

間投助詞「さ」と話題の関係

中俣尚己(京都教育大学)

1. はじめに

1.1 本研究の目的

本発表では筆者が構築した『日本語話題別会話コーパス：J-TOCC』を利用し、間投助詞「さ」の使用が話題の影響を大きく受けていることを示す。さらにコーパスから得られる様々な情報を元に間投助詞「さ」の機能について考察する。

1.2 『日本語話題別会話コーパス：J-TOCC』

『日本語話題別会話コーパス：J-TOCC』(中俣, 2021a; 中俣・太田ほか, 2021) 発表者らが2021年に公開した話題を統制した会話のコーパスである。J-TOCCは、日本語教育において話題を中心とした学習活動や教材開発を支援するため構築された。最大の特徴は同じ会話協力者のペアが15の話題について、5分間ずつ会話を行っていることであり、他の条件を統制したうえで話題が語彙・文法・談話ストラテジーなどに与える影響を検討することに特化したコーパスである。なお、J-TOCCの名称はJapanese Topic-Oriented Conversation Corpusの略である。「ジェイトック」と読む。

調査協力者は様々な話題について自然に話しやすい20代大学生の親しい友人同士の関係に限定した。大学院生は研究のテーマが万一含まれると容易に個人を特定できるため、全員が学部生である。関東地区と関西地区でバランスをとり、さらにそれぞれ「男女」「男男」「女女」の3グループを同じペア数だけ収録した。1グループあたり、20ペア(40名)が参加しており、話者の重なりはない。1話題あたりの延べ時間は10時間、語数はおおよそ11万語であり、話題間の差はない。全体では120ペア、240名となる。録音は2018年から2019年にかけて行われた。

話題の選定基準は詳しくは中俣(2021a)に譲るが「身の回りの話題」として「01 食べること」「02 ファッション」「03 旅行」「04 スポーツ」「05 マンガ・ゲーム」「06 家事」「07 学校」「08 スマートフォン」「09 アルバイト」「10 動物」「11 天気」「社会にかかわる内容も含む話題」として「12 夢・将来設計」「13 マナー」「14 住環境」「15 日本の未来」を設定した。全ての話題について5分ずつの会話を収録している。

コーパスは発表者のウェブサイトからダウンロードできる。文字化したテキストデータと、調査協力者の情報に関するフェイスシートがある。フェイスシートの内容は、録音地・性別・言語形成地(6~12歳の間に居住していた都道府県、複数回答あり)・話題親密度である。話題親密度とは、「それぞれの話題についてどれだけ詳しいか、あるいはどれだけ自信を持って話すことができたか」という概念であり5段階で評価してもらった。

1.3 話題と機能語

J-TOCCは話題と語の関係を調査し、日本語教育のタスクや教材作りに活かすことを目的としている。そしてここでの語は名詞や動詞といった実質語に限らない。日本語教育で「文法項目」と呼ばれることもある機能語もまた話題によって大きく出現数が変わることがわかっている。J-TOCCの解説論文である中俣・太田ほか(2021a)ではアスペクト形式の「てる」は話題によって大きく出現数が異なり、「05 マンガ・ゲーム」や「08 スマートフォン」に多いことを報告している。そして、この結果は他のコーパスを利用した調査(Nakamata2019; 中俣・小口ほか2021)とも一致し、進行のアスペクトマーカー「てる」は映画、ドラマ、動画などを見る話題と非常に相性が良いと結論づけられる。このように、一見話題とは無関係とも見える機能語と話題の関係を明らかにすることも、本プロジェクトの目的の一つである。

中俣(2021b)ではどのような機能語が話題の影響を受けているかを探索的に調査するために、15の話題別コーパスについて全ての語の対数尤度比(LLR)を計算した。その結果、格助詞に加えて間投助詞の「さ」が「日本の未来」に非常に多く出現するという結果が得られた。これは調査実施時の印象としても、この話題になった途端に「さ」が増加したと感じられるほどの変化である。そこで、以下2つのRQについて調査を行った。

RQ1: 「さ」の使用数の増減に影響を与える要因は何か? またその影響力はどれほどか?

RQ2: 会話の中で「さ」はどのように使われることが多いのか?

