



AAMT

Asia-Pacific Association for Machine Translation

Journal

No.32

March 2002

アジア太平洋機械翻訳協会

目 次

論文：	究極の翻訳方式の実現に向けて……………池原 悟 ……………	1
	リーズ大学における MT 活用法……………トニー・ハートレ……………	8
	機械翻訳の科学に向けて……………フランシス・ボンド……………	12
実験サービス紹介：	NTT「多言語情報アクセス技術」……………	22
新製品紹介：	東芝「The 翻訳プロフェッショナル V7.0」……………	26
	富士通「ATLAS V 8」……………	29
	〃 「ATLAS 翻訳パーソナル2002」……………	32
MT リンク集：	……………	24
事務局だより：	協会活動報告（2001年11月～2002年3月）……………	11
	第12回総会開催のご案内……………	21
	会員による著書のご紹介	
	「ブロードバンドで学ぶ英語」……………ジョナサン・ルイス……………	31
	AAMT 入会のご案内……………	34
	AAMT 入会のご案内（英文）……………	36
	編集委員長から……………井佐原 均……………	37

CONTENTS

Technical Essay：	“Towards the Realization of an Analogical Mapping……………S. Ikehara ……………	1
	Method for MT based on Semantic Typology”	
	“MT at Leeds Univ., UK” (English)……………T. Hartley……………	8
	“Toward a Science of Machine Translating” (Eng.)……………F. Bond……………	12
Trial Service：	“Multilingual Information Access” NTT Cyber Space Laboratories……………	22
New Products：	Toshiba “The Honyaku Professional V 7.0”……………	26
	Fujitsu “ATLAS V 8”……………	29
	〃 “ATLAS Honyaku Personal 2002”……………	32
MT Web Links：	……………	24
AAMT News：	AAMT Activities (Nov.,2001~Mar.2002)……………	11
	The 12 th AAMT General Meeting……………	21
	Publication by AAMT Members…………“Broadband English”……………J. Lewis……………	31
	Guide to AAMT Membership (Japanese)……………	34
	Guide to AAMT Membership (English)……………	36
	Message from the new chair of the Editorial Committee H. Isahara……………	37

究極の翻訳方式の実現に向けて

—— 類推思考の原理に基づく翻訳方式 ——

鳥取大学工学部知能情報工学科 池原 悟

1. はじめに

最近、計算機が大幅に安くなったこと、インターネットの発達によって、海外との情報流通量が格段に増加したことにより、機械翻訳システムへの需要が急増し、すでに膨大な数のシステムが使用されるようになりました。しかし、日英両言語のように言語族の異なる言語間の翻訳は難しく、依然として満足できる品質を得るのは困難な状況です。従来の方式技術は、すでに限界に近いレベルに達していると思われる、今後はパラダイムシフトを伴うような新技術の実現が期待されています。このような現状に対して、人工知能の分野からは、「言語を人間及び人間社会に深く拘わる現象として根底から問い直すべきだ」と言う指摘もあります。

私たちは、今まで意味解析型翻訳方式の研究を行ってきましたが、このたび、科学技術振興事業団の支援を頂いて、「セマンティック・タイポロジーによる言語の等価的変換と生成技術」の研究を開始することになりました。この研究は、「多段翻訳方式 (ALT-J/E)」で提案された「構造的な意味の単位」の概念を意味類型論 (セマンティック・タイポロジー) の観点から発展させるとともに、「等価的類推思考の原理」を言語に適用したもので、究極の翻訳方式への接近をねらっております。

本稿では、上記の意味解析型翻訳研究の達成点と問題点について触れたのち、原点に立ち返り、新研究の目指すところをご紹介します。

2. 意味解析型翻訳研究の達成点

機械翻訳で最大の問題は、言語の持つ表現と意味の豊かさに起因して、解析の過程で、多くの文法的、意味的曖昧さが発生することです。この問題を解決するため、すでに多くの試みが行われてきました。

例えば、原文入力に先だって、曖昧性の発生を抑ええるようにするための「前編集」や原文執筆の段

階で一定の制約に従って文章を書いてもらう「制限言語」の方法が提案されましたが、機械翻訳の簡便性や自然言語の自由度を損なう点が問題でした。

また、解析の段階で曖昧性の発生を抑止、もしくは解消する方法としては、用例翻訳、パターン翻訳、知識ベース翻訳の方法などが提案されています。最近では、音声処理での成功を受けて、隠れマルコフモデルを用いた統計翻訳の試みも行われています。しかし、いずれの方式も一長一短があり、なかなか満足する結果を得るに至っておりません。

ところで、従来実現された翻訳システムをみると、いずれもトランスファー方式が基本となっています。この方式は、構文解析結果を用いて原言語の表現構造を目的言語の構造に変換し、その後、原言語に使用された語を目的言語の訳語に置き換えることを基本としています。すなわち、文の構造は意味と独立に扱われ、文全体の意味は、個々の単語単位に翻訳された後、合成されることから、この方式は、意味の線形性を仮定した要素合成法 (重ね合わせの原理) を基本とする方法だと言えます。

自然言語は慣用表現の固まりだと言う翻訳家もおられるくらいです。現実の言語表現では、意味の線形性は保証されません。文を単語に分解する過程でしばしば全体の意味が失われます。言語表現の意味の非線形性は、言語の本質的な特徴でもあり、要素合成法の仮定は、従来の翻訳方式の最大の問題と言えます。

これに対して、私たちは、すでに表現の構造と意味の関係を考慮した方法として「多段翻訳方式」を提案しました。この方式は言語間の発想の違いを吸収することを狙っており、話者の概念化された認識を表す「客体的表現」と概念化されない感情や意志を表す「主体的表現」を別々に変換するパスを持っています。客体的表現の翻訳では、動詞と名詞の結合した基本的な言語表現に対して、構造的な意味の単位を設け、表現構造の持つ意味を失わないような翻訳を試みており、要素合成法の限界に挑戦した方

法と行うことができます。

ところで、この方式を実現するには、対象認識の形式に関する精密で膨大な規模の言語知識ベースを開発する必要があります。そこで第1歩として、単文レベルの表現を対象に言語知識ベースを開発しました。これは単語の意味的用法を3,000種の意味属性にまとめ、日本語動詞6,000語と名詞30万語からなる単文構造とその意味を14,000種類の日英翻訳規則として記述したものです。

この知識ベースを用いた翻訳実験では、「IPAL 計算機用日本語基本動詞辞書」の全例文約5,000文に対して90%の精度で訳語が正しく決定されたことが報告されています。また、誤りの原因の大半は、形態素解析失敗など、方式上の問題ではないことから見て、単文レベルでの意味解析の問題は、ほぼ解決できる見通しです。

しかし、この方法は以下の問題を持っています。一つは、原言語の表現構造に対して固定的に目的言語の単一の表現構造が対応づけられているため、文脈上、必ずしも適切な表現に訳出されるとは限らないことです。もう一つは、複文（埋め込み節を持つ文）や重文（接続構造を持つ文）レベルの非線形性が扱えないことです。

本研究では、これらの問題に挑戦しようとしているのですが、いずれも複雑で困難が予想されます。そこで次章では、言語の本質を再考する立場から、これらの問題にどのようにアプローチしようとしているかを述べます。

3. 言語の個別性と表現の意味類型化

3. 1 対象認識と言語の関係

言語翻訳において問題となるのは表現と内容の関係ですが、言語によってこの関係は様々です。このような言語の違いはどこから生まれるのでしょうか。

（1）対象認識と言語の個別性

言語処理をやっていると忘れてしまいがちですが、言語は、社会慣習の一つです。言語は話者の認識を表現する手段であり、認識は民族によって様々です。それぞれの民族の有する自然観および社会観の違いは慣習の違いを生み、言語の個別的な発展の基礎となっています。

例えば、緑なす平原に住む民族では、自然との一

体性を基調とする多神教の自然観が生まれ、諸物の中に調和して存在する対象の姿を描写するような表現や実体より属性をクローズアップするような表現が発達しやすいこと、また逆に、砂漠に住む民族では、人間と自然との戦いを基調とする一神教の自然観が生まれ、主体と客体の区別を明確にし、それらの関係を立体的に表すための表現が発達することが指摘されています。語彙の発達の点でも、自然環境による大きな違いが知られています。

このような民族による自然観と社会構造の違いは言語発生以前から営々と蓄積され、世代から世代へと受け継がれており、それらの構造の上に発達した言語は、必然的に個別的とならざるを得ません。

ところで、言語は、いったん、それが社会慣習として定着すると、対象認識の枠組みとしても使用され、それを母国語とする話者の対象認識の形成能力を豊かにすると同時に、対象認識の制約条件としても働くようになります。

このように、異なった認識を表現する手段として成長してきた言語は、再び、それを使用する人の認識の制約条件として働きますから、認識は言語集団において2重の意味で個別的なものとなります。

（2）対象認識の共通性と言語表現の保守性

前項の話と矛盾するようですが、どの民族も同じ地球の重力の下で生活する人間です。異なる地域の異なる環境であっても、最低限、衣食住によって生活を営む存在ですから、異なる言語間においても、対象の認識の仕方には、基本的な部分で多くの共通点が存在します。

この認識の共通性は、民族が文化的な共通基盤を持つほど大きくなります。例えば、漢字文化圏である中国と日本とは同時に仏教文化圏でもあり、この文化圏では、社会制度にも歴史的に近いものが存在しました。また、近代以降、西洋科学技術が普遍化し、異なる言語集団でも、物質の対象世界に対して、共通する認識を持つようになってきました。社会に対する認識においても、文化的交流の進展によって、言語間で共有する認識の範囲は増大しています。

ところで、このような言語間での対象認識の共通化は、必ずしも言語間での表現構造の共通化を意味しません。言語の表現構造は、文化交流による対象認識の共通化に追従できるほど柔軟ではありません。

言語は常に矛盾を内包した過渡的な存在で、時と

共に変化します。確かに、語彙レベルでは、造語や外来語の導入によって、概念化された新しい認識を比較的容易に取り込むことができます。しかし、表現構造の場合は、部分的な変更であっても多くの場合、体系そのものに矛盾を発生させます。

それは、言語が多次元的な対象の認識を一次元的な文字の列、または、空気の粗密波で表現せざるを得ないという宿命をもつからです。言語は、立体的構造を持った対象認識を一次元的な文字列上に模写する方法として、それぞれ独自の表現体系を発達させてきました。そのため、そこに新しい概念に伴う新しい表現構造をそのまま輸入しようすると、せっかく構築された枠組みに矛盾が生じてしまいます。

このように、時代と共に言語間で認識の共通化が進むのに対して、それを表すための表現構造は頑固で保守的です。これは機械翻訳方式を考える際の第1の検討条件といえます。本研究では、後に述べるように、この性質に着目して、共通的な概念を介した表現間の意味的対応付けの方法を考えます。

3. 2 言語表現の意味的非線形性

ここで、言語の非線形性の話に戻ります。言語は、他の自然現象や社会現象と同様、複雑系としての特徴を持っています。

その第1の点は、自己組織化された表現規則の存在です。言語規範が自然発生的な社会規範の一つであると言うことは、それが人間同士の相互作用を介して自己組織化されたことを意味します。現存する言語族間での表現構造の大きな違いは、発生段階での小さな違いから成長してきたものと考えられますが、これも複雑系の特徴の一つです。

第2は、すでに触れましたが、表現と意味の関係の非線形性です。これは、言語が非平衡な開放系として進化の途上にあることと関係しています。言語は、常に既存の枠組みで表現できないような新しい経験と認識にさらされています。そのような場合、比喩や類推の仕組みなどが使用されますが、同じ比喩や類推の表現が繰り返されると、それが慣習となり、新しい表現規則が成長します。このような規則は、既存の枠組みと調和しにくい場合が多いため、言語表現の意味的な非線形性は、拡大再生産されることになります。

このような表現の意味に関する非線形性は言語の本質で、これを扱うための枠組みを考案することが

新しい翻訳方式に課せられた第2の検討条件と言えます。本研究では、次節で述べるように、この問題に対して、非線形な表現を「構造的な意味の単位」として取り出し、意味類型化する方法を考えます。

3. 3 言語表現の「意味類型」

本節では、前節で示した2つの条件に適した方式を実現するための「意味類型」なる概念について説明します。

(1) 言語表現の形式と意味類型

各言語の語によって表される概念は、単なる名辞ではなく、母国語話者が対象を把握するための網の目だと言われています。これに着目すると、翻訳は、原言語の網の目（概念）で掘り取って表現された話者の認識を目的言語の網の目で掘り直し、改めて表現することだと言えます。

従って、翻訳が可能であるためには、対象とする言語間で網の目を共有するか、もしくは類似の網の目が存在することが必要となります。網の目が共通しない場合は、より低位の概念を組み合わせることによって対象概念を説明することになります。たとえば、民族固有の風俗、風習、しきたりなどは、これに相当します。これらの説明を翻訳と言うかどうかは人にもよると思いますが、ここでは翻訳不能とし、翻訳の対象外とします。

