

国立国語研究所学術情報リポジトリ

特定領域研究「日本語コーパス」平成21年度公開ワークショップサテライトセッション予稿集

メタデータ	言語: Japanese 出版者: 公開日: 2021-06-18 キーワード (Ja): キーワード (En): Balanced Corpus of Contemporary Written Japanese 作成者: 特定領域研究「日本語コーパス」総括班, General Headquarters Priority-Area Research "Japanese Corpus" メールアドレス: 所属:
URL	https://doi.org/10.15084/00003339

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 International License.



特定領域研究「日本語コーパス」

平成21年度公開ワークショップ サテライトセッション予稿集

平成22年3月14日

文部科学省科学研究費特定領域研究
「代表性を有する大規模日本語書き言葉コーパスの構築：
21世紀の日本語研究の基盤整備」

総括班
JC-G-09-02

特定領域研究「日本語コーパス」

平成21年度公開ワークショップ サテライトセッション予稿集

2010年3月14日（日）

Program [プログラム]

3月14日 (日)

13:00 ■開 会

13:00~13:10 ■領域代表者挨拶 前川 喜久雄

13:10~14:50 ■口頭発表

「認識的モダリティ形式を含む引用節と主節述部・被修飾名詞の意味的関連性

—〈(の) だろう〉〈(の) だろうか〉の考察—」佐藤 雄亮

「反対概念の非対称性の普遍性と限界

—日本語と中国語における〈上／下〉の非対称性を中心に—」徐 蓮

「『現代日本語書き言葉均衡コーパス』を利用した格助詞の正誤判定」大山 浩美

「『(さ) せていただく』と『いたす』の競合」東泉 裕子

14:50~15:20 ■招待発表

「文芸創作における日本語コーパス活用」光原 百合

15:20~15:30 ■ポスター内容紹介

15:30~17:30 ■ポスター発表

「カタカナ表記のゆれと意味 —WWW調査との比較—」吉田 充良

「日本語教育から見た均衡コーパス」李 在鎬、秋本 瞳

「『Yahoo! 知恵袋』を利用した『の』の曖昧性解消の試み」西口 純代

「『～がる』と感情を表す『～む』動詞の意味・用法」韓 金柱

「嫌悪を表す『いや』と『きらい』について」加藤 恵梨

「話し言葉、書き言葉、そしてオンライン言語をめぐって

—日本語全体像を捉える試みへのパイロット・リサーチ—」西村 由起子

「動作の方向を表す『～テクル』に対する要求条件の解明」石山 友之

「BCCWJを用いた『語の文体』の位置づけ」井上 次夫

「複合辞トイウにみる文法化の諸相」金城 由美子

「BCCWJにより選定した日本語教育語彙の検証

—経営学を専攻する日本語学習者を対象にして—」寺嶋 弘道

17:30 ■閉 会

Contents [目次]

領域代表者挨拶

前川 喜久雄	1
--------------	---

口頭発表

「認識的モダリティ形式を含む引用節と主節述部・被修飾名詞の意味的関連性 —〈(の) だろう>〈(の) だろうか>の考察—」	3
佐藤 雄亮	
「反対概念の非対称性の普遍性と限界 —日本語と中国語における<上/下>の非対称性を中心に—」	13
徐 蓮	
「『現代日本語書き言葉均衡コーパス』を利用した格助詞の正誤判定」	21
大山 浩美	
「『(さ) せていただく』と『いたす』の競合」	27
東泉 裕子	

招待発表

「文芸創作における日本語コーパス活用」	33
光原 百合	

ポスター発表

「カタカナ表記のゆれと意味 —WWW 調査との比較—」	35
吉田 充良	
「日本語教育から見た均衡コーパス」	45
李 在鎬、秋本 瞳	
「『Yahoo! 知恵袋』を利用した『の』の曖昧性解消の試み」	51
西口 純代	
「『～がる』と感情を表す『～む』動詞の意味・用法」	61
韓 金柱	
「嫌悪を表す『いや』と『きらい』について」	69
加藤 恵梨	
「話し言葉、書き言葉、そしてオンライン言語をめぐって —日本語全体像を捉える試みへのパイロット・リサーチ—」	73
西村 由起子	
「動作の方向を表す『～テクル』に対する要求条件の解明」	85
石山 友之	
「BCCWJを用いた『語の文体』の位置づけ」	91
井上 次夫	
「複合辞トイウにみる文法化の諸相」	101
金城 由美子	
「BCCWJにより選定した日本語教育語彙の検証 —経営学を専攻する日本語学習者を対象にして—」	107
寺嶋 弘道	

領域代表者挨拶

3月14日（日） 13:00～13:10

▶前川 喜久雄

領域代表者挨拶

前川喜久雄（国立国語研究所）

文部科学省科学研究費特定領域研究「日本語コーパス」は、広く一般に公開可能な書き言葉均衡コーパスを整備することと、コーパス日本語学の方法を基礎と応用の両面において開拓することを目標として、2006 年度以来 5 年計画で活動を行っています。現在、ちょうど第 4 年度の活動を終えるところですが、特定領域研究最大の目標である 1 億語規模の『現代日本語書き言葉均衡コーパス』の構築は順調に進捗しており、データそのものは近日中に 1 億語を超えることが確実な状況となっています。現代語コーパスの常として著作権処理の問題がありますが、こちらも何とか 5000 万語を超えるサンプルの処理が終了するにいたっています。

ところで、コーパス日本語学の振興については、特定領域研究の枠を超えて、一層広い範囲の研究者にデータを公開することが有益であると考えられます。そのため『現代日本語書き言葉均衡コーパス』のうち、著作権処理が終了したデータを領域外の研究者に対しても早期に提供することとし、2008 年以来、研究者向けのモニター公開を行ってきています。モニター版コーパスの規模は、2009 年 7 月には約 4500 万語にまで拡張されており、試用申し込み件数は 2009 年度だけで 500 件を超えています。

このモニター版コーパスを利用した研究発表の場として、昨年引き続き、本サテライトセッションを開催します。本セッションをコーパス日本語学発展の場として機能させ、あわせて『現代日本語書き言葉均衡コーパス』評価の場としても機能させるため、参加者のみなさまには活発な討議をお願いしたいと思います。また、明日と明後日に東京工業大学（大岡山）で開催される「日本語コーパス」の公開ワークショップにも是非ご参加ください。

口頭発表

3月14日（日） 13:10～14:50

認識的モダリティ形式を含む引用節と主節述部・被修飾名詞の意味的関連性
—〈（の）だろう〉〈（の）だろうか〉の考察—

▶佐藤 雄亮

反対概念の非対称性の普遍性と限界

—日本語と中国語における〈上／下〉の非対称性を中心に—

▶徐 蓮

『現代日本語書き言葉均衡コーパス』を利用した格助詞の正誤判定

▶大山 浩美

「（さ）せていただく」と「いたす」の競合

▶東泉 裕子

認識的モダリティ形式を含む引用節と主節述部・被修飾名詞の意味的関連性—〈(の) だろう〉〈(の) だろうか〉の考察—

佐藤雄亮（東京外国語大学大学院地域文化研究科博士後期課程）[†]

The Semantic Correlation between the Quotative Clause Containing Epistemic Modals and the Predicates in the Main Clause

Sato, Yusuke (The Graduate School of Area and Culture Studies, Tokyo University of Foreign Studies)

1. 本研究の目的と背景

本研究は、〈だろう〉類の各形式が「と」や「って」などを用いる従属節（以下、日本語記述文法研究会(2008)に倣い「引用節」と呼ぶ）内に現れる用例について、それと共に起する主節述部の意味、およびそれらの意味分類ごとに認められる用例数の異なりに各形式の意味が反映していること、そしてそれが意味の記述に対する実証的な根拠となることを論じる¹。考察においては BCCWJ（2008 年度版）の用例を対象にする。各用例の意味分類基準として国立国語研究所（編）(2004)（以下、『分類語彙表』と呼ぶ）を用いた。これは意味の記述により客観的基準を設けようという試みの一つである。

引用節との共起における差異に基づき、モダリティ形式の意味を論じる研究はあまり見られない。その中で、森山(1988)、藤田(1986, 2000)は本研究にとって参考にすべき先行研究である。

森山(1988)は上の観点を「報告動詞分析」と呼び、『文』ト～スルというように、文の発話的な意味をどう報告するかによって、その文の表現的な意味を考えようとするものである」と説明する(森山 1988: 249)。例として、「だろう」を用いた文が「[彼はもうすぐ大学をやめるだろう]と判断する／した。」と言い換えられることなどが指摘されている。

特定の主節述部が取る引用節に、どのような性質が認められるか考察を進めた諸論考が藤田(2000)に収められている。日本語には、「おはようと言った。」など、「述部が引用句の発言・思考と事実上等しい動作・状態を表す」典型的な引用と言えるもの（「第Ⅰ類の引用」）と、「「おはよう。」と鈴木が入ってきた。」などのように「述部が引用句の発言・思考と共存する動作・状態を表す」もの（「第Ⅱ類の引用」）の二種類が認められる(藤田 1986, 2000)。この内、第Ⅰ類に該当する例では、「言った」などの主節述部は、「おはよう」という具体的な発言・心内発話に対して、抽象化を行ったものと見ることが出来る(藤田 2000: 71-72)。一方、第Ⅱ類の引用に見られる「引用句と共存する動作・状態」は、ある程度の緩やかな

[†] sato.yusuke.djc@tufs.ac.jp

¹ モダリティ形式の従属節における出現の可否という文法的性質を指摘した研究として、南(1974, 1993)などが挙げられる（それらの現象から各形式の意味の説明も試みられている）。大規模なコーパスからの用例数やその割合の違いを基に、モダリティ形式の類型化を論じるナロック(2006)、Narrog(2009)も本研究は参考にしている。

条件さえ満たせば、文脈に支えられて大抵の表現が許容されてしまう(藤田 1986: 213; 2000: 224) ものとされるため、引用節と主節述部の意味上の結び付きは弱いと判断できる。

本研究は、これらの先行研究を参考に、モダリティ形式を含む引用節と共起し、「第Ⅰ類の引用」を構成する主節述部を分析することによって、各モダリティ形式(を用いた節)がどのような行為・状態を表すものと捉えられているか考察するものである。

2. 研究方法

2.1. 用例の収集方法

本研究では、宮崎(2002)、日本語記述文法研究会(編)(2003)などにおいて「認識のモダリティ」を表すとされた形式の中から、〈だろう〉を取り上げ、〈の(だ)〉が前接する〈の(だ)だろう〉と併せてそれらが引用節内に現れる用例を考察の対象とする。さらに、〈(の) だろう〉の用例の中には〈(の) だろうか〉に類する例が認められるため、それらについても考察する。上記の形式が(A)動詞もしくは形容詞に上記のモダリティ形式が接続し、(B)引用を表す「と」「って」、もしくは例示を表す「など」「なんて」がそれらに接続する例を考察の対象とする²。

引用に関連して、「という」や「との」を介した名詞修飾表現がある。この内、「何か盗んだのだろうと疑った／何か盗んだのだろう {という・との} 疑い³」のように【引用節—主節】との平行性が認められる例は考察対象とする。さらに、具体的な意味を失った名詞を修飾している例の中で、「問題が起こるだだろう {ということが／と} 懸念される」のような例ではどちらもほぼ同じ意味を表すと判断できるため、考察対象に含める。

用例の検索は、BCCWJ 内に含まれる『ひまわり』BCCWJ パッケージを用いた。各要素が取りうる形式的変異をカバーできるよう次の①～⑤全体を検索語とし、正規表現を用いた検索を行った。

- ①「動詞・形容詞終止形」の指定：[うくぐすつぬぶむるいただ]⁴
- ②〈の(だ)〉の諸形式：[のん]
- ③〈だろう〉類の諸形式：(だろう|でしよう|だろうか|でしょうか)
- ④文末で現れうる要素：[?!、'…] ※「0 回以上の出現」で指定
- ⑤引用表現の指定：(と|って|なんて|など)

2.2. 用例の分類基準

前節の方法で収集した用例について、引用節を受ける主節述部あるいは名詞句を、意味

² 「すぐに帰るように命じた」や「すぐに帰れ。私はこう命じました」など、〈ように〉やダイクシスを用いる例に関しては、今回は調査対象から除外した。

³ 以下、例文においてはモダリティ形式に当たる部分に下線を、引用節を受ける主節述部・被修飾名詞句に囲み線を付す。疑問語の現れる例では疑問語に二重下線を付す。

⁴ 〈なければならない〉のみ、前に接続する文字列を指定しなかった。母音語幹動詞の場合、〈なければならない〉に前接する未然形の末尾がそのまま動詞語幹と同じ形になる場合(見・なければならない、など)があり、一様な文字列指定ができないためである。

によって分類したものを本研究の考察材料とする。コーパス検索で得られた用例のそれぞれについて、「と（いう）」などを受ける主節述部・名詞句について、『分類語彙表』における「中項目」、さらにその下位の分類である「分類項目番号」を付与し、対応する項目名と共に各例の意味分類とする⁵。論じる内容によっては、さらに下位の分類として設けられる「段落」を利用する箇所もある。

3. 考察

まず、平叙型、疑問語疑問型、肯否疑問型という引用節の類型別の特徴を指摘し、次に、平叙型、疑問語疑問型での現象から、各形式の意味特徴を論じる。

3.1. 〈(の) だろう (か)〉と引用節の類型との関係

引用節における〈(の) だろう〉には、(1a, b)のような平叙型の例だけではなく、(2a)(3a)のように疑問語（何、誰、なぜ、など）を伴う疑問型の例もある。疑問型の例はそれぞれ(2b)(3b)における〈(の) だろうか〉から「か」を省略したものと捉えられている⁶。〈(の) だろうか〉の例には、(4a)(5a)のように疑問語を伴わない、肯否疑問型も存在するが、肯否疑問型における〈(の) だろうか〉は、「か」を省略すると文意が保てない（例(4b)(5b)）。

(1a) 彼が行く だろう と 考えた。

(1b) 彼が行く のだろう と 考えた。

(2a) 誰 が行く だろう と 考えた。

= (2b) 誰 が行く だろうか と 考えた。

(3a) 誰 が行く のだろう と 考えた。

= (3b) 誰 が行く のだろうか と 考えた。

(4a) 本当に彼が行く だろうか と 考えた。

≠ (4b) 本当に彼が行く だろう と 考えた。

(5a) 本当に彼が行く のだろうか と 考えた。

≠ (5b) 本当に彼が行く のだろう と 考えた。

本研究では以上を踏まえ、〈(の) だろう〉における平叙型と疑問語疑問型、〈(の) だろうか〉における疑問語疑問型と肯否疑問型を区別し、それぞれの型ごとに考察を進めることとする。まず、表 1 に、〈(の) だろう〉〈(の) だろうか〉の引用節内における用例数を示す。例数に加え、4 形式全体の合計数（右端）に占める割合も併せて示す。

〈だろう〉と〈のだろう〉は、平叙型と疑問語疑問型においてその割合が大きく異なり、平叙型では約 3:1、疑問語疑問型では約 1:7 の比率になる。〈だろうか〉〈のだろうか〉の疑問語疑問型の例も、その割合は約 3:7 と〈のだろうか〉の割合が多い。先行研究では疑問

⁵ 基本的には『分類語彙表』に基づいた意味分類を用いるが、筆者が改めたものがある。

(a)『分類語彙表』に無い語：筆者の判断によって、置き換えが最も適当と考えられる語の番号を付与する
(b)同一品詞内で複数の項目に出現する語（例：「訴える」は、「2.3123」「伝達・報知」、「2.3611」「裁判」に掲載）：同じ分類項目内に掲載される語などを参考に、各例における意味に最も合致する項目の番号を付与する
(c)派生関係にある語が特定の品詞で複数の分類項目番号に割り振られている場合：最も多くの品詞に共通する項目番号を付与する
(d)『分類語彙表』が考慮しているのとは違う意味の用例であると思われる場合：類似する語の分類項目番号に変更する：「法律では、居住するには免許を取得しなければならない」とある」における「ある」を「存在」から「書き」に変更する、など

⁶ 例えば、宮崎(2001: 16)では、「まず、「ダロウカ」については、普通体「ダロウカ」の他、丁寧体「デショウカ」や、「誰だろう？」のように wh 疑問文において「カ」が脱落したものを含む。」と述べられている。

語が述部以外の箇所に見れる場合、「質問のフォーカスが明らかなので、スコープの「の

表 1 〈(の) だろう〉〈(の) だろうか〉の用例数

	だろう		のだろう		だろうか		のだろうか		合計
平叙型	1177	73.8%	418	26.2%	—		—		1595
疑問語疑問型	25	8.0%	180	57.7%	31	9.9%	76	24.4%	312
肯否疑問型	—		—		71	36.8%	122	63.2%	193
合計	1202	57.2%	598	28.5%	102	4.9%	198	9.4%	2100

（だ）」は免除されやすい(野田 1997: 123)」という指摘がなされているが、コーパスでの現われを見る限り、疑問語と〈(の)だ〉は結び付きやすいと判断できる。

3.2. 平叙型引用節

本節では、平叙型引用節に見れる〈だろう〉〈のだろう〉の意味特徴を、共起する主節述部・名詞句の出現数や割合から論じる。まず、該当する用例の全体数を表 2 に示す。表には『分類語彙表』の「中項目」毎の用例数を挙げる。

表からは〈だろう〉〈のだろう〉双方において、最も大きな割合で用いられるのが"30"の「心」、次に多いのが"31"「言語」であることがわかる。以下ではこれらの用例の内、〈だろう〉〈のだろう〉に違いの現れるものを、「分類項目」を毎に確認していく。

表 2 平叙型引用節の用例数と割合

中項目	項目名	語例	だろう		のだろう	
13	様相	整理する	1	0.1%	0	0.0%
15	作用	まとめる	4	0.3%	1	0.2%
30	心	思う	951	80.8%	374	89.5%
31	言語	言う	203	17.2%	43	10.3%
33	生活	祝福する	1	0.1%	0	0.0%
35	交わり	否定する	7	0.6%	0	0.0%
36	待遇	教える	10	0.8%	0	0.0%
合計			1177	100.0%	418	100.0%

まず"3066"「判断・推測・評価」を取り上げる。各形式の用例数は次のとおりである。割合は平叙型引用節の全用例数に対する割合を示す。

表 3 平叙型引用節における"3066"「判断・推測・評価」の用例数と割合

分類項目	項目名	語例	だろう		のだろう	
3066	判断・推測・評価	想像する	75	6.4%	14	3.3%

〈のだろう〉の該当例について、『分類語彙表』の最下位の分類「段落」を確認すると、"01, 07, 08"の段落、具体的には「判断する」「評判する」「創造する」「察する」「推察する」「推測する」「読む」の例が現れた。一方、〈だろう〉には〈のだろう〉と共起したもの以外に、"03, 05, 06, 09, 10, 12, 17, 18"の例が出現する。〈だろう〉とのみ共起するのは「推定する」「決め込む」「妄想」「類推する」「予想（する）」「予期（する）」「見込み・見込む」「見通し・見通す」「想定する」などの例である。〈だろう〉と共起する語にのみ、「予想」「予測」「予期」「見通し」など、引用節の内容が未実現のものに偏ると考えられるものが含まれる点が特徴的である。

上に示した〈だろう〉〈のだろう〉が共起する語の違いは、〈のだろう〉の表す内容が既実現に偏り、一方〈だろう〉では、未実現の内容も表せることに原因が求められる。〈のだろう〉が表す推測内容について幸松(2007)の説明が参考になる。幸松(2007: 129)は、「改札が混雑しているから、電車がついたのだろう。(幸松(2007: 121 例 2))」などの例に対して、「「のだろう」にはもともと〈与えられた状況からその原因・理由・事情などを推測する〉という意味がある」と述べ、「原因・理由・事情」は、話し手が得た情報（この情報は「から」従属節に現れうる）よりも時間的に遡る事態であることがほとんどであると説明している。話し手が発話時点で得た情報よりも時間的に遡るということはつまり、その推測内容は既実現であることが多いということになる。一方〈だろう〉の場合、その内容が「原因・理由・事情」になることはほとんど無く⁷、推測内容の既実現への傾きももたらさない。

(6)同社のトップ層でも、一〇年くらいの時間をかけて、中国をコカ・コーラの世界における最大の市場にしようと狙っている。その目標を実現するために、コカ・コーラ社はこれからさらに現地化の努力に力を入れるだろう。コカ・コーラの中華カラーもこれからますます強化され、クローズアップされていくだろう、と予想がつく。（中国の心をつかんだ企業戦略―日米欧中の勝ち組に学べ）

(7)彼は早く父親を超えたいという思いから、店長のころから経営の改革に取り組み、新たにカラオケ店やケーキショップを開店するなど、事業もどんどん拡大していった。だが、それは流行に乗って儲かるだろうという見通しだけで出店した、戦略不在の拡大路線である。（「ありがとう」戦略―その一言が人生と経営を劇的に変える）

以上の違いが、「予想」などの語が〈のだろう〉ではなく〈だろう〉と共起することに影響しているのだと考えられる。

〈のだろう〉と〈だろう〉と主節述部・名詞句の分布において、他に特徴的なものとして、"3132"「問答」に関する例を取り上げる。「段落」を参考に聞き手への質問を表すものと、返答を表すものを分けると、〈だろう〉においては質問類が1例、返答類が13例とな

⁷幸松(2007: 131)の調査では、「原因・理由・事情」を推測する〈のだろう〉の例が1例のみ現れている（〈のだろう〉は40例）。ただし、これは「非常に珍しい(幸松 2007: 130)」とあるように、完全に自然な例とは見做しにくい。

空港を出発しても、平壤市内への沿道に花を手にした人波が続き、驚いた。60万人も出たというから、動員があっただろう。（幸松 2007: 131）

るのに対し、〈のだろう〉においては質問類が 10 例、返答類が 2 例と、割合が逆転する。各該当例を示す。

(9) 【〈だろう〉 質問類：1 例】 先日その子にそっくりの中 3 の男の子を見つけました。その子の友達に「彼モテるでしょう？」と尋ねると、全然モテないしチョコをくれる女の子もいないそうです。 (Yahoo!知恵袋)

(10) 【〈だろう〉 返答類：13 例】 「...なに？ 初めてはやっぱ、海の見えるコテージがいか、星空を見ながらとか、夢持ってる？」

女の子じゃあるまいし、そんなわけないだろう、と思わず言い返す。

(家庭のジジョー)

(11) 【〈のだろう〉 質問類：10 例】 それで、「お坊さん、どこへ行くんですか」と聞いたら「商売に行くんです」というのです。「商売って、何しに行くんですか。布教に行くんでしょう？」と私が聞くと、「中国で卒塔婆をつくってるんです」というのです。

(環境考古学のすすめ)

(12) 【〈のだろう〉 返答類：2 例】 なかのひとりの男がおれにニガーたちが盗むところを見たかどうか聞いた。(中略) ただニガーたちはおれの主人を起こしてしまったのではとおそれて、面倒なことにならないうちに逃げちゃえと思っていたのだろうと答えた。

(マーク・トウェインコレクション)

上記の質問類の用例では、〈だろう〉〈のだろう〉はこれまでの研究で確認要求機能と呼ばれた機能を担う⁸。〈だろう〉と〈のだろう〉の確認要求機能をめぐって違いが存在することはこれまでの研究では指摘されていないが、今回の調査結果からは特に確認要求機能を表しやすいのは〈のだろう〉であると判断することができる。

3.3. 疑問語疑問型引用節

本節では、疑問語疑問型引用節に現れる形式の意味特徴を論じる。疑問語疑問型引用節では〈だろう〉〈のだろう〉〈だろうか〉〈のだろうか〉の全てが現れうる。中項目ごとの用例数と割合を示した表 4 から、ここでもやはり「心」「言語」の用例数がその大半を占めることがわかる。

前節に続いて、「3132」「問答」の例から確認する。既に指摘したように、疑問語疑問における〈(の) だろう〉の例は、「(の) だろうか」から「か」を省略したもの、と論じられてきた。ところが、今回の調査結果ではこの『か』の省略の有無によって、用例数の分布・割合が異なることが明らかとなった。「問答」における用例数を表 5 に示す。

⁸ 確認要求機能を扱った先行研究は多い。その中で例えば安達(1999)は、確認要求機能を、典型的な質問文の持つ機能のうち、「話し手には命題内容の真偽判断、あるいはその命題を構成する情報の一部が欠けている」という「不確定性条件」が成立しない（つまり、話し手が述べる内容は話し手にとって確定した事態だと捉えられている）点に特徴があると定義している。

表 4 疑問語疑問型の用例数と割合

中項目	項目名	語例	だろう		のだろう		だろうか		のだろうか	
30	心	疑う	21	84.00%	160	88.90%	21	67.70%	54	71.10%
31	言語	言う	3	12.00%	18	10.00%	10	32.30%	20	26.30%
34	行為	徒労感	0	0.00%	1	0.60%	0	0.00%	0	0.00%
35	交わり	相談する	1	4.00%	1	0.60%	0	0.00%	0	0.00%
36	待遇	頼む	0	0.00%	0	0.00%	0	0.00%	2	2.60%
合計			25	100.00%	180	100.00%	31	100.00%	76	100.00%

表 5 疑問語疑問型に占める「問答」の用例数と割合

分類項目	項目名	語例	だろう		のだろう		だろうか		のだろうか	
3132	問答	聞き返す	2	8.0%	5	2.8%	6	19.4%	10	13.2%

以下の用例に見られるように、「問答」と疑問語疑問型の共起例では、主節述部・名詞句は質問類のもののみであり、返答類の共起例は認められない。これは、疑問語を用いて返答を行うことが不自然であるためだと説明できる。

(13)—樊遲が、知者といわれる人は、民を治めるのに、どういう方法を採るだろうと訊いた。それに答えて、子は仰言った。(後略) (孔子)

(14)男の人と食事したとき、お店を出たところで「私どのくらい出せばいいでしょうか？」と聞いていますが、それでいいんですよね？レジの前でお財布出したくはないのでお店を出たあとにしてるのですが。。。(Yahoo!知恵袋)

(15)実際は、この二次会に来ている人々のすべてが、「なぜ行くのか」などという問いを持つことなく、多田は「ああいう人だから」と了解しているのに、やはり私だけが「なぜ行くのだろう」という問いに十分に答えられないことが、妙に苛立たしかったのだ。
(有名であれ無名であれ)

(16)新入社員が「われわれの事業には、いったいどのような意味があるのでしょうか？」と課長に質問しても、「そんなことを考えている暇があったら、さっさと伝票の書き方でも覚えろ」と怒られるのが関の山でしょう。(経営戦略入門講座)

これらの用例についての表 5 を確認すると、「か」が現れる〈だろうか〉〈のだろうか〉では、「問答」の用例が占める割合がそれぞれ 19.4%, 13.2%なのに対し、〈だろう〉〈のだろう〉においてはそれぞれ 8.0%, 2.8%と、割合が低くなることがわかる。「問答」の全 23 例内における割合を見ても、「か」の現れる形式の用例が 16 例と、「か」の無いものの 2 倍以上となる。この結果からは、「問答」に関連する語と〈だろう〉類が共起する場合、機能の面からは省略が可能であるものの「か」を明示することが好まれることが明らかになったといえる。

次に、「3013」「安心・焦燥・満足」について論じる。この項目に関しては、平叙型、肯

否疑問型の場合と疑問語疑問型の違いについて確認する。次の表 6 に、各形式における「安心・焦燥・満足」の用例数と割合、加えて各形式の全体の用例数と割合を示す。

表 6 「安心・焦燥・満足」および全体における類型毎の用例数

		だろう		のだろう		だろうか		のだろうか	
平叙型	"3013"	13	100.0%	1	<u>5.6%</u>	—		—	
	全体	1177	97.9%	418	<u>69.9%</u>	—		—	
疑問語疑問型	"3013"	0	0.0%	17	<u>94.4%</u>	2	16.7%	7	29.2%
	全体	25	2.1%	180	<u>30.1%</u>	31	30.4%	76	38.4%
肯否疑問型	"3013"	—		—		10	83.3%	17	70.8%
	全体	—		—		71	69.6%	122	61.6%
合計	"3013"	13	100.0%	18	100.0%	12	100.0%	24	100.0%
	全体	1202	100.0%	598	100.0%	102	100.0%	198	100.0%

〈だろう〉〈だろうか〉〈のだろうか〉においては、平叙型、疑問語疑問型、肯否疑問型のそれぞれが全体の用例数に占める割合と、「安心・焦燥・満足」の例に占める割合とは大きな差を見せない。一方、〈のだろう〉では、全体の用例数に対して平叙型が約 7 割を占めるのに対し、「安心・焦燥・満足」の場合には平叙型が 1 例のみ、割合にして 5.6% となっており、極端に疑問語疑問型の用例に偏っていることがわかる。さらに、平叙型の用例は、(17)の「慰め」という「安心」類の例であるのに対し、疑問語疑問型の用例は「不安(だ)」「気になる」「心配(だ/する)」「気をもむ」など(例(18)(19))「焦燥」類の例であった。

(17) (母の様態が急変し、病院にたどり着いた時にはもう亡くなっていた)

担当医師が、遠くの息子さんに会えて安心されたのでしょうと慰めのつもりで言ってくれたが、僕は母の最期にも息子として認知されてはいなかった。(境界線上の人々)

(18) 自費と聞いて、安堵のなかにも不安がかくせない千太郎の表情を見つめながら、鯛子の入院費はいったいだれが支払うのだろうと、私の方も不安を覚えた。(青春医者のないしょ話)

(19) 「そのせいだったんだ。唐突に電話が切れたから、何が気に障ったんだろうって、心配したよ」(楽園の鳥 ―カルカッタ幻想曲―)

〈だろう〉においては、平叙型であれば次のように「安心」類の用例(4例)も、「焦燥」類の用例(9例)も認められる。一方、疑問語疑問型の用例は1例も認められなかった。

(20) 昔に比べたら、ずいぶん販路はせまくなったけれど、こういう高級品は、日本のつづくかぎり、需要は絶えないだろうと、店の前途には楽観している。(一筋の道)

(21)そして続けてアナウンスは言った「コンコルドがもうすぐ見えます。しばらく停車することにします」。

乗客の中には急いでいる人もいるだろうと私は心配した。(自家中毒)

これらの結果からは、特に「焦燥」類のものと共起する場合、引用節が平叙型(21)などであれば〈だろう〉が、疑問語疑問型((18)(19)など)であれば〈のだろう〉が用いられるという使い分けが成立していると判断できる。既に 3.1.で指摘したように、全体的に見ても疑問語と〈のだろう〉の結びつきは強いと考えられるが、主節述部・名詞句の意味によっては、単なる結びつきの強さというよりも、疑問語の共起の有無が両者を使い分ける条件になっているものが存在していることが明らかになった。

同様の関係は、「安心・焦燥・満足」の他に、次の"3014"「苦悩・悲哀」においても認められる。つまり、疑問語疑問型の用例では〈のだろう〉の用例のみが認められたのである。「苦悩・悲哀」においてはさらに、疑問語疑問型のみならず、平叙型においても〈だろう〉の用例は認められなかった⁹。該当する例を以下に示す。

(22)あとになってから、どうしてアメを渡されたときに、きっぱり断らなかったんだろう、なんで落着いて対処できなかったんだろう、と情けない気持ちでいっぱいになりました。(Yahoo!知恵袋)

(23)でも、それは他人ごとではないからわかるのだと思う。私にしても、「あんなに丈夫だった私が、どうしてこんなに足が弱くなったのだろう」と、話は違うが過去にとらわれて一人で嘆くこともたまにはある。(私の気ままな老いじたく)

(24)四〇年以上信子に抽象画を指導した末松によれば、信子は最初から自分の絵のスタイルをもっていて、それはずっと変わらなかったという。(中略)そして、初めの頃は緑系の色を使ったが、だんだん赤に移っていった。そして、いつもなんで抽象画をやるのだろうと悩んでいたという。(ビッグ・リトル・ノブ)

4. 本研究のまとめ

以上の本研究における指摘は、次のようにまとめられる。

- ・〈のだろう〉〈のだろうか〉と疑問語疑問型の結び付きが強い
- ・平叙疑問型において、未実現の内容に関連する「予想」「予測」「予期」「見通し」とは〈だろう〉のみが共起し、確認要求機能としては〈だろう〉よりも〈のだろう〉の使用例が多い
- ・疑問語疑問型において、質問類の表現と共起する場合には〈(の) だろうか〉という形

⁹ 他に、〈だろうか〉1例、〈のだろうか〉3例が認められるが、これらはどちらも肯否疑問型の用例であり、疑問語疑問型では〈のだろう〉の用例のみであった。

(a)けれど、「昇進うつ」は、ポストに見合っただけの働きが自分にできるだろうかという苦悩によって発症する。(「うつ」から元気になる本)

(b)私は子供が欲しいのですが、妻はそうではありません。(中略)仕方ないとは思いますが、全く欲しくはなくこのまま2人がいいと言われてしまうと、このままの人生これでいいんだろうかと悩んでしまいます。(Yahoo!知恵袋)

式による「か」の明示が好まれ、「焦燥」「苦悩・悲哀」を表す表現との共起においては、疑問語の有無が〈だろう〉と〈のだろう〉の使い分けの条件となっている

本研究では、先行研究には無い指摘を実証的データと共に示すことができた。文法的には自然な用例の中にも、他に比べてより多く認められるもの、逆にかなり限定された例しか認められないものが存在する。そのような点についても数量を基に論じられるのが、コーパスを利用する利点の一つである。さらに、先行研究においてなされてきた意味記述に対して、それを補強する実証的データを示しうるという点も本研究の大きな主張の一つである。

今回考察を示せたのは、〈だろう〉類全体から見れば一部の点についてのみである。〈だろう〉類に関して、本研究の考察方法を採用することで実証が可能になる他の現象、さらには他のモダリティ形式に関しても同様の考察方法を採用することが有効であると考ええる。今後の課題として考察を続けたい。

参考文献

- Narrog, Heiko (2009) *Modality in Japanese: The Layered Structure of the Clause and Hierarchies of Functional Categories*. Amsterdam; Philadelphia: John Benjamins.
- 安達太郎 (1999) 『日本語疑問文における判断の諸相』 東京：くろしお出版
- 国立国語研究所（編）(2004) 『分類語彙表（増補改訂版）』 東京：大日本図書
- ナロック、ハイコ(2006) 「従属節におけるモダリティ形式の使用」 『日本語文法』6-1: 21-37.
- 日本語記述文法研究会（編）(2003) 『現代日本語文法 4 第8部 モダリティ』 東京：くろしお出版
- （編）(2008) 『現代日本語文法 6 第11部 複文』 東京：くろしお出版
- 野田春美 (1997) 『「の（だ）」の機能』 東京：くろしお出版
- 藤田保幸 (1986) 「文中引用句『～ト』による『引用』を整理する」 宮治裕（編）『論集日本語研究（一）現代編』206-230. 東京：明治書院
- (2000) 『国語引用構文の研究』 大阪：和泉書院
- 南不二男 (1974) 『現代日本語の構造』 東京：大修館書店
- (1993) 『現代日本語文法の輪郭』 東京：大修館書店
- 宮崎和人 (2002) 「第4章 認識のモダリティ」 宮崎和人・安達太郎・野田春美・高梨信乃 『モダリティ』203-227. 東京：くろしお出版
- 森山卓郎 (1988) 『日本語動詞述語文の研究』 東京：明治書院
- 幸松英恵 (2007) 「時間的に逆行している推論に関する一考察」 『日本語文法』7(2): 120-136.

