



AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.15
June 1996

アジア太平洋機械翻訳協会

目 次

	インターネットと機械翻訳	1
翻訳の現場から	タルムード翻訳に参加して	3
ヒアリング	翻訳ソフトの市場動向	6
ヒアリング	Windows 版 PENSEE の商品化と市場への取り組み	8
研究会報告	機械翻訳における記号の取り扱いについて：例文評価研究会	11
技術早分かり	機械翻訳と品詞	18
研究機関紹介	奈良先端科学技術大学院大学情報科学研究科松本研究室	24
学会報告	言語処理学会第2回年次大会参加報告	26
新製品報告	Mebius ノート PC-A 400シリーズ搭載「英日翻訳支援ソフト」	30
	翻訳サーフィン V 1.0	32
	英和・和英辞書ソフトウェア「電辞海 V 2.0」(for Windows)	34
	「LogoVista E to J Pro V 3.0」「LogoVista E to J Internet」	36
協会便り	第6回通常総会の開催	5
	'95年度成果報告会の開催	5
イベント報告	ビジネスショー'96報告	10
国際会議	MT サミット VI	17
協会活動報告		33
イベント紹介	テクニカルコミュニケーションシンポジウム'96	39
	COLING-96の開催	40
	情報処理学会自然言語処理研究会	41
	早稲田大学の'96 PC Conference	表紙裏

おことわり

No. 14に掲載しました「機械翻訳関連ソフトウェア一覧表」の詳細をこの号に掲載すると予告しましたが、都合により次号にさせていただきます。お詫び方々ご連絡致します。

イベント紹介

早稲田大学の'96 PC Conference

早稲田大学理工学部キャンパスで7月5日(金)から7月7日(日)まで開催される PC Conference において、機械翻訳に関する分科会等が開かれる。

分科会：語学教育への利用

司会：立命館大 松田教授

機械翻訳を活用した技術英文セミナーの報告 佐良木 昌氏(佐良木技術翻訳事務所)

7月6日(土) 11:20~12:00 54号館 302

ポスターセッション：機械翻訳を活用した技術英文セミナーの報告

佐良木 昌氏(佐良木技術翻訳事務所)

7月6日(土) 12:30~14:30 55号館第1会議室

展 示：(株)高電社 日英翻訳ソフト 中国語、韓国／朝鮮語入力システム&フォント等

7月5(金) 13:00 ~ 7月6日(土) 17:00 55号館

主 催：大学生協 PC カンファレンス実行委員会

TEL 03-3382-3133 FAX 03-3384-8113

e-mail KYB 01731 @ niftyserve. or. jp ciec-jim @ univcoop. or. jp

インターネットと機械翻訳

九州工業大学情報工学部 教授 野村 浩郷

世はまさにインターネットの時代である。最近では、職場からはもちろんのこと、家庭からでも自在に使えるようになってきた。将来は、さらに、情報ハイウェイと呼ばれる超高速なコンピュータネットワークが実現されるので、コンピュータネットワークへの情報アクセスは日常生活で不可欠なものとなる。

インターネットや情報ハイウェイで送受される情報は、いわゆるマルチメディアの形をしている。マルチメディアとは、文章などのテキスト、音楽などのサウンド、および映像などのイメージを三要素とするものである。さらに、文字、音声、対話、多言語などマルチモーダルな形で情報の授受が行われる。これらにより、臨場感がありかつ現実感がある情報の提示が行われる。

インターネットや情報ハイウェイの特徴は、世界中のすべての職場や家庭が結ばれること、情報が世界中に分散されて格納されること、世界中の情報にいつでもアクセスできること、情報が双方向に送信されること、および情報がリアルタイムに発信され即時に受信されることなどである。これらにより、すべての情報が手元にあるかのごとく情報サービスが行われる。

コンピュータネットワークが提供する情報には、コミュニケーション情報、デリバリ情報、およびストアード情報の三つの種類がある。コミュニケーション情報とは、電話、Fax、オンライン購買などの通信型情報である。デリバリ情報とは、通知、広報、ニュースなどの配送型情報である。ストアード情報とは、図書館などの蓄積型情報である。

コンピュータネットワークへの情報アクセスは、このような状況の下で最も効率的かつ効果的に行われなければならない。それを実現するためには、情報アクセスを手助けする知的な支援機能が必要となる。そのような知的な支援機能を備えたコンピュータネットワークは知能ネット (IntelliNet) と呼ばれる。本稿では、文字を使って表現される情報である

テキストに焦点をあてて、知的な情報アクセスの機能すなわち知能ネットに要求される機能についての問題点をさぐることにする。

知的情報アクセスの機能としては、情報探索、情報検索、情報抽出、情報発見、情報加工、情報整理などがある。情報探索とは、どのような種類の情報がどの場所にあるかを探すことである。これにより、未知の種類の情報が見いだされることになる。情報検索とは、あらかじめ定められているトピックに関する情報を探すことである。情報検索の一種として、従来から、文献検索は多く使われている。情報抽出とは、ニュース記事などから主要な情報を選び出すことである。例えば、新製品に関するニュース記事から製品種別、販売企業名、性能、価格などの情報を抽出し、レポートを作成したりデータベースを構築したりする。情報発見とは、情報検索などのトピックとして設定すべき新しい情報を認識することである。これにより、調査漏れを補完できたり、社会の変化や発展に対応できるようになる。情報加工とは、情報検索や情報抽出した情報などを報告書の形に整形すること、要約を作成すること、さらに翻訳することなどである。これらにより、冗長な情報記述を排除でき量が少なく密度の濃い情報記述が整形された形でかつ母国語で得られることになる。情報整理は、情報加工した情報を分類すること、そしてファイリングすることなどである。コンピュータネットワークからは、多種・大量の情報が得られるので、情報整理の機能なくしては情報の洪水に溺れたり埋もれたりすることになる。

知的情報アクセスの機能は、さらに、情報発信を支援するためのものも必要である。それらには、情報記述、情報付加、および情報格納などに関するものがある。情報記述とは、情報を実際にテキストとして記述することである。そのとき知的なワープロが必要となる。知的なワープロは、テキストの内容の記述とテキストの形式のレイアウトの両方を支援する。情報付加は、関連情報の所在に関する情報を

付加したり、知的情報アクセスの処理を支援するためにテキスト構成や言語情報に関する情報を付加したりすることである。後者は、後で述べる研究・開発用のタグドテキストコーパスに深い関係を持つ。これらにより、知的情報アクセスが効率的に行えるようになる。情報格納とは、多種多様な情報をファイル群や知識ベース群として構造化して蓄積することである。構造化は、物理的に一つの場所に蓄積される情報に対してだけではなく、異なる場所に蓄積される情報をも含めて行う必要がある。これにより、情報探索や情報検索が効率的に行えるようになる。

これらの知的情報アクセスには、翻訳という処理が不可欠となる。なぜなら、コンピュータネットワークが全世界をむすぶものであり、したがってテキストはいろいろな言語で記述されているものであるからである。さらに、翻訳は、知的情報アクセスの形態がさまざまであるから、それぞれの形態に対応する形で行われなければならない。

インターネットの普及は、自然言語処理および機械翻訳に新しい局面をもたらす。そして、その新しい局面は、いま始まろうとしている。このような機械翻訳の新しい展開は、人類史上にただ一回だけあるといっても過言ではない。そのただ一回だけの転機がいま到来しようとしているのである。この転機をうまく捕まえて適切に対応することにより、機械翻訳の新しい時代が始まる。

機械翻訳の新しい展開は、第一に、インターネットへアクセスするためのマンマシンインターフェースから始まる。WWWブラウザに表示される情報を翻訳する機械翻訳システムはすでに市場にでている。これを知的情報アクセスの形態と目的にそって高度化・多様化することが必要である。情報スキャンのためには、必ずしも高品質な翻訳は必要ない。情報の量は膨大であるので、キーワードが母国語で表示されるだけで情報アクセスはかなり効率化される。このような機能の需要は、今後、爆発的に増えると思われ。

現在使われているWWWブラウザは、単一言語用のものである。これでは、いま始まった多言語情報の時代には対応できない。多言語用のWWWブラウザの標準化と開発を早急に行わなければならない。これは、知的情報アクセスにおける情報加工などのためのものをも含めて行う必要がある。現在の

WWWブラウザはマルチメディアのための基本機能は備えているが、将来的には、さらに、マルチモーダルなアクセスにも対応できるようにしなければならない。すなわち、マンマシンインターフェースは、マルチメディア・マルチリンガル・マルチモーダルな機能を持つものでなければならない。

機械翻訳の機能を高めるためには、情報発信のときに機械翻訳を手助けできる情報を付加することも重要である。情報発信をするときには、その情報の解釈を規定する「場」が設定されているはずである。「場」は、「対象に関する知識」と「心の状態」と「外部の状況」を三つの要素とする。情報受信のときに、この「場」を部分的にでも再現できる情報が付加されていると、自然言語処理や翻訳処理におけるいわゆる「曖昧さ」を大幅に削減でき、したがって、高度な処理が可能となる。研究や開発のために言語情報を付加した基礎的な言語データは、タグドテキストコーパスと呼ばれる。これを取り入れて、共通化・一般化された情報付加の標準を定めることは、現在のHTMLをはるかに越える効果をもたらす。HTMLは、現在のWWWブラウザのホームページを記述するのに使われている。HTMLは、図表を含む文章のレイアウト記述言語である。

機械翻訳の機能を高めるためのもう一つの方策は、「制限言語」の考え方を導入することである。制限言語の考え方には、文を表記するときの読点の使い方に制約をおくもの、文構成要素の係り受け解釈を物理的に最も近いものに仮定するもの、および対象に依存する言語表現上の特徴をモデル化するものなどがある。いずれも、自然言語処理や機械翻訳のときの最も困難な技術的問題点である「曖昧さ」をできるだけ排除することを目的にしている。制限言語の効果は極めて高いので、情報記述の段階と情報アクセスの段階の両方について活用していく必要がある。「制限言語」は、実用の面でも学術の面でも自然言語処理・機械翻訳にブレークスルーをもたらす。

インターネットの普及は、機械翻訳の新しい局面をもたらしつつある。この機会を見逃すことなく、機械翻訳の新たな展開を積極的に進める必要がある。翻訳の質をあげるためには長期に渡る学術的な研究が必要であるが、現在の技術に立脚した役に立つ機械翻訳を知能ネットの一つの機能として実現することが重要である。

タルムード翻訳に参加して

(株) ハイテック 代表取締役 岩下陽示

ユダヤ人の知恵の源泉と言われるタルムードを日本語で出版するという事業に翻訳者の一人として参加するようになったのは、平成4年の春のことです。もう既に丸4年間もこの仕事に掛かりきりの状態が続いています。この事業が実現するに至ったのには、数多くの方々の善意と熱意、そしてさまざまなご苦勞があったわけですが、その間の経緯はさて置いて、ここではわたし自身の経験したいろいろな問題や面白かったことなどを書いてみたいと思います。

ご承知のように、タルムードはユダヤ人にとってヘブライ語聖書と並んで最も権威ある書物です。内容は紀元200年頃に編纂されたミシュナと呼ばれる法典（成文律法である聖書と並んで権威あるものとされた口伝律法を成文化したもの）と、そのミシュナをめぐって行われた注解やさまざまな論争・問答の記録（ゲマラと呼ばれ、紀元3世紀から5世紀にかけて成立）で構成されています。

日本の歴史や古典と比較してみますと、大和朝廷が東北地方を除く日本本土の大半を統一したのが4～5世紀、古墳時代は3世紀から7世紀まで続きました。現存する日本最古の歌集である万葉集は淳仁天皇時代（759年）までの350年間に詠まれた4500首を集めているわけであり、古事記の成立が712年、日本書紀の場合が720年ですから、タルムードの時代がいかに古いかが分かります。

しかも、ミシュナよりも前に成立した聖書（キリスト教で言う旧約聖書）の内容が歴史的記述や預言、詩歌などであり、キリスト教の新約聖書のそれが同じく歴史的記述や書簡、預言などで成り立っているのに対して、タルムード（とくにゲマラ）の大部分は論争で構成されています。冗談に、ユダヤ人が3人集まると4つの政党ができる、と言われるほど、ユダヤ人の議論好きは有名ですが、なるほど、こんなところにその淵源があるのかと思ったりします。いわゆる腹芸で、お互いに口に出さなくとも何となく分かり合っ

た違いです。人にとって守るべき法（「ハラハー」、語源は「歩く」を意味する動詞）は何かについて、大勢の賢者（ラビ）たちがああでもない、こうでもない、と、丁々発止の議論を戦わせています。これが記紀が書かれるよりも200年以上も前に編纂されたのかと思うと、まったく驚いてしまいます。

最近、日本でもタルムードについての関心が次第に高くなり、紹介する書物も数多く出版されています。最近亡くなったフランスの哲学者エマニュエル・レヴィナスのタルムード講解も何冊か国文社から出ています。しかし、現在行われているタルムード刊行は、(株) ライブと凸版が中心となって、企業メセナとしてタルムード本文全巻を日本語化しようとするものです。既にメガラー、ケトゥボート、ソーター、スッカーの各篇が上梓されました。企業メセナであるため、市販は行われず、公立図書館で利用できます。

ミシュナはヘブライ語で書かれ、ゲマラの部分はアラム語で書かれています。本来ならば、これらの原語の専門家が直接日本語に訳せば良いのですが、何しろ膨大な量の文書であるため、到底十分な数の専門家を日本で見つけることはできません。そのため、英語訳を原本として何人かの翻訳者で手分けして和訳し、専門的な学者の先生方にチェックして頂く方式で作業が進められています。

私はメガラー篇の翻訳をお手伝いし、現在はベラホート篇の翻訳に従事しています。実際にタルムードの翻訳を始めてみて、さまざまな問題にぶつかりました。一つには、先にも述べたように、英語からの重訳であるため、原意を正確に把握することが難しい点です。しかし、幸いなことに、幾つかの英訳が既に存在していますので、比較対照することによって、かなり原意に近づくことができます。また、重要な点については、ポーランド生まれで後にアメリカに移住したラビであり、辞典編纂者でもあったマーカス・モルデカイ・ジャストロウ (Marcus Mordecai Jastrow, 1829-1903) が編纂した Dictionary

of the Targumim, the Talmud Babli and Yerushalmi and the Midrashic Literature (vol. 2 1886-1903) という優れた辞典があり、大いに助けられています。

しかし、一番大きな問題は、議論している賢者たちがお互いにその道の専門家であるため、文章の多くの部分が省略されており、文体がきわめて簡潔であることです。少し誇張した言い方をすれば、時にはまるで暗号のように意味を補って解説しなければならない場合があるということです。事実、タルムードを文化的背景の異なる日本語に翻訳することは不可能である、とわざわざ伝えに来たイスラエルの学者もいた、という話も耳にしました。確かに、15、16世紀も昔のユダヤ人の法律解釈、説話、論争によって明らかにされている彼らの思考法、そのメンタリティは、現代の日本に住むわたしたちとはあまりにも異質な部分が多いのです。

しかし、幕末から明治にかけて、わたしたちの先達は自分たちの小さな世界から目覚めて外界に目を向け、それまでの考え方とはまったく異質な世界観を日本に紹介しました。その際に、外国の考え方を日本人に理解できるように翻訳した先輩諸人は、言葉では表現できないような苦勞をしてくれたわけです。新約聖書が日本語に翻訳される場合にも、ヨハネ伝1:1の英訳で In the beginning was the Word とある原文は最初、「ハジメニ、カシコキモノゴザル」と訳されたそうですが、これなども「ロゴス」という概念とは無縁の当時の人々に理解して貰おうとして試みられた見事な翻訳であると思います。

日本の法律にしても、ローマ法に始まる欧米法を母法とするものですが、明治以来の先人たちは大変な努力を重ねてこれを継受し、偉大な文化財として形成するに至りました。この他、経済の分野でも福沢諭吉が複式簿記法を翻訳して導入した（彼は理解しなかったという説もありますが）ことは著名な例ですし、科学技術の分野に至っては、いちいち枚挙にいとまがありません。

タルムードの翻訳・刊行がこれからの日本の社会にどのような影響を持つに至るかはまったく分かりません。その内容が日本社会にどのように消化され、理解されるのかは、今後の問題です。わたし自身は学者でも何でもありませんから、今はただ嵐のように押し寄せる専門知識や術語の流れにおぼれそうになっている毎日を過ごしているだけです。しかし、

