

# 音声翻訳システムを介した対話における修正の種類

金城 由美子 \*†

竹澤 寿幸 †

菊井 玄一郎 †

長崎純心大学\*

ATR 音声言語コミュニケーション研究所†

## 1 はじめに

対話の参加者は、発話・聞き取り・理解のそれぞれの局面において問題が生じた場合、確認や繰り返し、言い直しなどの修正を行うことなどによって問題に対処し、協調的に対話をすすめて行く。修正は、対話の目的である意思の伝達、情報の共有や相互理解を達成するため、対話参加者が用いる重要なストラテジーの一つである。

音声翻訳システムを介した対話では、単言語・母語話者間で行われる通常の対話で起こる言い間違い、聞き取りの失敗、解釈の食い違いなどの問題に加え、認識誤り、翻訳誤りに起因するミスコミュニケーションが生じる可能性がある。音声翻訳システムを利用する対話参加者は、円滑に対話をすすめ、目標を達成するために、これらの障害も克服していく必要がある。本稿では、音声翻訳システムを介した日本語話者と英語話者の課題遂行対話を対象とし、修正の種類を明らかにする。

## 2 対話実験の概要

本稿で扱う対話データは、日本語話者と英語話者が音声翻訳システムを介して課題遂行対話を行うことにより得られたデータ (Machin-Aided Dialogues, 以下 MAD) である。一連の対話実験は、音声翻訳システムの設計および評価に利用するための音声言語コーパス収集の一環として行われ、音声翻訳システムを介した対話において、実際に利用される言語表現を収集し、音声認識・翻訳システムの評価やシステムユーザである話者のコミュニケーション行動を検証するために行われている [11, 5]。

本実験では、日本語話者、英語話者のペアが、店員と客などのそれぞれの役割と土産品の購入などの比較的簡単な課題を与えられ、音声翻訳システムを利用して、課題達成のための対話を行なう。それぞれの原発話 (J,E) は、音声認識が行なわれ、さらにその認識結果 (R) を入力として、翻訳が行なわれる。

(1) J: いらっしゃいませ、何をお求めですか。

R: いらっしゃいませ何を御求めですか

T: Good afternoon what would you like

E: hello. | do you have Diet Coca-Cola?

R: hello do you have diet\_coke won't

T: いらっしゃいませ。ダイエットコークがありますか。

自発話の認識結果は話者それぞれに与えられた PDA に文字表示され、相手発話の翻訳結果はヘッドフォンに出力される合成音声と PDA 表示との両方が与えられる。

音声認識には、日英ともに ATRASR[3] を使用した。日英、英日翻訳には、句構造等のパターンに基づく機械翻訳 HPAT[2] と文を単位とする D3[9] に最良翻訳選択器を組み合わせたシステムを使用した。日本語音声合成には XIMERA[4]、英語音声合成には AT&T Natural Voices™ を使用した。

話者	日本語話者、英語話者各 6 名
入出力機器	ヘッドホン付き接話マイクと PDA
課題設定	20 パターン (観光地での道案内等)
対話数	139 対話
発話数	2,507 発話 (日本語 1,306 発話, 英語 1,201 発話)

表 1 対話実験の概要

あらかじめ、音声翻訳システムの使い方については簡単な説明を与え、発話練習を行った。話者には次のようなインストラクションが与えられた。

- 大きめの声で明瞭に話す。
- 1 回の発話は 10 秒以内とする。
- 時々誤りが発生するが、確認や再発話をするにより対話を続ける。
- 時々処理に時間がかかることがあるが、その場合は少し待つ。

個々の課題における役割とタスクについては、課題ごとにビデオによるインストラクションを行った。対話

は、課題が達成した時点で終了するが、1 課題につき 10 分を制限時間とし、達成しなかった場合はそこで対話を打ち切った。音声認識の誤り、翻訳の誤りは生じるものの、制限時間内にほとんどの課題は達成された。