反対概念の非対称性の普遍性と限界 —日本語と中国語における〈上/下〉の非対称性を中心に—

徐蓮（北京日本学研究中心／お茶の水女子大学大学院）

Universality and Individuality of the Asymmetry of Opposing Notions: Comparison of the Asymmetry of "zyo" & "ge" in Chinese and Japanese

Xu Lian (Beijing Center for Japanese Studies/Ochanomizu University)

1. はじめに

〈上/下〉、〈大/小〉、〈深/浅〉などの反対概念は一对として捉えられ、対称をなしているはずだと思われるが、実際非対称現象がしばしば見られる。例えば、「*上仕事/下仕事」「大黒柱/*小黒柱」「深い森/*浅い森」など。同じ非対称性は中国語にもある。例えば、“历史上/*历史下”“大本营/*小本营”“深信/*浅信”など。反対概念の非対称現象は多くの言語で普遍的に存在していると言われている。それでは、一体、この非対称性がどの程度まで共通しているか。また、どの部分に文化的条件が付けられているのか。本稿では、日本語と中国語における静態意味素〈上/下〉を含めた表現^①の非対称性を例として、認知意味論の視点から、「BCCWJ2009 モニター版」（日本語）と「CCL コーパス」（中国語）に基づき標本調査と統計分析を行う。その上に、日本語と中国語における〈上/下〉概念の非対称現象を量的に記述し、その意味上と使用率上の共通点と相違点を検討する。〈上/下〉の基本義と派生義の非対称構造を明らかにする上で、非対称現象の限界を求める。さらに、その背後にある認知的モチベーションを探る。

2. 先行研究と残された課題

中国語の上下の非対称現象に関する研究がたくさんある。意味論や語彙研究の面からの考察（呂叔湘 1965, 赵元任 1968, 张燕春 1994, 范洁清 2002, 白丽芳 2006, 蒋华 2006 等）もあり、語用論と認知意味論からの考察（周统权 2003, 李文莉 2004, 张玥 2007, 唐秋艳 2007, 蒋宁 2007 等）もあり、通時的な研究（吴峰 2007）もある。一方、日本語では、上下の非対称性に対する研究はそれほど盛んに行われていないようである。瀬戸賢一（1995）、村木新次郎（1999）、徐蓮（2009）などそれに触れた研究が少数あるが、深い研究がなされていないようである。

反対表現の非対称現象は多くの言語で普遍的に存在しているという仮説はまだ多言語での検証が必要である。また、先行研究は主に質的研究である。しかし、非対称現象は単なる意味項目の非対称だけではなく、使用率上の差異も非対称の一種であるから、量的研究が必要になる。本稿の課題は二つある。

(1) 日本語と中国語における静態意味素〈上/下〉の非対称現象を量的に記述し、その意味上と使用率上の共通点と相違点を検討する。

(2) 非対称現象の認知的モチベーションを探り、反対概念の非対称性の普遍性と限界を求める。

①同形の紛れを避けるには、中国語の表現を“”、日本語の表現を「」、両言語の同形表現を〈 〉に入れることにする。静態意味素〈上/下〉とは動態意味を表さない形態素〈上〉〈下〉を含めた表現のことである。表記が〈上/下〉の単純語と“上面”“下水道”などの複合語が含まれている。“上升”“上擦る”などは動詞であるが、〈上〉が動態意味ではないので、これも研究範囲内にある。静態と動態の判定基準について、日本語は動詞の語尾を持っているかどうか基準としている。中国語は《汉语大辞典》の品詞判定基準を参照した。

3. 語彙構成上の非対称

〈上/下〉は語彙を構成する面では対称を成していない。《汉语大词典》と『広辞苑』における静態意味素〈上/下〉を含めた語彙全体を考察した結果、形式上あるいは意味上の非対称によって次のように分類できる。

表 1 語彙構成上の対称/非対称^②

種類	言語	例	意味項目数 (%)
(1) 形式上も意味上も対称	日	上姿 (直衣を着た姿) / 下姿 (冠も袍もつけない姿)	366 (26)
	中	上月 (先月) / 下月 (来月)	194 (16)
(2) 意味上は対称、形式上は非対称	日	上掛け (下染めをした後で、他の染料でその上を染めること) / 下染め, 下搔き (本染の前に準備として染めておくこと)	48 (4)
	中	上调 (抜擢する) / 下放 (幹部を下部に下ろす)	98 (8)
(3) 形式上は対称、意味上は非対称	日	上足 (上席の門弟) / 下足 (脱いだ履物; その履物の番をする人)	258 (18)
	中	上文 (前文) / 下文 (後文)	124 (10)
(4) 形式上も意味上も非対称	日	母上 (母親) / ? ? / 下仕事 (下準備としての仕事)	730 (52)
	中	上校 (大佐) / ? ? / 下脚料 (材料の切れ端)	809 (66)

本稿では、意味上の非対称、つまり (3) (4) を研究対象にする。比率から見れば、非対称現象は決して少なくなく、中国語では意味項目総数の 76%、日本語では 70% を占めている。一方、日本語の (3) 類の数は中国語の約 2 倍もあり、(4) 類は中国語より 10% も少ない。

4. 意味上の非対称

4.1 プロトタイプ意味の非対称

幼児の言語習得に関する研究と空間用語の歴史的発達に関する研究によれば、空間的方位を表す〈上・下〉はプロトタイプ意味である。(徐蓮, 2009: 23-24) 意味がプロトタイプを起点に、イメージ・スキーマを通じて拡張する。スキーマが TR (trajector)、LM (landmark) と PATH の三つの部分からなる。静態関係を示す場合、PATH はゼロになる。「BCCWJ2009 モニター版」と「CCL コーパス」から〈上/下〉を含めた例文を抽出し、無作為抽出で日本語と中国語の〈上/下〉の例文を 500 例ずつ取り出し、2000 例のミニコーパスを作った。ミニコーパスの例文を分析した結果、プロトタイプの〈上/下〉の静態スキーマは次の五種類にまとめる。

【非接触類】 TR は LM より高い/低い、LM と接触しない

- (A1) 頭 (LM) の上で一声高く鳥 (TR) が啼いて、平八さんは顔を上げました。
- (A2) 在赤道附近，南北回归线之间，太阳 (TR) 终年在头顶 (LM) 上移来移去。
- (A3) まひろの家の桤 (LM) の下で、後藤課長も含めて皆でお花見をする。(TR は「皆」)
- (A4) 得骗且骗，总不能真让管大爷把我 (TR) 赶出门睡在桥 (LM) 底下吧。

【接触類】 TR は LM より高い/低い、LM と接触している

- (B1) 見渡す限り机 (LM) の上、コピーマシンのわき、厚い判例集 (TR) がうず高く、乱雑に積み上げてある。
- (B2) 然后又是一杯浓茶，一杯牛奶，一块蛋糕 (TR)，整整齐齐地放在桌 (LM) 上，食物的香味弥漫了整个房间，一切跟往常一样。
- (B3) 毛布 (TR) は布団 (LM) の上にかけていますか？ 下にかけていますか？

②最後の欄では、《汉语大词典》と『広辞苑』において、各種類の意味項目の数と〈上/下〉の意味項目総数 (日本語 1402、中国語 1225) で占める比率が示されている。「？」は対称している語がないことを意味する。

(B4) 老人被搀扶走的时候，又掏出了 10 元钱 (TR)，塞在我的 枕头 (LM) 下面。

【部分類】 TR は LM の一部であり、LM の高い/低いほうにある

(C1) 不動明王独特の牙がなく、並びのよい 上歯 (TR) で下唇をかみしめ、朱のはげた木目が筋肉どろろにあらわれて美しい。(LM は「歯」)

(C2) 傣族男子上身 (TR) 穿短衫，下着长裤，冷天披毛毡，多用白布或青布包头。(LM は“身”)

(C3) これで表示される画面 (LM) の右下あたりに「1 枚あたりのページ数」っての (TR) があるので「2 ページ」にすればいいです。

(C4) 在收看电视节目时，在屏幕 (LM) 下方有时会显示 一条重要新闻或现场比赛简报，或通知等文字信息 (TR)。

【融合類】 TR は LM と融合している

(D1) その考えは白い布 (LM) の 上 に広がる インクのしみ (TR) のように俺の胸の中にどす黒く広がっていった。

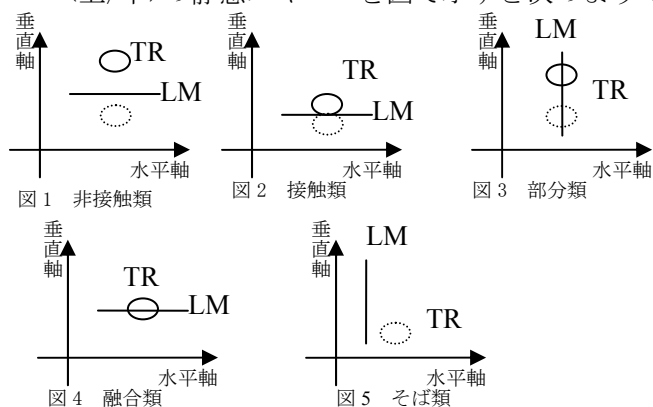
(D2) 谢军在宣纸 (LM) 上写了“安”“静”两个大字 (TR)，并开始和教练摆棋。

【そば類】 TR は LM のそばにあり、LM より低い。

(E1) 丘 (LM) の 下 は一帯のなだらかな 峡谷 (TR) で人家も少く、郊外の空気は甘く爽やかで新緑の匂いがふたりを包みこんでいた。

(E2) 由于尸体 (TR) 就放在脚 (LM) 下，看来再多问下去不太合适。

＜上/下＞の静態スキーマを図で示すと次のようである。



この五種類のスキーマのミニコーパスでの分布状況は下の図のようである。

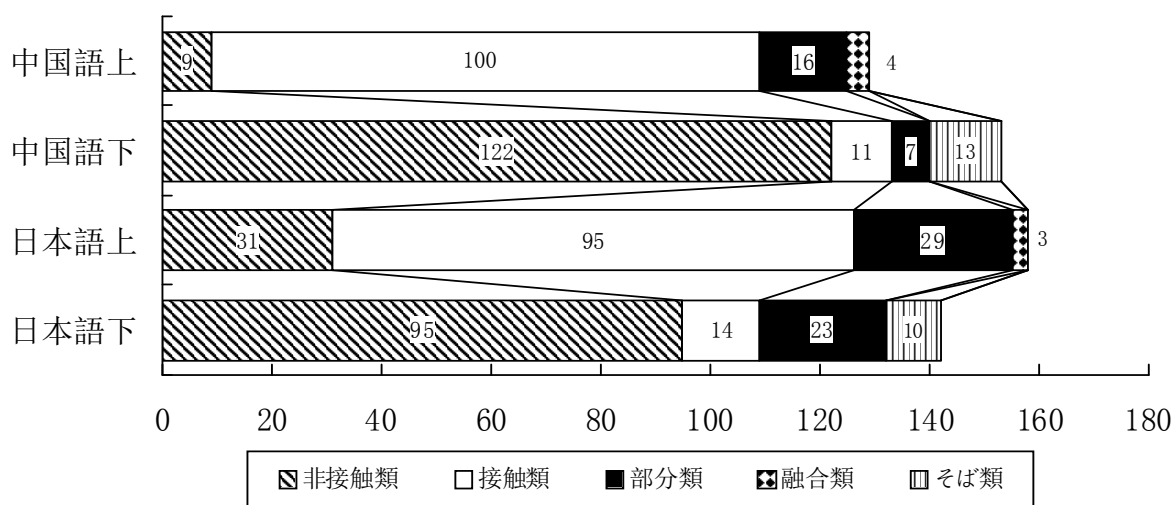


図 6 スキーマ分布図

プロトタイプ意味について、上の図から次の非対称性が見られる。

- ①非接触類スキーマでは、日本語においても中国語においても<上>の使用率が<下>よりずっと低い。中国語では上下の比率は1:14であり、日本語の1:3より著しい。
- ②接触類スキーマでは、日本語においても中国語においても<上>の使用率が<下>よりずっと高い。中国語では上下の比率は9:1であり、日本語の7:1よりやや著しい。
- ③部分類スキーマでは、日本語においても中国語においても<上>の使用率が<下>よりやや高い。中国語では上下の比率は約2:1であり、日本語の1.3:1よりやや著しい。
- ④融合類スキーマは日本語でも中国語でも少数ながら<上>の例文はあるが、<下>は見られない。
- ⑤そば類スキーマは日本語でも中国語でも<下>しかなく、<上>は使われていないようである。

4.2 拡張意味の非対称

<上/下>はプロトタイプ意味から、メタファーやメトニミーで他の意味に拡張してきている。《汉语大词典》、『広辞苑』とミニコーパスによって静態意味素<上/下>の主な拡張意味を次の表にまとめた。「→」はその上の意味からの再拡張を示している。括弧内の数はミニコーパスでの例文数である。

表2 拡張意味の分布表

拡張	目標域	意味	中国語		日本語	
			上	下	上	下
メ タ フ ァ ー	数量	多いことは上、少ないことは下	两杯以上 (16)	四十岁上下 (2)	百万以上 (16)	十人以下 (5)
	順序	前は上、後は下	上述 (18)	下文 (32)	上巻 (6)	以下の話 (8)
	社会地位	高いことは上、低いことは下	上司 (16)	下级 (11)	お父上 (27)	下位 (90)
		→影響されることは下		導引下 (8)		監督下 (17)
	品質	いいは上、悪いは下	后来居上 (4)	思想有高下 (1)	上機嫌 (27)	成績は彼の下だ (18)
	内外	表面は上、内部は下	风行水上	瘪下去的嘴 (1)	シャツの上にセーター (5)	下着 (8)
		→添付は上、基礎は下	在…的基础上 (16)	下意识 (1)	安い上にうまい (38)	下級生 (2)
メ ト ニ ミ ー		屋根は上、屋内は下	上覆	上漏下湿		
		頭は上			上傾き	
		湖沼は下		高高下下		
		足は下		下痿		
		陰部は下		下身		
		あごは下		丰下		
		排泄物は下				下の世話
		整体メトニミー		殿下		陛下(3)
文 法 化	所		脸上带着笑 (128)		路上 (4)	
	→方面		建筑史上 (67)		仕事上 (27)	

通用と 仮借	“上” = “常”、 下半身の衣類	進上			
	“下” = “假”、 仮借		下借		

<上/下>は主としてメタファーを通して意味を拡張してきている。垂直方位に対する認知で、数量、順序、社会地位、品質と内外関係が理解されている。その拡張意味はほとんど非対称的である。

- ①数量ドメインでは、日本語においても中国語においても<上>の使用率は<下>よりずっと高い。
- ②順序ドメインでは、両言語とも<下>が<上>を上回る傾向が見られる。
- ③社会地位ドメインでは、日本語では<下>の使用率が高いのに対し、中国語では<上>の使用率が高い。両言語とも「影響される」という意味の<下>が派生していて、その対称の<上>がない。
- ④品質ドメインでは、両言語とも<上>の使用率は<下>より高い。
- ⑤内外ドメインでは、<下>の使用率は<上>より高い。再拡張の「添付/基礎」の<上>の使用率は<下>より高い。

意味拡張のもう一つの主要手段はメトニミーである。メトニミーとは隣接あるいは関連している認知ドメインにおいて、顕著の物事で他の物事を取りかえることである。例えば、「上傾き」の「上」は位置で頭自身を指している。メトニミーによる<上/下>の拡張意味は主に古代語に使われているため、現代語のミニコーパスに例文がない。それらの拡張意味はほとんど非対称的である。

<上>は文法化され、場所を指すようになる。さらに抽象的な場所、つまり「～方面」を表す<上>も生まれる。このような使い方は非対称的である。また、使用率からみると、「上」の文法化の用法は「上」よりがずっと頻繁に使われている。

古代中国語では、発音の似ている漢字はよく仮借し合っている。「上」は“常”、“下”は“假”と通用していた。この意味の“上/下”も非対称的である。

5. 非対称現象の認知的モチベーション

5.1 非対称の生理的・心理的・物理的基礎

認知言語学によれば、人間は自分自身から世界を認知している。空間概念の認知も例外ではない。立つとき、人間の主な感覚器官は上半身にあり、上の空間がより感知しやすい。Ghent(1960,1961)は幼児が物の下の部分より、上の部分にもっと注意を払っていることを明らかにしている。(沈家煊 1999 : 181) 王祥榮 (2000)も母語習得の面から〈下〉より〈上〉のほうが認知しやすいことを立証している。<上>が<下>より認知しやすく、しかももっと広く使われていることは人間の生理構造によるものである。したがって、部分類の<上>の使用率は<下>よりずっと高いわけである。

<上/下>は垂直空間ドメインから内外ドメインに写像される際、表面は<上>であり、内部は<下>である。<下>の使用率は<上>より高い。認知心理学では、人間の体が容器である。空間はこの容器で「内」と「外」に分けられている。Clark (1974) と Clark & Brownell (1975) の心理学実験では、「内」を認知する時間は「外」より少ないことを明らかにしている。Clark (1973)によると、幼児が空間関係を認知する際、最も基本的な規則は、Bは容器だったら、AはBの「内」にあることである。(沈家煊 1999 : 180-181) したがって、「内」(<下>に対応)は「外」(<上>に対応)より基礎的であるため、もっと理解しやすく、もっと広く使われている。

言葉の意味というものは人間自身が存在している物理環境と切っても切れない関わりがある。そば類の〈下〉が対立した〈上〉がないのは、TRがLMの横の上方にあることは地球の重力に逆らうことである。LMと接触しない限り、このような空間関係が難しい。日常経験では稀であるゆえ、両言語において、〈上〉がないわけである。

5.2 顕現性 (Sense Salience) の影響

人間の認知は往々最も目立った、最も理解しやすい、記憶しやすい部分に注意している。言語においても、顕著の部分は表現の重点となる。ゲシュタルト心理学の実験によれば、この顕現性原則 (principle of prominence) は人間認知の重要原則の一つである。(赵艳芳 2001 : 99)

Clark(1974)と Carpe-nter & Just(1975)の研究では、顕著な参照物がない場合、人間は〈下〉ではなく〈上〉で相対方位を表す傾向がある。(沈家煊 1999 : 181) 融合類スキーマでは、LM の〈上〉と〈下〉という方位の対立が無くなり、その中和した所はより顕著な〈上〉に占められている。融合類の〈上〉が文法化して、場所を表すようになる。さらに、「～方面」という抽象的な場所も表すようになるが、〈下〉にはこの拡張がない。このような非対称はすべて顕現性原則に影響された結果である。一方、“上”の文法化程度が「上」より高いのは、“上”の場所を表す機能の一部は日本語においては、格助詞「に」に担われているからである。

LM は TR の方位を定める参照物である。その顕著な属性は〈上/下〉の非対称に影響している。接触類スキーマでは〈上〉の使用率が〈下〉より高く、非接触類スキーマでは〈上〉の使用率が〈下〉より低いのはこのためである。ミニコーパスでこの二つのスキーマの LM が次の表にまとめられる。

表3 接触類・非接触類スキーマの LM^③

スキーマ	上/下	言語	例
接 触 類	上	日	机/卓/テーブル*, 地*, 砂, (寝) 台*, 芝生/草, 畳, 石段…
		中	地*, 草席, 沙发, 椅子, 床*, 桌*, 座位…
	下	日	尻, 脚, 苔, 蒲団…
		中	脚, 高粱, 铁蹄…
非 接 触 類	上	日	頭, 市, 皿, 村, 門, 街, 土塀, 丘, 彼, お寺…
		中	头, 井台儿, 栏杆, 挡风玻璃, 人…
	下	日	地*, 木, 軒, ランプ/洋燈, 机*, ベッド*, 屋根, 帽…
		中	地*, 桌*, 矿井, 床*, 树, 屋檐, 灯, 木棚…

上の表から分かるように、接触類の LM は例文の文脈では「面」の特徴がある。それに対し、非接触類の LM は「点」と見なされ、〈上/下〉の方位はその「点」から広がり出る空間である。このような空間様態は人間の認知で最もよく見られる表れは傘あるいは屋根である。周知のように、「面」の重要な機能は支えることであり、〈上〉はその機能方位である。それに対し、傘と屋根の機能は遮ることであり、〈下〉はその機能方位である。幼児の空間認知に対する考察 (王祥荣, 2000) によれば、物の機能方位への認知は実際の方位に優先している。人間の認知においては、物の機能方位はいつも顕著であるので、LM の機能と一致している方位はより理解しやすく、また広く使われている。これが接触類と非接触類の〈上/下〉の非対称現象の認知的モチベーションである。上の調査によれば、日本語と中国語にはこの同じ非対称傾向があるが、中国語の非対称はもっと著しい。これは中国語ではもっと物の機能方位が重要視されているからである。

数量ドメインと品質ドメインでは、〈上〉の使用率が〈下〉に上回るのは、多いこと、いいことが目立つことに原因がある。それに対し、順序ドメインでは、人間が未来のことにもっと関心を寄せているため、〈下〉が目立ち、使用率も高いわけである。

5.3 不完全なマッピング

マッピングがそもそも数学概念であり、両集団の間の対応関係のことである。この対応関係で、一つの集団の成員が全員別の集団で対応している成員がある。(Fauconnier 1997 : 1) マッピングの概念は認知言語学に取り入れられている。メタファーもメトニミーも両ド

③*印のある LM は接触類にも非接触類にもある。これらの LM には「支える」と「遮る」と二種類の機能を同時に有しているからである。文脈によって機能が違い、スキーマも違う。

メインの間、あるいは一つのドメインの内部でのマッピングである。(李福印,2008:150-151)しかし、場合によって、マッピングしている成員もあれば、していない成員もある。また成員それぞれはマッピングの類型や目標ドメインが異なることもある。これは不完全なマッピングである。そのため、目標ドメインの成員が無くなったり、使用範囲が限られるようになったりして、非対称の原因となる。例えば、影響されることを表す<下>に対立した<上>がない。「添付」を表す<上>と「基礎」を表す<下>の使用率上の極度の非対称もそのためである。不完全なマッピングで、形式上対称しているが、意味上対称していない表現が生まれる。例えば「上足（上席の門弟）/下足（脱いだ履物；その履物の番をする人）」では、「上」は空間ドメインから品質ドメインへのメタファーのマッピングであるのに対し、「下」は物（脱いだ履物）から人（履物の番をする人）へのメトニミーのマッピングである。

5.4 認知ドメインの融合

Lakoff & Johnson は融合理論 (the theory of conflation) で幼児のメタファー連想メカニズムを説明している。メタファーには二つの段階があるとされている。第一段階は融合段階 (the conflation stage) である。二つのドメインは同じ活躍状態にあり、幼児は両ドメインを別々に体験するわけではなく、両ドメインを関連付けている。第二段階は分離段階 (the differentiation stage) である。両ドメインは始原ドメインと目標ドメインに分化される。例えば、褒められる時、幼児が抱き上げられ、怒られる時、幼児が大人の懷から地上に置かれる。したがって、幼児の認知では、良いことや嬉しいことのイメージは<上>のイメージに、悪いことや怒ることのイメージは<下>のイメージに重ねるようになる。それから、認知の発展につれて、この二つのドメインが分離され、良いことは<上>、悪いことは<下>というメタファーが生まれるわけである。(陳建生, 2008:141) 融合理論によると、<上>はプラスの意味素と、<下>はマイナスの意味素と複合語を作る。例えば、“上司”の“司”はつかさどることであり、社会地位が高いので、<上>のイメージと一致している。「下劣」の「劣」はおとっていることであり、品質が悪いので、<下>のイメージと一致している。このような複合語に対称している語（“*下司”“*上劣”）を作るのはどの言語においても難しい。

5.5 メトニミーと通用仮借

メトニミーによる拡張意味はほとんど非対称的であることはメトニミーの性質に関わる。<上/下>という物の方位でその物自身を表す時、取り替わられた物自身（屋根、頭、足、陰部など）は体あるいは部屋で唯一であり、上下の対称がないため、その拡張意味が非対称であることも当たり前である。

通用仮借は発音の近いことを基にしているので、通用している漢字の間には意味上の関りがない。<上/下>の対称は意味を基にしているので、このような拡張意味は対称していることはあり得ない。

5.6 社会文化的な要因

人間の認知は二つの源がある。一つは人間共通の身体・物質経験である。もう一つはある社会の文化的経験である。(張敏, 1998:99) 前者は認知の通言語的な共通性につながり、後者は認知の民族個性につながる。上述の認知的モチベーションは<上/下>の非対称の日中共通点の原因である。その相違点、つまり特有な非対称の背後には必ず社会文化的なモチベーションがある。

たとえば、明治維新以前、皇居は京都にあるので、京都あたりは「上方」と呼ばれている。しかし、この呼び方と対立しているのは「下方」ではなく、「江戸」である。それは、天皇のいる京都と幕府のある江戸が対峙しているからである。「上方」は空間ドメインから社会地位ドメインへのマッピングで際立っているが、「下方」は失敗した。同時に、社会文化の影響で「江戸」が「上方」の反対語となった。一連の派生語が生まれた。例えば、

「上方唄-江戸唄」「上方絵-江戸絵」「上方歌舞伎-江戸歌舞伎」「上方狂言-江戸狂言」「上方語-江戸語」「上方浄瑠璃-江戸浄瑠璃」「上方文学-江戸文学」など。仏教に関わっている「上座」から、「上座石」「上座部」など一連の派生語が生じた。同じく社会地位ドメインへの拡張であるが、「下座」は歌舞伎の専門用語であるため、「下座唄」「下座音楽」「下座台」「下座付」「下座帳」などの術語は「上座」の派生語の反対語になれない。文化に根を下している表現は複雑で、単純な空間的反対語で表せないわけである。社会文化的要因はある程度で非対称現象を促している。

上述から分かるように、社会地位ドメインでは、日本語と中国語では異なる非対称の傾向が見られることは目立つ。「上」の使用率は「下」より低く、“上”の使用率は“下”より高い。ミニコーパスで詳しく考察した結果、「下」の高い使用率は「下女」「廊下」「城下」などの表現に関わっていることが分かる。これらの表現は中国語に訳すと、“女仆”“走廊”“城市”となり、<下>の意味素が無くなる。中国語の例文で封建等級関係を表す表現も少なくないが、“下”という意味素をできるだけ避ける傾向が見られる。宗教に関する用語（“上仙”、“上清”、“上真”、“上果”等）や封建等級に関する用語（“上公”、“上嗣”、“上冢”、“上宰”、“上老”、“上相”等）は対立している<下>がない。この非対称は地位の高い者が目立つという認知メカニズムに主要原因があるが、社会文化的な原因もある。“上仙”を例にとってみよう。道教では様々な神様を“九仙”にまとめる。地位の高低を順に、“上仙、高仙、大仙、玄仙、天仙、真仙、神仙、灵仙、至仙”である。（杨秋生 2006）これらの称号はいずれもプラスなイメージである。宗教では、神様には等級があるが、マイナスのイメージにつながる<下>で名づけるわけにはいかない。地位の高いほうを目立たせることで等級をつける中国語に対し、日本語では違った方策が用いられるようである。

6. 結び

本稿では、語彙構成、意味及び使用率の面から、日本語と中国語における静態意味素<上/下>の非対称現象を対照した結果、反対概念の非対称現象は普遍的に存在していることを立証した。その普遍性を支えるのは人間の共通している認知メカニズムである。一方、日本語と中国語のような影響し合ってきた言語でも、非対称現象の限界がはっきり見られる。それは認知メカニズムの民族性の表れである。その背後には社会文化的なモチベーションがある。本稿でまとめた反対概念の非対称現象とその認知的モチベーションが、通言語的に適用できるかどうかはまだ検証が必要であるが、少なくとも一部の非対称現象が人間共有の認知メカニズムに根をおろしていて、多分多くの言語で共通していることはいえるだろう。この課題に関する他言語での検証を期待している。

文献（主要なもののみを挙げる）

- [1] 陈建生. 2008. 认知词汇学概论[M]. 上海: 复旦大学出版社.
- [2] 李福印. 2008. 认知语言学概论[M]. 北京: 北京大学出版社.
- [3] 沈家煊. 1999. 不对称与标记论[M]. 南昌: 江西教育出版社.
- [4] 王祥荣. 2000. 儿童语言中的“上”、“下”类方位词[J]. 安徽师范大学学报(4):568-573.
- [5] 徐莲等. 2009. 言語行為の認知的・語用的研究[M]. 北京: 大众文艺出版社.
- [6] 杨秋生. 2006. 试论何氏九仙信仰的文化构成[J]. 泉州师范学院学报(1):54-62.
- [7] 张敏. 1998. 认知语言学与汉语名词短语[M]. 北京: 中国社会科学出版社.
- [8] 赵艳芳. 2001. 认知语言学概论[M]. 上海: 上海外语教育出版社.
- [9] Fauconnier, Gilles. 1997. *Mappings in Thought and Language* [M]. Cambridge: Cambridge University Press.
- [10] Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites* [M]. Stanford: Stanford University Press.

『現代日本語書き言葉均衡コーパス』を利用した 格助詞の正誤判定

大山浩美（奈良先端科学技術大学院大学大学院生、情報科学研究科）[†]

Automatic Detection of the Grammaticality of Japanese Case Particles using the "Balanced Corpus of Contemporary Written Japanese"

Hiromi Oyama (Dept. of Information Science, Nara Institute of Science and Technology)

1. はじめに

近年、データ蓄積技術や情報検索技術の発達により様々な言語資料を収録したコーパスが数多く作成され、国語教育・日本語教育・辞書編纂・自然言語処理研究など様々な分野での活用が期待されている。その中で第二言語学習者の作文を集めた「学習者コーパス」は、学生へのフィードバック、教材研究、教授法の見直し等に利用でき、第二言語教育に役立てることができる。しかし、新聞や書籍などのコーパスと異なり学習者コーパスには正用だけでなく誤用も含まれており、教育に還元できるような調査分析をするためには正誤判定をする必要がでてくる。そのような事情から日本語学習者コーパスにおいて自動的に正用文と誤用文を判定するための機械学習を用いた正誤判定実験を行った。特に日本語文において文の意味に影響を与えやすく、日本語第二言語学習者が間違いやすいと考えられる格助詞に焦点をあてた。

自動的に正誤判定するための方法として機械学習法を利用するが、その機械学習をするためには、“教師”となるべきモデルデータが必要である。従来の正誤判定研究ではルールを抽出したり学習したりするモデルデータに関して、その文体、分野の差はあまり検討されてこなかった。今回、『現代日本語書き言葉均衡コーパス』から新聞(以下PN)、ヤフーブログ(以下OY)、検定教科書(以下OT)など文体、分野の異なるテキストデータをモデルデータとし、実験に使用した。そうすることにより文体、分野の差が機械学習による正誤判定の性能結果に影響するかを検証した。本稿では、5つの異なるモデルデータを用いた機械学習による日本語格助詞正誤判定実験について述べる。

2. 第二言語学習者の作文における自動正誤判定に関する先行研究

第二言語学習者の作文における誤用文を自動的に正誤判定する研究は、英文におけるスペルミスの正誤判定(Mays et al., 1991, Wilcox-O’hearn et al. 2008)、可算性(数えられる名詞)か、不可算性(数えられない名詞)かに関する誤りを判定する研究(Brockett & Gamon 2006, Nagata et al. 2006)、前置詞の正誤判定に関する研究(Chodorow & Leacock 2000, Chodorow et al. 2007, De Felice & Pulman 2007, 2008, Tetreault & Chodorow 2008)などがある。さらに、誤用タイプに特に着目せず、正用文と誤用文とを分類する研究もある(Sun et al. 2007)。和文における正誤判定研究では、主に格助詞を対象としている(今枝他 2003, 南保他 2007, 鈴木他 2006a, 2006b)。

[†] hiromi-o@is.naist.jp

表 1 先行研究について

	スペルミスの正誤判定	可算不可算名詞の正誤判定	前置詞の正誤判定	すべての誤用タイプ	格助詞の正誤判定
主な論文	Mays et al., 1991; Wilcox-O’hearn et al., 2008	Brockett & Gamon., 2006; Nagata et al. 2006	Chodorow & Leacock, 2000; Chodorow et.al 2007; Tetreault & Chodorow 2008; De Felice & Pulman 2007, 2008	Sun et al.,2007	今枝他2003; 南保他2007; 鈴木他2006a, 2006b
対象言語	英語	英語	英語	英語	日本語

日本語の格助詞に関する正誤判定において今枝他（2003）では、まず格助詞が書き誤られるルールに基づいて処理を行い（例えば、“を”が同文中に二つ書かれていたり、または連続してたりする場合）、そこで検出または訂正できなかった誤りに対して格フレーム照合を行う。格フレーム照合とは、NTT語彙体系(1997)の構文体系にのせてある動詞の取り得る格助詞の情報を利用して、入力文の格助詞の使用と照らし合わせる方法であり、合わなければ誤りと判定される。今枝他（2003）は、その手法により71%の誤り検出率(F値)と39%(F値)の誤り訂正率を得ている。ただし、今枝他（2003）では、実験の際に対象とする日本語学習者の作文の中で受身・使役・敬語表現を含まず、重文・複文も除外している点が様々な学習者の誤用文においても幅広く利用できるかどうか問題がある。南保他（2007）は、文節ごとに取り出した格助詞周辺の情報を最大エントロピー法¹という方法を用いる際の素性としている。例えば、「私は、りんごもみかんも好きです。」という文において「みかんも」をターゲットの文節とした場合、その周辺の文節である「りんごも」や「好きです」が機械学習する際の情報として利用される。南保他（2007）は、以下の方法で84%の再現率と64%の適合率を獲得している(F値は72%)。鈴木他（2006a, 2006b）の研究も同様に格助詞の誤りを検出するものであるが、目的は機械翻訳により生成された文に誤りがないかを調べるための実験であるため、前者2つの研究とは少し性格が異なる。南保他(2007)の実験においても、比較的単純な誤り文だけを対象としており、テスト文も84文にとどまっている。南保他(2007)の実験は、それまでの実験よりその精度を向上させているが、様々な学習者の誤用文においても利用できるようにするためには更なる実験が必要だと思われる。

上記のいずれの研究において手法の精度については述べてあるが、正用ルールを抽出したり学習したりする際のデータの質については言及されていない。本稿では文体、分野の差による正誤判定の精度への影響を調べるため、異なるテキストデータを使用して正誤判定実験を行った。

¹ 最大エントロピー法とは、確率を基にした情報分析の一つの方法であり、事前にわかっている確率を利用し未知の確率を計算する手法のことである。

3. 格助詞「を」における正誤判定実験について

3. 1. 実験データについて

日本人の書いたモデルとなる日本文を集めたコーパスとして『現代日本語書き言葉均衡コーパス』を利用した。その日本語コーパスは、現代日本語の書籍全般を対象として、全体で1億語を超える規模のコーパスの構築を目的とし、集積されている。このコーパスには様々なジャンルのテキストデータが収録されている。

機械学習を用いた格助詞の正誤判定タスクでは、文体、分野の異なりによる性能の差はまだ明確にされていない。文体、分野という要因が機械学習を用いた格助詞の正誤検出タスクに影響を及ぼすかどうかを調査するため、『現代日本語書き言葉均衡コーパス』から検定教科書 (OT)、ヤフーブログ (OY)、さらに新聞 (PN) などの異なるジャンルのサブコーパスを実験に利用した。数による公平性をきすため一万、五万、十万と事例数を一定にして実験に使用した。

さらに国立国語研究所で収集された『日本語学習者による日本語作文と、その母語訳との対訳データベース (以下、「作文対訳DB」²)』も利用した。これらのデータの中から、日本語教師等による日本語学習者の作文の添削 (ただし313ファイルのみ) 文において、添削された文において添削前の文を誤用文、添削後の文を正用文として利用した (ALL-CRR)。事例数は、2,923である。さらに、「作文対訳DB」には日本語母語話者による日本語作文も収集されているが、それも正用文として使用した (JP-ALL)。事例数は、13,321である (表2 参照)。

表 2 実験データについて

テキストデータ名	検定教科書	ヤフー ブログ	新聞	国研日本人作文	国研修正データ
	OT	OY	PN	JP-ALL	ALL-CRR
事例数	10,000	10,000	10,000	2,923	13,321
	50,000	50,000	50,000		
	100,000	100,000	100,000		

3. 2. 実験手順について

機械学習実験の手順について以下に示す。

1. 格助詞「を」、「が」、「に」、「で」、「へ」、「と」、「から」、「より」を実験対象格助詞とする。それらのいずれかを含む周辺単語数1～3個 (N1, N2, N3) を一事例とし、その周辺単語の表層情報、品詞情報、係り受け情報を機械学習のための素性として取り出す。
2. 「を」を使用している事例であれば+1、そうでなければ-1のラベルをふる。
3. 5つの実験データにおいてそれぞれ訓練用データとテスト用データ (訓練用データ

²このデータベースは、平成11～12年度科学研究費補助金「日本語教育のためのアジア諸言語の対訳作文データの収集とコーパスの構築」研究代表者：前田 (宇佐美) 洋) ならびに、平成12年度科学研究費補助金 (研究成果公開促進費・データベース) 「日本語学習者による日本語作文とその母語訳との対訳コーパス」の助成を受けて作成されたものである。<http://www2.kokken.go.jp/eag/wiki.cgi?page=taiyakuDBr%2Ftop>

の十分の一)に分ける。

4. 機械学習法 SVMs(Support Vector Machines)³を利用して訓練データにおいて正しく「を」を使用している事例とそうではない事例とを学習する。
5. 機械学習により作成された正用文モデルの精度をテストデータにより評価する。
6. 正用文を判定したモデルを利用して誤用文の判定をする。

3. 3. 実験評価

実験の評価は、F値を用いた。F値とは適合率と再現率との調和平均であり、適合率や再現率は実験結果の性能を評価する。今回の実験における再現率とは、検索結果として対象である格助詞「を」を使用した事例全体の集合の中でどれだけ検索対象事例が検出できているかを表す。ここでは、以下のように定義した。

$$\text{再現率} = \frac{\text{正しく検出された事例数}}{\text{格助詞「を」を含む全ての正用事例数}} \times 100$$

適合率とは、システムが正しいと判断して検出した事例のうちどれだけ正解事例を含んでいるかを表した値である。

$$\text{適合率} = \frac{\text{正しく検出された事例数}}{\text{システムが正しいと判断した事例数}} \times 100$$

F値の計算式は、以下の式で計算される。

$$F\text{値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

4. 格助詞「を」における正誤判定実験結果

実験は格助詞「を」を対象とし、それぞれ異なる5つのコーパスに対して、事例数が増加した場合（一万、五万、十万）、ウィンドウサイズをN1, N2, N3と変えた場合と、SVMsの多項式カーネルが2次と3次の場合、合計11×6通りの実験を行った。

表3で示されているように、5種類のテキストデータの中で最も結果がよかったのは、検定教科書を集めたデータで、事例数は十万のセットであった⁴。ウィンドウサイズによる差異は、テキストデータの種類によって違いがあり、OTにおいてはN2が最もよい結果を示している。

³ <http://chasen.org/~taku/software/TinySVM/>

⁴ 多項式カーネルの2次と3次ではいずれの場合も2次の方が良い結果を示したので、ここでは2次の結果のみを示した。

表 3 実験結果表-「を」正用文判定結果-

	OT			OY			PN			JP-ALL	ALL-CRR
	itiman	goman	jyuman	itiman	goman	jyuman	itiman	goman	jyuman		
N1	74.17%	81.08%	83.05%	56.16%	67.82%	72.69%	58.88%	74.61%	78.55%	61.21%	83.70%
N2	79.14%	84.14%	85.87%	63.89%	72.49%	76.55%	67.55%	79.39%	82.11%	68.31%	85.28%
N3	78.59%	83.88%	85.82%	61.74%	70.96%	76.05%	66.15%	79.07%	82.23%	61.82%	84.77%

表 4 は、誤用判定率のグラフである。誤用判定率においても正用判定率と同様に OT データ (N2, jyuman) が最も高い値を示している。以上の実験により格助詞「を」において検定教科書を用いた機械学習が最もよい性能を示すことがわかった。従来の正誤判定研究では、テキストの種類は考慮の対象とならなかったが、テキストデータの文体、分野の差が機械学習の性能の差に影響することがわかった。

表 4 実験結果表-「を」誤用文判定結果-

	OT			OY			PN			JP-ALL	ALL-CRR
	itiman	goman	jyuman	itiman	goman	jyuman	itiman	goman	jyuman		
N1	88.83%	91.46%	92.07%	86.83%	89.74%	91.50%	86.57%	90.53%	91.47%	87.74%	90.11%
N2	90.05%	92.46%	92.99%	88.83%	91.12%	92.79%	88.25%	91.98%	92.70%	89.84%	91.00%
N3	89.84%	92.22%	92.94%	88.86%	91.04%	92.80%	88.34%	92.01%	92.78%	88.27%	90.53%

5. まとめ

本稿では、5つの異なるモデルデータを用いた機械学習による日本語格助詞正誤判定実験を行った。機械学習を行う場合、モデルとなるテキストデータが必要であるが、先行研究ではモデルとなるテキストデータの文体、分野の違いによる正誤判定結果の差異を考慮されていなかった。今回、『現代日本語書き言葉均衡コーパス』による様々な文体、分野の違いのあるテキストを実験データとして使用した。その結果、それらテキストの違いが判定性能に影響を及ぼすことがわかった。さらに、5つの異なるテキストデータの中で検定教科書のテキストデータが良い結果を示すことがわかった。

今回は格助詞正誤判定というタスクにはテキストデータの文体、分野の違いが性能に影響するということはわかったが、格助詞正誤判定手法に影響するその他の要因、意味情報の付与などについては検証していない。今後は、それらの要因についても検証するつもりである。

文献

今枝恒治・河合敦夫・石川裕司・永田亮・榎井文人(2003)「日本語学習者の作文における格助詞の誤り検出と訂正」情報処理学会研究報告 コンピュータと教育研究会報告, 39-46, 情報処理学会

- 鈴木久美・Toutanova, K (2006a) 「機械学習による日本語格助詞の予測」 第12回言語処理学会予稿集, 1119-1122, 言語処理学会
- 永田亮・若菜崇宏・河合敦夫・森広浩一郎・梶井文人・井須尚紀(2006) 「可算／不可算名詞の判定に基づいた英文の誤り検出」, *J89-D(8)*, 1777-1790, 電子情報通信学会論文誌
- 南保亮太・乙武北斗・荒木健治(2007). 「文節内の特徴を用いた日本語助詞誤りの自動検出・校正」 情報処理学会研究報告 自然言語処理研究報告, 107-112, 情報処理学会
- 池原悟・宮崎正弘・白井諭・横尾昭男・中岩浩巳・小倉健太郎・林良彦 (1997)「NTT語彙体系 Vol.1.」, 「意味体系」, 岩波書店
- Brockett, C., Dolan, W., & Gamon, M. (2006). Correcting ESL Errors Using Phrasal SMT Techniques. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association of Computational Linguistics*, 249-256.
- Chodorow, M., & Leacock, C. (2000). An Unsupervised Method for Detecting Grammatical Errors. *Proceedings of the 6th Applied Natural Language Processing Conference*, 140-147.
- Chodorow, M., Tetreault, J., & Han, N.-R. (2007). Detection of grammatical errors involving prepositions. *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*, 45-50.
- De Felice, R., & Pulman, S. (2007). Automatically acquiring models of prepositional use. *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*, 45-50.
- De Felice, R., & Pulman, S. (2008). A classifier-based approach to preposition and determiner error correction in L2 English. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, 169-176.
- Sun, G., Liu, X., Cong, G., Zhou, M., Xiong, Z., Lee, J., et al. (2007). Detecting Erroneous Sentences using Automatically Mined Sequential Patterns. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. 81-88.
- Suzuki, H., & Toutanova, K. (2006b). Learning to Predict Case Makers in Japanese. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, 1049-1056.
- Tetreault, J., & Chodorow, M. (2008). The Ups and Downs of Preposition Error Detection in ESL Writing. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*.
- Wilcox-O' Hearn, A., Hirst, G., & Budanitsky, A. (2008). Real-word spelling correction with trigrams: A reconsideration of the Mays, Damerau and Mercer model. *Proceedings of 9th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2008)*, 605-616.

