

# 機械翻訳電子メールシステム

西野文人 中村直人

富士通研究所

電子メールを利用した機械翻訳システムについて報告する。いくつかの機械翻訳システムが既に商品化されている。しかし、機械翻訳システムは翻訳を業務としない一般の人にはほとんど利用されていない。これは現在の機械翻訳システムの質が充分でないということもあるが、それ以前に手軽に機械翻訳を使えるような環境がないために使用したことがないという人がほとんどである。多くの人に機械翻訳を利用してもらい、データや意見を収集し研究開発にフィードバックすることを目的として電子メールを利用した機械翻訳システムを作成した。本稿では、電子メールを利用した機械翻訳システムについて、その特徴や構成について述べ、実際の運用から得られた知見を述べる。

## A MACHINE TRANSLATION SYSTEM USING ELECTRIC MAIL

Fumihito NISHINO and Naohito NAKAMURA

FUJITSU LABORATORIES LTD.

1015, Kamikodanaka, Nakahara-ku, Kawasaki 211, Japan

This is a report on the machine translation system that we have developed for use with electronic mail. Some machine translation systems are already commercialized; however, presently few non-professionals use machine translation systems. One reason is that the quality of machine translation nowadays is not adequate. But the main reason is that there are few people who are in an environment where they can use a machine translation system. We developed a system for use with electronic mail so that many people could use it, and so that we could collect data and opinions, and thus receive feedback for research and development. In this paper, we describe the characteristics and the composition of the machine translation system using electronic mail and describe the findings obtained from actual operation.

## 1. なぜ機械翻訳システムが広く利用されないのか

いくつかの機械翻訳システムが既に商品化されている。しかし、機械翻訳システムは翻訳を業務としない一般の人にはほとんど利用されていない。翻訳を業務としない一般の研究者には機械翻訳システムは必要ないのだろうか。いや、一般の研究者であっても外国語で手紙や論文、マニュアルなどを書いたり読んだりすることがよくある。実際に95%の人(ただし富士通社内のUNIXユーザという偏りがある)が機械翻訳のニーズを認めている。それでは、なぜ機械翻訳システムがあまり利用されないのだろうか。それには次の3つの理由が考えられる。

- 1)手元に利用できる機械翻訳システムがない。
- 2)機械翻訳システムの使い勝手が悪い。
- 3)機械翻訳システムの翻訳精度が悪い。

確かに、現在の機械翻訳の能力はまだ充分とは言えず、どんな文章でも正確に翻訳できるというものではない。しかし、うまい利用分野を選べば役に立つはずである。それなのに、機械翻訳を使ってみただけで役に立たなかったと感じた人よりも、機械翻訳システムを本格的に使ってみたことがない、あるいは一度も使ったことがないという人の方が圧倒的に多い(一般よりかなり機械翻訳が使いやすい立場にいる前記の富士通社内の人間でさえ、わずか20%の人しか機械翻訳に触れたことがないという)。これは、まさしく、機械翻訳の能力の問題ではなく、機械翻訳使用のための環境の問題であろう。この利用環境の問題をもう少し細かく分析すると、次のような理由が考えられる。

- 機械翻訳システムが載っている計算機が手元にない。
- 機械翻訳システムと辞書をインストールするための資源(CPUの能力、メモリ容量、ディスク容量)が不足していて、インストールできない。
- 機械翻訳システムを載せようにもインストールに手間がかかる。
- 機械翻訳システムの起動の仕方がわからない。
- 機械翻訳システムが載っている計算機は、普段自分が利用している計算機ではない、あるいは文書形式が異なる。そのために、原文や翻訳結果の文書の転送、あるいは変換が必要である。
- 機械翻訳システムで用意されているテキストエディタの操作を覚えるのが大変だ。
- 機械翻訳した結果に対して何らかの処理を加えるためのツールが揃っていない。
- 機械翻訳の評判が悪いのでわざわざ使う気がしない。

このような問題点を解決するには、誰でもが手軽に機械翻訳を使えるような環境を整えることである。近年、計算機のネットワーク化が進み、UNIXネットワークあるいはパソコン通信により電子メールを利用している人が増大してきた。そこで、電子メールを利用することにより手軽に機械翻訳が利用できる機械翻訳電子メールを作成した。

