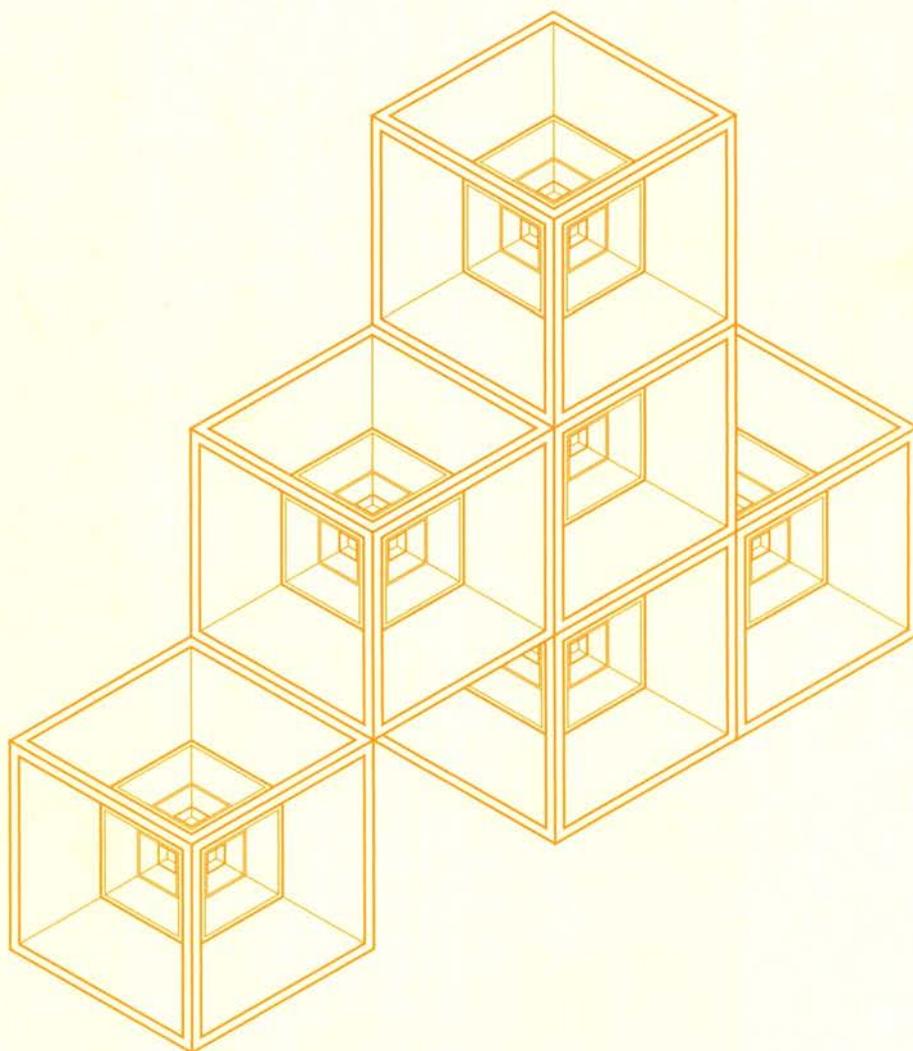


AAMT

Asia-Pacific Association for Machine Translation

Journal



December 2007 *No.41*

アジア太平洋機械翻訳協会

目 次

巻頭言：	第2回 AAMT 長尾賞を受賞して	隅田 英一郎.....	1
MT Summit 報告：	TMI 2007 と MT Summit XI に参加して	江原 暉将	2
	MT Summit XI 参加レポート	宮澤 信一郎.....	6
	MT Summit XI 参加報告	安田 圭志	9
	MT Summit XI and statistical machine translation.....	Jong-Hoon Oh.....	10
	TMI-07・MT Summit XI 参加報告	熊野 正.....	13
	TMI-07 と MT Summit XI に参加して	荒牧 英治	15
自社紹介：	SDL 社の自動翻訳ソリューション	小澤 清美	17
	機械翻訳の発展とともに	古賀 勝夫	20
研究紹介：	計算機支援による空の安全管理システム	寺田 昭.....	24
	KUI: A Self-organizing tool for multi-lingual WordNet construction	V. Sornlertlamvanich	31
イベント参加報告：	第17回日本翻訳連盟・翻訳祭パネル ディスカッションの報告	山本 ゆうじ.....	36
新製品紹介：	The 翻訳サーバ Enterprise Edition V3.0	東芝ソリューション(株)	40
	トラベルウォーカー 中国	(株)高電社.....	42
会員紹介：	AAMT 会員のひろば (第1回)		45
総会報告：	第17回通常総会および関連行事の報告.....		55
委員会報告：	英日機械翻訳に関するアンケート結果報告	AAMT 機械翻訳課題調査委員会	57
	AAMT Journal に関するアンケート集計結果.....	AAMT 編集委員会.....	65
活動報告：	協会活動報告 (2007年5月～11月)		69
編集後記：	宇津呂 武仁.....	71

CONTENTS

Foreword:	Message from the AAMT Nagao Award 2007 Winner	<i>E. Sumita</i>	1
MT Summit Report:	Report on TMI2007 and MT Summit XI	<i>T. Ehara</i>	2
	Report on MT Summit XI	<i>S. Miyazawa</i>	6
	Report on MT Summit XI	<i>K. Yasuda</i>	9
	MT Summit XI and statistical machine translation.....	<i>J. H. Oh</i>	10
	Report on TMI-07 and MT Summit XI	<i>T. Kumano</i>	13
	Report on TMI-07 and MT Summit XI	<i>E. Aramaki</i>	15
Message:	Automatic translation solution of SDL.....	<i>K. Ozawa</i>	17
	With the advance of MT.....	<i>K. Koga</i>	20
Report:	Computer aided aviation safety management system	<i>A. Terada</i>	24
	KUI: A self-organizing tool for multi-lingual WordNet construction	<i>V. Sornlertlamvanich</i>	31
Event Report:	Panel discussion at the 17 th JTF Translation Fair.....	<i>Y. Yamamoto</i>	36
New Product:	“The Honyaku Server Enterprise Edition V.3.0”	<i>Toshiba Solutions Corp.</i>	40
	“TRAVEL WALKER CHINESE”	<i>Kodensha Co. Ltd.</i>	42
AAMT Members:	AAMT Agora		45
General Meeting:	Report on the 17 th AAMT General Meeting and related events.....		55
Committee Report:	Questionnaire on English-Japanese MT	<i>Committee for seeking future direction of MT57</i>	
	Questionnaire on AAMT Journal.....	<i>Editorial Committee</i>	65
AAMT Activities	AAMT Activities (May 2007～November 2007).....		69
Editor's Note:	Message from the Chair of the Editorial Committee	<i>T. Utsuro</i>	71

第2回 AAMT 長尾賞を受賞して —これまでの MT とこれからの MT—

国際電気通信基礎技術研究所(ATR)
自然言語処理研究室室長 隅田 英一郎

1986 年の ATR 創立以来、音声翻訳一筋に研究・開発・事業化してまいりました成果が評価され、第 2 回 AAMT 長尾賞をいただきました。OB も含め関係者一同大変光栄に存じております。

ジャーナルの巻頭の 1 ページをいただきましたので、MT の研究・開発・実用化について、これまでの 4 半世紀を振り返り、これからの 4 半世紀に思いを馳せてみようとおもいます。

この間に 3 つの大波が地球を伝播しました。82 年に科学技術庁の機械翻訳プロジェクトの成功をうけて、ルールベースの MT の研究・開発が普及し、ブラビスの商用翻訳ソフトの発売を機に、各ベンダーの MT も商用化されました。翻訳品質の改善に専門用語の充実が有効であったので、規模は数万から数百万まで拡大しました。厳しい競争の中で、価格は、数百万から 1 万まで下がり、最近では、2000 円を切るものもあります。

81 年に長尾先生が提唱された用例ベース翻訳が、90 年前後に、京大、ATR で行われた研究をきっかけに、二つ目の大波として、日本だけでなく世界へ広がって行きました。この考えは、ルールベースの商用システムに取り入れられ、さらに、現在 NICT が中心になり実施している日中翻訳プロジェクトの基本方式として採用されています。

88 年に IBM が提唱した統計ベース翻訳は、難解な原論文、非力な計算機、対訳コーパスの不足、特許明細書でしか開示されなかった実行方法、類縁言語以外には有効でない方式限界のために、長らく鳴かず飛ばずでしたが、フレーズベース方式の提案、学習系・実行系のオープンソース化、コーパスの充実、100 倍の計算機能力を追い風に、第 3 の大波が起きました。今、研究分野では 9 割は統計ベース翻訳の論文です。バブルがピーク・アウトしているのか、まだまだ、伸びるのか、この局面での判断は難しいところです。

今、ちょうど、三つの大波が重なって、ルールベースと用例ベースと統計ベースの短所・長所がわかってきました。単一の方式では駄目だというのが現時点の妥当な見解です。3 方式をうまく融合できたときに最高の性能が実現するでしょう。

ここまでの機械翻訳には共通の課題があります。すべて文単位の翻訳であり文脈情報が利用できていない点です。これを実現するためには、頑健で計算効率が良い文脈処理の技術が不可欠ですが、まだ、技術の種もわからない状態です。次の 4 半世紀の取り組んでいくべき最重要課題の一つでしょう。また、音声翻訳については、五月雨式に翻訳する同時通訳技術が大きな課題として、まだ、未着手で残っています。

次の 4 半世紀も、翻訳・通訳という知的作業を自動化するための研究・開発・事業化の種は尽きません。ファイトが湧いてくるではありませんか？

TMI 2007 と MT Summit XI に参加して

諏訪東京理科大学 システム工学部
電子システム工学科 江原 暉将

機械翻訳関係の国際学会である TMI 2007 と MT Summit XI に参加したので、筆者が聴講した部分を中心に主観を交えて報告する。なお、参考文献の内容については Machine Translation Archive などを参照されたい。

1. TMI 2007

TMI 2007(11th International Conference on Theoretical and Methodological Issues in Machine Translation) は9月7日から9日にかけてスウェーデンの Skövde(ショブデ)で開かれた。Skövde はストックホルムから列車で2時間のところにある人口3万人程の静かな町で会議をするには絶好の場所である。さて会議は50名くらいの参加者を得て、Högskolan i Skövde(ショブデ大学)で開かれ、2件の招待講演、16件の口頭発表、12件のポスター発表で構成された。

招待講演1はウプサラ大学言語哲学科の Anna Sågvall Hein 計算言語学教授による Rule-based and Statistical Machine Translation with a Focus on Swedish であった。Hein 教授はスウェーデンにおける機械翻訳の草分けである。スウェーデン語と英語の間の機械翻訳について、規則方式(RMT)と統計方式(SMT)の得失を詳しく説明した。ヨーロッパ言語間の翻訳であるため、日英間とはまた違った得失が挙げられていた。またスウェーデン語で書かれた大学のシラバスを英語に翻訳するための機械支援翻訳システムについても話した。高等教育の国際化が進む中、教育分野での翻訳需要の高まりが感じられた。

招待講演2はアーヘン工科大学の Hermann Ney 教授による Statistical MT from TMI-1998 to TMI-2007: What Has Happened? であった。1998年にIBMのPeter BrownがSMTを発表して以来、SMTが機械翻訳研究の主流の一つになっている。講演ではSMTの歴史を振り返り、その構成要素を Training, Phrase extraction, Definition of the log-linear model, Generation or search の4つに

分けて現状を解説した。また最近のSMTは構文情報を利用したり処理単位を長くしたりなど、用例ベース翻訳(EBMT)やメモリーベース翻訳(MBMT)に接近してきていると分析した。

一般発表では、SMT(アライメントを含む)が14件、翻訳評価が4件と、この分野の活発さが目に付いた。

統計翻訳(SMT)では、ヨーロッパ言語間あるいは中英間のもが多く、句をベースにした方式が主であった。[Carpuat]では、句の意味を考慮して精度を向上させている。[Stroppa2]では、用例方式の特徴を素性としてSMTに取り入れている。[Rottmann]では、独英間などで生ずる遠距離の語順変更を前処理でおこなっている。[Andrés-Ferrer]では、ペイズ分類に基づいて定式化した各種の素性を組み合わせている。[Fung]では、2言語の意味役割を学習してSMTで用いている。[Carl] [Meler] [Dirix] [Sofianopoulos]では、翻訳先の単言語コーパスと対訳辞書だけを用いたSMT(METIS-ID)を提案して種々の言語対に適用している。[Sanchis]では、語順の異なるスペイン語バスク語間において前処理段階で目標言語の語順にそろえている。[Buch-Kromann]では、構文木を用いたSMTで交差が生じ処理ができなくなる場合を克服する方式を提案している。

アライメント(対応)の主な発表は以下の通りである。[Ushioda]では、2言語の構文解析を用いて句対応を求めている。[Kumano]では、ParallelではなくComparableな2言語コーパスから句対応を学習させている。[Hearne]では、句ではなく構文木の対応を求めている。

翻訳評価の主な発表は以下の通りである。[Mehay]では、構文的類似性を評価に盛り込む方式を提案している。[Paul]では、人手での評価結果を機械学習によって自動評価に利用している。

その他の発表の主なものは以下の通りである。[Aramaki]では、サポートベクターマシン(SVM)を用いて異表記を判定している。[Stein]では、手話の

翻訳に MT を利用している。[Somers]では、医療分野に NLP を利用している。

また、Is MT In Crisis?と題してパネル討論が行われた。討論に先立って MT が危機にあるかどうかを司会者が会場に問うたところ、そう思う聴衆は 2 名しかおらず、MT 技術が危機であるという議論にはならなかった。むしろ MT 開発会社や MT 研究のバブル状態が危機であるという話が出た。また中国の寄与が大きくなっているのが危機であるといった(半ば冗談の)話もあった。これに対して中国からの参加者が「13 億の人口があるから寄与が大きくて当然である」といなしていた。最後にパネリストからこの 10 年の進歩と今後のブレークスルーについて話があり、SMT が発展した点が進歩であり、今後は意味の利用など深い処理が期待されるといった話で締めくくられた。



TMI 2007 で招待講演する
Anna Sågvald Hein 氏

2. MT Summit XI

MT Summit XI は 9 月 10 日から 14 日にかけてデンマークの Handelshøjskolen i København (コペンハーゲン・ビジネス・スクール、CBS) で開かれた。10 日は 6 件のチュートリアル、11 日は 4 件のワークショップが行われ、その後の 3 日間は本会議であった。

2.1 チュートリアル

統計方式機械翻訳(SMT)のチュートリアルでは、Philipp Koehn 氏が句ベースの SMT について Kevin Knight 氏が構文情報の利用について話した。2 人とも熱のこもった講演を行い、SMT に対する気持ちの入れようが伝わってくる良いものであった。その中で Knight 氏が、「構文情報を使う

のはばかげている」という批判にもめげず研究を進め、最近になってその効果が現れてきていることを示していた。時には良いアイデアを導入しても性能がかえって下がることもあると、研究の難しさを示していた。

2.2 ワークショップ

特許翻訳のワークショップでは、世界知的所有権機関(WIPO)の Rachel Chrem 氏が招待講演を行い、ユーザー報告として欧州特許庁(EPO)の Wolfgang Taeger 氏と日本特許庁(JPO)の遠山敬彦課長補佐から報告があった。また日本特許情報機構(Japio)の林昭彦理事長から挨拶があった。一般講演は 6 件であり、参加者は 35 名であった。

Chrem 氏の講演では、WIPO の活動の紹介とその中での翻訳の役割について話した。IPC(International Patent Code)関係の文書の翻訳や PCT(Patent Cooperation Treaty)での翻訳の現状、PCT が収集したデータを利用する Patentscope というサイトについて解説した。PCT の中で作成している日中アラビア英独仏西の 7 言語ターミノロジーデータベースが特に興味深かった。これは出現した文脈を参照できるデータベースとなっている。

Taeger 氏は EPO での MT の利用状況と EMTP(European MT Program)について話した。前者では、日本および韓国が提供する日英と韓英のシステム出力を EPO で利用している状況を説明した。後者は、2004 年から開始したプログラムであり現時点では独西⇄英について実用している。次期には仏伊葡蘭露スウェーデン⇄英について実用化し、将来的には独⇄仏や非ヨーロッパ言語、特に中国語を対象にすることが考えられると話した。

遠山氏は JPO での MT の開発/利用状況について話した。特に IPDL(Industry Property Digital Library)や AIPN(Advance Industry Property Network)とそれらの中での MT の利用について説明した。IPDL は誰でも使える特許検索システムであり、AIPN は 29 カ国の特許庁に審査関連の情報を提供するものである。いずれも日英の MT が利用されている。

一般講演では、英韓、日英、英デンマークの MT 技術に関するものや請求事項を効率的に翻訳する支援システムに関する発表などがあつた。また関連して、展示ブースでは韓国電気通信研究院(ETRI)による英韓特許翻訳システムのデモが行われた。

MT が実用されている特許翻訳の分野は重要であり、次回の MT Summit でも特許翻訳ワークショップを継続しようということで締めくくられた。

2.3 本会議

さて、MT Summit の本会議は 4 件の招待講演、2 件のパネル、45 件の口頭発表、29 件のポスター発表で構成された。投稿論文数は 115 件とのことである。参加者は 37 カ国からの 271 名であった。

招待講演 1 はスイスの CLS Communication の Doris Marty-Albisser 氏による Client Centric Multilingual Information Leveraging –Scale the Skills であり、招待講演 2 は Microsoft Research の Stephen Richardson 氏による Microsoft MT: From Research to Real User であった。いずれも、国際企業での社内コミュニケーションや顧客とのコミュニケーションに、MT がどのように役立っているかについて講演した。CLS の報告では、ある企業では社員の 80% が週に 1 回以上 MT を利用しており、30% は毎日利用している、また使用者の 80% は MT の性能が業務に役立つレベルにあると考えているとのことであった。Microsoft からは、MT の利用により翻訳コストが 5~25% 削減できたと報告された。また最近発表したウェブ翻訳システム Windows Live についてのデモを交えた紹介があった。

招待講演 3 はエディンバラ大学の Philipp Koehn 氏による EuroMatrix –Machine Translation for all European languages であった。EuroMatrix は EU 議会での討論の内容を EU での使用言語である 11 言語間で翻訳することを目指したプロジェクトで、SMT をベースにし、研究成果を公開することを原則にしている。SMT の利点は訳語選択を精度良くできる点であり、巨大なデータを用いることで、少数例や慣用句にも対応できるようになった。一方、欠点である形態論や構文論の利用が難しい点を克服するために規則方式と組み合わせることが始まっている。組み合わせには以下のようなものがある。固有表現や数量表現など SMT で扱いにくい部分を RMT で事前に翻訳しておく。長距離の語順変更を規則方式の前処理で済ましておく。RMT に統計的後編集を組み合わせる。構文木そのものを SMT の枠組みで扱う。結論として、SMT は、語、句、構文、意味といった処理単位の階層を上がっている。これは RMT がかつて辿った道をもう一度辿

っていることになる。昔と違うのは、データ量が多くなったこと、計算機パワーが大きくなったこと、数学的基礎付けがしっかりしてきたことである、として締めくくられた。

招待講演 4 は NEC の奥村明俊氏による NEC Human Communication Technology: Speech Translation for Hand-held Devices であった。旅行会話を対象とした音声翻訳システムの歴史を振り返り、デモを交えてその進歩を紹介した。現在では空港や旅客機の機内での実証実験をするまでに進歩してきている。またブログ作成支援という場面で、ロボットと音声対話をするという内容のビデオを紹介した。かわいらしいロボットの発話は日本語で行われたが英語字幕を通じて十分伝わってくるものであり、会場の爆笑を誘っていた。総じて日本の得意分野である音声翻訳技術や娯楽性を伴ったロボット対話技術が良く紹介されていた。余談になるが、バンケットでの個人的会話の中で「有名な日本文化である漫画を機械翻訳してはどうか」という話が出た。人手翻訳が待ちきれない外国の読者がいるようである。

一般発表で筆者が特に興味を持ったものは以下のとおりである。

[Babych]は、利用できるリソースが少ない小言語に関する MT を開発するに当たって、当該小言語に近い大言語を中間言語とするシステムが有効であることを、ウクライナ語から英語への翻訳でロシア語を中間言語とする例で示した。

[Echizen-ya]は、翻訳結果と基準翻訳を比較するに当たって、共通部分の長さだけでなく、その出現位置も考慮する新しい翻訳評価尺度を提案した。また、この尺度は共通部分の検出を再帰的に行うことが特徴の一つである。

[Davis]は、語のアライメントの精度を測る新しい尺度を提案し、その尺度で測ると最終的な翻訳精度との相関が高いことを示した。

[Kirchhoff]は、翻訳対象の様々な性質(例えば、対象の属する分野やメディア、未知語率など)と翻訳結果の人手評価との相互情報量を計算し、相関の多寡を求めた。ここでの中英、アラビア英での実験では属するメディア(News wire, News group, Broadcasting news, Broadcasting comment の 4 分類)が最も相関が高かった。本研究に限らず、現在、米国で盛んに行われている中英とアラビア英の翻訳に関する発表が散見された。中英に比較してア

ラビア英の精度が相対的に高く、その理由がいろいろと議論されていた。欧州にとって中国よりもアラビアの方が文化的に近いことが一因なのかもしれない。

[Déchelotte]は、標準的 SMT デコーダのひとつである Moses をベースに改良を加えた。改良点の一つはニューラルネットを用いた連続的言語モデルである。余談になるがこの講演の途中でプロジェクターに不具合が生じ、直前に発表した Systran 関係の人が手伝って回復することができた。日ごろ研究上でしのぎを削っている SMT と RMT の研究者が助け合う場面であり、ほほえましかった。

[Gamallo]は Parallel ではなく Comparable なコーパスから対訳辞書を構築する手法を述べた。Comparable corpus は Parallel corpus に比較して大量にあるので、この技術は重要である。本発表の特徴は少量の Parallel corpus からテンプレートを抽出し、そのテンプレートを用いて Comparable corpus から対訳辞書を抽出するところにある。[Quirk]でも同様の発表をし、こちらは Parallel corpus から単語対応を求め、それを利用して Comparable corpus から対応する表現を求めるものである。

[Singh] では、UNL(Universal Network Language)での中間表現からヒンディ語を生成し、マラタ語から英語を介してヒンディ語へ翻訳するシステムに利用している。

パネル討論は2回行われた。パネル1は Machine Translation in Use である。その中で井佐原均 AAMT 会長から言語グリッドの紹介がなされた。企業活動や国防に MT を実用するのも重要かもしれないが、人の心と心をつなぐ本来的なコミュニケーションに MT を実用することは、一層重要であると筆者は考える。言語グリッドはその好例であろう。この言語グリッドや医療場面での MT の利用などが情報通信研究機構の展示ブースで紹介されており、この分野での日本の寄与を印象付けていた。

パネル2は The Prospects of MT Today: When and how will the apparent progress lead to breakthroughs in technology and commercialization?であった。SMT の進歩が著しい中でも、依然として実用面では RMT が使われている。SMT を実用にするために、Dekai Wu 氏は Meaningful SMT を提案していた。このパネルは本 MT Summit の最後に行われ、今後の研究の方向性

が決まったような気がした。

2.4 IAMT General Assembly

3日目の朝には、IAMT General Assembly が行われた。その席で、次期会長に AMTA の Laurie Gerber 氏が、副会長には井佐原均氏が決まったことが発表された。次回 MT Summit はカナダで2009年に行われる予定である。



MT Summit XI 組織委員長の
Bente Maegaard 氏(手前)

MT Summit XI 参加レポート

秀明大学 英語情報マネジメント学部
教授 宮澤 信一郎

1. はじめに

11 回目の MT Summit が 9 月 10 から 14 日までデンマークのコペンハーゲンで EAMT (ヨーロッパ機械翻訳協会) の主催で開催された。私は、前回のブリーチでの MT Summit X で初めて開催され今回 2 回目となる、日本が中心となって開催する特許翻訳のワークショップ “Workshop on Patent Translation” への参加が主たる目的である。

コペンハーゲンは古い街並みが続く美しい街である。国土が日本の 10% 位であるが、標高が最も高いところでも 173 m しかなく平坦な土地が続いている。首都コペンハーゲンの人口は約 183 万人 (デンマーク全体で人口約 543 万人) で、そのためか全体的にゆったりした感じを受けた。予想以上にコンパクトな都市で徒歩でも主要な所にほとんど行けたのは便利だった。車は日中でもライトを点灯していた。北欧やカナダではエンジンを始動させると自動的に点灯する灯火が義務付けられていて大幅に交通事故が減ったと聞いていたが、全車が点灯したらさぞまぶしいだろうと思っていた。現地ですべてに見てさほどでもなかったのが、これで大幅に事故が減るのであれば日本でも導入する価値が十分ありそうだ。また主要道路に自転車専用レーンがあって多くの自転車が走っており、地下鉄にも自転車ごと乗ることが出来るなどエコロジーにも力を入れている様子である。小さな地下鉄の駅は無人駅で切符の自動販売機らしきものは無く、どこで切符を買ったら良いか迷ったが、ホームの中ほどに自動販売機があり、しかも現金は使えずクレジットカードのみであるのには面食らった。

2. 会議概要

会場はコペンハーゲンビジネススクールである。地

下鉄の駅のすぐ側で交通の便はすごく良かった。参加者は 37 カ国から約 270 名であった。10 日には 6 つのチュートリアルが開かれたが、私はワークショップの会場を下見しただけでチュートリアルには参加しなかった。



写真1 会場風景 (1階受付、2・3階は食堂)

