

音声対話データの分析と発話理解への応用 Analysis and Application of Dialogue Corpus to Utterance Understanding

上條 俊一, 秋葉 友良, 伊藤 克亘[†], 田中 穂積

Shunichi KAMIJO, Tomoyosi AKIBA,
Katunobu ITOU[†] and Hozumi TANAKA

東京工業大学 [†]電子技術総合研究所
Tokyo Institute of Technology [†]Electrotechnical Laboratory

Abstract

This paper describes the analysis of transcribed user's utterances in the dialogues between human and machine obtained by Wizard of Oz simulation. The recorded data amounts to about 1,300 minutes. According to our analysis, only 26 percent of user's utterances included filled pauses compared with the human-human dialogue in which about 50 percent of utterances included them, and over 80 percent of filled pauses in human-machine dialogues were independent or found at the beginning of the sentences. Repairs were observed in 5 percent of the utterances and 73 percent of them were the repetitions or substitutions. We also report the analysis of other phenomena and suggest how to apply these results to utterance understanding.

1 はじめに

対話において、人間は考えながら発話を進めるため、発話を中断して考えたり、発話を言い換えたり、相手がわかっていると思われることを省略したりする。発話を中断するときには、しばしば間投詞などが挿入される。このような現象は通常書き言葉には見られないため、これまでの言語処理のアプローチは通用しない。そこで、これらの自由発話に見られる特徴を分析する試みがなされ [6, 9, 13], それらをもとにした対応策もいくつか提案されている [9, 10, 14]. これらの分析はいずれも人間同士の対話を収録したものについて行なわれている。しかし、人間は相手のふるまいによって態度を変えるため、人工物である機械を相手にした場合は、人間を相手にした場合とは異なった現象が見られる可能性がある [4]. したがって、システムを相手にした対話を収録したものを分析することは興味深いものがある。欧米ではむしろシステムを相手にした対話の分析の方が主流である [1, 2, 3].

我々は、構築した対話システムの認識部のみを人間が肩代りするが被験者にはその旨を伝えず、あた

かも機械と対話しているように思わせるなりすまし方式 (Wizard of Oz 方式) で、人間と機械の対話を収録した [5]. 本稿ではその対話データ中の人間の発話の特徴を分析し、人間同士の対話との比較を行なう。そして、その分析結果を発話理解に応用する方法について述べる。

2 自由発話文の特徴

対話のような自由発話文には、書き言葉には見られないさまざまな特徴がある。呼び方ととらえ方は、研究者によって若干異なるが、大きくとらえると、間投詞などの冗長表現、言い淀み、言い直し、倒置、省略にまとめられる。以下にこれらについて詳細に述べるとともに、分析の際のこれらの定義を示す。

冗長表現

自由発話には、対話の流れに影響を与えない冗長な表現がしばしば登場する。間投詞、不要語、つなぎ言葉などと呼ばれるものがこれに該当するが、ここではもう少し広くとらえ、話者にとって伝達する必要がない表現ととらえる。話し言葉では、場つな

ぎとして接続詞や代名詞、連体詞、終助詞などが使われることがある [11]。また、対話文ではひとりごともしばしば現れる。したがって、間投詞(感動詞 [12])の他に、意味のない接続詞、代名詞、連体詞、終助詞、ひとりごとを冗長表現に含めることにする。ひとりごとは、多少性質が異なるので、ここでは冗長表現からひとりごとを抜いた表現を冗長語と呼んで区別することにする。

言い直し

話者自身が言い間違いに気づいた場合や、さらに良い言い回しをしたくなった場合に、発話を中断して、その部分を発話しなおすことである。言い直すとき、「じゃなくて」のように言い直すことを陽に示す語句をとともなう場合がある。このような、言い直しを明示する表現を cue phrase(合図語)と呼ぶことにする。言い直しの種類としては、

- 繰り返し (repetition)
- 言い換え (substitution)(同一構文/意味カテゴリのもので言い直したもの)
- 挿入による詳細化 (insertion)
- 言い捨て (deletion)(言い直し前の語を削除)

