

# 質疑応答文に対する協調融合翻訳の適応検討

東海林 里仁、樽松 明 (電気通信大学),  
美馬 秀樹<sup>1</sup>、飯田 仁 (ATR 音声翻訳通信研究所)

## 1. はじめに

教育の国際化において自動翻訳システムを用いて質疑応答などが行えるインテリジェントレクチャーのような翻訳支援システムが求められている。

質疑応答のような会話文では主語が抜けたりすることが多いため、ルールと用例を協調させて翻訳を行う協調融合翻訳手法が有効である。[1][2][3] 協調融合翻訳方式は ATR 音声翻訳通信研究所で、旅行会話向けのプロトタイプシステムが存在するが、用例やパターン等の翻訳知識が旅行会話向けに特化されており、インテリジェントレクチャーなどの他分野に応用できるかは確認されていない。

本研究では前述のプロトタイプシステムを用いて質疑応答文のための翻訳知識を構築し、ある決まった範囲内での質疑応答翻訳システムの有効性を確かめることを目的とする。本論文では、実験を行い、変換辞書、意味辞書、変換知識の改良方法を示す。

## 2. 協調融合翻訳

協調融合翻訳手法は、話し言葉の特徴である自然な話し言葉によく見られる豊富な言い回しや、文法的に不完全な表現に対応するための手法として、句や文を構成する表現の依存関係を解析しつつ対訳用例を適応して好ましい用例を選択するメカニズムに基づいている。翻訳結果を効率的に構成するために、文法と用例を用いる翻訳手法のそれぞれの長所を取り入れた手法である。文を「X の Y」などのパターンに分解し、それぞれを意味

的距離の近さから対訳用例を決定し、文法規則を付加して、翻訳結果を出力する。具体的には、

- 1、形態素処理
- 2、原言語構造の解析
- 3、文の最尤構造の決定
- 4、対訳パターンの決定
- 5、単語レベルでの翻訳
- 6、生成処理

の順で処理が行われ、出力結果が導かれる。

## 3. 実験

### 3.1 使用データ

代表的な質疑応答文として 25 のサンプル文を用いた。例を表 1 に示す。

表 1: サンプル文の例

計算量はいくらになるか
計算時間はどのくらいになるか
誤りは何個あったか
どちらが正しいか
どうすればとけますか

期待される翻訳結果を用意し、どのような知識等を追加することで、正しい翻訳結果を導くことができるか実験を行った。旅行会話の辞書および翻訳知識をベースとして、逐次改良を行った。

### 3.2 変換辞書と意味辞書の改良

まず初期段階として、変換辞書、意味辞書の改良を行った。その結果、図 1 に示すように 25 文中 6 文を正しく翻訳できるようになった。

<sup>1</sup> 現 Dept. of Computing, Manchester Metropolitan University, Manchester M1 5GD, U.K.

### 3.3 既存パターンへの用例追加

新たな用例のパターンを追加することなく、旅行会話での翻訳で用いられていた既存のパターンへの用例追加のみを行った。従来の変換知識を利用するため、今後の弊害が最も発生しにくい方法なので、できるかぎりこの方法で解決したい。その結果 25 文中 12 文を正しく翻訳できた。

### 3.4 新規パターン構築

従来の翻訳パターンでは、表現できない言い回しなどが存在し、新たなパターンを追加する必要がある。従来の変換知識に悪影響を与えないように極力パターンの形に沿うように新たなパターンを構築する必要がある。この作業により、25 文中 17 文を正しく翻訳することができた。

### 3.5 LOCAL 辞書の追加

訳し分けや翻訳する際に日本語の中には表れない表現を付加するときなどは、その言葉特有に変換する辞書をそのパターン固有に作成する必要がある場合がある。前述の変換知識の追加と比較してその単語固有の対応をするため、やや closed であるが、訳し分けなどを行うために必要不可欠な知識である。この知識を追加することにより、25 文中 22 文を正しく翻訳することができた。

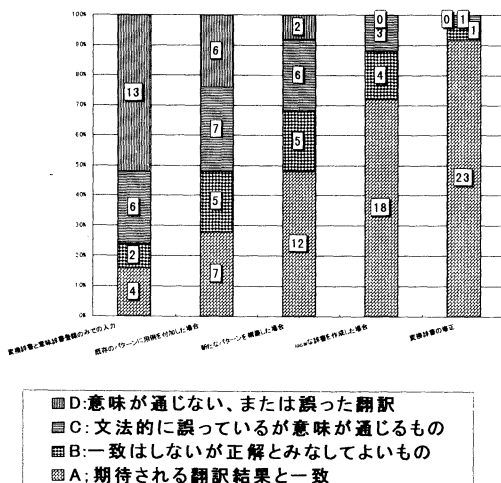


図1: 翻訳精度上昇の推移

### 3.5 変換辞書の変更

インテリジェントレクチャーのような翻訳を行う際に類似する英語表現であっても、変更したほうが適当である場合がある。この場合、変換辞書そのものを変更した。その結果 25 文中 24 文まで正しく翻訳できるに至った。