これまでにも電子メールを使用した機械翻訳サービスは商用化されている。しかし、これは原文と翻訳結果を翻訳依頼者と翻訳会社との間でやりとりするのに電子メールを利用しているというだけで、翻訳依頼文書を機械翻訳システムで処理できるようにするのは人手で行うし、翻訳結果も翻訳家がチェックして修正するというように、完全自動ではない。したがって、利用者からみれば翻訳家に頼んだのと同じ品質の翻訳結果を以前より安く手に入れられるものではあるが、人手がはいるので翻訳に要する時間も日単位であり翻訳コストも安くなったとはいえまだ手軽に使用できるものではない。ここで紹介するのは、一切人手がはいらない完全に自動な機械翻訳電子メールシステムである。

## 2. 機械翻訳電子メールシステムの目的

本システムの目的は以下のとおりである。

### (1) 多くの人に機械翻訳を利用してもらう。

多くの人に機械翻訳を利用してもらうことにより、利用者に現在の機械翻訳のレベルをわかってもらう。また、現在の機械翻訳は、翻訳精度の質は充分とは言い難いが、積極的に利用してもらうことで、機械翻訳のより有効な使い方を模索する。

### (2) 実験の場とし、データを収集する。

新たに作成したシステムの実験・評価の場として使用する。そして利用者のログをとり、それを分析することで利用者の誤りやすい部分の発見や現在の機械翻訳システムの問題点を探る。また、利用者の声を聞くことにより、研究開発にフィードバックさせる。そして、利用してもらった文書を整理することにより、文書データベースを作成する。

## 3. 機械翻訳電子メールシステムの概要

我々は機械翻訳システムATLAS IIをUNIX上に移植・部品化し、そのまわりにいくつかのツール群を作成することにより、電子メールを利用した機械翻訳システムを開発し、富士通社内のUNIXネットワークであるFNET及び商用のパソコン通信であるNIFTY-Serveの上で使用できるようにした。

このシステムにより機械翻訳を利用しようとする者は、翻訳したい文書を特定のアドレスに電子メールで送信するだけで、機械翻訳システムが翻訳した結果を電子メールにより受信できるようになった。この時、どの言語からどの言語へ翻訳するかや、翻訳結果をどのような書式にするかは本来のコメントとしての**Subject(標題)**欄に指定するようにした(図1)。

```
From: nisino@flab.fujitsu.co.jp
To: translate
Subject: je -1 翻訳テスト
          タイトルの認識
          本システムでは文を自動的に切り出します。その時タイトルも認識しますの
          で、上のタイトルが本文とつながって解釈されてしまうことはありません。
```

図1 電子メールによる翻訳依頼

システムはメールを受信すると起動し、そのメールヘッダを解析し、翻訳言語対や出力書式や返信先を調べる。本文は機械翻訳が実行可能なようにコード変換、文書構造解析、文切り出しなどの入力文整形が施され、機械翻訳にかけられる。機械翻訳結果は出力書式指示に従って出力整形され、電子メールとして指定されたアドレスに送り返される(図2)。

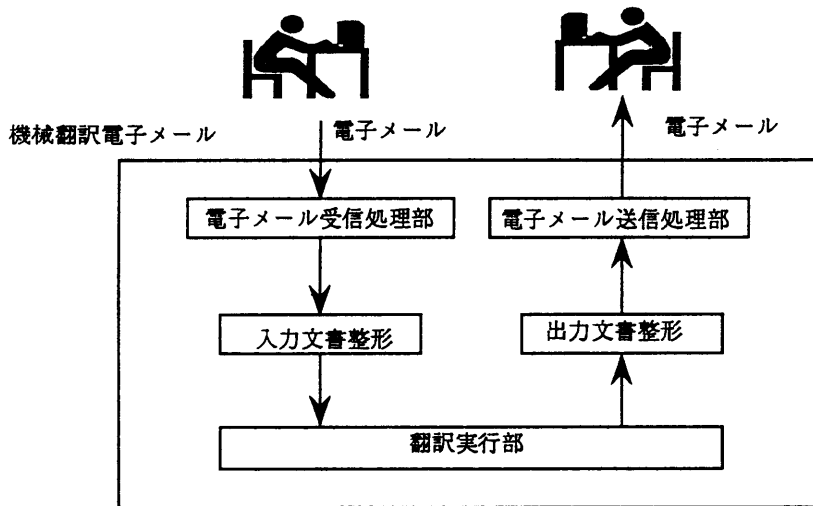


図2 構成

#### 4. 本システムの特長

本システムでは、利用者は翻訳したい文書のある特定の電子メールアドレスに送付するだけで、機械翻訳システムが翻訳した結果を返事として受信できる。これにより、上で指摘した多くの問題点は解決する。以下に本システムの特長をあげる。

