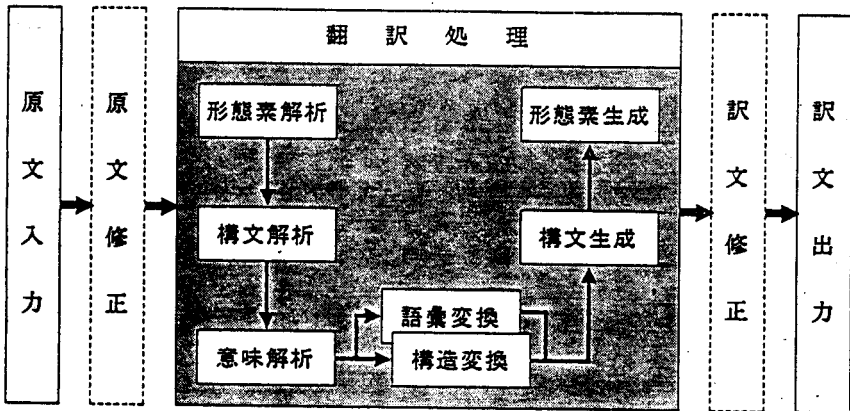


## 海外における機械翻訳の研究

NTT基礎研究所 野村浩郷

今日は海外における機械翻訳の研究、外国で行われています機械翻訳に関する活動の概要をお話し、具体的な話の前に、機械翻訳のプロセスの要点、海外の動向、現在私共が行っている研究の説明をしたいと思います。

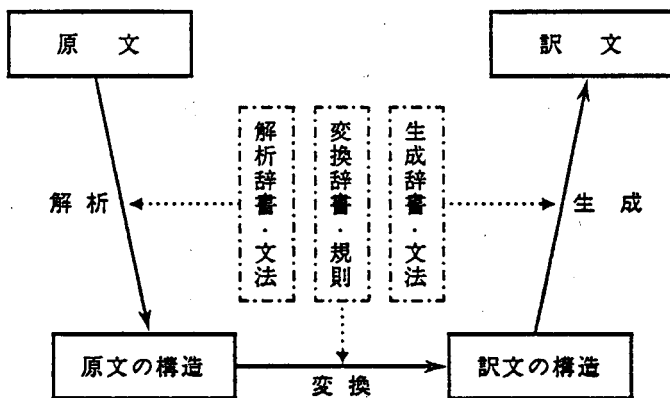
### 機械翻訳の処理の流れ



(資料1)

機械翻訳というのは、決してコンピュータの処理だけでなく、原文を機械に入れるプロセス、翻訳し終わったものをプリントするというプロセスも含むわけです。従って、翻訳の処理というのは、機械にとって翻訳しやすいような形に原文を人間が修正するというプロセスも入ります。その後でコンピュータの中へ文を入れ、形態素解析・構文解析・意味解析等を通じて原文の構文や意味の構造をコンピュータ内部に表現し、それを語彙レベル、構造のレベルで相手言語へ変換した後、構文生成・形態素生成を行い訳文を生成します。訳文が生成されたならば、それを人間がみて修正し、まともな訳文に修正するというプロセスをとり、そしてその後印刷をします。

## 翻訳処理の原理

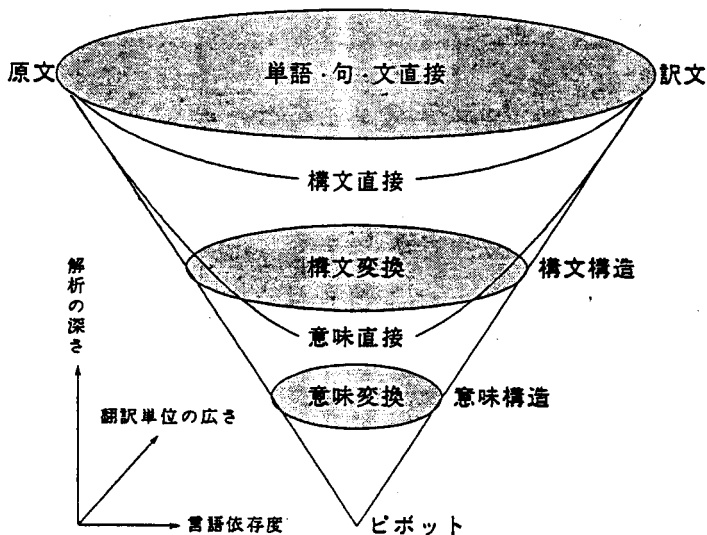


(資料2)

翻訳処理の流れは、(資料2)のように原文を解析し、訳文に適した構造に変換したうえで、訳文を生成するというのが大まかな流れです。その時には、解析では解析の辞書、文法などの言語に関する知識を使い、変換に関しては、1つの言語の表現から他の言語の表現になるわけですから、2つの言語をむすびつける知識が必要となります。それを変換辞書、変換規則などといい、訳文の方の言語に依存する生成辞書・文法に従い訳文を生成します。

ところで、機械翻訳の処理を実際にさせる時の方式はいろいろありますが、方式を構文直接や意味変換というふうなものですべてを規定するというようなことはそれほど意味がなく、翻訳をするには実際に何をしなくてはいけないかというのが最も重要です。

