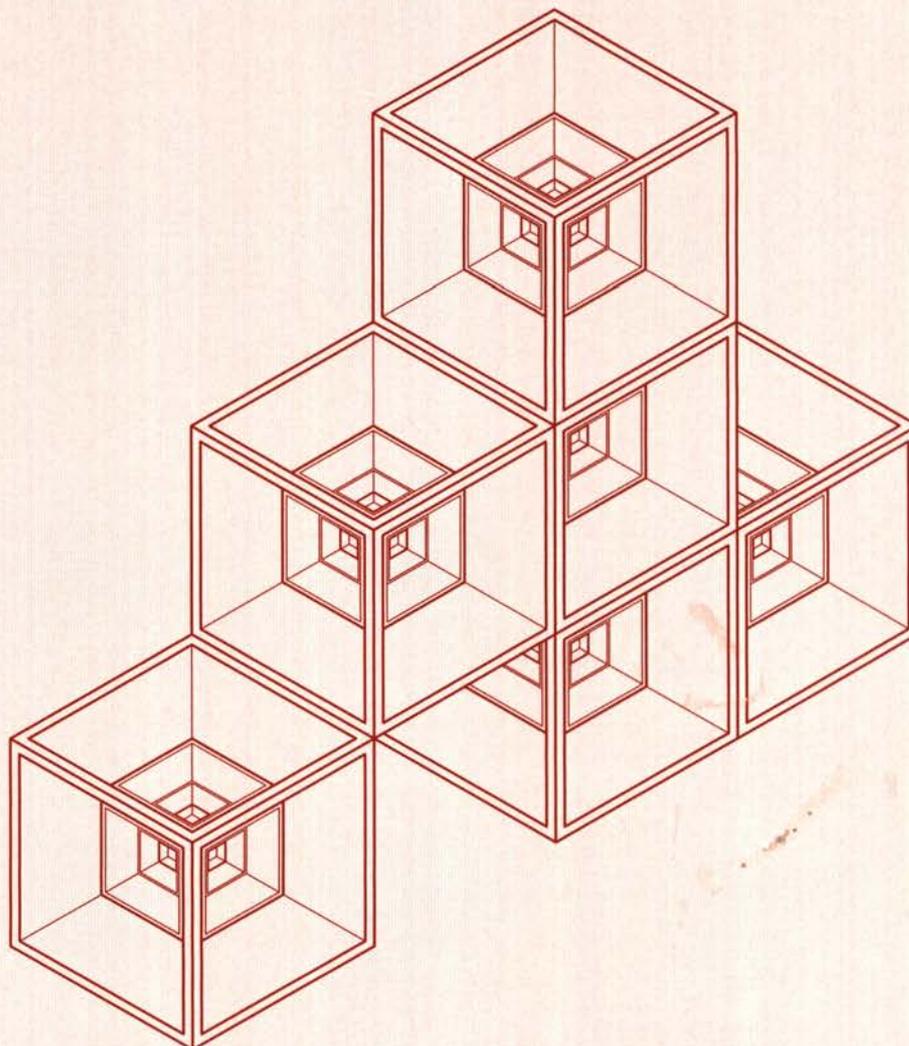


# AAMT

Asia-Pacific Association for Machine Translation

# Journal



December 2011 *No.50*

アジア太平洋機械翻訳協会

## 目 次

巻頭言：	機械翻訳の次の10年.....	出羽 哲也 ..... 1
シンポジウム報告：	MT Summit XIII 参加報告.....	二宮 崇、横山 晶一 ..... 2
	MT Summit XIII 参加報告.....	網川 隆司 ..... 4
	MT Summit XIII における「特許翻訳 WS」等に関する報告.....	森藤 淳志 ..... 6
	MT Summit XIII Technology Showcase 報告.....	熊野 明 ..... 8
	MT Summit XIII 参加報告.....	大倉 清司 ..... 10
	プロジェクト報告：	Machine Translation in China.....
レポート：	翻訳時の日本語表記に対する取り組み.....	田中 千鶴香 ..... 16
	Japan and the Global Language Services Market.....	D. A. DePalma ... 19
	ハイブリッドへの道：英⇄日ハイブリッド機械翻訳に対する Pangeanic と東芝の取り組み.....	..... Elia Yuste, Manuel Herranz & Alexandre Helle, Hirokazu Suzuki ..... 33
	多言語医療対話支援システムの開発と公開 ー用例対訳と機械翻訳の併用ー.....	..... 吉野 孝、福島 拓、尾崎 俊、東 拓央、宮部 真衣..... 40
	Web の多言語化と言語技術の標準化: W3C MultilingualWeb Limerick Workshop より.....	..... 林 良彦 ..... 45
	委員会活動報告：	AAMT Forum ホームページ刷新について.....
AAMT 会員のひろば：	AAMT 機械翻訳課題調査委員会 WG2.....	..... 52
	AAMT 機械翻訳課題調査委員会・第6回 AAMT 長尾賞を受賞して WG3.....	..... 共有化・標準化ワーキング グループ (UTX チーム)・山本 ゆうじ ..... 61
	予防と未然防止.....	島中 伸敏 ..... 63
事務局からのお知らせ：	協会活動報告 (2011 年 6 月～10 月)、総会案内.....	AAMT 事務局..... 70
編集後記	.....	..... 75

## C O N T E N T

Foreword:	"Next decade of machine translation".....	<i>T. Izuha</i> ..... 1
Symposium Report:	Report on MT Summit XIII.....	<i>T. Ninomiya, S. Yokoyama</i> ..... 2
	Report on MT Summit XIII.....	<i>T. Tsunakawa</i> ..... 4
	Report on MT Summit XIII --The report on the Patent Translation WS and the like.....	<i>A. Morifuji</i> ..... 6
	Report on MT Summit XIII.....	<i>A. Kumano</i> ..... 8
	Report on MT Summit XIII.....	<i>S. Ookura</i> ..... 10
	Project Report:	Machien Translation in China.....
Approach for Identifying Best Practices in Japanese Style Guide for Translation..		<i>Ch. Tanaka</i> ..... 16
Japan and the Global Language Services Market.....		<i>D. A. DePlama</i> ... 16
Going Hybrid: Pangeanic's and Toshiba's First Steps Toward ENJP MT Hybridization.....		..... <i>E. Yuste, M. Herranz &amp; A. Helle, H. Suzuki</i> ..... 33
Development and Release of Multilingual Medical Communication Support Systems		.....
Combination Effective Use of Parallels texts and Machine Translation.....		..... <i>T. Yoshino, T. Fukushima, Sh. Ozaki, T. Azuma, M. Miyabe</i> ..... 40
Report:	Multilingual Web and Standardization of Language Technologies.....	<i>Y. Hayashi</i> ..... 45
Committee Report:	AAMT Website Upgrade.....	<i>IWG</i> ..... 48
	Committee for seeking future direction of MT WG2.....	..... 52
	Committee for seeking future direction of MT WG3 Sixth AAMT Nagao Award Report ...	<i>Y. Yamamoto</i> ..... 61
AAMT Members:	How to Translate the Term of "Precaution" and "Preventive Action".....	<i>N. Hatanaka</i> ..... 63
	.....	..... <i>SunFlare Co., Ltd.</i> ..... 68
AAMT Activities:	AAMT Activities (from June to October, 2011), General Meeting.....	..... 70
Editor's Note:	.....	..... <i>T. Utsuro</i> ..... 75

## 機械翻訳の次の10年

株式会社東芝

研究開発センター 出羽 達也

過去10年、機械翻訳技術の研究開発は「ブーム」であったと言ってよいでしょう。統計翻訳という強力な推進力の存在は言うまでもなく、シーズ・ニーズ両面の様々な要因が「ブーム」を後押しし、停滞感すらあったその前の10年と比べると、研究者や論文の数、投入される研究開発資金の額は飛躍的に増えました。その結果、機械翻訳技術は長足の進歩を遂げ、実用に耐え得るシステムも公開されるようになりました。しかし、その膨大なニーズに比して現在の機械翻訳技術が成し得ることはまだまだ限られています。そこで、これからの10年もニーズに応え続けていくために何をすべきか、企業で技術開発に携わる立場から考えてみたいと思います。なお、本協会には翻訳を主たる業務とする会員の方が数多くいらっしゃると思いますが、本稿では業務や生活の手段として異言語障壁の克服を必要としているすべての人を機械翻訳技術の潜在ユーザに含めて考えています。

まず、特定のドメインで検証された技術を様々なドメインに低コストで展開する手法の確立が重要と考えています。例えば、旅行会話をモチーフとして開発した口語翻訳技術は、海外生産拠点での技術指導や国内自治体の在住外国人向け窓口業務への応用が考えられます。この時、コア技術の開発者でなければできない、あるいは多額の開発費と長い期間がかかるというのであれば、事業として成り立たせるのが難しく、なかなか応用が広がりません。これではせっかく開発した技術のポテンシャルが十分に活かされないこととなります。このようなコスト低減は、一義的には各企業の責任でありその結果が競争力の源泉となるべきものですが、まずは市場を広げ技術開発への再投資の下地を作り出すことが機械翻訳関係者全体の利益になると思います。既

に取り組みが進んでいるドメイン適応等の先駆的な研究に加えて多面的な取り組みがなされることを期待しています。

過去10年の機械翻訳技術の進歩には確かに目を見張るものがありましたが、基本的な翻訳精度向上という点において、もう一段のブレークスルーが必要と感じています。グローバルな企業活動が急速に拡大していく中で、対外業務の専門家ではない一般の技術者が海外の提携先や顧客との打ち合わせに出席し、ネイティブスピーカーを相手に厳しいやり取りをしなければならない、といったケースが増えています。このような人達から機械翻訳技術による支援を求められた時に、多額の開発費と長い開発期間を仮定したとしても十分な解決策の提示が難しいことが多々あります。もちろん、既存技術を巧みに組み合わせたソリューションが有効なケースもあります。しかし、多くの場合それだけでは不十分で、現状の技術レベルとそこからの漸進的な向上を前提とするだけでは、未充足のニーズの相当部分を取りこぼしてしまいます。基盤技術のブレークスルーについては大学や研究機関の成果に大きな期待を寄せていますが、企業なりのアプローチも追及していきたいと思っています。

ここまで読み返してみて、ごく当たり前のこと（あるいは逆に、できる根拠のない無理なことだと言われそうなこと）しか書いていないというためらいを禁じ得ませんが、企業の中であって、グローバルビジネスの最前線で苦闘するビジネスパーソンの切実なニーズを聞き、機械翻訳の研究開発の重要性を日々訴える中で痛感していることでもあります。次の10年も機械翻訳がそのニーズに応え続けていけるよう、本協会が提供する様々な場を最大限活用させていただきたいと思っています。

## MT Summit XIII 参加報告

愛媛大学 二宮 崇

山形大学 横山 晶一

第 13 回 機械翻訳国際会議 (Machine Translation Summit XIII、以下 MT Summit XIII と略称) は、2011 年 9 月 19~23 日、中国福建省廈門の廈門国家会計学院で開催されました。9 月 19 日にチュートリアル、9 月 20 日~22 日に本会議が開催され、9 月 23 日に第 4 回特許翻訳ワークショップが開催されました。会場は、中国に 3 か所 (他に北京と上海) ある国立の会計学院で、日本での会計大学院に当たります。設立は 2002 年で、会計士や会社の会計担当重役を養成する機関として、広大な敷地 (33.5 万 m<sup>2</sup>) の中に非常に整った設備 (建物面積 7.88 万 m<sup>2</sup>) が置かれています。参加者は 160 名を超え、地域別ではアジアが最も多く 112 名、アメリカが 23 名、ヨーロッパが 25 名、国別では、中国国内が最も多く 58 名、次いで日本 34 名、アメリカ合衆国 22 名、アイルランドと韓国各 6 名の参加者が集まりました。

私は、20 日の本会議から参加したため、残念ながらチュートリアルには参加できなかったのですが、非常に魅力的なチュートリアルが 4 件ありましたので、それらについて (概要からの抜粋で) 簡単に紹介いたします。19 日の午前中は Dekai Wu による「Syntactic SMT and Semantic SMT」の講演が行われ、従来の SMT における統語論(や意味論)の表層的な混合化ではなく、深い理論的なモデルに基づく混合化について解説が行われました。統語論については、様々な transduction grammar の解説を行い、inversion transduction grammar を実現するための足がかりとしての LTG、LITG、PLITG の良さを説明し、意味論については、従来の word sense disambiguation に対する phrase sense disambiguation の優位性について解説がありました。午前中のもう一つのチュートリアル講演は、

Mirko Plitt による「Productive Use of MT in Localization」という題目の講演で、ローカル化のための翻訳作業において、翻訳メモリの使用や機械翻訳の後編集による翻訳作業の実態について解説し、Autodesk 社における機械翻訳の取り組みについて解説がありました。午後には、Yanjun Ma、Yifan He、Josef van Genabith による「From the Confidence Estimation of Machine Translation to the Integration of MT and Translation Memory」の講演があり、後編集を想定した、翻訳メモリシステムと機械翻訳システムの統合について解説がありました。同じく午後の Alon Lavie による「Evaluating the Output of Machine Translation Systems」の講演では機械翻訳の人手による評価法と自動評価法を含んだ機械翻訳の評価法一般について解説がありました。

MT Summit XIII の一般講演は 110 件の応募論文のうち 73 件が採択され、うち 55 件が口頭発表、18 件がポスターでの発表でした。内訳は、研究論文が 90 件の応募に対して 62 件採択、ユーザースタディが 14 件の応募中 6 件採択、システムプレゼンテーションが 6 件中 5 件採択になります。一般講演は、3 つの平行セッションで行われ、統計的機械翻訳 (SMT) に関するモデルや学習の研究、翻訳の精度評価、文法理論、前処理、分野適応、翻訳支援など多岐にわたるテーマの発表がありました。

本会議初日の一般講演では、学習、機械翻訳のための前処理、音声翻訳の 3 セッションによる口頭発表がありました。学習のセッションでは SMT における MERT (Minimum Error Rate Training) パラダイムにおける不安定性を解消する研究が 2 件あり、パラメータ平均化による MERT 最適化の平滑化の研究、BLEU に対する相関係数も目的関数とし

て組み込んだ MERT との混合化の研究が発表されました。同セッションでは、さらに翻訳データのインスタンスに対する重み付けに関する発表がありました。機械翻訳のための前処理のセッションでは、前処理に関する研究や構文解析に関する研究が発表され、動詞と結びつくことによって特別な意味をもつ機能語(*particle*)まで含めた品詞解析を行うことによって翻訳性能をあげる研究、‘do not’における‘do’など対応する語がない時の単語消去に関する研究がありました。

本会議の二日目の午前中のセッションは、学習、機械翻訳を支える技術、翻訳支援の3セッションでした。学習のセッションでは追加学習データに対する適応や異なるドメインへの適応など新しいデータに対する適応の研究が発表され、新しく追加される翻訳データに対する効率的なパラメータ更新の研究や、二つのアライメントマトリックスに対する $\beta$ 分布のパラメータ消去によるベイズのドメイン適応の研究、および、統計機械翻訳における能動学習の研究の発表がありました。翻訳支援のセッションでは、参照翻訳を用いなくても計算できる翻訳スコアを用いて、規則に基づく機械翻訳 (RBMT) と SMT を混合したシステムの研究、人間の後編集のプロセスを模した Post-Editing Action (PEA) という枠組みを使って2つの異なる MT システムへの後編集を試みた研究、後編集すべき語を翻訳メモリを用いて同定する研究の発表がありました。

二日目の午前は続いて、モデル、言語学的知識に基づく機械翻訳、ユーザスタディの3セッションによる口頭発表が行われました。言語学的知識に基づく機械翻訳のセッションでは、中英のアライメントされたパラレルコーパスを用いて、中国語の単複形を自動的に推定する研究の発表がありました。この次の発表がキャンセルになったこともあり、単複をめくり、会場で様々な討論が行われました。ユーザスタディのセッションでは、Adobe などの企業において機械翻訳がどのように用いられているかについて解説が行われました。

二日目の午後は、モデル、ドメイン適応、マルチパス翻訳の3セッションがありました。マルチパス翻訳のセッションでは、語順を入れ替える前処理の規則を依存構造の解析結果から学習する研究、Phrase-Based SMT (PBSMT) において PBSMT を後処理に用いるパイプライン処理の新しい結合方法の研究、英日の翻訳にのみ適用可能な磯崎らの語順入れ替えの前処理を日英の翻訳にも適用可能とする研究が発表されました。その後、ポスター発表やシステムプレゼンテーションが行われました。

三日目の午前中は、モデル、コーパス、機械翻訳のための文法理論の3つのパラレルセッションにおける口頭発表と、それに続いてポスター発表とシステムプレゼンテーションがありました。機械翻訳のための文法理論のセッションでは、Dekai Wu による発表があり、彼らが提案した Linear Transduction Grammar が、ある範囲内では、いろいろな言語現象を合理的に説明できるという主張を、このセッションの発表が2件だったためか、1時間にわたりチュートリアル的に発表していたのが印象的でした。その後、同セッションでは、制約付き同期文法の研究の発表が行われました。

三日目の午後には、評価、システム組合せ、ユーザスタディの3セッションがありました。システム組合せのセッションでは、うまく翻訳できない原言語の文を言い換えの技術で変換する研究、粒度の細かい RBMT と SMT の混合化の研究、PBSMT における句テーブルを浅い RBMT から得られる辞書で増強する研究、N-best を利用するシステムでの解析誤りの伝播を防ぐためのハイパーグラフに基づく学習の研究発表がありました。

全体的には機械翻訳に対する性能向上の研究が多かったと思いますが、今回の会議では、特許翻訳を中心に、実社会における機械翻訳の利用法や環境整備に関する話を多く聞いたのが非常に印象的でした。現在すでに多くの企業で機械翻訳が実用的に用いられていて、今後、益々必要とされるということではないかと思います。

2011年9月19日(月)～23日(金)、中国・厦門(アモイ)の厦門国家会計学院(XNAI)において第13回機械翻訳サミットが開催されました。本会議は20日～22日の三日間にかけて行われ、19日にチュートリアル、23日に併設ワークショップとして第4回特許翻訳ワークショップが開かれました。機械翻訳サミットは隔年で開催されており、今回はAAMTの主催、中国・中国語情報学会(CIPS)および厦門大学の後援により中国では初めての開催となりました。

中国の58名を筆頭に、日本、アメリカ、欧州等世界各国からの参加者は160名を超えました。応募論文110編のうち73編が採択され、うち55件について口頭発表、18件についてポスター発表が行われました。また、論文の種別の内訳は研究論文62件、ユーザスタディー14件、システムプレゼンテーション6件でした。近年広く研究が進められている統計的機械翻訳のモデリングや訓練に関する研究の他、翻訳の精度評価、文法理論、コーパス、翻訳支援、音声翻訳等の多様なテーマにおける発表がありました。

私は20日の本会議から参加し、まず董振東教授(中国科学院)から基調講演がありました。翻訳に対する需要が大きいなか、現存する機械翻訳システムと実際の需要とのギャップを埋めるための人間中心の機械翻訳をテーマとして講演をされました。機械翻訳技術のうち人間の翻訳支援に必要な部分を分解し、システム側の知識(辞書、翻訳メモリ等)とユーザ側の知識(個人の語学能力、校正データ等)を統合して一つのプラットフォームとして組み上げることで、翻訳システムと翻訳のための知識を互いに洗練していく流れをつくるべきとの議論をされ、またその一例として格微軟件(Ge-soft)に

より開発された協調翻訳プラットフォーム、およびそれを用いた比較実験による翻訳能率の向上について紹介されました。

引き続き、Mike Dillinger氏(AMTA副会長、TOPs Globalization Consulting社長)から、“MT Everywhere: Next Steps”と題して招待講演がありました。機械翻訳システムの普及にもかかわらず、それほど多くの企業が機械翻訳を使いたがらない理由は何か。それは機械翻訳システムの価格が高いこと、職業翻訳者による抵抗、あるいは翻訳性能の不完全さのためではなく、機械翻訳そのものに対する理解が不足しているためであり、我々はもっとユーザに機械翻訳を理解してもらうための努力を必要があると説いています。例えば機械翻訳とは何かと問われた時に、「翻訳ソフトです」と答えると、ユーザは「人間みたいに翻訳してくれるソフトで、翻訳者や通訳の代わりになってくれるだろう」と期待します。そして実際使ってみると、「これを使ったら我が社の翻訳部門をリストラすべきか」といった誤解を生んだり、あるいは、「全然翻訳になってないじゃないか」といったような答えが返ってきたりしてしまいます。それに対して、機械翻訳とは「翻訳を能率化するソフトです」と答えれば、興味を持つユーザが増え、かつ期待した結果が得られるはずだということになります。また、異なる志向のユーザに対して個々に対応すべきであることや、機械翻訳システムを翻訳そのもののために使うのではなくユーザが求めるタスクに対して開発していくことが重要であると述べています。

この後、特許翻訳に関する特別セッションがあり、辻井潤一教授(マイクロソフト・リサーチ・アジア)から序論として特許機械翻訳に関する解説がなされました。特許機械翻訳を行う手段として大規

模並列コーパスの利用や統計的機械翻訳がある中、課題として大量の専門用語の扱い、特許文にありがちな長文の扱いの他、各国の特許を扱う機関間での協力も必要であるとの議論を展開しました。また、近年の NTCIR において特許翻訳タスクが開始され、特許向けの機械翻訳技術の開発が行われていることの紹介もされました。

この後、チェ・ユチョン氏（韓国特許情報院 (KIPI)）と Bruno Pouliquen 氏（世界知的所有権機関 (WIPO)）からそれぞれ招待講演がありました。チェ氏からは現状での韓国特許庁における機械翻訳の応用例として日韓、韓英および英韓の機械翻訳を用意し、機械翻訳を統合した特許検索システムの紹介と今後の取り組みについて述べられました。また Pouliquen 氏からは利用可能な資源として、英仏の並列特許出願文書コーパスである COPPA、9 か国語で利用可能な言語横断検索システムである CLIR、および WIPO の機械翻訳システムである TAPTA の紹介をされました。

本会議の招待講演は 22 日の最後にもう一件、Hans Uszkoreit 教授（DFKI/ザールラント大学）から、“Strategic MT Research in Europe: Themes, Approaches, Results and Plans”と題して行われました。Uszkoreit 教授は欧州における機械翻訳の指導的研究者であり、EuroMatrix、EuroMatrixPlus、TaraXÜ や META-NET といった様々なプロジェクトに参画しています。これらのプロジェクトの概観の他、現状の研究の進展と課題として、様々な面からの研究が進み新たなアプローチが提案されていく中で、カバレッジの低さや翻訳性能の向上速度の遅さ、機械翻訳の応用範囲の広さに対応するほどの資金が得られていない点等を指摘しています。また、機械翻訳結果のうちよく現れる誤りを含むもの、比較的修正が容易なもの、およびそれ以外の三つの方向に分けて研究を行うという提案や、欧州では外向きおよび欧州内での翻訳需要が比較的大きい点を述べて締めくくりました。

本会議の最後では、東京工科大学の飯田仁教授

に IAMT Award of Honor が贈られ、IAMT の次期会長 Andy Way 教授の挨拶と次回機械翻訳サミットの開催地ニースの紹介を行い閉会しました。

23 日には第 4 回特許翻訳ワークショップが開かれ、およそ 70 名の参加者を数え特許翻訳に対する関心の高さを伺わせました。午前中は欧州特許庁 (EPO) の Bertrand Le Chapelain 氏、中国特許情報センター (CPIC) の蔣宏飛氏、および日本特許庁の山本英一氏から各国での特許に対する機械翻訳への取り組みに関する招待講演が行われました。欧州特許庁では欧州の各言語から英仏独の各言語への翻訳が要求されており、翻訳システムの要素としてコーパス集、翻訳性能の評価、および翻訳ゲートウェイについて解説されました。中国では特許出願数の増大に伴い機械翻訳の活用を行っており、翻訳支援システムや翻訳システムを組み込んだ特許検索の紹介や、中英・英中に加え日中翻訳も開発中であることを述べました。また日本からは機械翻訳を利用したサービスとして、公開特許公報英文抄録 (PAJ) や特許電子図書館 (IPDL) の他、高度産業財産ネットワーク (AIPN) における日英機械翻訳の利用、日英・日中機械翻訳研究の促進等を紹介し、将来の展望として中国語や韓国語等での言語横断検索や概念検索のシステム作りを行っていることが紹介されました。

一般講演は 7 件あり、統計的機械翻訳の特許翻訳への適用やその改善、訳し分けへの対処、日本語機能表現の曖昧性の解決、特許翻訳における単語アライメント、および欧州の機械翻訳プロジェクト MOLTO の特許翻訳への応用についての発表が行われました。またワークショップの最後にはパネル討論が行われ、将来の特許機械翻訳に関する取り組みについて、特にアジア言語の扱いに焦点を置いた議論が交わされました。

全体を通して機械翻訳そのものの改善の話題が中心ではありますが、機械翻訳技術の社会への展開や環境整備などの応用面の重要性が強調されていたことが私としては印象に残った会議でした。

## MT Summit XIII における「特許翻訳 WS」等に関する報告

(財)日本特許情報機構 特許情報研究所

調査研究部長 森藤 淳志

### 1. はじめに

筆者は、平成22年10月から縁あって(財)日本特許情報機構(Japio)に在籍し、AAMT/Japio特許翻訳研究会を通じて、AAMTの活動に関与することができる機会に恵まれています。この度、MT Summit XIII(Xiamen)における特許翻訳に関するワークショップ(WS)と特別セッションのプログラム委員(Program committee)として関わり、MT Summitに初めて参加しました。

AAMTにおいて筆者は新参者の部類に入ると考えられますので、本稿では、まず筆者とMT Summitとの関わりについて簡単に触れたいと思います。また、第4回特許翻訳WS等の詳細な報告は他の参加者からなされるようですので、筆者からは今回のMT Summitについて補完的な報告をいたします。

### 2. 筆者とMT Summitの関わりなど

筆者とMT Summitの関わりは2007年に開催されたMT Summit XI (Copenhagen)に遡ります。

筆者は当時、日本特許庁にて、特許公報の日→英機械翻訳(MT)を一般に提供する「特許電子図書館(IPDL)・英語版」や拒絶理由通知などの審査書類の日→英MTを他国審査官に提供する「高度産業財産権ネットワーク(AIPN)」(2011年8月現在48の国・地域へ提供)を担当していました。

この頃に、第2回特許翻訳ワークショップでの講演の機会を頂戴し、派遣者の決定や講演内容の推敲を通じて、陰ながらMT Summitに関わりました(私の配下の遠山敬彦が参加・講演しました)。

### 3. 日本特許庁におけるMT活用と今回の講演

日本特許庁は1999年にIPDLを立ち上げ、その後まもなく、日→英MTを活用したサービスの提

供を開始し、さらに、2004年10月からAIPNの運用を開始しています。特許情報に機械翻訳を適用して、ネットでサービス提供することは、当時、他庁のサービスと比較して先駆的な試みでした。

