



AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.25
January 1999

アジア太平洋機械翻訳協会

目 次

エッセイ	鎖国時代の翻訳そして現在	1
論説	音声翻訳のための言語処理	5
技術早分かり	省略補完技術	8
新製品紹介	・ JICST 日英機械翻訳システム（成果公開版）	16
	・ 英日翻訳ソフト「EtoJ PRO V5.0」	18
	・ The 翻訳プロフェッショナル	20
国際会議報告	AMTA'98参加報告	22
MT Summit	MT Summit VII	2
事務局だより	委員の交替	7
	AAMT メーリングリスト	15
	協会活動報告	25

CONTENTS

Essay :	The Human Translation in the Isolation Era and Present Age in Japan	1
Paper :	Language Processing on Speech-to-Speech Translation	5
Technological Writings :		
	Supplementation for Omitted Elements in Machine Translation	8
New Products :	・ JICST JtoE Machine Translation System	16
	・ EtoJ Translation Software 「EtoJ PRO V5.0」	18
	・ The Translation Professional	20
International Conference :	AMTA '98 Report	22
MT Summit :	MT Summit VII	2
AAMT News :	Replacement of Committee	7
	AAMT Mailing List	15
	AAMT Activities	25

鎖国時代の翻訳そして現在

東京工業大学教授 AAMT 会長 田中 穂積

鎖国政策をとった江戸幕府は、外国からの情報を完全にシャットアウトしたわけではない。日本から異国に向けて開かれた唯一の窓口は、中国を除きオランダ一国であり、それも厳しい制約のもとでの交易が許されていたにすぎない。このように細々としたチャンネルではあったが、長崎の出島のオランダ商館からもたらされる異国の情報に幕府は耳をそばだてていた。異国からもたらされる情報が幕府にとって重要なものであることを認識していたからである。8代将軍吉宗が漢訳洋書の輸入制限を緩和し、青木昆陽らにオランダ語学を学ばせたことは、吉宗が新しい物好きの好奇心の旺盛な人柄であることに加えて、必要に迫られてのことであつたに違いない。

オランダ商館からは蘭書も手に入るようになり、細々とではあるにしても西洋のさまざまな知識が日本に流入する。それとともに蘭学が発展し、医学、薬学、天文学、暦学、植物学、地理学、数学、化学、兵学など多岐にわたる知見が日本にもたらされた。幕府はさらに蛮書和解御用という機関を設置し、そこで洋書の翻訳にあたらせ西洋の科学技術の吸収に一層努める。蛮書和解御用は日本における文献翻訳センターの第一号とよぶべきものである。

蘭書の日本語翻訳に対する蘭学者の寄与は勿論大きい。長崎奉行所とオランダ商館との意志交換を行うための通訳者、通詞の存在を無視することはできない。オランダ人と直接接する機会の少なかった大坂や江戸在住の蘭学者と違って、彼らは長崎で直接オランダ人に接触していたので、優れた通詞のオランダ語に関する知識は蘭学者のそれを上回っていたと思われる。実際彼らは通訳だけでなく蘭日辞書の編纂や蘭書の翻訳にも従事している。万有引力の法則や地動説を日本にはじめて紹介したのも長崎通詞の志筑忠雄であった。

江戸時代と言えば解体新書の前野良沢・杉田玄白

に触れない訳にはいかない。江戸詰めの中津藩医であつた前野良沢は、医学の知識をさらに深めるため蘭書が自由に読めるようになりたいとして、齢47にしてオランダ語を学ぶ目的で長崎への遊学を決意する。これは、江戸の蘭学者からの蘭学修得には限界があること、長崎にいる大通詞に直接蘭語を教えてもらいたかつたこと、そしておそらくはオランダ人にも直接会いたかつたからかもしれない。

長崎滞在中の彼は、ドイツ語からオランダ語に訳された解剖書「ターヘル・アナトミア」をオランダ商館員から1770年に手に入れる。日本初の腑分けを見学して山脇東洋が著した「蔵志」とよばれる人体解剖図がすでにあつたが、「ターヘル・アナトミア」の解剖図がそれとは異なることを知る。江戸に戻ってからその翻訳の準備として、以前にもましてオランダ語の研究にいそしむ。友人であつた玄白も偶然この書を手に入れており、意気投合した二人は彼らの弟子とともに、この書の翻訳に精力的に取り組む。玄白は翻訳チームのリーダーとしての資質は優れていたが、語学の研究に打ち込む姿勢に欠けるところがあつた。寝食を忘れ最後まで粘り強く翻訳に取り組んだのは良沢であつた。医学独特の専門用語の翻訳に苦しめられながら。

「解体新書」は前野良沢の血の滲む翻訳作業の結果に負うところが大きい。にもかかわらず「解体新書」の訳者に良沢の名はない。翻訳結果に満足せず訳者の一人となるのを固辞したからである。玄白は「解体新書」の出版後オランダ語研究から離れる。一方良沢はオランダ語の研究をさらに究め、医学以外の蘭書の翻訳にも没頭する。このあたりの事情は吉村昭著「冬の鷹」に詳しい。玄白は晩年名医の名をはしいままにする。ところが吉村の筆は、完べき主義で妥協を許さず偏狭なところもある良沢にあくまで優しい。

「冬の鷹」からは、専門用語蘭日対訳辞書がない状態での翻訳作業の困難さや、翻訳の原点が言語の研究にあること、通訳者からの教えも重要であることが読み取れる。これらは現在の機械翻訳システム開発者にとっても共通して言える事柄ではないだろうか。幸いなことに言語の研究や専門用語辞書の整備は当時と比べものにならないほど進んでいる。蘭日機械翻訳システムが今あれば、それと彼らの翻訳結果とを比べてみたい気もする。

随想風の巻頭言になってしまったが、ここで話題を変える。AAMTにとって今年最大のイベントは、9月にシンガポールで開催される第7回 MT サミットである。辻井潤一委員長の組織するプログラム委員会が、大変魅力的な内容をプログラムに盛り込んでいる。このサミット参加者は21世紀に向けた機械翻訳システムの新しい予兆を感じることができるだろう。サミット開催者の一人として、世界各国、特にアジアからは是非多数の参加者を得て会議を実りあるものにしたいものである。

MT Summit

MT Summit VII

アジア太平洋機械翻訳協会
会長 田中 穂積

アジア太平洋機械翻訳協会（AAMT）は、来年の9月に第7回機械翻訳国際会議「MT Summit VII」をシンガポールで開催いたします。

日本が提唱して始まった「MT Summit」は、1987年9月に箱根で第1回会議を開催したあと、2年ごとにアジア太平洋地域、ヨーロッパ、アメリカの順で開かれ、グローバルな機械翻訳の発展と交流に大きな成果をあげて参りました。

今回の会議はテーマを「MT in the Great Translation Era」といたしました。世界的な大翻訳時代を迎え、機械翻訳に対するニーズと期待が高まるなかで、私たちは何を研究開発しどんな情報とサービスを提供していくべきか、広くユーザーや市場の方々にも参加いただき、新しい時代の方針を論議していきたいと計画しております。

この会議が実り多いものとなるよう格別のご支援ご協力をお願い申し上げます。

1. 会議の名称 MT Summit VII（第7回機械翻訳国際会議）
2. 主催 アジア太平洋機械翻訳協会
3. 後援 通商産業省 （予定）
4. 協力 社団法人日本電子工業振興協会
財団法人国際情報化協力センター
5. 協賛
 社団法人日本翻訳連盟
 社団法人日本翻訳協会
 社団法人情報処理学会
 社団法人電子情報通信学会

社団法人人工知能学会
財団法人日本情報処理開発協会、
言語処理学会
テクニカルコミュニケーター協会

6. 会期

1999年9月13日（月）～17日（金）の5日間

9月13日（月）	Tutorials	Reception
14日（火）	Conference • Exhibition	
15日（水）	〃 〃	Banquet
16日（木）	〃 〃	
17日（金）	Workshop	

7. 会場

シンガポール
シンガポール国立大学内 Kent Ridge Digital Labs

8. 参加予定人員および出展予定数

参加200名（内訳：欧米50名、シンガポール50名、日本50名、他のアジア各国50名）
出展30社（内、日本から10社）

9. 予算の概要

収入および支出はシンガポールのローカルオーガナイズングコミッティが担当。
予算は概算で74,000シンガポールドル、約555万円の予定。

10. MT Summit VII組織委員会

名誉委員長	長尾 真	京都大学 総長
委員長	田中穂積	東京工業大学大学院情報理工学研究科 教授
副委員長	辻井潤一	東京大学大学院理学系研究科情報科学専攻 教授
	Dr. Low Hwee Boon	シンガポール Kent Ridge Digital Labs
委 員	小谷泰造	株式会社インターグループ 代表取締役
	丸山 武	富士通株式会社 代表取締役副社長
	吉川英一	日本電気株式会社 常務取締役
	浅田 篤	シャープ株式会社 常任顧問
	岡村 正	株式会社東芝 取締役上席常務
	松香茂道	株式会社日立製作所 専務取締役
	三木弼一	松下電器産業株式会社 取締役
	牛尾真太郎	沖電気工業株式会社 常務取締役
	田中達雄	社団法人日本電子工業振興協会 専務理事
	野村浩郷	九州工業大学情報工学部知能情報工学教室 教授
	石崎 俊	慶応義塾大学環境情報学部 教授
	山崎 章	財団法人国際情報化協力センター 専務理事
	小紫正樹	財団法人国際情報化協力センター シンガポール事務所長

酒井保良 株式会社国際電気通信基礎技術研究所 代表取締役副社長
 勝田美保子 社団法人日本翻訳連盟 会長
 Dr. Tong Loong Cheong シンガポール Kent Ridge Digital Labs
 Prof. Choi Key-Sun 韓国 KAIST
 Prof. Zaharin Yusoff マレーシア University Sains Malaysia
 Prof. Huang Changning 中国 Tsinghua University
 Dr. Su Keh-Yih 台湾 Behavior Design Corporation
 Dr. Wu Dekai 香港 University of Science & Technology

11. MT Summit VII 実行委員会

委員長	雨宮弘和	株式会社東芝
副委員長	森本秀樹	富士通株式会社
資金財務担当	奥平捨男	株式会社日立製作所
宣伝広報担当	種子田暁夫	日本電気株式会社
展示担当	樋口和雄	社団法人日本電子工業振興協会
会員調査団担当	菅野 淳	松下電器産業株式会社東京支社
渉外担当	山畑征四郎	株式会社インターグループ
広報協力	野村勇一	社団法人日本翻訳連盟
IAMT 行事担当	永田淳次	沖電気工業株式会社
関西地区担当	福持陽士	シャープ株式会社
広報協力	高橋省三	財団法人国際情報化協力センター

12. MT Summit VII プログラム委員会

委員長	辻井 潤一	東京大学 教授
委員	伊藤 悦雄	株式会社東芝
	梶 博行	株式会社日立製作所
	亀井真一郎	日本電気株式会社 C&C メディア研究所
	松井くにお	株式会社富士通研究所
	森本 秀樹	富士通株式会社
	飯田 仁	株式会社ソニーコンピュータサイエンス研究所
	白井 諭	株式会社 ATR 音声翻訳通信研究所