11 日には 4 つのワークショップが開催された。私は当然 “WI: On Patent Translation” に参加していたので、他のワークショップの様子は分からない。このワークショップでは前回に引き続いて *Proceeding* の編集作成を担当した。表紙と印刷は EAMT 側が行ってくれるので、何度か EAMT の担当者とやり取りを行って仕上げについて注文をつけた。そのため現地に行くまでどのような仕上がりになっているか心配であったが、綺麗な仕上げになっていて満足した。現地スタッフに感謝したい。ISBN コードを付けてくれるよう依頼していたが、それは出来ないとの返事で、コードが付かなかったことだけが心残りである。ワークショップには日本、デンマーク、アメリカ、韓国などから 37 名の参加者があり会場はほぼ満杯でこの分野への関心の深さがうかがわれた。午前中は、オーガナイザである山形大学の横山先生の開会挨拶に続き、招待講演 1 件、一般発表 2 件が行われ、午後はユーザから

の講演2件、Japio（財団法人 日本特許情報機構）理事長からの挨拶、一般講演3件があり、質疑応答も活発で盛況に終了し安堵した。次回の MT Summit はカナダになったが、カナダでも引き続き特許翻訳のワークショップを開催する意義は十分あると感じた。その際はぜひ参加したいものである。

18 時からは Opening Ceremony と Welcome Reception が、近くにある Frederiksberg 市の荘重な City Hall で開催された。市長や主催者の挨拶の後、ビッフェスタイルでの会食が始まり、民族楽器の演奏も加わり、和やかな雰囲気の中、夜が更けた。

12 日は、Opening session に続いて、招待講演が2件あった。大手翻訳会社 CLS Communication による“Client centric multilingual information leveraging - Scale the skills”と Microsoft Research による“Microsoft MT: From research to real user”である。前者は MT や TM がビジネスにどのように使われているかを、実際に即して現状を述べたものである。後者はマイクロソフトの MT への取り組みについて述べたものであり、これから本格的に実用化を図る予定であるようだ。13 日にも招待講演が2件あった。エジンバラ大学 Philipp Koehn 氏による“EuroMatrix - Machine translation for all European languages”と日本電気の奥村氏による“NEC Human communication technology: Speech translation for hand-held devices”である。前者では EU 域内における多言語の機械翻訳プロジェクト EuroMatrix についての多くの試みが紹介された。後者は NEC の旅行会話用ハンドヘルド音声翻訳装置の紹介である。

その後3つのトラック（“R&D”が2つと“User and practical systems”が1つ）に分かれて一般発表が行われた。一般発表は同じトラックで13、14両日も開かれ総計45件の発表があった。4時からは10件の Language Resources and Evaluation に関するポスターセッションが開かれた。ポスターセッションは13日も開かれ、こちらは Methodologies に関して18件であった。

13日は前述した招待講演、ポスターセッション、一般発表の他に、パネルディスカッションが開かれた。これは“Machine Translation in Use”というテーマで行われ、コミュニケーション、SMT（統計的機械翻訳）、ユーザ、評価の各々の立場から活発な議論が行われた。パネルディスカッションは最終日の14日にも行われ、こちらは“The Prospects of MT Today: When and how will the apparent progress lead to breakthroughs in technology and commercialization?”のテーマで行われたが、ブレイクスルーが早期に起こる見込みはないとの意見で一致していた。



写真2 ポスターセッション

夜は Conference Banquet がコペンハーゲンの東側に位置するアメリンボー宮殿近くの格式のあるホテルで開催された。各種挨拶の後、ジャズの演奏を聞きながら晩餐に移ったが、多くの参加者がおり、パーティドレスの女性も多く華やかな雰囲気だった。ここからは昼であればアンデルセンの人形姫の像もすぐ近くである。その像の手前には函館の五稜郭に良く似た五角形の城跡があったので、五稜郭はここをモデルにしたのかと思ったが、帰国後調べてみるとフランスの城がモデルのようだ。

14日はIAMTの総会が開かれ次回の MT Summit が2009年にカナダで開催されることが発表された。その後、上記の一般発表、パネルディスカッションと続き無事に終了した。

3. おわりに

この会議の途中で火災警報が鳴り響き、コペンハーゲンビジネススクールの学生も含めた全員が、一時避難させられ屋外が人で溢れかえるハブニングもあったが、機械翻訳の研究の最前線や問題点、ユーザのニーズなどが明らかになり、世界の機械翻訳研究者、開発者、ユーザが一同に会して情報を共有しあったことは大変意義深い。今後も MT Summit がより発展して機械翻訳の発展に大きく貢献することを願っている。最後になるが、いろいろ御支援いただいた AAMT や Japio の方々に感謝申し上げる次第です。

MT Summit XI 参加報告

情報通信研究機構/ATR 音声言語コミュニケーション研究所

安田 圭志

MT Summit XI は、2007 年 9 月 10 日から 9 月 14 日までの期間、デンマークコペンハーゲンの Copenhagen business school に開催された。自身にとっては 2001 年にスペインで開催された MT Summit-VIII 以来の 2 度目の参加であった。

本報告書では、MT Summit の本会議に加え自身が参加、発表を行ったワークショップ Using Corpora in Natural Language Generation and Machine Translation (UCNLG+MT) の模様について報告する。

ワークショップ UCNLG+MT

ワークショップ UCNLG+MT は、主に、The Attribute Selection for Generating Referring Expressions (ASGRE) のシェアードタスク型評価の参加機関による発表と、UCNLG+MT に関する一般講演の 2 つからなっていた。シェアードタスクにおいては、以下の 6 機関からの参加があった。

- ・ Computer Lab, Cambridge University, UK
- ・ Dublin Institute of Technology, Ireland
- ・ Universities of Twente and Tilburg, NL, and Macquarie University, Australia
- ・ University of Stuttgart, Germany
- ・ Universidad Complutense de Madrid, Spain
- ・ Tokyo Institute of Technology, Japan

一般講演においては、4 件の自然言語生成に関する発表と、3 件の機械翻訳に関する発表があった。

その他にも当ワークショップでは、Kevin Knight (ISI) による招待講演や、「Language Generation and Machine Translation」というタイトルで、パネルディスカッションが行われた。これらのセッションでは、現在の統計翻訳における諸問題や、統計翻訳の研究者と自然言語生成の研究者が

どのような点において協力できるかなどについて闊達な議論がなされた。

本会議

招待講演

本会議においては、以下の 3 件の招待講演が行われていた。

○Stephen Richardson (Microsoft): マイクロソフト社で進められている MT 関連の研究紹介と、マイクロソフト社内における MT の利用方法や、その効果について述べられた。

○Philipp Koehn (University of Edinburgh): Euro Matrix プロジェクトの紹介と、現状におけるヨーロッパ系言語対での機械翻訳の性能などについて述べられた。

○Akitoshi Okumura (NEC): 翻訳関連のみならず、音声処理や、ロボットなどの幅広い分野の製品開発についての説明がなされた。

私は、普段より研究関連の職務に携っている関係上、製品開発に関する発表は非常に新鮮であり、また、企業における MT 関連技術のニーズを知る上で非常に参考となった。

なお、招待講演の発表資料については、会議参加者以外でも以下より入手可能のようである。

<http://mtsummitcph.ku.dk/invitedpresentations/>

一般講演

一般講演においては、近年の MT 関連の研究のトレンドを反映し、統計翻訳関連の発表が多くなされていた。また、6 年前の MT Summit と比較し、ユーザや開発者による研究発表の割合が少なく、研究者サイドの発表の割合が増えていたように感じた。

MT Summit XI and Statistical Machine Translation

Jong-Hoon Oh

National Institute of Information and Communications Technology

MT Summit XI was held at Copenhagen, Denmark from September 10 to 14, 2007. About 250 researchers visited Copenhagen, one of the most beautiful cities in Scandinavia, to participate in the conference. Six tutorials and four workshops were scheduled on 10th and 11th September, respectively. Four invited speeches, two poster sessions, fifteen oral sessions (three sessions were parallel with three different topics at the same time) and two panels were scheduled in the main conference from 12th to 14th September.

I attended one of the tutorials, “Statistical Machine Translation (SMT)” organized by Kevin Knight (University of Southern California) and Philipp Koehn (University of Edinburgh). Philipp Koehn, who developed MOSES (the well-known open source phrase-based SMT system), talked about phrase-based SMT in the morning session. He began his talk with the word-based SMT and went into details of phrase-based SMT, including decoding, phrase translation table, and phrase-based translation model. Kevin Knight was in charge of the afternoon session entitled with “Syntax-based SMT: Can a Machine Translate without knowing what a verb is?”. He gave a very funny picture where ACL language engineers claim “Syntax will never work because syntax has been shown to make things worse!” and AAAI fellows claim “Syntax will never work because language is about the world thus you need semantics!” However, he went on to claim that we can do machine translation without knowing what a verb is but we need some grammar to improve phrase-based SMT. Even though we were surrounded by a forest of syntax trees during the tutorial, his talk about syntax-based SMT was very concise and clear and helped us to easily understand how the syntax helps a SMT system to translate more accurately.

There were plenty of invited speeches and oral sessions dealing with SMT at MT Summit XI – two invited speeches and five oral sessions. The SMT papers presented in the sessions covered word, phrase, syntax, and partially semantics. Robert Moore presented distortion penalty estimation and early pruning to improve decoding speed in phrase-based SMT. Rens Bod presented a machine learning approach to syntax-based SMT – unsupervised data-oriented translation (U-DOT). Marine Carpuat and Dekai Wu gave a talk “Context-dependent phrasal translation lexicons for SMT”. Their talk was about how to improve phrasal-based SMT using *full-sentence* and *dynamic* contexts..

The discussion at the panel sessions was also interesting. Five panelists participated in “Machine translation in Use” moderated by Eduard Hovy. Hitoshi Isahara claimed that machine translation has been useful in information dispatch, gathering, and communication. He gave us an example of a chatting system supported by machine translation system enabling communication between users with different native languages.

Margaret King claimed that every MT research group needs to get a high BLEU score in order to get research funds. She also expressed her concern that this may prevent some fundamental MT research from taking place. Three other panelists gave a talk about trends of MT systems from the 1970s to today. During the panel session, the claim that “users should be demanding” was made:

“In the last decades, machine translation systems become on-line and free of charge. We use the on-line machine translation systems in daily life. But we rarely complain of translation quality of the on-line translation system to the company or demand higher translation systems quality because it is free. Is it okay? Are you happy? Users should be demanding.”

Even though I use on-line MT systems frequently, I never recognized myself as the same inactive user as described above.



Fig.1 Panel “Machine translation in Use”

The second panel “The Prospects of MT Today: When and how will the apparent progress lead to breakthroughs in technology and commercialization?” moderated by Hans Uszkoreit was before the closing session of MT Summit XI. The moderator talked about the impressive progress in MT but he pointed out the skeptics’ claim “We are not much closer to a more powerful class of MT products than twenty years ago”. He gave us three questions seen from a viewpoint of commercialization.

- “Are there or will there be new commercially successful applications on the basis of recently created advanced technology or are there essential open problems that still need to be solved before?”
- “When and from where do you expect the needed solutions?”
- “Which approaches or mixes of approaches would best be suited for which types of MT applications?”

Gregor Thurmair, one of panelists, gave us an interesting opinion on the second question “When”. He claimed that “No major breakthroughs can be expected in the next five years as it was in the last decades. Instead, it will be slow but steady increase like in speech recognition”. Someone pointed out “commercial success and research success in machine translation are different”. Dekai Wu claimed that SMT should be “meaningful SMT”, in other words, self-organizing models to learn semantics for SMT. Even though panelists made a good attempt to answer the questions, time limits made it hard to do so fully.



Fig 2. Panel “The Prospects of MT Today: When and how will the apparent progress lead to breakthroughs in technology and commercialization?”

The MT Summit XI conference was closed with a short closing session. The MT Summit conference has given us a milestone in SMT research – SMT is still one of main streams in MT research and SMT paradigms are moving from pure phrase-based one to those based on syntax and semantics.

TMI-07・MT Summit XI 参加報告

NHK 放送技術研究所

熊野 正

9月10～14日にデンマークのコペンハーゲンで開催されたMT Summit XIと、その直前の7～9日にスウェーデンのショーブデ(Skövde)で開催されたTMI-07に参加してきました。私にとって初めて参加する機械翻訳の会議で、また初めての北欧旅行ということもあり、大変楽しく興味深い経験をすることができました。

TMI-07 (9/7-9, Skövde, Sweden)

TMI-07が開催されたショーブデは、スウェーデン南部、ストックホルムから鉄道で2時間ほどのところにある閑静な町です。ボルボの工場と軍の駐屯地、そしてこの会議の会場であるショーブデ大学がこの町の主な産業です。ショーブデ大学は1977年設立の比較的新しい大学で、市街地の中心部の緑あふれる広大な敷地に北欧的な洗練されたデザインの施設を擁する、美しいキャンパスがあります。

TMI (Theoretical and Methodological Issues in Machine Translation) は1985年に第1回会議が開かれ、今回で11回目となる国際会議です。チューニング競争から少し離れたところで機械翻訳の理論や手法をじっくりと議論することを目指し開催されています。今年も例年通り、参加者が全ての発表を聴き十分な議論をすることができるよう、口頭発表17件、ポスター・デモ12件と発表件数を絞り、シングルトラックセッションで十分な質疑時間をとるスケジュールが組まれました。参加者もそれほど多くなかった（正確な人数は把握していないが、60名程度？）こともあり、会議の雰囲気はとても親密で、“*Is MT In Crisis?*”と題したパネルセッションでもぎっくばらんな議論を楽しむことができました。

・ 私達のグループはこの会議で、統計翻訳技術を用



TMI会場のショーブデ大学

いて直訳でない文書対に表現対応を付与する手法に関する発表を行いました。

MT Summit XI (9/10-14, Copenhagen, Denmark)

MT Summitは、TMIの翌日より隣国デンマークのコペンハーゲンにあるコペンハーゲンビジネススクールを会場に開催されました。コペンハーゲンビジネススクールは世界中から学生を集める北欧最大規模のMBAスクールとのことで、ショーブデ大学の静かな佇まいとは打って変わって、モダンな校舎にたくさんの学生が行き来する傍ら、こちらも250名ほどの参加者が集う盛況な会議となりました。会議は10日に6件のチュートリアル、11日に4件のワークショップとレセプション、12～14日が本会議で45件の口頭発表と29件のポスター、4件の招待講演と2件のパネルがありました。

MT Summitは研究者だけのものではなく、翻訳システムメーカーや翻訳実務家に対しても開かれているという位置づけの下、チュートリアルやワークショップには機械翻訳システムの利用方法や製品開発法、特許翻訳についてなどのトピックが設けられていました。招待講演も実際にユーザに供されているシステムを紹介する話題が多く、また会議期間を



MT Summit 会場のコペンハーゲンビジネススクール

通じて、機械翻訳製品のデモ展示もありました。その一方で、本会議の発表内容には実務的なものはそれほど多くはなく、例えば翻訳支援ツールに関する発表が少ないことに不満を述べている参加者もいました。

両会議を通じての雑感

会議プログラムの発表タイトルを眺めると、もはや翻訳手法として統計翻訳を名乗らないものはほとんどないのではないかという状況になってきました。しかしその実態は、構文や、品詞などの語彙情報などの言語知識を取り入れたり、ヒューリスティクスをうまく取り入れたりする手法がほとんどで、これらは用例ベースやルールベースの翻訳手法とかなりの類似点を持っていると言えます。今後は、統計的な学習/決定手段をエンジンに、従来さまざまな翻訳手法で用いられてきた各種言語知識が統一的に扱われるようになり、統計・用例・ルールといった手法の分類はだんだん消えていくのではないかと想像しています。

また、翻訳性能評価手法についても、統計一辺倒のトレンドに変化が感じられました。従来統計翻訳の性能向上を牽引してきた BLEU などの自動評価手法は、それまで主流であった主観評価に比べると、手間はかからないが乱暴であり、妥当性に疑問を持つ人も以前より少なくなかったと思うのですが、今回、翻訳評価手法に関する発表件数が近年に多く

く、またパネルでも「意味があっているかを評価しなければならない」という主張がありました。今回、自動評価を言語的に豊かにする手法や、少ない回数的主観評価でシステムの性能を見積もる手法が提案されましたが、今後、評価手法の研究がどのように進むのか、興味深く見ていきたいと思えます。

機械翻訳技術は昔から着実に進歩していると思うのですが、それがなかなか本格的な実用に必要なブレイクスルーを超えられず、よいビジネスにならない、という問題は、私達機械翻訳研究者共通の悩みで、どちらの会議でも（ニュアンスは異なりますが）パネルの議題に上がりました。私は、あるパネリストの「機械翻訳は全自動翻訳だけを指すのではない」という発言に強く共感しています。今回の MT Summit ではひとところより寂しくなってしまった、機械翻訳技術を全自動翻訳以外の形で実用に供していくための研究が、次回以降また数多く発表されることを期待しています。

TMI-07 と MT Summit XI に参加して

東京大学医学部附属病院

荒牧 英治

はじめに

2007年9月の7日から9日にかけて Sweden は Skövde で TMI-07 が、その隣の国である Denmark で 10日から14日にかけて MT-Summit XI が開催された。二つの翻訳に関する会議が事実上ジョイントして開催されたわけであるが、TMI は、MT-Summit と比較して知名度が低いように思う。そこで本稿では、まずは、TMI の特色を MT-Summit と比較しながら紹介してから、今回の会議の様子をお伝えしたい。

TMI とは

【由来】TMI は、その名前だけからは MT の会議だと分かり難いが、“International Conference on Theoretical and Methodological Issues in Machine Translation”のアクロニムである。かくいう筆者も、今回が二回目の TMI 参加であるが、正式名称を初めて知った。会議名の一部になっている “Theoretical and Methodological Issues”への含意は、ケーススタディ的な内容ではなく、より学問的に翻訳の方法論を議論しましょう、との意味合いだと理解しているが、現在では MT-Summit も同様の傾向なので、もはや有名無実であろう。さて、略語に MT が入っていないのが、なんとも不思議である。

【規模】TMI は、MT-Summit と比較すると非常に小規模な会議である。発表は 29 件（口頭発表 17 件、ポスター12 件）であり、これは MT-Summit の 74 件（口頭発表 45 件、ポスター29 件）と比較すると半分以下の規模である。しかし、両会議とも開催日数は僅か一日しか違わない。

これは、TMI がシングルセッション（一つの会場でしか発表が行われない）であるためである。自分の発表を全ての人に聞いてもらえるというよい

面もあるが、会場に顔を見せないとサボっていることが丸わかりになってしまうという緊張感もある。

【レベル】さて、本質ではないと理解していても、どうしても採択率が気になってしまうのが、研究者の悲しい性であろう。これまでの MT-Summit では、採択率が公表されないことが多く、比較が困難であったが、今回は両会議とも submission 数が公開された。TMI の submission は 63 件で MT-Summit は約 110 件であるので、単純に考えると、口頭発表の採択率は TMI 27%、MT-Summit 46%、ポスターも含めた発表の採択率は TMI 41%、MT-Summit 67%となる。ここから考える限り、TMI はかなりレベルが高く、口頭発表だけに絞ってみれば、トップカンファレンスと比較しても遜色のない採択率である。勿論、採択率だけでなく、会議の特色、参加する研究者など様々な要素が考慮されるべきであるが、この採択率はある程度良質なペーパーを保障するものであろう。

【特色】前述したように TMI は小さい会議であり、一つしかない会場で三日間顔を合わせることになる。この環境ではすぐに全員の顔を覚えてしまうので自然と会話するようになる。むしろ、だんまりを決め込む方が難しいかもしれない。アットホームな雰囲気は TMI の大きな特色であり利点である。

会議の様子

以上のような TMI と MT-Summit であるが、今回の会議ではどのようなトピックが注目を集めたのだろうか。これについては、前回までの両会議と今回とは大きな変化があったように感じる。

ご存じの通り、最近の ACL や COLING などのトップカンファレンスでは SMT が圧倒的な勢いを誇っている。一方、これまでの MT-Summit では非

SMT 手法（ルールベースや用例ベース翻訳）も採択されバランスがとれていたように思う。言い換えれば、非 SMT 手法の研究の最後のオアシスがこれらの会議であった。

しかし、今回の両会議では、SMT の発表が非常に多かった。どこまでを SMT の研究とするのかその境界を定めるのは難しいが、発表タイトルに「SMT」または「Statistical Machine Translation」と明示的に記されているものだけでも両会議あわせて 26 件あった。一方、「Example-based Machine Translation」と銘打っているものは 3 件、「Rule-based」は 1 件と、その差は歴然である。さらに、「SMT」と明示的に述べていない場合でも、統計的アライメントなど SMT を想定した発表が多数を占めていた。このように SMT が広まったのは、基盤ツールである GIZA++ [1]や Moses [2]、SRILM [3]などが広く公開されているからであろう。これらのツールを組み合わせることにより、数日でベースラインとなる翻訳システムを構築でき、翻訳研究を開始することができる。数年かかってオモチャのような翻訳システムがやっと構築できた時代と比較すると夢のような状態である。

では、このまま SMT が主流となり、他の手法は消え去る運命なのだろうか？ その答えはイエス／ノーで答えられるような単純なものではないように思う。同じ質問を、Andy Way 氏(数年前に EBMT の本[4]を編集した人物)に投げかけたところ、「SMT と EBMT のどこが違う？ 違うところがあったら言ってみてくれ」と両手法の差異が、曖昧かつ無意味になっていることを指摘された。

実際に、今回、SMT の発表が多かったのは事実であるが、その中で目立ったのは統語的な情報の利用を行う研究であった。例えば、パーサーで両言語を解析し、その木構造をアライメントする知識を SMT に導入するといった類である。

このような考え方は EBMT の翻訳と矛盾するものでなく、もはや SMT の「S」の意味するもの



図1 招待講演の様子

は、人手を介するのではなくコーパスから学習しましょう、といったぐらいのニュアンスでしかない。この観点からは両手法で対立する点は皆無であるように思える。

今後、両手法がどのように交じり合い発展していくのか、次回の両会議が楽しみである。

おわりに

最後に、国際会議では英語という障壁のため、歯がゆい思いをすることが多い日本人研究者であるが、NECの奥村氏がユーモアに富んだ招待講演をされ、会場を沸かせていた(図1)。

会場にいる人を楽しませようとするサービス精神／ユーモアは万国共通のものであると実感し、多いに励まされた一瞬であった。

参考文献

- [1] <http://www.fjoch.com/GIZA++.html>
- [2] <http://www.statmt.org/moses/>
- [3] <http://www.speech.sri.com/projects/srilm>
- [4] Michael Carl and Andy Way, Recent Advances in Example-based Machine Translation, Kluwer Academic Publishers, 2003.

