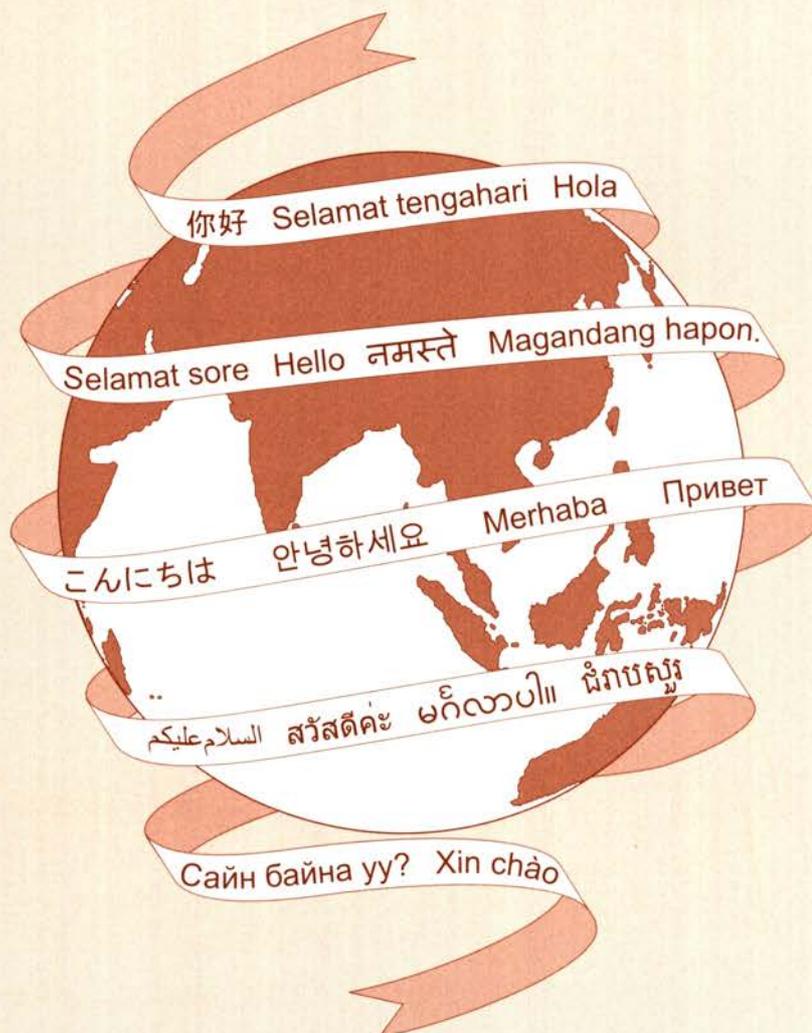


AAMT

Asia-Pacific Association for Machine Translation

Journal



October 2016 *No.63*

アジア太平洋機械翻訳協会

目 次

巻頭言：	「クラウドソーシング」に見る翻訳業界の現在（いま）と今後の展開..... 古谷 祐一.....2
記念講演報告：	「AAMT 設立までのこと」（講演者：長尾真先生） 目次 由美子.....4
記念講演：	拡大する翻訳ニーズと翻訳業界における機械翻訳への期待..... 東 郁男.....7
招待講演：	翻訳品質と TAUS DQF..... 西野 竜太郎..... 11
AAMT-MT フェア 2016 展示報告：	みんなの自動翻訳@TexTra のご紹介 内山 将夫.....15
AAMT-MT フェア 2016 展示報告：	機械翻訳技術を応用した読解支援情報付きニュース..... 田中 英輝.....17
AAMT-MT フェア 2016 展示報告：	サーバ型翻訳製品「ODMA 自動翻訳」と応用アプリのご紹介..... 片江 伸之.....18
AAMT 長尾賞：	産業応用に資する機械翻訳の現状と将来..... 栄藤 稔.....20
AAMT 長尾賞学生奨励賞：	中間言語モデルを用いた多言語機械翻訳の精度向上..... 三浦 明波.....23
連載：	機械翻訳事始め[VIII]..... 坂本 義行.....31
セッション報告：	言語処理学会テーマセッション 『文理・産学を越えた翻訳関連研究』開催報告 山田 優, 藤田 篤, 影浦 峯.....36
シンポジウム参加報告：	TC シンポジウム 2016（東京）参加報告 山本 ゆうじ.....41
プロジェクト紹介：	多言語情報発信支援の社会実装に関する研究 井佐原 均.....43
シンポジウム案内：	「第 26 回 JTF 翻訳祭」のお知らせ..... 古谷 祐一.....45
レポート：	産業翻訳分野における日英機械翻訳の実用化に向けて 徳田 愛.....47
製品・サービス紹介：	FUJITSU Business Application Operational Data Management & Analytics 自動翻訳 上柿 稔.....54
委員会活動報告：	これまでの AAMT Forum メールマガジン Vol.9 機械翻訳課題調査委員会 WG1,2.....57
事務局からのお知らせ：	第 26 回通常総会および関連行事の報告 AAMT 事務局.....62
事務局からのお知らせ：	協会活動報告（2016 年 5 月～2016 年 8 月） AAMT 事務局.....70
編集後記 宇津呂 武仁.....77

CONTENT

Foreword:	The Future Of The Cloudsourcing	<i>Y. Furuya</i>	2
General Meeting Report:	"At the time of the establishment of AAMT"	<i>Y. Metsugi</i>	4
General Meeting:	Expanding translation needs, the translation industry, and expectations	<i>I. Higashi</i>	7
General Meeting:	Translation Quality and TAUS DQF	<i>R. Nishino</i>	11
Exhibition Report:	Introduction of Min'na-no-Jido-Hon'yaku@TexTra	<i>M. Utiyama</i>	15
Exhibition Report:	Automatic Japanese News Reading Assistance Generation and its Application to News Services for the Foreign Population in Japan	<i>H. Tanaka</i>	17
Exhibition Report:	A server type translation product "ODMA automatic translation" and its application programs	<i>N. Katae</i>	18
AAMT Nagao Award:	The Present and Future of Machine Translation toward Industrial Applications	<i>M. Etoh</i>	20
AAMT Nagao Student Award:	Improving Multilingual Machine Translation using Pivot Language using Pivot Language	<i>A. Miura</i>	23
Series:	History of Machine Translation [VIII]	<i>Y. Sakamoto</i>	31
Session Report:	Translation-related studies across boundaries	<i>M. Yamada, A. Fujita, K. Kageura</i>	36
Symposium Report:	Technical Communication Symposium 2016 (Tokyo)	<i>Y. Yuji</i>	41
Project introduction:	Research on Societal Implementation of Supporting Multilingual Information Outbound	<i>H. Isahara</i>	43
Symposium Information:	Information for 26th JTF TRANSLATION FESTIVAL TOKYO 2016	<i>Y. Furuya</i>	45
Report:	Utilizing Machine Translation for Industrial Japanese - English Translation.....	<i>M. Tokuda</i>	47
Products:	FUJITSU Business Application Operational Data Management & Analytics Auto Translation.....	<i>M. Uegaki</i>	54
Committee Report:	AAMT Forum mail magazine vol.9	<i>WG1, WG2</i>	57
AAMT Activities:	General Meeting.....		62
AAMT Activities:	AAMT Activities (from May 2016 to August 2016)		70
Editor's Note:	<i>T. Utsuro</i>	77

「クラウドソーシング」に見る翻訳業界の現在(いま)と今後の展開

古谷 祐一

GMO スピード翻訳株式会社

翻訳業界における「クラウドソーシング」とは？

群集 (Crowd) と業務委託 (Sourcing) を組み合わせた造語で、特定の人々に業務をアウトソーシングすると対比される。クラウドソーシングは狭義では不特定多数の人間に業務を委託するという新しい雇用形態を指し、広義では、必ずしも雇用関係を必要とせず、不特定多数の人間により共同で進められるプロジェクト全般を指す場合もあり、その場合は、「Wikipedia」や「Facebook の多言語プロセス」が代表例として挙げられる。

狭義のクラウドソーシングでは、当該ビジネスモデルで急成長を果たした企業が数社存在している。翻訳産業の枠組みを超えて説明するのであれば、2014年9月18日に株式会社リアルワールド、2014年12月12日に株式会社クラウドワークスが相次いで上場を果たしている。クラウドソーシングの市場規模は2018年には1,800億円になるといわれており、日本の翻訳市場に迫る勢いだ。

クラウドソーシングは、デザイン、Web作成、ライティング、翻訳、動画作成といったありとあらゆる仕事を、クラウド(大衆)にアウトソースし、発注側と受注側をインターネット上で結び付け、手数料をとるビジネスモデルである。特筆すべき点としてはそのサービスの領域に「翻訳」というカテゴリーが存在しており、翻訳を依頼したいクライアントと翻訳作業者のマッチングが既に行われている。海外では、「Elastic」や「oDesk」が有名だが、実はこの2社は2013年12月18日に合併し、約30億円の資金調達に成功している。1年間にフリーランスに支払われた報酬は900億円を超えており、新たな市場が確実に形成されたといっても

過言ではない。

クラウドソーシングが将来性のあるビジネスモデルであることは間違いないが、多種多様な業態を取り扱っているだけに、各サービスの品質レベルを強化するには限界があり、品質が作業者に依存しているというマッチングビジネスの弱点を克服していくことが今後の課題になるだろう。

翻訳に特化したクラウドソーシング

翻訳に特化したクラウドソーシングとして代表的なサービスとして日本国内であれば、「Gengo」や「Conyac」、海外であれば、「One Hour Translation (米国)」や「Translated (イタリア)」、「Transfluent (フィンランド)」などのプレイヤーが新たな市場を開拓している。最初のケーススタディとして紹介したいのが、「Transfluent」だ。同社は、アメリカ大統領選挙でバラク・オバマ氏の公式 Twitter のスペイン語翻訳を手がけた事で、その名が世の中に知れ渡った。「翻訳者」と定義するかどうかの議論は別として、登録翻訳者数は5万人を超えており、80カ国の言語に対応している。ソーシャルメディア、ウェブ、アプリ、EC事業者向けに人力翻訳をAPIで提供しており、Facebook、Twitter、WordPress、ウェイボー、Magento、Zendesk、Unityといったサービスと連携している。ソーシャルメディアの投稿や商品説明、ユーザーレビューなどの翻訳は、リアルタイム性が求められるが、同社のサービスでは当該コンテンツがほぼリアルタイム(短い文章なら数分程度)で翻訳される。

こういった人力翻訳APIはTransfluentだけが突出しているわけではなく、事例は山のようにある。

「Gengo」と「Translated」は、YouTube と API 連携し、字幕翻訳をリアルタイムに依頼できるサービスを2年前から開始しているし、「Conyac」はバリュープレスと提携し英文プレスリリースの作成と海外300メディアへ配信可能なサービスを開始している。日本経済新聞では、全世界のクラウドソーシングの市場規模は13年に約2,000億円、18年には1兆円超に拡大すると予測を出している。伸び行く新規市場の中で、彼らは今後どのような展開をしていくのであろうか。今後の成長に期待したい。

アジア太平洋機械翻訳協会

AAMT 機械翻訳フェア（MT フェア）2016 創立 25 周年基調講演会

「AAMT 設立までのこと」

登壇者 長尾 真 IAMT、AAMT 初代会長

報告者： 目次 由美子

機械翻訳課題調査委員会

AAMT は 1991 年に「日本機械翻訳協会」として設立され、1992 年には事業拡大を目指して「アジア太平洋機械翻訳協会」と名称を変更した。機械翻訳に関する研究開発から製造、販売ならびに利用・活用に至るまで幅広く機械翻訳の普及啓蒙を推進する団体として、四半世紀にわたって活躍してきた。機械翻訳国際連盟（IAMT：International Association for Machine Translation）およびアジア太平洋機械翻訳協会（AAMT：Asia-Pacific Association for Machine Translation）の初代会長である長尾氏により、創立 25 周年を記念する講演が行われた。



MT (Machine Translation) 研究開発の始まり

1945 年頃にはコンピュータを使用した MT が発案され、1950 年代にアメリカとイギリスを中心に研究が始められたそうだ。1957 年にはソビエト連邦によって史上初の人工衛星「スプートニク」が打ち上げられ、これに衝撃を受けたアメリカはソ連の技術文献を調査する必要性を認識したとのこと。露英翻訳者が希少であることやコンピュータの普及なども手伝って、相当な投資を伴う開発が推進されたそうだ。アメリカでの MT 研究に刺激を受けた和田弘氏が、電気試験所からの派遣先であったマサチューセッツ工科大学から帰国した後、1956 年に日本で MT 研究を始めたそうだ。1958 年に完成させた MT システム「やまと」で単文を翻訳させ、1959 年にパリで開催された第 1 回情報処理国際会議にて発表したとのこと。

日米 MT セミナー

1960 年に設立された情報処理学会では、和田座長の下に MT 研究会が発足されたとのこと。計算言語学研究会は、現在も自然言語処理研究会として継続されているそうだ。1964 年には第 1 回 MT 日米セミナーが東京で開催され、カリフォルニア州立大学ロサンゼルス校やテキサス大学を含む約 10 カ所の MT 研究グループに所属する著名な研究者がアメリカから、そして長尾氏自身も参加したとのこと。翌 1965 年には同セミナーの

第2回がニューヨークにて開催され、同氏はアメリカ国内のMT研究機関の数カ所を訪問する機会に恵まれたそうだ。露英のMTデモを見る機会もあったが、当時のコンピュータにはメモリもなく、辞書はディスクに書き込まれていたそうだ。日米セミナーはその後も数回開催され、日本のMT研究に大きな刺激を与えたとのことだった。

1956年頃からは約5年にわたり多額の資金を伴うMT研究が推進されたが、実用には至らなかったことなどもあり表立って研究を継続することは難しく、自然言語処理研究の範疇としてMT研究が推進されていたそうだ。

計算言語学国際会議：Coling

1965年のMT日米セミナー開催時には、IFIP WCC (International Federation for Information Processing World Computer Congress)と第1回Coling (International Conference on Computational Linguistics、計算言語学国際会議)も開催され、ICCL (International Committee on Computational Linguistics、計算言語学国際委員会)がColingの主催団体として結成され、隔年開催が決定された。和田氏は日本からのICCLメンバーとなり、長尾氏も第3回から参加したとのこと。Colingは1980年には日本で初めて東京で開催され、京都を経て、本年12月に大阪での開催が予定されている。

電子協のMT調査研究会

1958年には、現在の電子情報技術産業協会 (JEITA)の前身である電子工業振興協会 (JEIDA)が設立され、1970年代に入って発足されたMT調査研究会の座長は長尾氏が務めたとのこと。大学の研究者や当時の有力なコンピュータメーカーが構成員として参加し、MTの世界の状況の調査や研究開発に関する議論を行うなど、大学のMT研究と企業の開発との橋渡しの役目を2000年代に入っても継続していたそうだ。

Muシステム開発とMTサミット

1982年から4年間行われたMuシステムは、アメリカからの強い要望を受けて、国として取り組んだ日米MTシステムの研究開発だったそうだ。当該の研究開発にはコンピュータメーカーのみでなく翻訳会社を含む多くの企業が参加し、日本のMT技術のレベルアップにも貢献したとのこと。1987年には、JEIDAを経由して通産省からの全面的な支援を受け、世界のMT研究者を集めた国際会議として第1回MTサミットを箱根で開催し、研究開発の成果を世界に向けて公表したそうだ。

IAMTの設立

1987年のMTサミットの成功を継続するべく、1989年には第2回がミュンヘンで開催され、国際的なMTの組織の形成が決定され、1991年にワシントンDCで開催された第3回にてIAMTが組織されたとのこと。IAMTの下にはAAMTを含む3つの地域組織が設けられ、ヨーロッパのEAMT、南北アメリカのAMTAと共に、ローテーションでMTサミットを主催することになった。

AAMT

1991年には日本機械翻訳協会が設立されていたが、IAMTの設立に伴って、1992年にアジア太平洋機械翻訳協会 (AAMT)として拡大し、日本を中心としてその他のアジア諸国にも呼びかけたそうだ。AAMTの25年に亘る発展においては歴代の会長 (田中穂積氏、辻井潤一氏、井佐原均氏、中岩浩巳氏)による尽力と、事務局の荻野孝野氏の協力によるものとのこと。2017年秋には名古屋で開催されるMTサミットの準備を開始していることも紹介された。

MT 研究開発の今後

MT システムの質の向上や、多言語に対応するための拡張が指摘された。テキスト翻訳における分野ごとの専門用語やコーパスの大規模な蓄積が必要であること、音声翻訳については多言語の音声データベースの整備が必要であることが示された。

翻訳専門家の知識の活用

翻訳専門家のノウハウをコンピュータに取り込むことの重要さも、人工知能的手法による蓄積・管理・拡張などを検討する必要性も指摘された。また、スマホのような簡便な装置で各種 MT を運用させることの重要性にも言及された。そして、30 年後には世界の言語障壁を十分低くできることを目指すと力強く語られた上で、努力重ねれば真に実現できると思うとして、講演を締めくくられた。

※ この記事は一般社団法人日本翻訳連盟発行『JTF ジャーナル』との同時取材です。

拡大する翻訳ニーズと翻訳業界における機械翻訳への期待

一般社団法人日本翻訳連盟

会長 東 郁男

今回 AAMT25 周年の総会、イベント開催にあたり、お祝い申し上げます。

また、これまで機械翻訳の開発・技術進展に尽力されて来られた方々に敬意を表しますとともに、日頃 JTF の活動に多大なるご協力をいただき、改めて御礼申し上げます。

日本企業による事業・コンテンツの海外展開の拡大にともなうアウトバウンドニーズや訪日観光客の増加に代表されるインバウンドニーズの増加にともない、翻訳ニーズは増加傾向にあります。その中で、お客様の翻訳に対する要求は、クオリティ、スピード、コスト面において、より多岐にわたってきています。

これらのニーズに対応するために、従来型の間人翻訳に加えて、ICT を活用したクラウド翻訳サービス、機械翻訳などの新しい翻訳手法が考案され、それらが拡大しつつあります。

翻訳業界として、クオリティ・スピードの向上、合理的でスマートなサービスが要求される中、機械翻訳における現状・課題、そして今後の期待について考察していきます。

●日本翻訳連盟 (JTF) の紹介

はじめに、一般社団法人日本翻訳連盟 (JTF : Japan Translation Federation) の概要について、ご紹介させていただきます。

JTF は翻訳に関わる企業、団体、個人の会員で構成される産業翻訳の業界団体です。1981 年に任意団体として設立され、1990 年の社団法人化以来 26 年、翻訳事業に関する調査・研究の実施、研修会・講演会の開催などの人材育成や翻訳関連の国際会議などへの参加

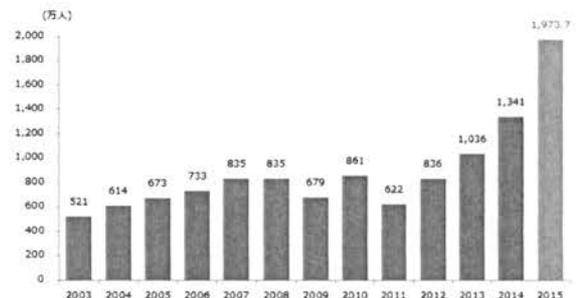
を通して、翻訳事業の振興と日本の経済社会の発展に寄与することを目的に掲げ活動しています。2016 年 3 月 31 日現在の会員数は 507 (法人 193、個人 313、賛助 1) となっています。

●翻訳市場の現状

総務省統計局の統計データによると、訪日外国人数は 2013 年以降増加傾向にあり、特に 2015 年は前年比約 1.5 倍と急激に増加しています (図表 1)。また、技術輸出額も 2011 年 (平成 23 年) を境に右肩上がりに増加していることから (図表 2)、翻訳に対するニーズは増加傾向にあると考えます。

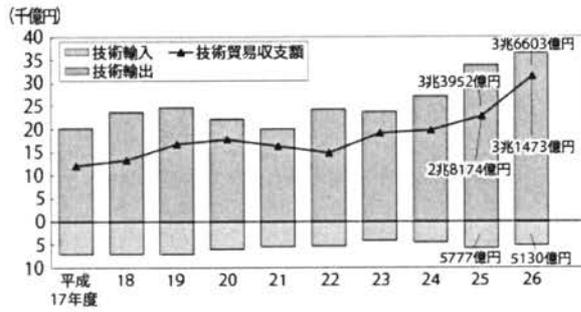
翻訳ニーズの内容としては、訪日観光客の増加にともなうメニュー、リーフレット、ウェブページなどの多言語化に代表されるインバウンドニーズ、および企業活動で発生するコンテンツのグローバル展開によるアウトバウンドニーズが挙げられます。

図表 1 訪日外国人数 (2003~2015)



統計データ出所：総務省統計局

図表2 技術輸出額（2005～2014）



統計データ出所：総務省統計局

● 翻訳産業・市場の変遷

市場のニーズ、ITの進展により翻訳対象のドキュメント・媒体、翻訳の作業環境も変遷しています。TAUS (Translation Automation User Society) のウェブサイトからの引用になりますが、時系列に翻訳産業の変遷を説明します。

1980年代には、紙媒体のドキュメントを、用語集を参照して翻訳していました。

1990年代には、電子ファイルを、翻訳支援ツール (CAT: Computer Aided Translation)・TM (Translation Memory) を活用して翻訳する方法が広まりました。

2000年代には、翻訳対象となるコンテンツのグローバル展開が進み、CMS (Contents Management System)、TMS (Translation Management System)、CAT およびワークフローシステム等の活用により、翻訳プロセス管理の仕組みが導入されました。

2010年代には、TM、CMS、TMSに加えMTを統合して活用することにより、組織や特定グループで共有するコンテンツの翻訳および言語資産の再利用が実現されてきました。

ITとインターネットの発達に伴って翻訳手法、翻訳対象コンテンツが変化しています。また、よりグローバルな社会になるにつれ、翻訳の言語ペアも増加してきたことがわかります。

さらに今後、2020年代にはビッグデータとカスタマイズされたMTを、どこでも誰でも活用できるような時代が到来すると予想されています。

紙媒体の手書き翻訳からタイプライター、ワードブ

ロッセサー・PCなどでのデータ保存・編集が可能となり、その後翻訳支援ツールの活用、そしてMTの活用と、今後翻訳業界にとって大きな変化の時期を迎えることになるかもしれません。

図表3 翻訳産業の変遷

	1980 翻訳	1990 ローカリ ゼーション	2000 グローバル ゼーション	2010 統合	2020 収束
翻訳業界の 関心事	ドキュメント	ソフトウェア	世界同時発売	システム統合	コピキタス化 (埋め込み)
翻訳するコ ンテンツの 種類	紙	電子ファイル	ウェブ	ソーシャル ネットワーク (共有)	個人使用
言語ペア	10	25	40	400	40,000
翻訳支援 ツール	なし	TMと用語管 理	ワークフロー (GMS)	MTと高度な レバレッジ	カスタマイズ MT (リアル タイム)
リソース	用語集	翻訳メモリー	翻訳管理シス テム (TMS)	限定された共 有データ	無制限のピッ グデータ

出所：TAUS「翻訳の未来はあなたの選択次第」

● 機械翻訳の活用状況

次に、翻訳業界内の機械翻訳の活用例を紹介します。

情報の公開範囲とコーパス量をマトリックスにして見ると (図表4)、①コーパス量の蓄積が進み、公開範囲が限定されていない文書については統計ベース機械翻訳 (SMT) が活用されている、②コーパス量の蓄積が少なく、公開範囲が限定されている文書については、ルールベース機械翻訳 (RBMT) が活用されている、ことがわかります。

ただし、例えば、UI (User Interface) など画面ごとの訳し分け (単語ベースでは前後の背景がわからない) などには今後の技術的な課題があるとも認識しています。

図表4 機械翻訳の活用状況

		公開範囲	
		限定	一般
コーパス量	多	(該当なし)	特許公報、マニュアルなど SMT
	少	社内資料、仕様書など RBMT	マーケティング資料、論文、 ソフトウェア画面など

●活用事例の紹介

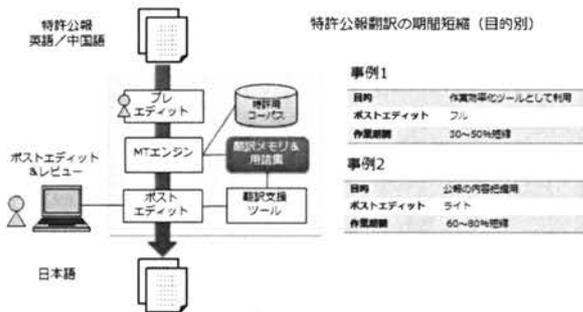
機械翻訳の活用事例として、情報が一般に公開されている特許公報ではコーパスも蓄積され、SMT が広く活用されています（図表5）。