しかしながら、今回のMT Summitで講演された、各庁(日本特許庁、世界知的所有権機関、欧州特許庁、韓国特許庁、中国特許庁)における機械翻訳の活用状況に関する内容を見る限り、その後の世界知的所有権機関など他庁におけるMT活用の取り組みの状況には目を見張るものがあります。例えば、韓国特許庁から、審査官と一般公衆に対して、日本特許公報の日→韓MTサービスが提供され、欧州特許庁からは多言語間のMTサービスが提供されています。日本特許庁の上記MTサービスは、日→英MTの精度改善にとどまっていて、日本企業等ユーザ向けに外国語→日本語MTサービスは未だ実現されていません(民間プロバイダが一部実施)。

### 4. 日本特許庁の国際知財戦略と今回の講演

経済産業大臣の諮問に応じて、経済及び産業の発展に関する重要事項を調査審議する産業構造審議会の下部には、特許等の産業財産権に関する政策審議を行う場として、知的財産政策部会が設置されています。さる平成23年7月19日に第16回産業構造審議会 知的財産政策部会が開催され、「国際知財戦略(Global IP Initiative)～国際的な知的財産のインフラ整備に向けた具体策～」(以下、「国際知財戦略」)が審議されました。

この国際知財戦略の資料中で、「日本語・英語以外の特許文献、特に急増する中国文献に対し、企業・特許庁ともに戦略の転換が必要」であること、「中国において、無審査登録の実用新案権に基づき、賠償を求められる事例が出現。分類や翻訳の整

備により、中国文献などを容易に把握できるように」することの必要性が報告されました。同資料中で、こうした現状を踏まえ、中→日、韓→日 MT 機能を備えた外国特許文献検索システムの整備を行うという方向性が示されています。

この点、今回の MT サミットにおける日本特許庁山本氏のプレゼンテーションでは「The JPO's New Search System」というスライドにおいて、「Cross-lingual search system (Chinese, Korean, etc.)」という事項が含まれていることから、日本特許庁は上記審議会での審議結果を実施に移す意向があることが伺うことができます。

Japio での研究活動を通じて、筆者も、日本のグローバル企業の知的財産部の方と意見交換をすることがあります。日本を代表するグローバル企業が、中国文献の未曾有の増加や訴訟件数の急増（訴訟大国・米国の 2 倍に到達）に対して苦慮している現状を垣間見るにつけ、日本特許庁の上記施策が実現されること、その際には、日本における中日 MT の研究成果が活用され、高品質な MT 結果が得られるようになることを希求します。

## 5. 本 MT サミットにおける Japio の貢献

Japio は、毎年 11 月に主催する特許・情報フェア&コンファレンス（東京）等において、世界的所有権機関や欧州特許庁などから特許情報専門家を招聘したり、別途、研修生を受け入れたりするなど、海外の特許庁とのネットワークを有しています。このネットワークを活用し、また、日本特許庁の力添えも頂戴しながら、Japio は、特許翻訳特別セッションと第 4 回特許翻訳 WS への各国特許庁から招待講演者の招へいを担当しました。

また、WS では、Opening Address を当機構の専務理事守屋が担当し、その中で、Japio の機械翻訳への取り組み（Japio コーパス、AAMT/Japio 研究会）に加え、多言語化に向けて中国語対応に着手したことを報告しています。

## 6. 開催地アモイ；観光地？ストイック？

廈門（アモイ：Xiamen）は、中国福建省の南東部、台湾の対岸に位置する島で、南西に隣接するコロンス島は、歴史的建築物が数多くあり、観光スポットとして多くの人を訪れます。他方、今回の会場となった Xiamen National Accounting Institute (XNAI) は、主に中国国内の会計担当者のための研修施設であり、観光スポットから隔離されたストイックな環境であることから、筆者は MT サミットに集中できました。下図：XNAI の全景図



## 7. むすび：謝辞と今後

今回の特許翻訳 WS 等は、成功裏に終了し、各国特許庁からの出席者にとっても有益であったと思います。この特許翻訳 WS 等の企画段階から当日の運営に至るまで、辻井委員長以下 AAMT/Japio 特許翻訳研究会の全員で対応してきました。特に、chair を勤めていただいた横山先生と co-chair の江原先生による、各国特許庁からの招待講演者に対する紳士的な対応も成功の一因と思います。委員長と両氏を始め関係者の皆様に深く御礼申し上げます。

最後に一言付記します。昨今の記録的な円高などもあり、日本の企業の海外展開はさらに加速することが予想されます。直近では、中国特許文献が注目されていますが、今後は、ロシア語などの特許文献への対応が必要となると予想されます。まずは、中国語での機械翻訳を成功させ、次のターゲット言語にも速やかに展開していくことが必要となると思います。今後の非英語圏 MT の発展に期待します。

## MT Summit XIII Technology Showcase 報告

東芝ソリューション株式会社

熊野 明

MT Summit XIII の本会議が開催された 9 月 20 日から 22 日までの 3 日間、機械翻訳に関連する技術展示を行う Technology Showcase が開かれた。論文発表会場 XNAI 会議棟の 1 部屋で、7 団体がそれぞれの技術や製品を紹介した。筆者は、自社の技術・製品の紹介のために 3 日間参加した。

学会参加者には、技術文書のローカライゼーション、海外の特許文書の検索など、機械翻訳の実運用に関心のある人が多く、論文発表や講演セッションの休憩中に、頻りに展示会場を訪れていた。参加者の関心は、製品の仕様だけでなく、しばしば翻訳エンジンに関する技術的な項目にも及び、展示説明者と積極的な議論を行っていた。

以下に各団体の展示概要を紹介する。(ABC 順)

### ○ Baidu (百度)

中国最大の検索サイトを運営する企業。Web から収集した対訳テキストを利用し、SMT(Statistical MT)をメインエンジンとした機械翻訳を開発し、ポータルサイトで公開している。

### ○ ETRI (Electronics and Telecommunication Research Institute)

韓国の研究機関であり、韓国語と英語、韓国語と中国語の双方向の機械翻訳技術を開発している。2005 年には特許文書を対象とした韓国語から英語への機械翻訳を、翌 2006 年には英語から韓国語への機械翻訳を製品化し、KIPO (韓国特許庁) にも技術を提供している。現在は技術文書だけでなく、話し言葉に対する機械翻訳の開発にも取り組んでいる。

### ○ Fujitsu Research and Development Center (富士通研究開発中心)

富士通が中国に設立した研究開発会社。機械翻訳の技術を利用して、中国語の特許を日本語で検索するなど、言語横断検索の技術を紹介した。

### ○ Microsoft Research

検索サービス Bing で公開している機械翻訳は SMT のエンジンを利用しており、Microsoft Office のアプリからも利用できる。CTF (Collaborative Translation Framework) は、3 種類のメンバーの分業によって文書翻訳を行なう体系。複数の User が翻訳した結果を、Moderator が確認、修正して承認する。Owner は全体を管理し、翻訳文書を完成する。

### ○ NICT

Web を利用することにより、複数の人の手で文書全体の翻訳を行なう“みんなの翻訳”のプロジェクトを紹介した。

### ○ Shenyang Global Envoy Software (瀋陽格微軟件公司)

自社の機械翻訳を活用した翻訳ビジネスを行っている中国企業。航空機マニュアルのような大規模なドキュメントに対し、専門のオペレータを集めたプロジェクトを組織化して、翻訳を行なう。前処理では、原文書から専門用語を抽出して訳語を決定する。後処理では、類似の原文に対する訳文の表現を統一し、全体で均質な訳文を実現している。

○ 東芝ソリューション

日本の企業として唯一の出展。日本語と英語、日本語と中国語の双方向の機械翻訳を、企業向けサーバシステム“The 翻訳エンタープライズ”、クラウドサービス Eiplaza/MT で提供している。パッケージソフト“The 翻訳プロフェッショナル”（日本語・英語双方向）では、SDL Trados Studio との連携機能により、プロの翻訳家に向けた翻訳環境を実現した。翻訳エンジンは、RBMT(Rule-based MT)と EBMT(Example-based MT)を組合せたものである。

複数の企業の社内文書ローカライゼーション担当者からは、SDL Trados Studio との連携に関して具体的な質問を受け、関心の高さが明らかになった。

今回は残念ながら 7 団体の参加にとどまったが、2 年後にはより多くの団体の出展を期待したい。

## MT Summit XIII 参加報告

AAMT 機械翻訳課題調査委員会  
標準化・共有化ワーキンググループ 大倉 清司



■写真：MT サミットの看板



■写真：XNAI の敷地

今回で 13 回目となる MT Summit が 2011 年 9 月 19 日から 23 日までアモイ国家会計学院(XNAI)で開催された。9 月 19 日は機械翻訳に関するチュートリアル、9 月 20 日から 22 日までが本会議、9 月 23 日が特許翻訳に関するワークショップであった。私の参加目的は、AAMT 機械翻訳課題調査委員会の標準化・共有化ワーキンググループで策定している UTX(\*)のバージョン 1.11 について発表することであった。今回はアジアでの開催だったが、欧米からの参加者も多かった。招待講演、論文発表の平行セッション、ポスター発表の他、テクノロジーケースとして、企業のデモなどが行われた。

会場は、アモイ空港から車で約 30 分のところにある。敷地内には教学楼（会議ホール、教室などがある施設）の他、ホテル、学生食堂がある。市街地から離れた場所であるため、学会に集中して参加できる環境であった。敷地内は山に囲まれたところで散策路もよく整備されており、水路もある。

水路のまわりは 30 分ほどで散策できるようだが、外周は軽いトレッキングができるような上り下りがある道で、1 周するには時間がかかりそうだった。

研修中と思われる中国人もたくさん見かけた。治安は非常によかった。



■写真：学会会場の建物（教学楼）

平行セッションの聴講者は 1 教室に 30 人～50 人ほどだった。しかし私の発表は本会議の最終日の最終セッションということもあってか、聴講者は少なく 20 人程度だった。UTX の意義は伝わったと思う。

発表タイトルは” UTX 1.11, a Simple and Open User Dictionary/Terminology Standard, and its Effectiveness with Multiple MT Systems” である。

(1)UTX の概要、(2)UTX 1.11 で導入された 4 つの term status および、UTX から 4 つの翻訳ソフト用ユーザー辞書への変換ツール、(3)UTX 辞書を使つての UTX 仕様書の自動翻訳結果の評価について説明した。



■写真：発表の様子

発表後の質疑応答は以下の通りだった。

Q：分野の中でも単語の曖昧性があるのではない  
か

A：曖昧性がないように分野を決める必要がある

Q：他のフォーマットからはユーザー辞書に変換  
できないのか

A：他のフォーマットは仕様が複雑で辞書の作成  
も翻訳者にとっては困難であり、変換ツールも少な  
いため難しい。

Q：すでにつくった 3 つの辞書とは

A：計算言語学、医学、法律。

Q：Web2.0 のように、辞書を集められないのか

A：そう考えている

「1 分野でも 2 つ以上の辞書が必要な場合があ  
ると思う」というコメントもあり、これに関しては  
今後検討していきたい。

発表全体としての感想であるが、全体的に統計翻  
訳に関する発表が多かった。しかし、翻訳精度向上  
というよりは、翻訳精度が悪いなりにどうやって自  
動翻訳を使っていくかという、実際的な発表が多か  
ったと思う。欧米では翻訳需要は劇的に増大してお  
り、自動翻訳なしでは考えられないほどだという。  
実際に自動翻訳を使っている企業からも発表があ  
り、聴講者は熱心に発表を聴いていた。

(\*)UTXについては以下のURLをご参照ください。

<http://www.aamt.info/japanese/utx/>

## Machine Translation in China

Qun LIU

Key Laboratory of Intelligent Information Processing  
Institute of Computing Technology, Chinese Academy of Sciences  
liuqun@ict.ac.cn

---

### 1. Introduction

As in many other countries, the way of machine translation in China is full of twists and turns. In this paper I will introduce the history of machine translation in China in three periods.

### 2. The first period

The earliest machine translation research in China was carried out from late 1950s to early 1960s. In December 1956, the Chinese government released its "1956-1967 Prospective Plan of Science and Technology Development"[1], in which "automatic translation" was listed as an important task in the item 41. In 1959, researchers in the Institute of Computing Technology (CAS-ICT) and the Institute of Linguistics (CAS-IL) of Chinese Academy of Sciences successfully conducted the first Russian-Chinese machine translation experiment in China. After that, many institutes continued machine translation research for many years.

### 3. The second period

- Research

The machine translation research was interrupted for about ten years because of the Culture Revolution. In mid-1970s, as part of the activities of a national project - the "748" Project, machine translation research was recovered in many institutes [3].

In early 1980s, personal computers were much power and cheaper and were ready to be used in machine translation research. Many institutes and universities were involved in machine translation researches from this period. Harbin Institute of Technology (HIT) and Northeast University (NEU) began their Chinese-English machine translation researches from mid-1980's. Nanjing University (NJU) started its Japanese-Chinese machine translation research in this period. CAS-ICT recovered its machine translation research from late 1980's. The institute of Automata, Chinese Academy of Sciences (CAS-IA) began its research on spoken language translation research and joined the Consortium for Speech Translation Advanced Research (C-STAR) in late 1990's.

From 1987, many Chinese institute and universities were involved in the "Joint study for Multilingual Machine Translation" project, which was supported by Japanese government, and aimed to an ambitious goal of high quality machine translation between five Asian languages, i.e. Japanese, Chinese, Thai, Malay and Indonesia Language [5]. This project lasted about 10 years.

Three machine translations systems obtained the China's National Science and Technology Awards. MT-IR-EC, an English-Chinese title and catalog translation system for journal publishing, which was developed by the Research Institute of Post and Telecommunication Science, won the first National S&T

Award as a machine translation system for the first time. KY-1, an English-Chinese machine translation system, which is developed by the Academy of Military Sciences, won the second class National S&T Awards. The KY-1 system was also the predecessor of the first commercial machine translation system Transtar, which was released by Chinese National Software & Service Co. Ltd. (CS&S) in late 1980s. The Huajian Chinese-English machine translation system, developed by Huajian Co. Ltd, a spin-off from CAS-ICT, won the first class China's National Science and Technology Award in 1996 [2,3].

In this period, almost all the machine translation systems adopt rule-based approach, while some of them took example-based approach as a complement.

- Evaluations & Resources

From 1995 to 2005, the expert group of "863" plan held six open machine translation evaluation in 1994, 1995, 1998, 2003, 2004 and 2005. In the early "863" evaluations, Peking University proposed a check-point based approach for automatic machine translation evaluation [17].

Some public available language resources were also established and released, for example, the Grammatical Knowledge-base for Contemporary Chinese (Peking University, PKU), the People's Daily Corpus for Chinese word segmentation and segmentation (PKU), the Hownet (a lexical semantic knowledge database developed Prof. Zhendong Dong), etc. An open source tool ICTCLAS (CAS-ICT) for Chinese word segmentation and part-of-speech tagging was also released in 2003. These resources and tools greatly promoted the machine translation and other natural language processing researches and applications.

- Applications

CAS-ICT and GSL Co. Ltd. (Hongkong) released the first electronic dictionary which supported machine translation between Chinese and English -- the Instant-Dict EC863 in 1992, which made great commercial success in the market [3]. There were also many machine translation systems appeared in the software market, such as the Huajian System (Huajian Group), the Transtar system (CS&S), the GaoLi system (GaoLi Co. Ltd. & the Institute of Linguistics, Chinese Academy of Social Sciences), the CCID system (CCID Group), the Kingsoft Quick Translation system (Kingsoft Co. Ltd.), the web-based Read-World Translation System (Sunshine Co. Ltd), etc.

Some research groups, such as HIT group and NEU group, also developed some application systems collaboratively with domestic or foreign companies.

#### 4. The third period

- Research

From 2004, some groups start their statistical machine translation (SMT) researches in China. These groups includes: the CAS-ICT group, the CAS-IA group and the Xiamen University (XMU) group. In 1995, these three groups held the first Symposium of Statistical Machine Translation (SSMT) in Xiamen, which became the precedent of the series of the annual China Workshop of Machine Translation (CWMT). From then on, the SMT researches grew very fast in China.

Many Chinese research groups (CAS-ICT, CAS-IA, HIT, NEU, Microsoft Research Asia, Toshiba China, etc.) published papers in the top conferences (ACL, EMNLP, COLING, etc.) and Journals (Computational Linguistics, Machine Translation) in MT area in recent years [4-16]. CAS-ICT proposed a series of new syntax-based translation model and approach, i.e. maximum-entropy bracketing transduction grammar model [5], tree-to-string model [6], forest-based approach [10,11], joint parsing and translation

approach [14], and dependency-to-string model [16]. These research works are referred or followed by many researchers in the machine translation community.

- Evaluations & Resources

From 2007, an open machine translation evaluation was held accompanying with almost every CWMT<sup>1</sup>, where all the training data and test data were shared on ChineseLDC<sup>2</sup>. In recent years, many new language pairs, include some minority languages in China, were include in the CWMT MT evaluation, such Mongolian, Tibetan, Uygur, Kazakh, Kirgiz, etc. More than 20 research groups have ever participated in the CWMT evaluations, including some groups from foreign countries or foreign invested enterprises in China, such as Systran (USA), Microsoft Research Asia (US-China), I2R (Singapore), Sharp Co. Ltd. (Japan), NICT-ATR (Japan), Dublin City University (Ireland), Fujitsu R&D Center (Japan-China), NTT (Japan), etc.

In 2006, five research groups, including CAS-ICT, CAS-IA, XMU, Institute of Software / Chinese Academy of Sciences (CAS-IS), HIT, joint released an open source phrase-based SMT system: Silk Road version 1.0 in the second SSMT. In 2011, NEU released a new open source machine translation system named NiuTrans.

Some Chinese MT groups gained good results in the international machine translation evaluations. In the NIST open MT Evaluations, CAS-ICT ranked No.5 in 2006 in Chinese-English track, and No.3 in 2009 in the progress test of system combination Chinese-English track. CAS-IA and CAS-ICT also ranked No.1 in the IWSLT spoken language translation evaluation for many times.

- Applications

Besides the traditional machine translation companies which adopt rule-based approach in their products, such as Huajian Group, CCID and CS&S, some companies or institutes began to adopt statistical machine translation products in real applications. Netease Co. Ltd. released its web-based Youdao Translation products in August 2008. Baidu Co. Ltd., the most big search engine company in China, also released its web-based machine translation system in 2011. CAS-ICT and the Eastlinden Co. Ltd. jointly developed a patent translation system and utilize it for large scale patent translation. In Beijing 2008 Olympic Games, machine translation technologies were also used to provide automatic or computer-aided translation services.

## 5. Conclusion

Machine translation has a history of more than 50 years in China. A number of institutes, universities and companies have been involved in machine translation researches and applications. In recent years, Chinese researchers achieved great progress in statistical machine translation research. We believe the machine translation researches and applications will have a better prospect in the future in China, along with the growing of the Chinese economy and the exchanges between China and the world.

## Acknowledge:

Thanks to Dr. Yajuan Lü for her valuable suggestions and information.

---

<sup>1</sup> <http://nlp.ict.ac.cn/new/CWMT/index.php>

<sup>2</sup> <http://www.chinese.org>

## References:

- [1] 科学规划委员会, 1956-1967 年科学技术发展远景规划纲要, 1956 年 12 月
- [2] 董振东, 中国机器翻译的世纪回顾, *中国计算机世界*, 2000 年第 1 期, 2000 年 01 月 03 日
- [3] 黄果, 重建巴比塔——冯志伟研究员谈我国机器翻译发展历程, *计算机世界报*, 1999 年 9 月 27 日
- [4] Yajuan Lü, Ming Zhou, Sheng Li, Changning Huang, Tiejun Zhao, Automatic Translation Template Acquisition Based on Bilingual Structure Alignment, *Computational Linguistics and Chinese Language Processing*, Vol. 6, No. 1, February 2001, pp. 83-108
- [5] Deyi Xiong, Qun Liu, and Shouxun Lin. 2006. Maximum Entropy Based Phrase Reordering Model for Statistical Machine Translation. In *Proceedings of COLING/ACL 2006*, pages 521-528, Sydney, Australia, July.
- [6] Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-String Alignment Template for Statistical Machine Translation. In *Proceedings of COLING/ACL 2006*, pages 609-616, Sydney, Australia, July.
- [7] Hua Wu and Haifeng Wang, Pivot language approach for phrase-based statistical machine translation, *Machine Translation*, Volume 21, Number 3, 165-181, 2007
- [8] Dongdong Zhang, Mu Li, Chi-Ho Li, Ming Zhou, Phrase Reordering Model Integrating Syntactic Knowledge for SMT, in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 533–540, Prague, June 2007.
- [9] Chi-Ho Li, Dongdong Zhang, Mu Li, Ming Zhou, Minghui Li, Yi Guan, A Probabilistic Approach to Syntax-based Reordering for Statistical Machine Translation, in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 720–727, Prague, Czech Republic, June 2007
- [10] Haitao Mi and Liang Huang. 2008. Forest-based Translation Rule Extraction. In *Proceedings of EMNLP 2008*, pages 206-214, Honolulu, Hawaii, October.
- [11] Haitao Mi, Liang Huang and Qun Liu. 2008. Forest-Based Translation. In *Proceedings of ACL 2008*, pages 192-199, Columbus, Ohio, USA, June.
- [12] Jiajun Zhang, Chengqing Zong, Shoushan Li, in: *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Manchester, UK, August 2008.
- [13] Hua Wu, Haifeng Wang and Chengqing Zong, Domain adaptation for statistical machine translation with domain dictionary and monolingual corpora, in *Proceedings of the 22nd International Conference on Computational Linguistics*, Manchester, UK, August 2008.
- [14] Yang Liu, and Qun Liu. 2010. Joint Parsing and Translation. In *Proceedings of COLING 2010*, Beijing, China, August.
- [15] Yang Liu, Qun Liu, and Shouxun Lin. 2010. Discriminative Word Alignment by Linear Modeling. *Computational Linguistics*, 36(3).
- [16] Jun Xie, Haitao Mi, Qun Liu, A novel dependency-to-string model for statistical machine translation, in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 216--226, Edinburgh, Scotland, UK. July 2011
- [17] Yu Shiwen, Automatic evaluation of output quality for Machine Translation systems, *Machine Translation*, Volume 8, Numbers 1-2, pages 117-126, 1993, DOI: 10.1007/BF00981248

## 翻訳時の日本語表記に対する取り組み

日本翻訳連盟  
標準スタイルガイド検討委員会  
委員長 田中 千鶴香

### 日本翻訳連盟 (JTF)

日本翻訳連盟は、産業翻訳にかかわる企業、団体、個人の会員からなる業界団体です。翻訳者（フリーランス・企業内翻訳者）、エージェント（翻訳会社）、ソースクライアント（最終需要者）が会員として加入しており、翻訳事業に関する調査や研究、情報の収集や提供を行い、資格能力審査（ほんやく検定）を定期的の実施しています。

複数の委員会活動があり、そのなかのひとつである「標準スタイルガイド検討委員会」では、翻訳時の日本語表記に関する課題に取り組んでいます。昨年度はオープンな Web フォーラム「SINAPS Forum」を運営して、和訳時の日本語表記について広く意見を募りました。現在、JTF 推奨「標準日本語スタイルガイド（翻訳用）」を作成中です。

### 翻訳と日本語スタイルガイド

一般に実用文を書く際には、用字・用語や文章表現をできるだけ統一する方がのぞましいとされます。表記が統一された文書は、読み手の理解を助けるからです。外国語から日本語に翻訳される実務文書にも、当然、同様の品質基準が適用されます。

また実務翻訳の場合、複数の翻訳者・翻訳会社が同じ業務（同一のマニュアルの翻訳など）に携わる場合があります。担当した翻訳者によって表記や表現にばらつきが生じないよう、あらかじめ表記や表現について一定の基準やルールを定め、それに従って翻訳を行う必要があります。

このような表記仕様を集めたガイドラインを「スタイルガイド」あるいは「日本語スタイルガイド」（和訳用の場合）と呼びます。スタイルガイドは、発注元（顧客）ごとに決まっている場合、翻訳会社が作成あるいは指定する場合、プロジェクト単位で作成される場合があります。

実務翻訳の現場で、日本語スタイルガイドが常に使用されるわけではありません。日本語スタイルガイドが最も頻繁に使用されるのは、情報技術（IT）分野の翻訳です。

IT 分野の翻訳では、ユーザー向けのマニュアルやオンラインヘルプなど、文字量の多いドキュメントをごく短期間に翻訳しなければならないことがあります。単一の翻訳プロジェクトに複数の翻訳会社/翻訳者がかわることが珍しくありません。

また 15 年ほど前から、翻訳メモリーツールの利用が進み、過去訳を効率的にリサイクル（再利用）するためにも、表記や表現のゆらぎをできるだけ避ける必要があります。

IT 分野の翻訳で使用されるスタイルガイドのなかには、表記や表現のルールを非常に細かく定め、ページ数が何十ページを越えるものもあります。場合によっては、スタイルガイドへの準拠が品質基準のひとつとされ、仕様に適合した表記や表現による翻訳が強く求められます。

一方、医薬や特許などのほかの分野では、IT 分野での翻訳時ほど、スタイルガイドが活用されません。発注元から文体や訳語が部分的に指定されることはあっても、用字や表記については、担当する翻訳会社/翻訳者の裁量にまかされるのが一般的です。

もちろん、常識の範囲で、表記や表現の不統一を避ける努力がなされますが、IT 分野のように、発注元から何十ページにも及ぶスタイルガイドが支給され、表記仕様への準拠が厳しく要求されることはほとんどありません。

分野によってスタイルガイドの使い方に違いが生じているのは、扱うドキュメントの種類の違いや翻訳メモリーツールなどの自動化ツールの活用度合いに関連があると思われます。

### 翻訳用のスタイルガイドで規定される表記仕様

代表的な表記仕様について調べるには、市販されている表記ハンドブックや日本語スタイルガイドを見るのが早道です。しかし、カタカナ語の表記など、表記の不統一が特に起きやすい表記仕様については、ハンドブックごとに仕様が変わっていたり、代表的な考え方を示すにとどまって、推奨する表記仕様が明記されていない場合があります。決定版となる日本語スタイルガイドがまだ存在しないのが現状です。

以下に、一般的な翻訳用のスタイルガイドで規定される表記仕様を項目別に示します。このなかには「半角文字と全角文字の間の半角スペースの有無」など、市販のハンドブックにあまり記載されない表記仕様があります。

#### 1. 文体

・敬体か常体か？

本文を、敬体（です・ます調）で書くか、常体（である調）で書くかを規定します。本文以外に、見出し、箇条書き、表中・図表内のテキスト、図表のキャプション、用語集のそれぞれについて文体が細かく規定される場合があります。

#### 2. 文字の表記と使い分け

ひらがな・カタカナ・漢字の表記、漢字の送りがな、複合語の送りがな、ひらがなと漢字の使い分け、カタカナの長音の有無、算用数字（位取りの表記）の表記、算用数字と漢数字の使い分けを規定します。