ところで、網の目となる「概念」は、一般に、対象を抽象化して得られるものです。特殊性と普遍性の統一体として、それ以上分割できない認識の単位だと言われています。しかし、その内部は必ずしも単純ではなく、様々な構造を持つことがわかります。

例えば、言語表現において、「山」や「川」のような実体概念は、そのまま、それを表象する単語に対応づけられるのに対して、「因果関係」のような関係概念は、原因と結果になる事象の関係が様々な形式を持った表現に対応づけられます。また、「比較」の概念では、「誰が」、「何を」、「何と」、「どんな点で」比較するかなどの要素がそれぞれの言語の表現に対応づけられます。

自然言語において、このような概念の表現に使用される形式は多彩です。どの言語にも多数の形式があり、認識の微妙な違いによってそれを使い分けています。そこで以下では、概念に対応づけるための表現の形式を有田潤に習って、「意味類型」と呼ぶことにします。

有田は、「意味類型は具体的言語表現の一段奥にある思考形式のごときものであるが、その性質上各自の母語をもってこれを考え理解し、また整理することができる」とし、さらに次のように言っています。

「もし諸言語に共通のものがあって、それを取り出すのが普遍的な言語理論の仕事であるとするなら、意味類型の考えかたこそ（名称はこれだけでなくよい）その可能性を開く唯一のものであろうと思う」。「自動翻訳機がもし本当に自然言語を翻訳しうるようになれば、その原理もおそらくここに求められるに違いない。」

（２）非線形な表現要素としての意味類型

「意味類型」は、言語表現をその表す概念を変えない範囲で抽象化したものです。言語表現から、線形要素（置き換え可能な要素）を捨て、非線形な部分を取り出したものと見ることができます。ここでは、文法的、意味的な制約付きの変数と字面からなるパターンで記述することにします。「意味類型パターン」のもっとも典型的な例は慣用表現です。ただし、慣用表現では、部分的な語句を置き換えた場合でも表される概念は変わってしまうため、表現全体をそのまま「意味類型パターン」とする必要があるります。

ところで、このような「意味類型」は言語に個別的で、日本語には日本語の「意味類型」があり、英語には、英語の「意味類型」が存在しますが、これも言語規範と言えます。一般に、表現は無限に存在しますが、人間の記憶能力からみて、それを生成するための言語規範の数は有限と考えられますから、「意味類型」の数も有限と考えられます。一般に、異なる言語間の個別的で多様な表現を意味的に対応づけることは、工学的に困難と考えられますが、有限の数の「意味類型」を対応づけることは、可能性があるとして期待されます。

4. 言語表現の意味的等価変換方式

前章では、言語表現の意味類型化に至る考え方について述べました。本章では、「等価的類推思考の原理」を応用した「言語の意味的等価変換」の概要について説明します。

4. 1 意味的等価変換方式の原理

（１）等価的類推思考の原理

かつて、人工知能の分野では、人間の知能の本質は、汎用的な思考規則とそれを用いた推論能力にあると考えられていましたが、1980年代以降、類推能力の役割が注目されるようになってきました。類推は、あるものを他の似たものに例えることによって思考する形式を持ちます。一般的な規則を用いない点に特徴があり、「覚える」、「思い出す」、「一般化する」、「特殊化する」等の機能が有機的に結合したものと言われている。

ところで、市川亀久彌は、「独創的研究」の中で、「人間の独創性は、類推思考の能力にある」として、「等価的類推思考の原理」を提案しました。これは創造的問題解決プロセスのモデルと言えるもので、式(1)に示すように、「異なる系での異なる現象が、ある一定の条件Cの下で共通点 ϵ を持つこと」を前提としています。

$$C(A_{\alpha} \overset{\epsilon}{=} B_{\beta}) \quad (1)$$

但し、「C」は条件、「 ϵ 」は共通見地、「 A_{α} 」は系 α 上の事象 A_{α} 、「 B_{β} 」系 β 上の事象 B_{β} を表します。

ここで、類推とは、ある系 α 上の事象 A_{α} （ソース）が与えられたとき、条件Cの下で ϵ なる共通性を持った事象系 β 上の事象 B_{β} （ターゲット）を思い浮かべることです。

ところで、翻訳では、原言語の表現の意味を理解したら、それを表すのに適した目的言語の表現を思い浮かべ、もっとも適切な表現を選ぶことが行われますから、翻訳も類推であり、式(1)を前提としていることが分かります。

式(1)を言語に適用した場合、その意味は以下の通りとなります。すなわち、 A_{α} 、 B_{β} を言語 α 上の表現、言語 β 上の表現とし、条件Cによって、それらの持つ意味に着目すると、両表現間に同一な、もしくは類似な意味 ϵ が存在することです。言い換えれば、言語 α 上の表現 A_{α} が言語 β の表現 B_{β} に翻訳できるには、言語 β の表現の中に、同じ ϵ の概念を表すような表現 B_{β} が存在しなければならないことになります。これは、異なる言語間での翻訳が成り立つための条件で、前章で述べた考えと同じです。

(2) 意味的等価変換方式の原理

すでに述べたように、異なる言語間で、無限大の数と言え表現を直接、意味的に対応づけることは工学的に無理です。そこで、これを有限の数の対応関係に縮退させる方法として、前章で定義した「意味類型」を使用します。本方式の原理を式(2)に示します。

$$A_{\alpha} \Rightarrow C (A_{\alpha}) \Rightarrow \epsilon \Rightarrow C (B_{\beta}) \Rightarrow B_{\beta} \quad (2)$$

式中の「 \Rightarrow 」は写像、「 ϵ 」は「論理的意味範疇」すなわち、「概念（真理項）」の集合、「 C 」は言語表現を意味類型化するための関数を表します。

$\alpha \neq \beta$ の場合は翻訳方式となり、 $\alpha = \beta$ の場合は同一言語内での言い換え方式となります。また、この式は「意味類型」間での変換、すなわち言語表現の非線形部分の変換を示していますが、線形な言語表現への適用することも何ら差し支えありません。

処理ブロック図を図1に示します。図1では、非線形な言語の表現構造を対象とした「意味類型知識ベース」による変換パスと併せて、従来の語彙変換

と同様の方法による線形要素（代入可能な要素として変数で表現される）の変換のパスを示しました。

(3)「意味類型」と「論理的意味範疇」について

上記の等価変換方式で、異なる言語の「意味類型」の写像を媒介するのは「論理的意味範疇」の仕組みです。「論理的意味範疇」は、それぞれの「意味類型」によって表現される「概念」の集合です。一つの「意味類型」が複数の個別的な概念を組み合わせた複合的な概念に対応する場合がありますので、ここでは、概念の真偽値によって複合的な概念を表現することを考えて、個々の概念を「真理項」と呼びます。

例を図2に示します。この図は、「比較」の概念を表す意味類型パターンの一部を示したもので、真理項は3段階に階層化されています。意味類型パターン間の写像では、このような原言語と目的言語の「意味類型」を収録した「意味類型知識ベース」を使用します。

4. 2 意味類型間の写像の方法

(1) 言語間のパターンの対応関係

原言語の「意味類型パターン」とそれに意味的に

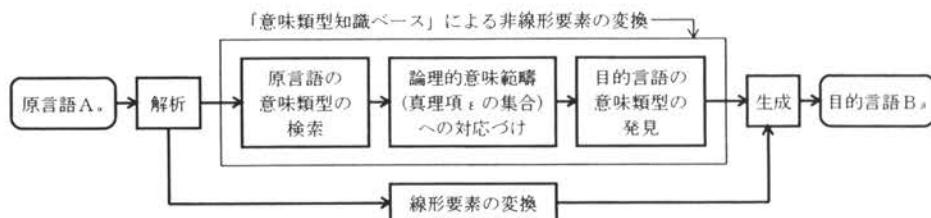
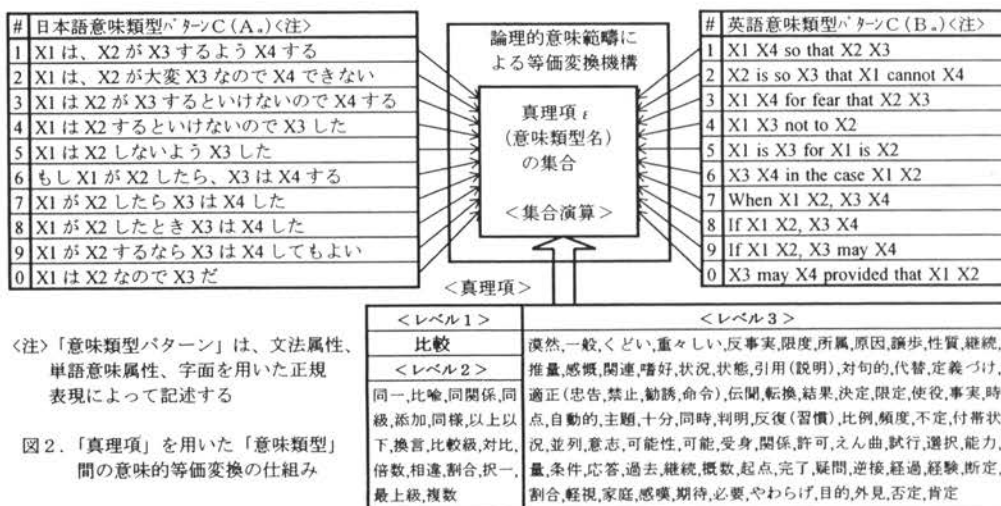


図1. 言語の等価変換方式の構成



対応する目的言語の「意味類型パターン」は、あらかじめ両言語に共通の「論理的意味範疇」の「真理項」毎に集められ、グループ化されます。これによって、両言語の「意味類型パターン」は、意味的に同等なグループ毎に対応づけられます。但し、複合的な概念に対応する「意味類型パターン」は、必要に応じて複数のグループのメンバーとします。

以上により、翻訳では、原文解析によって得られた原言語の「意味類型パターン」に対応する複数の目的言語のパターンの中から、最も適切なパターンが選択されます。次に、その方法を述べます。

(2) 静的選択と動的選択

最適なパターンは原則として動的に選択します。そのため、意味類型知識データベースには、パターンを選択するための文脈上の条件を記載し、翻訳実行時に得られた情報によって、適切なパターンが選択できるようにします。しかし、動的に決定できない時は、あらかじめ決定しておいたデフォルトパターンを選択します。

動的選択で使用する情報は以下のようなものです。

①文内情報：副詞により補足された意味、形容詞修飾による名詞の意味制約など、パターン要素に含まれない文要素の情報や疑問文、命令文、否定文、受動態文など文型に関する情報

②文間情報：文脈による代名詞の意味や接続詞で結合された文の情報など

③文外情報：フォーマル／インフォーマルの区別、書き言葉、話し言葉の区別、場面情報など

これらの情報は、従来の機械翻訳でゼロ代名詞の補完処理で使用されている情報とほぼ同様です。

4. 3 複文・重文翻訳への適用例

日英機械翻訳では、「日本語語彙大系」を用いることにより、単文レベルでの訳文品質は大幅に向上してきたのに対して、複文と重文の翻訳精度の向上は、緊急の課題です。そこで、本方式を複文と重文の翻訳に適用する例を図3に示します。

図3では、入力された日本語は、非線形要素を取り出すため、「意味類型パターン辞書」と照合され、文中で使用された「意味類型パターン」が抽出されます。その際、パターンに使用された変数に対応する日本文中の表現(「単語」、「句」、「節」)は線形要素リストとして保存されます。

次に、得られたパターンは、「論理的意味範疇」

による等価変換機構を通じて英語の「意味類型パターン」に写像されます。最後に、英語の「意味類型パターン」から英文が生成されます。このとき、線形要素リストに保存されていた日本語表現は、既存の翻訳機構によって翻訳され、英語「意味類型パターン」の該当する変数の値として使用されます。なお、複数の意味類型パターンを組み合わせた表現の場合は、より大きな構造から順に、本方式が繰り返して適用されます。

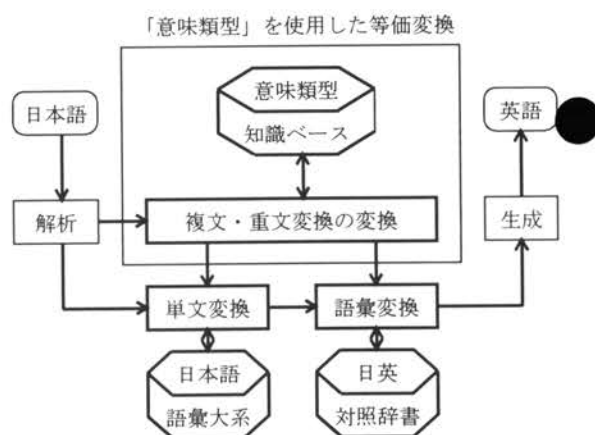


図3. 言語の意味的等価変換システム

5. 新方式の特徴と限界

5. 1 新方式の特徴と有効性

(1) 意味処理基本方式としての意義

従来、「計算機が、与えられた表現の意味を理解したかどうか」を検証する方法として、「意味を変えないで言い換えができるか」、もしくは、「与えられた複数表現の意味的同等性が判断できるか」のテストを用いる方法が考えられています。