##### 1)多くのユーザが利用可能

電子メールの送信先の機械上で機械翻訳を実行するので、自分の普段使用している機械に機械翻訳システムをインストールしなくても、ネットワークに繋がってさえいれば、機械翻訳が利用できる。

##### 2)簡単なユーザインタフェース

機械翻訳を実行するには単に電子メールを出すだけでいいので、電子メールの使い方を知っている人ならば誰でも利用できる。

##### 3)使い慣れたシステム上での作業

元の文書の入力、翻訳結果の編集・印刷などは、利用者の使用している機械の上で行うことになるので、利用者が使い慣れたテキストエディタ、ツールなどを利用することが可能である。

##### 4)最新システムの利用

機械翻訳システムはまだ成長期にあり、辞書・文法は日々改良されている。本システムでは機械翻訳システムはセンター側で一括管理されるので、利用者がインストールやメンテナンスをする必要はないし、常に最新の機械翻訳システムを利用することができる。

#### 5. 多くの人に利用してもらうには

機械翻訳システムを電子メールから利用できるようにすることで、上に示したような様々な問題点が解決されるが、しかし、単にこれだけでは不十分である。このようなシステムで重要なのは、ユーザに対して変な制限をつけないことであり、使う気をおこさせるような付加価値がついていることである。また、電子メールを利用しているということでメールシステム特有の管理機構が必要なものもある。以下では、本システムにおけるこれらの工夫を述べる。

### 5.1. 文書構造解析

ユーザは自分の作成した文書を変換しなければならないとしたらこのシステムを使用してくれないであろう。システムとしては、全角半角まじりの文書の処理やタイトルの認識、左づめ、センタリングなどの整形書式を抽出し、翻訳結果に反映することが大事になる。

この時、UNIXユーザはテキストエディタで作成した文書をそのまま送ってくるものがほとんどであり、そのような文書では大きな問題は生じない。しかし、NIFTY-Serveのようなパソコン通信ユーザの中にはワープロで作成した文書を変換した後に送ってくるものがある。その場合、ワープロでダブルスペース、左端設定のようなフォーマットや拡大文字は、変換後はただの空行あるいは空白としてしか表現されていない(図3)。多くの人に手軽に利用してもらうには、このような文書も解析することが必要である。

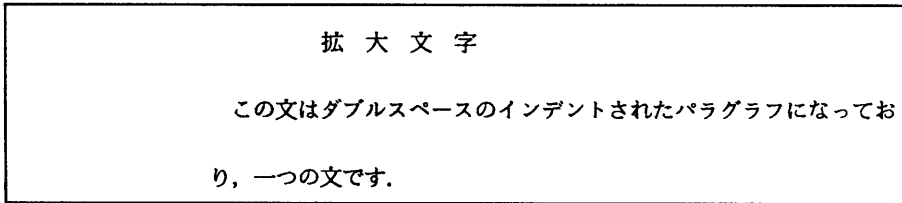


図3 フォーマットされている文書

### 5.2. 様々な出力書式

翻訳結果をどのように使用するかによって、どういう出力書式にしたら良いかということも変わってくる。エディタや印刷をユーザ側にまかせてあり、様々な使われ方をする翻訳メールシステムでは、様々な出力書式を用意することも重要である。我々のシステムでは、原文と翻訳結果の対訳を出力するか、あるいは翻訳結果だけを出力するかをオプションとして指定できるようにした。このとき、原文の文書構造を解析した結果をそのまま清書用コマンドとして翻訳結果中に埋めこんで返信する(図4)か、あるいは清書処理を実行した結果を返すかも選択できるようにしてある。