1 文字のアルファベットを全角と半角のどちらで表記するか、助数詞で「か」、「カ」、「ヶ」のどれを使うかなども規定します。
例：L字型、3か月/3カ月/3ヶ月
3. 文字間のスペース
全角文字と半角文字の間に半角スペースを入れるかどうかを規定します。
例：「自社の△Web△サイトの制作」「100△円玉」
4. カタカナ複合語
カタカナ複合語の区切り方を規定します。「中黒」「半角スペース」「何も入れない」の3種類のいずれかが選択されます。
例：「ローカル・エリア・ネットワーク」、「ローカル△エリア△ネットワーク」、「ローカルエリアネットワーク」
5. 記号の表記と用途
句読点、各種記号、かっこ類を全角と半角のどちらで表記するかを規定します。
6. 単位の表記
JIS規格で定められた単位記号を使用するか、カタカナ表記を使用するかを規定します。
例：「kg」、「キログラム」
7. 正確に伝わる文章、簡潔な文章を書くためのルール
命令文の使い方、機能動詞の使用、回避すべき表現についての指定、主語についての指定、能動態と受動態の使い方、可能表現の使い方などのルールを規定します。
8. 特別な表記
コンピューターなどの画面用語、キーボードのキーの表記、操作手順の文体、参照の文体などを規定します。

※例文中の「△」は半角スペースを挿入する位置です。

## 表記仕様のばらつき

発注元(顧客)ごとに、翻訳用の「日本語スタイルガイド」が決められ、それぞれに対する準拠が要求される一方で、発注元によって用語や表記仕様が異なり、標準化されていないという現状があります。

JTF 標準スタイルガイド検討委員会では、協力企業 6社(大手 IT 企業)のスタイルガイドを項目別にまとめた一覧表を作成しました。

その一覧表から、表記仕様に明らかなばらつきの見られる項目を抜き出して示します(一覧表の作成後に表記仕様を変更した企業があります。表内の表記例は最新ではない場合があります)。

### 1. カタカナ語末尾の長音記号の有無

6社の表記仕様に従って、英単語 computer、user、memory の訳語を表記した例を示します。

- 「コンピュータ (一)」の表記例から、A社、C社、F社の表記仕様は、平成3年6月28日内閣告示第2号『外来語の表記』に基本的に準じており、一方、B社、D社、E社の表記仕様は、1951年制定、2011年最新改正のJIS規格の影響を受けていることがわかります。

- ただし、「メモリ (一)」のような文字数の少ない単語になると、JIS規格の影響を受けたB社、D社、E社3社の表記例にさらにばらつきが生じます。
- この6社はいずれも「ユーザー」と末尾に長音を付ける表記を採用していますが、「ユーザ」と末尾の長音を省く表記も一般技術文書で広く採用されています。

A社	コンピューター、ユーザー、メモリー
B社	コンピュータ、ユーザー、メモリ
C社	コンピューター、ユーザー、メモリー
D社	コンピュータ、ユーザー、メモリー
E社	コンピュータ、ユーザー、メモリー
F社	コンピューター、ユーザー、メモリー

### 2. カタカナ複合語の区切り方

6社の表記仕様に従って、英語句 local area network の訳語を表記した例を示します。

- A社とE社：半角スペース区切り
- C社とD社：中黒区切り
- B社とF社：区切りに何も入れない

A社	ローカル エリア ネットワーク
B社	ローカルエリアネットワーク
C社	ローカル・エリア・ネットワーク
D社	ローカル・エリア・ネットワーク
E社	ローカル エリア ネットワーク
F社	ローカルエリアネットワーク

### 3. Internet provider の訳語の表記

前述の2つの表記仕様を組み合わせた例(カタカナ語末尾の長音記号の有無×カタカナ複合語の区切り方)として、英語句 Internet provider の訳語を表記した例を示します。

6社6通りの表記になっており、一致する表記がありません。

A社	インターネット プロバイダー
B社	インターネットプロバイダ
C社	インターネット・プロバイダー
D社	インターネット・プロバイダ
E社	インターネット プロバイダ
F社	インターネットプロバイダー

企業間で表記仕様にばらつきの多い項目は、ほかに、全角文字と半角文字の間のスペースの有無、句読点・記号・かっこ類の全角/半角表記があります。

## 企業間で表記仕様が不統一である理由

企業間で表記仕様が標準化されていない理由として、文章の表記スタイルが「企業文化」に密接に関係するという考え方があること、表記ガイドラインが機密文書として扱われるために情報の交換が進まないこと、すでに大量に蓄積されている大量の文書の表記スタイルを今から変更できないことを挙げることができます。しかし、各企業が日本語表記の標準化に反対する立

場をとっているわけではありません。むしろ、表記の標準化そのものには賛成という声をよく聞きます。各社の本音は「うちは変えない。相手がうちに合わせて変えてくれるのはOK」かもしれません。

そうであれば、たとえば表記統一ツールを導入するなど、企業のニーズに適合する方法での、自動化による表記の統一が可能になれば、将来、企業間での表記の違いに伴う問題が解消される可能性があると思われる。

### 表記仕様の不統一がもたらす影響

和訳時の表記の不統一がどのような影響を及ぼすかを具体例で紹介します。

次の図は、TAUS Data Association (TDA、<http://www.tausdata.org/index.php>)の共有翻訳メモリープール内を検索できる Web ページのスクリーンショットです。ソース言語に「英語」、ターゲット言語に「日本語」を指定し、検索ボックスに「graphical user interface」と入力して検索すると、この小さな画面だけで6つの異なる表記が表れます。

TDAは、クラウド上で翻訳メモリーを共有するための国際的な試みで、プールされている翻訳メモリーはダウンロードして使用することができます。

しかし、次の例のように表記の異なる訳文が含まれていると、翻訳メモリーを簡単に再利用できません。用語管理にも問題が生じます。これは、ほかの言語には見られない、日本語ならではの問題ではないかと思われる。

最新の自然言語処理技術をもってすれば、カタカナ表記に多少のゆらぎがあっても、検索や解析などの処理を実行するのは難しくないとされます。

しかし、文字の種類が多い日本語の表記のバリエーションは意外なほど多く、日々新しい言葉が増えていきます。また、特に実務翻訳のようなビジネスの現場で

は、後処理にかかる手間をできるだけ省くことが重要です。

最初に、文書や翻訳メモリーを作成する段階で、表記や表現に一定の基準や制限を設けることで、作業プロセス全体の効率化を期待できます。その結果、表記のゆらぎの少ない質の良い翻訳メモリーが大量にできれば、機械翻訳用のコーパスとして利用することも可能です。

### 表記統一の真のメリットとは

表記とは、見た目の美しさや読みやすさだけの問題だろうかと考えます。たしかに、表記が不揃いであると、読みにくい、落ち着かない、という意見を耳にします。表記が揃っている文書は、たしかに気持ちがよく、読みやすい。

では表記が不揃いであると、誰が(who)、どのようなどきに(when)、どのように(how)困るのでしょうか。個々の考え方は耳にするのですが、あいまいで十分に整理されていません。また、翻訳業界全体で課題として共有されていません。

実務翻訳における表記統一の真のメリットは、日本語にかかわる言語資産の作成と活用にあります。表記のゆらぎを防ぐことは、良質の機械翻訳用コーパスの作成につながり、ドキュメントの検索性や解析性の向上による文書の管理および活用方法の改善につながります。

同時に、原文の情報を正確にわかりやすく伝える、質の高い翻訳を提供する翻訳サービスの実現にも寄与します。

上記の観点から、日本翻訳連盟の標準スタイルガイド検討委員会では、実務翻訳の現場で広く活用できる「日本語スタイルガイド(翻訳用)」を作成しています。また表記統一の真のメリットが、翻訳業界内で共有され、活かされるように努力しています。

## Japan and the Global Language Services Market

*by Donald A. DePalma, Ph.D.*

*Chief Strategy Officer and Founder of Common Sense Advisory, Inc.*

---

The Global Market for Language Services .....	19
Asia Represents a Growing Presence in Language Services.....	23
Japanese LSPs Dominate Asian Language Service Sector .....	26
Pricing for Japanese Language Services .....	27
The Japanese Language Services Market .....	28
Japanese LSPs Are Increasing Staff.....	29
On Business Age, Growth Rates, and Most Importantly, Revenue.....	30
Additional Findings on Japanese LSP Characteristics .....	31
 Figure 1: Regional Distribution of Language Service Providers in 2011.....	 21
 Table 1: Projected Language Services Revenues and Regional Distribution.....	 23
Table 2: Japanese LSPs among the Top 50 Global Language Service Providers in 2011.....	25
Table 3: The Top 15 Language Service Providers in Asia .....	26
Table 4: Price Distribution – English to Japanese and Japanese to English .....	28
Table 5: Employee Size Ranges of LSPs in Japan for 2009 and 2010.....	29
Table 6: Number of Office Locations of LSPs in Japan in 2009 and 2010 .....	29
Table 7: Business Age of LSPs in Japan in 2009 and 2010 .....	30
Table 8: Average Reported and Expected Revenue of Japanese LSPs (in U.S. Dollars).....	30
Table 9: Japanese LSP Growth Rates.....	31
Table 10: Revenue Breakdown for Japanese LSPs.....	32

### **The Global Market for Language Services**

---

The last few years have posed enormous challenges to companies in every market sector, including language services and technology. At Common Sense Advisory, our annual market sizing report has found that the majority of language industry participants experienced continued growth following a period of global economic uncertainty from 2008 to 2010.

However, others have expressed their fears and hesitations about the market. This article reviews the global market for language services and where Japanese companies fit.

In our most recent report on the market, we took a wide and deep view of the language services sector.<sup>1</sup> Our research team spent approximately six months on primary data collection. First, we updated the global directory of language service providers (LSPs) that we developed in 2010. Then we spent three months on survey recruitment, in order to survey as many suppliers in the market as possible. Finally, we processed the data and applied our proprietary algorithm in order to size the market and identify the largest players.

Without a doubt, we have found that the language services market is global in nature. The 25,256 LSPs that we identified were located in 152 countries. Our search led us to identify translation and interpreting providers in many places, including some that are not the “usual suspects” for producing LSPs, such as Swaziland, Tajikistan, and Yemen. However, we found the largest numbers of language service providers where one would expect them to be located – Europe, North America, and Asia (see Figure 1).

---

<sup>1</sup> “The Language Services Market: 2011,” May 2011. A synopsis is available at <http://www.common senseadvisory.com/research.aspx>.

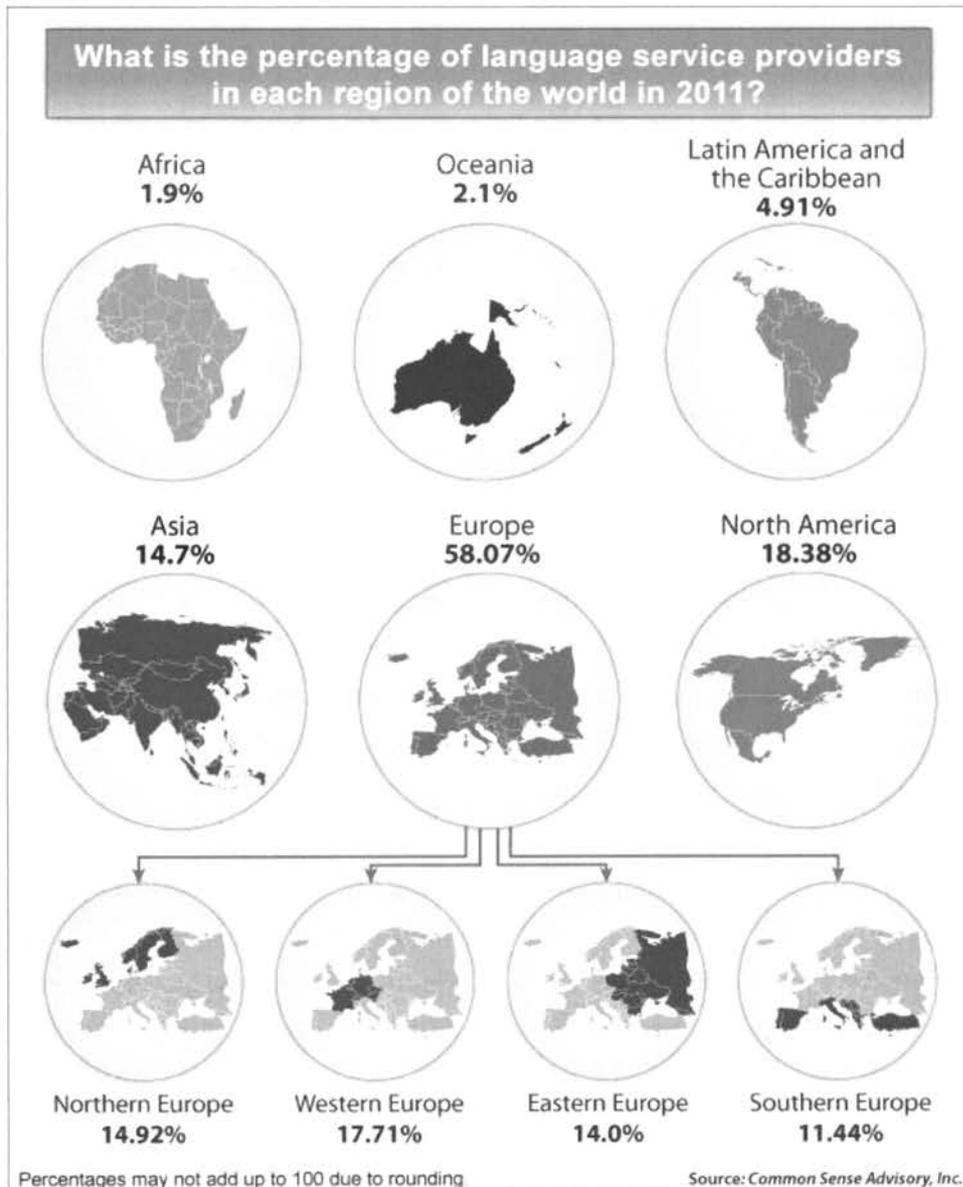


Figure 1: Regional Distribution of Language Service Providers in 2011

Source: Common Sense Advisory, Inc.

In 2010, Common Sense Advisory developed a proprietary market-sizing algorithm that includes 24 calculation areas.<sup>2</sup> Within each area, we used custom-programmed statistical software to carry out numerous computations within each category. To ensure consistency and allow for valid year-to-year comparisons from 2010 to 2011, we purposely made minimal coding updates to the algorithm itself.

<sup>2</sup> "Language Services Sizing Algorithm," May 2010. A synopsis is available at <http://www.common senseadvisory.com/research.aspx>.

We fed our algorithm three primary sets of data: 1) our directory of 25,256 LSPs, containing information about employee size ranges, country of origin, and other demographic details; 2) data from 912 unique companies that filled out our survey; and 3) a subset of the survey responses, which included 622 companies that provided us with exact revenue data for 2009 and 2010. For each of the firms reporting revenue in currencies other than U.S. dollars, we converted the firm's revenue according to the average daily exchange rate for its fiscal year. Currency conversion is one of the most time-consuming and tedious steps in our process, but an extremely important one for data homogeneity.

To obtain our global market size estimate, we identified the LSP size ranges in the total population. We took the dataset from our survey for the companies that provided revenue data for 2010, normalized in U.S. dollars using the average daily exchange rate for the fiscal year of each company. We then calculated the average revenue for firms in each size range to come up with total estimated revenue for all companies in each size group. Then we computed the percentage of sales to other suppliers in each size category, and we reduced the total in each size range by the amounts sold to other LSPs.

To calculate the rate at which the underlying market actually grew from 2009 to 2010, we measured the exact change in revenue for all companies who provide us with actual figures. For those companies that did not provide us with exact revenue, we used the revenue range midpoints. To determine the anticipated market growth rate for the next five years, we took the average of the 2009-2010 revenue ranges and the 2010-2011 expected growth rate. Finally, we calculated the estimated market share by region by computing the total LSP revenue contributed by each region as a percentage of total market size.

This year, the higher number of LSPs reporting actual revenue data enables us to reveal even more detailed information about the market itself, including its regional and sector-specific breakdowns. For 2011, Common Sense Advisory calculates that the market for outsourced language services is worth US\$31.438 billion in 2011.

Our analysis shows that the language services and technology market is growing at a slower pace than previously thought. In May 2010, we estimated that the underlying market was growing at a rate of more than 13%. This year, our study suggests that the actual market growth rate is nearly half that amount. It's important to remember that the growth rate for the market itself is not necessarily a good benchmark for companies to use as a basis for comparison with their own growth. Instead, the growth rate of 7.41% reflects the growth of the market at large – the

rates of growth for individual companies that are proactively going out into the market with a goal of increasing revenue are typically much higher.

While the overall market growth rate is 7.41%, we should note that this calculation takes into account three years of data – the current year, the previous year, and the projected year. Our model reduces the impact of reselling among language service providers, thus counting revenue just once. Therefore, actual growth rates for an individual firm frequently exceed our market growth rates because the calculations themselves are quite different.

To provide a broader view, what these two findings mean is that our five-year projections as published last year are nearly identical to this year’s projections, with the exception that the market is currently larger than previously anticipated, but remains on track to reach a similar size within the timeframe we provided earlier (see Table 1).

Region	Market Share	2010 US\$ M	2011 US\$ M	2012 US\$ M	2013 US\$ M	2014 US\$ M
North America	49.25%	14,415	15,483	16,631	17,864	19,188
Western Europe	21.13%	6,186	6,644	7,137	7,666	8,234
Northern Europe	12.71%	3,720	3,995	4,292	4,610	4,951
Asia	7.43%	2,175	2,336	2,509	2,695	2,895
Southern Europe	5.39%	1,577	1,694	1,820	1,955	2,100
Eastern Europe	2.84%	832	894	960	1,031	1,107
Oceania	0.66%	192	207	222	238	256
Latin America	0.32%	95	102	110	118	126
Africa	0.26%	77	83	89	96	103
Growth Totals	100.00%	29,268	31,438	33,768	36,271	38,960

*Table 1: Projected Language Services Revenues and Regional Distribution*

*Source: Common Sense Advisory, Inc.*

## **Asia Represents a Growing Presence in Language Services**

As Figure 1 shows, 14.7% of the world’s LSPs operate out of Asia. Of these thousands of providers, six Japanese language service providers appeared in our list of the 50 largest companies, as measured by revenue (see Table 2). Two of these, Honyaku Center and Ishida

Taiseisha Inc., exceeded the general market growth of 7.41% -- Honyaku grew 23.3% over 2010, while Ishida Taiseisha rose 7.8%.

	Company	HQ Country	2010 Revenue in US\$M	Employees	Offices	Status
1	Mission Essential Personnel	US	588.00	7494	20	Private
2	HP ACG	FR	460.00	4200	15	Public
3	Global Linguist Solutions	US	435.00	339	9	Private
4	Lionbridge Technologies	US	405.20	4500	40	Public
5	TransPerfect/Translations.com	US	252.44	1268	66	Private
6	SDL	UK	245.09	1800	60	Public
7	L-3 Linguist Operations & Technical Support	US	167.00	1016	1	Public
8	STAR Group	CH	145.00	850	43	Private
9	euroscript International S.A.	LU	124.13	1350	16	Private
10	ManpowerGroup	US	101.00	324	11	Public
11	RWS Holdings PLC	UK	96.92	503	12	Public
12	Yamagata Intech Corporation	JP	68.45	800	13	Private
13	CLS Communication	CH	67.78	550	19	Private
14	thebigword Group	UK	60.55	380	9	Private
15	Honyaku Center Inc.	JP	56.36	225	8	Public
16	Logos Group	IT	51.50	170	10	Private
17	Semantix AB	SE	48.51	202	10	Private
18	Welocalize, Inc.	US	44.71	442	10	Private
19	Moravia Worldwide	CZ	42.70	458	10	Private
20	hiSoft Technology International Ltd.	ZH	38.50	930	16	Public
21	Cyacom International, Inc.	US	37.40	442	4	Private
22	Invizion	US	33.00	290	1	Private
23	Språkservice Sverige AB	SE	32.30	40	2	Private
24	Crestec, Inc.	JP	30.88	360	15	Private
25	AAC Global	FI	30.54	238	14	Public

26	Merrill Brink International	US	30.02	100	10	Private
27	Pacific Interpreters	US	30.00	245	5	Private
28	ONCALL Interpreters & Translators	AU	28.39	85	5	Private
29	Language Services Associates	US	27.00	153	5	Private
30	CBG Konsult AB	SE	25.04	190	9	Private
31	Skrivanek Group	CZ	24.80	320	44	Private
32	CSOFT International	ZH	24.80	260	14	Private
33	Jonckers Translation & Engineering s.a.	BE	24.17	150	9	Private
34	Ishida Taiseisha Inc.	JP	24.10	1650	27	Private
35	PTSGI	TW	22.28	300	4	Private
36	Telelingua	BE	21.75	200	6	Private
37	Alpha CRC	UK	21.65	265	12	Private
38	Concorde Group	NL	21.25	75	2	Private
39	SeproTec Multilingual Solutions	ES	20.06	404	16	Private
40	LanguageWire	DK	19.11	75	7	Private
41	Verztec Consulting PTE Ltd.	SG	16.75	136	7	Private
43	Sajan	US	16.00	124	5	Public
44	Amesto Translations	NO	15.59	75	4	Private
45	Transvoice Sweden AB	SE	15.31	65	3	Private
46	TRANSN Co., Ltd.	ZH	14.76	360	6	Private
47	TOIN Corporation	JP	14.04	100	4	Private
48	Chizai Corporation	JP	13.77	75	3	Private
49	TextMinded, EEIG	BE	13.41	78	7	Private
50	Xplanation Language Services NV	BE	12.55	67	5	Private

Table 2: Japanese LSPs among the Top 50 Global Language Service Providers in 2011

Source: Common Sense Advisory, Inc.

## Japanese LSPs Dominate Asian Language Service Sector

Japanese LSPs dominate the list of the largest Asia service providers, taking nine of the top 15 positions (see Table 3). Japan has traditionally been an important region of the world for language services, and this year's ranking makes that fact clear. The Japanese earthquake and tsunami did not prevent most Japanese market leaders from participating in our survey – however, it's important to remember that the revenue listed here is for the 2010 fiscal year. If there is any financial impact from the natural disasters the country has experienced, it would not be reflected until next year. Still, it's quite impressive that Japanese companies make up nearly two-thirds of the companies on the regional ranking for Asia

	Company	HQ Country	2010 Revenue in US\$M	Employees	Offices	Status
1	Yamagata Intech Corporation	JP	68.45	800	13	Private
2	Honyaku Center Inc.	JP	56.36	225	8	Public
3	hiSoft Technology International Ltd.	ZH	38.50	930	16	Public
4	Crestec, Inc.	JP	30.88	360	15	Private
5	CSOFT International	ZH	24.80	260	14	Private
6	Ishida Taiseisha Inc.	JP	24.10	1650	27	Private
7	PTSGI	TW	22.28	300	4	Private
8	Verztec Consulting PTE LTD	SG	16.75	136	7	Private
9	TRANSN Co., Ltd.	ZH	14.76	360	6	Private
10	TOIN Corporation	JP	14.04	100	4	Private
11	Chizai Corporation	JP	13.77	75	3	Private
12	Human Science Co., Ltd.	JP	9.52	95	1	Private
13	WIP Japan Corporation	JP	9.21	31	5	Private
14	Lyric Labs	IN	3.68	46	4	Private
15	ida Corporation	JP	3.63	35	2	Private

Table 3: The Top 15 Language Service Providers in Asia

Source: Common Sense Advisory, Inc.

Of course, China is gaining ground, with companies like hiSoft, CSOFT, and TRANSN bringing in significant amounts of revenue. We also see entries from PTSGI (Taiwan) and Verztec Consulting (Singapore) on this year’s regional list. For the first time, our regional ranking also includes a company from India (Lyric Labs). We first mentioned Lyric Labs in our ranking of telephone interpreting providers.<sup>3</sup> We expect to see more companies from India joining the Asian ranking in future years.

## Pricing for Japanese Language Services

In a pricing study that Common Sense Advisory conducted in 2010, we surveyed 651 language service providers and freelancers from 73 countries on the prices for their most popular language pairs, yielding responses for 156 source-target combinations, including Japanese<>English (see Table 4).<sup>4</sup> We recognize the many differences between the pricing and practices of a freelance translator and a language services behemoth, so our analysis factors in company size and, in some cases, the location of their headquarters.

Table 4 includes prices by company size wherever we had two or more responses and the average of all responses (All). The table contains the number of responses for that language pair (Response Count). Finally, each row includes the range of prices normalized into U.S. dollars charged by companies of that size (Price Min for the minimum or bottom of the range, and Price Max for the maximum), the average price (Price Avg), and the standard deviation (Price Std). The highlighted cells in Table 4 are for the average prices across each group (Price Avg) and across all respondents (All)

Language Pair	Company Size	Response Count	Price Min	Price Max	Price Avg	Price Std
English to Japanese	Freelancer	6	0.13	0.26	0.21	0.05
	Small	74	0.08	0.58	0.27	0.09

<sup>3</sup> “Top 15 Telephone Interpreting Suppliers,” September 2009. A synopsis is available at <http://www.commonseadvisory.com/research.aspx>.

<sup>4</sup> “Translation and Localization Pricing,” July 2010. A synopsis is available at <http://www.commonseadvisory.com/research.aspx>.

Language Pair	Company Size	Response Count	Price Min	Price Max	Price Avg	Price Std
	Medium	52	0.13	0.45	0.25	0.07
	Large	28	0.15	0.49	0.28	0.07
	All	160	0.08	0.58	0.26	0.08
Japanese to English	Freelancer	16	0.03	0.24	0.11	0.05
	Small	22	0.05	0.55	0.27	0.15
	Medium	14	0.07	0.40	0.20	0.09
	Large	4	0.13	0.85	0.41	0.34
	All	56	0.03	0.85	0.22	0.16

*Table 4: Price Distribution – English to Japanese and Japanese to English*  
*Source: Common Sense Advisory, Inc.*

## The Japanese Language Services Market

Let's look at the Japanese supplier market for translation and interpreting services. Our research reviews the typical service mix, rate of growth, and average revenue per employee based on responses to our annual survey.<sup>5</sup> Our annual study of the language services market identified 600 translation and interpreting services companies in Japan (see "The Language Services Market: 2011," May11).

We received responses from 17 of these firms, who provided details about their reported revenue and other business characteristics for the previous fiscal year (2010). This is a very small sample, in which each respondent's data could have a significant impact on the overall results. However, these results paint a picture of the Japanese language services suppliers. In future years, we hope that the response rate will be much higher, thus providing a solid data-driven view of this market.

<sup>5</sup> "The Japanese Language Services Market," October 2011. A synopsis is available at <http://www.commonsenseadvisory.com/research.aspx>.

## Japanese LSPs Are Increasing Staff

Based on the data we collected, we see that the employee size range of Japanese LSPs increased this year as compared with the previous year. In 2010, more than a third of the respondents told us that they had fewer than 10 full-time employees (FTEs), whereas this year, only about 17% of respondents fell into that category (see Table 5). In 2009, only 15.4% of Japanese suppliers indicated they had more than 501 FTEs, whereas in 2010, this number increased by more than 50%. The average number of FTEs per Japanese LSP also increased: In 2010 it was 86.34; the previous year it was 73.75 full-time employees.

