



Title	テキストマイニングとデジタルヒューマニティーズ 2020 （冊子）
Author(s)	
Citation	言語文化共同研究プロジェクト. 2021, 2020
Version Type	VoR
URL	https://hdl.handle.net/11094/85194
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

言語文化共同研究プロジェクト 2020

テキストマイニングと デジタルヒューマニティーズ 2020

山	田	彬	堯
徐			勤
黄		晨	雯
福	本	広	光
王			鉦
田	畑	智	司

大阪大学大学院言語文化研究科

2021

言語文化共同研究プロジェクト

テキストマイニングと デジタルヒューマニティーズ 2020

目 次

田畑 智司	プロジェクトの目的と活動	1
山田 彬堯	江戸期から昭和期にかけての尊敬語のバリエーション 多項混合効果ロジスティック回帰によるベイズ統計学的 アプローチ	5
徐 勤	日本人中国語学習者の叙述文における言語的特徴の分析 —MF/MD 法を使って—	27
黄 晨雯	Top2Vec による小説の探索的研究 —程小青の作品解説を中心に—	43
福本 広光	19 世紀米国大統領演説における分離不定詞 — Abraham Lincoln の事例から—	55
王 鉦	中国人日本語学習者と日本語母語話者における多義動詞 「切る」のカテゴリー構造比較 —心理実験により意味分析の結果を検証する—	75

「テキストマイニングとデジタルヒューマニティーズ」 プロジェクトの目的と活動

本共同研究は、自然言語処理、コーパス言語学・計量言語学、数理統計学、データマイニング、機械学習など、諸分野の知見を有機的に統合した方法論を開発し、テキストマイニングを応用して人文学、言語文化学の諸問題にアプローチする、すなわち「デジタルヒューマニティーズ (Digital Humanities)」の実践と理論的精緻化の可能性を探る営みである。このプロジェクトは、2001 年度に岩根 久教授、緒方 典裕助教授、および筆者の3名でスタートした「電子化言語資料分析の方法論」を基礎とするが、2003 年度から名称を一部改め、言語文化研究科の大学院生もメンバーに加わった。2006 年度には三宅 真紀助教の加入を得て、対象言語も英・仏・ギリシャ語に広がった。2011 年には言語文化教育論講座に着任した今尾 康裕講師が加入した。2014 年度後期から、さらに Hodošček Bor 講師が加わった。そして、2019 年度をもって退職された岩根 久教授の後任として、2020 年度に山田 彬亮講師の着任があり、現在の陣容となっている。(職位はいずれも当時)。2016 年度から、プロジェクトの名称を、当該リサーチコミュニティの名称としてより相応しい「テキストマイニングとデジタルヒューマニティーズ」にアップデートしたが、研究の系統は創始時より常に一貫している。

「テキストマイニングとデジタルヒューマニティーズ」プロジェクトは大きく分けて二つの層で構成されている。一つは研究基盤となるコーパス、テキストアーカイブの開発・構築、もう一つは構築したコーパス、テキストアーカイブからのデータ抽出法研究、並びに得られた高次元の言語データの計量分析である。前者には英・仏語の文学作品や、聖書 (共観福音書) などの電子テキスト化、ロシア語政治演説コーパス、近代日本文学コーパスの編纂、マークアップ言語 XML による TEI (Text Encoding Initiative: デジタル化したテキストの国際互換規格の枠組) に準拠したタグ付けなど、人文学資料のデジタル化やマークアップ法、データ符号化方法論の開発などが含まれる。一方、高次元人文学データ分析の事例として、語彙・語法、コロケーション、意味構造、語用論などのレベルにおける言語使用の実態研究、高度な数理モデルや機械学習を応用した言語分析やテキストマイニング、文学作品の言語特徴の特定や、使用域間の言語変異や文体識別問題の考察、著者推定法の精密化研究を挙げることができる。

本プロジェクト班は言語文化研究科の専任教員5名と名誉教授1名 (今尾 康裕, 田畑 智司, Hodošček Bor, 三宅 真紀, 山田 彬亮, 岩根 久名誉教授), 当研究科博士後期課程在学生7名 (黒田 絢香, 黄 晨雯, 岡部 未希, 徐 勤, 福本 広光, 藤田 郁, 三野 貴志), 博士前期課程在学生3名 (田辺 まりか, 竹森 ありさ, 王 鈺), 研究生2名 (曹 芳慧, 李 晨婕) に加え, OG の土村 成美氏 (2021 年3月当研究科より博士学位取得)・浅野 元子氏 (2020 年3月博士学位取得)・杉山 真央氏 (2019 年3月博士学位取得), 本学非常勤講師の高橋 新氏, 南澤 佑樹氏 (本研究科博士課程修了), 摂南大学後藤 一章氏 (本研究科博士課程修了), 帝塚山学院大学八野 幸子氏 (本研究科博士課程修了), 数理・データ科学教育研究センターの上阪 彩香氏を主たる参加メンバーとしている。研究を遂行するために, コアメンバー以外も自由に参加できる月例の研究会・討論会, さらに統計数理研究所の言語系共同利用研究班との夏・春の合同セミナーの開催などを通して, 研究情報の交換, 論文や開発ツールのレビューなどを行っている。

2020 年度は, 新型コロナウイルスが世界中を席卷し, パンデミック下全ての研究会をオンラインで行わざるを得なくなった。前例のない研究会実施形式ではあったが, 次第に新たな形式に慣れていくとともに, 実は当プロジェクトの研究内容はオンライン発表に絶妙に適合しているのだということを実感した一年でもあった。もっとも, インフォーマルな雰囲気の中で情報交換を行う過程で思いがけない新たな着想, セレンディピティを生むオフライン, 対面での懇親の場を持つことができなかったのは残念であったけれども。2020 年度の本プロジェクト研究会開催記録を以下に記す。

2020 年度「テキストマイニングとデジタルヒューマニティーズ」研究会開催記録

第 1 回 2020 年 5 月 29 日開催 (Zoom を介したオンライン)

発表者・発表題目

全メンバー 2020 年度の活動計画打合せ

第 2 回 2020 年 6 月 26 日開催 (Zoom を介したオンライン)

発表者・発表題目

Hodošček Bor 「言語処理とアノテーション」

第 3 回 2020 年 7 月 17 日開催 (Zoom を介したオンライン)

発表者・発表題目

田畑 智司 「R パッケージ Stylo によるテキスト分析」

第 4 回 2020 年 8 月 21 日開催 (Zoom を介したオンライン)

発表者・発表題目

八野 幸子 「理科との連携による教科等横断的視点を取り入れた英語教育のための語彙研究
—植物に関する文脈に出現する語彙を中心に—」

第 5 回 2020 年 9 月 11 日開催 (Zoom を介したオンライン)

発表者・発表題目

林 希和子 「留学生 (高等教育機関進学者) にむけた共通語彙の解明
—日本留学試験問題の語彙コーパスの作成と分析—」
藤田 郁 「Tennyson の -ly 副詞再考」

第 6 回 2020 年 10 月 9 日開催 (Zoom を介したオンライン)

発表者・発表題目

山田 彬堯 “Multinomial Mixed-Effects Models and Linguistic Variation:
Competitions among Japanese Subject-Honorific Constructions”
竹森ありさ “An Analysis of Intensifying Similes with Color Adjectives”

第 7 回 2020 年 11 月 4 日開催 (Zoom を介したオンライン)

発表者・発表題目

黄 晨雯 「Top2vec による小説の探索的研究」
岡部 未希 「ディキンソンの詩における鉱石の描写—ダイヤモンドに着目して—」

第8回 2020年11月29日開催 (Zoom を介したオンライン) Japan-Korea Collaborative Symposium
on Digital Humanities として開催
発表者・発表題目

Tomoji Tabata “Language Action Types and the Semantics of Texts:
Using rhetorical annotation to classify texts into meaningful groups”
Hodošček Bor “Visualization of Classical Japanese Poetic Vocabulary”
Iku Fujita “DH Approaches ‘to Know’ Tennyson”
Ayaka Kuroda “Exploring the Network of Topics, Words and Documents:
Machine Learning Analysis for Finding Key-Words”

第9回 2020年12月4日開催 (Zoom を介したオンライン)
発表者・発表題目

田辺 まりか 「日英機械翻訳における日本語省略表現訳出分析と事前学習モデルの改良
Wikipedia 日英京都関連文書対訳コーパスを用いて」
徐 勤 「日本人中国語学習者の叙述文における言語的特徴の分析—MF/MD 法を使って—」

第10回 2021年1月15日開催 (Zoom を介したオンライン)
発表者・発表題目

三野 貴志 「There comes の共時的・通時的・類型論的分析」

第11回 2021年2月5日開催 (Zoom を介したオンライン)
発表者・発表題目

三宅 真紀 「IIIF (International Image Interoperability Framework) 画像 API を用いた新約聖書
写本研究との連携」
浅野 元子 「研究論文抄録の英日対訳パラレルコーパスの言語使用と句読法の特徴について
—教育応用を視野に」
福本 広光 「19世紀アメリカ大統領スピーチにおける分離不定詞構造
—ABRAHAM LINCOLN の事例から—」

第12回 2021年3月5日開催 (Zoom を介したオンライン)
発表者・発表題目

王 鈺 「中国人日本語学習者と日本語母語話者における多義動詞「切る」の意味構造分析
—心理実験により意味分析の結果を検証する—」
高橋 新 “Study on the Application of Stylometric Methods to Analysing English Translations of the
Bible: 17+ x Gospels of Mark and John”
今尾 康裕 「次世代 CasualConc 開発版デモンストレーション」

第13回 2021年3月20日開催 (Zoom を介したオンライン) 統計数理研究所言語系共同研究班
合同セミナー「言語研究と統計 2021」として開催
発表者・発表題目

竹森ありさ	「英語色彩語を含む強意直喩表現の分析」
福本 広光	「アメリカ大統領演説における分離不定詞」
今尾 康裕	「ジャンルの違いはコロケーションにどのように反映されるのか」
山田 彬堯	「ラレル尊敬の特徴は何か？」
黄 晨雯	「Top2Vec による小説の探索的研究」
岡部 未希	「Dickinson の詩における鉱石」
黒田絢香	「トピックモデルによる文学作品分析」
浅野 元子	「国際英語としての英語学術論文のコーパス研究 —医学分野への日本からの貢献を例に—」
高橋 新	「英語翻訳聖書間の計量的スタイル及び語彙選択の分析 —マルコ及びヨハネによる 両福音書 17 + α 翻訳の分析」
王 鈺	「中国人日本語学習者と日本語母語話者における多義動詞「切る」の意味構造分析 —心理実験により意味分析の結果を検証する—」
八野 幸子	「理科との連携による教科等横断的英語教育のための語彙・表現研究」
徐 勤	「日本人中国語学習者の叙述文における言語的特徴の分析 —MF/MD 法を使って—」

2021 年 4 月
研究代表者 田畑 智司

江戸期から昭和期にかけての尊敬語のバリエーション 多項混合効果ロジスティック回帰によるベイズ統計学的アプローチ

山田 彬堯

大阪大学大学院言語文化研究科

〒 560-0043 豊中市待兼山町 1-8

E-mail: a.yamada@lang.osaka-u.ac.jp

あらまし 本稿は、江戸から昭和にかけて使われた日本語尊敬語構文に対する、多変量統計解析の報告である。これまでの研究で尊敬語は語族の異なる様々な言語に分布していることが明らかになっているが、そのような海外の尊敬語とは異なり、日本語の特徴的な性質として知られるのが、同一言語システムの中に複数の尊敬語構文が競合している点である。このため、この特異なシステムを研究することで、尊敬語だけではなく、言語変異一般に関するより深い考察が得られると期待される。そこで、本研究では、日本語歴史コーパスのデータを用い、多項混合効果ロジスティック回帰分析を行い、「御...になる」「...なさる」「御...なさる」構文の選択傾向を調査した。明らかになったのは以下の点である：第一に、最も効果量の大きい固定効果は命令文か否かという文のモードである。第二に、明治期以降「御...になる」の使用が強まるが、それぞれの動詞によってこの全体の傾向に逆行するものもあれば、より強く「御...になる」を指向するものがある。第三に、尊敬語構文の選択にはジャンルごとの違いも存在し、国語教科書での「御...になる」の使用が顕著である。

キーワード ベイズ統計学、多項混合効果ロジスティック回帰分析、尊敬語、変異、通時的变化

The variation among the subject-honorific constructions from the Edo period to the Showa period:

A Bayesian approach using multinomial mixed-effects logistic regressions

Akitaka Yamada

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract This paper is a quantitative multifactorial study of the Japanese subject-honorific constructions used from the Edo period to the Showa period. Subject-honorific expressions are found in genealogically-unrelated languages, but unlike other languages, Japanese possesses several different competing constructions to encode the speaker's respect for the referent, and thus the examination of this peculiar system is expected to develop our understanding of subject honorifics, and the language variation in general. Using the data taken from the Corpus of Historical Japanese (CHJ), this study compares three subject-honorific constructions *o...ni nar*, *...nasar*, and *o...nasar* by conducting a Bayesian multinomial mixed-effects logistic

regression. The findings are as follows: First, the imperative shows the largest effect size indicating the difficulty of the *o...ni nar* construction's taking this sentence mood. Second, verbs show their own selectional tendencies. Finally, the selection of a subject-honorific construction is also subject to the genre; *o...ni nar* is abundant in textbooks of standard Japanese.

Keywords Bayesian statistics, multinomial mixed-effects logistic regression, subject honorifics, variation, diachronic development

1. はじめに

近年の研究の進展により、Ladakhi (Sino-Tibetan; Koshal 1987)、Classic Nahuatl (Uto-Aztec; Andrews 1975; Launey and Mackay 2011)、Magahi (Indo-Aryan; Alok to appear)、Maithili (Indo-Aryan; Yadav 1996)、Timucua (Isolate; Broadwell 2019)、アイヌ語 (金田一・知里 1936) など、語族を異にする言語の間に尊敬語 (主語の指示対象に対する敬意を表す動詞の活用/接尾辞) というべき表現が存在することが分かってきた。そして、これらの通言語的なデータは、(i) 形態統語的表出の仕方、(ii) 敬意の対象、(iii) 通時的な発達、(iv) 競合、(v) 他の文法形式との相互作用、といった視点から整理することができ、これらのパラメータによってそれぞれの言語の尊敬語に特徴的な性質を把握することができる (Yamada to appear)。

この視点から日本語の尊敬語を眺めたとき、とりわけ日本語に特異な性質として浮かび上がるのが競合 (competition) の存在である (菊地 1997、Yamada 2019、2020b、to appear)。現代韓国語では尊敬語接辞として *-si* のみが使われるのに対して、日本語では下記に見るように複数の構文が競合を見せる。競合は、とりわけ変異理論 (Variation Theory) の枠組みで研究対象にされてきた。ただし、これまで扱われてきた現象の多く、例えば d-deletion (Guy 1991)、補文標識 *que* の省略 (Cedergren and Sankoff 1974) などに代表されるような二項対立である。(1) に見られるように 3 つ以上の表現の競合する事例は、まれであり、この現象の研究は、敬語研究の枠組みを超えて、言語変異の研究一般に大きな貢献をすることが期待される。

(1) 競合する尊敬語形式

- | | | |
|----------|-----------|-----------|
| a. 到着される | b. ご到着される | c. ご到着になる |
| d. 到着なさる | e. ご到着なさる | |

一方で、国内の研究、特に国語学の伝統においても尊敬語の競合は多くの研究者の関心を集めてきた。とりわけ、その通時的な発達には精密な研究の蓄積があり、登場した時代の違いや他の文法要素と関わり合いには、記述・分析が進められている (山田 1959、辻村 1968、原口 1974、村上 2005、山田 2013、2015)。

¹ 本研究は 2020-2021 年度「研究活動スタート支援」#20K21957 の支援を受けている。また、2021 年 3 月 20 日 (土) に開催された統計数理研究所言語系共同研究グループ研究発表会「言語と統計 2021」(セミナーシリーズ No. 16) での発表を発展させたものであり、当日は、石川慎一郎先生 (神戸大学)、前田忠彦先生 (統計数理研究所) から大変有益なコメントいただいている。本論文の誤りはすべて筆者によるものである。

しかし、通言語的研究、あるいは通時的研究のいずれの枠組みであっても、統計的アプローチに基づいてこの尊敬語の競合という問題に迫った研究は未だ蓄積が浅い。だが、構文が競合するとは、構文の選択が確率的であるということであり、したがって、統計的な手法・視点が持ち込まれることで、これまでの研究では明らかにできなかった尊敬語使用の詳細が明らかになるのではないかと期待される。このような問題意識から、著者はすでに、現代日本語の尊敬語選択の問題について、『日本語書き言葉均衡コーパス (BCCWJ)』を用いた統計的分析を行い、細やかな構文選択の諸相を明らかにしてきた (山田 2019、Yamada 2020b)。しかし、用いたコーパスの性質上、その通時的な発達については考察の対象外となっていた。だが、なぜ所与の言語体系に競合が存在しているのか、という点を考える際に、どのような歴史的経緯で複数の構文が言語体系にその地位を確立させてきたのかという通時的な視点は欠かせない。

そこで、本稿では、対象を現代日本語から近代日本語へと変え、定性的な先行研究で指摘されてきた点がどの程度統計分析によって支持されるのか、そして先行研究では議論されてこなかった注目すべき事実にはどのようなものがあるのかを論じ、尊敬語の競合についてのより詳細で包括的な理解を目指す。

本稿の構成

本稿の構成は以下の通りである。2 節では、尊敬語の競合に関する先行研究を紹介し、それらの先行研究では深く論じられてこなかった問をリサーチクエスションとして明確に提示する。続く、3 節では、本研究で用いるコーパスデータについての説明をし、そのデータに対する記述統計学的な探索を 4 節で、推測統計学的な分析を 5 節で展開する。6 節において本研究のまとめと将来への展望を述べる。

2. 先行研究

伝統的な国語学の系譜に連なる研究では、尊敬語構文の歴史的発達／変遷について詳しい記述・分析がある (山田 1959、辻村 1968、原口 1974、村上 2005、山田 2013、2015)。以下その要点をまとめていく。

中世から 17 世紀

中古以来、「する」という動詞の補充形尊敬語として使われてきた「なさる」が、補助動詞用法を獲得したのは、中世末のこととされる (山崎 1963:109; 村上 2005:18-19)。ただし、『天草版平家物語 (1593)』『きのふはけふの物語 (元和古活字版)』『大蔵虎明本狂言集 (1642 年写)』など 16 世紀後半から 17 世紀に見られる初期事例は、「御... なさる」という形で用いられ、当初は接頭辞「御」の存在が義務的であったことが知られている。¹

¹ ただし、東国方言の資料である『雑兵物語 (1728 写; 成立したのは 1683 年以前とされる)』などでは「捨なさる」といった「御」を落とした形が報告されている (村上 2005:20)。

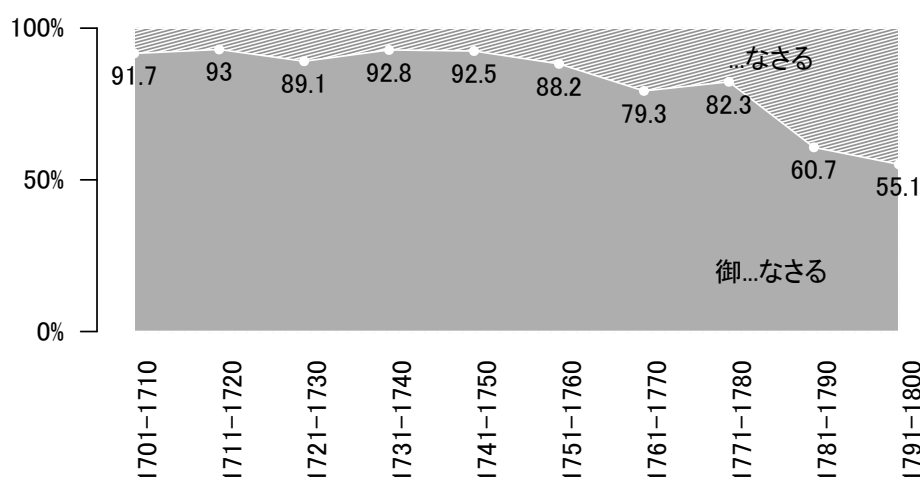


Fig. 1: 18 世紀における「御... になさる」対「... なさる」の使用比率（村上 2005:23 より作成）

18 世紀

「御」のとれた「... なさる」という形式が使われるようになったのは少し時代が下り 18 世紀ごろとされる。村上 (2005) は、18 世紀に刊行された 115 資料における「御... なさる」と「... なさる」を調査し、その使用頻度を刊行年によって 10 年ごとにまとめて報告している。表 1 はこの調査結果を積み上げ面グラフによって視覚化したものである。これを見ると、18 世紀の後半に「... なさる」形の使用が増加していく傾向が読み取れる。

村上 (2005: 24-26) は、「御」が義務的ではなくなった原因として二つの可能性を指摘している。第一は、「御」接辞の上接範囲の拡大である。「御」はもともと動詞そのものにつくことはなかったが、1750 年代（宝暦）ころから、直接動詞につく事例が増加する。動詞に「御」がついたりつかなくなったりする状況、すなわち「御」の着脱が自由にできる状況が生まれたことで、「御... なさる」構文の「御」も着脱が自由になった可能性がある。第二は、連用形命令法の登場である。「あそこへ行き」のような連用形命令形は 1750 年代を境に登場した表現であるとされる（村上 2003）。この「行き」のような表現が「行きなされ」の「なされ」が落ちたものという異分析を受け、「行きなされ」という命令形が認可され、続いて非命令形「行きなさる」が認可されるようになったというのが村上の提案する二つ目の可能性である。

これらの道筋はあくまでも仮説であり、その因果関係の証明は本稿の射程を大きく超えるが、しかし尊敬語が命令形と密接な関係にあるという視点は重要な着眼点である可能性がある。これは、尊敬語の使用が命令形と強い相互作用を持つためである。第一に、現代日本語では、(2)a-(2)c が示すように、一部の尊敬語は命令形では使えない（Svahn 2016、山田 2019、Yamada 2020a）。第二に、命令形を持つことが可能である (2)d-(2)e の表現でさえも、その尊敬語の意味が、平叙文や疑問文で生じる尊敬語の意味とは質的に異なることが知られている；平叙文や疑問文などでは「主語の指示対象が話し手よりも社会的に上である場面」、例えば、生徒が先生に対して話をする場面では尊敬語の使用が可能である（例：「先生がお話しなさる」と生徒が発言しても問題はない）が、命令文においては（聞き手と同一人物にな

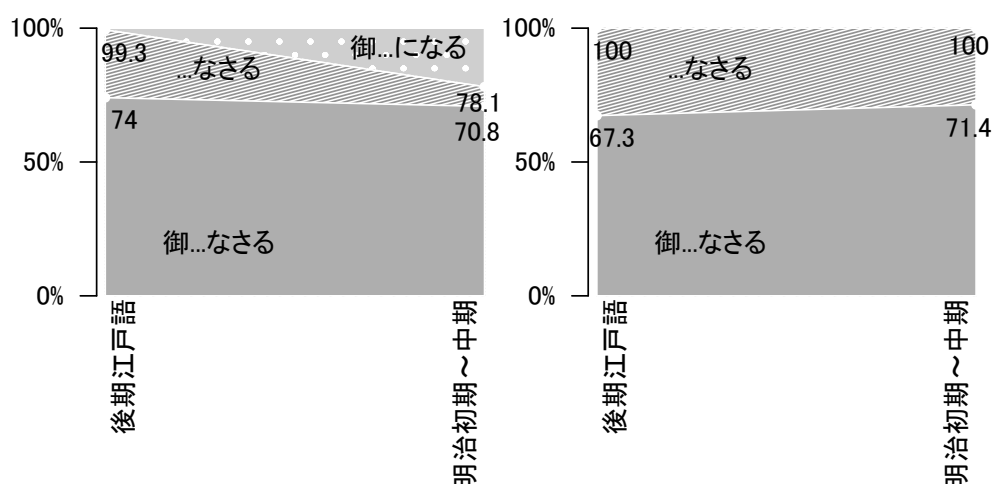


Fig. 2: 後期江戸語～明治中期の三つの尊敬語構文の使用比率（左：非平叙文、右：命令文；山田 2015:46、137 より作成）

る）主語の指示対象が話し手から見て社会的に目上の人である場合であるにもかかわらず、尊敬語の使用が不適切となる（例：(2)d-(2)e を、生徒が先生に使った場合など）。このように、尊敬語は命令文と強い相互作用を示すのであり、本稿でのちに見る統計モデルでも、この点を考慮し文のムード（Sentence mood）を説明変数として組み入れて考察を行う。

- (2) a. *仕事され！ b. *お仕事され！ c. *お仕事になれ！
 d.. 仕事なさい！ e. お仕事なさい！

19 世紀

19 世紀は「御... になる」という形式が日本語の尊敬語体系に加わった時期として知られ、この表現はとりわけ明治 10 年代後半から使用が拡大していったことが知られている（辻村 1968；山田 1959；原口 1974；山田 2013、2015）。山田 (2015) は後期江戸語（宝暦から幕末期）の洒落本、滑稽本、人情本、および明治初年から 30 年までの小説・講演速記資料から、当該時期の尊敬語形式の競合を分析している。この論文で収集された「御... なさる」「... なさる」「御... になる」の三形式の使用比率を示すと Fig 2 のようになり、ここから非命令文における「御... になる」の伸びが明確にうかがい知れる。また、Fig 2（右図）からは、(2) に示した命令文との相互作用が明治期においても存在していたことが分かる。

リサーチクエスチョン：記述統計学から推測統計学へ

ここまでまとめてきたことがらは、惜しめない手間と時間が注がれた国語学の枠組みに連なる先行研究の緻密な言語観察の賜物であり、これによって、尊敬語形式の大局的な移り変わりの傾向は十分明らかになってきたと言ってよい。しかし、その定量的な扱いについては、コーパス言語学／推測統計学の立場からさらなる発展が期待できる。例えば、適切な統計モデルを作ることで、(3) に掲げたようなリサーチクエスチョンに対し、より精密な答えを与えることができる。

(3) リサーチクエスション

- a. (問1) 他の要因の影響を統制した下での、各要因(例: モーラ制約、命令文の影響)の効果量はどの程度なのか?
- b. (問2) この節でまとめられた尊敬語構文の選択傾向は全ての動詞で一律に同じなのか? それとも、一部の動詞には大局的な言語変化に抗う傾向が見られるのか?
- c. (問3) この節でまとめられた尊敬語構文の選択傾向は全てのジャンルで一律に同じなのか? それとも、一部のジャンルでは大局的な言語変化に抗う傾向が見られるのか?

そこで、本稿では、コーパスデータを利用し、多項混合効果ロジスティック回帰モデルに基づいた分析から、この三つの問いを詳細に検討していく。

3. データ

形態素情報に加え、ジャンルや成立年など豊富な情報がアノテーションされた、大規模な歴史コーパスであることを重要視し、本研究では、国立国語研究所によって作成・公開されている日本語歴史コーパス(CHJ、国立国語研究所 2020)を分析対象として利用する。検索対象は本研究が対象とする「御...になる」「...なさる」「御...なさる」の使用が観察され始める江戸時代から、本コーパスがカバーする最後の年代である昭和時代までである。コーパスからのデータ抽出には中納言を利用し、下記の検索式を用いて、データを取得した。² 動詞だけではなく、名詞にも検索対象を広げているのは、(i)「ご紹介になる」のようなサ変名詞をも考察対象に入れるため、また、(ii)和語起源の事例であっても、名詞としてアノテーションを受けている事例があるためである。³

(4) 御[動詞/名詞]になる

- a. キー: 品詞 LIKE "動詞%" AND 前方共起: 語彙素="御" ON 1 WORDS FROM キー AND 後方共起: 語彙素="に" ON 1 WORDS FROM キー AND 後方共起: 語彙素="成る" ON 2 WORDS FROM キー
- b. キー: 品詞 LIKE "名詞%" AND 前方共起: 語彙素="御" ON 1 WORDS FROM キー AND 後方共起: 語彙素="に" ON 1 WORDS FROM キー AND 後方共起: 語彙素="成る" ON 2 WORDS FROM キー

² 平安期などの考察対象の時期を排除するため、すべての検索式には、続けて次の範囲指定を設けている。

(i) IN(時代名="5 江戸" OR 時代名="6 明治" OR 時代名="7 大正" OR 時代名="8 昭和")

³ 例えば、(i)の「頼み」は「名詞-一般」というアノテーションを受けている。

(i) 森さんが強つてお頼みになる。(60M 女世 1909_05027)

	御... になる	... なさる	御... なさる
動詞	1,276	2,515	584
名詞	757	543	290

Table 1: CSJ の検索結果

(5) (御) [動詞/名詞] なさる

- a. キー: 品詞 LIKE "動詞%" AND 後方共起: 語彙素="為さる" ON 1 WORDS FROM キー
- b. キー: 品詞 LIKE "名詞%" AND 後方共起: 語彙素="為さる" ON 1 WORDS FROM キー

これらの検索式で取得された事例は表 1 に示すとおりである。本研究では、以下の変更を施したうえで、これらのデータを使用する。⁴ 第一に、本コーパスのアノテーションには不統一な部分を修正する。和語動詞の連用形を名詞とコーディングしているものについては、動詞へ語彙素を変更し、その他の事例と統一化を行った。例えば、「お[わかり]になる」を「分かり」という名詞として扱っているものと「分かる」という動詞として扱っているものがあり、同一の語彙素として扱うためすべて「分かる」という動詞に統一した。また、「御覧になる」は「御覧」を一つの名詞として扱っているため、(4) の式では採取できない。そこで、(6) の検索式で別途検索し、ヒットした 90 件を表 1 のデータに追加した。

(6) キー: 語彙素="御覧" AND 後方共起: 語彙素="に" ON 1 WORDS FROM キー AND 後方共起: 語彙素="成る" ON 2 WORDS FROM キー

第二に、尊敬語とは言えない構文を排除する。例えば、「ご馳走/世話/供/為/用になる」は「御... になる」という形を取ってはいるものの、これらは主語に対する敬意を表す表現ではない。⁵ また、「お昼になる」のような美化語も「御... になる」として収集されてしまうがこれらも排除した（正月、昼、昼時、天気、流れ、大尽（大臣）、金持ち）。また、「為（す）る」という語彙素には、語彙素読みが「タメ」となっているものと「スル」となっているものがある。動詞であるのは後者であるので、前者の用例は除外した。「免」についても「免する」という非敬語用法がないため対象から外した。

第三に、尊敬語読みと非尊敬語読みが混在する可能性持つものも排除の対象とした。このような例には「お馴染み/蟲眞/仲間/知り合い/相手/友達/使いになる」が該当する。

第四に、明確なアノテーションの過誤であると考えられるものを排除した。これには「坊んさん」の「んさん」が「なさる」の活用と見なされたもの、「書き」の語彙素を「文」とし

⁴ (i) のような「二格主語構文」については、「ガ格」の指示対象を敬う他の大部分の用例と異なる点はあるものの、これまでの理論研究で重要な働きを果たしてきたことを踏まえ排除はしなかった。

(i) 浅倉君にしてもその他の諸君にしても先づもつて洋服が【お 似合ひ】 になるといつて差支ない。

⁵ これらは「先生のご{馳走/世話/供}になる」のように「ノ格」に対する敬意であると考えられる。

御... になる	... なさる	御... なさる
978	924	1,599

Table 2: 分析対象の事例

ている事例が該当する。「着せる」の語彙素が「着（ちゃく）する」と表記されているものがあり、これらを「着（き）せる」へ改めた。⁶

最後に、頻度が低い事例を排除した。これは、以下の議論では相対頻度に基づいて考察を進めるためである。例えば、ある動詞がコーパスの中で 100 回使われてすべて「御... になる」で使用された場合と、ある動詞が 1 回しか使われずそれがたまたま「御... になる」という形を取った場合、相対頻度を計算した場合は、どちらも得られた事例全てが「御... になる」を取った「御... になる」指向の強い動詞であるかのように解釈されるが、低頻度の動詞の場合はこの結論が必ずしも成り立つとは限らない。全体で、25 回以上使われた動詞に限定して議論を進める（これらの動詞は表 4 に掲載されている）。

以上の修正の下、Table 2 にその内訳の示された 3,411 件を分析の対象とする。

4. 記述統計学：ヒストグラム・散布図

本節では、統計的推測を行う前に、ヒストグラムと散布図を利用して分布の特徴を分析し、得られたデータの基本的な性質の予備考察を行う。

4.1. ヒストグラム

Table 2 に示されたヒストグラムは、これら三形式が時代ごとにどのくらい採取されたかを示している。なお、三形式を全て足し合わせた全体の採取頻度を表した左上の図は、CHJ における資料の時代間の偏りを示唆している。例えば、1830-1840 年代は前後の年代と比べ突出して尊敬語の採取件数が多いが、これはこの時期に特別に尊敬語が多用されたというよりは、入手可能な資料が豊富に存在していることを反映しての結果と解するのが妥当である。程度の差こそあれこのような入手可能資料の大小は他の年代にも当てはまり、このヒストグラムの形状をそのまま使用頻度の増減と解釈することには慎重であるべきである。

しかし、以上の留保を念頭に置いたうえでも、以下の点は確かであろう。(i) 先行研究の指摘通り、「御... になる」は明治期に入ってから登場し勢力を伸ばした構文である。(ii) CHJ において「なさる」「御... なさる」が確認されるのは 18 世紀に入ってからである。先述の Fig 1 に見られた観察の通り、先に「御... なさる」が使用されはじめそれを後追いつける形で「... なさる」構文が使われ始めたことが伺える。(iii) 明治期以降は、「... なさる」の使用が他の構文に比した割合では減少している。これは、Fig 2（左図）の結果とも符合する。