音声翻訳のための言語処理

株式会社 エイ・ティ・アール 音声翻訳通信研究所
白井 諭 隅田英一郎 柏岡秀紀 山田節夫

1. はじめに

エイ・ティ・アール音声翻訳通信研究所では、人が喋った言葉をコンピュータが外国語に翻訳し喋るシステムの実現を目指し、基礎技術の研究を行っている。要素技術として、人が喋った言葉を聞き取る音声認識、聞き取った言葉を他の言語に翻訳する機械翻訳、翻訳された言葉を音声として出力する音声合成、これらの処理を統合管理する音声言語統合処理の4つの構築を進めてきた。1998年10月末、旅行会話を対象として日英および英日の翻訳をほぼ実時間でこなすプロトタイプシステムの開発に成功した。

音声認識（ATR-SPREC）は、不明瞭な発声でも認識できる音響モデルを提案し、高精度の音声認識が可能である。機械翻訳（TDMT）は、話し言葉特有の言い回しとその対訳文に基づいてパターンと用例をデータベース化し、これと入力文の類似性を調べながら、覚えている表現を使い適切に翻訳する。音声合成（CHATR）は、音声データベースの中から適切な音声波形を選択し、必要に応じて信号処理を行なうことにより、極めて高品質の合成音声を生成する。このとき、音声認識で得た話者の発声の特徴パラメータも反映されるため、女性が喋れば女性の音声を合成する。

以下、本稿では音声翻訳を実現するために開発した機械翻訳を含む言語処理について述べる。

2. 音声翻訳における言語処理の課題

音声会話の認識結果を翻訳対象とすると、記述文を翻訳する場合に比べ、次のようなことが特に問題となる。

- a. はば瞬時に翻訳される必要がある
- b. 文法的に誤っている表現が多く含まれる
- c. 前提事項は言わないため断片的な表現が多い

- d. 音声認識結果には誤りが含まれることがある

音声翻訳を通して会話を行なう場合、喋ってからその翻訳が合成音声で聞こえるまでに長い時間を要しては会話は成立しなくなる。また、喋っている言葉には詳細に分析すると文法的に間違っている表現も少なくないが、その状態で会話は成立している。すなわち、文法的に誤った表現でもそれなりの翻訳ができることが必要になる。そこで、変換主導翻訳（TDMT）を提案し、これらの問題解決を図った。

会話は面前の相手との間で成立するものであり、両者の了解事項は改めて発話されないことが多い。特に日本語の発話では、「誰が」や「何を」が省略されるため、英語など主語や目的語の省略が許されない言語への翻訳は難しい。会話を対象に、これらの省略要素を自動的に補完する技術を実現した。

また、音声認識結果には認識誤りが含まれることがある。記述文翻訳では、しばしば前編集や後編集が行なわれるが、会話翻訳にはそのような介入は不可能である。TDMTでは覚えている表現を使い回して翻訳するが、誤りが含まれるとき、覚えている表現との類似性が低いという現象が生じることがわかった。そこで、その類似性に基づいて正しいと推定される部分だけを翻訳する部分翻訳処理を実現した。

3. 変換主導翻訳

変換主導翻訳（TDMT）として、実際に出現した表現を模倣して翻訳することにより、文法的に間違った表現であっても翻訳することができる方法の実現を目指した。すべての表現をあらかじめ登録するのは不可能であるため、特徴的な表現を一般化して登録する事前学習を行なっている。例えば、「中国の河」は「XのY」というパターンに一般化する。一般化すると元の意味が失われるので、このパターンに実際の表現「中国の河」とその対訳「river in

China」を一緒に登録する。翻訳システムに文が入力されるとその表現に当てはまるパターンを検索し、そのパターンに登録されている表現のうち入力された表現と意味的に最も近い表現を選択し、選択された表現を真似て訳文を作成する。

「京都に到着する時間を教えてください」という文が入力された場合、次のようにして訳文が生成される。まず「京都に到着する」のパターン「XにY」に着目する。このパターンには「大阪に着く」「3時に着く」「友人に会う」「ひどい目に遇う」などの表現が登録されているとする。シソーラスを用いることにより「京都に到着する」に意味的に近いものとして「大阪に着く」を選択する。その対訳「arrive in Osaka」を模倣して「arrive in Kyoto」を生成する。以下、「到着する時間」に着目し「the time when」を、「時間を教えて」に着目し「tell (me) the time」を、「教えてください」に着目して「please tell (me)」を順に生成する。これらを文法規則に従って組み合わせることにより、最終的に「Please tell (me) the time when (I) arrive in Kyoto」と翻訳する。

多くの翻訳システムでは規則型の翻訳方式に基づいており、文法規則に基づいて解析・変換・生成を行なっている。翻訳システムを改良するには、特に変換性能を向上させる必要がある。規則型翻訳では翻訳失敗を生じた原因となっている文法規則の改良を必要とするが、安易に規則を追加すると想定外の表現に適用され思わぬ副作用を招く恐れが大きく、専門家が十分に副作用等を検証することが必須条件となっている。これに対して、TDMTでは変換の単位となるパターンを見いだすのに必要な解析と上述の生成には文法規則を用いるが、変換はパターンごとに登録された翻訳例に基づいて実行される。従って、翻訳例を追加することにより、比較的容易にシステムの性能向上が達成される。

4. 省略格要素補完

前節の翻訳において、「Please tell (me) the time when (I) arrive in Kyoto」の括弧を付した部分は、入力文「京都に到着する時間を教えてください」には含まれていない。日本語の表現としてはこれらを言わないのが普通であり、英語の表現としては必ず言わなければならない。これらの要素を推定する技術が必

要となる。

翻訳の実例は、このような要素が盛り込まれた状態で表現されている。従って、対訳文からどのような場合に何を補えばよいかを規則として抽出することが考えられる。しかし、そのような規則を手作業で収集するのは容易ではないので、効率よく収集する方法が望まれる。そこで、自動学習の手法を応用して、対訳文から欠落している要素を推定するための決定木（判定規則と適用順序を体系化したもの）を作成し、それを使って必要な要素を推定する方法を考案した。

旅行会話の対訳データベースを用いた評価実験によれば、話し手、欠落要素のある動詞、尊敬などの待遇表現などを手がかりとした場合、決定木の作成に使用していない文に対して、文の主語では80%以上を正しく補完できるようになった。また、間接目的語等に対しても同様に補完できる。

5. 部分翻訳

現在の音声認識技術ではすべての会話を正しく聞き取るのは難しいため、認識結果には誤りが含まれることがあり、それにどのように対応するかが音声翻訳の成否を決める1つの課題である。しかし、人間同士が会話する場合にも聞き誤ることはあるが、聞き返したり、聞き取れたところを手がかりにして全体を推測したりするなどにより、会話は続けられる。音声翻訳を通して人間が会話する場合でも、聞き返しは人間が自由に行なうことができる。そこで、音声翻訳としては、正しく聞き取れたところだけを翻訳することにより、会話を継続させることを目指した。

「ホテルを予約したいんだけど」と発話された場合、これを正しく認識できれば、例えば「I'd like to reserve the hotel」と翻訳することができる。ここで「ホテル」を「蜚（ホタル）」と聞き間違えた場合、そのまま訳すと、翻訳としては正しくても意味不明になってしまう。しかし、「蜚」が間違いであると判定できれば、その部分を外した「予約したいんだけど」を翻訳し、「I'd like to reserve」を伝えることにより、不完全ではあるが会話を続けることができると考えられる。

音声認識では、ある単語の次にどのような単語が

現れるかを予測しながら認識を行なっている。そのため、単語の認識誤りがあっても、単語のつながり自体は自然なものとなっている。全体を通して初めて不自然であることがわかる。変換主導翻訳では、前述のように、パターンごとに用意された実際の翻訳例のうち最も類似性の高い翻訳例を真似て翻訳する。このパターンは比較的広範囲の表現を眺めていると考えることができる。

すなわち、認識誤りが含まれると、パターンの一部分だけ類似性が低いという現象が現れる。例えば、用例として「部屋を予約したいんだけど」「I'd like to reserve the room」が登録されている場合、「部屋」と「蛭」の類似性は低いが、残りの部分「を予約したいんだけど」の類似性は高い（この例では一致している）。

そこで、この類似性の低い部分を誤りと判定することにした。これにより、「I'd like to reserve...」(...

は誤りがあると判定され翻訳されなかった部分)が出力できるようになった。

表1：変換主導翻訳による翻訳率

	日英	英日	日韓	韓日	日独	日中
翻訳率(文単位)	78.5%	85.0%	94.5%	90.4%	75.5%	(開発中)
翻訳時間(平均)	0.4秒	0.3秒	0.3秒	0.4秒	0.6秒	

6. おわりに

変換主導翻訳は多言語翻訳にも適用可能である。現在、日英、英日のほか、日韓、韓日、日独の翻訳に適用するとともに、日中翻訳の開発を実現した(図1、表1)。現在、翻訳品質の向上を目指し、翻訳例の追加を進めている。

今後は、正しく認識されたと判定された部分だけを翻訳する部分翻訳を一步進め、本当はこのように発話されたはずだと推測する誤り修復を試みる予定である。

参考文献

- [1] 古瀬 蔵, 隅田英一郎, 飯田 仁:
“経験的な知識を活用する変換主導型機械翻訳”,
情報処理学会論文誌, Vol. 35, No. 3, pp. 414-425
(1994)
- [2] 脇田由実, 河井 淳, 飯田 仁:
“意味的類似性を用いた音声認識正解部分の特定法
と正解部分のみを翻訳する音声翻訳手法”, 自然言語
処理, Vol. 5, No. 4, pp. 111-125 (1998)
- [3] 山本和英, 隅田英一郎:
“決定木学習による日本語対話文の格要素省略補
完”, 自然言語処理, Vol. 6, No. 1, pp. 3-28 (1999)

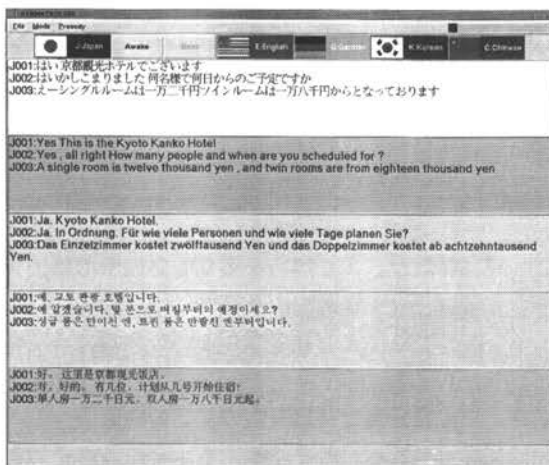


図1：変換主導翻訳による多言語翻訳システム

事務局だより

委員の交替

[市場動向調査委員会]

富士通株式会社

退任 森本 秀樹

就任 伊藤 文子

ソフトウェア事業本部第一ソフトウェア事業部
第四開発部 担当部長

[UPF 会議メンバー] および [インターネット WG]

株式会社 NEC 情報システムズ

退任 檜山 努

就任 濱田 和彦

オープン技術システム事業部
第二技術部 技術担当課長

省略補完技術

NTT コミュニケーション科学研究所 中岩浩巳

1. はじめに

通常、日本語などの自然言語では、相手が容易に推測出来る要素は、冗長な表現となることを避けるために文章上表現せずに、省略される場合が多くあります。しかし、このような自然言語における省略現象の出現傾向は、その言語の種類によって異なります。たとえば、日本語と英語を比較してみますと、日本語では、主語や目的語などの格要素は頻繁に省略される傾向にあるのに対して、英語では、格要素を明示する必要がある場合がほとんどです。例えば、次の新聞記事の日英対訳例文(1)では、

(1) <第1文>NTTは新型交換機を導入する。

→ “NTT will introduce a new model switching system.”

