

WWWを介した対話システムとの対話における 混乱の分析

熊本 忠彦[†] 伊藤 昭^{††}

[†] 郵政省 通信総合研究所 関西支所
^{††} 山形大学 工学部

〒 651-2401 兵庫県神戸市西区岩岡町岩岡 588-2
Tel.: (078) 969 2173, E-mail: kuma@crl.go.jp

あらまし 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舞うのか、ヒューマンファクタに関する解析は十分とは言えない。そこで本稿では、対話コンテスト DiaLeague'97 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を調べた。なお、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一つの試みとして実施されたコンテストであり、WWWページを介して行われた。

キーワード 自然言語対話システム, 対話コンテスト DiaLeague'97, ヒューマンファクタの解析

An Analysis of Instances of Confusion Observed in Dialogues with Our Dialogue System through a WWW Page

Tadahiko Kumamoto[†] and Akira Ito^{††}

[†]Kansai Advanced Research Center, Communications Research Laboratory, MPT
^{††}Faculty of Engineering, Yamagata University

588-2 Iwaoka-cho Iwaoka, Nishi-ku, Kobe, 651-2401 Hyogo, Japan
Tel.: (078) 969 2173, E-mail: kuma@crl.go.jp

Abstract 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, and was proposed as a method for objectively evaluating the overall performance of a dialogue system. This contest was held on the WWW (World-Wide Web) pages for one week, and many Internet users participated. The five dialogue systems had 728 dialogues with the users. We analyzed the 141 dialogues between our dialogue system and the users, and investigated instances of confusion observed in the 141 dialogues. We also show how the users behaved when the confusion occurred.

key words Natural language dialogue system, dialogue contest DiaLeague'97, human factor

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、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盛んに研究されている^{[1]-[3]}。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舞うのか、ヒューマンファクタに関する解析は十分とは言えない。

対話は対話者同士による協同作業であるため、対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^[4]と実動の対話システムを利用するもの^[5]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 [4] 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の特徴（間投表現や言い淀み、言い直しなど）を分析するには十分だが、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'97^[6]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一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、

もしくはシステム設計者による模擬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言える。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）はキーボード等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そのものの分析には適していると言える。

本稿では当コンテストで得られた対話（728 対話）のうち、我々が開発し、参加登録した対話システムとの対話（141 対話）を分析し、機械である対話システムとの対話において生じる様々な混乱に対し、ユーザがどのように振舞うのかを調べる。具体的には、（1）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、（2）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変わるのか、といったことを調べる。

2. 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^[6]。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出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。

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

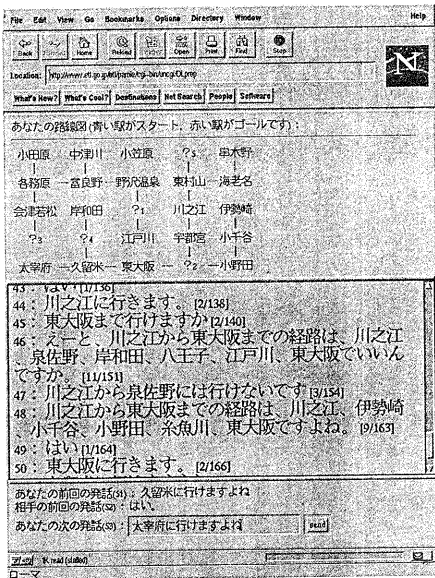


図1 WWWページを介した対話の様子

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、対話システム、対話の順番(先手/後手)、経路課題を適当に決め、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)は最初決められたとおりであった。

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、我々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。ちなみに、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、98.38点(175回)、37.50点(167回)、20.73点(60回;中途参加)であった。また、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。