### 3 修正の種類

修正 (repair) とは、間違いの訂正 (correction) だけではなく、発話内容の変更による言い直しや、発話内容の確認をする質問に対する相手話者への回答も含む。[8] においては、修正は i) 誰が行なうか (自己または他者) ii) 修正がどこで開始されるか iii) 修正がどこで行なわれるか、の 3 つの観点から分析が行われている。本稿では、自己修正を分析の対象とし<sup>\*1</sup>、修正対象に対し、どのような修正が行なわれるか、日本語、英語それぞれの修正の特徴を探る。また、通常の対話における修正と MAD におけるそれでは異なる特徴があることも示す。

#### 3.1 MAD における修正の概要

139 対話、2,507 発話中、修正が行なわれたのは日本語 139 発話、英語 134 発話の計 273 発話で、修正対象に対し、修正が a) 同一ターン、b) ターン移行スペース<sup>\*2</sup>、c) (修正対象を第一ターンとして) 第三ターン、のどこで行われるかによって、3 つに分類できる。

	日	英	計
同一ターン	27	7	34 (12.5%)
ターン移行スペース	62	39	101(37.0%)
第三ターン	87	51	138 (50.5%)

表 2 修正位置による分類

修正は速やかに行われるというのが原則であり、通常の対話においては修正対象となる要素の直後で、修正が行われる。つまり同一話者により修正対象と同一ターン内で行われるのが最も一般的な修正である。しかし、MAD では話者はそれぞれ音声認識システム、翻訳システムを介することにより、自分の発話内容について知っているが、相手話者に対し、どのような翻訳結果が示されているかはわからない。

(2) に MAD における典型的な修正の例を示す。発話 “may I help you?” が誤認識され、修正対象 (trouble source) となる。修正は、相手話者の「もう一度お願いします」や “could you repeat that again?” のような明

示的に再発話を求める発話により開始され (other repair initiation)、修正は修正開始発話の直後、修正対象を最初のターンとして数えた場合の 3 番目のターンにおいて、行なわれることが多い。

#### (2) 典型的な修正の例

T1 E: hello. may I help you?  
 <E> R: hello men help you  
 T: もしもし。男性はあなたに召し上がります。

T2 J: もう一回言って頂けますか。  
 R: もう一回言って頂けますか  
 T: Could you say one ?

T3 E: may I help you?  
 R: may\_i\_help\_you  
 T: いらっしゃいませ

#### 3.2 同一ターンでの修正

MAD において、3 つの修正のうち、同一ターンでの修正は 34 例 (日 27, 英 7) と最も少ない。同一ターンでの修正が起らないということは、言い誤りや、発話途中の内容変更などがないことを意味する。これは、話者が翻訳システムを利用するという状況から比較的整った発話をする傾向があること [7]、また、1 回の発話は 10 秒以内という制約があるため、発話内容をあらかじめ決定してから話し始めることによりもたらされるものと思われる。

通常の対話における同一ターンでの修正については、日本語と英語では基本語順や語構成の違いから、異なる特徴を示すことが自然対話を分析の対象として [1] で指摘されている。ここでは、[1] で述べられた同一ターンでの修正と MAD における同一ターンでの比較を行う。

##### 3.2.1 名詞途中、または名詞直後の修正

名詞の途中または名詞直後の修正の大きな違いの一つとして、目的語が修正される場合、英語では動詞句や主語まで遡って修正発話で再利用されることがあるのに対し、日本語では上位の構成素を含むような修正はない。MAD における発話においても、同様の例が確認された。{} は修正対象を示し、修正発話を下線部で示す。

(3) [uh] I'm sorry. {so you made the con} I'm sorry.  
you made the reservation for me?

