



AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.10

March 1995

アジア太平洋機械翻訳協会

機械翻訳は使える

沖電気工業 専務取締役 山本正隆

NII とか GII という情報ハイウェイが、マルチメディアと共に、最近の文明評論におけるキーワードになっています。GII はグローバルなネットワークの話のようですが、グローバルなコミュニケーションに翻訳は欠かせません。翻訳の仕事に何らかの係りを持って業としている私共にとって、今年が良い年だと思っておりました。

しかし神戸への初詣も済ませないうちに、突然神戸で大地震が起り、追風ムードに水を差すことになってしまいました。震災に遭われた多くの方々には心からお見舞い申し上げます。このとき、私共の会社でも神戸に支店があり、また多くのお客様もいらっしゃる、更に神戸地区に本人や家族の住んでいる従業員も大勢いたりして、連絡がつかずに不安だったものです。

また神戸は日本有数の港を持っており、いわゆる国際都市であります。被災された外国人の方々も多かろうと思われそうですが、特にあった非常時には言葉の問題でお困りになったのではないかと拝察するわけがあります。ここでまた、コミュニケーションとか翻訳

の問題を意識せざるを得ないことになってしまいました。

長尾先生が指導された機械翻訳の国家プロジェクトに私共の会社も参画させて頂いてから、かれこれ十年が経ちました。その間に人工知能(AI)ブームが起り、その先兵ということで機械翻訳も脚光を浴びました。その後AIも落ち着きを取戻し、同様に機械翻訳も、本協会の活動にみられる通り、着実な歩みをみせていることは御同慶のいたりです。

私も機械翻訳の周辺に係わるようになり、言葉の問題に興味を持ったのは幸でした。人間の知的活動、とりわけ知識と、それに基づく認識や判断即ち理性と呼ばれている活動のメディアが言葉であることがよく理解できました。『始めに言葉ありき』などが、分かったような気になれたわけです。

機械翻訳を巷で論じるとき、とても人間にはかなわないとか、文脈処理をしないと意味の翻訳はできないなど、ネガティブな発言がまだよくあります。しかしそれは、機械がどこまで人間を支援できるかという問題と、人間と人間のいわゆるコミュニケーションの

MT-SUMMIT V

開催月日 平成7年7月11日～13日

開催場所 EC本部へミサイクル(ルクセンブルグ)

テ	マ	*MTの経済的側面	*専門分野用語と電子化辞書
		*ビジネスインフラとMTの融和	*オンラインMTサービス
		*SME対応のPC版MT	*ユーザ側からみたシステム評価基準

参加料 250 ECU's .

主催 IAMT 後援 EC XIII部会 協賛 EAMT, AAMT, AMTA

連絡先 SEMA Groupe Benelux (Place du Champ de Mars 5, Bte 40, B-1050, Bruxelles)

Tel. 32-2-512-5317 / 32-2-508-5401 FAX. 32-2-512-1076

基本問題との区別ができない人達の発言ではないでしょうか。

力学の世界では、物体の位置と運動量は同時に決定することはできないといういわゆる不確定性原理があるそうです。同様に言語の世界にも、言葉（記号）と意味は同時に決定できないという原理があるように思われます。ある言葉の意味は長い時間をかけないと分らない。単語はそれが発せられた瞬間は意味が全く不確定ということです。文脈と呼んでいるのは、この不確定性原理に含まれていると思われます。

私共も一応グローバルなビジネスをしていますので、言葉の違う方々とお話をすることがあります。いわゆる外大出身の営業マンなどが通訳してくれる訳ですが、仕事のことはともかく、双方の趣味の話になったりすると、とたんにもどかしくなってきます。専門語はもとより、平易な言葉でも専門的な意味に使う言葉が通訳に通じなくなるのです。同好の士だったりすると、片言の方がよく通じたりするわけです。これは多次元空間論の変換の問題なのかも知れません。

不確定性原理の問題も、多次元空間の問題も、極めて高度な物理学の問題です。日常の私共の生活で自動車を走らせたり、飛行機を飛ばしたりしている世界では全く問題になり

ません。言葉と意味の問題も、全く同様です。確かに物理の問題より身近に遭遇する問題ではありますが、その分だけ人間は互にそのことを十分承知のうえでコミュニケーションしているのです。

すなわち、場合によって誤解もあり得ることが暗黙の了解になっているのです。

文明評論家を気取って難しそうなことを述べてきましたが、言いたかったのは、現在の機械翻訳のレベルは十分使いものになるということです。文学を訳そうという訳ではありません。文学などもともと原文でも難解なのです。ノーベル賞作家大江健三郎が良い例です。機械翻訳の問題ではありません。機械は確かに人間とは異なった癖を持っているかも知れません。しかしそれは賢明な人間はすぐ見抜いてしまいます。

今ビジネスの世界では電子メールが流行っています。これはキーボードを使いますので、慣れない私などどうしてもメッセージが電報調になります。機械翻訳にはもっとも馴染易い文であります。

案ずるより生むが易しいという諺があります。機械翻訳は今やその時代ではないでしょうか。使えば当然進化します。やはり今年は機械翻訳にとって楽しみな年のような気がします。

目次	機械翻訳にも価格破壊の波が… (翻訳システム、翻訳ソフトの動向)	(2)
	ヒヤリング・PC版ソフトで技術文献を翻訳して	(8)
	・機械翻訳を用いた翻訳作業の効率化の可能性	(17)
	・ホワイトカラーの知的生産性を考える	(24)
	・MT, HTの生産性比較	(20)
	マネージメント・トランスレータの選択	(12)
	海外レポート・シンガポールの機械翻訳研究	(36)
	・チベット語情報処理技術の現状と展望	(40)
	研究報告・EDR辞書の研究成果と今後の展開について	(44)
	技術早分かり・OCRによる文書テキストの読み取り	(32)
	研究機関紹介……NHK放送技術研究所	(30)
	新製品紹介…PC-Transerシリーズ(52) TransLand for Windows(54) こりゃ英和(56)	
	電辞海 for Windows(58) WD-01SW (60) Pensee for windows(62) j-London E/J(64)	

価格破壊は機械翻訳にも

MTのニュートレンド

はじめに

価格破壊とは単なる値引きやダンピングを意味するものではない。価格を構成する要素や生産、流通を含めたプロセスにおいて徹底した合理化をはかり、消費者ニーズにマッチした新たな価格体系を作りて行くものである。かつてバブル時代にはブランド指向に走り、また商品の2次機能や独自性、手作りなど希少価値のみを強調した高価格品が好まれた時期もあったが、バブル崩壊後購買力が低下してくるとともに市場ニーズも変化し、従来の価格体系を打ち破る新しい商品群が誕生して来ている。

米国で市販されている商品、これは輸入品、国産品を問わず日本よりはずっと安いと云われている。日本における流通経路やコスト、各種規制、円高、高い地価や物価高など国内価格におよぼす影響も少なくないが、これでは消費者離れることになるが、昨年後半期から機械翻訳にも価格破壊の波が押し寄せてきている。そこで最近の各社のカタログからその内容を調査してみた。各社のカタログには訴求点や表現力にかなりの差はあるが、あくまでも標記された内容に従ってその動向を探ってみた。

□種類と形態

まず名称についてであるが、機械翻訳システム、翻訳支援システム、英文作成システム、翻訳ワークベンチ、翻訳ソフト、翻訳ツールなど商品名もまちまちで、素人にはじつに分かりにくい呼称である。

その形態としてはホストコンピュータに接続して使用する翻訳支援システム、ワークステーションに翻訳ソフトを組み込みOCRやプリンタとシステム化して使用する翻訳システムや翻訳支援システム、パソコンハードに適合させた各種PC翻訳ソフト、更に簡易型の翻訳ツールなどの形態がある。

また日本で市販されているものは、その殆どが日英、英日翻訳であり、それも片方向のものや双方向翻訳のもの、さらには片方向のものを連結する双方向オプションもある。

ホストコンピュータタイプは市販されていないプロトタイプのものも稼働はしているが、市販機種は少ないのでワークステーションタイプとPCタイプについて考えてみよう。

□W/S翻訳システム

ワークステーションタイプの機械翻訳システムが商品化されてから既に11年を経過する。当初は文法規則や技術主導で商品化されただけに年々変化する言語現象にも合致せず、前処理や前編集技術も試行錯誤を繰り返し、人手翻訳に比べ翻訳品質も悪く、時間もかかるとのことで次第にホコリをかぶる存在となって行った。この原因には原文側にも一部問題があり、例えば日本語のもつ曖昧性、文法ルールにない主語や目的語の省略、冗長な言い回しなど正確無比な処理を行うコンピュータに理解できない問題も多々あった。これがため原文側に対し、機械にも十分理解できるテクニカルライティングや制限言語の導入などを要請しようとする動きもあったが、原文作成者側からすれば、機械のために余計な苦勞をしなければならないのかという反発もあった。

しかしメーカー研究、技術陣が翻訳現場の人々と一体となって改良に改良を積み重ね、中には年2回以上もバージョンアップが繰り返されるなどして現在では品質面で著しい改善がみられている。

また操作性についてもユーザインタフェースの改善で、メニューをみながら手軽に操作できるようになり、かつての「難しいもの」との印象は拭い去られ、素人でも1~2日練習すれば十分に使いこなせるという所までになった。

機能的にどのような改善がなされたかについては各社各様の対応があるが、翻訳速度は6万語/時間、帰社前に機械に原文をセットしておけば翌朝出社時には単行本の2~3冊文は翻訳できてしまう高速翻訳可能なもの、辞書の選択や訳語表現方法など翻訳環境を画面の表示に従って設定できるユーザカスタマイズ機能などを標準とし、長文分割、合成名詞、未登録後検出、冗長表現、接続表現、否定表現、否

定表現、受身表現、連帯中止表現、不適表現辞書の自動検出など、また無翻訳指定、フレーズ指定、品詞指定、文型指定などの前編集支援機能を有するものや訳語入れ替え、次候補呼び出し、日本語文の訂正、校正ができる後編集支援機能を備えたものもある。

これらの機能を駆使すれば人手翻訳で過ちを犯しやすい用字誤り(誤字、脱字、送りかな、)訳し漏れ、転記ミス、書式設定ミス、あるいは文体混合などは自動的に処理できる。

最近1年間ではある程度改良も定着し、新製品の発売頻度は落ち着いて来ている。またユーザサイドでも専任担当者の設置、前処理や翻訳ノウハウの蓄積、ユーザ辞書や専門辞書の構築も進展し、機械翻訳の実力の理解も深まり。また周辺技術の開発も進められ、十分に使いこなすまでに至っている。

しかしW/Sは価格的にはなお高価である。ハードの生産実績を調査すると平成6年1~10月では144千台、これは前年同期比で92%の落ち込み出荷価格は単純平均で167万円であり、前年同期の172万に比べ僅かながら下落してはいるもののOCRやプリンタなども取り揃えたとすればやはり慎重にならざるを得ない。

また企業内翻訳といえども出版社でもないのではそれほど毎日大量の翻訳量があるわけでもないし、また前、後編集にも結構時間がかかって、マシンがフル稼働するような状況でもない。最も機械の利用効率を上げるために、未知語検索用に、下訳用にと翻訳のプロセスに従って、何回も機械に原文や下訳を通すという使い方もされる。

□オフィスワープロ装填翻訳支援ソフト

日本語ワープロに英和辞典を内蔵した商品は過去に発売されていたが、これは英文のそれぞれの単語に語訳を訳出表示するというものであったが、近くオフィスワープロに本格的な翻訳支援ソフトを装填し、翻訳システムとして利用出来るものが発売されとの発表もある。

□パソコンハードの動き

パソコン翻訳ソフトは87年頃から開発、市販されていたが、当時はメーカーも少なく、ワークステーションレベルの蔭にかくれ、またワークステーションシ

ステムでも翻訳精度には批判が多く、まして機能の落ちるパソコンソフトは脚光をあびることが少なかった。

パソコンソフトが本格的に市場で注目されるようになったのは90年以降の事である。そしてその様相は94年以降大きく変化した。

まずパソコンハードについての動向であるが、バブル崩壊により92年、93年の生産、出荷実績は低迷したが、94年に入って需要は大きく上昇した。日本電子工業振興協会の発表によれば、94年第3四半期の総生産台数は112万2千台となり、前年同期比43%も増加している。この数値は自主統計に参加している23社の数値で、カバー率は95%といわれているから実質118万台の出荷とみられる。暦年(1~12月)での出荷台数は国内407万8千台で、この内国内出荷は300万5千台、輸出は107万3千台である。国内出荷の内ラップトップが104万1千台、その98%がノートブックタイプである。

またbit数では32bitが99.7%を占めており、16bitクラスは3千台の0.3%に過ぎず、高能力化している。

また国内出荷価格は10~20万クラスが6.4%、20~50万クラスが73%、50~100万クラスが20%を占めている。

一昨年の市場価格ではほぼ80~100万であったことから比較すれば大幅な価格ダウンになっている。さらに米国、台湾などからの輸入品も増加しており、価格競争は益々激化の様相を呈しつつある。それだけに消費者にとっては手頃で買い易くなったということで、従来のオフィスユースもさることながら、今後は一般家庭や教育市場にも多く普及して行くものと思われる。

このような価格ダウンの背景には半導体の高密度化が一層進行し、生産も好調であり、パソコンの機能もまた著しく向上し、需要を刺激していることに外ならない。因にB5サイズサブノートパソコンの標準価格は40万円前後となっている。

このようなハードの動向に対応してパソコン翻訳ソフトも大きな変化を示している。

□パソコン翻訳ソフトの動向

パソコン翻訳ソフトの好調な売れ行きに刺激されて、高電社やブラザーなど翻訳ソフト開発企業への新規参入が目立つことである。また従来翻訳システ

ムメーカーだった富士通などの企業もパソコンソフト部門に進出して来ていることがあげられる。

また価格的にはそれまで英日あるいは日英翻訳ソフトは198,000円程度で推移して来たものが、昨年夏頃から低価格化を目指すものと、価格は維持しながら高機能化を目指すもの、更にはこの双方をめざすものに分化して来ている。

ユーザサイドにおいては一昨年までは価格が同等のものが多かったため機能の対比でだけで商品選択を行えばよかったものが、最近では価格と機能の両面から商品の選択を行わなければならなくなった。反面自分の翻訳目的に適合した商品が合理的な価格で選べるようになったという利点も出て来た。

翻訳ソフトメーカーサイドでもこのような競争激化に伴い、市場ニーズに把握して行かなければ競争に勝てないとの判断から開発方向を絞り込む動きもある。例えば日英翻訳のみに重点をおいて開発しているブラザー製品や英日翻訳ソフトのみというカテナ製品もある。いたずらに対象言語を拡大するより、充実した中身の濃いソフトに仕上げるためとの狙いがあるのかも知れない。また最近の輸入の増加もあって、英日翻訳の需要が70%近くに達しているという市場動向から英日翻訳ソフトの新製品の増加も目立って来ている。

PC翻訳ソフトの価格面だけから見れば、現在市販されているものは日英翻訳での最低価格は98版で49,800円、Windows版翻訳支援システムで75,000円、パーソナルユースタイプで39,000円というものもあり、英日翻訳ソフトでは28,000円や39,000円というものがある。また翻訳ツールではあるが1万円を割り込んだ9,800円というものもある。こう見てくると翻訳ソフトも大幅な価格ダウン傾向が続いているというのが現状である。

先程も述べたように商品や価格が分化して来ているが、商品企画意図をに調べてみよう。

まずタイプ別に見て、Macタイプ、Windowsタイプがあり、WindowsタイプにもPC98系とDOS/V系が現在主流となっている。そして、翻訳ソフトはそれぞれのタイプにのみ適合するものやMac、Windows双方に適合するものなどがある。ユーザのハード使用機種を見越しての対応であろう。傾向としてはWindows版の増大が目立っている。

さて動作環境面ではやはり高性能なCPUを搭載した

機種に対応したものが必要であるが、最近のハードはその殆どが32bitクラスである。やはり10万にも及ぶ文法ルールを駆使しての翻訳であり、高性能なCPUがないと翻訳処理に時間がかゝるからである。しかしPC翻訳ソフトを利用しても、5万語/時間の高速翻訳が出来る機種もある。

また翻訳には大容量のメモリーが必要となる。翻訳する文書の種類にもよるが、出来るだけ大容量のメモリーのものを利用するほうが処理しやすい。

ソフトは設計者の思想の現れとも言われるが、翻訳機能については各社各様である。どのような機能が備わっているのかこれを羅列してみよう。

日英翻訳では翻訳モード（主語なし文の自動設定動詞表現方法の指定）一括翻訳、一文翻訳、部分翻訳、画面非表示の複数ファイルの翻訳、ステップ翻訳、訳例反映機能、エディタ機能（カット、ペースト、文字検索置換、文字編集）対話処理機能、訳語対応機能、訳語変更機能、未知語検索、シソーラス（類義語、反意語）文体指定、クリップボードのテキスト翻訳、ステップ翻訳、原文レイアウト保存機能など、機種によっては具備している機能、具備していないものもある。また特別機能としてはワープロで作成した文書をダイレクトに呼び込み翻訳することができるものもある。

英日翻訳での機能については翻訳モード設定（大文字小文字区分、複数文書ファイル一括翻訳、長文命令文翻訳方法、文体選択、助動詞表現）クリップボードテキスト翻訳、主語なし文の平叙文翻訳、並列句指定、非翻訳文指定、指定部分翻訳、一括翻訳部分翻訳、一文翻訳、画面非表示の複数ファイル翻訳、ステップ翻訳、訳例反映機能、別解釈機能、対話処理、エディタ機能（カット、ペースト、文字検索置換、文字編集、訳語対応機能、未知語検索、訳語変更機能、品詞指定、品詞変更機能、訳語学習、原文レイアウト保存機能、翻訳スピードコントロール機能、複数辞書の優先順位設定などの機能がある

これらPC翻訳ソフトを機能価格対比すると、機能が変わらず価格の下がったものや、機能を若干おとして価格を下げたものもある。機能価格とも引き下げたものはパーソナル版とか普及版、簡易版と呼んで発売されている。例えばAAMTジャーナルNo.9でも紹介したが、LogoVista E to J(価格194,000円)とE to J Personal(価格79,800円)の2機種の対比では

「E to J」は翻訳業務やプロフェッショナルな翻訳ツールを目指し、より専門的分野にも対応することを目的として豊富に専門辞書(別売)を用意しているが、一般のビジネスマンや学生などが使う場合は一般辞書と自分たちの使う言葉や頻度の高い特別な用語はユーザ辞書に登録して使えばよい」とし、機能面ではE to Jの一括翻訳できる英文の長さは80KB、121ワードであるのに対し、personalは8KB、30ワード、また別解釈機能やAI学習機能はなく、専門辞書は使用不可(但し基本辞書は10万語)別訳語の入れ替えは可能である。

一文の長さについては、技術論文など150語くらいは必要という人もいるが、語そのものに意味があるので極端なまでに短文化して翻訳しなければならないとする人もいる。日本の法律で一文が3頁にも及ぶ条文もあるようであるが、一読しただけでは理解はまず不可能であろう。このような極端な例は別にしても文の係受けや、文の前後関係からしても長ければ良いとばかりは言えない場合がある。

さらに翻訳業界を驚かせたのは昨年10月末に発売されたWindows版英日翻訳ソフト「こりゃ英和」、価格は1万円を切った9,800円のソフトである。マスコミ向けのパブリシティには「英日翻訳ソフト」とも標記されているが、「初心者ユーザをふくめた幅広いユーザに翻訳ソフトを使用して貰うことを目的とし年間2万本の販売を目標としている」との発表である。

基本辞書は6万7千語のほか学習辞書、局所解析用辞書も装備しているので、辞書引きソフトとしても使用可能、クリップボード監視機能、自動保存機能、ヒストリー機能などがあるが、詳細については新製品紹介頁に記載されているので参照して欲しい。

□ユーザの利用動向

情報収集や情報提供のための翻訳、特定、不特定を対象とした翻訳、原著者の意志、思想を忠実に再現する翻訳、原著にとらわれず自由な発想に基づいて行う翻訳などがあり、それぞれ翻訳品質についても差が出て来ても当然である。人手翻訳といえども決して100%の翻訳品質を誇るものなど有り得ないし、またその必要のないものもある。

パソコン翻訳ソフトを駆使し、プロの業務用としてマニュアルや技術文書の翻訳を実施している企業

もある。パソコンもかつてのオフコンを上回る機能を有するまで高機能化して来ている。

このような業務用に対し、産業翻訳では個人ユーザとしての利用が大半を占める。オフィスにおける業務が多様化、専門化し、さらに国際化も加速し、どのような部門でも英文の一枚や二枚は常に飛び交う世の中になって来ているが、業務の生産性をあげるため、文具感覚で機械翻訳が使用されている。

オフィスや技術者での国際コミュニケーションは専門用語は国際語共通語化し、コミュニケーションの背景や環境も平常から理解されているだけに、翻訳においてもそれほど品質は重視されず、文意や要旨さえ分かれば良い、それよりもコミュニケーションのスピードの方がより重要視される。このような実態からPC翻訳ソフトの需要が高まって来るとみて差し支えない。情報提供にはある程度の翻訳品質を必要とされるが、これがために下訳用としても利用される。また機械翻訳の持つ豊富な辞書機能を利用し、訳文の表現力を高める創作支援機器としての利用もある。更に対話翻訳を利用し、外人の読解支援など教育機器として、日本人の外文読解支援機器としても使用されている。これらについては昨年5月に当協会が実施したPC翻訳ソフトの利用実態調査からもその傾向を伺い知る事ができる。

おわりに

今後の機械翻訳についてはパソコンの高機能化は更に持続されて、現在のワークステーションの領域にまで迫るであろうし、これによりユーザインタフェースもさらに改善され、より高品質の翻訳ができるようになるのは間違いない。しかし言葉は年々変化して行くものであり、完全自動翻訳というものは望めず、人間がインタフェースとして介在して行くことには変わらない。したがって最高の品質を望むならば翻訳者の翻訳リテラシーを醸成していくことが重要である。またより良い翻訳環境を目標とするならば、互換性の確保や標準化をより進めて行く必要があるし、また翻訳インフラとしてのネットワーク利用や専門用語のみならず、用例、文例などの共用化、更には異文化個々に対応する幅広い翻訳環境を整備して行かねばならないと考えられる。

(事務局長 星野 禎 男)

PC 翻訳ソフト (略語:Win=Windows、M=メモリ D=ディスク)

メーカー	商品名	対象言語/システム	動作環境
富士通	ATLAS Win J/E V1	日英機械翻訳システム	Win. 10MB(M)35MB(D)
	ATLAS Win E/J V1	英日機械翻訳システム	Win. 10MB(M)35MB(D)
	ATLAS/Win	双方向翻訳オプション	Win. 3MB(D)
沖電気	PENSEE E/J	英日翻訳支援ソフト	Win. (UNIX系)8MB(M)40MB(D)
	PENSEE J/E	日英翻訳支援ソフト	Win. (UNIX系)8MB(M)40MB(D)
	PENSE for WindowsV2.0	日英/英日翻訳支援	Win. 8MB(M)40MB(D)
ノヴァ	PC-Transer/je for Win.	日英翻訳	Win. 8MB(M)20MB(D)
	PC-Transer/ej for Win.	英日翻訳	Win. 8MB(M)20MB(D)
	PC-Transer/ej for Mac	英日翻訳	Mac. 12MB(M)20MB(D)
	PC-Transer/je for Mac	日英翻訳	Mac. 12MB(M)20MB(D)
	PC-Transerカップル版	英日/日英翻訳	Mac、Win
	PC-Transer		DOS版、
ロゴヴィスタ	LogoVista E to J 2.1	英日翻訳支援	Win. 6MB(M)35MB(D)
	LogoVista E to J 2.1	英日翻訳支援	Mac&PowerMac6MB(M)35MB(D)
	LogoVista E to J Personal	英日翻訳支援	Win. (98, DOS/V系)9MB(M)25MB(D)
	LogoVista E to J Personal	英日翻訳支援	Mac&P. Mac 5MB(M)25MB(D)
ブラザー	TransLand	日英翻訳	Mac&P. Mac 10MB(M)20MB(D)
	TransLand	日英翻訳	PC-9801版 8MB(M)16MB(D)
	TransLand for Windows	日英翻訳	Win.
カテナ	こりゃ英和	英日翻訳ツール	Win. 6MB(M) 20MB(D)
	The TranslatorV2.5	英日翻訳	Mac, P&Mac 6MB(M)20MB(D)
	The Translator Mini 1.2	英日翻訳	Mac&P. Mac 5MB(M)20MB(D)
亀島産業	E-J BANK for Windows	英日翻訳支援	Win. (M)640KB以上 3MB(D)
	E-J BANK	英日翻訳支援	DOS/V 640KB(M)2MB(D)
	E-J BANK Jr	英日翻訳支援	
	J-E BANK for Windows V3	英文作成システム	Win. 4MB(M)4MB(D)
	J-E BANK V2	英文作成システム	MS/DOS 640KB(M)2MB(D)
イリスインター	Power Translator TM	英 独、仏、西/英語双方向	(プロ版) DOS版148,000円
	全	全	(普及版) DOS版 38,000円
エムティラボ	ブラビスJ/E V5.1	日英機械翻訳システム	DOS、DOS/V 20MB(HDD)
ソフトウェア技術	E tran J win	英日翻訳ソフト	Win. 5MB(M)
高電社	j・London/EJ	英日翻訳ソフト	Win. 12MB(M)50MB(D)
スピリット	X-EJ2/W for MS-Windows	英日自動翻訳	MS/DOS Win. 4MB(M)3MB(D)
	X-DIC/W For Windows	英日/日英双方向翻訳支援	MS/DOS, Win. 4MB(M)6MB(D)
	X-DIC V3.0	英日/日英双方向翻訳支援	MS/DOS DOS/V 640KB(M)6MB(D)
	X-EJ MK-2	英日自動翻訳システム	MS/DOS DOS/V 640KB(M)2HDD(D)
創歩人コミュニケーション	HARIS for Windows	英日翻訳支援ソフト	Win. MS-DOS 4MB(M)3MB(HDD)

価 格	基本辞書	主要機能・特徴
198,000円	13万語	テキストファイル一括翻訳、対話、一括翻訳、クリップボード翻訳、訳文スタイル選択
198,000円	13万語	テキストファイル一括翻訳、対話、一括翻訳、クリップボード翻訳、訳文スタイル選択
168,000円		本機をインストールすれば双方向翻訳が可能
198,000円	13万語	システム、経済、ビジネス用語標準装備、対話、一括、部分翻訳機能、未知語検索機能
198,000円	15万語	システム、経済、ビジネス用語標準装備、対話、一括、部分翻訳機能、未知語検索機能
198,000円	日英15万英日13万語	編集翻訳ファイル保存、辞書引機能、システム辞書拡張性、優先辞書指定、未登録語検索
198,000円	6万語	シソーラス15語標準装備、訳例反映機能、複数ファイル同時オープン、バックグラウンド翻訳
198,000円	6万語	翻訳速度5万語/時、訳例反映機能、ステップ翻訳、バックグラウンド翻訳、複数ファイルオープン
198,000円	6万語	翻訳速度12万語/時(P. Macの場合)訳例反映、バックグラウンド、ステップ翻訳、複数ファイル解放
198,000円	6万語	22万字/時(P. Macの場合)シソーラス15万語標準装備、訳例反映機能、複数ファイルオープン、
各346,000円		PC-Transer ej/jeを同梱
DOS(98用) 248,000円	DOS/V用 298,000円	
194,000円	10万語	多機能、ユーザインタフェイス(選択文翻訳、別訳語、対応語検索、品詞指定、優先辞書選択、
194,000円	10万語	オープンシステムとスピードコントロール、ユーザインタフェイス(選択文、別訳語、対応語他)
79,800円	10万語	英文長さ制限30ワード、別訳語入替可能、未知語検索、品詞指定、翻訳速度自動設定、
79,800円	10万語	Windows版と同じ
69,800円	5万語	意味トランスファ方式、ユーザフレンドリー操作性(別訳、カットコピー、ペースト、検索置換、
49,800円	4.3万語	再編集、再翻訳可能、一文逐次入力翻訳、テキストファイル読み込み
79,800円	5.2万語	クリップボード翻訳、ユーザ提携文辞書、コピー、カット、ペースト、検索置換
9,800円	6.7万語	翻訳可能文200文字、クリップボード監視機能、自動保存機能、豊富な辞書引機能、ヒストリー機能
49,800円	6.2万語	再翻訳、和文編集も可能、ステップ翻訳、豊富な編集機能(別訳語、訳語変更、品詞指定、検索、置換)
9,800円	2.5万語	ユーザズメモリ1MBで対応、編集機能(部分、別訳、訳語置換、単語対応、文末判定、文統合、分離)
75,000円	5.7万語	対話一括翻訳、2カ国エディタ、カット、コピー、ペースト検索、置換、移動、ワードラップ表示
98,800円	6万語	スペルチェック機能、文体指定、ヘルプ画面、ワープロ機能(カット、ペースト、置換、文字列検索)
39,000円	4.8万語	E-JBANKの普及版で辞書の拡張性、編集機能を簡素化、
75,000円	5.4万語(V2)	文章テンプレート機能、カット、ペースト、検索、置換、移動編集機能、一括対話翻訳切替可能
98,000円		ワープロ感覚編集機能(ワードラップ、カット、ペースト、文字列検索、置換拡張文字列置換)、
Mac版168,000円 OS/2版180,000円 UNIX版 480,000円		基本辞書25万語、一括、対話翻訳、Mac普及版は音声出力可能
Mac版 58,000円 Winds版58,000円 [DOS版640K、Win. 版1MB+EMS、Mac版2MB、OS/2 OS2のメモリが必要]		
190,000円	6万語	一括、部分、一文翻訳、自動和文チェッカー、未登録語検出機能、ワープロ機能
60,000円		熟語辞書1万、単語辞書5万、翻訳速度12千語/時、推敲規則、
98,000円	8万語	対訳エディタ、訳語選択/訳語学習、訳出指定、パターン翻訳機能、オンライン通信ソフト(別売)
98,000円	6万語	DOSとの互換、例文データベース装備、スペルチェッカー、学習機能、同時平行処理、辞書保守機能
96,000円	英和6万和英5.5万語	登録無限ユーザ辞書、並列処理、学習機能強化、例文データベース、スペルチェッカ
96,000円	英和6万和英5.5万語	高速辞書引き、強力エディタ、登録無限ユーザ辞書、英・和辞書の同時作成、学習機能
98,000円	6万語	構文ダイレクト方式、対話、一括ワープロ機能、辞書引きダイレクト、データの有効活用、
28,000円	4万語	ネットワークファイル指定可能、クリップボード、一文230字、任意翻訳範囲指定、別訳語候補表示

