



AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.16
September 1996

アジア太平洋機械翻訳協会

目 次

退任挨拶	会長退任にあたって	1
就任挨拶	機械翻訳システムの今後	2
通常総会	第6回通常総会の報告	3
事務局だより	役員および委員の紹介	4
成果報告	95年度成果報告会の報告	7
委員会報告	95年度市場動向調査委員会成果報告	9
ヒアリング	翻訳ソフト市場の現在	12
	インターネットと機械翻訳(その2)	14
技術早分かり	機械翻訳と文法	16
研究機関紹介	郵政省通信総合研究所関西先端研究センター 知的機能研究室	21
商品一覧表	機械翻訳関連ソフトウェア一覧表	24
新製品紹介	英文作成支援ソフトウェア「日英辞書引き君(単文変換版)」	28
	翻訳アダプタⅡ	30
イベント紹介	JTF 翻訳祭	32
	データショウ'96	20
新刊図書紹介	言葉でおぼえる Word95入門 for Windows	3
	自然言語処理	11
	プログラミングの壺	13
TV放映	徹底解剖自動翻訳ソフト	8
事務局だより	協会活動報告	23
	ホームページを開設	6
	アンケートにご協力のお願い	目次下

アンケートにご協力のお願い

AAMT 会員各位

AAMT 編集委員会

この機関誌が、すこしでも多く会員の皆様方のお役に立つようと、このたび、アンケート調査を行うことにいたしました。ご多忙のこととは存じますが、何卒ご協力のほどよろしく願いいたします。

以上

《宛先》

アジア太平洋機械翻訳協会事務局
〒105 東京都港区芝公園3-5-12
芝公園真田ビル3階
FAX: 03-5473-0569

(回答は'96年10月31日までお願いいたします)

会長退任にあたって

長尾 真

このたび、アジア太平洋機械翻訳協会会長を退任させていただくことになりました。2期4年間の会員諸兄のご協力に感謝申し上げます。新しい会長は、東京工業大学の田中穂積先生であります。いろいろと新しいアイデアを持っておられます先生のもとに、本協会が新しく発展してゆくことを期待いたします。

機械翻訳システムは、1980年代後半から90年代の初めまで多くの期待を担ってマーケット形成が行われましたが、翻訳品質の改善という課題を容易に解決できず、その後は苦しい時代を過ごしてきました。本協会もスタート時点では多くの期待を集めました。その後はどちらかというと縮小ぎみで今日まで推移してきております。協会を設立するとき、会員としては、翻訳会社のかたがたにたくさん入っていただき、機械翻訳システムの利用法についてのノウハウをいろいろと交換していただく場が提供できること、また、機械翻訳システムのメーカーとユーザーの間で意見を交換していただく公の場となることを期待していましたが、それが必ずしも成功したとは言えないのは残念なことであります。会長としての力不足を会員のみなさまにおわび申し上げます。

ただ幸いなことに、インターネットが最近急速に社会に広がってきたおかげで、インターネット上の情報の英日機械翻訳システムが注目されるようになってきております。一説によりますと、昨年一年間に約50万本の機械翻訳ソフトが売れ、本年度には150万本が売れるだろうと予測されております。これは機械翻訳システムが従来の大型コンピューター、ワークステーションからパソコンの上に乗るようになったこと、そして利用者が企業から個人に移って来つつあるというところに大きな原因があるでしょう。いずれにしても大変ありがたいことであり、ここまで機械翻訳システムを育ててこられた各メーカーの担当者の努力に頭の下がる思いがいたします。

いずれにしても、時代が急激に変化して来ていることは事実であります。こういった時に新しい会長

のもとに新しいアイデアで個人のユーザーにまで活動範囲を広げた新しい協会に脱皮して行っていただくことが大切かと存じます。

委員会および研究会も3つが新しく出発しなおし、会員のみなさまにとって真に役に立つ情報の収集と内容の検討、問題点の明確化と解決に向かって精力的な活動を開始してござっております。このように新しい会長と役員諸氏、新しい委員会活動の下に協会は再出発しておりますので、会員のみなさまのより一層のご協力をお願いいたしますと存じます。

機械翻訳のための基礎研究も、大学・研究所などで着実に進んでおります。形態素解析や構文解析の精度は飛躍的に向上しておりますし、訳語の適切な選択法や文脈を考慮した翻訳などの研究にも着実な成果が上がってきておりますので、そのうちに従来のものに比べて飛躍的に質のよい翻訳を可能とする、新しい考え方の機械翻訳システムが出現するでしょう。一方、21世紀に入れば国際化がますます浸透してゆくことは間違いありませんから、機械翻訳システムはますます要求されるようになるでしょう。そして日本語と英語の間の翻訳だけでなく、他の多くの言語、特に近隣諸国の言語間との翻訳システムが重要になってゆくものと思われまます。

言語は人間というものを代表し表現する道具でありますから、これを機械的立場から扱うことは簡単ではありません。チェスを指すプログラムは、これまで30年間の努力によってようやくチェスの世界チャンピオンと同等のところまで来たわけですから、言語を扱う能力の獲得には50年、100年くらいかかるとみなければならぬでしょう。しかし、努力をすれば人間に近づくことができるということは間違いのないことと存じます。息の長い努力が必要であります。そういった中で本協会の役割はますます重要になってゆくでしょう。

新会長のもとに本協会がますます発展してゆくことを期待いたします。

機械翻訳システムの今後

田中穂積

インターネットによる通信の便利さが分かるにつれて、これを使う人の数は爆発的に増えている。これは通称パソコンが家電製品なみに売れて、膨大な数のパソコンが家庭に入りつつあることと無縁ではない。光ファイバーによる情報ネットワークが各家庭にまで入り込んだ情報インフラが整備される時代が来れば、高品質の音声や映像、画像などの情報が、その所在を問わず、家庭にいながらにして容易にアクセスすることが可能な本格的なマルチメディア時代が到来する。

人と人との間のコミュニケーションの形態として最も自然な形態が、言語によるものであるとすれば、通信を行なうためにネットワーク上を駆け巡る情報の大切な部分は自然言語で書かれているとよい。インターネットにより、情報が世界のどこにあってもアクセス可能であるというお題目を唱えてみても、現実の情報の中身は、それぞれの国の言葉で書かれた自然言語の文であるから、アクセスした側からすれば、それを自分の国の言葉に翻訳する必要がある。最近、低迷していた機械翻訳システムの市場が、インターネットがらみで息を吹き返し活を取り戻しているのは自然の成行きであろう。地球規模のインターネットは、言語の相違を解消する手段としての機械翻訳システムにとって、絶好の追い風になっている。その背景として、ネットワークの他に、機械翻訳システムそれ自体の価格が急激に低下し、一般の人が容易に入手可能になったことも大きな要因としてあげられる。

機械翻訳の品質についてはどうだろうか。5年前の機械翻訳システムと比べて確かに品質は向上している。しかし到底十分であるとはいえない。それにもかかわらず使用者からの不満は以前ほど厳しいものがないという。価格が今少し高めればおそらく不満を漏らしたであろう人も、価格が安くなったために、この程度の価格だからこの程度の翻訳の品質でもしょうがないということで不満が少なくなったのであろうか。過去には、機械翻訳システムの使用者

からの厳しい注文が、翻訳の品質の向上に大きな影響を与えた。使用者側からは、購入した機械翻訳システムに今後とも引続き厳しい注文をつけてもらいたいものである。

機械翻訳技術についてはどうだろうか。残念ながら、技術としては80年代のものがベースになっている。最近の機械翻訳システムの翻訳品質の向上の多くは、辞書の内容の充実に負っている。機械翻訳システムの性能向上をはかるためのブレイクスルーとなる技術の研究は、おそらく意味の問題に深く踏み込むことになる。

最近、コンピュータ技術の発展により、これまでは考えられないほどの高速で多量のメモリーが安価に手に入るようになってきた。それを受けて、力づくの方法で機械翻訳を行なおうとする技術が盛んに研究されている。言語データに潜む統計情報を用いる方法はその一つである。翻訳例文を多量に収集し、それを利用しようとする方法もその一つである。これらの方法は意味の問題を直接扱う必要がないこともあって注目されている。筆者の偏見かも知れないが、皆が皆、この力づくのこの方法に取り組む必要はないとも思う。ところが現実には、必要以上に多くの人が、この力づくの方法に取り組んでいるのではないだろうか。困難ではあっても意味の問題を深く追求する研究が、ややなおざりにされているように思う。もっともこの困難な問題に地道に取り組む研究者技術者がいるべきだと思う。そこから、次世代の機械翻訳システムのためのブレイクスルーとなる技術が間違いなく生まれる。

こうした地道な研究を進める上で障害となるものは、意味理解に踏み込んだ計算機用の辞書のインフラである。意味理解に踏み込んだ辞書の開発は容易ではない。そこで既存のものとして、電子化辞書研究所で開発した辞書(EDR辞書)を利用することが考えられる。現時点でのこの辞書の品質は十分とはいえないが、品質を今少し高めれば、十分、次世代の機械翻訳システムを生み出す情報インフラとな

りうと思う。この時忘れてならないことは、辞書インフラは皆で使って厳しい注文を出し、世の中の標準となるように育て上げることである。EDR 辞書をその実験台とする価値はあると思う。

以上、最近考えていることを思いつくままに述べた。最後になったが、本協会に、機械翻訳の開発者、研究者、使用者とが活発に意見を交換するムードが

醸し出されれば、それが機械翻訳システムの未来にとって極めて望ましい効果をもたらすということを目指しておきたい。協会会員相互の意見交換の場として、本誌を役立てることも考えられる。シンポジウムを開くことも有効だろう。その方向に微力ではあるが会長として努力してみたい。微力ではありますが会員の皆様の御協力をよろしくお願いします。

通常総会

第6回通常総会の報告

第6回通常総会が、96年6月17日（月）16時5分から、東京都港区芝公園の機械振興会館地下2階大ホールで開催された。

出席者数の確認により本総会の成立が宣言されたのち、会則により長尾会長が議長に就任した。また、議事録署名人として長尾真会長、田中穂積理事および小谷泰造理事が選任された。議事に入り、

第1号議案 95年度事業報告および決算報告

第2号議案 96年度事業計画案および収支予算案

第3号議案 役員改選案

が順次提出された。それぞれ説明、審議ののちいずれも可決承認された。なお、総会案内では第4号議案として「会則の変更」が予定されていたが、その後の検討の結果変更の必要がなくなったため取り下げる旨議長から報告された。続いて出席者に上記以外の議案の提出が諮られたが、提案はなかった。

ここで約5分間の休憩に入り、この間に本日選任された新役員による臨時理事会が別室で開かれ、会長および副会長が選出された。

再開された総会で、新会長に田中穂積理事、副会長に小谷泰造理事と岡村正理事が選ばれたこと、また、後日大学関係者から1名副会長を選出する予定

であることが議長から報告された。このあと、田中穂積新会長が「来年のMTサミット、更にその次のアジアでのMTサミットへの対応を果たしてゆきたい。そのほかは長尾先生の敷かれた路線に沿って発展させたい。みなさまのご支援ご協力をお願いします。」とあいさつした。

これで議事は終了したが、引きつづき長尾前会長が退任のあいさつをした。

「もっと発展すると期待してやってきたが、すんなり世の中に受け入れられるまでにはならなかった。一般に新しい技術は芽生えてから実用になるには40年～50年かかる。日本の機械翻訳の技術力や実績は世界をリードしてきた。今後も新会長のもとでリーダーシップをとって発展させてほしい。

企業の立場からは利益が出ず苦しいかもしれないが、あせらず我慢して育ててほしい。10年～20年先には利益をもたらすと確信している。今後国際化が進み機械翻訳は欠かすことのできないものとなり当協会の重要性はますます増大する。発展を期待している。」

このあと、田中新会長から長尾前会長に記念品が贈呈され、総会の行事は17時に終了した。

新刊図書を紹介

■書名：言葉でおぼえるWord95入門 for Windows

著者：安藤 進 青山学院大学理工学部講師

発行所：株式会社富士通経営研究所

概要：

文章を書くことの多いビジネスマンやライターを念頭に、言葉の解説をまじえながらWordの世界を案内する。

約30分の拾い読みで、Wordになにができるのかについて、大体の感じがつかめる。

約1時間ほどで読み物として読み通せる。言葉の基本的な意味とコンピューターやWordの世界に特有な意味と用法がわかる。

例題を自分でやってみることによって、約3時間で一応のコツがのみ込める。

役員および委員の紹介

総会および理事会で新しい役員が選出され、今後2年間の体制が確立しました。(順不同)