マニュアルの場合、ボリュームの多さと頻繁なアップデートに対応するため、コンテンツ管理システムと連携して、SMT が活用される例があります（図表6）。

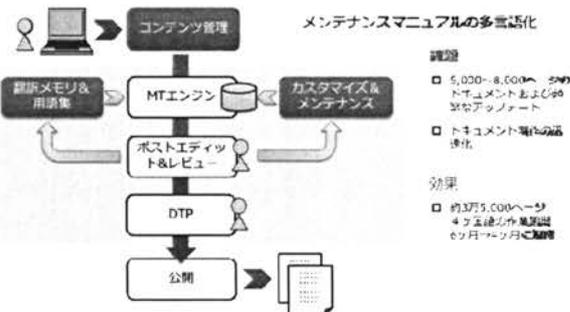
一方、公開範囲が限定されている社内資料などの文書の翻訳については、その翻訳ニーズが「早く・内容が理解できればよい」ということ、またコーパスの量も少ないということもあり、RBMT を活用しているケースがあります（図表7）。

効果として納期の短縮など作業の効率化が図れる一方、活用するケースにもよりますが、MT の前後でプレエディットやポストエディットなどの人的作業の工程も必要となります。そのため、品質基準の設定や作業ガイドラインの策定などの課題についても検討が必要です。

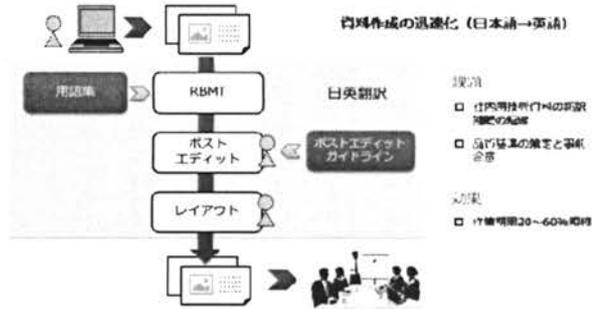
図表5 活用事例の紹介：特許公報



図表6 活用事例の紹介：マニュアル



図表7 活用事例の紹介：社内資料



●JTF の取り組み

JTF としては、今後も AAMT と協力して翻訳業界向けにセミナーを実施したいと考えています。

JTF 会員の間で機械翻訳への興味を引き起こし、印象的なイベントとなった2015年度第3回 JTF 関西セミナー「機械翻訳と向き合うときが来た -MT をもっと身近に、現実的に考える-」に代表されるように、数年前より AAMT と JTF 間の人的交流、セミナーの開催協力も進んでおり、JTF が毎年開催している「JTF 翻訳祭」では、AAMT の企画によるセッションを開催したり、相互の団体の監事担当者を選出したりするなどの人的交流を含めて、今後さらなる連携を図っていきたいと考えています。

また、2017年の秋に名古屋での開催が予定されている MT Summit においても JTF が何らかの形で協力できればと考えています。

●機械翻訳への期待と課題

機械翻訳は今後、インバウンドだけではなくアウトバウンドにおける翻訳にも積極的に活用されることが見込まれます。これらに対応するためには、コーパスのなお一層の蓄積に加え、MT エンジンのチューニングとトレーニング技術の進化、翻訳精度や処理速度の向上などの改善余地に期待が寄せられています。

翻訳業界としては、多様化するクライアントのニーズに合わせて、人手翻訳、CAT、MT、MT+PE（機械翻訳+ポストエディット）などの様々な手法を柔軟に活用して、コスト・スピード・発注のしやすさといった、

お客様のニーズを満たすことができるよう努力していく必要があります。

●最後に

今後、翻訳ニーズがより拡大していく中、翻訳業界として市場、クライアントの要求に応えられるキャパシティを確保していくためにも、AAMT と JTF がさらに連携していくことにより、企業・社会・経済のグローバル化に貢献できるのではないかと考えます。

本年 11 月 29 日（火）東京・市ヶ谷のアルカディア市ヶ谷にて第 26 回翻訳祭が開催されます。当日は AAMT ご協力のセミナーも予定されておりますので、是非 JTF のイベントにもご参加いただければ幸いに存じます。

最後に、AAMT 総会での講演に続き、AAMT ジャーナルへの寄稿の機会までいただきました関係各位の皆様方に御礼申し上げます。

翻訳品質と TAUS DQF

西野 竜太郎

合同会社グローバリゼーションデザイン研究所 / TAUS Representative

1. 翻訳の「品質」とは

翻訳業界において翻訳の「品質」が何を指すかは、立場によって異なる。長年の実績を持つ翻訳者であれば、自分の経験を基にして良し悪しを判断するかもしれない。最終読者の場合は読みやすさを重視するかもしれない。発注したクライアントは、社内の品質要件に適合しているかどうかで判断するかもしれない。受注した翻訳会社であれば、誤訳や用語集違反などのエラーの少なさを指標にするかもしれないし、「この料金でこれだけでできていれば質は十分」と考えるかもしれない。どの立場からの評価方法も一理あるように思えるが、これだけ異なっていると何かしら合意を形成するのは難しそうだ。

こういった状況の中、Fieldsら翻訳業界の関係者は経営学者 Garvin の論を参考にして、翻訳の「品質」を分類した^[1]。これによると、翻訳の品質は以下の5つのアプローチに分けられる。

- A. 超越的 (transcendental) : もともと哲学の言葉。良い文章に多く触れるといった経験を通して培われた力で、品質の良し悪しを直観的に判断
- B. プロダクトベース (product-based) : 製品やサービスの品質は原材料や特質によって測定が可能。計測可能な数値を用いるため、製品やサービスどうしの比較ができる
- C. ユーザーベース (user-based) : 製品やサービスの品質は、ユーザーのニーズ、要望、好

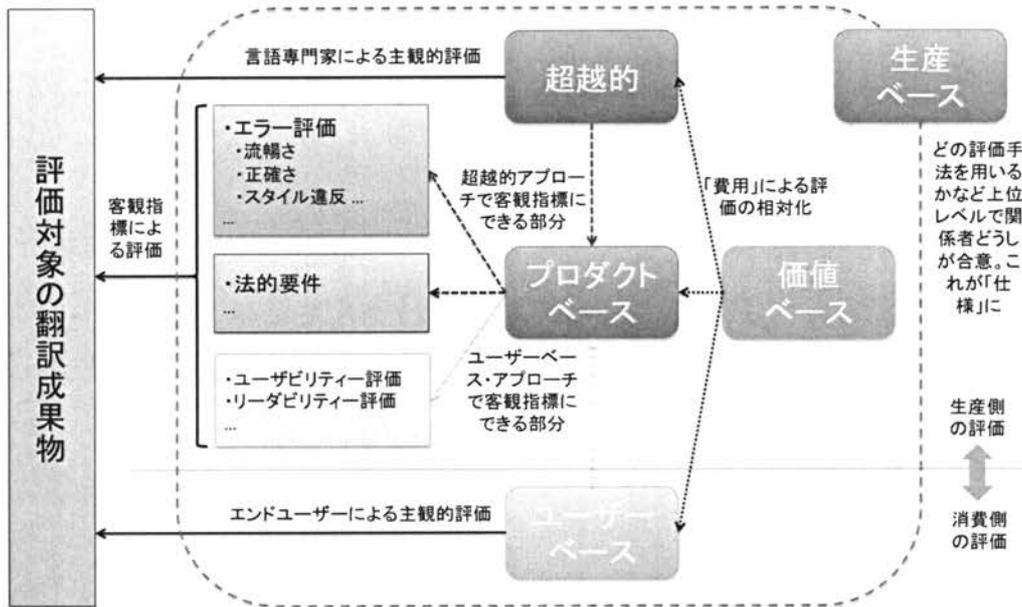
みを満たしている度合いによって決まる

- D. 生産ベース (production-based) : あらかじめ定めた要件や仕様をどの程度満たしているかで品質が決まる
- E. 価値ベース (value-based) : 費用と便益で品質を測定。便益が費用に比べてより大きいなら、製品やサービスにより価値があり、品質も高い

これらを用いると、異なる立場をうまく分類できる。まず、自分の経験で判断をする翻訳者は「A. 超越的」のアプローチを用いている。読みやすさを重視する最終読者は「C. ユーザーベース」、社内の品質要件で判断したクライアントは「D. 生産ベース」の考え方に基いている。また、翻訳会社が誤訳などのエラーの少なさ(数字)で評価する場合は「B. プロダクトベース」、料金(費用)との兼ね合いで品質を判断する場合は「E. 価値ベース」のアプローチであると言えよう。ちなみに Garvin は、これまで主張された「品質」の定義は、この5つのどれかに入ると述べている^[2]。

2. 品質5分類の関係

しかし、この5分類はお互いの関係がからみ合っていて見えにくい。例えばBのプロダクトベースでは直接対象の品質を測定しているのに対し、Eの価値ベースでは、そういった測定結果を費用という視点で相対化している。そこで5分類の関係を整理し、次の図にまとめた。



図の左端に「評価対象の翻訳成果物」がある。これに向かって「超越的」アプローチから実線が伸びている。これは言語専門家による主観的評価だ。主観的といっても悪い意味ではなく、長年の経験や独自の感性に基づく判断は重要である。そこからさらに下の「プロダクトベース」に点線が伸びている。超越的アプローチのうち客観指標にできるような部分がプロダクトベースに変換されるという意味である。例えば翻訳者などの言語専門家の判断を「エラー評価」のような客観指標に変換するケースだ。エラー評価には流暢さ、正確さ、スタイル違反などがあり得る。プロダクトベースでは、こういった比較的客観的な指標によって翻訳成果物を評価する。最下部に「ユーザーベース」があり、ここからも左端の評価対象に実線が伸びている。これはエンドユーザーによる主観的評価である。主観的と言ってもやはり悪い意味ではなく、実際のユーザーがどういう印象を持ったのかという情報には価値がある。こちらも同様にプロダクトベースに向かって点線があり、ユーザーベースのアプローチで客観指標にできる部分に変換される。例えばユーザビリティやリーダビリティが考えられ、こういった客観指標もプロダクトベースによる評価だと言える。

中央のやや右側に「価値ベース」がある。直接的に

翻訳成果物を評価するのではなく、費用という視点で評価を相対化する。そのため、超越的、プロダクトベース、ユーザーベースに点線が伸びている。最後に「生産ベース」がある。生産ベースでは、関係者間で合意した仕様や要件をどの程度満たしているかで品質を測る。どの評価手法を用いるか、検収時のしきい値はどうするか、費用はどうするのかといった点で合意したものが仕様である。この生産ベースも翻訳成果物を直接的に評価するわけではなく、ほかの4つのうちで何をどう組み込んで用いるのか、上位レベルで仕様として規定する。そのため、図ではほかの4つを点線で囲んでいる。

つまりまとめると、直接的に翻訳成果物を評価するのが「超越的」、「プロダクトベース」、「ユーザーベース」という3つであり、その3つを費用という視点から相対化するのが「価値ベース」、さらにそれら4つをどう用いるのかという上位レベルのアプローチが「生産ベース」だと言える。

3. 翻訳ビジネスでは何を用いるべきか

Garvin は論文[2]で、5つのアプローチは状況に応じて使い分けるべきだとしている。製造業の場合、市場

調査段階でユーザーベース、それをプロダクトベースの特性に変換し、そこから仕様を作って、製造段階で生産ベースを使うという例を挙げている。例えばアイスクリームであれば、まず市場調査でどのようなアイスクリームが好まれるのかを調べ、仮に乳脂肪が多いものが好まれるとしたら、それを乳脂肪分何パーセントというような特性に変換し、そこから仕様を作ってアイスクリームを製造するという流れである。

これは翻訳業界でも応用できるだろう。例えば、まず翻訳者のような言語専門家が普段主観的に判断している部分、つまり超越的な部分を分析し、プロダクトベースに変換する。典型的にはエラー分類だろう。こういったエラー分類は一から作るのではなく、後述のDQFのような業界のものも利用できるはずだ。次に、どの方法を用いて翻訳成果物を評価するのかといった点について関係者間で合意し、仕様を作る。最終的にはこの仕様を満たしているかどうかで品質を評価するという流れとなる。

翻訳ビジネスでは、仕様を満たすかどうかの「生産ベース」を用いるべきだと私は考える。これは翻訳業界では複数の独立した事業者（フリーランス翻訳者、翻訳会社、クライアント）がかかわることが多いからである。ビジネスではどうしても経済的な利害関係が発生する。例えばクライアント社内で検収を担当する言語専門家による評価は尊重すべきだが、判断が毎回揺らいで場当たりに修正を求められると、翻訳会社は困るだろう。また、厳しいスケジュールで納品を求められたのに、後になって「訳が読みにくい」などと低評価を下されたら、翻訳者としては辛いはずだ。どの評価手法を用いるかなど、あらかじめ評価の枠組み自体をまず上位レベルで合意する必要がある。実際、ASTM F2575 や MQM など海外のいくつかの品質基準では生産ベースの考え方が採用されている。もちろんビジネス以外の場面では、生産ベースの仕様を作成することなく、例えば超越的アプローチで評価しても問題ないだろう。独立した事業者どうしのビジネスだからこそ、仕様として合意する必要がある。

4. プロダクトベースのエラー評価

これまで述べてきたように、分類の1つに「プロダクトベース」があり、その中心的な手法に「エラー評価」がある。従来から翻訳業界で用いられてきており、品質評価の方法としては主流であると言えよう。一般的には、訳文中に含まれるエラーをカテゴリ別に重み付けしながらカウントし、合計点数がある値を超えると不合格にする方法が採られている。カテゴリには、流暢さ、正確さ、スタイル違反などがある。また、エラーが深刻な場合は10点、重度な場合は5点、軽度の場合は1点などと重み付けする。従来から用いられてきた方法としては、ローカリゼーション業界におけるLISA QAモデルや自動車業界におけるSAE J2450がある。

しかし近年、新しいエラー評価の方法が提唱されている。TAUSによる「DQF」（2012年）、EUが資金拠出したQTLaunchPadなどによる「MQM」（2014年）である。次のセクションではこのうちDQFについて詳しく説明する。

5. TAUS DQF とは

DQFとはDynamic Quality Frameworkのことで、エラー評価などの品質評価手法に加え、ツールも含めた大きな枠組みである。コンテンツの種類に応じて評価方法を変えるので「Dynamic」と呼ばれる。例えばマーケティング資料と特許文書では評価方法が違って当然であるが、従来の手法ではコンテンツの種類が何であっても同じ基準を用いる傾向があり、「one-size-fits-all」といった批判^[3]もあった。DQFはそういった批判に応えようとしているのである。

DQFの評価手順としては、まず「コンテンツ・プロファイリング」をウェブ上のツール（※注1）で実施する。この結果、コンテンツの種類に合った評価手法がいくつか推奨される。例えば「ユーザー・インター

フェイス・テキスト」の場合、ユーザビリティ評価やエラー評価などが推奨される。つまり DQF のエラー評価は、推奨される方法のうちの 1 つなのである。その後、推奨の方法で品質評価を実施するという流れになる。

6. DQF のエラー分類

前述のように、推奨された場合にエラー評価を実施するが、DQF は独自のエラー評価手法（分類と重み付け）を提示している（※注 2）。そのうちエラーの分類はレベルが 2 つに分かれている。時間があれば詳細レベルで評価し、簡略的に実施したければ上位レベルだけ用いればよい。以下では上位レベルの項目のみを紹介する。

1. 正確さ (Accuracy)
2. 流暢さ (Fluency)
3. 用語 (Terminology)
4. スタイル (Style)
5. デザイン (Design)
6. ロケール慣習 (Locale convention) ※注 3
7. 事実不一致 (Verity) ※注 4
8. その他

ちなみに DQF のエラー分類は、2015 年に「MQM」と統合されている。上記の分類は統合後のものである。統合の結果、今後より広く国際的に利用されることが期待されている。

7. まとめ

翻訳の「品質」が何を指すかは使う人によって異なるが、Garvin の方法で 5 つに分類できる。超越的、プロダクトベース、ユーザーベース、価値ベース、生産ベースである。そのうち、独立した事業者が複数関わる翻訳ビジネスでは「生産ベース」を用いるべきだ

と考える。生産ベースでは、どの評価手法を使うかなど、上位レベルで関係者が合意して「仕様」とする。その仕様を満たす度合いで品質を測ることになる。そういった仕様に組み込まれ、直接的に翻訳成果物を評価するのがプロダクトベースのアプローチで、中心的な方法にはエラー評価がある。そのエラー評価の新しい方法として TAUS の DQF が提唱されている。DQF と MQM のエラー分類項目は 2015 年に統合され、今後広く使われる可能性があると考えられる。

注 1 : URL は

<https://evaluate.taus.net/evaluate/content-profiling/profile-your-content> (2016/7/9 現在)

注 2 : URL は

<https://evaluate.taus.net/evaluate/taus-dqf-documentation#dqf-mqm-error-types> (2016/7/9 現在)

注 3 : 住所や日付などが翻訳対象地域の慣習に従っているかどうか、という分類

注 4 : 訳文が事実に合致しているかどうか、という分類。例えば、ある自動車のモデルに特定の機能がないのに、訳文に「ある」と書かれている場合

参考文献

- [1] Fields, P., Hague, D., Koby, G.S., and Melby, A. What Is Quality? A Management Discipline and the Translation Industry Get Acquainted. *Revista Tradumática*, 12, pp.404-412 (2014).
- [2] Garvin, D.A. What Does “Product Quality” Really Mean? *Sloan Management Review* 26, pp.25-43 (1984).
- [3] O'Brien, S. Towards a dynamic quality evaluation model for translation. *The Journal of Specialised Translation*, 17, (2012).

みんなの自動翻訳@TexTraのご紹介

内山 将夫

情報通信研究機構 先進的翻訳技術研究室

1. みんなの自動翻訳@TexTraの特徴

「みんなの自動翻訳@TexTra」¹は、国立研究開発法人情報通信研究機構（NICT）が開発した自動翻訳サイトです²。NICTの自動翻訳エンジンは、特許・論文・マニュアルなどの長文翻訳を得意とし、特許庁³や国立研究開発法人科学技術振興機構⁴との研究の連携も行っています。現在、日本語・英語・中国語・韓国語の4言語に対応し、今後に対応言語を拡張させる予定です。特許・論文・マニュアルなどのほか、汎用的な翻訳にも対応できる自動翻訳エンジンも搭載しています。

みんなの自動翻訳@TexTraでは、対訳や用語をユーザーが追加することにより、自分好みの自動翻訳エンジンが作成できます。また、追加されたデータは自動翻訳技術の向上に活かされます。みんなの自動翻訳@TexTraは、自動翻訳をみんなで育てるサイトです。



¹ <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>

²「特許やマニュアルなどの長文を正確に翻訳するための新技術（語順変換と訳語選択の独立実行による統計翻訳技術）の研究」に基づく自動翻訳エンジンを活用しています。この自動翻訳エンジンは、機械翻訳システムの実用化のための研究開発に貢献したことが認められ、2014年第9回アジア太平洋機械翻訳協会（AAMT）長尾賞を受賞しました。

³NICTと特許庁の特許文献の機械翻訳に関する協力の継続について（2016年4月1日）

⁴科学技術文献データベースの作成に「高精度自動翻訳システム」を導入（2015年3月30日）

2. 特許請求項に特化したエンジンの特徴

特許請求項は特許で一番重要ですが、特許請求項は特別なパターンを持つ超長文です。そのため、従来のSMTでは特許請求項の精度良い翻訳は不可能でした。

NICTはパターン翻訳をSMTに導入しました。まず、パターン解析をして、請求項の各部品を並び替えます。そして、各部品は特許翻訳エンジンで翻訳します。この翻訳には、特許庁との協力により作成されたMTエンジンを活用します。

次に翻訳例を示します。

原文：A rotary cutter comprising: a handle;

an arm extending from an end of the handle; a circular blade coupled to the arm; and a crust cutter member extending from the end of the handle, wherein an edge of the crust cutter member is sharpened and the crust cutter member extends in a plane that is substantially parallel to a centerline of the handle.

NICT 翻訳：ハンドルと；前記ハンドルの前記端部から延びるアームと；前記アームに連結された円形のブレードと；前記ハンドルの前記端部から延在する外皮切断部材と；を備え、外皮部材の縁部は鋭利であり、皮の切断部材が、ハンドルの中心線に実質的に平行な平面に延在している、ロータリカッタ。

上記翻訳例からわかるように、NICT翻訳は次の2点の特徴を持ちます。

- (1) 特許文特有の用語と表現に対応
- (2) 請求項文のスタイルに則った明確な記載

3. ビジネスのスピードを加速させ、コストの削減も可能に！

大規模特許対訳文（日英 2 億文、日中 1 億文）を利用した高精度な特許自動翻訳の活用で、海外の技術調査や情報収集が簡単になります。また、外国出願時に必要な特許明細書の下訳作成にも役立ちます。さらに、海外での事業展開に必要な製品マニュアルやリーフレット作成もスムーズにこなせます。また、幅広い分野で活用可能な自動翻訳エンジンで、インバウンド需要にも対応可能です。

4. 請求項翻訳を使ってみよう！

Step1. ユーザー登録してログイン



Step 2. 自動翻訳を選択



Step 3. 「JPO+NICT 請求項」にチェック



Step 4. 訳文が表示される



本サービスの利用は非商用利用に限定されたものです。商用利用のお問合せ先：ltg-info@knh.nict.go.jp

機械翻訳技術を応用した読解支援情報付きニュース

田中英輝

NHK 放送技術研究所

1. はじめに

日本に在住する外国人の数は年々増え、2015年12月には223万人、日本の総人口の1.8%に達しました。この比率は欧米諸国に比べて高いとは言えませんが、東京都新宿区では10%を超えるなど、地域によっては欧米並みの外国人の集中が見られます。

外国人の中にはニュースが難しくてよくわからないという方が大勢います。NHKではこのような外国人のために、外国語に加えてニュースを「やさしい日本語」に書き換えて提供するインターネットサイト、NEWSWEB EASYを2012年4月に始めました¹。

2. NEWSWEB EASY

NEWSWEB EASYのニュースは、ネットに公開されたニュースをやさしい日本語に書き換えて作ります。この作業は日本語教師と記者が協力して行い、日本語学習の初級レベルの単語と文法をできるだけ使うようにします。また長いニュースは半分以下に要約します。

さらにWebの画面には漢字へのふりがな、難しい単語への小学生用辞書の説明の付与、音声合成によるテキストの読み上げ、固有名詞のカラーハイライトなどの読解を助ける機能があります。

開始以来、NEWSWEB EASYは国内外の多くの方に閲覧されてきました。この中にはニュースを知るためだけでなく、日本語の勉強の教材として見ている方が大勢いることがわかってきました。

3. 読解支援情報付きニュース

NEWSWEB EASYではやさしく書き換えたニュースを提供していますが、新たに、元のニュースを提供

するための研究を始めました。NEWSWEB EASYには実現できていない機能がいくつかあります。まず元ニュースを要約して作るため、内容全部を知ることはできません。またニュースには慣用的な表現がたくさん現れますが、ほぼやさしい日本語に書き換えられているため勉強できません。さらに提供している辞書に冊子辞書の電子版をそのまま使っているため正しい説明を見つけにくい問題があります。

これらを解決するには元ニュースに機械翻訳技術を使って作成したやさしい日本語と外国語の訳をふりがなのように重ねて提供する「読解支援情報付きニュース」が有効だと考えています。この方式では元ニュースそのものの提供でき、それぞれの表現に適切な訳語を付けることができます。もちろん誤った翻訳の修正は必要です。

今回はニュースにやさしい日本語と韓国語の読解支援情報を自動的に重ねて表示するシステムと読解支援情報の付いたニュースのイメージを展示しました。図に示したイメージの右上のスライダーを使うと読解支援情報の量を調節できます。今後は日本語学習者や日本語教師のみなさんのご意見を聞きながら研究開発を進めていきます。



¹ <http://www3.nhk.or.jp/news/easy>

サーバ型翻訳製品「ODMA 自動翻訳」と応用アプリのご紹介

片江伸之
(株) 富士通研究所

2016年6月17日に、ホテルアジュール竹芝にて開催された「AAMT 機械翻訳フェア (MT フェア) 2016」にて、富士通株式会社から発売しているサーバ型翻訳製品「ODMA 自動翻訳」¹のご紹介と、それを用いた応用アプリの展示を行いました。

1. サーバ型翻訳製品「ODMA 自動翻訳」

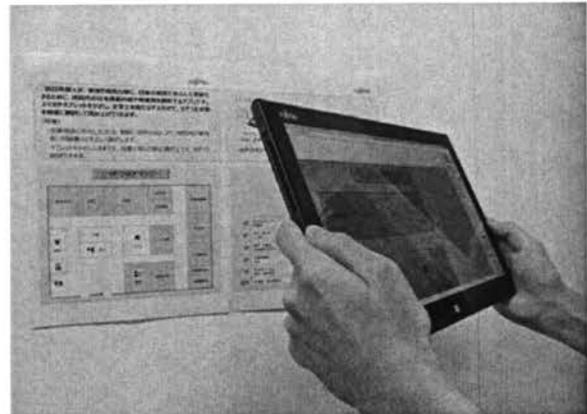
「ODMA 自動翻訳」は、ネットワークを介して使用する日本語／英語の自動翻訳ソフトです。社内に自動翻訳サーバを設置し、Web ブラウザでアクセスして翻訳機能を利用します。サーバに収録された共通の辞書を用いた翻訳を全社で利用できるため、社内文書を共通の訳語ルール、表記ルールに基づいて翻訳することが可能になり、会社全体で翻訳品質を統一できます。

本製品の詳細につきましては本ジャーナルの記事「FUJITSU Business Application Operational Data Management & Analytics 自動翻訳」に掲載されておりますので、そちらをご参照ください。