欧米やイスラエルには優れたユダヤ学、タルムード学の伝統がありますので、何とか現在の日本人の理解できる形でタルムードを紹介できれば、と願っています。

ただ、翻訳者の端くれとして、たとえ重要ではない、小さな事柄であっても、「言葉」に興味を持つ「職人」でありたいと思っています。卑近な例で恐縮ですが、例えば最近訳していたペラホートの一節に「キヌーアハ・セウダー」という単語が出てきました。英訳では dessert になっています。これを「デザート」と訳すことは簡単ですが、現代では日本でもデザートという概念はあたりまえの事だとしても、紀元5世紀のバビロニアの文書の訳語として使用することが適切であるか、いささかの抵抗を感じました。例のジャストロウの辞典を開いて見ますと、字義通りには「口を拭うもの」という意味であり、別の箇所では食後に口を拭うものとしてはパン以外には使用してはならない、などという規定があることが分かります（フリーン105a）。ここを訳しながら、日本語には料理のことを卑下して「お口汚し」という呼び名があることを思い出しました。そこで、5世紀の文書に使われる言葉として、食後のデザートというのはどうもうまくないので、「食後の茶菓」と訳すことにしたわけです。

どうしても良いような事ですが、なぜ、こんなことが問題になるのかといいますと、ユダヤ人の場合、食後（場合によっては食前）に食物について神に感謝する祝祷を唱える必要があるからです。そこで、どの種類の食物を食べるときには、どの種類の祝祷を唱えるべきか、ということを学者がいちいち議論する社会——これは、日常生活の細部に至るまで、細かく明示された規定を作り上げて、これに従って生きて行く社会であるわけです。われわれにとっては、いささか、窮屈で仕方がない感じはします。しかし、一つ言えることは、この様に言葉を大切にし、行動に責任を持つ社会においては、「ビデオを見せたかどうか記憶にはありませんが、そのような明示的な証拠があるのなら、見せたのでしょ」などという某テレビ会社のような言い逃れは存在できないのではないかと思います。

言葉よりも気持ちが通じ合うことを重視するという日本社会の特性は優れたものではあると思います。しかし、社会がますます開かれたものとなり、異文

化を持った世界中の人々との接触が次第に増大する時代にあつて、タルムードの賢者たちのように自由に議論する姿勢も大いに学ぶべき点がある気がします。日本の社会とは非常に異なる社会に属する文書であるだけに、タルムードを翻訳するという事は、ほんとうに精神的に刺激を受ける作業です。苦労はしていますが、楽しくもあります。

最後にタルムードを機械翻訳で訳すということについてですが、一つには構文があまりにも複雑であり、内容が専門的すぎるため、事前処理・事後処理に時間を取られ過ぎますから、現時点ではあまり効率的ではないと思われます。私としてもその方面の専門知識が乏しいし、たくさんの種類のソフトウェアを扱った経験ありませんので、確かなことを述べることはできません。しかし、乏しいながら経験した範囲で言いますと、どうも現在のソフトウェア

は複文を扱うのが苦手なようです。関係代名詞の that や which を指示代名詞や疑問詞としてしか扱わないようなので、できるだけ単文に処理してから訳させています。いきなり、機械翻訳を使わずに自分で訳すのと、機械翻訳をしてから処理をするのと、どちらが効率がいいか、現時点では何とも言えないと思います。しかし、もう少しソフトウェアが発達し、機能が強化されるならば、やがて十分に実用的なものとなるのではないのでしょうか。その日が早く来ることを大いに期待し、祈っています。



岩下 陽示 (いわした・ようじ)

1935年生。大分県中津市出身。日本ソールクリニック聖書学院卒。中央聖書学校中退。広島、横浜、京都などにおいてキリスト教の牧会に従事。1981年以降、宗教経済、技術関係の翻訳者となる。1986年、(株)ハイテック設立。現在、代表取締役。

協会便り

第6回通常総会の開催

開催日時： '96年6月17日(月) 16:00~17:00

開催場所： 機械振興会館 地下大ホール

(東京都港区芝公園3-5-8 会場 ☎03-3434-8211)

議 題： 第1号議案 95年度事業報告および決算報告案

第2号議案 96年度事業計画および予算案

第3号議案 役員の改選

第4号議案 その他、会員提案事項

協会便り

'95年度成果報告会の開催

開催日時： '96年6月17日(月) 13:00~15:45

開催場所： 機械振興会館 地下大ホール

(東京都港区芝公園3-5-8 会場 ☎03-3434-8211)

スケジュール：

13:00 開会

13:05~14:00 講演「21世紀における機械翻訳」 長尾 真会長 (京都大学教授)

14:00~14:30 成果報告① 市場動向調査委員会 石崎 俊委員長 (慶応義塾大学教授)

14:30~15:00 成果報告② 技術動向調査委員会 村木一至委員長 (日本電気 部長)

15:00~15:15 コーヒーブレイク

15:15~15:45 成果報告③ 例文評価研究会 横山晶一座長 (山形大学教授)

15:45~ 閉会

司会： 田中穂積理事 (東京工業大学教授)

参加費： 無料

翻訳ソフトの市場動向

株式会社ノヴァ 代表取締役 古賀勝夫

□ダウンサイジング

我国の商用の翻訳ソフトは、1985年に始まるが、始めの数年は主に大手コンピュータメーカーにより、汎用機を中心に展開された。もとより、汎用機ベースではコストパフォーマンスという意味では、普及すべくもなかった。

この傾向は、ワークステーションという高性能・低価格のハードウェアが一般に広がりだした1987年頃から、顕著に変化を見せはじめ、翻訳ソフトの中心は、ワークステーションに移っていった。しかし、それでもハード・ソフト一式で購入するとなると標準セットで600万円以上はするような価格で、一般に普及するにはなかなかハードルが高かった。

こうした中で、1980年代から普及を見せ始めていたパソコンが、8ビットから、16ビットへ、さらに32ビットへと高性能化し、CPU、メモリー、ハードディスクといったコンピュータリソースをふんだんに食う翻訳ソフトにとっても、パソコンでの稼動が充分現実的な段階になってきた。

1990年には、パソコンの主流が32ビットへと推移する時期に、1991年弊社ではPC98上で、初めてワークステーション版をパソコン版へ移植した英日翻訳ソフト「PC-Transer/ej」を発売した。

これは、当時ソフト価格として248,000円であり、従来の翻訳ソフトのニーズを大きく変え、「翻訳者が翻訳のため使う」というより、「英文情報の入手者が、内容を楽に、早く把握するため使う」という使い方が主流になってきた。言ってみれば翻訳ソフトを気楽に使う、それでいて、情報の流れは、大いに向上する、という効果を産み出したといえるのではないだろうか。

1992年以降は、着実にパソコン翻訳ソフトが主流を占めだし、現在に至っている。

(表1)

□低価格化傾向

1994年中頃までは、パソコン翻訳ソフトは、ロゴヴィスタ、沖電機、富士通と新規参入を入れながら、

価格的には20万円弱で推移していた。

これは、1994年11月にカテナが9,800円の英日翻訳ソフト「コリャ!英和」を出して一気にくずれ、以降次々と低価格翻訳ソフトが出現している。

この傾向は、パソコン翻訳ソフトを広く普及・認知させるということでは、大いに貢献していると言える。実際、パソコン翻訳ソフトの普及本数は、飛躍的に向上している。

今後もこの傾向は、続いていくことが想定されるが、製品系列としては、手軽に簡単に使える低額商品と、翻訳機能を充実させた高額商品とに分化していくのではなからうか。

(表2)

□インターネットの影響

翻訳ソフトの普及、低価格化に強い影響を与えているものに、インターネットの普及がある。インターネットは、90年に米国で始まり、日本でも93年より始まっているが、とりわけ95年以降、急速に伸びており、現在世界ではユーザー数8千万人以上、日本でも、50万人以上はかたいと言われている。

インターネットは、世界中の8000万人以上がネットワークを通して、お互いに瞬時に情報のやり取りができるという意味で、これまでにない画期的なインフラといえ、「情報革命」といっても大袈裟ではない。

このインターネットの世界では、90%以上が英語であるため、ユーザにとっては、従来よりはるかに日常的に英語に触れる場面が増えている。また、情報というものは鮮度があってこそ価値あるものであるが、英語情報をリアルタイムに理解していくという意味では、英語が多いということは日本人にとっては、大きなハンデである。

この状況下で、翻訳ソフトは気軽な速読用ツールとして、認知されだしており、その意味で、インターネットが翻訳ソフトの普及に対して、起爆剤になるきざしが見え始めてきた。

□今後の市場動向

1年半程前の段階では、翻訳ソフトの出荷本数は、月あたり2、3千本程度ではなかったかと予測されるが、最近ではその10倍以上は堅そうである。

この飛躍的な伸びは、翻訳ソフトの低価格化とインターネットの普及によるところが大であろう。

今後もインターネットユーザは毎年数倍は増え続ける見通しであるので、翻訳ソフトも、これに比例して増え続けることは確実であろう。

さらに、最近モデムやパソコンに翻訳ソフトがバンドルされるといったことが増えており、既に月当たり数万セットのオーダーには達しているものと思われる。この傾向も増え続けると思われるが、これには、落とし穴があり、ソフトメーカーにとっては、バンドル価格が余りにも低いため、採算がとれないというジレンマがある。このため、バンドル製品は、通常のパッケージ製品そのものではなく、その機能限定版や体験版として、それからのアップグレードサービスをねらうといった、傾向も出はじめてきた。

最後に翻訳ソフトの普及が今後見込まれる一方で、翻訳精度の向上に対するユーザの強い要望があることも事実であり、これに対してメーカー側は恒常的に努力をしていく必要がある。

表2. 翻訳ソフトの最新動向

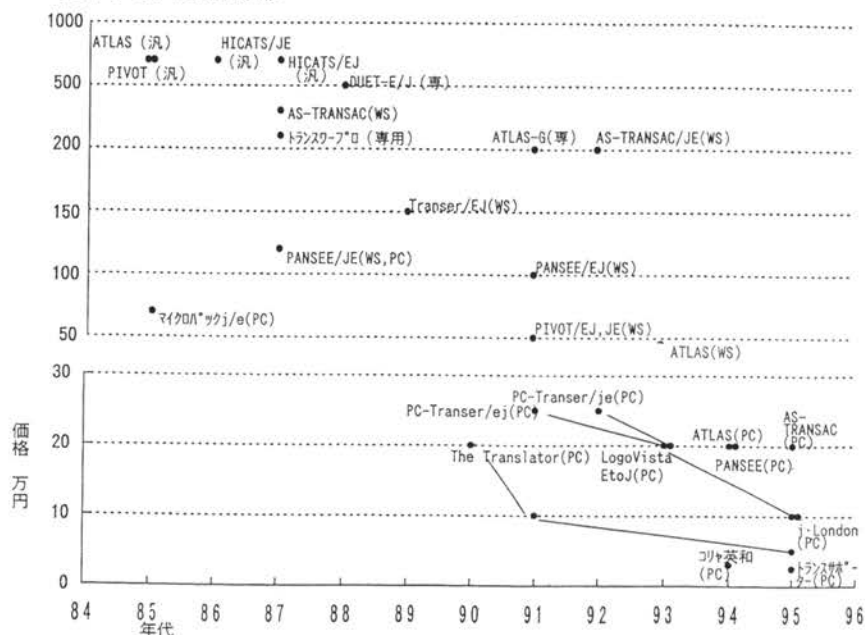
<汎用翻訳ソフト>

1993年	・ノヴァ 「PC-Transer/ej, je」	198,000 円
	・カテナ 「The Translator」 (英日)	98,000
	・ロゴヴィスタ 「LogoVista EtoJ」	194,000
	・亀島産業 「E-JBANK, J-EBANK」	97,000
1994年	4月・沖電機 「PENSEE」	198,000
	6月・富士通 「ATLAS/Win」	198,000
	11月・カテナ 「コリヤ! 英和」	9,800
1995年	2月・三洋電機 「ザ・国際人」	25,000
	3月・高電社 「J・London/EJ, JE」	97,000
	4月・東芝 「ASTRANSAC」	198,000
	6月・富士通 「アトラス E J・J E」	58,000
	7月・沖ソフトウェア 「パンセ V」	38,000
	8月・東芝 「ASTRANSAC」	63,000

<インターネット用英日翻訳ソフト>

1995年	8月・ノヴァ 「NetSurfer/ej」	29,800
	・亀島産業 「Transpad」	14,800
	9月・高電社 「Worldnet/EJ」	19,800
	10月・メディアビジョン 「Dr.サーブ」	14,800
	11月・沖電機 「パンセ・フォー・インターネット」	9,800

表1. MT製品推移



Windows 版 PENSEE の商品化と市場への取り組み

沖ソフトウェア（株）企画室 商品企画 石川 友隆

1 はじめに

沖ソフトウェア（株）（以下、弊社と略す）では、94年5月より、Windows 版 PENSEE を商品化発売しているが、本稿では商品化の背景とねらいを中心に述べ、市場への取り組みについても合わせて述べる。

2 商品展開

- 94年 5月 PENSEE for Windows V. 1
- 95年 4月 PENSEE for Internet サーバタイプ
- 95年 5月 PENSEE for Windows V. 2
- 95年 8月 PENSEE V
- 95年 11月 PENSEE for Internet パーソナルユース

3 商品化のねらい

従来、沖電気では大阪ガス、オーガス総研と機械翻訳システム PENSEE の研究開発および商品化を行ってきた。しかしながら、PENSEE のみならず、機械翻訳全般が、技術的にはともかく、ビジネスの面で成功してきたとは言い難い。一方、弊社が機械翻訳ソフトの商品化を着手したタイミング、すなわち93年の時点では、正に、Windows 3.1 が急速に普及しつつある状況であり、PC 上で手軽に個人が PENSEE を使えるようにしたら市場に十分受け入れられると判断したのである。

以下に、その商品化コンセプトを示す。

（1）パーソナルユース

従来は、研究機関や翻訳サービスをビジネスとする企業が主たるユーザであったが、個人がその人の使用目的に応じて自由に使える。

（2）手軽に

従来の高価なワークステーション（WS）で使うのではなく、既にオフィスに普及しているパソコン（PC）で使える。

（3）安く

従来、50～150万円という価格に対し、まずは、企業内のパーソナルユースをターゲットに想定し20万円未満とし、市場に受け入れられれば、10万円未

満も考える。

（4）使い易く

PC 上でアプリケーション（AP）を操作する場合、ユーザがいつも使い慣れている AP と同様な作法で操作可能であれば、事前学習を省いて直ぐに使い始められる。ここに、Windows 上で動作させる最大のメリットが存在する。

4 従来の問題点

従来の機械翻訳システムは、なぜビジネスとして成功できなかったかを考察し、弊社の商品化コンセプトに課題として反映させた。

（1）翻訳結果に対する、ユーザ・市場の期待レベルと現状の大きなギャップ

機械翻訳に関する研究や技術開発については、国内有数の研究機関や企業などが強力に推進してきた精一杯の成果であり、理想に対しては不十分であるが現実として受けとめざるを得ない。それにしても、ユーザ・市場の期待レベルが余りに高過ぎた。

（2）研究成果をそのまま商品とした

訳質が不完全である以上、前編集／後編集が必要であるが、研究開発指向で商品化した場合、通常、ユーザの立場にたった商品にはなりづらい。たとえば、訳質向上に主に精力を注ぎ込んだため、操作性に対する価値観が不足していた。また、辞書登録機能やワープロソフトとの連携機能などのサポート機能について不足していた。

（3）動作環境が機械翻訳にとって不十分

従来は、ハード面では、CPU、メモリなど性能、容量が非力、不十分であった。ソフト面では、GUI の出現・普及以前であり、とりわけインタラクティブな操作性の実現が困難であった。これらが理由となり、結果としてシステムの開発上の制限、都合をユーザに押しつける結果となった。

（4）高価格

研究機関や翻訳サービス企業などが主たるユーザであるため市場が小さく、研究コストを回収しようとする、価格的に高く設定せざるを得なかった。

価格が高ければ幅広く普及しない。普及しなければ、改良が進まないという悪循環に入っていた。世の中に普及させるためには、まず数をばらまく政策も必要である。

(5) 機能的不完全商品

完璧な訳が出せなければ、一般論として不完全商品であり、販売する立場である営業マンに言わせれば、まったく売りづらい商品であった。(また、現在も不完全である。)したがって、不完全さを前提とした商品設定と売り方、セールストークが必要であった。

5 Windows 版商品化に際して

前述の商品化コンセプト並びに従来の問題点を踏まえ、Windows 版商品化に際して以下の通り推進した。

(1) 訳質などの翻訳性能は、現状の水準を前提として商品設定

あくまで現状の水準を前提にして、不十分な点に対し適切なサポート機能を設定し、不完全さを補うものとした。ただし、翻訳性能の向上は当然継続しており、向上されればそのバージョンに対し新たに装備されるべきサポート機能を付加すればよい。