〈理事〉	会長	田中穂積	東京工業大学大学院情報工学研究科	教授
	副会長	小谷泰造	株式会社インターグループ	代表取締役
	副会長	岡村正	株式会社東芝	常務取締役 情報通信制御システム事業本部長
	副会長	辻井潤一	東京大学 大学院理学系研究科情報科学専攻	教授
	理事	野々内隆	株式会社日立製作所	専務取締役
	理事	吉原忠雄	富士通株式会社	常務取締役
	理事	浅田篤	シャープ株式会社	副社長
	理事	榊靖夫	沖電気工業株式会社	取締役 研究開発本部長
	理事	吉川英一	日本電気株式会社	常務取締役
	理事	長澤雅浩	松下電器産業株式会社	取締役 開発本部長
	理事	田中達雄	日本電子工業振興協会	専務理事
	理事	長尾真	京都大学 工学部工学研究科	電子通信工学専攻 教授
	理事	野村浩郷	九州工業大学 情報工学部	知能情報工学教室 教授
	理事	石崎俊	慶応義塾大学	環境情報部 教授
〈監事〉	監事	佐藤清俊	日本電子工業振興協会	常務取締役
	監事	勝田美保子	株式会社十印	代表取締役社長

委員会および研究会のメンバーは95年度とほぼ同じです。(順不同)

〈運営委員会〉

委員長	奥平捨男	株式会社日立製作所	情報事業本部公共企画部	部長代理
委員	羽下雄之輔	沖電気工業株式会社	研究開発本部	主席研究員
委員	鈴木等	シャープ株式会社	情報システム事業本部情報商品開発研究所	主任研究員
委員	雨宮弘和	株式会社東芝	情報通信制御システム事業本部	渉外担当課長
委員	村木一至	日本電気株式会社	情報メディア研究所	音声言語研究部 部長
委員	森本秀樹	富士通株式会社	法務・知的財産権本部	技術企画部 担当課長
委員	吉原通年	松下電器産業株式会社	東京支社	技術渉外部 開発推進担当部長
委員	寺尾卓	株式会社インターグループ	翻訳営業部	部長
委員	野村勇一	日本翻訳連盟	専務理事	(十印から出向)
委員	古澤章	日本電子工業振興協会	情報産業部	部長
委員	野村浩郷	九州工業大学	情報工学部	知能情報工学教室 教授
委員	石崎俊	慶応義塾大学	環境情報部	教授

〈編集委員会〉

委員長	野村浩郷	九州工業大学	情報工学部	知能情報工学教室 教授
委員	亀井真一郎	日本電気株式会社	情報メディア研究所	音声言語研究部 課長
委員	富士秀	富士通研究所	マルチメディアシステム研究所	メディア統合研究部
委員	福持陽士	シャープ株式会社	情報商品開発研究所	主任研究員

《技術動向調査委員会》

- 委員長 村木一至 日本電気株式会社 情報メディア研究所 音声言語研究部 部長
委員 森本康嗣 株式会社日立製作所 中央研究所 知能システム部 研究員
委員 小松英二 沖電気工業株式会社 マルチメディア研究所 マルチモーダルプロジェクト第1グループ
委員 平井徳行 シャープ株式会社 情報商品開発研究所 主任
委員 伊藤悦雄 株式会社東芝 東京システムセンター ソフトウェアプロダクト部 主務
委員 亀井真一郎 日本電気株式会社 情報メディア研究所 音声言語研究部 課長
委員 斎藤由香梨 齋富士通研究所 マルチメディアシステム研究所 メディア統合研究部
委員 川越 睦 松下電器産業株式会社 マルチメディア開発センター情報グループ 情報第1チーム
委員 立田種宏 ソニー・テクトロニクス株式会社 ソフトウェアリサーチ推進室 リサーチ・フェロー
委員 畠中伸敏 キヤノン株式会社 情報通信システム本部 担当課長

《市場動向調査委員会》

- 委員長 石崎 俊 慶応義塾大学 環境情報部 教授
副委員長 坂本義行 東京家政学院筑波女子大学短期大学部 国際教養科 講師
委員 大高政浩 株式会社日立製作所 ソフトウェア開発本部 関西設計部 UL 技師
委員 石川友隆 沖ソフトウェア株式会社 企画室 担当部長
委員 森口 稔 シャープ株式会社 情報システム事業本部 PC 事業推進センター 第1技術部副主任
委員 松岡多恵子 株式会社東芝 コンピューター事業統括部コンピューター商品企画 担当主任
委員 杉村領一 松下電器産業株式会社 マルチメディア開発センター 情報グループ 開発推進担当副参事
委員 時岡洋一 株式会社インターグループ ソフト開発部 部長
委員 立田種宏 ソニー・テクトロニクス株式会社 ソフトウェアリサーチ推進室 リサーチ・フェロー
委員 堂野前進 長瀬産業株式会社 総務部文書課 翻訳担当 次長
委員 菊池純一 青山学院女子短期大学 教授
委員 森本秀樹 富士通株式会社 法務・知的財産権本部 技術企画部 担当課長
委員 古瀬 蔵 ATR 音声翻訳通信研究所 第3研究室 主任研究員

《例文評価研究会》

- 座長 横山晶一 山形大学 工学部 電子情報工学科 教授
徳永信治 シャープ株式会社 情報商品開発研究所 係長
松平正樹 沖電気工業株式会社 研究開発本部 商品化推進部 担当係長
中 英康 日本電気株式会社 C&C マルチメディア事業推進本部 PC 通信開発部 システム企画
時岡洋一 株式会社インターグループ ソフト開発部 部長
熊野 明 株式会社東芝 情報通信システム研究所 研究第5担当 研究主務
芦崎達雄 日本科学技術情報センター 技術開発部技術開発課 課長代理

白木沢佳子 日本科学技術情報センター 技術開発部技術開発課 情報員
 川越 睦 松下電器産業株式会社 マルチメディア開発センター情報グループ
 第1チーム
 安藤 進 TCC 代表
 塩津 誠 富士通株式会社 川崎工場 2MW-3開発部 ATLAS プロジェクト
 内尾淑美 富士通株式会社 川崎工場 2MW-3開発部 ATLAS プロジェクト
 堂野前進 長瀬産業株式会社 総務部文書課 翻訳担当 次長

なお、本年度からの新たな活動に対処するため、下記の2グループが新設されました。

(1) UPF 会議

情報処理振興事業協会 (IPA) の、創造的ソフトウェア育成事業として、「日本語情報海外発信促進のための言語知識コンテンツ蓄積・流通支援ソフトの開発」を行ないます。

各自小規模に蓄積されている異言語コミュニケーション用知識を互いに共有利用できるユニバーサルプラットフォーム (UPF) と、それを流通促進させるソフトを開発しようというものです。

技術動向調査委員会のもとで、委員会委員のほか下記の人たちが参加して活動します。

リーダー 村木一至 技術動向調査委員会委員長 (日本電気株式会社)
 技術動向調査委員会委員 9名
 時岡洋一 株式会社インターグループ ソフト開発部 部長
 岡 俊行 株式会社ノヴァ 研究開発2部 部長
 厚川美和 カテナ株式会社 ソフトウェアプロダクツ事業部事業部長
 桧山 努 株式会社 NEC 情報システムズ 技術システム事業部 主任

(2) インターネット WG (インターネットワーキンググループ)

ホームページ開設のためのグループです。

リーダー 森本秀樹 富士通株式会社 法務・知的財産権本部 技術企画部 担当課長
 立田種宏 ソニー・テクトロニクス株式会社 ソフトウェアリサーチ推進室
 リサーチ・フェロー
 富士 秀 富士通研究所 マルチメディアシステム研究所 メディア統合研究部
 中島奈津江 松下電器産業株式会社 東京支社 技術渉外部
 アドバイザー 野村浩郷 九州工業大学 情報工学部 知能情報工学教室 教授
 オブザーバー 樋口和雄 日本電子工業振興協会 技術部 部長

ホームページを開設！

この度、(社)日本電子工業振興協会 (JEIDA) のご厚意により、ホームページを開設致しました。皆様方への諸連絡等に活用致します。

より良いものに育て上げたいと思いますので、会員の方々からのご意見や情報をお待ちしております。

<http://www.jeida.or.jp/aamt/>

'95年度成果報告会の報告

第6回通常総会に先立ち、96年6月17日(月)13時から機械振興会館地下2階大ホールにおいて95年度の成果報告会が開催された。昨年は非会員も参加した有料のセミナーであったが、今年は会員のみを対象とした無料の報告会であった。約70名が参加し、田中穂積理事(東京工業大学教授)の司会で進められた。

はじめに「21世紀における機械翻訳」と題して長尾真会長(京都大学教授)の講演が行われた。多くの例を引用してのお話であった。以下講演内容。

「機械翻訳は1950年代に研究が始まり80年代に実用化された。しかし、翻訳品質の面で世の期待に添えなかった。最近パソコンやインターネットの普及により、パソコンでの機械翻訳がかなり売れるようになった。品質の割によく売れているようで、将来の発展のために根本的に見直すべき時であると思う。

今の機械翻訳がもうひとつうまくいかないのは、西洋流解析に基づいているからだと思う。すなわち、分解してそれぞれを解くという方式であり、あまりにも局部的取り扱いになりすぎているのではないか。また、理論的にきれいに解こうとしすぎたのではないか。今後はできるだけ広く、かついくつかの要素を統一的に考えるべきである。

例えば、現在形態要素解析のプログラムは97%~98%の精度であるが、範囲を広げて連結成分も考えて解析すると99.8%ぐらいまで精度が上がる。更に徹底的にやれば99.9%まで上がるだろう。構文解析では、並列語句の解析がむづかしいが、人間は並列語句のバランスを重視している。ダイナミックプログラミングの方法でフレーズ間の相似度を調べることによって90%ぐらいまでいく。

さらに、文脈的な解析が必要であろう。指示詞の指すものの推定、単数と複数の区別、具体名詞か抽象名詞かの区別等を意味的に考えるのはむづかしいが、他の方法での推察が可能である。すなわち、文の表面の情報をうまく利用して、意味の世界である

と思われているところまで迫ることができる。例えばある語についていろいろなキーを考え、いろいろなヒューリスティックを適用して蓋然性を求め、これを加え合わせてその語の性質を推定することができる。

また、ひとつの文だけで解析するのではなく前にある文からの流れを把握して解析する。さらに、後続の文も考慮することによりあいまいなものが明確になる。これらをそれぞれヒューリスティックルールとして使うことによって70%~80%はいける。

また、従来の辞書では、例えば名詞の場合は意味内容に応じてそれぞれの翻訳語が並んでいるだけであり、動詞の場合はケースフレームを整理しただけのものであったが、これだけでは不十分である。語の間の関係、例えば上下関係、全体と部分の関係、原因と結果の関係等々を入れた知識辞書を作る必要がある。しかし、どのようにして知識を蓄積するかが問題である。人間が全部をやることは不可能である。いろいろな例を取り出しそれぞれの場合の関係性を自動的に調べ辞書を拡大していくことが非常に大事である。これは言語外の知識であり、これをいかに整理していくかがこれからの大きな問題である。ぜひ、がんばってやってほしい。

発話者の意図、発話者と聞き手との関係も、非常に大きな問題である。しかし、どのようにすればよいか、まだ案はもっていない。

もう一つの問題は文と文のつながりである。並列的、対比的、主題連鎖的、原因結果的等々いろいろのつながりがある。これらがうまく解析できるかどうかの問題である。このような分析は翻訳だけでなく、自動抄訳、情報検索などでも大切である。

結論として、日本語の場合形態素解析では現在99.8%の精度であるが近い将来99.9%になるだろう。dependency analysisは現在97%~98%と考えられるが、99%にする必要がある。書かれたものの文節単位であればいけると思う。case structure representationでは、現在約80%であるが90%にはできるだろう。

問題は、信頼できる語数の多い格フレーム辞書のないことである。NTT 研究所にあるが非公開である。NTT のトップに公表をお願いしており、近々公表されると期待している。discourse interpretation は現在70%ぐらいだが、80%が限界ではないかと思う。

機械翻訳の品質は、これらを総合した全体の精度で決まる。そうなるとこれらを掛け合わせねばならず、現在は43%程度になってしまう。将来、上述のことが達成されたとしても60%強である。しかし、文章を書くときにちょっと工夫してもらえると精度はぐんと上がる。

以上のように、今世紀末には60数%から70%には

なるだろう。そして21世紀初頭には、これらの知識をうまく持ち込んだ新たな機械翻訳、すなわち文脈情報処理を中心とした新しい考えにもとづく機械翻訳が出てくるのではないかと期待している。」

つづいて、石崎俊委員長（慶応義塾大学教授）による市場動向調査委員会、村木一至委員長（日本電気齏部長）による技術動向調査委員会および横山晶一座長（山形大学教授）による例文評価研究会の成果報告が行われた。報告の内容は、AAMT Journal No.15およびNo.16に掲載されているので、ここでは省略する。

