

C-4 話し言葉の経年変化の定量的記述 —『昭和話し言葉コーパス』 『日本語日常会話コーパス』『日本語話し言葉コーパス』を用いて—

佐野真一郎
慶應義塾大学

【要旨】本研究では、言語データの収録時期・スタイルの異なる3種類のコーパスを用いて、話し言葉の経年変化を定量的に捉えることを目的とした。調査対象としたのは、「凄い」の活用形、テイル形、「全然」の呼応、文末丁寧表現である。調査の結果、各現象について経年変化、及びスタイルによる影響が確認されると共に、それらの使用実態が明らかとなった。これらの結果から、日常語や比較的新しい用法ほど、1) 近年のデータに、2) くだけたスタイルに現れやすい、反対に伝統的な表現は、3) 往年のデータに多く、4) 改まったスタイルにおいて保たれやすいと言える。また、話し言葉・書き言葉といったジャンル差だけでなく、話し言葉の中でのスタイル差も言語の経年変化に影響を与えることを確認した。とりわけ、改まったスタイルは書き言葉と同様、言語使用の経年変化に対して一定の耐性を持つと考えられる。

1. はじめに

近年の大規模コーパスの拡充に伴い、これを用いた言語研究・分析が高まりを見せている。この流れの中で、話し言葉・書き言葉それぞれの中でも、性格を異にする複数のコーパスが利用可能となってきた。話し言葉を収録したコーパスに目を向けると、独話を中心とした『日本語話し言葉コーパス』(以下、CSJ とする)、様々な場面での雑談を収録した『名大会話コーパス』、日常場面での自発的な会話を映像資料付きで提供する『日本語日常会話コーパス』(以下、CEJC とする)、昭和期の録音音声をもとめた『昭和話し言葉コーパス』(以下、昭和コーパス、または昭和とする)などがある(佐野 2018)。このことにより、複数のコーパスを用いて、話し言葉・書き言葉それぞれの中でのスタイル差や時代差を比較することが可能となっている。

現代日本語において、これまでに格助詞の交替(例、太郎が／の買った本、南部 2007 他)、「ら抜き言葉」(Matsuda 1993 他)、など様々な言語変化が報告されてきているが、言語の揺れ・変化、中でも比較的新しい特徴は、話し言葉において観察されやすい。書き言葉では、校正・編集の過程でこのような特徴が修正され得ることもその理由の一つである。加えて、話し言葉特有の現象も存在する。これらは、話し言葉を対象として言語の揺れ・変化を研究することに妥当性が認められることを示している。また、話し言葉(音声)コーパスは、その性格上縦断的な研究に耐えるものを構築するには困難を伴い、近代語の音声を対象としたコーパスは長らく現れなかったが、『昭和話し言葉コーパス』は話し言葉の縦断的研究の可能性を提供する。以上を踏まえ、本研究では、言語使用の揺れ・変化が見られる「凄い」の活用形、テイル形、「全然」の呼応、文末丁寧表現に注目し、複数の話し言葉コーパスの分析を通してこれらの経年変化を定量的に記述する。

2. 対象とする現象

本節では、本研究において分析対象とした言語現象4項目について、その特徴を例を示しつつ紹介する。

2.1 「凄い」の活用形の揺れ

現代語における「凄い」は、「大変」「非常に」などと同様、程度性を強調する意味を持つ形容詞としての文法機能を持つ。例(1)、(2a, b)のように、従来「凄い」は体言接続の場合、連体形「すごい」となり、用言接続の場合は連用形「すごく」となる。しかしながら、例(2c, d)と表1に示すように、用言接続において体言接続であるはずの連体形「すごい」が観察されるようになっており、この揺れが近年専門家・非専門家の間で指摘されている(堀尾 2014 他)。

- | | | | |
|----------|------------------------------|------|-------------------|
| (1) 体言接続 | a. もう <u>すごい</u> 強風で | 名詞 | (CEJC: C001_001) |
| | b. <u>すごい</u> のしかないもんね | 準体助詞 | (CEJC : K004_008) |
| (2) 用言接続 | a. それは <u>すごく</u> 感じた | 動詞 | (CEJC: T004_006) |
| | b. <u>すごく</u> いいところに住んでるんですね | 形容詞 | (CSJ: S03F0224) |
| | c. あたしもきょうも <u>すごい</u> 迷ったけど | 動詞 | (CEJC: C001_012) |
| | d. <u>すごい</u> うまいですよ | 形容詞 | (昭和: dt00413) |

