

機械翻訳システムの応用

Practical Systems Using Machine Translation

山本 秀樹*
Hideki Yamamoto

* 沖電気工業株式会社 研究開発本部 関西総合研究所
Kansai Laboratory, R&D Group, Oki Electric Industry Co., Ltd.

1998 年 1 月 25 日 受理

Keywords: machine translation system, natural language processing.

1. ま え が き

1990 年前半までは数百万円以上のワークステーションやそれ以上の処理能力を持った大型計算機でしか動作しなかった機械翻訳システムが、計算機のハードウェアの処理能力の向上とソフトウェア技術の進歩によってパーソナルコンピュータで稼働するようになった。また、パソコン通信やインターネットの普及に伴って、海外との情報のやりとりが増えるにつれ、機械翻訳に対する一般の人の関心が高まってきている。

機械翻訳システムは、ほとんどの部分が自然言語処理技術によって構成されているという理由で、自然言語処理の典型的なアプリケーションとすることができる。自然言語処理は、言語処理学会が設立され、活発な活動を行なっていることからわかるように、盛んに研究が行なわれている発展途上の技術分野である。その典型的なアプリケーションである機械翻訳システムが出す翻訳結果が、完全でないことは想像に難くない。それにもかかわらず上記の理由で、関心が高まっており、また、いくつかのシステムが実用的に使用されている。

どのようにして不完全な技術が実用化されてきたのであろうか。それは、機械翻訳の利点を生かし欠点を補完するような技術が生み出されてきたためであろう。

本稿では、機械翻訳の様々な実用システムを説明し、実用化の際にどのような技術が補完されていったかについて述べる。さらに今後必要になるであろう技術について述べる。自然言語処理技術自体の展望は、本稿

の範囲外であるが関心のある読者は過去の学会誌に掲載された長尾論文 [長尾 96] を参考にされたい。

2. 電子メール用機械翻訳システム

2・1 電子メール機械翻訳サービスの誕生

電子メールから手軽に機械翻訳を使用できるようにした電子メール機械翻訳サービスは、1990 年前後から機械翻訳メーカーの社内実験、パソコン通信上の有料サービス、大学での実験サービスとして利用されていた [永田 95, 西野 90, 西野 95]。当時すでに機械翻訳システム自体は商用システムが存在したが、翻訳ソフトを実行できる計算機が大型計算機や専用機であったことと翻訳ソフト自体も高価であったことから、利用範囲は大量文書を翻訳するというマニュアル翻訳など用途が非常に限定されていた。

当時は、機械翻訳に対する潜在的需要が徐々に増えていった時でもあった。すなわち、日本語から英語への翻訳という面では、ワープロやパソコンによって作成された電子化された日本語原稿が増えてきたこと、英語から日本語の翻訳という面では、マニュアルなど電子化された英語のファイルがパソコン通信やネットニュースなどから直接入手できるようになってきたことがあげられる。

これらの新しい需要は、翻訳対象が大量でないこと、ファイル形式が様々であることから商用システムの導入はコスト的に適さないものであり、そのような中で、電子メール機械翻訳サービスは誕生した。以下では、90 年より著者たちが運用している Net

PENSÉE^{*1}システムの例を示す。

2・2 翻訳依頼メール

利用者は、翻訳してほしい文書ファイルをプレーンテキストに変換し、翻訳サービスを提供するサーバに翻訳依頼メールとして電子メールで送付する。翻訳サーバはメールを受けると翻訳システムを起動し、翻訳が終了すると翻訳結果を依頼者に電子メールで返送する。誕生当時からそうであるが、ワープロ専用機やワープロソフトのファイル間に完全な互換性は期待できず、だれでも簡単に使えるファイルというとプレーンテキストということになってしまう。

翻訳依頼メールには、原文以外に後述の翻訳オプションやマニュアル送付依頼を含めることができる。

2・3 翻訳オプション

対話的に翻訳結果を修正したりしながら利用する翻訳ソフトのもつ機能の中で、別解の表示など対話的に使用するオプション以外は、電子メール翻訳サーバでも指定できるようになっている。表1に翻訳オプションの例を示す。固有名詞の翻訳指定 (proper)、人称代名詞の訳出の制御 (you)、翻訳結果の文体などの指定 (imp,keitai) などである。人称代名詞の訳出の制御は、人称代名詞を訳出しないほうが読みやすくなるマニュアルの翻訳のときに使用する。

format オプションは翻訳対象の文書構造を指定するものである。L^AT_EX のファイルをきちんと翻訳するにはユーザが定義するマクロなどすべての L^AT_EX コマンドを解釈することが必要になるが、本システムでは論文執筆に使用する程度のコマンドを中心に処理できるようにしている。

format オプションの mail の指定は、翻訳対象が電子メールや NetNews 記事の文書のときに使用する。電子メールや NetNews 記事には、それ以前のメールなどが引用記号付きで引用されていることが多い。引用記号が単語や文の途中にあると、それぞれ形態素解析や構文解析が失敗し翻訳結果が乱れてしまう。また、引用記号を除去するだけでは翻訳結果を見た場合にどの部分が引用されたものでどの部分がそうでないのかわからなくなってしまう。本システムでは mail を指定すると引用記号を除いた部分を翻訳し、翻訳結果にまた引用記号を付加する。