(4) 同じ部屋の友達が、夜中に、{ お腹を }お腹が痛いと言っています。

また名詞の修正が日本語ではターン開始部が半分以上という指摘があったが、この点は 14 例中 2 例とあては

<sup>\*1</sup> 他者修正の例は見られない。また一般の対話においても他者修正はまれであるとされている。

<sup>\*2</sup> 先行ターンの完了点または次話者のターン開始点となる可能性のあるポイント。

まらなかった。

動詞複合要素の途中、または直後の修正

英語は動詞のみの修正は少なく、ターン開始部では起こらないという指摘があるが、MADにおいても2例の動詞修正はいずれも主語から繰り返されている。

(5) yes.{I'd like to stay.} yes. I'd like to stay tonight and tomorrow night

日本語の動詞要素の修正は、ターン開始部での動詞のみの修正が多い、との分析だったが、ターン開始部での動詞の修正は14例中3例しか確認できなかった。

(6) {かしこいま} かしこまりました。

また英語にない日本語の動詞要素の修正の特徴として、動詞複合要素内の形態素を修正するという例があげられている\*3。あてはまる例が確認されたが、1例しかなく動詞複合要素内の形態素「い」が「りません」に修正されたかどうかは判断が難しい。

(7) 電車で乗り換えなければな{い}りません。

同一ターンの修正は頻度が極端に少ないだけでなく、MADにおける日本語の修正は自然対話のそれと異なる点が目立つ。この問題については、次節で考察を行う。

### 3.3 ターン移行スペースでの修正

ターン移行スペースでの修正、つまり修正対象発話を含むターンの直後のターンでの修正は、101例(日62、英39)が確認された。人による書き起こしテキスト、音声認識および翻訳システムの記録から、修正が音声認識の誤りに起因すると判断できる例は、75例(日48、英27)、音声認識は正しいが翻訳に問題がある例は5例(日3例、英2例)であった。残る21例については、認識、翻訳ともに致命的な間違いはなかった。実験中、認識結果のPDA出力、翻訳結果の合成音声出力のトラブルが起き、これらが原因となっている可能性はある。

音声認識誤りがターン移行スペースでの修正の多数を占めるのは、自発話の翻訳の正誤がわからないのとは異なり、音声認識誤りが起きた場合、話者はPDA表示により、発話直後にそれを知ることができ、すぐに修正を行うことができるからであろう。

ターン移行スペースでの修正の特徴は、修正対象と修正発話の発話内容がほとんど同じものがほぼ8割を占め

ることである\*4。残りの約2割は修正により情報が付加されたものと削減されたものに二分される。

### 3.4 第三ターンでの修正

第三ターンでの修正は、日本語87例、英語51例計138例と最も多い。第三ターンでの修正の多くは(2)で示したように、相手話者が開始するものが日本語37例、英語67例計104例と大多数を占める。

修正を開始する表現の多くは、再発話を要求する定型的な表現であり、7割弱を占める。次のような表現が頻繁に使われるが、「わかりません」のような表現も同等の役割を果たす。

(8) 再発話を要求する定型的な表現

日本語 もう一度お願いします、もう一度言っていた  
だけますか、

英語 could you repeat that, please?, pardon me?,  
excuse me?

その他に、修正の開始となる表現は相手話者による問い返しであり、WH疑問文による一部情報の要求と、Yes/No疑問文による先行発話の全体または一部の情報の確認がある。WH疑問文による一部情報の要求の方が頻度が多いが、大きな違いとは言えない。

(9) WH疑問文による情報の要求

T1 E: I need to get money.

R: i need

T: 欲しいのですか。

T2 J: 何が欲しいのですか。

R:何が欲しいのですか

T: what would you like from me

T3 E: I need to get money from my credit card.

R: i knees to you get money from my credit\_card

T: 膝中私のクレジットカードの金を手に入れません。

(10) Yes/No疑問文による情報の確認

T1 J: この先を右に曲がると、すぐにその銀行はあります。

R: この先を右に曲がるとすぐにその銀行はあります

T: I have around the corner on the right

T2 E: turn right and go straight?

