ISSN 1883-1818

AAMT

Asia-Pacific Association for Machine Translation

Journal

June 2017 No.65 Special Issue

アジア太平洋機械翻訳協会

機械翻訳事始め

坂本義行
アジア太平洋機械翻訳協会理事

本別冊の発行にあたって

本別冊は、坂本義行先生からAAMTジャーナルNo.53（2013年6月）以来No.64（2017年3月）までご寄稿いただき、「機械翻訳事始め」として掲載してきたものを、別冊としてまとめ再収録したものである。

アジア太平洋機械翻訳協会ジャーナル編集委員会

目次

はじめに ・・・・・・・1

機械翻訳事始め I （AAMT ジャーナル 53 号、2013 年 6 月） ・・・・・・・2

機械翻訳事始め II （AAMT ジャーナル 55 号、2014 年 3 月） ・・・・・・・9

機械翻訳事始め III （AAMT ジャーナル 56 号、2014 年 6 月） ・・・・・・・17

機械翻訳事始め IV （AAMT ジャーナル 57 号、2014 年 10 月） ・・・・・・・25

機械翻訳事始め V （AAMT ジャーナル 59 号、2015 年 6 月） ・・・・・・・33

機械翻訳事始め VI （AAMT ジャーナル 61 号、2016 年 2 月） ・・・・・・・41

機械翻訳事始め VII （AAMT ジャーナル 62 号、2016 年 6 月） ・・・・・・・47

機械翻訳事始め VIII （AAMT ジャーナル 63 号、2016 年 10 月 ・・・・・・・60

機械翻訳事始め IX （AAMT ジャーナル 64 号、2017 年 3 月） ・・・・・・・65

おわりに ・・・・・・・72

はじめに

十数年前、協会の理事に推薦され、その中で歴史担当を仰せ付けられました。そして、2013 年の 6 月号から、4 年ほどかけて、9 回にわたって、日本を中心に世界の機械翻訳の開発を紹介してきました。

計算機という言葉が最初にもちいられたのは、1645 年に、かの有名なパスカルによって発明された機械式計算機、Pascaline と言われています。そして 1923 年に世界で初めて、翻訳機械の特許がロシア人のトロヤンスキーから出され、1950 年代に、旧ソ連や米国で、本格的に電子計算機を用いて機械翻訳の開発が始まりました。

日本でも、1950 年代後半から、九州大学の KT－1 や電気試験所の機械翻訳機「やまと」などの開発がスタートしています。私は、1965 年に電気試験所の入所し、COBOL を用いたプログラムによる日英の機械翻訳システムの開発に従事しました。

また、「やまと」については、1959 年にパリの第 1 回国際情報処理会議で、和田氏によって報告されており、米国とならんで、機械翻訳の開発拠点になってきました。

1960 年代から、日本では、計算機の高速化、小型化、記憶容量の高速・大容量により、1961 年には、情報処理学会に「機械翻訳研究委員会」が設立され、文書処理にも、文字の認識、要約機能、文献検索の研究・開発がすすめられました。また、日米間で、日米機械翻訳セミナーが開かれるようになりました。

1966 年に、米国政府から出された ALPAC レポートは、世界に衝撃を与えました。それは、米国という背景とも関係があるが、「機械翻訳の開発は必要ない。これからは、自然言語の研究が重要であり、マシンエイデッドな翻訳システムの開発が実用的である。」というものでした。この報告を受けて、日本では 60 年代から 70 年代にかけて、政府が「機械翻訳」に関するプロジェクトに予算をつけることはありませんでした。

1982 年から 4 年間、科技庁予算による本格的な機械翻訳システムの開発「Mu システムの開発が行われ、官民が協力して、行われました。その成果は、言語処理システム、辞書作成・管理システム、総合システムとして開発され、民間で商用システムとして、製品化されました。

1987 年から 8 年間をかけ、通産省の予算、60 億円をかけ「近隣諸国間の機械翻訳システムに関する研究協力」が、日本、中国、インドネシア、マレーシア、タイの 5 カ国を対象として、相互の言語に翻訳可能なシステムの開発が行われました。

さらに、1985 年から 5 年間をかけ、特許庁関連の情報に関する「日英自動翻訳システムに関する総合調査」が行われ、特許自動翻訳システムの将来構想として、知的所有権をめぐる国際動向と機械翻訳、特許総合情報システムの将来像について、調査と報告がなされました。

1.1 ことばの発生

ことばについて、「翻訳機械」[1]の訳者まえがきによると「旧約聖書によれば、創世のむかしは人類の言語は一つであった。それが現在のようにいろいろの民族が自分たちで通用する言語を話し、異民族相互間の意志の疎通をかくようになったのは、天までとどくことを意図して建設をはじめたいわゆるバベルの塔が神の怒りにふれたためであるという。

各民族が独自の言語を持って活動している現実の事態は、それぞれに固有の文化を築いていけるという大きな利点がある反面、相互の思想、意志の交流疎通という点で実用上多大の不便をともなう。この障害を軽減する手段として、だれにでも思いつく方法は即応性のある組織的な大規模の"翻訳"機関を常備することであるが、人間の翻訳者に依存する限り、このようなものの実現可能性はないと考えられる。したがって、翻訳者をわずらわすことなく、翻訳を高速でやってのける機械というものは、われわれの空想の中に生ずる一つの大きな夢であった。

この企てを具体的に示した最初のものはペ・ペ・トロヤンスキーが 1923 年に出した翻訳機械の特許であろう。」と述べられている。

1.2 機械式計算機

計算機といえば機械式計算機があり、その代表としてパスカルの機械式計算機があり、これについては、ウイキペディアによると[2]、ブレーズ・パスカルが 1645 年に発明した機械式計算機。Pascaline（パスカリーヌ）または Machine Arithmétique と呼ばれている。

パスカルは 1642 年、彼が 19 歳のときから計算機について研究を始めている。徴税官だった父の手伝いをしていたパスカルは、仕事を減らすための道具を作ろうと考えた。1642 年までにパスカルは 50 台もの試作機を作ったが、売れたのは 1 ダース強である。高価であったことと複雑であったこと（加えて加減算しかできず、減算のやり方が難しかった）が原因で、それ以上売れることはなく、パスカルは 1652 年に計算機の製作を止めた。

Pascaline は十進数を使った機械である。しかし、当時のフランスの通貨は十進数ではなく、イギリスのポンド、シリング、ペンスと似ていた。従って、金額を計算するのに Pascaline を使おうとすると、計算結果を更に変換する必要があった。1979 年フランスはメートル法に切り替えた。このとき、パスカルの基本設計に触発された職人が登場したが、彼らも商業的には成功しなかった。

最初の Pascaline は 5 個のダイヤルがあり、後には 6 ダイヤルや 8 ダイヤルのものが作られている。最大のもので 9,999,999 までの数値を扱うことが出来た。各ダイヤルは数値のうちの 1 桁に対応し、計算結果は上部の窓に表示される。歯車は一方向にしか回らないため、負の値を直接計算することはできない。減算をするには 9 の補数表現にして加算する必要があった。ユーザーを助けるため、9 の補数を入力すると上部の窓には元の数値が表示されるようになっている。

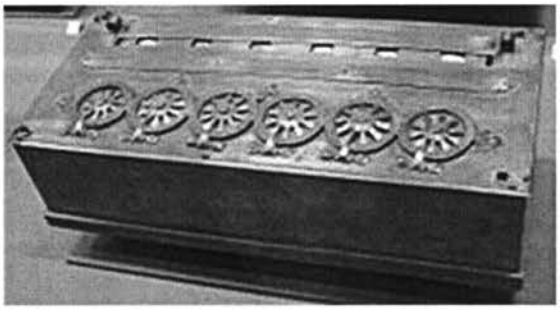


図 1 Pascaline

1.3 トロヤンスキーの発明特許証書

ペ・ペ・トロヤンスキーの発明特許証書については、玉木英彦・気安善市編「自動翻訳」[3]によると、

「以下のような特許が出願されているが、当時それを実現することはうまくいかなかった。これは、この目的に適した自動装置がまったくつくれていなかった、当然のことである。」と書かれている。

また、文献[1]によると、

「この翻訳機は、同時にいくつかの言語を翻訳し、遠方まで送ることを可能にするように見えた。しかしトロヤンスキーは 1939 年に依頼したソビエトの言語学者たちの支持も得られなかったし、1944 年には科学

アカデミーの自動化と遠隔制御機構学会の支持も得られなかった。」と書かれている。

発明特許証書

発明の名称 一つの言語から他の一つ、または同時に多数の言語への翻訳に際し、単語を選別しかつ印刷する機械

特許出願人 ペ・ペ・トロヤンスキー

出願年月日 1933年9月5日

Класс 15g, 20; 57d, 1

№ 40995

АВТОРСКОЕ СВИДЕТЕЛЬСТВО НА ИЗОБРЕТЕНИЕ

ОПИСАНИЕ

машины для подбора и печатания слов при переводе с одного языка на другой или на несколько других одновременно.

К авторскому свидетельству П. П. Троянского, заявленному 5 сентября 1933 года (спр. о перв. № 134430).

О выдаче авторского свидетельства опубликовано 31 января 1935 года.

Предлагаемая машина предназначена для подбора и печатания слов при переводе с одного языка на другой или на несколько других, одновременно и состоит, главным образом, из передвиж-

реводимому слову передвигают ленту так, чтобы соответствующее слово или ряд слов на нескольких языках установились точно против отверстия б, затем ленту затормаживают и спускают затвор

図2 発明特許証書

1.4 暗号解読から計算機の技術へ

第2次大戦中に発展した暗号解読の技術（文献1では）を基礎として、電子計算機が急速に実用化され出した1946年頃から研究が始められた。もちろん当初W.ウィーバーが考えていたような《翻訳とは、外国語という暗号で書かれた文章を母国語に解読することである》という単純な考え方だけでは通用しないようになっている。

英国人 A.D.ブースとロックフェラー財団のワレン・ウィーバーとが同時に翻訳問題に手を付けたのは1946年のことであった。第二次大戦中に用いられた暗号解読技術をこれに適用しようとウィーバーが提案したのに対し、ブースはディジタル計算機が十分な数のデータを蓄えることができるので、《辞書と首っ引きで》行えるような、単語を単語に移す型の逐語訳を実行する能力があることを強調した。文章論や語順やあるテキストのすべての単語の機械的な翻訳についてすら問題にならなかった。というのは当時の考え方は外国語のテキストを理解するために、学者の目の前に翻訳された主要な単語の系列を置けば、この題目に関する読者の深い知識により彼が手探りで自分の道を見つける役に立つであろうということであった。

プリンストンにおいてブースはブリテンとともに、計算機がその記憶装置中に記録された辞書を引き、かつ一本の穿孔テープで入力に導かれる一連の単語の逐語訳を与えることを可能にする《命令》について研究した。

1948年に英国人リチェンズは、語尾に自動文法解析の考えを主張した。このことは単に読者に単語の文法的役割を説明することで自動翻訳を改良するのみならず、電子辞書における探索の速度を高めるというのは、理論的にみて見出し語の全数を減少させうるからである。

1949年ウィーバーは言語ごとに見かけ上の相違はあるが、暗号術で認められ、かつ情報理論により認められている統計的不変量や、中国語学者 E.ライフらが歴史的つながりのない言語のあいだに認めた意義的不変量や、さらにライヘンバッハにより指摘された論理的不変量が存在することを強調した。こういった不変量は、人間の頭脳の基本性質の存在と、言語に共通な心理社会的な起源とに相応する、ブースとリチェンズとの仕事を参照して、ウィーバーは、単語に単語を当てる純粋の逐語訳はその性格上科学技術の研究にとって大きな役割を果たすであろうと主張した。しかも彼はずっと先に進み、直接の文脈を調べることで、意義のあいまいさを解決することを強く主張した。

言語の論理的要素は、計算機の論理回路で処理できると断言した。すなわちシャノンの情報理論は翻訳の問題に関し、統計的組織面で光を投げかけるのであって、彼はこの理論に照らして統計的語義論の研究を企てることを特にすすめた。おわりに彼は通信の道具である言語自体の性質に関する研究の問題を生彩のあるやりかたで提出した。

研究はいくらか分散しながら多方面に分散していった。オスワルドとフレッチャーはドイツ語の文章論構成の機械的解法を研究した。マサチューセッツ工科大学（M.I.T.）では、イスラエルの論理学者バー・ヒレ

ルが機械翻訳の目的で言語の研究に没頭した最初の人である。1952年の初めに、ロックフェラー財団はM.I.T.主催の翻訳問題に専念している言語学者と電子工学者との会議に財政的援助を行った。18人の参加者は次の二つの段階を熟考することに同意した。まず第一に、科学的題目を取り扱うテキストについて単語の頻度の研究、異なる言語間の等価単語の研究、電子記憶装置を用いる方法についての調査ならびに自動辞書のその他の技術的様相の調査を同時に行うことであった。その後で機械翻訳をするプログラムを作る目的での文章論解析に移ることができよう。文法的および文章論的な諸問題を解決するために必要な回路の研究は、当分の間さきのこととしてすぐに手を付けなくてもよい。多言語機械翻訳をつくることは、ある一つの言語Aからもうひとつの言語Bへ行う一方向の翻訳の経験から得られる結果をみてからすればよい。しかしながら、媒介言語、いわゆる《機械語》すなわち最終的にはあらゆる言語のあいだの回転板として役立ちうるものの可能性は除外されなかったのである。

1.5 1954年

前述の見解に刺激されて米国の研究が発展した。記憶、容量、語彙を制限することの効用、意味ならびに語尾の機械的確認などの研究がなされた。

1954年にジョージタウン大学のドスタートとガルビンとは、I.B.M.のシェリンダンと協力して250個の単語からなる辞書と、6個の文章論の法則を用い、IBM701電子計算機を使ってロシア語から英語への自動翻訳の仕事を首尾よく成し遂げた。

1.6 電子計算機 BЭCM[3)]

1955年の末に電子計算機で行った自動翻訳の最初の実験では、この実験用に専門的な数学のテキストの翻訳を考えて、あらかじめ952の英単語と1073のロシア単語を選んで辞書を作り、数学のテキストの翻訳のために使用されうるプログルラムを作った。

INTRODUCTION

When a practical problem is science or technology permits mathematical formation, the enhances are rather good that is leads to one or more differential equations. This is true certainly of the vast category of problems associated with force and motion. So that whether we want to know the future path of Jupiter is the heavens or the path of an electron in an electron microscope we resort to differential equations. The same is true for the study of phenomena in continuous media, propagation of waves, flow of heat, diffusion, static or dynamic electricity, etc., except that we here deal with partial differential equations.

図3 ミルヌの本の「序文」の最初の一節

ВВЕДЕНИЕ

ЕСДИ ПРАКТИЧЕСКАЯ ЭАДАЧА В НАУКЕ
ИЛИ ТЕХНИКЕ ДОПУСКАЕТ
МАТЕМАТИЧЕСКУЮ ФОРМУЛИРОВКУ,
АНСИ ДОВОЛЬНО ВЕЛИКИ ЧТО ЗТО
ПРИВОДИТ К ОДНОМУ ИЛИ БОЛЕЕ
ДИФФЕРЕНЦИАПЬНЫМ УРАВНЕНИЯМ,
ЭТО ВЕРНО БЕЗУСЛОВНО ДЛЯ ОБЖИРНОЯ
КАТЕГОРИИ ЗАДАЧ СВЯЗАННЫХ С СИЛОЙ И
ДВИЖЕНИЕМ,
ТАК ЧТО ХОТИМ ЛИ МЫ ЭНАТЬ БУДУЦИЯ
ПУТЬ ЮПИТЕРА В НЕБЕСАХ ИЛИ ПУТЬ
ЭЛЕКТРОНА Э ЭЛЕКТРОННОМ МИКРОСКОПЕ
МЫ ПРИБЕГАЕМ К ДИФФЕРЕНЦИАПЬНЫМ
УРАВНЕНИЯМ. ТО ЖЕ ВЕРНО ДЛЯ ИЗУЧЕНИЯ
ЯВЛЕНИЙ В НЕПРЕРМВНОЙ СРЕДЕ,
РАСПРОСТРАНЕНИЯ ВОЛН, ПОТОКА ТЕПЛА,
ДИФФУЗИЙ. СТАТИЧЕСКОГО ИЛИ
ДИНАМИЧЕСКОГО ЭЛЕКТТРИЧЕСТВА, И Т Д,
ЗА ИСКЛЮЧЕНИЕМ ТОГО ЧТО ИЫ ЭДЕСЬ
БУДЕМ РАССМАТРИВАТЬ ДИФФЕРЕНЦИАЛЪНЫЕ
УРАВНЕНИЯ В ЧАСТН ЫХПРОИЗВОДНЫХ.

図4 そのロシア語訳をタイプで印刷した形

プログラムは実際上辞書には関係なく適用できるものである。翻訳のためにミルヌの《微分方程式の数値解》から断章を選んだ。

図3に入力として「序文」の最初の一節の英文テキストを、図4に出力として語訳をタイプで印刷した形で示した。

1.7 最初の機械翻訳に関する国際会議[1)]

1956年10月にM.I.T.は、機械翻訳に関する最初の国際会議を招集した。約30人の英国、カナダ、米国の専門家達が集まり、ソ連科学アカデミーのD.J.パノフ博士は手紙でソビエトの研究についての重要な報告を行った。三つの主要な活動の中心は、米国、英国、並びにソ連であり、これらのものより規模の小さいグループはイタリアとスカンジナビアであった。問題は

もはや翻訳機械が可能であるということを研究成果によって確かめるということでなく、努力を集中して研究能率を有効にし、ばらばらに行われている仕事を総合的見地で研究を組織化することにあった。

1957 年 5 月にモスクワで科学技術会議が開かれ、56 論文が提出され、それらの中には、ドイツ語、英語、ハンガリー語、中国語からの自動翻訳に関するものや、フランス語からの実験的翻訳に関するものがあり、さらに主として数学的な諸方法を利用して、この学問を精密科学にするようなやり方で言語の研究を方向付ける必要性を立証した。

1958 年 5 月に、自動翻訳についての会議が開かれ、その結果 71 の提出論文の要約が刊行された [1]。

1958 年頃の機械翻訳の研究 [1]

1958 年には米国で、1 ダースばかりのグループが大小規模の活動的な研究がなされていた。

1.7.1 米国では

＊ハーバード大学では、エティンガーが露英自動辞書に関する研究を行っていた。

＊M.I.T.では、ロック、ユングブ、コムスキー等が文章構造、ドイツ語の文章論、方法論を研究していた。

＊ジョージタウン大学では、ドスタートとガルビンがロシア語の文章論と語彙論の研究を行っていた。

＊ミシガン大学では、クートスーダとコーフヘージがロシア語ならびに多意義語の問題を研究していた。

＊ワシントン州立大学では、ライフラー、マイクルセン、ヒルがロスアンゼルスの国際テレメータ会社と共同で、ロシア語とドイツ語の研究をしていた。

＊ランド会社では、ヘイとエドマンソンとが方法論に関する興味ある研究に従事している。

ワシントンの国立科学財団は、これらの研究に財政的援助を与え、これらの研究の調整に努力を払った。

1.7.2 英国では

＊ブース、ブランドウッド、クリーブがスフィールド財団の援助を受けて、バークベック単科大学で機械翻訳の方法論を研究し、さらに点字やフランス語やドイツ語についての仕事、ならびに翻訳機械の特性と原価に関する仕事を追及した。彼らは 1958 年に 'Mechanical Resolution of Linguistic Problems'[4] という著作を発行したが、これは言語データとアイデアとの豊かな著作であった。

＊ケンブリッジ言語研究グループでは、マスターマンが指導して、辞書編集と普遍文法論との研究を進め、辞書と文章論とに《機械辞書的熟語辞書》すなわち理念で分類した機械辞書の考えを適用することを探究した。

1.7.3 ソ連では

1956 年 9－10 月より刊行された『言語学の諸問題』は自動翻訳の問題に定期的な題目を提供している。

＊ソ連科学アカデミーの研究チームは、言語学者と電子工学者との間の協力についての指導原理と一般的な方法論との輪郭を明らかにした。

＊メルチュウクとクラギナ女史とは、フランス語からロシア語への翻訳に関する非常に興味深い研究を公表した。

＊コロレーフは辞書構成のための符号圧縮の問題について研究した。

＊ラズモフスキーは翻訳機械の自動プログラミングについて研究した。

＊モロシュナヤ女史はイエスペルセンとフリーズとの構造言語学上の仕事を種々の言語間の文章論的差異の解決に使用する方法を示した。

1.7.4 日本では

1950 年代に電気試験所機械翻訳「やまと」や九州大学の KT-1 が開発された。

1.8 機械翻訳機「やまと」[5]

電気試験所（現産業技術総合研究所）では、1958 年 1 月より和田弘（当時電子部・部長）の指示により、高橋茂を中心に、トランジスタ電子計算機 ETL Mark IV の基本論理回路を使用して、英日翻訳専用機「やまと」を試作することが計画された。翻訳のための辞書を収容

できる大容量記憶を持った電子計算機が手許になかったため、専用機として600万円の予算で製作された。

装置の製作は1958年11月に終了した。

基本設計

辞書: 英単語 3000語　150,000bit

=3,000語×6字×8bit

和単語 3000語　500,000bit

熟語 300語　20,000bit

その他　350,000bit

文字: 英文; ABC, abc, 記号、数字

和文; カタカナ、ABC、記号、数字

プログラム: 機械語 4,000語

Add Numeric, Sub Numeric, Add Chr,

Read in to AC,Store Word in VM,

Write Mark. Atop

プログラムメモリ 2ブロック 8,192語

1語 32ビット、1トラック=128語、

1ブロック=32トラック=4,096語

計算機の構成

本体: 基本回路の3種のプラグイン約1,000枚が用いられた

記憶装置: 磁気ドラム（電気洗濯機位の大きさ）

回転数 約3,000r.p.m（平均待ち合わせ時間は約 10ms）の低速型磁気ドラム 1台、1周 4,096bit

記憶容量 約82万bit=820,000bit=約100kB

回路: トランジスタ 500個、ダイオード 700個で構成

入出力装置:

1. 機械式テープ読取機（8単位） 10字/秒
2. プリンタ　8字/秒
   ABC、カタカナ、記号、数字
3. パンチャ　8字/秒
   ABC、abc、記号、数字
4. 万能入出力装置　8字/秒
   カタカナ、記号、数字

このうち、3と4は計算機と独立で働かすこともできる。

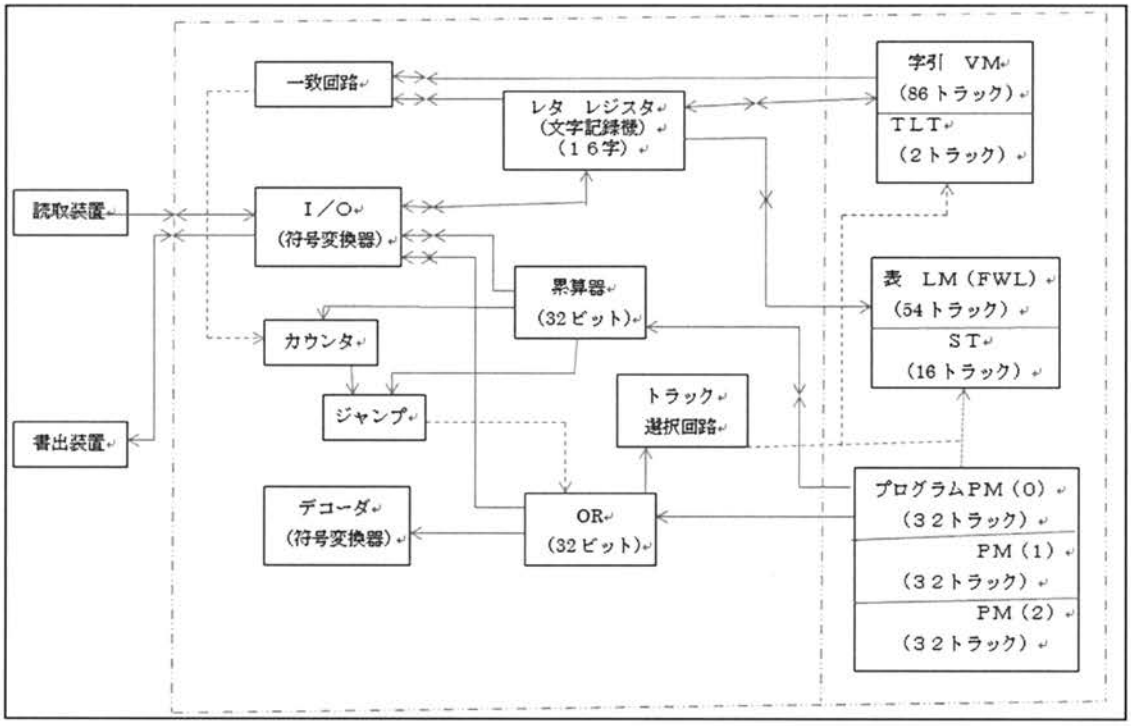


図3 「やまと」の計算機ブロック図

蓼沼良一を中心に、英文和訳の研究に着手し、研究の課題として、

1. 翻訳に適した機械装置を研究すること
2. この装置を用いて翻訳させること

であると記されている。

1959 年 2 月には「I like music」とパンチした紙テープを入力すると、「ワレガ オンガクヲ コノム」とプリンタに出力することができるようになった。

本研究の成果については、1959 年 6 月にパリで開催された第 1 回国際情報処理会議で和田氏によって発表された。

入力の英文	出力できた和文
1. This is a pen.	1. コレガ ペン ダ
2. Is this a pen?	2. コレガ ペン ダカ
3. Yes, it is a pen.	3. ハイ ソレガ ペン ダ
4. No, it is not a pen.	4. イイエ ソレガ ペン デナイ
5. What is this?	5. コレガ ナニ ダカ
6. This is my watch.	6. コレガ ワシノ トケイダ
7. It is a blue pencil.	7. ソレガ アオイ エンピツ ダ
8. Is that a window, too?	8. ソノウエ アレガ マド ダカ
9. What is it, then?	9. ソレカラ ソレガ ナニ ダカ
10. I have a pen.	10. ワレガ ペンヲ モツ
11. Have I a pen?	11. ワレガ ペンヲ モツカ
12. What have I?	12. ワレガ ナニヲ モツカ
13. I have two arms.	13. ワレガ 2 ウデヲ モツ
14. I have a camera in my hand.	14. ワレガ カメラヲ ワレ テ ノナカニ モツ
15. What have you in your hand?	15. ナレガ ナニヲ ナレノ テ ナカニ モツカ
16. But I have no brother.	16. シカシ ワレガ ナイ アニヲ モツ
17. Who is that gentleman?	17. ソノ シンシガ ダレ ダカ
18. He is Mr. Smith.	18. カレガ ミスター SMITH ダ
19. He is Betty's father,	19. カレガ BETTY ノ チチ ダ
20. He has some eggs.	20. カレガ イクラカノ タマゴヲ モツ
21. Is that girl your sister?	21. アノ オンナノコガ ナレノ アネ ダカ
22. The rooster is black.	22. オンドリガ クロイ ダ
23. There is a desk in the room	23. ソコニ ヘヤア ノナカニ ツクエガ アル
24. Is there a desk in your room?	24. ソコニ ナレノ ヘヤ ノナカニ ツクエガ アルカ
25. Where is the vase?	25. ドコニ カビンガ アルカ
26. There is a picture of my family.	26. コレガ ワレノ カゾク ノ エ ダ
27. I do not speak French.	27. ワレガ フランスゴヲ ハナサ ナイ
28. Do I speak English?	28. ワレガ エイゴヲ ハナスカ
29. Thank you.	29. アリガトウ
30. It is half past four.	30. ソレガ 4 ヲスギテ ハンブン ダ
31. It is twenty minutes past six.	31. ソレガ 6 ヲスギテ 20 フン ダ
32. We have breakfast at seven.	32. ワレガ アサゴハンヲ 7 ニオイテ モツ
33. Then I go to school.	33. ソレカラ ワレガ ガッコウ ニ イク
34. School begins at half past eight.	34. ガッコウガ 8 ヲスギテ ハンブン ニオイテ ハジマル
35. What do the boys do?	35. オトコノコガ ナニヲ スルカ
36. He sometimes helps me with my lessons.	36. カレガ ワレヲ トキドキ ワレノ カ トトモニ タスケル
37. Can you swim?	37. ナレガ オヨグ コトガデキルカ
38. Yes, I can swim.	38. ハイ ワレガ オヨグ コトガ デキル
39. I am walking.	39. ワレガ アルキ ツツアル
40. Am I walking?	40. ワレガ アルキ ツツアルカ
41. I am reading a story.	41. ワレガ ハナシヲ ヨミ ツツア カ
42. How do you do?	42. イカガ デスカ
43. He is the fastest runner in our class.	43. カレガ ワレラノ クミ ノナカニ モットモ ハヤイ ハシルモノ ダ
44. Come to the blackboard.	44. コクバンニ クルベシ
45. Write the names of the seven days of the week.	45. シュウ ノ 7 ヒ ノ ナマエヲ カクベシ
46. Let's have a raced.	46. キョウソウヲ モツ コトヲシヨウ ベシ
47. The crow forgot her cheese.	47. カラスガ サレノ CHEESE ヲ ワスレタ

機械翻訳「やまと」で翻訳できた例

1.9 KT-1[6)]

KT-1 は、我が国初の日・英・独 3 カ国語相互実験翻訳のために、九州大学が 1950 年代後半に開発した言語処理用計算機である。言語理論は直接構成要素法を採用し、処理方式は各国語特有の性質を表にまとめ、プログラムは各国語間のどの方向の翻訳にも共通にした特徴あるものであった。KT-1 ではおもに英語、ドイツ語、日本語間の操作について、原語の構文解析により原語を一度中間言語的な構文情報に変換する方法で、設計試作が行われた。KT-1 では 2 進法 6 チャネル（直列には 64 進法）の直並列方式をとり、単語は直列 8 桁を単位とした。検索命令やパターンマッチに工夫を

凝らし、文字列演算に優れていた。

論理回路にはトランジスタによるダイナミックフリップフロップ方式を採用した。記憶装置は記憶容量約10万ビット・回転数12,000回転/分の磁気ドラムを使用した。入出力には6単位符号を採用し、鑽孔紙テープによる方法を採った。

KT-1の設計は九州大学、製作は三菱電機が行った。単能の翻訳機ではなく、一般技術計算用機能を持たせ、その利用も考慮されていた。

現在、KT-1に使用された論理パッケージ、磁気ドラム、および入力データの鑽孔紙テープが、九州大学に保存されている。

なお、元九州大学、田町研究室の岡田直之氏の記憶では、次のことは事実です:

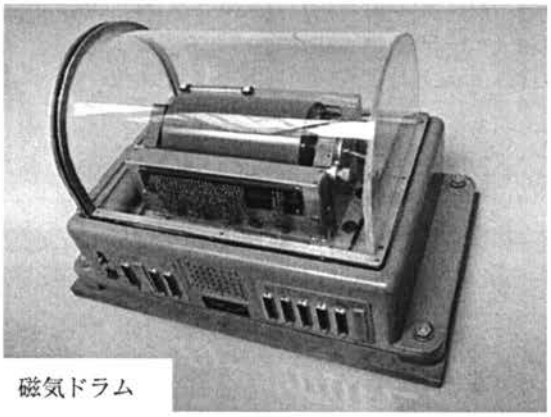
＊自動翻訳機 KT-1 の設計は、大野克郎、栗原俊彦、田町常夫の3教授が行った。

＊翻訳プログラムは、ほぼ田町常夫教授が担当した。

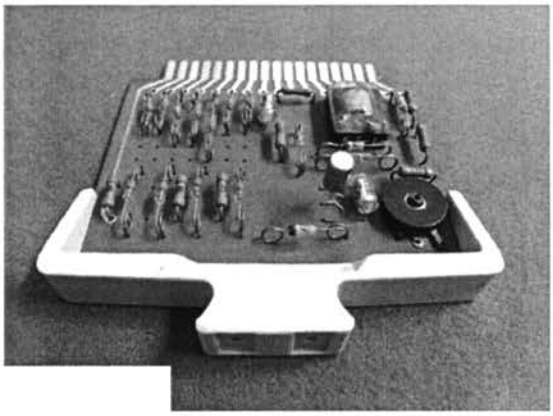
＊10万ビットの主記憶に単語辞書、文法規則、それに翻訳プログラムを機械語を用いて配列した.

