

1T-02 パターンベース翻訳システム PalmTree の構文解析

宮平知博

日本アイ・ビー・エム株式会社

1. はじめに

パターンベース翻訳システム PalmTree [1, 2, 3] は、同期文法 [4] の考え方に基づいて翻訳の解析から生成までを一つの枠組みで処理する英日翻訳システムである。共起処理 [5] や辞書の自動切り換え機能 [6] などにより翻訳精度は向上しつつあるが、翻訳パターン単体の単純なマッチングを繰り返すことで翻訳を実行する現在の翻訳エンジンの枠組みだけで対応しようとすると、様々な組み合わせを考慮して大量のパターンをあらかじめ用意しなければならない場合がある。これを解決するため、構文解析実行時に原文に適合するパターンを動的に生成する機能を実装したので、それについて報告する。

2. 実行時の動的パターン生成

構文解析実行時の動的パターン生成は、大きく2つに分けられる。パターン変形とパターン合成である。パターン変形は、既存のパターンを翻訳しようとする文章に合う形に変形させるものである。動詞句パターンの語順を入れ換えることで、関係代名詞節や受動態に適用可能なパターンを生成する。パターン合成は、既存の複数パターンを合成することで翻訳しようとする文章に合う形の新たなパターンを生成するものである。パターンを単純に組み合わせる従来の枠組みでは解析がうまく行なえなかった場合に対応することができる。

3. パターン変形

パターンベース翻訳では、翻訳パターンと原文の単純なマッチングによって構文解析を実行する。そのため、関係代名詞節や受動態で語順が変わると能動

文のために作成された動詞句パターンは使用されず、動詞のデフォルトの訳が出力されるだけだった。たとえば、

admire ~ for ~ = ~の ~に 感心する
という動詞句パターンが登録してあっても、

tha man whom I admire for ~
のような語順が異なる文にはマッチせず、admire のデフォルトの訳が使用されるだけだった。

3.1 関係代名詞節

目的語を先行詞とする関係代名詞節では動詞句の語順が変わってくる。それを考慮して、翻訳しようとする文が関係代名詞節を含むと判断される場合には、動詞句パターンを変形して目的語を前に出した形の名詞句パターンを構文解析実行時に生成する。たとえば、

admire:VERB:1 NP:2 for NP:3 → VP:1

VP:1 ← NP:2 の NP:3 に 感心する:1

という動詞句パターンから

NP:2 whom NP:5 admire:VP:1 for NP:3 → NP:2

NP:2 ← NP:5 が NP:3 に 感心する:1 NP:2

NP:2 NP:5 admire:VP:1 for NP:3 → NP:2

NP:2 ← NP:5 が NP:3 に 感心する:1 NP:2

NP:3 for which NP:5 admire:VP:1 NP:2 → NP:3

NP:3 ← NP:5 が 感心する:1 NP:2 の NP:3

NP:3 NP:5 admire:VP:1 NP:2 for → NP:3

NP:3 ← NP:5 が 感心する:1 NP:2 の NP:3

などの名詞句パターンが生成される。動詞の目的語を先行詞とする場合と共に、前置詞の目的語を先行詞とする場合にも対応している。

3.2 受動態

受動態の文の場合にも、語順が変わってくるため、動詞句パターンはそのままでは使用されない。それに対応するために、翻訳しようとする文の動詞が受動態になっていると判断される場合には、動詞句パターンを変形して動詞の過去分詞形に関連するパターンを構文解析実行時に生成して使用する。

たとえば、

admire:VERB:1 NP:2 for NP:3 → VP:1

VP:1 ← NP:2 の NP:3 に 感心する:1

という動詞句パターンから

NP:2 be:AUX admire:VP:1 for NP:3 → S:1

S:1 ← NP:2 が NP:3 に 感心される:1

NP:2 admire:VP:1 for NP:3 → NP:2

NP:1 ← NP:3 に 感心される:1 NP:2

NP:2 to be admire:VP:1 for NP:3 → NP:2

NP:1 ← NP:3 に 感心される:1 NP:2

being admire:VP:1 for NP:3 → VP:2

VP:2 ← NP:3 に 感心される:1

to be admire:VP:1 for NP:3 → VP:2

VP:2 ← NP:3 に 感心される:1

be:AUX NP:2 admire:VP:1 for NP:3 → S:1

S:1 ← NP:2 が NP:3 に 感心される:1

という文・名詞句・動詞句のパターンが生成され、翻訳しようとする文にマッチするかどうかチェックされる。

自動詞・他動詞の両方の用法を持つ動詞の場合には「名詞+自動詞過去」の原文があると「名詞+他動詞過去分詞」という生成パターンによる名詞句の訳になってしまうことがあるので、パターンのコストの調整によって自動詞過去の方が優先されるようにしている。また、受動態パターンがあらかじめ辞書に登録されている場合もあるので、その場合には、パターン生成は行なわない。

4. パターン合成

次の場合に、実行時に複数のパターンから動的にパターンを合成し、構文解析に使用している。

- ・ 2つのパターンの一部が重複する場合
- ・ 並列なパターンの一部が省略される場合

これらをそれぞれ、重複合成、等位合成と呼ぶ。

4.1 重複合成

パターンの原語側の一部が重複し、かつ、訳語側も同様に一部が重複する2つのパターンを合成し、新たなパターンを生成する。たとえば、

static RAM → NOUN

NOUN ← スタティックRAM

RAM card → NOUN

NOUN ← RAMカード

の2つのパターンから

static RAM card → NOUN

NOUN ← スタティックRAMカード

というパターンを合成して適用する。

従来は、(static) + (RAM card) と解析されて「静的なRAMカード」と訳されていた。

4.2 等位合成

等位接続詞の前後いずれかにパターンがマッチした場合に、その一部を等位接続している他方に追加するとマッチするパターンが存在する場合に、訳語側でその部分を補って等位接続したパターンを生成する。たとえば、

summer vacation → NOUN

NOUN ← 夏季休暇

winter vacation → NOUN

NOUN ← 冬季休暇

の2つのパターンから、

summer and winter vacation → NP

NP ← 夏季休暇と冬季休暇

というパターンを合成して適用する。

従来は、(summer) and (winter vacation) と解析されて、「夏と冬季休暇」と訳されていた。

5. まとめ

パターンベース翻訳エンジン PalmTree に新たに導入した構文解析実行時の動的パターン生成に関して述べた。パターンの変形により新たなパターンを生成するパターン変形と、複数パターンの合成により新たなパターンを生成するパターン合成を実装した。これらによって従来のパターンでは対応できなかった文の翻訳が可能になっている。しかし、まだ対応できないケースがあるので、パターン生成機能のさらなる強化を検討したい。

[参考文献]

- [1] Takeda, K., "Pattern-Based Machine Translation", Proc. of 16th Coling, Vol. 2, pp. 1155-1158, 1996
- [2] 渡辺, 武田, "パターンベース翻訳システム: PalmTree", 情報処理学会第55回全国大会, 1997
- [3] 宮平, 渡辺, 田添, 神山, 武田, "インターネット機械翻訳の世界", 毎日コミュニケーションズ, 2000
- [4] Shieber, S., and Shabes, Y., "Synchronous Tree Adjoining Grammars," Proc. Of COLING 90, 1990
- [5] 宮平, 渡辺, 武田, 那須川, "パターンベース翻訳システム PalmTree の文脈処理", 情報処理学会第57回全国大会, 1998
- [6] 宮平, 神山, 羽鳥, "パターンベース翻訳システム PalmTree の訳語選択", 情報処理学会第59回全国大会, 1999