

日本語モノログにおけるフィラーの心理的機能

太田 公子, 井佐原 均

独立行政法人情報通信研究機構 (NICT)

{kimiko,isahara}@nict.go.jp

1. はじめに

話し言葉に多く現れるフィラーは、それ自体特定の意味を持たないため、言語的な位置付けは定かではないし [1], 言語処理においても様々な処理の中で削除される場合が多い。そのように扱われる背景として、フィラーは、話し手の極度な緊張による癖 [2], あるいは、整理されていない話しの辻褃合わせとして頻出することが多く、話しの内容との関係が薄いと見なされているからである。しかし、フィラー自体冗長なものであっても、それをとりまく文脈との相互作用によって話しの内容の心理的機能が高まると考えられる。ここで心理的機能とは、話しの内容を予測あるいは推測するための手がかりとみなす。すべてのフィラーに手がかり的な機能があるとは言えないが、ある特定の文脈において、フィラーが特定の機能を持っていることが確認できれば、重要文の抽出や要約に貢献できると思われる。

2. データ

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(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(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(形態素解析の精度は99.5%)を分析及び実験対象データとした。ここで用いるフィラーは、形態素解析結果でFタグが付与されたデータである。

395講演中一講演あたりのフィラー平均数は175.5であった(最多発話数1264/一講演、最小発話数0/一講演)。フィラー出現率(フィラー数/形態素数×100)で表すと、平均6.5%である。また、フィラーの種類と内訳は表1に示す通りである。395講演中フィラー総数は68759、そのうち出現頻度の高いフィラーの種類を示した。

3. 実験

実験で用いたデータは、フィラー出現率が平均値近辺のもの40講演(フィラー総数は5933)で、学会講演17講演、模擬講演23講演である。被験者は各フィラーに対し、以下に示す4つの中から適当と思われる機能

表 1: フィラーの種類

	フィラーの種類	出現率 (%)
e	「えー」「え」	38
ano	「あのー」「あの」	17
ma	「ま」「まー」	16
eto	「えーと」「えーっと」 「えーとー」「えっと」	
sono	「えと」「えっとー」「えーっとー」 「その」「そのー」	6 5
	その他	18
		100

を選択する。もし、どうしても4つの機能に属さないとと思われる場合は、「その他」を選択する。被験者には以下に示す機能名とその説明以外の教示は呈示していない。また、被験者に呈示したデータは、CSJ書き起こしデータの50ms~200msのポーズで区切られた(改行された)もので、Fタグのみが付与された文章である。

間つなぎ:

発話の調子を軽く整えたり、次の言葉を発するまでの時間を稼いだりするような機能

トピックシフト:

トピック(話しの話題)が変化する所で、次のトピックをどのように進めていくか、その方向性を確かめるような機能

発話内容への引き込み(以下「引き込み」):

聴き手が講演に引き込むために、特に重要と思われる発話を強調するような機能

発話の和らげ(以下「和らげ」):

話し手が聴き手を思いやって、話しの負担を和らげるような機能

これらの機能は、山根 [3] がまとめた日本語の談話におけるフィラーの「話し手/聞き手からみたフィラーの存在意義」から抜き出したものである。山根は、フィラーの存在意義として24項目挙げているが、そのうち、間つなぎ(話し手)、境界指示(ここではトピックシフ

表 2: フィラーの種類と選択された機能の関係

	e		ano		ma		eto		sono	
	文頭	文中	文頭	文中	文頭	文中	文頭	文中	文頭	文中
間つなぎ	1614	525	461	455	514	509	322	104	94	124
トピックシフト	110	7	11	3	17	4	51	1		
引き込み	25	1	8	3	7	3	6		1	
和らげ					43	70				

トとした)(話し手及び聞き手), 発話内容への引き込み(話し手), 発話の和らげ(聞き手)を実験用に選択した。

被験者は, 日本語を母国語とする男女 30 名ずつの 60 名。うち半数は, ライターなど言葉に関する職に就いた経験のある者。

4. 結果

4.1 フィラーの種類と機能の関係

まず, 各フィラーに対し, 最も多く選択された機能を求めた。間つなぎ 5531 (93.2%), トピックシフト 213 (3.6%), 発話内容への引き込み 64 (1.1%), 発話の和らげ 113 (1.9%), その他 12 (0.2%)であった。一講演あたり, トピックシフトは約 5 個, 発話内容への引き込みは約 1 個, 発話の和らげは約 2, 3 個に相当する。機能別に振り分けられたフィラーの種類と出現位置は表 2 に示す通りである。

