

円滑な発話権制御のための 談話行為の連鎖に基づくフィラーの生起と形態の予測

Predicting Occurrence and Form of Fillers based on Dialog Act Pairs for Smooth Turn-Taking

中西亮輔* 井上昂治 中村静 高梨克也 河原達也

Ryosuke Nakanishi, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, Tatsuya Kawahara

京都大学 大学院情報学研究科
Graduate School of Informatics, Kyoto University

Abstract: In spoken dialogue systems, especially chatting systems for autonomous android robots interacting with human beings, smooth turn-taking function is one of the most important factors to realize natural interaction with users. Speech collisions often occur when a user and the dialog system speak simultaneously. This study presents a method to generate fillers, like “*e-tto*” or “*ano-*” in Japanese, at the beginning of the system utterances to indicate an intention of turn-taking or turn-holding just like human conversations. To this end, we analyzed the relationship between a dialog context and fillers observed in a human-robot interaction corpus, where a user talks with a humanoid robot ERICA remotely operated by a human. At first, we annotated dialog act tags in the dialogue corpus and analyzed the typical type of a sequential pair of dialog acts, called a DA pair. By considering adjacency pairs of the dialog acts, we also identified the DA pairs which could cause speech collisions frequently. Then, we build a machine learning model to predict occurrence of fillers and its proper form from linguistic and prosodic features extracted from the preceding and the following utterances. The results show that it is important for filler prediction to consider its DA pair because the effective feature set depends on the type of DA pair.

1 はじめに

近年、音声対話システムはスマートフォンで利用されるサービスの一つとして実用化され、広く知られるようになった [1]。これらの多くの対話システムはユーザ発話による開始が前提 (*push-to-talk* やマジックワードなど) となっている。また、対話システムの内部状態の表出には画面などのインタフェースが用いられている。しかし、自律型アンドロイドによる対話システムではこれらの前提条件や手法を用いることは難しい。さらに、インタフェースが人間としての存在感を持っている点から、豊かな音声表現や非言語行動も含めた、より自然なインタラクションを行うことが求められている [4]。特にアンドロイドの社会的役割としては、来訪客への応対や面接などが期待されている。

そのため、課題遂行型の対話システムと比べて、このような自然な対話を行うことを目指すシステムでは

円滑な話者交替が重要になる。円滑な話者交替を妨げる問題の一つに、複数話者が同時に発話を開始することで生じる発話の重なり (発話の衝突) がある。そこで、このような非円滑な話者交替が起こりやすい場面について、本研究では、対話システムが発話冒頭で場繋ぎの表現であるフィラーを発することによって発話権の獲得や保持の意思を表明することで発話の衝突を回避し、円滑な発話権制御を実現する枠組みを提案する。

フィラーに関する研究としては、後続する発話の複雑さと発話の意味的な切れ目の強さに注目してフィラーの出現率を調査したもの [5] や、フィラーの各形態がもつ機能に関するもの [6] などがあるが、これらはいずれも主に独話を対象としている。一方、フィラーによってユーザの許容可能なシステムの反応時間を調査した研究 [8] では、「えっと」というフィラーにはユーザがシステム発話の遅延に対して感じる悪印象を軽減する効果があることが示されている。

しかし、従来の研究では対話の流れを踏まえて、システムが多様な形態のフィラーを選択的に生成するこ

*連絡先: 京都大学 情報学研究科 知能情報学専攻
京都市左京区吉田本町
E-mail: nakanisi@sap.ist.i.kyoto-u.ac.jp

とは行われていない。そこで本研究では、談話行為の連鎖に注目することによって、適切なフィラーを選択することを目的として、二話者の対話コーパスを用いたフィラーの生起および形態の予測を行う。

第2節では、本研究で使用するコーパスとアノテーション、さらにアノテーションの集計結果について述べる。第3節では、まずコーパスにおける典型的な談話行為の連鎖パターンを1gramと2gramとを比較して分析し、発話衝突が起りやすい連鎖を隣接ペアの観点から理論的に考察することによって、フィラー予測の対象とすべき談話行為の連鎖を特定する。第4節では、予測対象の連鎖パターンに対してフィラーの生起の有無、および生起すべき形態の予測実験を行う。最後に第5節では、本研究をまとめ、結論を述べる。