以上

TV 放映

徹底解剖自動翻訳ソフト

—NHK サイエンスアイで放映—

96年5月18日23時からのNHK サイエンスアイは、機械翻訳をテーマに放映された。

- 秋葉原のパソコンショップに翻訳ソフトがたくさん並び、幅広い客層に買われている状況
- 大手コンピューターメーカーでの英訳の実験例
- 慶応義塾大学湘南藤沢キャンパスでの自動翻訳の授業風景
- NHK の機械翻訳ニュースの例
- 日本科学技術情報センターで、月間約3,000件の論文要約が英訳されて海外に送られている状況
- 翻訳者養成講座で、機械翻訳が必須になっていること
- 翻訳ソフトは、95年に約50万本売れ、96年には150万本売れると予想されていること。
- 翻訳能力は完全ではないが、どんな内容であるか分かる程度には訳せること。
- しくみは、単語に分けて動詞に注目し、使用頻度にもとづく点数から最も可能性の高いものを選び出す得点法が使われていること。
- 京大教授長尾真氏との対談：
 - ・ 機械翻訳の問題点は辞書の大きさにある。用例や条件が細かく書かれた辞書が必要で、5～10年かけて技術者が作っている。
 - ・ ソフトを選ぶには、使ってみて特に have や get のような広く使われる単語の訳を調べてみるのがよい。

- ・ 今後の発展のためには、辞書を良くして行かねばならない。
- 八王子の翻訳会社の状況と、社長桜井恵三氏の
 - ・ 「知識の共有と再利用に大変役立つ」という話
- ATR 音声翻訳通信研究所での研究の紹介
 - ・ 音声翻訳のしくみは、音声→波形分析→文字→機械翻訳→音声パターン
 - ・ 現在は日本語から英語と独語への翻訳のみを研究しているが、3年後には日、米、独、伊、韓のTV会議を実験する予定。
 - ・ センサーを取付けた手袋を用いた動作辞書が作られており、手話翻訳が研究されている。
 - ・ 声の質（例えば風邪ざみの声）、方言等も今後の研究課題
- 再度長尾真教授との対談
 - ・ 機械翻訳は現在は日英が主だが、他の言語も研究されている。ヨーロッパでは英独仏の翻訳に用いられている。
 - ・ 文脈を考えての翻訳が重要
 - ・ 用例をたくさん蓄積した辞書作成が重要
 - ・ 言語は本質的には文化に依存している。それを取入れた翻訳が必要。
 - ・ 特に日本語はあいまいな表現が多い。それを明確な文章にするだけでも翻訳の質は向上する。など、翻訳ソフトの現状が幅広く、分かりやすく紹介された。

’95年度市場動向調査委員会 成果報告

委員長 石崎 俊 (慶應大学)

1. はじめに

当協会は昨年には発足5年目を迎え、昨年度の後半から新たに「市場動向調査委員会」を発足させました。それまで活動していた需要予測研究会を統合して、委員会の活動範囲を広げ、下記のような活動を展開することを目的にしています。

1.1 委員会の構成と活動目的

本委員会には、機械翻訳システムの主なメーカーとユーザおよび中立的な機関から、現在13名の委員が参加しています。それらの委員の活発な活動によって、機械翻訳システムの動向およびそれをめぐる環境変化を調査して、コンピュータ機器やマルチメディア技術の進歩、ネットワーク技術の普及、海外の市場動向などに対応して、機械翻訳システムの発展および普及を図り関連産業の育成を目的とします。

当協会の下には、もうひとつ別の委員会として「技術動向調査委員会」があります。この委員会では機械翻訳システムの性能を向上させるための技術を調査します。しかし、市場動向調査委員会では、ユーザの側から見た機械翻訳システムへのニーズ、要望を調査して、パソコンやインターネットなどの環境の変化に対応した、最適なシステムの在り方も視野に入れて行きたいと考えています。

1.2 調査活動項目

(1) 機械翻訳システムの普及および利用実態の調査と分析

- ・分野別の翻訳環境とその動向調査
- ・海外市場の動向調査

(2) 機械翻訳システムと関連産業の発展および普及のための方策

- ・翻訳リテラシーの構築

機械翻訳システムの性能は長い期間にわたって少しずつ進歩し、普及が進むと思います。人間並の能力を期待すると、むしろ何

十年かかるか分からないですし、永遠の課題と言えるかも知れません。したがって、翻訳システムの基本的な使い方を確立し、支援システムとしての位置付けを明確にすることが重要だと思います。

利用分野拡大のための調査（学校、旅行などの分野の検討）

たとえば、大学生がゼミの英文資料の翻訳に使用したり、大学での語学教育用に使用した例があるそうです。また、旅行での携帯用のシステムや、手紙の翻訳、英作文支援システムなどの利用が考えられています。ユーザ利用技術向上のための講習会の開催以前より地方を中心にして講習会を開催して利用者の拡大とともに、機械翻訳システムの基礎的な使用法の確立と普及を図って来ています。

1.3 背景について

自然言語処理に関しては、10年以上前から（社）電子工業振興協会の専門委員会です。この電子協では、当初は自然言語処理の中でも機械翻訳システムに焦点を当てた活動をしていました。現在も続いている機械翻訳の国際会議MT-SUMMITはこの活動が主体となって始まったものです。

その後、本協会が機械翻訳に関する団体として設立され、さまざまな活動を活発化する中で、電子協の自然言語処理に関する委員会では、調査対象を従来の機械翻訳からテキスト処理、対話処理、マルチモーダル情報処理へと重心を移して来ています。

本委員会では、そのような自然言語処理技術の展開を踏まえて、インターネットなどのネットワーク技術やマルチメディア技術の進歩と普及を念頭に置きながら、機械翻訳システムの市場動向を調査して行く予定です。

東アジア諸国と共同で機械翻訳システムを開発する8年間のプロジェクトが平成7年3月に終了しました。中国、タイ、マレーシア、インドネシアと日本が協力して中間言語方式の機械翻訳システムを構築したものです。このプロジェクトの成果を利用する計画がいろいろあるそうです。このプロジェクトの実施によって参加各国の自然言語処理研究が進み、大学を中心にして多数の研究者が育成されています。韓国やオーストラリアなどの国も含めて、アジア太平洋地域での機械翻訳システムの研究活動や企業活動が、今後ますます活発化することが期待されます。

2. 活動予定

本委員会の立ち上げに当たって、具体的かつ優先度の高い活動項目、活動範囲などについて委員会で討議し、下記のようにまとめました。

(1) 機械翻訳システムの動向調査

- ・専門分野別、目的別、メディア別の調査と利用分野拡大策の検討。
例えば、医学・薬品関係、教育・旅行、産業用・個人用など。
- ・現在のシステムの問題点、ニーズの把握。
- ・調査方法はアンケートやヒアリングなどの他に、内外の新聞雑誌、パソコン通信などを駆使した情報の収集。
- ・価格動向の調査。
- ・インターネット関連の使用や応用に関する調査。

(2) ユーザ講習会

機械翻訳システムの上手な使用法の教育、インターネット上の新しい利用法などの講習を行います。

(3) IPA・創造的ソフトウェア開発プロジェクトの推進

技術動向調査委員会に協力して、本プロジェクトの推進を図り、機械翻訳システムの基盤的なプラットフォームの確立によって、機械翻訳技術の一層の展開を目指します。

3. 活動報告

近年のパソコンの翻訳ソフトの市場展開には目を見張るものがあります。本委員会でも今年に入ってから代表的なパソコン翻訳ソフトのメーカーである、カテナ(株)、(株)ノヴァ、沖ソフトウエア(株)の協力を得て、各社の最近の製品開発、販売動向についてヒアリングを行いました。

1980年代後半から市販され出した翻訳システムは、当初は汎用大型機を使用し、次いでワークステーション上で稼働するものが販売されたが、ハードウェア込みではまだ数百万円以上していました。翻訳の質から見ると、構文パターンを用いたものや、浅い意味情報を用いたものが多く、価格に対応して期待されていた翻訳の質には到達されていませんでした。本来は文脈を考慮した高品質なものが、高価格に対して要求されている訳です。

1990年代に入りますと、急速に高性能化したパソコンに載る翻訳ソフトが開発・販売され出して、20万円前後のシステムの普及が始まりました。従来の翻訳システムの場合は、翻訳結果を仕事に使い、できれば外部でも通用するものが期待されたのに対し、廉価版では、英語で書かれた情報の内容を手早く把握し、重要なものか不要なものかを判断するという目的が変わって来ました。安いものであれば、そのような目的でもいいわけです。

1994年11月から、1万円を切る英日翻訳ソフトが販売されると、これを契機に数万円代の低価格版が次々に売り出され、普及が促進されました。この金額ですと、個人が自宅のパソコン用に気楽に購入できるので、一挙にユーザの層が拡大しました。

さらに最近では、インターネットが注目を集めており、パソコンを用いて世界中の情報にアクセスできる状況になって来ています。日本での使用者も毎年飛躍的に増えています。例えば、筆者の所属する慶應義塾大学藤沢湘南キャンパス(SFC)では、毎年1000人近い新入生がいて、キャンパス全体では5000人近い学生がいます。全員がアカウントをもって、1年生の前期のうちに電子メールばかりでなく、インターネットの使用法などもマスターします。このような教育環境は今後急速に普及すると思いますので、文系・理系を問わずに、パソコン上で自由に電子メールを使い、インターネットを使いこなす社会人は急速に増えると思われます。昨年後半からイ

ンターネット対応の翻訳ソフトが売り出され、販売も好調だと聞いていますが、時宜を得たものとして理解できます。

このようにインターネットが飛躍的に普及しますと、インターネットの内容はほとんどが英語で書いているので、日本人にとって気楽に翻訳してくれるパソコン翻訳ソフトはますます普及すると思います。したがって、これらのシステムを使用する目的は、従来のもの比べてますます手軽に使うという面が強くなると思います。インターネットの画面をリアルタイムに翻訳して概要を把握するだけでいいわけです。

それでは、今後の翻訳ソフトはどのようなものが要求されるでしょうか。やはり、翻訳精度の向上が最も重要だと思われます。構文情報、意味情報、文脈情報と高度化すると、技術的な難しさは「指数関数的に」増えます。システムの価格と、翻訳に要する時間、使用者のニーズなどのバランスを慎重に考慮して、今後の翻訳システムの開発や販売の戦略が立てられて行くと思います。

4. おわりに

情報技術に関してさまざまな分野およびテーマで

国際的な研究協力が進んでいて、APECでも積極的な発言が出ていると聞いています。アジア太平洋地域で使われているさまざまな言語を対象に多言語情報処理に関する共同研究開発ができれば、お互いの文化の理解にも役立つでしょう。アジア圏でよく使われている漢字のコードの標準化を初めとする自然言語の表示及び入出力の問題は今後の重要な課題です。また、多言語を同時に処理するという意味から機械翻訳システムは格好の応用課題と言えるでしょう。

このような意味から、機械翻訳システムはこれからも魅力的な研究テーマだと思います。なかでも電子化辞書や概念辞書などの大規模なシステムの開発のためには、国際的な共同活動の必要性が高いと思われます。

また、教育関連のソフトウェアの開発も盛んになって来ています。自然言語に関する教育システムの開発ニーズが多く、将来の有望なマーケットでしょう。機械翻訳システムがCAIの一環として役立つには、どのような側面を強化する必要があるかについて、市場を十分に調査する必要があると思われます。

新刊図書を紹介

■書名：自然言語処理

岩波講座 ソフトウェア科学 15

著者：長尾 真 京都大学工学研究科言語メディア工学分野教授
京都大学附属図書館長
佐藤理史 北陸先端科学技術大学院大学
情報科学研究科助教授
黒橋禎夫 京都大学工学研究科言語メディア工学分野助手
角田達彦 京都大学工学研究科言語メディア工学分野助手

発行所：株式会社岩波書店

定価：4,800円

概要：

世界的パイオニアたちによる決定版！
人間は「ことば」と「概念」を操作して考える。マルチメディア時代にますます重要となる自然言語処理技術をやさしく解説。計算機を学ぶ人すべてに必携！

形態素解析、構文解析、意味解析などの基本的な自然言語処理の理論と技術をていねいに解説する。さらに、文脈解説、言語理解、文の生成、学習などの進んだテーマ、また辞書や情報検索、機械翻訳などの応用上重要なテーマについても実践的に紹介する。重要なアルゴリズムや手順などは、実際の役に立つように具体的に記述した、世界的な自然言語処理研究のパイオニアたちによる決定版。

翻訳ソフト市場の現在

カテナ株式会社 ソフトウェアプロダクツ事業部 事業部長 厚川 美和

当社が「コリヤ英和！」ウィンドウズ版を定価9,800円で発売したのが1994年10月。翻訳ソフトの低価格化の火付け役になりました。当時は、「コリヤ英和！」があまりにも安すぎるので高い翻訳ソフトが売れなくなる、と他社メーカーの方々は非常に懸念されていたようです。ところが、蓋を開けてみると、低価格化によって、翻訳ソフトが一般ユーザーの選択の範囲に入ってくることによって、翻訳ソフト市場自体の拡大に大きく貢献することになり、各社の売上も増大したようです。

過去に、翻訳ソフトの潜在市場を計るマーケティング調査を行ったことがあります。そこで出た結論は、翻訳ソフトのニーズはワープロなみに大きなものでした。この結果が今現実となっています。大きな潜在ニーズを、これまでなかなか顕在化できなかった理由はなんでしょうか。まず第一には、環境としてのパソコン普及がありますが、翻訳ソフト側にもいくつかの要素があります。