*1 PENSÉE(パンセ)は沖電気工業(株)と大阪ガス(株)とが共同開発した日英・英日機械翻訳システム。PENSÉEは、沖電気工業(株)、大阪ガス(株)および(株)オーヂス総研の登録商標。

*2 「Hideki>」という文字列

```

-----
To: mail-trans@aaa.bbb.oki.co.jp
From: Hideki Yanamoto (hyama@kansai.oki.co.jp)
X-Mailer: Mew version 1.70 on Emacs 19.28.1 / Mule 2.3
Mime-Version: 1.0
-----
VAR:
format mail
END:
EJ:
Hideki> I am a beginner of GA. I am searching for
Hideki> free GA system running on FreeBSD. Please tell
Hideki> me where to get it.

You can get it from ftp://xxxx.yyy.xxx.aaa/pub/GA1.1/.

It is a beta release. If you find any bugs, please
tell me as soon as possible.
END:

```

(1) 翻訳依頼のメール

```

-----
4496D 01/25 11:42J Operator [==== white backup program === Su
4499 M01/25 14:47J Net PENSÉE test output from PENSÉE 【ネットパンセの
-----
出力結果(翻訳文) output from Net-PENSÉE -----Ver 0.51 97.10.13-----

Hideki> 私は、GA(GA)の初心者である。私は、FreeBSD
Hideki> に関して動かす無料GA(GA)システムを探してい
Hideki> る。どうぞ、それを得る場所を私に話して下さ
Hideki> い。

あなたは、ftp://xxxx.yyy.xxx.aaa/pub/GA1.1/. からそれを得ることができる

それは、ベータリリースである。あなたがいかなるバグ
も見つけるならば、どうぞ、可能な限り早く私に話して
下さい。

-----依頼メール your request -----

```

(2) 翻訳結果のメール

図1 翻訳依頼と翻訳結果の例

翻訳依頼と翻訳結果の例を図1に示す。図1(1)は、文書構造情報の指定を含んだ引用記号*2付き電子メールの翻訳依頼である。(2)はその翻訳結果であり、電子メールの引用記号を除いた部分を翻訳している。

商用システムでは、この他のオプションとして、専門用語辞書指定、結果の送付先の指定などがある[ATLAS 94, KOSK 94, PCVAN 94]。

2・4 利用状況

Net PENSÉEは、90年より社内でも実験運用しており、継続して機能評価、改良を行なっている。平均すると約20件/日の翻訳依頼がある。内訳は、英文マニュアル、論文、電子ニュースの記事、手紙などがほとんどである。インターネットが普及した今となっては電子メールを使えない部所を探す方が難しいが、運用開始当時には部門内で電子メールの導入を計画している段階で、電子メールの利点としてNet PENSÉEの利用を挙げた部門もあった。

3. WWW用機械翻訳システム

3.1 開発の背景

World Wide Web(WWW)は、机上の計算機から簡単に世界中の情報にアクセスできることと、WWWサーバに情報をおくことで、潜在的に全世界のインターネットユーザに向けて情報発信が可能であることから急速に世界中に広まった。膨大なWWWの中から必要な情報を見つけるための支援として、情報検索や情報フィルタリングの研究が盛んに行なわれている。これらの研究成果を利用することであるWWWサーバに必要な情報があることがわかったとしても、その情報が自分の母国語で書かれたものでないと理解に時間がかかってしまう。また、情報発信側を考えると、自分の日常使用する言語で記述する場合は迅速に情報発信が可能であるが、他の言語、それも複数の言語への翻訳を行なった後に情報を提供するのには容易ではない。

機械翻訳をWWWサーバの情報検索に使用できるようになると、情報発信、情報入手の時間が短縮され、世界的な規模のネットワークとその上のWWWサーバを人類共通の資産として広く活用することが可能になる。以下では、WWW用機械翻訳システムに要求される機能と、ネットワーク上での機械翻訳の位置による3つの分類について述べ、さらに、実例としてWWW用機械翻訳システム PENSÉE for Internet について述べる [村田 96]。