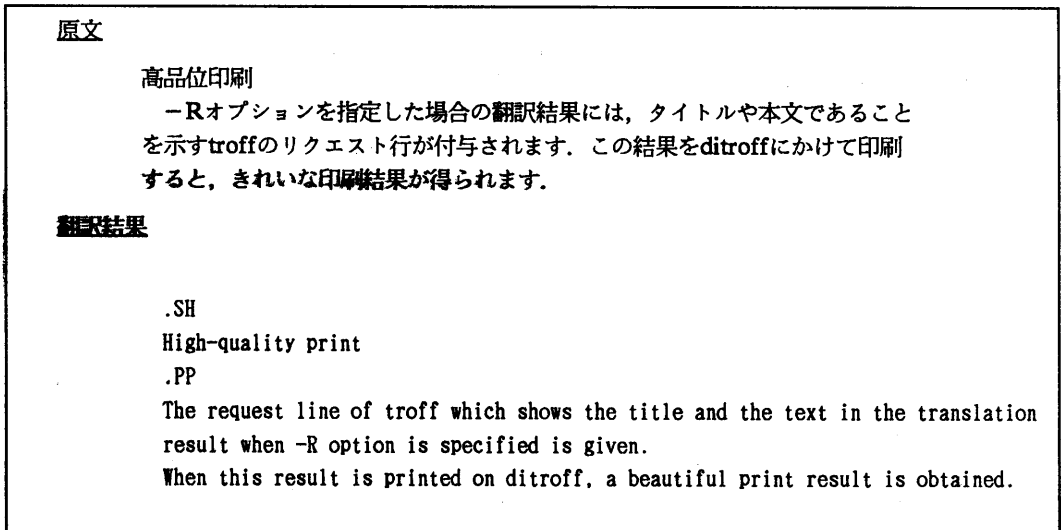


図4 清書コマンドの埋め込み

さらに、あらかじめ清書用コマンド付の文書を送信しそのまま機械翻訳にかけることも可能にした。これにより、文書中のテーブルや図を保存しながら、文書の部分のみを翻訳する(図5)ことも可能になった。

#### 原文印刷結果

troff用に作成された文書は、-RRオプションを指定することで、troffリクエスト行を保存したまま翻訳することができます。

**この行は大きい字で印刷されます。**

また、テーブル内の文も翻訳できます。

オプション	説明
-2	システムは原文と翻訳結果を返します。
-1	システムは翻訳結果のみを返します。
-R	システムはtroffのリクエスト行を付与した翻訳結果を返します。
-RR	システムはtroffの制御コードを保存して翻訳します。

#### 翻訳結果印刷結果

As for the document made for troff, it is possible to translate with troff request line preserved by specifying -RR option.

**This line is printed by a large character.**

Moreover, the sentence in the table can be translated.

Option	Explanation
-2	The system returns the original and the translation result.
-1	The system returns only the translation result.
-R	The system returns the translation result which gives the request line of troff.
-RR	The system translates preserving the control code of troff.

図5 清書コマンド込み翻訳

### 5.3. 私用辞書

機械翻訳品質を向上させるためには専門用語辞書・私用辞書が必要になることがある。専門用語辞書はセンター側で用意しておき、オプションでどの専用辞書を使用するかを指定すればよい。問題は私用辞書である。不特定多数の利用者・マシンを対象とする機械翻訳メールシステムでは、私用辞書を利用者側の機械上で機械翻訳用に保持し、検索することはできない。システムはメール本文中の単語、訳語、品詞を記述したコマンド行を解釈し、翻訳実行時に一時的に辞書登録をして機械翻訳を実行する。頻繁に私用辞書を使用する利用者はセンターに私用辞書使用の登録をし、専用辞書と同様に使用する。

#### 5.4. ジョブ管理

翻訳あるいは文書整形などの実行中にエラーが生じて実行継続が不可能になった場合、利用者からは何もわからない。管理者としても翻訳実行のプロセスを常時監視しているものでなければ、すぐにはなにかが起きていることに気がつかないかもしれない。管理者に自動的に連絡する手段を設けておくことが必要である。また、このようなエラー以外にも、一度に翻訳依頼が多数押し寄せた時やシステムの切り替えなどの際にも、翻訳をすぐ実行して返すことができない時がある。このような時、利用者からすると単に翻訳結果が戻って来ないというのでは利用者は不安である。システムから翻訳結果が遅延することのメッセージを利用者にも送ることが利用者を安心させる。

