

述部依存構造の類似性を利用した英日変換と

木村 直樹

松原 茂樹

小川 泰弘

稲垣 康善 (名)

1 はじめに

英日機械翻訳における変換処理では、英語構造をもとに日本語構造を作成するが、英語と日本語では一般に、構文的構造が異なる。そのため、構文変換のための規則が必要であり、特に、述部は、動詞や助動詞などといった多様な要素から構成されるため、変換処理が煩雑になることが多い。そこで本稿では、構造変換のための規則を用いることなく、英語述部を日本語述部に変換する方法について述べる。英語と日本語の間では、述部の依存構造が類似しているため [4]、英語述部における構成素を、対応する日本語構成素に置き換えさえすれば、日本語述部の構造を作成できる。ここで依存構造とは文の構成素間の支配、従属の関係を表す構造である。また、英日翻訳実験による本手法の評価について報告する。

2 英日述部の翻訳方法

本節では、英語述部を日本語述部に翻訳する方法について述べる。英語と日本語の間では、述部を構成する要素間の依存関係は類似しているため、英日の述部を構成する要素の対応関係を特定し、英語述部の各構成素を、対応する日本語構成素に置き換えることにより、日本語述部の依存構造を作り上げることができる。

本手法は、解析、変換、生成の3つのフェーズから構成される。各フェーズにおける処理は以下の通りである。

- 1) 解析 入力英語文の述部に対して依存構造を作成する。主辞情報が付与された CFG を用いて作成した句構造木をもとに述部構成素間の依存関係を決定する [5]。
- 2) 変換 英語述部の依存構造を日本語述部の依存構造に変換する。ただし、英語の動詞や助動詞は、語尾を変化させ、原形そのものもつ意味に加え、文法的機能を含んだ形で出現することがある。このため、動詞や助動詞を原形と文法的機能に分解する。文法的機能として、“[pres]”(現在形)、“[past]”(過去形)、“[pres-p]”(現在分詞形)、“[past-p]”(過去分詞形)がある。“[pres-p]”、“[past-p]”は、直前の be 動詞や完了形で用いられる“have”と結合する。図1に変換処理の例を示す。図中の矢印は、矢印の先の構成素が矢印の元の構成素に依存することを表す。“was”、“written”を原形“be”と過去形の“[past]”、原形“write”と過去分詞形の“[past-p]”にそれぞれ分解し、“[past]”を“[過去]”に、“be”と“[past-p]”を結合し“[受身]”に、“write”を“書k”に変換し、日本語述部依存構造を作成する。
- 3) 生成 日本語依存構造をもとに述部を生成する [2]。語順決定、接尾辞決定、形態素連結の順に行う。生成処理の例を図2に示す。日本語では各形態素はそれが依存する形態素よりも先に出現するため、述部内の語順は“書k”、“[受身]”、“[過去]”となる。“[受身]”、“[過去]”を接尾辞“-(r)are-”、“-(i)ta”にそれぞれ置き換え、最後にこれらを派生文法 [3] に従って接続し、述部を作り上げる。

3 評価

本手法の利用可能性を評価するため、本手法を GNU Common Lisp 2.2.2 を用いて実装し、述部翻訳実験を行った。文献 [1] で提示された助動詞の導出方法を参考に、規則数

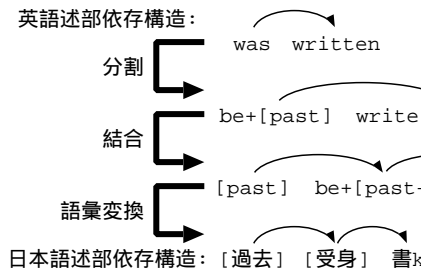


図1: 日本語述部依存構造への変換

37, 範疇数 39 の規模で文脈自由文法を生成し、動詞を対象とした依存関係をもとに主辞情報から抽出した 1000 文中に含まれる全 1247

94.9% に相当する 1183 述部が正しく構造変換に利用できることがわかった。“does not understand”のように“not”否定を表す副詞を含む述部では、両言語な問題に対して、“not”を除いた部分に形態素を生成処理において加えることが

4 まとめ

本稿では、英語と日本語の述部の依存した。構成素間での変換により英日述部翻訳実験の結果、94.9%の述部が正しく翻訳性を確認した。

参考文献

- [1] 荒木, 小野, 中野: 現代の英文法 9
- [2] 木村, 松原, 小川, 外山, 稲垣: 英日述部変換に基づく方法 - 情報処理学会第 60 回大会論文要旨集, pp. 491-49
- [3] 清瀬: 日本語文法新論 - 派生文法序論
- [4] 児玉: 語順の普遍性, 山口書店(1987)
- [5] 村瀬, 松原, 外山, 稲垣: 依存関係に基づく英日述部変換, pp. 491-49
- [6] 日本電子化辞書研究所: 英語共起辞書
<http://www.ijnet.or.jp/edr/TG-V>