「(さ) せていただく」と「いたす」の競合

東泉裕子（東京学芸大学留学生センター）

Rivalry between the Construction *-(a)se-te itadaku* and the Verb *itasu*

Yuko Higashiizumi (International Student Exchange Center, Tokyo Gakugei University)

1. はじめに

「(さ) せていただく」は、「基本的には、自分側が行うことを、ア) 相手側又は第三者の許可を受けて行い、イ) そのことで恩恵を受けるという事実や気持ちのある場合に使われる」(敬語の指針 p.40) 表現であるが、現代日本語においては使用範囲が広がってきており、許容度には個人差があることが指摘されている。また、この表現は主に謙譲語「いたす」の代わりに用いられているようである。

本研究では、『太陽コーパス』と『「現代日本語書き言葉均衡コーパス」モニター公開データ (2008 年度版)』を用いて 1890 年以降の「(さ) せていただく」と「いたす」の出現率の推移を調査する。

2. 先行研究

2. 1 敬語の分類

日本語の敬語は、大きく「素材敬語」と「対者敬語」とに分けることができる(辻村 1967)。前者は話題の人物に対する敬語、後者は聞き手に対する敬語である。また、『敬語の指針』では従来「謙譲語」と呼ばれていたものを「謙譲語Ⅰ」と「謙譲語Ⅱ」(丁寧語)とに分けている。表 1 は、主な敬語の分類法をまとめたものである。

表 1¹

辻村 (1967)	5 種類 (『敬語の指針』)		従来の分類
	尊敬語	「いらっしゃる・おっしゃる」型	尊敬語
素材敬語	謙譲語Ⅰ	「伺う・申し上げる」型	謙譲語
-----	謙譲語Ⅱ (丁寧語)	「参る・申す」型	
対者敬語	丁寧語	「です・ます」型	丁寧語
	美化語	「お酒・お料理」型	

2. 2 謙譲語と丁寧語

本研究の調査対象である「(さ) せていただく」と「いたす」が関係するのは「謙譲語Ⅰ」、「謙譲語Ⅱ」、「丁寧語」の 3 つである。これらの違いを『敬語の指針』に沿ってここで確認する。「謙譲語Ⅰ」は「<向かう先> (..... 相手側である場合も、第三者である場合もあ

¹ 表 1 の「5 種類」と「従来の分類」は『敬語の指針』p.13 に基づく。

る)に対する敬語」(敬語の指針 p.19)であるが、「謙譲語Ⅱ」は「<相手>に対する敬語」(敬語の指針 p.19)である。また、両者は「ます」との関係についても違いがある。前者は「ます」を伴わずに使うこともできるのに対し、後者は一般に「ます」を伴って使う(敬語の指針 p.19)。「謙譲語Ⅱ」と「丁寧語」の違いは次のとおりである。前者は「基本的には「自分側」のことを述べる場合に使い、特に「相手側」や「立てるべき人物」の行為については使えない」のに対し、後者は「「自分側」のことに限らず、広く様々な内容を述べるのに使える」(敬語の指針 pp.20-21)。

「いただく」は「謙譲語Ⅰ」であるが、「謙譲語Ⅰ」の基本的な働きに加えて、恩恵を受けるという意味も併せて表(敬語の指針 p.17)し、「(さ)せていただく」は「謙譲語Ⅰ」に分類される(菊地 1996, 1997a, 1997b など)。²

「いたす」は「謙譲語Ⅱ」であり、「お(ご)……いたす」は「「謙譲語Ⅰ」兼「謙譲語Ⅱ」の一般的な語形」(敬語の指針 p.28)に分類される。以上をまとめ、『敬語の指針』(pp.15-21、p.28)からの例を加えたものが表2である。

表2

	働き	該当語例	「ます」との関係
謙譲語Ⅰ	自分側から相手側又は第三者に向かう行為・ものごとなどについて、その向かう先の人物をたてて述べるもの	伺う、申し上げる、お目に掛かる、差し上げる、お届けする、御案内する、 <u>(さ)せていただく</u>	「ます」を伴わずに使うこともできる
「謙譲語Ⅰ」兼「謙譲語Ⅱ」		<u>お(ご)……いたす</u>	一般に「ます」を伴って使う ³
謙譲語Ⅱ	自分側の行為・ものごとなどを、話や文章の相手に対して丁寧に述べるもの	参る、申す、 <u>いたす</u> 、おる	
丁寧語	話や文章の相手に対して丁寧に述べるもの	です、ます	

² 菊地(1996, 1997a, 1997b など)では、「謙譲語Ⅰ」、「謙譲語Ⅱ」をそれぞれ「謙譲語A」、「謙譲語B」としている。

³ ただし、近藤(2008)は、謙譲語Ⅱの中には「ます」と接合しないものも多いと指摘している。

2. 3 「(さ) せていただく」と「いたす」

『敬語の指針』は「(さ) せていただく」の使い方について次のように解説している。

「(お・ご) …… (さ) せていただく」といった敬語の形式は、基本的には、自分側が行うことを、ア) 相手側又は第三者の許可を受けて行い、イ) そのことで恩恵を受けるという事実や気持ちのある場合に使われる。したがって、ア)、イ) の条件をどの程度満たすかによって、「発表させていただく」など、「…… (さ) せていただく」を用いた表現には、適切な場合と、あまり適切だとは言えない場合とがある (p.40)。

ア) の条件がない場合には、やや冗長な言い方になるため、「発表いたします。」の方が簡潔に感じられるようである (p.41)。

菊地 (1997a) は、「(さ) せていただく」の使い方には個人差や広がりがあること、恩恵や許可とはとらえられない場合にも使われるようになってきていることを指摘している。⁴そして、「(さ) せていただく」が「謙譲語Ⅰ」から「謙譲語Ⅱ」へと変わっていく可能性を指摘している。

また、「謙譲語Ⅱ」の「いたす」はサ変動詞に限られる上に連体修飾節内では使いにくいなどの制約があるのに対し、「謙譲語Ⅰ」の「(さ) せていただく」にはそのような制約がない (菊地 1996, 1997a, 1997b など)。「いたす」の代わりに、「(さ) せていただく」が使われやすいという面もありそうである。そこで、本研究は「(さ) せていただく」と「いたす」の競合について調査することにする。

3. 調査

3. 1 調査方法

『太陽コーパス』の「口語」と『「現代日本語書き言葉均衡コーパス」モニター公開データ (2008 年度版)』 (以下、『BCCWJ2008 モニター版』と略す) の「書籍」を用いて、1890 年以降の「(さ) せていただく」と「いたす」の出現数の推移を調査した。

全文検索システム『ひまわり』の「後文脈」機能を使い、「(さ) せていただく」については、「せていただ」「せて戴」「せて頂」をキーワードとして検索した。「いたす」については、「いたさ」「いたす」「いたせ」「いたそ」「いたして」「いたしま」「いたした」「いたし」とその漢字表記「致さ」「致す」「致せ」「致そ」「致して」「致しま」「致した」「致し」をキーワードとして検索した。(そして、検索結果から、例えば、「それにいたしても」、「～にいたしましては」、「どういたしまして」などの表現を取り除いた。)

⁴ 菊地 (1997a) を参考にして筆者が行ったアンケート調査の結果を本稿の末尾に「付録」として掲載する。

検索結果を 1890 年から出版年 10 年ごとに分けて集計したものを表 3 に示す。「(さ) せていただく」は『太陽コーパス』には 10 例しかなかったのに対し、『BCCWJ2008 モニター版』には 358 例あった。また、「いたす」と比較すると、「(さ) せていただく」は 1980 年代以降、出現率が増加していることがわかる。

表 3

		(さ) せていただく		いたす		合計	
		%	(出現数)	%	(出現数)		
太陽	1890 年代	0.0%	(0)	100.0%	(322)	100%	(322)
	1900 年代	0.6%	(2)	99.4%	(339)	100%	(341)
	1910 年代	3.8%	(5)	96.2%	(128)	100%	(133)
	1920 年代	0.8%	(3)	99.2%	(358)	100%	(361)
計		0.9%	(10)	99.1%	(1,147)	100%	(1,157)
BCCWJ	1970 年代	6.9%	(2)	93.1%	(27)	100%	(29)
	1980 年代	28.3%	(45)	71.7%	(114)	100%	(159)
	1990 年代	27.0%	(61)	73.0%	(165)	100%	(226)
	2000 年代	26.0%	(250)	74.0%	(710)	100%	(960)
計		26.1%	(358)	73.9%	(1,016)	100%	(1,374)
合計		14.5%	(368)	85.5%	(2,163)	100%	(2,531)

4. 考察

佐野 (2008a, 2008b) は「さ入れ言葉」に後続する要素の中では「させていただく」が支配的であり、「させていただく」という形が 1980 年代以降、爆発的に増加していることを検証している。本研究では「謙譲語 II」の「いたす」との競合という観点から「(さ) せていただく」を調査したが、同様の結果を得た。

さらに、「謙譲語 I」だった語が「謙譲語 II」になった例が歴史的にあるという (菊地 1996, 1997a, 1997b, 金水 2004, 2005, 森山 2003, 2006, Traugott & Dasher, 他)。同様に、「謙譲語 I」である「(さ) せていただく」は現代日本語において「謙譲語 II」の「いたす」と競合し、「謙譲語 II」へと拡大しつつあるものと思われる。

5. まとめ

本研究では、『太陽コーパス』と『BCCWJ2008 モニター版』を用いて 1890 年以降の「(さ) せていただく」と「いたす」の出現率の推移を調査し、前者が 1980 年代以降増加していることを報告した。『ひまわり』の「前文脈」機能を使い、「(さ) せていただく」と「いたす」に先行する要素をさらにきめ細かく調査する必要があるが、それは今後の課題としたい。

謝辞

本稿は、学習院大学歴史語用論ワークショップ（2009 年 3 月 11 日）、11th International Pragmatics Conference のワークショップ Historical Pragmatics: Socio-Cultural Motivations of Language Change（2009 年 7 月 14 日）、青山英語学研究会（Aoyama Lunchtime Linguistics）（2010 年 1 月 27 日）において発表した内容に加筆・修正を施したものです。ご助言をくださった方々に感謝申し上げます。

使用データ

国立国語研究所(2005). 『太陽コーパス』(国語研究所資料集 15), 博文館新社.
国立国語研究所(2008). 『現代日本語書き言葉均衡コーパス』モニター公開データ(2008 年度版), 国立国語研究所.

文献

- 菊地康人(1996). 『敬語再入門』丸善ライブラリー.
菊地康人(1997a). 「変わりゆく『させていただく』」月刊言語, 26:6 (1997 年 6 月号), pp.40–47.
菊地康人(1997b). 『敬語』講談社学術文庫.
金水敏(2004). 「日本語の敬語の歴史と文法化」月刊言語, 33:4 (2004 年 4 月号), pp.34–41.
金水敏(2005). 「日本語敬語の文法化と意味変化」日本語の研究, 1:2, pp.18–29.
近藤泰弘(2008). 「『敬語の指針』と敬語理論」日本語学, 27:7, pp.38–43.
佐野真一郎(2008a). 「『日本語話言葉コーパス』に現れる「さ入れ言葉」に関する数量的分析」言語研究, 133, pp.77–106.
佐野真一郎(2008b). 「国会会議録よるさ入れ言葉の分析」松田謙次郎（編）『国会会議録を使った日本語研究』, pp.159–184. ひつじ書房.
文化審議会(2007). 「敬語の指針」（平成 19 年 2 月 2 日文化審議会答申）(http://www.bunka.go.jp/kokugo_nihongo/bunkasingi/pdf/keigo_tousin.pdf).
森山由紀子(2003). 「謙譲語から見た敬語史、丁寧語から見た敬語史-「尊者定位」から「自己定位」へ-」菊地康人（編）『朝倉日本語講座 8 敬語』朝倉書店, pp.200–224.
森山由紀子(2006). 「日本語における対者敬語の成立-『古今和歌集』詞書にみる「ハベリ」文法化の過程-」語用論研究, 第 8 号, pp.93–108.
辻村敏樹(1967). 『現代の敬語』共文社.
Traugott, Elizabeth Closs, and Richard B. Dasher. (2002). *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

付録

	菊地（1997a） 社会人 103 名 適切 不適切		2009 年 学部生 64 名 適切 不適切	
ア （学生が教師に）すみませんが、先生の 本を使わせていただけないでし ょうか。	87%	10%	63%	38%
	(90)	(10)	(40)	(24)
イ （パーティーの出欠の返事で）出席さ せていただきます。	83%	13%	80%	20%
	(86)	(13)	(51)	(13)
ウ （結婚式で媒酌人が）媒酌人として一 言ご挨拶させていただきます。	76%	20%	91%	9%
	(78)	(21)	(58)	(6)
エ （結婚式での、新郎の友人のスピー チ）新郎とは十年来のおつきあいをさ せていただいております。	56%	36%	70%	30%
	(58)	(37)	(45)	(19)
オ （結婚式での、新婦の友人のスピー チ）私は新婦と三年間一緒にテニスを させていただいた田中と申します。	39%	49%	33%	67%
	(40)	(50)	(21)	(43)
カ （セールスマンが客に）私どもはこの たび新製品を開発させていただきました して……。	17%	74%	36%	66%
	(17)	(74)	(23)	(42)
キ （近所の人に）私どもは、正月はハワ イで過ごさせていただきます。	5%	90%	19%	81%
	(5)	(93)	(12)	(52)

菊地（1997a）：

社会人 103 人（20 代 21 人、30 代 18 人、40 代 45 人、50 代以上 19 人）対象

2009 年（筆者が 2009/4/17 と 2009/12/11 に実施したもの）：

大学生 64 名（10 代、20 代）対象

招待発表

3月14日（日） 14:50～15:20

文芸創作における日本語コーパス活用

▶ 光原 百合

文芸創作における日本語コーパス活用

光原百合（尾道大学）

On the Use of BCCWJ for Creative Writing

Yuri Mitsuhashi (Onomichi University)

1. はじめに

筆者は自身が小説を執筆しており、また大学生に文芸創作を指導している。その立場から、文芸創作および文芸創作指導において日本語コーパスをどのように活用することが可能かの検討を行ったので、現段階での案をいくつか紹介したい。

尾道大学で担当している文芸創作演習の受講生にも、現代日本語書き言葉均衡コーパス（以後、BCCWJと表記）の検索デモンストレーション版の使い方を指導したうえで、自分たちが文章を書く上でどのような使い方が可能かを考えさせたので、その中からもいくつか紹介する。

また、その過程で「このような機能があればより便利」という声もあがったので、あわせて記しておきたい。

2. 新しい表現の使われ方について

たとえば十年ほど前の時代を舞台にした小説を書く場合に、「携帯電話」というものがどの程度普及していたかを知りたくなったとする。生産台数・普及台数などを調べても、それが人々の間でどれくらい定着した存在となっていたかはなかなか実感にくい。その実感をつかむためには、BCCWJが非常に有効である。小説において携帯電話が使われている場面がごく普通に描かれていれば、その時期にはすっかり定着していたと判断して間違いないだろう。

また、「携帯電話」「携帯」「ケータイ」とバリエーションが増えていく様子も如実にわかる。たとえば、書籍に限定して検索したところ、2001年には「ケータイ」という表記が11件ヒットしたが、それ以前にはまったく見当たらない（なおこういった検索によって、ある作家が新しい表現を積極的に取り入れるかどうかの文体研究の手がかりになるかもしれない）。

また学生からの提案だが、具体的な事物だけでなく、かつて流行していたが今は廃れた言葉が、当時はどのようなニュアンスで用いられていたかを調べる手がかりにもなるだろう。

3. 辞書的定義だけでは決められない表現の機微について

ある表現が、用法の変遷によって辞書的な意味から多少ずれるケースは多い。たとえば

最近話題になった例だが、「姑息」という表現はもともと一時逃れ、その場しのぎといった意味で使われていた。しかし最近ではそこに、「ずるい」「卑怯な」というニュアンスが加わっていることが多い印象がある。その印象が自分だけの思い込みでないかどうか、BCCWJを使って検討することができる。

また、筆者が実際に行った例だが、「吹聴」という言葉は辞書を見ると、よいこと悪いことに限らず広く喧伝することに使えるようである（このあたりは多少、辞書によって違っているが）。しかし、どちらかといえば「よくないこと、根拠のないこと、言わぬほうがよいことを触れ回る」という意味で使われている印象があったため、BCCWJで検索してみると、やはりそのような用例が多かった。

ただ文章を指導する際、学生が〈用例が多いか少ないかによってその表現を使うかどうか決めなければならない〉という方向に考えないよう留意したい。誤解を避けて違う表現を探すか、多くの人が誤解する危険性があったとしてもあえてその表現を使うか（あるいは、誤解の余地を減らすため使い方に工夫するという手段もあるだろう）というところこそ、書き手の個性が現れるからである。BCCWJはその判断材料を提供してくれる有効なツールであることを理解したうえで活用したい（させたい）と考えている。

4. 言葉のニュアンスの研究について

多くの用例を一度に見ることができるコーパスは、これまで個人の内省に頼るしかなかった「ニュアンス」の判断について重要なツールになるのは、すでにご存知の通りである。

文芸創作を学ぶ教え子たちの間でも、たとえばひらがな・カタカナ・漢字で表記を使い分けた場合（「めがね」「メガネ」「眼鏡」等）文体上どのような違いがあるかを研究するのに便利であるという声が複数あがっていた。教える立場としても、これは有効な授業課題にできると考えているところである。

やや違う話題になるが、役割語の研究ですでになされているように「ある表現をどういうキャラクターが使っているか」の検索が簡単にできる点も、小説の執筆にとってはありがたい。

5. 今後に望む機能について

以下のことが可能になればさらに役立つのでは……

- ・ 該当作品の初出年の記載（現在、該当の書籍の出版年は記載されている）
- ・ ひらがなで検索して漢字表記もピックアップ（あるいはその逆も）
- ・ 検索結果を作家ごとにピックアップ
- ・ 作品のジャンル分け

ポスター発表

3月14日（日） 15:30～17:30

カタカナ表記のゆれと意味 ―WWW調査との比較―

▶吉田 充良

日本語教育から見た均衡コーパス

▶李 在鎬、秋本 瞳

「Yahoo! 知恵袋」を利用した「の」の曖昧性解消の試み

▶西口 純代

「～がる」と感情を表す「～む」動詞の意味・用法

▶韓 金柱

嫌悪を表す「いや」と「きらい」について

▶加藤 恵梨

話し言葉、書き言葉、そしてオンライン言語をめぐって

―日本語全体像を捉える試みへのパイロット・リサーチ―

▶西村 由起子

動作の方向を表す「―テクル」に対する要求条件の解明

▶石山 友之

BCCWJを用いた「語の文体」の位置づけ

▶井上 次夫

複合辞トイウにみる文法化の諸相

▶金城 由美子

BCCWJにより選定した日本語教育語彙の検証

―経営学を専攻する日本語学習者を対象にして―

▶寺嶋 弘道

カタカナ表記のゆれと意味—WWW調査との比較—

吉田 充良（日本大学文理学部国文学科）

Variation in Katakana-Spelling and the Meaning —Comparison with the WWW-based Research—

Mitsuyoshi Yoshida (Nihon University)

1. 要旨

松崎寛(1993.9)は、辞典類のカタカナ表記語を調査し、263 語についてその音韻の表記が外来語音寄りか、日本語音寄りかを示している。この先行研究が示すカタカナ表記のゆれ（ある拍が外来語音寄りか日本語音寄りか）が、使用場面によって意味の弁別機能を果たしているのではないかと考え、BCCWJ を使用し、カタカナ表記のゆれと意味の関係について調査を行った。

調査は、松崎寛(1993.9)で取り上げられている語の中から、十分に用例が確保された 23 語を調査語として選び出し、それぞれの語の外来語音寄り表記と日本語音寄り表記を「ひまわり」で検索し、検索結果の 100 件について用例確認を行った。

本稿は、コーパスを使用して表記のゆれと意味の関係を調べることができるということ、また、同時に行った WWW を日本語コーパスとした調査の結果と本調査の結果とを比較し、コーパスの違いによる調査結果の違いを示すことで、BCCWJ のコーパスとしての特徴の一側面を提示することを目指す。

2. はじめに

外来語のカタカナ表記に関する問題について、武部良明(1991.7)は「外来語の表記について問題になるのは、どのような仮名をどのように用いるかということである。」としている。外来語の表記に関する研究は、内閣告示『外来語の表記』に代表されるような、借用語がもつ外来音を日本語においてどのように表記すべきか、という問題を盛んに取り扱ってきた（安永実(1991.7)、武部良明(1991.7)）。外来語の表記に関するその他の研究としては、外来語表記上の特徴について論じたもの（石野博史(1984.5)）、そうした表記の時間経過による変化を論じたもの（宮島・高木(1974.12)）、語ごとによる外来語表記の定着度の違いを論じたもの（松崎寛(1993.9)）、外来語表記と発音に関して論じたもの（赤坂和雄(1972.10)）が挙げられる。

しかし、こうしてみると、カタカナ表記と意味について同時に論じられたものはないように思われる。そこで、本稿では、「コーパスに基づく日本語研究」（田野村忠温(2009.1)）によってこうした音韻や語種などの過去の調査とは違った視点から、「外来語のカタカナ表記のゆれが持つ役割」に関する極めてささやかな資料を示すことにする。

3. 調査概要

3.1. 調査目的

森田良行他編(1989)は、「表記がなぜゆれているのか、は（中略）個人の意識にかかわっていて、簡単には解明できない問題である」としているが、本稿では、個人の意識だけではない、カタカナ表記のゆれが持つ“機能”について、その実態の記述を目的とする。従来の研究が明らかにしてきた「外来語をどのように表記すべきか」という問題から離れ、「表記のゆれが何を引き起こしているか」ということについて明らかにしていきたい。

3.2. 調査語

本調査の調査語は、松崎寛(1993.9)で扱われる 263 語に従った。その 236 語の中から本調査に適するものを選び出すため、「定着度」と「ゆれの有無」の観点から、次の 2 つの基準を設定した。まず、松崎寛(1993.9)の 263 語を WWW で検索エンジン『Yahoo!Japan』（以下 Yahoo!）を用いて、「外来語音寄り表記・日本語音寄り表記」（松崎寛(1993.9)）をそれぞれ検索し、「どちらの表記でも 100,000 件以上のヒットがあるかどうか」を第一の基準とした。次にそうした語にゆれがあることを確認するため、こうして得られたヒット件数をもとに、「双方表記のヒット件数の合計を 1（100%）とした時、一方の割合が 15-85%のもの」を第二の基準とした。

この 2 つの基準をクリアして選び出した語について、さらに検索結果の先頭 20 件の用例を確認する予備調査を経て、本調査の調査語を選出した。

また、アンティック、アンチック、アンティークなど、ゆれとされる表記が 3 つ以上あるものについては、松崎寛(1993.9)が提示する表記に従った。

3.3. 調査方法

BCCWJ 調査では、BCCWJ モニター版 2009 に付属する全文検索システム「ひまわり」を使用して BCCWJ 内の用例を得た。調査語を「ひまわり」の“検索語”に直接入れ、そうして得られた語について、それぞれの文脈で“その語”がどのような意味で使用されているかを確認する。用例確認数は、得られた用例の先頭から 100 件を目安とした。

語によって傾向が異なるため、「5 例を超えたら一つのカテゴリ」というような基準は敢えて設定しなかったが、得られた用例を適宜意味ごとの用例数に従ってカテゴリライズし、作成されたカテゴリごとに用例数を集計した。そして、もう一方の表記でも、同様の作業を繰り返した後、双方表記（外来語音寄り表記と日本語音寄り表記とのペア）のカテゴリの内訳を比較し、「表記のゆれが“何らかの機能”を持つかどうか」を確認する。

WWW 調査での用例の検索は、検索エンジン「Yahoo!」で、一方の表記を指定したとき、もう一方の表記が検索結果に表示されないようにするため、キーワードをダブルクォートで囲む「フレーズ検索」と、キーワードの直前にマイナスをつける「マイナス検索」を組み合わせた、「ゆれを含まない検索」を採用した。たとえば、キーワード「ゼリー」の場

合、「ゼリー」「ジェリー」のようにする。これで、論理的には WWW 上の「ゼリー」の用例だけを検索結果として返してくることになる。

BCCWJ 調査同様、こうして得られた検索結果の用例の中で、それぞれの文脈で“その語”がどのような意味で使用されているかを確認する。用例確認数は、検索結果の先頭 100 件とした。

4. 調査結果

ここからは具体的な調査結果を示していく。

4.1. ジェリーとゼリー

“ジェリー”の表記と“ゼリー”の表記のペアは、今回調査した語の中で、「表記のゆれが意味の弁別機能を持つ」特徴を顕著に示した例である。“弁別”といっても、一方の表記が一つの意味しか持たず、表記のゆれが意味をはっきりと二分するというわけではない。もし、はっきりと二分するということになれば、“表記のゆれ”ではなく、“語形の違い”という観点が浮かびあがってくることになるのではなかろうか。“ゆれ”の中でも、特にカテゴリの用例数の偏りから、“弁別”の特徴を読み取った。

次の表は、BCCWJ 調査の「jelly」のカテゴリとその用例数の一覧である。表の一段目左の列から、「外来語音寄り表記」、「外来語音寄り表記の検索結果の総数」、「日本語音寄り表記」、「日本語音寄り表記の検索結果の総数」の順に並べた。また、二段目左の列から「外来語音寄り表記のカテゴリ」、「外来語音寄り表記のカテゴリの用例数」、「日本語音寄り表記のカテゴリ」、「日本語音寄り表記のカテゴリの用例数」の順に並べた。

表 1 jelly [BCCWJ]

●ジェリー	106	●ゼリー	153
人名	40	菓子	79
アニメタイトル	3	潤滑ゼリー	13
菓子	1	ゴミ	8
ゴミ	56		
[人名]	[24]		
[オレンジゼリー]	[18]		
[ランゼリー]	[8]		
[その他]	[6]		

“ジェリー”の表記で、中心的なカテゴリは「人名」なのに対して、“ゼリー”の表記では「菓子」のカテゴリがその中心である。したがって、“ジェリー”の表記では「人名」を、“ゼリー”の表記では「菓子」をそれぞれ選択しやすいといえることから、“ジェリー”と“ゼリー”の表記間には、意味の弁別機能が確認される。

次に WWW 調査の「jelly」のカテゴリとその用例数を示す。

表 2 jelly 【WWW】

●ジェリー	15,000,000	●ゼリー	36,500,000
人名	27	菓子	91
アニメタイトル	23	アニメタイトル	1
菓子	15	その他	8
その他	25		
ゴミ	10		

BCCWJ 調査の「jelly」同様、“ジェリー”の表記では「人名」のカテゴリが、“ゼリー”の表記では「菓子」のカテゴリがそれぞれの中心をなしている。“ジェリー”の「アニメタイトル」は、アニメ「トムとジェリー」のキャラクター「ジェリー」のことで、これは、「人名」に準ずる。

また、WWW の“ジェリー”では、「菓子」の用例が 15 例と BCCWJ 調査よりも多いが、これは、得られた用例の総数の違いと考える。WWW の“ゼリー”の「菓子」の用例数が 91 例なのに比べると WWW の“ジェリー”の「菓子」の 15 例は圧倒的に少ない。

したがって、WWW 調査でも BCCWJ 調査同様に、“ジェリー”と“ゼリー”の表記間に意味の弁別機能が確認された。

BCCWJ、WWW どちらにも共通していえることだが、“ゼリー”は、「オレンジゼリー」や「コーヒーゼリー」のような使用で、一般名詞としての傾向が強いのに対し、人名やアニメタイトルというのは固有名詞としての傾向が強い。データからは一般名詞の使われ方ではゆれが起きており、固有名詞の傾向の使われ方では表記の選択をしているということになる。

また、BCCWJ の用例の特徴として、「レンジェリー」や「アンジェリーナ」などの、“ジェリー”だけで一語をなさない「ゴミ」の多いことが指摘される。そもそも、BCCWJ の「ひまわり」と WWW の「Yahoo!」では、検索のシステムが異なるため、比べることは意味をなさないが、先頭 100 件に登場するものとして記述したときの特徴であると考えた。

以上が「jelly」の用例のコーパスごとの比較である。「jelly」については、「意味」の面で表記間の弁別機能がよく表れている。しかし、意味の弁別機能というのは、すなわち、「語形」の違いであるとも考えられる。むしろこの例ではそちらのほうが正しいかもしれない。内省に照らし合わせても、「ジェリーとゼリー」、は“ゆれ”というよりも“別語”と意識される。この点から、「表記のゆれと語形の違いのあいまいさ」が認識できる。

4.2. ウィルソンとウイルソン

「jelly」同様「Wilson」も「表記のゆれによる弁別機能」が確認された例であるが、“ジェリー”と“ゼリー”とは違った傾向を示した。また、コーパスの違いで結果が異なる例として着目される。まず、BCCWJ 調査の「Wilson」のカテゴリとその用例数の一覧を示す。

表 3 Wilson 【BCCWJ】

●ウィルソン	65	●ウイルソン	10
人名	59	人名	7

ウィルソン株	2	ウイルソン通り	2
スポーツブランド	1	ウイルソン株	1
その他	3		

どちらの表記でも、「人名」への使用が用例の中心になっていることから、“ウィルソン”と“ウイルソン”の表記間に意味の弁別機能は確認されなかった。

ただし、“ウイルソン”の表記の用例の総数が10例というのは比較するには少ない。これでは用例がやや不十分で“ウィルソン”の表記と比較した結果としては曖昧になってしまうように思われる。

また、これはBCCWJの特徴として、「Wilson」についてBCCWJにとられるテキストでは“ウィルソン”の表記が選択されやすいという傾向が読み取れる。

次にWWW調査の「Wilson」のカテゴリとその用例数を示す。

表 4 Wilson【WWW】

●ウィルソン	8,440,000	●ウイルソン	3,140,000
人名	37	スポーツブランド	75
スポーツブランド	35	企業名	7
企業名	13	ウイルソン病	7
ウィルソン株	6	人名	4
その他	8	その他	6
ゴミ	1	ゴミ	1

BCCWJではどちらの表記でも「人名」がカテゴリの中心をなしていたのが、WWWでは“ウィルソン”の表記でのみその傾向が表れる。

また、「Wilson」では、スポーツブランド「Wilson」の“ウィルソン”と“ウイルソン”のそれぞれの表記の用例数の差に着目した。「スポーツブランド」のカテゴリは、“ウィルソン”の表記で35例、“ウイルソン”の表記で75例と、それぞれの表記においてその占める割合が大きく異なる。スポーツブランド「Wilson」の社名は、トマス・E・ウイルソン社長に由来するが、「Wilson」社のホームページが検索結果の先頭に登場するのは、“ウイルソン”の表記であった。一方の“ウィルソン”の表記では、同じ「スポーツブランド」としての使用でも、スポーツ用品店のホームページで「Wilson」商品を取り扱っているということを示すものが多い。このことから、「Wilson」が社名として正式に採用しているのは、「ウイルソン」の表記で、このことが“ウイルソン”の表記でスポーツブランドの使用数が多いことにつながっている可能性が高い。

スポーツブランド「ウイルソン」は、「人名」からの転用だが、スポーツブランドとしての使用では「人名」の意味は損なわれていると考えられることから、“ウィルソン”と“ウイルソン”の表記間には意味の弁別機能が確認されるといえる。

「ジェリーとゼリー」では表記のゆれの弁別機能が「語形の違い」を示すと指摘したが、「ウィルソンとウイルソン」は「語形の違い」というよりも「表記のゆれ」という認識がある。「Wilson」についていえば、それは、「スポーツブランド」が、「人名」から「企業名」への意味の転用を経て、さらに「スポーツブランド」への転用をしているからである。「Wilson」という企業が「企業名」として意図的に「ウイルソン」の表記を選択したことで、

企業という団体は「ウイルソン」の表記を習慣的に選択し、それが影響力を持って「スポーツブランドの時はウイルソンの表記」という“ルール”ができているように見て取れる。しかし、それは企業という団体が習慣的に行っている“表記ルール”で、個人がブログなどでスポーツブランドの「Wilson」を書き表す際にはそのルールが適用されず、“ウィルソン”の表記でも「スポーツブランド」の用例が現れるものと考えられる。

表記のゆれによる意味の弁別機能には、「語形の違い」を表すもの、企業などの団体の中にある“表記ルール”の存在を表すものの、二通りが考えられる。

4.3. ヒンドゥーとヒンズー

「新聞記事の日本語には、語彙・文体・表記の面で新聞社特有の統制が加えられている」（前川喜久雄(2007.10)）。それは、新聞社という企業として出す文書にゆれがないようにするというひとつの言語行動である。同じ企業内でも、その文書を書く人が違えば、表記も異なってくる。そうした「ゆれ」を防ぐために、企業として統一したルールを定め、それに従って書きすすめられている。しかし、そうした企業外のもの、たとえばブログなど個人の書くものについては統一したルールが実際に見られないのであるから、結果として、新聞社などが出す文書の中の表記にはゆれはなく（ないはずで）、個人の書いたものについてはゆれが発生するということになる。そうしたことが言語使用の実態に表れた例が WWW 調査の「Hindu」である。

Yahoo!の検索結果の見出しの最後に、下図の丸で示す「- MSN 産経ニュース」のような表示のあるものを「ニュース記事」としてカウントした。

転倒事故で123人死亡 インド北部・ヒンズー教寺院 - MSN 産経ニュース

... ヒンズー教寺院. 2008.8.3 22:52. 3日、インド北部ヒマチャルプラデシュ州にあるヒンズー教のナйна ... インド北部ヒマチャルプラデシュ州にあるヒンズー教のナйна・デビ寺院で3日、多数の信者が転倒する事故が発生し、123人が死亡した。 ...

sankei.jp.msn.com/world/asia/080803/asi0808032253007-n1.htm キャッシュ

図 1 ニュース記事例

まず最初に、BCCWJ 調査の「Hindu」のカテゴリとその用例数の一覧を示す。

表 5 Hindu 【BCCWJ】

●ヒンドゥー	64	●ヒンズー	75
ヒンドゥー教関連	64	ヒンズー教関連	75

BCCWJ 調査では、どちらの表記でもすべてが「ヒンドゥー教」や「ヒンドゥー教徒」、「ヒンドゥー美術」などの用例をまとめた「ヒンドゥー教関連」の意味での使用だったことから、“ヒンドゥー”と“ヒンズー”の表記間に意味の弁別機能は認められなかった。

次に WWW 調査の「Hindu」のカテゴリとその用例数の一覧を示す。

表 6 Hindu 【WWW】

●ヒンドゥー	1,920,000	●ヒンズー	1,760,000
ヒンドゥー教関連	95	ヒンドゥー教関連	97
その他	5	[ニュース記事]	[23]
		その他	3

BCCWJ 調査と同様に、どちらの表記でも「ヒンドゥー教関連」のカテゴリの用例数がそのほとんどを占めていることから、“ヒンドゥー”と“ヒンズー”の表記間に意味の弁別機能は確認されないといえる。

しかし、検索結果の見出しの最後に「MSN 産経ニュース」などの表示がでてくるもの、ニュース記事、は WWW 調査の“ヒンズー”の表記にしか現れなかった。なお、“ヒンズー”に現れたニュース記事は 23 例で、おもに記事の見出しのものが引っ掛かった。“ヒンドゥー”の表記ではニュース場面への使用が特定されるものがなかったことから、“ヒンドゥー”と“ヒンズー”の表記間には、使用場面の弁別機能がある可能性が指摘される。

WWW 調査では、ニュース記事での使用による表記の緩やかな使用場面の弁別機能が確認されたが、BCCWJ 調査ではそうした特徴は確認されなかった。BCCWJ で、「Hindu」については用例数もそれなりにあるが、このようなデータの現れ方では、「表記がゆれている」という事実はわかっていても、「その先で起こっていること」は見えない。WWW で観察された「ニュース場面では“ヒンズー”の表記が選択されやすい」という特徴は、「新聞社などは表記に関する独自の規範を持っている」という前提があり、それが言語事実として現れているものが観察されたのに対し、もし、BCCWJ だけで調査を行った場合には、そうした言語事実を見落としてしまうことになる。こうした事実を踏まえると、WWW が持つデータの「多様さ」というものは評価されるべきものであると考える。

4.4. ウェートとウエート

外来語が同音意義を形成するケースはいくつかある。英語の「steal」と「steel」は英語では語形（表記）が異なることによって意味も異なるが、日本語ではどちらも「スティール（もしくはスチール）」と表記され、語形が等しいなかに「steal」と「steel」の意味を持つことになる。本調査においても、「weight」は一語で複数意味をもち、それがどのような使われ方をしているかが観察された。

まずは、BCCWJ 調査の「weight」のカテゴリとその用例数の一覧を示す。

表 7 weight 【BCCWJ】

●ウェート	79	●ウエート	172
比重・比率	8	比重・比率	87
重心	1	ゴミ	13
ゴミ	70	[クウェート]	2
[クウェート]	[66]	[ウエートレス]	11
[ウエートレス]	[4]		

“ウェート”の表記では「クウェート」、「ウエートレス」などのゴミが数多く見られたが、それらを除けばどちらの表記でも中心的な意味は、「比重・比率」である。ゴミを除くと用例数の差がかなりあるが、それでも“ウェート”と“ウエート”の表記間には意味の弁別機能は見られないといえる。

次に WWW 調査の「Hindu」のカテゴリとその用例数の一覧を示す。

表 8 weight 【WWW】

●ウェート	873,000	●ウエート	928,000
体重	26	体重	36
率	22	率	18
おもり	16	株の用語	17
株の用語	9	ウエートリフティング	8
重心	7	おもり	7
おもさ	7	おもさ	7
人名	1	その他	3
ゴミ	12	参照例・ゴミ	4

WWW 調査では、「体重」、「率」、「おもり」、「株の用語」、「重心」、「おもさ」のカテゴリで用例数の多少はあるものの、すべての意味がどちらの表記でも現れることから、“ウェート”と“ウエート”の表記間に意味の弁別機能は確認されないといえる。

カテゴリ一覧から WWW の結果は BCCWJ の結果よりも、ウェート（ウエート）一語が幅広い意味の用例を示すことがわかる。この点は WWW と BCCWJ とのデータの「多様さの差」として指摘される。

5. BCCWJ と WWW の比較

日本語の語種の中で「外来語」の比率はそれほど高くない。宮島達夫(2007.10)の雑誌の延べ語数の語種比率のグラフによれば、「外来語」の語種の比率は雑誌九十種調査(1956)で 2.9%、雑誌 70 誌調査(1994)で 10.7%である。

その「外来語」のなかで、「ゆれ」の双方表記の一定量の用例を確保することは容易なことではない。BCCWJ 調査の調査語と WWW 調査の調査語のそれぞれの総数の違いからも明らかのように、BCCWJ の示す用例数では、本調査のように総数の少ない「外来語」のさらに「ゆれ」の表記で大量の用例を必要とする場合には、「まだまだデータ数が足りない」ということが結論づけられる。現段階で公開されているデータは BCCWJ の一部であるということで、今後、その全体が公開されたとき用例数については十分得られるようになる可能性はあるが、現段階では、用例数の観点から、本調査のような研究テーマでは WWW の方

が有用であるといえる。

「用例の内容」についても、やはり WWW のデータの方が「多様さ」を持ち、本研究テーマからも有用といえる。WWW の調査結果で得られた「ニュース場面」の特定は、新聞社などの独自の表記規制という事実を言語の使用実態から再確認されたこと、「weight」をめぐる意味の広がりについて BCCWJ では確認されなかった意味が WWW ではその用例が多数得られたことなどからも、WWW のデータの有用性が確認できる。

調査を通じて、BCCWJ と WWW との用例の内容には決定的な違いがあった。それは、BCCWJ の「jelly」に現れたゴミの多さから説明することが可能であると考えられる。BCCWJ では一度「ゴミ」が現れるとそれと同じ「ゴミ」が何度も続き、結果的に、全用例の半数以上を「ゴミ」が占めるということが起きる。それは、そのいくつかの「ゴミ」の用例が書籍なら書籍の同じ出典の同一文章内に繰り返し現れるもの一つ一つが検索結果の一件一件に相当しているからである。同一文章内では別々の用例だとしても「同義」の場合が多い。一方で、WWW の用例では、同一出典からの複数用例というのはあまりなく、ほとんどの用例が別々の文章にふっと現れたものなのである。つまり、BCCWJ では複数の用例に対して使用主体が一人なのに対し、WWW では用例一つ一つに対して使用主体が一人ということなのである。したがって、BCCWJ の各表記の用例数は、その使用主体の数とイコールではないということになる。表記のゆれが「個人の意識による」ものであれば、用例は「一用例一使用主体」であることが本調査では望ましい。したがって、使用主体を「1」と数えれば、実質的な BCCWJ の検索性数は限りなく少ないということになる。こうなると、いくら用例数が増えたとしても、カタカナ表記のゆれは確認されても、肝心の意味については確認されることはなくなってしまう。

従って、少なくとも、本稿で扱ったようなテーマの調査には WWW コーパスは非常に有効になるということが改めて結論づけられる。

6. おわりに

「外来語のカタカナ表記」は日本語の全体から見ればごく少数だが、表記のゆれのように大量の用例が必要になる本稿のような調査では、BCCWJ よりも WWW コーパスが適するといえる。しかし、「WWW コーパス」が真の日本語の姿を示しているかはまた別の問題である。検索方法の工夫でいくつかの問題は解決されるように思われるが、WWW をコーパスとして使用する場合は各語がもつ性格に注意しながら使用しなければならないこともここに確認された。こうした注意さえすれば、WWW が持つ言語データの大量性という強みを生かして言語調査への応用が可能であるといえる。

本稿では、カタカナ表記のゆれが、ゆれによって指し示す内容を変えたり、使用場面を変えたりするという言語事実を、多数の用例に基づいて明らかにした。語によってその傾向は異なるが、一つ一つの語を見ていくことで“日本語の多重層性”を改めて実感した。

ゆれによって意味を弁別するということは、突き詰めればそれは「語形の違い」という

ことになる。ゆれはひとつの意味に対して複数の表記が存在するものだが、語形の違いは、ひとつの意味に対してひとつの表記が割り当てられているものであるからだ。「表記のゆれ」と「語形の違い」とは“あいまいさ”で繋がれている。武部良明(1991.7)は「正書法」の観点から、内閣告示『外来語の表記』が「複数の表記を可能にしたことについては、問題が残る」としているが、本稿で示した「表記のゆれが持つ機能」を考慮に入れば、表記のゆれは「ゆれていて良い」ものと考えられる。

7. 付記

本稿は、2010年1月12日に日本大学文理学部国文学科に提出した卒業論文「カタカナ表記のゆれと意味」の内容を一部抜粋したものである。実際の卒業論文では、WWW 調査で 39 語、BCCWJ 調査で 22 語の調査結果を掲載している。

8. 参考文献

- 石野博史(1984.5)「カタカナ語のかなづかい」『日本語学』第3号第5巻
- 荻野綱男(2004.2)「各種検索エンジンの実態と特徴」『日本語学』第23巻第2号
- 荻野綱男(2006.7)「検索エンジン Google の使い方と WWW コーパスによる日本語研究」城生 佰太郎博士還暦記念論文集編集委員会編『城生 佰太郎博士還暦記念論文集』（東京堂出版）
- 荻野綱男(2007.7)「コーパスとしての WWW 検索の活用」『月刊言語』第36巻第7号
- 荻野綱男(2009.1)「コロケーション辞書」『国文学解釈と鑑賞』第74巻第1号
- 武部良明(1991.7)「内閣告示『外来語の表記』の考え方」『日本語学』第10巻第7号
- 田野村忠温(2008.6)「日本語研究の観点からのサーチエンジンの比較評価——Yahoo! と Google の比較を中心に——」『計量日本語学』第26巻第5号
- 田野村忠温(2009.1)「コーパスと文法研究」『国文学解釈と鑑賞』第74巻第1号
- 前川喜久雄(2007.10)「コーパス日本語学の可能性——大規模均衡コーパスがもたらすもの——」『日本語科学』第22号
- 宮島達夫・高木翠(1974.12)「外来語の表記の変化とゆれ」『計量国語学』第71号
- 森田良行他編(1989)『ケーススタディー日本語の語彙』（桜楓社）
- 安永実(1991.7)「内閣告示『外来語の表記』のできるまで」『日本語学』第10巻第7号

日本語教育から見た均衡コーパス

李 在鎬 (国際交流基金) [†]
秋本 瞳 (麗澤大学大学院生)

Balanced Corpus from the Viewpoint of Japanese Language Education

Jae-Ho LEE (Japan Foundation)

Hitomi AKIMOTO (Graduate School of Language Education, Reitaku University)

1. はじめに

コーパスは、言語使用例の集合であるため、言語の運用サンプルを得る上で最適なリソースである (大曾 2006:3)。このことは、日本語教育においてもコーパスが注目される一因になっており、教育用語彙選定の基礎資料としての利用をはじめ、様々な方向での活用方法が模索されている (cf. 砂川 (他)(2008), 森山(2009))。

さて、コーパスの日本語教育における活用の意義や妥当性を考えるにあたっては、そもそもの前提として「コーパスのデータが日本語学習者にどの程度、理解されうるのか」について明確にしておく必要がある。例えば、「日本語能力1級学習者はコーパスデータを何割、理解するのか」といった問題が想定されるが、これに対する客観的データはこれまでの研究では示されていない。こうした事項に対する明示化ができてこそ、日本語教育におけるコーパスの利用価値が正当に評価される。また、運用面においても、学習のどのような段階でどう導入されるべきかの指針が立てられるであろう。