Number of Full-Time Employees (FTEs)	in 2009	in 2010	Year-to-Year Change
501 or more	15.4%	23.5%	52.6%
201 to 250	15.4%	5.9%	-61.7%
31 to 50	7.7%	11.8%	53.2%
11 to 15	15.4%	5.9%	-61.7%
6 to 10	23.1%	11.8%	-48.9%
2 to 5	7.7%	5.9%	-23.4%

*Table 5: Employee Size Ranges of LSPs in Japan for 2009 and 2010*

*Source: Common Sense Advisory, Inc.*

We also noted an increase in the number of LSPs that operated from five or more offices (see Table 6).

Number of Offices	in 2009	in 2010	Year-to-Year Change
1	30.8%	29.4%	-4.5%
2	30.8%	11.8%	-61.7%
3	7.7%	11.8%	53.2%
5 or more	30.8%	35.3%	14.6%

*Table 6: Number of Office Locations of LSPs in Japan in 2009 and 2010*

*Source: Common Sense Advisory, Inc.*

## On Business Age, Growth Rates, and Most Importantly, Revenue

About 35% of Japanese LSPs responding to this year's survey told us that they were more than 30 years old, which more or less corresponds with last year's data for the same age range. About 12% of start-ups figured in our survey (see Table 7).

Years in Business	in 2009	in 2010	Year-to-Year Change
1 to 2 years	7.7%	11.8%	53.2%
3 to 5 years	15.4%	5.9%	-61.7%
11 to 15 years	7.7%	11.8%	53.2%
16 to 20 years	7.7%	11.8%	53.2%
21 to 30 years	30.8%	17.6%	-42.8%
More than 30 years	30.8%	35.3%	14.6%

Table 7: Business Age of LSPs in Japan in 2009 and 2010

Source: Common Sense Advisory, Inc.

The average revenue per LSP in Japan reported for 2009 was US\$13,339,286. The number increased, but only slightly, to US\$13,857,143 in 2010, but this exceeded Japanese LSPs' estimation of US\$13,681,818, as reported to us in 2010. For 2011, Japanese suppliers estimate a marginal decrease in their revenue (see Table 8).

Reported Revenue 2009 (US\$)	Reported Revenue 2010 (US\$)	Estimated Revenue 2011 (US\$)	Revenue Average (2009-2011) (US\$)
13,339,286	13,857,143	13,321,429	13,505,953

Table 8: Average Reported and Expected Revenue of Japanese LSPs (in U.S. Dollars)

Source: Common Sense Advisory, Inc.

The average rate of growth for Japanese LSPs from 2009 to 2010 was 10.41%, which is more than a 100% drop from the previous year's growth rate (see Table 9). However, continuing their roller-coaster ride, Japanese language companies think they will grow at a rate of 118.33% from 2010 to 2011. By comparison, the overall market is currently growing at a rate of 7.41%. This change might be explained by the composition of the respondents, with some larger players that skewed the average higher than last year.

Japanese LSP Reported Growth Rate 2008 to 2009	Japanese LSP Reported Growth Rate 2009 to 2010	Japanese LSP Expected Growth Rate 2010 to 2011
24.59%	10.41%	118.33%

Table 9: Japanese LSP Growth Rates

Source: Common Sense Advisory, Inc.

## Additional Findings on Japanese LSP Characteristics

We observed several other interesting characteristics of Japanese LSPs:

- Average revenue per employee.** Companies often ask us how much they should be earning for each person they employ. In Japan, that amount is US\$30,026 – based on 2010 revenue and the number of employees reported to us for the same fiscal year. Last year, the expected earnings from each FTE for Japanese LSPs stood at US\$21,103. On a global scale, average revenue per employee for LSPs stands at US\$79,667. Again, the participation of larger suppliers may have affected this year’s results.
- Service mix.** Translation continued to be the main revenue generator, accounting for nearly 66% of the revenue earned by language service companies in Japan (see Table 10). The other double-digit revenue-earning service was software localization, at nearly 11%. However, both translation and software localization had contributed slightly more to the revenue in 2009. Website globalization increased nearly 90%, perhaps the result of increased activity by Japanese companies in promoting their goods in foreign markets. The laggards in Japan were telephone interpreting, video interpreting, interpreting tools/software, and business process outsourcing – all of which suffered a 90% to 100% drop in revenue.
- Re-selling.** Nearly 10% of Japanese LSPs’ sales were to other LSPs, according to our survey data. Last year, re-selling accounted for 20.59% of sales.

Service	Percentage of Total Revenue in 2009	Percentage of Total Revenue in 2010	Percent Change from 2009 to 2010
Translation	69.50%	65.95%	-5.10%
Software localization	11.44%	10.86%	-5.07%
Multimedia localization	6.38%	3.30%	-48.34%
Website globalization	3.88%	7.35%	89.00%
Translation tools/software	1.55%	0.60%	-61.42%
Transcreation	2.12%	0.90%	-57.64%
On-site interpreting	8.00%	14.25%	78.12%
Telephone interpreting	5.15%	0.33%	-93.52%
Video interpreting	0.80%	0.00%	-100.00%
Interpreting tools/software	0.30%	0.00%	-100.00%
MT post-editing	1.00%	0.81%	-18.18%
Subtitling	1.00%	0.66%	-33.33%
Voice-over/dubbing/narration	3.00%	2.22%	-25.92%
International testing/QA	2.66%	1.77%	-33.33%
Internationalization services	3.22%	3.50%	8.62%
Business process outsourcing	7.94%	1.44%	-81.81%

*Table 10: Revenue Breakdown for Japanese LSPs*

*Source: Common Sense Advisory, Inc.*

The Japanese economy held out for some time against the global recession, but felt the impact in late 2008 because of a sharp decline in business investment and global demand for Japan's exports. However, in 2010, while the economy grew at a rate of 3.9%, Japanese language services providers grew at a rate of 10.41%.

Year after year, we find that more companies are willing to share their business data with us for the purposes of estimating the market size. For our 2012 report on the global language services and technology market, we hope to see many more of Japan's LSPs and language software vendors respond to our survey. It's through such collaboration that we can estimate the size of the market and determine where it's growing – so that we can write articles like this to inform the participants about their industry.



## Going Hybrid: Pangeanic's and Toshiba's First Steps Toward ENJP MT Hybridization

### ハイブリッドへの道：英⇄日ハイブリッド機械翻訳に対する Pangeanic—東芝の取り組み

Elia Yuste, Manuel Herranz & Alexandre Helle  
*Pangeanic / B.I Europa*  
*PangeaMT Technologies Division*  
*Valencia, Spain*

Hirokazu Suzuki  
*Knowledge Media Laboratory*  
*Corporate Research & Development Center*  
*Toshiba Corporation, Tokyo, Japan*

---

#### 1. Introduction on our respective corporate history and business model

##### 1.1. Toshiba's history on machine translation

Toshiba group has been continuing researching and developing on machine translation (MT) technology over 30 years. Since 1985 when Toshiba put MT system to practical use, Toshiba group has been commercializing 'The Honyaku series', which are English-Japanese/Japanese-English translation software on PC, and 'The Honyaku server series', which are server/client-based MT systems for enterprises. As a part of the research and development, Toshiba has realized the system which alleviates users' burden for their customizing translational environment to improve translation quality by their user dictionary, translation memory, example-based MT, corpus-based word selection and so on. In 2009, Toshiba group received 'AAMT Nagao award' for these achievements<sup>1</sup>.

Toshiba's MT, as well as other commercial MT software in Japan, is rule-based MT (RBMT) system. This is good for reliable MT customizations through adding/modifying translation rules but might not meet current expectations of rapid and flexible ones. Therefore, the statistical method is a natural approach for expanding possibility of RBMT.

##### 1.2. Pangeanic's transformation from a translation agency into a global LSP, statistical machine translation solution customizer and MT customization enabler

Pangeanic/B.I Europa (Pangeanic for short, associate member of B.I Corporation) is a Spanish language service provider (LSP), which has been committed to providing high-quality, speedy, and cost-effective documentation translation solutions for more than 20 years. That is why Pangeanic is a no. 1 choice for major global players operating worldwide, in electronics, automotive, power generation, tool machines, software, medicine, biotechnology, and many other industries. With partner offices in Japan and China, Pangeanic can provide point-to-point professional language solutions to help any business become and stay competitive in the global marketplace. The entire organization is based on a simple mantra: "genuinely be close to the customer". This is not an empty statement. It led Pangeanic to join the world MT think-tank TAUS<sup>2</sup> in 2005, becoming one of the TAUS Data Association (TDA) founding members. Data sharing programs could only benefit all the company's

---

1 [www.toshiba.co.jp/about/press/2009\\_06/pr\\_j2201.htm](http://www.toshiba.co.jp/about/press/2009_06/pr_j2201.htm)

2 [www.translationautomation.org](http://www.translationautomation.org)

customers, current and future. The more data available, the better processes and human language technologies (HLT) could be developed with domain representativeness, speed and cost-effectiveness in mind. Research and development have remained essential ever since: a large percentage of Pangeanic's annual turnover is re-invested in language technology R&D applications and training programs.

After deploying and evaluating a number of commercial MT systems, Pangeanic created a Moses<sup>3</sup>-based statistical machine translation (SMT) system called Pangematic in late 2008. This was originally conceived for the LSP's own internal translation production needs – a development motivation that has remained since then and has made Pangeanic transform Pangematic, renamed PangeaMT in early 2010, into a truly customizable MT commercial offering tested and deployed by other language service providers and corporations in need of flexible, open-standards geared, cost-effective MT solutions. PangeaMT philosophy is quite different from standard MT historical players. Pangeanic's PangeaMT division team, in collaboration with Valencia Polytechnic University's ITI, has built several modules and pre- / post-processing scripts on top of Moses, as well as interfaces and retraining procedures in order to develop an industry-tested alternative<sup>4</sup> to off-the-shelf, more rigid MT products.

With a view to fostering even more accessibility to flexible and customizable MT solutions as well as further independence from MT providers for engine experimentation and retraining, Pangeanic officially released the **PangeaMT SMT DIY** Solution for E-FIGS<sup>5</sup> right before the *Localization World Conference* in Barcelona in June 2011. Its main asset consists of a fuzz-free automatic engine training routine, so there is no need to knock on the MT provider's door any more in order to pay for engine updating.

PangeaMT SMT DIY comes with a training data set structure where users simply leave their bilingual aligned files in TMX 1.4b format, and engine training and updating is done automatically, either on a time basis or incrementally (after 50Mb, 100Mb, 300Mb of additional data are detected, or even 1Kb!). This is paired with a Control Panel that shows engine listing, status and automatic BLEU quality metric percentage by engine's language direction, and enables unlimited users' creation and activity tracking, etc. As far as MT requesting and user-application interaction are concerned, there is an intuitive Google-like web-based translation panel and, most interesting, a %MT file translation feature whereby pre-translated XLIFF or SDL Trados TTX files will be translated after a certain % match has been leveraged<sup>6</sup> from the translation memory (TM) by the user. This is fully customizable by project, so one chooses exactly at what pre-translation level MT will come in, not touching any material leveraged from the TM. Other useful feature is the Glossary TXT file upload that fixes terminology and preferred expressions, such as *untranslatables* or DNTs<sup>7</sup>, in the background before MT output (quite a historic request to SMT developers).

---

3 [www.statmt.org/moses](http://www.statmt.org/moses)

4 See Yuste et al. (2010) for more information.

5 English into-from French, Italian, German, and Spanish. Offering this language bundle, some combinations thereof or other languages as per client's specifications (e.g. Brazilian/Iberian Portuguese, Nordic languages, etc.).

6 PangeaMT's TMX / XLIFF /TTX parsers facilitate the user's preference for any CAT tool as post-editing environment (a TMX or XLIFF editor can be used, and even Tag Editor, which will identify pre-translations as coming from MT! and stop at each segment. This allows users to leverage their TMs as well.

7 DNT is a localization industry acronym that stands for “do-not-translate” a given word or expression, usually due to product marketing and corporate identity related issues. A typical example is the brand name Apple. For a computational linguist, a DNT is somewhat similar to the concept of *named entity*.

Pangeanic also offers Data Consultancy and MT output post-editing (PE) services to clients, whose existing bilingual data may be unclear or scarce for engine training purposes and who are not familiar enough with PE procedures or lack PE staff, respectively. The PangeaMT team is keen to offer support to DIY customers in need of language-specific language and MT modelling and willing to keep them abreast with any PangeaMT-related future advancements. However, PangeaMT, in releasing the DIY solution, sets apart from other providers that impose word-based or user number limitations to their clients. Instead, PangeaMT SMT DIY users can make the most out of their domain- or client-specific bi-texts by mixing and experimenting with these data sets for unlimited in-domain or mixed domain SMT engines (e.g. safety/insurance and automotive, software and life sciences, etc.) and automatic engine retraining or updating.

## **2. Joint endeavor's beginning and rationale**

Toshiba Corporation's R&D MT scientific lead, Hirokazu Suzuki, and Pangeanic's CEO and main PangeaMT's evangelizer, Manuel Herranz, first met at TAUS User Conference edition in Tokyo in Spring 2010. A PangeaMT success real-world implementation was first being presented for an international but Asia-focused audience. Manuel Herranz delivered a rather more MT awareness-raising presentation at the Japan Translation Company festival of December'10 in Tokyo. These two translation industry events allowed for thorough discussions among both companies about how they could join efforts in MT hybridization.

While Toshiba was keen to get further acquainted with a more flexible and rapid system development entailed in an SMT paradigm and to explore partnership with a European MT customizer, Pangeanic wanted to rely on a partner with outstanding MT and language resources creation expertise to season their PangeaMT solution with linguistic information in case of implementations involving linguistically unrelated languages. The English-Japanese language pair was of particular interest for both companies for obvious reasons. At the time of writing this paper, both parties have already decided to meet up in Tokyo this November again. Besides, Suzuki and Yuste plan to present further progress of this work at the next edition of TAUS Focus on Asia, which will hopefully be held in Tokyo in Spring 2012.

## **3. Ongoing collaboration description**

MT hybridization, initially for the ENJP language pair in both directions, was the goal of this collaboration from day one: How to combine the power of statistics with the potential help of linguistic rules, and to determine whether this combination rendered positive indicators to justify and sustain the joint effort.

The first action items to agree upon were how to handle and process the significant amount of bilingual aligned ENJP corpora that the parties had at their disposal through Pangeanic's TAUS membership and what hybridization technique to devise and implement. Firstly, Toshiba would have to pre-process the data meaningfully, adding Japanese-related linguistic information, which would in theory lead to render better PangeaMT output. The hypothesis was formulated around the principle of nipponization in the pre-processing stage. In other words, if the EN text strings of the translation units (TUs) making up the engine training corpora could be arranged in a JP-like word order (WO), the system would probabilistically infer and produce output with a more accurate Japanese-like word order.

### 3.1. Experimentation methodology and initial findings

#### 3.1.1. PangeaMT's preliminary work with Japanese

As already mentioned, Moses lies at the core of PangeaMT. The PangeaMT team carried out prior and in parallel to this collaboration with Toshiba some experimentation<sup>8</sup> with the freely available KFTT<sup>9</sup> corpus. This was aimed at comparing different Japanese segmenters and analyzing the impact thereof on Moses. Soon it was found that:

[a] Moses yielded no results when no JP segmentation took place in the pre-processing stage, making this step therefore compulsory for future development work involving Japanese;

[b] while the *KyTea*<sup>10</sup> segmenter produced better results for the EN into JP translation, *Mecab*<sup>11</sup> led to better ones in the opposite direction. Nevertheless, since the results were pretty close, they would be examined under the statistical significance technique Paired difference interval (see Zhang 2004). This showed that the intervals were not significantly different in the JP to EN translation.

[c] After this preliminary work, the creation of language models (LMs) to obtain the probabilities with at least 5 grams using the library Ken-LM<sup>12</sup> was recommended, plus the *-unk* option<sup>13</sup>.

These preliminary work findings provided some guidance as to how to proceed with the data linguistically processed by Toshiba.

#### 3.1.2. Focus on the PangeaMT (Pangeanic)-Toshiba Tandem

In this section we focus on the actual experimentation carried out jointly by Pangeanic and Toshiba. As already said, we primarily set out to find out whether nipponization of the English, i.e. technique to reorder the English component of the training data in a Japanese-like WO, would lead to better MT output. To prevent Moses from reordering translations in the usual way, it was necessary to arrange for monotone translation (no reordering). The PangeaMT team would also take this opportunity to compare the usefulness of segmentation processing with Mecab and Toshiba's proprietary segmenter.

The training data set, which is kept in TMX 1.4b format, contains the following 5 partitions:

- Original English (*en-US\_original*)
- Nipponized English as processed by Toshiba (*en-US\_nipponized*)
- Original Japanese (this will not be used for our MT experimentation directly)
- Japanese as segmented by Toshiba (*ja-JP\_Tos-Seg*)
- Japanese as segmented by Toshiba plus katakana character normalization (*ja-JP\_Tos-Seg-Norm*)

---

8 Please contact Alexandre Helle for detailed information on this part of the experimentation.

9 [www.phontron.com/kfft](http://www.phontron.com/kfft)

10 [www.phontron.com/kytea](http://www.phontron.com/kytea)

11 [mecab.sourceforge.net](http://mecab.sourceforge.net)

12 [kheafield.com/code/kenlm/](http://kheafield.com/code/kenlm/) It proved to work slightly better under the latest Moses subversion release of 2011-04-28

13 This obliges the system to take into account unknown words as normal ones.

Each Translation Unit (TU) is now five-fold and structures data in the following fashion:

```

<tuv xml:lang="en-US" datatype="original">
<seg>Checked for broken pins.</seg>
</tuv>
<tuv xml:lang="en-US" datatype="nipponized">
<seg>broken pins for Checked .</seg>
</tuv>
<tuv xml:lang="ja-JP" datatype="original">
<seg>曲がったり折れたりしているピンがないかどうかを調べました。</seg>
</tuv>
<tuv xml:lang="ja-JP" datatype="TOS Seg">
<seg>曲がったり折れたりしているピンがないかどうかを調べました。</seg>
</tuv>
<tuv xml:lang="ja-JP" datatype="TOS Seg+Normalize">
<seg>曲がったり折れたりしているピンがないかどうかを調べました。</seg>
</tuv>
</tu>

```

The performed pre-processing consists of these steps:

- Extract corpus text, which results in five training files aligned at segment level.
- To handle our experiments fully, a sixth corpus is created out of the original Japanese text that is now segmented using *Mecab* (named as *ja-JP\_Mecab-Seg*).
- Merge the six training data in one (making use of separators), remove repeated sentences and shuffle.
- Split the corpus into 2,000 sentences for *Tuning*, another 2,000 for *Testing* and the remaining for *Training*.
- Split the latter yet again into 6 training sets.
- Finally, complete the pre-processing stage by carrying out the typical pre-processing for the English component.

No further Japanese pre-processing will be carried out as we aim to keep the segmentation.

	Sentences	Running Words	Vocabulary size	Out Of Vocabulary (OOV)	Perplexity (5-gram)
<b>Original English</b>	<b>Training</b>	5291144	82363810	423439	
	<b>Tuning</b>	2000	31221	4863	
	<b>Testing</b>	2000	30465	4827	0
<b>Original Japanese</b>	<b>Training</b>	5291099	16235690	4805975	
	<b>Tuning</b>	2000	6198	4425	
	<b>Testing</b>	2000	6162	4375	0
<b>Segmented JP → Mecab</b>	<b>Training</b>	5291099	111834935	186469	
	<b>Tuning</b>	2000	42272	3396	
	<b>Testing</b>	2000	41218	3360	0
<b>Segmented JP → Toshiba</b>	<b>Training</b>	5291144	66412862	113920	
	<b>Tuning</b>	2000	25018	2895	
	<b>Testing</b>	2000	24473	2898	0
<b>Segmented JP → Toshiba, with normalization</b>	<b>Training</b>	5291144	66412862	113959	
	<b>Tuning</b>	2000	25018	2899	
	<b>Testing</b>	2000	24473	2898	0

Statistical information on training data sets for joint experimentation

Six translation experiments were then carried out between the pairs below:

- en-US\_nipponized.ja↔JP\_Mecab-Seg
- en-US\_nipponized.ja↔JP\_Tos-Seg
- en-US\_nipponized.ja↔JP\_Tos-Seg-Norm
- en-US\_nipponized.ja↔JP\_Mecab-Seg
- en-US\_nipponized.ja↔JP\_Tos-Seg
- en-US\_nipponized.ja↔JP\_Tos-Seg-Norm

No translation between English, original or nipponized, and original Japanese was attempted, as it was known to us thanks to previous experimentation work (see section 3.1.1) that a Moses-based system such as PangeaMT would not function correctly if the Japanese data was not segmented. All the above translation experiments were performed on the basis of 5-gram language modelling and using the -unk option, as these were the parameters that yielded better results in the other experimentation:

```
$ ngram-count -order 5 -interpolate -kndiscount -unk -text xxxxxx -lm xxxxxx.lm.arpa
```

#### 4. Conclusion

After carrying out the experiments (see next chart for further details), one of the first conclusions is that the segmentation performed by Toshiba has produced less favorable results than the segmentation pre-processing with Mecab. This is likely caused by the aspect of Toshiba's segmenter which tends to segment sentence into larger parts than Mecab, but this will surely motivate us to investigate further about own segmentation-related improvements in the future.

As far as nipponization is concerned, we are pleased to say that English nipponization is really useful, particularly for the EN into JP direction.

Finally, it can also be seen that the katakana character normalization has not provided significantly better results.

Exp. No.	Experiment Specification	BLEU		TER	
		JP>EN	EN>JP	JP>EN	EN>JP
1	en-US_nipponized.ja↔JP_Mecab-Seg	<b>28.07</b>	<b>44.02</b>	0.563	0.405
2	en-US_nipponized.ja↔JP_Tos-Seg	27.34	34.50	0.593	0.501
3	en-US_nipponized.ja↔JP_Tos-Seg-Norm	27.30	34.79	0.594	0.501
4	en-US_original.ja↔JP_Mecab-Seg	<b>25.11</b>	<b>36.13</b>	0.671	0.495
5	en-US_original.ja↔JP_Tos-Seg	25.96	28.00	0.682	0.584
6	en-US_original.ja↔JP_Tos-Seg-Norm	25.70	27.79	0.683	581

*Experiments' results – most significant figures in boldface*

## 5. Future work

Prior to full commercialization of our joint MT system, there are a number of research-gearred implementation avenues that the Pangeanic-Toshiba tandem is willing to explore. More linguistic knowledge can be introduced in the models and the evaluation parameters. Syntax modeling is also another area of interest.

Phrase-based SMT translators, where one input sentence corresponds to an output sentence, are the current state of the art. Yet tree-based models, with inclusion of linguistic rules, are gathering momentum. Tree-based engines can in fact model more complex instances, such as  $ne..pas \rightarrow not$ ) and improve results in very distinct languages by making use of medium-range reorderings. Typically, their cons are speed sacrifice and larger RAM and HDD requirements. The Moses toolkit is compatible with tree-based models since 2009 while other toolkits, such as *Jane*<sup>14</sup> (RWTH Aachen) and *Joshua*<sup>15</sup> (John Hopkins University) are just compatible with tree-based models. Comparing phrase-based models paired with the reordering procedure at the pre-processing level that we have named nipponization when applicable to the English data component, with tree-based models under a Moses framework or any of the other mentioned tree-based toolkits would be interesting.

We would also be keen to assess the results that would be yielded by applying a reordering technique equivalent to nipponization for Korean, Chinese or Arabic in our wish to create a heterogeneous range of PangeaMT-Toshiba hybrid MT solutions.

### *Acknowledgements*

The authors wish to thank one another and their respective companies for their enthusiasm and support to stick to this collaboration project throughout the current year, particularly after the devastating natural disaster of March 11<sup>th</sup> in Japan.

### **References**

- Papineni, K. et al. (2002) BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proc. of the 20th ACL*. Association for Computational Linguistics. Issue: July, pages: 311-318.
- Yuste et al. (2010) PangeaMT – putting standards to work... well. In *Proceedings of The Ninth Conference of the Association for Machine Translation in the Americas – AMTA 2010*, Denver. Available at: <http://amta2010.amtaweb.org/AMTA/papers/4-04-HerranzYusteEtal.pdf>
- Zhang (Joy), Ying and Stephan Vogel (2004). Measuring Confidence Intervals for MT Evaluation Metrics. In *IEEE Journal Transactions on Medical Imaging – TMI*. Baltimore, MD, USA.

