶対、モテモテ少女になりたあ～い！！

¹ 日本語の学習者と教師のための Web システム「jReadability」の中の「日本語文章何度判定システム」を使用した。日本語の文章テキストを入力すると、その難易度を 6 段階で判定するシステムである。(李 2016) <https://jreadability.net/>

	<p>恋の五百円玉うらない</p> <p>■うらない方 五百円玉を1個用意してね。そして、イラストのように、左手の人さし指のツメで五百円玉を立て右手の人さし指のツメで、チョンと五百円玉をはじいて回転させるの。クルクルと回転した五百円玉が止まったとき、どっちの面が上になっていたかを確認するのを忘れずにね。</p>
男女	<p>つい先日の話だ。</p> <p>最近流行りの通信販売。例の新聞の日曜版の裏面などに、克明にズラリと商品が写真などで広告してあるやつ。あれをば何となく眺めているうちに、どうしても欲しくなった商品があった。</p> <p>よし、こいつひとつについてやれとばかりすぐ電話にとびついた。</p> <p>「ハイ、こちら一です」と出たのは、耳ざわりだけでわかるアルバイトギャルの声。</p> <p>「商品番号をおっしゃって下さい」といわれて答える。</p> <p>さらに「御住所と御名前、電話番号を郵便番号からどうぞ」ってんで、こいつにも律儀に返事をする。</p> <p>が、その次がいけない。</p> <p>「では生年月日とお年をどうぞ」</p> <p>「え！」</p>
動物	<p>5 動物たちの愛は血縁淘汰では解けない 血縁者でない仲間を多く助けたことこそ大切</p> <p>私が「警戒音を発して仲間を逃がす利他行動」を遺伝子レベルで考え、その犠牲になった個体の遺伝子が生きのこった兄弟の中に伝わるからいいのだという「血縁淘汰」の考え方にハテおかしいぞと思ったのは、すぐにいくつかの疑いが生まれたからである。</p> <p>だいいち、犠牲者の兄弟はそれで助かって、大へん結構だったが、多くの血縁者でない仲間たちはどう評価する？一彼らの方がはるかに多く助かっていることも多いだろうに。それは犠牲者の遺伝子とかかわりがないから、「派生的なもの」で、行動学的には無視するのか、考えなくていいのか、ソナバカナ、ということが一つ。</p>

各調査使用データの文章は、次のような文体的特徴を持っている。『物理』（以降、サンプルのタイトルは表2に掲載した略称で示す）は「専門家向きで、客観的で、硬く、くだけておらず、語りかけ性がない」文体である。『約束』は『物理』とはおおよそ逆の文体的特徴をもっており、「中高生向きで、主観的で、軟らかく、くだけており、語りかけ性がある」文体である。『約束』では丁寧体が用いられている。また、『占い』は『約束』と類似の傾向を持っているが、対象読者がより幼くなり、「小学生・幼児向きで、主観的で、軟らかく、くだけており、語りかけ性がある」文体である。『男女』は『約束』とほぼ同じ特徴を持ち、対象読者のみが異なっている。対象読者が異なることに起因して、『男女』は使用されている語に漢語や難度の高い語が増え、また、文体が普通体である。一方、『動物』は、「客観的で硬い」のに「くだけている」という点で、そのほかの4サンプルと異なっている²。馬場（2018）は、「専門度」「客観度」「硬度」「くだけ度」の4指標が相互に強い相関があることを明らかにしているが、その点から見ると『動物』はやや異質なサンプルである。『動物』は他の4サンプルより難易度がやや高めであるが、日本語学習者がこの種の文体をどのように把握するかを確認するために、調査対象とした。

² 一般的には、「客観的で硬い」文体は「くだけていない」場合が多く、『図書館サブコーパス文体情報』でも、「とても／どちらかといえば客観的」で「とても／どちらかといえば硬い」文体の文章2370サンプルのうち、「とてもくだけている」のは『動物』を含む2サンプルだけであったしかし、メイナード（2008）が「マルチジャンルという現象こそがごく頻繁に見られる言語のバリエーションの一種である」（p.14）と指摘するように、ひとつのディスコースの中に他のジャンルの特徴とされる表現が混用される例は少なくない。『動物』もそのような捉え方ができる可能性がある。

2.3 調査協力者

調査協力者は表4に示す12名で、2022年6月の調査時点で日本の大学または大学院に在籍している留学生である。日本語の習熟度を把握するために、SPOT90(小林 2015)を受験してもらった。SPOT90の成績目安は、「上級が81~90点」「中級が56~80点」となるため³、中級と判定された協力者H~Lも、中級の上位であると考えられる⁴。

表4 調査協力者一覧

協力者	母語	日本語 学習期間	日本 滞在期間	SPOT レベル	SPOT 点数
A	中国語	8年7か月	3年7か月	上級	88
B	トルコ語	5年6か月	0年9か月	上級	86
C	中国語	8年8か月	3年5か月	上級	85
D	中国語	6年3か月	5年8か月	上級	84
E	中国語	6年8か月	2年8か月	上級	83
F	韓国語	5年6か月	1年2か月	上級	83
G	ベトナム語	5年0か月	0年9か月	上級	81
H	中国語	6年3か月	2年8か月	中級	80
I	中国語	6年6か月	4年0か月	中級	80
J	韓国語	6年10か月	0年8か月	中級	80
K	中国語	6年0か月	5年9か月	中級	77
L	中国語	8年7か月	5年2か月	中級	72

2.4 調査実施手順

調査は、①背景調査シートの記入、②読解調査5種、③読解調査の回答に対するインタビュー、④SPOTの受験、という流れで行った。5種の読解は、調査協力者によって読む読解文の順序を変え、読解順序の影響が偏って出ないようにした。調査はテストではないので、わからない語や表現があってもかまわないことや、趣味で読書をするときと同じような読み方をしてもらうことを調査協力者に伝えた。そのため、辞書を使いたい場合は使ってもよいと伝えたが、辞書を使った者はほとんどおらず、また、使った場合も文体理解に影響を及ぼす語ではなかった。また、SPOTの受験は自宅で都合の良い時間に行ってもらった。

読解調査の設問は表5のとおりである。「専門的」「客観」「主観」「硬い」「軟らかい」「くだけている」「語りかける」という重要語については、調査前に各調査協力者に理解を確認した。

「◆内容がどれくらい理解できましたか。」という設問は、日本語学習者である調査協力者の文章理解の程度を知ることが目的としたものである。また、「◆この文章はどれに当てはまると思いますか。」の各設問は、「図書館サブコーパス文体情報」の指標を問うものである。「◆(3)~(5)について、どうしてそう思いましたか。」は、そのような指標の判断に至った語や表現、文法項目などを把握することを目的とした設問である。柏野(2013)によると、(1)で問う専門度は対象読者に想定される読解レベル(難易度)、(2)で問う客観度はテキスト作成意図を示しており、文体情報を表すのは(3)「硬度」、(4)「くだけ度」、(5)「語りかけ性度」である。そのため、文体情報の把握について知るために、(3)~(4)について詳細を問う。また、読解文は、BCCWJの図書館サブコーパスのサンプルとなる段階で、レイアウトやフォントの情報などが取り除かれている。現物の書籍であればイラストがついていたり、フォントやレイアウト等に工夫があったりすると思われるが、今回の調査は、日本語学

³ 筑波日本語テスト集 TTBJ のサイトに掲載がある。 <https://ttbj.cegloc.tsukuba.ac.jp/p1.html>

⁴ そのため、本稿のタイトルは両者をまとめて「上級」と仮称している。

習者が日本語の文章からどのような文体情報を取り出すかを知ることが目的としているため、レイアウト等の情報は調査から除外された状態で行った。

表5 読解調査の設問

<p>◆内容がどれくらい理解できましたか。○をつけてください。 とてもよく理解できた ・ だいたい理解できた ・ あまり理解できなかった</p> <p>◆この文章はどれに当てはまると思われますか。○をつけてください。 (1) とても専門的 ・ やや専門的 ・ 一般の人向き ・ 中学や高校生向き ・ 子供向き (2) とても客観的 ・ どちらかといえば客観的 ・ どちらかといえば主観的 ・ とても主観的 (3) とても硬い ・ どちらかといえば硬い ・ どちらかといえば軟らかい ・ とても軟らかい (4) とてもくだけている ・ どちらかといえばくだけている ・ くだけていない (5) とても語りかけられている感じがする ・ 少し語りかけられている感じがする ・ 語りかけられていない感じがしない</p> <p>◆(3)～(5)について、どうしてそう思いましたか。(あとでインタビューでも聞きますから、線を引くのが難しい場合はひかなくてもいいです) (3)について(硬い・軟らかい) そのように思った部分に_____を引いてください。 (4)について(くだけている) そのように思った部分に_____を引いてください。 (5)について(語りかけ) そのように思った部分に <input type="checkbox"/> をつけてください。</p>

2.5 指標の計算手順

各調査協力者が「◆この文章はどれに当てはまると思われますか。○をつけてください。」という設問で回答した結果から、各指標の数値を得ることとした。指標を数値に表す場合は、各項目の一番左端を1点とし、右に進むにつれて1点加算されることとした。たとえば、「客観度」において「とても客観的」は1点、「どちらかといえば客観的」は2点、「どちらかといえば主観的」は3点、「とても主観的」は4点である。

このように計算した各調査協力者の結果を基に、各サンプルの各指標における平均値を算出した。

3. 調査結果と分析

3.1 調査結果の全体像

まず、調査から得られた各指標の値を表6に示す。「図書館サブコーパス文体情報」の指標の値を「指標」列の「コーパス」の行に示す。また、「指標」列の「調査結果」の行に、各調査協力者の選択した指標の平均値を示す。また「差」の行には、「図書館サブコーパス文体情報」の指標と本調査結果で得られた平均値の差を絶対値で示す。

表 6 各指標の調査結果と「図書館サブコーパス文体情報」の指標との差

サンプル	指標	理解度	専門度	客観度	硬度	くだけ度	語りかけ度	差の合計
物理	調査結果	2.8	1.2	1.2	1.3	3.0	2.9	
	コーパス		1.0	1.0	1.0	3.0	3.0	
	差		0.2	0.2	0.3	0.0	0.1	0.8
約束	調査結果	1.0	3.8	3.3	3.2	2.3	1.5	
	コーパス		4.0	4.0	4.0	1.0	1.0	
	差		0.3	0.8	0.8	1.3	0.5	3.7
占い	調査結果	1.1	3.7	3.3	3.9	1.3	1.3	
	コーパス		5.0	4.0	4.0	1.0	1.0	
	差		1.3	0.7	0.1	0.3	0.3	2.7
男女	調査結果	1.5	3.1	3.8	3.2	1.5	1.8	
	コーパス		3.0	4.0	4.0	1.0	2.0	
	差		0.1	0.3	0.8	0.5	0.2	1.9
動物	調査結果	1.7	2.4	3.3	2.3	2.6	2.5	
	コーパス		3.0	2.0	2.0	1.0	3.0	
	差		0.6	1.3	0.3	1.6	0.5	4.3

※理解度：とてもよく理解できた 1、あまり理解できなかった 3

専門度：とても専門的 1、子供向き 5 客観度：とても客観的 1 とても主観的 4

硬度：とても硬い 1、とても軟らかい 4 くだけ度：とてもくだけている 1、くだけていない 3

語りかけ度：とても語りかけられている感じがする 1、語りかけられている感じがしない 3

3.2 読解サンプルからみた分析

表 6 をもとに、読解サンプルごとに分析を行う。

5 種の読解サンプルの差の合計を比較すると、「図書館サブコーパス文体情報」の指標と最も調査結果が近いものは、差が 0.8 の『物理』である。『物理』は、理解度は最も低いが、文体情報は母語話者とほぼ同様に把握できている。調査の回答およびインタビューからは、「Dirac 場」や「Fermi 統計」などの専門用語がわからないため、内容を理解することはできないが、定理が示されていたり文献情報が掲載されたりしていることから、「論文っぽいもの」として理解したと述べる調査協力者が多く、「硬い」「くだけていない」「語りかけていない」と判断することが可能だったとみられる。

一方、差の合計が大きい読解サンプルは差が 3.7 の『約束』と 4.3 の『動物』である。『約束』は「客観度」「硬度」「くだけ度」で、『動物』は「客観度」「くだけ度」で差が大きい。『約束』は口頭で行われたスピーチを文字に落としたような文章で、丁寧体で書かれている。また、『動物』は扱う内容が生物学的でやや専門性の高いものであるが、「ソナバカナ」のようなくだけた表現がちりばめられており、「客観的で硬い」のに「くだけている」という面で文体を理解するのが難しいと思われるサンプルである。差の合計は『動物』が最も高く、調査協力者がこの種の文体の理解に苦労したことがわかる。

差が 2.7 の『占い』と 1.9 の『男女』は上記 2 タイプの中間に位置する。『占い』で差が大きいのは「専門度」と「客観度」である。「専門度」では「中高生向き」と答えた調査協力者が多く、インタビューで「日本の女子高生には占いが人気であるから」という趣旨の発言が複数あった。「また、「客観度」では「客観的」と答えた調査協力者が多く、インタビューには「占いを科学的だと考えるかどうかは人による。占いを科学的だと考える人は、この文章は客観的だと思うだろう」という趣旨の発言が複数見られた。対象読者やテキストの作成意図などに関する部分は、それが読まれる社会文化的な理解とも関連することがわかる。また、『占い』には「いろ～んな」や「なりたあ～い」のようなくだけた表現や「モテモテ」

のようなカタカナ表記、オノマトペが多用されており、これらから「軟らかい」「くだけている」と判定した調査協力者が多かった。その結果、「硬度」「くだけ度」に関しては「図書館サブコーパス文体情報」の指標とほぼ同様の結果となっている。『男女』で差が大きいのは「硬度」で、文体情報に関わる指標である。

『約束』『動物』『男女』では文体に関わる指標で偏差が大きい。次の節では、これら3つの読解サンプルの文体情報に関わる「硬度」「くだけ度」の分析を行う。

3.3 指標からみた分析—「硬度」

本節では、文体情報に関わる指標を扱う。その中でも、3.2節で差が大きかった「硬度」「くだけ度」を中心に分析する。

まず、「硬度」について分析する。『約束』も『男女』も、「図書館サブコーパス文体情報」では「とても軟らかい」となっている。しかし、本調査では『約束』は「どちらかといえば軟らかい」を選んだ調査協力者がもっとも多く8名、次いで「とても軟らかい」が3名、「どちらかといえば硬い」が1名であった。また、『男女』では「どちらかといえば硬い」が4名、「どちらかといえば軟らかい」が2名、「とても軟らかい」が4名で、ばらつきが大きかった。

インタビュー等を通してみると、「軟らかい」と判断するに至った語や表現は、『約束』では「でしょう」や終助詞の「よ」を伴う表現が目立ち、「話しことばなので軟らかく感じる」

『です・ます体』は軟らかく感じる」といった意見が多かった。『男女』では「やつ」「よし」のようなくだけた語や、「ガッチャーン」のようなオノマトペが「軟らかい」と判断する根拠となっており、『約束』と同様に「話しことばなので軟らかく感じる」という意見が見られた。また、『男女』は「克明」「凄みのこもった詰問調」など、日本語学習者にとって難度の高い語が複数あり、「漢字語が多いので硬く感じる」「会話で使わない語／自分の知らない語は硬く感じる」と答えた調査協力者が見られ、それが「どちらかといえば硬い」という指標の回答が多かった理由であろう。

また、「硬度」を判断する際の文体の捉え方には、異なる意見が見られた。『約束』は丁寧体で書かれているが、『です・ます体』は敬語なので硬い」という意見と『です・ます体』は軟らかく、硬いのは『である体』という意見があった。また、『男女』は普通体で書かれているが、「普通体を軟らかく感じる」という意見があった。このような異なる意見は、異なる調査協力者間に見られ、同一の調査協力者内では読解サンプルごとに異なる見解に揺れることはなかった。

調査協力者によって異なるこのような文体理解は、媒体が音声である音声言語における「丁寧体／普通体」の役割と、媒体が文字である文字言語における「丁寧体／普通体」の役割の混同が関係しているように思われる⁵。音声言語においては、丁寧体は対人的な丁寧さを表し、普通体が非丁寧さを表す。つまり、丁寧体は形式的でくだけていないことを表し、普通体はぞんざいさや親密な対応であることを表す。一方、文字言語においては、丁寧体が読み手を意識した丁寧さを表す点は同じであるが、普通体はぞんざいさや親密さを示さず、文字言語における無標の文体である。つまり、丁寧体が形式的でくだけていないことから硬さを表すのは音声言語のみであり、普通体が軟らかさ（ぞんざいさ）を表すのも音声言語においてのみである。日本語学習者である調査協力者には、音声言語におけるその知識が、部分的に文字言語に適応されているのではないだろうか。

⁵ ここでは、音声・文字という媒体の違いを指し示すため、「話しことば」「書きことば」という用語を用いない。音声という媒体で表される言語を「音声言語」、文字という媒体で表される言語を「文字言語」と呼ぶこととする。

3.4 指標からみた分析—「くだけ度」

次に「くだけ度」について分析する。「図書館サブコーパス文体情報」の指標では、『約束』『動物』ともに「とてもくだけている」となっているが、本調査では『約束』は「くだけていない」が5名、「どちらかといえばくだけている」が6名、「とてもくだけている」が1名であった。『動物』では「くだけていない」が7名、「どちらかといえばくだけている」が5名であった。特に『約束』で判断が二分している。

「くだけているかどうか」という判断基準は、調査協力者によって大きく異なっており、いくつかのタイプに分けられることがわかった。1つ目のタイプは、「文が完結していないこと」や「何かが省略されていること」を「くだけている」と捉えるタイプである。そのような調査協力者の場合、『動物』の「～が一つ。」「ということ。」のような体言止めの文にコンピュータがないために「くだけている」と判断したり、『約束』の「もらえたらと思います」のように「たらしい」の「いい」が省かれていることを「くだけている」と判断したりした。2つ目のタイプは、文章の書式や構成に着目するタイプである。そのような調査協力者は、『男女』の「ソナバカナ」のように普段ひらがなで書かれるものがカタカナで書かれていることや、同じ内容を何度も繰り返すことを「くだけている」と捉えた⁶。3つ目のタイプは、表現を判断基準とするもので、「日常的な話しことばはくだけている」と捉えたり、「漢字語が多いのでくだけていない」と捉えたりした。また、「くだけ度」においても、調査協力者によって異なる文体の理解が影響を与えており、『です・ます体』は硬いと考える者は「くだけ度」でも『です・ます体』はくだけていないと判断した。

4. おわりに

本研究は、「図書館サブコーパス文体情報」の指標を用いて、上級日本語学習者が日本語の文章の文体をどのようにとらえるかをケーススタディとして調査し、その結果を考察したものである。調査の結果を、「図書館サブコーパス文体情報」の指標との差という数量的な側面と、インタビュー調査による質的な側面から分析した。その結果、文体的な指標では特に「硬度」「くだけ度」に理解の難しさが見られたが、その理解を生む過程が、調査協力者によって異なっていることが示された。丁寧体・普通体の理解に調査協力者による異なりがあることは注目し得る。ここから、日本語教育において「音声言語における丁寧体・普通体」と「文字言語における丁寧体・普通体」の違いが重要視されていない可能性と、そのことが文体の理解の違いを生じさせることが示唆された。

調査協力者はインタビューで「話しことば」「書きことば」という用語を何度も用いたが、それが何を表すと理解しているのか、捉えることが難しかった。ここには、「話しことば」という用語が、音声という媒体を指すだけでなく、くだけていることや改まっていないことを必然的に含むことによる難しさも関連している。また、自分が知らない語や漢字語を硬く感じる、オノマトペを軟らかく感じるなどの特徴が見られた。また、「客観的で硬い」のに「くだけている」文章の文体的な理解が難しいこともわかった。

日本語学習者に類義表現や文法的な表現を説明・記述する際に、文体情報は重要な役割を果たすと思われる。よりよい記述につなげるために、引き続き、学習者の文体把握に対する理解を深めていく必要があると思われる。

謝 辞

本研究は JSPS 科研費 18K12420 の助成を受けたものである。

⁶ 1つ目のタイプや2つ目のタイプの調査協力者は「くだけている」という語の理解が不十分であった可能性もある。調査前に各語の確認をしたが、調査の中身に関わるため、「くだけている」表現がどんなものかについては具体的に説明しなかった。しかし、調査協力者が上級～中級の後半であることを考えると、「くだけている」ことを十分学ばないことが現状の日本語教育であるとも考えることができる。

文 献

- 柏野和佳子 (2013) 「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), pp.43-53.
- 国立国語研究所 (2015) 『BCCWJ 図書館サブコーパスの文体情報』(2015 年公開第 1 版) <http://id.nii.ac.jp/1328/00003109/>
- 小林典子 (2015) 「SPOT」李在鎬編『日本語教育のためのテストガイドブック』くろしお出版, pp.110 -126
- 日本語記述文法研究会編 (2009) 『現代日本語文法 7 第 12 部談話 第 13 部待遇表現』くろしお出版
- 馬場俊臣 (2022) 『語の文体値データ』(2022 年 2 月公開第 1 版)
<http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/12130>
- 馬場俊臣 (2018) 「『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性」『言語資源活用ワークショップ発表論文集』3, 国立国語研究所, pp.241-256.
- 李在鎬 (2016) 「日本語教育のための文章難易度研究」『早稲田日本語教育学』Vol. 21, pp.1-16.

エンタメ小説における会話文の発話意図分析 —発話文表現文型辞書の改良に向けて—

夏目 和子 (名古屋大学)

佐藤 理史 (名古屋大学)

Analyzing Utterance Intentions in Entertainment Novels

Kazuko Natsume (Nagoya University)

Satoshi Sato (Nagoya University)

要旨

小説の会話文を生成するためのツールとして『発話文表現文型辞書』を編纂し改訂を重ねている。本辞書は、「発話意図」と「表現文型」と「話し方の特徴」で構成されており、「発話意図」の項目の抜本的な見直しが目下の課題である。本研究では、エンタメ小説の会話文を発話意図分析することによって、現在の辞書の発話意図 (60 種) で対応できない発話文はどれか、追加すべき発話意図は何か、どのような発話意図のラベルや体系を採用すれば、小説発話文生成時に使いやすいかなどを検討する。

1. はじめに

小説の会話文を生成するためのツールとして『発話文表現文型辞書』を編纂し改訂を重ねている (夏目 2017, 2019, 2022)。本辞書は、発話意図 (=発話の目的)、表現文型 (=発話意図を表現する言語形式)、話し方の特徴ベクトル (6 軸 20 次元) および例文で構成されている。2022 年 8 月現在の辞書のサイズは、発話意図が 60 種、話し方の特徴を付与した表現文型のエントリ数が 8,055、例文数が 23,489 である。発話文生成および変換における本辞書の役割について 2 節で紹介する。

『発話文表現文型辞書』の改良のために取り組むべき多くの課題のうち、本研究では「発話意図」項目の抜本の見直しに取り組む。まずは現在の辞書の発話意図の課題を 3 節で確認する。それらの課題の解決策をさぐるために、実在のエンタメ小説の会話文に対して、実際に発話意図を付与することを行なった。本稿ではその内容を報告する。

エンタメ小説の会話に発話意図を付与する際に特に留意した 4 点を以下に示す。

1. 発話意図をどのくらい細かく区別するか
2. 語用論的な意図をどう扱うか
3. 感情をどう扱うか
4. 発話意図がひとつに定まらない場合はどのように記述するか

2. 発話文生成における『発話文表現文型辞書』の役割

2.1 発話文表現文型辞書

『発話文表現文型辞書』は、小説の発話文生成を念頭に、「ある目的(発話意図)で発話する時、ある話し方で表すならば、この文型を使う」という情報を提供する辞書である。

発話文表現文型辞書の各エントリは、主に下記の 3 種類の情報で構成されている。

発話意図＝発話の目的
 表現文型＝発話意図を表現する際に使用する言語形式
 話し方の特徴ベクトル (6 軸 20 次元)

2.2 発話文生成

話文生成システム (刀山 2016) (木村 2018) (米田 2022) における『発話文表現文型辞書』の役割は、「発話意図」と「話し方の特徴ベクトル」が入力されると、それに適合した「表現文型」を提案することである。以下に例を示す。「話し方の特徴ベクトル」は、実際は6軸20次元のベクトルであるが、ここでは便宜上文字で表現する。

入力		提案
発話意図＝非難-実行	→	表現文型＝Vt なんて！
話し方の特徴＝女性的、強い、下扱い		

発話文生成システムは、受け取った表現文型 (Vt なんて!) と発話内容 (チョコレートを食べる) を HaoriBricks3 (佐藤 2020) で表層化し、発話文「チョコレートを食べたなんて!」を出力する。

2.3 発話文変換

『発話文表現文型辞書』は、「話し方の特徴」を変更すると、別の「表現文型」を提案する。以下に例を示す。

入力		提案
発話意図＝非難-実行	→	表現文型＝Vb やがって
話し方の特徴＝男性的 2、強い、 下扱い、くだけ、ぞんざい		

この場合、発話文変換システム (木村 2019) は、発話文「チョコレートを食べたなんて!」を「チョコレートを食べやがって」に変換する。

『発話文表現文型辞書』は、「発話意図」を変更すると、別の「表現文型」を提案する。以下に例を示す。

入力		提案
発話意図＝禁止	→	表現文型＝Vb んじゃねえぞ
話し方の特徴＝男性的 2、強い、 下扱い、くだけ、ぞんざい		

この場合、発話文変換システムは、発話文「チョコレートを食べやがって」を「チョコレートを食べんじゃねえぞ」に変換する。

3. 発話文表現文型辞書の発話意図の課題

本辞書の発話意図を選択し体系化するために参考にした資料とその内容を以下に示す。下線資料は、この度の発話意図見直しのために新たに参照したものである。

- a. 日本語学：表現意図およびそれに応ずる文表現の分類 (国立国語研究所 1960, 1963)
 モダリティに関する項目 (日本語記述文法研究会 2003) (益岡 1992)
 談話に関する項目 (日本語記述文法研究会 2009)

評価のモダリティ (高梨 2010)
 発話機能 (山岡 2008)
 伝達的な機能による発話の分類 (宮島 1995)

- b. 日本語教育： 発話内行為およびモダリティ (グループ・ジャマシイ 1998)
 表現意図 (市川 2019)
- c. 談話分析： 発話単位タグ (荒木 1999)
 談話行為タグ (伝 2015) (居關 2018)
 発話意図タグ (入江 2005)
 発話機能の連鎖組織 (筒井 2012)

いずれも有益な情報である。しかしながら、小説発話文を生成する際に発話意図として使用しにくいものも含まれている。例えば、モダリティ (またはムード) を表す「確言」「当為」「概言」「比況」や「示唆」など日常使わない専門用語はラベルとして使いにくい。また、主に談話分析で用いられる「情報要求」「情報提供」などの機能を表すラベルや、「評価」「確認」などやや抽象的で、それ単体で発話意図として用いるのはやや困難である。

現在の辞書の発話意図のリストを表1に示す。現状は、上記のように扱いづらいものは保留にしているため、本当に必要な発話意図が欠落している可能性がある。加えて、「推量」「断定」など情報の確からしさのモダリティに関しては、既に多くの表現文型を本辞書に収録しているが、小説発話文生成時に用いる発話意図としては問題がある。

表1 発話文表現文型辞書のグループ別発話意図

グループ	発話意図	発話意図の数	エントリの数
行為要求	命令、依頼-実行、依頼-非実行、勧誘、忠告、勧告、許可、禁止、勧め	9	378
意向伺い	申し出	1	26
態度表明	褒める、貶す、非難-過失、非難-実行、非難-非実行	5	212
意志表示	願望-実行、願望-非実行、願望-事態、願望-他者の行為、後悔-実行、後悔-非実行、決心-実行、決心-非実行	8	261
感情表出	—	0	0
情報伝達	説明-物事、説明-理由、説明-原因、説明-結論、説明-換言、伝聞	6	1,132
情報の確からしさ	推量-だろう、推量-のだろう、推量-ようだ、推量-みたいだ、推量-らしい、推量-かもしれない、推量-と思う、断定-言い切り、断定-にちがいない、確信-はずだ、否定的確信-はずがない、否定的確信-わけがない、不確定	13	5,714
質問	確認-肯否要求、確認-未知情報要求、確認-念押し、許可要求	4	103
応答	肯定、承諾-依頼、依頼-勧誘、受諾-申し出、否定、断り-依頼、断り-勧誘、断り-申し出、不明、未定	10	172
感動詞	よびかけ、別れ、感謝、謝罪	4	57
		60	8,055

現在の発話意図の主な課題を以下にまとめる。

- 全体の課題
小説発話文生成のための発話意図リストを作成するために
 1. 保留中の発話意図を解決する
 2. 足りない発話意図を見つける

- グループ別の課題
情報の確らしさ（モダリティ）：確らしさは発話意図ではない。表現文型エントリーを適切な発話意図に振り分ける。グループの見直しも必要である。
質問：質問の中心的形式は、yes-no 質問と wh 質問であるが、辞書の表現文型が文末述語（述部）を対象としているので、とりあえず保留にしてきた。
感情表出：どのような発話意図が必要か。現状はゼロ。
情報伝達：既存の発話意図の見直し。追加でどのような発話意図が必要か。

4. エンタメ小説の発話文への発話意図付与

前節で示した発話意図の課題に解決の糸口をさぐるために、実際の小説での会話文一つ一つに対して、発話意図を付与することを行なった。

4.1 対象データ

分析の対象としては、エンタメ小説を選択した。その理由を以下に示す。

1. プロットが明解で、会話の状況や場面設定がわかりやすい
2. 登場人物のキャラクタがはっきりしている
3. 会話が多い。作品がシリーズ化されることも多いので、特定の人物の会話文を大量に収集できる

これらのことは、今後、特定の話者の表現文型を多く収集し、将来、発話意図を自動で付与するという目標に役立つかもしれないと考えてのことである。

具体的には、以下の3小説の会話文を抽出して利用した。

結城浩『数学ガール秘密のノート：微分を追いかけて』（以下、【結城】と略記）

有川浩『図書館戦争』（【有川】）

三浦しをん『風が強く吹いている』（【三浦】）

4.2 方法

3冊の小説の巻頭から各 100 発話文を抽出し、それぞれに文型と発話意図を付与した。文型は、述語を含む文には HB3 の仕様の接続形式と主要部と末尾部を記述した。まず、発話意図を付与したのち、発話意図を 3 種類に区分した。

文型辞書の発話意図を付与

新規の発話意図を付与

不明

次に、不明に区分した発話文を対象に、その理由を分析しながら、複数の発話意図を付与するか、発話意図を付与しないかを決定した。すなわち、最終的に以下の 4 種類に区分した。

1. 辞書の発話意図を付与
2. 新規の発話意図を付与
3. 複数の発話意図を付与
4. 付与しない

5. 結果

表2に、エンタメ小説3作品の各100発話文、合計300発話文に対する発話意図の付与結果を集計した表を示す。種類の計は3作品の重複を除いた数を()内に示す。

表2 小説作品の各発話文に対する発話意図の付与結果

	作品	発話 文数	種類	発話意図 (回数)
1 辞書の 発話意図 を付与	【結城】	15	12	説明-(6)、依頼-(3)、確認-(2)、肯定-(2)、感謝、命令
	【有川】	20	13	説明-(5)、依頼-(2)、確認(2)、禁止(2)、忠告(2)、命令(2) 他
	【三浦】	30	21	よびかけ(4)、説明-(4)、命令(3)、確認-(3)、依頼-(2)、否定
	計	81	(30)	(2) 他
2 新規の 発話意図 を付与	【結城】	63	32	説明-(20)、wh 質問-(7)、了解-(5)、聞き返し(5)、意見-(3)、助言(3)、yes-no 質問(2)、指摘(2)、申し出-説明(2)、予告(2) 他
	【有川】	43	23	意見-(12)、評価-(5)、主張-(4)、報告-(2)、説明-(2)、同意要求(2) 他
	【三浦】	53	21	説明-(13)、wh 質問-(9)、紹介(9)、挨拶(4)、評価(4)、yes-no 質問(3) 他
計	154	(62)		
3 複数の 発話意図 を付与	【結城】	20	17	意見/不満(2)、説明-理由/助言(2)、了解/不本意(2)、wh 質問-何/反感、wh 質問/意見要求、wh 質問/説明要求 他
	【有川】	32	29	気遣い/嫌味(2)、評価/開き直り(2) 他
	【三浦】	15	12	説明-事情/言い訳(4)、wh 質問/怒り、wh 質問/非難 他
計	67	(57)		
4 付与し ない	【結城】	2		
	【有川】	5		
	【三浦】	2		
	計	9		
作品別 計	【結城】	100	62	
	【有川】	100	67	
	【三浦】	100	35	
合計		300	(151)	

5.1 発話意図の階層化

発話意図付与では、(-)を使用した階層化ラベルを用いた。発話意図の階層化について説明する。(-)を用いた発話意図の階層化は、発話意図を細分化するために、発話文表現文型辞書の初期から採用していた方法である。(-)を挟んだ左右の関係は様々ある。以下に例を示す。

1. 説明-理由：理由を説明する。例：「等しい速度で運動してるから」【結城】
2. 依頼-非実行：ある行為をしないように依頼する。例：「まあ、そう言わないで」【有川】
3. wh 質問-場所：質問の対象が場所。例：「どこで野宿してたんだ？」【三浦】
4. 肯定-確認：確認に対して肯定の応答をする。例：「できるよね？」に対して「できるよ、もちろん」【結城】

このように階層化を採用した理由は、その発話意図を表現する文型を特定するために、意図を細分化して、意味を限定する必要があったからである。例えば、「説明」の下位分析には「理由」の他に、「結果」、「前提」、「事情」、「状況」、「物事」などがあり、それぞれ表現する文型は異なる。

5.2 辞書の発話意図を付与

表2のグループ1は、既に辞書に収録されている発話意図を付与した発話文である。対象となった発話文数と発話意図の種類、および、その内訳を示している。なお、内訳では、発話意図の最上位の階層が同一のものは、まとめて集計している。

5.3 新規の発話意図を付与

表2のグループ2は、辞書に収録されていない新しい発話意図を付与した発話文である。新しい発話意図は、合計で62種類となった。

5.4 複数の発話意図を付与

表2のグループ3は、複数の発話意図を (I) で併記して付与した発話文である。付与した発話意図は、既存の発話意図か新しい発話意図かは問わない。どのような場合に複数の発話意図を付与するかを以下に示す。

1. 語用論的な意図を併記
例：wh 質問/非難「なんなんだ、その穴は」【三浦】
例：wh 質問/説明要求「どゆこと？」【結城】
2. 感情を併記
例：yes-no 質問/警戒「俺を脅そうっての？」【三浦】
感情のラベルは、基本的に (Ackerman 2020) を参照している。
3. 発話意図がひとつに定まらない場合、全て併記
例：拒絶-説明/取り消し-依頼「やっぱり説明はいいや」【結城】

これらのうち、前者の2つは、発話文の文型を定めるために併記が不可欠である。

5.5 発話意図を付与しない

表2のグループ4は、発話意図を付与しなかった発話文の数を示している。どのような場合に付与しなかったかを以下に示す。

1. 発話文が倒置 例：「あたしだけよこんな腕立て食らってんの！」【有川】
2. 述語がない 例：「…ハイジさん、ここが？」【三浦】
3. 記号のみの文 例：「？」【結城】

6. 考察

6.1 3作品に付与した発話意図の特徴

対象とした3作品のそれぞれの物語の特徴と、その作品に特徴的な発話意図を表3に示す。

表3 物語の特徴と特徴的な発話意図

作品	物語の特徴	特徴的な発話意図と「発話文」
【結城】	高校生の僕が中学生の従姉妹のユーリに数学を教える	<ul style="list-style-type: none"> ・説明-前提 「たとえば、直線の上を一つの点が動くとしよう」「つまり、点Pの位置xが1から4に変わったとする」 ・説明-手順 「ここまでで、点Pとその位置の話が終わり」「次は時刻だ」 ・意見/確信-はずだ 「『なーんだ、それだけのことなの?』って言うはずだよ」 ・意見/断定-言い切り 「ユーリにもちゃんとわかる」
【有川】	新人図書隊員の女の子が、厳しい教官や仲間たちと一緒に図書館の敵と戦う	<ul style="list-style-type: none"> ・指摘/怒り 「しかもコケたのゴールの後じゃないよっ!」 ・意見 「絞られてへこたれる程度なら辞めりゃいいんだ」「女子から採るとすれば彼女しかいないと思うけど」
【三浦】	箱根駅伝にチャレンジする大学生たちの話	<ul style="list-style-type: none"> ・説明-理由/言い訳 「だって、これしかないんだもん」 ・非難-過失 「気をつけろって言っただろ」 ・説明-事情/言い訳 「気をつけてはいたんだ」 ・wh 質問/怒り 「ホントにあんた、なんなんですか?」 ・wh 質問/非難 「なんなんだ、その穴は」

【結城】に特徴的な発話意図は、「説明」と「意見」のバリエーションである。これらには、バリエーションがあり、適切な発話文を生成するためには、細分化が必要である。発話意図の階層化は、このための手段を提供する。

【有川】に特徴的な発話意図は、感情を伴う強めの「意見」交換である。感情は、発話文の文型に影響するため、感情の併記が不可欠である。

【三浦】に特徴的な発話意図は、「wh 質問」や語用論的な意図を伴うものである。「wh 質問」に伴う感情や語用論的意図も、併記が必要である。

6.2 発話意図の体系化に向けて

今回の調査により、不足している発話意図が62種類明らかになった。これらの多くは保留にしていたものが多いが、まったく想定していなかった発話意図もある。小説の発話文に実際に発話意図を付与してみることは、発話意図の体系の妥当性を評価するためには不可欠である。

それぞれの小説で付与した発話意図の分布には、大きな偏りがあった。発話意図の体系は、拡張可能とすることが望ましく、階層化が不可欠である。階層的なラベル体系を用いることにより、必要に応じて発話意図の細分化が可能となる。

発話文によっては、発話意図をひとつだけ付与することは現実的ではなく、複数の発話意図を併記することが必要となる。特に、語用論的意味や感情は、併記することにより、適切な文型を選択できるようになると考えられる。

7. おわりに

本稿では、『発話文表現文型辞書』の「発話意図」の抜本的な見直しのために実施した、小説の発話文への発話意図付与について述べた。今後は、さらに、対象とする発話文を拡大し、辞書の「発話意図」の改訂案の策定を行う予定である。

エンタメ小説の会話文への発話意図付与は、付与するラベルの体系が定まれば、それほど難しくなくという感触を得ている。考えられる理由2点を以下に示す。

1. 小説の会話文の発話意図は、話者の発話意図というよりも、話者の口を借りた小説の筆者の意図なので、ストーリー展開がわかりやすいエンタメ小説は、発話意図もわかりやすい。
2. 小説の会話文の発話意図は、音声や表情などのマルチモーダルに頼ることができないので、熟練の作家ほど、文字情報だけで誤解なく意図が伝わるように書いている。

以上の理由から、適切な発話意図が付与された小説発話文 DB が実現すれば、小説発話文の生成以外にも、日本語教育や会話生成にも、役に立つ言語資源になることが期待できる。

謝 辞

本研究は、JSPS 科研費 JP18H03285、JP21H03497 の助成を受けたものです。

文 献

- 夏目和子・刀山将大・佐藤理史 (2017) 「発話文自動生成のための日本語表現文型辞書の作成」言語資源活用ワークショップ 2016 表論文集, pp.126-136.
- 夏目和子・佐藤理史 (2019) 「発話文表現文型辞書の設計と編纂」言語資源活用ワークショップ 2019 発表論文集, pp.295-312.
- 夏目和子・佐藤理史 (2022) 「話し言葉生成のための発話文表現文型辞書」言語処理学会第 28 回年次大会発表論文集, pp.1612-1617.
- 刀山将大・夏目和子・佐藤理史他 (2017) 「話者の特徴を反映した発話文生成器の作成」言語処理学会第 23 回年次大会発表論文集, pp.28-31.
- 木村遼・夏目和子・佐藤理史・松崎拓也 (2018) 「発話表現文型辞書を利用した多様な発話文生成機構」第 32 回人工知能学会全国大会, 2E2-02.
- 米田智美・佐藤理史・夏目和子他 (2022) 「話者の性格を反映した発話文の生成」言語処理学会第 28 回年次大会発表論文集, pp.1744-1748.
- 佐藤理史 (2020) 「HaoriBricks3: 日本語文を合成するためのドメイン特化言語」自然言語処理, 27:2, pp.411-444.
- 木村遼 (2019) 「表現文型を用いた発話文の解析と生成」名古屋大学大学院工学研究科修士論文 (未公開)
- 国立国語研究所編 (1960, 1963) 『話しことばの文型』国立国語研究所.

- 益岡隆志・田窪行規 (1992) 『基礎日本語文法』 改訂版, くろしお出版.
- 日本語記述文法研究会 (2003) 『現代日本語文法』 4 モダリティ, くろしお出版.
- 日本語記述文法研究会 (2009) 『現代日本語文法』 7 談話・待遇表現, くろしお出版.
- 高梨信乃 (2010) 『評価のモダリティ: 現代日本語における記述的研究』 くろしお出版.
- 山岡政紀 (2008) 『発話機能論』 くろしお出版.
- 宮島達夫・仁田義雄編 (1995) 『日本語類義表現の文法』 上: 単文編, くろしお出版.
- グループ・ジャマシイ編著 (1998) 『教師と学習者のための日本語文型辞典』 くろしお出版.
- 市川保子 (2019) 『日本語類義表現と使い方のポイント: 表現意図から考える』 スリーエーネットワーク.
- 荒木雅弘・伊藤敏彦・熊谷智子・石橋雅人 (1999) 「発話単位タグ標準化案の作成」 人工知能誌, 14:2, pp. 251-260.
- 伝康晴 (2015) 「対話への情報付与」 小磯花絵 (編), 『講座日本語コーパス3: 話しことばコーパス-設計と構築-』, 朝倉書店, pp.101-130.
- 居關友里子・門田圭佑・伝康晴 (2018) 「『日本語日常会話コーパス』への談話行為アノテーションの試み: タグ選択が困難な事例に焦点を当てて」 言語資源活用ワークショップ2018 発表論文集, pp.47-56.
- 入江友紀・松原茂樹・河口信夫・山口由紀子・稲垣康善 (2005) 「音声対話コーパスにおける発話意図タグの設計と評価」 電子情報通信学会論文誌 D, 88:10, pp. 2169-2173.
- 筒井佐代 (2012) 『雑談の構造分析』 くろしお出版.
- 結城浩 (2015) 『数学ガールの秘密のノート: 微分を追いかけて』 SBクリエイティブ.
- 有川浩 (2011) 『図書館戦争』 角川文庫.
- 三浦しをん (2009) 『風が強く吹いている』 新潮文庫.
- Angela Ackerman, Becca Puglisi・滝本杏奈 (2020) 『感情類語辞典』 増補改訂版, フィルムアート社.

幼児と保護者によるごっこ遊びの相互行為 —日常場面に関する知識の利用に着目して—

居關 友里子 (国立国語研究所) †

小磯 花絵 (国立国語研究所)

Pretend Play Interactions Between Caregiver and Child

Yuriko Iseki (National Institute for Japanese Language and Linguistics)

Hanae Koiso (National Institute for Japanese Language and Linguistics)

要旨

本研究では、国立国語研究所で現在構築中の『子ども版日本語日常会話コーパス (仮)』に格納予定である幼児1名(2歳6ヶ月~3歳1ヶ月時)と保護者の会話データを使用し、ここで行われていた日常場面をテーマとしたごっこ遊びのやり取りについて記述を行った。着目したのは、模倣される場面に関する日常的な知識がどのようにごっこ遊びのやり取りの中に出現し、利用されているのかについてである。会話データからは当該場面や特定の役割に紐付いた挨拶表現や定型表現、典型的な振る舞いやその連鎖が文脈の共有や役割の交渉、場面の切り替えの表示などに利用され、やり取りの足場となっている様子が観察された。馴染みのある場面からそうでない場面まで、子どもが現実の中で「当該場面らしさ」を見出したこれらの表現・振る舞いに関する知識を試行、調整する場としてごっこ遊びは機能していると考えられる。

1. はじめに

ままごとやおみせやさんごっこなどの「ごっこ遊び」は、子どもが好んで行う遊びの一つである。ごっこ遊びは想像遊びとされるが、遊びの素材、イメージ、テーマは具体的・現実的であり、日常の生活経験を反映しているとされる(明神, 2005)。本研究では、この子どもによる日常の経験や知識がごっこ遊びの中にどのように現れ、利用されているのかに注目する。現在、国立国語研究所が構築している『子ども版日本語日常会話コーパス (仮)』に格納予定である幼児1名(2歳6ヶ月~3歳1ヶ月時)のデータ内に生じていた、保護者とごっこ遊びを行った際のやり取りについて収録された音声・映像データを観察、記述し、ごっこ遊びにおいて日常場面に関する知識がどのように用いられているのかについて考察を行う。

2. 先行研究

「ごっこ遊び」の定義については各研究に幅がある。ここに含まれ得る要素としては、例えば大戸(1997)は「見立て」、「役割」、「話の筋立て」、あるいはガーヴェイ(1980)は「役割」「物」「動きのプラン」「状況」といったものを挙げている。また、ごっこ遊びを「子どもが何か物事のまねをして、いっしょに遊ぶこと」とであると捉える明神(2005)は、子ども時代の遊びのなかで、想像力がもっとも豊かに、活発に使用される遊びであり、想像力そのものであると述べている。松原ほか(2022)は、このごっこ遊びに見られる模倣が目に見えないものを再現するものであり、認知的にも発達した状態でなければ行えないこと、加えて、

† iseki@ninjal.ac.jp

この模倣が学びの一形態でもあることを指摘している。そして、実生活での学びにおいては観察学習が重要な位置を占めるとされ（松原ほか, 2022）、本研究が注目する、日常生活の中で経験する様々な場면을扱ったごっこ遊びは、この学習過程・成果を反映していると考えられる。

3. データ

ごっこ遊びの分析については実験的手法のほかに、実際の保育場面の参与観察などの手法を通して多くの知見が得られているが、本研究では実際に行われた会話の音声・映像データをもとに、やり取りの過程について詳細に見ていく。観察対象として使用したのは、国立国語研究所で現在構築中の『子ども版日本語日常会話コーパス（仮）』（小磯ほか, 2022）に格納予定のデータである。このコーパスは調査協力世帯の子どもが、両親、祖父母、友人たちなどといった子どもを取り巻く人々との間で交わした日常でのやり取りを、一定期間継続して録音・録画したデータを扱ったものである。このうち協力世帯の子どもの一人であるユウ（Y001）の日常で行っている相互行為に本研究は注目した。彼女が2歳6ヶ月から3歳1ヶ月までの間に収録された11時間25分の音声・映像データの中に生じていた、ユウと保護者（両親、祖母）による日常場面を模倣の対象としたごっこ遊びを抽出し、日常場面に関する知識がどのように利用されているのかについて観察、記述を行う。

子どもの月齢とごっこ遊びの関連については、見立て遊びやふり遊びをはじめのが2歳頃、ごっこ遊びの原型を示しはじめるのが3歳頃、4歳頃には明確なごっこ遊びを盛んに行うようになるとされる（小山, 1998; 多田ほか, 2009）。個人差はあると考えられるが、以上に照らすと今回扱う月齢である2歳6ヶ月から3歳1ヶ月という期間はごっこ遊びの原型ができるかできないかの時点に位置づけられるだろう。実際に調査対象児であるユウには、自在にはと言えないものの、持てる方略を活用し、やり取りの中で場面や役割の交渉を試みる振る舞いが観察される。以下では意図する結果が得られた場合と、得られなかった場合の両者を取り上げる。

4. ごっこ遊びにおける日常場面に関する知識の利用

まず目立って観察されたのが、当該場面や特定の役割に紐付いた挨拶表現や定型表現の使用である。また、言語表現以外の典型的な振る舞い（行為）についても同様にしばしば用いられている。これらは模倣される対象となる場面らしさを表現する基本的な振る舞いといえる。本節ではこれらがごっこ遊びの中で実際にどのように用いられているのかを、いくつかの事例を取り上げて見ていく。

4.1. 設定の共有と交渉

定型表現の使用を通してごっこ遊びの参加者が行っていることの一つとして、ごっこ遊びの設定の共有と交渉が挙げられる。以下の事例1はごっこ遊びを行いたい子ども（ユウ）が母親にそれを伝えるやり取りである。

[事例1] いらっしゃいませしたい（Y001_003, 2歳6ヶ月）

- 01 ユウ▶ いらっしゃいませしたい。
 02 母: いらっしゃいませするの?じゃお願いします。
 03 ユウ: いらっしゃいませ:。(声を張り上げて)
 04 母: は:い.=
 05 ユウ: =は:い.座ってここ.

- 06 母: はい.
 07 ユウ: いらっちゃいませ.
 08 母: じゃあ:(.)ドーナツください.
 09 ユウ: はい.

ここでユウはいわゆる「お店屋さんごっこ」のことを「いらっちゃいませする」と表現し、この場面設定は母によってすぐに了解される (L02)。続けて子どもは声を張り上げ「いらっちゃいませ」と発話することによって (L03)、この場面が開始したことと店員の役を担うのが自身であること、同時に母が担うのは客であることを提示している。

模倣対象となる場面に紐付いた定型表現、典型的な振る舞いは、より複雑な交渉にも使用されている。次に挙げる事例2はユウと祖母(モモ)によるやり取りである。この事例2ではユウと祖母の互いが担う役割に関する理解が行き違っており、その交渉がなされている。

抜粋以前、ユウと祖母はベビーサークルを家(「ユウちゃんのおうち」)に見立て、サークル内で遊んだりおしゃべりをしたりしていた。そんな中ユウは「お出かけに行く」と宣言し、おもちゃの車に乗ってベビーサークルから出ていく。ユウは加えて、母親や祖母が普段利用している精肉店「ニジイロさん」に行きたいと述べ、これを受けて祖母はベーコンとコロッセのおつかいを頼む。そして、車に乗って移動したユウは店に見立てたテントの前に車を止め、テント内に入っていった。01 行目はこのテント、つまりごっこ遊びにおける店の中からなされた発話である。

抜粋以前のやり取りからは、ユウはお遣いに行く役、つまり客としての役割を担うことが期待される。これに反して抜粋の01 行目、ユウが店に見立てたテント内から発話したのは「お持ち帰りでもいいですか?」という、店員が商品を包む際に用いる定型表現であった。祖母はこの時もベビーサークルの中、つまり家からこの発話を聞いている。なおテントは全体が幕に覆われており、二人は抜粋中、相手の姿を見ずにやり取りを進めている。

[事例2] 並んで (Y001_026, 3歳0ヶ月)

- 01 ユウ▶ はい.(0.6)お持ち帰りでもいいですか?:
 02 祖母: はい.持ち帰りま:す.
 03 (0.8)
 04 ユウ▶ いらっちゃいませ::ってユウちゃんが言う.
 05 祖母: はい.い↑らっちゃいませ:.ユウ[ちゃん_
 06 ユウ: [ねえ.
 07 ユウ: ↑ねえ.モモ・モモちゃんここ[に:,立って?:
 08 祖母: [うん.
 09 (0.6)
 10 祖母: え::ど:↓こ:.[よいしょ.(立ち上がりベビーサークルから出る))
 11 ユウ: [(立って)
 12 (0.4)
 13 ユウ▶ 立って人がいるから並んで?
 14 祖母: あそうなの?並ぶの知ってるんだ.
 15 母: ↑HAHAHA((別室から))
 16 ユウ: 人がいるからな[らんで?
 17 祖母: [あそお?はい.

01 行目の定型表現は店に見立てたテントの中から、対する02 行目はユウの家に立ってたベビーサークルの中からはなされており、二人はごっこ遊びの設定上全く別の空間にいる状況となっている。そのため祖母の反応(L02)はお遣いに行ったユウの問いかけに答えるものとも、店員からの問いかけに答えるものとも聞くことができ、この時点で互いがどのよう

な役として振る舞おうとしているのか、参加者たち自身がよくわかっていない状況である。

この文脈が錯綜した状況で、ユウは役割について明示的な交渉を開始する。ユウは自身の希望する店員役を、店員が用いる定型表現「いらっしやいませ:: (L04)」を「ユウちゃんが言う。(L04)」という表現によって提示しようと試みる。これを受けた祖母は、ユウが例示した定型表現を声色をやや高めに変え、演技がかった声色で発話する。このように祖母はユウの要求に実際に店員役として振る舞ってみせることで応えていることから、ユウが定型表現によって指示しようとしたものが店員役のことであると理解したことがわかる。その一方で、店員役の担い手についてはユウの希望を取り違えており、依然として互いが担う役についての共通理解が得られない状況が続く。

文脈を交渉する次の一手として、ユウは祖母を店に見立てたテントのそばに来るように促し、続けて「立って人がいるから並んで? (L13)」と依頼する。ここでは客という役割に結びついた典型的な振る舞いに関する知識が使用されているといえる。祖母はこの知識を妥当なものとして評価し (L14)、またやり取りを離れた場所から聞いていた母も、この振る舞いがリアルなものであるとして可笑しみをもって反応を示している。そしてこの依頼と同時に互いの役割を受け入れた祖母は、その後ユウの希望に沿ってテントの外に立ってユウの接客を待つ。

利根川 (2022) は「ごっこ遊びとして「医師になる」際に重要な「医師らしさ」とは、医師というポジションにあらわれるのではなく、医師として行為する中にあらわれる」と指摘している。ユウの発話にある「いらっしやいませと言う」「並んで待つ」とは、このような実際に当該場面を観察する中で得た、各役割のそれらしい振る舞いであり、ユウはそれを役割交渉の足場として使用しているといえる。

以上では模倣する対象となる場面に紐付いた定型表現、典型的な振る舞いの知識を利用して、文脈の交渉を行う様子を記述した。この時期のユウには「お店屋さんごっこ」や「店員」などといった、大人やごっこ遊びに熟達した子どもが端的に遊びの場面や役を指定する際に用いる語彙の使用は見られない。しかしながら、このような語彙を用いずとも「いらっしやいませ (する)」「お持ち帰りですか」「並ぶ」といった、実際の買い物で繰り返し聞きしたと思われる出来事をもとに、その場面や役割と結びつく言語表現や振る舞いを見出し、何もない状況から文脈を提示したり、参加者同士の理解が食い違ふといった込み入った状況を切り抜ける様子が観察された。

4.2. 場面の切り替え

子どもの遊びについて鹿嶋 (2004) は、次の発言をどうするかなどを明示的に表明する「明示的メタコミュニケーション」とそうではない「暗示的メタコミュニケーション」の両者を組み合わせて相互行為を展開していくとしている。挨拶表現や定型表現は、後者のタイプとして役のままやり取りの局面を表示することに用いることが可能であり、今回観察したごっこ遊びの中にもこのような使用が多く見られた。

4.1 節で見た「いらっしやいませ」はこのような使用も観察された表現の一つである。次に挙げる事例3は事例2の後に生じたやり取りである。01 行目、03 行目のユウの発話は、買い物場面の開始を表示するキューになっており、かつ店員として客に呼びかける挨拶表現でもあることから、続けて祖母が客として振る舞うことを促すものとなっている。事例3で祖母は一時的にユウのもとを離れており、ごっこ遊びはすぐに開始しなかった。そこでユウは「いらっしやいませ」を繰り返し、開始を促している。

[事例3] いらっしやいませ (Y001_026, 3歳0ヶ月)

01 ユウ▶ いらっ(.)しやいませ::

02 (5.0)

03 ユウ▶ いらっ(.)しゃませ::

04 祖母: は:い.待ってて:.

この他にもベビーサークルに入る際に「お邪魔します」という表現を用いることで、このサークルを家に見立て、家に住む者とそこに来た客のやり取りを開始したり、差し出されたおもちゃの食べ物を食べた後に「ごちそうさま」という表現で食事の終了を表示したりなどといった振る舞いが子どもおよび保護者の発話に観察された。これらの表現は「買い物場面を始める」「家の中の場面である」「おもちゃの食べ物に変化はないが、食べ終わったことにする」といった情報を明示的に述べることなしに、ごっこ遊びの役としての振る舞いを通して参与者同士で場面設定を共有することを可能にしている。そして、共有された設定や局面をもとに、ごっこ遊びは次の展開に進んでいく。

このように、ごっこ遊びでは当該場面のスクリプトが喚起され、挨拶や定型表現がやり取りの局面を表示・操作することに利用されている。このような働きは、挨拶や定型表現ほど決まりきったものに限らず、その場面を形成する一連の行為の流れと強く紐付いた行為によっても担われていると考えられる。事例4はその一例で、母親がユウのリクエストに応え、折り紙でごはん（茶碗に盛られた白米）を作ってやった際のやり取りである。ここでは折り紙遊びの中に短いごっこ遊びが挟み込まれている。冒頭の01行目と02行目は、折り紙が完成した直後であり、親子はごはんの完成を発話で確認し合っている。

[事例4] おいしい (Y001_025, 3歳0ヶ月)

01 ユウ: これごはんじゃない?

02 母: うん.ごはんできたできた.

