

# The Asia-Pacific Association for Machine Translation Journal



アジア太平洋機械翻訳協会

### 目 次

エッセイ	機械翻訳をめぐる現況について
ヒヤリング	システム化の観点から見た言語メディア技術の将来動向2
技術早分かり	人間の翻訳と機械翻訳
研究機関紹介	東京工業大学大学院情報理工学研究科 田中・徳永研究室11
関連記事紹介	ATA (アメリカ翻訳者協会) の状況
イベント紹介	Eighth International Unicode/ISO 10646 Conference
国際会議	LISA フォーラム参加報告
協会活動報告	10
国際会議	NLPRS'95参加報告
新刊図書紹介	「プログラムングの壷」(第 2 部 人間編)24
関連記事紹介	英日翻訳支援ソフトウェア (ザ・国際人)24
関連業界情報	翻訳業界の現状調査25
商品一覧表	機械翻訳関連ソフトウェア一覧表26
新製品紹介	英語読解ツール「訳師如来」(for Windows/for Macintosh) ·······28
新刊図書紹介	THE COMING INDUSTRY OF TELETRANSLATION目次下

### 新刊図書紹介

# THE COMING INDUSTRY OF TELETRANSLATION

# Minako O'Hagan

(Victoria University of Wellington)

### CONTENTS

Foreword, Prologue

Chapter 1: Communications Technology & The Language Problem

Chapter 2: The Translation Business and the Impact of MT

Chapter 3: Global Network of Communications

Chapter 4: Marrying Language Services to Telecommunications: The Coming Industry of Teletranslation

Chapter 5: A Teletranslation Service Epilogue; Glossary of Terms; References

### **AUTHOR INFORMATION**

Minako O'Hagan, originally from Japan, has worked in the translation industry in both the public and private sectors in New Zealand since 1984. Based on her experience as a translator, she carried out research and completed her MA at the Department of Communications Studies at Victoria University of Wellington.

MULTILINGUAL MATTERS LTD Frankfurt Lodge, Clevedon Hall, Victoria Rd, Clevedon, Avon, ENGLAND, BS21 7SJ.

☎ 01275 876519 Fax: 01275 343096

# 機械翻訳をめぐる現況について

富士通株式会社 常務取締役 吉原忠雄

この機会に、機械翻訳に関連して私が最近感じている事を述べさせて頂きたい。機械翻訳の今後の展開を考える上で、情報技術の動向と経済社会のグローバル化がキーになることは、衆目の一致するところであろう。

ここ数年のインターネット、マルチメディアの話 題の隆盛は、情報技術革命が大きなうねりで進行し ていることを感じさせる。このユーザ技術動向の大 きな流れに遅れることなく、機械翻訳システムをイ ンターネットやマルチメディアのサポートが可能な ようにチューニングするとともに、顧客満足に合っ た翻訳サービスを提供するという、高いハードルが 目前に突きつけられているのではないだろうか。更 に踏み込んで言えば、情報技術革命の進行の中で、 機械翻訳システムが広く一般の人々に使用される可 能性を我々が手中に収めつつある一方で、商品とし ての機械翻訳システムのアイデンティティをユーザ とサプライヤの間でどのように齟齬なく定められる かという課題に取り組む必要が出てくるのではない だろうか。PL(製造物責任)に繋がる翻訳品質保 証の世界がある一方で、パーソナルな満足範囲での 利便性の提供といったゲーム・ソフトに通じるよう な世界が他方にあり、難しいのはその両方の世界が サプライヤに対して隣接・混在している状況にある と言えるのではないだろうか。このような機械翻訳 システムのアイデンティティや、ユーザ・サプライ ヤの広範なコンセンサス形成に、当協会が果たされ る役割への期待も大きいものがあると確信する次第 である。

また、ハードウエア・ソフトウエア技術の進歩に基づく情報技術革命に加え、メガコンペティッションと称せられるまでに到っている経済社会のグローバル化も機械翻訳システムへの現在のインパクトに密接に連動しているものと思われる。コンパック・ショックと呼ばれたパソコン低価格化・大衆化の動

きも、生産・市場のグローバル化を背景としたコスト低減やオープン化が、日本においても米国並みのパソコン普及に向けて価格を平準化していると捉えることができよう。メーカによるパソコン価格の低廉化努力が、コンピュータ・ユーザの大衆化を促し、種々のアプリケーション・ビジネスの拡大に繋がることが期待される。とりわけ、当協会が普及を目指される機械翻訳システムについても、メーンフレーム、ワークステーションにとどまらず、広範なパソコン・ユーザにとっても有用なツールとして社会的認知が得られる絶好の機会と言えるだろう。ユーザの増加がアプリケーション機能・品質の向上につながることを期待したい。

また、経済社会のグローバル化は、情報通信技術の進歩とあいまって世界的な情報流通の量と速度を飛躍的に拡大・加速させている。しかしながら、情報受発信の基点においては、ヒューマン・インタフェースとして最も重要な役割を担うべき自然言語が、民族・国家に根ざす歴史的な壁を有し、情報の円滑な流通にとっての隘路となっている。わが国において、学校・企業を中心に比較的国際流通性の高い英語の教育の見直しや、社員の英語力向上に努める機運がたかまっているのも、グローバル化対応の象徴と言えるであろう。このような個人の語学能力の向上への依存のみならず、情報処理技術による隘路打開の手段として機械翻訳システムによるアプローチにも大きな期待がかかっている。

情報技術革命と経済社会のグローバル化という絶 好の機会を的確にとらえ、市場の要求に応えていく 上で、ユーザ・サプライヤ等の幅広い会員から構成 される当協会が果たす役割は益々重要になるものと 拝察する次第である。

# システム化の観点から見た言語メディア技術の将来動向

(出日本電子工業振興協会 自然言語処理実用システム専門委員会委員長 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長 井佐原均

### 1. はじめに

(社)日本電子工業振興協会では、わが国における自然言語処理技術の健全な発展を目指して、自然言語処理技術委員会(委員長:京都大学長尾真教授)を、また、その下に自然言語技術調査専門委員会(委員長:田中穂積東京工業大学教授)を設置している。さらに、具体的な活動を行なうために、実用システム専門委員会を初めとする3つの専門委員会を設置して活動を続けてきた。

これら委員会の活動の詳細については、平成7年 11月に開催された「マルチメディア・ネットワーク 社会における言語情報処理講演会」において、報告 された。本稿では、この報告のうち、実用システム 専門委員会の活動に関する部分を解説する。

### 2. マルチメディア・ネットワーク

マルチメディアとは何かについてはいろいろな意見がある。たとえば、画像・音声その他を組み合わせて、臨場感を持たせるということもマルチメディアの一つであると言われる。その場所に行かなくても、その場所にいるかのように、見、聞き、感じることができれば、それはそれで、素晴らしいことである。あるいは、専門家が目の前にいるかのように指導を受けることができる環境は様々な状況の下で有効に使うことができるだろう。

これらは、複数のメディアを足し合わせたバーチャルリアリティーの実現であると言える。しかし、可能であれば、その場所に行く方が、あるいは専門家と対面する方が、より良いことは言うまでもない。このような足し算のマルチメディアは、結局は現実の環境の代替物でしかないことになる。そこでの経験は新規のものではない。

マルチメディアを新しい環境であると考えるならば、従来は現実に手に入れることが不可能であった 環境を実現するものでなくてはならないだろう。こ こでは、現実世界をそのまま提示するのではなく、 必要な情報を選択的に提示する技術が有効となる。 現実の世界の模倣ではなく、情報の密度の高い仮想 的な世界を積極的に作ることが必要である。

マルチメディア・ネットワーク社会における言語 情報処理としては、情報の密度を高めるためのスク リーニング技術としての自然言語処理技術に着目し、 実用システムに展開する必要がある。

### 3. 自然言語処理技術を用いた実用システムの展開

ネットワーク社会における実用システムの開発に 当たっては、ユーザのニーズ(マーケットニーズ) を把握することが必要である。現時点でニーズの可 能性が顕在化しているものとしては、情報検索、イ ンターネット機械翻訳、電子新聞、電子ショッピン グ等が挙げられる。

情報検索においては、検索精度の向上と結果の適切な提示法の開発が必要とされる。電子新聞においては、既存の印刷媒体としての新聞とは異なったメディアとしての可能性の検討が必要である。ここでは、その双方向性、検索可能性、マルチメディアの利用が検討されよう。双方向性の存在は広告等においても、新たな可能性を開くことが期待される。広告主が容易にクライアントの情報を集め得るということは、すなわち、情報提供者が利益を得られるということであり、この種のメディアの発展には必須の要因である。

実用システムの実現には、自然言語処理の要素技術の研究の進展と同時に、情報公開やアミューズメントのあり方についての検討、また、データの鮮度をどのように保持するか等、考慮すべき問題は多い。本節では、これらの問題の内、実用システム開発のための要素技術、インターネット機械翻訳、機械翻訳システムの評価法、その他の実用システムの可能性、について述べる。

### 3.1. 実用システム開発のための要素技術

自然言語処理技術を用いた実用システム作成のための主要な要素技術には、大きく分けて知識と処理システムの二つがある。また、知識を作り出す(あるいは用例として直接用いる)元データとしての言語データに関する技術も重要である。さらに、個々の要素技術とは別に、システム全体としての頑強さも要求される。以下に、それぞれに含まれる要素を羅列する。近年、文脈自由文法を用いる構文解析手法が、遺伝子解析に用いられるなど、これら要素の新規の利用法が生まれる例もある。

(1) 言語知識として 辞書(単言語、二言語、多言語) シソーラス オントロジー ツール(作成、検索、編集)

(2) 処理システムとして 形態素解析 構文解析 意味解析 文脈解析 文脈解析 談話/対話解析 生成

(3) 言語データとして コーパス(単言語、対訳、タグ付き)タグ付けツール(形態素、構文情報、対訳)

### 3. 2. インターネットと機械翻訳

インターネットが社会のキーワード化するに伴って、インターネット対応の機械翻訳システムがいくつかの会社から発売されている。既存の機械翻訳システムをインターネット対応のシステムとするために、これらのソフトにはいくつかの特徴がある。一つには WWW ブラウザとの連携である。連携の程度には色々な段階があるが、原文と同一イメージの日本語ページを生成し、他ページへのリンク情報も保持するというものや、指定された範囲の文章を翻訳し、対訳ウィンドウに表示するものなどがある。また、実際の使われ方が、情報収集型であるので、

ある種の機能限定が行なわれることもある。速読用の翻訳を目指したり、前処理・後処理機能が省略されたりする。

また、PC上の低価格の機械翻訳システムが数多く現れてきた。英語で書かれ、かつ、電子化された文章が、インターネットを介して大量に入手できるようになったことと合わせて、機械翻訳の個人での利用が期待される。

実際、ここ数年間で機械翻訳に関する環境は大き く変わった。ワークステーション上のシステムが中 心であったものが、PC上のソフトが中心となって いる。これに伴って、ユーザも企業から個人へと広 がり、対象文書も技術文書から、様々な内容へと広 がっている。これらのシステムの特徴は低コストで ありながら、高性能であるということである。欧米 ではこのような安価なシステムは40以上もあり、大 量に販売されている。また、最近では音声出力と組 み合わされた機械翻訳システムも出現してきた。こ れらにより、現在では、それぞれのユーザが、思う がままのやり方で、様々な用途に機械翻訳を利用で きるようになった。これらのシステムは高級システ ムよりも有効に使われていく可能性がある。また、 単体のシステムとしてではなく、"component-ware" としての発展も期待される。

また別の意味での、機械翻訳とネットワークの関わりとしては、ネットワークを介しての機械翻訳の利用が考えられる。一つの利用法は、ネットワーク中の別の場所にある機械翻訳システムを利用するものである。ネットワーク対応のサーバ型の機械翻訳システムがいくつか発表されている。もう一つの利用法は、パソコン通信等を用いた機械翻訳サービスを利用するものである。

PC 化を中心とする欧米における最近の機械翻訳システムの動向については、7月にルクセンブルグで開催された第5回機械翻訳サミットでも報告された。この概要については、参考文献[1][2]を参照されたい。

3. 3. 機械翻訳システム品質評価用テストセット について

機械翻訳技術の発展に寄与するため、日本電子工 業振興協会では、開発者の立場からの機械翻訳シス テムの品質評価を行なうためのテストセットを開発し、公開した[3]。これは英日および日英機械翻訳システムの開発者が自己のシステムの不備を客観的に把握することを主たる目的として作成されたものであり、これまでの機械翻訳システム評価に関する多くの研究と、日本電子工業振興協会が開発してきた各種の機械翻訳システム評価基準の上に成り立つものである。

このテストセットは既存の機械翻訳システムを用いた評価実験によって改良を重ねて作られたものであり、開発者だけではなく、機械翻訳システムのユーザにも、有益な示唆を与えられるものと考えている。テストセットの例を図1に示す。