(2) 操作性の徹底追及

開発プロセスをスパイラル型とし、使い勝手は実際にユーザに使っていただき、ユーザの声を取り込んでさらに改良していくのが結局は良いモノになっていく道筋である。研究指向でできたモノを黙って使えでは使いづらいソフトにしかない。

(3) パッケージ指向

広く市場に普及することを前提とした価格設定とした。PC 向けパッケージソフトの価格は、メーカが設定するのではなく、市場が設定すると言われている。購入してもらえぬ価格設定は必須条件である。数多くのユーザに使われ、より多くユーザの声を次のバージョンに反映できたソフトが生き残るというのがパッケージ市場のメカニズムである。

(4) 開発体制

従来の研究開発指向の体制に加えて、ユーザ指向を強化した。具体的には、操作性がキーポイントになる事務処理/OA システム系のメンバと Windows を精通しているメンバを投入した。これにより GUI をフルに活用した操作性の実現や業界初の Word や一太郎との連携機能を装備させることができた。

(5) オープン指向

従来の専用クローズシステム指向からスタンダードな動作環境を前提としたオープン指向とした。当然、ワープロ、表計算、OCR ソフト、音声合成ソフトなど各種アプリケーションと連携、連動が可能である。

(6) ユーザカスタマイズ指向

パッケージであるから、個々のユーザに対し変更・追加は実施しない。その替わり、各種カスタマイズ機能や環境設定機能、ユーザ辞書などにより、ユーザ自身の使い方、ニーズにフィットさせることが可能である。

6 これからの市場への取り組み

市場動向として、PC の爆発的普及 (95年出荷台数は550万台、96年度は750万台と予想)、情報システムのネットワーク化、インターネットの急速な発展などがあげられる。とりわけ、インターネット WWW の普及は機械翻訳ソフトにとって強烈なインパクトとなっている。WWW で発信される多くの情報は英語ベースであり、日本人にとって英語を日本語化してくれるツールとしてインターネットサーフィンには必須アイテムとなりつつある。95年4月に、国内初 (たぶん、世界初) の WWW のページをレイアウトを保存しつつ英日翻訳させる PENSEE for Internet を発表発売した。その後、類似商品が他社からも発売され、ユーザから好評を得ている。このことは、情報が電子化、デジタル化され、ネットワークを通じて世界のあらゆる情報機器からアクセス可能になりつつあることを示し、したがって、今後益々、ユーザは英日 (今後、日英もニーズはありえる) 翻訳ツールを日常的に活用し、情報収集加工することになる。

このような市場動向の中で、我々は以下のことに取り組みたいと考える。

(1) 飛躍的な訳質向上

(2) 翻訳時間のリアルタイム化

(3) 辞書情報の共有、統合化、標準化

(4) 各種関連アプリケーション、ネットワーク環境などとの融合

(5) 翻訳エンジンを活用した各種目的別アプリケーションの商品化

7 まとめ

今後、21世紀にかけ (といってもわずか5年たら

ず)、情報機器は高性能化、大容量化が急激に進む中、本格的な機械翻訳の時代がやってくるわけであり、一方、自然語理解や知的情報処理に関する研究成果が機械翻訳に本格的に応用されれば、訳質が人間の能力に急速に接近することが予想される。この3年間は、各社とも基本的に、既に開発済みの翻訳エンジンに操作性、サポート機能での改良を加え、市場投入し、それなりのビジネス上の成果を得た。しかし、これらは本来市場に受け入れられるべき水準にあるにもかかわらず、操作性、価格面などの問題点がゆえに思い通りにならなかった部分を是正したに過ぎない。

しかし、今後の3年間で展望する時、機械翻訳の有用性が社会に認知されつつある以上、いよいよ機

械翻訳の本質的なブレイクスルーが期待されるわけであり、弊社もまた全力をあげる所存である。

PENSEEは大阪ガス(株)及び沖電気工業(株)の共同開発による製品です。

PENSEEは大阪ガス(株)、(株)オージス総研、沖電気工業(株)の登録商標です。

PENSEE for Windows及びPENSEE Vは(株)オージス総研、沖電気工業(株)及び沖ソフトウェア(株)との共同開発です。

PENSEE for Internetは沖ソフトウェア(株)が開発し、製品化したものです。

Windowsは米国Microsoft Corporationの登録商標です。

各製品名は、各社の商号、商標または登録商標です。

イベント報告

ビジネスショー'96報告

5月14日～17日の4日間、東京臨海部有明に新たに建設された東京ビッグサイトにて、ビジネスショー'96が開催された。今回のビジネスショーでは、メーカ、ソフトハウスから商品化されている翻訳ソフト、システムが多数出品され、来場者の関心を集めていた。それらを挙げると以下の通りである。

- ・三洋インフォメーションビジネス(株)
インターネット ザ・国際人 EJ V3.0 (WWW 対応レイアウト保存翻訳)
- ・日本電気(株)
Netscape/ej 英日辞書引機能付インターネットブラウザ
- ・松下電器産業(株)、九州松下電器(株)
インターネット対応英日翻訳ソフト Webpartner (WWW 対応レイアウト保存翻訳、参考出品)
- ・シャープ(株)
ノートパソコン Mebius A400シリーズ搭載のインターネット対応英日翻訳支援ソフト (WWW 対応レイアウト保存翻訳)

- ・(株)ノヴァ
Netsurfer/ej Ver. 2 (WWW 対応レイアウト保存翻訳)
- ・亀島産業(株)
インターネットの翻訳ツール Transpad for Windows (WWW 対応レイアウト保存翻訳、メール翻訳)
- ・クリエイイト大阪(株)
日→中機械翻訳ソフト 孫悟空

昨今のインターネットブームを反映して、各社とも、インターネット対応の WWW レイアウト保存翻訳 (WWW の英文ホームページを日本語で元のレイアウトのままブラウズできる機能) を強調した展示内容であった。また、従来、携帯型の翻訳ツールを販売していた富士ゼロックス(株)が、携帯電話と接続できる新製品を出品していた。

全般的には、今年のビジネスショーでは、翻訳ソフトの展示が活況を呈し、特に、一般ユーザーの間でのインターネットへの関心の高さを示しているようであった。

機械翻訳における記号の取り扱いについて

例文評価研究会

1. はじめに

例文評価研究会は、95年度の活動として、機械翻訳の現状を把握するために、次の2人の方からヒアリングを行った。

- ・井佐原均氏（通信総研、JEIDA 実用システム専門委員会委員長）
- ・桜井恵三氏（IBS 社長）

ヒアリングの結果や、その後の内部的な討論の結果を踏まえて、今年度は記号を含む翻訳困難文を取り扱うことにした。討議に参加したのは、以下のメンバーである。

座 長 横山晶一（山形大）
 委 員 芦崎達雄（JICST）
 安藤進（TCC）
 内尾淑美（富士通）
 川越睦（松下電器）
 熊野明（東芝）
 白木澤佳子（JICST）
 堂野前進（長瀬産業）
 時岡洋一（インターグループ）
 徳永信治（シャープ）
 中英康（日本電気）
 松平正樹（沖電気）
 事務局 田中孝
 （五十音順）

次節以下では、その結果について述べる。

2. 記号を含む翻訳困難文

2.1. 概要

機械翻訳の際、文の中の種々の記号が翻訳の障害になることは、実際に翻訳を行えばすぐに分かることである。しかしながら、記号の問題は、研究の対

象となりにくく、比較的軽視されてきた面がある。今年度の例文評価研究会では、この問題を取り上げ、現在の機械翻訳システムで文中の記号がどの程度処理できているのか、また、前処理（現在機械翻訳の障害となる記号はほとんどが人手によって前処理されている）を自動的にを行うにはどうしたらよいかについて検討した。

以下に示すものは、文中に現れる記号を分類し、記号を含む文についていくつかのシステムで機械翻訳を行った結果である。以前の報告では、すべてのシステムの結果を並記していたが、今回の報告ではいくつかのシステムによる翻訳結果の代表例を示したが、うまく訳せない、または訳語等が正しくないという場合（いくつかのシステムではできていないという場合には、日本語文のあとに#を付け、すべてのシステムでうまくいかない場合には、日本語文のあとに*を付ける。前者は、システムの対応の仕方によっては対処可能であり、後者はシステムの作り方を含めて検討する必要がある。なお、ここでは記号の扱いが正しいかどうかを問題にしておき、翻訳が正当かどうかは判断しない）には、他の多くのシステムも同様にうまく訳せないと考えてよい。

また、実際の翻訳の場合に記号処理対象範囲をどこまで（システム側で）考慮すべきかについての対応を考慮して、その対処につき、コメント（システム側として対処すべきもの）やアドバイス（現在のシステムでは対処が難しそうなので、ユーザ側の配慮を求めるもの）の形で加えてある。

2.2. 記号を含む翻訳困難文の分類

1. 項番

1.1 一般的な（数字による）項番

（1）項番とタイトルの間にスペースのない例

1. 機械翻訳の現状

（1）機械翻訳の現状

訳：（1）The present state of the machine translation.

[1] 機械翻訳の現状

【1】 機械翻訳の現状

1) 機械翻訳の現状

1: 機械翻訳の現状

1. 1 機械翻訳の現状

1-1 機械翻訳の現状

A. 機械翻訳の現状

©. 機械翻訳の現状

(外字を項番に使った場合)

α) 機械翻訳の現状

§ 2 機械翻訳の現状

* 2 機械翻訳の現状

(ア) 機械翻訳の現状#

< 1 > 機械翻訳の現状#

訳: The present state of the <1> machine translation.

(見かけ上括弧だが、実際には、大小記号)

アドバイス

上記の2文については、対応できるメーカーもあるが、項番とは見なせないことが多いので、避けた方がよい。

(2) 項番の次にスペースのある例

1. 機械翻訳の現状

(1) 機械翻訳の現状

< 1 > 機械翻訳の現状

訳: <1> The present state of machine translation.

[1] 機械翻訳の現状

【1】 機械翻訳の現状

1) 機械翻訳の現状

1: 機械翻訳の現状

1.1 機械翻訳の現状

1-1 機械翻訳の現状

A. 機械翻訳の現状

(ア) 機械翻訳の現状

©. 機械翻訳の現状

(外字を項番に使った場合)

α) 機械翻訳の現状

§ 2 機械翻訳の現状

* 2 機械翻訳の現状

アドバイス

(1) 一般的には、() の中を全角のひらがな

などにするのは避けた方がよい。

(2) 括弧も片括弧よりは、両括弧の方が望ましい。

(3) 括弧の後にスペース(現状ではメーカーによって1~2の違いがある)を入れると多くの場合にはうまくいく。

(4) 項番でピリオドを使用する場合には、ピリオドの後にスペースを入れた方が望ましい。

(5) 上の訳のように、< > のような項番と見なしにくいものも、後ろにスペースを入れるとうまく訳せる場合がある。

(3) 項番に似た表現

(1) を担当します

(1) を担当します

訳: Charge of (1) is taken.

アドバイス

上記の2文は、スペースを入れると項番と見なされ、正確な訳ができなくなる。

コメント

・ 文の要素である場合

この場合にはスペースを入れない。

1) 文中 ex. 私は(1)を担当します。
名詞句として扱う。

2) 文頭 ex. (1)を担当します。
次が格助詞または副助詞で、これらの次がひらがなでない時のみ、名詞句として扱う。

それ以外は、項番扱い。

1.2 記号による項番

・ 機械翻訳の現状

● 機械翻訳の現状

・ 8 MB のメモリ

○ 8 MB のメモリ

* 8 MB のメモリ

訳: * The 8-MB memory

・ 機械翻訳の現状

● 機械翻訳の現状

- ・ 8 MB のメモリ
- 8 MB のメモリ
- * 8 MB のメモリ

アドバイス

- ・ 1. 1. と同じように、全角のみしかない記号はなるべく使わない。

コメント

- (1) 後ろのスペース
一体で項番と見なす。
- (2) 文中に現れた記号
項番とはしない。
「・」は後述 5. の扱いで、他は名詞扱いとなる。

1.3 文字による項番

<条件 1> 節点 x で規則が適用されている#

訳：The rule is applied by <condition 1> joint x.

【条件 1】 節点 x で規則が適用されている#

条件 1) 節点 x で規則が適用されている#

条件 1： 節点 x で規則が適用されている#

<条件 1> 節点 x で規則が適用されている

【条件 1】 節点 x で規則が適用されている

条件 1) 節点 x で規則が適用されている

訳：Condition 1) The rule is applied by node x.

条件 1： 節点 x で規則が適用されている

アドバイス

- ・ 括弧の後にスペースを入れた方がよいのは、前と同じ。

コメント

- (1) 括弧または後ろの記号
一般的な（数字による）項番と同じ扱いをすればよい。
- (2) 範囲
名詞句と名詞句＋英数字を項番と見なす必要がある。
- (3) 項番内にも翻訳する必要がある。

2. 括弧

括弧全体についてのコメント

(1) 訳出位置

直前が名詞の場合

→名詞句の直後（中心語でなく最も後ろに現れる語の直後）

用言の場合

→中心語の後ろ（各訳語が分散して現れても同じ）

(2) 共通した扱い

括弧内部を別処理し、その訳文を上記訳出位置に出力する。

括弧外処理時は、括弧部を無視する。

2.1 省略（略称）

彼らは、朝鮮民主主義人民共和国（北朝鮮）を中心に活動を行なっている。

訳：They are acting around the Korean Democracy People Republic (North Korea).

国民総生産（GNP）の実績成長率は、きわめて低い。

訳：A results growth rate of the gross national product (GNP) is very low.

コメント

- ・ 二重翻訳にならないようにする。
- 辞書では省略語とそうでない語の訳は、極力変えておく。
- 同じになるとカットする。
- ・ 省略と認識する必要はない。

2.2 対訳

自然言語クラス（Natural Language Class）

訳：Natural-language class (Natural Language Class)

人工知能（Artificial intelligence, AI）

訳：Artificial intelligence (Artificial intelligence, AI)

コメント

- (1) 扱い
 - ・ ダブらないようにする。
 - 同じになるとカットすればよいが、同じにならないときは一般的には対応困難である。
 - 訳でない別の意味の英文が訳語かわからない

ときは処理しない

- (2) ダブったときに実害は特にない。しかしながら、訳が同じにならないときはまぎらわしい。

2.3 所属、年月、住所など

横山晶一（山形大）#

訳：Yokoyama Akikazu (Yamagata university)
(正しくは Shoichi)

実験動物では最長生存を達成した（7月1日）

訳：The longest survival was attained for the
experiment animal (July 1st).

教団付属医院（東京都中野区）の医師らは、診察を行っていない。

訳：The doctors of religious group attached hospital
(the Tokyo Nakano-ku) are not having a look.

コメント

- ・括弧について特別の扱いをしなくてもよい。

2.4 補足説明

膜を形成（再成長）する

訳：A film is formed (regrow).

格要素（特に主語と目的語）は、重要な意味をもっている。

訳：The case element (especially, subject and
object) has an important meaning.

入力ファイル（省略時には標準入力）をパターン検索して、結果を出力する。#

訳：It refers to the input file (inputting a standard in
case of omission) in the pattern and it outputs a
result.

検索情報（この集合体をリクエストと呼ぶ）を用意し、次の処理に進む。

訳：The retrieval information (This aggregate is called
a request.) is prepared and you go to next
processing.

コメント

- ・括弧について特別の扱いをしなくてもよい。

2.5 曜日

1月19日（金）

訳：Fri., January 19th

2月14日（水）・15日（木）*

アドバイス

- ・ユーザ側ではなるべく中黒を入れない。

コメント

- (1) 括弧の直前も含めて出てくる種類は限られているので、訳出方法も含めて特別な処理を行えばよいはずである。

- (2) (水)、(木) の訳が (water), (tree) になる場合がある。メーカー側では (water), (tree) といった訳にならずに、正しく訳出できるようにする必要がある。

2.6 漢字の読み

よい途（みち）を選んだ

訳：A good road (Michi) was chosen.

金融淘汰（とうた）の時代

訳：the time of finance selection (it obtained)

(参考) 前編集した場合

よい途を選んだ

訳：The good way was chosen.

金融淘汰の時代

訳：The ages of financial selection

コメント

- (1) 扱い

辞書に読みも持って、直前の語の辞書の読みを後方一致でマッチを取り、マッチしていれば、削除する。

- (2) 一般的には、人手による前処理以外に有効な方法はない。

2.7 要素となる場合

印刷は【印】と書いたキーを押して下さい

訳：Please push the key written on which
[sign] to the print.

