

英日機械翻訳における日本語述部生成

— 派生文法に基づく方法 —

木村 直樹^{†1}松原 茂樹^{†2†4}小川 泰弘^{†3}外山 勝彦^{†3†4}稲垣 康善^{†3}^{†1}名古屋大学工学部^{†2}名古屋大学言語文化部^{†3}名古屋大学大学院工学研究科^{†4}名古屋大学統合音響情報研究拠点 (CIAIR)

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

1 はじめに

構文トランスファー方式に基づく機械翻訳システムは、通常、解析・変換・生成の3つのモジュールから構成される。このうち、生成モジュールでは、目標言語構造をもとに目標言語文を作り上げるが、英日翻訳では日本語述部の生成方法が問題となる。すなわち、学校文法に基づく一般的な生成手法では、用言や助動詞が活用することを前提とするため、各形態素の語形を決定する処理が複雑になる[4]。

そこで本稿では、英日機械翻訳のための、派生文法[1]を用いた日本語述部生成手法を提案する。派生文法では日本語述部を、活用という語形変化としてではなく、語幹への接尾辞の接続として記述する。そのため、活用処理が不要となるとともに、形態素の接続に関する規則も単純となり、述部の生成が容易になる。

本手法では、英日翻訳の変換処理で作成する日本語構造として依存構造[2]を採用する。日本語依存文法の係り受け制約に従って語順を定め、派生文法の形態素接続規則に基づき述部を作り上げる。本手法を計算機上にインプリメントし、その実現可能性を確認した。

2 派生文法

本節では、派生文法[1]について概説する。なお、派生文法は音韻論に基づくため、以下では、日本語表記の一部にローマ字を用いる。

派生文法では、動詞の語幹は音素単位で扱われ、例えば、動詞「書く」の語幹は「書」ではなく、「書k」となる。動詞の語幹のうち、「書k」のように子音で終わる語幹を子音幹といい、「食べル」の語幹「食be」のように母音で終わる語幹を母音幹という。

学校文法における活用語尾や助詞、及び助動詞を、派生文法ではいずれも接尾辞として扱う。動詞の語幹に接続する接尾辞を動詞接尾辞という。動詞接尾辞は、新しい語幹を派生する派生接尾辞と新たな語幹を派生しない統語接尾辞に分類される。語幹と接尾辞の接続には制約がある。

また、用言の語形変化を語幹への接尾辞の接続として表

現する。例えば、「書カセタ」は子音幹「書k」、使役を表す派生接尾辞「-(s)ase-」、完了を表す統語接尾辞「-(i)ta」の3つの形態素から構成される。「-(s)ase-」、「-(i)ta」の括弧内の音素はそれぞれ連結子音、連結母音と呼ばれる。連結子音をもつ接尾辞が子音幹に接続する場合は連結子音が、連結母音をもつ接尾辞が母音幹に接続する場合は連結母音がそれぞれ欠落する。「書カセタ」の場合、子音幹「書k」に接尾辞「-(s)ase-」が接続し、連結子音「(s)」が欠落することにより、新たな母音幹「書kase-」を形成し、さらに接尾辞「-(i)ta」の接続では、連結母音「(i)」が欠落する。

3 派生文法に基づく日本語述部生成

本節では、派生文法を用いた日本語述部の生成処理について述べる。なお、本研究では、生成処理の入力として日本語依存構造を想定する。依存文法は語と語の関係をもとに構文を定めるため、語順の制約が比較的緩い日本語の文法として適している。本手法の手続きを以下に示す。

1) 語順の決定

日本語述部を構成する形態素の生起順序を依存文法に従って決定する。

2) 接尾辞の決定

派生文法における語幹と接尾辞の間の接続制約に基づき、接尾辞を決定する。

3) 形態素の連結

語幹と接尾辞を連結することにより述部を作り上げる。

このうち、手続き3)は、2節で述べた連結方法に従う。以下では、まず1)と2)について説明し、次に、本手法に基づく述部生成の例を示す。

3.1 依存文法に基づく語順の決定

依存文法では、文の構造を、支配と従属といった語と語の依存関係としてとらえる。なお、本稿では、文献[2]にならない、支配する語、従属する語をそれぞれ主要語、修飾語と呼ぶ。文の構成要素の生起順序は、修飾語が主要語よりも前に生起するという日本語の語順原則に基づき定める。日本語述部では、一つの形態素に依存する形態素は高々一つであるという性質があるため[2]、語順は一意に定まる。よって、日本語依存構造から形態素の生起順序を容易に決定できる。