これに対して、異なる言語間の翻訳機能と同一言語内での言い換え機能の実現を目指した本方式は、まさに、これらの質問に答えるためのものですから意味処理の基本技術といえます。

(2) 柔軟で適切な翻訳の可能性

厳密に言って、100%正確な翻訳は存在しません。翻訳は近似であり、近似精度をいかに上げるかが問題です。この点から見ると、本方式は、原言語と目的言語の概念把握の共通性の程度に応じた近似度での翻訳を可能としています。

ところで、一般に、原言語の意味に相当する意味を持つ目的言語の表現は複数存在します。従来の翻訳方式では、同一の表現は同一の変換パスを通り、同一の翻訳結果を生成するのが基本でしたが、本方式は、目的言語の「意味類型」の中から、文脈に応じて適切な表現形式を選択することを基本としており、柔軟な翻訳方式と言えます。

（３）従来の翻訳方式との違い

従来のトランスファー方式が、言語表現の意味の線形性を仮定した方法を基本としているのに対して、本方式は、意味的に非線形な表現の形式を「意味類型」として抽出し、それを目的言語に対応づける（Mapping であり、Transfer ではない）ことを基本としています。すなわち、要素合成法の限界に挑戦しています。

また、本方式は、言語間に「論理的意味範疇」を設ける点で、一見して、従来のインターリング方式と類似しています。しかし、インターリング方式では、言語共通の意味を表現するためのピボットを設け、原言語表現の意味のすべてをピボットを介して目的言語に翻訳しようとしていたのに対して、本方式の「論理的意味範疇」は、類型化された概念レベルで両言語の表現構造を対応づけるためのものであり、表現の持つ詳細な情報を変換するためのパスは別に存在します。

5. 2 新方式の限界について

本方式は、原理的に「意味類型」と「類推思考」の考えから成り立っています。単純に言えば、翻訳は、原言語の表現と目的言語の表現を意味的に対応づけることです。しかし、計算機の場合、原言語、目的言語共に無限に存在する表現を意味的に対応づけるのは工学的に困難です。そこで、本方式では、言語表現を意味類型化することによって、無限の対象を有限な領域に持ち込むことと、類推原理によって意味類型間の対応づけることで工学的な解決を図っています。

このような工学的調和の方法には必然的に限界が生じます。本方式の場合、あらかじめ言語表現に使用される対象認識の形式を網羅的に収集し体系化することが必要です。従って、適用対象は言語規範に直接媒介された表現、すなわち「直接的言語表現」に限定されます。比喩（直喩を除く）などの間接的な言語表現への適用はできません。

この点からみると、まだ究極の翻訳方式だと言うわけにはいきませんが、比喩理解の問題も、「類推思考」考えが手がかりとなりそうです。すなわち、比喩が認知される枠組みは、「例えるもの」、「例えられるもの」、「例えの根拠」の3つの基本要素から構成されますが、これらの要素の間には、「類推思考の前提」として4章で示した(1)式と同じ関係が成り立ちます。この場合、比喩理解は、与えられた同一言語の表現 A_α と B_β に対して、共通見地 ϵ を求める問題となります。

本方式で使用する「意味類型知識ベース」は、表現と意味に関する言語規範を体系化したものであるのに対して、比喩理解では、表現対象そのものに関する世界知識を必要とします。しかし、比喩で使用される世界知識は比較的限定されたものかもしれません。いずれにしても、そこまでできるようになったら 究極の翻訳方式と言ってよいのではないのでしょうか。本稿では、この研究が、その一步となることを期待して、題名をつけました。

6. むすび

「言語の意味的等価変換方式」について紹介しました。この方式は、数学的に見れば、言語表現の非線形領域を分解不能部分として抽出し、単位化することによって全体を線形近似に持ち込むもので、「局所的写像を介して大局写像を得る方法」と言えます。また、目的言語の複数の表現構造の中から、原言語の表現の意味を表すのに最も適した表現構造を類推し、その構造を使った訳文を作成しようとする点では、人間の翻訳の過程をシミュレートした方法です。

この方式の基礎となる「意味類型」と「類推思考」のアイディアは、起源の古いものですが、いずれも、自然言語処理では忘れられていました。従って、本方式は、忘れ去られていた思想と原理に光を当て、新しい方式技術としてよみがえらせようとするもので、機械翻訳の分野でのルネッサンスを目指したものと言えるかもしれません。

いずれにしても、この方式は、まだアイディアの段階です。実現する過程では、多くの難問が予想され、実際に実現可能かどうかは、これからの課題です。みなさんの知恵や経験を教えて頂き、是非、実現性を確認したいものです。

MT at Leeds University, UK

Tony Hartley

The use of language technologies to support professional linguists at work is the focal interest of the Centre for Translation Studies at Leeds University in England. Its activities cover both translation and conference interpreting, the two areas in which it offers distinctive Masters programmes. Within this broader setting, there is a particular interest in the evaluation of MT systems from the perspective of the user.

Leeds acknowledged the need to train today's translators in the use and evaluation of translation software by establishing, in 1996, the MA Applied Translation Studies (MAATS). It attracts students from Japan and China, as well as the Middle East, continental Europe and UK. In 2001, the Centre launched the MA Interpreting and Translation Studies (MAITS) to cater to interpreting trainees from a similar range of geographical origins. The principal languages covered by these two Masters programmes are : Arabic, Chinese, French, German, Greek, Italian, Japanese, Portuguese, Russian and Spanish. The students probably represent the largest and -- linguistically -- the most diverse group of trainee translators and interpreters in UK.

The Centre has a dedicated IT laboratory with a wide range of industry-standard translation software, and a state-of-the-art facility for conference interpreting equipped with double and single booths.

MT in the Masters programmes

A module in Machine Translation is available for students on both of the Masters programmes. Its goal is to enable them to evaluate the potential and limitations of MT through hands-on use of one or more commercial systems. The intention is that students acquire the knowledge needed for taking strategic decisions in a corporate environment about whether or how to use MT to assist with the gathering and dissemination of information. The module aims, therefore, to produce informed users rather than system developers. It does this through a mixture of seminars and laboratory sessions. Students are assessed on the basis of a practical project.

Some time is spent on understanding the challenges of Natural Language Processing in general and MT in particular. But students devote most of their efforts to creating and testing dictionaries for specific domains, and to conducting various evaluations of output quality. Since the details of dictionary creation may vary considerably from one system to another, students can observe directly the costs and benefits of relatively rich and relatively poor lexical representations. The evaluation of output quality may also take a comparative approach. The most

obvious object of the comparison is the MT systems, but it may equally be the evaluation techniques themselves. Students explore a range of techniques employed in exercises conducted by DARPA, JEIDA and others.

The MT applications available to students are those produced by Sail Labs, Sharp, Softissimo and Systran.

Masters programmes : overview

Most of the students who enter the course have obtained their first degree in foreign languages, but some have qualifications in other disciplines, such as law. Having expertise in some subject domain as well as in languages is a considerable advantage for professional linguists, so such applicants are encouraged. Many entrants have some experience of working after graduating with their first degree.

The compulsory core modules common to both programmes consist primarily of translation between English and the student's other language(s), with regular contributions from staff of EU institutions. Seminars on translation methods are designed to provide the conceptual tools of the trade. Short pre-sessional courses are offered in writing academic English and in achieving baseline skills in the use of the multilingual IT facilities.

Optional modules are available in Film Translation, a field of significant and rapid future growth, in several areas of Linguistics and in various other languages.

Those specialising in Conference Interpreting (MAITS) follow an intensive programme staffed largely by practitioners, including members of AIIC and staff interpreters from the UN and the European Commission.

Students whose first language is not English take a module targetting the high level of proficiency in public speaking expected of conference interpreters.

Those specialising in Applied Translation (MAATS) have a comprehensive introduction to computer-assisted translation tools (see below). They also follow a module in Technical Communication for Translators, in which they acquire the document design skills that are essential to practising translators, who must often improve on badly-written originals. For a selected software application, students design, test and revise (in English) user documentation and on-line help, using the dotCHM authoring environment for the latter.

All students devote the Summer to preparing a Masters Project, which may consist of a Dissertation on some aspect of translation or interpreting, or of two Extended Translations in different subject areas, with an accompanying critical commentary.

Computer-assisted translation

The Computer Tools for Translators module introduces MAATS students to an exceptionally wide array of applications. Rather than training them in the use of a single tool, the module aims to give them a critical overview of respective functionalities, strengths and weaknesses. It does this through a series of linked projects on which students work in multilingual groups, confronting all members

with the different challenges posed by, for example, Arabic, Chinese, Japanese and Russian character sets. They create translation memories and glossaries in one application, subsequently re-using them within others. Projects include both document translation and software localisation. Project management issues are addressed explicitly, and groups use the LTC Organiser application to schedule and monitor their tasks and resources.

The software tools currently in use are : Alchemy Catalyst, Deja Vu, Passolo, STAR Transit, IBM Translation Manager, Trados Translator's Workbench, Wordfast.

MT Research

Research activities are lead by Professor Tony Hartley, Director of the Centre. A number of projects are in progress. One, in collaboration with EPFL Lausanne, has made promising advances towards automatically ranking different MT systems in an ordering that is consistent with rankings based on human judgments. The first results, based on data from the 1994 DARPA MT evaluation exercise for French-English, appeared at MT Summit VIII. Further results, based on data from a new MT system and new human evaluations, will be presented at LREC 2002. The English output texts are scored according to features of the structural dependencies identified by Xerox's XELDA shallow parser. These features appear to be reliable and robust predictors, independent of the particular systems and subject domain. New experiments are planned to compare this research with recent work by IBM, and to conduct additional experiments on a large corpus of new source data, including fresh human evaluations.

Leeds is also working with the New Technologies Unit of the Joint Interpreting and Conference Service of the European Commission to study the impact of various technologies on interpreters at work. As the Commission attempts to support real-time spoken debate between European citizens and European institutions, so the project aims to establish how configurations of video-conferencing, simultaneous interpreting, chuchotage, sight interpreting and MT can best attain this objective. MT, on-line dictionaries and other web-based resources are becoming available in the interpreting booth, yet there is no clear understanding of their impact on the interpreters' other tasks.

Scheduled projects include : an investigation into the impact of MT output on Information Extraction performance (joint work with Professor Yorick Wilks, Sheffield) ; a study of the perceived usefulness of Japanese-English MT in the UK academic-scientific community ; and a contrastive analysis of selected genres of Chinese and English.

The Centre for Translation Studies is happy to discuss possible collaborative projects in MT design or evaluation with interested research groups. It is also happy to discuss with developers and vendors of software tools for translators how it might be able to bring their products to the attention of its current and future students.

Email cts@leeds.ac.uk
Web <http://leeds.ac.uk/cts>

References

Rajman, M & T Hartley (2001) Automatically predicting MT system rankings compatible with Adequacy, Fluency of Informativeness scores. Workshop on MT Evaluation, MT Summit VIII, Santiago de Compostela, Spain, September 2001.
Rajman, M & T Hartley (forthcoming) Automatic ranking of MT systems. 3rd International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas de Gran Canaria, Spain May/June 2002.

事務局だより

協会活動報告

(2001年11月～2002年3月)

理事会	3月26日	①02年度事業計画案 ②02年度収支予算案 ③決算理事会および総会日程 ④その他
MT SummitⅧ	9月18日～22日	EAMT 主催 (スペイン：サンチアゴ・コンポステーラにて)
運営委員会	12月11日	①収支予算状況討議 ②その他
	2月22日	①来年度収支予算討議 ②その他
	3月25日	①予算理事会準備 ②その他
編集委員会	1月10日	①前号反省 ②次号 (32号) 企画 ③その他
技術動向調査委員会	11月5日	①ホームページ強化討議 ②MT システム評価討議 ③その他
	12月6日	①ホームページ強化討議 ②MT システム評価討議 ③その他
	1月10日	①UPF 関連討議 ②ホームページ強化討議 ③MT システム評価討議 ④その他
	2月21日	①UPF 関連討議 ②ホームページ強化討議 ③その他
市場動向調査委員会	11月9日	①ユーザー向けアンケート計画討議 ②その他
	12月14日	①ユーザー／ベンダー向けアンケート計画討議 ②その他
	2月8日	①ユーザー／ベンダー向けアンケート計画詳細討議 ②その他
	3月8日	①今後の計画討議 ②その他
ネットワーク翻訳研究会	12月4日	①サミット報告 ②今後の計画討議 ③その他
	2月5日	①創価大／岡田先生「大量情報の統計的検定方法」 ②今後の計画討議 ③その他

Toward a Science of Machine Translating

Francis Bond

NTT Machine Translation Research Group
NTT Communication Science Laboratories
Nippon Telegraph and Telephone Corporation
bond@cslab.kecl.ntt.co.jp

1 Introduction

Despite decades of research, the quality of machine translation still does not reach the level of good human translations. Much recent research has focused on the automatic acquisition of lexical knowledge or transfer rules, while commercial development has concentrated on the important but non-core issues of integrating MT into DTP formats and HTML (Schütz, 2001). However, there is still an unmet need to improve the quality of machine translation itself.

