

## 話し言葉における接続詞「で」の特徴分析

太田 公子, 高梨 克也, 井佐原 均

独立行政法人 通信総合研究所

{kimiko,takanasi,isahara}@crl.go.jp

## 1. はじめに

話し言葉に頻繁に現われる接続詞「で」は、経時的に起こる事象を接続する接続詞で、「それで」という指示語系の接続表現から生じた表現であると言われている [1]. 添加型あるいは転換型としての機能があり [2], 一般的な表現である「それで」に対して、よりうちとけた会話的な表現であると位置付けされている. ここで、添加型とは、前の内容の同類や列挙を後に述べる型のことを指し、「そして」「さらに」「また」などがある. 転換型とは、前の内容から転じた別個の内容を後に述べる型で、「ところで」「さて」「じゃあ」などがある. また「で」は、前の文の内容を認めたくえで、そこから次に述べるべき事柄を展開させるときに用いる接続詞 [3] という解釈もある. しかし、書き言葉の場合の多くの接続詞とは異なり、話し言葉における「で」は複数の接続関係を担っていることが予想される. そこで、本研究では、『話し言葉工学』プロジェクトで作成している大規模な自発音声データベースを利用して、「で」の現われ方を検討し、さらに、その接続関係について実験的に検証した.

## 2. 接続詞「で」のふるまい

## 2.1 接続詞の出現頻度

データは「日本語話し言葉コーパス」(Corpus of Spontaneous Japanese, CSJ) を用いた. CSJ は、自発性の高いモノローグを主対象としたコーパスであり、学会における口頭発表 (以下「学会講演」) と、一般話者による個人的な内容に関するスピーチ (以下「模擬講演」) が中心となる [4]. この CSJ の中で、手作業で形態素情報が付与されたもの (83 時間 88 万形態素分、形態素解析の精度は 99.5%) を分析対象データとした. 表 1 にデータ中に現われた接続詞と個数を示す. \* が記された語の品詞は副詞であるが、これらは文間を接続する接続詞と同様の働きもあり、表 1 に示すように非常に出現頻度も高いことからここに載せた. 出現頻度の最も高い接続詞は「で」であり、全接続詞の半数以上を占める.

## 2.2 「で」の出現位置

「で」の出現位置は、文と文の間や述語のテ形の後に出現

表 1: 接続詞とその個数

で	6135	また	765	そして	667
例えば*	650	あるいは	424	特に*	330
んで	325	ただ	222	結局*	205
じゃ	201	しかし	189	更に	141
て	119	および	111	即ち	73
しかも	71	一方	70	から	64
ただし	54	かつ	36	もしくは	31
なお	29	さて	25	と	12
ないし	12	して	11	尤も	7
けれど	6	が	5	んじゃ	2
じゃあ	1	されば	1		

\* の品詞情報は副詞. 合計 9812 接続詞 / 88 万形態素

現するのが典型的であるとされている. また、フィラーのような『感動詞+「で」] や『「で」+他の接続詞』という位置関係があるという [5]. このように話し言葉における「で」は直前直後の他の接続表現と共起する可能性が高いことが特徴的である. 表 2 に、接続詞「で」と「例えば」の直前直後に現われた品詞の出現率 (前の品詞の出現個数 / 各接続詞の個数) を示す. 「で」の直前の品詞は、助詞か助動詞、直後は、感動詞か名詞である. わずかではあるが、直後には代名詞や連体詞もみられる.

- 1: それで何とか続けてこれた面もあるのかなと「で」歴史はあんまり得意じゃなかったから
- 2: マイナスを付けたもので定義されます「で」(F えー) 実際の刺激音の作成方法

## 2.3 「で」の使用の個人差

分析対象としたデータ、83 時間分の CSJ は、338 講演分に相当する. 各講演の長さは異なっているが、参考までに、一講演にどの程度の「で」が出現するかを図 1 にまとめた. 横軸が「で」の個数、縦軸が頻度を示す. 338 講演中、約半数が 0 個~15 個である. 全く出現しなかった (0 個) 講演は 42 講演であった. 一方、100 個も出現した講演もあることから「で」の使用には個人差があると考えられる.