### 3. 4. その他の実用システムの可能性

機械翻訳システム等の自然言語処理システムの開発過程で得られた自然言語処理技術を用いて今後作られるであろう実用システムとしては、文書作成の支援システムや、テキスト処理システム、対話を伴うマンマシンインタフェース等が考えられる。

支援システムは、さらに校正支援システム、文書 作成支援システム、教育支援システム等に分けられ る。校正支援システムの構成要素としては、スペル チェッカー、文法チェッカー、スタイルチェッカー が挙げられる。スペルチェッカーは、概ね実用化さ れているが、文法チェッカーについては課題が残っ ている。スタイルチェッカーは、未だ実用的なもの は現れていない。校正支援システムの持つべき機能 としては、形態素解析で単語を認定し、解析に失敗 した区間や、同音異義語・異表記が存在する語など に警告を出すというものなどが考えられる。

校正時における誤り検出および訂正については、表記の問題(漢字表記と平仮名表記の関係等)、表現の問題(文法的誤り、語用論的誤り等)、内容の問題に分けられる。さらに誤り訂正においては、適切な修正候補の提示が要求される。また、読み上げ校正システムや、ワープロ上での編集ミスに対応する手法の開発が期待される。

文書作成支援システムは、今後、もっとも必要性 が高まるものと言えよう。文法事項、辞書情報、関 連文書/データ、テキストテンプレート等を提示し、 文書を生成する過程を支援する環境は、情報発信型 の機械翻訳を補完するものとして、また、より積極 的な発想支援環境として必要となろう。このような システムは、外国語生成支援として、また、外国語 教育支援としても、有効となると思われる。

なお、テキスト処理に関しては検索システムや解析・理解システムが考えられる。これについては、 本稿では関連要素の羅列にとどめる。なお、情報供 給者用システムと情報利用者用システムの違いは常 に認識しておくべきことがらである。

検索 フルテキスト検索 類似テキスト検索 検索条件の自然言語入力 検索条件の自動修正

> (検索結果のフィードバック) 検索エンジン・ロボット (エージェント) interest profile

解析 情報抽出 分類 要約 キーワード抽出

### 4. インターネットからの情報の選択的抽出 (知的ニュースリーダの提案)

現在、ネットニュースを通じて多くの情報が流通し、多くの人々がその情報を利用している。その情報量は1日約1000件約2MBであり[4]、日々増加の傾向にある。利用者の多様化に伴い、複数のニュースグループ間を移動する話題も少なくない。その中で、全部の情報を調査し利用することはかなり困難である。

このような情報をより効率良く利用するため、ダイジェスト機構の実現[5]や知的ニュースリーダの提案[6]もされ始めている。しかし、それらはまだまだ未成熟であり、多くの問題点を含んでいる。また、ユーザのニーズを考えた場合、ダイジェストで満足する場合は必ずしも多くはなく、興味を持った内容については、関連する記事そのものを読みたいと考えられる。

ここでは、ネットニュース上の情報利用の問題点 を洗い出し、それらの問題を解決するため、話題の 連続性に着目して必要な情報を抽出する方法を提案 する [7]。

ネットニュースに流れる情報を効率的に利用するには大きく分けて2つの方法が考えられる。

- 1) ネットニュース全体において情報検索をする。
- 2) ニュースグループにわけて情報検索をする。

ネットニュースを利用する立場から考えると1)が好ましいが、ネットニュースにはニュースグループによっていくつかの特徴があり[8]それぞれ情報の形態も異るため、リファレンスやサブジェクト情報だけからの単純な検索では、必要な記事全てのダイジェスト生成や記事検索は困難である。

最近のニュースリーダではサブジェクトによって 記事関係を分類しているが、サブジェクトが記事内 容からずれていることも多く、単純な検索方法では 不要な記事が多数抽出されることがある。

また、ネットニュースを利用しているユーザは、ネットニュース全体にどんな情報があるかを見るよりも先に、カテゴライズされているニュースグループの中に興味のある記事があるかどうかを確認することが多い。しかし、1日の投稿量からすると、ある1つのニュースグループにおいてですら読まなければならない記事が多量にある。

これらの問題を解決するために、記事をある種の 対話テキストとみなし、引用部分やリファレンス、 及び使用単語、文章の意味的距離 [9] を参考に記 事の話題の流れを文脈情報として遡り、必要な情報 (話題) の記事を検索する方法を提案する。

システムイメージとしては、しばらくネットニュースを読んでいなかったユーザが、今日届いた記事群を眺めていて、その中に興味のある話題を含む記事を見つけ、その話題に関するこれまでの記事を過不足なく抽出できるようなシステムである。システムの動作を以下に示す。

- (1) ユーザが記事群の中に興味のある記事を見つける。
- (2) ユーザが興味を持った記事の意味的特徴を記事中の利用単語と、概念辞書を用いて得る。
- (3) サブジェクト、リファレンス情報を用いて、 記事の関連ネットワークを成長させていく。
- (4) 関連ネットワーク中の枝分かれについて、記

事に関するヒューリスティックスと、元記事の意味的特徴を利用して、関連の深い記事と そうでないものとに分類する。関連の深い記 事の関連ネットワークを成長させていく。

- (5) 元記事の意味的特徴をキーとして、リファレンスで継っていない関連記事をも抽出する。
- (6)(5) で得られた記事についても (3)(4) を 行なう.

このシステムの特徴は、ユーザの興味に合わせて動的に処理を行なう点にある。また、記事同士の意味的距離をはかるので、ニュースグループを気にすることなく必要な情報を参照することが可能となる。検討課題としては、以下のようなものがあげられる。

- (1) 一般語と専門用語が記事の流れの中で果たす 役割についてのヒューリスティックスの検討。
- (2) 意味的特徴の動的な選択方法の開発。

### 5. おわりに

本稿では、これからの知的システムに要求される ものを情報の選択能力であると考え、自然言語処理 技術の当該能力への適用という点に着目して述べた。 今後も、日本電子工業振興協会では、自然言語処理 技術の実用的テキスト処理システムへの適用につい て、調査検討を進めていく予定である。

### 参考文献

- [1] 井佐原均他:"MT-Summit V 開催報告" AAMT Journal No. 12, 1995
- [2] 日本電子工業振興協会自然言語処理技術委員会:"MT-Summit V 開催報告"電子工業月報10月号,1995
- [3] 日本電子工業振興協会95-計-17: "機械翻訳システム評価基準-品質評価用テストセット-"1995
- [4] fj の歩き方編集委員会:"fj の歩き方"オーム社開発局, 1995
- [5] 佐藤円他: "電子ニュースにおけるダイジェスト機構の実現"情処第49回後期全国大会, 1994

- [6] 岩爪道昭他: "電子掲示版における記事の自 動分類と議論の可視化-知的ニュースリーダ の提案"人工知能学会全国大会、1994
- [7] 小作浩美他: "話題関連性に着目した知的ニ [9] 小嶋秀樹他: "意味空間のスケール変換によ ユースリーダの提案"平成7年電気関係学会 関西支部連合大会, 1995
- [8] 佐藤円他: "投稿行動から見た電子ニュース fj."第23回 jus UNIX シンポジウム論文集, pp 53-65. 日本 UNIX ユーザ会、1994
  - る動的シソーラスの実現"信学技報、NLC95 -19.1995

```
2. 1. 1 多品詞(品詞認定)
                         (= Part of Speech Disambiguation)
2. 1. 1. 2 名詞/助動詞 (= Noun/Auxiliary verb)
【番号】2.1.1.2-1 (= 【ID No.】)
       The trash can was thrown away. (= [Example])
【例文】
        ごみカンは捨てられた。 (= 【Translation】)
【訳文】
       "can" が「カン/缶」のように名詞として訳されていますか?
【質問】
        (= [Q.] Is "can" translated as a noun?)
【訳出例】○(くず缶/ごみ容器/くず入れ)は(廃棄された/[投げ]捨てられた)。
        ×ごみは捨てられ得る。
        (= [Translation Samples] literally meaning:
             ves: The (garbage can/trash bin/litter bin) was (discarded/[thrown] dumped).
             no: The trash can be discarded. )
【関連文】The last will was opened.「最後の遺言書は開けられた。」
        (= [Related Examples] and the Japanese translation)
【参照項目】2.1.1.2-2, 2.1.1.2-3 (= 【Reference Items】)
        "can was" の並びから、"can" が助動詞でないことがわかる。
        (= [Explanation] The word order of "can was" shows that "can" is not an auxiliary verb.)
【番号】
       2.1.1.2-2 (= [ID No.])
       The trash can be thrown away. (= [Example])
【例文】
        ごみは捨てられ得る。 (= 【Translation】)
【訳文】
        "can" が「~できる/得る」のように助動詞として訳されていますか?
【質問】
        (= [Q.] Is "can" translated as an auxiliary verb meaning "has ability
                                                             to/has a possibility to"?)
【訳出例】○ (くず/ごみ)は(廃棄できる/[投げ]捨てられることができる)。
        × ごみカンは捨てられた。
        (= [Translation Samples] literally meaning:
             ves: The (garbage/trash) (can be discarded/[thrown] dumped).
             no: The trash can was discarded. )
【関連文】(= 【Related Examples】)
【参照項目】2.1.1.2-1, 2.1.1.2-3 (= 【Reference Items】)
【解説】 2.1.1.2-1とは逆に、ここでは "can" は名詞ではなく助動詞。
        (= [Explanation] In contrast to No. 2. 1. 1. 2-1, here, "can" is not a noun but
                                                                   an auxiliary verb.)
```

図1 品質評価用テストセット

# 人間の翻訳と機械翻訳

NEC 情報メディア研究所 亀井真一郎

### 1 はじめに

この「技術早分かり」シリーズでは、本協会が日本機械翻訳協会だった時代のJAMTジャーナル No. 2から AAMTジャーナル No. 12まで約4年間で、機械翻訳の基本となる各要素技術の解説を一通り完了しました。解説は、形態素解析技術、構文解析技術、文生成技術に始まり、辞書、機械翻訳の利用法、関連技術(音声認識技術、構成支援技術、OCR技術)までを含んでいます。

今回からはこれまでとは少し視点を変えて、人間の翻訳と機械翻訳を比較する観点から新たな技術解説を始めます。機械翻訳には「辞書」が重要な役割を果たしますが、機械翻訳用の辞書と通常私たちが使う学習用の辞書との違いはどこにあるのでしょうか。また機械翻訳の分野で用いられる重要な用語に「文法」がありますが、機械翻訳で使われる「文法」と私たちが学校で習う日本語や英語の文法はどこが似ていてどのような違いがあるのでしょうか。このような機械翻訳用の辞書と一般の辞書の違い、機械翻訳で使われる文法と学校文法の違い、さらに機械による翻訳と人間の翻訳の関係などの比較を通して、機械翻訳の要素技術を紹介してゆきます。

### 2 一般の紙の辞書と機械翻訳辞書

まず個々の単語や句に関して記述した「辞書」について考えてみましょう。人間が翻訳する際に辞書が必須のものであるのと全く同様に、機械翻訳にとって辞書は最も基本的な言語知識ベースです。

### 2.1 英和辞典と英日機械翻訳辞書

翻訳に使う辞書としてまず私たちがイメージするのは英和辞典です。そこでまず通常私たちが使う英和辞典と英日機械翻訳システムで使われる辞書の内容を比較してみます。英和辞典には次のような項目が書かれています。

- ・英単語の表記、発音、品詞、活用、構文
- ・日本語の訳語

- ・例文 (英語と日本語の対訳例文) 例えば [eat] という英語の動詞を例に取ると、
  - · 英単語の表記:eat 発音:[i:t] 品詞:V 活用:eat ate eaten eating 構文:S V O
  - ・日本語の訳語:食べる
  - · 例文 I eat bread 私はパンを食べる

といった具合です。英和辞典は英語の使い方を説明 したものですから、英単語については詳しい記述が あるのですが、日本語の訳語としては単に対応する 単語が書いてあるだけであまり詳しくないことがわ かります。日本人は日本語の知識が頭の中にはいっ ているので訳語の日本語について詳しい説明は必要 ありませんが、機械が翻訳する場合には、英語につ いての記述と同様な詳しさで日本語についても明示 的な記述が必要です。そこで通常、機械翻訳の辞書 には以下のような項目が書かれています。

- 英単語の表記、発音、品詞、活用、構文
- ・日本語の訳語の表記、品詞、活用、構文
- 英語と日本語の対応

例えば「eat」の場合には次のようになります。

- ・英単語の表記:eat 発音:[i:t] 品詞:V 活用:eat ate eaten eating 構文:S V O
- ・日本語の訳語の表記:食べる 品詞:動詞 活用:一段活用 構文: (誰か)が(何か)を.
- 英語と日本語の対応
  - S(主語) = 「が | のついた名詞句
  - O(目的語)=「を」のついた名詞句

ここで英語の構文というのはいわゆる SVO、SVOC といった文型を詳細化したものです。 SVOC の C は補語と呼ばれますが、その種類はたくさんあります。単純な形容詞や名詞のこともありますし、「in」や「with」などの前置詞のついた句の場合もありますし、that 節のような文相当の節の場合もあります。そこでそれらを詳細化した文型を指定する訳です。日本語に対しても同様な文型の情報が必要です。例えば「食べる」という動詞は「(誰か)が(何か)を食べる」のように助詞「が」「を」のついた名詞句を

