

WWW を介した対話システムとの対話における ユーザの振舞いについて

熊 本 忠 彦[†] 伊 藤 昭^{††}

対話は対話者同士による協同作業であるため、発話の形式や内容だけでなく、対話の進め方なども話し相手によって大きく異なる。しかしながら、機械である対話システムとの対話においてユーザがどのように振る舞うのか、ヒューマンファクタに関する解析は十分とはいえず、頑健な対話システムを開発する際の妨げとなっている。対話システムにユーザとの対話を通して何らかの課題を解かせ、その課題を解くまでの対話量を競うというコンテスト (DiaLeague'97) が WWW (world-wide web) ページを介して行われた。本稿では当コンテストで得られた対話 (728 対話) のうち、最も頑健であった対話システムとユーザとの対話 (141 対話) を中心に分析し、機械である対話システムとの対話においてユーザがどのように振る舞うのか、特に対話システムの頑健性に関する要因に焦点を当て調べた。その結果、(1) ユーザはデス・マス調で発話する、(2) ユーザは間接的な発話形式を採用することがある、(3) ユーザ発話数の増加にともない、異なり形態素数は対話の順番に関係なく増えているが、発話パターン数はユーザ先手のときの方がハイペースに増えている、(4) ユーザはシステムの発話パターンをまったくそのままの形で再利用しない、(5) ユーザは文脈から外れた予想外のシステム発話に対しても好意的に振る舞う、(6) 「わかりません」というシステム発話やシステム発話の反復に対して、ユーザは、自分の発話を修正して言い直すことよりも、発話内容そのものを変え、まったく別のことを発話する方を好む、(7) 発話理解や対話処理に失敗したときにはシステム発話の繰返しという対話戦略が有効であるといったことが分かった。

An Analysis of Dialogues between Dialogue Systems and Users through a WWW Page

TADAHIKO KUMAMOTO[†] and AKIRA ITO^{††}

A dialogue is a collaboration between dialogue participants. Therefore, identification of a conversational partner influences not only the form and contents of an utterance but also the context and expansion of a dialogue. The human factors in a man-machine dialogue, however, are not obvious enough to understand with regard to how people talk with a dialogue system.

DiaLeague'97 was the second dialogue contest in which a natural language dialogue system engaged in a dialogue with a human to solve a specific problem. Each of the dialogue systems that participated in the contest obtained a score according to the amount of dialogue with a contest participant. This contest was held on the WWW (world-wide web) pages for one week, and the five dialogue systems had 728 dialogues with Internet users. We analyzed mainly the 141 dialogues between the users and the robustest dialogue system, and investigated the dialogues at the utterance and dialogue levels. As the results, we found the followings: (1) users talked to a dialogue system in a polite manner, (2) users did not always make a sentence using the direct speech, (3) the number of different words increased similarly either in user first or system first, and the sentence patterns observed in a dialogue were richer in variety when the dialogue began with a user question or request, (4) users were not influenced by system sentence patterns in making a sentence, (5) users did not ignore an unexpected system utterance, (6) users preferred to change the contents of their utterance rather than to express it differently when the dialogue system said "I don't understand," and repeated a system question, and (7) dialogue confusion which was caused by the failure of spoken language understanding and dialogue processing was often recovered by repeating a system question.

1. はじめに

自然言語による対話は、人間にとって最も自然かつ容易な意志伝達手段であり、自分の信念や意図、外界の状況などを他者に伝えるのに適している。それゆえ、

[†] 郵政省通信総合研究所

Communications Research Laboratory, MPT

^{††} 山形大学工学部

Faculty of Engineering, Yamagata University

自然言語対話を実現する対話システムが人に優しいマン・マシンインタフェースとしてさかんに研究されている^{1)~9)}。しかしながら、機械である対話システムとの対話においてユーザがどのように振る舞うのか、ヒューマンファクタに関する解析は十分とはいえず、頑健な対話システムを開発する際の妨げとなっている。

対話は対話者同士による協同作業であるため、発話の形式や内容だけでなく、対話の進め方なども話し相手によって大きく異なる。このようなヒューマンファクタの解析を目的とした対話の収集には、主として Wizard of Oz 法によるもの^{10)~15)}と実動の対話システムを利用するもの^{16)~19)}とがある。前者の方法では、人間が対話システムに成りすまして被験者と対話するため、人間同士の対話に近い高度な対話を収集できるというメリットがある反面、実動の対話システムとの対話では生じるであろう様々な現象（発話理解や対話処理の失敗など）が対話に及ぼす影響を調べることができない。また、文献 10) において「ユーザはシステム使用の動機が十分でなく、実システムの使用時には考えられない勘違いも見られ」と述べられているように、被験者として集められた少数のユーザがシステムを実際の場面において利用するであろうユーザと同じ振舞いをするとは限らない。一方、後者の方法、すなわち実動の対話システムを利用して対話を収集する方法では、音声認識能力の限界から簡単なタスクしか設定できないため、ユーザ発話の音声言語的特徴（間投表現や言い淀み、言い直しなど）を分析するのには十分だが、ユーザの振舞いを解析できるほど多様な対話データを収集することができない。

そこで本稿では、対話コンテスト DiaLeague'97²⁰⁾ で得られた対話を分析対象とした。DiaLeague'97 は、対話システムの性能を総合的、客観的、定量的に評価するための 1 つの試みとして実施されたコンテストであり、あらかじめ参加登録された対話システムが人と自然言語（日本語）で対話し、与えられた経路課題を解くまでの対話量（得点 = 300 - 発話文節数の総和）を競うというものであった。このコンテストの特徴は対話システムとの対話を WWW（world-wide web）ページを介して実現したという点にあり、その結果、これまでの対話システムとの対話が被験者として集められた少数ユーザによる限定的なものであったのに対し、今回のコンテストで得られた対話は不特定多数の一般ユーザによる実対話であるといえる。なお、当コンテストにおいて、対話システムへの入力（ユーザ発話）はキーボードなどによるテキスト入力であり、対話シ

ステムからの出力（システム発話）は画面へのテキスト表示であった。そのため、ユーザ発話の特徴を音声レベルで分析することはできないが、その反面、テキストレベルでの分析や対話過程の分析には適しているといえる。