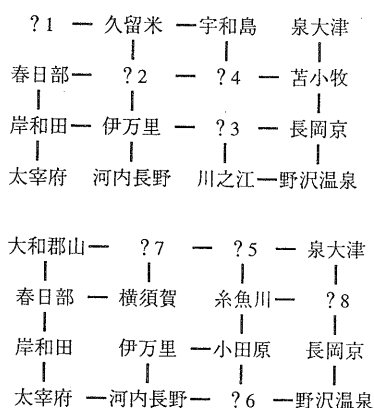
ここで、経路課題の例とそのときの対話を図2に示す。この対話は、13発話、42文節、178文字からなり、ユーザと対話システムには258点が与えられた。

3. 対話の混乱の分析

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、また文法的にも必ず正しいとは言えない。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、未定義な形態素に対してはその品詞や概念を推定すること[8]に、発話パターンの多様性に対しては自然言語処理部のロバスト性を高めること[9],[10]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得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を調べる。

3.1 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、唐突なことを発話したときに、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1にまとめる。対話システムが唐突に言った言葉は他にもいろいろ



(a) 先手 (ユーザ) の路線図 (上) と
後手 (システム) の路線図 (下)

- U 1: 泉大津から苫小牧へ行けますか?
 S 2: 苫小牧は何处ですか。
 U 3: 苫小牧は泉大津の下です。
 S 4: 苫小牧に行きます。
 U 5: 苫小牧から春日部へ乗り換え無しで行けますか?
 S 6: 泉大津の左の駅は何ですか。
 U 7: 宇和島です。
 S 8: 宇和島の左の駅は何ですか。
 U 9: 久留米です。
 S 10: えーと、苫小牧からの経路は、苫小牧、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。
 U 11: 久留米の下は横須賀ですか?
 S 12: はい。
 U 13: 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

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 1 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応答であり、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、システム発話が無かへへの不理解を示す「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、全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わかるよう

表 1 唐突なシステム発話に対するユーザの反応

S 発話	我々は今どこにいるんですか	
ユーザの反応	S 発話に応答した	9
	S 発話を見捨てるして発話した	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 発話に応答した	14
	S 発話を見捨てるして発話した	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 発話に応答した	22
	S 発話を見捨てるして発話した	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。

表 2 自分の質問を見捨れたときの
ユーザの反応

ユーザの反応	頻度
S 発話について質問した	4
S 発話に応答した	23
S 発話を見捨てるして発話した	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

に言い直すこ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、発話内容が全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全く異なる発話であった。

次に、対話システムがユーザの質問を見捨てる、対話の主導権を奪ったときのユーザの反応を調べてみた。その結果を表 2 に示す。表 2 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、質問または応答をしていることがわかる。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、全く無視したと言えるのは1回だけであった。但し、対話を中止したケースも1例だけがあった。

3.2 システム発話の繰り返しに対するユーザの反応

対話システムは、発話理解や対話処理の失敗に伴い、システム発話を繰り返すことがあった。そこで、システム発話の最初の繰り返しに対するユーザの反応を調べてみた。その結果を表3に示す。表3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、システム発話の繰り返し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4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回の繰り返し中80回が対話の修復に貢献しており、発話理解や対話処理を失敗したときにシステム発話を繰り返すという対話戦略が有効であることを示している。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応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たためであった。「対話が打ち切られた」という場合も6回ほどあった。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。

3.3 経路正解時、経路誤答時、文節数超過時の対話の特徴

経路正解時および経路誤答時、文節数超過による対話中止時の対話を分析し、それぞれの対話の特徴を比較した。具体的には、対話の混乱の原因となる現象の生起頻度を調べ、経路正解時の生起頻度を1とした場合の比を求めた。以上の結果を表5に示す。表5において「Sウソ/Uウソ」は対話システム/ユーザの発話内容に事実と反する内容が含まれている場合を表し、「S無視/U無視」は対話システム/ユーザが相手の質問を無視して、対話の主導権を奪った場

表3 システム発話の1回目の反復時におけるユーザの反応

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発話を繰り返した	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4 システム発話の反復が対話に与える影響