2. 応用アプリ

●なぞって翻訳

日本語で書かれた案内表示や書類などにスマホやタブレットをかざし、画面上の文字を指でなぞるだけで、なぞった文字を英語などの外国語に翻訳し読み上げてくれるアプリです。訪日外国人の方が、怪我や病気で日本の病院を受診されることを想定し、病院内の案内板や診療明細書を翻訳するデモをご覧いただきました。



①翻訳対象を撮影



②文字をなぞると即座に翻訳して読み上げ

●インスタントメッセージ翻訳

海外拠点とのコミュニケーション円滑化のために、日本語メッセージを相手の言語にリアルタイムに翻訳する機能を提供します。

Lync/ Skype for Business インスタントメッセージの画面において、母語でテキストを入力し、翻訳ボタンを押下するだけで翻訳された結果が相手に送信されます。相手からのメッセージは自動的に母語に翻訳されて表示されます。また、相手の言語と自身の言語での会話のログが並行して表示されますので、両言語で

¹<http://www.fujitsu.com/jp/solutions/business-technology/intelligent-data-services/ba/product/operational-data-management-and-analytics/auto-translation/>

会話の流れを確認することができます。複雑な操作をすることなく多言語コミュニケーションを実現することが可能になります。

最後に

当日はたくさんの方々にお越しいただき、意見交換をさせていただくことができました。誠にありがとうございました。

お問い合わせ先：

- ・お問い合わせフォーム

https://jp.fujitsu.com/cgi-bin/fjid/formoutput_cs.cgi?FMT=/contact/csform/csque11701/form0010/1

- ・富士通コンタクトライン（総合窓口）

TEL: 0120-933-200

注) Lync, Skype for Business は、米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です。



インスタントメッセージ翻訳の画面

産業応用に資する機械翻訳の現状と将来

栄藤 稔

株式会社 みらい翻訳

1. みらい翻訳の翻訳エンジン

機械翻訳の基本技術となっているルールベース機械翻訳(Rule-Based MT; RBMT)と統計的機械翻訳(Statistical MT; SMT)の技術を用いた複数のエンジンを組み合わせて機械翻訳を行っている(図1)。RBMTは文法規則や辞書から人手により翻訳ルールを登録し、ルールに基づき翻訳文を作成する。教科書的な文法ルール、構文が明確な文章であれば、翻訳しやすい。文法が似ているヨーロッパ語族間の文書間であれば、辞書登録により実用的な翻訳が可能である。語族が異なる語順が違う言語間でも、分野を問わず長文を大崩れせず翻訳できるが、一定レベル以上の精度向上は人手によるチューニングでは困難な為、訳質に不自然さは残る。そもそも、想定外の原文を翻訳で補うのが難しく、直訳で流暢ではないという弱点を持っている。コスト構造は「人手と辞書整備」となる。

一方で SMT では、大量の対訳文(コーパス)学習により、確率的な組み合わせ計算により翻訳文を作成する。口語等文法ルール、構文がくずれた文章でも翻訳しやすいという特徴がある。また特定分野のコーパスを準備することで、使用用途やカテゴリーに沿った訳出が可能という特徴もある。コーパスさえ用意できれば、訳質の向上が望める。短所はその裏返しであり、RBMTが必要としないコーパスが必要となる。翻訳したい文書の分野と、学習しているコーパスの分野が異なると、翻訳精度は低くなる。翻訳精度の向上には、コーパスの量と質が要求される。「コーパス整備」がSMTコスト構造となっている。

近年のビッグデータのブームにより、SMTがRBMTを翻訳精度で上回る例が多くなった。一方で、コーパス整備ができず、RBMTが活躍している事例も多い。図1

に示す、当社の協業先であるシストラン社翻訳サーバーでは、RBMT, SMT(いずれも複数)の組み合わせによるハイブリッド構成を採用している。エンジン周辺は、さらに辞書、前処理、後処理、翻訳メモリ(過去の翻訳事例データベース)、コーパス、翻訳対象の言語モデルデータベースを柔軟に組み合わせて翻訳を行う構成となっている。

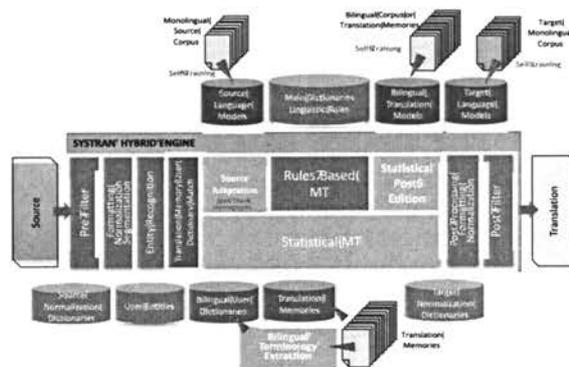


図1. SYSTRAN Enterprise Server.

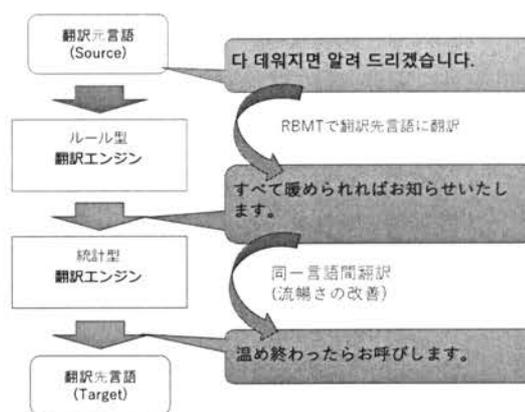


図2. ハイブリッド構成による流暢さの改善.

図1中 Statistical Post-Edit(SPE)とあるのは、

実質 SMT である。RBMT+SPE の動作例を図 2 に示す。この例では、韓国語の文章をルールベースにより日本語に変換する、その上で日本語から日本語への SMT を行う。同一言語間の機械翻訳とは奇妙な感じがするが、これにより流暢さが増す。RBMT の性能が良い局面でコーパスが数十万あれば良好な結果が得られる。RBMT と SMT の折衷案として高性能が期待出来る事例である。

2. 事前ならび替え技術について

SMT は話し言葉という短文では、少々不規則さがあっても自然な文章として翻訳できる一方で、マニュアルや特許、さらには契約書という長文の産業文書の機械翻訳には不向きという常識がこれまであった。この常識を打ち破ったのが、「事前並び替え」という新技術¹である。これにより、長文翻訳で SMT が RBMT を上回る目処が立った。図 3 上段にその例を示す。この例では、英語を日本語に機械翻訳する際、英語の主語(S)、述語(V)、目的語(O)の語順を日本語の語順 SOV に変換した後、SMT を行う。原文を対象言語の語順に事前に並び替える。この「前処理+SMT」は前述の「RBMT+SPE」に似ているが、長い文章翻訳に適している。前処理には様々な方法があり、事前並び替え自身、機械学習でコンピュータが習得可能である²。この並び替えの難しさは幾つかの文献^{3,4}で定量的に計測されており、それを図 3 下段に示す。文法構造の違いと文脈依存性から日本語から英語への変換が最も難しいことがわかる。

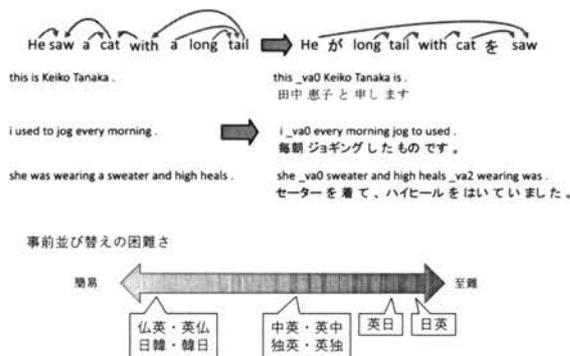


図 3 事前並び替え

3. TOEIC にみる性能

機械翻訳の性能を評価する方法は主観評価による 5 段階評価、BLEU 値⁵などがあるが、生活実感に基づいた評価が欲しい。そこで、TOEIC 試験による評価を行った。コンピュータが TOEIC 試験を受けて文章要約など高度な読解力を持つわけではないので、TOEIC 受験者の英作文能力との比較により、その英作文能力が TOEIC 何点相当かを試す^{6,7}。評価手順を以下に示す。

- ① 過去半年に TOEIC を受験した被験者を、100 点台刻みで 5 人ずつ召集する。
- ② ②の被験者に、辞書、翻訳ツール等の使用なしで、300 問の日本語原文を自力で英訳してもらう。
- ③ 評価対象エンジン(機械翻訳)で、②と同じ日本語原文を英訳する。
- ④ 日英バイリンガルの評価者に、②と③の翻訳精度を 4 段階でランク付けしてもらい、被験者 vs. 機械翻訳の優劣を決定する。
- ⑤ ④の優劣の結果から被験者ごとに機械翻訳の勝率を計算し、回帰分析により近似直線を得る。
- ⑥ ⑤の近似直線と勝率 0.5(=機械翻訳と人手翻訳が互角)の交点が、TOEIC スコア換算値となる。

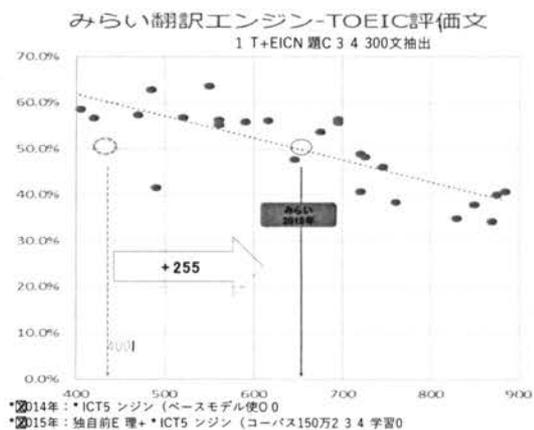


図 4 TOEIC 試験結果

みらい翻訳は国立研究開発法人、情報通信研究機構(NICT)より SMT のエンジン供与を受けている。この SMT エンジンを図 1 のサーバーに組み込み性能評価を行った。SMT の性能はエンジン単体では決まらず、前処理、使用コーパスによって異なることに留意して欲しい。

ここではTOEIC問題集から集めたビジネス会話を試験対象とする。旅行会話に比べて幅広いコーパスが必要となる。2014年、みらい翻訳成立時のNICTエンジンは事前並び替えなしの方式であった。ビジネス会話へのチューニングも不十分で、TOEIC400点台前半の受験者相当の英作文能力であった。2015年末、みらい翻訳独自の事前並び替え技術をNICTエンジンに組み込み、150万コーパスの学習をさせたところ、655点の受験者と同等の英作文能力を達成できた。事前並び替えがビジネス会話という実際の応用に有効であることを実証した。このスコアは英検2級相当の実力であり、多くの企業で「通訳をつけず一人で海外に行き、最低限の打ち合わせが出来るレベル」と認識されている。

4. 機械翻訳プラットフォームと産業応用

SMTは事前並び替え等の新技術の導入により、コーパスさえ揃えばこれまで達成不可能と思われてきた機械翻訳の性能を実現してきた。図5にNICTの事前並び付きSMT(長文翻訳エンジン)の特許翻訳例を示す。特許は直訳の綺麗なコーパスが集まるため、SMTの効果が得られやすい。この例では、1000万コーパスの学習を行っている。もはや、人手の翻訳と遜色ないレベルに達成していることが分かる。

SMTの実用化はいかに良質のコーパスを得るにかかっている。産業応用では、お客様各社の持つコーパスを学習に用いた個別チューニングが重要となる(図6)。みらい翻訳の強みは、自社の持つ基本データベースに加えて個社データを加えたチューニングシステムを保有していることである。図6のサイクルを回すことにより、日本において成功例を作っていくたい。電子メール、特許、マニュアル、契約文書の翻訳から、機械翻訳による世界的規模のデータマイニングによるマーケティングなど新たなイノベーションが期待できる。この10年間で、機械翻訳は日進月歩の成長を遂げた。今後はニューラルネットワークによる時系列データの深層学習モデルの応用も進むことが期待できる。今後おきるブレークスルーに期待したい。

原文: The electric motor 24 is attached and supported to the second housing 12 by a mounting bolt 28, and is driven by a controller (not shown) in correspondence to the detected torque of the torque sensor 23.

機 訳 動モータ24は、取付ボルト28により第2ハウジング12に取付け持され、トルクセンサ23の検出トルクに応じてコントローラ(図せず)によって駆動される。

手 訳 動モータ24は、取付ボルト28により第2ハウジング12に取付け支持さ、不図示のコントローラにより、トルクセンサ23の検出トルクに応じて駆動される。

原文: 根据本发明的聚合物共混物的导电性能进一步。例如,用商业产品抗静电化合物、软化剂或其它小分子吸湿性化合物来改进。

機 訳 本発明のポリマーブレンドの導電性することができ、例えば、市販の防止化合物、軟化剤または他の低分子吸湿性化合物を用いて改善する。

手 訳 発明のポリマーブレンドの導電性は、例えば、市販の帯電防止化合物、化剤、または他の小さい分子の吸湿性化合物と共にさらに改良できる。

図5. NICT長文翻訳エンジンの翻訳結果。

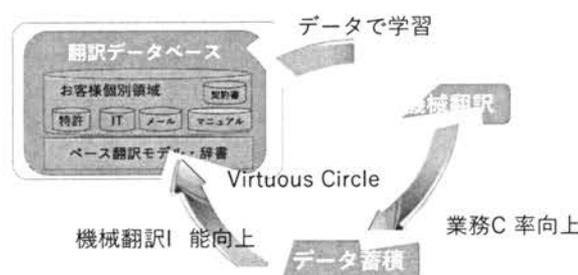


図6. 個別チューニング。

¹ Isozaki, H., Sudoh, K., Tsukada, H. and Duh, K., 2010, July. Head finalization: A simple reordering rule for sov languages. In Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR (pp. 244-251). Association for Computational Linguistics.

² Bisazza, A. and Federico, M., 2016. A Survey of Word Reordering in Statistical Machine Translation: Computational Models and Language Phenomena. *Computational Linguistics*.

³ Kudo, T., Ichikawa, H. and Kazawa, H., 2014. A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation.

⁴ Li, J., Kim, S.J., Na, H. and Lee, J.H., 2014. Postech's system description for medical text translation task. *ACL 2014*. p.229.

⁵ Doddington, G., 2002, March. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the second international conference on Human Language Technology Research* (pp. 138-145). Morgan Kaufmann Publishers Inc..

⁶ 菅谷史昭, 竹澤寿幸, 横尾昭男 and 山本誠一, 2001. 音声翻訳システムと人間との比較による音声翻訳能力評価手法の提案と比較実験. *電子情報通信学会論文誌 D*, 84(11), pp.2362-2370.

⁷ 安田圭志 and 菊井玄一郎, 人間の能力との比較による音声翻訳システム性能評価法の効率化に関する検討. *言語処理学会第12回年次大会*, P3-3 pp.252-255.

中間言語モデルを用いた多言語機械翻訳の精度向上

三浦 明波

奈良先端科学技術大学院大学

1. はじめに

統計的機械翻訳 (Statistical Machine Translation: SMT [1]) では, 学習に用いる対訳コーパスが大規模になるほど, 高精度な訳出結果を得られることが知られている [2]. 一方, 英語を含まない言語対などを考慮すれば, 多くの言語対において, 大規模な対訳コーパスを取得することは困難と言える. 特定の言語対で十分な量の対訳コーパスが得られない場合, 中間言語を介したピボット翻訳が有効な手法の一つである [3].

中間言語を用いる手法には様々なものが考案されている [4, 5, 6] が, 特に原言語・中間言語, 中間言語・目的言語の 2 つの翻訳モデルを合成し, 新しく得られた原言語・目的言語の翻訳モデルによって翻訳を行うテーブル合成手法で, 高い翻訳精度を得られることが知られている. ところが, 語義曖昧性や言語間の用語法の差異により, 原言語・目的言語間のフレーズ対応を正確に推定することは困難である.

図 1 (a) はテーブル合成手法によって対応を推定するフレーズの例を示しており, 図中では日本語とイタリア語のそれぞれにおける 3 つの単語が, 語義曖昧性を持つ英単語「approach」に結び付いている. このような場合, 原言語・目的言語間のフレーズ対応を求め, 適切な翻訳確率を推定するのは複雑な問題となる. その上, 図 1 (b) に示すように, 従来のテーブル合成手法では, 合成時に原言語と目的言語の橋渡しをしていた中間言語フレーズの情報が, 合成後には保存されず失われてしまう. この問題に対処するため, 本研究では原言語と目的言語を結び付けていた中間言語フレーズの情報も翻訳モデル中に保存し, 原言語から目的言語と中間言語への同時翻訳確率を推定すること



図 1 テーブル合成手法の例. および従来手法と提案手法の比較

によって翻訳を行う新しいテーブル合成手法を提案する. 図 1 (c) に, 本提案手法によって得られるフレーズ対応の例を示す. 本手法の利点は, 英語のように中間言語として選ばれる言語は豊富な単言語資源も得られる傾向が強いため, このような追加の言語情報を翻訳システムに組み込み, 翻訳の質を向上させられることにある.

中間言語フレーズの情報を翻訳時に役立てるため, 同期文脈自由文法 (Synchronous Context-free Grammar: SCFG [7]) を複数の目的言語文の同時生成に対応できるように拡張した複数同期文脈自由文法 (Multi-Synchronous CFG: MSCFG [8]) を用いて翻訳モデルの学習を行う. MSCFG による翻訳モデルを構築するために, 原言語・中間言語および中間言語・目的言語の SCFG 翻訳規則が格納されたルールテーブルを元に, SCFG ルールテーブルとしてではなく, 原言語・目的言語・中間言語の MSCFG ルールテーブルとして合成し, これによって中間言語情報を記憶す

る。訳出候補の探索時には、生成文の自然性を評価し、適切な語彙選択を促すために言語モデルを用いるが、目的言語の言語モデル（目的言語モデル）のみでなく、中間言語の言語モデル（中間言語モデル）も同時に用いた探索を行う。本手法の有効性を調査するため、欧州議会議事録を元にした Europarl 多言語コーパスから英語を中間言語とし、異なる4つの言語を用いて実験を行ったところ、すべての組合せにおいて従来手法よりも有意に高い翻訳精度が得られた。

2. 機械翻訳方式

2.1 同期文脈自由文法

本節では先ず、階層的フレーズベース翻訳 (Hierarchical Phrase-Based Translation: Hiero [7]) を代表とする様々な翻訳方式で用いられる SCFG について紹介する。SCFG は、以下のような同期規則によって構成される。

$$X \rightarrow \langle \bar{s}, \bar{t} \rangle \quad (1)$$

ここで、 X は同期規則の親記号であり \bar{s} と \bar{t} はそれぞれ原言語と目的言語における終端記号と非終端記号からなる記号列である。 \bar{s} と \bar{t} にはそれぞれ同じ数の非終端記号が含まれ、対応する記号に対して同じインデックスが付与される。以下に同期規則の例を示す。

$$X \rightarrow \langle X_0 \text{ of } X_1, X_1 \text{ の } X_0 \rangle \quad (2)$$

Chiang による SCFG の学習手法では、対訳文と単語アラインメントを元に自動的に SCFG ルールが抽出される。抽出された各々のルールには、双方向のフレーズ翻訳確率 $\phi(\bar{s}|\bar{t})$, $\phi(\bar{t}|\bar{s})$, 双方向の語彙翻訳確率 $\phi_{lex}(\bar{s}|\bar{t})$, $\phi_{lex}(\bar{t}|\bar{s})$, ワードペナルティ (\bar{t} の終端記号数), フレーズペナルティ (定数 1) の計 6 つのスコアが付与される。翻訳時には、導出に

用いられるルールのスコアと、生成される目的言語文の言語モデルスコアの和を導出確率として最大化するよう探索を行う。言語モデルを考慮しない場合、CKY+法 [9] によって効率的な探索を行ってスコア最大の導出を得ることが可能である。言語モデルを考慮する場合には、キューブ枝刈り [7] などの近似法により探索空間を抑えつつ、目的言語モデルを考慮した探索が可能である。

2.2 複数同期文脈自由文法

MSCFG [8] は、SCFG を複数の目的言語文の同時生成に対応できるように拡張された翻訳方式である。SCFG では同期規則中の目的言語記号列 \bar{t} が単一であったが、MSCFG では以下のように N 個の目的言語記号列を有する。

$$X \rightarrow \langle \bar{s}, \bar{t}_1, \dots, \bar{t}_N \rangle \quad (3)$$

通常の MSCFG 学習手法では、SCFG ルール抽出手法を一般化し、行アラインメントの取れた多言語コーパスから多言語同期規則が抽出され、複数の目的言語を考慮したスコアが付与される。

MSCFG で複数の目的言語モデルを考慮した探索を行う場合、言語毎に導出中の単語列を記憶し、組合せ毎に状態を区別する必要があるため、探索手順にも複数の手法が考えられる。SCFG の探索における単一の目的言語文を単純に複数の目的言語文に拡張して同時に展開する同時探索では、探索幅の制限により主要な目的言語文の多様性が失われてしまう可能性がある。そこで先ず、第一の目的言語文のみを考慮した組合せで探索して多様性を確保し、続いてその他の目的言語文との組合せに展開する逐次探索により、主要な目的言語を重視した効率的な探索が行える。

Neubig らによる MSCFG の学習手法では、原言語と複数の目的言語とで行アラインメントの取れた大規模な多言語コーパスが必要となるため、直接の対訳コ

一パスの取得性が限られている際には用いることができない。そこで次節からは、2つの SCFG ルールテーブルを合成することによって MSCFG ルールテーブルを得る手法について述べる。

3. ピボット翻訳手法

SMT において中間言語を用いて翻訳を行う手法は複数考案されており、以下の3種類に大別することができる。

逐次的ピボット翻訳手法 (Cascade):

入力文を原言語・中間言語の翻訳モデルで翻訳し、その出力を中間言語・目的言語文で翻訳する [3]

擬似対訳コーパス手法 (Synthetic):

原言語・中間言語あるいは目的言語・中間言語の対訳コーパスの中間言語を翻訳することで擬似的な対訳コーパスを得る [3]

テーブル合成手法 (Triangulation):

原言語・中間言語および中間言語・目的言語の翻訳モデルを合成することで原言語・目的言語の翻訳モデルを得る [4]

とりわけ、テーブル合成手法によるピボット翻訳で高品質な訳出結果が得られることが知られているため [5]、以降ではテーブル合成手法に的を絞って議論を行う。

3.1 従来のテーブル合成手法

Cohn らによるテーブル合成手法 [4] では、先ず原言語・中間言語および中間言語・目的言語の翻訳モデルを対訳コーパスによって学習し、それぞれをフレーズ (ルール) テーブル T_{SP} , T_{PT} のそれぞれにフレーズ対 $\langle \bar{s}, \bar{p} \rangle$, $\langle \bar{p}, \bar{t} \rangle$ が含まれるような中間言語フレーズ \bar{p} が存在する場合、新しい規則 $X \rightarrow \langle \bar{s}, \bar{t} \rangle$ を作成する。作成されたすべての規則について、フレーズ翻訳確率 $\phi(\cdot)$ と語彙翻訳確率 $\phi_{lex}(\cdot)$ を以下の数式に従

って推定することで原言語・目的言語のテーブル T_{ST} を合成する。

$$\phi(\bar{t}|\bar{s}) = \sum_{\bar{p}} \phi(\bar{t}|\bar{p})\phi(\bar{p}|\bar{s}) \quad (5)$$

$$\phi(\bar{s}|\bar{t}) = \sum_{\bar{p}} \phi(\bar{s}|\bar{p})\phi(\bar{p}|\bar{t}) \quad (6)$$

$$\phi_{lex}(\bar{t}|\bar{s}) = \sum_{\bar{p}} \phi_{lex}(\bar{t}|\bar{p})\phi_{lex}(\bar{p}|\bar{s}) \quad (7)$$

$$\phi_{lex}(\bar{s}|\bar{t}) = \sum_{\bar{p}} \phi_{lex}(\bar{s}|\bar{p})\phi_{lex}(\bar{p}|\bar{t}) \quad (8)$$

式 (5)-(8) は、以下のような条件を満たす無記憶通信路モデルに基いて確率周辺化を行っている。

$$\phi(\bar{t}|\bar{p}, \bar{s}) = \phi(\bar{t}|\bar{p}) \quad (9)$$

$$\phi(\bar{s}|\bar{p}, \bar{t}) = \phi(\bar{s}|\bar{p}) \quad (10)$$

ところが、現実には語の多義性や言語間の文法の不一致によって、これらの数式は正確ではない。結果として、ピボット翻訳は通常の機械翻訳よりも遥かに大きな曖昧性の問題を抱えている。

