

# 傾聴対話システムのための発話を促す聞き手応答の生成

## Generation of Responses Encouraging User Utterances in Attentive Listening System

石田真也\* 井上昂治 中村静  
高梨克也 河原達也

Masanari Ishida 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:** We investigate generating various and proper responses of not only backchannels but also questions for elaboration, partial repeat, and formulaic responses to improve dialogues by encouraging users to talk more in attentive listening. The proposed system selects responses among a question for elaborating the focus word and a partial repeat for the focus word. On the other hand, when there is not any focus words, the system selects responses among a question for elaborating the predicate and a formulaic response. To generate elaborating questions, classifiers are designed to determine the possible interrogative words for the focused word or the predicate, and features used are N-gram probability of the pair of an interrogative word and the focus word, or the pair of an interrogative word and the predicate. In computing the N-gram probability of compound nouns, two methods are compared by considering the modification structure. Finally, we evaluate the system using test data to show the effectiveness of the methods.

## 1 はじめに

近年、人と対話を行うシステムが日常生活においても盛んに利用されるようになってきている。対話の種類には、タスク遂行型対話と非タスク遂行型対話がある。従来は前者が取り上げられることが多かったが、近年、後者を対象とした研究も進められている [1][2]。非タスク遂行型対話の一つに、ユーザの話の聞き役となる傾聴対話 [3][4] がある。システムが傾聴対話を行うことにより、話し相手としてユーザの話を聞いて欲しいという欲求をみたすなどの効果が期待されている。

傾聴対話を実現するためには、ユーザがより多く発話するように促すふるまいをシステムが行う必要がある。これまで、システムの相槌に焦点があてられてきた。システムが相槌をうつ場面において、相槌の適切な韻律 [5] や適切な形態・タイミング [6] を決定することなどが取り組まれてきた。これらの研究では狭義の相槌による応答のみを対象としているが、傾聴対話においては、相手の発話を促すためにより多様な種類の応答を行うべきである。

そこで本研究では、相槌以外に、傾聴対話を実現するためのより多様な聞き手応答の生成を検討する。具体的には、ユーザの発話内容に応じて、慣用的表現で同意を示す「語彙的応答」、先行発話の一部を繰り返す「繰り返し応答」、先行発話の詳細を問う「掘り下げ質問」を行う。伝ら [7] によると聞き手応答は6種類とされているが、提案手法ではこのうちの「語彙的応答」と「繰り返し」、これに加えて「掘り下げ質問」を生成することを目指す。ユーザ発話とその焦点、およびシステム応答の例を図1に示す。これらの多様な聞き手応答の中から、与えられたユーザ発話に適切なものを自動的に選択する。

## 2 関連研究

本章では、傾聴対話に関する関連研究を紹介する。

### 2.1 聞き手応答

相槌については様々な定義がなされているが、ここではメイナード [8] の「話し手が発話権を持っている発話内で、話し手の発話に対して聞き手が発する、発話権の移動を伴わない発話」という定義を採用する。相槌の特徴として、打つタイミング、形態、韻律などが

\*連絡先：京都大学 大学院情報学研究科  
知能情報学専攻 河原研究室  
〒 606-8501 京都市左京区吉田本町  
E-mail:ishida@sap.ist.i.kyoto-u.ac.jp

- A** 焦点に関する掘り下げ質問  
 ユーザ「ゲームは気分転換にはもってこいです」  
 焦点:ゲーム  
 システム「どんなゲームですか？」
- B** 繰り返し応答  
 ユーザ「アマゾンから東野圭吾の新刊が着くんですよ」  
 焦点:東野圭吾  
 システム「東野圭吾ですか」
- C** 述語に関する掘り下げ質問  
 ユーザ「昼はサンドイッチとおにぎりを食べました」  
 焦点:なし  
 システム「どこで食べましたか？」
- D** 語彙的応答  
 ユーザ「何事も喜んで取り組む姿勢が大切だと思います」  
 焦点:なし  
 システム「そうですね」