表 2: 接続詞の直前直後に現われる品詞の出現率 (%)

品詞	で (直前)	例えば (直前)	で (直後)	例えば (直後)
1: 感動詞	3.85	35.85	26.67	26.46
2: 記号	0.03	0.00	0.64	2.15
3: 形状詞	0.03	0.00	0.75	0.62
4: 形容詞	0.33	0.15	0.70	1.38
5: 言い淀み	0.49	1.69	2.38	1.38
6: 助詞	32.49	36.92	0.26	0.31
7: 助動詞	58.61	14.77	0.65	2.15
8: 接続詞	0.46	2.15	2.95	0.31
9: 接尾辞	0.34	0.62	0.00	0.00
10: 接頭辞	0.00	0.00	1.03	1.54
11: 代名詞	0.03	0.31	16.07	7.08
12: 動詞	1.74	0.92	1.37	2.31
13: 副詞	0.13	0.62	9.21	2.46
14: 名詞	1.45	5.69	25.98	46.62
15: 連体詞	0.02	0.31	11.36	5.54

表 3: グルーピングの結果

グループ	直前の品詞	直後の品詞
1	助詞	感動詞
2	助動詞	感動詞
3	助詞	名詞
4	助動詞	名詞

そこで、前節で行った「で」の直前直後の品詞情報に注目し、どのような品詞の間に「で」が出現するかを講演ごとに求めた。その際、直前直後の各品詞にはそれぞれに対応した番号が付与される(表 2 参照)。338 (講演) × 133 (「で」の個数の最大) × 2 (品詞番号) の行列が作成されたが、講演ごとに個数が異なるため(欠損値となる)、一講演 10 個以上出現した講演を対象に、「で」の出現個数の中央値(26)をとり、さらに、中央値以下の講演は分析対象に入れないことにした。こうして残った講演数は、95 講演であり、95 (講演) × 26 × 2 の行列作成された。この行列にワード法によるクラスター分析を施し、グルーピングを行った結果、4つのグループができた。

グループ 1 と 3 の講演群は、直前の品詞が助詞、直後の品詞が感動詞か名詞、グループ 2 と 4 の講演群は、直前の品詞が助動詞、直後の品詞が感動詞か名詞というグループである。模擬講演の多くがグループ 1 か 3 に属し、学会講演の多くはグループ 2 か 4 に属していた。模擬講演では「で」の直前は「～ですね」「～と」などが多く、学会講演では「～です」「～ます」が多い。

### 3. 「で」の機能を知るための実験

接続詞の直前直後の品詞情報から「で」と類似したふるまいを示す接続詞は幾つかあったが、これらが、「で」と

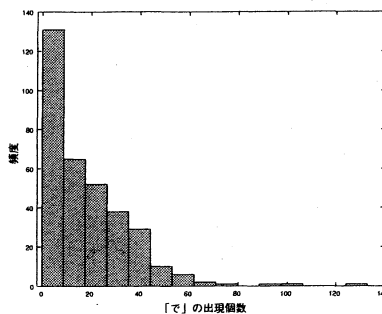


図 1: 「で」の出現個数の分布

表 4: 接続関係の種類 (佐久間 2000 p. 19 より)

- 1 順接型: 前の内容から当然予想される結果を後に述べる
- 2 逆接型: 前の内容から予想に反する内容を後に述べる
- 3 添加型: 前の内容の同類や列挙を後に述べる
- 4 対比型: 前の内容の対照や比較を後に述べる
- 5 同列型: 前の内容の反復や言い換えを後に述べる
- 6 転換型: 前の内容から転じた別個の内容を後に述べる
- 7 補足型: 前の内容を後で補充する

同様の機能を果たしているかどうかはわからない。そこで、「で」が発話された箇所を書き言葉として使用されている接続詞に置き換えるという実験を行うことにした。この実験により、被験者が何を手がかりとしているかを知ることができ、また、「で」と他の接続詞との関係も明確になるものと思われる。

#### 3.1 実験対象データ

前節で行ったグルーピングの結果を利用し、8 講演を実験対象データとした。すなわち、各グループから 2 講演ずつ選ぶが、その際、学会講演、模擬講演から「で」の発話が比較的多い講演を 4 講演ずつ選んだ。

#### 3.2 実験方法

被験者は 23 歳～47 歳の男性 8 名、女性 8 名、計 16 名である。被験者に、講演に出現する「で」を他の接続詞に置き換えるという作業を課す。「で」が出現した箇所には、「で」の代わりに接続詞「だから」「しかし」「そして」「一方」「すなわち」「ところで」「なぜなら」が挿入してあり、被験者はそれらの中から前後の文脈から考えて意味的に適切なもの一つを選択する。どうしても選択不可能な場合は「\*」を選択する。上記した 7 つの接続詞は、表 4 に示す型 [2] を代表する接続詞である。