PC版ソフトで技術文献を翻訳して (機械翻訳上の問題点)

ソニー・テクトロニクス(株) 立田 種 宏

プロフィール

私はMTの中身については素人である。それに翻訳が本業ではない。私の専門はCASE (Computer Aided Software Engineering)と呼ばれるソフトウェア開発技術で、それを日本に普及させることが業務目標の一つである。そこで、その関連学会や技術雑誌への投稿、あるいは技術解説のセミナー講師などを行っているが、それだけでは普及は難しい。何故なら、CASEが他のソフトウェア関連技術、例えばC言語やUNIXなどと同様に欧米で考え出された技術であるため、英語の関連書籍の翻訳出版が欠かせないためだ。そこで、著名な英語関連図書を出来るだけ数多く翻訳し日本で紹介する事が何よりも大切で、その一環として今回「オブジェクト指向設計の標準HOOD 3.1」(海文堂出版社刊)を翻訳した。

翻訳は、今回が初めての経験ではない。CASE関連の技術論文なら、1987年に経験して以来、5~6件翻訳しているし、CASE関連の技術書なら、「リアルタイムシステムの構造化分析」と「CASEツール導入の実践」の2冊(いずれも日経BP社刊)経験している。技術書の場合、いずれも部分的ではあるがMTを使っている。

私の英語力は知れているし、翻訳スピードも速いとは言い難い。それでも、長年続けてきたのは、翻訳でお世話になった出版社から、「英和技術翻訳に必要な素養は、先ず正しい日本語が書け、次に専門知識を有し、3番目に英語力があることだ」と教えられたためだ。ちなみに、「いわゆる翻訳家はこの逆の順序だと思っている人が多い」というのが、この出版社の弁であった。確かに、その出版社の翻訳物を読むと、日本語として読みやすい訳文が多いように感じられる。

さて、今回も含め、これまで私が経験してきた翻訳にビジネス性があるかということ、殆どあり得ないと言ってよいだろう。CASEの分野は読者層が限られており、今回は定価3000円の本が初版で1500部のみである。しかも再版の予定はないし、初版でさえ全て売れるかどうか難しい状況だ。しかしこれはCASEに限ったことではない。技術書が翻訳される場合、多くは初版3000部前後の出版で終わり、一部のヒット作のみ再版されているのが実状である。

C言語やUNIXのユーザマニュアルようになり人気が高かった翻訳書でも、数万部止まりだと聞いている。この事は、CASEの分野に限らず技術書の多くは、企業が別の恩恵を期待して赤字覚悟で翻訳を引き受けるか、もしくはボランティア精神に富んだ翻訳者が出現しない限り、未翻訳に終わってしまう事を意味する。

私は、この悪循環を何とか打破し、多品種少量翻訳の道を開きたいと考え、そのためのMTシステム実現を念頭に置いて今回の翻訳に取り組んでみた。

翻訳を行う上で欠かせない条件

この翻訳の特色は、英語がネイティブでないフランス人の手による英文書物という点にある。これがため長短色々である。

長所らしきものとしては、ウイットに富んだ英文や、こなれた英文があまり見られないため日本人にとって比較的読みやすいこと、さらには形式ばった英語であるため書式の似かよっているパターン化された箇所が数多く見られ、楽な翻訳箇所が多かったことが挙げられる。たとえば「The ODS of an ACTIVE OBJECT is an object descriptor composed of the following sections」などの比較的わかりやすい構文が多く、同時にこの文の大文字だけが別の単語に置き換わっている箇所があちこちに点在している。

これは、MT作業をする上でも楽であった。その意味では、非英語圏の英文著作物はMTを使用する場合の狙い目かもしれない。

その反面、英語がネイティブでないための問題点もあった。一つには、スペルミスの多さである。その多くはスペルチェッカーなどで救えるが、例えば an と on の区別のようにスペルチェッカーが素通りし、さらにその位置によっては「グラマティックマック (Grammatic Mac)」のような文法チェッカーさえエラーを見逃してしまう場合もある。また原著者の英語力が今一つゆえに、代名詞が何を指しているのか、前の文との関係や係受け等を見てもわからない上、その分野の専門知識を働かせても判断に迷う場合があった。スペルミスや英文法の問題は、原著者や原出版社の責任とも言えるが、そのままにしておいたのでは翻訳者の誤訳を招いてしまう。そもそも、CASEを日本で普及させるために翻訳するわけだから、意味的にいい加減な訳で放っておくことは出来ない。そのため、原著者への質問状や確認作業などにかなり手間隙を要した。

過去に2回、部分的なMTを経験したことを述べたがどちらも外注方式であった。その時は「前処理とMTの使用、及び後処理」を外注にお願いした。ところが外注先の担当者が情報処理の専門家というフレコミだったのに実際は違っていたり、途中で担当者が降板したため自分がリライトする時に一貫性のない2つの文体を苦勞しながら整合を取る羽目になったりと、いずれも散々な結果であった。昨今は、グループウェアと称して積極的に作業分担する向きも見られるが、たとえば前処理と翻訳との分担ならまだしも、一つの翻訳作業を複数の翻訳担当者で分けるのは、いかがなものだろう。リライトする側にとってみれば、担当者が途中で交代した時と同じように苦勞させられるだけである。

外注の場合、更に深刻な問題は外注費である。過去2回とも、100万円を越す支払を経験している。今回の場合、出版社から翻訳で得られる報酬は、その数分の1に過ぎない。つまり英文技術書の翻訳では殆どが、外注費の半分も賄えないのである。この現状を考えると、「出来るだけ多く英文技術書を翻訳する道を開きたい」という目標実現のためには、外注などの入手は避けざるを得ない。そこで今回は、

先ずは一人で全工程をこなせるMT翻訳を大前提に取り組んでみた。

オリジナルな電子ファイルの入手は、翻訳を引き受ける場合に欠かせない条件である。今回も原著書が出版された以上は、元の電子ファイルが存在している筈だ。仮に手翻訳となっても、元の電子ファイルがあればワープロなどで用語を文字一括変換し、その用語だけでもきちっとした訳語に置き換えることが出来る。もちろん、MTを使えば更にメリットが増す。そこで、電子ファイルの入手を絶対条件とししかも出来るだけ使いやすいフォーマットでのファイル獲得に努めた。

電子ファイルが入手できなければ、OCR認識という別手段が考えられるかもしれない。しかし、それだけは避けたかった。MTソフトよりも高価なOCRシステムを一式購入しても、100%のOCR精度は期待出来ないし、図表部分を除くテキスト部分だけを旨く抽出できるのかという点も疑問である。又、認識ミスを修正し、オリジナルデータに戻すというのも大変な作業だ。

昨今の翻訳の世界では、よほどの古文書や歴史的書物など特殊な場合を除いて、電子ファイル化されていない例はないと言ってよいだろう。仮に執筆者が手書きで作成したり、タイプライターを使ったりしたとしても、印刷屋か出版社など何処かに電子ファイル化されたメディアはある筈だ。このような状況下でOCRを利用することは、建物の建築に例えるなら、建築現場でその建築資材が過不足なく正確に組み上がっているにもかかわらず、それをわざわざ全部バラして資材に過不足がないか最初からチェックし直し、確認後に再び元通りになるよう組み上げているのと何ら変わりがない。それは、単に余計な作業であるばかりか、大量の労力や時間の浪費を伴う上、その間に資材の一部が欠けたり間違っ組み上げられたりして手直しするリスクもあるだろう。

このような馬鹿げた行為にも等しいOCRの利用は厳に戒めるべきであり、翻訳関係者は先ず、あらゆる手を尽くして最も扱いやすい電子ファイルの入手に努めるべきである。

電子ファイルを手に入れた後は、そのデータをフルに活用するため、翻訳の全工程をコンピュータ上で行うことにした。又、その間は出来るだけ手作業を

伴わないようにして、エラーの混入や時間の浪費を少なくする事を心がけた。更に、その作業の中心には是非MTを据えたいと考えた。それは今や、「MTは使いものにならないと傍観している時代」ではなく、「何としてでもMTの能力を最大限に引き出して使うべき時代」である、との認識に立っているためだ。あとは、無制限に金をかけられないので、費用対効果という点を考慮してMTを選択するだけであった。

パソコン上で動作するMTの選択

今回の翻訳についてはパソコン対応のMTしか考えなかった。W/SベースやUNIXベースのものは1桁、2桁も違う価格なので、とても手が出なかった。又、パソコン通信の翻訳サービスは、それ自体の使用料のみならず、大量のファイル転送などに伴うパソコンネットへのアクセス料金や通信費が米国などの諸外国に比べると高く、結構馬鹿にならないので見合わせた。

どのパソコンソフトを選ぶかに際しては、MTについての専門的知識をあまり持ち合わせていなかったが、自分自身10年来のMACユーザであるため、それ以外のパソコン環境は考えなかった。仮に他のパソコンで動作するソフトの方が優れていたとしても、劣っている分はMACの快適な操作環境が補って余りあるだろう、との思いもあった。当時MACで動作するソフトは3つしかなく、その中から最終的に、翻訳精度もあまり高くなく、専門雑誌の評価も一番低いが、価格は一番安いK社製のT(10万円以下)を選んだ。この選択は、むしろ価格最優先に基づいたものだが、それ以外に、10万円を切るソフトでも一冊の翻訳を立派に出来ることを示したい気持ちもあった。さらには、これがうまく実現すれば、もっと高価なMTを使うことで将来、より高品質の翻訳につながる可能性もあるだろうとも考えた。

翻訳作業の開始

電子ファイルの入手に際して、出版社によっては高額な代金を請求するケースもある。例えば、過去に経験した技術書の翻訳の場合、「XYVision(サイビジョン)システムで読み出せる磁気テープなら5

百ドル、それをIBM-PC上のフロッピに変換するサービスを受けなければ5千ドルの手数料が必要だ」、と出版元から言われたことがある。幸いにして、今回フランスの出版社からは、フレームメカ3.0フォーマットで無償入手できた。

ところが、このフォーマットは欧州ではポピュラーらしいが、フレームメカというソフトウェアを持っていなかったためそのソフトを購入するか別のツールを使ってベタのテキストファイルに変換する必要があった。調査したところ、Conversions Plus(DATA VIZ社製と)という変換ソフト(2万円)が市販されており、それで変換するのが一番安上がりだとわかった。

最終的には、そのツールを購入した後になってから、現地の出版社がベタのテキストファイルに変換したものを無償で送ってきたため、そのツールは無駄な買い物となったが、余分な変換作業は避けられた。

いずれにしても、最も使いやすいフォーマットの電子ファイルを安く(出来れば無償で)入手する努力を惜しまないことが、MTの利用には非常に重要であろう。

次にMT環境の整備に入った。このMTを本格的に利用するのは、今回が初めてだったため、MTと翻訳物との大雑把な相性を調べるために全文を前編集なしに一括変換させてみた。その結果いろいろな問題点が出て来た。これらの問題点はメーカーに細かくフィードバックしたところ、その都度親切な回答を頂戴したため、リスクの高そうな問題点については事前にかなり洗い出すことが出来た。

次は辞書の整備である。まずは、原著書の最後の頁に付いているINDEXから必要な用語を拾い出し、次々とユーザ辞書へ登録した。そんな事をしなくてもCASEの分野なら、市販されているコンピューター用語辞書などの専門辞書が使いそうに思えるかもしれないが、専門辞書の購入は全く考えなかった。それは金銭的な事以上に、専門辞書が役に立たなかったからである。その理由は、ある意味で出版社の独自性の問題であり、また訳者自身がそれに対抗できる力量を備えていないためでもある。例えば、以前の翻訳の場合、出版社から事前に、ユーザ辞書に相当す

を余儀なくされている。また、別の翻訳では、CASE業界での過去の使用事例や用語を踏襲して翻訳して欲しい、と要求されている。そして今回は、ある情報処理用語辞典に準拠して訳すことにした。いずれにしても、今後は出版社がMTの文化を強く意識して市販の専門辞書とうまく付き合わない限り、この分野で専門辞書が活躍する機会はないだろう。

ところで、今回のMTの場合、いくらユーザ辞書に登録しても、基本辞書が有している別訳が優先されるために、うまく訳出されない事例が多く見られた。そこで、このような訳語は、前処理の段階から日本語に置き換えて英文に埋め込む作業を余儀なくされた。幸い、このMTは日本語混じりの英語でも日本語を記号として認識し、そのまま日本語を通してくれるので、そのような措置で対応出来た。その結果、英文の中に適度に日本語が散りばめられている文章を読むことになり、英文が格段に把握しやすくなるという思わぬメリットもあった、ところが、基本辞書が優先されているかどうかの確認は、いちいちMTの対話モードで調べるしか方法がない。この作業には、かなり手間取った。もし、MTの基本辞書が、どこかの出版社の辞書を採用しているのならば、その事を公表すべきであろう。そうすれば、その辞書を片手にしながらでも、あるいはCD-ROM化された辞書があるのならそれを検索しながらでも、ユーザ辞書に登録すべきか否かの判断が効率よく下せるだろう。もし、どこの出版社にも無い独自の基本辞書を採用しているのなら、その基本辞書を簡単に確認できる機能が欲しいところだ。そのような機能無くして、効率よいユーザ辞書登録は望めない。

次に、前処理の必要な箇所を探すことにした。そこで、MTによる一括翻訳を行って、前処理すべき所を見つけ出すという方法をとった。英日翻訳の場合、前処理はやらないという人も多いが、今回は翻訳精度に目を瞑って格安のMTを選んでいるため、多少とも複雑な英文を入力すると意味不明な日本語になることは明らかである。意味の通じない日本語を読まされてストレスが溜まるような事は避け、出来るだけ前処理を重視した。

前処理で多く見られたのが、係受けの不明な箇所

る訳語集が手渡され、それに基づいて翻訳することである。しかし、and とか or など、そのパターンは限られている。MTの中でこの手直しを簡単にできる機能が求められる。他にも、細かな問題は色々あったが、様々な工夫を凝らすことによってMTはなんとか使える、との確信が得られた。

長期的課題

今回入手した電子ファイルは、原著のテキスト部のみなので、図や表の情報が全く拾えなかった。そのため、図表のテキスト部は手翻訳を余儀なくされ図表の作成は日本の出版社が原著を見ながら最初から行った。今後、原著の図表を含めた全ての情報が利用できるような体制が必要だろう。

今回MTがうまく使えた理由の一つに、英語がネイティブでない著者による英文著書という点がある。これは言い換えれば、制限言語的な英文が効を奏したとすることも出来る。そもそも技術書は文学書ではないので、制限言語的であっても問題ない筈だ。今後、MTをもっと活用してもらうためには、原著者への制限言語的な書き方の奨励、例えば原稿料の上乗せなどの方策が取られてもよいのではと思う。そのためにも、制限言語の仕様を標準化して公表する事が求められる。

訳出した箇所に対して、例えば推敲ソフトなどを使った日本語のスペルチェックを行う向きがあるが訳者が日本語をしっかりと読みさえすれば、日本語表現の誤りは大抵見つけられるものだ。それよりもむしろ原著書に対して正しく訳されたかどうかをチェックするシステムの必要性を感じた。

日本語訳だけをいくら眺めていても、英文の一部が抜け落ちて訳されている事や、単語の一部が誤訳である事に気付くことは少なく、かなり神経を使って英文と照らし合わせても、そのミスから完全に免れる事は難しい。そこで、例えば英日と日英の2つのMTを使って英→日→英と変換した後に、元の英文と訳文の英文を比較することで日本語訳の間違いをうまく見つけ出すようなやり方などを使って、英語から正しく日本語に置き換わったかどうかを見つけ出す検証システムの出現が待ち望まれる。

トランスレータの選択

はじめに

兵庫県南部地震は未曾有の被害をもたらした。地震発生後の情報収集や初動救援体制の遅れにより、改めて政府の危機管理体制の在り方が問われ、前例がないと動けない行政の対応に批判が集中している。

地震のように突然襲って来る大きな変化、その後2日間に亘って神戸の街を焼き尽くした大火災、次々に襲ってくる恐怖の中で、瓦礫と化した町中の避難所で空腹や寒さに必死に耐えながらも秩序を守りたくましく復興の決意を語る地域住民の言動に国内外から励ましと称賛の声が多く寄せられている。

大震災を教訓に危機管理体制の在り方についての見直しが進められている。

一方実業の世界に目を転じてみると、ディスクロージャという新語が叫ばれているが、金融業をはじめ翻訳業での経営面や業務面でも危機管理体制はたして万全なのであろうか。

過去の歴史に学ぶ

機械翻訳のメーカーでもある電機業界の過去の好、不況の波を振り返ってみよう。

電機業界では過去幾度か重大な不況を経験している。ポストオリンピックの昭和39年から40年にかけて、それまで順調に成長を続けて来た家電製品の普及が一巡し、オリンピックという特需の反動もあって販売不振に陥り、メーカー、流通在庫が増大、生産調整を余儀なくされた。当時はまだまだ生産面では労働集約的な生産形態が取られており、事務、営業などの間接部門に対し、工場で働く直接部門での比率が高かったが、工場では操業率を低下させ、余剰人員は販売の第一線に転属させ、また需要喚起のために様々な新規商品の開発に取り組むなど懸命な対策が講じられた。その後次第に海外をはじめ一般景気の回復とともに生産の省力化への取り組みが始まり、またコンピュータやカラーテレビの需要も台頭して再び高度成長路線を歩み始めたという歴史をもつ。

ついて昭和46年から47年にかけて、それまで1ドル360円という固定相場制で保護されてきた我が国

の国際為替相場は米国経済の弱体化によって、レート切り上げが断行された。増大傾向をたどってきた輸出産業にとっては国際価格の高騰となり、国際競争力を低下させ、まさにドルショックという危機に直面、さらに追い打ちをかけて石油危機が訪れた。

かつて石炭に依存していたエネルギーは次第にコストの安い石油への比重を高め、先進各国は産油国である中東からの輸入に頼ることが多くなったが、原油の枯渇化懸念や産油国の影響力の行使もあって原油価格が高騰し、これが生産エネルギーや電気、石油製品、化成品などの価格に跳ね返り、世界的にも物価は30%近く高騰し、消費購買力を著しく減退させた。これがため国際競争力の弱い商品は海外へ生産拠点を移行させるものも現れはじめたが、国内においては生産設備の自動化、物的流通の合理化、商取引流通の合理化など徹底したコスト合理化が実施された。その後石油情勢も高値安定ながら落ち着きを取り戻し、また間接部門の生産性向上策としてオフィスオートメーション(OA化)が叫ばれ、パソコンやワープロ、普通紙複写機、電卓などの新規製品が誕生し、事務の生産性向上に寄与した。

また国の輸出振興策もあって日本製品は世界各国へ浸透し、これがその後大幅な貿易収支の黒字を現出するとともにダンピング問題や輸入制限など国際摩擦を生じさせることにもなった。

昭和60年、プラザ合意もあって国際為替相場は変動相場制に移行し、当時1ドル240円前後であった円は次第になだらかな円高傾向をたどり、1年半の間に120円台まで上昇した。この間収益性確保のために企画された製品も生産工程に乗る手前で、波及的に続く円高の影響で収益面での確保が難しくなりその都度設計をやり直さざるを得ないなど製品企画と円高がいたちごっこを繰り返し、暫く新製品が創出されないという時期もあった。一層のコストダウンをはかるために部品での海外調達やコストの安い海外へ生産シフトさざるを得ない結果ともなった。

その後貿易摩擦を回避することもあって輸入促進や内需拡大など政策転換が行われたが、国際収支の

黒字の持続もあって、金余り現象はバブル経済へと進展し、その行き過ぎからバブル崩壊へとつながって行ったのである。

不況時の対策

需要低迷時には新しい需要創造のための新規商品の発掘や開発、販路の拡大、不採算商品や不採算部門の統廃合、在庫調整、設備投資の抑制、組織や人員体制の見直し、各部門における生産性の向上、徹底した経費削減、資金の社外流失の防止など諸々の施策が行われるのが普通である。例えば経費削減策としては管理可能経費の一律カットばかりではなく固定的経費の削減までメスが入られる。これらを徹底させるため、権限も一部制限され、上位管理者に権限が集約されることもある。

3Kといわれる出張交通費、広告費、交際費の大幅なカット、不用不急の固定設備や材料、消耗品購入の繰り延べ、昼食休憩時間の消灯の励行、近くの階へのエレベータ昇降禁止、コピー用紙は裏表利用コピーを削減しFDでのファイル励行、飛行機禁止で新幹線も自由席利用、タクシーの利用禁止、会議は短時間励行、社外通話は3分間励行、外注化抑制と内製化への転換、残業、休日出勤の禁止、収益部門への業務応援派遣など、給料カットか経費削減かの選択を迫られるケースある。中には金額はそれほど大きくなくとも経費削減に対する従業員マインドを駆り立てる狙いが込められて取られる対策も少なくないが全社員一丸となって難局に立ち向かう危機管理体制がとられるのが通例である。

過去の不況時の抜本的な対策としてとられてきたFA、RA、OA化への移行などの改革が断行されて、それが定着し、殆ど合理化されつくした段階でのバブルとその崩壊であり、今回の不況打開策としてはリストラの一環としてのホワイトカラーの人員体制の見直しがクローズアップした。勧奨退職や系列企業への出向、独立支援制度の設定、給与体系も年俸制の導入、厳しい生産性チェックと実績評価制度などかつての終身雇用制を覆す新たな取り組みが行われている。

翻訳業の対策

かつて経団連の会長であった土光敏男氏は「変化には3つの側面がある。1つは変化の断層性であり

過去の経験と知識の上にあぐらをかいてはならない。また変化には波及性がある。自分の専門に立てこもって、安心してはいけない。さらに変化には加速性がある。変化の並に乘るタイミングや、変化を先取りしていく必要がある」とその著書で述べ、平時における危機管理体制や意識を強調されている。

翻訳の歴史は古いが、翻訳業界の歴史は新しい。景気低迷の見え始めた92年頃から翻訳受注量は減少翻訳単価の低下、また納期の短縮や翻訳品質等の条件も一段と厳しさを増した。当初は暫く辛抱すれば景気も回復に向かうとの楽観論も聞かれたが、昨年実施された「翻訳フェア」や「翻訳の日」のシンポジウムなどでは経営的危機が強くアピールされた。

このような情勢下において、翻訳業では経費の削減は勿論のこと、顧客との接触頻度増加させて営業力の強化、受注先の拡大、顧客の選別、少量翻訳の受注、更には雇用調整まで実施したところも少なくない。また納期、コスト半減を目標に人手翻訳から機械翻訳への移行や文書のトータル処理の観点から翻訳、編集、印刷まで業務の多角化への取り組みも見られる。更に登録翻訳者の戦力アップを図るため研修を実施し、個別専門領域の強化を図ったところもある。いずれを選択するかはそれぞれ企業の置かれている環境の違いにより、その対処策も異なるのは当然である。

94年10月頃からは景気もようやく上向きの方向に転じたとの政府や民間シンクタンクからの発表があったが、これも企業の縮小均衡の上立った上での業績であり、実感としてはなお厳しい状況が続いており、更にこの度の阪神大震災で景気回復のテンポが遅れるとの見通しもあり、まだまだ予断を許さないし、また円高の追い打ちも見えかくれる。

これらマクロな一般景気情勢のみならず、バブル崩壊による翻訳業界固有の影響について、特に機械翻訳と翻訳需要の変化について考えてみよう。

企業内翻訳業務の変化

翻訳を業とする業界としては、翻訳サービス業、印刷業、さらに企業内翻訳に大別でき、90年の翻訳需要は4800億円前後、現在では7600億と推定される。

バブル崩壊に伴い企業内の翻訳業務にも様々な変化が見られる。先に述べたように、企業では資金の

外部流出や徹底した経費節減もあって翻訳業務を内部処理する傾向が強まったことである。

少なくとも30年くらい前までは海外取引は専門商社へ依存する傾向が強かった。しかし輸出の拡大によって企業は輸出専門部門の設置や海外に出先機関を設け、海外情報の収集や輸出業務を自ら実施するようになった。しかし当初の翻訳業務は専門部門で当然行うべきものとの考え方が強く、現在でもその風潮は根強く残っている。しかし海外との交流が輸出部門のみならず海外企業との技術提携や海外生産シフトなどにより社内各部門へと拡大し、マニュアルや説明書など大量翻訳業務を外部に依存するようになったが、バブル崩壊による経費削減の嵐はこの翻訳外注部門も例外ではなかった。

バブル期においては機械翻訳の操作性面や翻訳品質から「機械翻訳はホコリを被ったままのものが多し」との声が数多く聞かれたが、ユーザの品質に対する声を忠実に汲み上げてメーカー側は改善を加え、中には年間2回以上もバージョンアップした商品が創出されたり、ウインドウズ版の出現によって、機器操作性も抜群に向上したマシンが多く創出されている。

更に人材の有効活用から機械翻訳専任担当者が設置され、企業専用語や業界専門用語も蓄積され、また翻訳技能も向上、また習熟度や翻訳のリテラシーも高まって機械翻訳システムを十分に使いこなすところまで進んできている。

更に企業内の個人評価にも業績評価や生産性評価が存分に導入されたこともこの利用効率を増進させ翻訳依頼部門でも同じ観点から社内利用促進へ拍車を駆けてきたことは充分におしはかることができる。

86年における民間企業の海外長期滞在者は16万4千人であったが、91年には41万2千人にも増加した。これは駐在員交替等で企業内にも海外経験者が著しく増加したことを物語っている。また国際化の進展に伴い、海外からの技術者等の企業への受け入れも多い。更に80年の外国人登録者は78万3千人であったのに対し92年には128万1千人に増加しているし、長期滞在者は88年で8万1千人に達している。

また学校教育も87年台、92年台と段階を追って語学力は向上し、これら優秀な学生が続々と入社して来ていることも見逃せない。

当然のことながら企業の属する業界知識や海外の

出先機関との交流も活発で、分野別用語にも卓越しまた情報発生の背景や環境も熟知した専門集団が形成されつつある。

また社内外で作成される情報は電子化されたデータがLANで自由に利用出来る環境も整っており、ことに社内翻訳について云えばその内部情報量と伝達スピードは第三者の立ち入るすきを与えない位早い。

こう見てくると企業内翻訳が醸成される土壌は十分に整って来ていると見てよい。

「企業内翻訳は人件費から見て結局コスト高になるのでは」との見方もあるかもしれないが、翻訳の外注管理、支払い管理や外注折衝、ユーザ辞書の交付の手間などトータルの考えれば果たしてどのような計算になるのであろうか。更にマニュアル、説明書など商品が発売されるまでは社外秘のものも結構あり、外注先でのリスクの懸念もある。

オフィスパーソナル需要も

対外的な情報提供の形態をとる業務上のマニュアルや説明書といった翻訳物のみならず、自己の業務遂行のための外国文献の翻訳や専門書の翻訳など情報収集を目的にしたパーソナルニーズは件数的には圧倒的に増加している。

パソコン翻訳ソフトは90年頃から本格的に市場に出回り始めたが、発売当初の価格は20万円近く、これにハード本体を含めると結構高価な買い物となっていた。

パソコンハードは91年から93年にかけて需要は低迷したが、半導体の高密度化が一層進行し、当時100万円前後であったものが、94年には40万前後にまで価格ダウンし、この傾向は95年に入っても続いている。

またパソコン翻訳ソフトも94年の後半頃から需要の増大傾向を反映して販売価格は一時の半値以下となり、またウインドウズ版の出現により操作性が改善され、また本体機能の大容量化やニーズの先取りで翻訳機能も著しく向上し、オフィスの個人ユーザにとっては非常に買いやすいものになり、現在爆発的な売れ行きを示している。