2 コーパスとアノテーション

本節では、本研究で使用する対話コーパスとフィラーや長い発話単位、談話行為のアノテーションの認定基準、およびその集計結果を述べる。

2.1 対話コーパス

本研究で用いるデータは、人間の被験者とアンドロイドERICA[9]の初対面対話の音声と映像を収録したものである。ただし、アンドロイドについてはオペレータが遠隔で音声対話と非言語行動の操作を行っているため、以下では「オペレータ」と呼ぶ。オペレータ室のモニタには、被験者の顔とアンドロイドの動作が確認できるように設置されたカメラによる映像がリアルタイムで送られている。

収録は2016年9月に大阪大学で実施した。各セッションの対話時間は10分間程度で、合計39セッションである。オペレータは6名(20~30代の女性)、被験者は39名(20~60代の男性が16名、10~70代の女性が23名)である。

対話における参加者の社会的立場は、被験者が大学の研究室の来訪者、アンドロイドが研究室の秘書である。来訪者は何らかの目的を達成するために教授と面会する必要があるが、来訪時に教授が不在であるため、来訪者は10分程度秘書のアンドロイドと教授を待つ。そのため、基本的な対話の流れは、「来訪目的の説明」の後、「(教授を待つ間の)雑談」である。

オペレータの音声は卓上に設置されたスタンドマイクで、被験者の音声は被験者の足下に設置したガンマイクでそれぞれ収録したものをを用いた。映像は被験者とアンドロイドの両者が映るように配置したビデオカメラで収録した。

以下のアノテーションにはELAN[10]を使用した。

2.2 アノテーション

フィラーの定義は「言い淀み時などに出現する場繋ぎ的な表現」[11]とした。相槌は「話し手が発話権を行使している間に聞き手が送る発話権の移動を伴わない発話」[12]と定義されるため、談話行為タグを付与する話し手役割の発話からは除外する。

典型的な談話行為の連鎖とそれの中のフィラーの生起傾向を捉えるため、本研究では談話行為Dialog Act(以下、DA)タグを用いる。DAは[13]で定められている一般目的機能の分類に基づき、次の4つのクラスを定義した。まず、情報要求機能をもつ発話のDAをQuestion(以下、Q)とし、情報提供機能のInform、行為交渉機能の行為拘束型に属するOfferとPromise、指図型に属するRequestとInstructはまとめてStatement(S)とした。回答や受諾、拒否等の特定のDAに対する応答は全てResponse(R)とし、挨拶や御礼など、Q、R、Sのいずれにも属さないDAはOther(O)とした。ただし、Oのうち、特定次元機能のフィードバックに相当する表現は、主に「そうですね」などの語彙的応答であるため、これらは先行DAに対する応答Rと定義した。本研究ではDA認定の対象となる発話単位として、節単位[14]をもとに定義された統語的・談話的・相互行為的な発話単位である長い発話単位Long Utterance Unit(以下、LUU)[15]を用い、各LUUに対して一つずつDAタグを付与した。

DAタグ(Q、R、S、O)の認定基準の信頼度を検証するため、アノテータ間の一致度をCohenの κ 値[16]で評価した。2名のアノテータが3セッション(評価対象のDAは495個)に対して行った結果、 $\kappa=0.809$ と高い一致度が確認できた。

2.3 アノテーションの集計結果

DAの総数は全39セッションで計6441であった。オペレータの各DAを Q_O 、 R_O 、 S_O 、 O_O 、被験者の各DAを Q_S 、 R_S 、 S_S 、 O_S と表す。 Q_O 、 R_O 、 S_O 、 O_O の生起度数はそれぞれ758、1064、779、706、 Q_S 、 R_S 、 S_S 、 O_S の生起度数はそれぞれ267、1687、477、703であった。