SDL 社の自動翻訳ソリューション — 自動化と手動プロセスの統合 —

エスディーエルジャパン株式会社 SDL TRADOS Technologies

マーケティング マネージャー 小澤 清美

SDL 社は、北米、アジア、ヨーロッパ、中東に 40 以上のオフィスを持ち、ローカライゼーションソリューションサービスを提供する企業として長年確実な実績を挙げてきました。近年、TRADOS 社（翻訳メモリツールの開発・販売）、Pass エンジニアリング社（ソフトウェア ローカライゼーションツールの開発・販売）、Tridion 社（コンテンツ マネジメント システムの開発・販売）を買収し、翻訳、ローカライゼーションのプラットフォームの統合という視野を含めたソリューション提供に事業展開をシフトしてきました。2007 年、ビジネスドメインをグローバル情報管理ソリューションの提供と再定義し、ミッションとして、企業のグローバルグローバルブランドの一貫性の確立や市場投入期間の短縮を支援することを掲げました。

SDL 社では、クライアント企業に end-to-end のコンテンツライフサイクルを考慮することを推奨しています。単純にコンテンツを作成後、グローバル市場のお客様のことを考え、コンテンツを翻訳し始めるのではなく、コンテンツ作成の時点から戦略的であるべきだと考えています。グローバル情報管理（GIM）は、そのための戦略です。この戦略をビジネスプロセスに取り入れることで実質的なメリットを得た企業が数多くあります。

■ 今日の機械翻訳

今日のコンピュータの性能は、機械翻訳が出始めた 60 年ほど前に比べ、劇的に進化しています。統計的もしくはルールベースなど、それぞれのアプローチには固有のメリットとデメリットがあります。

しかし、最終的に考慮されるべきことは、翻訳品質のレベルがその訳文コンテンツが使用される「目的」に見合っているかどうかなのです。

■ 「要約」ツールとしての機械翻訳

コミュニケーションが行われる状況が違くと要求される品質のレベルも違ってきます。友人同士のメッセージを通じた雑談などでは、相手が大体どのようなことを言っているかがわかれば十分でしょう。この場合、メッセージの機械翻訳機能（SDL Chat Translator や Yahoo IM Translator など）を利用できます。また、外国語で書かれたメールや Web サイトから素早く特定の情報を収集したい場合、Yahoo! 翻訳、エキサイト翻訳などが便利です。機械翻訳が原文コンテンツの「要約」を素早く示してくれます。

企業の中には、この「要約」という使い方をサポートや知識ベースの情報提供システムに利用しているケースがあります。コンテンツの要点が、理解できる言語で示されることにより、知っているか、知らないか、が鍵になる状況に対応できる場合があるのです。CNH 社（農業機器の製造・販売会社）は、目的によって機械翻訳のレベルを使い分けることで効果を上げています。たとえば、社外で仕事をしているフィールドエンジニアが「要約」により、問題に関する情報を素早く得ることで解決策を見つけることが容易になります。

ここ数年、販売後のサポートの分野に機械翻訳を導入することへの関心が高まっています。知識ベースや社内システムには膨大な技術データが蓄積されています。しかし、通常これらの情報は原文言語でしか保存されていません。多くのグローバル企業では、お客様のローカル言語でのサポートの提供が当たり前のことになってきています。多言語でのサービスに携わる人々の中から「何も情報がないよりは、何かがあるだけいい」と感じる人々が増え始めているのです。従来の翻訳プロセスでは、時間とコストがかかりすぎて実用に適さない状況で機械翻訳が導入されはじめています。

一方、お客様が購入検討中の製品やサービスについて確実な情報を得たい、という場合はどうでしょうか。この場合、SDL社の提供する「ナレッジベース翻訳システム（KbTS: Knowledge-based Translation System）」がひとつのソリューションになります。

■ 機械翻訳 + マニュアル翻訳の統合

SDL社は、翻訳支援ツールのテクノロジーや翻訳、ローカライゼーションサービスに幅広い経験と実績を持っています。SDL社は、これらの知識を結集し、「ナレッジベース翻訳システム(以降 KbTS)」というサービスを3年半ほど前にローンチしました。今日までに多くのグローバル企業にこのサービスをご導入いただいています。高品質、低価格な翻訳サービスを可能にするKbTSは、グローバル企業の市場投入期間の短縮を舞台裏で支えています。

KbTSは、自動用語抽出、用語ベース、翻訳メモリ、翻訳管理システムという自動化の技術を機械翻訳と統合したものです。大量の翻訳をこなすのに適したコンピュータと十分な知識とスキルのある翻訳者による翻訳（後編集）を組み合わせることによって、機械翻訳のみではなしえない、高品質の翻訳サービスを可能にしました。

その品質はこれまでの実績が物語っています。今までに同サービスを導入されたお客様は、現在まで継続してコンテンツのメンテナンスなど必要時にご利用いただいております。品質にもご満足いただいております。お客様には、クライスラー、HP、マイクロソフトなど、グローバル市場で先駆的な企業が多く、導入ケースは増加し続けています。

■ RS コンポーネンツ社

RS コンポーネンツ社は、世界の研究所や開発室、またはメンテナンス用などに幅広い品揃えで少量の電子部品を提供する企業です。幅広い品揃えのため、カタログやWeb上に掲載されている部品についてのコンテンツボリュームは膨大で、グローバル市場への製品投入に時間がかかっていました。

KbTS導入以前は、各ローカルオフィスがローカルベンダーに翻訳を発注していましたが、マニュアル翻訳では、コストも時間もかかる上、拡張性に乏しく、翻訳ボリュームが大きくなると数社に分けて作業発注する必要がありました。このためプロセスが複雑になり、効率や生産性の向上が重要な課題になっていました。

KbTSと製品コンテンツ管理のStibo Step Catalogueというシステムを統合することで複数の市場で同時に製品を販売することが可能になり、市場投入期間を35%短縮、コストを27%削減することができました。

■ CNH社

世界最大の農業機器メーカー、CNH社は、24カ国に支社を持ち、160カ国に販売代理店ネットワークを持っています。

同社が機械翻訳を導入し、目的別に使い分け始めた背景には、この世界的な販売ネットワークがあります。

各国の販売代理店のサービスエンジニアは、リアルタイムで自国語を話さない CNH 社のサポートセンタースタッフとコミュニケーションする必要がありました。フィールドエンジニアは、機械翻訳と用語ベースを用い、質問を翻訳してサポートセンターに送ります。サポートセンターのスタッフは、回答をローカルの言語に機械翻訳して送り返します。このソリューションでは、自社のサービスに密接に結びつけた専用の用語ベースを使うため、通常の機械翻訳より品質の高い機械翻訳が可能になります。しかし、やはり人手による後編集を加えた品質にはかきません。より高い品質レベルが必要になる状況では、機械翻訳とその他の自動化プロセスおよび人手による後編集を加えた方法を使います。たとえば、ある販売代理店がスペインで故障した車両の修理方法を知りたい、などという場合、オンラインシステム使って自国語で関連情報を検索、閲覧できます。

CNH社はオンラインでサポート情報を提供することにより、電話サポートセンターを運営するよりコストを削減することに成功しました。

■ KbTS とこれからの機械翻訳ソリューション

一般企業にとって、各言語に翻訳する翻訳者を雇い、用語ベースを作成・管理し、さらに関連するツールやテクノロジーの統合を必要とする、KbTS のようなシステムの構築は易しいものではありません。要求に見合うレベルの品質で翻訳し、コンテンツの更新後も必要な品質を維持し続けるには、膨大な時間と経験を要します。KbTS は、お客様側で翻訳者を雇ったり、プロセスを構築したりする手間を省くアウトソーシングソリューションです。

KbTS の効果を特に実感できるコンテンツは、技術関連文書、知識ベース内のサポートコンテンツ、Web サイト、ウイルスの警告メッセージ、製品カタログなどです。

近年、KbTs に関心を持つ企業、KbTs を通して機械翻訳の可能性を再発見する企業が増えています。SDL社は、今後とも機械翻訳、翻訳メモリ、用語ベース、マニュアル翻訳を統合し、膨大な知識遺産を生かそうとする試みがますます増えていくであろうと予測しています。

本稿でご紹介した製品やサービスについての詳細は弊社サイトにてご覧いただけます。

■ KbTS についての情報、ケーススタディ

<http://www.sdl.com/jp/>

トップメニュー「製品」>「自動翻訳」

■ 翻訳メモリ、用語ベースについての情報

個人翻訳者様向け

<http://www.translationzone.com/jp/>

言語サービスプロバイダ様向け

<http://www.lspzone.com>

日本のお客様および初心者向けの情報

<http://www.sdltradosjapan.com>

機械翻訳の発展とともに

株式会社クロスランゲージ

古賀 勝夫

1. 弊社のあゆみ

当社の機械翻訳ソフトに関する歩みを踏まえながら、今後の方向について、コメントしたいと思います。

当社の設立は約 20 年前の 1987 年に遡ります。同年 5 月に(株)ノヴァという社名で、会社を設立し、1998 年にアジア展開に向けて(株)ノヴァ・アジアを設立、2002 年に両社を合併して(ノヴァ・アジアを存続会社)、社名を(株)クロスランゲージと改めて、今日に至っております。翻訳ソフト勃興の時代からその只中で育ってきた翻訳ソフト開発専門の会社といえます。

「PC-Transer」

当社が製品を初めて世の中に出したのは、1989 年にワークステーション版英日機械翻訳ソフト「TRANSER/EJ」を出したのが最初でした。続いて、1991 年に NEC PC9801 シリーズ (MS-DOS) 上で動く翻訳ソフト「PC-Transer」を発売しまして、大きく飛躍するきっかけとなりました。当時の CPU は 80286 でしたが、1 文を翻訳するのに平均 5 秒くらいかかっていたものです。本ソフトは、その後合計 15 回のバージョンアップを重ねまして、現在の「PC-Transer 翻訳スタジオ 2008」に至っております。

「分野特化型翻訳ソフト」

その間、翻訳エンジンの改良や辞書の追加等を行う一方で、2つの新しい試みを行いました。ひとつは、「分野特化型翻訳ソフト」の開発です。これは、ターゲットを定めずに汎用的な文に対応しようとした場合に、どうしても翻訳精度が今ひとつよくなることを考慮して、ニーズの高い領域から、1998 年に特許文対応翻訳ソフト「PAT-Transer」を開発しました。これは、特許文固有の慣用表現を翻訳

メモリーに登録したり、大量の特許用語を第 1 訳語でうまく出るように工夫することをはじめとして、特許文特有の長文を通常の翻訳アルゴリズムで翻訳しようとしてもうまく翻訳できないことに鑑み、文のある規則で切って短くして翻訳したりするカスタマイズを行って、精度向上を狙ったものです。

現在 7 回のバージョンアップを実施し、「PAT-Transer 2007」となっております。

同様に第 2 弾として 2001 年に医学用翻訳ソフト「MED-Transer」を開発しました。現在は、バージョンアップを 5 回実施し、MED-Transer 2007 に至っております。さらに第 3 弾として、契約書用翻訳ソフト「Legal Transer」を 2007 年 10 月に発売しております。これは、法律、契約書用の翻訳メモリーや辞書を準備するとともに、契約書に固有の表現を機械翻訳しやすいように置き換えるという「自動前編集」を行って翻訳させることによって、高品質の契約書翻訳を実現しております。

「多言語翻訳」への取組

もうひとつは、日⇄英以外の「多言語翻訳」への取組です。はじめは、中国、韓国への展開を目標に、英中翻訳ソフトを 1994 年から開発を開始し、1999 年から中国本土で「金山快訳」という名称で OEM 提供しました。また、韓国でも 2000 年から英韓翻訳ソフトを、2003 年から韓英翻訳ソフトを発売しました。日本では、2002 年に日中・中日翻訳ソフト「蓬莱」を発売開始し、現在は、「蓬莱 VER 5.5 (2008)」になりました。さらに、日中・中日翻訳の精度を飛躍的に向上させるべく、「蓬莱」とは違う新エンジンの開発に取り組み、日中エンジンは昨年度に完成し、現在、「Yahoo!翻訳サービス」に組み込んでいます。また、中日翻訳エンジンも現在開発進行中で、

来年度半ばにリリースを予定しています。

その他の「多言語」翻訳ソフトとして、日韓・韓英翻訳ソフトを2001年から「高麗」として発売開始し、現在、「高麗 VER 3.5 (2008)」となっております。そして、ヨーロッパ語（ドイツ語、フランス語、イタリア語、スペイン語、ポルトガル語）への取組として、日欧翻訳ソフト（英語を媒介言語として日英⇄英⇄日によって実現）を2002年から「Jx9 多国語」として発売開始し、現在は、「JxEuro」と改名して「JxEuro V2 (2008)」に至っております。

翻訳サーバー「Web-Transer」の取組

3つめは、翻訳ソフトをインターネットを介して多くの人が Web ベースで利用できる環境を提供するサーバー翻訳ソフトへの取組です。インターネットから無料で翻訳サービスが提供される時代を迎えて、当社では2001年に多言語翻訳サーバーソフト「Web-Transer」の発売を開始しました。同ソフトはその後、企業のイントラネット下で導入利用されるとともに、2005年から「Yahoo! 翻訳サービス」に採用され、インターネットから誰でも利用できるサービスとして、現在まで提供されております。同様に、2006年には「楽天 Infoseek 多言語翻訳サービス」に採用され、多くの人に利用されております。また2007年9月から、Mediba より、au 携帯電話から利用できる翻訳サービスの提供が開始されました。

「Web-Transer」は、現在9カ国語間で利用できますが、今後、翻訳エンジンの改良、新エンジンの追加をはじめ、言語の追加、翻訳メモリーの利用、ユーザー辞書の利用などの改良を計画しております。

2. 最新の活動

翻訳ソフト提供から翻訳サービス提供へ

今後は、インターネットや携帯からの無料翻訳サービスの提供の拡大などで、ますます多くのユーザーによる翻訳ソフトの利用が見込まれておりますが、我々翻訳ソフトのベンダー側としては、このような「無料翻訳サービス」時代を迎えて、従来のように

「翻訳ソフトを売る」というビジネスモデルからの転換を求められているとも言えます。そうした背景から、弊社では新しいいくつかの動きを始めております。

中国特許検索・翻訳サービス「WORLDPATENT.jp」

上記の時代の流れから我々が達したひとつの答えは、翻訳ソフト提供から翻訳サービス提供へ転換しなければならないということでした。

このようなとき、日本企業の中国進出が急激に拡大し、従来以上に中国マーケットが注目される中で、中国での特許侵害、特許防衛の課題が日に日に増してきている状況を鑑み、弊社のこれまでの特許翻訳へ取り組んで来た強み、および、人手による翻訳事業に取り組んできた強みを生かして、日本企業へ中国特許のオンライン提供サービスと翻訳サービスを結合させた中国特許検索・翻訳サービスを2006年初に企画しました。まず中国專利局の元にある知識産権局知識産権出版社(CNIPR)と中国特許提供に関する契約を行い、中国における特許更新と同じサイクルでの更新データの提供を受けて、サービスを提供できる体制を整え、「WORLDPATENT.jp」サイトで、2006年11月から試験サービスを開始し、2007年9月から有料でのサービスを開始しました。

これは、機能的には、IPC 番号や出願者、出願年月日などで中国特許を検索できることはもとより、日本語でキーワードを指定することにより、それを中国語に変換して検索できる言語横断検索機能を有し、かつ、検索した結果を中国語とともに機械翻訳した日本語を左右に併置して見せる機能を持っております。中日翻訳ソフトは、弊社の「蓬莱」を特許用にカスタマイズし、高頻度の特許用語を約10万語整備して、利用しております。

利用法としては、本サービスでまず、所望の中国特許をIPC 分類などで粗く検索し、機械翻訳結果を参照しつつ、ある程度の件数に絞ってそれを「抄録翻訳」(人手による翻訳サービス)にかけて、所望する

特許に絞って、全文翻訳にかける（人手による翻訳サービス）という流れを想定しております。このサービスのために、大連事務所を3年前に設立して中国特許専門の翻訳者を養成するとともに、上海、北京の翻訳会社とも提携をして、人手による中日特許翻訳サービスの体制を強化しております。

ここでは、翻訳ソフトは当社自身が利用して、ユーザへは中国特許の提供サービスというサービスモデルへの転換を行っております。

3. 今後の展開

翻訳ソリューションの提供

翻訳ソフトの最大のマーケットは人手による翻訳市場であろう、しかし、翻訳市場ではなかなか翻訳ソフトは受け入れられない、という現実から、自ら人手翻訳事業に取り組んでみようという試みを、2001年度から始め、現在まで続けてきております。その意識の一つは、弊社の翻訳ソフトや辞書を人手翻訳にできる限り活かすにはどうすればよいか、ということです。しかし、ユーザにとって、本当に大事なことは、「機械翻訳ソフト」を導入するか、あるいは、「人手翻訳」するかではなく、かかえている翻訳ニーズをどう解決するかであるということに痛感するようになりました。ユーザの抱えている問題を分析して、最適のソリューションを出すことこそ重要だということで、今年度から弊社の法人営業部の重要テーマとして、「翻訳ソリューションの提供」に取り組んでおります。

新しい「翻訳モデル」の創造

「翻訳」は、ユーザの目的に応じて、求める物が全然違っております。たとえば、「できるだけ早く安く内容を自分で理解することが目的」なら、その外国語をある程度理解できることは前提にするものの、現在の機械翻訳ソフトはかなり役に立っていると言えます。その対極にあるものは、「翻訳結果を他人に見せる必要がある場合」は、「正しい翻訳」が要求され、この世界では翻訳ソフトは一部ツールと

して利用されるものの、ほとんどは人手による翻訳で実行されています。しかし、我々が抱えている問題には、この対極の中間に位置するものが多数あります。たとえば、先の「中国特許検索・サービス」での場面を考えてみましょう。ユーザは、まず探したいものを逃さないように大きな網で検索をかけます。当然、多数の該当する特許が存在します（たとえば1000件とか）。その中から、本当に自分たちの注目する特許を探さなければなりません（たとえば10件とか）。ここで最初に見つけた1000件を全部翻訳させれば、大変な費用がかかります。そこで、最初の1000件に対し、検索結果を、機械翻訳結果もあわせて表示するようにしておけば、機械翻訳の精度に問題はあるものの、ユーザの求める特許に「まったく関係がなさそうなものを捨てる」という役には立ちます。この時点で、最初の1000件が1/5の200件になるなら、それはコスト的に大いに助かります。次は残る200件に対し、ある程度、精査していかなければなりません。それをやるために、「全文翻訳」の必要はなく、「抄録翻訳」の範囲で選別は可能でしょう。そして、この「抄録翻訳」の目的は、「選別」することにあって、「99%正しい翻訳ではない」ならば、何も専門分野の専門知識を持った翻訳者が要求されるわけでもありません。この範囲であれば、専門用語を充実させて機械翻訳を利用し、その結果を参考にしながら人間が翻訳をすることによって、時間的にもコスト的にも安く上げる工夫も成り立ちます。ユーザはその「抄録翻訳」の結果を見て、「自分たちの注目する特許」を見つけます（たとえば、20件とか）。そこで初めてこの20件を全文翻訳すればよいでしょう。

そのときは、かなり正確な翻訳が要求されるので、専門的な知識も要求されるでしょう。これは、従来の人手翻訳のコストがかかることはやむを得ません。

このプロセスを見ても分かるように、最終結果を求めるまでに要求される「翻訳の質」はかなり異な

っております。この途中段階には、「翻訳ソフト」や「辞書」が役に立つ場面は、多数残されていると思われまます。当然、それなりの「カスタマイズ」は要求されるでしょう。ここに、今後の「翻訳ソリューション」の新しいモデルの考え方が含まれていると我々は考えております。このためには、我々メーカー側は、もっともっとユーザーに近づいていかなければならないと思います。

「翻訳ツール」の「利用技術」について

初めての商用翻訳ソフトが世の中に出て、20年以上がたちました。弊社も初めて翻訳ソフトを世に出して、18年たちました。その間に、十数回バージョンアップを繰り返して、改良を重ねてきております。特に翻訳エンジンに関しては、毎回何らかの手を加えて、改良を重ねております。しかし、正直、ユーザーの視点に立った場合、翻訳エンジンの改良が果たして効果的と思われるかどうか、疑問になることがあります。もちろん、これは、もはや翻訳エンジンの改良は意味をなさない、と言っているではありません。そうではなく、エンジンの改良と並行して、もっと別の効果的なアプローチを必要としているのではないかということです。分野をしぼる「分野特化型」というアプローチはひとつの有力な方法と考えております。「分野」は、狭ければ狭いほど有効でしょう。「一企業」に特化するのも有効でしょう。辞書の問題にしても、現在のレベルで十分かといえば、まったく不十分だとも言えます。特定専門分野の適訳が出ないというのも頻繁に見受けられます。辞書の訳語をどのように充実させていくかというのは、今後も大きい課題です。メーカー側からすべて提供できるわけもないので、それではどのようにユーザー側からその知識を受けて吸収していく仕組みを作るかは重要なテーマです。また、インターネットからの検索が広く全世界から求められるようになってきている現代で、「検索」とからめて「翻訳」を見た場合、言語横断検索での辞書の有様や、概要把握にとって全文の機械翻訳結果より有効

な翻訳表示の仕方など、今後検討しなければならない課題もたくさん残されています。

こうした気づいている課題も数多くありますが、まだ我々が気づいていない課題もたくさんあると思われまます。それは、ユーザーの抱えている問題に近づいていくことで、もっと明らかになっていくのではないかと我々は考えております。そこからまた新しい「翻訳ツール」の利用の仕方が発見されてくるのではないかと期待して、我々はユーザーに近づいていこうと決意しております。

計算機支援による空の安全管理システム

株式会社日本航空

寺田 昭

1. はじめに

交通機関の安全、食の安全、医療の安全等、社会全体で安全で安心なサービスが求められている。安全とは、サービスを提供する側が守らなければならない最低限の安全性を保っていることであり、安心とは、サービスを提供される側、例えば、お客様や消費者が安心できる状態である。したがって、安全だけを提供するのは、サービスを提供する側の論理であり、安全で安心なサービスを提供しないような企業・組織は社会から批判を浴び、その姿勢を改めなければ、存続さえ認められなくなってきた。しかしながら、現実社会では、原因究明よりも犯人探しを行うような懲罰主義、安全よりも利益を追求するような利益優先主義、問題が起こってもそれを隠蔽するような隠蔽主義、問題の原因を根本から追求しないパッチ充ての対策にみられるように、安全・安心な土壌は、未だに未成熟である。

これまで安全であると思われてきたシステムでも、潜在する不安全要素が現れ、危険な状態になることがある。

トラブルの要因を調べてみると、トラブルの中には、前もって何らかの兆候があり、警告がなされていたものや、過去に同種のトラブルが起こっていたものが多く存在する。したがって、そのような情報を利用することにより、トラブルが起こる前に未然に防止できることが可能になる。

本稿では、航空分野に於いて、計算機支援により、トラブル情報を大量に蓄積したテキストから、不安全要素を発見し、安全対策を行う手法について述べる。

ボーイング社の統計によると、1970年代半ばまで減少していた航空機の事故率は、それ以降あまり低下していない。航空機の出現後、初期の頃は、ハードウェア

の発達・改良により事故率が大幅に減少したが、ソフトウェアに起因する事故率は、あまり減少していない。その原因の約6割は、人的要因にあると云われている。

人的要因をSHELモデルで説明したのが図1であるが、人(Liveware)を中心に、ハードウェア(Hardware)、環境(Environment)、ソフトウェア(Software)、人(Liveware)が関係しており、その関係からエラーが発生することを表示している。エラーの要因として、人

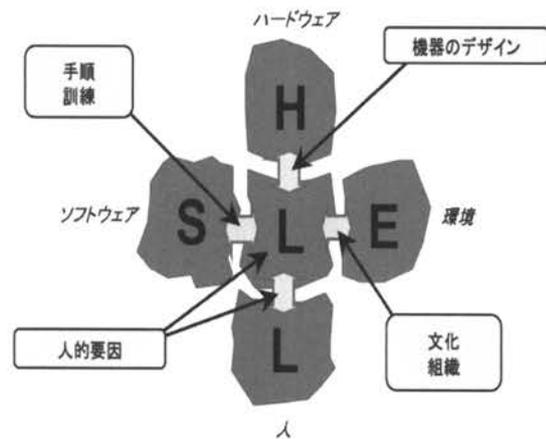


図1. SHELモデル

とハードウェアでは機器のデザイン、人と環境では文化・組織、人とソフトウェアでは手順・訓練、人単独または人と人では人的要因があることを表している。そして、人とそれぞれとのインターフェースがうまくいかない場合にエラーが発生する。

トラブルの影響度を考えると、事故、インシデント(軽微な事故)、ヒヤリハット(異常)に分けられる。ハインリッヒの法則によると、1つの事故の背後にその29倍のインシデントがあり、更にその300倍のヒヤリハットが存在すると云われている。

事故およびインシデントについては、その原因が国の事故調査委員会等で詳細に徹底的に調査される。それに基づいた対策は、同種の事故を起こさないためのものであり、Re-active(対症療法的)な安全対策と呼ばれている。この対策は、同種の事故の防止には役立つが、それ以外のものについては、あまり役立たない。Re-active な安全対策の問題点は、事例の数が少ないため、将来起こりうるトラブルを網羅できず、また、対策立案に時間が掛かる点である。

一方、Pro-active(予防的)な安全対策は、事故に至らなかったヒヤリハットの事例から、その要因を分析して、事故の芽を事前に摘み取る手法である。その背景としては、事故、インシデント、ヒヤリハットには、通常、複数の要因が関係しており、個々の要因には、共通のものが存在しているからである。

図2は、エラーを冰山表示したもので、水面に出ているのが事故のみで、それ以外の膨大な数のインシデント、ヒヤリハットが水面下に隠れている様子を示している。図2から、事故の要因A, B, C, Dの中で要因Aは、インシデントにも、ヒヤリハットにも出現していることが分かる。したがって、ヒヤリハットからトラブルの要因を抽出して、それを潰していくことにより、事故の要因を未然に潰していくことが可能になる。

(株)日本航空は、(株)富士通研究所、東京大学中川研究室と共同で計算機によるリスク・マネジメントの研究を行っているが、その成果および安全管理システムの中での利用法について報告する。なお、テキストマイニングのソフトウェアは、(株)富士通研究所で開発したものを使用している。

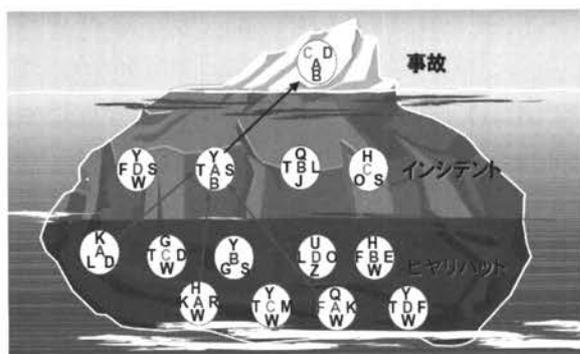


図2. エラーの冰山表示

2. 安全管理システム

日本航空社内で行っている Pro-active な安全管理システム(Safety Management System)について説明する。安全管理システムは、正確なデータに基づく情報からリスクを予想し、事前に対策をたてるものである。図3に示すように、データ収集、データ解析、リスク評価、予防対策の策定、予防対策の実行という PDCA サイクルがその中心をなしている。現在、これらの各ステップの運用は全て人手(専門の解析者)で行われているが、計算機支援という観点から、それぞれのステップを見ていくこととする。

2.1 データ収集

データは、主にトラブルを経験した人が提出するレポートを使用している。レポートには、天気等の定型項目とテキストでの自由記述の非定型項目から構成される。テキスト情報は、タイトルおよび本文からなっており、タイトルには、事例の概要、本文には、その

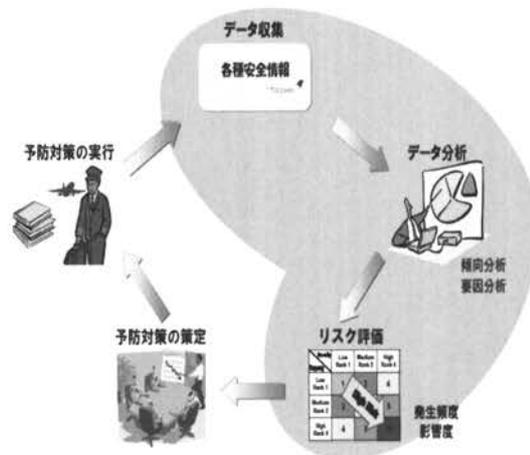


図3. 安全管理システム

詳細な状況、対策等が記述されている。

レポートの収集面からは、レポートを提出した人に不利益が及ばないような制度の構築が必要である。つまり、提出者のプライバシーの保護や提出することにより不利益の及ばない非懲罰制度の導入である。