⁶

(i) これは和洋服何れの上へ お着せになつても可愛いものです。

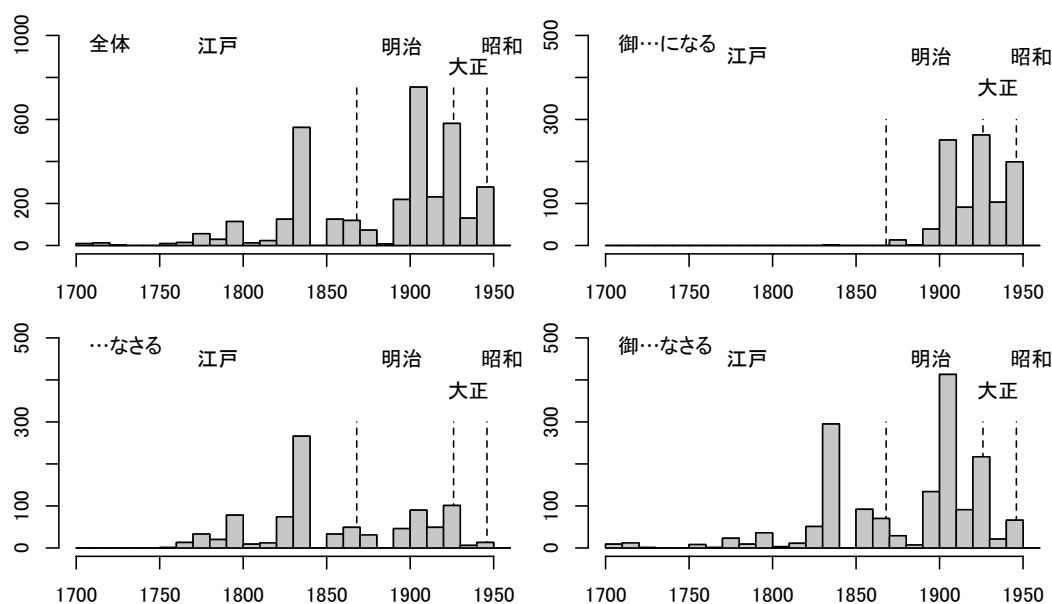


Fig. 3: CHJ における尊敬語形式の時代別採取件数の推移

4.2. 散布図

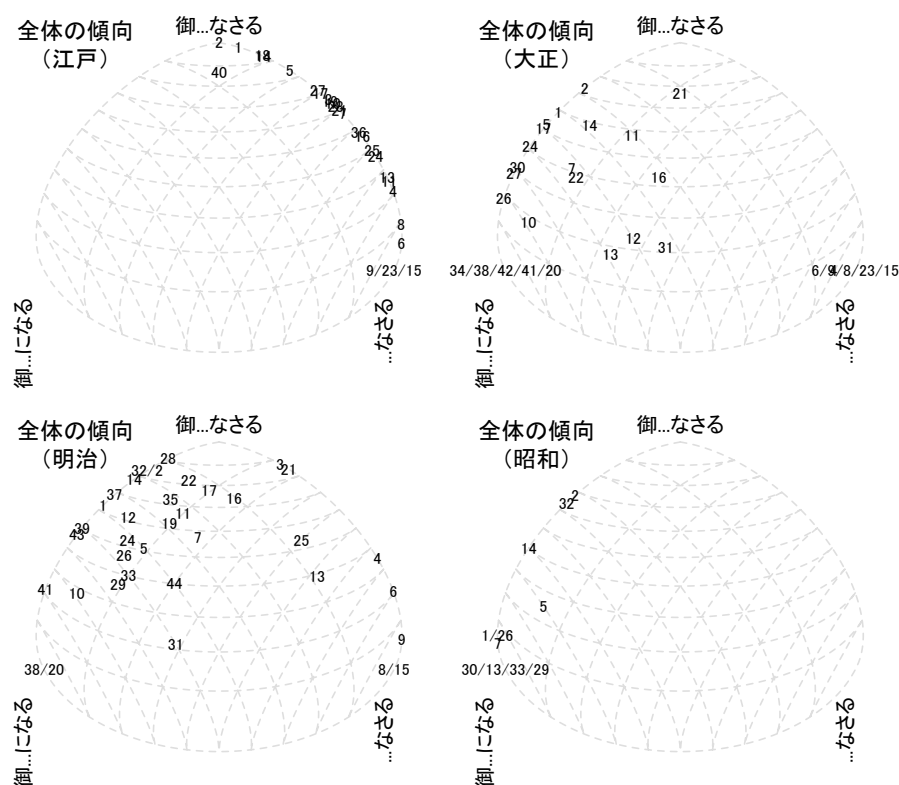
次に、散布図を用いて、各動詞が時代ごとにどのように使用されたのかを分析する。相対頻度に直された各動詞は、二次元の単体 (simplex) 上に分布する。この単体上の点間の距離を選択するためにはユークリッド距離では解釈が不自然になる (Yamada 2017)。そのため、ここでは距離尺度としてヘリンジャー距離 (Hellinger Distance) を用い、これに基づく散布図を Fig 4.3. に掲載した。以下の点が重要な着眼点として指摘できる。

一つ目は、明治以前の構文選択における動詞間の違いである。江戸時代は「御... なさる」と「... なさる」の二つの構文の間でバリエーションが形成されていたことは先述の通りだが、先ほどのヒストグラムからでは全ての動詞でこの二つの尊敬語形式に自由度があったのか、それとも一部の動詞だけが二種類の形式を取っていたのかについての情報はなかった。しかし、この散布図から、動詞ごとに構文選択の傾向に差が存在することが明確に示された。

二つ目は、明治期の「御... になる」構文の登場に際して生じた動詞間の傾向の違いである。江戸時代の散布図と明治時代の散布図を比較すると、多くの動詞で尊敬語形式の選択傾向に変化が生じ、分布の中心が図の左側にシフトしていることがわかる。しかしながら、すべての動詞で「御... になる」指向が強まったのかというとそうではなく、大局的な言語体系の変化に抗う形で、「(御) ... なさる」を取り続けている動詞も少数ながら存在することが確認できる。昭和期は CHJ の用例数が少ないため、情報が少ないが、大正期以降も同様の傾向が継続していると見ることもできる。

4.3. 限界：統計モデリングの必要性

一見、Fig 4.3. に描かれた散布図は、(3)b で提示したリサーチクエスションへの答えを与えてくれそうに思われるが、事はそう単純ではない。これは、記述統計的な整理には、以下の問題点が生じてしまうからである。



ID	動詞	ID	動詞	ID	動詞	ID	動詞
1:	出でる (416)	12:	出る (74)	23:	行く (50)	34:	帰る (32)
2:	覧 (409)	13:	思う (73)	24:	出す (50)	35:	有る (31)
3:	呉れる (208)	14:	休む (71)	25:	置く (48)	36:	案ずる (30)
4:	言う (183)	15:	来る (70)	26:	尋ねる (46)	37:	書く (30)
5:	返る (159)	16:	遣る (68)	27:	呼ぶ (44)	38:	見える (29)
6:	為る (142)	17:	入る (61)	28:	止す (41)	39:	立つ (28)
7:	聞く (124)	18:	上がる (60)	29:	通る (41)	40:	泊まる (27)
8:	見る (113)	19:	成る (57)	30:	考える (40)	41:	召す (27)
9:	居る (102)	20:	分かる (56)	31:	付ける (39)	42:	作る (27)
10:	話す (78)	21:	待つ (54)	32:	掛ける (34)	43:	出掛ける (26)
11:	仕舞う (76)	22:	上げる (52)	33:	笑う (33)	44:	怒る (25)

Fig. 4: 分析対象の動詞（括弧内は素頻度）

第一は、母集団と標本のずれである。あくまでこの図は標本（CHJ）の性質を表しているだけで、本当の興味関心である後期江戸語から明治期の日本語の性質がどの程度推測できるのかということに関する情報は含まれていない。標本に基づいて母集団を推測することが望ましい。

第二は、変数の間の関係（交絡）の可能性である。例えば、Fig 4 の大正時代では 23 番の動詞（行く）が他のいくつかの動詞とともに「... なさる」構文を多くする動詞として分布していることが分かるが、果たしてこの分布上の性質がこの動詞の独自性を反映しているのか、それとも別の要因に起因するのかまではわからない。可能性として、この CSJ に採取された

サンプルの「行く」の大部分が命令文で使われていたために、「...なさる」構文が指向された可能性もある。このように、散布図だけでは変数の交絡を判断することは難しく、統計モデルを立て、互いの影響を排除したもとの各要因の効果量を推定することが求められる。

これらの欠点を補うために、次節では統計モデルを立て、パラメータを推定しその結果に基づいて議論を深めていく。

5. 推測統計学：統計モデリング

5.1. モデル

本研究では、先行研究の知見を反映し表3に示された変数を統計モデルに組み入れる。尊敬語の選択は離散値を取るため、 j 番目の動詞が、 k 番目のジャンルで用いられた i 番目の事例の応答変数は、パラメータ $\mathbb{I}_{i(jk)}$ カテゴリカル分布に従うと仮定し、このパラメータは逆ロジット関数で線形モデルに結びつけられていると想定する。また「モーラ」という変数は、事例レベル（Population level）ではなく、動詞というグループレベル（Group level）における固定効果として扱うことが妥当であるため、階層構造を想定し、下記のような交差分類型（Cross-classified）の混合効果モデルを用いて分析を行う。

応答変数のベースラインには「御...になる」を据え、「...なさる」との比に関する説明変数には肩に添え字として b を、「御...なさる」との比較に使用する説明変数には肩に添え字として c をつけて区別している。

$$(7) \quad y_{i(jk)} \sim \text{Categorical}(\mathbb{I}_{i(jk)})$$

$$\mathbb{I}_{i(jk)} = \text{inv_logit}(\eta_{i(jk)})$$

$$\eta_{i(jk)} = \begin{pmatrix} 0 \\ \gamma_{\text{切片}}^b + \gamma_{\text{モーラ}}^b w_{\text{モーラ } j} + \beta_{\text{命令強}}^b x_{\text{命令強 } i(jk)} + \beta_{\text{命令弱}}^b x_{\text{命令弱 } i(jk)} \\ + \beta_{\text{江戸}}^b x_{\text{江戸 } i(jk)} + \beta_{\text{大正}}^b x_{\text{大正 } i(jk)} + \beta_{\text{昭和}}^b x_{\text{昭和 } i(jk)} + u_{\text{動詞 } j}^b + u_{\text{ジャンル } k}^b \\ \gamma_{\text{切片}}^c + \gamma_{\text{モーラ}}^c w_{\text{モーラ } j} + \beta_{\text{命令強}}^c x_{\text{命令強 } i(jk)} + \beta_{\text{命令弱}}^c x_{\text{命令弱 } i(jk)} \\ + \beta_{\text{江戸}}^c x_{\text{江戸 } i(jk)} + \beta_{\text{大正}}^c x_{\text{大正 } i(jk)} + \beta_{\text{昭和}}^c x_{\text{昭和 } i(jk)} + u_{\text{動詞 } j}^c + u_{\text{ジャンル } k}^c \end{pmatrix}$$

5.2. 推定方法

本研究では、ベイズ推測の枠組みで (7) に含まれるパラメータの推測を行う。Stan で記述したモデルを R から呼び出し実行し、ハミルトニアン・モンテカルロ法（Hamiltonian Monte Carlo）により推定を行った（R Core Team 2020; Stan Development Team 2020）。イテレーションの回数は 20,000 回、そのうち最初の 19,000 回をバーン・イン期間として事後分布の計算からは取り除いた。また、まびき（thinning）を 11 回ごとに設定し、収束を改善させた。走らせたチェーンは 4 つであり、各チェーンから 91 個ずつ、合計 364 個の事後分布からのサンプリングを得た。 \hat{R} の値が全て 1.03 以下に収まっていることを以て、これらが収束していると判断を下した。

変数名	性 質	レベル	内容
y	応 答	事例	選択された尊敬語構文を表す。ベースラインは「御... になる」に設定した。
$x_{\text{命令強}}$	説 明 (固定)	事例	要求という発話行為を持つ命令文を示す変数(例: 走りなさい)。事例レベルの固定効果として扱い、当てはまれば 1、それ以外は 0 として扱う。
$x_{\text{命令弱}}$	説 明 (固定)	事例	依頼など弱い要求に関わる発話行為を持つ命令文を示す変数(例: お走りになってください)。事例レベルの固定効果として扱い、当てはまれば 1、それ以外は 0 として扱う。
$x_{\text{江戸}}$	説 明 (固定)	事例	成立年が江戸時代に含まれるもの。事例レベルの固定効果として扱い、当てはまれば 1、それ以外は 0 として扱う。
$x_{\text{大正}}$	説 明 (固定)	事例	成立年が大正時代に含まれるもの。事例レベルの固定効果として扱い、当てはまれば 1、それ以外は 0 として扱う。
$x_{\text{昭和}}$	説 明 (固定)	事例	成立年が昭和時代に含まれるもの。事例レベルの固定効果として扱い、当てはまれば 1、それ以外は 0 として扱う。
$w_{\text{モーラ}}$	説 明 (固定)	グループ	一モーラか否かを示す変数。動詞というグループレベルの固定効果として扱い、一モーラであれば 1、それ以外は 0 として扱う。
$u_{\text{動詞}}$	説 明 (変量)		尊敬語構文が接続する本動詞(表 4 に挙げられた動詞)。グループを作る変量効果として扱う。
$u_{\text{ジャンル}}$	説 明 (変量)		各文が生成されたジャンルを表す変数。グループを作る変量効果として扱う。

Table 3: 変数一覧

5.3. 推定結果

5.3.1 固定効果の分析

Fig 5 は固定効果に対して想定した事前分布と推定結果として得られた事後分布を示している。事前分布としては弱無情報分布として平均が 0、標準偏差が 5 の正規分布を用意した。⁷ 図において末尾に b が付いているものが、「... なさる」対「御... になる」の比率をモデル化したものであり、末尾に c が付いているものが、「御... なさる」対「御... になる」の比率をモデル化したものである。三段階で表された灰色の濃淡は、それぞれ外側から 95%ベイズ確信区間、80%ベイズ確信区間、33%ベイズ確信区間を示している。中央の太線は MCMC サンプルの事後中央値を表している。値が負に振れれば振れるほど「御... になる」の割合が高く、正に振れれば振れるほど他方の構文への選択度合いが強まることを表している。

⁷ ロジスティック回帰の(偏)回帰係数は exp の肩の上に乗っているため値が 3 や 4 でもきわめて強い効果を持つ。特別なことがなければ 10 や 20 といった値が生じることはなく、この程度の幅の事前分布であっても十分機能する。分散を狭めることで収束時間の短縮や後述のように完全分離問題にも対応ができる。

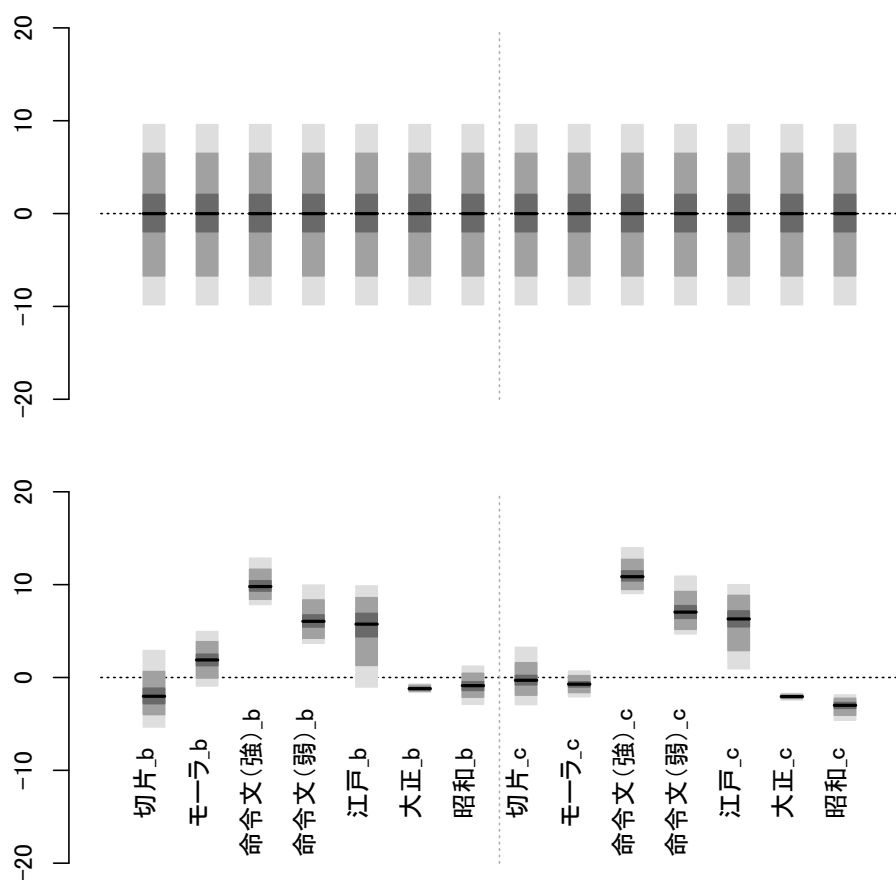


Fig. 5: 事前分布（上）と事後分布（下）の比較

ここから読み取れることは以下の通りであり、これらが(3)aに掲げたりサーチクエスチョンに対する本稿の答えとなる。第一に、切片は、基準とされた参照カテゴリーのふるまいとして解釈される。すなわち、非一モーラ語であり、非命令文であり、明治期に産出された事例のふるまいである。「御... なさる」構文との比較においては、値はほぼゼロを指し、他の要因がコントロールされたもとでは、両者の間に選択傾向の差はないと言える。一方、「... なさる」構文との比較においては、確信区間がゼロを含んでいることもあり、このCHJのデータだけから確定的な判断をすることには慎重さが求められる。ただし、分布全体は負の方向に傾いていることを鑑みれば、他の要因がコントロールされたもとでは「... なさる」よりも「御... になる」の方が明治期の平叙文（等）では選択される傾向があったという蓋然性がある。

第二に、モーラについては、他の要因を統制した際の一モーラ語の効果量は、「... なさる」構文との比較において、正に少し振れている。正に振れることは、先行研究の知見とも合致する。すなわち、「来る」「する」などの連用形が一モーラとなる動詞は「御」という接頭辞と接続することができない（例：*御来になる）という現代日本語に存在する制約が明治期においても存在していたことを示している。しかし、確信区間の幅 80%確信区間でゼロを

含んでおり、CHJ のデータからでは確定的な物言いにとどり着けるほどの確かさが無いとも言える。この結果は、一見すると不思議かもしれない。なぜならばこのモーラ制約は極めて強い制約だと考えられているため、コーパスのデータでもこれを反映した結果になるのではないかと予測されるためである。実は、この理由は動詞間のバリエーションに起因する（詳細は 5.3.2 節で後述する）。確かに大部分の一モーラの動詞ではこのモーラ制約が順守されているのだが、一モーラでありながらこの制約を満たさないものが存在するのである。このため、モーラ制約という固定効果ではそこまで明確な差が現れず、動詞という変量効果の残差の方にモーラ制約が現れる結果となっている。一方で、「御... なさる」構文のモーラに関する効果量はやや負に振れてはいるものの、これもゼロに近い値を取っている。これはモーラ制約は、同じ「御」という接頭辞を取る以上「御... になる」同様「御... なさる」構文にも等しく影響を与えているためであると考えられる。

第三に、命令文の強弱については、極めて明確な正の効果が存在している。これは、(2) に示したように「御... になる」がその他の二つと異なり命令形を取れないことに起因する。結果として、完全分離問題が生じる。すなわち、事前情報がない状態では、係数の値を 100 にすべきか 1,000 にすべきか判断が付かなくなるため、推定の収束に困難が生じるのである。しかし、ベイズ統計学では事前分布がこの完全分離問題を緩和させる機能を持ち、本研究では上述のように推定が収束している。ただし、完全分離が生じている状況を踏まえると、命令文に関しては、結果として出てくる値、すなわち 10 や 7 といった値そのものに実質科学的な意味があるというように解釈をするよりも、極めて強い効果（断定的な効果）を持っていることが示されたと判断をする方が望ましいだろう。

第四に、時代間のずれであるが、まず江戸時代はどちらの構文でも事後中央値が命令文並みに強い正の値を示している。「御... なさる」が「... なさる」構文よりも狭い確信区間を持っているのは、Figs 1~3 で見てきたように「... なさる」構文の使用頻度が江戸時代において「御... なさる」構文より少ないからではないかと考えられる。大正時代は、「御... なさる」「... なさる」の両方で、明確な負の効果量が得られている。すなわち、非命令文の非一モーラ動詞では、明治期と比べて大正期においてより「御... になる」を取る傾向が強まっていることを示している。昭和期でも同様の傾向が見られるが、昭和期は大正期と比べてサンプルサイズが小さいことを反映し、確信区間がやや広めに出ている。

以上のことから、この統計モデルの結果は先行研究が緻密に検証してきた点を全て適切に把握しているものであることが見て取れる。統計的分析と、非統計的分析の結果が一致した結論を示している点で、先行研究の調査も、そして、本稿の統計モデルの分析もお互いの信頼性を互いに高め合っていると言える。しかしそれだけではなく、以下に見るように、この統計モデルは、先行研究の分析よりもより詳細な情報を提示しているより情報量が高い。

5.3.2 変量効果の分析 (I) : 動詞

固定効果の分析から本稿の統計モデルの妥当性が検証されたところで、これまでの先行研究では未解明であった動詞間のばらつき、ジャンル間のばらつきといった問題、すなわち (3) の残りのリサーチクエスションに対しての検討を行う。

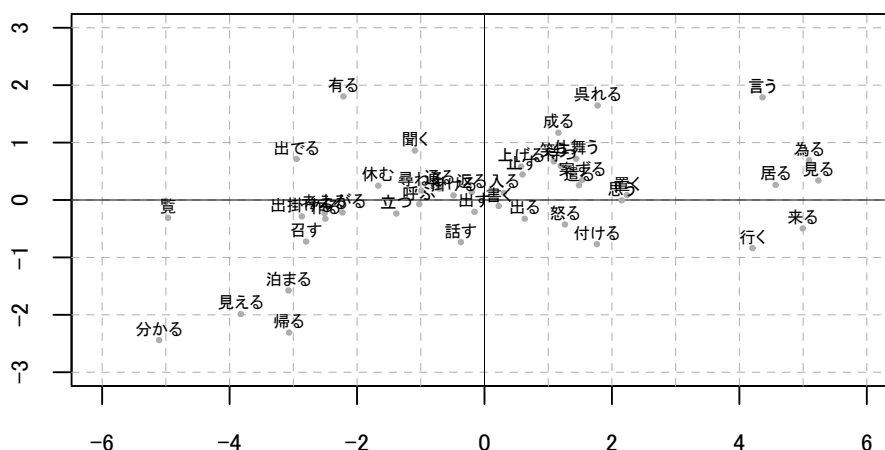


Fig 6 は、「... なさる／御... になる」における動詞の独自性 ($u_{動詞 1}^b, \dots, u_{動詞 44}^b$ の事後中央値；横軸) と、「御... なさる／御... になる」における動詞の独自性 ($u_{動詞 1}^c, \dots, u_{動詞 44}^c$ の事後中央値；縦軸) によって、CHJ で用いられていた 44 つの動詞の選択傾向を示した残差プロットである。散布図という点では Fig 4 と同じだが、Fig 4 とは異なり、先述の固定効果の影響を取り除いた後の動詞の独自性が示されている点で、Fig 4 の限界が乗り越えられている点に注意されたい。

図の原点に近い動詞ほど、その挙動が先の固定効果によってほとんど説明尽くされた、その意味でモデルの予測通りの挙動を示す動詞である（例：出す、書く、入る…など）。一方で、原点から離れた位置に存在する動詞は、先の固定効果だけでは説明できない独自性を持った動詞であり、注目に値する。紙幅の都合上、全てに注目をするのではなく、 x 軸においてその絶対値が4を超える動詞たちに限定をして、それらの独自性を考察する。

「...なさる」指向の強い動詞 1 : 一モーラの動詞

横軸の値が極端に正に振れている動詞たちがある。これらは、「御...になる」と「...なさる」を比較したとき、極めて高い割合で「...なさる」を取る動詞である。この中で「為(す)る、見る、居る、来る」は、実は一モーラの動詞である。モーラか否かはグループレベルの固定効果としてモデルに組み込まれていたもので、動詞の独自性として「...なさる」指向が残っていることに違和感を覚えるかもしれないが、実はこれは一モーラでありながら「...なさる」指向を持たない動詞が存在していることに起因していると考えられる。すなわち、「出(で)る」という動詞である。この動詞は、現代日本語でも「電話にお出(で)にならなかった」のように「御...になる」構文での使用が可能であり、CHJ から例を拾うと次のようなものがある。

- (8) スリツパ召して御廊下に御出になる。(60M 女世 1909_16077、さゆわ子『女学世界』, 1909 年)

	御…になる	…なさる	御…なさる
言う	0	28	12
行く	1	3	0
分かる	26	0	0

Table 4: 独自性の強い動詞たち（明治時代の非命令形での使用頻度）

この「出る」という動詞の存在は、これまで定説とされてきた「一モーラ制約（「御」という前接辞は連用形が一モーラになる動詞には使用できない）」という一般化に対する重要な反例となる。なぜ「為（す）る、見る、居る、来る」は一モーラ制約に従い、「出る」は従わないのかについて今後の研究で理論的に分析がなされることが俟たれる。

「…なさる」指向の強い動詞 2：その他

その他の「…なさる」指向が高い動詞に「言う」「行く」が挙げられる。Table 4 に、これらの動詞の明治時代における非命令文における頻度を示している。明治時代は多くの動詞が「御…になる」への依存度を強めた時代であり、非命令形という「御…になる」が生産的に使われた環境での使用頻度を観察することで、いかにこれらの動詞が大局的な構文選好傾向と異なるふるまいを示しているかが見て取れる。⁸ この二つの動詞の共通点としては、(i) 母音の「い」で始まること、(ii) 補充形尊敬語を持つことがすぐに挙げられるが、これらの要素が果たして、「…なさる」指向をもたらした直接的な要因であるかについては今後の考察が俟たれる。⁹

というのは、第一の点については、本研究で検討した動詞の中には「出でる」という「い」で始まる動詞がもう一つ含まれており、この動詞はむしろは「御…になる」への指向が強い。このため、音韻的な理由で「御」と衝突があるとは考えにくいからである。第二の補充形については、もちろん次のような理由付けが可能ではある：「江戸時代にすでに補充形が存在していたため、補充形を持つ動詞たちは明治時代に登場した新しい構文である『御…になる』を採用する必要がなかった。そのため、補充形を持たない動詞たちとは違い『御…になる』の頻度が増えなかった」。要は「事足りていたからだ」という仮説であるが、この仮

⁸ 「言う」と「行く」が「御…になる」で使われた事例は CHJ においては、下記の一例ずつである。ただし、「行く」については「お嫁に行く」という意味で「嫁く」という漢字が使用され、振り仮名として「イク」が使われている点で例外的なものと言えよう。

- (i) 言う
私を呼んで沁々とお云ひになつたことを考へますとね、(60M 婦俱 1925_06098 『婦人倶楽部』, 本田美禪 (作))
- (ii) 行く
安藤さんへ御嫁きになりまして後も、(60M 女世 1909_05010 『女学世界』 1909 年、三浦環 (作))

⁹ 尊敬語に補充形を持つ非一モーラの動詞としては、「呉れる>下さる」という動詞が存在し、この動詞も「おくれになる」という形式では使いにくい。ただし、「…なさる」「(御) …なさる」という形式では「おくれなさる」ではなく「おくんなさる」という音韻変化が生じることもあり、この動詞の独自性については詳細な検討が必要である。なお補助動詞用法「てくれる」については、山田 (2015) などに先行研究のまとめや分析がある。

説には次の検討課題がある。確かに、補充形が非補充形の登場をブロックするというブロック効果（Blocking Effect）は形態論の教科書的分析ではあるが、もしそうだとすると「なぜ補充形の存在は『...なさる』や『御...なさる』をブロックしないのか」という点が説明されなければならない（例：行きなさる、お言いなさる）。また、「言う」と「行く」を比較すると、(9)が示すように「言う」にだけは「御...なさる」の選択傾向が強い。この差がなぜ生じるのかという点についても考察が深められなければならない。なお、仮に補充形を要因として統計モデルに組み込む場合は、「モーラ」との相関が高いため多重共線性の問題が生じることが予想されるためモデルを立てる際に工夫が求められるだろう。¹⁰

- (9) 何處を歩いて居ただなんて、阿母さんも随分酷い事をお云ひなさるのね。』（60M 太陽 1901_01029、『太陽』1901年、広津柳浪（作））

「御...になる」指向の強い動詞：動詞「分かる」

これまで見てきた動詞とは対照的に、今回の分析で「御...になる」指向が最も強いと解釈されたのは「分かる」である。本コーパスでは1874年から1847年にかけて合計56例の使用が観察されるが、その全てが「御...になる」で使われている。現代日本語においても「お分かりになる」「分かりなさる」「お分かりなさる」の三つを比較してみると、非文とまでは言い切れないものの「分かりなさる」「お分かりなさる」の容認度は「お分かりなる」よりもはるかに低い。これまでの先行研究では、上に見た補充形やモーラ制約という問題については議論がなされていたが、この「分かる」の傾向を説明できるような要因は提案されていない。そこで以下複数の可能性について検討を行う。

第一に、「分かる」のような意志性（Volitionality）が低い認識動詞では、相対的に意志性の強い「する」起源である「...なさる」「御...なさる」よりも、相対的に意志性の弱い「なる」起源の「御...になる」との親和性が高いのではないかという、動詞の意味に注目した仮説が考えられる。しかし、Table 5に見るように明確な相関関係はみられない。

第二に、「分かる」の持つ特殊な格配列が影響を与えているのではないかという、動詞の統語的な特徴に注目した仮説が考えられる。すなわち、「分かる」は、「BがAを[動詞]」という形式を取る「知る」「理解する」などとは異なり「AにBが分かる」という構文を取る。この「二格主語」が「御...になる」指向を強めているのではないかという可能性である。類似した構文には、「見える‘can see’」「出来る」「知れ渡る」「有る」などがあり、本コーパスではTable 6に示された動詞が観測されている。上記の仮説が正しければ、これらの動詞ですべて「御...になる」指向が見られるはずである。Fig 6とTable 6から「分かる」に加え「出来る」「見える」という「二格主語」であり、かつ「可能」の意味をその語彙の意味に

¹⁰ むしろ、一モーラの動詞は「御」との使用ができないため、補充形が発達したと見ることができる。この場合は、要因間に明確な因果関係があるので補充形だけを説明変数に入れて結果を解釈する方法が望ましい場合がある。しかし、共時的に有効なこのブロック効果による説明が果たして歴史的にどこまで正しいのかについては検討が必要である。この解釈が正しければ『御...なさる』の成立の後に『御見になる』などが排除されるため『御覧になる』という表現が求められ成立した」というような予測となる。これが本当に当てはまるのかどうかについては歴史的な考証が欠かせないからである。

	御…になる	…なさる	御…なさる
分かる	56	0	0
知る	3	14	6
気付く	2	0	0
発見	0	2	0
理解	0	1	0

Table 5: 「分かる」と意味的に類似した動詞（非命令形での使用頻度）

	御…になる	…なさる	御…なさる
分かる	56	0	0
有る	8	0	22
出来る	12	0	8
見える	27	0	2

Table 6: 二格主語を取る動詞（非命令形での使用頻度）

持つ動詞にこの傾向は成立していると言っていだろう。¹¹ただし、「可能」の意味を持たない「二格主語」の可能な「有る」に関してはこの傾向ははっきりと出ているわけではない。「御…なさる」の使用が大きい点で、「分かる」とは異なる性質を表していると言え、仮説が十全に指示されたとは言い難い。

上記の考察を踏まえ、本稿では Fig 6 で顕著に示された「分かる」の「御…になる」指向は、その他の動詞にも共通して存在する明確な要因の存在を示しているというよりはこの動詞の独自性として扱われるべきだと、結論づける。

5.3.3 変量効果の分析（Ⅱ）：ジャンルの独自性

Fig 7(左図)は、「…なさる／御…になる」におけるジャンルの独自性($u_{\text{ジャンル } 1}^b, \dots, u_{\text{ジャンル } 6}^b$ の事後中央値;横軸)と、「御…なさる／御…になる」におけるジャンルの独自性($u_{\text{ジャンル } 1}^c, \dots, u_{\text{ジャンル } 6}^c$ の事後中央値;縦軸)によって、CHJ で用いられていた 6 つのジャンル区分の傾向を示した残差プロットである。

原点から最も遠く離れた位置に離れた、とりわけ横軸方向への逸脱が見られるジャンルに「国語教科書」である。第三象限に位置していることから「御…になる」への強い指向性を持つジャンルであり、その頻度の内訳は Fig 7 (右図) に示される通りである。この国語教科書における「御…になる」の使用が多いという事実は、次の二つの点で重要である。

第一に、国語教科書での多用は、他の尊敬語表現と比べた時に「御…になる」が高い敬意を持つ表現であった可能性を強く示唆している。本研究で対象とした明治期から昭和期の国語教科書には、天皇の権威付けを目的として、神話などが数多く掲載されている。下記の例文が示すように天皇に対して用いる敬語に「御…になる」形式が多用されていることを踏まえると、この新規形式が高い敬意を示す表現と認知されていた可能性が高い。

¹¹ Table 6 で挙げられた動詞については、多義に注意を払う必要がある。特に「見える」では come の意味と can see の二つの意味があり、さらなる考察には細かい分類が必要である。ただし、仮にこの多義を整理したとしても、「見える」に「御…になる」指向が強いことは変わらないであろう。