次に、対話の流れを考慮するために、隣接ペア[17]の概念を導入する。隣接ペアは異なる話者の隣り合う二つの発話から構成される。例えば、ある話者が質問を行った後に別の話者が回答を行う場合は、質問が連鎖の第一部分、回答が第二部分に相当する。隣り合った二つ組のDAによって構成されるDAの連鎖(以下、DA連鎖)は、本コーパス中に6402組含まれているが、そのうち発話の重なりがある箇所を除外した5080組を分析対象とした。対話セッション数は39であるため、対象となるDA連鎖を構成するDAそのものの総数は5119(=5080+39)個である。

フィラーの総数は 4292 であった。本研究で対象とするフィラーは、先行 DA と重なりがなく後続する DA の冒頭に生起する 1460 個（オペレータ：875 個，被験者：585 個）のみに限定する。これらは発話権制御に関係していると考えられるものである。先行 DA と後続 DA の間にフィラーが生起する DA 連鎖の割合は 28.7% (=1460/5080) であり，全フィラーのうち DA の冒頭に生起するフィラーの割合は 34.0% (=1460/4292) である。また，先行 DA と後続 DA の話者が異なる「話者交替」時の DA 連鎖が 2516 組，フィラーが 836 個であり，話者が同一である「話者継続」時の DA 連鎖が 2564 組，フィラーが 624 個であった。従って，話者交替時にフィラーが生起する割合は 33.2% (=836/2516)，であり，話者継続時にフィラーが生起する割合は 24.3% (=624/2564) である。

3 DA 連鎖に基づくフィラーの分析

本節では 2 つの観点からフィラーを予測する DA 連鎖の種類を限定する。1 つは典型的な連鎖で，もう 1 つは発話衝突の生起確率が高いと考えられる連鎖である。

3.1 典型的な DA 連鎖

アノテーションの集計結果から，対象となる DA の中で生起度数が大きいのは， R_S (1395)， R_O (826)， Q_O (558) である。分析対象となる DA の生起度数と総数 5119 個を用いて DA の 1gram を算出した結果を表 1，表 2 の最下行に示す。

この 1gram の値と 2gram の値とを比較することによって，典型的な DA 連鎖を特定する。ただし，フィラーは後続する DA の冒頭に付属する発話であると考えられるため，本稿では後続 DA を基準として，それぞれの先行 DA との間の 2gram を算出した。結果を表 1，表 2 に示す。表の各行は先行 DA を表し，各列は後続 DA を表している。後続 DA のうち，先行 DA と話者が異なる場合が「話者交替」，同一の場合は「話者継続」である。表 1 の先行話者はオペレータ，表 2 の先行話者は被験者である。表中の括弧内の数値は DA 連鎖の生起度数，最終行は各列に対応する DA の 1gram の値を表している。太字箇所は DA 連鎖の度数が大きいものである。網掛け箇所は後続 DA の 1gram と比較して生起確率が各列の中で最も大きい特徴的な DA 連鎖を示す。表 1，表 2 より，先行 DA から後続 DA への遷移には明確な偏りがあるとわかる。

表 1，表 2 中の太字箇所および網掛け箇所の DA 連鎖を考慮すると，本コーパスにおける典型的な対話の流れは図 1 のように表せる。図中の各 DA 連鎖に付与した丸囲み数字で示されているものは表 1，表 2 中のものと対応している。例えば，先行話者がオペレータの

遷移①はオペレータが質問 (Q_O) をして，被験者がそれに応答 (R_S) する連鎖に相当する。これはオペレータが隣接ペアの第一部分である質問を行い，被験者が第二部分でこの質問に応答するものである。

本研究と同様の ERICA を用いた遠隔対話で，以前に行った予備収録データを用いた分析 [18] と比較すると，今回と共通する典型的な DA 連鎖は図 1 のオペレータが先行話者となる①と⑥，被験者が先行話者となる⑤である。予備収録では，対話の主導権がオペレータに偏っており，隣接ペアの第一部分の発話の大半はオペレータによるものとなる。一方，本コーパスでは秘書と来訪者という設定のもとで対話を行っているため，混合主導の対話となり，典型的な連鎖パターンが多くなっている。そのため本研究では，オペレータと被験者を区別せずに分析および予測実験を行う。