コメント

- ・括弧部を名詞句と見なして扱う。

説明

ex. 大阪（梅田）で買った。

→括弧部削除しても、文が成り立つ。

印刷は【印】と書いたキーを押して下さい。

→括弧部削除すると、文が成り立たない。

3. 鍵括弧

被告は「薬物を投与して死亡した」と供述している。

訳：A defendant is stating "The medicine was prescribed for the patient and it died."

被告は『薬物を投与して死亡した』と供述している。

訳：A defendant is stating with "The drug was dosed, and you died." it.

「医療」については、予防の重視が強調された。

訳：The serious consideration on "medical treatment" of the prevention.

『医療』については、予防の重視が強調された。

訳：Valuing prevention was emphasized about 'Medical treatment'.

「社会保障体制の再構築」と題する勧告書

訳：Recommendation book entitled "Restructing of social security system"

『社会保障体制の再構築』と題する勧告書

訳：The advice document entitled "reconstruction of social security organization"

永野会長は「選挙運動を何もしない人が当選した。これは1つの革命である」と感想を漏らした。*

訳：Nagano president let out impression with "whereby it does nothing of election campaign whereby people were elected this is 1 revolution."

永野会長は『選挙運動を何もしない人が当選した。これは1つの革命である』と感想を漏らした。*

訳：The man with whom the Nagano chairman does no "election movements was elected. This leaked] and the comment which are one revolution.

(参考) 前編集した場合

被告は薬物を投与して死亡したと供述している。

訳：You are stating the defendant doxed a drug, and you died.

永野会長は次のように感想を漏らした。

選挙運動を何もしない人が当選した。これは1つの革命である。

訳：Chairman Nagano leaked the impression as follows. The person to whom anything did not electioneer either was elected.

This is one revolution.

アドバイス

・括弧を何重にもしないで2重までにする。

コメント

(1) 対象

<全角>「」『』 → (出力)'''

(2) 扱い

・括弧内部を別処理する。

・括弧外処理時は、括弧部をダミーの文とする。
(「と」の前) 他は、上記2. 7同様。

・括弧内に2文以上ある場合は、分割して、複数回、翻訳処理を通す。

・機械翻訳では、括弧内と括弧外との関係が解析できない。

・直接話法は翻訳が困難である。

4. 引用符 (クォート、ダブルクォート)

地下鉄駅構内などの“閉じられた”場所#

訳：The "It was closed." places such as the subway station premises

地下鉄駅構内などの”閉じられた”場所#

地下鉄駅構内などの‘閉じられた’場所#

地下鉄駅構内などの’閉じられた’場所#

地下鉄駅構内などの”閉じられた”場所#

地下鉄駅構内などの’閉じられた’場所#

ディスプレイに“P”が表示される

訳：“P” is displayed on display.

ディスプレイに”P”が表示される

ディスプレイに‘P’が表示される

ディスプレイに’P’が表示される

ディスプレイに”P”が表示される

ディスプレイに’P’が表示される

コメント

・強調と見なして無視する

・訳語を同様に囲む (分離した訳の場合は共に)

5. 中黒（中点）

ライブラリ・ファイル

訳：Library file

月曜日・火曜日

訳：Monday and Tuesday

ファイルを作成・更新する

訳：A file is created and updated.

月曜日と火曜日

訳：Monday and Tuesday

ファイルを作成し更新する

訳：The file is made and is updated.

コメント

（１）扱い

- ・スコープは前後とも直近の１単語のみとする。
- ・両側カタカナ（ダッシュ、ハイフン、スペースも同様）の場合は中黒を無視する。
- ・両側が名詞句の場合は並列扱いとする。
- ・両側が用言の場合、可能な限り並列扱いとする。

（２）扱う範囲

- ・共通部分を省略した表現、例えば

例文： 月・火曜日

は翻訳対象表現として認めず、「月曜日・火曜日」に書き換える方が望ましい。なぜなら、「月」＋「曜日」＝「月曜日」でないからである。

ただし、接頭語の並列表現、例えば、

例文： 新・旧役員

のように、「新（new）」＋「役員（officer）」＝「新役員（new officer）」となるケースは、入力表現として認められる。

6. 参照

統語構造を持つ構文がある〔２〕。#

訳：〔２〕 with the syntax with the syntax structure.
類型に整理した（表２）。#

訳：It arranged to the type（Table 2）.

推進団体（図10）には120の会員企業が参加している。

訳：120members of enterprises have participated in the promotion group（Figure10）.

アドバイス

- ・間にスペースを入れると有効である。

コメント

・訳出位置

- ex. 既に知られている。（１）→文末に置く
これは直後の文や句とつながったり項番と見なされたりする恐れがあるので認めない。
- ex. 既に知られている。（１）→文末に置く
- ex. 種々の研究があるが、松平〔２〕は…
→「松平」の直後に置く。

7. スラッシュ

約１／４の消費電力

訳：The power consumption of about 1／4

カートリッジの取り付け／交換時

訳：In installation / replacement of the cartridge

コマンドを解析／解釈します

訳：A command is analyzed and interpreted.

コメント

（１）扱い

- ・両側数字の場合→分数または日付
→区別しないでそのまま出力する。
- ・スコープは前後とも直近の１単語のみとする。
- ・置換可能語と見なし、「・」の場合の並列と同様に処理する。

（２）扱う範囲

5. 中黒（中点）同様に、

例文： 月／火曜日

は機械翻訳対象表現として認めず、「月曜日／火曜日」に書き換えることにするが、

例文： 新／旧役員は認める。

8. 文末記号（：などで終わる文）

ショックウェーブがやってきた！

訳：The shock wave has done!

home page を作ってみました：

訳：home page was made.：

コメント

- （１）一文切り出しプログラムの使用を行う。
翻訳用でないファイルやOCR読み込みファイル利用時に有効に活用する。

(2) 扱い

切りとり、訳文の文末に付けるようにする。
「？」は、翻訳処理時に疑問文と見なす。

9. 文中の箇条書き

新開発のプロセスでは、(1) P型の順に形成する、
(2) パターンを作り込む、(3) 膜を形成する、
という工程を採用しています。*

訳：In the process of the new development, the
process that (3) film which sets the (2) pattern
of the (1) P type formed in order is formed has
been adopted.

(参考) 前編集した場合

新開発のプロセスでは、次の工程を採用しています。

- (1) P型の順に形成する
- (2) パターンを作り込む
- (3) 膜を形成する

訳：The following process is employed in a newly
developed process.

- (1) Form in the order of P type.
- (2) Make a pattern.
- (3) Form a film.

アドバイス

箇条書きは文中に入れずに、独立した文として書く。

10. その他の記号

(機械翻訳テストはまだ行っていない)

(1) 記号「…」の例文

冒頭の文「この表では…示している。」において、…
補助動詞(みる、くる、しまう、…)を扱うために…

(2) 記号「～」の例文

他に、「～ている」「～できる」という表現に対して…

変数の2～n倍を、…

コメント

- ・ 5. 中黒(中点)同様に、

例文： 月～水曜日

は翻訳対象表現として認めず、「月曜日～水曜
日」

に書き換えることにする。

3. おわりに

記号についての問題点を取り上げ、その対処の仕
方等についての議論の経過を付した。現在の実際の
システムにおいては、対応のまちまちなものも多い
が、どこまでこのような記号に対処できるかについ
て今後さらに検討する予定である。

また、英日翻訳の場合には、ex., Fig. など、ピ
リオドを含む語がある場合に、一文の区切りの認識
が問題になることがあるが、これらについても今後
検討の予定である。

MT Summit VI

次の MT Summit は、97年10月29日から11月1日にかけて、米国の San Diego で開催されることに決定しました。

米国での MT Summit 開催は2度目です。前回の MT Summit III は、91年7月1日から4日まで Washington D.C. で開かれ、そこで IAMT および地域協会が設立されました。

次回の97年10月29日から11月1日という開催日は、San Francisco で開かれる ATA (American Translators Association) の年次大会の丁度前の週であり、また、San Diego の一番美しい時期です。

今から予定を立てていただき、多数ご参加ください。

なお、MT Summit VII は、1999年に AAMT が担当して開催することになっております。

以上

機械翻訳と品詞

(株) 東芝 研究開発センター 熊野 明

今回は、機械翻訳で使う品詞に関する話をします。品詞といえば、名詞・動詞・形容詞などを学校で習いました。機械翻訳といっても、このような品詞分類と全く異なるものを使うわけではありませんが、全く同じでもありません。また、品詞に関わりの深いものとして、活用変化があります。機械翻訳での品詞、活用の詳細な考え方は、システムによって異なりますが、計算機で翻訳するという目的が共通なので、共通の部分も多いはずです。

以下では、機械翻訳での品詞、活用を、学校文法での考え方と比較して紹介することにします。ある時には学校文法より詳細に、またある時には学校文法より単純に処理することで、機械処理の特徴が見えてくると思います。

1. 日本語の品詞

日本語においても英語においても、文中の語の品詞を決めることは、翻訳処理の最初のステップです。ここでは、日本語を機械翻訳で処理する上で、品詞の果たす役割を、色々な観点から述べます。

1. 1 学校文法での品詞・活用

私たちは学校の国語の授業で、名詞、代名詞、動詞、形容詞、形容動詞、副詞、連体詞、接続詞、感動詞、助詞、助動詞、といった品詞を習いました。活用変化としても、未然・連用・終止・連体・仮定・命令という6段を習いました。

ところが、ここで、疑問を持った人も多いのではないのでしょうか。

ー動詞の未然形には、「ない」につながる形と、「う・よう」につながる形があるし、連用形には、「ます」につながる形と、「た(だ)」につながる形がある。どうして同じ未然形、連用形なのだろうか。

ー上一段動詞「見る」の活用変化は「み、み、みる、みる、みれ、みよ(みろ)」と習ったけど、「み」は活用しない部分ではないか。

ー動詞に「ない」がつながるのは未然形で、形容詞に「ない」がつながるのは連用形。それも、動詞につながる「ない」は助動詞で、形容詞につながる「ない」は独立した形容詞。なぜ。

このような、かなり無理のある説明だったような気がします。もちろん、ある面では整理されていますが、それはあくまで人間の理解のためなのです。文法を理解するのは人間だから、常識的に判断されることを考えれば、これで十分なのです。

1. 2 機械翻訳文法における品詞・活用

ところが、機械翻訳では、常識は通じません。翻訳プログラムが教えられた辞書・文法だけに頼って、翻訳処理を行なうからです。

上で疑問を感じたことは、できるだけ客観的に整理して計算機に教えてやります。例えば、動詞では、未然形、連用形をそれぞれ2種類に分類したほうがいいでしょう。その結果、活用段は8段になります。上一段動詞と下二段動詞は、活用表の中で変化しない部分を語幹とみなせば、全く同じ活用種類にできます。

このように機械的に分類すべきものは、すべて独立したものとみなします。逆に、まとめられるものはまとめましょう。

つまり、ある面では「学校文法より詳細に」、ある面では「学校文法より単純に」なのです。

1. 3 形態素解析

文を意味のある構成要素に分割することを、形態素解析といいます。形態素という言葉は学校文法では聞いたことがないかも知れませんが、意味的にま

とまりのある最小の構成単位のことです。個々の形態素は、辞書に登録されているもので、それぞれ品詞が付与されています。

次の文を形態素解析してみます。

「彼女は外国に行きたいそうだ。」

辞書との照合だけでは、次のような区切りが考えられます。

- (1) 行き（動詞） | たい（助動詞） | そうだ（助動詞）
- (2) 行き（動詞） | たい（助動詞） | そう（副詞） | だ（助動詞）
- (3) 行き（動詞） | た（助動詞） | い（動詞） | そうだ（助動詞）
- (4) 行き（動詞） | た（助動詞） | い（動詞） | そう（副詞） | だ（助動詞）

この文の正しい解釈は（1）です。文法を知っている日本人なら、迷うことなく判断できるでしょう。というより、（2）、（3）、（4）の解釈を考えるほうが難しいくらいです。しかし、計算機では、（1）、（2）、（3）、（4）の解釈は、どれも同レベルの解釈候補です。では、計算機の気持ちで解析を進めることにしましょう。

最初に間違っていると判断できるのは、（3）、（4）です。動詞「行く」の連用形「行き」の直後には、助動詞「た」は接続しません。

（1）と（2）ではそれぞれ、助動詞「たい」を終止形とする解釈と、連体形とする解釈があります。

（2）はどうでしょうか。「たい」は終止形か連体形ですが、終止形の直後に副詞は接続できません。連体形と考えても、この後に連体修飾を受ける名詞がないので、正しくありません。

この例を見てもわかるように、機械的な処理では、人間が予想もしない可能性が出てきます。

1. 4 接続テーブル

上の形態素解析で、隣接する単語の正しさを判定するのに利用した情報は、接続関係でした。つまり、ある形態素の活用形が、ある形態素の直前に接続す

るかどうかをすべて知る必要があります。

これを計算機に教えるには、その全ての組み合わせをテーブルにしておくのが簡単な方法です。この方法は一般的に使われており、その接続関係をテーブルにしたものが「接続テーブル」と呼ばれるもので、日本語の形態素解析ではよく用いられます。下に示すものがその一部です。

前接続 \ 後接続	そ た う な た だ い だ い …				
	た	だ	い	だ	い …
カ行五段動詞					
未然形 1（書か）	—	—	—	—	○
未然形 2（書こ）	—	—	—	—	—
連用形 1（書き）	—	—	○	—	—
連用形 2（書い）	○	—	—	—	—
終止形（書く）	—	—	—	○	—
連体形（書く）	—	—	—	—	—
……					
助動詞「たい」					
終止形（たい）	—	—	—	○	—
連体形（たい）	—	—	—	—	—

前接続の欄は、接続の前に来る語尾の種類、後接続の欄は、接続の後に来る語の種類を示すものです。○で示したところが、接続OK、それ以外の組み合わせは接続できないということです。つまり、1行目は「カ行五段動詞未然形1（書か）には、『ない』が接続する」という意味です。このテーブルがあれば、全ての語の活用とそれに隣接する語の可能性がわかることになります。

接続テーブルの行数は、活用語の活用変化の数だけ必要ですが、五段動詞の終止形は活用行にかかわらず同じ接続をするので1種類にまとめることができ、一般的には200余り必要です。桁数は、助詞、助動詞、形式名詞、補助動詞などの種類、約100必要です。学校文法では、こんなテーブルは使いませんでした。「学校文法より詳細に」なのです。

1. 5 品詞の細分化

「今日日本人から聞きました。」

この文を、単純な辞書照合だけで解析すると、前半部分だけでも、次のような可能性が得られます。

- (1)「今日」「本人」「から」「聞きました」
- (2)「今日」「本」「人」「から」「聞きました」
- (3)「今」「日本人」「から」「聞きました」
- (4)「今」「日本」「人」「から」「聞きました」
- (5)「今」「日」「本人」「から」「聞きました」
- (6)「今」「日」「本」「人」「から」「聞きました」

(「聞きました」は「聞き」「まし」「た」と解析されませんが、ここでは省略します。)

「今日」「今」「日本人」「日本」「日」「本人」「本」「人」は全て名詞です。が、よく見ると文中の役割が異なるものがあることに気が付きます。名詞は後に助詞をとまって、文の主語のような要素になります。「日本は狭い。」の「日本」や、「本人から聞いた。」の「本人」はこの例です。しかし、「今日」という名詞は、「今日は寒い。」のように助詞を伴った用法の他に、「今日、彼に会った。」のような副詞のような用法もあります。この用法は、「今」にはありますが、「日本」や「日本人」「本人」などにはない性質です。

この性質は構文解析で重要です。そこで、「今日」「今」などは、普通の名詞とは区別して「副詞的名詞」とすることが一般的です。

この後、構文解析によって、正しい解釈を求めますが、正しい日本語の文として認められるのは、(1)と(3)でしょう。

1. 6 語分割

「それは本当かも知れない。」

この文を形態素解析してみます。学校で習った品詞・方法を使うと、次のように分割することになるでしょう。

それ	: 代名詞
は	: 副助詞
本当	: 形容動詞の語幹

か	: 副助詞
も	: 副助詞
知れ	: 動詞の未然形
ない	: 助動詞の終止形

かなり複雑です。

では、これを英語に翻訳することを考えます。普通、英語を習った日本人なら、「それは本当だ」は「It is true.」。それに、「かも知れない」が付くから、「It may be true.」。ごく自然な考え方です。

ここで注目したいのは、「それ」、「本当」は文中で独立した要素として「it」「(be) true」と考えましたが、「かも知れない」の部分は、せっかく「か(副助詞)」「も(副助詞)」「知れ(動詞の未然形)」「ない(助動詞の終止形)」と分割したのに、それを全く使っていないことです。

計算機でこの翻訳を行なうことも、同様のことが言えます。「か」+「も」+「知れ」+「ない」から may が得られるのではなく、「かも知れない」で一つのまとまりとして may と訳すのです。

機械翻訳の辞書では、このような場合、「かも知れない」は一つの助動詞(相当語)として扱うのが自然です。つまり、「学校文法より単純に」なのです。分解しても意味のない部分は、まとまったまま扱う。意味的なまとまりで分解する。これが計算機にとって、合理的な方法なのです。

次のような例もあります。

「によって」、「を用いて」、「のために」

などは、細かく解析すると、

「に(助詞)」「よっ(動詞【よる】の連用形)」「て(助詞)」

「を(助詞)」「用い(動詞【用いる】の連用形)」「て(助詞)」

「の(助詞)」「ため(名詞)」「に(助詞)」

のようになりますが、ここまで分解しても、処理を複雑にするだけです。「によって」「を用いて」「のために」をそのまま、助詞(相当語)として扱うほうが、機械処理としては合理的です。

1. 7 形態素生成

形態素解析は、入力文を形態素に分割する処理で、日英翻訳の最初のステップでした。これに対して、英日翻訳を行なう場合には、形態素生成という処理が必要になります。これは、英語を解析して得られた情報から、日本語を生成する過程です。例えば、

The program was not designed.