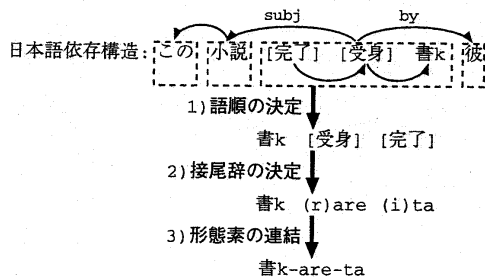


図 1: 述部生成処理の例

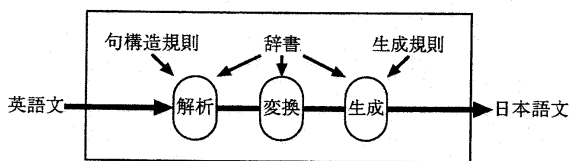


図 2: システム構成

3.2 接続制約を用いた接尾辞の決定

派生文法では、同じ意味をもつ接尾辞が複数存在するが、それらは接続先の違いにより使い分けられる。例えば、完了を表す接尾辞として「-(i)ta」, 「-katta」などがあるが、「-(i)ta」は動詞の語幹に接続し、「-katta」は形容詞の語幹に接続するという制約がある。そこで、この接続制約を用いれば接尾辞を決定できる。本手法では、日本語依存構造において、接尾辞の代わりにその意味表現を用い、生成処理において語順を決定した後で、その意味表現を接続先の語幹に適した接尾辞に置き換える。

3.3 述部生成の例

本手法における述部生成の例を図1に示す。まず、日本語依存構造において修飾語が主要語よりも前に生起するように語順を「書k」, 「[受身]」, 「[完了]」として決定する。次に、接尾辞の選択処理を行う。受身は派生接尾辞「-(r)are-」により表され、これは動詞の語幹に派生する。完了を表す接尾辞は複数存在するが、このうち、動詞の語幹に接続可能な「-(i)ta」を選択する。最後に、子音幹「書k」に接続する「-(r)are-」の連結子音「(r)」を除去し、同様に母音幹「書kare-」に接続する「-(i)ta」の連結母音「(i)」を除去する。これにより、日本語述部「書カレタ」を生成する。

4 英日機械翻訳への応用

前節で示した生成手法に基づく英日機械翻訳システムを作成した。システムは、図2に示す通り、解析、変換、生成の3つのモジュールから構成される。そのうち、解析処理では、入力英語文に対する依存構造を作成する。句構造木において主辞と補語といった語彙上の関係を求めることにより、構文情報のみから依存関係を決定できる[3]。変換処理では、英語依存構造をもとに、語彙変換により日本語依

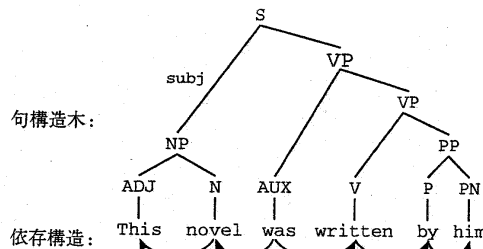


図 3: 句構造木および依存構造の例

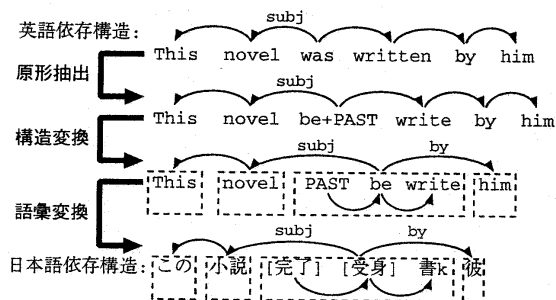


図 4: 日本語構造の例

存構造を作成する。動詞と助動詞については、出現形を原形と変化標識に置き換え、それらの間の依存関係を定める。

英語文「This novel was written by him.」に対する翻訳処理の過程を以下に示す。解析処理により、図3に示す句構造木および依存構造を作成する。変換処理ではまず、助動詞、動詞を原形に置き換える。「was」を原形「be」と過去を表す変化標識「PAST」に、「written」を原形「write」にそれぞれ置き換える。次に「be」と「PAST」の依存関係を定め、日本語文節間の依存関係を決定する。さらに、語彙変換を実施し、図4に示す日本語依存構造を作成する。生成処理では、助詞決定処理、語順決定処理を行い、3節で述べた方法に従って日本語述部を生成する(図1)。

5 おわりに

本稿では、英日機械翻訳の日本語生成処理において、派生文法に基づき述部を生成する手法を提案した。本手法では、活用処理を行う必要がないため、述部を構成する形態素の接続処理が簡潔になる。本稿で提案した手法をもとに機械翻訳システムを作成し、その利用可能性を確認した。今後は、本手法を話し言葉翻訳に応用する予定である。

参考文献

- [1] 清瀬: 日本語文法新論—派生文法序説—, 桜楓社 (1989).
- [2] 児玉: 語順の普遍性, 山口書店 (1987).
- [3] 松原, 渡邊, 外山, 稲垣: 英日話し言葉翻訳のための漸進的文生成手法, 情報処理学会研究報告, NL132-13, pp. 95-100 (1999).
- [4] 野村: 言語処理と機械翻訳, 講談社 (1991).