Kay et al. (1994, p 200), in an early study of Verbmobil, strongly recommend that the methods of professional translators are studied as part of research into machine translation. Despite these recommendations, there has been surprisingly little attention paid by the machine translation community to how humans achieve high-quality results. In this paper, I suggest several ways to improve the quality of machine translation, based on the best practices of human translators, as described in Nida's, (1964) *Toward a Science of Translating*. I call this approach Multi-Pass Machine Translation (MPMT), as it crucially relies on processing the text more than once. It is similar to the opportunistic approach of Gdaniec (1999) in that it sets out to use any means at hand, adding to or changing them as necessary. The resulting approach integrates much recent research into a single system.

2 Toward a Science of Machine Translation

Nida (1964, 246-247) sets out the following nine steps to be employed by a competent translator (with some steps omissible):

1. Reading over the entire document
2. Obtaining background information
3. Comparing existing translations of the text [if they exist]
4. Making a first draft of sufficiently comprehensible units
5. Revising the first draft after a short lapse of time
6. Reading aloud for style and rhythm

7. Studying the reactions of receptors by the reading of the text by another person [omissible]
8. Submitting a translation to the scrutiny of other competent translators [omissible]
9. Revising the text for publication

There is no question that machines operate in different ways from humans: an airplane, for example, does not fly in the same way as a bird. Nevertheless, translation is a task of manipulating texts written by humans for humans; therefore there is value in examining how humans approach the task (see Kay et al. (1994, pp 200–202) for further discussion).

In the following nine subsections, I will suggest how most of Nida's steps for human translators can be adopted by an MT system. I will concentrate on the translation of texts; interpretation is a different problem which requires different solutions. I will then combine them to produce a multi-pass machine translation system (§ 3).

2.1 Reading over the entire document

Before translation, a system should parse the entire document, in order to do the following things:

- Identify the source language (possibly per segment)
- Identify the domain
- Identify terms and named entities

Knowledge of the source language is an obvious prerequisite for machine translation, as it enables the system to select the appropriate grammar and lexicons. In many settings, this knowledge can be assumed, but the language is not normally identified on the web. Another problem that requires language identification is the reasonably common practice of interspersing text with text of one or more different languages.

Knowledge of domains has been shown to improve the quality of translation in at least two studies. Yoshimoto et al. (1997) using 41 hierarchically organized domains (which they call Subject Areas) marked on 9,000 words, were able to improve the translations of 12% of the badly translated nouns ($\frac{46}{383}$). Lange & Yang (1999) used 77 Domains (TERMinology CATegories) and 30 Topical Glossaries. In a first pass through they chose two appropriate domains (the domains with the greatest number of domain-tagged words). With these domains set, there were changes in 0–40% of translations, and the majority of the changes were improvements.

Identification of terms and named entries are both robust research fields in their own right. It is important for a translator to identify them, so that they can ensure they have the correct translations for that domain, as discussed in the next section.

2.2 Obtaining background information

Once the domain is known, and lists of terms and named entities have been produced, the system must try and find appropriate translations. This is often the most time-consuming task for human translators, especially for new domains and unknown proper nouns.

Identifying unknown proper nouns has been shown to give a 20.8% improvement in quality of translation of proper nouns (Yoshimoto et al., 1997). Knowing whether something is a proper noun is crucial to both the analysis of the text and the selection of the translation. Further, Kim et al. (2001) showed that creating a user dictionary of unknown words in a first pass, increased the accuracy of the morphological analysis for Korean in a second pass by 2.26%.

Translations of unknown words can be extracted from unaligned corpora (Tanaka & Matsuo, 1999), or from the world wide web (Grefenstette, 1999). Pinkham & Smets (2002) claimed that, if you could learn a transfer database by training on domain corpora then the only non-domain specific system dictionary needed was a small system dictionary of function words. In fact, their results showed that, while you could get comparable results without a general purpose system dictionary, using a wide-coverage system dictionary did improve quality slightly. Even so, the fact that a domain-specific induced dictionary was almost able to replace a general dictionary shows just how useful it is to learn domain-specific translations.

Systems that use stochastic models of source and target text, such as n-grams, or rankings for choosing translations and variations, should retrain on, or adapt to text of similar domains for their language models.

2.3 Comparing existing translations

One proof of the utility of comparing with existing translations is the well-attested benefits of the use of translation memories (Planas, 1999). Their value is so widely known that there is little need to say more. If existing translations are available they can also be used to extract rules (Yamada et al., 1995), or provide data for example-based machine translation. Further, a translation can help to guide a system's understanding of the source text by adding extra constraints on its parse, in the same way that a human may look to a translation to clarify the meaning of a section they find hard to understand. Wu (1997) shows a way of exploiting a translation to parse a text using an inversion transduction grammar.

2.4 Making a first draft

The actual translation process itself should use a good rule-based system with the correct domain dictionaries and unknown words and their translations identified and put into a user dictionary, possibly combined with a translation memory to translate previously encountered text. This creates the first draft.

2.5 Revising the first draft

There have been some suggestions to check for deceptive cognates (Isabelle et al., 1993), or omissions in translation (Russell, 1999), and to choose articles

(Chander, 1998). At present, these techniques are not yet so useful in practice, probably because the revision systems are not able to consider the meaning of the text. Naruedomkul & Cercone (1997) suggested an MT architecture that would first translate and then revise poor translations but they did not implement the revision part of the system.

One task that could be considered a revision is the addition of translator's footnotes or notes. Whether they are appropriate or not depends on the target audience. A translator's note is useful if a word without a good translation equivalent, but with a known definition, appears several times. In this case it should be glossed with a definition (the note) the first time and then transliterated in subsequent uses (possible in a different font). For example:

There were several *akah* (a kind of turtle) lying on the beach. . . . The next morning, I was awakened by an *akah* falling on to my sleeping bag.

It is also common for named entities to be introduced by an explanation in some reporting. For example, in non-Japanese English-language newspapers, NTT is often introduced with an explanation such as *NTT, Japan's dominant telephone company*, This is not necessary in a Japanese text, as everyone presumably knows that NTT is Japan's dominant telephone company, but becomes necessary in translation. It would be useful to add such notes either during the translation, as part of an initial revision, or as part of the final revision before publishing.

2.6 Reading aloud for style and rhythm

The style and rhythm of a text are analogous to its probability: if it sounds natural then it should be a likely sentence, and if it is an unlikely sentence it should sound unnatural. I take this step to suggest a translation model which generates all possible **variations** (translations with almost identical semantic restrictions) and chooses the best one based on a target language model. This is similar to the system proposed by Knight et al. (1994), who use n-grams as the target model.

2.7 Studying the reactions of other receptors

In order to decide which parts of the translation are of high quality, the text should be parsed by a target language parser, and scored in some way. Bernth (1999) has shown that a confidence index is useful when presenting text to end users, and suggested a score based on the translation process. This is a good idea, but requires detailed knowledge of the transfer process. A simple monolingual metric measured on the target text is more robust.

2.8 Submitting a translation to other translators' scrutiny

By using a confidence measure on the output text, it is possible to rank the outputs of more than one translation system. Callison-Burch & Flournoy (2001) trained a trigram language model on 2,000,000 words of English and used it to choose among the outputs of four Japanese-English systems. This gave an

overall quality of 74% compared to 70%, 58%, 40% and 27% for the individual systems. Choosing between two French-English systems gave even better results: 84% compared to 76% and 56%. From this we can see that it is well worth consulting other translators, should you have access to them.

2.9 Revising the text for publication

Finally, to prepare the text for publication, it is important to restore any markup. To do this the system must keep a record of source text equivalences and tags, as well as any text-style tags.

Markup itself may need to be translated. For example, Japanese text traditionally has no bold or italic variants, so emphasis would be provided by putting small dots next to the characters (called *bōten*). These should be converted to underlining or a different font in translation.

3 Multi-Pass Machine Translation

In this section I will describe two versions of a the Multi-Pass Machine Translation system. The first is a single engine system, the second is a multi-engine system, that is, a system that combines the outputs of more than one system.

3.1 Single Engine Multi-Pass Machine Translation System

The basis of a single engine Multi-Pass Machine Translation system is a semantic transfer system. The text is analysed with a rule-based parser, and the best parse chosen using a stochastically trained source language model. The output should be a semantic representation that is as language independent as possible. In practice, a full semantic analysis is an AI complete problem, and thus far from being solved. For the time being, we must resort to the Multi-Level Transfer method (Ikehara et al., 1991, 1996) where a text is analysed as far as is judged necessary, and then transferred to the target language. The transfer stage has a rule-based core with a transfer model to choose between alternatives. Finally, the target language text is generated from the target language semantic representation. In order to benefit from work in mono-lingual parsing, generation and lexicons, it is desirable to keep the analysis and generation systems as modular as possible. The enhancement of this system by the multi-pass model is shown in Figure 1. Although here I assume that the translation is done sentence by sentence, following the current state-of-the-art, the architecture is equally applicable to a discourse-based system such as that proposed by Marcu et al. (2000).

The multi-pass method relies critically on two things. The first is having enough processing power and training data to be able parse several times, and retrain models on the fly. Such data and power is only just becoming available. Exploiting it is part of the general trend to push more background work onto the computer, so that humans can use their time more productively. The second thing is an integrated mixture of rules and stochastic rankings, which I see becoming more and more dominant in natural language processing in general. In this sense I agree with Och & Ney (2001) who argue that “for future successful MT systems statistical machine translation will be a much more important

-
1. Identify the source text language (if unknown)
 - Check the language of each unit (sentence or paragraph)
 2. Pre-parse the text
 - (a) Extract named entities
Put named entities into local dictionary
 - (b) Run a morphological analyser to identify any unknown words
Add unknown word candidates to local dictionary
 - (c) Identify the genre and domain(s)
 3. Find translations for the named entities and unknown words
Enter them into local transfer and target dictionaries
 4. Train the source and target language target models using text from similar genre and domain(s)
 - If there are any existing translations, or translations from the same domain, train the transfer rules on them
 5. Translate using the local dictionaries, appropriate domain dictionaries, and the domain-trained source, transfer and target models
 6. Check the style of the resulting text using a target language checker
 7. Restore any markup from the source text.
 - Possibly add explanatory notes or footnotes

Figure 1: Single Engine Multi-Pass Machine Translation

approach than it is today". In my approach, however, I incorporate statistical models into a rule-based system, rather than add linguistic knowledge to a statistical model.

I hope that this architecture, and the advances in natural language processing on which it is based, will help to bring machine translation ever closer to the capabilities of human translators.

3.2 Multi-Engine Multi-Pass Machine Translation System

At the current level of success of machine translation systems, a multi-engine Multi-Pass Machine Translation system along the lines suggested by Callison-Burch & Flounoy (2001) seems worth building. This system translates a text with multiple translation systems, and then chooses the best output for each sentence. The passes could in fact be carried out in parallel, even on different machines. Callison-Burch & Flounoy (2001) used n-grams to select the best result from several rule-based systems. This could be a problem if a statistical system was included, as its results will always be smooth, thus an n-gram-based ranking would be biased towards its results. To choose fairly, the ranking would have to be based on fidelity as well as fluency.

However, for really high quality translation (equalling the capabilities of a good human translator), translating one sentence with one system and one sentence with another is not a good strategy. A good text must be coherent, written in the same register, with reference to entities depending on how they have been referred to in the past. It is not possible to do this by combining the output of multiple systems sentence by sentence. For the same reason, translation memories can not be expected to provide the best translations of complete texts, unless the whole text is in the memory. That is not to deny that they are extremely useful now, due to the poor performance of current machine translation systems.

Therefore, rather than combining the results of multiple systems as described above, in a perfect world a single-engine MPMT system would instead combine all of the knowledge (rules and lexicons) used by the multiple systems. This would be enhanced by any transfer rules and statistics it could learn from existing translations, and the result used to produce a single coherent translation. This is the closest to Nida's, (1964) strategy. The main problems to doing this are more social than technical. This is not to belittle the genuine technical problems involved with combining multiple knowledge sources, but rather to acknowledge that they are dwarfed by the legal problems of gaining access to the raw linguistic resources used in multiple machine translation systems.