03 ユウ: ((ごはんを手にとって眺め、折り紙の箸を取り出して食べる真似をする))

04 ユウ▶ おいし:い.((母に顔を向けながら))

05 母: おいし:い.よかった::

06 ((よけてあった折り紙の本を母が手元に引き寄せ、ユウと二人で眺める。ユウが次の作品をリクエストする。))

そして、ユウは母がつくった折り紙のごはんを手にとってしばらく眺めた後、過去に作られた折り紙の箸を無言で取り出し、ごはんを食べる仕草をする(L03)。唐突にユウによって開始された食事のごっこ遊びは、「おいし:い。(L04)」という評価を母に伝えることで母を巻き込み、母はその発話をそのまま繰り返すことによってユウの評価とごっこ遊びへの参与を受け入れる。

折り紙でごはんを作ってもらい、そこに食事という文脈との関連を見出すことで開始した食事ごっこの中で、ユウは作ってもらった食事を食べるだけでなく、さらにその末尾で食事の作り手に評価を伝える、という食事場面の典型的な行為の流れを実践して見せている。そしてこの発話に母はごっこ遊びの区切りを見出し、次の活動への移行を開始している。このような非常に短いごっこ遊びの中にも、ユウが日常場面に対して持つ知識が表示されていおり、それがその後の展開のきっかけとして利用されているといえる。

類似の例が以下の事例5である。子どもに見立てた人形(ハナ)を、母とユウが二人で世話するごっこ遊びが行われている。抜粋では母が人形の歯を磨いてやっており、母は「ハナちゃんおりこう::.(L04)」と人形を褒めることで歯磨きの終了を形成している。それを見守っていたユウも同じく、母とは別の表現で人形を褒める(「がんばったね.(L09)」)。「歯磨きを行い、最後に褒める」という流れが、この場面を形成する一連の行為の流れとして母親だけでなく子どもにも理解されており、このような共通理解がごっこ遊びの局面を表示すること、またそれを受け取ることに利用可能になっていると考えられる。

[事例5] がんばったね (Y001_029, 3歳1ヶ月)

- 01 ユウ: イ:もやる.
 02 母: ジャイ:の口してくださいイ::ごしごしごしごし.
 03 ユウ: ごっちごち[ごち.
 04 母: [はいハナちゃんおりこう::.
 05 ユウ: は::い.
 06 母: ユウちゃんもやる?(0.8)あとで.
 07 ユウ: うん.
 08 母: ((ユウに一声かけた後その場を離れる))
 09 ユウ:▶ ((人形を抱っこして))° がんばったね°

以上のように、当該場面に結びついた挨拶表現や定型表現、また提供された食事を食べて評価を述べたり、子どもの歯を磨いた後褒めたりといった、その場면을形成する一連の行為の流れといったものに対する志向が、保護者だけでなく子どもの振る舞いにもしばしば観察された。これらの振る舞いを用いることで模倣する場面らしさを表現し、同時に当該場面の中のごっこ遊びの局面を相手に、また自身にも示すことで、ごっこ遊びの展開を形作っているといえる。

5. 日常場面に関する知識を試す場としてのごっこ遊び

ごっこ遊びでは、日常の中で見聞きした様々な場面が取り入れられ得ると考えられる。模される対象となる場面に関して、参与者たちの共有知識がどのようなものなのかは具体的なごっこ遊びの進行の中で明らかになり、実際にどのような展開をたどるのかについてもこの中で参与者たちの交渉や調整を通して都度決定されていく。今回扱ったデータには「お店屋さんごっこ」のように名付けも含めてごっこ遊びに定着している場面や子どものお世話（歯磨き）をしてやるなど、名付けにくいけれどもしばしば題材として扱われる場面だけでなく、ごっこ遊びとしては馴染みのない、しかし日常でしばしば経験されている場面についても、子どもが自身の持っている知識や経験をもとに、ごっこ遊びの対象に取り入れようと試みる様子が観察された。

以下に挙げる事例6は、収録に用いるビデオカメラの存在をきっかけに、ホームビデオを撮る場면을模してごっこ遊びをしようとユウが父親を誘っている場面である。ごっこ遊びで扱う場面としてはややイメージしにくい場面ではあるが、ユウは4節の事例と同様、ホームビデオで行われる典型的な振る舞いを遊び相手である父に繰り返し提示し、この場면을共有しようと試みている。

抜粋前でユウは収録用ビデオカメラの背面に回り込み、機体背面についている液晶モニターを見て部屋や自分がどのように写っているのかに興味を示していた。この時両親は仕事の相談を行っており、ユウはこれに割り込んでごっこ遊びを開始しようとしている。

[事例6] たたたたーって来て (Y001_014, 2歳10ヶ月)

- 01 父: だからエクセルは間違[い° だって°.
 02 ユウ: [ねえ[パ:パ:]
 03 母: [う:んだから:]
 04 (0.2)
 05 母: PDFに合わせ[ていいんだよね?]
 06 ユウ:▶ [た た][た た た]た:って.hh ここまで:h座ってっ.
 07 父: [そ:うだよ]
 08 (0.4)

- 09 ユウ▶ ユウちゃんが:(1.0)パ:パおっきくなったねえって>できるか(ら).<
 10 父: ehehehe どうゆうこと?
 11 (2.4)
 12 ユウ: ねえママたたたた:って来てパパ.
 13 (1.0)
 14 父: パパ?
 15 (1.2)
 16 父: 上乗ってよ(上にじゃあ.)
 17 (1.8)
 18 父: ↓これでこようかなじゃあ.((床に寝そべったまま身体の向きを変える))
 19 (0.6)
 20 父: い::よいしょ.どうどうどう↑どうどう.来たよ.((匍匐前進でカメラに近寄る))
 21 (2.4)
 22 ユウ▶ も(う近く).(1.4)(もう近く.)
 23 (0.4)
 24 父: もう一回来んの?
 25 (1.8)
 26 ユウ: えっと_
 27 父: うん.
 28 ユウ▶ ここまで来る.
 29 ((中略5秒 父母で仕事に関するやり取り))
 30 ユウ▶ もういっこ.((「もう一步」前へ来るようにという意味))
 31 父: うん.((父が匍匐前進でカメラに近寄る))

仕事について話している両親に割り込んで父親の注意を獲得したユウは、「たたたたたた:って.hh ここまで:h 座ってっ. (L06)」という発話とともに、ビデオカメラに向かって歩み寄り、画面の中心に座ってカメラを見つめる動きを実演して見せることで父親に動きの指示を行う。続けて「ユウちゃんが:(1.0)パ:パおっきくなったねえって>できるか(ら).<という発話によって、ビデオカメラに歩み寄るのが子どもであり、この動画はその成長を記録し、動画を見る者に見せるためのものであるという場面設定を提示する。ここでは同時に、父親が子どもの役、ユウがビデオを見る(撮る)大人の役を担うという配役も示されている。ユウはそこで行われる一連の振る舞いを提示することによって、ホームビデオという言葉を用いずにごっこ遊びの文脈を共有しようと試みている。これを受けた父親はユウの指示がどのような意図でなされたのかについて理解できないことを表示し、その難解さを笑っている (L10)。

06 行目からなされている指示はごっこ遊びをするという文脈のないまま唐突に行われており、またホームビデオを撮る場面とその振る舞いというのもごっこ遊びの対象としてはやや喚起されにくいものであるといえるだろう。その後もユウは同様の指示を繰り返したり (L12)、移動後の位置など、より具体的な振る舞いを指定するなど試みる (L22,28,30)。対する父親はその指示に従って振る舞うものの、匍匐前進でカメラ側に近づいたり、何者かが迫ってくることを表現するような効果音をつけていたり (L20)、結局ホームビデオというユウの意図するごっこ遊びの文脈は共有されずにここでのやり取りは終了することになった。

以上では、データ収録に用いていたビデオカメラをホームビデオを撮るカメラに見立て、そこで行われる典型的な振る舞いに関する知識をもとにごっこ遊びに相手を誘ったものの、その文脈の共有には失敗した例を見た。ここで子どもは4節で見た事例と同様に、当該出来事について持つ知識を具体的な行動の指示として提示し、模倣する対象場面を共有しよう

と試みている。これが何の真似なのか、何ごっこと呼ばれる遊びなのかという説明はできなくとも、この場면을構成する振る舞いをごっこ遊びに取り入れようとしており、ここにも「ホームビデオらしさ」に対するユウの理解が示されているといえる。

ごっこ遊びでの扱いやすさといった基準があれば、ホームビデオを収録する場面は上級者向けのものであろう。しかし、子どもはこのような基準でごっこ遊びの対象場面を選んでいるとは考えにくく、扱いやすい場面、扱いにくい場面を区別せず、子どもたちの日常で得た知識や経験すべてがごっこ遊びの素材となり得ることが示唆される。ごっこ遊びを含む象徴遊びは子どもの生育の目安になるだけでなく、発達そのものを促すものとしても重要視されており、特にごっこあそびは一貫したルールに従って行動することなどを学ぶ機会となり得ると指摘されている（多田, 2009）。本研究で観察した事例からも、やり取りのその瞬間はごっこ遊びとは気づかない場合も含め、ここには子どもが生活の中で見聞きした知識や経験が織り込まれており、ごっこ遊びという場を通してこれらの理解が試行、調整され、様々な場面に結びついた具体的で適切な振る舞いが学ばれていることが示唆される。

6. おわりに

本研究では幼児と保護者間で行われた、日常場면을模倣の対象としたごっこ遊びの中で、日常場面に関する知識がどのように利用されているのかを見てきた。観察対象児は、当該場面や特定の役割に紐付いた挨拶表現や定型表現、典型的な振る舞いに関する情報を足場として、相手参加者とごっこ遊びの設定を共有したり、交渉したりすることを試みていた。同様に、開始や終了といったごっこ遊びの局面を共有するマーカとしても使用されており、これを用いることによってごっこ遊びの展開を促す様子が観察された。今回分析対象として扱った2歳6ヶ月から3歳1ヶ月という時期は、ごっこ遊びの原型が完成する前、あるいは完成して間もない時期であると考えられ、自身の頭で組み立てた想像の状況を口頭で自由自在に表現することにはまだ届かない段階であると予想される。しかしながら、大人が用いるような端的に遊びの場面や役を指定する語彙が使用できないとしても、子どもは先に挙げたような表現を用い、そして保護者がそれを汲み取りつつ交渉を行うことによってごっこ遊びは展開させられていた。やり取りに失敗する場合も含めて、ごっこ遊びという創造的な遊びの場は、日常生活の中で子どもが行っている観察学習の成果を試し、修正、定着させる学びの場として利用されていることが本研究の具体的な振る舞いからも示唆される。

今後は今回多く触れることができなかった保護者側からの働きかけの影響に注目した記述もあわせて行っていきたい。ごっこ遊びにおける親の関わり的重要性については既に多くの指摘がある（石川, 2021 など）。親からの働きかけに関する実験的手法を用いた量的研究（伴・内山, 2015; 2021 など）の興味深い知見は、日常の自然な文脈の中で生じた会話の音声・映像データを用いた質的研究を行うことによってさらに具体化させることができると考えられる。子どもとそれを取り巻く人々の振る舞いについて、日常の生の相互行為を収めた音声・映像を活用し、記述を進めていく予定である。

謝 辞

本研究は国立国語研究所のプロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」および科研費 20H01264, 20K20695, 22K13109 の研究成果を報告したものである。

文 献

- 石川洋子(2021)「保育者のふり遊びへの関わり：0～2歳児に焦点をあてて」『文教大学教育学部紀要』55, pp.111-121.
- 大戸美也子(1997)「ごっこ遊び: pretend play.」岡田正章・千羽喜代子(編)『現代保育用語辞典』pp.155-156. フレーベル館.
- ガーヴェイ, C. (高橋たまき訳) (1980)『「ごっこ」の構造：子どもの遊びの世界』サイエンス社.
- 鹿嶋桃子(2004)「遊びにおける即興的な相互行為とその分析視座.」『北海道大学大学院教育学研究科紀要』93, pp.159-196.
- 小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・藤越・西川賢哉(2022)『子ども版日本語日常会話コーパス』の構築」『言語資源ワークショップ2022 発表論文集』.
- 小山優子(1998)「幼児教育における質的研究の方法論的一試案：幼児のごっこ遊びの事例分析を通して」『保育学研究』36(2), pp.185-192.
- 多田幸子・大田紀子・井上聡子・杉村伸一郎(2009)「3歳児における保育者参加型ごっこ遊び：事例分析を通じた保育者の役割の検討」『幼年教育研究年報』31, pp.47-54.
- 利根川彰博(2022)「幼児のごっこ遊びにおける「役割」についての検討：幼児が「何者かになっている」ということの位置づけに注目して」『秋草学園短期大学紀要』38, pp.113-125.
- 伴碧・内山伊知郎(2021)「単独のふりシグナルは幼児のふりの理解を促すか：ふり場面と現実場面との比較から」『応用心理学研究』47(1), pp.37-46.
- 伴碧・内山伊知郎(2015)「大人によるふりシグナルの提示は子どものふり行動を促すか?」『心理学研究』86, pp.333-339.
- 松原乃理子・大滝茜・織壁佐和子・富田貴代・深沢佐恵香・森末一代・請川滋大(2022)「「ごっこ遊び」研究の傾向：保育実践を対象とした調査に着目して」『日本女子大学紀要 家政学部』69, pp.1-12.
- 明神もと子(2005)「幼児のごっこ遊びの想像力について」『釧路論集:北海道教育大学釧路校研究紀要』37, pp.143-150.

「一周回って」の意味・用法をめぐって : Twitter の投稿データを言語資源として

朴 秀娟 (神戸大学)

于 一楽 (滋賀大学)

On the Meaning and Usage of *issyuu mawatte* : using Twitter as a Language Resource

Sooyun Park (Kobe University)

Yile Yu (Shiga University)

要旨

本研究では、「一周回って」の意味・用法を明らかにし、それらの意味・用法間における関連性を示す。「一周回って」は、本来、「一周回るといふ動作を行って」ということを表すが、それ以外にも、「いろいろと試した結果／時間の経過とともに（原点に戻る）」、「時間の経過とともに（考え方・見え方などが変わる）」、「異なる視点から捉えると（考え方・見え方などが変わる）」、「(Xは)異なる視点から捉えようとしても（＝どう考えても）(Xだ)」のような意味・用法もあることを指摘する。また、「一周回って」が用いられる文の構文的・意味的特徴に基づき、上記の意味・用法間の関連性を示す。なお、本研究で用いたデータは Twitter の投稿データから収集したものであり、言語資源としての Twitter の有用性についても述べる。

1. はじめに

「一周回って」は、「回る」という動詞の中止形に、その動作の量を表す「一周」が組み合わさったもので、後には、通常、動詞述語が続く。「回る」という動作は、後続する動作との時間関係において、継起関係であることもあれば(例1)、同時関係であることもあるが(例2)¹、いずれにしても、回るという動作を実際に一周行うということに変わりはない。

- (1) 運動場を一周回って帰ってきた。
- (2) くりりと一周回って衣装を見せてくれた。

しかし、近年では、「一周回って」が次のように用いられている例が見られる。例1、2とは異なり、「一周回って」に後続する述語が必ずしも動詞とは限らず(例3~5)、一周回るといふ動作が実際に行われているとも言い難い。

- (3) ポテトチップスは一周回ってやっぱりうす塩味が一番だ。
- (4) 疲れすぎて一周回って眠くない。

¹ 動詞の中止形がもつ意味の詳細については、仁田(1995)、日本語記述文法研究会(編)(2008)などを参照されたい。

- (5) ネガティブは一周回ってポジティブだ。
- (6) すごすぎて一周回ってすごいとしか言えない。

例 3～6 のように用いられる「一周回って」は、新語や若者ことばとして扱われている。『現代用語の基礎知識』（自由国民社編）では、2019年版の「時代・流行」カテゴリーのうち、「若者」カテゴリーにおいて、「強調する言い方、評価する言葉」として「一周回って」を取り上げ、次のように説明している²。

一時代（10年ほど）経って、流行が一回りして。「一周回って、いま DA PUMP がダサカッコいい」。

「一周回って」が用いられると、時が経って以前のもものが再度よく見えてきたり流行ったりすることを意味するという解釈である。しかしながら、この定義では、例 3 については説明ができて、例 4～6 については説明ができない。そこで、本研究では、このような、従来の定義では説明しきれない「一周回って」の多様な使われ方に注目し、①「一周回って」の意味・用法の全体像を明らかにするとともに、②それらの意味・用法間の関連性について考察を加える。

本稿の構成は以下のとおりである。まず 2 節では、考察に用いたデータについて説明する。続く 3 節では、考察の結果明らかになった「一周回って」の意味・用法を示し、4 節において、それらの意味・用法間の関連性について述べる。最後の 5 節ではまとめを行い、「一周回って」の意味・用法に関する今後の展望について述べる。

2. 考察に用いたデータ：データ収集源としての Twitter の有用性と限界

本研究では、Twitter（ツイッター）の投稿データから用例を収集した。Twitter は、「インターネットを通じて 140 字以内の『つぶやき』を不特定多数にリアルタイムに発信し、自分で選択した他人の『つぶやき』を受信するサービス」（津田 2009：12）である。2006 年 7 月にサービスが開始した。本研究のデータ収集源として Twitter を用いた理由は以下のとおりである。

まず、Twitter では、いわゆる新語や若者ことばとされているものや、近年の流行語のデータを入手しやすい。「一周回って」は新語や若者ことばとして扱われていることから、若い世代が多く利用し発信をしている Twitter において多くのデータを収集できると考えた。総務省情報通信政策研究所（2021）の『令和 2 年度情報通信メディアの利用時間と情報行動に関する調査』によると、Twitter の全世代の利用率は、ユーザー同士の交流やコミュニケーションを主な目的とするソーシャルメディア系サービス／アプリの中では、Instagram（インスタグラム）と並んで最も高く（42.3%）、なかでも、10 代及び 20 代の利用率が高い（それぞれ 67.6%、79.8%）。実際に、検索したキーワードが Twitter においてどのくらい使われているかをグラフで表示してくれる、『Yahoo! JAPAN』の「リアルタイム検索」サービスで「一周回って」を検索してみると（検索日時：2022 年 6 月 14 日、17 時 40 分）、6 時間の間に 337 件、24 時間の間に 1,410 件、7 日間の間に 9,824 件、30 日間の間に 40,385

² ジャパンナレッジ版 (<https://japanknowledge.com>) を参照した（最終閲覧日：2022 年 8 月 23 日）。なお、「一周回って」の意味・用法全般について体系的に記述、または考察を行っている論考は、管見の限り、見当たらない。

件という結果となった。件数には、すでに投稿されたものを引用して再投稿しているもの（Twitterでは、「リツイート」や「引用ツイート」と呼ばれる）も含まれており、同一のデータが重複してカウントされていることもあるが、それを考慮したとしても、「一周回って」について、十分なデータ数を確保できると考えた。

次に、Twitterでは、従来のコーパスに比べて、より近年のデータを入手することができる。先に述べた『現代用語の基礎知識』（自由国民社編）は毎年刊行されているが、「一周回って」は、2019年版に掲載されており、比較的新しい表現であると考えられる。Twitterでは今もなお投稿が続いており、ユーザーが公開設定にさえしていれば、投稿した内容をリアルタイムで閲覧することができる。したがって、データの収集年代が特定の時期に限定されている従来のコーパスに比べると、Twitterでは、ごく最近になって使われ始めたことばであっても入手が可能であると考えた。実際に、現代日本語の研究に利用されることの多い『現代日本語書き言葉均衡コーパス 中納言版 (BCCWJ)』『国語研日本語ウェブコーパス 中納言版 (NWJC)』『日本語話し言葉コーパス (CSJ)』『日本語日常会話コーパス (CEJC)』『名大会話コーパス (NUCC)』『現日研・職場談話コーパス (CWPC)』を対象に「一周回って」を検索してみたところ³、「一周回って」は、NWJCに6例、CEJCに1例しか出現しなかった⁴。そのため、既存のコーパスでは「一周回って」のデータを収集することが困難であると判断し、Twitterからデータを収集することとした。

一方で、Twitterから収集したデータは、分析をするにあたり、次のような限界もはらんでいる。まず、投稿する際の文字数が140文字以内でなければならないという制限があるため、文脈が読み取りにくいことがある。Twitterの投稿データについては、岡島（2011）でも、「細切れの発言の前後がたどりにくく、用語の意味を考察しようとする際に不便である（p.78）」との指摘が見られるが、本研究で考察を行うのも意味・用法であるため、文脈が読み取れない場合、意味・用法の判断に迷うデータが多く、考察が難航することが予想された。次に、Twitterは2006年7月に開始したサービスであるため、それ以降のデータしか入手することができず、それよりも以前に使われ始めたことばである場合、初出年については検討することができない。意味・用法の判断に迷うデータが多くなる可能性と合わせて考えると、意味・用法ごとの出現数や出現時期の正確な提示が難しく、意味・用法の変化の観点から行う考察は難しいと思われた。このような限界点があるにもかかわらず、データの収集源としてTwitterを用いたのは、Twitterでは「一周回って」のデータを大量に確保できるという長所に重きをおいたからである。データ数が確保できれば、文脈が読み取れず判断に迷うデータが多くても意味・用法の洗い出しは可能であり、また、変化の観点から考察を行うことは難しくても「一周回って」が用いられている文の構文的、意味的な特徴に基づいて、各々の意味・用法間の関連性を示すことは可能であると考えた⁵。

本研究では、Twitter上の「高度な検索」機能を用いてデータを収集している。「高度な検索」は、キーワードに関する条件だけでなく、反応が多かった投稿や特定の日付に投稿されたものなど、条件を指定して検索できる機能である。「一周回って」は、Twitterのサービスが開始した2006年にも見られるが、分析の便宜上、2022年から過去10年間の投稿を対象

³ 文字列検索を用いて、「一周回って」「一周まわって」「いっしゅう回って」「一周周って」「いっしゅう周って」「いっしゅうまわって」「イッシュウマワッテ」の7つを検索した結果である。

⁴ 件数には、重複して現れた例は含まれていない。

⁵ Twitterを利用したデータ収集がもつ有用性や限界については、岡島（2011）、岡田（2013）にも詳しい。そちらも参照されたい。

を絞った。また、期間をなるべく等間隔にすべく、投稿日を毎年5月31日から6月1日までの二日間に指定した⁶。その結果出てきた投稿を上から順に目視で確認を行いながら、再投稿など、重複しているものは除き、各年100例ずつ抽出し、計1000例を対象に考察を行った⁷。

3 「一周回って」の意味・用法

「一周回って」の意味・用法は、構文的・意味的特徴から、次の5つに分けられる。

意味・用法A: 「一周回るといふ動作を行って」

意味・用法B: 「いろいろと試した結果/時間の経過とともに(原点に戻る)」

意味・用法C: 「時間の経過とともに(考え方・見え方などが変わる)」

意味・用法D: 「異なる視点から捉えると(考え方・見え方などが変わる)」

意味・用法E: 「(Xは)異なる視点から捉えようとしても(=どう考えても)(Xだ)」

以下、それぞれについて詳しく述べていく。

3.1 意味・用法A: 「一周回るといふ動作を行って」

「一周回って」は、本来、「一周回る」という実際の動作を行うことを表す。「一周」と「回る」固有の語彙的意味が失われておらず、有情物の物理的な距離の移動を伴う。そのため、実際に回るといふ行為を行う空間がヲ格の名詞として現れていたり(例8)、一周回る様態を修飾する副詞が伴われていたりすることがある(例9)⁸。「一周」と「回る」の間の語彙的な緊密性は弱く、ヲ格の空間名詞や副詞を「一周」と「回る」の間に挿入させても、文が表す意味は変わらない。(以下、各例の末尾に、投稿日及び投稿にアクセスできるURLを示す⁹。)

(7) 個人的に閣下は犬苦手そうだなあとか、そんなイメージあるなあ(陛下は犬好きそう/でもすぐ飽きそう)

近寄ってくるクリーパーに対してもシッシッ！って必死に追い払いそうだし、それでダメなら一周回って寝転がってお腹を見せてゴロニャンすれば逆に逃げて行くんじゃない？

人生はギャンブル...!!!

(2019/6/1、https://twitter.com/Dora_Anko/status/1134592464012976128)

(8) さっきゴミ捨て行って、ゴミを置いたらゴキブリがカサカサカサって逃げた。と思ったら僕の周りを一周回ってゴミ捨て場に戻ったのを見て、ゴキブリもちょっとは何かを考えて生きてるのか？と思った。

⁶ 日付はランダムに選定したものである。

⁷ Twitterには「一周回って」「一周まわって」「いっしゅう回って」「一周周って」「いっしゅう周って」「いっしゅうまわって」「イッシュウマワッテ」の7つがいずれも出現する。考察の便宜上、本研究では、「一周回って」のみに限定して考察を行う。

⁸ ここには転載していないが、例9には、ダンサーがステージを大きく回っている写真も合わせて掲載されていた。

⁹ 2022年8月24日現在アクセスが可能なURLである。ユーザーが投稿を削除したり非公開設定に変更したりした場合、閲覧することはできない。

(2015/6/1、<https://twitter.com/Takamixx/status/605061549493846016>)

- (9) 🐼さんで！めちゃ悩んだ！多分時間置いたらまた変わる気がするけど😓
- 1、エルボー1990
 - 2、トップロック！特に音ハメバチバチに決まってる時はテンション上がる🔗 (動画のやつとか！)
 - 3、技じゃないんだけど、🐼さんのムーブの最初に、大きく一周回って場の空気を一気に🐼さんに持ってくるのが好き！

(2021/6/1、https://twitter.com/kawauso_no_uso/status/1399515793344188416)

3.2 意味・用法 B：「いろいろと試した結果／時間の経過とともに（原点に戻る）」

移動を行う主体が有情物ではなく事柄になると、「一周回って」が用いられる文は、いろいろと試した結果、または、時間の経過とともに、もと(=原点)に戻ったことを表す。『現代用語の基礎知識』(自由国民社編)で示されている定義は、この<意味・用法 B>の「一周回って」にあたる。<意味・用法 B>の場合、<意味・用法 A>とは異なり、「一周回って」に後続する述語は、例 11 (好きだ)、12 (名曲だ)、14 (イオンだ)のように、必ずしも動詞述語とは限らないが¹⁰、動詞述語である場合、「戻る」「戻ってくる」「帰る」「帰ってくる」のような回帰を表す動詞であるのが特徴的である。また、「やっぱり、結局」といった副詞を用いて原点に戻ったことを示す例も多い(例 12、14)。

- (10) (31🐼) 逆転していた昼夜が一周回って戻ってきた感じ

(2015/6/1、https://twitter.com/soo_mei/status/605110511122841600)

例 11、12 のように、個人的な好みなど、話し手自身にまつわる事柄である場合、波線で示すように、話し手自身が時間をかけていろいろと試してみた結果、原点に戻ってきたことを表す。

- (11) (筆者注：香水の話をしている)

なんてゆーんかなー？重い系は好きかも💖色々付けてきたけど一周回ってテッパンやけどブルガリ全般と、クロエと、ヴェルサーチ好きー\(^o^)/たぶん爽やかやのに重いのが好きかも！わかる？w

(2015/6/1、<https://twitter.com/mackyhiiiiip/status/605077639749992449>)

- (12) aiko も Ao も他の好きなアーティスト全てに言える事だけど、このアーティストといえど**みたい曲だとかデビュー曲だとかは、最初はやっぱりすっごくハマって繰り返し聴くんだけど聴きすぎて敢えて語る「好きな曲」からは外れて、でも他のいろんな曲を愛でたあとに、一周回ってやっぱ名曲って思うの。**

(2016/6/1、<https://twitter.com/aeu1019xxx/status/605068817664860160>)

話し手自身にまつわる事柄でない場合は、例 13、14 のように、単なる時間の経過とともに、原点に戻ってきたことを示す。

¹⁰ 例 14 は、「イオンに戻った」と見ることも可能である。

- (13) ……いろいろな所のおさがりの服が家にある。
その中にほんとにずっと前のなんだろう。まだ着れる赤のギンガムシャツがあるんだけど、
前に偶然テレビで今年の流行がレトロとか聞いた。
なんか……
時代が一周回って戻った所を見た気分だ

(2015/6/1、<https://twitter.com/okanoue1tyoume/status/605156742742147072>)

- (14) また今度描いてみたいと思います!!
イオン津店はニチイとジャスコの共同出店→ジャスコが閉店しサティがジャスコの跡地にも出店→一周回って結局イオン という興味深い店舗です!!また今度書いてみようと思います!!

(2019/6/1、<https://twitter.com/jNI2Mu2YZK2eqt7/status/1134595448289648640>)

<意味・用法 B>の「一周回って」が使われる文と、<意味・用法 A>の「一周回って」が使われる文の違いは、主体が有情物か事柄かだけで、主体の移動する先が原点であるという点では共通している。これは、後に述べる<意味・用法 C~E>の「一周回って」とは異なる特徴である。原点は、本来あるべき姿 (例 10)、本来の好み (例 11、12)、一昔前に流行っていたもの (例 13)、最初にあったもの (例 14) など、様々である。この、原点に対する話し手の評価は、例 10、13、14 のように中立的か、例 11、12 のように肯定的である。原点を否定的に捉えている例は見られない。

3.3 意味・用法 C:「時間の経過とともに (考え方・見え方などが変わる)」

事柄が移動する先が原点ではなく他になると、<意味・用法 C>となる。<意味・用法 C>の「一周回って」が用いられる文は、時間の経過とともに、考え方・見え方などが変わってきたことを表す。移動先は、例 15~17 のように、最初の考え方・見え方とは逆のものとなる。例えば、例 15 では、「つまらない」から「楽しい (=つまらなくない)」へ、例 16 では、「理解できない」から「理解できる (かもしれない)」へ、例 17 では、「気味が悪い」から「可愛い (=気味が悪くない)」へと変化している。

- (15) いまフランス語やってるけどつままないって気持ちが一週回ってめちゃくちゃ楽しくなってきた(狂)

(2015/6/1、<https://twitter.com/roooute0/status/605094303090606080>)

- (16) 前までは「哲学ですね～」と言われると「……?哲学って何ですか??哲学の定義を教えてくださいよ話はそれからだ!」ってなっていたけど(なっていない)、最近は一週回って哲学かもしれないな……と思い始めた

(2020/6/1、<https://twitter.com/AOh3OhI0xBaiEAZ/status/1267140395781156864>)

- (17) LINE で見つけたホラースタンプ最高だな
最初気味悪かったけど一周回って可愛く見えてきた

(2021/6/1、https://twitter.com/tom_general1993/status/1399508075724230656)

<意味・用法 C>の「一周回って」が用いられる文は変化を表すため、次の例に示すように、「シテクル」、「スルようになる」、「Nになる」のような、変化を表す述語形式を伴って

いることが多い。いずれの形式も伴っていない例 21 も、「疲れすぎると一周回ってねれなくなるのな」のように言い換えることが可能である。

(18) 一周回って空腹感紛れてきた(しかしお腹すごい音で鳴ってる
(2019/6/1、<https://twitter.com/hinyanm/status/1134516080192266240>)

(19) 【スーツ】

クールビズが浸透してくるにつれて着崩すことに違和感がなくなってきたけれど、一周回ってしっかり着こなすことがカッコよく感じるようになりました。

#スーツ #見た目

(2022/6/1、<https://twitter.com/TomoakiSato10/status/1531781514354368512>)

(20) デジタルの波に便乗して便利なシステムを開発したかと思いきや一周回って物凄い面倒くさいものになってしまったパターン

(2017/6/1、https://twitter.com/mitui_yome/status/869932315647987712)

(21) 疲れすぎると一周回ってねれないのな

(2020/6/1、<https://twitter.com/ryoh80/status/1267229094090924033>)

また、考え方・見え方が変化するのには時間の経過とともに生じるものであり、過去と現在を対比するコンテキストにおいてよく使われる。

(22) なんか昔はいたる絵慣れなかったけど一周回って今は良いように思えるな…

あとは私服を描くときのセンスが…ね…

(2018/6/1、<https://twitter.com/ascyeet0811/status/1002334815012524032>)

(23) 海外の人が考えた日本描写に以前はびくっとしてたけど最近は一周回ってすんなり受け入れてる私。ゲンジが暗闇でもびかびか光るサイバーニンジャなのとかもうたまらないよ

(2016/6/1、<https://twitter.com/twnakki/status/737681568664195073>)

3.4 意味・用法 D:「異なる視点から捉えると (考え方・見え方などが変わる)」

考え方・見え方などの変化を表していても、時間の経過を伴わない、すなわち、同時点における同一の現象に対する考え方・見え方などの変化である場合は、「一周回って」は、「異なる視点から捉えると」という意味になる。例 24 は、1日に2回も川に落ちることについて、バカであるという見方もできるが、「珍しい」行為であるという視点を取ると、天才にも見えるということを表す。例 25 では、61分遅れは、捉え方によっては1分遅れとして捉えることができ、例 26 では、多くのチュートリアルをクリアしていかないといけない状況は、捉え方によってはゲームをしている状況とも捉えられることを示している。「一周回って」がなくなると、そのような意味は出にくく、いずれも、単に「天才」「1分の遅延」「新規ゲーム」と見なしていることになる。

(24) 1日に2回も川に落ちるのは一周回って天才だとおもう

(2016/6/1、<https://twitter.com/suisuisuiki/status/737683651425492992>)

(25) 61分遅れとかいう単語が聞こえたけど、もはやそれは一周回って1分の遅延なのでは

(2016/6/1、<https://twitter.com/hajime0083/status/737790376400982016>)

(26) 久々にログインしたら新しいチュートリアルだらけで一周回って新規ゲームみたいだった www

(2017/6/1、<https://twitter.com/limenikon/status/869939711392890881>)

同一の現象に対する2つの異なる見方を提示していることから、<意味・用法 D>の「一周回って」も、<意味・用法 C>の「一周回って」と同様に、対比のコンテキストで現れやすい(波線部分)。<意味・用法 C>と<意味・用法 D>の違いは、時間の経過の有無である。<意味・用法 D>の「一周回って」は時間の経過を伴わず、1つの現象について、2つの異なる視点から述べているということを表しているに過ぎない。

(27) 長芋の竜田揚げ🍡またもレシピを聴けて有り難い朝。

そして青空がくるかと思いきや、格好いいな選曲で驚き…

メロディと歌詞が何度聴いてもとんでもない曲だと思います😓

途方もない絶望のようにも、一周回って壮大な惚気にも聴こえる曲。

“あたしの泣いた顔はブス それを見る困った顔も格好いいな”

(2020/6/1、https://twitter.com/unagi_g_KIDS17/status/1267243578146738176)

(28) オーディオブックでナポレオンヒルの本買ったなら5千円以上して高いなと思ったけど、22時間も朗読したものを何度も繰り返し聴けるんだから一周回って激安だと気付いた。

(2021/6/1、<https://twitter.com/abazurebitch/status/1399512734203404289>)

3. 5 意味・用法 E:「(Xは)異なる視点から捉えようとしても(=どう考えても)(Xだ)」

<意味・用法 D>から考え方・見え方などの「変化」がなくなると、<意味・用法 E>になる。<意味・用法 E>の「一周回って」は対比のコンテキストでは用いられず、むしろ、当初の考え方・見え方がそのまま維持されることを表すコンテキストにおいて用いられる。<意味・用法 D>の「一周回って」が「異なる視点から捉えると」を表すとすれば、<意味・用法 E>の「一周回って」は「異なる視点から捉えようとしても(=どう考えても)」と言える。限定を表す「しか」や、副詞「やっぱり」を伴うことが多い。

(29) いや、本当にすごいなあ〜

凄いの一言で片付けちゃいけないかもだけど、一周回って凄いなしか出てこないのよね。

もういい言葉が見つかりません😓

ただただ同じ時代に生を受けてる事が、同じ時間を生きていられてる事が、自分の目で見届けられている事が有り難くてしょうがない🙏✨

(2022/6/1、https://twitter.com/lazimolala_BTS/status/1531712586428739585)

(30) #裏トーク

クライアントさんへの感謝しか書けない自分は、目立ちたい自分を抑圧してるとか。もっとメディア掲載を望む!とかせつかくならガイアからの情熱大陸、そこでプロフェッショナル出たいとか、書くべきかとか。悩んだんですけど、一周回って、やっぱり感謝なんすよね。

(2021/6/1、https://twitter.com/hiroki_miyamoto/status/1399495029383565313)

当初の考え方・見え方が維持されるコンテキストにおける「一周回って」は、形容詞と共

起すると、程度を修飾し、程度が大きいことを表すようになる。

(31) オールディーズが好きで。一番「アメリカ～！」っていう感じする上に、50・60年代のは一周回ってホットだと思う。

(2016/6/1、<https://twitter.com/SuzukiSara94/status/737670635032186880>)

(32) 髪切ってピアスバチバチに開けたい欲。

誰かジャニーズでいないっけ。

耳が埋まるくらい開いてる人って一周回って好き。

(2020/6/1、<https://twitter.com/shokotan38/status/1267235222275231745>)

4. 「一周回って」の意味・用法間の関連性

3 節で述べたことに基づいて、「一周回って」の意味・用法の全体像を示すと、以下の表のように整理できる。

表 「一周回って」の意味・用法の全体像

	A	B	C	D	E
意味・用法	一周回る という 動作を行って	いろいろと 試した結果 ・ 時間の経過と ともに	時間の経過と ともに	異なる 視点から 捉えると	異なる 視点から 捉えようと しても
主体	有情物	事柄	事柄	事柄	事柄
動作	原点に戻る	－	－	－	－
思考・視点	－	原点に戻る	逆になる	逆になる	維持
時間の経過	有	有	有	無	無

<意味・用法 A>は、「一周回って」の本来の意味である。有情物が「一周回る」という実際の動作を行い、もと(=原点)に戻ってくる様子を表す。

主体が有情物ではなく事柄になると、「一周回って」は物理的な移動ではなくなり、事柄の原点への回帰(意味・用法 B)や変化(意味・用法 C、意味・用法 D)、見え方・考え方の維持ならびに程度の強調(意味・用法 E)を表すようになる。<意味・用法 A>と同じく、もとに戻ってくることを表すと<意味・用法 B>に、もとの位置ではなく逆になると、<意味・用法 C>、<意味・用法 D>に、もとの位置でも逆でもなく、維持している場合は、<意味・用法 E>となる。<意味・用法 E>は、見方によっては<意味・用法 B>と捉えることもできそうであるが、<意味・用法 B>の場合、「戻る」ためには、ほかの方向へと一度出向く必要があり、また、もとに戻るまでに時間の経過を要するという2点において、<意味・用法 E>とは異なる。後者の時間の経過の有無は、<意味・用法 C>と<意味・用法 D

>を区別する要因でもあり、同じく、思考・視点が逆になることを表している、それが時間の経過を要していれば<意味・用法 C>に、時間の経過を要していなければ<意味・用法 D>となる。

5. おわりに

本研究では、「一周回って」の意味・用法ならびに、それらの意味・用法間の関連性について考察を行った。「一周回って」は、単に一周回るという動作を行うことを表すものであったが、近年では、事柄の原点への回帰や変化にも用いられ、その回帰や変化に対して話し手が何らかの評価をくだしているとも言える「一周回って」や、さらには、程度を強調している「一周回って」までも見られるようになってきている。様態を修飾していたものが、話し手の判断に関わるものや程度を修飾するものへと変化することは、副詞の変遷過程においてはよく見られる現象である¹¹。「一周回って」の意味・用法に見られる変化の方向性も自然なものであると言えよう。

また、今回収集した「一周回って」のデータの中には、次のような、意味・用法の判断に迷う例があった。事柄の原点への回帰や変化を言及しているとも、考え方・見え方の維持や、程度の強調を表しているとも言いがたい。

(33) 新型コロナ、一周回ってペニシリンで治らんかな。

(2020/6/1、<https://twitter.com/suziQ201611/status/1267218012618158081>)

(34) 聞いて聞いて、普段髪の毛流してるイケメン先生がおるんだけど、昨日前髪作っててまじ爆イケだった!!!!!!もう、一周回って結婚しよう👰🏻本当大好き💕

(2022/6/1、<https://twitter.com/juW3z6Ly3E1a1qA/status/1531764545257017344>)

2節でも述べた、Twitterの投稿データから収集した用例を分析する際の限界から、正確な数は出せないが、このような、<意味・用法 A~E>では説明がつかない「一周回って」は、2020年ごろのデータより増加しているように見受けられる。「一周回って」の使用をめぐっては例35のような興味深い投稿も見られており、ひとまず「一周回って」を使っておけば、話し手が述べようとしていることが、何らかの思考活動を踏まえてのものであることが示されるのかもしれない。「一周回って」の使用に見られる今後の変化に注目したい。

(35) 一周回って、を頭につけておけば、大概の発言は多少の深みがあるかのようになるのではないか。

(2009/9/23、<https://twitter.com/tak1000/status/4308201475>)

謝 辞

本研究はJSPS 科研費 22K00638、20K13004 の助成を受けたものである。

文 献

岡島昭浩 (2011) 「ウェブ検索の応用」 荻野綱男・田野村忠温 (編) 『講座 IT と日本語研究 6 ユーパスとしてのウェブ』 明治書院、pp.59-88.

¹¹ これらを指摘した研究に、播磨 (1993)、永澤 (2007) などを挙げることができる。

- 岡田祥平 (2013) 「Twitter を利用した新語・流行語研究の可能性—アイドルグループ「Sexy Zone」の略語を例に—」『新潟大学教育学部研究紀要 人文・社会科学編』6-1、pp.49-74.
- 自由国民社 (編) (2019) 『現代用語の基礎知識 2019 年版』自由国民社 (ジャパンナレッジ版 [<https://japanknowledge.com>], 最終閲覧日: 2022 年 8 月 23 日)
- 総務省情報通信政策研究所 (2021) 『令和 2 年度情報通信メディアの利用時間と情報行動に関する調査報告書』https://www.soumu.go.jp/main_content/000765258.pdf (最終閲覧日: 2022 年 8 月 24 日)
- 津田大介 (2009) 『Twitter 社会論 新たなリアルタイム・ウェブの潮流』洋泉社
- 永澤済 (2007) 「様態副詞から文副詞へ—日本語の副詞「明らかに」の歴史的変化を追って」『認知言語学論考』7、ひつじ書房、pp.97-115.
- 仁田義雄 (1995) 「シテ形接続をめぐって」仁田義雄 (編) 『復文の研究 (上)』くろしお出版、pp.87-126.
- 日本語記述文法研究会 (編) (2008) 『現代日本語文法 6 第 11 部 復文』くろしお出版
- 播磨桂子 (1993) 「「とても」「全然」などにみられる副詞の用法変遷の一類型」『語文研究』75、pp.11-22.

関連 URL

Twitter <https://twitter.com>

近現代語における「もちろん」の用法

東泉 裕子 (東洋大学)

高橋 圭子 (東洋大学)

Usage of *mochiron* in modern and contemporary Japanese

Yuko Higashiizumi (Toyo University)

Keiko Takahashi (Toyo University)

要旨

本研究では、各種コーパスを利用し、近現代語における「もちろん」の用法を調査した結果を報告する。調査の結果、次の3点が明らかになった。(i)19世紀末から20世紀初頭にかけて、中世以来中心的であった名詞述語用法の割合が減少し、副詞としての用法(副詞用法、譲歩用法、応答用法)が増加した。(ii)現代の会話では「もちろん」単独の形式や畳語の形式による感動詞的な応答用法が観察される。(iii)近現代には「～はもちろん(のこと)…も/まで(も)」という尺度含意用法や、「もちろん…〈逆接表現〉…」といった譲歩用法などが定型化した。

1. はじめに

現代日本語の「もちろん」には、主に(1)～(5)のような用法がある。(1)～(5)は「現代日本語書き言葉均衡コーパス(BCCWJ)」および「日本語日常会話コーパス (モニター公開版)(CEJC)」のものである¹。「もちろん」の品詞は、(1)・(5)は名詞、(2)～(4)は副詞である。

- (1) 博士論文は、修士論文より高度でなければならぬことはもちろんである。
BCCW, LBi3_00011, 910, 1994年
- (2) 尾形： #うん#え#来るっしょう#
菊地： #もちろん行きますよ#
尾形： #お#いいね#
CEJC, T006_006b, 48710, 2016年
- (3) このようにいうと、何か遠大な理想に邁進せよというように聞こえるかもしれないが、必ずしもそうではない。もちろんライフ・ワークを大きく持つこともいいかもしれないが、ささやかなことによるこびを見出すことも重要である。
BCCW, OB1X_00192, 54200, 2004年
- (4) 武井： #一個食べて平気なの#
佐竹： #もちろん#
矢口： #うん#うん#うん#
CEJC, T011_017, 93930, 2017年

¹ コーパス検索アプリケーション『中納言』で検索した用例には、順に、コーパス略称、サンプルID、開始位置、出版年/収録年などを記す。以下、用例の下線は稿者らによる。なお、CEJCの用例の「#」は、発話単位を示す区切りである。

- (5) オーストラリアの学校では、大学はもちろんのこと、高等学校から小学校の教育機関においても、英語以外の外国語授業を活発に行っているところがひじょうに多い。
BCCW, OT33_00100, 16550, 2006年

(1)はコンピュータを伴う名詞述語の用法である。(2)~(4)は副詞の用法である。(2)は一般的な副詞の用法である。(3)は副詞「もちろん」が「もちろん…〈逆接表現〉…」という談話構造で使用され、全体で「譲歩」の意味を表す用法である。(4)は副詞の感動詞的用法であり、相手の発話への応答として用いられている。(5)は「~はもちろんのこと…も」という定型化した「尺度含意」の表現である。本稿では、(1)を名詞述語用法²、(2)を副詞用法、(3)を譲歩用法、(4)を応答用法、(5)を尺度含意用法と、それぞれ呼ぶことにする。

本研究では、各種コーパスを用いて、近現代語の「もちろん」の使用実態を調査する。

2. 先行研究

2.1 「もちろん」の語史

「勿論/もちろん」という語が出現したのは中世である(藤堂(編)1978、諸橋 2018、『日本国語大辞典 第二版』)。中世の用例は文末での名詞述語用法が中心であったが、(2)のような副詞用法もすでに見出される(中山 1995、鳴海 2014、高橋・東泉 2019, 2020a)。また、(4)のような応答用法の先駆的例も観察される(高橋・東泉 2019, 2020a)。近世にはジャンルや分野を問わず広く用いられるようになり、(3)の譲歩用法や(5)の尺度含意用法も少なからず見出されるようになる(高橋・東泉 2019, 2020a, 2020b)。

2.2 現代語における「もちろん」の用法

(2)のような「もちろん」の副詞用法は、「誘導副詞」(渡辺 1971)、「話し手の主観的態度を表す副詞(SSA 副詞)」(森本 1994)、「叙法副詞」(工藤 2016)、「陳述副用語」の一つ(趙 2013)として研究されてきた。

(3)の譲歩用法については、趙(2013: 227)、飛田・浅田(2018: 548)などでも指摘されている。東泉・高橋(2020)は、現代語のコーパスを用いて、この用法を調査している。

(4)のような「もちろん」の用法を「応答表現」として分析した研究には、柏野(2019, 2020)がある。「応答表現」とは、「相手の発話に応じ、何らかの反応を返す表現」(柏野 2019: 368、2020: 331)のことである。柏野(2019, 2020)は、「うん」「はい」「ああ」などの狭義の感動詞だけでなく、「ですよ」「だろうね」などの文末表現に由来するものや、「いいね」「さすが」「すごい」「なるほど」「確かに」「了解」「もちろん」「だいじょぶ」「だめ」なども「応答表現」としている³。同じように、「もちろん」を「応答表現」として分析した研究には、森山(1989)、蓮沼(2018, 2021)などがある。

(5)のような尺度含意用法は、グループ・ジャマシイ(編)(1998: 586-587)、飛田・浅田(2018: 584)などに記述されている。

² 「~はもちろんだが」などの「~はもちろん+コンピュータ+〈逆接表現〉…」は譲歩の意味を表すが、本稿の調査対象では6例(いずれも表2のCHJ1874-1909雑誌・教科書)のみであり、便宜上、本稿では名詞述語用法に分類した。また、(5)の「もちろん」は名詞だが、(1)とは用法が異なるため別に分類した。

³ 日本語学会(編)(2018: 191)は、感動詞を、「ああ」「へえ」「ほほう」などの「原生感動詞」と「あれっ」「どれ」「ほんと」など他の品詞から転成した「転成感動詞」に分けている。前者が狭義の感動詞に相当する。

3. 調査データ

今回の調査に用いたコーパスは表 1 にまとめた。いずれのコーパスも国立国語研究所コーパス開発センターによるものである。

青空文庫パッケージ以外は、コーパス検索アプリケーション『中納言』を用いて検索した。短単位モードを使い、語彙素「勿論」で検索を行った。

青空文庫パッケージについては、全文検索システム『ひまわり』ver.1.7.2 を利用した。文字列検索で「勿論」「もちろん」をそれぞれ検索し、100 語ずつランダムに抽出した。

コーパスから得た用例は、目視で確認し、前後の文脈が不明のものは除去した。検索は、2021 年 11 月から 2022 年 4 月に行った。

表 1 調査対象コーパス

コーパス名 (略称)	調査対象期間	データバージョン	中納言バージョン	分析対象用例数
日本語歴史コーパス(CHJ)	1874–1925	2022.03	2.5.2	106
青空文庫パッケージ(AO)	1926–1962	2022-04-02	—	92
現代日本語書き言葉均衡コーパス(BCCWJ)	1976–2007	2021.03	2.4.5	233
昭和話し言葉コーパス(SSC) (丸山他 2022)	1952–1956	2022.03	2.4.5	135
日本語話し言葉コーパス(CSJ)	1999–2001	2019.01	2.4.2	117
日本語日常会話コーパス(モニター公開版)(CEJC)(小磯他 2020)	2016–2018	2021.03	2.4.5	177

4. 調査結果

4.1 近現代語の書き言葉のコーパスにおける「もちろん」

表 2 は、本研究で調査対象とした近現代語の書き言葉のコーパスにおける「もちろん」の用例をまとめたものである。CHJ の調査対象期間における「もちろん」の用例は「明治・大正-雑誌」および「明治・大正-教科書」に見出された。そこで、BCCWJ においても「出版・雑誌」および「特定の目的・教科書」を利用した。ただし、「出版・雑誌」は用例数が多いためコアデータのみを対象にした。また、「出版・雑誌」以外のコアデータの用例数を、調査対象期間「1976–2008」の列にまとめた。

表 2 から読み取れることは、19 世紀末から 20 世紀初頭にかけて、名詞述語用法の割合が減少し、副詞(副詞用法、譲歩用法、応答用法)としての用法の割合が増加したことである。また、譲歩用法や尺度含意用法などの定型化した表現が、近現代語ではある程度定着したことが窺える。

表2 近現代語の書き言葉のコーパスにおける「もちろん」

調査対象期間	1874-1909	1917-1925	1926-1962	2001-2007	1976-2008	
コーパス	CHJ	CHJ	AO	BCCWJ	BCCWJ	
サブコーパス	雑誌・ 教科書	雑誌・ 教科書	—	雑誌・ 教科書	雑誌・教 科書以外 ⁴	
名詞述語用法	15 24.2%	3 6.8%	1 1.1%	2 2.0%	2 1.5%	
副詞	副詞用法	20 32.3%	21 47.7%	65 70.7%	54 53.5%	89 67.4%
	譲歩用法	10 16.1%	12 27.3%	10 10.9%	21 20.8%	30 22.7%
	応答用法	1 1.6%	1 2.3%	3 3.3%	2 2.0%	1 0.8%
	小計	31 50.0%	34 77.3%	78 84.8%	77 76.2%	120 90.9%
尺度含意用法	16 25.8%	7 15.9%	13 14.1%	22 21.8%	10 7.6%	
合計	62 100%	44 100%	92 100%	101 100%	132 100%	

「もちろん」の応答用法は、少数ではあるが、小説などの会話において確認することができた。(6)は本調査における初出例であり、(7)はその次に古い例である。20世紀初頭には、「もちろん」の応答用法が会話で使用されていたことを窺い知ることができる。ただし、2例とも表2のAOの調査対象期間(1926-1962年)以前における用例であるため、表2の用例数には含めなかった。

(6) 「君、僕は昨夜からの疑問だがネ、梅子さんの胸底には若し、恋が潜んでるのぢや無からうか」大橋は莞爾と打ち笑み「勿論！彼女の心が恋愛の聖火に燃ゆること、抑も一朝一夕の故に非らずサ、...(略)... AO, 『火の柱』, 木下尚江, 1904年⁵

(7) 「いろんなものにひかれるのは結構じゃないか。つまりそれだけ、おぬしはあたまが広いのだからな。」そう出られれば「もちろん」と裕佐はいいたくなるのだった。AO, 『青銅の基督』, 長与善郎, 1923年

次の(8)と(9)も、「もちろん」の応答用法と同様の働きをしていると考えられるため、応答用法として数えた。

⁴ 「出版・雑誌」以外のコアデータは、「出版・新聞」、「出版・書籍」、「特定目的・白書」、「特定目的・知恵袋」、「特定目的・ブログ」である。

⁵ 青空文庫パッケージの用例には、順に、コーパス略称、作品名、著者、初出年を記す。なお、用例の下線は稿者らによる。

- (8) 宮城。 むむ、それで、けふ、やすんだんだな。
山口。 もち論でさ、まア、これを御覽じろ。

CHJ, 60M 女雑 1894_37015, 53090, 1894 年

- (9) 『...(略)... 私の無免許を罰しないと約束するなら、あなたの命を救ふ爲に、私の心理的施術をしてあげませう』と私は言つた。『勿論ですとも。ああ又痛くなつて来た、早くして下さい』と言ふから ...(略)...

CHJ, 60M 太陽 1925_04056, 36020, 1925 年

4.2 現代語の話し言葉のコーパスにおける「もちろん」

表3は本研究で調査対象とした現代語の話し言葉のコーパスにおける「もちろん」の用例数とその使用比率である。ただし、CSJは用例数が多いためコアデータのみを対象にした。

表3 現代語の話し言葉のコーパスにおける「もちろん」

調査対象期間		1955-1974	1999-2001	1952-1969	1986-2018
タイプ		独話		対話/会話	
コーパス		SSC	CSJ	SSC	CSJ & CEJC
名詞述語用法		1 1.1%		2 4.3%	3 1.6%
副詞	副詞用法	62 70.5%	54 50.5%	32 68.1%	81 43.3%
	譲歩用法	22 25.0%	53 49.5%	6 12.8%	54 28.9%
	応答用法			4 8.5%	47 25.1%
	小計	84 98.9%	107 100.0%	42 89.4%	182 97.3%
尺度含意用法		3 3.4%		3 6.4%	2 1.1%
合計		88 100%	107 100%	47 100%	187 100%

表3からは、「もちろん」は大部分が副詞(副詞用法、譲歩用法、応答用法)として使用されていることが確認できる。

4.3 現代語の話し言葉のコーパスにおける「もちろん」の応答用法

表4は、話し言葉のコーパスの会話における「もちろん」の応答用法をまとめたものである。(4)のような単独の形式の「もちろん」の他に、「もちろんもちろん」などの畳語、その他の形式を応答用法として数えた。なお、表4の数値は「あー」「えっ」「うん」などが先行するものも含む⁶。

⁶ これらの表現は、CEJCで「感動詞-一般」のタグが付与されているものとした。

表4 現代語の話し言葉のコーパスにおける「もちろん」の応答用法⁷

調査対象期間	1952-1969	1986-2018	用例
タイプ	会話		
コーパス	SSC	CEJC	
もちろん (もちろん、もちろんもちろん、他)	4 100%	34 72.3%	(4), (10), (11), (12)
もちろん+終助詞		2 4.3%	(13)
もちろん+コピュラ (+終助詞)		10 21.3%	(14), (15)
それはもちろん		1 2.1%	(16)
合計	4 100%	47 100%	

(10)~(16)は、応答用法の例である。(10)~(12)は、「もちろん」が2回以上繰り返されているものである。(12)ではフィラーの「まあ」が前接している。(13)では「うん」に続く「もちろん」に終助詞「ね」が付加されている。(14)・(15)はコピュラを伴い、(15)にはさらに終助詞が付加されている。(13)・(16)は、モノローグ的な発話の間で使用されていることから、相槌的な用法と考えることもできる。

- (10) 島村： #あとあのレンジでちんした野菜が基本よね#
 すみれ： #もちろんもちろん#
 はるな： #あー#レンジでちんした野菜って何# CEJC, K004_008, 38760, 2017年
- (11) 牧野： #もし来ていただけるととても心強いですけれど#
 由美： #もちろんもちろん#もちろんもちろん#
 牧野： #うーんと六日#六月六日# CEJC, T003_003, 29760, 2016年
- (12) 大場： #だから...(略)...普通の楽譜の感じで来られたら全部わかるってゆう#
 徹： #まあもちろんもちろんもちろん#
 龍之介： #うん# CEJC, T010_004, 84810, 2016年
- (13) 岩崎： #ほんとは...(略)...一応ディベディベートだから反対意見と賛成意見を
 やってみるってゆう#
 一ノ宮： #うん#
 松下： #うん#そう#
 田代： #うん#もちろんね
 松下： #そう#
 岩崎： #それでを見てどう思いましたか# CEJC, T004_013, 57940, 2016年

⁷ 「もちろん」の応答用法の用例は、CSJの「対話」では皆無だった。

- (14) 青井： #わたしもワインいただいていいですか#
 はるか： #もちろんです#
 由里子： #どうぞ# CEJC, T004_010b, 7560, 2016 年
- (15) 田辺： #ここってでもお酒とかって飲めたりするんですか#
 溝口： #もちろんだよ#
 由里子： #なんか#あつ# CEJC, T013_013, 7990, 2017 年
- (16) 神崎： #でやっぱどんなどんな人でもさやっぱ平和に生きたいとは思ってるし#
 徹： #それはもちろんもちろん#
 神崎： #みんな仲良くできたらいいなって絶対思ってるはずなんだよね#
 CEJC, T010_013, 73270, 2017 年

5 おわりに

本発表では、近現代語における「もちろん」の用法を、複数のコーパスを用いて調査した。その結果、19世紀末から20世紀初頭にかけて、名詞述語用法の割合が減少し、「もちろん」の副詞の用法(副詞用法、譲歩用法、応答用法)が増加したことが明らかになった。また、現代の会話では、「もちろん」単独の形式や畳語の形式の応答用法が観察された。さらに、「～はもちろん(のこと)…も/まで(も)」や「もちろん～、〈逆接表現〉…」などの定型化した表現が、近現代に定着したこともわかった。

今後、さらに調査対象を拡大し、用例を精査していきたい。

謝 辞

言語資源ワークショップ2022において示唆に富むご意見やご質問をお寄せくださった皆様に感謝申し上げます。なお、本研究は、日本学術振興会科学研究費補助金による基盤研究(C)「漢字文化圏における漢語の語用論的標識化」(研究代表:高橋圭子、課題番号:20K00650)、同「構文化としての定型表現の発達と談話構造の並行進化について—英語史からの検証—」(研究代表:柴崎礼士郎、課題番号:22K00610)の研究成果の一部である。

文 献

- 柏野和佳子 (2019) 『日本語日常会話コーパス』モニター公開版に見られる応答表現『言語資源活用ワークショップ発表論文集』4, pp. 386–380. 国立国語研究所
- 柏野和佳子 (2020) 『日本語日常会話コーパス』モニター公開版に見られる感動詞以外の応答表現』5, pp. 331–347. 国立国語研究所
- 工藤浩 (2016) 『副詞と文』ひつじ書房
- 小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉 (2020) 『日本語日常会話コーパス』モニター版の設計・評価・予備的分析『国立国語研究所論集』18, pp. 17–33. 国立国語研究所
- 高橋圭子・東泉裕子 (2019) 『勿論』考『言語資源活用ワークショップ発表論文集』4, pp. 128–138. 国立国語研究所
- 高橋圭子・東泉裕子 (2020a) 「語用論的標識としての『勿論』の歴史」『東洋大学人間科学総

- 合研究所紀要』22, pp. 197–208. 東洋大学人間科学総合研究所
高橋圭子・東泉裕子 (2020b) 「ヘッジとしての『もちろん』の歴史」『コミュニケーション文化』14, pp. 43–49. 東洋大学人間科学総合研究所
趙英姫 (2013) 「近現代の漢語副詞の成立」野村雅昭 (編)『現代日本漢語の探究』pp. 214–233. 東京堂出版
中山緑朗 (1995) 『平安・鎌倉時代古記録の語彙』東苑社
鳴海伸一 (2014) 「漢語形容動詞・副詞の品詞性と用法変化：通時的観点からみた近現代の特徴」『国立国語研究所共同研究報告 13-03 近現代日本語における新語・新用法の研究』pp. 56–75. 国立国語研究所
蓮沼昭子 (2018) 「自然談話における副詞の応答用法—「もちろん」「たしかに」「なるほど」を例に—」『日本語日本文学』28, pp. 1–16. 創価大学日本語日本文学会
蓮沼昭子 (2021) 「自然談話における副詞の応答用法—「もちろん」「たしかに」「なるほど」を例に—」『日本語日本文学』31, pp. 30–56. 創価大学日本語日本文学会
東泉裕子・高橋圭子 (2020) 「現代日本語における『もちろん』のヘッジ用法—現代語コーパスの用例より—」『アカデミック・ジャパニーズ・ジャーナル』12, pp. 26–43. アカデミック・ジャパニーズ・グループ研究会
丸山岳彦・小磯花絵・西川賢哉 (2022) 「『昭和話し言葉コーパス』の設計と構築」『国立国語研究所論集』22, pp. 197–221. 国立国語研究所
森本順子 (1994) 『話し手の主観を表す副詞について』くろしお出版
森山卓郎 (1989) 「応答と談話管理システム」『阪大日本語研究』1, pp. 63–88. 大阪大学文学部
渡辺実 (1971) 『国語構文論』塙書房

辞書・辞典

- グループ・ジャマシイ(編) (1998) 『日本語文型辞典』くろしお出版
藤堂明保(編) (1978) 『学研漢和大字典』学習研究社
日本語学会(編) (2018) 『日本語学大辞典』東京堂出版
飛田良文・浅田秀子 (2018) 『現代副詞用法辞典 新装版』東京堂出版
諸橋轍次(著) 鎌田正・米山寅太郎(修訂増補) (2018) 『大漢和辞典 デジタル版』大修館書店

関連 URL

- コーパス検索アプリケーション『中納言』 <https://chunagon.ninjal.ac.jp/>
国立国語研究所言語資源開発センター <https://clrd.ninjal.ac.jp>
Japan Knowledge Lib 『日本国語大辞典 第二版』 <https://japanknowledge.com/library/>
全文検索システム『ひまわり』 <https://csd.ninjal.ac.jp/lrc/index.php?himawari>

医学書テキストのたとえる表現 (2) —接尾辞「状」の特徴

本多 由美子 (一橋大学・国立国語研究所) †
三枝 令子 (元一橋大学)

Illustrating Expressions in Medical Texts (2) : Characteristics of the suffix “jou”

Yumiko Honda (Hitotsubashi University / NINJAL)
Reiko Saegusa (Hitotsubashi University)

要旨

本研究では医学書テキストにおける「たとえる表現」の一端を明らかにする目的で接尾辞「状」に注目し、用法を分析した。調査には医学書5冊(約450万語)のデータを用いた。分析の結果、前接する語には「S、線」など形そのものを表すもの、「粥、海綿」など質的な様子も表す語、「嚢胞、結節」などの部位が変化した形を表す語が多く、後接する語には「血管、結腸」などの体の部位や「陰影、硬化」など病態を表す語が多かった。「状」の後には名詞が最も多いが、3割程度は「の」を介した名詞修飾や「に、と」などを介した動詞修飾用法であった。『現代日本語書き言葉均衡コーパス』の書籍に付与されたNDC情報を用いて比べたところ、「状」は医学において多用される語であることが示唆された。また、同じ「たとえる表現」の接尾辞「様(ヨウ)」と比較した結果、「様」の前には疾患名のような状態全体を表す語が見られ、用法の違いが観察された。

1. はじめに

筆者らは、日本の医師国家試験を受験する外国人学習者への日本語支援を目的として、医学教材の開発を目指し、医学書テキストを対象とした言語学的分析を行っている。医学書テキストには、患者の症状や、検査結果などをもとに診断を行うための情報が記載され、そこには「どのような状態・様子であるか」を表す「たとえる表現」が見られる。たとえる表現は、「蕁状血管腫」や「網目状の紅斑」のように対象(血管腫、紅斑)を具体的に説明するのに用いられる。医学や医療の分野における先行研究では、用語の抽出や合成語の語構成に関する研究が多く行われているが、たとえる表現を分析した研究は「ような」と接尾辞「様(ヨウ)」のみであり、その他の表現については使用傾向が明らかではない。そこで、本研究では、医学書テキストのたとえる表現の1つである接尾辞「状」の分析を行う。

まず、「何」を「どう」たとえるのかという観点から、接尾辞「状」に前接する語(=どう)と後接する語(=何)に注目し分析を行う。次に、医学書テキストと一般的な書き言葉との比較をするために『現代日本語書き言葉均衡コーパス』(以下、BCCWJと略す)から抽出した「状」のデータとの比較を行う。さらに、同じくたとえる表現である接尾辞「様」との比較をとおして、「状」と「様」それぞれの特徴を明らかにしたい。

† nihonda[at]hotmail.com

2. 先行研究

接尾辞「状」は、「名詞に付いて、...のような形である、...に似たようすである、などの意を表す。」（『デジタル大辞典』）と説明されている。野村（1978：116、120）では、「状」は「的」や「様」と同じく、「...ノヨウス」「...ラシサ」といった意味をそえる「相言型」に分類され¹、「相言型」全体の特徴として、「～状ノ～」という形で体言を修飾することが多く、「ガ」・「ヲ」などの格助詞をともなうことはほとんどないと述べられている。「状」については、「のり状」という語例が紹介されているが、詳しい言及はされていない。

医学や医療に関する用語では、接辞の頻度が高いことが指摘されている（三枝他 2020、内山他 2018）。しかし、個別の接辞を扱った研究は、病名における接辞「性」を分析した東条他（2020）など、わずかである。

医学書テキストを対象に、たとえる表現を分析したものには三枝・本多（2021）がある。三枝・本多（2021）では、たとえる表現の「ような」と接尾辞「様」の用法を比較し、「ような」よりも「様」を用いた表現のほうが叙述性が薄まり一語性が強くなることや、「様」の前に疾患名が置かれる用法は医学特有の表現であることが指摘されている。

3. データ・調査手順

本稿では医学書テキストのデータに以下の5冊を用いた。①～④は長年版を重ね現場の医師によく利用されている書籍であり、⑤は医師国家試験の受験参考書として評価の高い書籍である。

