

3L-9

インターネット翻訳における ユーザインターフェース

株式会社 東芝
東京システムセンター
伊藤悦雄・武田公人

1. はじめに

我々は、従来産業翻訳家をターゲットとした機械翻訳システムの開発・改良を行なってきた。産業翻訳をターゲットとした場合は、翻訳結果を完成原稿として利用することが目的であり、原文も推敲が十分なされたものである場合が多かった。このため翻訳システムには後編集処理まで含めた生産性の向上が求められ、例えば翻訳失敗箇所を明示する機能などが必須であった。

しかし、近年機械翻訳を取り巻く状況は大きく変化した。PCの低価格化と高性能化、GUIの改善により、機械翻訳の利用者層が個人レベルまで拡大し、さらにインターネットの普及により一般家庭へ英語が侵入するなど、産業翻訳家以外にも機械翻訳が使用される要因が整ってきた。

個人ユーザの場合には発信情報は「話し言葉」、翻訳結果は意味が取れば良い（文書として完成させる必要がない）といった「読む翻訳」の需要が高い。特にこの傾向は個人間のコミュニケーションにおいて顕著である。

以上の状況に基づき、我々は、通信回線を通じキーボードで英語・日本語間で会話を行なう実験を通じ「読む翻訳」に関する必要な機能を検討した。本実験は、電話回線を使用し端末を直結して行なったが、インターネットを用いれば広ユーザ層による翻訳会話が行なうことが可能である。そこで、上記の検討結果をインターネット情報の翻訳に適用した場合の問題点を検討した。その結果としてインターネット用翻訳ソフトを開発したので報告する。

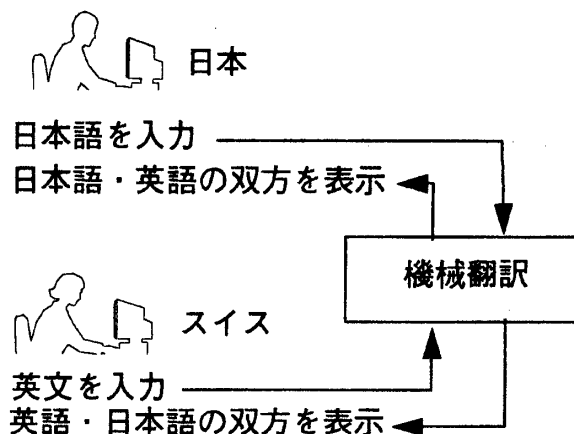
2. 「読む翻訳」の備えるべき要因

産業翻訳以外の分野における機械翻訳の応用として、我々はキーボード翻訳会話の実験を行なっ

た。構成は図1に示すように、キーボード会話の間に機械翻訳を挿入し、オペレータは互いに自国語で会話を行なえるようにしたものである。

この会話において、日本側は英文をある程度理解できるユーザ、スイス側は日本語を全く理解できないユーザと、ユーザ層が二分された。このため、日本側では翻訳結果を読み、理解できない（誤訳と思われる）部分は英文を参照するという使用形態となった。この場合、原文を参照するキーとなるために、誤訳の可能性のある部分をユーザに提示することが重要であった。一方スイス側では原文の参照は一切行なわれず、意味がわからない部分は問い返しが行なわれた。このような場合、結果を鵜呑みにしないためにも誤訳の可能性のある部分をユーザに提示することが重要であった。

以上の点から考慮して、原文を理解できるユー



実験期間：5日間

実験対象：のべ78名（1429発話）

図1 翻訳文字電話実験の概略

ザとできないユーザとは基本的に「読む翻訳」が備えるべき要因は共通であることがわかる。以下その要因を述べる。

2. 1. 翻訳信頼度の提示

自然言語の特性として曖昧性を全く持たない文章を作成することは困難であり、そういった文章を100%正確に翻訳することはさらに困難である。従って原文の曖昧性が高い場合など翻訳結果が保証されない場合には、その部分にユーザの注意を喚起する必要がある。さらにユーザが原文を理解できる場合には原文を参照できる機能が重要である。この場合、原文の表示方法として次の2種が考えられる。

- 1) 一画面中に原文と訳文を交互に表示する
- 2) 原文と訳文を別画面に表示する

前者は原文と訳文の対応箇所が判明しやすいという利点があるが、レイアウトが乱れるなどの弊害が多い。したがって、後者に原文と訳文の対応箇所が容易に判明する機能を付加することが好ましい。

2. 2. 誤翻訳の防止

「読む翻訳」の場合、原文を参照せずに意味の把握を行なうことが多い。このため、翻訳品質が特に重要となる。特に、肯定・否定の解釈の誤りを行なった場合にも文としては成立し、ユーザは疑問を抱かない場合があるので、このような誤りは致命的である。

2. 3 十分な翻訳速度

「読む翻訳」の場合、人間が端末上の文書を読む速度以上の表示速度が要求される。このとき、翻訳結果は同時に出力される必要はなく、順次出力される方が体感速度を向上できる。この場合の最低要求速度を決定するために実験を行なった。これにより、翻訳速度は20,000語/時間以上必要であることが判明した。

2. 4 多様な環境への適合

読む翻訳の場合、翻訳を行なうために新たな環境の整備をユーザには期待できない。そのため、既にユーザが構築している環境に翻訳システムが適合できなくてはならない。

3. インターネット用翻訳ソフト

以上の要因を考慮して、我々はインターネット上でWWW(World Wide Web)情報を翻訳するために適したソフトを試作した。以下、本ソフトの概要

を述べる。

3. 1 原文参照機能

文に複数の解釈のある場合および文の解析に失敗した場合には記号を提示している。ユーザはそれを参考に原文を参照できる。翻訳結果および原文に「翻訳ハイパーリンクタグ」を埋め込み、これをクリックすることによって翻訳結果から原文の対応箇所を呼び出せるようにした。また、翻訳結果には原文中のHTMLタグを復元し、レイアウトイメージ、ハイパーリンクが翻訳結果で保持され、原文と訳文は同じ感覚で参照できる。

3. 2 誤翻訳の防止

翻訳エンジンは産業翻訳用に開発したものをベースとし、WWW情報に頻出する口語表現に対応するために構文解析規則の強化、および語彙辞書の強化を行なった。

3. 3 翻訳速度の向上

辞書内容のキャッシュなどを行なうことにより、80,000語/Hの翻訳速度を実現した。

3. 4 代理サーバ方式

ブラウザを限定することのないように、翻訳プロキシ方式の翻訳方式を採用した。これは、インターネットとブラウザの間に通信制御部(疑似プロキシサーバ)を導入し、ブラウザへ流入するデータを直接翻訳する方式である。

4. おわりに

以上の様に我々は、「読む翻訳」について必要な機能を検討し、インターネット翻訳ソフト ASTRANSAC® for Internetを開発した。現在は、英日翻訳を使用して、日本語しか理解できなくても英語で書かれた情報を収集収集できることを主目的とした検討・開発を行なっているが、情報発信が今後の課題となる。現在情報発信用のツール (HTML文書作成ツール) をWebAssistant™として提供しているが、英文で日本のWEBサイトにアクセスできる翻訳ソフト、およびブラウザ上での翻訳会話用ツールの開発などを行なう予定である。

[参考文献]

旭岡他、「自動翻訳文字電話の翻訳について」第36回情報処理学会全国大会論文集