図 1: システム応答例

ある。これらの要素によって相槌が持つ機能や意味役割が変化する。

伝ら [7] は狭義の相槌だけでなく、これに機能的に類する表現をも含めた「聞き手応答」を対象として、それらを形態と生起位置にもとづいて分類する手法を提案している。具体的には、聞き手応答をその形態によって、応答系感動詞、感情表出系感動詞、語彙的応答、評価応答、繰り返し、共同補完の 6 つに分類した。次に聞き手応答の生起位置によって、発話ターン途中、第 1 部分、第 2 部分、第 3 部分、その他の応答位置、分類不能の 6 つに分類した。本稿で対象とする聞き手応答の生起位置は一般的な相槌の生起位置である「発話ターン途中」ではなく、「第 2 部分」や「第 3 部分」に相当する。

## 2.2 傾聴対話におけるシステム応答

相槌による応答は、傾聴対話におけるシステム応答で重要な役割を果たすものの一つである。傾聴対話を対象として、相槌の生起パターンの分析を行い、システムが生成すべき相槌の韻律的、形態的、タイミングの特徴を予測することを試みた研究もある。上里ら [5] は相槌の韻律、山口ら [6] は相槌の形態・タイミングについて研究を行っている。適切な相槌の生成は傾聴システムに必要であるが、それらのみで十分に適切な傾聴を行えるわけではない。[9] によると、傾聴に適しているカウンセラー行動は以下である。

1. 語られたことを繰り返す
2. もっと語るように問い返す
3. 話し手に共感し、気持ちを言葉にする
4. 相槌などにより聴いていることを示す

## 5. 語られた内容を要約する

これらのうち [5][6] の対象とされているのは 4 のみである。本研究では、上記の 1, 2, 4 をそれぞれ本稿では、繰り返し応答、掘り下げ質問、語彙的応答、と定義し、それらを実現することを目標とする。下岡ら [3] のシステムでは、上記の 1, 2, 3 にそれぞれ相当する「問い返し応答」、「繰り返し応答」、「共感応答」などが生成される。しかし、ユーザ発話に対して焦点解析を行わずに応答の種類を選択が行われており、想定外のユーザ発話に対する頑健性が欠如している。そこで本研究では、焦点解析を導入するとともに応答の種類を選択を統計ベースで行うことで頑健な応答生成を実現する。

## 3 システムの処理の流れ

本研究で提案する処理の流れを説明する。処理の概要を図 2 に示す。はじめに、ユーザ発話文に対して、形態素解析・述語項構造解析を行う。形態素解析には Chasen<sup>1</sup> を使用する。次に焦点解析は 3.1 節で示した手順で行う。焦点解析の結果、焦点が存在すると判定された場合、その内容語（名詞または名詞句）を抽出した上で、焦点に関する掘り下げ質問と繰り返し応答のどちらが適切かを N-gram 確率を用いて判定する（図 2 の焦点による判別器）。焦点がない場合には、述語に関する掘り下げ質問と語彙的応答のどちらが適切かを N-gram 確率を用いて判定する（図 2 の述語による判別器）。最後に、選択された種類の応答生成モジュールで応答文を生成し出力する。

### 3.1 焦点解析

傾聴対話においては、相手の発話を促すために、相手発話内の単語から焦点を抽出し、その焦点から話題をずらさずに応答をする必要がある。本研究では、繰り返し応答、掘り下げ質問の生成過程で焦点解析を行う。吉野ら [10] は Conditional Random Fields (CRF) を用いて焦点解析する手法を提案している。CRF の素性として品詞、各情報、出現順序などの 9 つの特徴量が用いられる。当該手法の文節レベルでの精度は 78.5%、文全体での焦点の有無の精度は 99.9% と報告されている。