表 1. 「凄い」の活用形の揺れ

活用形（規範）		活用形（現在の言語使用）
体言接続	連体形 <u>すごい</u>	連体形 <u>すごい</u>
用言接続	連用形 <u>すごく</u>	連用形～連体形 <u>すごく～すごい</u>

本稿では、この用言接続における「すごく／すごい」の選択を第一の調査項目として、実際の発話における分布を調べた。

2.2 テイル形の揺れ

テイル形については、動詞テ形に「イル」が接続するか「オル」が接続するかによって、「～している／～しておる」のようなテイルとテオルの間の揺れが見られる（服部 2009 他）。両者を構成する「イル」「オル」間には、語用論的・言語外的な意味・機能の違いがあり、「イル」に対して「オル」は古風な、尊大な、方言的な印象を与え、また「マス」を伴って丁寧な表現となるなど、スタイルに関わる違いがある。以下に、コーパスで観察された両形式の用例を示す。

- | | | |
|----------|---|-------------------|
| (3) テイル形 | a. ちびちゃんがめっちゃめっちゃ食 <u>って</u> いたような気がする | (CEJC: K001_003b) |
| | b. 今日はそのことを話したいと思 <u>って</u> います | (CSJ: S00F0066) |
| (4) テオル形 | a. そういう基本語というものを知 <u>って</u> おれば、非常に有効だと | (昭和: dt06317) |
| | b. 検討をしていき <u>たい</u> と思 <u>って</u> おります | (CSJ: A01M0110) |

実際の発話において、例 (3), (4) のような「テイル」「テオル」のどちらが現れるのか、その選択のパターンを第二の調査項目とした。

2.3 「全然」の呼応の揺れ

陳述副詞「全然」は、否定辞など特定の表現と共起する「呼応」と呼ばれる特性を持つ。その呼応する表現の種類に関しては変化が見られ、明治期には肯定的表現とも否定的表現とも呼応していたが、大正の終わりとから肯定的表現との呼応が見られなくなり、結果として現在の規範文法に記されているように、否定的表現との呼応が正しいとされるようになった。その後昭和後期ごろから肯定的表現との呼応が再度見られるようになり、近年それが増加している（鈴木 1993, 野田 2000, 服部 2010, 佐野 2012, Sano 2016 他）。つまり、例 (5), (6), (7) に示すように、「全然」の呼応パターンには現在揺れが見られるのである。

- | | | |
|--------|--|------------------|
| (5) a. | <u>全然</u> 伝 <u>わ</u> らない <u>み</u> たいな | (CEJC: T002_019) |
| | b. こういう責任を負うということは <u>全然</u> 存じ <u>ま</u> せん <u>で</u> | (昭和: dt06294) |
| (6) a. | 何か昔とは <u>全然</u> 違 <u>う</u> ん <u>で</u> す <u>け</u> れ <u>ど</u> も | (CSJ: S03F0108) |
| | b. 足を突っ込んだのは長い <u>ん</u> です <u>け</u> れ <u>ど</u> も、 <u>全然</u> 駄 <u>目</u> な <u>ん</u> です | (昭和: dt00472) |

- (7) a. あもう全然そんなのいいよ (CSJ: S01F0183)
 b. 私が集めてるだけなんで全然大丈夫です (CEJC: T007_007)

本研究では、例 (5) に示すような「全然」と否定辞との呼応を「ない類」、例 (6) に示すような「違う」「ダメ」などとの呼応を「違う・駄目類」、例 (7) に示すような「いい」「大丈夫」などとの呼応を「いい・大丈夫類」とし、実際の発話におけるこれら 3 種類の呼応の分布を第三の調査項目とした。

2.4 文末丁寧表現の揺れ

名詞や形状詞に後続する断定の助動詞として機能する文末丁寧表現には「です」の他に、「だ」の連用形に「あります」「有る」「在る」の連用形+丁寧の助動詞「ます」や「ございます」「有る」「在る」に対応する丁寧語「ござる」の連用形+丁寧の助動詞「ます」を伴う「であります」「でございます」などがあり、これらは丁寧度などによって選択される (蒲谷ら 1998 他)。

- (8) a. スーツケースがなければ十分歩ける距離です (CEJC: K001_014)
 b. 白い花が咲いてとても綺麗です (CSJ: S03F0383)
 (9) a. 主として、学校文法、文法教科書の問題でありまして (昭和: dt06309)
 b. 動作や結果といった概念も分析には有効であります (CSJ: A02M0107)
 (10) a. お待たせいたしました、ブレンドコーヒーでございます (CEJC: K001_014)
 b. そういう点では大変に環境もよく幸せでございました (昭和: dt06293)

