

スペシャル・セッション [対話音声の分析, 合成, 認識]
音声対話システムに対する利用者の発話の解析*

1-1-11

◎伊藤克巨 △田中穂積 (東京工業大学)

速水 悟 (電子技術総合研究所)

人間は、相手のふるまいによって態度を変える。ひとくちふるまいといっても、それを構成する要素は、口調などの表層的なものから、信念などの深層的なものまで多様である。したがって、利用者が計算機に対してどのような言い回しをするかや計算機の発話をどの程度認識・理解するかは、システムの応答の内容や応答にかかる時間、その他のイメージ(計算機の「表情」など)によって大きく変化すると考えられる。しかし、現時点ではそれらを研究するための材料すらほとんどない。

こうした問題を調べるために、われわれは、音声対話システムを構築した[1, 2]。本稿では、そのシステムで収録した利用者の発話を解析した結果を報告する。

1 対話システムの概要

本稿で述べる対話システムは、つくばから東京方面への交通案内をおこなう。答えられる項目は、目的地までの経路・所要時間・運賃だけで、目的地も7つしか用意されていない。しかし、これらの制限を加えることで、ほぼ実時間で利用者に応答することを可能にしている。

このシステムでおこなえる対話の例を以下に示す。

システム: どこに行きますか。
利用者: 新宿に行きたいんですけど。
システム: 新宿ですか。新宿へ行くには、常磐線と山手線に乗ります。
この経路について、のりかえ・所要時間・費用をお尋ね下さい。
利用者: どこでのりかえるんですか。
システム: 上野です。
他に何かございますか。
利用者: どのくらいかかりますか。
システム: 値段ですか、それとも時間をお尋ねですか。
利用者: 時間です。
システム: 1時間20分です。
他に何かございますか。
利用者: 上野から新宿までは何分かかりますか。
システム: 上野から新宿までは20分です。
(以下略)

システムが認識できる発話は以下に示すテンプレートの形で記述される文法で定義される。

<駅名><から><駅名><まで><どれくらい><かかる>

<駅名> = 東京 | 荒川沖 | 上野 | 大岡山 | ...
<かかる> = かかるでしょう | かかりますか | ...

文法は、対話システムの試験版を用いて、利用者との対話の様子を収録したのべ22名の42対話からなるデータをもとに作成した。また、文法を逸脱する発話に対しては、認識結果のスコアがしきい値以下ときには、その発話を棄却しているが、他には、間投詞や未知語に対する特別な処理はおこなっていない。

本稿では、このシステムを利用して収録した対話のうち、とくに利用者のいい回しに重点をおいて解析した。

*Analysis User Responses To A Speech Dialog System by ITOU Katunobu, TANAKA Hozumi (Tokyo Institute of Technology) and HAYAMIZU Satoru (Electrotechnical Laboratory)

2 収録条件

設定した対話の形態を以下に示す。

- システムと利用者が交互に発話
- 利用者の発話は一回につき一文
- 認識結果の音韻列と漢字かなまじり文の応答を示す
- 利用者の発話が終了してから応答する
- 一つの対話で一つの目的地しか扱わない
- システムが主導権をとる
- 利用者は協調的であると仮定する
- 現在の状態を明示的には示さない

対話は、以下のような形で記録した。

- 発話音声(切り出し後の波形データ)
- 認識結果(1位のみ)
- 応答内容
(以上の項目については光ディスク上に自動収録)
- 性別
- 年齢層

収録する際に仮定した対話の場面を以下に示す。

- 利用者は常磐線荒川沖駅にいる
- 利用者は目的地への行き方を知らない

これらの仮定と、システムにどのようなことができるかを知らせるために、簡単な利用者用ガイドを提示した。また、説明者が見本の対話を行うこともあったが、利用者に明示的に認識可能な例文や語彙を示すことはなかった。対話は、まわりにもかなり人がいる状況で収録しており、利用者が手に持つ型のマイクを用いてシステムとの対話をおこなうようにしたため、まわりの人の話し声が利用者の発話に重なることもあった。

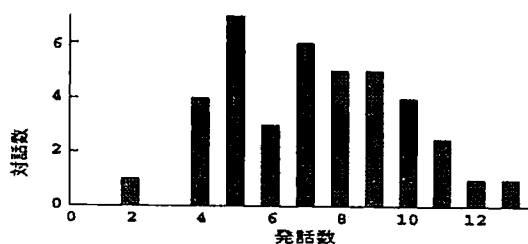
3 解析結果

1で説明したシステムを用いて、1991年11月8日から15日にかけて、38名(男性33名、女性5名)の44対話を収録した。被験者の年齢層は、20才代から60才代までとなっている。また、日本語を母国語としない被験者も1名含まれている。

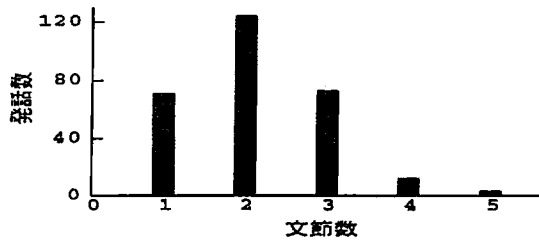
このうち、システムについて詳しく知っている被験者による対話と対話システムが動作しなかった対話を除いた39対話(285発話)について解析した。解析したデータのうち、システムとの初めての対話は33で、2度目の対話が5、5度目の対話が1あった。複数回の場合は、ほとんど間をおかずに収録した場合と、間をおいた場合の両方があった。