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

合を表している。「S反復」はシステム発話の繰り返しを表している。

表5を用いて、それぞれの対話の特徴を比較する。まずはじめに、経路正解時と経路誤答時を比べてみると、二つの対話の差がUウソとS反復に顕著に現れていることがわかる。経路誤答時には経路正解時の20倍強のUウソと5倍強のS反復が観測されている。しかしながら、SウソやS無視、U無視の生起頻度は、経路誤答時の方がむしろ少ない。次に、経路正解時と文節数超過時を比べてみると、文節数超過時の方が全体的に頻度が高く対話が混乱している様子が伺える。特にS反復の生起頻度が高く、文節数が増加する一因となっている。最後に、経路誤答時と文節数超過時を比べてみると、Uウソは経路誤答時の方が多いが、Sウソは文節数超過時の方が多い。経路誤答時の対話では、Uウソが対話システムの知識と矛盾しなかったため、目的駅に達することができている。ユーザはUウソを発話した後、訂正を試みる場面もあったが、対話システムには一度得た知識を修正する機能がなかったため、対話を修復することができなかった。

以上の分析の結果、ユーザとの対話が成功するか失敗するかは、Uウソの取り扱いに依っていることがわかる。我々の対話システムには、Uウソを検知し、誤った信念であることをユーザに告知する機能が備わっており、対話の混乱を回避できた場合も多かった。しかしながら、すべての状況には対応でき

表5 経路正解時、経路誤答時、文節数超過時の対話の特徴

	経路 正解時	経路 誤答時	文節数 超過時
対話数	92	11	5
S ウソ	1.47 (1)	0.74 (0.5)	10.54 (7.2)
U ウソ	0.88 (1)	18.38 (20.9)	7.25 (8.2)
S 無視	1.66 (1)	1.47 (0.9)	4.61 (2.8)
U 無視	4.80 (1)	2.94 (0.6)	9.22 (1.9)
S 反復	0.48 (1)	2.55 (5.3)	3.20 (6.7)

※S反復以外: 1,000 文節あたりの生起頻度

S反復: 1 対話あたりの生起頻度

※括弧内の数字は経路正解時の生起頻度との比を表す。

ていなかったため、経路誤答や文節数超過の原因となった。また、ユーザ発話がウソかどうかを判定できないとき、すなわち対話システムの知識と矛盾しないとき、そのまま自分の知識として登録してしまっていた。

4. 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、我々の対話システムとの対話 (141 対話) を分析し、機械である対話システムとの対話において生じる様々な対話の混乱に対しユーザがどのように振舞うのか調べた。その結果、(1) ユーザは、文脈から外れた唐突な内容のシステム発話に対しても好意的に振舞う。(2) 「わかりません」というシステム発話やシステム発話の繰り返しに対して、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、全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。(4) ユーザの事実と反する信念が対話システムに知識として登録されたとき、対

話は破綻した、といったことがわかった。

今後は、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

対話データ収集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、謹んで感謝の意を表します。

参考文献

- [1] 高間, 土肥, 石塚: 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, 溝口, 原田: 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, 情処学論, Vol. 35, No. 11, pp. 2265-2275 (1994).
- [3] 熊本, 伊藤: 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).
- [4] 中川, 山本: 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).
- [5] 黒岩, 武田, 井ノ上, 山本: 機械との対話における発話分析, 信学技報 (音声研究会), No. SP94-30, pp. 57-64 (1994).
- [6] 橋田, 伝, 長尾, 柏岡, 酒井, 島津, 中野: DiaLeague — 自然言語処理システムの総合評価一, 人工知能学会誌, Vol. 12, No. 3, pp. 390-399 (1997).
- [7] 妙木, 松本, 長尾: 汎用日本語辞書および形態素解析システム, 第 42 回情処全大講演論文集, No. 1C-9, pp. 3-17-3-18 (1991).
- [8] 大石, 松本: 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).
- [9] 元吉, 大場, 石崎, 井佐原, 横山: 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).
- [10] 今一, 松本: 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).