

日本語入力による英文作成支援システム

- 仮名漢字変換から仮名英語変換へ -

4B-1

赤峯享 奥村明俊 村木一至 亀井真一郎
(NEC C&C 情報研究所)

1. はじめに

日本人が英文を作成する機会が増え、日本語の発想から英文の作成を支援するシステムが望まれている [1, 2]。筆者らは、日本語からインタラクティブに英文を作成する際に必要な機能やインタフェースについての考察を行ってきた [1, 3]。ユーザが文章の作成に集中できる英文作成システムのインタフェースとしては、1) 語順の違い等の日本語と英語との不对応を吸収できる、2) 変換の操作が少ない(ユーザが変換の操作を意識しない)、3) 既存の文章作成システム(ワープロ)と融合できることが望まれる。

筆者らは、これらの要求を満たすインタフェースを、仮名漢字変換のインタフェースを拡張することで実現可能であると考へ、実際に、インタラクティブな日英変換機能を構造化エディタ上 [6] に構築した。また、この基本機能に、長文のパターンによる翻訳機能 [4]、辞書学習機能等 [5] を加え、英文作成支援システムとして統合した。

本稿では、まず、この英文作成支援システムの機能・全体構成を示す。次に、ユーザとシステムが協力し、インタラクティブな翻訳を行うための日英変換インタフェースについて述べる。一次元の文字列の変換である仮名漢字変換と比較すると、構造の変換を行う必要のある日英変換では、変換を行う範囲の認定が重要な問題となる。エディタの内部に単語・文節・文といった構造、及び、品詞等の属性を保持し、それを利用することで変換範囲の推定を行う。また、ユーザの文章作成のための発想を妨げることなく英語を得るために、自動的に変換のタイミングを計り、自動的に日本語から英語への変換を行う自動変換機能について述べる。

2. 英文作成システムの機能・構成

この英文作成支援システムの開発の目的は、ユーザが文書の作成そのものに思考を集中できる環境を構築することであり、以下のような機能をもつことを特徴としている。

- 仮名漢字変換と同様の日英変換インタフェース
- 翻訳パターンを利用することで長文を分割して変換
- 過去の操作履歴による自動学習機能
- 用例・電子化辞書等の外部データの参照

このシステムの構成を図1に示す。ユーザの入力に対して形態素解析を行い、単語・文節・文等の文の構造を品詞等の属性と合わせて保持する構造化エディタがシステムの基盤となる。日英の翻訳は、まず、日英変換モードにおいて、文の構造を解析して、翻訳を行うべき部分を認識・抽出する。次に、機械翻訳デーモンに抽出した文字列を送り、日本語から英語への翻訳を行う。機械翻訳デーモンは、既存の機械翻訳システムを利用しており、単語、句、文の翻訳が可能である。また、構造化エディタ内に、日本語、及び、変換結果の英語を共に保持することで、元の日本語を利用して翻訳を行うことや、変換結果を取り消して日本語に戻すことが可能となっている。

An English documentation through input in Japanese
- Kana-English conversoin -
Susumu Akamine, Akitoshi Okumura, Kazunori Muraki,
Shin-ichiro Kamei: NEC Corp.

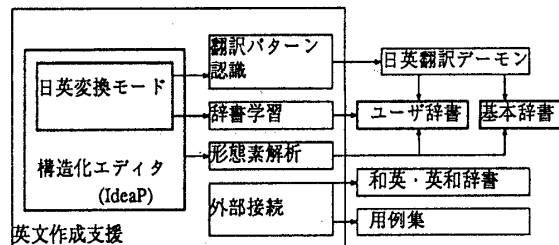


図1: システム構成

3. 日英変換インタフェース

日本語から英文を作成する場合、従来のワープロ等で用いられる仮名漢字変換の機能を拡張することによって、日本語から英語へのインタラクティブな変換を実現する。読みから表記への文字列の変換である仮名漢字変換と比較すると、構造の変換が必要である日英変換では、まず、単語として、次に句として、最後に文としてというように、変換の範囲を変え、同じ部分を何度も変換することが必要となる。従って、変換キーが押された場合、1) 変換の対象となる範囲を決定し、2) その範囲に対してどういう変換を行うかが問題となる。

図2に日英変換のイメージの1例を示す。この例では、「私は論文を」と入力した段階で変換キーを押した場合「論文」を対象として変換し、「書く。」と入力した後で文全体を対象として変換している。

```
私は論文を
↓ 変換 (変換範囲: 「論文」)
私は essay を 候補「essay,paper, ...」
↓ paper を選択、確定
私は paper を書く。
↓ 変換 (変換範囲: 「私は paper を書く」)
I write a paper.
```

図2 日英変換

3.1. 変換範囲

図2の「私は論文を」を入力した段階で変換を行う場合、仮名漢字変換のように「私は論文を」全体を変換範囲とすることも考えられる。しかし、ユーザが現在注目しているのは「論文を」の部分であると考えられる。また、「私は」の部分は自動詞が入力された段階で文全体を変換する場合に、変換の対象になることから、「論文を」の部分を変換範囲とするのが妥当であると思われる。そこで、今回、変換の範囲は注目している文節をヘッドとする句構造を推測して設定することとした。