の訳文として、

「そのプログラムは設計されなかった。」

を生成するとします。この文の述語「設計されなかった」は、「was not designed」の訳です。英語を少し知っている日本人にとっては、問題とも思わない処理でしょう。ところが、計算機で行なうには、考えねばならないことがあります。

一般にここで行なわれる生成処理は、「設計する」というサ行変格動詞に、過去、受け身、否定の様相を加える処理です。過去を示すには助動詞「た」または「だ」、受け身を示すには助動詞「れる」または「られる」、否定を示すには助動詞「ない」または「ず」を生成すればいいと予想できます。

しかし、それだけでは、「た」を生成するか「だ」を生成するか、過去と受け身と否定をどの順序に生成するかがわかりません。形態素の順序の組み合わせは $3 \times 2 = 6$ 通りあります。

- (1)「設計する」+過去+受け身+否定
- (2)「設計する」+過去+否定+受け身
- (3)「設計する」+受け身+過去+否定
- (4)「設計する」+受け身+否定+過去
- (5)「設計する」+否定+過去+受け身
- (6)「設計する」+否定+受け身+過去

過去、受け身、否定の形態素としてそれぞれ2通りのどちらを出すかと、その順序を考えると全ての組み合わせは $6 \times (2 \times 2 \times 2) = 48$ 通りになります。この中のどれか一つが日本語として正しい表現のはずです。日本語のネイティブなら、「設計されなかった」となることは容易にわかります。これは、上に示した組み合わせの中なら、

(4)「設計する」+受け身「れる」+否定「ない」+過去「た」

を選んだことになります。受け身の表現に「れる」を選んだのは、その前が「設計する」というサ変動詞だからです。否定が「ず」でなく「ない」なのは、その前が「れる」という助動詞だからです。同様に、最後の過去が「だ」でないのは「ない」と接続できるようにです。

実は、順序だけでは最終的な日本語は生成できません。「設計する」は「れる」の前だから未然形で、「れる」は「ない」の前だからやはり未然形で、「ない」は「た」の前だから連用形で生成しなければなりません。

ここまでの情報が揃って、正しい表現「設計されなかった」が得られるのです。

上の例は、各形態素が、学校文法と一致していました。では、

It may be true.

を

「それは本当かも知れない」

と訳すことを考えます。形態素生成では、「本当だ」という形容動詞に推量の様相を加えることになります。「本当かも知れない」と生成することを考えると、

「本当」+「か(副助詞)」+「も(副助詞)」
+「知れ(動詞の未然形)」
+「ない(助動詞の終止形)」

になります。しかしここで「か」、「も」、「知れ」、「ない」を独立した形態素として考える意味はあまりありません。この場合は、推量の意味を示す「かも知れない」という一つの形態素として考えたほうが、知識も処理も簡単になります。「学校文法より単純に」です。

2. 英語の品詞

Time flies like an arrow.

これは有名な諺ですが、機械翻訳の上では別の意味で有名な例文です。私たちが「光陰矢のごとし」と受けとめるこの文には、機械翻訳にとって構文的に曖昧性があるからです。常識を知らない機械翻訳では、「時蠅は矢を好む。」「矢のように蠅を計れ。」といった珍訳が出てきます。なぜでしょうか。

2. 1 多品詞語

これは、英語の多品詞に起因するものです。文中の time, flies, like は次のように複数の品詞をもっています。このような語を多品詞語と呼んでいます。

time: 名詞・動詞

flies: 名詞・動詞

like: 動詞・名詞・形容詞・副詞・前置詞・接続詞

これだけでも、 $2 \times 2 \times 6 = 24$ 通りの組み合わせがあります。

例文の正しい解釈は、time=名詞、flies=動詞、like=前置詞としたものです。英語では、このような多品詞語が多く、曖昧性を増やす原因になっています。time を名詞、flies も名詞、like を動詞として解析した結果が、「時蠅は矢を好む。」、time を動詞、flies も名詞、like を前置詞として解析した結果が、「矢のように蠅を計れ。」です。

英語では、time, fly のように、名詞にも動詞にもなれる多品詞語は、非常にたくさんあります。answer, check, desire, fall, help, load, mind, play, strike, walk, … このくらいは思い付くでしょうが、英語では、主な名詞はほとんど動詞としても使われるようです。cake, dog, husband, pencil, people, roof, stomach, tree, など、普通には名詞としか思えない語でも、大きな辞書を見ると、名詞のほかに動詞があることがわかります。

もっと長い文では、かなりの組み合わせになります。仮に動詞にも名詞にもなる語が1文に10語あれば、 $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 1024$ 通りになります。この組み合わせの中なら、正しいものを見極めて語と語の関係を明らかにするのが構文解析です。

2. 2 活用変化

ところで、名詞にも動詞にもなる語はとても多いのですが、文中に現われる形でどちらか判断が付く場合もあります。

He answered.

という文では、構文解析を行なうまでもなく、answer (ed) は動詞です。語尾に-ed が付くのは動詞の活用でしかないからです。語尾が活用すれば品詞が明確になるか、といえ、必ずしもそうではありません。上の諺で現われた flies は、fly の活用変化形ですが、これは名詞とも動詞とも判断が付きません。flies には、名詞 fly の複数形と、動詞 fly の三人称単数現在形がともにあるからです。

英語の活用変化は、日本語に比べると種類が少なく単純です。日本語では、動詞の語幹に様々な助動詞、あるいは相当語句が接続して、「設計されなかったかも知れません」のように、かなり長い表現がありえます。しかし英語では、-s (-es) が付いてさらに-ing が付いて過去の-ed 語尾になるということはなく、どれか一つの変化です。したがって、活用語尾を1種類外して語尾を補正すれば、OK です。例えば、名詞の複数変化を処理するのは、次の4種類くらいあれば大丈夫です。

- (1) -ves なら、es を取るか、s を取って、v を f で置き換える。
- (2) -ies なら、ies を取って y と置き換える。
- (3) -es なら、es を取るか、s を取る。
- (4) -s なら、s を取る。

これだけの規則があれば、英語名詞の複数形を処理できるのです。日本語の複雑な語尾処理と比べれば、かなり簡単です。

ところが、大きな問題を残しています。それは、不規則変化の語です。man の複数形が men、foot の複数形が feet、という変化形です。

これを処理するには、規則は必要ありません。文字どおり、不規則変化なので、規則にはなりません。その代わり現実のデータが必要なのです。man に対して men、foot に対して feet、…のように、全ての

不規則変化を辞書に入れておかねばなりません。

この問題は、不規則変化動詞に対しても同じことが言えます。

begin, began, begun
eat, ate, eaten
make, made, made
put, put, put
.....

英語の教科書や辞書の巻末に載っている不規則変化動詞の表を、すべて機械翻訳用の辞書に変換することになります。

2. 3 他動詞・自動詞

学校の英文法として習ったものに、英語の5文型というのがありました。

1. S + V
2. S + V + O
3. S + V + C
4. S + V + O + O
5. S + V + O + C

Sは主語、Vは動詞、Oは目的語、Cは補語です。

英語の辞書によっては、どの構文を取ることができかを示したものもあります。機械翻訳辞書には、このような分類が必要です。

英語の辞書を見ると、動詞は他動詞と自動詞に分けられています。目的語をとる動詞が他動詞、それ以外が自動詞でした。学校文法によると、1. と3. の動詞が自動詞、2, 4, 5の動詞が他動詞ということになります。

ただし、他動詞は目的語をとるからといって、2, 4, 5の構文を全部とれるわけではありません。4. の構文をとるのは、give, show など一部の他動詞です。また、5. の構文をとるのは、find, make など一部の他動詞です。このような、語の振る舞いの違いは、構文解析に重要な情報です。

そこで、機械翻訳の辞書では、4. の構文をとる他動詞、5. の構文をとる他動詞を、一般の他動詞と別の品詞にすることがあります。

品詞はもっと細分化することもあります。3. の構文をとる動詞でも、目的語にthat節をとるもの、疑問詞節をとるもの、単に名詞句しかとらないもの、などに分類することができます。それなら、3. の他動詞をさらに分類しましょう。

このように、構文解析の効率を考えると、品詞を細かく分類するほうが得策です。つまり、日本語と同様に「学校文法より詳細に」なのです。

2. 4 形態素生成

日英翻訳で英語の訳文を生成する処理の最後は、形態素生成です。その一つに名詞の複数形変化がありますが、規則変化をする名詞では、

sを付けるもの
esを付けるもの
語尾のyをiesに置き換えるもの
語尾のf (またはfe) をvesに置き換えるもの

この分類情報だけあれば十分です。(2. 2で用いた情報と同じものです。)あれば十分というのは簡単ですが、裏を返せば、全ての名詞に関してそれだけの情報が必要ということです。もちろん、不規則変化の名詞は、解析のところで使った表を参照することになります。

3. おわりに

機械翻訳で用いられる品詞に関する話を、活用変化を含めて紹介してきました。特徴のいくつかを、個別に述べたので、全体としてまとまりがなくなっただかも知れません。ただ、機械翻訳に関する品詞・活用の主な話題はできるだけ取り入れたつもりです。案外単純だ、と思う人もあれば、面倒な処理をしている、と思う人もあるでしょう。

人間の理解を前提として作られた学校文法の品詞・活用と、計算機処理のために使われている機械翻訳用の品詞・活用の扱い方の違いがわかってもらえればこの話はおしまいにできます。

奈良先端科学技術大学院大学情報科学研究科松本研究室

奈良先端科学技術大学院大学情報科学研究科教授 松本裕治

1. はじめに

奈良先端大は、平成5年4月に情報科学研究科の博士前期課程学生の受け入れを始めた新しい大学院大学である。現在は4年目にあたり、博士後期課程2年までの学生が揃っている。前期課程が2学年で約250名、後期課程が2学年77名が情報科学のあらゆる分野での研究、勉学に励んでいる。松本研究室の正式名は自然言語処理学講座であり、自然言語と知識情報処理に関する広い研究を行なっている。

2. 研究室の構成と活動

現在の人員構成は、教授1名、助手2名（宇津呂武仁、宮田高志）、教務職員1名（成田えりか）、学術振興会海外特別研究員（Kristiina Jokinen）、博士後期課程学生9名、前期課程学生18名、秘書1名である。また、関連企業との共同研究なども積極的に行なっている。

研究室では、毎週1回全体の研究会を開き、研究の進捗状況の報告や関連研究紹介などを行なっている。本年度からは定例の研究会以外に研究トピックごとの研究グループを作って小人数での活動も同時に行なっている。現在活動中のグループは、統計的言語処理、HPSGに基づく日本語文法、文生成、辞書および語彙知識、遺伝子情報、対話などである。また、研究グループではないが、データや言語処理ツールの整備のためのグループや基礎的な科目（離散数学、自然言語理解入門など）の勉強会も開かれている。夏休みの短期集中勉強会や合宿なども年中の行事として行なっている。遺伝子情報については、本学のバイオサイエンス研究科との共同研究を計画しており、定期的な交流を持つ予定である。

3. 研究内容

研究室でもっとも重視しているテーマは、テキストからの言語知識の獲得である。獲得すべき知識には様々なものがあり、日本語形態素解析のためのパラメータ推定、確率文法の拡張とパラメータ推定、動詞の格フレーム獲得と意味分類、名詞シソーラス構築、日英対訳テキストからの翻訳規則獲得、特定分野のテキストからの知識抽出など、色々な観点から研究を行なっている。これらのいくつかについて説明する。

大量の電子化テキストが入手可能になったものの、ほとんどが品詞情報などのないプレーンテキストであるため、様々な加工を行うための基本的なツール群が必要である。したがって、日本語テキストの形態素解析、統語解析の精度向上のための手法の研究、さらに解析済みテキストからの語彙知識獲得の研究を行っている。

日本語形態素解析では、HMMなどの統計的手法をもちいて、品詞接続や語の生起確率などのパラメータの推定を行なっている。精度の高い結果を得るために確率パラメータの初期値の影響や、文法的知識の導入、多出現語の語彙化やスムージングなどを行ない、従来開発してきたJUMANシステムを上回るパラメータが学習可能であることがわかった。ここまでの実験はJUMANと同様、連続2単語間の接続しか見ていない（bi-gram）が、必要に応じた長さのコンテキストを自動的に学習できるための方法を開発している。

確率文法については従来の統語規則の範囲だけでなく、句の主辞の意味カテゴリや句の長さの情報を考慮した確率文法の学習方法を提案し、従来の統語情報だけをを用いる場合よりも解析精度が向上することを示した。

日本語文の係受け解析がある程度可能であるとの仮定にたち、個々の動詞の表層的な格パターンを集積することにより、動詞の属する意味クラスを特定する研究を行なっている。動詞の意味クラスを決定することにより、それぞれの表層格がどのような深層格と対応するかが明らかになるとともに、意味クラスで共有する表層格パターンを記述しておくことにより、動詞の統語的な働きを予測することができる。EDR で公開された共起辞書の係受けデータを利用することにより、深層格の推定が80%以上の精度で行なえることを確認した。

対訳テキストからの翻訳規則抽出については、いくつかの分野の日英対訳データに対して、実験を行なっている。文対応が取れた日英両文を統語的に解析し、事前に統計的手法によって計算した日英単語対の類似度に基づいて最適な構造対応を取る方法を開発した。特定の語を含む大量の文の構造対応を蓄積することにより、その語に関する訳し分けの規則を自動獲得する研究を行なっている。統計的手法による日英単語対の類似度計算についても、最近単語対だけでなく、単語連続の対についても類似度計算ができるように拡張しており、これと統語解析による構造上の対応を取る手法を用いることにより、より詳細な翻訳規則抽出を試みている。

例に基づく翻訳では、対訳例を直接翻訳に用いることを考えているが、我々の方法は対訳例のどの部分が具体的に対応しているかを自動的に判定し、それをある程度抽象化して翻訳規則としたもので、その翻訳規則は入力文の解析の段階でも有用だと考えている。また、例に基づく翻訳では大量の対訳例文の検索に膨大な処理時間が必要であり、それが問題になっているが、我々は対訳例の量にほとんど影響を受けない高速な例検索の方法も提案している。

言語知識獲得以外のテーマでは、文法規則の制約に違反した文を非文として排除するのではなく、如何に制約を緩和して、尤もらしい解釈を得るかという問題について、汎用性の高い方法を提案している。従来は、文法的エラーの種類にしたがってエラー修正のための規則が書かれる方法などが用いられていたが、単一化文法に基づいて文法的制約を統一的な

方法で記述し、解析中の制約違反（コスト）とそれを無視した場合の解析結果（報酬）の損得（コスト－報酬トレードオフ）を考えることにより、自然な解釈を得る手法を提案した。

言語に関する様々な研究をサポートするためには、基本的な処理プログラムや言語解析支援ツール、言語データなどが重要である。ツールやデータは狭い研究室だけでなく、自然言語研究のコミュニティ全体で共有していくことが重要である。現在、研究室で公開しているツールには、日本語形態素解析システム JUMAN およびその解析結果視覚化システム ViJUMAN、構文解析システム SAX およびその解析結果視覚化システム VisIPS などがある。

現在、様々なレベルの共起を獲得するためのツール、複数の辞書間の語彙の変換をサポートするツール群などを構築中である。その他、データの共有化にも努力してきた。以前、情報処理学会の言語資源共有化研究グループで調査した言語データの調査結果などを研究室の WWW ページで公開などしている。

