AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.6
March 1994

アジア太平洋機械翻訳協会

1994 International Conference on Spoken Language Processing

音声言語処理国際会議

ＩＰＣＬＰ 94

人間と機械における音声言語の処理に関し、基礎から応用にわたる
重要な諸問題をとりあげ、最新の研究成果の発表と意見交換を通じて
この分野の学術の発展に寄与する事を目的とする。

日　　時　　平成６年９月１８日～２２日
場　　所　　横浜パシフィコ
（〒220 横浜市西区ミナト未来1-1）

共　　催　　日本音響学会・IEEE東京支部・欧州音声通信学会(ESCA)
協　　賛　　電子情報通信学会・情報処理学会・日本人工知能学会・IEEE信号処理学会
事 務 局　　㈱サイマル・インターナショナル ICSLP94事務局(03-3586-8691)

目　　次

エッセイ

奥深い翻訳の世界

機械翻訳に関して「使える」とか、「使えない」とかの様々な議論が世情を賑わしている。時には「こうすれば使える機械翻訳」というセミナーまで開催される有り様である。これらの議論の内容も混然としていて、機械の操作方法や処理手法から「使いこなせる」とか「使いこなせない」という場合もあり、また翻訳の精度の良し、悪しから評価するものや、翻訳そのものに関する考え方の相違、時には原文章のまずさを翻訳精度に転化してしまっているものまである。

翻訳にたずさわる人の呼称にも種々あって、呼称如何によってはそのステイタスも違う。即ち翻訳家、翻訳者、少しへり下って翻訳屋、更には機械翻訳従事者の中には翻訳技能者とか翻訳技術者などの呼称がある。

翻訳家は主として文芸翻訳を行っている人々で、原作者の思想や思考を感情豊かに表現し、きわめて文学性、文化度の高い翻訳を指向するものである。したがって原文の字句にはあまりとらわれず、独創性があり、読者のイメージをかきたて、感動を呼び起こす表現力が重視される。

翻訳者は主として産業翻訳に従事する人である。産業翻訳にもメディア翻訳（映画、ビデオ、歌詞の音声、字幕等の翻訳）とビジネス翻訳がある。またビジネス翻訳にも文科系、技術系がある。文科系翻訳には講演原稿とか契約書、提案書、報告書、連絡文書なども含まれる。言語、文字には長い歴史の間に培われた文化的、環境的、習慣的な背景が凝縮されており、同国人はそれを見ただけで即座にイメージとして理解することができるが、異文化人はこのような背景がないため、翻訳者は原文にない背景説明を加味して翻訳しようとする。これを「行間や余白まで翻訳する」と云うが、これも翻訳者の個性や表現力が大いに関係してくる。反対に文献や連絡文書などでは不要文、不要語は省略し、骨子、要旨のみ理解できれば良いことから粗訳、下訳だけですまされる事もある。然し技術系の文献、マニュアル、説明書などになると、要約や翻訳者の独自判断よる翻訳では困る場合があり、出来るだけ原文に忠実に、意図するところが正確に伝わるような翻訳が望ましく、逐語訳、直訳的なものになり易い。かかる観点から、現在の機械翻訳システムが直訳マシン的であり、マニュアルや説明書、技術文書など大量且つ高速翻訳に向いていると思われ、粗訳、意訳のためのものならば、パソコン翻訳ソフトの翻訳でもこと足りるといわれる所以である。

このように翻訳そのものについても種々の解釈ややり方、種類があり、明確な定義というものが見い出せないのではないかと思う。

さて、過日通勤途上の電車の中での出来事である。米国人少年と日本人少年が乗り込んで来た。小学校の同級生らしく、小綺麗な身なりをしていたが、米人少年は英語で、日本人少年は日本語で喋り始めた。最初はお互いに独り言を云っているのかと思って見ていたが、あまりにも繰り返し続くので、それとはなしに耳を傾けると小学校の課外授業についての会話をしているのである。そこへ例によってお節介焼きのオバタリアンが近ずいてきて、米国少年に「あなたは日本語が話せるの」と日本語で問いかけた。米国少年は小声で「スコシ…」とこたえた。更にオバタリアンは「エンパイアステートビルへ行ったことがある？　世界で一番高いビルよ」と尋ねた。米国少年はこれに黙って答えず、代わって日本人少年が答えた。「世界貿易センタービルに行ったことはあるヨ」と。一瞬白けたムードが漂ったが、オバタリアンはそそくさと次の駅で降りて行った。

英文を読んだり理解することは出来るが、英語で表現することは苦手という日本人は多い。現状の日英機械翻訳も同じである。反対に日本語が読め、理解出来ても、日本語で表現するのは苦手というアメリカ人もまた多い。お互いに得意とする母国語を用いて、意志の疎通をはかる事が今後増えて行くのであろうか、高学歴化、国際化が進展する中で、日常会話にもカタカナ英語が増え、スシ、テンプラなどの日本語が英語の中に浸透し、言語の融合化が進みつつある今日、尚更この感を強くする次第である。

（事務局長・星野禎男）

カントリーレポート

CS&SのLEDの紹介

（中国 国立ソフトウエアサービス公司/語源工程事業部）

副総工程師・語言工程事業部　総経理　関 維 忠

中国国立ソフトウェア／サービス公司(CS&S)の語言工程事業部(LED)は、1986年に設立された。

LEDはその中でも、研究開発(R&D)部門に属している。その主な業務内容は、市販の自然言語処理機器翻訳の研究開発である。

LEDは４つの下部組織、すなわち英中MT、中外MT、電子化辞書、そして国際協力プロジェクトMMTを擁している。

MMTは日本の通産省の主導で、CICCとの共同研究のために組織化されたものである。LEDでは、30人を超えるメンバーが、上記のR & Dプログラムに係わっており、現在、その大部分は言語学(MT)とソフトウェアテクノロジーのバックグラウンドを持つ研究フェローである。

LEDは、中国国内の国家MTプログラムと自然言語処理製品の開発に重要な役割を果たしている。第７次五カ年計画の施行に当たる1986～1990年に、中国政府は国家MTプログラムに数百万を投資した。私どもは、同計画において、国内部分のオーガナイザーとコーディネーターを努め得たことを非常に誇りに思っている。

また、LEDは中国の代表として国際共同開発プロジェクトMMTに参加し、プロジェクト遂行に向けて各方面と協調し、調整を行ってきた。

現在まで、LEDの活動を通じて、さまざまな成果が出てきている。例えば、1988年9月に中国初の商用ベースの英中MTシステム"Transtar"（トランスター）をリリースし、60パーセントを超える解析精度が得られ、翻訳スピードも毎時3000単語に達した。

1990年7月には、大規模ドキュメンテーションデータベースが開発された。この開発に関して、LEDから検索技術が提供された。

1991年3月には、1200語を超える中国語の動詞を含む、自然言語処理のための辞書を編集した。理論上の開発という面でも、我々はTranstarに関して、パーサおよびルールの能力と欠陥の再評価を行い、中国語パーサとルール説明システムを開発した。これらについては、すでにいくつかの論文で発表されている。

1991年以来、LEDはここに報告したように前述の方法に従い、更に研究開発を進めてきた。

A．英中MTシステム　Transtar

英語は分析的な抑揚言語に属する。これに対して、システムの品質はその解析の深さによって決まり、また、単語間の論理的意味関係もシステムの品質に影響する。パーサの目標は、解析される文の中の単語や語句の間の関係を反映する、１組の情報を作り出すことである。ツリー構造のルートがトランスファ処理において、カーネルと見なされ、各成分は言語対の両側に、想定された順番に従って配置される。

LEDグループは上記原理に従ってパーサと各ルールをまとめて、最適化した。その結果、新しい「Transtar－９２」の解析精度はＰＣ４８６マシーン上で実に70パーセント以上に向上し、速度も毎時3万語である。現在、Transtarの国内ユーザは数百人に達し、またアメリカ、シンガポール及び香港に販売されている。

B　共同プロジェクトMMT

我々はMMTプロジェクトのほぼ全プログラムに参加してきたが、中国側のオーガナイザーを務め、中国文解析と文生成の主な部分を担当してきた。
また清華大学、東北大学、南京大学の各機関と協力してMMT開発に従事してきた。研究開発の対象領域は、以下のとおりである。

すなわち、解析、翻訳支援、文生成、辞書サポート、技術辞書およびワープロ操作と OCR などの出入力処理である。1991年には、CICCと共同でMMTの国際シンポジウムを北京で開催した。常に日本側と連携を保ってきており、協調関係のムードは高まっている。今後もより一層このプロジェクト達成に向かって努力を重ねたいと考える。

C　中外 MT"SinoTrans"

概して、中外MTの特徴は、外-中MTと異なるところがない。しかしながら、中国語の構文解析は英語のものよりもはるかに難しいため、目標言語の解析精度は、幾分かの前編集処理を経ても４８６マシーン上でおよそ精度は70パーセントであり、処理速度は毎時およそ１万語である。

現在、1993年9月のSinoTrans発表以来、数人の国内ユーザがある。

D　辞書

周知のとおり、MTは、辞書によって提供される情報に基づいて処理を行う。そのため、MTの品質は主に辞書の整合性に依存する。辞書の場合、LEDは10種以上の英中専門辞書を開発した。 電子計算機工学、経済学、化学工学、貿易などがそれである。
累計の総語数は60万件に達した。中英辞書に関して４万語規模の一般領域辞書を制作した。また、中和専門辞書に構文/意味論的な情報を組み込んでいる。同時にまた、一般ユーザのために電子化辞書を開発しており、最初の製品は包括的英中辞書である。

この実用化にはそれほど時間を要しないと見込まれる。

中国におけるMT研究の開始は1950年代に溯り、当初はロシア語と中国語の組み合わせから始めた。このプロジェクトの進行は諸般の状況のため何十年間も中断されたが、再び70年代後半に再開された。中国のMT開発の黄金期は、1987年から始まったとういうこともできる。現在では、いくつかの組織や大学でMTに関する授業を開設しており、あるいはMTに関る研究を行っている。この傾向は今後も続くと予測される。

MTとして、英中システムに限らず、その他の言語例えば日中や独中、露中のMTが取り上げられる。

マーケティング戦略が、より重視されるようになった。

いくつかの企業が、独自の自己資金でＭＴに係わってきている。現在、Transtar、SinoTrans、露中システム及び中和システムの商業ベースの開発が進み市場で入手可能である。

理論上の研究の面では、LEDはMTに関しては世界のレベルに追いついたと言われるが、特に中国語文解析と文生成の点で精通している。中国は、数百万語規模のコーポラを達成したが、1995年末までにその50パーセントにタグ付けが完了する見通しである。同時に、種々の情報を持つ大規模辞書を作成中である。また、現在、事例を用いるタグ付けシステムを開発中である。

我々は、外国語と中国語の間で双方向的な翻訳の必要性を感じている。今日の努力研鑽は、将来必ず満足な結果を生むと期待される。この点について、今後のCS&SのLEDのMT研究開発のガイドラインは以下のとおりとなっている。

1. 我々は今後も積極的にMMTプロジェクトに参加する。中国側のすべての設定目標の達成を目指す。改良後のMMTシステムおよびLEDによって開発される商業MTシステムを用い、一般のために翻訳サービスセンターを開く予定である。
2. 知的製品であるMTシステムの開発は、我々が扱う諸言語の本質をどれほど熟知しているかにかかっている。この点を念頭に置き、解析と文生成について、中国語の正式文法、また中国語規則体系に十分配慮する。
3. 以前に取り上げなかった言語ペア 即ち、中国／ドイツ語や中国／フランス語のMTシステムに関して国内外の提携先を探している。

おわりに

現状では、世界人口の四分の一が母語としての中国語を使用している。また、栄光の中国文化を育んできたのはこの言語である。世界中の人々が中国についてもっと知りたがっており、また中国の人々も世界に関して知りたがっている。それにもかかわらず、中国語のパージングと生成テクニックは簡単に取得されない。しかも中国語からの外国語生成は別の重大な問題である。我々が直面するこれらの挑戦に超越するには、独自にベストを尽くすことと平行して、CICCが中国に設立を希望している OSI（コンピュータ間相互接続／相互運用計画）に協調し、積極的に参加するよう努めたい。広くMT世界における相互の技術的交流に大いに関心を持ちつつ、中国語／外国語MTシステムの輝かしい未来を信じて、今後も努力を続ける所存である。

研究報告

「JICST 頻度付き辞書とバイリンガルコーポラ」

JICST 技術開発部 芦崎達雄

平成5年11月22～24日、米国のワシントンにおいて「機械支援翻訳に関する日米国際ワークショップ」が開催されました。ＪＩＣＳＴから「JICST頻度付き辞書とバイリンガルコーポラ」について発表しました。

発表内容としては、JICSTの機械翻訳システムの概要とそれを利用した英文データベース(DB)の作成、機械翻訳辞書の作成、翻訳実行時辞書、頻度付き辞書、英文ＤＢからバイリンガルコーパスの作成等について行いました。

JICSTにおける機械翻訳システムの開発経緯としましては、科学技術振興調整費により研究・開発が行われた日英、英日の翻訳システム(μプロジェクト)において、翻訳辞書の開発を行ったことからでした。

μプロジェクトは、1982～86年の4年間にわたり、参加機関は、京大、電総研、RIPS、JICSTの４機関で開発しました。JICSTはμプロジェクトの研究成果をベースに、1986年から91年の５年間かけて、科学技術論文の抄録をを翻訳する日英機械翻訳システムの運用版システムを開発し、90年の夏から実際にJICSTの英文ＤＢ作成に利用しています。

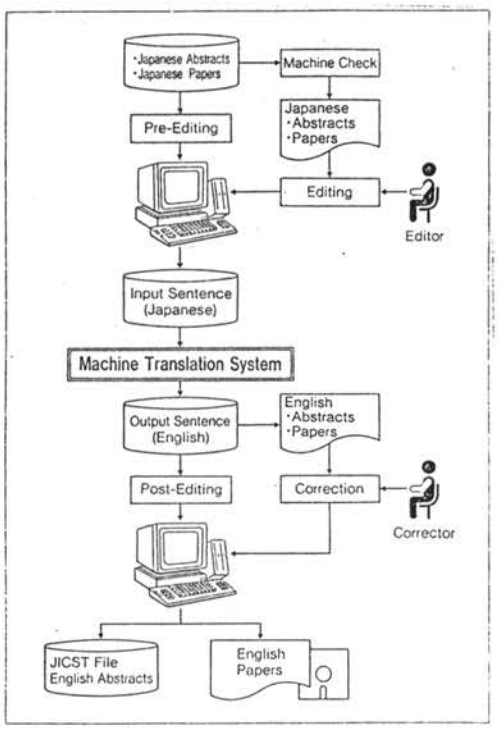


図1 JICST機械翻訳システム(J/E)