本稿では当コンテストで得られた対話（728 対話）のうち、最も頑健であった対話システムとユーザとの対話（141 対話）を中心に分析し、機械である対話システムとの対話においてユーザがどのように振る舞うのか調べる。特に、対話システムの頑健性に関する要因に焦点を当て、発話レベルの分析としては、(1) ユーザがどのような文形式や文末表現を用いて対話システムと話すのか、(2) ユーザ発話の多様性はどの程度なのか、(3) ユーザ発話はシステム発話の文形式や発話パターンにどの程度影響されるのかを調べ、対話レベルの分析としては、(4) 対話システムが文脈から外れた発話をしたり、相手の質問を無視して、発話したりしたときに、ユーザがどのように反応するのか、(5) システム発話が繰り返されたときに、ユーザの発話はどうか変わるのかを調べる。

2. 対話システムとユーザとの対話の獲得

本章では対話コンテスト DiaLeague'97 の概要を述べ、当コンテストで得られた対話システムとユーザとの対話の中から分析対象とする対話を選定する。具体的には、各対話システムの頑健性を調べ、最も頑健であった対話システムとユーザとの対話を分析対象とする。

2.1 対話コンテスト DiaLeague'97 の概要

DiaLeague'97 は、対話システムの性能を総合的、客観的、定量的に評価するための試みとして実施されたコンテストであり、1997 年 9 月 25 日正午から 10 月 2 日正午までの丸 1 週間開催された²⁰⁾。今回登録された対話システムは全部で 5 つあり、対話システムにアクセスしたユーザの延べ人数は 728 人であった。

このコンテストで両者に与えられた課題は経路課題と呼ばれるもので、いずれの対話者も正解を知らない、両者が互いに情報を伝え合って正解を見いださなければならないという相互型の課題であった。具体的には、対話システムとユーザのそれぞれに与えられた微妙に異なる鉄道路線図から両方の路線図に共通な経路を見つけるというものであった。すなわち、双方の路線図において駅の個数や配置、出発駅（スタート）、目的駅（ゴール）は同じであったが、線路のつながり方（結線パターン）や名称不明な駅の位置が異なっており、両者に共通な経路（正解経路）は 1 通りしかなかった。

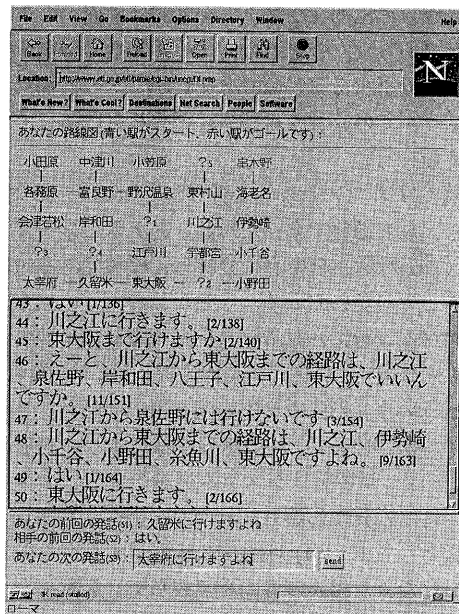


図1 WWW ページを介した対話の様子

Fig. 1 Man-machine dialogue through a WWW page.

コンテストは、ネットニュースやメーリングリストを通じて広くアナウンスされ、WWW ページ上で開催された。ユーザはまず対話サーバ (<http://www.etl.go.jp/etl/nl/dialeague/webdialogue/>) にアクセスし、エントリー名 (本名/メールアドレスである必要はない) を登録した。すると、対話サーバは、相手をする対話システム、対話の順番 (先手/後手)、課題に使用する路線図を適当に決め、図1に示すような3つのフレームからなる WWW ページを構成した。この WWW ページの一番上のフレームにはユーザ用の路線図が表示されており、右上の角にある駅が出発駅 (スタート)、左下の角にある駅が目的駅 (ゴール) となっている。また、記号「?」で始まる駅名は、その駅の名前が不明であることを示している。真ん中のフレームにはこれまでの対話の内容が表示されており、いつでも参照できる。一番下のフレームにはユーザ発話入力用のボックスがあり、キーボードによる日本語入力、マウスによるカット&ペーストを受け付ける。なお、発話入力は改行 (もしくは send ボタンを押すこと) によって終了したため、1 入力が 1 文とは限られていなかった。そのため、複数の文を一度に入力することもできたが、対話の順番 (ターン) は交互に割り当てられるよう決められており、割込みなどは許されなかった。

対話の得点は、

得点 = 300 - 発話文節数の総和
という計算式で求められ、ユーザと対話システムに同

じ得点が与えられた。なお、発話文節数は、対話サーバに組み込まれた形態素解析システムによって自動的にカウントされた。また、発話文節数の総和が 300 を超えたときは、その対話は打ち切られ、両者に得点 0 が与えられた。

以上のコンテストの結果、筆者らの対話システムが平均得点 123.29 点 (対話回数 141 回) を獲得し、第 1 位の成績を得た。第 2 位の対話システムで平均得点は 104.30 点 (185 回) であり、以下、第 3 位 98.38 点 (175 回)、第 4 位 37.50 点 (167 回)、第 5 位 20.73 点 (60 回; 中途参加) であった。なお、当コンテストではユーザの成績ランキングも求められており、優勝者の平均得点は 194.71 点 (対話回数 14 回) であった。以下、第 2 位 193.40 点 (10 回)、第 3 位 172.63 点 (19 回)、第 4 位 154.80 点 (10 回)、第 5 位 138.15 点 (13 回) と続いている。

