

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

Analysis on Form and Location of Fillers for Smooth Turn-Taking by Autonomous Android

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, smooth turn-taking function is one of the most important techniques to realize natural interaction with users. Fillers are presumed to play an important role in natural turn-taking in human dialogue. To this end, in this study, we analyzed a dialogue corpus where people talk with a humanoid android ERICA which was remotely operated. At first, we examined the occurrence rate of fillers at clause units in the case of turn-taking or not. The results suggest that we can see some differences among the roles of the participants. Next, some functions of each filler form were analyzed, which suggests a possibility that forms of fillers can be predicted using an adjacency 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]。これらは、発話内容を取り扱うタスク指向型対話システムであり、対話は一問一答を行う機械的なやり取りとなっている。一方で、雑談型対話システムの研究も進められている [2][3]。特に自律型アンドロイドへの実装を想定した雑談型対話システムでは、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、人間と自然なインタラクションを行う対話システムの実現が目指されている [4]。

人間と自然なインタラクションを実現するための重要課題の一つとして、円滑な発話権制御がある。人間同士の対話における発話権制御では、「えーと」や「あー」といった場繋ぎ的表現であるフィラーや、対話相手の発話権維持を促すために用いられる相槌といった対話中の振る舞いを生成することが重要であると考えられる。フィラーに関する研究としては、出現率に関するもの [5] や、機能に関するもの [6] が挙げられるが、これらはいずれも独話を対象としている。

そこで本研究では、対話におけるフィラーの生成や認識に基づく円滑な発話権制御を目的として、フィラーの生起位置および形態に関する分析を行う。第 3 節で

は、話者交替の起こり得る箇所において、話者交替が起こった場合と起こらなかった場合のフィラーの生起比率を比較する。第 4 節では、話者の立場ごとに用いられるフィラーを生起位置および機能の観点から比較する。第 5 節では、フィラーの各形態における特徴を生起位置および機能の観点から話者の立場ごとに比較する。第 6 節では、フィラー生成による発話権制御に利用できる知見を検討する。第 7 節では、今後の課題を述べる。これにより、システムがフィラーを生成する場合、話者交替が起こり得る箇所、またはそうでない箇所ではフィラーを打つべきか否か、さらにそこでどのような形態を用いるべきかが明らかになると期待される。

2 コーパス

本研究で、分析に用いるデータは、被験者とアンドロイド ERICA が対面で行う二者間での初対面对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マイクを通して遠隔で対話をしているため、以下では「オペレータ」と呼ぶ。

各セッションの対話時間は 10 分間程度で、合計 15 セッションである。オペレータは 1 名、被験者は 15 名である。本研究では、このうち 6 セッションを用いて分析を行った。オペレータに対して教示した対話の流れは、フェーズ 1「挨拶と居住地や趣味等の雑談」、フェー

*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

ズ2「ロボットについての雑談」である。音声は被験者の椅子の下に設置したワイヤレスマイクで、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、それぞれ収録した。書き起こしは、ワイヤレスマイクで収録された音声をもとに行った（以下のアノテーションには、ELAN[7]を使用）。

アノテーションに用いるフィルターの定義は、「言い淀み時などに出現する場繋ぎ的な表現」[8]とした。

発話権制御を目的とした本研究では、話者交替が起こり得る位置である TRP(turn relevant position)[9]を取り扱う必要があり、その基準としては、TCU(turn constructional unit)[9]の末尾が考えられる。しかし、TCUの認定は容易でないため[10]、本研究では、節単位[11][12]（以下、CU）の末尾（絶対境界および強境界）を近似的に用いることにする。ただし、倒置が生じている箇所では、倒置部の直前の節末と直後の位置の両方をCU末とする。また、CU末以外の箇所はCU途中と定義する。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」[13]と定義されるため、話し手役割における発話からは除外することとした。