---

14 [www-i6.informatik.rwth-aachen.de/jane](http://www-i6.informatik.rwth-aachen.de/jane)

15 [joshua.sourceforge.net](http://joshua.sourceforge.net)

## 多言語医療対話支援システムの開発と公開 ～用例対訳と機械翻訳の併用～

和歌山大学 システム工学部  
吉野 孝、福島 拓、尾崎 俊、東 拓央

東京大学 知の構造化センター  
宮部 真衣

### 1. はじめに

2005年、情報通信研究機構（NICT）で始まった言語グリッドプロジェクトをきっかけに、和歌山大学の吉野研究室では、多言語間コミュニケーション支援システムに関する研究開発をスタートしました。私たちは、特に、医療現場における外国人患者と日本人医療従事者間の言葉の壁を少しでも解消するためのシステムの研究開発に取り組んでいます。

AAMT Journal No.48（2010年11月）では、「M<sup>3</sup>（エムキューブ）とTackPad（タックパッド）－ 用例対訳を用いた多言語医療対話支援システム－」が掲載されました。このレポートでは、その後の1年間の進展内容と新しく開発したシステムについて紹介します。これまでの研究開発の背景や医療分野における多言語間コミュニケーションの問題については、前回の記事をご参照ください。

このレポートでは、まず、医療機関における様々な場所での利用を想定した携帯型多言語医療対話支援システム「ぶち通」と、多言語問診票作成システムを紹介します。これらの2つのシステムは、用例対訳<sup>1</sup>と機械翻訳を併用した点に特徴があります。次に、東京大学医学部附属病院に導入された多言語医療受付支援システムM<sup>3</sup>について紹介します。また、2011年3月末に公開したスマートフォン上で動作するM<sup>3</sup>（モバイル版M<sup>3</sup>）について紹介します。

### 2. 携帯型多言語医療対話支援システム「ぶち通」

携帯型多言語医療対話支援システム「ぶち通」は、医療機関の様々な場所で利用可能な携帯型の多言語医療対話支援システムです。Androidを搭載した携帯端末



図 1 携帯型多言語医療対話支援システム「ぶち通」の起動時の画面

上で動作します。図1は、ぶち通の起動時の画面です。ぶち通は、主に、入院時の医療従事者と外国人患者との対話の支援を想定したシステムです。医療従事者主導でシステムを利用します。

医療現場へ導入することを想定していたため、当初ぶち通では用例対訳のみの利用を考えていました。しかし、入院時には、医療に関する定型なコミュニケーションに加えて、医療従事者と外国人患者との多岐にわたる日常的な会話への対応が必要とされており、用例対訳のみでは対応が難しいことがわかりました。そこで、翻訳の正確性が求められる医療に関する会話には用例対訳を、それほど正確性を必要としない日常的な会話には機械翻訳を利用することで、多様な状況に対応しています。機械翻訳を利用した場合は、その入力文を多言語用例対訳共有システムTackPadへ送り、用例対訳の作成を依頼します。この仕組みにより、2回

<sup>1</sup> 用例対訳とは、予め正確に翻訳された多言語の対訳のことです。



図2 携帯型多言語医療対話支援システム「ぶち通」の画面例

目以降は、正確性の確保された用例対訳を利用可能にすることを狙っています。

ぶち通の機能を紹介します。ぶち通には、場面検索機能、翻訳機能、簡易応答機能、音声認識・音声合成機能（Google音声認識、言語グリッドを利用）、お気に入り登録機能などがあります。図2は、ぶち通の各機能の画面例です。現在、ぶち通は、日本語、英語、中国語、韓国・朝鮮語、ポルトガル語の5言語に対応しています。

図2(a)は、場面検索機能の画面例です。場面検索機能では、医療従事者が、対話場面によって用例対訳を絞り込むことができます。文字を入力することなく、階層化された対話場面を順番にクリックすることで、用例対訳の検索をできるようにしています。これらの場面検索機能の対話用例数は、5,390文（1,078文×5言語）あり、17の場面にカテゴリ分けされています。これらのカテゴリ分けは、医療従事者の協力を得て、適切に行ってもらっています。

図2(b)は、翻訳機能の画面例です。この機能は、場面検索機能を使わずに、音声認識を利用して、類似した用例対訳を表示する機能です。また、音声認識した結果を、機械翻訳によって翻訳して表示する機能もあります。機械翻訳結果の精度を利用者に判断してもらうために、折り返し翻訳結果も表示しています。

図2(c)は、簡易応答機能の画面例です。日本人医療従

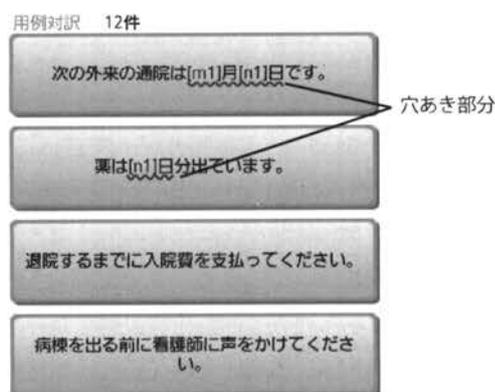


図3 穴あき用例の例

事者が、外国人患者に質問した場合、患者からの返答が必要となります。簡易応答機能は、外国人から必要な返事を得るために利用する機能です。より詳細な回答を得るための、詳細応答機能もあります。

図3は、穴あき用例の例です。穴あき用例とは、日付や金額など、数字が入ることが分かっている用例の該当部分に、予め穴（図3の[m1]や[n1]など）を割り当てることによって、用例対訳を用いながら、多くのパターンに対応できるようにした用例です。

現在、ぶち通は、医療従事者の意見を聞きながら、鋭意改良中です。2011年度中を目標に、実際の医療機関における試用を予定しています。

### 3. 多言語問診票作成システムの開発と公開

日本の医療機関では、初診の患者が来院した場合、患者の症状を知るために問診票の記入を求めます。また、医療従事者が患者の症状に関する詳細な情報を得るために、多くの問診票には自由記述の項目が用意されています。

そこで、私たちは、(1)各医療機関独自の多言語問診票の様式作成機能と (2)多言語問診票への症状入力機能を持った多言語問診票作成システムを開発しました。このシステムを使うことで、医療従事者と外国人患者それぞれの母語が表記された問診票を作成することができます。

図4は、多言語問診票の様式作成機能の画面例です。様式作成には、正確性の確保された用例対訳のみを用いています。事前に既存の紙の問診票を調査し、紙と同等の内容の問診票が作成できるようにしています。また、作成済みの問診票を改良して新たな問診票を作成できるようにすることで、問診票作成者の負担を軽減しています。

図5は、多言語問診票の入力機能の画面例です。患者は、基本的には選択式の項目から症状を選択しますが、当てはまる症状がない場合や、症状を詳しく記入したいときは自由に症状を書くこともできます。システムは、自由に記述された内容と同じ用例対訳がないか調べ、同じものがあれば用例対訳を、なければ機械翻訳を使用します。自動的に用例対訳と機械翻訳を切り替えることで、患者の入力の負担を軽減しています。また、図6のように用例対訳と機械翻訳を表示し、患者が意図する内容を選択できる画面も用意しています。図6には4つの候補が出ています。一番上が機械翻訳で、下の3つが入力文と類似した用例対訳です。文を修正することもできるため、多言語問診票に、より正確に入力することができます。

多言語問診票作成システムは、「こころの症状に関する問診」(日本語と英語、中国語、韓国・朝鮮語、ポルトガル語、ベトナム語間の翻訳に対応)と「放射能被ばくに関する問診」(日本語と英語、中国語、ポルトガル語間の翻訳に対応)の2つの問診票を公開しています。多言語問診票作成システムに症状を入力すると、2言語併記された問診内容がPDF形式で出力され、印刷して利用できます。PDF形式で出力された問診票の中で機械翻訳を使用した文は、赤字で強調されています。こ



図4 多言語問診票の様式作成機能の画面例



図5 多言語問診票の入力機能の画面例

れは、問診票を見る医療従事者に、機械翻訳による翻訳であることを伝え、注意を促すために行っています。

多言語問診票作成システムは、下記のURLから利用できます。

URL: <http://is.tackpad.net/>

2 目前症状(現在どのような症状がありますか?)

- 其他(その他)

原文	訳文(日本語)	再翻訳	
	朝起きた後に、頭がぼろっとするのはもうろうとしていて(もうろうとして)、全身力がない	早上起来之后, 头朦胧地要朦胧没有(朦胧), 全身力	選択中
早上起床后, 头昏昏沉沉, 浑身无力	⇒ 体がだるい	⇒ 浑身无力	使用
	⇒ 浮動性のめまいがする	⇒ 头昏眼花	使用
	⇒ のぼせる	⇒ 上火、头昏脑涨	使用

図 6 用例対訳と機械翻訳の翻訳結果選択画面の例

#### 4. 多言語医療受付支援システムM<sup>3</sup>(エムキューブ)の東京大学医学部附属病院への導入

2011年6月30日、東京大学医学部附属病院外来診療棟1階総合案内横に、多言語医療受付支援システムM<sup>3</sup>(エムキューブ、以下、病院設置型M<sup>3</sup>)が設置されました。図7は、設置されたシステムの写真です。病院設置型M<sup>3</sup>は、医療受付における外国人患者の支援を目的としており、タッチパネルを用いて操作することができます。病院設置型M<sup>3</sup>は、5言語(日本語、英語、中国語、韓国・朝鮮語、ポルトガル語)で受診の流れなどを案内することができます。

医療機関への導入には、医療機関内での機器導入の判断なども必要となり、時間を要します。各医療機関向けのシステムの調整(道案内データの作成や各病院個別の手続き内容・Q&Aの多言語化など)には、数ヶ月程度要します。また、医療機関側での作業もあるため、なかなかスムーズには設置できません。今回、導入の打ち合わせをしてから、実際に設置されるまで1年ほどかかりました。

今回の導入では、病院内の道案内経路の作成や写真撮影など、必要なデータの作成に関して病院側の積極的な協力が得られました。

#### 5. スマートフォン上で動作するM<sup>3</sup>(モバイル版M<sup>3</sup>)の公開

スマートフォン上で動作するM<sup>3</sup>(モバイル版M<sup>3</sup>)は、病院設置型M<sup>3</sup>の問診機能をスマートフォン上で利用できるようにしたものです。利用者自身のスマートフォン上で問診機能が利用可能になるとさらに便利になると考え、開発しました。

図8は、モバイル版M<sup>3</sup>の動作中の画面例です。モバ



図 7 東京大学医学部附属病院外来診療棟 1 階総合案内横に設置された多言語医療受付支援システム M<sup>3</sup>(エムキューブ)

イル版M<sup>3</sup>は、Androidを搭載した携帯端末とiPhone上で動作します。図9(a)は、言語選択画面の例です。モバイル版M<sup>3</sup>は、病院設置型M<sup>3</sup>で対応している言語に加え、スペイン語、インドネシア語、ロシア語、やさしい日本語でも利用することができます。図9(b)は、履歴機能の画面例です。モバイル版M<sup>3</sup>専用の機能として、選択した症状を記録できる機能を作成しました。症状を一度記録すると、記録した内容をいつでも見ることができます。

モバイル版M<sup>3</sup>には、3,708用例(412症状×9言語)があります。現在、無料で公開しています。

ダウンロードは下記のQRコード(URL)から行うことができます。公開して半年ですが、Android版は約10,000回、iPhone版は約5,000回、それぞれダウンロードされています。

Android Market:

<https://market.android.com/details?id=jp.m3.full>



App Store:

<http://itunes.apple.com/jp/app/m3-for-iphone/id427406934>



(a) Android 版 M<sup>3</sup> の画面



(b) iPhone 版 M<sup>3</sup> の画面

図 8 モバイル版 M<sup>3</sup> の画面例



(a) 言語選択画面



(b) 履歴機能画面

図 9 モバイル版 M<sup>3</sup> の言語選択と履歴機能の画面例

## 6. おわりに

このレポートでは、和歌山大学の吉野研究室を中心に研究開発している多言語医療対話支援システムについて紹介しました。

当初、医療分野での機械翻訳の利用は、全くできないものと思っていましたが、用例対訳の精度の高さと、機械翻訳の自由度の高さをうまく活かすことで、高い有用性をもつシステムとなる可能性が開けてきました。

今後は、実際の医療現場へ導入することで、その有用性の検証を行っていきたいと考えています。

## 謝辞

多言語医療対話支援システムの開発は、医療従事者の皆様、翻訳者の皆様、システムの画面設計や性能評価に関わった多くの学生など、非常に多くの関係者に支えられております。ここに記して謝意を表します。

多文化共生センターきょうとの重野亜久里氏、前田華奈氏には、本研究室で開発している全ての多言語医療対話支援システムにおいて、開発当初から現在に至るまで、システムの設計や改良、多言語化に、多大な協力を頂きました。

病院設置型M<sup>3</sup>とモバイル版M<sup>3</sup>、ぷち通、TackPad、多言語問診票作成システムの開発の一部は、総務省戦略的情報通信研究開発推進制度 (SCOPE) の助成により行われています。

# Web の多言語化と言語技術の標準化:

W3C MultilingualWeb Limerick Workshop より

大阪大学

林 良彦

## はじめに

2011年9月21日～22日にアイルランド・リムリック市で開催された W3C MultilingualWeb<sup>1</sup> Limerick Workshop に参加する機会を得た<sup>2</sup>ので、その模様を簡単に報告し、Web の多言語化(あるいは多言語 Web)と言語技術に関連する標準化について若干の所感を述べてさせていただきます。

## 1. W3C MultilingualWeb Project

ご存知のように World Wide Web Consortium (W3C) は、World Wide Web に関する技術の標準化を推進する団体である。HTML や XML といった基本的な規格はもとより、セマンティック Web に関連して、RDF や SPARQL といった規格の標準化も進めている。

全世界的な情報発信・流通の活発化に伴い、Web 上で各国語による情報を正しく表示し処理すること、さらには多言語の情報を同時に扱うことができる Web の多言語化への要請が高まっている。このような背景のもと、MultilingualWeb Project<sup>3</sup>は、EC によるファンディング、W3C によるコーディネートというスキームで設立され、Web の多言語化に関連する技術標準やベストプラクティスに関する情報交換を目的とする公開ワークショップを2年間に4回、開催することとしている。

## 2. Limerick Workshop 模様

今回リムリック大学をホストとして開催されたワー

クショップは、全4回のワークショップシリーズの第3回目にあたる。

### (1) ワークショップ第1日目

2日間の会期のうち初日は、基調講演の後、以下に示す6つのセッションが企画され、参加者からの報告を中心に議論が進められた。会議中は、Twitter による意見交換が並行して進められていたほか、IRC (Internet Relay Chat) によるほぼリアルタイムのまとめ情報の提供が行われていた。

- 基調講演 (Daniel Glazman, W3C CSS Working Group Co-Chair): Web の多言語化に対する要請の背景を整理して提示し、いかに HTML5, CSS3, EPUB3 といった規格がこの要請に答えようとしているかを述べた。
- 開発者 (Developers) セッション: 開発者の立場からの3件の発表があり、International Tag Set (ITS) の現状、表示におけるハイフネーション(ドイツ語などでは長い複合名詞が現れる)のほか、EU における LT-Web<sup>4</sup> (LT: Language Technology) の動向報告があった。翻訳に関連しては、特に翻訳によって生成されたコンテンツに対するメタデータの付与が議論の中心的な対象となっているようであった。
- クリエータ (Creators) セッション: コンテンツ制作側からの報告が3件行われた。ここでの中心的なキーワードの一つは、コンテンツマネジメントということであり、Contents Management System (CMS) の利用やコンテンツのローカライゼーションの事例が紹介された。

<sup>1</sup> “MultilingualWeb”は筆者のスペルミスではなく、そのように表記されている。

<sup>2</sup> 総務省戦略的情報通信研究開発推進制度(SCOPE)の援助を受けた。

<sup>3</sup> <http://www.multilingualweb.eu/>

<sup>4</sup> <http://cordis.europa.eu/fp7/ict/language-technologies/>

- ローカライザ (Localizers) セッション: ローカライゼーションに関連する 3 件の報告が行われた。ローカライゼーションにおけるキー要素は、速度、ボリューム、コストということで、いかに効率よくローカライゼーションを達成するかというのが最大の焦点とのことである。発表の中では、クラウドソーシング的なアプローチにより、いわゆるリソースに乏しいマイナーな言語の翻訳を行うという事例が注目を浴びていた。
- マシン (Machines) セッション: 言語処理と関連して、統計的機械翻訳によるトピックに特化した翻訳を経済記事における評価表現の解析に適用した経験、RDF を適用した自然言語処理ツールのデータ交換フォーマットの提案、辞書モデリングのための ISO 標準 LMF を適用した多言語意味辞書のアクセスサービスの実装(報告者による)の 3 件の報告が行われた。特に翻訳技術に関するのは最初の報告で、現状の翻訳サービスの実質的な標準である Google API はプリミティブ過ぎであり、ドメイン固有のより精緻な情報を扱えるようにすべきであることや、機械翻訳により生成された Web コンテンツに対するマークアップの必要性などが述べられた。
- ユーザ (Users) セッション: 2 件の報告が行われた。そのうちの 1 件は、パーソナライズ化された多言語 Web のためのコンテンツマネージメントに関するものであり、その際にキーとなるのは、コンテンツへの適切なメタデータの付与であることが強調された。これに関連して、人々、出来事といったエンティティに関するマークアップを行うためのスキーマ体系<sup>5</sup>の紹介があった。
- ポリシー (Policy) セッション: 多言語 Web のための技術標準、標準化活動に関する報告が 3 件行われた。ISO, W3C, OASIS といった複数の標準化機構・団体がお互いに重複するような標準化を進めていて、その整理が必要であるという考えが

示されたが、これは一筋縄ではいかないという印象が強かった。また、技術標準は実装されてこそ意味を持つというような、いまさらながらの議論が出ていたことも標準化に関する難しさを示しているように感じられた。

## (2) ワークショップ第 2 日目

第 2 日目はブレイクアウトセッションということで、グループ討論のテーマの提案と選定、グループ討論の実施、まとめの報告・議論という手順で進められた。編成されたグループは、(1) 標準化、(2) 翻訳コンテンツの標準化、(3) LT-Web、(4) 多言語ソーシャルメディア、(5) ベストプラクティスの 5 つであった。翻訳に関連しては、(2) において、翻訳文書をパッケージ化するコンテンツの標準化に関する議論が行われ、そのなかで Linport プロジェクト<sup>6</sup>の紹介が行われた。また、(3) においては、通常 CMS で管理されている翻訳コンテンツを翻訳システムの訓練データとして利用するという方向性が示されたほか、(4) では、翻訳業務にクラウドソーシングを適用する際の問題点などが提起された。

半日という限られた時間の中で活発な議論が行われたことに関しては、参加者だけでなく主催者側も大いに評価しており、2012 年 2 月にルクセンブルクで開催される予定の最終回のワークショップでもこのフォーマットが適用される見込みとなった。

## 3. 所感: Web の多言語化と言語技術の標準化

欧州での研究開発や技術標準化の動向を調査したり、現地の研究者たちと議論したりするたびに感じることであるが、欧州における multilingualism は、日々の諸活動に根ざした概念であることを今回も痛感させられた。また、言語技術 (language technology) や言語サービス (language service) といった用語も我々が抱きがちな純テクノロジー的なものではなく、これらの適用条件や利用環境を含めた幅広い概念であるように感じられる。

<sup>5</sup> <http://schema.org>

<sup>6</sup> <http://www.linport.org/>

その上で、広い意味での言語技術の標準化は、Webの多言語化(あるいは多言語 Web) のより良い実現へ向けて、避けては通れない、あるいは、待ったなし、の課題として認識されているようである。また、多言語と言っても必ずしも欧米系の言語だけが含まれるのではなく、日本語や中国語に代表されるようなアジアの言語に対しても相応の配慮がされているように感じられた(その端的な一例は HTML における Ruby (るび))。

翻訳に関連しては、機械翻訳システムや特に翻訳メモリが翻訳業務において普通に利用され、さらに CMS によってコンテンツやワークフローの管理ツールとして定着しているという現在の状況の上にコンテンツ(特に翻訳文書)に対する各種のメタデータの体系を整備し、これによりコンテンツの相互運用性(interoperability)を高めようという方向性が顕著になっていると感じた。

#### 4. Centre for Next Generation Localisation

上記ワークショップにも参加されていたリムリック大学の西尾直人さんが同大学のキャンパス、および、キャンパス内に設置されている Centre for Next Generation Localisation<sup>7</sup> のオフィスを案内して下さった。このセンターは、リムリック大学、ダブリン市立大学などのアイルランド内の4つの大学、および、世界的な10社(日本からは大日本印刷)が参加している産学連携的な性格を持つ研究開発組織であり、ローカリゼーションに関する活発な活動が展開されている。

日本ではローカリゼーションはまだ学問領域としては確立されていないように見受けられるが、彼の地アイルランドにおいては、機械翻訳や自然言語処理だけでなく、ヒューマンインタフェースや、さらには翻訳論といった文系的な領域を総合する学際的な領域として認知され、国の一つの重要な戦略的な領域としても位置づけられ、関連する教育プログラムにも展開されているようであった。例えば、リムリック大学と LRC

(Localisation Research Centre)<sup>8</sup> は、これまで提供してきた修士課程の MSc in Localisation and technology を Distant learning 形式にてより多くの学生に提供できるよう準備しているとのことである。

「なぜアイルランドでこのような研究が活発なのか?」という報告者の間に、西尾さんは「1980年代中ごろから、アイルランドではマイクロソフトやシマンテックなどの企業によるソフトウェア製品の生産・輸出が主要な産業となり、ソフトウェアのローカリゼーションが必然となる流れの中で、その研究開発、産業化の中心と位置づけられるようになった」というように答えてくださった。

筆者はこの分野の現状に詳しくはないが、現状の日本では依然としてローカリゼーションの必要性が低いのかもしれない。これには、日本語あるいは日本文化といった特徴も関連しているのかもしれないが、人材の交流などが多くの分野で進みむにつれ、状況は変わっていくものと考えられる。そうなったとき、日本におけるローカリゼーションの研究開発や実践も盛んになる可能性がある。そのとき日本においては、ローカリゼーションは学問的にどのような位置づけがされ、関連する研究テーマは大学ではどんな学部・学科に属することになるのだろうか?(ちなみに、上記のセンターは計算機科学科の配下に位置づけられているとのことであった)



**写真:** リムリック大学のキャンパスの中には、シャノン川がゆったり流れている。川岸には遊歩道も整備されており、快適な散歩が楽しめそうである。

<sup>7</sup> <http://www.engl.ie/>

<sup>8</sup> <http://www.localisation.ie/>. 1995年に Reinhard Schäler氏を中心に設立。2008年より CNGL と共同研究を継続中。

## AAMT Forum ホームページ刷新について

AAMT インターネットワーキンググループ

## 1. はじめに

現在の AAMT ホームページに刷新されてから 9 年が経過しました。9 年前の刷新時は、ホームページ管理における問題、デザインが「暗い」、「地味」という意見などに加え、英語と日本語のページしかなく、「アジア・太平洋」と言うことがはばかれるなどの理由から刷新が図られました。

刷新された AAMT ホームページは、アジア太平洋地区全体に関わる情報を記載するグローバルページと、各地域の情報を記載するローカルページから構成されるようになりました。そして、「AAMT のページを見れば、MT のことは全てわかる」ということを目指して、情報の拡充にも努めてきました。MT の歴史年表は、日本初と自負しています。今までは、読み物としての MT の歴史はありましたが、「〇〇年代には、こういう研究が盛んであった」とか「〇〇年頃には、こういう製品が普及した」とかという記載が多く、何年に何があったと明確な記載はありませんでした。また、翻訳ソフトのリスト、書籍、雑誌記事、新聞記事のリストも、AAMT の名に恥じない内容となっていると自負しています。

AAMT ホームページは、2011 年 9 月 26 日の時点で、総ページ数 255 ページ、外部へのリンク数 634 の規模となりました。現在、インターネットワーキンググループでは、より良い AAMT ホームページとすべく、再び刷新を検討しています。本稿においては、ホームページ刷新活動の途中経過を報告させていただきます。

## 2. AAMT ホームページの多言語化

今回のホームページ刷新の目的の一つは、AAMT ホームページの多言語への対応を図ることです。そこで、AAMT ホームページにおける使用言語を調査しました。

トップページが用意されている言語として英語、日本語、中国語、ハンダ語の四言語が挙げられます。これらの言語の比率は次のグラフ（図 1）の通りです。主たる使用言語として日本語が位置づけられます。

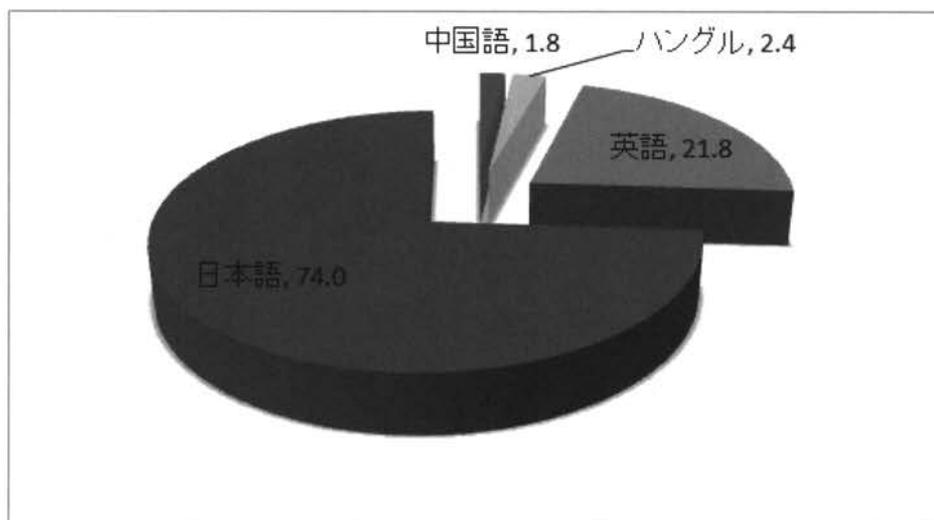


図 1: ホームページ内の言語比率 (%)

AAMT 韓国支部の発足を踏まえ、今後ハングルの比率が上がると予測されます。また、これらの四言語以外にも数多くの言語への対応が必要と思われます。その一端を示す事例として、AAMT ホームページ内のローカルポータルとして対応されている国や地域の範囲が挙げられます。現在、準備されているローカルポータルは、アジア太平洋地域の 13 カ国、インド、インドネシア、オーストラリア、シンガポール、ベトナム、マレーシア、ミャンマー、ラオス、韓国、香港、台湾、中国、日本となります。当初、各地域のページは各地域で作成して頂くという計画でしたが、残念ながら実現していません。また、タイのページは、プーケットでの MT サミットにあわせ、一時的にタイの担当者により作成されていました。

### 3. AAMT ホームページのデザイン

ホームページ刷新を検討するにあたり、AAMT ホームページの作成当時のことを調査し、その当時担当されていた方々（現在も担当されている方がほとんどのようですが）の様々な苦難に加え、作成方針などが AAMT ジャーナルに掲載されていることを知りました（「AAMT Web 作成ダイアリー」古川淳子、AAMT ジャーナル 33 号）。その記事によると、作成方針の一つとしてアジアを彷彿させるデザインをモチーフとすることが挙げられています。この方針に沿って作成されたトップページにはアジアを彷彿させる画像や色彩が配されるといった工夫が施されています（図 2）。

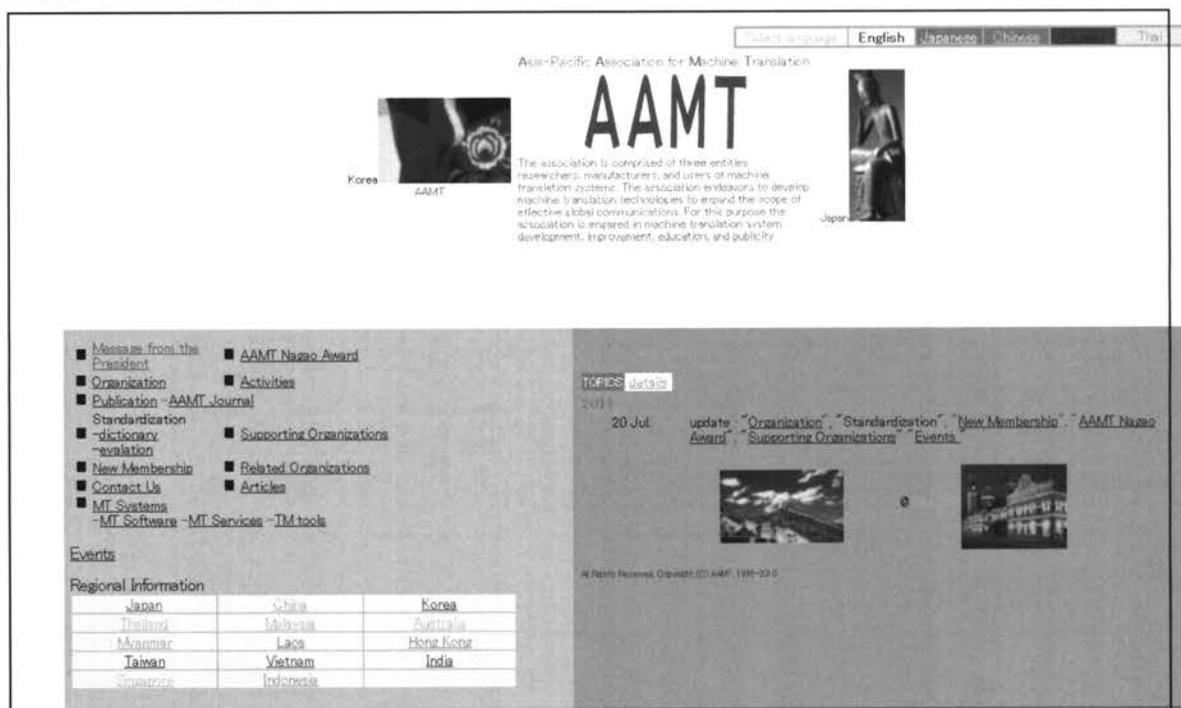


図 2 : AAMT ホームページ

AAMT の関連団体である EAMT (European Association for Machine Translation) のホームページ (<http://www.eamt.org>) (図 3) や AMTA (The Association for Machine Translation in the Americas) のホームページ (<http://www.amtaweb.org>) (図 4) と比較しても、このデザインは特徴的といえるのではないのでしょうか。

