

話者の役割を利用した音声翻訳

山田節夫 隅田英一郎 柏岡秀紀

E-mail: {syamada, sumita, kashioka}@itl.atr.co.jp

ATR 音声翻訳通信研究所

1 はじめに

最近様々な音声翻訳システムが提案されている [1, 6, 7, 8]。これらの音声翻訳システムを使って会話を自然に進めるためには、原言語を解析して得られる言語情報の他に言語外情報も使う必要がある。例えば、話者に関する情報 (社会的役割 (受付、客等) や性別等) は、原言語を解析するだけでは得にくい情報であるが、これらの情報を使うことによって、より自然な会話が可能となる。

言語外情報を利用する翻訳手法は幾つか提案されている。例えば、文献 [3] では、「spoken language pragmatic information」を使った翻訳手法を、また、文献 [5] では、「situational information」を使った手法を提案している。両文献とも言語外情報を利用した手法であり、文献 [5] では机上評価もしているが、実際の翻訳システムには適用していない。言語外情報である「pragmatic adaptation」を実際に人と機械とのインターフェースへの利用に試みている文献 [4] もあるが、これも音声翻訳システムには適用していない。これら提案の言語外情報を実際の音声翻訳システム上で利用するには困難な課題が数多くあり、解決するのは時間がかかると考えられる。

そこで、本稿では、下記の理由により話者の役割に着目し、実際の音声翻訳システムに容易に適用可能な話者の情報の利用について提案する。

- 音声翻訳システムの主たる活用の一つに、窓口での対話がある。そこでの「受付」と「客」のような会話では、「受付」がより丁寧に喋ったほうが会話が違和感なく進む。
- 音声翻訳システムでは、マイクなどのインターフェースによって発話者の役割が容易にシステムに取り込める。

本稿では、この役割に着目し、変換ルールと対訳辞書に、話者の役割に応じたルールや辞書エントリーを追加することによって、翻訳結果の「丁寧さ」を制御することを提案する。

旅行会話の未訓練 (ルール作成時に参照していない) 23 会話 (344 発声) で実験し、丁寧表現にすべき

かどうかという観点で評価した。その結果、本手法は再現率が 65%、適合率が 86% という精度であった。さらに、話者の情報に応じたルールや辞書エントリーをより充実させると、再現率が 86%、適合率が 96% になることを机上で確認した。したがって、本手法は実現性が高く、音声翻訳システムを使ってより自然な会話をするのに効果的であると言える。

2 「話者の役割」と「丁寧さ」

翻訳システムを使った会話をより自然にするために、訳文の「丁寧さ」が重要である。我々は、「話者の役割」と「丁寧さ」について着目し、その関係について調査した。「丁寧さ」を特に考慮せずに一般的な訳を出す翻訳システムが出力した結果に対して、話者の役割が「受付」である場合に、訳をより丁寧な表現にすることが望まれるかという観点で調査を行った。実際の調査に用いた翻訳システムは、変換主導型翻訳システム [7] (Transfer-Driven Machine Translation、略して TDMT) の英日版で、対象文は話者の役割が「受付」である 1409 発声とした。その結果、約 70% (952) の発声を丁寧にしたほうが良いと判断された。したがって、翻訳システムを使ったより自然な会話を実現するために、話者の役割に応じて「丁寧さ」を変えるのは、有用であると言える。

3 話者の情報を翻訳知識に組み込む方法

本節では、役割や性別のような話者の情報をどのように既存の変換ルールや対訳辞書で実現するかについて述べる。最初に、TDMT について簡単に説明し、次に、話者の情報を付与したルールや辞書エントリーについて述べる。

3.1 変換主導型翻訳 (TDMT)

TDMT は図 1 に示す変換ルールを使って、左から右へのボトムアップ型チャート解析法で構文解析を行っている [2]。様々な変換ルールによって、入力文の構文構造が複数の候補となった時は、意味距離計算によってその候補は絞られる。