- ① 『今日の治療指針 2018 年版』
- ② 『今日の診断指針 第7版』
- ③ 『新臨床内科学 第9版』
- ④ 『今日の小児治療指針 第16版』
- ⑤ 『イヤートート 2020 内科・外科編』

この5冊から本文テキストを抽出し、このうち句点で終わる行のみを対象に形態素解析を行った²。語数は4,510,648語（記号等・空白・改行を除く）である。以下の手順で分析対象の語を抽出したところ、1,921語となった。

- ① 形態素解析の結果から語彙素が「状」かつ品詞が「接尾辞-名詞的-一般」のデータを抽出する。
- ② 前接する1語（以下、「前接語」とする）が補助記号、空白、未知語のデータおよび後接する1語（以下、「後接語」とする）が未知語のデータを除外する。
- ③ 目視で「状」が「書状」を表す例および誤解析を除外する³。

¹ 野村（1978）では接辞性字音語基を「体言型」「用言型」「相言型」に3分類している。

² 形態素解析には MeCab の ver.0.996（辞書は UniDic-cwj-2.3.0）を使用した。

³ 「状」が書状を表す語は9語（紹介状8語、告訴状1語）であった。誤解析の例は、「胃腸炎症状」や「有症状」が「胃腸／炎症／状」「有症／状」と解析されるような、「症状」の誤解析で、296語であった。

4. 医学書テキストにおける「状」

4.1 前接語の特徴

「状」の前接語は、延べ語数 1,921 語、異なり語数 366 語であった。表 1 は、頻度 30 以上の語を高頻度順にまとめたものである。「例」の欄には、後接する 1 語あるいは 2 語を確認し、頻度の高い表現を挙げた。例えば、「S」は 85 例のうち 78 例が「S 状結腸」であった。

表 1 で最も頻度の高い語は「S」であるが、「状」全体に占める割合は 4.4% である。このことから、「状」には多様な語が前接していることがわかる。表 1 に挙げた語を見ると、「S」「顆粒」「線」のように、形そのものを表すものや、「索」「尾」「巢」など、形に特徴のあるもの、「粥」「海綿」のように、質的な特徴を想起させるものも見られる。また、「囊胞」「結節」は部位が変化した形を表す語である。「囊胞」は『医学大辞典』によると、内腔に液体や泥状物を含む袋のような形の構造物であり、外から見た形に加え内部の質的な様子も想起させる語である。

表 1 の「例」の欄では、語彙素の頻度に占める割合が 30% 以上のものを下線で示した。表 1 に示した前接語 12 語のうち、「S」「粥」「海綿」「索」「尾」の 5 語は、「S 状結腸」(78 例、91.8%)、「粥状硬化」(31 例、48.4%)、「粥状動脈硬化」(22 例、34.4%)、「海綿状血管腫」(25 例、50.0%)、「索状物」(15 例、31.2%)、「尾状核」(29 例、69.0%) という形式で用いられていた。このように、頻度の高い前接語の中には、さらに長い合成語を形成し、ある程度決まった形式で用いられるものがあることがわかる。

表 1 接尾辞「状」の前接語 (頻度 30 以上、N=1,921)

順位	語彙素	頻度	割合	例
1	S	85	4.4%	<u>S 状結腸 78</u>
2	粥	64	3.3%	<u>粥状硬化 31</u> 、 <u>粥状動脈硬化 22</u>
3	囊胞	55	2.9%	囊胞状に～13
3	顆粒	55	2.9%	顆粒状の～10、顆粒状沈着 10
5	海綿	50	2.6%	<u>海綿状血管腫 25</u>
6	結節	48	2.5%	結節状の～12
6	索	48	2.5%	<u>索状物 15</u>
8	尾	42	2.2%	<u>尾状核 29</u>
9	乳頭	35	1.8%	乳頭状腫瘍 5
10	線	34	1.8%	線状の～7
11	泡沫	32	1.7%	泡沫状の～10
12	巢	30	1.6%	巢状分節 (性～) 7

・下線：語彙素の頻度に占める割合が 30% 以上のもの

医学書テキストにおける「状」を用いた表現は、(1) の「顆粒状」のような視診のほか、(2) の「索状物」のように触診の際の情報を示す場合もある。

- (1) 主に糸球体係蹄に顆粒状かつ花卉を縁取るような形態でみられ、本症を診断する手がかりとなる (内科_腎疾患)

(2) 下腹部に連続する腫瘍や索状物を触知することもある。(診断_小外科疾患)

また、表 1 の語種を見ると、「S」を除くとほぼ漢語である。形態素解析の結果では「粥(カユ)」「尾(オ)」「巢(ス)」はいずれも和語に分類されているが、本調査で用いた医学テキストでは「粥状(ジャクジョウ)」「尾状(ビジョウ)」と読む。「巢」は頻度 30 のうち、「巢状(ソウジョウ)」が 28、「クモの巢(ス)状」が 2 であった。

表 1 には掲載されていないが、和語で頻度が高かった前接語は、「鎌(カマ)」24、「編目(書字形:網目)」21、「莓(イチゴ)」18 であった。外来語の前接語は高頻度順に、「リング」22、「ポリープ」18、「ゼリー」17、「ガラス」14、「ドーム」11、「バルーン」5、「スリット」5、「ドーナツ」5 であった。「リング」「ドーム」「スリット」「ドーナツ」のように形に特徴のあるものや、「ゼリー」「ガラス」のように質的な特徴を想起させるもの、「ポリープ」のように体内にできた病変を表す語が見られ、表 1 に見られる語と特徴が重なる。

4.2 後接語の特徴

表 2 は、状の後接語の品詞を見たものである。全体としては名詞接続が 54.3% と半数を超えている。2 位は助詞で、32% を占める。表 3 は、「状」の具体的な後接語を高頻度順にまとめたものである。延べ語数は 1,921 語、異なり語数は 207 語である。後接語で最も頻度が高いのは助詞の「の」であった。このことは、「状」の後には名詞が接続することが最も多いが、「の」を介した名詞接続も少なくないことを示している。「の」の次に多い後接語は同じく助詞の「に」で、「に」を介してさらに動詞に接続する(に拡張する 19 例、になる 16 例など)。表 3 の頻度 32 以上の高頻度の後接語 12 語の中で、助詞 4 語、名詞 6 語、さらに補助記号 2 つと、「状」の後接語は多様である。

表 2 「状」の後接語の品詞

順位	品詞	頻度	割合
1	名詞	1,042	54.3%
2	助詞	616	32.1%
3	補助記号	147	7.7%
4	接尾辞	53	2.8%
5	接続詞	31	1.6%
6	接頭辞	18	0.9%
7	助動詞	8	0.4%
8	記号	4	0.2%
9	形状詞	1	0.1%
9	空白	1	0.1%
合計		1,921	100.0%

表 3 接尾辞「状」の後接語

(頻度 32 以上、N=1,921)

順位	語彙素	頻度	割合
1	の	303	15.8%
2	に	198	10.3%
3	血管	83	4.3%
4	結腸	78	4.1%
5	,	61	3.2%
6	陰影	58	3.0%
7	硬化	35	1.8%
7	と	35	1.8%
9	を	34	1.8%
10	核	33	1.7%
10	突起	33	1.7%
12)	32	1.7%

表 3 の 5 位の補助記号の読点は以下の (3) に見るように並列用法である。また (4) に見

る12位の「」は、前に示された語の内容を「状」を用いて説明しているもので、独立用法と言える。表2の5位の接続詞は(5)のような「または」「あるいは」等、並列の接続詞が用いられている。

- (3) 表面は不均一で、乳頭状、結節状、ポリープ状など様々な形態の腫瘤形成がみられる。
(診断_産科・婦人科の症候)
- (4) 初期：発熱(階段状)(イヤート_感染症)
- (5) 混成石はコレステロールとビリルビンカルシウムの成分が結石の中に混在しエコーでは半月状または三日月状で強い音響陰影を伴う。(診断_腹部・消化器系の症候)

以上の「状」の後接語の観察から、医学書テキストにおける「状」は名詞複合語の一部を構成する場合もあるが、動詞を修飾する用法、並列的な用法、さらには説明をする独立的な用法もあることがわかる。

最後に、「状」に直接名詞が後接する場合と、「の」を介して後接する場合とを比較する。表4は、「状」に後接する名詞を高頻度順にまとめたものである。「血管・腫」や「結腸」は表1の頻度の高い前接語に対応するものだが、それ以外にも「陰影」「突起」「変化」「拡張」といった形や変化を表す一般語も後接語に使われているのがわかる。

表4 接尾辞「状」に後接する名詞(頻度19以上、N=1,042)

順位	語彙素	頻度	割合	例
1	血管	83	8.0%	海綿状血管腫 25、莓状血管腫 16
2	結腸	78	7.5%	S状結腸 78
3	陰影	58	5.6%	ガラス状陰影 12、腫瘤状陰影 8
4	硬化	35	3.4%	粥状硬化 31
5	突起	33	3.2%	鞘状突起 21、針状突起 6
5	核	33	3.2%	尾状核 29
7	動脈	24	2.3%	動脈硬化 22
8	変化	20	1.9%	変化が～4
9	拡張	19	1.8%	拡張を～5

「状」に「の」が後接する場合は303例(表3)だが、「の」に接続する名詞の異なり語数は182語と多く、多様な語が接続することがわかった。頻度の高い語は順に「高」12、「物」10、「腫瘤」9、「隆起」「形態」各7、「内容」「低」各5であった。この内の「高」は、「高信号」9という形式で用いられる(例 十字状の高信号)。「の」を介した「状の」には後接する名詞に疾患名を示す名詞はなく、すべて形のある物や変化であった。

4.3 『現代日本語書き言葉均衡コーパス』の書籍サンプルとの比較

一般的な書き言葉における「状」と医学書テキストにおける「状」との比較をするために、BCCWJの書籍サンプルに付与された日本十進分類法(NDC)の分類記号の情報を用いて比較することにした。NDCを用いたのは、「状」の使用は医学という書籍のテーマの影響が大

きいのではないかと考えたためである。

まず、BCCWJの「出版・書籍」および「図書館・書籍」のデータから、医学書テキストでの検索と同様に語彙素が「状」かつ品詞が「接尾辞-名詞的-一般⁴」の用例を抽出し⁴、前接する1語が補助記号の例および「状」が「書状」を表す例、誤解析を除外した。分析対象の語は、延べ2,648語、異なり735語となった。これ以降、本稿では「出版・書籍」と「図書館・書籍」の2つのレジスターを合わせて「書籍」と呼ぶ。

次に、書籍データと医学書テキストのデータを比較するために調整頻度を求めた。書籍データは「短単位語数 Excel データ (Version 1.1)」に掲載されているサンプルごとの語数を NDC 分類記号（第一次区分）別に合計した⁵。BCCWJの書籍サンプルにおける NDC 情報は増補されている（加藤他 2021）ため、増補後の情報を用いた⁶。表 5 は調整頻度を求める際に用いた医学書テキストおよび NDC 分類記号別の語数である。なお「中納言」では、NDC 分類記号の第一次区分を「ジャンル」と呼んでいることから、本稿でも「ジャンル」と呼ぶ場合がある。

図 1 は、「状」の 100 万語あたりの語数を、医学書テキストおよび書籍のジャンル別に示したものである。

表 5 医学書テキストおよび NDC 分類記号（第一次区分）別の語数

医学書	0 総記	1 哲学	2 歴史	3 社会科学	4 自然科学	5 技術工学	6 産業	7 芸術美術	8 言語	9 文学
5,358,397	1,682,302	3,500,118	6,712,080	15,380,038	5,268,094	4,708,999	2,565,422	4,627,744	1,185,236	22,555,019

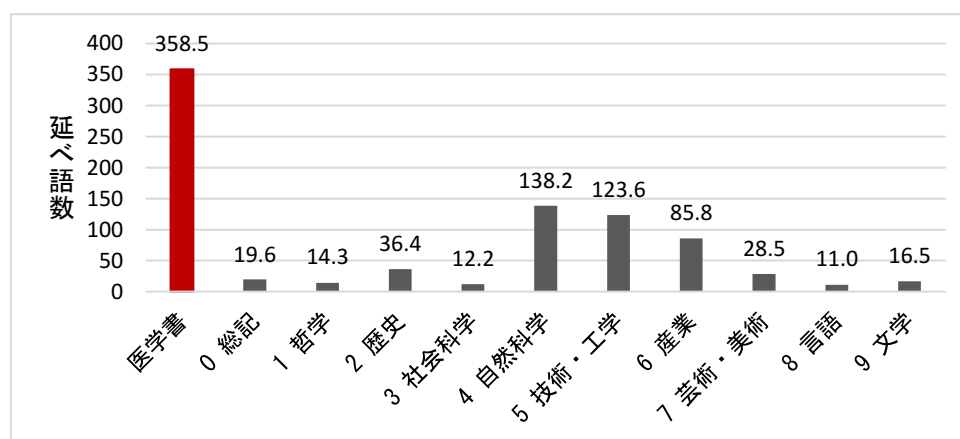


図 1 ジャンル別「状」の語数 (100 万語あたり)

⁴ 抽出には検索アプリケーション「中納言 2.6.0」（データバージョン 2021.03）を用いた。
⁵ 「短単位語数 Excel データ (Version 1.1)」に掲載されているサンプルごとの語数は、記号等を除外した値と、除外されていない値が示されているが、本節の分析で調整頻度を求める際には、記号等が除外されていない値を用いた。これは、「S 状結腸」のように、「状」の前接語の上位に記号が用いられる場合があることによる。医学書テキストの調整頻度を求める際も同様である。
⁶ 書籍サンプルにおける NDC 番号の増補データは加藤祥他「BCCWJ-NDC」にて取得した。
 (<https://github.com/masayu-a/BCCWJ-NDC/>)

図1の延べ語数は、医学書が358.5語と最も多い。BCCWJのジャンルで最も多いのは「自然科学」で138.2語、次いで「技術・工学」の123.6語であるが、医学書テキストは「自然科学」の2.5倍、「技術・工学」の2.9倍と、目立って頻度が高い。また、「自然科学」について、NDC分類番号の第二次、第三次区分を見たところ、割合が高い順に、「医学」55.6%、「動物学」12.5%、「地球科学・地学」8.8%であり、医学が「自然科学」の半数以上を占めていた⁷。これらのことから、たとえる表現の「状」は医学において多用される語であることが示唆される。表6は各ジャンルの頻度順で上位に見られた語である。頻度の低い語はたまたま使用された可能性もあるが、参考までに掲載する。なお、表6では示されていないが、すべてのジャンルに見られた語は「放射」と「帯」の2語であった。

表6 ジャンル別 「状」の前接語の上位語（括弧内：100万語あたり語数）

ジャンル	前接語（語彙素）
0 総記	缶 (1.78) /放射、帯、階段、テープ (1.19)
1 哲学	分子 (1.43) /渦巻き (1.14) /放射、スポンジ (0.86)
2 歴史	袋 (2.98) /板 (2.68) /放射、階段 (1.19)
3 社会科学	粉 (0.85) /箸 (0.46) /板、帯 (0.39)
4 自然科学	放射 (7.40) /帯、S (3.42) /円筒 (3.23)
5 技術・工学	放射 (4.88) /クリーム (3.82) /ペースト (3.61)
6 産業	顆粒 (3.90) /ペレット (3.12) /房 (2.73)
7 芸術・美術	粒 (1.73) /放射 (1.51) /板 (1.30)
8 言語	粒 (1.69) /放射 (0.84) /帯、喇叭、短冊、液体、ツリー (0.84)
9 文学	放射 (0.93) /螺旋 (0.62) /筒、ドーナツ、円盤 (0.40)

4.4 医学書テキストの「名詞-普通名詞-一般」における「状」

ここまで、形態素解析の結果を用いて品詞が接尾辞の「状」を見てきた。形態素解析の結果を見ると、品詞が「名詞-普通名詞-一般」の語の中にも「輪状」「斑状」のように、「状」を含む表現が見られる。これらの語は本調査における接尾辞「状」の用例ではないが、医学書テキストにおけるたとえる表現「状」の特徴を見るために確認しておきたい⁸。

調査方法は、医学書テキストから品詞が「名詞-普通名詞-一般」かつ語彙素の語末に「状」が用いられる語を抽出した。目視で「状」がたとえる表現ではないと思われる語⁹、前後の語が未知語のもの、誤解析を除外した¹⁰。分析対象の語は延べ4,563語、異なり62語であった。表7は頻度が上位の語である。

⁷ NDC分類記号は「NDC Navi」(http://jasmine.media.osaka-cu.ac.jp/ndc_navi10/)を参照した。

⁸ 『現代日本語書き言葉均衡コーパス』利用の手引 第1.1版の5章「形態論情報」には、「接尾辞-名詞的-一般」の「状」について、以下の注記が記載されている。説明中の「=」は「当該規定で注目している箇所」である。

注記：「[～の形・有り様]の意。漢語の1最小単位と結合したものは除く。(液=状)」

⁹ 「状」がたとえる表現ではないとみなし、分析対象から除外した語は次の7語である。「異状、形状、現状、症状、性状、病状、令状」

¹⁰ 誤解析とみなした語は「頭状」5語で「八ツ頭状」が「八つ/頭状」と解析されていた。

表7 「名詞-普通名詞-一般」のたとえる表現「状」を含む語（高頻度順、N=4,563）

順位	語彙素	頻度	割合	順位	語彙素	頻度	割合
1	甲状	2,802	61.4%	7	線状	88	1.9%
2	帯状	206	4.5%	8	環状	77	1.7%
3	輪状	153	3.4%	8	球状	77	1.7%
4	点状	119	2.6%	10	粒状	74	1.6%
5	斑状	96	2.1%	11	糸状	71	1.6%
6	網状	92	2.0%				

頻度の最も高い語は「甲状」で、全体の6割を占めている。「甲状」の後接語を見ると、「甲状」の97.6%（2,735語）は、「甲状線」「甲状腺癌」「甲状腺腫」のいずれかであり、「甲状腺」を含む語であった。「甲状」の「甲」は「兜（かぶと）」を表す（『日本国語大辞典』）が、現在では「兜（かぶと）」という意識は薄れ「甲状腺」という形式で一語化していると思われる。「甲状」を除いた2位以下の語を合計すると延べ1,761語であり、接尾辞「状」の延べ1,921語と比べて、少なくないことがわかる。表5の2～7位の「帯状」「輪状」「点状」「斑状」「網状」「線状」は、表1で1位の「S状」よりも頻度が高かった。

表7の各語彙素の左側の漢字を見ると、「帯」「輪」「点」「斑」「網」「粒」など、形に特徴のあるものが目立つ。これは表1で示された接尾辞「状」の前接語の特徴とも重なる。

5. 接尾辞「様」との比較

「様」は辞書で「名詞の下に付いて複合語をつくる。㊦ある物に類似していることを表す。…ふう。…のよう。」（『デジタル大辞泉』）と説明されるように、接尾辞「状」と類似した表現である。両者の相違を見るために、医学書のテキストから「状」と同様の手順で「様」の前接語と後接語のデータを抽出した。「皆様」の「様」のような「様」が敬称を表す例は目視で除外した。分析対象の語は1,340語となった。

5. 1 「状」と「様」の前接語の比較

「様」の前接語は延べ1,340語、異なり394語であった。表8と表9は前接語の内訳である。「状」よりも「様」のほうが延べ語数が少なく、異なり語数が多いことから、「状」よりも「様」のほうが多様な表現を形成していることがわかる。「状」および「様」の前接語の重なりを見ると、「状」のみ、あるいは「様」のみに見られる前接語は、延べで「状」が74.6%、「様」が78.7%と8割弱、異なりで「状」が79.8%、「様」が81.2%と約8割である。このことから、医学書テキストにおける「状」と「様」の前接語は一部重なりがあるものの、使い分けられていることがわかる。

表8 「状」の前接語の内訳

前接語	延べ		異なり	
	語数	割合	語数	割合
「状」のみ	1,433	74.6%	292	79.8%
「様」と重なる	488	25.4%	74	20.2%
計	1,921	100.0%	366	100.0%

表9 「様」の前接語の内訳

前接語	延べ		異なり	
	語数	割合	語数	割合
「様」のみ	1,055	78.7%	320	81.2%
「状」と重なる	285	21.3%	74	18.8%
計	1,340	100.0%	394	100.0%

以下、「状」のみあるいは「様」のみに見られた前接語、「状」と「様」のいずれにも見られた前接語を見ていく。

まず、表 10 と表 11 は「状」のみ、あるいは「様」のみに見られた前接語のうち、頻度 15 以上のものを挙げた。表 11 の「様」のみに見られた前接語の特徴としては、「感冒」や「インフルエンザ」「～病」などの具体的な疾患名が見られることが挙げられる。「様」の前接語で最も頻度の高い「症」の内訳を見ると、頻度が高い順に「心筋症」15 例、「失調症」9 例、「狭心症」8 例、「認知症」8 例と、疾患名が用いられていた。一方、表 10 の「状」の前接語に疾患名は見られない。また、表 11 の「様」の前接語「満月」はすべて「満月様顔貌」であり、「イヌボエ (書字形：犬吠)」は 15 例のうち、14 例が「犬吠様 (の) 咳、犬吠様咳嗽」であった。「犬吠」のように、視覚的に特徴がとらえられない語が用いられるのも「様」の特徴と言える。(6) ～ (8) は「感冒様」「心筋症様」「犬吠様」の例である。

- (6) ウイルス性心筋炎では発症の 1 ～ 2 週間前に感冒様症状がみられることが多い。(診断_循環器疾患)
- (7) C型肝炎ウイルスにより肥大型心筋症様病態をきたす可能性が示唆されている。(内科_循環器疾患)
- (8) 小児に起こる吸気時喘鳴，息切れ，嘔声を伴った犬吠様のせきが特徴である。(内科_呼吸器疾患)

表 10 「状」のみに見られた前接語

(頻度 15 以上、N=1,433)

順位	語彙素	頻度	順位	語彙素	頻度
1	粥	64	15	鞘	21
2	顆粒	55	16	撥	20
3	海綿	50	16	紡錘	20
4	索	48	16	浮腫	20
5	尾	42	16	地図	20
6	乳頭	35	20	樹枝	19
7	線	34	20	ニュウトウ	19
8	泡沫	32	22	放射	17
9	巢	30	22	雲	17
10	囊	27	24	半球	15
11	鎌	24	24	鋸齒	15
12	円板	23	24	兔糞	15
13	リング	22	24	蜘蛛	15
13	板	22			

表 11 「様」のみに見られた前接語

(頻度 15 以上、N=1,055)

順位	語彙素	頻度
1	症	73
2	感冒	46
3	腺腫	40
4	インフルエンザ	37
5	病	30
6	アナフィラキシー	23
7	喘息	22
8	インシュリン	19
9	満月	16
10	イヌボエ	15

「状」と「様」の前接語を比べると、「状」は形を表す語が目立ち、質的な特徴を表す語も見られる。他方、「様」は、疾患名をはじめ、対象を全体的にとらえる語が目立つ。「様」の前接語に疾患名が見られる用法について、三枝・本多 (2021) では、医師が患者の疾患を推定する際に用いられる、医学に特徴的な用法であることが述べられている。本調査も、そ

の指摘と重なる結果だと思われる。

次に、表 12 は、「状」と「様」の両者に重なって見られた前接語 74 語のうち、「状」と「様」の頻度の合計が 10 以上の 21 語である。全体的に「嚢胞」「結節」「腫瘤」「腫瘍」「瘤」「腫」「ポリープ」「嚢腫」「水腫」「疣」「憩室」など、部位が変化した形を表す語や、病変に関係する語がみられる。また、「ソーセージ」「風船」も含め、膨らみのある形をしたものを表す語が目立つ。

頻度を見ると、21 語のうち、「状」「様」いずれかの頻度が 1 の前接語が 10 語と約半数を占める。このことは、「状」と「様」の両者に用いられる語も「状」か「様」で使い分けられる傾向が見られる。

表 12 「状」と「様」の両者に見られた前接語

（「状」と「様」の合計頻度降順）

順位	語彙素	状	様	計
1	S	85	2	87
2	嚢胞	55	8	63
3	ガラス	14	37	51
3	結節	48	3	51
5	炎	1	37	38
6	腫瘤	18	9	27
7	腫瘍	4	21	25
8	編目	21	1	22
9	ポリープ	18	3	21
9	腫	6	15	21
11	瘤	12	8	20

順位	語彙素	状	様	計
12	苺	18	1	19
13	ゼリー	17	1	18
14	ソーセージ	1	16	17
14	嚢腫	14	3	17
16	水腫	3	11	14
16	疣	13	1	14
18	房	11	1	12
19	憩室	1	10	11
19	字	10	1	11
21	風船	1	9	10

表 12 で「状」と「様」の頻度がいずれも 5 以上の語は「嚢胞」「ガラス」「腫瘤」「瘤」「腫」の 5 語であった。この 5 語について、さらに前後の語を見てみると、「ガラス」では、「すりガラス状陰影」12 例、「すりガラス様陰影」24 例のように、同じ語で「状」と「様」が使われている例も見られたが、多くは (9) の「嚢胞状に～」や (10) の「嚢胞様構造」のように異なる語に用いられていた。

- (9) III型(congenital choledochocoele)は十二指腸壁内の胆管が嚢胞状に拡張し十二指腸内腔へ突出したものである。(内科_肝・胆・膵疾患)
- (10) 気体あるいは液体内容をもつ肺内の嚢胞あるいは嚢胞様構造を指し、結核・腫瘍・寄生虫・真菌などによる原因の明らかな空洞を除くもの(イヤート_呼吸器疾患)

なお、表 12 の語を表 10 の「状」のみで見られる語と比べると、表 12 には部位が変化した形を表す語や、病変に関係する語が目立つ。これは、「状」の持つ形をとるという特徴と、「様」の持つ具体的な疾患名を推定する際に用いられるという特徴の間に位置するものではないかと思われるが、詳細な検討は今後の課題としたい。

5. 2 「状」と「様」の後接語の比較

表 13 は、「様」の後接語を高頻度順にまとめたものである。延べ語数は 1,340 語、異なり語数は 186 語である。先の「状」の後接語を示した表 3 と比較すると、助詞「の」の頻度が最も高い点は共通している。ただし、「様」の後接語のほうが割合は高い。また「状」では 10.3% と高頻度であった「に」が「様」では 3.4% と低く、補助記号は 12 位までには現れず、動詞を修飾する用法や並列用法等が少ないことが見てとれる。すなわち、「様」は名詞が後接する割合が「状」より高いと言える。さらに、「様の」に後接する名詞は「症状」が 42 例と最も多く、「様」の後接名詞は、「の」を介する場合介さない場合を合わせて「症状」が 226 例と最も多い。一方、「状」と「症状」の組み合わせは一例もない。このことから「様」は「症状」という広範な意味を持つ語を修飾しうる全体的なありさまを示す語であり、「状」はあくまで具体的な形を形容するために使われる語と考えられる。

表 13 接尾辞「様」の後接語 (頻度 22 以上、N=1,340)

順位	語彙素	頻度	割合
1	の	283	21.1%
2	症状	184	13.7%
3	顔貌	48	3.6%
4	に	45	3.4%
5	皮疹	37	2.8%
6	陰影	31	2.3%
7	変化	29	2.2%
8	物質	27	2.0%
9	病変	26	1.9%
10	発疹	24	1.8%
11	作用	23	1.7%
12	所見	22	1.6%

6. まとめ

本研究では医学書テキストにおける「たとえる表現」の一端を明らかにする目的で接尾辞「状」に注目し、用法を分析した。その結果を以下の 5 点にまとめる。

1. 「状」の前接語には「S、線」など形そのものを表すもの、「粥、海綿」など質的な様子も表す語、「嚢胞、結節」などの部位が変化した形を表す語が多い。
2. 「状」の後接語で最も頻度が高いのは名詞であったが、助詞の「の」を介した名詞接続も少なくない。また、「に、と、を」を介した動詞修飾用法や読点や接続詞を用いた並列用法も見られる。
3. 「状」の後接語には「血管、結腸」などの体の部位のほか、「印影」「突起」「変化」「拡張」といった形や変化を表す一般語が使われている。
4. 書籍の NDC 分類記号を用いて分析した結果、「状」は医学において多用される語であ

ることが示唆された。

5. 同じ「たとえる表現」の接尾辞「様(ヨウ)」と比較した結果、「状」は形を表す語が多く用いられる傾向があるのに対し、「様」の前接語には疾患名、後接語には「症状」のような全体の状態やありさまを表す語が多く見られ、用法の違いが観察された。

医学書テキストに用いられるたとえる表現には、「状」「様」のほかにも、「一感」(イライラ感)、「一形」(三日月形)、「一大」(小豆大)などが見られる。今後は、これらの表現の分析も行い、医学書テキストにおけるたとえる表現の実態を明らかにしていきたい。

謝 辞

本研究は JSPS 科研費 JP18H00679 の助成を受けたものです。

文 献

内山清子・岡照晃・東条佳奈・小野正子・山崎誠・相良かおる(2018)「実践医療用語の語構成要素抽出の試み」『言語資源活用ワークショップ発表論文集』3、pp.463-467.

doi/10.15084/00001680

加藤 祥・森山 奈々美・浅原 正幸 (2021)「現代日本語書き言葉均衡コーパス」書籍サンプルの NDC 情報増補 : NDC 情報を用いた随筆の抽出と文体調査」『国立国語研究所論集』

(21) pp.65-84.doi/10.15084/00003437

国立国語研究所 (2015)「『現代日本語書き言葉均衡コーパス』利用の手引き第 1.1 版」

<https://clrd.ninjal.ac.jp/bccwj/doc.html>

三枝令子・本多由美子 (2021)「医学書テキストのたとえる表現—「よな」、接尾辞「様(ヨウ)」の特徴」『言語資源ワークショップ発表論文集』(6) pp.178-189.国立国語研究所.

doi/10.15084/00003491

三枝令子・丸山岳彦・松下達彦・品川なぎさ・稲田朋晃・山本一晃・石川和信・小林元・遠藤織枝・庵功雄 (2020)『医学用語の収集と分類』「日本語教育」176、pp.34-47.

東条佳奈・麻子軒・相良かおる・高崎智子・山崎誠 (2020)「病名における「-性」の分析 : 一般書籍との比較から」『言語資源活用ワークショップ発表論文集』5、pp.357-364.

doi/10.15084/00003175

野村雅昭 (1978)「接辞性字音語基の性格」『電子計算機による国語研究 9』pp.102-138.

doi/10.15084/00001057

調査資料

伊藤正男・井村裕夫・高久史磨 総編集 (2010)『医学書院医学大辞典 第2版』

岡庭豊・荒瀬康司・三角和雄 (2019)『イヤート 2020 内科・外科編』メディック・メディア

金澤一郎, 永井良三 総編集 (2015)『今日の診断指針 第7版』医学書院

高久史磨・尾形悦郎・黒川清・矢崎義雄 監修 (2010)『新臨床内科学 第9版』医学書院

福井次矢・高木誠・小室一成 総編集 (2018)『今日の治療指針 2018年版』医学書院

水口雅・市橋光・崎山弘 総編集 (2015)『今日の小児治療指針 第16版』医学書院

『デジタル大辞泉』小学館、「ジャパンナレッジ」利用 (2022年5月31日閲覧)

『日本国語大辞典』小学館、「ジャパンナレッジ」利用 (2022年5月31日閲覧)

関連 URL

加藤祥・森山奈々美・浅原正幸「BCCWJ-NDC」<https://github.com/masayu-a/BCCWJ-NDC> (2022年6月17日閲覧)

国立国語研究所 「「中納言」版 BCCWJ/短単位語数」<https://clrd.ninjal.ac.jp/bccwj/bcc-chu-suw.html>

打ち合わせにおける談話構造の修辞機能からの分析

田中 弥生 (国立国語研究所 研究系) *

An Analysis of Discourse Structure of a Meeting from a Rhetorical Function Point of View.

Yayoi TANAKA (National Institute for Japanese Language and Linguistics)

要旨

本研究は、修辞機能分析の分類法による日常会話分析の一環として、打ち合わせの談話の談話構造を修辞機能から検討するものである。修辞機能とは、「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、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。本発表では、『日本語日常会話コーパス』に収録されている地域活動ボランティアによるイベント企画の打ち合わせ談話を対象に、イベント企画にかかわる会場、日程、募集方法などの検討の段階について、修辞機能から談話構造を検討した。分析の結果、話題内容が変わるきっかけの発話の修辞機能の特徴が見られること、全体的に頻度の高い修辞機能があること、話題内容によって用いられる修辞機能があることがわかった。

1. はじめに

打ち合わせの談話に関して、職場の談話における文末表現、疑問表現、敬語、自称詞、談話進行や談話構造、談話管理など様々な観点からの分析が行われている(渡邊 2014a,b, 現代日本語研究会 2011)。また、雑談や相談など様々な談話の構造分析が行われている(ザトラウスキー 1993, 鈴木 2002, 筒井 2012)。本研究は、地域活動のボランティアによるイベント企画打ち合わせの談話について、修辞機能の観点から分析し、談話構造を検討するものである。修辞機能とは、いわゆる文章の技巧や説得法のレトリックではなく、ここでは、「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、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。この修辞機能は脱文脈の度合いと関連づけられている。ここでの文脈とは話者の「いま・ここ・わたし」とし、脱文脈度は「コミュニケーションが行われている時空とその発話内容との時間的・空間的距離の程度」とする。分析には、修辞機能分析の分類法⁽¹⁾を用いる。この分類法は、修辞ユニット分析(佐野 2010, 佐野・小磯 2011)を元に、日本語文法の枠組みで修正を加えたものである。テキストの分析単位(概ね、日本語文法の節に相当)ごとに、述部と主語・主題の分類から修辞機能が特定される。修辞機能の脱文脈度の高低により、一般的な内容か個人的な内容か、発話の時空に依存しない内容か依存する内容か、抽象的なことか具体的なことかなどを示すことができる。例えば、子供は目の前のことから話せるよ

* yayoi@ninja.ac.jp

(1) 現在手順書を執筆中である。

うになり、成長とともに過去のことや、明日のこと、その場にはいないおばあちゃんのことなど、時空を離れた発話ができるようになる。「それ、ちょうだい」と同じ時空にいる親に言うのは文脈化しており、「はやぶさは北海道新幹線だよ」は同じ時空にいなくても伝えられるため脱文脈度が高い。

これまで、児童作文、家族の談話、相談の談話、高齢者グループの談話など（田中ほか 2021, 田中・小磯 2020, 田中 2017, 田中ほか 2022）の分析から、目的や話題内容、状況によって、用いられる修辞機能が異なり、脱文脈度は推移することが明らかになっている。田中ほか（2020）では、手順を説明する談話における、説明の各ステップにおいて、目の前の手順説明に関わる修辞機能が用いられながらも、合間に個人の習慣や過去の経験や一般的なことが話され、脱文脈度が様々に推移することが明らかになっている。

本発表では、『日本語日常会話コーパス』に含まれる、公共施設での地域活動の仲間たちとのイベント企画打合せの談話を分析対象とし、それぞれの発話について、言語表現から修辞機能と脱文脈度を特定し、打ち合わせの話題内容によって、どのように発話が推移しているか検討を行う。以下、第2節で分析データと分析方法について説明し、第3節で分析結果と考察を述べ、第4節でまとめと今後の課題について述べる。

2. 分析データと分析方法

2.1 分析データ

本研究の分析対象として、『日本語日常会話コーパス (Corpus of Everyday Japanese Conversation, CEJC)』（小磯ほか 2022）に収録されている公共施設での地域活動の仲間たちとイベントの打合せ（T004 _ 017 17分、参加者7名）の書き起こしデータを用いる。イベントはボランティアの会を中心とする実行委員会と施設の管理者が協力して企画・運営されるもので、本データの参加者の1名が管理者で、それ以外はボランティアの会のメンバーである。

2.2 分析方法

修辞機能分析は、Rhetorical Unit Analysis(Cloran 1994, 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 2011)を元に、日本語文法の枠組みで修正を加えた分類法である。分析手順は、次のとおりである。

1. 分析単位（メッセージ）に分割し、分析対象を特定する。
2. 分析対象のメッセージについて発話機能を分類する。
3. 発話機能が「命題」のメッセージについて、時間要素と空間要素を分類する。
4. 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。

以下に手順の概要を示す。

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するが、連体修飾節は独立したメッセージとして扱わない。メッセージは「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。

2.2.2 発話機能・時間要素・空間要素

メッセージの種類が「主節」「並列」「引用」に分類されたメッセージについて、発話機能・時間要素・空間要素を分類する。表1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⁾。

表1 発話機能・時間要素・空間要素からの修辞機能と脱文脈指数の特定

定義	↑ 高 ↑空間的距離のレベル ↓ 低						一般化 [14]
状況外		報告 [9]	状況外回想 [10]	予測 [11]		推量 [12]	説明 [13]
状況内		↓ 参加	実況 [2]	状況内回想 [3]	状況内予想 [5]		観測 [8]
参加					計画 [4]	状況内推測 [6]	
空間要素	今ここわたし	← 低 ← 時間的距離のレベル → 高					
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				

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相手に働きかけたり、会話者同士の行為にかかわる発話内容が該当し、【行動】[1]と特定される。例えば、同じ時空にいる相手への「このトマト見て!」「お醤油を取って」のような行為や物を要求する場合である。「命題」は、情報を交換する陳述・質問で、「私はトマトが大好き」「このトマトは真っ赤だね」「トマトはナス科の植物だ」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る。図1に時間要素の分類を示す。「習慣・恒久」⁽³⁾「現在」「過去」「未来意志的」「未来非意志的」「仮定」に分類する。「太郎がトマトを美味しそうに食べている」は「現在」、「昨日食べたトマトは美味しかった」は「過去」、「来年はトマトを育てよう」は「未来意志的」、「私はトマトが大好き」は嗜好であるため「習慣・恒久」、「トマトはナス科の植物だ」は恒久的と判断し「習慣・恒久」に分類する。

空間要素は、話者のいる場所を基準として、メッセージの中心との空間的距離を示す要素で、主語、主題、述部の主体から判断する。図2に空間要素の分類を示す。「参加」「状況内」⁽⁴⁾「状況外」「定義」に分類する。「私」「あなた」が主語であれば「参加」、「太郎がトマトを美味しそうに食べている」は太郎が話者と同じ時空にいると考えられるので「状況内」、「昨日食べたト

(2) 以下、修辞機能を【】で、脱文脈指数を[]で示す

(3) 「習慣・恒久」には、属性、嗜好、評価も含む。

(4) 「状況内」には、話者の身体や所有物、思想、また、その談話の中で話題になっている事柄も該当すると考える。

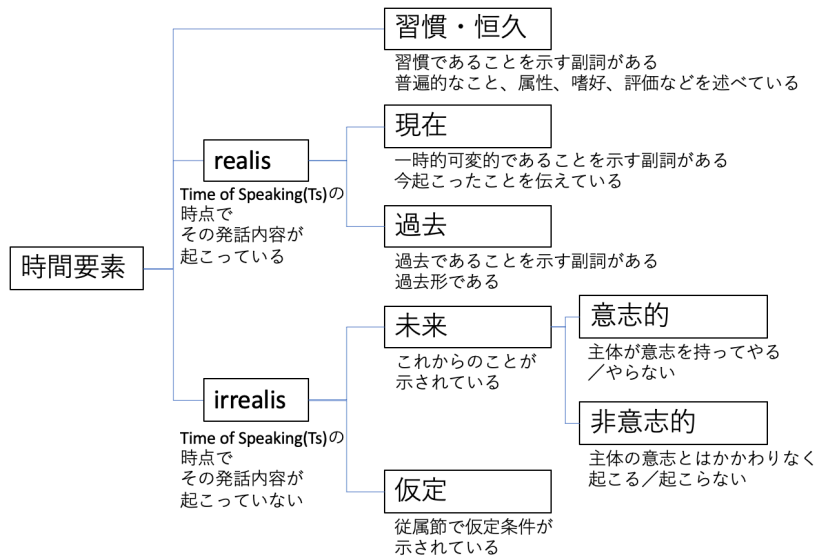


図1 時間要素の分類

「トマトは美味しかった」「トマトはあっちの八百屋のが美味しい」の「トマト」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」。「トマトはナス科の植物だ」は「トマト」という植物の性質一般を述べているので「定義」に分類する。

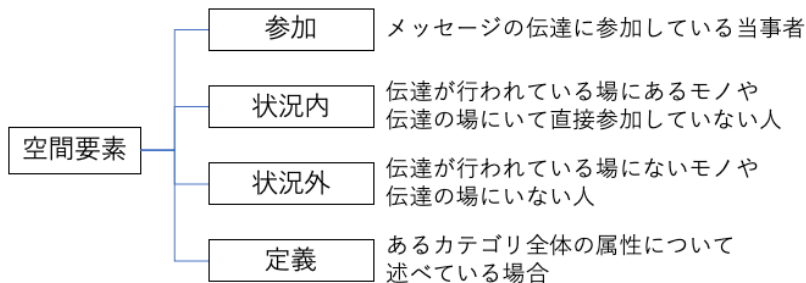


図2 空間要素の分類

2.2.3 修辞機能と脱文脈化度の特定

表1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。

3. 分析結果と考察

参加者ごとのメッセージ数を図3に示す。分析対象となる主節と並列は一ノ宮、塚田、土井の順で多く、他の参加者は相槌など、定型句類の使用が多いか発話が少ない。

分析対象の談話は、その話題内容から、1. 会場の検討、2. 日程検討、3. 募集、4. 原稿、5. 実地調査に分けられる。表2および図4に、話題内容ごとの修辞機能の頻度と割合を示す。全体

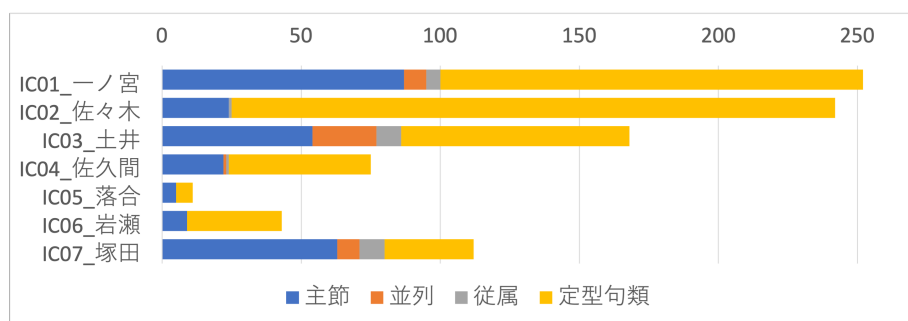


図3 参加者ごとのメッセージ数

では【説明】[13]と【観測】[8]が使用されており、1. 会場の検討と4. 原稿では同数なのに対して、3. 募集では【観測】[8]が【説明】[13]の約2倍、5. 実地調査では反対に【説明】[13]が【観測】[8]の約2倍となっている。また、1. 会場の検討と5. 実地調査では【自己記述】[7]が用いられている。

表2 話題内容ごとの修辞機能 [脱文脈指数]

修辞機能 [脱文脈指数]	1 会場の検討	2 日程検討	3 募集	4 原稿	5 実地調査	計
一般化 [14]	0	0	0	0	0	0
説明 [13]	14	5	17	9	37	82
推量 [12]	0	0	0	0	0	0
予測 [11]	1	2	0	2	7	12
状況外回想 [10]	7	0	6	0	0	13
報告 [9]	3	0	0	0	0	3
観測 [8]	14	7	33	9	17	80
自己記述 [7]	12	2	10	0	20	44
状況内推測 [6]	0	1	1	0	1	3
状況内予想 [5]	0	1	10	4	4	19
計画 [4]	7	0	5	3	8	23
状況内回想 [3]	3	1	4	0	0	8
実況 [2]	1	2	0	2	4	9
行動 [1]	0	0	0	0	1	1
計	62	21	86	29	99	297

話題内容ごと参加者別の修辞機能の出現を図5に示す。

図5からも、一ノ宮、土井、塚田の発話が多いことが確認できる。また、話題内容によって、会話に参加する人数が異なっていることがわかる⁽⁵⁾。

以下で、各話題内容ごとに、修辞機能の出現を見ていく。

⁽⁵⁾ この集計には図4の定型句類(相槌をうったり、「そうですね」と応答している発話)は含まれていない。

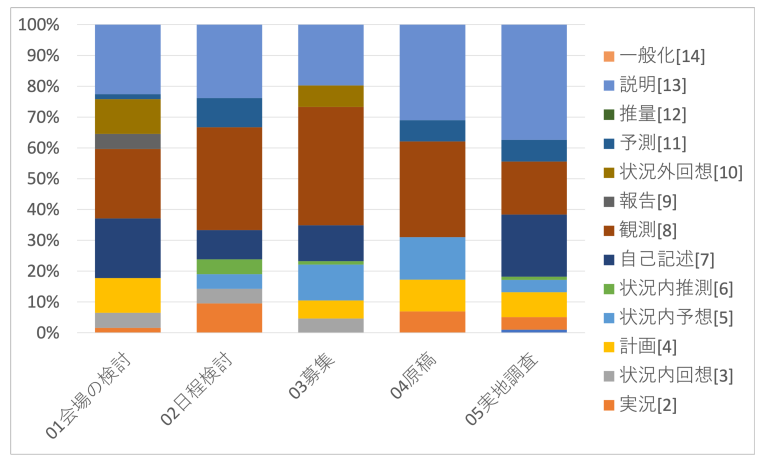


図4 話題内容ごとの修辞機能の出現割合

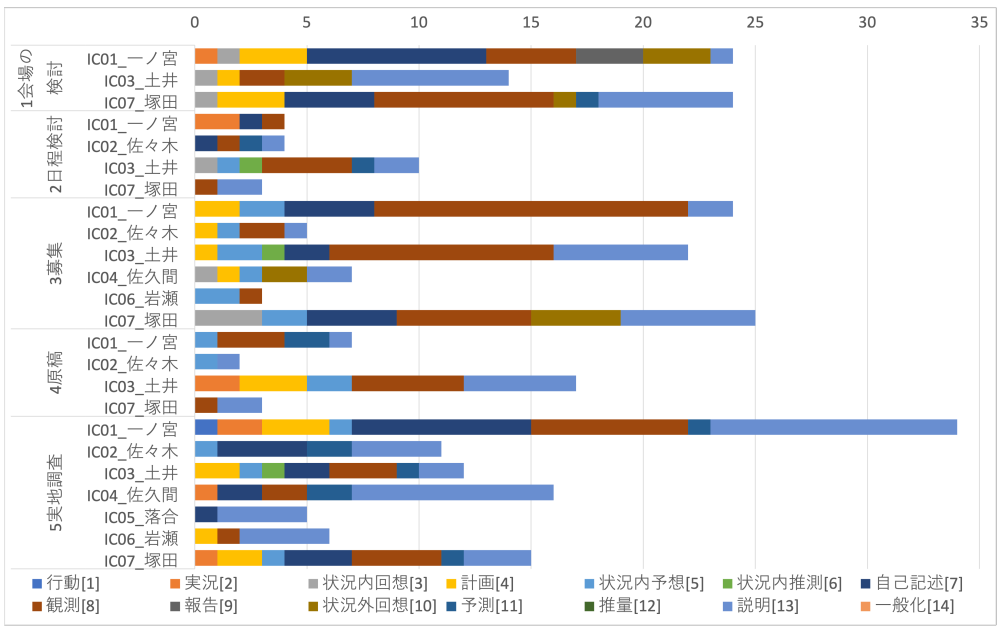


図5 話題内容ごと参加者別の修辞機能の出現

3.1 話題内容ごとの会話の流れと修辞機能

3.1.1 1. 会場の検討

1. 会場の検討の冒頭を、表3に示す。この話題内容の前に、このデータの始まりの部分として、「でもこれで（略）大体コースの形としては収まりそうな感じですね」「そうですね」という会話がある。その後、「それで」という話題を変える接続詞につづいて、それまで話していたコースの最後の場所の確認の発話があり、どこにするか、どのように予約するか、というやりとりが交わされる。なお、図5で示したように、1. 会場の検討では、実質的な発話は、一ノ宮、土井、塚田の3名によって交わされている。

冒頭の2つのメッセージについて、(1)に発話機能・時間要素・空間要素を示す。

表3 1 会場の検討の冒頭部分

発話者	発話	修辞機能 [脱文脈度]
IC01 一ノ宮	それで(R 第二)の区活: でラストでいいのかな:。	観測 [08]
IC07 塚田	いんじゃない?。楽で。	観測 [08]
IC02 佐々木	うん。	
IC03 土井	うーん。	
IC07 塚田	うん。	
IC03 土井	会場、最後までしてもね、鷺宮の時もそうでしたけど。やっぱ会場がね 最後がちがちになっちゃうと。	状況外回想 [10]
IC07 塚田	例の紙芝居やりましたよね?。	状況内回想 [03]
IC02 佐々木	うん うん。	
IC01 一ノ宮	やった。	状況内回想 [3]
IC02 佐々木	うん うん。	
IC01 一ノ宮	そう。	
IC07 塚田	一番いいですね?。	観測 [8]
IC02 佐々木	そうですね。	
IC03 土井	うーん。	
IC01 一ノ宮	えーと (R 峰松) さんが取ってくれたんだけど:。	状況外回想 [10]
IC07 塚田	そう そう そう。	
IC01 一ノ宮	今回は: だいじょぶかしら。	観測 [8]
IC07 塚田	え。それはだって 誰に取ってもらえるってゆってもいっぱいいるじゃん。	説明 [13]
IC01 一ノ宮	(R 第二)?。	観測 [8]
IC07 塚田	うん。	
IC03 土井	二カ月前: の抽選: 会に行くって感じ: ですか。	説明 [13]

(1) IC01 一ノ宮 それで (R⁽⁶⁾ 第二) の区活⁽⁷⁾でラストでいいのかな。

→ (命題&習慣・恒久&状況内 → 【観測】 [8])

IC07 塚田 いんじゃない?。楽で。 → (命題&習慣・恒久&状況内 → 【観測】 [8])

いずれも空間要素は省略されているが、「コースの内容は」のような語句が復元できると考える。打ち合わせの中で話題にしていることであるため、状況内に分類した。

この一ノ宮と塚田のやりとりを聞いて、管理者である土井は、過去の出来事を話し始め、塚田、一ノ宮も過去の出来事を話し、また、「今回はだいじょうぶかしら」と今回のイベントに話が戻る。この後、どのように申し込むか、誰に頼めるか、というやりとりが続くが、土井は管理者という立場から、施設側でどのような動きがあるかについて、主に【説明】 [13] で述べ、その中に、【状況外回想】 [10] 【状況内回想】 [3] で過去の出来事も交えながら伝えている。また一

(6) R は仮名を表す

(7) 区の施設。区民活動センター。

ノ宮と塚田は、誰に頼むのがいいかについて、【観測】[8]や【自己記述】[7]で話し合っている。表4に1. 会場の検討の終了部分の会話を示す。

表4 1. 会場の検討の終了部分

発話者	発話	修辞機能 [脱文脈度]
IC07 塚田	もう二カ月前に 抽選だから	説明 [13]
IC07 塚田	その前に押さえておかないとまずいんだよね。	観測 [8]
IC02 佐々木	二カ月前か。	
IC01 一ノ宮	うん。	
IC03 土井	そうですね そうですね。	
IC01 一ノ宮	そうだよね。	
IC02 佐々木	うーん。	

この話題は、塚田による「もう二カ月前に抽選だから」【説明】[13]「その前に押さえておかないとまずいんだよね。」【観測】[8]が、次の話題へとうつるきっかけとなる。

3.1.2 2. 日程検討

表5に2. 日程検討の冒頭部分の会話を示す。

表5 2. 日程検討の冒頭部分

発話者	発話	修辞機能 [脱文脈度]
IC03 土井	だからこれ えっと 日にち的には 一応十一月のどこかって感じですよ。	観測 [8]
IC01 一ノ宮	だ日 日程 決めたいよね。	実況 [2]
IC07 塚田	そうですね。	
IC03 土井	そうすると: 九月のたぶん一日ぐらいが抽選日になるから	予測 [11]
IC03 土井	その前ってことですよ。	観測 [08]

塚田の「もう二カ月前に抽選だから」【説明】[13]「その前に押さえておかないとまずいんだよね。」【観測】[8]を受けて、土井が「だからこれ えっと 日にち的には 一応十一月のどこかって感じですよ。」【観測】[8]と、具体的な日程の話題を始める。ほぼ同時に一ノ宮は「日程決めたいよね」【実況】[2]と、参加者共通の今するべきことについて述べている。続いて、土井が【予測】[11]【観測】[8]、塚田が【観測】[8]、土井が【状況内推測】[6]と見込みを述べ、また土井が【状況内回想】[3]で過去の実績を述べる。

この話題内容ではこの後、町巡り博覧会という他組織の運営する企画の影響について【説明】[13]で交わされたあと、3. 募集の話題にうつる。

3.1.3 3. 募集

表6に3. 募集の冒頭部分の会話を示す。土井の「あとはまあ募集の時期ですね」【観測】[8]という発話をきっかけに、話題内容が募集にうつり、時期や定員についてのやりとりが交わされる。一ノ宮と土井を中心とする【観測】[8]と、土井と塚田を中心とする【説明】[13]、前年度の状況【状況内回想】[3]なども交え、募集について話し合われる。

表 6 3. 募集 の冒頭部分

発話者	発話	修辞機能 [脱文脈度]
IC03 土井	あとまあ	観測 [8]
IC02 佐々木	そうですね。	
IC03 土井	募集の時期ですね。	
IC01 一ノ宮	大丈夫?。	
IC02 佐々木	募集も。	
IC02 佐々木	うーん。	
IC01 一ノ宮	うん。	
IC03 土井	とりあえず十一月: の:	
IC03 土井	えー あでもまあ。やっぱ 今回も 三十名	
IC03 土井	二十名 三十名ぐらい。	
IC01 一ノ宮	どうしよう。	観測 [8]
IC07 塚田	二十名。	
IC02 佐々木	二十名ですかね:。	

3.1.4 4. 原稿

表 7 に 4. 原稿 の冒頭部分の会話を示す。

表 7 4. 原稿 の冒頭部分

発話者	発話	修辞機能 [脱文脈度]
IC01 一ノ宮	と 原稿いつまでですか?:?	観測 [8]
IC03 土井	あ。えーとですね。	観測 [8]
IC03 土井	一応 九月の 頭で	
IC03 土井	あの ハ	
IC03 土井	えー なんだ 十月号にはこうゆうのを載せたいってゆう。	
IC03 土井	あ。違う。	
IC07 塚田	八月でしょう。	観測 [8]
IC03 土井	ごめんなさい。	計画 [4]
IC02 佐々木	うん。	
IC03 土井	えーっと 十月号に載せるから。	
IC03 土井	そうですね。	
IC03 土井	八月ですね。	
IC03 土井	八月の初旬になりますね。	
IC03 土井		

一ノ宮の「原稿いつまでですか?」【観測】[8] という発話から、話題が原稿のことにうつっていく。原稿を書くのは一ノ宮で、管理者の土井は、【説明】[13] によって原稿の締め切りやスペースのことなど情報を提供している発話が見られた。

3.1.5 5. 実地調査

表 8 に 5. 実地調査 の冒頭部分の会話を示す。

一ノ宮の「じゃ とりあえず七月の 集まる のはいつにしましょうか」【状況内予想】[5] という発話がきっかけになり、話題がイベントの事前の実地調査の日程調整に移った。「これ

表 8 5. 実地調査 の冒頭部分

発話者	発話	修辞機能 [脱文脈度]
IC01 一ノ宮	じゃ とりあえず七月の 集まる	
IC05 落合	うん。	
IC04 佐久間	うん。	
IC01 一ノ宮	これ歩きだよね。また。	観測 [08]
IC01 一ノ宮	あのは いつ: に しましょう。	状況内予想 [5]
IC02 佐々木	そうですね。	
IC04 佐久間	暑いな。	説明 [13]
IC02 佐々木	暑く。	説明 [13]
IC06 岩瀬	暑いな。	説明 [13]
IC01 一ノ宮	暑いな。	説明 [13]
IC02 佐々木	時間帯と。	
IC07 塚田	夜にしますか?。	計画 [04]

歩きだよね」【観測】[8]の自己確認が途中に含まれている。この話題内容では、【説明】[13]と【自己記述】[7]が他の話題内容よりも多い割合で用いられている。【説明】[13]は「(七月に歩くのは)暑い」や、「(朝は気温が)まだ涼しい」など、実地調査をするときの気候についての発話があり、【自己記述】[7]は、「午後だったら十六でもいいけど」「わたし八時からでもいいですよ」や「ずっと連休だから」など各自の日程に関する発話である。

3.2 談話構造

第3.1節で示したように、それぞれの話題内容では、話題の切り替えの際にきっかけとなる発話があった。1. 会場の検討では、コースの最後の場所を確認する発話が該当し、2. 日程検討では、「だからこれ日にち的には一応十一月のどこかって感じですよ」、3. 募集では、「あとまあ、募集の時期ですね」、4. 原稿では「原稿いつまでですか」と、いずれも【観測】[8]によって話題内容の切り替えがおこなわれていた。ただし、5. 実地調査については、「じゃ とりあえず七月の集まるのはいつにしましょう」【状況内予想】[5]であった。

1. 会場の検討 2. 日程検討 3. 募集 4. 原稿の【観測】[8]は、時間要素が習慣・恒久で、空間要素が状況内という組み合わせで特定される修辞機能である。本研究の分析対象データのような恒例のイベントの企画打ち合わせでは、「(イベントの)会場は」「(イベントの)日程は」「われわれが検討すべきことは」などが参加者が共通して理解していることとして「状況内」に分類され、それらがどういう状況であるかを確認したり述べていることから【観測】[8]が特定されている。

一方5. 実地調査の【状況内予想】[5]は「(事前の実地調査として)7月に集まる」という参加者が共通して理解している「状況内」のことについて、「いつにしましょうか」という参加者の意志にかかわる表現であるため【状況内予想】[5]が特定されている。これは、図5に示されているように、5. 実地調査では他の話題内容と異なり、打ち合わせ参加者全員が発話している

ことと関連し、5つの話題内容の中では、最も積極的に参加が求められる話題内容であるためとも考えられる。

冒頭部分以外については、【観測】[8]と【説明】[13]が分析データの基本となる修辞機能で、それ以外の修辞機能の出現は、話題内容に関わると考えられる。1. 会場の検討では、会場の予約について、誰に頼むのがいいか、頼めるか、知り合いがいるかについて、【自己記述】[7]や【計画】[4]によって話されている。2. 日程検討では、過去のイベントからの今回のイベントへの【予測】[11]【状況内予想】[5]などが用いられている。3. 募集では、話題に出た「町巡り博覧会」への応募と過去の実績について【自己記述】[7]【状況内回想】[3]【状況外回想】[10]によって話されている。4. 原稿では、提出や掲載の計画【計画】[4]や原稿が間に合うか【状況内予想】[5]などが話されている。5. 実地調査では、自身の都合を述べる【自己記述】[7]が多く用いられている。

このことから、分析対象の打ち合わせ談話の話題内容の中では、次のような構造があることがうかがえた。

〈【観測】[8]/【状況内予想】[5] → 【説明】[13]・【観測】[8] (・話題内容に応じた修辞機能)〉

話題の始まりでは、【観測】[8]か【状況内予想】[5]で話題が提示され、それ以降は、【説明】[13]と【観測】[8]という基本的な修辞機能に加え、話題内容に応じた修辞機能が用いられる、というものである。

4. おわりに

本稿では、イベント企画打ち合わせ時の談話データを対象として、それぞれの発話について言語表現から修辞機能と脱文脈度を特定し、談話の話題内容によってどのような修辞機能が用いられているか、その話題の中で特徴的な構造がみられるかを確認した。分析の結果、分析対象の談話で用いられる修辞機能は、【観測】[8]【説明】[13]が基本となっているが、1. 会場の検討、2. 日程検討、3. 募集、4. 原稿、5. 実地調査という話題内容によって、頻度は異なり、それ以外の修辞機能の使用も異なることがわかった。また、話題の切り替え時に、5. 実地調査以外で【観測】[8]、5. 実地調査では【状況内予想】[5]が用いられ、その違いは、参加者の意志の有無であることから、話題内容と打ち合わせ参加者の関わりによって異なることがうかがえた。本稿では、一つの打ち合わせデータの分析であるため、このことがイベント企画打ち合わせの特徴であるかについては、今後他の打ち合わせ談話を分析することによってさらに確認していきたい。

謝 辞

本研究は国立国語研究所のプロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」、科研費基盤 (B)(特設分野研究)(18KT0035)、および科研費基盤 (C)(19K00588) によるものです。

文 献

- 渡邊ゆかり (2014a). 「少人数企画会議における、とりまとめ役の談話管理スタイル1 – とりまとめ役の発話単位量とムーブの観点から –」 広島女学院大学日本文学, 24, pp. 1–23.
- 渡邊ゆかり (2014b). 「少人数企画会議における、とりまとめ役の談話管理スタイル2 – とりまとめ役のターン譲渡と獲得の仕方の観点から –」 広島女学院大学国語国文学誌, 44, pp. 1–29.
- 現代日本語研究会 (2011). 『合本女性のことば・男性のことば (職場編):』 ひつじ書房
- ザトラウスキーポリー (1993). 『日本語の談話の構造分析: 勧誘のストラテジーの考察』 くろしお出版, 東京
- 鈴木香子 (2002). 「ラジオの医療相談の談話の構造分析」 早稲田大学日本語教育研究:1, pp. 117–130.
- 筒井佐代 (2012). 『雑談の構造分析』 くろしお出版, 東京
- 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 (システムック理論) における談話分析 (修辞機能編)』.
- 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・佐尾ちとせ・宮城信 (2021). 「児童作文の評価に向けた脱文脈化観点からの検討」 言語処理学会 第27回年次大会 発表論文集, pp. 750–755.
- 田中弥生・小磯花絵 (2020). 「家庭での幼児の発話の修辞機能: 脱文脈化の観点からの検討」 言語資源活用ワークショップ発表論文集, 4, pp. 106–118.
- 田中弥生 (2017). 「相談における談話構造: 修辞機能と脱文脈化の観点からの分析」 言語資源活用ワークショップ発表論文集, 1, pp. 69–78.
- 田中弥生・小磯花絵・大武美保子 (2022). 「脱文脈化の観点から見た共想法に基づく高齢者談話の分析」 国立国語研究所論集:22, pp. 137–155.
- 田中弥生・浅原正幸・小磯花絵 (2020). 「手順説明談話における脱文脈化の諸相」 言語処理学会第26回年次大会発表論文集, pp. 720–723.
- 小磯花絵・天谷晴香・石本祐一・居關友里子・白田泰如・柏野和佳子・良子川端・田中弥生・伝康晴・西川賢哉・渡邊友香 (2022). 『『日本語日常会話コーパス』設計・構築・特徴』 6巻国語研究所「日常会話コーパス」プロジェクト報告書.
- C. Cloran (1994). “Rhetorical units and decontextualisation: An enquiry into some relations of context, meaning and grammar.” Unpublished doctoral dissertation, University of Nottingham Nottingham.
- C. Cloran (1999). “Contexts for learning.” Frances C (Ed.), *Pedagogy and the Shaping of Consciousness: Linguistic and Social Processes*. London: Continuum International Publishing. pp. 31–65.

