

# 話し言葉における助詞の出現頻度に対する話題の影響

なかまた なおき  
中俣 尚己 (京都教育大学)

## 1. はじめに

発表者は日本語教育における教材や教室活動の開発に資するために、話題が語彙・文法・談話ストラテジーなどに与える影響の研究を行っている。近年の日本語教育は文型積み上げ式のシラバスから、課題遂行能力を重視するシラバスへの移行が進んでいる。しかし、シラバスの中で様々な言語活動をまとめる際に「場面」や「話題」で複数の活動をまとめて提示することが多い。この時に、「話題」と具体的な言語行動の関係がわかっていると効果的な学習活動が行えると考え。また、従来型の文型積み上げ型シラバスを採用するにしても、会話には必ず「話題」が存在するため、文型と話題の関係は日本語教師の関心事である。

実質語の語彙が話題に従属するという考えはわかりやすく、山内(編)(2013)のような実用的な語彙表も作られている。一方で、いわゆる文法項目と呼ばれる機能語については話題に従属するようにも見えず、山内(編)(2013)でも話題に従属しないとしている。これに対して、Nakamata(2019)は、小規模な接触場面会話コーパスを用いた調査を行い文法形式にも話題の影響が現れることを明らかにした。例えばテンス・アスペクト形式は「ポップ・カルチャー」に多く使用される。ただし、この研究で利用されたデータは話題の統制がされていたわけではない。また、接触場面のデータであることから、母語話者どうしの会話でも話題により同様の偏りが見られるかは未検証であった。

そこで、新たに母語話者どうしの話題別会話コーパスを構築し、その偏りを検証した。結論から述べると、一部の機能語は話題により大きく出現頻度が変化することが確認できた。本発表ではその中でも、一見話題とは無関係に使用されているように見える助詞に議論を絞り、話題が会話中に出現する助詞に与える影響について報告する。

## 2. 資料「日本語話題別会話コーパス：J-TOCC」について

### 2.1 概要

「日本語話題別会話コーパス：J-TOCC」は、120組240名の調査協力者が、15の指定された話題につき5分ずつ会話をしたものを録音、文字化したコーパスであり、発表者のウェブサイトで公開されている(<http://nakamata.info/database/>)。J-TOCCの名称はJapanese Topic-Oriented Conversation Corpusの略で「ジェイトック」と読む。その最大の特徴は、話題以外の情報が統制されているため、純粋に話題の影響を調べるのに特化していることである。

## 2.2 調査協力者

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。録音は2018年から2019年にかけて行われた。

## 2.3 話題の選定

話題選定の基準としては前述の山内(編)(2013)に掲載されている100話題を出発点とし、コーパスサイズや教材としてのまとまりから15前後を目標にすることにした。選定の基準は、「a.大学生にとって身近である。」「b.初級日本語学習者むけとしてふさわしい。」「c.プライバシー上問題となる情報が多く出てこない。」「d.他の話題と近すぎない。」などを勘案し、100話題から消去法的に選び出した。最終的には表1の15話題を選定した。

表1 J-T0CCの15話題(中俣(印刷中))

番号	話題	参加者への指示内容	親密	具体
身の周りの話題				
01	食べること	例えば：好きな料理、外食。 ※料理を作る話は除く	I	A
02	ファッション	<指示内容なし>	II	B
03	旅行	例えば：行きたい場所、行ったことがある場所	I	B
04	スポーツ	例えば：運動の経験、スポーツ観戦	II	C
05	マンガ・ゲーム	アニメ、ケータイゲームを含む	II	B
06	家事	例えば：料理、洗濯、掃除	II	B
07	学校	小学校、中学校、高校時代の思い出	I	B
08	スマートフォン	例えば：機種、アプリ、SNS ※ゲームは除く	I	A
09	アルバイト	アルバイト経験、やってみたいアルバイト	II	A
10	動物	例えば：好きな動物、ペット	II	A
11	天気	例えば：最近の天気、温暖化	I	D
社会にかかわる内容も含む話題				
12	夢・将来設計	例えば：就職・結婚・家庭	III	D
13	マナー	公共交通機関でのマナーについて	II	C
14	住環境	都会がいいか、地方がいいか	III	A
15	日本の未来	少子化・高齢化をどう考えるか	III	D

## 2. 4 語数について

以下、表 2 に各話題の語数などの情報を掲載する。語数については中・長単位解析器 Comainu (小澤ほか 2014) という解析ツールで計量した。Comainu は国語研の「中納言」などで使われている「長単位」に相当する単位で語を分割するツールであり、「ている」などの日本語教育で用いられる複合辞や「勉強する」などサ変動詞が一つの単位として切り出される。

さらに、本発表に関係する情報として、動詞・名詞・形容詞(形状詞含む)・助詞の比率をそれぞれパーセントで示す。また、散らばりの度合いを変動係数(標準偏差÷平均)で示す。マクロな指標では話題ごとの 15 のサブコーパスは非常に均一的である。

表 2: J-TOCC の話題ごとの語数 (中俣(印刷中)を修正)