ここで、経路課題の例とそのときの対話を図2に示す。この対話は、13 発話、42 文節、178 文字からなり、ユーザと対話システムには 258 点が与えられた。

2.2 分析対象とする対話の選定

まず、対話システムの頑健性を表す指標として「対話成功率」を定義する。

$$\text{対話成功率} = \frac{\text{得点が 0 でない対話の数}}{\text{総対話数}}$$

この対話成功率という指標は、言い換えれば対話が破綻しなかった割合を表しており、コンテストで用いられた指標「平均得点」とは次のような関係にある。

$$\text{平均得点} = \text{対話成功率} \times \text{対話効率}$$

ただし、

$$\text{対話効率} = \frac{\text{総獲得得点}}{\text{得点が 0 でない対話の数}}$$

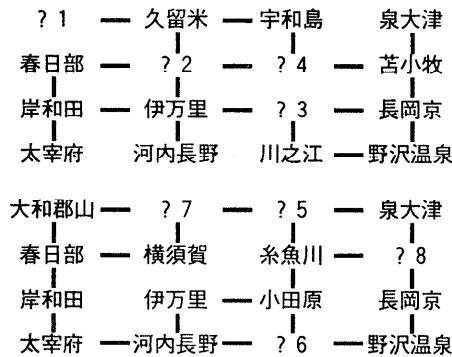
とする。

以上の指標を用いた分析結果を表1にまとめる。表1によれば、平均得点に基づくコンテスト順位と対話成功率に基づく順位は一致しており、筆者らの対話システムが対話成功率 65.2% で第 1 位の成績を得ている。そこで本稿では筆者らの対話システムとユーザとの対話 (141 対話) を中心に分析する。なお、対話効率は、対話が成功した場合に期待される得点 (300 - 発話文節数の総和) を表しており、得点が高いほど、少ない発話文節数で対話に成功することを意味する。

3. 発話 (テキスト) レベルの分析

3.1 発話パターンの抽出

対話システムやユーザの発話には駅名などの個別概念や表記上の揺れなどが存在したので、まずはじめに



出発駅＝泉大津，目的駅＝太宰府

(a) 先手(ユーザ)の路線図(上)と
後手(システム)の路線図(下)

- U1: 泉大津から小牧へ行きますか？
S2: 小牧は何処ですか。
U3: 小牧は泉大津の下です。
S4: 小牧に行きます。
U5: 小牧から春日部へ乗り換え無しで行けますか？
S6: 泉大津の左の駅は何ですか。
U7: 宇和島です。
S8: 宇和島の左の駅は何ですか。
U9: 久留米です。
S10: えーと、小牧からの経路は、小牧、糸魚川、宇和島、久留米、横須賀、春日部、岸和田、太宰府でいいんですか。
U11: 久留米の下は横須賀ですか？
S12: はい。
U13: はい。

(b) ユーザ(U)とシステム(S)との対話

図2 経路課題と対話の例

Fig. 2 Examples of route maps and dialogue.

表1 対話システムの頑健性

Table 1 Robustness of dialogue systems.

コンテスト 順位	対話 成功率	対話 効率	対話 回数
1	65.2%	189.0 点	141 回
2	54.1%	193.0 点	185 回
3	45.1%	217.9 点	175 回
4	17.4%	215.9 点	167 回
5	11.7%	177.7 点	60 回

それらの除去を行った。具体的には以下の方法で各発話の表記を変換した。

- 文字化けを含む発話文を「※文字化け※」と表記する。
- 駅名を「A」と表記する。
- 句読点や記号を削除する。
- 駅名の2個以上の連りを「A…A」と表記する。
- 方向概念(「上下左右東西南北」)を「(方向)」と

表2 ユーザ発話の文末表現の分類

Table 2 Classification of end-of-sentence expressions.

タイプ	表層表現の例	文数
感動詞	はい、いいえ	1095
マス	～ます、～ますか、～ません	985
デス	～です、～ですか、～ですね	705
断片文	名詞、～は、格助詞	54
スル	基本形、基本形＋の	21
依頼文	～ください	9
ンデス	～んですが、～んでしょ	8
ノデス	～のですか、～のですね	6
ダ	～だよ、～だってば	2
ンダ	～んだけど	1
ナノ	～なの	1
シテ	タ系連用テ形	1
その他	(文字化け、タイプミスなど)	42
合計		2930

表記する。

- 数字概念(序数を含む)を「(数字)」と表記する。
- 漢字表記とひらがな表記が混在しているものを漢字表記もしくはひらがな表記のいずれかに統一する。
- 間投表現を除去する。

以上の方法により、たとえば、「太宰府は春日部の右ですか。」という発話は「AはAの(方向)ですか」という表記に、また「えーと、富士見、寝屋川、東村山、習志野と行けますか？」という発話は「A…Aと行けますか」という表記に変換される。

なお、筆者らの対話システムとユーザとの対話(141対話)にはユーザ発話2,930文、システム発話2,898文が含まれていたが、上述の変換操作により、ユーザ発話(2,930文)は302個のパターンに、システム発話(2,898文)は48個のパターンに分類された。

3.2 ユーザ発話の文形式と文末表現

ユーザの対話システムへの発話は、基本的には対話システムが受理できるとユーザが考えている範囲内でなされる。そこで、本節ではユーザ発話の文末表現と文形式を調べ、対話システムに対するユーザの発話モデルについて考察する。

まずはじめに、ユーザ発話2,930文に対し、その文末表現を調べた。結果は表2のとおりであった。「感動詞」や「その他」を除くユーザ発話1,793文のうちの1,690文(94.3%)が「デス」と「マス」に分類されており、ユーザが比較的丁寧な物言いで発話していることが分かる。この結果は、第2位の対話システムとユーザとの対話(185対話)でも同様であり、ユーザ発話2,453文(表2の「感動詞」や「その他」に相当する発話を除く)のうちの2,302文(93.8%)がデス・マス調の発話であった。文献11)によると、対話

表3 ユーザ発話とシステム発話の文形式

Table 3 Forms of user and system utterances.

頻度 (割合)	平叙文	真偽 疑問文	疑問語 疑問文
システム 発話	559 (19.3%)	1517 (52.3%)	470 (16.2%)
ユーザ 発話	1061 (36.2%)	444 (15.2%)	210 (7.2%)

表4 駅の場所や駅名が分からないときのユーザ発話文の形式

Table 4 Forms of questions on the name and location of a station.