### 3.2 掘り下げ質問、応答の生成

焦点解析の結果に基づいて別々の処理を適用する。本稿では、焦点および述語とそれぞれの疑問詞との N-

<sup>1</sup><http://chasen-legacy.osdn.jp/>

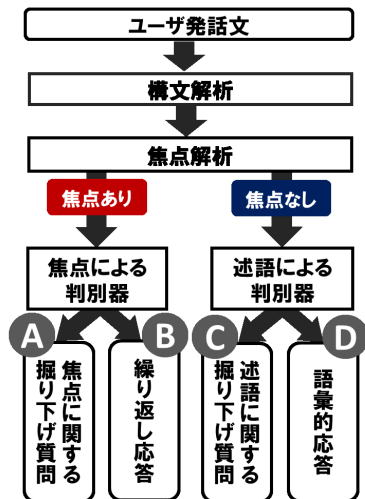


図 2: システム概要

gram 確率に基づいて応答発話の種類を適切さを判定する方法を提案する

### 3.2.1 焦点がある場合

焦点解析によって抽出したユーザ発話内の焦点となる単語を入力として、その焦点に関する掘り下げ質問を生成するのに適した疑問詞、もしくは疑問詞なしを選択する。当該判別器で候補とする疑問詞は 11 種類で次のように 4 つのカテゴリに分類される。

- ・どんな/どの/なんの/なにの/何の
- ・どこ/何処
- ・いつ/何時
- ・だれの/誰の

各カテゴリ毎で N-gram 確率の総和を計算し、その値が閾値以上となるカテゴリが存在する場合、その中で最大のものの代表的な疑問詞（各カテゴリの先頭に記述されている疑問詞）と焦点となる単語のペアから「A. 焦点に関する掘り下げ質問」を生成する。閾値以上となる疑問詞が存在しない場合は焦点となる単語のみで「B. 繰り返し応答」を生成する。

### 3.2.2 焦点がない場合

焦点が抽出されなかった場合は、ユーザ発話内の述語から、その述語に関する掘り下げ質問を生成するのに適した疑問詞、もしくは疑問詞なしを選択する。当該判別器で候補とする疑問詞は 33 種類で次のように 17 個のカテゴリに分類される。

- ・何を/なにを
- ・何に/なにに
- ・何から/なにから
- ・何まで/なにまで
- ・どこで/何処で
- ・どこへ/何処へ
- ・どこから/何処から
- ・どこまで/何処まで
- ・どこに/何処に
- ・誰を/だれを
- ・誰に/だれに
- ・誰から/だれから
- ・誰へ/だれへ
- ・誰と/だれと
- ・いつ
- ・いつまで/何時まで
- ・いつから/何時から

前節と同様に各カテゴリ毎に N-gram 確率の総和を求め、閾値を超えたカテゴリがある場合、最大のカテゴリの代表的な疑問詞と述語のペアから「C. 述語に関する掘り下げ質問」を生成する。どのカテゴリも閾値を超えない場合、「D. 語彙的応答」を生成する。

### 3.2.3 N-gram 確率の計算方法

本研究で用いる N-gram モデルは、Julius ディクテーションキット<sup>2</sup>に付属されているもので、現代日本語書き言葉均衡コーパス (BCCWJ) の全データから学習されている。疑問詞が 1 つの形態素からなるときは bi-gram 確率、2 つの形態素からなるときは tri-gram 確率を用いる。

また、焦点が複合名詞 (ex. 「グリル料理」) である時、単純に「疑問詞+複合名詞」で bi-gram 確率、tri-gram 確率を求めようとしても、そもそも複合名詞（「グリル」と「料理」のように）は分割されて辞書に登録されているため、ヒットすることはない。疑問詞 (x) と複合名詞 (a.b) との N-gram 確率  $P(a.b | x)$  として以下の 2 通りが考えられる。

$$P(a.b | x) = P(b | x) * P(b | a) \quad (1)$$

$$P(a.b | x) = P(a | x) * P(b | a) \quad (2)$$

本来「x a.b」という系列に対する N-gram 確率は (2) 式で与えられるが、実際に x は a ではなく b にかかる場合が多いと考えられるので、本研究では (1) 式と (2) 式を比較検討する。

## 4 評価実験

3 節で述べた通り、焦点解析は吉野ら [10] の手法を用いるため、掘り下げ質問や応答の生成に関して焦点解析結果は所与であるとし、評価実験を行う。2 つの判別器は互いに独立に評価を行う。評価の指標として用いるのは、再現率と適合率である。ある応答の種類 Y における再現率と適合率は以下で計算される。