提出者の負担を少なくするために、レポートの記入・提出を IT 化する入力支援システムやレポートを安

全に送信できるようなセキュリティー・システムも考慮する必要がある。

さらに、ヒヤリハットにもならないような通常の状態からの少しの逸脱を飛行中のデジタル・データの中から拾い、実際に運航に携わっていた運航乗務員にその時の状況をインタビューした内容をテキストにまとめたものも利用している。このデータを社内でFDM(Flight Data Monitoring)¹と呼んでいるが、このFDM データの占める領域は、ヒヤリハットにも至らない潜在的な危険要素が含まれているという意味で、図2のヒヤリハットの更に下の大きな領域である。このような通常状態からの少しの逸脱からトラブルの潜在要因を特定する安全対策は、Predictive(予見的)な安全対策と呼ばれている。

2.2 データ分析

データ分析は、現在、全て、人間の経験と記憶に頼って行われているが、データの分析部分は計算機の得意な部分であり、この部分を計算機支援により行い、大量のデータからの傾向分析、要因分析を行うことが可能となってきた。主に本文部分のテキストを対象としているが、自由記述型のテキストには、これまで分析者が気付かなかったような「気付き」情報やトラブルの要因が含まれており、定型情報では得られないような貴重な情報が含まれている。収集されたデータ量が膨大になってくると、人間が分析するには限界があり、計算機の利用価値の大きな部分である。したがって、そのような情報をテキストマイニング技術により大量のテキストから獲得する。そして、計算機の分析結果を人間が判断し利用する形になる。

データ分析は、どのような事例が発生しているのかの傾向を見る傾向分析、事例の発生原因は何かを探る要因分析の2つに大きく分けられる。

まず、傾向分析により、機種別・空港別の特徴、時系列的な変化傾向など不具合事例発生の全体的な傾向を掴むことにより、どのような事象に注目すべきかを

決定する。次に、要因分析では、注目すべき事象につき、人的エラーとそれを誘発した要因に注目する。トラブル事例では、原因-現象という因果関係に注目することが重要である。しかしながら、安全対策では、事例の原因-現象を発見するだけでは不十分で、その事例について、何故発生したかを突き詰めて、真の要因は何かを発見し、その防止策をたてることが重要である。

例えば、「T/O(Takeoff) FLAP 忘れ」が原因で「RTO(Rejected Takeoff: 離陸中止)」という現象が発生した場合を考えてみる。「RTO」の原因として「T/O FLAP 忘れ」を発見するだけでは不十分である。つまり、「RTO」を防止するための対策として、その原因である「T/O FLAP 忘れ」をしないようにと教育しても、精神的に「失敗するな」と云うだけで、根本要因を掴んでいないので、効果は少ない。なぜ、「T/O FLAP 忘れ」をしたのかの要因を探ることが必要で、その内容を見てみると、「通常操作を行わなかった」、「チェックリストで見逃した」というようなエラーが見えてくる。さらに、そのエラーがなぜ起こったかの要因を考察すると、「T/O FLAP をセットする時点でATC(航空管制機関)との交信があり、複数のタスクが重複した」というような要因が見えてくる。

そして、同種エラーについての要因を複数のレポートから集めることにより、少数のレポートでは記載されていないような要因も補完され、個々のエラーにどのような要因が多く関係しているかが分かり、エラーの防止策を策定することができるようになる。

以上のような分析は、現在の技術水準では、エラーやその要因の抽出を完全に自動化できない。人間が行う分析を計算機が支援する形になる。したがって、計算機の分析結果をグラフ、図解化して人間が理解しやすいように表示する可視化技術および様々な角度から分析できるような機能を備えていることが重要である。

2.3 リスク評価

企業の経済的、物理的な理由から、全てのリスクを一度に排除することはできない。したがって、どのよ

¹ 一般にFOQA(Flight Operational Quality Assurance)と呼ばれている。

うなトラブルから対策を立案するかの優先順位を付けなければならない。優先順位は、トラブルの発生頻度と影響度を勘案して付ける。もちろん、影響度が非常に大きなものでは、対策案を作成する前に運航停止にするようなものもある。

2.4 予防対策の策定

第2.3節のステップで選択された項目について、予防対策を策定する。

2.5 予防対策の実施

予防対策を実施するが、その方法は、単なる情報の提供のみのも、定期訓練・教育に取り入れ、Procedureを定着させるものまでである。定期訓練に取り入れた場合には、全員が受講するまでに1年間掛かる欠点があるので、対策の中で重要なものは、最初に情報を提供し、その後、訓練に取り入れるようなハイブリッドな手法が多く用いられている。

3. 要素技術

第2.2節で述べたデータ分析を計算機で行うための要素技術について述べる。

3.1 テキストマイニング技術

汎用のテキストマイニング技術を使用するが、因果関係を抽出するためのルールの付加やレポートの内容を代表するようなキーワードを抽出するために分野に特化したチューンアップが必要である。

本稿で使用する手法を以下に紹介する[2]。最初に、文書全体をクラスタリングする。クラスタの中には、複数の話題が含まれる場合があり、それぞれの話題に関するキーワードが混在し解釈が困難な場合や、複数のクラスタに含まれる話題が抽出漏れになる場合がある。したがって、クラスタの中身を表すキーワードを話題毎に分割して、分析者に提示する必要がある。そのようなキーワードを特徴キーワードと呼ぶことにする。特徴キーワードの値は、クラスタと正の相関関係にあるキーワードの χ^2 値を χ^2 検定により求め、クラスタ内の出現頻度との積により求める。そして、特徴

キーワードをランキングして、値の大きなもののみを使用する。特徴キーワードの分離は、特徴キーワード間の関連度グラフを共起関係により作成することにより行う(図4参照)。

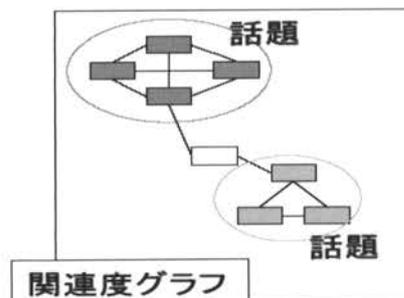


図4. 関連度グラフ

3.2 本文参照機能

注目する場所(エラー、要因など)の詳細について、分析者が本文を簡単に参照できるような機能や、キーとなる語句の前後によく出現する表現を参照できるような機能も必要である。後者については、PortableKiwi[6]を使用している。PortableKiwiは、対象としているコーパスに対して、ある言語表現を入力すると、その前後に現れる適当な長さの文字列[1]のうち、頻度の高いものから順に表示する用例検索システムである。

3.3 同義語の同定

航空分野では、名詞の同義語が多数使用されている。例えば、「トラブル」を「Trouble」や「TRBL」などと表現している。したがって、これらの同義語を同定するための辞書の構築が必要である。我々は、現在、計算機による同義語辞書作成支援システムを開発している[4][5]。

更に、名詞だけではなく、幅広い表現から同義なもの同定する必要がある。例えば、「バードストライク」と「鳥が航空機に衝突した」などである。

3.4 因果関係の抽出

第2節で述べたように、トラブルの分析には、因果

関係およびその要因の抽出が重要である。因果関係は、「による」などの因果関係を表す表層表現の前後を抽出したり、文の前後関係や時間関係などから抽出する方法がある。現段階の技術では、因果関係と要因を直接結びつけることは難しい。

4. 計算機によるデータ分析実験

本節では、航空安全レポート 6,432 件を分析対象とした実験結果について述べる[3]。

4.1 因果関係の抽出

因果関係の表現は、本文部分にはキーとなるような明示的な表現は少なく、その詳細度も異なっている。一方、タイトル部分には、因果関係が明示的に表現されている場合が多く、その詳細度もほぼ揃っている。したがって、タイトルから因果関係を、本文からその要因となるような特徴キーワードを抽出することにした。この処理では、因果関係とその要因は、直接、結びつかないが、因果関係の中で注目すべきものについて、特徴キーワードマップを描くことにより、その要因が読み取れるようにしている。

因果関係の抽出には、表層的な表現の「による」、「ため」、「DUE TO」の3つの表現を基本として、そのパリエーションを考慮してルールを作成した。その結果、全体の約半数の3,562のレポートから因果関係を抽出できた。更に、同義語リストを使用して、現象を統合することにより、現象の異なり数は、791になった(例えば、「DELAY」を「遅延」に言い換えるなど)。

4.2 因果関係の分析

抽出した因果関係のネットワーク表現を図5に示す。分かりやすいように、原因を☆印、現象を★印で表している。

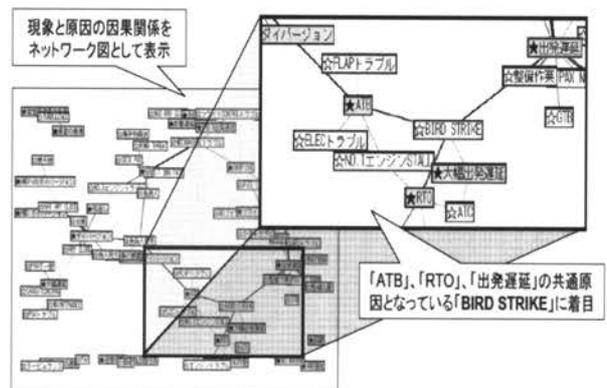


図5. 因果関係ネットワーク

図5から、レポートに出現した因果関係全体を俯瞰でき、その中から、「ATB(空中引き返し)」、「RTO」、「出発遅延」というような運航に大きな影響を与えている共通原因が「BIRD STRIKE」であると特定できる。

図6は、抽出した原因を空港別に集計した割合グラフである。この図から、羽田、福岡に「BIRD STRIKE」が多いことが読み取れる。

図5、図6は、それぞれ傾向分析を行っているが、図5では原因-現象に注目し、図6では原因に注目した分析になっている。このように、分析者がさまざまな切り口から分析をできることが重要である。

図7は、「BIRD STRIKE」が発生した98事例について特徴キーワードマップを描いたものである。図7から、

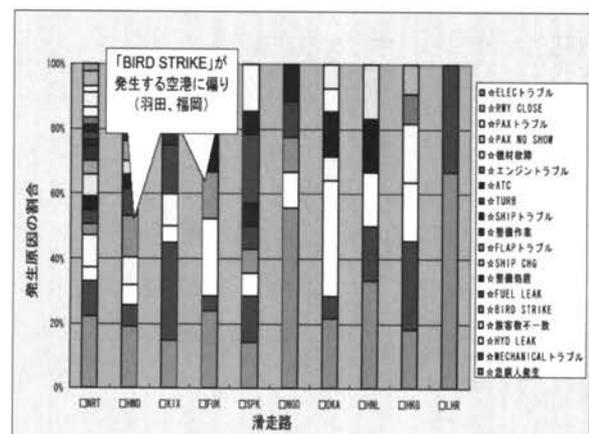


図6. 空港別の「原因」発生割合

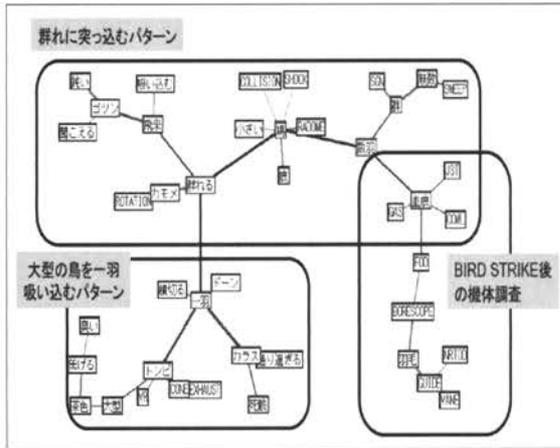


図7. 「BIRD STRIKE」発生事例の要因分析

「BIRD STRIKE」は、一羽の大型の鳥(カラスやトンビ)によるもの、小型の鳥(鳩など)の群れによるものという2つの発生パターンがあることが読み取れる。また、もう1つのクラスタは、「BIRD STRIKE」発生後の機体検査に関するものである。

5. まとめ

本稿では、航空分野の安全レポートを使用して事故が起こる前に未然にその要因を摘み取る Pro-active な安全対策を計算機支援により行う手法について述べた。

今後、解決しなければならない課題は多いが、分析実験結果に見られるように、現段階でも興味深い結果が得られている。

現在、経済産業省の情報大航海プロジェクトにより、より大量のデータを使用した実証実験、技術開発を行っており、良い成果が得られ、日本航空社内だけでなく、広く社会の安全・安心のモデル・サービスを提供できることを期待している。尚、本稿の内容は、情報大航海プロジェクト以前の成果である。

今後の課題として、以下が挙げられる。

- レポートからトラブルの要因を抽出する手法を紹介したが、トラブルに遭遇してもレポートを提出するかどうかは、当事者に委ねられている。したがって、安全意識を向上させて、レポート提出の意義を理解させるような努力が企業・組織に必

要である。第2.1節で述べたように、提出者の非懲罰、プライバシーの保護の確立が必要である。さらに、提出者によりレポートの記載内容の詳細度が異なるので、必要なアイテム、例えば、昼夜の区別が重要であれば、それを定型項目に含める等の改善が必要である。また、FDMデータに見られるように飛行データに閾値を設け、その値から逸脱した飛行についてのデータを収集することにより、ヒヤリハットよりも裾野の広いデータを収集でき、閾値を超えたデータを全て収集しているので、統計的解析に向けたデータが収集できる。

- 因果関係とそれに関係したエラー、要因を自動的に関連付けて抽出できるような技術の開発が必要である。
- 本稿で述べた手法は、航空分野だけではなく、幅広く他の分野に適用できるものと思われる。分野に関係なく共通の部分、分野別にチューンアップすべき部分を区別し、移植性のよいものを構築し、他の分野でも実験してみる必要がある。
- 日本語のテキストを対照としたが、他の言語でも同様な分析を行う必要がある。例えば、米国では、ASRS(Aviation Safety Reporting System)というレポートシステムがあり、大量のデータが蓄積されている。ヒューマンエラーは言語に関係なく共通のものと考えられるので、他言語での分析を行うことで、より多くのヒヤリハット情報を利用・共有できるようになる。
- テキスト情報を対照としたが、テキスト以外のマルチメディア情報の利用についても、今後、考慮していく必要がある。

参考文献

- [1] Tanaka-Ishii, K. and Nakagawa, H. (2005). A Multilingual Usage Consultation Tool based on Internet Searching ---More than search engine, Less than QA. In the 14th International World Wide Web Conference (WWW2005), pp. 363-371.
- [2] 齋藤孝広 渡部勇 松井くにお 寺田昭 齋藤隆

(2005). 航空安全情報からのトラブル発生パターンの抽出について. 言語処理学会第 11 回年次大会, pp. 69-72.

[3] 齊藤孝弘 葉師寺あかね 渡部勇 松井くにお 佐々木敏宏 寺田昭 斎藤隆 (2006). 航空安全情報分析ツール-因果関係に着目したレポート分析手法の提案-. 第 44 回 飛行機シンポジウム, pp. 493-500.

[4] 寺田昭, 吉田稔, 中川裕志 (2006). 文脈情報による同義語辞書作成支援ツール. 情報処理学会研究報告, pp. 87-94.

[5] 寺田昭 吉田稔 中川裕志 (2007). 計算機支援による同義語辞書作成ツール. 第 45 回 飛行機シンポジウム, pp. 1171-1177.

[6] 藤本宏涼, 吉田稔, 中川裕志 (2005). ローカルコーパスからのテキストマイニングツール: PortableKiwi, 言語処理学会第 11 回年次大会, pp. 97-100.

KUI: A Self-organizing Tool for Multi-lingual WordNet Construction

Virach Sornlertlamvanich, Thatsanee Charoenporn, Kergrit Robkop, Hitoshi Isahara*
 Thai Computational Linguistics Lab, NICT Asia Research Center, NICT*

Abstract

This paper describes a tool for multi-lingual WordNet construction, called KUI (Knowledge Unifying Initiator), which is a knowledge user interface for online collaborative knowledge construction. KUI is a sort of social networking system that unifies the various discussions following the process of thinking model, i.e. initiating the topic of interest, collecting the opinions to the selected topics, localizing the opinions through the translation or customization and finally posting for public hearing to conceptualize the knowledge. The process of thinking is done under the selectional preference simulated by voting mechanism in the case that there are many alternatives. By monitoring the history of participation of each member, KUI adaptively manages the reliability of each member's opinion and vote according to the estimated *ExpertScore*. As a result, the multi-lingual WordNet can be created online and produce a reliable result.

1 Introduction

The constructions of the WordNet (Fellbaum, 1998) for languages can be varied according to the availability of the language resources. Some were developed from scratch, and some were developed from the combination of various existing lexical resources. Spanish and Catalan Wordnets¹, for instance, are automatically constructed using hyponym relation, monolingual dictionary, bilingual dictionary and taxonomy (Atserias et al., 1997). Italian WordNet (Magnini et al., 1994) is semi-automatically constructed from definition in monolingual dictionary, bilingual dictionary, and WordNet glosses. Hungarian WordNet uses bilingual dictionary, mono-lingual explanatory dictionary, and Hungarian thesaurus in the construction (Proszeky et al., 2002), etc.

A tool to facilitate the construction is one of the important issues related to the WordNet construction. Some of the previous efforts were spent for developing the tools such as Polaris (Louw, 1998), the editing and browsing for EuroWordNet, and VisDic (Horak et al., 2004), the XML based Multilingual WordNet

browsing and editing tool developed by Czech WordNet team. To facilitate an online collaborative development and annotate a reliability score to the proposed word entries, we, therefore, proposed KUI (Knowledge Unifying Initiator) to be a Knowledge User Interface (KUI) for online collaborative construction of multi-lingual WordNet. KUI facilitates online community in developing and discussing multi-lingual WordNet. KUI is a sort of social networking system that unifies the various discussions following the process of thinking model, i.e. initiating the topic of interest, collecting the opinions to the selected topics, localizing the opinions through the translation or customization and finally posting for public hearing to conceptualize the knowledge. The process of thinking is done under the selectional preference simulated by voting mechanism in the case that there are many alternatives.

This paper illustrates an online tool to facilitate the multi-lingual WordNet construction by using the existing resources having only English equivalents and the lexical synonyms. Since the system is opened for online contribution, we need a mechanism to inform the reliability of the result. We introduce *ExpertScore* which can be estimated from the history of the participation of each member. The weight of each vote and opinion will be determined by the *ExpertScore*. The result will then be ranked according to this score to show the reliability of the opinion.

The rest of this paper is organized as follows: Section 2 describes the process of managing the knowledge. Section 3 explains the design of KUI for Collaborative resource development. Section 4 provides some examples of KUI for WordNet construction. And Section 5 concludes our work.

2 Process of Knowledge Development

A thought is dynamically formed up by a trigger which can be an interest from inside or a proposed topic from outside. However, knowledge can be formed up from the thought only when managed in an appropriate way. Since we are considering the knowledge of a community, we can consider the knowledge that is formed by a community in the following manner.

1. Knowledge is managed by the knowledge users
2. Knowledge is dynamically changed
3. Knowledge is developed in an individual manner or a community manner

¹ <http://www.lsi.upc.edu/~nlp/>

4. Knowledge is both explicit and tacit

The environment of online community can successfully serve the requirement of knowledge management. Under the environment, the knowledge should be grouped up and narrowed down into a specific domain for each group. The domain specific group can then be managed to generate a concrete knowledge after receiving the consensus from the participants at any moment.

Open Source software development is a model for open collaboration in the domain of software development. The openness of the development process has successfully established a largest software community that shares their development and using experience. The activities are dedicated to the domain of software knowledge development. SourceForge² is a platform for project based Open Source software development. Open Source software developers deploy SourceForge.net to announce their initiation, to call for participation, to distribute their works and to receive feedbacks concerning their proposed software. Developers and users are actively using SourceForge.net to communicate with each other.

Adopting the concept of Open Source software development, we will possibly be able to develop a framework for domain specific knowledge development under the open community environment. Sharing and collaboration are the considerable features of the framework. The knowledge will be finally shared among the communities by receiving the consensus from the participants in each step. To facilitate the knowledge development, we deliberate the process into 4 steps.

1. Topic of interest

The topic will be posted to draw the intention from the participants. The selected topics will then be further discussed in the appropriate step.

2. Opinion

The selected topic is posted to call for opinions from the participants in this step. Opinion poll is conducted to get the population of each opinion. The result of the opinion poll provides the variety of opinions that reflects the current thought of the communities together with the consensus to the opinions.

3. Localization

Translation is the straightforward implementation of the localization. Collaborative translation helps producing the knowledge in multiple languages in the most efficient way.

4. Public-Hearing

The result of discussion will be revised and confirmed by gathering the opinions to the final draft of proposal.

Figure 1 shows the process of how knowledge is developed within a community. Starting from posting 'Topic of Interest', participants express their supports

by casting a vote. Upon a threshold the 'Topic of Interest' is selected for conducting a poll on 'Opinion', or introducing to the community by 'Localization', or posting a draft for 'Public-Hearing' to gather feedbacks from the community.

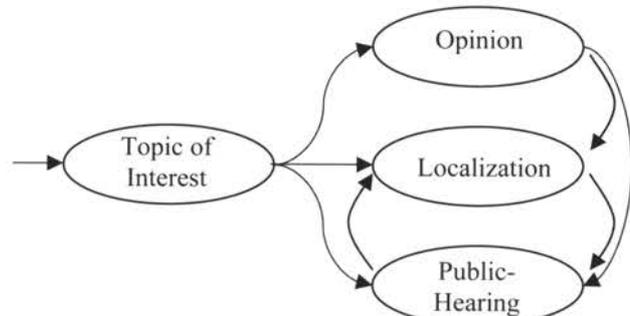


Figure 1 Process of knowledge development

The transition from 'Opinion' to either 'Localization' or 'Public-Hearing' occurs when the 'Opinion' has a concrete view for implementation. The discussion in 'Localization' and 'Public-Hearing' is however interchangeable due to purpose of implementation whether to adopt the knowledge to the local community or to get feedbacks from the community.

The knowledge creating is managed in 4 different categories corresponding to the stage of knowledge. Each individual in the community casts a vote to rank the appropriateness of solutions at each category. The community can then form the community knowledge under the 'Selectional Preference' background. On the other hand, the under-threshold solutions become obsolete by nature of the 'Selectional Preference'.

3 Knowledge User Interface for Knowledge Unifying Initiative

3.1 What is KUI?

KUI is a GUI for knowledge engineering, in other words Knowledge User Interface (KUI). It provides a web interface accessible for pre-registered members. An online registration is offered to manage an account by profiling the login participant in making contribution. A contributor can comfortably move around in the virtual space from desk to desk to participate in a particular task. A working desk can be a meeting place for collaborative work that needs discussion through the 'Chat', or allow a contributor to work individually by using the message slot to record each own comment. The working space can be expanded by closing the unnecessary frames so the contributor can concentrate on the task. All working topics can be statistically viewed through the provided tabs. These tabs help contributors to understand KUI in the aspects of the current status of contribution and the tasks. A knowledge community can be formed and can efficiently create the domain knowledge through the fea-

² <http://www.sourceforge.net/>

tures provided by KUI. These KUI features fulfill the process of human thought to record the knowledge.

KUI also provides a 'KUI look up' function for viewing the composed knowledge. It is equipped with a powerful search and statistical browse in many aspects. Moreover, the 'Chatlog' is provided to learn about the intention of the knowledge composers. We frequently want to know about the background of the solution for better understanding or to remind us about the decision, but we cannot find one. To avoid the repetition of a mistake, we systematically provide the 'Chatlog' to keep the trace of discussion or the comments to show the intention of knowledge composers.

3.2 Feature of KUI

- **Poll-based Opinion or Public-Hearing**

A contributor may choose to work individually by posting an opinion e.g. localization, suggestion etc., or join a discussion desk to conduct 'Public-Hearing' with others on the selected topic. The discussion can be conducted via the provided 'Chat' frame before concluding an opinion. Any opinions or suggestions are committed to voting. Opinions can be different but majority votes will cast the belief of the community. These features naturally realize the online collaborative works to create the knowledge.

- **Individual or Group works**

Thought may be formed individually or though a concentrated discussion. KUI facilitates a window for submitting an opinion and another window for submitting a chat message. Each suggestion can be cast through the 'Opinion' window marked with a degree of its confidence. By working individually, comments to a suggestion can be posted to mark its background to make it better understanding. On the other hand, when working as a group, discussions among the group participants will be recorded. The discussion can be resumed at any points to avoid the iterating words.

- **Record of Intention**

The intention of each opinion can be reminded by the recorded comments or the trace of discussions. Frequently, we have to discuss again and again on the result that we have already agreed. Misinterpretation of the previous decision is also frequently faced when we do not record the background of decision. Record of intention is therefore necessary in the process of knowledge creation. The knowledge interpretation also refers to the record of intention to obtain a better understanding.

- **Selectional Preference**

Opinions can be differed from person to person depending on the aspects of the problem. It is not always necessary to say what is right and what is wrong. Each opinion should be treated as a result of intelligent activity. However, the majority accepted opinions are preferred at the moment. Experiences could tell the preference via vote casting. The dynamically vote

ranking will tell the selectional preference of the community at the moment.