JICST MT System J/E

翻訳のシステムフローは図１の通りです。

JICSTの主な事業としては、国内外の科学技術論文について日本語の抄録を作成することです。機械翻訳を行う文としては、日本で発生した科学技術論文のタイトルと抄録です。論文の中には、英語でタイトルや抄録が含まれているものや、日本語のタイトル、抄録しかないというものもあります。これらを抽出・分類して、前編集プログラムを経て機械翻訳を行います。翻訳結果は、全件人手によって後編集チエックを行い、これをデータベース化しています。

MT using rate in JICST-E

	1990	1991	1992
Title	147,123 (63%)	156,511 (67%)	147,065 (60%)
Direct	82,649 (35%)	70,846 (30%)	73,448 (30%)
Human	48,982 (21%)	-	-
MT	15,492 (7%)	85,665 (37%)	73,617 (30%)
Tit+Abs	87,035 (37%)	76,104 (33%)	95,790 (40%)
Direct	77,607 (33%)	59,678 (26%)	82,459 (34%)
Human	506 (0%)	-	-
MT	8,922 (4%)	16,426 (7%)	13,331 (6%)
MT Total	24,414 (10%)	102,091 (44%)	86,948 (36%)
Total	234,158	232,615	242,855

表1 英文ＤＢにおける機械翻訳の割合

MT using rate in JICST-E

英文ＤＢの作成件数は、表１の通りです。

機械翻訳を利用して、データベースの作成を開始したのは1990年からです。1990年度の英文ＤＢの処理件数について説明すると、「Direct」とは著者の抄録をそのまま利用するもの、「Human」とは人手による翻訳を行ったもの、「MT」と記しているのが機械翻訳を利用したものです。「タイトル」というのは、タイトルと書誌的事項のみを英文にしたものです。タイトルだけでは、14万7千件あり、このうち機械翻訳を利用したものが1万5千件で7%に相当します。「タイトル＋抄録」というのは、タイトルと抄録を翻訳するものです。タイトル＋抄録では、8万7千件中機械翻訳を利用したものは8千9百件で4%、「タイトル」と「タイトル＋抄録」との合計では、23万4千件中で機械翻訳を利用したものは10%という実績でした。

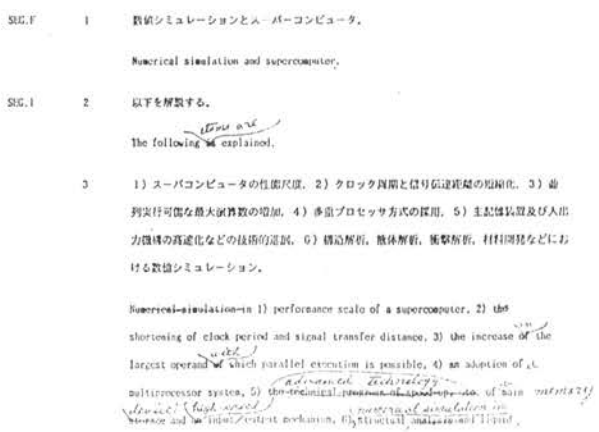
今のところ日本で発生した科学技術論文の抄録全てについて、英文抄録までは作成していません。英文抄録を作成する優先順位としては、政府関係出版物(一般には入手しにくい報告書類・公共資料)等が年間5千件位、その他日本が比較的進んでいると思われるメカトロニクス分野です。公共資料には、日本固有の農産物やかなり小さいレベルの地名とかが含まれており、また科学技術のほとんどの分野にわたっています。91年度では、「タイトル」だけでは15万７千件中機械翻訳で37%、「タイトル＋抄録」は７万６千件中機械翻訳で７%、合計で44%を機械翻訳を利用しました。92年度は総件数24万３千件中機械翻訳で36%処理をしました。1984年以来の英文データベースは、年間20万件前後、累計では現在200万件位になっています。（表２）

	produce	accumulate
84	18,454	
85	173,862	192,316
86	200,101	392,417
87	235,867	628,284
88	260,262	888,546
89	244,068	1,132,614
90	234,158	1,366,772
91	232,615	1,599,387
92	242,855	1,842,242

表２ JICST-Eファイルの作成・蓄積

後編集処理文例（図２）

機械翻訳した結果を後編集処理したものです。機械翻訳で十分処理できない単数、複数、前置詞等や構文の係り受けによる語順の変更等について、修正が入ります。一番難しいのは、係り受けや句並列ですが、時制についても、例えば「……について述べた」というような文章を過去形にするのか、現在形にするのか等は、まだ統一的な後処理が行われていません。後編集処理では、完全な日本語と英語の文というのがないので、２～４人の目が通れば最初に赤を入れ、次に青を入れ、更に黒を入れるという場合があり、後でみれば機械翻訳と大して変わりない


SIG.F 1 数値シミュレーションとスーパーコンピュータ.

Numerical simulation and supercomputer.

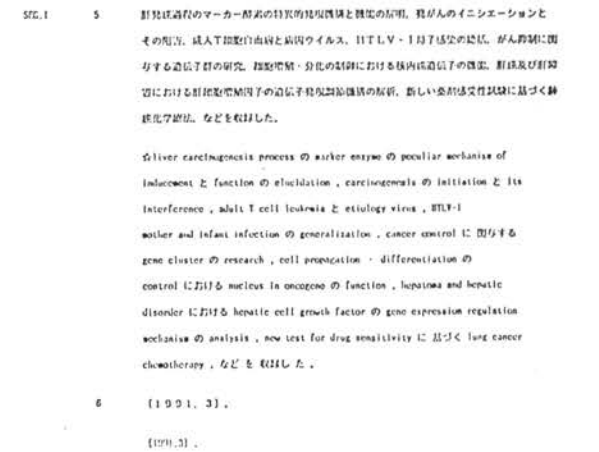
SIG.1 2 以下を解説する.

The following are explained.

3 1) スーパコンピュータの性能尺度. 2) クロック周期と信号伝達距離の短縮化. 3) 並列実行可能な最大演算数の増加. 4) 多重プロセッサ方式の採用. 5) 主記憶装置及び入出力機構の高速化などの技術的課題. 6) 構造解析, 流体解析, 振動解析, 材料開発などにおける数値シミュレーション.

Numerical-simulation-in 1) performance scale of a supercomputer, 2) the shortening of clock period and signal transfer distance, 3) the increase of the largest operand which parallel execution is possible, 4) an adoption of a multiprocessor system, 5) the technical problem of speed-up etc. of main advanced technology device (high speed numerical simulation in storage and the input/output mechanism, 6) structural analysis and flying

図２ 後編集処理例


SIG.1 5 肝発ガン過程のマーカー酵素の特異的発現機構と機能の解明, 肝がんのイニシエーションとその阻害, 成人T細胞白血病と病因ウイルス, HTLV-1母子感染の疫学, がん抑制に関与する遺伝子群の研究, 細胞増殖・分化の制御における核内癌遺伝子の機能, 肝疾及び肝硬変における肝細胞増殖因子の遺伝子発現調節機構の解析, 新しい薬剤感受性試験に基づく肺癌化学療法, などを収録した.

肝臓 carcinogenesis process の marker enzyme の peculiar mechanism of inducement と function の elucidation, carcinogenesis の initiation と its interference, adult T cell leukemia と etiology virus, HTLV-I mother and infant infection の generalization, cancer control に 関する gene cluster の research, cell propagation・differentiation の control における nucleus in oncogene の function, hepatoma and hepatic disorder における hepatic cell growth factor の gene expression regulation mechanism の analysis, new test for drug sensitivity に 基づく lung cancer chemotherapy, など を 収録した.

6 (1991.3).

(1991.3).

図３ 対訳辞書引き例

といったこともあります。

MTにとっての不適切例文

機械翻訳にとって適切でない文としては、まず長文で1文が150字以上、次に係り受けがはっきりしない文、そしてひらがな文字列が10文字以上続く文です。これらはいずれも翻訳処理が難しいことです。

機械翻訳処理を失敗した文は、日本語文のみ出力されるので、日本語だけで後編集者が英語に翻訳するのはかなり大変です。そこで、形態素生成の段階で日本語しかないという文については、再度形態素解析を行って、構文解析を行わずに、次に変処理を行いリスト出力を行います。名詞、形容詞等については、翻訳辞書の第一訳語を出力し、動詞と助詞については、日本語のままで、日本語の語順通りに辞書引き結果をを出力します。その出力結果を参照して後編集者が英語の構文を生成するのに役立つのではないかと思っています。(図3)

科学技術の抄録文は、日本語で書れた文章ですが、専門用語が多く、一般の人が内容を理解するのはかなり難しいことです。しかし、その分野の専門家が見れば、簡単に内容を理解できるようです。英語の抄録でも、その分野の専門家が見れば、名詞、形容詞、動詞等の重要語が翻訳されていれば、そこそこ理解できる程度のものではないかと思います。文意さえ解れば後編集で文章をきちんと生成しなくてもいいのではないかという意見もあります。現状の機械翻訳システムの場合は、一つしか訳語が出力できませんが、後編集者の参考のために同義語(他訳語)をリスト出力しています。

辞書の語数(表3)

翻訳辞書の規模は、名詞 (S&T)は、JICSTで作成した科学技術用辞書であり、名詞 (Medical) は、MEID辞書を JICSTの翻訳辞書構造に適応するように変換したものです。カッコ内のProper Noun は地名・機関名等です。動詞の語数は、日本語側で1万5千語程度、英語側では5千語程度です。科学技術分野ではサ変動詞が非常に多く使われており、例えば

「Aをデータ処理する」というときには、「Do ～ Data Processing」という代動詞＋名詞として翻訳するのではなく「Aのデータを処理する」といった構文として扱い、英語として読みやすい表現を採用しているので、日本語側の動詞の数が多くなっています。

「比較検討する」とか「比較研究する」は、「比較し検討する」のか、「比較に検討する」なのか判断が難しい語が多く見受けられます。動詞連続なのか副

P.O.S	JAPANESE	J-E	ENGLISH
NOUN(S&T)	352,016	352,160	294,250
NOUN(MEDICAL)	183,092	183,092	119,306
PROPER NOUN	(24,896)	(24,896)	(806)
PRONOUN	30	30	47
VERB	15,256	14,398	4,971
ADJECTIVE	6,795	6,795	7,088
ADVERB	451	451	2,123
DETERMINER	155	155	48
CONJUNCTION	87	87	34
POSTPOSITION	102	87	–
AUXILIARY VERB	115	115	9
AFFIX	279	252	–
PREPOSITION	–	–	158
ARTICLE	–	–	3
CARDINAL NUMBER	–	–	30
ORDINAL NUMBER	–	–	30
UNIT	–	–	5
TOTAL	558,348	557,594	428,097

表3 翻訳辞書語数

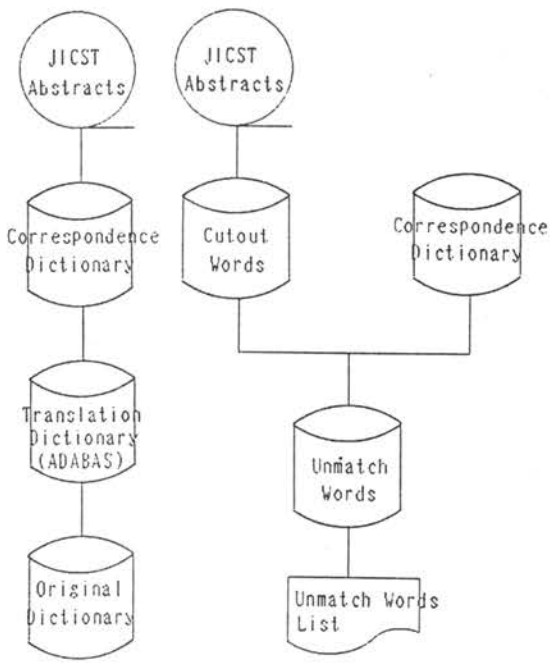


図4 各種翻訳辞書

図5 対訳辞書作成フロー

詞と動詞の組み合わせなのか、その時の個々の語の用法によります。そこで、実際に使用されている例文に当たって、辞書を作成し、文法的にも対応を採るようにしています。日本語の形容詞だけについていえば、あまり量的には多くないですが、比較的多い語としては、「……的」、「……型」という形容動詞で、なるべく一語で英語の形容詞で表現が出来るものを作成しています。このため、英語の形容詞がかなり多くなっています。

地名とか機関名は、変換辞書の訳語をそのまま使用しているので、生成辞書はほとんど作成していません。また、英語のアブリベーションは、そのままで翻訳しないで出力しているので、翻訳辞書はほとんど作成していません。ただし意味属性が付かないので今後は作成する必要があります。英語辞書は、日本語辞書の変換訳語が確定してから作成するので、おおむね1年遅れで作成しています。このため語数については、日本語名詞が55万8千語に対し英語名詞は42万8千語です。まだ、英語生成辞書が未作成の語が4～5万語あります。

各種翻訳辞書(図4)

JICSTの抄録から対訳辞書を作り、さらに機械翻訳側で使う翻訳辞書、それから翻訳実行時辞書の順で辞書を作成していきます。

対訳辞書作成、補語抽出(図5)

対訳辞書の作成方法としては、JICSTの抄録から切り出し文字(語)……カナ漢字変換をして言葉を切り出す言語処理のプログラムがあるので、これを既存の対訳辞書と照合し、未登録語を抽出してそれから翻訳辞書を作成します。

TERMJET(図6)