＊栗原の意志をつぎ吉田将は、単語と単語のつながりに意味が反映されるという立場で、大規模言語データの収集やかな漢字変換に努めた。

＊田町の翻訳プログラムを発展させ岡田直之は、言語や図形の持つ意味に注目し、語彙の意味分析や言語・図形翻訳などマルチモーダルな人工知能に進んだ。



磁気ドラム



製造者　九州大学
所有者　九州大学
史料所在地
〒819-0395 福岡県福岡市西区元岡 744
九州大学工学部百周年記念展示室『CENTENNIAL GALLERY』
公開情報 公開中

参考文献

1) エメール・ドラヴネ著、別所照彦、沢辺弘共訳：「翻訳機械」、文庫クセジュ、白水社 1964 年。

2) ウィキペディアから機械式計算機より、(http://ja.wikipedia.org/wiki/)

3) 玉木英彦、喜安善市編、「自動翻訳」、みすず書房、1960 年。

4) Booth (Andrew J.), Brandwood (L.) Cleave (J.P), Mechanical Resolution of Linguistic Problems, Butterworths Scientific Publications, London, 1958.

5) 蓼沼良一：「電子計算機による英文和訳の研究、電気試験所研究報告 醍 624 号」、1961 年。

6) 情報処理学会、コンピュータ博物館より、(http://museum.ipsj.or.jp/heritage/KT-1.html)

2.0 概要

情報処理学会ができ、その学会誌「情報処理」が発刊され、1960 年 7 月号、学会誌第 1 号に「計算しない計算機」という題で、工業技術院電気試験所の和田弘氏が論文を書かれている。[1)]

2.1 1960 年代前半の活動

2.1.1 文字の認識

その内容は、まず最初に文字を読む機械について、電子計算機が符号（code）だけしか読めないので、人間に読める文字が読めない現状は不便です。そこで機械に人間の文字を読ませようとする努力は、いくつかの方法が発表されている。

電気試験所では、800 個のゲルマニュームダイオードを用いて、普通の英文タイプにある文字や記号の合計 73 文字が読める装置を開発した。このような研究は、イギリスの Manchester 大学、アメリカの M.I.T. や Bell 研究所でも始められており、音声認識についても京都大学、東北大学、名古屋大学、電気試験所などで研究されており、これらを対象とする研究を pattern 認識（recognition）と呼んでいると述べられている。

2.1.2 外国語の翻訳

外国語を機械で翻訳させる研究は逐語訳から始まりました。欧米語はソ連を含めて、同じ系統から発しているので、一つの文章の中で単語の相互の位置には、あまり大きな違いがありません。

われわれは、semantic な研究よりも、まず構文的（syntactic）な研究が必要だと判断した。

英日翻訳を考え、文章を単純なものだけとした。訳文は仮名文字で、分かち書きして打たせた。この翻訳結果は、前に紹介している。そして、実用化のためには、vocabulary を増すことが必要ですが、研究としては欧米でやっているような semantic なものへ移行すべき段階に到達したと思っていますが、それにも machine learning の考えを必要とする。さらに、九州大学、教 learning の考えを必要とする。さらに、九州大学、教育大学でも研究が始められていると述べられている。

2.1.3 要約機

論文を要約（abstract）する必要があります。できることなら、よい論文は機械を通しても一語も減らないが、まずい論文は何も残らないようにしたいものです。

採用した方法は、英文の単語を意味のある単語と意味のない単語に分けて、意味のあるのは名詞だけとし、その頻度を計算し、ある数以上の頻度の単語を選び出す。論文を読み込み、単語がある数以上含まれている文だけを印字させる。

例えば、「猫の首に鈴をつける鼠の相談」という寓話をこの方法で、要約させると、

The cat killed some of the mice,

Put a bell on the cat.

But who is going to put the bell on the cat?

の 3 文が印字される。注目した単語が cat, mouse, bell の 3 語であって、上の文ではそれらの語が 2 個以上含まれていたためです。1,000 語程度の文章であれば、「やまと」を使うと、約 17 分で要約されます。

今後は syntactic な考え、単語のつながり、三段論法などを考慮に入れて改善するつもりですが、ここにも machine learning の問題が潜んでいる。

2.1.4 検索機

論文を読んで、今後の参考にしたいと感ずると、我々は著者名、雑誌名、その番号ページ、発行年月、論文の標題、および内容の要約とか備考などを記録、保存しています。これを機械化して、能率を高めることを企てました。これらの事項をテープに打って機械にいれると、磁気テープに記録、保存されます。必要な時に、それらのどれかの項をこの機械にいれると、該当する論文全部について、上記の事項を自動的に打ち出してくれる検索機（indexer）を最近完成しました。

将来、何とかして、同じものは一か所にまとめ、古いものは忘れて（整理して）、またひとつのことがいろいろの角度から取り出せるように、例えば、cross-reference というような概念を導入して、要するに合理的な整理方法を案出したいと念じております。

2.1.5 むすび

現在のところ計算機が利用されているのは、演繹的（deductive）に書かれた数値計算に限られています。定理を証明するような帰納的（inductive）なことはまだできません。

また、先ほどから述べてきたところの諸例は、いずれも多量のデータを整理して共通な本質ともいうべきものを求めたうえ、判断する性質のこと、すなわち帰納的なことでして、人間の学習過程に似た内容のことです。machine learning と呼ばれる所以です。

本日誕生した情報処理学会が壮年になるころのことを想うと、意見なり思想を表す文章も、この計算機にかかっているでしょう。司法から始まって、行政から政治までも、やがてこれらの機械によって処理されることも夢とばかり片づけられません。

それが実現されるためにはまず述語論理について、帰納的な理論が確立されることが望まれます。そしてその際にも、言葉(単語)は、element として残りましょう。単語の意味というものは数値と違って先に述べたように曖昧な点がありますが、数学、ことに代数学はわれわれの知っているような単純な数や文字を対象とすることから離れて、複雑な集合のような抽象的な対象を操作する方向に進んでいますから、単語や文章の解明に大いに参考になることと思われます。

ことに、日本語を科学的に考察し改善することは必要で、これを怠ると機械文明の進歩にもかかわらず、日本人の生活には向上しない面が残されることになりはしないかと案ぜられます。

2.1.6 まとめ

半世紀以上も前、トランジスタでやっと計算機ができ、プログラミング言語としては、機械語しかなかった時代に、言語処理の未来を予測した論文を投稿された先輩、和田氏の先見性に（私、坂本は）驚かされました。

2.2 1962 年

前掲の 2 年後、「情報処理」[2] の Vol.3、No.3、1962 年 5 月号に和田氏が機械翻訳の現状を解説されています。

2.2.1 現状

まず、前号で私が報告したような 1960 年以前の世界での研究・開発の状況報告を行い。

問題の所在として、文中の単語の位置がヨーロッパ言語間では、殆ど同一なのに比べ、日本語では、大きく異なると述べ、

機械翻訳においては両者は総合的に研究されねばならない。研究は

1. 外国語を整理する
2. 整理された内容を機械に教える
3. 訳文を合成する

の 3 段階に分けられる。各国の研究状況を見ると、逐語訳から一歩進んで、構文の分析に関心が持たれる時代になっている。自国語を合成することにはまだあまり力が注がれていないようだ。

2.2.2 言語を計算機へ

2.2.2.1 処理方法

文字の表示法として、アルファニュメリックの 2 進法表示と引き算によるその照合方法について述べ。つぎに単語辞書について、派生語を含めて見出しをどうするか。さらに、その配列方法。使用頻度順に配列するのも一つの方法であろうと述べられている。

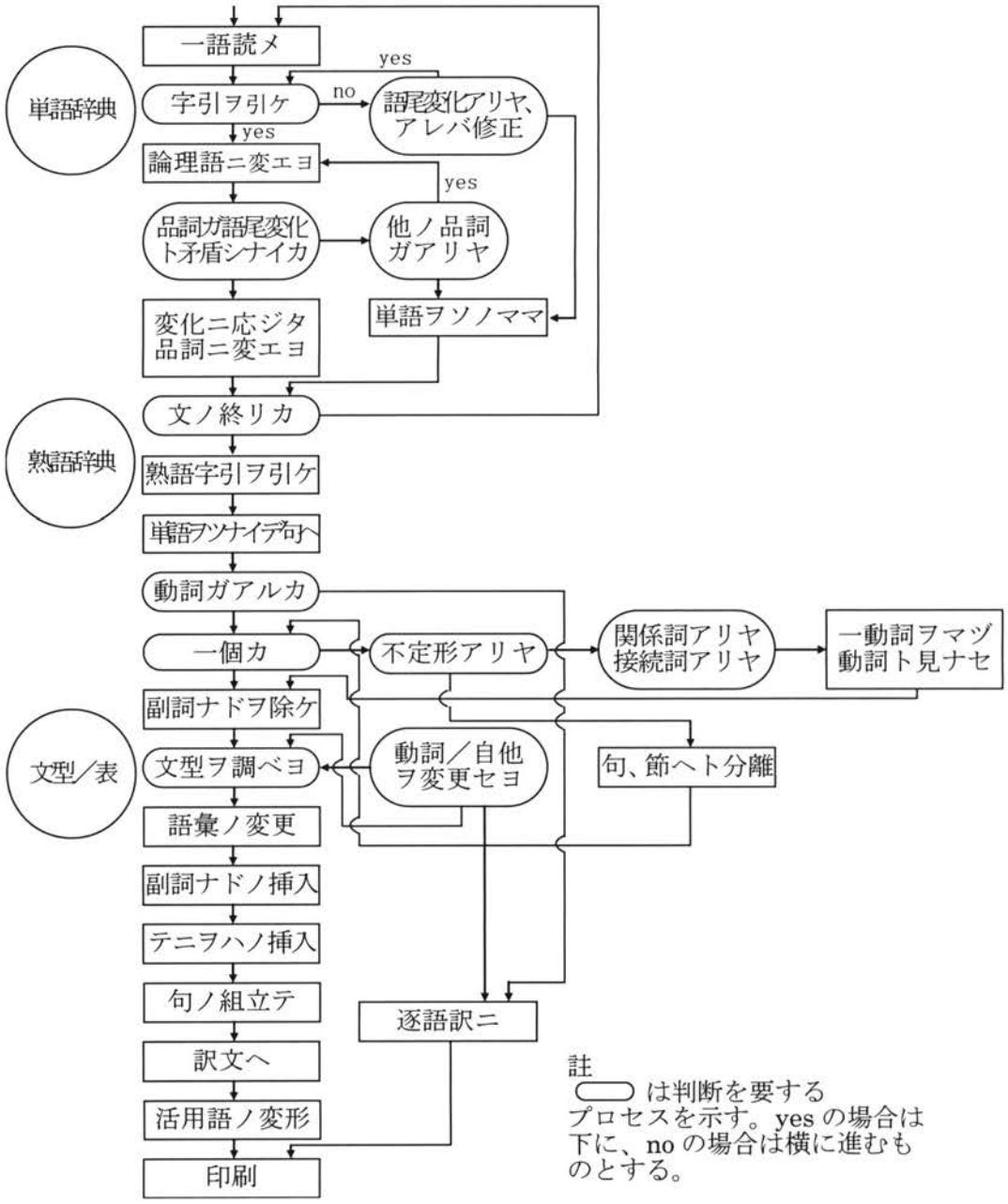
語尾処理には、サブルーチンという手法を用いると機械に便利だとも述べられている。

文型の決定では、単文の処理で、単文は主語、述語を含んでいて、その述語の性質で目的語あるいは補語

を必要とする。このことは四則演算がａｘｂのように必ず２個の数値を操作するのと違う点である。文型を機械的に確認させる方法として、table を用いることによって、容易に照合がとれる。ただし、多品詞語には苦しめられるとも述べられている。

処理手順を具体的に表示するために、流れ図をつくる必要がある。これを第１図に示す。さらに、語尾変化を処理するブロックを subroutine を使って行う。最後に、このような詳細図を機械の持つ「命令」の系列を使って表示しなければならない。命令を機械語ともいう。

2.2.2.2 プログラム



第１図

われわれが翻訳（電気試験所の「やまと」）に当たって使いたい命令形の単語は機械に備わっている命令には備わっていないし、これらでは表現できないように感じられる。これが機械翻訳の一つの隘路になっているようだ。しかし、これを機械でやらせようとするからには、命令で表せるような工夫が必要だ。

その要領を述べよう。"a x b"なる演算は、

$$\begin{array}{r} 23 \\ \times\ 14 \\ \hline 92 \\ 23 \\ \hline 322 \end{array}$$

が示すとおり、九九、左への桁送り、加算、という操作を積み重ねた結果である。このように単純な操作を集合として記述することを算法（algorithm）という。従って命令"乗算"がなくても、上記 algorithm ができれば、機械に乗算を行わせられる。普通は"乗算"と書けば、操作する電気回路が機内に組み込まれている。このような操作はプログラム software でも実行できるが、「やまと」では hardware で組み込んでいる。

翻訳でプログラムを作りたいときに、欲しい機械語がなくても前述の「字引を引け」のような疑似命令（pseudo code）を提供することは、プログラミングにとって大切な課題である。

2.2.2.3 現実の文章

一文中の各単語の品詞の決定では、1 品詞のもの、多品詞のもの、それに単語が辞書にない場合と品詞の決定には多くの困難がある。

辞書つくりでは、単語の見出し語をどのように登録するかにも問題があるとし、Washington 大学の Relfler は、全てを登録する形で、170,000 語を登録しているとし、California 大学の Lamb は、単語を分解し、lexeme（語幹、接頭辞、接尾辞、語尾変化）に分類し、15,000 を用意し、普通単語はこれらの組合せで作られて約 1,141 万語までできるはずだし、同大学が化学を分担している関係から、化学用語は、2,400 万語まで見出せる筈だという。

見出し語（entry）がどうであったらよいかは、多くの問題があるから、形態論（morphology）の議論が続くであろう。

文章は、文からなり、文は句（phrase）、節（clause）という概念で処理している。句には、形式上動詞を含まないもののほかに、不定詞、動名詞、分詞構文などがある。それらは一面、文中での働きに応じて、名詞句、形容詞句などと分類される。句は単語が連結されたものもので、このことは句が一つの結合語とでもいえる因子（string, package, fused unit などとよばれる)となって、他の単語と同格な立場で単文を構成していると見做せる。

節も同じで、coordinate clause は接続詞でつながれているから、重文と同じであり、単文を f と表わせば、f1＋f2 の形になる。

これに対して subordinate clause は日本人を苦しめやすいが、その clause は形式的には単文の形を備えており、句読点、接続詞、関係詞がその始まる位置を示している。そこでこの節を見いだせれば、節にも働きによって名詞節、形容詞節、副詞節などになるから、この新因子は他の単語とともに、全体としての単文の規則に従うことになる。

構文分析の例として、Georgetown 大学では、第 2 図のような手順でおこなっており、ソ連では、第 3 図のような手順で行われており、Indiana 大学では、多品詞語の品詞の決定することに重点をおいて、動詞の関係している品詞の確定できないものを第 4 図のような 6 種に絞って、その Subroutine をつくっている。さらに、Pennsylvania 大学では、文章に含まれている無数の表現を整理して、それらの翻訳形式を用意すれば、辞典とおなじようにそれを look up させることで翻訳の一解法になると述べられている。

1．字引を引く
2．形態の分析
　a.　熟語の調査
　b.　Syntagmatic の分析
　c.　挿入句の除外
　d.　分離
3．構文分析

第２図

1.　動詞の分析
2.　句読点の分析
3.　構文の分析
4.　名詞の分析
5.　形容詞の分析
6.　語順の変更

第３図

最後に和田の試案というものが紹介されている。

文章を構成する element は単文であると見做す。

関係詞、if, then, as, because....などがこれに含まれる。Comma の如きもこれに含める。

1．n/v (現在)	point, face
2．a/v (現在)	clean, complete
3．n/v （過去分詞）	cut, thought
4．a/v（過去分詞）	given, interested
5．a/n/v (現在分詞)	meaning, running
6．特長のあるもの	mine, like, well

第４図

次に修飾句などを棚上げすること、単文の場合とおなじ、このようにして簡素化された文に対して、別に用意してある基本文型を、

1. 文中に連接語を見出して、これを文型とあわせるか。
2. 文中に主格を見出して、これを文型と合わせるか。
3. 文中に動詞を見出して、これを文型と合わせるか、などで照合させてみて、文型の残余の部分と文章のこれに対応するところが、文型の全領域で合致するかどうか調べる。この調査には品詞の決まっていない語が文型の要求する品詞をふくむかどうかが含まれているであろう。

以上の操作を繰り返すことによって、文章全体の構文を決定できれば、品詞も決まることになる。なお、句については、動詞を含む句の基本形を用意する要がある。

2.2.2.4 関連ある研究

2.2.2.4.1 中間言語

機械翻訳は当初英語から日本語へというように、特定の国語間で進められてきた。しかし、どんな国語からでも翻訳できるように拡張したくなる日もあろう。その場合に、共通点も多いのだから、それらをまとめて記述できれば労が少なくて済むはずである。あるいは各国が自国語を分析した資料を作って相手国と交換すれば、直ちに MT ができることが望ましい。

MT はさきに記述したとおり、辞典の問題、形態の問題、および構文の解析によって基本文型（pivot language と呼ぶ人もある）まですれば、あとは自国語の合成である。これらの問題をどんな形で記述するかに当たって、参照すべき一つの基準があって、それが algorithm に便利な姿であると便利である。Esperanto があるように、このような基準となる言語を、中間言語（intermediate language）と呼ぶ。

これをどんなものにすればよいかはまだ多くの研究に俟たねばなるまいが、次に述べる述語論理がその方向を示唆してくれれば有難い。

2.2.2.4.2 述語論理

命題論理は、文と文との結合方式を明らかにしてくれたが、述語論理は一つの文の構造を研究して、量化した概念を取り扱えるようになってきた。また文が主語と述語だけの他に、いくつかの目的語を含んだ単文についてそれらの言葉の間の関係を追及してくれそうな傾向にある。

まだ、未完成であるが、文章が数式化できる暁には、その基準となる素質を備えていると信ぜられる。

また、Basic English にもふれ、L .A. Richards が 1943 年に提案した方法では、18 の述語で表現できるといわれ、これらを記号で表すとすれば、述語論理の手法を使えば、文章は名詞、修飾語と記号とで表示す

ることもできるのではあるまいか、とも述べている。

2.2.2.4.3 自動プログラミング

電子計算機の科学計算、事務計算に必要な program は、初期は machine language を使って克明に作られていたが、記号 program を経て、いまや Algol, Cobol などによる自動 programming の時代に入ろうとしている。各メーカーはこのための compiler をつくる競争をしているといってもよかろう。従って機械翻訳の研究と compiler の研究は近い。

このようにプログラムを楽に各思想が普及してくると、機械翻訳の分野でもこれに似たものが開発されるようになることは当然であろう。例えば、MIT の COMIT は単に翻訳だけでなく、数字でない文字をいろいろの目的で計算機で取扱う際に便利なように完成されたもので、要求があれば配布できるようになっている。

私（坂本）も電気試験所で、最初に学んだプログラミング言語は、COMIT であった。

2.2.3 あとがき

最後に、あとがきとして、我が国でも、MT が話題となるようになって、早くも 5 年になる。この方面に関心を持つ人が次第に増えてきていることは喜ばしい。これらの人はいずれも MT に寄与する専門知識を持っている。言語学者、論理学者、技術者にしても、同じ分野の人々だけで固まったのではグループとはなっても MT を推進させる力になりにくい。

MT をなるべく速く実用化に近づけるようにするには、それらの専門家の間の研究の方向に向く必要がある。そして事実その意志もあるように見受けられる。

そこで具体的な方策となるが、それには専門を異にする人々が自己の立場で話をすることから一歩前進して、共通の立場としてプログラムを中心とした話へと移ることは良いと思われる。もちろん言語学、機械、理論などに分類していることは差支えないが

と述べられている。

2.3 日米機械翻訳セミナ

2.3.1 第 1 回の日米機械翻訳セミナのニュース

第 1 回の日米機械翻訳セミナの内容が「情報処理」の 1964 年 5 月号のニュース欄に以下のように紹介されている。

機械翻訳セミナ（日米科学協力委員会）東京で開かる

1964 年 4 月 20 日から日米科学協力委員会第 2 部会（科学技術情報資料交換会）の主催による機械翻訳セミナが都道府県会館（東京）で開かれた。日本側議長丹羽保次郎、米国側議長 Franz L,ALT をはじめ日本側 26 名、米国側 11 人の参加者があった。

論文発表が 3 日半、将来の協力計画の討論が 1 日、その他見学などがあった。発表者は日本側 13 人、米国側 8 人で、両国における機械翻訳研究の現状調査の報告と、個別的な研究の報告とがあった。米国側の報告がメタ言語理論の観点からなされたものが多かったのに比し、日本側の報告は自然言語（とくに日本語）を取り上げたものが多かった。その他問題向言語、文字読取機、特殊計算機などについても触れられた。

おわりに 8 項目の勧告が採択された。これにもとづき、当学会またはその機械翻訳研究会が第 2 部会とのあいだでなんらかの了解に達するものと期待されている。勧告には高速度計算機の容易な利用、機械翻訳関係の日本語論文に英文の要約をつけること、カードやテープの形の両国語文章資料の交換、プログラムの交換などがあげられている。

日本側発表論文の予稿集は希望者に配布される（取扱者・西村・電試）。米国側発表論文および資料はおもな研究機関に配布保管されているはずである。

2.3.2 論文紹介

「情報処理」の 1964 年の 9 月号と 1965 年の 1 月号にその内容の概要が日本語で紹介されている。

最初に、丹羽保次郎氏の開会の挨拶が以下のように述べられている。

丹羽保次郎

Welcome Address

科学協力に関する日米委員会は 1961 年の池田総理とケネディー大統領との会談の結果生まれたもので、その第一回会合において 5 項目を選び、両国政府に勧告した。その中に科学情報と資料との交換の促進があるが、このために両国にそれぞれのパネルが置かれ、密接に連絡して仕事をすすめている。米国側パネルの主査は、最初は IBM 社の Dr. E. R. Piore であったが、今は NBS の Dr. A. V. Astin である。日本側は丹羽がつとめている。科学情報と資料のパネルでは、論文の抄録、目録索引に関する専門家の会議や一次出版物編集者の会合を企画した。この度の機械翻訳に関するセミナは日米両国における機械翻訳の研究の現状を検討し、今後の協力につき協議する目的で、第 2 回の日米科学委員会で提案せられた。

日本における機械翻訳の研究は、今まで物理学者、数学者、電子学者、言語学者、心理学者などの数グループにより、大学や、研究所などでおこなわれていた。これらの研究の連絡、討議のため、1962 年に機械翻訳研究委員会が発足し、日本における機械翻訳の研究者を網羅する体制ができ、今までに 10 回のシンポジュームを催した。今回の会合についても、この委員会に追うところが少なくない。

日本の国語、国学、国文は欧米のそれに比し、非常に異なっている。この事実は論文の和欧または欧和の機械翻訳の研究をいっそうむつかしいものにするが、一方それだけ機械翻訳研究の国際協力の必要性を強くするものといえよう。今回の会合が日米両国における機械翻訳の研究と両国の協力の推進に寄与することを望んでやまないものである。

以下、論文のタイトルは、

1. Franz L. Alt: A Survey of Automatic Translation in the United States
2. D,G. Hays: Summary of U.S. contributions
3. H.H. Josselson: Linguistic Basis of mechanical Translation. Contributions of Standard Linguistic Theory
4. W. S. Wang: The Linguistic Basis of Translation: a Transformational Approach
5. S. M. Lamb: Stratificational Linguistics as a Basis for Mechanical Translation
6. E. D. Pendergraft : Hardware for Mechanical Translation: Relation between Hardware and Linguistics
7. Arnold C. Satterthwait : Hardware for Mechanical Translation : Problem-Oriented Programing Language
8. E. N. Adams: Input Reading for Language Processing system
9. Osamu Watanabe : Verb Patterns for the Basic of Sentence Generation
10. Akihiro Nozaki : Mechanical Translation of Pre-edited Japanese
11. Tuneo Tamati : A Survey of MT Oriented Hardware in Japan
12. Toshiyuki Sakai: Models and Strategies for MT
13. Bunpei Koro : English-Japanese Translation Formulas
14. Hiroshi Wada : Status and Problems of Mechanical Translation in Japan
15. Makoto Nagao : An Approach to the General Theory of Natural Language
16. Seiichiro Ohnishi : Psych-Linguistic Studies of Mechanical Translation in Japan
17. Toshihiko Kurihara : Design of the Special Purpose Computer KT-2
18. Hirohiko Nishimura : Automatic Segmentation of Japanese Text
19. Isao Imai : Japanese Grammar of the Zeroth Approximation
20. Shokichi Iyanaga : Procedure for pre-editing Japanese

である。

さらに、第2回日米機械翻訳セミナが米国で行われ、その報告と座談会が「情報処理」に掲載されている。それは次号に。

参考文献

1) 和田 弘:「計算しない計算機」、p11~15,「情報処理」、Vol.1, No.1,1960.7.