日常会話における「状況づけられた語り」

白田 泰如 (国立国語研究所研究系) *

“Situated Tellings” in Everyday Conversation

Yasuyuki Usuda (National Institute for Japanese Language and Linguistics)

要旨

日常生活において、目の前で起こった出来事と結びつける形で、過去に自分が経験した事柄について語るということがなされる。本研究ではこの種の語りを「状況づけられた語り situated telling」と呼ぶこととし、以下の2点について検討する。1. この種の語りによって何が行なわれているのか。2. 具体的にどのような方法によって、語られる経験と目の前の出来事とが結びつけられているのか。この問題を検討するため、『日本語日常会話コーパス』からいくつかの会話断片を選び出して分析する。状況づけられた語りは、目の前で起こった出来事と類似の経験を有していることを示す。このことは、目の前で起こった一回限りの出来事を、何らかのパターンに基づいて「繰り返し起こること」として再編し、理解可能にしている。それにより目の前で起こった出来事を「繰り返し起こること」として理解することを主張するものであるといえる。そのために状況づけられた語りは、過去に経験された出来事の構成要素が目の前の出来事理解に即した形で組み立てられる。またそのことによって目の前の出来事が経験の語りによって理解可能になるといえ、相互反映的な関係が見出される。

1. はじめに

日常生活において、我々はしばしば、そのとき目についた物事、経験した出来事にいわば呼び起こされるような形でトピックを見出し、語りを開始するという事態に遭遇する。例えばデータ1の3行目から始まるIC01の語りは、その直前に起こった出来事(断片外)に関連する形でなされている。詳細は4において述べるが、この断片はレストランでの食事場面のものであり、3行目からの語りはIC01が「大好きでよく行く」という「ベルギービールの店」における「新入社員」の話だが、それはこの断片の会話が行なわれたレストランの店員のふるまい(についての参加者の理解)に関連づけられる形でなされている。

本稿では、このようにその場の状況に関連してなされる語りを「状況づけられた語り」と呼び、状況と結びつけて語ることで参加者は何をしようとしているのか、どのような仕方で状況と語りを結びつけているのか、について考える。また、その場の物事や出来事に関連づけて語りを行うことと、他の参加者が語った内容に結びつけて語りを行うこととの並行性を指摘し、その場の物事や出来事、あるいは先行する語りに対して、当該の語りがあるように位置付けられており、それがどのように観察可能になっているか、について述べる。

* usuda@ninjal.ac.jp

データ 1 [会話 ID: S002_007 1965.261 秒-2008.600 秒]

- 1 IC05 あん [まりわかんない [って [ね:(言いますよね:).
 2 IC02 ([あ [::).
 3 IC01 [でも [でも この [前: 俺 池袋で (.)
 4 IC04 [ね: 最近 だって みんな飲まない
 5 IC01 [え あの ベルギービールの店があるんですよ。=
 6 IC04 [ってゆうじゃないっすか。
 7 IC03 =へ [::。
 8 IC01 [で:もう大好きでよく行くんだけども そこも四月に新入社員が入った
 9 らしく [て
 10 IC03 [うん
 (0.5)
 11 IC01 初めてなもんで すいません (0.4) [えっと:ってゆって。
 12 IC03 [huhhuhuhu
 (0.3)
 13 IC01 あの それって でも 普通の栓抜きのやつだから (0.2) 開けるだけなのね。
 14 IC03 うん。
 15 IC01 で そのあと
 (.)
 16 IC04 あほんと かけたら お [いしいっすよ。
 17 IC01 [三人で m [ビックボ [トルを割るから [ついでく。
 18 IC02 [え?
 19 IC04 [か- 辛いのお [いしいっすよ。
 20 IC03 [あ::[::。
 21 IC02 [辛いのお [おいしい。
 (.)
 22 IC03 うん。
 (0.3)
 23 IC01 もうね (.) あの 一人目 泡だけ。
 24 IC03 aha[hahahaha
 25 IC01 [hehe
 26 IC04 [ん でも ちょっと薄いです。
 27 IC01 [で 二人目がすっげえうまうまっってて (.) [三人目は (0.2) なん (でか) 知らない (けど
 28 IC02 [なら これ。
 29 IC01 すごい) 底のほうに溜まった澱だけってゆう (.) すごい面白い状況になって。
 30 (.) 申し訳ございませんって言われても申し訳ございませんじゃねえだろうこれ
 31 って思ったもん。
 (0.4)
 32 IC03 hhha[hahaha
 33 IC01 [とゆうのは あった。
 34 IC03 .h [hh
 35 IC02 [ん?
 36 IC05 [ね: ちょっとというか。
 37 IC03 [え 一番おいしいやつ 飲みたいですよ?

2. 背景

2.1 会話における語り

語り手が過去に遭遇した出来事を、時系列的に順を追って話すことは「物語 (storytelling, Jefferson 1978, Mandelbaum 2012)」と呼ばれる。また、具体的な過去の出来事を時系列に沿って話すことに限らず、嗜好、主張、不満といったことならについて、複数の発話を費やして話すことを「語り (telling, Jefferson and Lee 1992)」と呼ぶ。いずれの場合においても、ある参加者が語りを成功裏に開始した場合、語りの途中で発話の権利が相手に移行することが適切な位置が生じたとしても、まだ語りが終わりそうにない場合、一般に発話権は移行せず、語りを聞いている参加者は最小限の応答を返すにとどまることが知られている (Jefferson 1978, Mandelbaum 2012, Schegloff 2007)。つまり語りの発話連鎖においては、通常のターン交代システムが停止される。そのことは会話において、語りが際立った連鎖環境であることを特徴づける。

2.2 第二の物語

Another way in which recipients indicate what they are making of a storytelling is by telling a "second" story (Ryave, 1978; Sacks, 1992: I: 764-72, II: 3-17, 249-68). These stories are built to show that they are touched off by and / or are picking up the point of the story to which they are responding (Sacks, 1992: I: 767-8). Sacks also noted that second stories are not just "touched off by," but are carefully fitted to and specifically "stand as analysis of" the prior (771).

(Sidnell and Stivers 2012:505)

第二の物語 (second story, Sacks 1992) とは上記の引用において述べられているように、先行する物語に触発される形で開始される語りを指す (Sacks 1992, Mandelbaum 2012)。先行する物語に対して類似の経験による第二の物語を語ることは、それまで聴き手であった参加者が、それまでの語り手が語った経験について理解したことを、単に主張するのではなく立証することができることとされる (Sacks 1992)。

2.3 本研究における議論との接続

既に語られた話に対して第二の物語によって応接することには、必然的にその先行する物語をどのような態度をもって捉えたかということが問題になる。従って第二の物語をめぐっては、Sacks (1992) で論じられるような第一の物語の理解の問題に加えて、成員性や共感の問題が論じられてきた。例えば、語られた経験に対して類似の経験を語ることにより、第二の物語の語り手は第一の物語の語り手と「共-成員 co-member」であること、すなわち同じカテゴリーに属する成員であることが立証され (串田 2001, 2005, 戸江 2018)、語られた事柄に対して共感的であることを示すこととなる (Ruusuvuori 2007, 安井 2012)。

一方、4 において見るように、物語ではなく出来事に応接して語られる物語は、その出来事がどのような物語と規範を同じくするものとして理解可能であるかが表出されることとなる。

同時に、態度が問題にならないわけではなく、出来事そのものではなく、その出来事がどのように扱われているのかに対しての態度が表出される。「第二の物語」の第一の物語との結びつきと、出来事とそれに対する語りとの結びつきには5において触れる。

3. データと方法論

本研究で扱うデータは、国立国語研究所より2022年3月に公開された『日本語日常会話コーパス(CEJC)』(Koiso ほか)である。CEJCは日常生活における会話の多様性をできるだけ反映し、さまざまな研究に利用可能な形で提供するため、音声および映像と文字起こしテキストを利用可能な形で提供するほか、以下のような特徴を備えるよう設計されている。

- **大規模**：200時間分の会話データ
- **代表性**：年齢・性別・属性・会話の種類 of 均衡性を考慮
- **検索性**：形態論情報(品詞、文中の位置、発話時間など)

上記の自然会話データについて、会話分析(conversation analysis, Sacks et al. 1974, Schegloff 2007)の方法論にもとづく分析を行う。会話分析とは、「人が日常生活の中で従事する多種多様な実践的諸活動——会話、会議、診察、面接、ゲーム、授業、接客等々——を構成する出来事や人びとの振る舞いが、いかにしてその場で常識的に合理的な理解可能性を備えるものとして成立しているか、この秩序を産出するための社会成員の「方法」(Garfinkel 1967)を、発話をはじめとする相互行為中の振る舞いの観察を通じて明らかにする」方法論である(平本 2018)。我々はやりとりを行いながら日常のさまざまな活動を行なっている。そうした活動の中のやりとりに用いられることばや身振りなどのふるまいは、すべてがそうではないにせよ、その活動を構成するひとつひとつの行為や活動全体を成り立たせるための参加者の指し手になっているものを含んでいる。会話分析が採用する分析方針は、どのようにしてそうしたふるまいが行為を構成する指し手になっているのかを明らかにすることである。

4. 分析

再度、1のデータ1を見られたい。この断片では仕事に関連する知り合い同士が5人で、レストランで食事をしている。断片に先行する部分で、熟練していない店員がワインの開栓に失敗しており、その次にワインを注文した際には件の店員は「こちらで開け」る(バックヤードで開けてくる)ことを申し出て、テーブルを離れる。それからしばらく件の店員のありうる失敗についての冗談を言い合う(「ばらっばらになった瓶を持ってきたら笑うよね」など)。断片はその直後からのものであり、1行目は、件の店員が飲酒の習慣をもたないためにワインの開栓のことが「わかんない」のだろうと述べているものである。なお、断片の会話において、IC01、IC03、IC05とIC02、IC04がそれぞれ分裂状態にあり、3行目から開始されているIC01の語りの聴き手になっているのはIC03とIC05である。

3行目の語りの開始部には「でも」が前置されており、直前の見解となんらかの形で逆接・対比的な内容が開始されることが予示されている。これに続いて「ベルギービールの店(5行目)」およびその「新入社員8行目」はいずれも、参加者が直前までに経験した出来事に沿うように物語を組み立てるために提示されているといえる。11行目において再現されている「新

入社員」の発話も同様に、店員の不慣れさを例示することで、直前の出来事と語られている経験が類似していることが例証されている。一方、続く13行目では、IC01の経験においては開栓には特段の技能を要さないことが、「でも」を伴って示されており、類似しない部分が明確にされている。続く15行目には「で」が前置され、その後の17行目から物語の中心的内容にさしかかることが予示される。23行目は laughable(Glenn 2003) として提示されていることは、23行目に対する反応である24行目からも見て取れる。一方、27行目および29行目には直後に明瞭な反応が生じておらず、31行目のあとで再び笑いが生じている(32行目)。

まず、この断片に先行する時点で話題にされている、参加者が会話時点で食事をしているレストランの店員のふるまいは、まさに話題にされているそのことにより、ある程度際立って特異なものとして扱われていると言える。3行目の「でも」は、まずその先行する文脈に対して、続く語りが対比的な位置づけにあることを示している。すなわち、その出来事が他に類を見ないほどの特異なものなのではなく、類例を自身の経験の中から見つけて語るができるくらいにはよくあることだという理解が示されている。そのように類似の経験を語ることによって、そのような理解の仕方を示すことができるのは、とりもなおさず我々が通常、出来事を理解するために用いている「ドキュメンタリー的方法 (documentary method of interpretation, Garfinkel 1967)」との並行性による。ドキュメンタリー的方法とは端的に以下のように説明される。

解釈のドキュメンタリー的方法とは、なにか特別な方法のことを指すのではなく、人々がものごとを理解するために日常的に行っている実践であり、現実には発生する発話や活動を、その背後に想定される規範のドキュメント (その規範についての 資料・事例) として取り扱い、規範と結びつけて解釈していく実践を指す。

(前田ほか 2007)

上の説明からもわかるように、我々は日々、原理的に一回限りの出来事に繰り返し遭遇する上で、それらをすべて互いにどれとも似ていない通約不可能なことがらとして扱っているということではなく、なんらかの仕方で類型にまとめて理解している。類似の経験を語るにより、目の前の出来事が、語られた経験と同種のものとして、ある累計のもとにまとめて理解できるものとして再構成しているといえる。33行目は、そのように目の前の出来事をあてはめることのできる類型の候補として、先行する経験の語りを行ったことを説明しているとみることができる。

このとき、注目に値するのは31行目における「もん」である。この「もん」はある種の理由を表す表現である。では、この「もん」が接続している30行目および31行目が何かの理由になっているのだろうか。そうだとすると、考えられるのは、語られた経験を通じて30行目および31行目のように「思った」ことがあるために、類似した目の前の出来事も、その類型のうちに理解することのできるものであるということを主張しているのではないだろうか。ただし、ここに理由を表す「もん」という語彙が配置されていることにより、語り自体がそのような主張をするための理由を提示するものになっているのかどうかは現時点では断定できない。

5. 考察

4で論じた現象と類似したものとして、「第二の物語」を語る事例を見てみたい。データ2は、2人の男性が職場(IC01の主宰する個人学習塾)で、業務外の雑談をしている場面である。断片の直前では、IC01の知人のフィリピン出身の人物の弟が、仕事をせず家にいて家族の女性の収入に頼って暮らしており、そのような「ハウスハズバンド」が「海外」ではある程度社会的に認知されているという話がされている。3行目は、そうしたライフスタイルは日本では受け入れられていない、という趣旨で開始される「第二の物語」である。

データ2 [会話 ID: T018_004 943.395 秒-944.063 秒]

- 1 IC02 (な [かなか] うーん。
- 2 IC01 [そっか:。
(0.5)
- 3 IC02 でも 日本では (なかなか) その制度 (.)
- 4 あの (0.3) (え) (0.4) 女性を働か () の うちの保育士の 母 保母ですけど:
(0.5)
- 5 あの: なんか昔テレビで: (0.6) 女性がけん- あの 研究職で: 男性がトヨタの (0.5)
- 6 あの: なんだっけ 技術職の人がいて: 結婚した時 女性がもう 研究職とですげえ
- 7 大成し (そう) だから: 男性が トヨタ辞めちゃったって話しがあつたんすけ [ど:。
- 8 IC01 [うん。
- 9 IC02 (0.3) うちの親 それすっげえ否定してましたからね 男性が (働か) ないとおかしいっ
て。
(0.5)
- 10 IC01 だ それ 理由 でも 理由はないじゃん 別にさ [なんとなく感覚でしょ?
- 11 IC02 [はい 感 [覚で
- 12 IC01 [うん。
- 13 IC02 そう (反対) してるんす [けど:
- 14 IC01 [だって別に:
- 15 IC02 はい。

2.2において述べたように、第二の物語は、先に語られた事柄を、自分も似た経験があるということを示して理解を立証するものである。従って、第二の物語は先行する語りを第二の物語の語り手がどのように理解したかが示されることになる。データ2は、IC01が先に語ったフィリピン人の知人の話に対して、自らの母親の話語ることで、自らの経験の範囲での先の話と理解する上での規範を共有するであろう物語を取り出して語ったものだと考えることができる。従って、ここでの第二の物語も、先行する事柄に対して自らの経験を語ることで、それをどのように理解したかを表出し、ある規範のもとに理解可能な事例としてそれらを並列する手続きになっているといえる。

ここにデータ1のような状況づけられた語りとデータ2のような第二の物語との並行性を指摘することができる。既に述べたように、どちらも先行する事柄をどのように理解したかを、自らの経験を語るという仕方では表出する方法である。また、データ1において行われていたのは、出来事から事象をピックアップして再構成し、それに沿う形で経験を物語として構築す

ることであった。これは既に第二の物語において指摘されている、単に似た話を語るのではなくて、第一の物語の詳細な分析に基づいて第二の物語が構築されるということと並行的である (Sacks 1992)。

加えて、4 では、データ 1 では単に共通の規範に基づいて理解可能な事例が表出されているのみならず、その規範のもとに先行することがらが理解可能であることが主張されていることの傍証として、語りの末尾に現れる「もん」に着目した。同様の理由を表す表現である「から」が、同様にデータ 2 の 9 行目にも現れている。この「から」も同様に、語られた事柄のような事例があることを理由に、先行する事柄がそれと並行的に理解可能であるということをも主張するものであるように聞くことができる。

6. おわりに

本研究では、直前に起こった出来事に関連して、それと類似の経験を語ることを「状況づけられた語り」と呼び、以下のことを指摘した。状況づけられた語りは、目の前の出来事が語り手の経験した別の出来事と同じ規範に基づいて理解可能であることを主張するものである。この、同じ規範に基づいて出来事を理解するという手続きは「ドキュメンタリー的方法」という、人々が日頃普通に行っている方法が可視的になされているものであるといえる。またそのような主張がなされていることの傍証として、しばしば状況づけられた語りにおいて、その経験を根拠として出来事を理解しようとする表現が出現することが指摘できる。

本文中で用いた転記記号

本文中で会話の転記に用いた記号を表 1 に示す。これらの記号は西阪 (2008) に拠る。

表 1 転記記号一覧

[発話の重なる開始 (重なるの終了が] によって示されることもある)
:	直前の音が引き伸ばされている
=	発話末と次の発話頭に付され、その二つの部分が隙間なくつながっている
-	直前の語が途切れている
?	発話末が上がる音調
(発話)	不明瞭な発話
(.)	おおむね 0.2 秒に満たないわずかな間
(数字)	おおむね括弧内の数字の秒数の間が空いている
> 発話 <	記号に挟まれた発話の部分が相対的に速く発話されている
< 発話 >	記号に挟まれた発話の部分が相対的に遅く発話されている
h	呼気音 (聞き取りに応じて他の子音が用いられることもある)
.h	吸気音

謝 辞

本研究は、国立国語研究所のプロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」(プロジェクトリーダー・小磯花絵) による成果に基づいて行われた。また日本学術振興会科学研究費交付金若手研究「日常会話コーパスを用いた「課題」に基づく会話の分析：定

量・定性の両面から」(研究代表者:白田泰如, 課題番号 20K13019) の助成を受けて行われた。

文 献

- Gail Jefferson (1978). "Sequential Aspects of Story Telling in Conversation." Jim Schenkein (Ed.), *Studies in the Organization of Conversational Interaction*. New York: Academic Press., Chap. 9 pp. 213–248.
- Jenny Mandelbaum (2012). "Storytelling in Conversation." Jack Sidnell, and Tanya Stivers (Eds.), *The Handbook of Conversation Analysis*. West Sussex: Wiley-Blackwell., Chap. 24 pp. 492–507.
- Gail Jefferson, and John R. E. Lee (1992). "The Rejection of advice: Managing the problematic convergence of a 'troubles-telling' and a 'service encounter'." Paul Drew, and John Heritage (Eds.), *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press. pp. 521–548.
- Emanuel A. Schegloff (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Jack Sidnell, and Tanya Stivers (Eds.) (2012). *The Handbook of Conversation Analysis*. West Sussex: Wiley-Blackwell.
- Harvey Sacks (1992). *Lectures on Conversation Volume 1 & 2*. New Jersey: Wiley-Blackwell.
- 串田秀也 (2001). 「私は - 私は連鎖: 経験の『分かち合い』と共 - 成員性の可視化」 社会学評論:2, pp. 36–54.
- 串田秀也 (2005). 「会話における参加の組織化の研究: 日本語会話における「話し手」と「共 - 成員性」の産出手続き」 博士論文 (未公刊), 京都大学大学院人間・環境学研究科.
- 戸江哲理 (2018). 『和みを紡ぐ: 子育てひろばの会話分析』 勁草書房, 東京.
- Johanna Ruusuvuori (2007). "Managing affect: integration of empathy and problem-solving in health care encounters." *Discourse Studies*, 9:5, pp. 597–622.
- 安井永子 (2012). 「接続詞「でも」の会話分析研究: 悩みの語りに対する理解・共感の提示において (高橋亨教授退職記念)」 名古屋大学文学部研究論集, 58, pp. 89–102.
- Hanae Koiso・Haruka Amatani・Yasuharu Den・Yuriko Iseki・Yuichi Ishimoto・Wakako Kashino・Yoshiko Kawabata・Ken'ya Nishikawa・Yayoi Tanaka・Yasuyuki Usuda・Yuka Watanabe "Design and Evaluation of the Corpus of Everyday Japanese Conversation." *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pp. 5587.
- Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson (1974). "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *Language*, 50:4, pp. 696–735.
- Harold Garfinkel (1967). *Studies in Ethnomethodology*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- 平本毅 (2018). 「会話分析の広がり」 平本毅・横森大輔・増田将伸・戸江哲理・城綾実 (編)

『会話分析の広がり』 ひつじ書房, 東京 pp. 1-33.

Phillip Glenn (2003). *Laughter in Interaction. Studies in Interactional Sociolinguistics*.
New York: Cambridge University Press.

前田泰樹・水川喜文・岡田光弘 (編) (2007). 『エスノメソドロロジー：人びとの実践から学ぶ』
ワードマップ 新曜社, 東京.

西阪仰 (2008). 『分散する身体：エスノメソドロロジー的相互行為分析の展開』 勁草書房, 東京.

関連 URL

『日本語日常会話コーパス』 <https://www2.ninjal.ac.jp/conversation/cejc.html>

コーパスからの複合動詞の自動抽出の試み —近現代作家の文学作品からの用例抽出を例に—

チャクマク ビルギル・ニハル (アンカラ大学/麗澤大学)[†]

千葉庄寿 (麗澤大学)^{††}

An Extraction of Japanese Compound Verbs from Corpus: An Example of Modern Literary Works

Nihal Cakmak Bilgili (Ankara University/Reitaku University)

Shoju Chiba (Reitaku University)

要旨

これまで谷崎潤一郎作品に出現する複合動詞をマニュアルにて抽出し、データベース化してきた。今後、さらに多くの谷崎作品から複合動詞を網羅的に取得するにあたり、検索作業の自動化を試みた。本稿では、既存の複合動詞データベース3件(野田 2013, 山口 2013, 国立国語研究所 2015)のデータを統合して検索・処理に使用し、谷崎作品に応用した結果に基づいて評価をおこなう。また、複合動詞と単純動詞を区別せず、一般動詞の語彙素として認定している UniDic の解析データの問題点の克服を試みるとともに、さらに、既存の複合動詞データベースに含まれない「新しい」複合動詞の候補を効率よく発見するための工夫についても紹介する。

1. はじめに：谷崎作品からの複合動詞用例抽出

複合動詞とは、動詞の前に何か別の単語を組み合わせたもので、日本語では、大別すると「嘆き悲しむ」、「笑い飛ばす」等の動詞+動詞型の複合動詞と「名づける」、「旅立つ」などの名詞+動詞型の複合動詞の2つのタイプに分かれる(由本 2013:59)。本研究で取り上げる複合動詞型は「動詞連用形+動詞」型である(「テ形」+動詞は扱わない)。代表的な先行研究には、日本語の「動詞+動詞」型複合動詞の研究として影山(1993:78-96)に代表される、「語彙的」複合動詞と「統語的」複合動詞との2分類が挙げられる。「語彙的」複合動詞の特徴は、2つの動詞が一体化して意味の不透明化や語彙化が進んでいることであり、レキシコンに登録されているものと考えられることができる(影山 1993:78)。一方、統語的複合動詞では、そのような意味の習慣化が見られず、2つの動詞の意味関係は完全に透明であり、2つは使用される際に統語的に組み合わせられるものと考えられる(同上)。

本研究では、既存の複合動詞データベース3件(野田 2013, 山口 2013, 国立国語研究所 2015)のデータを統合し、複合動詞のリスト(「複合動詞統合データベース」と呼ぶ)を作成した。そのリストに「茶まめ」を利用し UniDic による形態素解析を施し、検索処理のためのプログラムの参照元となるデータを作成した。

UniDic では、登録されている複合動詞は単純動詞から区別されず、一般動詞と同様に1つの語彙素として認定されている。さらに、UniDic においては、登録されていない複合動詞が2つの語彙素に分割されたり、ケースによっては誤解析されたりする。本稿では、コーパスから対象となる複合動詞をコーパスから網羅的に検索するための方策として、

[†] nihalcmmk@gmail.com

^{††} schiba@reitaku-u.ac.jp

UniDic に登録されていない複合動詞についても抽出できるよう検索プログラムを工夫した。

発表者によるこれまでの谷崎潤一郎作品における複合動詞の分析においては、短編小説『刺青(しせい)』、『少年』と『春琴抄』から複合動詞を手作業にて抽出し、用例データを作成してきた。このデータと本研究で作成した検索プログラムを用いて谷崎の作品全般からの用例抽出に実際に応用した結果を比較することで、検索ツールとしての評価をおこなうことができる。

発表者の研究の射程は統語的複合動詞を含めた「動詞+動詞」型複合動詞の全体像を記述することであり、研究課題にはどのような統語的複合動詞が文献にどの程度用いられているかといった問題も含まれる。既存の複合動詞データベースには統語的複合動詞が含まれず、主として語彙的複合動詞のみが収集されている(山口 2013, 国立国語研究所 2015)。そこで、その複合動詞データベースに含まれない統語的な複合動詞、およびこれらのデータベースにない「新しい」複合動詞を発見することも重要であり、本研究のプログラムが求められるところである。

2. 既存の複合動詞データベース 3 種の特徴

本稿では既存の複合動詞データベース 3 件(野田 2013, 山口 2013, 国立国語研究所 2015)のデータを統合し、検索のベースとなる複合動詞リストを作成した。使用されたデータベース 3 件はその目的や収集方法等が異なるので、まず本節では、これらのデータベースの特徴と統合の過程と結果について示す。

野田(2013)が「日本語教育のための資料として提示することを目的」に作成した複合動詞一覧は、約 2,400 語の複合動詞を含む。一覧には、国語辞書の複合動詞の語彙項目に加え、「辞書に載らない、意味の分かりやすいものをできるだけ拾」っている(野田 2013:38)。

『Web データに基づく複合動詞用例データベース』(山口 2019)とは、山口昌也氏によって作られたデータベースであり、2013 年より公開されている。用例データベースは日本語の複合動詞と、それを構成する動詞との意味的な関係を分析するための基礎データの構築することを目的とする。山口(2019)によれば、データベースは「動詞連用形+動詞」型の 3,371 語の語彙的複合動詞を含み、それぞれの複合動詞について表記、読み、語構成や用例等の情報が付加されている。用例収集の際には、JUMAN version 6.0 を用いて形態素解析し、構文解析・格解析には KNP version 3.01 を用いているが、収集された複合動詞が語彙的複合動詞であるかどうかは、影山(1993)による代動詞「そうする」による置換テストを用い(山口 2019:19-18)、最終的には手作業で選別している(山口 2019:22)。

『複合動詞レキシコン』²は国立国語研究所が作成・公開しているデータベースである。データベースには「現在の日本語でよく使われる」、2,759 語の「動詞+動詞」型複合動詞が含まれ、古語・古典語、特殊な専門分野や文学作品に出現が限られ一般性がない語彙は除外されている(ホームページより)。収録されているのは「動詞連用形+動詞」型のみで、「読んでみる、止めておく」のように前項動詞がテ形のものには含まれない。その一方で、テ形を含む「見て取る」のような意味が慣習化した語彙は収録されている。データベースには語彙的複合動詞だけが収録されており、「～始める、～かける、～忘れる、～合う」

¹ オンライン版のデータベースには、3,757 語が収録されている。

² <https://vvlexicon.ninjal.ac.jp/>

のような統語的複合動詞は除かれている。複合動詞が語彙的か統語的かの区別については、データベースのマニュアルに影山(1993)に基づく詳細な記述(3.3節参照)がみられる。データベースには、意味的・文法的情報も付与されている。

『複合動詞レキシコン』と『Web データに基づく複合動詞用例データベース』は語彙的複合動詞のみを含むが、野田の(2013)データベースには複合動詞の語彙的・統語的区別についての言及はない(3.3節参照)。

3. 複合動詞検索システムの構築

3.1 用例検索システムの設計と実装

本研究では、統語的複合動詞を含めた複合動詞の用例を網羅的に検索するため、2つの方策でツールを準備する。

- 複合動詞データベースの用例を統合し、これらに登録されている複合動詞をもれなく検索するためのプログラムを作成する。
- 統語的な複合動詞については、動詞連用形+動詞というパターンで検索をおこない、上記複合動詞データベースの項目を差し引いた項目を「可能性のある」複合動詞の候補としてコーパスから取得する。そのうえで、V2動詞(姫野によれば30、影山は27を挙げている)をリストアップし、検索された用例から統語的な複合動詞を取り出す。

上記複合動詞データベースや統語的複合動詞のパターンに含まれない、未知の複合動詞を検索するためのアルゴリズムとして、本研究では以下の方法をもちいて検討する。

- 複合動詞データベースの用例の UniDic による解析結果を参照し、誤解析や2語彙素以上の要素として解析されているデータをもちいて、(1)データベースに含まれず、(2)動詞連用形+動詞という連鎖としても解析されていないパターンを検索するためのアルゴリズムを検討し実際に検索を試みる。

検索システムの構築にあたり、本研究では UniDic (伝ほか2007)による形態素解析データをベースとすることにした。UniDic には、登録されている複合動詞は1形態素(短単位)で収録され、「動詞-一般」という一般的な品詞が付与され、従って複合動詞としてのマーカーがない。山口(2019)では、このことを踏まえて、UniDic ではなく Juman と KNP を解析ツールとして使用している。本研究では、古典作品・近現代むけの UniDic などが開発されている現状に合わせ、UniDic の解析を積極的に使用しながら、複合動詞の処理に必要な情報を追加付与する形でツールを整備することにした。

3つの複合動詞のデータベースに含まれる複合動詞は、一つのリストにまとめ、UniDic で解析をおこなって、検索のもとになる語彙データベースを作った。その結果、上述のとおり、1語彙素としての複合動詞、2語彙素以上に分けて解析される「動詞複合」、動詞+動詞として解析されないもの、さらには誤解析されるものなど、解析結果はさまざまであった。

特に問題となるのが、統合した複合動詞データベースを UniDic で解析したところ、表記によって誤解析されることである。例えば、複合動詞「有り余す」の前項要素「あり〜」

の品詞は UniDic によって「名詞-普通名詞-一般」として解析され、全体は 2 語彙素となる。一方、前項動詞がひらがなで書かれている場合（「あり余る」）、品詞は「動詞-一般」となり、1 語彙素として解析される。

さらに、前項要素が誤解析される例もある。例えば、「きっ立つ」は「切り立つ」の前項要素が音変化したものであるが、前項要素「きっ～」は（副詞）と誤解析される。本研究では、これらの解析違いのパターンを「そのまま」用例解析に活用することで、未知の複合動詞を含めた検索を試みることにした。

3.2 複合動詞データベースの統合について

2 節で紹介した 3 つの複合動詞データベースに登録されている語彙について、以下のようにより修正をおこなった。

1. 受け身形で登録されている複合動詞は基本形で統一した。

思いやられる、焼け出される、打ちひしがれる(2 語彙素→1 語彙素)、並び称される(3→2 語彙素：並び+称する)

2. 否定形で登録されている複合動詞は肯定形で統一した。

煮え切らない、数え切れない

また、野田の複合動詞リストには、いくつか語彙の重複があった（「使い分ける」「見出す」）ので修正をおこなっている。同様に、「抱きしめる」「抱き合う」「注ぎ込む」は前項動詞の読みの違いで異なる場所に 2 つリストされているが、本研究では UniDic で同一の語彙素に解析されることをふまえて 1 つの項目にまとめた。

3.3 統語的複合動詞の選定について

影山（1993:96）によると、統語的複合動詞の後項動詞は限られており、統語的複合動詞を形成すると思われる後項動詞は以下の 27 語である。

表 1：統語的複合動詞の後項動詞(27 語)

始動：～かける、～だす、～始める
継続：～まくる、～続ける
完了：～終える、～終わる、～尽くす、～きる、～通す、～抜く
未遂：～そこなう、～損じる、～そびれる、～かねる、～遅れる、～忘れる、～残す、～誤る、～あぐねる
過剰行為：～過ぎる
再試行：～直す
習慣：～つける、～慣れる、～飽きる

相互行為：～合う
可能：～得る

姫野 (2018:20-21) は、上記の影山 (1993 : 96) の統語的複合動詞 27 語にさらに 3 語を追加している(始動「～かかる」、未遂「～損ねる」、完了「～果てる」)。

統語的複合動詞の判断基準として、影山(1993)は以下の 5 つを提案している。

a. 代用形「そうする」(影山 1993:80)

「そうする」は意図的な行為を表すから、意味的に排除され、統語的複合動詞の場合は、前項動詞を「そうする」で代用しても全く問題が生じない。例えば「調べ終える」は、「そうし終える」のように適切に置き換えることができる。

語彙的複合動詞の場合、例えば「遊び暮らす」の前項動詞の代わりに「そうする」を置換することはできず、「*そうし暮らす」のような不適切な表現になる。

b. 主語尊敬語(影山 1993 : 83-84)

統語的な複合動詞のもう一つの特徴として、前項動詞では主語尊敬語が可能である。

(例) 「歌い始める」は、「お歌いになり始める」のように可能である。

一方、語彙的複合動詞では、「書き込む」の前項動詞を主語尊敬にすることは不可能である(「*お書きになり込む」)。

c. 受身形

複合動詞の前項動詞では受身形で活用することには、語彙的複合動詞には(「書き込む」vs. 「*書かれ込む」) 制限がある。それに対して、統語的複合動詞の場合は、制限がなく、「名前が呼ばれ始めた」のように前項動詞では受身形が可能である。ただし、統語的複合動詞がすべての前項動詞に受身形を許すわけではない(影山 1993:87)。なお、姫野 (2018 : 20) は使役形も受身形と同様に考えることができる(「書かせ始める」と指摘している)。

d. サ変動詞(影山 1993:88)

サ変動詞とは、「雑談する、テストする、立ち読みする」のような動詞類であり、語彙的複合動詞の前項動詞を同義的なサ変動詞と置き換えることはできない。統語的複合動詞の場合は、前項動詞としてサ変動詞に自由に置き換えることが可能である。

e. 重複構文(影山 1993:91)

統語的複合動詞は前項動詞に動詞重複を許すことができるが、語彙的複合動詞はできない。ただし、統語的複後動詞であっても、意味的な制限があれば不適格になる。

(例) 大臣はそれをひた隠しに隠し続けた。

*行方不明の子供を探しに探し歩いた。

以上、統語的複合動詞は、影山 (1993) が指摘している前項動詞の 5 つの文法的な特徴を使って判定することができる。これらの特徴のいずれか一つでももっている場合には、その複合動詞は統語的であると判断できる。逆に言えば、語彙的複合動詞には、これら統語的複合動詞の文法特徴は見られず、また意味が透明でないなどの意味的な制限がかかっていることになる。

本研究では、原則としてコーパスから取得された複合動詞の候補のうち、上記 30 の後項動詞を含む用例について、3 つのデータベースのいずれかに記載があれば「語彙的」、なければ「統語的」複合動詞として分類する。(実際には、語彙的複合動詞として分類される用例には、統語的な動詞としての解釈がふさわしいものがある(詳細は後述する)が、本研究では、この分析はおこなわない。)

一部の複合動詞は解釈により統語的・語彙的複合動詞の両方の特徴をもつと考えることができる。例えば、「～出す、～かける、～合う」のような複合動詞の後項要素は、語彙的複合動詞にも統語的複合動詞にも見える (国立国語研究所 2015, 姫野 2018:83-172)。本研究では、原則として複合動詞データベースに(前項動詞つきで)登録されている動詞はいったん「語彙的」複合動詞としてカウントし、統語的な複合動詞かどうかのチェックをおこなう追加分析を別におこなうことにする。

なお、野田のリストには、以下のような、統語的な複合動詞と考えられるものが含まれている。以下は前項動詞「する」の例であるが、下線の複合動詞は他の 2 つのデータベースには含まれておらず、意味が規則的であることから統語的複合動詞と見るのが妥当である。

仕上がる、仕上げる、し終える、し掛かる、し掛ける、しかねる、し損なう、し損じる、し出す、し尽くす、し付ける、し遂げる、し直す、し慣れる、し残す

なお、野田はリストの作成にあたり、語彙的・統語的の区別をおこなっていない。

3.4 語彙データベースに含まれない複合動詞の検出

データベースに含まれない複合動詞については、検索対象となるコーパスを UniDic で解析したデータから、以下のような手順で「未知の複合動詞」候補の用例の抽出をこころみる。

- 統合した複合動詞データベースや統語的複合動詞のリストに含まれない「動詞連用形+動詞」
- 前項動詞が動詞として解析されないケース：

3.1 節で述べたように、UniDic で複合動詞データベースに登録されている複合動詞を解析すると、複合動詞として 1 語彙素で検出されるもののほか、2 語彙素以上に分けて解析されたり、動詞+動詞以外の語彙素の組み合わせとして解析されるもの、さらには誤解析されるものがある。

そこで、2 形態素の前項動詞が名詞として解析されるものは前項要素として動詞も、また動詞として解析されるものは名詞も検索候補として加える。後者により、

「連用形+動詞」のような一般的なパターンでは検索されない複合動詞を検出することができる。

- 上記と同様に、UniDicの解析間違いがおこなっている前項要素について、後項要素として、データベースに共起している動詞以外が出現するパターンも検索する。以下の例では前項動詞が動詞ではなく記号、副詞や代名詞として解析されている。

表2：前項動詞がUniDicで動詞以外の品詞で解析される複合動詞候補の例

V1 書字形	V1 品詞	V2 書字形	V2 品詞	V1 語彙素	V2 語彙素
おん	記号-一般	出す	動詞-非自立可能	オン	出す
きっ	副詞	立つ	動詞-一般	きっ	立つ
ずり	副詞	下ろす	動詞-一般	ずり	下ろす
ずり	副詞	出る	動詞-一般	ずり	出る
ずり	副詞	落とす	動詞-一般	ずり	落とす
のっ	副詞	掛かる	動詞-非自立可能	のっ	掛かる
ぶち	副詞	飛ばす	動詞-一般	ぶち	飛ばす
ぶっ	副詞	千切る	動詞-一般	ぶっ	ちぎる
ぶっ	副詞	通す	動詞-非自立可能	ぶっ	通す
われ	代名詞	返る	動詞-一般	我	返る
息せき	副詞	切る	動詞-非自立可能	息急き	切る

3.5 検索システムによる用例検索の手順

作成した複合動詞データベースをもとに、コーパスから以下の手順で用例を抽出する。

- 検索対象とするコーパスをUniDicをもちいて解析しておく(4節参照)。
- 統合したデータベースの解析結果を使ってパターンマッチをおこない複合動詞の候補を含む用例を抽出する。その際、1語彙素、2語彙素、3語彙素以上の複合動詞データをそれぞれ検索するほか、動詞連用形+動詞のパターンも検索する。3.4節で示した候補となるパターンも検索する。
- 検出された複合動詞を含む用例は、UniDicの出力結果にコーパス名と出現位置(語彙素番号)、文内の位置(語彙素番号)を記し、複合動詞にあたる語彙素にフラグをつけた形で出力する。UniDicの解析済みデータの形で出力しておけば、その後KWICコンコーダンスの形式などに加工することができる。

4. 谷崎潤一郎作品からの複合動詞の抽出実験

4.1 谷崎作品について

上記で述べた複合動詞検索システムを用いて、現在複合動詞の用例を収集している文学

作品を実験データとして、複合動詞の抽出をおこなってみる。今回対象とする谷崎潤一郎（1886-1965）とその作品群について述べる。

谷崎は明治期末から亡くなる昭和 40 年までの非常に長い期間、執筆活動をおこなっている。その間発表された作品における言語表現には、分析において問題になる要素がいくつかある。まず一つは、カタカナ表記される文章を含むことである。例えば、谷崎潤一郎の作品「鍵」は、日記のスタイルで執筆し、妻が書いている部分は平仮名で、夫が書いている部分は片仮名で書き分けられる。それにより、後者の文章に含まれる複合動詞は、「引き留メタ」「坐り込ンダ」のように書かれており、一般に用例収集を困難にしている。

現在までに、用例を手作業で収集した谷崎の小説『刺青』『少年』『春琴抄』のほか、以下の作品について、作品テキストをコーパスとして複合動詞の自動検索をおこなう。

表 3：検索する谷崎潤一郎作品の書籍情報

作品名	作品名読み	発表年	仮名
刺青	しせい	明治 43 (1910)	新字新仮名
少年	しょうねん	明治 44 (1911)	新字新仮名
春琴抄	しゅんきんしょう	昭和 8 (1933)	新字新仮名
猫と庄造と二人の女	ねことしょうぞうとふたりのおんな	昭和 11 (1936)	新字新仮名
細雪 上巻	ささめゆき	昭和 18 (1943)	新字新仮名
細雪 中巻		昭和 22 (1947)	新字新仮名
細雪 下巻		昭和 22 (1947)～ 昭和 23 (1948)	新字新仮名
鍵	かぎ	昭和 31 (1956)	新字新仮名
瘋癲老人日記	ふうてんろうじんにつき	昭和 36 (1961)	新字新仮名

4.2 用例抽出の準備：テキストの前処理と UniDic による解析

今回作業するのコーパスの元データとなる本文テキストは『青空文庫』より取得した。青空文庫のフォーマットから本文以外の要素を取り除き、電子化にあたって付与されているルビは削除し、コメントとして記述されている特殊文字を以下のように Unicode 文字に変換した(以下は Python のコードの一部である)。

```

line = re.sub(r' | ([^ <]+) <[^\>]+> ', r'¥1', line)
line = re.sub(r' <[^\>]+> ', '', line)
line = re.sub(r'※ [#コト、1-2-24] ', '7', line)
line = re.sub(r'※ [#二の字点、1-2-22] ', 'と', line)
line = re.sub(r'※ [#「言+墟のつくり」、第4水準 2-88-74] ', '謔', line)
line = re.sub(r'※ [#「插」でつくりの縦棒が下に突き抜けている、第4水準 2-13-28] ', '挿', line)

```

```

line = re.sub(r'※ [# 「魚+鑢のつくり」、第4水準2-93-92] ', '鱻', line)
line = re.sub(r'※ [# 「奚+隹」、第3水準1-93-66] ', '雞', line)
line = re.sub(r'※ [# 「王+干」、第3水準1-87-83] ', '玕', line)
line = re.sub(r'※ [# 「勺<夕」、第3水準1-14-76] ', '匆', line)
line = re.sub(r'※ [# 「足へん+宛」、第3水準1-92-36] ', '踰', line)
line = re.sub(r'※ [# 「僵のつくり/糸」、第3水準1-90-24] ', '纍', line)
line = re.sub(r'※ [# 「口+它」、第3水準1-14-88] ', '咤', line)
line = re.sub(r'※ [# 「口+穢のつくり」、第3水準1-15-21] ', '噉', line)
line = re.sub(r'※ [# 「兵」の「丘」に代えて「白」、第3水準1-14-51] ', '貞', line)
line = re.sub(r'※ [# 「りっしんべん+刀」、第3水準1-84-38] ', '切', line)
line = re.sub(r'※ [# トモ、38-8] ', 'ドモ', line)
line = re.sub(r'※ [# トキ、72-13] ', 'トキ', line)
line = re.sub(r'※ [# 「口+云」、第3水準1-14-87] ', '呷', line)
line = re.sub(r'※ [# 「藹」の「言」に代えて「月」、第3水準1-91-26] ', '藹', line)
line = re.sub(r'※ [# 「さんずい+鬲」、U+6EC6、383-4] ', '漏', line)
line = re.sub(r'※ [# 「ころもへん+施のつくり」、第3水準1-91-72] ', '施', line)
line = re.sub(r'※ [# 「足へん+母」、U+27FF9、63-14] ', '踣', line)
line = re.sub(r'※ [# トキ、97-2] ', 'トキ', line)
line = re.sub(r'※ [# 「土へん+敦」、第3水準1-15-63] ', '墩', line)
line = re.sub(r'※ [# 「くさかんむり/奥」、第4水準2-86-89] ', '奠', line)
line = re.sub(r'※ [# 「日+麗」、第4水準2-14-21] ', '曬', line)
line = re.sub(r'※ [# 「齒+乞」、第4水準2-94-76] ', '齧', line)
line = re.sub(r'※ [# トキ、33-6] ', 'トキ', line)
line = re.sub(r'※ [# トキ、146-15] ', 'トキ', line)
line = re.sub(r'※ [# トキ、36-12] ', 'トキ', line)
line = re.sub(r'※ [# トキ、37-2] ', 'トキ', line)
line = re.sub(r'※ [# トキ、36-7] ', 'トキ', line)
line = re.sub(r'※ [# 「くさかんむり/嬰」、第4水準2-87-16] ', '夔', line)
line = re.sub(r'※ [# トキ、43-3] ', 'トキ', line)
line = re.sub(r'※ [# 「てへん+僉」、第3水準1-84-94] ', '檢', line)
line = re.sub(r'※ [# トキ、75-11] ', 'トキ', line)
line = re.sub(r'※ [# 「てへん+聿」、第4水準2-78-12] ', '擿', line)
line = re.sub(r'※ [# トモ、38-11] ', 'ドモ', line)
line = re.sub(r'※ [# 「てへん+𠂔のへん」、第4水準2-13-55] ', '擿', line)
line = re.sub(r'※ [# 「くさかんむり/生」、U+82FC、19-4] ', '莖', line)
line = re.sub(r'※ [# 「飲のへん+稻のつくり」、第4水準2-92-68] ', '飴', line)
line = re.sub(r'※ [# 「肆のへん+欠」、第3水準1-86-31] ', '欸', line)

```

このようにして準備した文献ファイルを UniDic により解析した。解析辞書には、現代語辞書を用いた。(UniDic で公開されている近現代口語小説、旧仮名口語といった古い文献用の電子辞書による解析結果の比較は稿を改めておこないたい。)

4.3 抽出結果 (1) 統合リストを用いた用例検索

統合した既存の複合動詞データベース 3 件(野田 2013, 山口 2013, 国立国語研究 2015)のデータの比較結果は以下の通りである。

- 3つのデータベース全てに含まれる複合動詞：1447 語
- 野田 (2013) のデータベースにしか登録されていない複合動詞：571 語
- 山口 (2013) のデータベースにしか登録されていない複合動詞：1432 語

- 国立国語研究所（2015）にしか登録されていない複合動詞：558 語

その他に、2つのデータベースにのみ収録されている複合動詞もある。以下にまとめて図で示す。

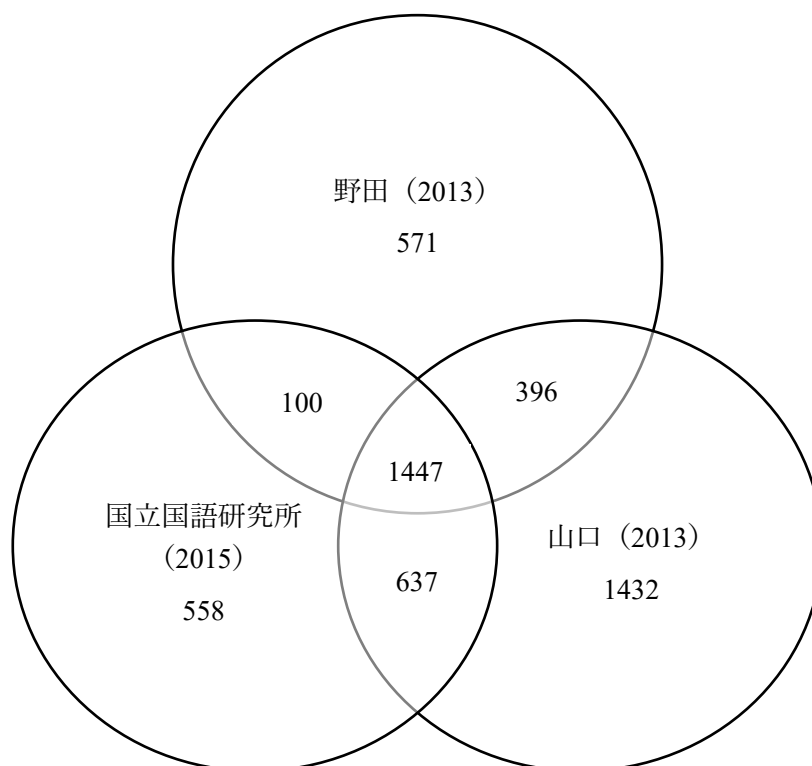


図1 データベースの収録語彙の比較結果

統合したデータベースを UniDic で解析した結果、統合したデータベースでは UniDic で 1 語彙素として登録されている複合動詞の合計は 7,873 であり、登録されていない複合動詞の合計は 1,312 であった。

表4：統合データベースに収録された複合動詞の UniDic の登録状況

	国立国語研究所 (2015) ³	山口(2013)	野田(2013)	総計
UniDic に語彙 登録あり	2,456	3,264	2,153	7,873
UniDic では 1 語彙素でない	303	648	361	1,312
総計	2,759	3,912	2,514	9,185

このようにして解析された「複合動詞統合データベース」の複合動詞を用いて谷崎コーパスを検索した結果、1 語彙素の複合動詞として 2,852 例(異なり 847 種類)、2 語彙素から

³ 表4の国立国語研究所(2015)のデータは、「選り抜く」のような読みが異なる複合動詞を分けて数えている。図1では読みの違いを相殺しているため、合計は2742となる。

なる複合動詞として 278 例(85 種類)の用例が検索された。また、2 語彙素で解析された結果の前項動詞について、データベースの解析結果(名詞+動詞)とは異なる品詞(動詞+動詞)で検索したところ、「落ち着き払う」が 1 例検索された。

表 5 UniDic による「落ち着き払う」の解析例(句点を省略している)

	書字形	語彙素	品詞	活用形	語種
データベースの解析	落ち着き	落ち着き	名詞-普通名詞-一般		和
	払う	払う	動詞-一般	終止形-一般	和
コーパスの解析	落ち着き	落ち着く	動詞-一般	連用形-一般	和
	払っ	払う	動詞-一般	連用形-促音便	和

(例) と、妙子がことさら落ち着き払った口調で云った。(『細雪』下巻: 37649-37661)

このような用例は、UniDic でデータベースの動詞を解析した結果をそのまま用いて検索すると取得できないことになる。同様に、谷崎の作品における「飛び出す」は 2 語彙素で「飛ぶ」と「出す」として、いっぽうデータベースに収録されている「飛び出す」は 1 語彙素で解析される。「着替える」も、谷崎作品では書字形が「着換える」として入っているが、この書字形は UniDic では 2 形態素となる。いっぽう、データベースの「着替える」は 1 語彙素で解析される。

さらに、今までマニュアルで作成した谷崎潤一郎作品の複合動詞データベース（『刺青(しせい)』『少年』『春琴抄』と『細雪上巻(約半分)』）を比べると、「複合動詞統合データベース」に入っていない複合動詞が見つかった。

例えば『刺青』に載せられている「軋み合う、誇り合う、疼き出す、薄らぎ初める、暮れかかる」のような複合動詞は 3 つのデータベースのいずれにも入っていない。それら 5 語の後項要素は統語的複合動詞の候補であるので、データベースに掲載されていないのは自然である。

このような例文も網羅的に検索するため、本研究では、「動詞連用形+動詞」という品詞パターンでの検索もおこない、用例の補完を試みている(3.3 節参照)。しかし、これらのうち、この補完的検索から得られたのは「誇り合う」「疼き出す」「薄らぎ初める」「暮れかかる」であった。「軋み合う」は UniDic による解析が今回の想定と異なってしまった。このような解析のゆれをどのようにコントロールし網羅的検索につなげるかが、課題として残った。

いっぽう、影山、姫野が挙げる統語的複合動詞の後項要素(3.3 節参照)にない「生き代わる、死に代わる、結い繞る(ゆいめぐる)」のような複合動詞も、手作業による用例検索で見つかっているが、これらも「複合動詞統合データベース」を使った今回の検索からは得られなかった。

「結い繞る」は統合データベースにはなかったため取得できなかったのであるが、「生き代わる」「死に代わる」が検索できない問題には、3.1 節で述べた原文の送りがなの表記

の問題があることがわかった。以下の UniDic の解析例にみるように、UniDic には語彙素として「生き変わる」「死に変わる」が登録されている。しかし、谷崎作品『刺青』に収録されているのは「生き代る」「死に代る」という異なる漢字表記であり、送りがなも異なっている。結果として、今回の実験では用例としての検索がうまくいかなかったことになる。UniDic は階層的見出しを導入しており、書字形情報をもつことで表記のゆれに柔軟であるという特徴をもつ(伝ほか 2007:108)が、複合動詞の検索において、漢字の異表記が検索において依然として課題であることがわかった。

表 6 : UniDic による漢字表記の異なる複合動詞の解析例(句点を省略している)

文境界	書字形	語彙素	品詞	活用形	語種
B	生き	生きる	動詞-一般	連用形-一般	和
I	代わる	変わる	動詞-一般	終止形-一般	和
B	生き変わる	生き変わる	動詞-一般	終止形-一般	和
B	死	死	名詞-普通名詞-一般		漢
I	に	に	助詞-格助詞		和
I	代わる	変わる	動詞-一般	終止形-一般	和
B	死に変わる	死に変わる	動詞-一般	終止形-一般	和

同様の例には「齧り着く」「噛み着く」「食い着く」「飛び着く」のように、現代であれば後項動詞を「付く」と表記するものなどが挙げられる。これらは統合データベースに収録されているものの、漢字表記の関係で検索できず、全て「動詞連用形+動詞」の用例として取得することができた。

なお、興味深いことに、上記表 6 で前項動詞が動詞として解析されている「生き代わる」であるが、「動詞連用形+動詞」による検索によっては該当する用例を取得できなかった。漢字や送りがなの異表記のため、前部要素が動詞ではなく名詞と解析されたことによるもので、このような、文脈によって解析パターンが異なるケースの存在は、網羅的検索のための方策を考える上での重要な課題である。

4.4 抽出結果 (2) 未知の用例の検索の結果

谷崎潤一郎作品における「動詞連用形+動詞」パターンを用いた検索では、異なり語数で 164 種類(延べ 1,722 語)の後項動詞が見つかった。その 164 種類の複合動詞の後項動詞の中には、統語的複合動詞候補として影山(1993)、姫野(2018)が挙げている 30 語種類中 22 語が含まれていた(述べ 786 語)。

作品から得られた統合的複合動詞候補：

～出す (234) , ～かける (106) , ～始める (78) , 切る (74) , ～付ける (55) ,
 ～合う (41) , ～過ぎる (35) , ～得る (34) , ～続ける (33) , ～かかる (30) ,
 ～かねる (17) , ～尽くす (9) , ～慣れる (9) , ～直す (8) , ～通す (7) , ～
 終わる (6) , ～終わる (5) , ～抜く (5) , ～損なう (5) , ～そびれる (5) ,
 ～果てる (5) , ～誤る (1) , ～忘れる (1) 。

統語的複合動詞を形成する以下の後項動詞は谷崎作品コーパスには見つからなかった：
～まくる，～損じる，～遅れる，～残す，～あぐねる，～飽きる，～損ねる

さらに、語彙的複合動詞と考えられる後項動詞として、後項動詞「置く」の例(「(依頼) 致し置く」)のような統合データベースに収録されていない動詞が複数見ついている。

5. 複合動詞検索システムの評価と今後の課題

本研究では、谷崎潤一郎作品からの複合動詞の用例を網羅的に収集する目的で構想された。既存の3件の複合動詞データベースを統合して複合動詞リストを作成し、そこに収録された全ての語彙的複合動詞を抽出するとともに、統語的複合動詞を含む、統合データベースに含まれない新たな複合動詞の用例収集も試みた。

統合データベースに含まれる複合動詞の用例の検索には、UniDic による解析を介しておこなった。複合動詞は UniDic において単純動詞と区別されていないため、その検索には工夫が必要である。UniDic の解析内容に準じた検索により、データベースにある多くの複合動詞はその用例が検索できるものの、網羅的な検索には至らないことが示された。本研究でおこなった副次的候補による検索で収集できた例文もあったが、解析内容を完全に網羅した検索には至らなかった。(小説『刺青』の用例データを用いた検索システムの評価については論文末の付録を参照されたい。)

収集された新たな用例のなかには、「～為(な)さる」「～下(くだ)さる」「～やはる」「～致す」「～申す」のような終助詞的要素と考えられる後項動詞も多く見出される。谷崎コーパスから収集された複合動詞の詳細は稿を改めて分析したい。

今後は本研究では実施に至らなかった、UniDic や Sudachi の辞書をカスタマイズし検索に利用する方法、さらに今回試みた、解析の副次的候補をさらに充実させる手法について、さらに検討を加え、より網羅的な例文の収集にむけとりくんでゆきたい。また、UniDic の品詞体系に含まれない複合動詞のマーカーをどのように解析情報に加えていくかという問題についても検討していきたい。

謝 辞

本研究は、国際交流基金の日本研究フェローシップにより共同研究として実現したものである。記して感謝する。また、ポスター発表の際、有益なコメントをいただいた参加者の皆さまに深く感謝申し上げます。

参考文献

- 影山太郎 (1993) 『文法と語形成』 ひつじ書房。
 伝 康晴, 小木曾 智信, 小椋 秀樹他 (2007) 「コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用」『日本語科学』 22, pp.101-123.
 野田時寛 (2013) 「日本語動詞用法事典について(4)—複合動詞一覧の試み—」『人文研紀要』 75, pp.31-62. 中央大学人文科学研究所。
 姫野昌子 (2018) 『新版複合動詞の構造と意味用法』 研究社。
 野村雅昭・石井正彦 (1987) 『複合動詞資料集』 国立国語研究所。
 山口昌也 (2019) 『Web データに基づく複合動詞用例データベース』の構築と評価『国立国語研究所論集』 17, pp.15-34.

由本陽子 (2013) 「動詞+動詞型の複合動詞」『レキシコンフォーラム No:6』(編: 影山太郎) pp.59-78. ひつじ書房.

