

話題に対する知悉度と言語表現の出現傾向の関係

もり あつし 森 篤嗣 (京都外国語大学)

1. 日本語話題別会話コーパス (J-TOCC) について

本発表では、筆者らが構築した『日本語話題別会話コーパス: J-TOCC』(以下, J-TOCC)を用いて、話題に対する知悉度と言語表現の出現傾向に関する分析をおこなう。J-TOCCは、1対1の自然会話を15の話題別にそれぞれ5分観ずつ収録したコーパスである。J-TOCCはJapanese Topic-Oriented Conversation Corpusの略である。

このコーパスの最大の特徴は、15の話題について大学生のペアに5分ずつ話してもらったことで、話題以外の条件が統制されていることにある。話題は身の回りの話題が11、社会に関わる話題が4である。1つの話題につき120ペア240人、合計10時間の会話が収められており、約11万語のコーパスとなっている。性別の組み合わせ(男×男40ペア、男×女40ペア、女×女40ペア)、録音地(東60ペア、西60ペア)の均衡を取っている。

そして、本発表の分析で中心となる「話題に対する知悉度」は、15の話題についての収録が終わったあとに、「今日話した、以下の15の話題のそれぞれについて、自分がどれだけ詳しいか、自信を持って語れたかを5段階の中から1つ選び、数字を○で囲んでください」という指示で、「詳しい・自信あり(5)」から「詳しくない・自信なし(1)」まで5段階で評価してもらったものである。したがって、「話題に対する知悉度という側面だけでなく、「今日の会話の出来映え」の自己評価という側面もあるが、本発表ではこれを話題知悉度として扱うこととする。

2. 話題知悉度の話題別分布

まず、15の話題に対する240人の評価の分布を話題別に示し、それぞれの話題の知悉度の240人の平均値を示す。

表1 話題知悉度の話題別分布

知悉度	1	2	3	4	5	平均値
01.食べること	1	9	84	95	51	3.78
02.ファッション	39	56	71	45	29	2.87
03.旅行	8	29	77	86	40	3.50
04.スポーツ	20	38	63	75	44	3.35
05.マンガ・ゲーム	17	23	48	78	74	3.70
06.家事	16	41	99	56	28	3.16
07.学校	5	8	88	90	49	3.71
08.スマートフォン	4	23	96	87	30	3.48
09.アルバイト	6	18	64	106	46	3.70

10.動物	10	36	82	62	50	3.44
11.天気	14	68	100	41	17	2.91
12.夢・将来設計	11	57	80	58	34	3.20
13.マナー	7	38	106	62	27	3.27
14.住環境	8	39	97	72	24	3.27
15.日本の未来	29	76	69	44	22	2.81
計	195	559	1,224	1,057	565	3.34

上記の通り、全体的に 5（詳しい・自信あり）に偏る傾向があり、1（詳しくない・自信なし）は少なくなっている。これは、15 の話題について 5 分ずつ会話をした後に、質問したため、5 分間それほど話まらずに会話できたということが影響を与えていると予測される。

しかし、やはり話題によって高低はあり、「01.食べること」「05.マンガ・ゲーム」など、大学生にとって身近な話題は高く、「11.天気」「15.日本の未来」などは低い。「02.ファッション」が低いのは意外である。

3. 話題知悉度の男女差

次に話題知悉度（最大値 5）を男女別に集計し、平均値と標準偏差、そして t 検定によって差を検証した。

表 2 話題知悉度の男女差

	男性		女性		t 値
	平均値	標準偏差	平均値	標準偏差	
01.食べること	3.76	0.83	3.79	0.85	-.308
02.ファッション	2.59	1.19	3.15	1.23	-3.565 ***
03.旅行	3.48	1.09	3.53	0.93	-.445
04.スポーツ	3.56	1.17	3.15	1.18	2.691 **
05.マンガ・ゲーム	3.91	1.09	3.50	1.28	2.662 **
06.家事	3.14	1.03	3.18	1.08	-.305
07.学校	3.83	0.84	3.59	0.95	2.021 *
08.スマートフォン	3.62	0.87	3.35	0.89	2.339 *
09.アルバイト	3.78	0.96	3.62	0.93	1.366
10.動物	3.35	1.17	3.53	1.03	-1.288
11.天気	2.77	0.88	3.06	1.06	-2.318 *
12.夢・将来設計	3.29	1.02	3.10	1.16	1.360
13.マナー	3.21	0.96	3.33	0.95	-.944
14.住環境	3.23	0.90	3.32	1.02	-.737
15.日本の未来	2.72	1.02	2.90	1.26	-1.238

* $p<.05$, ** $p<.01$, *** $p<.001$

男女で有意な差が見られたのは、15 話題中 6 話題であった。そのうち、「04.スポーツ」「05.マンガ・ゲーム」「07.学校」「08.スマートフォン」の 4 話題は男性が有意に高く、「02.ファッション」「11.天気」の 2 話題は女性が有意に高かった。