2) 和田 弘:「機械翻訳」、p144～154、「情報処理」、Vol.3,No.3.1962.5.

3) 「ニュース」、p178,「情報処理」、Vol.5.No.3,1964.5.

4) 「機械翻訳セミナ紹介」、p299~302,「情報処理」、Vol.5,No.5,1964.9.

5) 「機械翻訳セミナ紹介」、p47~55,「情報処理」、Vol.6,No.1,1965.1.

3.1 機械翻訳研究委員会の発足

日本で機械翻訳の研究が本格的に始まりました。そして、委員会の発足が1965年の「情報処理」[1]の「機械翻訳の現状と将来」和田弘、第5回情報処理学会大会における招待講演と題した記事の中に述べられています。

機械翻訳について話すよう交渉をうけました。近頃直接研究に従事しておりませんが、学会の機械翻訳研究委員会の委員長をいたしている関係からお引き受けしました。

1960年4月当学会が創立されました際、"計算しない計算機"(この記事については、前号で紹介しました)と題しまして、当時電気試験所でやっておりました機械翻訳の研究内容をお話したことがあります。

翌年の秋、学会に研究委員会の規則が制定され、"機械翻訳研究委員会"が設置されて以来、隔月に1回ずつひらかれて現在に至っております。どんな人がどんな項目の発表をなさっておられるかは、会誌の会告欄をご覧いただければわかるはずです。

この委員会はopenではありますが、関係の方には特に通知しているので、参集される方も自から定まっております。言語学を専攻される方は折角の熱意は持っていても機械の使い方がよくわからないので困っているような印象をうけましたので、希望者には電気試験所で計算機の使い方の講習をしたこともありますが、1962年には総説としてご参考になることをまとめたこともあります。(この文献も前号で紹介しました)

日米科学協力委員会で情報交換を担当しておられる丹羽保次郎先生は、科学技術の論文だけでも機械翻訳(以下略してMTとする)ができるようになればと念願されて、文部省の助成金によって俗称"丹羽委員会"を結成されました。学会のMT委員会には地方の方が旅費の関係で出席されないことと、理学部関係の方がかけていたので、これを補われたものでした。数回の会合で同委員会も軌道に乗った頃、MTのセミナーを開催することをアメリカ側と交渉され、1964年4月東京でアメリカから9名の来訪を得て計20名が論文を発表し、意見を交換し、日本のおもな研究の実地を訪問しました。これらの論文の要旨は学会誌に紹介されております。(この件も、前号で紹介しました)

その後、丹羽委員会は学会のMT委員会の強化改組を条件としてこれに合流の上、発展的解散をいたしました。

日米科学協力を担当する日本学術振興会には、MTの研究を推進するために、MT部会が設けられています。

3.2 世界の情勢

わが国では機械にかけた実験をしている研究機関としては、電気試験所、九大工学部、京大工学部、東大理学部、名大工学部、防衛庁技研、および電機大学などです。いずれも英語を和訳することを主としていますが、東大、電試では和文英訳の実験もやっています。言語学者も文法についていろいろ建設的な提案をしておられますが、それを骨子とした実験はまだ見当たりません。研究グループとしては、国語研究所、新潟大、京大、名大、教育大、紅露研などです。

一方アメリカでの研究状況はかなり変って来ております。われわれより一歩先んじて研究を始めてみたがすでに30億円の研究費を投じたにかかわらず、実用になったとはいい難いです。研究者自身が反省して、翻訳よりもまず自国語でもよいから言葉を研究してみる必要がある。それには計算機を使う必要があるということから、Computational linguisticsなる標題のもとに研究するように転じています。そして成功すればMTはもちろん、文献のabstract、情報検索などにも共通に貢献できると考えています。

こんな理由から、現在MTを中心に研究をしているところとしては、大学ではTexas, Calif.大学、Wayne State大学、Ohio State大学、Washington大学、Indiana大学、George-town大学。

その他の機関としては、National Bureau of Standards、Rand Corp.、Bunker-Ramoなどでして、Harvard大学、M.I.T.、I.B.M.などはさきに述べた転

向の色が濃いいようです。

（MTの解説：中略）

さてMT研究の現状であるが、どの研究者もその点に困っている。多くの説が提案されていること自体がその間の状況を物語っているわけである。いずれもこの説明しにくいと申した内容について、ある一面に重点を置いたものともいえる。

これらを念頭に研究グループの態度を整理してみると、まず次の3派に大別できる。

（1）経験派

従来の文法の中で骨格と思われるいくつかをまずプログラムして実験し、その結果から、追加改良して行こうというもので、Georgetown 大学など古くからやっているものはこれに属する。

（2）伝統派

既存の文法など言語についての知識をそのままプログラム化してゆく、集大成されている知識を数学的に記述すると、膨大なものになって、到底完成の見込みがあるまいとの見解に立ったものである。

（3）理論派

文章の構造をきめる態度として認められている説にConstituent structure 論と dependency structure 論とがある。前者は文を主部と述部とに2分し、主部をさらに名詞句とその他とに2分するなど、次々に2分してtree状の解析をするものである。一方は、動詞に着目して、主部、目的部、副詞修飾部がこれに依存的構成であるとして分析を進めるのである。また数学から出発したN. ChomskyはZ.S. Harrisの句についての研究を再構成してphase structureについての文法を確立した、これには変換理論（transformation rule）が併せて発表され、信者がふえている。またS. Lambは言語を単語より小さい単位のものの集合であると見做す層化理論（stratification rule）を主唱して異彩を放っている。それぞれの哲学に従っているが、共通していることは、数式化、形式化ということが根底にあることであろう。

電気試験所でのMTの研究で考えたことをこれらと較べてみると、これらの諸説はみんな含まれているような気がする。つまり諸派諸説と雖も、実体は紙一重の差しかないのではなかろうか。優劣を論ずるには実験結果に待つ以外にはあるまい。

こんなわけで、さきにも述べたとおり研究の中心はMTから英語そのものの研究へと動いている。その典型的な例はYngveが始めた文章の創作であろう。これは核となる文章をまずphrase structureの規則で作り、これに変換の法則を適用して変形して文章を打ち出させるものである。このためには辞典での単語に意味のほかにその語の持つ機能を含ませておく必要がある。京大の長尾君もこの実験を進めている。

このようなことを念頭に、日本語を見るとき、まず和文英訳の実験が始まったことは喜ばしい。東大の野崎君によれば、

I think that reader know well that sound is vibration of air but sound by which we do talk and that enjoy music is only a part of air vibration which we utilize. Supersonic wave about which I talk here from now on is vibration that is inaudible to our ear and of which frequency is high. I think that there is man that wonder then if such inaudible sound is useful therefore I shall describe about utility value of supersonic wave.

という訳文がだされています。原文は読まなくてもよろしいでしょう。原文にpre-editをして、inputしたものです。それにしても、日本語については、次のような諸問題がまだ手を付けられないことを申し上げねばなりません。

（1）文字：現在では、ローマ字あるいはカナ文字が使われているが、実際の文章は漢字が含まれているのみならず、英語のように分かち書きになっていない。単語辞典を作ることはMTの第一歩ですが、この辞典の見出し語を漢字にするか、カナ文字で行くか、分かち書きになっていない単語の両端をどうして求めるか大変なことでしょう。

(2) 構文：日本語と英語との間には大きな違いがある。英語は客観的な思想で書かれているが、日本語は人主思想で貫かれているようです。そこで主語となる人が明らかな場合は省略される。このようなものを追加したり、逆に削除したり、あるいは別途の形式をとると翻訳が容易になるような事例が乏しくない。

それにも増して、MT に適した文法が見当たりません。言語学者と計算機側の学者との協力がますます必要であると考えます。幸いそういう気運は盛り上がりかけていると見ています。

なお、1965 年 5 月には、I.F.I.P の大会が New York で開かれますが、その前の週に、この方面の国際会議が開催されることになっています。

新しい進歩が生まれることを期待したいと思います。

3.3 第 2 回日米機械翻訳セミナー

「情報処理」[2)]に第 2 回日米機械翻訳セミナーの報告会が行われ、その時の様子が、前述の和田氏からなされています。以下にこれを掲載します。

報告　和田　弘

1. セミナー参加者

日本側

和田弘（成蹊大）、坂井利之（京大）、
田町常夫（九大）、野崎昭弘（東大）、
長尾真（京大）、平松啓二（電機大）、
山田小枝（立正大）。

米国側

E.D. Pebdergraft (U.Texas), V. Yngve (M.I.T.),
D. Hays (Rand), H. Josselson (Wayne State),
S. Kuno(Harvard), C. Fillmore (Ohio State),
P. Garvin(Bunker Ramo), I. Sakai(Ohio State),

2. 見学旅行

5 月 5 日

(i) Bunker Ramo. (Canoga Park)
(ii) RAND Corp.(Santa Monica)

5 月 6 日

(iii) U. of Calif., Los Angeles 分校
(iv) System Development Corp.

5 月 10 日

(v) LRC, Univ. of Texas(Austin)

5 月 12 日

(vi) National Bureau of Standards (Washington D.C.)

5 月 13 日

(vii) Harvard Univ.(Cambridge)
(viii) M.I.T.

3. セミナー

5 月 17、18 日の両日、New York の Park-Sheraton Hotel で開催されたアメリカ側の出席者はさきに示した方々のほかに、N.S.F.の Dr. See、Mr.Pronko、RAND Corp.の Mr. Ziehe と Mr. Kay、Bunker Ramo の Mr. Penn。さらに通訳として日本人が 1 名加わりました。

われわれは論文を用意して行って発表しました。

すなわち

Input and Data Format of Japanese Texts:
和田弘
Procedure for the Analysis of Japanese Text:
坂井利之
Syntactic Description of Japanese Grammar:
田町常夫
On the Dictionary Preparation:
野崎昭弘
Japanese-English Translation regarded as Sentence Generation:
長尾真
Syntactic analysis of Japanese , by Using the Property of Dependency:
田町常夫、石原好宏

の6論文です。

アメリカ側は見学旅行の際に資料を提供してくれたため、セミナーではまとまった論文を中心に自由討論が大部分であったと申せます。

第2日目にN.S.F.と学振とを経由して日米科学協力委員会に提出する勧告書の草案を審議しました。議論が活発に出て、原案が相当修正されました。

勧告書の訳を記します。

勧告

昨年のセミナーでの勧告を再確認すると次の5項目を一層に明確にする好機であるとの前置きにつづいて、

(1) 研究者の交換は日米協力の中で大切なものである。現在の計画は1年間で成果を上げるには短すぎるから、延長するつもりで続けていこう。別の計画も奨励しよう。

(2) 以下の勧告を実行するには日米の学会（情報処理学会とAMTCL）との支持が要る。各学会は個人または委員会を任命して、それぞれのスポンサー機関との密接に連絡しよう。

(3) 文献などの交流をもっと促進しよう。日本の文献の題目をアメリカのThe Finite Stringのリストに掲載しよう。

(4) テキストの交換を便利にするために共通のcodeとformatを採用できるように考えよう。両国の学者はそれぞれの国語についての提案を出してもらいたい。

(5) 昨年と今年との2回のセミナーの結果、専門家による小規模の会合が有効なことがわかった。上記(3)および(4)の実体についてseminar, planning meetingあるいはworking groupが時々必要だ。少なくとも2年以内には次回の集合をしよう。日米の両学会はテーマと会合場所とを選定して、スポンサーに進言しよう。

3.4 国際会議

5月19～21日の3日間、同じPark-Sheraton HotelでInternational Conference on Computational Linguisticsが開催されました。日本はさきに6個の論文を送ったところ、京大の坂井、長尾両君によるsentence generationの論文が発表されることになっていました。論文は40ぐらい参加者に配布されました。2日目の夕、晩餐会があって主催者たるATMCLの会長Prof. Lehmanの講演がありました。そのあとで同氏の部屋に各国から1名ずつ代表が招かれまして、Whiskyをご馳走になりながら非公式に国際的な組織を作ることによって、こういう会議を開催することはどうだろうかと提案がありました。

ヨーロッパの人々はいずれも賛成しましたが、私は反対しておきました。本日ここにおられる久野さんも反対しておられたようです。しかし多数の意見によって推進することになり、その結果すでにProf. Lehmanから手紙が来ました。

それによると、FederationとCentral Committeeをつくることについて、各国の意見をまとめて7月中に返事が欲しいとのことです。もし大多数が賛成ならば、1967年には会議を開くことに予定したい、との趣旨です。返事を出さねばなりません。

質問 反対の理由は、何か義務でもできますか。

和田 できます。できれば、大した額ではないが費用を負担しなくてはなりません。さらに、会合の都度誰か一人は出ねばならず、できれば同一人でありたいものですが、無理でしょう。

国際的な会合はヨーロッパで開かれると思わねばなりません。旅費が大変です。それにヨーロッパの諸国では、国語がいずれも違うのですからどうしてもその辺が議論の中心になって、日本語はやはり議論の外におかれることになりましょう。

さらにこの動きを注目しますと、言語学者がその研究に計算機を利用しようといういわゆるcomputational linguisticsの人々が中心になっています。わが国でいえば計量国語学会に相当するものです。理工科系の人はすでにあるIFIP

の3年ごとに開かれるCongressの中でMTのための場をもっています。

また仮に反対しても、設立されるでしょう。加入してなくとも67年の会議には参加することはもちろんできると思います。久野さんも理由を一つ、

久野 アメリカの連中は肩書きを欲しがっている。またヨーロッパに行く理由を求めています。委員会ばかり沢山作っていて、意味がない。

座談会

出席者和田弘、坂井利之、久野瞭、野崎昭弘、
長尾真、小笠原林樹、伊藤正安、
ベーツホッファー

司会　中野道夫

3.4.1. M.I.T.

中野 はじめに、とくに印象に残ったことを伺いたいのですが

和田 アメリカでいま MT を一生懸命やっているところでは、テキサス大学だけです、たいていのものはもうあきらめて、N.S.F.から金をもらって、好きなことをしていると見ました。ただ日本人がぼやぼやしていると、アメリカ人の方が日本語の機械翻訳を先にやってしまうのではないかという心配はあります。

長尾 M.I.T.はYngve が彼のdepth hypothesisが日本語でどうなっているかと興味をもって尋ねていた。彼の部屋には、マルティアクセスのステーションがあって、ここで彼がつくったCOMITを呼び出して日本語のジェネレーションの実験をしてくれた。田町さん、山田さん、彼はここで日本語をやったのは初めてだといっていたが、やはりプログラミングシステムは完備している。だから、言語学者は言語のことを専心して考えればよい。つまり計算機とかプログラミングを使う方に関しては、非常に自由に使える。そういう環境にあるということはやはり良いことですね。

中野 マルティアクセスというと、実際使う人はどうするのですか、inputは何ですか？

野崎 簡単にいったら、オンラインのフレクソです。

和田 長尾さん、あなたもそれに似たようなことをおやりになっているようだけれども、能率についてくらべて頂けると…

長尾 それは比較にならないですね。彼等の場合だと、現に黒板に書いたものを、その場でポンポンとルールに打ち込めば、すぐ何百というようなセンテンスがパッパッとでてくるわけです。

小笠原 その出てくるセンテンスは立派なものですか。

長尾 MTをやっているのではなくて、文法の検定ですよ。ゼネレートされた文章は全部われわれが見て、文章の意味を持っていなければまだ作られた文法が完全でないということです。

中野 ゼネレートしてみれば、文法のどこがおかしか、すぐわかるわけですね。

> 花もあるきたい、みます。わたくしでも　わたくしでも　わたくしは花でも食べます。みます。あるきます。犬が花でも花をわたくしはみたい。
>
> わたくしがたべたい。たべたい。花は私でも犬もたべたい。犬も犬がわたくしでも犬はわたくしは　わたくしはでもみます。わたくしでもあるきたい。犬がたべたい。etc

小笠原 M.I.T.はいわゆるchomsky言語学が中心になっていると思うのですが、彼の文法はやはりそこで使われているか？それと、チョムスキーはMTに関心を持っているのですか。

和田 使っている人は多いでしょう。しかし、チョムスキー自身は、MTについての関心は、今は全然もっていない。興味はもっているが、とっくにやめた。

3.4.2. 彼我の差

中野 ハードウェアの面がうらやましいという声があったでしょう。

野崎 ハード、ソフト、データ伝送も含めて、要するに計算機については差が大きいですね。

和田 そうだ。彼等はとにかく、機械翻訳をやりたい時はいつでも機械をつかっている。日本人は何か頭が良さそうな顔をして、キョロキョロしているけれども、機械をいじりもしないのが多い。

坂井 MTの見学に行く前に、アメリカ帰りのMTの専門家でない人がいっていたことに、1950 年代の初めに、アメリカで非常にさかんだったことを、今日日本でヤイヤイやっている。ところがアメリカ自体は、エンジニアが機械にかけた、これは面白いとやってみたものは、言語学的にみて何の翻訳にもなっていなかった。それで今度は言語学者がそれはいけないというので、非常にもとの方へもどりすぎているのではないかと、すなわち、学問の研究というのは確かにオッシレーションはあるけれども、少しその反動が大きすぎるのではないかというような印象をもったわけです。もう一つは、アメリカは機械と送付とソフトウェアが進んでいて、確かに環境は十分なのだけれども、とくに新しいものが出たというのではなしに、日本でやっていても別に今からでも遅くはない。わたくし自身はそういう反省のもとで、初めから機械を使ってやっているけれども、機械を使った MT も、日本ではやっていける余地は十分あると感じて帰ったわけです。わたしの感じでは、ランドにしろ、テキサスにしろ、辞書なんかだとやはり何万という単語を入れている、久野さんの文法にしろ、われわれがいっているようなものより、相当ち密な内容をもったもんをもっている。それでいろいろ議論をしているわけで、だからいろいろな面もわかってきている。抽象的なことだけでないということはありますね。もちろん周囲環境は日本にくらべて恵まれている。考えようによっては、ああいう環境があれば、われわれでもある程度おもしろいことができるのではないかという気がする。そういうことです。

野崎 少し長い目でみた話だと、アメリカの方が計算機そのものばかりでなく、計算機についての教育が普及していることの差を感じました。ほんとによい MT を長期的な見透しで実現しようとしたら、だいたい大学における教育から考えないと結局進まないんじゃないかな。

中野 皆さん、だいたい量的な人員、設備、環境、それは確かに良いが、質的には日本と比べてどういうことはないと。

坂井 パーソナリティーがないという感じはわたくしはした。

和田 それをいうと危ないんだよ。一面の真理だが、坂井さんがさっきいわれたことですが、アメリカは「アツモノにこりてナマスを吹く」という感じがする。要するにMTをしていないのだから、MTがそんなに進歩しない。アメリカがCLをやっているから、日本もすべきだとは思わない。やはり MT を頑張れば良いと思います。

田町 小生の印象ではアメリカは計算機は進んでいる。Vocabulary もある。しかし MT や MT に関係のある CL の面では特にこれはと思うものにお目にかからなかったこと、率直にいうと、すぐれたアイデアでとられたデータがない。あるいはデータはあってもどうしてよいかわからないというのが現状で、特に意味に関しては役に立ちそうな具体論もデータもないようだということ…などです。これは日本に比べて劣っているというのではありませんが、その点役に立つデータをとることが必要で、数の上だけで“日本はよほど頑張らなければ”というのは当たらないと思います。

3.4.3. セミナーの印象

中野 今度はセミナーの印象をおきかせください。

和田 先に申しあげたとおりで、セミナーでアメリカ人は良い話をしませんでした。自分達の方が経験者のような顔をしていた。日本人がどのくらいのレベルの話をするか、まあ聞いてみてやろうということだった。日本は 1 年間にたいして進歩してないが、彼等が思っていたよりやっていたという印象を受けたのではないかな。

坂井 わたしはレゾリューション・コミティにいって、項目としてどういうことをうたうかを協議したのですけれども、その始め日からハッキリしたことは、俺のところにも人物交流のチャンスを与えろということで、これが文章にも強く出たことです。ところで日本人をどの程度欲しがっているかというと、自分らが支払うものに対して有利であればよいということで、よい言葉でいえば、cooperation でやりたいことが、ハッキリでていた、もう一つはテキサスの話でペンダクラフトが非常に大きな組織をつかってやっていると、それに対してガーヴィンが非常に違った反対の意見を出している。アメリカの中にもいわゆるフランス流の考え方が分かれていた。

中野 カンファレンスの方では、日本の講演に対して、良い質問は出ましたか。

長尾 その時は 30 分のもち時間をほとんど使ったので、その時は出なかった。しかし、たとえばランドのハーパーという人などは、自分のと非常によく似ているといっていました。ドイツのボンの MT の研究所のホッペが“アメリカ人は意味のことについてはほとんど関心をもっていないと、あれはけしからんことである”といってました。彼等も私達と非常によう似た考えをもってやっているのだそうです。

3.4.4. トランスフォーメーショナルグラマーの評価

小笠原 久野さんはトランスフォーメイショナルグラマーのその位置というのはどういうふうにお考えですか。

久野 やはりフレーズ・ストラクチャ・グラマーの時代は過ぎて、どのグループもそれではもう十分ではないということが分かって来た。型はいろいろ違うが、なんらかのかたちで Chomsky 流のトランスフォーメイションを組み入れている。

和田 Chomsky のグラマーは、日本語に直接適用できないから、つまらない、チョムスキーの本の付録に英語についてだ。

野崎 ある付録を例に挙げるのは拙いでしょう。

和田 それは知っている。シンタクテックストラクチャの初めの方に書いてあるマセマティカルな所はいいと思う。あそこなら日本語にアプライできる。

長尾 トランスフォーメイショナルグラマーは非常に有益ですネ。いろいろなことを説明するのに、だけれどもその…どういうふうに一般的なレベルで利用できるかということになると非常にむずかしいと思う。

皆はトランスフォーメイショングラマーの云々とかいういろんなことを意識していってるけれども実際上それを有益に使った人は未だにないのではないか。否定をとるとかいうくらいのことでネ。一見よくて皆とびついたけれども、にっちもさっちもいかなくなったという危険性をいつも考えるべきだといいたい。

久野 Chomsky がいいだしたから、トランスフォーメイションをいうのではなくて、要するに言語にそういう面があるので、それを使わないと、いろいろな面を描き出せない。結局説明が不足になるわけですよ。

3.4.5. IFIP での MT

長尾 IFIP の MT が少ないようですけれども、あれが何故そんなつまらんのが集まったか不思議ですね。

中野 それは日本だけですか。全体ですか？

長尾 全体ですね。

小笠原 それは、その前にMTCLの方にそちらの人達が主力をおいたということですか？

久野 MTの良いペーパーがないのです。

和田 あまりやってないのに、どうしてあるのだ。

中野 それが世界的傾向ですか？

野崎 MTという意味をせまくとれば、良いペーパーはそうはないんじゃないですか？

久野 一度フレームワークを出したら、それから1年後に文法がこれだけよくなったって書いてもしょうがないでしょう。

中野 おしまいにしましょう。ありがとうございました。

3.5 私事ですが

1965年4月といえば、私が電気試験所に入所し、電子計算機部情報処理特別研究室に配属された年です。その時の電子計算機部長が和田弘氏でした。

私は、1960年3月に大学の通信工学科を卒業し、4月から東京大学生産技術研究所、第3部（電気工学）に入所しました。研究所では、東大のカッパー、ラムダーのロケット打ち上げが行われていました。弾道を描いて飛ぶロケットです。

1961年9月から、1年間の予定で、フルブライト交換留学生として、この時に前出の丹羽保次郎先生に推薦状を書いて頂きました。米国、プリンストン大学大学院電気工学科に留学しました。ここで、ある日、大学にソ連の宇宙飛行士、チトフ氏が講演に来て、これを拝聴しました。また、大学院では、プログラムの講義を受けました。

その影響を受け、将来を考え、ロケットの仕事とソフトウェアの仕事について、検討しました。ロケットといえば、軍事目的、計算機を利用した平和目的はないかと調査をすすめ、私の英語の力を使えば、機械翻訳の仕事ができるのではないかと考えました。

帰国後、調査し、電気試験所で研究が行われていることを知り、公務員試験を受け、1965年から機械翻訳の仕事に従事することができました。

当時、研究室には、「やまと」が置かれて、電気洗濯機のような磁気ドラムが回っていました。しかし、友達に研究室の名前を話すと、情報処理とは、下水処理とか汚水処理みたいな研究か、計算機で言語を研究するのだと答えると、外務省の下請けで、暗号解読の研究をするのかと質問され、当時の日本人の情報処理にたいする認識はこんな程度でした。

ただ、昨年、「はやぶさの宇宙からの帰還」のニュースが流れたときには、ロケットの使い道について、浅学であったなあという残念な気持ちにもなりました。

参考文献

1) 和田弘：情報処理、Vol. 6, No. 3, 1965 pp. 126~128.

2) 和田弘：情報処理、Vol. 7, No. 1, 1966, pp. 37~41.