従えます。一方「住む」という動詞では「(誰か)が (どこか)に住む」のように助詞「が」「に」のついた名詞句を必要とします。このような情報が構文情報です。英語の「I eat bread」を日本語の「私がパンを食べる」に翻訳するにはこの構文情報の英語と日本語の間の対応を使います。「eat」は SVO という構文をとりますが、その S (主語)は「食べる」の「が」のつく名詞句に対応します。また O (目的語)は「を」のつく名詞句に対応します。このような英語と日本語の対応を辞書に明示的に記述しておくことで、適切な翻訳が行えるのです。

「I live in Tokyo」を「私は東京に住んでいる」と訳すためにはさらに別の情報が必要です。「live」は「住む」に対応しますが、翻訳するときには「住んでいる」のように「~している」の形にする方が自然です。通常日本語の「~している」の形は英語の進行形(be + ing 形)に対応することが多いのですが、「live」の場合にはそのままの形で「~している」の形に対応する訳です。このような情報は英和辞典には明示的に書かれていないことが多いのですが、機械翻訳用の辞書にはすべての動詞に対してこのような情報が記述されています。

### 2.2 和英辞典と日英機械翻訳辞書

英語を書くときに参考にするのは和英辞典ですが、 残念ながら和英辞典だけでは使いづらいと感じてい る方は多いと思います。ここでその理由を考えてみ ます。和英辞典には通常次のような項目が記述され ています。

- ・日本語の表記
- ・ 英語の訳語
- ・例文(日本語と英語の対訳例文)

例えば「食べる」を例にとると次のようになります。

- ・日本語の表記:たべる(食べる)
- · 英語の訳語:eat
- ・例文:私はパンを食べる I eat bread

特に問題なのは英語の訳語の部分です。和英辞典の英訳語としてはふつう単語が書かれているだけで、その単語の活用や構文は書かれていません。そこである一つの訳語(例えば「eat」)を使おうと思っても、その単語の活用形がわからなかったり、動詞の場合その動詞の構文が SVO だったか SVC だったか分からなかったり、SVC の場合どのような前置詞

をとればよいのかが分からなかったり、といったことがおこります。そこで和英辞典の次に英和辞典を引いてその単語の使い方を確認することが必要になり手間がかかってしまうのです。

次に機械翻訳のための辞書を考えてみます。機械にとっては日本語も英語もいわば外国語ですから、機械翻訳で使う辞書には、日本語に関しても英語に関しても一つ一つの単語の詳細な情報を記述する必要があります。日英機械翻訳用の辞書の場合には次のような項目が必要です。ちょうど英日辞書と同等な情報が書かれていることがわかります。

- ・日本語の表記、品詞、活用、構文
- ・英語の訳語の表記、品詞、発音、活用、構文
- 日本語と英語の対応

動詞「食べる」の場合には次のようになります。

- ・日本語の表記:たべる、食べる 品詞:動詞 活用:一段活用 構文: (誰か)が(何か)を...
- ・英語の訳語の表記:eat 品詞:V 活用:eat ate eaten eating 構文:S V O
- ・日本語と英語の対応 「が」のついた名詞句=S(主語) 「を」のついた名詞句=O(目的語)

このように機械翻訳のための辞書には通常の紙の辞書に明示的に書かれていないたくさんの情報が一つ一つの単語に対して書かれています。機械翻訳を使う時に基本辞書にない単語は利用者がユーザ辞書に登録しなければなりませんが、その登録は思ったより複雑な作業になってしまいます。その理由はここで述べた点にあるのです。

### 2. 3 表現辞典

私たちが英語を勉強する時、まず単語を覚え、次にだんだんとフレーズの使い方を覚えてゆきます。例えば「電話」を英語では「telephone」といいますが、「電話をかける」という表現は英語で何というのでしょうか。日本語の「電話」と「かける」をそのまま合成しても適切な英語の表現「make a call」は得られません。単語の使い方を一通り学習した後は、このような個々の単語の組み合わせでは得られない表現を集めた表現辞典が必要となるわけです。

機械翻訳の場合も同様です。機械翻訳の開発では まず単語の辞書の整備が行われました。そして現在 の機械翻訳システムでは何万語もの単語を基本辞書 の中にもっています。しかし単語の辞書だけ何十万 語に増やしても適切な翻訳は得られません。そこで 表現辞書の充実が現在も継続されています。

### 3 学校で習う文法と機械翻訳用文法

次に言葉の使い方を規則の形にまとめた「文法」 について考えます。文法の基本は品詞の体系なので 特に品詞を中心に、学校で習うものと機械翻訳で必 要なものを比較してみます。

### 3.1 英語の文法

英語の文法は詳細な分析研究が進んでいるので、機械翻訳でもその成果を基本にしています。学校で習う(従って通常の辞書や文法書に載っている)英語文法やそこで使われる品詞は分かりやすさのために簡略化してありますが、大部分は機械翻訳用にもそのまま使えます。しかしより高い精度で翻訳を行うには品詞の細分類が必要となります。例えば「副詞」は細分類が必要な品詞の典型例です。通常習う英文法では「very」も「quickly」も「therefore」も副詞と呼ばれ、辞書にも通常そう書かれていますが、これらの語の使われ方はそれぞれ全く異なります。

「very」は形容詞や、他の活用する副詞を強めるための語で「強調詞」と呼ばれます。「quickly」は動詞を修飾してその様子を表す語で「様態の副詞」と呼ばれます。「therefore」は前の文と次の文をつなぐ役目をする「接続の副詞」です。これらの語は文のどこの位置に置かれるかも異なっています。機械翻訳のための辞書にはこれらの品詞を詳細に分類し、それぞれの使われ方を一つ一つ記述してあります。機械翻訳システムはこのような詳細情報を使って英文の解釈・生成を行っているのです。

### 3.2 日本語の文法

日本語の文法はまだまだ分析を進めなければならない点が多く残っています。実は基本となる品詞に関してもまだ標準となる体系がないと言って良いと思います。中学校で習う日本語の文法では次の10の品詞を設定しています。

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

しかしこれだけの品詞では日本語の分析にとって不 十分であることはすぐにわかります。例えば「ここ 数カ月間」という表現は「ここ」「数」「ヶ月」「間」の 4つに分けられると思いますが、それぞれの単語の 働きを適切に表せる品詞は上記の10品詞の中にはあ りません。そこで上記の10の品詞の他に「接頭辞」「接 尾辞」「助数詞」等という品詞が設定されるのですが、 それらの新しい品詞にしても、そこに分類される語 は一様ではなく、さらに細分化が必要です。例に挙 げた「ここ」「数」「ヶ月」「間」を何という品詞として 扱うかに関してはまだ標準的な文法ができあがって いないのが現状です。

機械翻訳システムを開発するにあたって準拠すべき標準的な品詞の体系がありませんでしたので、機械翻訳の研究者たち自身が言葉の使われ方を分析し、品詞を定め、辞書を作成し、文法規則を記述するというプロセスを各システムごとに行なってきました。したがって現状では品詞の詳細は各システムごとに異なっています。今後は日本語の研究者や機械翻訳の研究者などが協力して正確でかつ分かりやすい日本語の品詞の標準化を行なってゆく必要があります。

### 4 人間の翻訳と機械翻訳

人間が翻訳をする場合と機械翻訳システムが翻訳 をする場合とでは、その方法も、そこで利用する知 識も異なります。人間の翻訳と機械翻訳の似ている 点と異なる点について考えてみます。

### 4.1 日英翻訳と英日翻訳

日本人にとっては、日本語を英語に翻訳する日英 翻訳の方が、英語を日本語に翻訳する英日翻訳より も困難です。日英翻訳の場合、結果の英語が日本人 にとっての母国語でないために翻訳結果の正確さを 保証することが難しいからです。

機械翻訳の場合にも、理由は少し異なりますが、 人間の場合と同様に、英日翻訳より日英翻訳の方が 困難です。日英翻訳の場合には、まず元の日本語の 文の構造を把握する(これを「日本語を解析する」 という)必要がありますが、日本語の解析にはいろ いろな困難な点があるのです。

まず日本語の表記の点が挙げられます。英語は各 単語をスペースで区切って書くので単語の認定に困 ることはありませんが、日本語の場合には単語の間 に区切れ目がありません。従って日本語の文を解釈 するにはまずその文を各単語に分ける(これを形態 素解析という)必要がありますが、この段階ですでに正しく単語分割が行えないことがあるのです。

単語分割の次には単語と単語の間の関係、つまり 文の構造を把握する(これを構文解析という)訳で すが、この段階でも日本語には難しい点があります。 日本語では文脈から分かる主語や目的語が省略され てしまう性質があるからです。人間にとっては明か であっても、機械にとって、省略されたものを回復 して解釈するのは容易ではありません。また前にも 述べた通り、日本語は基本的な文法に関してまだ整 備が進んでいないために元の日本語の解釈が正しく 行えないことも多いのです。

さらに次の段階、つまり日本語を英語に対応させる段階での困難な点として、冠詞や名詞の単数・複数の問題があります。日本語には冠詞も名詞の単数・複数の区別もないのですが、英語ではそれらを適切に区別しなければなりません。このような区別は人間の学習者にとっても難しく、まだきれいに規則化できていないので、現状の機械翻訳システムではこの種類の現象を正確に扱うことは困難なのです。

さらに日英翻訳システムを日本人が使う場合、システムが出力する英語が正しいのか誤りなのか判断 しにくいという点も日英機械翻訳システムを使うことを困難にしている一因です。

英日機械翻訳の場合は、日英機械翻訳の場合とちょうど逆になります。元の英語は単語に区切られていますし、主語や目的語の省略もありません。出力される結果の日本語も日本人にとっての母国語なので、多少訳語や語順がおかしくても、システムの誤りの癖を知っていれば何が書かれているのかを想像できることも多いのです。つまり機械翻訳システムに一括処理させて、文の大意を把握するという使い方が有り得るのです。現状で市販されて実際に使用されている機械翻訳システムとして英日システムが圧倒的に多いのはこのような理由からです。

### 4.2 文脈の活用

人間が文章を理解する時にはごく自然に文脈を利用しています。おなじ文面であっても前に書かれている内容によって解釈をごく自然に変えて読みとります。しかし機械で「文脈」を正確に扱うことができるようにするにはまだまだ研究が必要です。「文脈」の情報というのは直接単語や句として表現され

ているものばかりではなく、その場面についてあらかじめもっているべき知識や一般的な常識などが総合的に組み合わされたものであるからです。文脈すなわち文の流れがうまく扱えない現状の機械翻訳システムでは、通常、文章全体を一連のものとして翻訳するのではなく、文を一つ一つバラバラにして扱っています。したがって、文の前後関係によって翻訳しわけるといったことは実現できていません。文脈の研究とそれを利用して自然で適切に翻訳を行なう技術はまさに今後の重要な研究課題の一つです。

### 5 おわりに

近年は社会全体が国際化し、また世界規模の通信ネットワークが急速に発達してきており、一般の人々も外国語の情報に接したり外国語で文書を書いたりする機会が急速に増えています。機械による翻訳の技術あるいは翻訳支援の技術は今後ますます重要になるでしょう。

機械翻訳システムの研究・開発の歴史を振り返ってみると、機械翻訳システムが登場した当初はシステムに対する過大な期待がありました。人間の翻訳者が失業してしまうのではないかという心配まで聞かれました。しかし機械翻訳システムの実力が明らかになるにつれて、それまでの過大な期待に対する反動として機械翻訳に対する不当な非難の時代がやってきました。機械翻訳なんて全く使えないという訳です。しかしそのような過大な期待と不当な非難の時代を経て、最近ようやく機械翻訳が人間の翻訳を支援する技術であると正当に捉えられるようになってきました。

意味や文脈を正確に捉えて翻訳が行なえるシステムを作ることは機械翻訳の研究の究極の目標ですが、それがまだ完全には実現されていない現在は、翻訳する時すべてを機械まかせにすることはできません。しかし人間の得意なところは人間が担当し、機械の得意なところは機械が受け持つという使い方をすることで機械翻訳の技術を生かすことができます。

これからは、研究者が翻訳品質の向上を目指して 研究を継続するだけでなく、研究者と利用者とが一 緒になって、機械翻訳を使いこなすための利用技術 を研究・開発してゆくことがますます重要です。

# 東京工業大学大学院情報理工学研究科田中·徳永研究室

東京工業大学大学院情報理工学研究科教授 田中穂積

この春で研究室発足から早12年になる。当初学生が4名だった研究室も、現在では教授、助教授、助手各1名と学生21名を数える大所帯になり、幅広い興味をもって研究を進めることができるようになった。本稿では、自然言語処理に関連する現在の研究テーマをいくつか紹介する。