4. その他

奈良先端大は学部を持たない独立の大学院大学であるため、幅広い経歴をもつ優秀な入学生を募集している。企業からの派遣を含め社会人経験者が多いのも特徴である。優秀な学生の短期修了や支援財団による学生の海外の国際会議での発表の援助など、研究面での様々なサポートを行なっている。奈良先端大と松本研究室のホームページはそれぞれ次の通りである。

奈良先端大：

<http://www.aist-nara.ac.jp/index.html>

松本研究室：

<http://cactus.aist-nara.ac.jp/lab/home.html>

研究室に関する御質問があればメール

（matsu@is.aist-nara.ac.jp）まで気軽にお問い合わせ下さい。

言語処理学会 第2回年次大会 参加報告

NEC 情報メディア研究所 亀井真一郎、福島俊一、野村直之、赤峯享、田村真子
シェフィールド大学 若尾孝博

3月26日、27日の両日にわたり東京工業大学を会場として、言語処理学会の第2回年次大会が開催された。論文件数は約100件にのぼり、以下に記す18のセッションおよびポスターセッションに分けて発表が行なわれた。

形態素解析、解析1、2、頑健な解析、文生成、
コーパス1、2、3、4、5、
言語分析1、2、談話解析、対話理解、
機械翻訳1、2、情報検索1、2

発表件数の点ではコーパスを用いた言語処理に関する研究の数が多い。また従来の処理の枠組を越えて言語の本質に迫ろうとする談話解析、対話理解あるいは基礎的な言語分析の研究も活発である。さらにグローバル情報ネットワークの発達を反映して、機械翻訳の研究も依然盛んである。また情報検索の高度化を目指す試みもなされている。以下では各セッションごとにその分野の傾向を紹介する。なおA1～P1はセッション番号である。

会長挨拶：

大会第1日の昼食後は、長尾真教授に代わって、昨年に引き続き学会ホスト役で尽力されている東工大田中穂積教授の「会長挨拶」で始まった。まず、会員数が96年3月現在で正会員463人、学生会員約100人となったことが報告された。具体的な活動への参加状況としては、昨年の第1回大会が予想を上回る93件の投稿でご祝儀相場かと思ったら、今年は100件を突破し、嬉しい誤算となっている。参加者の分布を他学会と比較すると、産業界の活動が弱い割には大学が強いのが特色といえる。自然言語処理という単一分野のメンバーが一堂に会する場として今後とも順調な発展を期するため、会員へのインセンティブがある。例えば機関誌にページ数の制限がなく、ページ超過で高額な付加料金をとったりしない（会場笑）。本格的なネットワーク時代を迎えた現在、一

般の人が情報発信をするのに自然言語が最適メディアである。今後とも我々が頑張って非定型、不均質な情報の処理への取り組みを益々発展させていこうではないか、という前向きな提言で締めくくられ、参加者の共感を呼んだ。

招待講演：

本年の招待講演は、「日韓・韓日機械翻訳の現状と課題」と題する韓国科学技術院 K.S.Choi 助教授による翻訳の評価・分析法に関する講演であった。Choi 先生は86年にソウル大で PhD 取得直後、NEC 情報研で日韓・韓日機械翻訳の研究に従事しており、堪能な日本語で講演された。講演内容は現行の日韓・韓日機械翻訳商品を小説、新聞他から広く採った約100種類のコーパスで評価して技術の現状を分析したものである。韓日は高電社製品のみで他は日韓である。詳細に先立って、いくつかの評価技法の長所・短所が、評価コストや対象テキストの種類、客観性・再現性等の観点で整理された。評価結果は、予想通り法律経済、技術分野が好成績で、小説やエッセイなど、ひらがなが多いテキストが点数が低かった。解析の改良のポイントは、形態素解析と語義選択。「れる／られる」の曖昧性解消など、韓日間で対応が大きく異なる助詞・助動詞の問題点も強調された。英日間では複合名詞の並べ方等が訳文の品質を大きく左右するのとは対照的に、韓日間ではより本質的な分析を要する問題に正面から取り組むことが直接的に精度向上につながる。このことが参加者の共通理解となり、非常に有意義であった。

A1/A2：機械翻訳1、2

発表内容は、翻訳手法、辞書、評価、言語分析と多岐に渡っていたが、アプローチとしては、近年の一つの傾向である自動抽出が目立っていた。また、翻訳ソフトの実用化が進んでいることの現れか、理論中心の発表よりは、語彙の収集や、翻訳の質の評価実験、言語分析などが目立っていたことも興味深

い。日本語から手話への翻訳システムの研究もそのような興味深い発表の一つであり、今後もこのような実用システムに直結した研究がどんどん進められることを期待する。

A3：形態素解析

形態素解析は、その研究の歴史が古く、実用の場での改良も積み重ねられて安定した技術になっている。しかし、ベタ書き日本語テキストを単語に切るという自然言語処理（機械翻訳においても）の基本処理であるため、精度100%へ近づけるための努力がいまなお続けられている。特に最近増えている試みは統計／確率情報を用いた形態素解析であり、今回も6件中3件がそのアプローチであった（他の3件は中国語の入力方式が1件、最長一致法や連語登録による従来タイプの形態素解析の改良が2件）。この背景には、大規模な電子化テキスト（コーパス）が利用可能になり、それらから統計量／確率値が計算・学習できるようになったことがある。

A4：情報検索1

このセッションは、大量のテキスト情報の中からユーザーが必要とする情報を、簡単に、理解しやすい形で提供する方法に関するものが集まっていた。7件の発表の内、関連する単語間、節・章等のブロック間、文書間に自動的にリンクを張ることで可読性を上げようとするものが4件、重要語や話題を抽出し、要約等を行うことで可読性を上げようとするものが3件であった。また、対象とするテキストは、辞書、マニュアル、新聞記事、ネットニュース、物語の多岐に渡っていた。これらの技術は、大量の文書の電子化やインターネットの普及等で必ず必要となるものなので、今後の発展に期待したい。

A5：情報検索2

インターネットブームを背景に自然言語処理技術と情報検索技術の融合が急速に進み始めた。テンプレートを用いた要約（情報抽出）、類似検索、マルチモーダル検索、情報の可視化などに関する発表（5件）があった。インターネット上のドキュメントが機械翻訳の対象となる機会は急激に増加してきており、上記のような情報検索技術と組み合わせた機械翻訳の利用形態はますます重要になっていくにちがいない。

い。例えば、インターネット上の膨大なテキスト情報をすべて翻訳して読むのではなく、必要な部分にのみ絞り込んで（要約して）から翻訳するとか、検索要求を日本語で入れて他国語のテキストを検索するなどである。

A6：頑健な解析

自然言語処理の技術あるいは機械翻訳の技術を広い場面で使えるようにするためには、解析の頑健性を高めていかねばならない。文字認識によるテキスト入力を対象としたもの（2件）と、音声対話を対象としたもの（2件）があった。音声認識や文字認識における誤り・ノイズを含んでいても、高い精度で読み取り、解析することが目標である。認識結果を受け取って確率／統計に基づく選別を行なう後処理的なアプローチだけでは限界があり、パタン処理（音声認識、文字認識）と言語処理の連携を強めていくアプローチが今後重要になっていくものと思う。このような頑健な解析技術は、機械翻訳ユーザから見ると、自動通訳の実現、OCR入力による翻訳精度の向上、電子メールなどインターネット上の多種多様なドキュメントの翻訳などへの展開を支えるものであり、本学会においては、もっと力が入れられてもよい研究分野であると感じた。

B1：解析1

本セッションには、様々な解析手法、およびそのための言語知識の基本アーキテクチャの提案が集められた。全く斬新な手法というよりは、従来から印欧語で検証されていた手法、たとえば暗黙のうちに語順の制約を利用していた範疇単一化文法（CUG）に工夫を凝らして日本語に適用する、といった趣向の発表が中心であった。解析の無断の除去、語彙素性の分類とのインタフェースの見通しの良さ、実装のし易さなどが各々のメリットとして主張されたが、各手法を相互に比較して優劣を決するまでの具体的評価には至っていない。もちろん、言語学的に妥当な理論に、新たに並列化など、高速・軽量な実現法を考案したり、統計手法や知識獲得とのインタフェースを付ける試みなどは今後の展開を考えると大いに意味のあることである。文法の設計誤りや解析誤りの影響が局所化されたり、一定の領域でカバレッジの評価に言及した発表には好感がもてる。

B2：コーパス1

大規模コーパスから言語知識を自動的に獲得しようという研究は、現在の自然言語処理の研究の核をなしており、非常に盛んに行われている。このセッションでは、コーパスから統語規則を自動的に獲得しようとする試み、単語の類似度を獲得しようとする試み等が発表された。その中で、特に d-bigram (単語間の距離情報を取り入れた bigram) を用いた研究が関連する部門から4件も発表されて、目を引いた。bigram や相互情報量等を用いた統計ベースの自然言語処理が、今後どういう発展をしていくのかは、特に注目される。

B3：コーパス2

6件の発表が、コーパス/対訳表現の自動抽出(3件)と対訳コーパスの自動アラインメント(3件)の2つに分かれていた。自然言語処理のあらゆる分野で必要とされる言語知識をできるだけ少ない人手と時間で大量に蓄積することが重要になってきていることが伺えた。また、対象としているのも単語単位よりむしろ複数の単語と単語の組合せや句表現であるものが目立った。このような語いを大量に収集することは自然言語処理技術の向上に直接つながる研究であり今後も期待したい。

B4：文生成

英文の精訳よりもむしろ読解を支援するシステムでの利用を考えて、英語文の順序に沿った形で句・節単位で日本文にしている研究が注目された。従来の文単位での翻訳に因われることなく、句または節を単位として英文の順に則して訳して行くのは人間の同時通訳者が言う「頭ごなし」の訳し方である。原文が英語であることがユーザーに分かっていれば、読解を結構支援出来るのではないと思われる。

B5、6、7：コーパス3、4、5

共起語に関連する研究(3件)、表現パターン、専門語、事例の獲得に関する研究(5件)、コーパス作り(2件)などが見られた。

共起語データを使っての研究では、対象が名詞よりも、今まであまり研究されて来っていない動詞である。共起語として「を」格を取る名詞や副詞に注目して、それらが動詞の意味分類にどれだけ役立つ

かの検討がなされた。

件数からみれば、コーパスから例文や事例までを含む種々の表現を抽出して来るものが一番多かった。数は少ないもののコーパス作り(2件)の発表は興味あるものであった。現在入手出来る分析済み英語大規模コーパスでは付与されている情報があまり詳しくなく、そこから自動的に文法を抽出しようすると問題を生じることが知られている。そこで、今回発表のあった研究では、更に詳しい情報を付与しようという試みであった。品詞の数を増やし、品詞に付ける意味カテゴリーの数も増やしてある。詳しい情報を付けるとなると、その一貫性を保つことが課題となるし、時間もかかるが、完成した際には是非とも廉価で研究者に公開してもらいたいと要望が出ていた。

C1：言語分析1

自然言語処理でも特に機械翻訳における意味解釈の問題を解決するために言語現象を整理してモデル化しようという試みの発表が目立った。対象はさまざまであるが、比喩や同語反復表現など個別の言語現象を扱った発表、文と文をまたいで文脈を解釈しようという内容の発表などである。文の文字通りの解釈を越えて発話の場面や状況から意味を解釈するための研究は今後ますます重要になると思われる。

C2：談話解析

個々の発話の間の繋がりを機械に自動認識させようとする談話理解の分野は、理論言語学の方で後回しにされていることもあって、言語処理の古くて長い研究分野となっている。たとえ狭い範囲の現象であっても理論的に掘り下げようというのが従来の特徴であったが、今回の新傾向として、統計や実データの重視がここにも及んできた。その背景には、EDRの共起辞書、概念辞書が広く利用可能となり、比較的「柔らかい」知識を用いた実験が新たに可能となったことが挙げられる。今回のテーマは、概念的な手がかりによる名詞間の間接照応、文章フォーマットの特徴を用いた省略代名詞の照応、省略主語の同定、共起関係を利用した話題境界推定、重回帰分析とクラスター分析、あるいは決定木獲得アルゴリズムによるテキストのセグメンテーション、である。セグメンテーションとは、文章が単なる文の集まりで

なく、結束性のある「切ったら血の出る」有機体である点に着目して、「意味段落への分割」に類する処理を、1) 表層的特徴、2) 語彙的結束性に基づいて行おうとするものである。同様の技術が、自動要約や、読みやすさの評価支援にも結び付くため、今後の発展が期待される。

C3：対話理解

従来はもっぱら一つ一つの文が独立した処理単位として扱われてきた。それは基本要素である単語とその間の関係の認定の研究に中心があったためである。しかし最近、文解析技術の基盤が整備され、一文ごとの処理を越えて対話の構造を認定しようという研究が活発化している。このセッションでは対話処理のモデルに関する発表が4件、話し言葉の解析に関する発表が2件行なわれた。対話の処理には解析の各モジュールが並行処理を行う必要のあること、また話し言葉における言い直しの認定には、言い直した句で元の発話を置き換えた時の言語制約違反コストの減少を想定するとよいことなどが指摘された。

C4：解析2

文解析の精度を高めるためのさまざまな方法が提案された。従来型の文法に代わって文節間の共起を係り受け解析に用いる方法、複文の主語推定に述語の意味的性質「外部観察性」を利用する方法、統計的学習に意味制約を取り入れる方法などが提案された。また従来から提案されている方式の詳細な検討もなされた。名詞の多義性解消への場面知識の利用の検討、解析木に確率を付与する解析法によって解析結果がどの程度絞り込めるかの検討、バイグラム確率でnグラム確率を補完する方法の検討、格フレーム選択のための代表的手法の比較、などである。従来の解析法に用例、意味、確率などの知見を統合させる試みが着実に進みつつあることが示された。

C5：言語分析2

このセッションの4つの発表の分析対象は、話し言葉におけるゼロ格助詞の役割、「よ」「ね」「な」「ぞ」「ぜ」といった終助詞の機能、助詞「で」の意味分類、日本語の助数詞の意味的分類、と多岐にわたっている。しかしこれらは共通して、いわゆる機能語に関係し、また意味の問題に深くかかわる現象であ

る。従来の研究によって形態的、統語的処理は分析が進み処理法もほぼ確立したが、機能語の役割や意味の問題は依然不明とされている。本セッションの発表は、こうした問題が今後の解析精度の向上に必須の分野であること、また同時にその説明は容易ではないことを示している。

P1：ポスター発表

ポスター発表は、他の講演会場より1階上の3階に2部屋、ワークステーション等の機材有りとしに分かれて行われた。機材を使用する側では、対話システムのデモンストレーションや、World Wide Web上の情報整理システムが展示された。会場外のリソースに実際にアクセスしながらの展示は、95年度を特徴づける発表として好評を博した。ポスターを使っている発表は、人間の意味の理解やその速度、あるいは、複合語の構文特徴、意味的特徴、イントネーションとの間の相関関係を分析したものなど、実在の知識を深く分析したものが目立つ。一般講演のスライド・ショーで、さっさと通過されるのではなく、豊富な実例を掲載したポスターを好きなだけ好きな順にじっくり見ることが出来るのが好評であった。

パネル討論：

パネル討論は、「自然言語処理の今後」というテーマで、古瀬幸広氏、辻井潤一氏、郡司隆男氏、黒崎政男氏をパネリストとして、田中穂積氏の司会で行われた。最終日の最後の時間帯に行われたにも関わらず、席に座りきれない人がでる程の盛況であった。討論は、過去(80年代)の研究と、現在の研究を比較することで、今後を占うという形で行われた。この10年余りの変化として、(1)理論重視の研究から役に立つシステムの開発へ移行した、(2)言語学の理論が混乱した、(3)コーパスベースのアプローチが盛んになった、(4)情報検索等の新しい適用分野が台頭した、(5)研究の観点から見るとむしろ退行した、等の意見がみられた。最近、自然言語処理の研究に従来の活気がなくなったという話を良く聞くが、今回の大会の発表件数を見ても、研究者の数が減少しているとは思えない。今後、若い研究者の活気のある研究活動を期待したい。

Mebius ノート PC-A 400 シリーズ搭載「英日翻訳支援ソフト」

シャープ (株)

シャープでは、ノートパソコン Mebius の新製品 PC-A400シリーズに、インターネット WWW 情報や英文ファイル等の英文翻訳をサポートするソフトとして「英日翻訳支援ソフト」を搭載致しました。