#### 3.4 結果

##### 3.4.1 選択された接続詞

接続詞がどのように選択されたかを求める。まず、各講

表 5: 選択された接続詞

	*	だから	しかし	そして	一方	すなわち	ところで	なぜなら	計
高一一致率群	8	25	22	133	12	7	12	2	221
%	3.6	11.3	10.0	60.2	5.4	3.2	5.4	0.9	100
低一致率群	1	4	5	17	5	3	4	1	40
%	2.5	10.0	12.5	42.5	12.5	7.5	10.0	2.5	100
計	10	29	27	151	17	10	16	3	263
%	3.8	11.0	10.3	57.4	6.5	3.8	6.1	1.1	100

演、各「で」において8名が選択した接続詞をカウントし、一致率を求める。ここで求める一致率とは、8個の選択肢のうち、4者以上が一致しているものを高一一致率群、3者以下の一致を低一致率群とした。表5に結果を示す。被験者間で一致率の高い箇所が多く選ばれた選択肢は「そして」である。

### 3.4.2 高一一致率群の接続詞の分析

「で」の直前直後の要素の観点から分類し、被験者が最も多く選択した接続詞との関係の分析を行った。「で」の直前の要素として、テ形、文末（倒置を含む）、接続表現、その他（上記以外の要素）、そして、「で」直後の要素として、接続表現、その他とし、人手で分類を行った。また、品詞分類上は接続詞ではないが、意味的には接続関係を表しているもの（副詞など）や複数の形態素連続が固定的に一つの機能を担っている場合（「それで」など）などを考慮して、接続表現の判定は、品詞に因われることなく、より多くのマーカーを採用した。たとえば、直前ならば節末の助動詞や接続助詞、直後ならば接続詞だけでなく、副詞や指示詞などである。表6に分類結果を示す。

### 3.5 分類結果の考察

被験者間の一致率の高い選択肢は「そして」であったが（高一一致率群全体の60.2%の選択率）、特に、直前がテ形である場合には「そして」が選択される割合が高い。直前が文末の場合は、被験者間の選択肢に散らばりがあることから、様々な接続関係に解釈される可能性があることが示唆される。また、「で」の直前あるいは直後が接続表現の時、「そして」以外の接続詞が選択される割合が高くなる。特に、直前が接続表現の場合の「そして」の選択率は35.5%まで減少する。以下に例を挙げながら「で」とその前後の接続表現との関係について述べる。

- 直前が逆接の接続表現「けれども（けども、けど）」「が」の場合、逆接型の接続詞「しかし」が選ばれることが多い（例：それは何だってそんなんじゃないのかと思ったんですけど「で」→「しかし」（Fまー））。

- 直後に「それに対して」や「逆に」のような対比関係を示す接続表現がある場合には「一方」が選ばれる。また「ただ」「でも」のような逆接や保留を示す表現があるときは「しかし」が、「じゃ」のような帰結を示す表現がある場合には「だから」が選ばれるというように、「で」の解釈は直後の接続表現によって影響を受けやすい（例：詳細は説明できません「で」→「一方」逆に文書翻訳の方では）。

- 「まず」「次に」のような順序関係を表す表現や、「そうやって」「いよいよ」のような時間経過を示す表現がある場合には「そして」が選ばれ易い（例：セットされています「で」→「そして」まず正面の緑色の）。

- 「そして」が選択される場合、「で」の直前の文中の要素を指示する前方照応表現（代名詞や連体詞）が「で」の直後に共起していることが多い（例：位相を求めます「で」→「そして」それを(Fおー)逆フリエ変換して）。

- フィラー「ま」は副詞「まあ」からの転用であると考えられ、また「しかし」や「だから」と解釈された「で」の直後に多く出現することから、他のフィラーとは異なり、譲歩や一般化のような発話態度を表す機能を担っている可能性がある（例：近所だったんですね「で」→「だから」（Fま）それからも）。

話し言葉における「で」は書き言葉における「そして」と同様に添加の接続関係を担っていることが多い。しかし、前後にさまざまな接続表現が共起している場合には、添加以外の接続関係であると解釈される割合が多くなる。また、このような前後にさまざまな接続表現を伴っている場合の「で」の場合、接続関係の種類自体はこれらの接続表現によって表されているため、「で」自体の機能は必ずしも特定の種類の接続関係を表現することではないかもしれない。むしろそれは、発話産出上あるいは談話構造上の何らかの「境界」を示す機能を担うものであ