### 1. GLR パーザによる音声認識・形態素解析・構 文解析の統合

自然言語解析では、従来〈音声認識〉→〈形態素解 析>→<構文解析>→…のように段階的に進めてい た処理をどのように統合するかが重要な課題である。 当研究室では、この課題へのアプローチとして、音 素接続制約・形態素接続制約・構文制約を共通の知 識表現形式に変換し、音声認識から構文解析までを 統合的に実現する研究を進めている。この研究で 我々が着目しているのは一般化 LR 法で用いる LR 表という言語知識の表現形式である。我々はすでに, LR 表を利用すれば上の3つのレベルの制約を単一 の知識表現形式にコンパイルすることができること を明らかにしている。これにより一般化 LR 法とい う単一の枠組で音声認識から構文解析までを統合的 に実現することができる。この技術を適用した音声 認識の実験からすでにその有効性が確認されている。 さらに、我々はLR表に確率情報を組み込む方法も 開発しており、これにより音韻・形態素レベルの bigram や確率文法の統合的な利用が可能になった。 現在, この方式のツール化を進めており, これが広 く自然言語処理システムの研究・開発に利用される ことを期待している。

### 2. コーパスを利用した言語処理

自然言語処理技術の実用化に向けてクリアする必要 のある重要なハードルの一つが知識の大規模化であ る。大規模コーパスの利用はこの目的の達成に不可 欠なアプローチの一つと言えよう。当研究室では,その試みの一つとして,括弧付きコーパスから適用範囲の広い確率文脈自由文法を自動的に抽出する研究を進めている。これは,括弧で表現された構文木の中間ノードに非終端記号を自動的に割り当てる非終端記号を自動のである。内部ノードに割り当てる非終端記号は,主辞がつねに句の最後に位置する日本語の特徴を利用して自動的に決定する。また,抽出したりをは,主辞がつねに句の最後に位置する日本語の特徴を利用して自動的に決定する。また,抽出したりを追いる。これまでの実験から,冗長な規則を削除することによって文法サイズを縮小したり,品詞の細分化によって解析結果のあいまい性を抑制したりすることがある程度可能なことが確かめられている。

コーパスからは構文的制約の他にも多種多様な言語 知識を抽出できるだろう。その一つに語の共起に関 する制約があげられる。当研究室では、これに関す る事例研究として複合名詞解析の研究を進めている。 複合名詞という構造は造語力が極めて高いため、全 ての複合名詞をあらかじめ辞書に登録しておくこと は不可能である。そこで我々は、個々の名詞間の意 味レベルでの共起情報をコーパスから抽出し、これ と既存の辞書を併用することによって複合名詞を解 析する方式を開発した。この方式では、3つ以上の 語からなる複合名詞内の係り受け構造を各語の意味 クラスの共起頻度にもとづいて決定し, さらに名詞 間の意味的関係を各語の統語的・意味的特徴にもと づいて推定する。今後は、さらに詳細な意味的関係 の推定に挑戦するとともに、この技術を他の構文要 素の解析に応用する研究を進めたいと考えている。

また,事例ベースの動詞多義性解消の研究も手がけている。事例ベースの自然言語処理では,事例が十分に集まらない場合の対処と,大量の事例の獲得に要する人的コストの抑制が大きな課題である。前者の課題については,格要素によって動詞語義の弁別

に貢献する度合が異なることに着目し、これを解析 に反映させる方式を提案している。これまでのとこ ろ、この方式を用いると事例が少ない場合でもある 程度正解率を維持できるという実験結果を得ている。 事例獲得のコスト抑制に関しては、解析システムを 用いて半自動的に事例を獲得する際に、解析結果の 信頼性を定量化し、信頼性の低い事例のみ人間が修 正するという方式を考え、現在これについて実験を 進めている。

### 3. 対話における談話プランニング

対話では、多様な方向に遷移する話題を双方の対話者が正確に認識できなければならない。そのためには、相手の発話から話題を推定するための談話解析技術とともに、話題の遷移を相手に正確に伝達するための談話生成技術が必要である。当研究室では、そのための足がかりとして後者の技術に焦点をあて、談話マーカやメタコメントのような定型的な発話に加え、過去の談話を要約するような非定型な発話を生成する技術の研究を進めている。誤解や不理解の修復を目的とする副対話の構造と話題の遷移の伝達を目的とする発話の関係に着目し、対話コーパスの分析と試作システムでの実験による生成モデルの構築・洗練をめざしている。

### 4. 知的情報アクセス

電子化文書が急速に普及するにつれて、文書データベースから必要な情報を検索するサービスの需要は増大する一方である。当研究室では、知的情報検索実現のための第一歩として文書クラスタリングおよび文書自動分類に関する基礎研究を進めている。一般に文書間の類似性は単語の出現頻度などから得られる文書特徴ベクトルから計算されるが、我々はこの類似度の計算に理論的妥当性を与える確率モデルを提案しており、これを新聞記事のクラスタリングに応用する実験によってその有効性を確かめている。現在、このモデルの実用化をめざし、10万件オーダの文書を実用的な速度でクラスタリングする方式の開発に挑戦している。また、同じモデルを文書の自動分類や単語のクラスタリング(すなわちシソーラスの自動構築)に応用する研究も進めている。

情報検索の研究と並行して情報抽出の研究も進めている。情報抽出の研究に関しては米国の MUC が有名である。MUC のアプローチはあらかじめ決められたドメインのテンプレートを用意しておき,入力テキストを解析してそのテンプレートを埋める情報を抽出するというものである。我々の研究室では、テンプレートをあらかじめ人手で用意するのではなく,特定の分野の新聞記事の集合から半自動的にテンプレートを生成する研究をおこなっている。

### 5. 最近発表した博士論文

小林義行、コーパスを用いた複合名詞の解析、1996.

Hui Li, Integrating Connection Constraints into a GLR Parser and its Applications in a Continuous Speech Recognition System, 1996.

秋葉友良,音声対話システムにおける言語処理とユーザ・モデルに関する研究,1995.

乾健太郎,自然言語生成における相互依存的制約の 扱いに関する研究,1995.

Thanaruk Theeramunkong, A Study on Parallel Natural Language Processing, 1995.

堀内靖雄, 人間と協調する演奏システムに関する基 礎的研究, 1995.

伊藤克亘,連続音声認識システムに関する研究,1993.

徳永健伸,自然言語処理のための統語的・意味的知識に関する基礎研究,1991.

岩山真, 計算機による比喩理解モデルに関する研究, 1991.

沼崎浩明,並列計算モデルを用いた自然言語処理の 高速化に関する研究,1991.

### 関連記事紹介

### ATA(アメリカ翻訳者協会)の状況

ATA (アメリカ翻訳者協会、会長 E. Losa 女史) の会報によると、加入会員数は、5,719人 (95年9月現在) である。会員の中には通訳者も含まれている。

会員の身分はパートタイム契約者 34%、フルタイム契約者 32%、フルタイムプライベートセクター従事者 12%、事業主 9%、学者 8%、公務員 5%である。

専門分野ではビジネス 17%、美術人文学 12%、工業技術 11%、法律 10%、コンピュータ 10%、薬学 10%、 社会科学 7%、科学 7%、工学 7%の順となっている。

また、会員の母国語としては、英語 39%、スペイン語 24%、ドイツ語 10%、フランス語 8%、ロシア語 6%、日本語 4%、イタリア語 3%、ポーランド語 2%、アラビア語 1%の割合となっている。多民族国家であり、英語を公用語としていない米国の現状を物語っている。

(日本翻訳ジャーナル 11/1995から転載)

### イベント紹介

# Eighth International Unicode/ISO 10646 Conference

1991年に設立され、Unicode Standard の推進に貢献してきた"The Unicode Consortium"が、第8回国際会議を4月18~19日にホンコンで開催する。

Date April 18&19, 1996

Location

Hotel Furama Kempinski Hong Kong

One Connaught Road Central

Hong Kong

Tel: (852) 2525-5111 Fax: (852) 2845-9339

Conference Fees

Unicode members: US \$ 650,00 to include all (employees of)

corporate, associate and individual members

Non-menbers:

Prior to March15: US \$ 695.00

After March15: US \$ 750.00

to include all conference materials; proceedings; continental breakfast, refreshment breaks and luncheons for both days; cocktail reception and entry to the exhibit

area on Day 1 and a conference T-shirt.

Program: P14~15

The Unicode Consortium

P. O. Box 700519

San Jose, CA

95170 - 0519

Tel: (408) 777 - 3721

Internet:

unicode-inc@unicode. org

This Conference is managed by:

Global Meeting Services Inc.

3627 Princess Avenue

N, Vancouver, BC

Canada V7N 2E4.

Telephone: (604) 983-9157

Fax: (604) 983-9158

Internet: conference@unicode. org

http//unicode. org

# PROGRAM

# THURSDAY, APRIL 18th

	NTINENTAL BREAKFAST			
9:00 - 9:10 a.m. OPENIN	OPENING ADDRESS			
Lisa Moore, Chair, Confe	erence Editorial Review Board			
9:10 - 9:45 a.m. KEYNOTE I	PRESENTATION			
7 A 1 A 1 A 1 A 1 A 1 A 1 A 1 A 1 A 1 A	chiro Kida			
	nd Literature, Chairman, TOK Advisory Board			
9:45 - 10:20 a.m. PL	ENARY			
UNICODE	CHALLENGES			
Asmus Freytag, Preside	nt, Asmus, Inc., Seattle, WA.			
10:20 - 10:40 a.m. REFRESH	MENT BREAK			
WEB 10:40 - 11:15 a.m.	TUTORIAL			
<b>A1</b> (E/SD/P)	B1 (M/E/SD/SA/P/MK)			
USING UNICODE IN WEB APPLICATIONS	HOW TO DESIGN NON-LATIN PROBLEMS INTO YOUR			
Ken Fowles, Program Manager, Internet Explorer Group,	PRODUCT, PART I.			
Microsoft Corporation, Redmond, WA.	Richard Ishida, Global Design Consultant, Rank Xerox Technical Centre, Herts, UK			
11:20 - 12:00 Noon				
A2 (M/E/SD/SA/P/MK)	B2 (M/E/SD/SA/P/M)			
AUTHORING AND BROWSING THE WORLD WIDE WEB	HOW TO DESIGN NON-LATIN PROBLEMS INTO YOUR			
IN UNICODE	PRODUCT, PART II.			
Robert N. Goldfich, Accent Software International, Israel.	Richard Ishida, Global Design Consultant, Rank Xerox Technical Centre, Herts, UK			
12:00 - 1:30 p.m.	NCHEON			
43 (E/SD/P/M)	B3 (M/E/SD/SA/P/MK)			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  lim Welch, Director, Emerging Technologies, Bitstream Inc.,	THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 1			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc.,	THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Jim Weich, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA. 2:10 - 2:45 p.m.	THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 1 Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
43 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  lim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  44 (M/E/SD/P/M)	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1 Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK)			
43 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Ilim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  44 (M/E/SD/P/M)  UNICODE IN VISUAL COMPUTING.	THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 1 Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc.,	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1 Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK)			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc.,	THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK)  THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 2			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc.,  Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/P/M)  UNICODE IN VISUAL COMPUTING.  Gianni Mariani, Manager of Globalization R&D Lining Sun,  MTS and Michael Kung, MTS, Silicon Graphics, Inc.,  Mountain View, CA.	THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK)  THE UNICODE STANDARD, VERSION 2.0 : A TUTORIAL PART 2			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Jim Weich, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m. A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
43 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Ilim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  44 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2 Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
43 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Ilim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m. 44 (N/E/SD/P/M) JUNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  8:05 - 3:40 p.m.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Itim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/P/M) INICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m. A New Unicode based WWW Gateway for Browsing Mutt-lingual Information in Internet.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Jim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m. A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m. A5 (M/E/SD/SA/P/MK) A New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet. Ricky, Hung Tak Chan, Assistant Computer Officer, City	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc  Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m.  A5 (M/E/SD/SA/P/MK) A New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet.  Ricky, Hung Tak Chan, Assistant Computer Officer, City University of Hong Kong, Hong Kong.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc.,  Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/P/M)  UNICODE IN VISUAL COMPUTING.  Gianni Mariani, Manager of Globalization R&D Lining Sun,  MTS and Michael Kung, MTS, Silicon Graphics, Inc.,  Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m.  A5 (M/E/SD/SA/P/MK)  A New Unicode based WWW Gateway for Browsing  Multi-lingual Information in Internet.  Ricky, Hung Tak Chan, Assistant Computer Officer, City  University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3  Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
43 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Ilim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  44 (M/E/SD/P/M)  UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m.  45 (M/E/SD/SA/P/MK)  4 New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet. Ricky, Hung Tak Chan, Assistant Computer Officer, City University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m.  46 (M/E/SD/SA/P/M)	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3			
43 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Ilim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m.  44 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m. 45 (M/E/SD/SA/P/MK) A New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet. Ricky, Hung Tak Chan, Assistant Computer Officer, City University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m.  46 (M/E/SD/SA/P/M) CREATING A MULTILINGUAL UNICODE WWW	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3  Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT.  Jim Welch, Director, Emerging Technologies, Bitstream Inc.,  Cambridge, MA.  2:10 - 2:45 p.m.  A4 (M/E/SD/SA/P/M)  WINICODE IN VISUAL COMPUTING.  Gianni Mariani, Manager of Globalization R&D Lining Sun,  MTS and Michael Kung, MTS, Silicon Graphics, Inc.,  Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m.  A5 (M/E/SD/SA/P/MK)  A New Unicode based WWW Gateway for Browsing  Multi-lingual Information in Internet.  Ricky, Hung Tak Chan, Assistant Computer Officer, City  University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m.  A6 (M/E/SD/SA/P/M)  CREATING A MULTILINGUAL UNICODE WWW  HOMEPAGE.  Chong Chia Jen, Software Engineer, ISS, National University	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B6 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Jim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m. A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m. A5 (M/E/SD/SA/P/MK) A New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet. Ricky, Hung Tak Chan, Assistant Computer Officer, City University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m. A6 (M/E/SD/SA/P/M) CREATING A MULTILINGUAL UNICODE WWW HOMEPAGE. Chong Chia Jen, Software Engineer, ISS, National University of Singapore, Singapore.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B6 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 4  Asmus Freytag, President, Asmus, Inc., Seattle, WA.			
3:05 - 3:40 p.m.  A5 (M/E/SD/SA/P/MK) A New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet.  Ricky, Hung Tak Chan, Assistant Computer Officer, City University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m.  A6 (M/E/SD/SA/P/M)  CREATING A MULTILINGUAL UNICODE WWW HOMEPAGE.  Chong Chia Jen, Software Engineer, ISS, National University of Singapore, Singapore.  4:25 - 5:00 p.m.	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  WENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B6 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 4			
A3 (E/SD/P/M) PLATFORM INDEPENDENT FONT SUPPORT. Ilim Welch, Director, Emerging Technologies, Bitstream Inc., Cambridge, MA.  2:10 - 2:45 p.m. A4 (M/E/SD/P/M) UNICODE IN VISUAL COMPUTING. Gianni Mariani, Manager of Globalization R&D Lining Sun, MTS and Michael Kung, MTS, Silicon Graphics, Inc., Mountain View, CA.  2:45 - 3:05 p.m.  REFRESH  3:05 - 3:40 p.m. A5 (M/E/SD/SA/P/MK) A New Unicode based WWW Gateway for Browsing Multi-lingual Information in Internet. Ricky, Hung Tak Chan, Assistant Computer Officer, City University of Hong Kong, Hong Kong.  3:45 - 4:20 p.m. A6 (M/E/SD/SA/P/M) CREATING A MULTILINGUAL UNICODE WWW HOMEPAGE. Chong Chia Jen, Software Engineer, ISS, National University of Singapore, Singapore.  **EYNOTE I: THE IMPORTANCE OF MULTI- BUSINESS IN ASIA - AN INFOR	THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 1  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B4 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 2  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  MENT BREAK  B5 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 3  Asmus Freytag, President, Asmus, Inc., Seattle, WA.  B6 (M/E/SD/SA/P/MK) THE UNICODE STANDARD, VERSION 2.0: A TUTORIAL PART 4  Asmus Freytag, President, Asmus, Inc., Seattle, WA.			