本研究では、上記の課題に対して日本語能力試験 (Japanese Language Proficiency Test: 以下 JLPT) の出題基準と均衡コーパスを結びつける調査を行った。調査データとしては、JLPT 出題基準の語彙表¹と日本語の言語使用の総体とも言うべき二つの均衡コーパスを使用した。「日本語話し言葉コーパス (Corpus of Spontaneous Japanese: 以下 CSJ と呼ぶ)」(国立国語研究所)と「現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese: 以下 BCCWJ と呼ぶ)」(国立国語研究所)である。CSJ は話し言葉を代表するコーパスとして、BCCWJ は書き言葉を代表するコーパスとして位置づけ、調査を行った。調査方法としては、国際交流基金が開発した JLPT 語彙判定システムを用いて、コーパスデータをファイル単位で全文解析し、カバー率調査を行った。具体的には、コーパスのテキストデータを形態素解析したあと、JLPT 出題基準に照合し、各テキストに JLPT の語彙、すなわち日本語教育用の語彙がどの程度出現するのか、そして、どのレベルの語彙がどれだけ出現しているのかを調査した。調査の結果、CSJ では全体の 74%、BCCWJ では全体の 72%の語彙が JLPT の語彙であることが明らかになった。この結果を踏まえ、日本語教育における応用可能性を検討した。

[†] jhlee@gmail.com

¹ JLPT 出題基準の語彙表とは『日本語能力試験出題基準』(国際交流基金)として公開された資料で、1級から4級の日本語能力別に集計された語彙のリストである。1級語彙は異なりで7,800語、2級語彙は4,800語、3級語彙は1,500語、4級語彙は800語である。短単位相当の語彙のほかに、挨拶表現などの慣用句も多数含まれており、国内外における語彙教育の指針として活用されている。

2. データと方法

本研究において利用したデータおよび調査の手順について説明する。CSJは「その他」を除く4つのサブコーパスから先頭の50ファイルをサンプルとして決定した。BCCWJは2009年度のモニター版を利用し、CSJ同様に、先頭から50ファイルを選択し、調査を行った。ただし、国会会議録のみ1ファイルのサイズが大きいため、25ファイルを選択し、調査を行った。コーパスサイズを表1に示す。

表1 コーパスサイズの集計

コーパス	サブコーパス	ファイル数	文字数	文数	平均一文長
CSJ	学会発表	50	272,175	16,059	17.67
	朗読	50	22,161	2,078	11.00
	模擬講演	50	95,198	11,017	8.84
	対話	50	339,258	27,189	12.30
BCCWJ	書籍	50	267,162	9,080	34.09
	Yahoo!知恵袋	50	17,221	397	26.50
	白書	50	263,143	4,918	55.83
	国会会議録	25	1,649,019	17,197	83.17

表1のコーパスをファイル単位で解析した。解析は国際交流基金が保有する語彙チェックシステムを用いて行った。この語彙チェックシステムは、内部の形態素解析に基づいて入力テキストから語彙を取り出し、その結果をJLPTの出題基準にマッチングさせ、級別語彙を自動で集計する。なお、形態素解析はWATERSという商用解析エンジンで行っているが、40万語規模の解析用の辞書を持っている。また、解析用の辞書はJLPT用にカスタマイズされており、「初～」といった接頭辞から「お願いします」のような慣用句まで登録され、JLPTのために細かい部分までチューニングされている。

3. 結果

最初の調査として、JLPTの語彙がコーパス内の語彙のどの程度をカバーするのかを調べてみた。表2の結果となった。

表2 カバー率一覧表

コーパス	サブコーパス	JLPT 語彙の延べ頻度	JLPT 外の 語彙の延べ頻度	JLPT 語彙 のカバー率
CSJ	学会発表	43,620	19,403	69%
	朗読	5,224	690	88%
	模擬講演	19,340	5,741	77%
	対話	45,974	27,059	63%
BCCWJ	書籍	46,690	21,929	68%
	Yahoo!知恵袋	1,418	674	68%
	白書	58,073	20,646	74%
	国会会議録	295,480	80,833	78%

JLPT 語彙とは、各サブコーパス内で **JLPT 出題基準**としての語彙、すなわち日本語教育語彙である。一方、**JLPT 外の語彙**とは、**JLPT** の出題基準に含まれない語彙、日本語教育語彙ではない語彙である。**JLPT 語彙のカバー率**とは、全体の語彙における **JLPT 語彙**の比率である。たとえば、**CSJ** の学会発表コーパスは延べ 63,023 語で構成されているが、そのうち日本語教育用の語彙は 43,620 語含まれていることになり、そのカバー率は 69%ということになる。表 2 の結果を踏まえ、**JLPT 語彙**に対して、級別出現状況を集計してみた。図 1 の結果となった。

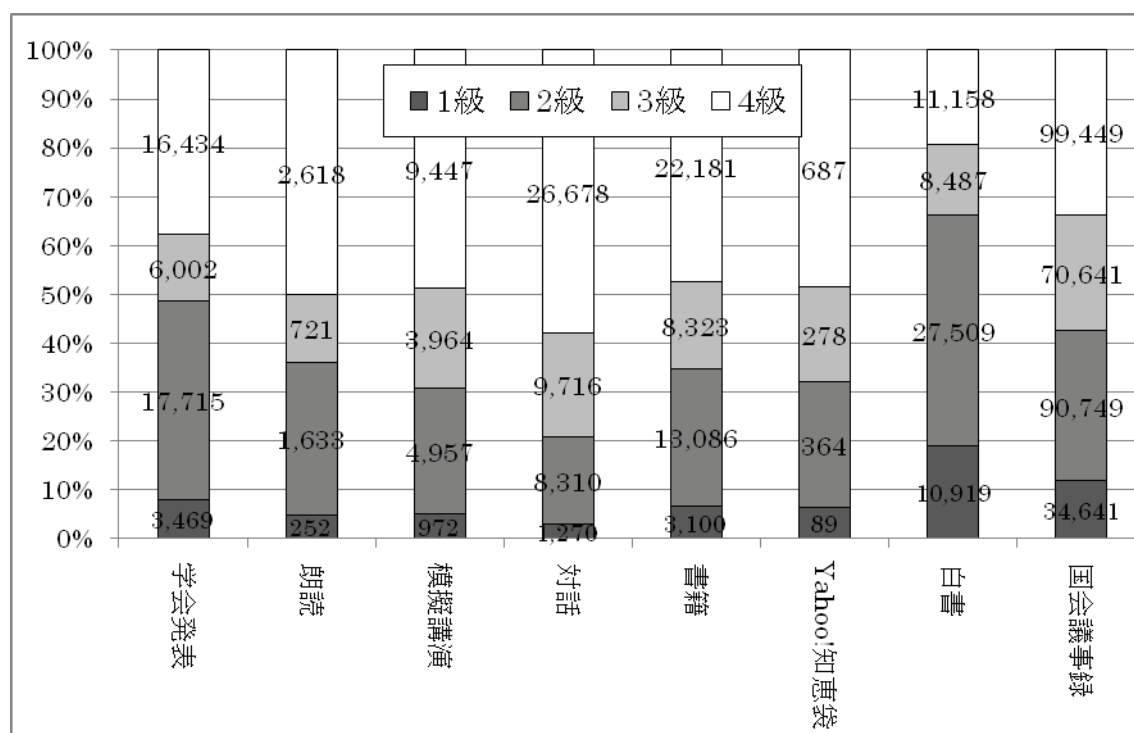


図 1 級別出現語彙の分布

表 2 および図 1 から以下の三点が明らかになった。

1. 話し言葉と書き言葉で平均 7 割のカバー率を示しているので、計算上は **JLPT** の 1 級合格者であれば、均衡コーパスの 7 割の語彙が理解できる。
2. 1 級語彙は、**JLPT** の項目数としてはもっとも多い半面、実際の使用としてはもっとも少ない。「白書」と「国会会議録」を除くいずれのサブコーパスでも使用率は 1 割以下と非常に少ない。
3. 下位級の中では、3 級語彙よりは 4 級語彙が多く、上位級の中には 1 級語彙よりは 2 級語彙が多い。ここから、4 級語彙が持つ基本語性が確認された反面、1 級語彙が持つ非基本語性（≒難解性）が示唆される。

4. 考察

表2の結果を受け、人手チェックの方法で JLPT 外の語彙に対して、詳細な調査を行った。表1のデータすべてに対して、人手チェックを行うことは、現実的でない判断し、サブコーパスから10ファイルを選び、語彙チェックシステムにかけ、JLPT 外と判定された1153語に対して「なぜ JLPT 外と判定されたのか」調べてみた。

表3 JLPT 外語彙の詳細

		未登録語彙	固有名詞	異表記	解析エラー	その他	合計
CSJ	度数	109	51	53	149	43	405
	%	26.9%	12.6%	13.1%	36.8%	10.6%	100.0%
BCCWJ	度数	378	98	65	198	8	748
	%	50.5%	13.1%	8.7%	26.5%	1.1%	100.0%
総計	度数	487	149	118	347	51	1153
	%	42.2%	12.9%	10.2%	30.1%	4.4%	100.0%

未登録語彙とは、純粹に JLPT 出題基準に登録されていない語彙を指す。固有名詞とは、場所や人名などの固有名に関わる語彙を指す。異表記とは、JLPT 出題基準が示す表記とは異なる表記をしているために、JLPT 外の語彙と判定されたものを指す。主として動詞類の仮名遣いに関して多く見られる。例えば、JLPT 出題基準では「手掛ける」になっているが、実際のテキストでは「手がける」になっているような例である。次に解析エラーとは、形態素解析システムが語彙の切り方を間違ったために、適切な判定ができなかったものを指す。最後に、その他とは、話し手の言い間違いなどが関わる例で、解析システムの問題ではないと思われる例である。

表3の結果に関連し、注目すべきは、書き言葉に関しては未登録の語彙が半分以上を占めている点である。具体例を示す。

- A) 外来語の例：アスパラガス，アリア，エコー，カスタラート，ギャラ，コミュニティ，シャンソン，スパン，スペクタクル，ソプラノ，チェンバロ，テノール，ハーブ，パトロン，ピンポイント，プロデューサー，レインジャー
- B) 漢語名詞類：韻律，烏帽子，雲高，王侯，果樹，歌曲，歌劇，花婿，荷役，海域，海況，海難，海氷，器楽，既設，機知，戯曲，宮廷，救難，教皇，橋梁，桑園，君主，粗飼料，短波，地霧，低地，低利，適時，配役，薄雲，畑作，微風，夫君，舞踏，保健，法面，防疫，幕間，未然，陽光，養蚕，養豚，裏作，老朽

以上の結果から、書き言葉に関しては、領域固有の語彙が多いため、日本語教育への応用に関しても慎重に行わなければならないことが示唆される。

次に、図1の級別の語彙の分布をめぐって、サブコーパスの分類を行ってみた。図1の

値を比率データに変換し、階層的クラスター分析を行った。クラスター間の距離は Ward 法を使用し、項目間の距離は平方ユークリッド距離で測定した。その結果、図 2 が得られた。

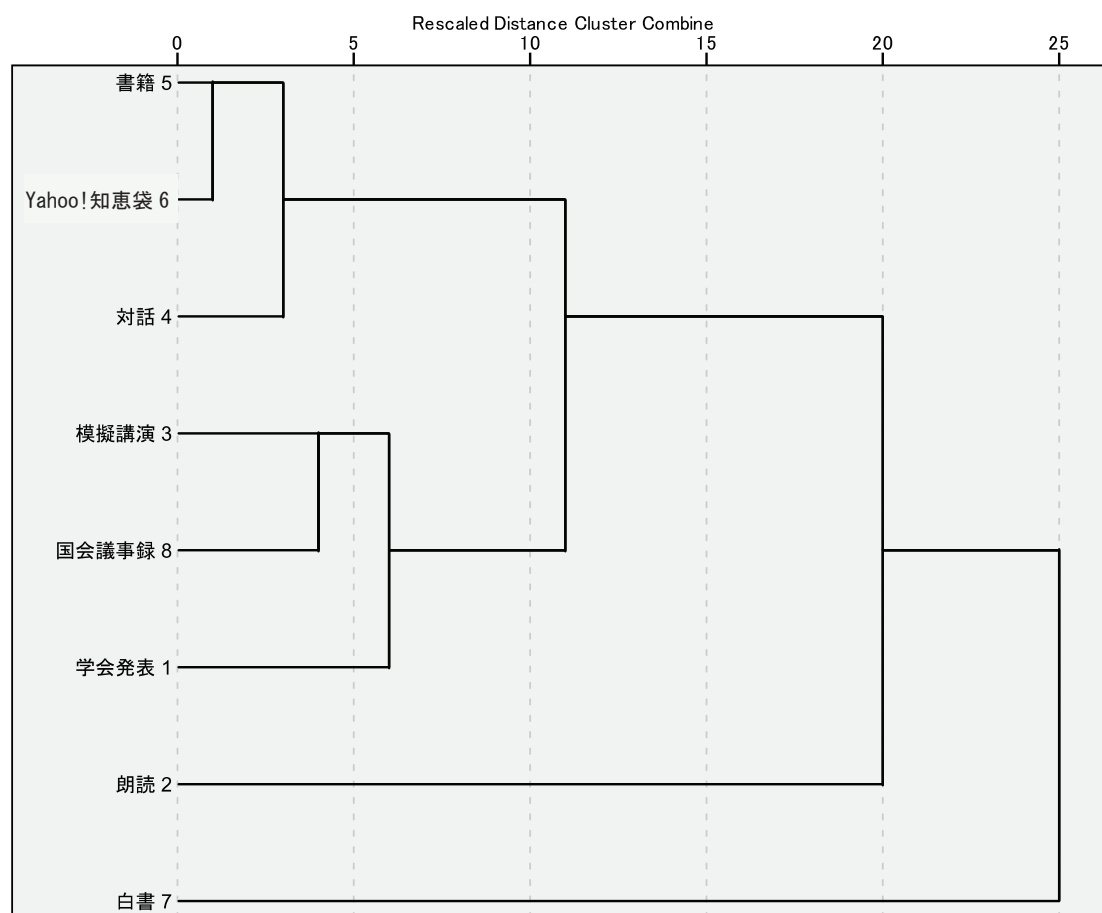


図2 コーパスのクラスター分析

グループ1の「書籍(BCCWJ)」、「Yahoo!知恵袋 (BCCWJ)」、「対話 (CSJ)」は、いずれも4級の語彙が全体の半分近く、3級語彙を入れた場合、全体の7割がカバーされるもので、比較的易しいテキストのグループとして理解することができる。一方、グループ2の「模擬講演 (CSJ)」、「国会議事録(BCCWJ)」、「学会発表 (CSJ)」は2級語彙が示す割合が大きいという特徴を持つ。このことから、比較的難しいテキストのグループとして理解することができる。一方、グループ3の「朗読 (CSJ)」とグループ4の「白書 (BCCWJ)」のコーパスはいずれからも切り離される結果となった。その理由として、「朗読(CSJ)」の場合、JLPT 語彙の分布パターンそのものは、グループ1とほぼ同一であるが、JLPT 外の語彙が1割弱となっており、非常に少ないという特徴がある。よってグループ1には統合されなかった。グループ4の「白書(BCCWJ)」は、JLPT 外の語彙が2割程度になっており、全体のカバー率は高いが、JLPT 語彙の分布を見た場合、1級、2級の語彙が全体の7割近くを占めており、他のテキストに比べて、異質であるということが示唆される。

5. 終わりに（日本語教育への応用）

本研究の調査結果から、以下のことが明らかになった。

- [1] 均衡コーパスにおける語彙の JLPT 出題基準のカバー率は平均 7 割程度である。
これは、L2 の標準的語彙研究が既知語彙 9 割を内容理解のベースラインと決めていることから考えた場合、決して高いとは言えない。
- [2] 書き言葉においては未登録語彙が多いが、これは領域固有の語彙が多いことに起因するもので、その教育的応用においては慎重でなければならない。
- [3] JLPT 語彙の分布からコーパスをグループ化した場合、「書籍(BCCWJ)」、「Yahoo! 知恵袋(BCCWJ)」、「対話(CSJ)」は 3 級までの語彙が主であり、比較的易しい語彙多く含まれているが、「模擬講演(CSJ)」、「学会発表(CSJ)」、「国会会議録(BCCWJ)」は 2 級までの語彙が主であり、比較的難しい語彙が多く含まれている。

文献

- 大曾美恵子(2006)「日本語コーパスと日本語教育」日本語教育, 130, pp. 3-10
砂川有里子(他)(2008)「代表性を有する書き言葉コーパスを利用した日本語教育研究」(特定領域研究「日本語コーパス」日本語教育班 平成 19 年度研究成果報告書)(課題番号: 18061001)
森山新(2009)「第 8 回日本語教育国際研究大会 言語学の分野を中心に」日本語教育, 143, pp.33-42

コーパス

- 「日本語話し言葉コーパス」(<http://www.kokken.go.jp/katsudo/seika/corpus/>)
「現代日本語書き言葉均衡コーパス」(モニター版)(http://www.kokken.go.jp/kotonoha/ex_8.html)

「Yahoo!知恵袋」を利用した「の」の曖昧性解消の試み

西口純代（徳島大学先端技術科学教育部国際連携教育開発センター）[†]

Disambiguating “No” by Yahoo! Chiebukuro

Sumiyo Nishiguchi (Center for International Cooperation in Engineering Education,
Graduate School of Advanced Technology and Science, The University of Tokushima)

1. 要旨

本研究では、2009 年度版の「コアデータ」の「Yahoo!知恵袋」に含まれている格助詞「の」の用例を ChaKi.NET 1.2β を利用して取り出し、「NP₁のNP₂」構文の意味を名詞間の意味関係に基づいて人手で分類した。その結果、NP₂ の語彙的意味の内在的特性、すなわち Pustejovsky (1995) の言うクオリア構造を NP₁ が修飾している (selective binding) 例が 28%を占めることがわかった (例「フジの連ドラ」)。さらに NP₂ を広義の関係名詞であると見なすと、NP₁ がその項となっている例が 24% (「胸の前」) ある。NP₂ の場所、時間、様態などの一時的特性を NP₁ が修飾する例が 7% (「オペラ座の怪人」) あることから、語彙意味構造に内在的特性のみならず、外延的要素を加える必要性を主張する。その他に、NP₂ がイベント項を持つ派生名詞で、NP₁ がイベントの動作主などの項を表している例が 14% (「社長の期待」)、NP₂ の形容詞的修飾の例が 12% (「アル中の俳優」)、NP₁ が数量詞である例が 4.8% (「ほとんどの家庭」) あった。

2. 研究の目的

「の」の意味の新たな分類に基づいてデータを分類し統計結果を得る。その結果に基づいて理論的な枠組みを修正する必要性を示す。

3. 調査の方法

3.1 手順

2009 年度版の「コアデータ」の「Yahoo!知恵袋」Data Base File (松本祐治氏提供, pc, 2009 年 12 月 12 日) に含まれている格助詞「の」の用例 3159 件を ChaKi.NET 1.2β を利用して取り出し、「NP₁のNP₂」構文の意味を名詞間の意味関係に基づいて人手で分類した。¹

3.2 分類法

I. Selective binding of qualia in NP₂

NP₂ の語彙的意味のクオリア構造の CONSTITUTIVE (part-whole 関係)、FORMAL (オントロジックのカテゴリー、形、色など)、TELIC (目的)、AGENTIVE (起源) 役割のうちのどれかを NP₁ が修飾している場合 (Pustejovsky 1995)。

- AGENTIVE role modification

[†] nishiguchi@cicee.tokushima-u.ac.jp

¹ 松本祐治氏に ChaKi の用法をお教え頂いたことに感謝を申し上げる。

- (1) フジの連ドラ (447)²
 (2) Docomo の FOMA (1157)

「連ドラ」の製作者また提供者は「フジ」テレビ、また「FOMA」の製作者は「ドコモ」であるので、各々「連ドラ」「FOMA」の AGENTIVE 役割を修飾していると考えられる。

- FORMAL role modification

- (3) 扇風機の羽根 (1084)

「羽根」は「扇風機」を全体とする部分であるので、FORMAL 役割を修飾していることになる。

II. NP₂ is a relational noun

- (4) 胸の前 (4179)
 (5) 潜水艦の中 (1824)
 (6) 扇風機の羽根の数 (1084)
 (7) ローンのこと (1336)
 (8) ダイエーの中内さんのバカ息子 (1394)

「前」「中」「数」「こと」「息子」はそれだけでは意味を成さない関係名詞であり、「胸」「潜水艦」「扇風機の羽根」「ローン」「ダイエーの中内さん」がそれぞれ、その項となっている。

III. NP₂ is a deverbal noun

- (9) a. 社長の期待 (4069)
 b. 社長が期待する。
 (10) a. ドラムブレーキの錆びの処理 (1910)
 b. ドラムブレーキの錆びを処理する。

「期待」は動詞「期待する」、「処理」は動詞「処理する」と同様、イベント項をもつ。(b)文が示すように、「社長」「ドラムブレーキの錆び」が期待イベントの主体役割、処理イベントの目的役割を担う。

IV. NP₁ is adjectival property of NP₂

(11)–(12)が示すように、限定形容詞は叙述形容詞として後置可能である。

- (11) a. アル中の俳優 (462)
 b. 俳優はアル中だ。
 (12) a. 無料のキャンプ場 (2078)
 b. キャンプ場は無料だ。
 (13) a. 扇風機の羽根 (1084)
 b. *羽根は扇風機だ。

Kuno (1973:25) は数量詞「一冊の」や「学生の」などの「の」格名詞句は叙述形容詞として後置可能であるので、実のところ属格の助詞ではなく、コピュラの「無料だ」の限定用法であると主張している。

- (14) a. 一冊の本
 b. 本は一冊だ。
 (15) a. 学生の田中
 b. 田中は学生だ。

² 括弧内の数字は ChaKi 出力時の sentence ID。

V. Referential module modification of NP₂

(16)の「近所の」、(17)の「真夜中の」は各々「整形外科」「海岸」の指示対象の一時的な場所、時間をあらわす。本論文では、指示対象の一時的な場所、時間、様態等を **referential module** という、クオリア構造とは別の構造を語彙的意味の一部として加える(後述)。「近所の」「真夜中の」は「整形外科」「海岸」の **referential module** を修飾している。

- 場所 Location

(16) 近所の整形外科 (3379)

- 時間 Time

(17) 真夜中の海岸 (3633)

VI. NP₁ is a quantifier

(18) ほとんどの家庭 (170)

(19) 他の人 (4012)

(20) かなりの人 (3875)

VII. Ga/no conversion in embedded clause

Harada (1971), Maki and Uchibori (2008) 等で取り上げられている「『が』『の』の格交替現象」である。例えば、(21)の関係節「あまり知識の無い」は「あまり知識がない」と置き換えられる。

(21) a. 小学6年のあまり知識の無い人では、ちょっと苦労しますよ。
(809)

b. 小学6年のあまり知識がない人では、ちょっと苦労しますよ。

VIII. Possession

(22) 相手の携帯 (5709)

(23) 自分のP C (883)

IX. Demonstrative

(24) それらのタイヤ (2066)

(25) どの地域 (4713)

X. NP₁ is a deverbal noun

(26) お勧めの公園 (5380)

「お勧め」は動詞「勧める」の名詞形であり、イベント項をもつ。「公園」はその目的語 theme argument である。

XI. NP₁ is a theme argument of deadjectival noun NP₂

(27) 旅の楽しさ (5395)

「楽しさ」は形容詞「楽しい」の名詞形で「旅」はその目的項 theme である。

XII. Adverb

(28) ただのまぬけ (5874)

XIII. Selective binding of qualia in NP₁

TELIC modification

(29) 全然違う学校の女の子 (3835)

「学校」は学生が勉強するための場所であり、「女の子」は勉強するというイベントの動作主 agent である。

3.3 分類上の問題点

3.3.1 種 kind を表す修飾語句 adjectival property と FORMAL 役割修飾

「長毛種の大型犬」の「長毛種の」は叙述形容詞として後置可能である。

- (30) a. 長毛種の大型犬 (5224)
b. 大型犬は長毛種だ。

種 kind を表す修飾語句は adjectival property と分類できるし、Pustejovsky の selective binding によって FORMAL 役割を修飾しているとも考えることもできるが、本研究では後置可能であるという統語的振る舞いを重視し、adjectival property として分類し、さらに kind というサブカテゴリーに分類した。395 件の adjectival property の用例のうち、53 例が種類を表すものであった。

3.3.2 同格の扱いについて

同格を表す「の」格も adjectival property として分類した。というのは、「二本の」などの数量詞と同様、叙述形容詞として後置できるからである。

- (31) a. インターハイなどの公式試合 (1880)
b. 公式試合はインターハイなどだ。

4. 結果と考察

4.1 統計的結果

分類の結果は以下のとおりである。表 1 では、「NP₁ の NP₂」構文の用例を、ややおおまかに分類した。4 クオリア構造のうちのどれを修飾しているか、また referential module のうちのどれを修飾しているかということには捉われない。

表 1 「の」格修飾の意味パターンと分布

selective binding of qualia in NP ₂	886	0.281448539
NP ₂ is a relational noun	777	0.24682338
NP ₂ is a deverbal noun	445	0.141359593
NP ₁ is adjectival property	395	0.125476493
referential module modification of NP ₂	244	0.07750953
NP ₁ is a quantifier	152	0.048284625
ga/no conversion	118	0.037484117
possession	45	0.01429479
demonstrative	32	0.010165184
NP ₁ is a deverbal noun	24	0.007623888
NP ₁ is theme of deadjectival NP ₂	23	0.007306226
adverb	6	0.001905972
selective binding of qualia in NP ₁	1	0.000317662
	3148	1

表 1 によれば、NP₂ のクオリア構造を「NP₁ の」が修飾している例と、NP₂ が関係名詞である場合が合わせて全体の半分以上を占める。NP₂ のクオリア構造修飾の例 (selective binding) が全体の 28 パーセント、NP₂ が関係名詞である場合が 24 パーセントを占める。その次に多いのが、NP₂ がイベント項を持つ動詞派生名詞である場合で、14 パーセント、また NP₁ が限定形容詞的である場合が 12 パーセントである。また、本論文の主張である referential module modification 外延構造修飾の場合が全体の 7.7 パーセントとなっている。以上の 5 タイプで全体の 87 パーセントを占める割合となる。

それ以外では「NP₁ の」が数量詞 quantifier であるとき、ga/no conversion 「が」「の」交替現象、所有関係、指示詞、NP₁ がイベント項を持つ派生名詞である場合、形容詞派生名詞の項である場合、副詞である場合などの用例が、その他の 10 パーセントほどを占める。

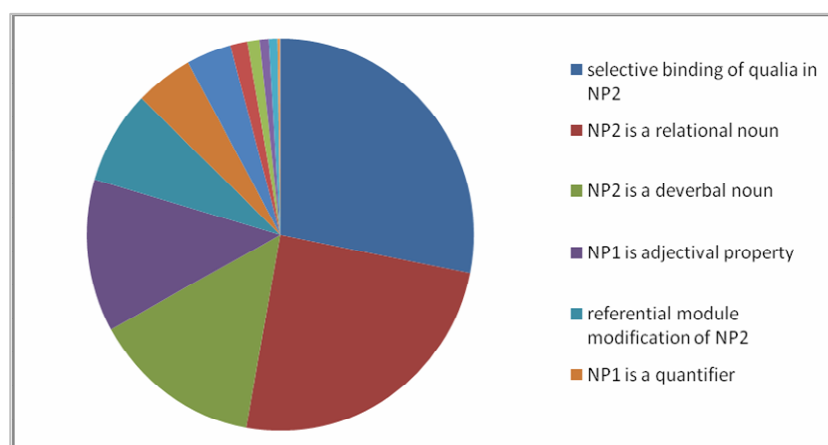


図 1 「の」格修飾の意味パターン分布の割合

表 2 は 4 クオリア構造 (FORMAL, CONSTITUTIVE, TELIC, AGENTIVE) のうちのどれを修飾しているか、また referential module の TIME, LOCATION, MANNER 等の下位構造のうち、どれを修飾しているかということも分類して個別の用例の頻度を求めたものである。

表 2 細分類した「の」格修飾の意味パターンと分布

NP ₂ is a relational noun		777	0.246823
NP ₁ is adjectival property		395	0.125476
selective binding of qualia in NP ₂	constitutive	322	0.102287
selective binding of qualia in NP ₂	telic	294	0.093393
NP ₂ is a deverbal noun	theme	244	0.07751
selective binding of qualia in NP ₂	agentive	220	0.069886
NP ₁ is a quantifier		152	0.048285
NP ₂ is a deverbal noun	agent	146	0.046379
referential module modification of NP ₂	location	131	0.041614
ga/no conversion		118	0.037484
referential module modification of NP ₂	time	112	0.035578

selective binding of qualia in NP ₂	formal	50	0.015883
possession		45	0.014295
demonstrative		32	0.010165
NP ₁ is theme of deadjectival NP ₂		23	0.007306
NP ₂ is a deverbal noun	time	18	0.005718
NP ₁ is a deverbal noun	theme	17	0.0054
NP ₂ is a deverbal noun	goal	12	0.003812
NP ₂ is a deverbal noun	location	11	0.003494
NP ₂ is a deverbal noun	source	9	0.002859
NP ₁ is deverbal noun	agent	7	0.002224
adverb		6	0.001906
NP ₂ is a deverbal noun	manner	4	0.001271
NP ₂ is a deverbal noun	instrument	1	0.000318
referential module modification of NP ₂	manner	1	0.000318
selective binding of qualia in NP ₁	telic	1	0.000318
total		3148	1

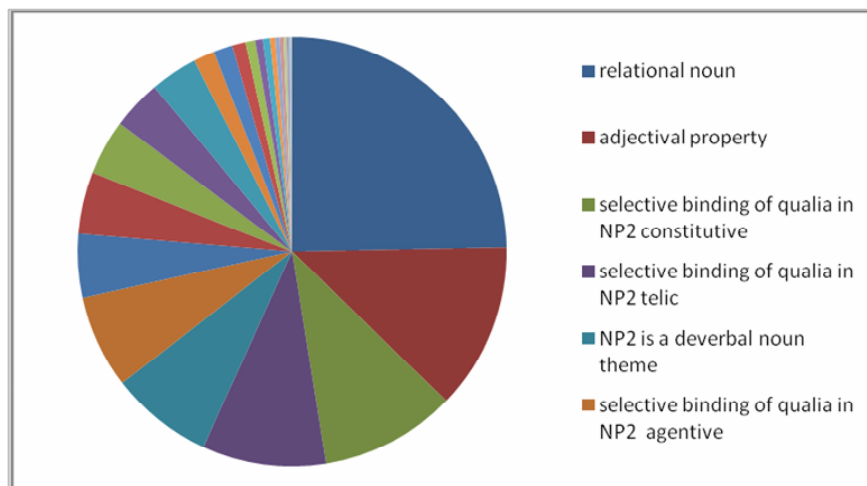


図2 細分類した「の」格修飾の意味パターンと分布

4.2 結果の考察

4.2.1 クオリア構造修飾の多用

「NP₁の」句による名詞修飾で一番多いのはクオリア構造の修飾である。クオリア構造は語の内在的特性をあらわすものである。

- AGENTIVE role modification

(32) 鳥の卵 (5249)

$[[\text{鳥の卵}]] = \lambda x[\text{egg}'(x) \ \& \ \text{AGENTIVE} = \exists e.[\text{bear}'(x)(\epsilon y.\text{bird}'(y))]]$

(33) カナダ大手新聞の統計 (5350)

[| カナダ大手新聞の統計 |] = $\lambda w, x[\text{statistics}'(x) \ \& \ [\text{AGENTIVE} = \text{make_act}(e) \ \& \ \text{theme}(e) = x \ \& \ \text{agent}(e) = \epsilon y.\text{major_Canadian_newspaper}'(y)]]^w$

- FORMAL role modification

(34) [| 詰め将棋の本 |] = $\lambda x[\text{book}'(x) \ \& \ \text{FORMAL} = \exists e[\text{hold}'(e) \ \& \ \text{agent}(e) = x \ \& \ \text{theme}(e) = i]]$

- TELIC role modification

(35) 評価の返信 (5589)

[| 評価の返信 |] = $\lambda x[\text{reply}'(x) \ \& \ \text{TELIC} = \exists e[\text{communicate}'(\epsilon y.\text{evaluation}'(y))]]$

5. 主張と提案

5.1 関係名詞の定義の拡張

一般的に関係名詞というと、「夫の両親」の「両親」、「彼女の友人」の「友人」などの例が当てはまる。Partee (1983/1997) が英語の *John's brother* などの英語の所有表現について指摘しているように、「夫の両親」の「夫」と「両親」、「彼女の友人」の「彼女」と「友人」との関係は「両親」「友人」が語彙的にあらず親子関係、友人関係を各々受け継ぐこととなる。

本稿ではさらに、「上」「下」「横」「間」などの場所的な位置関係をあらわす関係名詞も同様に扱う。空間的な位置関係を表すのに、英語では *in, on, under* などの前置詞が用いられるが、世界の言語では前置詞、後置詞の代わりに、日本語の「中」「上」「下」などに相当する関係名詞を使って位置関係を表す言語がある。例えば、Lillehaugen and Munro (2006) は、メキシコの Tlacolula Valley Zapotec また北米の Chickasaw の例を挙げている。

西山(2003)は「非飽和名詞」として「この芝居の主演」の「主演」、「太郎の上司」の「上司」のような名詞を挙げて、それらは「この芝居の」また「太郎の」のような「パラメータ」の値が決まらなければ意味的に充足しない不飽和な名詞であるとした。

それで本研究では、西山の言う非飽和名詞も関係名詞として扱う。「扇風機の羽根の数」の「数」、「ローンのこと」の「こと」、「飛行機の方」の「方」、「今年の年末」の「年末」、「主人公の名前」の「名前」などはみな「○○の」にあてはまる項が必要な関係名詞とみなす。普通名詞が一項述語であるので、これらの関係名詞は 2 項述語、また「間」のような名詞は 3 項述語である。

- (36) a. [| 間 |] = $\lambda x \lambda y \lambda z[\text{between}'(z)(y)(x)]$
 b. [| 歯と歯茎の間 |] = $\lambda x[\text{between}'(\epsilon y.\text{gum}'(y))(\epsilon z.\text{tooth}'(z))]$
- (37) a. [| 年末 |] = $\lambda x \lambda y[\text{end-of-the-year}'(y)(x)]$
 b. [| 今年の年末 |] = $\lambda y[\text{end-of-the-year}'(y)(t)]$
- (38) a. [| 名前 |] = $\lambda x \lambda y[\text{name-of}'(y)(x)]$
 b. [| 主人公の名前 |] = $\lambda x[\text{name-of}'(x)(\epsilon y.\text{hero}'(y))]$

5.2 Generative Lexicon 理論の拡張

5.2.1 Generative Lexicon 理論の限界

このセクションでは表 1 において 7.7 パーセントを占める referential module modification of NP₂ を形式化することを提案する。そのために Pustejovsky の GL 理論における語彙構造を拡張する。

(39)が示すように、GL 理論のクオリア構造だけでは、発話地点における一時的な場所関係をあらわす「オペラ座の怪人」、一場面をあらわす「真夜中の海岸」、一時的な特徴である乗り物、服装、持ち物などの特徴、例えば「バイクの彼ら」の意味を計算する術がない。今オペラ座に住んでいる怪人は生まれたときからそこに住んでいたのではないし (AGENTIVE 役割)、「真夜中の海岸」は真夜中に遊ぶために作られた公園でもない (TELIC 役割)。既存の理論ではそのような意味を導き出すことができない。

- (39) a. $[[\text{オペラ座の怪人}]] \neq \lambda x[\text{phantom}'(x) \ \& \ [\text{AGENTIVE} = \exists e[\text{born}'(e) \ \& \ \text{theme}(e) = x \ \& \ \text{location}(e) = \text{The Opera}']]]$
 b. $[[\text{真夜中の海岸}]] \neq \lambda x[\text{beach}'(x) \ \& \ [\text{TELIC} = \exists e[\text{recreational_activity}(e) \ \& \ \text{time}(e) = \text{midnight}']]]$
 c. $[[\text{バイクの彼ら}]] \neq \lambda x[g(1) = x \ \& \ [\text{AGENTIVE} = \exists e[\text{born}'(e) \ \& \ \text{manner}(e) = \text{with-scooter}']]]^3$

では、これらの一時的な特性とはなんだろうか。

5. 2. 2 非内在的特性と *Ex-/Nuovo* テスト

Busa et al (2001) によれば、イタリア語の接頭辞 *ex-* “former” と形容詞 *nuovo* “new” は変わりうる性質をあらわす名詞でなければ修飾できない。というのは、それらの語は現在あるいは過去においては指示対象はその特性を持っていなかったことを含意するからである。そのため、Busa et al (2001) はイタリア語の接頭辞 *ex-* “former” は語彙的意味として内在的な FORMAL 役割とは別の意味を修飾しているとしている。

- (40) a. *ex-musicista*
 former musician
 b. *#ex-uomo*
 former man
 (41) a. *#un nuovo musicista*
 a new musician
 b. *un nuovo orchestrale*
 a new orchestra player

(40)で *ex-musicista* は文法的であるのに *ex-uomo* は非文法的でないのは *uomo* “man”であるというのは恒常的で通常変わらない性質であるのに対して *musicista* “musician”であるという職業は変わり得るので、以前音楽家であった人がやめて今は別の仕事をしているかも知れないからである。それに対して、(41)が示すように、*nuovo* “new” という形容詞はそれよりももっと一時的な特質でないと修飾できない。音楽家というのは急になれるものではないので「新しい音楽家」とは言えない。しかし「新しいオーケストラの団員」は文法的である。オーケストラの団員というのは、なったり辞めたりするのが音楽家と比べてしやすいということである。

益岡 (2008) は「あの人は多忙だ」における「多忙だ」などの属性は「日本は島国だ」の「島国だ」などの「内在的特性」と比べるともっと時間的に限定されるとして「非内在的特性」と呼んでいる。Busa et al (2001)の *ex-/nuovo* テストをしてみると、確かに非内在的

³ 代名詞の扱いに関しては Heim and Kratzer (1998:241)に倣う。1 が指標 index で *g* が 1 を領域 domain 内に含む variable assignment であれば $[[\text{彼ら}]]^g = g(1)$ 。

性をあらわす性質は「最近」「以前」などで時間限定可能であるが、内在的特性は「最近」で修飾できない。

- (42) a. あの人は最近多忙だ。
b. あの人は以前多忙だった。

- (43) *日本は最近島国だ。

セクション4で **referential module modification** としている例のうち、(47)のように様態を修飾しているものは「以前」で修飾可能であるので非内在的特性である。

- (44) *以前近くのお店
(45) *以前オペラ座の怪人
(46) *以前真夜中の海岸
(47) a. 最近バイクの彼ら
b. 最近バイクの田中さん

場所をあらわす「近くの」、時間をあらわす「真夜中の」は「以前」で修飾できないが、(39)が示すように、恒常的な特質をあらわすものでもない。それゆえ、一時的な特性をあらわす修飾語との意味的合成を可能にするために、Pustevjosky の GL 理論の拡張を試みる。

5.2.3 Generative Lexicon 理論の拡張

Pustejovsky の4クオリア構造は語の指示対象の内在的で本質的な性質を捉えているが、それに加えて語の指示対象の外延的特徴も語彙項目の情報として付加する。GLy の語彙情報には、タイプ構造、項構造、イベント構造、クオリア構造が含まれており、その中でクオリア構造には **FORMAL**、**CONSTITUTIVE**、**AGENTIVE**、**TELIC** 役割が付与されている。

本論文が提案する拡張された語彙情報には、クオリア構造以外に外延的モジュール (**referential module**)、**EXTENSION (EXT)** 構造を付加する。EXT には **TIME**、**LOC**、**MANNER** 役割という下位構造がある。具体例においては、「オペラ座の怪人」の「オペラ座の」、「真夜中の海岸」の「真夜中の」はそれぞれ「怪人」「海岸」の外延的モジュール (**extensional module**) EXT を修飾していると考える。「バイクの彼ら」ではバイクに乗っているというのは指示対象の一時的特性であるので、イベントの様態項に当てはまる。

(48) 従来の GL テンプレート

(cf. Johnston & Busa 1996)

$$\left(\begin{array}{l} \alpha \\ \text{TYPESTR} = \{ \text{ARG}_1 = x \} \\ \text{EVENTSTR} = \left[\begin{array}{l} E_1 = e_1 \\ D-E_1 = e_2 \end{array} \right] \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG}_2 = y \\ D-\text{ARG}_1 = z \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{CONST} = \\ \text{FORMAL} = \\ \text{TELIC} = \\ \text{AGENTIVE} = \end{array} \right] \end{array} \right)$$



(49) Extended GL テンプレート

$$\left(\begin{array}{l} \alpha \\ \text{TYPESTR} = \{ \text{ARG}_1 = x \} \\ \text{EVENTSTR} = \left[\begin{array}{l} E_1 = e_1 \\ D-E_1 = e_2 \end{array} \right] \\ \text{ARGSTR} = \left[\begin{array}{l} D-\text{ARG}_1 = l \\ D-\text{ARG}_2 = t \\ D-\text{ARG}_3 = y \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{CONST} = \\ \text{FORMAL} = \\ \text{TELIC} = \\ \text{AGENTIVE} = \end{array} \right] \\ \text{EXT} = \left[\begin{array}{l} \text{LOC} = \text{in}(e_2, x, l) \\ \text{TIME} = \text{at}(e_2, x, t) \\ \text{MANNER} = \text{with}(e_2, x, y) \end{array} \right] \end{array} \right)$$

その結果、Pustevjosky のクオリア構造の selective binding と同様の仕組みで「NP₁ の NP₂」の意味を計算できる。例えば「オペラ座の怪人」の「オペラ座の」は「怪人」の指示対象の居場所をオペラ座とする。オペラ座で生まれた怪人ではなく（クオリア構造内 AGENTIVE 役割修飾）オペラ座で何らかの役割を果たす目的のためにいる怪人ではなく（クオリア構造内 TELIC 役割修飾）、オペラ座に存在する怪人である（EXT 構造内 LOCATION 役割修飾）。

- (50) a. $[[\text{オペラ座の怪人}] \neq \lambda x[\text{phantom}(x) \ \& \ [\text{EXT} = \exists e[\text{be-phantom}(e) \ \& \ \text{theme}(e) = x \ \& \ \text{location}(e) = \text{The Opera}]]]]$
- b. $[[\text{真夜中の海岸}] = \lambda x[\text{beach}'(x) \ \& \ [\text{EXT} = \exists e[\text{be-beach}'(e) \ \& \ \text{theme}(e) = x \ \& \ \text{time}(e) = \text{midnight}]]]]$
- c. $[[\text{バイクの彼ら}]]^s = \lambda x[g(1) = x \ \& \ [\text{EXT} = \exists e[\text{born}(e) \ \& \ \text{manner}(e) = \text{with-scooter}]]]]$

6. 結論

本研究では「Yahoo!知恵袋」の中から「NP₁ の NP₂」形の事例を取り出し、意味特性に従って分類した。その結果、GL 理論のクオリア構造の修飾の例、また広義の関係名詞に当たる用例が多いことがわかった。同時に、GL 理論を拡張すれば意味計算可能となるような例が少なからずあることがわかった。

文献

- Harada, Shin-ichi (1971) “*Ga/no* Conversion and Ideolectal Variations in Japanese,” *Gengo Kenkyu* 60: 25-38.
- Heim, Irene and Angelika Kratzer (1998) *Semantics in Generative Grammar*, Blackwell, Oxford.
- Johnston, Michael and Federica Busa (1996) “Qualia Structure and the Compositional Interpretation of Compounds,” *Proceedings of the ACL SIGLEX Workshop on Breadth and Depth of Semantic Lexicons*, 77-88, Kluwer, Dordrecht.
- Kuno, Susumu (1973) *The Structure of the Japanese Language*, MIT, Cambridge.
- Lillehaugen, Brook Danielle and Pamela Munro (2006) “Prepositions and Relational Nouns in a Typology of Component Part Locatives,”
[<http://www.linguistics.ucla.edu/people/grads/lillehaugen/LillehaugenMunro2006aho.pdf>].
- Maki, Hideki and Asako Uchibori (2008) “*Ga/no* Conversion,” Shigeru Miyagawa and Mamoru Saito eds., *The Oxford Handbook of Japanese Linguistics*, 192-216, Oxford University Press, Oxford.
- 益岡隆志 (2008) 『叙述類型論』、くろしお出版、東京。
- 西山佑司 (2003) 『日本語名詞句の意味論と語用論—指示的名詞句と非指示的名詞句』、ひつじ書房、東京。
- Partee, B. H. (1983, 1997) “Genitives: A case study,” Johan van Benthem and Alice ter Meulen eds., *Handbook of Logic and Language*, 464-470, Elsevier, Amsterdam.

