

日本語入力による英文作成支援システム

—長文パターンによる翻訳—

佐藤研治 赤峯享 村木一至
(NEC C&C 情報研究所)

4B-2

1. はじめに

現在、バッチ的に処理を進める機械翻訳システムとは別に、原言語を入力しつつ目的言語を生成するインタラクティブな翻訳方式に対する需要が高まっている。このような翻訳方式としては、ワープロから単に機械翻訳システムが呼び出せるものや、英和・和英等の辞典を呼び出せるだけのものが提案されている [1]。本研究所では、機械翻訳で計算機に任せていた言語の曖昧性をユーザに提示し、選択してもらうことで、より良い翻訳結果を得ようと考え、これまで日本語入力による英文作成支援システムの提案を行ってきた [2, 3]。

しかし、入力文が長くなると、日英の翻訳結果として語順が著しく異なる候補が現れ、これをそのまま提示するとユーザを混乱させてしまうという問題がある。本稿ではこの混乱を避けるため、重文や複文の切れ目として働く語彙に注目し、これによって定義される長文のパターンを用いたインタラクティブな翻訳方式を提案する。このパターンを用いた英語への変換と、重文や複文を構成する単文内の変換は独立に行うことができる。また、これらの候補のユーザへの提示・選択も独立に行うことが可能となる。このため、ユーザは日本語を入力しながら英語の語彙にも、文全体の構成にも注意を払うことが可能となる。

本稿では、この長文のパターンを用いた日本語入力による英文作成支援のインターフェースの提案を行う。

2. 日本語入力による英文作成の課題

日本語を入力しながら英文に変換していく際に、語順変更のタイミングが問題となる。全文が入力されていないうちに、翻訳語順を決定することは困難である。このため、我々が提案してきた英文作成支援システムは、入力しつつある状態では基本的に単語や句のみでしか変換を行わず、文が全て入力されてから語順の変更を行う。しかし、文全体が長くなると、その文全てが入力された後に全体の語順が一度に変化し、ユーザが戸惑うことが考えられる。また、語順に複数の候補がある場合、その文全体を表示しても、ユーザはその違いを把握し難く、選択するのが困難になる。

長文は、複数の単文とそれらの間の関係を示す接続語に分けられる。この接続語を切れ目としての強さ等で分類・整理が行い、この整理された情報を用いて、より正確に自然言語解析を行う手法が提案されている [4]。この手法により長文の切れ目となる接続語の認定を行うと、その語彙の前後の文をそれぞれ単文として解析することが可能となる。また、接続語前後の文の翻訳結果と接続語の訳語を組み合わせることで、文全体の翻訳を行うことが可能と

なる。この長文の切れ目となる接続語とこの接続語によって定義される文の構造パターンを長文パターンと呼ぶ。

長文パターンを用いると、重文や複文を構成する単文内での候補と、文全体の構造となる接続語の訳語と語順の候補の選択を独立にユーザに行わせることができる。ユーザは、訳語に注目しているときには単語や文節の候補を選び、文全体の構造に注目しているときには長文パターンの候補を選択することで、混乱することなく目的とする英文を得ることが可能となる。

3. 日本語入力による英文作成支援

3.1. 長文パターン

ここで長文パターンとこれに対する翻訳のパターンの例を表1に挙げる。

表1: 長文パターン

パターン	翻訳候補
(1) なので (2)	Because (1), (2). (2), because (1). Since (1); (2). (1) and (2). ...
(1) までに (2)	(2) by (1). (2) before (1). ...

表1では長文パターンを定義する接続語「なので」に対し、その接続語が現れる前の文を(1)で、後ろに続く文を(2)で示している。一つのパターンに対する翻訳候補が複数あるが、例えば「Because (1), (2).」では、「なので」の前に現れた文の翻訳結果を(1)に埋め込み、後ろに続く文の翻訳結果を(2)の部分に埋め込めばよいことを表している。(1)のパターンの開始点は、文頭か、またはそれ以前に現れた接続語の直後からとする。

3.2. 長文パターンによる翻訳インターフェース

日本語入力による英文作成支援システムでは、ユーザの文章作成と並行して、自動的に変換を進める自動変換を行う [5]。長文パターンを用いた翻訳では、その候補は入力中の文とは別の部分に表示され、さらにその中の第1候補が強調されて表示される。その第1候補がユーザにとって正しい英文であれば、ユーザの意図と計算機の処理が一致しており、ユーザに新たな選択動作をさせる必要はない。この場合、そのまま続く文を打ち続けることで、その第1候補を暗黙に認めたと捉えることができる。これにより、翻訳候補が表示された場合のユーザの選択肢は以下の3通りになる。

- 第1候補を暗黙で認め、次を入力する。

An English documentation through input in Japanese
— Translation by Long Sentence Patterns —
Kenji Satoh, Susumu Akamine, Kazunori Muraki
NEC Corp.