「前編集などの手間が大変で、翻訳品質も問題」と懸念する声もあるが、「専門用語はおてのもの、文意さえ掴めればいい」というケースも少なくない。殆ど完全に近い翻訳文であっても自分にとってあま

りに役に立たない情報では何の価値もない。それより翻訳品質は多少落ちてても、有効な情報はそれをヒントとしてアイデアをふくらませ、新しい創造力を生み出して行くことの方がより生産的である。

海外からの情報は緊急を要するものも少なくない。外注に出すにしても面倒な発注手続きや納期的にも問題がある場合が多い。また同業、同種の情報が多く、情報の背景や環境が文面に書かれていなくとも容易に把握出来る。この点外注し、翻訳者に環境説明する暇があるならば多少翻訳の品質に問題があろうと自分の判断で処理した方が早い。仮に検討違いの解釈があっても相手方とのやりとりの中でお互いに理解しあえるようになるものである。要は品質よりもスピードである、コミュニケーションにはリアルタイムが必要との考え方が強い。また相手側からの情報提供は兎に角分からせよう、正確に伝えようとの意識で丁寧に書かれるのが普通であり、わけの分からない文章は書く方に問題があるとの考え方が強い。

従来はかかる海外情報は分厚い書類の下にかくれて見過ごすか、放置されるケースが多く、ビジネスチャンスを逃すことも多かったが、PC翻訳ソフトの普及で情報消化量と情報の再生産量が増大したとの声が強い。かつてコンピュータや複写機は専用の部屋におかれ、専門要員によって集中運用されていたが、現在では部門毎あるいは個人にまで分散処理されるようになり、またそれぞれがネットワークで結合されるようになったが、機械翻訳も同じような経緯をたどっている。

このように翻訳を取り巻く環境もバブル崩壊を契機として大きく変わり、翻訳の世界の勢力分野も変化しつつある。

翻訳者の生き残り戦略

今年に入って翻訳需要は上向き傾向と云われるが日英翻訳、つまり輸出商品のマニュアルや説明書の減少に変わって、輸入拡大に伴う海外企業からの翻訳需要、つまり英日翻訳の増大が目立つ。本来日英に対し英日翻訳はコストも安く設定されているので仕事量の割りには売上が上がらないとの声もある。

またこの需要も決算期を控えての駆け込み翻訳もあろうし、また引き続き円高懸念もあり、このまま順調に推移して行くものか、またバブル期当時の全盛時代に戻るのは何時かは未だに確たる見通しも立

ちにくい。

並いる翻訳業の中で競争に打ち勝ち、優位性を保持していくには何と云ってもは顧客に信頼され、容認される企業としての個性が必要である。先ほども述べたように「納期」「コスト半減」宣言や、特許医療など特定分野では他社の追随を許さない「翻訳品質」を誇るとか、さらには翻訳の原点に立ち返り原文作成者の意志や思想を忠実に訳文に反映させる等、原文のもつ環境や背景説明まで踏み込んだ著述代行を手掛け、同業他社との差別化を図りながら収益確保に努めて行く事もアイデンティティを発揮することに繋がるのではないだろうか。翻訳企業である限り何でも大量にこなさなければならないという考え方も理解できるが、発注する側からすれば、やはり信用がおけて、尚且つ早く安く翻訳してくれる業者の方が良いのである。

付加価値性指向の料金体系を

翻訳料金は「枚数×単価」で設定される場合が多い。本来、知的集約産業である翻訳サービスが労働集約的価格体系を取っているのも腑に落ちない。翻訳の目的や種類によって付加価値性の高い翻訳、翻訳の質によって価格が設定されるべきであろう。一般ユーザの間では「翻訳料はまだ高い」という印象がある。翻訳者の人件費や辞書、一般経費を考えれば「とんでもない」との反論も当然出てくるだろう。購買力平価で換算しても日本の物価は世界的に高い。各種規制や地価、オフィス生産性の低さが人件費や物価の高騰を招いた要因ともいわれる。翻訳の生産性を上げるには価格的にも買い易くなった機械翻訳を導入すべきであろう。

「機械翻訳の翻訳品質なんてマダマダ」という翻訳者も多いが、機械翻訳は自動翻訳ではない。翻訳支援システムである。下訳として、また辞書引きにも大きな効力を発揮するし、翻訳者がインタフェースとなって訳文を手直し、粗訳で良いという顧客には簡単な前処理だけで訳文を仕上げて行くという方法もある。それぞれ顧客の翻訳目的の違いに応じて翻訳品質に差をつける考え方があっても不思議ではない。

知的資産の構築

翻訳業が知識産業であるならば知的資産はどこま

で構築されているのであろうか。翻訳ノウハウや翻訳技能は翻訳者の属人的なものであり、企業として資産化できにくいとの声もある。しかしその有能な翻訳者をいつまでも企業に縛り付けておくことは危険度が高い。翻訳料金が高く、翻訳価値を認めてくれる企業へとホッピングして行くことは有り得る。しからばこれらを機械翻訳（コンピュータ）の人間も及ばぬ絶大な記憶力を活用して知的資産を構築し翻訳者に共用させて行くことが生産性を高めることにつながるのではないだろうか。ユーザ辞書、専門辞書はもとより業界によって違う文体や言い回し、用例や定形不定形を問わず各種様式や書式、翻訳事例などを資産化し、再利用し、人が変わっても資産は目減りしないという対策が必要であろう。

ニーズの先取りへ

さらに受注翻訳から攻めの体制も必要である。日本の海外直接投資はいまだGNP比6%程度であり欧米の20%に比べればまだまだ低い。コスト吸収のためにこれからも海外生産や直接投資は増大して行く。新規投資先の情報はJETROあたりでもかなり資料化、頒布されてもいるが、個々の企業にとってはより詳細で具体的なものを必要としている。例えば海外進出にあたっての各種インフラの現況とか商品、販売、投資、税務上の規制、部品調達、生産、管理、品質労務、研修マニュアルなどの多岐にわたる情報を顧客に提供するサービスも必要であろう。いずれこれら対象国向けの翻訳の受注があるとすれば、翻訳を行う際の習慣、背景や環境理解に大きく役にたつことのもなる。

機械翻訳は原文を10万近い文法ルールによって解析し、目標言語に置き換え、目標言語の文法に従って生成、訳出するという過程を辿る。入力された原情報はコンピュータ処理で正確に、高速で処理されて行く。しかし入力されていない情報は、つまり原文にないデータは処理できないのである。従って機械翻訳は直訳、逐語訳であり、これがため比較的構文の簡単な技術書やマニュアル、説明書など原文忠実度の高い、大量翻訳に最適といわれてきた。そして最近では構文構造の比較的簡単なオフィス文書の翻訳処理にも広く使われるようになって来た。オフィス文書は部門により使用される用語が多岐にわたることもあって、専門用語をすべて装備するには費用も嵩

むが、基本辞書のみでの翻訳で、多少翻訳品質を犠牲にしても文意が分かれば良いとの使われ方もする。

しかし真の翻訳とは原作者の意志や思想を忠実に読者に伝えて行くことである。従って原文にない文書の行間や余白部分も翻訳しないと異文化人や不特定多数の人々に正確には伝わらない。まして格調の高い講演文や督励文、情感豊かな表現力に富んだ販促文、異感情、異環境の外人にも感動を呼び起こす翻訳文が必要で、直訳的な機械翻訳の出力文ではそのまま使えない場合が多い。これがため原文の文字だけではなく、原作者の心をも読みとる理解力、相手感情や感覚に置き換える展開力、さらにその表現力に知識を働かせるべきで、スペルチェックやリライトは機械に任せるべきであろう。これぞプロの翻訳といわしめるには作家的、コピーライターの、テクニカルライターの、時にはアジテーター的な要素をも駆使し、個性のある翻訳文化を構築していくことこそプロの翻訳者であり、これが増大しつつある企業内翻訳者との差別化にもつながる。これがためには訳文作成の推敲にも機械翻訳を利用すればよい。辞書引き自分流辞書の構築も可能、また用例や訳例を記憶させることもでき、再利用も可能である。

多民族国家のアメリカは英語が公用語ではない。これでは経費的にも無駄が多いとのことで英語を公用語とする州も現れ始めた。現在の機械翻訳は英日、日英が殆どであるが、標準的な英語が使用されている。しかし同じ英語でもロイヤルイングリッシュといわれる英国で通用してもニューイングリッシュと云われるインド、フィリピンなど、或いはアメリカでも通用しないケースはいくらでもある。プロの翻訳者はこれらを的確に区分して翻訳していくことも必要である。言葉や文字は生き物で時とともに変貌して行く。日本では科学技術分野だけでも新語が年間1万語も誕生するという。これを機械利用して記憶させ、的確にフォローしていくことも重要である。翻訳能力は感性もあろうが外語を読んだり、聞いたりする量に比例して向上するといわれるが、産業翻訳でも平時から業界の動向や業界環境など受注如何に拘わらず勉強して行く必要がある。不況とは云え国際化のなかで翻訳業務は無くなることはない。どのような方法を取って行くのか、これは翻訳者の選択の問題でもある。

（中小企業診断士 星野 禎 男）

機械翻訳を用いた翻訳作業効率化の可能性

NHK放送技術研究所 先端制作技術研究部 主任研究員 江原暉将

1. はじめに

機械翻訳の処理速度はきわめて速いので100%の正翻訳率があれば作業時間は短縮されるが、現実には品質面で種々問題がある。訳出された文を後編集をしたり、また原文にもどって前編集をしたり、訳語選択をするなど非常に手間がかかる。「どれくらいの正翻訳率があれば機械翻訳を利用して有効であるか」ということは誰も興味をもつ点である。

本文では外国語(英文)を母国語(日本語)に翻訳するケースを取り上げ、上記の問題を考察する。

2. 正翻訳率と翻訳作業時間の関係

外国語を日本語に翻訳する目的は種々考えられる。それをここでは、2つに集約する。1つは情報収集、つまり自分のために外国語で書かれた文書を読むということ、もう1つは本当に翻訳する、つまり他人に情報を提供するという場合である。この目的のちがいでによって手翻訳の場合と機械翻訳の場合の4つの利用方法がある(表1)。ここで考察するのは、情報収集と情報提供の各々の場合についての、手翻訳と機械翻訳の作業量である。

表1 考察する手順の種類

	情報収集	情報提供
手 翻 訳	[1]	[2]
機械支援翻訳	[3]	[4]

まず計量的な分析をするために、作業量を単位文書量あたりの翻訳作業時間としてとらえる。単位文書量としては、例えば日本語で400字詰原稿用紙1枚、英語ならば150ワード、A4で1頁などの単位である。この単位あたりでどれ位の時間がかかるかで検討を進めてゆく。

ここで、以下の記号を用意する。

r:reading
 t:translation
 w:writing
 p:editing (pre-editing post-editing)

 s:souce language
 t:target language

 h:hand
 m:machine

たとえば、原言語を読む時間(単位文書量あたり)は r_s と書き、機械で翻訳する時間は t_m と書く。

2.1 情報収集

まず情報収集の場合を考える。作業①にあたる手翻訳とは原文を読むということであるから、原言語を読む時間 r_s がかかる。

情報収集で機械支援翻訳をつかった③の場合は以下ようになる。まず、正翻訳率として原文を参照しなくとも意味がわかる文の率を考え、 A とする。機械支援翻訳の手順は、機械翻訳をおこなう時間(t_m)、訳出されたターゲット言語(日本語)を読む時間(r_t)がまずかかる。ところがそのうち、読んでも分からなかった部分の原文を再び読む必要があり、 $(1-A) \cdot r_s$ という時間がかかることになる。この①と③を比較する。機械支援翻訳を分子に、手翻訳を分母にとると、作業時間の比は

$$M_g = \frac{t_m + r_t + (1 - A) \cdot r_s}{r_s} \quad (1)$$

となる。

ここで、訳文を読む時間と原文を読む時間の比を

$$k = \frac{r_t}{r_s}$$

とする。この k の式を使って(1)式を書きかえると

$$M_g = \frac{t_m}{r_s} + k + (1 - A)$$

となる。機械翻訳の処理時間はきわめて速く、これをゼロと考えると

$$M_g = 1 + k - A$$

となる。結局 k と A の大きさがどうかということできまる。つまり原文を読む時間と訳文を読む時間の比 k が 1 であると、これは、英語を読むことと日本語を読むことがかわらない人であるから、M_g の値は 1 になるので、機械翻訳は当然のことながら必要がないことになる。一般に、日本人は英語を読む時間より日本語を読む時間の方が速いので k の値は 1 より小さくなる。仮にこれを 0.6 と仮定すると (1 + 0.6 - A) が 1 よりも小さくなる範囲の A、つまり、A > 0.6 が達成できていれば、機械支援翻訳が有効となる。

2.2 情報提供

手翻訳で情報提供する②の場合、原文を読む時間 (r_s)、日本語に翻訳する時間 (t_h) と訳文を書く時間 (w_t) という時間がかかる。実際翻訳上ではこれだけ厳密にわけるとはできないかもしれないが、便宜上このように区分して検討する。

情報提供を機械支援翻訳でおこなう④の場合は以下ようになる。まず、ここでの正翻訳率を後編集が可能である文の率 B と定義する。機械の出力をみてこれは後編集をした方が速いのか書き直すほうが速いかを判断するが、後編集をした方がよいと判断した場合の文の率である。作業手順にしたがって作業時間をもとめると、

原文を読む	r _s
前編集をおこなう	p _s
機械翻訳をおこなう	t _m
訳文を読む	r _t

という時間がまずかかる。そして文全体の内 B の部

分は後編集をおこなう。

後編集をおこなう B · p_s t

のこりの (1 - B) は再度読み直して、手翻訳をおこなって、更に訳文を書くという時間がかかる。つまり

再度読み直す (1 - B) r_s

手翻訳をおこなう (1 - B) t_h

訳文を書く (1 - B) w_t

となる。したがって機械支援翻訳をつかって情報提供をおこなう場合はこれらを合計した時間がかかる。

結局、情報提供の場合の手翻訳と機械支援翻訳の場合の作業時間の比は

$$M_p = \frac{r_s + p_s + t_m + r_t + B \cdot p_t + (1 - B)(r_s + t_h + w_t)}{r_s + t_h + w_t} = \frac{r_s + p_s + t_m + r_t + B \cdot p_t}{r_s + t_h + w_t} + (1 - B) \quad (4)$$

となる。M_p < 1、つまり M_p が 1 より小さければ機械支援翻訳の方が有利になる。ここで、

$$T_p = \frac{p_t}{r_s + t_h + w_t} \quad (5)$$

とおく。分母は手翻訳に要する時間、分子は後編集に要する時間(いずれも単位文書量あたり)である。また

$$T_m = \frac{r_s + p_s + t_m + r_t}{r_s + t_h + w_t}$$

とおく。この分母は手翻訳に要する時間、分子は機械翻訳のうち、後編集と手翻訳による再翻訳を除いた部分の時間(いずれも単位文書量あたり)である。T_p と T_m を用いて、4式および M_p < 1 の式を書き直し、整理すると、以下の条件が満足されれば機械支援翻訳の方が有利になることがわかる。

$$B > \frac{T_m}{1 - T_p}$$

これをグラフ化したのが(図1)である。B(正翻訳率)をパラメータとし、1.0, 0.8, 0.6, 0.4の場合をしめした。これらの直線より下の部分が機械支援をした場合に役に立つ領域である。

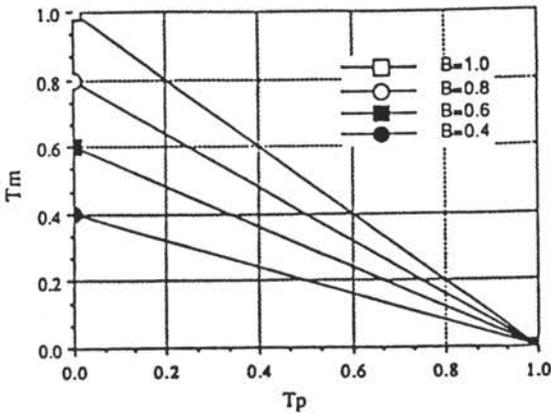


図1 機械支援翻訳が有利な領域(情報提供)

Tp(手翻訳と後編集の時間比)が1というのは、手翻訳と後編集の時間が変わらないということであるから、機械翻訳の利用は無意味となる。しかし手翻訳するよりも後編集をした方が半分位(0.5位ですむという場合)は正翻訳率の値に応じて、役に立つ領域が生まれてくる。これを具体例で示すと英日翻訳の場合で単位文書量を150ワードとし、英語を読む時間rsを2分、手翻訳する時間thを15分、日本語に書き直す時間wtを15分、後編集にかかる時間ptを20分、前編集時間psを5分、機械翻訳処理時間tmを30秒、出力(日本語)を読む時間rtを2分とすると

$$T_m = \frac{2 + 5 + 0.5 + 2}{2 + 15 + 15} = 0.30$$

$$T_p = \frac{20}{2 + 15 + 15} = 0.63$$

(後編集が63%位で出来る)

となり、

$$B > \frac{0.30}{1 - 0.63} = 0.81$$

ならば、つまり正翻訳率が81%あれば機械翻訳は有利となってくる。

3. おわりに

機械翻訳利用の有効性を達成するためにはどの程度の正翻訳率が必要であるかを、作業時間からみて定量的にもとめる方法をしめし、1例として、情報収集では、60%程度、情報提供では80%程度の正翻訳率が必要であることをしめした。ただし、はじめにものべたように翻訳の目的や手順はさまざまであるので、利用者は、現実の場合にそくして、式を変更し、計算しなおす必要がある。

また、機械翻訳利用の目的は、単に作業時間を短縮するためだけでなく、訳語の統一やセキュリティの確保などの目的もありうるため、ここでの評価法以外にも種々の評価基準があることはいうまでもない。

さらに、ここでは、外国語を母国語に翻訳する場合を考察したが、母国語を外国語に翻訳する場合もまた、手順がことなるであろう。その場合も式を変更することが必要であるが、基本的に、単位文書量あたりの作業時間で効率を評価することはこの場合も可能である。

報道ニュース文に対する使用経験から、現状の英日機械翻訳システムを広範なニュース文に適用した場合、正翻訳率は40%程度であると筆者らはとらえている。経済ニュースのような、かなり定型的な表現のおおい分野にかぎった場合でも、70%程である。ただ、これは、英日の場合であり、韓日ではもっと高い正翻訳率がえられるであろう。

近未来において、正翻訳率を向上させる有効な方法として、筆者は、共起辞書の利用があると考えている。機械翻訳の誤りのおおくは、構文解析の誤りと訳語選択の誤りである。語と語の共起関係を利用して、その双方の精度を向上させることができる。最近、共起辞書をユーザー登録できるシステムがあらわれてきたことは大変のぞましい。ただ、共起辞書は、単純に考えると、単語辞書の2乗のオーダーの規模となり、作成が困難であるという問題が生ずる。この場合も、分野を限定して、語彙量を制限することが効果的であろう。

HT、MTの生産性比較

東芝 情報・通信システム研究所 天野 真家

はじめに

機械翻訳については、アカデミックサイドでは原理的研究が主となるので、トータルシステムを作らなくても、翻訳できれば良い訳であるし、又、場合によっては、意味的に正しければ良いのであるが、商業翻訳、翻訳サービスサイドでは矢張りお客様が満足してくれる翻訳、あるいは翻訳システムでなければ商売にならない。

品質改善への取り組み

図1 機械翻訳システム



これがためどのような知恵を出そうかという問題であるが、その1つは100%機械を使ってやるという完全自動高品質翻訳と、もう1つは推敲しやすいシステムを如何に上手に作るかという事になる。

「下手なシステムを作るくらいなら、とにかく辞書がディスプレイに出て、あとはエディタで校正でき

るようにしてくれさえすれば良い」ということが翻訳者側では言われる。なまじ翻訳などをしてくれて、機械の文体で訳出されると、推敲しにくく困る、辞書だけ出してくれればいいんだ」という事である。

「使いやすいエディタがあればそちらのほうが助かるといわれる」こともある。しかし技術者の側からすれば、出来る限り機械でやって、その後、人間に委ねるという方法をとりたいという考えで、翻訳品質を向上させるという技術開発は熱心に行なわれている。

翻訳は日本語から外国語が出て来ればよい、外国語になれば良いと言うものではなくて、トータルで考えなければならない。私もいろいろ翻訳業の人と付き合ってきたが、翻訳者は1時間に大雑把に云々と400字詰原稿用紙2枚程度、早い人は4枚以上やる人もいるが、一般的には2枚しか翻訳出来ないということである。1日8時間とすると16枚程度しか仕上がらない。1枚の翻訳料金はユーザからの受取価格は種々あるが、高品質を売り物にしているところでは5000円、しかし実際は3,000円前後が一般であろう。このような現実即して、機械翻訳も価格的に対抗して行かねばならない。そこから翻訳コストの対比や計算が始まるのである。

翻訳プロセス

まず翻訳プロセスを考えてみよう。大袈裟に言えば翻訳工場みたいなものを目指したい。そうしないと人間翻訳には勝てない、つまりOCRで文書入力し、一括翻訳をさせて、後編集を行い、DTPで訳出するという一括流れ作業でやらないと人手翻訳には勝てない。

人間の場合、眼で文を読む＝翻訳であるため、OCRのような過程は必要ではないが、出力が人手になるので、そこで遅くなる。機械の場合、眼で読むところが弱い。OCRは認識率99.5%といわれているが、英文の

場合A4で1頁2,000字とすると、10字の誤りが出てくる計算になる。しかしこれは理想的な原稿の場合であって、悪い印刷や原稿の汚れなどがあると95%にも認識精度が落ちる場合もある。そうすると100字の誤りがでる計算になる。この誤りを捜し出すのがまた大変である。この場合、人間が眼で捜し出すことは意味がないので、スペリングチェッカーに掛ける。このようなことをやっているとまた時間が掛かってしまう。現状でもリーズナブルな価格のOCRの処理速度はA4判1枚で2～3分はかかる。スペリングチェッカーはそれほど時間はかからないが、誤りを発見してこれを修正するにも、1語で数秒はたちまちかかってしまう。これが100語もあれば15～20分もかかる計算になる。このような事を勘案して機械翻訳と人手翻訳のコスト比を出すため、翻訳を外注して、一方では機械翻訳+人間推敲で同じ文章を翻訳し、比較してみた。お金に換算しても機械翻訳の方が安くついたという経験をした。もっともこれは数年前の話ではあるが…。

図2 翻訳時間の推定

(例、英文A4サイズ100頁を翻訳)

	人 間	機 械
高品質のKF (所要時間)	翻訳者の適性	カスタマイズ
入 力	0 時間	4 時間
誤 読 修 正	0 時間	10 時間
翻 訳	200時間	6 時間
後 編 集	?	?
訳 文 入 力	0～50時間	0 時間
合 計	200～250時間	20+後編集時間

(KF:キーファクタ)

人間翻訳の場合、翻訳の質は、やはり翻訳者の素質にかかっているが、機械の場合、カスタマイズがどこまで出来るにかかっている。勿論カスタマイズのみならず、辞書や文法といった基本的な問題もあるが…、

翻訳時間の推定

人間の場合、入力時間は0時間、つまり入力=翻訳で

ある。A4を100頁入力するのに機械の場合、3～4時間はかかる。OCRの誤読修正についてはパラメータをどのように考えるかによって時間も大きく変わる。読み取り率が何%かによっても大きく変わるが、一般的には100頁で10時間位ではないかと思われる。これが人手翻訳の場合、ざっと言って、200時間位はかかるだろう。

機械翻訳では6時間というのは5年前の機械であって、現在では1/3とか1/4になっているのではないかと思うが、これはSUN3の1MPS程度の計算機を使用した翻訳機を基準にしたものなので、現在では30MPSとか100MIPSくらいなんでもないので、もっと早く処理出来る。しかしこの図でも分かるように、機械利用での翻訳だけの時間は、昔も今も全工程に占める時間はそれほど多くはない。ここは仮にゼロになってもたいしたことではない。機械翻訳における翻訳時間など無視できるといっても過言ではない。問題は後編集時間で図2では??で表示してあるが、これがどれ位になるかが一番問題になる人間翻訳だと訳文入力というのがあり、これも翻訳会社によってはワープロを打ちながら翻訳しているから翻訳時間の中に入れるのか、どうか分からないが、所要時間としては0～50時間かかるとして計算してみた。50時間というのは、A4判100枚を打つのに、最大5ストローク/秒打てるというもの、これは短時間のことであり(プロのワープロ入力の方からそんなことはなく、長時間連続できると聞いたこともあるが)普通一般的にはA4サイズ1頁で800字として15分で打つのは可成り熟練が必要であり、またそれほど長時間持続出来る作業でもない。それで計算しても一時間で4頁、100頁では25時間かかる。これにインデントやタブ、あるいは同音語選択などの編集があり、通常では2倍かゝるとしても差し支えないのではないかと思う。

翻訳中は無人

機械翻訳の場合、機械が稼働する時間というのは20時間程度しかなく、その中での翻訳時間というのは無人であり、無視してもいい時間であり、残された問題はOCRの誤読修正が一番重要な問題である。現在のOCRは性能も向上し、入力時間も4時間もかかる

ないが、しかし誤読修正だけは未だ問題として残る。翻訳のトータル処理としては後編集がもっとも厄介な問題である。これを100時間でやるとすれば、つまり1頁1時間となるが、人手翻訳に比べ機械翻訳は2倍の効率が得られる。もっと、後編集の手間を減らすには、結局カスタマイズが何処までやられているかと云うことになる。

図3 効率改善ファクタ

入力から出力までの全体的所要時間を削減する

時間ファクタ	極小化ファクタ
原文入力	OCR精度
前編集と訂正	スペリングチェック
翻訳	翻訳速度
後編集	翻訳品質
	マンマシンインタフェース
	保守ツール
最終編集	DTP

また他人の書いた下手な文章を書き直す事は、ゼロから書き直すよりも難しい。どうしても書いた人の文章に引っ張られることが多いので、むしろ完全に忘れて、ゼロから出発するほうが時間的には早い。ところが折角、訳文があるのでそれを利用しようという意識が働き、かえって相当骨が折れる事もしばしば経験することである。

結局どこへ頼るか云々とカスタマイズしかない。

カスタマイズが決め手

E BMT(例文方式MT)も結局はカスタマイズである。例文方式といっても無限の例文を入れることは出来ないし、一般的な例文などは役に立たない。機械では人間のように創造性に頼った翻訳はできないので特定の分野に絞った例文に頼るしかない。如何に多くの文例を集め、分野を絞って如何にカスタマイズ出来るかということが最大の問題と言える。

E BMTには2通りあると思っている。一つは全部例文でやってしまうこと、それは大学等で検証して貰えば良いと思う。しかし物を作っているところではそのような方法はとれないわけで、結局、一

般的な辞書、文法を使ったシステムの上にカスタマイズした部分として例文を入れるという、しかも関係のない分野の例文をもって来てても使い物にならないのであって、結局自分の所の文体でカスタマイズしていくしか方法がない。しかしこれは結構いやらしいもので、つまり対訳が必要になるが、意外に対訳というものはない。

日本語から英語のマニュアルを作る場合、その日本語を英語に訳しているわけではない。マニュアルには国々の文化、発想があり、例えば、米国には米国の方式があるので、それに合わせたレベルで対応して行かなければならない。同じ製品の日本語のマニュアルとは全然対訳になっていない場合もある。

ユーザにMTを販売し、どのようにして例文を入れこむかという、本当のところ難しいのである。たまたまある企業には翻訳部みたいな部署があって、自社の、いくつもの文例があれば可能かもしれないが、一般ユーザ段階できめ細かくこれを実施していくことは至難ではないかと思う。

質疑応答

Q. 機械翻訳の辞書の特徴はなにか

A. 機械翻訳の辞書というのは、汎用辞書と異なり、ある特定の方向性を持っているので、あまり大規模なものとは出来ない。反面、記述や品詞分けの細分、格パターンなど非常に細かな情報が盛り込まれているものだと思う。

Q. 翻訳業での機械翻訳利用はあまり芳しくないが...

A. 翻訳会社ではユーザが多岐にわたり、何でもこなさなければならないから、カスタマイズを行い難いと思う、従ってある特定の企業、特定の部門での利用が一番カスタマイズしやすいのである。

Q. OCRに対する評価が厳しいようだが、

A. 日英の場合、ワープロが普及し、原稿もFDで受け取ることができるが、英日の場合、一般的には中々FDを手に入れることは難しい。DTPの場合、色々

なフォーマットがあって、それに対応していくのは大変だ。

Q. 翻訳の精度について

A. 英日の場合、使い捨ての需要がどれくらいあるのか、つまり読んで捨てるという類いの翻訳だが、この読み捨ての翻訳データの需要量は全く拮拠していない。残念ながら今のところそれほど多いとは思えない。この場合どれくらいの品質なら許されるのかも検討して行かねばならない。INTERNETやWorld Wide Webなどの情報網が完備し、英文の情報が簡単に入手できる時代になったが、英文を格安のPC翻訳ソフトで翻訳してみるかという所まで行っていない。低品質機械翻訳の翻訳文では先入観が先に立って翻訳文そのものが受け入れられないケースが多いのが現実である。機械翻訳にはやはり高品質ということが未だに期待され、使い捨て需要は少ないと思う。機械翻訳を導入し、その特性を知った上で利用しているユーザならば、その修正箇所も判断できるし、それなりに勉強もしている。しかし一般の人達はやはり完全翻訳を期待しているのだから。