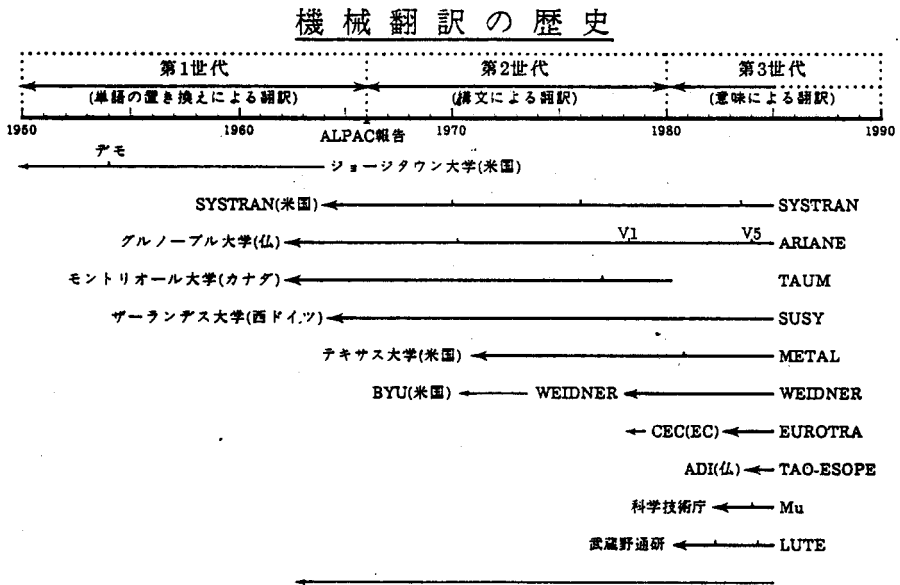
## 機械翻訳の方式



(資料3)

資料3が示すのは、1つは解析の深さ、これは原文に含まれている意味内容あるいは文の構造だとかをいかに精密に詳細に解析してから変換するかという問題です。もう1つは、翻訳の単位・単語だけを見て、単語だけ置き換えるという操作で翻訳するのか、名詞・句という単位で翻訳をするのか、文という単位で翻訳をするのか、パラグラフという単位で翻訳をするのか、というような翻訳単位の広さという側面もあります。さらに、解析の深さと関係しますが、言語に依存しないで理解するところの万国共通の概念、意味内容のどのあたりまで立ち入るかといったレベルの話があります。それによって単語・句・文を直接相手の方へもっていく方法と、文法に基づいた文の構造を直接相手の言語へもっていく方法、構造解析した後、構造の変換というのを行う方法と、意味をダイレクトにもっていく方法があります。その中で最も理想的な表現がピボットという方法ですが、これも必ずしも十分に深い意味を表わさない場合もあります。

では、機械翻訳の海外の動きに入っていきたいと思います。その前に機械翻訳の歴史を簡単に説明していきます。



(資料4)

機械翻訳の研究が開始されたのは1950年頃からで、電子計算機が世の中にでてきたのが1946年代でありますから、計算機が出た頃あるいはその後後に、すでに機械翻訳の研究は開始されていきました。計算機というのはその名の示すとおり計算をする機械です。英語でCOMPUTERとは計算をするという意味であり、CALCULATORは算数の計算をするという意味です。しかし、翻訳をさせるというのは計算ではありません。計算機を使う使い方は2つあり、1つは計

算をさせる事でもう一方は計算以外のことをさせる事です。翻訳を行うには人間の思考過程を計算機上で実現させる必要があり、翻訳を含む人工知能という分野は、計算以外のことを計算機で行います。翻訳の試みは、1950年から始まっていました。ところがこの時期の計算機は真空管でつくられており現在のパーソナルコンピュータと比べると見劣りがする計算機でした。従って小規模の機械で、人間の頭脳で行っている様な翻訳が実現できるはずがありません。それで初期の段階の機械翻訳は必ずしも成功しませんでした。成功しなかったというより、現在まで続いている、今後まだ研究していかなければならない事がたくさんあるという状況です。しかしながら、この計算機に翻訳をさせるという事は非常に大きな意味を含んでいると同時に、計算以外の計算機の使い方というのは、人工頭脳という言葉にそのままつながり、一つの夢でした。とくにアメリカでは、非常にたくさんの研究が行われています。

1950～1960年代の中頃までを第1世代と呼び、機械翻訳の研究者としては黄金の時代でした。非常にたくさんの研究費が出て、非常にたくさんの研究が行われたらしいです。

1954年、アメリカのジョージタウン大学とIBMとが共同研究していました機械翻訳の大デモンストレーションがアメリカで行われました。その翻訳は、ロシア語から英語への翻訳で、非常にセンセーショナルだったらしいです。その時の機械翻訳のシステムは、ロシア語の単語を250持っていて、ロシア語の1つの単語に対してたかだか2つの英語の単語が対応づけられ、それで翻訳をしようとしたわけです。これは、250のロシア語とそれらに対したかだか2つの単語が割り当てである辞書だけを使っていたので、ロシア語で書かれた論文を英語に翻訳はできないわけです。ところが、このシステムはその後どんどん拡張され、研究が進められ、特にアメリカの軍ではソビエトの情報を収集するために、かなりのロシア語の文献を翻訳しています。翻訳された文献は、国会図書館で12～13年前に見た時はとてもむつかしくて、英語の文の中にロシア語がどんどん入っています。つまり辞書にないロシア語の単語がそのまま出てきて意味がわかりませんでした。この時のもう1つの制限が、機械翻訳を行わせるためのバックグラウンドになる言語理論が発達しなかったといっても良いことです。言語学そのものは紀元前よりありましたが、機械の上では十分な能力を保証するための基礎理論までは発達していないという状況です。1960年の少し前くらいから、アメリカのMITのチョームスキーという言語学の先生が構造生成文法理論という提案をしました。彼の言語理論は1960年代に入り、一般にいわれるオートマトン理論(計算機の計算の理論的裏づけをする理論)と対応づけられて、言語のあるクラスと、オートマトンの処理できる能力とが1対1に関連づけられる形で発展し、それによって計算機言語、コンパイラの基礎理論が完成しました。この理論そのものは、いわゆる構文解析、記述された文法に基づいて、元の文の構文的な構造を解析するための手法を定式化していったわけです。従って、こういう構文解析が未発達であったということから、第1世代ではたくさんのチャレンジがあったにもかかわらず、この段階で機械