- 1) 翻訳の質
- 2) 価格
- 3) 英文の電子化の手間
- 4) 翻訳スピード
- 5) 使う人、使う場面のニーズの違いへの対応

翻訳ソフトの発展・成長の要素は、なんといってもも初期訳精度の向上です。当然のことではありますが、メーカーの関心は機械翻訳の最大課題「翻訳の質の向上」に集中してきました。しかし、現段階で、翻訳ソフトの商品としての完成度を言えば、まだまだ発展途上にあり未成熟なものです。ソフト以外の一般的な他の商品と比較しますと、ある意味では欠陥品とも言えるかもしれません。たとえば、電卓の機能で「〇〇ができる」と言えば、それは100%できるのであり、出来ないものは不良品です。お金を支払って購入する商品に対しての一般的顧客の認識からくる期待値と、現在の翻訳ソフトの品質(=精度)の間のずれはまだまだ大きいと言えます。翻訳ソフト開発メーカーは、この乖離を埋めるために翻

訳精度の向上に向けてエンジンの改良に努力をしてきました。機能面でも翻訳精度を向上させるための編集機能的なものの改良が主でした。翻訳精度は確実に一步一步レベルが向上していますが、時間がかかります。最近の特徴は、精度の向上以外に、使いがっての向上によりユーザーの満足度を上げていることです。その結果、潜在ニーズの掘り起こしにつながって大きな成果を上げています。

パソコンの普及によって機械翻訳をとりまく環境が変化し、翻訳精度以外の2)～5)の要素が潜在ユーザーを顕在化させ市場を拡大できる条件が整いました。特に5)については、機械翻訳の手順・使い方の発想が作り手側の中で固定化していましたが、「翻訳を仕事としない」一般ユーザー層のニーズへの対応が市場を上げました。

「コリヤ英和！」は単に低価格化への火付け役だけではなく、5)の要素に対しても潜在ニーズである「読むためのツール」としての位置付けをはっきりさせました。そのために不要な機能は複雑さを避けるためになくし、「読むツール」としてのニーズにマッチした使い方を実現する「クリップボード翻訳」という機能を付けています。一般ローエンドユーザーの目的は、翻訳自体ではなく、その英文情報の内容が知りたいのです。ですから、「翻訳」という過程には出来るだけ煩わされたくないわけです。

「NetSurfer」は一早くインターネットニーズに対応した商品ですが、これも5)の要素に着眼した商品です。インターネットのホームページを「読むツール」として、表示形式を原文そのままにすることでニーズに対応しました。もちろん、インターネット自体が、これまでの機械翻訳の大きな課題の一つであった「入力の手間をいかに省くか」を解決してくれたことが直接的な大きな原因です。と、同時にインターネットならではの使いがっての改善が翻訳ソフトの可能性を大きくしました。過去を振り返ってみると、「原文レイアウトをそのままに翻訳結果を出す事」は、以前からマニュアルなどの日本語版を

作成する方々のニーズの中にあつたことです。当時は、「翻訳ソフトがそこまで対応する必要があるのか」と考えたものですが、インターネットによって情報が提供されるメディアの概念自体の変化により、読むツールには当然の要素になりました。

電子化された情報はますます増えていくわけですが、それでもまだまだ紙媒体の情報が主流ですから、入力に関する概念を変える使い方の提案も一つの方向です。当社では、紙媒体の英文情報をちよつと拾い読みするためのツール、入力と翻訳が、ワンステップで行えるようなツールを提案しています。当社が提案しているのは、翻訳ソフト（ロゴヴィスタ「E to J Personal」）とペン型OCRスキャナ（「IRISPen」）をセットにした商品ロゴヴィスタ「E to J Personal スキャニングシステム」です。ペン型スキャナと聞くと、「ハンディスキャナのさらに小型版にすぎず、訳に立つのだろうか？」という印象を持つ方は多いでしょう。しかし、IRISPenは、形状だけでなく、機能面でもこれまでのOCRとスキャナの使い勝手のイメージを変えました。「イメージデータの取り込み→テキストデータへの変換→翻訳ソフトへの取り込み」というこれまでの3段階のステップを「スキャンする（＝ペンスキャナで英文をなぞる）」とい

う1回の操作で行ってしまいます。あとは、翻訳実行アイコンをクリックするだけで翻訳結果が得られます。この商品も、5) 番目の要素を発展させる提案です。翻訳市場の拡大には、翻訳に必要な全てのステップの課題を解決する必要があります。

「わざわざ翻訳する」という壁をなくす商品が求められています。翻訳の質は当然のこととして、また翻訳の質が向上すればするほど、使い勝手の良さを求めるニーズは高くなります。OpenDocやActiveX、Appletといった技術により、アプリケーションの在り方・使われ方まで違ってきます。過去の行動パターンとは違うデジタル時代の翻訳ニーズ＝英文と係わる場面を一早くキャッチして、ユーザに使い方をソリューションとして提案できる商品が潜在ニーズを掘り起こし市場を拡大していくでしょう。

翻訳の質については、現時点で表面に現れられている翻訳精度の善し悪しだけでなく、いかに成長性のある翻訳理論であるかが商品の将来性を決めます。

「継続的な翻訳精度の向上が保証された翻訳理論の確かさ」と「ソリューション提案」の両輪がそろふことで、翻訳市場は今後もさらに大きく拡大すると考えます。

新刊図書を紹介

■書名：プログラミングの壺

出版社：共立出版

原著：“Programming on Purpose”

原著者：P. J. Plauger

訳者：安藤 進

監訳者：石田晴久

定価：①設計編(1995年10月1日発行) 2,700円

②人間編(1996年1月25日発行) 2,700円

③技術編(1996年6月18日発行) 2,700円

概要：

原著のタイトルは「目的意識を持ってプログラミングしよう」といった意味合いだが、ちょうど《アラジンと不思議なランプ》のように、すばらしいアイデアがたくさん詰まっている《壺》にちなんで邦訳を決めた。

4半世紀にわたり、メインフレームからミニコン、パソコンへと進むコンピューターの世界を歩んできた著者が、その間の技術と人間模様を活写した珠玉のエッセイ集。米国人には異色のコンセプトで展開するソフトウエアに関する待望の哲学書でもある。

翻訳は読者と著者のインターフェースだというのが、訳者の信念だ。この一年間、著者とはほぼ毎日《電子メール》でやり取りした。第1集と第2集にも、読者から多くの問い合わせをいただき、技術的な問題については、その都度著者に確認して返事している。《インターネット社会》での新しい翻訳のあり方を提示できたと自負している。

問合せ先：E-Mail Ando.Susumu. <GFG02221@niftyserve.or.jp>

インターネットと機械翻訳（その2）

九州工業大学情報工学部 教授 野村浩郷

インターネットへのアクセスにおいては、好むと好まざるとに拘わらず外国語で書かれた文章に接することになる。そのため、インターネットが一般化すればするほど、外国語を読み・書きすることに関する言語障壁の困難な問題が顕在化してくる。いくら外国語になれている人でも、大量の文章を速く読んだり、必要な情報と不要な情報とを速く選別するためには母国語の方が効率が良いであろう。このような問題を解決するための直接的な方策は、いうまでもなく機械翻訳を利用することである。そのような新しい状況の下における機械翻訳への要請に応えるためには、現在の機械翻訳は「変わらなければならない」。その意味で、前号において、「機械翻訳は人類史上にただ一回だけくる重大な転機に直面している」と書いた。では、どのように「変わらなければならない」のであろうか。今回は、言語対と多言語表示の問題についてみていくことにする。

国際的な情報交換では一般に外国語が不可欠である。しかし、外国語といっても、ほとんどの場合は世界の共通語である「英語」で用が足りる。なぜなら、インターネットで提供される情報の多くは英語で記述されており、この現象は今後も続くと考えられるからである。さらに、インターネットにおける業務用の URL では、日本語圏でもその他の非英語圏でも、自国語のものに加えて「英語ページ」も用意するのが普通である。したがって、インターネットで使う機械翻訳は、英語から自国語への翻訳の需要が最も高いということになる。翻訳需要におけるこのような傾向は、インターネットというものに依存するものではなく、従来から一般的に見られる傾向である。

しかし、外国語での URL の用意には工数がかかるものであり、また日本のような非英語圏では英語の URL を書くだけの英語力が必ずしも十分でない場合も多い。しかも、インターネットがさらに一般化すると、業務上必須ではない URL をわざわざ英語で用意する割合は減少すると思われる。これは、

将来、非英語と自国語との間の機械翻訳の需要が増す可能性を示唆している。したがって、従来はそれほど重要視されていなかった非英語と日本語との間の機械翻訳への要請も、今後は高まってくると予想される。

機械翻訳は、多言語機械翻訳ではなく、2言語機械翻訳として実現することが重要である。また、文書の種類や内容の分野に制限がないオールマイティな機械翻訳としてではなく、翻訳の対象と目的に合わせて機能を適合させた機械翻訳として実現することが重要である。なぜなら、現在の機械翻訳の技術（もっと一般的に言うと自然言語処理の技術）はあまりにも低いレベルに留まっているからである。現在および近い将来の技術レベルでは、機械翻訳における自然言語処理の機能の飛躍に高い期待をかけるのは間違っている。機械翻訳は、初歩的多言語処理とコンピュータネットワークとによる効果を翻訳において最大限に活用するという観点で考えるべきものである。

英語の文章を読む場合と、英語で情報を発信する場合とでは、機械翻訳への要求は異なる。また、いずれの場合でも、情報アクセスや情報発信の目的などによって機械翻訳への要求は異なる。機械翻訳の技術は未だ極めて初歩的な段階にあるので、技術の現状の範囲内で最も効果的な使い方を実現しなければならない。将来、機械翻訳の技術が飛躍したとしても、高い品質の翻訳はコストが高くなるであろうから、コスト・パフォーマンスの観点からの妥協も避けることができない。例えば、情報スキャンのような目的では、高い品質の翻訳よりは高速な翻訳が要求される。インターネットは大量な情報をリアルタイムにオンラインで提供するものであるから、そのような要求はますます多くなる。翻訳された文章が文法的に不完全であっても、母国語で表現されたキーワードが一瞥できるだけでもその効果は高い。すなわち、品質の悪い機械翻訳でも有効に活用できる場面は多い。

一方、情報発信のための機械翻訳は、意味内容が正確であるばかりでなく、文法的に正しくかつ読みやすい文章に翻訳できなければならない。しかし、残念ながら、このような要請に答えることができる完全なる機械翻訳の実現の見通しは全くないといわざるを得ない。したがって、現在の機械翻訳の活用のし方と同じく、最終的に人間が翻訳を完成させるという方法をとらざるを得ない。

インターネットで機械翻訳を使うためには、ブラウザがインターフェースとして機械翻訳をサポートする機能を備えていなければならない。しかし、残念ながら、現在のブラウザは機械翻訳のサポートを全く考慮しないところから出発しているので、そのような機能を望むべくもない。したがって、機械翻訳機能を持つ多言語ブラウザというものを新しく開発する必要がある。以下では、機械翻訳のための多言語ブラウザについて、多言語表示とウィンドウ構成の問題を中心に見ていくことにする。

多言語表示の機能に関しては、上に述べた通り、まず英語と自国語の表示、そしてさらに非英語の外国語と自国語が表示できる必要がある。このとき、非英語の外国語は必要十分な数だけ選択・表示できるものでなければならない。現在のブラウザの多言語表示の機能は、まったく不十分でありかつ極めて不適切である。

多言語表示を不具合なく表示するためには、文字コードが適切に設定されていなければならない。機械翻訳の観点から定められた文字コードとしては、CICC 機械翻訳プロジェクトの文字コードが唯一のものであると言ってよいであろう。CICC 機械翻訳プロジェクトは、日本語、中国語、タイ語、マレーシア語およびインドネシア語の5ヶ国語の間の機械翻訳システムの開発に関する国際協力プロジェクトであった。このプロジェクトでは、公式な文字コードの形式として、4バイトコードを採用した。この4バイトコードでは、したがって、各文字はそれぞれ4バイトのコードで表される。4バイトの内、最初の2バイトは言語識別用に、そして残りの2バイトは文字識別用に用いられる。文字識別用コードはそれぞれの国で標準化されているコードを使う。日本語に対しては、例えば、JISコードを使う。中国語に対しては、中国で標準化されているものを使う。これにより、言語は単純明快な方法により識別でき、

文字はその国の文字と共通な方法により識別できる。したがって各国語は不具合なしに適切に表示でき、しかも、十分多くの数の言語が表示できる。

日本語や中国語の文字の数は、欧米語の文字の数に比べて桁が違うほどに多い。したがって、多言語表示の問題点は、欧米語のみを扱うところからは明確にできない。多言語表示の問題は、例えば日本が世界のイニシアティブをとって取り組むべき問題である。そのとき、CICCの文字コードに基礎を置くべきであろう。しかし、残念ながら、現在までにそのような観点からの取組・成果は見られないようである。