本研究では、例 (8) に示すような「です」、例 (9) に示すような「であります」、例 (10) に示すような「でございます」の使い分けを揺れと捉え、これを第四の調査項目として実際の発話における分布を調べた。

3. 方法

3.1 コーパス

ここで、本研究で利用した話し言葉を対象とした 3 種類のコーパスを紹介する。

『昭和話し言葉コーパス』(丸山 2016) は、1950 年代から 1970 年代まで (1950 年代が中心) 国立国語研究所で収集された録音資料約 50 時間分の音声を取りまとめ、コーパスとして一般公開するものである (現在構築中)。収録データは会話・雑談に加えて、講演などの独話も含む。約 60 年前の日常談話資料が話し言葉コーパスとして利用可能となることで、話し言葉の変遷に関して手掛かりを得ることができるだろう。本研究では、収録した録音資料約 50 時間分のパイロットデータを元に調査を行った。

『日本語日常会話コーパス』(小磯ら 2017) は、人工的な課題や状況における会話ではなく、多人数による日常場面で自発的に生じる会話や雑談約 200 時間を記録したものである。会話の情報は、書き起こしテキストと音声に加えて、映像でも提供される。本コーパスも現在構築段階にあり、データ収録は 2016 年から始まり継続中であるが、2018 年 12 月に 50 時間分の会話データがモニター公開された。本研究では、このモニター公開版を調査対象とした。

『日本語話し言葉コーパス』(前川ら 2000) は、約 651 時間・752 万語の自然発話を収録した大規模自発音声コーパスである。その大部分は「学会講演」「模擬講演」と呼ばれる独話によって構成されている。データ収録は 1999 年から 2003 年にかけて行われた。発話サンプルには、書き起こしテキストに加え、豊富な研究用付加情報 (アノテーション) が付与されているが、本研究ではより詳細なアノテーションが施された「コア」と呼ばれる約 45 時間・50 万語を調査対象とした。

本研究と関係のある各コーパスの特徴は、収録されている発話がいつのものであるかということと、その

発話がどのようなスタイル的特徴を持つかということであり、表2のようにまとめられる。

表2. コーパスの特徴

コーパス	データ収録時期		スタイル	
昭和	1950年代中心	旧	会話・雑談 (+講演などの独話)	中間
CEJC	2016~継続中	新	会話・雑談	くだけた
CSJ	1999~2003年	新	講演などの独話 (+対話)	改まった

データ収録時期の点で、CEJC, CSJ 中の発話は昭和コーパス中の発話よりも新しいと考えられる。また CEJC ではくだけた発話スタイルとなり、反対に CSJ では改まった発話スタイルとなる。昭和コーパスはその中間であると考えられる。本研究では、昭和コーパスを約 60 年前の話し言葉の代表、CEJC を現在のくだけた話し言葉の代表、そして CSJ を現在の改まった話し言葉の代表として捉えることとする。

3.2 用例検索

それぞれの現象について、文字列・形態論情報を参照し、コーパスから目的となるデータを以下の手順に従って検索・抽出した。

昭和コーパスには『ひまわり』(ver. 1.6) (山口 2014) を用いて転記テキストデータに Mecab 0.996 による形態素解析を施し、検索を行った。CEJC, CSJ には『中納言』(ver. 2.4.4) (小木曾・中村 2013) を用い、CEJC は提供されている全データ、CSJ は「コア」を対象としてオンラインで検索した。いずれの場合も、結果をエクスポート・ダウンロードし、Microsoft® Excel for Mac (ver. 16.9) で処理した。

ひまわりでは、言語現象ごとに概略以下のような文字列・条件を指定し、検索を行った。

- 【「凄い」の活用形】 検索文字列：基本形="すごい" or "ものすごい"
- 【テイル形】 検索文字列：基本形="て", フィルタ：基本形 1="いる" or "おる"
- 【「全然」の呼応】 検索文字列：基本形="全然"
- 【文末丁寧表現】 検索文字列：品詞¹="名詞", "連体詞", フィルタ：基本形 1="です"
検索文字列：基本形="だ", フィルタ：基本形 1="ある" or "ござる", 基本型 2="ます"