PC-A400シリーズは、インターネットを一般加入電話回線で利用するための28.8kbps モデムや、インターネットへの加入手続きソフト、WWW ブラウザソフト Netscape Navigator を搭載するなど、オール・イン・ワンタイプのインターネット対応パソコンです。

インターネットからリアルタイムに届く情報や、入手した文書ファイルの多くが英文であり、これらの情報を活用しやすくするため、初心者でも簡単に使用できる「英日翻訳支援ソフト」も合わせて搭載致しました。

この「英日翻訳支援ソフト」は、次に説明する「E/J for WWW」と「E/J for Text」の2つの翻訳支援ソフトで構成されています。

「E/J for WWW」は、WWW ブラウザソフトの1つである Netscape Navigator 1.12I[ja]を使ってインターネットの英文 WWW ページを翻訳する WWW 専用の翻訳ソフトです。WWW ページには HTML の制御文やリンク情報の URL が含まれていますが、通常の翻訳ソフトで WWW ページを翻訳すると、制御文やリンク情報まで翻訳してしまい、正しい翻訳結果が得られません。「E/J for WWW」は、これらの制御文やリンク情報を認識して翻訳しますので、WWW ページでも原文に対応した翻訳結果が得られます。

また、翻訳結果の和文ページを原文 (英文) ページと同じレイアウトで表示でき、またこれら2つのページを並べて表示できますので、見慣れたページも違和感なく見るができます。和文ページからもリンクをたどることができますので、わざわざ原文ページに戻る必要がありません。また一度翻訳が完了したページは、「E/J for WWW」のキャッシュファイルに書き込まれますので、再度同じページを



図1：原文画面例



図2：訳文画面例

表示させた場合、キャッシュファイルの翻訳結果を使いますので、すばやい表示が行えます。

「E/J for Text」は、英文テキストを簡単な操作で

翻訳することができるソフトで、Windows95のドラッグ・アンド・ドロップ機能を使い直観的な操作で起動／翻訳を行うことができます。ドラッグ・アンド・ドロップに対応するワードパッドや Microsoft Word 等の文書作成ソフトで範囲選択された英文テキスト、またテキストファイルとして登録されている英文であれば、選択文字列を「E/J for Text」のアイコンにドラッグ・アンド・ドロップするだけの操作で翻訳が行えます。また範囲選択する側のソフトウェアがドラッグ・アンド・ドロップに対応していない場合でも、テキストファイルアイコンを「E/J for Text」のアイコンにドラッグ・アンド・ドロップしたり、また翻訳させたい英文をクリップボードに取り込み「E/J for Text」を起動させるだけで英文翻訳ができます。

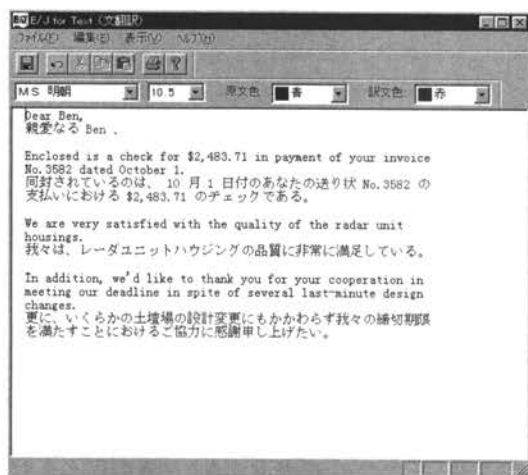


図 3：文翻訳画面例

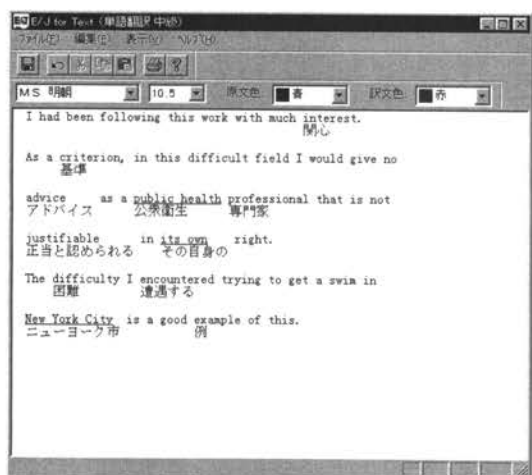


図 4：単語翻訳画面例

また「E/J for Text」の英文翻訳結果は、1 文単位で英文と和文を比較表示する「文翻訳」、フレーズに分けて翻訳表示する「フレーズ翻訳」に加え、単語に単語訳を対比させて表示する「単語翻訳」が行えます。

翻訳用辞書には、「基本語辞書」、「ユーザ辞書」「専門用語辞書」があります。

「基本語辞書」には基本的な単語として約10万9千語が登録されています。また優先する訳語の分野を「化学工学」「機械工学」「経済」「時事」「情報処理」「電子工学」「医学」「日常話題」の中から3つ選ぶことができます。

「ユーザ辞書」は、ユーザーが自由に作成できる辞書で、同時に2つまで使用できます。1つのユーザ辞書には、最大4万語の名詞が登録できます。

「専門用語辞書」としては、インターネット WWW ページの英文を翻訳するときに必要な単語が登録されており、インターネットからの英文情報の翻訳結果がより適切な日本語になるようにしています。

以上説明した「英日翻訳支援ソフト」は PC-A400 シリーズに搭載されており、ハイパフォーマンスな環境でご使用いただけます。

<PC-A400シリーズ>

CPU	Pentium プロセッサ120/133MHz
標準メモリ	16MB
表示	11.3型/12.1型 SVGA TFT 液晶ディスプレイ 800x600ドット、6万5千色表示
HDD	1 GB
FDD	3.5型1.44MB
CD-ROM	4 倍速標準搭載
モデム	28.8kbps ボイス FAX モデム搭載
OS	Windows95
WWW ブラウザ	

Netscape Navigator1.12i[ja]

Microsoft Internet Explorer2.0

注)「E/J for WWW」は、Netscape Navigator1.12i[ja] に対応しています

標準価格 55万円～70万円

<PC-A400シリーズに関するお問い合わせ先>

シャープ株式会社コンシューマーセンター

西日本相談室 TEL (06) 621-4649

東日本相談室 TEL (043) 297-4649

翻訳サーフィン V 1.0

富士通株式会社

1 概要

・利用者辞書は ATLAS と相互に利用できます。

翻訳サーフィンは、Microsoft Windows95上で動作する、WWW ブラウザ Netscape Navigator と連携して英日翻訳を行なうプログラムです。

HTML 文書のタグを翻訳結果にも反映し、英文と同じレイアウトで日本語を読むことができます。

2 特徴

2.1 高品質な翻訳

- ・高い翻訳品質を特徴とする ATLAS 英日翻訳エンジンを32ビット化して採用しています。
- ・インターネットで良く使われる単語を日本語／英語ペアで約1000単語新たに追加した、13万6千語の豊富な基本辞書を標準装備しています。

2.2 4つの翻訳方法

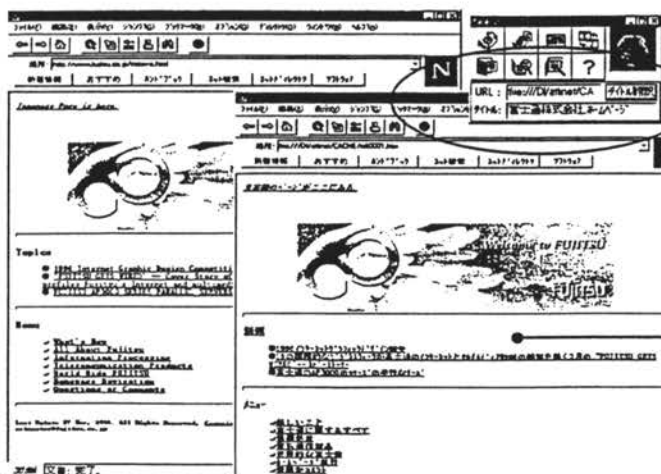
知りたい情報を効率良く翻訳するために、4つの翻訳方法を持っています。

1. 連携翻訳

Netscape Navigator で参照中のホームページをそのままの形で翻訳します。HTML タグを保存して翻訳しますので、英文ホームページと同じレイアウトで日本語を読むことができます。また、同一画面上に英文(原文)と和文(訳文)を交互に表示することもできます。

2. 一括翻訳

Netscape Navigator の表示中の英文をキャッシュに蓄えておき、後で一括して翻訳します。イ



<部分翻訳>

読みたい部分だけを範囲選択してスピーディに翻訳。

<一括翻訳>

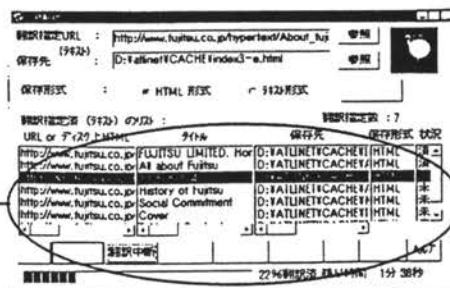
インターネット接続がオンライン状態で翻訳すると、時間課金がかかることが無視できない！そこで、指定したホームページをオフライン後に自動翻訳。

<タイトル翻訳>

表示されている英文ホームページのタイトル部分の翻訳。

<連携翻訳>

表示されている英文ホームページの1ページを同じレイアウトで翻訳。



インターネット接続時は翻訳したいページを指定するだけで、インターネットアクセスを終了後、まとめて翻訳すれば、ネットワークの接続時間が軽減でき経済的でしかも効率的です。

3. 部分翻訳

分量の多いページに対しては、必要な部分だけ範囲を指定して翻訳することができます。これにより読みたい部分だけを短時間に翻訳することができます。

4. タイトル翻訳

参照しているページのタイトル部分だけを翻訳しますから、見たいページかどうかを日本語ですぐに判断できます。

2.3 辞書引き機能

英語を読みながら、単語の意味を簡単に調べることができます。また、複数単語を一括して辞書引きし、対応する訳語を見やすく表示することも可能です（単語翻訳機能）。

2.4 翻訳チューニング

訳語や訳文のスタイルをチューニングすることができます。

1. 辞書の作成・選択

ユーザ独自の利用者辞書を作成することができます。また、この利用者辞書やオプションとして提供される「ATLAS 専門用語辞書 V3.0」の

24種類の辞書を必要に応じて利用することができます。

2. 訳文スタイルの変更

訳文スタイル（リンク情報翻訳、大文字で始まる単語の翻訳、文体）を変更することができます。

動作環境

対応 OS/機種	Windows95が動作する パーソナルコンピュータ
必要メモリ	16MB 以上
必要ディスク	31MB 以上
対応ブラウザ	Netscape Navigator V1.1以上/ 2.0（別途必要）

価格

12,800円（税別）

お問い合わせ先

富士通株式会社
クライアントサーバ推進本部
ソフトウェア販売推進部
〒222 神奈川県横浜市港北区新横浜 2-15-16
（日興不動産ビル）
Tel: 045-475-1956
受け付け時間: 9:00~12:00 13:00~17:00
月曜日～金曜日（祝日除く）

協会活動報告

理 事 会	3月28日	96年度事業計画案、96年度予算案の審議。委員会委員長の承認。
	5月24日	95年度事業報告案、95年度決算案および役員改選案の審議。
監 査 委 員 会	4月22日	95年度の監査。
	3月18日	96年度事業計画案と96年度予算案の審議。
運 営 委 員 会	5月16日	95年度事業報告案、95年度決算報告案および役員改選案の審議。総会および成果報告会についての討議。
	3月29日	AAMT J. No. 15の編集。ホームページの開設等について討議。
技術動向調査委員会	3月5日	創造的ソフトウェア、96年度活動計画等についての討議。
	5月21日	ユーザ辞書フォーマットの討議。セミナー展示会等についての討議。
市場動向調査委員会	3月8日	ヒアリング（沖ソフトウェア）。95年度の活動のまとめについて。
	5月15日	95年度活動報告のまとめ。96年度の活動計画。
例文評価研究会	3月26日	記号、固有名詞等の処理についての討議。
	5月8日	95年度活動報告のまとめ。今後の活動計画。

英和・和英辞書ソフトウェア「電辞海 V 2.0」(for Windows)

富士通株式会社

1 概要

「電辞海 for Windows V2.0L11」は、英和・和英合計で約32万語の基本辞書から単語の情報を取り出すソフトウェアです。更に、6分野合計300万語の専門用語辞書をオプションとして用意してあります。

2 特長

2.1 豊富で柔軟な検索方法

「電辞海」の最も大きな特長として、豊富で柔軟な検索が用意されていることが挙げられます。

特に、英語、日本語を意識せずに検索できることは大きな特長であり、文字列を入力するだけで「電辞海」が自動的に判断して検索を行います。

検索手段として、以下の9種類を用意しています。

・完全一致検索

検索したい単語と完全に一致した単語を検索します。

・前方一致検索

前半部が一致した単語を検索します。

・次頁／前頁検索

検索した単語の次または前の10単語を検索します。

・表記検索

日本語の検索では、読みだけではなく表記からも検索できます。

・選択語の検索

検索した内容から調べたい単語を検索することができます。

・原形検索

複数形や過去形のまま入力しても、英単語を自動的に原形にして検索します。

・クリップボードからの検索

クリップボードにコピーされているテキストを検索することができます。

・あいまい検索

記憶のあいまいなつづりを入力して検索するこ



とが可能です。入力から正しいつづりを推測するルールはカスタマイズ可能です。

・ 英文検索

英文を単語に自動分割して検索することができます。

2.2 強力な連携機能

クリップボードまたは DDE により、他のワープロ (MS-WORD、一太郎、OASYS など) の中から単語を簡単に検索することができます。

クリップボード連携では、クリップボード常時監視機能を設定することができます。この機能を設定すると、他アプリからテキストをクリップボードに複写するだけで、「電辞海」が自動的に辞書を検索して内容を表示しますので、少ないタッチ数で検索することができます。

2.3 バリエティに富んだ履歴機能

「電辞海」には、様々な特長を持った履歴機能があります。

・ 入力履歴

以前に入力した検索文字列をドロップダウンメニューから選んで検索することができます。

・ 表示履歴

以前に検索した検索結果が自動的に保存されていて、簡単な操作で再表示させることができます。

・ しおり機能

覚えておきたい検索結果にしおりを付けることにより、後で検索結果を参照することができます。

・ 電辞海メモ帳

検索結果をメモを取るイメージで電辞海メモ帳に複写することができます。参照はスクロールにより簡単にできます。

2.4 個人辞書のサポート

個人辞書を作成したり、個人辞書に対する登録、削除を行うことができます。大量の単語を登録する場合は、一括登録機能が利用できます。

2.5 複数辞書検索

「電辞海」に添付の基本辞書だけでなく、専門用語辞書や個人辞書などを接続することにより、複数の辞書を同時に検索することができます。

3 製品構成

製品名	語数	ディスク容量	価格
	万語	MB	円
電辞海 for Windows V2.0	32	20 (1.5)	24,000
情報・コンピュータ用語辞書 V2.0	25	14 (2.0)	各 50,000
科学技術用語辞書 V2.0	74	35 (0.2)	
機械工学・運輸輸送用語辞書 V2.0	29	14 (0.2)	
医学・薬学用語辞書 V2.0	59	30 (0.2)	
ビジネス・経済用語辞書 V2.0	40	17 (0.2)	
EDR 辞書 (日本電子化辞書研究所版) V2.0	73	114 (0.2)	

「電辞海」の製品構成は以下の通りです。

- ・ 注1) ディスク容量の () 内は辞書を CD-ROM 上に置いたときの必要ハードディスク容量です。
- ・ 注2) 全ての製品は CD-ROM で提供し、添付されるインストールフロッピーディスク (3.5", 720 KB) を使用してインストールする必要があります。

4 動作環境

対応機種	Microsoft Windows3.1以上が動作する機種。
	CPU は i386TM 以上を推奨
対応 OS	Microsoft Windows 3.1 または Microsoft Windows 95
	注) Microsoft Windows 95 では 32 ビットアプリケーションとして動作します。
必要メモリ	基本で 4 MB 以上 (OS 除く)
必要ディスク	ハードディスク辞書利用時: 20MB 以上 (OS 除く) CD-ROM 辞書利用時: 1.5MB 以上 (OS 除く)

お問い合わせ先

富士通株式会社

クライアントサーバ推進本部

ソフトウェア販売推進部

〒222 神奈川県横浜市港北区新横浜 2-15-16

(日興不動産ビル)

Tel : 045-475-1956

受付時間 : 9 : 00 ~ 12 : 00 13 : 00 ~ 17 : 00

月曜日 ~ 金曜日 (祝日除く)