## FRIDAY, APRIL 19th

IDEOGRAPHS 9:00 - 9:35 a.m.	GLOBAL SOFTWARE	COMPRESSION/CONVERSION
A1 (M/E/SD)	<b>B1</b> (E/SD/P)	C1 (M/E/SD/SA/P/MK)
ISO/IEC 10646 (UNICODE) AND KANJI DATABASE.	LOCALIZATION IN THE UNET++ APPLICATION	COMPRESSING UNICODE.
Eiji Matsuoka, Associate Professor, Tokyo Gakugei	FRAMEWORK.	Misha Wolf, Globalisation Manager, Reuters, London,
University, Tokyo, Japan.	Martin Dürst, Senior Research Associate, University of Zurich, Zurich, Switzerland	England.
9:40 - 10:15 a.m.		
A2 (M/E/SD/SA/P/MK)	<b>B2</b> (M/E/SA)	C2 (E/SD/P)
CULTURAL AND ENGINEERING ISSUES IN DEVELOPING CJK FONTS	LISA, THE LOCALIZATION INDUSTRY STANDARD	MAPPING BETWEEN UNICODE AND OTHER
Yoshihiko Sawada, Director/Vice President, DynaLab Japan	ASSOCIATION.	CHARACTER SETS: METHODS AND ISSUES. Lloyd Honomichl, Consultant Engineer, Novell, Inc.,
Ltd., Tokyo, Japan.	Andreas Schütz, Manager Product Internationalization, Software AG, Darmstadt, Germany.	Provo, UT.
10:15 - 10:35 a.m.	REFRESHMENT BREAK	
10:35 - 11:10 a.m. A3 (M/E/P)	<b>B3</b> (M)	C3 (E/SD/SA/P)
THE FIRST UNICODE CJK CHARACTER DICTIONARY.	CONVERSION OF A MULTINATIONAL SOFTWARE	UNICODE CONVERSIONS TO/FROM IBM CODE PAGE
Martin Dürst, Senior Research Associate, University of	COMPANY TO UNICODE - A CASE HISTORY TO-DATE.	AND CCSIDs.
Zurich, Zurich, Switzerland	Michael McKenna, Globalization Strategist, Sybase, Inc., Emeryville, CA.	Hossein Kushki, Advisory Engineer, NLTC, IBM Canad Laboratory ON, Canada.
11:15 - 11:50 a.m.		
44 (M/SD/P/MK)	B4 (M/E/SD/SA/P/MK)	
OVERVIEW OF THE WINDOWS NT EXTENDED KANJI	PRINTING IN FOUR DIRECTIONS.	
PROCESSING SPECIFICATION.	John Plaice, Université Laval, Quebec, Canada, and Yannis	
Shigeru Nagahisa, Group Manager, Microsoft Co., Ltd.,	Haralambous, Lille, France.	
Tokyo, Japan.		
L2:00 - 1:30 p.m.	LUNCHEON	
DESKTOP	UNIX	DATABASE
1:30 - 2:05 p.m. 45 (E/SD/SA/P)	B5 (M/E/SD/MK)	C5 (M/SD/SA/MK)
NTERNATIONALIZATION ON COPLAND.	DESIGN CONSIDERATIONS FOR MULTILINGUAL	UNICODE IN DISTRIBUTED COMPUTING
ohn McConnell, International Evengelist, Apple Computer,	APPLICATIONS	ENVIRONMENTS.
Cupertino, CA.		Michael McKenna, Globalization Strategist, Sybase,
		Emeryville, CA.
<b>2:10 - 2:45 p.m. 46</b> (E/SD/SA/P)	<b>B6</b> (E/SD/P)	C6 (M/E/SD/SA/P/MK)
NEXT GENERATION INPUT METHOD MANAGER.	MULTILANGUAGE INPUT MODEL ON THE X-WINDOW	THE EUROPEAN COMMUNITY AND UNICODE/ISO
lames Saunders, Scientist, Apple Computer, Cupertino, CA.	SYSTEM.	10646.
The state of the s	Shinobu Matsuzuka, Member of Technical Staff, Sun Soft,	Anthony Brickell, System Development Manager. The
	Mountain View, CA.	British Library, London, UK.
2:45 - 3:05 p.m.	REFRESHMENT BREAK	
3:05 - 3:40 p.m.		
47 (E/SD/SA/P)	<b>B7</b> (E/SD/P)	C7 (M/E/SA)
UNICODE ON COPLAND,	UNICODE AND INTERNALIZATION WITH UNIX AND THE	ADABAS/D, IMPLEMENTATION OF FULL UNICODE
Peter Edberg, Scientist, Apple Computer, Cupertino, CA.	X-WINDOW'S SYSTEM	SUPPORT.
•	Hideki Hiura, Software Architect, Sun Soft, Mountain View, CA.	Andreas Schütz, Manager Product Internationalization Software AG, Darmstadt, Germany.
25 33cm - \$3.5cm - 1		STANDARDS
<b>3:45 - 4:20</b> p.m. <b>48</b> (M/E/SD/P)	<b>B8</b> (M/E/SD/P)	C8
UNICODE AND THE WIN32 NLS API.	UNICODE SUPPORT WITHIN A UNIX ENVIRONMENT.	ISO 10646 - UNICODE : CONTINUED CONVERGENC
William Hall, Engineering Consultant, Novell, Inc.,	Jürgen Bettels, Globalization Consultant, Digital Equip.	Michael Ksar, Hewlett-Packard Company, Palo Alto, C
San Jose, CA.	Corp., Switzerland and Wendy Rannenberg, Technical Director, Digital Equip. Corp., Nashua, NH.	
1:25 -5:00 p.m. 49 (M/E/SD/P)		
UNICODE EDITOR APPLET OCX.		
Carl Forsander, VP Engineering, Gamma Productions, Inc., San Diego, CA.		
5:00 - 5:30 p.m.	Q & A	

# LISA フォーラム参加報告

AAMT技術動向調査委員会委員、兼、市場動向調査委員会委員 ソニー・テクトロニクス(株) 立田種宏

### 前書き:

昨年の10月1日~3日、シンガポールでLISAフォーラムが開催された。筆者は丁度その時期、アジア各国を回り、現地の機械翻訳の研究者達と技術交流をする予定になっていた。そこで、この機会を利用してLISAフォーラムにも立ち寄り、フォーラムの最終日に設けられたパネルにパネラーとして参加する事になった。ここでは、その時のパネルの様子を中心に報告する。

### 本文:

LISA (The Localisation Industry Standards Association: ローカライゼーション産業規格協会) は製品の現地化という視点から長年活動を続けている団体で、フォーラムの開催も今回が21回目となる。尚、LISA についてあまり馴染みのない読者は、本誌95年10月号のイベント紹介欄を参考にして欲しい。

今回のフォーラムのテーマは、『アジア市場でのローカライゼーションビジネスの展望』である。そのテーマのためか、初のアジア開催となった。フォーラムの初日、実質的な運営責任者である M. Anobile 氏から挨拶があり、計47企業90名の参加が告げられた。Microsoft、HP、Xerox など国際化対応に普段熱心な企業からの参加者が多く見られ、特に HP は、主要現地法人毎に1名以上参加者を送り込むほどの熱の入れようであった。

### 1 主なセッション

フォーラム2日目から、実質的な発表が始まった。 冒頭、フォーラムのホスト役であるStream International Asia Pacific 社副社長 Y. Chen 氏から、 ビジネスにおけるアジアの重要性が述べられた。氏 は、アジアにおける情報産業分野は他の産業と同様、 成長著しく、世界で最も成長を遂げた国と地域のト ップ5の内、4つはアジア(すなわち中国、韓国、 台湾、それに日本)であるとの認識を示した。そし て、その4つの国と地域でビジネスを成功させる為 には、ローカライゼーションへの取り組みが欠かせ ない事を強調した。

キーノート・セッションとして、Lotus Development ジャパン社の R. Yanagiya 氏が「ローカライゼーシ ョンにおいてアジアから学んだ事」というテーマで 話をした。氏は、ソフトウェア製品のローカライゼ ーションの際、現地で使われる文字との違いによる 苦労が絶えなかった様子を紹介した。特に、Lotus のソフトウェア製品そのものの日本語化は、製品マ ニュアルの日本語訳などとは比較にならないほど大 変な作業だったようだ。文字の扱い(Character enabling) と、文字セット (character set) の OS 対 応には随分と苦労したらしい。しかし、この事は逆 に、ローカライゼーションそのものがビジネスとし て意味を持っており、一つのビジネスチャンスであ る点も強調していた。尚、これらの苦労は、OS が 新しくなるにつれて少なくなってきているらしい。 このセッションでは、翻訳の観点からの話が面白く 聞けた。それは、よく言われるところの直訳の問題 点についてである。氏の経験によると、英中の翻訳 は、英日の翻訳に比べれば直訳の場合ずっと問題が 少ないとの事である。その大きな理由は、英語と中 国語の語順が同じ事に由来する為との指摘は、うな ずける。直訳での意味の通じ易さという観点から見 れば、英中の場合、現状の機械翻訳ソフトの翻訳精 度でも十分に通じる場面は結構ありそうだ。その一 方で、単語単位での翻訳の方に以外な落とし穴があ り、'cut'、'copy'、'paste'などのような OS でお馴染み の単語などは、その国の文化に依存した言葉に訳さ ないと原語のニュアンスが伝わらないとの指摘があ った。

以下、最終日のパネルに至るまで設けられたセッションを挙げておく。

「ビジネスとローカライゼーションに関する戦略」

HP ジャパン・プリンター部門 S. Yin

「2バイトコードの市場におけるローカライゼー ション」

SAP ジャパン社 T. Shirk

- 「1つの中国語と2つのバージョン」ソフトウェア AG 社 A. W. Pun
- ・「ローカライゼーションの体験で得られた事」 IBM China / Hong Kong S. Choy
- ・「ローカライゼーション」 Microsoft インターナショナル社 D. Kieffer
- 「技術的チャレンジに関するローカライゼーション戦略」