4.0 概要

機械翻訳の研究で、最も衝撃的な出来事は、1960年代に発表されたALPACレポートの発表でした。今回はこの内容について、報告したいと思います。

4.1 ALPACレポートの発表

米国 National Research Council, National Academy of Sciences of Division of Behavioral Sciencesに設置されたALPAC(Automatic Language Processing Advisory Committee)が、1966年米国政府において翻訳を機械化する目的で推進されていた機械翻訳開発への研究費援助の継続に関して審議した結果による勧告である。[1)]

ALPACレポートは、米国政府機関における翻訳および機械翻訳に関して広範囲な調査・検討を行いその結果を15個の項目にまとめている。さらに、各項目に関する詳細な考察およびデータを19個の項目にまとめて付録として付加している。

4.2 ALPACレポートの詳細

このレポートについて、日本電子工業振興協会機械翻訳システム調査専門委員会が1989年度に行った調査に基づいた報告書が出ている。[2)]

この報告書によると、

a) 人間翻訳 (Human Translation)

＜主張＞

翻訳者の選択、環境の整備、翻訳の方法の工夫により、翻訳作業の効率向上が図れる。

＜概要＞

人間が行っている翻訳作業について分析すると、次の点が指摘できる。

1) 人が翻訳する場合、よい翻訳をするには、①目標言語についての十分な(good)知識、②翻訳対象についての理解能力、③原言語についての適当な(adequate)知識、を有する人がおこなうのがよい。
2) 人が翻訳する場合、よい翻訳を得るには、よい辞書と参考文献が必要である。
3) 人が翻訳する場合、ツールとしては、口述録音機(dictating machine)やタイプライターを使うが、これらでは、図、式などを挿入した最終稿が出力できず、更に別の工程が必要である。それでも費用は翻訳に一番かかる。
4) Rapid dictated翻訳(意味をつかんで同時通訳的に目標言語で表現する)とfull翻訳との比較実験において、前者は品質がほぼ同程度で時間的には1/4で済む。

b) 翻訳者の雇用形態 (Types of Translator Employment)

＜主張＞

翻訳者の雇用形態は、当時存在した形態のどちらも、一長一短で雇用者側が満足できるものではない。

<概要>

翻訳者の雇用形態は、IN-HOUSE(社内でかかえる正社員翻訳者)とCONTRACT(フリーランサー)、の2種類である。

1) IN-HOUSE(正社員翻訳者)

翻訳側からみた利点：

正社員として保証されている。

雇用側からみた利点・欠点

利点：

① 必要時にその場で翻訳して貰える
② 依頼者／翻訳者間のよりよい協力体制
③ 必要時に迅速なサービスが受けられる
④ 保持が容易

欠点：

⑤ コストがかかる
⑥ ワークロード調整が時に困難
⑦ 全ての分野に精通した翻訳家を揃えることはできないので、一貫した良い翻訳は得られにくい

2) CONTRACT(フリーランス翻訳者)

翻訳側からみた利点：

自由に仕事を選ぶことができる。

希少言語で広い分野がカバーできれば収入も高い。

発注側からみた利点・欠点

利点：

① その場ですぐ翻訳を依頼できない

② 機密保持が困難

c) 科学言語としての英語（English as the Language of Science）

＜主張＞

英語母国語者にとって翻訳の需要はそれほど大きくない。

＜概要＞

英語で書かれた学術論文が圧倒的に多いため、アメリカを始めとする英語圏科学者が母国語以外の言語で書かれた論文を読んだり翻訳したりする必要性は他の母国語より低い。

d) 科学者がロシア語を学習するために必要な時間（Time Required for Scientists to learn English）

＜主張＞

ALPAC の関心事は、ソ連の科学技術文献を効率よく理解し、翻訳することにある。ロシア語翻訳を頻繁に行う者は、元の文書を読めるようにするほうが簡単で経済的である。政府職員向けには既にロシア語科学文献を読むための教育コースがあるので、これを活用すべきである。

＜概要＞

ALPAC は、ある場合には、ロシア語翻訳をよく利用する者が文書を原語のまま読めるように学習した方がより簡単で経済的であると確信する。ある論文は、科学者が2時間クラスを8ないし16回受講する間に、関心のあるロシア語文献を判別できるようになる、と指摘している。数式、図表、グラフから必要事項を抜き出すことができる場合もある。また、関心のある文献について口頭の部分訳で十分な場合もある。このように、科学技術系の読者が専門の外国語文献を利用するために、わずかの外国語の知識しか必要としないことは、一般に認められていることである。

事実、いくつかのよく知られた研究によって、科学者は200時間以内に専門分野のロシア語資料を読むための知識を得られることが指摘されている。また、その程度の知識を持ち合わせている科学技術者は増え続けている。

ロシア語テキストを読むための政府職員養成のための施設が既に存在しているが、これまで大々的に利用されることはなかった。国防言語研究所の西海岸分室では、この目的のために2種類（6週間・10週間）の教育コースとテキストを開発した。ALPAC は既に、国防言語研究所より生徒の登録を歓迎する旨の連絡を受けている。

e) 米国政府における翻訳の現状（Translation in the United States Government）

＜主張＞

JPRS（Joint Publications Research Service）による人間翻訳と空軍の機械翻訳システムを比較すると、コスト、所要時間、翻訳の質のいずれにおいても前者が優れている。また、JPRS は翻訳者を十分に確保しており、翻訳の増加にも十分対処できる。従って、政府機関における翻訳のために機械翻訳システムの利用を拡大することは不要である。

＜概要＞

米国政府における翻訳は、①業者への発注、②部内での翻訳、③JPRS（共同刊行物調査事業）の利用など種々の方法で行われている。④一部の機関では、Public Law 480（Agricultural Trade Development and Assistance Act 農産物通商開発援助法）の見返り資金による翻訳を利用している。また、主として空軍で、⑤機械翻訳システム（後編集付き）の利用が行われている。さらに、⑥NSF（National Science Foundation）の援助による雑誌の翻訳も行われている。

f) 米国政府における翻訳者の数（Number of

Government Translators)

＜概要＞

政府内部の翻訳者の正確な数は不明。

政府内での職制分類の「翻訳者」の地位が低いため職種として「翻訳者」が敬遠されている。

g) 翻訳のための予算（Amount of Spent for Translation）

＜主張＞

政府が翻訳に使っている金額は非常に少ない。

＜概要＞

翻訳を安全にするために用いられているいろいろな方法を考えると、軍関係が翻訳にさまざまな異なった代価（1,000 語あたり 9 ドル～66 ドル）を支払っていることは驚くべきことではない。

最初の会合で、翻訳にかかる金額の見積もりがあれば役に立つであろうということになった。以下がその見積もりである。

政府機関が翻訳に使った金額は、（1963～65 会計年度で、13.07 百万ドル）

これからわかるように、政府の研究、開発を援助している多くの事業と比べると、政府が翻訳に使っている金額は非常に少ない。

商業的翻訳は、アメリカ翻訳協会の理事によれば、約 750 万ドルである。

他に政府以外のインハウスの翻訳者によるものが約 200 万ドルと見られる。従って翻訳に使われる金額の見積もりは、合計で約 2,200 万ドルというところであろう。

h) 翻訳者あるいは翻訳の供給量は不足しているか（Is There a Shortage of Translator or Translations?）

＜主張＞

翻訳者あるいは翻訳の供給量は不足していない。したがって、機械翻訳の研究が必要であるという議論には結びつかない。

＜概要＞

ALPAC 委員会は、翻訳者あるいは翻訳の供給量は、多言語から英語への翻訳で見る限り、不足していないという結論に達した。これは以下のデータによる。

1) 翻訳者への報酬は 1,000 語あたり 6 ドルという安い金額であること、および翻訳者たちへのインタビューの結果。
2) JPRS（Joint Publications Research Service）は、約 4,000 人の翻訳者と契約しているが、月平均ではわずか 300 人たらずが仕事をしているにすぎない。
3) NSF（National Science Foundation）で翻訳している雑誌の種類と量。

1964 年度では 30 種類の雑誌を、一誌あたりたった 200 コピー配布しただけであった。また、米国議会の下院に設けられた委員会により、主に政府系の 45 機関に対する調査が行われた。回答のあった 25 件のうち、翻訳者は不足していないとしているのは 7 件で、不足しているとしているのは 9 件であった。しかし、その 9 件も原因は翻訳者の待遇が良くないせいだという意見が大部分であった。

i) 翻訳された文書へのアクセス（Regarding a Possible Excess of Translation）

＜主張＞

現在、翻訳は、ユーザーの声に関係なく行われており、ユーザーの実際に必要とする文献の翻訳に、支障をきたしている。これに対する解決策としては、現在のところ、ユーザーの要求する文献を如何に素早く、高品質で、経済的に、供給するかを検討すべきである。

＜概要＞

委員会の見解として、翻訳の不足の心配はなかったが、翻訳の過多の可能性の心配は、確かに存在したと述べている。ハッキリとした読者の存在しない文献を翻訳するのは、無駄であるばかりでなく、翻訳や情報の流れを詰まらせる。片端から翻訳を行うのは、確実に観覧される文献に限り、更に翻訳を追加する場合は、特に要求のあった場合に限るべきである。ユーザーの

声として、機械的に翻訳を行うよりも、実際にユーザーの要求のある追加の翻訳を急いで欲しいと云う声もあった。

また、機械的に翻訳を行った場合、原稿の質や、新しさから、必ずしも、その有用度が高いとは言えない。具体的には、ソ連の雑誌から取った論文を翻訳した場合、アメリカの雑誌に採用できるものが、3分の1以下になる場合もあった。

更に、この不要な翻訳量の多さより、本当に必要と読者が思い、要求している論文は、かなりの時間を経過しないと、翻訳が手に入らないのが現状である。

これに対する対策として、

1) 文献の要求のみの翻訳の作成

2) ある分野全体の現状の報告の作成

3) 個々のユーザーの要求に敏速に対応する仕組み作り

といった対応策を挙げる者もいるが、委員会としては、

1) 全ての技術論文を翻訳するのは価値がないが、

2) どの文献が字際に必要かを予測しようとしても無駄であるので、

現在は、要求される文献を如何に素早く、高品質で、経済的に供給するかに集中するべきであるというのが、委員会の結論である。

j) 翻訳に関する重要な問題点（The Crucial Problems of Translation）

＜概要＞

1) 品質

・翻訳の品質は、requester の要求に十分耐えうるものでなければならないが、多くの場合十分な品質の翻訳がなされていない。

・翻訳の品質の評価実験

・評価尺度 intelligibility informativeness

・人間訳すべてが、機械訳より優れていた。

・概して、機械訳は、最も良い訳から最も悪い訳までの評価点のほぼ中間に位置していた。

・翻訳の品質とコストの関係は、精密には分からない。

2) 速度

・翻訳の lag time はまだ改良の余地があるが、現状は以下のようになっている。

・平均的な文書に対する lag time

FTD の Automatic Language Translator 109 日 [AN/GSQ-16(XW-2)]

外部のサービス会社（JPRS） 65 日

・lag time の中には、当然、翻訳そのものだけでなく、ポストエディットや出版等にかかる日数も含まれているが、FTD の機械翻訳にはさらにマシーンにかけるための時間がかかっている。

3) コスト

・機械翻訳は、高品質を要求される外交文書や文学の翻訳には向かないが、科学文献等の翻訳を支援できるであろう。

・政府がどのような翻訳形態を採用するかは、結局はコストによる。

・翻訳コストの算出が示されている。

・米国においては、露英翻訳が中心であり、この先もそれが変わる見通しはない。

・露語の文献を現状の頻度で用いるならば、人材の露語教育に力を入れた方が良い。

k) 機械翻訳の現状（The present State of Machine Translation）

＜主張＞

機械翻訳は、後編集なしでは役に立たず、後編集をすると人手による翻訳より非経済的になる。

＜概要＞

1) 翻訳が非常に限られている。

簡単でかつ選別された文章の初期の機械翻訳（10年前のジョージタウン - IBM の実験）は励ましになったが、一般科学技術文献の機械翻訳（宇宙生物学に関するロシア語記事の4つの異なる MT システムの翻訳例）は落胆的である。

2) 品質が悪く、後編集なしでは役に立たない。

科学技術文献の生の機械出力は、ほぼ判読できるが、

時々誤解をまねき、また間違いもある。そして、読むのが遅く苦痛になる。

FDT 機械翻訳システムと人間が翻訳した、物理学及び地学のロシア語文書の読む速さ、正確さ、理解度の比較で、MT の生出力は 10～29%劣り、後編集された出力でも 3~23%劣っていた。

問題の多くは、構文と語順の不自然さにある。

3) 後編集をすると人手による翻訳より非経済的になる。

現在の機械翻訳は後編集なしでは役に立たず、後編集をすると人手による翻訳より非経済的となる。

4) 機械翻訳のレベルを高めるには機械が「理解」することが必要であるが、技術的な見通がない。

5) 機械翻訳は現在それほど必要でないので、そのような研究の必要性を感じない。

l) Mannheim および Luxembourg における機械援助の翻訳（Machine-Aided Translation at Mannheim and Luxembourg）

＜主張＞

専門用語集などの翻訳支援システムは翻訳の生産性向上に役立っている。一方、後編集のコストなどの評価はできなかったが、後編集付の機械翻訳の質は通常の人間翻訳よりも質が悪いという結果になった。

＜概要＞

当面は完全な自動翻訳の実現が難しいことから、翻訳支援システムへの関心が高まっている。ここでは翻訳支援システムに関係する事項を紹介している。

1) 翻訳支援システム例

The Federal Armed Forces Translation Agency（MANNHEIM, GERMANY）での実験では、適切な専門用語集を用いることにより翻訳の生産性が向上することが示された。

英語の技術文献をドイツ語に翻訳するのに、以下の2つの方法、

① 一般の辞書、参考文献を参照したり専門家の意見を聞いたりして翻訳をする従来の方法

② テキストに出現する専門用語集（対訳付きで出現順にリストされたもの）だけを参照して翻訳をする方法

を比較しており、後者の方が 5 割以上も効率が良く、翻訳ミスが 3 分の 1 になっている。

2) 後編集

通常の（人による）翻訳と、機械翻訳の出力を後編集する方法との比較実験を行っている（ロシアがから英語へ翻訳／被験者 23 名）。

後編集の難しさやコストについては結論を出せなかったとしているが、実験結果は以下のようになる。

① 後編集の難しさに関しては意見が分かれた。興味のある点は、翻訳速度の速い翻訳者の方が、後編集は難しいと感じている点である。

② 翻訳者の 3 分の 1 は、後編集するために原文が必要であるとし、他の翻訳者の多くはどちらとも云えないとしている。

③ 機械翻訳の後編集は、翻訳速度が速い翻訳者にとっては妨げとなり、翻訳速度が遅い翻訳者にとっては手助けになる傾向がある。

④ 通常の翻訳の速度と後編集の速度は、平均してほぼ同じである。

⑤ 翻訳者にとって後編集作業は退屈だが、翻訳の際に役立っていることもあるようだ（特に専門用語に関して）。

⑥ 翻訳者は修正を容易に行える環境が必要と感じている。

3) ポストエディット付き機械翻訳と通常翻訳の品質比較

JPRS（Joint Publications Research Service）と FTD（Foreign Technology Division Translations）による翻訳の質の比較（FTD の翻訳は後編集付きの機械翻訳）した結果、JPRS の翻訳の方が FTD のものよりも質が良かった。

m) 自然言語処理と計算言語学（Automatic Language Processing and Computational Linguistics）

＜主張＞

機械翻訳の開発に資金を投資したが成果が出ない。これは機械翻訳の基礎技術である言語学が未熟であることが原因である。そこで言語学者用に、言語データの整備、言語理論検証用のコンピュータ支援ツールを開発することが重要である。

＜概要＞

政府は過去 10 年間に 200 万ドルを機械翻訳の研究に投資した。これは 1 年間に政府が翻訳に支払う金額よりも多かった。

一方政府援助は少なかったけど、コンピュータ産業界はこの間マシーンの開発に重点をおき、電子化出版技術も発展させた。これは、マーケットが明白なためであった。

機械翻訳の研究は、コンピュータハードウェアの進展にはあまり寄与しなかったが、コンピュータソフトウェアには多大な影響をあたえた。何にもまして重要な機械翻訳研究の成果は、言語学に与えた影響である。

計算言語学を研究すれば、自然言語の研究に革命を与えるであろう。現在、理論をデータと付きあわせるためには計算機のサポートを要求している言語学者が急増している。一世代まえに一生かかっていた作業は今では数週間できる。この言語学の革命は、単に機械翻訳や解析技術の研究によるものではないが、これなくしては考えられなかった。

コンピュータは言語学者に多くの挑戦をつきつけたが、この挑戦に立ち向かうことにより、多くの分野に重要な貢献をもたらすことができるだろう。例えば、

1) 外国語を教えたり、
2) 言語の本質を教えたり、
3) 令や交信に自然言語を使ったり、
4) 特殊目的の人工言語を使ったり、
5) 人間の言語能力、思考等の心理学実験をしたり、
6) 翻訳とか検索の支援にコンピュータを利用したり、

に貢献できる。

これらのための言語学の研究には、コンピュータが不可欠である。しかし、コンピュータが言語データを容易に扱える一般的な方法はまだ開発されていない。従って今後重点的に行い、サポートしていかなければならない研究分野は、

1) 言語データを扱うための基礎研究
2) 言語学者の言語理論を記述し、それを検証するためのコンピュータ利用法

の 2 点である。

・プログラミング言語と言語学との関係

＜主張＞

プログラミングは言語学にとって実際的なモデルであり、言語学はプログラミング言語の設計に理論的な方策を与えており、将来も互いの影響が予想される。

＜概要＞

1) 言語学からプログラミングへ

①FORTRAN 以前

プログラミングは機械語で行われており、言語学の影響を受ける余地はなかった。

②FORTRAN ~ALGOL(1960)

言語学から文法と構文を借りて設計された。

③ALGOL60～ALPAC 発表当時

数理言語学理論とその記法および技法を借りている。

プログラミング言語の設計は、数理言語学のために以前より理論的な手続きでおこなわれるようになった。

2) プログラミングから言語学への影響

プログラミングは、言語学に“実際的な”モデルを提供し、言語理論に対し計算に関する課題（表現の効率性や計算速度）を与える。

3) プログラミングと言語学の将来への影響

あるプログラミングは、言語で書かれたプログラミングの実行は何を意味するか、というプログラミング言語の意味論に注目が集まるだろう。また、自然言語理論における意味論の研究からの数理言語学的モデルへの影響も生じよう。それは、プログラミングに影響を与えることになる。

・機械翻訳と言語学

＜主張＞

コンピュータの出現により、大量データによる言語理論の検証が可能になった。機械翻訳による検証は、従来の言語理論が不十分であることを指摘した。しかし、コンピュータを得たことにより、今後、言語学の大きな進歩が期待できる。

＜概要＞

ALPAC レポート提出の 10 年前にはほとんどの言語学者は、「syntax とは語順、活用、機能語、イントネーション、句読法のような表層構造だけを考察するものである」と考えていた。コンピュータの出現により、大量データを使って従来の理論を検証することができるようになり、それまでに得られていた文法で構文解析実験を行った。これにより、多くの言語学者は初期の構文に対する考え方が不十分であることに気がついた。

このように、コンピュータは理論を確認する上で不可欠なものである。言語学の最近の多くの変化は、理論的進歩とともにコンピュータによる実証実験によるところが大きい。まさに、「コンピュータは自然言語の解析の水準を変ええることができる」と云えよう。事実、以前より多くの言語学者がコンピュータの支援を求めている。

機械翻訳は上記 2 つの側面を持つものである。一つは、言語学的研究の側面であり、一つは研究によって得られた方法を実験によって検証する側面である。この検証により機械翻訳はいろいろな否定的結果を持ったが、このことが従来の理論の不十分さを言語学者に指摘した。

n) 翻訳改善への旅程（Avenues to Improvement of Translation）

＜主張＞

有益な機械翻訳を実現するには、まだ長期的な研究開発を必要とする。機械の助けは、人間翻訳にも、マシンエイデッドの翻訳にも重要な要素である。翻訳作業そのものだけでなく、文書処理、印刷等の、付随的業務にも改善の余地があり、それらの研究開発も重要である。目標とする品質により、翻訳作業の内容も、コストも大幅に変わる。研究開発投資の方針としては、現実的で、比較的短期的な目標に対して、引き締めた投資をすべきであると考える。

＜概要＞

一般の科学技術文献についてのマシンエイデッドな翻訳はできるが、有用な機械翻訳は実現されていない。

機械翻訳が貢献したのは、第一に言語学に対してであり、次いでコンピュータ・プログラムに対してである。翻訳は重要であるが、その需要は、小規模ながら有能な人々の活動によって満たされている。しかし、翻訳の業務には改善の余地があり、それに向けての活動を提唱する。翻訳の品質が重要であり、目標とする品質によりコストは大幅に変動する。したがって、正確さと品質についての客観的な評価を達成することが重要である。

機械の助けは、人間翻訳にも、マシンエイデッドの翻訳にも重要な要素である。FTD（USAF Foreign Technology Division）のデータによると、製作費（最終原稿の複製や製本の費用）が非常に大きい。翻訳ジャーナルにおける遅れは、翻訳よりも製作の方に起因しているようだ。

マシンエイデッドの翻訳はより良く、早く、安い翻訳への重要な一つの道かもしてない。人間の助けとして何が最も良いであろうか？——専門用語の選択、単語の辞書引き、あるいは租訳においてか、どのようにして、多段階の工程の遅れを無くすことができるのか？どのようにして製作費を切り下げられるのか？

マシンエイデッドの翻訳にとって自動文字認識が重要だと云われるが、FTD のデータではその効果は僅かであり、情報検索や郵便業務など、その技術がもっとも有効な分野において、その研究開発は、支持されればよい。

翻訳の改善のための研究開発に、いかほどの投資をすべきか？うまく運んでいる比較的小さなビジネスに、法外な投資をする理由はない。重要で、現実的で、比較的短期的な目標に対して、引き締めた投資をすべき

であると考える。

o) 勧告（Recommendations）

＜主張＞

第一に計算言語学を自然科学の分野の一つと見做し、科学的に研究すべきである。

第二に翻訳業務の改善を抜本的に行うべきである。

（直接言及していないが、工学的アプローチでの機械翻訳の研究には資金援助はしない。）

＜概要＞

本委員会は以下に述べる二つの異なった領域に資金を投入することを勧告する。まず第一に言語学の一分野としての計算言語学である。これは、機械の支援を受けるか受けないにかかわらず翻訳の実験を含め、パージング、文の生成、構造、意味論、統計、計量的言語データの研究である。言語学は自然科学と見做すべきであり、実用的な翻訳にたいしてすぐ目に見えるような貢献をするものと判断してはいけない。また具体的な提案事項は現代の言語学的研究の成果を正しく見極められ、しかも科学的な価値観に従ってこれを評価できるような人材によって検討されることが重要である。

もう一つの領域は翻訳業務の改善である。その研究対象として以下のようなものが挙げられる。

1) 翻訳を評価するための実用的な方法
2) 人手による翻訳作業の能率アップを図る手段
3) 様々の方法による翻訳の品質とコストの評価
4) 一度も目を通されることのないような不要な翻訳を阻止することを目的とした翻訳の利用実態調査
5) 雑誌と個別の文書の両方の翻訳対象として、全翻訳過程における遅延の研究と、その改善策の研究
6) 様々な種類の機械支援翻訳の相対的な速度とコストの評価
7) 翻訳の実施過程および編集過程の機械化促進
8) 翻訳過程全般に関する研究
9) 今、主に機械翻訳中の自動辞書引きのために用意されている用語集の改良作成を含め、翻訳者のための適切な手引き書の作成

上記の研究はすべて容認できる品質の程度の仕様化、および翻訳の迅速化、低価格化を目指すべきである。

4.3 おわりに

この勧告を受けて、1960 年代後半から、1970 年代にかけて、我が国では、政府が「機械翻訳」に関するプロジェクトに予算をつけることはなかった。それについては、次号に。

参考文献

1) Language and Machines: Computers in Translation and Linguistics, 1966.
2) 日本電子工業振興会編：「機械翻訳の開発と実用化に関する実態調査」、p21~31、1989,7.

5.0 概要

1966 年に米国から ALPAC レポートが勧告され、全世界に衝撃をあたえたことは、前回に報告したが、今回は、その後の状況について報告したいと思います。

5.1 MT から CL へ

情報処理学会では、1961 年 11 月に機械翻訳研究会の第 1 回が開催された。原則として、奇数月に一回開催された。1962 年 1 月から、研究会が研究委員会として、委員会制で発足することになった。

そして、1967 年度から、機械翻訳研究委員会は、Computational Linguistics 研究委員会、通称「CL 研究委員会」に改称された。

その結果と思われるが、1967 年度以降の、この委員会等で発表された論文を私が調査した範囲では、1976 年度までの、ほぼ 10 年間で、機械翻訳という言葉が用いられたタイトルは、全タイトル 75 件の中で、以下の 4 篇に過ぎない。

・1967.3.18、「ロシア語科学論文タイトルの自動翻訳」、岡本哲也

・1967.7.15、「Phrase Structure Language の機械翻訳への応用」、田町常夫

・1967.12、「電子計算機による和文英訳」、坂井利之他

・1970.5.16、「機械翻訳のための基礎的プロジェクト」、野崎昭弘

その他のタイトルには、日本語の分析、英語の分析、辞書について、かな漢字変換、文字発生装置、漢字の符号化、記号処理、意味処理、KWIC、質問応答といった言語の分析とシステム化といった CL 研究が主流となったことが分かる。

5.2 学会での発表状況

1966 年以降、1980 年までの約 15 年間に、学会誌「情報処理」[1]に掲載された機械翻訳に関する論文は、次の 6 篇だが、この中で、研究発表としての論文は、1)、3)、5)の 3 件だけだ。それも、おそらく、ALPAC レポートが、発表される前から研究なされてきたものの報告と考えられる。また、他の論文も言語処理、自然言語、といった言葉が中心となっている。機械翻訳を中心にした研究あるいは、プロジェクトは、皆無に等しいと言える。

1)「機械翻訳の一模型」、西村恕彦、Vol.8, No.2, 1967.3, No.6.

2)「最近の言語処理研究について」、坂井利之、長尾真、Vol.10, No.1, 1969. 1.

3)「電子計算機による和文英訳」、坂井利之、杉田繁治、渡辺明正、Vol.10, No.6, 1969.11.

4)「自然言語の機械処理」、栗原俊彦、Vol.14, No.4, 1973.4.

5)「専門分野を対象とした日英機械翻訳について」、首藤公昭、Vol.14, No.9, 1973.9.

6) 解説「機械翻訳」、長尾真、Vol.20, No.10, 1979.10.

これに対して、日本語処理、漢字処理に関する論文は、毎年のように発表され、電子計算機のハードウェアの高速化、コンパクト化にともない、ソフトウェアの充実とあいまって、日本語特有の処理技術が発表されてきた。

1)「漢字の陰極線管表示方式」、仲摩雅年他、Vol.9, No.6, 1968.11.

2)「データ自動圧縮編集機能をもつ KWIC 索引システム」、植村俊亮、Vol.9, No.6. 1968.11.

3)「文字列の配列順序についての問題」、西村恕彦、Vo.10, No.1, 1969.1.

4)「漢字情報処理特集について」、石田晴久、浦城恒雄、Vol.10, No.5, 1969.9.

5) 寄書「漢字テレタイプライタの使用経験」、西村恕彦、Vol.11, No.2, 1970.2.

6)「和文入力に関する一考察」、保原信、山本稔、Vol.11, No.2, 1970.2.

7)「計算機意味論の実験」、西村恕彦、岩坪秀一、Vol.11, No.3, 1970.3.

8) 報告「計算機による標準文字図形の発生」、Vol.12,

No.3, 1971.3.
9)「科学技術における情報検索の諸問題」、中井浩、Vol.13, No.6, 1972.6.

日本語特有の処理システムとして、ハード、ソフトの両面での処理に関しては、

1)「漢字かな混り文変換システム」、松下温、山崎晴明、佐藤文和、Vol.15, No.1, 1974.1.
2) 資料「タッチ打法による漢字入力」、川上晃、川上義、Vol.15, No.11, 1974.11.
3) 解説「姓名のカナ漢字変換システム」、田中康仁、Vol.16, No.3, 1975.3.
4) 解説「漢字プリンターの縦横変換方式」、面谷和生、Vol.16, No.5, 1975.5.
5) 解説「漢字情報処理における特殊漢字」、大倉信治、Vol.16, No.6, 1975.6.
6) 解説「漢字処理システムの一例(FCL)」、柿沼悌司、Vol.16, No.8, 1975.8.
7) 解説「高速漢字プリンタシステムについて」、黒崎悦明、Vol.16, No.9, 1975.9.
8) 解説「パターン合成による漢字入出力装置」、長谷川実郎、Vol.16, No.9, 1975.9.
9) 解説「S5000 漢字情報処理システムにおける文字コード」、菊池利徳、Vol.16, No.11, 1975.11.
10)「日本語文入力用カナ漢字変換システムの試作」、木村健、遠藤安彦、小橋史彦、Vol.17, No.11, 1976.11
11)「カナ漢字変換の一方式」、牧野寛、勝部康人、木澤誠、Vol.18, No.7, 1977.7.
12) 解説「漢字処理装置」、長谷川実郎、Vol.10, No.4, 1978.4.
13)「国語辞書の記憶と日本語文の自動分割」、長尾真他、Vol.19, No.6, 1978.6.
14)「清書プログラム」、杉原厚吉、Vol.20, No.11, 1979.11.
15)「漢字姓名入力システム」、野寄雅人他、Vol.19, No.3, 1978.3.
16) 展望「日本語のワード・プロセッシング」、河田勉、天野真家、Vol.21, No.8, 1980.8.
17) 解説「日本語文字発生方式」、野村仙一、小池博之、Vol.21, No.11, 1980.11.

自然言語処理を中心に、

1)「自然言語処理のためのプログラミング言語 PLATON」、長尾真、辻井潤一、Vol.15, No.9, 1974.9.
2)「速記反訳システム」、絹川博之他、Vol.16, No.6, 1975.6.
3) 談話室「外国語データ・バンクの構想」、披田野陽一, Vol.16, No.9, 1975.9.
4)「新しい記号処理概念による SNOBOL インタプリタ」、三上和敬、豊田順一、田中幸吉、Vol.16, No.6, 1975.6.
5)「ストリング処理用仮想計算機 VC/S とその上での SNOBOL3 の実現、国立勉、吉田雄二、福村晃夫、Vol.16, No.11, 1975.11,
6)「語彙の一分割法の提案－意味及び文脈情報を用いた日本語文の解析—名詞句・単文の処理」、長尾真、辻井潤一、田中一敏、Vol.17, No.1, 1976.1.
7)「意味及び文脈情報を用いた日本語文の解析—文脈を考慮した処理」、長尾真、辻井潤一、田中一敏、Vol.17, No.1, 1976.1.
8)「日本語文献における重要語の自動抽出」、長尾真、水谷幹男、池田浩之、Vol.17, No.2, 1976.2.
9) 報告「第 6 回計算言語学国際会議に出席して」、長尾真、辻井潤一、Vol.17,No.11, 1976,11.
10)「自然言語処理プログラム」、長尾真、辻井潤一、Vol.18, No.1, 1977.1.
11)「日本語情報処理の過去・現在・将来」、長尾真、Vol.19, No.2, 1978.2.
12)「日本語情報処理特集号の編集にあたって」、長尾真、田中穂積、Vol.20, No.10, 1979.10.
13)「情報科学と自然言語」、長尾真、Vol.20, No.3, 1979.3.
14)「べた書き文の分かち書きと仮名漢字変換」、牧野寛、木澤誠、Vol.20, No.4, 1979.7.

15)「英語論文の清書における英単語の自動文節に関する統計的方法、朝倉秀三、Vol.21, No.1, 1980.1.
16)「日本語文構造解析による自動インデクシング方式」、絹川博之、木村睦子、Vol.21, No.3, 1980.5.
17)「推論過程を Fuzzy 化した自然言語システム」、三重野博司、戸内順一、Vol.21, No.6, 1980.11.