3.2 提案するテーブル合成手法

前述の曖昧性の問題に対処するため、本研究で提案するテーブル合成手法では、関連する中間言語フレーズを追加の言語情報として記憶することで、曖昧性の解消に利用する。特に、中間言語フレーズ \bar{p} で確率周辺化を行い SCFG ルールに合成する代わりに、原言語・目的言語・中間言語のフレーズ対応を MSCFG ルールとして以下のように合成する。

$$X \rightarrow \langle \bar{s}, \bar{t}, \bar{p} \rangle \quad (11)$$

このような規則を用いて翻訳を行うことによって、同時生成される中間言語文を通じて中間言語モデルなどのような追加の素性を取り入れることが可能となる。式 (5)-(8) に加えて、目的言語と中間言語を同時に考慮した翻訳確率 $\phi(\bar{t}, \bar{p}|\bar{s})$, $\phi(\bar{s}|\bar{p}, \bar{t})$ を以下のように推定する。

$$\phi(\bar{e}, \bar{p}|\bar{s}) = \phi(\bar{e}|\bar{p})\phi(p|\bar{s}) \quad (12)$$

$$\phi(\bar{s}|\bar{p}, \bar{e}) = \phi(\bar{s}|\bar{p}) \quad (13)$$

原言語・中間言語間の翻訳確率 $\phi(\bar{p}|\bar{s})$, $\phi(\bar{s}|\bar{p})$, $\phi_{lex}(\bar{p}|\bar{s})$, $\phi_{lex}(\bar{s}|\bar{p})$ はテーブル T_{SP} のスコアをそのまま用いることが可能である。これら 10 個の翻訳確率に加えて、 \bar{e} と \bar{p} に含まれる非終端記号数を 2 つのワードペナルティとし、定数 1 のフレーズペナルティの合わせて 13 個のスコアが MSCFG ルールにおける素性となる。

このように中間言語を記憶するテーブル合成手法では、 $\langle \bar{s}, \bar{e} \rangle$ ではなく、 $\langle \bar{s}, \bar{e}, \bar{p} \rangle$ の全組合せを記録するため、従来より大きなルールテーブルが合成されてしまう。計算資源を節約するためには、複数のフィルタリング手法が考えられる。Neubig らによると、主要な目的言語 T_1 と補助的な目的言語 T_2 で MSCFG を用いた翻訳を行う際には、 T_1 -フィルタリング手法 [8] が効果的である。この手法では、原言語フレーズ \bar{s} に対して、先ず言語 T_1 において $\phi(\bar{e}_1|\bar{s})$ が上位 L 個までの \bar{e}_1 を残し、それぞれの \bar{e}_1 に対して $\phi(\bar{e}_1, \bar{e}_2|\bar{s})$ が最大となるような \bar{e}_2 を残す。

4. 実験的評価

4.1 実験設定

本研究の提案手法の有効性を評価するため、欧州議会議事録を広くカバーし、ピボット翻訳のような多言語翻訳タスクで広く用いられる Europarl コーパス [10] を用いて実験を実施した。本実験では、英語 (En) を中間言語として固定し、欧州の中でも特に話者数の多いドイツ語 (De)、スペイン語 (Es)、フランス語 (Fr)、イタリア語 (It) の 4 言語の組合せでピボット翻訳を行い、手法毎の翻訳精度を比較した。5 言語間の対訳コーパスを得るため、先ず Gale-Church アラインメント法 [11] によって行アラインメントの取れた多言語コーパス約 90 万文を取得し、そこから 1,500 文ずつを最適化と評価用に取り出した。それぞれの翻訳

モデルの学習には 10 万文、目的言語モデルの学習にも 10 万文を用いた。また、多くの場合、英語においては大規模な単言語資源が取得可能であるため、最大 200 万文までのデータを用いて段階的に規模の異なる中間言語モデルの学習を行った。翻訳結果の評価には、自動評価尺度 BLEU [13] を用い、各翻訳モデルは MERT [14] により、開発データセットに対して BLEU スコアが最大となるようにパラメータ調整を行った。提案手法のテーブル合成手法によって得られた MSCFG ルールテーブルは、 $L = 20$ の T_1 -フィルタリング手法によって枝刈りを行った。また、本実験の MSCFG を用いた探索では、目的言語文、中間言語文の順に組合せを展開する逐次探索を行った。

次節では以下の 6 つの翻訳手法を比較評価する。

Direct:

比較のため、中間言語を用いずに原言語・目的言語の直接対訳コーパスで学習した SCFG で翻訳

Cascade:

原言語・中間言語および中間言語・目的言語の SCFG モデルで逐次的ピボット翻訳

Tri. SCFG:

原言語・中間言語および中間言語・目的言語の SCFG モデルを合成し、原言語・目的言語の SCFG モデルによって翻訳 (ベースライン, 3.1 節)

Tri. MSCFG:

原言語・中間言語および中間言語・目的言語の SCFG モデルを合成し、原言語・目的言語・中間言語の MSCFG によって翻訳。「-PivotLM」は中間言語モデルを用いないことを示し、「+PivotLM 100k/2M」はそれぞれ 10 万文、200 万文で学習した中間言語モデルを用いることを示す (提案手法, 3.2 節)

4.2 翻訳精度の比較

表 1 に、英語を介したすべての言語対におけるピボット翻訳の結果を示す。実験で得られた結果は、ブートストラップ・リサンプリング法 [15] により統計的

Src	Trg	BLEU Score [%]					
		Direct	Cascade	Tri. SCFG (baseline)	Tri. MSCFG -PivotLM	Tri. MSCFG +PivotLM 100k	Tri. MSCFG +PivotLM 2M
De	Es	27.10	25.05	25.31	25.38	25.52	† 25.75
	Fr	25.65	23.86	24.12	24.16	24.25	† 24.58
	It	23.04	20.76	21.27	21.42	† 21.65	‡ 22.29
Es	De	20.11	18.52	18.77	18.97	19.08	† 19.40
	Fr	33.48	27.00	29.54	† 29.87	† 29.91	† 29.95
	It	27.82	22.57	25.11	25.01	25.18	‡ 25.64
Fr	De	19.69	18.01	18.73	18.77	18.87	† 19.19
	Es	34.36	27.26	30.31	30.53	† 30.73	‡ 31.00
	It	28.48	22.73	25.31	25.50	† 25.72	‡ 26.22
It	De	19.09	14.03	17.35	† 17.99	‡ 18.17	‡ 18.52
	Es	31.99	25.64	28.85	28.83	29.01	† 29.31
	Fr	31.39	25.87	28.48	28.40	28.63	† 29.02

表 1 各手法と言語の組合せにおける翻訳精度

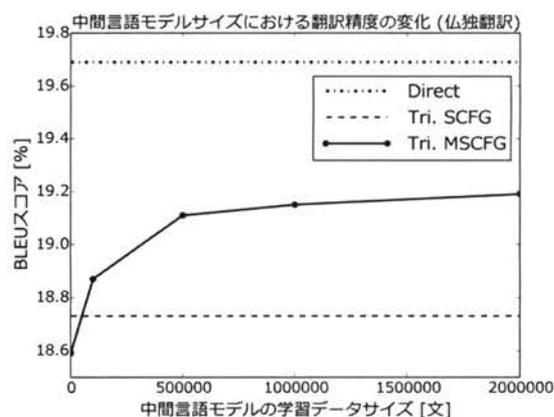


図 2 中間言語モデル規模がピボット翻訳精度に与える影響

有意差を検証した。太字はそれぞれの言語対において最も BLEU スコアが高いことを示し、短剣符は各ピボット翻訳手法の翻訳精度が従来手法 Tri. SCFG よりも統計的に有意に高いことを示す (†: $p < 0.05$, ‡: $p < 0.01$)。

評価値から、提案したテーブル合成手法で中間言語モデルを考慮した翻訳を行った場合、すべての言語対において従来のテーブル合成手法よりも BLEU スコアが上昇していることが確認できる。すべての組合せにおいて、テーブル合成手法で中間言語情報を記憶し、200 万文の言語モデルを考慮して翻訳を行った場合に最も高いスコアを達成しており、従来法に比べ 0.4 から 1.2 ほどの BLEU 値の向上が見られる。このことから、中間言語情報を記憶し、これを翻訳に利用するこ

とが曖昧性の解消に繋がり、安定して翻訳精度を改善できたと考えられる。

また、MSCFG に合成するが中間言語モデルを用いずに翻訳を行う場合を見てみると、SCFG に合成する場合よりも多くの言語対で僅かに精度が向上している。これは、追加の翻訳確率などのスコアが有効な素性として働き、適切な語彙選択に繋がったことなどが原因として考えられる。

4.3 中間言語モデルの規模が翻訳精度に与える影響

中間言語モデルのサイズがピボット翻訳精度に与える影響の大きさは言語対によって異なっているが、中間言語モデルの規模が大きくなるほど精度が向上することも確認できる。図 2 はフランス語・ドイツ語のピボット翻訳において異なるデータサイズで学習した言語モデルが翻訳精度に与える影響を示す。

4.4 曖昧性が解消された例と未解決の問題

本提案手法によって中間言語側で曖昧性が解消されて精度向上に繋がったと考えられる訳出の例を示す。

入力文 (ドイツ語):

Ich bedaure, daß es keine gemeinsame Annäherung gegeben hat.

参照訳 (イタリア語):

Sono spiacente del mancato **approccio commune** .
対応する英文:

I regret that there was no **common approach** .

Tri. SCFG:

Mi rammarico per il fatto che non si
ravvicinamento commune . (BLEU+1: 13.84)

Tri. MSCFG+PivotLM 2M:

Mi dispiace che non esiste un **approccio commune** .
(BLEU+1: 25.10)

I regret that there is no **common approach** .
(同時生成された英文)

上記の Tri. MSCFG+PivotLM 2M で導出に用いられた MSCFG ルールでは、イタリア語の「**approccio**」と英語の「**approach**」が結び付いており、生成される英文中の単語の前後関係から適切な語彙選択を促し、精度向上に繋がったものと考えられる。

逆に、提案手法では語彙選択がうまくいかず、直接対訳で学習した場合よりも精度が落ちた訳出の例を示す。

入力文 (フランス語):

Vous avez tout à fait raison et je vous **remercie**
d'avoir attiré l'attention sur ce point .

参照訳 (スペイン語):

Tiene usted toda la razón y le **agradezco** que nos
llame la atención sobre este punto .

対応する英文:

You are quite right and I **thank you** for drawing
our attention to this matter .

Direct:

Tiene usted razón y le **agradezco** que haya llamado
la atención sobre este punto . (BLEU+1: 56.00)

Tri. MSCFG+PivotLM 2M:

Tiene usted mucha razón y **gracias** por haber
conseguido la atención sobre este punto .

(BLEU+1: 38.91)

You have quite right and I **thank you** for having
courageously brought the attention on this point .
(同時生成された英文)

この例では、英単語「**thank**」に対応するスペイン語の動詞活用形「**agradezco** (私は感謝する)」と名詞の間投詞的用法「**gracias** (感謝, ありがとう)」に訳が分かれてしまっている。この例では品詞が異なっても意味としては大きく変化しておらず、単純に1つしか存在しない正訳と異なるためにスコアが下がっている。しかし、このような多品詞語によって生じる曖昧性の問題もピボット翻訳では多く直面し、品詞を誤ることで文法を誤る可能性だけでなく、異なる意味の文になる場合もある。

本節で示した例のように、本研究の提案手法でテーブル合成手法に中間言語情報と中間言語モデルを用いることで、導出に関する中間言語フレーズの単語の並びを評価して曖昧性を解消できた部分と、これだけでは不十分な部分があることが分かる。

4.5 品詞ごとの翻訳精度

前節では、英語の多品詞語の問題について触れた。また、英語は他の欧州諸言語と比較して、性・数・格に応じた活用などが簡略された言語としても有名であり、語形から統語情報が失われることで発生する曖昧性の問題もある。本節では、英語を介したドイツ語・フランス語の両方向の翻訳において、誤りの発生しやすい品詞について調査する。先ず、独仏・仏独翻訳における評価データの参照訳および各ピボット翻訳手法の翻訳結果に対して Stanford POS Tagger [16, 17] を用いて品詞付与を行い、参照訳と翻訳結果を比較して、語順は考慮せずに適合率と再現率の調和平均である F 値を算出した。

表 2 および表 3 は、各翻訳における高頻出品詞の正解率を表している。出現頻度は、参照訳中の各品詞の出現回数を意味し、丸括弧内に示された数値は、提案

品詞	出現頻度	F-Measure [%]		
		Direct	Tri. MSCFG +PivotLM 2M	Tri.SCFG
NC (一般名詞)	8,360	78.94	78.19 (+0.01)	78.18
P (前置詞)	6,553	92.58	92.46 (+1.02)	91.44
DET (定冠詞)	5,363	91.04	90.63 (+1.50)	89.13
PUNC (区切り記号)	4,258	93.89	93.77 (+0.44)	93.33
ADJ (形容詞)	3,725	71.36	69.85 (+0.67)	69.18
V (動詞)	3,324	69.52	63.86 (+1.06)	62.80
ADV (副詞)	2,238	81.98	77.81 (+0.34)	77.47

表 2 独仏翻訳における品詞ごとの翻訳精度

品詞	出現頻度	F-Measure [%]		
		Direct	Tri. MSCFG +PivotLM 2M	Tri.SCFG
NN (一般名詞)	10,813	78.04	75.42 (+0.53)	74.89
ART (冠詞)	4,502	90.38	87.21 (+1.24)	85.97
CARD (数詞)	3,849	94.56	94.38 (-0.44)	94.82
APPR (前置詞)	2,577	89.66	84.73 (+1.50)	83.23
ADJA (形容詞)	2,311	65.96	65.29 (+0.47)	64.82
NE (固有名詞)	2,005	77.52	75.34 (+0.47)	74.87
ADV (副詞)	1,848	74.32	70.13 (+0.18)	69.95
PPEF (再帰代名詞)	1,665	92.64	87.46 (+1.67)	85.79

表 3 仏独翻訳における品詞ごとの翻訳精度

手法とベースライン手法の F 値の差分である。結果は言語対依存であるが、特に目的言語に強く依存していることが明らかである。

表 2 の独仏翻訳の例では、提案手法によって、ベースライン手法よりも特に前置詞、定冠詞、動詞で F 値が大きく向上している。一般名詞、形容詞、動詞、副詞は重要な内容語であり、語彙選択の幅も広いから、どの手法でも全体的に F 値が低くなっている。一般名詞に関しては、通常のテーブル合成手法でも Direct と大きな差は出しておらず、そのため提案手法でもほとんど改善されなかった。一方で、動詞の F 値は提案手法で大きく向上しており、中間言語モデルによって語の並びを考慮して語彙選択を行うことで翻訳精度向上に繋がったと考えられる。しかし、それでも Direct には大きく及ばず、頻度のあまり高くない内容語の語彙選択を適切に行うことは非常に難しいことが分かる。一方で、機能語においては提案手法において Direct と近い F 値が得られた。

表 3 の仏独翻訳の例では、表 2 の場合と少し変わっており、冠詞、前置詞、再帰代名詞のような機能語で、

提案手法によって F 値が大きく改善されているものの、Direct と比べると大きな差があり、ピボット翻訳で精度が大きく落ちる原因と考えられる。冠詞や前置詞は機能語であるものの、ドイツ語では男性形・女性形に加えて、英語にもフランス語にも無い中性形の活用を持っているため、フランス語よりも活用の種類が多く、また冠詞や前置詞も格に応じた活用をすることが知られている。原言語を英語に翻訳した際には統語的情報が失われることが多いから、機能語にも活用幅があるような目的言語に対してはピボット翻訳が特に困難であることが多い。

5. まとめ

本研究の目的は、多言語機械翻訳における翻訳精度の向上を目指し、従来のピボット翻訳手法を調査、問題点を改善して翻訳精度の向上を図ることであった。テーブル合成手法によって他のピボット翻訳手法よりも安定して高い翻訳精度を得られることが知られているが、従来手法では中間言語情報が失われ、曖昧性により翻訳精度が減少する問題がある。この問題に対処するため、中間言語情報を記憶し、中間言語モデルを利用して自然な語順の語彙選択を促すことで精度を向上させるテーブル合成手法について提案した。本論文で提案した手法を用いて、欧州議会議事録コーパスの 5 言語のデータによってピボット翻訳の比較評価を行ったところ、すべての言語の組合せで従来のテーブル合成手法よりも高い翻訳精度が得られた。また、特に大規模な中間言語モデルを用いることで、より適切な語彙選択が促されて翻訳の質を高められることが分かった。しかし、提案手法でも、直接の対訳コーパスを用いて学習を行った理想的な状況と比較すると、精度の開きが大きく、中間言語モデルをさらに大きくするだけでは解決できないであろう点も示唆された。本提案手法で解決できなかった曖昧性の問題を調査すべく、特定の言語対で品詞ごとの単語正解率を求めたところ、多品詞語や格活用など異なる要因によって、語順を考

慮するだけでは解消されない問題が残されていることが分かった。今後の課題として、多品詞語や格活用などの統語的な曖昧性の問題に対して、構文情報を用いて対処していくことを考えている。

参考文献

- [1] Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, Vol. 19, pp. 263–312, 1993.
- [2] Christopher Dyer, Aaron Cordova, Alex Mont, and Jimmy Lin. Fast, easy, and cheap: construction of statistical machine translation models with MapReduce. In *Proc. WMT*, pp. 199–207, 2008.
- [3] Adrià de Gispert and José B. Mariño. Catalan-English Statistical Machine Translation without Parallel Corpus: Bridging through Spanish. In *Proc. of LREC 5th Workshop on Strategies for developing machine translation for minority languages*, 2006.
- [4] Trevor Cohn and Mirella Lapata. Machine Translation by Triangulation: Making Effective Use of Multi-Parallel Corpora. In *Proc. ACL*, pp. 728–735, June 2007.
- [5] Masao Utiyama and Hitoshi Isahara. A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation. In *Proc. NAACL*, pp. 484–491, 2007.
- [6] Xiaoning Zhu, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang, and Tiejun Zhao. Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs. In *Proc. EMNLP*, 2014.
- [7] David Chiang. Hierarchical phrase-based translation. *Computational Linguistics*, Vol. 33, No. 2, pp. 201–228, 2007.
- [8] Graham Neubig, Philip Arthur, and Kevin Duh. Multi-Target Machine Translation with Multi-Synchronous Context-free Grammars. In *Proc. NAACL*, 2015.
- [9] Jean-Cédric Chappelier, Martin Rajman, et al. A Generalized CYK Algorithm for Parsing Stochastic CFG. *TAPD*, Vol. 98, No. 133-137, p. 5, 1998.
- [10] Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, Vol. 5, pp. 79–86, 2005.
- [11] William A Gale and Kenneth W Church. A program for aligning sentences in bilingual corpora. *Computational linguistics*, Vol. 19, No. 1, pp. 75–102, 1993.
- [12] Graham Neubig. Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers. In *Proc. ACL Demo Track*, pp. 91–96, 2013.
- [13] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proc. ACL*, pp. 311–318, 2002.
- [14] Franz Josef Och. Minimum Error Rate Training in Statistical Machine Translation. In *Proc. ACL*, pp. 160–167, 2003.
- [15] Koehn, P. (2004). “Statistical Significance Tests for Machine Translation Evaluation.” In *Proc. EMNLP*, pp. 388–395.
- [16] Kristina Toutanova and Christopher D. Manning. Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger In *Proc. EMNLP*, pp. 63–70, 2000.
- [17] Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network In *Proc. NAACL*, pp. 173–180, 2003.

機械翻訳事始め[VIII]

坂本義行

アジア太平洋機械翻訳協会理事

今回は、日本を中心に、1987年度～1994年度（8年間）にわたって、アジア各国と協力して行われた「近隣諸国間の機械翻訳システムに関する研究協力」について、報告書^{1),2)}に基づいて、その概要を報告します。

1. はじめに

この研究は、財団法人 国際情報協力センターの共同研究として行われました。

2. 目的

機械翻訳システムに関する研究協力事業は、最新の情報処理システムを近隣諸国と共同研究開発すること、この共同研究開発から派生する母国語情報処理技術の基盤を確立すること及びその他さまざまな技術交流や文化交流を促進することを目的として実施された。

技術面では、日本、中国、インドネシア、マレーシア及びタイのアジア5カ国を対象として、これらのどの国の言語にも翻訳できる方式の研究開発を行うことを、経済面では、情報処理分野の文献を実用的な速さで及び正確さで翻訳を行うことを目標とした。

3. 概要

- ① 名称：Multi-lingual Machine Translation(MMT) Project
- ② 対象国：中国、インドネシア、マレーシア及びタイ

- ③ 期間：1987年度～1994年度（8年間）
1995年度、1996年度（2年間）フォローアップ

- ④ 予算：162百万円（87年度）
332百万円（88年度）
639百万円（89年度）
880百万円（90年度）
902百万円（91年度）
981百万円（92年度）
981百万円（93年度）
948百万円（94年度）
累計 5,825百万円
16百万円（95年度）
16百万円（96年度）

累計 5,857百万円

⑤ 内容：

中間言語方式により、日本、中国、インドネシア、マレーシア及びタイのアジア5カ国間の言語を対象として、これらのどの国の言語にも翻訳できる機械翻訳システムの研究開発を行うこと。

4. 実施体制

- ① 通商産業省からの委託を受けて、国際情報化協力センターが実施。
- ② 日本語側では、電子技術総合研究所、(株)日本電子化辞書研究所(EDR)、(財)情報処理相互運用技術協会(INTAP)、関連企業(沖電気工業、シャープ、東芝、日本電気、日立製作所、富士通、三菱電機)、学術経験者等の協力参加。

③共同研究の相手方：中国計算機軟件與技術服務總公司（CS&S）、インドネシア技術評価応用庁（BPPT）、マレーシア教育省（MOE）、タイ国立コンピュータ技術センター（NECTEC）等。（マレーシアについては、1995年にマレーシア国立翻訳研究所（ITNM）に移管、さらにマレーシア工科大学に移管）

5. 内容及び成果

このプロジェクトでは、5カ国語間の機械翻訳システムにおいて「中間言語方式」を採用した。研究開発の内容及び成果として、中間言語の仕様を作成、作成した仕様に基づく機械翻訳システムの開発及び電子化辞書・コーパスの作成、開発したシステムによる機械翻訳の実証実験等が挙げられる。

(1) 中間言語の仕様の作成

対象とする5カ国で表現された文章を独自の表現形式で表すための仕様（規則、記述方法等）を作成した。当初の仕様は、何度も見直しが行われ、改善が図られた。

(2) システム開発

作成された中間言語の仕様をベースにして、機械翻訳システムが開発された。このシステムは、入出力システム、翻訳支援システム、文解析システム、文生成システム、電子化辞書システム、文章ファイル管理システム、ネットワークシステム等から構成された。

(3) 電子化辞書及びコーパス作成

機械翻訳システムを実際に稼働させるためには電子化辞書が不可欠で、対象とした各国語について、基本語および専門用語の電子化辞書を作成した。また、実証実験の対象データとして、各国語で記述された3000の例文（コーパス）を作成した。

(4) 実証実験

開発された機械翻訳システム及び作成された電子化辞書を使用して、作成したコーパスを対象に各国語間での翻訳の実証実験を行った。最終的に、目標とした少なくとも50%の翻訳率をクリアした。

6. 経過

1987年度から1994年度までの8年をかけて、機械翻訳システムの研究開発を進めた。1995年度及び1996年度には、8年間の共同研究の成果をもとに、フォローアップ事業を実施した。

(1) 機械翻訳システムの基礎研究（1987、88年度）

最初の2年間で、機械翻訳システムを構成するいくつかのシステムと要素の基本仕様を設計し、試作した。約5000語の辞書と数十の例文を作成し、文法規則の仕様を設計した。また、限定された翻訳例文を使って、機械翻訳システムの処理能力を確認した。

(2) 機械翻訳システムの研究開発（1989～92年度）

次の4年間で、基本語辞書約5万語、情報処理関連の専門語2万5千語及び約3000の例文を作成し、文解析や文生成を制御する文法規則を試作した。また、追加した翻訳例文と任意の自然文を使って、機械翻訳システムの翻訳処理能力を確認した。

(3) 機械翻訳システムの評価と改良（1993、94年度）

最後の2年間で、(2)段階までに試作された機械翻訳システムを使って、翻訳処理能力実験を繰り返して、辞書、文法規則およびシステム全体を改善した。

(4) 機械翻訳のフォローアップ事業（1995、96年度）

8年間の共同研究の成果をもとに、研究に参加した

各国において、辞書・文法の改良・追加・機械翻訳システム実用化に向けた追加研究を独自に実施できるように支援した。

7. 各国の研究機関の状況

(1) 中国

CS&S は、本プロジェクトにより大きく発展した。また、本プロジェクトに参加した清華大学等の大学における言語情報処理に関する研究能力も向上した。

その後、CS&S は、独自に中一英、中一日の機械翻訳システムを開発した。

(2) インドネシア

BPPT は、この分野ではゼロからの出発で、大きく成長、発展した。開発した機械翻訳システムを維持し、インターネットによる電子化辞書、翻訳サービスを行った。また、独自に1997年から2002年までの自然言語処理に関する研究計画を策定し、実施した。

(3) マレーシア

ITNT は、機械翻訳システムの研究から撤退し、マレーシア工科大学へ引き継いだ。INTN の実施していたマレーシア語のスペルチェッカーの開発、電子化辞書の生産は、残念ながら中止となった。

(4) タイ

NECTEC は、1997年初めに設立10周年の記念行事を行ったが、本事業とともに発展した。独自にCD-ROMによる電子化辞書を開発・製品化（タイ語全般、タイ語用例、タイ語同義語・反義語、タイ語—英語、英語—タイ語）し、その改訂版を製作した。