コーパス

国立国語研究所 (2009) 『現代日本語書き言葉均衡コーパス BCCWJ モニター公開データ』。

「～がる」と感情を表す「～む」動詞の意味・用法

韓 金柱（東京外国語大学 大学院地域文化研究科博士後期課程）

The Meaning and Usages of -garu and Emotional Verb -mu

Han Jinzhu

(The Graduate School of Area and Culture Studies, Tokyo University of Foreign Studies)

1. はじめに

1.1 研究の目的

本研究は、「～がる」と感情を表す「～む」動詞との間に、どのような意味・用法の違いがあるかを探ることを目的とする。ここで、「～がる」とは、形容詞の語幹に「がる」をつけ、動詞化した形（例えば、「懐かしがる」「羨ましがる」など）のことであり、また本研究では、感情形容詞¹に形態的・意味的に対応する動詞（例えば、「懐かしい」に対して「懐かしむ」、「羨ましい」に対して「羨む」など）のことを「感情を表す『～む』動詞」と呼ぶこととする。なお、感情を表す「～む」動詞には、「恥ずかしい」に対応する「恥じる」、「欲しい」に対応する「欲する」などのように、必ずしも「～む」の形で終わらない動詞も含むものとする。

1.2 分析の対象

分析の対象とするのは、以下の（１）（２）に示すような「感情形容詞＋がる」および、感情を表す「～む」動詞である。（以下、分析の対象となる「感情形容詞＋がる」と感情を表す「～む」動詞に下線を付して示す。また特に出典を示さない例文は執筆者による作例である。）なお以下、先行研究の内容に言及する際には、「～がる」および「～む」動詞の表記は、それぞれ先行研究における表記に従い、随時「～がる」「ーがる」および「～む」「ーむ」「む」などとするにすることにする。

（１）彼女は自分の失敗を心の中で悔しがっている。

（２）彼女は自分の失敗を心の中で悔やんでいる。

1.3 用例収集について

本研究では、『現代日本語書き言葉均衡コーパス』モニター公開データ（2009 年度版）から用例を収集して考察を行う。今回は特にコーパスの検索対象を限定せず、「書籍」「白書」「国会会議録」「Yahoo!知恵袋」の全体から用例を収集した。その際、感情を表す「～む」動詞については、『日本語能力試験出題基準』の１級までの感情形容詞 85 語をもとに、形態的・意味的に対応する動詞を有する 15 語²を、検索の対象とした。

¹ 本研究での「感情形容詞」とは、「私は一。」の形で第一人称名詞句を主題として、感情表出文を作ることが可能である形容詞（「悔しい」「痛い」など）のこととする。

² 「痛む」「羨む」「惜しむ」「恐れる」「悲しむ」「悔やむ」「苦しむ」「恋する」「楽しむ」「懐かしむ」「悩む」「憎む」「はかなむ」「恥じる」「欲する」の 15 語を検索の対象とした。

2. 「～がる」と感情を表す「～む」動詞との違いに関する先行研究の記述の検討

西尾（1975）は、「羨ましがる」と「羨む」について、「羨ましがる」は「羨ましいという感情を言葉や表情に表したり、それを相手にも伝えようとする場面によく使われる」のに対し、「羨む」は「心の中だけの目に見えない事象でも使われる」としている。ただし、例文として「私は事情が自分の推察通りだったので、内心嬉しがりました」を挙げ、「～がる」には「感情が外には表現されていない場面での用例もあるので、必ずしも外発的な感情活動を表すとは言い切れない」と指摘している。

しかし、上記のように「～がる」と感情を表す「～む」動詞が、どちらも用いることができるという場合の違いについては、述べていない。まず、感情が外に表されているとする場合、例えば、「彼女は私のことを『いいなあ』と羨ましがった」と「彼女は私のことを『いいなあ』と羨んだ」との違いがどこにあるのかについては明らかではない。また、感情が外に表されていないとする場合、「彼女は自分の失敗を心の中で悔しがっている」と「彼女は自分の失敗を心の中で悔やんでいる」との違いがどこにあるのかについても述べられていない。

大曾（2001）は、「～がる」が、話し手によってとらえられた特定の人物の感情表出の描写に使われるのに対し、「～む」は、事物の性状規定、つまり感情的品定めに使われるとしている。しかし、「～がる」は、必ずしもある人物の感情“表出”の描写に使われるとは限らない。例えば、「彼女は顔には出さなかったが、心の中では悔しがっていた」のような場合の「悔しがる」は、主語に当たる人物「彼女」の感情が、必ずしも表出されていることを描写しているとは言えない。

大曾（2001）は、また、「苦しがる」と「苦しむ」は明らかに意味用法に違いがあるとし、『苦しがる』は肉体的な苦しみを表すのに対し、『苦しむ』は精神的・肉体的両方の苦しみを表す」と述べている。ただし、この記述はこの2語の違いだけにとどまり、この分析結果が他の「～がる」と感情を表す「～む」動詞との違いについても、当てはまるのかどうかは検討されていない。

王（2005）は、「～がる」の使用は、感情事象の関与者である「感情主」に密接に関わるものであり、「～がる」は感情主の行動や意志を中心に描写する機能を果たすのに対して、感情を表す「む」類動詞の使用は、むしろ「感情対象」に密接に関わっており、感情対象の性質・状態、およびそれによって引き起こされた感情行為を描写するとしている。

例えば、「花子は私のフィアンセを見て羨ましがった／羨んだ」の場合、感情主である「花子」を中心に描写する際には「羨ましがる」が使われ、感情対象である「私のフィアンセ」を中心に描写する際には「羨む」が使われるとしている。しかし、「感情主」を中心として描写しているか、それとも「感情対象」を中心として描写しているかを、何を基準に判断するのが十分に説明されていない。また、この場合、感情対象は「私のフィアンセ」ではなく「私」と解釈することも可能ではないかと考えられる。

以上のような記述は、①分析の対象として扱われている例が少ない。②概念の設定が明確にされていない。③両者の違いを明確にするだけの十分な記述が行われていない、という各点において十分でないと考えられる。

3. 「～がる」の意味・用法

「～がる」の意味・用法に関する先行研究としては、浅野他（1971）、西尾（1972, 1975）、森田（1988）、富田（1997）、黄（2004）、王（2005）、市川（2005）などが挙げられる。韓（印刷中）では、これらの先行研究の記述について問題点³を指摘し、様態の「そうだ」と比較して接尾辞「がる」の意味・用法について考察を行った。そこでは、「対象となる人物が示している外的な様子⁴」と「対象となる人物の内面⁵」、および「総合的な知識⁶」という3つの観点を立て、分析を行った。以下、第3節ではその分析結果をまとめて示す。

3.1 「～がる」の意味

「～がる」の意味は、「話者が、対象となる人物が示している外的な様子を、総合的な知識に基づいて、その人物の内面と関係付けてとらえていることを表す」ものであると考えられる。

3.2 「～がる」の用法

具体的には以下のような3つの用法を観察することができる。（以下、対象となる人物が示している外的な様子に、 を付して示す。）

まず「～がる」は、対象となる人物が示している外的な様子が、一般的に人が当該の形容詞で表される感情を持っている際に、それを示す様子の一種であり、話者がこの外的な様子と対象となる人物の内面とが「一致する」ものととらえている場面で用いられる。例えば、以下の（3）のようなものである。

- （3）椎野と禿も研究室に呼びつけて、「お前ら、恥ずかしくないのか」と叱りつけた。
じっと聞いていた椎野は拳を握りしめ、涙をぼろぼろ流して悔しがって、「私は今まで演劇なんてどうでもいいと思っていましたが、演劇、続けます。（『教室を路地に！』）

（3）で対象となる人物「椎野」が示している外的な様子「拳を握りしめ、涙をぼろぼろ流す」は、一般的に人が「悔しい」感情である際に示す様子の一種であると考えられる。ただし、この場合、対象となる人物が示しているこのような外的な様子は、あくまで話者がとらえた対象となる人物の様子であり、その人物が本当に当該の形容詞「悔しい」という感情を持っていて、それを外に表したものであるという保証はない。つまり、ここでの「～がる」という表現は、話者が、対象となる人物が示している外的な様子を、一般的な知識に基づいてその人物の内面と関係付け、その外的な様子は、対象となる人物の「悔しい」という感情の表出であるものととらえていることを表していると考えられる。この時、話

³ ①「～がる」という表現が用いられる時、主語に当たる人物が必ずしも内的にその形容詞で表される気持ちや状態にあるとは限らない、②「～がる」という表現が用いられる時、主語にあたる人物の「内的状態」と「外に表れた様子」は必ずしも一致するとは限らない、③「～がる」という表現が一人称で用いられる時、必ずしも、浅野他（1971）や黄（2004）が述べるように、自分自身の「以前」に抱いた感情・感覚（「むかし」のこと）をとらえて、それを客体的に眺めているとは限らない、④様態の「そうだ」との違いが明確に記述されていない、という各点において十分でないと考えた。

⁴ 対象となる人物が発した言葉、示した態度、動作、動き、表情などのこと。

⁵ 対象となる人物の心理状態のこと。

⁶ 話者が、外的な様子を、その内面と関係付けてとらえる際に用いる知識のこと。つまり、人がある感情を抱いている際に見せる表情や態度、行動などに関する一般的な知識、および話者の頭の中に蓄積されている対象となる人物に関する知識（その人物の普段のふるまい、またはその人物の背景など）のこと。

者は対象となる人物の内面と、その人物が示している外的な様子とは「一致する」ものととらえていると思われる。

二つ目は、対象となる人物が示している外的な様子は、一般的に人が当該の形容詞で表される感情を持っている際に、それを示す様子の一種であるとは考えられないものであり、話者が、この外的な様子と対象となる人物の内面とが「一致しない」ととらえている場面で用いられる。例えば、以下の(4)のようなものである。

- (4) 昨日の中居君のドラマフェス、リアルタイムで見れなくて、まだPのとこだけちょっと拾って見たただけなんだけど、アーチェリーはうまかったけど、バスケとボーリングは、なかなか、「スター山P」を見せることはできなくて、残念！Pは苦笑いして平静顔をしてたけど、きっと内心では、すんごく悔しがってたんだろうな～って。(http://blog.goo.ne.jp/foreverheart2424/)

(4)での対象となる人物「スター山P」が示した態度「平静顔をしてた」は、一般的に人が当該の形容詞「悔しい」で表される感情を持っている際に、それが外的にそのまま表れた様子であるとは考えられない。話者は、対象となる人物「スター山P」が示しているこのような外的な様子を、頭の中に形成されているその人物に関する知識を根拠として、その内面に関係付けていると考えられる。例えば、対象となる人物「スター山P」は有名人であり、普段負けず嫌いで、かつこれまでいつも優勝していたことを知っているなどである。つまり、話者は、対象となる人物が示している外的な様子(平静顔をしていること)を総合的な知識に基づき、その人物の内面に関係付け、その人物がやはり内面には「悔しい」という感情を持っているととらえ、「へがる」という表現を用いていると思われる。この時、話者が、対象となる人物の内面とその人物が今示している外的な様子とは「一致しない」ととらえていると考えられる。

三つ目は、対象となる人物が示している外的な様子が、一般的に人が当該の形容詞で表される感情を持っている際にそれを示す様子の一種であるが、話者が、この外的な様子と対象となる人物の内面とが「一致しない」ととらえている場面で用いられる。例えば、以下の(5)のようなものである。

- (5) A: 巨人関係者並びプロ野球関係者は、サッカーで日本が負けて喜んでいますか？
B: 表面上は悔しがっていても心の中ではほくそ笑んでいる事でしょう。特にナ○ネさん・・・。(http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q118451226)

(5)の場合、対象となる人物が示している「外的な様子」が具体的にどのようなものであるかは述べられていないが、「表面上」とされていることからわかるように、対象となる人物は、外的には、一般的に人が「悔しい」感情である際に示す何らかの態度を示していると思われる。この場合、話者は、対象となる人物が示している表面上の態度を、頭の中に形成されている対象となる人物に関する知識(例えば、対象となる人物の普段の

ふるまい、その背景など)に基づき、その人物の「内面」に関係付けて、それは、対象となる人物の「悔しいふりをしよう」という目論見の表れであるだろうととらえたことを表していると考えられる。この時、話者が、対象となる人物の内面とその人物が今示している外的な様子とは「一致しない」ととらえていると考えられる。

以上の3つの用法における共通点は、「対象となる人物が示している外的な様子を、その人物の内面に関係付けてとらえ、表している」という点である。

4. 感情を表す「～む」動詞の意味・用法

4.1 感情を表す動詞に関する先行研究の記述の検討

寺村(1982)は、感情を表す動詞について、「～ニ」という形の、感情の動きの誘因を表す補語をとるタイプのものと、「～ヲ」という形の、感情の向かう対象を表す補語をとるものの2つのタイプがあるとしている。本研究で研究対象とする感情を表す「～む」動詞は、ここでいう「～ヲ」という形をとるタイプに分類されている。

しかし、感情を表す「～む」動詞は、必ずしも「～ヲ」という形をとるとは限らない。例えば、「～ヲ」という形をとる動詞として挙げられている「苦しむ」「恋する」は、実例を観察すると、「～ヲ」という補語をとるというより、むしろ「～ニ」という補語をとることが多い。また、「悲しむ」などは「{失敗／力不足／親友の死}に悲しむ」のように「～ニ」という補語をとることも可能である。このように、感情を表す動詞には、「～ヲ」という補語と「～ニ」という補語の両方をとるものがあるということについては、清水(2007)などでも指摘されている。このように、感情を表す動詞は、必ずしも「～ヲ」という補語をとるものと、「～ニ」という補語をとるものの2つに、截然と分けられるとは言えない。

寺村(1982)はまた、「～ヲ」という形をとる動詞は、主体の感情が、対象をめざして「発動する」ことを表すとしている。このことは、直接受身文を作ると、主体は「ニ」のみならず、「カラ」もとることができるということで裏付けられている。

しかし、感情を表す「～む」動詞は、必ずしも直接受身文を作れるとは限らない。例えば、「彼女は失敗を悔やんだ」の場合は、「失敗は、彼女*ニ/*から悔やまれた」という直接受身文を作ることができない。このような場合、何をもって「発動性」があるとするのかが明らかではない。この場合、「彼女」の「悔しい」という感情は、むしろ「失敗」が原因となって、引き起こされたことを表しているのではないかと考えられる。

山岡(2002)は感情動詞を「感情表出動詞」(例えば、「むかつく」など)、「感情変化動詞」(例えば、「あきれた」など)、「感情描写動詞」(例えば、「苦しむ」「楽しむ」など)の3つに分類している。本研究で研究対象とする感情を表す「～む」動詞は、ここでいう「感情描写動詞」に分類されている。ここでは、「感情描写動詞」について「表情や態度として持続的に表出されている感情を、二次的・間接的に描写する」ものであるとしている。

しかし、感情を表す「～む」動詞は、必ずしも表情や態度として「表出」されている感情について描写するとは限らない。例えば、「彼女は顔には出さなかったが、心の中で悔やんでいた」のような場合には、「彼女」の感情は必ずしも表出されているわけではない。

以上、先行研究での記述を見ると、①感情を表す動詞は、必ずしも「～ヲ」という補語

をとるものと、「～ニ」という補語をとるものの2つに、截然と分けられるとは言えない。
②「発動性」についての概念が明確でない。③感情を表す「～む」動詞は、必ずしもいつも外に「表出」されている感情について描写しているとは限らない、という各点において十分でないと考えられる。

4.2 感情を表す「～む」動詞の意味

本研究では、感情を表す「～む」動詞の意味を、「主体がある対象に対して、当該の感情を抱く」ことを表すものであると考える。なお、ここでの「感情」とは、その動詞に形態的・意味的に対応する感情形容詞で表される感情や気持ちのことを指す。

4.3 感情を表す「～む」動詞の用法

4.3.1 主体が何らかの原因でそのような感情を抱く場合

まず、感情を表す「～む」動詞は、主体が何らかの原因でそのような感情を抱くことを表す際に用いられる。この用法においては、感情を抱く対象は「～ヲ」という補語によって表される場合も、「～ニ」という補語によって表される場合もあるが、いずれの場合も、その対象が、感情を抱く原因となっている⁷。例えば、以下の(6)～(8)のようなものである。(以下、感情を抱く対象に、 を付して示す。)

(6) ジョンダラーは性急に行動したことを悔やみ、自分が大人げないふるまいを見せたことを恥じ、ああも唐突に炉辺を出てきたからには二度と歓迎されないはずだ、とひとりで決めつけていたのだ。(『マンモス・ハンター』)

(7) 過去の出来事を振り返り、相手がこう思っていたんじゃないかなー、だから私はこうすればよかったんだーとか過去の自分の失敗を悔やんでしまいます。

(『Yahoo!知恵袋』)

(8) 重病に苦しむザビエルを放置して、臨終に立ち会いもしなかったアルメイダの情報は文字通り伝聞でしかないし、アタイデの息がかかった胡散臭い彼がザビエルの病死を強調するのも不自然な感じがする。(『ザビエルの謎』)

(6) で感情を抱く対象となる「大人げないふるまいを見せたこと」は「～ヲ」という補語によって表されている。しかしこの場合、対象となる「大人げないふるまいを見せたこと」は「大人げないふるまいを見せたことによって」に置き換えることが可能である。このことから、主体となる「ジョンダラー」が「大人げないふるまいを見せたこと」が原因で、「恥ずかしい」という感情を抱くことを表していると考えられる。(7) での感情を抱く対象となる「自分の失敗」も「～ヲ」という補語によって表されているが、主体となる「私」が「自分の失敗」が原因で「悔しい」という感情を抱くことを表すと思われる。

⁷ 山岡(2002)では、「ヲ格」と「ニ格」のそれぞれについて、「のせいで」に置き換えることができるものを「原因格」、置き換えることができないものを「対象格」としている。ここで、感情を抱く対象が、原因となっているか、原因とはなっていないかについての判断は、山岡(2002)に従うことにする。ただし、ここでは、「のせいで」の他に、「(な) ので」「によって」「のために」に置き換えることができるものについても、感情を抱く原因となつてゐるとする。つまり、「のせいで」「(な) ので」「によって」「のために」のいずれかに置き換えることができるものは、原因となつてゐるとし、置き換えることができないものは、原因とはなつてゐないとする。

感情を表す「～む」動詞は、(8)のように「～ニ」という補語をとる場合もある。ただし、この場合も(6)(7)と同様、主体となる「ザビエル」は「重病」が原因で「苦しい」という気持ちを抱くことを表すと考えられる。例えば、対象となる「重病に」を「重病なので」に置き換えることが可能である。

(6)～(8)を見ると、対象が「～ヲ」という補語をとるか「～ニ」という補語をとるかに関わらず、いずれも主体が、ある対象が原因となって、その対象に対してそのような感情を抱くことを表す。

4.3.2 主体が自発的にそのような感情を抱く場合

二つ目は、感情を表す「～む」動詞は、主体が自発的にそのような感情を抱くことを表す際に用いられる。この用法においても、感情を抱く対象は、「～ヲ」という補語によって表される場合も、「～ニ」という補語によって表される場合もある。ただし、この場合は、対象が感情を抱く原因とはなっていない。例えば、以下の(9)～(11)のようなものである。

(9) やがて戦争が終わり、餃子や鍋貼の製法を身につけた日本人が帰還し、当時の味を懐かしみ、餃子などの中華料理を作り始めました。(『Yahoo!知恵袋』)

(10) 平凡を欲し、平凡の中でしか潑刺と生きることの出来ない女なのだ。

(『夢見通りの人々』)

(11) 小督に恋する人は帝ばかりでなく、冷泉隆房は彼女が宮中に召される前から見初めて、召された後も恋心を断ち切れないありさまであった。(『洛中洛外漢詩紀行』)

(9)で、感情を抱く対象となる「当時の味」は「～ヲ」という補語によって表されている。しかし、この場合、対象となる「当時の味を」は「当時の味なので」に置き換えることができない。つまり、主体となる「日本人」が自ら対象となる「当時の味」を思い出し、自発的に「懐かしい」という感情を抱くことを表していると思われる。(10)での感情を抱く対象となる「平凡」も「～ヲ」という補語によって表されているが、主体となる「自分」が、対象となる「平凡」に対して、自発的に「欲しい」という感情を抱くことを表す。

この用法においても、(11)のように対象が「～ニ」という補語によって表されている場合もある。この場合も、(9)(10)と同様、主体となる「人」が「小督」に対して、自発的に「恋しい」という感情を抱くことを表すと考えられる。例えば、対象となる「小督に」を「小督のせいで」に置き換えることができない。

(9)～(11)を見ると、対象が「～ヲ」という補語をとるか「～ニ」という補語をとるかに関わらず、いずれも主体が自発的に対象に対してそのような感情を抱くことを表す。

5. 「～がる」と感情を表す「～む」動詞との相違点

まず、主体の「～ふりをしよう」という目論見の表れであることを表す場合には、「～がる」は使うことができるが、感情を表す「～む」動詞を使うのは不自然となる。例えば、以下の(12)のようなものである。

- (12) a. 彼女は表面上は敗戦を悔しがったが、心の中では喜んでいた。（「がる」の用法 3）
 b. ? 彼女は表面上は敗戦を悔やんだが、心の中では喜んでいた。

「～がる」は対象となる人物の外的な様子を、その人物の内面と関係付けてとらえていることを表すものであり、その人物が本当にそのような感情を持っているか、持っていないかについて述べるものではない。これに対して、感情を表す「～む」動詞は、主体の外的な様子については問題にせず、主体が当該の感情を抱くことを表すものである。そのため、(12) b のように、主体が心の中でそのような感情を持っていない場合について述べる際には、感情を表す「～む」動詞を用いると不自然となる。

次に、主体がそのような感情を持っていることを表す際には、「～がる」と感情を表す「～む」動詞は、どちらも用いることができる。例えば、次の (13) のようなものである。

- (13) a. 彼女は自分の失敗を心の中で悔しがっていた。（「がる」の用法 2）
 b. 彼女は自分の失敗を心の中で悔やんでいた。

「～がる」と感情を表す「～む」動詞は、いずれも人の感情に関連することについて述べるという点では共通している。そのため、(13) のように、人の心の中について述べる際には、どちらも用いることが可能となる。ただし、「～がる」は主体の内面と外的な様子とを関係付けて述べるものである。つまり、(13) a は話者が「彼女」の内面は外的な様子とは一致せず、やはり「悔しい」という感情を持っているととらえていることを表す。これに対して、感情を表す「～む」動詞は、主体がある対象に対してどのような感情を抱くのかについて述べるものである。つまり、(13) b は、「彼女」が「自分の失敗」に対して「悔しい」という感情を抱くことを表す。

参考文献

- 浅野鶴子他（1971）『外国人のための基本語用例辞典 第三版』文化庁
 市川保子（2005）「感情・感覚形容詞、感情動詞」日本語教育学会編『日本語教育事典』大修館書店 p. 202
 王安（2005）「接尾辞「～がる」の機能の再考」『北海道大学大学院文学研究科研究論集 5』北海道大学大学院文学研究科 pp. 241-26
 大曾美恵子（2001）「感情を表す動詞・形容詞に関する一考察」『言語文化論集 第 XXI 巻第 2 号』名古屋大学言語文化部・国際言語文化研究科紀 pp. 21-30
 韓金柱（印刷中）「接尾辞『がる』の意味・用法―様態の『そうだ』と比較して―」『東京外国語大学大学院博士後期課程論叢 言語・地域文化研究 第 16 号』東京外国語大学大学院 pp. 323-336
 黄其正（2004）「～がる」『現代日本語の接尾辞研究』溪水社 pp. 101-117
 国際交流基金（2004）『日本語能力試験出題基準』凡人社
 清水泰行（2007）「心理動詞の格と意味役割の対応・ずれ―『引用構文』における名詞句と引用節の意味関係から―」『日本文藝研究 58 (4)』関西学院大学 pp. 23-39
 寺村秀夫（1982）『日本語のシンタクスと意味 I』くろしお出版
 富田隆行（1997）『続・基礎表現 50 とその教え方』凡人社
 西尾寅弥（1972）『形容詞の意味・用法の記述的研究』国立国語研究所報告 44 秀英出版
 西尾寅弥（1975）「第六章『ぼくは悲しい』けれど『彼女は悲しがる』」『日本文法の見えてくる本』汐文社 pp. 81-96
 森田富美子（1988）「接尾辞『～がる』について」『東海大学紀要 8』東海大学留学生教育センター pp. 1-15
 山岡政紀（2002）「感情描写動詞の語彙と文法的特徴」『日本語日本文学 12』創価大学 pp. 23-54

嫌悪を表す「いや」と「きらい」について

加藤恵梨（名古屋大学留学生センター）

A Semantic Analysis of *iya* and *kirai*

Eri Kato (Education Center for International Students, Nagoya University)

1. 本稿の目的

日本語の形容詞は、プラスの感情を表すものよりもマイナスの感情を表すもののほうが多く存在する。このように、日本語ではマイナスの感情がさまざまな形容詞によって表現されるため、それらの意味の違いについて明らかにする必要があるが、管見の限り、辞書等の記述を見てもそれらの意味の違いが明らかにされているとはいえない。本稿で取り上げる「いや」と「きらい」においても同様である。よって本稿では、『現代日本語書き言葉均衡コーパス』（以下 BCCWJ）を用い、嫌悪を表す「いや」と「きらい」の意味及び意味の違いについて明らかにすることを目指す。

2. 先行研究の記述とその検討

形容詞の意味について網羅的に記述されている『現代形容詞用法辞典』で「いや」と「きらい」を見ると、以下のように記されている。

「いや」：物事や人を不快に感ずる様子を表す。主観的な嫌悪を表し、はっきりした理由はないことが多い。（p.66）

「きらい」：嫌悪を表す最も一般的な語の一つ。（p.198）

また、両語の意味の違いについては次のように記されている。

「いや」は「きらい」に似ているが、「きらい」が一般的な嫌悪感を表すのに対して、「いや」が表す不快感はその場かぎりのものであることが多く、主体や対象を限定しない。また、「きらい」は自分以外の人についても言い切りで用いられる点が「いや」と異なる。（p.67）

上の『現代形容詞用法辞典』の記述に「『いや』が表す不快感はその場かぎりのものであることが多い」とある。確かに下の例(1)はその場かぎりの感情であるということが出来るが、例(2)は「毎日朝起きて」とあるようにその場かぎりの感情であるとは考えにくい。

(1) 金属アレルギーは生まれつきのもの。

その金属に対して肌が「嫌だ！」って反応してしまうんです。（Yahoo!知恵袋）

(2) 毎日朝起きて出勤するのが嫌 （Yahoo!知恵袋）

また、「いや」は「主観的な嫌悪感」を表し、「きらい」は「一般的な嫌悪感」を表すという説明では、「いや」と「きらい」の意味を説明したとはいえないため、BCCWJ の例

を基に両語の意味について再考する必要がある。

3. 意味分析

以下では、BCCWJ を用い、両語の意味及び意味の違いについて考察する。

3.1 「いや」の意味について

3.1.1 別義1：〈五感に受けた刺激を反射的に回避しようとする〉〈さま〉

- (3) トマトの酸味がいやというのは、味覚からいえば普通ですよ。

大人とは味覚が違うので、ピーマンの苦味が嫌なのも普通のことです。

(Yahoo!知恵袋)

- (4) 嘉手納を例にとりますと、爆音がいやだと地主が逃げ出してしまうと、それをよいことに国が虫食い状態のように買い漁って、国有地が増えていくような気がします。

(『沖縄の基地問題』)

- (5) (猫の死体の) 見た目が嫌なら段ボールにでも入れておいて下さい。

(Yahoo!知恵袋、()内は引用者)

- (6) 小学校の頃に、父にいわれてサンマを焼いたことがあります、生臭いにおいが嫌だ
なという思いもあったし、(後略)。 (『歯医者で虫歯が治らない理由』)

- (7) めれてしまうと、べたつく感じがいやなので、ジェルなどは使わないです。

(Yahoo!知恵袋)

上の例(3)から(7)はそれぞれ、味覚、聴覚、視覚、嗅覚、皮膚感覚に受けた刺激を、反射的に回避しようとするさまを表す。そのためここでの表現は「この味がいやだ」というような言語表出だけではなく、しばしば「顔をしかめる」、「耳をふさぐ」、「鳥肌がたつ」といった無意識的な表出動作を伴うことが BCCWJ で確認できる。

3.1.2 別義2：〈ある行動をする、あるいは受け入れることを拒む〉〈さま〉

- (8) 毎日朝起きて出勤するのが嫌 (= (2))

- (9) 地方転出については、私は切にお止め申したけれども、自分が東京にいるのがいやで、
熱烈に地方転出を志望されたのである。 (『「文芸春秋」にみる昭和史』)

- (10) 他人に束縛されるのが嫌

安月給でこき使われるのが嫌 (Yahoo!知恵袋)

- (11) 私はなんだかんだと客に声をかけられるのがいやなので、あえてカウンターに座れないようにしたのだ。 (『赤い電車は未知への一步でした』)

「いや」の別義2は、例(8)や(9)のように「出勤する」あるいは「東京にいる」といったある行動をすること、または例(10)や(11)のように「他人に束縛される」、「安月給でこき使われる」あるいは「声をかけられる」といったある行動を受け入れることを拒むさまを表す。

3.1.3 「いや」の意味のまとめ

以上のように、「いや」は五感によって感知した刺激を反射的に回避しようとする、生得的な（広義の）感情に基づく嫌悪感を基本義とする。さらに、ある行動をする、あるいは受け入れるという気持ちが生じず、それを拒むという心理的な拒絶感も表すということができる。

3.2 「きらい」の意味について

3.2.1 別義1：〈経験を基に、ある対象に心理的な嫌悪感を抱く〉〈さま〉

(12) (前略)「古文」の授業はおもしろくない、したがって「古文」は嫌い、という図式ができあがってしまうのだと、私は考えています。 (『古典がもっと好きになる』)

(13) ポテトサラダは好き。美味いから。

フランスパンは嫌い。硬いから。 (Yahoo!知恵袋)

(14) 私はひどい運動音痴です。だからスポーツは大嫌い。 (Yahoo!知恵袋)

例(12)は「古文」の授業を受けておもしろくなかったという経験を基に、心理的に「古文」を受け入れることができないというさまを表している。同様に例(13)と(14)においても、「フランスパンを食べると硬かった」あるいは「スポーツをうまくすることができなかった」という経験を基に、心理的に「フランスパン」や「スポーツ」を受け入れることができないというさまを表している。よって、「きらい」の別義1は、経験を基に、ある対象に心理的な嫌悪感を抱くさまということができる。

3.2.2 別義2：〈ある対象をそれが属するカテゴリーの他のメンバーと比較し、心理的な嫌悪感を抱く〉〈さま〉

(15) 比較政治学を学ぶ者として、暇があれば、地球を歩いてきた。新しい人と出会い、風俗や文化のほんの片端を触らせてもらってきた。ベタベタした濃密な人間関係が嫌いだから、こたわりなく多くの国を歩けたのかもしれない。

(『定年後：「もうひとつの人生」への案内』)

(16) 私「すいません」という言葉が嫌いでした。

すまないと思うなら「すみません」と言ってほしかった。 (Yahoo!知恵袋)

(17) 「私、アメリカンは、コーヒーらしくなくて嫌いです。できるだけ濃くて...苦いくらいのものを高校生のころから飲んでいまして、コーヒー中毒と言われているの」

(『関西国際空港殺人事件：長編推理小説』)

まず例(15)は、「人間関係」というもののの中の「ベタベタした濃密な人間関係」に対して嫌悪感を抱いている。これは、「人間関係」というカテゴリーの中で「ベタベタした濃密な人間関係」をそのカテゴリーの他のメンバー（例えば、あっさりした人間関係など）と比較し、「ベタベタした濃密な人間関係」を心理的に受け入れられないさまを表している。同様に例(16)と(17)においても、「言葉」あるいは「コーヒー」というカテゴリーの中で、「すい

ません」あるいは「アメリカン」を他のメンバーと比較することにより、それらの対象を心理的に受け入れられないさまを「きらい」と表わしている。よって、「きらい」の別義2は、ある対象をそれが属するカテゴリーの他のメンバーと比較し、ある対象に心理的な嫌悪感を抱くさまである。

3.2.3 「きらい」の意味のまとめ

以上のように、「きらい」は経験に基づいて、ある対象に対する心理的な嫌悪感を表す。さらに、ある対象をそれが属するカテゴリーの他のメンバーと比較することによって、その対象に対して心理的な嫌悪感を表すということができる。

3.3 「いや」と「きらい」の意味の違いについて

上で述べたように、「いや」は生得的な（広義の）感情に基づく嫌悪感を基本義とし、さらにある行動に対する心理的な拒絶感を表す。それに対し、「きらい」は経験、さらにはある対象が属するカテゴリーの他のメンバーとの比較に基づいて、ある対象に対する心理的な嫌悪感を表すという違いがあるといえる。

4. 今後の課題

今後も日本語に多く存在するマイナスイメージをもつ感情形容詞の意味について考察していきたいと考える。

参考文献

- 莊巖舜哉（監訳）（1996）.『感情心理学』，ナカニシヤ出版.
飛田良文・浅田秀子（1991）.『現代形容詞用法辞典』，東京堂出版.
福田正治（2003）.『感情を知る 感情学入門』，ナカニシヤ出版.
森田良行（1989）.『基礎日本語辞典』，角川書店.
山田忠雄（編）（2005）.『新明解国語辞典 第六版』，三省堂.