翻訳を成功させることができませんでした。その時の技術というのが単語のおきかえを基本とする技術であります。それに対する批判あるいは評価としてアメリカでALPAC報告が出され、以降アメリカで一切の研究費を機械翻訳には出さないという決定が行われました。それから第2世代に入り、その間を暗黒の時代とって、研究者としては研究費が1円ももらえないという時代が続きました。ところが、言語理論の発達と計算機上での処理モデルの理論とがむすびついた研究が進み、この時期には70年くらいまでの間に構文解析の理論が非常に発達しました。すなわち、この間における計算機の規模、利用するための技術の発展が、第2世代をもたらし、構文による技術をかなり発展させたのです。ところが残念ながら、我々は文を書く、話をする時に、構文構造を相手に伝えることを目的として話をしているのではありません。即ち、名詞、助詞、動詞、形容詞の言い方をしても相手に伝わりません。構文解析というのは、名詞、助詞、動詞、形容詞というのが最初の名詞が主語の部分をなし、次の動詞が述語をなし、形容詞+名詞が名詞句をなして、例えば目的語という構造解析をするのでありますから、言語の持っている内容を表わすものではありません。我々が友達と話をする時何を伝えたいかと言うと、それは意味を伝えたいのです。従って言語の意味を扱わない限り、決して言語の処理のための高度な技術が開発されたとは言えません。1980年代から意味と言うことの研究が行われるようになってきました。但し、充分進んでいるとは申しがたく、むしろ、これから本当の研究が進められる時期であります。1980年代以降、第3世代と名前をつけていますが、この3つの世代と、具体的な研究、あるいは機械翻訳のための実験のシステムを書いたのがここにあります。

ジョージタウン大学のシステムを引きついで発展させ、商品化したのがアメリカのストララン会社で作っているシステムです。これはアメリカの軍の関係だとか、カナダのジェネラルモーター社、ECの欧州共同体本部とかでいろいろ使われています。もちろん制限された能力の範囲で使われています。

仏のグルノーブル大学では、1962年位に研究を開始し、現在までずっと研究を続けています。最近ではアリアンヌ (ARIANE) という翻訳の実験システムを開発して、78年に完成し、第5バージョンが1983年に作られています。

カナダのモントリオール大学では、1960年代の最初から研究を進め、途中から政府がカナダの国家プロジェクトとし翻訳の研究をやってきていますが、1980年～81年にかけてモントリオール大学の進捗状況を厳しく評価した結果、とても経済的に達成される見込みがないということと技術的に研究がまともに進んでないという理由でプロジェクトは現在中止されています。

西ドイツのザーランドス大学では、1960年代の始めから研究が開始され現在まで続いており実験システムのスージー (SUSY) というのが動いています。

アメリカのテキサス大学では1970年頃から研究を開始しずっと続いてきて政府の援助を受けましたが、途中から援助が打ち切れ、西ドイツのジーメン社が以降引きついで資金援助し研究が進められました。

アメリカのBYU（ブリガムヤング）大学では、モルモン教の聖書を世界各国語に翻訳するためにプロジェクトを作り研究を開始し続けてきました。しかし、その後研究者が二手に分かれ、両方とも研究をやめそれぞれ会社を作りました。1つはプログラマー、他の1つは言語学者の方であり、プログラマーの方はワイドナーという会社を、言語学者のほうはアルプスという会社を作りました。

EC本部では、1976年位に機械翻訳をやっています。新しい機械翻訳システム開発プロジェクトを1978年から検討し、正式にプロジェクトとして発足したのが1982年の11月からです。

フランス政府は、1982年7月から仏国家としてのプロジェクトを3年計画で開始しています。現在のシステムの概要を説明したのが資料5です。

### 機械翻訳実験システム(外国)

システム名 機	翻訳言語	辞書	翻訳対象	備考
ARIANE-78 Grenoble大学(仏)	独、英、仏、 露、その他	1万語	一般	1983年に第5版を作成。
SUSY Saarlandes大学(西独)	独、英、仏、露	13.5万語	保守マニュアル 新聞・雑誌の見出し	実験中。
TAUM-METEO Montreal大学(加)	英-仏	—	気象通報文	1977年から実用中。
METAL Texas大学(米)	独-英、英-独	1万語	技術文書	独-英システムは1983年から評価開始。 英-独ししては1983年より開発開始。
ITS BYU(米)	—	—	—	会話型翻訳援助システム。
EUROTRA CEC(EC)	英、仏、独、 伊、デンマーク (スペイン、 ポルトガル)	2万語(予定)	一般	パイロットシステムの開発が目的。 1982.11-1988末。
TAO-ESOPÉ ADI(仏)	英、仏	3千語(予定)	航空機マニュアル (英-仏) 情報処理文獻 (仏-英)	翻訳援助システムの作成を通じて ARIANE-78の技術を民間に技術移転する のが目的。 1983.7-1985.6。

BYU: Brigham Young University, CEC: Commission of the European Communities, ADI: Agence de l'Informatique

### (資料5)

仏のグルノーブル大学のシステムがアリアンヌ78 (ARIANE-78) と呼ばれ、言語としては独語、英語、仏語、ロシア語、その他の小規模な実験ができます。辞書は約1万語の大きさで、翻訳対象は特に限定されず、学術関係のドキュメント的なものを扱っております。

以上の中で、カナダ政府が資金援助を行いモントリオール大学で稼働している機械翻訳システムにトムメテオ (TAUM-METEO) というのがあります。このトムメテオというのは、気象通報文のみを英語から仏語に翻訳を行うシステムであり、1974、5年から始まり、77~78