3.1 発話数・文節数

まず、はじめに、一対話あたりの発話数の分布を示す。



平均は、7.3 だった。
つぎに、一発話あたりの文節数の分布を示す。



平均は、2.1 だった。
文節数 1 のものが多いため、平均が小さくなっている。これは、この対話が、「さようなら」などのあいさつで終了するようになってきているのが原因で、あいさつをのぞけば、単語や一文節からなる発話はほとんどなかった。

3.2 文法外の発話

発話が用意した文法で扱えない要因として、(1) 未知の言い直しを含む (2) 間投詞を含む (3) 言い直し・言い淀みを含む、の 3 つがあげられる。この 3 点について観察した結果を述べる。

3.2.1 未知の言い直し

未知の言い直しを含む発話は、全部で 113 あった。これは、全体の 40% にあたる。ここで、意味的ではなく、表層として辞書に存在しない自立語を未知語とよぶことにする。未知語を含む発話は 35 あった。これは、全体の 12% にあたる。その他は、語尾の活用が用意されたパターンでなかったもの、語の組み合わせが用意されてなかったもの、語順が用意されてなかったものであった。

未知語は全体で 18 種類あらわれた。未知語には、以下にあげる種類の語が含まれていた。

- 発音のゆれがある語 (3 語: 「いい」(「よい」は辞書に含まれる))
- 辞書に含まれる名詞の同義語 (2 語: 「お金」(「料金」は辞書に含まれる))
- 辞書に含まれる副詞の同義語 (2 語: 「いかほど」(「どのくらい」は辞書に含まれる))
- 終了のあいさつ (6 語: 「サンキュー」)
- システムが答えられない概念を持つ語 (1 語) 「確認」という語が発話されたが、システムは一度応答したことを確認する機能はもっていなかった。
- その他 (文法で扱える文をいい換えるときに使われる副詞・動詞) (4 語)

また、未知の言い直しと、利用者のシステムとの対話回数との関係を以下に示す。

	初めて	2 度目以降
未知の言い直しを含む発話	91 (39%)	22 (45%)
全体の発話数	236	49
未知の言い直しを含む対話	9 (27%)	3 (50%)
全体の対話数	33	6

2 度目以降のデータの量が不十分であるが、初めて接する場合の結果を見ると、特定の対話 (話者) によって未知の言い直しが発話されることが多い傾向が見られる。利用者は、自分の発話に未知語が含まれているかどうかを、システムの応答から明示的に知ることができないので、未知語を発話するかどうかは、システムに対する慣れよりは、個人差の方が大きいようである。

3.2.2 間投詞

間投詞を含む発話は、全部で 15 あった。これは、全体の 5% にあたる。出現した間投詞は以下にあげる 6 種類だった。

えー、え、えと、あー、あ、えーと

間投詞は出現する全ての発話で、発話の最初に一度だけあらわれていた。

3.2.3 言い直し・言いよどみ

言い直しを含む発話は、全部で 6 あった。これは、全体の 2% にあたる。これらの発話は、以下の 3 種類に分類される。

- 言いよどみ (4 例: 「じ、時間です。」)
- 言い誤り (1 例: 「もう一度、お金の、お金を教えて。」)
- 強調 (1 例: 「値段だつてば、値段。」)

3.3 認識率

用意した辞書と文法で扱えた発話は、全部で 161 あった。これは、全体の 56% にあたる。これらの発話に対する認識率は、63% だった。また、認識の正誤にかかわらず、発話に対する応答が正しかったものの割合を理解率と定義すると、文法で扱える発話の理解率は 83% だった。

全体の発話に対する理解率は 64% だった。未知の言い直しを含むもの場合は理解率が 41%、間投詞を含むもの場合は 26%、言い直しを含むもの場合は 33% であった。

4 まとめと今後の課題

実際に動作するシステムを用いて、計算機に対する利用者の発話を収録し、言い直しに重点をおいて解析した結果について報告した。

解析の結果、本システムに対する発話では、未知の言い直しを含む発話は全体の 40% とかなり多かったが、間投詞や言い直しはそれぞれ全体の 5%、2% とそれほど多くはなかった。しかし、今回は、同じ条件下での対話しか収録していないので、この結果がタスクに依存するものか、それとも計算機を相手にする場合の特徴かは、明らかではない。しかし、未知の言い直しを含む発話の多さは、対話システムを設計していくうえで、文法や辞書の構築の難しさを表していると考えられる。

今回収録したデータでは、利用者が話してからシステムが応答するまでの時間、また、システムが応答してから次の発話までの時間が全く記録されていなかった。また、条件がかなり固定化されており、タスクも 1 種類であったので、今後は、様々な条件下での対話を収録し解析していく必要があると考えられる。

謝辞

本稿のシステムで利用している音韻モデルの学習に用いたデータのうち、連続音声資料の一部は日本音響学会の研究用連続音声データベースの一部であり、関係各位のご尽力に感謝いたします。

また、日頃御討論いただく、東工大田中研の皆様、並びに電総研知能情報部音声研究室の皆様にも感謝します。

参考文献

- [1] 速水悟, 伊藤克巨, 田中和世. 音声対話システムの構築とそれを用いた会話音声収集. 信学会技術報告, 12 1991.
- [2] 伊藤克巨. 音声対話システム. 電子情報通信学会技術報告, Vol. SP92-38, pp. 23-30, 1992.