3.3 ExpertScore

KUI heavily depends on members' voting score to produce a reliable result. Therefore, we introduce an adjustable voting score to realize a self-organizing system. Each member is initially provided a default value of voting score equals to one. The voting score is increased according to *ExpertScore* which is estimated by the value of *Expertise*, *Contribution*, and *Continuity* of the participation history of each member. *Expertise* is a composite score of the accuracy of opinion and vote, as shown in Equation 1. *Contribution* is a composite score of the ratio of opinion and vote posting comparing to the total, as shown in Equation 2. *Continuity* is a regressive function based on the assumption that the absence of participation of a member will gradually decrease its *ExpertScore* to one after a year (365 days) of the absence, as shown in Equation 3.

$$Expertise = \alpha \frac{count(BestOpinion)}{count(Opinion)} + \beta \frac{count(BestVote)}{count(Vote)} \dots\dots(1)$$

$$Contribution = \gamma \frac{count(Opinion)}{count(TotalOpinion)} + \rho \frac{count(Vote)}{count(TotalVote)} \dots\dots(2)$$

$$Continuity = 1 - \left(\frac{D}{365}\right)^4 \dots\dots\dots(3)$$

Where,

$$\alpha + \beta + \gamma + \rho = 1$$

D is number of recent absent date ($0 \leq D < 365$)

As a result, the *ExpertScore* can be estimated by Equation 4.

$$ExpertScore = \left(1 - \left(\frac{D}{365}\right)^4 \right) \times \left\{ \begin{array}{l} \alpha \frac{count(BestOpinion)}{count(Opinion)} + \beta \frac{count(BestVote)}{count(Vote)} \\ + \gamma \frac{count(Opinion)}{count(TotalOpinion)} + \rho \frac{count(Vote)}{count(TotalVote)} \end{array} \right\} \dots\dots\dots(4)$$

If *ExpertScore* is less than one, it is reset to one. The value of *ExpertScore* is ranged between 1 to 365 according to the accuracy and the rate of contribution of each member. This means that reliable members are rewarded better score for each vote. However, the expertise of the member is decreased according to the continuity of the participation. By means of the *ExpertScore*, we can rank the opinions precisely and

yield reliable results, especially for the results produced by an online community.

4 KUI for Multi-lingual WordNet Construction

There are some previous efforts in developing Wordnets for Asian languages, e.g. Chinese, Japanese, Korean (Choi, 2003), (Choi et al., 2004), (Kaji et al., 2006), (KorLex, 2006), (Huang, 2007) and Hindi (Hindi Wordnet, 2007). The number of languages that have been successfully developed their Wordnets is still limited to some active research in this area. However, the extensive development of Wordnet in other languages is important, not only to help in implementing NLP applications in each language, but also in inter-linking Wordnets of different languages to develop multi-lingual applications to overcome the language barrier.

We adopt the proposed criteria for automatic synset assignment for Asian languages which has limited language resources. Based on the result from the above synset assignment algorithm, we provide KUI (Knowledge Unifying Initiator) (Somlertlamvanich, 2006), (Somlertlamvanich et al., 2007) to establish an online collaborative work in refining the WordNets.

KUI allows registered members including language experts revise and vote for the synset assignment. The system manages the synset assignment according to the preferred score obtained from the revision process. As a result, the community WordNets will be accomplished and exported into the original form of WordNet database. Via the synset ID assigned in the WordNet, the system can generate a cross language WordNet result. Through this effort, an initial version of Asian WordNet can be fulfilled.

Table 1 shows a record of WordNet displayed for translation in KUI interface. English entry together with its part-of-speech, synset, and gloss are provided if exists. The members will examine the assigned lexical entry whether to vote for it or propose a new translation.

Car
[Options]
POS : NOUN
Synset : auto, automobile, machine, motorcar
Gloss : a motor vehicle with four wheels; usually propelled by an internal combustion engine;

Table 1 A record of WordNet



Figure 2 KUI Participation page

Figure 2 illustrates the translation page of KUI³. In the working area, the login member can participate in proposing a new translation or vote for the preferred translation to revise the synset assignment. Statistics of the progress as well as many useful functions such as item search, record jump, chat, list of online participants are also provided. KUI is actively facilitating members in revising the Asian WordNet database.



Figure 3 KUI Lookup page

Figure 3 illustrates the lookup page of KUI. The returned result of a keyword lookup is sorted according to the best translated word of each language. The best translated word is determined by the highest vote score. As a result, the user can consult the WordNet to obtain a list of equivalent words of the same sense sorted by the languages. The *ExpertScore* provided in KUI will help selecting the best translation of each word.

5 Conclusion

KUI is a platform for composing knowledge in the Open Source style. A contributor can naturally follow the process of knowledge development that includes

³ <http://www.tcllab.org/kui>

posting in 'Topic of interest', 'Opinion', 'Localization' and 'Public-Hearing'. The posted items are committed to voting within the community to perform the selection preference. The results will be ranked according to the vote preference estimated by the *ExpertScore* for the purpose of managing the multiple results. 'Chatlog' is kept to indicate the record of intention of knowledge composers. A contributor may participate KUI individually or join a discussion group to compose the knowledge. We are expecting KUI to be a Knowledge User Interface for composing the knowledge in the Open Source style under the monitoring of the community. The statistical-base visualized 'KUI look up' is also provided for the efficient consultation of the knowledge. We introduce KUI for Asian WordNet development. The *ExpertScore* efficiently ranks the results especially in the case that there are more than one equivalent.

References

- Bernardo Magnini, Carlo Strapparava, Fabio Ciravegna and Emanuele Pianta, 1994. *A Project for the Construction of an Italian Lexical Knowledge Base in the Framework of WordNet*, IRST Technical Report # 9406-15.
- Christiane Fellbaum, (ed.), 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.
- Chu-Ren Huang, 2007. *Chinese Wordnet*, Academica Sinica, Available at <http://bow.sinica.edu.tw/wn/>
- Gabor Proszeky, Marton Mihaltz, 2002. *Semi-Automatic Development of the Hungarian WordNet*, Proceedings of the LREC 2002, Spain
- Hindi Wordnet, 2007. Available at <http://www.cfilt.iitb.ac.in/wordnet/webhwn/>
- Hiroyuki Kaji and Mariko Watanabe, 2006. *Automatic Construction of Japanese WordNet*, Proceedings of LREC2006, Italy.
- J. Atserias, S. Clement, X. Farreres, German Rigau, H. Rodríguez, 1997. *Combining Multiple Methods for the Automatic Construction of Multilingual WordNets*, Proceedings of the International Conference on Recent Advances in Natural Language, Bulgaria.
- Key-Sun. Choi, H.S. Bae, W.Kang, J. Lee, E. Kim, H. Kim, D. Kim, Y. Song1, and H. Shin, 2004. *Korean-Chinese-Japanese Multilingual Wordnet with Shared Semantic Hierarchy*, Proceedings of LREC 2004, Portugal.
- Key-Sun Choi, 2003. *CoreNet: Chinese-Japanese-Korean wordnet with shared semantic hierarchy*, Proceedings of Natural Language Processing and Knowledge Engineering, Beijing.
- Korlex, 2006. *Korean WordNet*, Korean Language Processing Lab, Pusan National University, 2007. Available at <http://164.125.65.68/>
- Louw, M, 1998. *Polaris User's Guide*. Technical report, Belgium.
- Virach Sornlertlamvanich, 2006. *KUI: The OSS-Styled Knowledge Development System*, Proceedings of The 7th AOSS Symposium, Malaysia.
- Virach Sornlertlamvanich, Thatsanee Charoenporn, Kergit Robkop, and Hitoshi Isahara. *Collaborative Platform for Multilingual Resource Development and Intercultural Communication*, Proceedings of the First International Workshop on Intercultural Collaboration (IWIC2007), Japan

第17回日本翻訳連盟・翻訳祭パネル ディスカッションの報告

秋桜舎

代表 山本 ゆうじ

パネル ディスカッションの経緯

去る2007年10月17日に、日本翻訳連盟（JTF）が主催する第17回翻訳祭が開かれました。

翻訳祭の主な内容は2つの講演とパネル ディスカッションから構成されており、私はパネル ディスカッションの司会を務めさせて頂きました。パネル ディスカッションのテーマは「ツール活用で品質と効率の向上を両立する」、副題が「～翻訳ソフト／翻訳メモリー ワークフローの現状と未来～」というものでした。このテーマでこれだけの規模のディスカッションが行われたのは、おそらく世界でも初めての画期的なことだったと思います。ここにその内容を簡単にご報告させていただきます（なお、この記事は音声入力で執筆しています）。

参加者からの注目度は非常に高く、おかげさまで翻訳祭は史上最多数の入場者となったそうです。入場者数制限により当日参加できなかった方が多数いらっしまったようで、大変恐縮です。翻訳祭の様子をすべて収録したDVDが、JTFから発売されます< <http://www.jtf.jp/>>。ご関心のある方はそちらもあわせてご覧いただくとよいかと思います。

翻訳業界についてはなじみがない方もいらっしやると思いますので、簡単に経緯を説明します。これまでの翻訳祭でも、翻訳ソフトや翻訳メモリーツールのブースが地下の会場で開かれ、ここ数年は盛況を呈していました。しかし、これら翻訳支援ツールは、まだ翻訳者によって十分活用されているとは言えないのが現状です。一部では「食わず嫌い」の根強い偏見も見られます。

特に方向性の異なる2つの翻訳支援技術、翻訳ソフトと翻訳メモリー ツールをどのように組み合わせていくかが、今後の実務翻訳の業界では非常に重

要になります。そのために、今回のテーマが選ばれたわけです。

パネリスト紹介

パネリストとして、翻訳発注者、翻訳会社、個人翻訳者、翻訳支援ツール メーカーのそれぞれの立場から、以下の方々にご参加いただきました。（順不同）

- 脇田 早紀子氏（日本アイ・ビー・エム株式会社 ナショナル・ランゲージ・サポート）
- 塩津 誠氏（富士通株式会社 ソフトウェア事業本部 ミドルウェアコンポーネント事業部 プロジェクト部長）
- 永島 和暢氏（ライオンブリッジ ジャパン株式会社 ソリューションズ アーキテクト）
- 岡野 徹氏（SDL トラドス ジャパン LSP Business Manager）
- 柳 英夫氏（株式会社サン・フレア GMS ソリューション事業部 チーフコンサルタント）
- 川原 勝氏（株式会社十印 サポートセンター兼 エンジニアリング部 シニア マネージャ）
- 久徳 省三氏（実務翻訳者、JTF 翻訳支援ツール 委員長）

パネル ディスカッションの流れ

パネル ディスカッションは、以下のような流れで行われました。

- MT と TM について概説
- 各社の翻訳支援ツールの紹介
- 論点の提示
- ディスカッション
- 質疑応答

最初に MT と TM について概説したのは、参加者

の背景が多様であり、また機械翻訳と翻訳メモリー
の特性について説明する必要があったためです。確
認したところ、今回の参加者のほとんどは翻訳メモ
リーを使っていましたが、翻訳ソフトを使っている
人が非常に少なかったようです。

特に「翻訳支援と自動翻訳」の違いについて説明
を行いました。翻訳業界では、機械翻訳による翻訳
支援については、残念ながらあまり理解されていま
せん。「機械翻訳といえば自動翻訳」という思い込
みが未だに根強いようです。「自動翻訳」では、「ボ
タン一つでソフトにお任せ」であるのに対し、「翻
訳支援」では「予測不可能な翻訳ソフトを指定通り
に動かす」に動かすという違いがあります。「翻訳
支援と自動翻訳」の明確に区別することにより、翻
訳ソフトの役割がはっきりします。これはつまり
「翻訳ソフトに何をさせるべきで、何をさせるべき
でないか」ということでもあります。

さまざまな翻訳支援ツール

続いて各社の翻訳支援ツールの紹介を行いまし
た。翻訳ソフトは別にして、以下のようなツールが
あります。

翻訳メモリー ツール

- SDL Trados (SDL)
- Logoport (LionBridge)
- TransAssist (インターメディア)
- Tratool (ロゼッタ)
- 対訳君 (MCL)

翻訳ワークフロー管理ツール

- Transit (シュタール)
- Idiom WorldServer (Idiom)

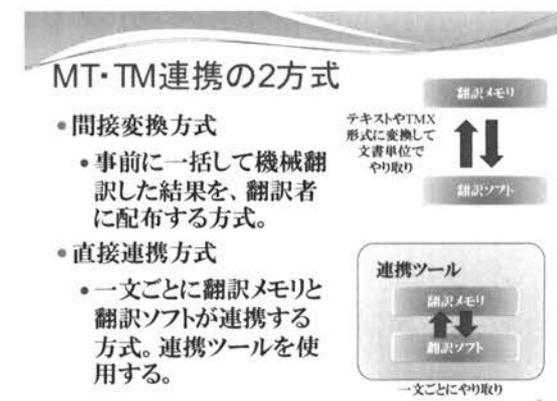
翻訳ソフト・翻訳メモリー連携ソリューション

- SATILA (秋桜舎)

これは網羅的なリストではなく、あくまで一例に
過ぎません。便宜的に以上のように分類しまし
たが、翻訳メモリー ツールが翻訳ワークフロー管理
ツールの役割を含んでいたり、その逆であったりす
ることもあります。

MT・TM 連携の 2 方式

このようにさまざまな翻訳支援ツールがある中
で、翻訳ソフトと翻訳メモリー ツールを連携させ
るには、以下のような 2 つの方式があります。



従来の方式は間接変換方式であり、翻訳者は翻訳
ソフトの出力だけを受け取ります。この場合、翻訳
者は翻訳ソフトを直接操作しないため、翻訳者が翻
訳ソフトの操作を覚える必要がないという利点があ
ります。ただし機械翻訳の過程に翻訳者が介入で
きないため、最初から十分にカスタマイズ・調整さ
れた辞書を使わない限り、きわめて完成度の低い出
力になります。また問題点が見つかって、それを
次の出力に反映させることができません。

これに対して直接連携方式では、翻訳者が翻訳ソ
フトを直接操作します。翻訳ソフトに任せ切りにす
るではありません。この場合翻訳者には高度な技
能が要求されます。しかし機械翻訳の過程に翻訳者
が直接介入し、問題点を修正してゆくことで、辞書
の完成度を高めてゆくことができます。私が秋桜舎
で開発している SATILA は、直接連携方式を採用し
ています。最終的に間接変換方式を使うとしても、
まずは直接連携方式により辞書の練成を行うこと
で、結果を大きく改善できます。

パネル ディスカッションの論点

このように議論の前提となる基本事項を確認し
た後で、パネル ディスカッションの論点を提示し
ました。今回は、クライアント、翻訳会社、翻訳者
それぞれの立場から、以下の論点について議論しま

した。

現状確認

- 翻訳ソフトとメモリーはどこまで活用されつつあるか
- 今なぜ翻訳ソフトなのか
- 翻訳ソフトで解決できる翻訳業界の課題

今後

- 翻訳ソフト導入の障害はなにか
- 翻訳ソフトとどのように組み合わせるべきか

キーワード

議論では以下のキーワードを提示しました。

- 高技能翻訳者
- 翻訳支援
- 「翻訳者にとっての翻訳支援とは」
- 用語集
- ユーザー辞書
- 用語管理者（ターミノロジスト）

論点・現状確認 1

「翻訳ソフトとメモリーはどこまで活用されつつあるか」

ここからが具体的な論点です。まず、SDL Tradosの岡野様から、企業での翻訳ソフトの実際の活用例をご紹介いただきました。専門化した役割分担、3～5倍程度の効率向上のグラフなど、具体例として非常に興味深いものでした。この件については本誌の別記事でも扱われていますので、合わせてご参照ください。

続いて久徳様からも、ご自身で翻訳ソフトを活用した和文英訳の場合の効率向上を表とグラフで紹介いただきました。日数が経過して辞書が蓄積されていくにつれ効率が上がることを具体的な数値で示されました。今後もさらにこのような事例の紹介が増えてくることで、翻訳ソフトの有用性が認識されることと思います。

また、翻訳ソフトの使われ方について、富士通の塩津様から「現状では読み捨て翻訳に使われることが多いが、今後は市場拡大のために翻訳者の方の意

見も取り入れていきたい」とのご解説・ご意見を頂きました。

IBMの脇田様からは「翻訳ソフトには関心があり、実験をしているがなかなかうまくゆかない。翻訳ソフトに感わされないように、翻訳者の訓練が必要」というご意見を頂きました。

サン・フレアの柳様からは、「現状でも、翻訳資産としての翻訳メモリーをクライアントと共有するのに困難がある。翻訳ソフトでも同じような問題が予想されるので、その解決が重要ではないか」とご指摘を頂きました。

ライオンブリッジの永島様からは、主にヨーロッパ言語間での、間接変換方式での実例をご紹介いただきました。生産性の向上についてはまだ評価中とのことでした。

十印の川原様からは、「目に見える翻訳市場以外にも社内文書の翻訳のニーズがあり、人間の翻訳とは違った使われ方がある」とのご意見を頂きました。

論点・現状確認 2

「今なぜ翻訳ソフトなのか」

次の論点については、まず久徳様から「翻訳ソフトは単にコストダウンのためだけに使われるものではない。IT分野以外でもより合理的な取り組みが必要だ」とのご意見を頂きました。

脇田様からは「誤訳、訳抜けのない、付加価値の高い翻訳を行うためにも翻訳ソフトを活用できるはず」とご意見を頂きました。

ここで、司会の立場から「翻訳ソフトの利点」として以下のような点を挙げて、パネリストの方のご意見を伺いました。

- 自動適用による用語の正確さ・効率化
- 言い換え表現の蓄積—再利用と共用
→教育的効果
- 専門辞書の単語数
→不適切な訳語はユーザー辞書でカバーする
特に、翻訳メモリーの文単位での翻訳知識の再利

用に加えて、以下のような意識フレーズの翻訳知識を辞書として蓄積し、自動的に共有・再利用できることをご紹介しました。

- “in the same boat”→「同じような状況にある」
- “be shuffled to the sidelines”→
「指をくわえて見ている」
- “for one”→「たとえば」
- “take an active part” →「積極的に参加する」

また、翻訳にかかわる立場の違いから、「翻訳支援技術はだれの利益になることか」という問題提起を行いました。これに対して、柳様から「単なる語数競争でない付加価値を生み出すことが重要」とのご意見を頂きました。永島様からは「翻訳メモリーが浸透してきたように、やがては翻訳ソフトを使える技能が業界で必須になるかもしれない」とのご意見を頂きました。岡野様からは「用語集が自分で準備できないメーカーに広まるのは難しいのではないか」というご指摘を頂きました。

脇田様のご質問に対し、翻訳ソフトの対訳エディタでの、適切な訳語選択のデモをお見せしました。川原様からは「(適切な訳語選択のための) カスタマイズの費用と労力はだれが負担するのか」というご質問を頂きました。これに対しては「余分な労力が新たに発生するのではなく、翻訳工程全体を見た場合、翻訳会社がその労力を負担しても元は取れるだけの効率化は可能である」とお答えしました。

さらに、司会の立場から、より体系的な「翻訳工学」に向けて、以下の事項が重要になるということをご紹介しました。

- 合理的な方法論
- 個人プレーではない、チームワークとしての実務翻訳
 - 用語集
 - 「実装可能なスタイル ガイド」
- 翻訳資産の共有と再利用

論点・現状確認 3

「翻訳ソフトで解決できる翻訳業界の課題」

紙数の関係上、以下は論点と司会からの情報のみを挙げます。

単純な置換ツールの限界として以下の問題点を挙げ、翻訳ソフトではこれらの点に対応可能であることを実際に実演してご紹介しました。

- 専門辞書がない
 - 辞書をすべて自分で作る必要がある
- 語順が原文のまま
 - 入れ替えが必要。修正が手間
- 活用形に対応できない
 - それぞれの活用形ごとに辞書作成？

論点・今後 1

「翻訳ソフト導入の障害はなにか」

翻訳ソフトの課題として以下の点を挙げました。

- 使いこなすには一定の技能が必要
英語・日本語・翻訳・パソコン
 - 翻訳精度の不足
 - = ユーザー辞書の活用が不十分
 - ユーザー辞書の不足
 - = きめ細かい制御が困難
- ここで、AAMT での共有ユーザー辞書仕様 UTX の取り組みについてもご紹介しました。

論点・今後 2

「翻訳ソフトと翻訳メモリーをどのように組み合わせるべきか？」

この他にも、パネリストの方々からの情報提供、ご提案、多数の興味深いやり取りが活発に行われました。ご関心のある方は、JTF ウェブサイトから DVD を入手して全体をご覧いただければと思います。

The翻訳サーバ Enterprise Edition V3.0

東芝ソリューション株式会社

1. はじめに

東芝ソリューション株式会社は、新たに中日／日中翻訳を実現したサーバ型翻訳システム「The翻訳サーバ Enterprise Edition V3.0」を発売しました。

中国に拠点を持つ日本企業では、現地情報の早期把握や中国の拠点との情報共有など、中日双方向の翻訳を効率的に行いたいというニーズが高まっています。これを受けて、当社は「統計ベース中国語解析技術」を用いた中日／日中翻訳を新たに実現しました。

翻訳エンジンに、統計ベース中国語解析技術を付加することにより、英日翻訳で使用しているきめ細かな訳し分けを活かした中日翻訳を実現しています。

The翻訳サーバは、大陸で広く使用されている簡体字と、香港を中心とした地域で使用される繁体字の両方に対応しています(表1)。また、中日／日中の双方向とも「自動車」・「ビジネス」・「情報・通信」・「電気・電子」・「機械」の5分野の専門用語辞書を装備しています。更に、ユーザ辞書・翻訳メモリ辞書の作成も可能で、新しい単語や用例を追加していくことができます。

2. The翻訳サーバ Enterprise Edition V3.0 の特長

(1) 高度な技術で中国語の翻訳を実現

中国語の文は、漢字のみで語間の切れ目がないうえに、一つの単語が複数種類の品詞をもつことも多いため、日本語や英語に比べて文の解析に大きな曖昧性が生じる可能性があります。このため、高速で高精度の機械翻訳を実現することは難しいとされてきました。今回当社は、さまざまな中国語の文書から単語の境界や品詞、構文構造に関する知識を収集し、文の解析にその知識を利用する「統計ベース中国語解析技術」を開発しました。英日翻訳で高い翻訳精度を誇る当社の

(2) 大量導入や業務アプリケーションへの組み込みが容易

The翻訳サーバは、クライアントとしてインターネット・エクスペローラを利用して翻訳を行えます。翻訳結果はインターネット・エクスペローラ上に表示する、あるいはメールで受信するという形で得られます。クライアントに専用ソフトウェアが不要なため、利用者が多数でも翻訳機能の導入が容易におこなえます。

また業務アプリケーション等に翻訳機能を組み込むインターフェースを提供しています。このため、利

表1：中国語の翻訳例

(簡体字の場合)		
中日	原文	以下简要介绍与企业的社会责任有关的活动的主要内容。
	訳文	以下に簡単に企業の社会責任と関係のある活動の主要な内容を紹介する。
日中	原文	保証書の規定にしたがって、販売店が修理させていただきます。
	訳文	按照保证书的规定请让销售店修理。
(繁体字の場合)		
中日	原文	從2000年引進市場起，累計出貨量已經超過4千萬台。
	訳文	2000年に市場に導入してから、累計出荷量は既に4千万台を超えた。
日中	原文	相互理解の促進のために、積極的な情報開示とコミュニケーションを行います。
	訳文	為了相互理解的促進，進行積極的信息公開和交流。

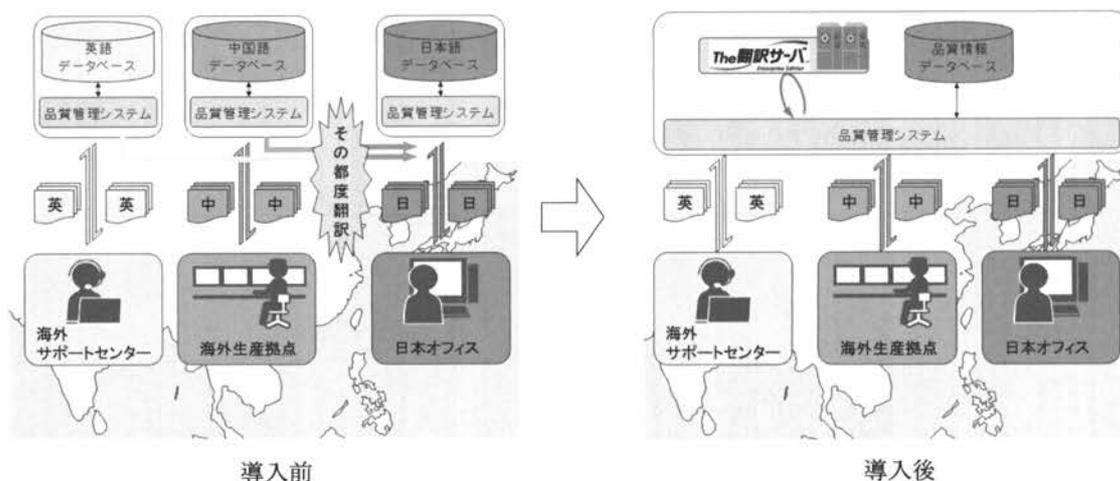


図 1：システムへの組み込みイメージ図

用している環境やワークフローを変えることなく、日本語文書しか扱えなかったアプリケーションで英語や中国語の文書が扱えるようになります。

例えば品質管理システムにおいて、データベースが日本語、英語、中国語で独立・併用されている場合、日本語環境から英語や中国語のデータベースを参照・閲覧するには、それぞれの言語での検索を行い、本文をその都度翻訳する作業が必要でした。このようなシステムにThe翻訳サーバを組み込んで、データベースへの登録時や参照時に英語や中国語を日本語に翻訳すれば、情報の検索や閲覧も日本語で行うことが可能となります(図1)。これによりデータベースの一元管理・共有、既存システムの多言語化が容易になり、日・英・中混在業務の効率アップが図れます。

(3) 英日/日英翻訳も強化

The翻訳サーバ Enterprise Edition V3.0では英日/日英翻訳も強化されています。

同義語や類義語といった言葉の概念を体系的に記録した「概念辞書」を利用して、文意に合った訳語を選択する「概念翻訳」を新たに搭載したほか、基本辞書が202万語に、専門用語辞書が27種類494万語に増加しました。

