AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.7
June 1994

アジア太平洋機械翻訳協会

機械翻訳セミナー

「機械翻訳の新たなる潮流」

日　　時　平成6年6月22日(水) A.M.9.30～P.M.4.00
場　　所　芝公園/機械振興会館(B2大ホール)
（東京都港区芝公園3－5－8）(会場連絡先　Tel.03-3434-8216/8217)

テ ー マ　1 パソコン翻訳ソフトのユーザ利用実態調査報告　（利用技術研究会報告）
2 翻訳市場、機械翻訳市場の将来需要予測　（需要予測研究会報告）
3 ユーザ例文、ネーティブ例文の評価　（例文評価研究会報告）
4 ユーザ主導 Example-Based MT の開発　（凸版印刷㈱）
5 ユーザ利用事例発表
①情報伝達の正確さ、前編集のノウハウ　（長瀬産業㈱）
②翻訳品質の向上、周辺翻訳ソフトの開発　（㈱データプロセス）
③パソコン翻訳ソフトで適用領域を拡大
6 講演「国際化の中で、翻訳需要は今……」　（日経国際ニュース）

申し込み期限　6月15日（FAX 03-3664-1352 へ申し込み下さい）
参 加 料　会員 @7,000円（94年度研究報告書込み）
主催・申込先　アジア太平洋機械翻訳協会 事務局

目　　次

・エッセイ　「道具としての機械翻訳」 …… 1
・技術開発　「ユーザ主導のＥＢＭＴの開発 …… 2
・技術動向　「ヨーロッパの現在のＭＴ活動」 …… 10
・カントリーレポート「BehaviorTran(台湾)」 …… 14
・技術早分かり　「統計手法を利用した機械翻訳」 …… 36
・翻訳の現場から「アンカーマンの目」 …… 24
・パソコン通信・翻訳サービスの利用法 ①ATLAS及び関連サービス …… 18
②PC-VAN及び関連サービス …… 26
・新製品の紹介　①ブラビスJ/E　（エムティラボ㈱） …… 30
②ＰＥＮＳＥＥ　for Windows　（沖電気/沖ソフト/オージス総研） …… 32
③E/J J/E ＡＳＴＲＡＮＳＡＣ　（㈱東　芝） …… 34
・通常総会開催案内(表3) 協会活動状況 (40頁) TCシンポジウム(17頁) アジアのイベント(29頁)

エッセイ

道具としての機械翻訳

松下電器産業株式会社　取締役開発本部長　長澤雅浩

商品の研究開発に携わる者として、常に、心に留め置いていることは、「お客様の声を聴く」ということです。

しかし、お客様は、商品が使われる現場のことは熟知されていても、どのような技術があり、その技術を生活や仕事の改善にどのように役立てることができるのかは技術の専門家ではありませんのでご存じではありません。

このため，お客様のニーズの多くは潜在化し、声や文字として表現されることは滅多にありません。

そこで我々は「お客様の声を聴く」ために実際に現場へ出向き、潜んだニーズを掘り起こすことを心がけているわけです。

計算機利用技術の一つである機械翻訳の研究開発においても事情は同じだと思います。

技術者は、複数の言葉を読み書きしなければならないお客様の現場の実態を充分に把握した上で、その作業を改善するための画期的な道具を提案して行くべきです。

機械翻訳の研究開発を考える前に、計算機利用技術の歴史に見え隠れする２つの研究開発アプローチを整理してみますと、一つは、計算機を、人間の作業を手助けする道具として活用するアプローチです。例えば、エディタや帳票計算プログラムなどはその代表例です。

もう一つは、計算機に人間の思考能力を代行させるアプローチで、計算機はユーザの思考能力そのものを補うように設計されています。ただ残念なことに人間の思考能力を実現しているアルゴリズムはまだ解明されておりませんので、限られた範囲で人間の思考を真似るような利用が一般的です。例えばエキスパートシステムがそうです。

この２つのアプローチは機械翻訳システムにも当てはまるようです。

機械翻訳システムをあくまで翻訳作業用の道具とするアプローチでは、計算機が手助けできる潜在ニーズは意外と多くお客様の周辺から見つかるのではないでしょうか。

一方、人間なみの翻訳能力を持つ機械翻訳システムについては、適用先を限定した利用を続けながら、アルゴリズムの研究を含めた息の長い明日への取組みを続ける必要があると思います。家電製品のように幅広く使われる商品として花開くにはまだ相当の時間がかかりそうです。

以上のことから考えますと、機械翻訳に取り組む技術者は、今日のニーズである、「道具としての機械翻訳」はどうあるべきかをしっかりと「お客様に聴き、潜在ニーズをつかむ」ことを心がけるべきだと考えます。

機械翻訳協会は、技術を知悉した技術者と翻訳現場の方々が一堂に会して，語り合える極めて有用な組織だと思います。活発な議論や情報交換を通じて、言語や文化の壁を乗り越えた今までには無かった道具類のアイデアが次々と生み出されるのではないでしょうか。

今後、通信技術などの発展に伴い、海外との複数言語によるやりとりは更に頻繁になることは間違いありません。

当社の社是「世界文化の進展に寄与せんことを期す」にも正に合致する翻訳技術が，世界中の対話を支える文化として近い将来認知されることを期待してやみません。

ヒアリング

ユーザ主導の機械翻訳

凸版印刷　生産技術研究本部　ＤＥ部　仁井正治

《現行機械翻訳システムとユーザの立場》

私どもはあるメーカの英日機械翻訳システムを使っておりますが、どうも使い勝手が悪く扱い難い、ということを感じております。その原因を考えると共に、ユーザとして対応できる方法はないのかと、色々考えた結果、ユーザの立場に立った機械翻訳システムを、私ども自ら研究開発するのが一番良い、との結論に達しました。それによって研究開発しましたのが、これからその概要を説明いたします「ユーザ主導型機械翻訳システムＴＯＰＴＲＡＮ」です。

市販の機械翻訳システムを使用していて、非常に問題になったことは、システムがメーカの論理に基づいて設計されており、翻訳能力の向上は、システムに内蔵された文法規則や構文規則などの規則に依存していて、私どもユーザは何もできないということです。翻訳能力の向上はメーカの研究開発者でないと行えないので、私どもはユーザの立場から、メーカに種々の改善要求を出します。メーカの方では、その都度規則を増やして対応してくれるのですが、時には規則同志が干渉しあって、要求したものは良くなるが、それまで良かった他のものが全く駄目になる、ということが起きて、次第に対応してくれなくなります。

もう一つは訳文を生成する時に、不自然な言い回しになることが多いことです。内容さえ分かれば良いというのであれば、それでも良いのですが、私ども印刷会社のような所では、翻訳したものを印刷物にしなければならないので、自然な表現でないと非常に困るのです。

それから前編集、後編集の場合に、品詞指定、書き換え、分割、訳語選択など、機械の都合に合わせた操作を行わなければなりません。余計な事をやらねばならない、と第一線の翻訳者はこれを嫌うのです。機械の都合のために余計なことをしなければならないのなら、機械を使わないで自分で翻訳した方が早い、と言うのです。前編集に要する労力に対して、得る結果は甚だ小さく、思ったようには良くならないというのが普通です。勿論良い結果を得るケースもありますが。このようなことから、第一線の翻訳者は、機械翻訳システムを使いたがらないのが現状です。後編集の結果は、次の翻訳に反映されません。最近はメーカで改良が加えられてきて、後編集の結果を生かすことのできるシステムが開発されてはおりますが…。

このようなことから、私どもは、メーカ主導型でない、ユーザの立場に立ったユーザ主導型の英日・日英機械翻訳システムＴＯＰＴＲＡＮを、研究開発したのです。

《ユーザ主導の機械翻訳》

それでは、ユーザ主導型として、どのようなシステムが求められているのか、第一線の翻訳者の意見をまとめてみますと、次のようになります。

(1) 仮に完全翻訳が不可能でも、下訳として使用可能な翻訳品質であること。－－既存の機械翻訳システムは、完全な翻訳を目指すためにシステムのアルゴリズムが複雑になって、かえって不自然な訳文を生成しています。実際の翻訳作業では、人間がやる場合でも、どしどし赤を入れて行くもので、必ずネイティブチェッカーがいて、訳文をチェックします。完全でなくても、下訳として使えるような自然な表現が多い場合には、翻訳が少々不十分であっても構わない、というのが第一線の翻訳者の意見なのです。

(2) 機械翻訳システムの内部を知らない普通の翻訳者が、システムの翻訳能力を向上させられること。－－市販のシステムは、文法規則に基づいて翻訳を行うので、構文的に綺麗な原文は、綺麗に翻訳できるようになっていますが、実際の翻訳では、構文的に綺麗な原文は少ないのが現実です。日本人が英語で書く論

文などは、文法的にきちんとした文章なので、綺麗な訳が出ることはありますが、通常の私どもが翻訳しなければならない英語では、先ずこれはありません。

(3) 一度後編集を行うと、それを学習して、次からは、同一文に関しては同一の訳文が出力できることも可能なこと。－－最近、市販の機械翻訳システムでは、対応されつつありますが、まだまだ不十分といえます。

(4) 前編集等、人手翻訳で行わない作業は行わなくて済むこと。

(5) システム操作性は、ワープロのレベルであること。－－現在、ほとんどの翻訳者は、ワープロを使って訳文を作成しています。勿論、お年寄りの大学教授で、ペンを使って原稿用紙に書かれる方もいらっしゃいますが、そのような方は、ごく僅かなのが実状です。ワープロは必需品として利用されていますので、システムの操作がワープロ並であれば、翻訳者にすんなりと受け入れて貰えるのです。

《ＥＢＭＴ方式》

ユーザ主導型の機械翻訳システムは、どのような機能を持たなければならないかが分かりましたが、どのようにすればシステムが実現できるのか、ということになります。私どもは、その方式として、事例ベースの機械翻訳方式（Example-Based Machine Translation：ＥＢＭＴ）に着目しました。ＥＢＭＴという言葉は、最近話題になっておりますので、既にご承知だとは思いますが、後の説明の関係上、ＥＢＭＴの特徴を、おさらいの意味で、説明します。

(1) ある文の翻訳を行う時、それとよく似た文の翻訳例を見つけ、それを模倣することによって翻訳を行う。

(2) 新しい翻訳例を追加していくだけで、翻訳能力を向上させることができる。

(3) 人間の訳した翻訳例を真似ることによって、こなれた訳を出力できる。

《ＴＯＰＴＲＡＮの目指す方向》

先程、私どもが、ユーザ主導型機械翻訳システムをどのように考えれば良いのか、第一線の翻訳者の意見をまとめて説明しましたが、実はそれらの意見は、調査研究の結果、このＥＢＭＴの考え方につながるものであることが分かりました。そこで私どもは、ＥＢＭＴ方式の機械翻訳システムを、研究開発することにしました。私どものシステムは、Toppan Translation Database System を略称して TOPTRAN と名付けております。Database という言葉が加わっているのは、ただ単なる翻訳システムではなく、データベース志向の翻訳システムである、という位置付けだからです。その目指す方向は、これはあくまで理想であり、いまだ完成されたものではありませんが、次のようになります。

(1) 翻訳能力を向上させるには、対訳データを学習させれば良い。

(2) 翻訳作業時、前編集は行わなくても良い。
－－勿論、制限言語などは不要です。

(3) 後編集した結果は、つぎの翻訳時に反映される。

(4) 翻訳品質は、第一線の翻訳者の下訳として使えるものであれば良く、必ずしも完全翻訳を目指さない。－－いま、私どもを初めとして、国内外で「事例ベース」ということで色々な研究開発が行われていますが、私ども以外は、どの発表を聞きましても、完全翻訳を目指して研究開発されているように思えます。これだけ多くの研究開発が行われているのに、何が原因で、プロトタイプのシステムでさえ、未だに開発されないのか、その原因がここにあると思っております。私どもは、必ずしも完全翻訳を望んではおりません。勿論、完全翻訳で出力される場合もありますが、うまく行かない場合もあります。私どもは、翻訳者が下訳として使えれば良い、というシステムを前提にしています。

(5) システムの操作性は、ワープロのレベルであれば良い。

(6) 英日、日英の双方向の翻訳を可能とする。

《ＴＯＰＴＲＡＮの特徴》

以上が目指す方向ですが、次に、私どもが今までに具体化した、ＴＯＰＴＲＡＮの機能について説明します。

■対訳登録

対訳登録とは、英文、邦文のペアをデータベースに登録することをいいます。勿論、対訳登録するのは、文でなくて、単語でも構いません。対訳登録しますと、英語または日本語の翻訳原文で、対訳登録されている文と全く同じ文が来れば、それを使ってそのまま翻訳できる状態になります。

対訳登録では、オペレータは英文と邦文のペアを用意して、システムに登録指示を出すだけで、対訳データベースに対訳データが登録されます。この時のデータベース・キーは、英語と日本語の形態素解析結果と文字列、品詞情報その他、及びパーツ解析（構文解析ではない）という概念に基づいた、解析結果を使用して作成され、自動的に付加されます。「図１　同一ファイル内の対訳登録例」は、一つのファイルの中に英語と日本語のペアがある場合で、これは英語、日本語の順になっていても良く、またその逆に日本語、英語の順になっていても良いのです。例えば、ファイルの中に Selectively plated with …と「6.3.1の項の計測方法により…」がペアで入っているとします。システムがこれを読み込みますと、英語と日本語の形態素解析を行って、未登録語のチェックを行い、対訳と共に、その結果を画面に表示してきます。対訳に、形態素解析の辞書にない単語があれば、「英語未登録語：」、または、「日本語未登録語：」の次の欄に、その単語が表示されます。それに対して、マルチ・ウインドウのシステムであれば、他のウインドウから、辞書ユーティリティを使用して、未登録語を辞書登録すると、これが消えます。消えた時点で、「コマンド＞」の所に「ｄ」を入力すると、対訳データが、自動的にデータベースに登録されます。オペレータのやることは、これだけです。

マルチ・ウインドウでないシステムの場合は、英語を登録するコマンドを入力すると、対訳登録ウインドウが変わって、英語の辞書登録ウインドウになり、日本語を登録するコマンドを入力すると、日本語の辞書登録ウインドウになります。未登録語の辞書登録が終わると、もとの対訳登録ウインドウに戻り、データベースに登録することができます。

```
TOPTRAN訳文登録&学習
h ヘルプ
--------------------------------------------------------------------------------
第31文

英文

Selectively plated with minimum of 150 micro inches of silver on the bonding are
a, measured as in section 6.3.1.

邦文

６．３．１の項の計測方法により、ボンディングされる部分に銀メッキが最低１５０マイ
クロインチの厚さで部分メッキが施されていること。

--------------------------------------------------------------------------------
コマンド>
英語未登録語:
日本語未登録語:
```

図１　同一ファイル内の対訳登録例

英語と日本語がペアで入っているのは極めて稀で、通常の場合は違うファイルに入っているケースが多いのです。例えば日本語から英語に訳す場合には、ワープロ入力された原文が、クライアントあるいは社内から、翻訳部門に送られて来ます。これを受けた翻訳者は、人手翻訳の場合、原文を見て英文ワープロでどんどん翻訳文を入力していくことになります。英語データがファイルで送られて来ると、それを翻訳者が見て、日本語ワープロで訳語を作って行く、これが人間翻訳の普通のやり方です。通常のファイルは、日本語と英語のファイルが分かれているのです。勿論、原文が紙に書かれていて、人手だけで翻訳を行う場合、原文をいちいちファイルに入力することはありませんが、対訳データの収拾のために、翻訳者たちは、暇なときに、原文をファイル入力してくれています。

英文と邦文が別々のファイルに入っている場合、英語のファイル名と日本語のファイル名を入力しますと、「図２　異なるファイルの登録例」のように、英語は左側に、日本語は右側に表示されると共に、システムが勝手に推測して英文と邦文を対応させ、それぞれの先頭に表示し、「対訳登録しますか？」と聞いて来ます。対応しているとき「ｙ」と入力すれば、そのままデータベースに対訳登録されます。対応が間違っていて、例えば、先頭に表示されている英文ではなく、次の英文が、先頭に表示されている邦文と対応している場合、「ｅ」を入力すれば、次の英文が一つ上に上がって先頭に表示されます。何故これをやるかといい

TOPTRAN訳文データベース登録

H ヘルプ　　英語文番号=31 邦文文番号=31　　訳文登録中

31 Selectively plated with minimum of 150 micro inches of silver on the bonding area, measured as in section 6.3.1.	6．3．1の項の計測方法により、ボンディングされる部分に銀メッキが最低１５０マイクロインチの厚さで部分メッキが施されていること。
32 Plating shall be uniform in appearance, smooth, bright in color, free from blisters, peeling, nodules, discoloration, contaminants and burn marks.	メッキは、見た目に均一で、滑らかで、光沢があり、ブリスター、剥離、こぶ、変色、不純物、ムラがあったりしてはならない。
33 There shall be no exposed base metal.	またベースの金属が露出してはならない。
34 Copper strike under plating shall be uniform in appearance,	メッキの下の銅の地金は、見た目に均一で、滑らかで、かつブリスターや、剥離、変色、不純物があってはならない。

対訳登録しますか？ (y/n/q/e/k/r/p/g/t/b/h/l/d/c)

図２　異なるファイルの登録例

ますと、一つの邦文に対して二つの英文、あるいは三つの英文が対応していたり、一つの英文に対して、いくつかの邦文が対応しているような場合、自動的に対応づけることは不可能ですので、オペレータが見て、その組み合わせを判断し、操作することになります。その結果が良ければ、「ｙ」を入力すると、データベースに対訳登録されます。これ以外にも、文を、分割したりくっつけたりする機能を、備えています。

以上が対訳登録の例ですが、対訳登録したデータは、完全に一致する原文の翻訳にしか、使用できません。実際の翻訳では、全く一致する場合は余りなく、かなり違うとか、良く似ていても、どこか違うのがほとんどです。このような場合のために、データベースに蓄積したデータを使えるようにするのが、対訳学習なのです。そして、これら学習された結果を組み合わせて、訳文を生成するのが翻訳なのです。