3.2 発話衝突が起こりやすい DA 連鎖

発話衝突は対話内でどちらの参加者が発話権を有しているかが不明確になりやすい箇所できりやすいと考えられる。隣接ペアを考慮すると，対話の主導権に極端な偏りが無い限り，現行話者が当該連鎖を継続するのか，あるいは他の参加者が新たな連鎖を開始するのかが曖昧である場合に発話権の保持者が不明確となりやすいと仮定できる。すなわち，一方ではこれまでの連鎖の継続（特に R の後の R）が可能であり，もう一方では新たに第一部分で連鎖を開始することも可能であるため，話者の交替と継続の曖昧性が高くなり，発話衝突の可能性が高くなると考えられる。

第一部分で生起可能な DA は Q, S, O であるが，O は挨拶や御礼などの定型表現に相当し，発話衝突の問題に対して重要でないため本研究では取り扱わない。したがって，発話衝突が起こりやすい DA 連鎖は R か S から Q または S に遷移するパターン（表 1，表 2 中の⑤から⑨）であると考えられる。そこで次節では，3.1 節で挙げた①から⑦の典型的な DA 連鎖と⑤から⑨の発話衝突の可能性が高い DA 連鎖両方を対象に予測実験を行う。

4 予測実験

本章では，前節で示した DA 連鎖を対象に，フィラーの生起および形態の予測を行う。なお，本研究では一般的なモデルを構築することを優先して，オペレータによる DA か被験者による DA かは区別せずにまとめてモデルを学習する。すなわち，例えば，「 $Q_O R_S$ 」と「 $Q_S R_O$ 」は「QR 交替 (①)」としてまとめて扱う。

4.1 予測カテゴリ設定

本研究における仮説は，生起するフィラーの形態には DA 連鎖の種類ごとに偏りがあるというものである。

表 1: 先行 DA の話者がオペレータであるときの bigram と生起度数

後続DA 先行DA		(話者交替) 被験者				(話者継続) オペレータ				計
		Q _s	R _s	S _s	O _s	Q _o	R _o	S _o	O _o	
オペレータ	Q _o	0.01 (3)	① 0.79 (441)	0.01 (7)	0.01 (8)	0.10 (57)	0.01 (5)	0.04 (23)	0.03 (14)	1.00 (558)
	R _o	⑤ 0.05 (41)	③ 0.11 (89)	⑧ 0.08 (66)	0.05 (42)	⑥ 0.16 (128)	② 0.29 (243)	⑨ 0.17 (141)	0.09 (76)	1.00 (826)
	S _o	0.05 (31)	④ 0.24 (158)	0.05 (29)	0.11 (72)	0.09 (58)	0.01 (3)	⑦ 0.39 (250)	0.07 (47)	1.00 (648)
	O _o	0.03 (14)	0.01 (6)	0.06 (30)	0.26 (130)	0.18 (91)	0.03 (17)	0.28 (142)	0.15 (77)	1.00 (507)
1gram		0.04	0.29	0.07	0.11	0.09	0.18	0.11	0.11	

表 2: 先行 DA の話者が被験者であるときの bigram と生起度数

後続DA 先行DA		(話者交替) オペレータ				(話者継続) 被験者				計
		Q _o	R _o	S _o	O _o	Q _s	R _s	S _s	O _s	
被験者	Q _s	0.05 (12)	① 0.78 (186)	0.02 (4)	0.03 (6)	0.08 (18)	0.00 (1)	0.03 (7)	0.02 (5)	1.00 (239)
	R _s	⑤ 0.16 (216)	③ 0.18 (247)	⑧ 0.07 (98)	0.07 (92)	⑥ 0.04 (59)	② 0.35 (488)	⑨ 0.09 (125)	0.05 (70)	1.00 (1395)
	S _s	0.12 (51)	④ 0.28 (116)	0.07 (30)	0.11 (45)	0.09 (39)	0.01 (2)	⑦ 0.27 (115)	0.05 (22)	1.00 (420)
	O _s	0.10 (50)	0.03 (16)	0.08 (41)	0.29 (139)	0.04 (17)	0.02 (9)	0.11 (55)	0.33 (160)	1.00 (487)
1gram		0.09	0.18	0.11	0.11	0.04	0.29	0.07	0.11	