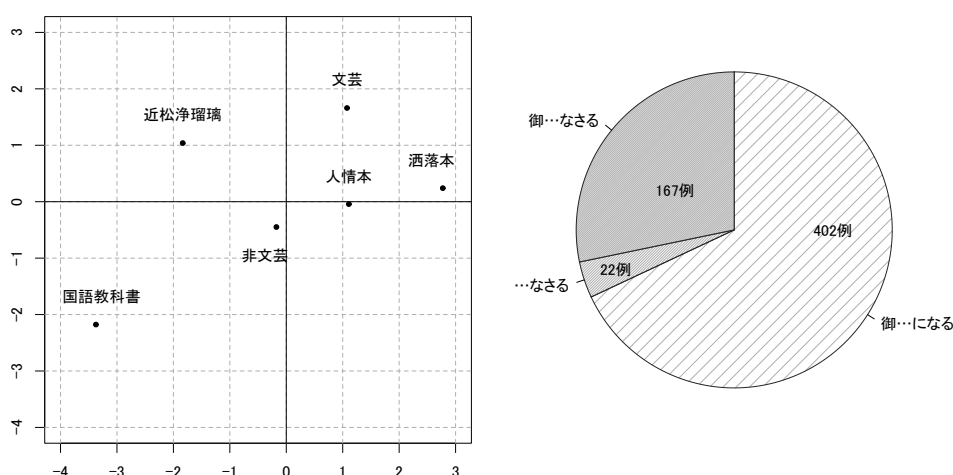


Fig. 7: ジャンルの残差プロット（左）と「国語教科書」における使用頻度の内訳（右）

	1904	1910	1918	1933	1941	1947
御...になる	31	19	50	103	111	88
...なさる	0	0	3	6	3	10
御...なさる	39	19	22	21	26	40

Table 7: 国語教科書における年代別の使用状況

- (10) a. 神武天皇は、このわるものどもをせめに、おいでになって、とうとう、まかして おしまひになりました。(60T 小読 1904_12B15、『小学校国語 1 期』、1904 年)
- b. この神さまはさきほどおとほりになつた神さまがたの弟の方です。(60T 小読 1910_22B18 『小学校国語 2 期』、1910 年)

先行研究では「御... になる」の明治期における敬意の高さは指摘があり（原口 1974:23）、ジャンルに注目した Fig 7（左図）の残差プロットはこの指摘を別の角度からも傍証したものだと言える。

第二は、国語教科書に掲載されたという点で、新規尊敬語標識である「御... になる」が登場して間もない明治時代からすでに規範的表現と見なされていたことが推察される。表 7 が示すように、「御... になる」は、1904 年（明治 37 年）の国語の教科書においてすでに尊敬語の代表格として使用されている。同様の指摘は辻村（1968）などにも見られ、本形式の広まりの早さがうかがい知れる。また、本稿の対象資料ではないが、明治政府によって沖縄に対する同化政策の一環で作成された「沖縄対話」（明治 13 年）という「標準語」の教科書においても「御... になる」が多用されていることが報告されており（辻村 1968、山田 1959、山田 2015）、この表現が明治期においてすでに規範的表現として目されていた可能性は高いと考えられる。

6. まとめと将来研究への示唆

本研究は、尊敬語の競合に関する先行研究で指摘されてきた観察結果を、統計モデルによって裏付けつつも、これまでには明らかにされてこなかった細かい動詞やジャンルの独自

性を明らかにしてきた。

第一のリサーチクエスションは、「他の要因の影響を統制した下での、各要因（例：モーラ制約、命令文の影響）の効果量はどの程度なのか」であった。他の変数の影響を除いた下で、最も大きな効果量を示したのが事例レベルの固定効果である文のモード（命令形か否か）である。また、時代やモーラに関しても一定の効果量を示していることが確認された。

第二のリサーチクエスションは、「動詞ごとの独自性はあるのか」というものであった。一モーラでありながら「御...になる」構文を取る「出る」、全体的な傾向に逆らい二モーラでありながらも「御...になる」の頻度が低い「言う」「行く」、そして他の要因からは説明が付きにくい極めて強い「御...になる」指向を見せる「分かる」といった動詞が指摘され、大局的な構文選択の変化を時代ごとに追う必要性に加え、独自の動詞の挙動に注目する意義が示された。

第三のリサーチクエスションは、「ジャンルごとの独自性はあるのか」というものであった。特異な挙動を示すジャンルとして「国語教科書」が指摘され、そこから、「御...になる」の敬意の高さ、規範性などがうかがい知れることが指摘された。

この分析結果は、今後の研究に対して、以下の示唆を持つ。第一に、日本語の尊敬語研究に対しては、(i) 文のモードとのかかわり、(ii) 動詞の独自性という二点がとりわけ重要な示唆となろう。(i) に関しては、一見すると、平叙文なのか疑問文なのか、はたまた命令文なのかという文のモードの問題は、主語に敬意を払うのか否かという点とは独立した文法項目に見えるかもしれない。しかし、命令文は主語の指示対象が聞き手と一致する文のモードであり、主語に関する発話行為に関わる点で、尊敬語とのつながりが生じる（Yamada 2019）。聞き手に対する発話行為に関わるという点では、「です／ます」のような聞き手敬語の存在も重要であり、今後の研究で三者の理論的なつながりについて整理されることが望まれる。

第三に、コーパスデータを用いた定量的言語変異に対して、階層レベルを組み入れた混合効果モデルを取り入れることの重要性が指摘できる。本稿の考察の中心をなした「動詞ごとの独自性」「ジャンルの独自性」は、変量効果をモデルに組み込むことにより、より精緻に分析ができる。尊敬語に限らず、類似した構文選択の問題を扱う場合に同種の方法が有効であると言える（Levshina 2016、2018）。

文 献

- [1] Alok, D. (to appear) The morphosyntax of Magahi addressee agreement. Manuscript.
- [2] Andrews, J. R. (1975) *Introduction to classical Nahuatl*. Austin: University of Texas Press.
- [3] Broadwell, G. A. (2019) Honorific usage in Timucua exempla. In *Preaching in New Worlds*, edited by T. Johnson, K. W. Shelby, and J. D. Young. New York: Routledge.
- [4] Cedergren, H. J. and D. Sankoff. (1974) Variable rules: performance as a statistical reflection of competence. *Language* 50/2: 333-355.
- [5] Guy, G. R. (1991) Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* 3(1): 1-22.
- [6] 原口裕 (1974) 「『おーになる』考」続貂『国語学』 96: 23-32.
- [7] 菊地康人 (1997 [1994]) 『敬語』講談社学術文庫.
- [8] 金田一京助・知里真志保 (1936) 『アイヌ語概説』岩波書店.

- [9] 国立国語研究所 (2020) 『日本語歴史コーパス』(バージョン 2020.3, 中納言バージョン 2.5.2) https://pj.ninjal.ac.jp/corpus_center/chj/ (2020 年 3 月 31 日確認)
- [10] Koshal, S. (1987) Honorific systems of the Ladakhi language. *Multilingua* 6/2: 149–168.
- [11] Launey, M. and C. Mackay (2011) *An introduction to Classic Nahuatl*. Cambridge: Cambridge University Press.
- [12] Levshina, N. (2016) When variables align: a Bayesian multinomial mixed-effects model of English permissive constructions. *Cognitive Linguistics* 27/2: 235–268.
- [13] Levshina, N. (2018) Probabilistic grammar and constructional predictability: Bayesian generalized additive models of *help* + (*to*) Infinitive in varieties of web-based English. *Glossa: a journal of general linguistics* 3/1/55: 1–22.
- [14] 村上謙 (2003) 「近世後期上方における連用形命令法の出現について」『国語学』54/2.
- [15] 村上謙 (2005) 「近世上方における補助動詞ナサルの変遷」『国語国文』74/2.
- [16] R Core team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- [17] Stan Development Team (2020) *RStan: the R interface to Stan*. R package version 2.21.2. <http://mc-stan.org/>.
- [18] Svahn, A. (2016) *The Japanese imperative*. Lund: Lund University Publications.
- [19] 辻村敏樹 (1968) 『敬語の史的研究』東京堂
- [20] Yadav, R. (1996) *A reference grammar of Maithili*. Berlin, Mouton de Gruyter.
- [21] 山田巖 (1959) 「明治初期の文献にあらわれた尊敬表現『お(ご)…になる』について」『ことばの研究』1: 201–214.
- [22] 山田里奈 (2013) 「明治 20 年代までにおける〈する・なる〉の尊敬表現形式—『お～なさる』、『～なさる』、『～だ』系を中心に—」『早稲田大学大学院教育学研究科紀要 別冊 21/1: 115–127.
- [23] 山田里奈 (2015) 『近世後期江戸語から明治期東京語における尊敬語表現研究』博士論文, 早稲田大学.
- [24] Yamada, A. (2017) A Reflection on the Clustering in Corpus Linguistics. 『日本言語学会第 154 回大会予稿集』: 58–63.
- [25] 山田彬堯 (2019) 「和語起源の動詞と競合する尊敬語構文」『計量国語学会第六十三回大会予稿集』: 66–71.
- [26] Yamada, A. (2019) *The syntax, semantics and pragmatics of Japanese addressee-honorific markers*. PhD Thesis. Georgetown University.
- [27] Yamada, A. (2020a) An OT-driven dynamic pragmatics: high-applicatives, subject-honorific markers and imperatives in Japanese. In *New Frontiers in artificial intelligence: JSAI-isAI International Workshops, JURISIN, AL-Biz, LENLS, Kansei-AI Yokohama, Japan, November 10–12, 2019 revised selected papers*: 354–369.
- [28] Yamada, A. (2020b) Multinomial Mixed-Effects Models and Linguistic Variation: Competitions among Japanese Subject-Honorific Constructions. In *Proceedings of the 10th conference of the Japanese Association of Digital Humanities (JADH 2020) “a new decade in digital scholarship: microcosms and hubs”*: 33–37.
- [29] Yamada, A. (to appear) Honorificity. In *The Wiley Blackwell Companion to Morphology*, edited by P. Ackema, S. Bendjaballah, E. Bonet, and A. Fábregas.
- [30] 山崎久之 (1963) 『国語待遇表現体系の研究—近世編』武蔵野書院

日本人中国語学習者の叙述文における言語的特徴の分析 —MF/MD 法を使って—

徐 勤

大阪大学言語文化研究科
〒560-0043 豊中市待兼山町 1-8
Email: u099322c@ecs.osaka-u.ac.jp

概要 本研究では、中国語母語話者の作文データを比較対象として取り上げ、多項目・多次元法（MF/MD 法）の分析視点から、日本人中国語学習者の叙述文における言語的特徴を考察する。まず、探索的因子分析で言語項目間の共起パターンを抽出し、統計的に有効な因子を7つ採用した。そして、7つの因子を言語変異の主な次元として捉え、それぞれを「書面的 vs. 口語的」、「詳細情報の叙述」、「文脈的複雑性」、「個人の意見」、「過去の叙述」、「ロジックの繋がり」及び「行為の叙述」とラベリングした。作文データの例文を組み合わせて分析した結果、日本人中国語学習者が書いた叙述文には、口語的な傾向があり、文脈的複雑性とロジックの繋がりが不足している一方で、単純な動詞で動作を描写する傾向と、過去の出来事を現在形で叙述してしまう問題点があるということも明らかになった。

キーワード 日本人中国語学習者, MF/MD 法, 言語的特徴

日本人汉语学習者記叙文写作語言特征研究 —基于MF/MD分析法—

徐 勤

大阪大学言語文化研究科
〒560-0043 日本大阪府豊中市待兼山町1-8

摘要 本研究以汉语母語者的作文語料作為比較對象，從多維度・多特征（MF/MD）方法的分析角度出發，考察了日本漢語學習者記敘文寫作的語言特征。首先，通過探索性因子分析提取出語言項目之間的共現模式，得到了7個具有統計學意義的有效因子。其後，將7個因子作為語言變異的主要維度，分別標注為“書面性與口語化”、“敘述細節信息”、“語句複雜性”、“表述個人意見”、“敘述過去事件”、“邏輯關聯性”、“描述動作行為”。最後，結合作文文本的例句進行定性分析，結果顯示，日本漢語學習者的記敘文具有口語化傾向，語言表達缺乏複雜性和邏輯關聯性，同時，學習者還偏好於使用簡單的動詞進行動作描寫，以及誤用一般現在時態來描述過去事件的問題。

关键词 日本人汉语學習者, MF/MD分析法, 語言特征

1. はじめに

MF/MD 法 (Multi-Features / Multi-Dimensional Approach : 多項目・多次元法) とは、主に因子分析を使用することで、言語項目間 (変数間) の複雑な共起関係 (co-occurrence) を明らかにし、言語変異を多次元的に捉える手法である。この分野では、Douglas Biber の研究成果は注目に値する。Biber (1988) は書き言葉の LOB コーパス (Lancaster/Oslo-Bergen Corpus) と話し言葉の LLC コーパス (London-Lund Corpus) を基に、新聞報道・論説・伝記・政府刊行物・学術散文・推理小説・ユーモア・即席スピーチ・準備したスピーチ等の 23 ジャンル (合計 481 個のサンプル) のテキストにおける 67 種の言語項目の出現頻度を集計し、因子分析によって言語変異を解釈できる 7 つの因子を次元 (Dimensions) として捉えている。

次元 1: Involved versus Informational Production

次元 2: Narrative versus Non-narrative Concerns

次元 3: Explicit versus Situation-Dependent Reference

次元 4: Overt Expression of Persuasion

次元 5: Abstract versus Non-Abstract Information

次元 6: On-Line Informational Elaboration

次元 7: Academic Hedging

しかし、次元 7 は言語項目数も統計的な有意性も非常に小さいため、実際の分析には使用されていない。Biber (1988) では「最終的に 6 因子が次元として提案されている」(McEnery & Hardie 2011:106 (石川 2014:159) 訳)。さらに、大学の言語使用域 (university registers) における英語の口語 (例: Office hours、Study groups 等) と書面語 (例: Textbooks、Institutional writing 等) の区別を明確にできるように、Biber (2006) は T2K-SW AL コーパス (the TOEFL 2000 Spoken and Written Academic Language Corpus) を用い、MF/MD 法によって抽出した 4 つの因子を言語変異の 4 つの次元と解釈している。

次元 1: Oral vs. literate discourse

次元 2: Procedural vs. content-focused discourse

次元 3: Reconstructed account of events

次元 4: Teacher-centered stance

Biber の研究成果は言語変異の研究分野に大きく貢献している。田畑 (2005) は「Biber の言語変異研究モデルは使用域の問題だけでなく散文体の通時的変化」、「英語以外の言語における言語変異の研究」(p.192-193) にも適用されると述べている。また、Biber の研究モデルは外国語言語学習にも応用されている。例えば、Friginal & Weigle (2014) は学習者の英作文の特徴を解釈するために、Biber (1988) の視点に基づき、英語学習者の作文を分析データとして取り上げ、探索的因子分析 (Exploratory Factor Analysis) によって 4 つの次元を抽出し、L2 ライティングのマイクロ変化を分析している。

次元 1: Involved vs. Informational Focus

次元 2: Addressee-Focused Description vs. Personal Narrative

次元 3: Simplified vs. Elaborated Description

次元 4: Personal Opinion vs. Impersonal Evaluation/Assessment

近年では、MF/MD 法を使って、中国語の書き言葉と話し言葉の言語使用域に関する言語変異の研究 (Zhang, 2012; Zhu, 2015; 劉, 2019) が見られるが、管見の限りでは、中国語教育分野では、MF/MD 法による中国語学習者の作文における言語的特徴に関する研究が見当たらない。そのため、これを研究テーマとして取り組みたい。

2. 方法

2.1. 目的

中国語ネイティブスピーカーの書き言葉と比べて、日本人中国語学習者が書いた中国語作文にはどのような言語的変異が見られるのかを、MF/MD 法で考察する。

2.2. 研究対象とコーパス

本研究では、日本人中国語学習者が書いた叙述文を研究対象として取り上げる。具体的には、大阪大学の中国語専攻に在籍しており、上級中国語作文授業を受講している大学3年生の日本人が書いた中国語作文を収集して電子化し、学習者（以下は「JP」とする）の作文コーパスにしたものを分析データとする。JP コーパスには、260 篇の叙述文テキストが含まれている。さらに、学習者の作文における言語的特徴と比較対照するために、本研究では中国国内の高校3年生¹が書いた叙述文（260 篇）も収集し、母語話者（以下は「CHN」とする）の作文コーパスにしたものを比較対象とする。

表 1 コーパスの基本情報

	N (総篇数)	総字数	総語数	異なり語数
JP	260	145,959	100,244	45,221
CHN	260	249,599	182,053	82,543

2.3. 言語項目

変数として用いられる言語的特徴 (linguistic features) は、名詞・動詞・形容詞・複文などの 21 カテゴリーからなる中国語の 111 種の言語項目 (Appendix I) である。111 種の言語項目における 58 種は Biber (2006) によるもので、主に英語と中国語に共通する言語的特徴（例えば、心理的動詞/mental verbs、副詞/adverbs 等）である。また、Biber (2006) の 4 つの次元には含まれていない 3 つの言語的特徴（最もよく使われる名詞/noun: most frequently used; 最もよく使われる動詞/verb: most frequently used; 最もよく使われる形容詞/adjective: most frequently used）も本研究の言語項目

¹ 張(2018)では、“高中三年级学生的汉语水平能够代表汉语母语者的语言水平”（訳：高校3年生の中国語レベルは、中国語ネイティブスピーカーの言語レベルを代表することができる）と述べているため、本研究は高校3年生の中国人が書いた作文を参照対象として取り上げる。

リストに設けている。中国語には節（例、that-clauses, to-clauses, relative clauses 等）がないものの、複文が用いられているため、黄と廖(2017) で考察されている 10 種類の複文は分析に含まれている。そのほかに、中国語の言語的特徴及びアカデミックと非アカデミック的な中国語に関する研究（Zhu, 2015; Zhang, 2012; 劉, 2019; 馮ら, 2008）を参考にして、40 種の言語項目（例：着/zhe, 了/le, 过/guo, 的/de, 地/de, 得/de, 口語表現（口語詞/oral words）, 二音節テンプレートで用いられる単音語（嵌偶単音詞/monosyllabic words used in disyllabic templates）, 古典的な中国語単語（古語詞/particular classical Chinese words 等）が言語項目の構成要素として追加されている。

2.4. 分析手順

(1) 単語分割とタグ付け

各コーパス内の作文テキストに対し、NLPIR で単語分割とタグ付けを行う。

(2) 言語項目の頻度集計

各テキストにおける各言語項目の出現頻度を集計し、テキスト毎の頻度表を作成する。そして、出現頻度を 1,000 語当たりの相対頻度に調整する。

(3) 因子分析

相対頻度のデータを使って因子分析を行う。本研究の因子抽出法は Unweighted Least Squares（重みなし最小 2 乗法）であり、因子軸の回転法は斜交回転法における Promax Rotation（プロマックス回転）である。

(4) 次元の解釈

因子分析によって統計的に抽出された各因子を主な次元として捉え、機能ラベルを付与して解釈する。

3. 結果と考察

本研究は111種の言語項目を設けたが、Biber (2006:182-183)² が述べている三つの理由に基づいて、一部分の言語項目を削除し、変数として因子分析に使用した言語項目は58種である。また、因子分析を行う条件として、サンプル数が観測変数の5倍以上になる必要がある（Gorsuch 1983:322; Biber 1988:65）。本研究のテキストの数（520）と観測変数（58）は因子分析の条件を満たしている。

因子分析を行う前に、KMOとBartlettの球面性検定でサンプルサイズの妥当性を検証する必要がある。本研究で得られたKMO値は0.756（Kaiser(1974)の基準³を参照）で、Bartlettの球面性検定

² Biber(2006:182-183):

(1) Some features were dropped because they overlapped to a large extent with other features.

(2) Some features were dropped because they were extremely rare.

(3) Some features were dropped because they shared little variance with the overall factorial structure (features with communalities below 0.15 do not have meaningful factor loadings on any factor).

³ Kaiser(1974):

0.90 = marvelous ; 0.80 = meritorious ; 0.70 = middling ; 0.60 = mediocre ; 0.50 = miserable ;
below 0.50 = unacceptable

の有意確立は $p < 0.001$ であったため、因子分析を行う妥当性があることが確認された。

表 2 KMO と Bartlett の球面性検定

KMO and Bartlett's Test	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.756
Bartlett's Test of Sphericity	Approx. Chi-Square 15658.272
	df 1653
	Sig. .000

そして、本研究は「説明された分散の合計（Total Variance Explained）」とスクリー基準(Scree criterion)で7因子解（図1）を採用した。最初の7つの因子の累積寄与率（表3：第1因子から第7因子まで順番に加算した寄与率）は約40.69%であるため、7つの因子でデータ全体の約40.69%を説明することができる。

表 3 説明された分散の合計

Total Variance Explained							
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	8.432	14.538	14.538	8.027	13.840	13.840	6.836
2	4.833	8.333	22.870	4.556	7.855	21.694	3.569
3	3.696	6.372	29.243	3.284	5.662	27.356	4.537
4	2.812	4.847	34.090	2.366	4.080	31.436	3.402
5	2.304	3.972	38.062	1.960	3.378	34.814	3.054
6	2.091	3.605	41.666	1.722	2.969	37.783	3.793
7	2.027	3.496	45.162	1.687	2.908	40.691	2.458
8	1.868	3.221	48.383	1.422	2.452	43.144	2.084

表 4 因子間相関行列

Factor	1	2	3	4	5	6	7
1	1						
2	-.26	1					
3	.33	-.38	1				
4	.02	.15	.30	1			
5	.20	-.14	.05	-.07	1		
6	.30	-.03	.48	.28	.19	1	
7	-.13	-.07	.00	.05	.30	.18	1

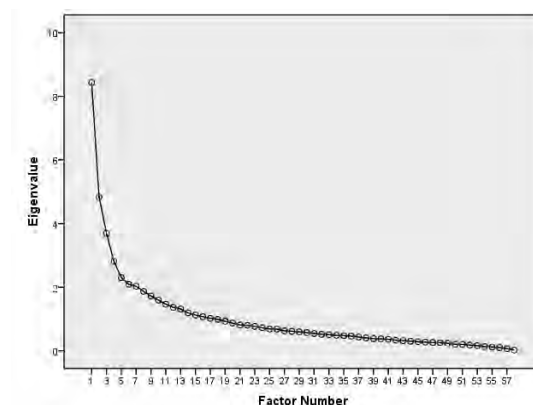


図 1 因子のスクリープロット

表4は因子軸をプロマックス回転した後の因子間相関の結果である。表4の相関係数を見ると、7つの因子は全般的に相互相関が低いことがわかった。つまり、7つの因子それぞれが、言語の変化の独立した次元を反映している。

3.1. 共起パターン

表5は58個の言語項目間の共起パターンである。このパターンは因子軸をPromax Rotationで回転した後のPattern Matrixの結果である。Biber(1988:87)によると、絶対値が0.30未満の負荷量 (Loadings) は、たとえ統計的に有意であっても、一般的には重要でないとして除外される。そのため、表5に因子負荷量の絶対値は0.30以上の言語項目しか表示されていない。また、負荷量の右側に括弧を付けている言語項目は、「他の因子にも重複して現れる項目で、他の因子における負荷量の方が高いため、因子得点の計算には使用されない」(田畑, 2002:83)。

表 5 言語項目の共起パターン⁴

Factor 1		Factor 3	
词汇多样性 (Lexical diversity)	.77	副词: 最常用 (Adverbs: most commonly used)	.75
助词“着” (Auxiliary: zhe/着)	.67	并列复句 (Parallel compound sentences)	.71
进行式动词 (Progressive verbs)	.65	副词: 表态度 (Adverbs: attitude)	.65
副词: 中度常用 (Adverbs: moderately commonly used)	.61	时间副词 (Adverbs of time)	.47
平均句长 (Average sentence length)	.61	假设复句 (Hypothetical compound sentences)	.41
形容词: 非常用 (Adjectives: rarely used)	.58		
状态词 (State words)	.53	Factor 4	
大小形容词 (Adjectives of size)	.42	副词: 表必然性 (Adverbs: inevitability)	1.02
助词“的” (Auxiliary: de/的)	.42	否定词 (Negative words)	1.00
嵌偶单音词 (monosyllabic words used in disyllabic templates)	.39		
方位词 (Position words)	.36	Factor 5	
时间副词 (Adverbs of time)	.33	过去式动词 (Past verbs (e.g., heard that/曾听))	1.01
.....		助词“了” (Auxiliary: le/了)	.90
因果复句 (Causal compound sentences)	-.47		
句首介词 (Prepositions in the beginning of a sentence)	-.52	Factor 6	
第一人代词“我” (First person pronoun: I/我)	-.64	条件复句 (Conditional compound sentences)	.90
		顺承复句 (Successive compound sentences)	.67
		假设复句 (Hypothetical compound sentences)	.47
Factor 2		Factor 7	
合偶双音词 (Couple disyllabic words)	.74	动作行为动词 (Action verbs)	.86
名动词 (Nominal verbs)	.67	动词: 最常用 (Verbs: most commonly used)	.71
平均词长 (Average word length)	.53	趋向动词 (Directional verbs)	.48
集体名词 (Collective nouns)	.39	
动词: 最常用 (Verbs: most commonly used)	.36	助词“的” (Auxiliary: de/的)	-.35
.....			
处所词 (Location nouns)	-.35		

⁴ 因子軸を Promax Rotation で回転した後の Pattern Matrix の結果における因子負荷量は回帰係数 (regression coefficients) であるため、因子負荷量の大きさが1を超えることが可能である (In Factor Analysis (or in PCA), what does it mean a factor loading greater than 1? - Cross Validated (stackexchange.com))。

3.2. 次元の解釈

表5における各因子を構成する言語項目に基づき、各因子の機能を解釈する。ここでは、第1因子から第7因子までを主な次元（Dimensions）として捉える。

3.2.1 次元1：書面的 vs.口語的

次元1のスコア及び次元1を構成している言語項目の相対頻度に、両コーパスの間で有意差があるかどうかについて、本研究は対応のないサンプルのT検定を行った。

表 6 対応のないサンプルの T 検定（次元1）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 1	260	-0.83	260	0.83	.000	3.32
词汇多样性 (Lexical diversity)	260	8.81	260	11.98	.000	2.09
助词“着” (Auxiliary: zhe/着)	260	1.13	260	9.42	.000	1.82
进行式动词 (Progressive verbs)	260	4.44	260	12.81	.000	1.45
副词：中度常用 (Adverbs: moderately commonly used)	260	3.52	260	8.23	.000	1.22
平均句长 (Average sentence length)	260	15.29	260	27.00	.000	1.45
形容词：非常用 (Adjectives: rarely used)	260	0.61	260	2.63	.000	0.95
状态词 (State words)	260	0.52	260	4.96	.000	1.43
大小形容词 (Adjectives of size)	260	3.13	260	5.66	.000	0.61
助词“的” (Auxiliary: de/的)	260	55.40	260	60.33	.001	0.30
嵌偶单音词 (monosyllabic words used in disyllabic templates)	260	18.24	260	23.01	.000	0.52
方位词 (Position words)	260	13.57	260	18.27	.000	0.64
时间副词 (Adverbs of time)	260	17.73	260	31.31	.000	1.45
因果复句 (Causal compound sentences)	260	8.81	260	3.29	.000	1.05
句首介词 (Prepositions in the beginning of a sentence)	260	6.57	260	2.19	.000	1.16
第一人代词“我” (First person pronoun: I/我)	260	54.44	260	27.66	.000	1.19

結果としては、すべて有意差 ($p<0.05$) があるということが明らかになった。また、言語項目の相対頻度の平均値について、正の負荷量を持っている言語項目は、JP（学習者）の方が過少使用し、負の負荷量を持っている言語項目は、JP（学習者）が過剰使用している。次元1では、正の負荷量が一番高いのは語彙多様性（词汇多样性/Lexical diversity）であり、負の負荷量が一番高いのは一人称代名詞“我”（First person pronoun: I/我）である。これは書面的と口語的な対比と言えるので、次元1に「書面的vs口語的」という名前を付ける。

- (1) 我喜欢吃中国菜，特别是回锅肉。我妈妈教我菜谱。我妈妈很会做菜，所以我想像我妈妈。我妈妈的得意菜是干烧虾仁，所以我想一起做。最近我也喜欢做点心。但是，因为失败了，所以我想再做。（JP_051）

文字通りの訳：私は中華料理、特にホイコーローが大好きです。私の母が私にレシピを教えてください。私の母は料理が得意なので、だから私は私の母のような人間になりたいです。私の母の得意な料理はエビチリで、だから私は一緒に作りたいです。最近

私は点心を作るのも好きです。しかし、失敗したので、だから私はまた作りたいと思います。)

学習者の作文には、一人称代名詞“我”と因果関係（因果復句/Causal compound sentences）を表している接続詞“因为”（～ので）、“所以”（だから～）等が多く使用されている。また、JPコーパスにおける語彙多様性（词汇多样性/Lexical diversity）、平均文長（平均句長/Average sentence length）等は、CHN（母語話者）コーパスより低いので、学習者の作文は口語的な傾向があるということが明らかになった。

3.2.2 次元2：詳細情報の叙述

次元2では、因子得点について、両コーパスの間に有意差があり、学習者コーパスの方が因子得点が高い。次元2に属している言語項目の相対頻度を見ると、正の負荷量を持っている言語項目は学習者の方が過剰使用している。負の負荷量を持っている言語項目は、両コーパスの間には有意差がないため、次元2を解釈する際、主に正の負荷量を持っている言語項目で解釈する。

表 7 対応のないサンプルの T 検定（次元2）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 2	260	0.39	260	-0.39	.000	1.05
合偶双音詞 (Couple disyllabic words)	260	17.19	260	12.47	.000	1.16
名動詞 (Nominal verbs)	260	16.23	260	9.45	.000	1.19
平均詞長 (Average word length)	260	1.45	260	1.37	.000	1.05
集体名詞 (Collective nouns)	260	11.71	260	5.44	.000	1.16
動詞：最常用 (Verbs: most commonly used)	260	142.84	260	127.92	.000	1.19
处所詞 (Location nouns)	260	6.84	260	7.27	.392	0.08

しかし、最もよく使われる動詞（動詞：最常用/verb: most frequently used）という言語項目は他の因子にも重複して現れる項目（表5）で、次元2の解釈に使用することができない。そのため、最初の4つの言語項目で次元2を解釈してみると、学習者の作文には、学習者の日常生活に関係がある集合名詞（集体名詞/Collective nouns）、例えば、「大阪大学」、及び名動詞（名動詞/Nominal verbs）、例えば、「影響」、「留学」などがよく出現しているので、次元2を「詳細情報の叙述」と名付ける。

- (2) 我有一个中国朋友，是大连理工大学的学生。她去年暑假在大阪大学学了三周日语。那时我当她的志愿者，帮她学习日语。她打算从十月开始在大阪大学一年的留学。今天她给我发短信。我从她发来的短信中知道了因为受新冠病毒的影响，她的留学期间缩短了。
(JP_068)

文字通りの訳：私は大连理工大学の学生である中国人の友人がいます。彼女は今年の夏、大阪大学で3週間日本語を勉強しました。その時、私は彼女のボランティアとして、日本語の勉強を手伝っていました。彼女は10月から大阪大学で1年間の留学をする予定があります。今日、彼女からメールが来ました。彼女のメールによると、新型コロナの影響で留学期間が短縮されたそうです。

3.2.3 次元3：文脈的複雑性

次元3では、因子得点について、両コーパスの間に有意差があり、学習者コーパスの方が因子得点が低い。最もよく使われる副詞（副詞：最常用/Adverbs: most commonly used）のほかの言語項目、例えば、接続詞“一边…一边…”（～しながら～する）を表している並列複合文（并列复句/Parallel compound sentences），“马上”（すぐ）を表している時の副詞（时间副词/Adverbs of time）などは、学習者の方が過少使用している。母語話者による例文を基に、次元3を「文脈的複雑性」と名付ける。つまり、学習者の作文には、文脈的複雑性が足りないと言える。

表 8 対応のないサンプルの T 検定（次元 3）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 3	260	-0.43	260	0.43	.000	1.04
副詞：最常用 (Adverbs: most commonly used)	260	63.34	260	63.82	.751	0.03
并列复句 (Parallel compound sentences)	260	2.89	260	5.57	.000	0.73
副詞：表态度 (Adverbs: attitude)	260	8.71	260	12.60	.000	0.59
时间副词 (Adverbs of time)	260	17.73	260	31.31	.000	1.45
假设复句 (Hypothetical compound sentences)	260	2.94	260	7.29	.000	1.12

(3)就在这时，汽车靠站了，上来一位抱着婴儿的阿姨，怀里的婴儿还在不停地哭闹，只见她一边哄小孩说“宝宝别哭了，马上就要到家了”，一边掏钱买票。（CHN_001）

訳：その時、バスは駅に引っ張られ、まだ泣いている赤ちゃんを抱えたおばさんがやってきて、「泣かないで、もうすぐ家に着くから」と赤ちゃんをなだめながら、切符を買うためのお金を取り出した。

3.2.4 次元4：個人の意見

次元4では、両コーパス間の因子得点には有意差がない。次元4に属している言語項目を吟味して、次元4を「個人的意見」と名付ける。

表 9 対応のないサンプルの T 検定（次元 4）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 4	260	-0.02	260	0.02	.582	0.05
副詞：表必然性 (Adverbs: inevitability)	260	11.64	260	13.34	.008	0.24
否定词 (Negative words)	260	10.29	260	12.23	.001	0.30

3.2.5 次元5：過去の叙述

次元5に属している言語項目は過去を記述する特徴であるため、次元5を「過去の叙述」と名付

ける。

表 10 対応のないサンプルの T 検定（次元 5）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 5	260	-0.21	260	0.21	.000	0.42
過去式動詞 (Past verbs (e.g., heard that/曾听))	260	17.59	260	21.36	.000	0.36
助詞“了” (Auxiliary: le/了)	260	11.49	260	14.59	.000	0.37

(4) 这次我们去鸡肉火锅店。其实，昨天晚上吃了鸡肉鸡蛋盖饭，我吃腻鸡肉。(JP_001)

文字通りの訳：今回私たちは鶏鍋のお店に行きます。実は昨夜、親子丼を食べました、私は鶏肉に飽きます。

(5) 时间一天天过去了，如今我已经长大，父母已步入中年，周围的一切都发生了很大的变化，但是唯一没变的是他们仍然吵架。(CHN_258)

訳：日が経って、私はすでに大人になり、両親もすでに中年になりました。周りのことはだいぶ変わってきたが、唯一変わっていないのは、両親が相変わらず喧嘩ばかりしていることです。

