

# 日英話し言葉翻訳のための漸進的な英語文生成手法

松原 茂樹†

岩島 恵一†

外山 勝彦†

稲垣 康善†

†名古屋大学言語文化部

†名古屋大学大学院工学研究科

nlp@inagaki.nuie.nagoya-u.ac.jp

## 1 はじめに

機械翻訳を介した自然な多言語間対話を実現するためには、原言語入力に対して同時進行的に目標言語を出力することが可能な漸進的話し言葉翻訳システムが不可欠である。しかし、そのようなシステムを、語順が大きく異なる言語間において実現する場合、翻訳の即時性の度合と翻訳精度との間のトレードオフが問題となる。例えば、日本語を英語に変換する場合、

- 日本語入力に対して単純な順送りでの翻訳結果を出力すると意味の通らない英語文となる (翻訳精度の低下)
- 日本語述部は文の最後に生起するので、入力文の終了まで英語動詞句を出力できない (翻訳の即時性の低下)

という問題が生じる。

これを解決するためには、翻訳システムが両言語間の語順の違いを吸収する必要がある。原言語文の入力に対して、それと同じ意味内容を持ち、かつ、同時進行的に作成することが可能な目標言語文が存在するならば、それを翻訳結果として採用することは一つの方法である。実際、英語から日本語への翻訳においては、語順の自由度が高い、省略や繰り返しなどの非文法的な言語現象が頻出する、などといった日本語話し言葉の特徴を活用した変換手法が提案されており [4]、対話翻訳実験によりその実現可能性が示されている [5]。

本稿では、日本語入力に対してできる限り同時進行的に英語文を生成する手法について述べる。特に、日本語と英語の構造に関する性質の違いを吸収可能な英語文を生成するためのテクニックを示す。また、日本語対話を用いた翻訳実験の結果について報告する。

## 2 漸進的な英語文生成法

漸進的翻訳とは、入力文をその出現順に従って、できる限り小さな単位で順次変換することであり、その実現可能性には両言語間における語の生起位置の違いが大きく影響する。日本語英語間における構造上の違いを表 1 にまとめる。漸進的な日英翻訳処理を実現するためのポイントは、

- 日本語文末辺りに出現する動詞や疑問・否定表現を英語文において生成する方法とそのタイミング
- 語の生起順序があらかじめ明確でない日本語に対して、それを語順が厳格な英語に同期的に変換する方法

である。本節では、図 1 の 1), 2), 3) の違いに対処するためのアイデアについて述べる。

表 1: 日本語と英語との構造的特徴に関する比較

項目	日本語	英語
1) 動詞の出現位置	文末の近辺	動詞句の先頭
2) 語順の自由度	比較的的自由	厳格
3) 疑問表現の位置	文末	文頭
4) 否定表現の位置	文末	動詞句の先頭
5) 省略	頻繁 (特に主格)	比較的少ない

### 2.1 英語動詞の予測

日本語と英語では動詞の生起位置が異なる。よって、日本語文が最後まで入力されないと、出力すべき英語動詞を決定することはできない。例えば、日本語文

(2.1) 彼は 東京へ 電車で 行きました。

において述語「行きました」が入力されないと、対応する動詞 “went” を出力できない。

しかし日本語では、文を構成する動詞以外の構成要素から、その文の動詞をある程度予測することができる [6]。特に、対話タスクを限定すると使用される動詞も制限されるため、予測精度も向上すると考えられる。日本語動詞が入力される前に、すでに生成された英語名詞句からその文の動詞を予測できるなら、それを出力することにより即時性を損なうことなく翻訳処理を行える [2]。

### 2.2 語順の入れ換え

英語の語順は一般に厳格であるが、副詞や副詞句の生起位置については比較的の自由度が高い。そこで、日本語文中の早い段階で出現し、かつ、副詞的な役割を果たす日本語後置詞句については、それに対する英語副詞 (句) を即座に出力しても、全体として意味の通る英語文となる可能性がある。例えば、日本語文

(2.2) 今朝、ケン は 散歩に 出かけた。

における「今朝」の入力に対して、英語前置詞句 “in this morning” を即座に出力すると、生成される英語文は、

(2.3) In this morning, Ken took a walk.

となり、漸進性を保持した英語文生成が可能となる。

### 2.3 受身形への態の変換

日本語では主格は頻繁に省略されるため、英語の主語を生成するためには、それを補完するメカニズムが必要になる。しかし、態の変換により日本語文の対象格を英語の主語とすることができれば、省略された日本語主格を必ずしも訳出する必要はなくなる。また、日本語対象格を英語主語として訳出できれば、英語動詞を出力するタイミングを遅らすことができ、動詞予測精度の向上も期待できる。

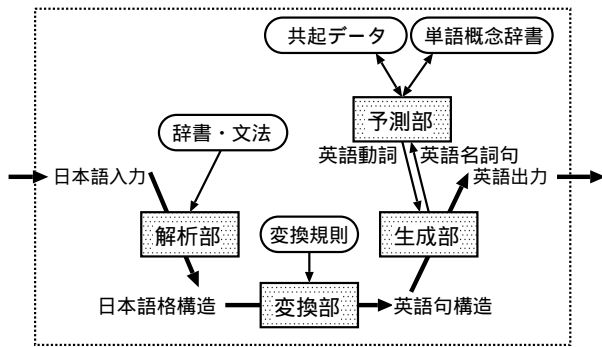


図 1: 漸進的な日英話し言葉翻訳システムの構成

## 2.4 イントネーションの利用

英語疑問文は通常、「Do ~ ?」のように文頭に疑問を表す語句が現れる。一方、日本語では「～ か?」のように文末に現れるため、日本語文全体が入力されないとそれが疑問文であるかどうかを判別できない。しかし、英語会話中では、疑問を示す語句を用いなくても文末のイントネーションを上げることにより、疑問文であることを表現することができる。