そこで、各 DA 連鎖ごとに設定したクラスのフィルターについて、フィルターに先行する発話と後続する発話から抽出される言語的な特徴と韻律的な特徴を用いて予測する。その際、すべてのフィルターの表層形をそのまま用いると異なりが大きくなりすぎるため、形態および機能を考慮してフィルターのクラスを定義する (表 3)。これらのクラスのうち、各 DA 連鎖において生起度数が最大のフィルターの形態クラスを「代表的なフィルター」とし、「それ以外のフィルター」、「フィルターをうたない」を加えた 3 クラスを設定した。ただし、フィルターの生起度数が小さい、またはある一つのフィルター形態に極端な偏りがある DA 連鎖では「フィルターをうつ (フ)」と「フィルターをうたない (無)」の 2 クラスのみを設定した。この方法で各 DA 連鎖に対して設定した代表的な形態について表 4 に示す。表中「- (ハイフン)」で示したものは 2 クラスに設定した連鎖である。

4.2 分類器と特徴量

今回は、フィルターをうつタイミングは先行発話の DA の終了後であるという前提のもとで予測実験を行う。また、後続発話はフィルターに続けてシステムが生成するものであるため、先行発話の DA だけでなく後続発話の DA も所与とする。

分類器は Random Forest (RF) を使用し、5 分割交差検定によって評価を行った。分類器については機械学習ライブラリの Scikit-learn[19] を用いた。RF でブートストラップ法によって生成する決定木の数は 10 に設定して学習を行った。

各カテゴリのサンプル数に大きな差があったため、学習時には最もサンプル数の少ないカテゴリにサンプル数を合わせて学習し、評価時には元の分布にしたがって評価した。評価には、適合率 (precision)、再現率 (recall)、F 値 (f-measure) を用いた。

学習に用いる特徴量は、先行 LUU (pLUU: preced-

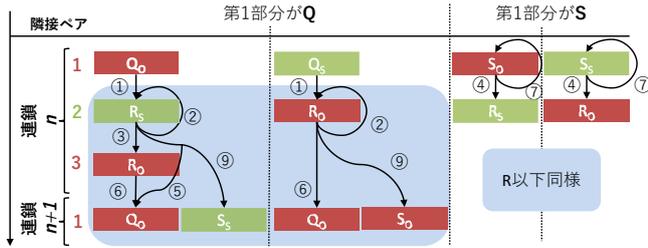


図 1: DA 連鎖を考慮した典型的な対話の流れ

ing LUU) から抽出される言語的特徴量 (L: Linguistic features) と韻律的特徴量 (P: Prosodic features), および後続 LUU (fLUU: following LUU) から抽出される言語的特徴量 (L) であり, 表 5 のようになる。

まず, 言語的特徴量のうち, DA は 2.2 節でアノテーションしたものを用いる。また, 先行発話末の品詞が話者交替と継続のそれぞれで特徴が異なること [20] が指摘されているため, 先行発話末の品詞を特徴量とした。その際, これが助動詞または終助詞である場合には, その表層形も特徴量として加えた。これは, 文末から得られる特徴の詳細度を高めることを目的としている。先行発話の長さは確定的であるため, 先行発話の単語数と文節数も特徴に加えた。同様に, 後続発話の冒頭の品詞も有効であると考えられるため, その品詞を特徴量に加えた。その際, 接続詞はフィラーとともに発話冒頭の特徴的な品詞であることから [21], 接続詞の場合には表層形も特徴量として加えた。さらに, 発話の統語的・意味的な切れ目の大きさも重要な特徴と考えられるため, 先行発話末の節境界も特徴量に加えた。これは, 節単位 [14] の分類に従い, 文末表現に相当する「絶対境界」, 文末表現ではないが統語的に大きな切れ目である「強境界」, これら以外の「その他」の 3 種類とした。一方, 後続する発話の長さはフィラーの生起に重要な特徴であることが指摘されているため [5], 後続発話の長さも特徴とした。ただし, 実際のリアルタイム対話システムでは後続発話の長さはフィラー生成時には正確に確定されていないものと仮定し, 後続発話の単語数と文節数を「長い」と「短い」の 2 値に量子化した。単語数と品詞の抽出には MeCab[22], 文節数の抽出には CaboCha[23] を用いた。