話題	延べ語数 異なり							
	延べ語数	(記号除)	語数	TTR	動	名	形	助
01. 食べること	145,361	108,022	6,338	0.0587	11.5	16.8	7.1	30.7
02. ファッション	148,909	112,627	6,111	0.0543	11.8	15.4	6.4	31.8
03. 旅行	147,386	110,123	6,660	0.0605	11.9	17.3	5.8	30.7
04. スポーツ	148,722	112,482	6,810	0.0605	12.0	16.4	5.6	31.5
05. マンガ・ゲーム	151,454	112,535	7,106	0.0631	12.6	16.0	4.8	31.0
06. 家事	147,950	111,915	6,141	0.0549	13.8	15.0	5.6	31.9
07. 学校	145,198	110,321	6,794	0.0616	11.1	17.4	5.6	30.7
08. スマートフォン	144,745	109,175	6,145	0.0563	12.5	14.6	5.4	32.0
09. アルバイト	147,120	111,168	6,791	0.0611	12.0	15.8	5.9	31.3
10. 動物	148,636	111,851	6,455	0.0577	11.8	16.0	6.8	31.2
11. 天気	146,245	110,713	6,242	0.0564	11.3	16.1	7.0	31.8
12. 夢・将来設計	141,101	107,460	6,034	0.0562	12.5	16.9	5.7	32.7
13. マナー	148,209	114,035	5,586	0.0490	13.3	14.4	5.5	33.2
14. 住環境	141,934	107,907	5,578	0.0517	10.5	16.8	6.7	33.9
15. 日本の未来	136,718	105,604	5,950	0.0563	12.2	15.5	5.2	33.8
合計(割合は平均値)	2,189,688	1,655,938	42,756	0.0258	12.1	15.9	5.9	31.9
変動係数	.024	.021	.069	.065	.077	.057	.111	.032

## 3. 研究方法

本研究は話題による単語の頻度の違いを明らかにするものであるが、そのためにただ頻度を比較するのではなく、「ある話題では多いが、他の話題では少ない」という特徴度に注目する。特徴度の指標として、話題ごとのサブコーパスのうち 1 つを当該コーパス、それ以外の 14 サブコーパスを合わせたものを参照コーパスとし、すべての語について対数尤度比(LLR)を計算した。これを 15 回繰り返し、各話題サブコーパスごとに LLR を算出した。また、変動係数も利用する。本発表では単独形式の助詞のみを対象とする。

## 4. 結果

### 4.1 格助詞

対数尤度比の一覧表を見ると、格助詞は全体的な傾向として表の右側、「社会にかかわる内容も含む話題」において特徴度が高くなる傾向があるようである。日常的な話題では「学校」で「の」と「から」が高いことが目立つ程度である。

それぞれの格助詞について、出現頻度と、LLR が高い話題（括弧内は LLR）、出現頻度の変動係数を示したものが表 3 である。スペースの都合で頻度 1,000 以下のものは省略した。

表 3 格助詞の結果

格助詞	頻度	特徴度の高い話題	変動係数
が	28,235	住環境(359), 日本の未来(205)	.145
を	8,035	日本の未来(204), 夢・将来設計(68)	.269
の	29,910	学校(125), 日本の未来(39)	.094
から	3,903	学校(89)	.220
に	27,706	住環境(82)	.105
と	32,992	住環境(47)	.083
で	20,063	マナー(30), 旅行(28), スポーツ(24)	.109

「が」が「住環境」「日本の未来」、「を」が「日本の未来」「夢・将来設計」に強い偏りを示したことが特徴的である。このうち「住環境」は「都会がいいか、地方がいいか」という比較の課題であったため、構文的に「が」の出現頻度が高まったとともに、「を」の出現が抑制(-120)されたと考えられる。表 1 でも動詞率が平均と比べて 1.5%ほど低く、この点が影響していると考えられる。一方で、これらの話題の共通点は「今、ここ」に関する話題ではないことであり、これが「が」や「を」の出現に影響を与えていると考えられる。

丸山(2015)による『日本語話し言葉コーパス』の調査では話し言葉ではどの格助詞も模擬講演より学会講演のほうが多いとされている。これはスタイルの違いであるが、同じ人間どうし話をした場合でも、話題によってスタイルは変わることを示唆している。

また、発表者は『名大会話コーパス』を目視で話題に分割し（中俣 2020）、今回と同様の手法で『日本語話題別語彙表』を作成・公開しているが、そちらのデータでは「が」については今回と同様の結果にはならなかったが、「を」については「政治」「宗教・風習」「人生・生き方」「調査・研究」などやはり硬い内容に特徴的であった。

格助詞の使用には方言差も考えられるが、J-TOCC の出現頻度を見る限り、東西差よりも話題による差のほうが影響が大きい。また、話者に注目すると、「が」はどの話題でも 99% の話者が最低 1 回は「が」を使用するのに対し、「を」は「食」の話題では 67.5%しか使用されていなかった。「を」のほうが話題の影響を強く受けていると言える。

これ以外では「の」「から」が「学校」の話題に特徴的。この話題は過去の話題であるため、「た」の特徴度も高い。「の」の特徴度が高くなった要因は不明であるが、「から」については「中学校から」「10 歳の頃から」のような時間表現で使われたためと言える。

