

## 機械翻訳文の読解に関する評価実験

5H-7

富士 秀、Eric M. Visser、長瀬 友樹†、川久保 賢一‡  
 (株)富士通研究所、†富士通(株)、‡富士通静岡エンジニアリング  
 {fuji||eric}@flab.fujitsu.co.jp

## 1 はじめに

機械翻訳の利用目的は、外国語文書を作成するための、いわゆる「書くための翻訳」と、外国語文書を読解するための、いわゆる「読むための翻訳」に分けられる。本論文では「読むための翻訳」を評価するための小規模実験を行ない、その中からどのような評価が必要となるかを考察した。

またこの評価結果の、読者の視点からの利用および開発者の視点からの利用について述べる。

## 2 従来の評価方法

従来の機械翻訳で中心となっていた「書くための翻訳」では、元々自力で翻訳することができる翻訳専門家のための能率向上ツールという色彩が強かった。このような環境では、1文ずつに翻訳作業を進めてゆくの、評価の方も1文単位の評価が中心となっており、翻訳対象文書としては操作マニュアルなどの限定された分野に関するものが多かった。

全般的には、従来の方法は開発者向けの評価が中心であったので、そこから導き出される評価結果（翻訳率など）は、一般の読者にとって有効な指針とならないことが多かった。

## 3 読むための翻訳に必要な評価

1文単位の処理が中心であった「書くための翻訳」とは違い、「読むための翻訳」では文書全体の読解が中心となる。このことから、訳文の読解に関する評価を行なう際に、文書全体での評価が必要となってくる。

また「読むための翻訳」では、利用者が翻訳者などの一部の人の限られておらず、極めて多様な人が多様な環境のもと多様な目的を持って翻訳を行なっている。そこで、このような多様

An Evaluation Method for Machine Translation  
 Systems Used as Scanning Tools  
 FUJI, Masaru  
 FUJITSU LABORATORIES LTD.

性が読解の評価結果にどのような影響を与えるかを分析する必要がある。

## 3.1 訳文の読解に関する評価

特に、文書全体の読解に関する評価を中心に考察した。

## 3.1.1 訳文の内容伝達度（翻訳精度）

これは、読者が訳文からいかに正確な読解をすることができるかを表した指標である。仮に訳文が文章として読みにくくても、内容が忠実に訳されていれば高い精度が得られる。翻訳精度は次のようにして測定できる。

- 原文と訳文の対応（主題の対応と、文のつながりの対応）
- 文書の内容理解を調べるような読解問題を作成し、読者が訳文を読んで読解問題に答える

## 3.1.2 文の理解容易度（翻訳明瞭度）

これは、読者がいかに苦痛なく文章を理解することができるかを示した指標である。翻訳内容の正確さは低くても自然に読める文章であれば高い明瞭度が得られる。次のような方法を採用した。

- 読解にかかる時間の測定
- 機械翻訳の訳文と、読者が期待する訳文との差の測定

## 3.2 読解の環境

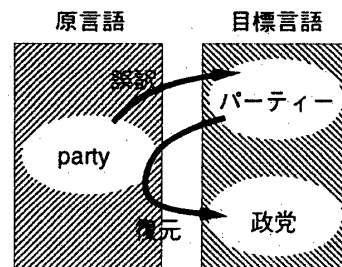


図 1.

### 3.2.1 読者

実験を進めるうちに、同じ訳文に対する評価でも、読解度は普遍的なものではなく読者の知識などによって大きく差が出てくることがわかった。例えば、図1.では、政治分野の記事において“party”という単語が機械翻訳システムによって「パーティー」と誤訳されるが、その訳語を見た読者が英語と政治の知識および文脈から「政党」という正しい訳語に到達する過程を示している。このような「復元」ができるかどうかは、各読者の以下のような能力に依存する。

- 外国語（原言語）の知識
- 翻訳対象文書の内容に関する知識
- 機械翻訳システムに関する知識

### 3.2.2 翻訳の目的

文書にはそれぞれ目的があるが、これを考慮しながら評価を行わなければ実用に即した評価はできない。全ての文書に対して同じ評価を施しても意味がない。

例えば計算機のコマンドのマニュアルの場合、その目的は、計算機の操作者が、必要とする機能を得るためのコマンドを入力できるようになることである。また、例えば新聞記事の場合は、読み手が興味のない記事を読み飛ばしながら所望の記事にたどりつき、必要な情報を入手することである。

## 4 評価結果の利用

以上述べた評価では、各読者の背景における訳文の読解についての評価が得られる。この評価結果を、読者からの視点および開発者からの視点で利用する場合について次に述べる。

### 4.1 読者の視点からの利用

読者の視点からすれば、ある環境において、訳文の読解から得られる精度と明瞭度が、原文の読解から得たものより高いかどうか、機械翻訳が有効であるかの基準となる。読者は、この有効さと翻訳操作の手間などの総合的な評価から、翻訳システムの利用価値を判断する。

手間とは、訳文を得るために必要な諸々の操作をいう。原文の入力が必要な場合には入力の手間、翻訳操作が自動でない場合には翻訳操作の手間、さらには、翻訳処理にかかる時間などもこれに含まれる。

### 4.2 開発者の視点からの利用

以上述べたような評価方法は、機械翻訳システムの翻訳精度向上のための開発に役立つ情報を獲得するために用いることができる。

翻訳精度を上げるためには、例えば、以下のような課題を克服する必要があるが、次にこれらの課題の優先度付けについて説明する。

解析（長文解析など） 照応関係 訳語選択 省略表現
------------------------------------

これら各課題を変数として複数の訳文を用意して、これら複数の訳文に対して読解評価を行なう。このときに、読解に大きな影響を与えた課題を重要課題として開発の優先度をあげる。また、この実験結果を読者や対象文種毎に集計することによって、開発のターゲットを絞ることができる。

例えば、長文解析の影響を調べるには、前編集なしの機械翻訳結果と、長文による誤りを前編集などで解決した翻訳結果の両方を用意し、それぞれについて評価実験を行なう。同様に、訳語選択の影響を調べるには、チューニングなしの翻訳結果と、訳語を完全にチューニングした翻訳結果の両方を用意し、これらを比較する。

## 5 まとめ

「読むための翻訳」のための評価について検討した。訳文の読解に関する評価では、文書全体の評価を導入し、また、翻訳の背景毎の分析の必要性について述べた。最後に、翻訳評価を翻訳精度向上に結び付ける試みについて述べた。

## 参考文献

- [1] Jim Baker et al. Machine translation evaluation methodology. In *MT Evaluation: Basis for Future Directions*, 11 1992.
- [2] Henry S. Thompson. "new directions: Automatic evaluation of translation quality: Outline of methodology and report on pilot experiment". In *MT Evaluation: Basis for Future Directions*, 11 1992.
- [3] John S. White. Approaches to black box evaluation. In *MT Summit V*, 7 1995.