更に、目的に応じた翻訳環境の条件設定を複数記録し、それらをメニューで簡単に選択できる「翻訳マルチ環境」を搭載しています。

3. 動作環境

対応OS：(全て日本語版)

Windows® 2000 Server SP1/SP2/SP3/SP4,

Windows® Server 2003 SP1/SP2, R2

SPARC版Solaris™ 9, SPARC版Solaris™ 10

- 商品の詳細は以下URLをご覧ください。
<http://mt-server.toshiba-sol.co.jp/>
- The翻訳、The翻訳サーバは、東芝ソリューション株式会社の登録商標です。
- Windows、Internet Explorer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。
- Solarisは、米国Sun Microsystems, Inc.の米国及びその他の国における登録商標または商標です。
- 記載されている商品の名称は、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。
- The翻訳サーバは、特許第3408291号、第3768105号その他の登録済み特許および出願中の特許が使用されています。
- 本製品の仕様は予告なく変更される場合があります。

トラベルウォーカー 中国

株式会社高電社

はじめに

株式会社高電社は、スマートフォンの OS として急速に普及している Windows Mobile に対応した旅行会話文例ソフト「トラベルウォーカー 中国」を開発、販売を開始いたしました。「トラベルウォーカー 中国」は、従来の電子辞書に収録されている文例集と、機械翻訳のそれぞれに見られる問題点を克服するために、自由度と精度を両立したソフトウェアです。

従来の文例集と大きく異なる点として、12の場面別に収録されている2,700文例のうち、約800文例で単語の置換や、日本語から中国語への翻訳が可能であることが挙げられます。また、翻訳は日本語の名詞(句)のみを対象としているため、機械翻訳の問題点である誤訳の発生率を最小限に抑制できます。

さらに、ネイティブレベルの中国語 TTS 機能も付属しているため、音声による意思疎通も実現します。

今後、ますますの普及が見込まれるスマートフォンなどの Windows Mobile 端末に「トラベルウォーカー 中国」をインストールして、中国語圏でのハンディなコミュニケーションが実現すると確信しています。



図1：スマートフォンでの利用例

表1：文例ソフト・機械翻訳・

「トラベルウォーカー 中国」の機能比較

	文例集ソフト	機械翻訳	トラベルウォーカー 中国
文章のカスタマイズ	×	○	○
文章の正確性	○	△	○
単語の翻訳	×	○	○
文章の翻訳	×	○	×

主な利用方法例

ここでは、「場面を選択」メニューから任意の文例を選択して、その文例をカスタマイズするフローを示します。

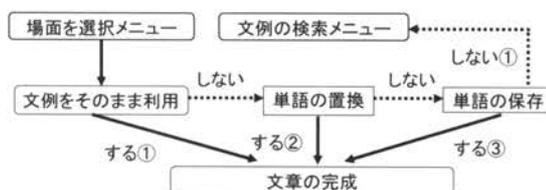


図2：文例のカスタマイズ有無による文章完成フロー

各フローの画面について、「空港」という場面に表示されている「免税店はどこですか。」という文例を例に説明する。まず、「場面から探す」→「空港」→「案内所」→「尋ねる」の順に選択し、図2の画面を表示します。

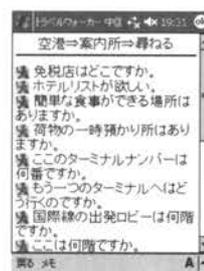


図3：文例選択画面

「する①」

初期値の文例をそのまま利用する場合には、日本語、中国語、ピンインが表示されます。

画面下左側のスピーカーアイコンをタップすると、中国語テキストを音声合成で朗読します。「登録」ボタンをタップすると、「お気に入り文例」として登録されます。

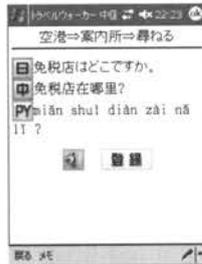


図4：文例を初期値のまま利用した例

「する②」

出口、トイレ、両替所などの単語が、置換可能な選択肢として表示されます。

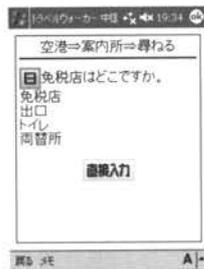


図5：置換できる単語例

「する③」

入力した日本語が中国語に翻訳されて、文例に挿入されます。



図6：「バス乗り場」という単語を入力した例



図7：「バス乗り場」が中国語に翻訳され、文章が完成した例

「しない①」

前述の「する①」～「する③」のいずれでも文章を完成できなかった場合には、「文例を探す」機能を利用して、特定のキーワードを含む文例を検索することによって、より目的とする文章にヒットする確率が高くなります。

翻訳をサポートする機能

ネイティブレベルの中国語 TTS

中国現地でのコミュニケーションに重宝する、ネイティブレベルの中国語音声合成エンジンを採用しています。

20万語を収録の「ミニ日中辞典」

日本語を入力して、該当する中国語の単語を検索、表示する機能です。複数の中国語単語が表示されることも多く、同義語の習得にも効果を発揮します。



図8：「呼ぶ」に相当する複数の中国語単語を表示した例

写真付き「グルメ辞典」

中国大陸のみならず、香港、台湾などの食文化を堪能する際に重宝するグルメ辞典です。800のメニューのうち、約500については写真も掲載されているので、レストランで店員に見せるだけでオーダーが可能です。



おわりに

その技術段階から、不完全な翻訳結果が散見される機械翻訳の弱点を補っただけでなく、従来の電子辞書に見られる文例集に単語翻訳機能を追加したことにより、ユーザーニーズに柔軟に対応できると考えます。中国との関係が深まるにつれ、ますます中国語翻訳への注目が高まることは必至であり、Windows Mobile を搭載したハードウェアの普及と共に、手のひらサイズのコミュニケーションツールとして「トラベルウォーカー 中国」がビジネスに、旅行にと広く受け入れられることを願っています。

動作環境

対応 OS

Windows Mobile2003 / 2003SE / 5.0 / 6 Classic /

6 Professional

提供メディア

256MB microSD カード

フォント

日本語 OS に依存

中国語 Unicode 対応 GB18030 明朝体

商品構成

256MB microSD カード (プログラム書き込み済み)

リカバリー用 CD-ROM・ユーザーズマニュアル

ユーザー登録ハガキ・商品構成確認書

対応機種例

ウィルコム W-ZERO3 シリーズ

ソフトバンクモバイル X01HT / X02HT

NTT ドコモ hTc Z

イー・モバイル EM・ONE

日本 HP iPAQ シリーズ

製品情報 URL

<http://www.kodensha.jp/soft/walker/rw/>

<http://plusd.itmedia.co.jp/mobile/articles/0712/07/news076.html>

http://k-tai.impress.co.jp/cda/article/news_toppage/37426.html

AAMT会員のひろば

現在、学者・研究者・翻訳者・流通業者など多岐にわたる AAMT 会員の交流は、年一回の総会や不定期のセミナー・研究会に限られている現状があります。もっと気軽にお互いを知ることができ、ニーズが合えば親密な交流を図ることが出来る場を AAMT Journal 誌面上で提供するべく、今号より「AAMT 会員自己紹介企画（仮題）」をスタートいたしました。会員の皆さまよりご寄稿いただいた、自己紹介・機械翻訳および翻訳業界に期待すること・AAMT へのご要望をご紹介します、会員間での新たな交流が生まれるきっかけとなることを期待しております。

第一弾の今号では、法人会員 1 社、個人会員 6 名の皆さまよりご寄稿いただきました。次号以降も本企画を継続していくことを考えておりますので、会員の皆さまからのご寄稿を心よりお待ちしております。本企画へのご寄稿・お問い合わせは AAMT 事務局(E-mail: aamt-info@amt.info)まで宜しくお願いいたします。

法人会員（敬称略）

会員名 財団法人 日本特許情報機構／JAPAN PATENT INFORMATION ORGANIZATION
自社紹介 財団法人 日本特許情報機構（JAPIO）は特許庁、独立行政法人工業所有権情報・研修館(INPIT)の委託を受けて公開特許公報英文抄録（PAJ）および米国特許明細書の和文翻訳抄録作成等の翻訳業務を行っています。これら特許翻訳業務において統一的な訳語の選定等、翻訳品質を向上させる目的で機械翻訳技術を利用しています。 PAJ は外国特許庁での先行技術調査や、日本人が米国出願した際の IDS（情報開示陳述書）として活用されています。これら翻訳結果は INPIT の特許電子図書館(IPDL)から無料で検索・参照することができますのでご利用ください。 特許電子図書館： http://www.ipdl.inpit.go.jp/Tokujitu/tokujitu.htm JAPIO のホームページも是非ご覧下さい。 http://www.japio.or.jp/
機械翻訳および翻訳業界に期待すること 機械翻訳技術がさらに発展して特許の翻訳業務の効率化・低コスト化に役立つとともにわが国の特許情報がより効率的に発信され、また我が国の産業界が外国の特許情報をより簡便に利用できるようになることを期待しています。 Japio は 2007 年 4 月、特許情報研究所を開設しました。当研究所を窓口として皆さんの力をお借りして特許翻訳の効率化・低コスト化の実現に努力したいと思います。

個人会員（敬称略・50音順）

会員名・所属
相沢 輝昭（フリー・広島市立大学名誉教授）／AIZAWA, Teruaki
E-mail
機械翻訳および翻訳業界に期待すること
『仏文和訳のこの頃』 自己紹介を兼ねて— なぜか「個人翻訳家」として入会し、ようやくそれらしくなってきた。このところずっと、ほぼ百年前パリで出版された T. de Wyzewa と G. de Saint-Foix によるフランス語の大著“Mozart”を日本語に翻訳する毎日。そんな古い本をなぜ、という問題はさておき、フランス語に堪能でない私がなぜ、という問には答えられる。フランス語やドイツ語は、私の大学時代、あこがれの言語だった。ところが慶応大学の鷺見洋一教授『翻訳仏文法』（ちくま学芸文庫、2003年）によれば「大学でフランス語を履修する学生数が激減した」「フランス文学専攻の在籍者数は、この20年間で、おおげさでなく半分以下から四分の一にまで落ちてしまった」とある。ドイツ語の惨状もそんなところだろう。一方フランス語もドイツ語も英語と同じく、あるいは英語以上に、豊かな文化遺産を誇っている。そうなれば、暴虎馮河のそしりを恐れず、語学の非専門家が立ち上がるしかないのではないか。ただし、素手でなく、できればコンピュータの手も借りたい。 以下、私の経験から、語学の非専門家とコンピュータとの関わりようを思いつくまま述べてみよう。 仏文和訳をやっていてつくづく実感するのは仏和辞書の充実である。大抵のことは辞書に載っている。日本におけるフランス語学のレベルの高さと言うべきだろう。ただ、 (1)成句がどの構成単語からでも引ける とありがたい。例えば、 <i>jamais de la vie</i> が <i>jamais</i> からでも <i>vie</i> からでも引ける、というようなこと。 用例に助けられることも多い。いっそ、 (2)任意のフレーズをキーとして大規模コーパスから用例が検索できる といい。和訳が付いてなくても、用例が列挙されるだけでも随分と参考になるはずだ。 フランス語の動詞変化にはまったく目がくらむ。人称・法・テンスに従って 87 通りに変化し、さらに現在分詞と過去分詞がある。非専門家には覚え切れない。つい、 (3)動詞活用形からそれに対応する人称・法・テンスをさっと教えてくれ と言いたくなる。 フランス語を日本語に「名訳」するには機械翻訳で言うところの <i>interlingual</i> 方式つまりは徹底的意識しかなさそうだ。構文を引きずった <i>transfer</i> 方式は特に仏文和訳には不向きと思われる。鷺見洋一『翻訳仏文法』には、J.ギュマン＝フレッシュールの『英仏比較統辞論—翻訳の諸問題』に従って興味深いアスペクトの英仏差が示されているが、ここでは川端康成『雪国』の冒頭文で日仏差を比べてみよう。 (4)国境の長いトンネルを抜けると雪国であった。 これを Armel Guerne は次のように訳している。 (5)Un long tunnel entre les deux régions, et voici qu'on était dans le pays de neige. これを再び日本語に直訳すれば、例えば、 (6)2つの領域の間の長いトンネル、そして私たちはここ雪国にいた。

ダイナミックな日本文(4)に対して、スタティックな仏文(5)。日仏翻訳(4)→(5)はいわば動的→静的変換である。結果的に、ゼロ代名詞補完が必要なくなっている。ちなみに Seidensticker の英訳では train が補完されて、

(7)The train came out of the long tunnel into the snow country.

鷺見先生の本にもマーク・トウェイン『トム・ソーヤー』の一節とその仏訳が載っていて、「まるで映画の数場面のように生き生きと進行する英文と比べあわせたら、仏訳はさながら古ぼけた記念写真か何かのようにどっしりと構えて動かない印象がないだろうか」と注釈されている。

仏文和訳に逆転して考えれば、ここには仏文(5)に対する和訳が2つある。(6)はちっとも面白くない transfer 訳、(4)は素晴らしい interlingual 訳。仏文和訳はすべからず(5)→(4)のようであるべし。訳者の想像力と一種のゲーム感覚が試されるところだ。

機械翻訳にとって「絶望的」な話を続けると、語学の非専門家である私が翻訳作業の拠り所になっている言葉が2つある。1つは松浦寿輝東大教授(芥川賞作家でもある)の「翻訳の質を決めるものは語学的知識の多寡ではない。他人の思考についてもっともよく考えようとする者が選び取る生の姿勢によって、翻訳はもっとも高度な倫理の表現となる」であり、もう1つは村上春樹の「何度も何度も読み直して、何度も何度も書き直して、人に読んでもらってまた書き直すということが続ける」である。

Interlingual の行き着く先には、ひょっとすると倫理があるのかもしれない。しかし、コンピュータからはいかにも遠い。また、再翻訳の都度、単なる次候補でない「別訳」を出してくれるようなシステムは作れるだろうか。訳文生成の問題のような気もするが.....

ネットワークには大いに助けられている。仏文解説にいよいよ窮するとグルノーブル GETA の Christian Boitet 教授にメールを出す。私のまったく知らないラテン語やイタリア語にもそれぞれメールの先に先生が就いてくれている。こないだは人名 Ruggiero と Roger の関係に戸惑い、結局、私のイタリア語の先生がイタリア人に質して、前者はイタリア名、後者是对応するフランス名とわかった。Jean-Johan-John に類する系列らしい。実際、固有名詞の翻訳にはいつも悩む。ついでに注文を出しておこう。

(8)各国語対応人名データベースの整備

Boitet 先生が来日されるとまとめて質問に答えて頂く。最近では昨年11月京都でお会いした。このときは、Mike Dillinger 博士(アメリカ機械翻訳協会会長)、田中英輝博士(NHK 技研)も同席して下さって、訳語を探しあぐねていた *grâce passionnée* の意味をあれこれ議論した。「情熱的優美さ」では落ち着かない。モーツァルトのドン・ジョヴァンニの影もチラついて私を惑わしているようだ。Dillinger さんは言われた。Grâce は女性の自然で優美な身のこなしのようなもの。自然なものに *passionnée* が係るのはおかしいが一種の oxymoron (矛盾語法) だろう。なるほど、論理的にまっとうな訳語ペアばかり考えてはいけないのだ。それ以来、ルーヴル美術館とその前庭のガラスのピラミッドがペアになって思い浮かぶようになった。

語学の非専門家とコンピュータとの関わりようをここまで書いてきて、本当のところ私は何が言いたかったのだろう、と改めて考える。英語一辺倒ではなく、文化的にメジャーなフランス語やドイツ語も忘れないで、という思いもある。とはいえ、商売を考えれば、今は仏日機械翻訳システムなど作れる状況にはない。が、今後を期して、それに備えることはできるのではないか。私のような語学の非専門家のためのコンピュータ支援システムと第2言語習得のための e-learning システムは1つにまとめられそうな気がする。そして、その構築には、これまで培ってきた機械翻訳のノウハウが大いに役立ちそうに思える。

<p>会員名・所属</p> <p>貝島 良太 (有限会社アトリエ・ワン) / KAIJIMA, Ryota</p>
<p>E-mail</p>
<p>自己紹介</p> <p>私は、2002年4月1日にAAMTの個人会員になりましたが、以前在籍した会社(当初の社名は株式会社国際ビジネスサービス、後に日立国際ビジネスに改名)で JAMT 発足時から法人会員になっていましたので最古参の会員ともいえます。前職で、当初私は翻訳部門を担当しておりました。その後開発部門を発足させ、用語集超活用ソフト HT³ (エイチティースリー)と改良版の SuperHT³ (スーパーエイチティースリー)を開発しました。2001年6月に同社を退職し、SOHOで引き続き同ソフトに関する業務をしています。</p> <p>以下、私と機械翻訳の馴初めと、そのソフトについて述べます。</p> <p>【私と機械翻訳】</p> <p>私が機械翻訳に初めて触れたのは1984年1月のことです。当時私は株式会社日立製作所の輸出営業所コンピューター部に在籍していました。新年早々に事業部長から突然呼ばれ、「明日から川崎市にあるシステム開発研究所に行ってもら。そこで開発中の日英機械翻訳を手伝ってこい。2年以内に立ち上げて、機械翻訳を事業の柱とする翻訳会社を設立したい」と言われました。</p> <p>それまで私は機械翻訳のことを SF 小説や鉄腕アトムの世界のことだと思っていました。専門知識がゼロなのでまずは辞退しましたが、「開発を横から見て、評価や改良の提案をすればよい」とのことでしたので結局は宮仕えの悲しさ、引き受けました。それから2年間、研究所通いが始まりました。</p> <p>当時は、パソコンはもちろんありません。研究棟にある端末機から、日本語を入力しておく、計算機センターにあるホストコンピューターで翻訳処理され、翌朝センターにプリントアウトを取りに行くというものでした。その紙の束を取ってきて、英語の出来を評価するのです。この2年間、シストラなど他社の機械翻訳のデモを見学したり、各社の開発者との意見交換をしたりしました。スタンフォード大学で開催された COLING'84 にも参加したり、テキサス大学の言語研究所に、英文生成に必要な英語の冠詞と不定冠詞の使い分け規則を作ってもらったり、New York の Smart 社を訪問し PEP (Plain English Program: 平易な技術英文を書くソフト)を導入する検討もしました。</p> <p>蛇足ですが、当時の日立の日英機械翻訳のニックネームは ATHENE/N (ギリシャ神話の学問の女神)でしたが、アテナは「(翻訳結果が)合ってねー」に通じ語呂が悪いからと、途中で HICATS JE (ハイキャッツジェイイー)という猫のような名前に変わりました。</p> <p>【機械翻訳の問題点】</p> <p>機械翻訳の質を上げるために文法規則や辞書の整備をしなければなりません。その規則と辞書を作るための例文に対しては、当然良い結果が出るようになります。しかし、人間にとって例文と同程度と思われる文章でも機械から見ると別物で、でたらめな訳文が出るのがしばしばあります。これを回避するためにさらに規則を追加しますが、次第に規則同士に矛盾が生じて自滅するようになります。</p> <p>そういう訳で当時から機械翻訳はエンドレスな開発となり、問題は山積でした。</p> <p>研究所での2年の刑期(?)を終えるにあたり「機械翻訳を翻訳事業に導入するのは時期尚早。経済性を無視すれば可能」という社内報告書をまとめました。確か、当時の試算で、機械翻訳は通常の翻訳者による翻訳の約2倍の費用がかかると思われました。手書き原稿からのテキスト入力費用、プレエディット費用、CPU 使用料、ポストエディット費用、</p>

更に高額なハードとソフトのリース料を見込むからです。特に、当時の汎用コンピューターの使用料は割高で1文章訳すのに10円から50円程度かかっていたと思います。さらに始末が悪いのは、機械にとって難解なものは相対的にCUP使用時間が長くなり、その挙句最後に日本語交じりの、訳文とはいえないものが出てくるのです。失敗したもののほど金がかかるという恐ろしいものでした。

1987年に前述の新会社が設立され、私は新会社の翻訳部門に移ることになりました。結局その報告書が功を奏し機械翻訳を導入しないで普通の翻訳業務を含むドキュメンテーション事業を立ち上げました。そして1991年JAMTの発足に当たり、最新情報を得るために法人会員になりました。

機械翻訳に携わる多くの方々に完成時期をお伺いすると、口をそろえたように「あと2年もあれば何とかなる」と言われました。あれから20年以上経つのですが、やはり「あと2年」はかかるのだらうと思います。

最大の問題点は自由入力された原文に簡単には対応できないことです。制限文法で書かれた原文であれば、まあまあ結果が出ますが、素人では使えません。自由入力文を制限文法で書き直すのに手間(お金)がかかりますし、書き直したら必ず良い結果が出るか分からないという点は永久に残ります。現在は、コンピューターの環境が20年前と比べて隔世の感がありますから、自分のパソコンで手軽な機械翻訳ソフトで遊んでみるすることができます。入力のしやすさや手軽さなど進化は十分感じますが、翻訳性能自体は20年前の研究所レベルと大差ないように感じます。

【HT³の誕生】

前職でのドキュメンテーション業務は、マニュアルの日英翻訳が大半を占めていました。多くの場合顧客からは原稿と共に用語集が支給されます。当然顧客は用語集にある訳語は翻訳文に使われていることを期待します。しかし、翻訳作業にはその分野に強い翻訳者を当てますから、翻訳者自身が用語をよく知っており、支給された用語集をあまり参照しないで翻訳をすることになります。そうすると、顧客指定の訳語があまり使われない翻訳になり、顧客からのクレームを受けることになります。

現状の機械翻訳は業務用にはまだ使えない。だから翻訳者に翻訳してもらわなければならない。しかし、用語集の訳語を思うようには使ってもらえない。用語集の用語が翻訳にきちんと使用されたかの検収にも多大な時間がかかる。

そこで考案したのが、次のような特徴を持つHT³でした。

- 翻訳原稿に訳語付与を完全に行う
- 翻訳結果の用語確認
- 用語集を日英/英日双方で使用(リバーシブル)
- 異表記にも対応
- 用語集はExcelで作成
- 複数の用語集を優先順位をつけて串刺しで使用
- 翻訳のヒントや解説を参照可能

このアイデアのユニーク性が認められ、1999年に日本、2000年に米国の特許を取得しました。

HT³は1996年6月ごろから開発を開始し、翌年1月に製品化しました。財団法人日本規格協会とタイアップし、1998年にJIS工業用語大辞典第4版対応のCD-ROM with HT³を発売いたしました。

【SuperHT³はこんなソフトです】

HT³のユーザーから、高速処理、図表や式の表示、ボールド/イタリック/上付き文字/下付き文字の文字属性への対応、用語の重複登録等々多くの要望が寄せられました。これらは改良では済まされるものではありませんでした。結局新しいプログラムを開発し、**SuperHT³**と命名、2001年3月に発売開始いたしました。日本規格協会からはJIS工業用

語大辞典第5版対応のCD-ROM with *SuperHT*[®]を受注し同年9月に発売開始いたしました。

訳語付与による翻訳支援に加え、最近 *SuperHT*[®] がマニュアルなどの日本語文書の用字表現の統一にも利用され始めました。例えば、「どうぞお座りください」「キップを下さい」のように「ください/下さい」の正しい使い分けが、辞書の作り方によってできるようになります。また、常用漢字 1945 字以外の文字を見つけることもできます。辞書の作り方次第で、部品番号と部品名の置換や付与、ある数字の組み合わせを乱数表の中から見つけることなどもできます。詳しくは <http://www.bow-wow.jp> をご覧ください。

以上、とりとめのない内容になってしまいましたが、用語集超活用ソフト *SuperHT*[®] に興味を少しでも持っていただければと思います。今後ともよろしく願い申し上げます。

ちなみに、アトリエ・ワンのワンは英語の 1 (One) ではなく、犬の鳴き声の「わん」で、英語名称は「Atelier Bow-Wow」です。愛犬「うめ」(1988-1999) にちなんで命名しました。

機械翻訳および翻訳業界に期待すること

機械翻訳の最大の問題点は、訳文を読んでみないと正確さがわからないという点です。日英/英日のように英語の知識が多少でもある場合は、自分で原文と訳文を見て評価ができますが、まったく知らない言語と日本語の場合は訳文が正しいのか判断がつきません。

「マーズ・アタック」という火星人が地球に攻めてくる SF コメディ映画があります。一人の科学者が宇宙の全言語に対応できる機械翻訳装置を作り、英語と火星語の翻訳をするところがあります。円盤で飛来し、地球に降り立った火星へからの人間からの最初のことば「Welcome」を火星語に翻訳すると火星人が突然激怒し、その場にいた大勢の人間を皆殺しにしたのをきっかけに、人間と火星人の間で戦争が始まるというものです。結局は人間が勝つのですが、翻訳結果が人間には分からないので、なぜ火星人が怒ったのかが分からず、人間も火星人のことを無礼な連中だということになったのです。本来の機械翻訳はまったく相手の言語が分からない状態で安全に使えなくてはならないはずで、現状の日英/英日機械翻訳を使って、合っているとかいなとか言っているようでは機械翻訳の先はまだまだ遠いといわざるを得ません。

機械翻訳に逆翻訳の仕組みを取り入れて、原文と逆翻訳の結果を同一言語で比較採点し、翻訳文の精度を推測できるようにしたらよいと思います。

会員名・所属

坂本 義行 / SAKAMOTO Yoshiyuki

E-mail

自己紹介

1. 現在無職

2. これまでに従事した機械翻訳に関する研究

1) 機械翻訳システム「やまと」の開発 (1965~1970)