Q. 翻訳品質率について

A. 品質が80%で我慢出来るかという聞き方で調査をやって、中々満足度は判断しにくい。実際に1頁2頁でも訳文を提示して、その満足度を調査するのでなければ、ただ単に%で調査されても判断しようがないだろう。

Q. 先程のOHPのデータはあくまでも英日のデータか

A. 英日の場合だが、日英でも同じ程度と考えていいのではないか。唯データそのものは大雑把なもので、原文の種類によって大きく揺れる事がある。日英の場合多少時間がかかるかも知れない。

Q. 通信ネットを介しての入力という将来見通しは

A. 動画像などはまだ回線容量の関係で難しいから、動画入りマニュアル等を送り、翻訳するというような新しいメリットは暫くはないのではないか。尤もISDNなど通信インフラが完備してくれば可能性もあるかもしれない。日本の高い通信費を考えればあまりメリットはない。

Q. 機械翻訳のオペレータもある程度、翻訳の資質を持たねばならない。翻訳にもドキュメンテーションと、抄録という目的があるのではないか。従ってオペレータも原文の解釈力や表現力、更に原文の書かれた環境や背景についても勉強しなければならない。

A. ジョイスの「ユリシーズ」という翻訳書があるがその誤訳を見つけだして、それを本にされたものがある。中には誤訳とはいえないものもあって、翻訳者の解釈が受け入れられないだけの話のもあるようである。実際にダブリンへ行って本の中に記述されている通りに歩いてみたら、その光景が翻訳と違うといっても、文章の曖昧性をいちいちその場所へ行って解決するなど、文学書では中々難しい。これはカルチャルリテラシーの問題で、言語上だけでは翻訳は出来ない。文学翻訳の場合は読者の想像性を引き立たせる事で翻訳の目的を果たせるが、マニュアルの場合はやはり事故にもつながりかねないので正確性が要求される。

《 K B & K S ' 9 5 》

Date April 10~13, 1995

Venue University of Twente, The Netherlands

Sponsored by JIPDEC and ECCAI

ホワイトカラーの知的生産性を考える

(財) オフィスオートメーション協会 常務理事 鈴木耀太郎

プロフィール

OA(オフィスオートメーション)協会は昭和56年に通産省の認可を得て設立、現在500社の会員を擁している。OA機器メーカーは勿論OAを推進しているユーザも数多く参加している。OA機器メーカーの意見も必要ではあるが、ユーザサイドのニーズが把握されている事が重要であり、双方の話し合いの場として、さらに教育や普及活動、海外での実態調査など、またワープロの普及促進のためのコンテスト、表彰などを行っている団体である。

1 ホワイトカラーの知的生産性とは仕事の価値を認めること

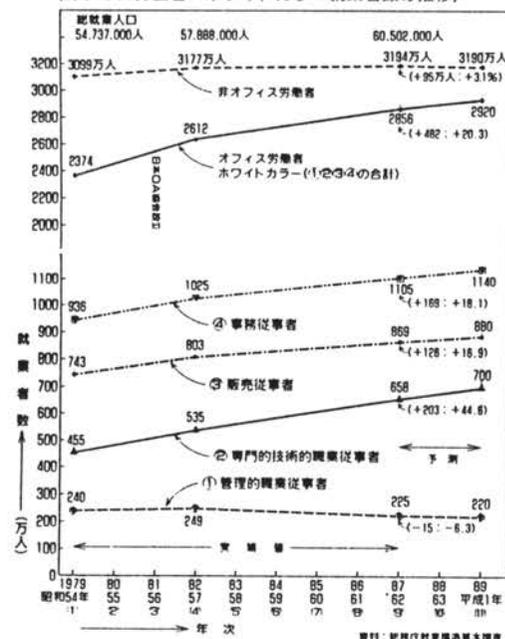
昭和55年にワードプロセッサが誕生したが、発売当初の価格は630万円というものであった。現在は10万円もしないものも多く出回っている。つまり価格は100分の1くらいになり、家庭用にも普及するようになった。エレクトロニクスの革命は価格が安くなるのが大きな特色といえる。その間賃金は決して下がることはない。明治以来100年間で1回も賃金は下がったことがない。これに対して機械の値段は100分の1に下がるという現象が起きている。これはやはりOA化、情報化が進んで来たもの、背景にある。この現象はパソコンも同じである。コンピュータが発売された当初の値段は2~3億はした。現在の大型コンピュータの価格はおよそ1億円であろうが、パソコンは100万円、つまり100分の1の価格であるにもかかわらず、能力的にはそれほど変わりはない。大型コンピュータは平行処理を行い、スピードが早いという利点はあるが、やはり同時に平行処理をやる機械は高い。パソコンは処理は遅くても同じくらいの能力を持っている。

米国ではコンピュータは現在6000万台稼働している。これに対して日本は1000万台しか普及していな

い。人口の比率からすれば米国は日本の2倍であるから、まだまだ日本での普及率は低い。また価格も日本の方が高いという問題もある。

これらハードの値段は時とともに下がっていくが人件費は下がったことがない。現在初任給は大学卒で20万位であろうが、10年前に比べればこれは5倍くらい上がったことになる。いわゆるホワイトカラーの人々がこれらの給与をもらっているが、これらの人々の生産性が10倍になったかという、あまり上がっていない。商品の原価は人件費と設備コストと材料との関係で決るが、この人件費の比率が極度に高くなっている。だから物価が下がらない。したがって最近では日本で生産するより人件費の安い海外で生産をする方がいいという現象が際立って

Knowledge Intensive Working 現象
(オフィス労働者：ホワイトカラー-就業者数の推移)



来ている。

生産の場合はそれでも良いが、サービス業とか人間がソフトを作る産業では海外生産ができない。したがって、これらの国内生産品の価格は高くなる。これを作っているのがホワイトカラーの人々である。ホワイトカラーの定義としては技術関係、管理職、販売職、事務職関係者などもこれに属するが、工場はもうコンピューターで動く時代となっているし、やっている仕事の内容は殆どホワイトカラーの仕事になっている。現在日本では6,000万人の就業者がいるが、その殆どの人がホワイトカラーの仕事をしているがその生産性は上がっていない。

スタンフォード大学で1974年に「工場と事務の生産性が極度に違う」という論文を発表した。

その数値は1960年から1969年の10年間に於いて、工場の生産性は約80%向上したが、事務の生産性は4%しか向上していないという内容である。日本ではどうか調べてみると、ほぼ同じ位という事がわかった。それがため生産性向上問題に取り組まなければならないということで、OA協会が設立された。

OA協会ではワープロコンテストを実施しているが、字を書くということは100年前も今も全く同じである。人間がいくら早く書いても50字/分しか字が書けない。これが2枚目、3枚目になっても早くならない。字を書くと言う点では、100年前からその生産性は変わっていない。ところがワープロコンテストでは早い人では150字/分であり、手で書く場合と比べれば、3倍早くなる。これはやはり機械を使うべきだと考えた。しかも出来上がったものは何回も繰り返し利用出来るし、修正も簡単にできるという利点もある。このワープロも今800万台から1000万台は普及している。パソコンと合計すれば2000万台近く普及していることになる。但し漢字ワープロと言うのは日本だけであり、外人からみれば不思議な商品の一つである。つまり押したキーと表示される字が違う(変換/次候補機能で次々違う単語が出てくる)からである。外国のタイプライターは打ったキーと表示される字は全く同じである。

このように機械のスピードと値段で普及が進展したが、しかしまだまだ生産性が上がったとは云いがたい。ホワイトカラーの知的生産性はソフト開発にしても第三者からはなかなか認めてもらい難い。し

たがって価値についての考え方を根本的に検討しないと、本当の意味での生産性の論議が出来ない。

つまり顧客がお金を出し、商品なりサービスを買うわけであるが、その時に品質なのか、時間なのか何にお金を払ったのかが分からないと知的な価値について判断出来ない。情報の価値というものの中々分からない。ユニークなものを作っても使う人にとっては価値は高いが、使わない人にとっては価値がないという性格をもっている。また情報はいくら使っても減らないという性格をもっている。コピーなんかすぐ出来るという性質もある。したがって絶対価値を測ることが難しいのである。単行本の場合でも編集費、印刷費、配送費などの原価の積み上げで、中々その執筆内容にたいする価値が判断し難い。コンピューターソフトの場合でもステップなど行で書いて行くが、行の多いほうがステップ数が多くなり、コストがかかるが、この命令言語は短い程早く処理出来るので、短いほうがいいのだが、長く書かないとお金が取れないので長く書く傾向がある。それだけに価値というものは曖昧で、これが無料と見做されるケースも多い。

海外では海賊版の書籍が多く出回っているが、印刷コストのみで、情報価値は全く計算されない。このように知的生産価値を、どのように認めていくか重要で、先ずその価値を認め合うようにならないと知的生産物の生産性というものが明確にならない。たゞその価値は測りたいものであるから、原価主義でやるというのが実情である。ただインターネットという世界通信ネットでは1回入ると誰でも使えるようになっている。ソフトの作り方などを入力すると誰でも使っている。知的所有権についての制限問題も難かしいものがある。実際経済活動や企業の中で仕事をしている人には賃金は支払われているのだから、何を価値として評価し、賃金を支払っているのかを測って行かないと、生産性向上ということが分りにくい。

現在リエンジニアリングということが叫ばれているが、天下のIBMでも40万人いた従業員が22万5千人に減っている。米国には失業保険はあるが公的健康保健制度は存在しない。失業すると医者にもかゝれない状況に陥る。米国では企業が保険会社医療保険をかけているが、IBMの場合解雇後も

5年間だけは企業が医療保険を負担している。米国でも人減らさばやりで、減らすと生産性が上がるといふ皮肉な現象が見受けられる。

10/17のBusiness Weekに「Be thinking One」という記事が出ていたが、これは知的業務の見直しをしようというもので、「知恵を使った人の失業はないが、知恵を使わない人の失業は増えている。また知恵を使った人でも価値のある知恵かどうか見極め、ホワイトカラーの生産性というものを厳しく考えて行こう」としている。

10/3のFoutuneには「Intelectural Capital」という記事で、「会社の中に知的資産がどれくらいあるかを測って、有価証券報告書に記載すべきだ、これのないものは投資価値がないと考えるべきだ」とある。「それぞれの会社が価値の自己評価をしなければならない」と論じている。このように知的資産を厳しく見て行こうという動きがある。

2 知的生産性は測ることから始める

知的生産性はやはり測らないと分からない。オフィスにおいて1人の人間が行う仕事を、時間的

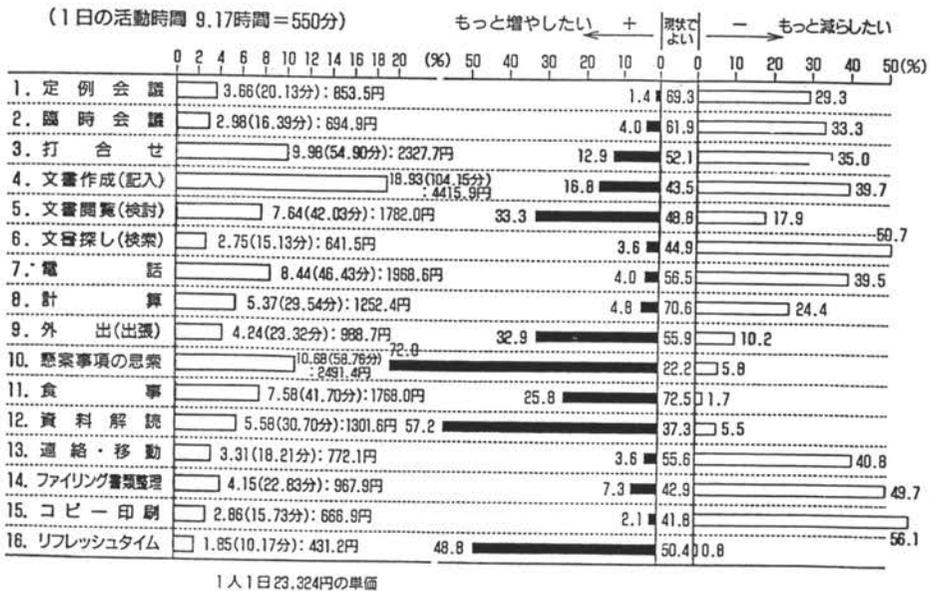
に分析した調査内容であるが、その業務の内、やりたいもの、やりたくないものも調べた。1日の内、やはり会議とか打ち合わせなどの時間が長く、30%近くを占める。1年間で会議で支払われた人件費は平均で83万円にも達する。

文書作成のコストは147万円、電話をかけている時間は42万円、考えている時間というのはコストでいうと53万5千円、ファイリングが20万円となる。これらが本当に生産性向上に繋がっているかという問題もある。これはやはり測らないと判らないことである。色々な業務処理が多すぎる、考える時間が少なすぎる、もっと考える時間を増やさない仕事と生産性をあげることは出来ない。米国でMajorling White-colourという論文がでていて、職種別に投入、産出ということで生産性を測っている。

3 ホワイトカラーの生産性を阻害するのは知識の陳腐化である

ホワイトカラーはやっている仕事はわかるが、何をもって価値を評価しているのか判らない。その知

オフィスにおける1日の活動時間と増減希望 (243人の回答)



的価値を上げようとする、情報や知識がないと仕事ができないという性質がある。情報を糧にして、また情報を発信しているのであるが、持っている知識や情報は陳腐化していくので、新しい情報を常に吸収していかないと劣化してしまう。情報を常にリフレッシュしているかという、必ずしもそうではない。データベースなど3年前のものしか入力されていないものも結構多い。国際為替相場担当など瞬時の情報を瞬時に処理して行かないと仕事にならない。いずれにしても情報が陳腐化すると、新しい情報は生まれず、また生産性も上がらないのである。

ドロッカーは知的生産性について「No Limits to Learning 勉強には限りがない」と成長の限界という書物の中で述べているが、知的業務をするには、常に新しい知識を用意しておかないと仕事にならないと指摘している。経済の成長に陰りが見えてきたときに問題になるのはやはり限りのない学習でありそれが知的価値を見い出だすことになると云っている。

4 ホワイトカラーの生産性向上手段はITを使いプロセスを短縮することである

情報技術とか情報機器は価格が下がって行くが、これら機器を使わないでホワイトカラーが無手勝流で生産性を上げようと思っても無理な話である。同じ価値のものを見い出すのであるならば結局量を増やすというのが1つの方法である。最近マルチウインドウなどが発売されているが、昔は命令言語を書かないと処理出来なかった。しかしこれらマルチウインドウを享受している老若男女は案外少ない。

昨秋Comdexに行った。ラスベガスにコンピューターのスーパーマーケットがある。米国では若い人より年寄りが買いに来ている。日本とは一寸違う現象である。使わないとその中身は判らない。日本では中間管理層があまりコンピューターを使ってくれないが、使わないとその良さも判らない。

昔は中央に大型コンピュータがあり、その下に端末がぶら下がっていて、ボタンを押すだけで利用できたが、現在はパソコンのようにある程度複雑な操作を必要とする機械があたえられるようになった。これで技術的に、分散、処理されるようになってきてい

る。一般的に道具がばらまかれ、ネットワークで結ばれ、その処理が世界中で同時に出来るようになっていくが、これを使いこなせる人は少ないのが現状である。しかしホワイトカラーの生産性を上げるといふのであれば、これを使いこなして行かねばならない。

業務のプロセスを短縮するというのであれば業務プロセスを改善すべきである。当たり前のことだが、顧客がテレビ買う時にそのコストの中身は何なのか材料費であり、加工費であり、物流費であるのであればその対価を払わねばならないが、テレビを作る会社の財務計算や、人事管理するコストにまで顧客は代金を支払う必要はないと思う。これらの経費は出来るだけ少ない方が良いのである。

日本企業の直間比率は昨年までは51%が間接経費である。ホワイトカラーの件数で、しかもコストが高い、顧客は何に対してお金を払っているのかという疑問が起こってくる。これが価格破壊の根源になっている。米国ではガソリンは25円/リッターである。日本は100円を越えるのが現状である。これは税制の問題で多少価格が高いというのはわかるが、米国では電気料金も、住宅も、食料も日本の半分の価格である。日米の生産性を購買力平価で、時間当たり単価で比較してみても、日本は米国の61.7%しか生産性がないと考えられる。つまり得た所得で買えるものの量と言うものは米国の61.7%しかない。米国と同じ所得を得たとしても日本では半分のものしか買えない。つまり如何に物が高いかということである。米国が間接費を落とした商品を提供してくると価格破壊が起こってしまうのも当然である。日本のものは全部高いという計算になってしまう。この原因は人件費、あるいは土地が高いということにある。空洞化し、安い人件費のところを物を生産するとすれば、日本は物を作らない仕事になってしまう。したがって知的生産性を上げない限り生きて行くことが出来なくなる。知的生産の価値を高めるには色々な方法があるが、不要な贅肉を落とすことも1つの方策だし、不要なプロセスを短縮すると、プロセスは半減するのである。

例えば山ノ手線はステンレスの車体の電車が走っているが、価格半分、寿命半分という電車である。従来の電車は20年は使えたが、これでは世の中の変

化に対応していけない、また20年間の補修部品の保有も必要で、コストがかかるものになる。電車といえども10年で使い捨てにする、そのかわり価格を半分にし、補修部品を持たないような電車がは走っている。これがためガラスは埋め込みで、窓は開かなくなり、またメンテナンスフリーの電車になっている。このように大幅に価格を下げるというのも1つの方法である。ところが物作りが海外へ出ていくと、日本で作るものは知恵の価値の認められないものばかりになってしまい、国内での生産は非常に難しくなる。

コンピューターもOSはすべて米国に依存している。マイクロプロセッサも米国依存であり、パソコンはこれらパーツを買って来たもので組み立てられている。日本で出来ない理由についてマイクロソフト社で聞いた話であるが、「日本語を使う者は日本人だけで、1億2千万人のためにソフトをつくることは出来ない、英語を使う人間は世界で7億5千万人もおり、その内10%の人がソフトを使うとすれば7500万人が使ってくれることになる。従って小さな日本人マーケットのためにソフトを開発は出来ない」とのことである。7500万を対象にしたコンピュータのOSが日本に入ってくると、これに合わせて作るしか方法がない。世界マーケットはマイクロソフト

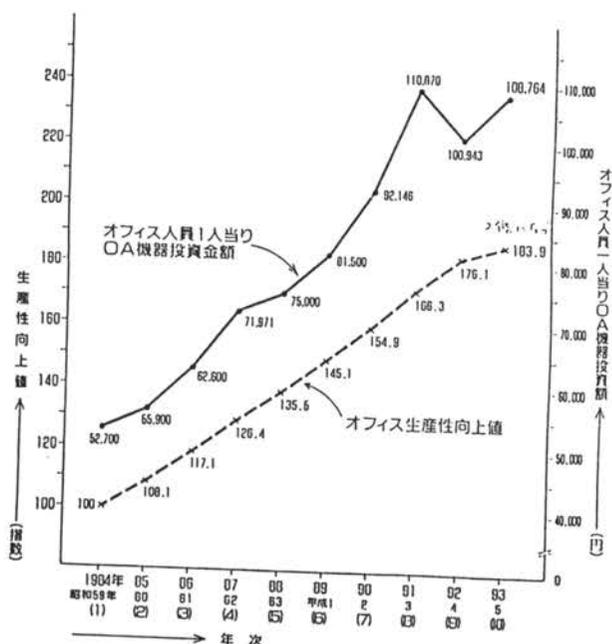
に押さえられてしまうのかということとほぼその傾向にあると言える。日本の技術は世界的に進んで来たというけれども、日本の優位、特性があるのではないかと思われるが、中々これがないのである。このような中で日本が知恵を価値にして作って行く力を持たなければならない。

5. ホワイトカラーの生産性向上は 限りなき学習にある

ホワイトカラーの仕事は難しいが、このようにのを作り出す創造力がなければならない。ところが学校教育とパソコンは同じなのである。つまり塾で子供を教育しているのは知恵を詰め込むことばかりやっている。いかに覚え込んだか、覚えた事だけをチェックしている。メモリが大きければいくらかでも知識は増えて行くが、覚えることばかりで育てられて来た人間に考えるようにしろと云っても難しい。このあたりから知的生産性の源泉という問題が出て来るわけで、OSが日本で出来ない、マーケットも小さいという問題をかゝっている。これをなんとか突破しないと、知的価値も出ないし、知的生産性も上がらない。これがため現在皆で知恵を絞っているのである。

OA機器投資金額とオフィス生産性向上値の推移
(オフィス人員1人当たり)

	対前年生産性 向上値(年率)	オフィス人員1人当り OA機器投資金額(円)
昭和59年	8.4%	52,700円
60年	8.1%	55,900円
61年	8.4%	62,600円
62年	7.9%	71,971円
63年	7.2%	75,000円
平成1年	7.1%	81,500円
2年	6.7%	92,146円
3年	7.4%	110,870円
4年	5.8%	100,943円
5年	4.5%	108,764円
平均 X	7.15%	81,239円



もうOSはあきらめてアプリケーションで頑張るしか方法がないということで、その1つとしての機械翻訳ソフトなのかもしれないが、このシステムが本格化する時代は来るのではないかと思う。やはり人手でやっていくとコストがかかり過ぎて駄目だということになると、MTなどで安く知的生産性を高めて行くことが必要になる。

ITが今後どこまで進歩するかによって、日本だけのマーケットで考えるのではなく国際間での日本語の文化というものを全く英語と同じように伝えるようにして行かねばならない。

日本ではこの10年間で500万人もホワイトカラーが増えている。これは1万人の会社が500社増えた勘定になる。新しい就業者の増え方どころではない。

OA化投資と生産性についてはこの10年で年率7%の生産性が上がってはいるが、1人当たり70万円という投資が見合うのか否か意見の分かれる所であるが、私見としては、大体見合ってお釣りがくる程度という感じである。ただ価格的に安いものが出てくると現行商品の値段もそうは上げられず、そのまま賃金が上がって行くと困った問題となる。これらは現在各社頭を抱えている問題でもある。しかし終身雇用という伝統的経営の中で米国のようにすぐレイオフとは言えないという問題もある。雇用か賃金かということろへぶつかる。雇用を確保するためには賃金をカットするしかないというワークシェアリング的な考え方も中々難しい。フォルクスワーゲンでは雇用確保のため週3日制にしたが、それがため賃金が下がってしまった。おかげで生産性も上がらない。つまりモチベーションが下がってしまったのである。

最近米国ではバーチャルコーポレーションという考え方が芽生えている。これは仮想企業とでも訳すべ

きなのか分からないが、初めてファーストバーチャルコーポレーションという会社が出現した。従業員は20人しかおらず、1000分の仕事をするという会社である。これはパソコン通信で社員が繋がっており1つの仕事が入ると専門家によるプロジェクトのようなグループが結成され、一斉にその仕事をこなし仕事が終了するとそのプロジェクトは解散するというシステムで、実際にコーディネートしている人は20人しかいないという企業である。だから仮想企業ということになっているのであるが、このような動きが米国では出て来ている。会社に属してはいないが、所得を得る事が出来るようなシステムである。設備もいない、人間の知恵だけで仕事をするとなれば、またバーチャルコーポレーションという存在も必然性をおびてくる。知恵を使った仕事の価値というものを認めると組織のどこに属しているかということはあまり意味を持たなくなる。このような仕事のやり方をして行くからには、限らない学習をやって行かないと知識が陳腐化し、仕事に有りつけなくなる。従って常に最新情報に触れていかなければならないし、自分の考え方を新しくしておかねばならないし、それによって生産性も上がるし、価値も認められる。いずれ日本もそうならざるを得ないし、これらを履行していきながらコストのやすい地域との価格差を詰めていかねばならない。米国の多くの会社が知的価値を極度に認めようとすると、知恵もっている人の汎用的な使い方が出来るのである。このような事態になってくると言葉も合理的に機械で処理して使えるようにしないと、人手がかかっていたのではコストがかかり過ぎて耐えられないということが日本でも起こってくるのではないかと思う。

タイ王国 第2回 自然言語処理国際会議

日 時 1995年8月2～4日

場 所 バンコック市 ハイアット セントラル プラザ

主 催 タイ王国科学技術環境省国立電子コンピュータ技術センター

照会先 Asanee Kawtrakul, Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Bangkok, 10900, Thailand (Tel. 662-579-4584 Ext. 182 FAX 662-579-6245)

N H K 放送技術研究所

江原 暉 将

1. まえがき

NHK放送技術研究所は昭和5年に設立され、以後、放送に係わる多くの研究を行ってきた。放送に「ことば」は不可欠であることから、当所でも、自然言語処理の研究を以前から行っている。かな漢字変換の研究から開始し、テキストや画像データベースの検索などを経て、現在は、機械翻訳を中心に研究している。

本文では、当所での機械翻訳に関する最近の研究から、いくつかを紹介しよう。

放送企業における翻訳の特徴として、

- ・テキストの分野が多岐に亘る。
- ・迅速かつ正確な翻訳が要求される。
- ・新語や固有名詞が頻出する。

などがあげられ、いずれも現在の機械翻訳では対応が困難な特徴である。

2. 基本動詞の訳語選択

機械翻訳の精度を悪くしている大きな要因は構文解析と訳語選択の誤りである。例えば、

I saw a girl with an umbrella.

は、少なくとも以下の4つの訳を持つ。

- a 私は傘を持った少女を見た。
- b 私は傘で少女を見た。
- c 私は傘を持った少女をのこ引きする。
- d 私は傘で少女をのこ引きする。

このうち、通常の解釈での正解はaである。この場合、bとdは前置詞句の係先に関する構文解析に誤りがあり、cとdはsawの訳語選択に誤りがある。

このように訳語選択は機械翻訳の大きな研究課題である。特に、放送分野は多岐に亘るため、適切な訳語選択が一層困難となる。この中で、makeやtakeなどの基本動詞は様々な意味で用いられ、出現頻度も高いため、その訳語を適切に選択することが翻訳精度の向上に大きく寄与する。このため、当所では英語基本動詞の訳語選択の研究を行っている。

基本動詞は、目的語として動作名詞を伴うこと



図1 機械翻訳を用いた日本語字幕作成例

がある。このような場合、例えば、

make an address

という表現では、makeは本来の意味である「作る」ではなく、Delexical Verbとして、「する」という意味になる。このような点に注目して、基本動詞の訳語選択に目的語名詞の動作性の有無を利用する方法を考案した[田中]。その結果、上の例は「演説を作る」ではなく、「演説をする」と正しく訳される。

動作名詞との共起に限らず、そもそも訳語選択のためには、訳語を決定したい語のみに注目していたのでは不可能であり、その語がどのような語と共起しているかなど、広い意味での文脈を利用することが不可欠である。しかし、文脈を利用すると、訳語選択のための規則集合が大きくなり、そのような規則集合を手で作成・管理することが困難となる。そのため、規則集合を翻訳対象テキストから自動的に学習する方法の研究を行っている[田中2]。

自動学習を行う元のテキストとしては基本動詞を含む英文とそれを日本語に翻訳した対訳テキストである。ここから、基本動詞の訳語選択を行うための決定木を学習する。このような自動学習によって訳語選択のための大規模な規則集合が効率的に作成できるだけでなく、従来、人間では気がつかなかった規則も得られることが分かった。

3. 英日機械翻訳での固有名詞処理

固有名詞を含む表現の翻訳精度を向上させるため当所の英日機械翻訳システムは、構文解析に先だっ

て、局所翻訳処理を実施している〔加藤〕。

固有名詞を含む表現は、未知語を含むことが多くまた複合語としての特殊な表現が多いために、通常の構文解析になじまないからである。この局所翻訳処理は精密な非終端記号を持った文脈自由文法を用いており、解析と同時に翻訳を行う点が特徴である。例えば、肩書き付きの人名として

Atlantic Council Vice President Job
Dittberner という表現は
未知語 COUNCIL VICE-PRESIDENT 未知語

のように解析され、同時に

Job Dittberner Atlantic 会議副議長
と翻訳される。また、President の訳語は議長、大統領、社長など多くあるが、Council と結び付くことで議長が選択される。このように、局所翻訳によって、きめ細かな訳語選択と未知語への対応が可能になった。

4. 自動短文分割と主語の補完

放送ニュース文には極めて長い文がある。例えば「通信所では、郵政省の免許がおりしたい、インテルサットの子備衛星を使って、埼玉県にあるKDD上福岡研究所との間で電話やFAX通信を中心におよそ2年間送受信実験を行い、実用化にこぎつきたいとしています。」などの例がある。このように長い文をそのまま、日英翻訳すると、誤訳になる可能性が高い。そこで、長いニュース文を自動的に短い文に分割する自動短文分割の研究を行っている〔金〕これは、パターンとアクションの組で記述された分割規則を用いるものであり、分割の結果、上の例は「通信所では、郵政省の免許がおりしたい、インテルサットの子備衛星を使います。」「埼玉県にあるKDD上福岡研究所との間で電話やFAX通信を中心におよそ2年間送受信実験を行います。」「実用化にこぎつきたいとしています。」と3文に分割される。

この例からも分かるように、分割を行うと主語の無くなる文が生ずる。このような文に対して、自動的に主語を補完する手法の研究も行っている。これは、主語を補完すべき述語の性質や、主語候補となる名詞句の性質を統計的に捉える方法であり、上の例では、「通信所」が分割後の第2文と第3文の主語として補完される。

5. 字幕作成システム

NHKでは、英日機械翻訳システムを使用して、一部の英語ニュースに日本語字幕を付与している。これは、衛星第1放送で毎平日午後5時20分から始まる機械翻訳ニュースの中で放送されている。字幕付与作業は、英語ニュースの聞き取り入力、機械翻訳、後編集による字幕作成、映像への字幕スーパー、の手順で日本人翻訳者によって行われる。現在の我々のシステムでは、翻訳者の作業能率の改善には至っておらず、あくまで、実験的な利用に留まっているが、当所での研究成果は随時現場のシステムに反映され、正翻訳率が向上していることも事実である。こうして出来た日本語字幕の例を上図に示す。

字幕翻訳作業では、英語ニュースを聞き取り入力しなければならないというネックがあったが、近年米国のテレビ番組に聴力障害者用の英語字幕データが付加された。この字幕データをパソコンに取り込むことによって入力作業の部分を大幅に改善できた。

6. あとがき

NHKでの機械翻訳の研究と実用化の現状について述べた。地球規模での環境の悪化や、民族紛争の激化が進む今日、世界規模で、様々なレベルのコミュニケーションを拡大することが要請されている。衛星による国境を越えたテレビ放送や、インターネットを介したコミュニケーションはその例であろう。しかし、ここに立ちはだかるのが言語の壁であり、機械翻訳は21世紀に向けて、重要な技術であると確信している。まだまだ課題の多い機械翻訳ではあるが、あらゆる場面を捉えて研究・開発を促進し、異言語間コミュニケーションの道具として、広く用いられるようになることを期待したい。

参考文献

〔加藤〕加藤直人ほか：英日機械翻訳における固有名詞処理、40回情全大、Mar. 1990.