■対訳学習

登録された対訳の英文、邦文それぞれを、単語レベル、句レベル、また文レベルに互いに対応可能な大きさに分解します。この分解したものを対訳パーツと呼びます。英語の対訳パーツ、日本語の対訳パーツで対応可能なものを、システムが今までの学習結果とパーツ解析結果をもとに推論して、自動的に対応させて表示してくれます。オペレータは、その表示を見て正しければ学習指示を行い、正しくなければ修正操作を行います。「図３　対訳学習例」は、対訳登録の例で説明した対訳を、学習させる場合の例で、この例を使用して、対訳学習の方法を説明します。

システムに対して登録命令を出すと、システムは内部で自動的に固有番号を対訳データに付与します。例えばこれが 3646 番としますと、対訳を学習する段階で、この 3646 という番号を入力します。 3646 が入力されると、システムは、3646 文の、対訳データとパーツ・リンク情報を表示し、「リンクＯＫ？　(y/n)」と表示してきます。これは、 selectively plated と「部分メッキが施されていること」という部分が対応しますか、と聞いてきているのです。更に、 minimum of に対して「最低」、 micro inches に対して「マイクロインチの」が対応しますか、と聞いてきているのです。＊印は、システムが分からない、と答えているのです。この例では間違えてはおりませんが、システムが間違って表示してくることもあります。

ＴＯＰＴＲＡＮでは、「パーツ」という言葉を使用していて、学会などで使用されている「トランスレーション・ユニット」という言葉を使いませんし、構文、ノードという言葉も使用していません。なぜ使用していないかといいますと、この英語と日本語の対応を、文法的に無茶苦茶にやっているからです。minimum of に対して「最低」、 in section に対しては「項の」を対応させていますし、 6.3.1 に対しては「６．３．１の」を対応させています。したがって、いわゆる「トランスレーション・ユニット」とは言えませんし、構文、あるいは、ノードという言葉とは、意味が異なるので、使わないのです。つまりこれがこのシステムの味噌なのですが、オペレータが、こうだ、と決めてやりますと、その通りになります。それがこのシステムの大きな特徴です。なぜこういうことをやるかといいますと、直訳調なら良いのですが、意訳調になると、構文解析など、文法に忠実な解析では、実際の翻訳をやるとうまく訳が出てきません。ここで selectively plated が、なぜ「部分メッキが施されていること」になるのかということは、理屈ではないのです。要するに、この英語を日本語に訳した時に、英語のどこが日本語のどこと、どのように対応するか、それが重要なのです。したがって、このシステムでは、敢えて「パーツ」という言葉を使っています。

システムは、「リンクＯＫ？　(y/n)」と聞いて

きていますが、この例では、with、measured、asについては、パーツ・リンクが行われていませんので、「ｎ」と答えております。「ｎ」が答えられますと、システムは、英語と日本語それぞれのパーツ情報を、番号と共に表示してきます。パーツは、どのようにして作られているかということは、ここでは説明を省略しますが、過去にシステムが学習した結果と、ある規則に則って作られています。オペレータはこれを見て、この例では、3番の with が、なにも対応していないので、英語リンクの3番 with に対して、日本語リンクの16、17番「厚さで」を対応させ、13、14番 measured as も、なにも対応していないので、日本語の3、4、5番「計測方法により」を対応させるように指示します。これらの指示が終わると、システムは、今修正指示された結果を取り入れて、再度パーツ・リンク情報を表示します。オペレータは、自分の望むパーツ・リンクが表示されるまで、この操作を繰り返すことができます。

修正指示が終了すると、システムは、リンク情報すべてを、対訳データベースに書き込んで行きます。オペレータのやることはこれだけです。システムの内部を知らなくてはなにもできない、ということは一切有りません。

■翻訳

次に、このような学習結果の情報を使って翻訳することになります。前編集の必要はなく、原文をそのままキー入力、またはファイル入力をします。入力原文に対して、対訳登録と同じ方法で、対訳データベース・キーが生成されます。生成されたキーをもとに、対訳データベースが検索され、類似度計算結果で一番近い対訳データ情報が引き出されます。

```
                    ＴＯＰＴＲＡＮ対訳学習
  第3646文の学習中
--------------------------------------------------------------------------------
3646 Selectively plated with minimum    ６．３．１の項の計測方法により、ボンデ
     of 150 micro inches of silver on   ィングされる部分に銀メッキが最低１５０
     the bonding area, measured as in   マイクロインチの厚さで部分メッキが施さ
     section 6.3.1.                     れていること。

パーツ・リンク情報
<<1:Selectively 2:plated 18:部分 19:メッキが 20:施さ 21:れ て 22:いる 23:こと >
>
<<3:with #####>>
<<4:minimum 5:of 12:最低 >>
<<6:150 13:１５０ >>
<<7:micro 8:inches 14:マイクロ 15:インチの >>
<<9:of silver 10:銀 >>
<<10:on 11:the bonding area 7:ボンディングさ 8:れる 9:部分に >>
<<13:measured #####>>
<<14:as #####>>
<<15:in section 2:項の >>
<<16:6.3.1 1:６．３．１の >>

リンクＯＫ？  (y/n) n
パーツ・リンク・リストを出力しますか？(y/n) n

英語パーツ情報
1:Selectively 2:plated 3:with 4:minimum 5:of 6:150 7:micro 8:inches 9:of silver
10:on 11:the bonding area 12:, 13:measured 14:as 15:in section 16:6.3.1

日本語パーツ情報
1:６．３．１ の 2:項 の 3:計測 4:方法 5:に より 6:、 7:ボンディングさ 8:れる 9:部
分 に 10:銀 11:メッキ が 12:最低 13:１５０ 14:マイクロ 15:インチ の 16:厚 さ 17:
で 18:部分 19:メッキ が 20:施さ 21:れ て 22:いる 23:こと 24:。

英語リンク：3
日本語リンク：16 17
英語リンク：13 14
日本語リンク：3 4 5
英語リンク：
```

図３　対訳学習例

ここでお断りして置きますが、日本で現在研究開発されているＥＢＭＴ方式のシステムでは、全て「シソーラス」を使用しています。が、私どものシステムでは、シソーラスは一切使っておりません。その理由は、シソーラスは誰が作るのか、シソーラスは万全なのか、という問題が解決されていないからです。

類似度計算結果で一番近い対訳データを引っ張ってくるのですが、普通は、翻訳に必要なパーツが、全部揃う訳でも有りません。足りないパーツがありますと、次に近い対訳データのパーツを引っ張ってきて、それでも足りないときは、更に次に近い対訳データ、というように、この繰り返しをやり、訳文を生成していきます。何か良い例がないかと思い、実際に翻訳を行った文の中を探しますと、びっくりするような綺麗な例がありましたので、ここではこの「図４　日英翻訳例」をもとに説明します。

「ボンディングされる部分に最低１００マイクロインチの厚さで、銀による部分メッキが施されていること」という原文を英語に翻訳する時、システムが類似度計算して、一番近い対訳データを引っ張って来たのが 3646 文です。よく似ていますが、どこか違います。この 3646 文の対訳パーツ・リンク情報は、＜３６４６文のパーツ・リンク情報＞に示されているものです。原文にあるものを、＜使えるパーツ・リンク＞で示されているものの中で探して見ますと、「部分メッキが施されていること」「厚さで」「最低」等はありますが「150」「計測方法により」等は、情報が違っていたり不要であったりしています。したがって、これらをうまく組み合わせる事により、＜訳文＞ Selectively plated with … を出力してきます。

＜原文＞

ボンディングされる部分に最低１００マイクロインチの厚さで銀による部分メッキが施されていること。

＜類似度計算の結果一番似ている文--->３６４６文＞

```
Selectively plated with minimum of 150 micro inches of
silver on the bonding area, measured as in section 6.3.1.
```

６．３．１の項の計測方法により、ボンディングされる部分に銀メッキが最低１５０マイクロインチの厚さで部分メッキが施されていること。

＜３６４６文のパーツ・リンク情報＞

```
Selectively plated  <----> 部分 メッキが 施さ れ て いる こと
with                <----> 厚さ で
minimum of          <----> 最低
150                 <----> １５０
micro inches        <----> マイクロ インチの
of silver           <----> 銀
on the bonding area <----> ボンディングさ れる 部分に
measured as         <----> 計測 方法 により
in section          <----> 項の
6.3.1               <----> ６．３．１の
```

＜使えるパーツ・リンク＞

```
Selectively plated  <----> 部分 メッキが 施さ れ て いる こと
with                <----> 厚さ で
minimum of          <----> 最低
150                 <****> １５０
micro inches        <----> マイクロ インチの
of silver           <----> 銀
on the bonding area <----> ボンディングさ れる 部分に
measured as         <****> 計測 方法 により
in section          <****> 項の
6.3.1               <****> ６．３．１の
```

＜訳文＞

```
Selectively plated with minimum of 100 micro inches of
silver on the bonding area.
```

図４　日英翻訳例

《システム構成》

図５はＴＯＰＴＲＡＮのシステム構成図で、ＴＯＰＴＲＡＮは、「対訳データ登録サブシステム」、「翻訳サブシステム」、「辞書管理ユーティリティ群」により、構成されています。「対訳データ登録サブシステム」の機能には、対訳データ登録、対訳学習、登録テスト、学習テスト、データベース管理があります。「翻訳サブシステム」」の機能には、英日翻訳、日英翻訳、英日翻訳印書、日英翻訳印書があります。「辞書管理ユーティリティ群」の機能には、英和辞書管理、和英辞書管理、英語形態素解析辞書管理、日本語形態素解析辞書管理、対訳パーツ解析辞書管理、辞書データベース送受信があります。

図６は、当社の機械翻訳グループが持っております機器構成です。ＴＯＰＴＲＡＮは、ＮＳＳＵＮ・ＳＰ２及びＡＳ４０６０上で稼働しています。

《評価基準と評価結果》

評価は、ＮＳＳＵＮ・ＳＰ２上で行いました。この資料の数字は古く、現在では大分変わっておりますが、それ以後正式な評価を行っていませんので、この数字で説明します。

評価時の対訳データは 4,359 対、内、英単語数は 44,533 語、学習対訳パーツは 12,268 個で実験を行いました。

「実験１」は、既に対訳データとして学習させたのと全く同じ分野の同種のデータを使って、英文、邦文それぞれ 50 文を抽出して行いました。「実験２」は、既に対訳データとして学習させたのと分野は同じだが、種類の全く異なるデータを用意して、同じく英文、邦文それぞれ 50 文を抽出して行いました。

下訳として使用可能と認める判断基準は、英日翻訳の場合、まず第一に、英単語及びフレーズに対応する日本語訳の表記が、文中で日本語として自然なこととし、意味が通じるだけでは不可といたしました。市販の英日機械翻訳システムでは、翻訳できるといっても、出てきた文をよく見ると、意味は通じるけれど、日本語としてはそうは言わないよ、という表現が多く見受けられます。ＴＯＰＴＲＡＮでは、分からなくはないけれども、というのは不可と致しました。しかし、下訳として使用可能という基準から、少々、単語とか文節の順序が狂っていても良い、また、格助詞その他の助詞は抜けていたり、間違っていても良い、としました。

日英翻訳の方も同様に、訳出された英単語及びフレーズの表記が、文中で英語として自然なこととし、意味が通じるだけでは不可としました。市販の日英機械翻訳システムを使って見ますと、変な英語が出て、このようなことを言うのかな、ということがよくありますが、この類は不可としました。ただし、前置詞 at, on, in などは抜けていても良い、あるいは間違っていても良い、接続詞 and, or についても、抜けていても良い、間違っていても良い、ということを下訳使用可能の条件としました。

その結果、原文に対する下訳として使用可能な訳文の割合は、「実験１」で、英日翻訳の場合 74%、日英翻訳で 76%、「実験２」では、英日翻訳で 18%、日英翻訳で 12% でした。但し、「実験２」では、訳が出ていないのではなく、訳出されてはいるが、自然さがない、意味は分かるけれども、自然でない、というものは全て落としたため低い数字になっております。

《今後の課題》

今まで一寸格好良い説明を行いましたが、まだ、バグその他種々の問題点が有り、日々改良を加えています。それから、対訳データベース中の対訳学習パーツの数が、このシステムの翻訳性能を決定しますので、対訳パーツの学習を、如何に短期間に低コストで大量にやるか、で知恵を絞っています。

システム全般に渡って、アルゴリズムに規則をできる限り入れないようにして、最小限の規則で動かしているのですが、やはりある程度規則を入れた方が良い場合も有りますので、アルゴリズムの対応を検討しております。

現在、対訳データは、総数で数千対のレベルですが、日々増加しております。現段階で、スピードの面では、特に意識したアルゴリズムになっていませんが、市販の機械翻訳システムに比べて、かなり速く、問題は出ていません。しかし、これが、数万、数十万になった時に、スピードはどうなるのか、まだ未知の問題として残っております。

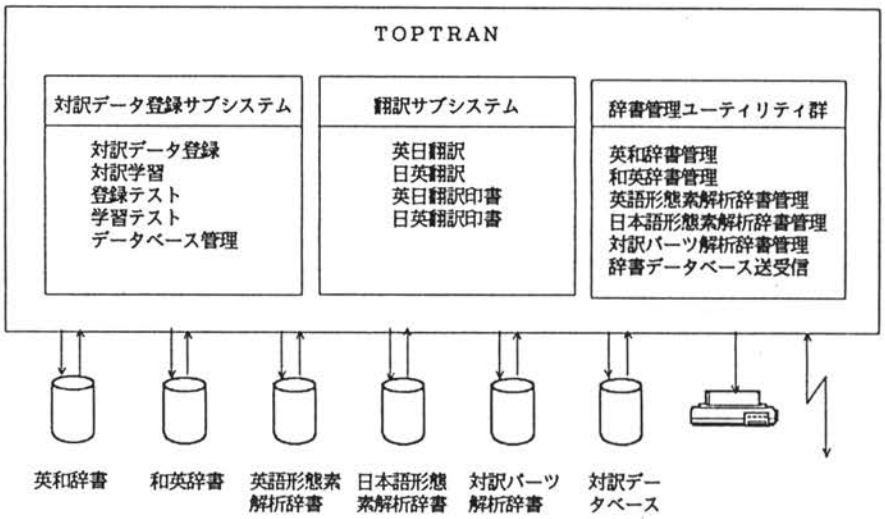


図5　TOPTRANシステム構成図

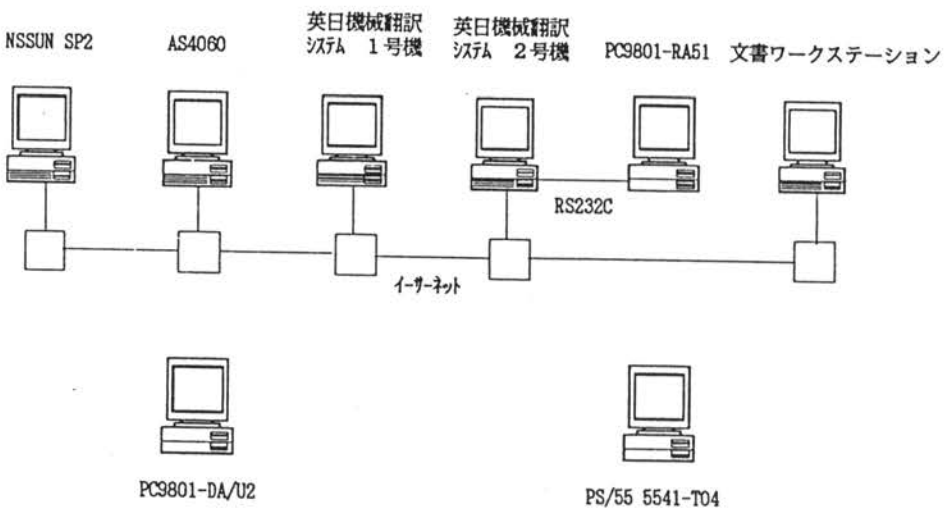


図6　機器構成

ヨーロッパの現在のＭＴ活動

UMIST・KIT　ラルフ・スタインバーガー（筆者紹介14頁）

１９９２年末をもって、１０年間にわたりヨーロッパ連合（ＥＵ）が膨大な努力を注いだＭＴ研究（ＥＵＲＯＴＲＡ）は終わったが、現在さえ、ＴＲＡＤＥとＣＡＴ２－ＥＤＳなどのヨーロッパにおけるＭＴ活動のかなりの部分はＥＵ主導であり、資金供与を受けている。これらのＥＵ関連の活動の他に、ヨーロッパのいくつかの企業が商業システムの開発に投資している。主だった企業プロジェクトとは、ＭＥＴＡＬ（ドイツ・シーメンス社開発）、ＥＵＲＯＬＡＮＧ（フランス・ソフトウェアハウスのＳＩＴＥ開発）などである。どちらのシステムにも、従来の第２世代技術が多かれ少なかれ使用され、非常に精巧なツールとの併用により、翻訳者の仕事をより効率化学する。新しく従来とは異なるアプローチを採用するのはＶＥＲＢＭＯＢＩＬプロジェクトで、これはドイツ連邦政府の開発技術庁ＢＭＦＴ（Ｂｕｎｄｅｓｍｉｎｉｓｔｅｒｉ ｕｍ ｆｕｅｒ Ｆｏｒｓｃｈｕｎｇ ｕｎｄ Ｔｅｃｈｎｉｋ）のスポンサーで開発されたものだ。
ＶＥＲＢＭＯＢＩＬは、音声言語翻訳を必要とするバイリンガルユーザの支援ツールを目指す。そのほかのＭＴ活動には、ＰａＴｒａｎｓ（デンマーク・ＥＵＲＯＴＲＡベースのインハウスシステム）、欧州シャープが開発する多元語商業ソフトウェア（英国・英語、ドイツ語、 フランス語、 日本語）ならびにマンチェスター大学で進められるＭＥＧ（英国・ＵＭＩＳＴ）、これはエグザンプルベース技法を用い、雪崩情報を英語、フランス語、ドイツ語およびイタリア語に翻訳する。