対訳辞書の作成方法としては、新規作成候補語として「科学技術教育」という見出し語について説明します。

```
科学技術教育

カガクギジュツキョウイク

#01 科学技術教育の要諦　だれが本当に心配しているのか

Decay in science and engineering education. Who re
lly cares?

AA

#02 非科学技術者に対する科学技術教育　次の世代への新しい挑戦

Education of technology to non-technologists - A c
hallenge for the next decade.

AA
```

図6 対訳辞書作成候補語

対訳辞書の作成候補語は、オリジナルが英語で書かれている論文から、和訳された日本語のタイトルに対してキーワードを切り出して、原文の英語は何であったのかというような作り方をしています。

「科学技術教育」に対応する英語の表現で、最初の例で「Science and engineering education」(#1)で次の事例では「Education of technology」(#2)が「科学技術教育」に該当します。このような方法で、対訳語を収集していきます。

対訳辞書例(図7)

「組織」の対訳辞書の例で、JICSTでは、約150位の科学技術分類コードを使用しており、抄録に付いている出典の分野コードを付記し、「tissue」とか「texture」、「structure」、「system」等の訳語を収集しました。対訳辞書は、あくまでも日本語と英語の組み合わせを作っておく辞書です。このような方法で30万語の対訳辞書を作成し、その内25万語位を翻訳辞書に取り込みました。現在はこの対訳辞書から翻訳辞書を作成するという方法では辞書を作成していません。その理由としては、対訳辞書というのは

```
【登録・更新年月日】  880701
【個　数】            8894
【見出し語】          組織
【フリガナ】          ソシキ
【対訳1】             tissue
【情報源分野コード】      AA00        EC00
                          EH00        EJ00
【対訳2】             texture
【情報源分野コード】      AA00        CA00
                          DC00        YB00
                          YC00
【対訳3】             structure
【情報源分野コード】      HB00        HC00
                          HD00        QJ00
                          YB00        YC00
【対訳4】             system
【情報源分野コード】      NA00
【対訳5】             organization
【情報源分野コード】      QC00
```

図7 対訳辞書例

```
*
*************************************************************
*      日英解析変換（名詞）辞書
*************************************************************
$$$
01 【見出しNo                ] LEXN-1000 C _ I-_ (1 1 1 ) A/12  NUM      _   _   (14) (CODEANCO
01 【更新年月日              ] UPDA-1100 C _ I-_ (0 1 0 ) A/12G NUM      _   _   (24) (UP_DATE
01 【作成者                  ] NAME-1200 C _ I-_ (0 1 0 ) K/12G _        _   _   (22) (USER_NAME
01 【見出し情報              ] LINF-2000 C G I-_ (_ _ _ ) _/_   _        _   _   (64) (J_MIDASBI
02 【  日本語見出し語        ] LEXF-2100 C _ I-_ (1 1 0 ) K/12G _        _   _   (12) (J_LEX
02 【  語基                  ] LEXB-2400 C _ I-_ (0 0 0 ) A/12G NUM      _   _   (24) (J_BASE
02 【  読み                  ] LEXL-2500 C _ I-_ (1 1 0 ) K/12G _        _   _   (12) (J_YOMI
02 【  異形語                ] LEXD-2700 C _ R-_ (0 1 0 ) K/12G _        _   _   (22) (J_IKEI
01 【形態素情報              ] MINF-4000 A G I-_ (_ _ _ ) _/_   _        _   _   (64) (J_MORPH_INF
02 【  品詞情報              ] PHEN-4310 A _ I-_ (1 0 1 ) A/2   NUM      PHEN PHEN (14) (J_FR_INF
02 【  状態情報              ] PSTA-4320 A _ I-_ (1 0 1 ) A/2   NUM      PSTA PSTA (14) (J_BC_INF
01 【構文意味情報            ] SINF-6000 A G I-_ (_ _ _ ) _/_   _        _   _   (64) (J_MSENT_INF
02 【  構文品詞              ] JPOS-6110 C _ I-_ (1 1 1 ) A/8   _        JPOS JPOS (11) (J_CAT
02 【  品詞細分類            ] JSPS-6120 C _ I-_ (1 1 0 ) A/8   _        JSPS _    (11) (J_SUBCAT
02 【  意味マーカ            ] SEMX-6200 A _ R-8 (0 1 1 ) A/2   _        SEMX SEMX (21) (J_SEM
02 【  変換情報              ] TINF-8000 T G R-_ (_ _ _ ) _/_   _        _   _   (64) (E_MIDASBI
03 【    J-UID               ] JUID-8150 T _ I-_ (1 0 1 ) A/2   NUM_     _   JUID (14) (USAGE
03 【  英語見出し語          ] ELEX-8200 T _ I-_ (1 1 0 ) K/12G _        _   _   (12) (E_LEX
03 【  英語品詞              ] EPOS-6130 T _ I-_ (0 1 1 ) A/8   _        EPOS EPOS (11) (E_CAT
03 【  英語品詞細分類        ] ESPS-6140 T _ I-_ (0 1 1 ) A/12  _        ESPS ESPS (11) (E_SUBCAT
03 【  分野コード            ] FILD-8300 T _ I-& (0 0 1 ) A/8   _        FILD FILD (21) (FIELD_CODE
03 【  意味マーカ            ] SSEX-8400 T _ I-& (0 0 0 ) A/2   _        SEMX _    (21) (J_SEM
03 【  意味細分類            ] ESEX-8500 T _ I-_ (0 0 0 ) A/2   _        JSEM _    (11) (J_SEM_CAT
03 【    E-UID               ] EUID-8120 T _ I-_ (0 0 0 ) A/2   NUM      _   _   (14) (E_UID
03 【  他の英語見出し語      ] ETTF-8450 T _ R-_ (0 0 0 ) K/12G _        _   _   (22) (OTHER_CHICE
03 【  後接修飾情報          ] PSTI-8700 T G R-_ (_ _ _ ) /     _        _   _   (32) (MODIFYING
```

図8 名詞翻訳辞書構造

ソースが英語の論文であり、現在JICSTで翻訳を対象としている文は、日本で発生した科学技術論文なので、未登録語は日本語に由来する語や最新科学技術用語が多いためです。現在は、主として後編集結果から未登録語を抽出して辞書を作成しています。

日英解析変換辞書（図8）

名詞翻訳辞書の構造です。対訳辞書の場合は、階層関係をほとんど持っていませんでしたが、翻訳辞書の場合、階層構造をはっきり定めました。
(01)がトップ概念で、最初に見出し情報があり、その下(02)に見出し語とか語句とか読みとか異形語が並んでいます。また(01)の中にも形態素情報とか構文意味情報などがあります。構文意味情報(01)の中で、構文品詞、品詞細分類、意味マーカ等が(02)レベルになります。変換情報(02)の中に J_UID、英語見出し語(03)で階層関係を表しています。またチェックデジット的なもので長さバイトなど構造的なものも表しています。また文法の参照記号もあります。

これはμシステムとほぼ同じです。見出し語だとJ_LEX、品詞ですとJ_CATなどがこの類いです。

日本語品詞属性値(図9)

品詞のバリューテーブルで、品詞分類（日本語の場合11個)とか 品詞細分類(約50個)を使用しています。品詞細分類とは、名詞でいうと固有名詞、普通名詞、サ変名詞、動作名詞(和語動詞)、格助詞的名詞、接続助詞的名詞、補文標識、副詞的名詞、形動語幹、マル化サ変（……化というような類いの名詞）です。

名詞辞書（図10）

名詞の翻訳辞書の構造に従った記述のフォーマットです。例えば「頭」という名詞では、形態素情報、構文意味情報の階層として意味情報、意味マーカ、動詞共起情報等、特に動詞共起情報については、かなり複雑な記述を行っています。例えば「人が頭を抱え込む」という例です。かなりイディオム的な用法

```
;
;*****************************
;       日本語品詞表
;*****************************
$$JPOS
NOUN:NOUN:名詞
PRON:PRON:代名詞
VERB:VERB:動詞
ADJA:ADJA:形容詞
ADJV:ADJV:形容動詞
ADVERB:ADVERB:副詞
AUX1:AUX1:助動詞
POSTP:POSTP:助詞
AFFIX:AFFIX:接辞・接語
DETERM:DETERM:連体詞
CONJUNC:CONJUNC:接続詞
$$

;*****************************
;       日本語品詞細分類表
;*****************************
$$JSPS
KOYUU:KOYUU:固有名詞
FUTUU:FUTUU:普通名詞
SAHEN:SAHEN:サ変名詞
DOUSA:DOUSA:動作名詞
KAKUTEKI:KAKUTEKI:格助詞的名詞
SETUTEKI:SETUTEKI:接続助詞的名詞
HOBUN:HOBUN:補文標識
FUKUTEKI:FUKUTEKI:副詞的名詞
KEIDO_K:KEIDO_K:形態語幹
KA_SAHEN:KA_SAHEN:マル化サ変

GIMOND:GIMOND:疑問代名詞
KINSYOUD:KINSYOUD:人称代名詞
SIJID:SIJID:指示代名詞

WAGOV:WAGOV:和語
SAHENV:SAHENV:サ変
KASAHENV:KASAHENV:化サ変
TOKUSYUV:TOKUSYUV:特殊

TEKI_KEI:TEKI_KEI:～的
SIKI_KEI:SIKI_KEI:～式
KATA_KEI:KATA_KEI:～型
HOKA_KEI:HOKA_KEI:その他の形容
KEIYO:KEIYO:形容詞

SUU:SUU:数
SUURYOU:SUURYOU:量詞
KAKSUU:KAKSUU:数詞
JOSUUS1:JOSUUS1:助数詞

SETTO_GO:SETTO_GO:接頭語
SETTO_JI:SETTO_JI:接頭辞
SETUB_GO:SETUB_GO:接尾語
SETUB_JI:SETUB_JI:接尾辞
```

```
@K01:0279744:00 (NOUN     FUTUU   ) 92-05-20
0001 [見出しＮｏ                    ] K01 LEXN 0279744
0002 [更新年月日                    ] K01 UPDA 920520
0003 [作成者                        ] K01 NAME 戸崎
0004 [見出し情報                    ] M01 JL1F
0005 [  日本語見出し語              ] M02 J_LEX 頭
0006 [  読み                        ] M02 LEXL アタマ
0007 [形態素情報                    ] K01 JM1F
0008 [  前接情報                    ] M02 PREN 1
0009 [  後接情報                    ] M02 POSTN 1
0010 [構文意味情報                  ] M01 JS1F
0011 [  日本語構文品詞              ] M02 JPOS NOUN
0012 [  日本語品詞細分類            ] M02 JSPS FUTUU
0013 [  意味マーカ                  ] M02 JSEM EP,EL
0014 [  変換情報                    ] M02 TINF
0015 [    Ｊ－ＵＩＤ                ] M03 JUID 1
0016 [  英語見出し語                ] M03 ELEX head
0017 [  英語構文品詞                ] M03 EPOS N
0018 [  英語品詞細分類              ] M03 ESPS COM
0019 [  分野コード                  ] M03 FILD AA00
0020 [  意味マーカ                  ] M03 SSEM EP,EL
0021 [  Ｅ－ＵＩＤ                  ] M03 EUID 1
0022 [  他の英語見出し語            ] M03 ETTF caput
0023 [  動詞共起情報                ] M03 VCIF
0024 [    用例                      ] M04 CESM 人が頭で頭を抱え込む
0025 [    日本語見出し語            ] M04 VCJL 抱え込む
0026 [    英語見出し語              ] M04 VCEL be
0027 [    Ｅ－ＵＩＤ                ] M04 VCID 2
0028 [    英語構文品詞              ] M04 VCEP V
0029 [    格フレーム                ] M04 CSFR
0030 [      日本語表層格            ] M05 JFAC が
0031 [      日本語深層格            ] M05 JDEP 主体
0032 [      英語表層格              ] M05 EFAC ＳＵＢＪ
0033 [      英語深層格              ] M05 EDEP AGT
0034 [      変換対応                ] M05 ECRP 主体
0035 [    格フレーム                ] M04 CSFR
0036 [      英語表層格              ] M05 EFAC ＡＴＲＳＵＢ
0037 [      英語深層格              ] M05 EDEP MAN
0038 [      変換対応                ] M05 ECRP ＄ ｐｕｚｚｌｅｄ
0039 [    格フレーム                ] M04 CSFR
0040 [      日本語表層格            ] M05 JFAC を
0041 [      日本語深層格            ] M05 JDEP 対象
0042 [      共起名詞                ] M05 JKE1 頭
0043 [      日本語位置              ] M05 JPS1 DEL
0044 [  動詞共起情報                ] M03 VCIF
0045 [    用例                      ] M04 CESM メンテナンスを頭に置く
0046 [    日本語見出し語            ] M04 VCJL 置く
0047 [    英語見出し語              ] M04 VCEL take
0048 [    Ｅ－ＵＩＤ                ] M04 VCID 4
0049 [    英語構文品詞              ] M04 VCEP V
0050 [    格フレーム                ] M04 CSFR
0051 [      日本語表層格            ] M05 JFAC が
0052 [      日本語深層格            ] M05 JDEP 主体
0053 [      英語表層格              ] M05 EFAC ＳＵＢＪ
```

```
      [見出し語                     ] 001 _ JL1F
0001  [  日本語見出し語             ] 002 K JL1F K 頭(する
0002  [  品詞字数                   ] 002 D GOSL A 2
0003  [  読み                       ] 002 K LEXL K ホンヤクスル
      [形態素情報                   ] 001 _ JM1F
0004  [  前接品詞型                 ] 002 K YTTF K サ変
0005  [  後接品詞行                 ] 002 K YL1R K サ行
      [  前接情報                   ] 002 D PREN A 2
      [  後接情報                   ] 002 D PSTN A 13
0006  [  品詞情報                   ] 002 D YHTA A 12
      [構文意味情報                 ] 001 _ JS1F
      [  日本語構文品詞             ] 002 K JPOS A VERB
      [  日本語意味細分類           ] 002 K JSPS A SAHENV
0007  [  格パターン                 ] 002 D CASE A V1
0008  [    自動詞／他動詞区分       ] 003 D V1TT A T
0009  [  アスペクト                 ] 003 D BASP K 結果
0010  [    態                       ] 003 D DVOC K 受身，可能，使役，である
0011  [    否定                     ] 003 D JNIL K 有
      [    文型変化                 ] 003 _ SINT
0012  [      受身                   ] 004 D PASS A 1
0013  [      使役                   ] 004 D CAUS A 0
0014  [      可能                   ] 004 D CAPA A 1
      [  格支配情報                 ] 003 _ JCFR
0015  [    用例                     ] 004 D CESM K 人
0016  [    日本語表層格             ] 004 D CJFC K が
0017  [    日本語深層格             ] 004 D CJDP K 主体
0018  [    意味マーカ               ] 004 K GCSM A GH,OF
0019  [    構文型                   ] 004 D CSYF A N
0020  [    必須格／任意格           ] 004 D JOBL A 1
      [  格支配情報                 ] 003 _ JCFR
0021  [    用例                     ] 004 D CESM K 英語
0022  [    日本語表層格             ] 004 D CJFC K を
0023  [    日本語深層格             ] 004 D CJDP K 対象
0024  [    意味マーカ               ] 004 K GCSM A IC
0025  [    構文型                   ] 004 D CSYF A N
0026  [    必須格／任意格           ] 004 D JOBL A 1
      [  格支配情報                 ] 003 _ JCFR
0027  [    用例                     ] 004 D CESM K 日本語
0028  [    日本語表層格             ] 004 D CJFC K に
0029  [    日本語深層格             ] 004 D CJDP K 終状態
0030  [    意味マーカ               ] 004 K GCSM A IC
0031  [    構文型                   ] 004 D CSYF A N
0032  [    必須格／任意格           ] 004 D JOBL A 1
      [変換情報                     ] 003 _ TINF
0033  [    Ｃｏｎｄｉｔｉｏｎ       ] 004 D COND A 1
0034  [    用例                     ] 004 D EXAS K 人がスワヒリ語を日本語に翻訳する
0035  [    英語見出し語             ] 004 R ELEX K translate
      [    英語構文品詞             ] 004 R EPOS A V
0036  [    Ｅ－ＵＩＤ               ] 004 D EUID A 2
      [    格フレーム               ] 004 _ CSFR
0037  [      日本語表層格           ] 005 D JFAC K が
0038  [      日本語深層格           ] 005 D JDCP K 主体
0039  [      英語表層格             ] 005 D EFAC K ＳＵＢＪ
0040  [      英語深層格             ] 005 D EDCP A AGT
0041  [      変換対応               ] 005 D ECRP K 主体
      [    格フレーム               ] 004 _ CSFR
0042  [      日本語表層格           ] 005 D JFAC K を
0043  [      日本語深層格           ] 005 D JDCP K 対象
0044  [      英語表層格             ] 005 D EFAC K ＯＢＪ１
0045  [      英語深層格             ] 005 D EDCP A OBJ
0046  [      変換対応               ] 005 D ECRP K 対象
      [    格フレーム               ] 004 _ CSFR
0047  [      日本語表層格           ] 005 D JFAC K に
0048  [      日本語深層格           ] 005 D JDCP K 終状態
0049  [      英語表層格             ] 005 D EFAC K = ｉｎｔｏ
0050  [      英語深層格             ] 005 D EDCP A GOA
0051  [      変換対応               ] 005 D ECRP K 終状態
```