頻度 (割合)	平叙文	真偽 疑問文	疑問語 疑問文
駅の場所	5 (3.0%)	41 (24.6%)	121 (72.5%)
駅名	28 (17.3%)	64 (39.5%)	70 (43.2%)

システム (wizard) との対話においてユーザは、単純かつ限定された child language を用い、断片文などの非文法的発話を多用すると述べられているが、今回の対話コンテストのような実動の対話システムとの対話ではシステムの解析失敗を避けるべく、より確実なデス・マス調の言葉を用いたものと考えられる。

次に、ユーザ発話とシステム発話の文形式を比較してみた。その結果を表3に示す。表3によれば、ユーザ発話はシステム発話に比べ、真偽疑問文や疑問語疑問文の割合が低く、平叙文の割合が高いことが分かる。この理由を調べるために、駅の場所や駅名が分からないとき、ユーザがどのような発話を行ったのか調べてみた。その結果を表4に示す。駅の場所や駅名が分からないときは、疑問語疑問文の形式で直接的に質問するのが最も単純な方法だが、表4が示すように、駅の場所に目星をつけた真偽疑問文の形式(「太宰府は春日部の右ですか」など)や、駅名を確認せずに間接的に言及する平叙文の形式(「伊勢原の一つ右へ行きます」など)も比較的使用されている。これは、対話システムとの対話においても人間同士の対話のような間接的な表現が用いられていることを示している。

以上の分析の結果、対話システムが受理できるとユーザが考えた発話の範囲は、表層表現上はデス・マス調の丁寧な発話であったが、発話形式としては直接的な表現だけでなく、間接的な表現も含んでいた。したがって、対話システムを設計する際には、ユーザがデス・マス調で話すことを前提に、直接的な表現だけでなく、間接的な表現も解釈できるよう注意する必要がある。特に、受理すべき疑問語疑問文に対し、そのパラフレーズの表現(平叙文や真偽疑問文の形式)も考慮する必要がある。

表5 発話の分類

Table 5 Classification of user and system utterances.

	トピック	システム発話	ユーザ発話
平叙文	移動欲求	227 (4)	4 (4)
	移動意志	98 (1)	117 (15)
	移動可能性	32 (2)	458 (54)
	駅名	57 (1)	348 (12)
	駅の場所	128 (3)	89 (13)
	経路	2 (1)	18 (9)
	現在地	6 (1)	0 (0)
	出発駅	2 (1)	9 (2)
	その他	7 (2)	18 (16)
	小計	559 (16)	1061 (125)
真偽疑問文	移動欲求	0 (0)	1 (1)
	移動意志	0 (0)	6 (5)
	移動可能性	1320 (4)	347 (46)
	駅名	0 (0)	8 (6)
	駅の場所	0 (0)	38 (2)
	経路	197 (6)	35 (10)
	その他	0 (0)	9 (9)
	小計	1517 (10)	444 (79)
疑問語疑問文	移動方法	0 (0)	2 (2)
	駅名	357 (3)	70 (11)
	駅の場所	81 (4)	121 (12)
	経路	0 (0)	1 (1)
	現在地	10 (1)	10 (7)
	出発駅	22 (1)	4 (3)
	経由駅	0 (0)	1 (1)
	その他	0 (0)	1 (1)
	小計	470 (9)	210 (38)
移動依頼		0 (0)	9 (8)
肯定的応答		211 (3)	915 (18)
否定的応答		123 (6)	243 (11)
その他		18 (4)	48 (23)
合計		2898 (48)	2930 (302)

※ただし、括弧内の数字は発話パターン数を表している。

3.3 ユーザ発話の多様性

本節では、対話システムとの対話におけるユーザ発話の多様性の程度を分析する。

まずはじめに、ユーザ発話 2,930 文およびシステム発話 2,898 文を、文の形式と文に含まれるトピックの種類で分類した。その結果を表5に示す。表5においてユーザ発話とシステム発話のパターン数を比べて見ると、ユーザ発話の方がパターン数において 6.3 倍 (302/48) ほど多様であることが分かる。同様に、第2位の対話システムとユーザとの対話においても、ユーザ発話の方がパターン数において 12.5 倍 (376/30) ほど多様であった。ここで、両方の対話におけるユーザ発話の多様性を比較するために、1パターンあたりの発話数を調べて見ると、筆者らの対話システムと対話したユーザの場合で 9.7 文/パターン (2,930/302)、第2位の対話システムと対話したユーザの場合で 10.2 文/パターン (3,824/376) とほぼ同じであった。また、ユーザが用い、システムが用いなかった発話パターンを調べて見ても、筆者らの対話システムと対話したユーザの場合で 67.9% (190/280)、第2位の対話システムと対話したユーザの場合で 67.4% (240/356) が利用頻度1の発話パターンであり、ほぼ同じ結果となっ

表6 「真偽疑問文－移動可能性」の細分類

Table 6 Subclasses of class “真偽疑問文－移動可能性”.

含まれている情報の種類	文数
出発駅, 方向, 距離, 目的駅	2 (1)
出発駅, 方向, 目的駅	3 (3)
出発駅, 目的駅, 制約	16 (7)
出発駅, 目的駅	150 (9)
出発駅, 基準駅, 方向	2 (2)
出発駅, 方向, 距離	2 (1)
出発駅, 方向	21 (3)
目的駅, 経由駅	2 (2)
目的駅, 方向	2 (1)
目的駅, 制約	3 (2)
目的駅	112 (8)
基準駅, 方向, 距離	1 (1)
基準駅, 方向	19 (3)
方向	11 (2)
経路 (駅の連なり)	1 (1)
合 計	347 (46)

※ただし、括弧内の数字は発話パターン数を表している。

ている。以上のことから、両方の対話におけるユーザ発話の多様性は同程度であったといえる。なお、筆者らの対話システムの発話パターンは、経路課題を解くうえで最低限必要な対話が行えるレベルに、対話が単調にならないよう言い方や文末表現に多少のバリエーションを加えて設計されており、その分発話パターン数が多くなっている。

次に、表5中の「真偽疑問文－移動可能性」に該当するユーザ発話(347文)を抽出し、文に含まれている情報の種類という観点で細分類してみた。その結果が表6であり、ユーザ発話(347文)がさらに15個のサブクラスに細分類されることが分かる。また、表6中の「出発駅, 目的駅」に分類された発話(150文)の発話パターンを見てみると、表7に示されるように表層表現上の多様性も観測された。表7において、出発駅を示す格には格助詞「から」しか使われていないが、目的駅を示す格には格助詞の「に、へ、まで」や提題助詞の「は」が使われている。また、疑問を表す助詞としては終助詞の「か、ね、よね」が利用されている。文法書などではこれら助詞のニュアンスの違いが詳しく述べられているが、実際の対話場面では区別なく利用されており、頑健な対話システムを設計する際の注意点といえる。

次に、対話数の増加、すなわちユーザ発話数の増加にともない、異なり形態素や発話パターン数の数がどう変化するのか調べた。ただし、形態素の認定は汎用日本語形態素解析システム JUMAN²¹⁾を用いて自動的に行ったため、一部の発話文では未定義語や誤った形態素解析結果が生じているが、未定義語は「未定義語」という品詞と仮定してカウントし、誤りもそのままカウントした。なお、ユーザ発話数は各対話の開始時

表7 「真偽疑問文－移動可能性－出発駅&目的駅」に分類された発話パターン

Table 7 Surface patterns observed in subclass “出発駅&目的駅”.