話し言葉、書き言葉、そしてオンライン言語をめぐって —日本語全体像を捉える試みへのパイロット・リサーチ—

西村由起子（東洋学園大学人文学部）[†]

On Variation across Speech, Writing and Language Online: A pilot research toward an “overall” approach to Japanese

Yukiko Nishimura (Faculty of Humanities, Toyo Gakuen University)

1. 導入と要約

本研究は、Biber (1988) が英語において追求してきた書き言葉と話し言葉におけるバリエーション研究を発展させ、日本語における話し言葉、書き言葉、そして、次第に広範化しているオンライン言語のバリエーションの態様を明らかにすることを目的とする。特に今日のコンピュータコミュニケーション上で用いられるオンライン言語が、話し言葉・書き言葉との計量的対比においてどの様に異なるのかに注目する。その意味で、丸山・田野村(2007)が指摘する「大局的視点」からのアプローチとなっている。

話し言葉、書き言葉そして、オンライン言語のバラエティとして、合計9種類の言語データを使用し、どの種類にも一貫して適用可能な、McCab および Unidic による形態素解析を適用する。それぞれにおける形態素の品詞別分布と、更に助詞・助動詞についてはその下位区分における形態素分布を取り出し比較分析する。

結果は二つに要約できる。第一に、「小説」以外のジャンルでは、形態素分布を見ると、オンライン言語と書き言葉にジャンル別のバリエーションが存在する。そして最も話し言葉に近いジャンルから、最も書き言葉的なジャンル（学術論文）まで、形態素分布が漸次的連続的に変化している。特にオンライン言語は話し言葉と書き言葉の両極の間にあることが明らかになっている。第二に、これとは対照的に「小説」のジャンルでは、オンラインのケータイ小説と印刷物に掲載された小説が形態素分布で見ると非常に似通っており、小説以外のジャンルとは際違った対照を示している。

2. 背景

日本語の計量分析研究は国立国語研究所による幾多の先駆的実証研究がある。宮島ら(1982)によって編まれた『図説日本語』は、それまでの日本語計量研究を網羅しその成果を伝えている。その出版から30年近く経た今日、コンピュータ技術の発展・普及はめざましく、研究者の言語研究にとってコンピュータが不可欠であることにとどまらず、一般話者にとっても、コンピュータを介して行う言語活動は多様な言語活動の中で重要な位置を占めている。しかしながら、日本語話者がコンピュータ上で使用する言語についての研究は、従来から蓄積されてきた話し言葉、書き言葉研究に比較すると、はるかに限られていると言わざるを得ない。

オンライン言語研究は、日本語より英語においてより早く、1990年代から行われてきている。オンライン言語と話し言葉・書き言葉との計量比較研究も、英語での研究(Yates 1996, Collot and Belmore 1996 等)が先行している。本研究は、オンライン日本語と話し言葉・書

[†]yukiko.nishimura@tyg.jp

き言葉との計量比較を行った Nishimura (2008a) 第4章を拡大、敷衍したものである。いわば『図説日本語』が今日改訂されることがあれば、話し言葉・書き言葉に加えて、当然含められるべきオンライン言語に関する項目の一端となり得るものである。

筆者はコンピュータ上の日本語を研究対象とし、オンライン言語のなかでも、インターネット掲示板（以下、BBS と略す）への投稿資料にみられる言語特徴（西村 2001, Nishimura 2003, Nishimura 2007a）や投稿者の言語行動（西村 2007b）を調査してきた。インターネットコミュニケーション上の言語資料は「書かれる話し言葉」と特徴づけることができ、書き手が語りかける口調を用い、話し言葉らしく読ませる工夫が随所に見られる。そうしたオンライン日本語資料が実際の話し言葉・書き言葉とどの様に異なるかに関する計量比較は、Nishimura (2008a) において行ったが、本稿では、そこで使用されたデータに加えて、『現代日本語書き言葉均衡コーパス』（以下 BCCWJ と略す）に含まれている、ヤフー！知恵袋（以下知恵袋と略す）で使用された言語も分析する。この追加によりオンライン日本語をより多面的に明らかにすることが出来るからである。更にオンライン日本語は「書かれる」言語である事に注目し、Nishimura (2008a) で扱った書き言葉データ（雑誌記事）に加えて、異なるジャンル（新聞社説、学術論文）における書き言葉との比較を行う。これにより、より幅広い観点からオンライン日本語を位置づけることが可能になる。更に、小説というジャンルも取り上げ、特に従来学術研究対象とされてこなかったケータイ小説を分析対象のオンライン言語に含める。「ケータイ」というコンピュータメディアで制作・消費されるケータイ小説と、従来からある、編集等のプロセスを経て紙媒体に印刷・出版後に購読される小説とがどのように似ているのか、違っているのかを比較し、更に小説以外のジャンルとも比較する。

具体的なデータの説明は次節で行うが、この度構築中の BCCWJ には、オンライン言語も、出版された小説も含まれているため、筆者の行ったオンライン言語との比較研究には、極めて有用である。こうしたコーパスの多様な言語サンプルに、更に新しい視点の言語データを加えて相互関係を分析することで、日本語全体像に至る大局的視点のアプローチへの一つのステップとしたい。

3. データ

本研究では9種類のデータを比較分析する。話し言葉データ1種類、([1]会話)、オンライン言語データ4種類 ([2]2ちゃんねる、[3]ヤフー！BBS、[4]ヤフー！知恵袋（以下「知恵袋」と略す）、[5]ケータイ小説）、書き言葉データ4種類 ([6]印刷小説、[7]雑誌記事、[8]新聞社説、[9]学術論文) で、以下にそれぞれについて説明する¹。

[1] 話し言葉データ：会話

本研究で使用した話し言葉データは、宇佐美(2005)監修『BTSによる多言語話し言葉コーパス—日本語会話1（日本語母語話者同士の会話）』で、主に大学生による対面会話の録音を言語研究のために転記されたものである。延べ1298分に及ぶ録音が宇佐美の開発によるシステムで転記されている。主に大学生の参加者が同性の相手とペアで日常的な話題について話している。このデータは、時間で約10.2%の電話会話と、9.2%の大学教師—学

¹ このうち、[1]会話、[2]2ちゃんねる、[3]ヤフー！BBS、[7]雑誌記事は、Nishimura (2008a)において使用したものである。

生間の論文指導における会話を含む。

[2] オンライン言語データ：2ちゃんねる BBS <<http://2ch.net/>>

2ちゃんねるの多数のボード・スレッドから、「映画」と「英語・留学」について議論しているスレッドを対象とした。このトピックの選択は、Nishimura (2008a)第5章以降でBBS参加者の言語行動分析に使用したことによる。メッセージの投稿時期は2003年6月で、映画トピックと英語トピックとで、それぞれ約3000メッセージを収集した。

[3] オンライン言語データ：ヤフー！BBS

2ちゃんねるからのデータと、比較可能にするため、映画スレッドと英語・留学スレッドから、2ちゃんねるのデータとほぼ同量の投稿メッセージを収集した。

[4] オンライン言語データ：ヤフー！知恵袋

本稿で新たに分析対象に加えた「知恵袋」からは、[3]と同様、映画と英語や留学および趣味に関する質問・回答を扱った。大カテゴリ「エンターテインメントと趣味」、と「教養と学問、サイエンス」から合計約2500組の質問と回答を収集した。

[5] オンライン言語データ：ケータイ小説

この種類では、広く読まれ人気のある作品として、ケータイ小説大賞受賞作品や、それらに匹敵するベストセラー作品計7点を選別した。著者、タイトル、などは、以下に記す²。

著者	タイトル	発行年	分析ページ	備考
1 十和	『クリアネス』	2006	1-269 (全)	「第1回日本ケータイ小説大賞」
2 美嘉	『恋空』	2006	1-200	ベストセラー
3 メイ	『赤い糸』	2006	1-200	ベストセラー
4 Kiki	『あたし彼女』	2008	1-430 (全)	「第3回日本ケータイ小説大賞」
5 ユウ	『ワイルドビースト』	2008	1-200	サイト連載時 ファン投票ランキング1位
6 パープル	『あしたの虹』	2008	1-100	瀬戸内寂聴著、連載時、ランキング1位
7 繭	『風にキス、君にキス』	2009	1-309 (全)	「第4回日本ケータイ小説大賞」

[6] 書き言葉データ：印刷小説

本研究でケータイ小説と比較のために使用される印刷小説は、ケータイ小説が若い年代の著者により最近書かれていることから、なるべく若い作家による最近の作品、という方針でBCCWJから以下の方法で収集した。まず、BCCWJ書籍サンプル情報より、書籍のNDCコード913(小説)とされた2640サンプルを選んだ。次にサンプル情報のC-コードにより、第1桁目が「8」である児童書は除外し、更に、第3・4桁目が、93(日本文学、小説・物語)以外の作品は除外した。その結果、最も若い1980年代生まれの作家による作品は8サンプル全て(2004,2005年発行)、1970年代生まれの作家による作品は2003年以降に発行の32サンプル、1960年代生まれの作家による作品は、2005年発行の48サンプル、合計88サンプルを選択した。

² これらの作品には以下のURLからアクセス可能である。

1. http://no-ichigo.jp/read/book/book_id/219907;
2. http://ip.tosp.co.jp/BK/TosBK100.asp?I=hidamari_book&BookId=1&SPA=200
3. <http://de-view.net/book/novel.php?aid=1&f=chapter&tid=1&%20pos=1>
4. http://no-ichigo.jp/read/book/book_id/148920
5. <http://ip.tosp.co.jp/bk/TosBK100.asp?I=WildBeast&BookId=1>
6. 現在非公開中
7. http://no-ichigo.jp/read/book/book_id/240186

[7] 書き言葉データ：一般雑誌記事

1996年から2007年の間に発行された『スクリーン』、『プレミア』、『英語教育』、『留学ジャーナル』、『週刊朝日』ビジネス雑誌数点で、合計文字数は456416文字である。2ちゃんねるとヤフーBBSで扱われているトピックと同種のトピックを扱っていることが雑誌記事の選択基準となっている。この中には、インタビュー記事も一部含む。

[8] 書き言葉データ：全国紙新聞社説

2005年1月、同年6月発行の『毎日新聞』、2006年10月発行の『朝日新聞』、2007年8月発行の『日本経済新聞』、および2009年12月—2010年1月発行のこれら3紙に掲載された社説合計36375語を使用した。特定の1紙に偏らないよう3紙を選び、発行時期に関してはなるべく分散させることを意図した。これらデータ収集に際して、『朝日新聞』と『日本経済新聞』はオンラインアーカイブを利用し、『毎日新聞』はCD-ROM版を利用した。

[9] 書き言葉データ：人文社会科学系学術論文

学術論文データとしては、査読付きジャーナルに掲載された論文、査読なしで紀要等に発表された論文、及び博士論文が含まれる。分野は人文(言語・教育)系や、社会科学(経済学)系で、合計41292語が収集された。数式、用例、等は除外した。

以上のように収集されたデータの分析手方法と結果を次セクション以降で述べ、考察する。

4. 分析手法

上記9種類の言語データを分析対象として、「茶まめ」(Version1.61)を用いて形態素の解析を行った。具体的には、「茶まめ」のオプションとして、どの種類にも一貫して適用可能な、MeCabおよびUnidic(Version1.3.11 Windows版)を利用した。9種類それぞれの言語データにおける形態素の品詞別分布を導出し、加えて助詞、助動詞についてはジャンルの特徴を見る上で有用であるためその下位区分における形態素分布を導出した。

各データは、形態素解析を行う前に、「クリーン」する必要がある。言語使用の実態として、文章を対象とするので、各データから、章やセクションのタイトル、図などのキャプション、見出し、ページ番号等は除外した。オンライン言語データでは、メッセージ番号や投稿時間など、メッセージ以外の部分は除外した。更に、全てのデータにおいて、言語データのみを対象とするため、句読点、括弧、他、記号類も形態素の集計から除外した。

品詞の欄に「接辞」(「超一」「一的」等)が含まれている。「接辞」については、接頭辞が付加される主要語により品詞が異なり、何らかの品詞に統合することができないので、MeCabの形態素解析により接頭辞・接尾辞とされた形態素を合算して「接辞」とした。

いったん解析を行った後、品詞認定で誤解析が散見されることがあり、その場合は、手作業で修正した。特に非標準的表記で書かれることの多いオンライン言語では誤解析が残っている可能性がある。

5. 結果

全体結果は、表1[品詞全体]、表2[助詞下位区分]、表3[助動詞下位区分]、小説に関する結果は表4[品詞全体]、表5[助詞下位区分]、表6[助動詞下位区分]にまとめられている。

表1. 各ジャンルにおける形態素分布:品詞全体

	小説以外							小説	
	話し言葉	オンライン言語			書き言葉			オンライン	書き言葉
品詞	会話	2ちゃんねる	Yahoo BBS	知恵袋	雑誌記事	新聞社説	学術論文	ケータイ小説	印刷小説
名詞	17.2%	28.9%	26.9%	29.1%	31.4%	37.4%	39.5%	22.8%	23.6%
接辞	2.7%	3.5%	3.5%	3.9%	4.1%	5.9%	5.8%	2.7%	3.1%
動詞	10.8%	12.5%	13.5%	13.6%	14.1%	12.8%	13.1%	15.7%	15.8%
形状詞	1.1%	1.3%	1.5%	1.2%	1.6%	1.3%	1.6%	1.6%	1.5%
連体詞	1.1%	0.7%	0.8%	0.7%	1.1%	0.6%	1.1%	1.0%	1.0%
接続詞	0.8%	0.3%	0.3%	0.3%	0.5%	0.2%	0.7%	0.3%	0.4%
助詞	32.7%	32.4%	32.4%	32.0%	32.6%	30.9%	28.7%	32.6%	33.3%
助動詞	12.6%	13.0%	14.2%	13.5%	9.2%	8.1%	6.6%	12.8%	12.7%
代名詞	4.2%	1.7%	1.7%	1.5%	1.5%	0.5%	0.7%	3.7%	2.9%
形容詞	2.7%	2.6%	2.0%	1.9%	1.7%	1.2%	1.1%	2.8%	2.3%
副詞	5.0%	2.5%	2.7%	2.0%	2.1%	1.0%	1.0%	3.0%	2.7%
感動詞	9.0%	0.5%	0.5%	0.3%	0.1%	0.0%	0.0%	1.0%	0.7%
計	100%	100%	100%	100%	100%	100%	100%	100%	100%
形態素数	248742	267561	251272	265255	278944	304531	283565	260386	288981

表2: 各ジャンルにおける形態素分布:助詞下位区分

	小説以外							小説	
	話し言葉	オンライン言語			書き言葉			オンライン	書き言葉
助詞	会話	2ちゃんねる	Yahoo BBS	知恵袋	雑誌記事	新聞社説	学術論文	ケータイ小説	印刷小説
格助詞	29.4%	44.2%	48.5%	48.7%	60.4%	65.0%	67.0%	49.3%	52.2%
係助詞	9.9%	15.2%	15.8%	13.7%	15.5%	15.8%	13.5%	14.6%	15.7%
接続助詞	14.7%	15.5%	15.9%	15.9%	14.3%	11.0%	14.6%	17.5%	16.8%
準体助詞	7.0%	6.3%	5.9%	6.3%	3.6%	2.2%	1.2%	4.3%	4.5%
副助詞	15.5%	8.5%	6.1%	7.1%	4.4%	5.3%	3.4%	6.6%	5.5%
終助詞	23.5%	10.3%	7.8%	8.3%	1.8%	0.8%	0.2%	7.8%	5.3%
計	100%	100%	100%	100%	100%	100%	100%	100%	100%
形態素数	81392	86613	81484	84917	90808	94125	81467	84919	96256

表3. 各ジャンルにおける形態素分布:助動詞下位区分

	小説以外							小説	
	話し言葉	オンライン言語			書き言葉			オンライン	書き言葉
助動詞	会話	2ちゃんねる	Yahoo BBS	知恵袋	雑誌記事	新聞社説	学術論文	ケータイ小説	印刷小説
ダ	39.1%	33.9%	22.2%	20.8%	44.6%	38.7%	46.5%	32.7%	33.5%
タ	18.4%	24.6%	19.6%	17.9%	30.0%	32.0%	23.3%	37.7%	37.6%
ナイ	9.2%	8.5%	5.5%	4.6%	8.1%	9.8%	7.0%	10.3%	7.7%
デス	10.0%	7.7%	18.4%	22.6%	1.8%	0.3%	0.1%	1.7%	3.7%
マス	2.5%	6.2%	18.1%	18.7%	2.1%	0.4%	0.0%	1.0%	3.3%
レル・ラレル	2.3%	3.7%	4.0%	4.1%	8.0%	10.1%	17.6%	4.4%	5.7%
テル・デル	10.7%	8.5%	4.9%	3.9%	0.3%	0.0%	0.0%	5.6%	2.0%
チャウ	1.6%	0.6%	0.5%	0.3%	0.1%	0.0%	0.0%	0.7%	0.4%
ベシ(文語)	0.1%	0.2%	0.2%	0.1%	0.3%	2.2%	0.4%	0.0%	0.2%
その他計	6.2%	6.2%	6.5%	7.0%	4.6%	6.4%	5.1%	5.9%	6.1%
計	100%	100%	100%	100%	100%	100%	100%	100%	100%
形態素数	31449	34792	35659	35881	25708	24659	18758	33444	36705

表4. 小説における「地の文」・「引用文」別形態素分布:品詞全体

品詞	ケータイ小説			印刷小説		
	引用文 (Q)	地の文 (N)	計	引用文 (Q)	地の文 (N)	計
名詞	20.2%	23.5%	22.8%	20.8%	24.7%	23.6%
接辞	3.3%	2.5%	2.7%	3.7%	2.9%	3.1%
動詞	13.6%	16.2%	15.7%	13.8%	16.5%	15.8%
形状詞	1.3%	1.7%	1.6%	1.2%	1.6%	1.5%
連体詞	0.8%	1.0%	1.0%	1.2%	0.9%	1.0%
接続詞	0.4%	0.3%	0.3%	0.3%	0.4%	0.4%
助詞	31.1%	33.0%	32.6%	33.5%	33.2%	33.3%
助動詞	14.6%	12.4%	12.8%	13.7%	12.3%	12.7%
代名詞	4.7%	3.4%	3.7%	4.1%	2.4%	2.9%
形容詞	3.2%	2.7%	2.8%	2.6%	2.2%	2.3%
副詞	3.1%	2.9%	3.0%	3.1%	2.6%	2.7%
感動詞	3.6%	0.4%	1.0%	1.9%	0.2%	0.7%
計	100%	100%	100%	100%	100%	100%
形態素数	52602	207784	260386	80491	208490	288981
Q・N割合	20.2%	79.8%	100%	27.9%	72.1%	100%

表5. 小説における「地の文」・「引用文」別形態素分布:助詞下位区分

助詞	ケータイ小説			印刷小説		
	引用文 (Q)	地の文 (N)	計	引用文 (Q)	地の文 (N)	計
格助詞	30.8%	53.7%	49.3%	40.4%	56.9%	52.2%
係助詞	12.2%	15.1%	14.6%	14.4%	16.2%	15.7%
接続助詞	18.1%	17.3%	17.5%	16.6%	16.9%	16.8%
準体助詞	6.7%	3.7%	4.3%	6.7%	3.6%	4.5%
副助詞	8.6%	6.1%	6.6%	7.5%	4.7%	5.5%
終助詞	23.6%	4.1%	7.8%	14.3%	1.7%	5.3%
計	100%	100%	100%	100%	100%	100%
形態素数	16373	68546	84919	26966	69290	96256
Q・N割合	19.3%	80.7%	100%	28.0%	72.0%	100%

表6. 小説における「地の文」・「引用文」別形態素分布: 助動詞下位区分

助動詞	ケータイ小説			印刷小説		
	引用文 (Q)	地の文 (N)	計	引用文 (Q)	地の文 (N)	計
ダ	35.8%	31.7%	32.7%	36.5%	32.2%	33.5%
タ	24.4%	41.7%	37.7%	22.4%	44.1%	37.6%
ナイ	10.7%	10.2%	10.3%	9.1%	7.1%	7.7%
デス	5.1%	0.7%	1.7%	8.9%	1.4%	3.7%
マス	2.9%	0.4%	1.0%	5.4%	2.4%	3.3%
レル・ラレル	2.7%	4.9%	4.4%	3.9%	6.5%	5.7%
テル・デル	10.0%	4.2%	5.6%	5.5%	0.5%	2.0%
チャウ	1.2%	0.5%	0.7%	1.1%	0.1%	0.4%
ベシ(文語)	0.0%	0.0%	0.0%	0.2%	0.2%	0.2%
その他計	7.1%	5.7%	5.9%	7.1%	5.5%	6.1%
計	100%	100%	100.0%	100%	100%	100.0%
形態素数	7690	25754	33444	11063	25642	36705
Q・N割合	23.0%	77.0%	100.0%	30.1%	69.9%	100.0%

6.1. 小説以外のジャンルの考察

小説以外のジャンルで品詞使用頻度分布を見ると(表1)、名詞、接辞、動詞の使用頻度が、話し言葉からオンライン言語を経て書き言葉に至るまでほぼ一貫して上昇していった

いる事が分かる。例えば名詞をとると、話し言葉では 17.2%、オンライン言語においては 26.9% - 29.1%、書き言葉では、31.4% - 39.5%となっている。

これとはちょうど逆に、感動詞、形容詞、副詞、代名詞では、使用頻度が、話し言葉、オンライン言語、書き言葉の順に下降していく。感動詞については、話し言葉の 9.0%が突出しており、オンライン言語で 0.5% - 0.3%、書き言葉では 0.1% - 0% となっている。形容詞、副詞においても、下降傾向に程度差はあるものの、それぞれ話し言葉からオンライン言語を経て書き言葉へと一貫して使用頻度が減っている。

この結果は、小説以外のジャンルでは、最もインフォーマルで即興的に発話される仲間内の話し言葉から、最もフォーマルで編集を経た情報量の多い学術論文まで、漸次変化する「連続体」と表現できることを含意している。そしてオンライン言語は話し言葉と書き言葉の中間であり「書かれた話し言葉」という性格を持っていることを示唆する。

ここで言う「連続体」の軸は、Biber (1988) の Multi-feature/multi-dimensional (MF/MD)モデルの Dimension 1: Informational versus Involved Production として考える事ができる。Biber (1988) は、この dimension について以下のように説明している。

The poles of this dimension represent discourse with interactional, affective, involved purposes, versus discourse with highly informational purposes, which is carefully crafted and highly edited. This dimension is very strong and represents a fundamental parameter of variation among texts in English. (p.115)

この dimension を本稿が対象とする言語コーパスの文脈で解釈してみよう。言語使用の目的には、「情報」や「情緒」を伝える事が含まれる。とすると、本稿の対象データでは、Biber の dimension は「情緒共有志向」から「情報伝達志向」へ向かう軸と考える事が出来る。

名詞の使用は「情報」の伝達が重要な役割であると解釈可能である。とすると情報を伝えることがより重要になる書き言葉で多くなることが予想される。これに対し、感動詞は「情緒」を伝えると考えるのが自然である。とすると、感情や雰囲気、ラポートを伝えるのがより重要となる会話では、感動詞が多くなることになる。会話における感動詞は、「うん」等の「あいづち」を含んでおり、会話管理の上で不可欠である（メイナード 1993）。ただ無意識的に発せられることが多いため、意識の上に上がりにくい。その為、能動的意識的になされる「書かれた話し言葉」であるオンライン言語ではそれほど多く出現しなかったと推測される。更に、形容詞・副詞については、会話でより重要となる気分や雰囲気を伝える形態素がこの品詞と解析されたと考えられ、又、直示語である代名詞に関しても、使用された時点・地点を基準にして指示される物事が決まることから、その頻度が会話で多いことは想像に難くない。このように、実証分析の結果は、Biber の MF/MD モデルからの予想と整合的である。

更に、助詞と助動詞についての結果も、Biber のフレームワークで説明可能である。まず助詞を見てみよう（表 2）。

一方で格助詞では、会話から学術論文に向かって、一貫して頻度が上昇している。会話の 29.4%から、オンライン言語で 44.2%-49.3%、書き言葉で 52.2%から、学術論文では、67%に至る。格助詞とは文中で名詞につき、その名詞と他の名詞との文法関係を示すもので、名詞（情報）の厳密な伝達に必要となり、書き言葉での分布が高くなることが考えられる。

他方終助詞と副助詞は、頻度は逆の下降パターンを示す。会話で終助詞：23.5%、副助詞：15.5%から、オンライン言語で終助詞：7.8% - 10.3%、副助詞：6.1% - 8.5%、書き言葉では終助詞：0.2% - 1.8%、副助詞：3.4% - 4.4%と頻度は下降している。終助詞は文や句の終りで用いられ、疑問・感動などを表し、「情緒」を伝える際には有用である。従って会話で多く使用され、オンライン言語でも、会話ほどではないにせよ、気分や雰囲気を伝えることにもウェイトが置かれるため、BBS でも多い。副助詞とは、『広辞苑』によると「体言・副詞などにつき、上の語に副詞の性質と職能とを帯びさせて下の用言の意味を限定する助詞」である。副詞的な機能を持ち、副詞が情緒を伝える場合でより多く使われる事から、品詞全体の場合と同様に、副助詞は会話から学术论文へ頻度が下降していると考えられる。

準体助詞も、終助詞、副助詞と同様に、会話から学术论文へと頻度が下降傾向にある。準体助詞とは、「の」に代表される一種の代名詞のような働きをする助詞である（例えば「大きいのを買った」における「の」）。従って情報を厳密に伝えることが重要なジャンルでは避けられると考えるのは自然である。このように書き言葉では情報を伝えることが重要であるため、終助詞、副助詞、準体助詞は低い水準となっていると考えられる。

次に、助動詞を見てみよう（表3）。品詞や助詞のレベルで観察された、頻度の上昇・下降パターンが見られる助動詞をまず挙げ分析する。その後でいくつかの助動詞の頻度分布から個々のジャンルに関して解釈可能な事柄を述べることにする。

まず、「テル・デル」という助動詞には、会話から学术论文へと、頻度の下降パターンが見られる。「テル・デル」は、「食べてる」、「遊んでる」のように、状態を表すが、書き言葉にはふさわしくないと考えられている。少なくともパソコン上で書く場合、マイクロソフトワードのスタイルチェッカーは「い抜き」表現として、書き言葉には不適切、との警告を発する。従って書き言葉では雑誌記事ではインタビュー記事も含まれるので非常に低いパーセンテージで出現してはいるが、新聞社説と学术论文では部内の編集を経た文章だけに、ゼロである。これに対してオンライン言語では、話し言葉程高くはないが出現しており、これはオンライン言語が「書かれる話し言葉」という話し言葉と書き言葉の中間形態である事を示唆している。

「テル・デル」とは逆に、「レル・ラレル」は書き言葉で頻度が高く、オンライン言語や話し言葉で低くなっている。「レル・ラレル」には、可能、尊敬、受け身などの用法があるが、書き言葉でも、学术论文で 17.6%と特に高いのは、客観的な響きを持たせるため、受け身の用法を用いるのではないかと考えられる。

個別ジャンルの特徴についてもいくつかの知見が得られている。まず、丁寧体「デス」の頻度分布を見ると、知恵袋で 22.6%と、非常に高くなっている。白井 (2009) は知恵袋の「デス」について、回答の最後に「デス」が多用されることを報告し、「「デス」で回答することが正しい形式と考えられている」(p.46)と述べている。「デス」の知恵袋における高水準がこの説明による理由だとすると、丁寧体「マス」の数値が、ヤフーBBS (18.1%) と知恵袋 (18.7%) とではほぼ同水準であることから、知恵袋とヤフーBBS では同程度の丁寧体助動詞を用いているのではないかと考えられる。「質問」と「回答」から成るコミュニケーションが行われる知恵袋ではあっても、このコミュニティがともにヤフーに属しており、利用者層も重なっている可能性は高い。西村 (2007b) ではヤフーBBS の投稿者は対人配慮が読み取れるポライトな行動をとっている事が示されたが、知恵袋においても、ヤフーBBS と同様に、投稿者が潜在的回答者となる読み手に対してポライトな行動をとっていることを示唆している。

これに対して、2ちゃんねるは同じオンライン言語ではあるが、丁寧体使用率が低く、知恵袋やヤフーBBSと非常に異なっている。この違いは、何らかの話題について意見を述べたり、投稿することを楽しむ2ちゃんねる掲示板は、ヤフーBBSと異なる性格を持つインターネットコミュニティと考えられる。スペースの関係上ここでは指摘だけにとどめ、詳細はNishimura(2008b, 2010)を参照されたい。

丁寧体「デス」に対する普通体「ダ」は、丁寧体比率の高いヤフーBBSと知恵袋では低く、話し言葉、書き言葉で全般に高い出現率である。これは、話し言葉、書き言葉とでは異なる理由から来ていると思われる。まず書き言葉においては、大人の書き言葉において、ダ体が無標であることに起因する。話し言葉においては、インフォーマルな親しい人間関係のコミュニケーションでは、丁寧体を用い距離をおくコミュニケーションは適切ではなく、距離を置かないスタイルがふさわしいと場合が多いため、と考えられる。

「新聞社説」で「当然」や「命令」等を表す「ベン」の頻度が2.2%と高く、他の全てのジャンルで0%台であるのと大きく異なっている。この数字は、新聞社が団体、個人などに行動を促すなどの、『主張』をしていることから来ていると考えられる。これはあるジャンルに特徴的に出現する助動詞が計量分析から明らかにされたケースとなっている。

6.2. 小説ジャンルの考察

前節では、小説以外のジャンルを見てきた。次に小説ジャンルを見ることにする。

表1から表3を見るとまず二つの事が分かる。第一に、品詞、そして助詞助動詞の下位区分の頻度分布をみても、オンラインの「ケータイ小説」と「印刷小説」はよく似ている。第二に、小説ジャンルの中でのオンラインと印刷媒体の差は、小説以外のジャンルでのオンラインと印刷媒体の差より、一見しても遙かに小さい。

しかしながら小説には、小説以外のジャンルにはない、登場人物の「会話」を引用した「引用文」と、作者がストーリーを語る「地の文」の複合体という性格がある。その為に、話し言葉に近い部分と、書き言葉に近い部分の混在が、品詞等の頻度分布の結果に影響を与えている可能性がある。この点を識別するため品詞等の頻度分布分析を「引用文」と「地の文」に区別して行う³。(表4[品詞全体]、表5[助詞下位区分]、表6[助動詞下位区分])。

まず、ケータイ小説と印刷小説の間になる微妙な違いは、前者では「地の文」で会話体に近いスタイルがより多く用いられているのに対し、後者の「地の文」は、より書き言葉に近い、という点にある。

会話的特徴を表す感動詞を見てみよう(表4)。印刷小説の「引用文」では1.9%の頻度に対しケータイ小説の「引用文」では3.6%である。但し実際の会話9.0%と比べると遙かに少ない。従ってケータイ小説は、より会話的にしようという意図があるものの、実際には感動詞の頻度は高くない。この事は「書かれた話し言葉」というオンライン言語の性格をよく表している。

会話に近くしたいというケータイ小説の傾向は「地の文」にも現れている。印刷小説では「地の文」では感動詞が0.2%という雑誌記事に近い水準に対しケータイ小説では「地の文」でも2倍の0.4%で、感動詞が相対的に多く出現している。

次に、終助詞を見る(表5)。ケータイ小説の「引用文」での頻度23.6%は、実際の会話での数値23.5%(表2)にほぼ等しいのに対し印刷小説の「引用文」では14.3%と大きく減

³樺島・寿岳(1965)においては、小説中の「引用文」比率を、文体を特徴づける指標の一項目としている。

少する。「地の文」を見ると、印刷小説においては 1.7%という雑誌記事に近い水準であるのに対し、ケータイ小説では 4.1%と高い。このようにケータイ小説では、登場人物どうしの会話ではない「地の文」において、あたかも著者が読者に語りかけるかのように、終助詞を使用していると言える。

最後に話し言葉的特徴の「テル・デル」に注目する（表 6）。ケータイ小説では登場人物の会話を引用した「引用文」で、実際の会話（10.0%）とほぼ同レベルの頻度（10.7%）で使用されている。このことは、ケータイ小説の場合、著者の意識的選択として、実際に話すことばを登場人物にも語らせている、といえるのではないかと考えられる。（この点は、感動詞の場合と異なる。感動詞は無意識に発せられ、そのため意識に上ることがない。そのためケータイ小説での頻度は実際の会話の水準に達しなかったのである。）さらに、「地の文」でもケータイ小説の「テル・デル」の出現頻度(4.2%)はヤフーBBSでの出現頻度(4.9%)並であり、会話的である。これと対照的に、印刷小説では「テル・デル」の出現頻度は、「引用文」でも実際の会話の半分程度(5.5%)、「地の文」では、雑誌記事(0.3%)の水準にまで落ち(0.5%)、低い数値となっている。このような結果から、ケータイ小説・印刷小説においてもともに、「引用文」、「地の文」で異なるスタイルが用いられるが、その差はケータイ小説のほうが「地の文」においてさえ会話的であり、印刷小説のほうが、「地の文」が書き言葉の性格をより多く示している。

次に以下において、(1)ケータイ小説と印刷小説の「地の文」と「引用文」に関する微妙な違いは何に起因しているのか、(2)ケータイ小説と印刷小説とでは、全体としてなぜそれほど違わないのかを考えることにする。

(1)に関してはそれぞれの製作・流通過程に要因があるのではないかと考えられる。印刷小説が編集のプロセスを経て印刷・出版後に読者に届くのに対して、ケータイ小説は著者がケータイサイトにアップロードし、読者は各自のケータイで読む、という基本的流通経路がある。ケータイ小説作家は、読者との距離が近く（吉田 2008）、読者からのフィードバックを直接受け、ケータイメールを書くかのように、読者に受け入れられやすいスタイルを「地の文」でも用いている、と考えられる。印刷小説ではこのような作家・読者間の相互作用は少ないと考えられ、編集者が出版社の方針や、書き言葉の規範に基づき、著者のスタイルを変更する場合もあり得る。「引用文」はともかく、「地の文」では両者におけるこれらの製作・流通過程が微妙な違いに関連していると考えられる。

(2)に関しては、本稿での印刷小説のサンプリングが関わっていると思われる。今回 BCCWJ から選択された印刷小説は著者生年代をなるべくケータイ小説作家のそれと近い、最近出版された作品にそろえた。ということはこの中に、印刷小説の中でも「ライトノベル」や「ジュニア小説」と呼ばれる青少年を読者対象とした若手作家による作品が含まれていることが予想される⁴。そのような作品や、若者向け雑誌などは「新言文一致体」（佐竹 1995）と呼ばれる、「仲間に向かって話すように書く」文体で書かれていることが報告されており（メイナード 2005）、本稿で分析したケータイ小説・印刷小説においてもともに「新言文一致体」が用いられていることは想像に難くない。BCCWJ 小説でも、ケータイ小説作家とは離れた年上の作家による作品や、やや古い作品を参照小説とすると、ケータイ小説との差はかなり大きくなる事が考えられる。

⁴ 実際含まれており、例えば、鏡貴也による『伝説の勇者の伝説』（2003 年、富士見書房、Sample_ID PB39_00315）や三雲岳斗による『ランブルフィッシュ』（2004 年、角川書店、Sample_ID PB49_00337）が榎本秋（2008）においてライトノベルとして論じられている。

7. 結語

本稿で使用した9種類の言語サンプルの品詞とその助詞・助動詞ではその下位区分の形態素分布により、話し言葉、書き言葉という対比において、インターネット上の言語が「生起率の連続的変化」(グラデーション)(前川 2008 p.25)を成していることが示され、話し言葉、書き言葉の中間に位置する事を計量的に明らかにした。またオンライン言語にも書き言葉にもバリエーションがあり、オンライン言語では、2ちゃんねる、ヤフーBBS、知恵袋、書き言葉では雑誌記事、新聞社説、学術論文との間の違いについても、計量的に明らかにした。小説のジャンルでは、オンライン言語と書き言葉のそれぞれに属するサンプルを取り上げ、他のジャンルと同様に比較分析した。これらは、メディアは異なっているものの、「小説」というジャンルでの共通性が大きいと見られ、形態素分布ではほとんど差がなかった。しかし、「引用文」・「地の文」の区別においてこの二者が区別され、ケータイ小説においては、「地の文」でも「引用文」に近い会話スタイルが多く用いられ、また、印刷小説では「地の文」ではケータイ小説より書き言葉の特徴を多く示す事が明らかにされた。

今後の課題として、分析対象データと方法論について述べる。本稿では、9種類のジャンルを比較分析の対象とした。これを英語での計量研究に照らすと、Biber (1988) は自身の開発した MF/MD モデルにおいて、書き言葉 17 ジャンル、話し言葉 6 ジャンル、計 23 のジャンルを分析(p.67)しており、Collot and Belmore (1996)では、Biber(1988)の MF/MD モデルに基づき更にオンライン言語 2 種類を加え、25 ジャンルを分析している。これらの英語での研究と比べると、本稿での 9 種類は、かなり少ないと言わざるを得ない。日本語において、更に異なるジャンルを加えることが今後の課題となる。将来ブログや学術論文が BCCWJ に含まれる計画があるとのことで(前川・山崎 2009) その構築が待たれる。

次に、今回の分析では、パイロット・リサーチとして形態素分布の傾向を見たにとどまり、統計的な仮説検定を行っていない。ジャンルの数、データ数の少なさから、今回得た結論は今回の分析サンプルに大きく依存している可能性があり、そこから一般的な結論を導くには十分な注意が必要である。

最後に今回の分析では方法論として Unidic/McCab による形態素解析を全データに対して行った。このアプローチは各形態素が属している品詞が明らかになり、また形態素情報を有しているため、利用範囲の広い分析手段である。今後は上でのべた、ジャンルという幅の拡大とともに、それぞれのジャンルで使われる言語特徴と、その共起関係などを含めた分析項目の精査・検討を経て研究目的にかなった方法を開発してゆく必要がある。

[謝辞]

本研究は、文部科学省科学研究費基盤研究 (C)「インターネット上の日本語及びその話者の言語行動に関する社会言語学的日英共同研究」(課題番号 21520448)による補助を得た。また、ウェブサイト「魔法の i らんど」広報部、野口恵氏には、当サイト掲載のケータイ小説の本研究での利用に関して快諾をいただいたことを記して謝する。

[文献]

- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge Univ. Press.
Collot, Milena and Nancy Belmore. 1996. Electronic language: A new variety of English. In Susan C. Herring (ed.). *Computer-mediated communication: Linguistic, social and cross-cultural*

- perspectives*. 13-28 Amsterdam: John Benjamins
- 榎本秋 2008 『ライトノベル文学論』 東京: NTT 出版
- 樺島忠夫・寿岳章子 1965 『文体の科学』 京都: 綜芸舎.
- 前川喜久雄 2008 「話し言葉と書き言葉」 『日本語学』 27(5) 23-33.
- 前川喜久雄・山崎誠 2009 『『現代日本語書き言葉均衡コーパス』』 『国文学 解釈と鑑賞』 74(1) 15-25.
- 丸山岳彦・田野村忠温 2007. 「コーパス日本語学の射程」 『日本語科学』 22 5-12.
- メイナード, 泉子・K. 1993 『日英語対照研究シリーズ2 会話分析』 東京: くろしお出版
- メイナード, 泉子・K. 2005 『談話表現ハンドブック』 東京: くろしお出版
- 宮島達夫, 江川清, 真田信治, 野村雅昭, 中野洋, 佐竹秀雄(編). 1982 『図説日本語—グラフで見ることばの姿』 東京: 角川書店.
- 西村由起子 2001. 「書かれる話し言葉--インターネット上に見られる新しいコミュニケーションスタイル」 東洋女子短期大学・東洋学園大学ことばを考える会編 『シリーズ言葉のスペクトル 対話』 27-57. 東京: リーベル出版
- Nishimura, Yukiko. 2003. Linguistic innovations and interactional features of casual online communication in Japanese. *Journal of Computer-Mediated Communication* 9 (1).
<<http://jcmc.indiana.edu/vol9/issue1/nishimura.html>>.
- Nishimura, Yukiko. 2007a. Linguistic innovations and interactional features in Japanese BBS communication. In Brenda Danet and Susan C. Herring (eds.). . *Multilingual internet: Language culture and communication online*. 163-183. Oxford: Oxford University Press.
- 西村由起子 2007b. 「電子コミュニケーションのポライトネス・インポライトネス」 岡本真一郎(編) 『ことばのコミュニケーション—対人関係のレトリック』 81-96. 京都: ナカニシヤ
- Nishimura, Yukiko. 2008a. *Aspects of Japanese Computer-Mediated Communication: Linguistic and Socio-Cultural Perspectives*. Sheffield, UK: Sheffield Hallam University, Unpublished PhD Dissertation.
- Nishimura, Yukiko. 2008b. Japanese BBS websites as online communities: (Im)politeness perspectives. *Language@Internet* 5 Special Issue: Data and methods in computer-mediated discourse analysis. <<http://www.language@internet.de>>.
- Nishimura, Yukiko. 2010. Impoliteness in Japanese BBS interactions: Observations from message exchanges in two online communities. *Journal of Politeness Research* 6 (1). 33-55. Special Issue: Politeness and impoliteness in computer-mediated communication.
- 佐竹秀雄 1995 「若者ことばとレトリック」 『日本語学』 14.11 月号 53-60
- 白井勢津子. 2009. 『『Yahoo!知恵袋』における『です』を用いた引用表現』 特定領域研究「日本語コーパス」平成20年度公開ワークショップサテライトセッション予稿集 39-46.
- 宇佐美まゆみ 2005 (監修) 『B T S による多言語話し言葉コーパス—日本語会話1 (日本語母語話者同士の会話)』 東京外国語大学,
- Yates, Simeon J. 1996. Oral and written linguistic aspects of computer-conferencing: A corpus based study. In Susan C. Herring (ed.). 29-46.
- 吉田悟美一 1908 『ケータイ小説がウケる理由』 マイコミ新書. 東京: 毎日コミュニケーションズ

動作の方向を表す「ーテクル」に対する要求条件の解明

石山友之（熊本県立大学大学院 文学研究科日本語日本文学専攻 博士前期課程）

When is the Direction Expressing Auxiliary Verb “-tekuru” Obligatory?