また, 話者交替と継続の区別には韻律的特徴が関係していること [20] や, 交替と継続の予測のために用いる特徴量として有効であること [26] が指摘されているため, 本稿ではフィラーの予測に先行発話末の韻律的特徴も用いる。F0 とパワーは STRAIGHT[24, 25] (XSX 法 [27]) を用いて抽出した。先行発話末 100ms から得られる値を用いて, F0 とパワーについてそれぞれ, 一次回帰係数, 平均, 最大, 最小を算出し, これらの特徴量とした。発話長は pLUU の開始から終了までの時間 (ms) とした。話速は LUU のかなの文字数を発話

表 3: フィラーのクラス

クラス	形態の例
固有系 (固)	えっとー, えー, えーっと, etc.
指示詞系 (指)	あの一, その一, この一, etc.
副詞系 (副)	まー, もー, なんか, etc.
気づき系 (気)	あ, え
その他 (他)	なんていうか, etc.
うたない (無)	

表 4: 各 DA 連鎖の代表的な形態

番号 (図1)	DA連鎖	代表的な形態
①	QR交替	気づき系
②	RR継続	指示詞系
③	RR交替	-
④	SR交替	-
⑤	RQ交替	気づき系
⑥	RQ継続	固有系
⑦	SS継続	-
⑧	RS交替	固有系
⑨	RS継続	指示詞系

長で割ったもので近似し, ポーズは pLUU の終了からフィラーがある場合にはフィラーの開始, フィラーがない場合に fLUU の開始までの時間 (ms) とした。

4.3 予測結果

3 節で特定した典型的な DA 連鎖と発話衝突が起こりやすい連鎖ごとのフィラーの予測結果 (F 値の平均) を表 6 に示す。特徴量の選択としては, 先行発話 (pLUU) の言語 (L) と韻律 (P), 後続発話 (fLUU) の言語のすべての組み合わせが可能であるが, 結果はそれぞれの DA 連鎖について用いた特徴量 (pLUU-L+P など) の組み合わせのうち, 最も F 値の平均が高かったもののみについて, 予測カテゴリ別の精度, 再現率, F 値を示している。代表的な形態の結果が斜線の箇所はフィラーをうつ/うたないの 2 クラス分類のものである。上段は話者交替時の連鎖, 下段は話者継続時のものである。

上段の話者交替時の DA 連鎖では後続発話 (fLUU) の言語的特徴 (P) が有効であることが共通しており, 話者交替時のフィラーの予測のためには後続発話の言語的特徴が有効であることが示唆される。これは, 交替時には DA で表される後続発話の種類がある程度決

表 5: 使用する特徴量

抽出対象	種類	特徴量	
先行発話 (pLUU)	言語 (L)	DA, 発話末品詞, 発話末表層形 (品詞が助動詞と終助詞の場合), 節境界, 単語数, 文節数	
	韻律 (P)	F0 パワー (発話末100ms)	一回帰係数 最大値 最小値 平均 発話長, 話速, ポーズ
後続発話 (fLUU)	言語 (L)	DA, 発話冒頭品詞, 単語数, 文節数 (量子化された値), 発話冒頭表層形 (品詞が接続詞の場合)	

定した状態で発話されるため、それと同時に用いられやすいフィラーも決まるためである。逆に、下段の話者継続時には先行発話 (pLUU) の特徴が有効である。特に、後続 DA が S である場合には韻律の特徴が有効であることが共通している。