さらに、言語理論から人工知能への問題として、

1) 講座「言語理論の最近の話題　I」、笠井琢美他、Vol.15, No.1, 1974.1.
2) 講演「人工知能研究についての二三の問題」、西野博二、Vol.15, No.6, 1974.6.

5.3 国立国会図書館での文献

国立国会図書館のインターネットによるキーワード検索[2)]によると、

「機械翻訳」のキーワードで得られた 500 件の中から、1980 年までの文献を抽出すると、以下の 40 の文献が得られた。この文献の著者を調べてみると、九州大学関係者が 15 名、京都大学が 6 名、電気試験所が 5 名と 26 件の発表がなされている。すなわち、日本では、これらの機関が中心に機械翻訳の研究が進められてきたことが分かる。

1)「言語の機械翻訳—1・2—」、栗原俊彦他、電気通信学会雑誌、46(11), 1963.11.
2)「日米科学協力委員会の機械翻訳セミナーについて」、渡辺修、計量国語学会、1964.06.
3)「機械翻訳の現況」、野崎昭弘、情報処理月例資料：4)、1965.
4)「機械翻訳の意味」、大坪一夫、言語学論叢、1965.05.
5)「機械翻訳の日本語に因んで」、井上増次郎、海星女子学院大学/短期大学紀要（通号 4)、1965.12.
6)「言語」、大河内一男、服部四郎、他、東京大学公開講座、1966.
7)「機械翻訳における品詞分類と法則抽出」、清野武、電気通信学会雑誌、50(2)、1967.02.
8)「A study on mechanical translation from English into Japanese」、杉田繁治、博士論文、1968.
9)「機械翻訳の現状」、田町常夫、エレクトロニクス 13-05、1968.05.
10)「機械翻訳における諸文法について—Chomsky の文法論を中心にして」、小熊均、都留文科大学研究紀要、1968.06.
11)「機械翻訳システムについての予測」、西村恕彦、電気試験所彙報、1968.10.
12)「機械翻訳のための英和文法の研究」、西村恕彦、電気試験所研究報告、1969.01.
13)「D-tree モデルによる機械翻訳のアルゴリズムと諸問題について」、石原好宏、九州大学工学集報、1969.08.
14)「英語文の単語および熟語に含まれる構文情報の機械翻訳について」、石原好宏他、1969.10.
15)「ソ連の機械翻訳研究」、石綿敏雄, 計量国語学会、1969.12.
16)「機械翻訳システムの開発」、西村恕彦、電気試験所研究報告、1970.01.
17)「機械翻訳プログラムの作成」、西村恕彦、電子技術総合研究所研究報告、1971.03.
18)「日英機械翻訳—1—」、栗原俊彦、九州大学工学部集報、44(4)、1971.08.
19)「日英機械翻訳について—2—述部の処理についての考察」、首藤公昭、福岡大学工学集報、1972.03.
20)「日英機械翻訳について—3—単語の意味分類と意味的基本構文」、首藤公昭、福岡大学工学集報、1972.08.
21)「トランジスタ回路に関する分野における日英機械翻訳-1-」、栗原俊彦、九州大学工学集報、1972.10.
22)「機械翻訳システムの研究」、西村恕彦、博士論文、1973.
23)「機械翻訳についての一考察」、奥村宣嘉、英米文学語学研究/関西学院編、（通号 2)、1973.
24)「日常言語の機械翻訳（英文)」、猪俣靖、北九州工業専門学校研究報告、1973.01.

25)「機械翻訳の研究」、渡部満彦、言語生活、1973.02.
26)「日英機械翻訳について—4—P－P関係および修飾構造の意味処理」、首藤公昭、福岡大学工学集報、1973.02.
27)「日英機械翻訳について—5—連語の処理および翻訳アルゴリズム」、首藤公昭、福岡大学工学集報、1973.05.
28)「D-tree モデルに基づく英日機械翻訳のための言語分析について」、石原好宏他、電子通信学会論文誌、D, 1974.07.
29)「日英機械翻訳について—6—述部処理システムPPSI について」、首藤公昭、福岡大学工学集報、1976.03.
30)「日英機械翻訳のための述部処理システム」、首藤公昭、鶴丸弘昭、吉田将、電子通信学会論文誌、D、1977.10.
31)「機械翻訳および機械援助翻訳（抄訳）」、W.J.Hutchins 著、中井浩抄訳、情報管理、1979.03.
32)「機械翻訳（日本語情報処理〈特集〉)、長尾真、情報処理、1979.10.
33)「機械翻訳を理解するための基礎用語」、辻井潤一、エレクトロニクス、1979.12.
34)「機械翻訳システムに関する基礎的研究」、長尾真、京都大学、1980.
35)「気象情報理解システム（ISBAR）による日英機械翻訳」、田町常夫、九州大学、1980.
36)「意味表現を仲介とする科学技術文献機械翻訳システムに関する研究」、堂下修司、京都大学、1980.
37)「機械翻訳と機械援助翻訳」、石綿敏雄、現代の図書館　18(3)、1980.09.
38)「機械翻訳論の試み」、石原好宏、山口大学工学部研究報告、31(1)、1980.10.
39)「着実に進展している機械翻訳（コンピュータ言語学（特集)、長尾真、科学朝日、1980.10.
40)「80 年代の機械翻訳システム（計算機はどこまで言語を理解できるか)」、辻井潤一、自然/中央公論社、1980.10.

以上のような調査から、ALPAC レポート前後にわたって、この当時、我が国では、計算機技術の進歩とともに、これを基に、自然言語処理の技術が、進展してきたことが分かる。

そして、トランジスタの進歩とともに、開発の中心が次に述べるような日本語ワープの開発に向かった。

5.4 日本語ワードプロセッサ JW-10[3)]

1978 年に東芝によって開発された。初の日本語ワードプロセッサ、1978 年 9 月に発表。出荷開始は、1979 年 2 月、当時、コンピュータで取り扱える日本語の文字は、特殊な機器を用いる一部システムを除き、カタカナが当たり前であった。一方、欧米では、タイプライターから進化したワードプロセッサが広く普及しており、日本語で利用できるワードプロセッサが望まれていた。

東芝では、総合研究所で森健一らによって基礎研究を重ねた文節指定入力によるかな漢字変換を、ミニコン TOSBAC-40L を母体にした事務机サイズのハードウェアに搭載し、日本語ワードプロセッサ JW-10 として発売した。文節指定入力とは、文章の読みを文節ごとに区切って入力する方法で、たとえば、「漢字を簡単に入力する」という文章を入力する場合、「かんじをかんたんに　にゅうりょくする」のように 3 つの文節に区切って入力する。JW-10 の入力方法には、文節指定入力のほかに、漢字指定入力があった。漢字指定入力は、漢字部分を指定して入力する方法で、先の例の場合には、「【かんじ】を【かんたん】に【にゅうりょく】する」のように入力する。かな漢字変換技術の開発にあたっての最大の課題は、同音異義語の選択であった。結局、文章の前後関係、使用頻度などを利用した方式とした。かな漢字変換で使用する辞書には、最大 80,000 語まで登録が可能で、使用者別に頻度管理を行っていた。標準登録語として辞書に登録することができたが、その語数は、普通単語が 10,000 語、固有名詞が 8,000 語であった。JW-10 が発表された時の新聞では、かな漢字変換の実用性を、「例えば... ウラ

ニワニニワトリ→裏庭に鶏」という表現を用いていた。

	JW-10	JW-10モデル2	JW-5
発売	1979年2月	1980年7月	1981年1月
価格	630万円～	340万円～	260万円～

東芝の研究開発チームを率いていた森健一氏の談話として[4]、

当初、手書きの書類を読み取る装置を作ろうとしていたが、役人から、「欧米の役所の生産性が高いのは、タイプライターを使って文書を作っているからだ」と指摘され、手書きの書類を作らなくて済むように、「ワープロ」の開発を目指した。

そのコンセプトは、

・手書きよりも早く日本語の文章ができる
・欧米のタイプライターのようにどこでも持っていけるポータビリティー
・文書ファイルに、電話線を通じてどこからでもアクセスできる

ここで問題なのは、「かな」あるいはローマ字で入力して漢字に直す時に日本語に多い同音異義語が問題です。入力したい単語を、1回目の変換でどれだけ出せるか、つまり第一変換率をどれだけ高くできるかがカギです。実験した結果、手書きより早くしようと思ったら、第一変換率が95%以上にならなければ駄目だということが分かった。

5.5 ALPACと私の研究

5.5.1 学生から東京大学生産技術研究所へ

1959年に大学4年生になり、卒業研究の指導を電電公社、武蔵野通信研究所の半導体研究部の川口室長の指導を受け、「トランジスタの高周波特性」の研究を行なった。卒業後、就職を川口氏の推薦で、東京大学生産技術研究所、第三部、野村民也教授の研究室で働くことになり、1960年4月に大学卒業後は、ここで働かせていただくことになった。野村先生は、当時「東大のロケット」と呼ばれていたロケットのテレメータリングの専門家でした。この後、渡米するまでの1年半の間、研究室の先輩の助手の下で、トランジスタ回路の設計を行なった。

東大での研究内容は、自動雷検出装置の製作で、落雷の際のインパルス電流を24時間監視し、磁気テープに自動記録する装置で、トランジスタ300個からなる回路の作成を先輩のもとで、お手伝いする仕事でした。当時、半導体製品は、米国がリードしていて、トランジスタ回路の設計を進めるには、米国の大学院に進んでみたいという願望が心の中に膨らんできました。そこで、大学4年のとき、「フルブライ交換留学生」と、プリンストン大学へ奨学生の申請を行ない、幸いにもプリンストン大学大学院電気工学科に奨学生として合格した。

六本木での研究が半年を迎えたとき、留学のときが来た。

5.5.2 フルブライト交換留学生として、プリンストン大学へ

ここからは、私の人生を決めた出会いです。アメリカで迎えた春の初め、ロシアで3番目の宇宙飛行士であるチトフ氏が大学の講演会に来ました。その講演を聴いて、私の心を悩ませることになりました。当時、私は東大のロケット打ち上げの一員でした。まだ、日本では、弾道を描くロケットを追尾し、そこから情報を得るレベルで、人工衛星、まして有人宇宙船の打ち上げなど、何十年後しか実現できないレベルだと思っていました。そこで、この私の留学の経験を生かした仕事はないものかと、考えました。

この大学には、すでに、デジタル・コンピュータが設置され、そのプログラムの書き方を授業で習いました。「今の日本にはないが、必ず、デジタル・コンピュータの時代が来るはずだ」。このコンピュータを応用した研究はないものかと日々考え、コンピュータを利用して、言葉を翻訳する研究はどうなのか。自分は日本語と英語なら少しはできる、そこで、日英翻訳をコンピュータで行なえないのか。このことについていろい

ろと調べ、日本で、その研究が始まっている、その目的のコンピュータが電気試験所で、トランジスタを使って試作されていることが分かった。日本に帰って、これを私の仕事にしよう！

5.5.3 公務員試験から電気試験所へ

1962年10月、留学を終えて、帰国後、公務員試験を受け、合格、電気試験所の面接を経て、1965年4月、夢にまで見た電気試験所に入所できました。研修期間の3ヶ月を終え、通産省工業技術院電気試験所電子計算機部情報処理特別研究室に配属されました。

私が入所したとき、電子計算機部長が和田弘氏、研究室長が蓼沼良一氏、研究室には西村恕彦氏が在室されていて、すでに英語から日本語へ翻訳する計算機「やまと」が研究室に置かれていました。これは、世界で最初に作られた翻訳をするための電子計算機でした。この装置は、トランジスタ3,000個を使い、箪笥のような大きさで、ブラウン管にそのプログラムの状態がビットで表示されるものでした。辞書を記憶するためのドラムは、当時の電気洗濯機位の大きさでした。これで、4Kビットの記憶容量です。今では、想像もつきません。正確ではありませんが、恐らく、400語位の英和辞書が蓄積されていたと思われます。この計算機が「やまと」と銘々されたのは、「I love you.」と入力すると、「ワレ ナンジヲ アイス。」と「大和言葉」で出てくるような翻訳機だからです。この翻訳機の操作を学びました。

しかし、この当時の日本人の常識は、ある日、友達と会い、一杯やりながら、今度入所した研究室の名前を教えたら、情報処理とは、下水処理みたいなもの、それは、ソフトウェアの研究だというと、それって、アンダウェアのような下着のことか、言語処理特別研究室とは、暗号解読のような研究をやるところかと質問を返されて、往生した記憶があります。今では、想像することもできない時代での、研究のスタートでした。

そして、機械翻訳の原理が理解でき始めた矢先に、ALPACレポートが1966年に出されたわけです。そして、次の年から、電気試験所の予算項目から機械翻訳という言葉は消えてしまいました。

そして、その後の10年間は、カナ漢字変換、KWIC、辞書の蓄積、検索、COBOLといったプログラミングといった自然言語処理の基礎研究と以下に記すような詰将棋や点字といった応用研究を行なった。

5.5.4 「将棋ゲームの指し手の理解モデルとオンライン処理について」[5)]

という論文をプログラム・シンポジュームで発表した。

このゲームを考えつくまでの経過を述べてみよう。日本語の計算機による処理を考えているとき、言語と将棋の比較の話を思い出した。確か次のような比較が述べられていたと思う。

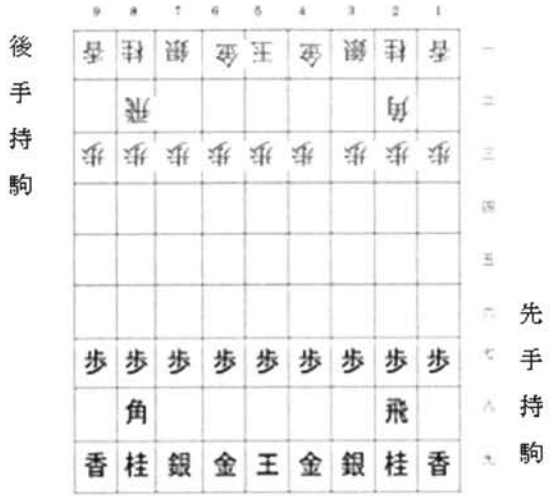
形態論－駒の動き（駒の働き）
統辞論－駒組、戦法
意味論－詰める（必死、詰めの手順、終盤の技法）
音韻論－手順の分析、統合による形態論への導入

そこで思い出されるのが、スイスの言語学者ソシュール（F. De Saussure）の一般言語学論義（Cours de Linguistique Generale, 1916）の中で語られている言葉「想像しうるすべての比較のうちで、最も適切なのは、言語の営と将棋の勝負とのそれである」（小林英夫訳）である。もちろんここでの将棋とはチェスを指している。

自然言語における意味の定義と比較して、少なくとも処理結果（ゲーム終了時；詰み）評価は明確であるという点に興味がひかれたのと同時に、チェスに比べて将棋についてのプログラム化は2、3の報告を除いて、その明確な手法の報告がなされていない。さらに、TSS画像端末による日本語（漢字）の表示が容易に行え、concurrent file処理が可能なシステムが利用できるということから、人間が学習するための将棋QAシステムを試作してみようと思い立ったのが、このプログラム開発の動機である。

・画像の制御

TSS 画像端末により局面を表示するには、TOSBAC5600 FORTRAN－Y をホスト言語として TSS 環境下で働く PLOT-10 と言うグラフィック・ソフトウェアを用いている。駒を表示する漢字パターンは直線で分割された最大 62 個の座標で表わされている。1 文字は文字パターンのファイル（3,000 種の文字）から漢字番号で索引し、PLOT-10 により表示する。なお、玉方の駒は漢字パターンを座標変換したものを用いている。



第 1 図　初期局面の表示例

5.5.5 仮名点字自動代筆システム[6)]

視覚障害者が言葉を理解するには、耳を介するか、点字を通してである。現在視覚障害者が本を読むには、直接、人（健常者）に読んでもらうか、またはこれを録音した磁気テープを利用するという方法と点字書・点訳書を読むといった方法がある。点字とテープは、その足らざるところをあい補い、視覚障害者の読書において、欠くことのできない役割を果たしている。

訓練された視覚障害者は、点字を放送のアナウンサーがニュースの原稿を読む以上の速度で点字を読める。しかし、視覚障害者のために点字があっても、点字で書かれた点字書は、非常に少なく、かつ高価である。年間新しく発行される通常の単行本は、2 万点を越える。新聞・雑誌を含めると、2 倍に達するとおもわれるが、点字書は 200 点前後に過ぎない。これ以外は、点訳奉仕者による点訳書で、手作りで作られるので、1 冊だけである。点字書も多くて 300 部程度である。これは視覚障害者の人口が 25 万人、そのうち視力が完全にない者が 10 万人、さらに点字を読める者が 5 万人といった少ない面もあるが、出版される点字書は、原本である普通図書に比べ、数倍から数十倍の価格にもなる。この原因は，発行部数が少ないこと、一冊の分量が多くなり、高価となるといった悪循環によるものである。たとえば、夏目漱石の「坊ちゃん」の文庫本が点字では全 2 巻に、また三省堂の国語辞典が全 32 巻にもなる。

点字と普通文字の世界を結ぶ過程に、計算機を導入することにより、正確で、高能率、経済的な処理体系、点字情報処理の実用化が待たれている。とくに、自動代筆と自動点訳のシステム開発が求められている。

視覚障害者が、直接点字タイプライターを打鍵することにより、点字紙上に点字モニタが得られると同時に、穿孔された 8 単位の紙テープが得られる。これを計算機への入力とし、プログラムにより普通文字へ変換し，出力する方法を自動代筆と呼ぶ。

点字の世界でも、普通文字の世界と同様に日本と欧米とでは，その表記法が異なっており、日本語は漢字・平仮名・片仮名といった多種の文字を使用している。しかし、先天的な視覚障害者に漢字を教えることは、非常に困難であるため、日本独自の Braille の 6 点点字に、仮名を当てはめた仮名点字表記法が一般に用いられている。既報の「Braille 符号と漢字の変換処理」において、長谷川氏考案の漢字点字符号系を用いる自動代筆システムについて述べているが、汎用性の面から、本システムは、仮名点字と普通文字の漢字仮名混じり文に変換するシステムである。

ここで採用した仮名点字体系は、できる限り日本点字表記法ならびに「点字毎日」の表記法に準ずるが，54 年度に改定の表記法及び表記法と矛盾する点は修

正を加えたものである。

「点字毎日」の 3 部に当たる約 10 万字を標本として入力し，on-line の TSS 端末上に処理結果を出力するシステムを開発した。なお、入力を効率的に行うために、汎用の複合用語に対する速記処理機能が付加されている。

計算機により、仮名点字文を普通文（漢字かな混じり文）に変換する自動代筆システムでは、仮名から漢字への変換での同音異義語の処理である。

本システムでは、約 8,000 語の自立語と 200 語ほどの付属語の辞書を用い、その接続情報により解析を行い、同音異字の選択を行った。標本として「点字毎日」約 10 万字を用いて、変換実験を行い、90%以上の変換率が得られた。ただしテキストは，文節レベルの分かち書きがなされている。さらに、詳細に分野別の変換辞書を付加することにより，95%程度の実用的なシステムの開発目標が得られた。

1966 年頃になると、日本でも、商用の電子計算機が発売されました。そこで、研究室では、この商用の計算機を用いて、機械翻訳のシステムの開発を始めました。プログラミング言語として、COBOL を用い、英語から日本語へ、入力文として、Bell Telephone Journal という技術文書を翻訳するシステムの開発を 5 年程かけ、5 名程の研究者で進め、新しい機械翻訳システム「やまと」を NEAC2200 の計算機上で COBOL による英日翻訳システムの開発をすすめ、これが完成したのは、1968 年でした。

5.5.6 研究室での機械翻訳システムの開発

つぎに、日本でのコンピュータの発展とともに、情報処理の時代がやってきました。英語で書かれた論文や技術レポートが増加する一方で、これを日本語に翻訳する技術翻訳家の数が圧倒的に不足する時代を迎えました。

そこで、科学技術庁が、国家プロジェクトとして、1982 年から、いわゆる「Mu プロジェクト」がスタートするわけですが、これは、次回に述べたいと思います。

参考文献

1) 情報処理学会、電子図書館、「情報処理」。
2) 国立国会図書館、国立国会図書館サーチ。
3) 情報処理学会、コンピュータ博物館、「日本語ワードプロセッサ」。
4) 産学官連携ジャーナル、科学技術振興機構、2012.9.
5) 第 18 回プログラミングシンポジウム、1977.1.
6) 第 16 回情報科学技術研究集会発表論文集、1979.10.

6.0 概要

今回は、日本で初めて、機械翻訳の国家プロジェクトとして、1982年度から90年度にかけて、科学技術庁の科学技術振興調整費で行われた、「日英科学技術文献の速報システムに関する研究」について、文献[1]と文献[2]をもとに、報告します。

6.1 はじめに

言語の機械翻訳は長年の夢であったし、現在もそうである。言語は人間の頭脳活動の最も具体的な表れであり、言語活動は人間の頭脳活動のメカニズムが解明されない限り、十分に解明されるところにはならないだろう。したがって、言語の機械翻訳という問題に直接取り組むことは無謀であるということができる。しかし1950年代の終わりごろから始められた機械翻訳の研究は、途中紆余曲折はあったものの、多くのことを明らかにしてきていた。したがって、これを総合化してしっかりしたシステムを作れば、ある程度使えるシステムが作れるだろうという期待をここ十年間くらい前から持っていたのである。しかしそれを具体的に示すためには、いくつかの条件が必要であった。それは、強力な計算機処理能力を使えること、かなりの人数の研究者による数年間の集中した研究開発作業が必要なこと、そのための研究費と研究体制をどうするかという問題であり、これは1大学の1研究室ではできるものではなかった。

一方、機械翻訳に対する社会的要請は、ここ数年非常に大きくなってきており、また、日本の科学技術活動の実態を諸外国に知らせることが、国としても必要となってきた。幸いにも国はこのような状況をよく認識し、科学技術会議は科学技術振興調整費の一部を用いて昭和57会計年度（1982年度）から3年間（その後予算などの都合で4年間に修正された）にわたって、「日本科学技術文献速報システムに関する研究」（通称:MUプロジェクト）を発足させることを決定した。

6.2 科技庁の研究の背景と成果

科学技術の高度化に伴い、研究者がより創造的かつ効率的な研究活動を進めていく上で、膨大な量の最新の科学技術情報を随時利用することが不可欠となってきた。このような科学技術情報のうち外国語で書かれた文献（海外科学技術文献）が全体の70%を占めており、これらの文献を利用する際には翻訳の需要が非常に強く、文献量の増大によりこれら文献の翻訳の機械化が求められている。

また、わが国の科学技術の発展に伴い、またこれらが国際的にも認められてきたことから、わが国の科学技術文献の海外への提供が強く求められてきているが、言語の障害により十分にこれに対応できない状況である。

これらの問題を解決するために、近年進歩の著しい情報技術を利用した日英科学技術文献の速報システムの実現が望まれている。

このため、科学技術庁は科学技術会議の方針に沿って昭和57年～60年度の4年間にわたり、科学技術振興調整費により、「日英科学技術文献の速報システムに関する研究」を実施しました。本研究は、諸外国との科学技術文献交流促進の必要性に鑑み、それら文献（抄録）の翻訳を効率的に行うために、科学技術用語辞書データベースを用いて科学技術文献を翻訳するための構文パターン方式による日英（英日）言語処理ソフトウェアの開発、科学技術分野の専門用語を収集、整理し、コンピュータで利用可能な形に編集した日英科学技術用語辞書データベースの開発、そして、これらを統合して利用する翻訳システムの開発、を目標に進められました。

そのうち、構文パターン変換方式による日英（英日）言語処理ソフトウェアの開発においては、強力な機械翻訳用ソフトウェアシステムGRADEを開発し、これによって日英・英日翻訳文法を作成し、また、形態素解析と形態素生成の各サブシステムを開発しました。日英科学技術用語辞書データベースの開発においては、名詞辞書として7万語の日英翻訳辞書、2万語の英日翻訳辞書、動詞辞書として5千語に及ぶ翻訳辞書を作

成した。また、総合システムの開発では言語処理システムと翻訳辞書を統合して日英・英日翻訳を行う総合システムの開発を行った。

昭和62年3月

科学技術庁科学技術振興局長　藤咲浩二

6.3 研究の趣旨

(1) 科学技術は、ますます高度なものとなり、その知識の質・量ともに加速度的に増大している。研究者がより創造的かつ広範な分野にわたる膨大な量の最新の科学技術情報を随時利用することが不可欠となっている。

(2) わが国においては、科学技術文献サービスは、日本科学技術情報センター、通商産業省工業技術院情報計算センター、農林水産省農林水産研究情報センター、大学等で行われているが、こうした機関では近年の科学技術の高度化、研究活動の活発化により取り扱う文献の数が極めて膨大なものとなっている。

そして、外国語で書かれた文献（海外科学技術文献）が、全体の70%を占めるという状況にある。

(3) 従来このような海外科学技術文献を国内で利用する際には翻訳の需要は非常に強く、多くの文献が翻訳され、利用されている。またこれらの文献をデータベース化する際には専門の翻訳技術者によりアブストラクト等の翻訳を行ってきた。

しかし、量の増大によりこうした作業の機械化、省力化が求められている。

(4) 一方、わが国の産業活動、科学技術活動が益々高度化し、国際化するに伴い、わが国の科学技術情報は益々国際的に利用されつつあり、科学技術情報活動の国際化が必須の問題となってきている。

(5) とくに、近年わが国の高度な科学技術水準に注目した欧米先進国や、わが国が研究協力等で重要な役割を演じている発展途上国等から、わが国の科学技術文献の提供が強く要請されているが、言語が障壁となって十分これに対応できない状況にあり、対応策が求められている。

(6) こうした問題を解決するために近年進歩の著しい情報技術を活用した日英科学技術文献の速報システムの実現が望まれる。

本研究は、現在実用化されているコンピュータ技術を前提に構文パターン変換方式による速報システムを構築することにより最終的には、科学技術文献アブストラクトの翻訳作業を支援することをねらいとする。

6.4 研究概要

諸外国との科学技術文献交流促進の必要性に鑑み、それら文献（抄録）の翻訳を効率的に行うため

① 科学技術用語辞書データベースを用いて科学技術文献を翻訳するための構文パターン変換方式による日英（英日も含む。以下同じ）言語処理ソフトウェアの開発。

② 科学技術分野の専門用語を収集、整理し、コンピュータで利用可能な形に編集した日英科学技術用語辞書データベースの開発。

③ これらを統合して利用しうる翻訳システムの開発、を行う。

(1) 言語処理システムの開発

① 構文パターン変換方式の開発

既存のコンピュータ技術を前提に、科学技術文献を言語処理する構文パターン変換方式の手法を開発する。

② 日英言語処理ソフトウェアの開発

構文パターン変換方式の手法を利用して、日本語から英語へ、また英語から日本語へ科学技術文献を言語処理するためのソフトウェアを開発する。

(2) 日ー英科学技術用語辞書データベースの開発

① 日ー英専門用語集の分析、対訳シソーラスの作成、既存文献からの対訳語の抽出及びこれらの辞書データへの編成

科学技術分野全般の日ー英専門用語集の分析、対訳

シソーラスの作成及び既存文献からの対訳語抽出を行い、これらのデータを日ー英科学技術用語辞書作成のための入力データに編成し、データベース化する。

② 科学技術用語辞書作成・管理システムの開発

日英科学技術用語辞書データベースの設計を行い、当該辞書データベースを作成・管理するシステムを開発する。

③ 翻訳のための動詞辞書の作成

文章に現れる動詞、形容詞等の翻訳辞書を作成する。

(3) 総合システムの開発

① ファイル管理システム及び編集ソフトウェアの開発

ファイル管理、更新を行うソフトウェア及び入出力時の編集を行うソフトウェアを開発する。

② 総合運用管理システムの開発

以上を総合的にシステム化した利用、運用のための管理システムを開発する。

(4) 研究推進

上記の研究の効率的推進のため研究実施計画、実施状況等必要な事項について連絡調整を行うとともに研究結果の検討を行う。

この目的にそって、科学技術文献の日英翻訳システム、英日翻訳システムの2つのシステムを作ることになった。文献の分野としては種々検討の末、われわれシステム開発者が理解しやすい電気工学分野をとることにし、日本語文献としては日本科学技術情報センターの電気工学分野の文献速報のタイトルおよび抄録文、英語文献としては、INSPECの同じ分野の抄録をとることにした。研究開発グループとしては、京都大学、電子技術総合研究所（ETL）、日本科学技術情報センター（JICST）、

工業技術院筑波情報計算センター（RIPS）の4者が協力して研究開発することになった。研究開発の分担は次のようである。京都大学は機械翻訳用ソフトウェアの開発、解析、変換、生成の文法規則の開発、解析・変換・生成の文法規則の開発・解析・変換・生成に用いる辞書情報の種類の決定、各単語に対してそれらの情報を与える作業基準の作成、変換部分の辞書データの一部の作成などを担当した。ETLは形態素解析、形態素生成のソフトウェアとその辞書の作成、解析のための用言辞書の作成を担当した。JICSTは主として体言辞書の作成を担当した。体言は一般用語のほかに大量の電気工学分野の専門用語を含む。RIPS は京都大学が作る機械翻訳システムに入力部分、テキスト出力部分、対話的修正部分、辞書の保守に関するソフトウェアなどを付け加え、機械翻訳のための総合システムを作ることになっている。

参考文献

1) 長尾真、辻井潤一、中村順一、坂本義行、鳥海剛、佐藤雅之、「科学技術庁機械翻訳プロジェクトの概要」、「情報処理」、Vol.26、No.10,1985.