3.2 WWW用機械翻訳に要求される機能

WWWサーバを検索しているユーザは、画面を眺めてみて、必要な情報がその文書にあるかそれともその文書の中のリンクが指す文書にあるのかを判断しているといえる。もし、その文書に必要な情報がある場合はその文書を丁寧に読むが、そうでない場合は、リンクをたどって次の文書を表示させるか、たどってきたリンクを後戻りし別のリンクをたどるか、あるいは全く別の文書のURLを入力する。ユーザが画面を見る際には、単に文字情報を読むだけでなく、画面にあらわれた図、文字の大きさやフォントが異なる部分、および他の文書へのリンクがある部分などを見ることによって文書を判断していると考えられる。

このような一連の操作の中で使用されるWWW機械翻訳に必要な機能について検討する。ユーザが指定した文書が外国語(例えば英語)の場合は、ユーザから翻訳要求があればそれをユーザの母国語(例えば日本語)に翻訳できることが必要である。その際、原文

の持つ、図、文字の大きさやフォントの違い、および他の文書へのリンクといった文書構造を翻訳結果に忠実に保存することが重要である。

次に、ユーザが翻訳結果のリンクをたどることを考えると、翻訳結果のリンク先は、原文の対応するリンク先をたどったときと同一の文書にいくべきである。さらに、リンク先の文書がもし外国語の場合はそれをまた翻訳する必要がある。ユーザがリンク元の文書を翻訳した場合は、リンク先の文書もまた翻訳すると考えられる。操作を簡単にするには、翻訳結果中のリンク先の文書が、外国語の場合は自動的に翻訳を起動するのが望ましい。

翻訳速度について考えると、ユーザが読みたいと思う文書はできるだけ速く翻訳できることが必要である。WWW用機械翻訳の場合、翻訳対象であるWWWサーバの文書はそう頻繁には変更されないと考えられる。また小規模なオフィスや学校などでWWWサーバを検索することを考えると複数の人が同じ文書の翻訳結果を要求することが予想される。いったん翻訳した結果を複数のユーザで共有できる機能があると翻訳速度に対する要求は満たすことができる。

上記の検討から、WWW用機械翻訳システムには以下の機能が要求される。

〔1〕 タグやレイアウトを保存した翻訳

WWWには、HTMLのタグだけでなく、空白、インデントあるいは改行によって整形された文書も格納されている。タグ付けされた文書には、タグの部分を翻訳せずに、それ以外の部分を翻訳し翻訳結果に原文と同じタグを埋め込む。

同様に、HTMLの<PRE>タグのようにタグ内では、空白、インデントがレイアウトに使用されている文書に対しては、翻訳前後でそれらを保存する。

〔2〕 言語自動判別機能

WWWのデータの記述言語(例えば英語か日本語かなど)を判別する機能である。システムは文書を翻訳すべきと判別すると翻訳を実行する。

〔3〕 事前翻訳機能

ユーザが翻訳操作を行なった後は、システムはできるだけ早く翻訳結果を出力するのが望ましい。そこで、原文を表示すると同時に翻訳処理を起動しておき、ユーザが読む必要があると判断すればすぐに翻訳中の結果を出力するようにする。すなわち、システムに対する翻訳の指定は、通常の機械翻訳のように翻訳処理の起動を意味するのではなく、翻訳結果の表示要求を意味するようにする。

[4] 蓄積翻訳機能

一度翻訳した翻訳結果を蓄積しておき、翻訳の要求があったときは原文が修正されている場合だけ翻訳し、そうでない場合は翻訳しないようにする機能である。

3.3 WWW 用機械翻訳の実装上の分類

WWW サーバとブラウザとが HTTP という標準の通信プロトコルで接続されている中の、どこに機械翻訳を組み込むかで、WWW 用機械翻訳システムは、大きく3つの型に分けられる(図2, [Yamamoto 95])。

WWW サーバ型(図2(1))は、WWW サーバ側に機械翻訳システムをおき、WWW サーバの外部コマンド実行機能(Common Gateway Interchange:CGI)として翻訳を実行する。ユーザは翻訳したいURLを翻訳システムに入力すると翻訳結果が得られる。翻訳結果からリンクの張られているページを翻訳する際に、毎回URLを入力しないですむようにするためには、翻訳システムは翻訳結果文書中のリンクをすべて書き換える必要がある。翻訳指定などの操作ボタンは、WWW サーバ型の翻訳システムからブラウザに送信する文書の先頭に追加することによって実現する。

proxy 型(図2(2))は、WWW サーバとブラウザとの通信路に機械翻訳システムをおくシステムである。WWW サーバの間のすべてのデータはWWW 用機械翻訳システムを中継することになる。すなわち、ブラウザからは翻訳システムが、proxy サーバとして見える。翻訳システムは翻訳対象のページに翻訳ボタンをつける。ユーザは翻訳ボタンを押すだけで翻訳結果が得られる。WWW サーバ型と異なり、翻訳したいURLを明示的に入力する必要は不要である。