学習者は、過去の出来事を記述する際、“動詞+了”（～した）を使うべき文脈において、“了”を忘れてしまったケースが多い。例(4)の文は、過去の出来事について述べているため、“这次我们去了鸡肉火锅店（今回私たちは鶏鍋のお店に行きました）”、“我吃腻了鸡肉（私は鶏肉に飽てしまいました）”とするのが意味的に正しい文である。しかし、動詞の後ろに“了”をつけ忘れることにより、過去の出来事を現在形で叙述してしまっている。また、JPコーパスでは、過去を表す動詞（過去式動詞/Past verbs）、すなわち「時の副詞+動詞」の形（例えば“已经+動詞”/すでに～）で完了した動作を描写するケースが少ない。次元5の因子得点についてのT検定の結果を合わせてみると、学習者の作文には過去の叙述は過少である。

3.2.6 次元 6：ロジックの繋がり

次元6に属している言語項目である「条件複文」、「承前複文」、「仮定複文」は全て文脈のロジックを表す言語項目である。次元6を「ロジックの繋がり」と名付ける。

表 11 対応のないサンプルの T 検定（次元 6）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 6	260	-0.41	260	0.41	.000	0.97
条件複句 (Conditional compound sentences)	260	1.00	260	3.13	.000	0.91
順承複句 (Successive compound sentences)	260	1.69	260	5.56	.000	1.20
假设複句 (Hypothetical compound sentences)	260	2.94	260	7.29	.000	1.12

JPコーパスでは、次元6を構成している三つの複文を過少使用しており、学習者の作文はロジックの繋がりが不足しているということが分かった。

- (5) 他说：“这虽然只是小事，但是如果没有人站出来指出他的错，那么他就会认为自己没有错，以后他就会一直这样下去的。”对啊！就是因为没有人肯出来指出错误，社会上才依然存在各种不良现象。(CHN_006)

訳：「些細なことです、誰も彼が間違っていることを指摘しなければ、彼は自分が間違っていないと思い、これからもそうあり続けるだろう」と彼は言いました。その通り。誰も前に出て誤りを指摘しようとしないからこそ、様々な悪事が社会に残ってしまうのです。

3.2.7 次元7：行為の叙述

次元7では、因子得点について、両コーパスの間に有意差があり、学習者コーパスの方が因子得点が高い。学習者の作文には、動作動詞（動作行為動詞/Action verbs）、例えば“吃”（食べる）/“出门”（家を出る）、及び最もよく使われる動詞（動詞：最常用/verb: most frequently used）、例えば“去”（行く）/“有”（ある）などのような難度の低い動詞がたくさん使われている。そのため、次元7を「行為の叙述」と名付ける。

表 12 対応のないサンプルの T 検定（次元7）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 7	260	0.11	260	-0.11	.010	0.23
動作行為動詞 (Action verbs)	260	81.38	260	77.82	.042	0.18
動詞：最常用 (Verbs: most commonly used)	260	142.84	260	127.92	.000	1.19
趋向動詞 (Directional verbs)	260	14.42	260	14.45	.966	0.00
助詞“的” (Auxiliary: de/的)	260	55.40	260	60.33	.001	0.30

- (6) 早上我出门的时候，没下雨。不过打完工的时候，已经开下起了毛毛细雨。我没有雨伞，被雨淋湿去石桥，因为今天我们的社团开忘年会。进了十二月我有很多酒会，前天也有打工的酒会。这次我们去鸡肉火锅店。其实，昨天晚上吃了鸡肉鸡蛋盖饭，我吃腻鸡肉。(JP_001)

文字通りの訳：朝、出かけたとき、雨は降っていなかったです。しかし、アルバイトを終えた頃には、すでに小雨が降り出していました。今日は私たちのクラブの忘年会があったので、私は傘を持たずに雨に濡れながら、石橋に行ってきました。12月に入ると飲み会が多くなり、おとといもアルバイトの飲み会がありました。今回私たちは、鶏鍋のお店に行きます。実は昨夜、親子丼を食べました、私は鶏肉に飽きます。

3.3. 可視化

各次元のスコアに基づき、Rader chartsで両コーパスの区別を可視化する。CHNによって書かれた叙述文と比較して、JPの叙述文には、口語的な傾向があり、文脈的複雑性とロジックの繋がりが不足している。また、学習者は、過去の出来事を記述する際、動詞の後に“了”を使うべきところで、“了”を使っていない場合がある。さらに、叙述文で動作を記述する際、学習者は単純な動詞を使用する傾向があることが明らかになった。

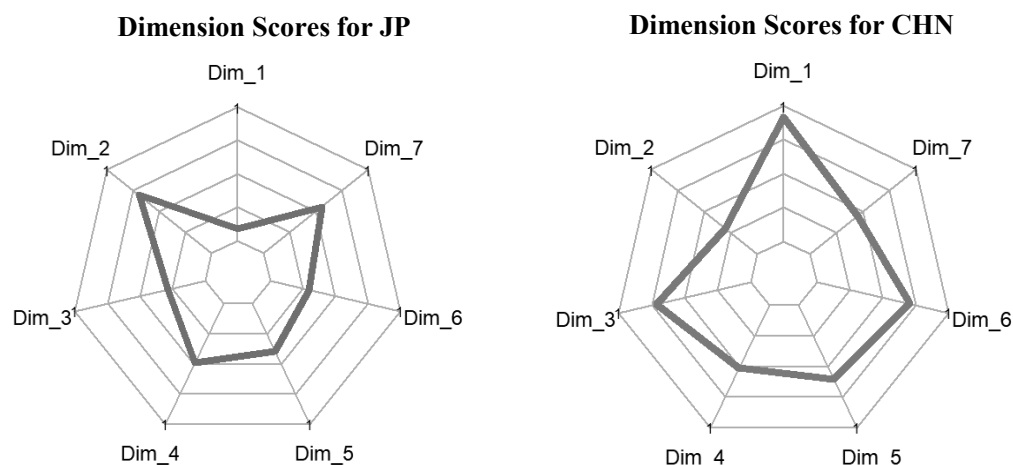


図 2 可視化図

4. 終わりに

本研究は予備的な研究として、MF/MD法の分析視点から、因子分析で7つの次元を抽出し、日本人中国語学習者が書いた叙述文における言語的特徴の考察を行った。

この研究では2つの点が不足していると考えられる。まず、Yong & Pearce(2013) がこの分析方法の制約の一つとして述べているように、因子のラベリングに問題がある可能性がある。つまり、次元を解釈する時に、因子内の変数を正確に反映していない場合がある。また、言語項目の頻度集計に問題があるおそれがある。例えば、中国語の複文はとても複雑であり、本研究では、主に接続詞のタグに基づいて複文の頻度を集計したが、すべての複文を集計できていない可能性もある。

今後の課題としては、上述の2つの不足を補うために、研究手法と分析視点を改善していく。また、コーパスデータを拡大し、ほかの種類の作文（例えば意見文、説明文）も収集して分析してみることが必要である。量的分析で得られた結果に基づき、作文テキストにおける典型的な言語表現を抽出し、質的分析を行いながら、日本人中国語学習者の作文における言語的特徴や傾向を明らかにしていきたい。

文献

- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, 14, 7-34.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing*, 26, 80-95.
- Gorsuch, Richard. (1983). *Factor Analysis, 2nd edition*. Hillsdale, NJ: Lawrence Erlbaum Association.
- Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22, 119-131.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, 9(2), 79-94.
- Zhu, X. (2015). *A multi-dimensional approach to register variation in Mandarin Chinese* (Unpublished masters' thesis). Zhejiang University, Hangzhou, China.
- Zhang, Z. S. (2012). A corpus study of variation in written Chinese. *Corpus Linguistics and Linguistic Theory*, 8(1), 209-240.
- 田畑 智司. (2002). 「Corpus-Based Stylistics--MF/MD 法による文体比較--」 『英語文体論の方法と射程』 言語文化共同プロジェクト. pp.75-89.
- 田畑 智司. (2005). 「コーパスに基づく文体論研究」 斎藤他 編『英語コーパス言語学：基礎と実践[改訂新版]』 第9章(pp. 183-206) 東京：研究者出版
- 石川慎一郎. (2014). 「(訳) 概説コーパス言語学—手法・理論・実践」 東京：ひつじ書房
- 冯胜利, 王洁, 黄梅. (2008). 汉语书面语体庄雅度的自动测量[An automatic feature checking algorithm for degree of formalities in written Chinese]. *语言科学[Yuyan Kexue]*, 7(2): 113-126.
- 黄伯荣, 廖序东. (2017). *现代汉语 (修订第六版) [Modern Chinese (6th Rev. ed.)]*. 北京: 高等教育出版社 [Beijing: Gaodeng Jiaoyu Chubanshe].
- 刘艳春. (2019). 汉语语体变异的多维度分析——基于 17 个语体 72 项语言特征的考察. [A multi-dimensional analysis to variations in Chinese registers: Based on a study on 72 linguistic features in 17 registers]. *江汉学术[Jiangnan Xueshu]*, 38 (3), 100-110.
- 张江丽. (2018). 汉语作为第二语言学习者笔语产出性词汇研究.[A Study on Written Productive Vocabulary of Chinese Second Language Learner]. *世界汉语教学[Shijie Hanyu Jiaoxue]*, 32(3), 417-430.

Appendix I

类别(Category)	语言特征(Linguistic feature)	例(Example)
1 名词及其特殊属类 Nouns & Special Noun Genus	1. 名词: 最常用 (Nouns: most commonly used)	人、社会、问题
	2. 名词: 中度常用 (Nouns: moderately commonly used)	微生物、军官、地主阶级
	3. 名词: 非常用 (Nouns: rarely used)	决赛、西红柿、写字台
	4. 抽象名词 (Abstract nouns)	奥秘、魅力、权术
	5. 立场名词 (Stance nouns)	倾向、意图、信心
	6. 心理名词 (Mental nouns)	悲欢、廉耻、嫌隙
	7. 大学学科专业分类词 (University subject classifications)	哲学、心理学、经济学
	8. 指人名词 (Personal nouns)	书法家、工程师
	9. 集体名词 (Collective nouns)	公社、陆军、编委会
	10. 具象名词 (Figurative nouns)	日记、酸奶、取景框
	11. 具象科技名词 (Figurative scientific and technological nouns)	暗礁、磁场、吞噬细胞
	12. 度量衡名词 (Metrics and measures)	尺寸、价码、储量
2 动词及其特殊属类 Verbs & Special Verb Genus	13. 动词: 最常用 (Verbs: most commonly used)	有、说、到
	14. 动词: 中度常用 (Verbs: moderately commonly used)	送来、被迫
	15. 动词: 非常用 (Verbs: rarely used)	掘、插嘴、试点
	16. 动作行为动词 (Action verbs)	唱歌、操作、挥洒
	17. 使役动词 (Causative verbs)	动摇、落实、复兴
	18. 存现动词 (Existing verbs)	升起、凝聚、失踪
	19. 心理动词 (Mental verbs)	按捺、忏悔、揣摩
	20. 肯定动词 (Affirmative verbs)	认可、查明、领会
	21. 交流动词 (Communicative verbs)	安慰、邀请、叙旧
	22. 推测性动词 (Speculative verbs)	臆想、推测、思索
	23. 可能性动词 (Possibility verbs)	似乎、觉得、看样子
	24. 趋向动词 (Directional verbs)	起、出来、下来
	25. 副动词 (Adverbial verbs)	带头、合力
	26. 形式动词 (Light verbs)	进行、给予、加以
3 形容词及其特殊属类 Adjectives & Special Adj.Genus	27. 形容词: 最常用 (Adjectives: most commonly used)	大、小、多
	28. 形容词: 中度常用 (Adjectives: moderately commonly used)	舒服、最低、得意
	29. 形容词: 非常用 (Adjectives: rarely used)	最优、无用、疲乏
	30. 大小形容词 (Adjectives of size)	大、微小、修长
	31. 常用形容词: 相关性 (Common adjectives: relevance)	大体的、全部的
	32. 状态词 (State words)	悠远、寥寥、活生生
	33. 区别词 (Distinguishing words)	特定、负面、崭新
4 数词和量词 Numerals & Quantifier	34. 数词 (Numerals)	一、5、千
	35. 数量词 (Quantifiers)	一些、两个、多次
	36. 量词 (Measure words)	这类、那种、各条
	37. 动量词 (Momentum quantifiers)	一番、两回、五场
	38. 时量词 (Time quantifiers)	两点、三周、十年
5 代词和代动词 Pronouns & Pro-Verbs	39. 第一人代词“我” (First person pronoun: I/我)	我
	40. 第一人代词“我们” (First person pronoun: we/我们)	我们
	41. 其他第一人称代词 (Other first person pronouns)	吾、咱们、晚生
	42. 第二人称代词 (Second person pronouns)	你、您、你们
	43. 第三人称代词 (Third person pronouns)	他、她、他们
	44. 代词: 它 (Pronoun: it/它)	它
	45. 普通名词+们 (Common nouns + plural/们)	朋友们、小鸟们
	46. 指示代词 (Demonstrative pronouns)	每、某、这儿
	47. 不定代词 (Indefinite pronouns)	某事、有人、任何
	48. 句首代词 (Pronouns at the beginning of a sentence)	我不信、你猜猜
6 副词及其特殊属类 Adverbs & Special Adv.Genus	49. 代动词 (Pro-Verbs)	再来一杯、来个比赛
	50. 副词: 最常用 (Adverbs: most commonly used)	也、不、就
	51. 副词: 中度常用 (Adverbs: moderately commonly used)	实在、不必、皆
	52. 副词: 非常用 (Adverbs: rarely used)	一概、一举、甚至于
	53. 副词: 表必然性 (Adverbs: inevitability)	甭、显然、势必
	54. 副词: 表可能 (Adverbs: possibility)	稍、大概、或许
7 介词及其短语 Prepositional Phrases	55. 副词: 表态度 (Adverbs: attitude)	难道、居然、不妨
	56. 句首介词 (Prepositions in the beginning of a sentence)	从、据、关于
	57. 介词短语 (Prepositional phrase)	向……、与……、由……

8	助词 Auxiliaries	58. 助词“的” (Auxiliary: de1/的)	的
		59. 助词“地” (Auxiliary: de2/地)	地
		60. 助词“得” (Auxiliary: de3/得)	得
		61. 助词“等”\“等等” (Auxiliaries: deng/等,dengdeng/等等)	等、等等
		62. 比况助词 (Situational comparative auxiliaries)	一样、似的
9	其他词汇属类 Other Lexical Genera	63. 语气词 (Modal particles)	啊、呢、焉
		64. 顺序词 (Sequential words)	首先、其次、总之
		65. 句内并列连词 (Syntactic conjunctions)	和、跟、以及
		66. 小品词 (Grammatical particles)	嗯、对对、行行
		67. 口语词 (Oral words)	今儿、够呛、没辙
		68. 儿化音 (Er hua words)	头儿、盖儿、手帕儿
		69. 嵌偶单音词 (monosyllabic words used in disyllabic templates)	极佳、围敌、返京
		70. 合偶双音词 (Couple disyllabic words)	禁止、购买、安装
		71. 古语词 (Classical words)	称之、岂不、颇有
		72. 增强语 (Amplification words)	深深、极其、颇为
		73. 模糊限制语 (Hedges)	至少、约摸、恐怕
10	名词形式 Noun Forms	74. 名物化 (Nominalization)	知名度、开放性
		75. 名动词 (Nominal verbs)	研究、储备、介绍
		76. 名形词 (Nominal adjectives)	方便、亲近、不同
11	状态形式 State Forms	77. 动词“是”作主要动词 (The verb 是/‘be’ used as the main verb)	尤其是、无疑是
		78. 动词“有”，表存现 (The verb 有/‘have’ used to express the existence)	有学者、有争议
12	时态和体标记 Temporal & Aspect Markers	79. 进行式动词 (Progressive verbs)	正看、玩耍中
		80. 过去式动词 (Past verbs)	曾听、刚刚来
		81. 助词“着” (Auxiliary: zhe/着)	着
		82. 助词“了” (Auxiliary: le/了)	了
		83. 助词“过” (Auxiliary: guo/过)	过
13	情态动词 Modal Verbs	84. 必要性情态动词 (Necessity modal verbs)	必须、一定、务必
		85. 情态动词：表未来 (Modal verbs: future)	会、要、应该
14	地点和时间状语 Adverbials of Place/Time	86. 时间副词 (Adverbs of time)	不断、偶尔、久久
		87. 副词：表处所 (Adverbs: Location)	到处、随处、遍地
		88. 处所词 (Location nouns)	内陆、海滨、前线
		89. 方位词 (Position words)	上、前、东南部
15	缩略形 Contractions	90. 缩略语 (Abbreviations)	北大、冬奥会
16	否定 Negatives	91. 否定词 (Negative words)	不、莫、未曾
17	独立语 Independent Words	92. 叹词 (Interjections)	噢、嗨、哎呀
		93. 拟声词 (Onomatopoeia)	嘭、咚咚、哗啦啦
		94. 插入语 (Parenthesis)	综上所述、相比之下
18	疑问句 Questions	95. 特指疑问句 (Interrogative Sentences)	哪些……、如何……
19	被动 Passives	96. 被动 (Passive)	被……、为……所……
		97. 无施事者被动句 (Passive sentences without doer)	被骂、挨打
		98. 介词“把” (Preposition 把/‘ba’ sentences)	把
20	复句 Compound Sentences	99. 并列复句 (Parallel compound sentences)	既……又……
		100. 顺承复句 (Successive compound sentences)	先……再……
		101. 解说复句 (Explanatory compound sentences)	即……、就是说……
		102. 选择复句 (Selective compound sentences)	或者……或者……
		103. 递进复句 (Progressive compound sentences)	不仅……而且……
		104. 条件复句 (Conditional compound sentences)	只要……就……
		105. 假设复句 (Hypothetical compound sentences)	如果……那么……
		106. 因果复句 (Causal compound sentences)	因为……所以……
		107. 目的复句 (Purpose compound sentences)	以便……、以免……
		108. 转折复句 (Turning compound sentences)	虽然……但是……
21	词汇丰度 Lexical specifications	109. 词汇多样性 (Lexical diversity)	/
		110. 平均词长 (Average word length)	/
		111. 平均句长 (Average sentence length)	/

Top2Vec による小説の探索的研究

—程小青の作品解説を中心に—

黄 晨雯

大阪大学大学院言語文化研究科

〒 560-0043 豊中市待兼山町 1-8

Email: u469911c@lang.osaka-u.ac.jp

概要 本稿では、Dimo(2020)によって提起された新たなトピックモデル Top2Vec を使用して、トピックという視点から小説の解説を試みることを主な目的とする。中国の民国時代に活躍した有名な推理作家である程小青の作品を分析対象として、実行結果をもとにトピックの構成単語、またトピックとテキストファイルの関係性などを考察する。程小青の作品全体を最も代表するトピックは社会問題をめぐる話題であり、冒頭部分において大概社会問題に関する思考や論述が導入的な話として好まれることを解明した。これは先行研究の質的分析においても検討されたことであるが、トピックの視点より客観的に裏付けしている。また、トピック間の関係性を可視化して、例えば封鎖空間を中心に事件を展開する形が好かれるなどのスタイル的なことも考察できる。文体的には言葉遣いが非常に異質な作品もトピックから見つけ出せる。さらに、他の作家との比較にあたって、Top2Vec の実用性がどれほどあるかを実験を繰り返して試みたところ、事前学習モデルを導入した Top2Vec の方がより汎用的だということも解明した。

キーワード トピックモデル, Top2Vec, 言語特徴

运用 Top2Vec 的小说探索性研究

—以程小青的作品解读为中心—

黄 晨雯

大阪大学大学院言語文化研究科

〒 560-0043 日本大阪丰中市待兼山町 1-8

摘要 本文旨在运用 Dimo 在 2020 年新提出的 Top2Vec 主题模型来尝试通过主题这个视角解读小说。主要的研究对象是中国民国时代非常著名的推理小说家程小青的作品。通过考察模型生成的各个主题的构成单词以及主题所对应的文档之间的关系来对小说进行分析。从结果上来看,最能代表程小青的是有关于社会问题的主题。而且通过对小说开头部分主题成分的分析可以看出大部分的作品都喜欢用对社会问题的思考和讨论来作为导入。这些在先行研究中也提到,但是通过主题的视角更能客观证明质性分析的结果。另外,通过主题两两关系的可视化图可以看出不少程小青在情节设计上的风格,譬如偏爱以室内空间作为事件场所等。通过对主题词和主题分布的分析也可以区分用词较为独特的部分作品。最后也探讨了运用 Top2Vec 进行

多个作者间主题分析的可能性。经过多次实验可发现，导入了预训练模型的 Top2Vec 在这方面颇有实用性。

关键词 主题模型, Top2Vec, 语言特征

1. 研究背景

近年、統計的テキスト解析の研究において、LDA を代表とするトピックモデルが頻用されている。Blei et al.(2003)によって提案された LDA(Latent Dirichlet Allocation)、つまり潜在ディリクレ配分法は、「各文書は潜在的にトピックを多数持ち、同じトピックに属する単語は同一文書に出現しやすい」と想定し、大規模な文書を構成する各々の文書に内在するトピックを推定するアルゴリズムである。黄 (2020) では LDA モデルを用いて中国近現代の推理小説作家 4 名の作品を対象に、トピックという観点より作家間の相違を探し出そうとしていた。その結果、民国時代に活躍した程小青と 21 世紀初頭の作家の間において時代背景によるトピックの違いがいくつかみられた。また、主人公やモチーフの設定に影響される特徴的なトピックについても検討した。ただし、LDA モデルの設定に関する問題などは解決されていない。例えば、LDA モデルではトピック数を研究者自身で指定しなければならないが、最適な数をどのように決めればいいのか、また stopwords にする単語（機能語など分析対象から取り除く語）も主観的なのではないかというような疑問がある。Dimo (2020) はこのような問題を解決するために新たな Top2Vec モデルを提起した。

Top2Vec は、Doc2Vec を使って文書をベクトルにしてから、UMAP で次元削減を行い、HDBSCAN で文書ベクトルを分類し、同じグループに分類された文書ベクトルの平均を取ったものを Topic Vector にした上で、Topic Vector から近い順にそれぞれのトピックの構成単語を決めるという仕組みになっている。Top2Vec の特徴としては意味的關係性を考慮した上に、短文にも性能が良いとされている。LDA と違って、1 つの文書が 1 つのトピックと対応し、トピック数を指定することもない。また stopwords がノイズになることが多いためトピックに配属されにくく、機能語などの前処理を行う必要がないと指摘されている。単語間の意味的關係性を考えたモデルなのでトピックの Information gain¹を評価したところ、訓練データにおいて LDA や PLSA など従来のトピックモデルより Top2Vec の方が遥かに高い得点を取ったことも証明されている。そこで、この新たな Top2Vec を使用して小説の解説を試みる。LDA モデルとの比較もしながら推理小説における Top2Vec の応用性と可能性を探ってみる。

2. データとモデルの実行

本稿では、中国の民国時代に活躍した有名な推理小説作家程小青(1893-1976)の作品を分析対象とし、Top2Vec を実行してみる。『漢揚居』という読書サイトに収録された 43 篇の程小青の作品をダウンロードし、Stanford CoreNLP という自然言語処理ツールを使って分かち書きと

¹ トピックがドキュメントをどれほど適切に記述しているかを評価する方法。Dimo Angelov (2020) TOP2VEC:Distributed Representations of Topics, Computation and Language, arXiv:2008.09470 pp.9-10

品詞タグ付与を行い、90 万語程度のコーパスを作り上げた。分析時に使用するデータは以下に示すような分割と品詞タグ付与が行われた文である。

我_PN 记叙_VV 我_PN 的_DEG 老友_NN 霍桑_NN 的_DEG 探案_NN 纪录_NN 已_AD 有_VE 好几_CD 十_CD 种_M 。_PU 一般_JJ 读者_NN 时常_AD 写_VV 信_NN 来_VV 寻找_VV , _PU 此外_AD 还_AD 有_VE 没_AD 有_VE 别的_DT 案件_NN 可以_VV 公诸_VV 同好_NN 。_PU 在_P 已往_AD 的_DEV 二十多_CD 年_M 中_LC , _PU 霍桑_NN 凭着_P 的_DEG 智力_NN , _PU 勇敢_JJ 的_DEG 精神_NN 和_CC 为_P 大众_NN 服务_VV 的_DEC 热忱_NN , _PU 所_MSP 经历_VV 的_DEC 疑难_JJ 案件_NN 何_AD 上_VV 一二百_CD 种_M , _PU 并且_AD 大半_CD 都_AD 记在_VV 我_PN 的_DEG 记事_NN 册_NN 里_LC 。_PU

(『一只鞋』より抜粋)

Dimo (2020) では訓練データとしてニュースコーパスが使われていたが、小説の場合は短いニュースと違って、作品のファイル分けが必要とされる。細かく分割すれば1つのトピックによって解釈されることが合理的だと思われる。トピックが変わる境目として考えられるのはパラグラフであるが、会話文がパラグラフになっていることが多いためパラグラフ単位の適切さが疑われる。そこでニュースの長さを参照として 100 語ごとにファイルを切り分けることにした。

単語を何一つ削除しないで実行した結果、人名が多数混じり込んでいることに気づいた。普通に考えれば人名は作品を特徴づける肝心な要素であるが、話題の一般性を求める場合は取り除いたほうが解釈しやすいと思われる。「_NR_」タグがつく固有名詞を除いたコーパスで実行した結果、59 個のトピックが出力された。トピックと対応するファイル数の大きい順から上位 10 個のトピックの主要単語とそれらを踏まえてつけたラベルを表 1 にまとめた。

表 1 上位 10 個のトピックを構成する主要単語 (Top2Vec)

Topic	Label	Keywords
0	社会問題	社会_nn 一般_jj 种_m 时代_nn 观念_nn 对于_p 传统_jj 能_vv
1	電話通知	电话_nn 到_vv 去_vv 消息_nn 往_p 打_vv 回来_vv 赶到_vv
2	仮説と推理	能_vv 推想_vv 假定_vv 想_vv 事实_nn 行凶_vv 这_dt 究竟_ad
3	外見と服装	穿_vv 身材_nn 脸形_nn 戴_vv 顶_m 双_m 穿着_vv 西装_nn
4	車	汽车_nn 辆_m 车子_nn 车_nn 转弯_vv 门面_nn 马路_nn
5	特定の言葉	她_pn 十分_ad 应该_vv 说道_vv 对_p 有点_ad 听到_vv
6	宅内の騒ぎ	听得_vv 声音_nn 外祖母_nn 睡_vv 楼下_nn 楼上_nn 惊醒_vv
7	対話	能_vv 你_pn 想_vv 团员_nn 实在_ad 太_ad 这_dt 事_nn
8	手紙	写_vv 张_m 信_nn 信箋_nn 字迹_nn 纸_nn 信封_nn 封_m
9	学業	读书_vv 年_m 毕业_vv 中学_nn 今年_nt 父亲_nn 大学_nn

トピックの構成単語をめぐる考察は 3.1 で行う。また、それぞれのトピックと対応するテキストファイルも Top2Vec の出力結果から検索できる。トピックとテキストファイルとの関係性などについても詳しい考察を試みる。

3. 考察

3.1 トピック



図 1：Topic0「社会問題」のワードクラウド

図 1 は程小青の作品を最も代表する（最も多くのテキストファイルを説明する）Topic0 のワードクラウドである。図 1 を吟味すれば、「社会」「一般」「觀念」「時代」「傳統」「態度」「教育」「民族」などの主要単語から、Topic0 は「社会問題」に関わる話題だと解釈して良いであろう。Topic0 が程小青の作品に共通した重要なトピックであるということは、程 (2014)などで指摘された「程小青が民国時代の社会問題を深刻に検討している」ということを裏付けたのではないかと考えられる。従来の質的分析において主張されたものをトピックの視点より一層客観的な論証を提供している。



図 2：Topic1「電話通知」のワードクラウド

また Topic1 では程小青が使う「電話」という媒介の役割を図 2 から覗くことができる。「電話_nn」（電話）、「消息_nn」（情報）、「说明_vv」（説明）のようなメッセージの伝達を表す語のほか、方向動詞を示す「到_vv」「去_vv」「往_p」「趕到_vv」「回来_vv」などが共起することから、「電話」と「移動」とのつながりが緊密で、電話が移動の合図となるプロットが好まれるのではないかと考えられる。

上位にあるトピックは大概多くの作品に出現しやすく一般性を持つと思われるが、Topic5 は上位トピックにおいて非常に異質である。図 3 からわかるように、Topic5 を構成する語彙はお互いに関係性が薄いように感じられ、特定の言葉遣いの集合なのではないかと考えられるが、話題として解釈することは難しい。そこで、コーパスにおける Topic5 の分布を見ると興味深い特徴が見られた。



図 3：Topic5 「特定の言葉」のワードクラウド

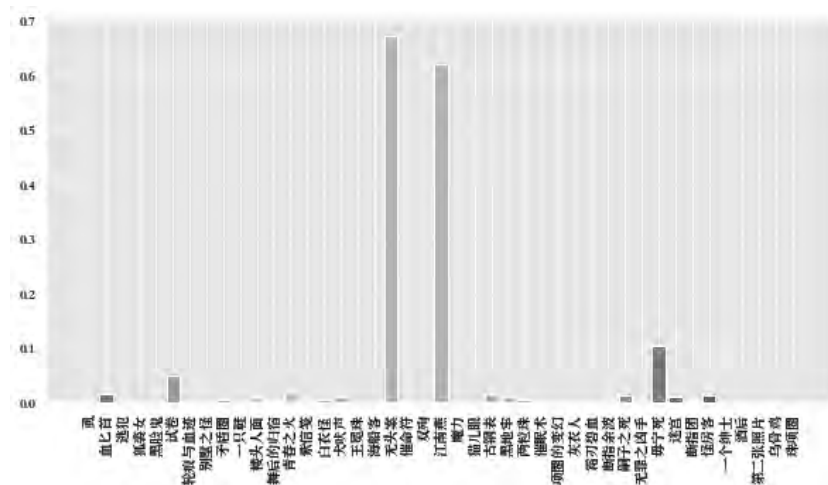


図 4：作品ごとの Topic5（特定の言葉）の割合

図4では各作品における Topic5 の分布を示している。『江南燕』と『无头案』を特徴づけるトピックだと判断して良いだろう。実は程小青の作品は出版する度に改訂されると知られている。ただし、程小青自身が詳しい説明をしていなかったため、どの部分、またどれほど書き換えたのかは追究することができない。『江南燕』と『无头案』は初版では古文で書かれたらしいが、現在読めるバージョンは口頭文になっているため改訂が行われたことは確実であろう。今までの研究では、「她_{pn}」という人称代名詞は程小青における使用率が非常に低いものに対して、現代作家に偏る特徴があることを既に解明した。Topic5 における分布の大差を成したのは現代で使われやすい言葉を改訂版において大量に使用しているからではないかと考えられる。

このような文体の相違が観察できたのも Top2Vec の見所なのではないかと思われる。なぜなら単語「她 pn」(彼女)は人称代名詞としてよく LDA モデルでは削除対象となっている。

「有点_ad」(少し)、「十分_ad」(とても)、「应该_vv」(べき)などもそうである。機能語でありながら作品を特徴づける語彙を見つけ出せることがメリットだと考えられる。

3.2 プロット

前節ではトピックを構成する単語を中心に検討した。この節では遠読の視点で作品のプロットをトピックで解説してみる。

3.2.1 冒頭部分

程 (2015) によると、西洋の探偵小説と違い、程小青は中国古代の芝居小屋で行われる講談を倣って聞く人を待つために、物語自体とそれほど関係のない話、例えば新聞記事や日常生活をめぐる議論などを最初にする傾向があるという。さらに、荊 (2014) によれば、近代探偵小説において物語は事件から始まる特徴がある。程小青の小説では通報、または依頼人が助けを求めてくるというプロットからストーリーが始まることが多いほか、冒頭で環境描写をすることも少なくない指摘された。そこでトピック的にはどのような特徴、あるいはパターンが見られるのかという疑問を抱きながら冒頭部分のファイルと対応するトピックの分布を考察する。

各作品の最初のファイル (最初の 100 語) に対応するトピックを統計すると、半分ほどのファイルが Topic0 (社会問題) に当てはまることが分かった。ただし、冒頭部分といっても 100 語以内に限ることはないと考えられ、テキストファイルを 10 個ごとにまとめて、最初の 10 個のファイル (「作品名+F0」で表記) における各トピックの割合と分布を図 5 を通じて考察してみる。

図 5 のヒートマップでは、セルにおける色の違いでトピックの比重を表している。左上のカラーバーを参照として、薄い黄色に近いほどトピックの占める割合が低い一方で、青色が濃いほど割合が高いとされている。横軸にはトピックの数字、縦軸には作品ファイルが並び、類似性のあるものが近くに配置される特性がある。

一番左に位置する Topic0 (社会問題) が半分以上の作品において高いウェイトを占めていることを図 1 から容易に読み取れ、Topic0 が非常に特徴的であることがわかる。また、Topic0 ほど明らかではないが、左から 4 列目の Topic3 (外見と服装) に関しては、U 検定² ($p\text{-value} < 0.0001$) を通じて計算した結果、その平均割合が Topic0 以外の全てのトピックとの間で統計的有意差があることが明らかになった。

要するに、社会問題と外見描写に関わるトピックが冒頭部分の特徴づけるトピックだと結論づけられる。特に社会問題に関する思考や論述が物語の導入的な話として好んで使われる傾向にあると言えるだろう。