図9 属性値　　図10 名詞辞書例　　図11 動詞辞書例

は、うまく記述出来ないこともあり、文法処理で対応しているものもあります。

動詞の場合も名詞と同様な階層構造で、主として格関係を記述しています。(図11)

例えば「人がスワヒリ語を日本語に翻訳する」といった場合、変換情報では 人、translate、スワヒリ語、into Japaneseと出力します。Japaneseの格に前置詞のintoを出力するという構造です。

深層格(表4)

日本語深層格は、34持っており、μシステムと同様です。しかし、全部をうまく使いこなしているとはいえません。

意味マーカ(図12)

意味マーカは、主なカテゴリーが13、サブカテゴリーが約60あります。名詞の意味マーカを作成した時点では、翻訳だけを意図して作ったのではなく、物事を分類するといった観点から作成されました。そのため具象物、抽象物等といった概念から分類されています。意味概念は、人によりブレが大きく、改良するのにはかなり難しい問題があります。

日本語名	英語名	用例
(1) 主体	SUBject	～が
(2) 対象	OBJect	～を
(3) 受け手	RECipient	～に与える
(4) 与え手	ORIgin	～から受ける，貰う
(5) 相手 1	PARtner	～と協調する，異なる，～に関連する
(6) 相手 2	OPPonent	～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
00 時間	DURation	5分間加熱する
01 場所	SPAce	～に位置する，～で発生する
02 場所・始点	Space-FRom	～から帰る
03 場所・終点	Space-TO	～へ送る，～に到達する
04 場所・経過	Space-THrough	～を通る，上空を飛ぶ
05 始状態	SOUrce	5.5%から6%へ引き上げる
06 終状態	GOAl	英語から日本語に翻訳する
07 属性	ATTrioute	適応性に富む，欠ける，乏しい
08 原因・理由	CAUse	事故で死ぬ，～から分かる
09 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	～から成る，～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	～に適する，備える，必要な
25 役割	ROLe	議長に選ぶ，～として用いる
26 内容規定	COnTent	～と呼ぶ，述べる，みなす
27 範囲規定	RANge	～について，～に関して
28 提題	TOPic	～は，～とは
29 観点	VIEwpoint	立場から，～の点で
30 比較	COmpaRison	～より大きい，～に劣る，～を上回る
31 附伴	ACCompany	とともに，～に伴って
32 度合	DEGree	5%増加する，3キロやせる
33 陳述	PREdicative	～である
00 その他	ETC	

表4　日本語深層格

動詞辞書(図13)

「翻訳する」という動詞のKWICで、動詞辞書を作成する時に使用したものです。また動詞評価の時も同様の形で出力し評価例文を抽出し、翻訳実行を行いました。動詞の活用語尾を展開させて文字列マッチングして、抄録文を抽出したものです。複合サ変動詞のように訳語がほぼ一つしかないものはあまり評価の必要性はありませんが、和語動詞のように訳し分けの多いものについては、かなり効果があります。連体修飾の動詞については、必ずしもうまくいきませんが、主動詞については実際文章を再度翻訳することで、辞書データの評価を高めて行く方法を採ってきています。翻訳辞書を作成する時に、このような例文を参照しながら、訳語とか格関係に着目して翻訳辞書を作成しています。

ルール辞書(図14)

翻訳実行時辞書の構成で、名詞の辞書だけでも、辞書のサイズが約1Gバイト位あり、このままでは大きすぎて翻訳実行時辞書には使えません。そこで2バイトのKEISコードから16進法にコード変換して、圧縮してオリジナル辞書を作成します。更にこのオリジナル辞書を基に、ルール辞書、プロパティ辞書、形態素解析辞書等の実行時辞書を作成しています。

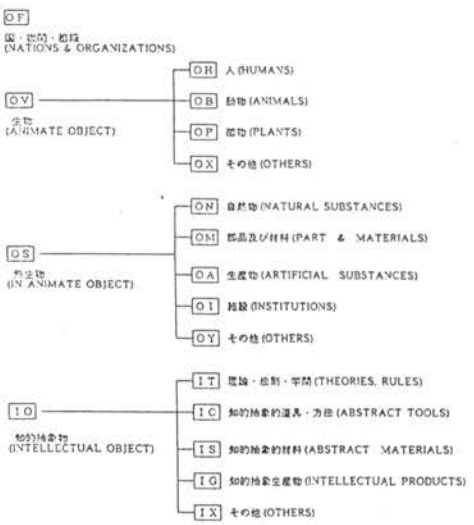


図12 名詞意味マーカ

その結果，ポリペプチド鎖は分子量３０万ダルトン以上と２２万ダルトン以下に	翻訳され
ルスＲＮＡ，エンドウリブロースｌ，５－二燐酸カルボキシラーゼｍＲＮＡが忠実に	翻訳され
また標準以外のコードへも対応でき，載込み	翻訳され
のプロコラーゲン鎖を同定したが，未同定のコラゲナーゼ感受性のポリペプチド鎖も	翻訳され
至適条件下で軟骨ｍＲＮＡを無細胞翻訳系で	翻訳し
ある言語の音声を認識し，他の言語に	翻訳し
有限要素法を用いる弾塑性解析のための大型計算機用のプログラムをパソコン用に	翻訳し
さらに，二次文献データベースとして，記述言語の指定があり，意味あるものは	翻訳し
この知識データを	翻訳し
翻訳は，翻訳パターンより係り受けの成立している単語を	翻訳し
本冊子を	翻訳し
複合利用問題における計画及び問題上の諸問題として，「複合利用問題」の一部分を	翻訳し
この方式を，ＣｅＬｉｓｐ系の会話プログラム可能な言語ＴＡＯに	翻訳する
的においては，社会的我用・便益の概念を用い，住民意識の変化を価値意識の変化に	翻訳する
うなものかを略述，ｎｅｔ，ａｎｎｏｕｎｃｅ，ｎｅｗｕｓｅｒｓの内容を日本語に	翻訳する
高い意味解析を行いながら技術文切から一般ビジネス文書まで，幅広い範囲の文章を	翻訳する
粉原料材を粉粒体が	摩耗さ
組にＢを添加することで，鉄系硬受の表面（摺摩耗性）を保存しながら，相手軸を	摩耗さ
，アルミナ，炭化けい素焼結体を引っかくとダイヤモンド先端に付着物が付いたり，	摩耗し
破砕機のライナが	摩耗し

翻訳され た
翻訳され た
翻訳され た情報はあらゆるシステムに接続可能であることなどである
翻訳され た
翻訳し た後，ポリペプチド鎖を電気泳動により分析
翻訳し ，音声に変換して出力する処理を，双方向に実時間で行う通訳電話について述べた
翻訳し て，枝柱式圧延での充填の効果に関する数値解析を試みた
翻訳し ，文字は翻字し，原表記は併記する原則がとられた
翻訳し ，整理することにより機械翻訳の多義性の解消ははかれる
翻訳し ，これらに対して重ね合わせ法を適用することにより訳語を得ることを特徴としている
翻訳し その大要を紹介
翻訳し た
翻訳する ことで実現する
翻訳する ことにより，複数の評価尺度を含む開発行為の総合評価が可能となった
翻訳する ことで，計算機を使った情報交換において何が問題となり，どんなルールが必要かを説明し
翻訳する 探記システムを組合した
摩耗さ せる現象と機構について述べ，粉粒体の性質が摩耗におよぼす主要因子，粒度，粒形，硬度，
摩耗さ せ易いという従来の欠点を克服した点が一大特徴であり，益産も開始した（３００万個／月）
摩耗し たり，破砕したりする
摩耗し ていく過程をそのまま計算機で追跡するプログラムを作った

図13 動詞ＫＷＩＣ

ルール辞書の事例(図15)

「愛する」という辞書の格パターンで、ルール辞書で、目で見える形で記述しています。文法と同じで意味マーカなどが入っています。動詞1万5千語プラス格パターンですから数量的には4～5倍はあると思います。動詞が呼ばれる度にルールが展開されるというような構造です。

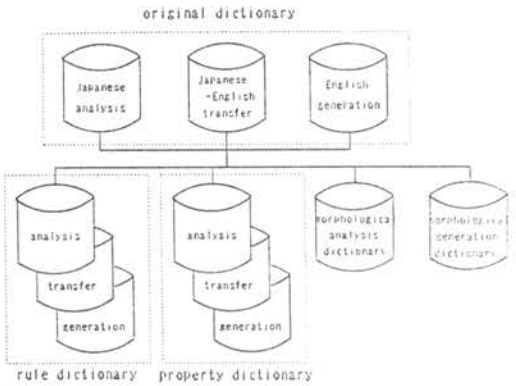


図14 翻訳実行時辞書

```
愛する_VERB_HENKAN.SG(#RI(愛する_VERB_HENKAN);)愛する_VERB_HENKAN.RR
(#VD(NAME=YR);#MC:(<#P);#P.#UD!=%;#SSO:#CDC(T,#P.#LX:#P.#UD:#HNK,(<#P>,(<YR
YR>;)愛する_V1_HENKAN.SG(#RI(愛する_V1_HENKAN1);)愛する_V1_HENKAN1.R
R(#MC:(<(#P #X));#K==#DS(<NOM1 NOM2 >,(<?>,LR);NOM1.#DP==SUB;NOM2.#DP==OBJ
NOM1.#EDC=EXP;NOM1.#ESC=SUBJ;NOM1.#EOB=YES;NOM2.#EDC=OBJ;NOM2.#ESC=OBJ1;
B=YES;#P.#EL="ｌｏｖｅ";#P.#EC=V;#P.#EU=2;)案ずる_VERB_HENKAN.SG(#RI(
Fフレマ _VERB_HENKAN);)案ずる_VERB_HENKAN.RR(#VD(NAME=YR);#MC:(<#P>;#P.#UD!=
#SSO:#CDC(T,#P.#LX:#P.#UD:#HNK,(<#P>,(<YR>);#CR:(<YR>;)案ずる_V1_HENKAN.SG(
案ずる_V1_HENKAN1.案ずる_V1_HENKAN2);)案ずる_V1_HENKAN1.RR(#MC:(<(#P
#X));#K==#DS(<NOM1 NOM2 >,(<?>,LR);NOM1.#DP==SUB;NOM2.#DP==OBJ;NOM1.#SM&&
=%;NOM2.#SM&&(OF,OV,OH,OB,OP,OX,ON,EL)!=%;#CR:**;NOM1 #EDC=EXP;NOM1.#ESC
M1.#EOB=YES;NOM2.#EDC=OBJ;NOM2.#EPL="ａｂｏｕｔ";NOM2.#ESC=A_COL2;NOM2.#
YES;#P.#EL="ｗｏｒｒｙ";#P.#EC=V;#P.#EU=1;)案ずる_V1_HENKAN2.RR(#MC:(<(#P
#X));#K==#DS(<NOM1 NOM2 >,(<?>,LR);NOM1.#DP==SUB;NOM2.#DP==OBJ;#CR:**;NO
AGT;NOM1.#ESC=SUBJ;NOM1.#EOB=YES;NOM2.#EDC=OBJ;NOM2.#ESC=OBJ1;NOM2.#EOB=
EL="ｄｅｖｉｓｅ";#P.#EC=V;#P.#EU=1;)圧する_VERB_HENKAN.SG(#RI(圧す
._VERB_HENKAN);)圧する_VERB_HENKAN.RR(#VD(NAME=YR);#MC:(<#P>;#P.#UD!=%;#S
CDC(T,#P.#LX:#P.#UD:#HNK,(<#P>,(<YR>);#CR:(<YR>;)圧する_V1_HENKAN.SG(#RI(圧
テキストの一部分を切り抜てました，切り抜ての発生した行数は20297行です
```