ブラウザ型(図2(3))は、ブラウザ側の計算機に機械翻訳システムをおく。ブラウザのメニューの中に機械翻訳の起動メニューができるようにブラウザと機械翻訳は密にリンクされる。

以下では、ブラウザに依存しないことおよび機械翻訳の使用をユーザに意識させないことを特徴とする proxy 型のシステム PENSÉE for Internet について説明する。

3.4 proxy 型システム PENSÉE for Internet

PENSÉE for Internet は、Sun^{*3}ワークステーション上に実装された proxy 型システムである。以下に PENSÉE for Internet を用いた処理の流れを示す(図3)。

(1) ユーザはブラウザを用いて PENSÉE for In-

ternet に対して URL を送る。

- (2) PENSÉE for Internet は、要求されたデータを HTTP プロトコルを用いて WWW サーバから転送する。
- (3) 転送したデータの言語を判別し、必要ならば翻訳機能を実行し翻訳結果を翻訳キャッシュとしてディスクに格納する。

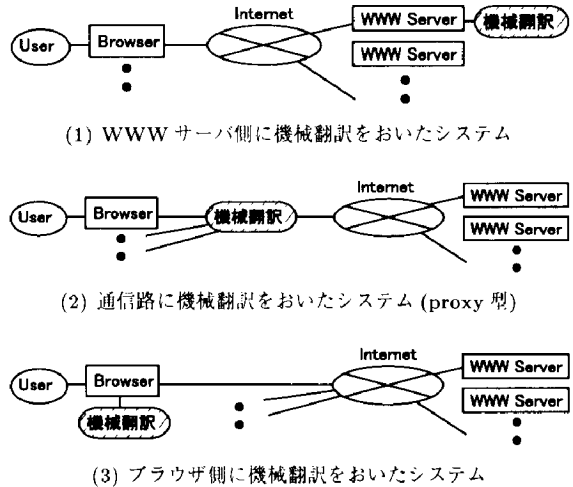


図2 WWW 用機械翻訳:W3-PENSÉE の3つの実現方式

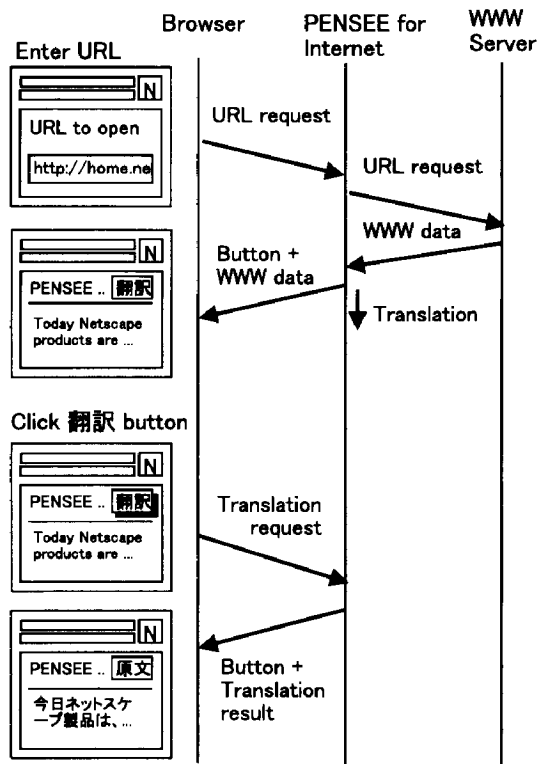


図3 処理の流れ

*3 Sun は、Sun Microsystems, Inc. の登録商標。

- (4) (3)と並行して、指定されたデータに、翻訳結果表示指定のボタンを付与しブラウザに転送する。
- (5) ユーザがブラウザを用いて他の文書を指定した場合は、その文書に対して同様の処理を行なう。また、ユーザが翻訳結果表示指定を指定した場合は、翻訳結果をブラウザに転送する。

図4(1)に英語のWWWサーバのデータ例を、図4(2)にその翻訳結果を示す。(1)の文書の先頭部分は、本システムが付加したヘッダである。「翻訳」は、翻訳結果へのリンクになっている。(2)の文書の「原文」は、原文へのリンクになっている。