さらに、話者交替[9]に関わるフィルターの機能を分析するため、実際に発話権が移行した箇所（以下、交替）および発話が継続した箇所（以下、継続）をそれぞれ認定する。これに基づき、各節境界について、先行節単位とは話者が異なる場合を交替、同じ場合を継続とした。フィルターについても、先行発話の話者とは異なる参加者によるものの場合には交替、話者が同一の場合には継続と認定する。また、フィルターが連続して生じた場合、2つ目以降のフィルターは全て継続とする。

コーパス中の対話データのうち、オペレータと被験者の質問の数は1対話の集計のみではあるが、それぞれ23対1と、主導権に大きな偏りが見られる。従って、本コーパスでの基本的な連鎖構造は、主にオペレータが隣接ペア[14]の第一位置である質問を行い、被験者が第二位置でこの質問に回答し、さらにオペレータが第三位置で再回答するという構造となる。

3 話者交替とフィルターの生起

3.1 分析対象

CU末での交替および継続のそれぞれの直後でフィルターが生起する割合を比較した。結果を図1に示す。なお、生起しているフィルターのうち、「あ」については、フィルターではなく相槌としての感情表出系感動詞[15]とみなす考え方もありうるため、別扱いにした。ただし、平板調で引き伸ばされた「あー」のうち、相槌でないものはフィルターとして取り扱うこととした。

ここで扱うCUの度数には、語彙的応答[15]や自己応答も含めているが、それぞれの頻度は170、31であっ

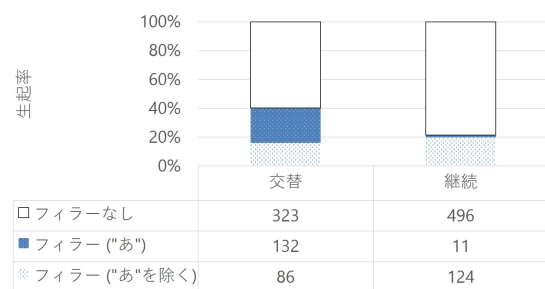


図1: 節単位末におけるフィルター生起率

た。自己応答とは、自分の発話の直後に現れる相槌のような表現のことである。語彙的応答の中には、「あ」で始まるものが84あった。

3.2 結果と考察

図1の通り、「あ」を除外しない場合、交替時にフィルターが生起する割合(40.3%)が、継続時の場合(21.4%)よりも大きいことがわかる。すなわち、CU末における話者交替時にはフィルターが関係している可能性が示唆される。ただし、「あ」を除外すると、フィルターが生起する割合は交替と継続でそれぞれ2割程度となり差がない。

したがって、交替と継続の間でフィルターの生起に何らかの差異があるかどうかについては、生起率の全般的な比較だけでなく、2節で述べた連鎖構造の偏りを考慮して、話者の立場の違いによる比較を行う必要がある。また、頻度ではなく、形態に関する比較も必要である。

4 話者の立場の違いに基づく分析

4.1 分析対象

第2節および第3節で述べた通り、今回のコーパスでは、オペレータ(O)と被験者(S)の間で主導権や連鎖構造が非対称であると考えられる。そこで本節では、フィルターの生起位置(CU末/CU途中)および機能(交替/継続)を分析する。オペレータと被験者のそれぞれの結果を表1、表2に示す。表中の数値はフィルターの頻度を表し、括弧内の数値はそれらが全体のうちに占める割合を表している。

4.2 結果と考察

まず、生起位置に着目し、CU末で生起するフィルターとCU途中で生起するフィルターの割合を比較すると、被験者よりオペレータの方でより大きな差が見られた。2節で示した基本的な連鎖構造の観点からは、オペレータは第一位置(主に質問)での発話頻度が高くなること、またそれに応じて、第三位置で発話する頻度も高

表 1: オペレータのフィラーの生起位置および交替/継続の頻度とその割合

位置 \ 話者交替	交替 (S to O)	継続 (O to O)	計
CU 末	112 (45.8%)	90 (36.0%)	202 (80.8%)
CU 途中	12 (4.8%)	36 (14.4%)	48 (19.2%)
計	124 (49.6%)	126 (50.4%)	250