Tomoyuki Ishiyama (Graduate School of Language and Literature, Prefectural Univ. of Kumamoto)

1. 先行研究と本発表の位置づけ

「ーテクル」については多くの研究がなされている（吉川(1976)、寺村(1984)、坂原(1995)など）。「ーテクル」には、「やってくる／生まれてくる」のように、「動詞のテ形＋クル」で一語になっているものや、「ご飯を食べて（それから）来る」のように「クル」が本動詞として機能するものがある。また、補助動詞としての「ーテクル」には、「空が明るくなってきた／今まで頑張ってきた」のようにアスペクトを表すものと、「ほこりが落ちてくる／あいつがこっちを見てくる」のように動作の方向を表すものがある。

本発表における考察の対象は補助動詞としての「ーテクル」のうち、動作の方向を表すものである。動作の方向を表す「ーテクル」は、省略すると不自然になる場合とならない場合がある。本発表は動作の方向を表す「ーテクル」の使用に関わる条件を探るものである。

2. データ

まず「書き言葉コーパス 2008 年度モニター版」（以下「書き言葉コーパス」）から動作の方向を表す「ーテクル」の用例を集めた。その後それぞれの用例に通し番号、本動詞、本動詞の自・他、「ーテクル」の省略の可否を入力し、データベースを作成した。

本動詞が自動詞・他動詞であるかの判断は、ヲ格を取れない動詞は自動詞（「自」）、取れるものは他動詞（「他」）、という基準に従っているが、「行く、歩く」などの移動動詞はヲ格を取りうるが自動詞に含めている。なお受身形や使役形は「受」、「使」とした。

「ーテクル」の省略の可否は、「ーテクル」を省略すると表現として不自然なもの、文意が大きく変わるもの、文法的に不適格になるものを省略ができないもの（×印）とし、「ーテクル」を省略しても（もちろん意味に多少の変化はあるが）表現として許容できるものを省略ができるもの（○印）とした。例えば、(1a)では「ーテクル」を省略すると、不自然な表現になるため「×」であるが、(1b)では「ーテクル」を省略しても「誰が誰に口答えするのか」という点が自然に解釈できるため「○」である。

- (1) a 目をあけたままの“死体の役”が{まわってきました／??まわりました}。
b 息子は最近、家で少しも言うことを聞かず、口答えばかり {してきます／します}。

データベースの一部を以下に載せる。

表 1 作成したデータベースの一部

番号	前文脈	動詞	自他	テク ル	後文脈	省 略
0002	宮崎県から完熟マンゴーが送っ	送る	他	てき まし	た、食べ方がわかります、美味しい食べ方を教えてください。	×
0048	彼氏がサークルの友達と女の子も3人含んで計7人で夏にキャンプに行きたいと言っ	言う	他	てき まし	た。	○
0092	相手の子は大きさに教師に訴え	訴える	他	てき まし	た。その子の親と話し合い、向こうに非があったという事で一応落ち着きました、教師の対応が許せません。	○
0162	昔の彼女、初めてのキスで口を開け	開ける	他	てき まし	た。	○
0202	さきほど友達が金貸して欲しいという内容のメールを送っ	送る	他	てき まし	た。理由は携帯が止まっちゃいそうだからだそうです。しかし、私は今、金を貸したくありません。	×
0206	兄が結婚相手連れできました。会って話をしたのですが、兄から話した内容について後日電話がかかっ	かかる	自	てき まし	た。相手は高卒です。	×
0219	昨日 彼氏がどんな夢を見たのかは分かりませんが、夜中いきなり『他に男おるやろー！！』って怒鳴っ	怒鳴る	自	てき まし	た。	○
0228	女友達が彼氏に「付き合うのを考えさせて欲しい」と言われ、泣きながら電話をかけ	かける	他	てき まし	た。	×
0777	ある受験関係の本に一流大学ほど入試傾向を変え	変える	他	てこ ない	と書いてあったんですがこれが本当なのですか？今まで一流大学から入試傾向をよく変えてくると思ってたんですが。。。。	○
0844	出品者でメールアドレスに自分の住所、本名、連絡先を書い	書く	他	てこ ない	人は評価がいい人でも不安ではないですか。	○
0972	それから、何年か後、 空には、数え切れないぐらいの宇宙船が宇宙の彼方から飛ん	飛ぶ	自	てき まし	た。	×

3. 考察

3.1 重点

書き言葉コーパスから集めた用例を見ると、(2)のように「ーテクル」を省略しにくい例があることがわかる。

(2) a さきほど友達が金貸して欲しいという内容のメールを {送ってきました／??送りました}。 (表1 202)

b この六名に「何か得意なことはあるか」という質問をすると、そのうちの一人から「パチンコなら誰にも負けない」という返事が {返ってきました／??返りました}。

c 兄が結婚相手を連れてきました。会って話をしたのですが、兄から話した内容について後日電話が {かかってきました／??かかりました}。

(表1 206)

d 女友達が彼氏に「付き合うのを考えさせて欲しい」と言われ、泣きながら電話を {かけてきました／??かけました}。 (表1 228)

(2)の「ーテクル」を省略すると、誰(どこ)に「送る、返る、かかる、かける」のかが曖昧になってしまうため「ーテクル」で動作の方向を示す必要がある。このことから、これらの例では「ーテクル」で表される「方向」に重点が置かれており、そのために「ーテクル」を省略することができないと考えられる。

一方、(3)では「ーテクル」を省略しても不自然にはならない

(3) a 彼氏がサークルの友達と女の子も3人含んで計7人で夏にキャンプに行きたいと {言ってきました／言いました}。 (表1 0048)

b あなたの近況報告をして、ご無沙汰してました、みたいな事を(引用者注：年賀状に)付け加えるだけでいいのではないのでしょうか？みんな、子どもが生まれると、子どもの事しか {書いてきませんが／書きませんが}、子ども以外の自分の最近の話ってないんですか？

c ある受験関係の本に一流大学ほど入試傾向を {変えてこない／変えない}と書かれていたんですが(引用者注：以下略)。 (表1 0777)

省略しても不自然にならないのは、(2)のように「ーテクル」で表される方向にはあまり重点がおかれていないためであると考えられる。言い換えると、本動詞で表される動作に重点が置かれている。

以上から本動詞か「ーテクル」のどちらに重点を置くかが「ーテクル」の省略の可否に関わっている可能性がある。つまり、本動詞に重点を置く場合は「ーテクル」が省略でき、「ーテクル」に重点を置く場合は省略できないと考えられる。

3.2 動作の方向の言語化

- (4) a 彼はぼくに、余計な詮索をしないよう、なぞを {かけてきました／かけました}。
- b あと、Bちゃんは私に {話しかけてきます／話しかけます}。
- c 子供と同じクラスの女の子（2年です）のお母さんが主人の元彼女なんです。（中略）子供同士も仲良しなんでまあまあ、仲良くしてるんですが、何かにつけて私と {張り合ってきます／張り合います}。
- (5) a 彼は余計な詮索をしないよう、なぞを {かけてきました／??かけました}。
- b あと、Bちゃんは {話しかけてきます／??話しかけます}。
- c （中略）子供同士も仲良しなんでまあまあ、仲良くしてるんですが、何かにつけて {張りあってきます／??張り合います}。

(4)では動作の方向（下線部）が言語化されており、また「ーテクル」を省略できる。しかし動作の方向を省略すると、「ーテクル」を省略できなくなる（(5)を参照）。つまり、動作の方向が現れているかどうか「ーテクル」の省略に関わっていると思われる。

(4)は、動作の方向を省略すると「ーテクル」が省略できなくなるという例であった。しかし(6)では動作の方向を省略できない（(6)、(7)参照）。

- (6) a お登勢は、お春の言葉を聞いて、十四郎に視線を {送ってきて／送って}、くすりと笑った。
- b 1927年、日本の子どもたちのもとへ、アメリカから1万2000体の人形が {送られてきました／送られました}。
- (7) a ??お登勢は、お春の言葉を聞いて、視線を {送ってきて／送って} くすりと笑った。
- b ??1927年、アメリカから1万2000体の人形が {送られてきました／送られました}。

(6)の動作の方向を省略すると、動作の方向が話者であるかのようにになってしまう。(4)との違いは動作の方向が話者かそうでないかという点であり、このことから、動作の方向が言語化されており、また方向が話者である場合は省略できると言える。言い換えると、動作の方向が話者である場合、動作の方向が言語化されない限り「ーテクル」によって動作の方向を明示する必要があるということになる。

3.3 本動詞から見て格関係が不適切なもの

- (8) a 宮崎県から完熟マンゴーが {送ってきました／*送りました}、食べ方が

わかりません、美味しい食べ方を教えてください。 (表1 0002)

- b この時期いろんな会社から封筒が来て株主総会の結果報告と会社の説明のパンフレットが {送ってきます／*送ります} よね。

(8)では、「一テクル」を省略すると文法的に不適格な文になってしまう。これは、本来ヲ格をとるべきである「完熟マンゴー、パンフレット」がガ格になっているためである。「一テクル」があっても文法的に適格なわけではないが、(8)でガ格になっている要因として考えられるのは以下の2点である。

まず、受身の表現と混同したという点である。「送る」対象をガ格にして自然なのは「完熟マンゴーが送られてきました」という受身の表現である。「一テクル」も受身形も事態を捉える視点を表すという点では共通しているために混同して、受身形のかわりに「一テクル」を使用したのかもしれない。

(9)も受身との混同だと思われる例である。

- (9) a その後タクシーにりましたが、タクシーの運転手に、「日本の女性は平気で人前でタバコ吸うね」って言ってきました。
b ビジュアル系の女友達にあたしのことをメンだメンと言ってきました。

ここでの「言う」の動作主は「タクシーの運転手、女友達」でガ格であるべきだが、ここではニ格になっている。受身の表現では動作主がニ格になるため、「一テクル」と受身を混同していると考えられる。

次に考えられるのは、「一テクル」の「クル」で表される移動の主体をガ格にしたという点である。「完熟マンゴーが送ってきました」の本動詞を抜くと、「完熟マンゴーがきました」となり、文法的には適格である。そのため、「きました」の主体として「マンゴー」をガ格にしたのかもしれない。

(10)も「一テクル」の「クル」の格関係が表れていると思われるものである。

- (10) a 回の土砂災害防止法案は、土砂災害が上部標高で起きて、その土砂が下部標高に {襲ってきて／*襲って}、そして被害を受ける方々、被害を受ける住宅地における対策ということでございまして、(略)。

本来「襲う」対象はヲ格であるが(10)ではニ格である。「クル」で表される移動の到達点をニ格で表しているものだと思われる。

ここで取り上げた例は、本動詞から見ると格関係が不適切である。受身との混同が見られるということは、「一テクル」と受身の表現で似ている点があると考えられる。また、「クル」の格関係が表れているのは、「クル」による移動に重点が置かれているためであると考えられる。

4. まとめ

- (11) a 「ーテクル」であらわされる移動の意味に重点が置かれる場合、「ーテクル」が使用される。また、本動詞に重点が置かれる場合、「ーテクル」は省略が可能になる。
- b 動作の方向が話者で、動作の方向が言語化されない場合は「ーテクル」を使用して方向を明示する。
- c 本動詞から見ると格関係が不適切な例が見られたが、これらは受身との混同や、「クル」の格関係が表れているためであると思われる。

以上が今回の考察でわかったことである。用例を見ると「ーテクル」の使用には他にも様々な要因が複雑に絡み合っていることが予想される。それを明らかにすることを今後の課題としたい。

参考文献

- 坂原茂（1995）「複合動詞『V て来る』」『言語・情報・テキスト』2 pp. 109-143 東京大学大学院総合文化研究科言語情報科学専攻
- 澤田淳（2008）『『動作の方向性』を表す『テクル』の分析―受動文との比較を含めて―』（日本語学会 2008 年度春季大会研究発表会発表要旨）『日本語の研究』4-4 pp. 149-150
- 寺村秀夫（1984）『日本語のシンタクスと意味Ⅱ』くろしお出版
- 山本裕子（2007）「<主観性>の指標としての『～テイク』『～テクル』」『人文学部研究論』17 pp. 67-81 中部大学
- 吉川武時（1976）「現代日本語動詞のアスペクトの研究」金田一春彦（編）『日本語動詞のアスペクト』 むぎ書房 pp.157-327

BCCWJを用いた「語の文体」の位置づけ

井上次夫（小山工業高等専門学校一般科）[†]

The Classification of "*Linguistic Register*" using BCCWJ

Tsugio INOUE (Dept.General Education,Oyama National College of Technology)

1. はじめに

高専生や留学生が書いた小論文・論文・レポート類に目を通すと、話しことばと書きことばの混用が見受けられることがある。このとき、語の文体が明確に位置づけされていれば、学生にとっても指導する側にとっても有益である。しかし、「俗語」「文章語」等の位相表示を行っている辞典類の記述だけにその解決策を求めるのでは質量ともに限界があり、けっして実用的であるとは言えない状況である。

そこで、本稿では宮島(1972)が示した語の文体の3分類、その後の宮島(1977)、田中(1999)等が示した5分類、さらに井上(2009a・2009b)による2分類及び5分類を参考に、『現代日本語書き言葉均衡コーパス（モニター公開データ 2009年度版）』（*Balanced Corpus of Contemporary Written Japanese*, 以下、BCCWJ2009と略す）を用い、実証的に語の文体の位置づけを図る方法について提案を行う。

2. 語の文体の分類

2-1 宮島（1972・1977）、田中（1999）

宮島（1972：709）は、単語の文体について「単語の文体というのは、文章全体としての文体をなりたせるような、個々の単語のもっている特徴のことである」としたうえで、単語の文体を大きく、俗語・日常語・文章語の3種類に分けた。そして、それらを以下のように定義した。

(1)「俗語」とは、もっぱらくだけた話しことばでだけもちいられ、書きことばの中や、あらたまった場面における話しことばには、あまりあらわれないもの

「日常語」とは、話しことば・書きことばを通じて、あらゆる言語生活の場面にもちいられ、特に文体上の制約がないもの

「文章語」とは主として書きことば、またはあらたまった場面における話しことばで使われるもの

また、宮島（1977：873）では、「文体的特徴からする単語の分類は、連続的であり、程度の差による」ことから、先の宮島（1972）の3分類の中にも程度差が認められるとして、日常語をさらに「くだけた日常語・無色透明な日常語・あらたまった日常語」に下位分類した。その結果、単語の文体として次の5種類を認めることとなった。

[†] inoue@oyama-ct.ac.jp

(2) 俗語・くだけた日常語・無色透明な日常語・あらたまった日常語・文章語

その後、田中（1999：229）では、さまざまな接続表現を5段階に分類し、会話的なものほど「うちとけた文体」、文語的なものほど「かたい文体」とする一覧表を示した。

(3) うちとけた←…会話的・話しことば的・一般・書きことば的・文語的…→かたい

例) けど・だけど・けれども・しかし・しかしながら

以上のように、語の文体は3分類から5分類へと進んではいるが、そもそも語の文体が連続的・相対的なものであることから、それぞれの語の文体の境界部分については依然として不透明な部分が残る、個人差も残る。

2-2 井上（2009a・2009b）

そこで、井上（2009b）は、論説文における語の文体としての適否という視点から、語の文体を大きく明確に2分類する提案を行った。これは、宮島（1977）の5分類を教育上、実用上の必要性から次のように分類しようとするものである。

(4) 俗語<くだけた日常語<無色透明な日常語<あらたまった日常語<文章語

（不等号「<」は書きことばらしさの強弱、「<<」は論説文に用いる語の文体としての適否の境界を表す）

そこでは、「くだけた日常語」と「無色透明な日常語」の境界部分を判定するに際し、BCCWJにおける語の出現数を用いている。例えば、「だいたい（大体）」と「おおむね（概ね）」の場合を例にみると、BCCWJ2008の「白書」「Yahoo!知恵袋（以下、知恵袋）」コーパスにおける出現数を求め、それを適否判断の材料とする。すなわち、論説文として適切な文体の側に「白書」を、論説文としては不適切な文体の側に「Yahoo!知恵袋」を位置づけ、それぞれにおける出現数の多寡により適否の判断を行う。

(5) だいたい（白書 8, 知恵袋 684）<<おおむね（白書 494, 知恵袋 23）

ただし、「だいたい」の「白書」での出現数 8、「おおむね」の「知恵袋」での出現数 23については、語の文体が「それぞれ、いちいちの単語のもっている特徴であって、それがどのような言語活動に現にあらわれているかということとは関係ない（宮島 1972：709）」ことから、語の文体の判断に大きな影響を与えるものではない点には注意されたい。

いっぽう、井上（2009a）では BCCWJ2008 を用いた、語の文体に関する5分類を提案した（表1参照）。そこでは、それが語の文体であることが明確になり、かつ語の文体とその内容とが結びつきやすくなることを目的として次のように命名した。

(6) 卑俗体・口頭体・汎用体・書記体・文語体^{注1}

このうち、「卑俗体」は通常、公式の場で使われない卑語・俗語・隠語・流行語等を含む。いっぽう、「文語体」には現代の日常生活では通常、触れることが少ない漢語・字音語・漢文訓読語的な語彙・雅語等を含む。これらは文体としての区別が比較的容易であると思われるが、区別に揺れが生じやすいのは「口頭体」と「汎用体」、「汎用体」と「書記体」の境界部分であり、言い換えれば「汎用体」の範囲であるとも言える。

^{注1} 井上（2009a）では「文章体」としたが、本稿では以下、「文語体」に改めた。

表1 「語の文体」5分類（井上 2009a）

印象	くだけた ←	うちとけた ←	普 通 →	あらたまった →	かたい
文体(井上)	卑俗体	口頭体	汎用体	書記体	文語体
文体(宮島)	俗語	くだけた日常語	無色透明な日常語	あらたまった日常語	文章語
文体(田中)	会話的	話しことば的	一般	書きことば的	文語的
ことば	話し言葉	話し言葉	話し言葉・書き言葉	書き言葉	書き言葉
主な場面	私的会話	日常会話	公私等の別なし	公的発言・論説文	論文・詩歌
語例	あつし	あたし	わたし	わたくし	小生
	けど	だけど	けれども	しかし	しかしながら
	φ	なかでも	とくに	とりわけ	なかんずく
	だって	φ	φ	というのは	なんとになれば

そうして、井上（2009a）ではそれらの境界、範囲を明確化するために BCCWJ2008 による語の出現数の調査を行い、(7)の結論を得た。

(7) 得られた出現数について、以下の特徴をもとに語の文体を検討する。

- ① 卑俗体は「白書」コーパスにおける出現数が 0 件である^{注2}。
- ② 文語体は「知恵袋」コーパスにおける出現数が 0 件である。4 コーパスにおける総出現数が相対的に少ない。
- ③ 汎用体は「知恵袋」「書籍」「国会」「白書」の 4 コーパスに出現し、総出現数が相対的に多い。特に「書籍」コーパスの出現数が多い。
- ④ 口頭体は「白書」コーパスにおける出現数が 0 件、または比較的少ない。
- ⑤ 書記体は「知恵袋」コーパスにおける出現数が 0 件、または比較的少ない。

上記(7)では、コーパスにおける出現数 0 件を一つの重要な特徴としている。その結果、「白書」コーパスに出現数 0 件の語の場合は卑俗体か口頭体、「知恵袋」コーパスに出現数 0 件の語の場合は文語体か書記体、そして「知恵袋」「書籍」「国会」「白書」コーパスいずれにも出現する場合は汎用体になる可能性が高いことを意味するものとなっている。しかし、出現数が 0 件でない語の場合については、その出現数が相対的に多いのか少ないのか、比較的多いのか少ないのか、依然として不明確であると言わざるを得ない。

2-3 BCCWJ2009 と語の文体

さて、BCCWJ2008 における「知恵袋」「書籍」「国会」「白書」コーパスと語の文体との関係をこれまで表 2 のように考えてきた。しかし、「国会」は一次的には話しことばであるが、基本的にあらたまった日常語でもあるため、いま、その取り扱いについては保留する。また、「書籍」については地の文と会話文が混在しているため、それを区別しながら調査することが実際には容易でない。さらに、BCCWJ2009 では「書籍」について新たに LB：流通

^{注2} 「白書」の出現数 0 件のものが卑俗体であるというのではない（口頭体の可能性も含んでいる）。②～⑤でも同様に、出現数は各文体の特徴について述べるものである。

実態サブコーパス（の書籍）、OB：特定目的サブコーパスの中のベストセラー、PB：生産実態サブコーパス（の書籍）と下位分類されたこともあり、それぞれにおいて地の文と会話文を区別しながらの調査はいつそう難しいため、いま、その取り扱いには保留する。

以上のことから、本稿では残された「知恵袋」と「白書」を主に取り上げ、前者を卑俗・口頭体、後者を書記・文語体と対応させ、汎用体を含めた 3 分類における語の文体の位置づけを図る方法について次章で検討する。

表 2 コーパスと語の文体

コーパス	語の文体		位相	総数
Yahoo!知恵袋	卑俗体・口頭体	くだけた日常語	話し言葉	約 500 万語
国会議事録	書記体	あらたまった日常語	話し言葉・書き言葉	約 500 万語
書籍	汎用体	無色透明な日常語	話し言葉・書き言葉	約 700 万語
白書	書記体・文語体	文章語	書き言葉	約 500 万語

3. 語の文体の位置づけ

ここでは、BCCWJ2009 を用い、語の文体を A 層：卑俗・口頭体、B 層：汎用体、C 層：書記・文語体のいずれかに位置づけようとする。先行研究の挙例を中心に、意味・用法が似ていると思われる 52 組 230 語を無作為に取り上げ、以下の手順で調査を行う。

ア BCCWJ2009 の「知恵袋」「白書」コーパスにおける各組の語の出現数を求める。

イ 「知恵袋」対「白書」の出現比率を求め、各組の語を次のように 3 層に分ける。

	知恵袋	白書
A 層（卑俗・口頭体）：	80%以上～100%	0 %以上～20%未満
B 層（汎用体）：	20%以上～80%未満	20%以上～80%未満
C 層（書記・文語体）：	0 %以上～20%未満	80%以上～100%

ウ 「知恵袋」対「白書」の出現数に基づき、各組の語を散布図により可視化する。

エ 上記ウの語の傾き ($\tan \theta$)、線分の長さ (r) を求め、数値により位置付ける。

オ 上記エに基づき、各組の語の文体を A 層～C 層のいずれかに位置づける。

3-1 出現比率による位置づけ

調査の対象とした 52 組 230 語を上記イの比率基準により 3 層に分類すると、表 3 のようである。E 層は出現数 0 の語で 15 語あった。E 層を除く

表 3 調査語の内訳

3 層の計 215 語における各層の割合は表 3 の括弧内のようにであった。ただし、これは今回の調査語の構成に過ぎず、一般に各層の語の文体がここから推測できるというものではない。しかし、各層にどのような語が属しているかを具体的に見ておくことは有益である。

そこで、A 層～C 層の 3 層をいずれも含んでいる組

層	文体	該当語
A	卑俗・口頭	138 (64.2%)
B	汎用	42 (19.5%)
C	書記・文語	35 (16.3%)
E	保留	15

のうち主な語を以下に例示する。①～⑩の語は各層の間で対応関係にある。

(8) A層：卑俗・口頭体（「知恵袋」出現率 80%以上）

①しょっちゅう、②だぶる、③なるだけ、④このごろ・ちかごろ、⑤だいたい・たいがい、⑥ただ、⑦じわじわ・だんだん・すこしずつ、⑧でも・けど・けれども・ところが、⑨わりかし・わりと、⑩すごく・とても・とっても

B層：汎用体（「知恵袋」「白書」出現率 20%以上 80%未満）

①つねに、②かさなる、③できるだけ、④最近、⑤ほぼ・ほとんど・おおよそ・およそ、⑥ただし・もともと、⑦徐々に、⑧しかし、⑨いくぶん、⑩非常に

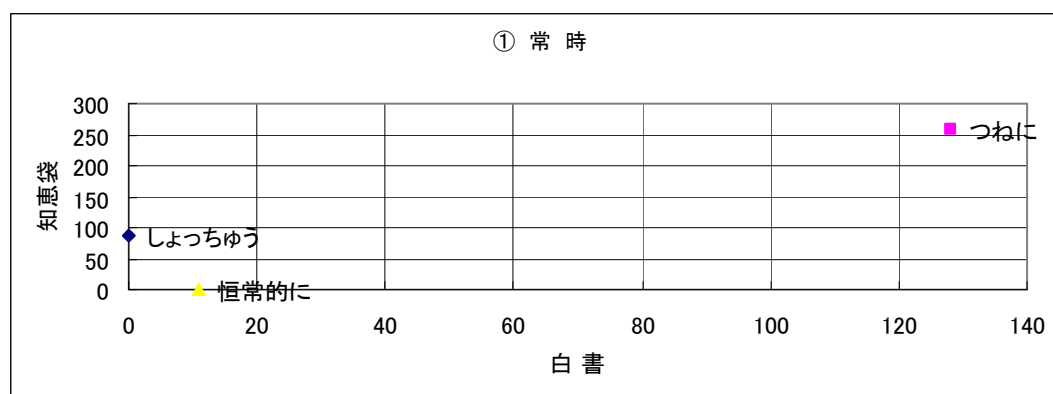
C層：書記・文語体（「白書」出現率 80%以上）

①恒常的に、②重複する、③できるかぎり・可能なかぎり、④近年、⑤おおむね・約 1（便宜上「約 1・約一」で調査）、⑥なお、⑦次第に、⑧しかしながら・しかるに、⑨やや・比較的、⑩きわめて

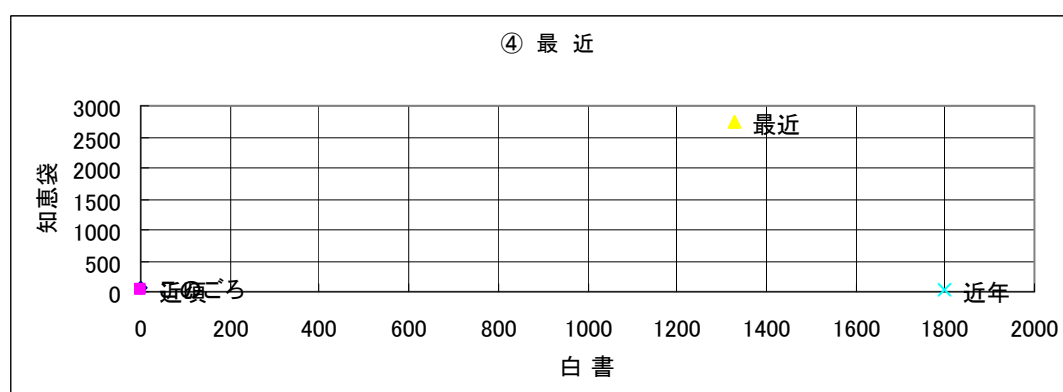
3-2 散布図による位置づけ

ここでは、調査語の「知恵袋」対「白書」コーパスにおける出現数に基づき、各組の語の散布図を作成し可視化を試みる。以下、前項の①～⑩のうち①常時（「しょっちゅう」類）、④近時（「このごろ」類）、⑥注意・補足（「ただ」類）、⑦漸進（「じわじわ」類）を例として取り上げ、縦軸を「知恵袋」、横軸を「白書」とする散布図を示す（3. ウ）。

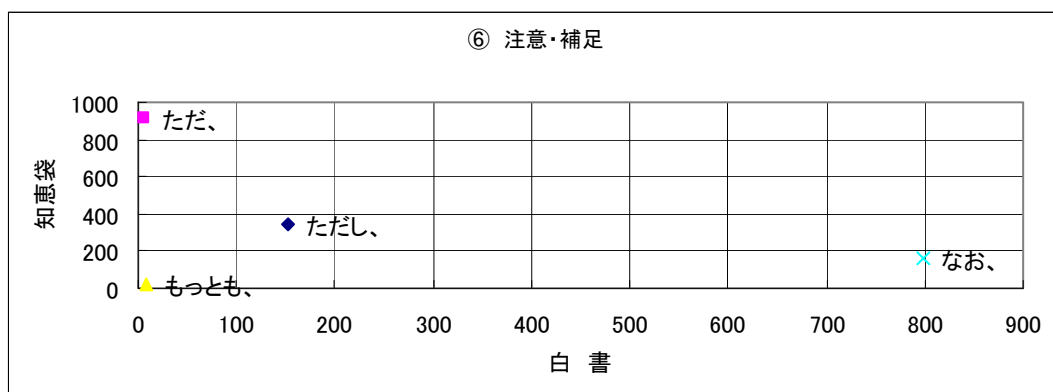
① 常時表現（しょっちゅう・つねに・恒常的に）



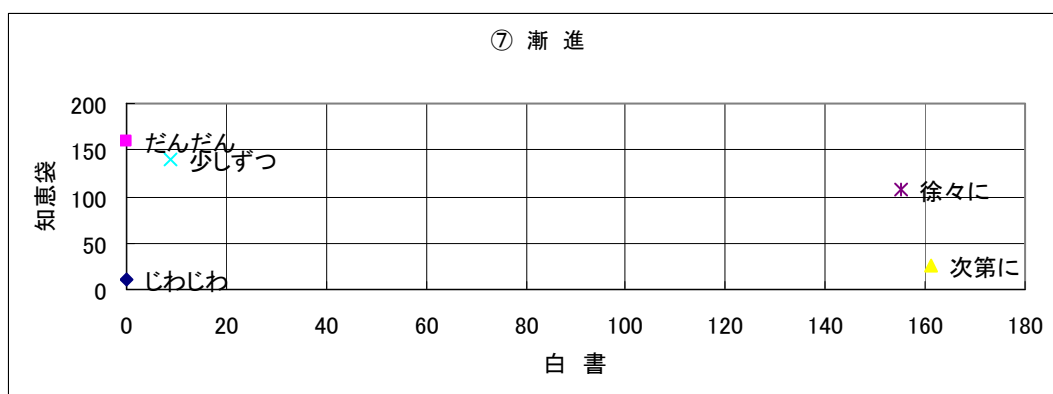
④ 近時表現（このごろ・ちかごろ・最近・近年）



⑥ 注意・補足表現（ただ・ただし・もっとも・なお）



⑦ 漸進表現（じわじわ・だんだん・すこしずつ・徐々に・次第に）



上掲の散布図①～⑦の 4 枚を見ると、各組の語の相対的位置が把握できる。縦軸に近いほど A 層：卑俗・口頭体、横軸に近いほど C 層：書記・文語体である。

(9) 各語の文体の位置関係

- ① 常時表現 しょっちゅうくつねにく恒常的に
- ④ 近時表現 このごろ・ちかごろく最近く近年
- ⑥ 注意・補足表現 ただ・もっともくただしくなお
- ⑦ 漸進表現 じわじわ・だんだんくすこしずつく徐々にく次第に

ただし、例えば、上記④近時表現の「このごろ」と「ちかごろ」の文体差の有無・程度、⑦漸進表現の中間に位置する「すこしずつ」は A 層：卑俗・口頭体のように見えるが B 層：汎用体ではないか。また「徐々に」は B 層のようだが C 層：書記・文語体ではないか、などといった点は散布図だけでは不明瞭であると言えよう。そこで、その課題に答えるため各語の数値による文体の位置づけを次項で図ることとする（3. エ）。

3-3 極座標による位置づけ

点 P が x-y 平面上にある場合、点 P の位置は座標 (x, y) で表せるいっぽう、極座標を用いて P (r, θ) と表すこともできる。r は基準点からの距離、θ は x 軸となす角である。ここでは、この極座標の考え方をを用いて、各語の位置を次のように表すこととする。

(10) 語W (r , θ) の位置

① r : 便宜的に、 $r/10$ で表す

② θ : 便宜的に、 $\tan \theta$ (傾き) で表す

すると、上記①により語の使用規模の大小、②により文体の傾きの程度を数値化することができる。なお、語Wの文体を考える際には、①の r (使用規模) よりも②の θ (その語の文体らしさ) のほうがより重要であると考えられるため、以下、本稿では先の(10)を改め、②を先に表示する方式の(11)を用いることとする。表4は常時表現を例に語の位置を示す。

(11) 語Wの位置 $W(\tan \theta, r/10)$

表4 常時表現

	語	白書	知恵袋	$\tan \theta$	r
1	しょっちゅう	0	87	∞	8.7
2	つねに	128	257	2	28.7
3	恒常的に	11	1	0.1	1.1

この表では、「しょっちゅう (∞ , 8.7)、つねに (2.0, 28.7)、恒常的に (0.1, 1.1)」であることから、先の(9)と同様に次の位置関係が成り立つことが確認される。以下、不等号 (書きことばらしさの強弱、文体の程度) は、 $\tan \theta$ の値を基準にしている^{注3}。

(12) しょっちゅう (∞ , 8.7) < つねに (2.0, 28.7) < 恒常的に (0.1, 1.1)

それでは、ここで先に前項3-2で掲げた2つの課題について検討しよう。

まず、④近時表現「このごろ」と「ちかごろ」の文体差の有無・程度についてであるが、各語の位置を調べてみると「ちかごろ (∞ , 2.8) < このごろ (25.0, 5.0)」である。このことから、ともに文体はA層：卑俗・口頭体に属することが確認される。そして、敢えて言えば「ちかごろ」よりも「このごろ」のほうが文体はかたく、使用規模も大きいということになる。

次に、⑦の漸進表現のうち「すこしずつ」はA層：卑俗・口頭体のようだがB層：汎用体ではないか。また「徐々に」はB層のようだがC層：書記・文語体ではないか、という点について検討する。

(13) じわじわ (∞ , 1.1)・だんだん (∞ , 15.9) < すこしずつ (15.4, 13.9)

< 徐々に (0.7, 18.9) < 次第に (0.2, 16.3)

(13)によると、「すこしずつ」と「徐々に」の $\tan \theta$ は、それぞれ15.4と0.7で大きいと思われるが、この差は語の文体差にどの程度影響すると考えればよいのだろうか。つまり、 $\tan \theta$ の範囲をA層：卑俗・口頭体、B層：汎用体、C層：書記・文語体という語の文体の分類にどのように対応させればよいか、ということが次の課題となる。

そこで、あらためて θ の範囲を確認すると、本稿の場合、次のとおりである。

(14) $0 \leq \theta \ll \pi/2$ ($=90^\circ$)

^{注3} $\tan \theta$ の値が小さいほど、書きことばらしさは強くなり、書記・文語体に近くなる。

いま仮に、 θ の範囲を出現比率の場合と同様の割合により 3 分割してみると、次のような対応関係が推定できる。

表 5 $\tan \theta$ に基づく語の文体①

θ	$\tan \theta$ (傾き)	語の文体
$72^\circ \leq \theta < 90^\circ$	$3.1 \leq \tan \theta < \infty$	A層：卑俗・口頭体
$18^\circ \leq \theta < 72^\circ$	$0.3 \leq \tan \theta < 3.1$	B層：汎用体
$0^\circ \leq \theta < 18^\circ$	$0.0 \leq \tan \theta < 0.3$	C層：書記・文語体

さて、この対応関係はどの程度まで妥当で有効なものであるのか、(8)の掲載語を対象サンプルとして調査を行い、(8)の分類と比較してみる。次の(15)では、(8)の掲載語の位置について、層の区分は変更せず、層内に複数ある語については、まず $\tan \theta$ 、次に $r/10$ の降順に不等号を用いて並べ換えた。下線は、表 5 の $\tan \theta$ の範囲に含まれない語。

(15) A層：卑俗・口頭体 ($3.1 \leq \tan \theta < \infty$)

①しょっちゅう ($\infty, 8.7$)、②だぶる ($\infty, 0.1$)、③なるだけ ($\infty, 0.4$)、④ちかごろ ($\infty, 2.8$) < このごろ (25.0, 5.0) ・、⑤たいがい ($\infty, 4.2$) < だいたい (85.5, 68.4)、⑥ただ (131.0, 37.6)、⑦だんだん ($\infty, 15.9$) < じわじわ ($\infty, 1.1$) < すこしずつ (15.4, 13.9)、⑧けど ($\infty, 6.1$) < でも (2878, 287.8) < ところが (4.7, 11.6) < けれども (4.5, 0.9)、⑨わりかし ($\infty, 0.1$) ・ わりと (26.5, 5.3)、⑩すごく ($\infty, 101.7$) < とっても ($\infty, 19.5$) < とても (145.8, 145.8)

B層：汎用体 ($0.3 \leq \tan \theta < 3.1$)

①つねに (2.0, 28.7)、②かさなる (0.5, 2.1)、③できるだけ (1.4, 19)、④最近 (2.1, 305.3)、⑤ほとんど (2.3, 119.8) < おおよそ (1.8, 2.3) < およそ (0.8, 4.2) < ほぼ (0.4, 92.6) ・ ⑥ただし (2.3, 37.6) < もっとも (2.2, 2.2)、⑦徐々に (0.7, 18.9)、⑧しかし (1.2, 135.4)、⑨いくぶん (0.6, 0.8)、⑩非常に (3.5, 58.3)

C層：書記・文語体 ($0.0 \leq \tan \theta < 0.3$)

①恒常的に (0.1, 1.1)、②重複する (0.1, 1.2)、③できるかぎり (0.2, 7.0) < 可能なかぎり (0.0, 8.5)、④近年 (0, 179.8)、⑤約 1 (0.2, 102.0) < おおむね (0.0, 49.3)、⑥なお (0.2, 81.4)、⑦次第に (0.2, 16.3)、⑧しかしながら (0.1, 27.6) < しかるに (0.0, 0.5)、⑨比較的 (0.2, 64.5) < やや (0.2, 45.2)、⑩きわめて (0.1, 66.6)

この結果から、まず、先の課題の⑦「すこしずつ」「徐々に」の文体については「すこしずつ」がA層、「徐々に」がB層であることが確認できる。

次に、下線の語「非常に (3.5, 58.3)」の $\tan \theta$ 値はB層の範囲を超えているためこの語はA層に属することになる。しかし、超えた値が 0.4 と小さいため、層の境界となる $\tan \theta$ 値の設定に課題がないわけではない。そこで、⑧の「けれども (4.5, 0.9)」をA層であると認めるならば、暫定的にA層とB層の $\tan \theta$ の境界値を(16)のように変更 (3.1→4.0) しておくことも可能である。しかしながら、今後、さらに多くの語について調査を行いつつ、層の境界となる $\tan \theta$ 値の最適値を求めていく必要があることは言うまでもない。

(16) $\tan \theta$ に基づく語の文体②

$4.0 \leq \tan \theta \rightarrow$ A層：卑俗・口頭体

$0.3 \leq \tan \theta < 4.0 \rightarrow$ B層：汎用体

$0.0 \leq \tan \theta < 0.3 \rightarrow$ C層：書記・文語体

4. まとめ

以上の語の文体の位置づけを「すごく・とっても・とても・非常に・きわめて」の5語を例にまとめておく。

表6 「すごく」類の文体

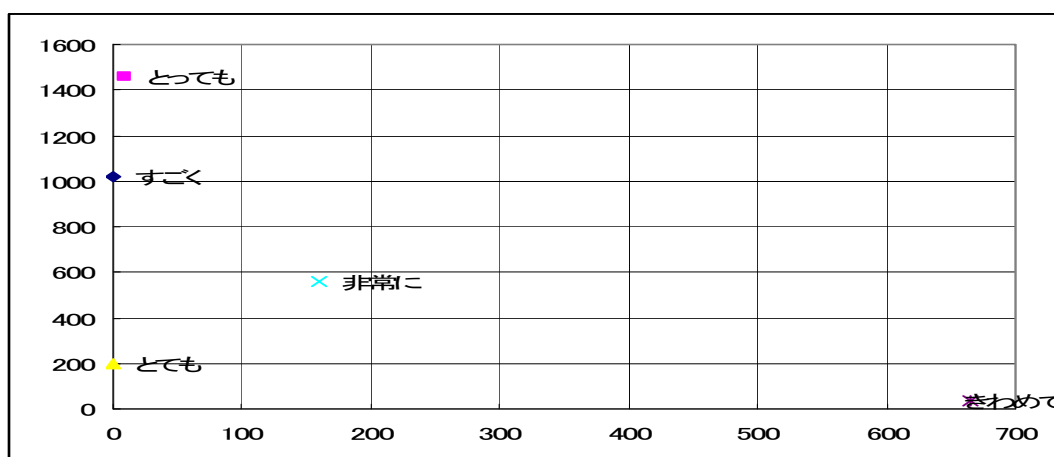
	①白書 a	知恵袋 a	②白書 b	知恵袋 b	③ $\tan \theta$	r	④文体
すごく	0	1017	0	100	∞	101.7	A層
とっても	0	195	0	100	∞	19.5	A層
とても	10	1458	0	100	145.8	145.8	A層
非常に	160	561	22.2	77.8	3.5	58.3	B層
きわめて	665	37	94.7	5.3	0.1	66.6	C層

表6を見ると、①白書 a・知恵袋 aは各語の出現数を表す。「すごく」「とっても」はともに白書0件につき井上(2009a)によれば、A層：卑俗・口頭体に属する。

②白書 b・知恵袋 bは両者の出現比率を表す。「すごく」「とっても」「とても」の3語は白書20%未満（知恵袋80%以上）につきA層：卑俗・口頭体、「非常に」は白書・知恵袋20%以上～80%未満につきB層：汎用体、「きわめて」は白書80%以上（知恵袋20%未満）つきC層：書記・文語体に属することになる。

③ $\tan \theta$ は文体の傾き（その語の文体らしさ）、r は距離（使用規模）を表す。「すごく」「とっても」「とても」の3語は $4.0 \leq \tan \theta$ につきA層：卑俗・口頭体、「非常に」は $0.3 \leq \tan \theta < 4.0$ につきB層：汎用体、「きわめて」は $0.0 \leq \tan \theta < 0.3$ につきC層：書記・文語体に属することになる。さらに、これら5語を散布図で示すと次のようになる。

(17) 散布図「すごく・とっても・とても・非常に・きわめて」



④以上のことから、各語の文体を次のように位置づけることができる。

A層：卑俗・口頭体…とつても、すごく、とても ($80\% \leq \text{知恵袋}$ 、 $4.0 \leq \tan \theta$)

B層：汎用体…非常に ($20\% \leq \text{白書} \cdot \text{知恵袋} < 80\%$ 、 $0.3 \leq \tan \theta < 4.0$)

C層：書記・文語体…きわめて ($80\% \leq \text{白書}$ 、 $0.0 \leq \tan \theta < 0.3$)

なお、ここではいずれの語も結果的にそれぞれの出現比率と文体の傾きの範囲が一致しているが、そうではない場合、どのように語の文体の判断を行うか、すなわち出現比率と文体の傾きのどちらをどのように重視するか、という問題が生ずるが、この点については今後の課題である。また、A層とC層の下位分類についても今後の課題とする。

5. おわりに

本稿では、BCCWJ2009 に基づき実証的に語の文体の位置づけを図るための分類法を提案した。そして、意味・用法が近似する語の文体の位置づけについて、知恵袋・白書の各コーパスでの出現比率及び散布図・極座標が有力な手だてとなり得ることを示した。なお、このことは任意の1語について、散布図・極座標によりその文体を位置づける扉を開くものでもある。しかし、今後、それらの手だての精度、使用法についてはさらに多くの語について調査を進め信頼性を高めていく必要がある。

文献

- 井上次夫(2009a). 「日本語コーパスに基づく『語の文体』の明確化」 文部科学省科学研究費特定領域研究「日本語コーパス」『平成20年度公開ワークショップサテライトセッション予稿集』 pp.109-118.
- 井上次夫(2009b). 「論説文における語の文体の適切性について」『日本語教育』141, pp.57-63.
- 荻野綱男(2006). 「WWWによる単語の文体差の研究」『日本語学会2006年度秋季大会予稿集』, pp.139-146.
- 菊沢季生(1930). 「国語位相論」『国語科学講座Ⅲ』, 明治書院 pp.2-67.
- 後藤斉(2001). 「日本語コーパス言語学と語の文体レベルに関する予備的考察」『東北大学文学研究科研究年報』 pp.200-214.
- 島本基(1990). 「語の位相」『講座日本語と日本語教育』7, 明治書院 pp.298-322.
- 田中章夫(1978). 『国語位相論』, 明治図書.
- 田中章夫(1999). 『日本語の位相と位相差』, 明治図書.
- 徳川宗賢・宮島達夫(1980). 『類義語辞典』20版, 東京堂出版.
- 前川喜久雄(2008). 「話し言葉と書き言葉」『日本語学』27-5, pp.23-33.
- 宮島達夫(1972). 『動詞の意味・用法の記述的研究』, 秀英出版.
- 宮島達夫(1977). 「単語の文体的特徴」『松村明教授還暦記念国語学と国語史』 pp.871-903.
- 宮島達夫(1988). 「単語の文体と意味」『国語学』154, pp.78-88.
- 宮島達夫(2008). 「文章の文体と単語の文体—国研コーパスを利用して—」『近代語研究』14, pp.375-386.