2. 間投助詞「さ」の先行研究

2.1 間投助詞と終助詞の区別

間投助詞「さ」についてはあまり多くのことがわかっているわけではない。また、森田(2007)など終助詞「さ」との間に境界を引かないという立場も目立つ。一方で、大江(2017)は統語的地位、任意性、相互作用性、人物像などの観点から終助詞と間投助詞を区別する必要性を論じている。発表者は間投助詞「さ」を終助詞「さ」と区別する必要があるとの立場に立つ。森田(2007)の主張は自然談話では「文末」を定義することが難しいという事実がその論拠となっている。この点については正しいが、「さ」がどのような環境に出現するかという点に注目した場合、大江(2017)も示している通り、顕著な違いが存在する。今回の調査で収集したデータから結論を先取りして述べると自然会話における「さ」は以下の2つの特性を持つ。これは「ね」とは大きく異なる。

- (1) 述語句に接続する場合、そのほとんどが接続助詞の後に接続する。動詞のル形、タ形、テイル形など主節相当部分には接続しない。働きかけや疑問のマーカ―の後も接続しない。
- (2) 感動詞にも接続するが、「あの」や「なんか」などフィルター系の感動詞に限られる。応答詞や挨拶には接続しない。これらは「文を区切る」働きのある語には「さ」は接続しないと言うこともできる。ただし、前接形式ではなく、あくまでも「区切る」という意識による問題であることが以下の(3)と(4)の比較からわかる。
- (3) やっぱり { ϕ /さ/ね}, 時代の流れだからさ。(E-112-2M: うん) 昔, 20年前, 30年前に比べて, 平均気温1度, 2度上がるみたいな感じじゃん. やっぱり。(東日本・男男・天気)
- (4) W-119-2M: 甘い好き? W-119-1M: 甘い大好き. 甘い大好きよ. W-119-2M: やっぱり { ϕ /*さ/ね}。(西日本・男男・食)

また、接続助詞を伴わずに述語句に接続すると、「男性」という人物像が想像される。一方で今回のデータでは「さ」は女性の方が使用が多いことがわかった。この点からも終助詞「さ」と間投助詞「さ」は異なる存在であると言える。

- (5) 明日ってサ, 雨降るかな?—明日はきっと晴れるサ。(大江, 2017:97)

2.2 間投助詞「さ」の機能

「さ」の機能については野田(2002:276)では「だんていず, 当然のこととして, あるいはとりあえずのこととして提示することを表す」, 富樫(2011)では「計算終了の標示」としている。また、森田(2007)は「ね」や「さ」の役割について、「何かしらの区切りをつけ, 「反応の機会の場所」(西阪, 2005)を明示する」としている。

3. 一般化線形混合モデル(GLMM)を用いた「さ」の使用頻度に与える要因の検討

3.1 調査の概要

ここでは、「RQ1: 「さ」の使用数の増減に影響を与える要因は何か? またその影響力はどれほどか?」を調べるため、一般化線形混合モデル(以下 GLMM)を用いて J-TOCC のデータを分析した結果を報告する。対数尤度比から言えることはあくまでもある話題に有意に多いという一点のみで、影響力の大きさまではわからない。話題以外の要因の影響も不明であり、また全話者の数値を合算して計算しているため、個人差が無視されている。そこで、統計モデルを利用し、「話題」「地方」「性別」「話題親密度」の説明変数がどのように「さ」の頻度に影響を与えているのかを計算した。具体的には J-TOCC のうち最も「さ」の使用数が少なかった「旅行」のデータと、最も使用数が多かった「日本の未来」のデータを使用した。正規表現で「さ」の後に句読点に来る例を集め、話者ごとに集計した。「さー」「さあ」なども含める。

3.2 ポアソン分布と一般化線形混合モデル(GLMM)

ポアソン分布は二項分布の極限をとった形であり、「単位時間に一定の確率で生起する事象が何回生起するか」という回数の確率分布はポワソン分布に従うとされる。1日に受け取るメールの件数や、1時間にある道路を通る車の台数はポワソン分布に従う。間投助詞「さ」も談話を巨視的に捉えれば、任意の場所にランダムに生起していると考えられる。生起しない環境は極めて稀であるからだ。ここでは、「さ」の出現はポアソン分布に従うとする。ポアソン分布をグラフで書くと様々な山型のグラフになるのだが、その形は λ という一つのパラメータによって決まり、この値を推定することが統計モデルを作ることの目標となる。また、この λ は平均回数に等しい。