R: turn right and go straight

T: 右に曲がってまっすぐ行って

T3 J: 右に曲がった所にあります。

R: 右に曲がった所にあります

T: I have around the corner on the right

\*3 「じゃ何時ごろにくだ{し}そう?」。

\*4 プロミネンスなど発話の仕方は異なることが多い。

## 4 修正のストラテジーと制約

初期の対話実験では、タイピストの入力を音声認識システムの代用としていたが、修正の内訳については、第三ターンでの修正が多く、次にターン移行スペース、同一ターンでの修正が少ないという本実験と非常によく似た傾向を示している [6]。つまり、MAD における修正の代表的なタイプは下記のような連鎖を持つ第三ターンでの修正ということになる。

### (11) MAD における修正の連鎖

Turn 1: 修正対象発話  
Turn 2: 修正開始  
Turn 3: 修正発話

第三ターンでの修正は、できるだけ速やかに自己修正を行うという修正の原則に最も遠いものに見えるが、自発話の翻訳結果を知ることができない以上、相手発話の情報を頼りに修正を行うという方法を取るしかない。

また、音声認識の誤りに関しては、認識結果の表示を頼りに、自発話直後のターン移行スペースでの修正が速やかな対応、ということになる。

最後に、同一ターンでの修正が少ない、ということは、同一ターンでの修正が MAD における修正のストラテジーとして役に立たないということではなく、なぜ同一ターンでの修正を行なわなくて済むのか、またなぜ同一ターンで行われる修正の形式が通常の対話における修正と異なる特徴を示すのか、という観点から、話者に対する制約とそれに対するストラテジーの関係を捉えていく必要があると思われる。

本実験では、発話音声、PDA 表示、合成音出力などの情報チャンネルの制限を行っていないが、[10] の実験では相手話者の音声、翻訳結果の PDA 表示の制限を組み合わせての対話実験を行っている。実験の結果、結果として誤訳の頻度は情報チャンネルが制限された場合でも、誤訳の生じる発話の頻度にほとんど変わりがないことがわかった。ただし、情報チャンネルの制限を受けている場合、制限を受けない場合と比べて、発話単語数が日本語、英語ともに少なくなることがわかっている。このような制約とストラテジーの関係が、同一ターンでの修正の特徴にも見出せるかどうか今後の課題である。

謝辞

本研究は通信・放送機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したも

のである。

## 参考文献

- [1] M. Hayashi. *A Comparative Study of Self-Repair in English and Japanese Conversation*, Vol. 4, Stanford, 1994. CSLI Publications.
- [2] K. Imamura. Application of translation knowledge acquired by hierarchical phrasealignment for pattern-based mt. In *TMI2002*, pp. 74–84, 2002.
- [3] 伊藤, 葦莉, 實廣, 中村. 音声認識統合環境 ATRASR の概要と評価報告. 日本音響学会 2004 年秋季研究発表会講演論文集 I, pp. 221–222, 2004.
- [4] H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda. Ximera: A new tts from atr based on corpus-based technologies. In *ISCA 5th Speech Synthesis Workshop*, pp. 179–184, 2004.
- [5] 菊井, 竹澤, 鈴木, 西野. 自動翻訳システムを用いた日英対話データの収集. 言語処理学会第 9 回年次大会発表論文集, pp. 529–532. 言語処理学会, 2003.
- [6] 金城, 竹沢, 菊井. 翻訳システムを介した対話における修正のストラテジー. 人工知能学会研究資料 (SIG-SLUD-A303-07), pp. 39–44. 人工知能学会, 2004.
- [7] 水島, 竹澤, 菊井. 翻訳システムを介した対話音声の発話スタイルについて—自然発話、朗読発話との関係—. 第 3 回話し言葉の科学と工学ワークショップ講演予稿集, pp. 135–142, 2004.
- [8] E. A. Schegloff, G. Jefferson, and H. Sacks. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382, 1977.
- [9] E. Sumita. Example-based machine translation using dp-matching between word and sequences. *Proc. ACL-2001 Workshop on Data-Driven Methods in Machine*, pp. 1–8, 2001.
- [10] 竹澤, 水島, 菊井. 翻訳チャンネルを介した対面コミュニケーションにおける各種情報チャンネルの影響について. *インタラクシオン* 2005.
- [11] 竹澤, 菊井, 鈴木, 西野. コーパスベース音声翻訳研究のための対話データ収集. 音声言語情報処理, 45(12):71–76, 2003.