関連 URL

国立国語研究所 (2015) 『複合動詞レキシコン』 <https://vvlexicon.ninjal.ac.jp/>
 『Web 茶まめ』 <https://chamame.ninjal.ac.jp/>
 『Web データに基づく複合動詞用例データベース』 <https://csd.ninjal.ac.jp/comp/>

付録: 谷崎『刺青』(1910)を用いた複合動詞の用例の検証メモ

谷崎の小説『刺青』(1910)には 54 種類の複合動詞が現れる(うち 3 種類は今回の検索プログラムで新たに抽出された)が、その中には今回使用している 3 つの複合動詞データベースのいずれにも見つからないものもある。以下に複合動詞の用例として手収集したリストをもとに、検索プログラムを用いて収集した用例との差異際をまとめる。

- UniDic には 1 語彙素の動詞として登録されているものの、複合動詞のデータベースには収録されておらず、用例として抽出できなかった複合動詞は以下の 3 つである: 打ち倒れる、明け放れる、軋み合う
- データベースと漢字や送り仮名や漢字と仮名の違いがあるもの:

	『刺青』に現れた複合動詞基本形	DB における見出し	UniDic の解析結果	DB に見つからなかった理由	例文として検索できたか	備考	DB に登録されているか
1	喰いしばる	喰いしばる (K,N,Y)	1 語彙素 「食い縛る」	漢字の違い	1	1 語彙素 「食い縛る」	1
2	繰り展げる	繰り広げる (K,N,Y)	1 語彙素 「繰り広げる」	漢字の違い	1	2 語彙素の候補として検出 「繰り広げる」	1
3	見馴れる	見慣れる (K,N,Y)	2 語彙素 「見慣れる」	漢字の違い	1	「見慣れる」 2 語彙素の候補として検出	1
4	惹きつける	引きつける (N,Y) 引き付ける (K)	1 語彙素 引き付ける	漢字の違い	1	1 語彙素 「引き付ける」	1
5	脹れ上がる	膨れ上がる (K,N,Y)	1 語彙素 「膨れ上がる」	漢字の違い	1	「脹れ上る」 2 語彙素の候補として検出	1

6	うち込む	打ち込む (K,N,Y)	1 語彙素 「打ち込 む」	漢字と仮名 の違い	1	2 語彙素の候 補として検出 「内 込む」	1
		撃ち込む (N,Y)					
7	探りあてる	探り当てる (K,N,Y)	1 語彙素 「探り当て る」	漢字と仮名 の違い	1	1 語彙素 「探り当て る」	1
8	さし置く	差し置く (K,N,Y)	1 語彙素 差し置く	漢字と仮名 の違い	1	1 語彙素「差 し置く」	1
9	出来上る	出来上がる (K,Y)	1 語彙素 「出来上が る」	送り仮名の 違い	1	1 語彙素「出 来上がる」	1
10	踏みつける	踏みつける (N,Y)	1 語彙素 「踏み付け る」	漢字と仮名 の違い	0	-	1
		踏み付ける (K)					
11	堪えかねる	耐えかねる (N)	2 語彙素 「耐え兼ね る」	漢字の違い	0	-	1
12	刺り込む	刺さり込む (Y)	2 語彙素 「刺さり込 む」	送り仮名の 違い	0	-	1
13	見出す	見出す (N)	1 語彙素 「見出だ す」	送り仮名の 違い	1	1 語彙素 「見出だす」	1
		見出だす (K)					
		見いだす (Y)					
14	見つめる	見つめる (N,Y)	1 語彙素 「見詰め る」	漢字と仮名 の違い	1	1 語彙素 「見詰める」	1
		見詰める (K)					
15	待ち構える	待ち構える (K,Y)	1 語彙素 「待ち構え る」	漢字と仮名 の違い	1	1 語彙素 「待ち構え る」	1
		待ちかまえ る (N)					

- データベースと語彙の違いが観察されるもの：

『刺青』に 現れた複合 動詞基本形	DBにおける見 出し	UniDicの解析 結果	DBに見 つからな かった理 由	例文と して検 索でき たか	備考
-------------------------	---------------	-----------------	---------------------------	-------------------------	----

1	生まれ出づ	生まれ出る (K,N,Y)	1 語彙素 「生まれ出る」	語彙の違い	1	2 語彙素の候補として検出 「生まれ出る」
2	なし終える	し終える (N)	2 語彙素 「し終える」 「為る」 「終える」	語彙の違い	1	2 語彙素の候補として検出 「なし終える」
3	打ち捨る (うちやる)	打ち捨てる (K,N,Y)	1 語彙素 「打ち捨てる」	語彙の違い	0	-
4	打ち倒れる	ぶっ倒れる (N)	1 語彙素 「打っ倒れる」	語彙の違い	0	-
5	縛(ゆ)いつける	縛(しば)り付ける (K,N,Y) cf. 結い付ける	1 語彙素 「縛り付ける」	語彙の違い	0	-

- データベースに入っていないもの：

	『刺青』に現れた 複合動詞基本形	例文として 検索できたか	候補として検出	備考	DBに登録されているか
1	具え始める	1	2 語彙素 「具え始める」	-	-
2	誇り合う	1	2 語彙素 「誇り合う」	-	-
3	薄らぎ初める	1	2 語彙素 「薄らぎ始める」	-	-
4	暮れかかる (暮れかゝった)	1	2 語彙素 「暮れ掛かる」	-	-
5	躍りたつ	1	2 語彙素 「躍り経つ」	-	-
6	疼き出す	1	2 語彙素 「疼き出す」	-	-
7	死に代わる	1	3 語彙素 「死 に 変わる」	-	-
8	生き代わる	0	-	-	-
9	得堪える	0	-	-	-
10	明け放(はな)れる	0	-	UniDic 「明け離れる」	-
11	軋み合う	0	-	UniDic にあり	-

- データベースに入っているもの：

	『刺青』に現れた複合動詞基本形	DBにおける見出し	UniDicの解析結果	例文として検索できたか	備考	DBに登録されているか
1	引き出す	引き出す (K,N,Y)	1 語彙素	1	-	1
2	引き入れる	引き入れる (K,N,Y)	1 語彙素	1	-	1
3	引き立てる	引き立てる (K,N,Y)	1 語彙素	1	-	1
4	引っ込める	引っ込める (K,N,Y)	1 語彙素	1	-	1
5	喰いしばる	喰いしばる (K,N,Y)	1 語彙素	1	1 語彙素 「食い縛る」	1
6	繰り展げる	繰り広げる (K,N,Y)	1 語彙素	1	2 語彙素の候補として検出 「繰り広げる」	1
7	繰り返す	繰り返す (K,N,Y)	1 語彙素	1	-	1
8	見つめる	見つめる (N,Y)	1 語彙素 「見詰める」	1	1 語彙素 「見詰める」	1
		見詰める (K)				
9	見守る	見守る (K,N,Y)	1 語彙素	1	-	1
10	見出す	見出す (N)	1 語彙素 「見出だす」	1	1 語彙素 「見出だす」	1
		見いだす (Y)				
		見出だす (K)				
11	見馴れる	見慣れる (K,N,Y)	2 語彙素	1	2 語彙素の候補として検出 「見慣れる」	1
12	見入る	見入る (K,N,Y)	1 語彙素	1	-	1
13	見つかる	見つかる (Y)	1 語彙素	1	1 語彙素 「見付ける」	1
14	さし置く	差し置く (K,N,Y)	1 語彙素	1	1 語彙素 「差し置く」	1

15	似通う	似通う (N,Y)	1 語彙素	1	-	1
16	持て囃す	持て囃す (Y)	1 語彙素	1	-	1
17	惹きつける	引きつける (N,Y)	1 語彙素 「引き付ける」	1	1 語彙素 「引き付ける」	1
		引き付ける (K)				
18	とり出す	取り出す (K,N,Y)	1 語彙素	1	1 語彙素 「取り出す」	1
		採り出す (Y)	2 語彙素			
19	取り出す	取り出す (K,N,Y)	1 語彙素	1	-	1
20	出来上る	出来上がる (K,Y)	1 語彙素	1	1 語彙素 「出来上がる」	1
21	掻き立てる	掻き立てる (K,N,Y)	1 語彙素	1	-	1
22	うち込む	打ち込む (K,N,Y)	1 語彙素	1	2 語彙素の候補として検出 「内 込む」	1
		撃ち込む (N,Y)	1 語彙素			
23	待ち構える	待ちかまえる (N)	1 語彙素 「待ち構える」	1	--	1
		待ち構える (K,Y)				
24	探りあてる	探り当てる (K,N,Y)	1 語彙素	1	1 語彙素 「探り当てる」	1
25	注ぎ込む	注ぎ込む (K,N,Y)	1 語彙素	1	-	1
26	脹れ上がる	膨れ上がる (K,N,Y)	1 語彙素	1	2 語彙素の候補として検出 「脹れ上る」	1
27	追いかける	追いかける (N,Y)	1 語彙素	1	1 語彙素 「追い掛ける」	1
28	追い返す	追い返す (K,N,Y)	1 語彙素	1	-	1
29	通りかかる	通りかかる (N,Y)	1 語彙素	1	1 語彙素 「通り掛かる」	1
30	突き刺す	突き刺す (K,N,Y)	1 語彙素	1		1

31	つきのける	突きのける (N)	1 語彙素	1	1 語彙素 「突き除ける」	1
32	のぞき込む	覗き込む (K,N,Y)	1 語彙素	1	1 語彙素 「覗き込む」	1
33	肥え太る	肥え太る (Y)	1 語彙素	1	-	1
34	抱きしめる	抱きしめる (N,Y)	1 語彙素	1	1 語彙素 「抱き締める」	1
35	立てかける	立てかける (N)	1 語彙素	1	1 語彙素 「立て掛ける」	1
36	流れ込む	流れ込む (K,N,Y)	1 語彙素	1	-	1
37	老い込む	老い込む (K,N,Y)	1 語彙素	1	-	1
38	踏みつける	踏みつける (N,Y)	1 語彙素 「踏み付ける」	0	-	1
		踏み付ける (K)				
39	堪えかねる	耐えかねる (N)	2 語彙素	0	-	1
40	刺り込む	刺さり込む (Y)	2 語彙素	0	-	1

今回の用例検索プログラムで抽出できた複合動詞の用例は、手作業で抽出した用例中 86.27% (51 例中 44 例)であった。検索できなかった用例のいくつかは、語彙の書字形のバリエーションをふまえた検索対象の拡張など、今後のプログラムの改善で収集できる見込みであるが、全ての用例の収集には、未だ課題が残っている。

● まとめ(『刺青』に出現する複合動詞の検索結果)

手作業で収集した用例リストにあり	DB にあり	書字形の違いなし	検索できた	25	51
			検索できなかった	0	
	書字形の違いあり	検索できた	12		
		検索できなかった	3		
DB になし	検索できた		7		
	検索できなかった		4		
なし	DB になし	検索できた		3	3
合計					54

以上

日本語学習者のコーパスツールの使用実態 —作文での産出に着目して—

寺嶋 弘道 (立命館アジア太平洋大学)

板井 芳江 (立命館アジア太平洋大学)

Use of Corpus Tool in Writing by Japanese Language Learners

Hiromichi Terajima (Ritsumeikan Asia Pacific University)

Yoshie Itai (Ritsumeikan Asia Pacific University)

要旨

本研究では、「ライティングにおけるコーパスツール活用モデル (寺嶋・板井, 2021)」を取り入れ、日本語学習者が作文を書く際のコーパスツールの使用実態を調査した。分析の結果、作文においてコーパスツールが使用された回数の中央値は 5 回 (最大値: 14 回、最小値: 2 回) であった。また、作文で使用された表現には、作文前に作成した語彙ネットワークから取り入れられたもの、作文を書いている間に検索されたものがあった。最も多く産出されたのは「名詞+助詞+動詞」のパターンで、中級前半レベルと中級後半レベルの言葉で構成されたコロケーションであった。さらに、その適切さを分析したところ、コロケーション、あるいはコロケーションと共に使用された文法項目が原因で誤用と判断されたもの、コロケーションが使用された節において誤用と判断されたものがあり、「混同」による誤用が多いことがわかった。

1. はじめに

日本語学習者 (以下: 学習者) のライティングでは、「私の国では車がすごく震える」という誤用¹が見られる。これらは「混同」に分類される誤用 (市川編, 2010) で、こうした誤用を減らすためには、言葉の共起関係 (以下: コロケーション²) の知識を事前に学んでおくか、何らかのリソースによって適切なコロケーションを調べる方法を身に付けておくことが必要である。例えば、コロケーションを調べられるツールの一つとして、NINJAL-LWP for TWC³というものがある。このツールは、レキシカルプロファイリングの機能を備えたコーパスツールで (赤瀬川他, 2016)、このツールを用いることで「車が揺れる」というコロケーションを見つけることが可能である。しかしながら、学習者がコーパスツールの機能を十分に活かせるとは限らない。ツールを使う場合、学習者は「車が+動詞」のパターンで表示される大量の情報の中から、「車が揺れる」という表現を探すことになる。候補が絞れておらず、読めない漢字、意味が理解できない言葉がツールに表示されれば、「揺れる」を見つけることは容易ではないだろう。

¹ 筆者が指導する日本語クラスで「自国の交通」というエッセイを書いてもらった際にフィリピン人の学習者から産出された誤用であった。

² 本研究では、実質語同士のつながりのことをコロケーションと呼ぶ。例えば、「本を読む」の場合、名詞「本」と動詞「読む」という実質語で構成されたコロケーションと考える。

³ 筑波大学で構築された 11 億語のコーパス「筑波ウェブコーパス」(Tsukuba Web Corpus: TWC) を検索するためのツールで、検索には、国立国語研究所と Lago 言語研究所が共同開発したコーパス検索システム NINJAL-LWP が使用されている。

2. 先行研究

先行研究では、学習者のコーパスツールの使用がライティングに効果があると報告がされている。Hodošček 他（2011）は、「なつめ⁴」を使用することが作文の共起表現の産出に有効であるかを検証したところ、意味、構文、レジスターにおいて効果があったと報告している。一方、寺嶋・板井（2021）はコーパスツールに関わる体系的な教育や支援カリキュラムにつながるような実践報告がないことを指摘し、学習者向けのコーパスツールワークショップでの実践を報告している。このワークショップは基本的なコーパスツールの使い方を学んだ後、ツールを用いたライティング活動を行うもので、「ライティングにおけるコーパスツール活用モデル」（図1）が取り入れられている。

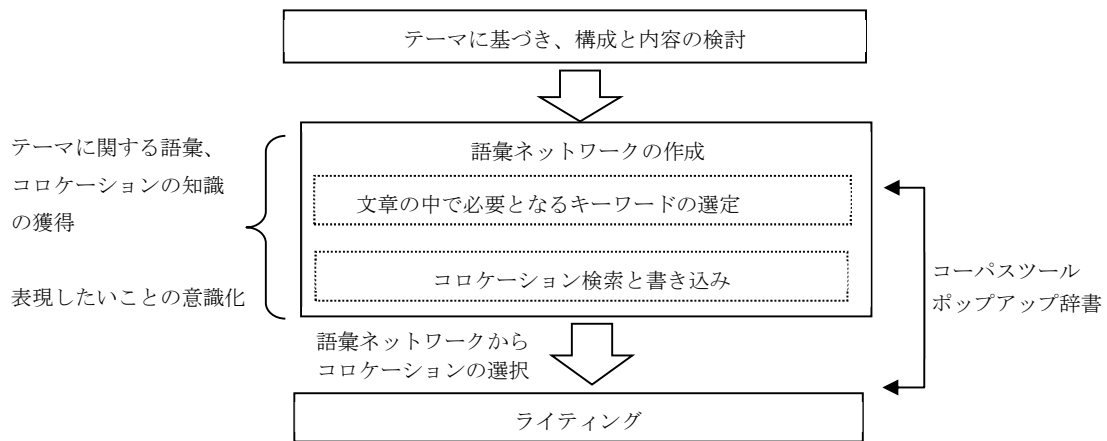


図1 ライティングにおけるコーパスツール活用モデル（寺嶋・板井, 2021）

このモデルは、①コーパスツールで表示される漢字の読み方や意味を理解するために、ツール上で使用できるポップアップ辞書を用いる、②ライティング前にテーマに関する語彙ネットワークを作成し、テーマに関する語彙・コロケーションの知識を獲得する、③語彙ネットワークの作成によって表現したいことを意識化する、④語彙ネットワークからコロケーションを選択し、ライティングを行う、⑤ライティング時にも必要に応じてコーパスツールを使用する、という特徴がある。語彙ネットワークとは、コーパスツールを用いて、キーワードの前後のコロケーション、コロケーションと共に使用される助詞などの文法項目を書き込むものである。たとえば、寺嶋他（2021）のワークショップでは、「経験」をキーワードとし、前につながる表現として「立派な・貴重な」、後ろにつながる表現として「を積む・を活かす」という書き込みが観察されている。

これまでの先行研究を見ると、さらに調査すべき課題もあると考える。例えば、寺嶋他（2021）では語彙ネットワークから作文に使用された表現がどのくらいあるか報告されているものの、作文を書いている間にどの程度、コーパスツールが使用されたかは明らかにされていない。また、コーパスツールを用いた場合に、どのような表現が作文で産出されるか、さらに詳しい調査が必要だろう。

3. 本研究の目的

本研究では、先行研究の課題を踏まえ、「ライティングにおけるコーパスツール活用モデル（寺嶋・板井, 2021）」を取り入れ、日本語学習者が作文を書く際のコーパスツールの

⁴ レキシカルプロファイリング機能を持ったコーパスツールの一つである。「なつめ」は日本語学習者向けのツールとして開発されている（阿辺川, 2011）。

使用実態を調査した。具体的には、学習者がどのようにコーパスツールを使用したか、作文でどのような産出をしたかを明らかにした。調査対象としたのは、コーパスツールワークショップの参加者 17 名（中級レベル 2 名、中上級レベル 15 名）であった。対象者の出身の国・地域は、ベトナム 6 名、韓国 2 名、アメリカ、インド、ウズベキスタン、フィリピン、ブラジル、マレーシア、モンゴル、中国、台湾各 1 名である。

4. コーパスツールワークショップ

2020 年 7 月 11 日、ZOOM を用いて、コーパスツールワークショップを行った。参加者は国内の大学の日本語クラスで募集した。学習者の日本語レベルは、中級～中上級レベルであった。ワークショップの目標はコーパスツールの使い方を理解し、「ライティングにおけるコーパスツール活用モデル（寺嶋・板井, 2021）」で 500 字以内の作文を書くことであった⁵。作文のテーマは、「私の将来のキャリア」であった。

ワークショップでは、まず、コーパス及びコーパスツールの説明をし、コーパスツールを用いた言葉の選択・文法の選択の練習、コーパスツールと辞書を用いて意味を理解する練習を行った。次に、語彙ネットワークの説明を行った後、一つのキーワードを提示し、その語彙ネットワークを作成する練習を行った。その後、作文のアウトラインの作成と共有、語彙ネットワークと作文の作成という流れで実施した。

今回、使用したコーパスツールは NINJAL-LWP for TWC（以下 NLT）で、ポップアップ辞書として rikaikun⁶を用いた。参加者は普段、英語で講義を受けていることから和英辞書である rikaikun でも問題がないと判断した。また、各タスクは進捗状況を把握するため、OneNote Class Notebook 上で行った。

5. 調査結果と考察

5.1 コーパスツールの使用回数と使用方法

表 1 は、各学習者が NLT を用いて語彙ネットワークに書き込みした数（以下：「語彙ネットワークへの書き込み数」）、NLT で調べた表現を作文に使用した数（以下：「作文への使用数」）、「作文への使用数」のなかで、語彙ネットワークから使用した表現の数（以下：「語彙ネットワークからの使用数」）を示したものである⁷。

まず、「語彙ネットワークへの書き込み数」は中央値が 10 で、最大値が 35、最小値が 6、「作文への使用数」は中央値が 5 回で、最大値が 14 回、最小値が 2 回であった。また、「語彙ネットワークからの使用数」は中央値が 3 回（最大値：10 回、最小値：0 回）であった。

表 1 からわかるように、今回の調査では、JL8 を除き、最低 1 回は語彙ネットワークからの表現を使用したこと、JL16 以外の学習者は、語彙ネットワークからの表現だけではなく、作文を書いている間にコーパスツールで調べた表現を作文に取り入れたことが確認できる。

表 1 コーパスツールの使用数

学習者	語彙ネットワークへの書き込み数	作文への使用数 (語彙ネットワークからの使用数)
JL1	16	7 (6)
JL2	10	9 (4)
JL3	11	9 (5)
JL4	13	14 (10)
JL5	35	7 (5)
JL6	10	5 (1)
JL7	23	5 (4)
JL8	10	4 (0)
JL9	11	3 (2)
JL10	9	5 (1)
JL11	10	2 (1)
JL12	7	5 (2)
JL13	9	14 (2)
JL14	6	5 (2)
JL15	25	8 (6)
JL16	9	3 (3)
JL17	6	6 (4)

⁵ 語彙ネットワークの作成に 30 分、作文を書くのに 30 分の時間を取った。

⁶ Ereik Speed 氏によって開発された Google Chrome 用の和英辞書アプリで、Google Chrome の画面でカーソルを近づけると、意味がポップアップする。Chrome ウェブストアからダウンロードできる。

⁷ 学習者にコーパスツールを使用した個所に、ハイライトをしてもらった。

図2は、各学習者の「作文への使用数」と作文の文字数の関係を図で表したものである。「作文への使用数」が中央値の5回であった学習者（JL6、JL7、JL10、JL12、JL14）は、JL6を除き、200字から300字の範囲で作文を書いている。特徴的な学習者だったのは、JL4、JL13、JL15である。JL15は文字数が少ないが、「作文への使用数」が8回と多かった。また、JL4、JL13の文字数は中程度であったが、「作文への使用数」が14回と多かった。一方、「作文への使用数」が少なかったのは、JL9、JL11、JL16であった。これらの学習者は文字数も同程度であった。

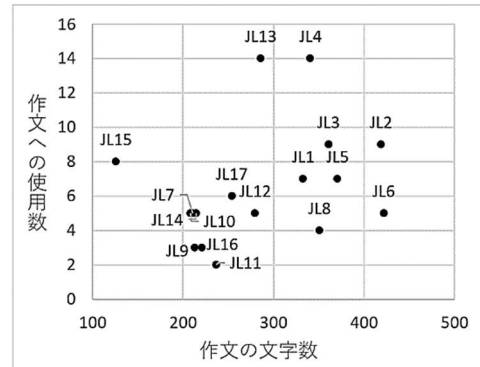


図2 「作文への使用数」と作文の文字数

5.2 使用パターンとコロケーションのレベル判定

次に、学習者がNLTを用いて、どのような表現を作文に取り入れたか分析した。表2は、NLTの分類に従い、使用パターンをカウントしたものである。最も多かったのは、「知識を習得する（JL3）」のような「名詞+助詞+動詞」で構成されたパターンで、全体の77%を占めていた。次いで、「人種差別（JL8）」のような「名詞」と「名詞」のパターン、「幅広い交流（JL15）」のような「形容詞」と「名詞」のパターンが多かった。

では、学習者はどの程度の難易度のコロケーションを作文に使用したのだろうか。表3はコロケーションに含まれる2語を「日本語教育語彙表（日本語学習辞書支援グループ, 2015）」を用いてレベル判定を行い、レベルの組み合わせに基づいてカウントしたものである⁸。

表2 使用パターン

名詞+助詞+動詞	86
名詞+名詞	7
形容詞+名詞	4
形容動詞語幹+な+名詞	3
動詞+名詞	3
名詞+の+名詞	3
名詞+的+な+名詞	2
名詞+名詞+接尾	2
接頭辞+名詞	1
合計	111

表3 コロケーションのレベル判定

中級前半・中級後半	33
中級後半・中級後半	15
初級後半・中級前半	12
中級前半・上級前半	12
中級前半・中級前半	9
中級後半・上級前半	6
初級前半・初級後半	5
初級後半・中級後半	5
初級後半・上級前半	4
初級前半・中級前半	3
初級前半・中級後半	3
初級前半・上級前半	2
初級後半・初級後半	1
上級前半・上級前半	1
合計	111

最も多かったのは、中級前半・中級後半（33）のコロケーションの組み合わせで、例えば、「経験を生かす（JL7）」、「文化を尊重する（JL13）」、「関係が築ける（JL1）」といったものが見られた⁹。次いで多かった中級後半・中級後半（15）の組み合わせでは、「急速な発展（JL15）」、「発展を目指す（JL15）」、「財産を管理する（JL2）」、中級前半・上級前半（12）の組み合わせでは、「分析を行う（JL10）」、「将来を見据える（JL16）」、「学術を研究する（JL1）」のようなものが見られた。

⁸ 日本語教育語彙表にない場合は、筆者らでレベル判定を行った。

⁹ 下線部はNLTで検索された言葉

また、初級レベルの簡単なコロケーションも検索されており、初級後半・中級前半 (12) の場合は、「試験に合格する (JL6)」、「良い計画 (JL2)」などがある。最も簡単なコロケーションの組み合わせだと思われる初級前半・初級後半 (5) の場合は、「会社に入る (JL13)」、「勉強を頑張る (JL17)」¹⁰などが見られたが、これらのコロケーションは全て作文を書いている最中に検索されたものであった。

5.3 語用分析

次に作文の中で NLT が使用された部分が適切であるかを分析した。適切さの分析では、NLT を用いて取り入れたコロケーション、あるいはコロケーションと共に使用された文法項目が原因で誤用と判断されたもの (以下:「コロケーションでの誤用」)、コロケーションが使用された節において誤用と判断されたもの (以下:「節での誤用」) に注目した。誤用については、市川編 (2010) の「脱落、付加、誤形成、混同、位置、その他」の分類を用いた。また、節については丸山他 (2016) で示された「主節、従属節:連用節、補足節、連体節」という分類を参考にした。例えば、JL12 は「日本人と交流したくて英語も使える仕事がしたいから、日本で国際的な企業を働きたい。」という文の主節において、NLT で調べた「国際的な企業」という表現を用いている。この場合、「国際的な企業」という表現は適切だが、「日本の国際的な企業で働きたい。」と修正することができる。したがって、「コロケーションでの誤用」はなく、「節での誤用」があると判断した。

今回、「作文への使用数」は全体で 111 あったが、そのうち 40 が「コロケーションでの誤用」、あるいは「節での誤用」が起きていると判断した。ただし、1 つの使用場面でも複数の誤用が起きている場合もあるため、その場合は、複数の誤用としてカウントした。表 4 はタイプごとの誤用数をカウントしたものの、表 5 は実際の誤用例である。

表 4 NLT 使用時における誤用数

コロケーションでの誤用					節での誤用				総数
脱落	付加	誤形成	混同	その他	脱落	付加	誤形成	混同	
4	4	5	14	1	6	2	2	12	50

表 5 NLT 使用時における誤用例

タイプ	誤用例		誤用の原因
コロケーションでの誤用	脱落	十分経験積んだ後で、～ (JL10)	「を」の脱落
	付加	根深いの問題について詳細に調べて、～ (JL5)	「の」の付加
	誤形成	日本語に興味を抱きて、(JL7) グローバルなビジネスに入りたい (JL12)	「抱いて」の誤形成 グローバルの誤表記
	混同	ビジネスアナリシスはニーズを定義し、(JL11) 興味を湧くようになれば～ (JL7)	「把握」と混同 「が」との混同
	その他	職場の雰囲気をよく感じて、(JL1)	文脈と合わない表現
節での誤用	脱落	私グローバルに活躍し (JL12)	「は」の脱落
	付加	日にほんごで苦手がなくなる (JL17)	「に」の付加
	誤形成	多いな勉強量が要求される。 (JL6)	「多くの」の誤形成
	混同	グローバルなビジネスに入りたい (JL12)	「携わりたい」との混同

¹⁰ 作文を書いている間に検索されたものであったため、どちらが検索語か判断できなかった。

表4からわかるように、「コロケーションでの誤用」において、最も多かったのは混同である。混同は表5のような言葉の混同（例：定義→把握）や助詞の混同（例：興味を湧く→興味が湧く）以外にも、文体、活用の混同等があった。次いで、多かったのは誤形成で、活用の誤用や誤表記が原因となっていた。一方、「節での誤用」でも混同が多く、誤用の原因も「コロケーションでの誤用」と類似していた。次いで多かったのは脱落で、助詞や形式名詞等の脱落が見られた。

5.4 語彙ネットワークと作文

次に、学習者がどのような語彙ネットワークを作成し、作文でどのような産出をしたかを見てみる。最も「作文への使用数」が多かった一人がJL4であった。図3はJL4が作成した語彙ネットワークで、図4はJL4の作文である。図4の_____（一重下線）は語彙ネットワークから使用された表現で、_____（二重下線）は作文を書いている間にコーパスツールを使い、調べた表現である。JL4の場合、「語彙ネットワークへの書き込み数」は13だが、そのうち10を作文で使用しており、効率的に語彙ネットワークの表現から作文に取り入れたことがわかる。そのため、語彙ネットワークの段階から、作文で何を書くかということが明確であったと考えられる。

	贅沢	贅沢品（良い意味ではないと感じるので、ハイエンドに変わる）
	ハイエンド	ハイエンド製品 ハイエンド市場
	賞金	を稼ぐ earn awards を獲得する achieve 100万円の賞金
	自習	に取り組む make effort に明け暮れ all the time, morning and evening
優れた 取材した	ルポルターージュ	活動
	キャリア	を積み上げる を磨く

図3 JL4の語彙ネットワーク

私が積み上げたい将来のキャリアはコンテンツストラテジストとして活動することである。具体的には、女性のファッション雑誌にハイエンド勢品の広告を載せるという仕事に携わりたいと思う。ゼミに入ってから、日本の雑誌についての研究を始める予定である。卒業後、日本のハイエンド市場で経験を十分に積んでから、ベトナムに帰国したいと思っている。なぜ贅沢品のコンテンツストラテジストになりたいかというと、キラキラなものを見ることが大好きからだ。_____に入学した後、様々な賞金を稼ぎ、選者の前に自分のキャリアについて発表したり、エッセイで書いたりした。最近、100万円の賞金を獲得し、将来のキャリアに進んだ。今後キャリアを磨くため、自習に明け暮れ、自分でハイエンド製品に関するルポルターージュ活動を行いたいと思う。

図4 JL4の作文

一方、語彙ネットワークからの使用がゼロだったのがJL8である。図5を見ると、JL8は2つのキーワードについて様々な表現を書き込んでいることがわかる。しかし、語彙ネットワークに2つのキーワードしかなかったこと、作文前に語彙ネットワークの表現をどのように使用するか十分にイメージしていなかったことから、作文を書いている間に使用する機会が見つけられなかったと考えられる。

苦労した、くろうした、 troubles・hardship 幅広い、はばひろい、 extensive	経験	が浅い、があさい have little experience が乏しい、がとぼしい、meagre が不可欠だ、がふかけつた、 indispensable
誤った、あやまった、 to misunderstand 優れた、すぐれた、 excellent	理解	を深める、をふかめる to have a better understanding を促進する、をそくしんする、 to encourage を促す、をうながす、 to prompt

図5 JL8の語彙ネットワーク

将来、政府で働き、外交官になりたい。なぜ外交官になりたいかというと、中学生からずっとネットで様々な国の人と話していた。その人々から様々な文化や見方を習い、非常に面白く、楽しかった。「なぜその人をそのように考えるか」「なぜそのようなことがわるい」な考え方をできおうした。異文化や異なる意見を理解することが初めてでき、自分が手に届く世界が広がった。その後、旅行の興味も引き付けた。美味しいな食べ物を食べたり、美しい景色を見たり、伝統的な服を着たり、歴史的な場所を行ったりすれば、ある国の文化や習慣を理解ことをできる可能性があるはず。しかしながら、世界中で戦争や人種差別などがあり、楽しく旅行と異文化を習うことが時々ない。そのため、卒業後、アメリカの政府の中にインターンシップし、様々な経験を積み重ねるつもりだ。

図6 JL8の作文

JL4 や JL8 以外にも特徴的な学習者がいる。例えば、JL13 の場合、語彙ネットワークの使用が2回と少ないが、作文を書いている間に調べたものが12回あった。JL13 が語彙ネットワークから用いたのは、「文化を尊重する」、「文化を取り入れる」といった中級前半から中級後半の表現であったのに対し、作文を書いている間に調べたものは、「店を開く」、「会社に入る」、「菓子を売る」などで、初級レベルの簡単な表現であった。一方、作文の文字数が126文字と少ないが、8回使用したのが JL15 であった。JL15 の場合、「幅広い交流」、「交流を促進する」、「急速な発展」、「発展を目指す」、「貢献を果たす」など、中級後半から上級レベルの表現が多く使用されていた。また、これらの表現は全て、語彙ネットワークから取り入れられたものであった。

全体を見ると、作文を書いている間にコーパスツールを使用した場合、初級レベルの簡単な表現を探す傾向が強かったわけではない。また、簡単な表現でも、学習者にとって必要な検索であれば、ツールの機能が活かされていると考える。しかし、JL13 のように語彙ネットワークを作文に活かせないという問題については、改善方法を検討するべきだろう。

6. まとめと今後の課題

本研究では「ライティングにおけるコーパスツール活用モデル(寺嶋・板井, 2021)」を用いた場合、学習者によって差異があるものの、作文で使用された表現の中に、語彙ネットワークから取り入れられたもの、作文を書いている間に検索されたものがあることが確認できた。また、作文で最も使用されたのは、「名詞+助詞+動詞」のパターンで、中級前半レベルと中級後半レベルの言葉で構成されたコロケーションであった。さらに、その適切さを分析したところ、コロケーション、あるいはコロケーションと共に使用された文法項目が原因で誤用と判断されたもの、コロケーションが使用された節において誤用と判断されたものがあり、「混同」による誤用が多いことがわかった。

学習者の誤用を見ると、コーパスツール上で表記や活用、助詞、さらには用例を確認する習慣を身に付けておくことで誤用を避けられるものもあっただろうと考える。学習者がそうした観察や分析が自然にできるようになれば、より自律的な学習が可能になる。そのためには、継続的に使用し、コーパスツールの使用に慣れることが必要だろう。

また、学習者の中には語彙ネットワークに書き込んだ表現をうまく取り入れられる者とそうではない者が見られた。このモデルの有効性をさらに高めるためには、語彙ネットワークの段階で作文がイメージできるようになる必要がある。例えば、語彙ネットワークから取り入れた表現を使った短文を作り、その短文を作文に使用するというプロセスを加えることで、より語彙ネットワークからの表現を取り入れられる可能性があるのではないだろうか。今後はそうした改善を加えたモデルについて検証をしていきたい。

謝 辞

本研究は JSPS 科研費 (18K00701「コーパスツールユーザーを育成するための教材開発とその検証」) の助成を受けたものである。

文 献

- 阿辺川武・ホドシチェク ボル・仁科喜久子 (2011) 「語の共起を効率的に検索できる日本語作文支援システム『なつめ』の紹介」言語処理学会 第17回年次大会 発表論文集, pp. 595-598.
- 赤瀬川史朗・プラシャント パルデシ・今井新悟 (2016) 『日本語コーパス活用入門』大修館書店.
- Hodošček Bor・阿辺川武・Bekeš Andrej・仁科喜久子 (2011) 「レポート作成のための共起表現産出支援－作文支援ツール『なつめ』の使用効果－」専門日本語教育研究 13号, pp. 33-40.
- 市川保子 (編) (2010) 『日本語誤用辞典 外国人学習者の誤用から学ぶ日本語の意味用法と指導のポイント』スリーエーネットワーク.
- 丸山岳彦・佐藤理史・夏目和子 (2016) 「現代日本語における節の分類体系について」言語処理学会第22回年次大会発表論文集, pp.1113-1116.
- 日本語学習辞書支援グループ (2015) 「日本語教育語彙表 Ver 1.0 (<http://jhlee.sakura.ne.jp/JEV/>よりダウンロード可能)
- 寺嶋弘道・板井芳江 (2021) 「ライティング力の向上を目指したコーパスツールワークショップの試みとその考察」APU 言語研究論叢 6号, pp. 49-65.

関連 URL

- Erek Speed 『rikaikun』 <https://chrome.google.com/webstore/detail/rikaikun/jipdnfibhldikgcjhfnomkfpcebammhp?hl=ja>
- 筑波大学・国立国語研究所・Lago 言語研究所『NINJAL-LWP for TWC』 <https://tsukubawebcorp.us.jp>

児童作文における書き出しと結びの分析

加藤 恵梨 (愛知教育大学教育学部)

A Study of the Opening and Closing Lines of Elementary Students' Compositions

Eri Kato (Aichi University of Education)

要旨

本研究は2019年度の「清流環境作文コンクール」応募作品を電子化した『清流環境作文コーパス』の中の、「イタイタイ病」に関して書かれたものを調査対象とし、発達段階において、児童が作文の書き出しおよび結びにどのような形式や内容等を好んで用いるのかを明らかにするため、2年生、4年生、6年生が書いた470編を分析した。その結果、それぞれの学年に次のような特徴があることがわかった。2年生は、自身の過去の経験・行動・見聞について書くことで文章を始め、自身の今後の行動について述べて文章を終えているものが多い。4年生は、自身の過去の経験・行動・見聞やイタイタイ病について述べることで文章を書き始めているが、結びでは今後の抱負など、物事に対する考えや思いを述べているものが多い。6年生は、書き出しと結びに自身の考えや思いを述べているものが多いがそれだけではなく、書き出しと結びの内容が一貫しており、書き出しと結びを読めば文章の意図が推察できるようなものも見られる。

1. はじめに

本稿は児童作文における書き出しと結びに注目し、学齢があがるにつれてそれらの表現にどのような変化が見られるのかを明らかにすることを目的とする。

児童にとって、書き出しに何を書くかを考えるのは容易ではない。太田(1965: 10)は、子どもが作文を書く時に最初にぶつかる大きな壁が書き出しであるとし、「手紙などでも、主文の起しことばによって、そのあとに続く文章が決まってくるように、作文でも、最初の文の書き出しで文章全体の調子が決まり、結論の方向まで示されるとさえいわれている。」と述べている。また、田近・井上・中村(編)(2018: 133)にも、「『書き出し』をどう始めればよいかわからないという子どもがいる。徒競走の時でも、スタートの時の失敗は後々までなかなか取り返しがつかないものである。」という記述がある。このように、書き出しは児童にとって難しいものであるが、文章全体に関わる重要なものであることが指摘されている。さらに、櫻本(1996: 404)も書き出しの重要性を述べ¹、小学校段階では「書き出しは、文章全体の構成にかかわる」「書き出しと結びを読めば文章の意図がほぼ推察できる」「『相手意識や目的意識をはっきりさせる』ために役立つ」という役割があると考え、書き出しの指導にあたるのが適当だと述べている。

以上から、作文の結尾だけでなく、冒頭の書き出しも、文章全体に関わる重要な文であることがわかる。本稿では、2019年度の「清流環境作文コンクール」応募作品を電子化した『清流環境作文コーパス』の中の、「イタイタイ病」に関して書かれた作文を調査対象とし、発達段階において、児童が書き出しおよび結びにどのような形式や内容等を用いているのかを調査・分析する。なお、書き出しは作文の冒頭の1段落を指すこともあれば、1文を指すこともあるが、本稿では、書き出しを作文の冒頭の1文とする。同様に、結びについても、作文の結尾の1段落ではなく、最後の1文とする。

¹ 櫻本(1996: 404)は書き出しを「冒頭の1文ないしは1段落」としている。

2. 書き出しについての先行研究

結びの類型についての先行研究は管見の限り見られないため、以下では書き出しの類型に関する先行研究を概観する²。

2.1 書き出しの類型について

児童作文の書き出しの類型と、その学年的推移について分析した研究に、太田(1965)がある。太田(1965: 11)は、書き出しの類型を表1のように示している³。

表1 太田(1965)の書き出しの類型⁴

1. 人物的発想	自己中心的発想	ぼくはおとうさんといっ ⁵
	人物中心の発想	おかあさんはチョコレートが好きで
	共同体的な発想	わたしたちのクラスは今年4月に
2. 日・時的発想	限定した期日から	去年のお正月の夜のことでした
	特定な時から	給食の時でした
	不特定な時から	ある日 この間は
	形式的な時から	毎年 毎日
3. 場所的発想	場所、場面的発想	グラウンドでドッチボール大会が
4. 自己行動からの発想		朝ごはんを食べていると
5. 他人行動からの発想		(記述なし)
6. 事象・状態からの発想	説明的情景描写	どんよりくもって降りそうな空です
	擬声語からの描写	「リーン」とベルの音がした
7. 心理状態からの発想		「しまった」と思ったがもうおそかった
8. 抽象的な説明からの発想	抽象的な文	待ちに待った運動会がやって来た
	引用文からの発想	「天は人の上に人をつくらず…」と
9. 会話からの発想		「今から版画を彫ります。」と先生が ぼくたちの頭を

² 田近・井上・中村（編）（2018:37）は、「文章の冒頭と結尾の型」について次の8つを挙げている。

- ①主題・要旨・結論・提案などを述べる（冒頭にも結尾にも）。
- ②主要な題材・話題について述べる（冒頭にも結尾にも）。
- ③粗筋・筋書きを述べる（冒頭にも結尾にも）。
- ④筆者の立場・意向・執筆態度などを述べる（冒頭にも結尾にも）。
- ⑤本題の内容を規定し、本題に枠をはめる（冒頭にも結尾にも）。
- ⑥導入として、時・所・登場人物を紹介したり、前置きの話を置いたりする（冒頭）。
- ⑦本題と関連のある事柄や感想などを、しめくくりとして述べる（結尾）。
- ⑧以上のような冒頭または結尾を置かず、本題を構成する一部として書き始め、また、書き終えるもの。

このように、結尾の型を示している研究はある。

³ 太田(1965: 11)は表1について、「すべての作文が全部この類型にあてはまるというわけではない」と述べている。

⁴ 表1は、太田(1965)のp.11の記述と別表1をもとに作成した。

⁵ 太田(1965: 11)があげている例には文末まで示されていないものがある。また、「5. 他人行動からの発想」のように、例があげられていないものもある。

表1のように、太田(1965)は書き出しを9つの類型に分類している。それらの中にはさらに分類されているものもあり、例えば「2. 日・時的発想」は「限定した期日から」「特定の時から」「不特定の時から」「形式的な時から」という4つに下位分類されている。なお、最右列に示しているのは各々の分類の例である。

また、太田(1965: 11)は、書き出しの一般的傾向、その標準性を把握するために、『年刊児童文詩集 (日本作文の会編・応募入選作品集)』1960年版、1961年版、1962年版、1963年版、1964年版について読書感想文を除いた全作品を分析している。その結果、「1. 人物的発想」(36%)と「2. 日・時的発想」(40%)がその大半を占めており、「1. 人物的発想」については、「自己中心的発想」が高学年に進むに従って「人物中心の発想」へ移行しているとともに、「共同体的な発想」も学年を追って順次増加している。このことから、「子どもたちの社会連帯的意識の成長がつかめる」と述べている。一方、「2. 日・時的発想」については、2年(49%)と3年(50%)が最高であるが、他の学年でもみな30%を越えていることから、「『子どもの作文は、過去のできごとを過去形で書いたもの。』と言われるのも当然で、作文を書く最初は期日や時間を思い起こすことがすでに通念となっているようだ」と指摘している。

2.2 児童作文の書き出しに良く見られる文

表1を用いることで、児童作文における書き出しの多くが分類できると思われるが、その他に、テーマを取り込んだ文が児童作文の書き出しに良く見られることが指摘されている。宮城・松崎(2015: 159)は松崎(2015)を踏まえ、「作文を書く際に提示されたテーマを取り込んで『〇〇は』とするのは、小・中学生の“書き癖”になっている」と述べている。本稿が調査資料とするコンクール応募作文においても、事前に作文のテーマが示されているため、書き出しにテーマを取り込んだ文が多く見られることが予想される。

3. 調査資料について

分析に先立ち、本稿が用いる調査資料について述べる。本稿が調査資料とするのは、『清流環境作文コーパス Ver.1.0』である。これは2019年度の「清流環境作文コンクール」応募作品を電子化した、小学校1年生から6年生の作文による112万形態素規模のコーパスである。平文テキストに加えて、形態論情報(短単位・長単位)、係り受け情報、節境界ラベル、モダリティ情報などの言語的情報が付与してある(今田・宮城・阿部・清水 2021: 1)。また、コーパスは内容により、「A イタイイタイ病に関するもの」「B 自然環境、周りの環境についてのもの」「C 実験、環境保全、調べ物、学習的なもの」「X 分類不能、複数にまたがるもの」に分類されている⁶。本稿では、「A イタイイタイ病に関するもの」を中心に、イタイイタイ病について書かれている作文のうちの、2年生、4年生、6年生が書いた470編(2年6編、4年75編、6年389編)を調査する⁷。2年生、4年生、6年生を低・中・高学年の代表とし、発達段階において児童が書き出しおよび結びにどのような形式や内容等を用いているのかを分析する。

なお、コンクール応募作文の応募要項には、作文のテーマとして「イタイイタイ病について感じたこと、考えたこと」とあり、応募作品の字数は、1・2年生が600字以内、3・4年生が1000字以内、5・6年生が1200字以内となっている。

⁶ 分類はコーパスの作業者が主観的に判定したものである(今田・宮城・阿部・清水 2021: 2)。

⁷ 「A イタイイタイ病に関するもの」のファイルの中には、イタイイタイ病とは関係のない内容の作文も含まれている。そのため、筆者が他のファイルの全作文についても内容を確認し、イタイイタイ病について書かれていると判断した作文を抽出し、調査対象とした。

4. 分析

以下では、まず、本稿が調査対象とする作文の書き出しと結びの類型を示す。それをもとに、年齢があがるにつれてそれらにどのような違いがあるのかについて分析する。

4.1 書き出しについて

4.1.1 本研究における書き出しの類型

表1を参考にし、調査資料の書き出しの類型を示したものが表2である。

表2 書き出しの類型

大分類	中分類	例
1. 人物的発想	自己中心的発想	私はイタイイタイ病について調べました。
	人物中心の発想	みなさんは「イタイイタイ病」を知っていますか。
	共同体的な発想	私たちはとてもめぐまれていると思います。
2. 日・時的発想	特定な時から	きょう、イタイイタイびょうしりょうかんに行きました。
	不特定な時から	昔、この富山市婦中町にある神通川にカドミウムというものがたくさんふくんで流れてきました。
3. 場所的発想	場所から	富山県は、水がきれいでおいしいです。
	場面から	社会の授業で四大公害病について勉強しました。
4. テーマからの発想	テーマから	イタイイタイ病は、神通川の上流から流れてきたカドミウムが原因の病気だ。
	テーマに関連する事物から	公害病を二度と起こしたくない。
5. 事象・状態からの発想		ここに一枚の写真があります。
6. 心理状態からの発想		「今、起こったらどうになってしまうのだろう・・・」語り部さんのお話を聞いたとき私はこう思いました。
7. 抽象的な説明からの発想		真実を真実として伝えていくことがぼくたちの使命だと思う。

表2のように、本稿の調査資料の書き出しは「1. 人物的発想」「2. 日・時的発想」「3. 場所的発想」「4. テーマからの発想」「5. 事象・状態からの発想」「6. 心理状態からの発想」「7. 抽象的な説明からの発想」という7つに大別できる。

ただし、表2は表1と全く同じではなく、大きく異なる点がある。まず大分類について、調査資料には「自己行動からの発想」「他人行動からの発想」「会話からの発想」が見られなかったため、削除した。また、宮城・松崎(2015)が指摘していたように、作文のテーマを主語として書かれた文が調査資料にも見られたため、大分類に「4. テーマからの発想」を加え、その中分類を「テーマから」と「テーマに関連する事物から」とした。前者はイタイイタイ病を主語として書かれた文を指し、後者はイタイイタイ病に深く関わりをもつ事物を主語として書かれた文を指す。次に中分類について、表1では「2. 日・時的発想」を4つに下位分類していたが、そのうちの「限定した期日から」と「特定な時から」には大きな違いがないので一つにまとめ、「特定な時から」とした。「形式的な時から」は調査資料に見ら

れなかったため、削除した。また、「3. 場所的発想」は表1では「場所、場面的発想」とされていたが、「場所」と「場面」は異なるため、「場所から」と「場面から」に分けた。さらに、表1の「事象・状態からの発想」の「擬声語からの描写」も見られなかったため、削除した。よって、表2の「5. 事象・状態からの発想」は、「説明的情景描写」を指す。加えて、表1の「抽象的な説明からの発想」の「引用文からの発想」も見られなかったため、削除した。

4.1.2 学年別による書き出しの種類

次に、学年によって書き出しの種類に違いがあるのかを見ていく。学年別による書き出しの種類の割合を示したのが表3である。

表3 学年別の書き出しの種類

大分類	中分類	2年 (6編)	4年 (75編)	6年 (389編)
1. 人物的発想	自己中心的発想	2 (33.3%)	45 (60%)	266 (68.3%)
	人物中心の発想	1 (16.7%)	0	2 (0.5%)
	共同体的な発想	0	0	20 (5.1%)
2. 日・時的発想	特定な時から	1 (16.7%)	3 (4%)	18 (4.6%)
	不特定な時から	0	1 (1.3%)	7 (1.8%)
3. 場所的発想	場所から	1 (16.7%)	1 (1.3%)	11 (2.8%)
	場面から	0	0	4 (1.0%)
4. テーマからの発想	テーマから	1 (16.7%)	20 (26.7%)	45 (11.6%)
	テーマに関連する事物から	0	4 (5.3%)	4 (1.0%)
5. 事象・状態からの発想		0	0	1 (0.3%)
6. 心理状態からの発想		0	1 (1.3%)	10 (2.6%)
7. 抽象的な説明からの発想		0	0	1 (0.3%)

各学年の作文数が異なるため、単純に比較することはできないが、すべての学年で最も多いのは「1. 人物的発想」の「自己中心的発想」であり、「わたし／ぼく」から書き始められている。

4年生に注目すると、二番目に多いのは「4. テーマからの発想」の「テーマから」であり、「イタイタイ病は」で文が始められている。4年生の書き出しは、「1. 人物的発想」と「4. テーマからの発想」で約9割を占める。

6年生においても、4年生と同様、「1. 人物的発想」と「4. テーマからの発想」で約8割を占める。しかし、6年生が4年生と異なる点は、「1. 人物的発想」のうちの「共同体的な

発想」が用いられている点である。また、6年生では「6. 心理状態からの発想」によって書き始められている文も増えている。

このように、年齢があがるにつれて「わたし／ぼくは」といった自己中心的な発想による書き出しや、「イタイイタイ病は」といったテーマからの発想による書き出しが少しずつ減り、多角的な発想で文章を書き始めようとしていることが伺える。この点は、2.1 で取りあげた太田(1965: 11)の分析結果と大きく異ならない。以下では、各学年の特徴について詳細に見ていく。

4.1.3 2年生の書き出しの特徴

4.1.2 で見たように、2年生は「1. 人物的発想」のうちの「自己中心的発想」からの書き出しが最も多いのであるが、その内容を見ると次の(1)のように、「過去の経験・行動・見聞」について書かれている。

(1) わたしは、イタイイタイびょうについて知らなかったのかぞくに聞いてみました。

また、「4. テーマからの発想」の「テーマから」においても、書き出しは「イタイイタイ病は」で始まっているが、その内容は(2)のように、「過去の経験・行動・見聞」について書かれている。

(2) イタイイタイびょうは、ニュースで言ばは聞いたことがあるけど、どんなびょう気か、いつおきたか、どこであったか知りませんでした。

このように、2年生の書き出しの内容は、自身の「過去の経験・行動・見聞」(5 (83%))が大部分を占めるのが特徴である。

4.1.4 4年生の書き出しの特徴

4年生の書き出しは「1. 人物的発想」と「4. テーマからの発想」で約9割を占めるが、その内容を見ると、2年生と同様、「過去の経験・行動・見聞」(43(57%))が最も多い。それだけではなく、次の(3)のような「これから書くことの予告」をしているもの(6(8%))⁸、(4)のように「感じたこと・思ったこと」が書かれているもの(3(4%))も見られる。

(3) わたしは、安全な水の大切さについて考えてみました。

(4) わたしは、イタイイタイ病になった人をかわいそうだと思います。

以上から、4年生の書き出しの内容は、自身の「過去の経験・行動・見聞」だけではなく、自身が「感じたこと・思ったこと」を述べたり、後の文とのつながりを考えて「これから書くことの予告」をしたりと、2年生よりも書き出しに工夫が見られる。

4.1.5 6年生の書き出しの特徴

6年生の書き出しにおいても「過去の経験・行動・見聞」(159(41%))が最も多いが、その他に、4年生では見られなかった、次の(5)のような「願望」を述べたもの(58(15%))や、(6)のように「読み手への問いかけ」(4(1%))がされているものが見られる。

⁸ 石黒(2017: 242)は、冒頭部に出現し、後続文脈の展開を規定する予告文として、「疑問により予告文」「宣言による予告文」「列挙による予告文(数字)」「列挙による予告文(多様)」「結論による予告文」という5種を挙げている。(3)も予告文と考えることができる。

- (5) ぼくは、イタイタイ病などの公害病が二度とおこらないでほしいです。
 (6) みなさんは「イタイタイ病」を知っていますか。

4.1.2 で、6年生が4年生と異なる点は、「1. 人物的発想」のうちの「共同体的な発想」や「6. 心理状態からの発想」が増えていることであると述べたが、「自己中心的発想」においても、自身の願望や考えを述べているものが多い。このことから、4年生の時よりも、より多角的な発想で文章を書き始めていることがわかる。

4.2 結びについて

4.2.1 本研究における結びの類型

次に、調査資料の結びについて分析する。本稿では結びの文末表現に注目し、表4のように分類した。なお、結びの多くが「～と思いました／と思います」で終わっている(2年生4(57%)、4年生44(59%)、6年生78(20%)。そのため、そのような場合には「～と思いました／と思います」に前接する表現にもとづいて分類を行った。

表4 結びの類型

分類	例
1. 願望	水をムダづかいせずに、大切につかいたいと思いました。
2. 意志	神通川はこれからもきれいな川で、そして、イタイタイ病をわすれないようにします。
3. 意見・主張	これからもきれいな水が飲めるように一人一人努力しなければいけないと思いました。
4. 感じたこと・思ったこと	イタイタイ病のことを知り、自然の大切さが学べてよかったです。
5. 考え	二度と公害を繰り返さず美しく豊かな環境を未来に引き継ぐこと、それが私たちの使命だと思いました。
6. 事象の説明	この工事は二千十二年に終わりました。
7. 前で述べたことのまとめ	これがぼくが出した答えです。
8. 読み手への呼びかけ	みなさんも、この先イタイタイ病を起こさない、ちょっとした気づかいを心がけましょう。

表4のように、調査資料の結びは「1. 願望」「2. 意志」「3. 意見・主張」「4. 感じたこと・思ったこと」「5. 考え」「6. 事象の説明」「7. 前で述べたことのまとめ」「8. 読み手への呼びかけ」という8つに大別できる。調査資料の結びの多くが、分類1から5のように、自身の考えや思いを述べて文章を締めている。

4.2.2 学年別による結びの類型

次に、学年によって好まれる結びに違いがあるのかを見ていく。学年別の結びのタイプの割合を示したのが表5である。

表5 学年別の結びの種類

分類	2年 (6編)	4年 (75編)	6年 (389編)
1. 願望	3 (50%)	37 (49.3%)	209 (53.7%)
2. 意志	1 (16.7%)	3 (4%)	19 (4.9%)
3. 意見・主張	1 (16.7%)	11 (15%)	55 (14.1%)
4. 感じたこと・思ったこと	1 (16.7%)	19 (25%)	56 (14.4%)
5. 考え	0	5 (6.7%)	37 (9.5%)
6. 事象の説明	0	0	9 (2.3%)
7. 前で述べたことのまとめ	0	0	3 (0.8%)
8. 読み手への呼びかけ	0	0	1 (0.3%)

表5についても各学年の作文数が異なるため、単純に比較することはできないが、すべての学年で最も多いのは「1. 願望」である。「～たい」という表現が多く用いられ、年齢があがると、「～ていきたい」が今後の抱負を述べるのに使われている。また、6年生では、「～てほしい」という表現を用い、他者への要望を述べているものも見られる。

4年生で二番目に多く用いられているのは、「4. 感じたこと・思ったこと」である。「よかったです」という表現で、イタイタイ病について知ることができたことに対する感想が述べられているものが多い。

6年生で二番目に多いものも「4. 感じたこと・思ったこと」であるが、「3. 意見・主張」がほぼ同率である。6年生においても、「4. 感じたこと・思ったこと」ではイタイタイ病について学ぶことができ「よかったです」などと述べられているものが多い。しかし、「3. 意見・主張」については、4年生では「～なければならぬ／なくてはいけない」という形式を用い、強い主張がなされているが、6年生では「～が私たちにできることだ」「～につながっていくのだ」といった表現を用い、4年生に比べると控え目に主張がなされているという特徴がある。また、6年生では、2年生や4年生ではなかった「6. 事象の説明」「7. 前で述べたことのまとめ」「8. 読み手への呼びかけ」の文が見られるという特徴もある。このように、6年生になると、自身の考えや思いを述べて文章を終えるという方法以外で書かれた文章も見られる。

4.2.3 2年生の結びの特徴

4.2.2 で見たように、2年生の結びは「1. 願望」が最も多いのであるが、その内容は次の(7)のように、自身が実際に行えそうな身近な行動・活動をあげ、それを行いたいと述べているものが多い。その傾向は、次の(8)のような「2. 意志」の内容にも見られる。(8)でも、「私にできることから少しずつ実行していこう」と述べられている。

(7) 水をムダづかいせずに、大切につかいたいと思いました。

(8) むだに水を使わない、海や川をよごさないなどまず私にできることから少しずつ実行していこうと思いました。

このように、2年生の結びの内容を見ると、自身の身近な行動・活動に関して書かれている。2年生は、書き出しも自身の過去の経験・行動・見聞についてであることから、自身の過去の経験・行動・見聞によって文章を始め、自身の今後の行動について述べて文章を終えるというように、自己中心的発想が根本にあると考えられる。

4.2.4 4年生の結びの特徴

4年生の結びにおいても「1. 願望」が最も多く、全体の約半数を占めている。4年生では次の(9)のように、これからも継続的に行動・活動を行いたいという抱負を述べているものが多い。そのため、「～ていきたい」という表現が多く用いられている。また、2年生では実際にできそうな身近な行動・活動をあげ、それを行いたいと述べているものが多かったが、4年生では次の(10)のように、実現を望む行動・活動の内容が壮大であるものが多い。さらに、次の(11)のように、自身だけで行動するのではなく、「みんなで」行動したいと願っているものが見られる点も2年生と異なる。

- (9) ぼくはまだ子どもだけど、これからもイタイイタイ病のことを調べて、次の若い人たちに伝えていきたいと思いました。
- (10) ぼくは、イタイイタイ病のげんいんのカドミウムのたいさくをしてイタイイタイ病をなくしたいと思いました。
- (11) みんなで自ぜんを大切に、もう二度とかなしい思いをしないかんきょうをつくっていききたいです。

このように、4年生では自身の身近なことについて述べるだけでなく、もっと広い範囲に目を向け、叶えるのが難しそうな大きな願望についても述べている。

二番目に多い「4. 感じたこと・思ったこと」では、次の(12)や(13)のように、自身が当事者の立場になったらどのように感じるかを想像し、相手の立場で「感じたこと・思ったこと」を述べているものが多い。

- (12) わたしのお母さんも三十五で、わたしをうんだのでこわいと思いました。
- (13) イタイイタイ病のことについて調べていて、ぼくが思ったことは、ほねがおれたら歩けないし遊べないからすごくかわいそうだなと思いました。

以上のように、4年生は、書き始めは自身の過去の経験・行動・見聞やイタイイタイ病についてであるが、結びでは、大きな願望や、当事者の立場になって想像したことなど、より広い視点で物事についての考えを述べることで文章を縮めている。このように、4年生では身近なことから出発し、より発展的な内容で文章を終えているものが見られる。

4.2.5 6年生の結びの特徴

6年生の結びにおいても「1. 願望」が最も多く、全体の半数以上を占めている。6年生も4年生と同様、今後継続的に行動・活動を行いたいという抱負を述べているものも多く、「～ていきたい」という表現が多く見られる。しかし、次の(14)のように実現を望む行動・活動の内容として自身が取り組みそうなことをあげているものが多い点が4年生と異なる。6年生になると、大きな願望を語るだけではなく、実際に自身に何ができるのかを考え、そのことを文章の最後に表明することが多いと考えられる。

- (14) まずは自分ができることから取り組んでいききたいです。

また、次の(15)のように、「～してほしい」という表現を用い、他者へ望むことを述べているものも多い。このように、6年生では自身がしたいことだけではなく、他者へ望むことも示されている。

- (15) 平和をずっと続けるということは、単純に聞こえるかもしれないけれど、そのうらで、これらの方々が必死に地域を守ってくれていることを忘れないでほしいと思います。

4.2.2 で「3. 意見・主張」について、4年生では「～なければならない／なくてはいけない」という形式を用い、強い主張がなされているが、6年生では「～が私たちにできることだ」「～につながっていくのだ」といった表現を用い、4年生に比べると控え目に主張がなされていることを指摘した。このことも先ほど述べた、「6年生になると、大きな願望を語るだけではなく、実際に自身に何ができるのかを考え、そのことを文章の最後に表明している」とことに関連していると言える。次の(16)のように、6年生は実際に自分たちに何ができるのかを考え、行動しようとするため、意見・主張を述べる際にも内容がより具体的で現実的なものとなっている。

- (16) 私たちはこのことを次の世代へまた、次の世代へと語りつぎ、絶対に忘れないようにするのが今、私たちにできることだと私は思いました。

その他にも、6年生では、次の(17)のように読み手へ注意を促して文章を締めているものや、(18)のように前に述べたことのまとめをしているものも見られる。

- (17) みなさんも、この先イタイイタイ病を起こさない、ちょっとした気づかいを心がけましょう。
(18) これがぼくが出した答えです。

6年生の書き出しでは、自身の過去の経験・行動・見聞だけではなく、考え・思いを述べたり、読み手への問いかけがされたりしているが、結びにおいても、書き出しと対応させるかのように、自身の考え・思いを述べて締めくくられているものが多い。よって、次の(19)や(20)のように、書き出しと結びの内容が一貫しており「書き出しと結びを読めば文章の意図がほぼ推察できる」（櫻本 1996: 404）ようなものも見られる。

- (19) 【書き出し】 私たちはとてもめぐまれていると思います。
【結び】 これらのことを実行し、公害を絶対にわすれず、そして今楽しく生活できている幸せさをわすれずに生きていきたいと思いました。
(20) 【書き出し】 私がイタイイタイ病について感じたことは、この世の中はどんどん便利になっていくけれど、人々はこのせいでアレルギーや病気になっていくのだなと感じました。
【結び】 だから私は病気に気をつけながら、この便利な世の中を元気に生きていきたいです。

5. おわりに

本稿では2019年度の「清流環境作文コンクール」応募作品を電子化した『清流環境作文コーパス』の中の、「イタイイタイ病」に関して書かれたものを調査対象とし、発達段階において、児童が書き出しおよび結びにどのような形式や内容等を好んで用いるのかについて分析を行った。その結果、2年生、4年生、6年生の書き出しと結びに次のような特徴があることがわかった。

2年生の特徴

- ・自身の過去の経験・行動・見聞によって文章を書き始め、自身の今後の行動について述べて文章を終えることが多い。自己中心的発想が文章を書く際の根本にあると考えられる。

4年生の特徴

- ・自身の過去の経験・行動・見聞やイタイタイ病について述べることで文章を書き始めているが、結びでは、叶えるのが難しそうな大きな願望について述べたり、当事者の立場になって想像したことを表現したりするなど、広い視点で物事についての考えを述べ、文章を縮めている。このように、身近なことから出発し、発展的な内容で文章を終えているものが見られる。

6年生の特徴

- ・自身の過去の経験・行動・見聞だけではなく、考え・思いや、読み手への問いかけなどから文章を書き始めているものが多く見られる。また、結びにおいても、書き出しと対応させるかのように、自身の考え・思いを述べて締めくくられているものが多い。よって、書き出しと結びの内容が一貫しており「書き出しと結びを読めば文章の意図がほぼ推察できる」(櫻本 1996: 404)ようなものも見られる。

今後は、1年生、3年生、5年生の作文についても調査・分析し、児童作文の書き出しと結びの実態を明らかにしたいと考える。

謝 辞

本研究は、科研費基盤研究(B)「コンクール応募作文を資料とした児童の文章作文能力の評価と発達段階の包括的研究」(2020年-2024年、代表者:宮城信)による成果の一部です。

文 献

- 石黒圭(2017). 「説得力のある全体構造の作り方」石黒圭(編)『わかりやすく書ける作文シラバス』くろしお出版, pp.225-244.
- 今田水穂・宮城信・阿部藤子・清水由貴子(2021). 「清流環境作文コーパス Ver.1.0」
- 太田康治(1965). 「作文指導 その反省と考察—書き出し文による発想の類型及びその傾向について—」『教育実践研究集録/新潟県立教育センター』2, pp.9-14.
- 櫻本明美(1996). 「説明的表現の指導—書き出しの工夫に着目して—」『神戸親和女子大学研究論叢』30, pp.400-422.
- 田近洵一・井上尚美・中村和弘(編)(2018). 『国語教育指導用語辞典〔第五版〕』教育出版
- 松崎史周(2015). 「中学生の作文に見られる『主述の不具合』の分析—出現傾向から学習者の表現特性を探る—」『解釈』61, pp.12-20.
- 宮城信・松崎史周(2015). 「児童作文における『理由述べ』表現」『全国大学国語教育学会国語科教育研究:大会研究発表要旨集』129, pp.157-160.