〔金〕金淵培ほか：日英機械翻訳のための日本語長文自動短文分割と主語の補完、情学論、June 1994.

〔田中〕田中英輝ほか：基本動詞と動作名詞の組合せ表現の英日機械翻訳手法、情学論、Feb. 1993.

〔田中2〕田中英輝：意味コード付き対訳データからの訳し分け情報の自動学習、49回情全大、Sept 1994.

OCRによる文書テキストの読み取り

星 沢 由 明

盤 東 芝 情報通信システム研究所 第3研究所

今回は機械翻訳システムの入力部に必要となる重要な機能の1つである文書テキストの自動読取の技術について紹介します。

この技術には英文テキストリーダと日本語テキストリーダがありますが、この両者にはほぼ同等な技術が使われています。ここではこれらを区別することなく説明して行きます。

テキストリーダは雑誌・書籍・マニュアル等の一般文書を自動的に計算機に入力するためのものであって、レイアウト理解機能とオムニフォント文字認識機能を備えたものです。レイアウト理解機能とは様々なレイアウトの文書から文字列を抽出し、正しい読み順を判定する機能のことであり、オムニフォント文字認識機能とは様々なフォントで印字された文字を高い精度で読み取る機能のことです。

1. はじめに

10年以上前から、印刷物の紙面を計算機に自動的に理解させる文書画像理解技術が多くの研究者によって研究されてきましたが、数年前の段階では実用化の段階には至っていませんでした。ところが近年になって、ニーズおよびシーズの両面から、このような文書画像理解技術を実用化しようという気運が高まり、次々と製品化がなされるようになってきました。

ニーズとしては、以下の点が挙げられます。すなわち、新聞、雑誌の記事やマニュアル等の印刷物を対象としたデータベース構築、機械翻訳、電子出版などの文書処理が活発になってきたのが第一点です。このような文書処理の活発化は、これまで人間が手作業で行ってきたキーボードによるコンピュータ入力作業の自動化を促すようになってきました。

一方、シーズとしては、CPUの高速化・メモリの大容量化に代表されるハードウェア技術の急速な進歩と、画像処理技術・パターン認識技術の進歩などのソフトウェア面での進歩が挙げられます。これ

ら両者があいまってテキストリーダの技術基盤を支えています。

以上のような状況の中で英文テキストリーダや日本語テキストリーダが開発されてきました。

2. テキストリーダの特徴

従来から帳票（伝票）入力作業の自動化には汎用OCR（光学的文字認識装置）と呼ばれる装置が用いられてきました。しかしこのOCRの場合、取り扱える帳票の書式（フォーマット）、印刷文字の書体（フォント）に大きな制限がありました。

すなわち、帳票には文字間ピッチや行間ピッチが固定した単純なフォーマットが要求され、文字フォントは予め定められた1種類のみでなくてはなりません。このようなOCRは特定の用途に限定して用いられていたのです。一方、一般の印刷物は文書フォーマットも文字フォントも多様であり、これらを限定したり特定したりすることはできません。ですからこれらをオペレータの介入無しに自動的に取り扱うことは不可能でした。

これに対してテキストリーダは、以下のような非常に融通性のある特徴をもっています。

(1) レイアウト理解

図表を含み、文字列が多段に組まれた文書から文字列部分を自動的に抽出し、その構造をも理解することができます。これにより、複数の記事から構成されるような複雑な文書を自動的に取り扱うことが出来ます。

(2) 高度な文字切り出し

一般の文書では文字間隔は不定であることが普通であり、文字間の接触も非常に多いのですが、テキストリーダは、文字行から文字を正しく切り出す機能を持っています。さらに図1に示すような特殊な文字行も自動的に判定して読み取ることが出来ます。

- (a) **RICHARD LANDRY**
- (b) If you can offer
- (c) Too
- (d) **W**ith economic growth expected to continue, the focus shifts to inflation. But here, too, the key is moderation. Productivity rebounded in

図1 特殊な文字行の例

(3) オムニフォント認識

タイプライター、レーザープリンター、写植印刷等の任意のフォントを高精度で読み取ることができま

(4) 高速処理

英文テキストリーダではソフトウェアのみでも毎秒数十文字程度の認識速度がほぼ常識になってきています。一方、日本語テキストリーダでもソフト版が商品化されていますが、高速・高精度を要求する場合には特別なハードウェアが必要となると考えられます。

3. テキストリーダにおける文書画像処理

テキストリーダの処理手順の一例を図2に示します。以下で順を追って説明していきます。

3.1 画像の取り込み

まず、印刷物のページ全体をスキャナにより光学的に読み取ります。

3.2 写真・図表領域の除去

次に、文書画像データにおいて、横方向に近接する黒画素を融合します(黒画素融合処理)。次いで、黒画素連結成分(黒のかたまり)を抽出し、それを取り囲む最小矩形を求めます。文字列領域における黒画素連結成分の多くは黒画素融合処理により文字

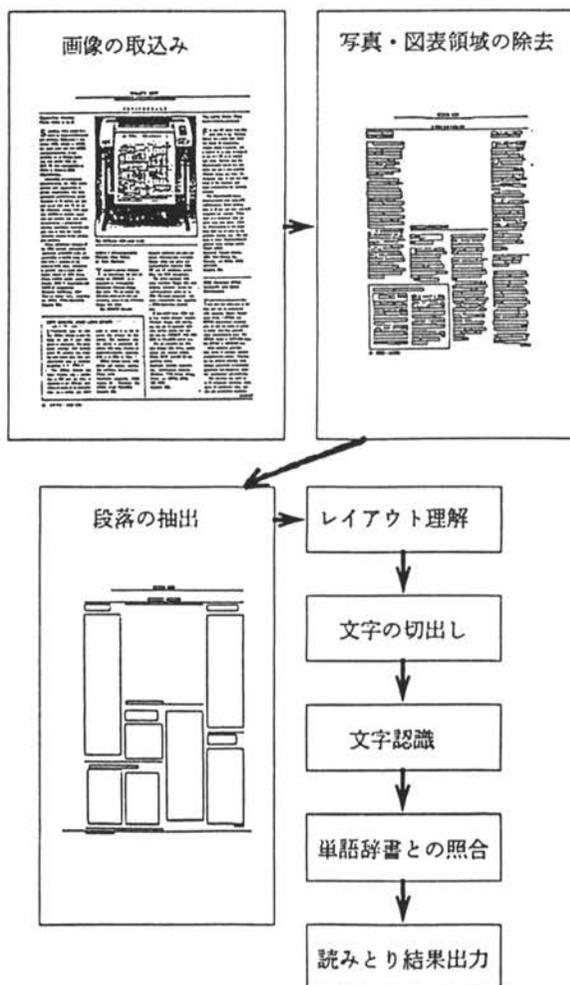


図2 処理の流れ

間の白領域が塗りつぶされるので、この黒画素連結成分を囲む最小矩形は単語を代表していると考えられます。

次に連結成分矩形の形状、及びその中の黒画素分布の複雑さを調べ、各連結成分を、文字列、図表・写真、水平・垂直線、ノイズに分類します。

なお、黒画素分布の複雑さは黒画素融合処理が実行された回数により簡略的に求めることができます。横方向に延びる細長い連結成分の場合、それが文字列であるかあるいは傾いた水平線であるかはその矩形の形状だけでは区別が付きませんが、黒画素分布の複雑さを調べるにより、正確に分類すること

Toshiba's Powerful, Compact Laser Printer Now Shipping

New York City, Oct. 10, 1989 — Toshiba America's Computer Systems Division is now shipping the PageLaser 6, a low cost, 6 page-per-minute, general business laser printer with a suggested retail price of \$1,899. It will be on display at Info '89 in booth 405 at the Jacob K. Jacobs Convention Center, New York City, Oct. 10-13, 1989.

"The market's immediate reaction to this product has surpassed our expectations. Users quickly realize that the PageLaser 6 delivers the powerful, total printing solution they have been seeking at a very affordable price," noted Mario Diaz, product manager.

"Featuring a space-saving, compact design, and weighing 35 pounds, the PageLaser 6 is designed for general office applications where quick and quiet high-quality laser printing is desired. To serve the office user's needs for a total solution, the PageLaser 6 offers a clean, straight-through paper path, sacrificed in some other recent laser printer introductions. This is essential for printing on many envelopes, labels and other special stocks," Diaz said.

"With HP LaserJet Series II emulation and two HP LaserJet compatible font cartridge slots, this printer offers the ultimate in HP LaserJet Series II compatibility — software applications written for the HP LaserJet Series II as well as HP font cartridges and soft fonts are compatible with the PageLaser 6," according to Diaz.

"In addition to its industry standard HP LaserJet Series II emulation, the PageLaser 6 includes an IBM Proprinter XL24 emulation to further maximize software compatibility and ensure turn-key operation in more environments without the need to purchase expensive options. As a result, end-users who have invested in software applications and fonts can save the cost of buying additional ones for PageLaser 6," Diaz noted.

Resident fonts include Courier,

Courier Bold, Prestige Elite and Line Printer. All fonts are HP LaserJet Series II compatible and are provided in portrait and landscape orientations. The printer supports both Centronics parallel and RS-232 serial interfaces as standard.

The printer has a suggested average usage rate/duty cycle of 4,000 pages per month and prints letter-size pages at 4 pages per minute. Memory of 512KB and resolution of 300 x 300 dots per inch are standard. Input and output trays have a capacity of 150 sheets. The input tray is fully adjustable from postcard to legal size paper, and supports a variety of special stocks, including envelopes.

Another special feature of the printer is its expandability. Op-

tions to expand the printer's memory area available in 1, 2 and 4MB upgrade kits, providing an upgrade path for end-users who require enhanced capabilities after their initial purchase, according to Diaz.

An available option designed for further flexibility is a detachable tray for face-up printing. This allows better stacking when using envelopes or heavier media with the straight paper path.

"Designed with easily accessible drum and toner components, the printer uses separate consumables. This contributes to a low operating cost for the unit, since the drum does not have to be discarded when the toner has to be replenished," Diaz stated.

"Engineered with a life of 300,000 pages, the PageLaser 6 will last almost twice as long as most printers in its class," Diaz said.

Toshiba America Information Systems, Inc., Computer Systems Division, 9740 Irvine Blvd., Irvine, Calif. 92718 (800) 457-7777.

Toshiba Debuts Desktop PC in U.S. Market

New York City, Oct. 10, 1989 — Toshiba America's Computer Systems Division today introduced the 80386-based T8500, a high performance, small footprint desktop computer — its first system in an emerging family of IBM-compatible desktop systems. It will be displayed at Info '89, Booth 405 at the Jacob K. Jacobs Convention Center today through Friday.

Toshiba, a leader in the portable PC arena, utilized its high-performance, miniaturized technology in the packaging and design of the T8500, on display in Booth 405, Jacob Javits Convention Center, site of INFO '89.

"As an industry leader, Toshiba

is committed to providing 'systems-oriented' solutions for diverse computing environments. This introduction responds to the need for industry compatible products that address a full range of high performance capabilities — expansion, networking, high performance, management and multiple operating system support," said Tom Sherrard, Director of PC marketing.

"We will continue Toshiba's leadership role in the portable market, as well as enhance our research and development to include additional products that complement our portables," said Sherrard.

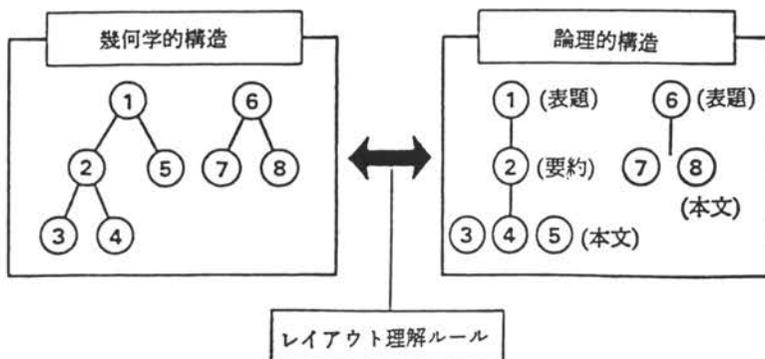


図3 レイアウト理解の枠組み

ができます。

3.3 段落の抽出

次に文字列を代表する矩形について、垂直方向にこれらを重ねあわせることより段組を決定します。さらに段組内の文字列矩形を統合して文字行を抽出します。そして、各段組において垂直方向に近接する文字行をブロック（段落）としてまとめます。

3.4 レイアウト理解

上記で求めたブロック間の論理的な関係をレイアウト理解機能により求めます。これにより、各ブロックには標題、要約、本文、あるいはヘッダー、フッター、キャプション等の属性が付加されます。

レイアウト理解機能は次の方式で実現されます。

(1) 文書構造の木表現

一般に文書のレイアウトは中身を読まなくても、一見ただけでその論理的構造を抽出できるように構成されています。ここでいう論理的構造とは、記事としてのまとまりや、その中のタイトルやパラグラフ間の論理的な相互関係を記述するものであり、図3に示すような木構造により表現することができます。

1つの記事は1つの木に対応し、タイトルやパラグラフは木のノードに対応します。また、枝は各パラグラフやタイトルのまとまり方に関する従属関係を記述しています。

これに対し、画像処理手段により文書の幾何学的構造を作成することが出来ます。

幾何学的構造は論理的構造と同様に図3に示すような木構造によって表現されます。

この木において各ノードは同一段組に位置する文字行のかたまりを示し、枝は各段組の位置関係を示しています。

(2) 木の交換によるレイアウト理解の実現

ここで、「レイアウトを理解する事」を「論理的構造木を抽出する事」と定式化します。

レイアウトに関する様々な知識、例えば、「左上から右下の方向に読み順がある」

などの知識をルールとして幾何学的構造木に作用させて交換することにより論理的構造木を獲得することができます。この様にレイアウトに関する一般的

知識を「木構造変更のルール」として表現するというのはテキストリーダ開発で出てきた新しいアイデアです。

3.5 文字切り出し

各文字行においてはまず黒画素連結成分（コンポーネント）を抽出します。

この時、コンポーネントは多くの場合1文字を代表していますが、接触文字等の存在により複数の文字を代表している可能性もあるわけです。そこで、まず、コンポーネントを1文字と仮定して文字認識を行いません。その時、その認識結果が受け入れられれば、それをそのコンポーネントの認識結果とします。

もし受け入れられなければ接触文字とみなして、すべての認識結果が受け入れられるようにコンポーネントを分割します。

3.6 オムニフォント文字認識

一般の印刷物に用いられている活字の種類には多種多様なものがあります。

その原因の第一は文字のフォントで、多くの種類が知られています。

現在アメリカには1500種類に及ぶフォントがあり、ヨーロッパ全体では5000種類に及ぶといわれています。日本にも相当数のフォントが流通しているものと思われます。フォントの主な系統として、英文ではローマン、イタリック、ゴシック等があり、それらは更に多様なフォントに細分化されています。

活字の種類が多くなる原因の第二はフォントの変形です。要するに、同じフォントでもサイズ、縦横比、線の太さにそれぞれ多数の変形があると言う事です。

この様な多様な活字の種類で印刷された文字を読み取る技術がオムニフォント認識技術です。これはマルチフォント認識とも言われます。

次に文字認識の代表例として、複合類似度法をとりあげてみましょう。

この方式は低品質シングルフォント活字認識をターゲットとして1970年頃に開発されたもので、その後、通常のOCRや音声認識に応用されました。この方式をオムニフォント認識に対応させるために

シンガポールの機械翻訳研究

I S S (シンガポール国立大学システム科学研究院) 董 振 東

はじめに

シンガポールにおける機械翻訳の研究・開発はいくつかの恵まれた条件に支えられている。すなわち

1. シンガポールは多民族、多言語国家であり、公用語の英語の他に中国語 (Mandarin)、マレーシア語 (Malay)、タミール語 (Tamil) が一般に用いられている。
2. シンガポールはアジア太平洋地区の貿易、観光情報サービスの中核の一つであり、言語交流の障害の克服が必須というニーズが高まっている。
3. シンガポール政府は情報技術の発展を非常に重視しており、この分野への投資比重も多く、また「情報技術2000年 (IT2000)」と称する長期発展計画を制定している。このようなニーズ環境が機械翻訳の技術研究を後押ししているのである。シンガポールは今年建国30周年を迎えたばかりの若い国家であり、当然機械翻訳の研究や発展の歴史

は諸外国にくらべ短い、若い国には若いだけの大きな潜在力と可能性を持っている。

本稿では、シンガポールの機械翻訳の研究・開発の歴史と現状について概述し、その特徴を紹介、考察したいと思う。

1. 当地の機械翻訳の歴史

シンガポールでの機械翻訳の研究開発は1987～88年に始った。まさに機械翻訳研究が世界中で大流行した時期である。大規模な国際プロジェクトとして EUROTRA、CICC多言語機械翻訳システムなどの共同研究が行われ、多くの商品化された機械翻訳システムが出現したのもこの時期である。

1987年には日本で第1回国際機械翻訳サミット (MT SUMMIT I) が開かれた。そうした国外の潮流とシンガポール国内のニーズ、潜在的可能性がシンガポールの機械翻訳研究を軌道に押し上げたといえる。シンガポールの機械翻訳の歴史の中でも以下の各シ

←(前頁からの続き)

以下に述べる方法で辞書(各文字カテゴリの規準となる参照パターン)を作成します。

- (1) 多数のフォントのサンプルを集め計算機に蓄える。
- (2) 印刷や画像入力時に起こると予想される変形操作の加えたパターン(太め、細め、位置ずれなど)を計算機内で自動的に生成する。
- (3) 生成された多数のパターンに対し統計解析を使用して辞書を作成する。

この方法により、多様なフォントの混在した文書の高い認識率で読み取ることが可能となりました。

3. 7 単語辞書との照合

文字認識された後、文字を単語に統合しますが、この時、単語の中に文字認識処理で疑わしいと判定された文字がある場合は、単語辞書との照合により

最終結果を確定します。ここでいう疑わしい文字というのは、'l', 'I', 'I' や '0', '0' 等の類似文字、およびノイズ・つぶれ・かすれなどで劣化した文字のことです。これらについては単語照合によって認識精度をさらに高めることができます。

これについては自然言語処理におけるたとえば形態素解析などの技術を応用してさらに高度な後処理を実現し、高精度化もはかられています。

3. 8 読み取り結果の出力

最後に読み取り結果を各種ワープロ・ソフトのファイル形式に変換します。この際、文字サイズ、フォント(イタリック、ボールド)、アンダーラインの存在範囲も属性として付加されて出力されます。変換できるワープロ・ソフトの種類は豊富に揃えられているのが普通です。

(東芝RDC・CI研 黒沢、有吉、辻本)

システムは特筆すべき成果である。

1. 1 CANTENA

1987年、シンガポール国立大学システム科学研究院 (Institute of Systems Science, ISS) の鄭奮興 (H. H. Teh) の強力な指導と後押しで、またシンガポール国立大学情報システム・コンピューター科学学部 (Department of Information Systems and Computer Science, DISCS) の陳樹霖 (C. C. Tan) の直接参画を得て、1988年シンガポール新聞業持株グループと協同でシンガポール初の機械翻訳システムの研究開発への取り組みが始った。システム名をCATENA (Computer Aided Translation of English News Articles) といい、英語-中国語片方向翻訳機能を持ち、対象専門分野を金融業界ニュースに絞り、ルール・ベースの直接法 (Rule-based, direct approach) を採用しているのが主な特徴である。

CATENAのプロトタイプは1990年に完成した。

1. 2 C-SHALT

1990年、ISSとIBMは機械翻訳システムの共同研究開発に関する契約に調印した。その目的は実用的な英漢機械翻訳システムの構築であり、研究は2段階に分けて実施された。

プロジェクトはISS側で進められ、余垂珠 (Jane EE) が機械翻訳グループの責任者となった。システムはC-SHALT (Chinese-System for Human Assisted Language Translation) と呼ばれた。

SHALTというのはIBM東京基礎研究所が開発した英漢機械翻訳システムであるが、典型的なルール・ベース型であり、置換法 (transfer approach) を採用したシステムである。C-SHALTの英語分析もSHALTと同様にPEGと呼ばれる文法、すなわちPLNLP (Programming Language for NLP) に基づく英語文法を採用している。ソース言語を分析して得られた中間表現を深淺両層に分けて分析を加える。すなわち浅層で統語構造 (syntactic structure) を解析する。後者はC-SHALTの中国語生成の起点となる。生成は2つの子モジュールを経て行われるが、その1つは生成規則 (Generation Planning) で、主に中国語の単語序列と虚詞運用が

管理される。

2つ目は生成実現 (Generation Realization) である。語彙転換は主に英漢辞書の助けを得て行われる。このシステムは第一段階 (16カ月) を経た時点でもかなり大きな進展が見られ、実用に向けた潜在力を予見させるまでになったが、惜しいことにIBMによるPLNLPグループが解体したため、プロジェクトの継続進行が不可能になってしまった。

1. 3 SAGE MT

SAGE (Software Aided Group Environment) はシンガポール国立大学が開発した決定サポートシステムである。SAGE MTは多元的文化グループ体の決定サポートを実現し、SAGEに多言語をサポートさせることを目的として、陳樹霖が彼の学生を指導、指揮して開発したSAGE英漢機械翻訳システムである。同システムはISSの湯龍祥 (Tong Loong Cheong) が設計開発したTapestry NLP Toolkitすなわち自然言語処理ツールを開発環境として利用している。湯龍祥はISSに入る前から機械翻訳研究に従事しており、有名なJEMAH英語-マレーシア語機械翻訳システムの設計者でもある。

SAGE MTはプロトタイプシステムに過ぎないが、Tapestryが多言語処理にも使える可能性を証明したシステムであった。

2. 現状と特徴

上述したようにシンガポールにおける機械翻訳の研究に従事している主な組織はDISCSとISSであるが、前者は教育システムや研究生育成とリンクしており、学術性に重点が置かれている。陳樹霖は1993年に日本の京都大学に行き、例文ベース (example-based) の機械翻訳研究に従事したことがある。後者はプロセス性を明瞭に打出しており、産業ニーズに力点を置いている。ISSは強力な技術力を持ち、ソフト及び言語方面の人材が合理的に配置された機械翻訳研究チームを構成しており、シンガポールの機械翻訳研究の中心的かつ代表的存在となっている。

現在ISSの研究方向と目標は以下の諸点に置かれている。すなわち、

1. 多言語処理であること。特に英語、マレーシア

語、中国語。

2. 機械翻訳実務と緊密に結びつき、できる限り早く機械翻訳センターを建設できるよう努める。
3. 信頼できる言語資源と成熟した技術を依り所とし、新技術を吸収して、既存システムに代る次世代となるシステムを構築する。ISS既存システムで、現在発展しつつあるか、すでに使用されているものには以下の3システムがある。

ISS EC英漢システム：

これはC-SHALTプロジェクトがストップした後、ISS機械翻訳グループが北京 Pacific 会社と共同開発したシステムである。同システムはコンピュータや金融財務等専門分野を含む10余万語の辞書を持っている。C言語でプログラミングされ、パソコンやワークステーション上で動作し、処理は高速で、適応性や拡張性の面でも優れている。

Tapestry 英語-マレーシア語システム：

英語、マレーシア語の辞書（英語は6万語）と英マ変換辞書を持っている。Tapestry NLP Toolkit をベースとする。

METS マレーシア語-英語システム：

性格はTapestry英マシステムと同じ。搭載辞書であるマレーシア語辞書には現在1万5千語の見出し語が登録されている。

ISSにおける機械翻訳研究の主な特徴：

2. 1 実用向き

1993年後半、ISSは新開発したISS EC英漢機械翻訳システムをMTF (Machine Translation Factory) というソフトに詰め込んで、英語から中国語（簡体字あるいは繁体字）への翻訳サービスを開始した。最近ではTapestry英語マレーシア語システムをMTFに詰め込んでいる。MTFというコンセプトはISSの劉斐文 (Low Hwee Boon) が提起したもので、彼は翻訳という作業も工業化の過程を経るに違いないという認識を持っていた。すなわち、一枚の紙、一本のペンで手作業で行われていた翻訳がコンピュータのサポートを経て工業的に生産され、未来の機械翻訳センターは翻訳工場となるだろうという考えである。今日、機械翻訳はまだ満

足できる訳文品質を得られていないが、すでに多くの機械翻訳システムが実際の翻訳実務に導入されており、また各種各様の商品化されたシステムが市場に出回っている。おもちゃ的翻訳システムと実用翻訳ソフトは翻訳量の違いだけに留らないというのは多くの人が認識しているところである。

一つの実用システムを翻訳実務に導入する場合、例えば1冊のユーザーズ・マニュアルを翻訳する場合、直面する問題解決の困難さは、おもちゃ的システムから実用システムへの発展を試みることと較べても、恐らく勝るとも劣らないだろう。

MTFの翻訳作業のプロセスには主に以下のものがある。テキスト抽出 (text extraction)、テキスト分割 (text segmentation)、翻訳前フィルター (pre-translation filter) 述語識別 (term identification)、機械翻訳、訳後編集、フィルター復元 (filter reformation)、本文復元 (documentation reformation) 等である。

これから分ることは、我々がこれまで鋭意研究してきた「機械翻訳」というのは、最も重要な部分ではあるが、翻訳全作業中の一部分に過ぎないということである。翻訳前フィルターの中には訳文再利用 (translation reuse) という重要なソフトがあり、正確マッチング (exact matching)、近似マッチング (approximate matching) という技術を用いて既存訳文の再利用を実現している。このようすることで、時間と労力が節約できるだけでなく、訳文の一致性が保証される。これは重要である。

また述語識別は機械翻訳前に抽出可能な述語を最大限度抽出し、それをユーザー辞書構築に供することで述語の一致性を確保できるようになっている。実際の翻訳実務の中に多くの新たな問題が発見でき、そこから新技術を研究、採用して、実用に役立たせることができる。

実用を経て問題を発見し研究を深めていくというやり方はISSの機械翻訳研究における第一の特徴と言える。ISSは長年にわたる固定ユーザーを持っているが、その中にInformix Asia/Pacificのようにシンガポールに本社を持つ多国籍企業やSingalabといった国内企業がある。

2. 2 技術の総合利用

周知の通り、機械翻訳は一種の総合技術であり、

NLP中の主要技術をほとんど含んでいる。ISSはこれらの技術の総合利用に相当な注意を払っている。湯龍祥が指導して開発したTapestryは機械翻訳研究に源を発するNLPツールで、これは以下の各部分から構成されている。すなわち、文法記述言語(grammar writing language)、辞書加工器(dictionary processor)、語義分類加工器(ontology processor)、自動タグ付け器(automatic tagger)、語法分析器(morphological analyzer)、単語パターン適合器(lexical pattern matcher)、テキスト記述言語(text description language)、意味フレーム加工器(meaning frame processor)等である。これらは機械翻訳システムの開発に用いられるが、それだけに留らず、例えばテキスト分類(text categorization)、テキスト内容抽出(text abstraction)等その他のNLPアプリケーションにも有効なサービスを提供できるものであることが、実践的に証明されている。