表 2: 被験者のフィラーの生起位置および交替/継続の頻度とその割合

位置 \ 話者交替	交替 (O to S)	継続 (S to S)	計
CU 末	106 (24.1%)	128 (29.1%)	234 (53.2%)
CU 途中	7 (1.6%)	199 (45.2%)	206 (46.8%)
計	113 (25.7%)	327 (74.3%)	440

くなること、逆に被験者は質問に対する第二位置での応答が多くなるということによると考えられる。この点に関連して、オペレータのフィラーは各条件のうちでCU末での交替が最も多いが、112個のうちの98個は「あ」であり、さらに後続する発話が「そうですか」などの語彙的応答であることが多かった。これは、「あ」が第三位置（応答）という特定の連鎖上の位置で用いられていることを表している。そのため、以降の分析では、「あ」は例外として除外して考える必要がある。

次に機能に着目し、交替時に生起するフィラーおよび継続時に生起するフィラーの割合を比較すると、オペレータではほぼ同じ割合で生起しているのに対して、被験者では継続が多くなっている。特に被験者については、CU途中での継続が4つの条件の中で最も多いとわかる。これも同様に、オペレータの第一位置に相当する発話の後、被験者は第二位置に相当する発話を行う頻度が高いことによると考えられる。具体的には、第一位置には質問が多く、被験者による第二位置は質問に対する回答となるため、ターンが長くなり、CU途中でも考えながら発話しているということが生起率の高さをもたらしていると考えられる。同様に、CU途中での継続の次に度数が多いのが、CU末での継続であるということからも、被験者のターンは長く、考えながら発話をしていることが示唆される。

さらに、両者に共通して、CU途中での交替は極端に少ない。CU途中での交替は、統語的意味的な切れ目の弱い箇所聞き手が発話権を獲得することを意味するが、割り込みとなる可能性も高いものであること

から、度数が少ないことは首肯できる。度数も極端に少ないため、以降の分析対象では扱わない。

5 フィラー形態ごとの生起特徴

5.1 形態の分類

前節の分析結果から、立場によって用いられるフィラーの違いがあることが示唆されたため、本節では、各フィラー形態ごとの生起位置および機能の分析を行う。フィラーの形態を表記をもとに、下記のようにまとめることによって、フィラー生成時の適切なフィラー形態の使い分けの基準を得ることを目指す。括弧は省略可能であることを表す。ただし、度数が一桁であったものは全て「その他」に分類した。

固有系：え(ー) え(ー)(っ)と

応答詞系：そうですね(ー) う(ー)ん

指示詞系：あの(ー) その(ー) こ(ー)

副詞系：ま(ー) も(ー) なんか(ー)

オペレータおよび被験者のそれぞれのフィラーについて、各形態の中で3つの条件（CU末交替、CU末継続、CU途中継続）が占める割合を図2、図3に示す。また、各条件ごとに多く出現している形態の順位と度数を、表3、表4に示す。ただし、いずれにおいても4.2節で述べた理由から「あ」は除外している。

5.2 固有系

まず、図2と図3からは、「え(ー)」は、CU末に生じやすいという傾向がオペレータと被験者とで類似していることがわかる（CU末交替+CU末継続）。しかし、各条件内での形態の順位からは、オペレータに関しては、CU末での交替および継続の度数が高いのに対して（表3）、被験者にはこの傾向は全く見られない（表4）。そこで、データを確認したところ、CU末での交替の「え(ー)」は引き伸ばしのない「え」であることがほとんどであり、これらは4.2節で述べた「あ」と同様に、語彙的応答の前につくものであった。一方、CU末で継続の場合には、引き伸ばしのある「えー」であることが多く、第一位置での質問を考えなければならないオペレータに特徴的なものであると考えられる。

次に、図2からは、「え(ー)(っ)と」はオペレータではCU末の継続の割合が高いことがわかる。これはCU末の継続での引き伸ばしのある「えー」と共通している。逆に被験者では、交替時の割合が多く（図3）、第二位置で応答する内容をターンの冒頭で考えながら発話している際に表出するものと考えられる。