文頭とは, 被験者に与えたフィラー文, 一文の頭の事を指す。間つなぎやトピックシフト, 引き込みは, 文頭に「えー」や「え」として現われるものがほとんどで, 和らげは, 文中の「ま」や「まー」が多い。トピックシフトが選択された文例として, 『(F えー) 続きまして』が挙げられる (F え) か (F えー) かは話者の癖によるが, 話題が変わるということを知り手にもわかってもらえるよう, 少し間を置く意味で, 長く発話する (F えー) の方が多い。表 2 に示すトピックシフトの e の内訳は (F え) 40 (F えー) 70 であった。さらに, フィラーの後は『続きまして』のように次に発話する内容がこれまでと異なることを促す言葉であることが多い。引き込みは, 『(F えー) この結果から日本人学習者は』というように, これまで述べてきた内容の整理を行う前や, 前に述べた内容の中で重要と思われることを再度述べるような所に出現している。和らげは, 引き込みと同様, 前に述べた内容の補足のような場所に出現しているが, 『(F まー) 意外と仲良くやってみました』というように, フィラー後の発話の一文は文末であったり, 『ちょっと (F ま) 汚い』『ただ (F まー) 急行は止まりません』のように, 副詞を伴っている場合が多い。

4.2 談話タグとの関係

CSJ コーパスには, 節単位に談話タグが施されたデー

表 3: 談話タグと機能との関係 (%) (* は談話目的, # は談話下位目的)

	間つなぎ	トピックシフト	引き込み	和らげ
*	6.4	18.1	8.4	6.2
#	8.1	10.7	8.1	7.0

タが 40 講演ある。ここでは, そのうち 10 講演を対象に, 談話タグと実験で選択された機能との関係について調べた。まず, 実験で用いたフィラー文と, 談話タグが施された文と一致する文を抽出した。談話タグは数種類あるが, ここでは, “DS_Purpose” (談話目的) “DS_SubPurpose” (談話下位目的) の 2 種類のものと一致するフィラー文を抽出した。次に, 談話タグと一致したフィラー文に施された機能の選択率を求めた。表 3 にその結果を示す。10 講演中の機能結果付き全フィラー文を 100%としたとき, 談話タグと一致したフィラー文の各機能の選択率を示してある。トピックシフトが 18%を占めており, 談話タグとの一致が認められた。

4.3 重要文タグとの関係

CSJ コーパスには, 要約を行うための重要文抽出結果が施されている。CSJ コーパスで規定された文単位に, 3 名の被験者による抽出結果がタグ付けされており, 文全体の 10%要約及び 50%要約を目標とした重要文がそれぞれ選択されている。ここでは, それらの重要文の中に存在するフィラー文を抽出した。対象とした講演は, 談話タグ抽出時と同様のものである。表 4 に結果を示す。10 講演中の機能付き全フィラー文を 100%としたとき, 重要文タグと一致したフィラー文の各機能の選択率を示してある。被験者や要約率によって異なっているが, 全体的に, トピックシフトより, 引き込みや和らげとの一致が高い。

4.4 談話タグ/重要文タグ以外のフィラー文との関係

この節では, 前節 4.2 及び 4.3 で取り挙げた談話タグ, 重要文タグのどちらにも一致しなかったフィラー文と各機能との関係を調べた。表 5 に結果を示す。和らげとの一致が認められる。比較的重要度の低いフィラーは, 間つなぎの次に和らげが当てはまる可能性が高いことが示唆される。

表 4: 重要文タグと機能との関係 (%) (1_10 とは、被験者番号 1 の者の 10 % 要約)

	間つなぎ	トピックシフト	引き込み	和らげ
1_10	15.4	13.3	15.3	14.7
1_50	11.0	23.6	19.5	21.4
2_10	14.7	13.1	15.2	14.3
2_50	59.1	53.2	60.6	57.3
3_10	21.6	20.5	22.3	22.0
3_50	48.8	42.9	49.4	48.2

表 5: 談話タグ、重要文タグ以外のフィラー文と機能との関係 (%)