2. 3 技術の整合

上述の技術の総合利用で、ある種の技術は多種にわたる異なる分野にも広範に利用できるということを述べたが、これは技術の多重再利用と言われる。

一方、技術の整合とは元々単一のものであった専用技術を合理的に集成し、それによって一つの新しい技術を形成したり、新しい技術を開拓したりする

ことである。前述の3種類の主要な機械翻訳システムはソフト、言語材料(lingware)ともすべて異なっており、互換性もない。こうした形では資源の浪費、労働の重複が出て来ることは避けられないが、その解決方法は技術整合に求めることができる。

1993年以来、ISSの自然言語処理研究部(現在、多言語処理研究部と改称。MLP Program)はこの点を重要視してきており、技術整合は今や既定戦略となっている。技術整合という考え方には、成熟した技術や長年苦勞して蓄積してきた資源を最大限に再利用を図るという意味も含んでいる。現在MLP Programではマスター辞書(Master Dictionary)と呼ばれる大型の辞書データベースを構築が進められている。また近い将来、Tapestry-Lexと命名されたワードプロセッサもISSの新機械翻訳システムの構成部分となるだろう。それらは個別に開発されたものであるが、別の特定システムに依存することなく、当然他の言語処理にも利用できる。

〔筆者紹介〕董 振 東

シンガポール国立大学システム科学研究院研究員。かつてCICC多言語機械翻訳研究プロジェクトの中国側技術責任者を担当(1993~94年)EDRで英語分析研究に従事、初めて商品化された英漢機械翻訳システムの設計者。

英語/ウクライナ語対訳コンピューター辞書

「Easy Ukrainian for Windows 3.1」

- 特 徴
- ・英語2万語に対応するウクライナ語3万5千語搭載、発音とアクセントを併記
 - ・英語の単語単位でウクライナ語訳が検索できる
 - ・アプリケーション英語記述を入力し、ウクライナ語訳がつけられる
 - ・訳語は同意語/地名まで網羅
 - ・ヘルプ機能はオンラインで作動

適用機種 MS/DOS版(Windows Ver 3.1)

ハードディスク 空きメモリ1.2 MB以上

フォーマット 3.5インチ(1.44MB)または5インチ(1.2MB)

価 格 \$139.00

販 売 社 POMIC REG(4359, Venise, Montreal, P.Q. HISIG CANADA ☎ (514) 728-3612)

チベット語情報処理技術の現状と展望

青海師範大学計算中心 徳熙嘉措 (Bdhe Shis Rgya Tshogs)

チベット族の基本概況

(1) チベット族の地理分布

チベット族は主に地球の第3紀に形成された青藏高原の降雪地域で生活している。チベット族は6、7万年前からすでに青藏高原に定住している。ヒマラヤ造山運動により、青藏高原はこの10万年間で徐々に海拔2千余メートルから現在の平均海拔4千余メートルまで隆起している。

チベット族は狩猟-農業-牧畜の3段階の特殊な歴史時期を経て、青藏高原の熱帯-温帯-半寒帯といった歴史的な地理環境の変化の中で、生産労働と生活を営み、今日まで繁栄してきた民族である。

現在チベット族は約600万人おり、中国国内で約500万、中国のチベット、青海、甘肅南部(甘南)、四川西部(原西康)及び雲南西北部に分布している。

言語と生活方式によって衛蔵(Tibet)、安多(Ando)と康巴(Kang Ba)の3つの区域に別れるが全体は蔵区(Tibet)と総称される。

国外在住のチベット族は約100万人で、インド(25万)、ラダク(30万)、カシミール(10万)、アメリカ(8万)、スイス(6万)、ベルギー(2万)、カナダ(5千)、ネパール(1万)、及びフランス、ドイツ、イギリス、スペイン、オランダなどでほとんどすべての国にチベット族の人がいるといっている。

(2) チベット族の経済生産

国外のチベット族は在住する国の人々と共に生活し、当地の様々な経済活動に参加している。中でもインド、ネパールのチベット族は自分たちで組織的に生産を行い、工場を営んでいる。

国内のチベット族の密集した地区では、牧畜生産が営まれ、青藏高原特有のヤクとチベット山羊を放

牧している。肉の価格が低いため、収入は少ない。河谷の比較的温暖な地区では農業あるいは半農半牧生産が行われている。都市のチベット族は漢族と同様に各業種に参加したり学習している。

(3) チベット族の歴史と文化

チベット族は約紀元前6、7万年~1万年までは青藏高原で生活しており、多くの旧石器の文化遺跡を残している。紀元前1万年~5千年の間は石器時代で狩猟と農業時代である。紀元前5千年前あたりから青藏高原の気候が寒冷になったため、徐々に牧畜生産へと移行していった。紀元4千年前にはチベット族はも青*(禾+果)や小麦を植え(カヤック文化)紀元前3700年前ころには銅器を使用しているが、これは中国の中原地区より600年早い。

3700±200年以前にはすでに古代チベット語を持ち(青海玉樹石刻)、アジアの3つの文化帯である北方オールドス文化帯、中原夏文化帯、南方蔵緬文化帯のうちの蔵緬文化帯の主な発祥地であった。南方の多くの民族の伝説で、彼等の祖先はコンロン山から降りてきたと伝えられている。

紀元前のチベット区は象雄王朝によって1千年統治された。7世紀~9世紀はトルファン王朝が統治した。中国の唐朝と東アジア2大王国を形成し、その中で、7世紀初には黄河上流で吐谷渾を敗り、唐朝と戦い、青海で唐軍40万を敗った後、両軍は境界を設けて、以後友好的に交流した。これ以後チベット族はインドから仏教を取り入れ、当地のボン教(Boo)と融合させ、蔵伝仏教を形成した。

10世紀から12世紀は封建領主による割拠時期である。この時期仏教が大きく発展した。13世紀にはモンゴル人が青海に侵入し、チベットを割拠したタングート帝国を成立させた。17世紀になるとトルファン帝国が成立して蒙古人による統治から独立し、満州族と連盟した。

1904年には英軍の侵入を敗り、半独立国家を

成立させ、1936年には中ロ英蔵条約により独立が確定し、第2次大戦時には中立地区となった。そして1951年に中華人民共和国に帰属した。

チベット族は仏教を依拠として発展し、チベット族伝統文化を形成、完成させてきた。哲学（宗教）人文科学、チベット医学、天文暦、芸術、文学等、今日までに豊かなチベット族の伝統文化を残してきた。3千余巻に上る大蔵経は各種經典を集成した百科叢書であり、有名な100万行の詩歌「格薩王伝」は有史以来最も長編の文学作品である。

チベット語の文字は表音式の組合わせ文字である

チベット族は7世紀に従来使用していた各種文字（象雄文を含む）を統一的に訂正し、松*干布の大臣吞米桑不札により制定され、無頭のチベット文字の字母が作られた。これは従来の象雄文より簡潔で美しい（彼が発明したチベット字母はインドのある文字からの借用である等と誤認識している人もいる）ものであったが、さらにいくどかの改訂を経て、現在のチベット文字が形成された。

チベット字母には2セットある。1つは当用チベット字母で、これは30個の声母文字と9個の韻母文字を持つ。もう一つはサンスクリット語名詞の音訳に用いられるサンスクリット転写用チベット文字で、3千の声母があり、16個の簡単な韻母と28個の複合韻母からなる。

チベット語の字母によるチベット文字表記は、上加字母、基本字母、下加字母、韻母の4部分から構成されるが、これは漢字の偏旁部首と似ている。漢字の音と部首は無関係であるが、チベット語はハングルと同様に字母を上下に組合わせて字符を作り、さらに前加字母、後加字母、再後加字母を加えて、音素を表す1文字を構成する。字音は構造中の字母の組合わせによって決定される。つまり、チベット語の文字は字符の組合わせからなり、字符は字母の結合で構成され、全体で一つの不可分の独立したブロック文字が形成されているのである。チベット語の文字の構造の分らない人は、チベット文字は英語のように音節を組合わせた文字だと勘違いしている人が多い。

以上のように、情報処理する際、チベット語の最小単位は文字であって、字母でないということを知

らなければならない。

チベット語の文法はインドのサンスクリット語と同じく、ラテン語系に相似しており、動詞の語尾変化、変格などの形式がある。チベット・ビルマ語系に属する。チベット・漢語系と見るのは誤りである。

チベット族は閉ざされた青藏高原上で独自の完成された文化体系を形成してきた。それは中国文化、西洋文化と同格である。哲学、精神科学方面では相当豊富な内容を持っており、中でも密教には哲学、認識論、思惟学、方法論、弁証法、霊智学、情緒学心理学、靈感学、人体機能学等、人類が21世紀に精神文明世界を構築していく上で必要となる有益な智恵の宝庫となっている。

チベット語情報処理の現在の状況

電子技術の発達、中国語処理技術の影響、チベット族内の科学、文化、技術の発展要求、チベット語処理技術の研究開発は一定の発展を遂げてきた。もちろん多くの困難にも遭遇してきたが、過去10年の間、少なからぬ研究成果や製品が生み出されている。

国内でチベット語情報処理の研究開発には、青海師範大学計算センター、青海民族学院物理学部、西北民族学院少数民族語学部、中央民族大学コンピュータ学部、チベット大学、チベット電信局、北大新技術開発公司、中国チベット学センター、中国航天部（宇宙航空省）、山東維坊計算機廠、深せん藍海公司等、多くの研究機関が従事しており、OS、ワードプロセッサ、レーザー写植、翻訳システム、音声合成等の方面で多くの仕事をしている。

(1) チベット語OS

現在までに開発、導入使用されているチベット語OSはすでに10種類にのぼるが、そのうち、青海師範大学の徳熙嘉措教授が1968年に開発したチベットCDO S2.00は実用に耐えるレベルに初めて到達したチベット語情報処理用の高機能のチベット語OSである。その後、航天部の羅聖儀、青海民族学院の錢伝強、アメリカのprices、イギリスのブタン等が異なる入力方法のチベット語OSを続々と誕生させた。しかし、文字バンクの文字のカバー率や

字体種についてはさらなる増強が必要であった。

(2) ワードプロセッサシステム

1986年、青海師範大学の徳熙嘉措がチベットCWSを開発した。これは英文ワークステーション用のチベット語ワープロシステムである。1992年には西北民族学院の于洪志と熊濤が藍海ワープロシステムLHWPSを開発した。これは機能が大幅に向上されており、文字バンクに一部ではあるがサンسكريットチベット転写用文字が含まれていた。航天部の羅聖儀が開発したチベット語軽印刷システムでは入力方法が増えて4種類（区位コード、字形コード、打丁文字代用コード、数字コード）となっている。

中国チベット学センター、青海民族印刷廠、山東維坊計算機廠が共同開発した華光チベット語レーザー写植システムでは、チベット文字バンクが4000文字に達し、フォントも美しい。これは1994年に図書や新聞の写植製版印刷に導入されている。

北京大学方正公司のレーザー写植システムはチベット語を乗せ変えることができ、フォントも正字体ゴシック、太ゴシック、アートフォントなど多彩だが、入力方法はチベット語、サンسكريット語ごとに1文字ずつ入力する方法で、入力速度は遅い。

現在ある各種システムにおいては、文字バンクは当用チベット文字とサンسكريットチベット転写用文字の2種類からなり、前者はすでに固まっているが、後者は文字使用頻度の統計がなされていないため、完全なものになっていない。

文字が分字点を持つと文字バンクは約30%増大する。チベット文字では統計の結果さらに分字点を持つ文字を294個増加する必要がある。そうすることでソフト間の互換性がよくなり、画面表示、印字ともきれいに整った文字となり、また情報量も23%減らすことができる。分字点を持たぬ文字バンクでは、画面表示、印字とも整然さが損われ、しまりがない上に、ソフト間の互換性も劣り、情報量もいたずらに増大する。

(3) 翻訳システム

1990年、青海師範大学計算機センターが英語

チベット語の音訳（外国語の読みを自国文字に転写）システムを開発した。英語の専門名詞、借用する科学技術名詞、地名、国名、人名等をチベット語に変換するもので、チベット語の名詞の音訳規範化に寄与するものであった。

1992年にはさらに中国語チベット語音訳システムを開発した。これは借用する中国語語彙を規範化された「ワード対ワード」で音訳するものである。

現在、西南民族学院はテキストの中国語チベット語翻訳システムの研究開発を進めており、青海師範大学の徳熙嘉措は英語チベット語の技術文献翻訳システムの研究開発に取り組んでいる。

機械翻訳を行う際のベースグラウンドとして重要なポイントは電子辞典が充実しているかどうかである。現在チベット語におけるテクニカルタームは不統一であり、かつ不足している。青海師範大学民族学部の桑元太らはすでに数学、物理、化学の3分野のテクニカルタームのチベット語辞典を完成させており、青海民族学院では今後数年内に生物、地理、医薬、計算機、農業、交通、商業、金融、統計、食品、薬物、密教等十数分野の専門用語辞書の完成を目指している。

機械翻訳のベースグラウンドとしてはもう一つ、チベット語文法、句法に対する精密な研究が重要であるが、現在それに取り組んでいる人はいない。

(4) 音声合成

1992年、チベット電信局がラサ音声合成システムを開発し、電信事業に応用されている。これによりチベット語を話せない漢族が電信業務において電話オペレーションを行う際の言語上の問題が解決された。

(5) チベット語情報処理の国際標準

チベット語情報処理及び交換用情報コード、文字ドット、コンピュータ用ディスクなどの中国国家標準と国際チベット語文字コード標準については、現在、その研究並びに確定作業がチベット語情報処理標準専門家グループによって進められている。専門家グループはチベット大学、青海師範大学、西北民

族学院、中央民族大学、深セン安華公司等で構成され、チベット民族委員会が牽引的役割を担い、青海師範大学がグループの長となって、チベット語情報処理に関する中国国家標準の研究に取り組んでいる。現在すでに小文字集、中文字集の標準草案が提出されており、今は大文字集のコーディングが検討中である。

アメリカは小文字集の国際標準を打出し、イギリスもそれに類似した小文字集案を提出し、ISOによる選定を目指している。

チベット文字の字母字体は文字の発音に対して表音作用を持つが、字母で構成された文字、構成された長方形の文字ブロックはそれ以上分解不能な独立した完全体であり、ハングルと非常に似ている。このため約6000個の文字を有する大文字集が必要となる。さもなくばチベット語情報処理上の文字に長さの異なる文字コードが発生し、画面表示、印字にも大きな困難をもたらすことになる。この点はタイ語の状況とやや似ている。組合わせ式文字の特徴を重視しなければならない。

チベット語情報処理の展望

チベット族の伝統文化は世界の人類文化にとっても得がたい宝庫の一つである。チベット族の伝統文化を現代化された情報処理技術を利用して全世界に紹介することは、21世紀の精神文明の建設のためにも、軽視できない意味を持っている。

チベット族は現代科学技術教育を進展させることを急務としており、そのためにはコンピュータ処理技術の導入、普及が不可欠で、我々は現代の情報処理分野の先進レベルに追い付かなければならない。

我々は出来る限り早くWINDOWS上のチベット語プラットフォームを開発し、良好なチベット語情報処理環境を提供し、また手書き文字識別機能を持つレーザー写植システムも開発し、文字バンクを完全なものとし、フォントを多用に整えなければならない。また積極的に英語-チベット語、中国語-チベット語の双方向テキスト機械翻訳システムを開発し、チベット族の伝統文化と現代科学技術文化との間の交流を促進したい。

人々のチベット語情報処理の重要性に対する認識は日に日に深まっており、注目を集めている。あと

数年も経ぬうちに世界各国でチベット語情報処理関連のソフトやマルチメディアなどの電子出版物が研究開発を経て世に出てくるにちがいない。我々はその確信している。

《筆者プロフィール》

1938年、中国チベット区安多（青海）生れ。チベット人。蘭州大学物理学部卒業後、物理学講師、計算機助教授、計算機教授などを歴任。

蘭州大学、青海師範大学物理学部、計算機センターでICの研究開発、チベット語OSの開発、英語-チベット語機械翻訳の研究に従事。

現在シンガポール国立大学システム科学研究所（ISS）で多言語処理システム（MASS）を研究中。またチベット語情報処理中国国家標準専門家グループの一員としてチベット語情報処理標準化の研究に参画。

BDHE-SHIS RGYA-THOGS 德熙嘉措（趙晨星）

教授 青海師範大学計算センター
中国青海省西寧 青海師範大学計算センター
中国、青海、貴徳

1962年 蘭州大学物理学部卒業後、大学に残って教員となり、ICとCPUを研究するかたわら、電磁学と無線を教える。

1975年 青海師範大学講師。電子技術と近代物理実験を教える。

1984年 青海師範大学計算センター副主任、助教授。研究生と共にコンピュータ・アプリケーションを研究

1986年 世界初のチベット語OSを開発。チベット語序類コード入力方法を発明。

1992年 特別手当が支給される「国家級専門家」となる。中国チベット語情報処理専門家グループグループ長。チベット語情報処理、チベット語科学技術史などの論文20余篇を発表。

EDR の研究成果と今後の展開について

(株)日本電子化辞書研究所(EDR)横井俊夫・末松博

1 はじめに

EDR電子化辞書プロジェクトは、1985年度から1994年度の9年間にわたるコンピュータによる言語処理用の辞書の開発を目的としたプロジェクトである。資金の7割は日本政府の出資という準国家プロジェクトではあるが、実行の主体となったのは代表的な日本のコンピュータメーカー8社である。

欧州で行われている辞書開発がどちらかといえば言語学や辞書学の出身者によって推進されているのに比べ、EDRはコンピュータ・サイエンスの出身者を中心に進められた。

言語学的に精緻な辞書を作るというより、言語処理用の大規模な辞書を短期的に作り上げるということが目的だったからである。当初の目的は達成されつつあるが、今後は精度の向上を含め言語学的観点からの改良が必要と思われる。

EDR電子化辞書の特徴を手短かに説明するために欧米における代表的な同種の活動と簡単に比較してみる。

その代表的な活動としては、辞書を中心としたものとして Acquilex [1]、Genelex [2]、Multilex [3]、Comlex [4]、シソーラス(意味記述)を対象とする WordNet [5]、コーパスを対象とする British National Corpus [6, 7, 8, 9]、Penn Treebank [10] がある。

比較によって云えることは次のようなことである。

- (1) EDR電子化辞書の基本的な設計思想は辞書、シソーラス、コーパスの全体を含む統合化された言語データとして電子化辞書を構成することである。ただし、この考え方は最終成果物として十分に成功しているとは言いがたいが、これからの電子化辞書にとっては重要な考え方である。
- (2) 構文的な情報の記述に関しては、ごくポピュラーなすでに広く使われている考え方を採用した。

これは、そのようなものでも言語処理用としては十分に役立つこと、多数の作業者の確保を容易にするためである。Acquilex のように最新言語理論を採用するというようなことはしなかった。

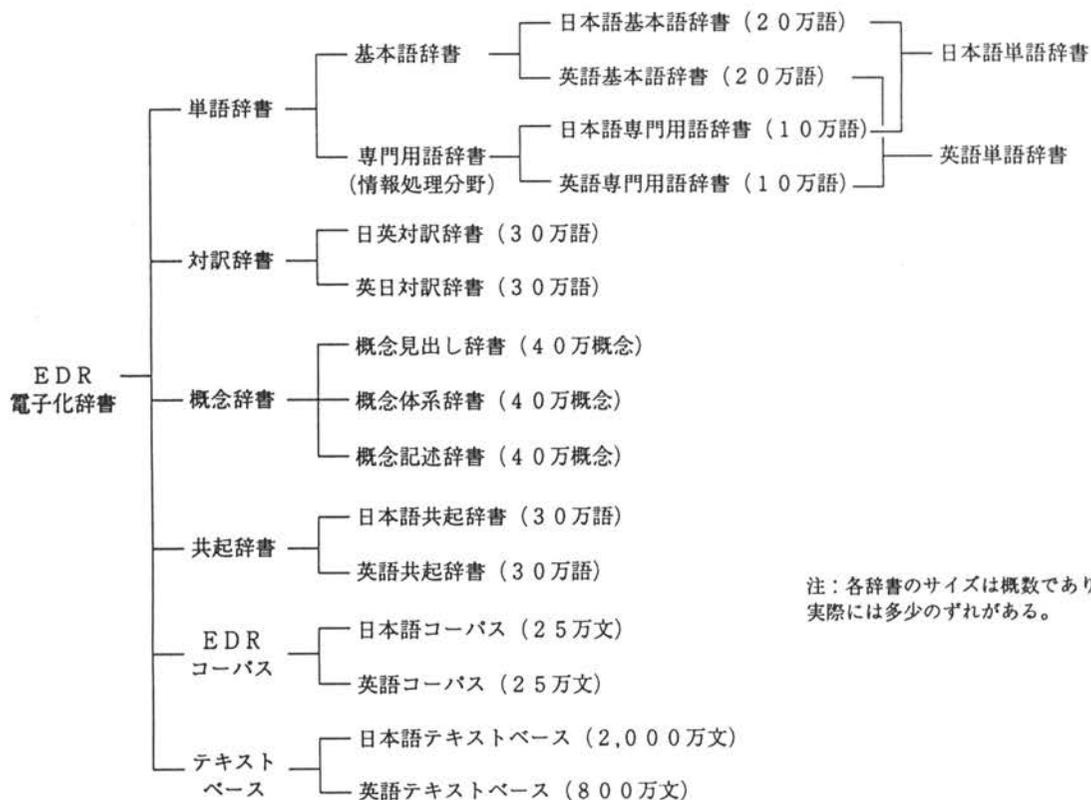
Genelex や Comlex よりももっと保守的であるかもしれない。新しい理論による辞書作りは、これからの課題と考えている。

- (3) データはすべてオリジナルに作成し、既存のMRD(Machine-Readable Dictionary機械可読辞書)を直接利用するということはしなかった。これは著作権上の問題もあってこのような方針をとった。これに関しては Comlex のように既存のものに準拠した方が効率的であったと思われる。(2)の方針から、既存のものも十分に無理なく利用できるからである。
- (4) 日本語と英語を対象とし、出来るだけ言語に共通の枠組みを追求した。しかし、十分であったとはいえない。Multilex の試みなどを十分に比較し再検討する必要がある。
- (5) 意味記述に関しては synonymy(同義)、hyponymy(下位)、entailment(含意) WordNet はこの他に antonymy(反義) meronymy(全体部分)、troponymy(比喩)について行っている。文脈処理等のためには、WordNet のような試みも必要である。ただし、WordNet よりもより言語処理の立場から行われている。
- (6) コーパスはとりあえず辞書開発用、言語処理用であり、British National Corpus のように言語現象を捕捉するものとして設計されたものではない。この点では Penn Treebank のような考え方に近い。タグ付けに関しては形態素、構文木、意味解析について試みているが、意味解析の品質等に関しては今後の課題が多い。

以下、2章で、EDR電子化辞書の構造（仕様）を概略する。3章では今後の改良・拡張の方針とするために、EDR電子化辞書の成果・経験を整理し、

一般化した電子化辞書である汎電子化辞書について簡単にふれる。EDR電子化辞書の設計方針はこの3章によってより良く理解されるものと思う。

2 辞書の構造



注：各辞書のサイズは概数であり
実際には多少のずれがある。

図2. 1 EDR電子化辞書の構成

2. 1 概要

EDR電子化辞書は単語辞書、対訳辞書、概念辞書、共起辞書の4種類の辞書とEDRコーパス、テキストベースから構成されている（図2. 1）。

単語辞書は、語彙の違いによって基本語辞書と専門用語辞書に分けられる。

基本語辞書は、言語の違いによって各々20万語の語彙を持つ日本語基本語辞書と英語基本語辞書に分けられる。

専門用語辞書は情報処理分野に関するものであり、言語の違いにより各々10万語の語彙を持つ日本語専門用語辞書と英語専門用語辞書に分けられる。

単語辞書の基本的役割は、単語と概念（意味）との対応関係を記述し、この対応関係が成り立つときの文法的特性を与えることである。

対訳辞書は、言語対の違いによって、各々30万語の日英対訳辞書と英日対訳辞書から構成される。対訳辞書は、日本語と英語の単語見出し間の対応関係の情報を与える。

概念辞書は単語辞書に語義として導入された40万の概念についての知識が記述され、情報の種類によって、概念見出し辞書、概念体系辞書、概念記述辞書の3つに分けられる。概念見出し辞書は概念に説明を施したものであり、概念体系辞書は40万の概念に対して、それらの間の上位下位関係を記述し

たものである。上位下位関係とは概念間の包含関係であり、一種のシソーラスと見なすことができる。

概念記述辞書は文中に共起する概念間（2項）の意味的關係（動作主、道具、場所、等）を整理したものを記述したものである。

共起辞書は、各々30万語の日本語共起辞書と英語共起辞書から構成され、言葉の言い回しに関する情報を2項関係で記述したものである。

EDRコーパスは、各々25万文の日本語コーパスと英語コーパスから構成される。大量の用例を収集し、形態素レベルから、構文・意味レベルまで分析して得られる言語データである。

テキストベースは、2,000万文の日本語テキストベースと、800万文の英語テキストベースから構成される。新聞、雑誌、百科事典、等からの文章を蓄積したものである。

EDRコーパスにおける文はここから抽出される。

2.2 EDR電子化辞書の仕様の概略

次に各辞書の仕様を概略する。詳しい仕様は本稿の最後に案内する「EDR電子化辞書発表会」にて発表予定であるのでそちらを参照されたい。

単語辞書の各レコードは、単語見出しや発音を始めとした形態に関する情報、品詞や表層格を始めとした文法的機能を表す情報、語義に関連した意味的情報、等から構成される。冊子体の辞書と構成は似ている。

対訳辞書の各レコードの構成は単語辞書を簡略化したものであるが、訳語に関する情報が備わっている点が異なる。この情報には訳語の種別（同義語、パラフレーズ、逐語訳、ローマ字/カナ、説明文）訳語、そして訳語の品詞が記述されている。

概念辞書は概念見出し辞書、概念体系辞書、概念記述辞書から構成される。概念見出し辞書の各レコードは、概念識別のための標識、代表語と上位語を組み合わせて概念への見出しとした概念見出しや、自然言語による説明から構成される。概念体系辞書の各レコードは、概念を概念間の上位下位関係（概念体系）で表現したものである。例えば、学校という概念は上位概念として組織、建物、機能を持ち、下位概念として小学校や大学などを持つものとして表現される。

概念記述辞書の各レコードには、他の概念との意味的關係（動作主、道具、場所、等）がタプル（概念-関係子-概念）として記述されている。

共起辞書の各レコードは、共起句を見出しと捕らえた句見出し、共起の状況、共起句の構成要素、表層の共起関係、そして概念の共起関係から構成される。

句見出しには単語の表記と共起関係子との組み合わせを記述する。共起状況にはコーパス中に共起した状況を頻度、例文という形で示したものを記述する。

共起句の構成要素情報には、共起句を構成する要素の情報を、形態素、かな表記（英語の場合には原形表記）、品詞、概念見出し等を用いて記述する。表層の共起情報には、表層の共起関係を表すタプルを記述し、概念の共起情報には、概念の共起関係を表すタプルを記述する。

EDRコーパスの各レコードの構成は共起辞書と基本的に同じである。共起状況の代わりに出典情報が記述され、文が出現したテキストが記述される。文の構成要素情報には、単語辞書を用いて文を検索した結果が表示され、共起辞書における構成要素情報と同様の情報が記述される。形態素列と別解析候補が記述され、文の形態素切りの結果と複合語・慣用表現と見なすべきものの候補を示す。依存構造を基本とする構文木が記述され、さらに、フレームないし、グラフにより概念間の関係を記した概念関係表現のデータが記述される。

テキストベースの各レコードは、新聞、雑誌等から獲得した文章を記述した文章見出しと、SGML (Standard Generalized Mark-up Language) によりコード化した文章を記述したテキスト構造情報より構成される。

3 汎電子化辞書に向けて

3.1 電子化辞書の立体構造

第1章で紹介された海外の電子化辞書に関連する様々な活動を考え合わせた場合、電子化辞書の基本構造を図3.1のように立体的に纏めることができる。この基本構造は、記述の単位（語・文・文章・・・）、記述のレベル（表層・深層・・・）、言語の種類（日本語、英語、・・・）の3軸に沿って成

り立つ。表層レベルに於いては、言語情報の構成単位（語・句・文・文章・文書）に沿ってそれぞれサブ辞書が設定される。

このサブ辞書は要素とそのコンテキスト、要素とその説明という関係で相互に関係し合うことになる。また、このサブ辞書は言語の種類軸に沿って各言語ごとに設けられ、言語間では広義の対訳によって関係し合うことになる。サブ辞書の辞書レコードは表記の構造に関わる形態的情報、構文的な所属カテゴリや分解構造に関わる構文的情報、意味の説明や

表現に関わる意味的情報、分野や状況に関わる運用的情報から成る。深層レベルは、意味表現の機構を何にするかによっていくつかの階層が設定できる。

この概念レベルにも原則として言語情報の構成単位に沿ってそれぞれにサブ辞書が設定される。このサブ辞書間の関係や言語間のサブ辞書間の関係は原則として表層レベルのものと同様である。

