

協調問題解決タスクの対話収集と発話の機能の分析

中里 収 森川 恵美 白井 克彦

早稲田大学 理工学部

現段階で実現されている音声対話システムの多くは、タスク遂行上必要となる情報があらかじめ決まっているような簡単な課題に限られている。利用者は、システムに質問したり、質問に答えたりしてタスク遂行上必要な情報を伝える。しかし、私たちの実際の会話の中で観察される発話形態の中には、単にタスク遂行上必要な情報のやりとりだけでなく、対話自体をスムーズに進めるための様々な情報が含まれている。対話システムにおいて、自然で柔軟な発話を生成できるような発話生成機構を構築するためには、人が対話の状況の中で、どのような要因によって発話形態を使い分けているのかを明らかにする必要がある。本稿では、種類の異なる3つの対話データの例を示し、対話データの収集の方法と、データの分析方法について議論する。

Collection of Cooperative Problem Solving Dialogue and Analysis on Functions of Their Utterances

Shu NAKAZATO Emi MORIKAWA Katsuhiko SHIRAI

School of Science and Engineering, Waseda University

In this paper, we show three examples of collected dialogue and discuss the method of collection and analysis on the dialogue. Most of the former spoken dialogue systems can deal with only simple tasks in which the required information has determined. In those systems, responding to the systems' questions, the user provides the information. However, in more actual conversation made between human speakers, there are not only the information that is necessary for the task, but more various information needed to go on the dialogue. To build a spoken dialogue system which can realize natural conversation, we must analyze human-human communication and functions of their utterances.

1. はじめに

音声認識、音声合成技術と計算機能力の向上に伴い、音声インタフェースは実用の段階に来ている。しかし、人間のように対話をするシステムは実現されているとは言いがたい。

現段階で実現されている音声対話システムの多くは、タスク遂行上必要となる情報があらかじめ決まっているような簡単な課題に限られている。利用者は、システムに質問したり、質問に答えたりしてタスク遂行上必要な情報を伝える。

しかし、私たちの実際の会話の中で観察される発話形態の中には、単にタスク遂行上必要な情報のやりとりだけでなく、対話自体をスムーズに進めるための様々な情報が含まれている。

対話システムにおいて、自然で柔軟な発話を生成できるような発話生成機構を構築するためには、人が対話の状況の中で、対話自体を成立させているそのような情報に注目する必要がある。

音声によるコミュニケーションは、私たちの日常生活の様々な場面でなされているが、それを研究対象とし、データを集めようとするとき、いくつか考慮すべき点が出てくる。

対話を収集するためには、他人の会話を盗聴するわけにはいかない。会話の状況を設定し、模擬対話の収録実験を行なう必要がある。この対話状況の設定は、対話内容に影響を与える。例えば、目上の人に対する発話は言葉使いが丁寧になるし、親しい者同士の対話では、共有している知識を前提として対話が進められることが多い。このように、どのような人がどのような状況で、どんなタスクで話をするのかによって、対話内容は異なってくる。

また、特に機械との対話を考えていく場合は、人間の自然な音声対話に存在する制約以外の制約が問題となるので、両者の制約を分けて考えることとそれらの間の相互の問題とを議論する必要がある。

日常何気なく行なわれる対話をあらためて考えたとき、発話される言葉はかなり状況に依存している。意識するしないにかかわらず、そこでは対話する状況に合った発話が行なわれている。このときの状況から与えられる制約には、様々なレベルと内容が存在する。本稿では、それらの制約を1) 物理的環境、2) 対話者の属性と関係、3) 対話の目的、という3つの視点からとらえることとし、3章以降では収集してきた対話データをもとに、発話形態の違いについて触れる。6章では発話の機能タグを用いた分析の例を報告する。

2. 対話コーパス収録で考慮する点

対話コーパスを収集する際、全く制約をかけない対話を対象とするのが理想のように思われるが、後で分析することを考慮すると、全く制約をかけない対話は扱いづらくなることもある。対話自体が理解できなかったり、対話状況の再現が困難であったりする。分析の用途に合った対話を生じさせるように、研究者は意図して条件を設定する。筆者は、対話内容に影響する制約を大きく次の3つに分けて考えている。

2.1 物理的環境の制約

対話内容に影響を与える制約の1つは、対話が行なわれる場面での物理的な環境である。音響的な要因はもちろん、対話者の位置関係や、通信に用いるデバイスなどもこの制約に含まれる。分析しやすいデータを収集する場合には、それらの影響を排除するために、静かな部屋で、相手と同じものが見えない状況にすることが多い。