に分けることができる。

言い淀み

発話中に話者が躊躇して、不自然な無音区間を作ったり、語尾を伸ばしたりする現象を言い淀みと呼ぶことにする。言い直しとの重複を避けるため、言い直しにともなう言い淀みは除くことにする。

言い誤り

助詞の誤り、言い間違いなどが話者によって訂正されていないのに聞き手が状況から理解した非文法的な表現を指す。これには、

- 言い間違い(固有名詞の聞き間違いなどによる)
- 助詞の誤用(省略は除く)
- 文脈の乱れ(よじれ、重複など)

などがある。

倒置、省略

日本語では語順の制約が緩いために、倒置を明確に定義することは難しい。また、対話文では発話が文脈に依存するため、省略も明確に定義することはできない。ここでは、山本ら [14] の定義に従い、倒置を「1文内で後ろから前にかかっているもの」、省略を名詞文節の助詞落ちであるとする。

3 対話データの分析

2節の定義をもとにして、対話データの分析を行った。使用したデータの詳細と分析結果を次に示す。

3.1 使用した対話データ

1節で述べたように、我々は作成した対話システムを使って、Wizard of Oz 方式で対話の収録実験を行った。そこで収録されたデータの概要を表 1 に示す。なお音声データの場合、句読点があるわけではないので、文や発話を定義することは容易ではない。表の発話数は書き起こした者の主観によって分けたものであることに注意されたい。

表 1: 対話データ

ドメイン	渋谷の情報提供(主に道案内)
対話数	197
被験者数	40名(男23, 女17)
発話数(被験者)	5,625
収録時間	約1,300分間

我々はこれらの対話をすべてテキストに書き起こした。分析はこのうち、チェックが終了した133対話(29名, 856分)について行なった。

収録実験では被験者に5つの課題を与えたが、そのうちの4つは被験者がシステムの説明を受ける側であったのに対し、残りの1つは役割を交替し、被験者がシステムに説明するというものであった(詳細は文献 [5] 参照)。前者(以下タイプ1)は人間がシステムを利用する状態に近いのに対し、後者(以下タイプ2)は人間同士の対話に近い。したがって、以下の分析ではこれらを分けて扱うことにする。なお、

タイプ1は106対話(2,775発話)、タイプ2は27対話(1,014発話)あった。

表 2: 対話時間の分布

タイプ 1		タイプ 2	
対話時間	対話数	対話時間	対話数
1 から 3 分	28	1 から 3 分	0
3 から 5 分	38	3 から 5 分	0
5 から 7 分	20	5 から 7 分	4
7 から 9 分	4	7 から 9 分	8
9 から 20 分	14	9 から 20 分	15
20 から 41 分	2	20 から 41 分	0

3.2 分析結果

対話データの分析結果を、項目ごとに述べる。

冗長表現

冗長表現を含む発話の数は、タイプ1で726発話、タイプ2で445発話あった。したがって、単純に計算するとタイプ1で発話全体の26%、タイプ2で発話全体の44%に冗長表現が含まれることになる。村上の分析[13]では、文の49.4%に冗長語(ひとりごととは含まない)が含まれていた。これと比較すると、タイプ1の場合の冗長表現が現れる割合はかなり少ない。これは、タスクが道案内が中心であり、ユーザが説明するような発話が少ないことも影響しているが、機械を相手にしていることから、できるだけ簡潔にわかりやすく話そうという意識が働くことも少なからず影響しているものと思われる。ユーザからシステムに向かって「私の言ったことわかります(上がり調子)」と聞く発話が見られたこともそのあらわれと思われる。

分析データ中に出現した冗長表現の数を表3にまとめる。

ひとりごとを除いて考えると、平均してタイプ1では1発話あたり1.4個、タイプ2では1発話あたり1.3個の冗長語が含まれることになる。ひとりごとを含まない冗長語の出現頻度を表4に示す。なお、表記はかなりばらつきがあったが、促音と長音が異