表 3: オペレータのフィルターの条件別の順位と度数

条件 \ 順位	1	2	3	4	5	計
CU 未交替	え(-) [4] あの(-) [4]		ま(-) [2] その他 [2]		え(-)(っ)と [1] なんか(-) [1]	[14]
CU 未継続	え(-) [16]	あの(-) [13] なんか(-) [13]		こ(-) [11]	え(-)(っ)と [7] ま(-) [7]	[79]
CU 途中継続	あの(-) [10]	こ(-) [9]	も(-) [4]	その(-) [3] その他 [3]		[34]

表 4: 被験者のフィルターの条件別の順位と度数

条件 \ 順位	1	2	3	4	5	計
CU 未交替	ま(-) [19]	う(-)ん [13]	その他 [11]	なんか(-) [9]	え(-)(っ)と [8]	[72]
CU 未継続	ま(-) [43]	う(-)ん [17]	なんか(-) [15]	その他 [10]	その(-) [9]	[127]
CU 途中継続	ま(-) [63]	なんか(-) [39]	その(-) [22]	あの(-) [15]	その他 [14]	[196]

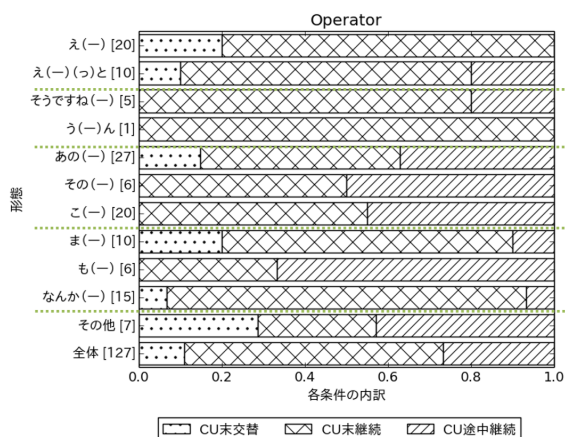


図 2: オペレータの各フィルター形態の生起特徴 (大括弧内の数字は当該形態の合計頻度を表す。)

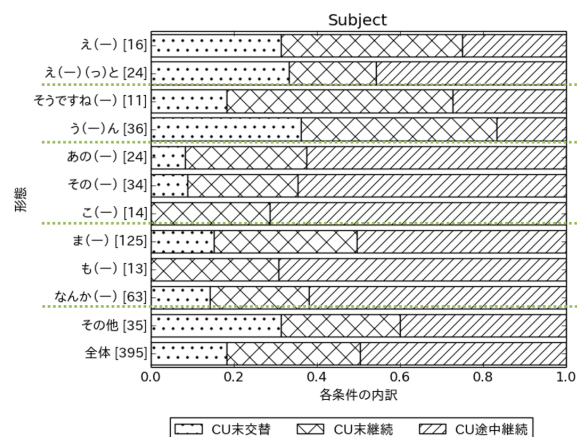


図 3: 被験者の各フィルター形態の生起特徴 (大括弧内の数字は当該形態の合計頻度を表す。)

5.3 応答詞系

図 2 と図 3 の度数からは、オペレータでは、被験者と比べて生起度数が非常に低いことがわかる。すなわち、応答としての第二位置の発話が多い被験者のフィルターが持つ特徴であると考えられる。特に「う(-)ん」は、被験者内での条件別の比較で全体より交替の割合が非常に大きく(図 3)、応答詞的な性質を持っていると考えられる。

5.4 指示詞系

「あの(-)」は、継続が多いという点で、オペレータと被験者の傾向が類似していることがわかる(図 2、図 3)。しかし、オペレータに関しては、条件別の順位で見ると CU 未での交替における度数も高いことが特徴的である(表 3)。これは、第一位置で新たな質問を開始することが多いためであり、「あの(-)」は円滑