図15 ルール辞書

頻度辞書作成フロー(図16)

JICSTの抄録から切り出し語を作り、翻訳辞書と照合させ、頻度付き辞書を作成する流れ図です。

切り出し語では、従来JICSTでの言語処理プログラムのタームジェットとか形態素解析といった言語処理システムは使用せず、１文字ずつスキップさせた文字列集合を作成しました。

頻度辞書の切り出し語(図17)

「頻度情報付き翻訳辞書」は13文字列あります。これを10～2文字の文字列に切り出し、それぞれを一つのタグとして翻訳辞書と照合します。

翻訳辞書にはリファインされた用語があり、文字列集合と照合しました。 1カ月2000万組の文字列集合を、一か月分ずつ加算して行き、2年分照合を行って頻度情報を付与しました。

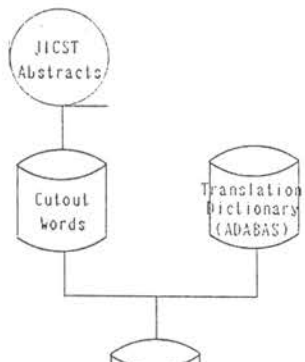


図16 頻度辞書フロー

頻度順の日本語見出し(図18)

このような文字列を切り出すと、サ変動詞の用法も拾ってしまうこともあります。またノイズとしてシスとシステムなどではシス（化合物のシストランスファ)のほうが頻度が高くなっています。

図17　切り出し語

頻度数	日本語見出し	ヨミ
[00175672]	こと	コト
[00154334]	より	ヨリ
[00130230]	検討	ケントウ
[00112595]	シス	シス
[00106622]	ステム	ステム
[00106323]	システム	システム
[00104084]	研究	ケンキュウ
[00097123]	結果	ケッカ
[00095815]	ため	タメ
[00092761]	開発	カイハツ
[00078210]	技術	ギジュツ
[00077694]	設定	ソウテイ
[00074083]	リン	リン
[00073925]	特性	トクセイ
[00068235]	構造	コウゾウ
[00068147]	影響	エイキョウ
[00067269]	また	マタ
[00067164]	処理	ショリ
[00065188]	解析	カイセキ
[00064451]	示し	シメシ
[00063282]	効果	コウカ
[00062644]	起因	キイン
[00061939]	実験	ジッケン
[00060089]	試験	シケン
[00058538]	方法	ホウホウ
[00057089]	調査	チョウサ
[00056746]	制御	セイギョ
[00056727]	プロ	プロ
[00055932]	物質	ブッシツ
[00055903]	変化	ヘンカ
[00054651]	評価	ヒョウカ
[00053016]	利用	リヨウ
[00052343]	問題	モンダイ
[00052312]	特徴	トクチョウ
[00050555]	これ	コレ
[00049785]	ロー	ロー
[00048341]	治療	チリョウ
[00049059]	もの	モノ
[00048757]	考え	カンガエ
[00048435]	反応	ハンノウ
[00047922]	スト	スト
[00047329]	装置	ソウチ
[00046527]	比較	ヒカク

頻度数	日本語見出し
[00039455]	ラン
[00038936]	設計
[00038604]	使用
[00038040]	データ
[00037856]	抄録
[00037122]	データ
[00036958]	関係
[00036950]	情報
[00036389]	ネル
[00036034]	投与
[00035920]	クリ
[00035657]	なか
[00035269]	管理
[00034897]	思考
[00034635]	分析
[00034329]	結晶
[00034210]	使用
[00033822]	検定
[00033654]	それ
[00033009]	診断
[00032744]	材料
[00032441]	遺伝
[00032231]	実施
[00031730]	方式
[00031412]	臨床
[00031392]	形成
[00031247]	表面
[00030905]	改善
[00030425]	採点
[00030302]	濃度
[00030147]	化学
[00030089]	応用
[00030057]	分析

図18 頻度順リスト

頻度情報(表５)

頻度情報としてのデータで、90年度は、日本論文が21万件で翻訳辞書が34万語の時に、約11万語に頻度情報が付きました。91＋92年分では64万件の文献に対して16万語の頻度情報が付きました。翻訳辞書の作成では、90年までは対訳辞書を基に作成していました。対訳辞書は、英語の論文をベースに辞書を作成しましたので、日本固有の動植物や食品等の語があまり登録されていなかったからです。91年以降は、日本語の論文をベースに翻訳辞書を作成していて、日本語抄録文との照合なので、頻度情報付きの用語が増加したと思っています。

頻度辞書のサンプル（図19)

JICSTで提供できる30万語の翻訳辞書のサンプルです。日本語見出し語、英語見出し語の他に、頻度情報を持っています。頻度情報では、実際の文字列が出現しなかった語や、長い文字、例えばアイアンフォメーション等は10文字を越えているので、頻度情報は付かない語もあります。この頻度情報の利用としては、翻訳辞書に採用するしないの用語の選択の他に、ワープロにおける同音異義語の優先順位決定に使用して、変換効率を上げることに役立つのではないかと思います。

バイリンガルコーパス(図20)

これはバイリンガルコーパスで、JICST英文ＤＢの

FREQUENCY	'90	'91 (RATIO TO '90)	'91+'92 (RATIO TO '90)
10,000 ～	92	169 (1.84)	442 (4.80)
5,000 ～ 9,999	165	272 (1.65)	560 (3.39)
1,000 ～ 4,999	1,346	1,952 (1.45)	3,404 (2.53)
500 ～ 999	1,371	1,956 (1.43)	3,294 (2.40)
200 ～ 499	3,346	4,705 (1.41)	7,893 (2.36)
100 ～ 199	4,425	6,394 (1.44)	9,528 (2.15)
50 ～ 99	6,923	9,628 (1.39)	14,262 (2.06)
20 ～ 49	14,283	20,273 (1.42)	26,482 (1.85)
10 ～ 19	16,148	21,612 (1.34)	25,174 (1.56)
5 ～ 9	20,212	25,996 (1.29)	28,362 (1.40)
1 ～ 4	44,458	50,402 (1.13)	45,240 (1.02)
TOTAL	112,769	143,359 (1.27)	164,641 (1.46)
NUMBER OF WORDS IN DICTIONARY	345,692	364,660 (1.05)	364,660 (1.05)
NUMBER OF ABSTRACTS (JAPANESE)	213,737	318,258	640,684

TABLE　FREQUENCY DISTRIBUTION OF TERMS IN JICST FILE
(1990 , 1991 and 1991+1992)

表５ 頻度分布表

作成結果です。後編集結果は人手でチェックして修正、入力しました。これを日本語とのランゲージペアを作ったのでものです。人手によっても、よく似た文章も、人によっては訳語や構文が変わってしまうこともあり、中々正しい日本語と英語を収集することは難しいです。

なお92年度版の電気編の情報処理という分野については半年分だけ抽出したバイリンガルコーパスのサンプルがあります。実際はこれを見て大いに検討して欲しいと思います。

英語見出し語 (1)	Indigo blue
用語No	M010013162
更新年月日	19901128
頻度	8
日本語見出し語	アイアン
読み	アイアン
(1)分野コード	AA00
英語見出し語 (1)	iron
用語No	M010013163
更新年月日	19901128
頻度	0
日本語見出し語	アイアンフォーメーション
読み	アイアンフォーメーション
(1)分野コード	DE00
英語見出し語 (1)	iron formation
用語No	M010013164
更新年月日	19901128
頻度	0
日本語見出し語	アイアンロー
読み	アイアンロー
(1)分野コード	AA00
英語見出し語 (1)	iron law
用語No	M010106449
更新年月日	19901128
頻度	5
日本語見出し語	愛育
読み	アイイク
(1)分野コード	AA00
英語見出し語 (1)	tender nurture
用語No	M010013165
更新年月日	19901128
頻度	0
日本語見出し語	アイウニア
読み	アイウニア
(1)分野コード	GP00
英語見出し語 (1)	glasses
用語No	M010013166
更新年月日	19910528
頻度	4
日本語見出し語	アイウエオ順
読み	アイウエオジュン
異形語	あいうえお順
(1)分野コード	AA00
英語見出し語 (1)	alphabetical order
用語No	M010106450
更新年月日	19901128
頻度	2
日本語見出し語	愛煙家

図19 頻度付辞書

(質疑応答)

Q: 頻度情報は何らかの形で公開提供する予定は?

A: 対訳辞書は30万語で内部利用で1200万円、翻訳システムに組み込む場合は3600万円で提供しています。その内で頻度情報が付いた16万語については安価で提供できないか検討中です。ただ今のところJICST のシステムは汎用大型機用なので W/S、PC用としては提供できないという問題があります。

Q: コーパスの提供は?

A: コーパスにどれだけの価値があるかは各社にとって違いがあると思います。無料では提供できないと思いますが、価格については未定です。

Q: 新語というのはどれ位発生するものですか?

A: 年間1万語程度の修正・追加を行っていますが、異表記語も多いです。科学技術分野を対象にしていますが、中にはイタメシとかボデコンみたいな最近の流行語まで入ってきても辞書に登録しています。

本来日本語の論文を対象にしていますが、日本のことだけが書かれているのではなく、海外の紹介記事などもあります。

さらに翻訳が難しい表現もあります。建築では夢を語る表現「永遠のかなたに美を追求したとか」、新車のキャッチフレースでは、「感性をくすぐる車」等という表現なども入ってきます。

2) MPUとメモリのスピードギャップを埋める高速DRAM.
2) High-speed DRAM which fills the speed gap between MPU and memory.
3) 磁気ディスク代替を目的とした廉価版メモリ.
3) Cost-oriented memory aiming at the replacement of magnetic disc.
4) システムの高性能化と小型化を図る特殊メモリ.
4) Special memory for system performance enhancement and size reduction.
本報では、各種ASメモリの基本構成、アプリケーション、ASメモリの開発動向を記述する.
This paper describes basic configuration and application of various AS memories and AS memory development trends.
91113119200O1
知的情報処理システムの現状と将来.
Present situation and future of AI information processing system.
主として応用上の観点から現在の知的情報システムの状況を調査した.
The current state of AI information system was surveyed mainly from the viewpoint of application.
AI技術の代表的応用事例について調査し、そこからAI技術の応用の現状と問題点、これからの動向をまとめた.
Typical applications of AI technology were surveyed, and present situation, problems and future trends of AI technology application were summarized ba

図20 バイリンガル コーパス

研究機関紹介

株式会社　富士通研究所

本ジャーナルで富士通研究所の自然言語処理の研究紹介を行ったのは、約2年前である。その時は、当社で研究開発し、事業部（富士通株式会社：当社の親会社）で商品化した機械翻訳システムATLAS を中心に、当社の研究開発活動を紹介した。日英翻訳から始め、独語、西語、韓国語等の多言語翻訳システムへと研究展開したことを述べた。またネットワーク化社会に対応した、商用パソコン通信ネットワーク上での機械翻訳サービスについても言及した。これは、言わば、機械翻訳あるいは自然言語処理の研究開発の幹に当たる部分であり、今回はその枝に当たる部分の研究活動を紹介したい。

まず最初に手がけたのは、OCR 文字認識後処理である。機械翻訳システムは、当然のことながら、電子化され、コード化された文章を入力し、それを翻訳するものである。 しかし、現実の翻訳作業の現場では、必ずしも翻訳したい文書が電子化・コード化されているとは限らず、紙しか存在しない場合も多い。その場合の重要なシステム構成要素はOCR である。そこで、自然言語処理技術を利用したOCR の強化に乗り出した。利用した自然言語処理技術は、形態素解析という技術である。形態素解析とは、ごく簡単に言えば、文を単語単位に自動的に分割することである。OCR 後処理とは、この形態素解析技術を使ってOCR の文字認識結果を吟味し、より確度の高い認識結果を得る技術である。この技術によりOCR 文字認識だけの場合に比べ、認識率が5%以上も改善出来ることを確認した。

次に取り組んだのが、オンライン辞書システムの研究開発である。これは、計算機がネットワーク化されたLAN 環境で、外国から来た電子メールを読む際や日本語文書を作成する場合に、その読解／作成中の文書を見ながら、辞書引きが簡単に行えるようにしたシステムである。辞書情報としては、日英・英日対訳や用例等が引き出せる。非常に簡単な操作で辞書情報を引き出せるようにするため形態素解析の技術を応用した。ユーザは、動詞の活用といったことを一切意識することなく（従来の辞書引きはこれを意識させるものが多かった）、画面上に表示されている文書中の単語の一部分だけを指定するだけで正しい単語の対訳情報や用例が取り出せるように設計されている。機械翻訳がこなれた訳を出せるようになるには、まだ相当の期間を要する。その間、このような辞書引きが機械翻訳の弱点を補ってくれると考えている。

機械翻訳の弱点を補う他の方法としては、対訳文書例をそのまま使うことが考えられる。当社では、英文ビジネスレターの作成の際、例文ベースを段落単位に管理し、それをビジネス用件に沿って組み合わせる組み合わせ式ビジネスレター作成支援システムの研究開発を行った。例えば、ある製品を売り込みたい時の手紙の組み立て方の典型は、製品の開発経緯から始まり、その製品の売り込みへ、さらに返事の依頼へと移り変わる。このシステムを使えば、ユーザは、このような段落遷移を適切に組み合わせ英文ビジネスレターを簡単に作成することができる

また前回の時も触れたが、機械翻訳の有効な利用分野としては、データベース検索との組み合わせがある。外国語データベースを検索し、その結果出てくる外国語文書を機械翻訳し、母国語で読めるようにすることである。この場合、出てきた結果が自分の本当に欲しい情報かどうかを、先ず判断できれば良く、それほどの翻訳精度は要求されない。斜め読みしてみて、出てきた結果が重要なものであれば、さらに精緻な翻訳を専門家に依頼すればよく、その時点で、重要でないものは、そのまま何もせず捨ててしまえるからである。ところが、ここでの問題の一つは、データベースから出てくる結果が長大な文章となり、読む時間が長くなる場合があることである。