- 第1候補が誤りなので、候補を選択する。
- 候補で入力に誤りに気づき、訂正する。

ただし、「暗黙の承認」によって選択された候補はユーザが単に見ていなかっただけであるという可能性を含んでいるため、暗黙で認めた候補を基に続く文章の候補が決定される事態が発生した場合には、ユーザに明示的に「暗黙の承認」を確認させる。

以下では、実際の例を用いて、インターフェースの説明をする。今、「今日は水曜日なので、彼は来ないだろう。」という入力を考える。ユーザが「今日は水曜日」まで入力した時点で、画面には「Today は Wednesday」と表示され、続いて「なので」を入力すると、以下のように候補が提示される。

Today は Wednesday なので

Because today is Wednesday, ...
 ..., because today is Wednesday.
 Since today is Wednesday, ...
 . . .

この際にユーザは第1候補が適切かどうか考え、このままで良ければ続きを入力する。そして最後まで入力すると、「Because today is Wednesday, he will not come.」という結果を得る。

また、ユーザが第1候補は適切でないと考えた場合は、例えば第2候補を選択し、以下の状態になる。

Today は Wednesday なので

..., because today is Wednesday.

これに引き続き最後まで入力すると、「He will not come, because today is Wednesday.」を得る。

3.3. 長文パターンの複合

長文パターンにより入力文を区切り、それぞれの部分で独立した変換を行うとき、パターンが1文中に2つ以上現れた場合には注意を要する。例えば「彼が尋ねたので笑いながら私は答えた。」と「彼が笑いながら尋ねたので私は答えた。」という文を考える。この場合、「笑いながら」は、前者では「私」が行った行為であるが、後者では「彼」が行った行為である。このような場合の主題・主格の共有認定法として、接続助詞の切れ目としての強さのレベルに注目した方法が提案されている[6]。本英文作成支援システムも、文の切れ目としての強さを用いて主題・主格の共有認定を行う。

しかし、日本語を入力しながら英文を作成している場合には、ユーザが英語での語順を想定しているために、単に入力された日本語をそのまま訳したのでは正しくない場面がある。例えば、「彼が尋ねたので笑いながら」という入力が行われても、ユーザが「笑いながら」の行為者として「彼」を想定している場合が起こりうる。このような場面が起こることをあらかじめ考えておき、文の切れ目の認定および主題・主格の共有認定はデフォルトの処理とし

て用いる。さらに、文の構造に対する候補として、文の切れ目が通常の日本語文とは異なるような組み合わせも用意することにした。

以下実例を用いて説明を行う。通常は、まず「彼が尋ねたので」まで入力すると、以下のように表示される。

He が ask したので

Because he asked, ...
 ..., because he asked.
 Since he asked, ...
 . . .

さらに続いて「笑いながら」を入力すると、文の切れ目と主題・主格の共有の認定が行われ、「笑いながら」が後ろに続く文に係るような候補が第1候補となる。さらに、第2候補以降に、「笑いながら」の行為者を「彼」とした候補も生成する。以下の例では3, 4番目の候補が「彼」を行為者とした候補に該当する。

He が ask したので laugh しながら

Because he asked, ... while laughing.
 ..., while laughing, because he asked.
 Because he asked while laughing, ...
 ..., because he asked while laughing.

ユーザが英語での語順を想定して、「笑いながら」を「彼が尋ねたので」の後ろに書いた場合、ユーザは第3, 4番目の候補を選択することで目的の表現を得ることが可能である。

4. おわりに

本稿では英文作成支援システム上での、重文や複文を分割する長文パターンを利用した翻訳方式とそのインターフェースを提案した。本方式では、文が長くなれば長くなるほど、長文パターンの候補の表示が多くなり、画面上で見にくくなる可能性が出てくる。このため、候補の画面上でのレイアウトや、候補の表示量によって、本方式の有効性が左右されると考えられる。現在、これらのヒューマンファクタを検討するため、本稿で提案した長文パターンによる翻訳を実際に英文作成支援システム上で実現し、評価を行っている。

参考文献

- [1] 相沢他「双方向機械翻訳システムを利用した外国語ワードプロセッサ」, 情処第44回全国大会 No.3, pp.147-148, 1992.
- [2] 赤峯他「日本語入力による英文作成支援」, 情処第43回全国大会 No.3, pp.205-206, 1991.
- [3] 赤峯他「日本語入力による英文作成支援インターフェース」, 情処第44回全国大会 No.3, pp.261-262, 1992.
- [4] 亀井他「Lexical Discourse Grammarの提案」, 信学会 NLC, 456-7, pp.1-5, 1986.
- [5] 赤峯他「日本語入力による英文作成支援システム」, 情処第46回全国大会 4B-1, 1993.
- [6] 土井他「語彙文脈法 (Lexical Discourse Grammar) とその大域的係り受け決定への応用 (II)」, 信学会 NLC, 91-29, 1991.