入力された日本語文は、次の文節の表記が確定された段階で形態素解析が自動的に行われ、単語毎に品詞等の情報が与えられ文節にまとめられている。変換キーが押された時の注目文節、および、それより前にある文節の情報が手掛かりとして、構造の切れ目を推測し、デフォルトの変換範囲の推定を行う。

表1に注目している文節の要素とそれに対する変換の範囲を示す。例えば、「私は / 彼の / 赤い / 本を」という入力に対して、「本を」に注目して変換を行う場合、連体の部分をまとめて、「彼の赤い本」の部分が変換範囲の対象となる。

表1: デフォルトの翻訳範囲

注目文節	変換範囲 [開始位置, 終了位置]
体言 + 助詞	名詞句 [連体接続の先頭, 注目文節上の体言 (助詞を除く)]
用言 (連体以外)	節 [連体以外の用言の直後, 注目文節の用言 (助詞を除く)]
用言 (連体)	名詞句 [助詞「は」の直後または連体以外の用言の直後, 次の文節の体言 (助詞を除く)]
句点を含む	文全体 (文頭, 句点)
その他	注目文節

3.2. 変換方式

変換範囲が決定された場合、その範囲内における変換方式について述べる。変換範囲内での変換方式としては、変換キーが押された時に、1) まず、単語毎に変換を確定していき、順に広い範囲での変換を行っていく方式と、2) いきなり広い範囲で変換を行う方式が考えられる。それぞれの方式の短所を以下に示す。

方式1):

- ・周囲の要素を考慮せずに変換を行ってしまう。
- ・ユーザが確定するものがあまりに多くなる。

方式2):

- ・部分的な変換結果を確認することができない。
- ・翻訳ミスがあった場合、変換結果をキャンセルし、変換範囲を狭めて、変換をやり直さなければならない。

1)の場合、ユーザの負担があまりに多くなり、ユーザの文章作成のための思考を妨げてしまう。また、動詞とその対象(目的語)のように依存関係のあるものに対して、システムの知識を利用して変換することが困難である。そこで、ユーザが明示的に変換キーを押して変換を行う場合、2)の方式を採用した。

3.3. 自動変換

次に、変換キーを押さずに変換を行う自動変換方式について述べる。前節の1),2)の方法の短所を取り除き、「部分的な変換結果の確認を行うことを可能」とし、「他の要素を考慮した翻訳を可能」にするために、変換を行うタイミング、および、範囲は以下のようにする。

- ・文節が確定した段階で、常に変換を行う。
- ・変換は、できるだけ広い範囲(前節のデフォルト範囲)で行う。

また、ユーザ操作からユーザの意志をくみ取ることで、訳し分けを可能とするために、確定方法により確定の強さを以下のようなレベルに分ける。上位のレベルは、下位のレベルの確定を取り消すことを可能とする。ただし、確定を取り消す場合には、ユーザの明示的な判断を必要とする。

- ・レベル1: ユーザが明示的に確定した。(ユーザが確定キーを押す等の操作をした。システムの警告を無視して通常の操作を行った。)

- ・レベル2: システムが他の要素を考慮して優先的に選択し、ユーザが異義を唱えずに通常の入力を行うことで承認した「暗黙の承認」。
- ・レベル3: システムがその要素単独で候補を選択し、ユーザが暗黙に承認した。

日英変換は、レベル1の確定が行われるまで、元の日本語を考慮して翻訳を行う。また、レベル3の確定は、他の要素を考慮した選択がなされることで自動的に取り消される。

この方式によって、「彼はかぜを引いた。」を変換する例を以下に示す。例1では、レベル3の確定である「かぜ→wind」の確定が、「引く」を考慮した選択で取り消される。例2では、レベル1の確定である「かぜ→wind」の確定は取り消されない。例1では、変換のためのキーを全く使わずに、例2では、一度、確定キーを押すことで訳し分けが可能となる。

<例1>

```

he はかぜを
↓
自動変換
he は wind を
↓
入力 (wind をレベル3で確定)
he は wind を引いた。
↓
自動変換 (英文をレベル2で選択)
He caught a cold. (wind → cold の変更を警告)
↓
入力 (レベル1で確定)
He caught a cold. それで

```

<例2>

```

he は wind を
↓
確定操作 (レベル1で確定)
he は wind を
↓
入力
he は wind を引いた。
↓
自動変換
He pulled the wind.

```

4. おわりに

日本人が日本語の発想から英文を作成する場合に、その作成を仮名漢字変換と同様のインタフェースで支援するシステムの設計を行った。現在プロトタイプシステムを構築して、この日英変換方式の評価を行っている。今後は、ユーザインタフェース等の評価・改良を行い、特定の分野(ビジネスライター等)に英文作成の対象を絞り、実用システムとして構築する予定である。

参考文献

- [1] 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
- [2] 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, 1991
- [3] 赤峯他, 「日本語入力による英文作成支援インタフェース」情処第44回全国大会 No.3, pp.261-262, 1992
- [4] 佐藤他, 「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 -」, 情処第45回全国大会 4B-2, 1992
- [5] 安藤他, 「日本語入力による英文作成支援システム - 辞書学習 -」, 情処第45回全国大会 4B-3, 1992
- [6] 野村他, 「言語の構造単位を保持した文章執筆支援システム」, 自然言語処理研究会, 1991