

2E-2

機械翻訳前編集支援ツールの開発

—— 指針と概要 ——

野村浩一, 伊藤悦雄, 木村真理子, 武田公人, 平川秀樹

(株) 東芝 総合研究所

1.はじめに

翻訳前編集作業は機械翻訳システムの効率的な運用に有効であるが、一般ユーザにとって翻訳前編集のノウハウを習得し、正しく適用することは必ずしも容易ではない。このため翻訳前編集のノウハウを計算機側に持たせ、これにより翻訳前編集作業を支援するシステムが望まれており、すでに各種の前編集支援ツールが提案されている[1][2]。

我々は、翻訳業務全体のコストを最小化するには、前編集支援ツールにどのような機能が必要であるかという観点から支援ツールの開発を行っている。本稿では、本支援ツールの開発にあたっての指針、およびシステムの概要について述べる。

2.機械翻訳のための前編集支援

2.1 翻訳前編集の特徴

機械翻訳の前編集においては、一般読者ではなく機械にとって分かりやすい曖昧性のない文書を作成する必要がある。このため機械翻訳のための書き替え基準と一般読者のための書き替え基準(校正支援システムなど)とは異なっており、翻訳前編集作業には以下のような特徴がある。

- (1)ある曖昧性を解消するためにユーザが文を書き替えたときに、ユーザの気付かない新しい曖昧性が発生している可能性がある。
- (2)機械翻訳システムが正しく解釈を行うためにはどのように文を書き替えば良いのかを、ユーザが事前に判断するのは困難である。また、意図に反した書き替えを行う危険性もある。
- (3)ユーザが曖昧性があると判断し文の書き替えを行っても、機械翻訳システムが行う解釈には何ら影響を与えないことがある。

さらに、機械翻訳システムは、人間に比べ曖昧性解消能力が低いため、翻訳前編集の作業量は一般読者を対象とした編集作業量に比べて多くなる。

2.2 翻訳前編集機能の設計指針

前節で述べたような特徴を考慮して、我々は次の3項目を

機械翻訳前編集支援ツールの設計指針とした。

(1) 対話的処理をベースとする。

機械翻訳のための前編集支援は、機械翻訳システムが解決できない曖昧性の解消をユーザの助けを借りて行うものである。ある曖昧性を解消するために新たに曖昧性が発生したり、以前に検出されていた曖昧性が解消されたりするなど、文の書き替え処理はその都度状況が変化する動的処理である。このため、支援ツールは対話的処理をベースとする。

(2) 具体的に書き替え候補を提示する。

支援ツールは曖昧性を検出した場合には、書き替え候補を提示し、ユーザはこれらから適切な候補を選択することで文の修正を行う。これによりユーザの判断も短時間で済み、かつ意図に反した書き替えを避けることができる。

(3) 効果的な書き替えから適用する。

ユーザの操作は最小限にしたい。このため訳文への影響が大きい部分から書き替えを行うこととする。

このためには、機械翻訳の解釈を書き替え操作に反映することが必然的に必要になる。なぜなら、解消すべき曖昧性は機械翻訳システムの能力に依存するからである。現時点では解釈が失敗するという理由でユーザの指示を必要とする文でも、将来的には解析できるようになる可能性がある。

3.前編集支援ツール

3.1 システム構成

前章で述べた指針をふまえた機械翻訳前編集支援ツールを作成した。第1図はそのシステム構成図である。

支援ツールは、文を解析して書き替え候補やガイダンスなど(書き替え情報)を作成する曖昧性検出部、ユーザが実際に文の修正を行うインタフェース部、さらに各部を総合的に制御するコントロール部から構成される。曖昧性検出部は機械翻訳の解析部の情報を用いる情報抽出部と、ユーザ定義が可能な置換辞書を用いる辞書制御部からなる。

情報抽出部は機械翻訳システムの解析処理結果から翻訳前編集支援に必要な形態情報や構文構造を抽出し、これらに対して曖昧性検出用の規則を適用することにより曖昧性の検出を行う。さらに、文の書き替えのために有効なガイダンスや書き替え情報を作成する。

Pre-editing support tool for Japanese-English Machine Translation

Kouichi NOMURA, Etsuo ITOH, Mariko KIMURA, Kimihito TAKEDA, Hideki HIRAKAWA

Toshiba

置換辞書は表層、形態レベルの置換語句を定義した辞書である。ユーザはこの辞書に置換を行う語句を登録することができる。

インタフェース部は、解析部より得られた書き替え情報を用いて、実際に文を書き替えるためのガイダンスを提示する。翻訳結果に大きな影響を与える項目から順にユーザに提示する。