<sup>2</sup><https://github.com/julius-speech/dictation-kit>

表 1: 評価データセット例

発話番号	ユーザ発話	焦点有無	焦点/述語	疑問詞番号	疑問詞	応答例
1	私は技術書ばかり読んでいます	1	技術書	1	どんな	どんな技術書ですか？
2	海辺で採りたての貝を食べたい	1	貝	1	どんな	どんな貝ですか？
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	私はピザが好きです	1	ピザ	3	どこの	どこのピザですか？
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	私はアップテンポな曲が好きです	1	曲	4	誰の	誰の曲ですか？
⋮	⋮	⋮	⋮	⋮	⋮	⋮
26	実は今日から九州に行くのです	1	九州	5	x	九州ですか
⋮	⋮	⋮	⋮	⋮	⋮	⋮
62	昼はサンドイッチとおにぎりを食べました	0	食べる	10	どこで	どこで食べましたか？
⋮	⋮	⋮	⋮	⋮	⋮	⋮
70	最近暑いから水につかると気持ちいいです	0	気持ちいい	23	x	そうですか
⋮	⋮	⋮	⋮	⋮	⋮	⋮

$$\text{再現率} = \frac{\text{応答の種類 } Y \text{ の例文におけるデータセット内の正解疑問詞と判別器が選択した疑問詞との一致数}}{\text{データセット内の応答の種類 } Y \text{ の例文の数}}$$

$$\text{適合率} = \frac{\text{応答の種類 } Y \text{ の例文におけるデータセット内の正解疑問詞と判別器が選択した疑問詞との一致数}}{\text{判別器が応答の種類 } Y \text{ としてラベリングした例文の数}}$$

## 4.1 評価用データセット

本稿の手法で生成される聞き手応答を評価するためのデータを 200 例作成した。このうち 100 例（データセット  $\alpha$ ）は手法の開発とチューニングの際にも利用し、残りの 100 例（データセット  $\beta$ ）は完全な評価用とする。以上のことから、実験では、データ  $\alpha$  とデータ  $\beta$  を分けて評価を行う。

評価データの作成には雑談対話コーパス [11] を用いた。このコーパスからユーザ発話を 200 例抽出した。抽出基準は、コーパス内のユーザ発話のうち、発話の内容が客観的事実、自身の経験、自身の感情に関するもので構成されており、述語を含むものとする。次に、抽出した 200 例のユーザ発話それぞれに対して「焦点の有無」、「焦点となる単語（もしくは述語）」、「焦点（述語）と相性のよい疑問詞」のアノテーションを行う。この焦点となる単語の選択は、ユーザ発話に含まれる名詞または複合名詞の中で、ユーザ発話内の述語の要素であることや、出現頻度が低いものであることなどから総合的に判断している。この条件を満たすような名詞がない場合は、焦点なしとする。適切な疑問詞の選択は、焦点（述語）と相性がよいと思われる疑問詞を候補の中から、文脈を考慮して行う。適切な疑問詞がない場合は、繰り返し応答（焦点ありの場合）か語彙的応答（焦点なしの場合）を生成する。

上記の手順で作成した評価データの一部を表 1 に示す。

## 4.2 焦点ありの場合の評価

3.2.1 節で述べたように、焦点に関する判別器は、ユーザ発話に焦点がある場合を対象として、A. 焦点に関する掘り下げ質問か、B. 繰り返し応答のいずれかを生成する。

この判別の問題は、評価データに記述されている各焦点の単語に対して、応答で使用するのにふさわしい疑問詞の種類のラベルをつける問題とみなすことができる。そこで、判別器によって決定されたラベルが評価データのラベルと一致する場合を正解とする。

本稿では、複合名詞を一つの名詞としてそのまま N-gram 確率を求める場合（以下、複合名詞非分割）と、3.2.3 節で述べた式のように分割して求める場合（式 (1) の分割、式 (2) の分割）の 3 通りで実験を行う。