この特徴的なデザインを踏襲しつつ、新たなデザインを検討しています。AAMT ホームページのデザインに関しては、会員の方々から「もっと機械翻訳らしく」、「もっと学際的に」と色々な意見が寄せられていると聞きま

した。これらのデザインに関する要望を具現化するには程遠い芸術性しか持ち合わせない担当者の一つの手立てとして、関連団体のホームページに倣い、名称をロゴのように冠する手立てを検討中です（図5）。



図3：EAMT ホームページ

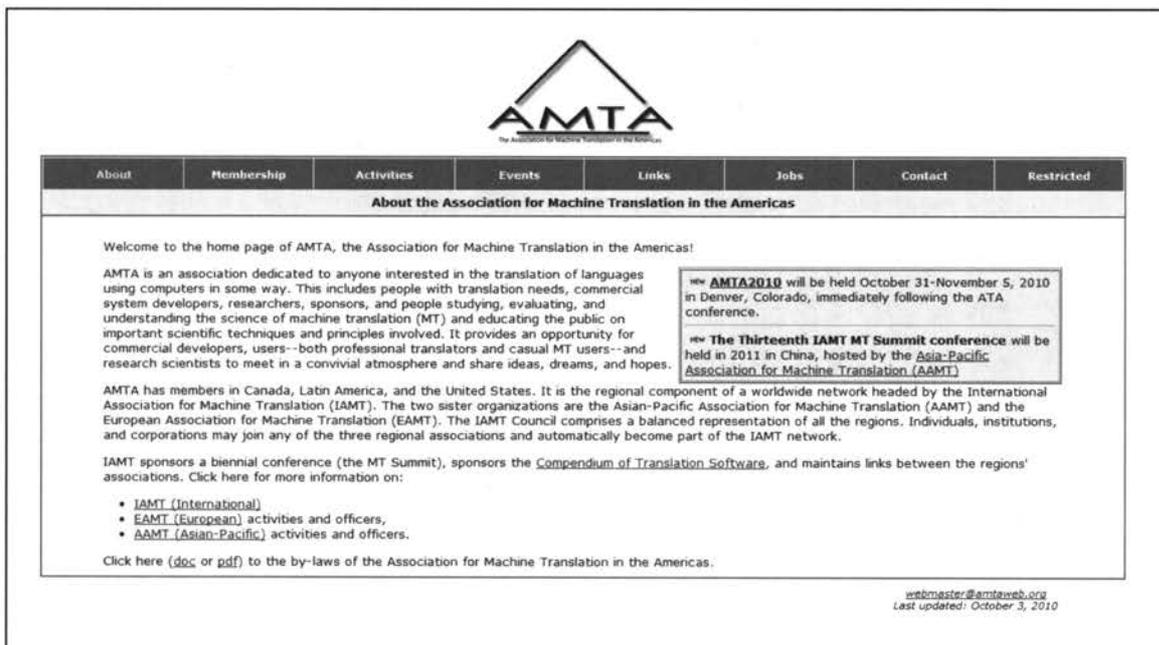


図4：AMTA ホームページ

# Asia-Pacific Association for Machine Translation

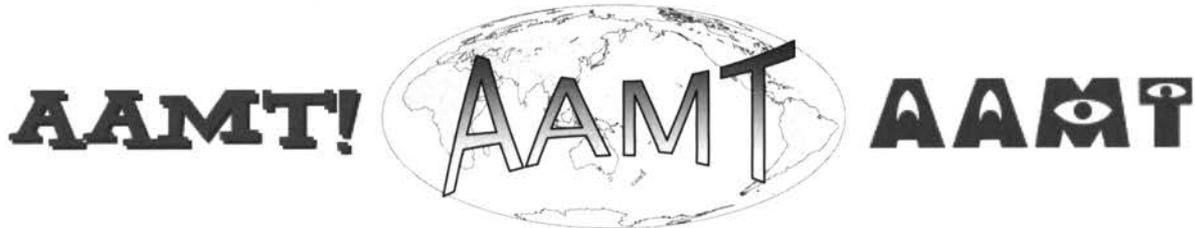


図5：デザイン例

## 4. その他の検討項目

上記の多言語化とデザインに加え、以下の項目が検討課題として挙がっています。

- ① テキストエディタ等でメンテナンスを行えるようにする。
- ② 各ページのダウンロード容量を少なくする。
- ③ 新着情報に対して「New」をつけるなどの編集をせずに新規箇所を明示する。
- ④ ユニバーサルデザインに配慮する。
- ⑤ AMTA や EAMT のようにグローバルナビゲーション機能を設ける。
- ⑥ サイト内検索機能を設ける。

## 5. おわりに

本稿では、インターネットワーキンググループが取り組んでいる AAMT ホームページの刷新について途中経過を報告させていただきました。より良いホームページとなるよう更に検討を重ねていきます。

今後、さらに AAMT ホームページを発展させるために検討すべき課題として、新規情報を収集することに特化したグループの策定、あるいは担当を設けることが挙げられます。これまで AAMT ホームページに掲載する情報の多くをインターネットワーキンググループで収集してきました。しかし、AAMT ホームページの維持が任務であるインターネットワーキンググループが、規模が大きくなったホームページの維持に加え、兼務として新規情報を収集することが困難となってきました。今回のホームページ刷新を機に AAMT ホームページの管理体制の再構築のきっかけになればと思います。

## 英日・日英機械翻訳に関するアンケート結果報告

機械翻訳課題調査委員会ワーキンググループ 2

## 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。本稿では、2011 年（2010 年度）のアンケート調査結果について報告する。

2011 年の調査は、定点観測と言える従来の基本項目（アンケート回答者自身の属性情報、機械翻訳の必要性、利用頻度、利用目的、分野など）に加え、利用者辞書登録の利用方法をテーマにアンケートを実施した。なお、2009 年（2008 年度）アンケートでは携帯電話における機械翻訳の利用をテーマにして通常の回答ページ以外に携帯電話用の回答ページも設ける工夫をしたが、翌年 2010 年には通常通りの方法に戻している。今回も 2010 年をほぼ踏襲する形で通常通りの方法でアンケートを実施している。

以下に 2011 年のアンケート実施概要を示す。

- アンケート期間
  - 2011 年 3 月 1 日～3 月 21 日（21 日間）  
期間中に 7 日間延長
- アンケート進呈品
  - 抽選で 30 名にクオカード 3,000 円分、自由記述設問の最優秀者 1 名に特賞クオカード 10,000 円分。  
※贈呈品にふさわしい価格帯の機械翻訳ソフト製品が無かったため、やむを得ずクオカードのみの進呈とした。
- アンケート回答者数
  - 77 人（うち、機械翻訳を必要としている

と回答した人は 56 人）

- なお、過去のアンケート回答総数は、2003 年 1,413 人、2004 年 1,720 人、2005 年 1,018 人、2006 年 1,061 人、2007 年 835 人、2008 年 1,495 (946) 人、2009 年 1,811 人 (1,059 人)、2010 年 1,022 人 (745 人)。

※括弧内は機械翻訳を必要とする人数。

今回、アンケート回答者が激減する異常事態が発生したが、これについては、「3. アンケート回答者激減について」で述べる。

## 2. 設問項目

主に以下に示す項目について質問した。

- 日英・英日機械翻訳ソフト・サービス全般に関して（日英・英日機械翻訳の必要性とその理由、機械翻訳の利用形態・用途・利用頻度・対象分野、機械翻訳に対する満足度等）
- 利用者辞書の利用方法（利用者辞書を利用しているか否か、利用者辞書の登録語数、利用者辞書を利用することによる作業効率向上の度合い等）
- 翻訳ソフトの有効的な利用方法の提案
- 回答者の基本属性について（年齢層、性別、職業、英語力など）

「翻訳ソフトの有効的な利用方法の提案」は新規に追加した項目であり、現在実践しているあるいは今後やってみたい翻訳ソフトのユニークな使い方を、文章で自由に回答してもらうものである。

## 3. アンケート回答者激減について

今回のアンケートでは、これまで例年 1,000 人以上回答者数を得ていたのが、77 人と大激減した。回答者

数が伸び悩んだため、アンケート期間の途中で期間を1週間延長したが挽回できなかった。以下に原因と思われる理由をあげる。特に①が決定的に大きく、その次に②がある程度影響したと推察している。

- ① 400名以上のアクセスがあったyahoo 懸賞がサービスをやめてしまったこと(そのことをWG2が把握できていなかった)
- ② 回答数が伸びる回答期間終盤に東日本大震災が発生した
- ③ アンケート贈呈品に機械翻訳ソフトを含めなかった(例年、半数は機械翻訳ソフトを希望)
- ④ アンケートの最初に正確を期すための注意書きを負やしたため、回答意欲が削がれた

図 3-1~3-3 に、どこでアンケートを知ったかという問いについての回答と、回答者数推移について、前回 2010 年の結果と今回 2011 年の結果を示す。

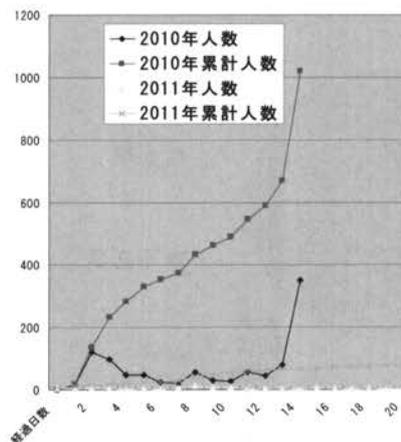


図 3-3 2010年(前回)と2011年(今回)の回答者数推移

今回、回答者数が激減し、英語能力の高い回答者の割合が例年になく高くなった。調査結果も、そのことが影響して例年にない変化が生じた。この後の節ではその点を踏まえて解説する。なお、回答者層の変化の詳細については、「6. 回答者に関する調査結果」を参照いただきたい。

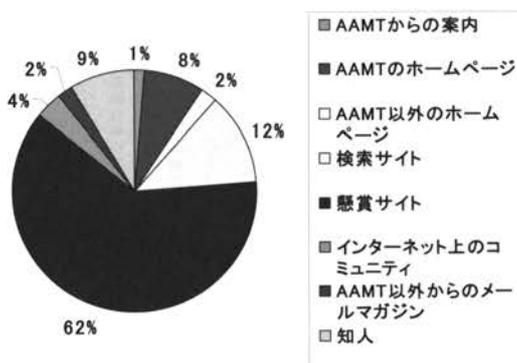


図 3-1 2010年(前回)：どこでアンケートを知ったか

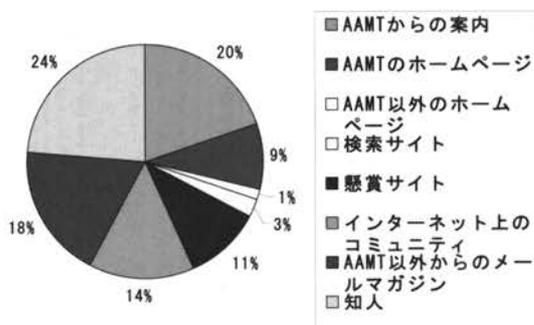


図 3-2 2011年(今回)：どこでアンケートを知ったか

#### 4. 機械翻訳全般に関する調査結果

本節では、機械翻訳全般に関する調査結果として、日英機械翻訳の必要性の有無とその理由、機械翻訳の利用形態・用途・利用頻度・対象分野、機械翻訳に対する満足度等についての調査結果を示す。

##### 4.1. 日英・英日機械翻訳の必要性の有無とその理由についての調査結果

図 4-1 に(有償・無償か、また、Web 上のサービスかパソコンソフトかを問わず)日英・英日機械翻訳が必要か否か、また、必要・不要それぞれの理由を下記に示す。機械翻訳は、75% (59 人) が必要と応え、25% (19 人) が不要と回答している。この比率自体は、2010 年の73%が必要、27%が不要という回答とほぼ同じである。

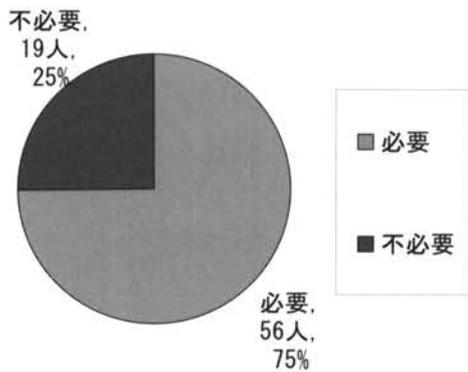


図 4-1 日英・英日機械翻訳が必要か

日英・英日機械翻訳の必要・不要それぞれの理由について、2010年の結果と2011年の結果を、図 4-2～図 4-5 に示す。いずれも、理由は複数選択可能な回答である。今回2011年の結果では、日英・英日機械翻訳が必要な理由としては、英語の能力はあり生産性向上に使う（概要把握、下訳作成など）場合が例年より大幅に増えた。また、2010年調査では機械翻訳を不要とする理由として圧倒的多数を占めていた「英語を話す機会はない」が、2011年調査では激減し、「自分で英語を読める」「自分で日本語を英語にできる」が多数を占めるようになった。

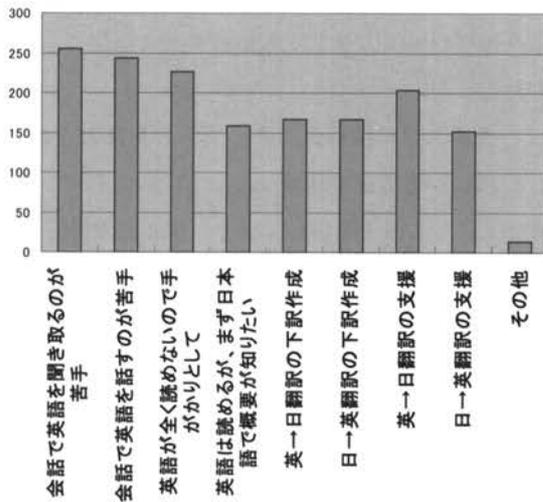


図 4-2 2010年機械翻訳が必要な理由（複数選択可）

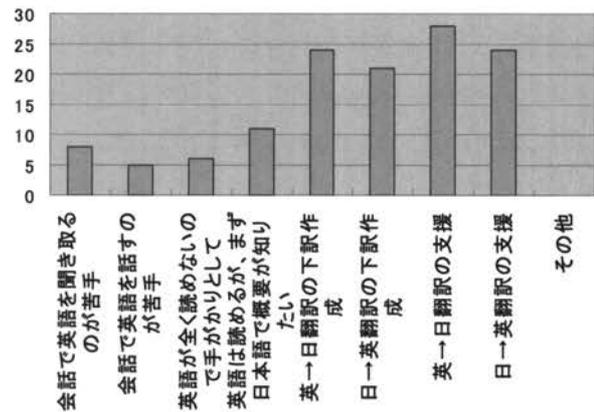


図 4-3 2011年機械翻訳が必要な理由（複数選択可）

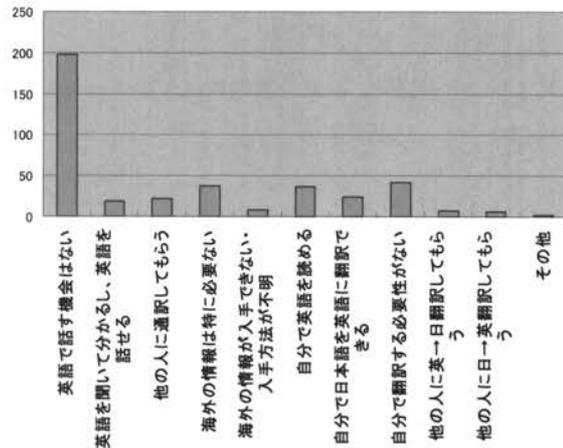


図 4-4 2010年機械翻訳が不要な理由（複数選択可）

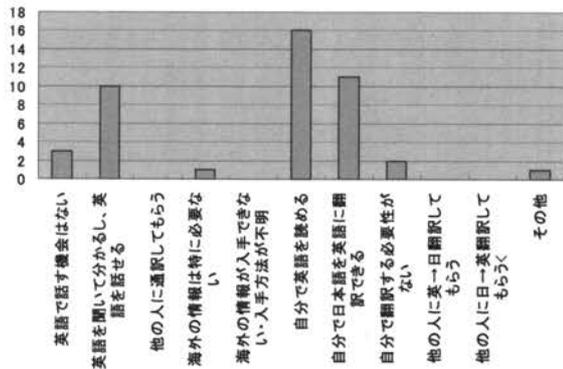


図 4-5 2011年機械翻訳が不要な理由（複数選択可）

#### 4.2. 日英・英日機械翻訳ソフト・サービスの利用状況

以後、機械翻訳が必要という人の機械翻訳ソフト・サービスの利用状況を示す。

図 4-6、4-7 に、最もよく利用する機械翻訳ソフト・サービスについて、2010 年、2011 年の結果を示す。2011 年の結果では無料の翻訳サイトの利用率が 20%減り、逆に市販の PC ソフトの利用率が 20%増えている。英語能力の高いユーザが、本格的に翻訳ソフトを使うには無償の翻訳サイトでは機能不足であることが窺われる。

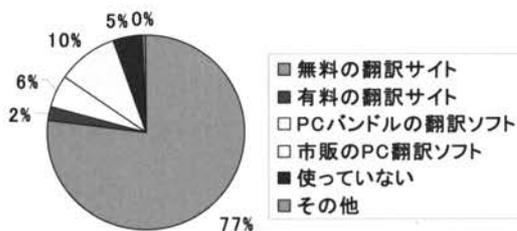


図 4-6 2010 年最も利用する機械翻訳ソフト・サービス

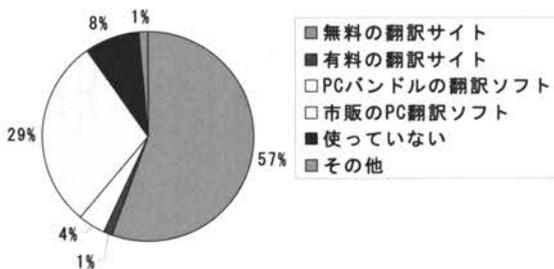


図 4-7 2011 年最も利用する機械翻訳ソフト・サービス

図 4-8、4-9 に、最もよく利用する機械翻訳ソフト・サービスに対する満足度について、2010 年、2011 年の結果を示す。高機能の市販 PC ソフト利用者の多い 2011 年の結果で、翻訳精度だけでなく、意外にも、使いやすさ、機能面でも 2010 年より辛口の評価がついている。満足度は、絶対的な評価というより価格との対比で決まってくる面があるため辛口の評価になること、下訳支援など業務として本格的に機械翻訳ソフトを使う人が多

いことから要求水準が高いことなどが理由として想定される。

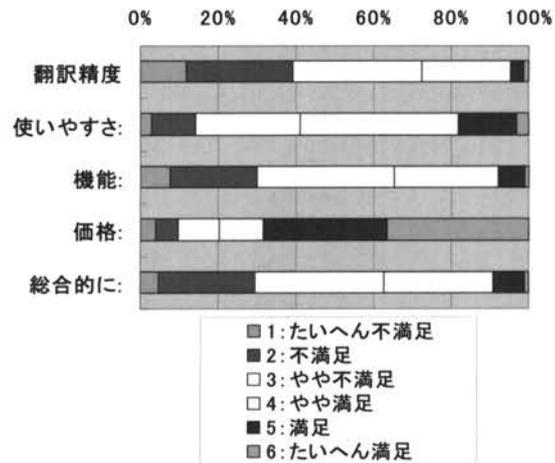


図 4-8 2010 年最もよく利用する機械翻訳ソフト・サービスに対する満足度

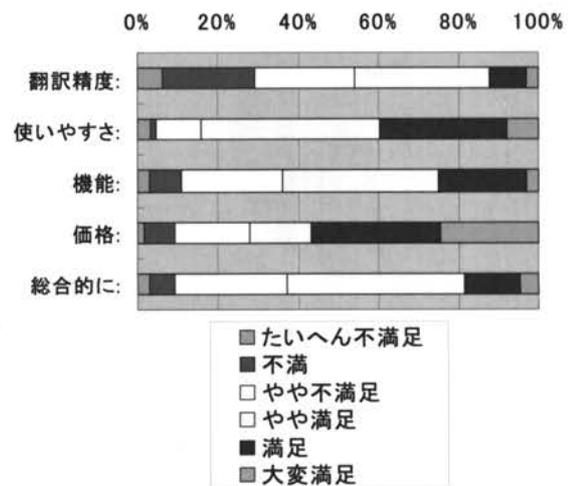


図 4-9 2011 年最もよく利用する機械翻訳ソフト・サービスに対する満足度

図 4-10、4-11 に最もよく利用する翻訳ソフト・サービスの利用頻度、図 4-12、4-13 にそのソフト・サービスを利用することによる作業時間短縮の割合について、2010 年、2011 年の結果を示す。2011 年の方が使用頻度が高い人が多いが、いずれも、月に数回以上、つまり定期的に機械翻訳ソフト・サービスを利用している人が 85%以上いることがわかる。また、2011 年は「使わな

いと全く翻訳できない」人の割合が15%以上減っているが、85%以上の人々が20%以上の作業時間短縮の効果を認めているという点で、英語能力の高低、用途に関わらず、一定の導入効果を認められていることがわかる。

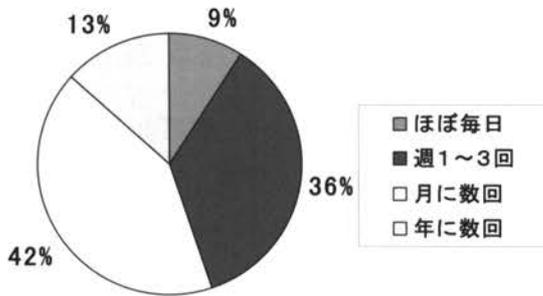


図 4-10 2010 年最も利用する機械翻訳ソフト・サービスの利用頻度

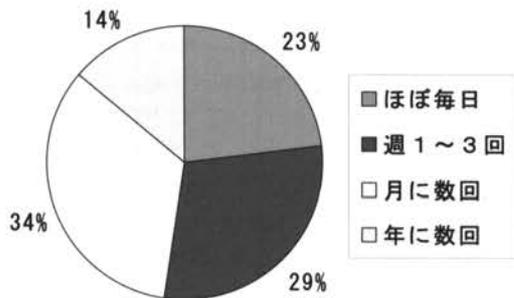


図 4-11 2011 年最も利用する機械翻訳ソフト・サービスの利用頻度

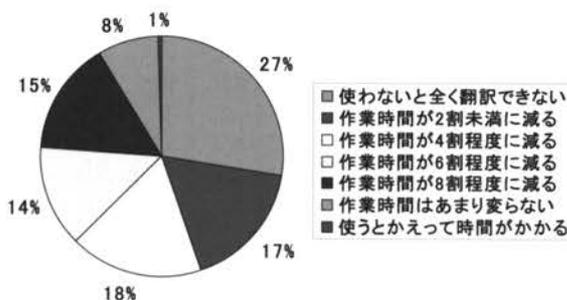


図 4-12 2010 年最も利用する機械翻訳ソフト・サービスによる作業時間短縮

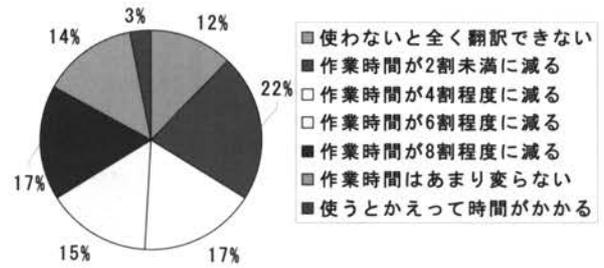


図 4-13 2011 年最も利用する機械翻訳ソフト・サービスによる作業時間短縮

図 4-13、4-14 に、翻訳対象の分野について、2010 年と 2011 年の結果を示す。翻訳を必要とする人が 2010 年調査で 745 人、2011 年で 56 人であることを考慮すると、2011 年は産業、社会科学、自然科学などの割合が高い。回答者の層の違いがここにも現れていると言える。

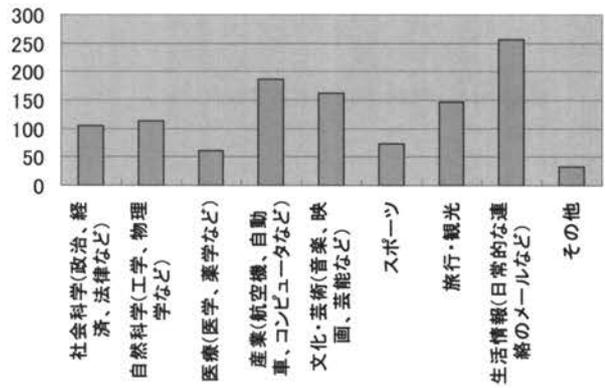


図 4-13 2010 年最も利用する機械翻訳ソフト・サービスにおける翻訳分野 (複数選択可)

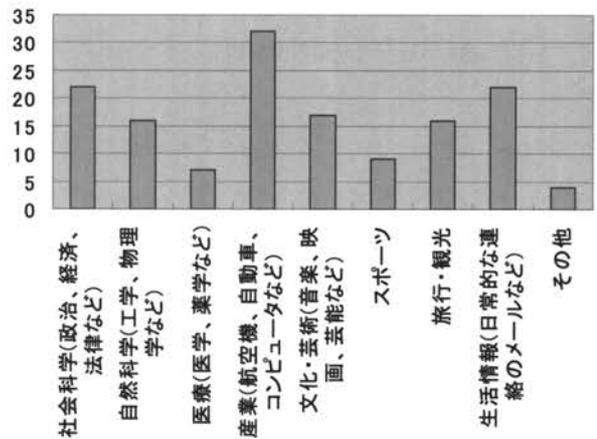


図 4-14 2011 年最も利用する機械翻訳ソフト・サービスにおける翻訳分野 (複数選択可)

## 5. 利用者辞書の利用についての調査結果

本節では、利用者辞書の利用についての調査結果を示す。

図 5-1 に、利用者辞書に語（文）を登録しているかどうかの割合を示す。前回の 2010 年調査では、登録していない人が 82%、登録している人が 18%だった。この数字の変化も、回答者層の変化や市販の PC ソフト利用者が増えたこと（無償の翻訳サービスは通常辞書登録機能を持たない）に依るものだろう。