4 Conclusion

As processing power has increased it is becomingly more and more feasible to process a text more than once when translating. This is how the best human translators produce their translations, and doing so provides a way for machine translation systems to improve their quality. Therefore, I propose a new approach to lead us on to the goal of fully automatic high-quality translation: Multi-Pass Machine Translation.

Acknowledgments

I have discussed the ideas in this paper with many people. I would particularly like to thank past and present members of the NTT Machine Translation Research Group, Timothy Baldwin, Ann Copestake, Mark Gawron, Claudia Gdaniec, Kyonghee Paik and Emmanuel Planas. An earlier version of this paper was presented at the Proceedings of the MT Roadmap Workshop at TMI-2002 <www.eisnet.org/roadmap-tmi2002.html>, and has been revised in light of the extremely interesting discussion there.

References

- Bernth, Arendse: 1999, 'A confidence index for machine translation', in *Eighth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-99*, Chester, UK, pp. 120–125.
- Callison-Burch, Chris & Raymond S. Fournoy: 2001, 'A program for automatically selecting the best output from multiple machine translation engines', in *Machine Translation Summit VIII*, Santiago de Compostela, pp. 63–66.
- Chander, Ishwar: 1998, 'Automated postediting of documents', Ph.D. thesis, University of Southern California, Marina del Rey, CA.
- Gdaniec, Claudia: 1999, 'Using MT for the purpose of information assimilation from the web', in *Workshop on Problems and Potential of English-to-German MT systems*, TMI, Chester, UK, <http://www.research.ibm.com/people/g/cgdaniec/tmi99.html>.
- Grefenstette, Gregory: 1999, 'The WWW as a resource for example-based MT tasks', in *Translating and the Computer 21: ASLIB'99*, London.
- Ikehara, Satoru, Satoshi Shirai & Francis Bond: 1996, 'Approaches to disambiguation in ALT-J/E', in *International Seminar on Multimodal Interactive Disambiguation: MIDDIM-96*, Grenoble, pp. 107–117.
- Ikehara, Satoru, Satoshi Shirai, Akio Yokoo & Hiromi Nakaiwa: 1991, 'Toward an MT system without pre-editing – effects of new methods in ALT-J/E–', in *Third Machine Translation Summit: MT Summit III*, Washington DC, pp. 101–106, (<http://xxx.lanl.gov/abs/cmp-1g/9510008>).
- Isabelle, Pierre, Marc Dymetman, George Foster, Jean-Mard Jutras, Elliot Macklovitch, François Perrault, Xiabao Ren & Michel Simard: 1993, 'Translation analysis and translation automation', in *Fifth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-93*, Kyoto, pp. 201–217.
- Kay, Martin, Jean Mark Gawron & Peter Norvig: 1994, *Verbmobil: A Translation System for Face-to-Face Dialog*, no. 33 in CSLI Lecture Notes, Stanford, CA: CSLI.
- Kim, Seonho, Mansuk Song & Yuntae Yoon: 2001, 'Proper analysis of unknown words using local dictionary', in *19th International Conference on Computer Processing of Oriental Languages: ICCPOL-2001*, Seoul, pp. 439–444.
- Knight, Kevin, Ishwar Chander, Matthew Haines, Vasileios Hatzivassiloglou, Eduard Hovy, Masayo Iida, Steve K. Luk, Akitoshi Okumura, Richard Whitney & Kenji Yamada: 1994, 'Integrating knowledge bases and statistics in MT', in *Proceedings of the 1st AMTA Conference*, Columbia, MD.

- Lange, Elke D. & Jin Yang: 1999, 'Automatic domain recognition for machine translation', in *Machine Translation Summit VII*, Singapore, pp. 641–645.
- Marcu, Daniel, Lynn Carlson & Maki Watanabe: 2000, 'The automatic translation of discourse structures', in *The 1st Meeting of the North American Chapter of the ACL: ANLP-NAACL-2000*, Seattle, pp. 9–17.
- Naruedomkul, Kanlaya & Nick Cercone: 1997, 'Steps toward accurate machine translation', in *Seventh International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-97*, Santa-Fe, pp. 63–75.
- Nida, Eugene A.: 1964, *Toward a Science of Translating*, Leiden, Netherlands: E. J. Brill.
- Och, Franz Joseph & Hermann Ney: 2001, 'What can machine translation learn from speech recognition', in *MT 2010 — Towards a Road Map for MT*, Santiago de Compostela, MT Summit VIII Workshop, pp. 26–31.
- Pinkham, Jess & Martine Smets: 2002, 'Machine translation without a bilingual dictionary', in *Ninth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-2002*, Keihanna, Japan, pp. 146–156.
- Planas, Emmanuel: 1999, 'Formalizing translation memories', in *Machine Translation Summit VII*, Singapore, <http://www.kecl.ntt.co.jp/icl/mtg/members/planas/MTSummitVIIconf.ps>.
- Russell, Graham: 1999, 'Errors of omission in translation', in *Eighth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-99*, Chester, pp. 128–138.
- Schütz, Jörg: 2001, 'Blueprint for MT evolution: Reflection on "Elements of Style"', in *MT 2010 — Towards a Road Map for MT*, Santiago de Compostela, MT Summit VIII Workshop, pp. 9–13.
- Tanaka, Takaaki & Yoshihiro Matsuo: 1999, 'Extraction of translation equivalents from non-parallel corpora', in *Eighth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-99*, Chester, UK, pp. 109–119.
- Wu, Dekai: 1997, 'Stochastic inversion transduction grammars and bilingual parsing of parallel corpora', *Computational Linguistics*, 23(3): 377–403.
- Yamada, Setsuo, Hiromi Nakaiwa, Kentaro Ogura & Satoru Ikehara: 1995, 'A method of automatically adapting a MT system to different domains', in *Sixth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-95*, Leuven, pp. 303–310.
- Yoshimoto, Yumiko, Satoshi Kinoshita & Miwako Shimazu: 1997, 'Processing of proper nouns and use of estimated subject area for web page translation', in *Seventh International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-97*, Santa Fe, pp. 10–18.

★ ★ ★ ★ ★ ★ ★

TMI-2002 On-line Conference Proceedings

The TMI-2002 conference and workshop proceedings are now on-line, see:

<<http://www.kecl.ntt.co.jp/events/tmi>>.

★ ★ ★ ★ ★ ★ ★

第12回通常総会開催のご案内

アジア太平洋機械翻訳協会の第12回通常総会を下記のとおり開催いたしますので、ご多用中とは存じますが、是非ご出席くださいますようお願い申し上げます。

記

1. 日 時 2002年6月21日（金） 13：00～13：45
2. 会 場 千代田区神田駿河台3-11 中央大学駿河台記念館6F 670会議室
（JR「御茶ノ水駅」徒歩3分、地下鉄千代田線「新御茶ノ水駅」徒歩3分）
3. 議 題 第1号議案 2001年度事業報告（案）
第2号議案 2001年度決算報告（案）
第3号議案 2002年度事業計画（案）
第4号議案 2002年度予算（案）
第5号議案 役員改選案
会員提案事項、その他

4. 総会当日の全体プログラム

6月21日通常総会のあとに、会員の皆様との交流、懇親を目的として次のような行事を予定しております。ぜひ多数ご参加くださいますようお願い申し上げます。

行 事	内 容	時 間
総 会	第12回通常総会 （参加 AAMT 会員のみのみ）	13：30～13：45
活動報告会	当協会各委員会からの活動報告 （参加費無料、どなたでも参加可）	14：00～15：00
講 演 会	①「機械翻訳の社会的インパクト」 一橋大学助教授 ジョナサン・ルイス 氏 ②「音声翻訳の展望と情報の獲得・利用」 （株）ソニーコンピュータサイエンス研究所 飯田 仁 氏 （①②とも参加費無料、どなたでも参加可）	15：00～17：00
懇 親 会	立食形式懇親パーティ （参加費4千円 どなたでも参加可、当日お支払い） 会場：中大駿河台記念館1F レストラン “プリオール”	17：20～19：00

申込方法：E-mail またはファクスで、住所、氏名、電話番号、勤務先及び懇親会参加の有無を AAMT 事務局までご連絡ください。

（会員の皆様には追って資料及び、出欠回答のハガキを郵送いたします。）

E-Mail：aamt 0001@infotokyo.ne.jp 又は aamt 0002@infotokyo.ne.jp

Fax： 03-3518-6472

ポータルサイト「goo」上で「多言語情報アクセス技術」を実験

ー英語・韓国語・中国語のページを日本語で検索・表示
さらに、日本語を英韓中へ翻訳する技術ー

日本電信電話株式会社

NTTサイバーコミュニケーション総合研究所

1. はじめに

NTTサイバーコミュニケーション総合研究所では、グローバルな情報流通へ向けた研究開発を進めてきています。その1つに「言語の壁」を越えた情報アクセスを支援する技術があります。この技術に関して、多くのユーザが利用する実サービス環境下で技術検証を行うこととなりました。具体的には、株式会社NTT-Xがサービス提供しているポータルサイトの「goo」上で、NTT-X社の協力を得て翻訳実験を行っています。期間は、平成14年3月から7月31日までです。

2. 実験サービスの概要

今回の実験では、「goo」(<http://www.goo.ne.jp>)にアクセスし、「翻訳」をクリックすれば、以下の実験サービスを無料で利用できます。翻訳実験のトップページは図1のような構成になっています。

(1) 翻訳検索実験サービス

日本語・英語・韓国語・中国語で書かれたページを、日本語で検索し、日本語で読むことができます。また、英語・日本語で書かれたページを、英語で検索し、英語で読むことができます。具体的には、図1の翻訳検索の項目にある「キーワード入力」の欄にキーワードを入力します。

(2) Web ページ翻訳実験サービス

閲覧したいWeb ページのURLを入力し、「英語→日本語」「日本語→英語」「韓国語→日本語」「日本語→韓国語」の翻訳パターンを指定することで、ページ全体を翻訳することができます。具体的には、図1のWeb ページ翻訳の項目にある「URL 入力」に翻訳したいページを指定します。なお、4月中を目処に「日本語→中国語」「中国語→日本語」のWeb ページ翻訳実験サービスも開始する予定です。

(3) テキスト翻訳実験サービス

日本語文章(500文字以内)を入力して、英語または韓国語へ翻訳でき、逆に、英語文書、韓国語文書を入力して日本語に翻訳することができます。具体的には、図1のテキスト翻訳の項目にある「文章入力」に翻訳したい文書を直接入力します。なお、4月中を目処に日中間のテキスト翻訳実験サービスも開始する予定です。

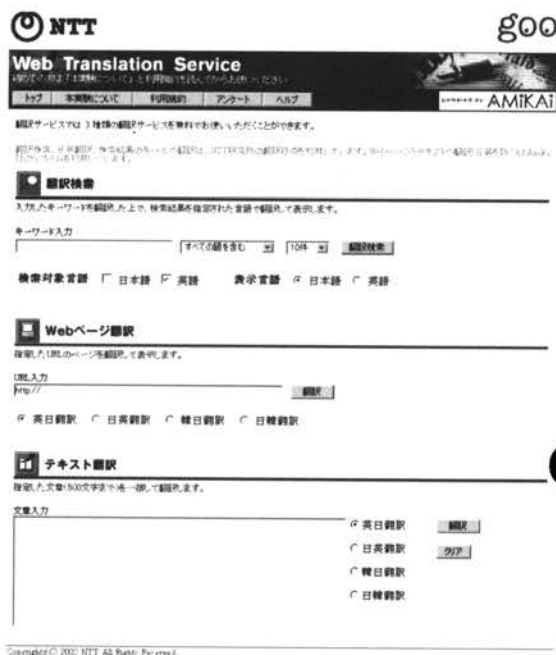


図1 翻訳実験トップ画面

3. 具体的な技術項目

今回の実験では、超高速インターネット・ロボット検索エンジン/検索語句翻訳エンジン/タイトル翻訳エンジン/Web ページ翻訳エンジン/テキスト翻訳機能で構成される「多言語情報アクセス技術」の実環境における有効性・有用性の検証を行い

ます。特に、以下の(1)(2)の翻訳エンジンが、既存の各種翻訳サービスと異なりそれぞれの翻訳対象の特徴にあわせた処理を行うことが大きな特徴です。また、新たに「韓国語で書かれた Web ページのデータベース」及び「中国語で書かれた Web ページのデータベース」をつくる点も特徴です。