発話パターン	文数
A から A に行けますか	75
A から A へ行けますか	35
A から A には行けますか	14
A から A に行けますね	10
A から A まで行けますか	7
A から A に行けますよね	6
では A から A まで行けますか	1
A から A まで行けますよね	1
A から A へは行けますか	1
合 計	150

基準に累積加算されている。

ユーザ発話数の増加にともない異なり形態素数、発話パターン数、異なり未定義語数の推移をそれぞれ図3、図4、図5に示す。これらの図より、異なり形態素数と異なり未定義語数は、対話の順番がシステム先手であるかユーザ先手であるかに関係なく、同じようなペースで増加しているが、発話パターン数はユーザ先手のときの方がハイペースであることが分かる。発話パターン数の推移においてユーザ先手のときの方がハイペースなのは、ユーザ後手の場合は、システムの質問に対し、「はい」、「いいえ」または単純な平叙文(たとえば、「Aです」や「Aの(方向)です」など)で答えることも可能だが、ユーザ先手のときは、質問文を自分で作る必要があったためと考えられる。なお、当コンテストにおいて対話の主導権は固定されていなかったで、「システム先手(ユーザ先手)」は対話システム(ユーザ)が対話の主導権を持つことを必ずしも意味しない。しかしながら、筆者らの対話システムとの対話では全141対話中95対話で主導権の移動が行われておらず、システム先手(ユーザ先手)の対話では多くの場合、対話システム(ユーザ)が質問側、ユーザ(対話システム)が応答側であった。

ここで、比較のために、第2位の対話システムとの対話における発話パターン数の推移の仕方を調べてみた。結果は図6に示されたとおりであり、筆者らの対話システムと対話したユーザの場合とほぼ一致している。

3.4 システム発話のユーザ発話への影響

本節では、ユーザ発話がシステムの発話パターンにどの程度影響されているのかを調べる。

システム発話とユーザ発話に共通な発話パターンは表8に示された22種類しかなかったが、その22種類中にシステム発話の82.5%(2,392/2,898)、ユーザ発話の61.8%(1,811/2,930)が含まれており、一見

異なり形態素数

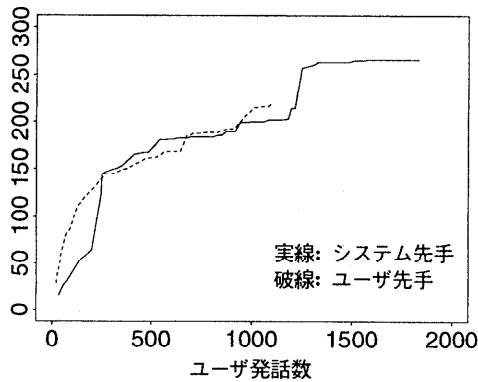


図3 第1位の対話システムとの対話における異なり形態素数の推移

Fig. 3 Increase of different morphemes in dialogues with our dialogue system.

異なり未定義語数

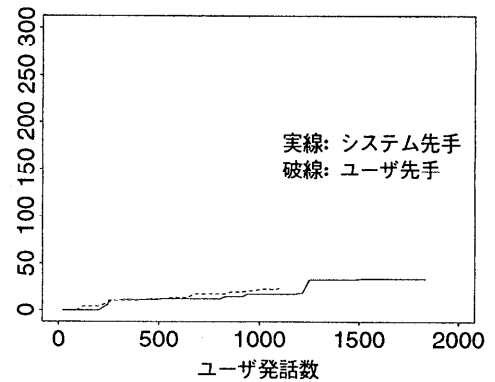


図5 第1位の対話システムとの対話における異なり未定義語数の推移

Fig. 5 Increase of different undefined words in dialogues with our dialogue system.

発話パターン数

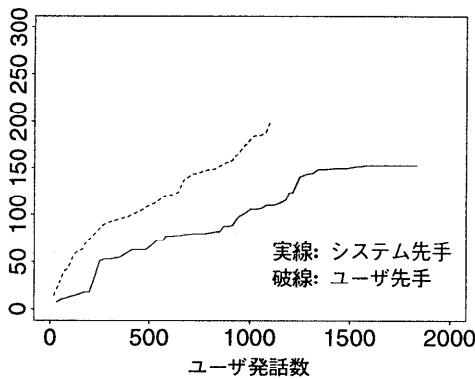


図4 第1位の対話システムとの対話における発話パターン数の推移

Fig. 4 Increase of sentence patterns in dialogues with our dialogue system.

発話パターン数

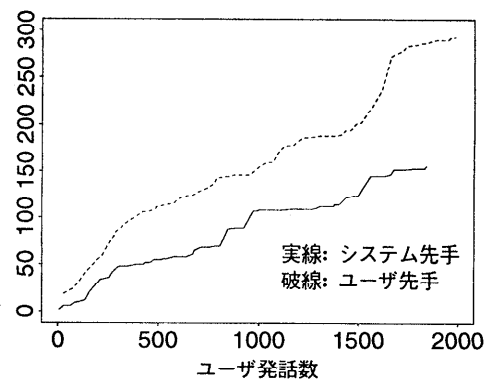


図6 第2位の対話システムとの対話における発話パターン数の推移

Fig. 6 Increase of sentence patterns in dialogues with the second dialogue system.

