

FEP 型英文作成支援ツール

— 日英構文変換部 —

5H-2

山端 潔 石原 佳典* 土井 伸一 亀井 真一郎 佐藤 研治 浜田 和彦*

NEC 情報メディア研究所 *NEC 情報システムズ

e-mail: yamabana@hum.cl.nec.co.jp

1 はじめに

FEP 型の英文作成支援ツール [1, 2, 3] の日英構文変換部について述べる。このツールは、かな漢字変換プログラムとアプリケーションの中間に位置し、前者から日本語文字列を受け取って英文に変換し後者に送信することにより、インタラクティブな英文作成を支援する [1]。日英構文変換部は、日本語文字列の形態素解析と辞書引きが終了後に FEP 部から起動され、日本語英語混じり文を受け取って英文に変換し、FEP 部に送信する機能を持つ。

対話的な訳語選択インタフェース [2] により、原則として変換前に指定可能なすべての英訳語が指定されている。また、変換範囲も、原則として用言一つを含む範囲以下に限定されており、これを越える場合変換を多段に行って結果を合成する。本稿では、このようなインタフェースに合致するように実現された構文変換部の構成と特徴について述べる。

第2節では、日英構文変換の基本的な操作手順を示す。第3節では、実現した構文変換部の構成と特徴について説明する。第4節はまとめである。

2 構文変換の手順

構文変換の基本的な手順を説明する [3]。

日本語文字列が入力されると、ツールの変換ウィンドウ上で単語に分割され、辞書引きされる。辞書引きされた語は自動的に英訳語で置き換えられ、日本語英語混じり文として表示される。入力文の下の赤線は、構文変換部に送られる区間を示している。ユーザは、必要に応じて訳語と選択区間を変更してから、構文変換を指示する。すると赤線で示された範囲の単語列が構文変換部に送られて英語表現を返す。

図1に、「私は彼が買った本を読んだ」を「I read a book which he bought」に変換する手順を示す。一度に変換できるのは、用言を一つ含む区間に限定されてい

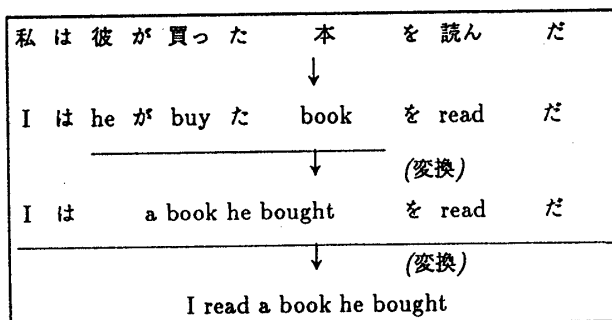


図1：対話的構文変換の手順

るため、入力文のように複数の用言を含む文は、用言を一つ含む範囲ごとに変換を多段に繰り返して全体の訳を得る。

3 日英構文変換部

構文変換部は、日本語解析部、英語生成部、FEP インタフェース部から構成される。

解析部は、補強文脈自由文法に基づくパーシングを行いながら、助動詞列、助詞列、数量表現など部分的な表現のまとめあげと生成を行う。これらの処理が終了すると、名詞句、単文、埋め込み句の区間が順次切り出されて生成部に送られる。生成部は受け取った単語列を英語表現に変換する。最終的に全体が一つの英語にまとまったら、変換結果を FEP 部に送信し処理を終了する。生成結果が複数得られた場合、優先度が上位のものを一定個数まで送信するようになっている。

インタフェースに関連して、構文変換部は以下のような特徴を持つ。

1. 原則として変換前にすべての訳語（品詞を含む）が指定されており、構文変換はこの指定に従う必要がある。
2. 変換範囲は、最大で用言一つを含む区間である。これを越える文は、用言一つを含む区間ごとに変換を繰り返して結果を得る。
3. 多段変換および変換結果のキャンセルをサポートしている。

A FEP-type English Writing Support Tool
 — Japanese to English Conversion Module —
 Kiyoshi YAMABANA, Yoshinori ISHIHARA*, Shinichi DOI,
 Shin-ichiro KAMEI, Kenji SATOH, Kazuhiko HAMADA*
 NEC Corporation, *NEC Informatec Systems

3.1 事前の訳語選択

原則として、変換前にすべての訳語¹がユーザにより指定されており²、構文変換部はこの指示に従わなければならない。ユーザ指定の訳語が必ず使用されることは、指定訳語以外はシステムに辞書ロードしないという単純な方法で保証している。