<第2文> (φ₁が) 自己診断機能を搭載、

(φ₂が) 200システムを設置する予定だ。

→ “The new model switching system is equipped with a self checking function and NTT is planning to install 200 systems.”

第2文の前半の主語(φ₁)と後半の主語(φ₂)が省略されていますが、その英訳文では、各々の省略要素に相当する部分それぞれが、“The new model switching system”、“NTT”と明示的に翻訳されています。ですから、機械翻訳においてこのような言語間の省略傾向の違いを克服し、自然で適切な訳文を生成するために、省略された要素を補完する技術が大変重要となります。

さて、この省略補完技術ですが、機械翻訳システム中で実現することを考えると、いくつかの特徴的な機能を備える必要が出てきます。具体的には、機械翻訳では、次の3種類の機能を備えたものが望まれます。

(a) 翻訳する必要がある省略箇所を認定する技術

これは、本当に明示的に翻訳が必要となる省略箇所がどこかを認定する技術です。これによって、訳出が不要な省略箇所に省略補完処理を適用する無駄をなくするとともに、省略補完処理が必要とな

る省略箇所を少なくすることで省略補完処理の失敗による誤訳箇所を少なくすることが出来ます。

(b) 省略箇所に省略された要素を補う技術

これは、(a)で認定された訳出に必要な省略箇所の補完要素を認定する技術です。日英機械翻訳では、特に主語や目的語などの格要素が頻繁に省略されるので、格要素の省略補完技術が特に重要となります。

(c) 補われた要素を適切な表現に翻訳する技術

これは、(b)で補完された要素を、その前後の情報を参考にして、適切な訳語表現に翻訳する技術です。これによって、補完した要素を冗長でない自然な表現で翻訳することができます。

ここでは、日英機械翻訳で特に重要となる格要素の省略補完技術に着目し、以上3種類の要素技術について、機械翻訳システム中で実現するという観点から解説していきしたいと思います。

2. 機械翻訳からみた解析必要な省略箇所の検出^[1]

ここでは、日英機械翻訳システムで省略箇所を検出する際に考慮しなければならない3種類の点について解説します。

2.1 日本語表現と英語表現の発想の違い

機械翻訳で省略箇所を検出する場合には、その省略箇所が、日本語表現と英語表現の発想の違いによる表現のしかたの違いにより補完が不要なものがあります。たとえば、

(2) <第1文>A社は2月1日、ハードディスク装置を発売する。

→ “Company A will put on sale the hard disk device from February 1.”

<第2文> (φがφを) 月4百台生産する。

→ “φが produce 400 unit of φを per month.”

という、文章の第2文では、用言「生産する」の主語と目的語が省略されています。しかし、より自然な表現で、英語に翻訳するためには、この文は、

(3) <第2文> 月産は4百台だ。

→ “Monthly production is 400 units.”

というような表現に書き替えたほうが自然な英訳となるとともに、「生産する」の主語と目的語の省略への補完処理をする必要がなくなります。このような表現の場合には、被翻訳文の構文解析時に、英語的な日本語構造に書き替えたうえで^[2]（原文自動書き替え）、省略箇所を検索することによって、補完処理が必要かどうかの切り分けが可能となります。

2.2 日本語と英語の構文構造の違い

日本語では必須格となる格要素でも、英語に翻訳すると任意格となるような用言が存在します。たとえば、

(4) A（施設）でB（動物）を飼う

→ “At A, B are being kept.”

という表現は、「飼う」の主語（が格）が省略されていますが、日本語で主語がなくても、“A raise B.”という表現で翻訳することが可能です。このような表現で翻訳するためには、翻訳システムの構文解析に用いる格パターンを、英語用言の訳語毎に用意して、日本語解析時に英語の構文構造を決定しておくことが有効です^{[3][4]}。これによって、英語で必須格とならない要素は日本語では必須格と判断しないので、省略箇所の検出が正確に行なえます。

2.3 日本語と英語の態の違い

日本語では省略されている要素でも、より自然な英語に翻訳するためには、態を変換することによって、訳出が不要となる場合があります。例えば、

● 日本文が受動態化されている文

この場合、英語表現としても受動態化することによって、補完が必要な省略箇所を限定できます。

● 日本語では受動態となる表現が、英語では能動態となるような用言からなる文

例えば、

(5) AがB（文書）に掲載される。

という表現は、「AをBに掲載する」の受け身表現であり、「掲載する」の主語（が格）が省略されていますが、英語では、日本語での主語がなくても、“A appear in B”という表現で翻訳するほうが日本語のニュアンスに近い訳が生成できます。このような場合には、翻訳システムの構文解析に用いる格パターンを、英語の用法別に用意して、日本語解析時に、受け身の様相を伴った場合の必須格要素を英語の表現で必須格となる要素のみに制限することによ

って、照応解析が必要となる省略箇所が限定できます。

以上の3種類の条件を用いることによって、訳出が必要のため補完処理の必要な省略箇所が自動認定できます。

3. 日本語格省略要素への補完技術

ここでは、日本語文中で省略された要素に対して、どの要素を補完すればよいか認定する処理である省略補完技術について述べます。本技術は、機械翻訳だけではなく、様々な自然言語処理でも重要な技術であるため、従来から様々な方法が提案されています^{[5]~[9]}。これらの従来の省略補完方法は大きく、助詞や特定の動詞のタイプなどの比較的表層的な情報を用いたタイプと、文の流れや、実際に起こり得る動作の流れをルール化した規則（プランやスクリプト）を導入したタイプに分けられます。機械翻訳への応用を考えたとき、前者は利用する情報が表層的なものに限られているため、解析精度の面で問題があります。例えば、前述した(1)の例では、第1文の主語が助詞「は」によって主題となっており、第2文は、前半の文章の主語と後半の文章の主語が省略されていますが、これらの2つの省略箇所のうち、前者は前文の目的語である「新型交換機」が、後者は前文の主語である「NTT」が補完要素となっています。このように、文章中に主題化された要素が1つだけしか存在しない場合でも、その主題となった要素が省略箇所への補完要素となるとは限りません。このような場合には、主題となっているか否かの情報だけではなく、もっと文の意味の情報を導入することが必要となります。また、従来手法の内、後者は、常識や世界モデルなどの膨大な知識を前提にシステムが構築されているので、幅広い分野のテキストを対象とした実用レベルでの翻訳システムに適用するにはあらかじめ用意する知識の量が非現実的な規模となり、その実現はきわめて困難であります。

ところで指示される側の要素によって分類すると、機械翻訳システムで解析が必要となる省略箇所は次のような3種類に分けられます。

(ア) 補完要素が文章中の他の文に存在する省略箇所（文間照応）：例えば、(1)の例文

(イ) 補完要素が同一文内に存在する省略箇所（文内照応）：例えば、

(6) 彼は方程式を解いて (φが) 答えを出した。

→ “He solved the equation and calculated the answer.”

(ウ) 補完要素が文章中に存在しない省略箇所 (文章外照応) : 例えば、

(7) (φが) 東京に行きたい。

→ “I want to go to Tokyo.”

これら3種類の省略箇所に精度良く補完するためには、個々の省略箇所の種類による特徴を活用した省略補完手法を用いる必要があります。また、これら3種類の省略格要素に補完するための補完ルールは、相互矛盾が起きないように、ルールの適用順序を考慮する必要があります。

ここでは、分類した用言の意味属性^[10]、様相表現、接続表現を用いて、これら3種類の格要素の省略箇所の補完要素を決定する機械翻訳に適した省略補完手法を紹介します。この方法は、機械翻訳システムのように対象分野を限定することが出来ない文章を対象とする場合でも、分類したカテゴリを用いた翻訳対象分野とは独立したルールを記述すればよいので、少ない知識量で省略補完処理が出来ます。

3.1 日本語省略格要素の文間補完技術^[11]

本節では、補完要素が文章中の別の文に現れる省略格要素への補完手法を紹介します。

3.1.1 新聞記事文での省略現象の傾向

例文(1)のような、補完要素が別の文に現れるタイプの格要素省略現象が頻繁に出現する新聞記事文30記事のリード文、102文を対象にして、格要素省略箇所(93件)とその補完要素の出現位置を調査して、このタイプの傾向を分析してみました。その結果、第1文に補完要素が現れる省略格要素の数が全体の86%(80件)、このうち、第2文以降の省略箇所の補完要素が第1文にある例が71件(76%)にものほりました。これは、新聞記事では第1文が記事全体の概要を示す重要な情報を含むことが多いので、その格要素が補完要素となる傾向の強いことを示しています。また、主語が省略されているが、その補完要素が助詞「は」で主題となった格要素や主語ではない場合が15件(16%)もあることから、主題化された格要素から補完要素を選択する手法や、助詞の種類によって補完要素の優先順位を決定する手法だけでは、補完要素を選択するための情報としては不十分であることが分かりました。さらに、この15件の例を詳細に見てみると、補完要素を支配するのは

「結成する」「発売する」「開設する」と言うような、目的語の要素を新しく場面に導入するタイプの用言であり、格要素が省略された文の用言は、「LANだ」のような名詞述語や、「所属する」のような属性を示す用言となっているという傾向がありました。以上から、このようなタイプの省略格要素の補完要素を決定するためには、用言の意味を分類して、省略格要素の補完要素をその用言意味属性の関係から推測することが必要であることが分かりました。

3.1.2 用言意味属性の分類^[10]

3.1.1項で述べましたとおり、従来の手法では対応できない種類の省略格要素も、意味情報を活用することにより正しく補完できることが期待されます。また、実用的翻訳システムで取り扱える規模の意味情報だけでも正確な省略補完処理を可能とするために、用言の持つ意味属性を分類・体系化して、この属性間の関係から、省略格要素の補完要素を決定する手法が提案されています。ここでは、少ない知識量でも正確な省略補完を行うために提案された、用言意味属性の体系を紹介します。

ここで紹介する日本語の用言分類では、文献^[12]の用言分類をもとにして、以下の2つの観点から拡張しております。

(1) 用言が持つ動的特性

用言そのものが持つ意味による分類：用言単体が明示されることにより、文脈情報として、どのような種類の行為がなされ、その発話の結果どのような状況になるかという観点で分類。例えば、