有名なヨーロッパのＭＴプロジェクト
ＥＵＲＯＴＲＡは、１０年以上かかってヨーロッパ共同体（ＥＣ）の１２カ国すべての参加を得た共同作業を１９９２年に終えた。のべ２００名の研究者が平行して研究を行い、３７,５００万 ＥＣＵ（ヨーロッパ通貨単位）をかけたこのプロジェクトの成果発表は、１９９２年にフランス（ナント）で開かれたコンピュータ言語学の国際会議ＣＯＬＩＮＧで提示された。

ＥＵＲＯＴＲＡは、多言語のルールベースＭＴの将来性とリミットについて、統計的手法とトランスファベース手法を使用して調査するという目標があり、非常に野心的な研究プロジェクトだった。
ＥＵＲＯＴＲＡが負った政治上の制約とは、平等性を理由に、ＥＣ加盟の１２カ国すべてがかかわらなければならなかったということである。そのため、９つの公式ヨーロッパ共同体言語すべて（英語、ドイツ語、フランス語、オランダ語、デンマーク語、スペイン語、イタリア語、ポルトガル語、ギリシア語）という、およそ管理可能な範囲を越えた数の言語を同時に扱う必要があったのである。これら９つの言語を各々ソース言語とターゲット言語として用いると、言語対は７２、翻訳方向が１４４にのぼる。これに対して、商業翻訳ソフトウェアの大部分が扱うのは２つの言語であり、ソフトウェア開発者は単一の翻訳方向だけを提供する場合が多い。

最終的研究成果評価委員会（英国人ブライアン・オークレー座長）が設定したＥＵＲＯＴＲＡの主要な成果の１つとして、ヨーロッパ全域に数百人の専門ＮＬＰ学者を組織化し訓練するプランがあった。具体的な目標とは、それまでＮＬＰに関する経験が浅い、もしくは全く持たないポルトガル、スペイン、アイルランド及びギリシアなどの国へ、ＮＬＰ知識を移植するすつというもので、これは首尾よく達成された。

前出の評価委員会によりこれまでにＥＵＲＯＴＲＡが蓄積した資料と経験について、さらに多くの実用化を図るため、より小規模のフォローアッププロジェクトの開始が示唆された。ＥＣでも、ＥＵＲＯＴＲＡその他の研究開発プロジェクトにおける学界と産業界の両方にかかわる多国籍プロジェクトを共催

した経緯から、この提案を認めている。プロジェクト参加企業に各社の研究開発費用の５０%以上を分担させることで、ＥＵは、アプリケーション指向の研究を促し、かつ、企業側がアカデミックなノウハウから利益を得させようとした。

多言語アクションプラン（ＭＬＡＰ）の予算枠によって、ＥＵは２つのＭＴプロジェクト、ＴＲＡＤＥとＣＡＴ２－ＥＤＳに共同出資した。これらはいずれも１９９４年に２カ年計画で開始されたもので、研究課題は、ＥＵＲＯＴＲＡタイプのシステムからプロトタイプを開発し、より少数の言語対について、さらに制限された言語領域（サブランゲージ）を用いる場合も、この方式でリーズナブルな翻訳結果が得られると示すことである。

トランスレーションデモンストレータ（ＴＲＡＤＥ）は、専門化した社会保障ドメインの英語とスペイン語とイタリア語のテキストを、他の言語（６つの翻訳方向）のいずれにも翻訳する。ＴＲＡＤＥは、スペインのソフトウェアハウス Ｃｅｎｔｒｏ ｄｅ Ｃａｌｃｕｌｏ ｄｅ Ｓａｂａｄｅｌがコオーディネーションを行い他のパートナーとしてＵＭＩＳＴ（英国・マンチェスター大学）と Ｆｕｎｄａｃｉｏ Ｂｏｓｃｈ Ｇｉｍｐｅｒａ’（スペイン・バルセロナ）と Ｇｒｕｐｐｏ ＤＩＭＡ’（イタリア・トリノ）とＩＮＰＳ（イタリア社会保障機関）である。Ｅスターと呼ばれる Ｇｒｕｐｐｏ ＤＩＭＡ’ 開発のもので、ＥＵＲＯＴＲＡから派生した形式を採用している。

ＴＲＡＤＥは、３つの言語に対し、ＥＵＲＯＴＲＡ文法と辞書の最適化されたバージョンを利用する。既存の資源の再利用の他に、辞書ツールの利用で、専門辞書を各言語あたり５０００単語のデモンストレーションサイズに拡大できる。その他の機能として、長文を分割するテキスト・セグメンテータなど、ユーザインタフェースが開発される予定だ。さらには、形式はＥＵＲＯＴＲＡと異なり、不当な入力や、未知語を含む文をも扱えるように強健になる。

ＭＬＡＰ枠で資金供与を受けた第２のプロジェクトは、ＣＡＴ２－とＥＤＳと呼ばれ、ドイツの会社エレクトロニク・データ・システムズ（ＥＤＳ）がコオーディネーションした。このうちＣＡＴ２は、ＥＵＲＯＴＲＡに参加した「応用情報科学研究所」（ＩＡＩ＝ドイツ・ザールブリュッケン）によって開発されたもので、やはりＥＵＲＯＴＲＡファミリーの一つである。その他、ＵＭＩＳＴ（イギリス）とフランスパリ第七大学のタラナ研究所が開発に参加している。ＴＲＡＤＥ同様の機能を備えるものとなるだろうが、ＣＡＴ２－ＥＤＳチームは１９９５年中には、ドイツ語・英語・フランス語相互のＭＴとして、オペレーティングシステムメッセージとソフトウェアマニュアルの翻訳専門の、プロトタイプを提示するだろう。

より少ない言語対を対象として使用する場合、また専門化された分野のサブランゲージに適用するとき、ＥＵＲＯＴＲＡのような形式が成功するという点は、最近、デンマーク語・英語のＭＴシステム ＰａＴｒａｎｓによって、示された。
ＰａＴｒａｎｓは実際にＥＵＲＯＴＲＡ形式を基礎に開発され、デンマーク語と英語の文法を用いた。ＴＲＡＤＥやＣＡＴ２－ＥＤＳと異なり、ＰａＴｒａｎｓは、デンマーク言語技術センター（デンマーク・コペンハーゲン）が単独で開発した。１言語対だけを翻訳対象として、原言語は英語に制限している。
ＰａＴｒａｎｓの研究資金はＥＵ供与ではなく、デンマークの特許翻訳企業Ｌｉｎｇｔｅｃｈおよび開発者であるＣＳＴの自己資金である。２年の開発期間の後、１９９３年の終わりにインハウスのＭＴシステムを配布した。開発者とクライアントの両方が翻訳結果に満足したことから、サブランゲージへと研究対象を拡大する計画である。

これらのアプリケーション指向プロジェクト開始の他に、ＥＵは、中・長期の目標として、ヨーロッパ圏内のＭＴソフトウェアその他のＮＬＰ応用技術の開発を上げている。これらの目標とは、共同開発を通じて、９つのＥＵ公用語すべての標準化を図り、言語資源を作成すること、および、それらの取得・生成・適用およびそれらを多くの目的のために応用するためのソフトウェアツールの開発を含む。この言語学的インフラは、ＥＵ圏内のすべての企業と研究機関の利用に供することで、研究者と開発者のために非常に貴重な資源となるだろう。現在のＥＵに

おけるＭＴ関連の諸活動の成果が、時間とお金の節約を助けるだけにとどまらず、さらに将来の文法の開発と、辞書及びツールの再利用を促すよう、望まれる。ＥＵにおけるこの方向の貢献の例として、ＡＬＥＰとＬＳ－ＧＲＡＭの２つを取り上げたい。

ＡＬＥＰ（高度言語エンジニアリング・プラットフォーム）は言語対象をトレイシングしデバッグし見るためのツールを持つ形式で、グラフィックのユーザ・機械インタフェースとテキスト処理システム他、ずべてのＮＬＰ開発に必要なツールがある。１９９１年以来開発を支持するＥＣ委員会（ＣＥＣ）がＡＬＥＰに期待するのは、堅固かつ携帯可能で広く実用可能なソフトウェア・プラットホームの確立であり、専門化されたアプリケーションのプロトタイプの開発にこれを提供することだ。ＣＥＣでは、産学一致してこのプラットホームを使用し、多くのさまざまなＮＬＰアプリケーションを開発すれば、それら研究開発の結果は、ヨーロッパ全域でコンパチブルかつ再使用可能な状態となるという理想像を描いていた。すなわち、この共通のプラットホーム利用を通じて、学界および企業の研究機関の間にシナジーを助成するわけである。ＡＬＥＰの最終バージョンは今年中に実用化されるだろう。

ヨーロッパでは、完全なNLP研究のインフラを確立するため、もう一つの主要なプロジェクトLS-GRAM（EC言語のための大規模文法）が進められてきた。研究期間２年のこのプロジェクトは、9つすべての公用語の大規模なドキュメンテーションを含む、コア文法をヨーロッパNLP学界へ提供することを目指した。文法はALEPを用いて書かれ、それらは「基本セット」としてALEPを使用するすべての研究開発機関に分配される予定だ。ALEPとコア文法の利用が、ヨーロッパ全域の産学協同をさらに助長し、またそれがアプリケーション開発に対しても、トレーニングや研究、実用化を意識したものを生む基盤として魅力あるものとなるよう、望まれている。1994年開始の同プロジェクトの中心となった「応用情報科学研究所」（IAI＝ドイツ・ザールブリュッケン）はじめ、エセックス大学（イギリス）、Fundacio Bosch Gimpera'（スペイン・バルセロナ）スツットガルト大学（IMS-CL・ドイツ）が参加した。1993年までにECが言語学関連のプロジェクトに対して行った資金供与は総額1,5億ECUに上る。内訳は、MLAP（1977年～）、EUROTRA（1982～92年）、LRE（1991年～）、そしてESPRIT（1984年～）関連である。これらの諸活動は、1994年から98年にかけて「第４次基幹プログラム」に編入され、予算規模４億ECUを投入して継続される。

しかしながら、ＥＣの一般的原則として、加盟諸国諸地域が独力では行えない研究活動、すなわち標準化の開発と各の研究努力の連携を図ることに助成対象が決められている。商業的指向の強いプロジェクトは、企業自身が資金供給しなければならない。ここで取り上げる主要なMTプロジェクトは、既存のMETALシステム（ドイツ企業シーメンス開発）EUROLANG（フランス中心の商用プロジェクト）およびドイツの国家研究プロジェクトVERBMOBILである。

METALの開発は1960年代半ばにアメリカ・テキサス州のオースティン大学で始まった。1978年より、同プロジェクトはドイツ・ミュンヘンに本拠を置くシーメンス社の支援を受け、80年には同プロジェクトは同社に移籍された。同社はドイツ語－英語の市販版をリリースしたのは89年になってからである。また92年まで、同システムは稼働システムが限られており、記号リスプマシンもしくはシーメンス社のワークステーションをソフトウェアと同時に購入するほかなかった。そのため、価格的にも25万DM（ドイツマルク＝約1,500万円）となり、販売実績は20台以下であった。

その後、シーメンス社は翻訳対象言語対を拡大、英語、ドイツ語、フランス語、オランダ語、スペイン語、デンマーク語を加えた。また、SUNワークステーションに移植、現在では約5.5万DM（約330万円）で販売している。

METALの特徴は、フルオートマチックな翻訳の提供に加え、ユーザに便利なツールを多く用意した点だ。大型の専門辞書、セミオートマチックな辞書アプデート機能、また後処理のための機能などが含まれている。ユーザにとっては、図表を含む入力文の書式が保存されるため、翻訳結果が入力文と全く同じ書式で得られる点が、時間短縮の大きな味方である。

シーメンス社は、フランスの SITE 社が率いるEUROLANGの開発にも参加しているが、このプロジェクトには10以上のヨーロッパの企業および大学が名前を連ねる。1992年～1994年の見込み予算は7億3500万 ECUである。1994年末までにSITEは、ドイツ語、英語、フランス語、イタリア語、スペイン語（その他多くの言語が計画中）の全自動MTのみならず、完璧な「翻訳者のためのワークベンチ」の完成と提供を計画している。全自動翻訳は、フランスのAriane(GETA)システムとMETALシステムの両方で行われる。

METALがもたらすツールやその他の利点以外にも、EUROLANGは、人間翻訳の支援として、理想的なワークベンチの役割を担う。一般辞書、専門辞書ともオンラインでいくつも提供する上、用語管理ツール、あるいは前回翻訳したテキストを検索するツールを提供できるからだ。特に後者は仕事のむだな重複を防ぎ、例えば修正やデータ更新を行った文書の翻訳には有効である。用語管理ツールはデータベースとして同システムを利用する複数の翻訳者に利用され用語統一に役立つ。EUROLANGは、SUNワークステーション上を走るだろう。

もう一つ、より研究指向の高いプロジェクトとしてVERBMOBILがある。これはドイツ人工知能研究センター（DFKI＝Deutsches Forschungsinstitut fuer Kuenstliche Intelligenz）といって、ドイツ政府（BMFT＝Bundesministerium fuer Forschung und Technik）が資金供与するものだ。VERBMOBILの目標は大胆で、対面型の会話のために、携帯用のドイツ語英語日本語翻訳支援のプロトタイプを作成することである。2年の準備研究の後、プロジェクトは1993年に公式に始まり、2001年終了を計画している。VERBMOBILはさしあたり国家プロジェクトであるが、アメリカと日本人のパートナーとの共同研究が計画されている。日本では、同様のプロジェクトを行っている京都の ATR音声翻訳研究所との協力が検討されている。アメリカでは、カーネギーメロン大学、スタンフォード大学のCSLI及びバークレイの国際コンピュータ科学研究所(ICSI)との共同研究が準備された。ドイツ国内での研究提携先は、シーメンス社、ダイムラー社、テレフンケン社及び9大学が含まれる。

VERBMOBILは、電話の会話の自動逐次翻訳システムとは規定されていない。むしろ、面と向かってともに英語を話す日本人とドイツ人に対して、会話の細かい表現に困ったときに支援するものである。この二人が何か英語表現で問題を感じたらすぐその場でVERBMOBILを動かすと、日本語かドイツ語の単語の英訳を即座が与えられるわけだ。これらは整った文の形である必要はない。翻訳プロセスは増加つまりシステムは直ちに荒翻訳を提供することができるし、もっと時間があるなら、さらに洗練された翻訳を代わりに提供することもできるのだ。

最初のデモンストレーション版は1995年リリースの予定だが、その会話のシチュエーションは、2人のビジネスパートナーの会合という設定で、目の前にカレンダーを広げて次のミーティングをいつ行うか話し合っている。このバージョンから2001年の完成版までに，アプリケーションドメインは徐々に広げられるだろう.

このプロジェクトの主要な目標は非常に野心的で音声の認識と合成、ダイアローグと知識処理及び機械翻訳など、高度な情報テクノロジーと自然言語処理の各分野を、統合化することである。最初の4年で、BMFTの資金は6,000万DMに達した。またVERBMOBILへのドイツの総合的な投資は、たぶんプロジェクト終了までに12,000万ECUに届くだろう。

これらのプロジェクトの他に、ヨーロッパではNLPの分野で多くの他の活動が行われている。多くのアプリケーションは、研究成果を即座に実用化する方向であり、例えば文法と文型のチェッカー、自動テキスト分類機能及び知的情報検索などが上げられる。しかしながら主にEU主導だが、多くは中・長期的な目標達成を目指している。将来的な投資として、大きいコーポラや他の言語学のデータ収集、評価テクノロジーの開発、既存の資源の再利用性の検討、コーポラからの知識取得の自動化、その他を含むアクティビティが行われている。ヨーロッパには、言語障壁に打ち勝つことにおいて経済的、政治上の大きな関心がある。いくつかの不可避な退行があったにもかかわらず、健全なNLPインフラの構築を目指し、協調して努力することを通して、ヨーロッパ諸国はこの問題を解決しようとする強い姿勢を示している。

BehaviorTran:

―― その現状、見とおし、および原理――

台湾国立清華大学　電気工学科　張 景 新・蘇 克 毅

1. 編成と歴史的概観

ビヘービアトラン英中機械翻訳システム(Behavior Tran:旧称アークトランArchTran)は、台湾すなわち中華民国で着手されたこの種の研究で最初のものです。また、世界で初めて商品化した英中システムのひとつでもあります。初めに商品化したシステムは1989年にリリースし、現在翻訳サービス・センターの中核として稼働しています。

ビヘービアトランに関する研究は、1985年5月に、台湾の国立清華大学(NTHU)とビヘービア・テック・コンピュータ社（BTC)との共同作業として開始されました。この作業は、ＮＴＨＵの教授　蘇 克 毅が指導し、地元の大学と研究機関の数人のプロのコンサルタントがサポートしました。研究の拡大にあわせ、このMTシステムの改良を目的として1988年2月に設立した、ビヘービア・デザイン社(BTC)R&Dセンターへ移しました。

ビヘービアトランMT R&Dグループの特徴は、MT R&Dグループであるというだけではなく、内部共同機械翻訳サービス・センターとしての役割も果たすという点です。ここでは、商業化した顧客サービスとMT研究チームに対するリアル・タスク・フィードバックを行います。現在、およそ 30人のR&Dスタッフが R&Dセンターで働いており、そのうちおよそ10人はMT研究グループに所属します。MT、翻訳サービス・センターは、専門のプロジェクト・マネージャーと専従ポスト・エディタで操作します。パートタイムのポスト・エディタは、公共の電話ネットワークを通してセンターとリンクされます。このような編成は、MT環境における実際の翻訳作業の全プロセスについて調査するための良い環境となっています。

このR&Dセンターは、学界と強く結びついており、この地域およびいくつかの国際的な団体におけるコンピュータ言語学会のアクティブなメンバーでもあります。国際会議、MT 専門誌および一般のNLP学界誌に、50編を超える技術論文を発表してきました。技術論文の発表は、ここでは定期的に行っています。また、ROCLING協会(中華民国計算語源学言語学会)の設立メンバーの一員でもあり、関連事業にも活発に参加しています。さらに、北アメリカのいくつかの有名な研究室と、数件の国際的技術提携や資源共有関係を結んでいます。

