

音声対話データから見た自由発話の言語的性質

樽松 明 山本英明

電気通信大学電子工学科

1 はじめに

実用的な音声入力システムを実現するには、普通に話されている言葉(以下、自由会話と呼ぶ)に対応した言語モデルを作る必要がある。音声や言語の研究を遂行するために、基礎となるデータベースの重要性は、広く認識され、日本語におけるこのための努力がこれまで重ねられてきている[1],[2],[3],[4]。話し言葉のデータベースについては、これまでは、主として模擬会話による対話データ、あるいは利用者とシステムとのやりとりの収録がおこなわれているが、まだデータの規模は小さい。

今後の音声による対話の研究においては、従来の音声や言語の個別のデータだけでは不十分であり、音声対話としての研究開発に役立つ自由発話の会話音声データベースの構築が不可欠であるといえる。音声対話データベースは、自由発話の音声認識・理解、対話の理解、対話システムにおける発話メカニズムなどの研究を想定する場合には、データベースは、統計的データをもとにした対話プロセスのモデル化が可能になるようにしなければならない。これには、音声対話現象を広い範囲でカバーするデータの量の拡大と、対話現象の基本となる表現や談話プロセスを包含した質のよさが必要である。

本稿では、スケジューリングタスクの自由発話音声対話データベースの構築のためのデータの収集とその言語的性質について言語モデルの元となる品詞バイグラム、品詞の出現頻度に着目し、言語的特徴を分析した結果を述べる。

この研究は、文部省科学研究費補助金重点領域研究「音声対話」ならびに国際コミュニケーション基金の援助によるものである。

2 スケジューリングタスクによる音声対話

データの収集

2.1 スケジューリングタスク

(1) 収集方法

音声会話は、接話型ヘッドフォンマイク(SENNEHEISER 410)を使用して、DATテープに収録した。録音は計算機のある部屋で行なった。このため、収集データには若干の雑音がある。話者間の会話以外の情報伝達をなくすため、お互いの顔を見えなくした。音声は2チャンネル独立に録音した(例えば話者AはLチャンネル、話者BはRチャンネル)。会話が自然になるようにするために、会話実験の前に雑談をしてもらった。

(2) 会話タスク

会話は2人の話者がそれぞれのスケジュール表を見ながら、会議等の日程を決めるスケジューリングタスクによって行なった。会話の条件は次のとおりである。

- 話者は大学の学生、職員で、男15人、女3人。1人1~3回会話を行なって貰った。
- スケジュール表は1ヶ月のうち、2週間程度埋まっているものを複数とおり用意した。これを会話毎に適宜選択した。
- 何のための会議かなどの細かな設定は、会話実験の前に被験者どうしで自由に決めてもらう。
- 学生が話者の場合、会話内容が発散する可能性があるので、教官の秘書となったと仮定して会話を行なって貰った。

会話内容の設定は、話者の自由である。会話内容をあまり制約すると、考えることが狭くなって話し方が片寄ってしまうので、カレンダーによるスケジュールという枠組みのなかで自由に話してもらうようにする。一方、枠組みのない全く自由な会話にすると、話題が広がりすぎて音声対話処理

に有用な言語情報が得られないということになる。カレンダーの制約以外には、話し方については常識的な話し方をしてもらうこと以外には自由に発話する。

2.2 会話データの分析

収集した音声データを、漢字かなまじり文に文字表記した。収録したスケジューリングタスクによる自由会話音声データ（以下、スケジューリング対話）を分析した。対話数は13対話である。

(1) 文字表記

データベースをもとに、音声的あるいは言語的に統計処理を行うには、データに必要な範囲のタグ情報を付与することが必要である。書き起こしの表記は、収録された音声データの内容を記述するもので、かな漢字による表記とローマ字およびかな表記を行っている。ローマ字は、日本以外の国でこのデータベースを利用する際の便宜を考えている。ローマ字およびかなによるテキストの表記には、形態素解析でなされる区切りを参考にして文節に近い区分で区切る。

書き起こしには、人によるノイズと雑音などの種々のノイズに相当する音声以外の音も表記する。人によるノイズには、間投詞、咳、吸息、吐息、笑い声、口で鳴る音などがある。雑音には、キーのクリック音、ペンでたたく音、ドアの音などがある。また、会話の途切れの沈黙、言い直し、言い間違いもそれぞれマークをつける。会話の重なりが生じた箇所では、重なり始めた音と終の音に記号を付加する。あいづちの場合はあいづちの入った場所に記号を付加する。