(8) 「持つ」 : <所有>

「開発する」 : <生成>

「持つ」という用言は<所有>の行為を行なうという文脈情報を提示するのに対し、「開発する」という用言は、何かを<生成>する行為を行なうという事実の文脈情報を提示します。

(2) 用言の格に対する関係

用言の支配する格が用言に対してもつ役割による分類：格要素がその用言とともに明示される場合、その格要素が文脈の中でどのような役割を果たすかの観点で分類。例えば、

(9) 「完成する」→N1が 完成する

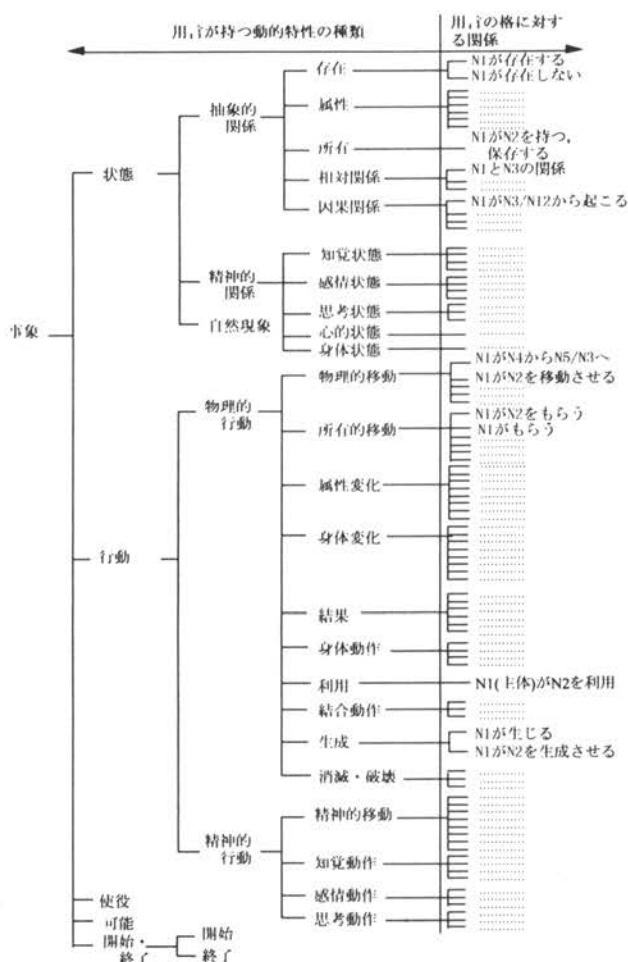
: <N1を生成する>

「開発する」→N1がN2を開発する

: <N1がN2を生成する>

「完成する」、「開発する」とともに生成の行為を示す用言であるが、「完成する」はN1を生成するのに対し、「開発する」はN1がN2を生成することを示します。この基準で分類した用言の概念体系を図1に示します。

図1 用言意味属性体系



3.1.3 省略補完手法

表1 用言意味属性による補完要素決定ルール

省略格要素の条件	補完要素の条件	用言関係	決定
用言意味属性	格	用言意味属性	補完要素
所有	ガ格	所有的移動 & 開始	ヲ格
思考動作	ガ格	所有的移動 & 開始	ガ格
.....	...		

補完要素が省略箇所との文とは別の文中にある文間照応省略格要素の補完処理は、格要素が省略されている文と、そこに補うべき要素がある文との意味的関係を、前述の用言意味属性の対により認定し、省略格要素の補完要素を認定します。このルールの一部を表1に示します。

3.1.1項で述べましたように、新聞記事リード文では第1文が記事全体の内容を概説する文章構造を持っており、第1文内の単位文の格要素がそれ以降の文の省略箇所に補完すべき要素となる場合が多くあります。ここで紹介する手法では、このような単位文を主題単位文と呼んで、この主題単位文の用言意味属性によって、以降の文の省略箇所へ補完すべき要素を選択します。

例えば、(1)の例文では、第一文の解析後、解析結果より主題単位文が抽出されます。この例文では、第一文が単位文で構成されるので、その文の解析結果が主題単位文として文脈情報保持部に保持されます。次に、第二文の解析結果より、「搭載する」と「予定だ」の主語が省略されていることが分かります。同一文内にはそれらの補完要素が存在しないので、主題単位文中の格要素が補完要素候補となります。「搭載する」と「予定だ」の用言意味属性はそれぞれ“所有”、“思考動作”であり、主題単位文の用言の用言意味属性は“所有的移動&開始”であるので、表1のルールにより用言意味関係は「詳細説明」「方針決定」が成立し、主題単位文の目的語、主語がそれぞれの補完要素となります。

前述しましたように、用言意味属性の分類体系の属性値で記述した省略格要素の補完要素の決定ルールを用いることによって、汎用的な少ない知識量で処理が実現できるので、実用的で分野の限定が困難な機械翻訳への応用が可能となります。

3.1.4 評価

本手法をNTTの日英翻訳システムALT-J/E^{[13][14]}(Automatic Language Translator - Japanese-to-English)上に実現して、3.1.1項で用いた新聞記事を対象に、2種類の性能評価を行ないました。まず、照応解析ルールを整備した標本実験では、機械翻訳の訳出に必要な省略格要素の大半(95%)が正しく補完されることが分かりました。次に、前述した実験で使ったルールをそのまま適用した別の標本の実験では、本手法が適用対象とする文間照応省略格

要素については83%の補完要素が正しく認定されました。さらに、若干の不足ルールを追加した後は、その正解率は95%に向上しました。以上の結果、機械翻訳における本手法の効果が実証され、省略された格要素の大半が復元できるという見通しを得ることができました。

3.2 日本語省略格要素の文内補完技術^[15]

3.2.1 照応解析手法

接続語のタイプ、用言意味属性、様相表現のタイプの3種類の制約に着目して、同一文中に補完要素を持つ省略格要素の補完要素を決定する手法について、それぞれの制約の有効性を中心に解説します。

(1) 接続語による制約

接続語は、省略箇所と同じ文内にある補完要素を決定するうえで最も強力な制約となります。これは、接続語のタイプに応じた格の共有に関する制約にもとづくものです。文献^[16]や文献^[17]でも提案されていますとおり、日本語の接続語は、格要素の影響範囲を制限するものがあります。例えば、南は、日本語接続語をA、B、Cの3種類に分類し、主題「は」を伴うハ格も助詞「が」を伴うガ格も共有するような「つつ」や「ながら」のような継続を示す接続語をA類、ハ格は共有するがガ格は共有しないような「ので」や「たら」のような条件を示す接続語をB類、ハ格もガ格も共有しないような「けれど」や「けど」のような接続語C類と呼びました。この分類によるとA類の接続語を伴う複文で、この接続語をはさんだ片方の単位文のガ格が省略されもう片方の単位文にガ格が存在する場合には、この省略格要素の補完要素はもう片方の単位文のガ格の要素となります。このような、接続詞の種類による格要素の特徴を、補完要素の決定に活用することができます。

(2) 用言意味属性による制約

省略格要素又はその補完要素が、埋め込み文又は引用文内にある場合の文内省略補完は、3.1の文間省略補完技術と同様に、埋め込み文又は引用文の表現と文内のその他の表現との意味的關係を認定することが必要となります。この意味解析に、文中の用言の用言意味属性が有効となると期待されます。例えば、

(10) 港区は資本参加すると(φが)言った。

と言う表現では、引用文の用言「参加する」はガ格の属性変化を示し、用言「言う」はガ格の要素が引

用文の内容を精神的移動させるという用言意味属性を持ちます。このような用言意味属性の属性対によって、この文は「助詞ハで主題化された要素が属性変化を起こす」という情報をこのハ格の要素が伝える」という意味をもつことが認定でき、この省略箇所の補完要素が「港区」となると決定できます。

また、(1)で述べた接続語のタイプに曖昧性を持つ接続語(例えば、接続語「て」は、動作の様態を示すA類の意味、時間関係や原因を示すB類の意味、並列表現を示すC類の意味がある)の曖昧性解消に、接続語前後の用言意味属性が有効な場合があります。例えば、曖昧性のある接続語「て」を含む文

(11) 彼は成長して(φが)立派な紳士になった。

では、用言「成長する」用言「なる」の用言意味属性は共に属性変化であることから、属性変化することによって属性変化するという関係を示す文であると認定できるので、原因理由の関係を示すことができます。これにより、用言「なる」のガ格の省略箇所の補完要素を用言「成長する」のハ格である「彼」と決めることができます。

(3) 様相表現による制約

様相表現は、省略格要素の文章外補完要素を決定するうえで最も強力な制約となります。この表層表現による制約は、文内省略補完にも有効となります。例えば、

(12) 彼が天文クラブ員なので(φが)あの星を知っているだろう。

という表現では、接続語「なので」がB類であるので、同一文内でガ格の要素が必ず共有するとはいえないため、接続語のタイプのみでは補完要素が決定できません。しかし、推量を示す様相表現「だろう」が用言「知る」に伴っているので、ガ格は話者や筆者以外の要素が来ると予測できるので、「彼」が補完要素となります。

3.2.2 評価

ここで説明した手法を日英翻訳システムALT-J/E^{[13][14]}上に実現して、日英翻訳システム評価用例文(3718文)^[18]中の補完要素が同一文内にある省略格要素139件を対象に、解析ルールを整備し、解析精度の評価実験を行なった。その結果、ここで述べた3種類の制約条件を用いた場合、それぞれの条件が文内補完処理に有効に働き、対象とした省略格

要素が再現率（補完が必要な省略箇所の内、正しく補完されたものの割合）98%、適合率（提案手法で補完された省略箇所の内、正しく補完されたものの割合）100%の精度で正しく補完できました。本手法を、従来の代表的な手法である Centering アルゴリズム^[6]（再現率74%、適合率89%）と比べても、再現率、適合率共に十分高ことがわかりました。特に、適合率100%と提案手法で補完されたものの全ては正しかったことから、誤補完によって本来の意味と違う訳文が生成されることがなく、本手法が機械翻訳システムでの実現に適することがわかりました。

3.3 日本語省略格要素の文章外補完技術^[19]

3.3.1 照応解析手法

様相表現、用言意味属性、接続語の制約に着目して、文章中に補完要素が明示されていない省略格要素の補完要素を決定し、適切な英語表現に変換する手法について、3種類の制約の有効性を中心に解説します。

（1）様相表現による制約

3.2.1（3）項で述べましたとおり、ある種の様相表現を伴う文では、様相表現が省略格要素の文章外補完要素を決定するうえで最も強力な制約となります。例えば、ガ格が省略されている場合には、様相表現「～したい（ ϕ want to～）」（希望）や「～してほしい（ ϕ want ϕ to～）」（三人称希望・使役）を伴うと、省略格要素は“I”と訳され、様相表現「～してはいけない（ ϕ must not～）」（禁止）や「～するべきだ（ ϕ should）」（義務）を伴うと、省略箇所は“you”と訳されます。このように、文脈により文章内の要素が補完要素となる場合以外で、様相表現を持つ文では、この文中の省略格要素はその様相表現による制約によって一意に補完要素が推定できます。