通商産業省工業技術院電気試験所 電子計算機部 情報基礎研究室
商用計算機 NEAC2200、COBOL プログラムで JCM 論文の英和翻訳、
漢字テレタイプタイターで出力、辞書 8,000 語

2) 科学技術庁振興調整費のプロジェクト「日英科学技術文献速報システムの開発」(Mu システム) (1982~1986)

通商産業省工業技術院電子技術総合研究所ソフトウェア部 言語処理研究室
科学技術振興調整費
日英相互の機械翻訳システムの開発に従事
京都大学工学部 長尾研究室、JICST、RIPS と共同研究を行った
日本語 JICST 文献抄録、英語 INSPEC 抄録
日英・英日翻訳、トランスファ方式、UTI-LISP、辞書 10 万語

3) 「近隣諸国間の機械翻訳システムの開発」(CICC プロジェクト) (1987~1995)

通商産業省工業技術院電子技術総合研究所知能情報部 自然言語研究室
アジア諸国間の翻訳システムの開発に参加
インドネシア、マレーシア、タイ、中国と日本語相互間の翻訳
中間言語方式、UNIX マシン、C プログラム
辞書 基本語 5 万語、 専門語 2.5 万語
特に、日本語の意味特徴の研究、中間言語方式による翻訳結果の評価を中心に研究

3. AAMT での活動

- 1) 市場動向調査委員会委員長 (1998 年~2006 年)
- 2) 機械翻訳白書「一機械翻訳—21 世紀のビジョン」編集委員 (1998~2000)
- 3) 機械翻訳課題調査委員会委員 (2006~)

機械翻訳および翻訳業界に期待すること

1. 翻訳を特定の専門分野と特定の文書に限定した実用システムの開発
2. 各システムに適合した入力文の仕様書あるいはガイドライン
3. 私用辞書 (専門辞書) の共有化 (AAMT で始めているが)

会員名・所属

田中 穂積 (中京大学・教授) / TANAKA Hozumi

自己紹介・機械翻訳および翻訳業界に期待すること

ルールベースとコーパスベース

この十年間、言語処理学の全国大会における発表会ではコーパスベースの研究発表が席卷し規則ベースの研究が肩身の狭い思いをしている、というより殆ど皆無であるといってよい状況にある。自然言語処理の初期の段階では言語学者もこの分野に興味を持つものも多く自然言語処理の研究者と言語学者との協調がそれなりにうまくいっていたケースも多々見られたようである。コーパスベースの研究成果がルールベースのそれを上

回り、文の精密な解析がルールベースだけでは無理であると思われた時期もあった。

これには二つの理由が考えられる。第一に、処理すべき文の量が大量になるにつれて、多種多様な文を規則で処理する文法の開発が難しくなってきたこと。どうしても規則に漏れが避けられず、言い換えると網羅性の問題がでてきたことが挙げられる。限られた人手で大規模な文法を開発するには限界が見えてきたのである。第二に、言語学者の興味は、文法によって一般性を失わずに言語現象を深く捉えることに在る。網羅性の問題ではない。

言語の文法性、規則性が統計的なデータとして反映されることは確かであるが、しかしどの統計データがどの言語現象を説明しているかを特定することが困難である。言語学者の言語学的な分析はここで行き詰る。第3に文法作成にしても計算機上で動く（動作する）文法でなければならないが、その意味を言語学者に理解させるのが難しい。文法が巨大になればなるほど解析結果に多数の曖昧性が含まれること、その数を減らすことが自然言語処理の高速化に直結することを理解してもらうことが必要である。しかしこれらは言語学者には言語学とは直接関係ないこととして関心がない。これは自然言語処理の立場からは極めて重要であるにも関わらず、要するに興味の対象が異なるのである。

以上の理由から、言語学者が作った大規模な文法を開発されたという話は聞かない。文法を開発するどころか文法学者が自然言語処理の分野から離れつつあると言う話はよく聞く。自然言語処理の研究者が日本語処理のために自分で日本語文法や英語文法などを作っているのが現状である。機械翻訳システムなどではそのような文法が開発されているのだろうが、公開されていないのが殆どなので比べて見ることができない。システムごとに個別の文法が作られているのだろうが、ずいぶん無駄なことをしているような気がする。

ルールベースの研究がもたもたしている隙に、コーパススペースの研究では着々と成果を挙げている。ますます研究者の目がコーパススペースの研究に向けられてきている。筆者はルールベースの研究の良さを皆に認めてもらうためには、大規模なCFG規則をまずはじめに作ってみたいと思っている。すでに中京大で研究をはじめているが、これは依存構造付きのコーパスからCFG規則を機械的に抽出するのである。依存構造付きのコーパスとして黒橋禎男らの開発した京大コーパスを用いる。その方法についてはまだ実験中であるが句構造付きのコーパスではないのでそれからどのようにしてCFGを(機械的に)抽出したらよいかの問題である。そこにモンターギュー文法の考え方が使えるのでないかと考えている。いずれ機会を見て発表したいと考えている。コーパススペースの考え方を利用したルールベースの自然言語処理を行おうとする試みである。

会員名・所属

長尾 真 (国立国会図書館・館長) / NAGAO, Makoto

E-mail

自己紹介

機械翻訳は研究として面白く、人間頭脳の働きをモデル理論的立場から解明することができるし、実用システムとしてビジネスにつながって行くだろうといったことを越えて、世界中の人達が自由にコミュニケーションすることによって世界に平和をもたらす最強の道具であるという使命感をもって取り組んで来た。

これから15~20年もすればかなり良いシステムが出来ることは間違いないので、これを世界中が使うことによって少しでも平和な世界が実現されることを望んでいる。

国立国会図書館長としては、世界中の知的財産資源の共有、各国、各地域の固有の文化の認識と尊重が機械

翻訳を通じて実現されることを期待している。

機械翻訳および翻訳業界に期待すること

かな漢字変換の日本語入力方式が今日社会に広く受け入れられるようになるのに約 20 年かかったことを考えれば、機械翻訳システムが一般に使われるようになるためには 50～60 年の開発・改良が必要だろう。したがって、2020 年～2030 年には十分に使えるシステムが出来るものと考えている。息の長い努力が必要である。

AAMT への要望

AAMT はアジア・オセアニア圏の機械翻訳関係者の情報交換の場として作ったが、会費の問題、情報交換のための言語の問題（日本語だけでは日本だけになってしまう。）などで、限定された活動になっている。この問題を解決することがこれからの大きな課題である。

アジア・オセアニア各国で機械翻訳への関心は高まっているので、お互いに知恵を出し合って AAMT の活動を高めてゆくことを期待したい。

会員名・所属

松尾 昭 (エム ティ ラボ株式会社) / MATSUO, Akira

E-mail

自己紹介

(※'05 年に AAMT ジャーナルに投稿した記事の中の一部とほぼ同じです)

日⇄英翻訳ソフト『訳せ!!ゴマ』の開発会社であり、翻訳サービスを行うエム ティラボ(株) 代表

私が機械翻訳ソフト(MT)の開発の世界に足を踏み入れてから 20 年以上になります。23 年前の 1984 年、当時工学系の大学院生だった私は、優秀な研究者たちが、その貴重な研究時間の多くを外国語論文の翻訳に割いているという現実を何とかしたいと感じていました。ある朝、何気なく手に取った朝日新聞の 1 面トップ記事「ベンチャー会社が、世界で初めて日英翻訳ソフトの製品化に成功」に目が釘づけになりました。これからは、高まる国際化の波と言語の壁を、コンピュータという道具により乗り越える時代が来ることを確信し、その会社、ブラビスインターナショナル(株)に就職しました。

数年後にブラビス社がバブル崩壊と共に倒産した後、その資産を数人の社員で出資して買い取り、MT 専門の開発会社、その名もエム ティ ラボ(株)を設立したのが 15 年前。その後、MT 製品『訳せ!!ゴマ』シリーズを旧エー・アイ・ソフト（現在はエプソン販売(株)に合併）と共同開発・販売させて頂き、現在に至っております。

ブラビスもエム ティ ラボも、スタッフに英語や言語のプロ(リンギスト)を SE・プログラマーより多く雇用し、MT の開発だけでなく人手による産業翻訳(HT)の業務もやり、そのノウハウやスキルを MT に活用し、「教科書的な文ではなく、活きた英語の言い回しを扱える」MT の開発を目標にして来たという特色があります。

エム ティ ラボは、現在、翻訳サービスを業務の主体としています。

AAMT は、前身である JAMT の創立以来の個人会員です。

機械翻訳および翻訳業界に期待すること

HT への MT の活用は、時代の趨勢であり、欧米言語間などで徐々に進んでいます。弊社の現状では、一部、自社の MT 用専門辞書データを利用して、社内ツールによる専門用語の自動的な辞書引きと、それによる用語の統一を行っている程度ですが、MT・HT 両方の現場に長年携わって来た者として、ぜひ行いたいと思って

います。

MTと既存TM製品との融合や、ローカライズのような特定分野へのMTのカスタマイズなどのご提案もしております。

弊社のMT資産に関心のある方や、HTへのMTの活用に必要な予算や時間にお付き合い頂ける翻訳会社様など、ビジネスパートナーとして関心がある方がいらっしゃいましたら、ぜひ御連絡ください。

AAMTへの要望

MTサミットやその他の会議を、観光地でのお祭りのなもの・研究者の発表の場が主な会でなく、より実用的なものとし、安価に利用できるようにして頂ければと思います。翻訳現場の団体のJTF(日本翻訳連盟)など、他の、より実用的な団体と共同で、何かできないでしょうか？

第 17 回通常総会および関連行事の報告

AAMT 事務局

当協会の第 17 回通常総会が 2007 年 6 月 11 日（月）13 時 30 分より東京・御茶ノ水の総評会館で開催されました。総会后、各委員会からの報告会、2 名の講師による講演会、そして第 2 回 AAMT 長尾賞授与式と受賞者による記念講演会が盛況のうちに行われました。

第 16 回通常総会

1. 開会の辞
2. 会長挨拶（独）情報通信研究機構 自然言語グループリーダー 井佐原 均
3. 来賓挨拶
4. 出席会員の確認
5. 議事録署名人の選出
6. 議案
第 1 号議案 2006 年度事業報告（案） 第 2 号議案 2006 年度決算報告（案）
第 3 号議案 2007 年度事業計画（案） 第 4 号議案 2007 年度収支予算（案）
第 5 号議案 会則改定について その他
※すべての議案は、
7. 閉会の辞

報告会

- | | |
|-----------------------|------------------------|
| 1. 機械翻訳課題調査委員会 | 委員長 内元 清貴（（独）情報通信研究機構） |
| 2. AAMT/Japio 特許翻訳研究会 | 副委員長 中川 裕志（東京大学） |
| 3. インターネット WG | リーダー 富士 秀（（株）富士通研究所） |
| 4. 編集委員会 | 委員長 宇津呂 武仁（筑波大学） |

講演会

講演 I：「JAL の安全対策とテキスト処理」

寺田昭 氏（（株）日本航空インターナショナル）

講演 II：「KUI to assist online multilingual text development」

（講演予定者のソムラートラムワーニッチ・ウィラット氏が欠席のため、共同研究者である佐原 AAMT 会長が代理で講演）

※ 上記 2 講演の関連論文が本誌に掲載されております。

第2回 AAMT 長尾賞授与式・記念講演会

受賞者：(株)国際電気通信基礎技術研究所 (ATR)

音声言語コミュニケーション研究所

受賞理由：統計と用例とに基づくコーパスベース翻訳技術の研究・開発ならびに
携帯電話を用いた多言語音声翻訳サービスの事業化

選考委員長：飯田仁（東京工科大学）

選考委員：石崎 俊（慶應義塾大学）

Kei-Sun Choi（KAIST）

中川 裕志（東京大学）



▲スピーチする長尾真国立国会図書館館長



▲受賞記念講演を行う ATR 隅田英一郎氏

懇親会

本会後の懇親会は副会長の飯田仁先生（東京工科大学）による乾杯の音頭で開会し、多数の参加者が楽しいひとときを過ごしました。会の最後では、AAMT 前会長であり現理事の辻井潤一先生（東京大学）より今後の機械翻訳分野における展望とそれに伴って必要とされる AAMT のますますの活性化についてお言葉を頂戴しました。

「英日機械翻訳に関するアンケート」結果報告

AAMT 機械翻訳課題調査委員会

1. はじめに

AAMT 機械翻訳課題調査委員会 WG2(調査・広報・啓蒙)では、その前身である市場動向調査委員会の時代から調査活動の一環として、年一回インターネット上で一般の機械翻訳ユーザを対象としたアンケートを行っている。AAMT の Web サイトに特設ページを設け、機械翻訳に関するいくつかの設問に回答してもらうという形式である。回答者の中から抽選で 25 名程度の方に商品(市価 1 万円程度の翻訳ソフトパッケージ)を進呈している。

毎年の実施時期(1 月下旬から 2 月上旬)や実施期間(12 日間)は一定であり、告知・宣伝の方法も毎年ほぼ同じ、設問内容も毎年少しずつ改良してはいるが、大きく変えることはしていないので、市場の経年変化を定点観測的に知ることができると考えている¹。

以下では 2007 年の調査結果について報告し、2003 年からのデータも適宜使って分析を試みる。なお、各年のアンケート回答総数は、2003 年 1,413、2004 年 1,720、2005 年 1,018、2006 年 1,061、2007 年 835 である。

2. 設問項目

以下に示す項目について質問を行った。

- ・回答者の基本属性について(メールアドレス、年齢層、性別、職業、英語力、アンケートに回答するのは何回目か)
- ・Web 上の機械翻訳サービスについて(利用したことがあるか、利用の頻度・目的・用途、使う(使わない)理由)
- ・MT ソフトに関して(利用したことがあるか、利用の頻度・目的・用途、使う(使わない)理由)
- ・機械翻訳全般に関して(品質や使い勝手が良くなっているか、今後に期待するか、メーカーに依存しない共通辞書を利用したいか)

3. 全体に関する調査結果

年齢(図1)に関しては、2006 年に較べて 2007 年に目立った傾向は見られないが、全体を通しては 20 代、30 代が減少し、50 代が増加するという高年齢層へシフトする傾向が見て取れる。性別(図2)に関しては、調査開始以来一貫して女性の比率が増加している。これらの結果は、インターネットユーザ一般に対して我々が持つ常識的な推測と一致しており、特に機械翻訳に特徴的なことではないと考えられる。

¹ このことは毎年回答者数に大きな変動のないこと、年による極端なばらつきが見られず、一定の傾向を示すケースが多いことによっても裏づけられる。

図1:年齢 (%)

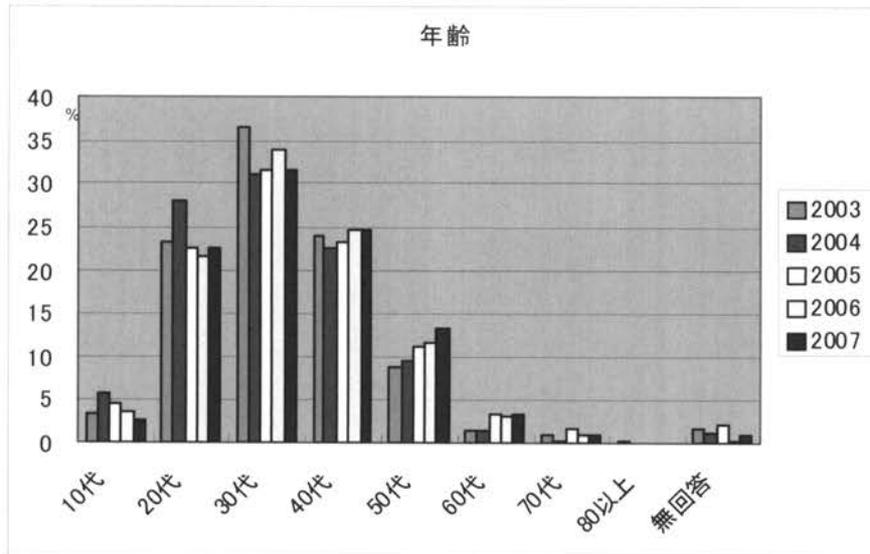
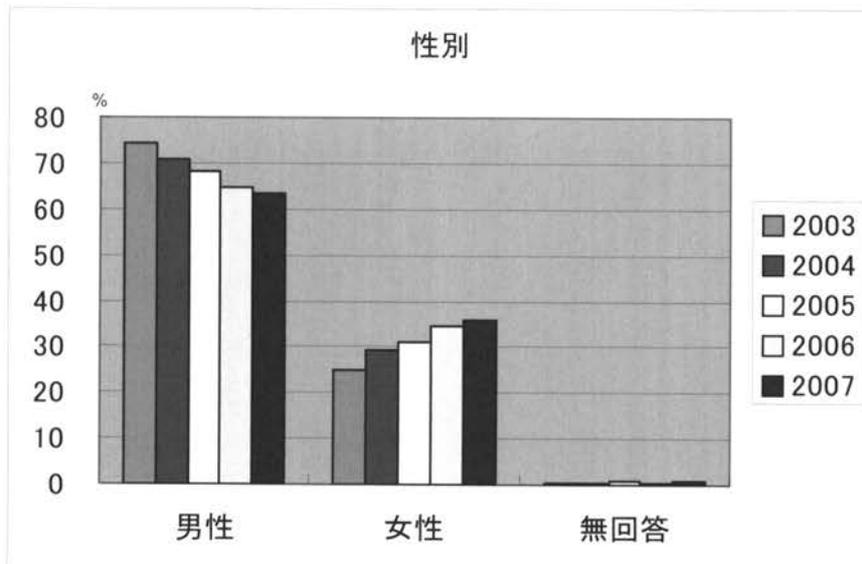


図2:性別 (%)



職業(図3)について見ると、「翻訳業」が年々増加しているのが目を引く。次の質問は、回答者の英語能力の水準を、英検/TOEICスコアに換算する形で自己評価してもらったものだが(図4)、スコアの高い人が増加傾向である。前項と併せて、英語力の高い人の機械翻訳への関心が増えつつある、と見ることができる。

図3: 職業(%)

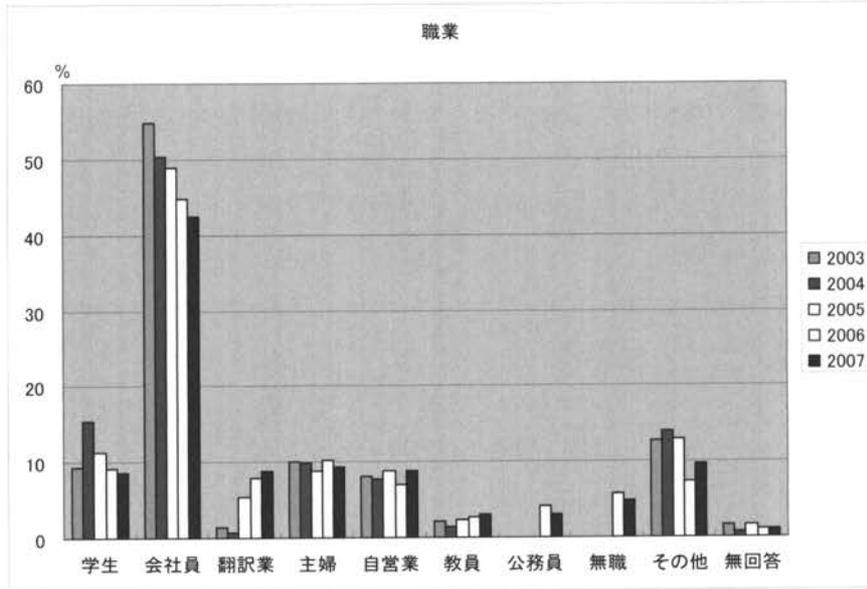
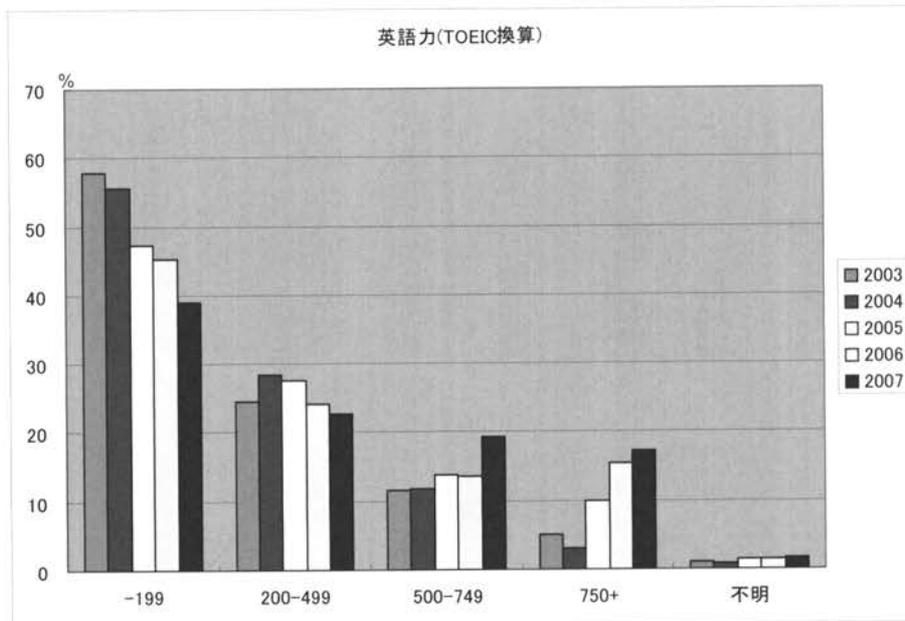


図4: 英語力 (%)



4. Web翻訳に関する調査結果

Web翻訳の使用経験として、「知らない」、「使ったことがない」は減少傾向にある(図5)。「Web翻訳を利用しない

理由」を問う設問(図 7)で、「翻訳サービスのサイトを知らない」や「サービスの使い方が分からない」が減少していることと併せ、Web 翻訳サービスの認知度が増していることがわかる。ただし、表 1 に示す通り、利用頻度を問う設問で「試しに数回」(この選択肢は 2006 年より追加した)が多く、日常的に使っている人は少ない。利用しない理由として増加しているのは、「翻訳精度が低い」である。前項の結果に見られるように、翻訳者等の英語力の高い回答者が増えたことが要因の一つと考えられる。

どうい分野の文書を翻訳対象としているかを問う設問では、昨年の調査から「特許」「契約書」という選択肢を増やしたが、これらは 1%前後といったところで、あまり多くなかった。今年の調査では「医療」を追加したが、これは 5.15%あった。多数を占めたのは、「産業」(18.08%)、「文化・芸術」(8.62%)、「自然科学」(8.38%)、「生活情報」(7.90%)といったところであった²。

Web 翻訳を利用する理由(図 6)に関しては、「概要が知りたい」や「下書き作成」といった回答も一定数見られるが、やはり「無料」という要因が大きいようである。

図5: Web翻訳の利用有無 (%)

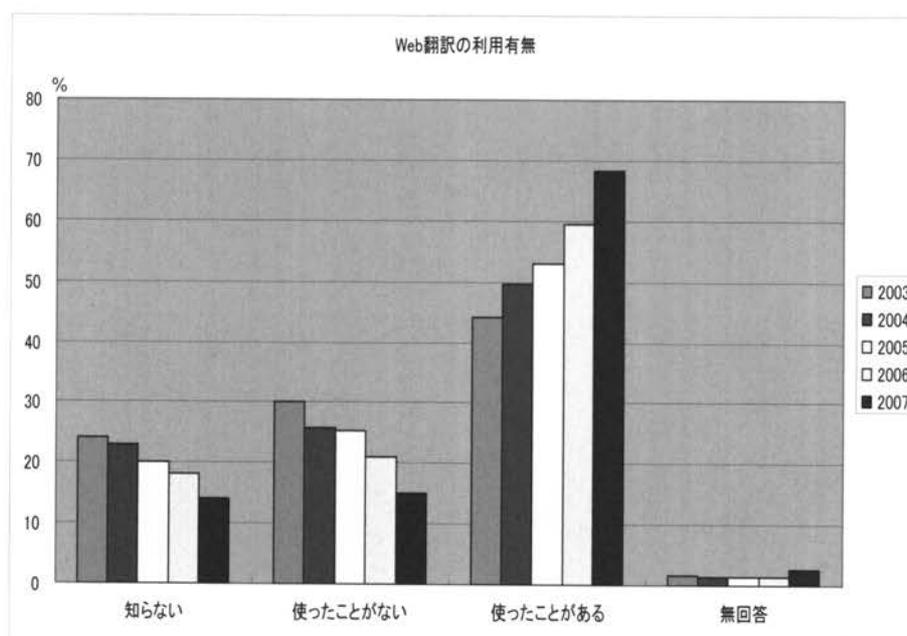


表1: Web翻訳の利用頻度 (%)

	ほぼ毎日	週 1-3 回	月 1-3 回	年 1-3 回	試しに数回	無回答
2006 年	3.20	13.20	19.98	8.48	22.71	32.42
2007 年	5.15	16.41	22.28	9.94	22.75	23.47

² この設問は、年によって分野の分け方を若干変えているため、経年変化を見るのが難しい。

図6: Web翻訳を使う理由 (%)

