

国会会議録コーパスを用いたオノマトペ使用の地域比較

Area comparisons of mimetic words usage using Japanese Diet Records

平田佐智子*¹
Sachiko Hirata-Mogi

中村聡史*²
Satoshi Nakamura

小松孝徳*²
Takanori Komatsu

秋田喜美*³
Kimi Akita

*¹ 東京大学
University of Tokyo

*² 明治大学
Meiji University

*³ 明治大学
Meiji University

*⁴ 大阪大学
Osaka University

Onomatopoeia is believed to be particularly frequent in some dialects of Japanese. In this study, we analyzed whether there is really a difference in the use of onomatopoeia among the eleven major areas of Japan using Japanese Diet Records. The results suggested that there is some difference among the frequencies of the use of particular onomatopoeia in each area, especially in emphatic onomatopoeia. Further research combined with dialectic, microscale perspective will be needed.

1. はじめに

「近畿圏出身者はオノマトペの使用頻度が他地域よりも高い」という素朴理論に対して、発言者の出身地域によってオノマトペ使用頻度が実際に異なるのかどうかを検討した。データとしては、過去 20 年分の国会会議録の発言内容に含まれるオノマトペと、発言者の出身地情報を用いた。その結果、オノマトペ全体としては地域差があまり見られないが、ある種類のオノマトペにおいて顕著な地域差が見られることがわかった。

2. オノマトペ使用頻度の地域差とは

2.1 オノマトペ

オノマトペは日本語語彙の重要な一部を占め、漫画や商品の宣伝などに幅広く使用されることばである。また、その使用においても文脈や状況との関係が指摘されている。例えばオノマトペが公的な立場における目上の対象にあまり用いられない半面、身内など私的な関係であれば目上・目下関係なく用いられることを示されている[平田 12b]。このように、オノマトペには使用される対象が限定されるという特徴があるが、使用者に関しても特徴がみられるという素朴理論が存在する。

2.2 オノマトペと地域差

「近畿圏出身者はオノマトペの使用頻度が他地域よりも高い」という素朴理論は、直感的に感じられるものである。また大阪弁／関西弁や大阪文化に関する研究においてもオノマトペにまつわる記述がいくつか見受けられる。(方言調査の一環として、発言に含まれるオノマトペ数を比較し順位付けを行っているが、近畿圏はオノマトペ使用がそこまで多くなく、東北や関東地方において比較的多いとする報告を行っている[三井 07])。しかし、この素朴理論に対して、実際にある特定の地域でオノマトペの使用頻度が高いのかどうかを実証的に示した例は現在のところ見当たらない。

同様の問題に対して Web 調査を行い、全国 47 都道府県に居住する 15～65 歳の男女を対象に、オノマトペを普段どの程度使用するかに関して尋ねた研究がある[平田 12a]。調査の結果、オノマトペ使用頻度を地域毎に比較した場合、有意な差は

見られなかったが、「最もオノマトペを使用していると考えられる地域」の順位を尋ねた場合、近畿圏が他の地域を大きく差をつけて 1 位となった。この結果から、「近畿圏出身者はオノマトペの使用頻度が他地域よりも高い」という素朴理論は実際の使用頻度を反映しておらず、むしろマスメディアから得られる「近畿圏出身者」のステレオタイプから波及したものであると結論づけた。しかし、この調査の問題点として、個々人の主観的な使用頻度を報告させたため、実際の使用頻度を必ずしも正確に反映していない可能性がある点や、そもそも Web 調査の回答者が「オノマトペ」を正確に理解していたかどうか不明な点が挙げられる。

2.3 本研究の目的

本研究では、先行研究[平田 12a]において問題点として挙げられた「オノマトペ使用頻度の主観性」を解決するため、実際に行われた会話データベースからオノマトペを抽出し、発言者毎のオノマトペ使用頻度を算出した。このようにして得られた客観的な使用頻度と発言者の出身地情報を合わせることで、オノマトペ使用頻度の地域差についてより詳細に検討することを目的とする。

3. 方法

3.1 国会会議録データベース

本研究では、話し言葉のデータベースの中でも数十年単位でのデータ蓄積がなされている国会議事録データベースを分析対象として選択した。

国会議事録データベースは「国会会議録検索システム (<http://kokkai.ndl.go.jp/>)」として Web 上で一般公開されており、1947 年の第 1 回国会から最新の議事録まで検索することができる。その範囲は衆議院・参議院本会議から両院協議会や合同審査会にまで及んでいる。本研究では 1992 年から 2012 年までの衆議院・参議院本会議の議事録に限定し、各発言者のオノマトペ使用頻度と算出した。