8. フォローアップ活動

CICC としては、これまでの10年間の研究開発の実績、培われた人的ネットワーク等を活用して、関係国との間で、機械翻訳に関する情報交換等を独自に実施した。具体的には、①研究者、研究機関リストの作成・提供、②電子メールやインターネット電子会議による情報交換、③各国レポート等のホームページによる情報提供、④本プロジェクトに関する小冊子の作成等を行った。

また、機械翻訳に関する新しいプロジェクトの可能性を検討するなど、関係機関との連携を図った。

9. 背景

当時、通商産業省は、アジアにおけるコンピュータ産業およびAI技術の発展を支援するために、ワークステーションタイプの計算機を日本の各メーカーから、各国に提供し、各国がその計算機上にソフトウェアを開発する技術を習得するために、機械翻訳ができる環境を構築することを支援するという目的があったという話を、私は耳にしたことがある。

【参考文献】

- 1) CICC 事業実績、P48-50, (財) 国際情報処理協力センター、平成18年6月。
- 2) CICC のホームページより、国際共同研究開発、「近隣諸国間の機械翻訳システムに関する研究協力 (MT)、
<http://www.cicc.or.jp/japanese/kyoudou/mt.html>

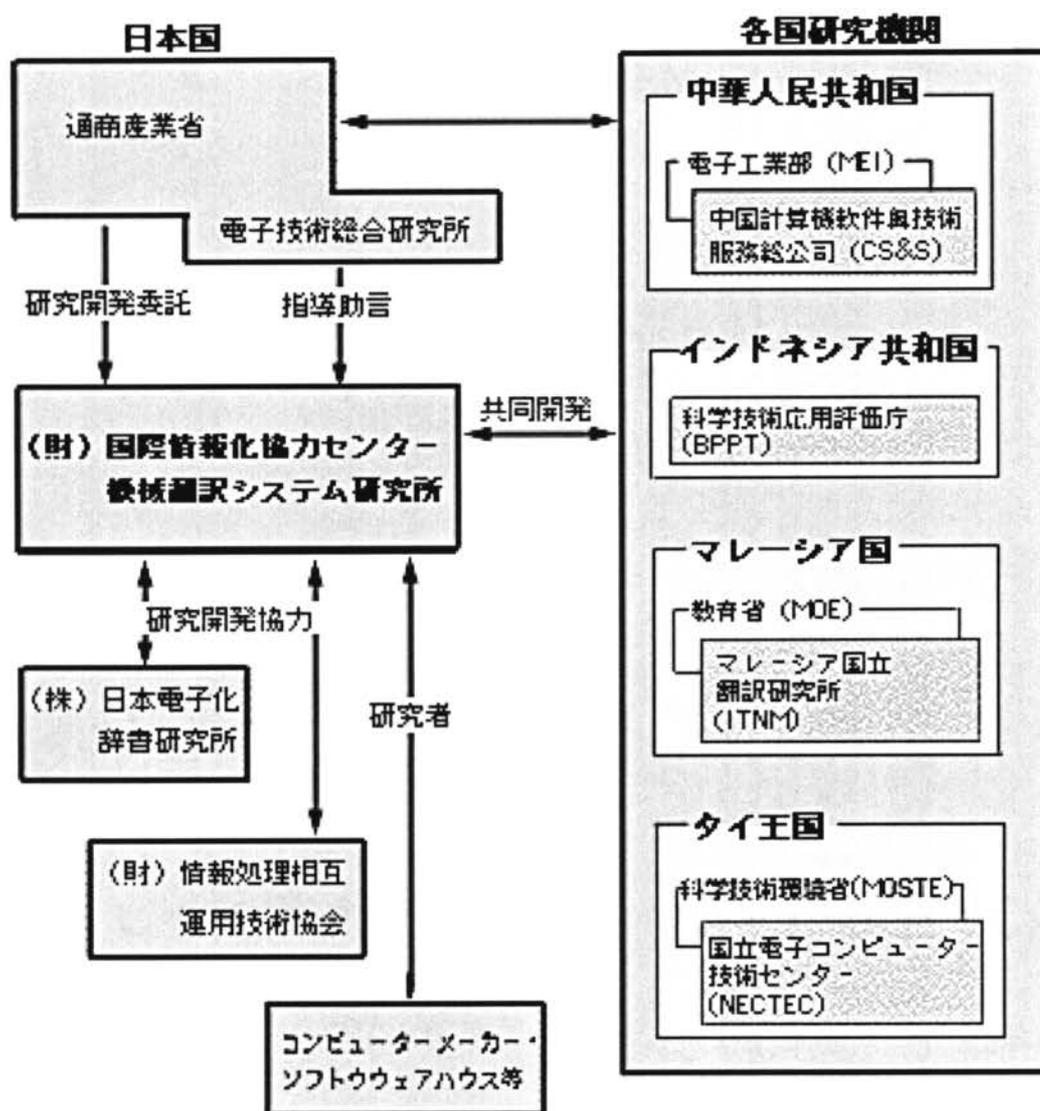


図1 体制図

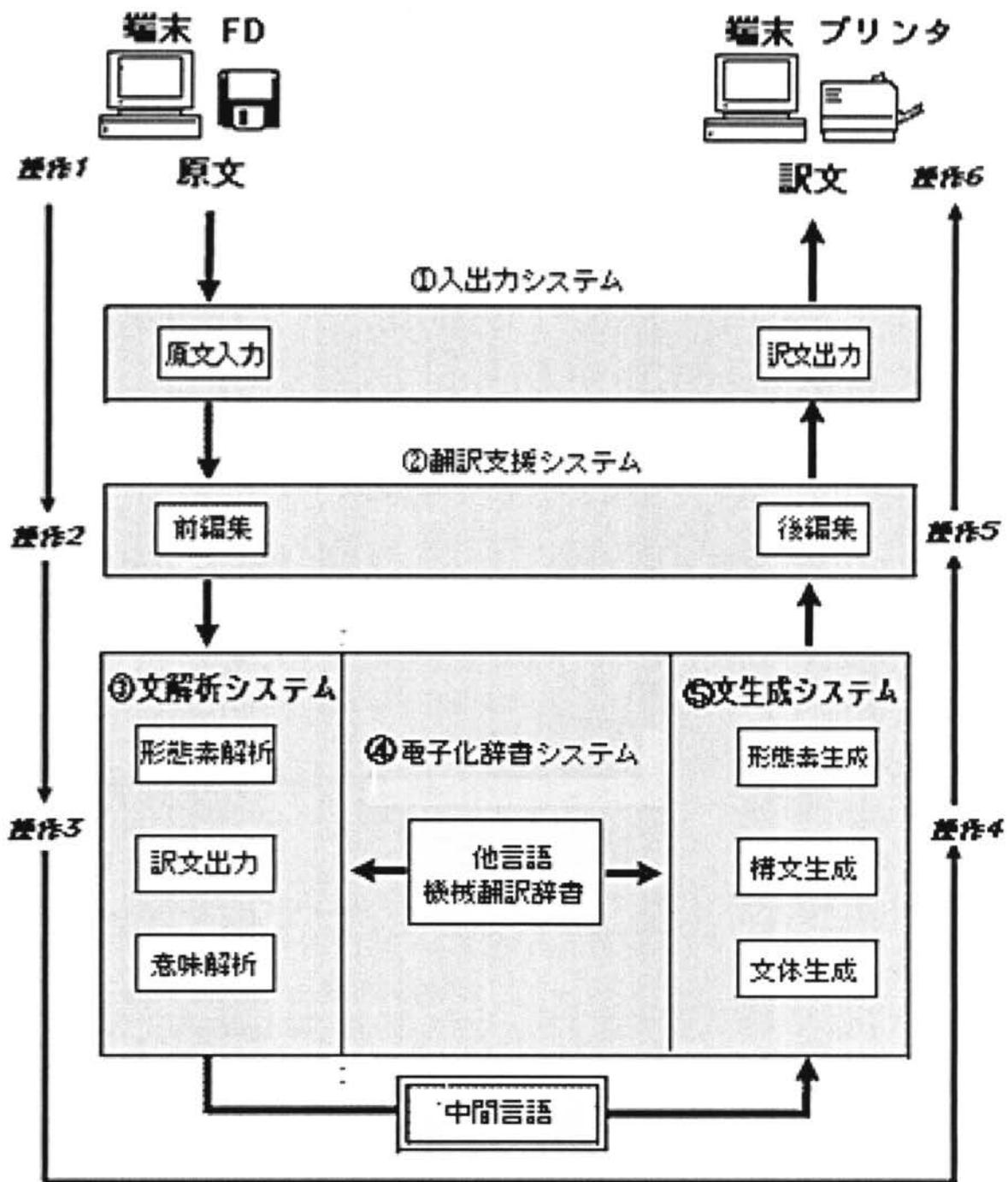


図2 機械翻訳システムの構成と翻訳の流れ

言語処理学会テーマセッション

『文理・産学を越えた翻訳関連研究』開催報告

山田 優 藤田 篤 影浦 峯
 関西大学 情報通信研究機構 東京大学

1. はじめに

2016年3月の言語処理学会第22回年次大会(NLP2016@東北大学)¹では、『文理・産学を越えた翻訳関連研究』と題したテーマセッションが開催された。本稿では、同セッションの提案者として、その時の模様を振り返りながら、自然言語処理(NLP)と翻訳研究(TS)を取り巻く状況をまとめ、今後の研究と活動の方向性を提示する。

テーマセッションの目的は、自然言語処理研究者だけでなく、翻訳者や翻訳ツールの開発者、翻訳研究者が文理・産学の垣根を越えてニーズとシーズを共有することであった。各分野のプレイヤーに属する技術、資産、人材、問題意識は、他分野プレイヤーにとって新たなニーズやシーズとなりうる。同セッションを通して異なる分野のプレイヤーが交流し共同研究や有機的な連携が促進されることを期待した。

上記年次大会ウェブサイトのセッション E5、E6、E7にあるように、本テーマセッションでは合計12の発表があった。これらは以下の3つに大別できる(図1も参照されたい)。

a) 自然言語処理・翻訳研究

NLPを中心とした情報処理・分析の研究、翻訳研究、機械翻訳(MT)や関連技術の進展に関与する研究者

b) 翻訳産業

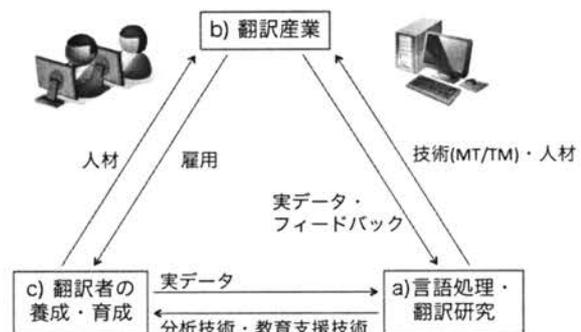
翻訳会社(LSP)、翻訳者などの実務者

c) 翻訳者の養成・育成

大学のコース、職業訓練校の教育者

発表すべてが、きれいに分けられるわけではないが、このように図式化することで、プレイヤーの守備範囲と関係が把握できる。以下では、セッションでの発表内容を図1の範疇a、b、cに位置づけつつ、2者間ごとのインタラクション(すなわち三角形の各辺)に着目して、研究課題、連携の可能性、拓かれた問いとそれに対する議論を詳述する。

図1: 各プレイヤーの関係図



2. テクノロジーの創出と運用: 工学的な産学連携

まず図1三角形の右辺に着目する。ここには、a)自然言語処理/翻訳研究とb)翻訳産業の連携がある。平たく言えば、アカデミアと産業、いわゆる産学連携である。

a)自然言語処理/翻訳研究は、言語処理や翻訳に関する基礎研究を行うので、その技術は、少なくとも理

¹ <http://www.anlp.jp/nlp2016/>

屈の上では翻訳産業で応用できるはずである。また少し技術研究の範囲を広げて考えるならば、業務管理や翻訳準備（前処理等）への応用も考えられる。翻訳研究は、人間と機械のインタラクションやその文脈（社会的構造）を研究対象とする。つまり、a) 自然言語処理・翻訳研究にとっては、b) 翻訳産業は、技術の出口であり研究題材の場の一つでもある。これに対して、b) 翻訳産業は、翻訳を実践する現場であり、彼らにとって、a) 自然言語処理・翻訳研究はキモになる道具・資料となりうる。

産業側からの発表として、河野[4]は、MTを始めとする技術革新をいかに活用していくのが要になると主張した。例えば、翻訳メモリ（TM）を使った作業のうち、70%以上のセグメントではMT訳を下訳として利用しているという実情がある。ただし、業界内ではMTを活用して得られた効率性向上とその報酬に関して翻訳会社側とプロ翻訳者の間で十分な相互理解がなされておらず、翻訳者の間ではMTに対する否定的意見も少なくない。MTはまさに破壊的イノベーションであり、これからの時代の新しい顧客や価値観をもたらすものであるから、いま乗り越えなければならない壁も高い。一方で、ポストエディット（翻訳後編集、PE）を含めて、MT結果を人間（の翻訳者）が利用するものとみなす限り、人間支援という視点を重視したシステム開発も必要である。アカデミアにはこのような視点での研究開発が期待される。

プロ翻訳者の多くがMTに否定的な意見を持つ状況において、プロの翻訳者ではない人向けのCATツールを提案するのがヤラクゼン[7]である。主にポストエディットに最適化したMTとユーザインタフェースを開発し、漠然とした「翻訳品質」よりも「ユーザー満足度」を重視したツールの開発に務める。すでに1万人以上のユーザーを抱える同サービスは翻訳コーパスも蓄積しており、研究者との連携も期待される。

宮田ら[6]は、プリエディット（翻訳前編集）、すなわちMTによる翻訳を前提としたオーサリング支援を検討している。この研究では、起点言語の話者に、満

足のいくMT訳が得られるまで、原文を試行錯誤にプリエディットさせて、その工程を記録し、修正履歴の分析を通じてプリエディットの全体像と有用性を見出すことを試みている。

人間翻訳者の振舞いは、作業モード（例えば入力メソッド）により異なる。Carlら[2]は、キーボード入力による人手翻訳（HT from scratch）、ポストエディット（PEMT）、音声認識入力（dictation）の3つのモードを比較し、作業効率の観点から、音声認識入力での翻訳がPEMTに迫ることが示唆された。この他にも、PEにおける人間の負荷の指標として作業中の平均ポーズ時間が有用である（相関が高い）ことが示された[5]。この指標を、逆にMT品質の測定に応用できるのではないかという意見も交わされた。

発表が示すように、a) 自然言語処理・翻訳研究とb) 翻訳産業の連携は拡大しているが、まだまだ発展の余地がある。今後は翻訳産業界とアカデミアの更なる連携に期待したい。実データを用いて実世界の問題を解決するというスタイルの連携は、アカデミアにとっても、(若い) 研究者を呼びこむ意味で有益である。また、翻訳産業界のテクノロジーの高度化に伴い、SEあるいは社内開発のために理系人材の活用が活発化することも予想される。

3. 翻訳産業が求める人材と翻訳者の育成: 文系人材の市場

次に左辺のb) 翻訳産業とc) 実務翻訳者の養成・育成の関係を見る。上述したようにa) 自然言語処理・翻訳研究を前提としたテクノロジーの発展により、その進化に対応すべく、b) 翻訳業界で求められるスキル、能力、人材も変化している。

MTを始めとする技術の展開は、翻訳産業の現場だけでなく、翻訳教育およびその背景にある、翻訳者に求められる力とは何かという理論的な問題にまで影響を与えている。とりわけ、大学院レベルでの翻訳教育はあまりなされておらず、翻訳産業との結びつきもあ

まり強くない日本の大学では、翻訳教育カリキュラム全体の設計と同時にその中に翻訳関連技術をどう位置づけるかについても検討しなくてはならない状況に置かれている。これは、翻訳関連技術も考慮した見通しのよいカリキュラムを構成する観点からチャンスでもあるが、一方で、翻訳実務を考慮した基本的な翻訳教育実践がそれほど進んでいない状況でそうしたカリキュラムが仮に作られたとしても砂上の楼閣になりかねない危険性が存在しているということでもある。

また、翻訳者として生計を立てていくのは、理系のSE等が翻訳産業で働くのとは異なり、様々な意味でハードルが高い。そもそも職人芸的な要素が強い仕事であるのに加え、近年多様化が進む業務形態への対応も求められる。例えば、PEMTへの対応を考えた場合、そのための訓練をどのように教育カリキュラムに折り込むべきかは白紙の状態である。それ以前に、PEMTで生計を立てることができるという見通しを作ることが、まず翻訳業界に求められているとも言える。

教育側はこうした業界の実情をふまえた人材育成に迫られている。近年のISOの翻訳サービス規格（ISO 17100）の発行を契機に、国内でも専門職教育に対する必要性の意識が高まっている。立教大学を中心に「翻訳通訳教育研究会」が立ち上がり、専門職訓練を意図した学科・コースもようやく開設しつつある。また、翻訳教育実践を統合的にサポートするシステムの開発と利用を目的とするプロジェクトなども進められている。

カリキュラムの作成に際しては、すでに大学での翻訳者養成が社会に浸透している国・地域における取り組みが参考になる。例えば、モンレー国際大学（現、ミドルベリー国際大学）における翻訳訓練プログラムでは、これまでは語学力とドメイン知識が中心であったが、現在はそれらに加えて翻訳管理能力やTMなどのツールなどのトレーニングが必須となっている。また翻訳産業界では、ローカリゼーション翻訳管理（マネージメント能力）に特化したプログラムの修了生に対するニーズが高まっているという。

現時点では、教員不足の問題、特に現状では b) 翻訳業界の実務翻訳者が c) 大学教育現場で教える人材になるためのコンピテンシーの整理、という問題が未解決である。欧米では両者が固く結びついている。日本でも同じような連携の実現が急がれる。

本セッションでは、業界の現場と教育についての問題が確認できたが、実際に b) 翻訳産業と c) 実務翻訳者養成が連携した発表はなかった。このことは、実践レベルで両者の連携があまり進んでいないという大きな問題を示している。

4. 翻訳教育におけるテクノロジーの活用: 文理融合・教育工学的アプローチ

3節の状況を想定して、最後に三角形の下辺である a) 自然言語処理研究・翻訳研究と c) 実務翻訳者養成の関係を見る。今回のテーマセッションでは、翻訳者の養成と翻訳関連技術、翻訳者に求められる力の関係をめぐる共同研究に関する発表も行われた。

影浦ら[3]は、産業翻訳のプロセスで標準的に採用されており、ISO17100でも前提とされているプロジェクト型翻訳のプロセスをエミュレートしつつ、翻訳の力を総合的に高めるための翻訳教育/学習システム「みんなの翻訳実習」の背景と理念を紹介した。そこでは、翻訳者の力として、狭義の「翻訳力」とプロジェクト管理やプロジェクト参加者間のコミュニケーションも含めた「翻訳者力」を区別し、システムが「メタ言語」の修得を通じた翻訳プロセスに対する反省的認識を促す「足場」を提供することで、翻訳力と翻訳者力を同時に涵養する見通しが示された。

上述の「みんなの翻訳実習」では、欧州の翻訳教育プログラムを通して開発された校閲カテゴリ（MeLLANGE）をベースに、新たに英日・日英翻訳への適用を想定した校閲カテゴリが設計された。豊島ら[8]は、実際の翻訳教育現場において校閲カテゴリを用いるための方法論的検討を行った。翻訳の学習者が作成した英日翻訳における校閲箇所のアノテーション

および校閲カテゴリ分類を通じて、校閲カテゴリを一貫して付与すること、ひいては校閲カテゴリが教育者と学習者の間の意思疎通のための「メタ言語」として利用しうることが示された。豊島らの調査をふまえ、山本ら[10]は、翻訳の初学者（学部学生）を対象とした翻訳演習講義において校閲カテゴリを用いた添削を繰り返し実施し、学習者の習熟状況の変化を分析した。その結果、校閲カテゴリを用いることによって反省的な認識が促進され、用語集に従うことや直訳調の翻訳を避けるべきであることは効果的に習得できたこと、一方で、原文の内容を正しく理解して過不足なく適切に目標言語に移す（誤訳をしない）ことや想定されるレジスタに合わせて訳すことなどは短期間では習得できないことが明らかにされた。これらは、翻訳実務段階での OJT や問題改善においても有用な知見と言える。

翻訳者を取り巻く環境や翻訳者に求められる要件の変化にいち早く対応すべく、山田ら[9]は、大学・大学院における翻訳テクノロジーに関する教育を促進するためのオンライン教材「翻訳テクノロジー」を開発している。本発表では、機械翻訳(MT)、MT後編集(PE)、翻訳メモリ(TM)を題材として取り上げ、動画およびクイズ形式の確認問題からなるeラーニング教材を提供していることが紹介された。山田らは、実際の講義においてこの教材を用いるとともに、翻訳テクノロジーに関する学習者の意識についても調査した。そして、MTを単体で用いる際の期待値の低さに対して、欧州で標準的な翻訳フローになりつつあるPEを学ぶことにより、道具としてのMTの可能性に気づく者も多いことを明らかにした。

5. 研究発表と討論を通じて拓かれた問い、暫定的な解の案

以上、言語処理学会第22回年次大会のテーマセッション『文理・産学を越えた翻訳関連研究』の模様を、各プレイヤー間の連携と問題・可能性について整理・

言及しながら俯瞰した。翻訳産業がテクノロジーの高度化とともに高効率化を目指す上で、文理・産学の連携が今後も重要であることが再確認できた。

本論考でプレイヤー別に整理して見えたこととして、特筆すべきは、b) 翻訳産業と c) 実務者養成の連携、すなわち翻訳教育における産学の連携が、実際にはかなり少ないということである。一方、文理・産学の共同研究自体は実施されてきていること、「教育」分野においても文理が連携した活動は存在することが参加者間で共有された。

今後、例えば「深層学習によって取って代わられる業務もあるのか」という議論や課題も出てくると予想される中で、翻訳教育における産学連携は急務と思われる。

6. 今後の我々の活動予定

我々は、今回の言語処理学会テーマセッションのような企画を、今後も継続的に実施したいと考えている。まずは、2016年9月に日本通訳翻訳学会(JAITS)第17回年次大会(@同志社大学)で本活動の近況と今後の課題について発表する。2017年3月に筑波大学で開催予定の言語処理学会第23回年次大会でも2度目のテーマセッションを開催したい。2017年の秋には名古屋でMT Summitが開催されるので、ここでもワークショップなどを行って、産学の連携強化等を図りたい。

いずれにしても、状況としては海外先行、日本後発であるので、世界の最先端での新規性を求めるよりは国内の意識向上がねらいである。

謝辞

同テーマセッションの提案者である武田珂代子先生・立見みどり先生(立教大学)にも、本稿の内容をご確認いただいた。

参考文献

- [1] M. Carl, I. Lacruz, M. Yamada, and A. Aizawa. Measuring the translation process. 言語処理学会第

- 22 回年次大会, 発表論文集, pp. 957–960, 2016.
- [2] M. Carl, I. Lacruz, M. Yamada, and A. Aizawa. Comparing spoken and written translation with post-editing in the ENJA15 English → Japanese translation corpus. 言語処理学会第 22 回年次大会発表論文集, pp. 1209–1212, 2016.
- [3] 影浦峽, M. Thomas, A. Hartley, 内山将夫. みんなの翻訳実習における「足場」と翻訳力・翻訳者力～みんなの翻訳第 6 報～. 言語処理学会第 22 回年次大会発表論文集, pp. 857–860, 2016.
- [4] 河野弘毅. 機械翻訳ビジネスの成功要件あるいは MT で金持ちになる方法. 言語処理学会第 22 回年次大会発表論文集, pp. 965–968, 2016.
- [5] I. Lacruz, M. Carl, M. Yamada, and A. Aizawa. Pause metrics and machine translation utility. 言語処理学会第 22 回年次大会発表論文集, pp. 1213–1216, 2016.
- [6] 宮田玲, 藤田篤, 内山将夫, 隅田英一郎. 機械翻訳向け前編集の事例収集と類型化. 言語処理学会第 22 回年次大会発表論文集, pp. 869–872, 2016.
- [7] 坂西優. 非翻訳者向け CAT ツール「ヤラクゼン」の仕組みと課題. 言語処理学会第 22 回年次大会発表論文集, pp. 961–964, 2016.
- [8] 豊島知穂, 田辺希久子, 藤田篤, 影浦峽. 翻訳教育での利用を意識した翻訳エラー分類体系の再構築. 言語処理学会第 22 回年次大会発表論文集, pp. 861–864, 2016.
- [9] 山田優, 立見みどり, 武田珂代子. 「翻訳テクノロジーを学ぶ」: 教材オンライン化の現状と展望. 言語処理学会第 22 回年次大会発表論文集, pp. 953–956, 2016.
- [10] 山本真佑花, 田辺希久子, 藤田篤. 翻訳学習者の学習過程におけるエラーの傾向の変化. 言語処理学会第 22 回年次大会発表論文集, pp. 865–868, 2016.