しかし、現実には得られたデータは理想的なポアソン分布とは合致しない(久保, 2012)。本研究のデータもまた然りである。これは現実には「個人差」が存在し、結果、「過分散」という現象が生じるためである。

理想的なポアソン分布では平均と分散は等しくなるはずであるが、本データの平均は4.3、分散は26.2である。間投助詞「さ」の使用頻度について個人差があると仮定することは妥当であろう。

そこで、個体差を含んだデータを分析できるGLMMを実施した。これは説明変数の効果はポアソン分布に従い、それ以外の個体差の効果が正規分布に従うと仮定し、二つの確率分布を混合して分布を作りパラメータを推定する手法である。計算にはR(3.6.3)とglmmMLパッケージを使用した。なお、作られたモデルの「当てはまり良さ」はAICという指標で計算でき、小さいほど良い。個体差を考慮しないGLMではAICが3367だったのに対し、GLMMを使うと1295となり劇的に改善される。

3.3 結果と考察

統計モデルでは利用する説明変数の数を変えて最適なモデルを検討するが、今回は「話題」「地方」「性別」「話題親密度」の4つの変数を全て組み込んだモデルが最もAICが小さくなった。その結果を簡易的に表示する。

表1 GLMMの結果(抜粋)

	切片	話題が旅行なら	場所が西日本なら	性別が男性なら	話題親密度×	
$\lambda = \exp($	1.48074	-0.74438	-0.42379	-0.25860	+0.07045)

λ の計算は、条件に合わせて足していく。例えば話題が「日本の未来」であればそのセルは計算しなくてよい。仮に色のついたセルの和を α とすると、 $\exp(\alpha) = e^\alpha$ (e はネイピア数。およそ2.718)を計算すればその値が「さ」の平均使用回数 λ である。試みに「日本の未来・東日本・女性・親密度2」について計算すると $\lambda = \exp(1.48074 + (2 \times 0.07045))$ となり、5.06となる。この条件にあうデータの実際の平均値は5.7であるため、おおむねうまく推定できている。

表1の結果から言えることは以下の通りである。①最も影響を与えるのは話題であり、難しい話題であると「さ」の使用が増える。②東日本のほうが「さ」をよく使用する。③女性のほうが「さ」をよく使用する。④話題に詳しい人間ほど「さ」をよく使用する。難しい話題でかつその話題に詳しいほど「さ」をよく使用するというのは重要な性質である。

4. 「さ」の前接語にからみた「さ」の機能の検討

ここでは、「RQ2: 会話の中で「さ」はどのように使われることが多いのか?」を調べるため、全ての「さ」の例を目視し、前接語の成分を調査した。結果を表2に示す。話題間により成分の違いは見られなかった。つまり、「日本の未来」の話題では「さ」は全体として多く使われるのであって、特定の用法が増えるわけではないということである。

表2 「さ」の前接成分の集計(N=2,081)

成分	実数	割合	顕著なパターン
名詞句	676	32.5%	それさ、これさ、とかさ、とかがさ、のさ、ってさ
述語句	828	39.8%	各種接続助詞、ていうかさ、みたいなさ
副詞句	230	11.1%	もうさ、結構さ
接続詞	190	9.1%	でもさ、だってさ
感動詞	152	7.3%	なんかさ、そのさ

以下、成分ごとに見ていく。名詞句では「さ」以外の助詞が付加されている例が多かったが、特に多かったのは「とか」である。例としては(5)格助詞を伴うもの、(6)格助詞を伴わず格関係を持つもの、(7)他の名詞と並列関係にあるもの、(8)格関係を認めにくく、主題相当句と考えられるものなどがある。

(5) 育児休暇とかがさ、何かと取りにくいとかさ。(西日本・女女・日本の未来)

(6) 何か男の人好きじゃん、(E-315-2F: そうだね) 列車とかさ、, 何か男のロマンじゃないけど、結構楽しいらしくて、(東日本・女女・旅行)

(7) でも実際さ、おじさんととかさ、お婆さんとかで結婚してない人いるからさ、(東日本・女女・日本の未来)

(8) 海外って、その、インドネシアとかってさ、結構そのトイレの文化とかが違うからさ。(西日本・男男・旅行)

「とかーさ」をコロケーションと見立て、コロケーションの強度の指標であるMI-Score計算すると2.98であった。これは一般的に意味のあるコロケーションの基準3にわずかに届かないが、「がーさ」1.99、「もーさ」1.70、「はーさ」1.61、と比べると結びつきが強い。