3.2 分析

国会会議録のデータは、基本的に会議の日(和暦)、会議の参加者、発話した人の名前、その発話自体からなる。また、発話者や発話内容は上から下に時系列に沿って並んでいる。そこで、まず国会会議録の各会議録から日時情報を抽出し西暦に変換する。また、発話者および発話者の役職、発話内容を抽出し、発話内容を 1 文ずつに分解することで、発話 ID(発話を一意に識別するもの)、発話日、発話者名、発話者の役職、発

話文からなるデータベースを構築する。ここで、発言内容を1文ずつに分解するため、句点または国会会議録作成者により改行されている部分、話者が変わるまでの部分を1文として処理する。データベースに格納された発話文数は48,512,596文で、発話文字数は3,437,337,516文字となった。

上記のデータベースに対して、分析のためのオノマトペリストから検索語を1つ取得し、検索語を含む発話文をすべて取得する。また、全ての発話文に対応する話者情報を付与し、さらに話者情報が同一である発話文に関しては同じ話者の発言と見なし、各話者の総発言数・オノマトペの含まれる発言数を算出した。データベースに対してこの処理を、すべての検索語について行った。

3.3 対象としたオノマトペ

検索対象とするオノマトペとして、先行研究[Kakehi 96]をもとに作成したオノマトペリストを使用した。なお、検索の際はオノマトペ以外の文字列が抽出されるのを防ぐ目的で、オノマトペ+「と」(例:オノマトペ「とんとん」は「とんとんと」の形で検索する)という形式を取った。

3.4 発言者

国会会議録に含まれる発言者から、1000件以上発言している発言者を抽出した。そして、抽出された各発言者の出身地(あるいは出生地)をWeb上のリソース(Wikipediaや本人の公式サイトなど)を用いて調査した。その結果、国会会議録中で1000件以上発話している5803人中3606人(62.12%)の出身地が判明した。出身地が判明し、かつ出身地が国内であった話者のみを分析対象とした。

4. 結果

出身地が判明している発言者を先行研究[平田 12a]で用いた国内11エリアを基準に分類した。その結果、表1のような分布となった。

出身エリア	人数
北海道	69
東北	155
関東	492
甲信越	99
北陸	64
東海	155
関西	262
中国	120
四国	73
九州	189
沖縄	20

表1 各エリアを出身地とする発言者数

4.1 結果1(リスト全体を対象とした場合)

図1は、オノマトペリスト全体を検索対象とした場合の各エリアの頻度を色によって表した図である。エリアの色が赤に近いほどオノマトペが高頻度で使用され、青に近いほど低頻度で使用されていることを示す。描画にあたり、各県のオノマトペ出現数/各県の総発言数、各発言者におけるオノマトペ出現数/総発言数の各県平均を求め、偏りを抑えるためこれらの数値を最大値で正規化し、足し合わせた値を求め、得られた値に基づき描画を行っている。

各エリアの色がほぼ類似しているため、リストに含まれるオノマトペ全てを検索対象とした場合、地域によるオノマトペ使用の偏りは観察できなかったといえる。そこで、結果2ではリストに含

まれるオノマトペの中から、特徴を持ったオノマトペに絞り、探索的に分析を行った。

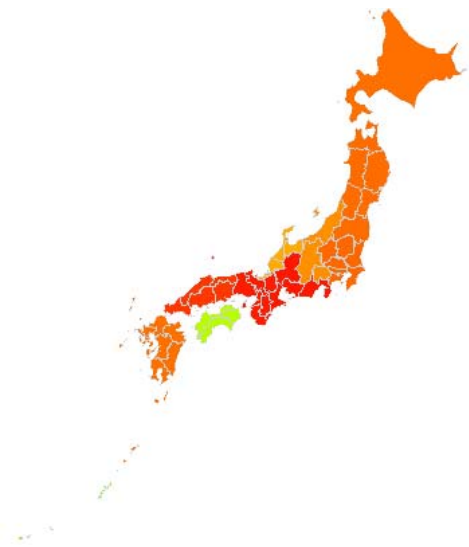


図1 各エリアのオノマトペ全体の使用頻度

4.2 結果2(特定のオノマトペ/それ以外)

図2及び図3は、それぞれ強調系オノマトペと、比較対象としてオノマトペではない副詞を対象とした場合のヒートマップを示している。



図2 強意系オノマトペの使用頻度

強意系オノマトペとは CVCCVri(Cは子音、Vは母音を表す)の形を持つ「しっかり」「はっきり」「じっくり」などのオノマトペである。この系統のオノマトペに限って分析を行うと、関西エリアが最も頻度が高い結果となった。