「つまり」による換言が促す理解の範囲について

櫻井芽衣子 (日本工業大学)

The Scope of Influence that Promotes Understanding by Paraphrasing

“*Tsumari*”

Meiko Sakurai (Nippon Institute of Technology)

要旨

日本語記述文法研究会 (2009) 『現代日本語文法 7』によると、「つまり」は先行部を具体的に説明したり要点をまとめたりするものである。文脈を考慮に入れ、「つまり」による換言の様相を『現代日本語書き言葉均衡コーパス』で見ると、後続部は必ずしも先行部を分かりやすく言い換えるとは限らないが、先行部に情報を追加し、先行部だけでなく、まとまりのある文章として理解を深める場合があることを指摘する。先行部の理解を促す換言と、文章の理解を促す換言、これらの換言は、石黒圭 (2001) 「換言を表す接続語について——『すなわち』『つまり』『要するに』を中心に——」『日本語教育 110 号』で指摘される換言の二つの目的に沿うものであることも確認する。

1. はじめに

換言の接続表現「つまり」は、表現を平易にする、具体化・抽象化する、観点を変える等、先行部を様々な形で言い換えて示すもので、「つまり」によって提示された後続部は、聞き手の理解を促すような先行部の解釈である (日本語記述文法研究会 2009)。換言された結果、先行部についての理解は深められる。

- (1) 実家では甲斐犬、つまり、オオカミのような毛色と立ち耳が特徴的である中型の日本犬を飼っている。

換言された後続部によって先行部「甲斐犬」のイメージが明確になり、理解しやすくなる。先行部「甲斐犬」の知識が不十分な読み手にとって、理解を促す換言であるといえる。逆に、「甲斐犬」に詳しい読み手にとっては、なくてはならない換言ではない。

その一方で、同じ「つまり」を使った換言であっても、「つまり」を含む一文だけでは理解が促されるとはいえない換言もある。

- (2) 隣家から甲斐犬、つまり伝統的に現在の山梨県の辺りで飼育されていた猟犬が柵を越えて家の庭に侵入し、柴犬のポチと仲良くじゃれていた。

先行部「甲斐犬」の換言として誤っているわけではないが、読み手によっては後続部を不要な情報であると感じ、その結果、却って理解を妨げるおそれがある。しかし (3) のような後続文脈がある場合、換言が重要な意味を持つ。

- (3) 隣家から甲斐犬、つまり伝統的に現在の山梨県の辺りで飼育されていた猟犬が柵を越えて家の庭に侵入し、ポチと仲良くじゃれていた。南アルプスの岩場を駆け巡って狩りをしていたため遺伝的に飛節が発達しており、跳躍を得意とする甲斐犬にとって、庭を区切る程度の柵を飛び越えることなど訳もない。何度も侵入を繰り返したことで、ポチもすっかり慣れている。

「山梨県」「猟犬」という後続部の記述が、後続文脈「南アルプスの岩場を駆け巡って

狩りをしていた」「跳躍を得意とする」へと繋がり、何度も柵を飛び越えているため隣家の犬と仲が良いという文章全体が理解しやすくなる。先行部「甲斐犬」に詳しい読み手であっても、換言されていた方が内容を読み取りやすくなる。

(1) の換言は先行部の理解を促すものであるといえ、(2) の換言は (3) との比較において明らかであるように、文章レベルでの理解を促すためのものであるといえる。(2) について、(1) と同じように換言を含む文だけを取り出して分析しては掘り切れないことがある。一見意味のないような換言であっても後続文脈を補うと内容理解が進むことから、換言を分析する際、文脈を考慮することは必要であると考えられる。

本発表では文脈も考慮しながら換言を分析し、換言は、読み手にとって必ずしも理解しやすいように先行部を言い換えるわけではないが、先行部に関する何らかの情報を追加するものであること、後続部が文脈の内容に裏付けられているか否かという観点から、三つの換言があることを指摘する。また、それぞれの換言によって、理解を促す範囲が異なることにも触れる。

2. 考察対象

考察の対象としたのは、接続表現「つまり」の前後に、「時」「人」あるいは、そのいずれかを主要部とする名詞句が来るものである。『現代日本語書き言葉均衡コーパス』(以下、BCCWJ) からコーパス検索アプリケーション『中納言』を用いて収集した。「時」「人」に限定したのは、多様な先行部・後続部ではなく、類似の性質を持つものを比較するためである。条件を揃えるためには同一の固有名詞や一般名詞を対象とすべきだが、用例数が少なかったため、国立国語研究所(2006)を参考に高頻度で出現するものを対象とした。

中納言の長単位検索を用い、名詞あるいは名詞句同士の換言を見るために、「つまり」から3語以内に「時」もしくは「人」を含むという前方共起条件、あるいは、10語以内に「時」もしくは「人」を含むという後方共起条件を設定して検索した。抽出されたもののうち、名詞(句)同士の換言であることが確認できるものを選定している。

3. 換言の様相

「つまり」の前後で言い換えている名詞(句)のうち、「つまり」の前に提示されているものを先行部、後のものを後続部と呼ぶ。また、「つまり」を含む文より前の文脈は先行文脈、「つまり」を含む文に続く文脈を後続文脈とする。

3.1 先行部の理解を促す換言

まずは、換言によって先行部の理解が深められる例を見る。

(4) は、換言を含む下線部のうち二か所が省略されているが、それぞれ「シェイクスピア」「千六百(九十四年)」であろう。シェイクスピアが生まれて30年経った時が西暦何年であったか、換言によって示されている。

(4) おれが教わった英文学の教授なんてものはさ、いちいちこまかいこと説明するのが面倒臭いからって、省略する部分を『>』『>』でやってしまう。ちょっとやってみましょうか。『シェイクスピアの大喜劇時代といわれる>は、だいたい>の三十歳の時つまり>九十四年から六年ぐらいで、機智の喜劇、たくらみの>>、ロマンチック>>などを>したが、この>の>>はやはり『真夏の夜の>』で、特に三つの>>の>>人物に>しているフォルスタッフなどは>>>>上類の

ない、性的性格である』あははははははは。笑っちゃうよね。でもこういう教授は本当にいたのです。今もやってるかもしれません。

BCCWJ サンプル ID: OB3X_00110 筒井康隆(著)『文学部唯野教授』

二重鍵括弧で括られており、シェイクスピアの話題はここで終了している。換言を含む一文だけを取り出しても理解でき、この限られた範囲においては前後の文脈と意味的な連続性を持つものとはいえない。西暦を述べている後続部は、先行部についての知識を増やし、先行部の理解を促す換言であると判断できる。

(5) もまた、先行部の理解を促す換言であると考えられる。二か所の下線部において、それぞれ「官軍」「賊軍」の辞書的な説明がなされ、その結果、先行部の知識が十分でない読み手は新たな情報を得ることとなる。この換言は、文脈から切り離しても成立する。

(5) けれど、宗教が政治と結びつくことは、いちばん怖い。国家権力によって、人々の心への強制が始まり、内心の自由を奪う。国家の意図する方向に、国民を引っ張っていくことができる。それが戦前の国家神道。侵略戦争に国民を駆り立てた恐ろしいからくりであった。その中心が、靖国神社である。／明治維新以来、官軍つまり天皇のために死んだ人のみを、「神」として祀り、賊軍つまり天皇に反抗して死んだ人は排除する。

BCCWJ サンプル ID: PB53_00218 平山知子(著)『家裁弁護士 ミモザの花言葉のように』

これらの換言は、先行部に対して後続部が情報を追加し、先行部の理解を促している。換言を含む文と前後の文脈との意味的な結びつきが重視されなければ、追加された情報は、先行部の分かりやすくするものにすぎない。

3.2 先行文脈に支えられる換言

続いて、後続部が先行文脈の内容に支えられ、先行文脈の内容を確認することとなる換言を見る。これは、必ずしも先行部を分かりやすく言い換えるとは限らない。しかし、後続部の成立が先行文脈に依っており、先行文脈があることで自然な換言となっている。後続部が先行文脈の内容を踏まえているため先行文脈の内容を確認することにもなり、先行文脈からのまとまりのある文章として理解しやすくなる。

(6) 物件価格の八十%までしか融資を受けられない、と聞いてがっかりしている人もいるだろうけど、チョット待って。残りの二十%は「頭金」で払えばいいのだ。じつはこの頭金を入れるということが、買う人、つまりローンを返済する人にとっては、たいへんラクになることなのだ。

BCCWJ サンプル ID: PB33_00046 らくらく持ち家委員会(著)『ゼットイ失敗しないマイホーム購入大満足ガイドブック』

先行部の「買う人」と後続部の「ローンを返済する人」は、先行文脈が無ければ関連付けることは難しい。先行文脈では、「物件」「融資を受けられない」「頭金」等の記述から、高額商品を現金一括払い以外の方法で購入する場合のことを説明していることは明らかである。先行文脈においては、購入時にローンが前提となっており、そのため、「買う人」が「ローンを返済する人」へと換言される。またこの後続部は、先行文脈の内容を確認することとなり、先行文脈から続く内容をまとまりのあるものとして理解しやすくなる。

(7) つまり、人間の意識を超えたところに、真実の世界がある。(略) 哲学はすべて汝自身を知るところから始まりました。ところがそこから先が東洋思想のい

いところで、「自己をならふ」ということは、「自己をわするるなり」。自分をならうというのはふつう、自分を対象化して、ああでもない、こうでもない吟味を加えること、あるいは自分の真理はどうだとかいって自分を分析することです。ところが、「自己をならふ」というのはそういうことではなく、「自己をわするるなり」です。(略) 自分をとことん考えていくためには、そんなに自分を大事にしてはだめだ。思い切って自分自身を、全部捨ててしまいなさいというのです。自分を大事に考えるという我執、執着を捨ててしまえ。自分を忘れた時つまり、自意識というものがなくなってしまった時に、「万法に証せらるるなり」です。万法というのはすべてのルール、秩序、すなわち真実ということです。宇宙の真実。自分というものを捨て去った時に、宇宙の真実と一致することになります。

BCCWJ サンプル ID: LBp1_00010 栗田勇(著)『道元いまを生きる極意』

(7) の先行文脈では、自分を考えるためには「自分自身」や「自分を大事に考えるという我執」を捨てることが求められていると説明されている。ここでいう「自分自身」「自分」とは、確かに私が存在していると感じる自己意識や自我であり、それを捨てることにより、自分の本質が明らかになる、ということである。この先行文脈の記述を踏まえて、「万法に証せらるる」のは「自意識というものがなくなってしまった時」と換言されていると考えられる。先行文脈に基づき後続部へと換言されることにより、長い先行文脈との関連が明らかになり、まとまりのある文連続として理解しやすくなる。

3.3 後続文脈に支えられる換言

最後に、後続文脈に支えられており、それ故に後続文脈も含むまとまりを理解しやすくなる換言を見る。

(8) 本当に強い人とはどんな人だろうか？誰がきても負けない人？ボクシング世界チャンピオン？K1世界チャンピオン？否。強い人とは負けない人。つまり戦わない人である。人はそれぞれなのだ、同じ土俵に上がる必要はない。つまり、妻に小言を言われたら、反論せずに寝たフリや子供をだしにその場から去る。恋人と喧嘩をする人は恋人があなたに嫌な事をした時に向き合うからだ。遅刻してきたら、「遅刻ムカつくんですけど」と言い、恋人が言い分けしてきたら、無言でその場を去ろう。1番いけないのは水掛け論。柳のようにしなると吉。

BCCWJ サンプル ID: OY01_01398 「Yahoo! ブログ」

(9) 六十歳以上のシルバーを雇用するときは、とくにこの労働時間の管理は重要です。私の存じ上げている企業の中で会計検査院から年金の不正受給だと指摘を受けたところがありました。その検査の仕方はこうでした。特定の人物の名前を数人あげて、その人間の「賃金台帳とタイムカードもしくは出勤簿を持ってこい」というのです。会計検査院は国税庁と社会保険庁とチームを組んで年金の不正受給を調べています。その調査は不正受給の可能性のある人(つまり収入の多い人)を事前にリストアップしていますから、かなりの中するようです。六十歳以上六十五歳未満の人は、厚生年金の在職老齢年金を受けることができますが、その年金は賃金の多寡に応じて減額されることになっています。しかし、勤務時間と勤務日数が少なく社会保険に加入していない場合は、その減額がありません。つまり年金を全額受給しながら賃金ももらえるわけです。問題は社会保険に該当するフルタイムで働いておりながら社会保険に加入せず、年金を満額受給しているケースです。こ

れは年金の不正受給になります。

BCCWJ サンプル ID: LB13_00069 北見昌朗(著)『パートさんに正社員以上の仕事をしてもらう本 北見式実践マニュアル』

(8) では、「負けない人」であるためにはそもそも「戦わない人」であればよいということとこのように換言されていると考えられ、これ以降の後続文脈において「戦わない人」の具体的な行為が列挙されている。もし換言されていなければ、先行部「負けない人」と後続文脈との関係が掴みにくくなり、読み手にとって疑問を抱かせる文章となる。一方、後続部「戦わない人」と後続文脈の関係は明確であり、後続文脈があるからこそ、後続部のように換言できるといえる。文章の内容もまとまりのあるものとして理解しやすくなっている。

(9) では、先行部「不正受給の可能性のある人」が、関連を掴みにくい「収入の多い人」に換言されている。「不正受給の可能性のある人」と「収入の多い人」をと直接関連付けることは難しいが、この関係は後続文脈で説明されている。在職老齢年金の受給やその金額の条件には「賃金の多寡」があり、不正受給が起こりうるのは、加入すべき社会保険に不正に入らない場合である。そして、社会保険は、「フルタイム」ならば加入しなくてはならない。この後続文脈に支えられて、先行部「不正受給の可能性のある人」は、「フルタイム」で働く「収入の多い人」へと換言されると考えられる。単なる先行部の理解に留まらず、文章レベルでの深い理解をもたらすこととなる。

なお、(7) の後続部は後続文脈の内容に支えられたものでもある。「自意識というものがなくなってしまった」は、後続文脈「自分というものを全て捨て去った」という記述があったからこそ成り立つといえる。

4. 文脈も含む検討の必要性

文脈から独立しており先行部を分かりやすくする換言、後続部が先行文脈あるいは後続文脈の内容に裏付けられ、連続するまとまりのある文章としての理解を促す換言を見てきた。

以上の換言は、「つまり」の前後だけを抜き出した場合、先行部の理解が深まるとは言いがたいものがある。以下に、下線部のみを挙げる。

- (4') 　の三十歳の時つまり　の九十四年
- (5') 　官軍つまり天皇のために死んだ人／賊軍つまり天皇に犯行して死んだ人
- (6') 　買う人、つまりローンを返済する人
- (7') 　自分を忘れた時つまり、自意識というものがなくなってしまった時
- (8') 　負けない人。つまり戦わない人
- (9') 　不正受給の可能性のある人 (つまり収入の多い人)

(4') (5') と比べて (6') ~ (8') は、「つまり」の先行部と後続部との関係が明確とはいえず、却って疑問を抱かせるおそれのある換言となっている。にもかかわらず、(6) ~ (8) のように文章レベルでの理解を促している点は、文脈も考慮した場合、換言が単なる先行部の言い換え以上の働きを負っているといっていよう。

先行部のみの理解を促す換言と、文章レベルでの理解を促す換言とでは、性質が異なるように見えるが、換言の働きそのものは先行部に関する情報の追加であることは共通している。換言した後続部が、文脈を踏まえたものであるか否かという点において異なっており、それ故、結果的に文章がより分かりやすいものとなると考えられる。

この把握の仕方は、言い換える理由には理解上の要請と文脈上の要請があるという石黒

圭（2001）の指摘に沿う。理解上の要請とは、一つの表現内容について異なる二つの表現を提示することで受け手の理解を促すために換言が成り立つというものであり、文脈上の要請とは、文章の流れを円滑にするために換言が必要とされるということである（石黒 2001）。先行部の理解を促す（4）（5）の換言は理解上の要請によるもので、文脈上の要請による換言が（6）～（9）である。（6）（7）は、先行文脈との流れを円滑にするための換言であり、（8）（9）は後続文脈との流れを円滑にするための換言である。

接続表現の先行部と後続部との関係だけに注目するのではなく、前後の文脈も考慮に入れて考察する必要がある。

5. まとめと今後の課題

換言そのものは情報の追加という働きを担っており、換言の前後で述べられている内容が後続部を支えるか否かという観点から、三つの換言を指摘できることを述べた。特に、文脈から切り離しても先行部と後続部との関係が明確である換言と、そうでない換言とでは、性質が異なるように見えるが、分析の際に文脈を考慮することで、より一般化して換言の働きを捉えることが可能になる。このことから、文脈という観点は文法研究に欠くことはできないといえる。

ここで、換言が文脈から独立していた（4）に注目したい。換言そのものの操作ではないため、文脈から切り離しても先行部を分かりやすくするものであることに変わりはないが、文脈によっては、後続文脈まで意味的な連続性のあるまとまりとして捉えることとなり、文章の理解を促すと考えられる。

（4）シェイクスピアの大喜劇時代といわれるは、だいたいの三十歳の時つまり
りは九十四年から六年ぐらいで、機智の喜劇、たくらみの、ロマンチック
などをしたが、このはやはり九十五～九十六年頃に執筆された『真夏の夜の』で、その頃世界的には、……

破線部分が、書き加えられた後続文脈である。シェイクスピアの年齢が西暦に換言されることにより作品の執筆年も自然な表現となり、「世界的には」に続くと考えられる出来事をイメージしやすくなる。（4）のように換言されていれば、文章がまとまりのあるものとして理解しやすくなるだろう。辞書的記述に基づく換言であり、後続部が文脈を前提としているものでなかったとしても、文脈との意味的な連続性が認められることもあるといえる。

必ずしも意識されるとは限らないが、読み手は換言を手掛かりに読解する。書き手の意図は、記述から推測することしかできず、換言の後続部に書き手の体験に基づく知識が反映されている可能性もある。推測にすぎない書き手の意図と、読み手の思考の展開は、換言を分析する上で区別しなくてはならない。文脈の他、必要な観点として今後の検討に生かす必要があると考える。

参考文献

- 石黒圭（2001）「換言を表す接続語について——『すなわち』『つまり』『要するに』を中心に——」『日本語教育 110号』pp.32-41
 国立国語研究所（2006）「現代雑誌 200 万字言語調査語彙表 公開版 ver.1.0」
 日本語記述文法研究会（2009）『現代日本語文法 7』くろしお出版

関連 URL

コーパス検索アプリケーション『中納言』<https://chunagon.ninjal.ac.jp/>

自発対話音声に対する叫び声アノテーション

白鳥 恵大*, 大久保 港, 松田 匠翔, 有本 泰子 (千葉工業大学情報科学部)

Scream and shout annotation for spontaneous dialog speech

Keita Shiratori, Minato Okubo, Takuto Matsuda, Yoshiko Arimoto

(Faculty of Information and Computer Science, Chiba Institute of Technology)

要旨

叫び声は突発的な感情表現を示す音声現象の一つである。先行研究では、自発対話音声に含まれる叫び声を感情表出系感動詞と区別して定義していた。しかし、先行研究の定義を基に叫び声と感情表出系感動詞のアノテーションを行っても、言語表現が似ている音声現象を音響的特性のみで区別する必要があるため、この二つの音声現象を区別することは困難であった。そこで、叫び声と感情表出系感動詞を区別するために改めて叫び声 (scream) の定義を行った。また、発話の特徴と叫び声の特徴を併せ持った音声を発話と叫びの共起 (shout) として区別した。これらの定義を基に自発対話音声に含まれる音声を収録した音声資料に対して叫び声アノテーションを行った。複数人でアノテーションした際の一致率算出を行って新たな定義と先行研究の定義との比較を行う。さらに、叫び声の事例をいくつか示し、自発的な叫び声がどのような音声言語現象として発せられているかについて考察する。

1. はじめに

叫び声は突発的な感情表出を示す音声現象であり、なにかに襲われたときに危機的状况を表したり、スポーツ観戦などで場を盛り上げたりと様々な場面で発声される。叫び声に関する研究はまだ少ないものの、自発的な叫び声を収集する基礎的な研究 (Mori and Kikuchi 2020) から、叫び声検出 (Laffitte et al. 2016) や叫び声合成 (土井敦也・有本泰子 2022) といった工学的応用研究まである。例えば、叫び声を検出し異常事態の発声を検知する監視システムの構築を行った研究がある (Laffitte et al. 2016)。監視カメラの死角からでも叫び声を検出することで映像では確認できない異常事態を知ることができる。他にも、叫び声を音声合成する研究 (土井敦也・有本泰子 2022) が実現可能となれば、スポーツ観戦時に機械が叫び声を発することで場を盛り上げるなど、機械と人間とのコミュニケーションをより豊かなものにすることができる。このように、叫び声の工学的応用研究は様々な場面で我々の社会生活で役に立つ。しかし、これらの研究を達成するためには、研究の資料となる自発的に発声された叫び声の音声資源が必要となる。叫び声は日常生活の中で頻出する音声現象ではなく、その資料を効率的に収集するのは非常に困難である。

* shiratori@mac-lab.org

Mori and Kikuchi (2020) は、この課題に果敢に挑戦し、叫び声を多く含んだ自発対話コーパスを作成した。このコーパスでは、叫び声は非言語音の中でも急激な感情表出を示す社会的シグナルの一つであり、感情表出系感動詞の中でも、様式化の程度が低く、話者の制御下にある度合いが低い（思わず発せられた）ものとしていた。しかし、Mori and Kikuchi (2020) の叫び声の定義では、音声の音響的特性のみで区別する必要がある、同じような言語情報を持つ感情表出系感動詞と叫び声を区別することは困難であった。また、叫び声には「うわー」や「きゃー」といった言語的な意味を伴わない叫び声と「助けて」や「やったー」といった言語的内容を伴った発話と叫びが共起した音声が存在しているが、この二つを明確に区別して定義していない。

本研究では、音声対話中に出現する叫び声を改めて定義する。提案する定義と Mori and Kikuchi (2020) での定義とでアノテーションを実施し、一致率の比較を行った。また、提案した定義を基に、自発対話音声中に出現する叫び声はどんな音声かを調査するため、叫び声の分節音の分析、発話と叫びが共起した際の言語分析を行った。これらの分析によって、自発的な叫び声がどのような音声言語現象として発せられているかについて考察する。

2. 音声資料

本研究では、アクションゲーム音声コミュニケーションコーパス (AGSC) (Mori and Kikuchi 2020) を使用した。AGSC は 2 人 1 組でゲームをプレイしている時の音声を収録している。会話中の自発的な叫び声を得るため、アクション性の高い 2 種のゲーム (First-person shooter(FPS) ゲームである OverWatch (Blizzard Entertainment, Inc.), サッカーゲームである FIFA16(Electronic Arts Inc.)) を使用し収録された。コーパスには、24 名 (男性 12 名, 女性 12 名) の音声量子化ビット数 16bit, サンプリング周波数 48kHz で収録されている。一人当たり平均 60.7 分, 全体で 728.4 分の音声コーパスである。AGSC は praat によって、収録した音声に対して発話番号や発話内容、先行研究の定義による叫び声・感情表出系感動詞がアノテーションされている。言語音・非言語音が 400ms 以上の途切れがなく連続している区間を発話のセグメントとし、発話のセグメンテーションと発話の書き起こしは大学生 6 名によって行われ、最終的に著者によって修正されている。

この音声資料の 24 名分全ての音声を使用して、叫び声のアノテーションを行った。

3. 叫び声アノテーション

3.1 叫び声の定義

叫び声の定義は、Mori and Kikuchi (2020) の定義 (以下、先行研究の定義) を参考に改めて定義した。感動詞は、フィラー、感情表出系感動詞、応答表現、挨拶表現、呼び掛け・掛け声の 5 つに分類される。そのうち、感情表出系感動詞は驚いた時や落胆したときなどに発する表現である。Mori and Kikuchi (2020) は、叫び声を次のように定義していた。

- 叫び声は非言語音の中でも急激な感情表出を示す社会的シグナルの一つであり、感情表出系感動詞の中でも、言語的な様式化の程度が低く、話者の制御下にある度合いが低い

(思わず発せられた)もの

叫び声は急激な感情表出を示す社会的シグナルであり危機的状況下などの突発的なイベントに対して発声される。そのため、文脈や状況を考慮してアノテーションを行う必要があるが、先行研究の定義では考慮されていない。そこで、

- 叫び声は予想外の出来事によって話者が無意識に発したもので、韻律あるいは声質が特異なもの

とし、叫びのうち聞き手が単語として意味を理解できないもの (scream) と定義した。

また、叫びには様式化の程度が低いものに加えて、「助けて」や「やったー」といった発話と叫びの共起によって言語内容を伴った叫び声が存在するが、先行研究の定義ではこの現象を説明できていない。笑いの研究では、発話と笑いの共起である音声 (speech laugh) を発話や発話を伴わない真の笑い声と区別して定義されている (Nwokah et al. 1999)。そこで、叫びのうち聞き手が単語として意味を理解できるもの (shout) を定義に加え、単語としての理解の可否で scream と区別した。

3.2 アノテーション

アノテーションは praat を用いて話者の音声のみを使用して行った。AGSC の TextGrid には3つの層があり、そのうち utterance 層は発話番号、word 層は発話内容、affectburst 層は叫び声 (scream, shout)、感情表出系感動詞が記述されている。word 層は、発話内容以外にも、叫び声には {shout} というラベルが付与されている。

この affectburst 層に追加・変更し、提案した定義による叫び声および感情表出系感動詞を記述した。提案した叫び声の定義をもとに、scream と shout および感情表出系感動詞を区別するためのラベルを定義し、アノテーションに使用した。提案した定義でのアノテーションで使用したラベルとその定義を以下に示す。

- {s}：予想外の出来事によって話者が無意識に発したもので、韻律あるいは声質が特異なものうち、聞き手が単語として理解できないもの (scream)
- {shout}：予想外の出来事によって話者が無意識に発したもので、韻律あるいは声質が特異なものうち、聞き手が単語として理解できるもの (shout)
- {a}：感動詞のうち感情表出系感動詞 (scream と shout は含まない)
- {o xxx}：{s} や {shout}, {a} のどれかにあたるが、音量が小さいなどの定義に合わないもの (xxx に定義に合わない理由を記述)

提案した叫び声の定義の妥当性を検証するため、まずは G011.L(男性話者, 58分)のデータのみを使用してアノテーションを実施した。先行研究の定義でのアノテーションで使用したラベルとその定義を以下に示す

- {s}：感情表出系感動詞の中でも、言語的な様式化の程度が低く、話者の制御下にある度合いが低いもの
- {a}：感動詞のうち感情表出系感動詞

大学生2名に、先行研究での {a} と {s} ラベル、提案した定義での {a} と {s} ラベル、それぞれに対してアノテーションを行わせ、アノテータ間的一致率を求めて比較した。提案した定

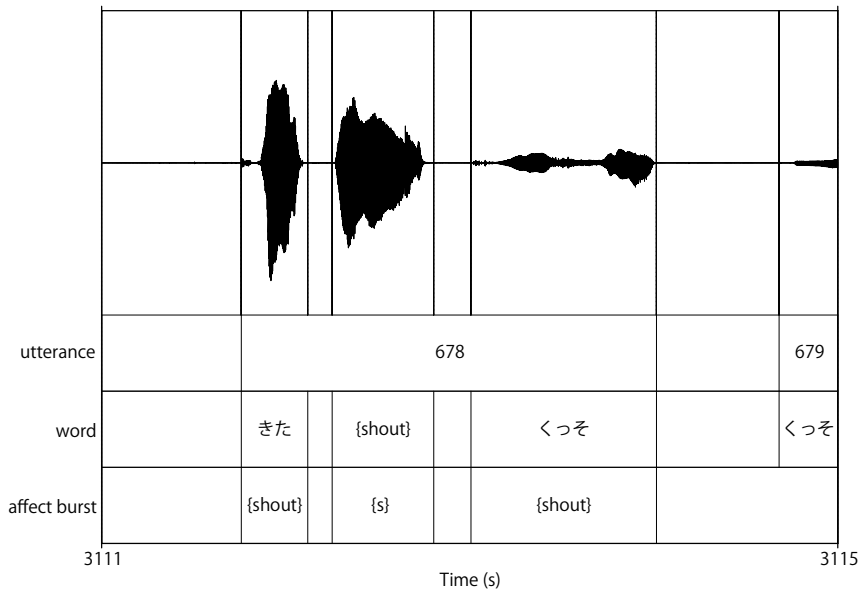


図1 praat アノテーション画面の例

義により一致率が向上したことを確認したのち、叫び声の研究に従事している大学生2名および大学院生1名の合計3名によってAGSCのすべてのデータに対してアノテーションを実施した。一人当たり8名分のデータに対してアノテーションさせている。判断が難しい音声に対しては、アノテーションを行った3名でディスカッションを行い、最も適切だと思われるラベルへ振り分けた。

AGSCのpraatのアノテーション画面の例を図1に示す。1段目は音声波形、2段目はTextGridによる書き起こしの内容を示している。

3.3 結果

提案した叫び声の定義の妥当性を検証するため、評価の指標として、正解率とF値を算出した。正解率とF値を求める際は、ラベルが付与された各区間の開始時刻と終了時刻の差が0.2s以内であれば一致しているとした。先行研究の定義の{s}と{a}ラベル、提案した定義の{s}と{a}ラベルのアノテーションの一致率を表1に示す。先行研究の定義の{a}ラベルの正解率は30%、{s}ラベルの正解率は29%、提案した定義の{a}ラベルの正解率は55%、{s}ラベルの正解率は58%であった。先行研究の定義の{a}ラベルのF値は46%、{s}ラベルのF値は45%、提案した定義の{a}ラベルのF値は71%、{s}ラベルのF値は74%であった。

AGSCの全データに対してアノテーションをした結果として、各ラベルが付与された区間の数を話者ごとに表2に示す。AGSCには、男性話者によるscreamが534個、shoutが278個、女性話者によるscreamが856個、shoutが364個、合計してscreamが1390個、shoutが642個含まれていた。

表 1 先行研究の定義と提案した定義のアノテーション一致率

	先行研究		提案	
	{a}	{s}	{a}	{s}
正解率	0.30	0.29	0.55	0.58
F 値	0.46	0.45	0.71	0.74

3.4 考察

先行研究の定義より、提案した定義による {a} ラベルの正解率は 25%、F 値は 25% 高く、{s} ラベルの正解率は 29%、F 値は 29% 高かった。提案した定義は先行研究の定義よりも安定したアノテーションが可能であることがわかった。先行研究の定義と提案した定義の異なる点は文脈や状況を考慮したことである。このことより、自発対話音声における叫び声は、突発的なイベントに対するリアクションであるという視点を追加することで他の音声との区別がより可能になると考えられる。また、先行研究の {s} で 2 名のアノテータが一致せず、かつ提案手法の {s} で一致した音声は、話者が無意識に発声したかどうか曖昧な音声が多かった。先行研究では対象となる叫び声の前後の文脈は考慮せず、その叫び声の音声情報のみでアノテーションを行っているため、曖昧な音声は判断がつきにくい。提案した定義では音声の直前の会話内容から予想外の出来事の有無を確認し、ある程度無意識に発声したかどうか判断できるため、アノテータ間でより一致した判断が得られた。

男性話者よりも女性話者の scream の数は 322 個多く、shout の数は 86 個多かった。今回の参加者の一月当たりの平均ゲーム時間は、男性話者が 92.5 (SD=60.47)、女性話者が 29.17 (SD=42.15) であり、ゲーム経験の差が叫び声の表出数に差を生み出した可能性がある。つまり、今回の参加者は、ゲームの内容になじみ深く、ある程度どんなことが起こるか想像することができるため、あまり叫び声をあげることがない参加者が男性に多く、また、ゲームの内容をあまり知らず、ゲームの展開を予想できなかったために、予想外の出来事が発生したときに、叫び声を上げた参加者が女性に多かった可能性がある。また、同じ男性話者でも G009.R は scream が 5 個、shout が 0 個に対して G011.L の話者は scream が 133 個、shout が 55 個と大きな差が見られた。そのため、叫び声の発声しやすさは性差ではなく個人差があると考えられる。

4. 叫び声の言語現象の分析

4.1 分析の目的

自発的な叫び声がどのような音声言語現象として発せられているかについて調査した。叫び声がどのような言語音や単語で発せられるかあるいは発せられやすいかを、コーパスを用いて研究された例はまだ見ず、それらを明らかにして、工学的な応用研究に役立てることが求められる。例えば、叫び声を合成する際に、どのような音韻で叫び声を構成すれば最も叫び声らしくなるかの判断の一助となる。まずは、scream がどのような音素で発声されているかを明ら

表2 ラベルが付与された音声区間の数

話者	性別	{a}[個]	{s}[個]	{shout}[個]
G001.L	男	137	17	7
G001.R	男	106	34	15
G003.L	女	239	89	24
G003.R	女	254	70	43
G004.L	女	275	61	25
G004.R	女	410	66	72
G005.L	女	145	10	3
G005.R	女	80	35	9
G006.L	女	208	113	22
G006.R	女	123	131	37
G007.L	女	167	52	34
G007.R	女	55	55	22
G008.L	女	80	11	9
G008.R	女	364	135	64
G009.L	男	51	16	2
G009.R	男	35	5	0
G010.L	男	145	35	23
G010.R	男	209	50	19
G011.L	男	283	133	55
G011.R	男	58	62	27
G012.L	男	114	94	55
G012.R	男	79	35	26
G013.L	男	156	21	20
G013.R	男	156	21	20
男性話者	男	1627	534	278
女性話者	女	2400	856	364
合計		4027	1390	642

かにする。また、発話と叫びが共起する shout では、「助けて」など意味のある単語で発せられるのか、「が」や「を」などの助詞のような意味のない単語で発せられるのかを明らかにする。さらに、shout では、話をしている途中で叫んで shout となるのか、叫んだ (scream) 後に shout となるのか、それとも発話と叫びを同時に発声するのかを明らかにする。

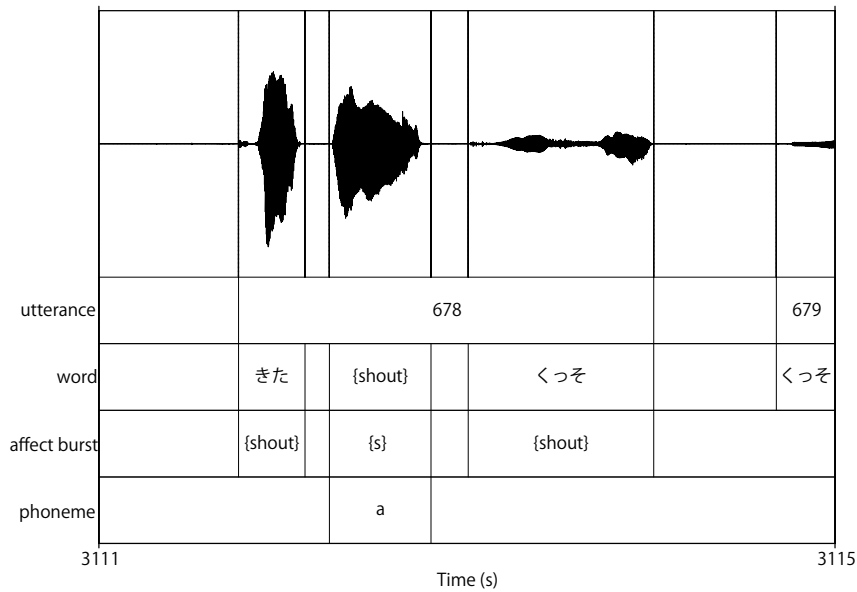


図2 分節音アノテーションの例

4.2 分析方法

分析に使用したデータは、男女それぞれで叫び声の数が多かった G006.R, G008.R, G011.L, G012.L の4名のデータを使用した。使用したデータには、{s}の音声は493個、{shout}の音声は211個含まれていた。

scream がどのような音素で発声されているかを明らかにするため、アノテーションしたscream に対して分節音アノテーションを行い、各分節音の出現率を求めた。分節音アノテーションの例を図2に示す。

phoneme 層を新たに追加し、affectburst 層に {s} のラベルが付与されている区間に分節音ラベルを記述した。分節音を判断する際に、いずれかの母音に該当するがどの母音にするか判断できない音声については X ラベルを付与した。各分節音の出現率の統計的な差を確認するために、分節音ラベルを母音と子音に分類し、帰無仮説を「すべての母音の発生率に差はない」、「すべての子音の発生率に差はない」としてカイ二乗検定を行った。

shout を発するとき、意味のある語で叫んでいるか意味のない語で叫んでいるかを調べるため、shout の冒頭が内容語であった数と機能語であった数を集計した。アノテーションを行った {shout} ラベルの部分の発話内容について MeCab による形態素解析を行い、出力された品詞を内容語と機能語に分類した。帰無仮説を「内容語と機能語の数に差はない」として二項検定を行った。

さらに、shout の直前の音声を集計し、「これさー、やばいー」のように発話から shout が発声されるか、「わー、やばいー」のように scream から shout が発声されるか、「やばいー」のように発話と叫びが同時に開始されて (shout の直前には無音区間がある。あるいは shout の直前に shout がある) shout が発声されるかを調査した。アノテーションを行った {shout} ラベルの部分の直前 0.2s の区間を調査し、その区間の音声が発話か scream かそれ以外かで、

発話から開始した shout か、scream から開始した shout か、発話と scream が同時に発声した shout かを分類した。発話から shout が発声されるか、scream から shout が発声されるか、発話と scream が同時に開始し shout が発声されるかの帰無仮説を「shout の直前の音声に差はない」としてカイ二乗検定を行い、調査した。

4.3 結果

scream の分節音の集計結果を、母音と子音に分けてそれぞれ図 3 と図 4 に示す。図 4 の子音のグラフは、2 個以上出現した分節音のみを掲載した。母音では/a/が 392 個と最も多く、/i/は 43 個、/u/は 69 個、/e/は 76 個、/o/は 79 個、/X/は 18 個であった。子音では、/h/で 46 個、/w/で 41 個、/y/で 16 個、/r/で 4 個、/d/で 2 個、/n/で 2 個であった。母音の出現率のカイ二乗検定の結果、 $\chi^2(5) = 852.82, p < 0.001$ となり、帰無仮説が棄却され「すべての母音の出現率に差がある」ことが分かった。子音の出現率のカイ二乗検定の結果、 $\chi^2(6) = 117.22, p < 0.001$ となり、帰無仮説が棄却され「すべての子音の出現率に差がある」ことが分かった。

shout の冒頭における内容語・機能語の集計結果を図 5 に示す。内容語であった数は、160 個、機能語であった数は 51 個であった。内容語と機能語の数に対する二項検定の結果、 $p < 0.001$ となり、帰無仮説が棄却され、内容語が機能語に対して有意に多いことが分かった。

shout の直前の音声の集計の結果を図 6 に示す。発話から開始されたものは 18 個、scream から開始されたものは 69 個、発話と叫びが同時に開始されたものは 124 個であった。shout の直前の音声に対するカイ二乗検定の結果、 $\chi^2(2) = 79.92, p < 0.001$ となり帰無仮説が棄却

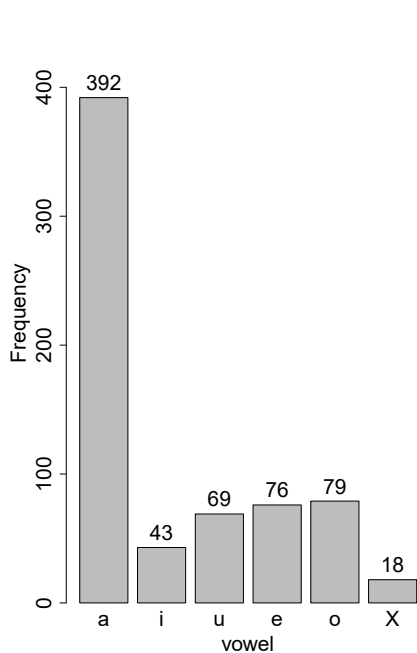


図 3 母音の集計結果

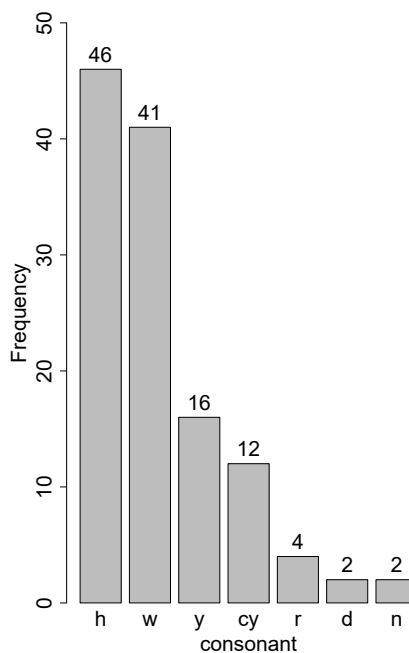


図 4 子音の集計結果

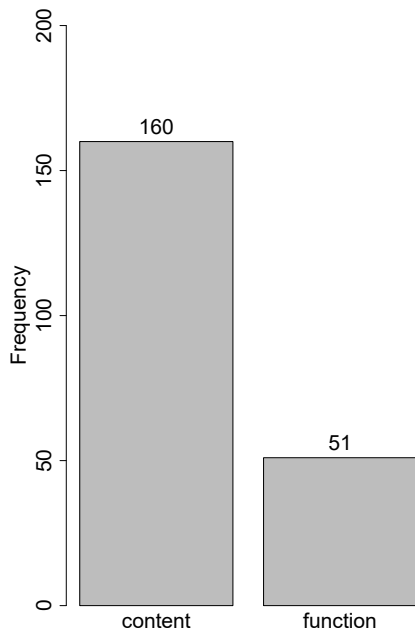


図5 shout が内容語か機能語かの集計結果

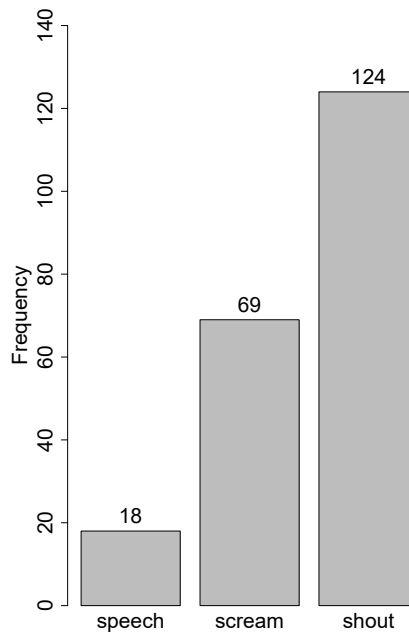


図6 shout 直前の音声の集計結果

され「shout の直前の音声に差がある」ことが分かった。

4.4 考察

scream の音声の分節音の分析によって、叫び声は母音では/a/, 子音では/h/や/w/で発声されやすいことが分かった。なかでも、分析対象となる 493 個の scream のうち母音/a/は 392 個と圧倒的に多く、ほとんどの scream が/a/で開始されていることが分かる。我々が思い浮かべる scream の典型的な音韻列は「きゃー」であるが、図4に示した子音の集計結果からも明らかなように/ky/で発声される scream はほとんどない。自発対話音声における scream には「きゃー」といったような叫び声はほとんど存在せず、単純に「あー」といった scream が多く発声されることが示唆される。

shout の内容語・機能語の分析では、言語内容を伴った叫び声は内容語で発声されやすいことが分かった。shout の音声は、「行け」や「取れ」といったゲーム内の自分が操作しているアバターに対して指示を出す際に思わず発することが多かった。このことより、応援や鼓舞をする強い気持ちが叫び声となって表出されると考えられる。

shout の直前の音声の分析では、発話から変化するのではなく叫びから変化するのではなく、発話と叫びの現象が同時に生じやすいことが分かった。shout の音声は、「行け、行け、行け」のように短い shout の区間を何度も繰り返すことが多かった。このことより、同じ言葉を繰り返し叫ぶことで伝えたい内容を強調していると考えられる。また、予想外の出来事によって scream が発声され、その後 shout が発声されることが多かった。scream によって予想外の出来事に対するリアクションをし、shout によってその状況を踏まえて伝えたい内容を発声

している可能性がある。

5. おわりに

叫び声検出や叫び声合成などの工学的な応用研究に対話中に自然に表出した叫び声を利用することを目的とし、自発対話音声における叫び声の定義を提案した。また、叫び声がどんな言語表現で表出するか調査するため、基本的な言語分析を行った。その結果、提案した叫び声の定義は Mori and Kikuchi (2020) の定義よりもアノテータ間の一致率が高く、比較的安定して叫び声ラベルを付与できることが分かった。言語内容を伴っていない叫び声である scream は母音 a で発声されることが多いことが分かった。言語内容を伴った叫び声である shout は、内容語で、発話と叫びの現象が同時に引き起こされやすいことが分かった。叫び声がいつ、どのように発声されるかを明らかにすることで、機械が対話の中で自然に叫び声を発声し場を盛り上げるなど、機械と人間とのコミュニケーションを円滑にできる。今回はゲーム場面での自発音声を扱ったが、今後は他の自発対話音声について叫び声の分析をし、今回の結果との差異を比較していきたい。

謝 辞

AGSC を快くご提供くださった宇都宮大学の森大毅准教授に感謝する。本研究の一部は公益財団法人中山隼雄科学技術文化財団設立 30 周年研究助成、および JSPS 科研費 JP22K18477, JP22K12107 の助成を受けた。

文 献

- Hiroki Mori, and Yuki Kikuchi (2020). “Gaming Corpus for Studying Social Screams.” *Proc. Interspeech 2020*, pp. 3132–3135.
- Pierre Laffitte, David Sodoyer, Charles Tatkeu, and Laurent Girin (2016). “Deep neural networks for automatic detection of screams and shouted speech in subway trains.” *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6460–6464.
- 土井敦也・有本泰子 (2022). 「WaveNet による叫び声合成の実現に向けたコンテキストラベルの検討」 日本音響学会 2022 年春季研究発表会講演論文集, pp. 951–952.
- Eva Nwokah, Hui-Chin Hsu, P Davies, and Alan Fogel (1999). “The integration of laughter and speech in vocal communication: A dynamic systems perspective.” *Journal of speech, language, and hearing research : JSLHR*, 42, pp. 880–94.

学術論文と論説文における「だから」の使用の比較

向坂卓也 (中国・外交学院) †

Comparison of the Usage of "Dakara" in Academic Papers and Editorial Texts

Takuya Mukozaka (China Foreign Affairs University)

要旨

日本語文章表現の教材には「だから」は話し言葉であり、書き言葉文では使用されないとしているものと、書き言葉であり、書き言葉文では使用されるとしているものがある。一口に「硬い文章」といってもジャンルや分野によって特徴があると考えられる。そこで本研究では J-stage の 15 学会誌の学術論文と webronza の 5 分野の論説文における「だから」を含む順接接続詞の使用状況を調査した。理系学術論文では「だから」の使用が稀だが、文系学術論文では分野によっては使用され、論説文では各分野で使用されていることがわかった。「だから」は主観的な結論付けを表す接続詞であり、著者の主観的結論付けが行われる場合に「だから」が使用される。「だから」が使用されないジャンルや分野では、話し言葉であるため使用されないのではなく、結論付けの際に客観性が要求されるため使用されないのである。

1. はじめに

日本語文章表現の教材には「だから」は話し言葉であり、書き言葉文では使用されないとしているものと、書き言葉であり、書き言葉文では使用されるとしているものがある。教材によってなぜこのような違いが生じるのだろうか。これらの教材はいわゆる「硬い文章」を書くことを学習目標としている。しかし、一口に「硬い文章」といってもジャンルや分野は様々であり、それぞれに特徴があるのではないだろうか。そこで本研究では「硬い文章」とされる学術論文と論説文の各分野における「だから」を含む順接接続詞の使用状況を調査する。

学術論文については国立研究開発法人科学技術振興機構が運営する学術論文の電子公開システムである J-stage に収録されている 15 学会誌各 50 本 (全て著者が異なる) の学術論文、論説文については朝日新聞社が運営する言論サイト webronza に収録されている 5 分野各 50 本 (全て著者が異なる) の論説文を調査対象とする。

2. 日本語文章表現教材における「だから」の取り扱い

日本語文章表現教材における「だから」の取り扱いについて調査した。教材が対象とする学習者、目標とする文章¹、「だから」を話し言葉としているか、書き言葉としているかについて表 1 にまとめた。

①～⑭が「だから」を話し言葉、⑮～⑱が「だから」を書き言葉としている教材である。論文・レポートを書くことを目標としている教材は、いずれも「だから」を話し言葉としている (①②, ④～⑫)。小論文を書くことを目標とした教材は「だから」を話し言葉としているもの (③, ⑫～⑭) と書き言葉としているもの (⑮～⑱) がある。「だから」を書き言葉であるとしている教材は、いずれも高校生向けに小論文を書くことを目標としている (表 1)。

† xiangban_zhuoye[at]yahoo.co.jp

¹ 教材が目標とする文章を「論文・レポート」と「小論文」に 2 つに分類した。「論文・レポート」として一まとめに目標とする教材が多いことから、「論文・レポート」として分類することにした。

表1 日本語文章表現教材における「だから」の取り扱い

「だから」を話し言葉としている教材								
	教材名	話し言葉	書き言葉	対象			目標	
				留学生	大学生	高校生	論文・レポート	小論文
①	『大学生と留学生のための論文ワークブック』（1997）	だから／それで	ゆえに／それゆえ／したがって	○	○		○	
②	『改訂版 留学生のための論理的な文章の書き方』（2003）	だから	したがって	○			○	
③	『小論文への12ステップ』（2008）	だから／ですから	そのため／したがって	○				○
④	『留学生のためのここが大切文章表現のルール』（2009）	だから	したがって	○			○	
⑤	『留学生と日本人学生のためのレポート・論文表現ハンドブック』（2009）	だから	したがって	○	○		○	
⑥	『論文・レポートの基本』（2012）	だから	そのため		○		○	
⑦	『改訂版 大学・大学院 留学生の日本語④論文作成編』（2015）	だから／ですから	そのため／そこで／したがって	○			○	
⑧	『学生のレポート・論文作成トレーニング - 改訂版 -』（2015）	だから／なので	そのため／それゆえ／それで		○		○	
⑨	『レポート・論文を書くための日本語文法』（2016）	だから／それで	そのため／したがって	○			○	
⑩	『21世紀のカレッジ・ジャパニーズ』（2018）	だから／それで	そのため／したがって	○	○		○	
⑪	『思考を鍛えるレポート・論文作成法第3版』（2019）	だから	したがって		○		○	
⑫	『日本語を学ぶ人のためのアカデミック・ライティング講座』（2020）	ですから／だから／なので／そして／それで／で	そのため／したがって／そこで／その結果／よって／ゆえに	○			○	○
⑬	『改訂版 何を書けばわからない人のための小論文のオキテ』（2020）	だから／なので	そこで／したがって			○		○
⑭	『7日間で合格する小論文』（2021）	なので／だから	したがって／そのため			○		○
「だから」を書き言葉としている教材								
	教材名	話し言葉	書き言葉	対象			目標	
				留学生	大学生	高校生	論文・レポート	小論文
⑮	『大学入試小論文の完全攻略本』（2010）	なので	だから／よって／ゆえに／したがって			○		○
⑯	『小論文 これだけ！書き方超基礎編』（2013）	なので	だから			○		○
⑰	『小論文をひとつひとつわかりやすく』（2016）	なので	だから／したがって			○		○
⑱	大修館書店『国語表現 改訂版（平成30年）』（2018）	なので	だから／したがって			○		○

3. 「だから」に関する先行研究

前章では日本語文章表現教材には「だから」を話し言葉としているものと、書き言葉としているものがあることについて述べた。「だから」を話し言葉としている教材では書き言葉文では「だから」ではなく「したがって」「そのため」「それゆえ」「よって」などの書き言葉を使うように指示されている。

では、「だから」と「したがって」「そのため」「それゆえ」「よって」「ゆえに」にはどの

ような違いがあるのだろうか。

そこでまず、第1節で「だから」と書き言葉において「だから」に換えて使用されるべきであるとされる「したがって」「そのため」「それゆえ」「よって」「ゆえに」などとの違いについて言及した先行研究を見ていくことにする。

また、論文・レポートを書くことを目標としている教材は、いずれも「だから」を話し言葉としているが、小論文を書くことを目標としている教材は、「だから」を話し言葉としているものもあれば、書き言葉としているものがある。そこで第2節でジャンルの違いによって「だから」の使用状況がどのように異なるのかについて論じた先行研究を見ていくことにする。

3.1 「だから」とその他の順接接続詞の意味の違いに関する先行研究

ひげ (1987) は「だから」を「はなし手の意志や主観による論理のたて方」とするいっぽうで、「したがって」を「はなし手 (かき手) はある (いくつかの) 事実を根拠として、そこから客観的にみちびきだされる結論についてのべる」ものであるとする (pp.55)。

石黒 (2016) は「だから」「ですから」「だからこそ」を書き手の判断で主観的にむすびつける「主観の接続詞」(p.26) とし、後件の文末に指示や要求を表す表現が使われ、客観性を重んじる論文などではあまり使われないことを指摘している。「したがって」「よって」「それゆえ」を主観を交えず、論理的必然を表す「必然の接続詞」(p.28) とし、論理性を重視する学問によく使われる表現であると述べている。「そのため」を「原因の接続詞」(p.30) とし、前件が後件の原因であることを示すのに適しているものであるとしている。

沖森 (2016) は「だから」について、後件に話し手の意志や主観的な判断が示されることがあるとしている (p.89)。一方、「そのため」は前に述べた事柄から引き出される客観的な事柄を次に導くものであり (p.92)、「したがって」も前件から導き出される客観的な事柄を次に導くものであり (p.97)、いずれも文末に話し手の主観的な表現は一般的に用いられず、また書きことばで用いられる表現であることをしている。さらに「ゆえに」の系列 («それゆえ」「それゆえに」「そのゆえ」「そのゆえに»), 「よって」の系列 («それによって」「それにより») を前述の内容を受け、必然的にある結論に帰着するというように論理的に次の事柄を導くものであるとしている (p.98)。

以上の先行研究においては「だから」は主観的な結論付けを表す接続詞であり、「したがって」「そのため」「それゆえ」「ゆえに」「よって」などが客観的な結論付けを表す接続詞であるという違いについて言及されている。

3.2 ジャンルの違いによる「だから」の使用状況に関する先行研究

劉 (2006) は CASTLE/L (日本語教育支援システム研究会) が開発した教材テキストデータベースに収録されている論説文における「だから」の使用回数を調査した結果、「だから」が 103 回、「したがって」が 133 回であり、「だから」は「したがって」よりも使用回数が少ないものの、学術論文とは異なり、著者の主観が許される論説文では「だから」が用いることができると述べた。話し手が「だから」を用いる時は「自分の中に働いている因果関係の知識が聞き手の中に成立していない (または成立しない) と仮定する場合」だが、客観性が求められる学術論文では「論証に必要とされる因果関係の知識が読み手の中にも成立すると仮定している」ため、学術論文では「だから」は用いられないことを指摘している。ただし、劉 (2006) は学術論文における「だから」の使用回数については調査していない (p.133)。

向坂 (2019) は紀要論文・博士論文における「だから」の使用状況 (使用の出現数と使用者数) を調査した。そして「理系分野の学術論文では「だから」の使用が見られなかったが、文系分野の学術論文では一定数の使用が見られた。数値などの客観的な証拠に基づいて研究する理系分野とは異なり、文系分野では客観的な証拠に基づいた研究しているとは限らず、著者の主観性が入り込む余地がある。そのことが文系分野の学術論文において「だから」を出現させていると考えられる」と論じている (pp.186-187)。

向坂（2019）は学術論文の使用状況を調査したが、調査対象が文系分野は文学（63本）、社会学（43本）、言語学（57本）、経済学（46本）の4分野209本の論文であったが、一方、理系分野は「理工学」に一本化しており、しかも19本と限定的であった。そこで本研究ではより幅広い分野におけるより多くの学術論文を調査対象とし、さらに論説文も調査対象とすることで学術論文と論説文における「だから」を含めた順接接続詞の使用状況を比較していくことにする。

4. 学術論文と論説文における順接接続詞の使用状況の調査

4.1 学術論文と論説文の違い

本研究では学術論文と論説文における「だから」を含む順接接続詞の使用状況を調査するが、その前に学術論文と論説文の違いについて述べる。

国語教育研究所（1991）は「広くは事柄を論ずる文章」としての「論文」を三つに分類している。「(1)ある事について自分の意見を述べ、論議する文章。広い意味での議論文で、政治、外交、経済、社会などの一般的な問題を論じるのが、「論説文」「批評文」や「調査報告」であり、「(2)専門的な問題について研究目的、経過、結果をまとめ、新しい見解を提示する文章」が「学術論文」や大学の「卒業論文」などであり、「(3)入社試験、大学入試に、特定の項目、作成条件を付して出される課題について論ずる文章」が「小論文」や「課題論文」である（p.872）。

国語教育研究所（1991）の定義を参考にすれば、学術論文は(2)、論説文は(1)に該当する。学術論文と論説文の違いは「専門的な問題」についてか「一般的な問題」についてか、そして「新しい見解を示す文章」か「自分の意見を述べる文章」かであると考えられる。

4.2 学術論文と論説文における順接接続詞の使用状況の調査方法

本研究では以下の方法で学術論文と論説文における「だから」を含む順接接続詞の使用状況を調査する。

- ①J-stage（国立研究開発法人科学技術振興機構が運営する学術論文の電子公開システム）に収録されている15学会誌の著者が異なる各50本の学術論文、論説文はwebronza（朝日新聞社の言論サイト）に収録されている5分野の著者が異なる各50本の論説文を対象とする。
- ②著者が特定できる学術論文及び論説文を対象とし、共著などの著者が特定できないものは除外する。
- ③「だから」²の他にも順接接続詞である「それで」³「なので」「よって」「ゆえに」「それゆえ」「そのゆえ」「その結果」「そのため」「このため」⁴「したがって」の使用件数を調査する。また、各分野における使用者数（一人の著者が一度でも接続詞を使えば使用者として計数する）を調査する。
- ④J-Stageに収録されている学術論文はpdfファイルをダウンロードし、pdfの検索機能を使い、順接接続詞を検索する。pdfファイルをword化し、wordの文字カウント機能により文字数を計数する。webronzaの論説文はwordに貼り付け、wordの検索機能を使い、順接接続詞を検索する。wordの文字カウント機能によりの文字数を計数する。
- ⑤敬体が使用されている学術論文や論説文（講演記録、インタビュー形式等）は除外する。
- ⑥学術論文や論説文における直接引用の部分（かぎかっこ引用、ブロック引用等）は調査

2 「だから」は「だからこそ」「だからか」も対象とするが、逆接的な用法である「だからといって」「だからというわけではないが」は対象外とする。向坂（2021）ではJ-Stageに収録されている学会誌の学術論文における「ことから」「から」「だから」の使用件数を調査したが、このときは各40本であった。またこのときは「だからといって」「だからというわけではないが」も対象としていた。

3 「それで」は順接の用法を対象とし、指示詞「それ」+助詞「で」は対象外とした。

4 「そのため」「このため」は順接の用法を対象とし、目的の用法は対象外とした。

対象から除外する。

⑦学術論文は学会誌ごと、論説文は分野ごとの順接接続詞の使用件数を計数し、さらに10万字あたりの使用件数を算出する。

⑧学術論文は学会誌ごと50著者、論説文は分野ごと50著者における順接接続詞の使用件数を計数し、使用者の比率を算出する。

4.3 学術論文における順接接続詞の使用状況の調査結果

J-stage に収録されている15学会誌の学術論文における順接接続詞の使用件数を調査した結果が表2である(上段が使用件数, 下段が10万字当たりの使用件数)。各学会誌における順接接続詞の使用件数とその比率をまとめたのが表3である(上段が使用者数, 下段が使用者比率)。

表2 J-stage の学術論文における順接接続詞の使用件数(上段: 使用件数 下段: 10万字当たりの使用件数)