表層レベルのサブ辞書の意味的情報部分を一段と抽象化したものを概念と呼ぶが、これを各種の意味的な関係を用いて定義するのが役割である。

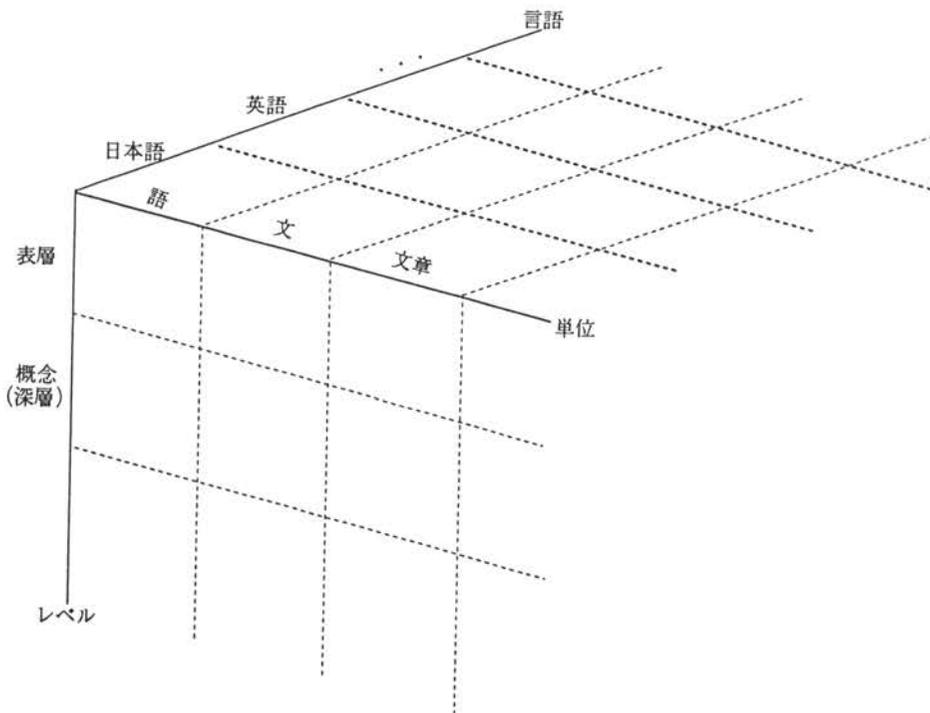


図3. 1 電子化辞書の立体構造

3. 2 EDR電子化辞書の位置付け

EDR電子化辞書の13個のサブ辞書を、この基本的コンセプトに照らし合わせて再整理すると図3.2に示すような関係になる。

記述の単位の観点から見た場合、語レベルに対応する辞書として単語辞書、対訳辞書、概念見出し辞書、概念体系辞書の4種類の辞書を持つ。句レベルに対

応する辞書として共起辞書と概念記述辞書の2種類の辞書を持つ。文レベルの辞書としてはコーパスを持ち、文章レベルの辞書としてテキストベースを持つ。記述のレベルという観点から見た場合、表層に位置するものは単語辞書、対訳辞書、共起辞書、コーパス、テキストベースであり、深層に位置するのは概念見出し辞書、概念体系辞書、そして概念記述辞書である。また、言語の種類という観点からは、

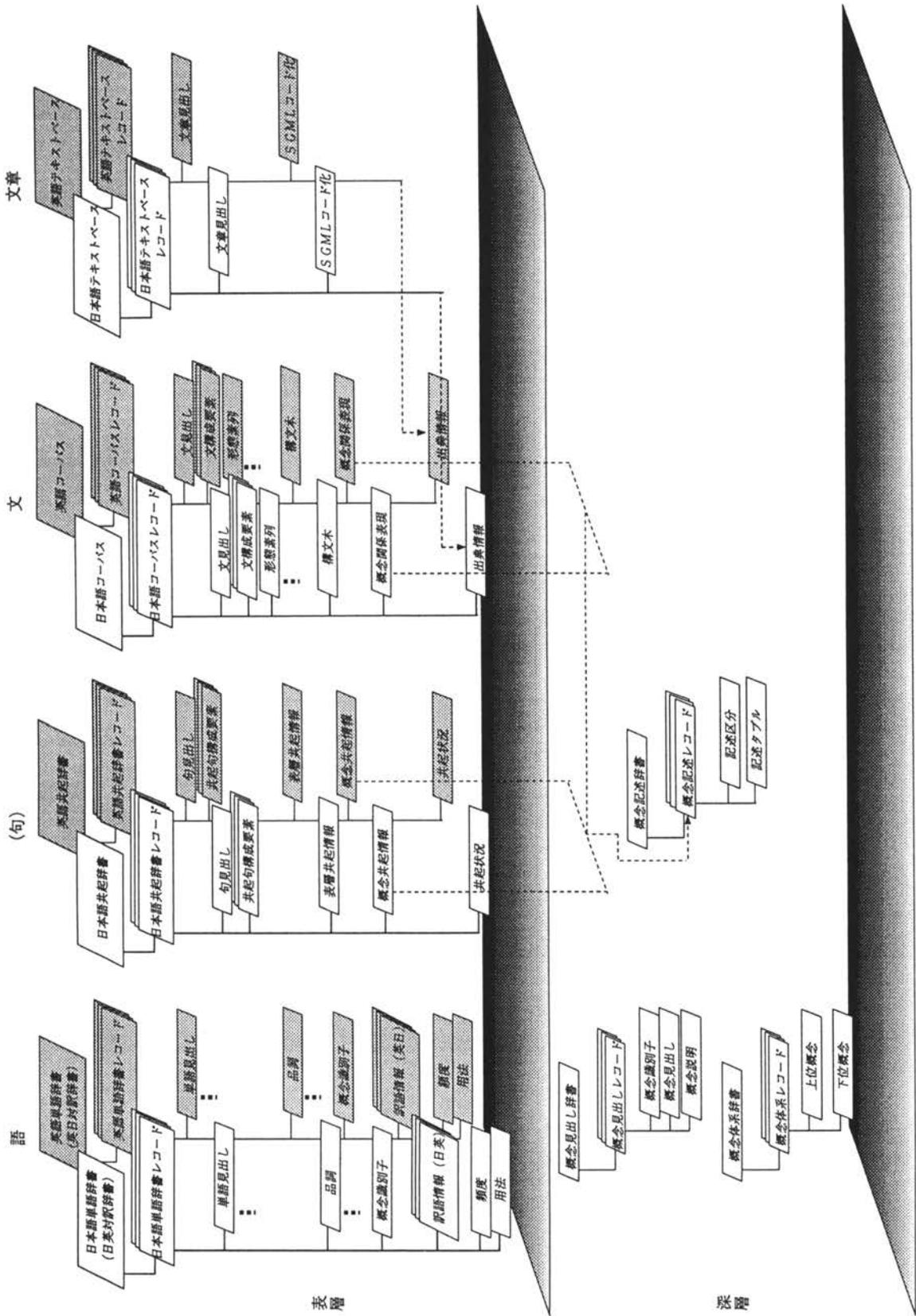


図 3. 2 EDR 電子化辞書の位置付

日本語と英語に関して記述されており、それぞれの言語に対して、単語辞書、対訳辞書、共起辞書、コーパス、テキストベースが用意されている。

表層レベルにおいては、各記述の単位（語、句、文、文章）に対応して、形態的情報として、単語見出し、句見出し、文見出し、文章見出し等があり、構文的（構造）情報として、品詞、表層共起情報、構文木、SGMLコード化がある。意味的情報としては見出しに対応する意味を識別するものとして、概念識別子、概念共起情報、概念関係表現がある。運用的情報としては、頻度・用法、共起状況、出典情報がある。また対訳関係を記述する訳語情報がそれぞれの方向の対訳辞書にある。

深層レベルにおいては、概念の単位を示し、単語辞書を初めとして他の辞書を関係付ける概念識別子が概念見出し辞書に記述され、概念間の上下位関係は概念体系辞書に記述されている。更に、文中に共起する概念間の意味的關係は、コーパス中の概念関係表現を基にして共起辞書の概念共起情報、そして概念記述辞書の概念記述レコードとしてタブルに整理されている。

3. 3 汎電子化辞書に向けた今後の課題

EDR電子化辞書を記述の単位、記述のレベル、言語の種類という3軸により整理し、各サブ辞書の位置付けならびに関係を見直してきたが、そこにはこれから汎電子化辞書に向けて取り組まなければならない課題が明確に提示されている。まず、記述の単位から見た場合、更に安定した言語データを確保するために文章の単位を越えたレベルとみなせる文書レベルのテキストの大量獲得が必要である。記述のレベルという観点からは、深層レベルにおける概念辞書のデータの安定化には、文レベルの深層辞書が必要である。意味ネットワークとしての概念関係表現を大量に収集整理し、現行の2項を越えた概念間の関係を得る必要がある。また、概念間の関係を記述するものとしては現在、シソーラス情報である概念体系情報（同義：synonymy、下位：hyponymy、含意：entailmentに相当）と、共起概念間の意味的關係である概念記述情報しか記述されていないが、更に精密な情報を与えるためには、反義（antonymy）全体部分（meronymy）、比喩（troponomy）、等の関

係を概念辞書に記述する必要がある。言語の種類という観点からは、日本語と英語との対応関係が現在語レベルでのみ取られているが、対訳の正確さを求めるためには語同士の関係ばかりではなく、句同士や文同士の対訳関係、更にはこれらの単位間の複雑な対応関係を求めなければならないことが判明しており、対訳関係を記述の単位に沿って大きくして行く必要がある。

4 まとめ

プロジェクトの成果としては、CD-ROM化され外部提供される辞書ばかりではなく、多くの試行錯誤の経験の蓄積というものがある。その何点かを以下にまとめ結びとする。

(1) いうまでもないことではあるが、言語学や辞書学の分野の協力が重要である。言語処理用ということから、辞書の枠組みの設計はコンピュータ・サイエンス側にイニシアティブを与えるべきであるが、中身に関しては、言語学や辞書学の方にイニシアティブを与えるべきである。

(2) これも言うまでもないことではあるが、それぞれの言語に関して、それを母国語とする人々や機関の協力を得ることが重要である。EDRでは日本語と英語を対象にしたが、電子化辞書の日本語にかかわる部分に比べ英語にかかわる部分の品質が少々落ちる。もちろん、EDRとしても、英語を母国語とする多くの人々の熱心な協力を得たが、もう一つ大きな仕組みを作る必要がある。

(3) はるかに低コストで開発する方法が必要である。電子化辞書は大規模で精度が高いものを適切な期間内に出来るだけ低コストで開発されなくてはならない。言語処理技術の研究開発や言語処理応用システムの開発・実用化には、大規模な電子化辞書が必須ではあるが、それらによって得られるゲインは当面のところそれ程大きくないからである。

現在、EDR電子化辞書は評価研究のため、国内外の約60ヶ所で実験的に利用されている。

1995年3月末には、商用版の提供が順次開始される。販売条件は、国内外を含め、出来るだけ多くの所で利用していただけるようなものになる予定である。なお、テキストベースはEDR内部の利用に限られ、ライセンスの対象外である。辞書利用の

詳しい内容は3月29日(水)、東京北青山 TEPIA
ホールにて「EDR電子化辞書発表会」を開催する

ので参加希望者は事務局(千葉または五月女、TEL:
(03) 3798-5521)まで連絡されたい。

《 参 考 文 献 》

- [1] Briscoe, E.: "Lexical Issues in Natural Language Processing," Natural Language and Speech, Springer-Verlag (1991).
- [2] Normier, B. and Nossin, M.: "Genelex Project: EUREKA for Linguistic Engineering," Proceedings of the International Workshop on Electronic Dictionaries, EDR TR-031, Japan Electronic Dictionary Research Institute, Tokyo, pp. 63-69 (1991).
- [3] Serasset, G.: Recent Trends of Electronic Dictionary Research and Development in Europe, EDR TM-038, Japan Electronic Dictionary Research Institute, pp. 25-61 (1994).
- [4] Grishman, R. and Macleod, C.: COMLEX 0.2 - A Draft Specification for a Lexical Knowledge Base, Computer Science Department, New York University (1992).
- [5] Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., and Teng, R.: Five Papers on WordNet, CSL Report 43, Cognitive Science Laboratory, Princeton Laboratory (1990).
- [6] BNC: British National Corpus - Fact Sheet (1992).
- [7] BNC: British National Corpus - Spoken Corpus Design Specification (TGAW 14) (1991).
- [8] BNC: British National Corpus - Written Corpus Design Specification (1991).
- [9] Leech, G.: British National Corpus - TGDW08: Revised Proposal for Basic Grammatical Tagset (1991).
- [10] Marcus, M., Santorini, B., and Marcinkiewicz, M.: "Building a Large Annotated Corpus of English: The Penn Treebank," Computational Linguistics, Vol. 19, No. 2, pp. 313-330 (1993).

21世紀の情報基盤

EDR 電子化辞書発表会

日 時 平成7年3月29日(水) 午後3時～5時
場 所 TEPIAホール(テピア 4階)
(東京都港区北青山2丁目8番44号) (☎03-5474-6111)
(営団地下鉄 銀座線 外苑前下車 青山ラグビー場そば)

発 表 会 1 EDR電子化辞書の紹介
2 EDR電子化辞書の頒布について

(定員) 250名

展 示 会 CD-ROM展示、説明(2階ホワイエ)

参 加 料 無 料

問い合わせ・申し込み 日本電子化辞書研究所(港区三田1-4-28三田国際アネックス)

☎ 03-3798-5521 FAX 03-3798-5335

オンライン機械翻訳セミナー ……佐々木技術翻訳事務所(名古屋市中区葵)では昨年12月からパソコン通信PC-VANの「草の根ネットワーク在宅ワーク」というSIGで機械翻訳セミナーを開講している。このセミナーはTranser/ejを使用し、機械翻訳の解説や現状、機械翻訳利用上のノウハウや問題点もとりあげ、また質問にたいして開発側からの回答も出されている。同じSIGで提供されているホワイトハウス・プレスリリースの原文と佐々木技術翻訳でリライトした機械翻訳文を教材として使用している。またオンラインだけではなくオフライン用のテキストも発行している。(連絡先☎052-935-0287)

タイMMT'95開催 ……多言語機械翻訳シンポジウムが3月8日、バンコクのNECTEC(AAMT会員)で開催された。これは日本の政府開発援助として過去8年間にわたり、インドネシア、マレーシア、中国、タイの4カ国との技術移転を目的とした共同研究開発が終了し、その成果を公に発表したものである。5カ国語同時翻訳のデモンストレーションも行われた。また共同研究に参加した各国の研究者の開発レポートの発表や「最終ゴールはグローバルMT」と題する講演も行なわれた。なお日本側においては昨年12月に成果発表会が開催されているが、現在日本で市販されている機械翻訳がトランスファー方式であるのに対し、このシステムは中間言語方式を採用しており、国際会議など同時多言語の資料や議事録翻訳に大きな効果を発揮するものと期待されている。

韓日、日韓の翻訳実績 ……千葉県松戸市の翻訳家兼通訳の金基東さんは毎年、年間の翻訳実績を関係先に公表しておられるが、韓日翻訳の内容としては「韓国科学物質の規制法案に関する説明資料」「多語調表示装置」「紙幣類の取り扱い装置及び取り扱い方法」「コードレス電話機の技術的条件」など、また日韓翻訳は「伝統芸能で訪れる四季」「国連地域環境子供サミット・イン島根」「カラオケビジネスフェア'95(幕張メッセ)」「韓国中小企業育成事業研修セミナー」「真空処理装置及びこれを用いた生膜装置と

生膜方法」その他多数である。日本と韓国との情報の質の違いが垣間見れる面白いデータである。また金さんは英韓、韓英の翻訳もやっておられる。(連絡先☎0473-85-9645)

電腦21研究会活動 ……電腦21研究会(主宰者・アメリカ大使館小宮山氏、長瀬産業堂野前氏)は「電腦の可能性についてユーザの立場から研究し、また異文化の交流と相互理解の増進」を目的に91年に発足し、すでに5年目を迎える。研究会は四半期毎に開催され、研究の方向は主として翻訳リテラシーの向上にある。毎回専門家を招いてヒアリングを実施した後熱心な討論が繰り返されている。最近の講演内容としては「機械翻訳のゴミ掃除ソフト」(嶋本東洋モートン取締役)、「クリエイティブライティングへのモチベーション」(片山外国語教室片山氏)、「ヨーロッパの機械翻訳の現状」(朝日新聞森本氏)、「異文化コミュニケーションの面白さ」(ユニカルインターナショナル佐々木代表)など興味深い盛り沢山な構成である。この他機械翻訳ソフトウェアやシステムメーカーも招いて、製品動向についての研究も行われている。(連絡先☎03-3459-0071)

バベル翻訳奨励賞表彰式 ……バベル翻訳奨励賞は新人翻訳家の登竜門とするために設けられた奨励賞で今年で19回目を迎える。英日、中日、韓日、日韓の4部門があり、小説や経済問題の課題にたいして審査された。また併せて第5回バベル国際翻訳大賞の審査も行われた。これは日米、日中部門で当該国の日本語への理解を深めることもその目的の一つとされている。毎年各国から多数の応募があり、審査する先生もその選考に頭を悩ますとのことであるが、この表彰式は3月29日に東京市ケ谷のアルカディア市ケ谷で午後5時から開催される。

この事業は主催が髹バベル、部門共催が韓国翻訳家協会、また日本貿易振興会、国際文化フォーラムなど多数の団体が後援している他当協会も後援している。

(連絡先 髹バベル ☎03-3295-2306)

英日/日英翻訳ソフト PC-Transer シリーズ

株式会社 ノヴァ

1991年PC98版英日翻訳ソフトを発売以来パソコンで使える本格的な機械翻訳ソフトとして定評をいただいているPC-Transerシリーズも約5年を経て、日本語・英語OCRソフトも含めた多様な製品群が完成、発売されています。過去本誌に個々の新製品についてご紹介していますが、今回は新製品も含め総括的に現時点での製品をご紹介いたします。

1) 英日翻訳ソフトPC-Transer/ej

(Windows版、Mac版) 198,000円

高速・高精度の翻訳を実現、使い易いユーザインタフェースと以下のような豊富な後編集機能が特長です。

●訳語対応・変更

翻訳後、瞬時に訳語の対応、別訳語表示・変更が可能です。変更された訳語を学習することができます。

●品詞変更

第一訳で品詞の取り違えがあったとき、簡単に品詞解釈を指示・再翻訳させることができます。

●別解釈

文法的に解釈可能な翻訳結果を参照引用することができます。

●ステップ翻訳(図1)

修飾関係が不明瞭なとき、句・節およびその品詞解釈を指定して部分訳を実行、全体訳に反映することができます。

この製品については、

武舎広幸著 「パソコン英日翻訳ソフト活用法」

PC-Transer/ej 実践マニュアル

株式会社プレントリスホール出版発行

株式会社トッパン発売 (3,800円)

に詳しく説明されています。この書籍はPC-Transer/ejの体験版CD-ROM (Windows版、Mac版)が添付されており、翻訳結果の出力・印刷はできないものの翻訳の精度・速度などを評価することができます

2) 日英翻訳ソフトPC-Transer/je

(Windows版、Mac版) 198,000円

科学・技術分野の英文作成に、またマニュアルの翻訳に、好評の日英翻訳ソフト。上述した英日の後編集と同等の機能を持つほか、英語シソーラスで実績のある米国インフォソフト社のシソーラス辞書

(15万語)を標準装備。別訳語の選択に類義語を参照・引用することができます。

最近発行された下記書籍でも PC-Transer/jeのことが詳しく紹介されており高い評価をいただいております。

諏訪邦夫、岩瀬良範監修

加山裕高、長田理、岩瀬良範、諏訪邦夫著

「知的生産のためのパソコン技法」

データ処理と論文・レポート作成法

技術評論社発行 (2,600円)

図1



3) 英日、日英用専門語辞書

(Windows版、Mac版) 30,000~80,000円

PC-Transer用の専門語辞書です。ソフト本体には約6万語の基本辞書が装備されていますが、専門分野の用語を補うためにこの辞書を使用します。18分野の辞書が現在販売されています。(英日用と日英用は別製品)

(対象分野) コンピュータ、機械工学、電気・電子、医療・医学、化学用語、バイオテクノロジー、ビジネス、数学・物理、地球環境、金属、海洋・船舶、航空・宇宙、土木・建築、原子力・エネルギー、防衛、貿易、金融、法律

4) 日英文字認識ソフト (図2)

J-Scan+ (Windows版、Mac版) 69,000円

英日、日英翻訳のための原文入力にはOCR(文字認識)ソフトの利用が効果的です。また、翻訳のみならず印刷文書のパソコンへの取り込みには幅広い需要があり、当社でもこうした需要に応えるべく標記製品を開発、このたび販売を開始しました。

製品の特長として、

- 日本語、英語の活字文字の認識が可能。
- 日本語で99.5%、英語で99.9%と高精度(原稿の鮮明さに依存します)
- 69,000円と画期的な低価格
- TWAIN、プラグインインタフェースに対応しておりほとんどのスキャナが利用可能。
- TIFF、BMP、PICTなどのイメージファイルの認識が可能。

などがあげられます。

5) 英文OCRソフト

英語まるかじり (Windows版、Mac版) 9,800円

上記のJ-Scan+より、英語部分を抜き出したのがこの製品です。

当社も含め、英日翻訳ソフトが各社より販売されておりますが、原文入力のための英文OCRはかなり高価でした。そこで原文入力にポイントをしぼり機能をシンプルにして低価格にしたわけです。

機能は、上記のJ-Scan+に準じます。

6) PC-Transerカップル版

(Windows版、Mac版) 346,000円

PC-Transer/ejとPC-Transer/jeの同梱版です。

7) PC-Transerトリプル版

(Windows版、Mac版) 398,000円

PC-Transer/ejとPC-Transer/jeにさらに、J-Scan+を同梱しました。カップル版、トリプル版ともにそれぞれのソフトを単体で購入するよりも低価格になっています。

以上、製品の概要です。英日と日英、さらに日本語・英語のOCR。また対応機種としてのWindows版、Macintosh版と製品がそろってまいりました。

今後ますます、性能・機能の向上に努めたいと考えております。

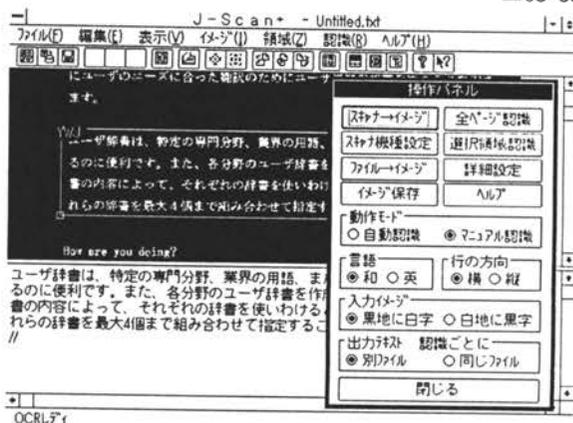
お問い合わせ先

株式会社 ノヴァ システム営業部

〒160 東京都新宿区荒木町23鈴商ビル301

☎03-3351-3356 FAX:03-3351-5766

図2



日英機械翻訳ソフト TransLand for Windows

ブラザー工業株式会社

インターネットなどの電子ネットワークが国境を越えて接続されるようになり、海外とのコミュニケーションの頻度は増え、ますます重要になってきています。

ネットワークの端末として使われているパソコンには視覚的に使いやすいユーザーインターフェイスが普及しつつあります。

そのユーザーインターフェイスの中核になるのが Windows だろうと言われていました。

当社は TransLand の 98 DOS 版を 1993 年に発売し、Mac 版を 1994 年に発売しました。

その翻訳精度の良さは多くのユーザーの支持を得てきております。今回、翻訳精度をより向上させ、且つ、より親しみやすいユーザーインターフェイスで Windows 対応版を発売致します。

Windows 版の特長を以下に紹介します。

■豊富な基本辞書と高品質の翻訳精度

52,000語の豊富な基本辞書。評判が良い翻訳精度にさらに磨きをかけました。

■クリップボード翻訳（新機能）

クリップボードを介して、他のアプリケーションソフトウェアの中から翻訳が可能です。クリップボードにコピーされた文章の翻訳を自動的に開始し、翻訳結果をクリップボードに返すことができます。あなたが普段お使いになるワープロの中から翻訳を行うことができます。つまりいつものワープロに高レベルの翻訳機能が装備できます。

■ユーザーに親切的な操作性

日本語と英文は左右対称に表示し、翻訳し、編集することができます。どなたでもコピー、カット & ペースト、検索や置換など、ワープロのように簡単に編集できます。

■柔軟な翻訳操作

以下の機能などで柔軟に翻訳ができます。単語単位で英訳を表示する「単語翻訳」の機能翻訳した英文で単語の別訳を表示したり、その単語を別訳の単語に置き換える「別訳表示」の機能

■テキストファイルの読み込み

テキストファイルへの出力

既存の日本語テキストファイルを直接読み込んだり、クリップボードを介して取り込むことができます。OCRソフトがテキストファイルを出力できるならば、そのテキストをそのまま利用できます。また、翻訳結果は他のソフトウェアで利用することができます。

■ユーザー単語辞書

専門用語、業界用語や特別な単語はユーザー単語辞書へ最大30,000語まで登録が可能です。複数のユーザー単語辞書を切り替えて使用できます。きめ細かな設定により正確に翻訳ができます。

■ユーザー定型文辞書（新機能）

挨拶文、慣用句等、普段よく使う定型文を最大100文登録できます。定型文内にはワイルドカードを使うことができます。例えば「第*図参照」という日本語に対して「Refer to figure *。」を英文に登録しておきますと「第20図参照」という日本語を「Refer to figure 20。」に翻訳できます（20の部分は何でもかまいません）。

■翻訳速度

[Pentium 90MHz] のパソコンを使うと1文を1秒～10数秒で翻訳します。(翻訳速度は文章の複雑さ、CPUの性能、メモリ容量により異なります)

■専門用語辞書（別売）

ライフサイエンス(医学・生物学を中心に用語を収集) 29,000語

対応する機種と動作条件

Windows3.1の動作するマシン(HDD 20Mが必要です)
CPU486以上推奨

■価格 79,800円(3月末発売)

※この製品に関する問い合わせ先

株式会社エクシング TAKERU事務所
TEL 052-824-2493

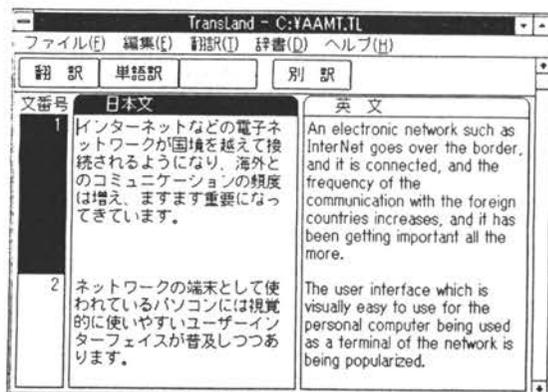


図1 対訳編集画面



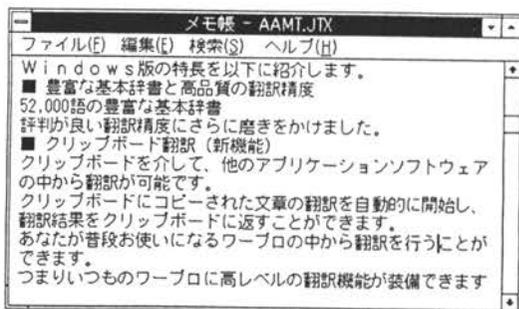
図2 ユーザ単語辞書一覧画面



図3 ユーザ単語辞書登録画面



図4 クリックボード翻訳画面



Windows版 英日翻訳ツール 『コリャ英和！』

カテナ株式会社

1. 翻訳システムから翻訳ツールへ

従来、機械翻訳システムは本格的に「翻訳する」ことがその使命でした。翻訳をやっているときにあれば便利かもしれない機能はどんどん詰め込まれ、その結果、豊富なメニュー、豊富なオプションを誇る巨大な本格的機械翻訳システムが続々誕生してきました。

翻訳の専門家であれば、このような多機能なシステムから必要なものを選び、自分用のシステムとして快適に使いこなせるかもしれません。しかし、一般のユーザにとってそのような機能を活用する機会はきわめて希なことと言えるでしょう。分からない英語に出会ったとき、気軽に辞書が引けたり翻訳してみたりできることの方がずっと重要なのです。

翻訳結果をよくするために、前編集したりするのはまっぴらごめんという方が大部分なのではないでしょうか。本格的で多機能なシステムは、複雑すぎて近寄りたくない印象さえ持たれるかもしれません。潜在的需要の大きさが報告されながら、なかなか機械翻訳市場が拡大しないのは、この辺に遠因があるようにも思えます。