（2）用言意味属性による制約

用言「もらう（get）」「やる（give）」のような受給表現、「行く（go）」や「来る（come）」のような移動表現等は、格要素が省略されている場合には様相表現が伴わなくても補完要素が決まります。例えば、ガ格が省略された場合には、「もらう（get）」の補完要素は“I”と訳され、「来る（come）」の補完要素は“I”以外（例えば“you”）と訳されます。これら用言は、筆者や話者とガ格となる要素との相対的

関係を暗に示しており、これらの性質により文章外補完要素が推定できます。また、用言の種類のみや様相表現の種類のみでは補完要素が決まらなくても、用言の種類と様相表現の双方から補完要素が決まる場合もあります。

（3）接続語による制約

3.2.1（1）項で述べましたとおり、日本語の接続語には、接続語を挟んだ前後の単位文において、1つの格要素を両方の単位文の格要素として共有することを強制するものがあります。この特徴を利用して、例えば、A類の接続語を伴う複文において、この接続語をはさんだ2種類の単位文のガ格がともに省略されている場合には、様相表現や用言意味属性の制約によって片方のガ格の補完要素が決まった場合、他方の補完要素もそれと同一であると推定できます。

3.3.2 評価

本手法を日英翻訳システム ALT-J/E^{[13][14]}上に実現して、日英翻訳システム評価用例文（3718文）^[18]中に含まれる補完要素が文章中にない省略格要素を対象に本手法の解析精度の評価実験を行ないました。まず、この文章外省略格要素中、出現頻度の高い5種類193件を手で分析し解析ルールを整備し、解析精度の評価実験を行なった結果、前述の4種類の制約条件でその全ての省略格要素の補完要素が正しく決定されました。その結果を詳細に分析すると、用言意味属性は、様相表現と同等以上の効果を持つという結果が得られました。また、比較的簡単なルールだけでも、高い精度で補完要素が決定出来ることが分かりました。さらに、前述の例文中の全ての文章外照応省略格要素371件から自動的に省略補完ルールを作成し解析精度を評価した場合、省略補完ルール作成に活用した省略箇所に応用した場合（ウィンドウテスト）で99.2%、省略補完ルール作成に活用していない別の省略箇所に応用した場合（ブラインドテスト）でも87.6%という高い割合で正しく解析できることが分かりました。

4. 補完要素の翻訳技術^[20]

3.で解説した省略補完技術によって、省略された格要素に補うべき要素が高い精度で認定できることが分かったと思います。しかし、日英翻訳システムでは、訳文が原文の内容を反映して適切に翻訳さ

れているかが最も重要になります。ですから、補完された格要素を英語に変換する場合にも、そのまま翻訳するだけでは、英語側の統語論的・語用論的制約によって、冗長な表現となったり、首尾一貫性が保たれなかったりして、適切な訳語が得られない場合があります。ここでは、実際の省略要素の変換条件を目的言語の言語現象別に分類し、その結果を元に提案された自然な訳語を得るための生成方法を解説します。

4.1 補完格要素訳出法

補完格要素を英語に翻訳する際には、英語側の様々な制約によってそのまま翻訳するだけでは不適切な表現があります。これらの表現は、補完要素の訳出表現別に次のように分類できます。(なお、()内の訳文は補完のみを行い、この補完格要素変換規則を適用しなかった場合の訳文です。)

(1) 訳出不要表現

(1-a) 接続語の種類による制約

特定の接続語を伴う複文では、接続語の前後の格要素が同じ要素となる場合、片方の要素を訳出しない方が自然な訳文となる場合

制約条件：接続関係の種類

- (13) Aは(「Aが」)Bを買収するためCを結成する。

→ “A organize B to acquire C.”

(“A organize B so that A acquire C.”)

- (14) AはBを開発、(「Aが」)販売を始めた。

→ “A developed B and began sales.”

(“A developed B and A began sales.”)

(1-b) 現在分詞表現による制約

動詞の種類により、補完した要素を明示せず、現在分詞として主語の不要な表現で翻訳した方が自然な訳文となる場合。

制約条件：英語動詞の種類

- (15) Aは、(「Aが」)Bを利用してCを開発した。

→ “A developed C, using B.”

(“A used B and A developed C.”)

(2) 代名詞化表現

(2-a) 冗長表現の回避のための制約

省略箇所と補完要素が同一文か接続する文にあらわれる場合に、両方とも文字通りに訳すと、冗長な訳文になってしまう場合。

制約条件：照応要素間の構文構造関係

- (16) A社はBを開発した。

(「A社が」)2月1日から販売を始める。

→ “A Corp. developed B.”

“They will begin sales on February 1 st.”

(“A Corp. will begin sales on February 1 st.”)

(3) 定冠詞付与表現

(3-a) 照応関係明確化のための制約

省略箇所と補完要素が同一文か近い文にあらわれ、両方とも文字通りに訳すと、冗長な訳文となってしまうが、両者の間には他の要素が存在するため、後で出現した要素を代名詞化して訳すと、その代名詞が何を指示するか曖昧になってしまう場合。

制約条件：照応要素間の構文構造関係

- (17) A社はBを開発した。

(「A社が」、「Bを」)2月1日から販売する。

→ “A Corp. developed B.”

“They will put the B on sale from February 1 st.”

(“A Corp. will put B on sale from February 1 st.”)

ここで示しました変換規則は、上記のように、補完要素と被補完要素が含まれる構文木の構造および意味属性によって記述することが出来ます。これらの制約を活用することによって、補完した要素を英語でより自然な表現で翻訳することが可能となります。

5. おわりに

日英機械翻訳での重要な技術である省略補完技術について、特に機械翻訳での実現を考慮に入れた手法を紹介させていただきました。この省略補完を含めた文脈処理の研究は、機械翻訳への応用に限らず近年盛んに研究されており、例えば1999年度のACLの年次大会でも多数の特別セッションが組まれるなど、大変ホットな研究分野です。ですからここで紹介させていただいた技術以外にも多数の研究がなされております。今後はこの文脈処理技術にも、他の自然言語処理と同様に、コーパスなどの言語資産を活用した手法が多数提案されてきて、より一層活発な研究が行われるようになると予想されます。ですから、興味をお持ちの方にとっては、今後もこの分野の研究動向には目が離せないものとなるでしょう。

参考文献

- [1] 中岩、池原：日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析、情報処

- 理学会論文誌、Vol.34、No. 8、pp.1705-1715 (1993).
- [2] 白井、池原、河岡、中村：日英機械翻訳における原文自動書き替え型翻訳方式とその効果、情報処理学会論文誌、Vol.36、No. 1、pp.12-21 (1995).
- [3] 池原、宮崎、横尾：日英機械翻訳のための意味解析用知識とその分解能、情報処理学会論文誌、Vol.34、No. 8、pp.1692-1704 (1993).
- [4] 池原、宮崎、白井、横尾、中岩、小倉、大山、林：日本語語彙大系、岩波書店 (1997).
- [5] M. Kameyama : "A Property-sharing Constraint in Centering", Proc of ACL'86, pp.200-206 (1986).
- [6] M. Walker et. al. : "Centering in Japanese Discourse", Proc. of COLING'90 (1990).
- [7] K. Yoshimoto : "Identifying Zero Pronouns in Japanese Dialogue", Proc. of COLING'88, pp.779-784 (1988).
- [8] 堂坂：語用論的条件の解釈に基づく日本語ゼロ代名詞の指示対象同定、情報処理学会論文誌、Vol.35、No. 5、pp.768-778, (1994).
- [9] 堂坂：日本語省略解析のための手法とその実現法、情報処理学会自然言語処理研究会資料、47-3、(1985).
- [10] 中岩、池原：日英の構文的対応関係に着目した日本語用言意味属性の分類、情報処理学会論文誌、Vol.38、No. 2、pp.215-225 (1997).
- [11] 中岩、池原：日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析、情報処理学会論文誌、Vol.34、No. 8、pp.1705-1715 (1993).
- [12] 高松・西田：動詞パターンと格構造に基づく英日機械翻訳、電子通信学会論文誌、Vol.J69-D、pp.815-822, (1981).
- [13] 池原、宮崎、白井、林：言語における話者の認識と多段翻訳方式、情報処理学会論文誌、Vol. 28、No.12、pp.1269-1279 (1987).
- [14] 中岩、池原：翻訳通信に適した意味解析型翻訳技術、NTT R&D、Vol.44、No. 5、pp.443-448 (1995).
- [15] 中岩、池原：語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析、自然言語処理、Vol. 3、No. 4、pp.49-65 (1996).
- [16] 南：現代日本語の構造、大修館書店 (1974).
- [17] 田窪：統語構造と文脈情報、日本語学、Vol. 6、No. 5、pp.37-47 (1987).
- [18] 池原、白井、小倉：言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成、人工知能学会誌論文、Vol. 9、No. 4、pp.569-579 (1994).
- [19] 中岩、白井、池原：日英機械翻訳における語用論的・意味論的制約を用いたゼロ代名詞の文章外照応解析、情報処理学会論文誌、Vol.38、No.11、pp.2167-2178 (1997).
- [20] 中岩：日英翻訳システムにおける日本語省略補完格要素の訳出方法、1990年人工知能学会全国大会、pp.323-326 (1990).