2) 科学技術庁科学技術振興局、昭和57年度～60年度　科学技術振興調整費「日英科学技術文献速報システムに関する研究」。

表1 年 次 計 画

研究項目	1982年度	1983年度	1984年度	1985年度
(1) 言語処理システムの開発				
① 構文パターン変換方式の開発	基本・詳細設計	システム開発		
② 日英言語処理ソフトウェアの開発	基本・詳細設計	システム開発		
(2) 日ー英科学技術用語辞書				
① 日ー英専門用語集の分析、対訳シソーラスの作成、既存文献からの対訳後の抽出及びこれらの辞書データの編成	データの収集・分析・データベース化			性能試験
② 科学技術用語辞書作成・管理システムの開発	基本・詳細設計	システム開発		運用試験
③ 翻訳のための動詞辞書の作成	データの収集・分析・データベース化			
(3) 総合システムの開発				
① ファイル管理システム及び編集ソフトウェアの開発	基本・詳細設計	システム開発		
② 総合運用管理システムの開発	基本・詳細設計	システム開発		
(4) 研究推進				
所 要 経 費(合計)	132百万円	163百万円	162百万円	161百万円

表 2　研究実施体制及び所要経費

(単位：千円)

研究項目	担当機関	1982 年度	1983 年度	1984 年度	1985 年度
(1) 言語処理システムの開発		78,934	57,513	67,310	24,921
① 構文パターン変換方式の開発	通商産業省工業技術院電子技術総合研究所(一部を京都大学に委託)	38,267 (30,587)	46,748 (43,148	56,541 (48,901)	20,150 (19,139)
② 日英言語処理ソフトウェアの開発	通商産業省工業技術院電子技術総合研究所	40,667	10,765	10,769	4,771
(2) 日－英科学技術用語辞書データベースの開発		46,761	79,262	72,313	111,061
① 日－英専門用語集の分析、対訳シソーラスの作成、既存文献からの対訳語の抽出及びこれらの辞書データへの編成	科学技術庁研究開発局（日本科学技術情報センターに委託）	4,377	21,234	39,601	64,917
② 科学技術用語辞書作成・管理システムの開発	科学技術庁研究開発局（日本科学技術情報センターに委託）	38,544	29,700	13,200	5,280
③ 翻訳のための動詞辞書の作成	通商産業省工業技術院電子技術総合研究所	3,840	28,328	19,512	40,864
(3) 総合システムの開発		4,960	25,736	21,199	23,628
① ファイル管理システム及び編集ソフトウェアの開発	通商産業省工業技術院計画課(情報計算センター)	1,920	18,501	16,103	10,704
② 総合運用管理システムの開発	通商産業省工業技術院計画課(情報計算センター)	3,040	7,235	5,096	12,924
(4) 研究推進	科学技術庁研究調整局、振興局	1,023	703	869	892
合　　計		131,678	163,214	161,691	160,502

表3　研究推進委員会

委　員	所　　属
○長　尾　　真	京都大学　工学部教授
相　沢　輝　昭	NHK　放送技術研究所画像研究部主任研究員
石　綿　敏　男	茨城大学　教養部教授
猪　瀬　　博	東京大学　工学部教授
内　山　英　夫	(財)日本特許情報機構　常務理事
柏　木　　寛	通商産業省　工業技術院電子技術総合研究所電子計算機部長
草　薙　　裕	筑波大学　文芸・言語学系教授
坂　本　義　行	通商産業省　工業技術院電子技術総合研究所ソフトウェア部主任研究官
佐　藤　征　夫	科学技術庁　振興局管理課情報室長
鳥　海　　剛	日本科学技術情報センター　参事
中　井　　浩	常盤大学　人間科学部コミュニケーション学科教授
中　島　邦　雄	通商産業省　工業技術院計画課長
野　村　浩　郷	日本電信電話公社(株)　基礎研究所情報通信基礎研究部言語処理研究グループリーダー
藤　原　鎮　雄	千葉大学　理学部教授
淵　一　　博	(財)新世代コンピュータ技術開発機構　常務理事・研究所長
水　谷　静　夫	東京女子大学　文理学部教授
村　田　賢　一	情報処理振興事業協会　技術センター特別研究員
渡　辺　定　久	通商産業省　工業技術院計画課研究情報管理企画室運営班長

（注：○は推進委員長、1986年3月現在）

7.0 概要

前回に続き、今回は、「日英科学技術文献の速報システムに関する研究」について、文献[1]と文献[2]をもとに、その続きを報告します。

7.1 システム設計の考え方

かなり複雑な文にたいしても対応できるような機械翻訳システムを非常な短期間に4つの研究機関の協力によって作らねばならないということは非常な難問であった。現状の技術の統合によってシステムを実現するということが目標で、それ以上の研究的要素はなるべく持ち込まないということが前提とされたが、小さなモデルで少数の文を翻訳する研究のレベルとはまったく異なったシステム開発であるから、未知の部分はかなりあり、それらに対しても対処できるシステムでなければならないなどのことからシステム設計の考え方として次のような方針をとった。

(1) 未知語や未知の構造の存在によって処理を途中で放棄することなく、不完全であっても、最後まで処理を行って翻訳結果を出すシステムであること、これは翻訳文の後修正をする場合の大切な考え方である。

(2) 研究開発の発展途上において明らかになってくるであろう未知の言語現象にも十分対処できるように、機械翻訳用のソフトウェアシステムは強力な表現機能をもつように設計する。ソフトウェアは解析・変換・生成のすべての段階で用いることができ、種々の異なった性質の言語が扱えるように処理過程の制御がきめこまかくできるシステムを作る。

(3) 解析・変換・生成、それぞれが膨大な規則になるために、多人数で平行に開発ができる必要がある。したがって、部分文法という考え方を導入し、部分文法ごとに独立に開発できるようにする。こうすることによって、翻訳の誤り原因の発見なども早くなることによって、翻訳の誤り原因の発見なども早くなり、システム全体の管理がしやすくなる。

(4) 将来の多言語間機械翻訳の可能性を考えて、解析と変換。生成の3つの段階を明確に分ける。それらの段階をつなぐインターフェイスの構造をできるだけ明らかにする。

(5) 言語は少数の基本的な規則によって説明されるものではなく、膨大な例外によって成り立っているものである。したがって、できるだけ広い言語現象に対して適用できる規則を探すことは当然であるが、各単語に固有の言語現象がうまく扱えるようなメカニズムを作ることを考えた。

(6) 辞書情報としては、現在の機械翻訳に直接必要なものでなく、将来必要となるであろう情報をも入れた。できるだけ中立的で理想的な形のものを作ることが必要である。一方、大量の単語に対して正確な情報を与えるためには、複雑な辞書構造をとることはよくない。さらに単語固有の言語現象を十分に記述できることも必要である。このような矛盾する要求に対処する方策を考えた。

(7) 言語の解析の基本には格文法の考え方をとる。これは日本語の解析に特に適しているだけでなく、英語その他のヨーロッパ諸言語の解析にも広く用いられ来つつある方法である。格文法をとるために各要素の設定と、各格要素にどのような名詞がはいりうるか記述するための意味素性の集合を決定しなければならない。これは日本語のほかに英語、その他のヨーロッパ諸言語の特性を知って作る必要がある。

7.2 ソフトウェアシステム

前章で述べた種々の言語記述に関する要請から、文法記述のためのソフトウェアシステムとしては、木構造を木構造に変換する規則を採用することにした。木構造の各ノードは任意の数の属性・属性値対をもつことができるようになっている。この部分に性・数・格の一致、意味の整合性などを記述する。このように複雑な木構造を取り扱い、しかも文法規則、辞書構造とともに定形的なものにはできないということから、基

本となるプログルラム言語として LISP を採用することにした。LISP の採用については、PL／1、PASCAL などの他の言語との比較検討をしたが、今回の研究においては翻訳速度よりは言語現象を十分に記述できることに重点を置いたこと、さらに短期間にこのソフトウェアシステムを開発しなければならないことなどの要因から LISP の採用を決定した。

われわれの開発した GRADE (Grammar Describer) と名づけた文法記述言語の書き換え規則の部分は次のような構造を持っている。文法を表現する書き換え規則としては、「左辺の木→右辺の木」という形に書かれる。この規則の左辺の木は、翻訳のあらゆる中間段階にある翻訳対象文の構造（この節ではこれを単にデータと呼ぶ。これも一般には木構造をしている）との間でマッチングがとれるかどうか調べられ、データのある部分でマッチングがとれたら、その部分が規則の右辺の木構造に書き換えられる。

木構造のマッチングにおいて次のような機能が存在する。(1) 属性・属性値対がついていれば、それもマッチングの対象となる。(2) 変数を木構造ノードに書くことができる。(3) 木の枝の左右の順序が任意であることを指定できる。(4) 規則の左辺の木のマッチングに関して複雑な条件が必要な場合には、それを LISP プログラムで書くことができる。(5) この規則の部分で単語辞書内に書かれている単語固有の規則を呼び出し動作させることができる。

文法規則の数は、たとえば日本語の解析文法を考えても数 100 から 1000 近くになるから、解析のあらゆる中間段階でこれらすべての規則の適用可能性を調べるというは大変な手間になる。また、名詞句を解析するための規則、文の構造を決定するための規則などは自ずから適用される時点が違っていて、言語学的に自然な順序で解析を進めるがよい。したがって、文法規則すべてを一様に扱うのではなく、異なった文法現象に関する規則をそれぞれ別々に集めて、これらを部分文法と名付けることにした。単純名詞句を解析する部分文法、並列名詞句を解析する部分文法、埋め込み文構造を決定する部分文法などである。そして、1 つの部分文法を 1 つのノードに対応させ、ノードを有向グラフ構造につないで部分文法の適用順序を規定することにした。これを部分文法ネットワークと呼んでいる。上記の名詞句、並列句、埋め込み文などの処理部分は、実際はすべてこの部分文法ネットワークとなっている。文法全体は部分文法の組み合わせから構成される。

そこで、GRADE にはいくつかの書き換え規則をまとめて 1 つの部分文法をとる機能、その部分文法の実行後、次に適用すべき部分文法を条件づき指定する機能をもっている。さらに 1 つの部分文法の中の書き換え規則の適用の仕方を指定することもできる。すなわち、書き換え規則をデータの左から適用していくか、右から適用していくかの指定、さらにマッチングが成功して書き換えが行われたあと、さらに同一規則の適用の可能性を調べてみるか、あるいは成功した規則を用いず、次の規則を試みるかという違いを指定できる。書き換え規則の表現の仕方と規則適用のモードとの間には関係があって、うまくこれらを調整すれば能率のよい解析ができるのである。

文の解析を行うと通常非常に多くのあいまいさが生じる。すなわち、1 つのデータに異なったいくつかの構造を対応させることになる。われわれの場合には、できる限り広い範囲を一度にみて意味のあいまいさの生じることを防いでいるが、それでもあいまいな構造は生じる。文法規則を書くときに、ある構造と別のある構造が同じ部分に生じた場合に、それらは同じ句として解釈されることを積極的に表現することができるように PARA ノードと称する特別なノードを作り、システム側がその解釈を自動的に行うよう GRADE を作ってある。

GRADE の制御部は 2 つの部分からなる。1 つは人間にわかりやすい形で書かれた上記の書き換え規則をパターンマッチングしやすい内部表現に変換する変換部であり、もう 1 つは内部表現に変換された書き換え規則をデータに適用して処理を進める実行部である。これらはすべて LISP で書かれていて種々の支援プロ

プラムも入れて1万行をはるかに超える複雑なプログラムとなっている。

GRADEには、この他に辞書システムとの結合をする部分がある。辞書データは膨大であるので現在のところ大型計算機の標準ランダムアクセス・ファイルシステムを利用して記憶されている。したがって、必要な辞書データはファイルから読み出してLISP形式で主記憶の上に展開しGRADE文法の使用に供するようになっている。

GRADEに付属した支援プログラムには種々のものがある。辞書データの誤りをできるだけ少なくするために、パンチされた辞書データの誤りをデータの初期入力のときにできるだけ発見するプログラムが作られており、また、計算機内の辞書データの変更、追加などのためのプログラムなどが代表的なものである。4つの研究機関が文法、辞書データ作成を分担しているため、計算機間のコード変換をはじめとして、お互いのインターフェイスをうまくまとめるためのプログラムがいろいろ必要で、こういったなんでもないところに予想以上の手間がかかった。GRADEの詳細は他にゆずる。

7.3 辞書情報とそのシステム化

機械翻訳において辞書が大切なものであることはますます認識されてきつつある。われわれの場合には、短期間の間に大量の辞書データを整備蓄積しなければならなかったために、いくつかの工夫を必要とした。それは、辞書データをできるだけ一目でわかる形の形式で表し、むだな情報の記入をせず、最小限度の記入ですむようにすること、計算機のことを知らない一般の人でも記入ができることであり、さらに多くの人に平行に作業をしてもらうときにできるだけ質をそろえるために、辞書情報の記入のための作業手続きを詳しく書いたガイドラインを作成したことにある。

辞書は使用の上から入力文解析のための辞書、言語間変換のための辞書、出力文生成のための辞書、および形態素解析のための辞書、形態素生成のための辞書の5つに分けられる。これらの辞書のうち、言語間変換の辞書の部分を除いて、他の1つの中立的な形の辞書として作ることができ、そこから使用目的に応じた辞書を生成して使用する。したがって、単一言語の辞書としては、この中立的な辞書を作ることになり、辞書管理、すなわち辞書データの追加、変更、削除などはすべてこの中立的辞書において行うことにしている。

辞書は使用上から入力文解析のための辞書、言語間変換のための辞書、出力文生成のための辞書、および形態素解析のための辞書、形態素生成のための辞書の5つに分けられる。これらの辞書のうち、言語間変換の辞書の部分を除いて、他は1つの中立的な形の辞書として作ることができ、そこから使用目的に応じた辞書を生成して使用する。したがって、単一言語の辞書としては、この中立的な辞書を作ることになり、辞書管理、すなわち辞書データの追加、変更、削除などはすべてこの中立的辞書において行うことにしている。

辞書データは現在約55,000語計算機に入れてあるが、その内容は動詞約3,000語、形容詞、形容動詞約1,000語、普通名詞および電気工学を主とする科学技術用語約5万語、連体詞、副詞その他の付属語約800といったところである。動詞、名詞の主なものは翻訳を対象としたJICSTの抄録の最初の3,000文に現れるものをとり、それにJICSTで別途抽出した11万語の用語辞書から選んで作成した。

辞書に記入すべき情報はどのような文法概念を用いてどのような解析を行うかということに依存することはやむをえないが、できるだけ中立的、一般的に作ることを心がけた。文法としては格文法をとることにし、格の種類として表1に示すものを採用した。ある動詞の種々の格スロットにどのような名詞が入りうるかを規定する方法としては、意味素性としては図1−1、図1−2に示すものを採った。これらの意味素性のいくつかのものの AND,OR 結合によって、動詞の格スロットに入りうる名詞を規定する。

辞書情報は記入フォーマットをきめ、記入のための作業基準書を作って多人数での作業が可能なよう工夫をした。記入フォーマットは名詞、動詞、形容詞、副

詞その他の品詞について作られている。辞書は日本語（解析、生成に供用）辞書、日英変換のための辞書、英語（解析、生成）辞書などが用意されているが、その詳細は他にゆずる。

7.4 機械翻訳システム・マンマシンインターフェースの作成

現在の機械翻訳システムは、今回のプロジェクトで作成したシステムを含めて、何らかの形式で人間の援助を必要とする。また、辞書等の翻訳に必要な言語データも、対象分野に応じて適宜用意する必要があり、この点でも柔軟なマン・マシンインターフェースを準備する必要がある。このようなことから、今回のプロジェクトでは、機械翻訳システムを中核に持つ総合システムとして、次のようなシステムを開発した。これを図2に示す。

① テキストの準備のためのエディタ(前編集エディタ)
② 入力テキストと出力テキストを相互に参照しながら、出力テキストを編集するためのエディタ（後編集エディタ）
③ 汎用辞書とユーザ用の個人辞書とを分離して管理する階層的な辞書管理システム
④ 辞書フォーマットの記述を直接編集する画面エディタ

特に、④の辞書ファーマット・エディタは、辞書作成者が辞書データの内部的な管理形式とは独立に、辞書の記入フォーマット用紙のイメージを直接操作できるようにしたものであり、一般用辞書・ユーザの個人辞書の双方を編集することができる。

また、この総合システムは、京大・ETL・JICSTの各研究組織で開発されたソフトウェア、及び言語データを相互に結合し、形態素解析・構文解析・変換・構文生成・形態素生成を、バッチ運用・TSS運用の各モードで実行できるようにしたものであり、翻訳システム開発時の各種テストも可能なように工夫されている。

7.5 翻訳結果の評価

翻訳結果の訳文の質の評価法は、できるだけ主観的な判断を避け、評価者個人による評価のバラツキを少なくする必要がある。今回のプロジェクトでは、

① 理解容易性:訳文だけを読んだときの理解の容易さ
② 忠実度：訳文と原文との意味のずれ

という2つの尺度を設け、各尺度ともに訳文の修正作業と関連づけること、評価法を詳細に規定したマニュアルを用意することにより、評価値が個人の主観に依存する度合を極力低くすることとした。この評価法は、本プロジェクトの翻訳結果の評価に用いただけでなく、いくつかの他の翻訳システムにおいても採用されるようになっている。

理解容易性の基準を表2に、忠実度の基準を表3にしめした。

日英翻訳システムの性能評価は、翻訳実験の進行とともに2回に分けておこなわれ、それぞれ3,000文と10,000文の翻訳結果が評価された。10,000文の翻訳結果は（表4に示す)、3,000文の評価に比べ、未知語(未登録語)が頻出するなどの理由から多少悪くなっているが、いずれの場合も、理解容易性、忠実度ともに80%前後が翻訳結果として許容範囲に入っていると判断され、両尺度ともに許容範囲となったのは60%を超えている。

この結果は、原文の前編集を全く行なわなかった場合であり、論文抄録文の原文自体が長く、複雑で理解困難なことを考えると、前編集、後編集を前提とする現在の機械翻訳システムの性能としては十分であると考えられる。

英日翻訳システムの性能評価も、日英の性能評価法とほぼ同じ方式で行なわれ、翻訳結果3,000文の評価結果（表5に示す）もほぼ同じであった。

なお、日英翻訳システムの対象テキストは、JICST発行の論文抄録誌から、また、英日翻訳システムのテキストは、INSPEC抄録誌から、いずれも、無作為に

採られ、原文の前編集は全く行なっていない。

言語間についての機械翻訳用辞書のあるべき姿を言語理論的、計算言語理論的に研究する必要がある。

7.6 研究の波及効果及び問題点

7.6.1 研究の位置付け及び波及効果

本研究が開始されるまで、機械翻訳の研究は小規模で限定されたモデルがいくつかの大学や研究所で行なわれていた程度であった。

外国においては、少数の商用システムがあったが、いずれも 1960 年代の考え方の基づいた古い形式のもので十分な能力をもつとは言いがたいものであった。研究としてはカナダで英仏機械翻訳システムの研究が行なわれていたが行き詰まりを見せていた。一方ヨーロッパ共同体では EUROTRA 計画という多言語間機械翻訳の研究がようやく開始されようとしていた。

いずれにしても日本語とヨーロッパ諸言語、特に英語との間の機械翻訳は全く行なわれていなかった。一方欧米からは日本の目覚しい科学技術の進展の基礎をなす科学技術情報の入手が盛んに要求されるようになり、日英機械翻訳システムの開発が国として必要であるとされた。

このような日本語と英語との間の機械翻訳システムの開発の必要性は特に電子計算機関係の企業で必要と感じられて来ていたが、従来の技術開発とは異なり、言葉という複雑なものをどのように処理していったらよいかについて明確な方針がたたないという状況にあった。

このような状況の中で、本研究が果たした役割りは次のように要約される。

(1) 科学技術文献の抄録の日英、英日機械翻訳によって文献内容の情報の伝達の可能性を明らかにした。

(2) この機械翻訳システムを科学技術文献抄録のデータベースに接続することによって、翻訳機能を持ったデータベースサービスの可能性を示した。

(3) 日本が国として文献速報のための機械翻訳システムの研究開発を開始し具体的成果をあげつつあるということは、ヨーロッパ共同体、カナダ政府等における機械翻訳システム開発に大きな刺激をあたえた。

(4) 本研究は世界各国の計算言語業界に影響を与え、最近各国でこの方向の研究が活発におこなわれるようになって来た。これは計算言語学国際会議論文集、いくつかの学術雑誌にも現れてきている。米国カーネギーメロン大学にも機械翻訳研究所が最近作られた。

〔参考〕

JICST では、科学技術振興調整費による研究の成果を活用し、実用機械翻訳システムを開発する計画である。

これは、JICST のオンライン情報システム (JOIS) で、1986 年 10 月より国内外にサービスする予定の JICST 英文データベース（日本文献の抄録、キーワード等を英文化したデータベース）作成を支援するための日英翻訳システムと、同じく JOIS でサービスしている外国のデータベースの和訳のための英日翻訳システムより成る。

開発期間は、日英翻訳システムが 1986 年度より 3 年間、英日翻訳システムが 1988 年度より 2 年間の予定である。開発完了時の辞書の語数は日英、英日翻訳システムそれぞれ 30 万語の見込みである。

なお、JICST 英文データベースは現在手作業で英訳し、かつ当面タイトルのみの英訳が主であるので、少しでも早く機械翻訳システムを開発し、抄録文まで含めた英文データベースを作成することが望まれる。

7.6.2 社会、経済に対する貢献の状況及び可能性

本研究は文献抄録のもつ情報の機械翻訳を通じての速報システムの可能性を示すためのシステムを作り、その可能性を実証することにあったため、現在の段階では学問世界、企業等における製品化等に対して貢献

はしたが、社会一般、経済界に対して具体的に貢献するところまで行っていない。

しかし、この研究を通じて自然言語処理技術がこれからの情報社会の中核的技術であり、その中でも機械翻訳システムは日本語ワードプロセッサの次に来るべき大型商品であり、この技術を用いた人間と機械との対話システムの作成もすみやかに開発すべきものであることが広く認識されて来ており、そのうちに「言語産業」と称すべき産業分野が形成されてゆくものと期待されるようになって来た。

7.6.3. 今後の問題点

(1) 言語は膨大な例外の集積であるとでも言うべきものであるから、文法、辞書の不断の改良、精密化が不可欠であり、またこれを行なうことによって、質の改善された翻訳結果を得ることが期待されている。

(2) 本研究で開発したシステムはあくまでも機械翻訳の実用への可能性を示すためのプロトタイプシステムであったから、実用システムとして用いるためにはそれなりの改良を行い、システムの作り直しをする必要がある。

(3) 本研究で開発したシステムは科学技術論文の抄録を対象としているため、そこにはほとんど現れない疑問文、命令文、感嘆文などの翻訳のための文法、さらに一般の文章の翻訳のための辞書データの蓄積など、使用目的に応じた改良・拡充を行う必要がある。

(4) 上記(3)と同様な観点から、現在対象としている電気工学関係のテキストから、さらにその他の科学技術分野の文章が翻訳できるよう、専門用語辞書の充実を行なう必要がある。

(5) 日英・英日翻訳システムにおける解析文法、生成文法、変換文法、変換辞書はそれぞれ独立であるが、これが両システムで共通に使えるかどうか、研究を行なう必要がある。

(6) 代名詞の指すものの推定、省略句の推定、さらに文脈処理、文章理解の方法など機械翻訳のための基礎研究を今後強力に進める必要がある。単一言語、多言語間についての機械翻訳用辞書のあるべき姿を言語理論的、計算言語理論的に研究する必要がある。

参考文献

1) 長尾真、辻井潤一、中村順一、坂本義行、鳥海剛、佐藤雅之、「科学技術庁機械翻訳プロジェクトの概要」、「情報処理」、Vol.26、No.10,1985.