4. 話題知悉度と各話者の発語数との相関

冒頭でも述べたように、話題知悉度は 15 の話題についての収録が終わったあとに 5 段階で評価してもらったものである。すなわち、「今日の会話の出来映え」の自己評価という側面が強く出ていれば、仮説として「たくさん話せた話題ほど、話者は知悉度を高くつける」という相関が見いだせる可能性がある。そこで、本節ではこの仮説を検証するために、話題知悉度と各話者の発語数との相関を検討してみたい。

各話者の発語数を求めるために、中・長単位解析器 Comainu (小澤ほか 2014) により形態素解析をおこなった。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という特徴を持つ。各話者の発語数の形態素解析結果から「記号」「補助記号」を除き、発話量を見るために延べ語数を求め、発話の豊かさを見るために異なり語数も求めて、話題知悉度との Pearson の相関係数を算出した。

表 3 話題知悉度と各話者の発語数との相関

	延べ語数平均	異なり語数平均	延べ語数相関	異なり語数相関
01.食べる	468.89	170.45	0.058	0.076
02.ファッション	489.16	173.35	0.040	0.009
03.旅行	477.88	173.51	0.104	0.156
04.スポーツ	488.86	175.33	0.148	0.112
05.マンガ・ゲーム	488.18	174.22	0.105	0.207
06.家事	486.32	179.60	0.202	0.178
07.学校	480.13	171.99	0.068	0.058
08.スマートフォン	474.35	170.77	0.039	0.124
09.アルバイト	484.37	178.11	0.145	0.167
10.動物	484.48	174.57	0.053	0.088
11.天気	479.38	175.74	0.109	0.088
12.夢・将来設計	466.94	169.75	0.225	0.220
13.マナー	497.18	172.95	0.063	0.111
14.住環境	469.66	165.46	0.132	0.116
15.日本の未来	460.20	168.17	0.152	0.120

話題知悉度と延べ語数との相関を見てみると、最高でも 0.225 と高いとは言えない。「たくさん話せた話題ほど、話者は知悉度を高くつける」とは言えないと考えられる。ちなみに、

話題知悉度と異なり語数との相関も最高 0.220 と高いとは言えず、「使用語彙が豊富であるほど、話者は知悉度を高くつける」というわけでもないことがわかる。

5. 話題に対する知悉度と言語表現の出現傾向の関係

ここから、本題の話題知悉度と言語表現の出現傾向の関係について述べていきたい。まずは全体の傾向をつかむために、全ての話題を統合し、それぞれの語の出現数を使用人数で割って 1 人あたりの各語の使用数を調べた。

表 4 は、総出現数上位 100 語に範囲を限って、1 人あたりの使用数と知悉度（5 段階）について Pearson の相関係数で 0.85 以上の 34 語を品詞別に抽出したものである。

表 4 話題に対する知悉度と言語表現の出現傾向の関係

	語彙素	1	2	3	4	5
助詞-格助詞	の	7.32	7.59	7.96	8.32	8.78
	と	8.41	8.64	8.99	9.41	9.96
	で	5.09	5.29	5.34	5.78	6.25
	に	7.17	7.57	7.68	7.73	8.14
	が	6.67	7.73	7.82	7.82	8.54
	から	0.98	0.94	1.00	1.18	1.26
副詞	結構	0.70	0.76	0.80	0.82	0.88
	もう	2.80	2.92	3.33	3.18	3.45
	こう	0.83	1.00	1.02	1.07	1.08
	そう	9.07	9.07	9.11	9.32	9.75
動詞-一般	食べる	0.14	0.27	0.65	0.79	0.88
	行く	2.19	2.08	2.51	2.83	2.95
	やる	3.20	3.01	3.35	3.69	3.93
助動詞	た	10.71	11.18	12.28	12.72	14.14
	だ	15.74	16.48	16.53	16.54	17.36
	てる	7.21	6.96	7.29	7.48	7.75
助詞-接続助詞	て	7.50	7.96	8.86	9.10	9.99
	し	1.55	1.69	1.67	1.79	1.92
	から	3.55	3.93	3.86	4.03	4.17
助詞-副助詞	って	6.66	7.05	7.07	7.19	7.33
	か	13.10	13.10	13.91	13.95	14.08
	や	0.59	0.63	0.65	0.70	0.67
感動詞-一般	そうそう	1.02	1.15	1.30	1.34	1.31
	あ	2.07	2.29	2.31	2.61	2.48
	え	2.03	2.42	2.27	2.46	2.60
接続詞	で	0.96	1.02	1.44	1.55	2.02
	だから	1.26	1.37	1.32	1.40	1.57

形容詞-一般	凄い	0.94	1.02	1.10	1.19	1.36
形状詞-一般	好き	0.64	0.68	0.87	1.09	1.32
助詞-係助詞	は	8.37	8.85	8.71	9.03	9.23
助詞-準体助詞	の	4.99	5.19	5.16	5.31	5.33
助詞-終助詞	の	2.52	2.73	2.75	2.92	3.08
代名詞	あれ	1.08	1.16	1.17	1.16	1.34
連体詞	あの	1.10	1.27	1.41	1.51	1.53