データセット  $\alpha$  に含まれる 100 文のうち焦点があるものは 59 個、データセット  $\beta$  に含まれる 100 文のうち焦点があるものは 66 個であった。判別器の閾値については、掘り下げ質問の再現率を大きくするためにほとんど 0 に近い値に設定したが、それによる適合率の著しい低下は見られなかったため、その値を閾値として最適なものとした。また、データセット  $\alpha$ 、 $\beta$ 、 $\alpha + \beta$  の 3 通りそれぞれにおいて評価実験を行い比較を行ったが結果に有意な差は見られなかった。データセット  $\alpha + \beta$  を用いて複合名詞の処理方法を変化させた結果を表 2 に示す。

まず、複合名詞をどちらの方法で分割しても、掘り下げ質問の再現率が 10% 以上改善した。この結果より、複合名詞分割の処理は有効であることがわかった。

ただし、掘り下げ質問の再現率自体は必ずしも十分に高いとはいえない。この理由として、そもそも N-gram 確率を求めるためのコーパスに出現しない名詞（ex. 「ヨット」、「お化け」、「かき氷」等）は N-gram 確率が求められないので、N-gram 確率のみの素性では

表 2: 焦点ありの場合の実験結果 (データセット  $\alpha + \beta$ :125 サンプル)

応答の種類	複合名詞	再現率	適合率
A 焦点に関する掘り下げ質問	式 (1) の分割	29/52 = <b>56%</b>	29/46 = <b>63%</b>
	式 (2) の分割	30/52 = <b>58%</b>	30/52 = <b>58%</b>
	非分割	24/52 = <b>46%</b>	24/38 = <b>63%</b>
B 繰り返し応答	式 (1) の分割	61/73 = <b>84%</b>	61/79 = <b>77%</b>
	式 (2) の分割	58/73 = <b>79%</b>	58/73 = <b>79%</b>
	非分割	63/73 = <b>86%</b>	63/87 = <b>72%</b>

表 3: 焦点なしの場合の評価実験結果 (データセット  $\alpha + \beta$  で 75 サンプル)

応答の種類	データセット	再現率	適合率
C 述語に関する掘り下げ質問	$\alpha$	2/5 = <b>40%</b>	2/14 = <b>14%</b>
	$\beta$	1/5 = <b>20%</b>	1/7 = <b>14%</b>
	$\alpha + \beta$	3/10 = <b>30%</b>	3/21 = <b>14%</b>
D 語彙的応答	$\alpha$	27/36 = <b>75%</b>	27/27 = <b>100%</b>
	$\beta$	24/29 = <b>83%</b>	24/27 = <b>89%</b>
	$\alpha + \beta$	51/65 = <b>78%</b>	51/54 = <b>94%</b>

不十分であることが考えられる。

### 4.3 焦点なしの場合の評価

3.2.2 節で述べたように、述語に関する判別器は、ユーザ発話に焦点がない場合を対象として、C. 述語に関する掘り下げ質問と、D. 語彙的応答のいずれかを生成する。この判別の問題は評価データの各文中の述語に対して、応答に使用するのにふさわしい疑問詞の種類をラベルを付与する問題と捉えることができる。そこで、判別器によって選択された疑問詞の種類が評価データのラベルと一致する場合を正解とする。

結果を表 3 に示す。判別器での N-gram 確率の閾値は、データセット  $\alpha$  で最適なものを探し、データセット  $\beta$  にも適用する。具体的にはデータセット  $\alpha$  において出現した特定の述語と疑問詞との N-gram 確率が閾値をクリアできる最大値に設定している。これより値を小さくしても、掘り下げ質問の再現率は上がらなかったことと、適合率が著しく下がっていったことにより、この閾値を最適とした。データセット  $\alpha, \beta$  において焦点なしと認定されたものはそれぞれ 41 個、34 個であった。

掘り下げ質問に関して、再現率と適合率のいずれも低く、あまりよい結果を得られなかった。掘り下げ質問自体は多数 (21 個) 生成されていることから、評価データの作成者が想定した適切な疑問詞の順位と、判別器が N-gram 確率によって決定する疑問詞の順位と