一方、訳語指定は、解析時の曖昧性解消や生成のプランニングに有効に利用することができる。一例として、「急いで」のように副詞としての用法と動詞としての用法を持つ語に対し、本システムでは英訳語の指定を見て判別を行っている。具体的には、品詞の曖昧性解消フェーズで、訳語が副詞ならば副詞解釈、動詞ならば動詞解釈としている。

訳語指定、特に品詞の選択を誤った場合、有用な変換結果は得られないが、その場合には変換結果をキャンセルして訳語選択・構文変換をやり直してもらう。

3.2 変換範囲の指定

構文変換を保証するのは、単文など、用言一つを含む区間以下に限定されている。また、変換すべき範囲は事前にユーザが指定する³。単文を越える範囲は用言一つを含む区間の変換を多段に繰り返すことで、全体の直訳を得ることができる。

この特徴に対応して、構文変換部は、名詞句、単文、埋め込み単文の3種類の入力のみを処理するように構成されている。さらに、各単語の係り先は、原則として、係受け可能でもっとも近い右側の要素とする。このため解析部は非常に単純である。

それ以外の係受けを指定したい場合、例えば「AのBのC」を「Aの「BのC」」と解釈したい場合は、次に述べる多段変換の機構を用いて、まず「BのC」を変換して一つの名詞句とし、それから「Aの「BのC」」を変換する、という手順による。

3.3 多段変換とキャンセル

前述したように、変換可能な範囲に制約があるため、一文は何度にも分けて多段に構文変換され、変換区間は一般には入れ子になる。

このため、一度変換した結果を外側の制約に従って生成しなおさなければならないことがある。例えば、「私は彼が本を読むのを助ける」という文は、最初に埋め込みの区間「彼が本を読む」を変換し、次に全体を変換するという手順を踏むが、最初の変換で得られるのは“he

¹ 正確には品詞込みで指定される

² システムに選択を任せることもできるが、その場合、単純に辞書中の先頭の単語が採用される

³ システムが推定したものをユーザが確認、変更する

reads a book” という文である。これを「助ける = help someone to do」と組み合わせて “I help him to read a book” を得るためには、最初の変換結果を不定詞句として生成し直す必要がある。また、FEP部は、変換結果を任意の深さまでキャンセルする機構を持っている。多段変換結果を途中までキャンセルして再度構文変換する場合に対応するには、以前の構文変換結果は何度でも再利用できる必要がある。

このような変換結果の再利用と再生成を可能にするために、構文変換部は、生成結果の英語構文木をすべてデータベースに保持している。格納された構文木はFEP部と構文変換部に共通のIDで識別され、仮想的な単語として扱われる。外部の制約にもとづく再生成は、構文木中に保持された生成途上の情報を添えて生成部を呼び出すことで可能になる。

英語構文木は一文の処理の間保持されているが、新しい文が入力されると以前の構文木はすべて破棄されデータベースが初期化される。

3.4 本方式による日英変換の利点

このツールは、対話的な翻訳支援システムの一つと見なすことができる。この観点での最大の特徴は、訳語選択のみならず原文の曖昧性解消や生成スタイル選択をも、訳語選択と変換範囲の指定という二種類のインタフェースだけで行う点にある。この方式は、精度が高いがインタフェースが繁雑であるという、対話型翻訳が持つ課題に対する一つの解答を示すものと考えている。

4 まとめ

試作したFEP型英文作成支援システムにおける日英構文変換部の特徴と構成について述べた。構文変換部は、1) すべての訳語が事前に指定されている、2) 変換範囲は用言を一つ含む範囲に限り、これを越える範囲は多段に変換を積み重ねる、というインターフェースの特徴を最大限に活かすように構成されている。これらは曖昧性解消や生成プランニングの情報を与え、構文変換部の簡素化に有用である。さらに、対話的変換をスムーズに行うために、3) 多段変換や変換のキャンセルに対応した機構を有する、のも特徴である。

参考文献

- [1] 亀井他: 「FEP型英文作成支援ツール — 外国語情報発信の効果的インタフェース —」第51回情処全国大会 5H-3 (1995).
- [2] 土井他: 「FEP型英文作成支援ツール — 訳語選択のユーザインタフェースと辞書記述 —」第51回情処全国大会 5H-1 (1995).
- [3] Kazunori MURAKI, Susumu AKAMINE, Kenji SATOH and Shinichi ANDO. “TWP: How to Assist English Production on Japanese Word Processor.” Proceedings of COLING 94 (1994).