また、動詞を修飾する言葉であれば地域によって使用頻度の偏りが出るかどうかを確認するため、「つい」「また」「十分」などのオノマトペではない副詞を対象に分析を行った。図3はこれら副詞の使用頻度を表しており、各エリアにおいて類似した色合いとなっている。この結果から、副詞はオノマトペのように各エリアにおける使用頻度の偏りは見られないことがわかった。

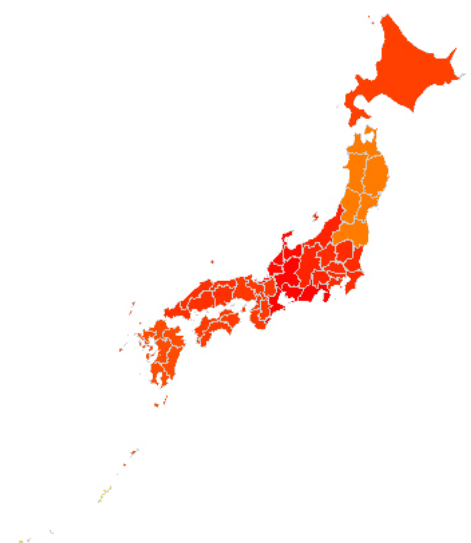


図3 各エリアにおける副詞の使用頻度

5. 考察

5.1 本研究よりわかったこと

本研究は、先行研究[平田 12a]を受けて、客観的なオノマトペ使用頻度を元に、オノマトペ使用の地域差が存在するかどうかを検討することを目的とした。国会会議録データベースを用いて行った分析により、オノマトペ全体として使用頻度を分析した場合は地域毎の偏りは見られないが、ある種類のオノマトペに関しては明確な使用頻度の地域差が見られることがわかった。また、非オノマトペ副詞の使用頻度の分布と比較すると、オノマトペの使用頻度は地域によって大きく偏りが生じることがわかった。

オノマトペ全体では頻度の偏りが見られないのに対し、ある種類(本研究における強意系オノマトペ)において地域差が見られたことに関しては、テレビ番組の分析を行った研究[田原 01]の中で、関西では「ガーッ」「パーッ」などのオノマトペが多用されるという結果が得られたことと一致する。つまり、「関西ではオノマトペが多用される」という認識に含まれるオノマトペは、ある種類のオノマトペのみが含まれる、という可能性である。この可能性については、関西ローカル番組などから得られる会話データを分析することで、詳細が明らかになると考えられる。

5.2 展望

本研究にはいくつかの問題点が存在する。一つ目は、本研究で対象とした国会会議録に含まれる発言者の出身地の全てが判明しているわけではない点である。また、様々なリソースを用いたため、出身地の定義をこちらが指定することができず、出身地と出生地が混在している場合も存在する。さらに、リソースの情報が合っていない、該当する地域の言語アイデンティティを持っているかどうかは定かではない。比較的古い年代の発言者は、地方から大学進学と同時に上京するパターンが多く見受けられたが、現代に近づくにつれて幼少期から国外を含め各地を転々とする傾向が見られ始めるため、言語形成期における厳密な居住地を個人に対して突き止めるのは非常に難しい。

これらの問題を避けるため、方言学ではあくまでマイクロ視点に徹し、少ないサンプルによる記述を主に行うが、出身地と言語の

問題に対し、あえて大規模データを用いるマクロ視点を取った本研究は非常に挑戦的であったといえる。今後マイクロ視点による研究と組み合わせることで、マクロ視点による研究の利点や問題点の解決に繋がることを期待する。

参考文献

- [Kakehi 96] Kakehi, H., Tamori, I., and Schourup, L.: *Dictionary of Iconic Expressions in Japanese*, Berlin/New York: Mouton de Gruyter (1996)
- [三井 07] 三井 はるみ, 井上 文子: 方言データベースの利用, 小林隆(編), 方言学の技報, pp.29-89, 岩波書店, (2007)
- [平田 12a] 平田佐智子, 秋田喜美, 小松孝徳, 中村聡史, 藤井弘樹, 澤井大樹: オノマトペに対する意識の地域比較論文, 第26回人工知能学会論文集, (2012)
- [平田 12b] 平田 佐智子, 日常会話におけるオノマトペ使用に関する調査, 日本心理学会第76回大会論文集, (2012)
- [田原 01] 田原 広司: ピヤッとちぎってシャッと渡す—関西弁のオノマトペ—. 月刊言語, Vol. 30, pp. 24-25 (2001)

謝辞

本研究は、科学研究費補助金: 学術研究助成基金助成金(若手研究(A)) (課題番号: 23680006) (中村) の補助を受けて行われた。