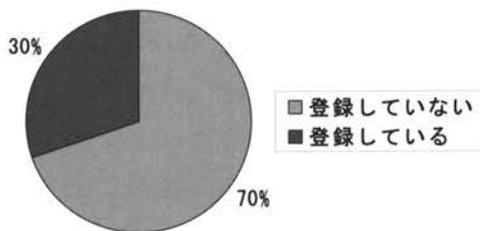


図 5-1 利用者辞書に語(文)を登録しているか

図 5-2、図 5-3 に利用者辞書に語（文）を登録していない人にその理由を聞いた結果を、2010 年と 2011 年について示す。「登録作業が分からない」が 10%減り、「登録作業が面倒」が 9%増加している。2011 年は回答者層の変化を受けて、利用者辞書の有効性は認識しているが、「登録作業が面倒」という人が増えたということである。機械翻訳において、利用者辞書への語（文）登録は、利用者側でできる数少ない翻訳品質向上策の一つだが、機能の使いやすさという点で工夫すべき点がある機能ということがわかるという過去の調査からの結論と合致する結果となった。

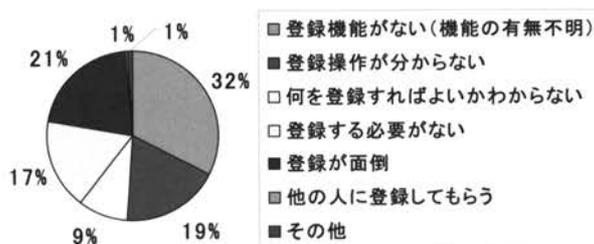


図 5-2 2010 年利用者辞書に語(文)を登録していない理由

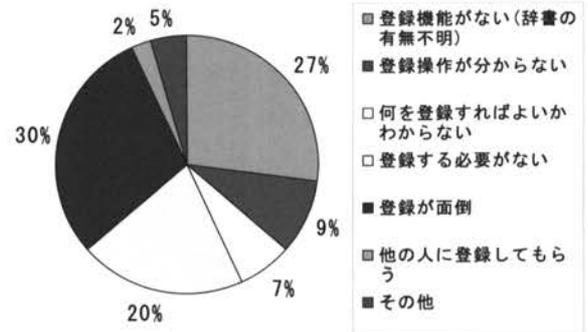


図 5-3 2011 年利用者辞書に語(文)を登録していない理由

次に、図 5-4、図 5-5 に、利用者辞書を利用している人が利用者辞書に登録している語（文）数を、2010 年と 2011 年について示す。2010 年では 100 語以上の登録が半数に達しないのに 2011 年では 70%を占めていること、2010 年は 1,000 語以上の大量登録が約 15%だったのが、2011 年では 50%に達することなどが大きく目を引く。上級者ほど利用者辞書に真剣に手間暇をかけた使っていると言える。

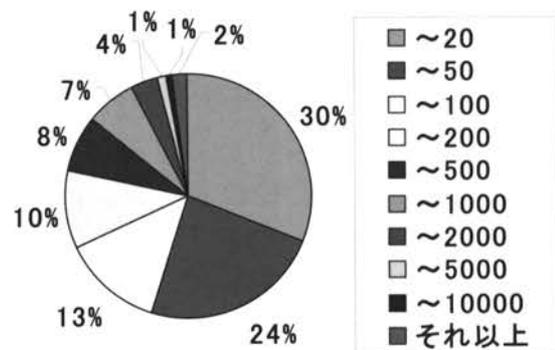


図 5-4 2010 年利用者辞書に登録している語(文)数

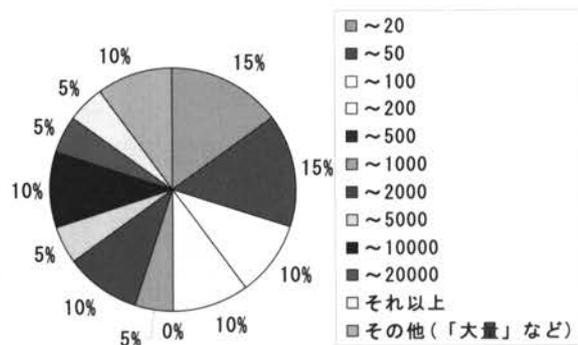


図 5-5 2011 年利用者辞書に登録している語(文)数

次に、図 5-6、図 5-7 に、利用者辞書の利用と作業効率の向上の割合の結果を、2010 年と 2011 年について示す。2010 年に比べて 2011 年は改善度合いが全般的に大きく向上している。たとえば、2010 年の結果では 8 割以上の作業効率向上が 10% だったのに対し、2011 年の結果では 40% 近くに達している。また、「あまりかわらない」が 2011 年は 2010 年と比べて 18% から 10% に減っているのも目につく。2011 年の調査では、利用者辞書への登録をまじめに行っている回答者が多く、そのことが反映している結果と言えるだろう。なお、利用者辞書の利用によって「かえって悪くなった」という回答は皆無であったことは、2010 年も 2011 年も同じである。

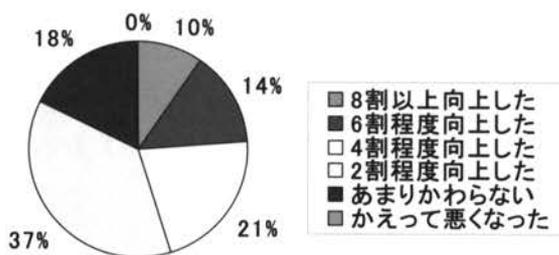


図 5-6 2010 年利用者辞書登録前と後での作業効率の変化

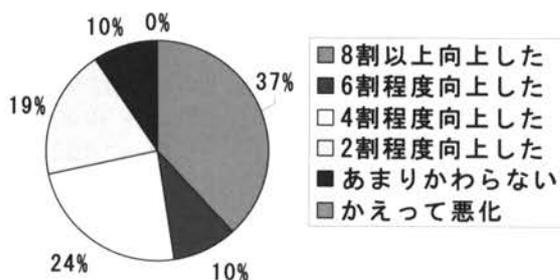


図 5-7 2011 年利用者辞書登録前と後での作業効率の変化

図 5-8 に、AAMT の UTX 辞書についての認知度を示す。これは、今回新規の設問である。今回は回答者に翻訳者が多いにも関わらず、認知度が余り高いとは言えない。ただし、UTX 辞書に対するコメントには、好意的なものや、利用者辞書の使い方の難しさ（連語の登録に難がある、業界や企業によって用語が違う、翻訳ソフト

による品詞体系・属性情報の違い）を問うコメントが多く、関心の高さが窺われる。

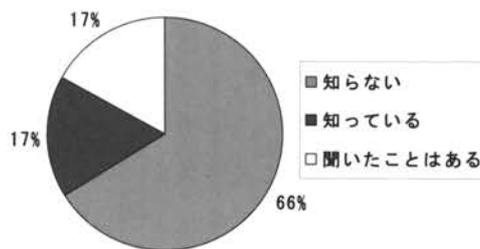
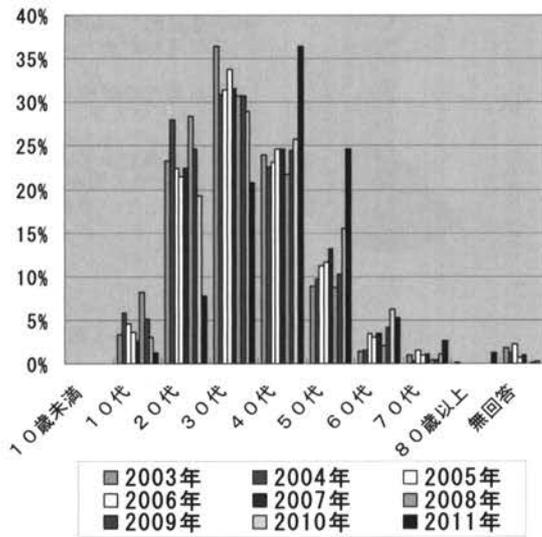


図 5-8 UTX 辞書の認知度

## 6. 回答者に関する調査結果

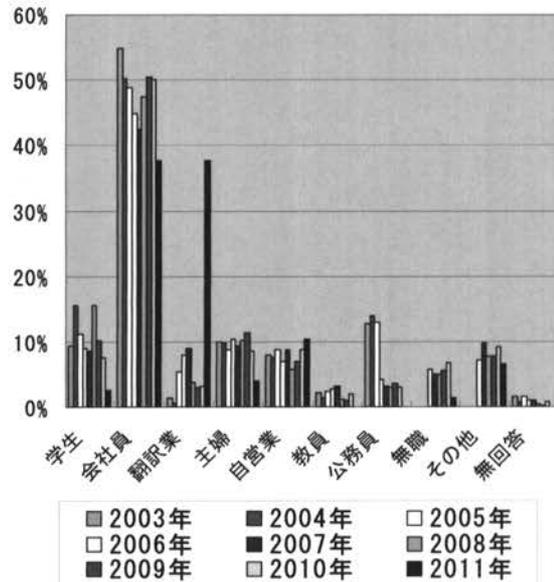
本節では、回答者の基本属性（年齢、性別、職業、英語力など）を示す。

図 6-1 に年齢、図 6-2 に性別、図 6-3 に職業、図 6-4 に英語能力（横軸は TOEIC スコア）毎の回答者の割合を、過去の年ごとに示す。なお、グラフがみづらいものについては、表も付す。英語能力が例年ほぼ 10% 以下 TOEIC スコア 750 点以上の層が 50% を超えたこと、また、例年 10% 以下の翻訳業が 40% 近くになっているのが大きな特徴である。



	2003	2004	2005	2006	2007	2008	2009	2010	2011
10歳未満	0%	0%	0%	0%	0%	0%	0%	0%	0%
10代	3%	6%	5%	4%	3%	8%	5%	3%	1%
20代	23%	28%	22%	21%	22%	28%	25%	19%	8%
30代	36%	31%	31%	34%	32%	31%	31%	29%	21%
40代	24%	23%	23%	25%	25%	22%	24%	26%	36%
50代	9%	10%	11%	12%	13%	9%	10%	15%	25%
60代	1%	2%	3%	3%	3%	2%	4%	6%	5%
70代	1%	0%	2%	1%	1%	0%	0%	1%	3%
80歳以上	0%	0%	0%	0%	0%	0%	0%	0%	1%
無回答	2%	1%	2%	1%	1%	0%	0%	0%	0%

図 6-1 年齢



	2003	2004	2005	2006	2007	2008	2009	2010	2011
学生	9%	15%	11%	9%	9%	16%	10%	8%	3%
会社員	55%	50%	49%	45%	42%	47%	51%	50%	38%
翻訳業	1%	1%	5%	8%	9%	4%	3%	3%	38%
主婦	10%	10%	9%	10%	9%	10%	11%	9%	4%
自営業	8%	8%	9%	7%	9%	6%	7%	9%	10%
教員	2%	1%	2%	3%	3%	1%	1%	2%	0%
公務員	13%	14%	13%	4%	3%	3%	3%	3%	0%
無職	0%	0%	0%	6%	5%	5%	6%	7%	1%
その他	0%	0%	0%	7%	10%	8%	8%	9%	6%
無回答	2%	1%	2%	1%	1%	0%	0%	1%	0%

図 6-3 職業

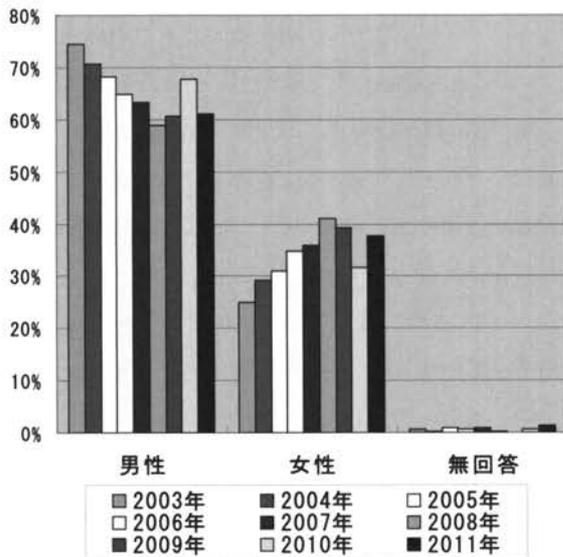


図 6-2 性別

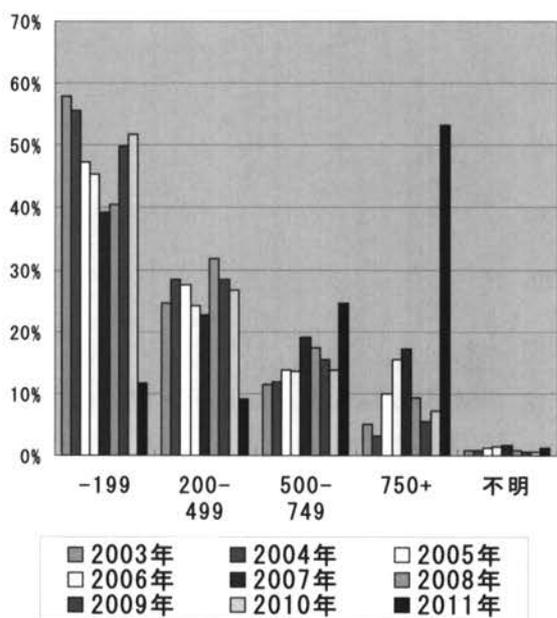


図 6-4 英語能力

また、図 6-5、図 6-6 に、今回新規の設問として、翻訳への関わり方、翻訳ソフト・サービスに求める翻訳品質について質問した結果を示す。これらも、回答者層の変化によっては異なる結果が出るのが想定される。

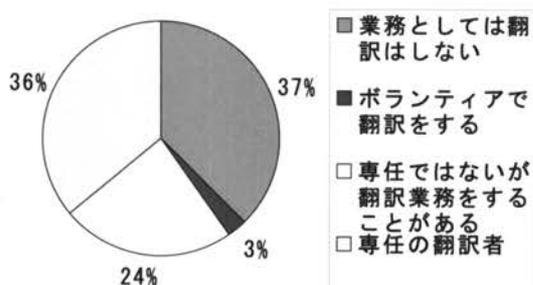


図 6-5 翻訳にどのように関わっているか

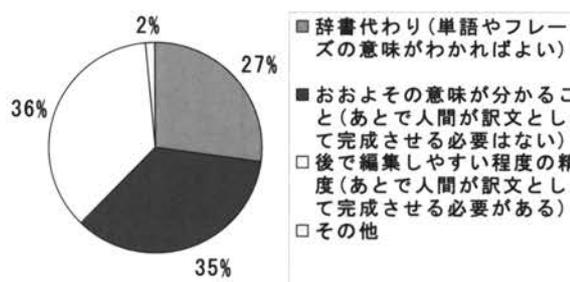


図 6-6 翻訳ソフト・翻訳サービスに求める翻訳品質

## 7. まとめ

機械翻訳に関するアンケートを実施した。本年は回答者激減という想定外の事態に陥ったが、回答者層の変化を通じて、新しい知見を得ることができた。

- 翻訳ソフト・サービスについては、翻訳品質など、不満が多々あるものの、仕事に、また、プライベートに、定常的に利用する人も多く、作業時間の短縮については一定の効果を認められている
- 利用者辞書への語(文)の登録については、英語能力の高い人ほど真剣に取り組んでおり、それにしたがって作業効率向上も高まっている。ただし、利用者にとってのハードルの高さなどの面で課題があると思われる。

今後、機械翻訳研究者/ベンダーに上記のアンケートに基づいた研究・開発・販売を期待すると共に、今後も機械翻訳技術利用の現状やニーズに係る調査を続け、機械翻訳をより役立つものとしていきたい。

また、2011年から追加した「翻訳ソフトの有効的な利用方法の提案」については、著作権上の問題のない方法で別の機会に報告したい。

## 第6回 AAMT 長尾賞を受賞して

機械翻訳課題調査委員会ワーキンググループ3  
共有化・標準化ワーキンググループ (UTX チーム)  
リーダー：秋桜舎 代表 山本 ゆうじ



### 受賞のお礼

先日の AAMT 総会にて、AAMT 機械翻訳課題調査委員会、共有化・標準化ワーキンググループが第6回 AAMT 長尾賞を受賞しました。「ワーキンググループ3」と言った場合、説明が必要になるので、短く「UTX チーム」と呼んでいます。小文にて、受賞報告とさせていただきます。

受賞理由は以下の通りです。「翻訳支援のためのシンプルでオープンな辞書仕様 UTX を開発、公開し、機械翻訳精度の向上ばかりでなく、辞書の interoperability および翻訳支援へも有効な手段を広く提供している。今後、ローカライゼーション、オープンソース、教育、行政、医療、法律などのさまざまな分野で活用が期待されるとともに、国際標準となる素地を備えている点で高く評価できる。」

長尾真先生、そして UTX チームの活動を認めて、さまざまな形で支えていただいた皆様、月例会合で毎回会場をご提供いただいているインターグループ様に、改めて心からお礼申し上げます。

### UTX チームのメンバー

UTX チームのメンバー各位の貢献について簡単

にご紹介します。富士通研究所の大倉委員には、言語処理学会、LISA China Focus、そして最近では 2011 年 9 月にアモイで行われた MT Summit XIII など、多くの国内外の会議で、UTX について発表いただいています。東芝ソリューションズの島津委員には、貴重なご意見とともに、月例会合で、毎回、議論を手際よくまとめて議事録に残していただいています。ラーニング コンサルタントの加藤委員には、ボイスカウト専門用語辞書への取り組みと、英語についてのアドバイスをいただいています。ワーキンググループ前リーダーである、沖電気工業の村田委員には、標準化に関する経験豊富なお立場から貴重なアドバイスをいただいています。南洋理工大学（シンガポール）の Bond 委員には、ハワイでの国際会議発表、UTX 辞書および変換ツールのご提供、そして英語についてのアドバイスをいただいています。インターグループの伊藤委員には、月例会合の会場やプロジェクターの手配でお世話になっています。

UTX については、こちらのページで詳細を解説しています。仕様と無料辞書もダウンロードできます。あわせてご覧ください。

<<http://www.aamt.info/japanese/utx/>>

## 翻訳現場と研究・開発者のずれを埋める

プロ翻訳者の支援をする「翻訳支援」という点では、翻訳現場と研究・開発者のずれは依然として大きなものがあります。機械翻訳研究・開発者の方にとって、プロの翻訳現場は、縁がないことも多いでしょう。翻訳現場を知らなくても、研究・開発は可能です。しかし、翻訳現場での知見や、翻訳者の発想は、研究・開発に本来、役立てることができるはずで、UTX チーム メンバーにもお勧めしていることですが、自社で翻訳ソフトを開発されている企業の方は、ぜひその**自社製品で、翻訳作業を試していただければ**と思います。必ずなにか感じることがあるはずで、そして、なぜ UTX が必要なのか理解していただけるはずで、

私は、2011 年 2 月に LISA Open Standards Summit (ボストン) で UTX について発表し、各国の翻訳業界の方と意見交換を行いました。その際、改めて日本の翻訳業界がガラパゴス状態であることを実感しました。欧米では、企業はもちろんですが、政府機関で、翻訳支援技術に関する理解と意識が高いことが特徴です。Linport <<http://linport.org/>>という新しい翻訳工程の標準化のプロジェクトも推進されています。ただ、このような高度な標準化は、翻訳メモリー、用語集、スタイル ガイドなどが整備されている企業や組織でこそ意味を持ちます。翻訳発注企業でさえ、自社製品の用語集を持っていないことが多いのが、日本の現実です。統計機械翻訳では用語集は不要になるのでしょうか？ 企業の翻訳では、なにをどう訳すかという基準は、どれだけ統計機械翻訳が進歩しても変わりません。つまり、「用語集」は、今後とも企業での翻訳が必要です。

11 月 29 日には、JTF (日本翻訳連盟) 翻訳祭が開催されます。<<http://goo.gl/LS5Mv>>

機械翻訳研究・開発者の方が、翻訳の現場に触れる貴重な機会です。AAMT はブースを出展しますが、これは、非常に有意義と思われます。私も、UTX について発表を行います。

## UTX は用語集作成の最適な出発点

UTX は、用語集作成の最適な出発点です。翻訳現場で使われている SDL の製品に、用語管理ツール MultiTerm があります。ただし、用語集のない状態から MultiTerm を活用するのは容易ではありません。先日、MultiTerm と Excel の間の同期ツール Excelling MultiTerm という製品を見つけました。Kaleidoscope というオーストリアの会社の製品で、フリーランス翻訳者用は 300 ユーロ、企業向けは 1500 ユーロします。このことから 2 つのことが分かります。まず、海外では、このようなニッチ的な製品が存在するほど、用語管理ツールの活用が進んでいること。そして、MultiTerm よりも Excel のほうが用語集を管理・作成しやすいということです。UTX では、Excel で編集できながら、Excel 形式そのものよりも汎用性があり、さまざまな翻訳ソフトで使えます。

UTX は、今後とも、翻訳現場に即した、実用的な仕様であることを目指しています。私事ながら、インターグループで、2011 年 9 月に翻訳ツールの講習を実施させていただきました。その際に、UTX の想定ユーザーに近い受講生の方に UTX をご紹介しました。このような取り組みを続け、翻訳現場ですばやく用語集作成をすることで、専門用語を正確に訳す UTX の地盤が固められます。

現在、UTX チームでは、多言語をサポートする次バージョン 1.20 の仕様を文書化する作業を行っています。UTX を活用できる具体的な翻訳プロジェクトをご存じの方は、ぜひ AAMT 事務局までお知らせください。<[aamt-info@aamt.info](mailto:aamt-info@aamt.info)>

なお、Facebook で、どなたでも気軽に参加できる「翻訳工学」のページを立ち上げました。ぜひ「いいね」をクリックしてご参加いただき、UTX、専門用語集作成などのトピックを含め、翻訳工学全般に関する、開かれた情報交換、意見交換の場としてご活用いただければと思います。

<<http://goo.gl/wtktf>>

## 予防と未然防止

東京情報大学大学院 総合情報学研究科

教授 畠中 伸敏

### はじめに

本稿では失われた日本人の心を、「予防と未然防止」の日本語訳を通して、訳の整合性で垣間見た。執筆者の独善に満ちているかも知れないが、成否の判断は読者に委ねる。

ところで、本稿を一読されて物足りないと思われた方は、筆者の著作になる次の二冊も、お読み頂けると、さらに内容をご理解頂けるものと思う。

- ①「環境配慮型設計—エコが征す商品開発（仮）」単著 畠中伸敏，日科技連盟出版社。近々に発売予定。
- ②「予防と未然防止（仮）」編著 畠中伸敏，日本規格協会，3月発売予定。

### 1. 崩れる日本の安全・安心神話

日本には「津波」という言葉がある，記録のある最古の例は貞観11年5月26日(869年7月13日)に発生した貞観地震で，慶長16年10月28日(1611年12月2日)に発生した慶長三陸地震は津波では最大高さは20mである。その後も，明治三陸地震(1896年)，昭和三陸地震(1933年)を経験し，津波の最大高さは38.2m，28.7mである。英語には「津波」の訳はなく，ラフカディオ・ハーンが「Tsunami」という語を用いた。中国にも該当する言葉が無く，日本で生まれた言葉である。それだけ，日本では「津波」を経験している。

平成23(2011)年3月11日14時46分に，太平洋三陸沖を震源として，マグニチュード9.0，最大震度7の東北地方太平洋沖地震が発生した。関東大震災のマ

グニチュード7.9を上回り，世界でも4番目の大地震となった。岩手県大船渡の津波の最大高さは30mを超えた。2005年11月15日の三陸沖の地震では，福島第一原子力発電所は持ちこたえて，原子力発電所の威厳を保った。今回の大地震での津波では，原子炉の冷却機の予備電源を支える燃料タンクを海岸近くに置いていたために津波により破壊され，原子炉の冷却が不能となった。冷却不能となった原子炉が温められ，水素を発生し水素爆発を起こした。放射性物質を放出し，約200km離れた東京の金町の取水場の水は，放射性物質が210ベクレルの許容範囲を超えて汚染され，東京都は乳幼児用に自然水4本を配布した。千葉県でも流山市にある千葉北浄水場，千葉市柏井浄水場も汚染された(130ベクレル)。そればかりか，建築ミスにより，予備電源として東北電力からケーブルを引き込んでいたが，津波により切断された。

原子力を設計した人は，「今回の津波は想定外である」と主張したが，過去の津波の経験が生かされているとは思えない。安全・安心は何によりも優先されるべきもので，完全なエンジニアのミスで明らかに人災だと主張するものがある。

コンピュータのプログラムを作成するときに，「最適化」という言葉があるが，安全・安心は最適化されるものではなく，なによりも優先されるべきもので，設計の安全率の考えが，いつのまにか喪失していた。

### 2. 予防と未然防止

#### 2.1 日米とEUとの考えの違い

防止と未然防止は，なんら定義することなく使用される場合が多い。当たり前のこととして使用されてい

て、誰も疑問を挟まないが、アメリカ人とヨーロッパ人では、少し考え方が違う。EC 条約第 175 条の手続きで採択された廃電気電子機器(WEEE: Waste Electrical and Electronic Equipment, ウィーと発音)指令では、言葉として防止の用語が使用され、一見して、予防と防止の区別が付かないが、予防の概念が導入されている。予防(Precaution)は発生確率が未知で被害や危害の程度が分からない時に使用する。例えば、地震のために、懐中電灯や非常食を用意する。また、防止(prevention)は発生確率が既知で、被害や危害がある程度、知りうる状態の時に使用される。例えば、自動車が来るので、左右を確認して道路を渡る。

具体的には、EC 条約第 95 条の手続きで採択された有害物質使用制限 (RoHS 指令: Restriction of Hazardous Substances, ローズと発音)指令にあるプラスチックの材料として使用される難燃剤の PBDE(ポリ臭素化ジフェニルエーテル)の取扱いである。アメリカ人は、難燃剤の PBDE と人体との因果関係が証明されていないことから、火災の発生の方が重要とし、PBDE の使用は、具体的な使用禁止の措置を講じていない。一方、EU の RoHS 指令では、PBDE は人体への蓄積性があることから、平成 18 年(2006 年)以降に上市する電子機器や電気製品に含有することを禁止している。なお、EC 条約第 95 条が選択されると、EU 指令より厳しい国内法は制定できない。また、EC 条約第 175 条が選択されると、環境保全そのものが目的であり、EU 指令より厳しい国内法が制定できる。

ほんの 30~40 年前は、海は広く、空は限りなく広いことから、生産段階で出る残さや使用済みの廃液は海や川に流し、煙は空に排気していた。その結果、地球温暖化の人類史上最大の公害をもたらし、後世の人間が、その付けを払うことになる。

経営者の中には、電子機器や電気製品から煙が出て火事になっても、どうせ製造物責任の問題は、お金で形がつくものとして、何%かの確率で発生する火事まで、未然に手を打つ必要がないと考える経営者がいる。

しかし、人体上の被害については、お金で賄えるものではない。一旦、人的被害が発生すると、営々と築き上げた企業としての信用が失墜し、企業イメージを低下させる。

## 2.2 リオ宣言の第 15 原則(予防原則)

EU の環境規制に関する基本的考え方は、平成 4 年(1992 年)、リオで開催した気候変動枠組条約(UNFCCC: United Nations Framework Convention on Climate Change)のリオ宣言の第 15 原則(予防原則: Precautionary principle)で、環境を保護するための予防原則が宣言された。

### Rio Declaration on Environment and Development

#### Principle 15 <sup>[1]</sup>

In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation.