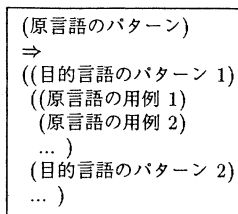


図 1: 変換ルールのフォーマット

変換ルールは、図 1 のように、原言語のパターン、目的言語のパターン、原言語の用例からなる。原言語のパターンは変項と構成素境界からなる [2]。ここでの変項は、X や Y などの大文字の記号で表し構成素に対応する。構成素境界は、機能語あるいは左の構成素の品詞と右の構成素の品詞を付けた記号（品詞バイグラムマーカと呼ぶ）である。品詞バイグラムマーカは、構文解析をする前に対象の構成素間に挿入される。例えば、“accept” が Verb で “payment” が Common Noun なので、例 1 では、品詞バイグラムマーカ $\langle V-CN \rangle$ が、“accept” と “payment” の間に挿入される。目的言語のパターンは、原言語のパターンに対応した変項と訳語から構成される。ここでの変項は、x や y などの小文字の記号で表し、例えば、X と x のように、原言語のパターンでの変項と対応している。原言語の用例は、そのパターンにマッチした訓練文（パターン作成時に参照した文）で実際に出現した語である。

例 1

英語：We accept payment by credit card
 標準：私達はクレジットカードでの支払いを受け付けます
 丁寧：私達はクレジットカードでのお支払いをお受けします

図 2 では、原言語のパターンは $(X \langle V-CN \rangle Y)$ 、変項は X、Y、目的言語のパターンは $(y \text{ “を” } x)$ や $(y \text{ “に” } x)$ 、目的言語の用例は、 $((\text{“receive”}) (\text{“payment”}))$ や $((\text{“take”}) (\text{“picture”}))$ でそれぞれ表されている。図 2 の変換ルールは、原言語が $(X \langle V-CN \rangle Y)$ にマッチしたら、X と Y に対応した入力語が最も意味的に近い用例の組が付いている目的言語のパターンを選択せよという意味である。例えば、“accept $\langle V-CN \rangle$ payment” がマッチしたとすると、最も近い用例の組は $((\text{“receive”}) (\text{“payment”}))$ なので、目的言語のパターンは $(y \text{ “を” } x)$ が選択される。このように、入力された語によって適切な目的言語のパターンが変換ルールによって選ばれる。

目的言語のパターンが選ばれた後、図 3 の対訳辞書を用いて、そのパターンに応じた訳を決定する。例え

ば、“accept $\langle V-CN \rangle$ payment” の場合は、“支払いを受け付ける” という訳になるが、“支払い” と “受け付ける” は図 4 の辞書引きの結果から、また、“を” は目的言語のパターン $(y \text{ “を” } x)$ から来ている。

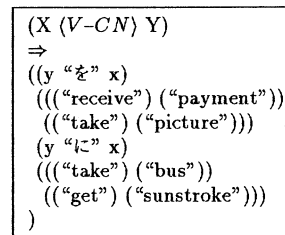


図 2: 変換ルール例

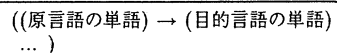


図 3: 対訳辞書のフォーマット

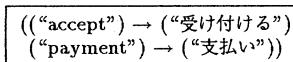


図 4: 対訳辞書例

3.2 話者の情報付きルール及び辞書エントリー

話者の情報付きルールや辞書エントリーを記述できるように、既存の変換ルールや対訳辞書のフォーマットを図 5 と図 6 のように変えた。図 5 では、目的言語のパターンはパターンの条件によって、また、目的言語の単語は単語の条件によって変えることができる。それぞれの条件に話者の情報からの条件を記述する。例えば、図 7 に示すように、“：単語の条件 1_1 ” を “:s-role clerk” と定義することができる。これは、話し手の役割が「受付」という条件である。したがって、もし話し手が「受付」とわかれば、“accept” は “お受けする” と訳される。ここで注意することは、話し手が「受付」であっても、常に “accept” の訳は “お受けする” ではないことである。つまり、図 5 の “目的言語の単語 1_1 ” は、そこで定義されている変換ルール内のみで有効なのである。もしも話者の情報に応じた訳が常に同じであれば、図 6 の対訳辞書に登録する必要がある。図 5 と図 6 の “：デフォルト” は、どのような条件にも合わなかった時にマッチする条件である。条件は、例えば最初に “：パターンの条件 1_1 ”、次に “：パターンの条件 1_2 ” というように、上から順に調べられ、最後にデフォルトのものが選択される。

```

(原言語のパターン)
⇒
(((目的言語のパターン 11) : パターンの条件 11
 (目的言語のパターン 12) : パターンの条件 12
 ...
 (目的言語のパターン 1n) : デフォルト)
((原言語の用例 1)
 ... )
(((原言語の用例 1) → (目的言語の単語 11) : 単語の条件 11
 (原言語の用例 1) → (目的言語の単語 12) : 単語の条件 12
 ...
 (原言語の用例 1) → (目的言語の単語 1m) : デフォルト)
 ... )
(((目的言語のパターン 21) : パターンの条件 21
 ... )))