	分野	字数	だから	それで	なので	よって	ゆえに	それゆえ	そのゆえ	その結果	そのため	このため	したがって	
文系	『日本文学』	日本文学	729,399	25 3.4	1 0.1	0 0.0	2 0.3	1 0.1	9 1.2	0 0.0	9 1.2	8 1.1	1 0.1	28 3.8
	『日本語の研究』	日本語学	865,913	1 0.1	0 0.0	0 0.0	34 3.9	1 0.1	4 0.5	0 0.0	36 4.2	73 8.4	4 0.5	80 9.2
	『日本語教育』	日本語教育学	940,700	1 0.1	0 0.0	0 0.0	22 2.3	3 0.3	4 0.4	0 0.0	103 10.9	77 8.2	9 1.0	43 4.6
	『年報社会学論集』	社会学	904,634	13 1.4	0 0.0	0 0.0	8 0.9	14 1.5	59 6.5	0 0.0	31 3.4	79 8.7	12 1.3	57 6.3
	『宗教哲学研究』	宗教哲学	828,695	11 1.3	5 0.6	0 0.0	13 1.6	6 0.7	87 10.5	0 0.0	17 2.1	50 6.0	3 0.4	118 14.2
	『刑法雑誌』	刑法学	701,881	12 1.7	0 0.0	0 0.0	7 1.0	8 1.1	62 8.8	0 0.0	26 3.7	64 9.1	5 0.7	127 18.1
	『国際政治』	国際政治学	1,114,644	2 0.2	0 0.0	0 0.0	10 0.9	5 0.4	33 3.0	0 0.0	40 3.6	110 9.9	8 0.7	71 6.4
	『国際経済』	国際経済学	984,529	4 0.4	2 0.2	0 0.0	35 3.6	4 0.4	23 2.3	0 0.0	65 6.6	61 6.2	22 2.2	122 12.4
	『日本経営学会誌』	経営学	1,056,199	1 0.1	0 0.0	0 0.0	42 4.0	13 1.2	60 5.7	0 0.0	69 6.5	159 15.1	34 3.2	145 13.7
	文系合計		8,126,594	70 0.9	8 0.1	0 0.0	173 2.1	55 0.7	341 4.2	0 0.0	396 4.9	681 8.4	98 1.2	791 9.7
理系	『数学』	数学	1,747,464	4 0.2	0 0.0	0 0.0	133 7.6	20 1.1	59 3.4	0 0.0	10 0.6	72 4.1	11 0.6	403 23.1
	『日本内科学会雑誌』	内科医学	648,009	0 0.0	0 0.0	0 0.0	14 2.2	1 0.2	2 0.3	0 0.0	34 5.2	25 3.9	7 1.1	49 7.6
	『化学と生物』	農芸学	635,408	0 0.0	0 0.0	2 0.3	5 0.8	1 0.2	14 2.2	0 0.0	75 11.8	41 6.5	10 1.6	63 9.9
	『日本化学会情報化学部会誌』	情報化学	649,547	2 0.3	0 0.0	0 0.0	3 0.5	2 0.3	3 0.5	0 0.0	15 2.3	48 7.4	15 2.3	80 12.3
	『海の研究』	海洋学	1,440,749	1 0.1	0 0.0	0 0.0	11 0.8	0 0.0	5 0.3	0 0.0	126 8.7	79 5.5	52 3.6	95 6.6
	『鉄と鋼』	鉄鋼学	909,423	0 0.0	0 0.0	0 0.0	11 1.2	2 0.2	3 0.3	0 0.0	39 4.3	70 7.7	21 2.3	121 13.3
	理系合計		6,030,600	7 0.1	0 0.0	2 0.0	177 2.9	26 0.4	86 1.4	0 0.0	299 5.0	335 5.6	116 1.9	811 13.4

「だから」の使用状況は、文系と理系で異なった様相を示している。文系では「だから」の10万字当たりの使用件数が0.9件だが、理系では0.1件である(表2)。使用者比率では

「だから」を使用している著者は文系が9.8%であるのに対し、理系は2%である（表3）。

文系全体の順接接続詞では「したがって」が9.7件と最も多く使用されており、「だから」（0.9件）の使用件数は「したがって」の使用件数の9.2%である。しかし、日本文学では「だから」の使用件数が3.4件であり、文系の他の分野と比べて多い。また日本文学の学術論文の順接接続詞では「したがって」の3.8件に次いで「だから」が3.4件使用されており、「だから」の使用件数は「したがって」の使用件数の89.4%である。社会学、宗教哲学、刑法では「だから」の10万字当たりの使用件数が1.3~1.7件であるが、日本語学、日本語教育学、国際政治学、国際経済学、経営学では0.1~0.4件と少ない（表2）。

使用者数でみると、日本文学の学術論文における「だから」の使用者比率は26%である。この割合は論説文全体における「だから」の使用者比率である28.4%に近い数値である。また社会学では著者のうち22%、宗教哲学では14%、刑法では12%の著者が「だから」を使用している。日本語学、日本語教育学、国際政治学、国際経済学、経営学は使用者比率が2~6%と少なくなっている（表4）。

論文・レポートを作成するための教材では、「だから」が話し言葉であり、論文・レポートでは使用すべきではないと説明されている。確かに理系学術論文では「だから」の使用が稀だが、文系学術論文では分野によっては使用されているのである。

表3 J-stageの学術論文における順接接続詞の使用者数と使用者比率（上段：使用者数 下段：使用者比率）

	分野	著者数	だから	それで	なので	よって	ゆえに	それゆえ	そのゆえ	その結果	そのため	このため	したがって
『日本文学』	日本文学	50	13	1	0	1	1	7	0	7	7	1	12
			26.0%	2.0%	0.0%	2.0%	2.0%	14.0%	0.0%	14.0%	14.0%	2.0%	24.0%
『日本語の研究』	日本語学	50	1	0	0	12	1	4	0	20	20	4	21
			2.0%	0.0%	0.0%	24.0%	2.0%	8.0%	0.0%	40.0%	40.0%	8.0%	42.0%
『日本語教育』	日本語教育学	50	1	0	0	11	3	3	0	35	34	5	19
			2.0%	0.0%	0.0%	22.0%	6.0%	6.0%	0.0%	70.0%	68.0%	10.0%	38.0%
『年報社会学論集』	社会学	50	11	0	0	4	4	19	0	22	34	9	27
			22.0%	0.0%	0.0%	8.0%	8.0%	38.0%	0.0%	44.0%	68.0%	18.0%	54.0%
『宗教哲学研究』	宗教哲学	50	7	2	0	6	6	28	0	11	14	3	36
			14.0%	4.0%	0.0%	12.0%	12.0%	56.0%	0.0%	22.0%	28.0%	6.0%	72.0%
『刑法雑誌』	刑法学	50	6	0	0	4	4	19	0	15	28	1	34
			12.0%	0.0%	0.0%	8.0%	8.0%	38.0%	0.0%	30.0%	56.0%	2.0%	68.0%
『国際政治』	国際政治学	50	1	0	0	8	3	16	0	27	36	4	24
			2.0%	0.0%	0.0%	16.0%	6.0%	32.0%	0.0%	54.0%	72.0%	8.0%	48.0%
『国際経済』	国際経済学	50	3	2	0	5	2	10	0	25	22	9	32
			6.0%	4.0%	0.0%	10.0%	4.0%	20.0%	0.0%	50.0%	44.0%	18.0%	64.0%
『日本経営学会誌』	経営学	50	1	0	0	9	4	17	0	27	33	13	39
			2.0%	0.0%	0.0%	18.0%	8.0%	34.0%	0.0%	54.0%	66.0%	26.0%	78.0%
文系合計		450	44	5	0	60	28	123	0	189	228	49	244
			9.8%	1.1%	0.0%	13.3%	6.2%	27.3%	0.0%	42.0%	50.7%	10.9%	54.2%
『数学』	数学	50	3	0	0	27	12	12	0	8	28	9	44
			6.0%	0.0%	0.0%	54.0%	24.0%	24.0%	0.0%	16.0%	56.0%	18.0%	88.0%
『日本内科学会雑誌』	内科医学	50	0	0	0	6	1	2	0	18	18	3	18
			0.0%	0.0%	0.0%	12.0%	2.0%	4.0%	0.0%	36.0%	36.0%	6.0%	36.0%
『化学と生物』	農芸学	50	0	0	1	5	1	5	0	34	14	5	26
			0.0%	0.0%	2.0%	10.0%	2.0%	10.0%	0.0%	68.0%	28.0%	10.0%	52.0%
『日本化学会情報化学部会誌』	情報化学	50	2	0	0	2	2	2	0	8	27	10	25
			4.0%	0.0%	0.0%	4.0%	4.0%	4.0%	0.0%	16.0%	54.0%	20.0%	50.0%
『海の研究』	海洋学	50	1	0	0	7	0	3	0	38	26	19	28
			2.0%	0.0%	0.0%	14.0%	0.0%	6.0%	0.0%	76.0%	52.0%	38.0%	56.0%
『鉄と鋼』	鉄鋼学	50	0	0	0	7	2	3	0	23	27	11	32
			0.0%	0.0%	0.0%	14.0%	4.0%	6.0%	0.0%	46.0%	54.0%	22.0%	64.0%
理系合計		300	6	0	1	54	18	27	0	129	140	57	173
			2.0%	0.0%	0.3%	18.0%	6.0%	9.0%	0.0%	43.0%	46.7%	19.0%	57.7%

4.4 論説文における順接接続詞の使用状況の調査結果

朝日新聞社言論サイト webronz に収録されている 5 分野（「文化・エンタメ」「科学・環境」「政治・国際」「経済・雇用」「社会・スポーツ」）における順接接続詞の使用件数を調査した結果が表 4 である（上段が使用件数，下段が 10 万字当たりの使用件数）。5 分野における順接接続詞の使用者数とその比率をまとめたのが表 5 である（上段が使用者数，下段が使用者比率）。

webronz の論説文における「だから」の使用件数は 10 万字あたり 8.7 件であり，順接接続詞の中では最も多い。また学術論文（文系が 0.9 件，理系が 0.1 件）に比べると「だから」が多く使用されていることがわかる。また（表 4）。また使用者比率で見ても「だから」は 28.4% であり，最も使用者数が多く，使用者比率で見ても学術論文よりも「だから」が多く使用されていることがわかる（表 5）。

表 4 webronz の論説文における順接接続詞の使用件数（上段：使用件数 下段：10 万字当たりの使用件数）

分野	字数	だから	それで	なので	よって	ゆえに	それゆえ	そのゆえ	その結果	そのため	このため	したがって
文化・エンタメ	262,997	30	0	0	0	0	3	0	2	11	4	8
		11.4	0.0	0.0	0.0	0.0	1.1	0.0	0.8	4.2	1.5	3.0
科学・環境	250,137	13	2	2	0	1	0	0	19	22	13	16
		5.2	0.8	0.8	0.0	0.4	0.0	0.0	7.6	8.8	5.2	6.4
政治・国際	276,833	19	0	0	0	6	2	0	8	16	5	11
		6.9	0.0	0.0	0.0	2.2	0.7	0.0	2.9	5.8	1.8	4.0
経済・雇用	261,933	26	1	0	1	1	1	0	5	9	8	9
		9.9	0.4	0.0	0.4	0.4	0.4	0.0	1.9	3.4	3.1	3.4
社会・スポーツ	269,858	27	1	0	0	1	6	0	7	9	0	11
		10.0	0.4	0.0	0.0	0.4	2.2	0.0	2.6	3.3	0.0	4.1
合計	1,321,758	115	4	2	1	9	12	0	41	67	30	55
		8.7	0.3	0.2	0.1	0.7	0.9	0.0	3.1	5.1	2.3	4.2

表 6 webronz の論説文における順接接続詞の使用者数と使用者比率（上段：使用者数 下段：使用者比率）

分野	著者数	だから	それで	なので	よって	ゆえに	それゆえ	そのゆえ	その結果	そのため	このため	したがって
文化・エンタメ	50	24	0	0	0	0	3	0	2	9	2	7
		48.0%	0.0%	0.0%	0.0%	0.0%	6.0%	0.0%	4.0%	18.0%	4.0%	14.0%
科学・環境	50	9	2	2	0	1	0	0	13	15	11	14
		18.0%	4.0%	4.0%	0.0%	2.0%	0.0%	0.0%	26.0%	30.0%	22.0%	28.0%
政治・国際	50	13	0	0	0	3	2	0	7	13	3	9
		26.0%	0.0%	0.0%	0.0%	6.0%	4.0%	0.0%	14.0%	26.0%	6.0%	18.0%
経済・雇用	50	14	1	0	1	1	1	0	3	7	7	8
		28.0%	2.0%	0.0%	2.0%	2.0%	2.0%	0.0%	6.0%	14.0%	14.0%	16.0%
社会・スポーツ	50	11	1	0	0	1	6	0	6	8	0	7
		22.0%	2.0%	0.0%	0.0%	2.0%	12.0%	0.0%	12.0%	16.0%	0.0%	14.0%
合計	250	71	4	2	1	6	12	0	31	52	23	45
		28.4%	1.6%	0.8%	0.4%	2.4%	4.8%	0.0%	12.4%	20.8%	9.2%	18.0%

「文化・エンタメ」「政治・国際」「経済・雇用」「社会・スポーツ」の分野では順接接続詞の中で「だから」が最も多く使用されている。理系分野である「科学・環境」の分野では「その結果」「そのため」「したがって」に比べると「だから」の使用件数が少ないが，最も使用されている「そのため」8.8 件と「だから」5.2 件の割合は 71.8% であり，「だから」の使用件数が 0.1 件の理系学術論文に比べると多く使用されている（表 4）。「科学・環境」分野の論説文における「だから」の使用者の比率は 18% であり，論説文の他の分野に比べると

少ないものの、「だから」の使用者の比率が2%の理系学術論文と比べると、「科学・環境」の分野で「だから」が多く使用されていることがわかる（表5）。

webronzaでは「文化・エンタメ」「科学・環境」「政治・国際」「経済・雇用」「社会・スポーツ」の5分野に分類されている。しかし、例えば「科学・環境」の分野で感染症について述べつつ、感染症の政策についても論じている論説文があり、「政治・国際」の分野にも関連した内容が書かれている。学術論文に比べると専門分化されておらず、他の分野に跨る内容の論説文も見られる。

6. 学術論文と論説文における「だから」

前章で見てきたように学術論文では分野によっては「だから」が使用され、論説文ではどの分野においても「だから」が使用されている。以下、学術論文や論説文で使用される「だから」の用例を見ていくことにする。

6.1 学術論文における「だから」

表2で見ると、日本文学では10万字あたりの「だから」の使用件数が3.4件であり、他の分野と比べて、「だから」が多く使用されている。また表3で見ると日本文学、宗教哲学、社会学、刑法の分野で「だから」の使用者数が10%以上である。これらの分野では著者が文学作品や文化・社会現象の解釈を行うことがあるが、前件で理由を述べ、後件で著者が解釈をする際に「だから」が使用されることがある。以下、(1)~(4)がその用例である。

- (1) おそらく、二日の日の宴は、国司たちは各地より参集した郡司たちの接待に余念がなかったはずであり、家持をはじめとする国司たちは、その接待に気を使ったものと思われる。だから、家持たちは、郡司の招待宴が終わると安堵したはずである。（『日本文学』2014年63巻5号）
- (2) また既に触れたように曳山祭においてシャギリが農村部からの雇いシャギリによるものであって、山組によって行われるものではなかったため、山組間の競合の対象ではなかったことが挙げられる。だからこそ、山組間のシャギリの協力も問題とされにくかった。（『年報社会学論集』2016年29号）
- (3) 死者のいくばくかの実在性は、私たちによる死者についての語りのなかに今なお生き生きとしている。だから、その死者を傷つけてはいけなく、その死者についての語りをそれ自体として尊重しなければいけない、ということである。（『宗教哲学研究』2019年36巻）
- (4) 人は本来「弱い」存在であり、だからこそ各種の基本権の保障など法による保護を必要とする。（『刑法雑誌』2016年55巻3号）

向坂（2019）で論じたように、「だから」は主観的な結論付けを表す接続詞だが、学術論文の著者による結論付けは一つの解釈であり、必然的な結論とは限らず、著者の主観性が入り込む余地があるため「だから」を出現させていると考えられる。一方、理系分野では数値などに基づいて結論を出すことが求められているが、数値は客観的なものであり、著者の主観が入り込む余地がなく、またそこから得られる結論も必然的なものであるとされるため、理系の学術論文において「だから」の出現を妨げていると考えられる。

6.2 論説文における「だから」

論説文とは一般的な問題について自分の意見を述べ、論議する文章である。(5)~(7)は前件で理由を述べ、後件で著者が意見を主張する際に「だから」が用いられている用例である。論説文では、自分の意見を述べるのが重要であり、(5)~(7)で述べられていることは主観的な意見である。学術論文（とくに理系）では数値などに基づいた客観的な根拠を挙げることを要求されているが、論説文では学術論文に比べると客観性が要求されておらず、そのことが「だから」の使用を可能にしていると考えられる。

- (5) そしてこの「万人が認める事実」を提供するのが「科学」である。だから、デモクラシーにサイエンスリテラシーが必須なのである。(webronza 科学・環境 2021年4月13日)
- (6) お寺の未来は、そう明るくはない。だからこそ、檀家を含めた、関係者みんなの力を結集して、将来のことを考えるべきだと思う。(webronza 文化・エンタメ 2022年5月24日)
- (7) しかし、その「相続」から排除されてきた子どもたちがいる。それが「あるはずのものがない」親たちの現実である。だから簡単に「虐待する親」とは呼べないと私は思う。(webronza 社会・スポーツ 2021年12月10日)
- また、文系学術論文と同じように、社会や文化の現象についての著者の解釈の際に「だから」が用いられている用例が見られる。以下(8)~(10)がその用例である。
- (8) こうして完成した映画は、満若監督のキャラクターが乗り移ったかのように、まことに饒舌である。そしてその饒舌さゆえに、単純な要約を許さない。だから観た人は、それぞれに受け止め、自らの思いを持ち帰って解釈をするしかない。(webronza 文化・エンタメ 2022年5月19日)
- (9) 安倍・菅政権は、政権交代が次第に遠くにかすんでいくという実感を持たなかった政権でもあった。だからこそ、安倍首相は2014年、17年と小刻みに衆議院を解散し、政権基盤を確保しようとしたのである。(webronza 政治・国際 2022年5月3日)
- (10) ちなみに「シェール革命」でシェールガス、シェールオイルの生産を増やし、いまや世界最大の産出国となった米国の自給率は106%、北海油田がある英国も75%と高い。だからこそ米英はロシアに遠慮なく経済制裁をかけられるという面もある。(webronza 政治・国際 2022年4月27日)

7. なぜ小論文では「だから」の使用が可能なのか

向坂(2019)は日本語非母語話者向けの日本語学習教材において「だから」を話し言葉としている一方、日本語母語話者を対象とする国語表現の教材で「だから」を書き言葉であるとしており、日本語非母語話者向けの教材と日本語母語話者向け教材の比較をもって「接続詞「だから」の文体差について日本語教育と国語教育の間に共通した認識がない」と述べている(pp.174-176)。

しかし、これは日本語教育と国語教育の認識の違いではなく、教材がどのような文章を書くことを目標としているかの違いであろう。表1で見ると、論文・レポートを書くことを目標としている教材ではいずれも「だから」が話し言葉であるとされ、小論文を書くことを目標としている教材では「だから」が書き言葉とされているものと話し言葉とされているものがある。すなわち、小論文では「だから」の使用を可能であると説明している教材があるということである。

伊集院・高野(2020)(日本語文章表現教材『日本語を学ぶ人のためのアカデミック・ライティング講座』)では、「小論文」とは「与えられた課題に関し、主張(意見・立場・提案など)を根拠を挙げて論理的に伝えることを目的とする文章」であり、「根拠」は書き手の頭の中の知識や一般的知識、社会的常識でよい。資料やデータに基づく客観的証拠の提示や文献の引用はなくても問題ない」と説明している。一方で「レポート」とは「設定したテーマに関し、文献や資料を読んだり、調査を行ったりして情報を収集し、それらの根拠に基づいて結果と考察を論理的にまとめた文章」であり、主張の根拠としては「客観的な情報・データを収集することが必須」と述べている(p.11)。

論説文も小論文も「自分の意見を論理的に述べる」ことが重要だが、学術論文やレポートに比べると資料やデータなどの客観性の要求は高くない。

また、石黒(2012)(日本語文章表現教材『論文・レポートの基本』)では「小論文は時間制限と戦いながら試験場で書く文章です。そのため、その場で思いついた発想を自分なりに論理を組み立てて面白く書けばよく、厳密な検証は要求されません」とし、「小論文」では

「ウソ」は「ある程度許容」されると述べている。一方、「論文」や「レポート」では「ウソ」は「認められない」としている (pp.3-4)。石黒 (2012) が言う「ウソ」とは厳密な検証を経ていない不正確な内容であると考えられる。

小論文は入学試験や就職試験の科目であり、通常は制限時間内に完成させなければならぬように資料閲覧が禁止されているため、資料やデータに基づく客観的な根拠を示すことが物理的に困難である。

小論文は、学術論文やレポートに比べると客観性の要求が低いいため、主観的な結論付けを表す「だから」の使用が可能なのである。

8. まとめ

日本語文章表現教材には「だから」は話し言葉であり、書き言葉文では使用されないとしているものと、書き言葉であり、書き言葉文では使用されるとしているものがある。学術論文と論説文における使用状況を調査したところ、理系学術論文では使用が稀であること、文系学術論文では分野によっては使用されていること、論説文では使用されていることがわかった。

「だから」が使用されないジャンルや分野の書き言葉文で「だから」が使用されないのは、「だから」が話し言葉であるためではない。「だから」は主観的な結論付けを表す接続詞であるため、理系学術論文のように客観性の要求が高いジャンルや分野では使用されないのである。「だから」が許容されるか否かは、書き言葉文であるか、話し言葉文であるかではなく、客観性の要求が高いかどうかである。小論文を書くことを目標とした教材で「だから」を書き言葉であると説明しているものがあるのも、小論文は学術論文やレポートに比べると客観性の要求が高くないためであると考えられる。

日本語文章表現教材では「「だから」は話し言葉であり、書き言葉文では使用されない」と説明しているものがあるが、学習者はどのような場合が主観的な結論付けで、どのような場合が客観的な結論付けなのかを判断できない場合もあるので、教育上の配慮として言い切った形で説明していることも考えられる。しかしこの説明は「だから」が主観的な結論付けを表す接続詞であるという本質を見えなくしている。教員としてはこの本質をよく理解し、著者である学習者が主観的な解釈や意見を述べているのか、それとも客観的な証拠に基づいて結論付けを行っているのかを見極めた上で学習者を指導する必要があると考えられる。

本研究では学術論文と論説文で「だから」が使われる例として、著者が現象を「解釈」をする場合と著者が主張を「意見」を述べる場合を挙げたが、どのような場合に学術論文や論説文で「だから」が使われるのかについてはさらなる考察が必要であり、今後の検討課題とする。

参考文献

- 石黒圭(2012).『論文・レポートの基本』,日本実業出版社.
- 石黒圭 (2016).『接続詞の技術』,実務教育出版.
- 伊集院郁子・高野愛子(2020).『日本語を学ぶ人のためのアカデミック・ライティング講座』,アスク出版.
- 沖森卓也(2016).『文章が変わる接続語の使い方』,ベレ出版.
- 国語教育研究所(1991).『国語教育研究大辞典普及版』,明治図書.
- ひげひろし(1987).「「それで」「だから」「したがって」」『教育国語』88, pp.46-59.
- 向坂卓也(2019).「学術論文における接続詞「だから」の使用」『日語偏誤与日語教学研究』4, pp.172-188
- 向坂卓也(2021).「学会誌論文における原因理由を表す接続表現「ことから」」『日本語教育

方法研究会誌』 28 (1), pp.112-113.

劉怡伶(2006). 「接続語「だから」の意味・用法 - 前件と後件に因果関係が認められる「だから」を中心に -」『世界の日本語教育』 16, pp.125-137.

参考教材

アカデミック・ジャパニーズ研究会(2015). 『改訂版 大学・大学院 留学生の日本語④論文作成編』, アルク.

石黒圭・筒井千絵 (2009). 『留学生のためのここが大切 文章表現のルール』, スリーエーネットワーク.

石黒圭(2012). 『論文・レポートの基本』, 日本実業出版社.

伊集院郁子・高野愛子(2020). 『日本語を学ぶ人のためのアカデミック・ライティング講座』, アスク出版.

伊藤博貴(2016). 『小論文をひとつひとつわかりやすく』, 学研プラス.

井下千以子(2019). 『思考を鍛えるレポート・論文作成法第3版』, 慶應義塾大学出版会.

河守晃芳(2021). 『7日間で合格する小論文』, 学研プラス.

神崎史彦(2010). 『大学入試小論文の完全攻略本』, 文英堂.

桑田てるみ・江竜珠緒・押木和子・勝亦あき子・松田ユリ子(2015). 『学生のレポート・論文作成トレーニング - 改訂版』, 実教出版.

小森万里・三井久美子(2016). 『レポート・論文を書くための日本語文法』, くろしお出版.

鈴木鋭智(2020). 『改訂版 何を書けばいいかわからない人のための 小論文のオキテ55』, KADOKAWA.

大修館書店(2018). 『国語表現 改訂版 (平成30年)』, 大修館書店.

友松悦子(2008). 『小論文への12のステップ』, スリーエーネットワーク.

二通信子・佐藤不二子(2003). 『改訂版 留学生のための論理的な文章の書き方』, スリーエーネットワーク.

二通信子・大島弥生・佐藤勢紀子・因京子・山本富美子(2009). 『留学生と日本人学生のためのレポート・論文表現ハンドブック』, 東京大学出版会.

浜田麻里・平尾得子・由井紀久子(1997). 『大学生と留学生のための論文ワークブック』, くろしお出版.

樋口裕一(2004). 『小論文 これだけ!書き方 超基礎編』, 東洋経済新報社.

深澤のぞみ・濱田美和・深川美帆・札野寛子・松田佳子(2018). 『21世紀のカレッジ・ジャパニーズ』, 国書刊行会.

関連 URL

J-stage 国立研究開発法人科学技術振興機構 <https://www.jstage.jst.go.jp/>

webronza 朝日新聞社 <https://webronza.asahi.com/>

少数言語のデジタルアーカイブ：PhoPhoNO と BantuDArc

李 勝勲 (国際基督教大学) †

倉部 慶太 (AA 研)

品川 大輔 (AA 研)

Digital archives of understudied languages: PhoPhoNO and BantuDArc

Seunghun J. Lee (International Christian University)

Keita Kurabe (ILCAA)

Daisuke Shinagawa (ILCAA)

要旨

大言語を対象とした様々なデジタルアーカイブに基づく研究が進展する一方で、少数言語を対象としたデジタルアーカイブの構築とその活用はまだ十分に進んでいるとはいえない。本稿では少数言語を中心に著者らが構築したデジタルアーカイブを紹介し、少数言語を対象としたアーカイブ化に関して議論する。一つ目はチベット・ビルマ系の5言語に関する資料を公開するアーカイブサイト 'PhoPhoNO'、もう一つはバントゥ系の5言語の資料をアーカイブ化したサイト 'Bantu Language Digital Archive (BantuDArc)' である。各サイトは言語に関するメタデータ、地図、そして言語資源から構成される。音声資料を含む個別のデータ項目には固有のIDが付与され、申請によってアクセスを認められれば、利用者はそれらデータを研究資源として利活用することができる。

1. はじめに

本稿は少数言語を対象にした二つのアーカイブサイトを紹介することを目的とする。少数言語を対象とする大規模デジタルアーカイブとして国際的に広く認知されているサイトには、イギリスの危機言語アーカイブ (ELAR: Endangered Languages Archive)、ドイツの DOBES アーカイブ (Documentation of Endangered Languages)、オーストラリアの PARADISEC (Pacific and Regional Archive for Digital Sources in Endangered Cultures) などが挙げられるが、それ以外にも近年では南アフリカの SADiLaR (South African Centre for Digital Language Resources; 南部アフリカ諸語を対象とする) やアメリカ・テキサス州の AILLA (Archive of the Indigenous Languages in Latin America; ラテン・アメリカの少数言語を対象とする) など、特定の地域や言語群を対象としたアーカイブも構築されている。日本では国立国語研究所が提供する日本語を対象とした例えば「危機言語データベース」などのアーカイブや東京外国語大学アジア・アフリカ言語文化研究所 (AA 研) の情報資源利用研究センター (IRC) が管理・維持するアジア・アフリカ地域を中心とした言語文化に関するアーカイブがある。

アーカイブの構築に際しては、保存されているデータの質だけでなく、ユーザビリティの観点からデータの整理形式についても配慮する必要がある。さらに、データの共有範囲や

† seunghun@icu.ac.jp

データそのものに関する質的な説明も求められる。以下、二節では南アジアと東南アジアの言語をアーカイブする PhoPhoNO の、三節ではアフリカの 5 つのバントゥ諸語の言語資料アーカイブである BantuDArc の、データ構造とその意義について紹介する。

2. 南アジア・東南アジア言語のアーカイブ (PhoPhoNO)

2.1 背景

このアーカイブのもととなるデータは、日本学術振興会国際共同研究事業（スイスとの国際共同研究プログラム）Phonetics, Phonology and New Orthographies in Roman and Indigenous Script: Helping Native Language Communities in the Himalayas (PhoPhoNO) によって収集された。2017 年から 2020 年までのプロジェクト期間中、インド・シッキム州のデンゾン語 (Drenjongke)、ネパールのタマン語 (Tamang)、ブータンのゾンカ語 (Dzongkha)、ミャンマーのビルマ語 (Burmese) など 4 つのチベット・ビルマ系言語、およびミャンマーで話されるモン・クメール系のモン語 (Mon) の言語データを収集した。

本アーカイブは、それらフィールドワークによって収集した一次資料の音声データをアーカイブ化したものである。元のデータはそれぞれのプロジェクトの目的に従って処理されている。そこで、アーカイブ構築に際しては、アーカイブ化に適したデータ処理を行った。なお、録音に参加した調査協力者からはアーカイブ関連の承認を事前に取得済であった。アーカイブ化には AA 研の情報資源利用研究センター (IRC) から 2019 年度と 2020 年度に支援を受け、PhoPhoNO プロジェクトによって収集された音声データを処理することで、デジタルアーカイブを構築した (<https://phophono.aa-ken.jp>、図 1)。グーグルフォームによる申請によって研究目的の利活用であることが承認されれば、ファイル・サーバへのリンクが送信されるシステムになっている。

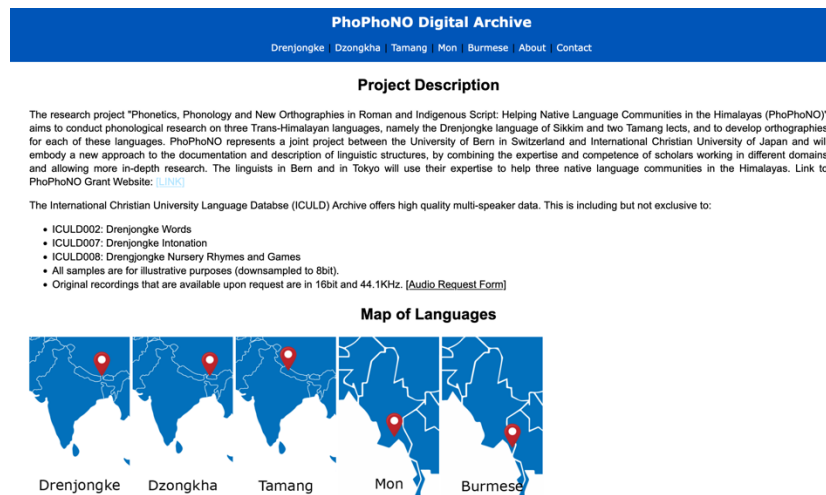


図 1 PhoPhoNO サイトのメインページ

表 1 PhoPhoNO アーカイブ内のデータベース

言語	ICULD	サイズ	長さ	内容
デンゾン語	1~9	2836.8 MB	314.7 分	音声・テキスト・プロット
ゾンカ語	10~11	99.7 MB	16.5 分	音声・テキスト・プロット
タマン語	12~15	9340MB	309.75 分	音声・テキスト・プロット
ビルマ語	32	5890MB	74.9 分	音声・テキスト・EGG
モン語	31	345MB	9.28 分	音声・テキスト・プロット

2.2 PhoPhoNO アーカイブの構造

このアーカイブは表 1 に示されている 5 つの言語のデータベース (ICU Language Database, ICULD) から構成されている。デンゾン語の内容については Baldoria et al. (2020a) および Kunzang Namgyal & Lee (to appear, 2022) に詳しく述べているので、ここでは他の 4 つの言語のアーカイブを中心に説明する。また Baldoria et al. (2020b) では 4 つの言語のトークンの確認ができる。

ブータンの公用語であるゾンカ語のデータベースには一人の話者が発話したゾンカ語の単語と音節 (syllabary) の録音 (ICULD-0010) と、ゾンカ語の個別の分節音を示すための録音 (ICULD-0011) が収録されている。タマン語のデータベースには多数の話者による東部方言 (ICULD-0012 と-0013) および西部方言 (ICULD-0014 と-0015) の録音が収録されている。これによって、二つの方言の母音の発音又は単語の発音の相違を対照的に把握することが可能になる。ビルマ語のデータベース (ICULD-0032) には声調 (tone) と発声 (phonation) に関する特徴を示すための録音があり、唯一のモン・クメール系言語であるモン語のデータベース (ICULD-0031) には発声に関する諸特徴を検証するためのデータが収録されている。

2.3 PhoPhoNO アーカイブの意義

デンゾン語の場合、これまで研究利用が可能な音声データ自体が存在しなかったため、PhoPhoNO によって初めて、電子的に利用可能な言語資源のアーカイブが構築されたことになる。ゾンカ語、タマン語、ビルマ語、モン語についても、国際的に電子データの公開が多かれ少なかれ進められているものの、PhoPhoNO アーカイブは独自資料に基づくアーカイブであり、また既存データを補完する役割も果たす。将来はこのデータベースの資料を利用した分析論文が出版されることが期待される。

3. バントウ系言語のアーカイブ : Bantu Language Digital Archive (BantuDArc)

3.1 背景

バントウ諸語は日本における長期の研究蓄積を有する言語群の一つである。BantuDArc はその研究成果の一部としての音声資料を一箇所に集約することを目的とする。6 つの南部バントウ諸語の形態統語論調査によって収集され、2021 年に公開された音声資料を含むアーカイブ (Bantu Microvariation Digital Archive; <https://renelda.aa-ken.jp/about.html>)、およびツォンガ語を含む統語構造の声調に対する影響に関するプロジェクト (ECPPPT: Effects of Syntactic Constituency on Phonology and Phonetics of Tone” Digital Archive; <https://ecppt.aa-ken.jp/sub.html>) のアーカイブを作成した経験に基づき、2022 年に構築された。PhoPhoNO アーカイブ同様、BantuDArc も AA 研の IRC からの支援を受けて、データのアーカイブ化作業およびウェブサイト構築が可能になった。

BantuDArc は当初から収録言語を拡張することを目的に構築されている。現在は南アフリカのツォンガ語 (S53) と東アフリカのルワ語 (E621A)、ベンデ語 (F12)、ルンディ語 (JD62)、さらにスワヒリ語 (G42) の録音ファイルを申請ベースによってダウンロードすることが可能である。

3.2 BantuDArc アーカイブの構造

BantuDArc のデータベース構成は表 2 のとおりである。Praat で生成された各音声ファイルのプロット (具体的には spectrogram と pitch track) はウェブサイトから制限なしでダウンロードすることができる。音声ファイルに関しては、PhoPhoNO 同様、研究目的であることが申請によって確認できる場合のみアクセスが可能になる。データベースの内容は Johnson

et al. (2022)で確認出来る。

表 2 BantuDArc アーカイブ内のデータベース

言語	ICULD	サイズ	長さ	内容
ツォンガ語	41	724.2 MB	186.41 分	音声・テキスト・プロット
ルワ語	44	229.3 MB	49.66 分	音声・テキスト・プロット
ベンデ語	42~43	966.2 MB	199.8 分	音声・テキスト・プロット
ルンディ語	46	137.7MB	31.57 分	音声・テキスト・プロット
スワヒリ語	45	333.2MB	96.51 分	音声・テキスト・プロット

ツォンガ語のデータベース (ICULD-0041) には一人の話者による発話がトークン化されて収録されている。多様なデータの内容に関する形態統語論レベルの記述情報は Lee et al. (2022) に整理されている。ルワ語 (ICULD-0044) には時制や相をパラメーターとする多様で複雑な動詞の変化形が網羅的に録音されている。ベンデ語は 2 つのデータベースから構成されている。ReNeLDA プロジェクトのために収集した形態統語論に関するデータ (ICULD0042) に加え、ICULD-0043 にはさまざまな語形式が記録されている。具体的には、名詞の場合は単数形と複数形、また形容詞や指示代名詞によって修飾される構造が、動詞の場合は人称や名詞クラスなどの文法範疇に基づく活用パラダイムが録音されている。ルンディ語 (ICULD-0046) は語の活用形や形態統語論的分析を目的とした例文データを収録し、ICULD-0045 にはスワヒリ語の付加詞 (particle) のイントネーションを分析するために収集された発話の録音が保存されている。

3.3 BantuDArc アーカイブの意義

バントゥ諸語の記述的資料は多く蓄積されているが、バントゥ諸語に特化した録音資料をダウンロード出来るアーカイブはほとんど存在しない。そのような状況をふまえて BantuDArc はあらゆるバントゥ諸語を対象とし、研究者と現地コミュニティの人々の双方が言語資源にアクセスできる場を提供している。

現在、ツォンガ語に関しては Lee et al. (2022) に、ルンディ語に関しては Shinagawa et al. (2022) に、それぞれ収録データに基づく言語学的な分析結果が提示されている。これらの成果は、母語話者の著者との共同出版としての意義も併せ持つ。

4. おわりに

本稿では二つのアーカイブ PhoPhoNO および BantuDArc を紹介した。研究成果の一部としての録音ファイルを基盤に構築されたこれらウェブサイトは、より広範囲の研究者のみならず母語話者コミュニティに対する資源的貢献となることをも意図して構築されている。

謝 辞

本研究は東京外国語大学アジア・アフリカ言語文化研究所 (AA 研) の情報資源利用研究センター (IRC) の支援を端緒とするものであり、本稿は科研費の国際共同研究加速化基金 (B) 「Microvariation in Bantu languages of South Africa: Building theories from typology data」及び AA 研共同利用・共同研究課題「通言語的観点からみた音声類型論 (PhonTyp)」の支援を受けて行われたものである。

文 献

- Baldoria, Yukki, Audrey Lai, Hannah Lee & Tomoko Monou (eds.) (2020a) *ICU Working Papers in Linguistics 11: ICU Language Database Series 1: PhoPhoNO Digital Archive 1*. Tokyo, Japan: ICU LingLab.
- Baldoria, Yukki, Audrey Lai & Rachel Liu (eds.) (2020b) *ICU Working Papers in Linguistics 12: ICU Language Database Series 2: PhoPhoNO Digital Archive 2*. Tokyo, Japan: ICU LingLab.
- Johnson, Kiara, Sayaka Manabe, Chrisanne M. Tuaño. (eds.). (2022) *International Christian University Working Papers in Linguistics 21: Bantu Language Digital Archive*. Tokyo, Japan: IUC LingLab.
- Kunzang Namgyal & Seunghun J. Lee (2022, to appear) Drenjongke (Bhutia) databases: a collaborative project with language experts. Ms.
- Lee, Seunghun J., Babane Morris Thembhani & Madala Crous Hlungwani. (2022) *Aspects of Xitsonga Grammar*. Tokyo, Japan: ILCAA.
- Shinagwa, Daisuke, Yuko Abe, Seunghun J. Lee & Chérubin Mugisha (2022) *Selected topics of Kirundi Grammar: A micro-typological perspective*. Tokyo, Japan: ILCAA.

関連 URL

PhoPhoNO アーカイブ
BantuDArc アーカイブ

<https://phophonno.aa-ken.jp/>
<http://BantuDArc.aa-ken.jp/>

「小中高大生による日本語絵描写ストーリーライティングコーパス」 (JASWRIC) の構築 : L1/L2 日本語研究の新しい資料として

石川慎一郎 (神戸大学)

友永達也 (神戸大学附属小学校), 大西遼平, 岡本利昭, 勝部尚樹, 川嶋久予,
岸本達也, 村中礼子 (神戸大学附属中等教育学校)

Japanese Students' L1 Story Writing Corpus (JASWRIC): A New Dataset for Analysis of L1/L2 Japanese

Shin'ichiro ISHIKAWA (Kobe University)

Tatsuya TOMONAGA (Kobe University, Elementary School), Ryohei ONISHI, Toshiaki
OKAMOTO, Naoki KATSUBE, Hisayo KAWASHIMA, Tatsuya KISHIMOTO, Reiko
MURANAKA (Kobe University, Secondary School)

要旨

本稿は、「小中高大生による日本語絵描写ストーリーライティングコーパス」(JASWRIC)の構築過程と概要を報告する。JASWRICには、700名の小中高大生による約13.6万語(短単位)のL1日本語作文が収録されている。全データは、ダウンロード版とオンライン版(JASWRIC Online)の2系統で公開される。一般公開されているL1の子どもの作文コーパスがほとんどない中で、JASWRICは、L1日本語の発達過程を調べる有益な資料となるだろう。また、JASWRICのデータは、「多言語母語の日本語学習者横断コーパス」(I-JAS)で採用されたストーリーライティングのプロンプトを使って集められている。このため、JASWRICは、I-JASと併用することで、L1/L2対照研究の参照データとしても使用可能である。

Abstract

This article introduces *Japanese Students' L1 Story Writing Corpus* (JASWRIC), which includes approximately 136,000-word L1 Japanese writing by 700 primary school, secondary school, and college students. The collected data is available as a download version as well as from the online corpus query system called *JASWRIC Online*. Considering that there have been almost no L1 Japanese students' essay corpora publicly available, we could expect that JASWRIC will be used as a valuable resource for the study of the development of children's L1 Japanese knowledge and skills. In addition, JASWRIC can also be utilized as reference data in L2 Japanese acquisition studies because its data is collected by the picture-based story writing task common to that adopted in *International Corpus of Japanese as a Second Language* (I-JAS). Adoption of the common picture prompts guarantees the validity in a contrastive study of L1 and L2 Japanese based on I-JAS and JASWRIC.

1. はじめに

「多言語母語の日本語学習者横断コーパス」(I-JAS) (迫田ほか、2020) がリリースされたことで、学習者によるL2日本語の習得過程については、大型コーパスを用いた実証的研究が広く行われるようになってきた。一方、母語話者の子ども(本研究では、広く未成年者を指す)のL1日本語の発達過程を直接の研究対象としたコーパス研究は、管見の限り、必

ずしも多くない。これには、研究基盤となる公開コーパスの少なさが影響している。

もちろん、過去において、子どもによる L1 日本語の産出データの収集例がなかったわけではない。まず、話し言葉について言えば、Talk Bank System のサブモジュールと位置付けられる CHILDES (Child Language Data Exchange System) に、幼児の L1 日本語発話等を集めた 12 種の資料が掲載されている。ただ、PaidoJapanese 資料 (2~5 歳児 85 名の発音) と Okayama 資料 (2~4 歳児 130 名と親との対話) を除くと、大半は 1 名ないし 2~6 名程度の小規模なデータである。これらは、音声学、発達心理学、会話分析など、特定の関心に即して集められた小規模データであり、一般的な意味でのコーパスとみなすことはむづかしい。ただ、近年になって、子どもの発話の体系的な収集も始まっており、小磯ほか (2020) では、8 世帯 10 名の子どもの会話と、幼稚園に通う園児の会話を集める「子どもの会話コーパス」の開発計画が紹介されている。研究の成果が期待される場所である。

次に、書き言葉について言うと、データ収集の対象は、文字の習得が前提になるため、必然的に、小学生以降となる。小学校では作文を書かせる指導が広く行われており、データ自体は随所に存在しているわけだが、言語学ないし言語習得研究の目的でそれらを体系的に収集し、本人および保護者の許諾を得て公開の研究資料とすることのハードルは高い。とくに、小学校の国語授業で書かせることが多い自由作文には、子どものプライバシー情報が出やすいという問題もある。

初期の作文収集事例として、島村 (1987) は、千葉県の 3 つの小学校に在籍する 2 年生・4 年生・6 年生約 360 名が「わたしの学校」「先生」「ともだち」というテーマで書いた作文を集めた。その一部 (約 34 万語) は、後に「学校課題作文コーパス」として電子化されている (今田・宮城、2020)。国立国語研究所 (1989) は各地の教育委員会などの編纂による小学生文集 10 種 (10 年分) から作文 47 万語を集めた。また、成田ほか (1995) は、小学校 1 年生から高校 2 年生と、大学生および大学留学生 520 名が「手」という同一のテーマで書いた「同題作文」を収集した。本資料を用いた研究としては、成田ほか (1995) に加え、村上・田中 (1997) および田中 (1998) がある。

近年では、より体系的な作文データの収集が試みられている。2010 年代前半に作られた各種作文コーパスを紹介した富士原ほか (2016) のまとめによれば、永田ほか (2010) は小学校 5 年生児童の読書ブログ約 4 万語を集め、坂本 (2010) は全国の小学校のウェブサイト公開されていた作文約 1 万本、約 123 万語を集め、鈴木ほか (2011) は中学校 1 年生から高校 2 年生の作文 25 万語を集めた。また、阿部ほか (2017) は、上述の成田ほか (1995) で集められたデータのうち、小中学生が書いた約 5 万語の作文を電子化し、あわせて、同じ題目に基づいて現代の小中学生が書いた作文約 23 万語を集め、総計 28 万語の「『手』作文コーパス」を構築した。最近の特筆すべき成果として、宮城・今田 (2018) は、小学校 1 年生から中学 3 年生までの児童・生徒が「ゆめ」および「ぼくの／私のがんばったこと」というテーマについて授業内で書いた作文約 5,300 編を集め、約 164 万語の「児童・生徒作文コーパス」を構築した (作文数・語数は今田 (2020) による)。なお、本コーパスは、富士原ほか (2016) において 70 万語の「小・中学生通年作文コーパス」として言及されていたものである。本コーパスについては、作文中の「問題例」の抽出を可能にする「問題例検索システム (仮称)」が開発中とのことである (砂川、2020)。

こうした過去の作文収集はいずれも価値ある研究実践と言えるが、(1)小学生・中学生の作文が大半で、高校生以上の作文を含めた資料が少ない、(2)L2 研究との接合性を考慮した資料が少ない、(3)緩やかなテーマで書かせた自由作文が多く、内容を統制して集めた資料

が少ない、(4)一般公開されているコーパスがほとんど存在しない、といった制約もある。この意味において、坂本が2010年時点で述べた「日本の子供の書き言葉コーパスは非常に少ないという現状」はその後大きく変化していないようである。

こうした現状をふまえ、本稿第一筆者（以下、筆者）は、(1')小学校1年生から大学1年生までの13学年の児童・生徒・学生に、(2'-3')既存の日本語学習者コーパスで使用されたものと同じイラストを使い、その内容を描写させるストーリーライティング形式で作文を行わせ、(4')集めたデータをダウンロード版・オンライン版として全面的に一般公開する、という基本コンセプトのもと、136,635語（短単位）からなる新たなコーパスを構築した。本稿は、以下、当該コーパスの構築過程と概要について報告する。

2. JASWRIC の開発経緯

2.1 目的

「小中高大生による日本語絵描写ストーリーライティングコーパス」(Japanese Students' L1 Story Writing Corpus : JASWRIC/ジャスリック)は、2つの研究目的に基づいて開発された。1点目は、JASWRICを単独で使用することで、学齢期の子どものL1日本語の書き言葉の発達過程を調査することである。2点目は、1,000人の日本語学習者と50名の成人日本語母語話者の産出データを収集しているI-JASのストーリーライティングデータと併用することで、L1の子どものL1の成人間の比較や、L1の子どものL2学習者の比較を行うことである。後者について、学習者コーパス研究では、学習者の産出を、参照基準となる母語話者の産出と比較して学習者の過剰・過少使用を特定する中間言語対照分析 (contrastive interlanguage analysis) が広く行われているが、少数の母語話者データを参照基準とすることには批判もある。I-JASの母語話者データをJASWRICで拡張することにより、参照基準が多様化し、I-JASを用いたL1/L2対照分析の妥当性と安定性が向上するという効果も期待される(石川、2022)。

2.2 データ収集の過程

JASWRICのデータ収集には、関西圏に所在する国立大学法人が設置する3つの学校(小学校、中等教育学校、大学)に在籍する児童・生徒・学生が参加した。参加者による作文とアンケート回答データは、以下の5期に分けて収集された(表1)。

表1 データの収集プロセス

段階	時期	収集したデータ
1期	2022年1~2月	小3~小6の作文収集
2期	2022年4~5月	中1~高3の作文・アンケート収集
3期	2022年6月	小1~小2の作文収集
4期	2022年6~7月	小3~小5(新小4~小6)のアンケート収集
5期	2022年7月	大1の作文・アンケート収集

研究倫理審査において承認(次節参照)が得られた時期の関係で、小学校3~6年生のデータを先行収集し、以後、中学生・高校生データ、小学校1~2年生データ、大学生データの順で収集を進めた。小学校3~5年生については、作文収集とアンケート調査を別個に実施したが、そのほかの学年では両者を同時に実施した。

2.3 研究倫理対応

とくに、成年に達していない児童・生徒の作文データを公開目的で収集する際には、研究倫理面でも慎重な対応が求められる。この点をふまえ、小中高生からのデータ収集について

は、前節で示した1～4期のそれぞれにおいて、研究計画（コーパスの構築と公開を目的としていることを明記）を法人の設置する研究内容審査委員会に提出し、「人を直接の対象とする研究の審査」承認を受けた。

また、実際のデータ収集にあたっては、小中高生の保護者向けに、文書でプロジェクトの趣旨を説明し、以下の点を示して同意を求めた。表2は第3期に使用した依頼フォームの一部である。

表2 小中高生保護者向け依頼文での説明内容

本調査の参加にかかる確認事項
(1) 本調査への参加は任意です。
(2) 本調査への参加の有無や、提出物の内容が学校での評価などに影響することは一切ありません。
(3) この用紙（裏面）に記名の上、作文を提出いただくことで、本調査への参加と、作文データの公開にご同意いただいたものと取り扱わせていただきます。
(4) 作文を提出した後で、それを取り消したくなった場合は、研究代表者に連絡してください。提出済みの作文を公開データベースから取り除く処理を行います。
(5) 収集した用紙は電子的にスキャンし、書き起こしを済ませた後、適切な方法で廃棄します。作文本文は電子的データベースとして公開されます。

大学生については、全員が成年（18歳）に達していることから、上記と同様の趣旨を文書で学生本人にのみ示し、同意を求めた。

小中高生の参加者には一切の報酬を与えておらず、ボランティアとしての参加となっている。そのため、学年間で参加者数に差が生じている。一方、大学生参加者には、仕事の内容と少額の報酬を提示した上で参加者を募った。

2.4 参加者

参加者の内訳は以下のとおりである。中等教育学校において生徒の学年は1～6年と呼称されているが、以下では、一般的な校種区分に従い、中学校1～3（中1～中3）年および高等学校1～3年（高1～高3）と記載している。また、データ整理の必要上、すべての学年（grade: G）を通してG01～G13という学年コードを設定した。なお、ストーリーライティングは後述のように2つの課題からなるが、一方しか回答していないデータはコーパスから削除している。これにより、最終的な参加者人数は700名となった（表3）。

表3 JASWRIC 参加者人数

学年	人数	作文数	学年	人数	作文数
G01 (小1)	37	74	G08 (中2)	112	224
G02 (小2)	43	86	G09 (中3)	45	90
G03 (小3)	35	70	G10 (高1)	90	180
G04 (小4)	27	54	G11 (高2)	86	172
G05 (小5)	58	116	G12 (高3)	30	60
G06 (小6)	62	124	G13 (大1)	53	106
G07 (中1)	22	44	以上合計	700	1,400

前述のように、小学校3～6年生の作文は2022年1～2月に先行収集され、そのほかの参加者の作文は、2022年の4～7月に収集された。このため、小学校3～6年生については当該年度の後半で、そのほかの学年では当該年度の前半でデータが収集されたことになる。年度内での執筆時期の数カ月のずれが産出に決定的な影響を与えることは少ないと思われる

が、13年間の発達過程を精密な計量的モデルとして議論する場合には、たとえば、G1、G2、G3.5、G4.5、G5.5、G6.5、G7、G8...のように、小学校3～6年生のデータのみ年次数を割り増して分析することも考えられるだろう。

2.5 作文の収集

作文コーパスでは、様々なタイプの作文が収集対象にされるが、今回のプロジェクトでは、2.1節で述べた目的をふまえ、I-JASで実施された2種のストーリーライティングタスク(「鍵」をテーマにした4枚の連続イラストに基づく作文と、「ピクニック」をテーマにした5枚の連続イラストに基づく作文)を実施することとした。自由作文に比べ、絵描写文は内容的なぶれがきわめて小さく、対照研究に適した資料と言える。I-JASの2種のイラストの使用については、コーパス開発者の迫田久美子氏より事前に許諾を得た。

I-JASでは、同じイラストについて、先にストーリーテリング(口頭描写)を行ってからストーリーライティング(コンピュータ上での作文)を行う手順になっているが、JASWRICでは、作文タスクのみを実施している。

I-JASのオリジナルのイラストでは、キーワードが日本語・平仮名・英訳で示されている。「鍵」については、鍵(かぎ/key)、ケン(けん)、マリ(まり)、梯子(はしご/ladder)、警官(けいかん/police man)の5語が、「ピクニック」については、ピクニック(ぴくにっく/picnic)、ケン(けん)、マリ(まり)、バスケット(ばすけっと/basket)、犬(いぬ/dog)、地図(ちず/map)の6語がキーワードである。今回のプロジェクトでは、英訳はすべて除去した。そのうえで、小学校1～2年生用のプロンプトでは、キーワードの漢字は「犬」を除いてすべて平仮名で表記し、「犬」は漢字に平仮名を添え、カタカナ語はカタカナに平仮名を添えた。小学校3年生～大学1年生用のプロンプトでは、英訳を除いたほかは、オリジナルのキーワード表記をそのまま使用した。

以下の図1-2は、小学校3年生以上の参加者に配布したプロンプトである。I-JASのオリジナルの調査デザインに合わせ、書き出しの指定文(漢字使用は学年によって調整)を作文欄の冒頭に明示し、続きを書かせるようにした。

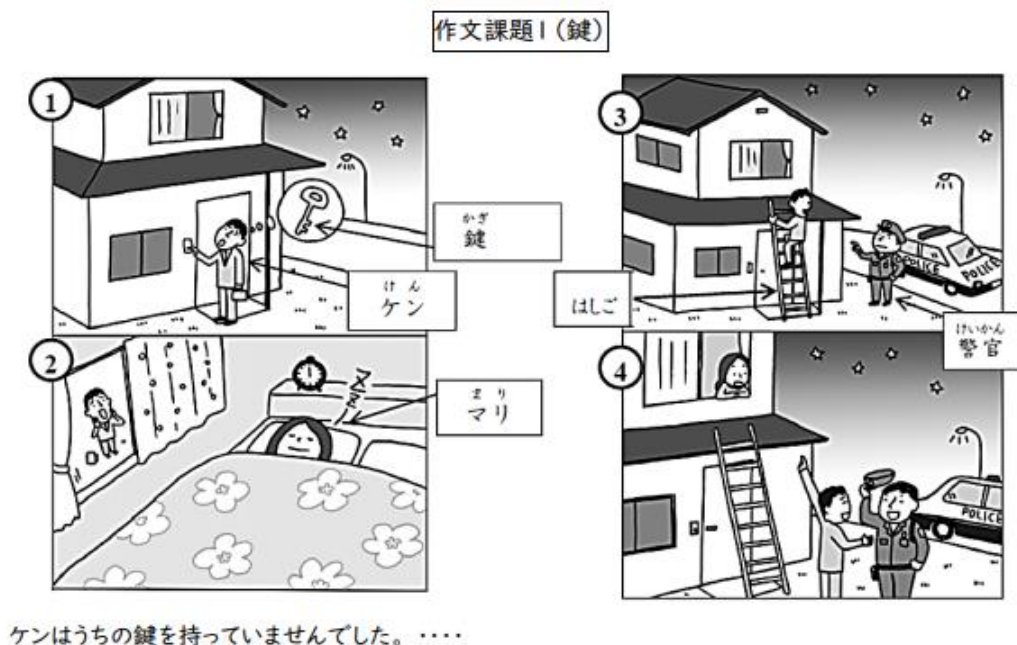
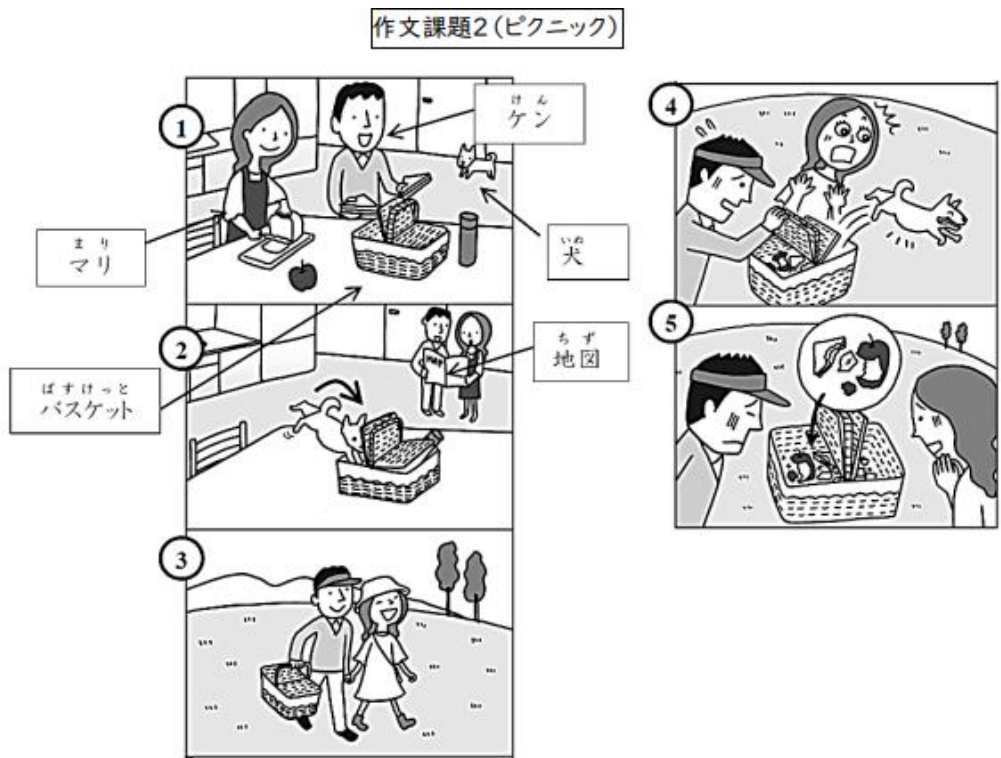


図1 ストーリーライティング用のプロンプト(「鍵」)



朝、ケンとマリはサンドイッチを作りました。…

図2 ストーリーライティング用のプロンプト（「ピクニック」）

以下は、学年グループ別の課題指示文である。小学校3年生～高校3年生用では、すべての漢字にフリガナを付けている（表4）。

表4 参加者向け課題指示文

学年	指示文
小1～2	えをよくみてください。 (1～4/1～5 のじゅんばんにならんでいます)。かきだしのことばにつづくよう、4つ (5つ) のえがあらわすストーリー (おはなし) をかいてください。
小3～高3	絵 [え] をよく見 [み] てください。書 [か] き出 [だ] しのことばに続 [つづ] くよう、4つ (5つ) の絵 [え] が表 [あらわ] すストーリー (お話「はな」し) を書 [か] いてください。
大1	下記の2種の絵をよく見てください。書き出しの言葉に続くよう、①～④ (または①～⑤) の絵が表すストーリー (お話) を書いてください。

執筆方法に関して、小学校1年生～高校3年生では、教員を通じて用紙を配布し、用紙内の指定箇所に筆記具で記入するよう指示した。大学生は、プロンプトを電子ファイルで配布し、作文はオンラインのサイト上に記入するよう指示した。いずれの場合も、字数制限と時間制限は設けず、平仮名の使用は問題ないと示した。また、参加者本人が独力で作文するよう示した。小中高生調査では、児童生徒への指示に加え、保護者向けの依頼文中にも注意事項を明記した（表5）。

表5 小中高生保護者向け依頼文での課題指示 (小1~2 保護者用)

保護者用の課題指示
作文の執筆にあたって、習っていない漢字はひらがなで書いていただいで問題ございません。時間制限は設けておりませんが、所要時間はおよそ5~10分程度と予想されます。なお、学年間の比較が調査の趣旨でありますことから、児童ご本人以外の方のお手伝いはお控えいただきまして、ご本人が独力で書いてくださるようお願いを申し上げます。

2.6 アンケートデータの収集

参加者の属性データとして、小学校6年生を除く全参加者に対して、「本を読むことが好きですか?」および「文を書くことが好きですか?」という2つの質問を与え(質問文中の漢字使用は学年によって調整している)、5段階での回答データを収集した。小中高生については、客観的な学力データの収集を行っていないが、2つの質問によって、読書や作文へのなじみ度は推定可能である。

小学校1~2年生および中学生~大学生においては、2項目のアンケートを作文用紙に記載し、作文と同時に提出させた。ただし、前述のように、先行してデータ収集を行った小学校3~6年生については、作文時にアンケートを実施していなかったため、作文収集の4~5カ月後に、アンケートのみを別個に実施した。アンケート実施時点で学年が進行していたため、一定数が法人設置学校以外に進学した6年生についてはアンケートデータが取られていない。

また、大学生の参加者には、上記のアンケートに加え、2022年1月に実施された大学入学共通テストにおける国語(現代文)のスコアレンジ(100点満点中10点刻み)を申告させた。参加者の性別は、産出への影響もある程度予想される場所であるが、性多様性への配慮の観点からあえて尋ねていない。

2.7 作文の文字化

小中高生の作文は手書きでなされたため、作文用紙をスキャンして専門業者に回送し、書きおこしを依頼した。書きおこしの質にぶれが出ないように、作業者は20年以上の作業経験を有する専業者1名のみ限定した。書きおこしにあたっては、以下のルールを定め、作業者に指示した(表6)。