2.2 対話者の属性と関係の制約

対話者（たいていの場合2人を想定することが多い）の属性と、社会的な関係も対話内容に影響する要因となる。年齢、性別、出身地などの個人の属性と、教師と学生、知らない者同士といった対話者間の社会的な関係が対話形態に表れる。普段から知っている者同士の場合には、共有知識が多いため、第三者が対話を聞いても分からない場合があるので、注意が必要である。

2.3 対話の目的（課題）の制約

自発的で自然な対話を収集するためには、実験で与える対話の課題も慎重に選ばなければならない。対話が自然でなく、収録のための「演技」にならないようにするには、被験者が本当にやる気になるような課題が望ましい。著者らは、後に述べるような協調問題解決タスクを用いているが、ゲーム性のある、目的のはっきりした課題であると、実験室環境下においても被験者が課題に熱中できるからである。しかし、日常から離れた、あまりに人工的な課題を与えてしまうと、対話を収集する際に課題自体の説明が必要になり、その説明の仕方が対話内容に影響することも考えられるので、特に注意が必要である。

3. 対話データの収集（1）フィールド調査

著者らは、今までに、実験室ではなく、対話が本当に必要となる条件での実対話のデータも収集してきた[1]。被験者には特に制約を設けないため、対話内容、言葉使い等の統制がとれるとは限らない。そこで生じる対話は様々な要因の影響が表れるが、日常の発話現象を調べるにはこのフィールド調査が有効である。

3.1 経路決定の対話

対話収録に協力してもらった被験者は同じ研究室の学生で、お互いによく知っている者同士である。実際に自動車に乗ってドライブする状況において、目的地を設定し、そこで生じる経路決定の対話を収録した。以下にその例を示す。

対話例

U: 横浜のベイブリッジに行きたいんですけど、どう行ったらよろしいのでしょうか？

S: うーんと、ドライブですか？

U: あ、はい。

S: ドライブ。じゃあ、えーと、ベイブリッジを見るというのが目的だね。

U: そうですね。

S: うーん、じゃあ、時間は何時でもいいということ。

U: そうですね、ただ、できるだけすいている方がいいんです。

S: あ、はい。それじゃあ、まず、ベイブリッジっていうのは、首都高速の上にあるんですけど、

U: はあ。

S: まあ、どこからか首都高速に乗らなければいけないんだけど。ええ、高速を使いたいですか、それともなるべく使わないようにしたいですか？

U: うーん、なるべく使わない方がいいです。

S: そうか、じゃあ、そうすると、今は、現在地は高田馬場だから、

(以下略)

Uは運転手、Sは情報を持ったナビゲーターの役割をしている。不要語、省略、あいづち、割り込み等が頻繁に現れる対話である[1]。

4. 対話データの収集（2）人間対機械

人間対機械の対話で、ある程度スムーズな対話データを収録するために Wizard of OZ 方式などと呼ばれる模擬対話実験が行なわれている[2]。これはユーザーに分からないように、機械を装った人間が話し相手を務め、そこでの対話内容を収集するといったものである。システム製作中の段階で、どのような発話を許容するかを決める上では有効なデータが得られる。

4.1 カーナビゲーションタスク

被験者はこのシステムを初めて使う大学生13人であり、それらを6人と7人との2群に分けた。それら被験者の2群は実験のタスクは全く同じものであり、教示だけが異なっている。どちらの群にも、タスク終了後のアンケート調査を目的だと伝えた上で、ひとつの群には音声対話システムと対話すると教示し、陰で人間が操作していることがわからないように配慮した。もう一方の群にはコンピューターを通じて、人と対話することによってタスクを行なうようにと教示した。この実験では前者を「対システム群」、後者を「対人間群」と呼ぶ。

本実験では、各被験者に対して、練習は行わず、実験の前にシステムの使用法についてのビデオ（約6分）を見せ、口頭で使い方を説明した。このビデオの中の対話内容はごく単純なもので、不要語・いい直しなどは含んでいないものとした。図1に収録システムのレイアウトを示す。

実験で行なうタスクは、現在地から目的地までの経路設定である。システムが示す経路の候補の中で、最もすいている道を選ぶ。このとき、システムには3本の候補を持っていることを伝えてあった。