「あれ」「あの」は、話し手も聞き手も既に知っている人や事柄を指して言うときに使われる。その意味で知悉度との相関が高いことは頷けるが、理論的に考えられてきたことが、実証的にデータとして示される意義は大きい（会話例の括弧内は知悉度）。

E-219-2M(3) : 100 ローのさあ、①あの、あれすごくない、②あの、アップルタルト 100 円。

E-219-1F(4) : ああ。

E-219-2M(3) : あれ、やばくない。

E-219-1F(4) : やばい。でも、え、でも、ちょっと値上がりしてたよ。

E-219-2M(3) : え、あ、そうなの。

E-219-1F(4) : 110 円ぐらいなってた。

E-219-2M(3) : あ、110 円。

E-219-1F(4) : そうそう。

指示詞として既知のものを指す使い方は②である。一方、「あの」は言いよどみにあたる①も多数出現していた。つまり、「あの」と知悉度との関係は、既知のものを指す指示詞の用法だけでなく、談話において「話題を持ちかける話者が何かを思い出すフィラー」としての出現数も多く、話題のイニシアティブとの関係も大きい。

また、上記の例の最後には、「そうそう」も出現している。「そうそう」は、まずは自分自身が知識を披露し、相手がそれに応じるという談話構造が定式化している結果が現れている。「あの」のイニシアティブとの関係も含め、本発表のような定量分析が談話分析に貢献する可能性も考えられる。

「食べる」は、「01.食べる」を中心に「03.旅行」「06.家事」など限られた話題にしか出現していない。しかし、自身が「食べる」ことを話題に出すことは、主体的な自己開示であり、知悉度との関係が高いという結果となった。

E-301-2F(4) : うーん。私はね、何か【人名：3人称】がめっちゃ今、画面バキバキなの。何これって言って、ほんとにね、見るじゃん。(E-301-1F : うん) 見にくくて、(E-301-1F : うん) 何でって言ったら、何か後輩の女の子とラインしてて、その子が何か最近、別れたらしいの、(E-301-1F : うん) 後輩の男の子と。それで、何かその飲みの席で、何かその女の子から、あ、ライン来たみたいな (E-301-1F : うんうんうん) 言って、で、みんなに見せてたら、(E-301-1F : うん) 【人名：3人称】の元カノが (E-301-1F : うん) パって見て、うわーって言って、ガーンって。

E-301-1F(4) : ほんとー？

「て」は機能語であり、これまでの語とは異なる。知悉度と接続助詞「て」の相関の高さは、一文の長さとの関係が考えられる。

W-108-1M(5)：宮島行くのが大好きで。

W-108-2M(4)：ああ、ああ、宮島。

W-108-1M(5)：そう。あすこ、うん。カキがすごいおいしかった。

一方で同じく機能語である「た」は、定延(2008)で検討されているように、「自らの体験を語る」こと、話題に対する知悉度や自信との関係が興味深い。

E-311-1F(4)：ああ、素敵、素敵。いいね。そうなんだ。え、合唱ってさ、(E-311-2F：うん) 私も何か、結構、その、吹奏楽部の先生がもともと声楽科の (E-311-2F：うん) 音大の人で、結構合唱とか (E-311-2F：ああ) 歌うことに関しても、厳しかったんだけど。あの、目をこう開いたりとか、(E-311-2F：ああ) あと何かさ、こう、上からぶら下げられてる。

E-311-2F(5)：うん、つら、つられているような気分だよね。

「こう」は、相手に何かを説明するときに、具体的なジェスチャーを伴って発せられる。いわば説明の時の専用形式であるとも言える。

6. まとめ

本発表では、筆者らが構築した『日本語話題別会話コーパス：J-TOCC』(以下、J-TOCC)を用いて、話題に対する知悉度と言語表現の出現傾向に関する分析をおこなった。前提として、今回の調査の話題知悉度と各話者の発語数との相関が見られないことを述べた。

そして、本題として話題に対する知悉度と言語表現の出現傾向の関係を検証し、「①理論的に考えられてきたことが、実証的にデータとして示せること」、「②話題に関する定量分析が、談話分析に貢献できる可能性があること」を明らかにした。

これから自然会話コーパスは益々充実していくと思われる。一般的には、自然会話コーパスに話題タグは付いていないと思われるため、話題に対する知悉度と言語表現の出現傾向に関する分析から、話し手の知悉度を推測できる指標が開発できると望ましい。書き言葉でも、話し言葉でも今後のコーパス研究は「誰が書いたか／話したか」により重きが置かれる方向に進むと思われるため、コーパスデータから書き手／話し手の特性が推定できる指標の開発は有用であると思われる。

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』582-585.

定延利之 (2008) 『煩悩の文法 体験を語りたがる人びとの欲望が日本語の文法システムをゆさぶる話』ちくま新書.

付記

本研究は JSPS 科研費基盤研究(B) 「話題が語彙・文法・談話ストラテジーに与える影響の解明 (研究代表者：中俣尚己)」(課題番号 18H00676) の助成を受けた。記して感謝申し上げます。