日本で使われている代表的な現在のブラウザは、例えば、日本語文字、韓国文字、あるいは中国文字のいずれかと欧州文字が選択できる。欧州文字としても、少数の選択がある。このような言語選択の機能は、機械翻訳のために用意されたものでもなく、またユーザの便を考えて用意されたものでもない。すなわち、機械翻訳のためには全く不十分かつ不適切である。したがって、ブラウザの多言語表示の機能は、根本的に設計し直す必要がある。

日本語文字のコードの種類としては、代表的なものに、EUCコード、SJISコード、およびJISコードの3種類がある。このように、日本語に対して3種類ものコードの種類があるため、ブラウザの表示で文字化けが生じるという極めて不具合な現象が起こっている。現在のブラウザの中には、文字コード選択機能が不適切であるため、日本語文字が適切に表示できないものもある。このような問題も含めて多言語表示の問題を解決するためには、日本としては、まず、日本語文字コードを統一する必要がある。

ウィンドウ構成については、ブラウザの言語画面分割、言語別ブラウザクローン作成、および逐次の言語対照表示などが考えられる。これらが十分な情報の量を表示でき、かつインターネットへのアクセスの対象や目的によって、そして個人の好みによっていずれもが容易に選択・移行できるようになっていなければならない。そのためには、画面は、現在のもの（例えば、1280x1024ピクセル）よりももっともっと高精細な表示能力を持っていなければならない。

機械翻訳と文法

(株) 富士通研究所 メディア統合研究部 富士 秀

1 文法とは？

さて、そもそも文法とは一体何でしょう。まずは機械翻訳は傍らに置いておいて、文法そのものについて考えてみます。

ちょっと手元の辞書を引いてみるとこんな風に載っています。「単語の文構成上の性質や文などの構造・運用に見られる法則の体系」。まあ、平たくいうと「単語がつながって文を作っていくときの、つながり方の規則をまとめたもの」といったところでしょうか。しかし、実際には文法という用語はかなり曖昧に使われていて、規則として書けるような事柄は何でも文法と呼ばれるような場合もあります。この記事の中でも、一般的な感覚では文法とは言えないような言語現象も取り上げていますが、大目に見てやってください。

言語の中には規則化し易い事柄としにくい事柄があり、それらが複雑に絡み合っているわけですが、今回は、規則化し易いものが文法だと考え、これに関してお話をしたいと思います。

2 文法って何の役に立つの？

なぜ文法などという面倒臭いものが必要なのでしょう？そもそも私達が日本語の文章を読む時に「ええと、この単語は名詞で、この単語が動詞だから…」などという読み方をする人はいないでしょう。私達が、日常普通に生活していく上で、文法が必要となる機会はそう多くはないと思います。

とすると、文法が使われる場面は日常生活以外ということになるわけですが、具体的にはどんなところで使われるのでしょうか。

● 国語教育

学校教育の国文法では、まさに文法というものがお目見えします。中学時代の国文法の教科書を引きづ

り出してくると、次のような記述があります。「国文法の学習を通して健全な日本語の使用ができるようにその基礎を研かなければならない。」でも、日本語を使う環境に生まれてきた人で、国文法を勉強するまではまともな文章が書けないと言う人は少数でしょう。辞書だって問題なく引けます。「何でもこの勉強するのだろうか」と思いながらもテストの点を取るために学習した人もいないのでしょうか。

● 外国語教育

外国語の授業では文法は大活躍します。例えば、英語の授業で「book (本)」は名詞で「walk (歩く)」は動詞だと言われれば、なるほど何故かピンとくるものです。日本語の文法にもそういう似たような概念があるので、このように文法を用いて説明されると分かり易いわけです。また、外国語と日本語の仕組みの違いを認識するのも役に立ちます。「英語には冠詞があるけど日本語にはない。」などというのはこの例でしょう。聞いただけでアレルギー反応を起こす人もいるかも知れませんが「関係代名詞」などもこの例です。ここでは、日本人が外国語を学習する際のことを述べましたが、外国人が日本語を学習するような場合でも同様に文法は活躍します。

ちなみに、学校教育の英語教育を受けても実際に使えるようにならない、と言われて久しくなります。これは、あまりにも文法を重視し過ぎたため、その内容はかなり「翻訳の勉強」に近いようなものになっていたからではないかとの指摘もあります。逆にそのお蔭で、学校教育の外国語の理論が翻訳に適したものになっているかも知れません。

● 機械翻訳

文法のもう一つの活躍場所は機械翻訳です。機械翻訳では、上の外国語教育のところ述べたような、原言語と目標言語の類似性や非類似性を利用して翻

訳を行ないます。例えば、「英語のこういう構造は日本語ではこういう風な構造に対応している」というような規則を用意しておいて翻訳を行なうこととなります。

多くの機械翻訳システムでは、何種類かの異なった文法セットが必要になります。一つ目は、原言語の構造を理解するための、いわゆる解析文法と呼ばれるものです。二つ目は、理解した原文の構造を目標言語の構造に変換する文法です。これは、「文法」という表現をされない場合もあるでしょう。そして、三つ目は、変換された構造から、目標言語の文法に沿った文章を作り出す、生成文法といわれるものです。

以上、文法が使われる場面を考えてみました。文法というものが、言語間の変換を行なう、つまり翻訳をするためには都合の良いものであることはおわかりいただけたと思います。

3 人間のための文法と機械のための文法

文法は翻訳のためには頼もしい存在なのですが、同じ文法でも人間のための文法と機械のための文法はどちらも様子が違ってきます。それでは、これら2種類の文法はどこがどう違うのでしょうか？

やはり決定的に違うのは、人間向けの文法は非常にシンプルにできているのに対して、機械向けの文法は何でも事細かに記述されていることでしょう。人間は曖昧なものを処理したり、一般化したり、推論したりすることができるので、文法でもそれを見越して必要以上に細かい説明をする必要がないのです。例えば、人間なら“I go to school.”の“school”の変わりに“the city centre”等と書かれたようなものでも難なく理解してしまいますし、“I go to school everyday.”が来てもちっとも驚きません。母国語の知識をフル活用するなどして、何のためらいもなく読んでしまいます。これに対して、機械に教える文法としては「この文の目的語としては、場所を表す名詞が来なければならない。場所を表す名詞とは…」のように本当にこと細かく記述してやる必要があるのです。

解析文法と生成文法が分かれているのも、そんな機械翻訳の文法の特徴を表しているのでしょう。普通の日本人であれば、英語をマスターすれば、ある程度は日本語から英語への翻訳と英語から日本語への翻訳の両方を同じようにこなせるようになります。つまり、英語が理解できると、ある程度は英語が書けるようにもなります。しかし、機械の場合は、英日翻訳システムができたからといってそんなに簡単には日英翻訳システムが作れるわけではありません。人間のように融通が効かないのです。

別の見方をすると、人間が文を処理する場合には、シンプルな文法理論と人間の融通性の高い曖昧性処理能力の両方が使われるのです。これに対して、機械翻訳で処理する場合には、シンプルな文法理論にどのくらい融通性を持たせられるかが勝負を決めるのです。

なお、前号や前々号の記事でも話があった通り、文法以外にも、辞書や品詞なども事細かな記述が必要になってきます。全くもって、融通の効かない機械にものごとを教えることは容易なことではないのです。

4 機械翻訳における文法の生い立ち

それでは、機械翻訳の文法はどのようにして生まれてきたのでしょうか？

4.1 理詰めのアプローチ

これまでお話ししてきたように、外国語教育などでお馴染みの人間向けの文法は、機械翻訳に限らず、翻訳を行なうには都合の良いものです。そこで、当初開発された機械翻訳は、この人間向けのシンプルな文法をベースに作られてきました。人間向けの文法がシンプルなのに対して、機械向けの文法は細かい記述が必要なので、文法の規則数はどんどん増えました。

また、最初は基本的な文法規則に対しての例外を記述していくような形で開発が進められるわけですが、このようにして規則の数が増えてくると、規則の間に矛盾が生じてきました。「例外処理に例外処理を重ねて文法を充実させていけば、そのうちに完璧な

翻訳ができようになるさ。」という夢は崩れ去ったのです。

この理由としては、言語というものは、文法のように単純に規則化できる要素ばかりでなく、もっと文全体のパターンに根ざしている要素や、複数の規則間のバランスに根ざした要素などが含まれている、ということが考慮されていなかったからでしょう。

4. 2 直観的アプローチ

以上のようにして、理詰めのアプローチでは、満足な翻訳結果が得られないという苛立ちがつのってきました。こんな中で、もっと実務的な翻訳のプロセスに近いような方式が提唱されるようになってきました。この方式では、例えば“the change of the company's name takes effect”に対して「新しい社名が有効になる」というような訳文の対を大量に用意しておきます。そして、入力文が与えられたら、その入力文に近いような対訳対を検索してきて訳文を取りだし、うまく組み合わせることによって出力を得ます。例えば、“the change of the school's name took effect”のような入力があったら、蓄積してある対訳対の中から上述の対が最も類似したものであると認識し、「新しい校名が有効になった」というような出力をします。

このような方式だと、想定される入力文に近いような対訳対が大量にあれば、非常に自然な訳文を得ることができます。この方法は、一見して文法が必要ないように思われますが、実は突き詰めていくとそうでもないことがわかってくるのです。

上の例で、英語の単語の“company”は例えば、“school”や“president”などでも構わないわけです。このことを表現するためには“the change of the xxx's ...”の“xxx”は特定の名詞によって置き換えられることを記述できなければならないわけですが、そのためにはやはり名詞などの文法カテゴリが必要になってきます。これは単純な例ですが、このようにして考えて行くと、直観的なアプローチでも文法に近いようなものが必要になってくるのがわかります。

まとめると、どのようなアプローチで開発を行なうとしても、文法的な要素を使うことは避けて通れないことであり、また文法的でない要素も同時に使う必要があるのです。

5 機械翻訳における文法の現状

さて、それでは実際の機械翻訳における文法はどのようなになっているのでしょうか。

現在の翻訳システムでは、基本的な文法に対応する処理はほぼ行なうことができるようになっていと言えるでしょう。また、文法規則ではカバーしきれないような曖昧さを含んだ文もある程度はこなせます。しかし、やはり安心して見ていられるのは、単純な文法に則った文が入力されたときでしょう。単純な構造の短い文ならかなり綺麗に訳せるものが多いのもこのためです。逆に、機械翻訳システムが苦手とする文の例を次にいくつか挙げてみました。

●長文

一般的に、短い文は構造が単純なものが多いのですが、長い文は必ずと構造も複雑になってきます。学校文法は、このような複雑な構造の文には対応していません。機械翻訳システムは、もともとと学校文法を基に作られているので、従来は長文を苦手とするシステムがほとんどでした。最近では、長文に焦点を絞ったような研究開発が行なわれてきており、訳文の品質も徐々に上がってはきていますが、まだまだこれからといったところでしょう。

●並列句を含む文

並列句を含む文とは、例えば、「太郎と花子の犬」のような文です。機械処理するには「(太郎と花子)の犬」という可能性と、「太郎と(花子の犬)」という可能性のどちらがより妥当な解釈なのかを判断する必要があります。恐らく人間がこの文を読むと、直観的に「(太郎と花子)の犬」というような解釈をしたいと思います。機械に処理させる場合にはうまく処理ができるような規則性(文法)を見つけてこれをシステムに組み込んでやる必要があります。

上の例の並列句を処理するためには、「太郎」と「花子」の単語間の類似性を調べたり、「太郎の犬」と

いう組合せの妥当性を調べたりしなければなりません。つまり「並列になる要素は類似性が高い」とか「それぞれの要素の妥当性が高い」などの複数の条件が必要となります。そして全体として、なるべく多くの条件がなるべく十分に満たされるような解が妥当性の高い解と言うことになります。

これはいわゆる「文法」とはイメージが違いますが、規則的なところもあるので文法と考えることもできるでしょう。また、文の構造を認識する上でも重要なので、この意味でも文法とは密接な関係にあります。しかしながら、学校文法などではあまり取り上げられることはなく、これまでは規則もあまり明文化されることはありませんでした。

文は長くなればなるほど、並列句を含む割合が増えてきます。また、日本語にも英語にも同じような現象が見られ、どちらの言語でも難しい問題となっています。

以上の例は、規則化できないことはないが、単純な規則とはなりにくいという意味では、文法とは呼べないかもしれません。いずれにせよ、これらの他にも、文法開発の上で障害になっていることは色々あり、現在の機械翻訳システムの開発で取り組まれています。

6 文法の限界

これまでのところ、機械翻訳が既に扱えるもの、そして、開発を進めればそのうち扱えるようになるであろうものについてお話ししました。しかし、まだあまり手も付けられていない要素もたくさんあります。勿論、文脈処理や意味処理のように規則にはなりにくいものもありますが、以下では規則になり得るがあまり取り組まれていないような要素の例を挙げてみました。

6.1 翻訳ノウハウ

例えば英語を日本語に翻訳する際、英語の単語が日本語の単語に置き換えられ、かつ日本語の単語が文法的に正しく並べられていたとしても、人間の翻訳者が書いたようなこなれた訳文を得ることはできません。翻訳者の書く訳文は、訳文の文章としての読み易さや自然さを追求しているからなのです。ちな

みに、学校の授業では原文に忠実に訳すことに重点が置かれているあまりに、一応文法的に正しければ読み易さはさほど重要視されません。

本職の翻訳家は、訳文を読み易くするためのノウハウを色々持っていますが、その大部分は経験によって培われたものであり、明文化されていないことがほとんどです。とはいえ、ある程度規則化が可能なものもあり、これらも次第に機械翻訳システムに取り入れられていくことでしょう。

規則化し易いものの例としては、訳文を組み立てる際の語順の処理が挙げられます。語順の処理とは、訳文を組み立てる際に原文の語順になるべく近くなるような語順で出力するような処理です。一般的に、文章中の単語は話がスムーズに流れるように順序付けられているのですが、不用意な翻訳をすることによって原文の語順が乱されてしまうことが起こります。このような事態を防ぎ、読み易さを追求するために語順の処理をする必要があります。

例えば、次のような英語の新聞記事を日本語に翻訳することを考えます。

"Imports this year are expected to slow from last year as a result of government measures," Mr. Johnson told the press in Canada, where he has been attending the conference since two weeks ago.