本システムでは、ユーザ辞書の登録/検索/削除もブラウザ上からできるようにし操作の統一をはかった。辞書は翻訳結果と同じように翻訳サーバのある場所で管理し、複数ユーザの間で共有できるようになっている。

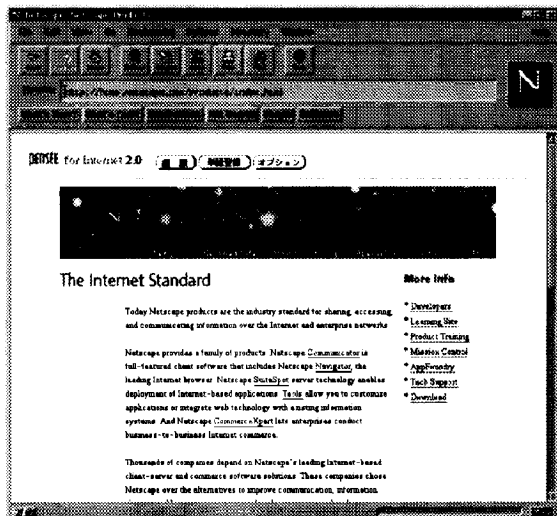
3.5 評価・考察

[1] 事前翻訳と蓄積翻訳について

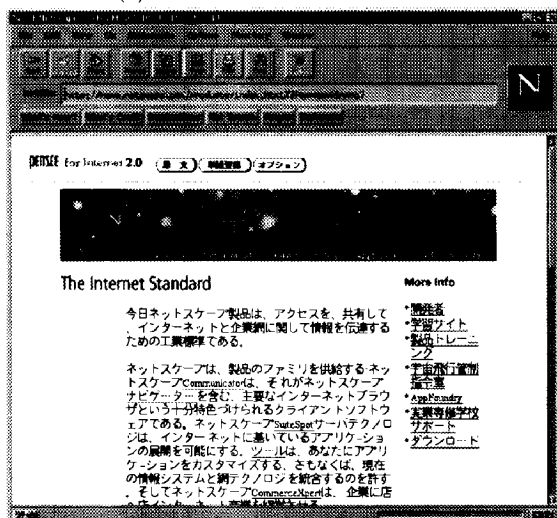
本システムは、実際の翻訳時間よりもユーザが翻訳時間を短く感じるように、事前翻訳と蓄積翻訳を導入した。評価は実際に本システムを稼働させたときのユーザのアクセスログを解析することによって行なった[村田96]。アクセスログには、ユーザがアクセスした(原文、訳文を含む)URLとその時刻、また翻訳の開始・終了時刻とそのURLが記録されている。約1カ月の実験では、翻訳ボタンが押された回数のうち、同じページに対して翻訳ボタンが押された回数の割合は1/3であった。また、事前翻訳したURLの1/2に対し、ブラウザから訳文の要求があった。

ユーザが訳文を要求した際に事前翻訳が完了していないと、システムは途中までの翻訳結果を返す。もしユーザが全ての翻訳結果を必要としているならばRELOADボタンを押せば、その時点の翻訳結果が表示される。RELOADを押さない場合は途中までの翻訳結果で必要な情報を得られたと考えられる。従ってそのユーザに対してシステムはリアルタイム翻訳(ほとんど時間がかからない翻訳)を実現したといえる。アクセスログの評価によると約70%の翻訳要求に対してリアルタイム翻訳を実現した。

ハードウェアの速度向上により、同じアルゴリズムで翻訳する場合は年々翻訳時間は短くなってきている。しかしながら、より詳細に意味情報を扱ったり、文脈情報を扱ったりするようになると、翻訳時間は長くなるだろう。従って、本システムに搭載した事前翻訳と蓄積翻訳は今後も必要になると考えられる。



(1) 英語のWWWサーバのデータ例



(2) 翻訳結果

* a. Netscape, Netscape Navigator および Netscape のロゴは、米[国]およびその他の諸[国]の Netscape Communications Corporation 社の登録商標です。Netscape のロゴ, Netscape の製品名とサービス名は、Netscape Communications Corporation 社の商標です。(一部の国では、登録商標となっています。) Netscape Navigator の著作権は、Netscape Communications 社に帰属します。

図4 WWW用機械翻訳システムの翻訳結果

[2] WWW用機械翻訳の利用形態について

従来の機械翻訳システムは、翻訳システムを使用する人と、その翻訳結果を利用する人が異なる場合が多かったと考えられる。そのため、より翻訳精度をあげるための前編集支援や後編集支援に力が注がれてきた。一方、WWW用機械翻訳は、翻訳システムを起動する人が翻訳結果をその場で理解し検索を進めるといった利用形態をとる。すなわち、これまでの前編集支援や