TC シンポジウム 2016（東京）参加報告

山本ゆうじ（秋桜舎代表）

機械翻訳課題調査委員会ワーキンググループ3（UTX チーム）リーダー

写真 1TC シンポジウム 2016（東京）会場・工学院大学



1. はじめに

本稿では、TC（technical communication）シンポジウム 2016 に、AAMT WG3（UTX チーム）リーダー山本ゆうじ（筆者）が参加した結果を報告する。2016 年 8 月 24、25 日の両日、TC シンポジウムが東京の工学院大学で開催された（写真 1）。本会議は、テクニカルコミュニケーター協会（JTCA）が毎年開催している。今回は、AAMT から筆者が聴講者として参加した。なお口頭でのみ確認した情報もあるため、万が一聞き間違いなどがあった場合はご容赦いただきたい。

2. 発表と会議

TC シンポジウムのトピックの種類としては、マニュアルの他に翻訳、機械翻訳、DITA などが多かった。DITA は、マニュアルなどのために文書を構造化して記述するための XML ファイル形式である（参考：<<http://dita-jp.org/?cat=13>>）。シンポジウム参加者の傾向は、セッションにより異なると思われる。一例として、後述の TC カフェで挙手による確認ではメーカ

ーの人がかなり多かった。

8 月 24 日午前は「セキュリティ対策を織り込んだ翻訳業務の ICT 化に関する考察」を聴講した。株式会社情報システムエンジニアリングの若林夏樹氏による講演である。セキュリティー課題についての説明で、クラウド機械翻訳サービスを利用することによる情報流出の例などが示されていた。ユーザー側としては単に「ウェブ上の機械翻訳を使用しているだけ」という意識で、危険性にまったく気づかないまま機密情報もらしてしまう。私見であるが、数行のセンテンスだけでなく、使用規約によっては、特定企業に対する大量のテキストマイニングにより、意図しない情報収集がされてしまう可能性もある。このような情報流出は、一度発生してしまうと企業にとって致命的な問題となり、回復することは困難である。PC のセキュリティーは、技術の進歩により改善されているものの、ウイルスやマルウェアも進化している。ユーザーのセキュリティー意識の改善も重要だが、抜本的にセキュリティーを改善するにはクラウドの利便性を犠牲にするしかない。機械翻訳は必要でもクラウド機械翻訳が必須でないのであれば、ルールベース機械翻訳も選択肢に

入ると思われた。また UTX 用語集形式で用語を管理し、機密用語のフラグを設定することにより、用語レベルでのチェックを行うことができる。

24 日午後には、「テクニカルコミュニケーション分野における専門用語の統一活動—Terminology of Technical Communication—」を聴講した。ドイツを中心とするテクニカルコミュニケーション組織 tekom の Michael Fritz 博士による講演である。これは、AAMT の UTX 用語集形式標準化、また AAMT から筆者が参加している ISO/TC37 (ターミノロジー) 委員会での活動と関係しているため興味深かった。たとえば「用語集は積極的に配布すべき」という主張など、筆者が日ごろから考えていることと近い点が多かった。tekomp 用語集では、「好ましい」用語と「禁止」用語の区別も行うそうだが、これは UTX での承認語と禁止語に似ている。将来的にはウェブで tekomp 用語集が公開されるとのことである。

その後、基調講演の「テクニカルコミュニケーションのバイアス」を一部聴講した(すべてのセッションを同時聴講できないため)。ストーリーが TC にも重要という主張は興味深かった。

また、パネルディスカッション『『できるテクニカルコミュニケーター』を育てるには?』を聴講した。質疑では、筆者は、用語管理に関する考えを伺った。「TC と翻訳では、分かりやすさという観点からも用語管理は重要ではないか」という問いに対して、ヤマハ株式会社の石川秀明氏からは「日本語では人間味を出すためにあえて統一していない。英訳するときは統一する」という回答を頂いた。

8 月 25 日午前には、AAMT でもたびたび講演・寄稿いただいている株式会社ヒューマンサイエンスの「日英翻訳でも機械翻訳は使える! ~事例からわかる日英機械翻訳活用のポイント~」を聴講した。「ある日英案件では統計機械翻訳よりルールベースのほうが精度が良かった」という結果が紹介された。案件の性質、分野、分量などさまざまな要因にもよるものの、筆者の見解と基本的に一致する。

8 月 25 日午後は TC カフェに参加した。質疑を歓迎するインタラクティブな方式のカジュアルな雰囲気、マニュアル作成の現場の声が聞けて、内容としては最も興味深かった。設計側の意図がくみ取れないなど、ライターの悩みが挙げられた。このセッションは、内容をより細かく分けて、詳しく知りたいと感じた。

また、会場ではマニュアルコンテストの結果が展示されていたが、そこでの出品作品の評価ポイントも読んでいて興味深かった。一般人には違いが分かりにくいこともあるが、分かりやすい工夫がされているマニュアルに対する評価には頷けるものもあった。

3. まとめ

シンポジウム全体では、AAMT で行う MT フェアなどに参考になるかもしれない、いくつかの気づきがあった。最大で 11 ものトラックが並行しているため、関心があっても聴講できないセッションが多かった。また資料をダウンロードする仕組みがない。紙で頂いた配付資料の量が増え、重たい上に後で電子化処理が必要になる。セッションのトピックにはマニュアルの電子化がさかんに取り上げられているのに、残念なことだった。AAMT でのフェアやセミナーなどでは、並行トラックを減らし、PDF での資料配付は事前にしっかり準備しておくのがよいと思われる。

1 セッションの時間は、1 または 2 時間であった。情報を限られた紙数や時間で、確実かつ簡潔に、要領よく伝えるのがテクニカルコミュニケーションであるなら、内容によっては 1 セッションが 30 分くらいでもよいと感じた。

総合的には、TC シンポジウムに参加してさまざまな情報が得られ、極めて有意義だった。マニュアルは、翻訳の対象文書の種類としては一部でしかないかもしれない。ただ「よい説明文章とは」という問題意識は「よい翻訳とは」ということにもつながる。機械翻訳のトピックも多かったため、AAMT とのより密接な協力関係が構築できれば双方に有意義と思われる。

多言語情報発信支援の社会実装に関する研究

井佐原 均

豊橋技術科学大学

1. はじめに

スマートフォンやパソコンで、テキストを入力したり、テキストを読むという事は日常化している。機械翻訳の精度向上によって、テキスト翻訳による情報の受発信は急速に社会に普及していくと思われる。

2020年に東京オリンピック・パラリンピックが開催されることを受け、海外からの観光客の増加に向けて、観光情報をはじめとする地域の情報の多言語化は重要である。場所的・質的・量的に出来るだけ多くの情報を発信したいが、人手による翻訳はコストとスピードの点で困難であり、機械翻訳の活用が必須である。機械翻訳システムによって、観光ガイドブックに載っているような主要な観光地に関する情報だけではなく、小規模な旅館、神社、土産物店などからもリアルタイムの情報発信が可能になる。

これは観光分野に限定されるものではない。国際競争力強化にはウェブ等での情報発信の多言語化が必須であり、工業製品や地域の特産品の輸出など、波及効果は膨大である。現時点では汎用の機械翻訳システムは、特に日本語からの情報発信の観点からは、精度が十分とは言えないが、Neural Machine Translation や、より多くのデータを共有する枠組みにより、大幅な精度向上が期待できる。関連組織間のデータ共有により機械翻訳を用いた翻訳プロセスの完成度を高めることが必要である。すなわち、対象分野の大量の対データを利用することによる性能の向上を行い、効率良い多言語情報発信を実現することにより、国際競争力を向上する。

個々の組織の文書の提供（公開）が、機械翻訳の精度の向上につながり、個々の組織や地域のより大きな

ビジネスチャンスに結び付く。多くの企業は翻訳の速度や精度を競っているわけではないのだから、企業同士が翻訳という非競争領域では協力しあう（データを提供しあう）ことにより、それぞれの企業が競争領域での個々の競争力の強化に集中できる。

2. 多言語情報発信支援に向けて

国立大学法人豊橋技術科学大学では、新しい価値を創造し、地域社会の活性化に貢献する研究拠点として、平成28年4月1日に技術科学イノベーション研究機構を設立した。ここでは国内外の研究機関や企業とのマッチングファンド形式により、特定分野の最先端を切り開くとともに、研究成果の社会実装・社会提言力の強化を目指している。機構のプロジェクトの一つとして、多言語情報発信支援の社会実装に関する研究が行われている。この研究の概要を以下に示す。

既に述べたように、我が国の国際競争力強化のためには多言語による情報発信が必須である。我が国産業の市場拡大や、年々増加傾向にある海外からの来訪者へのサービス向上、そして一層の海外来訪者増のためには、多言語による様々な情報へのアクセスを可能にすることが重要となる。一方で情報通信技術（ICT）がより高度化するにつれ、情報の更新頻度は高くなっている。事前に情報を翻訳しておく静的な翻訳では現代の様々なニーズを満たすことは時間とコストの両面において現実的ではなくなっている。また、日本は頻繁に自然災害に見舞われる国でもあり、防災・災害に関する情報などにおいては、リアルタイムに多言語に翻訳し、正確な情報を提供できることが重要である。

このような課題に対する解決策の一つとして、我々は、日本マイクロソフト株式会社、株式会社ブロードバンドタワーと協力し、機械翻訳システムを中心とする自然言語処理システムによって、品質の高いリアルタイム翻訳サービスを実現し、社会実装することを目指している。三者がそれぞれのテクノロジーを持ち寄り、AI・機械学習による多言語コミュニケーションの実現に向け協働を開始した。

この内容は2016年6月21日に東京で記者説明会を開催して広報した。当日は40名近い記者の参加があり、日刊紙6紙への記事掲載や30件を超えるネットニュースへの掲載を得た。共同研究のイメージを図1に示す。

この協働では、今後、大幅な増加が見込まれる海外からの来訪者に向けて、観光情報をはじめ、滞在中に必要な医療、災害など、様々な言語による情報をリアルタイムに提供する翻訳サービスを、東京オリンピック・パラリンピックが開催される2020年までにインターネット上の様々なサービスで活用可能にすることを目指している。

三者は、実社会における様々なシーンでのAI・機械学習の活用促進を目的に、機械学習の品質向上に必要な情報の収集やビッグデータの構築を多くの組織や業種の協力を得ながら協働で推進する。得られたデータを分析することで翻訳サービス品質の向上を図る。

3. おわりに

機械翻訳は、我々の生活から言語障壁を取り除き、我々と世界をつなぐ重要な技術になりつつある。機械翻訳技術を用いて、地域や企業の情報を世界に発信することにより、地域や企業のステータスを向上し、地域や企業と世界をつなぐグローバル社会を実現することが可能であろう。

本研究開発の成果により、企業・産業においては翻訳データの共有による関連組織間のネットワークの強化と、非競争領域での協力による、競争領域でのグローバル化が実現できる。

多言語化された文書を介したコミュニケーションによって、世界の人々との距離を近づけることが可能となる。また本稿では触れなかったが、世界中のボランティアによって後編集作業を行うなど、多言語化するプロセスによって、世界の人々の距離を近づけることが出来よう。情報技術による多言語コミュニケーションが世界を繋ぎ、その影響は社会・文化に広がる。科学技術が牽引する新たな社会システムの実現が可能となろう。

機械翻訳システムが社会・文化のイノベーションに貢献することを願っている。

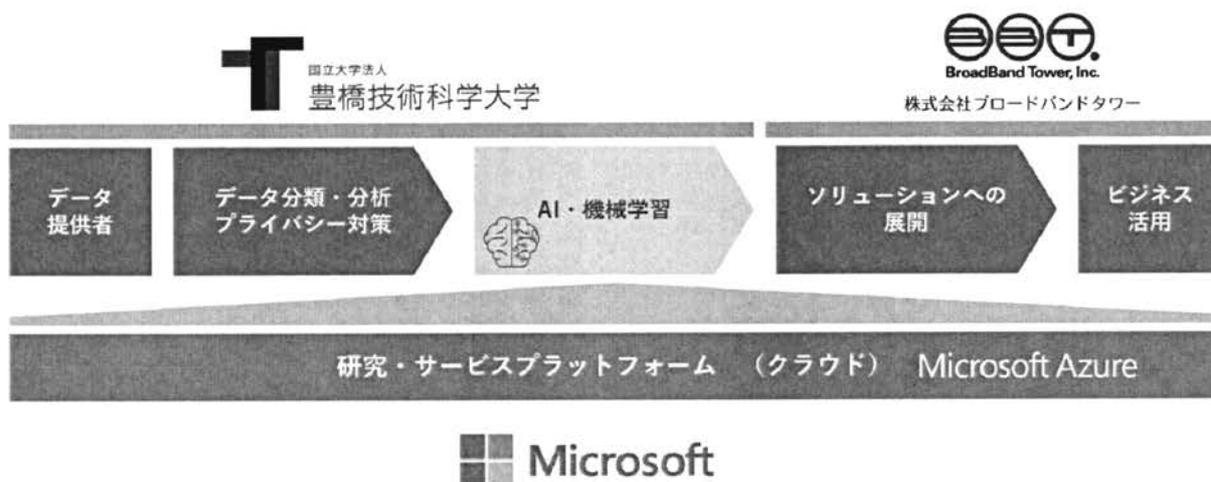


図1 3者共同プロジェクトのイメージ

「第26回 JTF 翻訳祭」のお知らせ

一般社団法人日本翻訳連盟

JTF 理事・第26回翻訳祭企画実行委員長 古谷 祐一

<第26回 JTF 翻訳祭開催のご案内>

☆☆☆ 第26回 JTF 翻訳祭 ☆☆☆

【日時】

2016年11月29日(火) 9:30~21:00 (開場 9:00)

日本翻訳連盟では、来る11月29日(火)、東京のアルカディア市ヶ谷(私学会館)を会場として、恒例の「JTF 翻訳祭」を開催いたします。AAMT様はじめ各関連団体より後援をいただいております。

今年掲げるテーマは、「考えよう、翻訳のこと～ともに歩む翻訳の未来へ～」です。分科会形式による「講演・パネルディスカッション」及び「プレゼン・製品説明コーナー」、「翻訳プラザ(展示会)」、「交流パーティー」などの催し物を用意しております。交流パーティーは参加者同士の交流促進や情報交換の場としてご活用いただければ幸いです。

翻訳祭は翻訳者、翻訳会社、クライアント、翻訳支援ツールメーカー、翻訳スクール等の業界関係者が一堂に会するイベントです。翻訳祭を通して、参加者の皆様に新たな人脈形成やビジネスの機会が生まれることを願っております。

翻訳祭にぜひともご来場くださいますよう、よろしくお申し込み申し上げます。

【場所】

「アルカディア市ヶ谷(私学会館)」東京
※JR・地下鉄「市ヶ谷駅」より徒歩2分

【主催・運営】

一般社団法人日本翻訳連盟
第26回 JTF 翻訳祭企画実行委員会

【後援】

経済産業省・アジア太平洋機械翻訳協会(AAMT) 他

【概要】

JTF翻訳祭は今年で第26回を迎えます。今年からJTFの定款に「通訳」の2文字が刻まれ、翻訳・通訳を含めた業界団体として更なる拡大に期待が集まっています。同時に2020年の東京オリンピックの開催や訪日外国人増加による翻訳・通訳需要の拡大など、業界を取り巻く環境が次第に変化していく中、業界関係者の連携をより一層深めることが必要とされています。

JTF誕生から四半世紀が経過した今だからこそ、「翻訳のこと」を再考し、ともに歩みながら輝かしい翻訳の未来を引き寄せることが重要だと感じ、「考えよう、翻訳のこと～ともに歩む翻訳の未来へ～」を今年のテーマに掲げました。

さて、「翻訳の原点」とはどこにあるのでしょうか？人それぞれ考え方は異なりますが、私は「翻訳者」にありたいと思います。翻訳会社、翻訳支援ツールメーカー、ソースクライアントのいずれでも、また統計的機械翻訳の学習プロセスでも、「翻訳者」の存在があってこそビジネスが成り立ち、技術は革新します。このような考えから、今年の企画実行委員は私を除いて全員「翻訳者」で構成されています。「翻訳者」の視点で業界関係者のバランスを配慮しながら、今までにない新しい翻訳祭を目指し、良質なコンテンツを集めるための議論を重ねてきました。

用意したのは24の講演&パネルディスカッション。翻訳プラザでは、製品のプレゼンコーナーに加え、海外を含む約40社が展示ブースで自社の製品

を紹介し、見たことのないツールや新機能を見ることができるかもしれません。

そして交流パーティー。300名を超える業界関係者が集い、情報交換やビジネスマッチングの機会として活用されています。

なお、今回の翻訳祭では、AAMTの会長である中岩氏がモデレーターを務め、同団体メンバーである長瀬氏、須藤氏、河野氏が参加されるパネルディスカッションが催されるほか、翻訳プラザにてAAMT様にご出展いただきます。

翻訳祭企画実行委員会全員がこだわって準備しているJTF翻訳祭。今年の翻訳祭は何かが違うと感じていただけたら幸いです。

【総来場者数（見込み）】

900名（前年度実績 890名）

【プログラム】

- ・開場／受付開始 9:00～
- ・トラック 1～6 [6会場]9:30～18:00
全 24セッション
- ・プレゼン・製品説明コーナー
9:30～17:00 全 6セッション
- ・翻訳プラザ（展示会）
9:00～17:30
- ・交流パーティー
18:30～21:00（150分）

【お申し込み】

下記サイトからお申し込みください。

<http://www.jtf.jp/>

●翻訳プラザ（展示会）

出展 40社（2016年9月現在）

翻訳プラザは、翻訳業界の企業が参加する展示会です。来場者と出展企業が交流し、情報交換する場です。AAMT様も出展されます。どうぞ3階の展示

会会場にお立ち寄りください。

<展示内容>

- ・翻訳支援ツールメーカーの製品デモ
- ・翻訳会社との翻訳相談（翻訳発注・翻訳者登録等）
- ・翻訳学校の翻訳講座紹介
- ・出版社の翻訳書籍販売

●プレゼン・製品説明コーナー

今年も展示会に隣接して「プレゼン・製品説明コーナー」を併設します。プロジェクター・スクリーンを備えた特設会場となっています。

- ・1セッション 60分／6セッション
- ・プレゼンコーナーを増設し、85名収容可能！

※翻訳プラザ、プレゼン・製品説明コーナーへの参加は無料です（事前登録不要）。

●交流パーティー

情報交換を行い、これからの業界の展望などを語り合う機会となります。昨年に続いて「ほんやく検定1級合格者の表彰式」を行います。お仕事のための人脈作りや最新の業界の動向をつかむためにぜひご参加ください。

<ボランティアスタッフ募集中！>

募集内容は、「受付」「会場司会」「会場アシスタント」「書記」「写真」です。ボランティアスタッフにご協力いただく方の「講演・パネルディスカッション」参加料は無料です（パーティーは有料）。詳細はJTFウェブサイトにて随時掲載いたします。

【お問い合わせ】

一般社団法人 日本翻訳連盟 事務局

<http://www.jtf.jp>E-mail:info@jtf.jp

産業翻訳分野における日英機械翻訳の実用化に向けて

徳田 愛

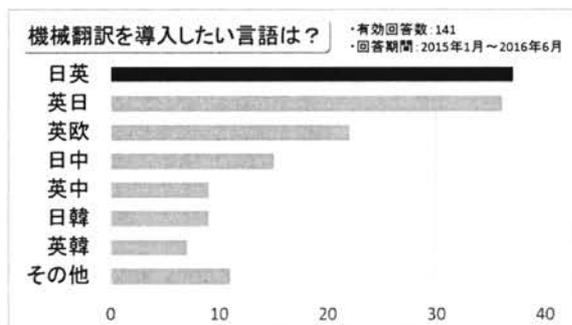
株式会社ヒューマンサイエンス ドキュメントソリューション部

1. はじめに

株式会社ヒューマンサイエンスでは、産業翻訳分野の機械翻訳導入コンサルティングを実施している。国内でも英語からヨーロッパ言語、もしくは英語から日本語といった言語ペアを中心に、機械翻訳を導入する企業が増えてきている。一方、日英機械翻訳については、日英の言語構造の違いが原因で、機械翻訳の精度が十分なレベルには達しておらず、実用化が遅れているという状況があった。しかし、日英機械翻訳についても、

- ・日本語と英語のマニュアルを同時リリースするために、翻訳期間を短縮したい
- ・大量にある社内の仕様書やサポート文書を効率的に翻訳したい
- ・社内のドキュメントをとりあえず英語に翻訳したいといった理由から、ここ数年日英機械翻訳を導入したいというニーズが高まっている。

弊社で2015年1月から2016年6月までに実施した機械翻訳セミナーでも、「機械翻訳を導入したい言語は？」という質問に対して、「日英翻訳」という回答が141人中37人で一番多いという結果になった。



こうした「日英機械翻訳」へのニーズの高まりを受け、弊社では産業翻訳分野での日英機械翻訳の実用化に向

けた品質向上やワークフロー構築に取り組んできた。その結果、日英機械翻訳であっても、条件やワークフロー次第でコスト削減や翻訳期間の短縮を実現することが十分に可能であることがわかっている。

そこで、今回は産業分野のドキュメントを使用した日英機械翻訳の検証事例を紹介し、日英機械翻訳実用化に向けての課題を考察する。

2. 日英機械翻訳検証事例

検証を実施したのは、製造業分野の2つのプロジェクトである。

●2つのプロジェクトの概要

	プロジェクト①	プロジェクト②
分野・対象製品	製造業・精密機器	製造業・FA機器
対象	技術文書	HTMLマニュアル
ボリューム	約100万文字/年間	約40万文字/年間
コーパス	約90000センテンス (約300万文字)	約38000センテンス (約130万文字)

プロジェクト①は、設計者や開発者が書いた精密機器の申請書や仕様書などの技術文書が翻訳対象である。プロジェクト②は、FA機器のHTMLマニュアルが翻訳対象である。

統計ベースのエンジンに読み込ませることのできるコーパスに関しては、プロジェクト①はプロジェクト②に対して、2倍以上の分量があった。一般的に統計ベースのエンジンに読み込ませるコーパスは、分量が多ければ多いほど、エンジンの精度が高くなると言われている。そのため、コーパスの分量が多いプロジェ

クト①の方がエンジンのトレーニングという点では有利と考えられる。

●使用したエンジン

この2つのプロジェクトのデータを用いて、以下の3つの機械翻訳エンジンを使用して、品質の比較検証を実施した。

エンジン A :

国内企業が販売しているルールベースエンジン

エンジン B :

海外企業が販売している統計ベースエンジン

エンジン C :

オープンソースの統計ベースエンジン

●品質評価方法

2つのプロジェクトの検証結果については、製造業の機械分野の翻訳を専門にしている日英翻訳者3名により、対象ドキュメントをランダムに抽出した50文に対して、一文ずつ以下の四段階での評価を実施した。分類の定義は次の通りである。

Excellent	ポストエディット不要 文法、構文が正しく、訳文を読めば、原文の意味が正確に理解できる
Good	部分的にポストエディットが必要 文法、構文に多少のエラーは含まれるが、原文の意味が正確に理解できる
Medium	ポストエディットが必要 文法、構文に多少のエラーがあり、原文の意味が正確に理解できない
Poor	再翻訳が必要 文法、構文に重大なエラーがあり、原文の意味が理解できない

3. 検証結果

2つのプロジェクトでの検証結果は次のグラフのよ

うになった。

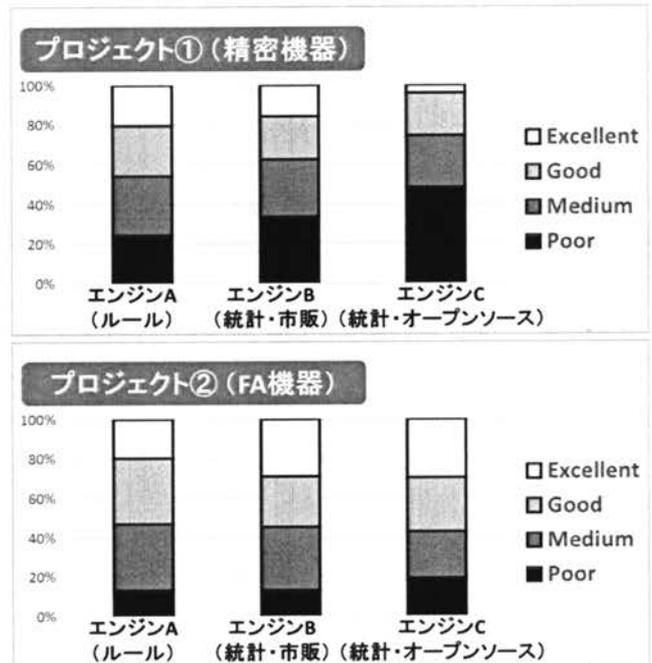


図2 各プロジェクトでの品質評価結果

プロジェクト①は、全エンジン平均で「Excellent」と「Good」の占める割合が全体の36%、プロジェクト②では55%という結果になった。プロジェクト②の品質であれば、目標品質は「人と同じレベル」に維持しながらも翻訳コストおよび翻訳期間は34%削減できる試算1となり、機械翻訳後にポストエディット(訳文修正)を実施すれば、十分に実用的に使用可能なレベルと言える。一方、プロジェクト①の品質では大部分の訳文を大幅に修正する必要があるため、実用的に使用するの難しいレベルである。

4. 検証結果の考察

プロジェクト①と②では、対象ドキュメントの種類やコーパスの分量という条件が異なっていた。一見すると、日英機械翻訳の導入は難しいと言われることの多い「マニュアル」が対象であり、コーパスの分量が少ないプロジェクト②の方が、品質が低くなるのではないかと考えられるが、結果はプロジェクト②の方が

1 試算結果については、品質評価結果より、各品質レベルで訳文修正にかかる負荷に基づいて、弊社独自の計算式を使用して算出。

品質が良い結果となった。

この結果から、単純に対象ドキュメントの種類やコーパスの分量だけが、日英機械翻訳の品質を左右するわけではないことがわかる。

そこで、日英機械翻訳の品質の違いを生む要因を探るために、プロジェクト①とプロジェクト②について、「対象ドキュメント」と「コーパス」という2つの観点で比較を行った。その結果、2つのプロジェクトでの品質の違いを生んでいた大きな要因として、

- ・対象ドキュメントの日本語の書き方
- ・コーパスの品質と特化性

という2つのポイントがあることがわかった。この2つのポイントについて、詳述する。

5. 日本語の書き方の違い

プロジェクト①とプロジェクト②では、対象ドキュメントの「日本語の書き方」に違いが見られた。プロジェクト①では、プロジェクト②に比べて一文が長く、意味が曖昧な文章が多い傾向があった。

●一文の長さ

プロジェクト①とプロジェクト②の一文の長さを比較したところ、プロジェクト①の方が長い文が多い傾向にあった。以下に示すのは、各プロジェクトの対象ドキュメントの文字数の平均値と中央値、そして、各文の文字数ごとの割合を示したグラフである。

	プロジェクト①	プロジェクト②
平均値 (文字)	32	28
中央値 (文字)	28	25

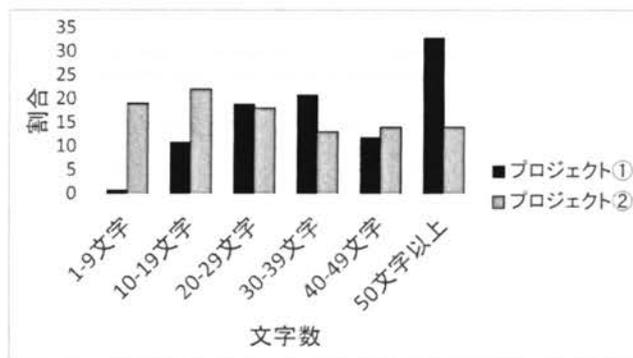


図3 原文一文ごとの文字数の分布

2つのプロジェクトの文字数の平均値、中央値には大きな差はないが、プロジェクト①では50文字以上の文が占める割合が極端に多くなっているのがわかる。一方、プロジェクト②では10文字から19文字をピークに、文字数が長くなるにつれ、全体に占める割合が減っている。日英機械翻訳の場合、文法構造の難しさから、一文が長くなればなるほど主語や目的語などを正確に認識するのが難しくなり、特に50文字以上になると正確に翻訳するのは極めて難しいとされている。実際に、50文字以上あった文のうち、87%はPoorもしくはMediumの品質であった。以下に、実際にプロジェクト①でPoorと評価された文を例として紹介する。

原文
また、本製品には、本体をコントロールするアプリケーションと、先述のプログラムの起動、及びプログラムのアップデート機能をコントロールするアプリケーションで構成されている。
訳文 (機械翻訳結果) ※Poor レベル
Moreover, an application which controls a main body, and an application that controls starting of the previously-mentioned application program and an upgrade function are consisted of by this product.