表 4: 述語に対する掘り下げ質問の疑問詞抽出結果

○	「見る」、「遊ぶ」、「終わる」
△	「やる」、「食べる」、「行く」、「する」
×	「かける」、「干す」、「塗る」

が一致しないことが考えられる。また、評価データでは述語だけではなくユーザ発話全体を考慮に入れて疑問詞が選択されているのに対して、判別器は文脈を考慮せず述語のみの情報で疑問詞を選択していることもその要因と考えられる。

そこで、評価データセットにおいて掘り下げ質問が適切と判断された 10 個の述語に関して、その述語と疑問詞のペアの N-gram 確率の情報に基づいて、正解となる疑問詞を (順位はともあれ) 抽出できているのかを調べた。その結果を表 4 に示す。「△」に分類された述語に関しては、適切な疑問詞が唯一でないことから、不正解と判別されたと考えられる。この改善策として、評価データを作成する際に、正解となる疑問詞を複数個選択することが挙げられる。「×」に分類された述語に関しては、使用頻度の低さからそもそも N-gram 確率が登録されていないことが不正解の原因と考えられる。

なお、語彙的応答の再現率・適合率や、合計の再現率・適合率は必ずしも低いわけではない。すなわち、気

の利いた応答はできていないが、対話の破綻の頻度は必ずしも高くないともいえる。多様で気の利いた聞き手応答である掘り下げ質問とより無難な語彙的応答との最適な比率については今後の課題としたい。

## 5 おわりに

本稿では、傾聴対話システムのために相槌だけでなく、焦点に関する掘り下げ質問、繰り返し応答、述語に関する掘り下げ質問、語彙的応答という4種類の聞き手応答を生成する方法を検討した。これらのうち掘り下げ質問については、ユーザ発話内の焦点となる単語の有無と、焦点となる単語またはユーザ発話内の述語と疑問詞とのN-gram確率とをもとに応答文を生成する手法を提案し、その精度を評価した。その結果、焦点がある場合の掘り下げ質問に関しては複合名詞分割の処理を行うことで精度が改善した。焦点がない場合の掘り下げ質問に関しては、精度は高くなかったが、語彙的応答で補うことができた。

## 謝辞

本研究は、JST ERATO 石黒共生ヒューマンロボットインタラクションプロジェクトの支援を受けて実施されたものである。

## 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, pp. 45–51, 2013.
- [2] 東中竜一郎. 傾聴対話システムに向けた取り組み. 人工知能学会 言語・音声理解と対話処理研究会, Vol. 70, pp. 65–70, 2014.
- [3] 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. 言語・音声理解と対話処理研究会, Vol. 58, pp. 61–66, 2010.
- [4] 横山祥恵, 山本大介, 小林優佳, 土井美和子. 高齢者向け対話インタフェース：雑談継続を目的とした話題提示・傾聴の切替式対話法. 情報処理学会研究報告 音声言語情報処理研究会 (SLP), Vol. 2010-SLP-80-4, , 2010.
- [5] 上里美樹, 吉野幸一郎, 高梨克也, 河原達也. 傾聴対話における相槌の韻律的特徴の同調傾向の分析. 言語・音声理解と対話処理研究会, Vol. 70, pp. 7–13, 2014.
- [6] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, Nigel Ward, 河原達也. 多様な相槌をうつ傾聴対話システムのための相槌形態の予測. 人工知能学会研究会資料, Vol. 75, pp. 1–6, 10 2015.
- [7] 伝康晴. 対話への情報付与. 小磯花絵 (編), 日本語コーパス 3: 話し言葉コーパス-設計と構築-, pp. 101–130. 朝倉書店, 東京, 2015.
- [8] 泉子・K・メイナード. 会話分析. くろしお出版, 1993.
- [9] 三島徳雄, 久保田進也. 積極傾聴を学ぶ-発見的体験学習法の実際. 中央労働災害防止協会, 2003.
- [10] 吉野幸一郎, 河原達也. Conversational system for information navigation based on pomdp with user focus tracking. *Computer Speech Language*, Vol. 34, pp. 275–291, 2015.
- [11] 東中竜一郎, 船越孝太郎. Project next nlp 対話タスクにおける雑談対話データの収集と対話破綻アノテーション. 人工知能学会 言語・音声理解と対話処理研究会, Vol. 72, pp. 45–50, 2014.