

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

6Q-8

岩島 恵一 松原 茂樹 外山 勝彦 稲垣 康善

名古屋大学大学院工学研究科

{iwashima,matu,toyama,inagaki}@inagaki.nuie.nagoya-u.ac.jp

1 はじめに

自然な対話を対象とした精度の高い翻訳処理システムが実現されている [3, 7]. 特に, 対話翻訳システムには, 入力された原言語を実時間で翻訳し, 即座に出力することが求められる. それに対して, 話者の発声途中の段階でそれまでの入力に対する意味を確定し, その対訳を即座に生成する手法が提案されている [6]. しかし, 節などを単位として意味を確定するため, 翻訳の即時性が十分に実現されているとは言い難い. 即時性のより高い翻訳処理を実現するために, 原言語入力に対して同時進行的に目的言語を生成することが考えられる [4].

ところが, 日本語を英語に翻訳する場合, 両言語は語の生起順序が異なるために, 日本語単語の入力に対して対応する英語表現を即座に出力できないことがある. 特に日本語では, 動詞が文の最後に現れるため, 対応する英語動詞を日本語文の入力が終わるまで出力できず, その結果, 翻訳処理の漸進性が大きく損なわれる可能性がある.

そこで, 本稿では, 名詞の組合せと動詞との共起関係を用いて, 既入力の名詞から訳出すべき英語動詞を予測することにより, 漸進的な日英翻訳を実現する手法を提案する. また, 対話コーパスをもとに名詞の組合せと動詞との共起関係規則を自動的に作成し, それらを用いて動詞の予測実験を行った. その結果, 本手法が漸進的な日英翻訳に利用可能であることを確認した.

2 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. 解析部, 変換部, 生成部ではそれぞれ, 日本語構造の構築, 英語構造への変換, 英語表現の生成を日本語の入力に対して同時進行的に実行する. 予測部では, 生成部から英語表現を随時受け取り, それと共起関係規則を用いて英語動詞を決定し, その結果を生成部に渡す.

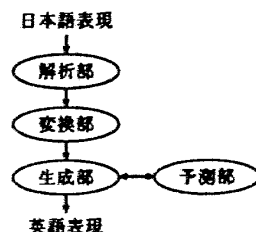


図1: 漸進的な日英話し言葉翻訳システムの構成

3 共起関係規則を用いた動詞の予測手法

3.1 漸進的な日英翻訳における基本的問題

漸進的な日英話し言葉翻訳システムを実現する上で, 両言語間の動詞の生起位置の違いは, 翻訳の漸進性を損なう重大な要因となる. 例えば, 日本語文

(3.1) 彼は東京へ電車で行きました.

に対し, 英語文

(3.2) He went to Tokyo by train.

を漸進的に生成する場合を考える. まず, 日本語入力「彼は」に対して, 対応する英語表現「He」を出力できる. ところが, それ以降の入力「東京へ」, 「電車で」にそれぞれ対応する英語表現の出力は, 英語の語順に関する制約のために, 日本語動詞「行きました」の入力とその翻訳出力が行われるまで実行できない. すなわち, 入力に対する翻訳結果の出力が遅れる.

これに対して, 入力「東京へ」に対して訳出すべき英語動詞「went」を予測し, 英語表現「went to Tokyo」を出力することにより, 翻訳の漸進性を確保することが考えられる. 既入力の名詞だけから単に動詞を予測することは一般には困難であるが, 対話タスクを定めれば出現する文のタイプがある程度限定されるため, 名詞の組合せと動詞との共起関係を用いることにより, 動詞の予測が可能となる [5].

3.2 共起関係規則の作成

英語動詞の予測に用いる名詞の組合せと動詞との共起関係規則は, 対話コーパスから自動的に作成した. コーパス中の各英語対話文から抽出した名詞, 前置詞および動詞を一つの共起関係規則とする. なお, 名詞はEDR 単語概念辞書 [2] を用いて概念化しておく. 例えば, 英語文

(3.3) The trip to Tokyo costs you 130,000 yen.

から作成される共起関係規則は,

(3.4) (0de5c9, ϕ) (1f39a7, to) (2dc303, ϕ) (0e24d3, ϕ) \Rightarrow cost
 である。ここで, “0de5c9”, “1f39a7”, “2dc303” および “0e24d3” は, それぞれ名詞 “trip”, “Tokyo”, “you” および “yen” の概念識別子である。概念化することにより, 名詞 “trip” がそれと同じ概念に属する名詞 “tour” と入れ替わった英語文

(3.5) The tour to Tokyo costs you 130,000 yen.