目標として、スタンドアロンの機械翻訳システムを提供するのではなく、ほぼ内部(インハウス)翻訳の環境に近いものをユーザに提供することです。これは設立以来のポリシーですが、世界の商用MTの運用の成功例にならい、「本物の解決策」を求める顧客に翻訳サービスを提供する最良の方法とは、顧客をMTシステムの保守経費などで煩わさずに内部翻訳を提供することだとの結論に至りました。

その結果、VAN（付加価値通信網)ベースの翻訳サービス・センターを創設しました。サービス形態としては、地方または公共のデータ・ネットワークを通じて電子文書が顧客、サービス・センターおよび

筆者紹介(10頁ヨーロッパの現在のＭＴ活動)

ラルフ・スタインバーガー博士 はTRADE及びCAT-2EDSのUMISTグループの代表とテクニカルマネージャを務める。MTとの取り組みは1989年からで、METAL(シーメンス社＝ドイツ・ミュンヘン) CAT2(IAI＝ドイツ ザルブリュッケン大学)EUROTRA(UMIST＝イギリス・マンチェスタ大学と各プロジェクトで研究した。
1994年4月より九州工業大学にてリサーチフェローとして来日中

ポスト・エディタの間を行き来します。これによって、時間、コストおよび機密保護の保証に関する顧客の負担を大いに削減します。また、顧客が自分たちで知識ベース（例えば辞書）を保守する経費もかかりません。

現在、ビヘービアトランは国際的に有名なコンピュータ企業数社を含む顧客層を確立しました。これらの顧客は、ほとんどがソフトウェア、ハードウェア、PCまたはワークステーションの分野では世界のトップランクの有名企業です。

ビヘービアトラン利用は、まずコンピュータ・マニュアルおよび関連ドキュメントの領域が上げられます。そして、機械や医学など他の技術分野についても、さまざまな私的、公的機関に対してサポートしています。

このシステムは、UNIXワークステーションの環境で開発されています。すなわち、C言語で書かれています。未処理の翻訳は、ほとんどPC上で後編集されます。PCは、翻訳サービス・センターのネットワークを通して翻訳サーバに接続しています。また高性能のサポート・ツールを翻訳ワークステーションにパッケージし、OCR、専用テキスト・エディタ、内部辞書、biテキスト、DTPなどの支援機能が統合されています。これらサポート・ソフトウェアは、BDCの他のR&Dグループからの提供もしくは共同開発で実現しました。これらの機能には、中英二カ国語の文書作成ワークベンチが組み込まれています。特徴は、中英二か国語のデスクトップ出版システムとして、テキスト・式・表・グラフィックおよびイメージ作成用の十分な機能をもつことです。

2. 一般的、技術的見通し

ビヘービアトランの研究は、従来のトランスファベースのMT構造から着手しました。システムでは、多くのルールがコード化され、種々の言語学的課題を処理します。しかし、システムが拡大すると、このようなルール・ベースのアプローチではいくつかの問題にぶつかることがわかっています。特に、あいまいな知識を扱うための客観的選択基準を欠くことと、複雑で不規則な知識の扱いが困難な点が指摘されています。ルールに関しても、時々例外が生じます。システム工学のレベルでは、異なる時間、異なる人々の間の、多量できめ細かい知識の一貫性を維持するのは困難です。種々の文献で提案されたように、言語学の知識を習得する系統的方法はまだだれも確立していません。

このように、人間が介入する多量のきめ細かい知識の習得は、費用と時間がかかります。したがって研究開始から2年後には研究の焦点はコーパス・ベースのアプローチに移りました。これによって、ビヘービアトラン・システムは、従来のMT方式と新しいコーパス・ベース・テクノロジーの統合型、すなわち、現在MT界で「ハイブリッド方式」と呼ばれるものになっています。このような構造を持つビヘービアトランは、混成のパージング戦略（いわゆるボトム・アップ・パージングとトップ・ダウン・フィルタリングの併用）と得点つきパージング・メカニズムおよびコーパス・ベースで統計学指向のパラダイム(CBSO)を採用しています。

ビヘービアトラン・システムの研究は、コーパス・ベースで統計学指向であるという特徴があります。したがって研究は、言語モデル・パラメタを取得するための系統的かつ自動的な方法を設計し、あいまいさを分析するための一定の確率的得点機能をもつ選択基準を使用する方向に向かっています。

CBSO言語モデルの開発において特に強調されるのは、あいまいさを分析するために、完全な意味解析ではなく、識別機能を使用することです。この原理は、必要なあいまいさ分析機能をもち、非常に小さなコストで知識の取得と保守を行う有効なコーパス・ベース・テクノロジーに通じるものです。コンピュータの部分では、言語パラメタを未知のテキストにも適用できるように、ビヘービアトランのCBSOアプローチはコンピュータ・モデルの強健さに重点を置いています。

これらをすべて考慮して、ビヘービアトランR&Dチームは、辞書解析、構文解析、意味解析、確率的トランスファおよび生成のモデル、統一された確率的な骨組みのもとのすべてのものについて、種々の選択基準を確立しました。また、これらのモデル用の種々の言語パラメタ上でより良い判断をするために、識別および強健指向のテクニックも確立しています。

3. 二か国語のコーパス集をもつ両方向トランスファに向けて

ビヘービアトランの操業以来、顧客のほとんどが翻訳出力を公に「出版できる」ことを強く求めてい

ることがわかりました。さらに、ターゲット言語の文書のスタイルが選択可能であればと望む顧客もあるでしょう。これは、通常、原語の文書のスタイルと異なるスタイルで翻訳出力が得られるためで、顧客によっては翻訳されたデータを、長年にわたって開発してきたスタイルと一致させたほうが使いやすいためです。　これらから、顧客が原語のスタイルを保持するよりも、ターゲット語をより好ましいスタイルで作成することの方に関心があること、そして、システムを自分たちのスタイルに適合させたいという強い要求をもつことがうかがわれます。また顧客が非常に関心を示す事柄として、それぞれの顧客の独自の要求に合うようにシステムを調整するのに費やされる時間と、顧客のフィードバックへの応答にかかる時間も非常に重要です。

現在のMTシステムがまだ広く利用されていない主な理由の１つは、生成された出力が原語に強く影響される、すなわち、スタイルは単純にターゲット語のネイティブスピーカーが使うような文体にはならないという事実にあります。MTシステムの従来のトランスファと生成のプロセスは、特定のトランスファユニット上に手作業でコード化したトランスファルールに従うので、これは当然のことです。トランスファ操作は、原語からターゲット語へと設計されるので、本質的には単方向です。すなわちトランスファルールは、原語の文法と「正当」で「自然」なターゲット語の文法とによって生ずる構文的構造間のマッピングが最良になるように調整されているわけではありません。

また、もう１つの問題は、適切なトランスファユニットとそのような手法によるトランスファ操作の正当なシーケンスを確認することです。有限個数の基本的ユニットにトランスファユニットを配置する系統的手法がなければ、トランスファルールまたはトランスファ操作を完全にまた有効に確立するのは困難です。さらに、そのようなプロセスのほとんどはパラメタ化することができないので、顧客のフィードバックを顧客の特定のスタイルに合う入力として使用することはできません。したがって、それぞれの可能なトランスファ操作に選択得点をつけることによって、顧客の選択に合うようにシステムを調整するのは困難です。

そこで、ビヘービアトラン・チームの目標は、２つの言語の間の双方向のトランスファおよび生成のモデルを構成し、トランスファおよび生成のプロセスを、高度にパラメタ化されたフィードバック制御の、トランスファのユニットと操作が得られる自動的なアプローチをもったシステムにすることです。

そのような目標を達成するために、原語の文法とターゲット語の文法のそれぞれによって生じる翻訳の組の注釈つき構文ツリーを、標準化されたバージョンのこれら注釈つき構文ツリーに転換する方法を我々は考えました。こうすれば、そのようなツリーの組の間の基本的トランスファユニットは自動的な手法で得られます。トランスファ得点と生成得点をそのような配置されたトランスファユニットに関して定義し、最良のマッピングまたは好みのマッピングを見つけます。

そのようなモデルでは、トランスファ操作が有限個数のトランスファユニットだけに制限される確立が高まります。さらに、標準化された構文ツリーはそれぞれの文法によって作成されるので、このメカニズムによって選択されたターゲット語の構造は、原語の文法に影響されません。したがって、マッピングは本質的に双方向で、どちらの側からも始めることができます。

構文的構造を基本的トランスファユニットに分解し、トランスファと生成の得点のための一定の確率的なモデルを使用することによって、トランスファと生成のプロセスをパラメタ化するのは簡単です。そのようなパラメタ化されたトランスファモデルを用いて、ビヘービアトランが採用する手法は特定の顧客のスタイルにトランスファパターンを調整する潜在的能力を増すことになります。したがって、我々はこの手法を強く支持しています。

4. 今後の目標

10年間近く機械翻訳の研究開発に携わった後、ビヘービアトランMTシステムの商用運用を通して、機械翻訳をより実用的に、また広く使用されるようにするためには、新しいコーパス・ベースの手法が非常に重要な役割を果たすと確信するに至りました。要するに、人間は一般的な言語をモデル化する能力を十分もっているが、コンピュータは大規模なデータ処理に有効です。したがって、よく認識されている言語学の現象を利用して、人間が確率的な言語モデルを構成し、統計学的に十分確立されたテクニックをもつ大きなコーパスから、確率的なモデルの言

語パラメタを評価することが適切です。

現在、この方向で、いくつかの高度なテクノロジーが開発中です。特に、自動コーパス変換、注釈、自動モデル改良および、よりパラメタ化された言語モデルが研究課題になっています。

参考文献

[Chang 93]=Chang, J.-S. and K.-Y. Su, "A Corpus-Based Statistics-Oriented Transfer and Generation Model for Machine Translation"(「コーパス・ベース、統計学指向の機械翻訳のためのトランスファおよび生成モデル」、TMI-93論文集 pp.3--14、1993年)

[Chen 91]=Chen, S.-C., J.-S. Chang, J.-N. Wang and K.-Y. Su, "ArchTran:A Corpus-Based Statistics-Oriented English-Chinese Machine Translation System"(「アークトラン：コーパス・ベース、統計学指向の機械翻訳システム」、第3回機械翻訳サミット論文集 pp.33-40、1991年)

[Su 90]=Su, K.-Y. and J.-S. Chang, "Some Key Issues in Designing MT Systems(「MTシステム設計のキーポイント」、機械翻訳vol.5, no.4, pp.265-300、クリューワーアカデミック出版、1990年)

[Su 92]=Su, K.-Y and J.-S. Chang, "Why Corpus-Based Statistics-Oriented Machine Translation(「なぜコーパス・ベース、統計学指向の機械翻訳か」、TMI-92論文集pp.249-262、1992年)

[Su 93]=Su, K.-Y. and J.-S. Chang, "Why MT Systems Are Still Not Widely Used?(「なぜMTシステムがまだ普及しないか」、機械翻訳vol.7, no.4, pp.285-291、1993年)

筆者連絡先

所在地：

E-Mail：

テクニカルコミュニケーション

第6回TCシンポジウム

開催月日　平成6年8月25日（木）～26日（金）
開催場所　国学院大学新宿キャンパス（JR新宿駅西口より徒歩4分）
テーマ　ユーザーインターフェースをささえるTC技術

内容　特別セミナー　文章作法、認知心理学、PL、レイアウト技法
事例発表　マニュアル改善事例、研究活動等
セッション　マルチメディア開発、ビジュアル表現、機械翻訳
文章校正法、制作技術他
マニュアル・カタログ展示会

主催　テクニカルコミュニケーター協会
問い合せ先　東京都新宿区北新宿4-8-16（Tel 03-3368-4607　FAX 03-3368-5087）
協力　(社)日本翻訳連盟、(社)日本翻訳協会、アジア太平洋機械翻訳協会

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

1．はじめに

1990年10月に当社が初めてパソコン通信ネットワークを利用した機械翻訳サービスをNIFTY-Serveで開始して以来、それに付随するサービスや国内の大手パソコン通信ネットワークで次々と同様なサービスが提供されて来ています。表1は現在行われているパソコン通信機械翻訳サービスと関連サービスの一覧です。

パソコンの普及と相まって、いつでも、スピーディに、安価な料金で利用できるパソコン通信機械翻訳サービスは多くの利用者からご支持を頂き、今や新しい文化として定着した感があります。そこで本号では前号(1992.1No.4)のサービス概要紹介に続きシステム構成、利用方法等についてご紹介します。

表1　ＡＴＬＡＳ機械翻訳サービス、及び関連サービス

平成６年４月現在

ネットワーク	サービス名	サービス内容	対象言語	翻訳所要時間[1]	料　　金[2]	サービス提供会社
NIFTY-Serve	ATLAS 機械翻訳	機械翻訳のみ	日英、英日	A4版１枚 15 分	日英：１文字１円　英日：１文字２円	富士通ﾄﾞｷｭﾒﾝﾄｻｰﾋﾞｽ
	インターグループ後編集室	機械翻訳の後編集	日英、英日	受信後１～７日（文字数による）	日英： 500字まで1200円 50字毎 100円 英日： 200字まで1000円 50字毎 200円	インターグループ
		ネイティブ翻訳チェック	日英	受信後１～７日（文字数による）	500字まで 600円 50字毎 70円	
	バベル後編集	機械翻訳の後編集	日英、英日	受信後１～７日（文字数による）	日英：1000字まで2500円 500字毎1250円 英日： 400字まで1800円 250字毎 900円	バベル・インターナショナル
		ネイティブ翻訳チェック	日英	受信後１～７日（文字数による）	1000字まで2500円 500字毎1250円	
	イリス機械翻訳	機械翻訳のみ	英←→仏 英←→独 英←→西	受信後 24 時間	英→仏独西：１文字２円 最低 1000円 仏独西→英：１文字３円 最低 1000円	イリス・インターナショナル
	日韓機械翻訳	機械翻訳のみ	日韓	受信後 24 時間	原稿： 200字まで1000円　１文字５円	高電社
ASCII-NET	ATLAS 機械翻訳	機械翻訳のみ	日英、英日	A4版１枚 15 分	日英：１文字１円　英日：１文字２円	富士通ﾄﾞｷｭﾒﾝﾄｻｰﾋﾞｽ
	インターグループ後編集室	機械翻訳の後編集	日英、英日	受信後１～７日（文字数による）	日英： 500字まで1200円 50字毎 100円 英日： 200字まで1000円 50字毎 200円	インターグループ
		ネイティブ翻訳チェック	日英	受信後１～７日（文字数による）	500字まで 600円 50字毎 70円	
PC-VAN	IBS 機械翻訳	機械翻訳のみ	日英、英日	A4版１枚 20 分	日英:1文字 0.6円 英日:1文字 1.5円	アイ・ビー・エス
		エディット付	日英、英日	受信後 24 時間	日英：１枚 1500円 最低 3000円 英日：１枚 1200円 最低 3000円	
		リライト付	日英、英日	受信後 24 時間	日英：１枚 4500円 英日：１枚 3600円	
	日韓機械翻訳	機械翻訳のみ	日韓	受信後 24 時間	原稿： 200字まで1000円　１文字５円	高電社
EYE-NET	英日機械翻訳	機械翻訳のみ	英日	受信後 24 時間	１文字１円3 円 最低 2000円	ノヴァ

注 1) 概略の時間を示す。依頼数、翻訳量および休日にかかるかどうかで変わる。詳細は、それぞれのサービス案内を参照されたい。
注 2) 日韓機械翻訳以外は、翻訳結果の文字数・枚数で料金が計算される。枚数換算など詳細は、それぞれのサービス案内を参照されたい。

2．サービスの内容

パソコン通信機械翻訳サービスの翻訳システムは市販の機械翻訳プログラムを使用しています。従って、その翻訳品質は一般の機械翻訳と同等です。図１に夏目漱石の「我輩は猫である」の英訳結果と図２にオー・ヘンリーの「最後の一葉」の和訳結果を載せます。

一般的に機械翻訳は、口語や会話文、詩歌、文学作品等の翻訳には向かないと言われています。１文が短く、省略や曖昧な表現の少ない技術文書やマニュアルの翻訳に利用されると、良い結果が得られます。特に、大量の文書を翻訳した場合、用語や表現の統一が保てるという長所があります。

また、その品質は使用する辞書に左右されます。

夏目漱石『我輩は猫である』より
1. 我輩は猫である。
1. I am a cat.
2. 名前はまだ無い。
2. There is no name yet.
3. どこで生まれたか頓と見当がつかぬ。
3. *** (*S) does not guess with haste where born.
4. 何でも薄暗いじめじめした所でニャーニャー泣いていた事だけは記憶している。
4. In the dim place sloppily, anything memorizes only the crying (*nyahnyah*) thing.

オー・ヘンリー『最後の一葉』より
1. In a little area west of Washington Square, many little streets cross each each other.
1. ワシントンSquareの小さい領域の西では，多くの小さい街の人々は各互いに交差する.
2. One street crosses itself once or twice.
2. ある街の人々は2回自分に交差する.
3. An artist once found that there was something good in this street.
3. 芸術家は，一度，この通りで何か良いものがあることがわかった.
4. If someone came to this street to get money for paints, paper and canvas, he might meet himself comming back without having gotten one cent.
4. だれかが絵の具のためのお金，ペーパー及びキャンバスを得るために，この通りにこんな事になるならば，彼は1つのセントを得ないで後部をcommingする自分に会うだろうに.

図1　翻訳結果例

当社のパソコン通信機械翻訳サービスでも、専門用語辞書が用意されており、利用者が指定できるようなっています。さらに、利用者固有の用語を翻訳の際に使用できるように名詞訳語指定機能も提供されています。図3に技術文書を専門用語辞書と名詞訳語指定機能を使用して翻訳した結果を載せます。

また、翻訳結果の利用形態に合わせて翻訳結果の形式を次の4種類の中から選択することが出来るようになっています。

(1) 文番号付き対訳
(2) 文番号無し対訳
(3) 翻訳結果のみ
(4) 文節毎の対訳

この他、翻訳結果をパソコン通信に加入している複数の人々に配達する機能（転送サービス）やパソコン通信には加入していないが、ＦＡＸならあるという人に翻訳結果をＦＡＸで転送する機能（ＦＡＸ配信サービス）も提供されています。

3．システム構成

当社の現用のパソコン通信機械翻訳サービスシステム構成図を図4に示します。

翻訳依頼文書

PATH（パスの設定）
このコマンドは、外部コマンドを捜すためのディレクトリを指定します。パスが設定されていない場合、コマンドプロセッサは外部コマンドをカレントディレクトリから捜します。

基本機能だけで翻訳した結果

PATH (setting of passing)
This command specifies the directory to search for an external command. If passing is not set, the command processor searches for an external command from a current directory.