†()内は冗長語の数

表 3: 冗長表現の数

種類	冗長表現の数	
	タイプ 1	タイプ 2
間投詞(感動詞)	905	525
接続詞	115	44
ひとりごと	94	120
その他	7	16
合計	1,121 (1,027) [†]	705 (585) [†]

なるだけのものは同じものとして扱い、表では代表的なものをのせておいた。

表 4: 冗長語出現頻度

種類	頻度		
	タイプ 1	タイプ 2	計
えー	217	161	378
えーと	172	101	273
んー	172	94	266
あ	128	41	169
じゃあ	73	9	82
んーと	29	24	53
ふーん	43	3	46
では	29	7	36
えーとねえ	12	15	27
あー	19	5	24
は	16	5	21
(その他)	117	120	237
計	1,027	585	1,612

表の上位11位までで85%の冗長語を網羅している。この結果は、人間同士の対話を分析した結果[8, 9, 10, 13]と大差はない。

次に冗長表現の出現位置を調べてみると、表5のようになった。なお、出現位置は上にあるものを除いたもので、例えば単語の区切れは文節の区切れを含んでいないものとする。

これから、タイプ1では86%、タイプ2では69%の冗長表現が単独で、あるいは文頭に現れていることがわかる。この割合も、人間同士の対話(39.41%[9])に比べると非常に多くなっている。タスクと分類の

表 5: 冗長表現の出現位置

出現位置	個数	
	タイプ 1	タイプ 2
単独	109 (9.7%)	86 (12.2%)
文頭	859 (76.6%)	403 (57.2%)
文節の区切れ	77 (6.9%)	149 (21.1%)
単語の区切れ	8 (0.7%)	10 (1.4%)
単語の途中	9 (0.8%)	5 (0.7%)
文末	59 (5.3%)	52 (7.4%)

仕方が異なるため、単純に比較はできないが、人間は機械を相手にすることで、曖昧に発話せず、文を頭で完成させてから発話する傾向があることを反映しているとも考えられる。また、単語の区切れや単語の途中に現れることは非常に少なかったことから、冗長表現は文節より大きなまとまりの間に挿入される傾向があると結論づけることができる。

言い直し

言い直しを含む発話の数は、タイプ 1 で 133 発話 (4.8%)、タイプ 2 で 84 発話 (8.3%) だった。また、言い直しの総数は、タイプ 1 で 162 回 (1.2 回/発話)、タイプ 2 で 121 回 (1.4 回/発話) であった。Bear ら [1] は、人間同士の対話の言い直し率が 34% なのに対し、彼らのコーパス (WOZ で収録) では 9 語以上からなる発話の 10% が言い直しを含んでいたと報告している。O'Shaughnessy の分析 [3] でも、発話の約 10% に言い直し現象が見られた。我々のコーパスの言い直しの頻度もこれにほぼ等しいことがわかる。

次に、言い直しの種類と頻度を表 6 にまとめる。

これから、言い直しの 38.9%/33.9% (タイプ 1/タイプ 2) が単純な繰り返しをとまない、ともに 38.9% が言い換えであり、10.5%/15.7% (タイプ 1/タイプ 2) が挿入語句を持つことがわかる。佐川ら [10] の分析にならって、言い換えも繰り返しとして認めると、挿入も繰り返し部分を持つことから、言い直しの約 88% が繰り返しをとまなうことになり、これは彼らの分析結果 (約 85%) とほとんど一致する。また、O'Shaughnessy の分析では、言い直しの 52% が繰り返しをとまない、26% が挿入語句を持ち、22% が新しく言い換えられたものであった。これから、言い直しを検出す

表 6: 言い直しの種類と頻度

種類	頻度	
	タイプ 1	タイプ 2
繰り返し (同一語句)	13	21
(言いかけた語)	50	20
言い換え (助詞のみ)	12	9
(言いかけの語句)	19	7
(語句)	10	14
(文節)	18	16
(文)	4	1
挿入による詳細化	17	19
言い捨て	19	14
合計	162	121