日本語が長く複雑であるために主語が正しく認識さ

れず、主語となるべき「本製品」が英文では末尾に置かれて主語として認識されていない。他にも文法エラーなどがあり、一から訳し直しが必要な Poor レベルの品質である。

この例文と同様に、プロジェクト①では「長い文が多い」ことが、品質が低下した要因の一つとなっていた。

●意味の明確さ

プロジェクト①とプロジェクト②では、意味の明確さという点でも違いが見られた。

プロジェクト①の日本語の書き方を分析したところ、全体の 32%の文章に以下のような特徴が見られることがわかった。

- ・主語・目的語が不明確
- ・係り受けが不明確
- ・具体的な動詞が使用されていない

一方、プロジェクト②では、上記のような問題のある文は全体の 2%であった。プロジェクト①では、日本語で主語や目的語が不明確であったり、係り受けが不明確であったりする文が多かったために、機械翻訳後の訳文で誤訳になっている傾向があった。実際に、プロジェクト①で意味が曖昧な文章のうち、93%は Poor もしくは Medium の品質であった。

以下に、実際にプロジェクト①で Poor と評価された文を例として紹介する。

原文
本プログラムは、ABC 操作に主に解析、充電電源のコントロールの制御をおこなう。
訳文（機械翻訳結果） ※Poor レベル
This program mainly controls control of an analyzation and a charge power supply to ABC operation.

「ABC 操作に」の「に」という助詞の用法が曖昧であり、複数の意味に解釈できる書き方になっている。本来は「ABC 操作において」という意味で訳されるべきであるが、機械翻訳結果では「ABC 操作に対して」という意味で訳されてしまっている。

また、動詞も「コントロールの制御をおこなう」という不自然で具体的なでない表現が使われており、機械翻訳結果でも「controls control」という不自然な表現が使用されている。

この例文のように、プロジェクト②では、日本語の書き方が曖昧な文が多かったことが、品質が低下した要因の一つとなっていた。

●日本語改善による効果

「一文が長い」「意味が曖昧」という問題があるプロジェクト①の原文を改善すると、訳文の品質が向上するのかを検証するため、プロジェクト①の原文を以下の観点で書き換えた。

- ・長い文は意味の切れ目に応じて区切る
- ・曖昧な助詞が使われている箇所を修正する
- ・修飾関係が不明確な箇所を修正する
- ・曖昧な動詞を具体的に修正する

以下は、プロジェクト①で問題があった日本語について、一文を短く意味が明確になるように書き換え、機械翻訳結果の品質を評価した結果である。

原文（改善前）
また、本製品には、本体をコントロールするアプリケーションと、先述のプログラムの起動、及びプログラムのアップデート機能をコントロールするアプリケーションで構成されている。
原文（改善後）
また、本製品は、以下で構成されている。
・本体をコントロールするアプリケーション
・先述のプログラムの起動およびアップデート機能

をコントロールするアプリケーション
訳文（改善後・機械翻訳結果） ※Good レベル
Moreover, this product consists of below. ・ Application which controls a main body ・ Application which controls starting of a previously-mentioned program, and an upgrade function

長く複雑だった文を、箇条書きを使用し3文に分けることで、各文の構文が正しく認識しやすくなり、訳文の品質が向上した。日本語の改善前は Poor の品質だったのに対し、改善後は Good の品質にまで向上している。

次のグラフは、プロジェクト①で最初に評価した100セグメントを対象に日本語の書き方を改善し、改善前と改善後の機械翻訳の品質を比較した結果である。

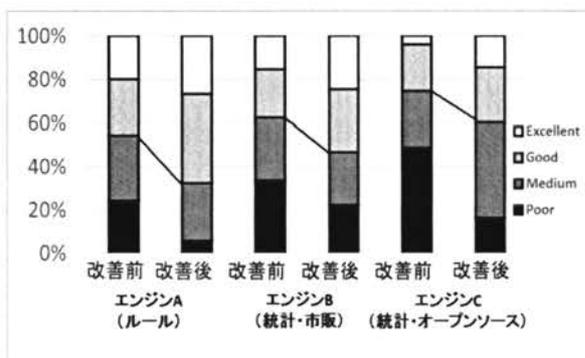


図4 プロジェクト①

日本語書き換え後の品質向上結果

いずれのエンジンでも、日本語を改善した後のほうが Excellent と Good の文が占める割合が増えているのがわかる。特にルールベースのエンジンでは Excellent と Good の文が全体の7割を占めるようになり、大きな品質改善が見られた。

これらの結果から、日本語の書き方が不適切であったという点が、プロジェクト①で品質低下を招いた大きな要因であったことがわかる。

●「日本語の書き方」の考察結果

日本語は語順を入れ替えたり、目的語や主語を省略したりすることができるため、文法構造が厳密に定まっている英語と比べて、機械翻訳が文の構造を正確に認識するのが難しい。

そのため、一文が短く、意味が明確な文になっているかどうかを機械翻訳の品質を向上させるために重要である。

英語からヨーロッパ言語への翻訳の場合でも原文の品質は重要であるが、英語の文法構造は厳密に定まっているため機械翻訳エンジンが文の構造を認識しやすく、また英語とヨーロッパ言語の文法構造が近いいため、原文の品質が機械翻訳の品質に与える影響は日本語に比べると小さい。そのため、例えばライティングのトレーニングを受けていない開発者や設計者が執筆した技術資料などのドキュメントと、プロのライターが執筆した取扱説明書などのドキュメントで、機械翻訳の品質の差は大きく出ないことが多い。

一方、日本語では原文の品質が機械翻訳に与える影響が多いため、わかりにくく冗長な文章が多い傾向にある技術資料などのドキュメントよりも、テクニカルライティング視点で書かれている取扱説明書やマニュアルの方が機械翻訳で良い品質が期待できると言える。

6. コーパスの品質と特化性

次に、統計ベースのエンジンに読み込ませたコーパスを比較した。プロジェクト①とプロジェクト②で違いがあったのは、コーパスの「品質」と「特化性」である。

●コーパスの品質

各プロジェクトのコーパスの品質については、以下の3つの観点で分析を行った

- ・品質に問題がないか（誤訳、訳漏れがないか）
- ・一文の長さに違いはないか
- ・訳し方に違いはないか

コーパスに誤訳や訳漏れなどが含まれていると、機械翻訳の品質も低くなるが、どちらのプロジェクトも、コーパスに誤訳、訳漏れなどのエラーはほとんど確認できなかった。また、「一文の長さ」についても、違いが見られなかった。しかし、「訳し方」という点で2つのプロジェクトでは違いが見られた。

●コーパス内の文章の「訳し方」

コーパス内の文章の「訳し方」について比較をしたところ、プロジェクト①では意識された訳文が多いのに対し、プロジェクト②では逐語的に訳された訳文が多いという傾向があった。

次に示すのは、プロジェクト①のコーパスに登録されていた原文・訳文の例である。

原文
ポイントされている全てのグラフにおいて先述の線が平行になるように処理が実行される。
訳文
The previously-mentioned lines on all the graphs are parallel.

こちらの例文では、日本語にある要素を一言一句訳出するのではなく、日本語をわかりやすく言い換えた上で翻訳されている。実際、この訳文を日本語に戻すと「全てのグラフにおいて先述の線が平行になる。」という意味となり、「ポイントされている」や「処理が実行される」という要素が訳出されていないのがわかる。

統計ベースのエンジンでは、日本語と英語のそれぞれの単語の品詞を解析して、日本語と英語で対応する品詞同士をひも付けながら対訳を学習するが、このようにコーパス内の文章が意識されていると、日本語と英語で対応しない単語が出てきてしまうため、正確な対訳が学習しにくくなってしまいます。プロジェクト①では意識が多かったためにエンジンが正確な対訳を学習することが阻害されてしまい、機械翻訳の精度が低下していたことが一つの原因として考えられる。

このように、プロジェクト①とプロジェクト②では、コーパス内の文章の「訳し方」に違いがあった。プロジェクト①では、意識されている文章が多く、正確な対訳を学習するのに適していなかったことが、エンジンの精度が低くなっていた原因として考えられる。

●特化性

コーパスについて、品質に加えて違いが見られたのが、「特化性」である。コーパスの「特化性」とは、コーパスの内容が対象ドキュメントの内容や分野にどれだけ特化したものになっているかということである。コーパスの特化性が高いほど、対象ドキュメントに適した用語や表現、文体で翻訳できるため、機械翻訳の精度が向上すると言われている。

今回のプロジェクトで分析したところ、プロジェクト①のコーパスはプロジェクト②に比べて、特化性が低いものであった。

	プロジェクト① (約 90000 センテンス)	プロジェクト② (約 38000 センテンス)
製品群	19 製品	1 製品
種類	10 種類	1 種類

図6 各プロジェクトのコーパスの内容

プロジェクト①のコーパスには、一つの製品だけではなく、19もの製品群のドキュメントの翻訳内容が含まれていた。また、ドキュメントの種類についても、技術資料やサポート文書、仕様書、評価書などの様々な種類のドキュメントの内容が含まれていた。そのため、コーパス内で複数の用語や表現、文体が混在していて、対象ドキュメントとの類似性が低かった。その結果、適切な表現や用語を使用して翻訳されていないという問題があった。

一方、プロジェクト②のコーパスでは、対象ドキュメントと同じ製品群で、ドキュメントの種類も同じHTMLマニュアルだった。そのため、コーパスの中の利用用語や表現、文体が統一されており、対象ドキュメン

トとの類似性が高かった。その結果、対象ドキュメントに適した表現や用語を使用して翻訳されていた。

プロジェクト①ではコーパスの分量は多いものの特化性が低く、対象ドキュメントとの類似性が低かったために、機械翻訳の精度が低下していたと考えられる。

●「コーパスの品質と特化性」の考察結果

今回の結果から、エンジンの精度を上げるためには、コーパスの分量だけでなく、コーパスの訳し方や特化性も重要であることがわかる。英語とヨーロッパ言語の場合は言語構造に近いが、日本語と英語では言語構造が大きく異なるために、コーパス内の文章の訳し方が統計ベースエンジンの品質に大きく関わってくる。単純に誤訳や訳漏れがないかといった翻訳品質だけではなく、コーパスが対訳を学習するのに十分な品質になっているかということが重要である。

また、特化性が高く、かつ対象文書との類似性が高いコーパスを使用した方が、適切な用語や表現で翻訳ができるため、機械翻訳の精度を高めることができる。将来的に機械翻訳導入を考えるのであれば、現時点から製品群やドキュメントの種類ごとにコーパスを管理しておくことで、特化性の高いコーパスを蓄積していくことができる。

7. 検証結果のまとめ

以上の検証結果から、日英機械翻訳においては、「対象ドキュメントの日本語の書き方」と「コーパスの品質および特化性」が品質を左右する重要なポイントであることがわかる。日本語は文法構造が複雑なため、機械が文法構造や文章の意味を理解しやすい原文になっていることや、対訳を学習しやすいコーパスになっていることが機械翻訳結果で良い品質を得るためには重要である。

8. 日英機械翻訳の実用化に向けて

「対象ドキュメントの日本語の書き方」と「コーパスの品質および特化性」が、日英機械翻訳においては

特に重要になる。しかし、産業分野のドキュメントでは、機械翻訳を視野に入れたプロセスが構築されていないため、日本語のドキュメントの品質が低かったり、コーパスがメンテナンスされていなかったりして、機械翻訳導入のボトルネックとなっているケースも多い。そのため、産業翻訳分野で日英機械翻訳を活用していくためには、単純にエンジンの精度の向上だけでなく、原文のドキュメントの改善や品質の良いコーパスの蓄積といったドキュメント作成プロセス全体の改善が課題となると考えられる。

また、産業翻訳ではマニュアルや技術資料といった定型文が多く、ボリュームがあるドキュメントを効率的に翻訳するために、翻訳支援ツールを使用して過去の翻訳を流用しながら翻訳することが多い。一方、現状、国内で市販されている日英の機械翻訳エンジンは、Web サイトや社内文書を翻訳することを想定に開発されたものが多く、翻訳支援ツールと連携できる日英機械翻訳エンジンの数が限られている。今後、翻訳支援ツールと連携が可能な機械翻訳エンジンが増えることが、産業翻訳分野の日英機械翻訳導入を促進させるためには不可欠である。

こうした課題を解決することで、産業翻訳分野での日英機械翻訳の実用化をより加速させることができるだろう。

弊社では、今後も機械翻訳エンジンメーカーや研究機関とも協同しながら、産業翻訳分野における実務に即した日英機械翻訳のプロセス構築に取り組み、国内のさらなる機械翻訳の普及に貢献していきたい。

徳田 愛 (とくだ めぐみ)

連絡先: m-tokuda@science.co.jp

プロフィール:

機械翻訳コンサルタントとして、国内企業に向けた機械翻訳の導入やプロセス構築のコンサルティングを行う。機械翻訳に適した原文作成のコンサルティングも実施。

FUJITSU Business Application

Operational Data Management & Analytics 自動翻訳

氏名 上栴 稔

富士通株式会社

1. はじめに

海外企業との協業、海外への工場進出、海外ベンダーからの調達など、近年日本企業では、海外とのコミュニケーションの機会が増えています。また、TPP（環太平洋戦略的経済連携協定）や東京オリンピック・パラリンピックが目前に迫り、ますますグローバルコミュニケーションの必要性が高まると考えられています。

「FUJITSU Business Application Operational Data Management & Analytics 自動翻訳」（以下、ODMA 自動翻訳）はグローバルコミュニケーションを支援する高いセキュリティを担保しながら利用できる高機能な翻訳ソリューションです。

2. 課題の解決

■日本語特有の文法体系や業界特有の専門用語を、よりよい品質で翻訳したい

グローバル展開を狙う日本企業にとって大きな障害になっているのが「言葉の壁」です。日本語は、漢字やひらがな、カタカナなど多様な文字種を持ち、外国語とは文法体系も大きく異なります。

また、業界ごとに特有の専門用語があり、同じ言葉であってもまったく意味が異なるケースも珍しくはありません。例えば、「ドライバー」は運送業界では「運転手」、製造業界では「ネジ回し」、IT 業界では「ハードウェアを制御するソフトウェア」、ゴルフ業界では「クラブの一種」となります。このように日本語への翻訳、または日本語からの翻訳は、対応する言葉に置き換えるだけでは意味の通じない文章になることが珍

しくありません。

「ODMA 自動翻訳」は構文と単語の意味を把握しながら翻訳する「意味処理方式」や、日本語と英語の言葉の違いに着目し、主語・目的語を入れ替えたり、態や時勢を替え自然な訳文を出す「概念変形」を採用しています。

■セキュアな環境で安全に翻訳サービスを使いたい

企業における情報セキュリティの重要性はますます高まっており、機密情報や個人情報を外部に漏洩させないよう、さまざまな対策が取られています。

一方、インターネット上の翻訳サービスを使ったことから機密情報が漏洩する事故も発生しています。これは、自動翻訳サイトに入力した文章を第三者が閲覧できる状態になっていたため起こった問題でした。そのため、企業では情報漏えい対策として、無料の Web 翻訳サービスの利用を禁止する傾向が強くなってきました。

■海外展開に向けてグローバルコミュニケーションを活性化させたい

海外企業との協業を成功させるには、グローバルコミュニケーションの活性化が必要です。言葉によるコミュニケーションが円滑に進めば、ビジネスのグローバル展開を成功に導くことができるでしょう。最近では電子メールや Office ファイルだけでなく、チャット（インスタントメッセージ）も、ビジネス現場で活用されるようになりました。チャットはリアルタイムにメッセージを送り合うサービスであるため、スピーディーな翻訳が必要になります。

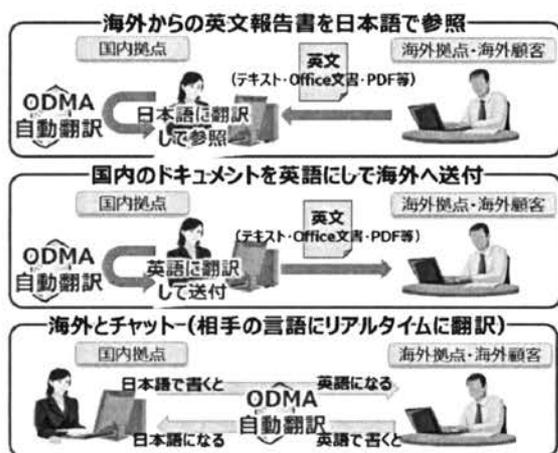
ビジネスで活用される、さまざまなコミュニケーションツールやファイルを自動で翻訳する「ODMA 自動翻訳」なら、セキュアな環境のもと、正確でわかりやすい文章に翻訳し、グローバルコミュニケーションの円滑な推進を支援します。

3. ODMA 自動翻訳とは

■Web ブラウザ、Microsoft Office や PDF ファイル、チャット等で利用できる自動翻訳ツール

「ODMA 自動翻訳」は、ネットワークで使用する日本語/英語の自動翻訳ソフトです。社内に自動翻訳サーバを設置し、Web ブラウザでアクセスして翻訳機能を利用できます。

クライアント PC にアドインツールを導入することで、Microsoft Office や Adobe Acrobat などのアプリケーション上からの利用も可能です。さらに、リアルタイムにメッセージを送り合うチャット (Lync / Skype for Business) での利用にも対応しています。



■大量の文書でも、統一した訳語ルールで翻訳が可能
サーバに収録された共通の辞書を使って全社で利用できるため、社内文書を共通の訳語ルール、表記ルールに基づいて翻訳することが可能になります。部署の違いや文書の量が原因で、訳語の不統一が起こりやすくなりますが、「ODMA 自動翻訳」を用いれば会社全体で翻訳の品質を統一できます。

4. 主な特長

■高品質翻訳：30年に渡る日本語翻訳技術研究の集大成

日本語は、世界的にも特異な文法体系を持っていると言われており、「ODMA 自動翻訳」は、富士通の30年に渡る日本語翻訳の研究・開発で培われた技術が、存分に投入されています。

日本語の文法体系に基づき、字句や構文を分析して翻訳していますので、「日本語→英語」、「英語→日本語」の双方において、正確でわかりやすい翻訳文の生成が可能です。その精度は、第三者評価においてもトップの成績を収めています。

■翻訳最適化：業種別の専門用語や企業特有の言い回しにも対応できるカスタマイズ辞書

「ODMA 自動翻訳」は、25分野 500万語以上の専門辞書に対応し、その業界に特有の用語も正しく翻訳できます。これは、日本語翻訳ソフトではトップレベルの単語数となっています。さらに、カスタマイズ辞書を作成する辞書登録サービスにより、お客様の会社特有の言い回しや用語も適切な英文に、また英文も各社特有の言葉を用いた日本語に翻訳できます。

■セキュリティ：社内のオンプレミス環境で情報漏えいを防止。翻訳ログも自動で削除

「ODMA 自動翻訳」は、翻訳機能を提供するサーバを社内に設置して利用します。外部に情報を持ち出さないので、情報漏えいを心配することなく、安全に自動翻訳を活用できます。

翻訳サーバには、「翻訳原文」「翻訳結果」のログを残さない「自動削除」機能も搭載しています。ログからの情報漏えいを防止できます。逆に、「社内でどのような翻訳がされたか」を確認するために、ログを残す設定も可能です。また、ユーザー認証機能により利用者の制限も可能です。

■グローバルコミュニケーション支援：Office 文書や PDF、チャットにも対応

アドインツールを導入することにより、ビジネスで

一般的に使われている Microsoft Office の文書や PDF ファイルを、アプリケーション上からボタンひとつで翻訳できます。特に PDF ファイルでは翻訳の際、文字量が増減してもレイアウトが崩れにくいような工夫も行っています。さらに、インスタントメッセージツールの Lync / Skype for Business にも対応。日本語・英語間でインスタントメッセージを、リアルタイムに翻訳できます。

■社内実践による実績：富士通社内で3年以上利用され、活発なグローバルコミュニケーションを実現

「ODMA 自動翻訳」は、富士通の社内実践によって蓄積されたノウハウが集結され、その信頼性が証明されています。全世界の従業員の35%強にあたる6万人以上が日々活用し、年間400万回近い翻訳を処理しています。その間、「ODMA 自動翻訳」のシステムは一度もダウンすることなく、ノンストップで稼働してきました。こうした経験、ノウハウを踏まえ、企業での活用を促進する「定着支援サービス」も提供しています。

5. あとがき

・「Operational Data Management & Analytics」は富士通のビッグデータソリューション製品シリーズです。

・Microsoft、Windows、Office、Lync、Skype for Business は、米国 Microsoft Corporation の米国およびその他の国における登録商標、または商標です。

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

・その他各種製品名は、各社の製品名称、商標または登録商標です。

・本サイトに記載されているシステム名、製品名等には、必ずしも商標表示（®）を付記していません。

・本内容は予告なく変更する場合があります。

これまでの AAMT Forum メールマガジン Vol.9

機械翻訳課題調査委員会 WG1、WG2

はじめに

AAMT Forum を通じて配信したメールマガジンのバックナンバーをお届けします。今回は第 75 号から第 83 号です。

今回は、様々な翻訳サービスの開始や新製品の開発に関する記事に加え、翻訳サービスの実証実験に関する記事が三つありました。

まず、すでに行われた実験としてタクシー乗車という場面での実証実験についての記事です。実験を通じて機械翻訳の導入に好意的な意見が多く得られたようです。同時に課題も多く確認できたようで、追実験が 12 月まで行われるとのことでした。

残り二つがこれから行なわれる実験についての記事です。医療、接客の場面での実験であり、その結果に期待を寄せたいと思います。

尚、メールマガジンとして配信すべき情報がございましたら、是非お知らせください。また、お気付きの点なども AAMT 事務局までお寄せください。よろしく願いいたします。

では、メールマガジン第 75 号から第 83 号を再びお届けします。

AAMT Forum メールマガジン

2016/06/03 配信 vol.075

こんにちは。AAMT Forum メールマガ担当です。初夏の候、いかがお過ごしでしょうか。過ごしやすい日々が続いていましたが、今年も梅雨が近づいてきました。湿気や大雨への対策は済ませましたでしょうか。では、メールマガジン第 75 号をお届けします。