Frame Technology 社 L. Richardson

- ・「シンガポールの IT 産業」 シンガポール国立コンピューター庁 K.
- 「アジアでのローカライゼーション製品管理の確立」

Sybase ジャパン社 M. Kokubu

「主力の DTP/WP パッケージをどのようにアジアの言語に対応させたか」

Stream International 社 G. Benoit

「多言語マルチメディア・ローカライゼーション に対する知的財産の創造」

FACT International 社 S. Forth

「ベンダーから見た日本のローカライゼーション」

Pacifitech 社 B. Myers

### 2 パネル

3日目最終日の最後のセッションとして、我々のパネル・ディスカッションが行われた。テーマは「アジアにおける機械翻訳の有効性の評価」である。ソフトウェア AG の A. Schuetz 氏がパネルの議長を務め、シンガポール国立大学の J. Ee 女史、長瀬産業の堂野前進氏、それに私の3人がパネラーとして壇上にあがった。

各パネラーの発言概要や持ち時間など大方の事は、 パネラー間の電子メールのやり取りで打ち合わせ済 みであった。それは、議長の Schuetz 氏が「機械翻訳は役に立たない」というデビル役を演じて問題提起をし、我々パネラー側が反論する形で各自、機械翻訳の有効性を説いていくというシナリオであった。そこで私は、その筋書きに基づいたプレゼンテーション用スライドを日本で用意し、PowerPoint でプレゼンテーションすべく愛用の PowerBook にスライド用データを積んで会場に乗り込んだのだが・・・パネルの前日、会場のプロジェクターに PowerBookを繋いでビデオミラーリング・チェックをしている時、フォーラムの運営責任者である Anobile 氏から思わぬ依頼が舞い込んだ。「講演者が一人キャンセルになったので、あなたのプレゼンテーション時間を倍あげよう。そこでお願いだが、明日までに追加分の配布資料を用意してもらえないか?」

おかげでその日は、宿に戻ってから一歩も外に出ることもなく資料の作成に追われる羽目になった。それでも、ハードディスクに残っていたデータの残骸をどうにかかき集めて何枚かスライドを用意し、ホテルの部屋の電話口と PowerBook とを音響カブラーで繋ぎ、PowerBook 内蔵のファックスモデムからAnobile 氏が泊まっているホテル宛に FAX 送信する事で何とか乗り切った。

さて、当日のパネルに話を戻すと、先ずデビル役の Schuetz 氏が、事前の台本通り次のように切り出した。「数年前、我が社を含む多くの会社が機械翻訳 ソフトを購入した。しかし我が社では、いろいろと 分析調査をした結果、機械翻訳ソフトは役立たずと の結論に達し、とっくにお払い箱となっている。機 械翻訳ソフトも、その当時に比べれば多少なりとも 進化したとの話もあるが、その真偽のほどは如何な ものか? これからパネラー達に語ってもらお う!」

そこで、まず最初のパネラーとして立った Ee 女史が、シンガポール国立大学 ISS(Institute of Systems Science)での実績を基に、機械翻訳の有効性を述べた。ISS には機械翻訳サービスを行う部門があり、女史はそこの責任者である。そこでの機械翻訳サービスは英語から他言語への翻訳で、今のところ中国語(簡体中文と繁體中文の両方可能)、マレー語、そ

して日本語が扱えるそうだ。女史は、機械翻訳を有 効に使うためのコッとして、一度翻訳した文章はグ ループ分けをして保存しそれをうまく再利用する事、 また、それを効率よく行うために専任者を置く事を 挙げた。そうする事によって、文章の一貫性と翻訳 の生産性が高まるとの見解である。特に、一度翻訳 で使った用語は用語集として利用する事が肝心で、 その際には、文中から該当する用語を効率よく探し 出す検索ソフトウェアが重要な役割をするとの見解 を示した。また、現状の機械翻訳ソフトを有効に使 うためには、後編集が不可欠である事も述べた。最 後に、女史の結論として、機械翻訳は用語の統一、 文章の再利用、翻訳のターンアラウンドの速さの点 で手翻訳よりも数段優れており、大量の文書、特に 技術文書の翻訳には最適であるとの認識を示した。 そして、ISSでの2年間にわたる機械翻訳サービス の実績から、機械翻訳は有効な手段であると締めく くった。

次に私がパネラーとして立ち、事前に用意していた 事例を幾つか交えながら機械翻訳ソフトの有効性を 述べた。その後は、昨日急場しのぎに用意したスラ イドを使いながら、今後の機械翻訳を展望してみた。 そこでは、ここ数年の間に劇的な価格破壊が起こっ た事、また翻訳対象物の入手に手入力や OCR 入力 を必要とするような従来型の世界からインターネッ ト等の普及によって機械翻訳ソフトが直に扱える世 界へと移行しつつある事などを、幾つかの統計デー タを交えながら紹介した。そして、これらの流れは まだ続くと思われ、その結果、機械翻訳ソフトの普 及にもさらに拍車がかかるであろうとの予測を示し た。そして、これまで一部の人達だけの特殊ツール に思われ勝ちであった機械翻訳ソフトが、簡単な辞 書引きや、文章のアウトラインを素早く認識する為 のツールとしても十分に利用価値があり、これから 多くの人にとって極めて身近な存在へと変わってい くであろう点を強調した。事実、日本では最近、低 価格帯の機械翻訳ソフトを中心に爆発的な売れ行き を示している事をつけ加えておいた。その一方では、 従来のように高い翻訳精度を期待する需要も根強く あり、その分野に対する機械翻訳ソフトの改善も 徐々にではあるが進むだろうとの見地も示した。

最後にパネラーとして立った堂野前氏も、機械翻訳 ソフトが十分に役立っている事を数々の事例を交え て示した。そして、高品質の翻訳を実現する為には 後編集が大切である点を説き、機械翻訳ソフトに過 度な期待を持つ事は禁物であるとの認識も示した。

### 3 フォーラムの終わりに

パネルの後、運営責任者の Anobile 氏がフォーラム を総括して挨拶し、LISA は今後、様々なコンサル ティンググループや規格協会の意見を仰ぎながら、 ベンダー側と利用者側とのバランスをうまく保って 運営を続けていく事を表明した。また、調査研究報 告の作成など幅広い活動の展開も進めていく事もつ け加えた (その一環として、レポート形式による LISA OA Model の案内があるのでそれを次頁に掲載 しておく)。また、これからは、このアジア太平洋地 域からより多くの参加者を募るべく、この地域の会 員が興味を示すような内容の催しを検討したいとの 事である。さらに、フォーラムとは別の形でのワー クショップを、例えば韓国で開催する考えのある事 も表明した。最後に、今年(1996年)6月の香港で の LISA フォーラム開催を提案し、このフォーラム を締めくくった。

尚、最近届いた M. Anobile 氏からの私信によると、 今年アジアで予定している LISA フォーラムは香港、 もしくは日本での開催になるとの事である。

### 後書き:

今や情報産業も、その生産基地を日本等からアジア 各国へシフトさせている。その一端を如実に物語る 場所はパソコンショップであろう。北京の中間村、 バンコックのパンティップラザ、香港の旺角、尖沙 咀、灣仔、中環、シャムスイポー等、情報産業の成 長著しい地域には秋葉原級のパソコンショップが在 存する。ここシンガポールも、福南中心という大型 ビルの中にパソコンショップが数十店舗入居してお り、アジアの熱い動きを感じ取る事が出来る。

### The LISA OA Model

### A Standardized Quality Assurance Model for Software Localization

LISA is pleased to announce the availability of the LISA Quality Assurance Model, which is designed to assist the localization research, production and quality control processes for software developers and service vendors.

The LISA QA Model provides guidelines and metrics for localization quality. The documentation and templates help clients and vendors implement a synchronized QA system at multiple production sites. All registered users will be published by LISA to familiarize more localisers with the model and to increase networking.

The LISA QA Model is the result of localization professionals working together to influence the industry with standards, tools and quality procedures.

### The LISA QA Model

A Standardized Quality Assurance Model for Software Localization

- \* Localization Process Quality Control
- \* QA Checkpoints
- \* QA Procedures for:
- Software
- Packaging
- Documentation
- Computer Based Tutorials
- \* Printed and Online Documentation QA
- Language Quality Criteria
- Formatting Quality Criteria
- Functional Criteria Printed and Online Doc
- Language Quality Criteria for Software

- Formatting Quality Criteria for Software
- \* Software Quality Checklists
- Language
- Formatting
- Country Compatibility
- \* QA Procedures for:
- How to Use QA Form Templates
- Error Criteria
- Quality Definitions
- QA Form Template Samples
- QA Checklists
- \* Document Quality Checklists
  - Language
- Formatting
- Functional Art & Design
- \* Project Management Follow-up Forms
- Client Follow-up Form
- Vendor Follow-up Form
- \* Sampling
- Types of Sampling
- Acceptance Sampling
- Acceptance Sampling Procedures

..............

If you would like to order the LISA QA Model, please contact:

Jane Barney, LISA Administration

2 bis rue Ad.-Fontanel

CH-1227 Carouge

Geneva, Switzerland fax: +41 22 301 5761)

e-mail: manobile@divsun.unige.ch

### 協会活動報告

運 営 委 員 会 1月25日 各委員会・研究会の状況報告。創造的ソフトウエア育成事業提案内容の 説明。'96年度計画概要案の説明。

> 2月16日 '95年度収支見通しの報告と検討。創造的ソフトウエア育成事業ヒアン グおよび第3次公募説明会の報告。

編 集 委 員 会 1月22日 AAMT Journal 13号の反省。14号以降の編集についての討議。

例 文 評 価 研 究 会 1月19日 今後の活動についての討議

2月23日 記号・固有名詞等の処理についての討議

市場動向調査委員会 1月23日 機械翻訳の市場動向についてのヒアリング2件

・カテナ株式会社 販売企画部長 厚川美和氏

・株式会社ノヴァ 代表取締役 古賀勝夫氏

# NLPRS'95 参 加 報 告

(株)日立製作所、基礎研究所 ソフトウェア研究プログラム、主任研究員 丹羽芳樹

第3回自然言語処理環太平洋シンポジウム

(Natural Language Processing Pacific Rim Symposium '95, 略称 NLPRS'95) が12月4日から7日にかけて4日間、韓国ソウル市のソフィテル・アンバサダーホテルで開催された。会議委員長は KAIST (Korea Advanced Institute of Science and Technology) の Kim, Gil Chung 教授、プログラム委員長は東京工業大学の田中穂積教授、そして組織委員長が KAIST の Choi, Key-Sun 教授という顔ぶれである。採択論文数はポスター51件を含めて116件、会議への参加者は約270人という盛会であった。また初日には招待講演5件とデモンストレーション2件があり、最終日には言語工学を主題としたポストワークショップが開かれた。

次回の NLPRS'97はタイのプーケットで97年12月の 第2週に開催される予定である。また併せてポスト ワークショップがマレーシアのペナンで開催される。 さらに次々回の NLPRS'99は中国北京市で開催され る予定である。

### 1. はじめに

自然言語処理環太平洋シンポジウムは、91年11月のシンガポール、93年12月の福岡と経て今回のソウルが3回目の開催となる。主に太平洋を取り巻く地域、なかでも東アジア地域における自然言語処理研究者間の交流を促進することを目的として隔年で開催されており、回を追うごとに内容・規模ともに充実してきておりその重要性が増しているように感じられる。

3日間のプログラムは初日に招待講演とポスターの一部、2日目と3日目が一般発表とポスターという構成であった。また最終日には京福宮の国立民俗博物館に場所を移して「言語工学(Language Engineering)」に焦点を当てたワークショップが開か

れた。

### 招待講演

初日には次の5件の招待講演があった。

- Derivation of Definition Primitives from A Monolingual Dictionary, Huang Changning (Tsinghua University, PRC)
- Language Comprehension as Abduction, Abduction as Reasoning Under Uncertainty, Eugene Charniak (Brown University, USA)
- "You shall know a word by the company it keeps —J.
   R.Firth" Kenneth Ward Church (AT&T)
- · Digital library in the 21st Century 長尾真(京都大)
- · Prospects for Spoken Dialogue Systems in a Multimedia Environment 古井貞熙(NTT/東工大)

Huang 教授の講演は辞書における定義される語と定義する語との参照関係に注目し、何らかの意味での基本語のセットを構成しようとする研究である。基本語のセットとはそれらの単語で定義できる語を順次加えていくことにより全ての単語の定義が可能となるような極小のセットのことである。そのようなセットの言語的な意味については疑問もあるが、グラフ論的な立場から興味深い面もあるのではないかと思われた。

Charniak 教授の講演は論理をベースにした言語理解に関する研究であり、解釈に曖昧性がある場合には既存知識からの推論コストが最小になるような解釈を取るのが良いという趣旨であったと理解している。Church 博士の講演は単語の意味はそれと共に使われる単語から推定されるはずである、という Firth の命題を巡る統計的な話である。単語間の共起度を測る尺度として相互情報量、またそれとは対照的に