## 4.2 接続助詞・準体助詞

接続助詞と準体助詞の結果を表4に示す。

表4 接続助詞・準体助詞の結果

助詞	頻度	特徴度の高い話題	変動係数
て	31,946	学校(130), 家事(80), マナー(70), 動物(43)	.149
し	6,251	住環境(113)	.191
の	18,656	マナー(99), 日本の未来(30)	.102
と	3,485	住環境(36)	.198
から	14,206	家事(26)	.092
けれど	16,872	マナー(20)	.058
ば	1,672	夢・将来設計(15)	.191

接続助詞では、並列や継起を表す「て」「し」の偏りが大きい。「学校」は過去に起きたことを時系列順に話すこと、「家事」は動作の連続を話すためであろう。また、「し」は形容詞や状態動詞との相性が良い。このことが、住む場所を比較・評価する話題の「住環境」での出現頻度の高さにつながったと言える。反対に、「て」は「住環境」では抑制されている(-156)。因果・条件は並列に比べると話題の影響を受けない。

## 4.3 係助詞・副助詞

係助詞と副助詞の結果を表5に示す。

表5 係助詞・副助詞の結果

助詞	頻度	特徴度の高い話題	変動係数
か	49,416	マナー(77), 動物(43)	.065
も	21,500	住環境(75)	.084
は	31,764	住環境(66), 夢・将来設計(49), 家事(35)	.080
って	25,594	日本の未来 (36), マナー(44), 夢・将来設計(40)	.096
たり	2,015	家事(41), マナー(38)	.284
くらい	4,318	天気(37)	.162
だけ	2,681	家事(24), アルバイト(24)	.244
しか	1,451	学校(23)	.301
まで	2,199	マンガ・ゲーム(17)	.251
なんか	1,133	学校(9) *特徴語とは言えない	.183
や	2,274	日本の未来(2) *特徴語とは言えない	.094

全体として LLR は高くない。「か」は準体「の」と同様に「マナー」に多い。「も」は「し」同様に住環境に多い。「って」は「社会に関わる話題」に多いという特徴が見られる。「たり」は並列であり、「家事」「マナー」は「て」と共通する。「くらい」は数量との共起が多く、結果「天気」に多くなった。

#### 4. 4 終助詞

終助詞の結果を表 6 に示す。

表 6 終助詞の結果

助詞	頻度	特徴度の高い話題	変動係数
さ	13,866	日本の未来(300), マナー(211), 天気(112)	.267
の	10,119	動物 (74)	.168
な	29,664	夢・将来設計(48), 住環境(66)	.070
か	18,517	夢・将来設計(55), 住環境(25)	.077
け	1,770	マンガ・ゲーム(37), 旅行 (34), 天気(33)	.379
ね	26,439	日本の未来(34), 天気(29)	.057
よ	16,937	天気(24), 食べること(14)	.086
ねん	3,413	動物(22), 家事(21)	.181
もの	2,373	天気 (14)	.166
や	1,194	ファッション(13)	.164
わ	2,950	マンガ・ゲーム(7) *特徴語とは言えない	.149

「さ」が目立つ。これは音声を聞いただけで、明らかに間投助詞「さ」を多用するスタイルに切り替わったと気づくほどの特徴である。これらの話題は何らかの「話しにくさ」に関係しているかもしれない。全体的な LLR の値は低いが、「天気」が目立つ。

#### 5. おわりに

J-TOCC の分析の結果、助詞の出現頻度も話題によって変化することがわかった。その要因は様々である。話題が直接助詞に影響をしているというより、過去テンスの出現が多い話題には格助詞「から」や接続助詞「て」が多くなる、あるいは静的述語が多い話題では「も」や「し」が多くなるというように、話題は述語を通して助詞に影響することが多い。しかしながら、「日本の未来」で書く格助詞「を」や間投助詞「さ」が多いということは、単に述語の影響とは考えにくく、むしろ話題が文体に影響を与え、これらの助詞の出現頻度はその文体の一部を為しているのではないかと考えられる。今後は、話題別の述語と助詞の関係についてさらに定量的な調査を進めていきたい。

#### 参考文献

- 小澤俊介、内元清貴、伝康晴 (2014) 「BCCWJ に基づく中・長単位解析ツール Comainu」『言語処理学会第 20 回年次大会論文集』
- Nakamata, Naoki. (2019) Vocabulary Depends on Topic, and So Does Grammar, *Journal of Japanese Linguistics*, 35(2)
- 中俣尚己 (2020) 「「自然会話コーパス話題アノテーション情報 ver1.0 」について」<https://www.gsk.or.jp/files/catalog/GSK2020-B/readme.pdf>
- 中俣尚己 (印刷中) 「日本語話題別会話コーパス : J-TOCC」『計量国語学』 33(1)
- 丸山直子(2015) 「コーパスにおける格助詞の使用実態—BCCWJ・CSJ にみる分布—」『計量国語学』 30(3)
- 山内博之 (編) (2013) 『実践日本語教育スタンダード』 ひつじ書房.