(1) 検索語句翻訳エンジン

ユーザの検索語句を適切に翻訳します。これまでの商用翻訳検索サービスでは、1つの訳語で検索がなされていますが、NTTとNTT-Xでは「どの訳語を検索に利用するか」が検索の精度及びユーザの満足度に大きく影響すると考えています。多くの場合、ユーザが検索時に入力する語句は数個の単語から成っており、しかも、その1つ1つの単語にはそれぞれ複数の訳語があります。そこで本エンジンでは、ユーザが入力した検索語句を、まず複数の単語に分解し、それぞれの単語に対して複数の訳語の可能性を許しながら、これらの訳語の組み合わせの中から統計的に適切な組合せを求め、他国語での検索要求式を生成します。これは日本初の試みです。

(2) タイトル翻訳エンジン

検索結果において表示される各ページの「タイトル」を翻訳するエンジンです。Web ページのタイトルは独特の表現パターンを持っており、体言止のような表現(名詞句)となっていることが多いなど、一般の文章とはスタイルが異なります。このような言語表現の翻訳を適切に行うために、タイトル原文の部分々々に、タイトルの訳語として適した語句を複数当てはめ、更にこれらの組合せの中から統計的に最も適切な訳語を選ぶことによりタイトルを生成します。これも日本で初めての試みです。

(3) Web ページ翻訳エンジン

Web ページの本文を翻訳します。実験開始当初は「英日翻訳」「日英翻訳」「韓日翻訳」「日韓翻訳」の各機能を、また、4月中を目処に「中日翻訳」「日中翻訳」機能の実験も開始し、アンケート調査等によ

り各機能の有効性・有用性を評価します。「日英翻訳」では、NTTが研究を行ってきた「ALT-J/E」と呼ぶシステムを利用しています。これ以外の言語ペアの翻訳については、Amikai 株式会社の ASP サービスを利用しています。

(4) テキスト翻訳機能

テキストを翻訳する機能です。実験開始当初は「英日翻訳」「日英翻訳」「韓日翻訳」「日韓翻訳」の各機能を、また、4月中を目処に「中日翻訳」「日中翻訳」機能の実験も開始し、アンケート調査等により各機能の有効性・有用性を評価します。ここでも、「日英翻訳」にはNTT研究所の「ALT-J/E」を利用し、それ以外の言語ペアの翻訳については Amikai 株式会社の ASP サービスを利用しています。例えば、「日英翻訳」に関しては、図2のような結果となります。



図2 テキスト翻訳例(日英翻訳)

4. 今後の予定

実験期間終了後は、本実験で得られたデータを基に、ビジネス化に向けた検討を進め、本サービス開始へと繋げる予定です。

翻訳実験トップページへのアクセス先

● <http://jupiter.nttrd.com/>

機械翻訳関連 Webニュース/マガジン リンク集

Webニュース/Webマガジンおよび機械翻訳関連のサイトを紹介します。本ページの内容は、AAMT ホームページにも記載され、随時更新しております。なお、本ページに記載されているページへのアクセスにより発生した被害をAAMTは補償いたしません。(2002/4月現在。順不同)

翻訳ソフトのページ - Green and White

<http://homepage2.nifty.com/oto3/>

個人向け翻訳ソフトの総合サイト。品質評価が充実

翻訳ソフトの学力試験

<http://www.clio.ne.jp/~kuwa/honyaku.htm>

投稿ベースの翻訳結果比較/MTIに関するチャットルーム開設

Yahoo! ニュースランキング

http://news.yahoo.co.jp/ranking/bcn/cpt/biz1_w.html

翻訳ソフトの販売シェアランキング(毎週更新)

Sales Week 3200

http://ma.nikkeibp.co.jp/cgi-bin/sw32_mn.cgi

翻訳ソフトの販売シェアランキング(インターネット/OAツール区分 毎週更新)

ISIZE PC相談室

<http://www.isize.com/pcsoudan/>

翻訳ソフトのクチコミ評価

南花台東小学校

<http://www5a.biglobe.ne.jp/~nankadai/>

学校教育への翻訳ソフト導入例

CICC 国際共同研究開発

<http://www.cicc.or.jp/japanese/kyoudou/mt.html>

近隣諸国間の機械翻訳システムに関する研究協力報告

中央大斉藤ゼミ長谷川氏HP

http://chubun.lite.tamacc.chuo-u.ac.jp/saitolab/forum98/hasegawa_m/honyaku.html

機械翻訳概要

All About Japan 英語の学び方・活かし方

<http://allabout.co.jp/education/english/closeup/CU20020412/index.htm>

ちょっとした工夫で翻訳ソフトは使いやすくなる!

Business Computer News

<http://www.computernews.com/>

翻訳ソフト・OCRソフトの販売傾向

Yomiuri On-Line Bit by Bit

<http://www.yomiuri.co.jp/bitbybit/bbb07/231901.htm>

日米間の小学生ネット交流 自動翻訳機能が活躍

Translation Journal (Vol.6, No.2)

<http://www.accurapid.com/journal/20mt.htm>

Useful Machine Translations of Japanese Patents Have Become a Reality

Nikkei BP Network

<http://smallbiz.nikkeibp.co.jp/members/QA/20020121/100705/>

英文ホームページを翻訳するソフトの選び方について教えて

PCW ARENA

http://arena.nikkeibp.co.jp/rev/soft/gaz/soft_179/index.shtml
1万円以下で買える翻訳ソフト5製品

未来生活研究所 / 大河原家家族会議

<http://www.sankei.co.jp/advertising/itnews-labo/kadenseikatu/33/01.html>
同時通訳マシンはすぐそこ！

BizTech Special

http://premium.nikkeibp.co.jp/english/frame2_3.html
主要オンライン翻訳サイトをチェック！ 翻訳サービスは本当に使えるのか？

Global Information,Inc プレスリリース (2001/08/30)

http://www.gii.co.jp/press/ab7638_jp.shtml
2006年末までに世界の機械翻訳市場は1億1,700万ドル規模に

日経コンピュータ IT Proニュース (2001/06/12)

<http://itpro.nikkeibp.co.jp/free/NC/NEWS/20010612/5/>
2007年のパソコンは日英即時通訳が可能

All About Japan Close Up(2001/05/30)

<http://allabout.co.jp/career/dictionary/closeup/CU20010530/index.htm>
メールだって大丈夫！ オンライン翻訳サイトで海外へ！

CNET Japan Software Review (2001/03/26)

http://cnet.sphere.ne.jp/JustIn/jp_soft/010326/
特集翻訳ソフトは人を超えるか！？

Jkapan.internet.comデイリーリサーチ(2001/03/19)

<http://japan.internet.com/research/20010319/1.html>
自動翻訳の発達に期待83%

ASCII DIGITAL BUYER(2001/01/12)

<http://db.ascii24.com/buyer/review/soft/translate/2001/01/12/621428-000.html>
徹底解析 厳選定番英日／日英翻訳ソフト10本＋α

INTERNET Watch(2000/10/30)

<http://www.watch.impress.co.jp/internet/www/article/2000/1030/trans.htm>
実践比較！ オンライン翻訳サービス

Mainichi INTERACTIVE カヴァーストーリー(2000/08/09)

<http://www.mainichi.co.jp/digital/coverstory/archive/200008/09/1.html>
コンピューター翻訳は人に代われるか

BizIT : 日経BP:USニュースフラッシュ(2000/05/30)

<http://bizit2.nikkeibp.co.jp/wcs/usn2/article/20000530/08.shtml>
世界の翻訳市場予測

HotWired Japan(1999/08/17)

<http://www.hotwired.co.jp/bitliteracy/guest/990817/textonly.htm>
コミュニケーション拡張装置としての機械翻訳

ZDNet Japan(1999/07/16)

<http://www.zdnet.co.jp/news/9907/16/chat.html>
ネットは言語の壁も乗り越えるか？

バベルオンライン講座(1999/02/26)

<http://www.babel.co.jp/mtsg/online/jta/>
翻訳のための機械翻訳の活用方法

AAMTホームページ

<http://it.jeita.or.jp/aamt/index.htm>

新製品紹介

翻訳品質・支援機能をさらに強化した英日／日英翻訳ソフト

「The翻訳プロフェッショナル V7.0」

株式会社 東芝



1. はじめに

東芝は、辞書語数の増強による翻訳品質向上と、辞書登録のための用語抽出と品詞情報の自動判別による辞書構築支援機能、および表示画面のマルチウィンドウ化による翻訳支援機能を強化した「The翻訳プロフェッショナル V7.0」を2月8日より発売しました。

「The翻訳プロフェッショナル V7.0」は、本格的な翻訳業務向けソフトとして、原文・訳文の対訳文の登録により、類似した文を参照できる翻訳メモリ機能や英文作成中の単語補完等を行う英論文作成支援機能、複数のパソコンによる翻訳作業を支援するためにサーバ上でユーザ辞書を共有できる機能等をサポートしており、Microsoft® Office XPをはじめとするオフィス文書からホームページ、メールあるいはイメージスキャナ入力した印刷物までの多様な原文を翻訳できる機能なども備えています。また、同日に発売された「英日／日英特許翻訳オプション V2.0」により、特許文書の翻訳精度をさらに向上させることが可能です。

2. 主な機能

(1) 操作性が向上した対話翻訳

The翻訳プロフェッショナルのメイン機能である対話翻訳では、同一ウィンドウ内に関連情報の表示エリアを設置し、類似文検索、確認翻訳、訳語選択、EPWING 辞書検索が行えるようになった他、翻訳方向の切り替えも簡単になりました。また、別々の原文・訳文ファイルを文の対応付けをしながら読み込む文アライナー機能を搭載、翻訳メモリへ登録するデータの作成が容易に行えます。Word 文書、Excel シート、PowerPoint®スライド、PDF ファイルからテキスト部分を抽出して読み込めるので、原文データ形式の選択の幅が広がりました。

(2) 過去の資産活用にも有効な翻訳メモリ機能

今までに蓄積したノウハウを生かした翻訳を実現する、翻訳メモリ機能を搭載しています。旧バージョンの対訳データや、文アライナー機能によって原文・訳文ファイルの文の対応付けをしながら読み込んだデータ、さらに TRADOS®等の製品で採用されている TMX 形式データを登録することができ、これを対話翻訳の類似文検索機能で検索、検索結果の訳文を翻訳結果として利用できます。

また、文中の特定の単語のみが異なる場合は、穴空き例文としてパターン辞書に登録することもできます。

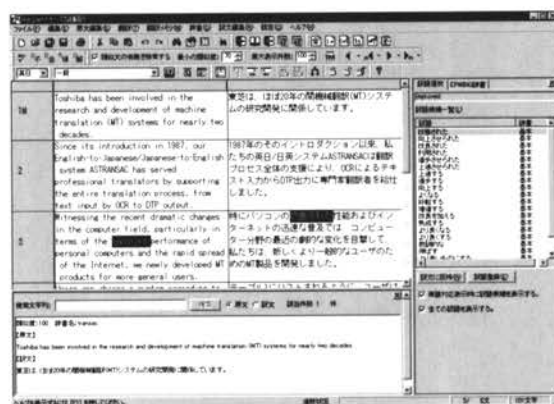


図1：対話翻訳画面

(3) ユーザ辞書作成を容易に行えるインテリジェント辞書登録機能

辞書データを作成する際に手間がかかるのは、品詞情報などの設定です。原文を読み込み用語抽出を行うと、インテリジェント辞書登録画面に辞書への登録を推奨する語と推定訳や品詞等の情報がセットされて出力されます。ここで必要な修正を加え、すぐ辞書登録を行うことが可能です。また、辞書登録後は、ユーザ辞書一覧編集機能で見出し語や訳語の編集も自由に行えます。

(4) 希望の解釈を選べる別解釈機能

複数の解釈を持つ文を翻訳すると意図した結果が最優先されない場合があります。別解釈表示機能ではそれぞれの解釈による翻訳結果を表示し、その中から適切な解釈を選択して訳文に適用できます。

(5) オフィス連携翻訳

翻訳対象となるデータがテキストファイルではなく、かつレイアウトを崩さない翻訳が要求される場合も少なくありません。The翻訳プロフェッショナルでは社内文書などに多く使用されている Word、Excel をはじめとする Microsoft®Office のツールに翻訳ボタンを埋め込んで、レイアウト情報を残したまま翻訳できます。また、Word での翻訳時には 1 文ずつ内容を確認しながら翻訳するステップ翻訳機能と、英日/日英特許翻訳オプションを組み込むことによって特許文書としての翻訳を行える機能を追加できるようになりました。

(6) ブラウザ翻訳

世界の情報を入手するために利用されるホームページ。翻訳したいページが表示されたら Internet Explorer や Netscape 6 に組み込まれた翻訳ボタンをクリックするだけで翻訳結果が得られます。さらに、翻訳パレットを起動を起動することにより、翻訳ボタンをクリックすることなく、翻訳結果が表示される自動翻訳も行えます。また、Internet Explorer の画面では、選択範囲の翻訳結果を対訳表示したり、画面上の単語を辞書引きする機能もあり、欲しい部分だけを翻訳することができます。