複合辞トイウにみる文法化の諸相

金城 由美子（長崎純心大学人文学部）[†]

On Grammaticalisation Process and Functions of “*toiu*”

Yumiko Kinjo (Faculty of Humanities, Nagasaki Junshin Catholic University)

1. はじめに

引用の動詞「言う」を含む複合辞トイウは、文末において伝聞を表す助動詞用法と、連体修飾節と共起し、連体助詞として機能する助詞用法を持つ。トイウが複合辞としての統語的・意味的機能を持つ一方で、「言う」の動詞としての機能は失われておらず、さまざまな用法が共存している。本稿では、トイウの助動詞用法と助詞用法を概観した上で、共起する節の種類とトイウの機能の関係について論じることを目的とする。助詞用法のトイウについては、トイウが共起する節の種別により非限定修飾など異なる機能を果たすことを示す。本稿での分析対象として、『現代日本語書き言葉均衡コーパス』モニター公開データ（2009年度版）（以下、BCCWJ2009 モニター版）を利用する。

2. 複合辞トイウと文法化

トイウをはじめとして、引用の動詞「言う」を用いた複合辞が多数存在し、さまざまな統語的・意味的機能を果たしている。砂川（2006）では、動詞「言う」を用いた複合辞を実質語から機能語への文法化として捉え、その文法化の事例には中核的なものから周辺的なものまでさまざまなレベルが認められ、それらが同時に並列的に存在し、重層的な様相を呈している点が指摘されている。また機能語へと変化する実質語は、もともと実質的な意味が希薄な語が多いが、「言う」は「する」や「なる」に比べて実質的な意味（「言葉で表す」「言葉を発する」）を持ちながら、助詞や助動詞に相当する多種多様な複合辞を作り上げているという、特殊性も指摘されている。

本稿では動詞「言う」を含む複合辞トイウの重層的な性質に注目し、節と共起する複合辞トイウの統語的・意味的機能を「実質語としての性質の喪失」と「機能語としての性質の獲得」という観点から、トイウの助動詞用法と助詞用法について考察を行う。本稿では、トイウの助詞用法のうち「節＋トイウ＋名詞」の形式について論じ、「名詞＋トイウ＋名詞」については扱わない。またトイウが共起する節のモダリティに注目し、その統語的・意味的機能について論じる。

砂川（2006）であげられている実質語から機能語への過程で起こる変化として、次のような変化が挙げられている。

- (1) 実質語の喪失過程（砂川 2006, p.28）
 - i) 実質的意味の希薄化
 - ii) 統語機能の喪失

[†] kinjo@n-junshin.ac.jp

iii) 形態的な固定化

iv) 音声的な摩滅

喪失過程で起こる意味の希薄化などについては、各用法の例とともに述べる。

トイウの助動詞用法、助詞用法ともに、節との共起があるものについて論じるが、その際、共起する節の階層構造からみた種類に注目する。節の種類については、野田 (2002) の分類に準じる。

(2) 主文述語の階層構造からみた節の種類 (野田 2002, p.12)

述語の 語幹	ヴォイス	アスペクト	肯定 否定	テンス	対事的 ムード	対他の ムード
閉め	られ	てい	なかつ	た	ようだ	ね

助動詞トイウは対事的ムード、助詞トイウは対他のムードの節と、異なる階層の節と共起することを示し、両者が異なる統語的機能を持つことを示す。

3. トイウの助動詞用法

文末のトイウは、伝聞のムードを表すとされる。他から聞いて得た知識を推論を加えず、そのまま伝えるムードで、過去の表現は持たない (益岡・田窪 1989)。

(3) 所蔵する作品の9割は、地元の人々からの寄贈だったという。(PB12_00286)

「いう」を「いった」にした場合、発言者が明示されていなくても、過去の発話について述べる文と解釈されてしまう。助動詞用法においては、もともと動詞が持っていたテンスという統語機能を喪失し、結果的に形態的にも固定化したものと捉えることができる。

また伝聞のトイウは「といわれる」ほぼ同義である。受動態を含む「といわれる」と助動詞トイウは、動作主を欠くという統語的性質を共通にしている。

トイウによる伝聞では、情報源を「～によれば／よると」で示すことができる (益岡・田窪 1989)。「言う」の動作主は「ものを言う」のような慣用句を除けば人に限られているが、情報源は人に限らず、書籍なども可能である。

(4) これは光恵さんによると京の文化だという。(LBq1_00027)

(5) 『平家物語』によれば、寿永元年(一一八二)の源平合戦に際し、恵日寺の宗徒は平家に味方して、木曾義仲追討のため信濃へ出兵したという。(PB42_00286)

(5)は「言葉を発する」という実質的な意味が希薄化し、動作主を持つという統語機能が失われたため、情報源として無情物が可能になったとみることができる。また(5)のトイウは、「とある」と置き換えが可能である¹。

(3)～(5)の例に見られるように、助動詞トイウは主題を含む節と共起することができる。また対事的ムードを含む節とも共起可能である。ただし、この場合の対事的ムードは話し手の主観を表すものではない。

(6) 家光には他にも愛した青年大名がいたらしいという。(OB1X_00118)

¹ 書籍からの引用に「とある」が利用される場合、情報源は「に」で示されることが多い。

対他的ムードを含む節と伝聞のトイウは共起しない。対他的ムードが含まれる場合は、節が発話と解釈され、伝聞にはならない。

(7) 家光には他にも愛した青年大名がいたらしいねという。

動詞「言う」の引用構文で引用句トは対他的ムードの節と共起するが、助動詞トイウは対他的ムードの節と共起しない。これは、統語的機能の喪失の一種と考えられる。

トイウの伝聞という助動詞用法は、「言葉を発する」という実質的意味の希薄化と、動作主、対他的ムードとの共起という統語機能の喪失によりもたらされたものであるといえる。

4. トイウの助詞用法

連体修飾構造において、連体修飾節と被修飾名詞の間に、トイウの介在を許すもの、許さないものがあり、連体修飾節の種別とトイウの介在可能性については多くの論考がある。本稿ではこれらの論考を踏まえた上で、助詞用法のトイウの機能を考察する。

寺村 (1975) は、連体修飾節を被修飾名詞が修飾節中の述語に対する補足語の関係にある「内の関係」(補足語修飾節) と、修飾節が被修飾名詞の内容を表す「外の関係」に大別した。「外の関係」の連体修飾節には、相対名詞修飾節と内容節(内容補充修飾節)の二種類がある。内容節は、トイウの介在を必須とするものと任意とするものがあり、相対名詞修飾節はトイウの介在を許さない。本稿では、主としてトイウの介在が起こる連体修飾節について論じる。

4.1 トイウと補足語修飾節

いわゆる「内の関係」にあたる補足語修飾節は、多くの場合トイウが介在しないものとされている。トイウが介在する場合は、伝聞と解釈される(山崎・藤田 2001, 日本語記述研究会 2008)。

(8) サンマを焼くというおやじ (山崎・藤田 2001, p. 119)

(8)は「おやじがサンマを焼くという」という伝聞の意味を持つ文に対応する連体修飾表現であると考えられる。助動詞トイウは、補足語修飾節内の要素であると考えられる。

補足語修飾節内という環境において、動詞「言う」と助動詞トイウの解釈があいまいになる場合がある。

(9) 不眠症だという女を診察した。

(9)は、「不眠症だと主張する女」と「不眠症だといわれる女」の二通りの解釈が可能である。前者は動詞「言う」、後者は助動詞トイウの解釈である。砂川 (2006)において「多動性障害といわれる子供を診察した」に「話し手の判断を保留しているかのような意味合い」があるとの指摘がされているが、どちらの解釈においても、補足語修飾節が話し手以外の判断を示している点は共通であり、第三者の判断を表す伝聞の表現への連続性がみてとれる。

4.2 トイウと内容補充修飾節

連体修飾節が被修飾名詞と補足語の関係を持たず、被修飾名詞の内容を説明するタイプのものを内容補充修飾節と呼ぶ。内容補充修飾節は、被修飾名詞のタイプによりトイウの

介在が必須のものと任意のものがある。まずトイウの介在が必須である引用系名詞が被修飾名詞である場合について、次にトイウの介在が任意であるコト系名詞が被修飾名詞である場合について、それぞれ助詞トイウの機能についてみていく。

4.2.1 引用系名詞

発話に関する名詞「言葉」「発言」「指摘」や、思考に関する名詞「考え」「新年」「意見」などを引用系名詞と呼ぶ（益岡 2002）。引用系名詞を被修飾名詞とする場合、トイウは必須の要素である。

(10) この案は本当に実現可能なのかという質問 （益岡 2002, p. 105）

トイウは対他格的ムードを持つ内容補充節と共起する。この点は動詞「言う」の統語的性質を引き継いでいるといえる。

トイウは「名詞＋トイウ＋名詞」という形式で連体助詞としての機能を果たす。丹羽(1993)はこのトイウの機能を「名前を同定するトイウ」とありとし、「「X トイウ Y」は名前を同定する「Y ヲ X トイウ」が連体関係で用いられたものである。」と説明している。つまり補足語修飾節と同じ形式のままで、実質語としての性質を失い、形態の固定化が起きたものと考えられる。一方、引用系名詞が被修飾名詞となる場合、内容補充節が補足語修飾節と同じ形式を取ることはない。引用構文が補足語修飾節になる場合、被修飾名詞になるのは主格である動作主のみであり、ヲ格要素も被修飾名詞とならない。引用句は副詞節であり、副詞は補足語修飾節の被修飾部となることはない。

(11) この案は本当に可能なのかと質問した議員

(12) a. この案は本当に可能なのかと文句を言った議員

b. *議員がこの案は本当に可能なのかと言った文句

つまり引用系名詞と共起するトイウは、名前を同定するトイウと共通の連体助詞としての統語機能だけでなく、内容補充という関係で節と被修飾名詞を結ぶという異なる統語機能も持っていると考えられる。

4.2.2 コト系名詞

内容補充節による修飾を受ける名詞のうち、引用系名詞をのぞいた「事実」「騒ぎ」「結果」などの名詞をコト系名詞と呼ぶ（益岡 2002）。コト系名詞が被修飾名詞となる場合、トイウの介在は任意である。益岡(2002)では、トイウが介在しない場合、被修飾名詞は内容補充節によって限定を受けるのに対し、トイウが介在する場合は限定を行うという性質を持たないものとしている。

(13) A: どんな仕事を探していますか。

B: 外国人に日本語を教える仕事を探しています。 （益岡 2002, p. 111）

(14) A: どんな仕事を探していますか。

B: ?外国人に日本語を教えるという仕事を探しています。 （益岡 2002, p. 111）

またトイウが介在する場合のほうが、内容補充節の情報の重要度が高いものとしている。

(15) 私が特に強調したいことは、この問題が既に何度も検討されてきているという事実である。(益岡 2002, p.112)

(16) ?私が特に強調したいことは、この問題が既に何度も検討されてきている事実である。(益岡 2002, p.112)

これらの事実と、連体形式をとみなわない補足語修飾節が限定修飾を行うことを考え合わせると、トイウが非限定修飾を行うのに対し、補足語修飾節やトイウの介在しない内容補充節は限定修飾を行うものとして区別することができる。

コト系名詞では、トイウが介在せず内容補充節が限定修飾を行う場合と、トイウが介在し非限定修飾が行われる場合がある。トイウが介在した場合、内容補充節は被修飾名詞を限定するといった従属的な役割でなく、被修飾名詞から独立した情報的価値を持つものとして存在すると考えられる。

コト系名詞が対事的ムードと共に起する例は出現の可能性が低いが、まれに見られるとされている(益岡 2002, p.115)。

(17) 他の投手では1イニングでも押さえるのが難しいだろうという場面
(益岡 2002, p.115)

引用系名詞とコト系名詞に共に起するトイウの共通の統語的・意味的機能として、「といった」と対立をなすという点がある。「A という N」はNの具体的な内実を示す同格的な関係を形成するのが基本であるのに対し、「A といった N」はNの具体的な内実を唯一それだけとするのではなく、幅を持たせて示す関係を形成するとされている(山崎・藤田 2001)。

4.3 トイウと名詞節

名詞節は、その節全体が名詞の働きをしているものであり、「～こと」、「～の」、「～か」という形になっている(野田 2002)。

(18) だれとでも仲良くなれるのが私の特技です。(野田 2002, p.11)

「～こと」、「～の」の名詞節内部には、助詞のトイウが現れる。

(19) だれとでも仲良くなれるというのが私の特技です。

(20) 自分が単なる一個人でないということを、これほど痛切に感じたことはなかった。
(LBj9_00202)

コトやノは形式名詞であり、名詞として実質的意味を持たないので、引用系名詞やコト系名詞と異なり、連体修飾節との関係を内容補充という意味関係で捉えることができない。名詞節におけるトイウは、(15)でみたような情報の重要度を示す取り立て詞のような働きをしていると考えられる。

また名詞節に現れるトイウは「といった」との対立を持たないという点も、引用系名詞やコト系名詞と異なる点である。

5. 今後の課題

本稿では、複合辞トイウの機能の重層性に注目し、助動詞用法と助詞用法のそれぞれに

ついて論じた。助詞用法については、共起する節の性質により少しずつ異なる機能を果たしており、「機能語としての性質の獲得」の様々なレベルをみることができる。しかし、共起する節の分類については不十分な点が多い。特に、共起する節のモダリティに関しては、詳細な検討を要する。

本稿の分析では、補足語修飾節とトイウの介在しない内容補充節の機能を限定修飾としたが、談話における被修飾名詞の特性により、これらの修飾節が限定を行わない場合もある。例えば、補足語修飾節の被修飾名詞が固有名詞など限定を必要としない名詞句の場合は、補足語修飾節は背景的情報となる。連体修飾節をとると、文がコンパクトにまとまると同時に、主節の事態に対する背景的な事態を被修飾名詞に先立って述べるため、その事態が背景であることを明示的に示しやすいという利点が指摘されている（日本語記述文法研究会 2008）。

(21) さっきまで暴れていた子どもたちが、ふと見ると寝ていた。

（日本語記述文法研究会 2008, p.91）

限定修飾と非限定修飾の違いについては、談話中での機能という観点に加え、トイウの介在を許さない連体修飾節（「私が食べた残り」「さんまを焼く匂い」など）を含めた連体修飾節全体の枠組みの中で検討していく必要がある。

文献

- 金城由美子 (2009) 「書き言葉と話し言葉における『というか』について」 特定領域研究「日本語コーパス」総括班『特定領域研究「日本語コーパス」平成 20 年度公開ワークショップ サテライトセッション予稿集』 pp. 63-68.
- 金城由美子 (2010) 「トイウによる連体修飾節を含む引用構文について—話し言葉としての国会会議録を対象に—」 岸本秀樹編『ことばの対照』くろしお出版. p. 233-243.
- 砂川有里子 (2006) 「「言う」を用いた複合辞—文法化の重層性に着目して—」 藤田保幸・山崎誠編『複合辞研究の現在』和泉書院. pp. 23-40.
- 寺村秀夫 (1975) 「連体修飾のシンタクスと意味—その 1—」 日本語・日本文化 8. (『寺村秀夫論文集 I』(くろしお出版) 所収.)
- 日本語記述文法研究会 (編) (2008) 『現代日本語文法 6 第 11 部 複文』くろしお出版.
- 丹羽哲也 (1993) 「引用を表す連体複合辞「トイウ」」 人文研究大阪市立大学文学部紀要 45:1 pp.25-60.
- 野田尚志 (2002) 「単文・複文とテキスト」『日本語の文法 4 複文と談話』岩波書店. p.3-62.
- 藤田保幸 (2006) 「複合辞研究の展開と問題点」 藤田保幸・山崎誠編『複合辞研究の現在』和泉書院. pp. 3-19.
- 益岡隆志・田窪行則 (1989) 『基礎日本語文法』くろしお出版.
- 益岡隆志 (2002) 「複文各論」『日本語の文法 4 複文と談話』岩波書店. pp.65-116.
- 山崎誠 (1993) 「引用の助詞「と」の用法を再整理する」研究報告集 14 (国立国語研究所), pp.1-22.
- 山崎誠・藤田保幸 (2001) 『現代語複合辞用例集』国立国語研究所.

BCCWJ により選定した日本語教育語彙の検証

ー経営学を専攻する日本語学習者を対象にしてー

寺嶋弘道（立命館アジア太平洋大学言語教育センター）

Surveying and Selecting Specialized Vocabulary for Japanese Learners Majoring in Business Administration

Hiromichi Terajima (Center for language education faculty, Ritsumeikan Asia Pacific University)

1. はじめに

2008 年から国立国語研究所で構築されている「日本語書き言葉均衡コーパス」（以下 BCCW）が応用できる分野の一つとして、日本語教育語彙の選定がある。BCCWJ のような代表性を持った大規模コーパスを用いれば、母集団の量的な構造が推測でき、ある特定分野で特徴的に使用される語（以下 特徴語）を統計的な指標で明らかにできる。このような大規模コーパスを用いた方法は、英語教育の分野では、大学英語教育学会基本語改訂委員会編（2003）による「JACET8000」があり、日本語教育分野においても期待できるものである。

しかし、「従来、コーパスから語彙頻度表を多角的に作成するというような作業は、新聞の面種ごとに語彙調査を行った試みなど少数例を除き、個人レベルではほとんど実現は不可能であった」（丸山 2009, P35）と指摘されているように、大規模コーパスから多角的に語彙頻度表を作成し、それを日本語教育語彙の選定に応用した例は多くない。したがって、BCCWJ からどのようなサンプルを取りだし、語彙頻度表を作成するのか、頻度情報をどのように語彙選定に応用するのかをこれから検討していく必要がある。

そこで、本調査では BCCWJ を用いて経営学を専攻する日本語学習者のための語彙選定を試み、その選定した語彙の効果を検証することで、サンプルの抽出方法及び選定に用いる指標について考察したい。想定した日本語学習者は日本語能力試験 3 級レベルの語彙力を持った学習者で、日本語で経営学を専攻し、就職活動のために日本経済新聞を読むことを目指している学習者である。

本調査では、まず、BCCWJ 内の経営学及び経済学に関連したサンプルからそれぞれの特徴語を尤度比検定（以下 LLR）という方法で抽出し、両分野に共通した特徴語と各分野のみに認められる特徴語を明らかにする。次に、ばらつきを示す指標としてエントロピー値を用い、特徴語の中でも、BCCWJ 内で幅広く使用されているものを優先的に選定する。最後に、その選定した語彙が大学で使用される経営学の教科書及び日本経済新聞をどの程度カバーできるかを検証し、サンプルの抽出方法及び選定に用いる指標について考察したい。

2. 経営学と経済学の関係

本調査では日本語で経営学を専攻し、就職活動のために日本経済新聞を読むことを目標にしている学習者を対象にし、語彙を選定する。経済問題に関連した記事の割合が大きい日本経済新聞を選んだ理由は、日本で就職活動するさいに学生に最もよく読まれる新聞の一つであるからである。したがって、本調査では、経営学だけでなく、経済学に関連した語彙も選定する必要がある。

田中（1998）は経営学が個別資本に着目し、生産、販売、財務、労務といった諸活動や諸現象を対象としていると述べ、経営学の学問的性格を経済学と比較している。同研究では経済学が社会資本に着目し、物価の形成や変動、国民所得の増減、景気循環といった諸現象を対象とする学問であると述べ、人間社会の経済的構造を研究する点、経済的な諸概念を共有している点において経営学と経済学の両分野は共通の基盤があり、お互いに依存し、作用し合っていると指摘している。そして、経営学と経済学を統一する科学として、

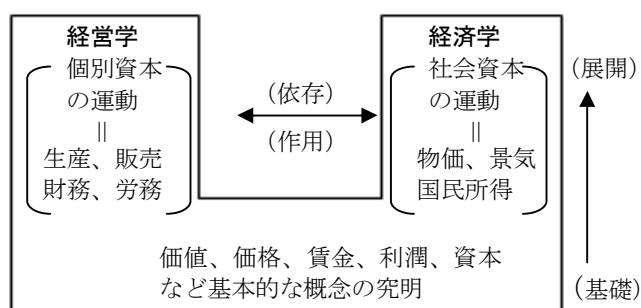


図1 経済科学の構造 出典:田中（1998）,p84

経済科学という枠組みを用い、図1のような関係を示している。つまり、経営学と経済学が隣接した学問として存在し、諸概念を共有しているという考えに従えば、両分野には共通して使用される語があり、それぞれの分野でのみ特徴的に使用される語もあると考えられる。

3. 調査方法

3.1 使用データ

本調査では、『現代日本語書き言葉均衡コーパス』モニター公開データ（2009年度版）（以下 BCCWJ2009 モニター版）を用いた。このコーパスの特徴は母集団の特性が反映されるようなサンプリングを行い、統計的な代表性が保障されていることである。本調査では、このデータのうち、9427 サンプルの書籍データを用いた。ただし、統計的な正確性を期すため、ランダムに指定した1文字から1000文字の範囲を抽出した固定長サンプルというデータを使用している。この書籍データを Unidic で形態素解析を行った結果、空白及び記号を除いた総語数は12,217,563語であった。以下、この9427サンプルを集めたものを書籍全CPと呼ぶことにする。

次に、9427 サンプルから経営学及び経済学のジャンルに該当するサンプルを取り出した。BCCWJ2009 モニター版には、書籍に日本十進分類法（以下 NDC）の番号が付与されている。そこで、本学で開講されている経営学及び経済学に関わる科目を対象とし、その科目のシラバスで明示された教科書と参考文献のNDC番号を調査した。経営学で対象とした科目は、コア科目と呼ばれるものである。コア科目とは、本学で履修を強く推奨し、経営学の専門領域を学ぶ上での基礎と位置づけられているものである。具体的には「経営学」、「会計学

I・II」、「金融論」、「マーケティング論」、「人的資源管理論」、「生産管理論」、「企業研究入門」の8科目である。一方、経済学は「経済学入門」、「現代の日本経済」「ミクロ経済学」、「マクロ経済学」を対象にした。ただし、偶然性を排除するため、複数の指定がある NDC 番号のみを対象にした。以下、前者のサンプルを集めたものを経営学 CP、後者を経済学 CP と呼ぶ。表 1 はそれぞれのコーパスの詳細である。空白及び記号を除くと、経営学 CP の総語数は 576,152 語、経済学 CP の総語数は 144,510 語であった。

表 1 経営学 CP と経済学 CP

	NDC	詳細	サンプル数
経営学 CP	335	経営	86
	336	経営管理	183
	338	金融	106
	366	労働経済・労働問題	50
	602	産業史・事情・物産誌	6
	675	マーケティング	20
経済学 CP	331	経済学・経済思想	22
	332	経済史・事情・経済体制	89

3.2 選定のための指標

あるコーパスに含まれる特徴語を統計的な方法で抽出するためには、量的な構造を比較するためのコーパスが必要である。本調査の場合であれば、書籍全 CP がその役割を持つ。特徴語を抽出するため、本調査では大学英語教育学会基本語改訂委員会編（2003）による「JACET8000」、橋本（2009）でも用いられている LLR を用いた。LLR を求めるためには、表 2 の分割表と＜LLR の計算式＞が必要である。表 2 における対象コーパスとは経営学 CP や経済学 CP のこと、参照コーパスとは書籍全 CP のことである。対象コーパスで特徴的に使用されるほどその統計量は高くなる。ただし、両コーパスのうち、どちらへの偏りがあるかを明確にするために、内山他（2004）のように、表 2 の分割表で $ad-bc<0$ の場合に計算式に -1 を乗じる補正を行った。

表 2 統計量を求めるための分割表

	対象 コーパス	参照 コーパス	計
語 W の頻度	a	b	a+b
語 W 以外の頻度	c	d	c+d
計	a+c	b+d	a+b+c+d=(n)

c = 対象コーパスの総語数 $- a$ 、 d = 参照コーパスの総語数 $- b$

＜LLR の計算式＞

$$\begin{aligned} \text{LLR} = & 2(a\log(a)+b\log(b)+c\log(c)+d\log(d)-(a+b)\log(a+b)-(a+c)\log(a+c) \\ & -(b+d)\log(b+d)-(c+d)\log(c+d)+n\log(n)) \end{aligned}$$

寺嶋（2009）は LLR がカイ 2 乗検定やイエーツの補正公式と比べ、統計量を抑える指標であることを報告しているが、LLR は低頻度であっても特徴的に使用されれば、有意な統計量を示すことがある。そこで、本調査では、中山他（2007）でも用いられたエントロピーという指標も使い、特徴語の中でも特に書籍全 CP で広く使用されている語彙を選定することにした。エントロピー値は見出し語がより多くのサンプルにばらついて出現すれば、その値が高くなり、どれぐらい広く使用されているのか、偏って使用されているのかを知ることができる指標である。エントロピー値は、以下の計算式で求められる。

<計算式>

$$H_w = -\sum_i \frac{f_{wi}}{F_w} \log_2 \frac{f_{wi}}{F_w} \quad ※F_w = \text{語 } W \text{ の総頻度} \quad f_{wi} = \text{各サンプル内での語 } W \text{ の頻度}$$

たとえば、サンプルが 5 つあり、見出し語 W が各サンプルに 4 回出現すると、見出し語 W のエントロピー値 H_w は以下の<計算方法>で求めることができる。5 つのサンプルに 4 回ずつ出現すると、2.32 ビットという値が得られるが、表 3 のケース 1 のように、各サンプルの頻度が全て 2 回になっても同じ確率で出現することは変わらないため、エントロピー値は変わらない。しかし、ケース 2～6 のように、各サンプルでの出現確率にばらつきがあると、エントロピー値は低くなる。

<計算方法>

$$\begin{aligned} H_w &= -(4/20 \times \log_2(4/20) + 4/20 \times \log_2(4/20) + 4/20 \times \log_2(4/20) + 4/20 \times \log_2(4/20) + \\ &\quad 4/20 \times \log_2(4/20)) \\ &= 2.32 \text{ ビット} \end{aligned}$$

表 3 ばらつきの違いによるエントロピー値の変化

	ケース 1	ケース 2	ケース 3	ケース 4	ケース 5	ケース 6
サンプル 1	2	3	2	5	7	10
サンプル 2	2	2	2	2	2	0
サンプル 3	2	2	3	2	1	0
サンプル 4	2	2	3	1	0	0
サンプル 5	2	1	0	0	0	0
エントロピー値	2.32	2.25	1.97	1.76	1.16	0.00

4. 語彙選定の手順

まず、経営学 CP と書籍全 CP、経済学 CP と書籍全 CP の組み合わせで LLR を計算し、統計量が 3.84 以上の語を抽出対象とした。LLR はカイ 2 乗検定と同じ臨界値を用いるため、自由度 1 の分割表の場合、統計量が 3.84 以上あれば、有意な結果だといえる。また、特徴の中でも書籍 CP で広く使用されている語彙を抽出するため、書籍 CP におけるエントロピー値を計算し、4 ビット以上である語を抽出対象にした。図 2 に抽出の流れを示している。

最終的に、経営学と経済学の共通の特徴語が 831 語、経営学の特徴語が 1186 語、経済学の特徴語が 556 語、合計 2573 語を選定することにした¹。しかし、本調査では日本語能力試験 3 級レベルの語彙を習得していることを前提としているため、この 2573 語にはそれらの語は含んでいない。また、数詞や固有名詞、助詞、助動詞なども対象から外した。表 4 は選定した語彙のうち、経営学と経済学の共通特徴語、50 語を示している²。

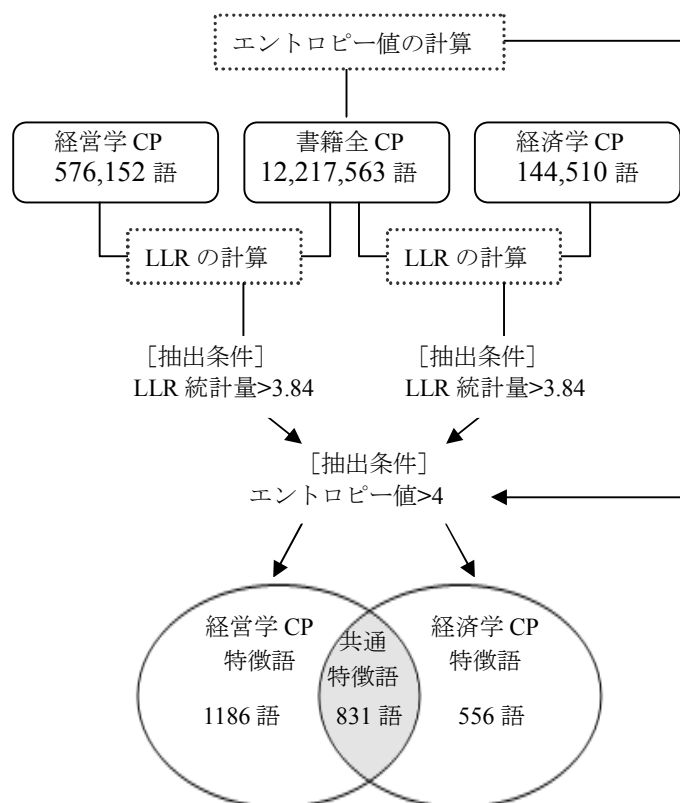


図 2 選定の流れ

表 4 経営学と経済学の共通特徴語

	語彙素	品詞	頻度			LLR		EPY
			書籍 全 CP	経営学 CP	経済学 CP	経営学 CP	経済学 CP	
1	企業	名詞-普通名詞-一般	4008	1840	414	5031.0	1151.9	8.7
2	投資	名詞-普通名詞-サ変可能	1164	658	122	1998.3	342.3	7.2
3	経営	名詞-普通名詞-サ変可能	2000	780	98	1943.0	152.3	8.4
4	金融	名詞-普通名詞-サ変可能	1132	616	150	1836.7	481.4	7.1
5	労働	名詞-普通名詞-サ変可能	2502	814	190	1807.7	429.6	7.9
6	資産	名詞-普通名詞-一般	920	508	24	1525.4	16.1	6.6
7	取り引き	名詞-普通名詞-サ変可能	1020	510	56	1457.6	97.0	7.2
8	市場	名詞-普通名詞-一般	1636	570	226	1323.9	741.7	8.0
9	事業	名詞-普通名詞-一般	2452	654	48	1259.1	17.2	8.2
10	者	接尾辞-名詞的-一般	19216	1962	264	1227.2	23.6	11.3

	語彙素	品詞	頻度			LLR		EPY
			書籍 全 CP	経営学 CP	経済学 CP	経営学 CP	経済学 CP	書籍 全 CP
11	業務	名詞-普通名詞-一般	872	410	42	1134.7	64.2	7.0
12	戦略	名詞-普通名詞-一般	738	374	22	1076.2	18.5	6.8
13	会計	名詞-普通名詞-サ変可能	598	336	12	1017.3	4.6	5.8
14	株式	名詞-普通名詞-一般	878	372	64	972.2	140.3	6.4
15	社員	名詞-普通名詞-一般	782	342	14	909.9	3.9	6.9
16	的	接尾辞-形状詞的	35032	2670	952	868.7	690.7	11.9
17	計算	名詞-普通名詞-サ変可能	1164	382	28	853.1	16.2	7.5
18	管理	名詞-普通名詞-サ変可能	2082	488	42	846.6	16.3	8.3
19	金利	名詞-普通名詞-一般	410	256	74	813.9	277.0	4.6
20	資金	名詞-普通名詞-一般	1050	348	80	782.3	181.3	7.7
21	金	接尾辞-名詞的-一般	1618	414	54	771.5	53.6	7.9
22	預金	名詞-普通名詞-サ変可能	414	238	34	728.4	81.3	5.5
23	利益	名詞-普通名詞-一般	1248	356	50	720.7	62.5	7.9
24	商品	名詞-普通名詞-一般	1322	352	86	676.7	172.5	7.9
25	資本	名詞-普通名詞-一般	1270	342	290	663.2	1203.5	7.3
26	基準	名詞-普通名詞-一般	1128	324	28	659.2	17.2	7.8
27	化	接尾辞-名詞的-サ変可能	7002	818	326	641.2	481.5	10.3
28	財務	名詞-普通名詞-一般	256	186	8	631.2	7.2	5.5
29	コスト	名詞-普通名詞-一般	554	238	24	626.9	32.8	6.6
30	製品	名詞-普通名詞-一般	1032	300	82	615.7	191.6	7.2
31	ビジネス	名詞-普通名詞-一般	674	248	42	596.1	81.3	7.0
32	組織	名詞-普通名詞-サ変可能	2086	406	50	595.2	28.7	8.5
33	パーセント	接尾辞-名詞的-助数詞	3236	492	192	552.6	356.5	8.7
34	組合	名詞-普通名詞-一般	762	248	16	550.7	6.9	6.1
35	型	接尾辞-名詞的-一般	2122	388	60	534.5	46.4	8.2
36	株	名詞-普通名詞-一般	476	200	10	520.2	4.3	5.9
37	銘柄	名詞-普通名詞-一般	206	152	6	519.0	4.9	4.4
38	制度	名詞-普通名詞-一般	2416	402	118	500.9	182.8	8.5
39	収益	名詞-普通名詞-一般	380	178	12	491.6	11.0	6.1
40	機関	名詞-普通名詞-一般	1690	332	48	491.1	37.4	8.4
41	システム	名詞-普通名詞-一般	1694	332	78	490.0	113.8	8.1

	語彙素	品詞	頻度			LLR		EPY
			書籍 全 CP	経営学 CP	経済学 CP	経営学 CP	経済学 CP	書籍 全 CP
42	性	接尾辞-名詞的-一般	8598	828	182	463.2	79.7	10.7
43	社	接尾辞-名詞的-助数詞	362	164	16	444.8	22.3	6.1
44	上場	名詞-普通名詞-サ変可能	198	134	10	441.3	16.0	4.8
45	株主	名詞-普通名詞-一般	604	194	60	427.0	162.8	5.4
46	価値	名詞-普通名詞-一般	1886	326	82	423.4	112.5	8.4
47	人材	名詞-普通名詞-一般	400	160	16	404.4	20.0	6.6
48	費用	名詞-普通名詞-一般	752	204	16	397.8	7.1	7.2
49	従業	名詞-普通名詞-一般	444	164	20	395.1	28.5	6.5
50	ベンチャー	名詞-普通名詞-一般	202	124	10	391.2	15.7	4.4

※EPY=エントロピー

5. カバー率調査による検証

最後に、選定した語彙が大学で使用される経営学の教科書及び日本経済新聞をどの程度カバーできるか、そのカバー率を調査し、サンプルの抽出方法及び選定に用いた指標について考察したい。本調査でいうカバー率とは、選定した語彙が対象となるテキストをどれぐらいカバーできるかという割合のことである。なお、カバー率の調査では、①経営学 CP と経済学 CP によって選定した語彙 (2573 語)、②経営学 CP だけから LLR が高い順に選定した語彙 (2576 語)、③経済学 CP だけから LLR が高い順に選定した語彙 (2575 語)、の 3 つを比較したい³。①-②で共通に含まれる語の割合は 59%、①-③は 54%、②-③は 24% であった。なお、本調査は、日本語能力試験 3 級レベルの語彙を習得した後、どのような語彙を教えるべきかという立場で進めているため、選定した語彙に 3 級レベルの語彙を加えたもので、カバー率を調査した。

5.1 経営学の教科書のカバー率

経営学の教科書を選ぶにあたり、インターネット上でシラバスが公開され、経営学の講義を持つ 30 の大学を調査し、経営学の講義で使用されている教科書を調査した。その結果、複数の大学で使用されている教科書が 5 冊あることがわかった。参考にした講義は、各大学で最も基礎と位置付けられる経営学の講義である。本カバー率調査の対象にしたのは、乱数を発生させたページから 7000 字分である。使用した書籍を表 5 に、カバー率調査の結果を表 6 に示す。

表 5 調査対象とした経営学の教科書

書籍番号	書籍情報	ページ	採用数
1	経営能力開発センター (2006) 『経営学検定試験公式テキスト 1 経営学の基本』中央経済社編	212	4 大学

2	片岡信之他（2008）『はじめて学ぶ人のための経営学入門』文真堂	161	2 大学
3	小松章（2006）『経営学（第2版）』新世社	90	3 大学
4	津田真徴（1977）『人事労務管理の思想』有斐閣	94	2 大学
5	奥林康司他（2007）『経験から学ぶ経営学入門』有斐閣	23	2 大学

表 6 経営学の教科書のカバー率

	書籍 1	書籍 2	書籍 3	書籍 4	書籍 5
①経営学 CP 及び 経済学 CP による選定	87.7%	90.1%	88.0%	83.0%	91.5%
②経営学 CP のみ による選定	84.0%	85.0%	85.8%	75.3%	85.7%
③経済学 CP のみ による選定	75.7%	81.6%	78.4%	74.0%	85.0%

どの書籍においても、③<②<①という順でカバー率が高くなることがわかる。経営学の教科書に対するカバー率調査であるため、経済学 CP から選定した語彙が経営学 CP から選定した語彙よりカバー率が低くなったのは当然のことだといえよう。しかし、両分野に共通している特徴語、エントロピー値が高く、幅広く使用される特徴語を優先的に選定することで、経営学の教科書に対するカバー率が上がる可能性が高くなることも確認できる。

一方、①でカバーできない語の中には、「向ける」、「触れる」「互いに」など専門的な特徴が見られないが、日常的によく使用されるようなものが多いことがわかった。今後、書籍全 CP から使用頻度が高い語彙を取りだし、それを選定した語彙に加えることで、さらに高いカバー率が得られると考えられる。

5.2 日本経済新聞のカバー率

次の調査では、日本経済新聞の朝刊（全国版）の一面記事を対象にした。2005 年 1 月 1 日～2009 年 12 月 31 日の過去 5 年間の全ての日に対し、乱数を発生させ、5 日分の記事を対象とした。一面にある記事の数は 7 つ～8 つである。対象とした記事の日付とカバー率調査の結果を表 7 に示す。

この結果から①は、どの日にちにおいても最も高いカバー率が得られることがわかるが、②と③に関しては、日にちによって順位が変わることがわかる。これにより、日本経済新聞には、経済学で使用されるような語彙が含まれることはもちろん、経営学で使用されるような語彙も多く含まれており、両分野を組み合わせると選定することが高いカバー率を得られる方法だといえる。しかし、経営学の教科書に対するカバー率と比べると、日本経済新聞のカバー率は全体的に低いといえる。①では、経営、経済に関連した語彙は非常に高い確率でカバーされているが、「閣僚」、「首脳」など政治に関連した語があまりカバーできていないようである。また、カバーできてい

なかった語として、「累積」、「貶す」、「複線」などの難易度が高く、書籍全 CP で使用頻度が低い語も多く見られた。今後、このような語彙をどう扱うのか、扱うのであれば、どのような視点で抽出していくのかを考える必要がある。

表 7 日本経済新聞のカバー率

	2008/12/30	2009/12/8	2005/2/12	2006/3/6	2007/5/9
①経営学 CP 及び経済学 CP からの選定	82.3%	78.5%	79.8%	75.4%	79.7%
②経営学 CP のみによる選定	75.1%	69.7%	74.1%	65.6%	77.1%
③経済学 CP のみによる選定	75.4%	70.3%	67.4%	66.6%	68.0%

6. まとめ

以上、本調査では BCCWJ を用いて経営学を専攻する日本語学習者のための語彙選定を試み、その選定した語彙の効果を検証し、サンプルの抽出方法及び用いた指標について考察した。

その結果、語彙選定のプロセスの中で、経営学と経済学に共通して使用される経済科学語彙の存在を明らかにすることができ、経営学に関連した語彙だけでなく、経済学にも関連し、書籍全 CP で広く使用されている語彙を優先的に選定することで、経営学の教科書、日本経済新聞をよりカバーできることがわかった。しかし、カバー率調査の結果、専門的な特徴が見られないが、よく使用される語彙や政治分野の語彙などが不足していたことがわかった。今後、調査を続け、必要だと考えられる語彙を本調査で得られた語彙に加えていきたいが、そのような語彙をどのように取り出すかは、さらなる調査が必要である。

また、LLR とともに用いたエントロピー値がカバー率を上げる可能性があることもわかった。本調査でエントロピー値がどの程度の効果を発揮したのか、引き続き調査が必要であるが、今後も複数の指標を用い、効果的な語彙選定の方法を検討していきたい。

注

1. この調査での言語処理、指標計算は、Active Perl5.10、Microsoft Office Access2003 を用いた。
2. この調査では短単位での選定であるため、実際に学習者に示す場合は、長単位での提示を検討する必要がある。
3. 同順位のものがあるため、数が若干異なっている。

文献

内山 将夫, 中條 清美, 山本 英子, 井佐原 均 (2004) 「英語教育のための分野特徴単語の選定尺度の比較」『自然言語処理』11 巻 3 号, 165-197
 大学英語教育学会基本語改訂委員会編 (2003) 『大学英語教育学会基本語リスト JACET List

of 8000 Basic Words』大学教育学会

田中照純（1998）『経営学の方法と歴史』ミネルヴァ書房

寺嶋弘道（2009）「日本語教育語彙を選定するための統計的指標　－尤度比検定、カイ 2 乗検定、イエーツの補正公式の特徴－」『ポリグロシア』17 巻,71－83

中山恵利子,桐生りか,山口昌也（2007）「第 3 章 新聞に見る基幹外来語」『公共媒体の外来語「外来語」言い換え提案を支える調査研究（国立国語研究所報告 126）』, 343－354

橋本直幸（2009）「BCCWJ を利用した日本語教育語彙リスト作成の試み」『特定領域研究 日本語コーパス平成 20 年度公開ワークショップ（研究成果報告会）予稿集』 特定領域研究日本語コーパス総括班, 183－190

深尾百合子（1999）「専門日本語教育」研究に期待するもの『専門日本語教育研究』 第 1 号, 6－9

前川喜久雄,山崎誠（2009）「現代日本語書き言葉均衡コーパス」『国文学解釈と鑑賞』74 巻 1 号,至文堂, 15－25

丸山岳彦（2009）「現代日本語書き言葉均衡コーパスから見えるもの」『日本語教育』140 号, 日本語教育学会, 26－36

使用コーパス

国立国語研究所（2009）『現代日本語書き言葉均衡コーパスモニター公開データ（2009年度版）』

書 名 特定領域研究「日本語コーパス」平成21年度公開ワークショップ サテライトセッション予稿集
発行日 平成22年3月11日
発行者 文部科学省科学研究費特定領域研究「日本語コーパス」総括班
<http://www.tokuteicorpus.jp/>
連絡先 〒190-8561 東京都立川市緑町10-2
大学共同利用機関法人 人間文化研究機構 国立国語研究所コーパス開発センター内
電 話 042-540-4300 (代表)
文書管理番号 JC-G-09-02
