音声言語情報処理 17－5
（1997. 7. 18）

調理行動に伴う対機械対話収録実験

酒寄哲也・望主雅子・小島裕一・佐藤奈穂子
(株)リコー・情報通信研究所
{sakayori masako koji nahoko}@ic.rdc.ricoh.co.jp

あらまし 機械に対する発話を集めるための対話音声収集実験を行った。これは音声対話を機械のユーザー・インタフェースに用いるための基礎的知見を得る事が目的である。想定される使用場面に近い状況で発話を得るためにいくつかの工夫を行った。タスクへの動機付けを確保し話題の発散を防ぐために、実作業を伴う状況での実験を行った。作業としては調理を選んだ。その理由は人工的・恣意的でなく日常的に馴染みがある、用語が安定し語彙的にも音声認識の可能性が高い範囲である、手が塞がっている、濡れているなど、音声を主たるモードとする事に必然性がある、などである。システムのレベルを現実的なものに近づけつつ広範な対話現象を得るため、音声認識部分のみをオペレーターが代行し、その他の部分を全て自動で行うWOZ方式を採用した。被験者にはタスクとの親和性が高い主婦を選び、実験環境にも実際の厨房を用い、被験者には新しい調理器具の試用という教示を与えた。
実験中の観察、アンケート、得られたデータの簡単な分析によって、発話はスピーカーに向かってなされやすい、焦点は基本的には詳細化の方向でながれる、ユーザー・システム間での常識や指示語概念の不一致がある場合に再質問が起こることなどがわかった。

Collection of Utterance during Cooking Tasks Using Spoken Dialogue System

SAKAYORI Tetsuya MOCHINUSHI, Masako, KOJIMA Yuuichi, SATOH Nahoko
Information & Communication R&D Center, RICOH Co. Ltd.

Abstract In this article, we report an experiment for collection of human utterance to navigation system. Utterance wear collected with "Wizard of OZ" system. We select cooking tasks to give the motivation to subjects. 20 housewives gave utterance in a kitchen.
We observed moving of the focus in topics. The focus moves to one of its details usually. And it moves form topic which has no detail to other detail of that's parent topic.

1.はじめに

我々は音声対話を機械のユーザー・インタフェースに用いるための研究開発を行っている。そのために機械と対話する場合の人間の発話を知る材料として、対話音声データを収集する必要がある。まったく自由な対話は発散的になりやすく分析が難しいため、このような実験では何らかのタスクを与えて対話の範囲を限定することが行われる。ただし、人工的･恣意的なタスクでは被験者の動機づけ

が弱いなどの問題点があり、実際に想定される使用場面での発話と異なるものになる危険性がある。

そこで、我々は出来る限りリアルな状況での対話現象を知るために実行動に伴う対話の研究を行っている。このような場面での人間対人間の対話データの観察結果についてはすでに報告した[1]。ここでは人間対機械で行った対話収集実験について報告する。実験結果の詳細な分析はまだ完了していないため、本報告では主として実験方法および観察結果について述べる。なお、言語現象に関する分析結果については別途報告する[2][3]。

2.実験システム

対話収録実験のために作成した WOZ 実験システムについて説明する。今回の収録実験の目的は、音声対話インタフェースを備えた機械に対する、人間の発話現象を模擬的に集める事である。ただし機械といってもさまざまなレベルが想定する事が出来る。ＳＦ映画や漫画に登場するような高度な知性を持った機械を模擬して得られた対話は、近い将来我々が開発できる機械に対するものとは異なると予想される。このため実験システムにはオペレーターの役割を極力少なくし、現在の技術レベルで実現可能な範囲の自動化を出来るだけ盛り込む事によって現実的なレベルのシステムを模擬する事を目指した。ただし、現在の音声認識技術をそのまま用いると、許容できる文型や語彙などに大幅に制限を加えざるを得ず、本来の目的である広範な発話現象に集めることが難しくなる。

そこでユーザー発話を認識し焦点を同定する部分だけをオペレーターが肩代わりし、その他の部分はすべて自動化するものとした。特に認識されたユーザー発話の焦点からタスク知識を用いてシステムを発話を生成する部分を完全に自動化する事によって、システムの印象を大きく左右するシステム発話のレベルが現実的なものとなり、被験者に妥当なシステム像を伝える事が出来る。