## 4. 考察

### 4.1 訳し分けの問題

一つの日本語に対応する複数の英語表現が存在し、その会話が発せられた状況や文脈の違い等により、訳語が異なることがある。今回は、旅行会話向けの翻訳システムの影響を受け、質疑応答翻訳に不適切な翻訳を行う場合や、前後関係から訳し分けが必要になる場合などが見受けられた。今回は LOCAL 辞書の追加を行うことで対応したが、今後とも翻訳の誤りは続くと思料される。

### 4.2 文法的に誤った結果を示す場合

#### ・単数複数、冠詞の誤り

単数、複数の問題点としては、特に用法によって count、uncount になる名詞においての冠詞の付け方を誤るケースが多い。現段階では、前後の単語との意味的つながりから単数、複数を判断することが適当である。このために、すべてのパターンにその単語のための LOCAL 辞書を設定する必要がある、汎用性が低い膨大な知識を追加する必要がある。しかし、この手法である程度の翻訳精度向上は確実にできる。count、uncount を分離するための新たな符号を付加する今後とも修復が必要なたびに用例を追加して対応することが望ましい。しかしながら、ある程度分離できる名詞に関しては、変換辞書に登録する際に冠詞を含めて登録することにより (uncount の場合は空白の冠詞を登録する)、翻訳精度を上昇させる手法も考えられる。

#### ・その他

時制を誤るケースや類似する用例の影響を強く受け、文法を誤るケース等が見受けられた。今後とも、このような誤りは新しい翻訳を行うことで発生することは予想されるが、新たな用例やパターンをその都度追加することで当面は解決可能である。

### 4.3 木構造を決定する際の問題

木構造を修復できないケースも見受けられた。次の例は今回唯一正しく翻訳できなかったケースである。

例「誤りは何個あったか」

→"How many an error were there?"

"errors"と翻訳されるべきところが"an error"となった。"How many"のあとに続く名詞は複数形にするルールを定義すれば解決すると思える。しかし、この場合は、

1. 「誤り」は「ある」
2. 「何個」 「ある」

と構造的分割され、それぞれのパターン内で「誤り」→"error" 「何個」→"how many"と翻訳されている。つまり、ルール内で「誤り」と「何個」の間にはつながりはなく、「何個」のあとの名詞は複数形にするといったルールでの解決法は、この場合できない。構造決定する段階で「誤り」と「何個」が結びつくように用例、ルールを工夫することで解決することは「何個」と「ある」の結びつきが強い、ため、困難であった。この結びつきを強くしている原因となっている用例は「N 便」と「ある」であった。この用例はおかしな用例ではないので、削除することでの解決することは好ましくない。疑問文における文法的な誤りは今後多数発生する事が予想される。新たな対応策として、木構造の決定に意味的つながりや対訳用例だけにこだわらず、日本語そのものを正しい日本語に一度変換してから翻訳にかけ手法を提案する。例えば、「誤りは何個あったか」→「何個の誤りがあったか」という変換プロセスを行うことである。

また、木構造を決定する際に基本的に2分木であるのだが、翻訳の都合上3分木のパターンが存在する(XはYのZ等)。3分木の構造は2分木で構造を決定したい場合に悪影響を及ぼすケースが見受けられた。前述の正しく翻訳をできなかったケースも、3分木のパターンを作成すればとりあえず解決はできるのだが、他の翻訳への悪影響が強いため極力使用しない方向でいくべきである。

### 4.4 敬語表現、新たな表現の追加

敬語表現、新たな表現を付加する際に正しく翻訳できないケースが見受けられた。新しいパターンやLOCAL辞書を作成することで対応可能であるが、今後ともこのようなケースが発生することが予想され、逐次改良する必要がある。

## 5. まとめ

質疑応答文を用いて協調融合方式を適応し改良を行った結果、25文中24文まで正しく翻訳することができた。工夫をすることで、極端にClosedな知識を加えることなく非常に高いレベルまで翻訳を可能であることより、インテリジェントレクチャーなどにおいて、ある決まった枠内での質疑応答に実用化する目安がたった。今後は単語量を少しずつ増やして的確な翻訳知識を追記し、翻訳精度を上げて範囲を広げることが重要である。さらには英日翻訳、音声認識への対応が今後の課題となる。

### 参考文献

- [1]古瀬蔵、隅田英一郎、飯田仁  
経験的知識を活用する変換主導型翻訳。情処学論、Vol.35, No3, pp.414-425, 1994. [Furuse 94]
- [2]Furuse, O. and Iida, H. Constituent Boundary parsing for Example-Based Machine Translation, Proc. of Coling'94, pp.105-111, 1994
- [3]古瀬、美馬、山本、Paul、飯田：「多言語話し言葉に関する変換手動翻訳システムの評価」言語処理学会第3回年次大会発表論文集、(1997)