年に完成しました。1977年以降カナダで実際に使われています。各地に散らばっている気象通報台から通信回線を使ってダイレクトに気象通報文が入りそれが人手を介さずこのシステムの中に入れられ、フランス語に翻訳されて、これが新聞社、テレビ会社に送られ人々に流れています。これは気象通報文に限られていますから使われている単語の数や及び文の形は決まっております技術そのものは一般の文を翻訳する形とは違っています。

### 市販翻訳システム(外国)

システム名 機 種	翻 訳 言 語	辞 書	速 度 計 算 機	備 考
SYSTRAN Systran社(米)	英→仏、英→露、英→独、 英→日、英→スペイン、 英→伊、英→アラビア、 英→ポルトガル、 英→ペルシア、独→仏、 独→スペイン、中→英	10万語～ 2万語～	百万語/時 IBMメインフレーム	バッチ処理型翻訳システム
WEIDNER Weidner社(米)	英→仏、英→独、 英→スペイン、 英→ポルトガル、 英→アラビア	数千語	4千～8千語/時 PDP11/23～ PDP11/70	会話型翻訳援助システム
ALPS Alpelt社(米)	英→仏、英→独、 英→スペイン、英→伊	—	—	会話型翻訳援助システム 電子辞書
LOGOS Logos社(米)	独→英	—	—	バッチ処理型翻訳システム
SMART Smart社(米)	英→仏、英→独、 英→スペイン、英→伊	—	—	電子辞書
SPANAM (米)	スペイン→英	—	—	電子辞書
TITUS Institut Textile de France(仏)	仏、英、独、イタリア	1.2万語・	—	翻訳援助システム 制限文法、限定語彙 会話型 ビボット方式

#### (資料6)

次に市販されている機械翻訳システムの種類をみていただきましょう。

シストラン(SYSTRAN)というのは、シストラン社が発表している内容で、言語対は16の言語対があります。それは英→仏、英→露、英→独、英→日、英→スペイン、英→イタリア、英→アラビア、英→ポルトガル、英→ペルシア、独→仏、独→スペイン、中→英が発表されています。具体的にどのくらい使われるかというのはわかりません。辞書の大きさは約10万語、12～3万語くらいから、英日の場合は2万語ちょっとくらいまで持っています。翻訳速度は、公式には1時間当たり100万語の速度です。

ワイドナー(WEIDNER)でも、英→仏、英→独、英→スペイン、英→ポルトガル、英→アラビアを扱い、辞書の大きさが数千語で、速度が1時間当たり4千～8千語くらいであり、計算機は小さい計算機を使います。システムの型は会話型、インタラクティブの型です。

このように機械翻訳システムの公式発表があった時には、このシステムが現実にとり得るだけの翻訳能力があるのかということに注意しなければいけません。例えば、ワイドナー社のシステムの正式の名前は翻訳援助システムという名前がついています。これは翻訳システムではなく、翻訳者が翻訳をする作業を支援するシステムであり、あくまでも翻訳の主体は人間にあります。さらにこれをもっと徹底させたものが、BYUのアルプス(ALPS)という会社のシステムで、こ

これはまさしく翻訳援助システムというより電子辞書に近いもの、即ち計算機の中に辞書が入っています。人間が翻訳をする時に、普通に辞書をひきながら翻訳をし、そのひく辞書の中味が計算機の中に入っていて、この単語の訳は何か?という時に、ディスプレイで中味が見れます。みれるだけでなく、その中に訳が書いてあり、その訳文の1部に使いたいと思うとその文をもってこれます。いわゆるオフィスオートメーションの中に電子辞書が入っているシステムです。

次にログス (LOGOS) というシステムですが、これはアメリカ軍がベトナム戦争をやっていた時に、ベトナムの兵士にアメリカの軍事兵器を使わせる訓練のため、兵器のマニュアルを英語からベトナム語へ翻訳をしたいということではじめたプロジェクトによるシステムです。これは急いで研究を行ったため余りたいたことができず、機械翻訳をやるよりもベトナムの兵士に英語を教える方が早いのではないかという結果でした。最近、別の観点からドイツ語から英語への翻訳ということで開発を始め、これが比較的質が良いということです。さらに機能が低いものがスマート (SMART) システムといいまして、これは翻訳システムと言えるかどうかよくわかりませんがワープロに少し辞書が入っているというくらいのもので、とても機械翻訳システムですとは言えません。それと同じレベルのものにスパンム (SPANAM) というのがあります。今申し上げたものとは反対に、全く別のものがあります。これは、仏のパリにティタス (TITUS) という名前のシステムがあります。これは、仏語、英語、イタリア語、独語の4つの言葉の文章を作成するための援助システムで、これは翻訳システムと言うよりもこの4つの言語で同時に文章を作るワープロだと言ってもよく、4つのうちの1つの言語で入力し、4つの言語で出力されるというシステムです。

### 市販翻訳システム(日本)

システム名 機	翻訳言語	辞書	速 度 計 算 機	備 考
SYSTRAN 日本Systtran社	英日	2万~	百万語/時 IBMメインフレーム	バッチ処理型翻訳システム
WEIDNER Bravice社	日-英	4万語	3千語/時 DEC LSI11/73	会話型翻訳援助システム セグメント直接
ATLAS/I 富士通	英-日	5万語	6万語/時 Facom Mシリーズ	翻訳援助システム 構文直接 作業手順書、技術資料、技術論文 訳文修正用の翻訳エディタ 辞書エディタ
ATLAS/II 富士通	日-英	5万語	6万語/時 Facom Mシリーズ	翻訳援助システム 構文変換
ATHENE 日立	英-日	7万語	2万語/時 日立M-240H	翻訳支援システム 経済分野を対象(欧米の経済新聞、経済通信ニュース、市況データ、マーケットコメント) 会話型(和文、英文、多義表示をみながら校正) 構造解析による直訳(形態素解析、構文解析、中間語表現、構文合成、日本文生成) キー入力
VENUS 日電	英-日	--	--	--