に対しても, 同一の規則により対応することができ, “tour”, “Tokyo” などとこの規則とから動詞 “cost” を予測できる。また, 前置詞句 “to Tokyo” の中の名詞 “Tokyo” に対しては, “(1f39a7, to)” のように前置詞 “to” を規則内に記述する。これにより, 動詞が名詞の組合せ “He”, “Tokyo” と共起する2つの英語文

(3.6) He came from Tokyo.

(3.7) He went to Tokyo.

から, 異なる規則を作成することができ, 入力 “He”, “to Tokyo” から動詞 “went” を予測できる。

3.3 予測部における動詞の予測手続き

予測部では, 入力の各時点において生成される名詞あるいは名詞と前置詞との組合せに対して, 共起関係規則を適用することにより動詞を決定する。動詞の予測手続きを以下に示す。

【動詞予測手続き】

step1 概念識別子が C_n の名詞 n のみが入力された場合, “(C_n, ϕ)” を含むすべての共起関係規則に得点を加える。概念識別子が $C_{n'}$ の名詞 n' と前置詞 p との組合せが入力された場合, “($C_{n'}, p$)” を含むすべての共起関係規則に得点を加える。

step2 得点が最も高い規則に含まれる動詞を予測結果として出力する。

生成部から名詞あるいは名詞と前置詞との組合せが入力されるたびに, この手続きを適用する。入力文に出現した名詞あるいは名詞と前置詞との組合せを含む共起関係規則は得点が高くなるので, それまでの入力に対して適切な動詞を決定できる。

3.4 動詞の予測実験

共起関係を用いた動詞予測手法の有用性を確認するために, ATR 対話データベース [1] 中の電話による旅行の申し込みをタスクとするバイリンガル対話コーパスを用いて予測実験を行った。このコーパスの中の15対話(1883発話)をもとに共起関係規則を作成した。また, 動詞予測システムへの入力として, 上述の15対話とは異なる7対話(450発話)を使用し, 英訳された名詞あるいは名詞と前置詞との組合せを, コーパス中の日本語文の生起順序に従って入力した。また, 動詞予測手続きの step1 において, 共起関係規則に対して与える得点はすべて1点とした。

実験の結果, 生成部が動詞を必要とする時点における動詞の予測精度は53.1%であった。これは, 漸進的な日英翻訳において, 本手法が基本的に利用可能であることを示している。

4 動詞の予測に基づく漸進的な翻訳処理

本手法を用いることにより漸進的な日英翻訳処理が可能となる。日本語文(3.1)を漸進的に翻訳することを考える。まず入力「彼は」に対して “He” を出力できるが, 次の入力「東京へ」に対応する英語表現 “to Tokyo” を出力するためには, 動詞が出力されている必要がある。そこで, それまでの入力と共起関係規則とを用いて動詞を予測する。この時点までの名詞の組合せ “He” および “to Tokyo” から “went” を予測することにより, 入力「東京へ」に対して英語表現 “went to Tokyo” を即座に出力できる。次の入力「電車で」に対して英語表現 “by train” を出力することにより, 日本語文(3.1)から英語文(3.2)への漸進的な翻訳処理を実現できる。

5 おわりに

本稿では, 限定された対話タスクの中で, 名詞の組合せと動詞との共起関係を利用して英語動詞を予測することにより, 漸進的な日英翻訳を実現する手法を提案した。名詞あるいは名詞と前置詞との組合せの入力に対して, それを含む共起関係規則に対して得点を加算することにより, それまでの入力の組合せに対して適切な動詞を決定できる。実験の結果, 本手法の有用性を確認した。現在, 解析部, 変換部, 生成部(図1)における漸進的な処理について検討している。

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 自動翻訳電話研究所 (1990).
- [2] EDR 電子化辞書仕様説明書 (第1.5版), (株)日本電子化辞書研究所, (1995).
- [3] Furuse, O. and Iida, H.: Cooperation between Transfer and Analysis in Example-Based Framework, *Proc. of COLING '92*, pp. 645-651 (1992).
- [4] Inagaki, Y. and Matsubara, S.: Models for Incremental Interpretation of Natural Language, *Proc. of SNLP '95*, pp. 51-60 (1995).
- [5] 岩島, 松原, 稲垣: 漸進的な日英機械翻訳手法の検討 - 用例を用いた動詞の推測に基づくアプローチ -, 電気関係学会東海支部連合大会講演論文集, p. 312 (1996).
- [6] 美馬, 古瀬, 飯田: 同時通訳システムの実現に向けた漸進的翻訳処理, 情報処理学会第54回全国大会, pp. 11-12 (1997).
- [7] Sumita, E. and Iida, H.: Experiments and Prospects of Example-based Machine Translation, *Proc. of ACL '91*, pp. 185-192 (1991).