5 結論

本研究では、談話行為 (DA) の連鎖を用いて対話の流れを考慮することによって、アンドロイドの対話システムによる適切なフィラーの選択的な生成を目的と

して、フィラーの生起および形態を予測するモデルを構築し、モデルの評価実験を行った。

3節では DA および DA 連鎖の生起確率を確認し、本コーパスにおける典型的な DA 連鎖を示した。また、隣接ペアの構造を理論的に考慮することにより、発話衝突が起こりやすい連鎖を特定した。4節では3節で特定した種類の DA 連鎖のそれぞれを区別した予測モデルを構築し、予測結果を評価した。その結果、(1) DA 連鎖を考慮することで話者交替時と話者継続時のそれぞれについてフィラーの予測に有効な特徴量に違いがあること、(2) 先行発話と後続発話、さらに言語的特徴と韻律的特徴のそれぞれがフィラーの生起および形態の識別に有効であることが確かめられた。

最後に、今後の課題としては次のような点が考えられる。今後は、発話衝突および発話衝突に起因する言い直しという問題に対するフィラーを生成することの有効性を検証するための評価実験を行う予定である。

参考文献

- [1] 河原達也. 音声対話システムの進化と淘汰: 歴史と最近の技術動向. 人工知能学会誌, Vol. 28, No. 1, pp. 45-51, 2013.
- [2] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 人工知能学会研究会資料, SLUD-A903-11, pp. 61-66, 2010.
- [3] 東中竜一郎. 雑談対話システムに向けた取り組み. 人工知能学会研究会資料, SIG-SLUD-B303-14, pp. 65-70, 2014.

表 6: フィラーの予測実験結果 (精度, 再現率, F 値)

DA連鎖		QR交替 (①)			RR交替 (③)			SR交替 (④)			RQ交替 (⑤)			RS交替 (⑧)		
分類クラス		気/他/無			フ/無			フ/無			気/他/無			固/他/無		
特徴量	抽出対象	fLUU			fLUU			fLUU			fLUU			pLUU+fLUU		
	種類	L			L			L			L			L		
		精度	再現率	F値	精度	再現率	F値									
フィラー クラス	代表的な形態	0.33	0.34	0.33	/	/	/	/	/	/	0.51	0.60	0.55	0.17	0.65	0.26
	その他	0.40	0.75	0.52	0.59	0.64	0.61	0.57	0.72	0.63	0.28	0.76	0.41	0.35	0.57	0.39
	うたない	0.85	0.66	0.74	0.74	0.69	0.72	0.80	0.69	0.74	0.72	0.32	0.44	0.84	0.44	0.57

DA連鎖		RR継続 (②)			RQ継続 (⑥)			SS継続 (⑦)			RS継続 (⑨)		
分類クラス		指/他/無			固/他/無			フ/無			指/他/無		
特徴量	抽出対象	pLUU+fLUU			pLUU+fLUU			pLUU+fLUU			pLUU		
	種類	L			L			L+P			P		
		精度	再現率	F値	精度	再現率	F値	精度	再現率	F値	精度	再現率	F値
フィラー クラス	代表的な形態	0.26	0.68	0.37	0.29	0.58	0.34	/	/	/	0.16	0.32	0.21
	その他	0.16	0.26	0.20	0.22	0.55	0.31	0.27	0.40	0.32	0.23	0.46	0.30
	うたない	0.86	0.53	0.65	0.78	0.39	0.51	0.77	0.64	0.70	0.70	0.38	0.48