#### リオ宣言の第 15 原則(予防原則) <sup>[2]</sup>

環境を保護するため、予防的方策は、各国により、その能力に応じて広く適用されなければならない。深刻な、あるいは不可逆的な被害のおそれのある場合には、完全な科学的確実性の欠如が、環境悪化を防止するための費用対効果の大きい対策を延期する理由として使われてはならない。

(環境庁・外務省訳)

予防原則の宣言中の「完全な科学的確実性の欠如」により、対策が遅れ被害が増大した事例には、日本の

水俣病がある。

リオ宣言の第15原則の予防原則は、EUと日米では若干、違い、EUは積極的で、1970年代には、西ドイツの環境政策で「予防的に環境保全を行う」ことを目的とした。2002年のヨハネスブルクサミットの化学安全に関する宣言においては、日米は予防原則の言葉の使用を拒否し、若干、ニアンスを変更するために、予防の方策（precautionary approach）の文言の使用を提案した。予防原則そのものの概念には、潜在的に有害な活動は阻止すべきという概念があり、これは「疑わしきは罰す」論的と揶揄された。最終的には、科学的確実性が不完全な場合でも、科学的にできるだけ可能な評価を行い、受入れ可能なリスク水準を配慮した上で、予防原則を選択肢とすることに合意した。

米国は予防原則の定義に慎重な態度を取ってきたが、1998年米国のウィスコン州ウィングスブレッド会議センターで開催された「予防原則に関するウィングスブレッド会議の宣言」では、予防原則の実行方法と実行への障壁について、危害への脅威、科学的な不確実性、予防行動の3つの要素を取り上げ、次の予防原則を宣言した。現在のところ、予防原則の定義とされている。

#### 予防原則(Precautionary principle)<sup>[3]</sup>

When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause-and-effect relationships are not fully established scientifically.

活動が人の健康と環境に対して危害を及ぼすおそれがあるときには、たとえ、その因果関係が科学的に十分立証されていないとしても、予防的手段が行われるべきである。

(環境庁・外務省訳)

#### 2.3 予防と未然防止の違い

未然防止については、地球規模の汚染が問題となり、1972年、ストックホルム国連環境会議の宣言で、「防止」が謳われた。

#### 未然防止(Preventive action)<sup>[4]</sup>

##### Declaration of the United Nations Conference on the Human Environment

##### Principle 7

States shall take all possible steps to prevent pollution of the seas by substances that are liable to create hazards to human health, to harm living resources and marine life, to damage amenities or to interfere with other legitimate uses of the sea.

##### 人間環境宣言 2.原則 【海洋汚染の防止】

(7)各国は、人間の健康に危険をもたらす、生物資源と海洋生物に害を与え、海洋の快適な環境を損なわない、海洋の正当な利用を妨げるような物質による海洋の汚染を防止するため、あらゆる可能な措置をとらなければならない。(環境庁・外務省訳)

一般的には、不可逆的な蓋然性と不確実性がある場合は、「予防」の対象であり、明らかな知見が得られる場合は、「防止」の対象であるといわれている。

環境制約の予防の観点は、科学的知見によらず、深刻なあるいは不可逆的な被害の恐れのある場合には、何らかの対策を講じることを必要としている。

ところで、品質マネジメントシステムの規格のISO 9001の2008年版では、preventive action (ISO 9001では予防処置と訳される)は、“action to eliminate the cause of potential nonconformity or other undesirable potential situation”と定義され、「起こり得る不適合又はその他の望ましくない起こり得る状況の原因を除去するための処置」<sup>[5]</sup>と訳されている。一

一般的に、「未然防止」は、クレームや不具合が発生しないようにあらかじめ処置することと考えられる。つまり、科学的な知見に基づいて、不具合やトラブルが発生しないように、あらかじめ原因を突き止めて処置することである。

また、製造物責任法第4条では、「当該製造物が引渡した時における科学又は技術に関する知見によっては、当該製造物にその欠陥があることを認識することができない」場合は免責事由となる。製造物責任(PL)の開発行為の抗弁と称し、製品を生産した段階で調査した結果、科学的知見が得られないとすると、生産者の責任が免れる。

このように、「予防」という概念は、「未然防止」の概念よりは広い意味で捉える必要があり、概念は似ているようで異なる。

## 2.4 日本の考え方(PRTR法)

日本国内では、「特定化学物質の環境への排出量の把握等及び管理の改善の促進に関する法律」(PRTR法:Pollutant Release and Transfer Register法)が、平成13年(2001年)3月30日に施行され、「国等による環境物品等の調達の推進等に関する法律」(グリーン購入法)が平成13年(2001年)4月に施行された。

有害物質の届出を促進する制度のPRTR法の考え方は、アメリカ流儀に近い。「科学的知見の状況を踏まえ、環境の保全上の支障を未然に防止することを目的」<sup>[6]</sup>とし、「未然防止」に主眼を置いている。

### PRTR法第一条(目的)<sup>[6]</sup>

この法律は環境の保全に係る化学物質の管理に関する国際的協調の動向に配慮しつつ、化学物質に関する科学的知見及び化学物質の製造、使用その他の取扱いに関する状況を踏まえ、事業者及び国民の理解の下に、特定の化学物質の環境への排出量等の把握に関する措置並びに事業者による特定の化学物質の性状及び取扱いに関する情報の提供に関する措置等を講ずることに

より、事業者による化学物質の自主的な管理の改善を促進し、環境の保全上の支障を未然に防止することを目的とする。

(太字と下線部分は著者が付記)

同第二条第二項で、「第一種指定化学物質」を定義し、同第5条第二項では、第一種指定化学物質の排出量と移動量に関して、主務省で定める事項の届け出の促進を法制化している。同第一四条では、指定化学物質等を他の事業者提供あるいは譲渡するときは、当該指定化学物質の性状及び取扱いに関する情報を文書または磁気ディスクでの交付の促進を法制化している。

また、虚偽の申告に対しては罰則を定めているが、制度として考えられ、企業の自主的な活動を主体として、虚偽の申請以外は、届け出をしないからといって、具体的な罰則の取り決めがない。

## 3. 日本人の心の喪失

ところで、EUの環境管理の基本原則は次の四つからなる<sup>[7]</sup>。

- ①「予防原則(Precautionary Principle)
- ②「未然防止原則(Prevention Principle)
- ③「発生源での対応原則(Ratification At Source Principle)
- ④「汚染者負担原則(Polluter Pays Principle)

それぞれ、原文が日本語に訳されているが、IEC(国際電気標準会議)の環境配慮設計WG国際主査の市川芳明氏は、Prevention Principleを事前防止原則と訳し、Precautionary Principleを予防原則と訳した。

本稿では、Preventive actionを未然防止と訳したが、これを、品質マネジメントシステムの規格のISO 9001の2008年版の訳に従うと、preventive actionを予防処置と訳し、Precautionary Principleに対応する相応しい訳の創出が難しく、整合性を失う。

したがって、Precautionary Principleは予防原則と

訳し、Preventive action は未然防止と訳すことが妥当と考える。なお、未然是“まだ起こらないこと”であり、事前は“何かが起こる前”であるから、事前防止原則と未然防止原則は同義である。

この四つの原則はワンセットであり、予防原則だけが必要であるとか、未然防止原則さえ実施していればよいというものではない、未然防止原則と予防原則の両方の概念が対となるものである。

これはヨーロッパ流の考え方ではあるが、日本人の遺伝子には津波などの被害に遭った経験が、先祖代々、刻みこまれて来た筈である。しかし、日本人は米国と永く付き合ううちに、いつのまにか、「備えあれば憂いなし」の考え方を喪失し、福島第一原発の事故を引き起こしたものと推測する。

#### 参考文献

- [1] The United Nations Conference on Environment and Development, *“Rio Declaration On Environment And Development”*, June 1992.
- [2] 環境庁・外務省監訳, “アジェンダ 21 実施計画’97”, エネルギージャーナル, 1997.
- [3] The Science and Environmental Health Network, “THE PRECAUTIONARY PRINCIPLE IN ACTION A HANDBOOK”, 1998.
- [4] 環境省訳, “国連環境会議(ストックホルム会議: 1972年)人間環境宣言”, 環境省, 1972.
- [5] 品質マネジメントシステム規格国内委員会監修, 「ISO 9001:2008」, 日本規格協会, 2009.
- [6] 青山義充他 7 名, 「小六法」, 有斐閣, 2004.
- [7] 市川芳明編著, 「EuP 指令入門」, 社団法人産業環境管理協会, 2006.

# AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第九回を迎えることができました。今号では、法人会員一社、個人会員一名の皆さまからのご寄稿をいただいております。

独自のお取組みのご紹介、機械翻訳研究への提言、AAMT の活動へのご要望など、今回も貴重なご意見をお寄せいただきました。

AAMT Journal では今後も引き続き、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。

ご寄稿・お問い合わせは AAMT 事務局(E-mail: AAMT-info@AAMT.info)まで宜しくお願いいたします。

---

## 法人会員（敬称略・50 音順）

---

### 会員名

株式会社サン・フレア/SunFlare Co., Ltd.

---

### 自己紹介

株式会社サン・フレアは、最先端の産業科学技術と文化の翻訳事業で培われた経験と実績を基に、創業 40 周年を迎えた今年、これまでの翻訳事業に加え、ドキュメントの作成、デザイン、コンサルテーションなどの総合的なドキュメンテーション事業を基軸とするビジネスモデルを確立し、「ドキュメント総研®」として新しい一歩を踏み出しました。これに伴い、これまでの当社の取り組みを総括し、力強く躍進するイメージの新コーポレートブランドロゴマークを発表します。また、新しいビジネスコンセプトである「ドキュメント総研®」を具現化する試みとして、Web サイトをリニューアルしました。<http://www.sunflare.com/>

私たちの強みは、多様な分野の専門性を備える「人財」です。さまざまな業界・業種にてドキュメンテーション実務に携わってきたエキスパートが、ドキュメント制作のライフサイクルのあらゆる場面で生じる課題に対し、実現可能なソリューションを提供します。「言語力」をコアコンピタンスとする「提案力」、「デザイン力」、「IT 力」、「教育力」の 5 つの「力」で、目的を果たすドキュメンテーションを実現します。

## MT/翻訳とのかかわり

### MT および翻訳業界に期待すること

かつて機械翻訳システム開発の最先端の一翼を担っていると自負していた立場から、現在はその経験を活かしそれを利用する立場に、その立ち位置を変えている一人として機械翻訳について日頃感じている個人的見解を述べてみたいと思います。

身近な自然言語処理技術のひとつとして、仮名漢字変換があります。今から 30 年程前、それは専用の装置、いわゆる日本語ワードプロセッサ（ワープロ）として存在していました。当時、私はその生みの親といわれる某総合

電気メーカーの研究所に入社配属されました。研究所の片隅には、その1号機と同型の机大の古めかしい大型装置がまだ現役として鎮座していました。すでにワープロは小型テレビ大の大きさのものが主流の時代ではありませんでしたが。さらにはキーボード大のワープロや、現在のノート PC よりも小型の試作機なども研究所内には存在しており、非常に驚いた記憶があります。大学ではどちらかというと数学に近い分野の形式言語理論の文脈依存言語の構文解析のアルゴリズムを日々考えていた世界とは違い、ものづくりの現場の凄さを実感しました。その後、ワープロは専用装置から日本語入力ソフトウェアとしてパソコンへ標準搭載されるようになり、今は携帯電話にも当たり前の機能のひとつとなっています。仮名漢字変換技術の精度向上を例えていうなら、その処理単位が点（単語）から線（文）へ、線から面（ドキュメント）へと変化していったといえます。すなわち、単漢字変換から文節変換へ、文節変換から連文節変換（一文まるごと変換）へ、そして文脈（話題）を考慮して同音異義語を最適に変換してくれる技術へと進化していきました。

一方、機械翻訳技術はその方式が規則に基づく方式からコーパスに基づく方式へ主流が移行しつつあるように感じます。しかしながら、開発当初から現在までその処理単位ははまだ文単位に留まっているような気がします「与えられた一文の情報のみで相手言語への翻訳を試みる」。今後、更なる翻訳精度の向上には本格的な文脈処理技術を考慮した翻訳技術へと進化していくことを期待しています。現状の機械翻訳システムも以前に比べたら格段に翻訳精度は向上しており、けっして「使えない」のではなく、「上手く使える」ところをみんなで考える必要があるのかも知れません。何かとスピードが求められる昨今、コストを下げその分、量を確保するという「大量短納期」の要求があります。先に述べた計算機への入力手段の進化も一因と考えますが、世界的に処理すべき電子化ドキュメントの量が様々な分野で加速度的に増えてきており、以前に比べて機械翻訳を求めるニーズは潜在的に非常に多くなってきていると考えます。

そのため、今後、機械翻訳を意識したドキュメントの品質管理や用語管理という側面が非常に重要になってくるのは明らかです。そこには多言語への翻訳もスムーズに処理したいという要求が内在しています。すなわち、機械翻訳で培われた言語処理技術の活躍の場は、与えられたドキュメントを単に処理するという立ち位置から「ドキュメント作成の上流工程から翻訳を含む下流工程すべてに関わってくる」ことになるものと確信しています。ワープロ同様に、機械翻訳のみならず、その要素技術（形態素解析、構文解析、用語抽出など）をその手段として利用すること、それが当たり前の時代はもう目の前に来ていると肌で感じています。グローバルな視点でみるとかなりのスピードで翻訳業界も変化しており、日本（語）だけがその流れに取り残されないよう皆様と一緒に頑張って進化していきたいと考えています。弊社も創立40周年を迎えた今年、「ドキュメント総研®」として進化し、新たな一步を踏み出しました。今後ともどうぞよろしくお願い致します。

株式会社サン・フレア 執行役員 安達 久博

## AAMT への要望

設立20周年を迎え、おめでとうございます。本協会のさらなる発展を期待しています。

最近「元気がない」、「変化をあまり好まない傾向が強い」といわれている日本（人）にあって、

AAMT に集う皆様方との交流が、ものすごいスピードで変化している世界にチャレンジする意気込みを常に共感できる場であり続けることを強く望みます。今後ともどうぞよろしくお願い申し上げます。

# 協会活動報告

(2011年6月～2011年10月)

## 第21回通常総会

2011年6月13日

- |            |               |       |               |
|------------|---------------|-------|---------------|
| 第1号議案      | 2010年度事業報告(案) | 第2号議案 | 2010年度決算報告(案) |
| 第3号議案      | 2011年度事業計画(案) | 第4号議案 | 2011年度収支予算(案) |
| その他・会員提案事項 |               |       |               |

### ● 報告会

2011年6月13日

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

### 講演会

○講演:

『Resource sharing for research on large scale NLP, Text Mining and machine Translation』  
辻井 潤一氏(Principal Researcher, Microsoft Research Asia)

### 第6回 AAMT 長尾賞受賞式・記念講演会

受賞者: AAMT (アジア太平洋機械翻訳協会)

機械翻訳課題調査委員会共有化・標準化ワーキンググループ

- ▶ 山本 ゆうじ (リーダー) 秋桜舎
- ▶ 伊藤 肇 株式会社インターグループ
- ▶ 村田 稔樹 沖電気工業株式会社
- ▶ Francis Bond 南洋理工大学 (シンガポール)
- ▶ 島津 美和子 東芝ソリューション株式会社
- ▶ 大倉 清司 株式会社富士通研究所
- ▶ 加藤 マイケル 孝仁 ラーニング コンサルタント

受賞理由: 「ユーザー辞書共通フォーマット UTX」

翻訳支援のためのシンプルでオープンな辞書仕様 UTX を開発、公開し、機械翻訳精度の向上ばかりでなく、辞書の interoperability および翻訳支援へも有効な手段を広く提供している。今後、ローカライゼーション、オープンソース、教育、行政、医療、法律などのさまざまな分野で活用が期待されるとともに、国際標準となる素地を備えている点で高く評価できる。

### 懇親会

2011年6月13日 ホテルアジュール竹芝 12F 白鳳

## 決算理事会

2011年6月13日

- |            |               |       |               |
|------------|---------------|-------|---------------|
| 第1号議案      | 2010年度事業報告(案) | 第2号議案 | 2010年度決算報告(案) |
| 第3号議案      | 2011年度事業計画(案) | 第4号議案 | 2011年度収支予算(案) |
| その他・会員提案事項 |               |       |               |

## 機械翻訳課題調査委員会

2011年7月1日(2011年度 第3回)

- ① 前回委員会の議事録の確認
- 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

2011年8月4日(2011年度 第4回)

- ① 前回委員会の議事録の確認
- 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

2011年9月9日(2011年度 第5回)

- ① 前回委員会の議事録の確認
- 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

2011年10月17日(2011年度 第6回)

- ① 前回委員会の議事録の確認
- 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

## 編集委員会

2011年7月28日(2011年度 第2回)

- ① AAMT Journal No.50の編集進捗状況について
- ② AMT Journal No.50の追加記事の確認・企画
- ③ No.50の企画について

## インターネット WG

- ① AAMT ホームページの更新（毎月）
- ② AAMT Forum メイリングリストの管理
- ③ AAMT Forum への情報発信
- ④ 委員会活動におけるメイリングリストの管理
- ⑤ 会員専用ホームページ更新にむけた準備
- ⑥ その他事務局ネットワークのインフラ管理全般

## AAMT/Japio 特許翻訳研究会

2011年6月10日（金）（2011年度 第3回）

1. 前回議事録の確認
2. 今年度の活動予定について（横山副委員長、江原副委員長）
  - MT サミット・特別セッション関連
  - 特許翻訳ワークショップ関連
3. 各グループの研究内容について（専門用語抽出関連：宇津呂委員他）
4. 研究会ホームページ開設の件（事務局）
5. 次回の開催について

日 時：2011年7月22日（金）（2011年度 第4回）

議事概要：

1. 前回議事録の確認
2. MT サミット関連報告（横山副委員長、江原副委員長）
3. 特別セッション関連
4. 特許翻訳ワークショップ関連
5. Japio のニーズについて（Japio）
6. 研究会ホームページ開設の件（事務局）
7. 次回の開催について
  - 開催の日時
  - 主な議題

日 時：2011年10月14日（金）（2011年度 第5回）

議事概要：

1. 前回議事録の確認
2. MT サミット関連報告（辻井委員長、横山副委員長、江原副委員長、二宮委員、Japio）
3. Extracting Pre-ordering Rules from Chunk-based Dependency Trees for Japanese-to-English Translation（呉研究員）
4. 事後並べ替え型日英統計翻訳（須藤研究員）
5. 特許庁・辞書データ整備に関する調査について（Japio）

## 第21回通常総会および関連行事の報告

AAMT 事務局

当協会の第21回通常総会が2011年6月13日(月)13時20分より・ホテルアジュール竹芝にて開催されました。総会后、各委員会からの報告会、講演会、そして第6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21回通常総会

1. 開会の辞
2. 会長挨拶 豊橋技術科学大学・情報基盤センター教授・井佐原 均
3. ご来賓挨拶
4. 出席会員の確認
5. 議案  
第1号議案 2010年度事業報告(案) 第2号議案 2010年度決算報告(案)  
第3号議案 2011年度事業計画(案) 第4号議案 2011年度収支予算(案)  
その他・会員提案事項
6. 総会後の組織、人事について(案)
7. 閉会の辞

### 報告会

- 開会挨拶 会長・豊橋技術科学大学・情報基盤センター教授・井佐原 均
1. 機械翻訳課題調査委員会 委員長 長瀬 友樹 ((株)富士通研究所)
  2. AAMT/Japio 特許翻訳研究会 副委員長 横山 晶一 (山形大学)
  3. インターネットワーキンググループ 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 :

「Resource sharing for research on large scale NLP, Text Mining and machine Translation」

辻井 潤一氏 (Principal Researcher, Microsoft Research Asia)

## 第6回 AAMT 長尾賞授与式・記念講演会

受賞者： AAMT（アジア太平洋機械翻訳協会）機械翻訳課題調査委員会

共有化・標準化ワーキンググループ

山本 ゆうじ（リーダー） 秋桜舎

伊藤 肇 株式会社インターグループ

村田 稔樹 沖電気工業株式会社

Francis Bond 南洋理工大学（シンガポール）

島津 美和子 東芝ソリューション株式会社 大倉 清司 株式会社富士通研究所

加藤 マイケル 孝仁 ラーニング コンサルタント

受賞理由：「ユーザー辞書共通フォーマット UTX」

- ▶ 翻訳支援のためのシンプルでオープンな辞書仕様 UTX を開発、公開し、機械翻訳精度の向上ばかりでなく、辞書の interoperability および翻訳支援へも有効な手段を広く提供している。今後、ローカライゼーション、オープンソース、教育、行政、医療、法律などのさまざまな分野で活用が期待されるとともに、国際標準となる素地を備えている点で高く評価できる。

選考委員長：飯田 仁（東京工科大学）

選考委員：横山 晶一（山形大学）

Virach Sornlertlamvanich（タイ NECTEC）

Key-Sun Choi（韓国 KAIST） 宇津呂 武仁（筑波大学）

隅田 英一郎（情報通信研究機構）

推薦者：井佐原 均（豊橋技術科学大学）

同意人：吉見 毅彦（龍谷大学）



授賞式

### 懇親会

本会後の懇親会は、多数の参加者にお集りいただき、学識経験者、研究者、翻訳家等、幅広い分野の方々が活発な意見交換を行い有意義な交流の場となりました。ご参加誠に有難うございました。

AAMT ジャーナル編集委員会委員長  
筑波大学 システム情報系 知能機能工学域  
宇津呂 武仁

AAMT ジャーナル 50 号をお送りします。

今号の巻頭言は、東芝研究開発センターの出羽達也様よりご寄稿頂きました。

今号では、第六回の AAMT 長尾賞の選考結果を受けまして、受賞者である AAMT 機械翻訳課題調査委員会 共有化・標準化ワーキンググループ(UTX チーム、代表：秋桜舎山本ゆうじ様)より、受賞理由となった「翻訳支援のためのシンプルでオープンな辞書仕様 UTX」の活動についての紹介記事をご寄稿していただきました。

また、9月にアモイ(中国)におきまして開催されました第 13 回機械翻訳サミット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。

その他、海外からは、中国科学院の Qun Liu 教授より、"Machine Translation in China"と題しまして、中国における機械翻訳の歴史について、また Common Sense Advisory, Inc.の Donald A. DePalma 博士より、"Japan and the Global Language Services Market"と題しまして、世界および日本の翻訳市場の状況について、それぞれご寄稿いただきました。また、国内からは、日本翻訳連盟 標準スタイルガイド検討委員会委員長の田中千鶴香様より、「翻訳時の日本語表記に対する取り組み」について、ご寄稿していただきました。

一方、AAMT 内の活動報告として、機械翻訳課題調査委員会から、英日・日英機械翻訳に関するアンケート結果報告についてご寄稿していただきました。その他、「AAMT 会員のひろば」の企画におきましては、個人会員の紹介文 1 件、法人会員の紹介文 1 件を掲載しました。

# AAMT

Asia-Pacific Association for Machine Translation

## AAMT 入会のご案内

AAMT は、機械翻訳の発展を目的として、機械翻訳の研究者、開発者、製造者、利用者が集まった任意の組織です。委員会による定期的な調査研究をはじめ、機関誌の発行、シンポジウム、セミナー等各イベントの開催など幅広く活動を行っています。

機械翻訳にご関心のあるすべての方にご入会をお勧めします。

### \*\* AAMT 会員の特典 \*\*

#### 1. AAMT Journal の購読ができます。

会員には、機関誌である AAMT Journal（年 2～3 回発刊予定）が送付されます。購読料は年会費に含まれています。

#### 2. 機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機械翻訳関連の情報をお届けします。

MT 新製品、新サービスの紹介、国際会議、シンポジウムのお知らせ、WEB での MT 関連記事の紹介など盛りだくさんです。

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

機械翻訳に関する言語資料の調査、広報、標準化活動に参加したり、AAMT Journal や会員専用メーリングリストで、自社製品、サービスの紹介を行うことができます。

#### 4. 関連機関の主催する国際会議に参加できます。

IAMT の主催で隔年開催される MT Summitをはじめ、AAMT、AMTA\*、EAMT\*\* の主催する会議やワークショップに参加できます。

AMTA\* : Association for Machine Translation in the Americas

EAMT\*\* : European Association for Machine Translation

#### 年会費は以下の通りです。

法人会員：入会金 1 口 10,000 円 年会費 1 口 50,000 円

個人会員：入会金 1,000 円 年会費 5,000 円（学生は学生会費 1,000 円）

ご関心のある方は、事務局までお問い合わせください。

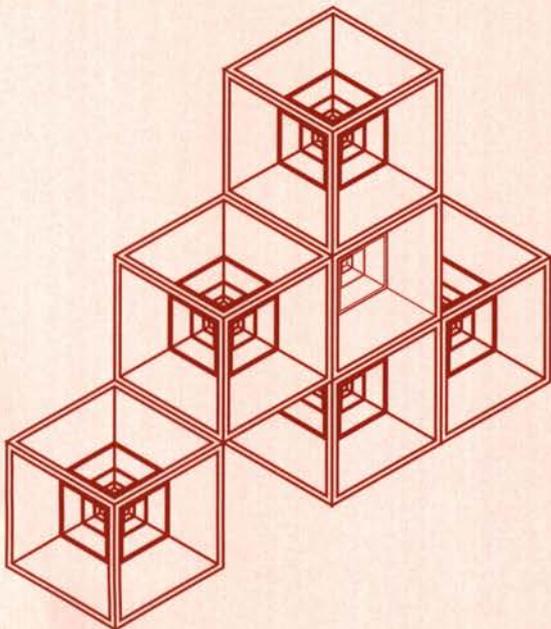
アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

電子メール：[aamt-info@aamt.info](mailto:aamt-info@aamt.info)



# AAMT



## AAMTジャーナル No.50

---

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：〒171-0014 東京都豊島区池袋2-55-2

鈴木ビル3階（株）日本システムアプリケーション内

phone：03-5951-3961 fax：03-5951-3966

編集委員会：宇津呂 武仁 小谷 克則 大倉 清司

鈴木 博和 三浦 貢 村上 嘉陽

事務局：神崎 享子 荻野 孝野

印刷所：株式会社ナビックス

---

Asia-Pacific Association for Machine Translation

c/o Japan System Applications Co., Ltd.

3F, Suzuki Bldg., 2-55-2 Ikebukuro, Toshima-ku, Tokyo #171-0014 Japan

Phone:+81(0)3-5951-3961 Fax:+81(0)3-5951-3966

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