ここで、関係詞の“where”を典型的な学校文法に則って訳すと以下のように語順が大幅に変わります。

ジョンソン氏は、2週間前から会議に出席しているカナダで記者団に対して「政府の抑制策の結果、今年の輸入は昨年より鈍化する見通しだ。」と述べた。

これに対して、原文の語順を意識した翻訳では次のような訳文が得られます。

ジョンソン氏は記者団に対し「政府の抑制策の結果、今年の輸入は昨年より鈍化する見通しだ。」と言明した。同氏が2週間前より出席しているカナダにおいて述べたもの。

6. 2 翻訳後処理

出版されるような翻訳物では、ただ単に翻訳者が原文を目標言語に置き換えただけでは品質的に不十分です。翻訳物が使われる分野において自然に理解されるような文章に直す作業が、専門家によって行なわれます。また、翻訳物が対象としている読者に応じたレベルの文章に書き換えるなどの作業が必要となります。例えば、原言語の方では当たり前に使われているような概念が、目標言語の方では一般的でないとする、単純にその概念を表す訳語を見つけるだけでは不十分で、多少意味を補足するような記述を埋め込むような努力も必要となります。

このような高度な処理を規則化して機械翻訳に取り込むのはかなり難しいとは思われますが、高品質の訳文を得るためには不可欠な要素であることには違いがありません。

7 最後に

言語の要素の中で、規則化できる要素を中心に考え

てみました。機械翻訳システムでは、学校文法レベルの規則はかなり扱えるが、それ以上については現在努力中といったところでしょう。また、将来的にも取り組むべき課題はまだまだ沢山あります。

今回は、文法という観点に絞って話をしましたが、文法の精度ばかりを上げることが究極の目標ではないことも考える必要があります。必ずしもプロの翻訳家がつくり出すような完璧な訳文が必要でない場合もあるのですから。

文法は言語を構成する一つの要素に過ぎないことを認識し、機械翻訳システムをトータル的に把握して、どんな要素を重点的に開発すべきかを考える必要もありそうです。

(株) 富士通研究所 メディア統合研究部
自然言語処理研究グループ
富士秀 (FUJI, Masaru)

fuji@flab.fujitsu.co.jp, NIFTY-Serve: RHD02514

イベント紹介

アジア最大規模の情報通信システム総合展

データショウ'96

会 期 1996年10月14日(月)～17日(木)・4日間 10:00～17:00

会 場 東京ビッグサイト(有明:東京国際展示場) 東展示棟1、2、3、6ホール
〒135 東京都江東区有明3-21-1

入 場 料 無料

主 催 社団法人 日本電子工業振興協会(JEIDA)
通信機械工業会(CIAJ)

協 力 社団法人 日本パーソナルコンピュータソフトウェア協会(JPSA)

後援(予定) 通商産業省
郵政省
日本電信電話株式会社
国際電信電話株式会社

同時開催 データショウ'96フォーラム
CALC Japan '96
PC カード WORLD '96
アクセシビリティコーナー
「パソコン・現在・過去・未来」
(JEIDA/JPSA 共同情報発信プロジェクト)

郵政省通信総合研究所関西先端研究センター知的機能研究室

知的機能研究室長 井佐原均

1. はじめに

通信総合研究所関西先端研究センターは、東京小金井に本所を置く通信総合研究所の支所として、平成元年に兵庫県神戸市に創設され、情報・物性・バイオの3部門について、基礎研究を中心に未来創造型の先端技術の研究開発を進めている。

その中で、知的機能研究室では、学習・推論・感性など、人間の持つ高次知的機能をモデル化し、その知見を基に高度な自然言語処理システムを構築する研究を行っている。

言語理解の研究においては、人間と同様に予測を用いながら理解を深めていく手法、表層的な意味だけではなく言葉の奥の深い意味までも理解する手法の研究を進めている。

言語生成の研究においては、状況・文脈に応じて語彙選択を適切に行う柔軟な形での文生成の研究を進めている。

また、人間の知的活動を支援することを目標に、個人のニーズに応じたネットワークニュースの検索を行う知的ニュースリーダーの開発を行っている。

これらの研究の基盤として、神経回路網モデルを用いた知識の研究、音楽情報や敬語表現に関する感性の研究、辞書・コーパス等の言語データの整備、さらには、他言語への応用としてのタイ語処理の研究を併せて行っている。

以下では、これら、知的機能研究室での研究の概要について紹介する。

2. 深い意味理解の研究

私たちは、言葉を理解する場面では、言語知識、世界知識等の様々な知識を用いて、深い意味理解を行っている。本研究では、これまで、比喩、駄洒落、同語反復表現等の様々な修辞表現を対象に、その検出と理解手法を研究してきた。

たとえば、「鳥は鳥だ」という同語反復表現は、

文字通りに解釈すれば、全く新たな情報を付与していないことになる。しかしながら、この文がペンギンを話題として発話されていた場合には、「通常、ペンギンは鳥ではあるが飛べないものと認識されているが、鳥の1種であることには違いがないのであるから、少くくらは飛べるだろう」といった解釈が必要となる場合もある。このような処理を行う場合、同語反復を検出すること、同語反復表現の示し得る幾つかの役割のうちのどれに対応するかを決定すること、そして最終的に意味解析を行うこと、等が必要となる。

このような知識を用いた処理は、人間の言語理解においては本質的なものであり、今後は照応解析や省略の処理への応用を行う予定である。

3. 柔軟な形での文生成の研究

我々の行う発話は非常に多くの要因に影響を受けている。例えば、「直前にどのような発話が行われたか、どのような場面・状況において発話しようとしているのか、発話の相手はどのような人で、発話者とはどのような関係にあるか」などがそうであり、これらの要因によって発話文の表層の形が異なってくる。

このような発話の文脈依存性あるいは状況依存性を考慮することは、表層文を生成する際に、柔軟な表現を可能にするためにも、文の結束性を保つためにも必要である。

本研究ではこのような発話の性質を考慮しながら文の形を決定する要因とその性質を明らかにし、柔軟な形での文生成システムの構築を目指している。

4. 知的ニュースリーダーの開発

インターネットの普及に伴い、ネットワークニュースの投稿数は増加の一途をたどっている。ネットワークニュースには、新聞記事型（アナウンス型）

のニュースグループと、討論型のニュースグループとが存在する。従来からの情報検索技術や要約技術の応用として、新聞記事型のニュースグループを対象に、キーワード検索を中心とする研究が行われてきた。

しかしながら、ネットワークニュースの中心は討論型のグループである。また、検索入力としてキーワードを用いる手法は、既知の情報へのアクセスには有効でも、人間の発想を積極的に支援するものとしては、使いづらい。

ここでは、討論型のニュースグループを対象に、記事の流れ（情報文脈）という概念を導入することにより、ユーザが興味を持った記事自体を検索のキーとし、その記事に関連する記事を過不足なく提示する手法を開発している。また、本研究では、単に基礎研究にとどまらず、システム化も併せて行う予定である。

5. 知識表現の神経回路網モデル

本研究では、記憶・学習・自己組織化のような基本機能を持つ神経回路網モデルの研究から着手し、それらを統合することによって、人間が持つような、

(1) 効率的な記憶と素早いアクセス、(2) 曖昧性へのロバスト性、(3) 修正が簡単、(4) 自己組織や一般化が可能などの特徴を有する知識表現システムを構築することを目標としている。これまでに、パターン復元だけでなくパターン分離もできる適応型連想記憶 AAM を提案し、それを用いることによって、上記特徴の 1 と 3 を実現する意味ネットワークモデルを構築した。

このモデルは、属性に基づく素早いアクセスを特徴としており、人間の言語理解における多重視点の問題（たとえばワープロは、筆記具、重量物、商品、工業製品、等の視点から見ることができる）に対応できる自然言語処理向きの知識表現に発展させることが可能であると考えている。

6. 感性情報処理の研究

音に関する感性情報処理の研究として、ふたつ以上の楽器による演奏音を対象とした自動採譜アルゴリズムの研究、及び、演奏の音高評価に関する工学

的モデルの研究を行なっている。

この内、自動採譜アルゴリズムの研究については、楽器個々の音響モデルを用いないアルゴリズムを提案し、バイオリン二重奏音やフルート二重奏音に対し、高い正解率で演奏された音符を推定することができた。ただし現在のところ、分離抽出された個々の音の音源同定（すなわち、どちらの楽器の音か）はまだ対象としていない。

音高評価に関する工学的モデルの研究については、トップダウン的な聴覚心理モデルを提案した上で、旋律の複雑さ、旋律中の音の予測のしやすさ、及び音高ずれ検知のしやすさの間の依存性を、いくつかの聴覚心理実験により明らかにしつつある。

また、言語における感性情報処理の研究として、心理実験に基づく待遇表現の丁寧さの計量化を通じ、敬語表現発話におけるメンタルモデルの研究を行なっている。

7. タイ語コーパスの品詞自動決定

一つの単語はしばしば複数の品詞を持ち得るが、その単語が一旦文に組み込まれると、前後の品詞（文脈）に応じて、その単語の品詞が唯一に決まる場合が多い。このような品詞の決定は人間には可能な作業であるが、大量の文から構成されるコーパスに人手で品詞をタグづけするのは極めて複雑でコストのかかる仕事である。本研究はニューラルネット技術を応用し、タイ語のコーパスの品詞自動タグづけシステムの構築を目指すものである。

ここではニューラルネットの持つ柔軟な学習能力を用いて、その言語に対する知識を予め持つことなく、精度の高いタグ付けを行うことを目指している。本研究から得られる知見は、単にタイ語の処理に留まらず、人間の言語獲得過程の研究にも示唆を与えられるものと期待している。

8. 言語データの作成

上記の品詞自動決定の研究に用いるテキストとして、タイ語で書かれた工学系の報告書を対象に、品詞タグを付与したコーパスを作成している。原文を語切りした後、電子辞書を参照し、各単語に可能な全ての品詞を付与する。その後、人手によって、そ

の文中での実際の品詞を決定する。このコーパス作成は、通信総合研究所と、タイ王国国立電子・計算機技術センター（NECTEC）との研究協力の一環として行われている。

またRWCデータベースワークショップ・テキストグループでは、毎日新聞の記事4年分を対象に1億形態素からなるタグ付きコーパスを作成し、公開しているが、知的機能研究室では、これと協調し、テキストグループが作成した形態素タグ付きコーパスの中で、人手修正を経たものから、用言と名詞句（名詞+助詞）の共起関係を抽出している。この作業は、別途作成したツールを用いて、人手で行われており、高品位のものである。このデータは辞書作成において、選択制限等の情報を記述するための基礎データとして利用できると考えている。

知的機能研究室が作成している、これらのデータは、原則として、研究用に一般公開する予定である。

9. 他分野との研究協力

知的機能研究室では、従来の研究分野の枠に囚われずに、常に広い範囲に興味を持って、新しい分野にも積極的に取り組んでいこうと考えている。

言語学に関しては、神田外語大学を拠点とするCOE形成プログラム「先端的言語理論の構築とその多角的な実証」に情報処理グループとして参加し、神経言語学・神経医学グループと共に、言語理論の実証に挑戦している。

また、郵政省からの委託研究として、脳神経科学の研究に関して東京都神経科学総合研究所と、人間の知的活動の支援環境に関して（株）リコーと、協力して研究を進めている。これらの研究協力は、たとえば、失語症の克服支援、あるいは、その知見に基づく自然言語処理メカニズムの開発等につながっていくものと期待している。