² 対応のない 2 群のデータに対して、差がないかについて検定する手法。

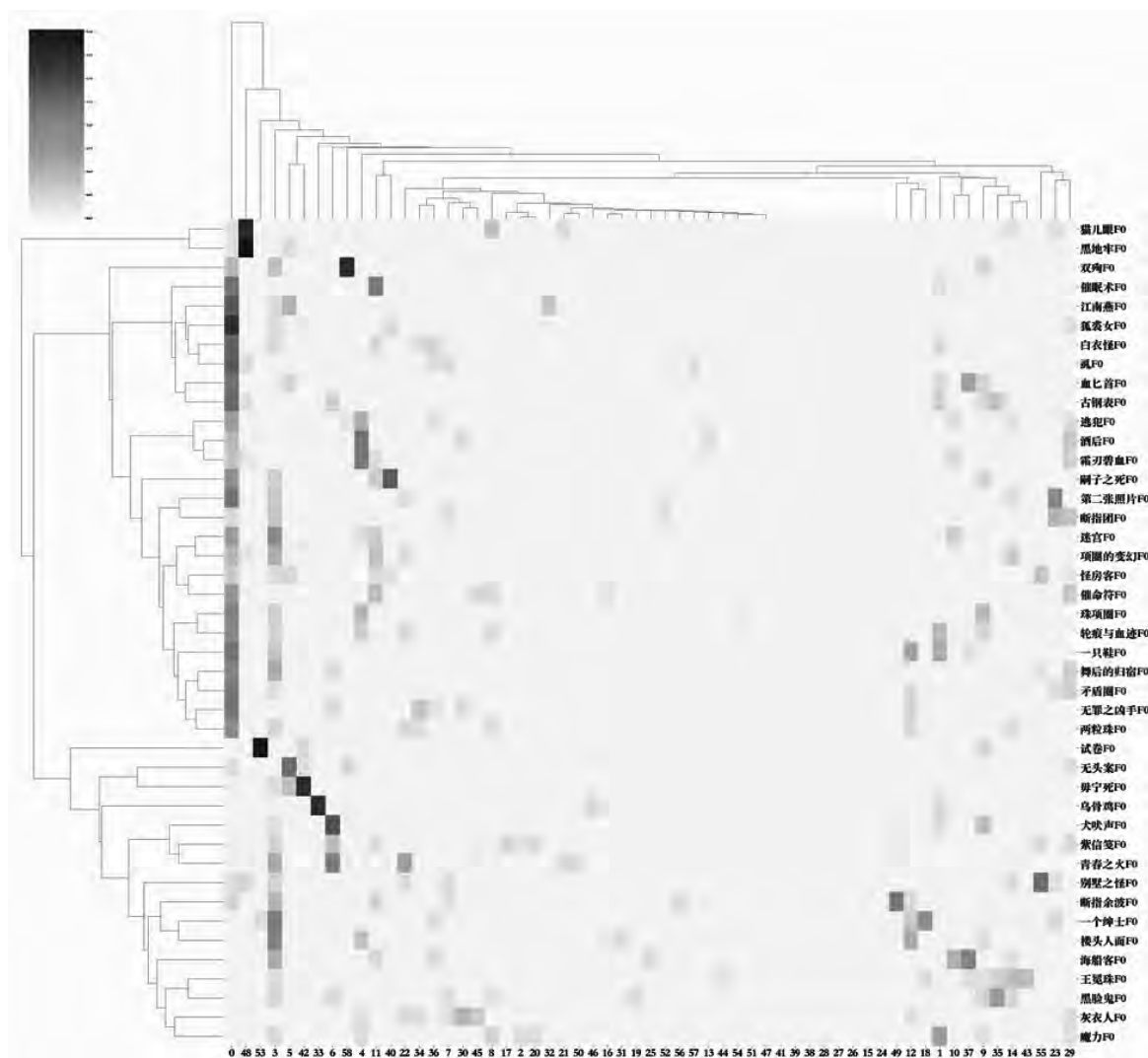


図5 トピックと作品ファイルのヒートマップ

3.3.2 2-gram Topic Chain

Top2Vec はコンテキスト的に語彙の関係性を考えた上に、各テキストファイルは 1 つのトピックによってしか解釈されない特徴があるため、トピック間の共起関係を見ることが興味深く感じられる。ここでは n-gram 言語モデルを適用し、単語ではなくトピック間の 2-gram を探してみる。表 2 には集計結果の上位 5 つの項目を示している。上位に並ぶ 2-gram は基本的に前出トピックと後出トピックが同じものになっている。プロット的に同じトピックが続く可能性が高いということは容易に理解できる。ただし、ここでは異なるトピックの関係性に注目したいので、同じトピックが継続するパターンを除いた 2-gram を表 3 で提示し、また共起関係を可視化したネットワークグラフ図 6 を作画した。

表 2：トピック 2-gram（上位 5 つ）

ランク	トピック 2-gram	共起頻度
1	特定の言葉 - 特定の言葉	273
2	社会問題 - 社会問題	165
3	車 - 車	122
4	手紙 - 手紙	105
5	宅内の騒ぎ - 宅内の騒ぎ	89

表 3：前後が異なるトピック 2-gram（上位 5 つ）

ランク	トピック 2-gram	共起頻度
1	電話通知 - 車	40
2	車 - 外見服装	36
3	社会問題 - 仮説と推理	32
4	電話通知 - 社会問題	31
5	対話 - 社会問題	30



図 6 共起ネットワーク（共起頻度 15 以上）

図 6 では他のトピックとの共起が多いほど、円が大きく表示される。トピックを繋ぐエッジについては、共起頻度が高いほど太く描かれる特性がある。概ね大きなトピックをめぐって展開されていることが分かる。「電話通知」と「車」の共起度が最も高いことから、4.1 で検討した電話と移動との関係性も裏付けしているのではないかと考えられる。右側に位置する「社会問題」「仮説と推理」「対話」などの共起が緊密になっており、推理と社会問題の検討が会話の中心的な主題になっていることが考えられる。また、事件の場所を示唆する「室内」「宅内の騒ぎ」「家

具」などから、封鎖空間を中心に事件を展開する形を好む程小青のスタイルが反映していると解釈するのが妥当であろう。

4. 他作家との比較

Top2Vec を通して、トピックをめぐる諸々な考察ができることは既に証明されている。では、作家間のトピックの相違などを見つけ出すことができるかどうかを解明するために、推理小説作家である雷米の作品『心理罪』シリーズ全5作を取り上げ、上述した前処理を行った後でコーパスに追加し、モデルを実行してみた。結果としては表4に示したような2つのトピックが出力された。

表4：程小青と雷米のデータコーパスから生成されたトピック

Topic	Keywords
0	但_ad 便_ad 些_m 虽_cs 使_vv 因着_p 吗_sp 瞧_vv 已_ad 自然_ad 我_pn 分明_vv 道_vv 忽_ad 答道_vv 仍_ad 眼光_nn 我们_pn 委实_ad 又_ad 既然_cs
1	她_pn 看到_vv 看着_vv 让_vv 盯着_vv 女孩_nn 的_dec 身边_nn 笑笑_vv 表情_nn 看看_vv 感到_vv 情况_nn 仍然_ad 刚刚_ad 边_ad 而_ad 自己_pn 看_vv 秒_m

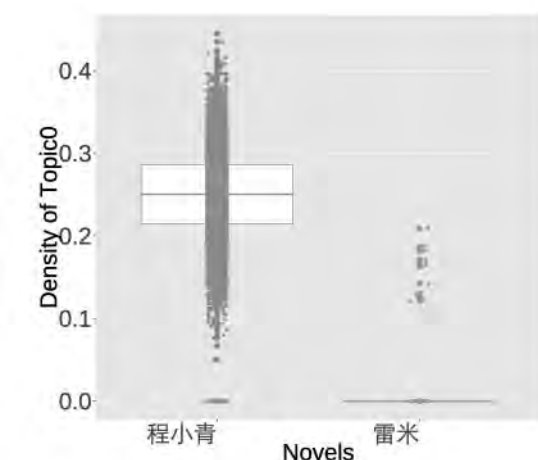


図7：Topic0 のジッター図

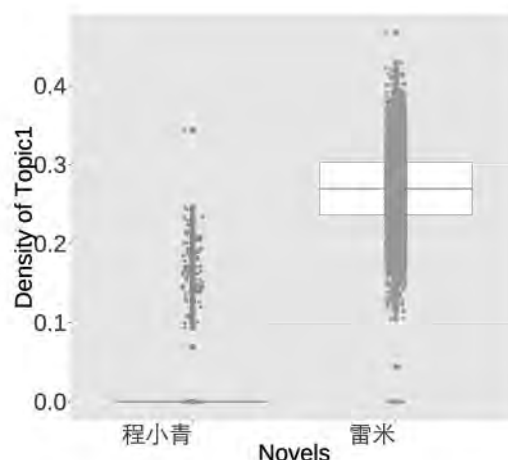


図8：Topic1 のジッター図

図7と図8は程小青と雷米の作品を2グループに分けて当該トピックの分布状況をジッター図で表すものである。図7と図8から明らかなように、Topic0とTopic1はそれぞれ程小青と雷米を特徴づけている。実際、トピックというより、作家ごとの特徴語のまとめと言っても過言ではない。作家間の言語特徴の相違を容易に考察できることを高く評価すべきであるが、細分化したトピックも当然望まれる。

大きく2つに分かれる理由として考えられるのは、コーパスがやや小さいため、同じ作家のテキストファイルはベクトルが非常に類似しているため、偏りやすくなる傾向にあることだ

ろう。ならば、ベクトルの処理において Doc2Vec ではなく、既に公開された訓練済みの事前学習モデルを使用して偏りの問題を解消できるかどうかを試みる。

タグがついたままだと、事前学習モデル³では大概未知語として処理されるため、実行の結果はやはり乱暴であった。よって、タグを除き、単語間にスペースをいれるだけのコーパスを取り上げた。その結果、51 個のトピックが出力された。表 5 に挙げたように、生成されたトピックはほぼ質が良くて非常に説明しやすい。

表 5 上位 10 個のトピックを構成する主要単語（事前学習モデル使用）

Topic	Keywords
0	霍桑不答 疑点 霍桑应道 疑案 疑难 嫌疑 霍桑道 可疑 霍桑 惊疑 怀疑 疑惑 侦查
1	昨晚上 昨夜 昨晚 午后 今晚 今夜 下午 当晚 中午 半夜 深夜 傍晚 上午 昨天
2	怎么 什么样 到底 不动声色 不到 什么 怎么样 下面 无声 怎样 底下 跌倒 落在
3	把门 敲门 大门 看门 门上 后门 房门 开门 门开 前门 车门 门外 进门 门前 门房
4	警局 警察们 警方 警察 公安局 报警 警署 警官 警士 警车 干警 公安 机关 巡官
5	车上 车里 车厢 车辆 黄包车 开车 汽车 司机 车子 车行 车窗 停车场 警车 吉普
6	教室 教师 教授 老师 上课 校园 小学 学生们 学校 同学 学生 中学 研究生 年级
7	杀人案 谋杀 杀人犯 凶手 凶案 谋害 疑案 杀死 杀人 侦探 自杀 犯案 刑事 嫌疑
8	婚姻 结婚 婚约 媳妇 不可能 想不到 不到 不行 不解 不觉 不是 没法 太太 没有
9	女孩子 女孩 少女 女儿 女生 怎么 无声 什么样 不动声色 女子 怎么样 什么 到底

このように、Doc2Vec の代わりに事前学習モデルを使用した Top2Vec を実行することで、作家間の特徴をより細かく考察することが可能になる。

5. おわりに

本稿では、Top2Vec の結果をもとに、程小青の作品全体におけるトピックを考察し、程小青の好みとスタイルを覗くことができた。また、Top2Vec による多数の作家間の比較に関する可能性も検討してみた。Top2Vec によって従来とは異なる視点から小説作品を解説することができるといえよう。

今後の課題としては、まず作家間のトピック考察を細かに行うつもりである。中国の作家のみならず、中日間の作家対比も望まれる。多言語のコーパスにおいて Top2Vec の実用性を試したい。また、タグを取り除くことによって単語の異なり語数も変わってくる。タグつきのコーパスとつかないコーパスをそれぞれ実行したモデルの結果を Information gain で評価して、最適なコーパスを作り上げたい。

³ 今回使われる事前学習モデルは「distiluse-base-multilingual-cased」という多言語対応の事前学習モデルである。

文献

- [1] 程海燕 (2015) 「论程小青侦探小说的本土化」 修士論文 安徽大学
- [2] 黄晨雯 (2020) 「中国のミステリー小説におけるトピック解析の試み」 『言語文化学』 1-17
- [3] 荆華 (2014) 「新文学侦探小说(1914-1949)叙事模式研究」 修士論文 遼寧大学
- [4] 周楠 (2015) 「近代侦探小说中的都市元素研究」 修士論文 上海師範大学
- [5] クジラ飛行机 (2016) 『Python によるスクレイピング&機械学習[開発テクニック]』 ソシム株式会社
- [6] 小林雄一郎 (2017) 『R による優しいテキストマイニング：機械学習編』 オーム社
- [7] 田畑智司 (2012) 「テキストマイニングからテキスト分析へ：Collins との共著作品における Dickens の文体」 『電子化言語資料分析研究 2011-2012』 3-17
- [8] 田畑智司 (2017) 「FLOB コーパスの意味構造：確率論的トピックモデルによる言語使用域の特徴づけ」 『言語文化共同研究プロジェクト 2016』 5-21
- [9] Blei, D.M., Ng, A. and Jordan, M. (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3:993-1022.
- [10] Blei, D.M. (2012) Probabilistic Topic Models. *Communications of the ACM*, 55(4): 77-84.
- [11] Dima Angelov (2020) TOP2VEC:Distributed Representations of Topics, *Computation and Language*, arXiv:2008.09470
- [12] Graham, S. and Milligan, I. (2012) Review of MALLET, produced by Andrew Kachites McCallum, *Journal of Digital Humanities*, 2(1).
- [13] Hofmann, T. (1999) Probabilistic Latent Semantic Analysis, *UAI'99 Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 289-296
- [14] Jockers, M. and Mimno, D. (2013) Significant themes in 19th-century literature, *Poetics* 41: 750-769

19 世紀米国大統領演説における分離不定詞 — Abraham Lincoln の事例から —

福本 広光

大阪大学大学院言語文化研究科

〒 560-0043 豊中市待兼山町 1-8

E-mail: fragilesunshine1632@gmail.com

概要 本研究では、米国歴代大統領による演説をコーパスとして、アメリカ英語における通時的な分離不定詞の実態・用法の一端について多角的に考察することを目的としている。本稿では、その予備的調査として、主に 19 世紀後半に活動した第 16 代大統領 Abraham Lincoln に特に着目して分析を行う。歴代の米国大統領の各演説に品詞タグを付与し、分離不定詞構造を抽出し、Lincoln が米国大統領の中では誰よりも先んじて分離不定詞を「多用」したという歴史的な頻度の傾向を明らかにした。事例研究においては、Lincoln の分離不定詞がどのような品詞的用法のもとでどのような統語パターンと共に用いられているかを分類し、構成要素としての特定の副詞に着目して文脈を観察するなど、複数の視座から分析することにより、彼の用いた分離不定詞の特徴的な文体やレトリックについて、新たな知見をいくつか提示することができた。

キーワード 分離不定詞, 米国大統領演説コーパス, Abraham Lincoln, レトリック

An Analysis of split Infinitive in the 19th Century US Presidential Speech: With Special Reference to Abraham Lincoln's Speeches

Hiromitsu Fukumoto

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract The present study attempts to discuss a part of the diachronic usage of split infinitive in American English from various angles, using the presidential speech text data included in the Corpus of Presidential Speeches (CoPS). As a preliminary survey, in this paper, we focus on Abraham Lincoln, the 16th president who was active in the middle of the 19th century. In order to extract the structures, we add POS-tags to each text of the speeches in CoPS and use a corpus analysis application, namely CasualConc. Due to the influence of prescriptivism, split infinitive seldom appeared in the data of presidents who took office by the mid-19th century; but this investigation reveals that Lincoln was the first US president to use the construction frequently in his speech. In the case study, we classify the part-of-speech usages (i.e., NOMINAL usage; ADJECTIVE usage; ADVERBIAL usage) of Lincoln's split infinitives. We also observe

the context focusing on specific adverbs, such as *not* and *so*. The multiple perspectives enable us to present several new findings on the style and rhetoric characteristic of the split infinitives used by him.

Keywords Split infinitive, Corpus of Presidential Speeches, Abraham Lincoln, rhetoric

1. はじめに

1.1. 導入

本稿では米国歴代大統領の演説を言語資料として (1) または (2) のような、不定詞マーカー *to* と原形動詞の間に副詞が挿入された、いわゆる分離不定詞 (split infinitive) のアメリカ英語における用法の一側面について、特に第 16 代大統領 Abraham Lincoln による事例を中心に分析する。

- (1) We need **to better understand** the impact of the relentless pressure on the health of those who attend to Alzheimer's patients. ¹ (TIME Magazine Corpus; 2005)
- (2) I think it is pretty common for people **to not know** what they want to do for a living until they are doing it. (COCA; WEB; 2012)

分離不定詞は一般的に規範主義の観点からは非難される語法である²が、現代英語ではやや寛容な立場をとる文法書が増えてきたこともあり (Curme 1931; Crystal 1984; Quirk et al. 1985; Huddleston and Pullum 2002 etc.), 再評価されつつある。とはいえ、改まった場面や状況においては使用を控えるよう指示する考え方も根強く存在し、その是非については依然議論の対象となっているのが現状である。

1.2. コーパスに基づく分離不定詞に関する先行研究

これまでの分離不定詞に関するコーパスを用いた研究には、Corpus of Contemporary American English (以下 COCA) や British National Corpus (以下 BNC) などの大規模コーパスを用いた英米語間の分離不定詞の比較や英語圏とアジア圏の英語における分離不定詞の比較など、主に現代英語を対象とした分離不定詞の使用実態の大枠を捉えることを目標とするものが多数見受けられる (Mitrascu 2009; Mikulová 2011; Perales-Escudero 2011; Supakorn 2012; Jang and Choi 2014; Johansson 2015)。しかし量的分析・質的分析のいずれにおいても通時的観点に立った実証的研究は決して十分ではない。³ 他にも、一部の小説、雑誌、新聞などを用いてテキストジャンルを限定した調査研究はいくつか存在する (後藤 2003a; 2003b; 2003c; Albakry 2005) が、演説を分析対象としたものは管見の限り例がない。米国で発行された小説や新聞等と同様、大統領による演説もその時代

¹ 以下、用例中の強調は筆者によるものである。強調箇所以外の用例中の表記は全て原文通りに記載している。

² Crystal (1995: 194)によると、1986年にBBC放送のラジオ番組 *English Now* に聴取者から寄せられた便りを調査した結果、文法に関する不満の一つに分離不定詞が含まれていたという。

³ コーパスを用いた、イギリス英語における分離不定詞の通時的研究には Calle-Martín and Miranda-García (2009)がある。その他、筆者はこれまでも Corpus of Historical American English (以下 COHA)を用いたアメリカ英語における分離不定詞の通時的使用について試験的な調査分析を行っている (福本 2018)ものの、さらなる考察が必要となる。

におけるアメリカ英語の枠を十分に投影したものであると考えることができるのではないだろうか。

1.3. 研究の目的

上記の内容を踏まえて、本研究の目的は、これまで英語史ないしは英語学研究の対象として注目されることが少なかった米国大統領演説コーパスを活用し、検索して得られた用例を参照しつつ、歴代米国大統領による分離不定詞使用の様相の一面に関して頻度や用法を中心として検討することである。本稿では、その予備的事例研究として、Abraham Lincoln に焦点を当てる。

2. コーパスの概要・調査手法

データソースとして、本調査では Grammar Lab による Corpus of Presidential Speeches (Brown, 2016, 以下 CoPS) 所収の全電子テキストを用いる。CoPS は、18 世紀後半の初代大統領 George Washington から 21 世紀初頭の Barack Obama までの米国歴代大統領による演説の(トラン)スクリプトを中心に集積したコーパスである。ここには就任演説や一般教書演説をはじめ、大学でのスピーチ、記者会見、テレビ討論など多様なジャンルから含まれている。「収録ファイル数や種類について人物ごとに偏りが見られる」といった、問題点と思われる箇所も散見されるものの、「米国歴代大統領によって発信された政治的資料」という単一ジャンルの史的データベースとして機能すると判断した。本コーパスを資料として用いた言語研究は現在のところ、ほとんど存在しないように思われる。しかし特定の大統領について集中的に文体などを分析する共時的研究はもちろん、ポリティカルディスコースなどに関わる通時的研究にも使用可能であると予想される。構成の一例として、以下の表 1 では、第 4 章にて用いるデータに含まれる演説の情報について具体的に示したものである。最右列は、該当の演説に生起する分離不定詞の用例数を表している。

表 1: CoPS に含まれる Lincoln によるデータ情報一覧⁴

File No.	Date	Name of Speech	Appearance frequency
0	Oct 16, 1854	At Peoria, Illinois	6
1	Jun 16, 1858	A House Divided Speech	6
2	Jul 6, 1852	Eulogy on Henry Clay	2
3	Feb 11, 1861	Farewell Address	0
4	Feb 27, 1860	Cooper Union Address	1
5	Mar 4, 1861	First Inaugural Address	3
6	Jul 4, 1861	July 4th Message to Congress	3
7	Dec 3, 1861	First Annual Message	1
8	Jan 1, 1863	Emancipation Proclamation	0
9	Aug 26, 1863	Public Letter to James Conkling	1
10	Dec 1, 1862	Second Annual Message	2
11	Dec 8, 1863	Third Annual Message	3
12	Dec 6, 1864	Fourth Annual Message	1
13	Nov 19, 1863	Gettysburg Address	0
14	Mar 4, 1865	Second Inaugural Address	0

これらのデータの中から、分離不定詞のような特定の品詞の語句を組み合わせた表現構造のみを収集するにあたり、各テキストファイルに含まれる全単語に品詞タグ付けアプリケーション CasualTreeTagger (今尾, 2019)により BNC タグの一つである CLAWS5 タグセットに基づいた自動品詞タグ付与を行った。その後コーパス分析アプリケーションである CasualConc (今尾, 2020)を用いて、分離不定詞に該当する構造を検索・抽出した。具体的な検索式は“to_TO0 *_AV* *_V*I”, “to_TO0 *_XX* *_V*I” および “to_TO0 *_P* *_AV* *_V*I” である。⁵

3. 各大統領における分離不定詞の頻度およびその変遷について

以下の表 2 は上記の調査によって各大統領に関するデータから抽出できた分離不定詞の頻度数を、便宜的に大統領就任順に並べたものである。表中には粗頻度に加え、各人のサブコーパスに含まれるおよその総語数と、その数値から算出した 1 万語あたりの平均調整頻度を記載している。

⁴ データの中には、当時口頭ではなく書面形式で伝達されていたとされる (cf: Teten 2003: 337) “Annual Message” (後の State of the Union Address (一般教書演説))や “Public Letter to James Conkling” といった、文書や書簡などの一般的な意味での「演説(speech)」と呼ぶことが適切か判断しづらいものも含まれていることに注意すべきであろう。ただし本稿ではこれらを区別することなく、全て「演説の一種」とであるとみなした上で、等価なものとして扱っている。

⁵ 前者 2 つに関しては、実際の検索においては “to_TO0 *_AV|XX* *_V*I”を用いた。

表 2: 米国の歴代大統領と収集された分離不定詞の全用例数一覧^{6,7,8}

Name of Presidents	Raw frequency	Tokens	Adjusted frequency	Name of Presidents	Raw frequency	Tokens	Adjusted frequency
George Washington	0	31,841	0.00	Benjamin Harrison	12	78,174	1.54
John Adams	0	14,762	0.00	Grover Cleveland (2)	31	81,880	3.79
Thomas Jefferson	0	40,367	0.00	William McKinley	10	93,896	1.07
James Madison	0	36,374	0.00	Theodore Roosevelt	7	198,208	0.35
James Monroe	0	50,310	0.00	William Howard Taft	4	120,465	0.33
John Quincy Adams	0	37,122	0.00	Woodrow Wilson	0	80,334	0.00
Andrew Jackson	1	158,255	0.06	Warren G. Harding	1	28,983	0.35
Martin Van Buren	0	64,984	0.00	Calvin Coolidge	0	75,138	0.00
William Henry Harrison	0	8,447	0.00	Herbert Hoover	12	89,572	1.34
John Tyler	0	70,100	0.00	Franklin D. Roosevelt	0	132,430	0.00
James K. Polk	1	105,247	0.10	Harry S. Truman	0	37,033	0.00
Zachary Taylor	0	11,430	0.00	Dwight D. Eisenhower	0	18,047	0.00
Millard Fillmore	0	39,821	0.00	John F. Kennedy	11	160,877	0.68
Franklin Pierce	1	63,956	0.16	Lyndon B. Johnson	44	251,755	1.75
James Buchanan	0	81,730	0.00	Richard M. Nixon	3	68,139	0.44
Abraham Lincoln	29	96,094	3.02	Gerald R. Ford	3	40,609	0.74
Andrew Johnson	0	99,826	0.00	James Carter	6	102,162	0.59
Ulysses S. Grant	17	105,102	1.62	Ronald Reagan	30	215,260	1.39
Rutherford B. Hayes	13	68,684	1.89	George H.W. Bush	11	99,938	1.10
James Garfield	0	2,974	0.00	William J. Clinton	23	152,839	1.50
Chester A. Arthur	1	51,190	0.20	George W. Bush	20	108,921	1.84
Grover Cleveland (1)	20	77,059	2.60	Barack Obama	54	198,699	2.72

変化についてさらに見やすくするため、各大統領による分離不定詞の平均頻度について図示したのが図 1 である。

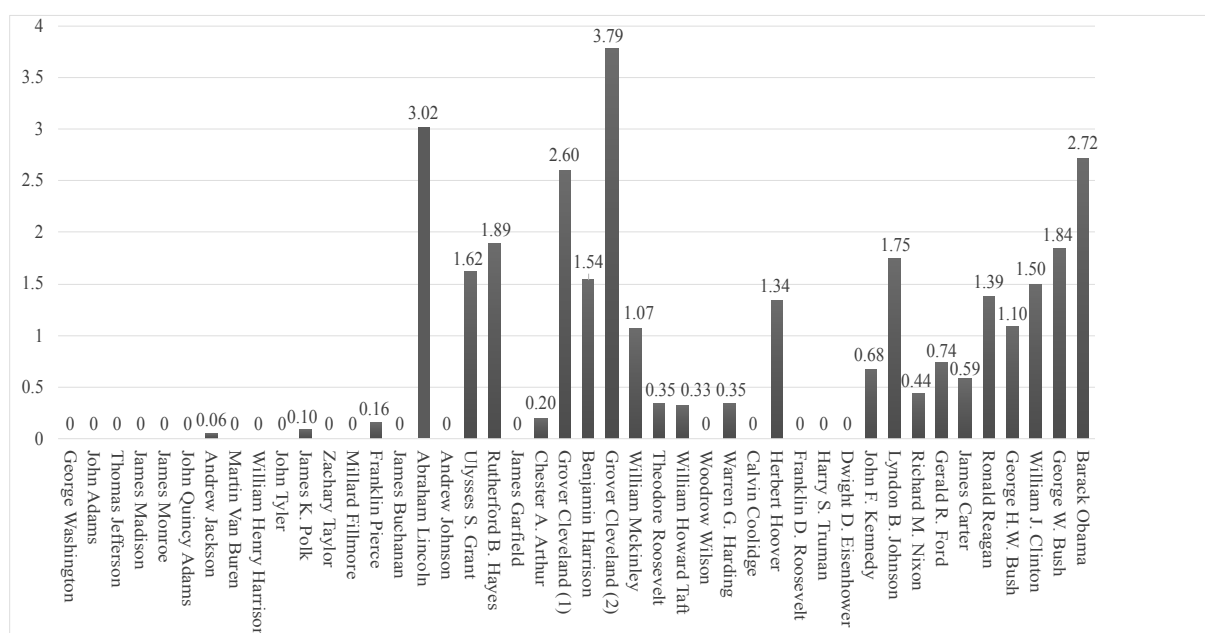


図 1: 米国大統領演説に出現する分離不定詞の頻度数の変遷 (per 10,000 words)

⁶ CoPS には大統領在任期間外においてなされた演説のデータも収録していることがある。即ち大統領就任よりも前の演説であれ、その人物の政治的経歴に影響したものであれば全て含まれている。そのため必ずしも時系列順になっておらず、ある大統領による演説が行われた期間が他の大統領在任期間と重なっている場合もある。

⁷ Cleveland は 1885-1889 年と 1893-1897 年において大統領を務めた。コーパス内ではその区別は特になされず、第 1 期も第 2 期も一つのデータの中にまとめて組み込まれていたが、ここでは分けて考え、Cleveland (1) と Cleveland (2) としている。

⁸ 表中の“Tokens” (総語数) については、コーパスデータをダウンロードした元サイトに記載されているものとは数値が異なることがほとんどであるが、ここでは先述のファイルに含まれる全ての未加工テキストを CasualConc の Word Count 機能にかけることで得られた語数の値を採用している。

George Washington をはじめ、18 世紀の大統領のデータからは分離不定詞を含む用例は一件も抽出されなかった。本コーパス中最初の事例は、第 7 代大統領 Andrew Jackson によって 1834 年に使用された、*both A and B* 構文の一部を構成した以下の文である。

- (3) ...then will the violation of privilege as it respects that House, of justice as it regards the President, and of the Constitution as it relates **to both be** only the more conspicuous **and** impressive.

(1834; Protest of Senate Censure)

その後は散発的に確認されているが、19 世紀前半も演説において分離不定詞を使用する大統領はほとんどおらず、その例外となりうるような場合でも、頻度は極めて低いものであった。仮に 1 万語あたり 1 回以上分離不定詞を使用することを「多用する」と表現するならば、今回のデータにおいて最初に「多用した」のは Abraham Lincoln である。⁹ そして 19 世紀後半の大統領は、数の多寡はあれども自由に、分離不定詞を「多用」する傾向にある。つまり彼が米国大統領の分離不定詞使用について大きな転換点となったことは間違いないだろう。とはいえ少なくとも 20 世紀後半になるまではかなり多くの用例が得られる場合もあれば、一例も見られない場合もあり、人物によって使用/不使用の差が明白に分かれるという結果となった。¹⁰ 20 世紀後半、特に John F. Kennedy 以降の大統領はある程度慣習的に分離不定詞を使用しており、分離不定詞の使用が演説というレジスターにおいても定着しつつあることが見て取れる。

本章では通時的な量的観察を中心に行い、Lincoln が分離不定詞を他の大統領よりも先んじて「多用」したことを指摘した。次章ではその事実に基づき、彼が実際に用いた分離不定詞の用例について特徴的な側面についての分析を行う。

4. Lincoln の演説における分離不定詞の事例分析 ¹¹

4.1. 品詞的用法・統語パターンによる分類

本章では、Lincoln の演説中に出現する分離不定詞の事例について、その実態や文体的特徴を質的に分析する。彼が分離不定詞をどのように用いているのかを判断するための一つの指標として、*to* 不定詞の品詞的用法（いわゆる名詞的用法、形容詞的用法、副詞的用法）の分布から捉えたい。

⁹ Lincoln が分離不定詞を多用している背景について、岩根久教授、田畑智司教授、山田彬亮講師（いずれも大阪大学）より、彼の出身地域の言語傾向（方言）や受けた教育環境など、その出自と何らかの関わりがあるのではないかという御示唆を頂戴した。本稿では詳しくは立ち入らないが、更なる検討の価値がある事項である。

¹⁰ 大統領の在任期間が人によって異なるということもあり、本コーパスに収録される各大統領のサブコーパスに含まれるテキストの「数」や「種類」にはかなりのばらつきがある。しかしながら、既に述べたように本研究では CoPS 所収の全てのテキストを分析対象としており、各大統領の分離不定詞の使用頻度の差異を各サブコーパスの「総語数」によって調整しているため、その点は大きな問題とはならない。

¹¹ 言うまでもないことだが、今回行った調査は Lincoln によって用いられた全ての分離不定詞を網羅的に収集したものではなく、あくまで CoPS 所収のデータに限定したものである。調査対象となる資料の数を増やせばさらに多くの用例を収集できる可能性がある。例えば、本データには収録されていなかったものの、Visser (1966: 1042)において、分離不定詞の用例の一つとして 1862 年 7 月 12 日に行われた “Appeal to the Border States in behalf of Compensated Emancipation” より、以下の一節が引用されている。

“How much better *to* thus *save* the money which else we sink forever in the war.” (Visser 1966: 1042)

元の *to* 不定詞のもつ機能について考えることは、その変異形たる分離不定詞の機能について判断することにつながるため、このやり方も一つの分析手法として有効であろう。更にそれらを、どのような統語パターンのもとで用いているかに関して分類する。

4.1.1. 名詞的用法の分離不定詞

表 2 で示したように、CoPS の Lincoln データからは全部で 29 の分離不定詞の用例が抽出され、そのうち 19 例 (約 65.5%) が名詞的用法によるものであった。¹² この割合の高さは *to* 不定詞全体の分布にも言えることかもしれないが、少なくともこの結果からは名詞的用法による用例が極めて多いことが示される。本小節では名詞的用法をさらに区分けし、その下位カテゴリーとして見られた統語パターンについて示す。

● Object (S + V + split infinitive)

以下のような、先行する本動詞の目的語として機能するものが特に目立ち、全部で 7 例あった。

- (4) a. I wish further to say, that I do not propose to question the patriotism, or to assail the motives of any man, or class of men; but rather **to strictly confine** myself to the naked merits of the question.
(1854, At Peoria, Illinois)
- b. But even if we fail **to technically restore** the compromise, it is still a great point to carry a popular vote in favor of the restoration.
(1854, At Peoria, Illinois)

ここでの本動詞には *consent, fail, prepared, promise, propose, require, wish* がある。これらには *to* 不定詞が後続してコロケーションをなす傾向があり、「要望や提案」あるいは「約束」などといった政治的スピーチに必須の動詞が占める割合が大きいの。ほぼ同様のパターンとしては、1 例のみであるが、形式目的語 *it* と共起する用例もあった。

● Object (S + V + it + split infinitive)

- (5) we find it impossible **to not believe** that Stephen and Franklin and Roger and James all understood one another from the beginning, and all worked upon a common plan or draft drawn up before the first lick was struck.
(1858, A House Divided Speech)

本調査の対象となった分離不定詞では、あらゆるカテゴリーの中で、この「目的語」としての機能を持つ用例が最も多く発見され、名詞的用法全体の中でもおよそ 4 割を占めている。Lincoln は「先行動詞 + *to* 不定詞」の形において分離不定詞を用いることが多かったものと思われる。他には、主語として機能するパターンが頻出した。以下のような形式主語構文の真主語として活用されることが多く、4 例が収集された。

¹² 分析対象とした分離不定詞の全用例について、用法の詳細とともに本論末に掲載した。(Appendix 参照)

● Subject (It + V + split infinitive)

- (6) a. It would be very agreeable to me, **to thus meet** my old friends, at my own home; but I cannot, just now, be absent from here, so long as a visit there, would require.