コントロール部は、曖昧性検出部とインタフェース部との間のデータ変換などを行う。

3.2 曖昧性検出部

曖昧性検出部は以下に示す機能を有している。

(1) 曖昧性の検出項目に対して優先度を付与する機能

一文が持つ曖昧性は1種類だけであるとは限らない。これらの曖昧性を一度にユーザに示したのでは、ユーザはどの書換えから行えば良いのかが分からず混乱することがある。

このため、訳文結果に大きな影響を与える曖昧性から解消することとし、影響の大きな順番に優先度を与える。

コントロール部は、この優先度を用いてユーザに提示する順番を決定する。

(2) 機械翻訳システムがデフォルトで行う解釈を示す機能

機械翻訳システムが行う解釈がユーザの望む解釈と一致する場合には、ユーザは原文の変更を行う必要がない。

このため、機械翻訳システムが行う解釈をユーザに示すことにより、ユーザは解釈が一致しない場合にのみ文の書き替えを行うので、修正の手間が軽減できる。

(3) 具体的なガイダンスを作成する機能

機械翻訳システムの利用者は、自然言語の専門家というわけではない。このため、ユーザに対して文法知識を要求すべきではない[3]。文に即したガイダンスをユーザに示すことにより、書き替えの支援を行うべきである。このため解析結果から得られた形態、構文、意味情報を用いて具体的なガイダンスを作成する。

(4) 書き替え候補を生成する機能

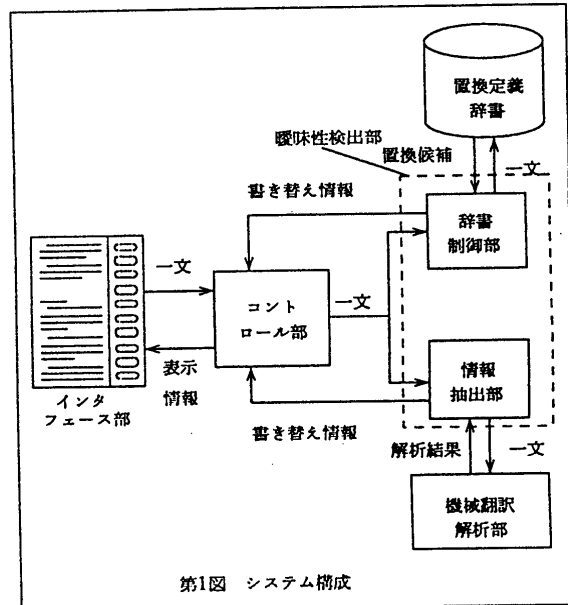
ガイダンスにしたがってユーザが原文を修正するときに、エディタを用いて直接原文を修正するのは非効率である。本機能は機械翻訳解析結果から得られた各種情報から書き替え候補を作成する。この候補をユーザに提示しユーザの選択に従って自動的に原文を修正する。

3.3 インタフェース部

インタフェース部は、以下の方針に基いて作成されている。

(1) マウス操作によりほとんどの作業を行う。

ユーザの操作は、なるべく簡単であることが望ましい。このため、可能な限りマウスのみで操作を行うことにする。た



第1図 システム構成

だし、この場合、マウスで選択する対象がどのような意味を持つのかを正しくユーザに伝えなければならない。

(2) 誤った操作への対応手段を用意する。

ユーザは常に正しく操作を行えるというわけではない。そのため、ユーザが誤った操作を行った場合でも、少なくとも直前の状態に復帰できるようにしておく必要がある。また、変更をすべて無効にしたい場合に備えて、初期状態に復帰できるようにする。

4. おわりに

本稿では、機械翻訳前編集支援システムの指針、概要について述べた。本システムは、現在当社のワークステーション、AS-4000シリーズ上で動作している。今後の課題は、情報抽出部の規則の充実により有効な書き替え候補を作成すること、および無効な書き替え候補の数を減少することである。これにより、さらに前編集の手間を低減し機械翻訳全体の効率向上を計っていく予定である。

参考文献：

[1] 平井 他 「日英機械翻訳用前編集支援システムに関する一考察」(情処学会 第35回全国大会)

[2] 西野 他 「機械翻訳使用のための統合的環境」(情処学会 NL研資 85-13)

[3] 長尾 他 「制限文法に基づく文書作成援助システム」(情処学会 NL研 44-5)