専門辞書（情報処理と電気）と名詞訳語機能（パス＝path）を使用し翻訳した結果

PATH (setting of path)
This command specifies the directory to search for an external command. If the path is not set, the command processor searches for an external command from the current directory.

図２　専門辞書と名詞訳語を使用した翻訳例

翻訳依頼文書の収集及び翻訳結果の送信にはパソコン通信の電子メールシステムを使用しています。日英および英日の機械翻訳サービス用に、それぞれ専用のＩＤが用意されています。依頼者が翻訳依頼文書を送信すると、このＩＤのメールボックスにその翻訳依頼文書が格納されます。パソコン通信のホストシステムと当社の機械翻訳システムとは電話回線で接続されており、機械翻訳システムが定期的にメールボックスを調べています。メールボックスに翻訳依頼文書があると機械翻訳システムはその文書を自動的に収集し翻訳して、その翻訳結果を依頼者のＩＤに自動的に送信します。このように翻訳依頼の収集から翻訳結果の送信まで全自動処理を行っています。

NIFTY-ServeとASCII-netとの相違は、パソコン通信のホストシステムと機械翻訳システムとの接続に前者では一般の電話回線を使用しているのに対して後者では専用電話回線を使用していることです。

このように電子メールシステムを使っていることから、パソコン通信のホストシステムと当社の機械翻訳システムは、互いにシステム的な影響を及ぼすことは全くありません。例えば、パソコン通信のホストシステムが定期点検などで停止しても、機械翻訳システムは翻訳依頼文書を収集できないだけで、ホストシステムの運用が再開されると再び、翻訳依頼文書の収集を始めます。逆に、機械翻訳システムがシステムダウンをしても、パソコン通信のホストシステムになんら影響を与えません。この場合は、翻訳依頼文書が収集されないので、機械翻訳システム用のメールボックスに翻訳依頼文書が滞積されます。勿論、機械翻訳システムが回復すると自動的にその滞積している翻訳依頼文書を収集し翻訳して、その翻訳結果を送信します。また、それぞれのシステムへの機能追加なども互いに独立して行えます。

翻訳システム（ソフトウェア）は富士通のＳシリーズワークステーションに搭載され、複数のワークステーションがＬＡＮで接続され機械翻訳システムを構成しています。１台のワークステーションに日英翻訳システムと英日翻訳システムの両方が搭載されています。また、その中の１台にはパソコン通信ネットワークとの通信システム（ソフトウェア）をも搭載しています。この通信システムが、複数のワークステーションの翻訳処理状況を管理していてメ

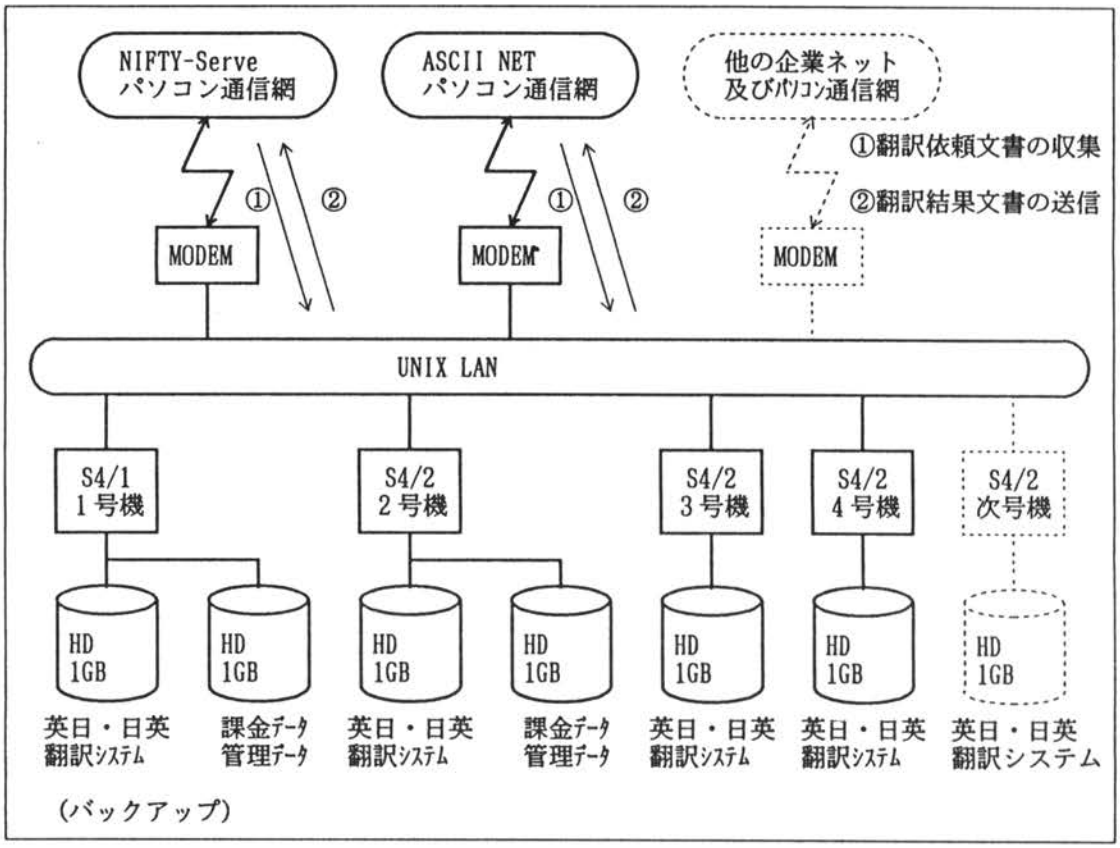


図 3 パソコン通信機械翻訳システム構成

ールボックスから収集した翻訳依頼文書を翻訳処理の終わったワークステーションに順次翻訳処理依頼をします。従って、1台のワークステーションは1つの日英翻訳処理又は英日翻訳処理を行います。各々のワークステーションはその翻訳処理を完了次第その翻訳結果を通信システムに返します。通信システムはこの翻訳結果を順次依頼者のメールボックスに送信します。

翻訳システムは複数のワークステーションに搭載されていることから、いずれかのワークステーションがダウンしても、そのワークステーションが動作しないだけで全体的な処理が止まることはありません。一方、通信システムを搭載したワークステーションには予備機が用意されており、現用がダウンした時直ちに切替えられるようになっています。また課金データ、管理データ等の重要なデータを2重化したハードディスクの両方に格納するようにしてあります。従って、それぞれのワークステーションの切離しや増設は、全体のシステムを停止することなくできます。

4．利用方法

（1）翻訳依頼文書の送信

最も基本的な翻訳依頼の送信方法をNIFTY-Serveを例に説明します。

つぎの例文を英日翻訳依頼をすることにします。

This is a computer.

That is a FAX.

専門辞書は「情報処理」と「電気」の2つを指定します。

翻訳結果は「対訳番号付」でもらいます。

＊NIFTY-Serveにアクセスするとつぎのメニューが表示されます。

→ プロンプト（>）の後ろに「GO ATLASMT」と入力して下さい。

```
NIFTY-Serve      TOP
 1. サービス案内・検索              2. 電子メール
 3. 掲示板                          4. CBシミュレーター
 5. フォーラム                      6. ニュース／スポーツ／天気予報
 7. 企業／経済／人物／行政          8. 科学／技術／翻訳
 9. 書籍／教育／就職               10. 趣味／旅行／生活
11. 占い／クイズ／ゲーム           12. ワープロ／コンピューター
13. ショッピング／ステーション     14. 海外データベース(INFOCUE)
15. コンピュサーブコーナー         16. ビギナーズコーナー
17. 他ネット接続サービス            E. 終了
>GO ATLASMT
```

図4　TOPメニュー画面

＊つぎのATLAS機械翻訳メニューが表示されます。

→ プロンプト（>）の後ろに「4」を入力して下さい。

```
ATLAS 機械翻訳     ATLASMT
 1. サービス案内                    2. オンラインマニュアル
 3. 日英翻訳（日本語->英語）(¥)    4. 英日翻訳（英語->日本語）(¥)
 5. Q&A 事例集                      6. お問い合わせメール
>4
```

図5　ATLAS機械翻訳メニュー画面

＊つぎの送信方法選択画面が表示されます。

→ プロンプト（：）の後ろに「1」を入力して下さい。

```
送信方法（1:メニュー選択  2:直接送信  E:中止）
：1
```

図6　送信方法選択画面

＊つぎの出力形式選択画面が表示されます。

→ プロンプト（：）の後ろに「1」を入力して下さい。

```
出力形式（1:対訳番号付  2:対訳番号無  3:訳文のみ  4:段落で出力  E:中止）
：1
```

図7　出力形式選択画面

＊つぎの専門用語辞書選択画面が表示されます。

```
専門辞書（1:使用する  2:使用しない  E:中止）
：1
```

図8　専門用語辞書選択画面

＊つぎの専門用語辞書指定画面が表示されます。

→ プロンプト（：）の後ろに「50,51」と入力して下さい。

```
専門辞書の選択
 10. 経済      (2万 3千語)    50. 情報処理   (3万語)
 30. 自動車    (1万 8千語)    51. 電気       (1万 7千語)
 31. 輸送      (2万 2千語)    52. 機械       (2万 8千語)
 32. プラント  (1万語)        53. 土木       (1万 5千語)
 40. 物理      (1万 5千語)    54. 工業化学   (1万 4千語)
 41. 数学      (3万 1千語)    55. 生化学     (1万 5千語)
 42. 生物      (8千語)        60. 医学       (23万1千語)
 43. 天文      (1万 3千語)    61. 薬学       (1万 8千語)

番号を選択してください(3つまで選択できます)
複数の場合は優先度順にカンマで区切ってください(例:10,30)
：50,51
```

図9　専門用語辞書指定画面

＊つぎの専門用語辞書確認画面が表示されます。

→ 希望する番号が正しく受け付けられていたらプロンプト（：）の後ろに「1」を入力して下さい。

```
確認  50,51 (1:OK  2:NG)
：1
```

図10　専門用語辞書確認画面

＊つぎの本文入力画面が表示されます。

→ 翻訳を依頼する文章を入力して下さい。入力が終了したら最後の行の次の行の先頭に/Eと入力して下さい。

```
本文 (300 行まで  終了は行頭で/E)
This is a computer.
That is a FAX.
/E
```

図11　本文入力画面

＊つぎの修正確認画面が表示されます。

→ 本文の修正が必要でなければ、プロンプト（：）の後ろに「2」を入力して下さい。

```
修正（1:修正する  2:しない）
：2
```

図12　修正確認画面

＊つぎの送信確認画面が表示されます。

→ プロンプト（：）の後ろに「1」を入力して下さい。

```
送信（1:送信する  2:しない）
：1
```

図13　送信確認画面

＊つぎのメッセージが画面に出力されて、翻訳依頼は完了です。

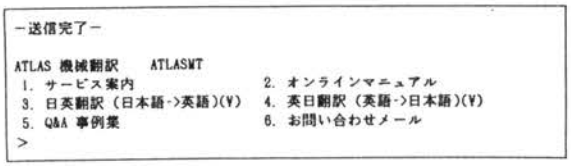


図14　送信完了画面

＊ここで一旦NIFTY-Serveとの接続を切ってしばらく待ちます。

（2）翻訳結果の受信

翻訳結果の受信は、電子メールを読む場合と全く同じです。

＊NIFTY-Serveにアクセスすると翻訳結果が着信している場合、TOPメニューに次のメッセージ表示されます。

→　プロンプト（>）の後ろに「MAIL」と入力して下さい。

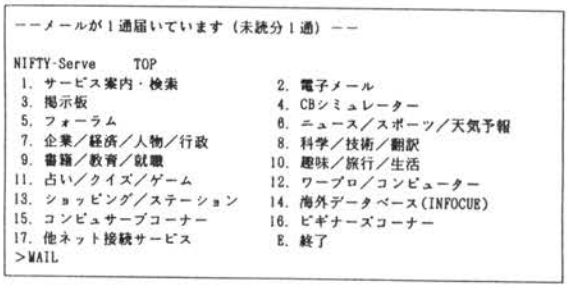


図15　TOPメニュー画面

＊つぎの電子メール画面が表示されます。

→　プロンプト（>）の後ろに「1」と入力して下さい。

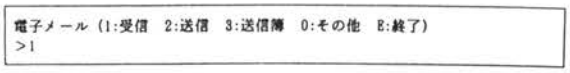


図16　電子メール画面

＊つぎの着信メール一覧画面が表示されます。

→　プロンプト（>）の後ろに「1」と入力して下さい。

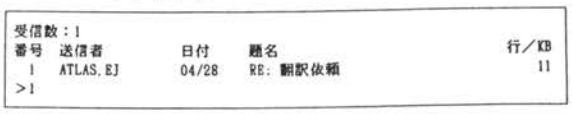


図17　着信メール一覧

＊つぎのように翻訳結果が表示されます。

→　翻訳結果を読み終わったらプロンプト（:）の後ろに「1」と入力して下さい。

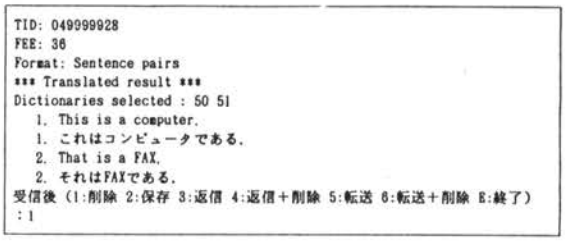


図18　翻訳結果表示画面

＊つぎの電子メール画面が表示されます。

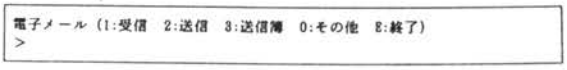


図16　電子メール画面

＊これで翻訳結果の受信が終わりましたのでNIFTY-Serveとの接続を切ります。

（3）名詞訳語の指定方法

専門辞書指定や翻訳結果の出力指定はメニューから選択することができますが、名詞訳語はメニューで指定することができません。翻訳依頼文書の先頭に、つぎのように記入して翻訳依頼をします。つぎの例の最初の2行が名詞訳語指定です。

翻訳依頼文書例

.dw/dog/ワンワン/
.dw/cat/ニャンコ/
She likes dogs.
He likes cats.

翻訳結果

1. She likes dogs.
1. 彼女はワンワンが好きである.
2. He likes cats.
2. 彼はニャンコが好きである.

この例は英日翻訳の場合ですが、日英翻訳の場合にはつぎのように指定します。

.dw/日本語(原語)/英語(訳語)/

5．利用状況

翻訳の現場から

アンカーマンの目

菊地　義明

編集者として訳文に目を通すようになってから、ほぼ２５年になる。となると、一目にはもうベテランの域に達したかに見えるが、それほど単純でないのがわれわれの世界である。

バブル経済時代は、その恩恵に浴した銀行、証券不動産関連のプロジェクトが主流で、国際会議も多く、量的には仕事の心配をする必要もなかった。でも、この頃は不況の波をもろに受け、業界全体が沈滞傾向にあるのは、何とも寂しいかぎりである。

現在、業界の英日翻訳料は、モノにもよるが400字１枚当たり￥3,000～3,500だ相場で、ここ10年、この数字はかわっていない。しかも、このようなご時世では 、さらにダンピングを強いられることも少なくないと聞く。そうなると、特に翻訳会社を通して仕事をしているプロ翻訳家の生活は、果たして成り立つのだろうか。

このような事態に私が一番懸念するのは、それに伴って「翻訳の質」も落ちることである。これまでにも、翻訳料が安いがために“ 量産 ”を強いられる翻訳者は多く、結果として誤訳、悪訳が量産されてきた事実がある。誤訳、悪訳、迷訳の語源は、そもそも翻訳者の力量とプロとしての 意識の欠如であり、そのアプローチの仕方にも問題がある。とはいえ翻訳者に対するクライアントの認識が深まり、翻訳者たる者がもっと優遇され、私の希望的観測では「翻訳者の質」も高まるものと思われる。

ここで編集作業の目安を示すと、私の場合は1時間あたり400字原稿1枚、1か月で約150～200枚のペースである。ここ15年は、このマイペースを崩さず、年間2,000枚の作業量を維持してきた。これを上まわると1字1句検討できなくなり、適当に“ 作文 ”する傾向のある翻訳者の術中におちいってしまうからだ。

翻訳のチェックポイントは、まず原文に忠実に訳出されているかどうかである。それに文体と誤字脱字、用字用語の統一、文全体のリズムと流れ、事実関係、表現表記法など多岐にわたる。

原文と訳文は、それぞれ付かず離れず、不即不離の関係が望ましいが、文化の違いでどうしても適語を見出せないこともある。というより、その連続である。その時は周囲のブレーンを勝手に動員して、赤っ恥を書くことは度々ある。また、それでも解決できないため、やけ酒を飲み、カラオケを歌ってストレスを解消したつもりでフトンにもぐり込んでから、突如としてその答えが夢のなかにもうろうと浮かび上がることもある。

（23頁より続き）

サービス開始以来、平均約百件／日の依頼があり総依頼数は10万件を超えていますが、景気の後退の影響を受け、昨年度は一昨年度に比較し１割程依頼件数が少なくなっています。この間の月毎、週毎、及び時間毎の依頼件数の動向を見ますと、経済界で二八枯れと言われますが月による有意差は見られません。１週間の内では、土曜日と日曜日それに休日が平日の約５～６割り程度になっています。１日の内では午前11時頃と午後３～５時頃に依頼件数の多くなっています。このようなことから、企業としての利用が主であると思われます。