図 2：Internet Explorer 画面

(7) メール翻訳

海外との取引に多く使われる電子メール。受け取ったメールを保存せずに直接翻訳することが可能です。翻訳結果は原文と訳文が対訳表示される対訳ビューワに表示され、再翻訳も行えます。

(8) マルチラインクイック翻訳

マウスでポイントした部分を翻訳できます。複数行にわたる文の先頭と最後を自動的に検出して 1 文を翻訳できるようになりました。単語単位での翻訳も、従来通り行えます。

(9) 翻訳結果を確かめる確認翻訳

日本語を入力して翻訳した結果が自分の意図したのものになっているかを、翻訳結果を再翻訳することで確かめることができます。入力してから実行する対訳ビューワと、1 文ずつ入力しながらの翻訳入力で使用できます。

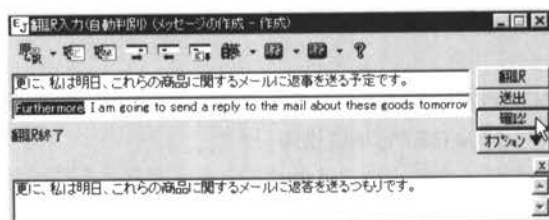


図 3：翻訳入力画面

(10) PDF ファイルのレイアウトを保持して翻訳が可能な Adobe®Acrobat®連携翻訳

PDF ファイルのまま翻訳することができ、翻訳結果は元の文書のレイアウトに復元します。さらに、PDF ファイルを RTF ファイルに変換する機能により、Word などの RTF ファイル対応ソフトを使って訳文を編集することができます。また、Adobe® Acrobat®がなくても Adobe®Acrobat®Reader™でテキスト部分を抽出し、翻訳を行うことが可能になりました。

(11) 複数の辞書をまとめるユーザ辞書マージ機能

複数に分散して作ったユーザ辞書や他の人から貰った辞書を一つにまとめることができます。また、ユーザ辞書をサーバ上に置いて共有することも可能です。

(12) 語彙の豊富な辞書を搭載

当社の従来製品に比べ英日26万語→27万語、日英25万語→基本辞書27万語に基本用語辞書を増加させるとともに英日は26万語（ビジネス、情報、電気・電子、機械、化学、医学・バイオ、航空、住所、ショッピング、オークション、エンターテインメント、グルメ、スポーツ、トラベル、キャンパスライフ）、日英では24万語（ビジネス、情報、電気・電子、機械、化学、医学・バイオ、航空、住所）の分野辞書を搭載して翻訳精度を高めています。また、「Section 1」、「第1章」といった数字部分を可変にして辞書登録したり、日英翻訳では、カタカナ表記の微妙な違いを認識することもできます。

(13) 英文入力に英論文作成支援機能

既にある日本語を英語入力する際に、単語の先頭数字文字を入力すると元の日本語から単語を推測して候補をリストします。また、和文の読みをローマ字入力することでも候補となる単語が表示されるので、辞書を引く手間を省けます。

(14) 研究社のリーダーズ英和辞典、新英和・和英中辞典で EPWING 検索

翻訳家を始め広く使われている上記辞典を電子辞書として搭載しています。対話翻訳の画面や専用のツールで単語検索にご利用いただけます。

(15) 高機能 OCR ソフト ExpressReader を搭載

紙の文書を翻訳するために、OCR ソフト ExpressReader を搭載しました。印刷文書をテキストに変換する作業の軽減がはかれます。

(16) 英日／日英特許翻訳オプション

本オプションを「The翻訳プロフェッショナル V 7.0」に追加インストールすると、従来の翻訳ソフトでは対応が難しい、独特の表現を持つ特許文書（明細書など）を、特許文書向けの設定で翻訳できるようになります。特許独特の文を翻訳しやすい単位に分割し、特許専用文法で解析・翻訳するため、特許文書の翻訳品質を向上させることができます。本オプションは、英日101万語、日英102万語の語彙を搭載した専門用語辞書も装備しています。

3. 動作環境（全て日本語版）

対象機種：PC/AT 互換機、PC98シリーズ

CPU：Pentium®以上及び Celeron®以上を推奨

メモリ：40MB 以上（NT/2000の場合は48MB 以上）

ディスク：最小90MB～フルインストール460MB

OS：

Windows®98/98SE/Me,

Windows NT®Server/Workstation4.0（SP 6 a 以上）,

Windows®2000Server/Professional,

Windows®XP Home Edition/Professional

対象ソフト：

Word, Excel, PowerPoint®, FrontPage®, Access,

Internet Explorer, Outlook®, Outlook®Express,

Netscape 6, Netscape Navigator,

Netscape Collabra, Netscape Messenger,

Becky! インターネットメール*, EUDORA Pro*,

Lotus Notes, AOL Instant Messenger, ICQ,

NetMeeting, LaLaVoice2001, ViaVoice,

一太郎, Adobe Acrobat, Adobe Acrobat Reader,

*自動受信メール翻訳のみ

※対象ソフトの詳しいバージョンについては製品情報 URL をご覧ください。

●製品情報 URL

<http://www.hon-yaku.toshiba.co.jp/>

●カタログご請求先 URL

<https://cn.toshiba.co.jp/reg/honyaku/form.htm>

●Microsoft、Windows、Outlook、FrontPage および PowerPoint は米国 Microsoft Corporation の米国およびその他の国における登録商標です。

●Netscape は米国 Netscape Communication Corporation の米国およびその他の国における登録商標です。

●Pentium、Celeron は Intel Corporation の米国およびその他の国における登録商標または商標です。

●Adobe、Adobe Acrobat は、Adobe Systems Incorporated（アドビシステムズ社）の商標です。

●TRADOS はトラドス・ジャパン（株）、TRADOS Corporation の商標です。

●The翻訳プロフェッショナル、The翻訳は株式会社東芝の日本における商標です。

●その他、本文中の商品名称はそれぞれ各社が商標として使用している場合があります。

ATLAS V 8

富士通株式会社

1. はじめに

「ATLAS」は従来より企業、官庁、学校関連の「実務翻訳」の現場から圧倒的なご指示をいただいております。今回、翻訳品質・機能性をアップし「ATLAS V 8」として登場です。今回のバージョンアップでは新たに、PDF ファイル翻訳、メール翻訳等の機能が加わりました。

業界最高水準の翻訳品質

「ATLAS V 8」では、翻訳ソフトの要でもある翻訳品質をあげるため、「基本辞書」と「翻訳エンジン」の両面で改善を図りました。「基本辞書」は、英日56万語、日英58万語を収録。また、「翻訳エンジン」にも大幅な改良を加え、16%以上の翻訳精度向上を実現いたしました。

2. 機能

対訳エディタ

対訳エディタは、翻訳したい原文と翻訳結果となる訳文を左右に表示し、編集作業を行いながら翻訳を行うための機能です。新規単語の登録や、ユーザー独自の翻訳例の登録などの多彩な機能で、効率のよい翻訳作業を実現します。

Microsoft Office のファイルや PDF ファイルも直接読み込むことができます。

・ 訳語選択と学習機能

単語の訳を変更したい場合、訳語一覧の中から



対訳エディタ（訳語選択）

別の訳語を選択することができます。ここで変更を行った単語の訳は、次回以降の翻訳時に優先的に訳語として出力されるようになります。同様に訳例登録も行えます。

・ 確認翻訳

ATLAS V 8では新たに、「確認翻訳機能」を搭載しました。訳文上で右クリックをし、確認翻訳を選択することで、翻訳結果の有効性を検証することができます。

ホームページ翻訳

Internet Explorer または Netscape と連携して、日本語のホームページを英語に、英語のホームページを日本語にクリックひとつで翻訳します。より簡単な操作方法として、Internet Explorer 5.01以上がパソコンにインストールされている場合、Internet Explorer のツールバーに ATLAS の翻訳機能がアドインされるようになりました。



ホームページ翻訳

MS-Office/Acrobat との連携

ATLAS は実務で頻繁に使われるアプリケーションとの連携機能を搭載しています。Microsoft Office (Word、Excel、PowerPoint) や OASYS、一太郎な

どのワープロソフトとの連携機能に加え、今回新たに、Adobe 社の Acrobat とも連携し、PDF ファイルも翻訳できるようになりました。原文ファイルのレイアウトを変えることなく、また、テキストボックスなどのオブジェクトや表の内容もそのまま翻訳されるので、効率的に訳文ファイルを作成できます。



Acrobat 連携翻訳

メール翻訳

受信したメールや作成中のメールを手軽に翻訳できるようになりました。お使いのメールソフトが Outlook Express または Lotus Notes Mail の場合、ATLAS インストール後、メール翻訳のツールバーが追加されます。



メール翻訳

マウス翻訳

読みたいところにマウスポインタを合わせ、クリック一回で即座に翻訳できるようになりました。ほとんどのアプリケーション上でお使いいただけます。



マウス翻訳

キータイプ翻訳

入力した文章を即座に翻訳し、テキストエディタや電子メールなどに訳文を挿入できる機能が加わりました。入力と同時に翻訳が実行されるので、Eメールやチャットのときも大変便利です。

ファイル翻訳

複数のファイルを指定し、自動的に連続翻訳することができます。テキスト、Microsoft Office、OASYS、一太郎、HTML ファイルなど従来のファイル形式に加え、PDF 形式のファイルにも対応しました。パソコンの空き時間を利用すれば、大量の文書でも効率よく翻訳できます。

単位換算機能

通貨、長さ、面積、体積、温度、速度などの単位を、各国の単位へと換算して表示します。

専門用語辞書

専門分野向けに25分野、英日211万語・日英215万語の専門用語辞書をオプションで用意しております。また、ATLAS V8本体の「電子辞書」用の専門用語辞書データも5分野227万語を収録しています。

専門用語辞書一覧

分野	英日	日英
1. 情報処理	135,000語	137,000語
2. 電気・電子	92,000語	93,000語
3. 物理・原子力	165,000語	16,7000語
4. 機械	76,000語	77,000語
5. 工業化学	128,000語	130,000語
6. プラント	61,000語	63,000語
7. 土木・建築	51,000語	53,000語
8. 金属	43,000語	45,000語
9. 地学・天文	69,000語	71,000語
10. 輸送	59,000語	60,000語
11. 自動車	41,000語	42,000語
12. 軍事	24,000語	26,000語
13. 農林水産	67,000語	69,000語
14. 生物	158,000語	160,000語
15. [医学] 生化学	141,000語	143,000語
16. [医学] 薬学	62,000語	63,000語
17. [医学] 解剖学	54,000語	55,000語
18. [医学] 疾患・症状	179,000語	180,000語
19. [医学] 精神医学	29,000語	30,000語
20. [医学] 医療機器	51,000語	53,000語
21. 金融・経済	52,000語	54,000語
22. 法律	8,000語	9,000語
23. ビジネス	130,000語	132,000語
24. 人名・地名	204,000語	205,000語
25. 環境	31,000語	33,000語

3. 動作環境

動作 OS : 日本語 Windows XP/Me/98/95/2000/
NT4.0

必要メモリ: 32MB 以上 (64MB 以上推奨、OS 含む)

必要ディスク容量: 翻訳ダブルパック 187MB
EJ/JE 182MB

4. 価格 (税別)

ATLAS EJ V 8 58,000円
ATLAS JE V 8 58,000円
ATLAS 翻訳ダブルパック V 8 88,000円
ATLAS 専門用語辞書フリーパック V 8 78,000円
ATLAS 翻訳スーパーパック V 8 128,000円

既に「ATLAS」の旧版をお持ちのお客様には、バージョンアップ製品を販売いたします。詳しくは下記の URL をご参照ください。

<http://software.fujitsu.com/jp/atlas/versionup.html>

■お問い合わせ先

富士通株式会社

ソフトウェア事業本部ビジネス統括部 マーケティング部

〒100-8143 東京都千代田区大手町 2-6-2

TEL: 03-3548-3795

受付時間: 9:00~12:00/13:00~17:00

(土日祝日除く)

URL: <http://software.fujitsu.com/jp/atlas/>

事務局だより

AAMT 会員による著書のご紹介

●「ブロードバンドで学ぶ英語」 光文社新書

著者: ジョナサン・ルイス (一橋大学助教授)

ホームページで聴いて、学ぶ——ブロードバンド時代、英語を学ぶには最適の環境が整った。インターネット・ラジオ局のホームページを開き、ボタンをクリックするだけで、世界中の放送が聴ける。初級から上級まで、語学学習のポイントと、自分に合ったインターネット・ラジオ局の選び方を具体的に紹介する。「インターネット・ラジオで人生が変わった!」著者が聴きまくった末の徹底ガイドで、自分に合ったインターネット・ラジオ局を探そう。