そこで、この問題に対処するため、原文書をコンパ

ユーザヒアリング

機械翻訳の活用事例

長瀬産業㈱ 翻訳室長　堂野前　進

長瀬産業ではどのように機械翻訳を利用しているか、企業内翻訳の実情、機械翻訳の導入と運用、大量機械翻訳の手順、問題点、機械翻訳用の辞書についてお話したいと思います。

企業内翻訳

長瀬産業は翻訳会社ではありません。染料問屋から出発し、１６０年の歴史をもつ化学品専門商社です。一般企業、特に商社では、英語屋はいらない。仕事を覚える過程で英語や外国語を習得するのは良いが仕事中に辞書を引いたり、文献や専門誌を読むのは感心しないと考える管理職が多いと思います。

翻訳は正規の業務として認められず、従って仕事中に翻訳の必要が生じた場合、家に持ち帰り、苦労して翻訳するのが普通であり、会社の机にむかって辞書を引いたり、文献を眺めたりしている者は早晩サラーリマンとして落第の評価を受けると思います。

しかし、社史を読むと、合成染料の輸入を開始した明治初期には、リヨンに駐在事務所を設立し、フランス語の文献や情報を収集、日本語に翻訳をして、立派な技術資料を定期的に刊行して、合成染料の普及に努めていたことが解ります。

ひとつの産業が起り、新しい技術や知識が発展して行く時、翻訳の必要性は高く、また、企業の利益への貢献が期待されるので、このような時には、企業は、金に糸目をつけずにパンフレットやリーフレットを作り技術、知識、商品を普及させようとします。

しかし、ある程度、そのような技術、知識、商品が普及すると、翻訳は必要でなくなり、セールスマンが口頭で顧客に情報を伝えていけば良いことになります。儲かるようになった商品とか業界では、翻訳のニーズは低く、これから大いに発展する新分野に、新しい概念とか新製品を導入する業界に翻訳のニーズがあると思います。

現在では、当社でも、情報、通信、コンピュータに関連する技術分野での翻訳が主流であり、この分野では、新しい技術、知識、概念を基に、辞書に登録されていない専門用語が生まれては、消えています。特に、この分野と既存の成熟した技術分野との境界領域に新しいビジネスが広がり、それらの海外情報を吸収同化するためにも、大量の技術資料、文献を短期間に翻訳（理解）する必要があります。

このような学際、業際分野で、複数の技術領域の専門用語に精通し、内容を正確に理解するには、専

（⇦前頁より続き）

クト化する技術の研究開発を行った。これは、原文書の意味を大きく変えることなく、原文書を短くする技術である。基本的方法は、各文の中心となる名詞や動詞を残し、修飾語句を取り除くというものである。この基本方法をいろいろな角度から精錬し技術開発を行った。英文記事の対象にした実験では、意味を変えることなく原文書を約６割にコンパクト化することが出来た。

以上のように、当社では、機械翻訳に関わる周辺技術の研究開発に取り組んできた。この研究開発のベースにある考え方は、「身近に使える自然言語処理」である。自然言語処理は、意味解析や文脈処理に代表されるように大変難しく、哲学的な要素まで含んだものとなっている。このような幹に相当する技術に対しては、地道な息の長い研究開発が必要になる反面、成果としては世に出にくい。そこで、枝とも言うべき、「もう一工夫で成果が出せるもの＝使えるもの」も考えていくという姿勢である。

当社としては、今後ともこの両方の道を追求し、世の中に役立つ技術を開発していきたい。

（杉山健司）

門分野が細分化され、非常に、狭く、深くなりつつある現在では、機械翻訳の助けなしでは、不可能に近いと思います。そして、新製品を導入、発売する時、一時的に、短期間で膨大な資料を翻訳しマニュアルを作成するには、機械翻訳が不可欠になると思います。

機械翻訳導入と運用

機械翻訳は20年程前に米国へ行った時、20年後には機械翻訳（翻訳電話）が普及しているとの記事を見たことがあります。まさか自分がこれを使用することになるとは思いませんでした。平成元年、翻訳作業の効率化を目的に、シャープのDUETを導入し、約6か月、独りで、徹底的に、前処理技法の開発、ユーザー辞書の構築、最適運用システムの検討を行いました。この期間使用した試験用のテキストは米国大使館広報、文化交流局報道部から配布されるオッフィシャル　テキストであり、その成果として、日本初の機械翻訳[ブッシュ大統領決断のスピーチ]電脳翻訳研究会訳を出版することができました。これを機縁に、同好の人達が集まり、年4回の勉強会を継続しています。

6か月間の試行錯誤の結果、DUETにPC98とAXパソコン2台を接続し、PC98上でもDuetを動かせるようにして、外部よりパソコン通信で受け取った原稿をDUETに転送して、3名の専任要員が同時にDUETとAX端末2 台を使用して、同一翻訳作業を分散並列処理することにしました。初仕事はコンピュータケミストリーの輸入ソフトのマニュアルの翻訳で、OCR で読み込み、DUETで翻訳、ワープロで仕上げを行いました。最初の挫折はOCR の読み取りと図形処理です。そこで、在宅勤務者のワープロ入力ネットワークを組織し、FAXで原稿を送付し、パソコン通信で返信して貰い、それを機械翻訳にかける工程にしました。

英日機械翻訳導入後、3年目に日英機械翻訳と情報処理通信関係の翻訳を目的とし、東芝のASTRANSACを導入し、イーサネットを利用して、IBM 互換機やＰＣ９８とＬＡＮを組み、大規模に構築した１００万語に近い辞書を共有して、本格的に、社外に対しても、機械翻訳サービスの提供を開始しました。

さらに、昨年からはシャープの日英システムをスタンドアロンで2台導入し、効率的な英文作成方法を検討しております。

人員体制は私を含め男子2名、女性3名、それに在宅勤務者3名です。

お客さんのニーズは種々ありますが、大量の文書を短期間に、安く翻訳して欲しいというのが最も多く中身については問題があるかもしれませんが、取り敢えず、専門用語を正確に翻訳することを前提にお引き受けしております。それでも、表現の巧拙は別にして、内容を正確に伝達するには後処理に、かなりの時間をとられます。

また、原文をテキストファイルというかマシンリーダブルにするにはかなりな苦労がいるわけです。出来るだけフロッピディスクやパソコン通信で送付して貰い、それを機械で翻訳して、フロッピーで納品することを前提に、短期間の大量翻訳のニーズに答えることができるようになりました。

大量機械翻訳の手順、問題点

大量の技術文書を短期間に翻訳するには、まず、昼間、文章の単文化や前処理、未知語の辞書登録を徹底的に行い、夜間、10時間から20時間連続機械翻訳をします。その結果、訳語の訂正や、意味不明な部分を洗いだし、さらに前処理を行い、再度夜間バッチ処理を行います。この工程を繰り返えすことにより、後編集やリライトにあまり労力をかけないで内容を正確に理解できる程度の訳文を得ることが出来ます。

機械翻訳の利点の一つは、これら前処理や訳語の訂正、表現の書き替えの殆どを全文もれなく、瞬時に一括して行うことができる点です。問題点は、大量の文書になると、原文のテキストファイルも数人の人間が入力しているため、人によっては半角文字を使ったり、表現や表記法が統一されていない場合が多く、修正に時間がをとられたり、機械が読み取れない部分があり、再度、入力をしなければならないこと、また、ＤＯＳのテキストファイルとして頂いたフロッピー自身がどうしても読み取れずに、そのフォーマットを突きとめるのに苦労することが往々にしてあります。

また、このような大量機械翻訳の場合は、技術内容を完全に理解し、自然な日本語や英語で表現するのは不可能です。あくまで、依頼先での目的に沿った専門家によるリライト、再編集が前提であり、如何にその人達の負担を軽くできるかが勝負である翻

訳支援サービスなのです。機械翻訳とはこのようなものだとお客さん自身がその意義を理解して、機械翻訳の成果を積極的に利用して、自分たちで好きなようにマニュアルを作ることが、当たり前になり、フロッピーでの依頼、納品が一般化すれば、機械翻訳のマーケットは拡大すると思います。

さらに、現在の機械翻訳システムは、自然言語処理だけを対象にしており、情報の収集、理解、伝達とか、マニュアル、論文、報告書作製など、翻訳作業の最終目標を考慮していないので、周辺機器との接続や他の情報処理ソフトとのインターフェースに問題があります。企業に英語屋は要らないと言われるのは、手段を目的と混同するからであり、翻訳も企業の情報活動の手段であり、一部である以上、通信機能を充実するとか、ＧＵＩを採用するとか、徹底した使い易いバイリンガルな翻訳機能付きワープロを完成するとか、改良発展の余地が、現在の翻訳精度でも、十分存在すると思います。

また、利用者側でも、人間でも機械でも、完全無欠の等価訳などこの世には存在しない。表現は翻訳できないし、する必要もない。大切なことは、表現の背後にある意味内容、情報、すなわち、データや事実と事実に対する見解を正確に理解し、伝達することであり、文書を完成した作品と見なさず、明確さこそ、文の命であり、明確さは、直接的な視野と簡潔で正確な表現から生まれるという立場に立つ必要があると思います。

機械翻訳用の辞書

機械翻訳システムを有効に活用するには、ユーザによるカスタマイゼイションが必要であり、その中心は、ユーザ辞書の構築とその選択組み合わせにあります。あまり大きな辞書を作成しても、翻訳精度をかえって落とす場合があります。あまり小さな辞書では、翻訳できません。基本辞書、専門辞書、ユーザ辞書の最適な組合せ方があるはずです。

また、どのような組み合わせ方をしても、かならず、未知語があり、訳語の定着していない語句や、合成語、省略語、頭字語に遭遇します。

一方、どのように専門用語が多く、難解な技術論文でも、使用語数の総計はそれほど多くはないはずです。どのような博識な専門家でも、一テキストで何十万語も、使用できるはずがありません。圧倒的に基本語辞書に登録されている日常語が多いはずです。最初に入手できる該当分野の専門辞書やユーザ辞書でスクリーニングすれば、多くても、未登録語は三桁の単位だと思います。そのような観点から、当社では、分野別に、一対一の対訳辞書を１００万件近く、あらかじめ入力して、毎日、翻訳の度に、さらに追加しており、辞書が成長し続けています。しかし、医学分野などでは、数十万語を入れても当たる時はすごく当たるけれども、当たらない時には２～3 割しか当たらないこともあり、さらに効率の良いユーザ辞書専門辞書の構築の仕方、組み合わせの仕方を再検討し、系統的に使用できるようにしたいと考えています。基本的には、機械翻訳用の専門ユーザー辞書はシソーラスではなく、コパースでなければならないと思います。専門辞書の領域には、訳語の定着した専門用語がすべて登録されており、その分野の概論とか通論とかいわれる標準テキストは完全に翻訳でき、さらに、その上に、細分化された特殊分野の狭い研究分野ごとに、ユーザー辞書をテキストごとに登録して、マージして行くのが良いと思います。

（ 質疑応答）

Q: 辞書の活用について

A: 医学は別にして、普通の技術文献ではどんなに立派な先生でも何十万語の言葉を使い分けることは有り得ず、３万語程度の専門辞書と２－３千語のユーザー辞書を使用すれば　追加すべき単語は数百の単位にではないかと思います。

Q: 専門用語数について

A: 専門用語はあまり細分化せず、１分野2～3万語程度あれば実用に耐えるのではないかと思います。ユーザ辞書も１テキスト数千語程度を整備して行けばいいのではないかと思います。

Q: 年数回の電脳研究会の内容は？

A: コンピュータ、言葉に興味があること、それから人工知能と人間の知性がこれからどのように係わりあったら良いのかと云う事など好きな事を発表して貰っております。

トピックス

日本への留学生数……総務庁行政監察局編による「留学生10万人をめざして」という資料が昨年末に発売になった。この資料によると平成４年には日本への留学生は４万８千人に達しており、昭和58年の4.7倍に達しているという。これら留学生の出身地域別ではアジアが93.8%でもっとも多く、北米は2.8%、欧州は2%という比率になっている。アジアのなかでも中国は42.1%、韓国23,9%、台湾12.6%となっている。これらの留学生に対するアンケートの結果、その専攻別では自然科学が41.7%、社会科学27.6%、人文科学18%その他となっている。地域別の専攻の比率は欧州が人文科学系が多いのに対し、自然科学は全地域とも比率が高くなっている。

来日前の日本語学習の状況……70 %の留学生が来日前に日本語学習の経験があると答えているが、アフリカは来日後に日本語を学習したとのことでありまた中近東でも来日前の学習経験は40 %である。

現在の日本語の会話能力は専門的な議論に参加できるというものが全体の42.5%、日常生活に必要な会話ができるが46%、自分の意志は何とかつたえられるが9,5%、ほとんど話せないは1.9%に留まっている。

また書く能力については日本語でレポートが書けるが63.3%、ひらがなだけでしか書けないが29.5%、ほとんど文章が書けないが4%である。これらのことから日本語教育の改善、充実を望むものが24.2%と多くまた留学生のための特別講義の増加を希望するものが26.7%の多きに達している。

日本人観……日本人はよく働く、勤勉と答えたものは17.8%、外国人に対して親切が16.9%あるのに対して外国人に心を開かないが11.4%、欧米人とは付き合うが東洋人を軽視するが4.4%、外国人に対して偏見があるが2.9% 、人情が薄い1.4%という厳しい印象を漏らしている。

政府は「2000年の留学生10万人をめざして」一層の基盤整備と各種支援施策の充実をはかって行く方針との事である。

産業翻訳市場規模……２月４日日本翻訳協会主催の「翻訳フェア'94」が開催された。この中で「産業翻訳のニーズと品質管理」と題してレクチャーがあった。93年当時の翻訳業の売上は1,300億、印刷業等の非翻訳業は1,700億、企業内独自翻訳は 1,000億、合計4,000億規模と推定される 最近の円高や輸出産業の翻訳ニーズの低迷により30 %近い売上ダウンになっている。反面輸入産業のニーズ増加は見られるもの落ち込みをカバー出来る程の売上を維持できない状況にあるとの厳しい報告があった。

アジアのイベント（問い合わせ先……国際情報化協力センター・普及部☎03-3457-0941）

「Infocomm China '94」
- 月　日　1994年4月6日～9日
- 場　所　中国 Shenzhen International Exhibition Center
- 主　催　Shenzhen International Exhibition Center ,PANASIA Convention & Exhibition Limited.