さらにユーザーには発話行為限界と話題限界の制約を与えた。これらによってシステムの現実的限界を設定するとともに制御の簡単化をはかった。話題限界については次節で説明する。以下発話行為限界について述べる。タスクの性格上、発話行為の意図はユーザー発話については情報要求、システム発話については情報提供のみに集中するものと考えられるので、本システムではこれらに発話行為を限定した。これにより対話管理から発話行為に関する制御が除かれ、話題に関する制御に特化される。話題に関する制御はユーザー発話の焦点を同定することと、システム発話の焦点をしかるべき位置に移動することである。

このためシステムは音声認識部に代わってユーザー発話の焦点を指定するオペレータ・インターフェース部、その結果とタスク知識とからシステムとして焦点を移動する焦点遷移部、焦点の位置、タスク知識、言語規則からシステム発話を生成する発話生成部、得られたシステム発話を発声する音声合成部から実現された。このうちオペレーター・インターフェース部はGUIによる操作画面のみであり、音声合成部は市販の規則合成ソフトを使用したので、焦点遷移部と発話生成部について以下に説明する。

2.1.焦点遷移部

まず、本システムにおけるタスク知識の扱い方について述べる。ここでのタスク知識は実作業をリアルタイムに支援するためのものである。作業は通常いくつかの工程に細分化する事が出来る。ある工程に関わっている時

点では、それ以外の工程に関する話題に比べ、現に関わっている工程に関する話題が焦点となることが非常に多くなると予想される。このため、タスク知識は細分化された工程毎に分割して持ち、ユーザーによる現工程終了報告以外では現工程外への焦点遷移は行わない。これが前節で述べた話題限界である。

タスク知識は当該工程内で焦点となりうる話題の集合として用意する。各話題は工程名を根とする木構造の節に配置される。節間の枝は話題間関係の種類をデータとして持つ。この様子を図 1に示す。

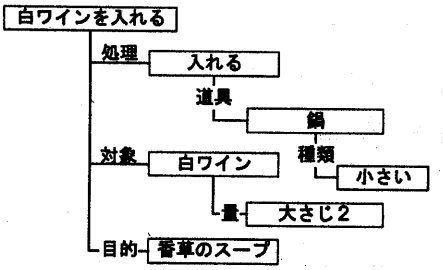

図 1 タスク知識の例

ユーザー発話があった場合にはオペレータによって判断し指定された話題へ焦点を移す。ユーザー発話がない場合には、現焦点との関係、過去の発話履歴、工程内での重要度などから工程内全話題についてスコアを求め、これによって次焦点となる話題を選択する。

2.2.発話生成部

新しい焦点への遷移が行われると、焦点を中心とした一定範囲内の話題から、過去の発話履歴などを考慮して発話対象話題が選択される。発話生成部では選択された話題の内容と関係に基づいて発話文を生成する。文の生成にはテンプレートを用いる。選択された話題セットに最も一致するテンプレートが選択される。

3.実験方法

実験タスクは調理作業とした。理由は用語が安定している、身近な作業で違和感がない、結果が具体的ですぐに分かるのでタスクへの積極性が期待出来る、などである。料理はアボカドの温製半熟卵添えベアルネーズソース[4]という、一般の人はまず調理経験がないと思われるフランス料理を選んで、事前知識レベルの統一と発話量の増加をはかった。

実験場所は業務用ではあるが通常の厨房を用いる事で実験室的雰囲気を避けた。実験スタッフは進行係り、オペレータ、記録係兼サブオペレータの 3 人である。実験環境の概要を図 2に示す。

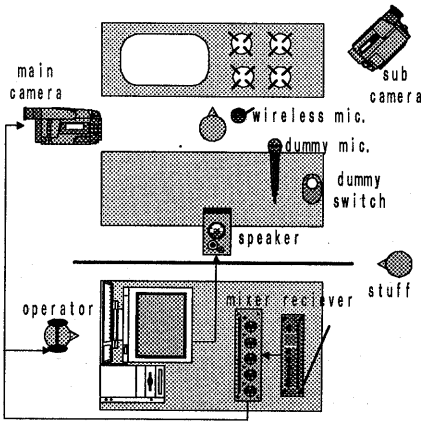

図 2 実験環境概略

被験者へには実験について、音声で調理のナビゲートをする新し器具の試作品のモニターと説明した。購入して自宅の台所で使っていると思って使用してもらうよう指示した。そのため必ずしもシステムの指示通りにする必要はなく、自分のスタイルで進めて構わない。実験前にはメニューの簡単な説明以外、作り方に関する知識はまったく与えず、完成写真を見せるにとどめた。