るには、繰り返しの部分を検出することが有効であると思われる。

次に、冗長表現をとまなう言い直しについて調べた結果を述べる。言い直しの直後に冗長表現が現れた回数は、タイプ 1 で 25 回 (言い直しの 15.4%)、タイプ 2 で 31 回 (言い直しの 25.6%) であった。現れた冗長表現は、「えー」「あー」といった比較的短いものが多かった。また、言い直すことを陽に示す cue phrase (合図語) について調べた。分析データ中の cue phrase を表 7 に示す。

表 7: cue phrase

表現	頻度
ごめんなさい	3
じゃない (/よ/や)	3
ていうか	2
いや	1
違うな	1
いいですいいです	1

合図語はタイプ 1,2 合わせて 11 例しか現れなかったことから、頻度は非常に少ないものと思われる。しかも、「いや」や「いいです」は別の意味にも解釈されうる。したがって、言い直しの検出に冗長語や合図語を用いることは難しそうである。

言い淀み

言い淀みは、タイプ 1 に 137 回、タイプ 2 に 140 回観測された。これらの出現位置を表 8 に示す。また、種類とその頻度を表 9 に示す。

表 8: 言い淀みの出現位置

出現位置	個数	
	タイプ 1	タイプ 2
文節の区切れ	8	40
単語の区切れ	110	82
単語の途中	19	18
計	137	140

表 9: 言い淀みの種類と頻度

種類	頻度
無音区間の挿入	178
語尾の長音化	82
長音 + 無音区間	17
計	277

言い淀みは、ほとんどが単語の区切れ、特に名詞の後に助詞が来るときの名詞の後が多かった。また、その種類はほとんどが不自然な無音区間の挿入であった。書き起こしデータを分析したため、無音区間の長さなどは調べられなかった。

言い誤り

対話データ中に、言い誤りは 47 個観測された。その内訳は、固有名詞の聞き間違いなどによる言い間違いが 34 個でもっとも多く、続いて助詞の誤用 (7 個)、文脈の乱れ (6 個) の順であった。合成器の音声が開きとりにくかったためか、固有名詞の言い間違いが目立った。

倒置、省略

対話データ中に、2 節で定義した倒置は 19 個、省略は 15 個存在した。

4 発話理解への応用

音声対話システムを構築するにあたり、音声認識と言語処理をつなげる試みは各所でなされているが、自由発話に対応するには両者を統合した処理を行なう必要がある [7]。認識手法としては、大きく分けて left-to-right 型の構文主導によるものと、ワードスポッティングなどの意味主導によるものに大別できる。我々は前者の認識器を用いて対話システムを試作した。しかし、分析で取り上げたような現象には対応できていない。そこで、今回の分析結果をもとにして、自由発話の特徴に対応する手段について項目別に考察する。

倒置と省略については、認識用の文法に記述することも可能であるし、文脈の情報などを使って解釈も可能と思われるので、対応は比較的容易である。

冗長語に関しては、小林ら [9] によって、頻度の高いものを単語として登録する手法の有効性が示されている。しかし、perplexity の増加をまねくことから、むやみに登録すると認識率が低下してしまう。データの分析結果より、冗長語は文頭に現れやすいなどといった出現位置に関する情報が得られたことから、これらを認識用の文法に反映させることで、認識精度を向上できるであろう。

言い直しに関しては、必ずしも単純な繰り返しとは限らないので、構文主導で扱うのは難しい。したがって、単語や文節のスポッティングとの併用など、意味主導のアプローチを行なう必要がある。

冗長表現と言い直しには、そもそも出現位置を特定するのが難しいという問題がある。それらの出現を音響的な情報をもとに特定する試みが Bear, O'Shaughnessy, Hirschberg らによってなされている。冗長語の存在や言い直しの有無がわかれば、N-best の候補から適切なものを選びやすくなる。今回の分析では、これらの音響的、韻律的な特徴は調べなかったが、それらについて考察することは有益であると思われる。