「COMPUTEC '94」
- 月　日　1994年5月12日～16日
- 場　所　中国 上海市上海商城中心(SHANGHAI CENTER)
- 主　催　China Instrument Society China Computer Industry Association

「China Conputer World EXPO」
- 月　日　1994年9月28日～10月1日
- 場　所　China International Exhibition Center Beijing PRC
- 主　催　IDC Computer Company (Hongkong)

「ASIAN IT EXPO '94」
- 月　日　1994年9月28日～10月1日
- 場　所　香港 HongKong Convention Center (Hongkong)
- 主　催　ADSALE Exhibition Services Ltd.(Hongkong)

機械翻訳の今後の技術動向

(Future Trend of Machine Translation Technology)

通産省電子総合研究所 主任研究官 井佐原 均

今回は、機械翻訳システムの新しい技術動向について紹介します。

現在市販されている機械翻訳システムは、（当たり前だといわれるかも知れませんが）文法規則と辞書を使って、入力文を解析し、翻訳出力を生成します。これまで、このシリーズで行なってきた機械翻訳技術の解説も、このようなやり方に沿ったものでした。このようなやり方を、規則（辞書の記述も規則の一つです）を用いる手法という意味で、Rule-based の手法と呼びます。

このような手法で作られた機械翻訳システムの性能を向上するためには、これまでこのシリーズで説明されてきた構文解析や意味解析といった要素技術をさらに深く研究していくことが必要です。

また、実際にシステムが使う文法規則や辞書に書く情報を出来るだけ詳細なものにいくことも必要となります。

これまでは、翻訳システムが用いるさまざまな規則は、人間が自分の言語的直観に基づいて作成してきたわけです。つまり、自分が言葉を理解したり、話したりする時に使う知識を、規則の形に書き下ろしていたのです。しかし、規則が増えてくるにつれて、取り扱う情報が非常に大きなものとなり、全体としての見通しが大変悪くなります。そのため、このような人手に頼る手法だけでは、処理しきれなくなります。規則がどんどん増えていくと、一人の人間が全体を理解することが難しくなり、その結果、「どうしてこの訳がでてきたのか？」という素朴な質問にシステム開発に直接関わった人ですら答えられないということになります。直感に基づいて規則を作成している以上、どの程度の量の情報を集めれば、翻訳には十分なのかということに対する理論的なきちんとした答えはないのです。

また、規則は単に集めれば良いというものではなく、その規則をその時々に応じて適切に使うことが必要となります。つまり、機械翻訳に必要とされる知識には、規則（言語情報）そのものと、ある時点で適用できる規則が複数あった場合に、それらのうちのどれを使えば良いかを決める適用性の情報（優先順位付け）とがあることになります。これらを区別して、それぞれを適切に表現できる枠組が確立していないことも、見通しを悪くする原因となっています。

現実のさまざまな文書を翻訳することの困難さはこのような問題点を解決し、きちんと使えるようになった規則化された知識が不足していることから来ているわけです。しかしながら、実際に世の中に存在する膨大な文書を対象に、その翻訳に必要とされる知識を人手で抽出し整理することは、ほとんど不可能でしょう。「それではどうすれば良いのだろうか？」この問いに対する答として、計算機に大規模なコーパスを直接処理させようという考え方が提案されました。

この手法の提案は、１０年前に遡る［１、２］のですが、当時は大規模なコーパスを解析するのに十分な能力を持った計算機は大変高価なものでした。最近の計算機の高速化・低価格化により、大量のデータを比較的容易に処理することが出来るようになってきたため、事例を利用した（あるいは統計的手法を用いた）自然言語処理と呼ばれるこの手法が脚光を浴び始めたのです。

この手法の機械翻訳への応用としては、まず、事例を用いて翻訳を行なうシステム（Example-based Machine Translation）があげられます。

例を使って説明しましょう。play という英語の単語を日本語に翻訳する場合には、play の目的語に応じて、「（スポーツを）する」「（楽器を）演奏する」など、適切な訳語を選ばなくてはなりません。これを従来からの規則を用いた手法で翻訳する場合は、

"play スポーツ" --> "スポーツをする"
"play 楽器" --> "楽器を演奏する"

といった訳し分け規則と

piano　（楽器）　ピアノ
violin　（楽器）バイオリン
tennis　（スポーツ）テニス
golf　　（スポーツ）ゴルフ

といった、辞書（訳語と意味素性）とを用意することになります。この程度の情報の場合は、どちらも非常に簡単で分かりやすいのですが、一般の雑多な文章をきちんと翻訳するためには、もっと多くの情報を記述することが必要になり、規則を表現する枠組も、非常に複雑なものになります。

一方、事例を用いた翻訳システムの場合には、対訳とシソーラスを準備します。対訳として、

I play tennis.　　私はテニスをする。
I play the piano.　私はピアノを演奏する。

シソーラスとして、

楽器
　ピアノ
　バイオリン
　トランペット
スポーツ
　球技
　　テニス
　　ゴルフ
　　卓球

というものが与えられていたとしましょう。

ここで、"I play golf." という文が入力された場合、ゴルフが、テニスとピアノのどちらに、より似ているかをシソーラスを使って調べます。この場合には、明らかにテニスの方に似ていますから、システムは、"I play golf." は "I play the piano." よりも、"I play tennis." の方に似ていると判断し、「テニス」を「ゴルフ」に置き換えて、「私はゴルフをする。」という訳文を生成します。

このようなシステムにおいては、対訳コーパスを検索し入力文に最もよく似た例を選び出す処理が一番重要な部分です。各単語の「似ている度合」を計算するためには、シソーラスを使う場合が多く、シソーラスの木構造の中での単語間の距離を使って計算します。

ただ、この例のように、単語が一つだけ異なっているような文（対訳例）がコーパスの中に含まれている場合は、現実の文章の翻訳では稀でしょうからほとんどの場合、適当な対訳を組み合わせて、一つの文を作ることが必要になります。このためには、入力と例との間で一致していない部分について、再帰的にコーパスの検索を繰り返す手法［3］等が提案されています。

事例を用いた手法の機械翻訳への応用としては、このようにコーパス中の対訳を使って直接翻訳をするものばかりではなく、事例から知識（あるいは統計的確率）を自動的に獲得しようとする試みも行なわれています。コーパスを使って、翻訳に必要な規則を作り出すわけです。この場合、原文を何らかの形で解析する必要があるわけですが、当然、完璧な解析は出来ません。誤った解析結果が含まれていたり、解析自体に失敗したりすることもあります。しかし、十分に大きなコーパスを用いていれば、全体としては、間違った解析結果よりも、正しい解析結果がたくさん含まれることになり、意味のある情報を取り出すことが出来ます。たとえば、簡単な構文解析によって、文中の動詞に係る名詞を見つけ、動詞と名詞の共起の頻度を調べることにより、動詞の持つ格フレームを自動的に獲得する研究［4］などが行なわれています。

また、自動翻訳を目的とはせずに、翻訳支援システムとしての使用を目的とするものもあります。翻訳したい文に類似した用例を検索し表示することにより、翻訳支援を行なうのです。ユーザが翻訳したい文を入力すると、システムはそれと類似した文を対訳コーパスから選びだし、その対訳と合わせて表示します。ユーザは、それを参考にして、自分が翻訳したい文の訳文を作成するのです。もちろん、類似した文をどうやって選び出すかが問題となるわけですが、あらかじめ形態素解析されたコーパスを用いるものや、前処理をされていない対訳コーパスを対象に、文字列の類似度を効率的に計算する手法［5］などが提案されています。

事例を利用したシステムの特徴として、ユーザがコーパスに対訳を追加するだけでシステムの能力を向上することができるということがあげられます。

システムの内部構造や自然言語処理に詳しくない人でも、対象とする言語を知っているならば、正しい対訳を追加していくことによって、自分の言語知識に沿って、システムの改良を行なうことが出来るのです。もし、あなたが、play という単語を翻訳する時に、「演じる」と訳した方が良い場合があると知っているならば、上で示した例に、

I play Romeo.　　ロミオの役を演じる。

という対訳を付け加えれば良いのです。

一方、規則に基づく手法では、システムを改良するために規則を追加する場合には、まず規則の記法を理解していなくてはなりませんし、規則を一つ追加したことによってシステム全体が崩壊してしまったりしないように、注意しなくてはなりません。例えば、"a → b" という文法規則と "b → a" という文法規則を入れたような場合、解析時にループを作ってしまうために、システムが暴走してしまうことがあります。事例を用いる手法では、用いるコーパスが十分に大きければ、一つの対訳を追加することによって、そのような崩壊が起こることはありません。ゴルフは、テニス、ピアノ、ロミオのうちでは、やはりテニスに一番近いでしょうし、"I play the violin." という入力に対しては「私はバイオリンを演奏する。」と翻訳してくれるでしょう。

勿論これは、システムの性能が簡単に向上するということでは、必ずしもありません。コーパスを大きくしていくことはそれ自体、大変な労力が必要なものです。しかし、このようなコーパスは、個々のシステムとは独立したものであり、一旦作られたコーパスは他のシステムでも用いることが出来ます。あるいは、既に何らかの形で存在する対訳集を使うことも出来るでしょう。

このような、事例を用いる手法は今後一層の発展が期待される分野ですが、それと同時に、既に規則化されている情報が使える場合には、それを利用した方が効率的な場合も当然あるわけです。したがって、今後は、規則を用いた手法と事例を用いた手法とのそれぞれの特長を活かして、両者を融合することが検討されていくでしょう。そのようなシステムでは、言語処理の各時点で利用できる情報のうちで、最も特定化されている情報を用いるという方針で処理が進められることになります。このような手法は人間の言語処理過程と類似しており、高速な自然言語処理が可能となります。

機械翻訳においては、たとえば、完全に一致する例が見つかれば、それを用いて翻訳を行なう。なければ、類似した用例を利用する。それも困難な場合には、規則に基づいて解析・生成を行なう、というようになるでしょう。先ほどの例でいうと、"I play tennis." という文が入力された場合には、完全に一致する文がありますので、そのまま対訳中の訳文が出力されます。挨拶文の翻訳などで、このようなことがよく起こると思われます。また、"I play golf." という文が入力された場合には、すでに述べたように、事例 "I play tennis." を使った翻訳が行なわれます。さて、もし、コーパスに疑問文が含まれていなかった場合に、"Do you hate sheep?" という文が入力された場合には、（少なくとも疑問文固有の部分については）文法規則を用いて解析し、訳文を生成します。もちろん、疑問文を解析する規則をシステムが持っていなくてはなりませんが。

マニュアル等の翻訳では、一旦翻訳した文書の修正版を翻訳することが、しばしば起こります。そのような場合には、以前に翻訳したものをできるだけ使いたいと考えるでしょう。この手法は、そのような状況に対応できることにもなります。

このように考えてきますと、従来からの規則を用いる手法も、やはり重要であるといえます。これまで自然言語処理の研究においては、自然言語の持つさまざまな曖昧性を計算機が正しく理解し、それを解消するためには、意味情報あるいは文脈情報が必要であるということが強調されてきました。そのために、表層の言語現象を詳細に検討するという立場が軽視されていた点があります。その結果、簡単なシソーラス程度の知識と構文解析によって構文構造をかなり正確に把握できるような場合があるにも関わらず、それを意味に関わる困難な問題として放置しておいた面も見られます。詳細な、そして当然膨大な意味情報を用いる意味解析は、もちろん重要な技術ですが、人間が文章を読む場合には、文の内容を完全には理解していなくても、文章の構造（並列句や係り受けなど）をある程度、判断できる場合があることから、人間は言葉の理解過程において、「浅い」意味を有効に用いているものと思われます。このような点に注目した研究の例としては、従来、

深い意味処理の対象とされていた名詞句あるいは連用中止による述語の並列構造を、文節同士の類似性を計算することにより、かなりの精度で決定できることを示したもの［6］などがあげられます。

なお、自然言語処理における困難な課題を避ける手段の一つとして、取り扱う領域を限定しようという考え方もあります。この場合、実際には、領域を限定することによって、取り扱う知識の範囲だけではなく、取り扱う言語表現（あるいは言語能力）をも制限していることになります。特定の場面で発話される文は、その言語で可能なすべての言語表現の中では、ほんの一部となります。そのような制約があることによって、辞書中に表層の情報に対応した知識を書き込んで行くことが可能となります。自然言語処理において、辞書の記述が主要な知識源であることは、どのような手法を用いる場合にもいえることであり、現実の文章から着実に辞書記述を改良して行くことは、最も重要な作業の一つです。

《参考文献》

［1］M. Nagao, Some Rationales and Methodologies for Example-based Approach, Proc. of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing, 1992