また、利用者からすると、使用方法が不明だったり、スタッフへの要求などがある場合もある。そのような時のために利用者がスタッフに質問や要求をしたりマニュアルを参照できるようにする機能が必要である。本システムでは **Subject** のコマンドとしてこれらのコマンドを用意した。

#### 6. 機械翻訳電子メール試行実験結果

社内のUNIXネットワークであるFNETおよび商用パソコン通信であるNIFTY-Serveで社内ユーザに限定して日英の機械翻訳電子メールサービスを行った。以下ではこれらの試行実験結果について述べる。

##### 6.1. 利用者

限られた範囲でのアナウンスにもかかわらず、これまでに約250人の人が機械翻訳を利用した。中には機械翻訳を使うために初めて電子メールを利用したという人もいた。今までの機械翻訳システムでは使う環境がなかった人や使う気になれなかった人が、このようなシステムならちょっと利用してみようという気になれるということであろう。

##### 6.2. 文書の種類

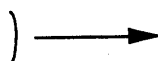
まず、利用者がどのような文書を翻訳したかという点、仕様書、操作説明書、機能説明書といったマニュアル類が約40%を占め、約25%が手紙、約10%が論文であった。これまでの機械翻訳システムは大量のマニュアルの翻訳工数の削減といったことが主目的であった。しかし、一般の利用者としては、手紙、論文といった翻訳の要求が多いことを示している。これらは翻訳というよりも外国語文書作成であり、その途中ではとりあえず英語になっていればよいということが多い。このような文書の翻訳を翻訳家に依頼していたのでは時間もかかるし、コストもかかってしまう。今回のような手軽に使用できる機械翻訳電子メールはこのような応用に有用である。また翻訳の質も、従来はマニュアル翻訳をはじめとする既存の文書の翻訳が中心であり、そのために省略された主語の補いであるとか並列句のあいまい性の解消だとかに苦労していたが、利用者がはじめから機械翻訳にかけるともりで文を作成すれば、このような問題も自ずからある程度解消される。また、利用者は何回かの翻訳実験により、どのような日本語を書けばよいのかも学習してくれる。

##### 6.3. 利用者の評価

本システムを利用した人の評価は以下に示すとおりである。

#### 翻訳品質に対する利用者のコメント

大変役に立った	6%
結構役に立った	49%
あまり役に立たなかった	34%
使いものにならない	11%



実務に利用した	60%
実務に利用しなかった	40%

#### 今後の使用

使いたい	95%
使う気にならない	5%

現在の翻訳品質でも役に立つと言っている者が55%もいる。今回は役に立たなかったという者も、日本語の原文の悪さを認識しており、今後使用したいという者は95%もいる。電子メールという手軽さ・速さにより、ほとんど手間がかからなければ、今の機械翻訳システムでもゼロより良いということであろう。

応答時間に関しては、多くの人が結構速いと感じていたようである。これは、翻訳結果を端末の前でじっと待っていたのでは遅いと感じるものでも、電子メールという形で依頼することによって、他の仕事をしたり、一般の人間相手のメールとの比較から速いと感じるのかもしれない。

#### 7. おわりに

本システムは多くの人に利用してもらうことができた。これにより、利用者の声を聞き、文例の収集など研究開発へのフィードバックができた。システムは多くの人に利用してもらうことによって改良され、使いやすいものとなっていく。機械翻訳システムももっと多くの人に利用してもらうことが必要であろう。そして、コミュニケーションの補助としての使い方など、新しい使い方をさがしていき、機械翻訳の使える場を増やしていくことも大切であろう。

#### 謝辞

UNIXとNIFTY-Serveとの接続及びフォーラム運用にあたりご協力いただいた下浦部長(エヌ・アイ・エフ株式会社)、野々垣主席担当部長、大浦氏、シス技の方々には謝意を表します。

#### 参考文献

- [1] 機械翻訳の開発と実用に関する実態調査, 社団法人 日本電子工業振興協会 (1989)
- [2] FUJITSU S Family Formatting Documents, 53SP-1756E-1 (1988)
- [2] FUJITSU S Family Using NROFF and TROFF, 53SP-1755E-1 (1988)