間つなぎ	トピックシフト	引き込み	和らげ
16.4	13.0	12.6	17.9

4.5 被験者の各機能に対する解釈のゆれ

この実験では、4 機能の簡単な説明を被験者に呈示しただけなので、被験者間でそれらの解釈は相当ゆれていると思われる。選択された機能のほとんどが間つなぎに偏っているため、間つなぎ以外の機能間におけるデータの信頼性を考える必要がある。そこで、データの散布度を表す測度として、次式によって定義される平均情報量 (H) を用いる [5]: $H = -\sum_i^k P_i \log_2 P_i$ ここで、 k は機能数、 P_i は各機能の相対度数を比率によって表したものである。この指標を用いることにより、機能間のデータの散布の程度を知ることができる。今、データは一つのフィラーに対し、4 つの機能 + その他の 5 つの機能間で散らばっている。最も散らばっている状態とは、被験者の回答が 5 つの機能にまんべんなく振り分けられており ($H = 2.3219(\log_2 5)$)、どの機能が生起するかがまったく予想できない、非常に曖昧な状態を言う。逆に一つの機能に集中している場合は、 H が 0 に近い値になる。

図 1 に実験で使用した全フィラー文に対する H の度数を示す。最も度数の高い H は、 $H = 1.5$ 付近で、4 機能の散布の度合として次に挙げるようなものが多い。

『(F えっとー) 絶対音感群と部分的な絶対音感群では』 [28,9,22,1,0] 『(F まー) そうやって予備校に通っていて』 [27,25,4,4,0]

文例に続くかぎ括弧内の数値は、間つなぎ、トピックシフト、引き込み、和らげ、その他の選択度数である。第一の文例では、文例のフィラーが間つなぎか、引き込みか、どちらか判断に迷う。また、第二文例においても、間つなぎか、それともトピックシフトと考えてよいか判断が難しい。そこで、判断基準を設けるため、

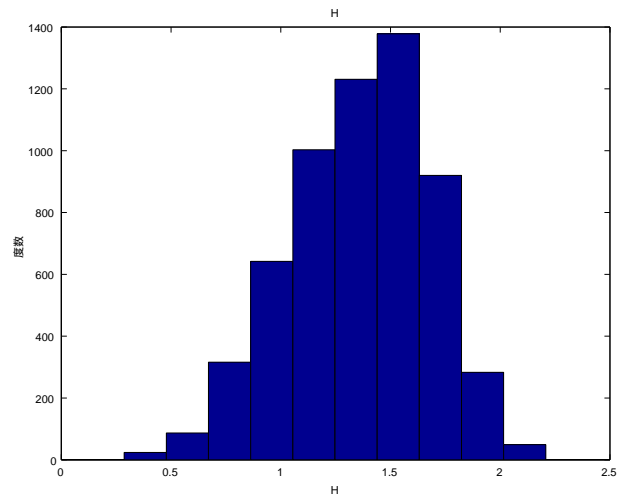


図 1: H の度数分布

各フィラー文の H に条件を与える。

まず、機能結果付きフィラー文を H の値の低い順に並べる。各フィラー文に対し、選択された機能の度数を f とすると (今、10 講演の機能結果付き全フィラー文のデータを対象としている);

$$\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{15} \\ f_{21} & f_{22} & \cdots & f_{25} \\ \vdots & & & \vdots \\ f_{N1} & f_{N2} & \cdots & f_{N5} \end{bmatrix}$$

なる行列ができる。4.2, 4.3, 4.4 節で求めた選択率は、

$$F_{ij} = \left\{ \sum_{i=1}^N f_{i1}; \sum_{i=1}^N f_{i2}; \sum_{i=1}^N f_{i3}; \sum_{i=1}^N f_{i4}; \sum_{i=1}^N f_{i5} \right\}$$

のとき、 $\sum_{i=1}^N f_{ij} / \sum F_{ij} \times 100$ として求めたものであった。その選択率の i を一列ずつシフトさせ、 H と対応させたものが、図 2, 3 である。こうすることによって、各機能の H に対する選択率の変化がわかる。

図 2 のトピックシフト、図 3 の引き込みはそれぞれ指数関数的に減少傾向にあり、 H が安定している箇所以外は選択の程度として曖昧な区間になっている。すなわち、トピックシフトにおいては、 $1.4 \leq H < 1.8$ 、引き込みは、 $1.0 \leq H < 1.8$ 、和らげは、 $0.8 \leq H < 1.7$ の区間が曖昧である。したがって、この区間では、間つなぎ以外の機能の選択を優先させることにする。また、 $H > 1.8$ のデータは信頼性が低いということで削除する。すべての機能結果付きフィラー文に対し、以上のような条件を適応させると、表 6 のようになった。トピックシフトや引き込みの選択率が上昇したことにより、考慮の範囲が広がった。