定価 (本体700円+税)

ATLAS 翻訳パーソナル2002

富士通株式会社

1. はじめに

本格的ビジネス翻訳ソフト ATLAS V 8 と同じ翻訳エンジンを継承し、個人ユーザを対象にした「ATLAS 翻訳パーソナル」がバージョンアップいたしました。

新製品の「ATLAS 翻訳パーソナル2002」は、最新の時事用語、固有名詞など約30万語増の英日56万語、日英58万語を収録し、業界最大語数の基本辞書と業界最高水準の翻訳品質をご提供します。グローバルビジネスの現場で、もはや必須となる「海外のホームページ」や、「英文メール」の翻訳など、スピード重視の翻訳作業に威力を発揮します。

2. 機能

シンプル&スピーディな操作

「ATLAS 翻訳パーソナル2002」は、ATLAS V 8 の翻訳エンジンを継承し、シンプル&スピーディな操作がコンセプトです。ご家庭や職場でも簡単に使える翻訳ソフトです。



インターネットアップデート

インターネットアップデート

ATLAS 翻訳パーソナル2002をご利用のお客様は最新の時事用語が入った辞書や便利なツールをホームページ上から無料でダウンロードできるようになりました。名前と E-mail アドレスを登録するだけで翻訳辞書の自動更新ができ、最新の用語を用いた翻訳を実現します。

ホームページ翻訳 (IE 連携)

日英・英日のホームページ翻訳が一層便利になりました。Internet Explorer のツールバーに ATLAS の翻訳機能をアドイン。Internet Explorer 5.01以上がパソコンにインストールされている場合は、ブラウザを起動するだけで、翻訳ソフトを意識せずに翻訳が行えるようになりました。



ホームページ翻訳 (英日翻訳の例)

韓日ホームページ翻訳

Internet Explorer 上で、韓国語ホームページを日本語に翻訳できるようになりました。インターネット上の Web ページはレイアウトをそのまま保存して翻訳します。サッカー辞書もついていますので、今話題の海外のサッカー選手の名前も正しく訳せます。



ホームページ翻訳（韓日翻訳の例）

◇クイック ATLAS◇

ATLAS にはタスクバーに常駐し、簡単な操作ですばやく ATLAS の翻訳をご利用いただける「クイック ATLAS」機能があります。以下にその機能のいくつかをご紹介します。

■メール翻訳

Outlook Express または Lotus Notes Mail の場合、翻訳ツールバーが自動起動し、メールをすばやく翻訳します。作成中のメールは、本文中に入力した文章を翻訳結果と置き換えます。受信したメールの翻訳結果は、別ウィンドウで表示します。

■キータイプ翻訳

「入力した文字列を即座に翻訳したい」という時に、文字入力と同時にリアルタイムに翻訳を実行し、その結果を入力中のウィンドウに貼り付けます。英文のメールを書くときなど大変便利です。

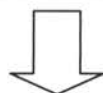
■マウス翻訳・マウス辞書

「この部分を翻訳してみたい」という時に、マウスクリックのみですばやく翻訳結果や辞書引きの結果を表示します。メール、ホームページ、ワープロなどあらゆるアプリケーションでご利用いただけます。

■クイックファイル翻訳

エクスプローラの右クリックメニューに「ATLAS

で翻訳」を追加します。選択したファイルを開くことなく即座に翻訳できます。



キータイプ翻訳の例

ATLAS 翻訳パーソナル+ビジネス用辞書バック2002

ATLAS 専門用語辞書フリーバック V8 の中から、特に人気の高いビジネス向けの辞書をセレクトしました。

ATLAS 翻訳パーソナル+ビジネス用辞書バック2002

分 野		収録単語（語）	
		英日	日英
IT 分野	情報処理	135,000	137,000
	電気・電子	92,000	93,000
	ビジネス	130,000	132,000
人名・地名	人名・地名	204,000	205,000
学術分野	環境	31,000	33,000
	生物	158,000	160,000
	農林水産	67,000	69,000
	地学・天文	69,000	71,000

3. 動作環境

動作 OS : Windows XP/Me/98/95/2000/NT4.0
 必要メモリ : 32MB 以上 (WindowsXP では64MB 以上)
 ディスク容量 : 187MB 以上（本体のみ）

4. 価格（税別）

ATLAS 翻訳パーソナル2002 8,800円

ATLAS 翻訳パーソナル+ビジネス用辞書パック2002
29,800円

下記対象製品をお持ちのお客様には、下記のとおり、バージョンアップキットをご提供いたします。

ATLAS 翻訳パーソナルグレードアップキット2002
4,800円

*対象製品：

・ATLAS 翻訳パーソナル V1.0

・翻訳サーフィン V1.0～V6.0

OASYS2002に添付されている

・ATLAS 翻訳パーソナル V1（体験版）

以下の他社ソフト

・AI ソフト／訳せゴマシリーズ

・ログヴィスタ／一発翻訳シリーズ

・ソースネクスト／本格翻訳シリーズ

・シャープ／翻訳これ一本シリーズ

・東芝／The 翻訳インターネットおまかせ翻訳

・IBM／インターネット翻訳の王様シリーズ

・アスキー／翻訳ピカイチシリーズ

ATLAS 翻訳パーソナル+ビジネス用辞書パック
グレードアップキット2002 21,000円

*対象製品：上記対象製品に加えて以下の製品

・ATLAS 翻訳パーソナル+ビジネス用辞書パック
V1.0

・ATLAS 翻訳パーソナル2002

■お問い合わせ先

富士通株式会社

ソフトウェア事業本部ビジネス統括部 マーケティング部

〒100-8143東京都千代田区大手町 2-6-2

TEL：03-3548-3795

受付時間：9：00～12：00／13：00～17：00

（土日祝日除く）

URL：http://www.software.fujitsu.com/jp/atlas/

事務局だより

AAMT 入会のご案内

AAMT は、機械翻訳の発展を目的として、機械翻訳の研究者、開発者、製造者、利用者が集まった組織です。委員会による定期的な調査研究をはじめ、機関誌の発行、セミナーや講習会の開催など幅広く活動を行っています。

機械翻訳にご関心のあるすべての方にご入会をお勧めします。

*** AAMT 会員の特典 ***

1. AAMT ジャーナルと MT News International (MTNI) が購読できます。

会員には、機関誌である AAMT ジャーナル（年 2 回発行）と、上位機関である IAMT (International Association for Machine Translation) の発行する MT News International（年 3 回発行）が送付されます。購読料は年会費に含まれています。

2. 機械翻訳関連の最新情報がメールで送付されます。

会員向けのメーリングリストで、最新の機械翻訳関連の情報が送付されます。

MT 新製品の紹介、国際会議やフォーラムのお知らせ、新聞・雑誌の MT 関連記事の紹介など盛りだくさんです。

3. 機械翻訳に関する講演会、講習会を開催しています。

この講習会、講演会に無料、あるいは、割引した参加料で参加できます。

4. AAMT が組織する委員会や調査活動に参加し、機械翻訳や翻訳に関心のある方との交流を深め、知見を広めることができます。

特に、法人会員は、機械翻訳に関する言語資料の標準化活動、機械翻訳市場の動向調査などに参加したり、AAMT の定期刊行物、ウェブサイトを使った製品の紹介を行うことができます。

5. 関連機関の主催する国際会議の参加費に割引が適用されます。

IAMT の主催で隔年に開催される MT Summit をはじめ、AAMT、AMTA*、EAMT** の主催する会議やワークショップに会員割引価格で参加できます。

AMTA はアメリカ、EAMT はヨーロッパの姉妹機関です。

AMTA* : Association for Machine Translation in the Americas

EAMT** : European Association for Machine Translation

6. AAMT および関連機関の出版物が会員価格で購入できます。

【最近の出版物】

- ・「ー機械翻訳ー 21世紀のビジョン」

(会員 2,500円、非会員 3,500円 / A 4 版 340ページ日本語)

- ・ Compendium of Translation Software (by Prof. John Hutchins)

(会員無料、非会員 2,000円 / 85ページ 英文)

☞ 年会費は以下の通りです。

法人会員：入会金 1 口 10,000円、年会費 1 口 50,000円

個人会員：入会金 1 口 1,000円、年会費 1 口 5,000円

☞ ご関心のある方は、事務局までお問い合わせください。

入会申込書をお送りいたします。

アジア太平洋機械翻訳協会 (AAMT)

〒101-0062 東京都千代田区神田駿河台3-11

三井住友海上別館ビル 3 階

(社) 電子情報技術産業協会 (JEITA) 内

TEL : 03-3518-6418 FAX : 03-3518-6472

E-mail : aamt 0002@infotokyo.ne.jp

AAMT invites you to join

The Asia-Pacific Association for Machine Translation, AAMT, is an association committed to supporting the advancement of machine translation, and is composed of researchers, developers, manufacturers, and users in the field. Membership is open to anyone who is interested in machine translation.

*** * * AAMT Membership Benefits * * ***

1. AAMT Journal and MT News International (MTNI)

Members receive the AAMT Journal (published semi-annually by AAMT) and MT News International (published three times a year by AAMT's parent organization, IAMT = The International Association for Machine Translation). Subscriptions are included in the membership fee.

2. Latest news and information by e-mail

The latest news related to machine translation is sent to members via e-mail. It covers a wide range of topics including the latest products, announcements for related conferences, and MT-related articles from newspapers and magazines.

3. Discounts on lectures and courses

Members can attend lectures and courses held by AAMT at no charge or with reduced fees.

4. Networking with professionals who share your interest

Members can participate in regular committee meetings to exchange views and share ideas. Corporate members can participate in broader activities such as MT market surveys and standardization projects of language data. Corporate members also have opportunities to introduce their products in AAMT's publications and on its official website.

5. Reduced fees for conferences held by related organizations

Members can participate in conferences and workshops organized by AAMT, IAMT, AMTA* and EAMT** with reduced fees. The MT Summit is organized by IAMT every two years. AMTA* is our sister association in the Americas and EAMT** is our sister association in Europe.

AMTA* : Association for Machine Translation in the Americas

EAMT** : European Association for Machine Translation

6. Special member prices for publications of AAMT and its related associations

[Recent Publications]

- ・ Machine Translation – Vision for the 21 st Century– (340 pages in Japanese)
(Member price : 2,500 yen, Non-member price : 3,500 yen)
- ・ Compendium of Translation Software by John Hutchins, president of EAMT
(Member price : Free, Non-member price : 2,000 yen / 86 pages)

✉ Membership Fees

Corporate Membership – Entrance Fee : 10,000 yen, Annual Fee : 50,000 yen

Individual Membership – Entrance Fee : 1,000 yen, Annual Fee : 5,000 yen

- ✉ If you are interested and would like to have more information, please send e-mail to aamt.0002@infotokyo.ne.jp or contact :

Asia-Pacific Association for Machine Translation (AAMT)
c/o JEITA Mitsui Sumitomo Kaijo Bldg., Annex, 3F
3-11, Kanda-Surugadai, Chiyoda-ku, Tokyo
101-0062, Japan

Phone: +81-3-3518-6418

Fax : +81-3-3518-6472

編集委員長から

編集委員会委員長 井佐原 均

今回の「AAMT ジャーナル」はいかがでしたでしょうか？ これまでの AAMT ジャーナルとは少し誌面を変えてみました。

本号では機械翻訳に関わる最新技術や動向を十分な誌面を使って、深く解説する記事を書けることにしました。また、外国での機械翻訳の現状を紹介する記事も掲載しています。機械翻訳技術の現状を詳しく知りたいというニーズに少しでもお答えできればと思っています。

これまでの「AAMT ジャーナル」は日本語による記事がほとんどでしたが、アジア太平洋機械翻訳協会 (AAMT) は、その名前が示すように日本のみではなく、広くアジア太平洋諸国も含めた組織です。日本語による情報が減らないように心がけつつ、これからは英語による記事も増やすことを考えています。

今回はその第一歩として、英文の論文 2 件を取り入れました。

(① “MT at Leeds Univ., UK” (P8～P11) ② “Toward a Science of MT” (P12～P120))

これからも、誌面を新しくする試みを続けていこうと思います。“Machine Translation News International (MTNI)”と相補的に進めていくことも考えるべきかもしれません。

皆様からのご意見やご提案をお待ちしております。

AAMT
ジャーナル

No. 32
(Mar. 2002)

発行所	行地	アジア太平洋機械翻訳協会 〒101-0062 東京都千代田区神田駿河台3-11 三井住友海上別館ビル3F
		TEL : 03-3518-6418 FAX : 03-3518-6472
		E-mail : aamt 0001@infotokyo.ne.jp
		ホームページ : http://it.jeita.or.jp/aamt
編集委員会		井佐原 均 大倉清司 山端 潔 熊野 明
事務局		村田勇造 高田佳代子
印刷所		伸光写植印刷株式会社