表 6: 分類結果 (横軸は直前の要素, 縦軸は直後の要素, 数値は選択数を示す)

	テ形	文末	接続表現	その他	計
接続表現	* 0	* 4	* 1	* 1	* 6
だから	1	だから 9	だから 2	だから 0	だから 12
しかし	0	しかし 12	しかし 4	しかし 1	しかし 17
そして	9	そして 50	そして 4	そして 0	そして 63
一方	1	一方 5	一方 0	一方 0	一方 6
すなわち	0	すなわち 4	すなわち 0	すなわち 0	すなわち 4
ところで	1	ところで 5	ところで 0	ところで 0	ところで 6
なぜなら	0	なぜなら 1	なぜなら 0	なぜなら 0	なぜなら 1
その他	* 0	* 2	* 0	* 0	* 2
だから	2	だから 8	だから 1	だから 2	だから 13
しかし	1	しかし 2	しかし 2	しかし 0	しかし 5
そして	23	そして 40	そして 2	そして 5	そして 70
一方	0	一方 3	一方 1	一方 2	一方 6
すなわち	0	すなわち 2	すなわち 0	すなわち 1	すなわち 3
ところで	0	ところで 6	ところで 0	ところで 0	ところで 6
なぜなら	0	なぜなら 1	なぜなら 0	なぜなら 0	なぜなら 1
計	* 0	* 6	* 1	* 1	* 8
だから	3	だから 17	だから 3	だから 2	だから 25
しかし	1	しかし 14	しかし 6	しかし 1	しかし 22
そして	31	そして 90	そして 6	そして 5	そして 133
一方	1	一方 8	一方 1	一方 2	一方 12
すなわち	0	すなわち 6	すなわち 0	すなわち 1	すなわち 7
ところで	1	ところで 11	ところで 0	ところで 0	ところで 12
なぜなら	0	なぜなら 2	なぜなら 0	なぜなら 0	なぜなら 2

ると考えられる<sup>1</sup>。発話産出の観点からは、「で」の前後にフィラーが共起しやすいため、「で」にもフィラーと同様の性質があると考えられる。談話構造の観点からは、「で」が談話セグメントの境界に相当する箇所によく現れるといえる。実際、「で」が「ところで」の意味で解釈された箇所では、「○○というのは」のような「話題導入表現」が多く出現する(例:作成しています「で」→「ところで」群遅延というものは)。

#### 4. おわりに

接続詞は談話内の様々な単位の境界とその意味関係を表すものであるため、接続詞の機能が明確になれば、話し言葉の文区切り [6] や談話セグメント境界の認定 [7]、要約 [8] といった言語処理技術に貢献できると考えられる。今後の課題として、「で」で結合されている文または談話セグメントの間の意味論的・語用論的關係についても考慮した分析を行う予定である。

#### 参考文献

- [1] 鈴木一彦, 研究資料日本語文法第 4 巻, 明治書院

<sup>1</sup>表 5 の「転換型」に相当するが, 特に「転換型」については, 前後の二文間の関係だけでなく, 談話のより広い範囲を視野に入れた分析が必要である。

(1997).

- [2] 佐久間まゆみ, 「接続表現 (1) 寺村秀夫編」, ケーススタディ日本語の文章・談話, おうふう (2000).
- [3] 森田良行, 「それで, で, だから」日本語教育辞典, 大修館書店, 364 (1987).
- [4] 古井貞熙, 前川喜久雄, 井佐原均, “科学技術振興調整費開放的融合研究推進制度—大規模コーパスに基づく『話し言葉工学』の構築—”, 日本音響学会誌, 56(11) (2000).
- [5] 土屋菜穂子, “対話コーパスを用いた言い淀みの語の統語論的考察”, 言語処理学会第 5 回年次大会発表論文集, 251-254 (1999).
- [6] 高梨克也, 丸山岳彦, 内元清貴, 井佐原均, “話し言葉の文境界—CSJ コーパスにおける文境界の定義と半自動認定—”, 言語処理学会第 9 回年次大会発表論文集 (2003).
- [7] 森本郁代, 高梨克也, 竹内和広, 小磯花絵, 井佐原均, “話し言葉コーパスへの談話構造タグ付与”, 言語処理学会第 9 回年次大会発表論文集 (2003).
- [8] 野畑周, 関根聡, 内元清貴, 井佐原均, “話し言葉コーパスにおける文の切り分けと重要文抽出”, 第 2 回話し言葉の科学と工学ワークショップ講演予稿集, 93-100 (2002).