

日英機械翻訳システム JETS における翻訳パターン 生成支援環境*

2K-2

渡辺 日出雄†

日本アイ・ビー・エム株式会社 東京基礎研究所‡

1 はじめに

我々は現在 Example-Based Approach (EBA)[2] をトランスファー部分 [5] に用いた日英翻訳システム JETS を開発している。従来のルールベースの翻訳システムではユーザーが翻訳パターンを登録するのはあらかじめ決められたパターンに限定されているのが多いが、我々のシステムでは EBA の特徴をいかして任意の翻訳パターンを登録できることを目指している。

ユーザー¹に翻訳パターンの登録をしてもらうといっても、パターンは構文解析された内部構造のペアとして記述しなければならないが、いきなりこれをユーザーに行なってもらうのは難しい。そこで、この負担を軽減するべく、翻訳例から翻訳パターンを（半）自動的に取り出そうという試みが最近盛んになってきている。手法としては、(1) 翻訳対を構文解析して対応付けをするもの [1, 4] と、(2) システムによる翻訳結果と正しい翻訳との比較から必要な翻訳パターンを見つけ出すというもの [6] がある。

本稿では、上記 (1) (2) の手法の簡単な説明とそれらを実現した翻訳パターンの生成支援システムについて報告する。

2 翻訳対の対応付け

これは入力された日本語文と英語文を解析して対応する部分を関連付けるという処理を行なう。我々は、[1, 4] に比べると非常に簡単な手法を用いている。

まず、日本語文と英語文はそれぞれの言語用のパーサー²を持ちいて依存構造へ変換される。その後、日英の辞書を用いて対応するノードどうしを関連付ける。その際に、(1) ルートノードどうしは必ず関連付けをし、(2) 英語側に対応するものが複数ある場合は日本語側でそのノードの近傍にある別のノードで既に対応関係が付けられているものを探し、その対応する英語に近いものを選ぶ、というヒューリスティクスを用いている。

*An Assisting Environment for Creating Translation Patterns in Japanese-to-English MT System JETS

†Hideo Watanabe (watanabe@trl.vnet.ibm.com)

‡IBM Research, Tokyo Research Laboratory

¹ここでのユーザーとはユーザーサイトでの翻訳システムの管理運営をする人を想定している。決して、エンドユーザーのことではない。

²日本語のパーサーは JETS のものを用い、英語のパーサーはやはり我々の研究所で開発中の英日翻訳システム SHALT2 のものを用いている。

3 正しい翻訳結果との比較

これは MT システムの翻訳結果とユーザーが入力した正しい翻訳結果を比較することにより現在のシステムに欠けている翻訳パターンを取り出すものである。

MT システムが用いた翻訳パターンは既に分かっているので日本語の入力構造 (G_i) と英語側の出力構造 (G_j) との対応関係も予め分かっている。そこで、まず前節で述べた手法を用いて G_i と正しい英語の解析結果 (G_0) との対応関係を求める。次に、MT システムが用いたそれぞれの翻訳パターンについてそれが $G_i - G_0$ ペアでも用いることが出来るかどうかを調べ、用いることが出来ない場合その $G_i - G_0$ ペア中での対応する翻訳パターンと、それらの翻訳パターンではカバーできない英語側の部分構造を一つにして新たな翻訳パターンとする。

4 翻訳パターン生成支援環境

今まで述べてきた二つの手法は完全ではないのでユーザーの訂正が必要となる。これを容易にするため翻訳パターン生成支援環境 TranPet を作成した。

図 1 上に任意の日本語と英語から翻訳パターンを生成しているところを示している。最初に日本語と英語のソースがパーサーで解析され、真中の行のウィンドウに解析結果が木構造の形で表示されている。一番下のウィンドウには対応するノードのペアが表示されている。木構造の任意のノードの上でクリックすると対応する相手側のノードが反転し容易に対応関係が分かるようになっている。対応関係や解析結果に誤りがある場合は直接内容を編集できるようになっている。

図 1 下は正しい翻訳との比較により翻訳パターンを生成するところを示している。左上のウィンドウに日本語の依存構造、その下に英語の依存構造、右側の真中のウィンドウに正しい英語の依存構造を木構造で表示している。これらもやはりノードをクリックすることにより対応するノードが反転するようになっている。解析結果や対応関係に誤りがなければ、"Find Pattern" ボタンを押すことにより必要な翻訳パターンが得られる。