- [4] 井上昂治, 河原達也. 自律型アンドロイド Erica のための音声対話システム. 人工知能学会研究会資料, SLUD-B502-5, pp. 21-24, 2015.
- [5] M.Watanabe. Features and Roles of Filled Pauses in Speech Communication: A corpus-based study of spontaneous speech. Hitsuji Syobo Publishing, 2009.
- [6] 川田拓也. 日本語フィラーの音声形式とその特徴について—聞き手とのインタラクションの程度を指標として—. 博士課程学位論文, 2010. <https://doi.org/10.14989/doctor.k15563>
- [7] 前川喜久雄. 自発音声中のフィラーの特性に関する予備的分析: 位置と高さの分析. 音声研究, Vol. 16, No.3, pp. 106-107, 2012.
- [8] T.Shiwa, T.Kanda, M. Imai, H.Ishiguro and N.Hagita. How quickly should communication robots respond?. International Journal of Social Robotics, vol. 1, pp. 153-160, 2009.
- [9] Koji Inoue, Pierrick Milhorat, Divesh Lala, Tianyu Zhao and Tatsuya Kawahara. Talking with ERICA, an autonomous android. SIGdial Meeting on Discourse and Dialogue (SIGDIAL), pp.212-215, 2016.
- [10] H.Brugman, A.Russel. Annotating Multimedia/Multi-modal resources with ELAN. In Proc. LREC, pp. 2065-2068, 2004.
- [11] 小磯花絵, 西川賢哉, 間淵洋子. 転記テキスト. 『日本語話し言葉コーパスの構築法』, pp. 23-132, 2006. http://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/02.pdf.
- [12] 泉子・K・メイナード. 会話分析. くろしお出版, 1993.
- [13] B.Harry, J.Alexandersson, J.Carletta, J.Choe, A.Fang, K.Hasida, K.Lee, V.Petukhova, A.Popescu-Belis, L.Romary, C.Soria and D.Traum. Towards an ISO standard for dialogue act annotation. Seventh conference on International Language Resources and Evaluation, pp. 2548-2555, 2010.
- [14] 丸山岳彦, 高梨克也, 内元清貴. 節単位情報. 『日本語話し言葉コーパスの構築法』, 国立国語研究所, pp. 255-322, 2006. http://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/05.pdf.
- [15] Japanese Discourse Research Initiative. 発話単位ラベリングマニュアル *version 2.0*. <http://www.jdri.org/resources/manuals/uu-doc-2.0.pdf>.
- [16] J.Cohen. A coefficient of agreement for nominal scales. Educational and psychological measurement Vol.20, No.1, pp. 37-46, 1960.
- [17] 伝康晴. 隣接ペア, 多人数インタラクションの分析手法. オーム社, pp. 82-94, 2009.
- [18] 中西亮輔, 井上昂治, 中村静, 高梨克也, 河原達也. 自律型アンドロイドによる円滑な発話権制御のためのフィラーの生起位置と形態の分析. 人工知能学会研究会資料, SLUD-B503-11, pp. 61-66, 2016.
- [19] F.Pedregosa, G.Varoquaux, A.Gramfort, V.Michel, B.Thirion, O.Grisel, M.Blondel, P.Prettenhofer, R.Weiss, V.Dubourg, J.Vanderplas, A.Passos, D.Cournapeau, M.Brucher, M.Perrot and E.Duchesnay. Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research Vol.12, pp. 2825-2830, 2011.
- [20] H.Koiso, Y.Horiuchi, S.Tutiya, A.Ichikawa, and Y.Den. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. Language and speech, Vol. 41, No. 3-4, pp. 295-321, 1998.
- [21] 伝康晴. コーパス言語学的手法による音声対話の分析: 認知・相互行為背反の観点から. 電子情報通信学会技術研究報告, Vol. 116, No. 185, pp. 19-24, 2016.
- [22] T.Kudo, K.Yamamoto and Y.Matsumoto. Applying Conditional Random Fields to Japanese Morphological Analysis. Conference on Empirical Methods in Natural Language Processing, pp. 230-237, 2004.
- [23] T.Kudo and Y.Matsumoto. Japanese dependency analysis using cascaded chunking. In CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 PostConference Workshops), pp. 63-69, 2002.
- [24] H.Kawahara, I.Masuda-Katsuse and A.de Cheveign'e. Restructuring speech representations using a pitch-adaptive time - frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds, Speech communication, Vol. 27, No. 3, pp. 187-207, 1999.
- [25] H.Kawahara, M.Morise, T.Takahashi, R.Nisimura, T.Irino and H.Banno. TANDEM-STRAIGHT : A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation, Proc. ICASSP, pp. 3933-3936, 2008.
- [26] 大須賀智子, 堀内靖雄, 西田昌史, 市川熹, 音声対話での話者交替/継続の予測における韻律情報の有効性. 人工知能学会論文誌, Vol.21, No.1, pp. 1-8, 2006.
- [27] H.Itagaki, M.Morise, R.Nisimura, T.Irino and H.Kawahara. A bottom-up procedure to extract periodicity structure of voiced sounds and its application to represent and restoration of pathological voices. Proc. MAVEBA, pp. 115-118, 2009.