## 2.5 言い直しの活用

2.1 節から 2.4 節で示した方法は、いずれも有用であると考えられるが、動詞の予測を誤る、不適切なタイミングで訳出する、などといったことが起こる可能性があり、どれも完全ではない。本稿では、誤った英語表現の生成を翻訳システムによる言い誤りとみなす。すなわち、人間による対話のように、システムが言い誤りに気付いた時点で言い直しを生成することにより、結果として正しい翻訳結果を作り上げる。

## 3 漸進的な日英話し言葉翻訳システム

漸進的な日英話し言葉翻訳システムの構成を図 1 に示す。システムは、解析部、変換部、生成部、及び予測部という 4 つのモジュールから構成されており、日本語文節単位で処理を実行し、順次英語を生成する。すなわち、文節が入力されるたびにそれまでの入力に対する日本語格構造を作成し(解析部)、それに変換規則を適用することにより英語句構造を順次作り上げる(変換部)。さらに、前節で示したアイデアに基づく英語文生成手続き [3] に従って、順次、英語表現を出力する(生成部)。

なお生成部では、英語句構造の中に新たな名詞句が格納されるたびに、それを予測部に送る。予測部では、名詞句を EDR 単語概念辞書 [7] を用いて概念化し、かつ、対話コーパスをもとに作成した共起データを参照することにより、それまでに入力された名詞句と共起する可能性が最も高い英語動詞を抽出し、それを生成部に送る。ただし、実際に出力する英語動詞は、動詞を出力すべきタイミングであると生成部が判断した時点で予測部が抽出した結果である。

## 4 翻訳実験

本稿で提案した英語文生成手法に基づく漸進的な日英翻訳の実験システムを作成した。ATR 対話データベース [1]

表 2: 翻訳正解率に関する実験結果

項目	誤りの原因	発話数	割合 (%)
正解		140	54.1
不正解	主格の省略	79	30.5
	解析誤り	33	12.7
	構文的曖昧性	7	2.7

の中の電話による旅行の申し込みをタスクとする 4 つの対話の解析のために、540 語からなる辞書を構築した。また、同様のタスクをもつ 15 対話から抽出し、EDR 単語概念辞書 [7] を用いて概念化した名詞句と動詞との共起関係を示す 1,883 規則を、英語動詞を予測するための共起データとして登録した。実験は、日本語話者による 259 発話に対して行った。動詞予測に関するオープンテストの結果、予測精度は 53.1% であり、日英翻訳における英語動詞予測の利用可能性を確認した。翻訳正解率に関する結果を、翻訳誤りの主な原因を含めて表 2 に示す。正解に分類された発話の 57.1% に対して、2 節で示した英語文生成手法を用いており、本手法の有効性を確認できた。

本手法では、図 1 の 4), 5) に対処していないが、特に、5) が漸進的な日英翻訳における最大の失敗要因となった。省略された語句を、その文の中の情報のみを用いて、発話の早い段階で補完することは難しく、文脈情報の効果的な活用が有効であると考えられる。さらに、日本語動詞が入力されていない段階で、日本語名詞句の格を決定することもまた困難であり、その影響が「解析誤り」や「構文的曖昧性」といった失敗原因として表われている。

## 5 おわりに

本稿では、漸進的な日英話し言葉翻訳のための英語文生成手法について述べた。漸進的で、かつ、意味の通じる英語文を生成することが可能な日英翻訳システムの実現に、本手法が有効であることを確認した。

作成した実験システムが処理可能な日本語発話は、単文に限られているとともに、話し言葉の特徴である文法的不適格性もまた扱うことができない。今後は、漸進的な日本語話し言葉解析手法について検討する必要がある。

## 参考文献

- [1] 江原 他: ATR 対話データベースの内容, Technical Report TR-I-0186, ATR 自動翻訳研究所 (1990).
- [2] 岩島, 松原, 外山, 稲垣: 英語動詞の予測に基づく漸進的な日英話し言葉翻訳手法, 情報処理学会第 56 回全国大会講演論文集 (2), pp. 281-282 (1998).
- [3] 岩島 恵一: 漸進的な日英話し言葉翻訳システムに関する研究, 名古屋大学大学院修士論文 (1998).
- [4] Matsubara, S. and Inagaki, Y.: Utilizing Extra-Grammatical Phenomena in Incremental English-Japanese Machine Translation, *Proc. of TMI-97*, pp.31-38 (1997).
- [5] 松原, 浅井, 外山, 稲垣: 不適格表現を活用する漸進的な英日話し言葉翻訳手法, 電気学会論文誌 C, Vol.118-C, No.1, pp.71-78 (1998).
- [6] 水野 的: 日英同時通訳研究ノート, 通訳理論研究, Vol.5, No.2 (1995).
- [7] (株) 日本電子化辞書研究所: EDR 電子化辞書仕様説明書, 第 1.5 版 (1995).