な発話権制御のために重要な形態であると考えられる。この点については、第 6 節で改めて論じる。

次に図 2、図 3 から、「その(-)」に注目すると、オペレータ内では「あの(-)」と比べて、生起度数が非常に低いのにに対して、被験者では、「その(-)」の方が多く、さらなる分析が必要である。

最後に、「こ(-)」は、オペレータでの継続の度数が高いため(表 3)、オペレータの方に特徴的であると言える。事例では、第一位置で質問を多くするオペレータが、不足した情報を新たに付け足すことによって、質問を拡張するのに用いている場合が多い。以下に対話例を示す。

O: お好きなスポットはありますか。こ一星を見たりするのに。

5.5 副詞系

「ま(ー)」について、表3、表4の条件別順位を比較すると、継続において、オペレータではほとんど見られないこと、逆に被験者では最も多く出現していることがわかる。被験者では、特にCU途中が最も多く出現し、副詞に近いという「ま(ー)」が持つ文法的性質と一致していることがわかる。

「も(ー)」は、オペレータ、被験者とも継続のために用いられるという役割が安定して見られる(図2)。

「なんか(ー)」は、オペレータではほぼCU末での継続でのみ用いられるのに対して(図2、表3)、被験者では比較的どの条件でも多く出現している(表4)。したがって、オペレータでは隣接ペアにおける第三位置の次の連鎖の第一位置の冒頭、被験者では第二位置(交替/継続)で用いられていると考えられる。

6 適切なフィラー生成に向けて

前節の分析結果に基づき、適切なフィラー生成に向けて、生起位置と機能によって選択されやすい形態について整理する。連鎖構造上の各位置で特徴的に生起するフィラーの形態をまとめたものを図4に示す。

これまで述べてきたように、今回の対象データでは、オペレータが第一位置で質問、被験者が第二位置で応答、オペレータが第三位置で主に語彙的応答を行うというパターン化された連鎖構造が見られる。そのため、これらの位置の間での話者交替については曖昧さが比較的少ない。これに対して、第三位置の語彙的応答の直後では、(a)被験者が直前の第二位置での応答の続きを話す、(b)オペレータが第三位置での応答を拡張する、(c)オペレータが次の連鎖の第一位置となる質問を行う、といった様々な展開の可能性があり、以下の対話例のように、これらのそれぞれについて、冒頭部分に特徴的なフィラーが現れやすいと考えられる。(a)、(b)、(c)はそれぞれ図4中の記号と対応している。

- (a) 連鎖_nの第三位置 O: あ、そうなんですか。
連鎖_nの第二位置の拡張 S: なんか、この前にも見た
(O Sのターン交替) んですけど、
- (b) 連鎖_nの第三位置 O: あ、そうなんですか。
(沈黙)
連鎖_nの第三位置の拡張 O: なんか、あの、手は動か
(Oのターン継続) ないんですけどね。

- (c) 連鎖_nの第三位置 O: 食べてみたいです。
連鎖_{n+1}の第一位置 O: あのー、ところでなんで
(Oのターン継続) すけど、

また、話者交替の曖昧さは相対的には少ないと考えられるものの、この位置以外にも、第一位置での質問が拡張される「こ(ー)」が用いられるといったことも観察される(5.4節)。

さらに、より一般的な現象として、交替のために生じたフィラーによって話者交替が起きなかった対話例を以下に示す。括弧内はオーバーラップの生じた箇所を表す。

- S: 踊りたいダンスとかありますか。
(フォークダンスとか)。社交ダンスとか。
O: (えーとですね)

ここでは、被験者による質問の拡張とオペレータによる応答の開始との間で同時発話が生じているが、オペレータの発話の冒頭がフィラー以外の言語的内容をもつ要素の場合と比べ、フィラーでの重複は問題が少ないと考えられる。