中納言では、言語現象ごとに以下の文字列・条件を指定し、検索を行った。その他「検索動作」等の設定はデフォルトのまま操作していない。

- 【「凄い」の活用形】 前方共起条件：語彙素="凄い" or "物凄い", キー：指定なし
- 【テイル形】 前方共起条件：語彙素="て", キー：語彙素="居る"
- 【「全然」の呼応】 前方共起条件：語彙素="全然", キー：指定なし
- 【文末丁寧表現】 キー：品詞=名詞, 形状詞, 代名詞, 準体助詞, 後方共起条件：語彙素="です"
キー：品詞=名詞, 形状詞, 代名詞, 準体助詞, 後方共起条件：語彙素="だ", 語彙素="有る" or "御座る", 語彙素="ます"

3.3 分類

次に、コーパスから抽出した用例について、本研究の対象とならないものを除き、数量的分析ができる形に整理・分類した。その際、ひまわり・中納言での品詞情報に基づく機械的な分類を基にして、文脈を参照して一例一例吟味して分類した。言い直し・言い指し・繰り返しなどは対象から除いた。

¹ 事前の調査から、「です」「であります」「でございます」が交替可能となる先行要素は名詞、連体詞であるため、ひまわりでの検索条件中の品詞を「名詞」「連体詞」と指定した。同様の検索結果を得るため、中納言では、名詞、形状詞、代名詞、準体助詞と指定した。

【「凄い」の活用形】抽出した用例のうち、体言接続のもの、「すごかった」「すごさ」、また係り先の無いものを除いた。また、助動詞「です」は「すごくです」とはならず「すごいです」にしかならないため、これを後続要素とする用例は除いた。残りを、「すごく」「ものすごく」、または「すごい」「ものすごい」「すげえ」「すんげえ」を含むに分類した。

【テイル形】抽出した用例のうち、1)「ない」が後続する場合、テイル形に限られる(*「ておらない」)²、2)「ず」が後続する場合、テオル形に限られる(*「ていず」)、尊敬の「ラレル」が後続する場合、テオル形に限られる(*「ていられる」)。これらはテイル・テオル両形式の出現可能性が確保できないため除いた。

【「全然」の呼応】コーパスから抽出した用例を、その呼応表現に基づき、文脈を参照しながら「ない」「違う・駄目」「いい・大丈夫」の3類に分類した。

【文末丁寧表現】上掲の3種類の検索条件から得られた用例を、その語彙素に基づきそれぞれ「です」「であります」「でございます」の3類に分類した。

4. 分析

以上の手続きを経て、4項目合計 31,205 件の用例が得られた。これを分析用データセットとする。調査項目ごとに分析を行い、R 言語 (R Development Core Team 1993–2019) を用いた混合効果回帰分析により、分布特徴の有意性を検定した。用例の選択を従属変数、コーパスの種類を目的変数とし、話者・単語を変量効果とした。多重比較には Steel-Dwass 法、Tukey 法を適用した。

4.1 「凄い」の活用形の揺れ

3つのコーパス全ての間で統計的有意差(1%水準)が認められた。図1に示す昭和コーパスと CEJC・CSJ との違い($t=15.84, t=3.24$)は、約60年間で「凄い」の活用形の用法が変化していることを示しており、「すごい」が増加してきていると考えられる。また、CEJC と CSJ の違い($t=18.74$)は、スタイルの違いが「すごい」「すごく」の選択に影響していることを示しており、「すごい」はくだけた発話に、「すごく」は改まった発話に親和性が高いと考えられる。

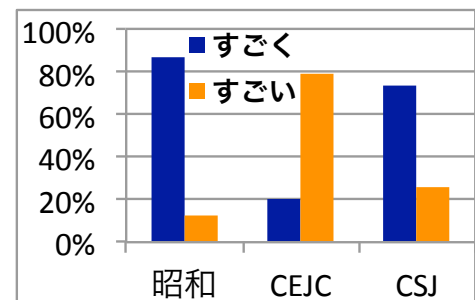


図1. 「すごい・すごく」の分布

4.2 テイル形の揺れ

3つのコーパス全ての間で統計的有意差(1%水準)が認められた。図2に示す昭和コーパスと CEJC・CSJ との違い($t=10.65, t=21.65$)は、約60年間でテイル形の用法が変化していることを示しており、「テイル」が増加してきていると考えられる。また、CEJC と CSJ の違い($t=4.46$)は、スタイルの違いが「テイル」「テオル」の選択に影響していることを示しており、「テイル」はくだけた発話で、「テオル」は改まった発話においてより高い頻度で用いられると考えられる。