システム発話とユーザ発話にはパターンの強い類似性があるように思われる。しかしながら、この類似性は、ユーザがシステムの発話パターンを再利用したというよりも、その発話パターンが経路課題を解くうえで典型的であったために生じたのかもしれない。実際、パターンの種類としては全ユーザ発話パターンの7.3% (22/302) にしかすぎず、共通な発話パターンにおける発話文数の分布も大分異なっている。そこで、駅の場所を質問するときの発話パターン「Aは何処ですか」と駅名を質問するときの発話パターン「Aの(方向)の駅は何ですか」を例に、ユーザ発話とシステム発話の使われ方を調べてみた。その結果、「Aは何処ですか」は、対話システムが33回、ユーザが108回用いているにもかかわらず、同じ対話で双方が用いたの

は9回だけであり、しかも、うち3回は対話システムよりも先にユーザが発話している場合であった。一方、「Aの(方向)の駅は何ですか」は対話システムが215回用いたにもかかわらず、ユーザは4回しか利用していなかった。なお、以上の傾向は第2位の対話システムとユーザとの対話でも観測される。システム発話とユーザ発話に共通な発話パターンは20種類あったが、その中にシステム発話の89.6% (3,374/3,765)、ユーザ発話の55.8% (2,134/3,824) が含まれていた。しかしながら、パターンの種類としては全ユーザ発話パターンの5.3% (20/376) にしかすぎず、共通な発話パターンにおける発話文数の分布も大分異なっていた。

以上のことから、ユーザ発話の半分以上がシステムの発話パターンと同じパターンであるのは、ユーザが

表 8 システム発話とユーザ発話に共通な発話パターン
Table 8 Common surface patterns between user and system utterances.

ユーザ発話	システム発話	発話パターン
867	167	はい
299	57	A です
225	76	いいえ
108	33	A は何処ですか
75	43	A から A に行けますか
69	252	A に行けますか
66	4	A の (方向) です
54	98	A に行きます
8	45	A から A までの経路は A…A でいいですか
8	2	A から A までの経路は A…A です
6	159	A から A に行けますよね
6	2	A からです
4	215	A の (方向) の駅は何ですか
4	83	A から A までの経路は A…A ですよ
3	28	A には行きません
2	52	A は A の (方向) です
2	15	分かりません
1	866	A に行けますよね
1	121	A の (方向) にある駅は何ですか
1	58	A から A に行きたいんですけど
1	10	我々は今何処にいるんですか
1	6	A から A に行きたいんですが
1811	2392	合 計

システムの発話パターンをまったくそのままの形で再利用したというよりも、単に経路課題を解くうえで典型的な発話パターンを用いたためと考えられる。ただ、ユーザが用い、システムが用いなかった発話パターンを両者に共通な発話パターンと比較してみると、助詞が異なるだけのものや余分な語（接続詞や副詞など）が挿入されているだけのものも多い（表 7 参照）。したがって、受理すべきユーザ発話の範囲を決める際には、典型的な発話パターンおよびそのバリエーション（特に助詞の多様性と挿入語の有無）を考慮する必要がある。また、他の対話システムが用いた発話パターンや他の対話システムで受理された発話パターンを利用しているケースも観測され、この点も考慮する必要がある。

4. 対話レベルの分析

ユーザの対話システムに対する発話は、人間同士の対話のときと同様、十分多様であり、また文法的に必ずしも正しいとはいえない。したがって、対話システムの実用レベルでの利用に際しては、未定義な形態素や想定されていない発話パターンに対処するための枠組みが必要となってくる。従来の研究では、未定義な形

表 9 予想外のシステム発話に対するユーザの反応
Table 9 User reaction for out-of-context system's questions.

S 発話	我々は今何処にいるんですか	
ユーザの反応	S 発話に回答した	9
	S 発話を見捨てて発話した	1
S 発話	何処から A に行くんですか	
ユーザの反応	S 発話について質問した	4
	S 発話に回答した	14
	S 発話を見捨てて発話した	2
	タイプミスだった	2
S 発話	A の (方向) の駅は何ですか	
ユーザの反応	S 発話について質問した	7
	S 発話に回答した	22
	S 発話を見捨てて発話した	1
S 発話	わかりません	
ユーザの反応	自発話の一部を修正した	2
	自発話の内容を変更した	10
	文字化け、タイプミスだった	3

※表中の数字は生起頻度を表す。

態素に対してはその品詞や概念を推定すること^{22),23)}に、発話パターンの多様性に対しては自然言語処理部の頑健性を高めること^{24)~26)}に重点が置かれている。しかしながら、各個人の言葉に対するイメージは微妙に異なっており、その意味や役割、解釈を一意に決定することは容易ではない。加えて、対話の混乱の原因となりうる現象の多種多様さのため、個別に対応することも困難と思われる。人間同士の対話でも未知語や非文法的発話が観測され、対話の混乱の原因となっているが、そのようにして生じた混乱は対話を通して修復されることが多い。そこで本章では、コンテストで得られた対話（141 対話）を分析し、様々な対話の混乱に対してユーザがどのように反応するのか、対話レベルのヒューマンファクタを解析する。

4.1 予想外のシステム発話に対するユーザの反応

まずはじめに、対話システムが文脈から外れて、予想外の発話をしたときに、ユーザがどのような反応を示すのか調べた。その結果を表 9 にまとめる。対話システムがした予想外のシステム発話は他にもいろいろあったが、ここでは出現頻度の高かった 4 種類の発話を対象とした。表 9 より、システム発話が質問文のときのユーザ発話は、多くがそのシステム発話に対する質問または回答であり、システム発話を見捨てるというケースは少なかった。一方、システム発話が何かへの不理解を示す「わかりません」という言葉であるとき、ユーザは、直前の自発話を修正するというよりも、発話内容そのものを変え、まったく別のことを発話した。人間同士の対話であれば、「わかりません」という言葉に対しては相手が分かるように言い直すこ

表 10 自分の質問を無視されたときのユーザの反応
Table 10 User reaction when our dialogue system ignored a user utterance.

ユーザの反応	頻度
S 発話について質問した	4
S 発話に回答した	23
S 発話を無視して発話した	4
無入力（空行を入力）だった	1
対話を打ち切った	1
合 計	33

とが一般的と思われるが、機械相手の対話では異なる対応をしている。このような現象が生じたのは相手に対する配慮が必要ないためと考えられる。また、「自発話の内容を変更した」場合の 10 回中 5 回は、発話内容がまったく異なるにもかかわらず、構文構造は同じであり、単純に単語を入れ替えただけの発話であった。これは、話し相手である対話システムが「わかりません」と発話する原因の中から、構文解析失敗の可能性が除かれていることを意味している。なお、他の 5 回は、内容的にも構文的にもまったく異なる発話であった。

一方、対話システムは、ユーザ質問に対する発話理解や対話処理に失敗したとき、その質問を無視して、対話の主導権を奪うことがあった。そのときのユーザの反応を調べてみた。結果を表 10 に示す。表 10 より、対話システムがユーザの質問を無視した場合でも、ユーザは好意的に対応し、質問または回答をしていることが分かる。システム発話を無視するケースも 33 回中 4 回ほどあったが、うち 3 回は直前の自発話を修正するための発話であり、まったく無視したといえるのは 1 回だけであった。ただし、対話を中止したケースも 1 例だけがあった。