単語間の非類似度を測る尺度としてt-スコアが紹介され、実際の統計データに基づいてこれらの尺度が有効であることが示された。

長尾教授の講演は将来のディジタル図書館に関する 提言である。従来の図書館と電子図書館での決定的 な差は、従来は書籍の蓄積と検索が主な機能であっ たのに対して電子図書館ではさまざまな角度からの 電子的な読書支援が加わることであるとの指摘があった。また目次を利用した検索は実用的な観点から 重要な技術であることが実感できた。

古井教授の講演は音声対話システムについての全貌から技術的な詳細までをほぼ網羅する講演であった。音声認識技術としては HMM技術ベースのNoisyHMMが強力なツールであるとの紹介があり、応用例として示された電話番号案内(7万6千語認識)は説得力のあるものであった。

### 2. 一般発表

一般発表は2本のパラレルセッションで行なわれ合計66件の発表があった。全体は16のセッションで構成され、その内訳は以下の通りである。

セッション名 発	表件数
形態素	4
構文解析 (1~4)	16
意味論	4
談話・語用論	4
生成	4
音声・対話(1,2)	8
レキシコン	4
コーパス (1, 2)	8
機械翻訳	5
情報検索(1,2)	9

発表者を国別でみると次のようになり、日本と韓国 からの発表が大部分を占める結果となっている。

日本	33	中国	3
韓国	18	香港	2

アメリカ	2	カナダ 1
イギリス	2	シンガポール 1
台湾	1	スェーデン 1
インド	1	フランス 1

発表内容は自然言語処理全般をほぼ網羅したものとなっているが、大きく分けて、応用にウェートを置いたものと要素技術にウェートがあるものがほぼ同数で30件前後ずつ、その他に認知心理や計算論などにウェートがあるもの5件という分布であった。

応用システムという観点から分類すると以下のようになり、やはり機械翻訳に関する研究が大きなウェートを占めていることが分かる。それ以外では文書検索や分類といった文書処理関係への応用と対話システムに関する研究が盛んである。

応用ターゲット	件数
機械翻訳	8
対話システム	3
文書検索	3
文書分類	2
要約システム	1
校正支援	2

また技術的なキーワードで目だったものとしては以下のようなものが挙げられる。

技術課題	件数
曖昧性(多義)解消	7
未知語の認識	3
格パターンの学習	3

以下各セッションから筆者の印象に残っている発表をいくつかピックアップして紹介する。もちろん全ての発表を聞いた訳ではなく、また理解できたのはさらにその一部になるので重要な論文を漏らしてしまっている可能性が大きいことはあらかじめご了解頂きたい。

コーパス 大規模コーパスから語彙知識の自動獲得

を試みる研究は現在最も盛んな分野であり、いくつかの興味深い研究があった。東工大の白井らは構文構造が付加されたコーパスから自動的に文脈自由文法を抽出する方法について論じ、構文構造内のノードに自動的に非終端記号を割り当てて曖昧性のない文法を抽出したと報告している。

香港の WenJie Li らは文字 n-グラムの手法を中国語に適用し、未知語の同定を高精度で行なえることを示した。また未知語に関しては他のセッションでもいくつかの発表があり、未知語同定を情報検索におけるキーワード抽出に応用する研究(韓国の Hyuk -Chul Kwon等) や話し言葉における未知語の解析を扱った研究(豊橋技科大の Ohtani等) もあり、ホットな研究ターゲットになっているという印象を持った。

談話理解と語用論 ある意味で統計処理と対極をなすセッションかもしれない。言語の深い意味に関わる現象を研究者の頭で精密にルール化しようとするのが基本戦略になっているようである。横浜国立大学の西沢、森らから2件の発表があり、日本語の省略主語の推定問題について、条件文などの特定の環境下での詳細な分析が行なわれている。多くの例を整合的に説明するルール化が試みられており研究の蓄積が感じられた。

機械翻訳 機械翻訳に関しては一般発表は5件に絞られていたが、ポスターには10件の発表があり多くの投稿があったことが伺われる。一般発表の中で特に新鮮な印象を持ったのはNTTの池原らによる動詞格パターン対の必要量推定に関する研究である。辞書などから半自動的に収集される数量と、動詞の種類を絞り人手を掛けて思いつく限りのものを収集した数量とを比較して全体数を割り出すもので、推定値は約2万5千パターンということであった。

レキシコン セッション名はレキシコンとなっているが実際には意味論のセッションと言っても良いかも知れない。

奈良先端大の宇津呂の研究は日英対訳コーパスを用いて日本語動詞の用例の自動分類をめざしたもので、

目的語などに取られる名詞のクラスと動詞の英訳語 との対応付けを最適化することにより高精度な分類 が可能になることを示している。

なお筆者もこのセッションで語彙的曖昧性の解消に 関する結果を発表した。文脈と意味との関連度を共 起統計に基づいて計算するもので、関連度の最も高 い意味が採用される。それほど高い精度が得られる 訳ではないが完全自動化ができるのがこの方法の魅 力であると考えている。

意味論 韓国のJeong-Mi Choらによる"Korean Verb Senses Disambiguation Using Distributional Information from Corpora"は目的語として取る名詞の種類によって動詞を分類する手法を韓国語に適用した研究であり、これにより動詞の多義解消が高精度で行なえることを示した。

またこれは構文解析のセッションの発表であるが、 山梨大学の福本は前置詞の係り先の曖昧性解消にコーパスから得られる3語対の情報を利用する手法を 提案している。統計的スパースネスを解消するため にシソーラスに類する情報を利用している点がポイントと思われる。

ポスターセッション 分野別の件数は表の通りであった。

発表件数
10
7
8
6
4
3
3
2
6

また初日夕方のポスターセッションでは次の2件の デモが行なわれた。

- · TRANSIT and TermStar
- · j.Seoul/JK\_data01

デモについての詳細は省略するが、TRANSIT と TermStar は Star (Software – Translation – Artwork – Recording) 社出展の機械翻訳システム (TRANSIT) と辞書構築・検索ツール (TermStar)、またj.Seoul は Koden 社出展の Windows95上で動作する韓日機械翻訳システムである。

### 3. ポストワークショップ

NLPRS'95に引き続きポストワークショップが7日の午前9時から午後5時まで国立民俗博物館にて開催された。ポストワークショップは93年のNLPRSから開催されており、今後の重要な研究テーマ、効果的な研究推進、国際協力などについての意見交換を目的としている。93年のテーマは自然言語処理の将来展望(Future NLP)であった。今回のテーマは言語工学(Language Engineering)であり、12件の発表があった。

### プログラム

- System, Grammar and Parsing Mechanism in MT Zaharin Yusoff (University Sains Malaysia)
- · Richly Tagged Text Corpus 野村浩鄉(九州工業大学)
- ・Conceptual Graph Language, Vilas Wuwongse (Asian Institute of Technology, タイ)
- · Some Ideas about Translator's Workstation
  Development
  周明 (Ming Zhou) (Tsinghua University,中国)
- · Intelligent Information Access to Text on Information Highway,野村浩鄉(九州工業大学)
- · Global Communication and Language Barrier 江原暉将(NHK 放送技術研究所)
- · Characteristics and Development Strategies of Language Engineering
- Changning Huang (Tsinghua University, 中国)
- A Brief Survey of Recent Trend of SIGNLP of IPSJ 新田義彦(日本大学/日立)
- · Truly Successful Natural Language Processing

Requires a Combination of an Engineering and a Human-Computer Approach
Michael Zock(LIMSI、フランス)

- Computational Linguistics & Language Engineering
   --Roles of Theories in Language Engineering
   -- 辻井潤一(東京大学)
- ・Language Engineering on the Highway: New Perspectives for the Multilingual Society, Ruslan Mitkov (University of Wolverhampton, イギリス)
- · Language Engineering, Information Architecture and Document Interchange Format: Distributed Documentation, Key-Sun Choi (KAIST, 韓国)

1日のプログラムとしては大変盛りだくさんであり、テーマである言語工学、(あるいは「工学としての自然言語処理」と言った方が適切かも知れない)に関してバラェティに富んだ観点から議論を聞くことができた。産業界においては自然言語処理技術を必要とする力強い応用ターゲットが求められる今日、時宜を得たテーマであったと思う。

具体的な提言としては、九州工大の野村教授からタグ付きテキストコーパスの重要性に関する提言があり、また KAIST の Choi 教授からは韓国語を対象とした研究開発のための共通プラットフォームである KLIP の紹介があった。その他一つ一つの発表に特色があり、全てをご紹介できないのが残念であるが、当日の議論で特徴的であった+LE (Language Engineering)と-LE の対比について一端を紹介したい。

一つはプログラム7番のTsinghua (清華) 大学のChangning Huang 教授による「言語工学の特色と開発戦略」と題した発表である。表1は Huang 教授による+/-LEの対比を示したものであり、異論を持たれる方も多いかと思われるがそれだけに示唆に富むものと考えられるのでご紹介させていただいた。筆者はこの発表を聞きながら+/-LE の差は一番目の動機と目的の項目だけではないかという印象を持った。それと付け加えるとすれば、研究やアウトプットを評価する尺度として性能と同様に定式化の単純さ、あるいは理論的な根拠を評価するか(-LE) しないか (+LE) ではないかとも考えた次第

である。

しかしながら、10番目の発表の辻井教授によればそれは正しくないということになる。すなわち「伸びる+LE は+CL(Computational Linguistics)すなわち +理論であるはずである」と聞き、そうかもしれな いと思い直した次第である。このような議論から何か具体的なものが生まれるということはないのだろうが、少くとも私にとっては現実的効果(+LE)と理論的根拠(+CL)の両立を目指す意識を喚起された点で有意義であった。

表 1 Huang 教授 (Tsinghua Univ.)による+LEと-LEの比較

	+LE	-LE
動機・目的	理論的な関心	市場性
カバレージ	限定領域	汎用・大規模
方法論	Rationalism	Empiricism
知識獲得	直観ベース	観測(統計)ベース
計算モデル	プロダクションシステム	n-グラム, マルコフモデル
開発費	小	非常に大

### 新刊図書の紹介

(AAMT 会員の出版物)

書 名:「プログラミングの壷」

(第2部 人間編)

原 著: Programming on purpose

著 者:P.J.Plauger 訳 者:安藤 進

監訳者:石田晴久(東京大学)

定 価:2,700円 出版社:共立出版

Cコンパイラの開発者、ソフト会社の経営者、C言語標準化委員会委員長など多彩な経歴の持ち主が、いろいろな人との出会い、様々な出来事について書いたエッセイ集。著者独自の人間観察に基づくテーゼが面白い。「ソフトウェア設計編」に続く3部作の2冊目。

原著のタイトルは「目的意識を持ってプログラミングする」であるが、ちょうど、アラジンの不思議なランプのように、素晴らしいアイディアがたくさん詰まっている壷にちなんで邦訳を決めた。第3部は今年中に発行の予定。

### 関連記事紹介

### 英日翻訳支援ソフトウェア(ザ・国際人)

日経エレクトロニクス No. 654 (1996年1月29日号) の特集記事「1996年版読者が選ぶエレクトロニクス・ヒット商品」によると、三洋電機は、上記商品を自社の1995年のヒット商品のトップに選んでいる。また、1996年に期待する商品として「翻訳支援ソフトウェアの新バージョン」をトップにあげている。同社の機械翻訳に対する強い意気込みが感じられる。

### 翻訳業界の現状調査

社団法人日本翻訳連盟(JTF)では、平成7年度全国中小企業団体中央会の補助事業として「翻訳業界の取引契約の実態調査」を実施した。その内容が、平成8年2月8日東京都勤労福祉会館で開かれたJTF講習会において「翻訳ビジネス環境への対応」という題目で報告された。また、平成8年2月22日に日本経済新聞社大阪本社で開催された「翻訳のつどい」でも、「翻訳業界の現状と今後の翻訳の方向を模索する」と言う題目で報告された。近く報告書にまとめられるが、概要を述べる。

ユネスコの調査によると、1986年(少し古いが)の翻訳図書数は全世界で61,531冊であった。それを原書の言語別に見ると、英語が29,294冊で断然多く、露語・仏語・独語が $5\sim6$ 千冊で続いている。5位はイタリア語の1,754冊、6位はスカンジナビア語の1,718冊である。日本語は232冊しかない。

一方、翻訳された図書数を国別に見ると、旧ソ連と西ドイツがともに8,000冊強、日本は2,875冊で第3位である。英国は914冊、米国は21冊と非常に少ない。日本で出版された翻訳図書の大半は英語からで2,295冊あり、2位の独語224冊、3位の仏語214冊を大きく引き離している。

部門別にみると、全世界で文学が3,0545冊で約半 分を占める。(日本でも1,227冊で、翻訳書総数2,875 冊の約43%である。)つづいて応用科学の7,215冊、 社会科学の6,987冊が占める。