10. おわりに

私たちは日々の生活の中で、言葉を通して様々な事柄を考え、伝え、助け合って生きている。知的機能研究室では、この「言葉」を中心とする人間の活動を様々な観点から研究することによって、人間の持つ知性を解明していきたいと考えている。

初期の言語能力の獲得から始まって修辞表現の理解に至る人間の言語機能は、どのような仕組みで作られているのだろうか？人間の感性に関して、音楽や言葉は、どのような情報を持ち、何を伝えているのだろうか？人間の知識はどのような枠組みで表現されているのだろうか？

このような基礎的な問いかけから、機械翻訳、情報検索、コーパス処理といったシステム化技術に至るまでの幅広い研究の中で、いくらかでも、言葉の本質に迫ることができればと思っている。

本稿に対する、ご意見、お問い合わせをお待ちしております。（isahara@crl.go.jp）

事務局だより

協会活動報告

総	会	：	6月17日	95年度事業報告書および決算報告、96年度事業計画および予算案、役員 の改選の3議案の審議承認。新会長、新副会長の紹介						
理	事	会	：	6月17日 会長および副会長の選任 7月18日 辻井潤一理事の副会長選任（書面審査）						
編	集	委	員	会	：	7月8日 AAMT Journal No.15の反省とNo.16の編集。アンケートについての討議				
市	場	動	向	調	査	委	員	会	：	7月17日 96年度活動方針、ヒアリング計画についての討議。IPAプロジェクトの 報告
成	果	報	告	会	：	6月17日 講演「21世紀における機械翻訳」 長尾 真 会 長 成果報告（1）市場動向調査委員会 石崎 俊 委員長 （2）技術動向調査委員会 村木一至 委員長 （3）例文評価研究会 横山晶一 座 長				

機械翻訳関連ソフトウェア一覧表

(100万円以下)

(順不同)

ソフトウェア名		販売会社	OS	翻訳対象言語	販売価格
1	訳師如来	CSK	Windows	英 → 日	6,000円
			Mac 7.1	英 → 日	6,000
2	ブラビス J/E ブラビス J/E SASUGA	パシフィックアイ& エムティラボ	Windows	日 → 英	58,000
			Windows	日 → 英	9,800
3	Trans Land	ブラザー	Mac 7.1	日 → 英	69,800
			Windows	日 → 英	39,800
4	トランスサポーター E J/C トランスサポーター E J/W トランスサポーター JE	三洋インフォメーシ ョンビジネス	Windows	英 → 日	25,000
			Windows	英 → 日	29,800
			Windows	日 → 英	25,000
5	WD-01 SW E/J for Text E/J for WWW	シャープ	ワープロ書院	英 → 日	150,000 }バンドルソフト
			Windows	英 → 日	
			Windows	英 → 日	
6	日韓機械翻訳システム 韓日機械翻訳システム	日立情報ネットワー ク	MS-DOS 5.0	日 → 韓	178,800
			MS-DOS 5.0	韓 → 日	178,800
			MS-DOS 5.0	韓 ↔ 日	298,800
7	PIVOT / JE PIVOT / EJ 英日翻訳アダプタ 英日辞書引き君 日英辞書引き君	NEC	EWS-UX / V	日 → 英	500,000
			EWS-UX / V	英 → 日	500,000
			Windows	英 → 日	14,800
			Windows	英和	9,800
			Windows	和英	9,800
8	ATLAS EJ for Windows ATLAS JE for Windows 電辞海	富士通	Windows	英 → 日	58,000
			Windows	日 → 英	58,000
			Windows	英和・和英	24,000
9	パンセ・ブイ英日版 パンセ・ブイ日英版 パンセフォーインターネット WWW サーバーモデル"パンセ"	沖ソフトウェア& オージス総研	Windows	英 → 日	38,000
			Windows	日 → 英	38,000
			Windows	英 ↔ 日	58,000
			95/NT	英 → 日	9,800
			solarisl. X	英 → 日	148,000
10	X-EJ 2/W X-EJ X-DIC	スピリット	Windows	英 → 日	98,000
			MS-DOS 3.1	英 → 日	29,000
				英 ↔ 日	48,000
11	Logo Vista E to j Logo Vista E to j Personal The Translator-Mini コリヤ英和	カテナ& ロゴヴィスタ	Windows	英 → 日	97,000
			Mac	英 → 日	97,000
			Windows	英 → 日	39,800
			Mac 7	英 → 日	39,800
			Mac 7.1	英 → 日	14,800
			Windows	英 → 日	9,800
Mac	英 → 日	9,800			
12	うちの翻訳屋さん	長瀬産業	Windows	英 → 日	246,000
			Mac 7.1	英 → 日	246,000
			Windows	日 → 英	246,000
			Mac 7.1	日 → 英	246,000

ソフトウェア名	販売会社	OS	翻訳対象言語	販売価格
13	ASTRANSAC Sun WS ASTRANSAC C / S ASTRANSAC for Windows ASTRANSAC / Web	東芝	UNIX OS / AS Windows(NT) Windows(NT) Windows(NT) Windows(NT) OS / AS	英 ↔ 日 700,000円 英 ↔ 日 340,000 英 ↔ 日 340,000 英 → 日 63,000 日 → 英 63,000 英 ↔ 日 98,000 英 ↔ 日 200,000
14	J•London / EJ J•London / JE J•London ツインセット J•ソウル日韓版 J•ソウル韓日版 WorldNet / EJ	高電社	Windows Windows Windows MS-DOS 3.1 MS-DOS 3.1 Windows	英 → 日 49,800 日 → 英 49,800 英 ↔ 日 78,000 日 → 韓 390,000 韓 → 日 298,000 英 → 日 19,800
15	Dr. Surf for Windows	メディアビジョン& 九州松下電器	Windows	英 → 日 14,800
16	Transpad E-JBANK for Windows	亀島産業	Windows Windows	英 → 日 14,800 英 → 日 75,000
17	Translation Manager / 2	日本 I B M	OS / 2	英 → 日 785,000
18	孫悟空	クリエイイト大阪	MS-DOS 3.1	日 → 中 190,000
19	Net Surfer / ej PC-Transer / ej PC-Transer / je マイトランサー EJ Super Transer / ej	ノヴァ & 山種産業	Windows Windows Mac DOS(PC 98) DOS(DOS / V) Windows Mac DOS(PC 98) DOS(DOS / V) DOS(PC 98) DOS(DOS / V) DOS(PC 98) DOS(DOS / V) DOS(PC 98) DOS(DOS / V)	英 → 日 29,800 英 → 日 198,000 英 → 日 198,000 英 → 日 248,000 英 → 日 298,000 日 → 英 198,000 日 → 英 198,000 日 → 英 248,000 日 → 英 298,000 英 → 日 98,000 英 → 日 98,000 英 → 日 448,000 英 → 日 448,000
20	Power Translator Pro. Power Translator Pro. Power Translator Pro. Power Translator Pro. Power Translator Power Translator Power Translator Power Translator 中国語 Power Translator CD-ROM	イリス インターナ ショナル	Windows DOS Mac DOS Windows Mac DOS DOS Windows	独・仏・西 148,000 独・仏・西 148,000 独・仏・西 148,000 露 ↔ 英 148,000 独・仏・西 58,000 独・仏・西 58,000 独・仏・西 38,000 英 → 中 28,000 独・仏・西 58,000
21	英日翻訳支援システム for Windows 英日翻訳支援システム 日英翻訳支援システム	日立製作所	Windows 95 UNIX UNIX	英 → 日 49,000 英 → 日 1,000,000 日 → 英 1,000,000

メモリー	基本辞書	周辺機器	問 合 せ 先		
			電 話	F A X	担 当 部 署
16MB～	8万語		0423-40-6893		東京システムセンターソフトウェアプロダクト部
13MB～ 12MB～	8万語 9万語	プリンタ プリンタ、OCR	06-628-8880	06-628-2351	本社
640kB～ 640kB～ 13MB～	7万語 5万語 10万語				
8MB～	7万語		03-3222-3908		(株)メディアビジョン
			03-3798-4838		
			03-3546-4154		
	3万語		03-3351-3356 03-3820-1121 06-348-0484		(株)ノヴァシステム営業部 山種産業 ク
			03-3542-0950		

英文作成支援ソフトウェア 「日英辞書引き君（単文変換版）」

NEC

NECではこのたび従来の機械翻訳ソフトとは異なる対話的インタフェースで日本人の英文作成を支援するソフト「日英辞書引き君（単文変換版）」を開発しました。このソフトは当社のパソコン「CanBe」「ValueStar」等にプレインストールされているほか、パッケージソフトとしても販売しています。

(Windows95版：9,800円)。

1. 「日英辞書引き君（単文変換版）」の役割

近年、社会全体の国際化および世界規模の電子ネットワークの発達により、海外と英語で情報をやりとりする機会が急増しています。英文読解支援としては英日翻訳ソフトが多数商品化されています。

しかしながら日本から海外への情報発信の環境整備はまだ進んでいません。日本人が英語の文書を作成する場合、ある程度の英語力を持った人であっても文意を伝えるのにどの英単語が適切か、スペルはどうだったか、この動詞はどのような文型をとるのか、など様々な疑問を抱きます。それらの疑問を解消するには、英作文の資料や和英・英和辞書などを参照することになりますが、こうした作業は通常面倒で、英文情報発信の阻害原因となっていました。

近年は電子化された英和／和英辞書が市販されていますが、ワープロやエディタ上の文書作成作業と関係をもっていないために、残念ながら英文作成の有効な助けにはなっていませんでした。また従来の一括処理型日英翻訳ソフトでは操作者の持っている言語知識と機械処理とがうまく融合しておらず、結果の英文の品質の向上に結び付きませんでした。

そこでこれらの問題を解決したのが「日英辞書引き君（単文変換版）」です。

2. 「日英辞書引き君（単文変換版）」の特徴

● 仮名漢字変換と同様のインタフェース

このソフトは、日本語ワープロで使われる仮名漢字変換の操作法を日本語英語変換に拡張したものです。仮名漢字変換と同様の「変換」「選択」の繰り返しによって、日本語を順次英語に変換してゆきます。このためユーザは仮名漢字変換の感覚でソフトとや

りとりしながら、英文を組み上げていくことができます。

入力された日本語は、単語の変換、句の変換、文の変換という段階を経て英文に変換されます。一度に変換できるのは動詞を一つ含む単文の範囲までですが、関係節や接続詞で結ばれた長い文に関しては、その中に含まれる単文範囲の変換を繰り返すことにより全体を英文に直すことができます。図1に変換の段階を示します。

● 訳語選択用補助機能およびCD-ROM 辞書検索機能

仮名漢字変換の場合には選びたい漢字がユーザの頭にあらかじめイメージされていますが、英語の単語の場合には、ふさわしい単語が判断できないのが普通です。例えば「料金」に相当する英語の単語の場合、「charge」「fee」「fare」などの類義語のどれが今書いている文書に適切なものかはすぐにはわかりません。そこで本ソフトでは、英訳語候補を表示する際に、それらの使い分けを示した候補選択用の補助情報も一緒に表示します。例えばサービスや労働に対する料金には「charge」が、運賃の場合には「fare」

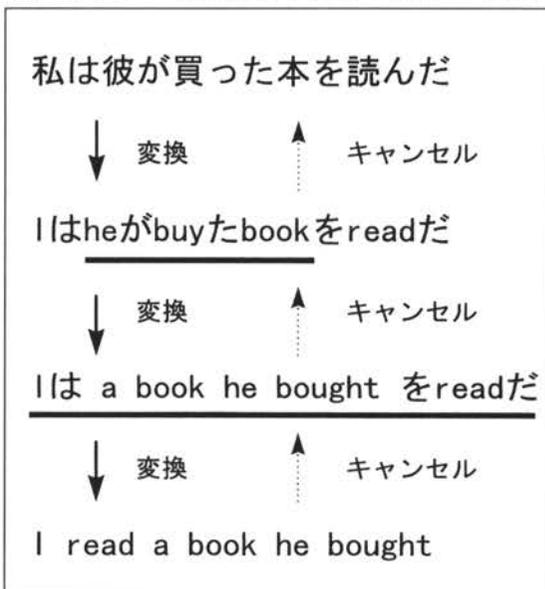


図1 「日英辞書引き君（単文変換版）」の変換段階

が使われるといった用法説明が補助情報として記述されていますので、ユーザはそれを元にして適切な英訳語を選択できます。

しかしそれでもまだ不十分で、その単語の使われる構文や文脈などを英和辞典の例文などにより確認したい場合があります。この要求に答えるため本ソフトは、市販されているEP-WING規約に準拠したCD-ROM英和・和英辞典をワンタッチで検索・表示する機能を備えています。候補訳語の一つに焦点を合わせ、CD-ROM検索ボタンを押すと、CD-ROM辞書が引かれてその単語の辞書内容が別画面に表示されます。CD-ROM辞書の内容は複数表示することが可能ですので、類義語の辞書内容を読み比べて、最適な単語を選択できます。辞書検索のために別のソフトを起動して検索単語を入力しなおす必要がないため、辞書を調べる負担が大幅に軽減されます。