5 おわりに

自由に用例を翻訳パターンとして追加できるという EBA の特徴を活かして、任意の翻訳パターンの生成を支援するツールと、MT システムの翻訳結果と正しい出力から

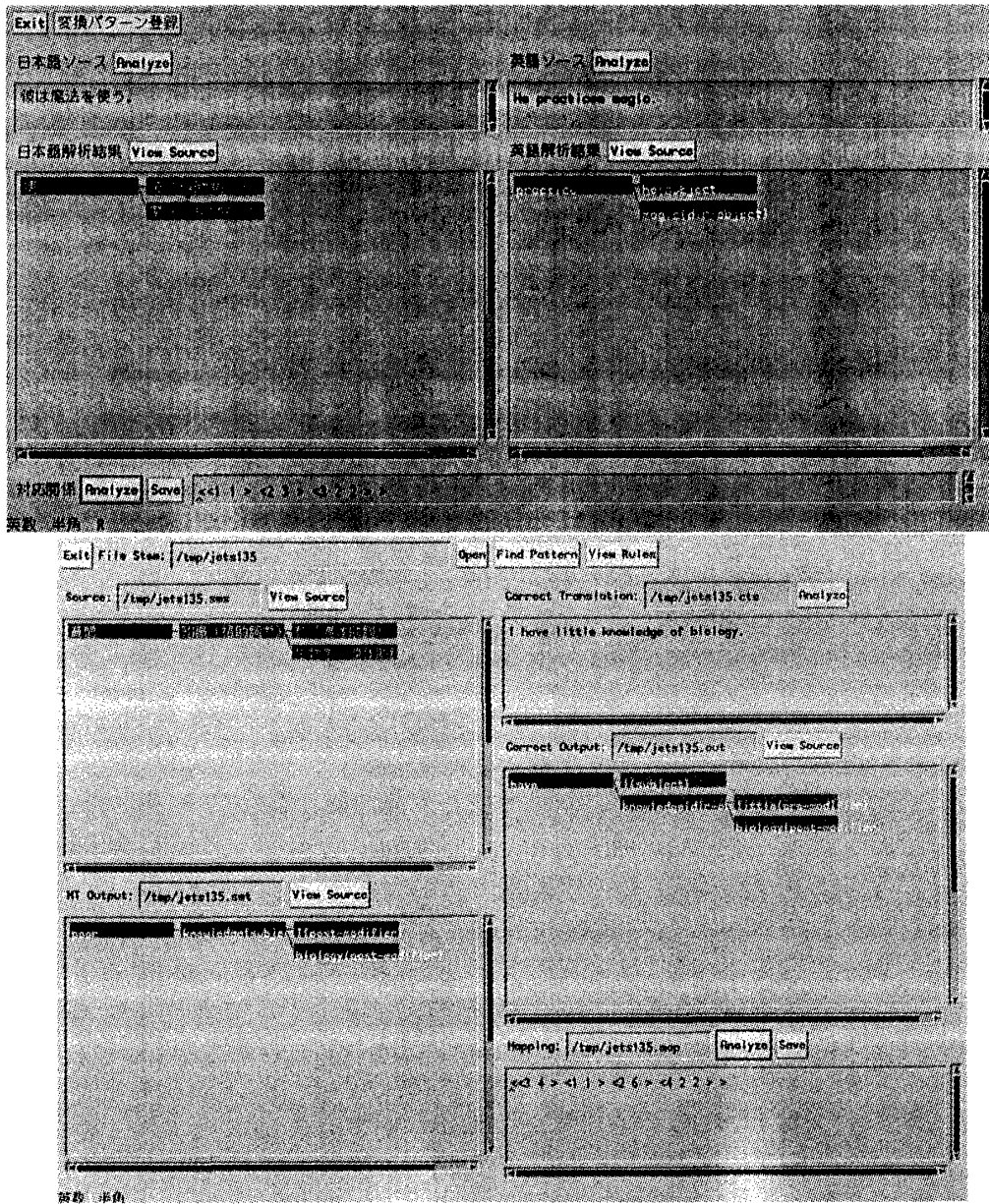


図 1: 任意の翻訳パターンの生成 (上) 及び比較による翻訳パターンの生成 (下) の画面

現在のシステムに足りない翻訳パターンの生成を支援するツールを作成した。従来 EBA をトランスファー処理に用いる場合の問題の一つとして用例を生成するのが大変であることが指摘されていたが、自動的に生成する手法とこのような支援環境を組み合わせることによりかなり容易になってきたといえる。

今後の課題としては、やはりユーザーの訂正作業を少なくするためにも自動的に生成するアルゴリズムを改良していくことが必要である。また、ユーザーインターフェースの部分についても、より使いやすいものにしてゆくつもりである。

参考文献

[1] Kaji, H., Kida, Y., and Morimoto, Y., "Learning Trans-

lation Templates from Bilingual Text," Proc. of 14th Coling, 1992

[2] Nagao, M., "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle," Elithorn, A. and Banerji, R. (eds.): *Artificial and Human Intelligence*, NATO 1984

[3] Sato, S. and Nagao, M., "Toward Memory-Based Translation", Proc. of 13th Coling, 1990

[4] Utsuro, T., Matsumoto, Y., and Nagao, M., "Lexical Knowledge Acquisition from Bilingual Corpora," Proc. of 14th Coling, 1992

[5] Watanabe, H., "A Similarity-Driven Transfer System," Proc. of 14th Coling, 1992

[6] Watanabe, H., "A Method for Extracting Translation Patterns from Translation Examples," Proc. of 5th TMI, 1993