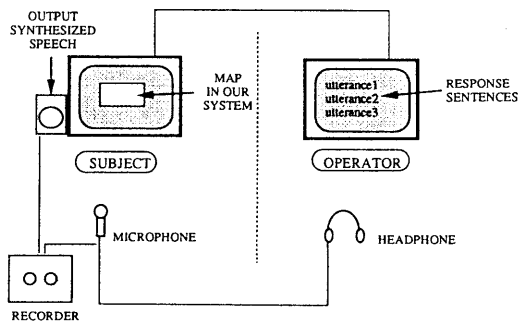


図1: 対話収録システム

システムからの応答文は、システム役を務める人間が、被験者の意図に合わせて、あらかじめ用意してある発話文の候補を選び、合成音で出力した。発話の候補は対話的にタスクを行なう上で必要となる文を約30文選びだしたもので、それらを組み合わせることでシステムの発話とした。

4.2 結果

録音テープをもとに全ての発話を書き起こし、集計、分析した。13人の被験者の総発話数は1039発話であった。

対話例

S：目的地はどこですか。
U：大崎です。
S：目的地は大崎駅ですね。
U：はい、そうです。
S：例えばこのような経路があります。
U：えっと、地図を大きくして下さい。
S：はい、分かりました。
S：これでいいですか
U：はい、いいです。
U：えっと、地図を下に移動して下さい。
S：はい、分かりました。
S：これでいいですか。
U：はい。
U：えー、どこが混んでいますか。
S：西麻布の近くが混んでいます。
(以下略)

1人の被験者が1回のタスクを終了するのにかかった平均の発話数は、システム群で25.8回、対人間群で27.3回であり、2群間に有意な差はみられなかった。

2群間で有意な差が見られたものは不要語の回数、ポーズの長さである[3]。

対話例を見ても分かるように、先ほどの経路決定の対話とはだいぶ異なる内容となっている。不要語、省略などは見られるものの、あいづち、言い直しなどがほとんど生じない。また、1発話ごとにポーズが現れ、発話権(ターンテイキング)がはっきりしているのも特徴である。

5. 対話データ収集(3) 人間対人間

協調作業中の会話では、比較的個人の思考過程が言語的に現れやすく、また、被験者は実験室状況下においても、課題に熱中すると自然な発話になりやすいことが知られている[2]。

タスクとしてはクロスワードを対話を通じて解く課題を取りあげた。クロスワードパズルとは、マスの区画の中に、与えられた言葉のヒントから推理した語を縦横に交差させて埋めていくパズルである。よく知られているパズルであり、データ収集の際に教示が簡単であり、被験者は課題に熱中しやすい。

5.1 クロスワードパズルタスク

被験者は大学生16人、お互い知っている者同士のペアである。解き方についての指示は与えず、自由に課題を進めてもらう。ただし、パートナーとの意思伝達は音声のみとした。今回のデータ収集では、8×8文字の同じクロスワードパズルを2人に与えた。この時、縦と横のヒントを2人に別々に与えた。そうすることによって、2人の被験者はヒントをお互いに伝え合い課題を進めていかなければならない。この課題の特徴としては、双方にほぼ同じ量のヒントがあるため、情報の流れが双方向であること、課題を達成するのに、ある程度知識が要求されること、などである。図2に収録の様子を示す。

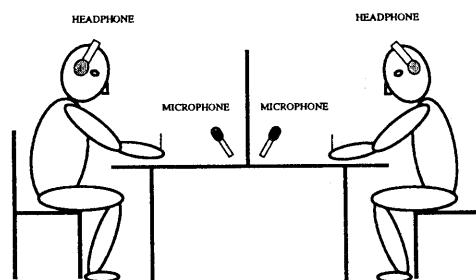


図2: クロスワードパズルタスク対話の収録

以下にその対話例を示す。

対話例

A：たてのね、
 B：うん。
 A：7。
 B：が、なんじゃ？
 A：好評につき、この商品は全て、ホニャララです。4文字？
 A：売り切れじゃない？
 B：売り切れ？
 A：7の縦。
 B：売り切れ。んーとね、じゃあ、4の横。
 A：4の横？
 B：磯野家の長男。かつお？
 A：かつお？
 B：かつお。
 A：フランス語でマロン、13の縦。
 B：13の縦。
 A：フランス語でマロン。
 B：栗？
 A：栗。
 (以下略)

5.2 分析と考察

8組のタスクでの総発話数は1583発話、1回のタスクでの1人の平均発話数は99.8発話であった。

あいづち、言い直し、不要語、省略、といった発話現象が、このタスクでは頻繁に生じていた。

このタスクの場合、1つの答を出すまでが1つの単位となっており、対話全体が構成されている。以下この単位を対話ユニットと呼ぶ。

このタスクの場合、これらのユニットが繰り返されて全対話が成り立っている。ユニットの遷移は1タスク中に平均30.6回であった。

ユニット同士の関係は入れ子構造になっているわけではなかった。このクロスワードパズルタスクの場合、問題同士に全体・部分という関係がなく、どの問題からも着手できるという問題の性質が対話の構造にも影響したと思われる[4][5]。