●充実した慣用表現辞書

日本語を英語に訳す時に困る表現に慣用表現があります。例えば「電話する」という意味の「電話をかける」という表現には「かける」という動詞が使われていますが、その英訳は「hang」や「lack」ではなく「make」(「電話をかける」→「make a call」)でなければなりません。そこで本ソフトでは基本辞書約10万語の他にこのような日本語の慣用表現辞書

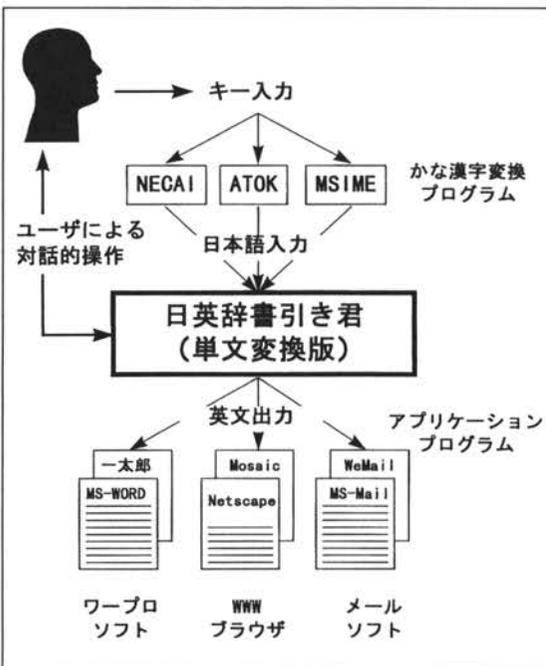


図2 「日英辞書引き君 (単文変換版)」の使用環境

約1万5千表現を標準装備しました。この慣用表現辞書は日本語らしい言い回しを英語らしい言い回しに直す際の強力な手助けとなります。また「電話をかける」の例でも「壁に電話をかける」の場合の「かける」は文字通りの用法ですが、このように同じ表現に文字通りの読みと慣用表現の読みとの両方がある場合には、通常の訳語選択と全く同様の統一的カーソル操作で一方から他方へ切替えられるようになっています。

●任意のワープロやエディタと組合せ可能

「日英辞書引き君 (単文変換版)」は仮名漢字変換と同様に他のワープロやエディタなどと組み合わせることで使えるフロントエンド・プロセッサ型のソフトウェアです。Windows95上のワープロやエディタなど皆さんの慣れ親しんだ任意のソフトウェアから起動をかけることができ、結果は元のワープロやエディタに取り込まれます。仮名漢字変換ソフトも現在お使いのものをそのままお使いいただけます。すなわち、現在の文書作成環境に本ソフトをアド・オンするだけで、英語作成支援環境を作り上げることができるわけです。電子メールソフトと組み合わせれば、そのまま海外との英文手紙のやりとりに活用できます。このような他ソフトとの連携の良さは、従来の翻訳支援ソフトにはなかった利点です。図2は「日英辞書引き君 (単文変換版)」を使用する際の構成図です。

3. 「日英辞書引き君 (単文変換版)」の動作環境

本ソフトウェアは以下の環境で動作します。

本体	PC-9800シリーズ
OS	Windows95日本語版
CPU	i486SX20MHz以上 (33MHz以上を推奨)
メモリ	2.5MB以上 (3.5MB以上を推奨；OSを除く)
ディスク空き容量	25MB以上 (OSを除く)

<お問い合わせ先>

NEC パソコンインフォメーションセンター
PC-9800シリーズ担当
〒108 東京都港区芝浦4-16-23
アクアシティ芝浦ビル
電話 03-3452-8000 (代表)

翻訳アダプタ II

NEC

1. 翻訳アダプタ II の概要

インターネットが広く普及することによって、ホームページを読むだけでなく、個人でホームページを作成し、電子メールをやり取りする機会も増えています。これに伴い、海外との接触も増え、英語と接することも多くなってきました。

画面上の英語を読むのは誰しも苦痛でしょう。紙の辞書を引けば時間もかかり、あまり効率的ではありません。そこで、最近英語を読むための翻訳ソフトが数多く登場し、市場に広まっています。しかし、最近の翻訳ソフトを見る限りでは、受信側が英語を読むためというニーズを満たすものではあっても、英文を書くという発信側のニーズには十分応えていません。

例えば、ホームページを作成して情報を発信したり、海外へ電子メールを送ったり、個人輸入などで自ら英語を書かなければならないことがあります。従来の翻訳ソフトは英日翻訳といった一方通行のソフトなので、こうした、日本語を英語に手軽に翻訳する目的には役に立ちません。また、英文を書いてニュースグループに投稿したり、通信販売などの問い合わせ、苦情のメールを出すと言っても、困ってしまうことがよくあります。

今回登場した翻訳アダプタ II は、こうしたニーズに応える英日&日英双方向翻訳ソフトです。

最近流行の、ホームページをそのまま翻訳するものとは違って、ワープロをはじめ、エディタや WWW ブラウザや電子メールソフトなどのさまざまなソフトとともに利用できます。

主な機能としては、翻訳、辞書引き、例文検索、書式検索があります。すべての機能が日英、英日両方

で動きます。

使い方はとても簡単です。辞書引きや翻訳したい箇所をマークして、キーを押せば（またはマウスの右ボタンクリック）、辞書を引いたり翻訳したりすることができます。辞書引き結果や、翻訳結果は図 1 のように表示されます。また、必要があれば、翻訳結果や辞書引き結果をワンタッチでワープロなどのアプリケーションに戻すこともできます。

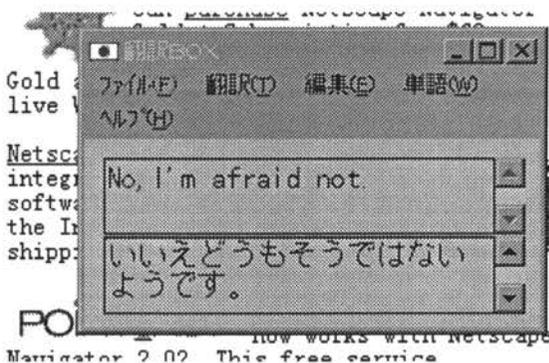


図 1 翻訳ボックス

添付されている例文や、書式検索のサンプル集は、英文レターなどを書くときに便利なものです。日英翻訳、日英辞書引きなどを利用して、英文作成は可能ですが、特殊な英語表現、ユーザの用途に合った表現、レターの構成などを反映させたい場合には、この例文集とサンプル集が役立ちます。用途別に書式を検索し、サンプル集を雛形とし、翻訳機能や辞書引き機能を使って、書き換えれば誰でも簡単に英文レターが作成できます。サンプル集はインターネットの世界で著名な、元 CNN ニュースキャスターの西森マリー氏の監修した品質の高いものとなっています。

収録されているサンプル集は、「英文ホームページ」「メール各種」「ニュースグループへの投稿」「通信販売における問い合わせ、照会、発注、苦情、キャン

セル」など、インターネット利用上の様々な場面を広範囲にわたってカバーしています（図2参照）。なお、例文集、サンプル集ともユーザが追加することができます。

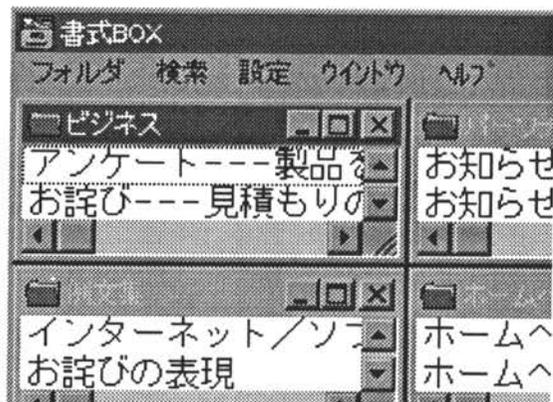


図2 サンプル集

このように、翻訳アダプタⅡはWWWブラウザだけでなく、メーラー、ワープロ、エディタなど使える範囲が広く、必要な箇所をマークするだけで、翻訳もしくは辞書引きが簡単にできる、つまり、英文の読解と作成を楽にしてくれる便利なツールです。

2 豊富な辞書

日英9万語、英日9万語の基本辞書を内蔵しており、ユーザ辞書は登録語数が無制限（ディスク容量の範囲内）で自分専用の辞書が簡単に作れます。専門用語辞書も31分野60万語と充実（別売り）しており、幅広い分野の翻訳に使えます。なお、この辞書は翻訳、辞書引き共通の辞書となっています。

3 翻訳と辞書引き

日英、英日両方とも複数の文を一度に翻訳できます。キーボードから入力した文も同様に翻訳が可能です。英日では、訳文文体の指定（である調・ですます調）や命令文の訳出の仕方、「you」の訳出の仕方、日英では省略された主語の訳出の仕方、などきめのこまかい翻訳制御ができます。

4 CDROM 検索ソフトとの連携機能

市販のCD-ROM 検索ソフトと連動して、ワープロ上の単語を検索することもできます。操作は翻訳アダプタⅡの辞書引き操作と同じでワンタッチで検索ソフトを起動させることができます。

5 システムの構成

OS	Windows3.1 (95動作確認済み)
CPU	i486 66MHz 以上推奨
マシン	NEC PC-9821または IBM PC/AT 互換機
解像度	640X480以上推奨
メモリ	16M バイト以上推奨 (スワップ25M バイト以上推奨)
ディスク	最小構成 約20M バイト (辞書はCD-ROM) 最大構成 約90M バイト (HDD にインストール時)
標準価格	9,800円

お問い合わせ先

EC 推進本部/汎用アプリケーション事業部
TEL 東京 03(3456)8343 (ダイヤルイン)
FAX 東京 03(3456)6348
〒108 東京都港区芝浦4-14-22
(NEC 大東田町ビル)



翻訳アダプタⅡ

J T F 翻 訳 祭

——つながりマッセ！心とこころ

超（スーパー）テクノロジー時代の翻訳——

日 時 1996年10月24日（木）13：15～20：30
 25日（金）10：00～17：00
 場 所 アイビーホール青学会館（〒150 東京都渋谷区渋谷4-4-25）
 主 催 日日本翻訳連盟（JTF）
 後 援 通商産業省（予定）、日日本ユネスコ協会連盟 ほかに10団体
 協 力 アジア太平洋機械翻訳協会（予定）

プログラム

10月24日（木）

13：15～13：30 開会のあいさつ 日日本翻訳連盟会長 勝田美保子
 13：30～14：30 ブロードキャスター ピーター・バラカン氏とともに考える
 ——“グローバル・ヴィレッジ”は実現が可能なのだろうか——
 14：40～15：20 講演「空飛ぶ文字事情」 JAL 国際線外国人フライトアテンダント
 15：30～16：45 パネル討論会「地球ネットワークの中のアジア事情と日本」
 18：00～20：30 懇親パーティー

10月25日（金）

10：00～12：00 基調講演（インターネット・サーフィン）
 PART 1 「だからインターネットはおもしろい」
 ナビゲーター：小林 典子（日航ビジネス、JTF 会員）
 技 術：村林亜希子（富士通、AAMT 会員）
 PART 2 「翻訳お役立ちツールたち」
 ナビゲーター：野上 員生（ベストバージョン、JTF 会員）
 技 術：村林亜希子（富士通、AAMT 会員）
 13：30～14：10 講演「ようこそ産業翻訳の世界へ」八代登志江（イカロス出版、JTF 会員）
 14：20～15：00 講演「翻訳のための PL & CS」石上 進（国際事業開発、代表取締役社長）
 15：10～16：20 講演「私の翻訳支援環境のスタイル」宇野みれ（企業内翻訳者）
 16：50～17：00 クロージング・メッセージ 日日本翻訳連盟副会長 大澤水紀雄

展示会

日 時 1996年10月24日（木）13：00～18：30
 25日（金）9：00～15：00
 場 所 青学会館ミルトス

機器実演展示、雑誌・パンフレットの展示、企業案内
 翻訳よろず相談コーナー

参加料 JTF の会員および学生：3,000円 一般：5,000円
 《団体チケット：10,000円（会員、学生：4名まで、一般：3名まで）》
 懇親パーティー：一律5,000円

申込期限 1996年10月18日（金）

申込方法 所定の参加申込書により郵便または FAX で申し込む

申込・問合せ先 〒104 東京都中央区八丁堀2-8-1 日日本翻訳連盟
 TEL：03-3555-6365 FAX：03-3552-1784

AAMT
ジャーナル

No. 16
(Sept. 1996)

発行	アジア太平洋機械翻訳協会
所在地	〒105 東京都港区芝公園3-5-12 芝公園真田ビル3階
	TEL: 03-5473-7135/7136 FAX: 03-5473-0569
	E-mail: KYN02317@niftyserve.or.jp (NIFTY)
編集委員会	野村浩郷(委員長) 亀井真一郎 福持陽士 富士 秀
事務局	田中 孝 神野千恵子
印刷所	伸光写植印刷株式会社