事務局日より

AAMT メーリングリスト

AAMT事務局は、2月初旬に法人・個人会員、AAMT委員を対象としたメーリングリストを開設いたします。このリストの使用を通し、会員間、会員とAAMTとの連絡・交流を深めていきたいと思っております。メーリングリストメンバーの登録は、e-mail アドレスをお持ちの方が対象となりますので、事務局にまだアドレスをご連絡いただいていない方は、aamt0003@infotokyo.or.jp へお知らせいただきますようよろしくお願いいたします。

利用方法と注意事項については、サービス開始の際にe-mailでお知らせいたします。発信する内容は、委員会活動報告、記事・執筆の推薦者の募集、ヒアリングのお知らせ等事務局からの情報や、リストメンバーから投稿された情報・文章です。メンバーからの投稿は、aamt0003@infotokyo.or.jp で受け付け、AAMT事務局が内容を確認後メンバーへ配信します。AAMT委員・会員の皆さまからの投稿をお待ちしております。

JICST 日英機械翻訳システム（成果公開版）

科学技術振興事業団

1. はじめに

科学技術振興事業団 科学技術情報事業本部（JICST）では、これまで日本で発生した数多くの科学技術情報を、独自に開発した日英機械翻訳システムを駆使して英文データベース化し、世界中に提供してきました。

JICST 日英機械翻訳システム（成果公開版）は、その蓄積技術をパソコン用翻訳ソフト（CD-ROM）として提供するものです。一般用語はもちろん、科学技術用語、医学用語、地名、機関名など約60万語のシステム辞書により、高品質な翻訳を支援いたします。

特に、難しい専門用語が多い科学技術分野では信頼性の高いレスポンスが得られます。

2. 特長

JICST 日英機械翻訳システム（成果公開版）は、Windows95および WindowsNT 上で稼働します。処理速度は、15,000語／時間、辞書数は約60万語で、翻訳処理を行うための各種機能を備えています。

2.1 Windows で簡単に翻訳

メニューバーからマウスでクリックすることにより、ダイレクトに用途に応じた機能を実現しました。翻訳画面に直接文章を入力するほかに、作成したテキスト形式文書を直接読み込むこともできます。

2.2 必要に応じた翻訳処理

文章を一括して翻訳するほかに、一文単位の翻訳や、必要な部分だけの翻訳など、用途によって使い分けることができます。

2.3 見やすい翻訳結果

画面左側に日本語、右側に英語を表示しているので日本語と英語の対応をとりながら翻訳が可能です。翻訳画面中の日本語に対応する訳語を表示することができます。

2.4 約60万語のシステム辞書

一般用語の他に JICST シソーラス、科学技術用語、医学用語、地名、機関名など約60万語のシステム辞書を用意しました。化学、機械、電気、情報等の科学技術分野や、医学分野に対して高品質な翻訳を実現します。

2.5 ユーザー辞書で柔軟な対応

システム辞書とユーザー辞書を組み合わせてよりきめ細やかな翻訳を実現します。人名等の未登録語もユーザー辞書に登録することにより、即座に翻訳結果に反映させることができます。

2.6 文体の指定も可能

目的の文体に近い翻訳結果を簡単に得られます。例えば、主語のない文については主語を設定することや、命令文にするといった文体の変更ができます。

3. 動作環境

Windows95、Windows NT（ver.4.0以降）が正常に動作するパソコン

使用メモリ	32MB 以上
空きディスク	200MB 以上
解像度	800×600以上のディスプレイ推奨
販売価格	定価（税込み）126,000円

（注1）インストール時に CD-ROM ドライブが必要になります。

（注2）ユーザー辞書に大量の単語を登録をする場合には、これ以上のディスク容量が必要になります

平成11年3月末まで、キャンペーン販売価格100,800円（税込み）で提供中です。

お問い合わせ先

科学技術振興事業団科学技術情報事業本部
情報提供部 ヘルプデスク

〒102-0081 東京都千代田区四番町 5-3

TEL: 0120-521022 FAX: 0120-521033

URL: <http://www.jst.go.jp/announce/>

MTCD/MTCD.html

e-mail: helpdesk@mr.jst.go.jp

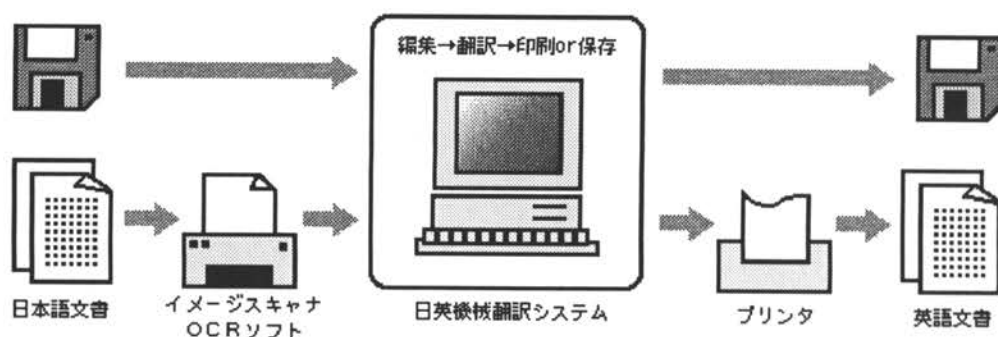


図1 機械翻訳処理の流れ

JST日英機械翻訳システム - Sample		
ファイル(F) 編集(E) 翻訳(T) ユーザー辞書(U) 表示(V)		
Sample		
No.	和文	英文
1 *	1. パソコンを用いた小規模監視制御システム。	1. The small-scale supervisory control system using personal computer.
2	情報処理分野におけるダウンサイジングという大きな流れは上下水道施設の監視制御システムにおいても例外ではない。	Large flow of the downsizing in the information processing field is not the exception in the supervisory control system of the water and sewage facilities.
3	このような状況の中でコストパフォーマンスに優れたパソコンを用いた小規模監視制御システムを構築した。	The small-scale supervisory control system using personal computer which was excellent in the cost-performance in such situation was constructed.
4	パソコンを用いることにより汎用アプリケーションソフトウェアを適用することが可能で、グラフィカルな監視や操作性に優れたシステムとなる。	It is possible to apply the general-purpose application software by using the personal computer, and it becomes a system which is excellent in graphical monitoring and operability.
5	また、スタンドアロンなシステムからNTT公衆回線を用いたシステムまで、ユーザは用途・規模に応じて最適なシステムを選択できる。	And, the user can choose the system which is optimum in proportion to application and scale to the system which used the NTT public line from the stand-alone system.

図2 翻訳画面イメージ

英日翻訳ソフト「EtoJ PRO V5.0」

ロゴヴィスタ株式会社

1. はじめに

翻訳精度を中心に高い評価を頂いている Logovista の英日翻訳ソフト「EtoJ」シリーズをバージョンアップし、1998年10月にバージョン5.0を発売いたしました。

「EtoJ」シリーズには、本格的なプロユース向けの「EtoJ PRO V5.0」とパーソナルユーザ向けの「EtoJ PERSONAL V5.0」などがありますが、ここでは、「EtoJ PRO V5.0」をご紹介します。

今回のバージョンアップでは、プログラムを翻訳エンジンおよび各機能単位にコンポーネント化し、多重処理も可能になりました。

また、翻訳ソフトで最も重要とされる翻訳精度についても翻訳エンジンの改良が図られ、さらに翻訳精度がアップしております。

2. 新機能

2-1. ホームページ翻訳

プロキシ方式によるインターネットホームページの翻訳ができるようになりました。

翻訳結果は、日本語のみ、英語日本語上下対訳のいずれかの形式で、原文レイアウトをそのまま保持し表示されます。

2-2. Word 連携翻訳

MSWord97、MSWord98のメニュー、ツールアイコンにEJ 翻訳機能を組み込み、MSWord 上の指定した範囲の英文を翻訳し、上書き、上下対訳、新規文書ウインドウ表示の3つの形式で翻訳結果を出力することができます。

2-3. 対話型翻訳の拡充

対話型翻訳 (TranslationAssistant) では、翻訳を実行しながら、別訳語、別解釈、品詞設定などの編集作業を並行して操作することができるようになりました。

2-4. ユーザ辞書登録機能の拡充

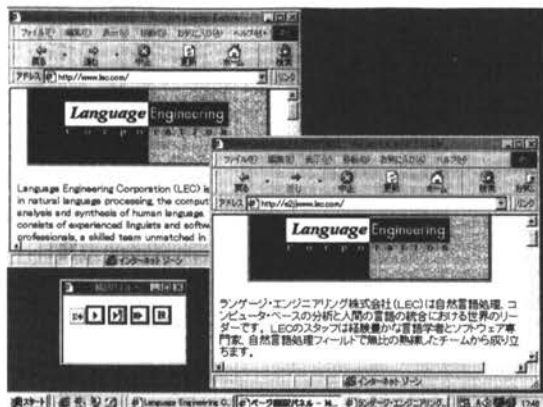
ユーザ辞書の登録が以前の登録内容に加え、詳細な登録ができるようになりました。

特に、動詞に関する登録の種類が増え、例えば「put ~off」などの動詞と前置詞の間に単語が挿入される複合語の登録ができます。

2-5. 後編集機能の拡充

対話型翻訳におけるユーザインタフェースの改良が以下のような点に行われました。

- ・ 翻訳結果の別解釈画面の表示方法を変更し、翻訳結果の相違を比較し易くしました
- ・ 新たに英和文上下対訳表示画面を追加



＜図-1 ホームページ翻訳＞



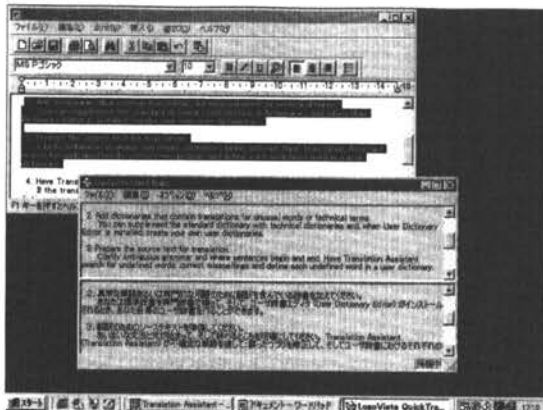
＜図-2 対話型翻訳 (別解釈表示)＞

3. その他の特長

「EtoJ PRO V5.0」は、様々な翻訳ニーズに対応できるよう、4つの翻訳方式がサポートされています。

メインとなる対話型翻訳 (TA) は、英文テキストを読み込み、翻訳、編集、辞書引きなどを対話形式で行えます。

クリップボード翻訳は、各種アプリケーション上のテキストをコピーすることによりクリップボード経由で翻訳を行い、ただちに翻訳結果を独自のウィンドウ上に表示します。

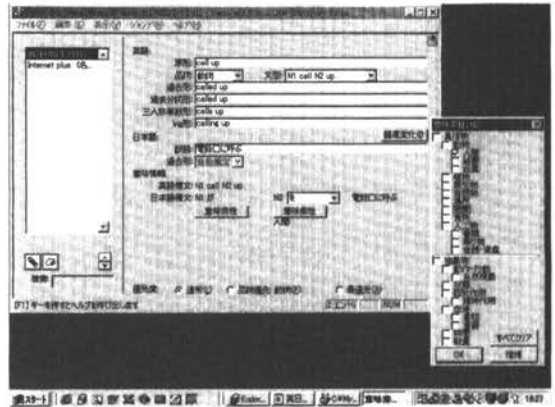


＜図—3 クリップボード翻訳＞

ホームページ翻訳は、ブラウザに表示されたインターネットホームページの英文をページ単位に翻訳します。形式は、日本語のみ、上下対訳を選べます。ファイル一括翻訳は、ファイル単位に1ファイルまたは、複数ファイルの翻訳を一括して行います。複数の文書ファイルをまとめて翻訳したい場合に大変便利な機能です。

さらに、翻訳精度を高めるためのユーザ辞書登録編集機能は、今回動詞、形容詞などで詳細登録ができるよう拡張されました。

また、意味素性情報 (human (人間)、abstract (抽象物) など) が開放され、辞書登録時に意味素性を定義することで、より精度の高い翻訳結果が得られるようになりました。



＜図—4 ユーザ辞書登録＞

ファジーマッチ翻訳は、定型的な英文と和文を事例ファイルに登録しておき、この事例ファイルを使用して、次のような置換、翻訳が行えます。

事例文に完全に一致する場合やよく似ている文を登録された日本語文をベースに数字、日付などを自動的に解析し、置き換えてくれます。

＜例＞

事例ファイル登録文

Thank you for your offer of July20.