産業翻訳業界は、昭和61年2月の日本標準産業分類では、「他に分類されない専門サービス業」の中に入っていたが、平成5年10月にはじめて「翻訳業」として分類された。その間"高度成長""輸出振興""海外生産"等にともない企業での翻訳需要は増加して来た。今回調査した企業の85%が翻訳ニーズがあると答えており、残り15%は国内のみを対象とする企業であった。また、はじめのころは、技術部門や貿易部門など一部の部署での需要であったが、今や企業内の全部門で翻訳需要が発生している。

翻訳需要は国内企業からは減少し、外資系企業からは増加の傾向にある。それにともない、和英翻訳

が減り英和が増している。国内企業からの減少は、 企業内での処理が増加しているためと思われる。外 資系企業の84%は東京・神奈川に集中しており、ほ とんどの企業が好調に売上高を伸ばしている。しか し、84%の企業がビジネスコストの高いことを指摘 しており、また、語学堪能者・研究者等の人材難を 問題点として上げた企業も多い。

翻訳需要を産業分野別にみると、トップは加工組立分野、2位が化学・医薬分野で、この2分野だけで4~5割を占める。翻訳文書の種類でみると技術資料と製品マニュアルが圧倒的に多い。翻訳言語のほとんどは日英・英日である。

翻訳会社の特徴は、専門に特化した小規模企業が 非常に多いことである。大きな創業投資や設備投資 がいらず、市場参入が容易なためと考えられる。外 部翻訳者を多く抱え、社員の翻訳者はコーディネー ター役を行っている企業の多いのも目につく。

回答をもらった翻訳者の約7割は翻訳を始めたのが1986年以降で、急激な需要増にともなって翻訳業界に入ったものと思われる。身分は多彩で、社員翻訳者あるいは個人事業翻訳者として活動しておられる人のほかに、兼業・副業として翻訳をしている人も多い。理科系の翻訳者では企業経験者が多く、翻訳力に加えて専門分野の知識が生かされている。

翻訳取引の状況をみると、取引契約を文書で締結する場合は非常に少なく、また、発注・受注の仕様書等の作成も少ないようである。場合に応じて柔軟に取引されていると思われる。また、受発注の単価もかなり巾があり、翻訳の内容・難易度・品質等によって決められているようである。

原稿の交付方法としては、FAXと郵送によるものが多く、パソコン通信の利用はまだ少ないようである。郵送には紙が多く用いられているが、フロッピーの利用も増えつつある。

今後、高学歴化・海外経験者の増大・海外生産の 進展・国際ネットワークの拡充等の翻訳環境の変化 に対応して、翻訳業界は、翻訳品質の向上・ネット ワーク化・契約の文書化等を進めねばならない。

# 機械翻訳関連ソフトウェアー一覧表

(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	パシフィックアイ&	Windows	日→英	58, 000
-	ブラビス J / E SASUGA	エムティラボ	Windows	日→英	9, 800
	// Exign B oncount		WINGONG.		0,000
3	トランスランド	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/JC	三群インフォメーションヒ゜シ゜ネス	Windows	英 → 日	25, 000
	トランスサポーターE/JW		Windows	英 → 日	29, 800
	トランスサポーターJE		Windows	日 → 英	25, 000
5	WD-018W	シャープ	ワープロ書院	英→日	150, 000
6	日韓機械翻訳システム	日立情報ネットワーク	MS-DOS5. 0	日→韓	178, 800
3500	韓日機械翻訳システム	CONTRACTOR ATTACKNESS AND AN	MS-DOS5. 0	韓→日	178, 800
	14.		MS-DOS5. 0	韓←→日	298, 800
7	PIVOT/JE	NEC	EWS-UX/V	日→英	500,000
	PIVOT/EJ	2005 C	EWS-UX/V	英→日	500,000
	英日翻訳アダプタ		Windows	英→日	14, 800
	英日辞書引き君		Windows	英和	9, 800
	日英辞書引き君		Windows	和英	9, 800
8	ATLAS EJ for Windows	富士通	Windows	英→日	58, 000
	ATLAS JE for Windows		Windows	日→英	58,000
	電辞海		Windows	英和・和英	24, 000
9	パンセ・ブイ英日版	沖ソフトウェア&	Windows	英→日	38, 000
	パンセ・ブイ日英版	オージス総研	Windows	日→英	38,000
		6000 8000807.0000000	Windows	英←→日	58,000
	パ ンセフォーインターネット		95/NT	英→日	9, 800
	WWWサーハ・ーモテ・ル"ハ・ンセ"		Solarisl. X	英→日	148, 000
10	X-EJ2/W	スピリット	Windows	英 → 日	98, 000
10	X-EJ		MS-DOS3. 1	英→日	29, 000
				30.50 11.	
11	LogoVista E to J	カテナ	Windows	英→日	97,000
			Mac	英→日	97, 000
	LogoVista E to J		Windows	英→日	39, 800
	Personal		Mac 7	英→日	39, 800
	The Translator-Mini		Mac 7.1	英 → 日	14, 800
	コリャ英和		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
5.73			Mac 7.1	英 → 日	246,000
			Windows	日→英	246,000
			Mac 7.1	日 → 英	246,000

次号で、各システムの詳細を掲載する予定です。

	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	東芝	UNIX	英 ←→ 日	700,000円
	ASTRANSAC C/S		OS/AS	英 ←→ 日	340, 000
			Windows(NT	英 ←→ 日	340,000
	ASTRANSAC for Windows		Windows(NT	英 → 日	63, 000
			Windows (NT	日 → 英	63, 000
			Windows (NT	英 ←→ 日	98, 000
	ASTRANSAC/Web		OS/AS	英 ←→ 日	200, 000
14	J·London/EJ	高電社	Windows	英→日	49, 800
	J·London/JE		Windows	日 → 英	49, 800
	J·Londonツインセット		Windows	英 ←→ 日	78, 000
	J・ソウル日韓版		MS-DOS3.1	日→韓	390,000
	J・ソウル韓日版		MS-DOS3.1	韓→日	298, 000
	WorldNet/EJ		Windows	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	亀島産業	Windows	英 → 日	14, 800
	E-JBANK for Windows	100 TO 10	Windows	英 → 日	75, 000
17	Translation Manager/2	日本IBM	OS/2	英 → 日	785, 000
18	孫悟空	クリエイト大阪	MS-DOS3.1	日 → 中	190, 000
19	NetSurfer/ej	ノバ&	Windows	英→日	29, 800
	PC-Transer/ej	山種産業	Windows	英 → 日	198,000
			Mac	英 → 日	198,000
			DOS (PC98)	英 → 日	248,000
			DOS (DOS/V)	英 → 日	298,000
	PC-Transer/je		Windows	日 → 英	198,000
			Mac	日 → 英	198,000
			DOS (PC98)	日 → 英	248,000
			DOS (DOS/V)	日 → 英	298,000
	マイトランサーEJ		DOS (PC98)	英 → 日	98,000
			DOS (DOS/V)	英 → 日	98,000
	SuperTranser/ej		DOS (PC98)	英 → 日	448,000
			DOS (DOS/V)	英→日	448, 000
20	PowerTranslator Pro.	イリス インターナショナル	Windows	独・仏・西	148, 000
	PowerTranslator Pro.		DOS	独・仏・西	148, 000
	PowerTranslator Pro.		Mac	独・仏・西	148, 000
	PowerTranslator Pro.		DOS	露←→英	148, 000
	PowerTranslator		Windows	独・仏・西	58, 000
	PowerTranslator		Mac	独・仏・西	58, 000
	PowerTranslator		DOS	独・仏・西	38, 000
	PowerTranslator中国語		DOS	英 → 中	28, 000
	PowerTranslatorCD-ROM		Windows	独・仏・西	58,000

# 英語読解ツール「訳師如来」(for Windows/for Macintosh)

株式会社 CSK

### 1. 英語読解ツールとは?

近年、翻訳ソフトの品質はめざましくよくなりま した。ちょっとした文章ならすぐに完璧な日本語に 翻訳してくれます。

しかし、まだ長い文章をきれいな日本語に翻訳することは難しく、翻訳ソフトががんばって訳出した 読みにくい日本語を読むことは避けられません。

そこで、単語毎にその単語の訳を表示し、英語が ある程度分かっている方なら簡単に英語が読めるよ うになる英語読解ツール「訳師如来」が誕生しまし た。

ただの「辞書引き」ソフトではありません。「訳師如来」には、翻訳ソフト同様に構文解析のアルゴリズムが使用されており、各英単語の品詞を決定し、その品詞にふさわしい対訳を表示します。

### 2. 画期的な20万語の辞書を装備

市販の英和辞典には普通5~8万語が収録されています。また、市販されている翻訳ソフトでは7~10万語が基本辞書として提供され、あとは有料のオプション辞書となります。

「訳師如来」の辞書は、人名、地名、科学・経済の専門用語を含む20万語を収録しています。

新藤田田 TO MAN (1) 279.7(2) 表示の ラウドラ(2) A57(1) (2) (2) 表示の ラウドラ(2) A57(1) (2) (2) 表示の ラウドラ(2) (2) A57(1) (2) A5

図1:あんちょこ表示

多くの方に使っていただけるよう豊富な辞書と機能 を、低価格で提供します。

### 3. 読みやすい「あんちょこ」表示(図1)

「訳師如来」の大きな特徴の一つにこの「あんちょこ」表示があります。

「あんちょこ」表示は、英単語の下にそれぞれの 対訳を表示します。学生の頃、英語の教科書の文章 の意味の分からない単語の下に、その単語の訳を書 いたことがあるかと思いますが、そのイメージです。

「あんちょこ」表示の他に「挿入」表示、「置換」 表示と3種類の表示が用意されています。

### 4. 分かりやすい、簡単操作

「訳師如来」は、英文を読み込んでボタン一つですぐ単語翻訳を実行します。読み込める英文は、テキスト形式のファイルですが、直接エディタに入力することも可能です。他のアプリケーションからならコピー&ペーストでデータを持ってきて実行することが出来ます。

5. 邪魔な対訳は出さず、訳語の変更も簡単(図2)

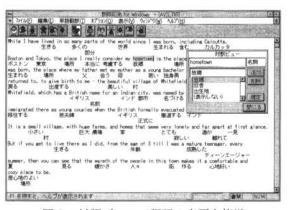


図2:対訳ビューで訳語の変更も簡単

「訳師如来」は、画面が繁雑にならないよう、"this" や "and"など、邪魔になる対訳は表示しません。単語毎に対訳を出すかどうかを設定することも出来るので、覚えた単語や知っている単語の対訳は出さない、ということも可能です。また、表示した対訳が気にいらない場合、「対訳ビュー」を使って簡単に別の訳語を選びなおすことが出来ます。

選びなおした訳語は、次回の単語翻訳から最優先 されます(学習機能)。

### 6. 単語の新規登録も可能(図3)

「訳師如来」は、自分で単語を登録することもできます。登録は、単語の変更と同じく「対訳ビュー」上で行います。辞書にない英単語で「対訳ビュー」を開くと「未定義語」として表示され、そのまま英単語と訳語、品詞を登録できます。訳語登録機能が自動的に起動するので、簡単にユーザが必要とする語を増やしていくことが出来ます。

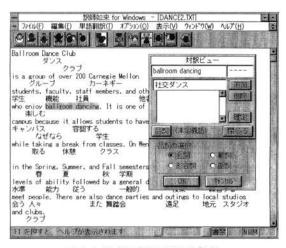


図3:単語の新規登録も簡単

### 7. お求めやすい低価格で

以上に述べたような機能がついて、価格は Windows 版・Macintosh 版共に6000円(消費税込み) と低価格でお求め頂けます。

### 8. 訳師如来の動作環境

### ●Windows 版

対応機種: MS-Windows3.1日本語版が

稼動する機種(#1)

メモリ : 8 MB 以上 HDD : 30 MB 以上

周辺機器: 要CD-ROMドライブ

# 1) Windows95上での使用はサポートしておりません。

### ●Macintosh 版

対応機種: 漢字 Talk7.1以上が稼動する

機種

メモリ : 6 MB 以上 HDD : 30 MB 以上

周辺機器: 要CD-ROMドライブ

### 9. ご購入について

「訳師如来」は書籍です。全国の大手書店、パソコンショップの書籍コーナーでお求め頂けます。また、お近くのお店でお求めになれない場合、パソコンネットの Nifty-Serve 内 CSK パッケージブティックや通信販売でお気軽にお求めになれます。

- ●「訳師如来」通信販売(T•ZONE) 0120-234-392
- ●Nifty-Serve 内 CSK パッケージブティックコマンド CSKPB

### 10. お問合わせ先

株式会社 CSK 訳師如来サポートセンター 〒169 東京都新宿区西早稲田 1 -21-1 早稲田大学西早稲田ビル 5 F

TEL 03-5286-7201 FAX 03-5286-7202

### お問合わせ時間

月~金(祝日を除く)9:00~18:00

# 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. 1 4 (March. 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 野村浩縣(委員長) 杉山健司 亀井真一郎 福持陽士 仁井正治

事 務 局 田中 孝 神野千惠子

印 刷 所 伸光写植印刷株式会社