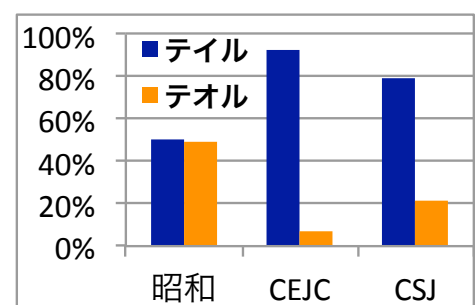


図2. 「テイル・テオル」の分布

4.3 「全然」の呼応の揺れ

「全然」の呼応については、CEJC と他の2つのコーパスとの違いは認められたものの(1%水準)、昭和コーパスと CSJ との間に有意な差はなかった。図3が示すように、CEJC と昭和コーパス・CSJ の違い($z=-5.59, z=-3.81$)は、CEJC では「ない類」に比べて「いい・大丈夫類」の比率が比較的高いが、昭和コーパス・CSJ

² 昭和コーパスには「ておらない」が6例あるが、分析の統一のため除いた。

では「いい・大丈夫類」の比率は低く、「ない類」が支配的である点にある。このことから、近年の話し言葉の方が、またくだけた発話スタイルの方が「いい・大丈夫類」が使用されやすいと考えられる。また、昭和コーパスとCSJに差が見られなかったことは($z = -0.43$, n.s.), 近年の改まったスタイルの話し言葉の特徴が、約60年前の話し言葉のそれと同様であることを示している。これは、改まったスタイルが言語使用の経年変化に対して一定の耐性を持つことを示唆している。

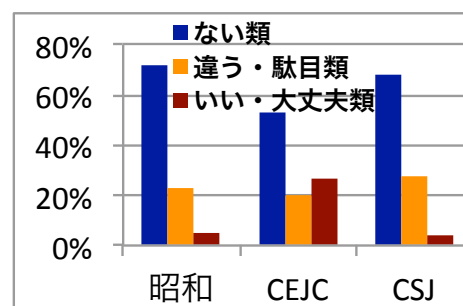


図3. 「全然」の呼応表現の分布

4.4 文末丁寧表現の揺れ

文末丁寧表現では、昭和コーパスとCEJC・CSJとの違いは認められたものの(1%水準), CEJCとCSJとの間に有意な差はなかった。

図4が示すように、昭和コーパスとCEJC・CSJとの違い($z = 8.24$, $z = 7.15$)は、昭和コーパスでは「です」が多いものの、「であります」

「でございます」も一定の比率で使用が認められる。一方、CEJC・CSJでは「です」が支配的で、それ以外の表現はほとんど見られない。このことは、以前は文末丁寧表現が複数の表現から選択されていたが、近年の話し言葉では「です」に統一されてきていること

を示している。また、CEJCとCSJに差が見られなかったことは($z = 0.95$, n.s.), 「です」への統一傾向がスタイルの影響を受けないということを示唆している。

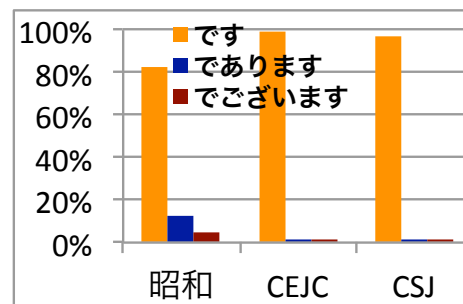


図4. 文末丁寧表現の分布

5. 考察

前節の分析結果を表3にまとめる。まず、全体的な傾向として、どの言語現象についても年代の影響を受け、経年変化が認められた。一方、スタイルについては現象ごとに違いが見られた。「凄い」の活用形、テイル形、「全然」の呼応はスタイルによる分布の違いがあるが、文末丁寧表現にはその違いはなかった。

表3. 各言語現象と経年変化・スタイルとの関係

	「凄い」の活用形	テイル形	「全然」の呼応	文末丁寧表現
年代の影響	+	+	+	+
スタイルの影響	+	+	+	-

また、経年変化・スタイルの影響の方向性について一般化すると、日常語や比較的新しい用法ほど、1) 近年のデータ(CSJ, CEJC)に、2) くだけたスタイル(CEJC)に現れやすい、反対に伝統的な表現は、3) 往年のデータ(昭和コーパス)に多く、4) 改まったスタイル(CSJ)において保たれやすいと言える。