実験中は視覚的情報はなく音声による情報提示のみによってナビゲートした。また、オペレーターもヘッドフォンを通したユーザーの発話に反応するのみで、衝立てによって視覚による状況把握などは出来ない状態においた。「終わりました」「次は」などの次工程の説明を要求するユーザー発話を受けると、システムは次工程の説明を始めるが、その間あるいはその後ユーザーはその工程について自由に質問する事が出来る。予め用意されているタスク知識でこたえられない場合は「すみません、わかりません。」というシステム発話が返される。主な実験条件を表 1 実験条件概要に示す。

表 1 実験条件概要

タスク	フランス料理調理作業	
	調理時間	約 30 分
	平行作業	2〜3
	工程数	約 35
被験者	30〜50 代の主婦 20 人	
実験環境	業務用厨房	
事前教示	調理器具試作品のモニター	
制御条件 1	システム発話の長短	
制御条件 2	発話スイッチの有無	

2 つの実験条件を変化させて違いを観察した。1 つはシステム発話 1 文に盛り込む話題数すなわちシステム発話の長さで、長い発話では 1 工程内で必要な情報をほぼ最初の 1 発話で説明する。これに対し焦点と隣接の話題のみからなる短い発話では数単語程度の断片的な文に区切って説明する。インターフェースの形としては前者はバッチ的、後者は対話的なものとなった。

もう一つの制御条件はユーザーが発話する際の制約で、一方は胸元のタイピン型ワイヤレスマイクから自由に発話できるものとし、一方はスイッチを押してから机上のマイクに向かって発話するものとした。ただしこのスイッチとマイクはともにダミーである。

4.実験を通して得られた知見

実験は対話データを得る事が目的であったが、それ以外にも興味深い観察結果が得られたので報告する。

ダミーマイクの有無とは無関係に、ほぼすべての被験者がシステム発話が聞こえるスピーカーに向かって発話していた。図 2 実験環境概略からわかる様に、被験者がシンクやコンロに向かって作業している場合、スピーカーは背中側にある事になるが、この場合でも振り向いて発話していた。このことから人間は発話に際して、やはり目にみえる対象を求めることが示唆される。

システム発話の長さによる違いとしては、長いシステム発話によるバッチ的インターフェースでは、作業時間も短く効率的な情報伝達がなされた。これに対し短いシステム発話による対話的インターフェースでは「賢い」「かわいい」などの印象が持たれている事が実験後のアンケートよりわかった。

5.ユーザー質問に関する簡単な分析

収録された対話データは全てテキストへの書き起こしを終わっているが、詳細な分析はこれからである。ここではユーザーからシステムに対する質問発話に関する簡単な分析結果を報告するにとどめる。なお、言語現象については別途報告を行う。

5.1.話題の詳細化

ユーザーからシステムへの質問 408 発話にその焦点の種類を表わすタグを付与した。タグはシステム発話に用いたデータ構造に準じて図 3のような階層的なものを用いた。

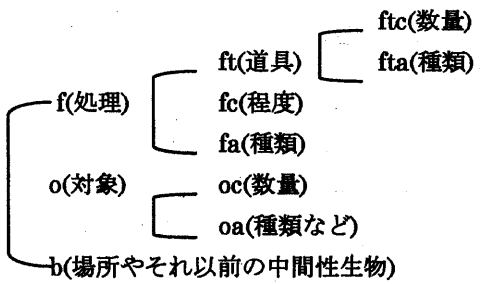

図 3 ユーザー質問発話に付与した話題タグ

このタグの木構造において、ユーザー質問発話の焦点の前焦点からの関係、すなわち話題木上での焦点遷移パターンを分析した結果が図 4、図 5である。

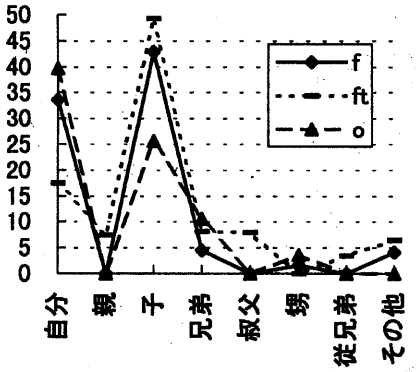

図 4 末端以外の話題からの焦点遷移

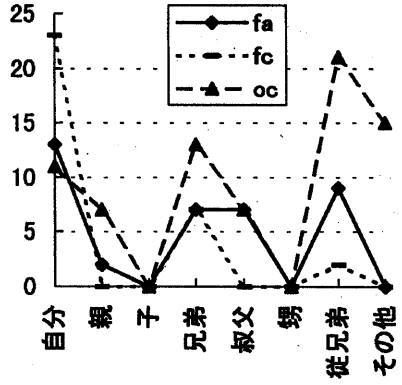

図 5 末端の話題からの焦点遷移

これらのグラフから末端以外の話題からは子へ、末端の話題からは兄弟や従兄弟の関係にある話題への焦点移動が多いことが読み取れる。このことから焦点は話題の詳細化の方向へ移動して、末端まで来ると戻って別方向の詳細化へ向かうという遷移パターンがうかがえる。