(2) 形態素解析

音声対話データのタグ付けのための辞書定義は、まだ定着しているとはいえない。話し言葉においては、普遍的な言語の定義がむづかしい。特に表現がゆれたり乱れたりするため、文法的な情報のみでは、形態素種別を人間により判定するのも安定にはできない。意味情報も加味していく必要がある。

音声対話データベースをもとにして、形態素カテゴリーの統計的性質を言語モデルとして利用することを考えると、形態素カテゴリーの種類はあま

り少ないと統計量が有効に働かない。また、あまり細かすぎると精度や安定性が保てなくなる。したがって、形態素解析プログラムの品詞体系カテゴリー程度の分類が適当と思われる。

形態素解析は文献 [5] [6] をもとに行なった。品詞の種類は27種類とした。品詞の内容は表3に示す。

なお、話し言葉の会話の音声認識や言語処理をする上で、助詞や助動詞の品詞カテゴリーを、学校文法のカテゴリーわけよりより長いもの（接続されたもの）をとるほうが好ましいという点もある。

例　でしょ（助動詞）う（助動詞）→でしょう（助動詞）

3 会話データの分析結果

3.1 統計量と言語的特徴

会話内容の統計的データがどのように現れるかを調べた [8]。13対話中に現れた単語の延べ数（単語総数）、異なり単語数、品詞バイグラムの種類などの会話データの総合的な結果を表1にまとめた。ここで、発話とは、同じ話者が続けて話した内容とした。発話数はかなりばらついている。短い対話では15発話で終わっているが、長い対話では147発話になった。品詞エントロピーは、品詞の出現確率より求めたものである。表1より、単語総数に対する異なり単語数の割合は、11.1%であった。

次に、品詞バイグラムの上位8位を表2に示す。表中の DUMMY は、その後、もしくは前にある品詞が、文頭もしくは文末であることを示している。ATR 対話データベース [1] との比較をおこなったところ、単語総数と異なり単語数との比、品詞の出現頻度におけるエントロピーは、会話のスタイルによらずほぼ同じであることがわかっていく [8]。また、表2の8つの品詞バイグラムの内、「数詞 接尾語」「感動詞 DUMMY」は、スケジューリングタスクにおいて、顕著なバイグラムのようである。比較的自由に発話する対話データベースであれば、大局的に見れば会話内容による言語的性質の違いは少ないと予想できる。

スケジューリング対話の品詞の出現頻度を表3に示す。

表1 スケジューリング対話データの統計量

対話数	13
発話数	1025
単語総数	6739
異なり単語数	750
1発話の平均語数	6.5
品詞バイグラム数	285
品詞エントロピー	3.85

表2 品詞バイグラム

バイグラム	出現率 (%)
DUMMY 感動詞	7.7
普通名詞 格助詞	4.8
感動詞 DUMMY	4.0
数詞 接尾語	3.8
助動詞 終動詞	3.6
格助詞 普通名詞	3.3
格助詞 本動詞	3.2
助動詞 接続助詞	3.0

表3 品詞ごとの異なり単語数の割合 (%)

固有名詞	0.68	サ変名詞	0.47
形容名詞	0.37	普通名詞	14.02
数詞	4.85	代名詞	2.26
本動詞	6.62	補助動詞	2.33
形容詞	1.05	副詞	4.88
連体詞	0.77	接続詞	1.41
感動詞	11.39	間投詞	0
助動詞	13.90	格助詞	11.50
準体助詞	1.47	係助詞	3.16
副助詞	0.65	並立助詞	0.03
接続助詞	6.40	終助詞	4.51
連帯助詞	0	引用助詞	0.03
接頭語	0.88	接尾語	6.19
その他	0.18		

本動詞が発話中に1つ含まれている割合は約50%である。本動詞の種類は、119種類である。

一方、英語の be 動詞の表現に対応して日本語で用いられる助詞「は、が、の、を、に」の出現は、総計 927 で、これらが出現単語数に占める割合は約 10% である。その内訳は、「は：229」「が：152」「の：406」「を：58」「に：82」である。文を構成する主な要素は、名詞、助動詞、格助詞であることがわかる。

3.2 話し言葉特有な語、表現

文末が完結していないで、名詞や、途中でとまっているものがある。また、現在の若い人の間で話される話し言葉として、短縮語や、本人同士がわかるローカルな用語があらわれる。これらは、未定義語となる。感動詞や副詞の使い方には、個的な癖がある。