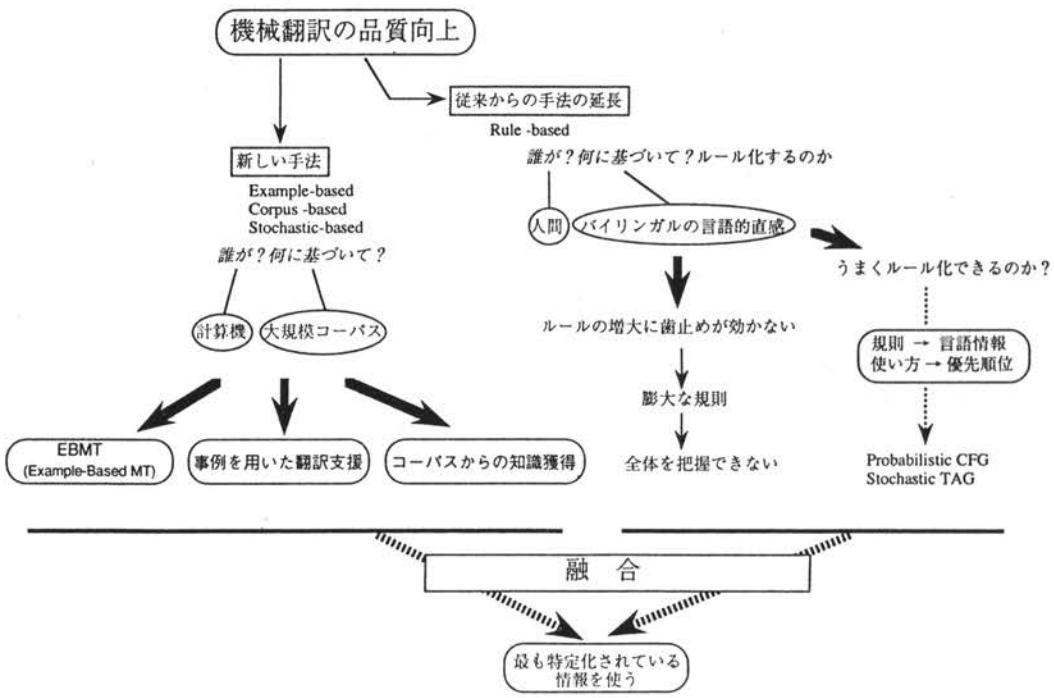
［2］M. Nagao, A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, in A. Elithorn, ed Artificial and Human Intelligence, Elsevier, 1984

［3］佐藤理史、ＭＢＴ２：実例に基づく翻訳における複数翻訳例の組合せ利用、人工知能学会誌、Ｖｏｌ．６、Ｎｏ．６、１９９１

［4］R. Grishman et al., Combining Rationalist and Empiricist Approaches to Machine Translation, Proc. of Fourth International Conference on Theoretical and Methodological issues in Machine Translation (TMI-92), 1992

［5］S. Sato, Example-Based Translation Approach Proc. of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing 1991

［6］M. Nagao et al., Dynamic Programming Method for Analyzing Conjunctive Structures in Japanese, Proc. of COLING'92, 1992



新製品の紹介

英日機械翻訳システム Ｖ2.1 『ＤＵＥＴ　Ｑｔ』

シャープ株式会社

シャープは６年前に業界で初めてＯＣＲ付きの機械翻訳専用システムを発売し、その使い易さや高い翻訳精度で多くのオフィスでお使い戴いております。

また、３年前には省スペース性を追求した業界最小、最軽量の専用システム『ＤＵＥＴ　Ｑｔ』を販売、２年前には、待望の双方向翻訳を実現する日英機械翻訳システムも開発、ＭＴメーカのパイオニアとして躍進し続けています。

そして、この度、当社にお寄せいただいた数々のご意見を参考に、英日機械翻訳システムの最新バージョン(Ｖ２．１)をリリース、従来にも増して高度な翻訳品質と機能の充実をはかりました。

［翻訳精度向上のための改良と辞書の拡充］

1 新構文解析方式（並行型構文解析）の採用
　複数の解析候補から最良の候補を選択する並行型構文解析を採用。
　従来の方式に比べ、より良い翻訳結果を訳出します。

2 優先解釈の枠組みの導入
　英文の形などをチェックする枠組み／規則を導入し、最適な形を優先的に判断します。

3 訳語優先解釈の導入
　基本後辞書に「格パターン優先度」などの情報を追加し、より柔軟な訳語選択が可能になりました。

4 基本辞書の拡充（１）
　主に動詞に対して、分野情報や型情報、意味情報を充実させ、その翻訳を強化しました。

5 基本辞書の拡充（２）
　新構造辞書の採用による訳語選択。
　基本語辞書の構造を大幅に変更し、従来では持っていなかった詳細の情報により、きめの細かい訳語選択を実現しました。

6 基本辞書の拡充（３）
　大量の評価文から抽出した単語や熟語、またその同類語や反対語、派生語にいたるまでを追加しました。また主要な人名も追加しています。
　（見出し語数８９、０００語）

［機能の充実］

1 ユーザ辞書、専門辞書の複数同時使用
　従来１つまでしか使えなかったユーザ辞書、専門用語辞書をそれぞれ２つまで使用して翻訳することが可能です。同じ種類の辞書内では優先順位をつけられます。

2 ユーザ辞書の登録可能語数の拡大。
　最大約８万語まで登録することが可能になりました。

3 訳語表示、参照時の交換辞書の拡大
　同じ単語が選択されている別の辞書に登録されている場合でも一画面上で確認したり、入れ替えることが可能です。

4 訳語反転表示機能の強化
　分割翻訳（句単位の翻訳）となった場合でも、英語と日本語の単語間の対応関係が参照できるようになりました。また、訳語を調べたり、学習させることも可能です。

5 新学習方式の採用
　従来では利用できなかった後置語などの情報も学習結果に自動的に反映されるようになりました。
　　Ｉ　ｍｅｅｔ　ｈｉｍ.

meet「会う」から「迎える」に学習させる
（従来）　私は、彼に迎える。
（V2.1）私は、彼を迎える。

6 範囲指定候補機能

文中の一部分だけの解析を変更させることが可能になりました。他の箇所は固定して正しい訳を得ることができますので、短時間でほしい訳が見つかります。

7 モード設定項目の追加

主語の無い文を命令形で訳すか、終止形で訳すかの設定や、文を一文ごとに切り出す場合の記号を追加して設定する等、合計4項目の設定を追加しました。

8 分野選択設定機能

化学工学、機械工学、経済、時事、情報処理、電子工学、医学の7分野から翻訳する文章の分野にあわせて、3つの分野までを選択することができます。

9 印刷環境設定項目の追加

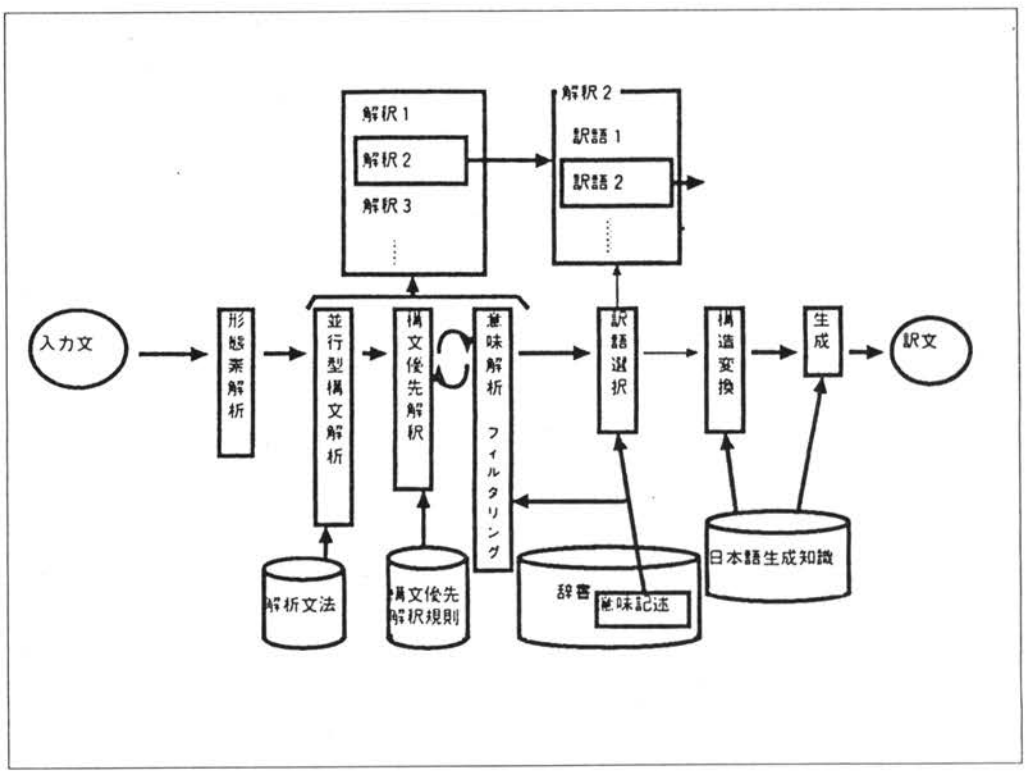
原文／訳文のレイアウト変更やシングル改行かダブル改行かの選択/設定が可能になりました。

10 マウス操作の改良

マウスのダブルクリックで、一文翻訳後簡単に単語訳を画面上に呼び出すことができるよになりました。

販売元：シャープシステムプロダクト（株）
翻訳システム営業部　03-3267-5753
問い合わせ先：営業推進部　043-299-8302
［シャープ㈱　コンピュータ商品企画部 蒲］

英日機械翻訳システム　V2.1 DUET　Qtの仕組み



翻訳現場見学記

機械翻訳の利用に当たっては前処理、後処理は不可欠である。これがため各社とも利用マニュアルを発行したり、実務講習会を実施している。しかしユーザサイドにも個々に開発された利用技能があり、アンケート調査等では把握出来ないものも多い。そこで翻訳現場に埋もれた処理手法を発掘し、これを整理、分類してみることも必要との考えから、利用技術研究会メンバーを中心に翻訳現場見学会を実施した。

①長瀬産業㈱翻訳室

当社の翻訳スタッフは常勤社員５名、在宅勤務者３名の構成で最近は需要量は落ちて来ているが、月間平均200頁の機械翻訳を行っている。翻訳文書は契約書、技術文献など社内の全ゆる文書であるが、やはり量的にはマニュアル類が多い。また社内だけに止まらず、社外からの依頼に応じ翻訳も行っているが、入力原稿はFDでの提供を呼びかけているが、社外の依頼文となると矢張り手書きのものも多く、在宅勤務者が前処理しながらデータ化し、それをパソコン通信で送信して貰い、機械処理している。 現在シャープ、東芝のマシンを使用しているが、それぞれ癖があり、前処理のやり方も異なるという。前処理のやり方は徹底した短文化にある。「句、節単位で処理する方が翻訳精度が高く、誤りの修正もやり易い。
短文化された句や節を後処理でどう繋ぎあわせていくかはやはり語学力が必要である。」と。蓄積された専門辞書は100万語に達しているとの事である。

②㈱アイ・ビー・エス

当社は機械翻訳業であり、約30台のパソコン、MAC、5台の翻訳機と10台の翻訳端末、W/Sがイーサネットで繋がっている。数百頁にも及ぶ長文は数人の作業者が章や節単位で分担する。最初は全員が依頼文の何割かの文章の専門用語を入力し、その後これらの辞書を使って、担当の章や節の前処理(係り受けや短文化)を行い、機械翻訳を行う。これにより用語の統一も出来る。ネイティブ部門では原文のスペルチェックと出来上がった翻訳文の化粧をする意味での後処理を行う。マニュアルの場合、1頁の中で図や写真の占めるスペースが多く、翻訳すべき文章が少ないこともあり、インターリーフとも併用(試用の段階)し、翻訳だけではなく周辺業務のトータル処理も試行中である。

訪問調査で多くの資料の収集が出来たが、研究会ではこれらを整理し報告書に纏める事にしている。

「COLING '94」

（第15回計算言語学国際会議）

開催月日　　平成６年８月５日（金）～９日（火）
開催場所　　京都　都ホテル
開催日程　　招待講演、論文発表（４セッション並行開催）、パネルディスカッション
　　８月５日（金）　　開会式・論文発表・レセプション
　　８月６日（土）～９日（火）論文発表
　　８月８日（月）　　バンケット（夕刻）
　　８月７日（日）　　インフォーマルミーティング

主　催　　第１５回計算言語学国際会議組織委員会
共　催　　ICCL(International Committee on Conputational Linguistics)
協　賛　　電子情報通信学会 情報処理学会 人工知能学会 日本ソフトウエア学会
　　日本情報処理開発協会 日本電子工業振興協会 アジア太平洋機械翻訳協会

協会活動状況報告

予算理事会　3 月23日 ①94年度事業計画案 ②94年度事業予算案③その他案件

運営委員会　1月25日 ①収支状況報告②会員の動向③IAMT理事会報告④研究会活動報告⑤その他
2月22日 ①協賛事業報告②研究会活動報告③94年度重点推進事業検討④その他
3月11日 ①予算理事会上程議案検討②研究会活動報告③共催事業検討④収支見込検討

利用技術研究会　1月21日 ①ユーザ利用実態調査質問項目の検討②辞書共有化方策の検討③翻訳トータル手法の検討④利用実態ヒアリングの実施
2月23日 ①ユーザ利用実態調査方法の検討②機械処理ノウハウの調査手法の検討③調査日程について

システム評価研究会　1月28日 ①情報交換「JEIDAのMT調査研究の現状について」②「翻訳困難例文の摘出とその言い換え例について」の作業方法と作業日程について③会名変更について
3月9日 ①翻訳困難例文の機械処理結果について②機械処理後の言い換え例の検討

需要予測検討会　1月27日 ①収集調査資料の概要説明②需要予測モデルについて③ＭＴ新市場の可能性についての検討
3月1日 ①収集資料の概要説明②需要構造の変化について③MT需要予測手法について④需要予測モデル④ＭＴ需要予測基準年の潜在需要の見直しについて⑤パソコンＭＴ市場の動向について

翻訳現場見学会　3月23日 訪問先「長瀬産業㈱」……利用技術研究会メンバー
①翻訳の処理実態について②翻訳体制について③翻訳需要の現状について④機械翻訳に関する情報交換
3月24日 訪問先「㈱ＩＢＳ」
①翻訳作業の処理実態について②文書のトータル処理について③翻訳作業体制についての情報交換

協賛事業
「94翻訳フェア」2月4日（主催日本翻訳協会）於:東京YMCAホテル
①セミナー(技能向上、品質向上)②講演会③パネル討論④機械翻訳デモ

新入会員の紹介

和田　肇
Leung Siu Wai
岡崎洋三

AAMTジャーナル
NO. 6（MAR／94）

発　　行　アジア太平洋機械翻訳協会（略称：ＡＡＭＴ）
所 在 地　〒103東京都中央区日本橋小舟町4-10 大宮ビル
☎ 03-3664-5637/5638　FAX 03-3664-1352
編集委員会　野村浩郷(委員長) 亀井真一郎 信田恵壱 杉山健司
事 務 局　星野禎男 西郷容子