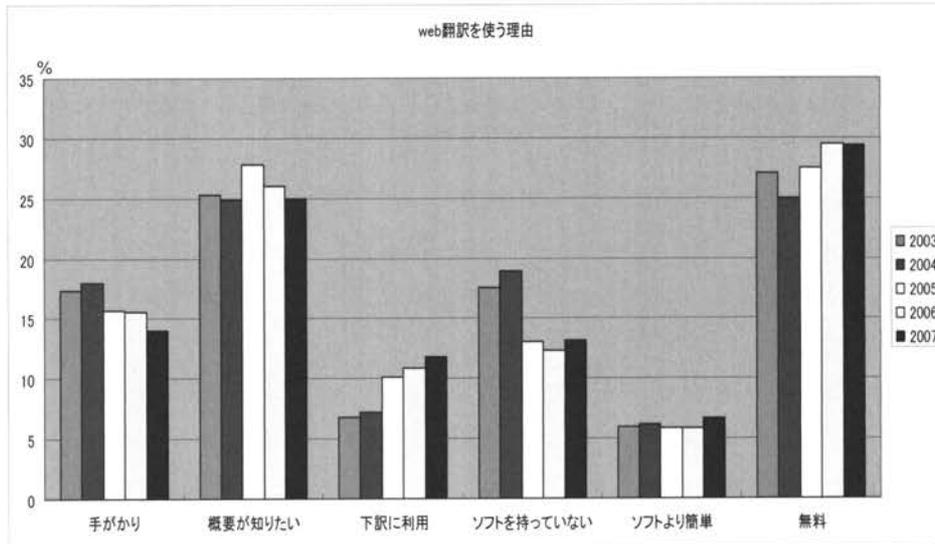
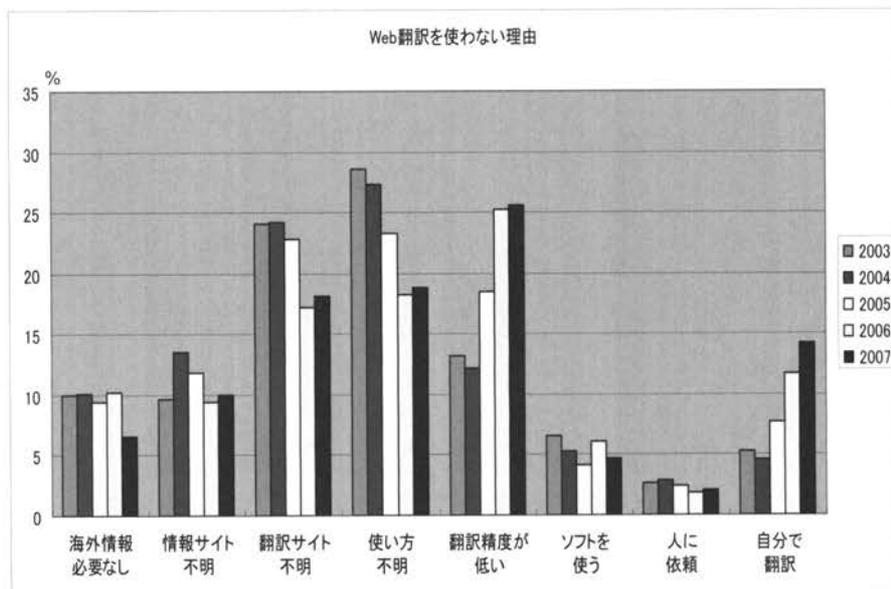


図 7: Web 翻訳を使わない理由 (%)



5. MTソフトに関する調査結果

MTソフト(バンドルとパッケージソフト)に関する調査を始めたのは2005年からなので、データは3年間のものである。

利用頻度や翻訳対象分野については、Web翻訳と比べて目立った傾向の違いは出て来なかった。利用目的としては仕事で使うという人の割合がWeb翻訳より若干多いとか、翻訳対象としてWeb(URL指定)が少なく一般のドキュメントが多いといった違いが見られた。いずれもWebとパッケージの特性の違いを考慮すれば、うなずける結果である。

バンドルソフトを使う理由(図8)はWeb翻訳と比べて特に目立った特徴は見られない。(Webに頼らず、敢えて)パッケージ製品を購入した理由(図9)として、「ドキュメントの機密保持」などもあり得るかと考えたが、あまり多くはなかった。ただし、それなりの機能・品質であれば、たとえ高価であってもパッケージを購入するという層は少数ながら確実に存在するようである。

MTソフトを利用しない理由(図10)で「Web翻訳ソフトで十分」が増加に転じ、トップになったのは注目される。翻訳ソフトパッケージの店頭販売の不振が伝えられ、Web上の翻訳サービスに食われているのではないかとする向きがあるが、それが裏づけられた格好である。

図 8: バンドルソフトを使う理由 (%)

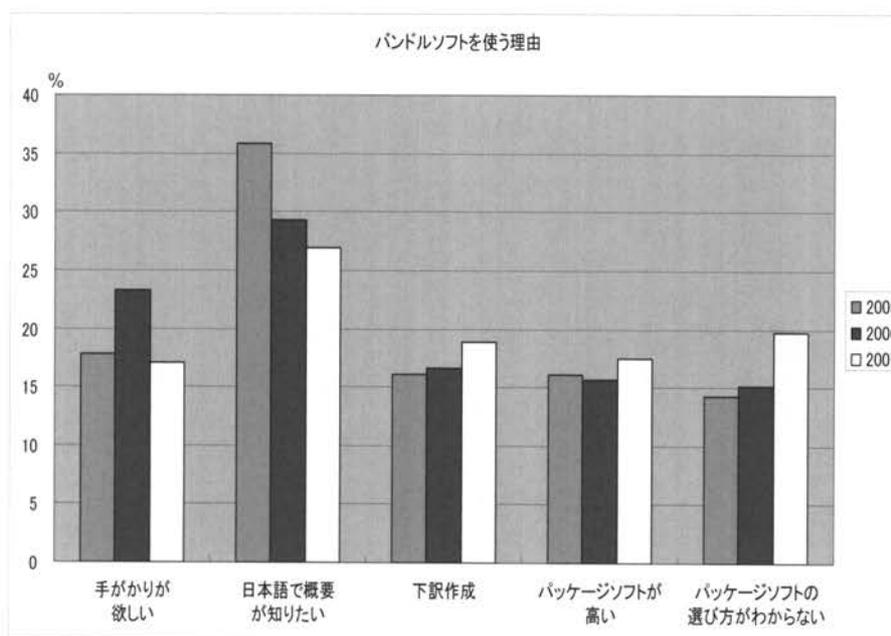


図 9: パッケージソフトを購入した理由 (%)

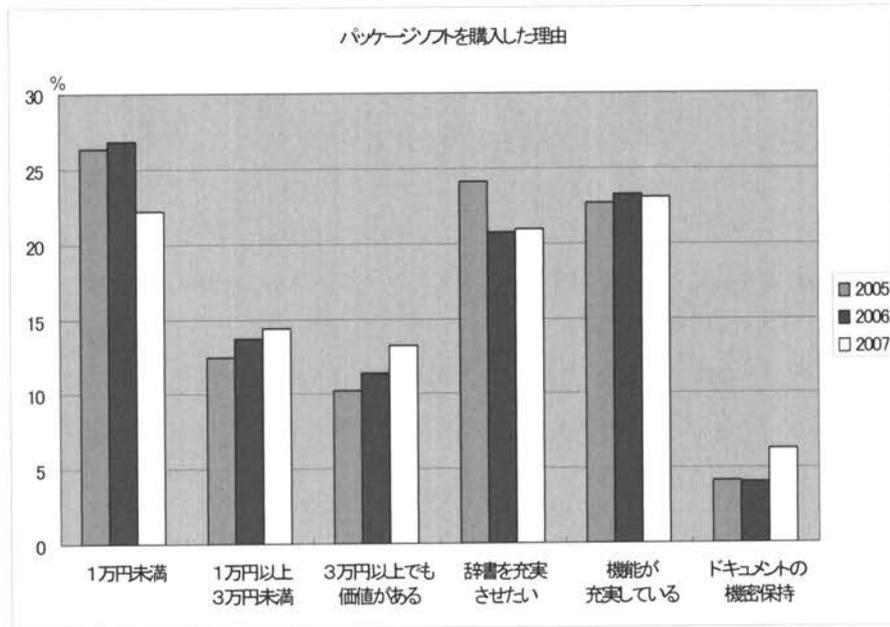
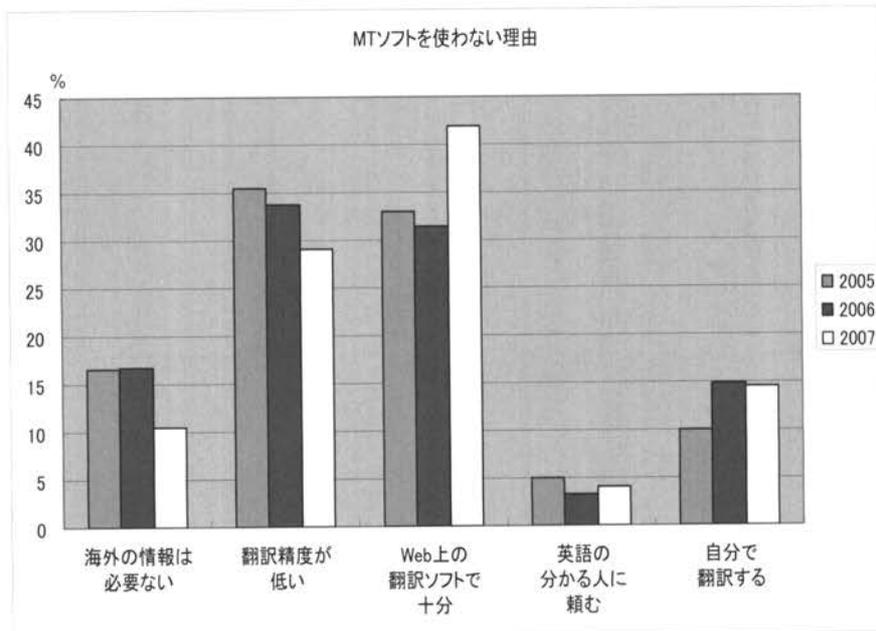


図 10: MTソフトを使わない理由 (%)



6. 機械翻訳全般に関する調査結果

本項は今年から新たに追加された項目である。まず、最近の機械翻訳について、品質と使い勝手がよくなっているかどうかについて、それぞれ「全くその通り」から「全くそう思わない」まで5段階で評価してもらった。結果はどちらの項目も肯定的な評価が過半数を越え、今後の機械翻訳について「期待する」と答えた人も70%弱に達した。

また、「期待する」と回答した人には何に期待するか、「期待しない」と回答した人にはその理由は何かを、自由記述形式(これも今回初めて導入した)で書いてもらった。回答結果を適当なキーワードで大まかに分類したのが図11、図12であるが、期待するにしろしないにしろ、ユーザの最大関心事は翻訳の精度であるということが改めて示された格好になった。それ以外では専門用語辞書の充実を望むユーザが多かったのが注目される。

最後の設問として、現在AAMTで規格の策定を進めている「メーカーに依存しない共通辞書」へのニーズを調べる質問を行った。結果は70%を越えるユーザが「利用したい」と答え、期待の大きさを窺わせた。

図 1 1： 期待する理由 (%)

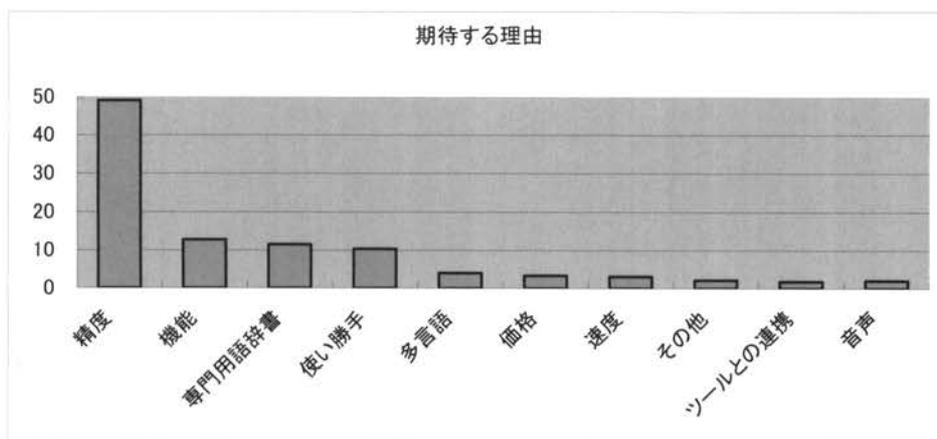
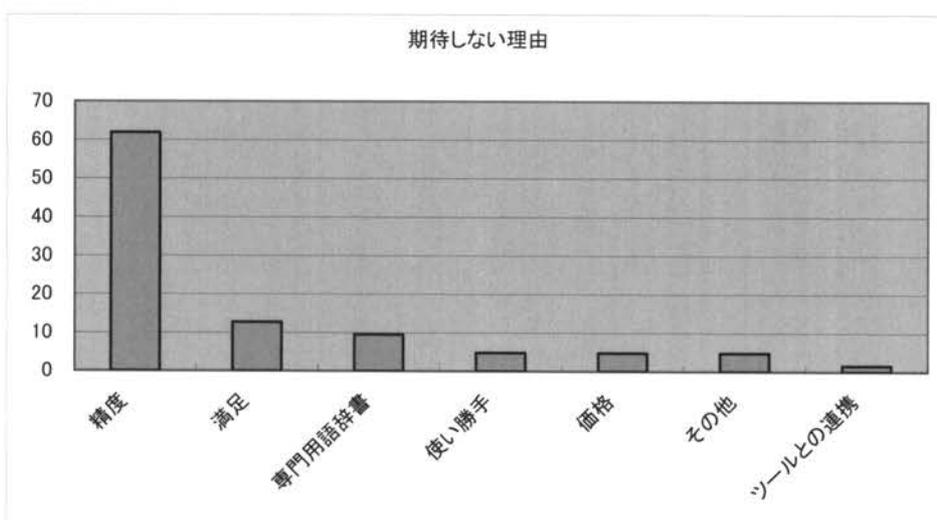


図 1 2： 期待しない理由 (%)



「AAMT Journal に関するアンケート」集計結果

AAMT 編集委員会

AAMT Journal を皆さまにとってより有益な情報提供媒体とするべく、2007 年 6 月の通常総会時に「AAMT Journal に関するアンケート」を実施いたしました。その結果、13 名の皆さまより貴重なご意見を頂戴いたしましたので以下にご紹介させていただきます。ご回答いただきました皆さまに感謝申し上げます。頂戴いたしましたご意見に関しましては、今後の編集方針を決める際の重要な項目として、委員会にて検討を重ねております。

AAMT 編集委員会では、今後も皆さまからのご意見をお待ちしております（連絡先：AAMT 事務局 aamt-info@aamt.info）。

Q1 あなたご自身について

Q1-1 あなたは AAMT 会員ですか？

法人会員のメンバー	4
個人会員	8
会員ではない	1

Q1-2 あなたは、機械翻訳についてどういう立場でしょうか？（複数回答可）

大学など、非営利で研究する立場	4
メーカーなど営利団体で研究・開発する立場	4
企画・営業販売	2
利用する立場	6
雑誌社など広める立場	0
その他	1

Q2 AAMT Journal について

Q2-1 AAMT Journal を読まれていますか？

毎号必ず読んでいる	6
興味がある記事だけ読んでいる	6
届いているが読んでいない	0
届いていない	1

Q2-2 AAMT Journal の企画の中で、良かったと思う企画はどれですか？その理由もお書きください。（複数回答可）

a	巻頭言	2
b	海外報告	6
c	シンポジウム・学会等イベント報告	6
d	新製品紹介	5
e	MT サービス一覧	6
f	研究・プロジェクト・その他調査報告	8
g	MT の歴史	2
h	事務局活動報告	3
i	特に良いと思う企画はない	0

理由：

- ・ MT 技術現状の理解に役立っています。（選択項目：b,c,d,e）
- ・ 他誌にはない情報であるため。（選択項目：b,c）

Q2-3 AAMT Journal の企画の中で、良くなかったと思う企画はどれですか？その理由もお書きください。（複数回答可）

a	巻頭言	0
b	海外報告	0
c	シンポジウム・学会等イベント報告	0
d	新製品紹介	1
e	MT サービス一覧	0
f	研究・プロジェクト・その他調査報告	0
g	MT の歴史	0
h	事務局活動報告	0
i	特に良くないと思う企画はない	12

理由：

- ・ 情報自体は有益ですが、広告的になりがち。評価（課題や問題点も）を付けて（または後追いで可）掲載していただければ価値があると思います。（選択項目：d）

Q2-4 現在（2007年6月）の一年あたりの発行号数（年2号程度）について、どのように思われますか？

ちょうどよい	5
少ない	6
多い	0
無回答	2

◆「少ない」と答えられた方は、一年につき何号程度が適当と思われますか？

3号	2
4号	3
5号	0
6号	1

Q2-5 現在（2007年6月）「会員特集号」を企画中です。会員の皆さまから自己紹介、機械翻訳業界への期待、AAMTへの要望などをご寄稿いただき、会員間での交流を促進しようという趣旨です。このような企画にご関心がありますか？（注：これはアンケートを実施した2007年6月時点での状況です。その後、企画の形式を変更し、通常号に会員紹介企画を含めることになりました。）

大変関心があり、寄稿したい（もしくはすでに寄稿した）	5
まあまあ関心があり、もう少し具体的な趣旨や細かい形式が決まれば寄稿したい（もしくはすでに寄稿した）	2
あまり関心がない	3
無回答	3

◆「大変関心がある」もしくは「まあまあ関心がある」を選択された方の中で、会員特集号に向けた提案があれば、ぜひお書きください。

- ・ MT研究者と開発者の第一人者とされる方々の座談会。
- ・ 顔写真も入れたらよいと思います。
- ・ 翻訳業の人々のMTに対する声を聞きたい。会員だけでなく、広く翻訳業の人々の考えに興味がある。

◆「あまり関心がない」を選択された方は、その理由をお書きください。

- ・ ジャーナルに合っていない？
- ・ 「会員特集号」には大変関心がありますが、寄稿云々とは異なるように思います。具体的になった時点でお返事をしたい。

Q2-6 今後、新たにAAMT Journalに含めてほしい企画はありますか？ある場合はその内容もお書きください。

ある	8
ない	1
無回答	4

◆ 今後含めてほしい内容：

- ・ 翻訳者からの意見。
- ・ MT の訓練コースについて。
- ・ MT 向け文章のガイドラインについて。
- ・ MT システムのユーザコメント（固有名詞は伏せて）。
- ・ 翻訳製品の評価（質的に）。
- ・ 新しい翻訳評価法の紹介。
- ・ ユーザの利用経験談。
- ・ いろいろな解説記事を載せていただけるとありがたい。
- ・ 逆翻訳でのチェックについて（知らない外国語が絡む翻訳が正しく行われたかのチェック）。
- ・ 原文例から機械翻訳を行い、最終訳文に仕上げるプロセスを載せてほしい。
- ・ 中国、韓国、その他の国の人々の声。

Q2-7 その他、AAMT Journal に対してご意見があればお書きください。

- ・ 毎回でなくてよいが、英語版を作成してほしい。
- ・ AAMT 会員専用ホームページでバックナンバーを閲覧できるようになり、便利になりました。
- ・ 英文ジャーナル等の英文記事は MT を利用して和文紹介をしていただければと思います。

協会活動報告

(2007年5月～2007年11月)

第17回通常総会

2007年6月11日

- | | | | |
|--------|---------------|--------|---------------|
| ①第1号議案 | 2006年度事業報告(案) | ②第2号議案 | 2006年度決算報告(案) |
| ③第3号議案 | 2007年度事業計画(案) | ④第4号議案 | 2007年度収支予算(案) |
| ⑤第5号議案 | 会則改定について | ⑥その他 | |

報告会

2007年6月11日

- | | |
|--------------|---------------------|
| ①機械翻訳課題調査委員会 | ②AAMT-Japio 特許翻訳研究会 |
| ③編集委員会 | ④インターネットWG |

講演会

2007年6月11日

講演 I : 「JAL の安全対策とテキスト処理」

寺田昭 氏 (株)日本航空インターナショナル)

講演 II : “KUI to assist online multilingual text development”

(講演予定者のソンラートラムワーニッチ・ウィラット氏が欠席のため、
共同研究者である井佐原 AAMT 会長が代理で講演)

第2回 AAMT 長尾賞授与式・記念講演会

2007年6月11日

受賞者：(株)国際電気通信基礎技術研究所 音声言語コミュニケーション研究所

受賞理由：統計と用例とに基づくコーパスベース翻訳技術の研究・開発ならびに
携帯電話を用いた多言語音声翻訳サービスの事業化

懇親会

2007年6月11日 山の上ホテル レストラン・ラヴィ

決算理事会

2007年6月11日

- | | | | |
|--------|---------------|--------|---------------|
| ①第1号議案 | 2006年度事業報告(案) | ②第2号議案 | 2006年度決算報告(案) |
| ③第3号議案 | 2007年度事業計画(案) | ④第4号議案 | 2007年度収支予算(案) |
| ⑤第5号議案 | 会則改定について | ⑥その他 | |

機械翻訳課題調査委員会

2007年5月18日 (WG②「調査・広報・啓蒙」個別委員会) 総会での報告について

2007年6月1日 (第2回)

- | | |
|--------------------------|-------------|
| ①前回委員会の議事録の確認 | ②総会での報告について |
| ③各WGの活動について (各WGに分かれて議論) | |

- ④活動内容の報告（各 WG から）
- ⑤活動内容についての議論
- ⑥まとめと次回委員会について

2007年7月9日（第3回）

- ①前回委員会の議事録の確認
- ②総会での報告について
- ③各 WG の活動について（各 WG に分かれて議論）
- ④活動内容の報告（各 WG から）
- ⑤活動内容についての議論
- ⑥まとめと次回委員会について

2007年8月7日 無料翻訳サイトへのヒアリング

2007年8月10日（WG③個別委員会）

- ①UPF3の正式名称について
- ②UPF3 パートナー募集について
- ③辞書作成・利用ツールについて
- ④UPF3 メーリングリストについて
- ⑤Doc Fest M24 について
- ⑥オープンソース系への打診の担当について
- ⑦WG③のメンバーについて

2007年8月20日（第4回）

- ①大阪大学 林良彦先生によるご講演
- ②各 WG の活動について（各 WG に分かれて議論）
- ③活動内容の報告（各 WG から）
- ④活動内容についての議論
- ⑤まとめと次回委員会について

2007年9月15日 WG③が Mozilla イベント Doc Fest M24 に参加

2007年10月2日（第5回）

- ①WG③Mozilla Doc Fest M24 参加報告
- ②各 WG の活動について（各 WG に分かれて議論）
- ③活動内容の報告（各 WG から）
- ④活動内容についての議論
- ⑤まとめと次回委員会について

2007年11月5日（第6回）

- ①前回委員会の議事録の確認
- ②総会での報告について
- ③各 WG の活動について（各 WG に分かれて議論）
- ④活動内容の報告（各 WG から）
- ⑤活動内容についての議論
- ⑥まとめと次回委員会について

編集委員会

2007年7月24日 AAMT Journal No.41 企画について

2007年10月23日 AAMT Journal No.42 企画について

インターネット・ワーキンググループ

- ①AAMT ホームページの更新（毎月）
- ② AAMT Forum での情報発信（月数回）
- ③委員会活動におけるメーリングリストの管理
- ④その他ネットワーク関連のインフラ管理全般

その他

2007年8月9日 Translators Association of China 代表団が AAMT 事務局を訪問

2007年10月19-20日 飯田仁副会長が AAMT 中国支部（仮称）設立に向けた協議のため
中国科学院（北京）を訪問

AAMT ジャーナル編集委員会委員長

筑波大学大学院システム情報工学研究科

知能機能システム専攻

宇津呂 武仁

AAMTジャーナル41号をお送りします。

今号では、まず、第二回のAAMT長尾賞の選考結果を受けまして、これまで長年に渡り音声翻訳研究の第一線で活躍してこられましたNICT/ATRの隅田様に、巻頭言をご執筆頂きました。

また、今号の大きな企画としまして、2007年9月にコペンハーゲン(デンマーク)・ショーヴデ(スウェーデン)におきまして開催されました第11回機械翻訳サミットおよび機械翻訳に関する理論的・方法論的課題に関する国際会議TMIへの参加報告を掲載しました。このサミットの併設ワークショップでは、AAMT/Japio特許翻訳研究会のメンバーが中心となって特許翻訳ワークショップを開催しましたので、このワークショップの報告も掲載しました。

今号に先立ちまして、2007年6月には定例の通常総会が開催されましたが、今号ではその報告をさせて頂くとともに、総会におきまして招待講演を頂きましたJALの寺田様からのご寄稿、および、総会にて行いました、AAMTジャーナルに関するアンケートの集計結果を掲載しました。また、機械翻訳課題調査委員会からの記事として、英日機械翻訳に関するアンケートの結果報告を掲載しております。

本ジャーナルは、AAMTの機関紙という位置付けでありますので、AAMT会員に向けたやや個別的な取り組みとして、今号より、「AAMT会員のひろば」と銘打ちました、会員自己紹介のページの企画を始めました。今号にはその第一弾を掲載しております。本企画に際しましては、今年度、会員の皆様に自己紹介ページ用の原稿を依頼させて頂きました。原稿の集まりのほどはまだ十分ではありませんので、今号の企画を参考にして頂いて、是非、更なる原稿のご寄稿のほどをお願い致します。

Memo

AAMT

AAMTジャーナル No.41

発行：アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

住所：〒619-0289 京都府相楽郡精華町光台3丁目5番地

独立行政法人 情報通信研究機構 自然言語グループ内

phone：0774-93-4625 fax：0774-93-4627

編集委員会：宇津呂 武仁 大倉 清司 熊野 明

松川 淑子 村上 嘉陽

事務局：神崎 享子 和泉 絵美

印刷所：株式会社ナビックス

Asia-Pacific Association for Machine Translation

c/o Computational Linguistics Group, NICT

3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan 619-0289

Phone:+81-774-93-4625 FAX:+81-774-93-4627

URL:<http://www.aamt.info>