さらに、交替が自明な位置でも、フィラーが先行発話の末尾と重複する現象が多く観察された。こうした箇所におけるフィラーの重複も問題の少ないものであり、自分の発話の開始タイミングを調整していると考えられる。今後は、重複している区間の韻律的特徴や持続時間の分析も必要である。

7 今後の課題

前節では、フィラーによる発話権制御について、フィラーの生起位置の観点から検討を行ったが、発話権制御をリアルタイムで行うことを考えると、生起の位置だけでなく、タイミングの分析も必要である。そのため、先行発話末からフィラーの生起までの無声休止区間の持続時間や先行発話末から次の実質的な発話の生起までの持続時間を分析することによって、この区間の中での適切なフィラーの生起タイミングを特定することが課題である。

また、フィラーの話者交替における役割を明らかにするためには、交替でフィラーが生起しない場合についても分析する必要がある。第3節で分析したように、話者交替時にフィラーが生起する割合は、CU末では約40%であるため、フィラーが生じていない交替箇所でも何が起こっているかも分析しておく必要がある。そこで、これらの箇所を確認したところ、被験者による第二位置での冒頭では、先行発話内容の繰り返しが多く見られた。この連鎖上の位置では、話者交替がある程度自明であるため、繰り返しはフィラーと同様の場繋ぎ的な役割を果たしていると考えられる。さらに、繰り返し

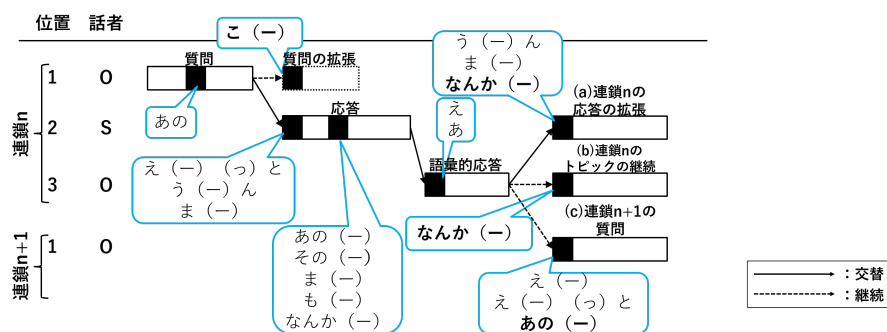


図 4: 連鎖構造上のフィラーの生起位置と形態
(太字の形態は本文中で対話例を紹介しているもの)

しの直後にはフィラーの生起が多く観察された。以下に対話例を示す。

O: 運動 は得意ですか。

S: 運動。まー、得意ではないですね。

同様に、この位置では語彙的応答も多く出現していた。これらは、フィラーの生起可能性がある箇所であり、フィラー以外の表現を用いて間をもたせているものであると考えられるが、このような事例に関して、フィラーとの比較分析が必要になると考えられる。

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.
- [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.
- [7] H. Brugman, A. Russel. Annotating Multimedia/ Multi-modal resources with ELAN. In Proc. LREC, pp. 2065-2068, 2004.
- [8] 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.
- [9] H. Sacks, E. A. Schegloff and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. Language, Vol. 50, No. 4, pp. 696-735, 1974. (西阪仰 (訳) 「会話のための順番交替の組織 - 最も単純な体系的記述」. 『会話分析基本論集: 順番交替と修復の組織』世界思想社, 2010, 5-153.
- [10] 榎本美香. ターン構成単位, 坊農真弓, 高梨克也 (編), 多人数インタラクションの分析手法. オーム社, pp. 68-81, 2009.
- [11] 高梨克也, 内元清貴, 丸山岳彦. 『日本語話し言葉コーパス』における節単位認定. 『日本語話し言葉コーパス』同梱マニュアル, 2004.
- [12] 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.
- [13] 泉子・K・メイナード. 会話分析. くろしお出版, 1993.
- [14] 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.
- [15] 伝康晴. 対話への情報付与. 小磯花絵 (編), 講座日本語コーパス 3: 話し言葉コーパス-設計と構築-, 朝倉書店, pp. 101-130, 2015.