(1863, Public Letter to James Conkling)

- b. The extensive blockade has been constantly increasing in efficiency as the Navy has expanded, yet on so long a line it has so far been impossible **to entirely suppress** illicit trade.

(1863, Third Annual Message)

また、類するカテゴリーとして、*of* や *for* を伴う用例があり、具体的に以下の 2 つが見つかった。

● It is (was)... for (of)... split infinitive

- (7) a. Finally, I insist, that if there is ANY THING which it is the duty of the WHOLE PEOPLE **to never entrust** to any hands but their own, that thing is the preservation and perpetuity, of their own liberties, and institutions.

(1854, At Peoria, Illinois)

- b. Plainly enough now, it was an exactly fitted niche, for the Dred Scott decision **to afterwards come in**, and declare the perfect freedom of the people, to be just no freedom at all.

(1858, A House Divided Speech)

これらの用例を「形式主語構文の真主語」として同一のものと考え、本パターンも名詞的用法の下位カテゴリーとしては比較的多く、全体の約 3 割を占めている。他に、文頭位置に配置されて主語として機能するパターンに分類されるものは以下の 1 例である。

● Sentence Head (split infinitive +V)

- (8) **To now abandon** them would be not only to relinquish a lever of power, but would also be a cruel and an astounding breach of faith.

(1863, Third Annual Message)

主語としての使用に際しては、形式主語構文の真主語として使用されるパターンがそのほとんどを占めていたものの、単独で文頭に配置するパターンが僅かながらも確認されたことは興味深い。名詞的用法で文頭に分離不定詞を用いられているのは CoPS のデータにおいてはこの 1 例のみであり、彼に特徴的な用法の一つである。分離不定詞が補語として用いられる用例もあり、目的格補語に 3 つ、そして主格補語に 1 つの合計 4 つの用例があった。

● Objective Complement

- (9) In your political contests among yourselves, each faction charges the other with sympathy with Black Republicanism; and then, to give point to the charge, defines Black Republicanism **to simply be** insurrection, blood and thunder among the slaves.

(1860, Cooper Union Address)

● Subjective Complement

- (10) Auxiliary to all this, and working hand in hand with it, the Nebraska doctrine, or what is left of it, is to educate and mould public opinion, at least Northern public opinion, **to not care** whether slavery is voted down or voted up.

(1858, A House Divided Speech)

以上の調査結果をまとめると、名詞的用法の割合が非常に高いことがわかった。統語パターンから用法を下位分類すると、本動詞に後続する形での目的語としての使用や形式主語構文の真主語としての使用が多いのは確かである。その一方で、各カテゴリーの用例数等に極端な偏りがあるわけではなく、多様な用法のもとでバランス良く使用していたことも窺える。主語として文頭位置に配置される(8)のような用例は、本調査においては彼単独のものであることもわかった。

4.1.2. 形容詞的用法の分離不定詞

形容詞的用法に区分されるものは、全部で4例(13.8%)あり、大きく2つのパターンに分けられる。即ち「名詞に後続してそれを修飾するもの」「主格補語としての役割を持つもの」である。

● NOUN + split infinitive (Objective Complement)

以下のような、特定の名詞の後に置かれ、それを修飾するための「目的格補語」としての役割を持つ分離不定詞が3例確認できた。

- (11) Having never been States, either in substance or in name, outside of the Union, whence this magical omnipotence of “State rights”, asserting a claim of power **to lawfully destroy** the Union itself?

(1861, July 4th Message to Congress)

● Subjective Complement

1例のみであるが、主格補語としての機能を持つ用例についても抽出した。

- (12) And as it is **to so go** at all events, may we not agree that the sooner the better?

(1864, Fourth Annual Message)

名詞的用法の場合と比較して形容詞的用法のもとで分離不定詞が使われている場合は少ないものの、それでも一定数の用例が見受けられた。

4.1.3. 副詞的用法の分離不定詞

副詞的用法に区分されるものは全部で6例(20.7%)であった。特定の統語パターンがあるわけではないが、意味的にこのカテゴリーに区分けされると判断されたものが多くあった。

● Set Phrase (so as to/ in order to)

「動作の目的」を意味する *so as to* のセットフレーズとともに用いられている用例が1例確認された。

- (13) “let us be more specific--let us amend the bill so as **to expressly declare** that the people of the territory may exclude slavery.”

(1858, A House Divided Speech)

● The Others

統語パターンとしては上記のいずれにも区分されないが、意味的に *to* 不定詞が副詞としての働きをしているものも複数発見された。

- (14) The Treasury report including a considerable sum now which had previously been reported from the Interior, sufficiently large **to greatly overreach** the sum derived from the three months now reported upon by the Interior and not by the Treasury.

(1862, Second Annual Message)

4.1.4. 4.1 節のまとめ

本節では、米国大統領として初めて分離不定詞を「多用」した人物である Lincoln が、どのようなやり方で分離不定詞を用いたのかについて、品詞的用法やその下位区分である統語パターンに基づいて分類を行った。不定詞用法の中では、名詞的用法としての使用が極めて多く、中でも本動詞に後続する目的語としての使用あるいは形式主語構文の真主語としての使用が多かったが、形容詞的用法や副詞的用法にもそれぞれ少数ながら一定の用例を確認することができた。ただし、分離不定詞の用法についてはその性質上、一般的な *to* 不定詞の用法と共通している部分が多々ある。そのため、「今回指摘した諸現象は（当時の）*to* 不定詞一般に見られる現象とは同様のものか？」「指摘事項は他のジャンルと比較した時に「演説」というジャンル特有のものか？」「他の大統領と比較した場合には同じことが言えるのか？」といった事柄については未だ疑問の余地が残るため、それらの調査・考察については今後の課題とする。

4.2. Lincoln に特徴的な分離不定詞再考：頻出した副詞を軸として

本節では前節と少し視点を換え、構成要素となる副詞に焦点を当て、特徴的なパターンの分離不定詞にアプローチする。表 3 は、分離不定詞各用例の構成要素をなす副詞を、頻度数と共に全てリストアップしたものである。

表 3：分離不定詞の構成要素となった副詞およびその頻度数（括弧内は粗頻度）

<i>not, so</i> (5x), <i>lawfully, now</i> (2x), <i>afterwards, correctly, duly, either, entirely, ever, expressly, finally, greatly, never, simply, strictly, technically, thus, yet</i> (1x)

意味的カテゴリーとしては、「様態」・「時」・「否定」を表す副詞が構成要素として目立っている。本節では特に、データに頻出した *not* ならびに *so* を伴う分離不定詞について集中的に扱う。¹³

4.2.1. 否定辞 *not* を伴う分離不定詞について

Lincoln のデータにおける分離不定詞の特徴の一つに、*not* を *to* と原形動詞の間に挿入するパターンの分離不定詞が他の大統領よりも多い、ということがある。19 世紀の大統領では、“*to*+*not*+原形動詞”の構造を用いていたのは彼ただ一人であり、20 世紀以降で見ても他の使用者はごく限られている。具体的には、Reagan が 2 回、Lyndon Johnson, Obama が 1 回ずつと、その頻度も極めて低い。片や Lincoln のテキストからは 5 回抽出され、その演説から唯一 3 回以上このタイプの構造が観察されている。この傾向がいかに関特殊であることを示すための参考調査として、アメリカ英語での書き言葉テキスト(fiction, popular magazine, newspaper, non-fiction, TV/Movie script) 中心のデータ

¹³ 渡辺秀樹教授 (大阪大学) に、表 2 には他の大統領が用いていない副詞パターンがあるのではないかと示唆をいただいた。御指摘の通り、分離不定詞の要素として、コーパスからは彼にのみ出現する副詞も散見される。(具体的には、*afterwards, correctly, duly, lawfully, technically, yet* がそれに当たる) 中でも *afterwards, lawfully, yet* を用いたものは一般的にも周辺的な事例と言え、その分析は興味深いものとなることが予想されるので、今後稿を改めて論じる必要があろう。

においてこの構造がどれほどの頻度で用いられていたか確認すべく、COHA を用いて 200 年間の頻度の推移を調べると、図 2 のような結果が得られた。

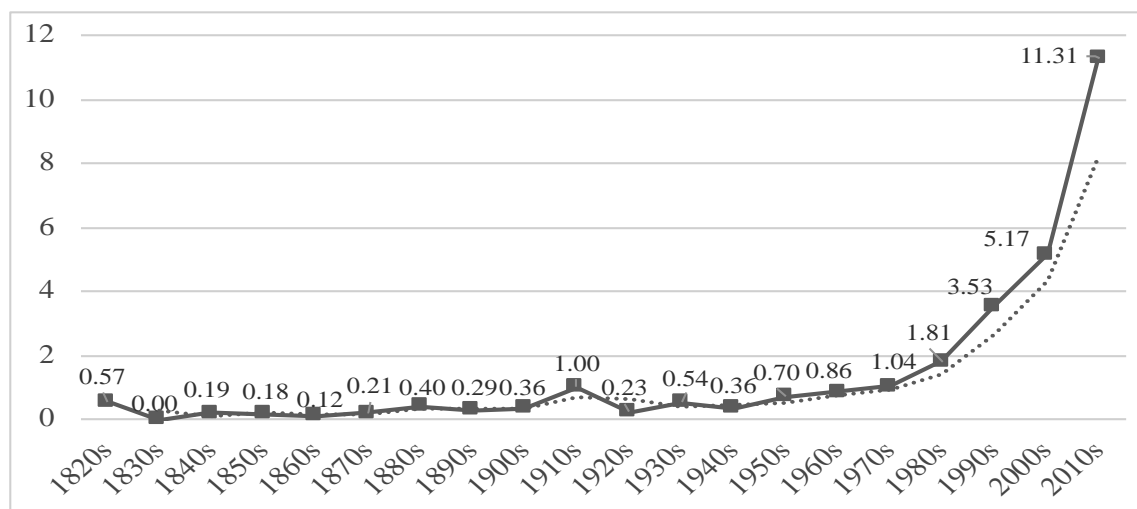


図 2: 1820 年から 2019 年までのアメリカ英語における“to not DO”構造の頻度変化 (pmw)¹⁴

否定辞 *not* を伴う分離不定詞自体、いずれの時代も高頻度とはいえない非常に周辺的な用法であるが、Lincoln が政治家・大統領としてキャリアを積んでいた 1850—1860 年代は特に低い（それぞれ 100 万語当たり 0.18 回, 0.12 回）。頻度が相対的に上昇し始めた 20 世紀に入ってから、Curme (1931: 459) は “[E]specially *not* clings to the old position before *to*” と、Jespersen (1940:335) は “[N]ot and *never* are often placed before *to*” と、更に Schibbye (1965: 26; note) に至っては “[C]ertain adverbs however can never be placed between *to* and the infinitive: *just, only, merely, not.*” とまで言及している。*not* を伴う分離不定詞は、当時刊行された文法書を見る限り、分離不定詞の中でも非常に稀な存在であり、人によっては半ば非文法的用法とも取れるような扱いをしていたと推測できる。実際、アメリカ英語一般において比較的大きな上昇傾向が見られ、これが使用されるようになってきたのは 20 世紀末以降である。故に、量的観点からも、19 世紀中頃に既にこれを多用していた Lincoln の特異性が際立っている。¹⁵ このタイプの分離不定詞をどのように用いていたのかについて更に検討すべく、以下にその用例を全て示す(なお (5), (10) は 4.1 節での用例の再掲である)。

- (5) ... in such a case, we find it **impossible to not believe** that Stephen and Franklin and Roger and James all understood one another from the beginning, and all worked upon a common plan or draft drawn up before the first lick was struck. (1858; A House Divided Speech; 再掲)

¹⁴ 図中の点線は頻度推移の移動平均を表している。(図 3 も同様)

¹⁵ OED Online で *not* を伴う分離不定詞を含む引用文を検索したところ、1650 年以降、1848 年までは該当する用例が一例もなく、1848 年 (s.v. *machine*, n. 5e), 1872 年 (s.v. *have*, v. 43b (b)) には 1 例ずつ確認できたものの、その後 1912 年 (s.v. *table dance in table*, n) になるまで用例が全く見受けられなかった。Visser (1966: 1040) には、該当する用例は 1650 年以降、1884 年になるまで引用されておらず、その次に出てくるのは 1904 年の用例である。これらの事実も当時、この構造が非常に低頻度なものであったこと示し、主張を支持する根拠となっている。

- (10) Auxiliary to all this, and working hand in hand with it, the Nebraska doctrine, or what is left of it, *is* to educate and mould public opinion, at least Northern public opinion, **to not care** whether slavery is voted down or voted up. (1858; A House Divided Speech; 再掲)
- (15) Missouri would not yield the point; and congress -- that is, a majority in congress -- by repeated votes, *showed* a determination **to not admit** the state unless it should yield. (1852; Eulogy on Henry Clay)
- (16) But if it is a sacred right for the people of Nebraska to take and hold slaves there, it is equally their sacred right to buy them where they can buy them cheapest; and that undoubtedly will be on the coast of Africa; provided you will *consent* **to not hang** them for going there to buy them. (1854; At Peoria, Illinois)
- (17) While it is fortunate that this great interest *is* **so** independent in its nature **as to not have demanded and extorted** more from the Government, I respectfully ask Congress to consider whether something more cannot be given voluntarily with general advantage. (1861; First Annual Message)

(15)が形容詞的用法, (17)は副詞的用法, その他は名詞的用法のもとで用いられていた。“to not”に後続する原形動詞としては *admit, hang, believe, care* といった動作動詞が主である。一方, (17)のように完了不定詞が後続する用例も一つ見受けられ, 多様な文脈においてこのタイプの分離不定詞が用いられていることがわかる。

いずれの用例も分離不定詞の前に先行する本動詞がある(用例中イタリック)が, 分離不定詞を用いることで *not* は *to* 不定詞中の原形動詞を修飾している, ということが明らかである (cf: Quirk et al. 1985: 497; Huddleston and Pullum 2002: 581–582)。他の特徴として, (5)は否定語と共起することで二重否定を形成し(「信じない」ことが不可能である, 即ち「信じないわけにいかないということがわかる」という意味である), (17)は「程度」を表す構文の *so...as to* の構文と共起している(「大変独立した性質を持つものであり, 政府から要求や強要をされることがなかった」の意味)ということがある。(用例中太字)

これ以外の要因として, Crystal (2006:126) は“...[I]t [split infinitive] is popular because it is rhythmically more natural to say. The basic rhythm of English is a ‘tum-te-tum’ rhythm – what in main tradition in English poetry is called an *iambic pentameter*, with strong (*stressed*) and weak (*unstressed*) syllables alternating...”と述べ, 英詩の自然なリズム, 即ち「弱強五歩格」を構成することもできるとして, レトリックから見た分離不定詞の有用性を認めている。この中では 用例 (5)と(15)の分離不定詞の箇所付近では「弱強弱強」の韻律が連続して使用されており, 「弱強五歩格」のリズムではないが自然な音調を整えるという点から分離不定詞は機能しているともいえよう。¹⁶

¹⁶ 渡辺秀樹教授 (大阪大学) より, 英語は「前置詞 (もしくは冠詞) + 名詞」などの語順が主体の言語であり, 殊に散文において自然なリズムとして機能するのは「弱強」のリズムが連続する場合である (従って Crystal による *iambic pentameter* という表現は誤解を招きうる)。分離不定詞を用いることにより, 「弱強」のリズムが連続しているのであれば, 分離不定詞が英文の自然なリズムを構成するのに貢献していると判断できるので, ここでそれについて特筆するべきではないか」との御教示をいただいた。分離不定詞と韻律との関わりについて分析する上で重要となりうる視点であり, ここに特に記す。

(5)' ...to not believe that Stephen and Franklin and Roger and James ...

(15)' ...to not admit the state unless it should yield.

この構造を用いていた他の大統領、即ち Lyndon Johnson, Reagan, Obama による用例からは、分離不定詞の中心的な使用動機としては文の曖昧性や文の不自然さを解消し、メッセージを適切に伝えるという試みによるものであることが判断できる。一方で、例えば Lyndon Johnson の “to not provoke”などは韻律に配慮した結果使われた用例ではないかと推測できる。¹⁷

to 不定詞の内容を否定する、not を伴う分離不定詞は 20 世紀後半になるまで一般的にはほとんど使用されていなかったが Lincoln の演説では比較的頻出したことから、本小節にて興味深い事実として分析対象とした。使用の動機は、文の曖昧さや不自然さを解消するという、どちらかといえば「普通」なもので、それを様々な文脈において繰り返し使っていたといえる。それ以外にも、英語の自然な韻律との関わりについて指摘できる用例が一定数あり、分離不定詞は彼なりの意味的・音韻的レトリックの一つとして機能させるものでもあったのではないだろうか。

4.2.2. 副詞 so を伴う分離不定詞について

先述の通り、Lincoln の演説に特徴的なもう一つの分離不定詞として、to 不定詞の間に so を嵌入させたものも挙げられる。本調査に使用したデータにおいて、“to+so+原形動詞”の構造を他に使っているのは 19 世紀の大統領からは Grant が 2 回, Hayes が 1 回, Arthur が 1 回, Cleveland が 5 回, (内訳: 第 1 期に 2 回, 第 2 期に 3 回) Benjamin Harrison が 2 回。20 世紀以降は、Theodore Roosevelt から 1 回, Taft から 1 回抽出されているが、この後は一例も抽出されていない。彼はこれを 5 回使用しており、歴代大統領の中では Cleveland と並んで最多である。このパターンについても、参考として先程と同様、COHA を活用して史の変遷を調査・観察すると図 3 のようになる。

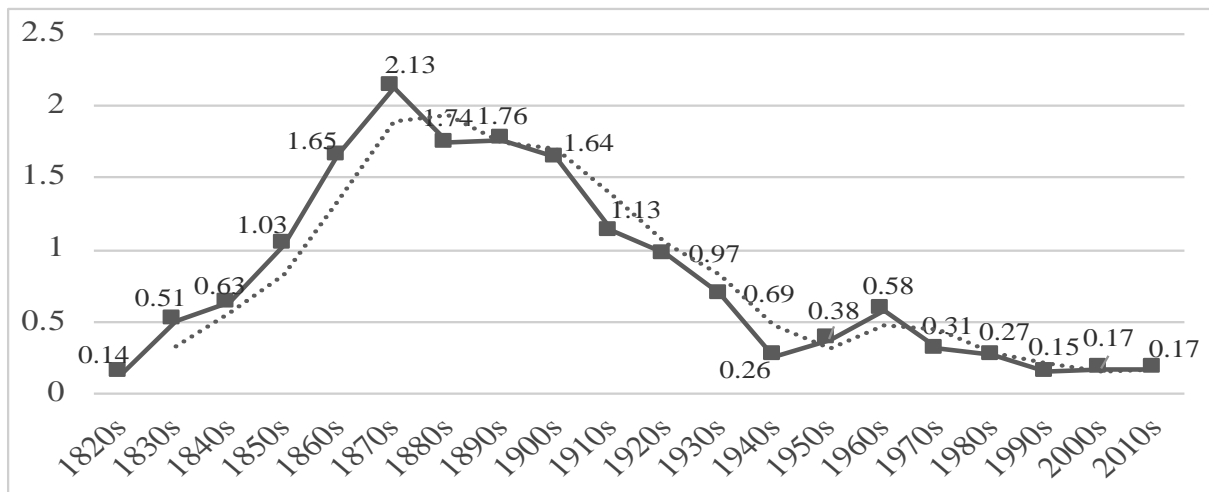


図 3: 1820 年から 2019 年までのアメリカ英語における“to so DO”構造の頻度変化 (pmw)

¹⁷ I realize that someone might indicate, because the suggestion in all its entirety wasn't carried out, there might be some difficulty between us, but my object in life has always been to not provoke fights, but to prevent them, if possible.

(Lyndon Johnson; 1964, Press Conference at the State Development)

このタイプも、200 年間を通して頻度が少ない周辺的な事例であり、全体的に前節にて扱った *not* を伴うタイプと同程度か、やや少ない頻度で推移している。変化傾向は *not* の場合と真逆に近く、19 世紀半ばから後半にかけて徐々に増加しているが、1870 年代をピークに減少の傾向にある。1910 年代を境に 100 万語あたり 1 回も使われなくなり、20 世紀後半には更に大きく頻度が下がっている。従って、19 世紀後半から 20 世紀初頭にかけて主に使用されていたが、現在ではほぼ使われない用法である。1850–60 年代はちょうどこの用法が増加傾向にある時期であったため、頻度的には Lincoln が多用していてもおかしくない。この調査結果は Taft 以降の大統領が一例も用いていないこととも矛盾しないものであった。用法についてさらに質的な分析を行うべく、*so* を伴う分離不定詞の全用例を詳細に観察する(ただし、(12)は 4.1 節の用例の再掲である)。

(12) And as it is **to so go** at all events, may we not agree that the sooner the better?

(1864; Fourth Annual Message; 再掲)

(18) After several failures, and great labor on the part of Mr. Clay **to so present** the question **that** a majority could consent to the admission, it was, by a vote, rejected, and as all seemed to think, finally.

(1852; Eulogy on Henry Clay)

(19) On Judge Douglas' motion a bill, or provision of a bill, passed the Senate **to so extend** the Missouri line.

(1854; At Peoria, Illinois)

(20) It was believed, however, that **to so abandon** that position under the circumstances would be utterly ruinous;...

(1861; July 4th Message to Congress)

(21) Care has been taken **to so shape** the document **as to** avoid embarrassments from this source.

(1863; Third Annual Message)

(20) は名詞的用法、(12) は形容詞的用法、それ以外は副詞的用法のもとで用いられている。“to+so”に後続する原形動詞としては *extend*, *present*, *abandon*, *shape*, *go* といった、動作動詞が主に用いられている。この *so* を用いた分離不定詞には、複数の用法パターンがある。(12), (19), (20)の用例を観察すると、例えば(19)は、Missouri line を拡張するという法案が上院で可決されたことについて、*so* を分離不定詞の要素とすることで、「そのようにして(拡張する)」という意味の様態の副詞として機能させており、(12)や(20)でも同様である。一方、例文中に太字で示したように、(18)は「目的」を表す *so...that* 構文（「クレイ氏には、大多数の人がその承認事項に同意できるよう、問題点を**提示するの**に多大な苦労をおかけした後,...」）（訳はいずれも筆者による）、(21)も「目的」を表す *so...as to* 構文の一部（「この情報についての**困惑を避けるべく、文書を作成することに注意が払われた**」）として使用されている(どちらも「A をするために B を行う」という文脈で用いられている）。他の大統領の用例と比較すると、Cleveland による 1 例¹⁸ 以外は、概ね同じ用法のもとで使われて

¹⁸ In conclusion I earnestly invoke such wise action on the part of the people's legislators as will subserve the public good and demonstrate during the remaining days of the Congress as at present organized its ability and inclination **to so meet** the people's needs **that** it shall be gratefully remembered by an expectant constituency. (Cleveland1; 1886, Second Annual Message)

「人々の要望に**とても沿うものであったので**、有権者と予想される方々に感謝の念を持って記憶されるだろう」の意。程度を表す *so...that* 構文の構成要素の一部となっている。

いたため、このパターンについても、他の大統領と同様な「普通」の用法を「繰り返し」利用していたということになる。

本調査により新たに確認できた事項は、Lincoln はどの大統領よりも先んじてこの構造を多用していたこと、そしてその用法は「様態をあらわす *so*」ならびに「*so... that* 構文や *so... as to* 構文」といった、特定の構文の構成要素」としての使用に分類されることである。

4.2.3. 4.2 節のまとめ

本節ではまず、得られた分離不定詞の構成要素となった副詞について傾向を概観した後、量的特徴のある、換言すれば特に高頻度な分離不定詞のパターン 2 種類について重点的に考察を行った。ここでは、下記の 2 点が主たる知見として挙げられる。

- ・ *to* 不定詞を否定する“*to not V*”の構造は、当時一般には殆ど用いられていなかったが、Lincoln の演説テキストには比較的頻出し、曖昧性の解消や韻律との関連が見られた。
- ・ 主に 19 世紀末から 20 世紀にかけて用いられていた“*to so V*”の構造は、単独で様態の意味を表す用法の他に、*so... that* 構文や *so... as to* 構文の構成要素の一部としても機能していた。

5. おわりに

本稿では、これまで英語学的な分析の対象となることが決して多いといえなかった演説（スピーチ）のジャンルから、Abraham Lincoln による分離不定詞の用例を対象に、頻度や用法などについて検討を行った。そして 19 世紀中頃以降後半にかけてのアメリカ英語での分離不定詞使用の様相の一端について文体的な記述分析を行い、彼が用いていた分離不定詞の特徴に関して、いくつかの新たな知見を提示した。

米国大統領の演説は、その世界的な影響力により小説や新聞などよりも言語的に保守的な傾向にあると仮定する¹⁹と、規範文法の全盛期である 19 世紀前半に分離不定詞が散発的にしか見られなかったことは理解できるが、Lincoln がその傾向を打ち破り分離不定詞を多用していた、という事実は非常に興味深い。英語史的観点からは分離不定詞が一般的に使われるようになったのは 19 世紀中頃から 20 世紀初頭にかけてであるという指摘が多い（cf: Visser 1966: 1036 etc）。そしてこれも本調査によるデータと客観的な整合性があるものであった。

今後の課題として、比較対象とする大統領の人数を増やして本稿と同様の調査を行い、英語文体論の考え方を援用しつつ「米国大統領の演説における分離不定詞」というテーマを軸に通時的研究を行いたい。また、本発表の分析対象は演説のみであるので、他のジャンルの用例と比較して今回の分析対象とした分離不定詞が特徴的なのかということについても今後検討し、論じてゆく必要があるだろう。

¹⁹ 図 1（本稿 3 節）のデータを元に、「分離不定詞の使用を許容するか、しないかについての判断は各大統領によって大きく異なる」という点、さらにそのことから「（米国大統領の）演説は、ジャンルとして見たときに日常言語とは一線を画すもので、言語的に保守的な傾向を持つ可能性が高い」という内容について、田畑智司教授（大阪大学）に御示唆いただいた。

謝辞

研究を進めるにあたり、田畑智司先生（大阪大学）をはじめ共同研究プロジェクトの先生方や院生メンバーの皆様には有益なご助言をいただいた。また、本稿執筆にあたり、草稿段階から渡辺秀樹先生（大阪大学）に多大なご指導を賜ったことも特筆すべき点である。この場をお借りして、深く感謝申し上げたい。

Appendix: Abraham Lincoln によって使用された分離不定詞の用法別全用例（本稿言及順）

<名詞的用法>

Object (S + V + split infinitive)

- ❖ I wish further to say, that I do not propose to question the patriotism, or to assail the motives of any man, or class of men; but rather **to strictly confine** myself to the naked merits of the question.
(1854, At Peoria, Illinois)
- ❖ But even if we fail **to technically restore** the compromise, it is still a great point to carry a popular vote in favor of the restoration.
(1854, At Peoria, Illinois)
- ❖ But if it is a sacred right for the people of Nebraska to take and hold slaves there, it is equally their sacred right to buy them where they can buy them cheapest; and that undoubtedly will be on the coast of Africa; provided you will consent **to not hang** them for going there to buy them.
(1854, At Peoria, Illinois)
- ❖ But clearly, he is not now with us--he does not pretend to be--he does not promise **to ever be**.
(1858, A House Divided Speech)
- ❖ we can see the place in the frame exactly fitted and prepared **to yet bring** such piece in--in such a case,
(1858, A House Divided Speech)
- ❖ One party to a contract may violate it -- break it, so to speak; but does it not require all **to lawfully rescind** it?
(1861, First Inaugural Address)
- ❖ I will venture to add that to me the convention mode seems preferable, in that it allows amendments to originate with the people themselves, instead of only permitting them to take or reject propositions originated by others not especially chosen for the purpose, and which might not be precisely such as they would wish **to either accept or refuse**.
(1861, First Inaugural Address)

Object (S + V+ it + split infinitive)

- ❖ we find it impossible **to not believe** that Stephen and Franklin and Roger and James all understood one another from the beginning, and all worked upon a common plan or draft drawn up before the first lick was struck.
(1858, A House Divided Speech)

Subject (It + V + split infinitive)

- ❖ It was believed, however, that **to so abandon** that position under the circumstances would be utterly ruinous; ... (1861, July 4th Message to Congress)
- ❖ It would be very agreeable to me, **to thus meet** my old friends, at my own home; but I cannot, just now, be absent from here, so long as a visit there, would require. (1863, Public Letter to James Conkling)
- ❖ It is of the first importance **to duly consider and estimate** this ever-enduring part. (1862, Second Annual Message)
- ❖ The extensive blockade has been constantly increasing in efficiency as the Navy has expanded, yet on so long a line it has so far been impossible **to entirely suppress** illicit trade. (1863, Third Annual Message)

It is (was)... for (of)... split infinitive

- ❖ Finally, I insist, that if there is ANY THING which it is the duty of the WHOLE PEOPLE **to never entrust** to any hands but their own, that thing is the preservation and perpetuity, of their own liberties, and institutions. (1854, At Peoria, Illinois)
- ❖ Plainly enough now, it was an exactly fitted niche, for the Dred Scott decision **to afterwards come** in, and declare the perfect freedom of the people, to be just no freedom at all. (1858, A House Divided Speech)

Sentence Head (split infinitive +V)

- ❖ **To now abandon** them would be not only to relinquish a lever of power, but would also be a cruel and an astounding breach of faith. (1863, Third Annual Message)

Objective Complement

- ❖ The foregoing history may not be precisely accurate in every particular; but I am sure it is sufficiently so, for all the uses I shall attempt to make of it, and in it, we have before us, the chief material enabling us **to correctly judge** whether the repeal of the Missouri Compromise is right or wrong. (1854, At Peoria, Illinois)
- ❖ In your political contests among yourselves, each faction charges the other with sympathy with Black Republicanism; and then, to give point to the charge, defines Black Republicanism **to simply be** insurrection, blood and thunder among the slaves. (1860, Cooper Union Address)
- ❖ Must they be allowed **to finally fail** of execution, even had it been perfectly clear that by the use of the means necessary to their execution some single law, made in such extreme tenderness of the citizen's liberty that practically it relieves more of the guilty than of the innocent, should to a very limited extent be violated? (1861, July 4th Message to Congress)

Subjective Complement

- ❖ Auxiliary to all this, and working hand in hand with it, the Nebraska doctrine, or what is left of it, is to educate and mould public opinion, at least Northern public opinion, **to not care** whether slavery is voted down or voted up. (1858, A House Divided Speech)

<形容詞の用法>

NOUN + split infinitive (Objective Complement)

- ❖ Missouri would not yield the point; and congress -- that is, a majority in congress -- by repeated votes, showed a determination **to not admit** the state unless it should yield. (1852, Eulogy on Henry Clay)
- ❖ To avoid misconstruction of what I have said, I depart from my purpose not to speak of particular amendments so far as to say that, holding such a provision **to now be** implied constitutional law, I have no objection to its being made express and irrevocable. (1861, First Inaugural Address)
- ❖ Having never been States, either in substance or in name, outside of the Union, whence this magical omnipotence of "State rights", asserting a claim of power **to lawfully destroy** the Union itself? (1861, July 4th Message to Congress)

Subjective Complement

- ❖ And as it is **to so go** at all events, may we not agree that the sooner the better? (1864, Fourth Annual Message)

<副詞の用法>

Set Phrase

- ❖ "let us be more specific--let us amend the bill so as **to expressly declare** that the people of the territory may exclude slavery." (1858, A House Divided Speech)

The Others

- ❖ On Judge Douglas' motion a bill, or provision of a bill, passed the Senate **to so extend** the Missouri line. (1854, At Peoria, Illinois)
- ❖ After several failures, and great labor on the part of Mr. Clay **to so present** the question that a majority could consent to the admission, it was, by a vote, rejected, and as all seemed to think, finally. (1852, Eulogy on Henry Clay)
- ❖ While it is fortunate that this great interest is so independent in its nature as **to not have** demanded and extorted more from the Government, I respectfully ask Congress to consider whether something more cannot be given voluntarily with general advantage. (1861, First Annual Message)
- ❖ the Treasury report including a considerable sum now which had previously been reported from the Interior, sufficiently large **to greatly overreach** the sum derived from the three months now reported upon by the Interior and not by the Treasury. (1862, Second Annual Message)
- ❖ Care has been taken **to so shape** the document as to avoid embarrassments from this source. (1863, Third Annual Message)

文献

- [1] Albakry, M.A. (2005) *Style in American newspaper language: Use and usage*. Unpublish Doctoral Dissertation Thesis, Northern Arizona University.
- [2] Calle-Martín, J. and Miranda-García, A. (2009) “On the use of split infinitives in English.” In: A. Renouf and A. Kehoe(eds.) *Corpus linguistics: refinements and reassessments*. Amsterdam; New York: Rodopi: 347–364.
- [3] Crystal, D. (1984) *Who Cares About English Usage?* Harmondsworth: Penguin.
- [4] Crystal, D. (1995) *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.
- [5] Crystal, D. (2006) *The fight for English: How language pundits ate, shot, and left*. Oxford: Oxford University Press.
- [6] Curme, G.O. (1931) *A Grammar of the English language. Volume3: Syntax*. Boston; London: DC Heath.
- [7] 福本広光. (2018) 「アメリカ英語での分離不定詞の使用実態調査と考察：使用された splitter を中心に」 渡辺秀樹 (編). 『大阪大学言語文化共同研究プロジェクト 2017 レトリック、メタファー、ディスコース』大阪大学言語文化研究科: 75–87.
- [8] 後藤弘. (2003a) 「分離不定詞構造について」『現代英語の文法と語法 一実証的研究一』73–91. 東京: 英宝社.
- [9] 後藤弘. (2003b) 「否定詞による分離不定詞構造」『現代英語の文法と語法 一実証的研究一』92–114. 東京: 英宝社.
- [10] 後藤弘. (2003c) 「英米の新聞英語における分離不定詞構造」『現代英語の文法と語法 一実証的研究一』115–126. 東京: 英宝社.
- [11] Huddleston, R. and Pullum, G.K. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- [12] Jang, Y. and Choi, S. (2014) “Split Infinitives in English: A corpus-based investigation.” *Linguistic Research* 31: 53–68.
- [13] Jespersen, O. (1940) *A Modern English Grammar on Historical Principles: V*. London: George Allen&Unwin.
- [14] Johansson, S. (2015) “Thou Shalt Not Split...?: A Corpus-Based Study on Split Infinitives in American English.” (<https://www.diva-portal.org/smash/get/diva2:791006/FULLTEXT01.pdf>)
- [15] Mikulová, H. (2011) “Split infinitive–Corpus analysis.” Unpublished Master’s thesis, Palacký University in Olomouc.
- [16] Mitrasca, M. (2009) “The split infinitive in electronic corpora: Should there be a rule?” *Concordia Working Papers in Applied Linguistics*, 2: 99–131.
- [17] Perales-Escudero, M. D. (2011) “To Split or to Not Split: The Split Infinitive Past and Present.” *Journal of English Linguistics* 39 (4): 313–334.
- [18] Quirk, R., Greenbaum, S., Leech, J. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

- [19] Schibsbye, K. (1965) *A Modern English Grammar*. London: Oxford University Press.
- [20] Supakorn, P. (2012) “The English Split Infinitive: A Comparative Study of Learner Corpora”. *Language Learning* 2.4: 21–32.
- [21] 高木八尺・斎藤光 (訳). (1956) 『岩波文庫 リンカーン演説集』 東京：岩波書店.
- [22] Teten, R.L. (2003) “Evolution of the Modern Rhetorical Presidency: A Critical Response,” *Presidential Studies Quarterly*, vol.33, no.2 (June 2003) 333–346.
- [23] Visser, F. T. (1966) *An Historical Syntax of the English Language: Part2*. Leiden: Brill.