7月20日のお申込みありがとうございます。

==>

Thank you for your offer of August 9.

8月9日のお申込みありがとうございます。

4. 動作環境

- ・ OS Windows98、Windows95、WindowsNT V4.0
- ・ 使用メモリ 16MB 以上 (推奨32MB 以上)
- ・ ハードディスク必要量 45MB 以上

5. 価格

- ・ EtoJ PRO V5.0 97,000円 (税別)
- ・ EtoJ PERSONAL V5.0 39,800円 (税別)

6. お問い合わせ先

カテナ株式会社ソフトウェアプロダクツ事業部

TEL 03-3615-4603

FAX 03-3615-3991

製品情報 URL

カテナ <http://www.catena.co.jp/sp/>

ログヴィスタ <http://www.logovista.co.jp>

The 翻訳プロフェッショナル

株式会社 東芝

1. はじめに

東芝は、「The 翻訳」シリーズ最高機種として Microsoft®オフィスやインターネットのホームページに加え、印刷された文書など、様々な翻訳作業を可能にした英日/日英翻訳ソフト「The 翻訳プロフェッショナル V4.0」を発売しました。

The 翻訳プロフェッショナルは OCR ソフトを搭載したことにより、印刷された文書を読み取って翻訳できます。さらに、語句などをマウスで指定するだけで素早く翻訳できる「クイック翻訳機能」や、クリップボードに翻訳したい範囲をコピーするだけで翻訳できる「ワンタッチ翻訳機能」を装備しています。既に翻訳された対訳文書を登録するだけで、次から翻訳に反映できる「翻訳メモリ機能」や、頻繁に使われる文章をテンプレート（穴あき例文）として登録できる「パターン翻訳機能」などの翻訳支援機能も充実しています。

加えて、マイクロソフトオフィス文書やインターネット情報の翻訳のために The 翻訳オフィス V1.0、The 翻訳インターネット V3.0の機能をフル装備し、パーソナルユースから業務まで様々な用途で活用できます。

さらに、アジア太平洋機械翻訳協会（AAMT）制定のユーザ辞書標準仕様形式や CSV 形式（Comma Separated Value Format）のデータを翻訳辞書に登録できる「一括登録機能」や、翻訳辞書へ登録した方が良い単語を既存の文書から自動的に抽出する「辞書構築支援機能」も搭載しており、今までの翻訳資産を有効に活用することができます。

2. 主な機能

（1）レイアウト自動認識機能付き OCR

認識率99.5%*1のOCR（文字認識）ソフト ExpressReaderLE により、印刷された文書を紙面から読み取って翻訳することが可能です。また、レイアウトの自動認識機能を有し、翻訳結果に原稿のレイアウトを反映できます*2。

（2）翻訳資産を活用できる「翻訳メモリ」

既に翻訳されている内容を翻訳資産として活用できます。CSV 形式データや対訳データで保存された文書を「翻訳メモリ機能」で登録し、新しく翻訳する文に適用できます。また既存の翻訳データの応用範囲が広そうな場合、既存のデータを「パターン翻訳機能」を用いて登録すれば、テンプレート（穴あき例文）として有効活用できます。

（3）機械翻訳に重要な辞書作りをサポート

登録すると翻訳結果が良くなりそうな語句を文書から自動的に取り出し、辞書登録をサポートする「辞書構築支援機能」を搭載しています。対訳文書があれば登録する訳語の候補まで自動推定できます*3。

また、表計算ソフト標準の CSV 形式や、AAMT 制定の翻訳ソフト対訳ユーザ辞書標準仕様（UPF: Universal PlatForm）のデータを用いて一括登録することも可能です。

（4）市販辞書が参照できる「EPWING 検索」

言葉の意味を詳しく知りたい時や語句の用例を知りたい時などには標準規格 EPWING Ver2 の電子辞書を参照できます*4。翻訳者の用途に応じた辞書を参照することで効率良く翻訳を行えます。

（5）ちょっと翻訳したい時の「クイック翻訳」

画面上の語句にマウスカーソルを近づけるだけで翻訳できる「クイック翻訳機能」*5や、ドラッグ

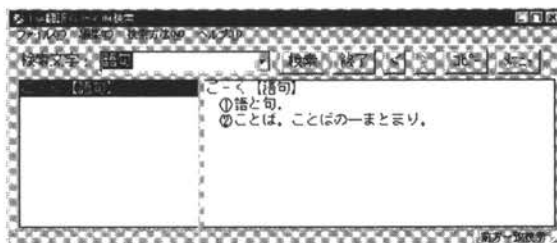


図1：EPWING 辞書検索

した範囲をクリップボードにコピーするだけで翻訳できる「ワンタッチ翻訳機能」で手軽に翻訳を行うことができます。

(6) インターネットやMS オフィスにも対応

ホームページや電子メール*6、ニュース*6、チャット*3の翻訳などが可能です。ホームページは「オートパイロット機能」により指定時刻に自動的に収集し翻訳できます。また、認証機能付きなのでホームページに設定されているパスワードにも自動応答できます。

さらに Word や Excel、PowerPoint®のデータを、そのまま画面上で翻訳可能です。また一太郎の画面からも翻訳できる他、テキストファイルやRTF、HTML形式のファイルも翻訳できます。

(7) その他の機能

原文と訳文を対応付けながら翻訳できる「対話翻訳機能」により、一文ずつ確認しながらA4版換算

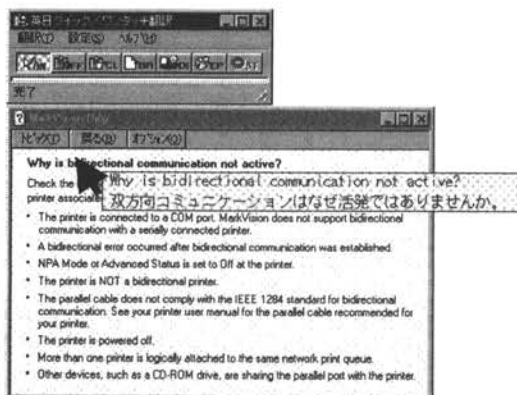


図2：クイック翻訳機能

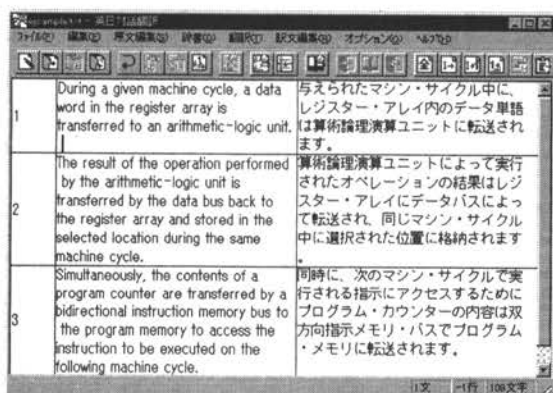


図3：対話翻訳機能

約300ページに及ぶ8,000文を翻訳、編集できます。

また、英日基本辞書24万語と日英基本辞書14万語を標準装備し、他に最大約20万語が登録できるユーザ辞書機能を採用しているので精度の高い翻訳が可能です。さらに、オプションの専門用語辞書を使用すれば、より高度な翻訳結果が得られます。

3. 動作環境 (全て日本語版)

ハードウェア：PC/AT 互換機、PC98シリーズ

CPU：Pentium 以上推奨

DISK：35MB (片方向)、ExpressReaderLE 60MB

MEM：32MB 以上 (WindowsNT の場合40MB 以上)

ソフトウェア：Windows95/98、WindowsNT4.0

連携可能ツール：

WORD95/97/98、Excel95/97、PowerPoint98、

Netscape Navigator3.0-4.05、Internet Explorer3.0

-4.02、Netscape Messenger4.0-4.05、Microsoft

OutLook Express4.0、Netscape Collabra4.0-4.05、

Netscape Conference Text Chat4.0-4.05、Microsoft

NetMeeting2.0-2.01、一太郎8/9

お問い合わせ先

株式会社東芝 コンピュータ・ネットワーク

プロダクト事業部 (03-3457-2725)

製品情報 URL：<http://eiplaza.toshiba.co.jp>

カタログご請求先：<http://eiplaza.toshiba.co.jp/>

pcserver/astransacctl.htm

*1：当社調べ。認識率は原稿の状態で変化します。

*2：一部復元できない情報もあります。

*3：英日・日英双方向が必要

*4：文字データのみ

*5：翻訳できないデータもあります。

*6：英日のみ

・Microsoft、Windows、Windows NT、PowerPoint は米国マイクロソフト社の米国ならびにその他の国における商標です。

・Pentium は Intel Corporation の登録商標です。

・その他の名称は、各社の商標、登録商標の場合があります。

AMTA'98参加報告

水野順子 (Language Engineering Corporation)
井佐原 均 (通信総研)

1. はじめに

平成10年10月28日から3日間、米国ペンシルバニア州ラングホーンにおいて、アメリカ機械翻訳協会により AMTA'98 (Third Conference of the Association for Machine Translation in the Americas) が開催された。今回の会議は Machine Translation and the Information Soup (MT in a growing field of language technologies) というタイトルで行われた。タイトルに「soup」とつけるだけあって、多言語間の翻訳に大きな関心が向けられていた。ただし、多言語といっても、欧米語の範囲を越えるものは例外という雰囲気は否定できない。

会議は、一般発表の他、Featured Speakers として、

The internation business of MT

— Gaston Bastiaens (Lernout & Hauspie)

Integrating MT with other technologies

— Louis Monier (Alta Vista)

Situating MT in the Information (Language Technology) Soup!