3.3 会話の文体

会話の文体は、会話の相手が誰であるかによって大きく変化する。たとえば、親しい友人との会話の場合、発音が曖昧になったり、文法がいい加減になったりする。一方、目上の人物や初対面の人と話す場合は、発音、文法ともに比較的性格になる。収集した対話データの中から例を示す。

友人との会話

A：あーね。行こ行こ言って全然行っていないよ。
B：そうそうそう。
A：うーんうーん。
B：なかなか。だからどうせ 火、金なんて、学校きてるでしょ。

秘書どうしという設定での会話

B：#h-# 水曜日はセミナーがはいってまして。
A：あー、そうなんですか。#pencil-#
B：午後の4時まで空いてませんので #h-#。#pencil-#
A：あ、じゃー無理ですね。
B：この日もちょっと、うーん。
(#pencil-#：鉛筆が転がる音、#h-#：息を吸う音、吐く音)

4 まとめと今後の研究課題

自由発話の音声対話データからみた、大局的なとらえた言語的性質をのべた。まだデータ数が少ないため、確定的にはいえない点があるが、品詞カテゴリーの出現傾向として、興味ある結果がえられた。

今後の研究課題として、以下の項目があげられる。

1. データの量の拡大

スケジューリングタスクにおける自由発話音声対話データの量を拡大して収集を行う。この際、話者へは、自由発話とはいえ、話者に協力的に話すよう前もって説明をしておくようにして、ある程度まとめた話し言葉会話のデータとする。

2. 付加情報のタグging

対話音声に対する付加情報のタグgingの方法については、できるだけ人手をかけない半自動的方法を検討する。付加情報の記述のフォーマットは、国際的な標準化に対応して、TEI (Text Encoding Initiative) に準拠する。

3. 自由発話の音響モデルの学習

収集したデータを用いて、自由発話の Hidden Markov Model による音響モデルの学習を行い、自由発話認識用の発音辞書を作成する。

4. 言語の統計的性質の抽出

自由発話のデータの分析を行い、言語の統計的性質を抽出する。読み上げ文のものと比較検討する。今後データ量の拡大に伴い、品詞バイグラムや音声認識における言語モデルに有用となる統計量の分析を進める予定である。さらに、日本語対話の自由発話の言語モデルの作成をおこなう。

5. 韻律情報

韻律情報は、話し言葉の会話の特徴を担う情報として、欠かせないものである。特にピッチ周波数と音韻の継続時間は、貴重であるので、韻律情報を付加する。これを、話し言葉の文節ごとの区分わけのために利用する。

6. 発話意図

話し言葉では、話し手の意図が、文末の表現として表わされ、発話意図として、そのタイプを抽出できる [7]。スケジューリングタスクにおける自由発話音声対話では、二人の話者の会話によりスケジュールの相談をするので、発話意図のタイプは聞き手への事実の伝達である「Inform」と、応答やあいずちであ「Response」がほとんどであって、疑問文や質問文は少ない。これは、タスクの性格によるものと思われる。発話意図情報を抽出してこの情報を付加する。

参考文献

- [1] 江原, 井ノ上, ほか, " ATR 対話データベースの内容", 日本音響学会誌, ATR テクニカルレポート, No.T-I-186, (1990)
- [2] 板橋秀一, " 文部省「重点領域研究」による言語データベース", 日本音響学会誌, Vol.48, No.12, pp.894-898, (1992)
- [3] 小林, 板橋, 速水, 竹沢" 日本音響学会研究用連続音声データベース", 日本音響学会誌, Vol.48, No.12, pp.888-893, (1992)
- [4] 匂坂芳典, 浦谷則好, " A T R 音声, 言語データベース", 日本音響学会誌, Vol.48, No.12, pp.878-882, (1992)
- [5] " 大字林", 三省堂, (1985)
- [6] 浦谷則好, 田代, 山田, 松本, " ATR 音声対話データベース用日本語形態素解析ユーザマニュアル" ART テルニカルレポート, IT-0009, (1993)
- [7] 森元, 村上, " 音声対話における言語現象", 日本音響学会誌, Vol.50, No.7, pp.55 8-562, (1994)
- [8] 山本, 樽松, " 音声対話の特徴分析", 平成6年度音響学会春季研究発表会発表予定, (1995)