2) 学技術庁科学技術振興局、昭和 57 年度～60 年度科学技術振興調整費「日英科学技術文献速報システムに関する研究」。

表1　日本語格ラベル一覧表

(1)主体	SUBject	ーが
(2)対象	OBJect	ーを
(3)受け手	RECipient	ーに与える
(4)与え手	ORIgn	ーから受ける,奪う
(5)相手 1	PARtner	ーと協議する、異なる、ーに関連する
(6)相手 2	OPPonent	ーから保護する,独立する
(7)時	TIMe	1980 年に
(8)時・始点	Time-From	5 月から
(9)時・終点	Time-TO	来年まで
(10)時間	DURation	5 分間加熱する
(11)場所	SPAce	ーに位置する、ーで発生する
(12)場所・始点	Space-From	ーから帰る
(13)場所・終点	Space-TO	ーへ送る、ーに到達する
(14)場所・経過	Space-Through	ーを通る、上空へ飛ぶ
(15)始状態	SOUrce	5.5%から 6%へ引き上げる
(16)終状態	GOAl	英語から日本語に翻訳する
(17)属性	ATTribute	適応性に富む、欠ける、乏しい
(18)原因・理由	CAUse	事故で死ぬ、ーから分かる
(19)手段・道具	TOOl	イオン法で、ドリルで
(20)材料	MATerial	ペーストで作る
(21)構成要素	COMponent	ーから成る、ーで構成する
(22)方式	MANner	並列に、10m/sec で
(23)条件	CONdition	焦点深度で決まる
(24)目的	PURpose	ーに適する、備える、必要な
(25)役割	ROLe	議長に選ぶ、ーとして用いる
(26)内容規定	ConTent	ーと呼ぶ、述べる、みなす
(27)範囲既定	RANge	ーについて、ーに関して
(28)提題	TOPic	ーより大きい、ーに劣る、ーを上回る
(29)観点	VIEwpoint	立場から、ーの点で
(30)比較	COmpaRison	ーより大きい、ーに劣る、ーを上回る
(31)随伴	ACCompaniment	ーとともに、ーに伴って
(32)度合	DEGree	5%増加する、3 キロやせる
(33)陳述	PREdicative	ーである

注）英語名中、大文字の部分（3 字）を略称とする。

表２　理解容易性の基準

1	文意は明瞭で疑う余地はない。文法、用語用法、文体ともに適切で、なんら修正の必要はない
2	文意は明瞭で理解できるが、文法、用語用法、文体上に多少の問題が見られ、1には至らない。ただし、こうした欠陥を改善するにあたっては出力文だけ見て外人評価者単独で修正を施すことができ、日本人評価者の援助を必要としない。
3	全体的文意は把握できるが、文法、用語用法上の問題のために細部理解に自信が持てない。こうした問題を取り除き、さらに文体面の改善を施そうとすると、修正方法がひとつに定まらず、入力文の意味するところを日本人評価者に確認しなければならない。
4	文法、用語用法上の問題が多いので、相当な思考の後わずかに文意を仮定できるか、もしくはほとんど把握できない。こうした問題を解決して文体を改良するよう手を加えるより、新たに翻訳をやり直した方が時間も早く、質も向上すると思われる。
5	まったく理解できない。どんなに熟考、検討しても文意をとることは不可能である。

表３　忠実度性の基準

0	入力文の構造は出力文に忠実に再現されていて、外人がみても明瞭で、修正する必要がない。
1	入力文の構造は出力文に忠実に再現されていて、また外人にも明瞭に理解できるが、多少の修正は必要である。ただし、その修正は与えられた修正文だけにもとづき、外人が単独で行える。
2	入力文の構造は出力文に忠実に再現されているが、語の置き換えを必要とする。
3	入力文の構造は出力文にほぼ忠実に再現されているが、句と句の関係、時制、数、態、副詞の位置などに誤りがある。また、単語が重複して出力されているなどの誤りがある。
4	入力文の構造は出力文にあまりよく再現されていない。句の要素の欠落、節と節、節と句、あるいは節内の関係に誤りがある。
5	入力文の構造は出力文に良く再現されていない。節や句の欠落があるが、一応文として成立している。
6	入力文の構造はまったく出力文に再現されていない。主語や述部の欠落があるため、文として成立していない。ただし、表題のような名詞句にあっては、その主名詞（日本語で最後に位置する名詞）、またはこれを直接修飾する節・句の動詞的要素が欠落している。

表 4　日英翻訳文の評価結果（10,000 文）

理解	忠実度									
容易性	0	1	2	3	4	5	6	不備	合計	割合
1	2058	79	618	376	303	144	63	8	3649	32.1
2	52	863	667	503	336	200	99	5	2725	24.0
3	36	111	688	839	1012	346	407	22	3461	30.4
4	2	3	94	181	394	140	248	9	1071	9.7
5	1	0	34	38	118	49	227	2	469	4.1
合計	2149	1056	2101	1937	2163	879	1044	46	11375	
割合	18.9	9.3	18.5	17.0	19.0	7.7	9.2	0.4		

表 5　英日翻訳文の評価結果（3,000 文）

理解	忠実度									
容易性	0	1	2	3	4	5	6	不備	合計	割合
1	592	59	85	47	54	8	12	1	860	29.7
2	10	285	53	54	52	4	8	0	466	16.1
3	6	4	437	106	144	11	47	1	756	26.1
4	1	0	144	59	121	20	119	2	466	16.1
5	1	0	38	17	44	4	236	4	346	11.9
合計	612	348	757	183	415	47	424	8	2894	
割合	21.1	12.0	26.1	6.3	14.3	1.6	14.6	0.2		

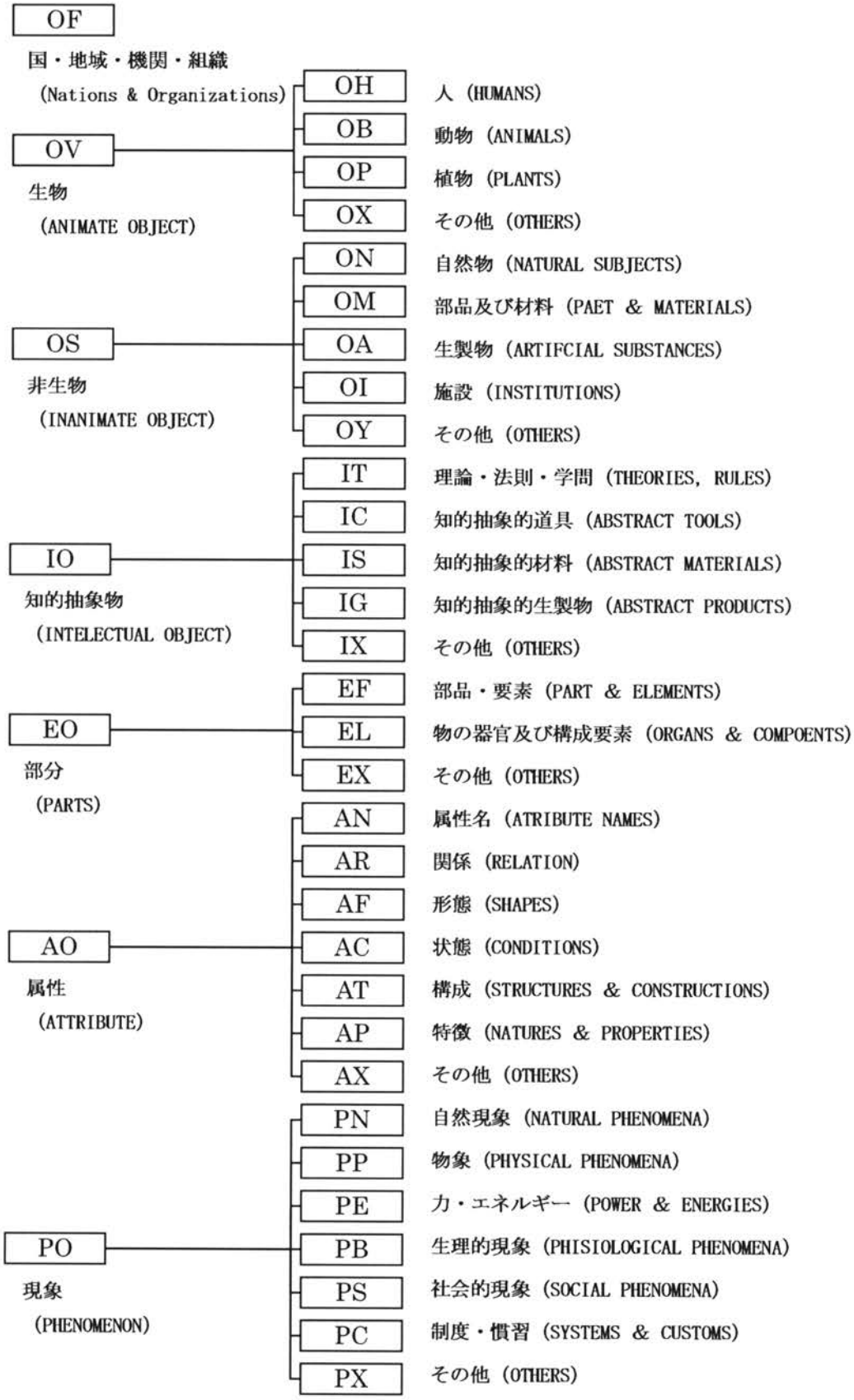


図 1-1　名詞意味マーカー体系表 (1)

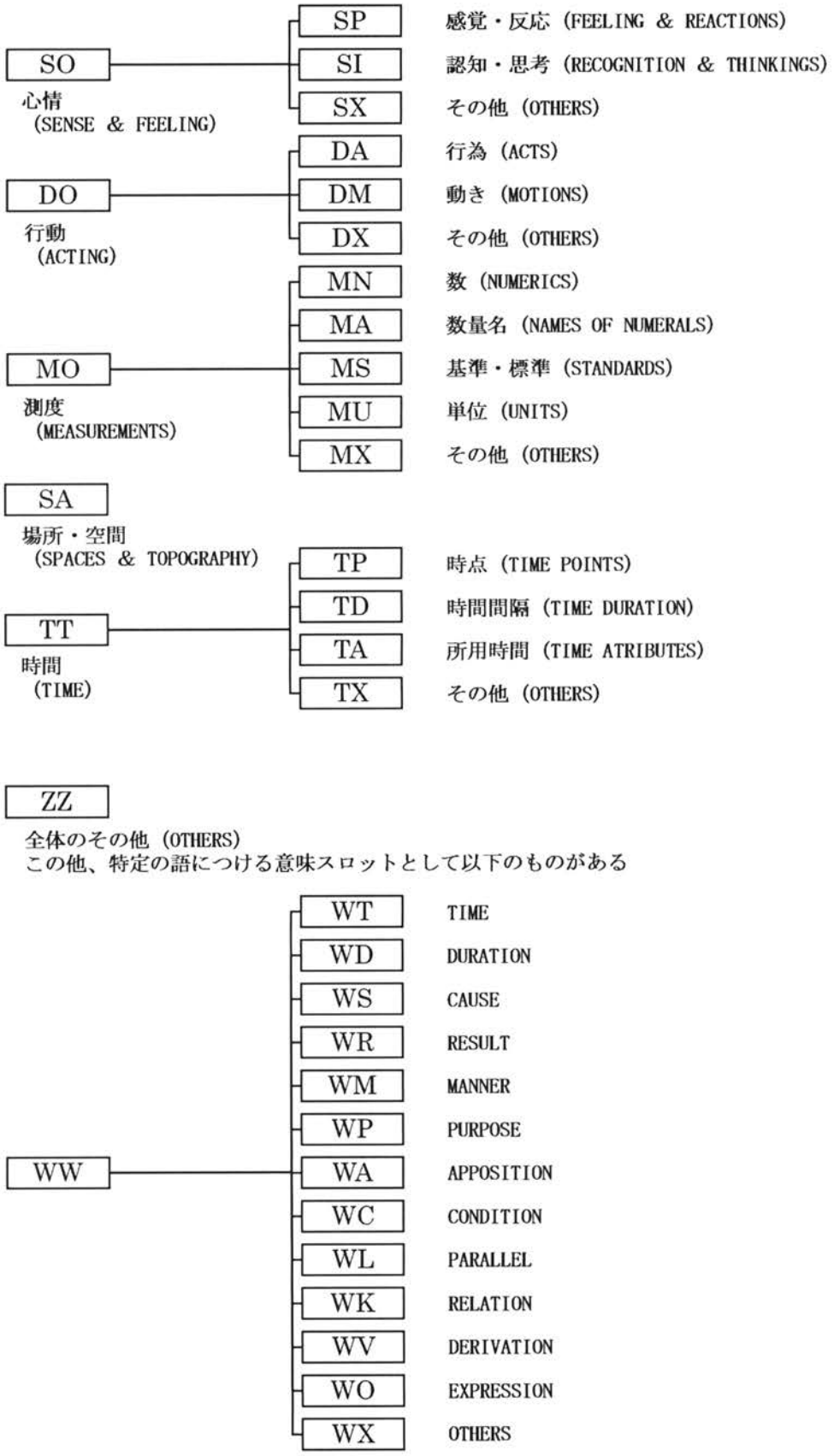


図 1-2　名詞意味マーカー体系表 (2)

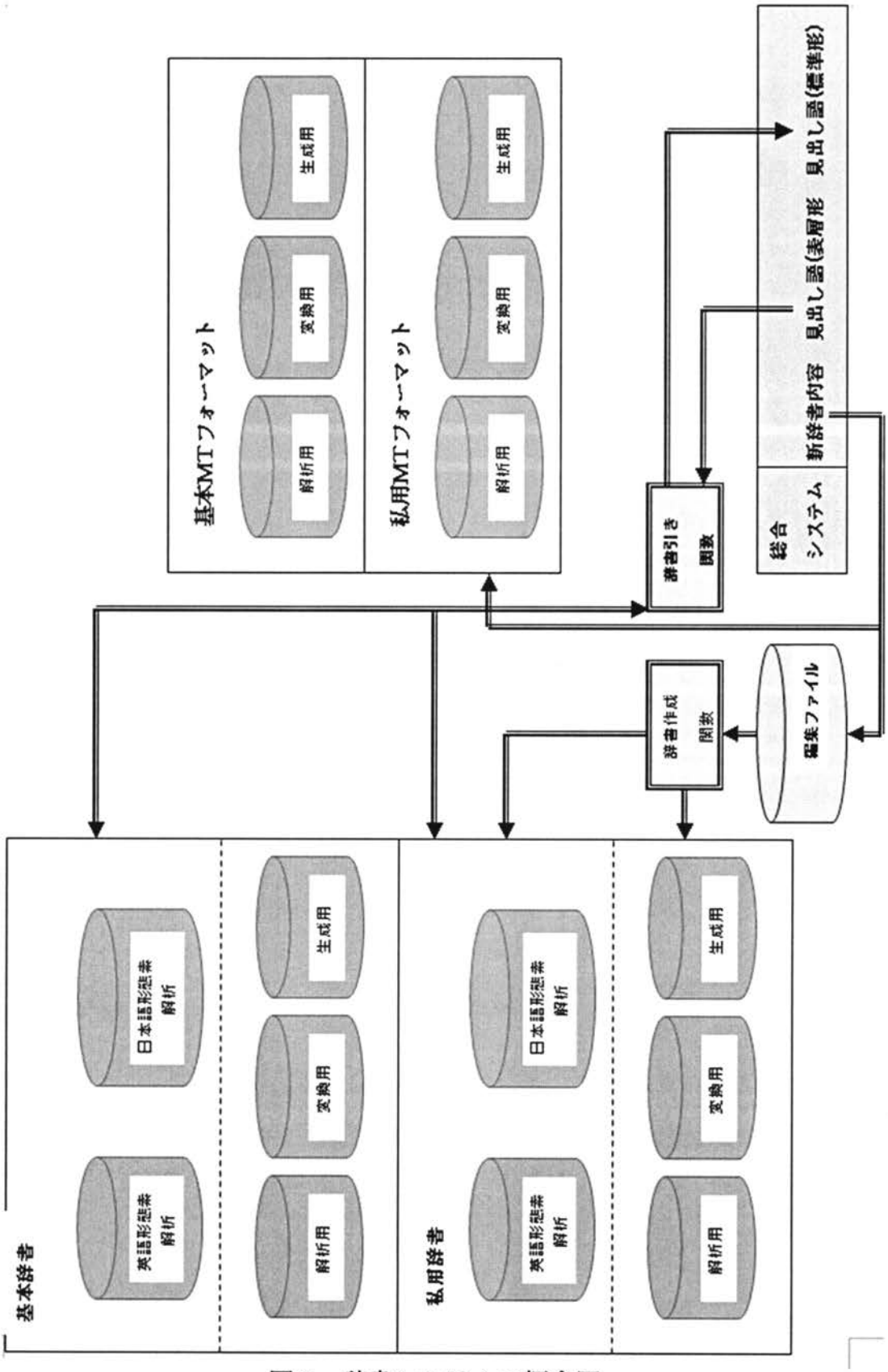


図３　辞書システムの概念図

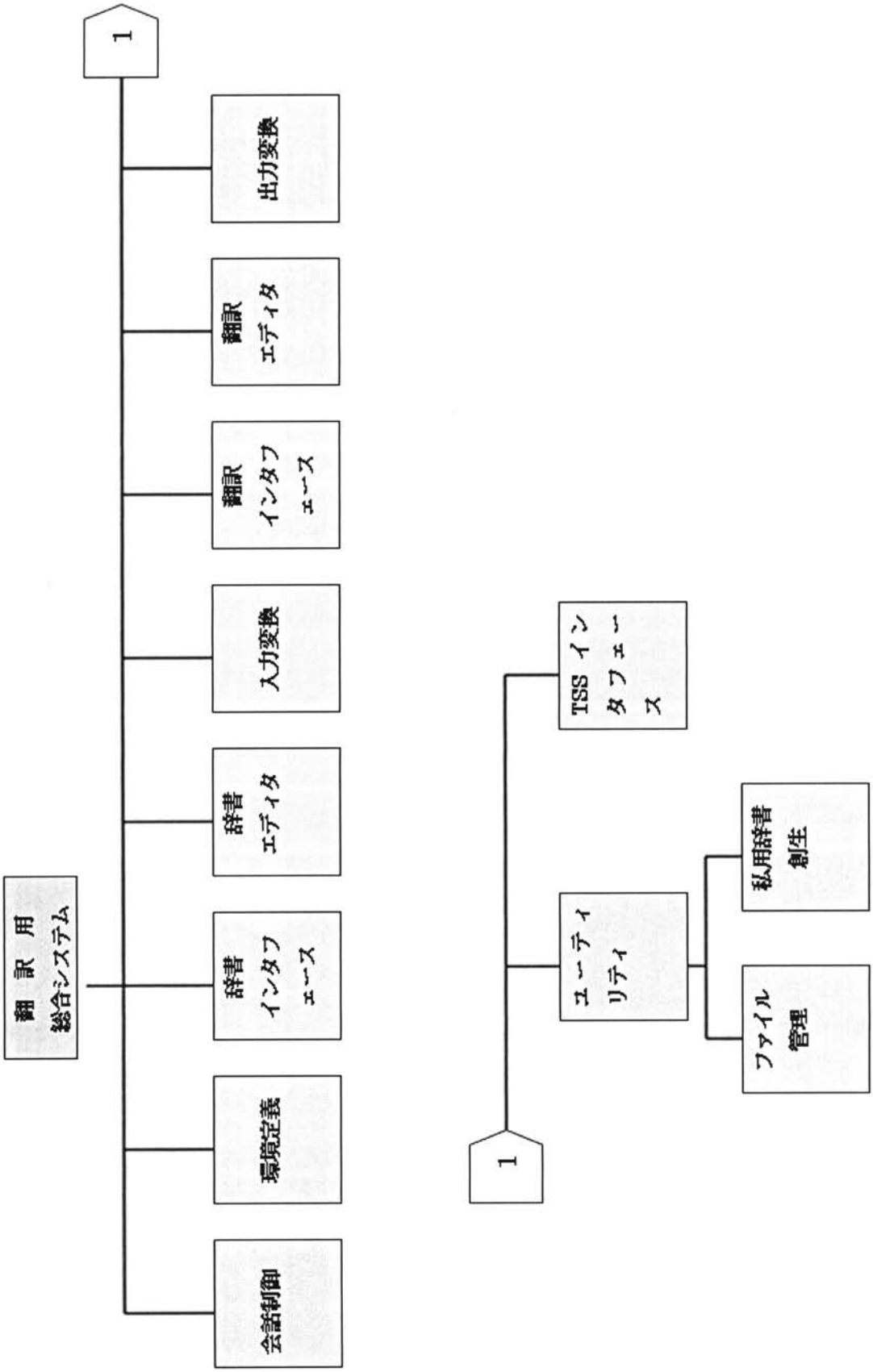


図 4　翻訳総合システム構成図

8.0 概要

今回は、日本を中心に、1987 年度～1994 年度（8 年間）にわたって、アジア各国と協力して行われた「近隣諸国間の機械翻訳システムに関する研究協力」について、報告書[1),2)]に基づいて、その概要を報告します。

8.1 はじめに

この研究は、財団法人国際情報協力センターの共同研究として行われました。

8.2 研究の目的

機械翻訳システムに関する研究協力事業は、最新の情報処理システムを近隣諸国と共同研究開発すること、この共同研究開発から派生する母国語情報処理技術の基盤を確立すること及びその他さまざまな技術交流や文化交流を促進することを目的として実施された。

技術面では、日本、中国、インドネシア、マレーシア及びタイのアジア 5 カ国を対象として、これらのどの国の言語にも翻訳できる方式の研究開発を行うことを、経済面では、情報処理分野の文献を実用的な速さで及び正確さで翻訳を行うことを目標とした。

8.3 研究の概要

① 名称：Multi-lingual Machine Translation (MMT) Project

② 対象国：中国、インドネシア、マレーシア及びタイ

③ 期間：1987 年度～1994 年度（8 年間）
1995 年度、1996 年度（2 年間）フォローアップ

④ 予算：162 百万円（87 年度）
332 百万円（88 年度）
639 百万円（89 年度）
880 百万円（90 年度）
902 百万円（91 年度）
981 百万円（92 年度）
981 百万円（93 年度）
948 百万円（94 年度）
累計　5,825 百万円
16 百万円（95 年度）
16 百万円（96 年度）
累計　5,857 百万円

⑤ 内容：
中間言語方式により、日本、中国、インドネシア、マレーシア及びタイのアジア 5 カ国間の言語を対象として、これらのどの国の言語にも翻訳できる機械翻訳システムの研究開発を行うこと。

8.4 実施体制

① 通商産業省からの委託を受けて、国際情報化協力センターが実施。

② 日本語側では、電子技術総合研究所、（株）日本電子化辞書研究所（EDR），（財）情報処理相互運用技術協会（INTAP）、関連企業（沖電気工業、シャープ、東芝、日本電気、日立製作所、富士通、三菱電機）、学術経験者等の協力参加。

③ 共同研究の相手方：中国計算機軟件與技術服務総公司（CS&S）、インドネシア技術評価応用庁（BPPT）,マレーシア教育省（MOE）、タイ国立コンピュータ技術センター（NECTEC）等。（マレーシアについては、1995 年にマレーシア国立翻訳研究所（ITNM）に移管、さらにマレーシア工科大学に移管）

8.5 内容及び成果

このプロジェクトでは、5 カ国語間の機械翻訳システムにおいて「中間言語方式」を採用した。研究開発の内容及び成果として、中間言語の仕様を作成、作成した仕様に基づく機械翻訳システムの開発及び電子化辞書・コーパスの作成、開発したシステムによる機械翻訳の実証実験等が挙げられる。

(1) 中間言語の仕様の作成

対象とする5カ国で表現された文章を独自の表現形式で表すための仕様(規則、記述方法等)を作成した。当初の仕様は、何度も見直しが行われ、改善が図られた。

(2) システム開発

作成された中間言語の仕様をベースにして、機械翻訳システムが開発された。このシステムは、入出力システム、翻訳支援システム、文解析システム、文生成システム、電子化辞書システム、文章ファイル管理システム、ネットワークシステム等から構成された。

(3) 電子化辞書及びコーパス作成

機械翻訳システムを実際に稼働させるためには電子化辞書が不可欠で、対象とした各国語について、基本語および専門用語の電子化辞書を作成した。また、実証実験の対象データとして、各国語で記述された3000の例文(コーパス)を作成した。

(4) 実証実験

開発された機械翻訳システム及び作成された電子化辞書を使用して、作成したコーパスを対象に各国語間での翻訳の実証実験を行った。最終的に、目標とした少なくとも50%の翻訳率をクリアした。

8.6 経過

1987年度から1994年度までの8年をかけて、機械翻訳システムの研究開発を進めた。1995年度及び1996年度には、8年間の共同研究の成果をもとに、フォローアップ事業を実施した。

(1) 機械翻訳システムの基礎研究(1987, 88年度)

最初の2年間で、機械翻訳システムを構成するいくつかのシステムと要素の基本仕様を設計し、試作した。約5000語の辞書と数十の例文を作成し、文法規則の仕様を設計した。また、限定された翻訳例文を使って、機械翻訳システムの処理能力を確認した。

(2) 機械翻訳システムの研究開発(1989~92年度)

次の4年間で、基本語辞書約5万語、情報処理関連の専門語2万5千語及び約3000の例文を作成し、文解析や文生成を制御する文法規則を試作した。また、追加した翻訳例文と任意の自然文を使って、機械翻訳システムの翻訳処理能力を確認した。

(3) 機械翻訳システムの評価と改良(1993、94年度)

最後の2年間で、(2)段階までに試作された機械翻訳システムを使って、翻訳処理能力実験を繰り返し、辞書、文法規則およびシステム全体を改善した。

(4) 機械翻訳のフォローアップ事業(1995, 96年度)

8年間の共同研究の成果をもとに、研究に参加した各国において、辞書・文法の改良・追加・機械翻訳システム実用化に向けた追加研究を独自に実施できるように支援した。

8.7 各国の研究機関の状況

(1) 中国

CS&Sは、本プロジェクトにより大きく発展した。また、本プロジェクトに参加した清華大学等の大学における言語情報処理に関する研究能力も向上した。

その後、CS&Sは、独自に中—英、中—日の機械翻訳システムを開発した。

(2) インドネシア

BPPTは、この分野ではゼロからの出発で、大きく成長、発展した。開発した機械翻訳システムを維持し、インターネットによる電子化辞書、翻訳サービスを行った。また、独自に1997年から2002年までへの自然言語処理に関する研究計画を策定し、実施した。

(3) マレーシア

ITNT は、機械翻訳システムの研究から撤退し、マレーシア工科大学へ引き継いだ。INTN の実施していたマレーシア語のスペルチェッカーの開発、電子化辞書の生産は、残念ながら中止となった。

(4) タイ

NECTEC は、1997 年初めに設立 10 周年の記念行事を行ったが、本事業とともに発展した。独自に CD-ROM による電子化辞書を開発・製品化（タイ語全般、タイ語用例、タイ語同義語・反義語、タイ語—英語、英語—タイ語）し、その改訂版を製作した。

8.8 フォローアップ活動

CICC としては、これまでの 10 年間の研究開発の実績、培われた人的ネットワーク等を活用して、関係国との間で、機械翻訳に関する情報交換等を独自に実施した。具体的には、①研究者、研究機関リストの作成・提供、②電子メールやインターネット電子会議による情報交換、③各国リポート等のホームページによる情報提供、④本プロジェクトに関する小冊子の作成等を行った。

また、機械翻訳に関する新しいプロジェクトの可能性を検討するなど、関係機関との連携を図った。

8.9 背景

当時、通商産業省は、アジアにおけるコンピュータ産業および AI 技術の発展を支援するために、ワークスステーションタイプの計算機を日本の各メーカーから、各国に提供し、各国がその計算機上にソフトウェアを開発する技術を習得するために、機械翻訳ができる環境を構築することを支援するといった目的があったという話を、私は耳にしたことがある。

参考文献

1) CICC 事業実績、P48-50,（財）国際情報処理協力センター、平成 18 年 6 月.