### (資料7)

参考までに日本のメーカーが新聞等に公式に発表しているシステムを、発表通りにお話ししましょう。シストランは日本語と英語の翻訳のシステムで、1983年の夏か秋に発表してい



ます。まだ販売するというところまでではないにしても、そろそろ翻訳サービスを開始する状況になっているということです。ワイドナー社は日本の翻訳会社のブラビスという会社を買収し、そこで日本語→英語への翻訳システムを発表しています。この辞書は約4万語と言われていますがワイドナーのシステムは基本的には4千～5千語の辞書の範囲内で使うというものです。なぜならば、人が知っている単語はそれほど多くはなく、普通に使っている単語は2千～3千、多くて4千～5千と言われており、人によって使用する単語が違うにもかかわらず何万語もの辞書を常に用意しておくのは無駄で、その人が必要とする単語だけシステムの中に入れていけばよいという考えからです。辞書を大きくすればそれだけシステムの値段が高くなるので、ミニコン、パソコンを使ってできる様にすれば良いと言うのが基本的な考え方です。4～5千語の辞書で普通は充分であろうところから、違う人が使う時にはその人専用の辞書に入れかえて使えば良いというのが基本になっています。

現在の機械翻訳システムは、正確に翻訳をするかどうかとか、あるいは翻訳後の文章、例えば日本語から英語に翻訳をした場合には本当にいい英語になっているのかということが非常に難しい問題になっています。

即ち、原文で述べられていることの意味内容がこうだということを大多数の人が認めるならば、その内容が訳文の方にも一応表現されていなければなりません。例えば原文の方では、「長崎から福岡まで150kmです」と書いてあるのに、訳文の方では「長崎から福岡まで1500kmです」とか、又「東京から札幌まで150kmです」になるとこれは間違いであり、そのレベルで原文で述べられていることをどうにか訳文で表わさなくてはなりません。そのためにはどのようにしたら良いかということを一生懸命やっているのが、現在機械翻訳の研究開発技術のレベルであると考えておいていた方がよろしいと思います。例えば、谷崎潤一郎の小説は訳せませんし、まして利害関係がからむ様なものは安易にはできないわけです。富士通が発表しているものにATLAS-IとIIというのがあります。富士通の2つのシステムは、処理する方式が全くちがっています。ATLAS-Iは非常に単純な処理を基本にしておりますが、それに対して、ATLAS-IIは比較的夢の多い考え方で作りあげています。しかし当然のことながら、商品にする際には研究でいろいろ考えているところの部分的な機能しか実現できません。このシステムは1985年6月頃発売開始だということです。日立はアテネあるいはアティーネ(ATHENE)という名前で出しています。これもどちらかというとならATLAS-Iに近い処理の方式で、訳語の生成のところでもかかり受け関係をどこに設置するかを翻訳のプロセスとしているシステムです。機械翻訳システムということでお話しましたが、機械翻訳あるいは自然語処理を考える時に、「原文がどのように翻訳されるのか」や「訳文の中に冠詞はちゃんと出ているのか」という話ばかりでは実際に何もできないわけです。その1つの例を資料8に示します。

## ターミノロジデータバンク

名 前	前 関	サービス 開始年	言 語	語 数	分 野	備 考
EURODICAUTOM CEC(EC)		1973	独、英、仏、 伊、オランダ、 デンマーク、 スペイン、 ポルトガル	30万語	技術、科学、経済、 法律	端末からオンライン、オフライン で使える EURONETからも使える
LEXIS Bundessprachenamt (西独)		1966	独、英、仏、 露、伊、 ポルトガル、 オランダ、 ポーランド	140万語	通信、電子などの 技術分野 数学、化学、物理、 自然科学などの科学 分野 防衛分野	端末からオンライン、オフライン で使える
TEAM Siemens社(西独)		1967	独、英、仏、 露、伊、 スペイン、 ポルトガル、 オランダ	100万語	電子、情報処理、 通信、経済など	端末からオンライン、オフライン で使える 色々な高品質のプリンタが接続さ れている
TERMIUM Canada政府(加)		1975	英、仏	60万語	50の分野	端末からオンライン、オフライン で使える 電話からも使える

### (資料 8)

ターミノロジデータバンクとは専門用語のデータベースのことで、専門用語とは非常にたくさんあり、物理学ならば物理学に関する言葉を、化学なら化学、医学ならば医学とその数は非常に多いのです。残念ながら日本では言葉というのが大切にされていないのかどうかわかりませんが、こういう試みは従来はなかったといってもいいくらいです。しかし外国では非常に進んでいて、ECがユーロディオートム(EURODICAUTOM)という専門用語集を計算機の中に作っています。そして端末から必要に応じて、ある言葉の訳を検索できるようになっています。その中には8つの言語が入り、単語の数としては30万語くらいです。2番目は西ドイツ政府が作っているレキス(LEXIS)、3番目が唯一民間会社ジューメン社が作っているティーム(TEAM)、一番下は、カナダ政府が作っているターミアム(TERMIUM)というものです。ユーロディオートムCECは専門の翻訳者が2千人いて、この2千人が毎日翻訳をしています。カナダ政府にも千人、西独では3百人の専門の翻訳者がいて毎日翻訳をしています。しかしその人達は大学の専攻は言語学で、翻訳の専門家であり、決して物理学の専門ではありません。又音楽や化学の専門家でもないのですが、それぞれの文献やドキュメントも翻訳しなければなりません。その時に専門用語の意味はよくわからないし、専門用語を勝手に訳したら怒られます。従って専門用語を正確に訳し、かつ前のページと訳を統一するということが必要であるためにターミノロジデータバンクを行っています。イギリスではこのような研究が遅れており、1982年にイギリス政府がある大学に諮問し、その大学の先生が3年かけて、英語のターミノロジに関する事とイギリス政府としての取り組み方の答申を行っています。その答申の中には、英語はすでに英語圏内で言葉が通じなくて、カナダの英語とイギリスの英語はちがいが、もちろんアメリカの英語やオーストラリアの英語、さらにはインドの英語ともちがうという指摘があります。