コーパス・辞書 (いずれも 2021 年 4 月 8 日最終閲覧)

- Brown, D.W. (2016) *Corpus of Presidential Speeches*. Retrieved from <http://www.thegrammarlab.com>
- Davies, M. (2021 updated) *The Corpus of Historical American English (COHA): 400 million words, 1820–2019*. (<https://www.english-corpora.org/coha/>)
- Oxford English Dictionary* (Online) (<https://www.oed-com.remote.library.osaka-u.ac.jp:8443/>)

コンピュータソフトウェア・コンコーダンス

- 今尾康裕. (2019) CasualTreeTagger (Version 1.0)
 (<https://sites.google.com/site/casualconcj/yutiriti-puroguramu/casualtreetagger>)
- . (2020) CasualConc (Version2.1.6)
 (<https://sites.google.com/site/casualconcj/>)

中国人日本語学習者と日本語母語話者における多義動詞 「切る」のカテゴリー構造比較

—心理実験により意味分析の結果を検証する—

王 鈺

大阪大学言語文化研究科

〒560-0043 豊中市待兼山町 1-8

Email: wangyu_ogyoku@yahoo.co.jp

概要 本研究では、日本語における典型的な多義動詞「切る」の意味のカテゴリー化はどのようなようになるのか、さらに、学習者と母語話者における「切る」の意味構造にはどのような違いがあるかを明らかにすることを目的とする。力動性という認知理論に基づく意味分析に加え、中国人日本語学習者と日本語母語話者を対象に心理実験を行い、その質的分析の結果を統計的アプローチから検証した。その結果、意味分析では、「切る」のスーパー・スキーマを抽出し、各意味を4つの意味パターンに再分類した。心理実験では、4つのパターンの妥当性を検証し、さらに検討すべき点を抽出した。また、母語話者と学習者の持つ「切る」の意味構造が異なることが明らかになった。具体的には、母語話者は目的語項の特徴によるまとまったカテゴリー構造を持つのに対し、学習者は、動詞と目的語項の関係を十分に認識せず、本動詞と複合動詞の意味の関連性を捉えにくく、カテゴリーが構造化していない傾向が示された。

キーワード 多義性, 力動性, 分断動詞, カテゴリー化

A comparative study of categorical structures of the polysemous verb *kiru* “cut” between Chinese learners and native speakers of Japanese:

A psychological experiment to verify the semantic analysis

Yu Wang

Graduate School of Language and Culture, Osaka University

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract The purpose of this study is to investigate how meanings of the typical Japanese polysemous verb *kiru* “cut” are categorized. The study also aims to claim how differently the semantics of *kiru* is perceived between learners and native speakers of the Japanese language. In addition to semantic analysis based on the cognitive theory of force dynamics, I conducted a psychological experiment with native speakers and Chinese learners of Japanese in order to verify the findings from the qualitative analysis through a statistical approach. As a result of semantic analysis, the super-schema of *kiru* was extracted with each meaning reclassified into four semantic patterns. In the psychological experiment, the validity of the four patterns was verified, suggesting a need for further investigation. The experiment revealed that native

speakers and learners have different semantics perception of the verb *kiru*. While native speakers tended to show a relatively more coherent set of semantic categories based on the features of the object argument, learners did not seem to fully recognize the relationship between the verb and the object argument. Learners had difficulty capturing the relationship between the meanings of the main verb and the compound verb with learners' categorization less likely to be structured.

Keywords polysemy, force dynamics, verb of cutting/severing, categorization

1. はじめに

外国語教育において、多義語の習得は非常に難しいとされている。現代日本語における数多くの多義動詞では、切断・分離の意味（以下、分断動詞）を表す「切る」は日本語教育の早い段階で扱われる基本動詞であり、学習者にとって主に二つの習得上の困難がある。

1 点目は、複数の意味をもつ「切る」の各意味とそのつながりに関することである。例えば、「リンゴを切る」は鋭利なナイフで対象をいくつかの部分に分けることである。「トランプを切る」は整理されたトランプを一続きになっているものとして捉え、手で対象を混合するという意味を表す。また、「10 秒を切る」は測定値が基準値である 10 秒を突破することを表し、動作主と道具格が明確ではなく、意味がより抽象的になる。母語話者はこれらの語義をひとつのカテゴリー構造として捉えられるのに対し、中上級学習者にとってさえも、各語義とその間のつながりを自然に理解することは至難の業である。

2 点目は、「切る」との共起制限である。現実世界においては、固体以外のものを切ることは物理的には可能ではないと考えられる。それゆえ、「切る」に関する動作の許容範囲は固体のみと考えられ、固体である場合は「切る」と共起でき、液体や気体である場合は「切る」と共起できないと判断する傾向が見られる。しかし、「泥を切る」、「砂糖を切る」のように、固体であっても「切る」と共起しにくい例が存在し、「野菜の水気を切る」、「飛行機は雲を切って飛ぶ」のように、液体と気体であっても「切る」と共起できる例も少なくない。そのため、「切る」の性質のみで容認度を判断するのは不適切であると考ええる。

そこで、本研究では、分断動詞「切る」を取り上げ、認知言語学の観点からその意味・機能を分析し、学習者と母語話者の持つ「切る」のカテゴリー構造はそれぞれどのようなものかを検討していく。

2. 先行研究

2.1. 多義動詞「切る」の意味分類研究

認知言語学の立場では、多義語はプロトタイプである中心義が何らかの動機づけに基づき語義が拡張し、ひとつの放射状カテゴリーが形成されると考える (Lakoff 1987: 91-114)。多様な意味・機能を持つ多義語の各意味をどのようにカテゴリーと把握するのは重要な問題である。以下、「切る」の意味カテゴリー化と意味分類に関する研究を概観する。

吉村 (2003) は「切る」のプロトタイプの意味を「連結・結合している状態のものを分かつ」

と定義すると同時に、「切る」のスキーマ的意味をほぼ同等に扱い「連結の分断」であると捉えた。また、連続体のどこで分断動作を行うかという点に注目し意味分類を行った。その結果、＜最初（「スタートを切る」など）＞、＜途中（「手を切る」など）＞、＜最後（「押し切る」など）＞という三つのグループが分かれ、切れ目の位置によってメタファーによる意味拡張が生じると指摘した。

許（2008）では、メタファーに加え、文脈的要素と百科事典的知識が語義間の関連性に関わりと指摘し、＜別義＞、＜文脈的別義＞、＜百科事典的意味＞の認定を行った。また、一部の語義を類義語である「割る」、「批判する」、「下回る」と比較した結果、「切る」にはこれらの類義語との意味範囲の違いが見られた。

森山（2012）は動詞のコロケーションを重視し、「切る」の中心義を「〔（意志を持つ）人〕が・〔刃物などの道具〕で・〔一続き物〕を・〔（力を加えて）分断する〕」としている。また、森山（2015）は心理的な手法から知見を得て、内省分析による13個の語義を意味拡張の動機づけにより、＜中心義とメトニミー拡張群＞、＜中心義からのメタファー拡張群＞、＜メトニミー拡張義からのメタファー拡張群＞にまとめ、各拡張義と中心義の間の意味拡張経路を明らかにした。

栗田（2018）では視覚スキーマを用い、＜切り分け＞、＜切り取り＞、＜切り捨て＞という動作の三段階による分類を行った。それに加え、複合動詞である「V1+切る」と本動詞の「切る」の意味を同一の意味構造に捉え、複合動詞と本動詞の意味の有契性を示した。

しかしながら、上述の「切る」に関する意味分類の研究は主に各研究者の理論や使用経験に基づいた内省分析のものであるため、意味記述と意味分類などでは一致しないことも多く、プロトタイプとスキーマ的概念との関係の解明には至っていない。これらのみの結果では信頼性と妥当性に疑問が残っていると考えられる。

2.2. 多義動詞の習得研究

第二言語習得研究（SLA）では、従来から語彙習得は文法習得に比べ遅れる状況にある（長友1999:12）。また、実際の海外の日本語教育現場では、授業時間が限られるため、語彙習得は文法習得より軽視される傾向があり、語彙習得は学習者による独学の場合が多い。その際、外国語環境における学習者は教科書の単語リストや辞典で未知の単語を調べている。しかし、教科書や各辞典は、語義間の関連性を明記せず、単に各語義を並べて記述しているものが多くなっている。実際の使用場面を持っていない学習者は丸暗記のみで各語義を習得し、記憶の負担が非常に重くなるだけでなく、長い時間使用しないと忘れてしまう。それゆえ、多様な用法を持つ語の意味を各々習得するより、カテゴリーとして理解させる必要がある（今井1993:243-253）。これらを念頭に、本節では認知言語学の観点からの多義語の習得研究について概観する。

Kellerman（1979:37-57）では、プロトタイプ理論に基づき、オランダ語を母語とする英語学習者にL1であるオランダ語 *breken* をL2である英語 *break* への翻訳可能性判断テストを行った。その結果、L1のプロトタイプ性が高い語義はL2への翻訳がしやすくなる傾向が見られた。また、鐘（2016）はKellerman（1979）を踏まえ、中国人日本語学習者を対象とし、受容性判断調査によ

り、「切る」の各語義においてプロトタイプ性が高い語義のほうが学習者に受容されやすいことを解明した。

今井（1993: 245-253）は多義語の内部的な意味構造と学習者のもつ意味表象に着目し、日本語母語話者に wear の用例を意味の類似性に基づいてグループ分けするように指示した。上位のカテゴリでは、学習者と母語話者は大まかに同様の概念でグループ化しているが、より下位のカテゴリでは、母語話者のもつ意味表象は安定した内部構造を持っているのに対し、学習者の意味表象はまとまりがない構造を持っている。ひとつの原因として、学習者は類似性に基づくメタファーなどの動機づけを理解しにくく、プロトタイプ語義から展開した放射状カテゴリを十分に捉えられないと指摘した。また、松田・黄（2011: 41-57）でも個々の語義間の類似性・近接性による意味拡張を説明するプロトタイプ・アプローチは学習者にとって必ずしも認識しやすすくないことを指摘した。何を類似性・近接性と考えるかは母語話者と学習者の言語文化的背景によって異なる可能性があり、学習者は母語話者が認識できる類似性・近接性などを理解しにくい場合が存在する。このため、松田・黄（2011）は異文化理解を射程に入れた語彙指導を提案し、プロトタイプ・アプローチを採用せず、日本語の多義動詞「とる」と中国語の多義動詞「取（qu）」を例に複数の意味・概念をひとつの包括概念に取りまとめたコア理論の知見を示した。

2.3. 先行研究の問題点

以上で概観したように、従来から、多義語の意味と習得に関する先行研究は多くなされている。様々な観点から「切る」の意味・機能、意味展開のメカニズムおよび学習者の多義語習得の状況を考察してきた。ただ、意味記述と意味分類では、解明されていない点が残っている。「切る」の核となる概念および「切る」の動作主、被動作主、力の作用方式に関する制約は十分に検討されていない。全体的な意味の広がりを見野に入れた意味記述を追求するため、新たな意味体系を構築する必要がある。また、認知言語学からの考察は主に研究者の少数例による内省的なものであり、質的分析の上で、統計的なアプローチから検証する研究は必ずしも多くない。認知理論に基づいた意味分析は量的な観点から裏付けられるか、学習者は母語話者のもつカテゴリ構造とどう異なるかはまだ検討の余地が残されている。次節からはこれらの課題を解決するための理論的枠組みを説明する。

3. 理論的枠組み

3.1. 力動性

山梨（2000）は言語の形式と意味は、外部世界の知覚、経験を基盤とする様々なイメージ・スキーマによって動機づけられていると指摘した。また、図1のように、言語の概念体系の形成は、普遍的な「身体的基盤」と固有的な「社会・文化的視点」に関わり、具体的経験領域、イメージ、イメージ・スキーマという三つの段階を経て、抽象度が上がると述べた。「切る」という動詞の背景には、力という要素が無視できないため、力のスキーマに関する重要な認知理論のひとつを紹介する。

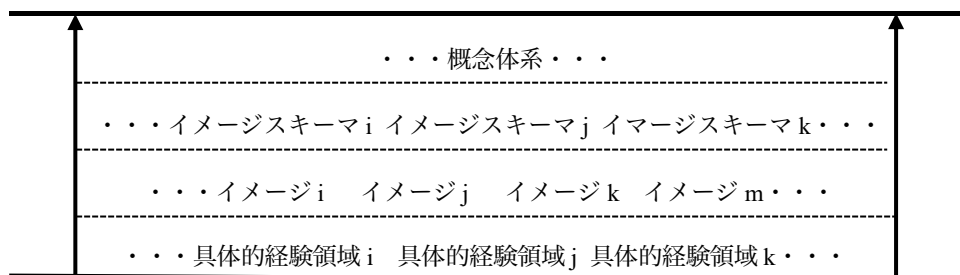


図1 概念体系の普遍性と固有性 (山梨 2000: 157)

Talmy (1985) では、力動性 (Force Dynamics) という認知理論を提唱している。力動性とは、力という観点から見た個体の相互作用のことである (松本 2003: 60)。力の構造においては、二つの対立した力の参加者の力のバランス (強・弱) によって相互作用の結果状態が異なると捉えている。この力構造の中で、本来的に活動や静止の傾向をもつ存在は主動体 (agonist) と呼ばれ、主動体に対抗する力を加える存在は対抗体 (antagonist) と呼ばれる。図2で力動性の図式を示す。図式aと図式bは力動性の基本的な通常状態のパターンとして挙げられる。例文(1)を参照されたい。目的語項 ((1a) の「空気」、(1b) の「丸太」) である主動体は活動・静止の傾向を持つが、主語項 ((1a) の「扇風機」、(1b) の「留め金」) である対抗体から継続的に強い力を加えられると、力が行使された期間中に本来の内在傾向性と逆の傾向が生じた状態にある。すなわち、このような事象では、使役と使役の結果が同時に生じ、拡張使役 (extent causation) と呼ばれる。一方、図式cと図式dのように対抗体 ((1c) の「ピストン」、(1d) の「遮断バルブ」) から主動体 ((1c) の「油」、(1d) の「ガソリン」) に力を行使しない状況から力を加える状況になることを通し、力が行使された期間中に主動体の状態変化が見られる事象がある。このような使役における変化のプロセスが見られた事象は開始時使役 (onset causation) と呼ばれる。

Talmy (2000) によると、分断・破壊を表す事象は、対象の状態変化を含み、その変化の原因・様態と変化のプロセスとして構成される。それゆえ、「切る」が表す事象は開始時使役のパターン (図式c, d) にあてはまると言える。そこで、開始時使役のパターンの力構造に基づき、「切る」の核となる概念と意味・機能を考察する。

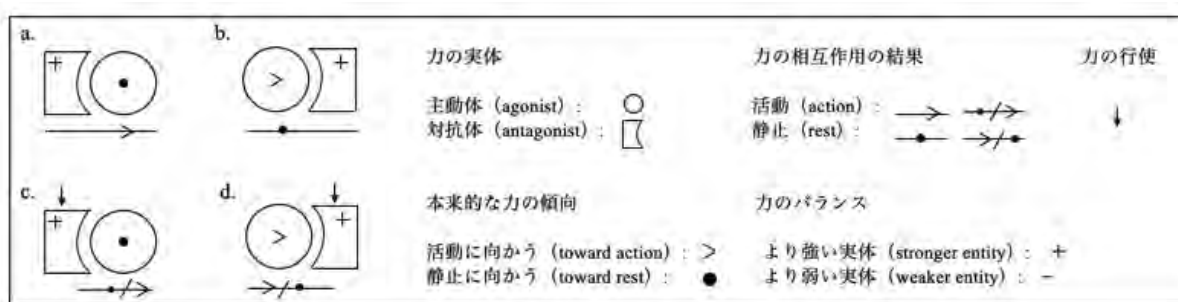


図2 力動性のパターンの図式

- (1) a. The fan kept the air moving.
(扇風機が空気を動かし続けた)
- b. The brace kept the logs from rolling down.
(留め金が丸太が転がり落ちるのを防いだ)
- c. The piston made the oil flow from the tank.
(ピストンが油をタンクから流れ出させた)
- d. The shutoff valve stopped the gas from flowing out.
(遮断バルブがガソリンが流れ出すのを止めた)
- (松本 2003: 60-65)

3.2. 概念メタファー

Lakoff & Johnson (1980) によると、メタファーはレトリックの言語表現だけではなく、人間の思考や行動などの日常の営みのあらゆるところに浸透しており、概念体系の大部分がメタファーによって成り立ち、日常の活動の仕方に構造を与えている。概念メタファー (Conceptual Metaphor) は具体的なメタファー表現と区別され、2つの概念領域間の構造的な対応関係に関わる。すなわち、抽象的なドメインを具体的なドメインで理解する認知的な仕組みにあてはまる概念レベルのメタファーは概念メタファーと呼ばれ、その概念メタファーを言語的に具体化したものはメタファー表現と呼ばれる。

Lakoff et al. の MetaNet team (2012-) により作成された概念メタファーのデータベース MetaNet Metaphor¹ によると、< CAUSES ARE FORCES > と < IMPEDIMENTS TO ACTION ARE ANTAGONISTIC FORCES > という概念メタファーが存在する。本研究はこれらのメタファーに従い、「切る」に関する事象において、< 行為は力 > という概念メタファーを提起し、< 先行する行為・既存する状態は主動体による力 > と < 行為に関する妨害・状態に対する破壊は対抗体による力 > をその概念メタファーのサブメタファーとして位置付ける。

4. リサーチデザイン

4.1. 目的と研究設問

先行研究と残されていた課題を踏まえ、本研究の目的は、日本語の多義動詞「切る」の各語義と意味構造を分析することで、「切る」の核となる概念と意味構造、目的語項と主語項および力の作用方式に関する制約、各意味間のつながりを解明すること、また、その認知理論に基づく意味分析の妥当性を心理実験によって検証し、学習者と母語話者における意味構造の違いを明らかにする点にある。そのため、本研究は以下の三つの課題を設定した。

¹ データベースの最終修正日は2016年8月7日である。

- RQ1 「切る」の核となる包括的概念と意味構造はどのようなになるのか。
- RQ2 意味分析の結果と心理実験による結果に違いがあるのか。
- RQ3 中国人日本語学習者と日本語母語話者の持つ「切る」の意味構造に違いがあるのか。

4.2. データ

4.2.1. 意味分析用のコーパスデータ

田中（1990）はコア理論を提唱している。コア理論に基づくことばの意味空間は、図3に示すように、様々なレベルの概念が集積している立体構造とみなされている。この意味空間において、多義語の複数の用法のそれぞれの意味・概念を一つに取りまとめた包括概念は「コア」であり（松田・黄 2011:44），Langacker（1987）によるスーパー・スキーマと一致する。本研究は、現代日本語書き言葉均衡コーパス（BCCWJ）に基づき、各辞書や各種の先行研究で言及された意味分類基準と語義項目を幅広く取り込み、表1のように、意味空間の概念レベル1にあてはまる「切る」の用例を抽出し、中心義を含み15個の語義として整理した。これらの用例を使用し、意味分析でそれぞれの語義に関する意味記述の詳細を説明し、意味分類とカテゴリー化を行った。

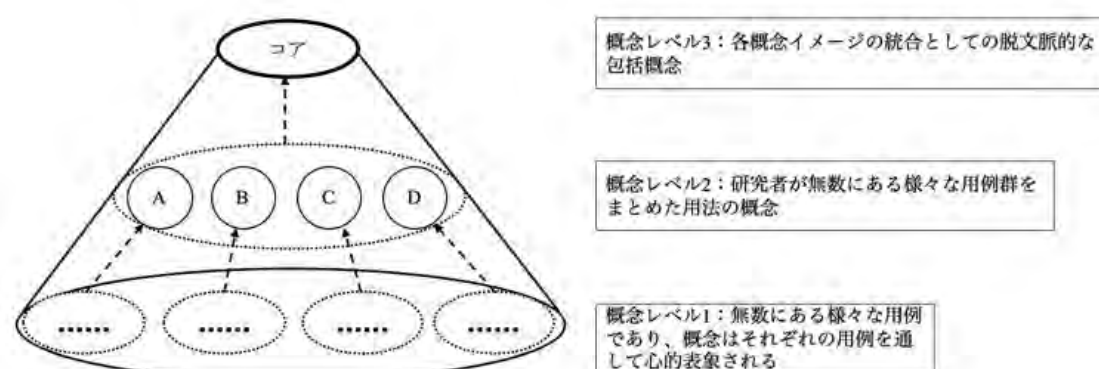


図3 コア理論に基づく意味空間（松田・黄 2011: 44）

表1 「切る」の意味空間における用例

語義	コーパスデータ
0. 分断義	(2) 熱湯に包丁をつけ、かるく水気を取ってから、一気に [ケーキ] ² を切る。
1. 開封義	(3) ホテルの部屋に帰って、車のなかで渡された [封筒の封] を切ると、なかから金字を刻印したダーク・ブルーの東独のパスポートが出てきた。
2. 切除義	(4) H2 ブロッカーの潰瘍治癒率は、八十%と効率で、潰瘍で [胃] を切る人が激減したことが確かです。
3. 切取義	(5) 依頼主に [領収書] を切る時は、印刷代+デザイン料・雑費を含めた金額を書いたらいいのでしょうか。

² 「切る」の目的語項を全て [] によって表記する。複合動詞の前項動詞を目的語項と同等に扱う。

4. 殺傷義	(6) 高橋与惣右衛門は敵兩人と切結び、猪太夫の眼前に於いて「敵一人」を切る。
5. 創造義	(7) そこで床を高くして床面に「炉」を切ること、つまり高床式住居にすることを考案した。
6. 混合義	(8) こんな遊びを日本でもやるのかなと彼は言いながら、「トランプ」を切ってテーブルに伏せて置き、上から一枚とって、表を見せた。
7. 取捨義	(9) 豆腐はキッチンペーパーに包み、電子レンジで三十秒加熱して「水気」を切る。
8. 中断義	(10) 就寝時や外出時には「電気」を切る。
9. 批判義	(11) 流麗な文章で、痛烈に当時の「世相」を切っていた。
10. 横断義	(12) 今は南からの黒潮に乗ってスナメリたちが「波」を切る。
11. 回転義	(13) バスが急な「ハンドル」を切るたびに、私の体は激しく左右に引張られた。
12. 突破義	(14) ようやく、と言ったが、パイフウの抜き撃ちは平均で「コンマ一秒」を切る。 (15) 日本語能力試験まで残り「一ヶ月」を切りました。
13. 完遂義	(16) みんなで声を出し合い、最後まで全力で「走り」切ると抱負を力強く話しました。
14. 極度義	(17) 「疲れ」切るまで暗い海を泳いで、それで神さまのところへ行けるならいいじゃないのって、自分に言い聞かせたわ。

4.2.2. 心理実験のデータ

心理実験では、日本語母語話者と中国人日本語学習者各30人を対象にカード分類法による類似性判断テストを行い、「切る」の語義各2つの用例で類似性判断テストのカードリストを作成した（表2を参照）。意味分析との一致性を高めるため、各語義の用例aは意味分析で用いた用例を使用した。ただし、テストの難易度と時間も実験の精度に影響を与えるため、被験者の負担にならないように、動詞の目的語項に重点に置き、用例のコロケーションのみを使った。

表2 類似性判断テストのカードリスト

語義	用例	中国語訳
0. 分断義	a. ケーキを切る。	切
	b. 紙を切る。	剪
1. 開封義	a. 封筒の封を切る。	打开, 剪开
	b. 袋を切る。	打开
2. 切除義	a. 胃を切る。	切除
	b. 爪を切る。	剪
3. 切取義	a. 領収書を切る。	发行
	b. 小切手を切る。	发行
4. 殺傷義	a. 敵を切る。	杀
	b. 手を切った。	切 (到)
5. 創造義	a. 囲炉裏を切る。	修筑
	b. 型紙を切る。	剪出
6. 混合義	a. トランプを切る。	洗
	b. 麻雀を切る。	洗

7. 取捨義	a. 野菜の水気を切る。	去除, 甩干
	b. 部門を切る。	去除
8. 中断義	a. 電気を切る。	关
	b. 縁を切る。	断绝
9. 批判義	a. 世相を切る。	批判
	b. 政治を切る。	批判
10. 横断義	a. 波を切る。	冲破
	b. 雲を切って飛ぶ。	划破
11. 回転義	a. ハンドルを切る。	使...转弯
	b. 球を切る。	使...转弯
12. 突破義	a. 100メートル走で10秒を切った。	打破
	b. 一ヶ月を切る。	低于
13. 完遂義	a. マラソン大会で最後まで走り切る。	V1+完
	b. 本を読み切った。	V1+完
14. 極度義	a. 彼は心身ともに疲れ切った。	V1+極(了)
	b. 部屋は冷え切った。	V1+極(了)

4.3. 分析手順

BCCWJ コーパスデータから抽出した用例を使用し、力動性の力構造に基づき、「切る」の各語義に関する意味分類を行い、語義間の有契性と共通した包括概念（スーパー・スキーマ³）を探った（RQ1）。意味分析の妥当性を量的な観点から検証するため、質的分析に加え、日本語母語話者と中国人学習者各 30 人を被験者とし、それぞれのグループに心理実験を行った。上記の「切る」に関する用例が書いてあるカードを提示し、「切る」の意味が似ているものをグループ（グループの数、および各グループ内のカードの数は自由とした）に分けるように指示することで類似性判断テストを行った。得られた結果で R によるクラスター分析を行った。その後、意味分析による意味構造が心理実験の結果と一致しないところを再検討し（RQ2）、母語話者と学習者のカテゴリー構造では違いがどのようになるかを考察した（RQ3）。

5. 結果と考察

5.1. 意味分析の結果

5.1.1. 力動性による再分類

多義語は複数の語義を持ち、文字通りの具体的な語義が中心となり、そこから意味拡張が起こる。この中心義には、①文字通りの意義であること、②具象的な意義であること、③なるべく他の多くの意義の前提になることという 3 つの判断基準がある（瀬戸他 2017: 74）。この基準に従い、本研究は特に動詞とその目的語項のコロケーションに注目するのに加え、「切る」の力の作用方式も考慮するため、先行研究と少し異なった「[鋭利な刃物などの道具]で・[集中した力]

³ 用語を統一するため、本論文は包括的概念を表すとき、「スーパー・スキーマ」を採用する。

で・「一続きもの」を分断する」を「切る」の中心義に据える。また、「切る」の対象が本来的にある特定の状態にあるために、一種の内在傾向性が捉えられる。それゆえ、「切る」の目的語項を主動体とし、目的語項に動作の影響を与える主語項を対抗体としている。このような力動性の力構造に基づき、主動体の状態変化に着目し、コーパスデータから整理された表 1 の 15 個の語義を以下の 4 つの意味パターンに分類した。

【パターン 1 外在連続性】：力によって外在連続性を破壊し移動する。

「外在連続性」とは、「切る」の対象は外在的に一続きになっている性質を持ち、安定した状態にあり、ある機能を果たすに足る「全体」としてのまとまり、統合性を持っていることである。一旦その外在連続性を破壊すると、対象の外観に加え、状態と機能にも変化が生じる。特に本来の機能が喪失されるのがそのパターンの特徴である。例えば、(18) のように全体に相当する「ケーキ」を幾つかの部分に分けることで、「ケーキ」の本来の連続体としての状態を失い、機能もある程度変わるようになる。最後に全体の安定した静止状態から各部分が自由に移動できる状態になる。しかし、(19) の「泥」と「砂糖」は同様に分断動作を行っても、分離された対象はまだつながっている全体として「泥」と「砂糖」と認識でき、ただ分断された前の量と比べ少なくなるだけである。外在状態と機能にもほぼ変化が生じていないため、「切る」と共起できない。

- (18) 熱湯に包丁をつけ、かるく水気を取ってから、一気に[ケーキ]を切る。(表 2-(2)再掲)
(19) *[泥]を切る。/*[砂糖]を切る。(作例)

また、(20)、(21)、(22)、(23)、(24)では、「切る」の意味はプロトタイプである<分断義>と異なり、分断動作+ α になる。これらの拡張義では、プロトタイプの分断動作に加え、開ける(開封義)、除去する(切除義)、利用する(切取義)、殺傷する(殺傷義)、創造する(創造義)ことなどの後続する出来事があるため、中心義からメトニミーによって拡張される。また、封は手紙の一部分であり、患部となる胃は人間の身体部位であり、発行された領収書は全体から取られた一部分であるように、「切る」の対象はあるもの全体の部分であるとみなされるため、本来にある外在連続性(部分と全体の連続性)が捉えられる。その外在連続性により、対象に含まれた安定した静止する傾向にある性質がプロトタイプの<分断義>と共通するため、同様に外在連続性のパターンに分類した。図 4 でパターン 1 の力動性の構造を示す。

- (20) ホテルの部屋に帰って、車のなかで渡された[封筒の封]を切ると、なかから金字を刻印したダーク・ブルーの東独のパスポートが出てきた。(表 2-(3)再掲)
(21) H2 ブロッカーの潰瘍治癒率は、八十%と効率で、潰瘍で[胃]を切る人が激減したことが確かです。(表 2-(4)再掲)
(22) 依頼主に[領収書]を切る時は、印刷代+デザイン料・雑費を含めた金額を書いたらいいいのでしょうか。(表 2-(5)再掲)

(23) 高橋与惣右衛門は敵兩人と切結び、猪太夫の眼前に於いて「敵一人」を切る。

(表 2-(6)再掲)

(24) そこで床を高くして床面に「炉」を切ること、つまり高床式住居にすることを考案した。

(表 2-(7)再掲)



図4 外在連続性の力構造パターン

【パターン2 内在秩序性】：力によって内在秩序性を破壊し移動する。

「内在秩序性」とは、「切る」の対象には、外在的に見えない体系性や秩序性がある一面が存在することである。(25)の「トランプ」は本質的に紙であるが、「紙を切る」のようなプロトタイプ的<分断義>と異なり、実際に分断動作を伴わず、対象の外在連続性を破壊することは含意しない。そのため、<混合義>の対象は<分断義>の対象のように一続きものに見立てられるが、主動体の内在傾向性の体现の仕方が異なる。トランプの用途と機能を考えると、実際にトランプは普通の紙ではなく、人間が意図的に数値やマークなどを付け、体系を設けるカードである。そのため、トランプには一種の内在秩序性が存在している。その内在秩序性により、トランプは本来的に静止する傾向にあると捉えられ、対抗体からの力でその安定した秩序性を壊すことによって、対象が混沌とした活動状態になるという状態変化が想定できる。

(25) こんな遊びを日本でもやるのかなと彼は言いながら、「トランプ」を切ってテーブルに伏せて置き、上から一枚とって、表を見せた。

(表 2-(8)再掲)

また、(26)のように、「切る」の対象「水気」は上記用例の対象である固体と異なり、明らかに液体であるが、「切る」と共起できる。この語義は、分断するように対象に「分離させる動作を行った→対象が本来の物体内から分離された→分離された対象を取り去った」という連続的なスクリプトを表し、分離動作が実行された後の取捨が問題となるため、<取捨義>と名付ける。一方、(27)の対象「電気」は物理的なものというより、目に見えずかなり抽象的なものである。状態が連続していて、ある機能を果たすものを分断することを通しその機能を中断する事象を表す。その物理的なものを対象とする<取捨義>と、より抽象的なものを対象とする<中断義>は共に、(21)「胃を切る」のような<切除義>から、実際に患部を切るように、不要なものをはらい捨てる意味を含意し、メタファーによって展開されると考えている。また、(28)の「世相」

は同様に抽象的なものであるが、表された事象と拡張経路が＜中断義＞と異なり、刃物で人間の肉体に傷害を与える＜殺傷義＞から、鋭利な言葉や文章などで対象に抽象的な傷害を与える＜批判義＞に拡張する。それゆえ、これらの語義は、パターン1の中心義からのメトニミー拡張群と比べ、プロトタイプからより遠く離れている。