「LogoVista E to J Pro V.3.0」「LogoVista E to J Internet」

カテナ株式会社 ロゴヴィスタ株式会社

ロゴヴィスタでは、この度翻訳支援ソフト「LogoVista E to J」製品をシリーズ化し、「LogoVista E to J Pro V.3.0」、「LogoVista E to J Personal V.3.0」とインターネット専用の新製品「LogoVista E to J Internet」を新発売いたしました。

LogoVista E to JシリーズはMacintosh版および Windows版があり、Windows95、Windows NT、Windows3.1の各環境でお使いいただけます。

《LogoVista E to J シリーズをお使いになると》

- ・ 短時間に英語から日本語への高精度な翻訳結果と訳語の統一が図れます。
- ・ 製品取扱説明書、技術文書、ビジネス・レター等、広範な種類の英文文書の翻訳ができます。
- ・ 翻訳作業の時間と経費を大幅に節約することが可能です。

* * * * *

「LogoVista E to J Pro」は、ロゴヴィスタ株式会社と米国ランゲージ・エンジニアリング・コーポレーションの共同開発により、最先端を行く言語理論に基づいて開発された英日翻訳ソフトです。

本システムは以下の特長をそなえております。

- ・ 統計言語理論に基づいて英文を解析し、単語間の構文関係を推定するパーサー
- ・ 意味関係を推敲した上で最適な翻訳結果を選び出す意味処理機構

- ・ こなれた日本語訳を作り出す日本語生成機能
- ・ 総語数63万語以上の辞書（一般辞書プラス専門辞書）

詳細な文法および意味情報が記述されている一般辞書には、英語の基本語彙を網羅する約12万8000語が登録されています。22分野の専門辞書（オプション）には、ビジネス分野および科学技術分野に必要な専門用語として総数約50万語以上が収録されています。またユーザ独自のユーザ辞書はいくつでも作成し、翻訳で使用することができます。

「LogoVista E to J Pro」は英文の前処理を行わずにそのまま翻訳したり、ユーザの選択や解釈を取り入れて、いわばユーザと対話しながら翻訳結果を作り出すことも可能です。「LogoVista E to J Pro」は、ユーザとの対話によって得られた別訳語の選択情報をそのつど学習し、将来の翻訳の際に反映させることができます。ユーザが「LogoVista E to J Pro」を使えば使うほど、ユーザの望む翻訳結果が得られるようになっていきます。

《価格》

97,000円



写真1 LogoVista E to J Pro 環境設定画面



写真2 LogoVista E to J Pro 別訳画面

《動作環境》

<Macintosh and Power Macintosh 版>

本体：Macintosh Classic II 以降の各機種、Powerbook 140以上または Power Macintosh

OS：漢字 Talk7.1以上

ハードディスク空き容量：25MB 以上（仮想メモリ含まず）

使用メモリ：推奨10MB 以上

<Windows 版>

本体：Windows3.1, Windows95, WindowsNT3.5日本語版が起動する DOS/V 互換機、NEC PC98シリーズ互換機

OS：Windows3.1, Windows95, WindowsNT3.5

ハードディスク空き容量：27MB 以上（仮想メモリ含まず）

使用メモリ：13.5MB 以上

《LogoVista E to J Pro Ver. 3.0の主な特長》

下記に示すような多彩な機能と優れたユーザーインターフェイスをサポートしています。

・「環境設定」機能

翻訳オプション（文書タイプ、文体など）、辞書（ユーザー辞書、専門辞書など）、レイアウト等の環境設定を保存

・「翻訳速度設定」機能

翻訳速度と精度の最適バランスを設定

・「未知語検索」機能

スペルミスなどの未知語を検索

・「別解釈」機能

構文、意味解析を中心とした別解釈が可能

・「別訳語」機能

英語と訳語の組み合わせでの別訳語置換が可能

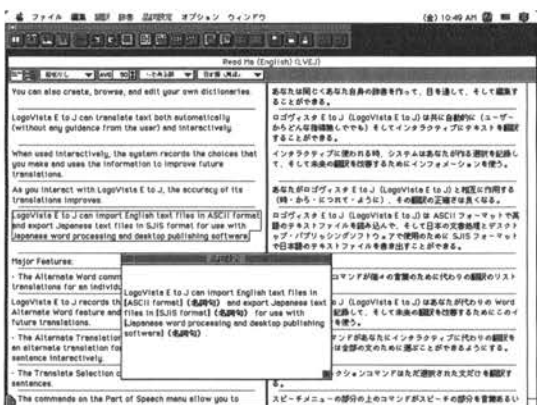


写真3 LogoVista E to J Pro 品詞設定画面

・「学習」機能

別訳語選択を学習し、学習の有無、強さの設定も可能

・「品詞設定」機能

品詞や品詞句を設定

・簡単なユーザー辞書登録機能

英文テキストボックスからツールボタンで辞書編集画面を表示

・簡単な辞書引き機能

英文テキストボックスからツールボタンで辞書の内容を表示

・「印刷」機能

翻訳された和文のみ出力、対訳（英文・和文を左右または上下に出力可能）および未知語リスト出力

・「クリップボード翻訳」機能

ワープロなど他のアプリケーション上から、テキストをドラッグし編集コマンドの[コピー]または[カット]により、LogoVista E to J Pro で翻訳を実行しクリップボードウィンドウに対訳で表示

・「SGML/HTML テキスト翻訳」機能

SGML 文書やインターネットのホームページのテキスト形式である HTML 文書を翻訳

翻訳精度を高める辞書機能として、次のような辞書が用意されています。

・一般辞書：基本的な用語を中心に英和約12万8000語を収録

語を収録

・専門辞書：オプションとして22分野の専門辞書が使用可能（総語数約50万語以上）

・ユーザー辞書：名詞、動詞（自動詞／他動詞）、形容詞、副詞、前置詞、前置詞句、接続詞が登録可能。



写真4 LogoVista E to J Pro 対訳表示画面

・ユーザ辞書の編集：必要に応じて簡単に登録、編集が可能。またテキストファイルから一括登録が可能

・辞書の優先度：複数のユーザ辞書が同時に使用可能であり、一般辞書、複数の専門辞書とともに優先順位の設定が可能

・事例ファイル：頻繁に使う定型の英文とその訳文を事例ファイルに登録、以後、翻訳時にその訳文を使用

《専門辞書》

CD-ROM 版にて下記の22分野を収録した専門辞書をオプションで販売

価格 29,800円

専門分野

ビジネス一般、法律、科学技術一般、航空宇宙、農業、応用化学、応用物理、建築、生物、化学、土木、コンピュータ、地球科学、電気電子通信、材料資源、機械、医学、船舶海洋、物理、都市工学、動物

* * * * *

新製品「LogoVista E to J Internet」をご使用いただくと、インターネットサーフィンをお楽しみいただきながら、英語のホームページを日本語に翻訳してお読みいただくことができます。「LogoVista E to J Internet」は最もポピュラーなWWWブラウザ Netscape Navigatorと組み合わせて使用し、ホームページ全体の翻訳、選択した英文テキストの部分翻訳、またはヘッダーやリンクタグだけの翻訳を行うことができます。

また、簡単な操作で辞書引きを行うこともできます。さらに、Netscape「メール」ウィンドウやNetscape「ニュース」ウィンドウのテキストをドラッグして

部分翻訳したり、辞書引きすることができます。インターネット時代の英文を読む強力なツールとしてご活用ください。

《価格》

7,980円

《動作環境》

<Macintosh and PowerMacintosh 版>

本体：Macintosh Classic II 以降の各機種、Powerbook 140以上またはPower Macintosh

OS：漢字Talk7.1以上

ハードディスク空き容量：25MB 以上（仮想メモリ含まず）

使用メモリ：for Mac 7 MB 以上、

for PowerMac 9 MB 以上

<Windows 版>

本体：Windows95、WindowsNT3.5日本語版が起動するDOS/V 互換機、NEC PC98シリーズ互換機

OS：Windows95、WindowsNT3.5

ハードディスク空き容量：27MB 以上（仮想メモリ含まず）

使用メモリ：10MB 以上

《対応ブラウザ》

Netscape Navigator Ver. 1.1以上

《LogoVista E to J Internet の主な翻訳機能》

・ページ翻訳

ページ全体を翻訳します。

・ヘッダー／リンクタグ翻訳

ページ内のヘッダーやリンクタグだけを翻訳することができます。

・部分翻訳

ページの一部を翻訳するときは、翻訳したい英文テ

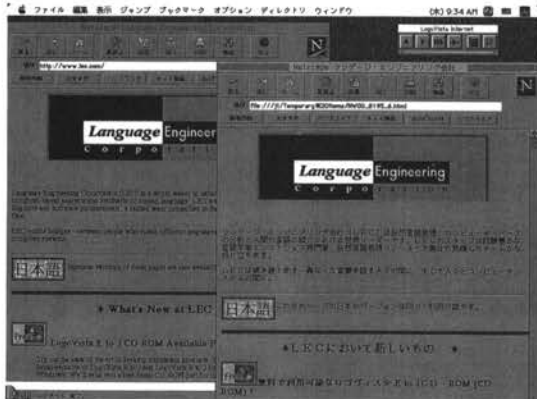


写真5 LogoVista E to J Internet ページ翻訳画面

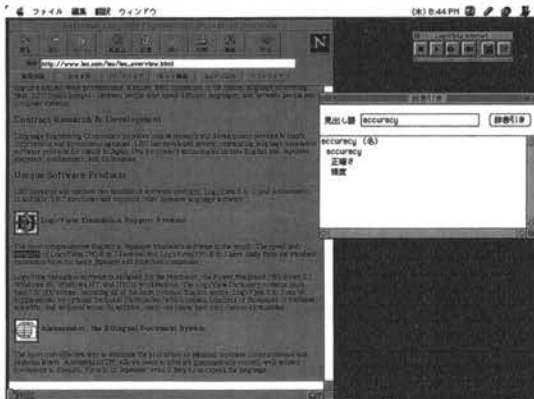


写真6 LogoVista E to J Internet 辞書引き画面

キスト部分をドラッグして選択します。指定した部分の翻訳結果を表示します。

《辞書引き機能》

辞書引きするには、次の2つの方法があります。

・ホームページで単語または語句を選択します（一般辞書には、“computer”のような単語や“United States of America”のような複合語が登録されています）。選択した語句が辞書の中で見つかり、その語句と訳語が[辞書引き]ウィンドウに表示されます。

・メニューから[辞書引きウィンドウ表示]コマンドを選択します。表示された[辞書引き]ウィンドウに語句を入力し、[辞書引き]ボタンをクリックします。

このように、「LogoVista E to J Internet」は、簡便な操作で手軽に利用できる英日翻訳ソフトです。

上記製品に関する問い合わせ先

《販売元》

〒206

東京都多摩市落合1-15-2

カテナ株式会社 ソフトウェアプロダクツ事業部

TEL: 0423-38-1118

FAX: 0423-38-1079

《開発元》

〒135

東京都江東区潮見2丁目10-24

ロゴヴィスタ株式会社

TEL: 03-5690-8531

FAX: 03-5690-1290

Internet: lvej@logovista.co.jp

WWW: <http://www.cet.co.jp/logovista>

イベント紹介

テクニカルコミュニケーションシンポジウム'96

日 時: 8月29日(金)・30日(土) 10:00~17:00

場 所: 新宿 工学院大学

主 催: テクニカルコミュニケーター協会

後 援: 通商産業省

協 力: アジア太平洋機械翻訳協会の他、新聞社、出版社、学会等30諸団体

基調テーマ インターネット時代のTC技術

基調講演 朝日インターネット誌「DOORS」編集長 矢野直明氏
「インターネット時代のTC技術について」

分科会 ライティング、ビジュアル、CALS、SGML、インターネット、オンラインマニュアル、翻訳、CSを中心に11分科会をパネルディスカッション形式で開催。

事例・研究発表 大学、メーカー、テクニカルライター、エディターの事例・研究成果38件。

特別セミナー テクニカルコミュニケーションに携わる初心者がすぐ役立つ6講座

展示会・情報コーナ マニュアルコンテスト応募マニュアル、情報雑誌が2日間展示される。

参加料金 入場券 会員 15,000円、非会員 20,000円、学生1日 3,000円

特別セミナー券(1講座分)

会員 10,000円、非会員 15,000円、学生会員 10,000円、学生非会員 15,000円

問合せ先 テクニカルコミュニケーター協会事務局

TEL. 03-3368-4607 FAX. 03-3368-5087

COLING-96の開催

COLING-96 Conference 1996年 8 月 5 日 (月) ~ 9 日 (金)
 Tutorial Programme 1996年 8 月 2 日 (金) ~ 3 日 (土)
 Workshops 1996年 8 月 4 日 (日)

会場 コペンハーゲン大学 (デンマーク)

[参加費]

PRE-COLING-96 :

Tutorials, 2 days :

Participant DKK 2,100
 University Employee DKK 1,400
 Student DKK 800

Workshop DKK 100

COLING-96 :

Participant DKK 3,900
 University Employee DKK 2,900
 Student DKK 1,900

Accompanying Person DKK 500

[招待講演]

Prof.W.J.M.Levelt (Max Planck Institute for Psycholinguistics,The Netherlands) :

A Theory of Lexical Access in Speech Production

Prof.B.Grosz (Harvard University,USA) : Discovering the Sounds of Discourse Structure

[特別講演]

Prof.M.Kay (Stanford University and Xerox Palo Alto,USA : President of ICCL) :

Title to be announced

Summary Session COLING-96 and Future Perspectives of CL

Prof.C.Boilet (Grenoble,France : ICCL member), Prof.M.Nagao (Kyoto,Japan : ICCL member),

Prof.A.Zampolli (Pisa,Italy : ICCL member), Prof.Y.Wilks (Sheffield,UK : ICCL member)

[パネル]

“Is Speech Language?”

Coordinators : Joseph Mariani (LIMSI/CNRS,France), Steven Krauwer (OTS,The Netherland) 他 4 件

[チュートリアル]

Linguistic Theory and Lexical Semantics in Machine Translation

Lori Levin (Carnegie Mellon University,USA) 他 3 件

[ワークショップ]

Fourth Workshop on Very Large Corpora

Topics : Innovative Uses and Applications of Large Corpora Contact : Eva Ejerhed,Dept.of Linguistics,

DGL,University of Umea, S-90187Umea,Sweden,

E-mail : WVLC-4 @ling.umu .se

or

Ido Dagan,Dept. of mathematics & Computer Science,Bar Ilan

University,Ramat Gan52900, Israel, E-mail : dagan@bimacs.cs.biu.ac.il

[問合せ先]
 COLING-96 :
 c/o DIS Conference Service Copenhagen A/S
 2 C, Herlev Ringvej
 DK-2730 Herlev
 Denmark
 Telephone : +45 4492 4492, Telefax : +45 4492 5050,
 E-mail : dis-con@inet.uni-c.dk
 COLING-96 :
 c/o University of Copenhagen
 Njalsgade 80
 DK-2300 Copenhagen S
 Denmark
 Telephone : +45 3532 9090, Telefax : +45 3532 9089,
 E-mail : coling96@cst.ku.dk

イベント紹介

情報処理学会 自然言語処理研究会

第114回 自然言語処理研究会

(電子情報通信学会「言語理解とコミュニケーション研究会」との共催)

日 程： 平成8年7月18日(木) 9:45~17:30

19日(金) 9:45~17:00

会 場： 豊橋技術科学大学 A1棟2階, A1-201号室

[連絡先]

豊橋技術科学大学 知識情報工学系 増山 繁

〒441 豊橋市天伯町雲雀ヶ丘1-1

Tel : 0532-44-6894 (ダイヤルイン)

Fax : 0532-44-6873

E-mail : masuyama@tutkie.tut.ac.jp

URL : <http://www.smlab.tutkie.tut.ac.jp/masuyama/>

* 1日目(18日)終了後、デモと懇親会予定。

* 発表件数は23件

* 以後の自然言語処理研究会の予定

開催日	96年9月12・13日	96年11月18・19日	97年1月	97年3月
開催場所	慶大(藤沢) 《情報学基礎研究会と合同(もしくは連続)開催》	奈良先端大	(東京)	長崎大学
テーマ	自然言語処理と情報検索 ・文書処理	「コーパスを用いた現実的な日本語研究」(予定)		
発表申込締切日	96年6月28日	96年9/6ごろ(予定)	96年11/6ごろ(予定)	97年1/7ごろ(予定)

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

●公開時間=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

データショウ基調講演会

CALS Japan'96

PCカードWORLD'96

アクセシビリティコーナー

JEIDA/JPSA共同プロジェクト(仮称)

AAMT
ジャーナル

No. 15
(June, 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 仁井正治
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