また、問い合わせや専門辞書の使用状況より情報処理に関する翻訳が多くなっています。

日英翻訳と英日翻訳を比較しますと、１件当たりの翻訳量は英日翻訳が多く、翻訳に数時間かかるような大量の翻訳依頼もあります。これは、英文マニュアルの和訳に利用されているものと思われます。

翻訳とは、言うなれば語と語、句と句、文と文にそれぞれ対応する“言葉の選択”である。その選択が適切に行われれば申し分のない翻訳を期待できる。ただし選択を誤ると、読者泣かせの悪文となる。悪文は、往々にして主述関係がはっきりせず、非論理的で筋が通らず、句読点の打ち方もでたらめで、読みづらいことも多い。

読みやすく分かりやすい文章を作るには、以上の点に注意する一方、文の長さやつなぎ言葉、カタカナとひらがなと漢字のバランスにも目を向ける必要がある。抽象度の高い表現はできるだけその抽象のレベルを下げ、難解な用語もできるだけ避けて、日常一般の言葉をつかうようにするとよい。また、日本語として不自然な表現はしないようにしなければならない。これには「の」や「彼、彼女」の多用、能動態や受動態のつかい分け、名詞の単複の訳し分けも含まれる。これを無視すると、いわゆる翻訳調になる。

このほかにも、チェッカー泣かせのポイントがいくつかある。そのうちとくに重要なのは、事実関係の調査確認である。これには人名、地名、組織団体名、役職、肩書などの固有名詞のほか、数量や単位、年号、歴史的事実が含まれる。たとえば会社名ひとつにしても、株式会社が会社名の前にくか後にくるかは、アイデンティティの問題として重要である。しかも、この作業を怠ると、翻訳そのものの信頼性、信憑（しんぴょう）性が問われることにもなる。

英日の翻訳について言えば、翻訳とは英語を日本語に置き換えることすぎない。しかし、置き換えるにあたっては、上述したさまざまな要素を考慮し、書く要素を総合的に組み合わせて行かなければならない翻訳の楽しみと苦しみは、まさにそのプロセスにある。このプロセスを大切にして、また肝に銘じて翻訳に取り組めば、翻訳の質は数段、向上するものと思われる。

最近は、「超訳」なるものを売り物にしている出版社があるという。しかし、きちんと翻訳できれば超訳など必要ないし、そういうものがあること自体おかしい、というのが私の持論である。最後にこのような超訳に目を光らせるのもチェッカーの役目であり、アンカーマンの仕事でもある。

著者の略歴

菊地　義明（きくち・よしあき）

翻訳家、翻訳監修・編集者。
（株）アイ・エム・シー代表　取締役
昭和17年1月、仙台生まれ。東京外国語大学インドパーキスターン科ヒンディ語部出身
（株）現代文化研究所・海外情報担当主任研究員を経て、昭和54年独立
主にモービル石油（株）のオピニオンリーダー誌「モービル文庫」の制作にあたる一方、サイマル・インターナショナル翻訳制作部の翻訳監修・編集を手がける。

このような利用状況より、機械翻訳サービスを新しい道具として、利用者個々が利用方法を工夫して使ってわれているものと思われます。

6．おわりに

パソコン通信は利用しているが、パソコン通信に「機械翻訳サービス」があることを知らない人が大勢います。ASCII-Netでは、サービス開始当初よりTOPメニューに「翻訳」が表示されましたが、NIFTY-Serveでは本年4月よりTOPメニューに表示されるようになりました。このようにまず「パソコン通信機械翻訳サービス」を知って頂くようにする一方、翻訳品質の向上に向けて専門辞書の拡充を図っていく所存であります。

（株）富士通ドキュメントサービス
第二翻訳部機械翻訳センター課
（石橋　正章）

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

NECでは1990年12月より、弊社のパソコン通信サービス「PC－VAN」において弊社の機械翻訳システムPIVOTを使用した「IBS機械翻訳サービス」を提供してまいりました。

当初は、「エディット（前編集）付き」のみでサービスしておりましたが、1992年12月に「機械翻訳のみ」および「リライト」を追加、および専門用語辞書の選択が可能となりました。さらに、1993年6月には、FAXで送信された原稿をOCRで読み取り、電子化された原稿を機械翻訳にかける「OCR機械翻訳サービス」を追加するなど、機能的・性能的にサービスの向上を行ってまいりました。

ここでは、PC－VANにおける機械翻訳サービスの内容及び利用方法をご紹介します。

IBS機械翻訳サービスのご紹介

IBS機械翻訳サービスには、英文和訳／和文英訳それぞれに翻訳レベルの異なる3種類の翻訳サービスとOCR読み込みサービスがあります。

1. 機械翻訳サービス

機械翻訳サービスは、依頼者（パソコン通信の利用者）が、電子メールにて原稿を受付BOXに送信し、その原稿を機械翻訳にかけ、電子メールにて依頼者に返送するサービスです。
このサービスには、和文英訳および英文和訳があり、下記の3種類の翻訳レベルが　あり、それぞれ特長を持っています。

1)機械翻訳のみ

利用者が依頼した原稿にいっさい手を加えず、そのまま機械翻訳にかけ、出力結果に関してもそのまま返信するサービスです。

◎料金が安価

翻訳結果に対し1文字あたりの安価な料金設定のため利用しやすい。

◎処理が早い／いつでも利用が可能

人手による書き直しがないため翻訳処理が早く（A4　4枚程度で約20分程度）、年中無休で利用が可能です。

×依頼原稿の文書形態により翻訳品質が不安定

依頼原稿の文書形態により、機械翻訳に適していない場合があります。機械翻訳が苦手な文書形態で依頼された場合、翻訳品質が著しく低下する場合があります。
依頼者自身でプリエディットを行うことにより、翻訳結果が向上します。

2)エディット付き

依頼原稿を翻訳にかける前に、人手により翻訳結果の品質が良くなるよう、IBS社の機械翻訳サービスのノウハウを生かした修正を加えます。ただし、翻訳結果に関してはいっさい手を加えません。

◎翻訳品質が向上する

翻訳品質が良くなるよう、原文を修正するため、翻訳後の結果が向上します。

・翻訳不可能部分が発生する

翻訳結果をそのまま返信するため、辞書に未登録のため変換不可能な単語や文法が発生します。

・処理に時間がかかる

人手による作業が介在するため、依頼を受けてから結果を返信するまで約1日かかります

3)リライト付き（英文和訳、和文英訳）

エディット付きに加えて、翻訳結果についてもネイティブスピーカによるリライトを行い、自然な文章に書き直して返信します。

◎翻訳結果が自然な文章で返信される

最終的に人手で書き直すので、文法誤りや翻

訳不可能な単語が極力少なくなる。

・処理に時間がかかる

人手による作業が介在するため、依頼を受けてから結果を返信するまで約１日かかる。

4)共　通

ＩＢＳ機械翻訳サービスでは、基本辞書（日英９万語　英日７万語）の他に、３１分野６０万語以上の専門用語辞書を追加指定できます。また、自分専用の辞書（ユーザ辞書）を作成し、保管ができるので、次回の翻訳依頼時にそれを指定することも可能です。

このユーザ辞書とは、専門用語辞書に登録されていないような固有名詞やある分野のみで使用される専門用語等の単語と訳語を、自分専用の辞書として持つことができ、ＰＩＶＯＴで翻訳するときにこの辞書名を指定することにより、専門用語辞書と同様に使用することができます。

ユーザ辞書の作成方法は、大変簡単で、単語と対訳を２つ以上のスペースで区切った対訳表を作成し、ＩＢＳへ送信するだけで、あとはＩＢＳ側でユーザ辞書を作成します。また、この辞書の管理は、ユーザＩＤと自由につけることのできる辞書名とで管理されますので、一人で色々な分野の辞書を持つことが可能です。

2. ＯＣＲ読み込みサービス

印刷されている文書をＦＡＸで送信し、これをＯＣＲ装置でワープロ文書化するサービスです。

パソコン通信の機械翻訳では依頼原稿を電子メールで送信するため、原稿はワープロ文書でなければなりませんが、このサービスを利用すると、プリント文書を機械翻訳で処理することができます。

ＩＢＳ機械翻訳サービスの利用法

ＩＢＳ機械翻訳サービスを利用するには、パソコン通信を始めなければなりません。

1. 必要な機材

パソコン通信を始めるためには、下記の５点が必です。

1)パソコン

通信機能のついたパソコン／ワープロ

2)モデム

パソコンと電話回線を接続するための装置

3)通信ソフト

4)電話回線

5)パソコン通信（ＰＣ－ＶＡＮ）への加入

ＩＢＳ機械翻訳サービスの利用方法［図１］

ここでは、ＩＢＳ機械翻訳サービスにて「和文英訳」の「機械翻訳のみ」を利用する方法をご紹介します。

英文和訳やエディット付き、リライト付きも同様の手順でご利用いただけます。

1)ＩＢＳ機械翻訳サービスを利用するためには、専用のメニューを呼び出します。

Ｊ　ＩＢＳ<リターン>

このコマンド入力により、ＩＢＳ機械翻訳サービスのメインメニューが表示されます。ここで２番の「ＩＢＳ機械翻訳サービス（¥）」を選択します。

2)料金案内に続いて、翻訳の種類を選択します。

２番の「和文英訳」を選択します。

3)翻訳レベルを選択します。

１番の「機械翻訳」を選択します。

4)使用辞書の選択

翻訳に使用する”専門用語辞書”やユーザ辞書を３つまで選択します。（，で区切ります）

5)翻訳依頼原稿の送信

使用する辞書の確認後、翻訳を行う原稿を送信します。

（送信には、無手順とファイル転送プロトコルのＸＭＯＤＥＭおよびＱｕｉｃｋ－ＶＡＮが利用できます。）

6)翻訳結果の取り出し

依頼された翻訳原稿をＩＢＳにて翻訳処理を行い、翻訳結果が電子メールにて返信されます。

ＩＢＳ機械翻訳サービスの上手な使い方
［和文英訳］

ＩＢＳ機械翻訳サービス（特に機械翻訳のみ）を上手に利用するために、下記の点に注意して文書を作成します。

1. 文の切れ目を句点（。）で切る。
 1)「？」や「！」で終了している文は、「。」に置き換えます。
 2)引用の”「」”を使用した場合には、”」”の外に”。”を置くようにします。

2. 自立語は漢字、付属語はひらがなで書く。
 1)動詞、名詞、形容詞、副詞は、通常漢字で書くものは漢字で書きます。
 2)それ以外の通常漢字で書くもの以外はひらがなで書きます。
 例）われわれ　→　我々
 ～出来る　→　～できる。

3. 数字はなるべく算用数字で書く。

4. 長い文は分割する。
 1)６０文字を越える文はなるべく分割する。
 2)述語が３つ以上ある文は、述語や接続詞の前後で分割する。
 3)項目の羅列がある場合は、箇条書にする。

5. 不要な言い回し／曖昧な表現は取り除く
 1)尊敬や謙譲表現は使わない。
 2)「こと」、「もの」、「ふう」、「よう」、「ほど」が含まれている言い回しには曖昧な表現が多いので極力避ける。
 3)「で」、「だ」、「ある」、「する」、「なる」は色々な意味で用いられるので、なるべくそれぞれの意味に限定された用語で置き換える。

6. 口語文や会話文、詩歌、俳句、短歌などは使用しない。

ＩＢＳ機械翻訳サービスの利用料金

ＩＢＳ機械翻訳サービスを利用する場合、下記の料金が必要となります。

1. ＰＣ－ＶＡＮ通信料金
 ＰＣ－ＶＡＮサービスに接続している間発生する料金
 ２０円／３分（３００－２４００ｂｐｓ）
 ２０円／１分（９６００－１４４００ｂｐｓ）

2. 機械翻訳料金
 機械翻訳の結果に対して利用料金が計算されます
 1)英文和訳（日本語に対して計算　1枚=360文

機械翻訳のみ	１文字	1.5円
エディット付き	１枚	1,200円
リライト付き	１枚	3,600円

 2)和文英訳（英語に対して計算　1枚=180語）

機械翻訳のみ	１文字	0.6円
エディット付き	１枚	1,500円
リライト付き	１枚	4,500円

 ※エディット付き、リライト付きには１件あ
 最低料金　3,000円がかかります。

 3)ユーザ辞書

月額保管料	3,000語毎	1,000円

 ※辞書毎に計算します。

3. ＯＣＲ読み込みサービス
 1)和文OCR(1文字あたり)

スペルチェック	あり	なし
200DPI	0.4円	0.8円
400DPI	0.4円	0.6円

 2)英文OCR(1語あたり)

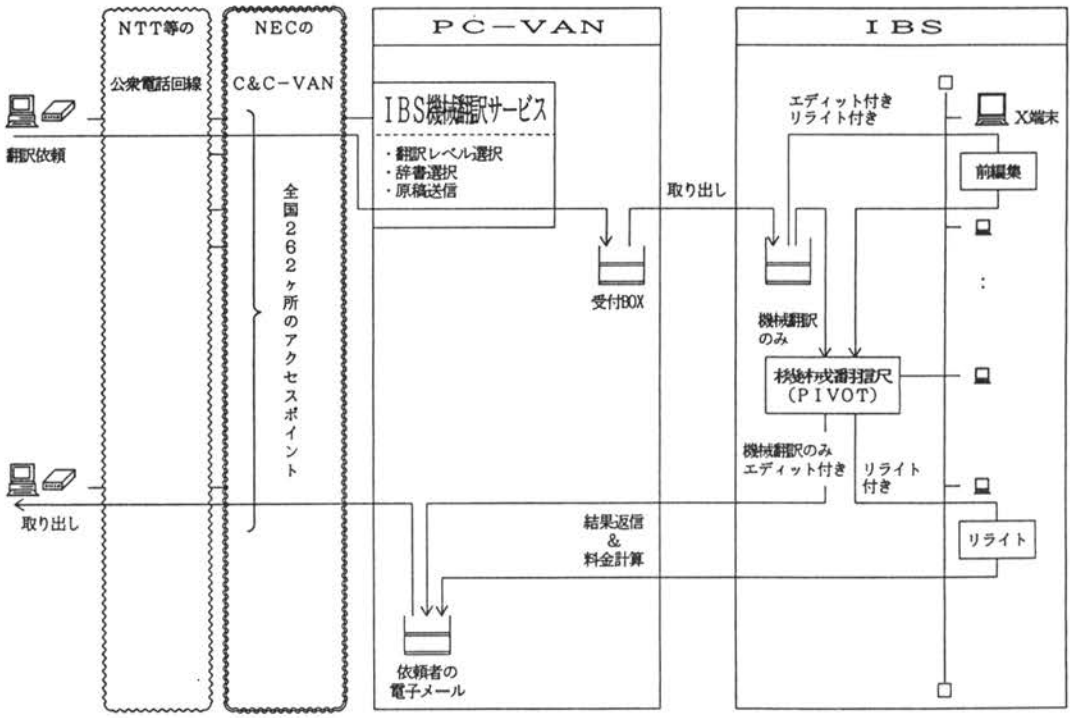
スペルチェック	あり	なし
200DPI	2.9円	1.1円
400DPI	2.1円	1.1円

 ※最低料金　300円がかかります。

※200DPI　G3モード
　400DPI　G4モード

問い合せ先：
日本電気株式会社　ＰＣ－ＶＡＮ事務局
〒１０８－０１ 東京都港区芝五丁目７番１号
電　話　　０３－３４５４－６９０９

システムイメージ図



アジアのイベント

（詳細問合せ先…ＣＩＣＣ 03-3459-0941）

SECOMM'94 MALAYSIA

月　日　1994年10月4日～6日
テーマ　NETWORK-ENABLING THE FUTURE
主　催　MALAYSIA NATIONAL COMPUTER CONFEDERATIONN (MNCC)
連絡先　Tel;603-775-1576　FAX;603-774-7026

SIGAPORE INFORMATICS '94

月　日　1994年10月13日～16日
場　所　World Trade Center, Singapore
主　催　Singapore Federation of The Computer Industry

CHINA INTERNATIONAL CONSUMER ELECTRONICS EXHIBITION

月　日　1994年11月8日～13日
場　所　Shanghai Exhibition Center, Shanghai, P, R, CHINA

日英機械翻訳システム
ブラビスJ/E ver5.1

エムティラボ株式会社

【開発の狙い】

日本があらゆる分野で世界でトップレベルの領域に達したことにより日本語を英語に翻訳する量が増大してきました。このような状況において、翻訳作業をコンピュータ化しコストや労力を省こうという発想から当ソフトは開発されました。ブラビスJ/E ver5.1は１９８４年に世界で初めて発売された日英翻訳システムの最新版であり、実績・品質ともに他社の追随を許しておりません。

【翻訳機能】

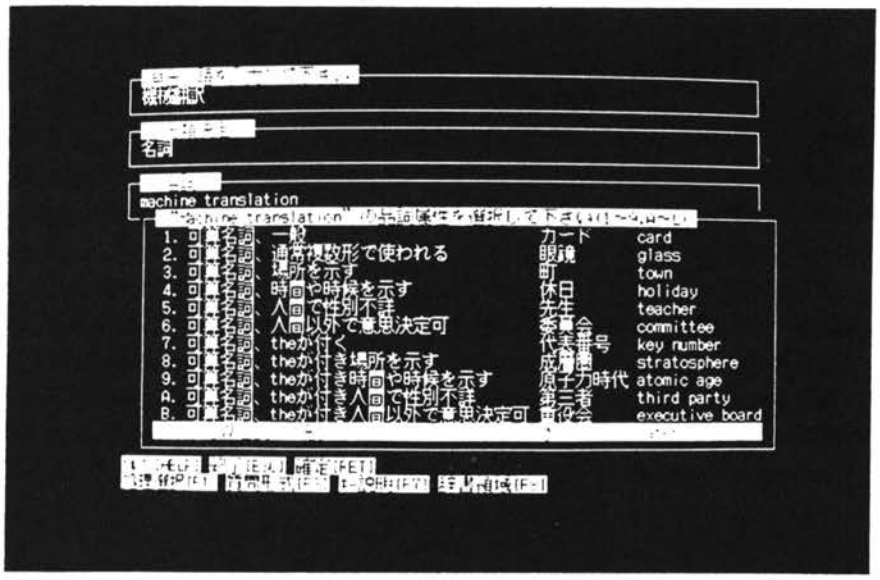
即時と一括の翻訳機能があります。一括翻訳では、搭載されているワープロを使って日本文を作成した後翻訳します。その際、全文翻訳・範囲翻訳・１文翻訳のどれかを選択することができます。一括翻訳ではファイルの長さに制限がなく、同時に１６ファイルまでの指定が可能です。またテキスト形式のファイルであれば他のワープロソフトで作った文章も翻訳可能です。　１～２文程度の短い文章の翻訳には即時翻訳が便利です。日本語を入力後、即座に英文が出力されるので事前にワープロで文章を作る必要がありません。一度に即時翻訳できる日本語は４００文字までです。