当社では、これらの点を鑑みてもっと気軽に使える翻訳ツールが必要であると考えました。ゲームやパソコン通信等をやっている出会ったちょっとした英語を翻訳したり、辞書引きしたり出来るツールです。こうして、Windows版英日翻訳ツール「コリャ英和！」が誕生しました。機能は必要最小限に絞り込みました。翻訳システムであればあって良い機能もどんどん削られました。一時的な翻訳が中心ということでファイル保存機能もついていません。

(もっとも、Windowsですので、コピーして他のアプリケーションにもって行けばそこからファイル保存は出来るわけです。) そのかわり、従来の翻訳システムでは実現されていなかった「クリップボード監視機能」(後述)がつけました。標準価格も気軽に購入できるよう 9,800円 としました。

このように初めての試みも多く大冒険ではありましたが、おかげさまで好評を持って受け入れてもらっているようです。

2. 翻訳エンジンは本格派

値段も安く、ツールに徹した造りになっているため、エンジンも手抜きしているとは考えられがちですが、実は翻訳エンジンには本格的なものを用いています。

もともと当社にはunixワークステーションで高速・高精度との定評のある「STAR」という翻訳システムが有り、「コリャ英和！」にはこの翻訳エンジンをそのまま載せています。

この手法は、Mac版で定評をいただいています「The Translator」シリーズでも用いたものです。「STAR」のエンジンがコンパクトで、もともとマルチプラットフォーム対応の設計となっているため、開発コストが大幅に削減できたことも幸いでした。

システム辞書の語数も、先日バージョンアップしたThe Translator Ver. 2.5 (6万2千語) よりもさらに増え、6万7千語にパワーアップし精度も上がっています。

3. それぞれ独立した翻訳と辞書引き

翻訳と辞書引きはそれぞれ独立したアプリケーションとなっています。これにより、ユーザが辞書引き中心でも翻訳中心でも軽快に利用できます。(もちろん、翻訳中に辞書引きするのは自動的に辞書ツールが呼び出されますので、めんどろな手続きは一切不要です。)

4. クリップボード監視機能

この機能は、たとえばゲームをやっていたり電子

メールやNewsを読んでいて分からない英語に出会ったとき、その部分をクリップボードに入れる（コピーする）だけで自動的に辞書引きや翻訳してくれるものです。使い方も簡単で、メニューの「クリップボード監視」をチェックしておくだけです。あとは、お好きなゲームで遊ぶなり電子メールなどを読んで下さい。そして調べたい英語が出てきたら、それをコピーしてみましょう。

この機能により、普段はめだたないが必要なときに登場してくれる翻訳家や辞書引き専門家がデスクトップの片隅にいてくれるわけで、一度使えばその便利さを納得していただけるでしょう。

もちろん、英文テキストを読み込んで翻訳するという翻訳システム的な使い方も出来ます。

5. 翻訳ツール

翻訳ツールの主な特徴をご紹介します。これで全てではないのですが、前述したように、複雑で近寄りたくなるのは避けたいのです。

・左右／上下の分割可

最近では、対訳形式の左右分割がはやっているようですが、敢えてどちらの分割もできるようにしました。片方だけとすべき明確な理由が見いだせなかったためです。

・5枚の翻訳パッド

最大5枚までの翻訳画面が使えます（うち1枚は、クリップボード翻訳と兼用）。これらは独立していて、翻訳ツールの終了時に自動的に保存されます。（前述のファイル保存無しをこれで若干補っています。）1パッドに英文、訳文それぞれ32KBまで入れることが出来ます。

・一括／対話型どちらの翻訳も可能

英文の一部または全部を選択して連続翻訳することも、一文ずつ対話的に翻訳することも出来ます。

・英語混じりモード翻訳が可能

辞書にない未知語は当然原語のまま訳に出て来ますが、もっと積極的に基本的な名詞と動詞を訳さない翻訳が可能です。これはもともと「STAR」に有った機能をそのまま「The Translator」も含め受け継いでいるのですが、特に専門用語の多い文を専門用語なら原語のままでもよい人が読む場合きわめて有用な機能で、変に日本語にするよ

りよほど読みやすいとの評価を頂いております。

・後編集と辞書登録

だんだんツールよりも本格的なイメージになってしまっていますが、訳語と原語の対応、訳語置換などが出来ます。学習辞書への登録もできます。

6. 辞書ツール

辞書ツールの主な機能は以下の通りです。

・活用形での辞書引き

原形だけでなく活用形でも辞書引き出来ます。特にクリップボード辞書引きのときに有用です。

・曖昧検索が可能

はじめの数文字を入力すると該当する単語リストが表示され、リスト中の語をダブルクリックするだけで辞書引きが出来ます。

・ヒストリー機能

辞書引きした語を最大10個まで覚えています。

7. 機械翻訳の普及に向けて

現在の機械翻訳はまだ残念ながらよちよち歩きの段階です。機械翻訳にかなりの期待を持って見る人は、皆一様に翻訳結果に失望するようです。訳文を直すくらいなら、自分ではじめから翻訳した方が早いと言った声も聞かれます。

「コリャ英和！」が普及することで、現在の機械翻訳のレベルを多くの人が知るようになれば、過大な期待はなくなります。逆に、現段階での水準をふまえた上での導入の検討となれば、失望することがないだけ話が進みやすいと言えます。

もちろん、機械翻訳の水準が上がりユーザが上質の翻訳結果で選択できるようになることが理想ですが、それまでは「コリャ英和！」に普及の先導をしてもらいましょう。

【問い合わせ先】

〒206 多摩市落合1-15-2

カテナ株式会社

ソフトウェアプロダクツ事業本部

カスタマーサービス部(松田健生)

☎0423-38-1118

英和・和英辞書ソフトウェア「電辞海 for Windows」

富士通株式会社

1. いま何故、辞書ソフトウェアか？

近年、国際化にともない、学生などの個人レベルから企業のビジネスのレベルにおいてまで、海外から情報を入手する機会が著しく増加しています。そのような海外からの情報を翻訳する場合、専門的な用語（技術用語など）は、従来だと、机の上の分厚い辞書をいちいち引かなくてはなりません。パソコンが個人市場に浸透した最近では、このようなニーズに対応して、CD-ROM辞書など、パソコン上で単語を検索できるソフトウェアが数多く商品化されています。しかし、現在提供されている辞書ソフトウェアの多くは、単語数が絶対的に少なく、専門用語を十分に網羅していないため、一部検索できない単語があるなど、不十分な点がありました。

『電辞海 for Windows』は、基本用語20万語、ビジネス・経済用語から医学・薬学用語まで5種類220万語の専門用語を収録した英和・和英辞書ソフトウェアです。合計240万語を収録した辞書（広辞苑11冊分）は業界随一であり、従来のCD-ROM辞書などでは検索できなかった用語を、パソコン上で簡単に検索することができるので、英文に慣れていない学生から技術英語・商業英語の翻訳家、企業・研究所で日々海外の最新情報に接している企業内個人まで、幅広く翻訳にご利用いただけます。

また、本製品は、訳語の他、類義語や用例を表示

できる7種類の多彩な検索機能を提供しています。例えば、ワープロ連携機能を使えば、海外のデータベースで検索した英文テキストを画面に出しながら、その中の単語を指定して、ワンタッチで辞書引きすることができます。複数形や過去形などの活用した単語も原形に直して検索されますし、用例を引用することもできるので、翻訳作業を非常にスムーズに進めることができます。ワープロで作成した和文の手紙などから英文を作成するような場合も、和文中の単語を次々と検索しながら、しおりをはさんだり、類義語を表示させることができ、英文作成作業を強力にバックアップします。

さらに、従来提供のCD-ROM辞書では不可能だった、自分だけの用例や訳語を辞書に書き込むことができるので、辞書にノウハウを蓄積させながら、快適な文書作成環境を作り上げることができます。

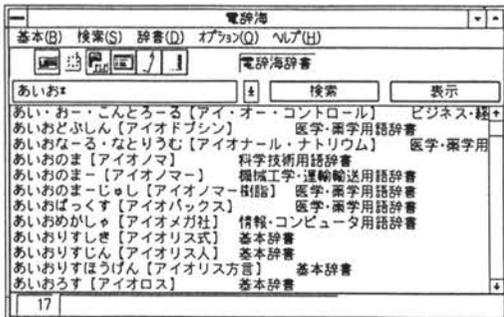
2. 電辞海 for Windowsの特長は？

(1) 充実した基本辞書

英和・和英計20万語の基本辞書を備えており、訳語の他、類義語や用例を表示することができます。

(2) 豊富な専門用語辞書

5分野合計220万語の専門用語辞書をオプションとして提供しています。専門用語辞書では、訳語と使用分野が表示されます。業界一の豊富な語彙



で幅広い分野の用語をカバーしています。

(3) 多彩な検索機能

複数形や過去形などの活用形から検索できる原形検索、ページをめくるように次々と表示できる次ページ検索など7種類の多彩な検索機能を持っています。

(4) 個人辞書の利用

ユーザごとに専用の辞書を作成することができます。新語やユーザ固有の単語を登録したり、検索した内容に特殊な訳語や自分だけの用例を書き込むことができます。

(5) 複数辞書の同時検索

検索キーを入力するだけで、電辞海基本辞書、電辞海専門用語辞書(5分野)、個人辞書の複数の辞書から同時に検索することができます。

(6) ワープロ連携

スクリプトやマクロを利用して他のワープロ中の単語を簡単に検索することができます。また、クリップボードを経由して、他のウィンドウ上の単語を検索することもできます。

(7) 検索の記録

検索を行った単語の履歴が残ります。その中で特に、後で参照したい単語に対しては「しおり」をはさむことができます。

(8) 印刷

検索した単語の内容を手軽に印刷することができます。

3. 利用形態は?

- ・英文読解時の英和辞書として
- ・英文執筆時の和英辞書として
- ・翻訳ノウハウの蓄積による翻訳支援ツールとして



- ・用例、類義語の表示による文書作成支援ツールとして

4. 利用によってどんな効果が出ますか?

- ・紙の辞書で苦勞して検索する手間が省けます。
- ・ワープロからワンタッチで検索。スマートな翻訳作業が可能です。
- ・登録機能により、自分だけの辞書を作成することができます。

5. 動作環境

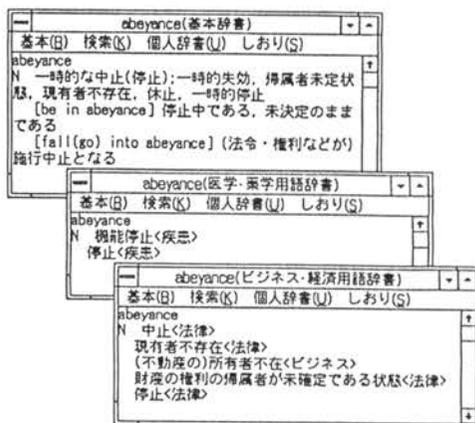
日本語Microsoft Windows V3.1が動作する機種
CPUがi386TM以上のマシンを推奨
メモリ 5.5MB以上
ディスク容量 16MB以上

6. 価格

電辞海 for Windows (基本20万語添付)
4万8千円
情報・コンピュータ用語辞書 23万語 6万円
科学技術用語辞書 73万語 6万円
機械工学・運輸送用用語辞書 29万語 6万円
医学・薬学用語辞書 59万語 10万円
ビジネス・経済用語辞書 42万語 6万円

[お問い合わせ先]

富士通株式会社 ソフトウェア販売推進部
〒222 横浜市港北区新横浜2-15-16
(日興不動産ビル)
TEL 0120-247-301 (フリーダイヤル)
お問い合わせ時間
月～金(除祝日) 9:00～12:00 / 13:00～17:00



英文翻訳支援ソフトウェア WD-01SW

シャープ株式会社

英文翻訳支援ソフトウェアWD-01SWをオフィスワープロWD-MF01のオプションとして発売を予定しております。

1. 企画意図

当社ではビジネスマンを対象に英語の使用状況についての調査を実施致しました。

- ・約65%の人が英語をよく使用、使用内容は文献や資料を読むことが一番多い
- ・約84%の人が英語で困っている、内容は翻訳に時間がかかるのが一番多い

となりました。

この結果を参考に、商業翻訳用としてではなく、ビジネスマンの英語に携わる業務の手助けを行い、業務の時間短縮、効率アップを提案できる商品として企画致しました。

2. 製品の特長

本商品は、翻訳専用機として評価の高い当社のワークステーションベースの機械翻訳システム『DUET QT』のエンジンを採用しておりますので、高い翻訳精度を持っており、更にビジネスマンに手軽に使って頂き、業務の効率アップを図って頂くための操作性を重視しております。

【システム構成】

オフィスワープロWD-MF01をベースにイメージスキャナJX-330M、レーザープリンタWD-06LP(A3)/WD-02LP(B4)/WD-07LP(A4:4月発売予定)、光磁気ディスクドライブが利用出来ます。

【書院WD-MF01との融合】

WD-01SWはオフィスワープロWD-MF01の優れた操作環境と融合し、入力から翻訳支援処理、編集、保存、検索まで一連の流れでデータの活用が出来ます。従いまして翻訳支

援の部分だけではなく、文書作成/文書処理といった総合的な観点からもパソコンソフトや翻訳専用機にない優れた操作性を実現しております。

【入力】

書院WD-MF01の優れた入力環境で、キーボードからの入力、通信を利用した入力は勿論、印刷物をスキャナから読み込み簡単にワープロに取り込みます。

更に、スペルコレクターも搭載しておりますので、入力した英文のチェックも簡単です。

【翻訳支援】

1. 訳振り機能

英文を完全に日本語に翻訳するのではなく、単語ごと、またはフレーズごとに訳を行い、各単語、各フレーズの下に表示します。特に単語訳振りについては訳を行う単語のレベル(当社で独自に設定)を3段階に設定でき、自分の英語力に合わせて訳を表示することができます。

従いまして、この機能は英文を中心に読み進めた時に、ちょっとこの単語の意味を知りたいといった場面に効果的で、スピーディーに意味を理解できます。

2. 翻訳支援機能

英文を日本語に翻訳し、結果を画面下半分に出力します。対象範囲の指定は、全文一括は勿論、範囲指定をしたり一文だけを切り出して処理をすることが可能です。※

また、出力結果はそのまま利用することもできますし、原文(英文)に取り込んで自由に編集することもできます。※

※:訳振り機能も同様です。

3. 英和辞典機能

英文の単語を指定し逐次辞書引きを行い意味だけでなく品詞や変化形、熟語、類語対語等の表示を行います。

【基本機能】

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。
2. 基本辞書
105,600語
3. ユーザー辞書
2万語（名詞／形容詞）：辞書設定の操作

性を重視

※発売時に商品の仕様変更になる場合もあります。

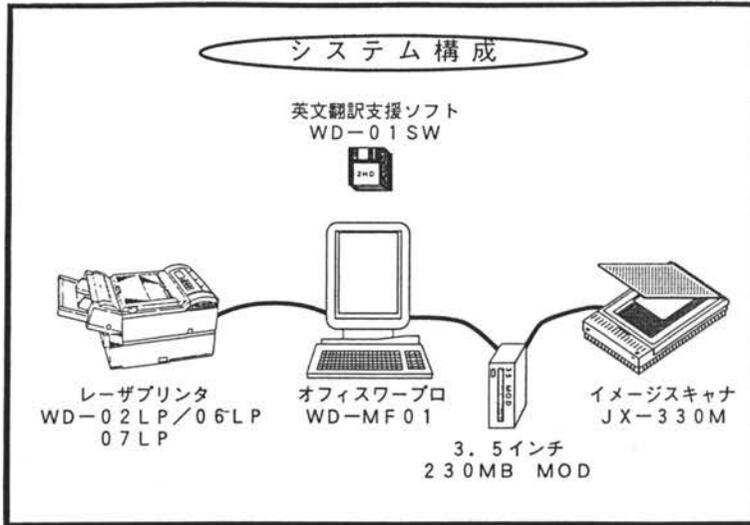
問い合わせ先

シャープ株式会社コンピュータシステム営業部

☎ 043-297-1221

コンピュータ商品企画部

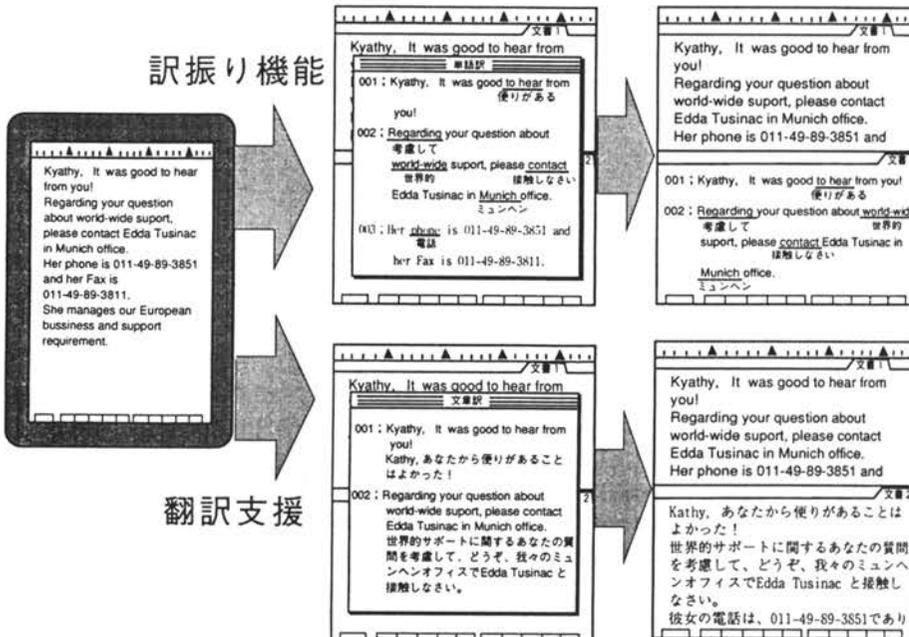
☎ 07435-3-5521



画面イメージ

訳振り機能

翻訳支援



日英/英日翻訳支援ソフトウェア PENSEE for Windows V2.0

沖電気工業株式会社 製品開発部 沖ソフトウェア部

PENSEE for Windowsは、(1)お手持ちのパソコンで(2)使いたれたワープロで、(3)手軽に使えることを目的として開発された日英/英日翻訳支援ソフトウェアです。今春にはより使い易く、高品質な翻訳を目標にバージョンアップを予定しております。

ワークステーション版で高い翻訳品質との評価をいただきましたPENSEEの翻訳エンジンをWindows上に移植し、Windowsの操作環境で本格的な翻訳ソフトを実現しました。また、Wordからの翻訳実行やEXCELでの辞書編集も可能です。バージョンアップで一太郎Ver6からの翻訳実行も可能となりました。

<特長>

● Windowsでの簡単操作

専用エディタは、Windowsの操作性を活かし、メニューバーによる編集、翻訳、ファイルへの保存など、初めての方にも操作は非常に簡単です。Ver2ではツールバー表示を実現し、さらに使い易くなりました。また、単語単位で原文と訳文の対応をとれるようになりました。

● 辞書引き機能

PENSEEの辞書を利用した辞書引き機能を装備しました。これにより柔軟な訳文を作成できます。

● PENSEEの高品質をそのまま利用

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

● 訳質に差がつくシステム辞書の拡張

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

7分野約10万語の専門用語辞書を用意しております。Ver2では登録内容を見直し、より高品質な翻訳を実現します。

● 柔軟なユーザ辞書指定

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

● パソコンでワークステーション並みの高速翻訳

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

● 簡単辞書登録のための未知語検索

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

● Word、一太郎との連動

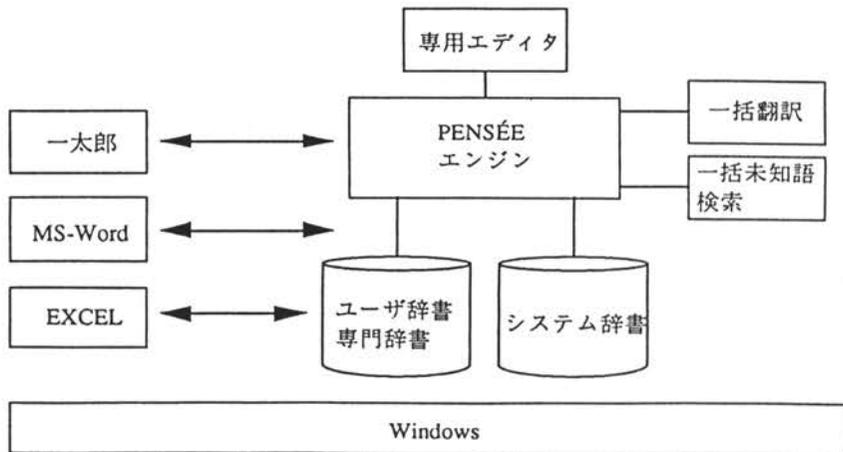
Wordや一太郎のマクロ機能を利用して直接翻訳を実行することができます。テキスト以外にWordや一太郎のファイル形式の文章を翻訳することができます。

● Windows上のアプリケーションとの連携

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

● 一括翻訳、一括未知語検索

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



PENSÉE for Windows の構成

<仕 様>

対象言語：日→英/英→日

ハードウェア：

日本語Windows3.1が稼働する機種

メモリ：8MB以上

ディスク：40MB以上（片方向）

OS：Windows 3.1

システム辞書：

日英：約15万語

英日：約13万語

専門用語辞書（オプション）：

日英：7分野、約10万語

英日：7分野、約10万語

（電気、情報通信、医療、機械、化学など）

ユーザ辞書：

語数無制限（ディスク範囲内）、

最大16個まで同時使用可能

処理速度：約15,000語/時間（平均）

（機種によって異なります）

翻訳オプション：

日英：文体（標準/命令）、文区切り文字指定

英日：文体（標準/敬体）、固有名詞訳出

（英語のまま/訳語のみ/訳語（英語）/

英語（訳語））、文区切り文字指定

他ソフトとの連携：

Word 5.0（マクロ機能による翻訳実行）

EXCEL 4.0（マクロ機能による辞書の編集）

一太郎6.0（マクロ機能による翻訳実行）

<価 格>

PENSÉE for Windows ￥198,000（片方向

専門用語辞書（7分野パック）￥48,000（片方向

※ PENSÉE は、沖電気工業（株）、大阪ガス（株）オージス総研の登録商標です。

※ MS及びMicrosoftは、米国マイクロソフト社登録商標です。

※ 一太郎は、ジャストシステム社の登録商標です。

<問い合わせ先>

沖ソフトウェア株式会社

営業部 営業第2グループ

TEL. 03-3454-7831

FAX. 03-3798-7088

株式会社オージス総研

アドバンストシステム本部

アドバンスト営業部

TEL. 03-3779-2541

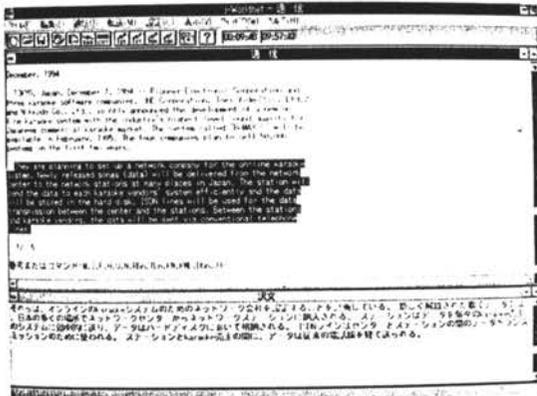
FAX. 03-3779-2543

日本語WINDOWS 3.1対応 新次元の英日翻訳ソフト
「j・London E/J」

髙 電 社

業界初のオンライン翻訳機能を実現

AP、ロイター、ダウジョーンズ等の英文情報が日本語で読めるニューコンセプトの翻訳システムです。



オンライン翻訳画面

■オンライン翻訳機能

オプションのパソコン通信ソフトから、「j・London E/J」の翻訳機能を起動し、通信中に翻訳を行うことができます。これによりパソコン通信から英文データベースにアクセスし、取り込んだ英文データを、ネットワークに接続したまま、その場で翻訳し通信画面とあわせて訳文を参照することができます。

■対訳エディタ

原文と対訳が左右対照に表示されるので原文と訳文を見くらべながら、前編集→翻訳→後編集がスムーズに行えます。

■訳語選択、訳語学習

翻訳後の訳語の変更を辞書から簡単に選択できます。さらに選択した訳語は次回の翻訳から第一訳語として学習させる事ができます。

■訳出指定

訳文に対して、訳し方を指定できます。

1. 訳文体指定 : である調、ですます調

2. 命令文文体指定 : 下さい/すること/する
3. Youの訳出指定 : 主語Youの訳出(する/しない)を指定

■ユーザー辞書登録

名詞、動詞、形容詞、副詞、単位の指定や各種活用形の指定ができます。登録語数には制限がなく、最大3つまでユーザー辞書を設定できます。

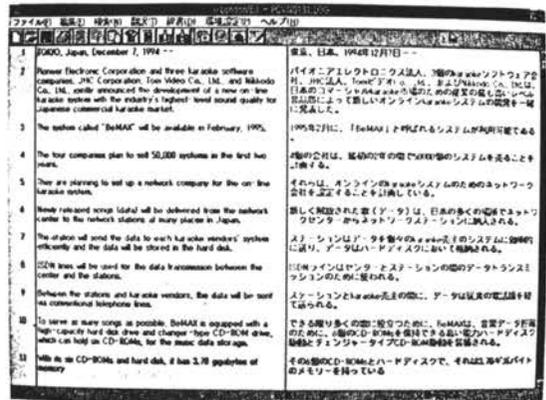


図 対訳エディタ

■オプション専門用語辞書(31分野)

医学・薬学	208,000語	200,000円	化学	11,400語	10,000円
ビジネス	46,400語	40,000円	原子力	4,000語	10,000円
環境	3,100語	10,000円	光学	2,800語	10,000円
金融	2,700語	10,000円	物理	3,700語	10,000円
法律	4,200語	10,000円	プラント	17,700語	20,000円
軍事	22,700語	20,000円	鉄道	1,600語	10,000円
航空・宇宙	78,500語	70,000円	船舶	10,900語	10,000円
情報処理	48,000語	40,000円	繊維	3,900語	10,000円
金属	6,100語	10,000円	自動車	10,500語	10,000円
パイオ	19,200語	20,000円	気象	1,900語	10,000円
建築・土木	13,600語	10,000円	ガス	8,800語	10,000円
電気	23,700語	20,000円	LAN	3,200語	10,000円
エネルギー	3,500語	10,000円	科学一般	20,800語	20,000円
機械	20,900語	20,000円	ANSI用語	5,100語	10,000円
数学・論理	3,000語	10,000円	MIL用語	11,700語	10,000円
計測	2,400語	10,000円			

(表示価格には消費税は含まれていません)

<動作環境>

- OS/日本語MS-WINDOWS 3.1以上
- CPU本体/日本語MS-WINDOWS 3.1以上が動作可能なもの
- メモリ/12MB以上
- ハードディスク/50MB以上の空き容量が必要
- プリンタ/日本語MS-WINDOWS 3.1以上で使用可能なもの
- マウス/日本語MS-WINDOWS 3.1以上で使用可能なもの
- フロッピーディスク/3.5インチ1.44MBが読み込み可能

なもの●供給媒体/3.5インチ(1.44MB)

価格:98,000円

*オンライン翻訳のための専用パソコン通信ソフト「j・Worldnet」が、限定5000本迄、無料で添付されています。

商品の問い合わせ先

髙電社(大阪市阿倍野区昭和町3-7-1)

☎06-628-2351 FAX 06-628-2351

協会活動報告

- 第2回理事会 3月23日 ①95年度事業計画案②95年度予算案③委員会規定④事務所移転案⑤理事退任審議⑥決算理事会、通常総会日程審議
- 運営委員会 1月18日 ①委員会規定審議②組織強化③IAMT関連④財務状況報告
2月22日 ①95年度事業計画案②事業予算案審議③会員の動向④事務所移転案⑤技術小委員会設置⑥理事会開催要領
3月13日 ①95年度事業計画案②95年度予算案③IAMT会費納入④技術小委員会報告
- 技術小委員会 3月10日 ①研究会検討テーマの発掘②テーマ概要の意見交換
- 編集委員会 2月20日 ①AAMTジャーナルNo10編集方針②英文版の取り扱い③MTニューズレター対策
- 需要予測研究会 1月24日 ①ヒアリング「MT, HTの生産性比較」(東芝天野氏)②MT生産性向上の要因分析手法
2月15日 ①ヒアリング「MTの品質評価」②MT生産性要因討議③生産性測定手法検討
3月24日 ①MT生産性の数値データの検討②予測手法の検討
- 例文評価研究会 3月27日 ①翻訳困難例文の言い換え例②言い換え例に基づくネイティブチェック結果と検討

AAMT
ジャーナル
No. 10
(MAR. 1995)

発行 アジア太平洋機械翻訳協会
所在地 〒103 東京都中央区日本橋小舟町4-10 大宮ビル 3階
☎03-3664-5637/5638 Fax 03-3664-1352
E-Mail (PC-VAN) JRD08020 (NIFTY-SERVE) KGE00013

編集委員会 野村浩郷(委員長) 杉山健司 亀井真一郎 福持陽士
事務局 星野禎男 西郷容子