6. 発話の機能に着目した分析

この対話コーパスの中の4対話について、表1に示すように、発話の機能という観点から各発話に対してタグを付けた。タグの種類は結局24種類必要であった(表2)。

表1: 対話例

発話表現	発話機能
A：えーと	呼びかけ
B：はい	あいづち
A：タテの1番で	話題提示
B：はい	あいづち
A：ヒントが徳川家の紋所 これアオイだね。	問題提示 答
B：そうですね	答に同意 ユニットの区切り
B：横の7番が	話題提示
A：うん	あいづち
B：おかわりです	答
A：はい	答に同意 ユニットの区切り

ここで発話の機能とは、単語や句から機械的に定めたものではなく、対話の文脈も考慮して、意味的に付与したものである。1発話に対して複数のタグを許し、また1単語に対しても複数のタグが付与されることもある。

6.1 機能タグを用いた分析

この、クロスワードパズルタスクで用いたタグは、一般的なものではなく、このコーパスを理解する上で必要となるものを付加しており、タグの種類はこのタスクに依存している。他の種類のコーパスにおいても同じ方針で、タスクに依存したタグづけをしていくことにより、タスクに依存するタグと依存しないタグを分類できると考えている。

これらの機能タグを用いて、出現頻度や遷移確率などを調べることができ、他のタスクとの比較が定量的に出来る。また、対話コーパスの中の機能タグがどのように移り変わっていくのかを調べ、次発話の予測や、対話の流れのモデル化が可能だと考えている。

表 2: 機能タグと各タグ出現回数 (4 対話中)

タグ	出現回数
呼びかけ	3
あいづち	41
話題提示 (場所の指定)	117
問題提起 (ヒント)	57
答	112
答に同意 (ヒントを知っていて)	33
異議	2
質問, 問いかけ	40
聞き返し	11
答の確認	73
場所の確認	26
ヒントの確認	10
答に同意 (ヒントは聞いてない)	9
質問に対する肯定	9
答の訂正	3
独り言 (思考中であることを示す)	6
独り言	13
以前のユニットに対する異議	3
情報・知識不足	9
懐疑的同調	4
問題解決のストラテジーの提案	8
ストラテジーに同意	4
復唱	83
ユニットの区切り	236

7. おわりに

以上の異なる条件での対話データを3種類示し、それらの発話形態の違いが確認できた。人間と機械の対話では、ターンテークキングがはっきりして分析がしやすい反面、あいづち、聞き返しなどがほとんど現れず、自然な対話としては不十分であることが分かった。

人間同士のクロスワードタスクの場合には、フィールド調査で確認できたあいづち、復唱による確認などが現れ、言語的な情報に加え、対話の制御や情報の重要度を対話場面に提示していることが確認された。

対話を自然な形で成立させるために、対話者はそれらの「対話をうまく進める上で有用な情報」をお互いに共有し、「話す」「聞く」の時間的同期や「何を話すか」「何を聞こうとしているか」の発話内容の共有を可能にしていると考えられる。音声対話システムが自然で柔軟な発話を生成するための機構として、ここで示

したような現象に注目し、取り入れていくことが今後の課題となるであろう。

参考文献

- [1] 菊池 英明, 小林 哲則, 白井 克彦: “自然な模擬対話を収録するために” 音声対話のモデル化とその機械処理に関する総合的研究 研究成果報告書, pp. 109-118, (1992)
- [2] René Amalberti, Noëlle Carbonell, Pierre Falzon: “User representations of computer systems in human-computer speech interaction” *Int. J. Man-Machine Studies*, 38, pp. 547-566, (1993)
- [3] 中里 収, 亀山 晋, 白井 克彦: “音声対話システム使用時のユーザー発話の分析” 第8回 人工知能学会全国大会論文誌, pp.93-96. (1994)
- [4] 中里 収, 田中 修一, 白井 克彦: “協調問題解決における発話の役割”, 音講論 3-P-18, pp.197-198, (1995).
- [5] 中里 収, 亀山 晋, 田中 修一, 白井 克彦: “協調作業における音声対話データの収録と分析” 音講論 1-Q-8, pp.157-158, (1994).