```

図 5: 話者の情報付き変換ルールのフォーマット

```

(((原言語の単語 1) → (目的言語の単語 11) : 条件 11
 (原言語の単語 1) → (目的言語の単語 12) : 条件 12
 ...
 (原言語の単語 1) → (目的言語の単語 1k) : デフォルト)
 ... )

```

図 6: 話者の情報付き対訳辞書のフォーマット

図 7 のルール、図 8 の辞書エントリーを用いると、例 1 の文は、話者が「受付」の場合、最終的には「私供はクレジットカードでのお支払いをお受けします」と訳出される。

4 実験

実験に用いた TDMT（英日版）は、話者の役割を考慮する前では、約 1,500 の変換ルール、約 8,000 の対訳辞書エントリーが登録されていた。「受付」用に「丁寧さ」を向上させるために、約 300 ルール、約 40

```

(X (V-CN) Y)
⇒
((y “を” x)
 (((“accept”) (“payment”))
 ((“take”) (“picture”))))
 (((“accept”) → (“お受けする”) :s-role clerk
 (“accept”) → (“受け付ける”)))
)

```

図 7: 役割情報付き変換ルールの例

```

(((“payment”) → (“お支払い”) :s-role clerk
 (“payment”) → (“支払い”))
 ((“we”) → (“私供”) :s-role clerk
 (“we”) → (“私達”)))

```

図 8: 役割情報付き対訳辞書の例

の辞書エントリーを修正した。

「受付」用に修正した変換ルールと対訳辞書を用いて未訓練である 23 会話（344 発声）を対象に実験を行った。音声翻訳の入力は本来音声であるが、今回の実験では、翻訳のみの評価を行うため書き起こしテキストを形態素解析したものからの入力とし、話者の情報も同時に与えた。

実験の評価をするために、実験に使った 23 会話（このうち、話者が「受付」は 199 発声、「客」は 145 発声）に対して、話者の情報を適用する前の TDMT の翻訳結果を 2 つに分類した。一方は丁寧にすべき表現を含む場合、もう一方は丁寧にすべき表現を含まない場合である。表 1 にその調査結果を示す。

表 1: 表現を変える必要性の調査結果

表現を変える必要性	発声数
あり	104
なし	166
対象外 *	74
合計	344

* 74 の発声に関しては、「丁寧さ」を判断するには不適切な翻訳結果であったので、本稿では対象外とした。

同じ対象文に対して、話者の情報を適用する前と後の TDMT の結果を「丁寧さ」という観点で印象がどう変わったかで評価した。表 1 の表現を変える必要性の有無別にまとめたものを表 2 に示す。

表 2: 話者の役割を使った翻訳の評価結果

表現を変える必要性	結果	発声数
あり (104)	良い	68
	同等	5
	悪い	3
なし (166)	差分なし	28
	良い	0
	同等	3
	悪い	0
	差分なし	163