また、対話管理部における誤認識への対応も重要であろう。言い直しを意味主導で扱うと構文的制約が緩くなるので認識誤りが増えることが予想される。言い淀みや言い誤りも、辞書項目に対応するものがないため、未知語になったり、誤認識されたりする場合がある。これらに対処するには、認識誤りに対してロバストな処理を用意して、対話がいきづまる

のを防ぐ必要がある。

5 まとめ

Wizard of Oz 方式で収録した人間と機械との対話データ中の、自由発話文の特徴を分析し、人間同士の対話での分析結果と比較した。その結果、人間は機械を相手にすると、冗長表現や言い直しは減少する傾向があり、冗長表現の位置も、単独か文頭になりやすいことが分かった。今後はこれらの現象の音響的な面の分析を行うとともに、意味主導で発話を理解する手法について検討する。

謝辞

対話データの書き起こしに協力して下さった田中・徳永研究室の諸氏、並びにデータ処理に御協力いただいた同研究室の白井清昭氏に感謝致します。

参考文献

- [1] J. Bear, J. Dowding, and E. Shriberg. "INTEGRATING MULTIPLE KNOWLEDGE SOURCES FOR DETECTION AND CORRECTION OF REPAIRS IN HUMAN-COMPUTER DIALOG". In *the Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 56-63, 1992.
- [2] J. Hirschberg and C. Nakatani. "A SPEECH-FIRST MODEL FOR REPAIR IDENTIFICATION IN SPOKEN LANGUAGE SYSTEMS". In *EUROSPEECH*, pp. 1173-1176, 1993.
- [3] Douglas O'Shaughnessy. "CORRECTING COMPLEX FALSE STARTS IN SPONTANEOUS SPEECH". In *the Proceedings of ICASSP*, pp. I:349-352, 1994.
- [4] Sharon Oviatt. "PREDICTING SPOKEN DISFLUENCIES DURING HUMAN-COMPUTER INTERACTION". In *the Proceedings of International Symposium on Spoken Dialogue*, pp. 53-56, November 1993.
- [5] 伊藤克亘, 秋葉友良, 長谷川修, 田中和世. 音声対話システム構築のための実対話データ収録実験. 音声言語情報処理, No. 2-6, pp. 35-42, July 1994.
- [6] 大塚裕子, 岡田美智男. 自然な発話における「言い直し」の現象について. 日本音響学会講演論文集, No. 2-4-14, pp. 45-46, March 1993.
- [7] 河原達也, 荒木雅弘. Spontaneous speech の理解のための処理モデル. 情報処理学会研究グループ資料, No. 93-SLP-1-9, pp. 49-53, 1993.
- [8] 小林聡, 山本幹雄, 中川聖一. 問投詞, 言い直し等の出現に関する音響的特徴. 情報処理学会研究グループ資料, No. 93-SLP-1-2, pp. 7-10, July 1993.
- [9] 小林聡, 甲斐充彦, 山本幹雄, 中川聖一. 問投詞の出現位置の特徴分析と音声認識システムの評価. 情報処理学会研究グループ資料, No. 92-SLP-3-4, February 1993.
- [10] 佐川雄二, 大西昇, 杉山昇. 自己修復を含む日本語不適格文の分析とその計算機による理解手法に関する考察. 情報処理学会論文誌, Vol. 35, No. 1, pp. 46-52, January 1994.
- [11] 八王子市議会事務局議事課(編). 議会会議録作成のための詞文の整理基準について(副題「話しコトバと書きコトバのかけ橋」). 八王子市議会事務局議事課, 1993.
- [12] 益岡隆志, 田窪行則. 基礎日本語文法-改訂版-. くろしお出版, 1992.
- [13] 村上仁一. ATR 対話データベースの内容と分析. 日本語学, Vol. 11, No. 10, pp. 73-81, 1992.
- [14] 山本幹雄, 小林聡, 中川聖一. 音声対話文における助詞落ち・倒置の分析と解析手法. 情報処理学会論文誌, Vol. 33, No. 11, pp. 1322-1330, November 1992.