— Gary Strong (National Science Foundation)

MT-Related technologies :

Information Retrieval

— Gregory Grefenstette (Xerox Research)

Multilingual Text Summarization

— Jaime Carbonnell (Carnegie Mellon University)

The future of the lexicon :

— Pascale Fung (Hong Kong University of Science and Technology)

— Dan Melamed (University of Pennsylvania)

の7件の講演が、また、パネル討論として、

Breaking the Quality Ceiling

(moderator : David Farwell)

The forgotten Majority : Neglected Languages

(moderator : Laurie Gerber)

The AMTA Seal of Approval

(moderator : Eduard Hovy)

の3件が行われた。

本稿では、会議中に行われた機械翻訳の認証に関するパネル討論に関連する話題に触れたのち、いくつかの発表について触れる。なお本稿は筆者らの聴講した部分についての私見であることをあらかじめお断りしておく。本会議の論文集は

“Machine Translation and the Information Soup”
Lecture Notes in Artificial Intelligence 1529

D. Farwell, L. Gerber and E. Hovy (eds.) Springer
ISBN3-540-65259-0

で入手できる。

2. パネル討論

(1) AMTA SEAL of Approval

機械翻訳であるか否かを AMTA の認証という形で保証できるか、というテーマのパネル討論であった。昨年、ドイツの展示会で、単なる辞書引きツールを、99ドル/言語でいろんな言語の MT があります！と宣伝した会社があるとのことであり、このような会社に自社の製品を機械翻訳だと言わせないために、AMTA で機械翻訳かどうかの認定をしようという話であった。米国の主要な機械翻訳メーカーは、このような会社の存在を重大な問題だと認識しているようである。実際のパネル討論では、何を評価するのか (品質か、方式か) ということが問題となり、立場によって意見が分かれた。時間の制約から十分には討論ができなかったが、とにかく、認証を前向きに考えようということで、まずは AMTA が中心となり、最終的には IAMT での認証としようとの結論となった。

このパネル討論では、事前にパネラー間でいくつ

かの話についてメールで討論が行われていた。そこで話題となったものとして、上記の辞書引きツールの件の他、87年に中間言語方式のMTでアメリカの特許を取った会社があり、この会社の代理人が最近、誰かがホームページ上で公開していたツールに対して特許侵害だから消せと言う手紙を送ったというものがあつた。もちろん、本気で裁判で争えば、覆せる内容のようではあるが、アメリカというのはいろんなことの起こる国ではある。。。

(2) Breaking the Quality Ceiling

機械翻訳の性能の限界を超えるための手段を模索するパネルであつた。アメリカだけの学会で閉じていないで国際的な連携を模索すべき、という意見や、機械翻訳の評価法の確立を目指すべきだとの意見が述べられていた。

3. 一般発表

(1) Machine Translation in Context

(Kurt Godden, General Motors)

機械翻訳のユーザの立場からの発表であつた。自動車部品を比喩に用いて、外側は素晴らしいが機能的に不十分な、“overfeatured”という評価ポイントが面白い。スタイルなどは気にしない、マニュアルを読んで正しく修理できればよいというユーザには、きめ細やかなインタフェースは不要というのは、同種の文書を大量に翻訳する必要があるヘビーユーザからの論であるが、見落としがちな点である。

(2) Can Simultaneous Interpretation Help

Machine Translation (Dan Loehr, Georgetown University)

機械翻訳が、同時通訳者から学び取るべきことは何か、という問題提起である。実際に同時通訳者にインタビューして検討している。個々の単語にこだわらない、完全主義を目指すのは無理と心得よ、ひとかたまりずつで捉える、予測せよ、などが得られている。結局、深い意味解析と文脈解析を通した、逐語訳でない翻訳が望ましいというところに落ち着くのだが。。。機械翻訳には厳しい指摘である。

(3) SYSTRAN on AltaVista: A User Study of Real-Time MT on the Internet (Jin Yang & Elke Lange, SYSTRAN Software Inc.)

インターネット上で無料の翻訳サービスを行い、ユーザの反応を調査した。大量の翻訳テストができ、多くの示唆が得られ、無料だったがペイしたと考えている。翻訳できる言語ペアが決まっているので、いったんある言語に翻訳してから、さらに別の言語に翻訳する形になる場合もあつたが、その不便さを訴えるユーザより、そのようなことができること自体に感激するユーザの方が多かったとのことである。概ね、良好な反応を得たようである。

(4) Spicing Up the Information Soup: MT and the Internet (Steve McLaughlin & Ulrike Schwall, Lernout & Hauspie)

多言語翻訳をサポートすると言っても、通常は2言語間翻訳を複数用意していることを指すことが多い。また、開発する側からすると、ソース言語の解析が最もコストがかかる重要な部分であるので、1つの言語を複数の言語に翻訳する1対多翻訳の方が取り組みやすい。しかし、通常のユーザにとっては、自由に操ることができる言語は1つの母国語である。ここでは、そういったユーザの立場からの多言語翻訳、つまり、元の言語が何語であれ、自分が読みたい1つの言語に翻訳してくれる多対1翻訳を提案している。ユーザの立場に立った、最も望まれる多言語翻訳の形態であろう。今まで、あまり検討されてこなかったのは、その実現が極めて困難なせいかもしれないが、あえてその困難に挑戦する姿勢は評価できよう。

(5) Finding the Right Words: An Analysis of Not-Translated Words in MT (Florence Reeder & Dan Loehr, The Mitre Corporation)

翻訳されなかった語を分析し、それらの多くが、未知語によるものではなく、入力エラーや入力部分の解釈の誤りなどによるものであることを確認した。入力情報に関する知識があれば、知的な前処理を行うことにより多くの未翻訳語を防げることを提言している。ある意味で古い論点ではあるが、電子メールアドレス形式、HTMLのマーキング、多言語間のアルファベット形態の違いなど、入力情報に関し

ての必要な知識が多岐に渡るようになった点は、現在の流れを反映していると言えるだろう。

(6) Demo of Alis Translation Solutions :

Overview (N. Cote, Alis Technologies Inc.)

他社で開発された種々の翻訳システム、および人間による翻訳など、さまざまな選択肢を駆使して、ユーザにとって最適の翻訳環境を構築することができるよう、各種のツールを提供する。昨今のように、選択肢が多岐にわたり、何が必要なのか、どう組み合わせればいいのか、などが必ずしも明瞭でないような状況では、特に新規に機械翻訳システムを使いたいというユーザには便利であろう。具体的にどの程度のことができるのかははっきりしないが、翻訳エンジンは自社では開発せずに、もっぱらそれらを有機的に組み合わせていこうという発想はユニークである。

(7) Logos 8 System Description

(Brigitte Orliac, LOGOS Inc.)

この発表の主旨は、メンテナンスを容易にするために、すべての辞書データを関係データベースで書き直したという点にある。辞書データの処理はすべて終了したが、文法規則に関しては計画はあるものの、まだ関係データベースに書き直すには至っていないとのことであった。5年前から検討を始め、実際にコンバージョンが始まったのが1年前である。非常に大変な作業で、発表者は満足感にあふれていた。ユーザには見えないが、メンテナンスの容易さは、今後ますます複雑になっていく機械翻訳システムにおいては重要な要素であろう。

(8) Improving Translation Quality by Manipulating Sentence Length (Laurie Gerber, SYSTRAN Systems Inc.; Eduard Hovy, Information Sciences Institute, USC)

文型がまったく異なる日英翻訳の場合は特に、文の構造を捉えることは1つの大きな鍵となる。1文中に多くの述語が出現すると構文の把握が困難になるため、長文については事前に分割してよりよい訳文を生成できるようにする。前回までの試行では精度の向上が見られたが、今回の実験(20語以上の文

について分割を試みる、分割後の各チャンクは7語以上であること、「～し(て)」、「～について」などの出現箇所で区切る、等の基準)ではあまり芳しい結果は出なかったとのことであった。にもかかわらず、聴衆は好意的な反応を示していた。目的も方向性も間違っておらず、必要なのはより優れた方策を発見することだからであろう。

本発表を聴いての感想であるが、日本電子工業振興協会が作成中の日英対訳コーパスは、日本政府発行の白書と、その英訳を対象データとしており、英訳に際して、日本語文が分割されている。この分割結果を用いた場合の翻訳実験も面白いと思われる。また、シストランの日英・英日翻訳システムのデモも行われていたが、真面目に作っていることをうかがわせるものであった。

(9) A Comparative Study of Query and Document Translation for Cross-language IR (Douglas Oard, UMIACS, University of Maryland)

多言語間の情報検索についての研究発表であった。情報そのものを1つの言語に翻訳した後に検索するか、情報検索コマンドを翻訳して検索するかの検討を行っている。検索用の質問文は、特に短いと、構文が必ずしも文法的に正しいとは限らないので、あまりいい結果につながらない場合があるが、いずれにせよ、機械翻訳は多言語間の情報検索にも有意義な役割を果たすことは確かであろう。

4. おわりに

アメリカ機械翻訳協会では、AMTA'94、'96、'98と機械翻訳サミットの年々に会議を開いている。今回の会議においても、一般発表、パネル討論のほか、システムのデモンストレーションや、最近学位をとった若手研究者の招待講演を2件行うなど積極的な会議運営を行っている。また、機械翻訳システムの認証について、パネル討論といった公開の場で議論を行うなど、認証自体の可否はともかく、機械翻訳技術の健全な発展および広報のために非常に重要な活動を率先して行っている。アジア太平洋機械翻訳協会の今後の活動にも多くの示唆が得られよう。

協会活動報告

(1998年10月～12月)

運営委員会	10月28日	①シンガポールサミットの準備 ②機械翻訳白書の作成
	11月26日	①シンガポールサミットの準備 ②MT 市場調査への取組み
	12月18日	①第1回臨時理事会および MT Summit VII組織委員会の報告 ②シンガポールサミット企画 ③MT 市場調査および機械翻訳白書
市場動向調査委員会	10月2日	①MT 市場調査の進め方および調査方針
	11月13日	①MT 市場調査アンケートに関する審議
	12月18日	(技術委員会と合同会議) ①MT 市場調査(プロバイダー)の実施状況 ②ユーザー向け MT 市場調査の進め方 ③機械翻訳白書への取組み
技術動向調査委員会	10月5日	①UPF 活動成果の活用方法 ②機械翻訳白書の作成 ③MT サミットへの取組み ④'98年度下期活動予定
	11月13日	①機械翻訳白書の作成 ②UPF 活動
	12月18日	(市場委員会と合同会議) ①MT 市場調査(プロバイダー)の実施状況 ②ユーザー向け MT 市場調査の進め方 ③機械翻訳白書への取組み
編集委員会	12月7日	①AAMT Journal No. 24の反省 ②AAMT Journal No. 25の企画 ③AAMT Journal No. 26以降の新トピック
	10月8日	①ヒアリング(沖電気松平正樹氏) ②今後の活動内容
	11月26日	①JST 提供の対訳データ評価 ②HTML 文書翻訳上の問題点 ③ヒアリング(松下電器川越睦氏)
ネットワーク翻訳研究会	12月22日	①ヒアリング(ATR 柏岡秀紀氏) ②対訳データの評価 ③HTML 文書翻訳上の問題点 ④機械翻訳白書への取組み
インターネット WG	12月10日	①メーリングリストの進め方
MT Summit VIIプログラム委員会	10月30日	①MT Summit VII組織編成案の確認 ②シンガポール出張報告 ③プログラム・展示会の検討
	11月27日	①プログラムの検討
MT Summit VII組織委員会	12月18日	①MT Summit VII 企画書の確認 ②小委員会の設置
AAMT 委員長懇談会	10月8日	①機械翻訳白書の作成 ②シンガポールサミットへの取組み
機械翻訳白書 WG	12月4日	①機械翻訳白書の企画・作成方針 ②編集・発行スケジュール ③実行組織の設置と推進担当者

AAMT
ジャーナル

No. 25
(Jan. 1999)

発行 アジア太平洋機械翻訳協会
所在地 〒105-0011 東京都港区芝公園3-5-12 芝公園真田ビル
TEL : 03-5473-7135 FAX : 03-5473-0569
E-mail : aamt 0001@infotokyo.or.jp
ホームページ : <http://www.jeida.or.jp/aamt/>
編集委員会 野村浩郷(委員長) 大倉清司 奥西稔幸 山端 潔
事務局 中島泰雄 鈴木 博 木谷 恵
印刷所 伸光写植印刷株式会社