良い：翻訳結果の印象が良い

同等：翻訳結果の印象が同等

悪い：翻訳結果の印象が悪い

差分なし：翻訳結果に差分がない

以下に示す再現率と適合率で本手法を評価した。

$$\text{再現率} = \frac{\text{翻訳結果が良くなった発声数}}{\text{丁寧表現にするべき発声数}}$$

$$\text{適合率} = \frac{\text{翻訳結果が良くなった発声数}}{\text{翻訳結果に差分のあった発声数}}$$

その結果、再現率は65% (= $68 \div (68 + 5 + 3 + 28)$)、適合率は86%(= $68 \div (68 + 5 + 3 + 0 + 3 + 0)$)であった。翻訳結果の印象が悪くなったものは、実験に使った翻訳システムの日本語生成処理に起因するものであった。翻訳結果の印象が同等のもの、表現を変える必要があるのに変らなかったものの原因は大きく2種類に分けることができた。1つは、実験に用いた翻訳システムでは、その発声の動作主が誰(何)なのかわからないことに原因がある。話者の役割が「受付」であっても、丁寧にする対象の語の動作主が「受付」の場合はその語は丁寧な表現とはならないが、「客」の場合は丁寧な表現となり得るからである。もう1つは、「受付」用のルールや辞書エントリーがまだ不十分なことに原因がある。

後者の原因は、比較的容易に解決できる問題であり、もしも十分にルールや辞書エントリーが登録できたとすると、再現率と適合率を向上させることができる。これを調査したところ、再現率が86%、適合率が96%まで向上できることが分かった。したがって、本手法は実現性が高く、音声翻訳システムを使って自然な会話をするのに効果的であると言える。

5 まとめ

原言語を解析するだけでは得にくい話者の役割を使った翻訳手法について述べた。この役割は音声翻訳システムの中では容易に得ることができ、本手法では、この話者の役割を利用して訳し分けができる。今回利用した役割は「受付」と「客」であるが、話し手が「受付」の時はより丁寧な表現にすることで実現している。

既存の変換ルールや対訳辞書に「受付」用のルールと辞書エントリーを追加し、旅行会話を対象とした英日翻訳システムに適用した。未訓練である23会話(344発声)を対象に実験し評価したところ、再現率が65%、適合率が86%の結果となった。「受付」用のルールや辞書エントリーをより充実させると、再現率が86%、適合率が96%に向上することが分かっている。したがって、本手法は実現性が高く、音声翻訳システムを使ってより自然な会話をするのに効果的であると言える。

話者の役割だけでなく、訳し分け対象の語の動作主が分からないと訳し分け困難な場合がある。そこで、動作主を決める方法を検討する予定である。また、より精度良く翻訳結果を出力させるために、「受付」用のルールや辞書エントリーをより一層充実させていく予定である。さらに、容易に得られそうな言語外情報(例えば、性別等)を音声翻訳システムで利用する手法についても検討していく予定である。

参考文献

- [1] Thomas Bub, et al. Verbmobil: The combination of deep and shallow processing for spontaneous speech translation. In *the 1997 International Conference on Acoustics, Speech, and Signal Processing: ICASSP 97*, pp. 71-74, Munich, 1997.
- [2] 古瀬蔵, 山本和英, 山田節夫. 構成素境界解析を用いた多言語話し言葉翻訳. 自然言語処理, Vol. 6, No. 5, pp. 63-91, 1999.
- [3] Keiko Horiguchi. Towards translating spoken language pragmatics in an analogical framework. In *Proceedings of ACL/EACL-97 workshop on Spoken Language Translation*, pp. 16-23, Madrid, 1997.
- [4] Susann LuperFoy, et al. An architecture for dialogue management, context tracking, and pragmatic adaptation in spoken dialogue system. In *Proceedings of COLING-ACL'98*, pp. 794-801, Montreal, 1998.
- [5] Hideki Mima, et al. A situation-based approach to spoken dialogue translation between different social roles. In *Proceedings of TMI-97*, pp. 176-183, Santa Fe, 1997.
- [6] Carolyn Penstein Rosé and Lori S. Levin. An interactive domain independent approach to robust dialogue interpretation. In *Proceedings of COLING-ACL'98*, pp. 1129-1135, Montreal, 1998.
- [7] Eiichiro Sumita, et al. Solutions to problems inherent in spoken-language translation: The ATR-MATRIX approach. In *the Machine Translation Summit VII*, pp. 229-235, Singapore, 1999.
- [8] Jae-Woo Yang and Jun Park. An experiment on Korean-to-English and Korean-to-Japanese spoken language translation. In *the 1997 International Conference on Acoustics, Speech, and Signal Processing: ICASSP 97*, pp. 87-90, Munich, 1997.