表6 書きおこし作業マニュアル

書きおこし作業指示 (中高生データ処理の指示書の一部)
(1) 入力範囲 <ul style="list-style-type: none"> 生徒が手書きで書いた作文本体(※欄外のメモや、注記などはカット:原則として青でマーク済) 書き出しの印刷された指示文は入力不要(ケンはずちのカギを持っていませんでした/朝、ケンとマリはサンドイッチを作りました) コマ番号を示す①、②などの番号を生徒が入れている場合、それらはカットして入力
(2) 生徒による修正の扱い <ul style="list-style-type: none"> 二重線や塗りつぶしではっきり消しているものは消されたものとみなして入力しない 消して書き直しているものは、書き直したほうを入力する 一度文を書いてから、山形マークなどで別要素を挿入しているものは挿入されたものとして入力。山形記号の反映は不要 私はずちから追記:イヌと>遊びに行った → 入力例: 私はイヌと遊びに行った
(3) 誤字・読みにくい文字の扱い <ul style="list-style-type: none"> 軽微な誤字や悪筆(点が足りない、線が一本多いなど)は、常識的に読み替えて入力(見る→見る)

- ・明確な誤字は【誤】タグをつけてそのまま入力、（大無し→【誤】大無し）
- (4) 記号・特殊表記
 - ・記号もそのまま入力（！？など、すべて全角で。！については直線・斜線・太い細いなどの区別があるが一律で！でOK）
 - ・「おま w」（おまえ、の意味のネット用語）などもそのまま入力
 - ・本来は「」が必要なセリフに、「」記号がない場合が散見されるが、ママで入力
- (5) 段落扱い
 - ・明確な改行がない場合は続けて入力
 - ・明確に行替えがある場合は改段として処理

以下は、小学校1年生の児童作文の現物と、入力後のデータの一例である（図3）。

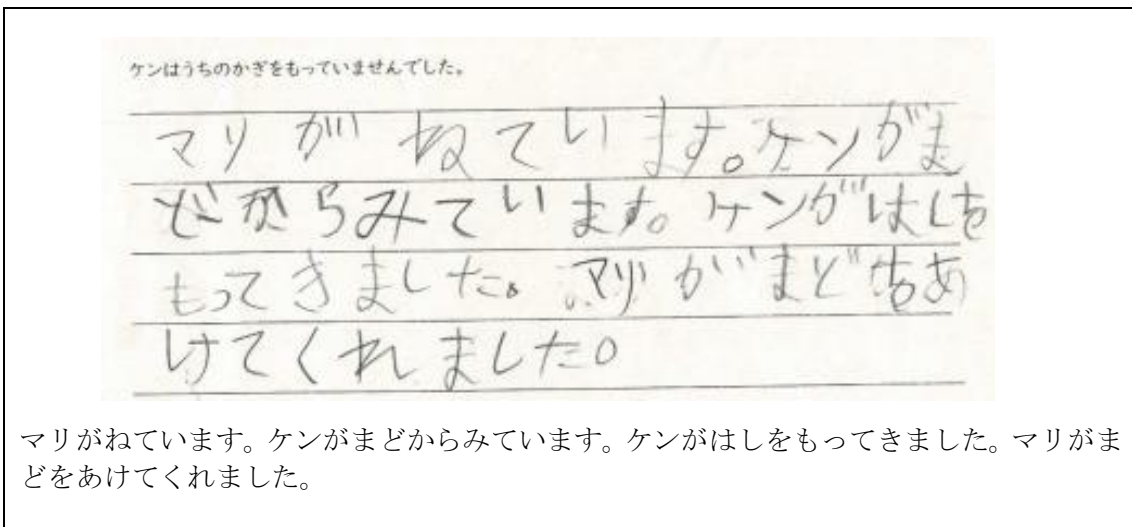


図3 小学校1年生の「鍵」作文の児童手書き原稿と書きおこしデータ（G01_Key_009）

2.8 データファイルの作成

個々の作文ファイルはUTF-8（BOMあり）形式で保存し、以下の基準で命名を行った（表7）。

表7 ファイル命名ルールと実際の例

凡例	例
学年コード_トピックコード_参加者連番.txt	(1) G01_Key_032.txt（小1、032児童の「鍵」作文） (2) G10_Pic_006.txt（高1、006番生徒の「ピクニック」作文）

2.9 テキスト校閲と形態素解析

形態素解析には、国立国語研究所が提供する「Web茶まめ」を使用した。前処理として、「半角・全角変換」と「数字処理」オプションを追加し、辞書は現代語を使用した。当初、素起こしのテキストデータに対して形態素解析を試みたが、とくに小学生低学年の児童はほぼすべてを平仮名で表記しており、かつ、句読点を使うことが少ないため、解析精度は著しく低かった。

そこで、解析の精度を上げるため、全データに対して、筆者が手作業で校閲を加えた。以下は、実際に行った校閲のタイプ別事例である（表8）。

表 8 元テキストに対する校閲の例

校閲タイプ	元テキスト	校閲後テキスト
(1) 平仮名の漢字化	<ul style="list-style-type: none"> ・ねむってたられかがよんできた ・通ほうされて… ・と中 ・じゅう人 	<p>眠ってた<u>ら誰</u>かが呼んできた</p> <ul style="list-style-type: none"> ・通報されて… ・途中 ・住人
(2) 平仮名のカタカナ化	<ul style="list-style-type: none"> ・ちやいむをならしました… ・ぴくにつく ・まりはねてたからけんは… ・ぐーすかぴーすかねています 	<ul style="list-style-type: none"> ・チャイムをならしました… ・ピクニック ・マリは寝てたからケンは… ・グースカピースカ寝ています
(3) 文節区切りスペース挿入	<ul style="list-style-type: none"> ・さんどいっちがありませんでしたおなかぺこぺこで… 	<ul style="list-style-type: none"> ・サンドイッチがありませんでした [スペース] お腹がぺこぺこで…
(4) エラー修正	<ul style="list-style-type: none"> ・きずきません ・事【誤】 状を説明すると… 	<ul style="list-style-type: none"> ・気づきません ・事情を説明すると…
(5) 段落削除	<ul style="list-style-type: none"> ・バスケットの中から犬がとび出てきました。【改段】マリはとてもおどろきました。【改段】2人がバスケットの中を見ると… 	<ul style="list-style-type: none"> ・バスケットの中から犬がとび出てきました。マリはとてもおどろきました。2人がバスケットの中を見ると…

(1) については、漢字で書くのがふつうであると判断した箇所と、平仮名で書いた場合に異なる解釈の余地が残ると判断した箇所（あける→開ける／空ける）を対象とした。とくに、「通ほう」「と中」「じゅう人」など、部分的に漢字表記されているものについては、そのままでは解析が失敗することが多いため、全体を漢字で表記するようにした。

(2)については、カタカナで書くのがふつうであると判断した箇所のほか、特殊なオノマトペと判断されるものについても地の文との切れ目を示すため、カタカナ表記に変更した。

(3)については、使用した句読点の数などを調査対象にする研究が想定されることから、句読点を校閲で追加することは行わず、スペースを補って、語の切れ目を示した。なお、補ったスペースは形態素解析後、再度消去している。

(4)については、書きおこし作業者が【誤】タグを付与した箇所のほか、MS Wordの校閲でエラーと指摘された箇所を中心に修正を行った。複数の参加者に共通して見られたエラーとしては、以下のようなものがある（表9）。

表 9 参加者による特徴的なエラー

タイプ	元テキスト	校閲後テキスト
(1) 「は」／「わ」の混同	・けんわ	・ケン <u>は</u>
(2) 「を」／「お」の混同	・はしごお	・はしご <u>を</u>
(3) 促音の不足・過剰挿入・挿入場所誤り	<ul style="list-style-type: none"> ・のぼて ・さんどいっち ・はいた 	<ul style="list-style-type: none"> ・のぼ<u>っ</u>て（登って） ・サンドイッチ ・は<u>い</u>った（入った）
(4) 四つ仮名の誤り	・きずいて	・気 <u>づ</u> いて
(5) 漢字の誤り	<ul style="list-style-type: none"> ・事状 ・不信（者） ・窓が空いている 	<ul style="list-style-type: none"> ・事情 ・不審（者） ・窓が<u>開</u>いている

筆者による以上の校閲は、後で検証・追加できるよう、すべて、Wordの校閲機能を用い

て記録した（図4）。

G ケンは、かぎ鍵がないので、チャイムちやいむをならしましたが、マリは、でて出てく
れませんがした。マリは、ねて寝ているので、きこえ聞こえませんでした。ケンは、近所き
かみから、はしごをもってきてしまいました。けいさつ警察の、サイレンきいれんに、マ
リは、おき起きてケンの、ほうをみたら、ケンが、マリのほうをゆびさして、びっくりしま
した。←

G ねかつて眠つてたらだれ誰かがよんできた。はしごでうえ上にのぼってたらけいさつか
ん警察官におこられ怒られた。けいさつかん警察官はゆるしてくれた。←

G ケンはおおごえ大声できけ叫びました。けれどマリには声が届きこえがときませんで
した。それでケンははしごをかけてのぼろうとしました。そのときけいさつかん警官にみつか
ってしまいました。ケンはわけ訳をはな話しました。どうやらけいさつかん警官がゆるしてくれ
ました。そのうえ上からはマリがみて見えています。←

G ケンは、うちにはいれませんが、ケンはマリにいれて はいれましたが、マリは
ねて寝ていてきづき気付きません。ケンは考えてかんがえて、はしごをもってきてマリのち
や部屋へのぼろうとします。けいさつかん警官がきて、どろぼう泥棒だとおも思われてしまいま
す。さゆう理由をはな話していたら、マリがおき起きてドアをあけ開けてくれました。←

G マリがおうちでねて寝てるとけいさつケンがかえつて帰つてきました。かぎ鍵をわすれ忘れ
たので、けいさつ警察に叱られてちかられてしまいました。これはわたしのちや家です。と
いったらけいさつ警察があやま謝ってくれました。←

図4 筆者による元データへの校閲追加の例

全 1,400 作文に対する校閲箇所は、延べで 9,090 箇所であった（4,631 箇所の挿入と、4,459 箇所の削除）。これらの校閲は、筆者が目視で気づいた範囲で、かつ、解析精度を高めるといふ目的に即して行ったもので、参加者の作文に見られる問題点を網羅的に検出・修正する性質のものではない。

校閲により、解析精度は、完全ではないものの、かなり上昇した。たとえば、上記にある「ちやいむ」「さいれん」「けいさつかん」の部分について言うと、平仮名表記の元テキストではそれぞれ「ちやう／無」「然（さ）／入れる」「警察／館」と誤解析されていたが、校閲テキストではいずれも正しい解析がなされている。V1.0 の公開に先立ち、人名である「ケン」と「マリ」の品詞情報の不統一のみ修正したが、そのほかは、自動解析結果のまま、手作業による修正は加えていない。

2.1.0 データの概要

以上の手順を経て、公開用のデータが完成した。表 10 は JASWRIC (V1.0) に含まれる作文の総語数（短単位解析による形態素の数）である。

「鍵」作文の総語数は約 72,000 語、「ピクニック」作文の総語数は約 65,000 語、合計約 136,000 語となる。近年の大規模なコーパスから見れば、必ずしも大きな数字ではないが、統制性の高い共通の絵描写ストーリーライティングであることを考えれば、個体差を超えた全体傾向を議論するに足る量のデータが集められたと言える。なお、今後、形態素解析結果の修正がなされた場合、語数は若干増減する可能性がある。

表 10 JASWRIC V1.0 の収録語数データ

Grade	人数	総語数			1人あたり語数		
		Key	Pic	All	Key	Pic	All
G01	37	2328	2446	4774	62.9	66.1	129
G02	43	3084	2683	5767	71.7	62.4	134.1
G03	35	3884	3461	7345	111	98.9	209.9
G04	27	3023	2742	5765	112	101.6	213.5
G05	58	6045	5611	11656	104.2	96.7	201
G06	62	6080	5451	11531	98.1	87.9	186
G07	22	2839	2158	4997	129	98.1	227.1
G08	112	10499	9597	20096	93.7	85.7	179.4
G09	45	4623	4323	8946	102.7	96.1	198.8
G10	90	10811	9343	20154	120.1	103.8	223.9
G11	86	9597	8561	18158	111.6	99.5	211.1
G12	30	3545	2980	6525	118.2	99.3	217.5
G13	53	5700	5221	10921	107.5	98.5	206.1
All	700	72058	64577	136635	102.9	92.3	195.2

3. JASWRIC の公開

3.1 ダウンロード版

JASWRIC (V1.0) は、2022年8月よりダウンロード版としての公開を開始した。ダウンロードデータには、Raw Data と Edited Data の2つのサブフォルダが含まれる。

前者には、個別作文の素起こしテキストファイル (txt) 1,400種、学年別・トピック別にテキストファイルを集約したマージテキストファイル (txt) 26種、また、小中高生 647名による手書き作文のスキャン画像 (jpg) 1,294枚 (図5) が含まれる。スキャン画像のデータは、子どもの文字の獲得研究などにも応用可能であろう。

後者には、形態素解析用に校閲を加えた文書ファイル (docx) が含まれる。文書ファイルには、Word の校閲機能を用い、素起こしのテキストに加えた編集履歴がすべて記録されている。

このほか、形態素解析済みデータ (tagged)、参加者属性シート (Participant Survey)、参考文書 (readme) が同梱されている。

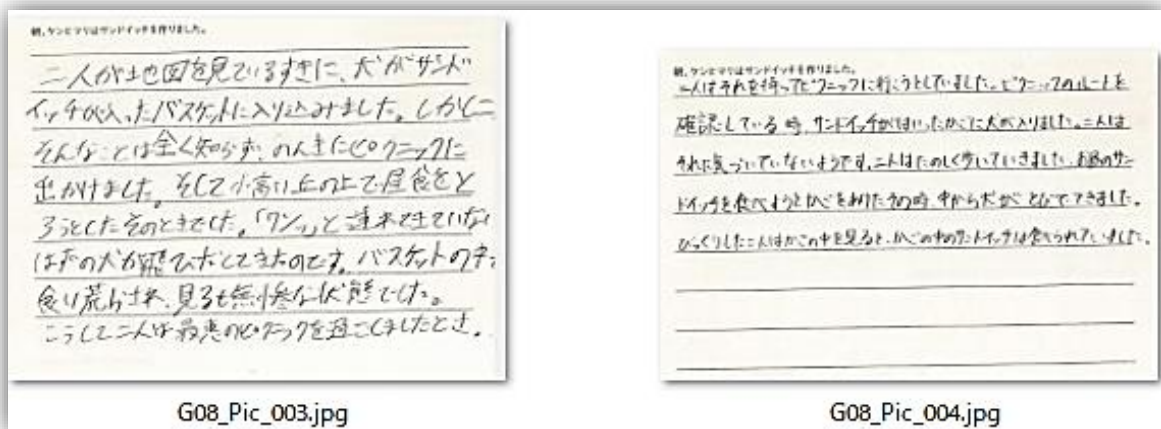


図5 手書き作文の画像ファイル (中2生徒の作文の一部)

3.2 オンライン検索版

3.2.1 概要

JASWRIC は、筆者の研究室でこれまでに開発した各種コーパスのオンライン統合検索サイト (Kobe Univ. Ishikawa Lab. Online Corpus Retrieval System) 上でも公開されている (図 6)。

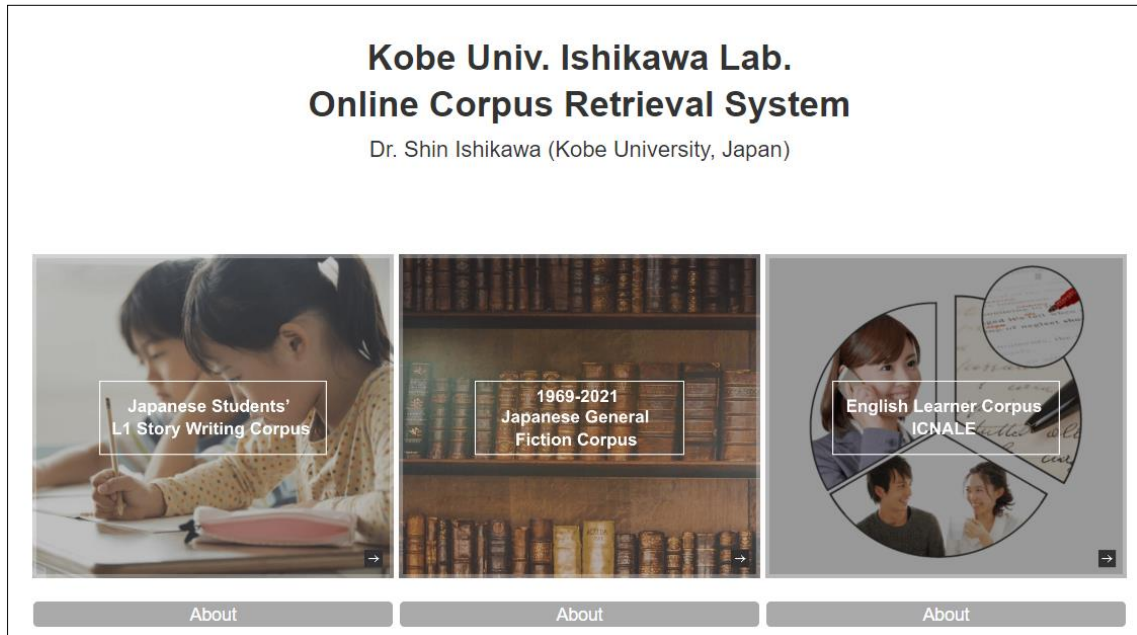


図 6 Kobe Univ. Ishikawa Lab. Online Corpus Retrieval System トップページ
(左から順に、JASWRIC、日本語小説コーパス JFIC、英語学習者コーパス ICNALE)

留意すべきは、オンライン版は、語彙研究や品詞研究などでの利用を念頭に置いているため、校閲作業後のデータを使っていることである。つまり、低学年児童の作文中の平仮名などは適宜漢字に直されており、このほか、大きなエラーなども修正されている。このため、漢字使用率や、誤字・誤記などを主たる関心対象として研究を行いたい場合は、前述のダウンロード版に含まれる生データを、直接、分析するほうが便利であろう。ただし、後述するように、オンライン版であっても、検索結果から、手書き作文の画像ファイルへのアクセスが提供される設計となっている。

オンライン版 (JASWRIC *Online*) では、子どもの日本語作文の多角的な分析が可能になるよう、既存コーパス用に研究室で開発した検索システムを内部的に改修して運用している。基本的な検索タイプは共通化されており、(1)KWIC 検索 (検索対象語を中央に、文脈を左右に配置して一覧表示)、(2)共起語検索 (当該語の左右 3 語範囲で位置ごとの高頻度共起語を表示、統計値は頻度・ t スコア・対数尤度比・相互情報量から選択可能)、(3)語彙リスト作成 (書字形 [=表層形] または語彙素に基づく語彙頻度表を作成・出力)、(4)頻度グラフ作成 (検索対象語の学年別頻度変化をグラフで表示、頻度は総語数で自動的に調整頻度で出力)、(5)特徴語検索 (ターゲットデータとレファレンスデータに内包されるすべての語の頻度を比較して、ターゲット側で有意に多い、または少ない語をキーワードとして一覧で表示。統計値はカイ二乗値と対数尤度比から選択可) の 5 種の機能を提供する。

図 6 の左端の写真をクリックすると、ユーザー登録の画面に遷移し、必要事項を記入すると、図 7 に示す、KWIC 検索用の検索条件指定画面が表示される (以下、本稿の画像は開発版のシステムから取ったもので、運用版では若干の変更が生じる可能性がある。また、実際の画面の一部を加工して表示している場合がある)。

KWIC	共起語	語彙リスト
検索語	<input type="text"/> ? <input type="button" value="POS"/>	
学年	<input checked="" type="checkbox"/> 小学校 <input checked="" type="checkbox"/> 中学校 <input checked="" type="checkbox"/> 高校 <input checked="" type="checkbox"/> 大学 <input checked="" type="checkbox"/> 1年 <input checked="" type="checkbox"/> 1年 <input checked="" type="checkbox"/> 1年 <input checked="" type="checkbox"/> 1年 <input checked="" type="checkbox"/> 2年 <input checked="" type="checkbox"/> 2年 <input checked="" type="checkbox"/> 2年 <input checked="" type="checkbox"/> 3年 <input checked="" type="checkbox"/> 3年 <input checked="" type="checkbox"/> 3年 <input checked="" type="checkbox"/> 4年 <input checked="" type="checkbox"/> 5年 <input checked="" type="checkbox"/> 6年	
トピック	<input checked="" type="checkbox"/> 鍵 <input checked="" type="checkbox"/> ピクニック	

図 7 KWIC 検索の検索条件指定画面 (一部)

上部にあるタブを押すことで、5つの検索機能を切り替えることが可能で、選択中の機能は赤で表示される。KWIC 検索の条件設定画面について言うと、まず、1行目にある「検索語」に調べたい語（または語の連鎖）を入力する。JASWRIC Online では、上記の「検索語」を含め、コーパス分析の基本概念や基本技術に関わる用語にはリンクが用意されており、リンクを押すことで、解説などが表示される。次に、2行目の「学年」で、調査したい学年を設定する。上部にある「小学校」「中学校」「高校」「大学」をクリックすると、それぞれ、自動的に、下部にある全学年が指定されるが、必要に応じてその一部を検索対象外にすることもできる（たとえば、小学校の低学年にあたる1~2年生のみを調べるなど）。最後に、3行目の「トピック」で、ストーリーライティングにおける2種類の課題のうち、調査対象としたいものを、一方または両方とも、指定する。

3.2.2 検索語の入力方法

どの検索を行う場合であっても、検索語の入力方法は同一であるが、ここでは、KWIC 検索に即して、検索語の入力の流れを確認しておきたい。

まず、検索は、書字形の単位で行われる。ゆえに、検索ボックスに「犬」と入れれば「犬」を含む用例が、「する」を入れれば「する」を含む用例が抽出される。ただし、語彙素をまとめているわけではないので、「する」を入れても、その活用形（「さ(れる)」「し(ない)」「すれ(ば)」など）は抽出されない。

活用形を同時に検索したい場合は、縦棒 (|) を使って OR 検索を行う。たとえば、「さ|し|する|すれ|しろ」と入力すれば、「する」とその活用形を同時に検索することができる。

JASWRIC Online は、書字形を、形態素（短単位）単位で区切ったデータを元データとしているため、連語や複合形を検索する際は、形態素の切れ目ごとに、半角のスペース（以下では[sp]で表示）を挟む必要がある。たとえば、「見ている」という表現を探すには、「見[sp]て[sp]いる」と入力し、「見ていた」を探すには「見[sp]て[sp]い[sp]た」と入力する。、自分の探したい表現がどのように区切られるかわからない場合は、「Web 茶まめ」(2.9 節参照) に当該表現を入力し、どのように区切られるか確認すると便利だろう。

品詞情報を検索条件に加えることもできる。検索ボックスの右側にある[POS]を押すと、

以下の品詞選択画面がポップアップ表示される（図 8）。



図 8 品詞指定画面（大区分に名詞を選んだ場合）

品詞選択画面は上段（大区分）と下段（小区分）の二階層構造になっており、上部（赤色で表示）で品詞の大区分を指定すると、下部（グレーで表示）に当該品詞の小区分が表示される仕様となっている。下段に表示されるボタンのいずれかを押すと、当該品詞を示すコードが自動的に検索ボックスに入力される。なお、[ALL]とは全小区分の同時指定のことである。

たとえば、上図で示されている名詞を例にして言えば、小区分のうち、固有名詞を選べば絵課題の登場人物名である「ケン」や「マリ」が、普通名詞を選べば「鍵」や「チャイム」が、数詞を選べば「二（階）」などが抽出される。品詞と書字形を連語として指定することもできる。たとえば、「[名詞] は」と入れると、名詞の後に「は」が後続する例が抽出される。

検索語の入力にあたって、注意すべき点は、異表記が別語扱いになることである。たとえば、「鍵」と「カギ」は表記が異なるため、どちらか一方を入力した場合、他方は出てこない。両方を検索するには、OR 検索を用いて「鍵|カギ」のように入力する必要がある。なお、OR 検索は、連語の一部に対しても使える。たとえば、「ケン|マリ[sp]は|が|の|を|に」と入力すれば、「ケンは」「マリが」「ケンの」など、「ケン」または「マリ」の直後に主要な助詞が後続する例を同時に検索することができる。

以下は、主な検索ルールを一覧にしたものである（表 11）。

表 11 JASWRIC Online 上での検索指定の例

検索タイプ	検索指定の例	出力例
通常	鍵	鍵
OR 検索指定	鍵 カギ	鍵、カギ
連語・複合形指定	し[sp]て[sp]いる	…している
品詞指定（大区分）	[名詞]	鍵、チャイム、梯子…
品詞指定（小区分）	[名詞-固有名詞]	ケン、マリ、日本、ポチ…
品詞+書字形指定	[名詞][sp]は	鍵は、チャイムは、梯子は…
OR 検索組み合わせ指定	ケン マリ[sp]は が	ケンは、ケンが、マリは、マリが

コーパス分析に詳しくないユーザーの場合、検索語の入力ルールはやや複雑に思えるかもしれないが、検索ボックス右側の[?]ボタンを押せば、用例付きの解説をオンラインで読むことができる。なお、本システムの検索精度は、データベースの元になっている形態素解析の精度に制約されており、完全なものではない。

3.2.3 KWIC 検索

以下、JASWRIC Online で提供されている各種の検索機能について概説する。まず、KWIC 検索は、関心対象語の文脈内での振る舞いを調べるために用いられる。図 9 は、全データを対象に「入っ[sp]た」を KWIC 検索した際の結果画面（一部、以下同）である。

Sorting : 1st key <input type="text"/> 2nd key <input type="text"/> 3rd key <input type="text"/>				Sort	
	二人は、地図	バスケットに	入った	ことには、行き	G03_012_Pic
	、サンドイッチ	が	入った	バスケットをも	G03_012_Pic
	て、	の中	入った	ことに	G03_017_Pic
	決めまし	の食べ物	入った	バスケットに入	G03_024_Pic

図 9 KWIC 検索の結果画面の一部（検索語：「入っ た」）

これらの各行をコンコーダンスラインと呼ぶ。コンコーダンスラインの中央には検索対象語が表示され、左右に一定の長さの文脈が表示される。無作為な順序で並んだコンコーダンスラインから傾向を読み取ることはむづかしいが、たとえば、「入った」の左隣ないし右隣に来る共起語の 50 音順で全体の行を並べ替えると（ソーティング）、同じ共起語が続けて並ぶこととなり、全体を概観することで、検索対象語がどのような語と一緒に生起しやすいかを直観的に読み取ることができる。

ソーティングの基準は、上記の画面上部の 3 つのキーで指定する。たとえば、第 1 キーを L1（左隣 1 語目）、第 2 キーを L2（同 2 語目）、第 3 キーを L2（同 3 語目）と指定すれば、「入った」の左隣 1～3 語目を基準として並べ替えが行われる。

また、各行の右端には参加者コードが表示されている。たとえば、上記の 1 行目の G03_012_Pic は、小学校 3 年生 (G03)、学年内コード 12 番の児童で、これがピクニック (Pic) 作文の一部であることがわかる。ダウンロード版に含まれる参加者メタデータを参照すれば、当該の児童が、読書と作文に対してそれぞれどの程度の親しみを感じているか、アンケート結果を確認することができる。

各行の左端にあるアイコンは、手書き作文の画像ファイルへのリンクを示す。たとえば、1 行目左端のアイコンを押すと、以下の画面がポップアップ表示される（図 10）。

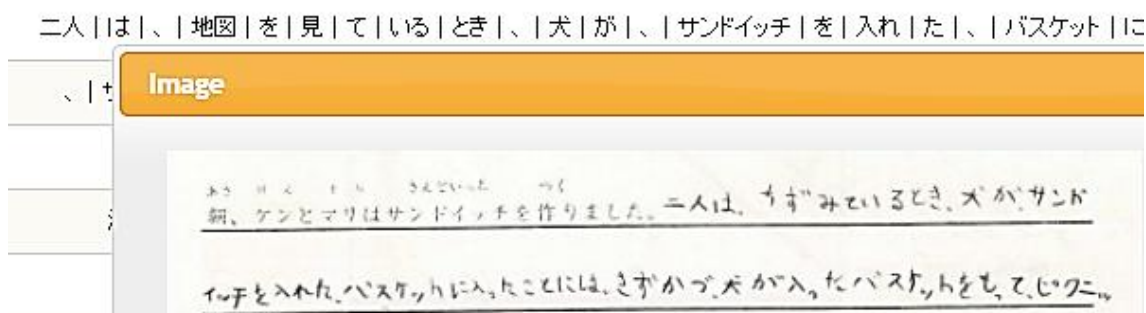


図 10 KWIC 検索結果画面で手書き画像を表示させた場合の例

こうした手書きデータを見ることで、実際に参加者が書いた作文を確認しながら分析を進めることができる。なお、大学生についてはワープロ上で作文したため、手書き画像データへのリンクはない。

3.2.4 共起語検索

共起語検索とは、検索対象語が出現しやすい近傍部の語彙的環境を明らかにする目的でなされる。以下は、全データを対象に「する」の共起語を検索した際の結果画面である（図11）。

		Raw Frequency		t-score		Log-Likelihood		Mutual Information	
		L1	0	R1	R2	R3			
78 (5484)	と	90 (4327)	する 475	と	382 (4327)	警官	53 (930)	が	55 (46)
26 (322)	説明	55 (240)		場所	17 (141)	に	31 (5913)	警官	48 (9)
25 (276)	を	36 (5484)		こと	15 (701)	マリ	15 (2596)	は	27 (41)
23 (810)	確認	28 (207)		ため	12 (231)	警察	13 (226)	に	23 (59)
15 (5913)	そう	8 (206)		公園	5 (111)	犬	10 (1596)	マリ	17 (25)
11 (8117)	ピクニック	6 (810)		の	4 (4061)	中	9 (1205)	犬	17 (15)
8 (105)	た	5 (8117)		ケン	3 (2132)	を	9 (5484)	から	14 (10)

図 11 共起語分析結果画面の一部（検索語：「する」）

検索結果の上部には、4種の統計量の切り替えボタンがあり、選択中の統計量が赤色で表示されている。デフォルトでは、粗頻度（raw frequency）の降順で各位置における高頻度共起語が示される。上記を見ると、左側要素との組み合わせでは「～とする」「～をする」「～をしようする」のほか、「説明する」「確認する」のような漢語動名詞が多いことがわかる。また、右側要素との組み合わせとしては、接続詞となる「すると」が最多で、そのほか、「～する場所」「～する公園」のような名詞修飾、また、「～すること」「～するため」といった抽象的名詞を含む連体表現が多いことも確認できる。

粗頻度は、共起を観察する際の一般的な基準だが、必要に応じて、*t* 統計量、対数尤度比、相互情報量に切り替えて結果を比較することができる。これらはいずれも共起の強さを示す尺度で、共起形の頻度を分子に、共起を構成する各要素の単独頻度を分母に置いて計算されるが、計算の詳細には差がある（石川、2021）。3種の統計量のうち、相互情報量（mutual information score）は、頻度の情報を対数化して圧縮するため、絶対的な頻度は低くても結び付きの強い組み合わせが上位に評価されやすい。たとえば、上記の条件で、統計量を相互情報量に切り替えると、L1 位置では、「プレゼント」などが上位に来る。この場合、「プレゼントする」という共起形の頻度は1例のみであるが、「プレゼント」の単独頻度も1例である。つまり、「プレゼント」という語はそもそも出現すること自体が極めてまれであるのに、出現する場合には必ず「する」を伴うものとして、上位に評価される。相互情報量は低頻度だがユニークな語の組み合わせを探したい時に有用である。

3.2.5 語彙リスト作成

語彙リスト作成は、指定した参加者群別、トピック別に、使用された語と頻度の一覧を調べる目的でなされる。この時、語の単位として、書字形または語彙素のいずれかを選ぶことができる。書字形を選べば表記や活用形は区別され、語彙素を選べばそれらは集約される。以下は、トピックを「鍵」に限定し、書字形単位でリスト作成を行った際の結果画面である（図12）。

Word ▲	Word ▼	Raw Frequency ▲	Raw Frequency ▼
て		4381	682.43
た		4185	651.89
を		2750	428.36
に		2706	421.51
まし		2401	374.00

図 12 語彙リストの一部（トピック：「鍵」、書字形単位）

結果画面の上部には、4つのボタンがあり、デフォルトでは粗頻度（raw frequency）による降順（▼）が選ばれているが、目的に応じて、昇順（▲）に切り替えたり、あるいは、語の50音順（word）による昇順や降順に切り替えたりすることも可能である。

リストを見ると、接続助詞の「て」や、過去助動詞の「た」、格助詞の「を」「に」、丁寧形助動詞「ます」の活用形である「まし」などがとくに多いことがわかる。個々の頻度は粗頻度のほか、100万語あたりの調整頻度（PMW）でも表示される。たとえば学年間で同一語の頻度を比較するといった目的であれば、調整頻度の参照が必要になる。

同じデータに対して、語彙素を単位として語彙リストを作成すると、3位に「ます」が入る。これは、書字形で5位であった「まし」の頻度に、「ませ（ん）」や「ます」の頻度が繰り込まれるためである。

3.2.6 頻度グラフ作成

頻度グラフ作成は、検索対象語の出現頻度の学年間変化を観察する目的でなされる。以下は、全データを対象として、「そして」の学年別頻度を調査した際の結果画面である（図 13）。

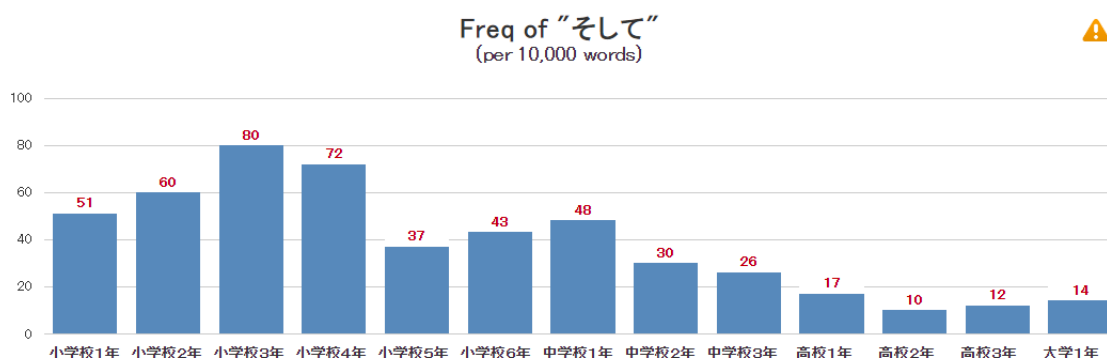


図 13 「そして」の学年別頻度変化（1万語あたり調整頻度）

若干のぶれはあるものの、「そして」の頻度は、およそ小学生で多く、中学生・高校生・大学生になるにつれて減少していく。詳細な分析は本稿の狙いを超えるが、今回の結果に限って言えば、低学年の子どもほど、「そして」を用いて、内容の連続性を明示的に強調しがちであるように思える。砂川（2020）は、L1 児童作文に「接続助詞などで節が連ねられ、だらだらと長く続いて脈絡が読み取りにくい文」が多いことを指摘しているが（4.2 節参照）、小学生による「そして」の多用も砂川の指摘とも呼応する。

3.2.7 特徴語検索

特徴語検索とは、対照群（レファレンス）と比較した場合に、調査対象群（ターゲット）の側で顕著に過剰ないし過少に使用される語を調べる目的でなされる。ターゲット側での過剰と過少は、それぞれ、レファレンス側での過少と過剰を示す。こうした研究では、一部対全体（例：「小1」対「小1～大1」）を比較することが多いが、性質の異なる2群に焦点を当て、一部対一部（例：「小1」対「大1」）を比較することもある。

ターゲットとレファレンスの指定は、検索条件設定パネルで平易に実行できる。以下は、ターゲットに小学校1～3年生を、レファレンスに高校生～大学生を指定し、トピックは両課題、語の単位は書字形とした場合の設定例である（図14）。

データ	調査群(target)	参照群(reference)
	<input checked="" type="checkbox"/> 小学校 <input checked="" type="checkbox"/> 1年 <input checked="" type="checkbox"/> 2年 <input checked="" type="checkbox"/> 3年 <input type="checkbox"/> 4年 <input type="checkbox"/> 5年 <input type="checkbox"/> 6年 <input type="checkbox"/> 中学校 <input type="checkbox"/> 1年 <input type="checkbox"/> 2年 <input type="checkbox"/> 3年 <input type="checkbox"/> 高校 <input type="checkbox"/> 1年 <input type="checkbox"/> 2年 <input type="checkbox"/> 3年 <input type="checkbox"/> 大学 <input type="checkbox"/> 1年	<input type="checkbox"/> 小学校 <input type="checkbox"/> 1年 <input type="checkbox"/> 2年 <input type="checkbox"/> 3年 <input type="checkbox"/> 中学校 <input type="checkbox"/> 1年 <input type="checkbox"/> 2年 <input type="checkbox"/> 3年 <input checked="" type="checkbox"/> 高校 <input checked="" type="checkbox"/> 1年 <input checked="" type="checkbox"/> 2年 <input checked="" type="checkbox"/> 3年 <input checked="" type="checkbox"/> 大学 <input checked="" type="checkbox"/> 1年
トピック	<input checked="" type="checkbox"/> 鍵 <input checked="" type="checkbox"/> ピクニック	<input checked="" type="checkbox"/> 鍵 <input checked="" type="checkbox"/> ピクニック
語の単位	<input checked="" type="radio"/> 書字形(表層形) <input type="radio"/> 語彙素	

図14 特徴語検索の検索条件指定画面

以下は、書字形（左側）および語彙素（右側）を単位として、高大生と比較した場合の小学校低学年児童の特徴語を抽出した結果である（図15）。

Chi2				Log-Likelihood			
過剰使用		過小使用		過剰使用		過小使用	
Word	Statistic	Word	Statistic	Word	Statistic	Word	Statistic
いい	121.71	の	81.18	言う	182.30	二人	52.62
たら	115.33	2人	70.32	そして	122.60	然し	35.83
そして	98.52	に	49.90	さん	74.34	二	32.50
いっ	66.70	しかし	42.67	御	62.03	階	30.12
よん	59.99	階	39.38	ます	52.16	為	27.51

図15 特徴語検索結果画面の一部（高校生～大学生に対する小学校1～3年生の特徴語）

検索結果の上部には、個々の語の特徴度、つまりは、レファレンスに対するターゲット側での過剰・過少度を示す2種の統計量の切り替えボタンがあり、選択中の統計量が赤色で表示される。デフォルトでは、カイ二乗統計量（Chi2）の降順で過剰使用語と過小使用語が表示されている。

書字形での分析によると、小学校低学年児童は、前述の順接接続詞「そして」のほか、仮定形助動詞の「たら」や、「言う」の活用形を過剰使用する一方、所有格・目的格を示す格助詞の「の」や「に」、逆接接続詞の「しかし」、また、登場人物を代名詞的に指し示す「2人」などを過剰使用することがわかる。また、語彙素での分析によると、低学年児童は、上記の傾向に加えて、丁寧さを示す接辞の「御～」、丁寧形語尾の「ます」、丁寧な人物呼称である「～さん」を過剰使用し、行為の目的を示す「為」などを過剰使用している。実際の研究では、こうして得られたキーワードを再度 KWIC 検索や共起語検索にかけ、これらの語の用法を詳しく確認していくことが必要である。なお、統計値については、対数尤度比の使用も可能である。対数尤度比は、対数を使用して頻度の情報を圧縮するため、一部対全体比較のように、ターゲットとレファレンス間でデータ量の差が大きい場合に使用が推奨される。

4 データの概要と研究展望

4.1 L1 日本語発達研究での利用

書き言葉の L1 日本語の発達研究に関しては、語・連語・文法項目などの使用状況の学年別変化の調査が想定される。以下に示すのは、JASWRIC に含まれる 13 学年分の「鍵」作文の中から、2 年間隔で、小 1、小 3、小 5、中 1、中 3、高 2、大 1 の作文サンプルを抽出した結果である（校閲なしの素起こしテキストを表示している）。

- (1) ケンは、かぎがないので、ちやいむをならしましたが、マリは、でてくれませんでした。マリは、ねているので、きこえませんでした。ケンは、きんじょから、はしごをもってきてしまいました。けいさつの、さいれんに、マリは、おきてケンの、ほうをみたら、ケンが、マリのほうをゆびさして、びっくりしました。(小 1、G01_001)
- (2) マリはよるの 12 じだからぐっすりねむりについていました。だからはしごでまどからはいろうとしたらけいかんがきてケンをちゅういしてケンがおえりてきてじじょうをはなしたらマリがおきてけいかんがなっとくしておわり。(小 3、G03_001)
- (3) 家のチャイムをならしてもだれも出てきません。「2 階でねているのか」と思ってマリをいくら呼んでも聞こえていません。しかたなくケンは庭のはしごを使って家のカーテンが開いている窓をたたこうとしました。「あともう少し」というところで警官にあやしまれ、事情聴取されました。わけを話すと警官も分かってくれたらしく、その声が 2 階にも聞こえたらしくマリが気づいてくれました。(小 5、G05_001)
- (4) ケンは言いました。「どうしよう。今日はマリちゃんと今後について話し合う予定なのに…」そこでケンは大声でさげんでマリに気づいてもらおうとしました。しかしぐっすりねてしまっているマリには届きません。そして考えあぐねた結果、自分のカバンからはしごを取り出しました。そうこのやけに小さなカバンはまほうのカバンだったのです。そしてそのはしごをやねにかけ、登ろうとしたその時、「何してるんですか」と声をかけられました。ふりむくとそこには警官が。「あやしいものではないんです、ただ鍵を…」とケンは説明しましたが、警官は、「分かっています。ただ私の話を聞いて下さい。あなたが結婚しようとしている「マリ」こと佐藤魔利は、結婚さぎ師です。佐藤に新たなカモができたと分かったので調査していました、あなたのことです。」「ええっ!?!」「あのマリちゃんが?」「はい。今後について話しあうと言って金をまき上げる、彼女のさぎの方法です。」「そうだったのか…ありがとうございます。今日鍵を忘れていたおかげで金を取られずに済みました…ありがとうございました、」「あのマリがさぎ師だったなんて…」二人がほほえみあった時、窓が開き、マリが「何

してるのケンちゃん、おそかったわね。明日のピクニック、用意してるわよ。」「では明後日にもう一度うかがいます」と言い残し、警官は去って行きました。(中 1、G07_001)

- (5) マリに声をかけたが、マリはねていたので反応がありませんでした。しかたなく、はしごを使ってあいている二階のまどから中に入ろうとしましたが、警官が勘違いして、呼び止められました。その後警官と和解し、その音でマリも起きてしまいました！(中 3、G09_001)
- (6) 家の二階にいるマリに助けを求めますが、寝ているため、返事がありません。そこでケンが、はしごを使い、二階のマリに気づいてもらおうとしたところ、不審に思った警官にとがめられました。ケンが事情を説明し、警官の誤解を解いていると、マリが起きてきて、外のような音を窓からのり出して確かめました。(高 2、G11_001)
- (7) 夜中だったのでインターホンを鳴らしても、家の中にいるマリは気づいてくれませんでした。ケンはマリの寝室の窓が空いていることに気付いたので「開けてくれ〜！」と呼びかけましたがマリは返事をくれません。なぜならマリはぐっすり眠っているからです。ケンはしばらく呼び続けましたが、庭にはしごをおいてあったことを思い出しました。はしごを開いている窓に立てかけ登っていると、たまたま通りかかった警察に呼び止められました。散々説明しても納得してくれませんでしたでしたが、その話し声でマリが目覚め、ケンは無事、家に入ることができました。(大 1、G13_001)

これらを一瞥するだけで、表記・語彙・文法などの各面において、L1 の子どもの書く日本語の質が連続的に変化していることが確認できるであろう。

また、学年間の変化は、計量的にも議論することができる。図 16 は、2 種のストーリーライティング作文を統合した際の 1 人あたりの総語数と、句読点・普通名詞・動詞・形容(状)詞・副詞・助詞・漢語・外来語頻度を自然対数に変換し、学年別の変化パターンを整理したものである。

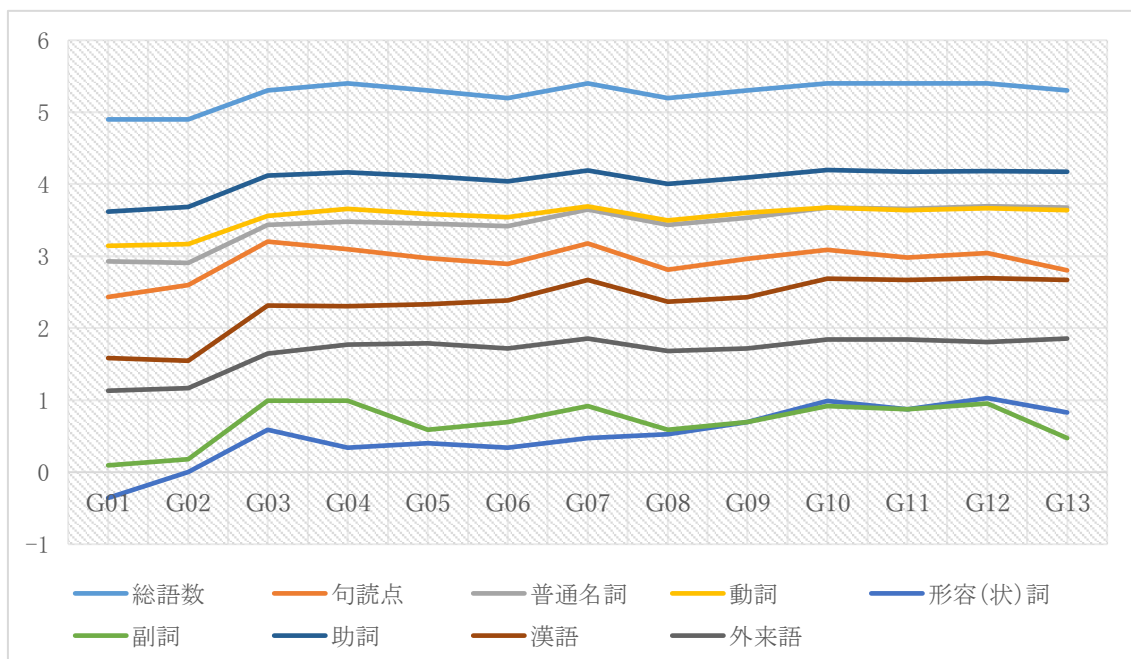


図 16 13 年間の総語数および各種頻度の変化 (自然対数変換)

13 学年中、データ量が最も少ない中学校 1 年生 1 (G07) の値が高めに出る傾向があるが (G07_001 の「鍵」作文をはじめ、いくつかの作文が例外的に長く、このことが全体の値に影響していると思われる)、その点を除いて考察すると、(1) 小学校 1~2 年生 (G01~G02) 間では、形容 (状) 詞・句読点頻度が増加しているが、そのほかの指標における変化の度合いは小さいこと、(2) 小学校 2~3 年生 (G02~G03) 間ではほぼすべての指標で増加が認められること、(3) 小学校 3 年生~大学 1 年生 (G03~G13) の間では、形容 (状) 詞が (相対的に) 線型的な増加を見せ、副詞が高校 3 年生~大学 1 年生 (G12~G13) 間で減少傾向を示す一方、そのほかの指標は大きく変化しないこと、などが確認できる。これらは、L1 日本語の書き言葉の各面の発達が単純な線型的伸長としてとらえにくいことを示しているだろう。また、小 2 と小 3 の間に見られる相対的に大きな伸長は、書き言葉の L1 発達における「発達の節目」がおよそこの時期にある可能性を示唆しているようにも思える。

同じデータを使えば、品詞や具体的語彙項目の頻度変化に限らず、誤用や不適切な用例の出現度の変化を探ることも可能である。従来、L1 日本語の作文研究では、作文技術の巧拙を議論することはあっても、第二言語習得論 (SLA) 的な意味合いにおいて、日本語としての正用性や誤用性を議論することは少なかった。この点に関して、砂川 (2020) は、「児童・生徒作文コーパス」に含まれる小学校 2 年生および 3 年生の作文を分析し、文法や語彙に関するはっきりした誤用は少ないものの、繰り返しや、「だらだら文」が多く見られることを明らかにしている。また、表現上、問題となりうる事例をタイプ別に分類したところ、文法 (25.9% : 助詞・ヴォイス・従属節の誤用など)、表現 (21.3% : 繰り返し、接続関係・指示対象の不明瞭性、こなれない表現など)、話し言葉使用 (19.3% : 接続詞、各種語彙など)、だらだら文 (12.2% : 接続助詞による複数の節の連続)、ねじれ文 (10.7% : 主述照応の不備) などが多かったと述べている。また、砂川 (2022) では、同コーパスに含まれる小学校 2 年生、小学校 5 年生、中学校 2 年生の 3 群の作文を分析し、「不具合」な箇所の合計数が 40→25→12 のように変化することが示されている。砂川の研究で採用された分析の手法は、そのまま、JASWRIC のデータにも適用が可能である。

4.2 L2 日本語習得研究での利用

共通のプロンプトと条件で産出を集めていることから、JASWRIC のデータは、I-JAS のデータと組み合わせると統合分析にかけることもできる。I-JAS にも、年齢分布に配慮した 50 人の成人母語話者 (20 代 : 19 人、30 代 : 14 人、40 代 : 13 人、50 代 : 4 人) のデータが用意されているが、JASWRIC を使うことで、より多様な母語話者データを参照基準として学習者の日本語の諸相を探ることができる。とくに、低学年の母語話者児童と初級学習者、高年次の母語話者と上級学習者の比較は多くの研究上の可能性を有する。

以下、低学年児童 (小学校 1 年生) と初級学習者 (トルコ語母語、J-CAT スコア=101)、また、高年次生 (大学 1 年生) と上級学習者 (中国語母語、J-CAT スコア=320) の作文を順に比較してみよう。

- (8) それでケンとマリがちずをみていたら、そのあいだに、いぬが、ばすけつとのなかのさんどいっちを、ぜんぶ、たべてしまいました。それで、ケンとマリがびくにつくについて、ついたらばすけつとのなかをみてみたら、ばすけつとのなかからいぬがでてきました。それでばすけつとのなかをみたらなにもありませんでした。(G01_011)
- (9) マリさんとケンさん食べ物をピクニックに行くために作ります。地図をみるときいぬはバスケットの中にあるたべものをたべましたマリさんとけんさんはわからない時行きました。ピクニックをする所へ行ってバスケットのなかに見ました。たくさんびっくりしました。(TTR16)

小学校 1 年生の作文は、大半が平仮名で書かれているため、非常に幼い印象を与えるが、

言語的な問題はほぼ皆無で、継続アスペクト (みていた/している)、補助動詞用法 (みてみたら/してみる/してしまう)、複合動詞 (できました) など、日本語教育で難度が高いとされる表現も自然に出現している。一方、初級学習者の動詞使用は、「目る (→見る)」や「見ました」のように単純時制に偏重しており、細かな文法・表現のエラーも数多く認められる。これらは、母語話者の初期段階と学習者の初級段階が一定の重なりを持ちつつも、質的な異なりを有することを示唆するだろう。

次に、高年次生と上級学習者を比較する。

- (10) 彼らはサンドイッチやリンゴをバスケットに入れ、ピクニックの準備をしました。しかし、彼らが地図を確認している隙に犬がバスケットの中に忍び込んでしまいました。何も知らないケンとマリはそのバスケットを持ってピクニックに出かけます。着いた先でバスケットを開けた途端、中から犬が飛び出し、二人はとても驚きました。バスケットの中身は、犬に食べられてしまっていて、二人は残念に思いました。

(G13_004)

- (11) ずっと前から計画してきたから、二人とも楽しみにしています。二人は犬を飼っていますが、犬を連れていくわけにはいきません。実は二人がどこかに行けばいいかと迷っていますので、ちょっと地図を見て場所を決めているうちに、犬はこっそりとバスケットの中には入りました。それでは、二人は歌を歌ったり、笑ったりして、手を握って一緒にピクニックに行きました。昼になると、色々遊んだ二人ともお腹が空いて、バスケットを開けて昼ご飯を食べようとしていたとたん、犬はバスケットの中からどっさりと飛び出して、二人ともびっくりしました。すると、バスケットの中を見ると、用意してきたおいしいサンドイッチもリンゴも食べられてしまいました。ああ、本当についていないですね。(CCM35)

大学生作文も動詞の形態的多様性によって特徴づけられるが (継続相: 確認している、受け身: 食べられて、複合動詞: 忍び込んで・飛び出し、テシマウ形: 忍び込んでしまいました/食べられてしまっていて)、上級学習者もこの点で大きな違いはなく、多様な動詞形を使用している (継続相: しています・飼っています・決めている・食べようとしていた、受け身: 食べられて、複合動詞: 連れていく・飛び出して、テシマウ形: 食べられてしまし(い)、テクル形: 計画してきた・用意してきた)。一方、いくつかの違いも認められる。ここでは、上級学習者作文の問題点を3点指摘しておきたい。1点目は接続表現の過剰使用である。学習者はストーリーの流れを強調しようとして「それでは」「すると」「昼になると」「見ると」といった接続表現を多用するが、大学生作文に同様のつなぎ言葉は出てこない。学習者が使用した接続表現の多くは本質的に不要なもので、削除が妥当である。2点目はコロケーション上の不整合である。学習者は「どっさりと」「飛び出(す)」と書いているが、2つの語は通例つながらない (「どっさに」との混同か?)。3点目は「途端」の不適切な使用である。「途端」は、「何かをする・何かが起こると、ほとんど同時に、急激な変化・因果関係のある事柄・偶然一致の事柄などが起こることを表し、習慣性のない過去の出来事の客観的な描写・事情説明をする」もので (『研究社類義語使い分け辞典』)、前件には時間の特定された具体的・特定のイベントが来る。この点に関して、母語話者用例にある「バスケットを開けた」+「途端」という組み合わせは自然だが、学習者用例にある「食べようとしていた」+「途端」の組み合わせは不整合である。

このように、JASWRIC に含まれる幅広い母語話者データを併用することで、I-JAS を用いた学習者研究・習得研究はさらなる広がりを見せるだろう。また、I-JAS に含まれる 1,000 名の学習者作文と、JASWRIC に含まれる 700 名の L1 作文があれば、大規模な計量的比較研究も可能になると思われる。この方向性については稿を改めて詳しく論じることとしたい (石川、2022)。

5. まとめ

以上、本稿では、JASWRIC の構築過程と、集められたデータの概要について報告を行った。JASWRIC は公開型のコーパスであるため、分析結果を第三者が検証することができる。また、データの使用者は、著作権法の定める範囲において、それぞれの研究目的に沿ってデータを加工・修正したり、新たな情報を付与したりすることもできる。さらには、同じプロンプトを使用して独自のデータ収集を行うことで、リサーチコミュニティ全体で、相互比較可能なコーパスデータのさらなる質的・量的拡大を図っていくことも可能であろう。

コーパス言語学の普及が進む中で、これからのコーパスは、従来のような閉じた完成物ではなく、外部に開かれ、他者を巻き込み、逐次的に変化・発展していく動的な資料に変容していくのではないかと予想される。JASWRIC が、こうした新しいタイプのコーパスとして、多様な研究者・研究機関とのコラボレーションを触発し、L1/L2 の日本語研究の進展に貢献することを期待したい。

注

(1) 本研究は、第一筆者が研究企画・データ収集統括・大学生データ収集・収集データ整理・検索システム開発・データ分析を担当し、第二筆者以下が小学生・中学生・高校生のデータ収集の実務を担当した。

(2) 本稿で示した分析事例は、2022年8月1日現在の JASWRIC のベースデータ (V1.0) に基づく。今後、形態素解析データの修正や JASWRIC Online の検索仕様の修正などが行われた場合、結果が変動する可能性があるので留意されたい。

謝辞

本コーパスの開発は、子どもの作文収集を試みた先行研究に多くを負っている。坂本真樹氏・富士原紀絵氏・宮城信氏の3氏には、各氏の構築されたコーパスの内容に関する照会に対して、丁寧なご教示をいただいた。今田水穂氏からは、富士原氏・宮城氏らと構築されたコーパスの最新の語数について情報をいただいた。砂川有里子氏からは、「児童・生徒作文コーパス」のデータを用いて開発中の「問題例検索システム (仮称)」についての情報をいただいた。各氏に深く御礼を申し上げる。また、本コーパスの最大の特徴は、I-JAS のストーリーライティングタスクのデザインを踏襲してデータ収集を行った点にあるが、プロンプトの使用を快諾くださった I-JAS 開発者である迫田久美子氏に改めて深く感謝したい。本コーパス開発を着想する直接のきっかけになったのは、I-JAS のプロンプトを用いて国内児童の産出を収集・分析された松隈杏梨氏の研究 (2021) であった。類似したコンセプトに基づくデータの収集についてご理解をくださった松隈氏、ならびに松隈氏の指導教員である丸山岳彦氏に御礼申し上げる。

本研究は、パフォーマンス評価データの体系的収集をテーマとする第一筆者の科学研究費 (20H01282) プロジェクトの成果の一部である。また、神戸大学附属学校部校種間連携部門プロジェクト「幼児期から児童期における発達・教育研究」と連携して実施されたものである。関係各位の支援に感謝申し上げます。

文献

- 阿部藤子・今田水穂・宗我部義則・富士原紀絵・松崎史周・宮城信 (2017) 「児童生徒の『手』作文に於ける経年変化の計量的分析：1992年と2016年の作文を比較して」『言語資源活用ワークショップ発表論文集』1, 234-247. <http://doi.org/10.15084/00001478>
- 石川慎一郎 (2021) 『ベーシックコーパス言語学』(第2版). ひつじ書房.
- 石川慎一郎 (2022) 「L2 日本語学習者を対象とした中間言語対照分析における参照基準の拡張—『多言語母語の日本語学習者横断コーパス』(I-JAS) と『小中高大生による日本語絵

- 描写ストーリーライティングコーパス』(JASWRIC)の連動分析の試み—『計量国語学会第66回大会予稿集』43-47.
- 今田水穂 (2020) 『『児童・生徒作文コーパス』形態論・係り受け情報データ Ver.1.6』.
- 今田水穂・宮城信 (2020) 「学校課題作文コーパスの構築」『言語資源活用ワークショップ発表論文集』5, 103-113. <http://doi.org/10.15084/0000314>
- 小磯花絵・居關友里子・柏野和佳子・角田ゆかり・田中弥生・宮城信 (2020) 「子どもの会話コーパスの構築に向けて」『言語資源活用ワークショップ発表論文集』5, 157-163. <http://doi.org/10.15084/00003155>
- 国立国語研究所 (1989) 『児童の作文使用語彙 (国立国語研究所報告 98)』東京書籍.
- 坂本真樹 (2010) 「小学生の作文コーパスの収集とその応用の可能性」『自然言語処理』17(5), 75-98.
- 迫田久美子・石川慎一郎・李在鎬 (2020) (編) 『日本語学習者コーパス I-JAS 入門: 研究・教育にどう使うか』くろしお出版.
- 島村直己 (1987) 「児童の漢字使用: 課題作文の漢字含有率から」『研究報告集 8 (国立国語研究所報告 90)』77-94. <http://doi.org/10.15084/00001105>
- 鈴木一史・棚橋尚子・河内昭宏 (2011) 「作文コーパスからみる生徒の使用語彙」『特定領域研究「日本語コーパス」平成22年度公開ワークショップ (研究成果報告会) 予稿集』343-350.
- 砂川有里子 (2020) 「『児童・生徒作文コーパス』に見られる問題例の調査報告—低学年児童の場合—」『日本語教育連絡会議論文集』33, 127-135.
- 砂川有里子 (2022) 「『児童・生徒作文コーパス』を用いた長文の文構造調査」『日本語習熟論学会第1回大会研究発表予稿集』17-24.
- 田中美也子 (1998) 「作文能力発達に関する縦断的研究その三: 段落意識形成の過程に関する一考察」『語学文学』36, A1-9.
- 永田亮・河合綾子・須田幸次・掛川淳一・森広浩一郎 (2010) 「作文履歴をトレース可能な子供コーパスの構築」『自然言語処理』17(2), 51-65.
- 成田信子・宗我部義則・田中美也子 (1995) 「作文能力発達に関する縦断的研究その一: 小学生から大学生に至る同題作文の分析」『国語科教育』42, 183-192.
- 富士原紀絵・宮城信・松崎史周 (2016) 「児童生徒作文の基礎的研究: 児童生徒作文コーパスの構築と活用」『子ども学研究紀要』4, 9-20.
- 松隈杏梨 (2021) 「JSL 児童生徒の接続詞使用における課題—I-JAS 学習環境別データの分析と調査の実施から—」学習者コーパス研究会 (2021年12月26日実施) 発表資料.
- 宮城信・今田水穂 (2018) 「『児童・生徒作文コーパス』を用いた漢字使用能力の発達過程の分析」『計量国語学』31(5), 352-369.
- 村上博之・田中美也子 (1997) 「作文能力発達に関する縦断的研究その二: 同題作文における漢語表現の発達」『国語科教育』44, 105-114.

参考 URL

- 「小中高大生による日本語絵描写ストーリーライティングコーパス」(JASWRIC)
<https://language.sakura.ne.jp/jaswric/>

書名 言語資源ワークショップ 2022 発表論文集

発行日 2023年2月28日

発行者 国立国語研究所言語資源開発センター

<https://clrd.ninjal.ac.jp/>

連絡先 〒190-8561 東京都立川市緑町10-2

大学共同利用機関法人 人間文化研究機構 国立国語研究所言語資源開発センター

電話 0570-08-8595 (ナビダイヤル)

(C) 2022 Center for Language Research Development, NINJAL