以上の分析結果より、目的指向型の対話においてユーザ質問に対する発話理解や対話処理に失敗したときは、「わかりません」という不明確な情報を流すより、既知の目的の達成に向けて対話を進める方が有効であるといえる。もちろん「わかりません」だけでなく、「(何々が)わかりません」と対象を明確に知らせるという戦略も考えられるが、対話システムの内部構成を知らない人に対して、ある特定の処理（形態素解析や構文解析、意味解析、対話処理）の成否などを知らせても、ユーザを混乱させるだけであり、好ましい戦略とはいえない。

4.2 システム発話の反復に対するユーザの反応

対話システムは、ユーザ回答に対する発話理解や対話処理に失敗したとき、システム発話を繰り返すことがあった。筆者らの対話システムとユーザとの対話

表 11 システム発話の 1 回目の反復に対するユーザの反応
Table 11 User reaction for questions first repeated by our dialogue system.

ユーザの反応	頻度
最初の U 発話の一部を修正した	22
最初の U 発話を反復した	8
自発話の内容を変更した	66
その他（文字化けなど）	4
合 計	100

表 12 システム発話の反復が対話に与える影響
Table 12 Influence of repetition of system utterances on a user.

ユーザの反応	頻度
正常な対話文脈への復帰に役立った	67
新たな対話文脈ができた	13
絶望的な対話状況に陥った	14
対話が打ち切られた	6
合 計	100

(141 対話)において、このような場面は全部で 100 回あり、1 回の場面でシステムが発話を反復した回数は平均 1.48 回、最多 9 回であった。

まずはじめに、本節ではシステム発話の 1 回目の反復に対するユーザの反応を調べて見た。その結果を表 11 に示す。表 11 によれば、ユーザの反応は、「わかりません」に対するとときと同様、自発話を修正して言い直すというよりも、発話内容そのものを変えてしまう傾向にあった。

次に、システム発話の反復が対話の修復に貢献したのかどうかを調べた。その結果は表 12 のとおりであった。「正常な対話文脈への復帰に役立った」場合と「新たな対話文脈ができた」場合とを合わせて、100 回の繰返し場面中 80 回が対話の修復に貢献しており、発話理解や対話処理に失敗したときにシステム発話を反復するという対話戦略が有効であることを示している。ただし、「正常な対話文脈への復帰に役立った」とは、ユーザが反復されたシステム質問に対する回答を行い、その発話理解と対話処理に対話システムが成功したことを意味する。一方、「新たな対話文脈ができた」とは、ユーザがシステム質問を無視して、質問を行い、その発話理解と対話処理に対話システムが成功したことを意味する。このユーザ質問、システム回答により、対話の主導権はユーザに移り、今までの対話に対し入れ子対話が生じている。なお、「絶望的な対話状況に陥った」りした場合も 14 回ほどあったが、うち 10 回はユーザが事実と反する回答をしたためであり、残りの 4 回は対話システムがユーザ発話の理解に失敗し、事実と反する命題を知識として登録してしまったためであった。「対話が打ち切られた」という場合も 6 回は

どあった。そのうちの1回は文節数超過によるものであったが、残りの5回については打ち切られた理由も不明であり、より多くの事例を集め、分析する必要がある。今後の課題とする。

5. ま と め

自然言語対話システムにユーザとの対話を通して何らかの課題を解かせ、その成績によって対話システムの性能を競うという対話コンテスト (DiaLeague'97) が WWW ページを介して行われた。

本稿では、当コンテストにおいて収集された対話 (728 対話) のうち、最も頑健であった筆者らの対話システムとユーザとの対話 (141 対話) を中心に分析し、機械である対話システムとの対話においてユーザがどのように振る舞うのか発話レベルと対話レベルとで調べた。その結果、発話レベルの分析では、(1) ユーザは対話システムに対しデス・マス調の文末表現を用いる。(2) 話し相手が機械であるにもかかわらず、間接的な発話形式を採用することがある。(3) ユーザ発話数の増加にともない、発話パターンや異なり形態素の数も増加した。このとき、異なり形態素数は、対話の順番がシステム先手であるかユーザ先手であるかに関係なく、同じようなペースで増えたが、発話パターン数はユーザ先手のときの方がハイペースであった。(4) ユーザは、システムの発話パターンをまったくそのままの形で再利用していないといったことが分かった。一方、対話レベルの分析では、(1) ユーザは、文脈から外れた予想外のシステム発話に対しても好意的に振る舞う。(2) 「わかりません」というシステム発話やシステム発話の反復に対して、ユーザは、自分の発話を修正して言い直すことよりも、発話内容そのものを変え、まったく別のことを発話する方を好んでいる。(3) 発話理解や対話処理を失敗したときには、システム発話を繰り返すという対話戦略が有効であるということが分かった。

発話スタイルや対話の進め方などには個人差がある。対話データをユーザごとに分類し、対話に現れる個人的特徴をユーザモデルとして定量化したい。また、各特徴量の1対話中における変化や複数対話における変化、すなわち学習によるユーザモデルの変化を定式化したい。今後の課題とする。また、本稿で述べた対話の混乱に対する分析結果から、そのような混乱を修復するためのメタな対話戦略ルールを定式化し、対話システムに実装する予定である。