後編集支援を使用しない新しい利用形態の機械翻訳を開発したといえる。

〔3〕 ブラウザ依存性について

本システムは、HTTP を用いてサーバからデータを受けとり必要ならば翻訳を行ない HTTP でブラウザに翻訳結果を送信する。従って原理的にはブラウザに全く依存しない方式である。

〔4〕 機械翻訳からの方式の独立性について

本システムは、機械翻訳と文単位でインタフェースをとっているためどんな機械翻訳でも接続可能である。すなわち、機械翻訳の方式に独立している。また言語についても日英、英日だけでなく他言語間の翻訳システムも接続可能である。対応する言語の数が増えても、翻訳結果の言語ごとに翻訳キャッシュを持つようすることで方式を変更する必要はない。

4. 改版文書用機械翻訳システム

4.1 開発の背景

大量のマニュアルの翻訳には、機械翻訳を利用することが増えてきている。機械翻訳システムは、高速性、低コスト、訳語の統一がはかれることが利点としてあげられるが、現状の翻訳品質では多くの場合人手による修正が必要となっている。マニュアル翻訳の場合、新製品のマニュアルだけでなく、改良した製品のマニュアル(改版マニュアル)を翻訳するという需要も多い。このような場合、改版マニュアルを最初から機械翻訳し人手で修正するのではなく、人手で新旧マニュアルの差分を見つけ、変更のあった部分のみを新たに翻訳し、その他の部分は旧版の訳文を用いるという方法がとられている。しかしながら文書量が多くなると相違点を見つけ出すのは容易ではない。以下では文書の旧版原文とその翻訳結果を利用して、文書の新版の翻訳を効率良く行なう改版文書用機械翻訳システムについて述べる(図5)[介弘 95a, 介弘 95b]。

4.2 新旧原文対応

改版されたマニュアルは、旧版と全く同じ文、一部が変更された文、および新たに追加された文からなる。

一部が変更された文、すなわち類似文とは、文を構成する語の中で同じ語の割合が多いものである。本システムでは、この割合はユーザがあらかじめ指定する。類似文の探索は、改版時に大幅に段落構成が変更されることもあるので、文書全体を探索する必要がある。探索空間を狭くするために、本システムではファイル間の相違検査法の一つである Heckel による対応付け

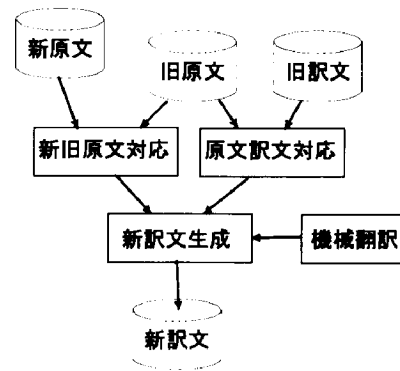


図5 改版文書翻訳システムの構成

[Heckel 78] を拡張した方法を採用した。具体的には、文書内に唯一存在している文同士の対応を基準とし、それに隣接している同一文や類似文は対応していると判断する。

4.3 原文訳文対応

原文訳文対応とは、原文文書と訳文文書との間でどの文がどの文に対応しているかを定めることである。原文の一文が必ずしも訳文の一文になっているとは限らないため、単純に先頭から一文ずつ対応していると決めることはできない。文の対応付けは、翻訳知識の自動獲得[北村 97]などに必要なため、様々な研究が行なわれている[春野 97, 宇津呂 93]。これらの研究ではべた書き文書を対象としているがマニュアルでは通常何らかのフォーマット情報がついている。本システムでは、フォーマット情報として文間の改行やインデントの情報を評価関数に用いた。

4.4 新訳文生成

新旧原文対応処理で得られた情報によって、以下のいずれかの処理を行なう。

〔1〕 「変更なし」の場合

原文と訳文の対応付けが正しく行なわれているかを確認するために旧原文と旧訳文の対応する部分を表示する。ユーザは必要に応じて修正を行なう。旧訳文ファイルから、原文に対応するデータを新訳文ファイルに追加する。

〔2〕 「変更」の場合

旧原文、新原文、旧訳文の対応する部分を画面上に表示し、ユーザに新訳文の入力を促す。ユーザが入力した訳文を新訳文ファイルに追加する。

〔3〕 「追加」の場合

新原文を機械翻訳で翻訳し画面に表示する。必要が

あればユーザが修正する。結果を新訳文ファイルに追加する。

4・5 評価

対訳文書として UNIX マニュアル (英文 589 文, 日本文 615 文) を用いて文対応付けの評価を行なった。システムが示した対応組の中で正しく対応がとれている組を評価したところ文書フォーマットを利用した場合は 99%, しなかった場合は 94% の精度であり, 文書フォーマット情報を利用することの有効性が示せた。