5.2.再質問の理由

既出の話題が再び焦点となる再質問は全体の 39%であった。このような再質問に至ったユーザーの事情(背景現象)とその原因を分類した結果図 6 再質問の背景現象と原因がである。

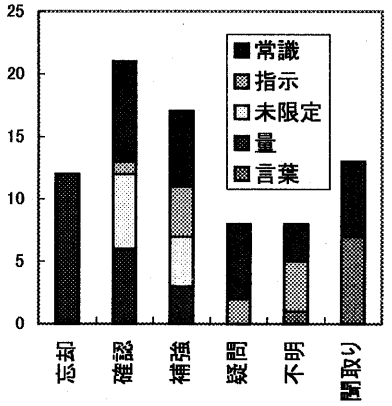

図 6再質問の背景現象と原因

グラフの横軸は再質問の背景となった現象を言語、音声、ビデオなどのデータから判断した分類であり、その意味を表 2に示す。

表 2 再質問の背景原因の分類

忘却	一度受理した情報を他の作業の間などに忘れてしまった
確認	受理した情報を確認する
補強	受理した情報だけでは足りないためさらに情報を要求する
疑問	提示された情報は一応言葉としては受理したが納得がいかない
不明	音としては聞き取れたがその文脈での言葉の意味が理解できない
聞取り	音自体が聞き取れていない

一方原因については表 3のように分類した。

表 3 再質問の原因の分類

常識	提示情報がユーザーの常識とずれている
指示	システムが使った名称がユーザーに受け入れられていない
未限定	システムが使った言葉では十分に意味が限定できず曖昧性が残る
量	大きさ、程度などに関する情報
言葉	音韻的、意味的に理解困難

分類された背景現象にはグラフの左から右へ、システム提示情報の受理性が低くなる順位があると考えられるが、この情報の受理性と原因の関係について考察する。受理性が高い場合、量にかかわる話題が再質問されることが多い。これは数量は受理しやすいが記憶しにくい情報であるためであると考えられる。逆に受理性の低い場合の再質問は卵黄、温製など聞取りにくい言葉や、エストラゴンなど耳慣れない言葉などに起因することが多い。

また、どんな場合においても常識のずれや指示詞の問題から発生する再質問は多い。これはユーザーモデルを把握し、そのモデルの確からしさを認識することが対話システムにとって重要であることを示す。

常識とずれた例としてはアボカドを温める、アボカドをペーパータオルで拭く、アボカドと卵を同じ鍋で加熱する、などであった。ソースを作る過程での「濾す」という動作も常識とずれるというほどではないが、受理性の高い再質問として現れている。また逆に自分の常識に強く引っ張られた例として「ボウルで冷水で冷やす」の「ボウル」が「氷」に聞き間違われたものがあった。

名称が確定的でないものをどう表現するかという問題は、このタスクでは中間生成物の表現に集中した。つまりさまざまな材料を混合してソースを作るとき、いつの時点からそれをソースと呼べるかという問題である。これを回避するために「鍋に混ぜてください」という表現も用いたが、鍋が複数ある場合にこれでは十分限定できなかった。

6.おわりに

実際に調理行動を行っている人を対象に、音声対話インタフェースを通して機械から実作業に必要な情報を得る実験を行い発話データを収集した。タスク、被験者、環境、システムなどに現実的要素を盛り込む事によって、実使用状態に近い発話データを得る事が出来た。

発話データを書き起こし話題に関する簡単なタグ付けを行った。詳細化に向かう焦点遷移のパターンと再質問の原因について大まかな傾向を把握した。

今後更に詳細な分析を行い実際の音声対話インターフェースの設計に活用する予定である。

[1] 望主,酒寄,小島,佐藤,「ナビゲーション対話における省略文の分析」,情報処理学会研究報告 音声言語情報処理,SLP-13-5,(1996)

[2] 望主,佐藤,酒寄,小島,「実作業を伴う音声模擬対話システムのユーザー発話現象」,言語処理学会第3回年次大会,(1996)

[3] 望主,酒寄,小島,佐藤,「実作業に伴う対機械対話の発話現象」,情報処理学会研究報告 音声言語情報処理,SLP-17-5,(1997)

[4] Matt Kramer, "Making Sense of WINE", 白水社,p205,(1996)