2) CICC のホームページより、国際共同研究開発、「近隣諸国間の機械翻訳システムに関する研究協力（MT）、http://www.cicc.or.jp/japanese/kyoudou/mt.html

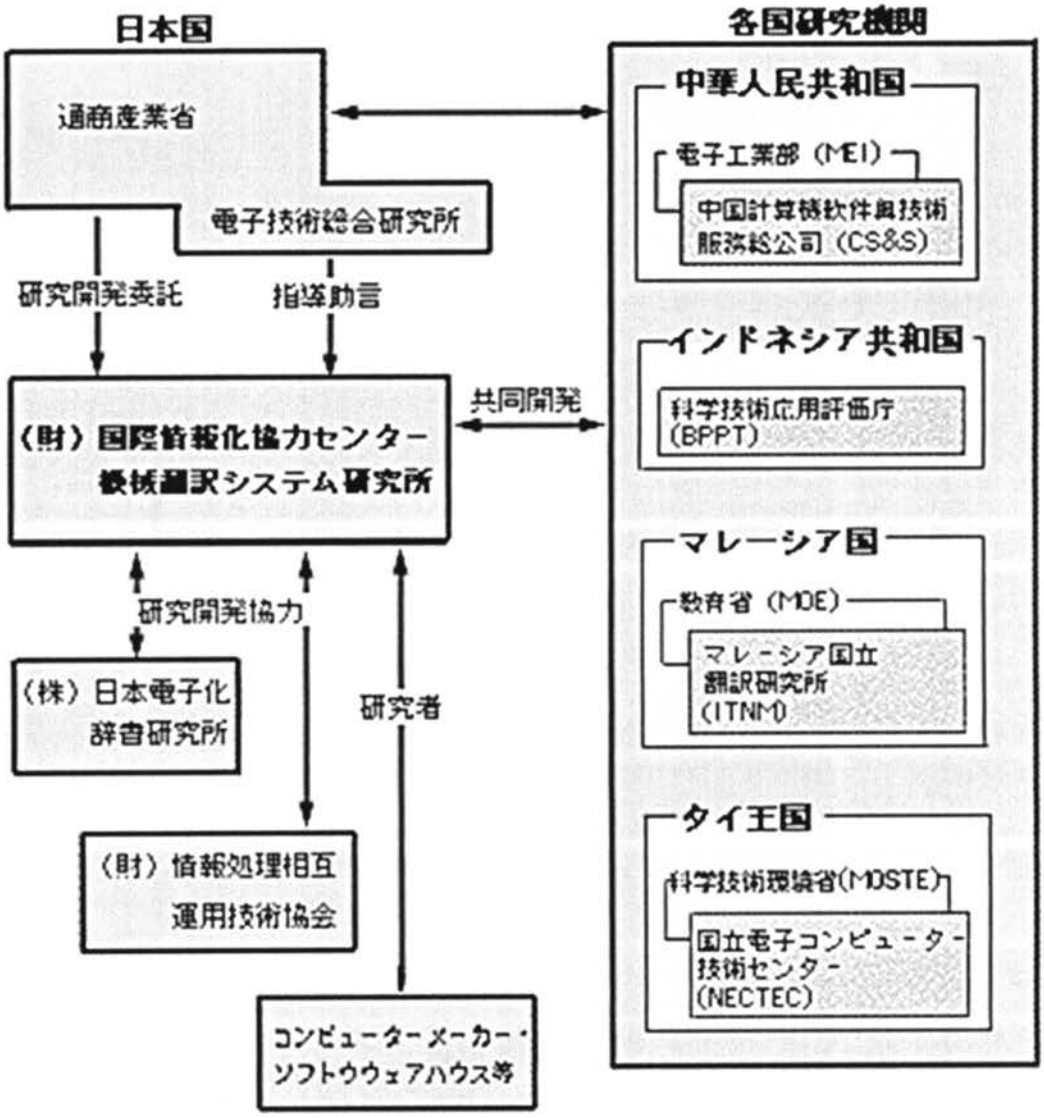

日本国
通商産業省
電子技術総合研究所
研究開発委託
指導助言
(財) 国際情報化協力センター
機械翻訳システム研究所
共同開発
研究開発協力
(株) 日本電子化
辞書研究所
研究者
(財) 情報処理相互
運用技術協会
コンピューターメーカー・
ソフトウウェアハウス等
各国研究機関
中華人民共和国
電子工業部 (MEI)
中国計算機软件與技術
服務総公司 (CS&S)
インドネシア共和国
科学技術応用評価庁
(BPPT)
マレーシア国
教育省 (MDE)
マレーシア国立
翻訳研究所
(ITNM)
タイ王国
科学技術環境省(MOSTE)
国立電子コンピューター
技術センター
(NECTEC)

図1　体制図

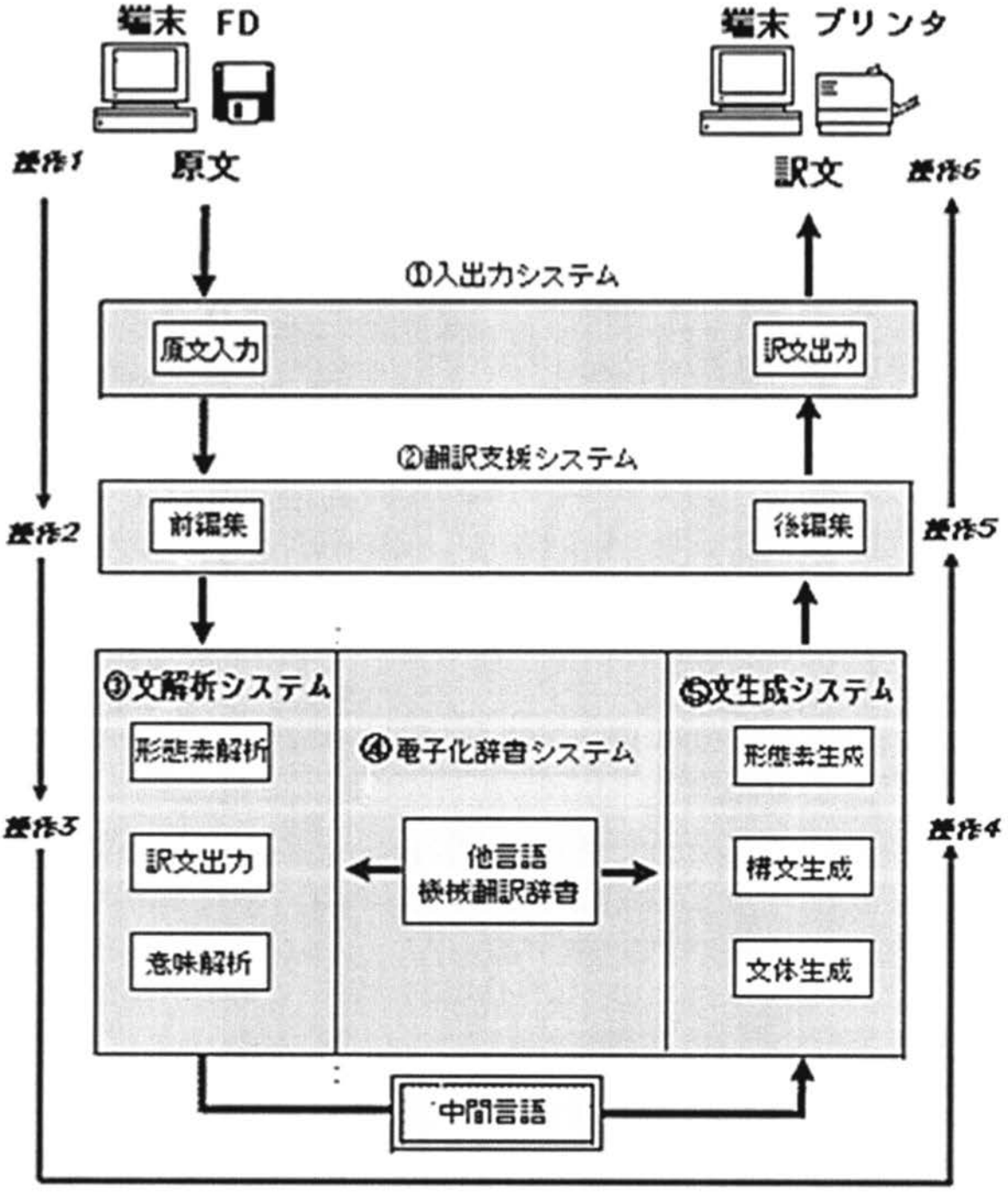


図２　機械翻訳システムの構成と翻訳の流れ

9.0 概要

今回は、特許庁を中心に行われた総合調査について、昭和60年度～63年度（1985～89）に行われた特許関連の情報に関する「日英自動翻訳システムに関する総合的調査」について、報告書[1]に基づいて、その概要を報告します。

9.1 はじめに

この研究は、財団法人　未来工学研究所に設置されましたプロジェクトを中心に調査が行われました。

プロジェクトのメンバー

[自動翻訳システム調査研究委員会]

委員長　草薙　裕
筑波大学／文芸・言語学系　教授

委員　坂本　義行
工業技術院／電子技術総合研究所
主任研究官

野村　浩郷
九州工業大学／情報工学部　教授

荻野　綱男
筑波大学／文芸・言語学系　講師

嶋田　祐輔（財）日本特許情報機構／
システム部　部長

福田　光男
(株) 特許データセンター／技術情報部　部長

内田　裕士
(株)富士通研究所／ソフトウェア部　部長代理

[未来工学研究所]

西澤　利夫　技術政策研究グループ　主任研究員（主幹）

八ッ橋　武明　研究企画室　主任研究員

小泉　則彰　技術政策研究グループ　主任研究員

前間　孝則　第1研究部門　主任研究員

太田　究三郎　第一研究部門　研究員

清水　房子　第一研究部門　研究員

[研究協力]

佐和田芳郎　京都大学／教育学部　助手

9.2 調査研究の目的

特許情報の国際的流通の必要性が論じられる中、我が国では公開特許の一部について英文抄録を作成し、海外の要求に応えていくうえで作業量とコストの問題が看過できなくなってきている。翻訳業務における要員確保を容易にし、人件費その他のコストの上昇を緩和するために、早期に対策を講じる必要がある。また、特許情報とともに、我が国の実用新案についてもその技術内容の紹介への要望も高まってきており、何らかの対応が早急に必要になっている。

一方、近年、機械翻訳システムの開発が進展し、実用に耐える能力を示すようになってきた。このような情勢のもと、昭和60年度以来、特許庁では公開特許抄録の翻訳実験、導入波及効果の分析、日本語ガイドラインの策定やコスト評価等、特許分野への機械翻訳システム導入について、課題の抽出・整理・対処方策の検討をすすめてきた。これらの成果を受けて、昭和63年度には実用新案書誌データの機械翻訳に焦点を置き、特許・実用新案情報の機械翻訳の実システムの課題と展望を検討した。

9.3 調査研究の方法

本調査では公開特許抄録の翻訳実験、波及効果の分析、日本語ガイドラインの策定やコスト評価を行い、特許分野への機械翻訳システム導入について、その課題の抽出・整理・対処方策の検討をすすめてきた。しかし、機械翻訳システムのための専門用語辞書の問題は未解決のまま残されており、後編集コストとの関係で辞書構築の範囲を明らかにすることがテーマのひとつになっている。一方、特許庁の政策課題として、「実用新案書誌」の翻訳がクローズ・アップされている。そこで、機械翻訳の導入が最も近いと思われる実用新案書誌について、辞書の問題に留意しつつ、機械翻訳導入の展望と課題について詳細に検討することにした。

① 自動翻訳用辞書整備におけるターミノロジーの検討
複合語が容易に創出されることは日本語の特徴である。これは概念の創発にとって有利な条件であるが、語彙管理上は問題がある。このため、辞書管理の観点から未登録語の抽出実験を行い、登録について検討する。

② 実用新案書誌データ等の自動翻訳の実証調査
1) 機械翻訳実験：実際の運用における機械翻訳の問題点を洗い出すため、実用新案書誌の翻訳実験を行う。また複合語を中心とする語彙登録がどこまで必要か、またコスト的に許されるかを、後編集作業との関係で分析する。
2) ワークステーション翻訳試行：コンピュータ技術の高度化に伴い、機械翻訳のシステムのワークステーション搭載可能になっている。そこで、昭和 61 年度の翻訳実験で用いた特許抄録により、ワークステーション上の翻訳システムの性能確認を行う。

③ 特許自動翻訳システムの基本構想と将来展望
過去 3 か年の調査と本年度の実用新案書誌データ翻訳実験の結果を踏まえて、特許情報分野への自動翻訳システムの導入のあるべき姿、その課題について総合的に検討する。

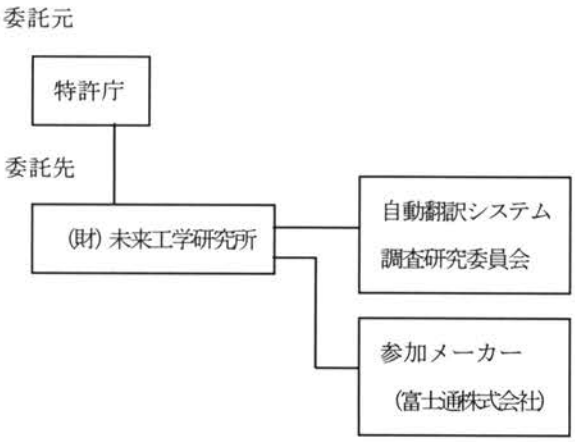


図 1.　調査研究体制図

9.4 特許自動翻訳システムの将来構想

9.4.1 知的所有権をめぐる国際動向と機械翻訳

(1) 国際動向と特許行政の課題

最近、日本の技術水準向上と国際的な技術交流・協力の高まりの中で、海外諸国に日本の技術情報を要望する声が高まっている。しかし、言語上の障害が我が国の国際的技術交流を大きく妨げている。特許の分野では公開広報抄録の 7 割が英語に翻訳されているのみであり、経済摩擦の進行のもとで、このままでは今後、国際交流上さらに大きい問題が惹起される可能性が強い。

他方、近年の情報処理技術の発展には目覚ましいものがあり、日本語・英語間の機械翻訳については、実用システムが開発・市販される段階に到達している。さらに、その高度化には多様な挑戦がなされており、相当な困難が予想されるものの、機械翻訳の有効性は今後徐々に増大するものと思われる。このような背景のもとに、特許庁では特許情報分野への日英機械翻訳の導入の可能性を考慮し、そのためのシステム評価と導入方法の検討を開始するに至っている。

ここで、現行特許制度の直面する大きな課題として次の 2 点を指摘することができる。まず第 1 に近年の出願件数の増加である。現在のペースで出願が増大し、それにつれて審査請求件数が増えていくと、迅速・的確な権利設定が不可能になる。現在、特許庁では、約 900 名の審査官によって、特許、実用新案あわせて年間 50 万件の出願処理を行っており、処理に要する期間は平均 2.5 年である。しかし、特許庁では処理する期間がさらに長期化するものと予測している。

特許制度の第 2 の大きな課題は、国際間の特許摩擦への対応である。1980 年代以降、世界経済の回復と、それにも拘わらず改善の進まない雇用情勢や構造不況産業の増大、貿易不均衡などを背景に各国で保護主義的風潮が高まり、貿易摩擦、経済摩擦が激化している。一方、1970 年代年代後半以降、先端技術開発が急速な進展をみせ、世界経済におけるそのウェイトの増大が激しい国際競争を引き起こすようになっており、ここに工業所有権に関する国際的摩擦、いわゆる特許摩擦が顕在化してきた。

このような情勢に鑑み、特許庁ではいくつもの総合的施策を展開しつつある。まず第1はいわゆるペーパーレス計画の推進である。ペーパーレスシステムは、特許庁への全出願書類ならびに審査に必要な資料をすべてコンピュータファイル化し、また、事務処理、審査・審判、広報発行等の各事務システムを統合して効率的な運用を図るものである。ペーパーレスシステムにより特許庁各部署での審査・審判の高度化・効率化は言うまでもなく、社会的な情報インフラ整備や出願企業の情報通信システム装備を前提に、出願者の特許情報管理技術、ひいては国際的企業競争力の向上が可能となる。

総合的施策の第2は各国の特許制度が有機的に機能する国際的枠組みの樹立である。このため、特許庁では国際間の特許制度及び運用の調和や開発途上国の特許整備協力に意をつくしてきた。しかし、特許摩擦の中には互いの国の特許制度・運用の相違の他、総合の理解不足に基づくものも多い。我が国は各国工業所有権制度の調和を主張し、また日米貿易委員会では、米国国際貿易委員会（ITC）の調査は日米の先端技術交流を阻害しかねないとの懸念を表明してきてもいる。

いわゆる日米欧の三極特許協力では各国間での共同プロジェクトが推進されている。とくに、コンピュータシステムの開発等で成果が上がっており、現在は次のものが進行中である。

① 特許出願処理の研究
② 特許バックファイル交換
③ 電子データ交換の標準
④ 化学データの標準化
⑤ 自動化サーチ手法
⑥ 分類システムの相関関係
⑦ 特許サーチ結果の交換
⑧ 特許情報の交換
⑨ 特許情報普及のガイドライン
⑩ 特許運用の調和（ハーモナイゼーション）
⑪ 自動翻訳

わが国にとって、特許制度の国際化の問題は、制度の問題と相互理解の不足との二重の課題である。特に後者については、わが国の特許情報の国際的流通の促進が大きな課題になっている。その意味でも、1985年の東京会合で「自動翻訳」がプロジェクトとして採択されたことは特筆に値する。

総合的施策として、最後に以上の各施策を可能にし、総合的な特許行政高度化を図るための特許法制度の全般的見直しがある。すでに特許庁は一般会計から特別会計に移行したほか、一定件数までの緊急審査の導入、出願における複数発明記述の許可、（財）日本特許情報機構（JAPIO）の発足や（財）工業所有権協力センター（IPCC）の設立等、外郭団体の整備統合もすすめられている。本件「特許自動翻訳システム導入」の推進も特許情報システムの整備・拡充の動きの一環といえる。

(2) 特許自動翻訳システムの展開

特許自動翻訳システムは広汎な展開を示す可能性がある。その第一は明細書全文翻訳への機械翻訳の適用である。国際的な技術開発競争の激化に伴い、外国への特許出願は今後ますます増大することが予想される。機械翻訳システムの辞書が英抄翻訳の実施を通して各分野に充実してゆけば、その利用は大いに考えられてよい。権利設定を目的とする特許文の翻訳が、リーズナブルな程度の前編集・後編集で可能となるかどうかがポイントであろう。

第2の可能性として、ペーパーレスシステムとの連動を前提に、英文のフレキシブルテキストデーターベース（FDB）の構築が考えられる。ペーパーレスシステム稼働下であらゆる出願書類が磁気媒体で提出されるとすれば、それ自体で日本語 FDB となるが、機械翻訳により、さらにその英語版が生成されるわけである。これが海外の特許審査における、あるいは海外の出願人にとってのわが国の開発技術の公知性を高める役割を果たすことは明らかである。また外国出願に直結しない以上、必ずしも厳密な翻訳でなくても良く、

この点でも機械翻訳が適合する可能性がある。一方で、キーワードの利用による技術解析手法も長足の進歩をとげており、英文 FDB の外国との相互交換は世界的な技術水準向上に資するものとなるであろう。

第3の可能性として、機械翻訳の実施経験が、現在、特許庁で研究されている出願の自動分類・自動解析システムの開発に貢献しうることがあげられる。機械翻訳技術は高度な意味処理の取り組みの方向に向かっているが特許文への現行機械翻訳技術の適用から、特許文特有の意味処理の課題が抽出される可能性がある。

最後に、特許分野における大規模な機械翻訳採用が機械翻訳技術全体の開発に貢献しうることを指摘しなければならない。機械翻訳は一定の水準での実用化を実現しつつ、一方で大きな発展の余地を残してもいる。特許分野は、そのマーケットとしての巨大さゆえ、この技術体系の成熟に大いに貢献しうるが、逆に技術の成熟によって特許分野が得るものも、そしてその社会的波及効果も大きい。ここでは、機械翻訳システムの「技術としての可能性」をふまえつつ、社会的に可能な投資の範囲を見極めていくことが重要であろう。

9.4.2 特許総合情報システムの将来像

(1) 特許情報システムの現状と動向

1990 年 10 月に予定されている特許庁ペーパーレスシステムの本格稼働に伴い、従来ペーパーベースで処理されていた受付から審査までの一連の事務処理が合理化される。最終的には、電子出願システムの定着により、受付・審査・公開といった一連の特許業務が基本的に全面機械化される方向にあり、情報管理・審査業務に関する業務の効率化が実現されるものと思われる。

しかし、トータルな特許業務としては、まだいくつかの課題が残されている。たとえば審査業務の遂行上必要となる各種分類コード付与業務、欧米諸国への情報公開に関する業務については現状では手作業で行われており、これらの効率化を目的とした機械化の検討が急務となっている。欧米からの情報公開要求への対応にも、今後さらに迅速化が求められることが予想される。この点で機械翻訳システムの果たす役割は大きい。

(2) AI システムの機能

このように、特許庁のトータルシステムとしては、受付、審査業務のシステム化以外に分類コード付与、翻訳、情報公開等を含めた総合的なシステムの果たす役割は大きい。

a. 分類コード自動付与システム

個々の出願に対して国際特許分類で定められた分類コードを付与する業務は、受付と審査業務の中間に位置しているが、現状では人手によって行われており、今後の出願数の増加による負担の増大を考慮すると、何らかの機械化処置が模索されるべきである。しかし本業務には専門的な知識・経験が必要であり、これをカバーするシステムは、現在は存在しない。

しかしながら、概念に関する知識情報処理については日本電子化辞書研究所（EDR）はじめいくつかのプロジェクトで基礎研究がすすんでおり、たとえば EDR の「概念辞書」にあたるものが全特許分野で構築されるとすれば、自然言語処理とくに構文解析技術の適用によって、人間の判断に近い分類システムが構築される可能性がある。

b. 特許自動翻訳システム

特許自動翻訳システムは、磁気媒体化された特許文書を入力として日英翻訳を行うものであり、ペーパーレスシステムのサブシステムとして位置づけられうる。しかし、機械翻訳システムとしては、①後編集等を含む運用体制の整備、②辞書等のシステム環境の整備といった翻訳システム固有の運用条件が必要となり、さらに③翻訳結果の管理、④情報公開機能（対外的な情報公開を支援する情報検索サービス等）に関しても考慮していかなければならない。

これらの業務システム AI（Artificial Intelligence）的な要素が強く、ペーパーレスシステムとは基本的に目的・守備範囲が異なっていると考えられるため、基

幹システムとのデータ連携を密にしつつも独立したシステムとして構築すべきであろう。すなわち、バックエンドシステムとしての特許AIシステムである。

(3) 特許総合情報システムのトータルビジョン

図2は、特許総合情報システムにおけるAIシステムの位置づけ、分類コード自動付与システムと機械翻訳システムの機能関連を図示したものである。機能ビジョンは次のとおりである。

まず、出願書類はペーパーレスシステムに電子的に受け入れられる。ペーパーレスシステム上での一連の審査過程と並行して、特許抄録や実用新案書誌が作成される。出願書類は「分類コード自動付与システム」によりコードが与えられ、抄録は機械翻訳される。この際、機械翻訳と分類コード付与は、共通の自然言語処理システムや専門用語辞書にサポートされる。出願書類と抄録類はデータベース化され、AIシステム内の情報検索システムからの検索を行う。その際、与えられた分類コードやキーワードが検索の重要な手がかりとなる。

(4) 特許総合情報システムの長期的スケジュール

図3は、特許総合情報システムの長期的スケジュールを示したものである。仮に1989年度を第0年度とすれば、1995年度ごろには機械翻訳システムを含む、特許総合情報システムの本格稼働が期待される。

9.4.3 将来への課題

汎用コンピュータ向け機械翻訳システムは、商用化されてわずか数年である。各メーカーで開発が本格化したのは1980年ごろと思われる。今後10年を考えると、システムの発展には相当のものが予想される。10年後、仮に現在とは比較にならない高度な処理が可能になっているとすれば、翻訳コストの削減、前編集・後編集コストの削減が大きく期待されよう。外国への特許出願支援等におけるマーケットの拡大もありうる。もちろん機械翻訳は商用システムとして稼働しつつもいまだ開発途上にあり、可動部分を活用しつつ、社会的に可能な投資がさらに続けられることが必要である。

機械翻訳システムは、自然言語処理という人工知能技術の最先端に深い連関を持っており、わが国の、さらには全世界の技術開発の最先端の課題に挑戦するものである。この観点からすれば、わが国の民間企業の共同事業として特許自動翻訳に対応したほうがよいとも考えられる。また、辞書の構築についてもメーカーの共同事業とし行ないうる可能性が大きい。

機械翻訳システムの市場開拓や、その効用を最大限に発揮させるとの観点から、機械翻訳のための新たな中核的組織を設立することや、あるいは既存の組織・団体を活用すること、さらには民間企業に業務を委託することなども考えられる、いずれの場合も、その公共的価値の巨大さ等の点から、国と民間との協力が必要とされよう。たとえば、特許抄録の日英自動翻訳システムが十分な辞書を伴って完成するとすれば、次のような各種の市場開拓が可能となる。

① 明細書全文の翻訳
② 英抄データベース、英文フルテキストデータベースのリリース
③ 自動分類・自動解析システム開発
④ 翻訳ノウハウを活用する外国出願コンサルティング
⑤ 特許関係以外の科学技術文献の日英翻訳
⑥ 辞書サービス

海外からの要望が高まっている実用新案書誌の翻訳については、本調査によって明らかなになったように、現段階から十分に機械翻訳システムの実稼働に移ることができよう。

参考文献

1) 財団法人　未来工学研究所：日英自動翻訳システムに関する総合調査 IV 平成元年3月.

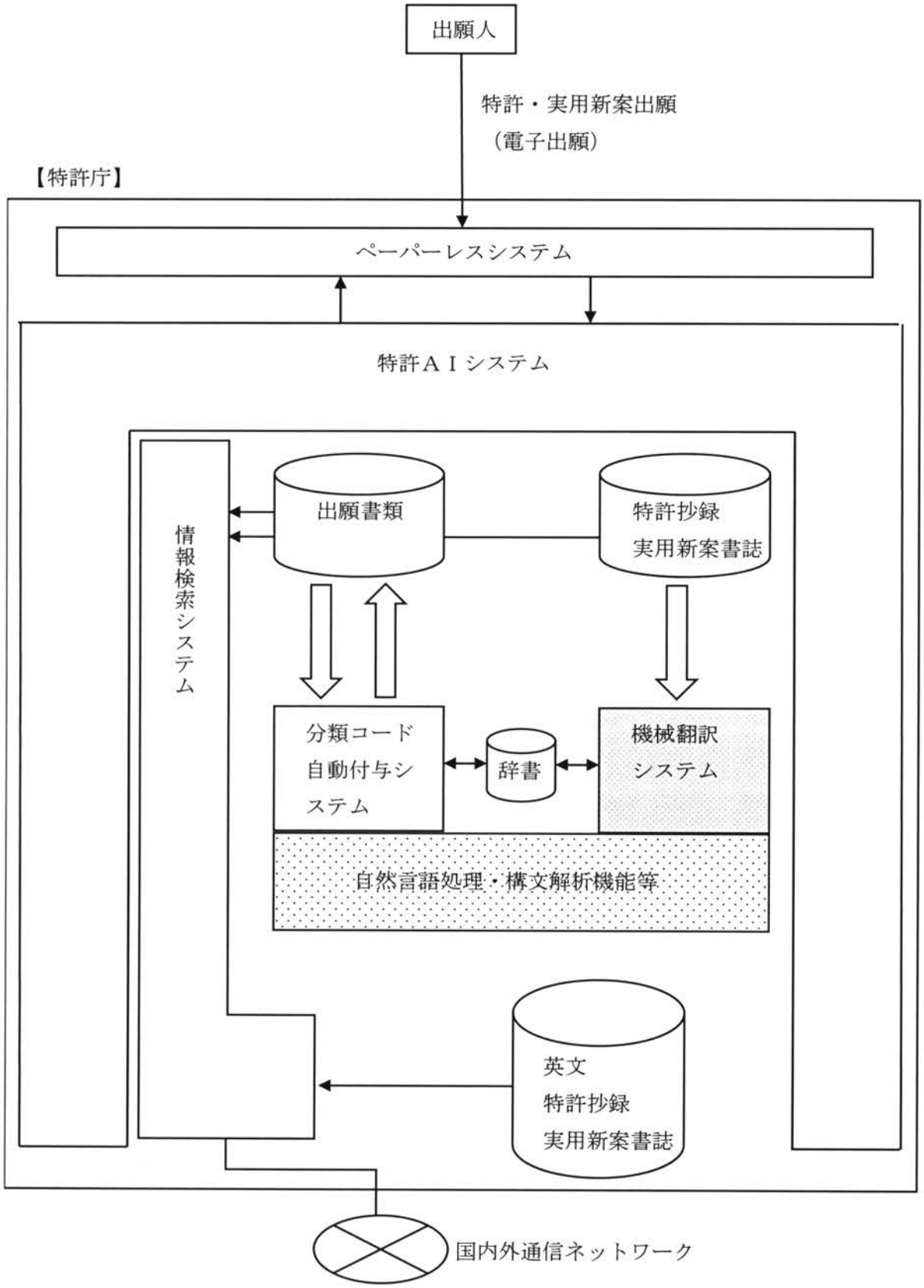


図2　特許総合情報システム概念図

システム　　期	0	1	2	3	4	5	6	7
1. 特許自動翻訳システム		←→		←				
2. 分類コード自動付与システム				←	→		←	

図3　特許情報総合システム長期的スケジュール

おわりに

本誌では、坂本義行先生からご寄稿いただいた「機械翻訳事始め」を、AAMT ジャーナル No.53（2013 年 6 月）以来前号の No.64（2017 年 3 月）までお届けしてきました。9 回に渡った連載が終了するのに際し、今回の連載のきっかけを思い出しました。それは坂本先生より頂戴した論文集でした。

坂本先生がご自身の喜寿に向けてまとめられた論文集『淹究』（エンキュウ、学問を広く究めるの意）には先生のこれまでの研究が記されていました。その論文集を拝読すると、機械翻訳そして自然言語処理の研究の軌跡がまるで歴史年表をたどるように頭に広がりました。と同時に、この感動を是非他の研究者の方々と共有できないか、という思いに至り、先生に機械翻訳の歴史に関する原稿を提案致しました。坂本先生にはご快諾いただけましたが、単発の記事で掲載できる範囲はやはり限られてしまうという問題に突き当たっていたところ、坂本先生の方からも連載記事としての扱いをご提案いただき、読切記事から連載記事へと企画変更する運びとなったのです。

その後、東京オリンピック・パラリンピックの開催が決まったことは皆様の記憶に新しいと思いますが、それに伴い、機械翻訳の研究も新たな展開を迎えております。

新たな展開を迎える機械翻訳の現状の一端を確認すべく、連載記事の題名（邦題：機械翻訳事始め、英題：History of Machine Translation）を翻訳してみました。利用した機械翻訳は、対応言語がアファル語（Afar）からズールー語（Zulu）までの 140 語、そして、その精度や利便性など高い評価を受けている Google 翻訳です。翻訳結果を見渡すと、邦題より英題の方がうまく訳せているようでした。とはいえ、こちらが確認できる言語はほんの数語であり、翻訳に原語の表現が残っていると翻訳失敗が分かるという程度ですが。どちらの言葉も問題なく訳される日が楽しみです。そして、この機械翻訳「事始め」から、1980 年代以降の機械翻訳「事」、そして、未来への機械翻訳「事納め」、つまり、長尾先生が常日頃おっしゃっている、バベルの塔に象徴される言葉の壁を打破する日、に向けて編集活動に邁進いたします。

AAMT ジャーナル編集委員会委員 小谷克則

AAMT

AAMTジャーナル 65号－特集号「機械翻訳事始め」

発行：アジア太平洋機械翻訳協会（AAMT）
ホームページ：http://www.aamt.info
住所：〒171-0014　東京都豊島区池袋2－55－2鈴木ビル3階
(株)日本システムアプリケーション内
phone：03-5951-3961　fax：03-5951-3966
編集委員会：宇津呂 武仁　小谷 克則　大倉 清司
阿部 さつき　釜谷 聡史　河野 弘毅　目次 由美子
表紙(図部分)デザイン：阿部 さつき
事　務　局：神崎 享子　荻野 孝野
印　刷　所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)
c/o Japan System Application Co., Ltd.
Suzuki Building 3F 2-55-2, Ikebukuro, Toshima-ku Tokyo 171-0014, JAPAN
aamt-info@aamt.info