謝辞 対話コンテスト DiaLeague'97 を企画・実行し、筆者らの対話システムとインターネットユーザと

の対話の場を与えてくださった、電子技術総合研究所 橋田浩一博士をはじめとする DiaLeague 実行委員会の皆様に、つつしんで感謝の意を表します。

参 考 文 献

- 1) 高間康史, 土肥 浩, 石塚 満: 擬人化エージェントにおける音声対話を通じての協調的応答戦略の自動学習, 人工知能学会誌, Vol.12, No.3, pp.456-465 (1997).
- 2) 山田耕一, 溝口理一郎, 原田直樹: 質問応答システムにおけるユーザ発話モデルと協調的応答の生成, 情報処理学会論文誌, Vol.35, No.11, pp.2265-2275 (1994).
- 3) 佐川雄二, 大西 昇, 杉江 昇: 自然言語対話システムにおけるユーザモデルの更新に伴う対話の再プランニングに関する考察, 情報処理学会論文誌, Vol.35, No.6, pp.1042-1049 (1994).
- 4) 熊本忠彦, 伊藤 昭: 支援対話の解析に基づく対話処理方式の提案, 信学論, Vol.J77-D-II, No.8, pp.1492-1501 (1994).
- 5) 大平栄二, 阿部正博, 小松昭男, 市川 熹: 情報検索における柔軟な対話制御方式, 信学論, Vol.J76-D-II, No.3, pp.586-595 (1993).
- 6) 西山敏雄, 大山芳史: データベース検索における協調的な自然語対話処理と評価, 信学論, Vol.J73-D-II, No.4, pp.625-632 (1990).
- 7) Kaplan, S.J.: Cooperative Responses from a Portable Natural Language Query System, *Artif. Intell.*, Vol.19, No.2, pp.165-187 (1982).
- 8) Motro, A.: FLEX: A Tolerant and Cooperative User Interface to Databases, *IEEE Trans. Knowledge and Data Engineering*, Vol.2, No.2, pp.231-246 (1990).
- 9) Wilensky, R., Arens, Y. and Chin, D.: Talking to UNIX in English: An Overview of UC, *Comm. ACM*, Vol.27, No.6, pp.574-593 (1984).
- 10) 中川聖一, 山本誠治: 音声対話システムの構成法とユーザ発話の関係, 信学論, Vol.J79-D-II, No.12, pp.2139-2145 (1996).
- 11) Guindon, R.: How to Interface to Advisory System? Users Request Help with a Very Simple Language, *Proc. CHI'88*, Washington D.C., USA, pp.191-196 (1988).
- 12) 長谷川修, 伊藤克亘, 秋葉友良, 速水 悟, 田中和世: 高度対話システムの実現のためのヒューマンファクタの解析, 人工知能学会全国大会 (第8回) 論文集, No.S3-2, pp.97-100 (1994).
- 13) 望主雅子, 酒寄哲也, 佐藤奈穂子, 小島裕一: 調理行動に伴う対機械対話の発話現象, 情報処理学会研究報告, No.SLP17-6, pp.33-38 (1997).
- 14) Moore, R. and Morris, A.: Experiences Collecting Genuine Spoken Enquiries using WOZ

- Techniques, *Proc. DARPA Speech and Natural Language Workshop*, New York, USA, pp.61-63 (1992).
- 15) Hirschman, L., Bates, M., Dahl, D., Fisher, W., Garofolo, J., Pallett, D., Hunicke-Smith, K., Price, P., Rudnick, A. and Tzoukermann, E.: Multi-Site Data Collection and Evaluation in Spoken Language Understanding, *Proc. ARPA Human Language Technology*, New Jersey, USA (1993).
 - 16) 黒岩真吾, 武田一哉, 井ノ上直巳, 山本誠一: 機械との対話における発話分析, 信学技報, No.SP94-30, pp.57-64 (1994).
 - 17) 金沢博史, 瀬戸重宣, 新地秀昭, 竹林洋一: 音声自由対話システム TOSBURG II におけるデータ収集と評価 環境, 信学技報, No.SP93-114, pp.1-6 (1993).
 - 18) Bernstein, J. and Danielson, D.: Spontaneous Speech Collection for the CSR Corpus, *Proc. DARPA Speech and Natural Language Workshop*, New York, USA, pp.373-378 (1992).
 - 19) Pao, C. and Wilpon, J.: Spontaneous Speech Collection for the ATIS Domain with an Aural User Feedback Paradigm, *Proc. DARPA Speech and Natural Language Workshop*, New York, USA, pp.43-48 (1992).
 - 20) 橋田浩一, 伝 康晴, 長尾 確, 柏岡秀紀, 酒井桂一, 島津 明, 中野幹生: DiaLeague—自然言語処理システムの総合評価, 人工知能学会誌, Vol.12, No.3, pp.390-399 (1997).
 - 21) 妙木 裕, 松本裕治, 長尾 真: 汎用日本語辞書および形態素解析システム, 第42回情報処理学会全国大会論文集, No.1C-9, pp.3-17-3-18 (1991).
 - 22) 伊藤慶明, 木山次郎, 岡 隆一: 文スポッティング音声認識における部分文認識と未知語処理方式, 信学論, Vol.J77-D-II, No.8, pp.1531-1542 (1994).
 - 23) 大石 亨, 松本裕治: 表層表現による日本語動詞句のアスペクトの推定, 自然言語処理, Vol.4, No.4, pp.41-60 (1997).
 - 24) 元吉文男, 大場健司, 石崎 俊, 井佐原均, 横山晶一: 未定義語を含む文の多段階構文解析法, 信学論, Vol.J72-D-II, No.10, pp.1672-1679 (1989).
 - 25) 今井宏樹, 田中穂積, 徳永健伸: 一般化 LR 法を用いた非文の処理, 人工知能学会全国大会(第8回)論文集, No.20-4, pp.539-542 (1994).
 - 26) 今一 修, 松本裕治: 文法的不適格文処理のための統合的枠組み, 人工知能学会誌, Vol.12, No.3, pp.404-411 (1997).

(平成 10 年 6 月 1 日受付)

(平成 10 年 12 月 7 日採録)



熊本 忠彦 (正会員)

昭和 41 年生。昭和 63 年筑波大学第三学群情報学類卒業。平成 2 年同大学院理工学研究科修士課程修了。同年、郵政省通信総合研究所入所。現在、同関西先端研究センター知識処理研究室主任研究官。平成 8 年筑波大学博士(工学)。頑健な話し言葉対話システムの実現を目指している。電子情報通信学会、人工知能学会、言語処理学会各会員。



伊藤 昭 (正会員)

昭和 25 年生。昭和 47 年京都大学理学部物理学科卒業。昭和 54 年同大学院理学研究科博士課程修了。理学博士。同年、郵政省電波研究所(現通信総合研究所)入所。以後、通信ネットワーク、知識処理、対話システム、ヒューマンインタフェース、エージェントモデル、コミュニケーションの認知機構などの研究に従事。同関西先端研究センター知識処理研究室長、研究調整官を経て、現在山形大学工学部電子情報工学科教授。電子情報通信学会、人工知能学会、日本ソフトウェア科学会、ACM、AAAI 各会員。