つまり、英語というのはイギリスの言葉であるはずなの、英語圏内の中でイギリス語の占める比重は非常に小さくなっていて、イギリス人は英語がわからなくなるのではないか、英語を話して意味がわからなくなる時期がくるのではないか、ということからイギリス政府としては努力をすべきであるという答申がきています。

そこで話をしぼり、ECが機械翻訳に関して具体的にどのような活動しているかということを中心に御紹介させていただきます。それは3つあります。1番目はターミノロジーデータバンクに関する事、つまり先程のユーロジコタムに関する事です。2番目が商用の機械翻訳システムを改良して、EC内での翻訳サービスの実験を行うことです。3番目が新しい機械翻訳システムをつくるプロジェクトであり、シストラン社のシステムを1976年に購入し、5年間をかけて辞書を開発した後1981年から翻訳サービスの実験を開始しています。これは3ヶ国語の言語対を追加して行ったもので、さらにこのシステムはフランスの国立の研究所及びドイツの原子力研究センターでも使われています。また専門の翻訳者の中には機械が好きでない方もいますからCECでは英→仏への翻訳システム、仏→英への翻訳システムと、英→伊への翻訳システムを提供し実験を行っています。この中で仏→英のシステムはよく使われていますが、英→仏のシステムは余り使われていません。これは外国では、翻訳者が翻訳をする際には、自分がnative language (母国語) とする言葉にむけてしか翻訳しないためです。あるいは翻訳者でも、自分がnativeな言語に向かってしか翻訳せず、このため日本語→英語は翻訳する時には英語をnativeとする人しか翻訳しません。しかし日本人が翻訳する場合には両方しますが外国人は決してしません。仏→英語システムが使われているのは、英語を母国語とする人は機械を使うことがあまり嫌いでないためです。翻訳者によく使ってもらおうと頭を悩まし、1年365日をほとんどこれに費やしているのが、CECでの実験を担当していますヘイトンのピボットという人でこの人がいつも考えていることです。シストランの翻訳では原文がフランス語で、これを機械にかけ、その後専門の翻訳者が修正し最終的に英語となります。この場合、A4版1ページ当たり専門の翻訳者は30分かけて修正しますが、従来はもっと時間がかかっていました。つまり、翻訳を専門の翻訳者に頼むと3週間～4週間かかっていたわけです。そこでEC側としてはもっと早くしようということで、機械を使うということを提案したところ2日～3日に短縮されました。その時専門の翻訳者は、計算機が翻訳した翻訳文を修正するのに1ページ30分以下におさえて下さい、30分かけたらそこで止めて下さいという形をとっています。従来じっくり時間をかけた修正のしかたが(資料9、10、11、12)です。

翻訳例 一原文(フランス語)一

II. LE BUDGET 1983

Au moment où s'achevaient les pourparlers au sein des instances du Conseil au sujet de la déclaration commune précitée, les Etats membres étaient déjà saisis de l'avant-projet de budget de la Commission pour l'exercice 1983. Cet avant-projet énonçait un accroissement des dépenses en engagements de 671,7 Mios ECU.

Au moment même cependant où la Commission transmettait cet avant-projet de budget, elle a saisi simultanément le Conseil d'un avant-projet de budget rectificatif et supplémentaire n° 1, qui était appelé à financer de nouvelles politiques et de nouvelles actions de la Communauté dans des domaines aussi divers que la recherche et l'aide au développement à partir d'économies réalisées dans le secteur du FEOGA comme de l'excédent de compte de gestion de 1981.

L'ampleur de cet avant-projet de budget rectificatif portait sur, grosso modo, 487 MUCE et, comme il s'agissait essentiellement de dépenses non obligatoires, ces dernières allaient sensiblement accroître la marge du Parlement européen. Par ailleurs, les décisions de fond pour un certain nombre d'actions proposées faisaient défaut.

Devant cet état de choses, le Conseil, après avoir longuement examiné les possibilités qui s'offraient à lui d'accepter un projet de budget rectificatif et supplémentaire réduit lors de sa session du 28 juillet, lorsqu'il eut à se prononcer à la fois sur le budget 1983 comme sur ce budget rectificatif et supplémentaire a rejeté ce dernier, tout en invitant simultanément la Commission à procéder à des virements pour les transferts qui lui semblaient, à lui Conseil, prioritaires (exemple : Fonds Social).

(資料 9)

翻訳例 一粗訳(英語)一

II. THE 1983 BUDGET

At the moment where were completd the negotiations within the Council's subordinate bodies about the above-mentioned statement common, the Member States were already seized of the preliminary draft budget of the commission on 1983. This preliminary draft stated an increase in expenditure in engagements: of 671.7 Mio\$ ECUS.

At the moment even however where the Commission transmitted this preliminary draft budget, it approached simultaneously the Council concerning a and additional n preliminary draft amending budget 1, which was bound to finance new policies and new actions of the Community in as various fields as research and the development assistance from savings made in the sector of EAGGF like a management accounts surplus of 1981.

The width of this preliminary draft amending budget carried on, roughly speaking, 487 MILLION EUA and, as it was primarily non-compulsory expenditure, these last appreciably will increase the margin of the European Parliament. In addition, the basic decisions for a number of actions suggested were missing.