【辞書作成機能】

基本辞書として６万語が搭載され、オプションとして３１分野の専門辞書が用意されています。さらにユーザーが自由に４万語まで登録することが出来ます。登録はワープロの単語登録のように簡単に行えます。機械翻訳において大変重要となる単語の属性についても、メニューの中から選択するだけで済むようになっています。

属性とは名詞なら固有名詞で人名であるとか一般名詞で数えられないとかを指します。これらを正確に登録することにより、関係代名詞や前置詞などの用法を判断することが出来ます。翻訳ソフトウェアを使いこなすためには辞書登録が不可欠であると言えます。

また単語だけでなく、複数の単語からなる「信頼



できる友」のような表現を登録することも可能です。よく使う表現を登録しておけば、自分の欲しい英文表現がいつでも得られます。

【日本文チェッカー機能】

より良い英文を引き出すためには入力された日本文を簡潔明瞭にする必要があります。日本文チェッカーは、ある程度この作業を自動的に行います。また、日本文チェッカーは文中に含まれる未登録語の検出も行います。それらを事前に辞書に登録することで、さらに良い英文が引き出せます。

【ワープロ機能】

使いやすいワープロを内蔵しています。日本語の入力には、使い慣れたＦＥＰを自由に組み込めます。翻訳や辞書登録、辞書検索はすべてこのワープロ画面から行えるます。翻訳ソフトを意識することなく簡単に操作できます。

【翻訳例】

原　文：

今日本では暖冬が珍しくない。これは単に一時的かもしれないが地球温暖化による現象かもしれない。

しかし明らかに二酸化炭素やフロンは上空に溜まっている。最近ＮＡＳＡの調査によれば、オゾン層の破壊は南極だけでなく北半球一帯の成層圏に及んでいる。オゾン層は地球を紫外線から保護している。

オゾン層破壊の原因となっているフロンは宇宙で１００年以上生き延びる。今その放出を止めたとしてもかつての正常な水準に戻るためには１００年かかる。その事態は非常に深刻である。

翻訳文：

A warm winter is not unusual in Japan now. It may be a phenomenon by global warming although this may be only temporary.

However, obviously carbon dioxide and chloro-fluoro carbon are accumulated in the air. Recently according to a survey of NASA, the destruction of an ozone layer is reaching in the stratosphere of all over a northern hemisphere in addition to the South Pole.

An ozone layer is protecting the earth from ultraviolet rays. Chloro-fluoro carbon that is a cause of ozone layer destruction survives for more than 100 years in the universe.

It takes 100 years to return in a previous and normal standard granted that the discharge is stopped now.

The situation is very serious.

【その他】

◆OS:日本語DOS（DOS/Vも可）
◆HDD:20MBの空きが必要（基本システムの場合）
◆価格:本体 ¥190,000・オプション辞書(３１分野)¥12,800 - ¥45,000
◆サポート:専用電話・ＦＡＸを開設(無料)

【お問い合わせ先】

株式会社 パシフィックアイ
〒532　大阪市淀川区西島6-9-20
新大阪GHﾋﾞﾙ7F
◆PHONE(06)885-1871
◆FAX(06)885-0720

PENSÉE for Windows

沖電気工業（株）・沖ソフトウェア（株）・（株）オージス総研

PENSÉE for Windowsは、
(1)お手持ちのパソコンで
(2)使いなれたワープロで、
(3)手軽に使えることを
目的として開発された日英／英日翻訳支援ソフトウエアです。

ワークステーション版で高い翻訳品質との評価をいただきましたPENSÉEをWindows上に移植し、Windows の操作環境で本格的な翻訳ソフトを実現しました。また、Wordからの翻訳実行やEXCELでの辞書編集も可能です。

＜ 特 長 ＞

● Windowsでの簡単操作

専用エディタは、Windowsの操作性を活かし、メニューバーによる編集、翻訳、ファイルへの保存など、初めての方にも操作は非常に簡単です。

● PENSEEの高品質をそのまま利用

従来より高い翻訳品質との評価を頂いておりますPENSEEをそのまま利用しておりますので、高品質な訳文が得られます。

● 訳質に差がつくシステム辞書の拡張

従来のシステム辞書に加え、今回ビジネス分野と経済分野の専門用語辞書を標準で装備しており、語数は合計で日英約15万語、英日約13万語です。他に7分野約10万語の専門用語辞書を用意しております。

● 柔軟なユーザ辞書指定

ユーザ辞書と専門用語辞書を優先順位をつけて最大16個まで同時に指定できます。

● パソコンでワークステーション並みの高速翻訳

1時間に平均約15,000語（最大50,000語）の速度で翻訳しますので、1ページ（A4）あたり数十秒で翻訳が完了します。

● 簡単辞書登録のための未知語検索

未知語検索結果をCSV形式に出力することができます。CSV形式を扱えるEXCELマクロを使用することにより大量の単語編集をスピーディーにかつ簡単に行なうことができます。

● Wordとの連動

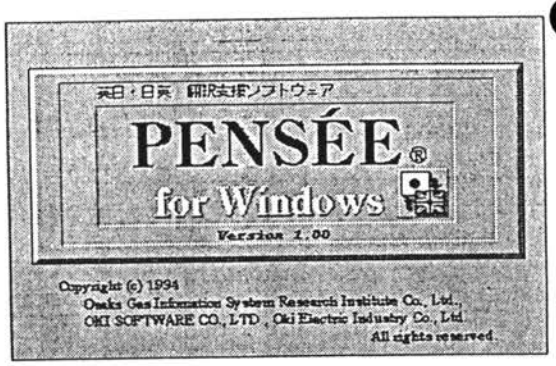
Wordのマクロ機能を利用して直接翻訳を実行することができます。テキスト以外にWordの.DOCの文章を翻訳することができます。

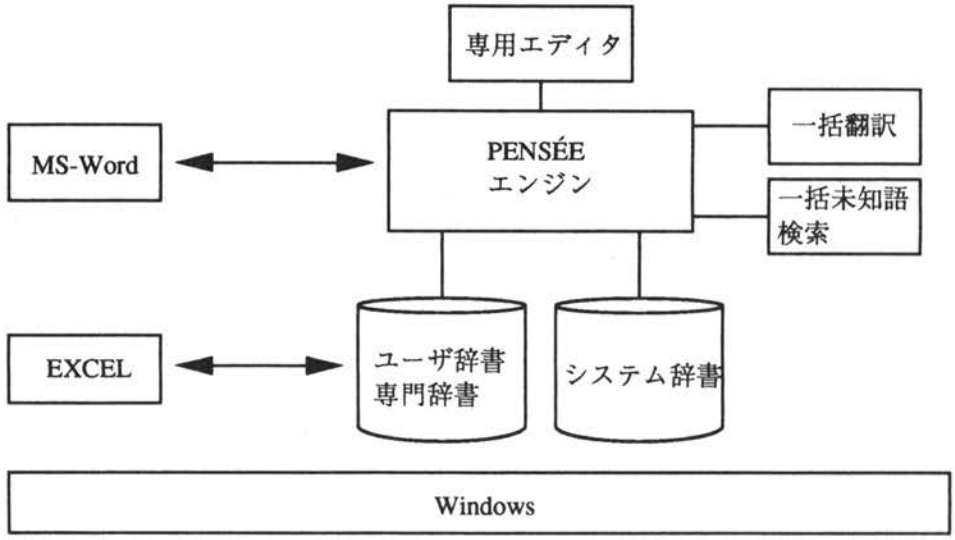
● Windows上のアプリケーションとの連携

Windows上の各種アプリケーション(OCR,DTPなど)とクリップボードを介して原文、訳文の受け渡しをすることができます。

● 一括翻訳、一括未知語検索

一括翻訳、一括未知語検索のバッチ機能を利用して、最大16個のテキストをバックグラウンドで実行させておくことができます。その間、他の作業を行うことができます。





＜ 仕 様 ＞

対象言語：日→英／英→日
ハードウエア：
　　日本語Windows3.1が稼働する機種
メモリ：8MB以上
ディスク：40MB以上（片方向）
OS：Windows 3.1
システム辞書+専門用語辞書(ビジネス、経済分野)
　日英：約15万語
　英日：約13万語
専門用語辞書（オプション）：
　日英：7分野、約10万語
　英日：7分野、約10万語
　（電気、情報通信、医療、機械、化学など）
ユーザ辞書：
　　語数無制限（ディスク範囲内）
　　最大16個まで同時使用可能
処理速度：約15,000語／時間（平均）
　　（機種によって異なります）
翻訳オプション：
　日英：文体（標準／命令）、文区切り文字指定
　英日：文体（標準／敬体）、固有名詞訳出
　　（英語のまま/訳語のみ/訳語（英語）/
　　英語（訳語）、文区切り文字指定
他ソフトとの連携：
　　Word 5.0（マクロ機能による翻訳実行）
　　EXCEL 4.0（マクロ機能による辞書の編集）

＜ 価 格 ＞

PENSÉE for Windows
　　¥198,000（片方向）
専門用語辞書(7分野パック)¥ 48,000（片方向）

※ PENSÉE は、沖電気工業(株) 大阪ガス(株)
　　（株）オージス総研の登録商標です。
※ MS及びMicrosoftは、米国マイクロソフト社の
　　登録商標です。

＜問い合わせ先＞

沖ソフトウェア株式会社営業部 営業第2グループ
TEL. 03-3454-7831　FAX. 03-3798-7088

株式会社オージス総研アドバンストシステム本部
アドバンスト営業部
TEL. 03-3779-2541　FAX. 03-3779-2543

英日／日英機械翻訳支援システム
ＡＳＴＲＡＮＳＡＣ （エイエストランザック）

株式会社 東　芝

東芝は古くから自然言語処理技術を研究しており、その成果の１つとしてＡＳＴＲＡＮＳＡＣを1986年よりＡＳシリーズ(*1)上で発売してきました。

さらに、1993年よりSun WS へのソフトウェアのみの販売をしてきました。

また、1992年にはＡＳＴＲＡＮＳＡＣの翻訳精度、優れたインタフェースが認められ、（財）ソフトウェア情報センターより「'92 Software Products of the Year 」の「知識情報処理部門」を受賞しました。

ＡＳＴＲＡＮＳＡＣにはオプションソフトウェアとして、ＯＣＲ機能も装備しております。英文を読取る「ＡＳ－Ｒｅａｄｅｒ」、日本語を読取る「日本語ＡＳ－Ｒｅａｄｅｒ」を利用することで、原稿入力の時間を短縮することができます。

この度、ＡＳＴＲＡＮＳＡＣのV4.2をリリース、それに伴いＯＣＲ機能もバージョンアップを行い、さらに翻訳の品質向上、機能の充実をはかりましたので特徴、最新バージョンをご紹介いたします。

＜ＡＳＴＲＡＮＳＡＣの特長＞

・翻訳スピードは10人力
1時間あたりの翻訳語数はなんと約 60,000語(*2)です。
英文なら A4・1 枚(250words)、和文なら400 字詰め原稿用紙１枚を１分程度で高速処理します。しかもワークステーションの性能向上にともなって翻訳スピードがさらにアップします。

・10万の文法ルールと意味トランスファ方式
6万語の基本辞書や10万の構文ルールの採用、高度な学習機能など東芝の自然言語処理技術を結集しました。また、意味トランスファ方式ですから、訳文の正確性が一層向上します。

・豊富な分野に対応できる専門辞書とユーザ辞書
情報処理、機械、政治経済、化学などの分野の専門用語を提供しています。またユーザ辞書には最大20万語を登録でき、複数のユーザ辞書を作成することができます。

・ユーザカスタマイズ機能
ユーザ特有の文書をより高精度かつ均質に作成するために、辞書の選択や訳語の表現方法などの翻訳環境を画面の指示に従って簡単に設定できます
また、設定した翻訳環境を保存して、文書の種類に応じて使い分けることもできます。この機能とユーザ辞書をあわせて活用すれば、よりユーザ事情に適した翻訳ができます。

・原稿をダイレクトに読み込めるOCR 機能
業界最高水準の東芝のOCR 技術を応用しました。印刷された英語の文書も日本語の文書も手元のスキャナから簡単に入力できます。英文で99.7% 、和文で99.5% という極めて高い認識率を実現しました。さらに、自動レイアウト解析により複雑なレイアウトの原稿や縦書きの原稿も認識できます

・優れたマンマシン・インタフェース
業界標準GUI であるOPEN LOOK に準拠していますのでコンピュータが初めての方でもアイコンとマウス、そしてマルチウインドウを使って効率的に翻訳作業を行えます。

そのうえ英語から日本語、日本語から英語という翻訳の方向をボタン１つで切り替えることも可能です。

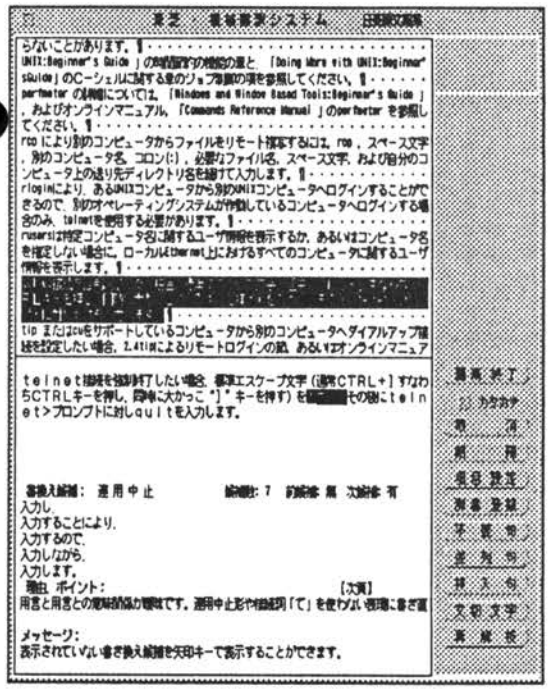
＜最新バージョンのご紹介＞
★ ＡＳＴＲＡＮＳＡＣ（V4.2）

主なバージョンアップの内容：

・日英翻訳前編集支援機能

機械翻訳に不可欠なプレエディットを半自動化しました。従来マニュアルで説明していたプレエディットのポイント項目の一部を自動的に検出します。書換え候補、ガイダンスメッセージをユーザに提示します。検出項目は、長文、合成名詞、ひらがな語、辞書未登録語、冗長表現、接続表現、否定表現、受身表現、連体中止表現、不適表現辞書です。

この機能の追加により、プレエディットの時間短縮及び翻訳結果の品質の向上がはかれます。



・英日翻訳辞書編集機能

ユーザ辞書に登録されている単語や新規に登録する単語の見出し語、訳語に対する複雑な属性設定、優先度の変更をウィンドウ上で容易に行うことができます。

この機能の追加により、複雑な属性設定などができるので、訳し分けができるようになり、翻訳品質の向上がはかれます。

・DOC Maker3(*3) 文書翻訳機能

DTPツールDocMaker3 のMIF 文書を翻訳できます。従来のDocMakerより一段と高機能なDTP ツールDocMaker3 で作成した文書を翻訳し、翻訳結果をMIF に保存し、訳文をDocMaker3 で文書を呼び出すと原文のレイアウトを再現できます。もちろんDocMakerの文書にも対応しています。

この機能の追加により、翻訳後の編集の手間が省けます。

・ファイルマネージャーインターフェース機能

OpenWindows のDeskSet のファイルマネージャからのDrag&Drop により文書登録できます。

この機能の追加により、簡単に翻訳の文書の登録ができます。

★ＡＳ－Ｒｅａｄｅｒ（V2.2）／
日本語ＡＳ－Ｒｅａｄｅｒ（V1.2）

主なバージョンアップの内容：

・イメージ取込み機能

従来はテキスト部分のみ認識していましたが、イメージも取込んで、テキスト・イメージ混在のDOCMakerまたはDOCMaker3 文書として出力できるようになりました。

この機能の追加により、翻訳後の編集の手間が省けます。

・イメージズーミング機能

領域指定の段階で、読取り具合がわかるようにイメージを拡大して表示できます。

この機能の追加により、読取り具合が認識する前にわかるので、スキャナの濃度設定の調整が容易に行えます。

(*1) 東芝のEWS(SunWS のOEM)

(*2) AS4080 /SPARCstation 10使用時

(*3) DocMaker およびDocMaker3 はFrameMaker(FrameMaker はFrame Technology Corporationの登録商標です）のMIF と互換性があります。

お問い合わせ先：

（株）東　芝
コンピュータ事業統括部
(03)3457-2725（ダイヤルイン）

技術早分かり

「統計手法を利用した機械翻訳」

NTT 情報通信網研究所　永田昌明

1 はじめに

皆さんは ``Back to the Future'' という映画を御覧になりましたか？　主人公の高校生マーティの友人で、科学者のドクは、30年前のある朝、つまずいて頭を強く打った瞬間に、タイムマシンのエンジンを実現する画期的なアイディアを思いつきました。しかし、それには雷に匹敵するような非常に大きなエネルギーが必要で、当時の技術ではとても実現不可能でした。

ところが、原子力の進歩がドクのアイディアを実現可能にします。ドクは「原子爆弾を作ってやるから」と言って、リビア人ゲリラをだまし、プルトニウムを手に入れます。そして遂に、あのタイムマシン「デロリアン」を完成させ、マーティとともにショッピングモールの駐車場で（この辺りが如何にもアメリカらしい）タイムトラベルの実験を行ないます。