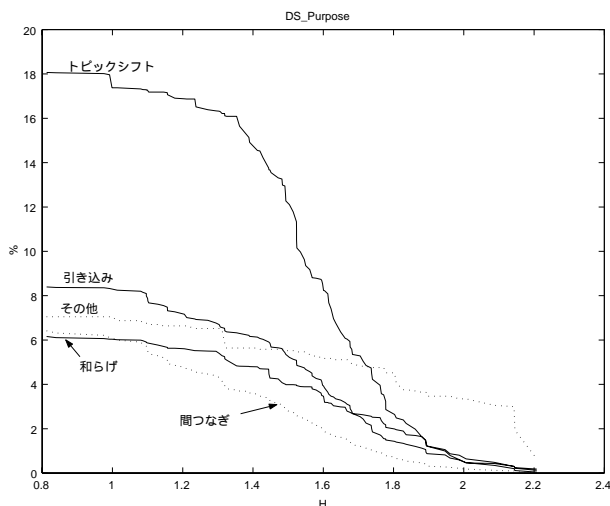


図 2: 談話タグと一致したフィラー文の選択率の推移

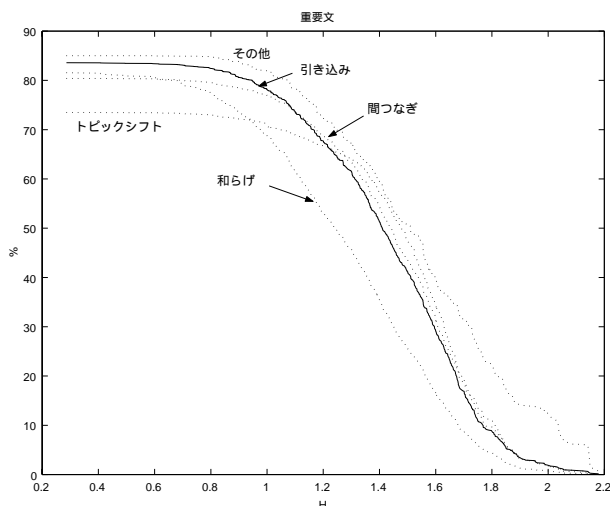


図 3: 重要文と一致したフィラー文の選択率の推移

また、談話タグ、重要文タグと一致したフィラー文のフィラーの種類を条件なし、条件付きで見ると、 H の条件を考慮することにより、重要文の（F ま）の割合が上がった（表 7）。

5. まとめ

トピックシフトとなるフィラーは、（F えーと）や（F えー）、引き込みや和らげは、（F ま）や（F まー）であることが示唆された。どちらのフィラーも、ポーズ後の文頭に出現する。文中に現れる（F ま）は、比較的重要度が低い。間つなぎと他の機能との違いは、フィラー後に現れる言葉である。トピックシフトは、フィラー + [最初に;最後に;今回;今まで;まず;それでは;それから;そこで]、引き込みや和らげは、フィラー + [例えば;実際に;そんな;それで;そのように;そう

表 6: 選択比率 (%) (* は条件なし, # は条件付き)

	間つなぎ	トピックシフト	引き込み	和らげ
*	93.2	3.6	1.1	1.0
#	63.5	19.9	15.7	0.9

表 7: 談話タグ、重要文タグと一致したフィラー文中のフィラーの種類の割合

	談話タグ (条件なし)	談話タグ (条件あり)
えーと	23.4	22.7
えー	15.9	18.6
え	16.8	15.5
ま	6.5	6.2
まー	11.2	11.3
あのー	15.0	13.4
その	3.7	4.1
その他	7.5	8.2
	重要文タグ (条件なし)	重要文タグ (条件あり)
えーと	10.5	8.0
えー	14.9	14.6
え	9.5	1.9
ま	23.6	30.5
まー	11.6	14.6
あの	7.3	6.6
あのー	9.5	10.3
その他	10.9	13.6

いう; そういった; これが; これは; このように; こう
いう; この; ここで]が多い。

参考文献

- [1] 田窪行則, “感動詞の言語的位置付け,” 言語, 34(11), 14-21 (2005).
- [2] M.G. Lallgee and M. Cook, “An experimental investigation of the function of filled pauses in speech,” Language and Speech, 12, pp.24-28 (1969).
- [3] 山根智恵, 日本語の談話におけるフィラー, くろしお出版 (2002).
- [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 - 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] 森 敏昭 編, 心理学のためのデータ解析テクニカルブック, 北大路書房 (1998).
- [6] M. Watanabe, “Fillers as indicators of discourse segment boundaries in Japanese monologues,” Proc. of Speech Prosody 2002, pp.691-694 (2002).
- [7] M. Swerts, A. Wichmann, R. Beun, “Filled pauses as markers of discourse structure,” Proc. of Int'l Conf. on Spoken Language Processing, pp.1033-1036 (1996).