In fron of this state of affairs, the Council, after lengthily having examined the possibilities which offered themselves to accept him a draft supplementary and amending budget reduced at its meeting on 28 July, when it had to come to a conclusion at the same time about the budget 1983 like on this pectifying budget and additional rejected the latter, while asking simultaneously the commission to proceed to transfers for the transfers which seemed to him, him the Council, priority (example : Social fund).

(資料 10)

翻訳例 一最終翻訳(英語)一

II. The 1983 BUDGET

By the time that the negotiations within the Council's subordinate bodies on the above-mentioned joint statement were concluded the Member States had already received the Commission's preliminary draft contained an increase in expenditure on commitments of 671.7 million ECU.

However, at the very moment when the Commission was forwarding this preliminary draft budget, it received from the Council a preliminary draft supplementary and amending budget No 1, which was intended to finance new Community policies and projects in such various fields as research and development aid from savings made in the sector of the EAGGF such as the management account surplus in 1981.

The funds covered by this preliminary draft amending budget amounted to approximately 487 million ECU and, as this was primarily non-compulsory expenditure, it would have increased appreciably the European Parliament's margin for manoeuvre. In addition, the basic decisions on a number of proposed projects were lacking.

In view of this state of affairs the Council discussed at length the possibilities open to it of accepting a reduced draft supplementary and amending budget at its meeting on 28 July, when it had to make a decision on the 1983 budget at the same time as on this amending and supplementary budget, rejected the latter, and at the same time asked the Commission to make transfers for the funds which the Council considered to have priority, such as the Social Fund).

(資料1 1)

翻訳者による校正例

Sir Derek EBBW, on behalf of the producers, proposed that Sir Derek EBBW should put forward the name of Mr André Robert. After recalling that the last Luxembourg presidency had been held by the candidate, Mr Robert, he wished to pay tribute to the success of Mr André Robert. (applause)

MR ANDRÉ ROBERT was elected President of the Committee (unanimously).

THE CHAIRMAN invited the producers to elect the members of the committee. He recalled that the producers had elected Mr Cassmann and Mr Jueth as Vice-Presidents of the committee.

MR ANDRÉ ROBERT and Mr Jueth were elected (unanimously) Vice-Presidents of the committee.

THE CHAIRMAN invited the committee to elect the members of the Bureau. He recalled that the producers had elected Mr Cassmann and Mr Jueth as Vice-Presidents of the committee. It was necessary to elect three producers: two members and the President.

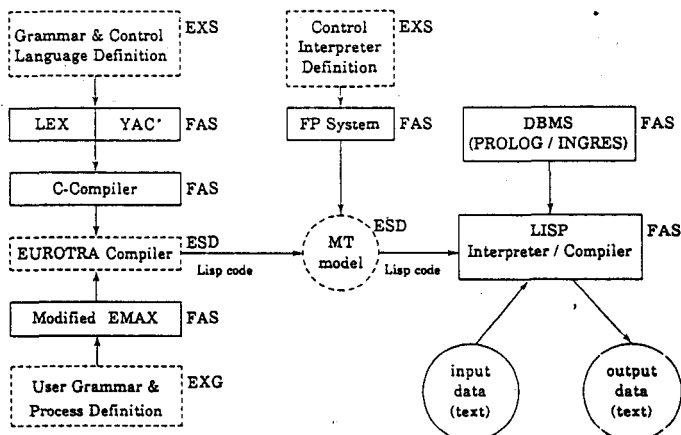
MR EBBW, on behalf of the producers, proposed that Sir Derek EBBW should put forward the name of Mr André Robert and that under the rules of the committee the producers should elect Mr Robert as President of the Bureau. He recalled that the producers had elected Mr Cassmann and Mr Jueth as Vice-Presidents of the committee.

(資料1 2)

ECの翻訳者が実際に自分の手で修正しました。これはどちらかと言うと修正箇所が少なめだと思えますが、このような形で修正します。機械が翻訳した結果はワープロに入っていますので、翻訳者は修正したのをワープロに打ちこんで最終的な翻訳とします。

次にECが1982年11月から5.5ヶ年計画で開始しました新しい機械翻訳システム開発プロジェクト ユロトラ (EUROTRA) の概要を簡単に説明しましょう。このユロトラは、1978年から予備研究を開始していましたが、実際にECの閣僚が決定したのが1982年の11月です。そこから全体を3つの期に分けてこのプロジェクトを進める予定です。第1期が1984年の11月で、次の2年間で2500の語彙を持ったシステムを開始する計画を持っています。そしてそれを基にし、2万語の語彙を持つ中規模のシステムを1988年の中頃までに開発し、民間会社が実用システムを作り上げる予定になっていますが、非常に大変なことです。ECの加盟国は7ヶ国ですのでユロトラは7つの言葉の間で任意の翻訳ができるのを目標としています。しかし新たに2ヶ国加盟国が増えましたので、合計9ヶ国語の間ですべての翻訳を行わなければなりません。これはECの使命です。ちなみにECというのは欧州共同体ですが、年間予算の半分は翻訳に使っているといわれています。これはすごいことです。日本政府が翻訳にかけているお金はごく少なく、このプロジェクトは約1100人の言語学の専門家が中心になって各国の大学が話をすすめているわけです。現在ソフトウェアの基本設計が終わっているはずですが、例えばこれは VAXの11/780を使って点線部分を作り、実線部分は VAX上にありますUNIXの機能を利用して行っています。

### EUROTRAのソフトウェア構成



EXS: External Specification of Software System  
 EXG: External Definition of Grammars  
 ESD: Eurotra system devices  
 FAS: Facilities Available on UNIX

実線の部分はすでに動いている(計算機はVAX11/780) 図中、左側がEurotra System言語のシンタックス部分の定義、中央がそのセマンティクスの定義、右側が実際の翻訳実行系

(資料13)