そこへ、だまされたことに気づいたリビア人が彼らに報復するために武装して襲撃してきます。追われる身となったマーティは、デロリアンに乗り込み、必死の思いでアクセルを思いきり踏み込みます。すると ...。

続きは映画でお楽しみ頂くとして、ここで皆さんに注目して頂きたいのは、昔は実現不可能だったドクのタイムマシンのアイディアが、要素技術（原子力）の進歩により実現可能になったことです。

Warren Weaver は、1949年に、統計的手法による機械翻訳を提案しました。彼は Shannon とともに、現在の情報理論の基礎を築いた人物です。1950 年代から 60 年代にかけては、 empiricism と呼ばれる大規模な言語データを統計的に分析する方法論が計算言語学の主流を占めていました。しかし、当時は利用できる言語データの量が少なかった上に、計算機は高価で処理能力も小さく、このアプローチはやがて行き詰まってしまいます。

そこへ、あの Chomsky が登場します。彼は人間の言語能力を近似するモデルを作成することを目標とし、Wh-移動などの長距離依存が扱えない Shannon の N-gram モデルの代案として独自の文法理論を提案しました。彼の理論のように、言語現象の基底に存在する原理を探求する方法論はrationalism と呼ばれます。そして、この方法論が、その後今日まで計算言語学の主流を占めてきました。

しかし、Chomsky 流のアプローチでは、N-gram モデルが捉えていたような人間の言語運用の情報がうまく扱えません。このことは、機械翻訳システムにおいて、一つの文に対して非常に多くの解釈が発生し、その大部分は日常的には使用しない解釈であることからも明らかです。

時は流れ、1980年代に入ると、計算機の性能が急激に進歩し、多くの文書が電子化され始めます。ひと昔前の大型計算機以上の処理能力を持つワークステーションを研究者が一人で占有し、一年分の新聞や数十万語の辞書が CD-ROM で手に入る時代になりました。そう、ドクにとってのプルトニウムです。こういう歴史的変遷を経て、現在の自然言語処理では「統計に基づく手法」あるいは「コーパスに基づく手法」が再び主流の座を取り戻しつつあります。いわば「Shannon の逆襲」ですね。

自然言語に対する統計的なアプローチは、まず音声認識と形態素解析の分野で成功を収め、現在は、構文解析や語彙意味論の研究が物凄い勢いで進んでいます。また最近では機械翻訳への適用も盛んです。

統計的な手法に基づく機械翻訳の基本的な枠組みは、アメリカの IBM の T. J. Watson 研究所の音声認識グループにいた Peter F. Brown によって形成されました Brown-et-al90, Brown-et-al, 93 。

以下では、Brown et al., 93に基づいて、フランス語から英語への翻訳を例として、この統計的な機械翻訳について解説します。

2 言語翻訳の確率・統計的モデル

一般に, あるフランス語の文 f に対する翻訳としては、様々な英語の文が考えられ、そのどれを選ぶ

かは好みの問題です。統計的機械翻訳では、全ての英語の文 e が f の翻訳になりうると考え、全ての文の対(e , f)に対して確率 Pr(e|f) を割り当てます。ベイズの法則によれば次の関係が成り立ち、

$$Pr(\mathbf{e}|\mathbf{f}) = \frac{Pr(\mathbf{e})Pr(\mathbf{f}|\mathbf{e})}{Pr(\mathbf{f})} \qquad (1)$$

分母は e に依存しないので、Pr(e) Pr(f|e)を最大にする e を選べば、フランス語を英語に翻訳する際の誤りを最小にできます。

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} Pr(\mathbf{e})Pr(\mathbf{f}|\mathbf{e}) \qquad (2)$$

一般に、英語文の事前確率 Pr(e) を言語モデル確率(language model probability)、英語文が与えられた時のフランス語文の条件付き確率Pr(f|e) を翻訳モデル確率 (translation model probability)と呼びます。式(eq:ml-translation)より、統計的機械翻訳を実現するには、言語モデル確率の推定、翻訳モデル確率の推定、最尤英語単語列の探索という三つの問題を解かなければならないことがわかります。

ここで「フランス語を英語に翻訳するならPr(e|f)を推定すれば済むではないか」という疑問が生じるかもしれません。しかし、実験的には Pr(f|e)よりも式(2)を使った方が翻訳の質が良いのです。翻訳モデルPr(f|e)は、フランス語の単語列に対して語彙的にも配列的にもほぼ対応する英語の単語列に対して大きな確率を与えるのに対して、言語モデルPr(e)はフランス語との対応は関係なく文法的に正しい英語の単語列に対して大きな確率を与えるので、式(2)を使うと両者が協調的に働き、良い結果を生むようです。

3 二言語コーパスの対応付け

言語モデル Pr(e) は、音声認識で使用するものと同じであり、単言語コーパスから推定できます。しかし、翻訳モデル Pr(f|e)を統計的に推定するためには、大量の翻訳例が必要です。

同じ内容を二つの言語で記述したテキストを並行テキスト (parallel text)、または二言語コーパス(bilingual corpus)と呼びます。並行テキストは、機械翻訳だけではなく、二言語辞書編集(bilingual lexicography) などにとって非常に有益な言語資源なので、近年、文レベルや単語レベルの対応付けを並行テキストに対して自動的に付与する方法の研究が非常に盛んです。並行テキストとしては、カナダの国会議事録である Hansards が有名です。 この Hansard コーパスは、機械可読な1億語以上の英仏並行テキストとして、統計的機械翻訳の研究に広く用いられています。

翻訳例を得るためには、まず、並行テキストを文レベルで対応付ける必要があります。これを文対応付け(sentence alignment)と呼びます。Hansard コーパスでは、約90 は一つの英語文が一つのフランス語文と対応しますが、二つの文が一つの文に対応したり、対応する文が存在しないことも稀ではありません。

そこで、「対応する文同士の長さはほぼ等しい」という性質を用いて、動的計画法により並行テキストの文の対応を高い精度で付ける手法が提案されました。[Brown-et-al 91 a, Gale/Church91。]

これに対して、各文に含まれる単語の情報を用いる方法も提案されています[Kay/Roescheisen 93。] 英語とフランス語は同じ語源の単語(cognate word)が多いので、語彙情報は対応付け精度の向上に有効ですが、処理速度にやや問題があるようです。

現在では、単語・名詞句・述語構造など、文の様々な構成要素の間に対応を付与する研究が行なわれていますが、次節では、翻訳モデルを解説する準備として、単語の対応付け(word-by-word alignment)について説明します。

4 単語の対応付け

互いに相手の翻訳になっている文(単語列)の対を翻訳(translation)と呼び、(Quelle heure est-il? | What time is it?) というように、二つの文を縦棒で区切り括弧で囲むことで表します。二つの文の間の単語の対応付けとは、フランス語の各単語について、それと対応する英語の単語を示すものです。

図 1 に単語対応の例を示します。図 1 では英語の単語とフランス語の単語の対応を線で表していますが、この線のことを接続(connection)と呼びます。単語対応は、(Le programme a été mis en application | And the(1) program(2) has(3) been(4)

implemented(5,6,7))のように、英語の単語の後ろにそれが接続されるフランス語の単語の位置を示すことで表します。

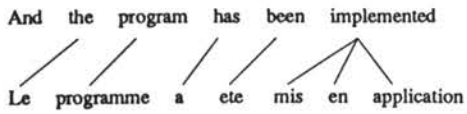


図 1: 単語対応付けの例 (Brown et al., 93 より)

統計的機械翻訳では、全ての対応付けは(ある確率を伴って)正しいと考えます。例えば、(Le programme a été mis en application |And (1,2,3,4,5,6,7) the program has been implementd)も可能な対応付けの一つですが、その確率は図 1 よりもずっと小さくなります。 (f|e) の全ての対応付けの集合を A(e,f) と表し、その中で最も確率が大きい対応付けをビタビ対応 (Viterbi alignment)と呼び、V(e|f) と表します。

5 翻訳モデルの推定

言語現象は非常に変化に富んでいるので,Pr(f|e) の分布を正確に求めることは実際には不可能です。そこで、言語翻訳に十分な精度を持つPr(f|e)の近似モデル (翻訳モデル) を如何に構築するかが統計的機械翻訳の中心的研究課題になります。ここでは、最も代表的な翻訳モデルである Brown et al., 90 を紹介します。

英語の the は、通常はフランス語の le, la, l' などに翻訳されますが、フランス語では省略されることもあります。また英語のonlyは通常はseulement に翻訳されるが、二つの単語 ne…que に翻訳されることもあります。そこで、ある単語対応において英語の単語 e が接続されるフランス語の単語の数を e の繁殖力(fertility)と呼ぶことにします。

また、フランス語文字列の最初の単語は英語文字列の最初の単語に接続されることが多いですし、フランス語の最後の単語は英語の最後の単語に接続されることが多いです。しかし、フランス語の単語と、それが接続されている英語の単語が離れた位置に現れることもあります。この現象を歪み(distortion)と呼ぶことにします。

以下では、条件付き確率 Pr(f|e)を翻訳 (f,e)の尤度と呼ぶことにします。フランス語単語列 f と英語単語列 e が互いに翻訳になっており、その間の単語対応が a であるような同時確率分布 Pr(f,a,e) を考えると、翻訳 (f|e) の尤度は、全ての単語対応 A(e,f) に関する条件付き確率 Pr(f,a|e) の和として表せます。

$$Pr(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} Pr(\mathbf{f},\mathbf{a}|\mathbf{e}) \qquad (3)$$

Pr(f,a|e) をモデル化するために、翻訳過程が次の3つのステップから構成されていると仮定します。(1) まず、英語の各単語について、それが接続されるフランス語の単語数を決める。(2) 次に、これらのフランス語の単語を決める。(3) 最後に、フランス語の単語列の中でのこれらの単語位置を決める。そして、この翻訳過程に対応して、次の 3 つの翻訳モデルのパラメタを定義します。

・繁殖確率(fertility probability)n(ϕ|e):英語の単語 e がϕ個のフランス語の単語に接続される確率

・翻訳確率(translation probability)t(f|e):英語の単語 e がフランス語の単語 f に翻訳される確率

・歪み確率(distortion probability)d(j|i,l,m): 英語単語列の長さが l で、フランス語単語列の長さが m の時に、英語の単語位置 i がフランス語の単語位置 j に接続される確率

もし、これらの翻訳モデルのパラメタが求まれば、ある対応付けに関するPr(f,a|e) を計算できます。例えば、図 1の対応付けの確率は、以下のようになります。

```
Pr(f,a|e)=
  n(0|and)
  n(1|the)t(le|the)d(1|2,6,7)
  n(1|program)t(programme|program)d(2|3,6,7)
  n(1|has)t(a|has)d(3|4,6,7)
  n(1|been)t(ete|been)d(4|5,6,7)
  n(3|implemented)t(mis|implemented)d(5|6,6,7)
     t(en|implemented)d(6|6,6,7)
```

t(application|implemented)d(7|6,6,7)

もし、Pr(f,a|e)が求まれば、式(3)より、全ての対応付けに関する和から翻訳の尤度 Pr(f|e)が求まります。従って、何らかの方法で翻訳モデルのパラメタを推定できれば、翻訳の尤度が求まるので、入力に対する最尤な翻訳出力や、入力と出力の間の最尤な対応付けを決めることができます。

統計に基づく手法では、訓練データと呼ばれる翻訳の集合の尤度を最大にするように翻訳モデルのパラメタを決定します。これを最尤推定法と呼びます。

ここで、翻訳の集合の尤度は、個々の翻訳の尤度の積です。一般に、統計に基づく手法では、最尤推定値を求めるのに EM（期待値最大化: Expectation Maximization）アルゴリズムを用います。このアルゴリズムでは、パラメタに対して適当な初期値を与え、前回のパラメタの推定値を用いて新しいパラメタの推定値を求めるという手順を繰り返します。ちょうど高校の数学で習う漸化式の要領です。この繰り返しにより、パラメタは(局所的な)最適値に収束することが証明されています。

しかし、実際には、式(3)は非常に大量の計算を必要とするので、Brown et al., 93 では、ビタビ対応 V(e,f) の近傍(neighborhood)という概念を導入し、比較的大きな確率を持つ対応付けだけに関する和で EM アルゴリズムの計算を近似しています。

最後に、図2[fig:trans-fertile-prob] に、英語の the の翻訳確率と繁殖確率をHansard コーパスから求めた結果を示します。これより the は le, la, les, l' などに翻訳されるが、フランス語には翻訳されない(繁殖力が 0 になる) 場合もあることがわかります。

f	$t(f\|e)$	ϕ	$n(\phi\|e)$
le	0.497	1	0.746
la	0.207	0	0.254
les	0.155		
l'	0.086		
ce	0.018		
cette	0.011		

図2 theの翻訳確率と繁殖確率(Brown et al., 93 より)

6 おわりに

如何でしょう? 機械翻訳への統計的アプローチについて少しは理解して頂けましたか? Brownの手法は、自己組織的（self-organizing）であること、すなわち、対象とするデータから言語モデルを自動的に学習することに重点を置いているので、言語学的に見ると翻訳のモデルとしては稚拙に感じるかもしれません。しかし、その副産物である対応付き二言語コーパスや翻訳モデル確率は、既存の機械翻訳システムにとっても非常に有益な情報を含んでいます。

ちなみに Peter Brown は、その後、確率・統計モデルの腕を買われて IBMから金融業界へ転身し、大金持ちになったそうです。こんなところにもアメリカンドリームが潜んでいるあたり、やはり、この国は懐が深いと思います。

参考文献

[1]Brown P. F. et al., ``A Statistical Approach to Machine Translation'', Computational Linguistics, Vol.16, No.2, pp.79-85, 1990.

[2]Brown P. F. et al., ``Aligning Sentences in Parallel Corpora'', ACL-91, pp.169-176, 1991.

[3]Brown P. F. et al., ``The Mathematics of Statistical Machine Translation: Parameter Estimation'', Computational Linguistics, Vol.19, No.2, pp.263-311, 1993.

[4]Gale, W. A., and Church, K. W., ``A Program for Aligning Sentences inBilingual Corpora'', ACL-91, pp.177-184, 1991.

[5]Kay, M., and R o scheisen, M., ``Text-Translation Alignment'', Computational Linguistics, Vol.19, No.1, pp.121-142, 1993.

永田昌明　　ＮＴＴ情報通信網研究所メッセージシステム研究部
Masaaki NAGATA, NTT Network Information Systems Laboratories
(Tel) 0468-59-2796 (Fax) 0468-59-3428
(E-mail) nagata@nttnly.ntt.jp

協会活動状況報告

第1回理事会	5月10日	①93年度事業報告案②93年度決算報告案③役員改選案④その他
監査	4月21日	①93年度収支計算書②期末貸借対照表③財産目録④詳細帳票類
運営委員会	4月26日	①理事会上程議案審議②総会、セミナー開催案③会員の動向
財務委員会	4月12日	①決算案検討②９４年度収入見直し
利用技術研究会	4月7日	①パソコン翻訳ソフト実態調査②ユーザ訪問調査結果報告
	5月13日	①ヒアリング「プロトコル」②報告書のまとめ方
	6月1日	①項目別報告内容検討
例文評価研究会	4月22日	①ヒアリング「凸版ＥＤＭＴ」②例文評価のまとめ③論文提出審議
	6月3日	①報告書案②ネイティブチェック文例評価
需要予測研究会	4月14日	①普及外部要因検討②パソコン翻訳ソフト調査検討
	5月26日	①調査結果検討②将来予測数値検討③内外普及要因検討
翻訳現場見学会	4月6日	浜松「データプロセス研究所」

新会員の紹介

吉田敬一
西田幸平
新垣元
望月伸浩
松田欣治
博田良
市村正弘
松井浩
大塚晴久
藤原光彦
河野悟
井上宗通
赤嶺良彦
平田篤己
Foo Chee Chye Michael
Mark Homnack
Tag Young Moon

AAMT
ジャーナル
No. 7 (June/1994)

発行　アジア太平洋機械翻訳協会
所在地　〒103 東京都中央区日本橋小舟町 4-10 大宮ビル 3 階
☎ 03-3664-5637 / 5638　FAX 03-3664-1352
E-Mail(PC-VAN) JRD 08020 (NIFTY-SERVE) KGE 00013

編集委員会 委員長 野村浩郷（九工大）Hirosato Nomura
亀井真一郎（NEC） 杉山健司（富士通）
信田恵壱（シャープ）
事務局　星野禎男　西郷容子

第４回通常総会

日　　時　平成６年６月２２日　午後４時～５時　（機械翻訳セミナー終了後）
場　　所　機械振興会館 Ｂ２ 大ホール
（東京都港区芝公園 3-5-8　会場連絡先 03-3434-8216）

議案(骨子)　**第１号議案　93年度事業報告及び決算報告(案)**

①利用技術、例文評価、需要予測、制限言語など 延べ27回に及ぶ研究会活動、機械翻訳サミット、チュートリアル、ワークショップ、翻訳現場見学会、研究成果報告会等の事業活動、AAMTジャーナル、MTニューズレター、技術文献頒布等の広報活動について報告する

②当期収入15,210千円(内会費収入14,364千円)支出14,651千円、収支差額560千円、前期繰り越し金15,737千円と共に時期に繰り越す案件

第２号議案　94年度事業計画及び予算案

①情報発信基盤強化のため調査研究活動を積極化、ユーザの利用技術向上をはかるための講習会の開催、利用ガイドラインの設定、関連団体との連携による普及活動の活発化、国内外の研究、技術、利用情報の収集を行い、広報誌等を通じて機械翻訳に関する情報の提供などの活動を推進して行く。

②当期収入32,417千円(内繰り越し15,733千円) 支出は調査研究活動費、事業費、事務諸費23,210千円の支出予算案

第３号議案　役員の改選案

任期（２年）満了に伴い、理事及び幹事の改選を行う
会長、副会長は新理事会で互選される

第４号議案　その他会員提案事案

・尚総会欠席者、代理出席者の委任状提出期限は６月１５日です。
・通常総会会場へは営団地下鉄 日比谷線 神谷町駅下車 徒歩7分
都営地下鉄 三田線 御成門駅下車 徒歩7分
都営地下鉄 浅草線 大門駅下車 徒歩12分