また、液体と抽象的なものはともにパターン1で述べた外在連続性が捉えにくい状態にある。(26)の水気は野菜の内部に含まれた成分であり、(27)の電源と(28)の「世相」自体には一定の体系や組織で構成され、安定した秩序性を構成するとみなすことが可能である。そのため、＜混合義＞と同様に、主動体の内在傾向性が内在秩序性で示される。これらの語義は同様にパターン2の内在秩序性にあてはまるが、さらに状態変化の経路の側面から考えると、語義の間に微妙な差異が存在する。他の語義における＜静止→活動＞という形の状態変化と違い、＜中断義＞の対象である電流や人間関係などは、内部組織が互いに行き交う形で特定の機能を実現するため、本来的に静止状態ではなく、流動する傾向にあると捉えられる。その流動的な電流や人間関係などに対抗体からの強い阻止力を適用すると、主動体が活動状態から静止状態に変化する。それゆえ、＜中断義＞はパターン2において、より周辺的な位置を占める。パターン2の力構造を以下の図5に示す。

(26) 豆腐はキッチンペーパーに包み、電子レンジで三十秒加熱して〔水気〕を切る。

(表 2-(9)再掲)

(27) 就寝時や外出時には〔電気〕を切る。

(表 2-(10)再掲)

(28) 流麗な文章で、痛烈に当時の〔世相〕をきっていた。

(表 2-(11)再掲)

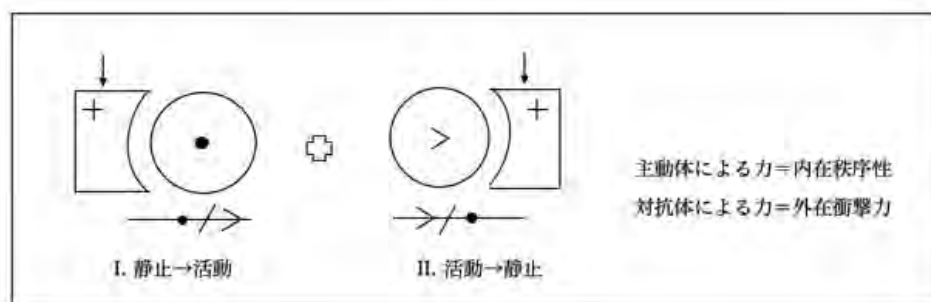


図5 内在秩序性の力構造パターン

【パターン3 物理抵抗性】：力によって物理抵抗性を克服し移動する。

「物理抵抗性」とは、「切る」の対象には動作への抵抗力が存在することである。(30)と(31)の「波」や「雲」は(19)の「泥」と「砂糖」などと同様に、分断してもその外在的な性質を壊さず、外在連続性が捉えにくい、「切る」との共起が容認される。一般的に、その＜横断義＞は中心義からメタファーによる展開で説明されるが、メタファーのみで解釈すると、「??水を

切る」と「波を切る」，「? 空気を切る」と「雲を切る」の容認度の違いは解釈しにくくなってしまう。したがって，「切る」の対象が液体と気体である場合，対象の性質に関する何らかの制限があると想定できる。液体や気体の物質は固体とは異なり，自由に流動できるため，密度がある程度濃い場合，対象自体には物理的な抵抗力が見られる。すなわち，「波」や「雲」などには濃い密度があるため，外からの力に対し，ある物理抵抗性が存在している。より集中的な強い力を対象に加え，主動体である「波」や「雲」が静止状態から活動状態に変わる。そのため，「波」，「雲」，「霧」などは許容されるが，濃い密度が含意されない（すなわち，主動体による力が捉えにくい）水や空気などは特定の文脈がないと，「切る」と共起しにくいと想定できる。

一方，(32)のように，文脈によって「空を切る」は容認される場合は実際に少なくない。これらの例文では，ある勢いとスピード感を伴っている点で共通している。力の強さの側面から考えると，ゆっくり分断動作を行うことと比べ，勢いとスピード感は道具である鋭利な刃物と同様に動作の手段と様態を表し，力の強さと衝撃を際立たせる。前述の通り，パターン3における「切る」の対象はある物理抵抗性を持っている。その抵抗力が存在するからこそ，分断目的を実現するための対抗体からの力は普通の力ではなく，主動体のもつ抵抗力ほどの集中された力を行使する必要がある。そのため，「空気」に関する用例に含意された勢いとスピード感は同様に，「切る」の対象が通常ある程度の抵抗力を持つことを裏付けることができる。

- (29) ?? [水] を切る。 / ? [空気] を切る。 (作例)
- (30) 今は南からの黒潮に乗ってスナメリたちが [波] を切る。 (表 2-(12)再掲)
- (31) と同時に稲妻が鋭く [雲] を切ってはしった。 (BCCWJ)
- (32) シュンっと [空] を切る音を立てて小石は飛び，見事，物の怪の後頭部に命中した。 (BCCWJ)

また，進行方向を変えたり，回転を与えたりする<回転義>は<横断義>の力関係と類似する。物体は同じ速度と方向を保ち続けようとする物理的な慣性を持つため，外力の作用に対し，物理的抵抗力があると想定できる。そのため，(33)のハンドルに加え，運転中の舵，球などは「切る」との共起が許容される。同じ速度や方向で運転するのは相対静止の状態にあたり，連続的な強い力によって相対静止の状態から活動した状態になると捉えられる。また，<横断義>と<回転義>の間に，「気体を勢いよく「切る」のような動作を加える→回転を与えたり進行方向を変えたりする」という連続的なスクリプトが想定できる。すなわち，「横断義」の「空気を勢いよく押し分ける」というプロセスで「回転義」の「回転を与えたり，進行方向を変えたりする」という結果を表し，メトニミーによる展開というパターン3内の語義間のつながりが捉えられる。図6でパターン3の力構造を示す。

- (33) バスが急 [ハンドル] を切るたびに，私の体は激しく左右に引張られた。 (表 2-(13)再掲)

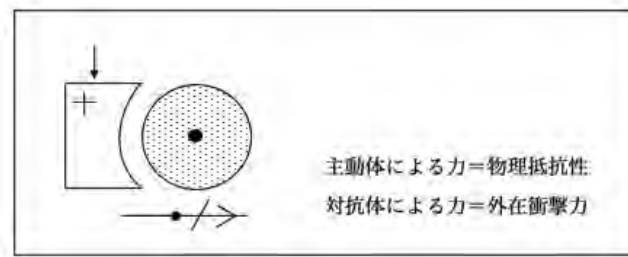


図6 物理抵抗性の力構造パターン

【パターン4 心理傾向性】：力によって心理傾向性を突破・違反し移動する。

「心理傾向性」では、人間が心理的に、行為の達成に対して強い抵抗力を持つと認識・予想する事柄が「切る」の対象となることを表す。このパターンには、二つの下位分類が存在している。タイプ1は事柄の達成を人間が期待しているが、意識的にこれに達成するのはあまり簡単ではなく、何らかの困難を伴うと認識する心理的傾向性が存在するものである。それに対し、タイプ2はその事柄達成を人間が期待していないものである。

(34)のように、「切る」の対象が数値である場合、基準値より下回るという意味を表している。数値が時間とともに起点から継続的に減少するとき、その減少経路がグラフ上で基準値を示す線を切って測定値になる。(30)の対抗体「スナメリ」が主動体「波」を切るように前方に進むことと移動様態で類似しており、＜横断義＞からメタファーの展開であると捉えられている。しかし、「数値を切る」は対象の基準値より下回るという意味のみではなく、人間の心理的傾向性が含意されている。例えば、(34)の用例には、「100メートル走で9秒99」の成果を達成する前に、「10秒」という壁を人間が意識している。したがって、「10秒を切った」は「10秒を下回った」と比べ、行為を実現したとき、目的達成というプラスのニュアンスを伴う。また、この＜突破義＞を表す場合、(35)の「ようやく」などの副詞と共に起する用例も多い。すなわち、「切る」の対象には、達成しようとする行為に何らかの妨害が存在し、その妨害はパターン3における物理的な妨害ではなく、人間が心理上・認識上に予想した妨害である。

一方、(36)「時間を切る」が表す事柄は人間にとって望ましくない心理的傾向性が見られる。この事象には、動詞句に表された事柄が起きてほしいが、起きにくいと思うというような心的葛藤が存在せず、心理的な側面は単純であるため、タイプ2のほうに合致している。その心理傾向性は「期限が切れる」という事象に対する抵抗性と理解できるので、事象達成と事象達成への心理的妨害という対立関係が捉えられる。このような心理的妨害は複合動詞の意味にも見られるため、本動詞と複合動詞の語義間のつながりも窺わせる。以下この側面から本動詞と複合動詞の有契性を見ていく。

(34) 100メートル走で[10秒]を切った。 (作例)

(35) ようやく、と言ったが、パイフウの抜き撃ちは平均で[コンマ一秒]を切る。

(表2-(14)再掲)

(36) 日本語能力試験まで残り「一ヶ月」を切りました。 (表 2-(15)再掲)

「切る」は複合動詞である場合、語彙的複合動詞と統語的複合動詞、両方とも存在している。語彙的複合動詞の場合はほぼ本動詞「切る」の中心義と同様であり、分断・終結の意味を表している(姫野 1999)。接辞的に統語的複合動詞として用いられる場合は前項動詞の性質によって、大まかに二つのグループに分けられる。前項動詞が継続動詞の場合、行為を完遂することを表すのに対し、前項動詞が瞬間動詞である場合、極度の状態に達することを表している⁴。(37)のように、＜完遂義＞では前項動詞が表す事象の完成は期待されるが、一瞬で簡単に目的を達成する事象ではなく、人間がそれを達成するまで、恐らく長い時間がかかり、または途中で困難や苦痛を伴うと認識する事象である。そのため、タイプ1にあてはまる。一方、＜極度義＞は(38)「疲れ切る」のように、前項動詞「疲れ」が表す事象は人間にとって明らかに望ましくないという心理的傾向性が見られる。タイプ2に該当する場合が多い。図7でパターン4の力構造を示す。

- (37) みんなで声を出し合い、最後まで全力で「走り」切ると抱負を力強く話しました。 (表 2-(16)再掲)
- (38) 「疲れ」切るまで暗い海を泳いで、それで神さまのところへ行けるならいいじゃないのって、自分に言い聞かせたわ。 (表 2-(17)再掲)

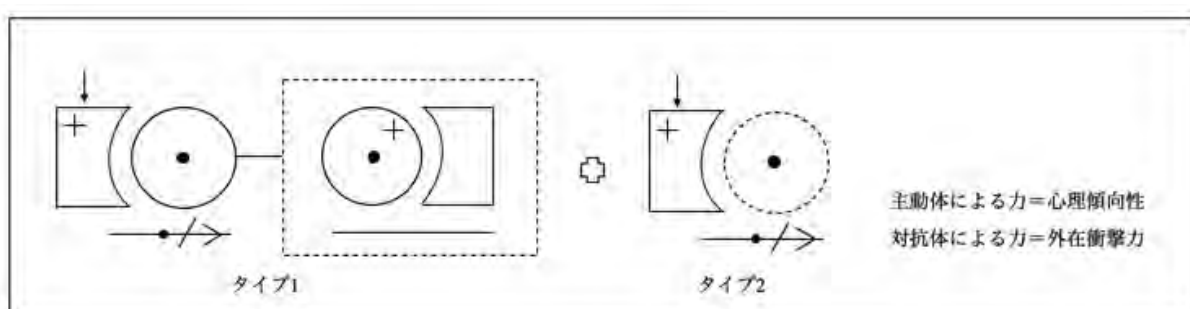


図7 心理傾向性の力構造パターン

5.1.2 「切る」のスーパー・スキーマ

5.1.1 節で力動性の構造に基づき「切る」の語義を4つの意味パターンに分類した。より詳細で背景化したものを考えると、パターン1の外在連続性とパターン2の内在秩序性には対象の安定した状態(主動体)とその状態への破壊(対抗体)が、パターン3の物理抵抗性とパターン4の心理傾向性には行為(対抗体)と行為への妨害(主動体)という対立関係が捉えられる。また、意味記述が示すように、いずれの意味パターンからも、力によって対象に含まれた内在傾向性を

⁴ 継続動詞が＜極度義＞であり、瞬間動詞が＜完遂義＞を表す場合も存在する(姫野 1999)が、本研究は主に典型的な場合を論じる。

何らかの形で破壊することを通し、対象の状態を変えるという共通の事象が抽出できる。力動性の構造を踏まえ具体的に言えば、力による使役移動（静止→活動）という状態変化を表す場合が多い。それゆえ、「切る」のスーパー・スキーマを「力による使役移動という状態変化」とする。

5.1.3 「切る」の意味構造

以上の考察により、力動性に基づく「切る」の4つの意味パターンと語義間の関連性を整理すると、図8のような意味構造になっている。

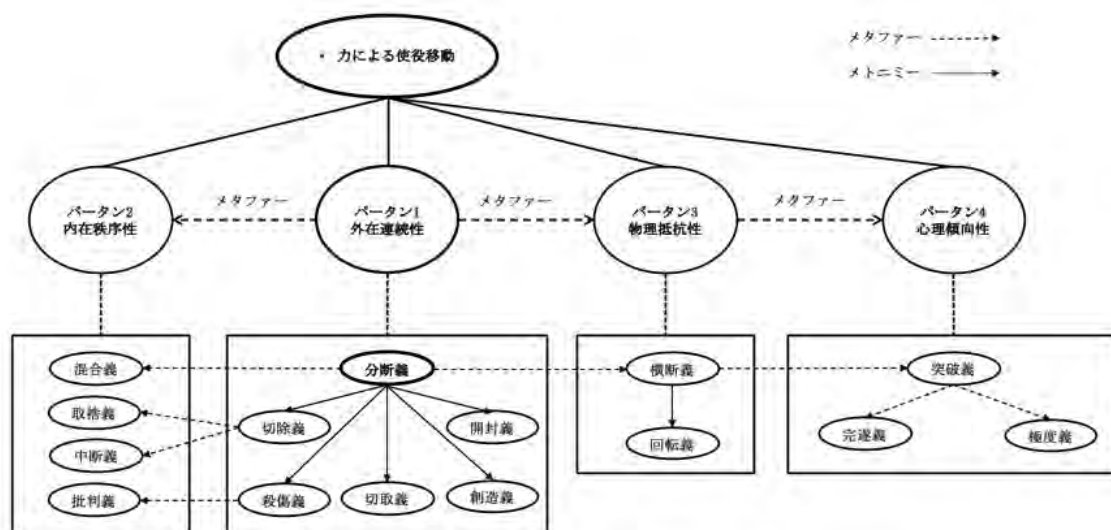


図8 「切る」の意味構造⁵（意味分析の結果を基に）

5.2. 心理実験の結果

心理実験では、母語話者と被験者それぞれのグループに類似性判断テストを行い、クラスター分析を行った。結果は図9と図10のようになった。

5.2.1. 母語話者の結果

母語話者の結果では、最長の定常状態の箇所にカッティングポイントを置くと、15個の「切る」の語義を含む30個の例文は大きく2群に分かれた。群内のコロケーションの内容を考察すると、左側の群は袋、紙、ケーキ、封、領収書、小切手、胃、爪などの目的語項を含むため、プロトタイプの概念である分断動作に近づいた語義のクラスターである。すなわち、中心義と中心義からのメトニミー拡張群である。右側の群は部門、麻雀、トランプ、世相、政治、水気、縁などの目的語項を含むため、中心義からより離れた拡張義のクラスターである。

さらに、2群の内部構造を分析すると、左側の群では、2つの下位であるクラスター（図9のI

⁵ 全体的な中心義と中心的な意味パターン＜外在連続性＞およびスーパー・スキーマを太線で示す。

と II を参照) が形成された。クラスター I は<分断義>、<開封義>、<切除義>、<切取義>、<殺傷義>のコロケーションを含み、<創造義>を除いた意味分析のパターン 1 の外在連続性と対応する。<創造義>のコロケーションは独立したクラスター II になった。「切る」の対象「囲炉裏」, 「型紙」は本来にあるもの全体の部分である点に着目するため、<創造義>が外在連続性のパターンに分類された意味分析の結果とずれていた。動詞と目的語項の関係から考えると、(17) 「ケーキを切る」の対象「ケーキ」が分断動作の対象であるのと異なり、「型紙」は分断動作の直接的な対象ではなく、直接的な対象の「紙」を分断して得られた結果である。すなわち、<創造義>では動作の結果が焦点化され、動作の対象が言語化されない。このため、<創造義>を他の語義と区別し、別途に考察する必要がある。また、Levin (1993) によると、john cut | a pattern out of paper / a slice of cake / a key | のように、「切る」の<創造義>は英語の cut にも存在している。しかし、このような意味用法は中国語の分断動作に見られない。

そして、右側の群では、3つのクラスター (図9のIII, IV, Vを参照) が形成され、それぞれ意味分析におけるパターン2の内在秩序性、パターン3の物理抵抗性、パターン4の心理傾向性と一致している。このため、母語話者のクラスター分析の結果は力動性理論に基づいた意味分析の妥当性のある程度検証できると言える。

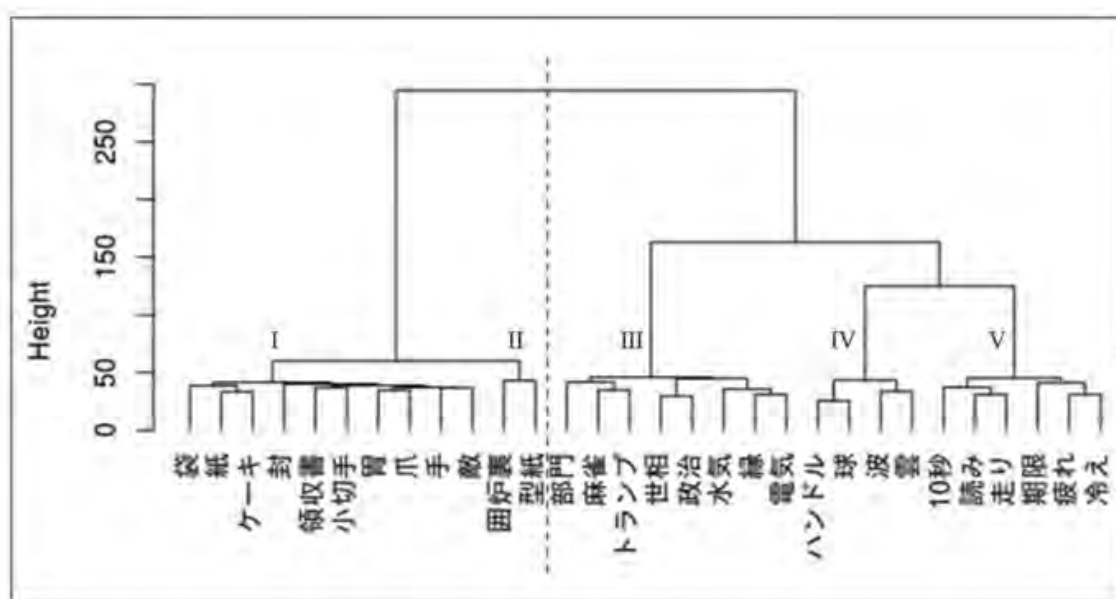


図9 JNSにおける「切る」の意味のカテゴリー化

5.2.2. 中国人学習者の結果

学習者の結果では、最長の定常状態の箇所にカッティングポイントを置くと、母語話者と同様に大きく中心義に近づいたクラスターと中心義に相対的に離れた拡張義を含むクラスターの2群に分かれた。

しかし、さらに2群の内部構造を分析すると、母語話者のカテゴリー化との差が見られる。ま

ず、左側の群では、＜創造義＞は独立したクラスターにならない点で母語話者の結果と一致していない。中国語の分断動詞においては、動作結果に焦点を当てた意味用法が存在しないため、中国人学習者はこの語義を十分に認識していないことがひとつの原因であると考えられる。右側の群では、母語話者の意味分析のパターンと一致していた3つのクラスターは、学習者の場合では同様なパターンは見られず、大きく2つの下位であるクラスター（図10のクラスターIIとIIIを参照）が形成された。クラスターIIは「切る」の複合動詞の語義を含む。複合動詞の語義が本動詞からの意味拡張経路を捉えにくくなり、本動詞と複合動詞の意味のつながりを意識していないため、学習者の方は多義語の本動詞の語義と複合動詞の語義を分けて認識する傾向が見られた。このため、複合動詞が独立したクラスターになった。クラスターIIIは部門、期限、縁、電気、政治、世相、カーブなどの抽象的なものを含むが、球、波、トランプ、麻雀、雲などの具象的なものも見られる。また、「切る」のような分断動作と類似する動作を行う語義とそうでない語義両方とも含んでいる。全般的な意味分類が少し混乱し、カテゴリー化していない傾向が示された。このクラスターの分類基準は明らかになっていない。

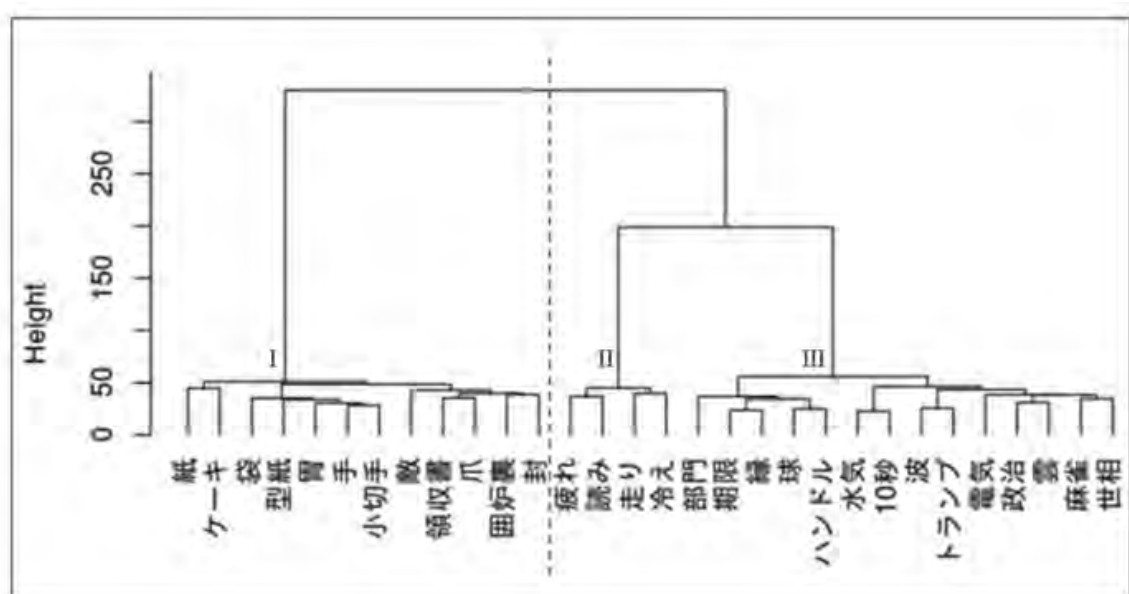


図10 CJLにおける「切る」の意味のカテゴリー化

6. まとめ

6.1. 得られた知見

本研究では、BCCWJ コーパスから抽出した用例で「切る」の意味のカテゴリー化に力動性に基づいた質的分析と心理実験による量的分析を行った。以下提示したリサーチクエスションの順に従い、得られた知見をまとめる。

まず、RQ1（包括的概念と意味構造）に関しては、力構造により「切る」の包括的概念である

スキーマは「力による使役移動という状態変化」である。また、そのスキーマの言語環境によるバリエーションとして、【パターン1 外在連続性】，【パターン2 内在秩序性】，【パターン3 物理抵抗性】，【パターン4 心理傾向性】という4つの意味パターンを立てた。その中に、中心義を含む外在連続性のパターンを4つの意味パターンの中心に据えた。語義間のつながりとして、従来の先行研究が示唆するメタファーやメトニミーなどの動機づけに加え、力構造の要素のパラメータを導入し、＜力動性+メタファー・メトニミー＞の形で関連づけた。図8に意味分析の結果に基づいた「切る」の意味構造を示した。

次に、RQ2（意味分析と心理実験による結果の差）に関しては、図9の母語話者のクラスター分析の結果が示すように、意味分析と心理実験の結果の間に、一致しないところも見られるが、心理実験における母語話者の結果では、力動性に基づいた4つの意味パターンのカテゴリー化と同様の傾向が見られた。そのため、力動性に基づいた意味分析の結果は統計的な手法である程度検証できる。意味分析と心理実験の結果を総合的にまとめた結果が図11のような意味構造である⁶。

最後に、RQ3（母語話者と学習者における意味構造の差）に関しては、図9と図10のカテゴリー化の結果を比較すると、母語話者と学習者のもつ「切る」の意味構造には違いが存在することが明らかにされた。学習者は具体的な分断動作を伴う語義が中心的語義群になったという点で一致しているが、下位のクラスターでは差が見られた。全般的に、母語話者のカテゴリー構造では、5つのクラスターが形成されたのに対し、学習者の方は3つのクラスターが形成され、カテゴリー化の基準が明らかにならなかった。

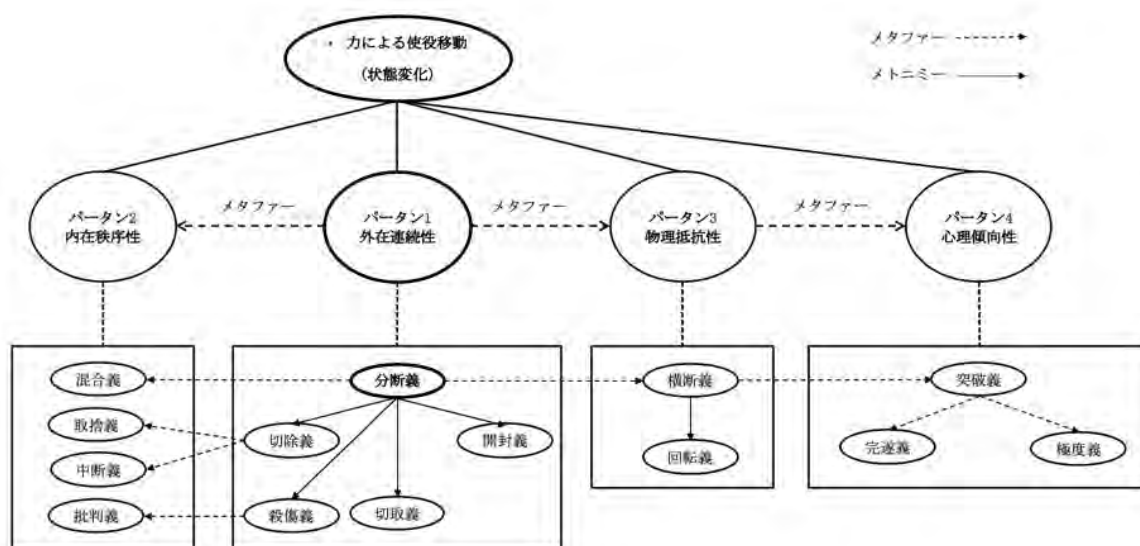


図11 「切る」の意味構造（意味分析と心理実験の結果を基に）

⁶ ＜創造義＞をどのようにこのカテゴリー構造に位置付けるかは今後の課題とする。

6.2. 今後の課題

以上で本研究の知見を整理した。しかし、本研究は力動性を用いた語彙研究を試みたもので、理論分析と心理実験のデザインでは、さらに検討すべき点は多く残っている。具体的には、今後の課題として、以下の4つを提出する。

一つ目は、コーパスに基づく考察の改善である。本研究はBCCWJの例文で分析を行ったが、少数例に基づく考察は説得力に欠けると考える。今後の研究で、意味構造分析を精緻化していくため、コーパス言語学の知見を援用し、対象用例のサンプルを抽出することを通し、より文脈的要素の特徴に注目する。また、「切る」の意味展開をさらに細かく分析し、以上の分析に立てる本研究の主張の妥当性を検証し、意味の関連性と意味展開のメカニズムを明らかにする。特に目的語項が被動作主ではない＜創造義＞はどのように位置付けるかを解明する。

二つ目としては、類型論の観点からの比較である。力動性理論は動作の参与者である主動体と対抗体における状態変化を示しているが、動作のプロセスである力の作用方式や強さなどを反映するものは管見の限りない。そのため、これらの点について別途に考察する必要がある。日本語の分断動詞群に属する他の多義動詞である「割る」、「裂く」、中国語の分断動詞「切(qie)」, 英語の分断動詞 cut, フランス語の分断動詞 couper との比較を行い、各言語における力の概念の捉え方と力動性のパターンの違いを考察することを通し、「切る」の意味・機能をさらに明確にする。

三つ目は、心理実験による検証の改善である。本研究の心理実験の部分は試行的なものとして、さらに改善すべき作業が多く残っている。本研究で行ってきた意味のカテゴリー化という実験では、カード分類法で被験者に意味が似たカードを同一のグループに分けるように指示した。しかし、本動詞と複合動詞を区別せず同時に提示すると、純粹に意味の類似ではなく、おそらく形の類似に着目する被験者もいるため、実験の精度に影響を与えている。また、意味のカテゴリー化以外に、語の概念形成と第二言語習得研究における動詞の意味習得に関する三つの認知プロセス（松田 2000）を踏まえ、他の三つの実験を加えようと考えている。一つ目の実験では意味の典型化というプロセスに関して、母語話者と学習者の産出する用例から典型概念を推測し、産出状況を明らかにするための文産出テストを行い、すぐに思いつく短文を五つ作るように求める形式を通し、母語話者と学習者の持つ「切る」のプロトタイプ概念と獲得された直感を探る。二つ目の実験は意味の一般化というプロセスを考察する容認性判断テストである。実際の場面と特定の文脈においては、「切る」の各意味・用法がどの程度容認されるかを確認し、学習者がその多義的意味用法を十分に使用できるかを考察する。また、意味拡張の動機づけによって語義の習得の難易度が異なるかを明らかにする。三つ目の実験は形式的には二つ目の実験と同様に容認性判断テストを行うが、意味の差異化という認知プロセスに注目する。日本語の分断動詞群における類義語は同じ場面と文脈では、いずれも容認されるか、それとも、適用される場面が異なるかを明確にする。また、学習者はどのくらいそれらの類義語を使い分けることができるかを分析する。さらに、意味分析との連続性を高めるため、今後の実験デザインにおいて、力のイメージ・スキーマと概念メタファーに関するメタフォリカル・コンピテンス (MC) の検定を以上の実験と組み合

わせようと考えている。

四つ目は、日本語教育への応用の可能性である。学習者と母語話者とのプロトタイプ概念とスキーマの概念および意味構造の違いを解明する上で、母語以外の要因を探る。その要因から得られた知見で、認知意味論による実際の日本語教育の指導における応用的な側面を探る。

文献

- Kellerman, Eric. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition*. 2(1), 37-57.
- Lakoff, George. (1987). *Women, Fire, and Dangerous Things: what our categories reveal about the mind*. Chicago: The university of Chicago Press.
- Lakoff, George. & Johnson, M. (1980). *Metaphors we live by*. Chicago: The university of Chicago Press.
- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Vol.1: Theoretical Perspective*. Stanford: Stanford University Press.
- Levin, Beth. (1993). *English Verb Classes and Alternations: A preliminary Investigation*. Chicago: University of Chicago Press.
- Talmy, Leonard. (1985). Force dynamics in language and thought. *In papers from the twenty-First Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.
- Talmy, Leonard. (2000). *Toward a cognitive semantics Volume2: Typology and process in concept structuring*. Cambridge, MA: MIT Press.
- 今井むつみ (1993) 「外国語学習者の語彙習得における問題点：言葉の意味表象の見地から」『教育心理学研究』41(3), 245-253. 教育心理学会.
- 石川慎一郎 (2017) 「現代日本語における「デ」格の意味役割の再考—コーパス頻度調査に基づく用法記述の精緻化と認知的意味拡張モデルの検証—」『計量国語学』31(2), 99-115.
- 許永蘭 (2008) 「「切る」の多義分析」『言葉と文化』9, 303-320.
- 栗田奈美 (2018) 『視覚スキーマを用いた意味拡張動機づけの分析』春風社.
- 瀬戸賢一・山添秀剛・小田希望 (2017) 『解いて学ぶ認知意味論』大修館書店.
- 鐘慧盈 (2016) 「L2「切る」の意味構造がその習得に及ぼす影響」『日本認知言語学会論文集』16, 555-560.
- 田中茂範 (1990) 「認知意味論—英語動詞の多義の構造—」三友社出版.
- 長友和彦 (1999) 「第二言語としての日本語習得研究—概観, 展望, 本科学研究の位置づけ—」『第2言語としての日本語の習得に関する総合研究』平成8年度～平成10年度基盤研究(A), 課題番号 08308019, 9-14.
- 姫野昌子 (1999) 『複合動詞の構造と意味用法』ひつじ書房.
- 松田文子 (2000) 「日本語学習者による語彙習得：差異化・一般化・典型化の観点から」『世界の日本語教育 日本語教育論集』10, 73-89. 独立行政法人国際交流基金.
- 松田文子・黄一君 (2011) 「コア図式論を用いた中国語動詞「取 (qu)」の意味記述—異文化理解を射程に入れた語彙指導の可能性に向けて—」『岡山大学大学院教育学研究科集録』146, 41-57.
- 森山新 (2012) 「認知意味論の観点からの「切る」の意味構造分析」『同日語文研究』27, 147-159.
- 森山新 (2015) 「日本語多義動詞「切る」の意味構造研究—心理的手法により内省分析を検証する—」『認知言語学研究』1, 138-155.

山梨正明（2000）『認知言語学原理』くろしお出版.

資料

MetaNet team. (2012-). MetaNet Metaphor Wiki.

https://metaphor.icsi.berkeley.edu/pub/en/index.php/MetaNet_Metaphor_Wiki.

国立国語研究所（編）（2011）「現代日本語書き言葉均衡コーパス」国立国語研究所.

執筆者紹介

田畑 智司（たばた ともじ）	大阪大学大学院言語文化研究科	教授
山田 彬堯（やまだ あきたか）	大阪大学大学院言語文化研究科	講師
徐 勤（Xu Qin）	大阪大学大学院言語文化研究科	博士後期課程
黄 晨雯（Huang Chenwen）	大阪大学大学院言語文化研究科	博士後期課程
福本 広光（ふくもと ひろみつ）	大阪大学大学院言語文化研究科	博士後期課程
王 鈺（Wang Yu）	大阪大学大学院言語文化研究科	博士前期課程

言語文化共同研究プロジェクト 2020

テキストマイニングと デジタルヒューマニティーズ 2020

2021 年 5 月 31 日 発行

編集発行者 大阪大学大学院言語文化研究科