5. 今後の展望

以上, 機械翻訳を応用したシステムについて述べた。本章ではこれまでの経験を踏まえ, 実用システム開発の立場から新たな実用システムを開発する上で重要な研究テーマについて述べる。

5・1 文以外の情報を用いた翻訳品質向上

機械翻訳システムの扱う対象がべた書き文書, マニュアル, WWW, 電子メールといったように時代とともに変化している。翻訳品質を向上させるために, 翻訳対象の性質をうまく利用する研究が必要である。本稿で示した電子メールの引用記号の処理もその一例である。引用記号を正しく識別できないと, 通常の翻訳品質を保持できなくなってしまう。他には, 図表混在の文章において, 図や表から何らかの知識を抽出し翻訳に利用することなどが考えられる。

5・2 機械翻訳の新しい利用場の発掘

WWW で使用されている言語の 82.3% は英語 [Babel 97] であり, 英語がインターネットの共通言語になっているが, その一方で, インターネットは多言語時代に入ってきている [西垣 97]。従って, 今後は世界各地で英語と様々な言語との間の機械翻訳が開発され, 本稿と類似のシステムが実用化されると予想される。

また, インターネットの急速な普及により, WWW のような新たなアプリケーションが今後も次々に誕生していくだろう。その中に, どのように機械翻訳を組み込むかは重要な課題である。この課題に取り組む上では, 機械翻訳はコミュニケーションのツールであり, ユーザに機械翻訳の存在を意識させないで使ってもらうことを考えることが重要であると考え。さらに, ユーザモデルを持ちユーザの嗜好にあわせて翻訳品質・翻訳時間などを変化させる機能も必要になるだろう。

5・3 機械翻訳の部品化

国内では, アジア太平洋翻訳協会を中心に機械翻訳のユーザ辞書の共通フォーマットが策定されている [伊藤 97]。このフォーマットが普及すればユーザはいったん作成したユーザ辞書を複数の翻訳システムで簡単に利用できるようになる。もし, ユーザ辞書だけでなく機械翻訳の翻訳知識や実行モジュールがソフトウェア部品として再利用可能になれば, 別々に開発されたモジュールを組み合わせて機械翻訳システムを構築したり, 機械翻訳モジュールを他のアプリケーションに流用するといったことが可能になるだろう。現在著者たちは, 部品として再利用することを念頭に, Java^{*4} で機械翻訳を開発している [OKI 97]。

6. おわりに

本稿では, 機械翻訳の様々な実用システムを説明し, 実用システムを開発する上で今後必要な技術について述べた。昨今の計算機の処理能力の向上とインターネットの普及により, 機械翻訳の分野で次々と新しい応用システムが生まれた。他の AI の分野でも同じように新しい応用システムが生まれるのではないかと思う。本稿が機械翻訳に携わる方だけでなく, 新しい応用システム研究開発に取り組まれている方に参考になるところがあれば幸いである。

◇ 参考文献 ◇

- [ATLAS 94] パソコン通信・翻訳サービス: ATLAS 機械翻訳サービス及び関連サービス, AAMT Journal, pp.18-25 (1994).
- [Babel 97] Alis Technologies and the Internet Society: Web Languages Hit Parade, <http://babel.alis.com:8080/palmars.html>(1997).
- [春野 97] 春野雅彦: 辞書と統計を用いた対訳アライメント, 情報処理学会論文誌, Vol.38, No.4, pp.719-726(1997)
- [Heckel 78] Heckel, P.: A Technique for Isolating Differences Between Files, CACM, Vol.21, No.4, pp.264-268(1978).
- [伊藤 97] 伊藤悦雄, 村木一幸, 松山 努, 赤羽美樹子, 齊藤由香梨, 平井徳行, 亀井貞一郎: 機械翻訳ユーザ辞書の共通フォーマットの設定 - アジア太平洋機械翻訳協会における活動中間報告 -, 言語処理学会第 3 回年次大会発表論文集, pp.19-22(1997)
- [介弘 95a] 介弘達哉, 下畑さより, 松下久明: 差分翻訳システムにおける対訳文書の文の対応付け, 1995 年電子情報通信学会総合大会講演論文集, pp.122(1995).
- [介弘 95b] 介弘達哉, 下畑さより, 松下久明: 改訂文書の翻訳を支援する差分翻訳システム, 情報処理学会第 51 回全国大会講

*4 Java およびすべての Java 関連の商標およびロゴは, 米国およびその他の国における米国 Sun Microsystems, Inc. の商標または登録商標です。