「全然」の呼応の分布から読み取れることとして、「全然」+否定的表現という規範は、くだけたスタイルでは肯定的表現も受け入れるように変わってきているが、改まったスタイルでは維持されていた。このことから、改まったスタイルは書き言葉と同様、経年変化に対する耐性を持つと考えられる。このように、話し言葉・書き言葉の違いはもちろんのこと、話し言葉内でのスタイル差も言語の経年変化に影響を与えることが示された。

その他、用例ごとの特徴については、「テオル」は昭和コーパスでは高い頻度で使われ、様々な形式があるが、CEJCになると頻度は低く、「ております」といった定型句のような形でのみ残っていた。「ございます」についても同様で、CEJCでは「こちら～でございます」のような形だけであった。反対に、昭和コーパスで

のみ観察された「ござんす」、また「てございます」や形容詞連用形ウ音便に後続する「美味しゅうございます」など、過去に定型句のように使われていたものも、CEJC・CSJでは見られず、これらが衰退してきていると考えられる。これらから、伝統的な用法から革新的な用法へと交替する言語変化の過程で、多様性が失われ、定型句などにのみ残るといった傾向が確認できた。

6. まとめ

本研究では、話し言葉コーパス 3 種類の比較・分析により、言語使用の経年変化、及びスタイルの影響を定量的に捉えた。個々の変化の詳細を記述すると共に、共通の傾向も観察された。同時に、言語変化の分析の際には、話し言葉・書き言葉の違いだけでなく、それぞれの中でスタイルの違いを考慮することの重要性を示す結果となった。本研究では、各言語現象が示す経年変化の全体的な傾向を把握することを主眼としたため、用例別・文脈別の細分化などは行わなかった。従って、今回の結果を足掛かりに、用例ごとの分析の深化が今後の課題となる。また、話し言葉の経年変化の理解には、本研究で取り上げた 4 つの言語現象だけではなく、他の現象についても分析を進め、現象間の共通点・相違点の比較・検討が必要となるだろう。

【参考文献】

- 小木曾智信・中村壮輔 2013.『『中納言』の使い方』前川喜久雄（編）『コーパス入門』（講座 日本語コーパス 1）159-169.朝倉書店
- 蒲谷宏・川口義一・坂本恵 1998.『敬語表現』大修館書店
- 小磯花絵・居關友里子・臼田泰如・柏野 和佳子・川端良子・田中弥生・伝康晴・西川 賢哉 2017.『『日本語日常会話コーパス』の構築』『言語処理学会第 23 回年次大会発表論文集』 775-778.
- 佐野真一郎 2012.「全然の変化を分析する」, 日比谷潤子編『はじめて学ぶ社会言語学』, 227-247. ミネルヴァ書房
- 鈴木英夫 1993.「新漢語の受け入れについて―「全然」を例として―」松村明先生喜寿記念会編『国語研究』 428-449, 明治書院
- 南部智史 2007.「定量的分析に基づく「が/の」交替再考」『言語研究』 131: 115-149.
- 野田春美 2000.「「ぜんぜん」と肯定形の共起」『計量国語学』 22-5: 169-182.
- 服部匡 2009.「「～シテイル」と「～シテオル」―戦後の国会会議録における使用傾向調査―」『計量国語学』 27-1: 1-17.
- 服部匡 2010.「「全く」と「全然」の使用傾向の変遷―国会会議録のデータより―」『同志社女子大学総合文化研究所紀要』 27: 162-176.
- 堀尾佳以 2014.『若者言葉にみられる言語変化に関する研究』博士論文, 九州大学
- 前川喜久雄・龍宮隆之・小磯花絵・小椋秀樹・菊池英明 2000.『『日本語話し言葉コーパス』の設計』『音声研究』 4-2: 51-61.
- 丸山岳彦 2016.『『昭和話し言葉コーパス』の計画と展望 ―1950 年代の話し言葉研究小史―』『専修大学人文科学研究月報』 282, 39-55.
- 山口昌也 2014.「全文検索システム『ひまわり』を用いた既存言語資料の活用方法の検討」『第 6 回日本語コーパスワークショップ予稿集』, 151-156.
- Matsuda Kenjiro 1993. Dissecting analogical leveling quantitatively: A case of innovative potential form in Tokyo Japanese. *Language Variation and Change*, 5-1: 1-34.
- R Development Core Team. 1993–2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sano, Shin-ichiro. 2016. A corpus-based analysis of the shift in licensing condition of NPI in Japanese. paper presented at *Sociolinguistics Symposium 21*, at University of Murcia, Murcia, Spain.