最も多かったのは述語句でその大半が従属節である。主節にはほぼ出現しない。また、以下のような連体節または引用節相当のパターンも見られた。

- (9) 何か、嫌じゃない、俺らの世代だけ、何かこの、暗黒時代みたいなさ。W-112-1M：●。ばり言われてんもんな。
(西日本・男男・日本の未来)
- (10) 何かその、そんな重要というか、何ていうの、流れ、流れ次第というかさ、例えば周りが3人産んでたら3人産むし(東日本・男女・日本の未来)

接続詞では「でもさ」と「だってさ」が多い。

- (11) E-206-2M：えー。少子化とかさ、っつか、ぶっちゃけ実感ないからさ、あんまり分かん、何か、高齢化は(E-206-1F：確かに)分かるけど、少子化ってさ、身近じゃなさ過ぎて。
E-206-1F：でもさ、うちの中学校さ、2クラスしかないんだよ。
- (12) W-217-1F：え、楽しくないって。もう絶対元気なうちに死にたい。
W-217-2M：いや、俺はもう生にしがみついでいく。
W-217-1F：だってさ、周りがみんなおらんくなってさ、死んでくやん。さみしいやん。

特に「だからーさ」のMI-Scoreは0.91であるのに対し、「だってーさ」は3.62であり、「だって」の後に「さ」が出現しやすくなっている。「でもーさ」も2.77とやや高かった。「だって」と「でも」に共通するのは直前の相手の意見に同調していないという点である。一方、「さ」を伴わない「だから」は相手の意見に同調する場面も多く見られた。

副詞は「もうさ」「結構さ」が多いものの、他にも様々なパターンが見られ、大きな偏りがあるとは言えない。「もうさ」はフィラーのようにも使われている。

感動詞は最も大きな偏りを見せた。全体の73.7%が「何かさ」、10.5%が「そのさ」であった。「何か」は話したい内容が頭の中にあり、言葉を選んでいる時に発せられるフィラーである。

5. おわりに

RQ1については、性差・地域差よりも話題が「さ」の使用頻度に大きな影響を与えていることが明らかになった。難しい話題、そしてその話題について詳しい人ほど「さ」を多用している。文体や相手との関係が変わらなくても、話す内容が変わるだけで間投助詞の使用が変化することが示された。また、「さ」が出現する環境には話題による違いはないが、共起しやすい語として「とか」「何か」「でも」「だって」などがある。2つの結果を合わせると「さ」は話し手はまだ他にも話す内容が残っており、言葉を選んでいっているという状況で使われやすいと言える。相手の意見に同調しない場面で使用し、断定を避けるという態度も確認できた。

謝辞 本研究の遂行にはJSPS18H00676の助成を受けた。また、GLMMについて国立国語研究所の岩崎拓也氏から有益なコメントを頂いた。ここに感謝する。

参考文献

- 久保啓弥(2012). データ解析のための統計モデリング入門 岩波書店
- 森田笑(2007). 終助詞・間投助詞の区別は必要か 言語, 36(3), 44-52.
- Nakamata, Naoki (2019). Vocabulary Depends on Topic, and So Does Grammar *Journal of Japanese Linguistics*, 35(2), 213-234,
- 中俣尚己(2021a). 日本語話題別会話コーパス：J-TOCC 解説資料
http://nakamata.info/database/j-tocc_document.pdf
- 中俣尚己(2021b). 会話における助詞の出現頻度に話題が与える影響 日本語学会 2021年度春季大会予稿集.
- 中俣尚己・小口悠紀子・小西円・建石始・堀内仁(2021) 自然会話コーパスを基にした『話題別日本語語彙表』, 計量国語学, 33(3), 194-204.
- 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣(2021). 「日本語話題別会話コーパス：J-TOCC」 計量国語学, 33(1), 205-213.
- 西阪仰(2005). 分散する文一相互行為としての文法 言語, 34(4), 40-47.
- 野田春美(2002). 終助詞の機能 宮崎和人・安達太郎・野田春美・高梨信乃 モダリティ くろしお出版 pp. 261-288.
- 大江元貴(2017). 間投助詞の位置づけの再検討：終助詞との比較を通して 語用論研究, 19, 90-99.
- 富樫純一(2011). 終助詞「さ」の本質的意味と用法 日本文学研究, 50, 150-138.