次に非常に典型的な別のシステムを御紹介します。

先程ティタス (TITUS) のシステムを申し上げましたが、これはももとは翻訳システムではなくて文書作成システムだったのです。この会社はフランスのパリにありますが、フランス語、イギリス語、英語、独語、スペイン語で同じ内容の文書を作る必要があったということです。文書をどれか一つの言語で入れて、それに非常に単純な処理、生成文法を使って処理が行われます。それを解くピボット言語は4つの言語のどれにも依存しない言語で、これは計算機の中のビットパターン、つまりアセンブラで書かれていのでとてもピボット言語といえるものではありません。それをファイルとしてとっておき、プリントしたいと思ったら非常に高速に4つの言語での文がさっと出てくるシステムで、1970年の始めからやっています。これには2つの制限があり、1つは限定語彙、つまり語彙の数を制限して使用できる単語が決まっています。もう1つは、文法制限です。これは制限言語といって、この文法にかなう文しか入れませんというように徹底しています。このシステムは、タイプインしかできません。原稿を入れる時、タイプで打たなくてはならず、タイプをしながらシステムがモニタリングします。文法にかなわない文が入ってきたならば、システムが"ピー"と鳴り修正しなければいけないので、ワープロで文を入れた時は正しい文になっています。そうでない限り動作しないというシステムです。

先程のBYUシステムについて御紹介します。BYUのメルビィという人は1970年ぐらいから機械翻訳を行っていますが、徹底していて機械翻訳というのは出来ない、出来るはずがないといっています。だいたい1つの文に2ヶ所以上まちがいがあって、2ヶ所以上人間が修正しなければいけない様な訳文を出す機械翻訳システムは使えないというのがここでやっている人の持論です。そこで、援助システムを徹底したのが3つのレベルです。第1レベルは完全にワープロですが、専門用語が計算機の中にはいっていますので、それを検索する機能は用意するもので、つまりワープロ+辞書です。第2レベルは辞書で、辞書に一応英和、和英を使えば訳語が書いているか、どの訳語が正しいのかわかりませんが、それをディスプレイ上で見せ、この中から人間が選びます。第3レベルは理想ですが、機械翻訳システムが一応やってくれて、人間が翻訳されたのを見て修正するものです。だからこの人は、機械翻訳を使うために3つのレベルでのソフトウェアシステムを提供しますという形でやっています。私共のところでは、機械翻訳そのものは研究の目的ではなく、計算機で意味理解をさせることをやっています。一応機械翻訳システムをツールとして作っていて、その名前をLUTEといっており、これはLanguage Understand Translator Editorの略で「言語の意味を理解し、変換し、編集する」ということです。LUTEというのは楽器の音色のことで、私共は日夜、楽器の音色をいかによくするかということを追及しています。私自身、人工知能の研究をやっていますが、約10年前に「それは意味理解、problem-solving (問題解決)、knowledge-learning (知識学習)

の3つの基本的な問題である」という前提の基で研究を進めています。その中で意味を理解するというのは、やはり1つの最も基本となることです。とりあえず、計算機に意味を処理させるというのは何かという研究を進めており、そのアプリケーションとして機械翻訳を考えています。それを進めるためのもう少し具体的な課題ですが、まず意味（セマンティックモデル）というものをとらえます。そのためには言語学的な話もあり、ワールドモデル、これはコミュニケーションするときのお互いの間に仮定される知識、背景、状況、又、知識を頭の中で記憶していくためには、計算機の中で表現しなくてははいけませんので、この表現形式としてのメモリーモデル、以上の3つが基本となります。セマンティックモデル、これをデータベースの方へ応用したものをノーリッジベース識別といますが、識別には言語内に存在する言語知識、三つ以上の言語間に存在するinter-linguistic knowledge、それから言語そのものではなく、いわゆる言語外の知識（extra-linguistic knowledge）で、これがknowledge baseを構成する処理では解析・変換を行います。こういうものを集めると意味を理解するシステムができ、それができるとアプリケーションとしてquestion-answering（質問応答）ができ、テキストプロセッシングワープロより高度な機能というところへ応用できるでしょう。そのknowledge baseのところがあって、ソーステキストのプロセスとトランスフォーメーション、ジェネレーション訳語が生成されますが、knowledge baseとして解析辞書、ジェネレーション辞書、文法、意味的な関係だとか、ストラクチャーだとか、それをリンクしているデータ、それから人間の一般的な常識、専門知識、それから2つの言葉の間を概念のレベルで使う lexical transfer dictionaryと、言語の構造のレベルで結びつけるstructural transfer dictionary、ある単語が表わす概念というのは言語により異なり、ずれがあります。このずれがなければ非常に簡単ですが、構造ももちろん違い、複雑な関係をセマンティックなリレーションとして辞書の中で表わしています。例えばある文を解析した時に計算機の内部に、文の意味的な構造として作られます。これは、「子供が成長すると、大人が持つ技能を持つことができる」という文を解析し、計算機内部に表現されているのを非常に単純化します。

私たちのシステムではこういう風に日本語→英語の場合、「1つの概念を確立した子供がその問題を解くことができる」という入門的な文に達し、日本語→英語、英語→日本語で翻訳ができる状態になっています。我々は、こういうのを進めるために辞書をつくるということで、マンマシンインタフェースというのが非常に重要となってきます。それから人間の能力は限られていますから、人間が一見みわたして一度で判断できる量的なものには限界があります。その能力を増長しない限り、言語は非常にたくさんのデータを使わなければなりません。

複雑で多様なものが使われていなければならないが、人間の頭脳はやはり能力が限られているので、それを手助けしてくれるシステムが必要だという観点からいろいろな機能を持ったサポートシステムを作りながら、我々は研究を続けています。以上です。

（昭和59年12月6日講演）