表1 電子メール機械翻訳サービスの翻訳オプションの例

オプション名	値 [既定値]	意味
proper	j / e / [je] / ej	英日の固有名詞の翻訳形式を指定します j: 日本語のみ, e: 英語のみ je: 日本語(英語), ej: 英語(日本語)
you	[on] / off	off のとき you, your を訳しません
imp	on / [off]	on のとき日英翻訳で命令調に翻訳します
keитай	on / [off]	on のとき英日翻訳で敬体調に翻訳します
format	[plain]	空白, タブ, 改行で整形されたファイルを翻訳します
	latex	LaTeX のファイルを翻訳します
	sgml	SGML のファイルを翻訳します (注2)
	mail	電子メールを引用記号を残した状態で翻訳します (注3)
	html	WWW のファイルを記述している HTML(Hypertext Markup Language) のファイルを翻訳します (注2)
	rtf	Rich Text Format のファイルを翻訳します (注2)
	raw	余分な空白や改行は無視。文書フォーマット用タグを通常の単語と認識して翻訳します
para	on / [off]	on のとき原文と訳文を一文ずつ対訳出力します
fill	[on] / off	para on の時の対訳結果と format raw の時の翻訳結果に対して, on のとき翻訳結果に見やすいように改行を入れます
udic	on / [off]	on のとき英日の原文から2回以上出現した2語以上の名詞句をユーザ辞書の形式で出力します
status	on / [off]	on のとき現在の翻訳変数の値を表示します

(注1) このユーザ辞書形式の訳語を自分好みに編集し次の翻訳から使用してみてください。

(注2) 文書構造のタグの中には, システムが正しく処理できないものがいくつかあります。翻訳後の後修正をお願いします。

(注3) 大抵の引用記号には対応しています。もし対応していない引用記号がありましたらご連絡くださるようお願いいたします。次のバージョンアップで考慮します。

演論文集, pp.3-83 - 84(1995).

[北村 97] 北村美穂子, 松本裕治: 対訳コーパスを利用した対訳表現の自動抽出, 情報処理学会論文誌, Vol.38, No.4, pp.727-736(1997)

[KOSK 94] <http://www.okisoft.co.jp/OSG/KOSK/pnsemail.htm>.

[村田 96] 村田稔樹, 山本秀樹, 永田淳次: WWW のリアルタイム翻訳を可能とする機械翻訳システム, 電子情報通信学会論文誌, Vol.J79-B-1, No.5, pp.364-372(1996).

[長尾 96] 長尾 真: 新しい機械翻訳のための自然言語処理, 人工知能学会誌, Vol.11, No.4, pp.500-506(1996).

[永田 95] 永田淳次, 山本秀樹: ユーザフレンドリな機械翻訳システム: PENSÉE, 沖電気研究開発, Vol.62, No.2, pp.23-26(1995).

[西垣 97] 西垣 通: 多言語時代を迎えたインターネット, 世界, 1997年10月号, pp.136-144(1997).

[西野 90] 西野文人, 中村直人: 機械翻訳電子メールシステム, 情報処理学会 自然言語処理研究会, 75-5(1990).

[西野 95] 西野文人, 中村直人: ネットワーク時代の機械翻訳, 自然言語処理の応用に関するシンポジウム, pp.89-96(1995).

[OKI 97] 100% Java で記述された機械翻訳システム, <http://www.oki.co.jp/OKI/RDG/JIS/java/pensee/index.html> (1997).

[PCVAN 94] パソコン通信による翻訳サービス: PC-VAN・IBS 機械翻訳サービス, AAMT Journal pp.26-29(1994).

[宇津呂 93] 宇津呂武仁他: 対訳辞書を用いた対訳文対応および未知語の推定, 自然言語処理における実働シンポジウム論文集, 電子情報通信学会, pp.140-143(1993)

[Yamamoto 95] Yamamoto, H., Murata, T., and Nagata, J.: W3-PENSÉE: WWW machine translation system that supports the comfortable Internet surfing, Proc.

of International Symp. on Digital Libraries, pp.159-166(1995).

著者紹介



山本 秀樹(正会員)

1984年京都大学工学部電気工学科卒業。同年沖電気工業(株)入社。機械翻訳システム, 知的CAI, エキスパートシステム, データベースマシンの研究開発に従事。知識処理システムの実用化に興味を持つ。1993年情報処理学会研究賞受賞。1995年東京大学博士(工学)。情報処理学会, 電子情報通信学会, 日本ソフトウェア科学会, 言語処理学会, ACM, AAAI, AIED, AACE 各会員。

<hyama@kansai.oki.co.jp>