■Facebook では、毎月 8 億人のユーザーが自動翻訳を利用している

現在 Facebook は、毎日 200 億件のテキストを翻訳しており、毎月 8 億人、即ち全 Facebook ユーザーの約半分が、翻訳された記事を読んでいるそうです。

<http://jp.techcrunch.com/2016/05/24/20160523facebook-translation/>

■定評ある英・日に加え、ヨーロッパ 5 カ国語、ロシア語、中国語、韓国語を手軽に翻訳！

ログヴィスタ株式会社は、Windows 用 10 カ国語翻訳ソフトの最新版「コリャ英和！一発翻訳 2017 for Win マルチリンガル」を、2016 年 6 月 24 日（金）より発売します。

http://www.logovista.co.jp/LVERP/information/news/2016-0631-kry2017win_m.html

AAMT Forum メールマガジン

2016/06/15 配信 vol.076

こんにちは。AAMT Forum メールマガ担当です。北は東北まで梅雨入りし、雨の日も多くなってきました。今年の梅雨は蒸し暑い日が続くそうですが、湿気対策は万全でしょうか。では、メールマガジン第 76 号をお届けします。

■NTT ドコモが企業向け接客翻訳サービスの提供を開始

株式会社 NTT ドコモは法人企業を対象とした翻訳サービス「はなして翻訳 for Biz」の提供を開始しました。

<http://inboundnavi.jp/docomo-hanashite-honyaku-forbiz>

■クロスランゲージ、ホームページ自動翻訳サービスに「HTTPS 翻訳」オプションを導入

株式会社クロスランゲージでは、2016年7月1日よりホームページ自動翻訳サービスの上級版において、オプションサービスとして「HTTPS 翻訳」の機能を提供します。

http://internet.watch.impress.co.jp/docs/release/20160527_759365.html

AAMT Forum メールマガジン
2016/06/29 配信 vol.077

こんにちは。AAMT Forum メルマガ担当です。そろそろ太陽が恋しいこの頃、いかがお過ごしでしょうか。曇り空や雨がっていますが、日に日に暑くなってきており、夏が近づいているのを感じますね。では、メールマガジン第77号をお届けします。

■マイクロソフトら、2020年までにAIで訪日客向けリアルタイム翻訳提供へ

豊橋技術科学大学、日本マイクロソフト、ブロード

バンドタワーの3者は6月21日、AI・機械学習による多言語コミュニケーションサービスの実現に向けた協働を開始すると発表しました。

<http://news.mynavi.jp/news/2016/06/22/158/>

■Instagramが投稿やコメントなどの翻訳機能を追加す
すでに利用可能なユーザーも多数

月間のアクティブユーザーが5億人を超えたと発表したばかりの写真SNS「Instagram」が、翻訳機能を段階的に追加すると発表しました。

<http://japanese.engadget.com/2016/06/23/instagram>

AAMT Forum メールマガジン
2016/07/15 配信 vol.078

こんにちは。AAMT Forum メルマガ担当です。盛夏の候、いかがお過ごしでしょうか。今年は7月に入って台風1号が発生し、猛烈な台風となりましたが、発生が観測上で2番目に遅いそうです。発生が遅いと7~9月にまとまりやすいそうですので、この夏は台風に要注意ですね。では、メールマガジン第78号をお届けします。

■Facebook、複数の言語に自動翻訳して投稿する機能を導入

Facebookは米国時間7月1日、投稿を公開すると、ユーザーが選択した複数の言語に自動翻訳してくれる新しいソフトウェアを導入しました。

<http://japan.cnet.com/news/service/35085269/>

■ドコモ、「単語翻訳アプリ」「インタビュー翻訳アプリ」開発

NTT ドコモは、訪日外国人とのコミュニケーションに活用できる二つの翻訳アプリを開発しました。7月～9月にかけて横須賀市で実証実験を行い、実用化を目指します。

<http://k-tai.watch.impress.co.jp/docs/news/1008951.html>

AAMT Forum メールマガジン

2016/07/29 配信 vol.079

こんにちは。AAMT Forum メールマガ担当です。いよいよ夏本番となるこの頃、暑さに参っていないでしょうか。関東では平成6年並の渇水が続いているようです。いろいろと水を使えば涼しいですが、節水も心に留めておきましょう。では、メールマガジン第79号をお届けします。

■音声翻訳プラットフォーム「はなして翻訳 for Biz プレミアム」を提供開始

株式会社 NTT ドコモは、法人企業向けに、音声認識・機械翻訳・音声合成機能をまとめて利用できる翻訳プラットフォームの提供を開始しました。

https://www.nttdocomo.co.jp/info/news_release/notice/2016/07/14_00.html

■会話支援アプリ、店舗で活用 京都銀行

京都銀行は、訪日外国人や聴覚障害者との会話を支

援するアプリを搭載したタブレット端末を京都市内17店舗で7月21日から導入しました。

<http://kyoto-np.co.jp/economy/article/20160714000114>

AAMT Forum メールマガジン

2016/08/04 配信 vol.080

こんにちは。AAMT Forum メールマガ担当です。真夏の日差しが照りつけるこの頃、いかがお過ごしでしょうか。8月に入り、ますます暑くなってきました。外出する際だけでなく、室内でも熱中症には注意しましょう。では、メールマガジン第80号をお届けします。

■Google 翻訳にも使われている「統計的機械翻訳」のしくみを覗いてみよう

統計的機械翻訳の基本的な仕組みと、その統計的機械翻訳において広く使われている言語モデル法について解説記事です。

<http://codezine.jp/article/detail/9547>

■『食べ物世界の共通言語だ。』

Google 翻訳による多国籍レストランが NY に限定オープン他言語同士の橋渡しをする Google 翻訳と、食べ物の普遍性をリンクさせた、ユニークな試みが行われました。

<http://adgang.jp/2016/08/128799.html>

こんにちは。AAMT Forum メールマガ担当です。暑い日が続いておりましたが、皆様は熱中症になど罹っていないでしょうか。今年は台風の発生が遅い年でしたが、ついにやって来ましたね。これから続々と発生するようですので、台風の情報には注意しましょう。では、メールマガジン第 81 号をお届けします。

■KDDI、富士登山口で訪日外国人に多言語案内を開始、登山者にはモバイルバッテリー配布

KDDI は、富士山におけるより安心して快適な登山をサポートする取り組みの一つとして、3 つの富士山の登山口にて、多言語音声翻訳システムを活用した訪日外国人への多言語案内を開始しました。

http://denpanews.jp/etc/20160812_10611.html

■五山送り火「DJ ポリス」リアルタイム翻訳実施へ

ヤマハは、京都府下鴨警察署と連携し、8月16日に行われた京都五山送り火において、マナー順守や混雑緩和案内などの警察官の呼びかけを多言語化する実験を実施しました。

<http://smartphone.r25.jp/case-file/155917>

こんにちは。AAMT Forum メールマガ担当です。初秋の候、いかがお過ごしでしょうか。初めて東北の太

平洋側に上陸した台風 10 号でしたが、皆様大事無かったですでしょうか。大雨で涼しくなったと思ったら、熱中症に必要な程の暑さが戻ってきたりと、健康管理には注意しましょう。では、メールマガジン第 82 号をお届けします。

■ビジネス/科学技術分野向け翻訳ソフト「翻訳ピカイチ V15」を 2016 年 9 月 9 日(金) 新発売!

株式会社クロスランゲージは、Windows 対応の英日・日英翻訳ソフト「翻訳ピカイチ V15 for Windows」を、2016 年 9 月 9 日(金)より販売を開始します。

<http://prtimes.jp/main/html/rd/p/000000086.000002804.html>

■6割好意的 KDDI、鳥取のタクシーで実験 / 鳥取

鳥取市内の外国人観光客向けタクシーを使い KDDI が実証実験を進めていた音声翻訳システムの結果がまとまりました。KDDI は見つかった課題に対し、7月から12月まで2度目の実験を実施し、解消を図ります。

<http://mainichi.jp/articles/20160824/ddl/k31/040/482000c>

こんにちは。AAMT Forum メールマガ担当です。野分の候、いかがお過ごしでしょうか。夏も盛りを過ぎてだんだんと涼しく、過ごしやすくなって来ました。一方で、台風の発生が相次いでいます。東北、北海道

では大きな被害もでしたので、十分に注意しましょう。では、メールマガジン第 83 号をお届けします。

■医療現場で多言語音声翻訳の実証実験を開始

NICT と富士通株式会社は、実際の医療現場で多言語音声翻訳システムを活用した実証実験を、2016 年 11 月から 2017 年度末までの期間に行います。

<https://www.nict.go.jp/press/2016/09/09-2.html>

■JTB とパナソニック、多言語案内など訪日客向けサービス 2 種の実証実験開始

多言語での観光案内や接客を支援する「JAPAN FITTER」と、大型手荷物を配送して手ぶら観光を実現する「LUGGAGE-FREE TRAVEL」の実証実験を 2016 年 9 月 1 日から 10 月 31 日まで実施します。

http://itpro.nikkeibp.co.jp/atcl/news/16/083102530?s_cid=bpn_TopTl

第 26 回通常総会および関連行事の報告

AAMT 事務局

当協会の第 26 回通常総会が 2016 年 6 月 17 日（金）13 時より・ホテルアジュール竹芝にて開催された。総会后、招待講演会、さらにモデレーテッド セッションが行われた。そして第 11 回 AAMT 長尾賞授与式と第 3 回 AAMT 長尾賞学生奨励賞授与式と、引き続いての AAMT 長尾賞学生奨励賞および長尾賞受賞者による記念講演会が盛況のうちに行われた。各参加者数は下記に示す通りである。

総会の参加者：個人会員：出席 20 名、委任状参加者 25 名、
法人会員：出席 17 法人（法人会員参加者 26 名）、委任状参加 2 法人

報告会・展示会・講演会・AAMT 長尾賞関連の式及び講演会：

個人会員： 28 名
法人会員： 61 名
会員外一般：13 名
合計： 102 名

懇親会：

個人会員： 16 名
法人会員： 39 名
会員外一般：11 名
合計： 66 名

第 26 回通常総会

1. 開会の辞

2. 会長挨拶

名古屋大学・中岩浩巳

3. ご来賓挨拶

4. 出席会員の確認

5. 議案

第 1 号議案 2015 年度事業報告（案） 第 2 号議案 2015 年度決算報告（案）

第 3 号議案 2016 年度事業計画（案） 第 4 号議案 2016 年度収支予算（案）

第 5 号議案 理事・役員改選（案） その他・会員提案事項

6. 閉会の辞

創立 25 周年記念講演会

- ・「AAMT 設立までのこと」

長尾 真

(IAMT、AAMT 初代会長)

- ・「拡大する翻訳ニーズと翻訳業界における機械翻訳への期待」

東 郁男

(一般社団法人日本翻訳連盟 代表理事・会長)

招待講演会

- ・「翻訳品質と TAUS DQF」

西野 竜太郎

(合同会社グローバリゼーションデザイン研究所、TAUS Representative)

AAMT からのお知らせ

- ・「MT Summit XVI 開催について」 中岩 浩巳 (AAMT 会長)

展示会

みんなの自動翻訳@TexTraR

国立研究開発法人 情報通信研究機構

Globalese 多言語機械翻訳エンジン

株式会社川村インターナショナル

機械翻訳技術を応用した読解支援情報付きニュース

NHK 放送技術研究所

未来を拓く！インバウンド翻訳ソリューション

～訪日観光客 4000 万人時代に向けて～

株式会社高電社

新製品 新製品パッケージシリーズの展示

株式会社クロスランゲージ

サーバ型翻訳製品「ODMA 自動翻訳」と応用アプリご紹介

富士通株式会社

AAMT 各委員会活動報告会

(本年度の委員会活動報告は展示形式にて行われました。)

- | | | | |
|-----------------------|------|-------|-------------|
| 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 | 須藤 克仁 | (NTT) |
| 3. インターネットワーキンググループ | リーダー | 富士 秀 | (情報通信研究機構) |
| 4. 編集委員会 | 委員 | 小谷 克則 | (関西外国語大学) |

2016 年度総会後の組織・人事について

人事

- ・ 会長、副会長、委員長、ワーキンググループリーダー、事務局長は下記の通りとする。
- ・ (敬称略：50 音順)

会長	中岩 浩巳 (名古屋大学)
副会長	飯田 仁 (東京工科大学名誉教授)
	小谷 泰造 (株式会社インターグループ)
理事	安達 久博 (株式会社サン・フレア)
	井佐原 均 (豊橋技術科学大学)
	石崎 俊 (慶應義塾大学名誉教授)
	宇津呂 武仁 (筑波大学)
	熊野 明 (東芝ソリューション株式会社)
	黒橋 禎夫 (京都大学)
	坂本 義行
	隅田 英一郎 (国立研究開発法人 情報通信研究機構)
	辻井 潤一 (国立研究開発法人 産業技術総合研究所)
	田中 英輝 (NHK放送技術研究所)
	長尾 真 (国際高等研究所)
	永田 昌明 (日本電信電話株式会社)
	中村 哲 (奈良先端科学技術大学院大学)
	星野 正広 (富士通株式会社)
	守屋 敏道 (一般財団法人日本特許情報機構)
	横山 晶一 (山形大学名誉教授)
	Key-Sun Choi (韓国 KAIST)
	Virach Sornlertlamvanich (タイ SIIT)

監事	種子田 暁夫 (日本電気株式会社)
	古谷 祐一 (一般社団法人日本翻訳連盟・GMOスピード翻訳株式会社)
退任	川村 みどり (一般社団法人日本翻訳連盟・ 株式会社川村インターナショナル)

機械翻訳課題調査委員長	長瀬 友樹 (株式会社富士通研究所)
AAMT/Japio 特許翻訳研究会委員長	辻井 潤一 (産業技術総合研究所)
編集委員長	宇津呂 武仁 (筑波大学)
インターネットワーキンググループリーダー	富士 秀 (情報通信研究機構)
事務局長	神崎 享子 (豊橋技術科学大学)

組織

AAMT機械翻訳課題調査委員会

退任： 森 岳至（日本電信電話株式会社）

島村 修司（インターグループ）

平田 周（ユニバーサルコンテンツ）

新任： 秋葉 泰弘（日本電信電話株式会社）

高橋 博之（株式会社クロスランゲージ）

編集委員会：

新任： 目次 由美子（LOGOStar）

新任： 河野 弘毅（ポストエディット東京）

その他委員会、ワーキンググループとも現状どおりとする。

以上

第 11 回 AAMT 長尾賞・第 3 回 AAMT 長尾賞学生奨励賞授与式
及び記念講演会

第 11 回 AAMT 長尾賞

受賞者：株式会社 みらい翻訳

受賞理由：

機械翻訳の基本技術となっている RBMT や SMT などを統合して、翻訳精度や速度に見合った自動翻訳サービスを提供する初めての機械翻訳プラットフォームを開発して、B2B ビジネスを実現した点が長尾賞にふさわしい。

SNS 向けの自動翻訳市場への今後の展開も大いに期待できることを補足する。

選考委員長：飯田 仁（東京工科大学名誉教授）

選考委員：Key-Sun Choi（韓国 KAIST）

宇津呂 武仁（筑波大学）

梶 博行（元静岡大学）

隅田 英一郎（情報通信研究機構）

推薦者：安達 久博（株式会社サン・フレア）

同意人：小林 照二（株式会社フィート）

授賞記念講演



第3回 AAMT 長尾賞学生奨励賞

受賞者：奈良先端科学技術大学院大学・博士後期課程

三浦明波

受賞対象論文：

Graham Neubig(ニュービック)の指導のものに、2016年度奈良先端科学技術大学院大学の修士論文としてまとめられた以下の論文が受賞対象論文である。

「中間言語モデルを用いた多言語機械翻訳の精度向上」

受賞理由：

本論文は、翻訳対象となる言語対に十分な量の対訳データが得られない場合に、第三の言語を中間言語として利用するピボット翻訳の新しい実現法を提案し、翻訳精度を大きく改善できることを示している。提案法は、今後、英語を介した日本語とアジア言語の翻訳のような語順が大きく異なる言語対への適用が期待できるなど、ピボット翻訳に関する新しい研究の方向性と今後の発展の可能性を示す優れた論文であり、AAMT 長尾学生奨励賞にふさわしいと考える。

選考委員長：永田昌明（日本電信電話株式会社）

選考委員：

鄭 育昌（富士通研究所）

藤田 篤（情報通信研究機構）

山本 和英（長岡技術科学大学）

推薦者：Graham Neubig（奈良先端科学技術大学院大学）

授賞式



協会活動報告

(2016年5月～2016年8月)

第26回通常総会

2016年6月17日

- | | | | |
|-------|---------------|------------|---------------|
| 第1号議案 | 2015年度事業報告(案) | 第2号議案 | 2015年度決算報告(案) |
| 第3号議案 | 2016年度事業計画(案) | 第4号議案 | 2016年度収支予算(案) |
| 第5号議案 | 理事・役員改選(案) | その他・会員提案事項 | |

報告会

2016年6月17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

創立25周年記念講演会

- ・「AAMT 設立までのこと」

長尾 真

(IAMT、AAMT 初代会長)

- ・「拡大する翻訳ニーズと翻訳業界における機械翻訳への期待」

東 郁男

(一般社団法人日本翻訳連盟 代表理事・会長)

招待講演会

- ・「翻訳品質と TAUS DQF」

西野 竜太郎

(合同会社グローバリゼーションデザイン研究所、TAUS Representative)

第11回 AAMT 長尾賞授与式・記念講演会

受賞者：

株式会社 みらい翻訳

受賞理由：

機械翻訳の基本技術となっている RBMT や SMT などと統合して、翻訳精度や速度に見合った自動翻訳サービスを提供する初めての機械翻訳プラットフォームを開発して、B2B ビジネスを実現した点が長尾賞にふさわしい。

SNS 向けの自動翻訳市場への今後の展開も大いに期待できることを補足する。

第3回 AAMT 長尾賞学生奨励賞

受賞者：

奈良先端科学技術大学院大学・博士後期課程

三浦明波

受賞対象論文：

「中間言語モデルを用いた多言語機械翻訳の精度向上」

受賞理由：

本論文は、翻訳対象となる言語対に十分な量の対訳データが得られない場合に、第三の言語を中間言語として利用するピボット翻訳の新しい実現法を提案し、翻訳精度を大きく改善できることを示している。提案法は、今後、英語を介した日本語とアジア言語の翻訳のような語順が大きく異なる言語対への適用が期待できるなど、ピボット翻訳に関する新しい研究の方向性と今後の発展の可能性を示す優れた論文であり、AAMT 長尾賞学生奨励賞にふさわしいと考える。

懇親会

2016年6月17日 ホテルアジュール竹芝 21階フレンチレストラン『ベイサイド』にて行なった。

決算理事会

2016年6月17日

- | | | | |
|-------|---------------|-------|---------------|
| 第1号議案 | 2015年度事業報告(案) | 第2号議案 | 2015年度決算報告(案) |
| 第3号議案 | 2016年度事業計画(案) | 第4号議案 | 2016年度収支予算(案) |
| 第5号議案 | 理事・役員改選(案) | | その他・会員提案事項 |

機械翻訳課題調査委員会

2016年5月13日(金)(2016年度 第2回)

① 各WGの活動について(各WGに分かれて議論)

(WG1,2)

- ・総会資料の確認
- ・今年度の進め方について
- ・テストサイト進捗
- ・その他

(WG3)

- ・総会の事業報告と事業計画の確認
山本委員の原案を確認。多少の表現調整を行い、最終的な確認完了。
- ・UTX1.20仕様検討
山本委員の最終案に対して他の委員が行ったコメント(前回の宿題)のうち、大きな論点に対して議論。
- ・ISO会議(コペンハーゲン)
山本委員がUTXを使用してのISO用語の各国語の訳語付与、多言語用語集作成を提案するために、参加を希望しており、その是非について議論。
- ・その他
- ・宿題

全体での意思統一が必要な論点を論点シートに追加する。

② 全体会議

- ・各WGの検討事項共有
- ・AAMT翻訳フェアの準備について
- ・MT Summit2017の準備について
- ・その他

2016年6月9日(木)(2016年度 第3回)

① 各WGの活動について(各WGに分かれて議論)

(WG1,2)

- ・テストサイトの状況
- ・翻訳フェアについて

- ・総会資料の確認
- ・今年度の進め方について
- ・その他

(WG3)

- ・ISO 会議(コペンハーゲン)
 - 山本委員が ISO 会議で、UTX を使用しての ISO 用語の多言語用語集作成を提案したいとの件について議論。
- ・UTX1.20 仕様検討
 - 各委員のコメントを反映した最新版について、いくつかの論点について議論。
- ・宿題
 - 山本委員が「1.4 Field, field value, and field item」を書き直し、他の委員はそれを確認して意見を出す。

② 全体会議

- ・各 WG の検討事項共有
- ・AAMT 翻訳フェアの準備について
- ・MT Summit2017 の準備について
- ・WAT への協力について
- ・その他

2016 年 7 月 13 日 (水) (2016 年度 第 4 回)

① 各 WG の活動について (各 WG に分かれて議論)

(WG1,2)

- ・今年度の具体的活動計画
- ・テストセット評価について
- ・その他

(WG3)

- ・UTX1.20 仕様検討
- ・宿題
 - 7/25 までに UTX1.20 仕様書の全体を見直す (最終確認)

② 全体会議

- ・各 WG の検討事項共有
- ・AAMT 翻訳フェアの振り返り
- ・MT Summit2017 の準備について
- ・その他

・インターネットWG

- ① 総会・AAMT 機械翻訳フェア前後で生じた更新事項に関するコンテンツ更新
- ② MT Summit XVI に向けたホームページ立ち上げの検討および作業依頼
- ③ AAMT ジャーナル目次情報等の通年コンテンツの掲載・更新
- ④ 利便性・プレゼンス向上に向けたサイト構成見直しの検討

編集委員会

- ① AAMT Journal No.63 の執筆依頼及び編集作業

AAMT/Japio 特許翻訳研究会

2016年5月13日(金)(2016年度 第1回)

1. 日本特許情報機構 大塩様からのご発表
2. 平成27年度報告書の紹介(2)
3. AAMT 総会での研究会活動報告
4. 平成28年度 AAMT/Japio 特許翻訳研究会 活動計画書依頼のお知らせ
5. 議事録の確認
6. その他
7. 次回の開催について
 - ・開催の日時(場所) 6月24日予定
 - ・主な議題

AAMT/Japio 特許翻訳研究会 拡大評価部会

2016年5月13日(金) (2016年度 第1回)

1. 平成28年度の拡大評価部会の活動計画
2. 各グループの活動発表
 - ・自動評価グループ
 - ・昨年度提案した「係り受け比較法」またはその改良によるスクランプリングへの対応のIMPACTなどの既存評価手法による比較実験
 - ・深層学習を用いた翻訳自動評価手法の試み
 - ・人手評価グループ
 - ・テストセットグループ
3. その他
4. 次回開催日予定 9月予定

AAMT/Japio 特許翻訳研究会

2016年 6月24日(金) (2016年度 第2回)

1. 議事録の確認
2. 京都大学黒橋研究室 修士2年 大谷さんのご発表
「機械翻訳のための人手評価の統合方法」
3. 28年度 AAMT 総会での研究会活動報告
4. 28年度 AAMT/Japio 特許翻訳研究会 活動研究計画書発表

5. 第4回特許情報シンポジウム 日程決定のお知らせ

2016年11月25日金曜日決定

6. 新メンバーの推薦について

7. その他

8. 次回の開催について

- ・開催の日時（場所）7月29日（金）予定
- ・主な議題

AAMT/Japio 特許翻訳研究会

2016年 7月29日（金）（2016年度 第3回）

1. 議事録の確認

2. マイクロソフトディベロップメント株式会社 吳先超様のご講演

対話システムを構成する諸技術について「女子高生人工知能りんなの対話システム」

3. 第4回特許情報シンポジウム 日程決定の今後の予定のお知らせ

シンポジウム開催場所変更のお知らせ。

グランパークプラザ田町にて開催決定

（グランパークプラザ田町 <http://www.granpark.jp/>）

4. 新メンバーの推薦について

5. その他

6. 次回の開催について

- ・開催の日時（場所）9月9日（金）開催予定
- ・主な議題

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 63 号をお送りします。

今号の巻頭言は、GMO スピード翻訳株式会社 古谷祐一様より、御寄稿を頂きました。

AAMT の活動として、今号に先立ちまして、2016 年 6 月に開催されました総会におきまして、AAMT 初代会長の長尾真先生より、AAMT 創立 25 周年基調講演を賜りましたが、今号におきましては、機械翻訳課題調査委員会・AAMT ジャーナル編集委員会の目次由美子様より、長尾真先生の御講演内容についての貴重な御寄稿を頂きました。

同総会におきましては、日本翻訳連盟(JTF)会長の東郁男様より、AAMT 創立 25 周年記念講演を賜りましたが、今号におきましては、御講演内容についての貴重な御寄稿を頂きました。

加えて、同総会におきましては、合同会社グローバリゼーションデザイン研究所・TAUS Representative の西野竜太郎様より貴重な御講演を賜りましたが、今号におきましては、御講演内容についての貴重な御寄稿を頂きました。

あわせて、AAMT 長尾賞の選考結果を受けまして、受賞者である株式会社みらい翻訳 栄藤稔様より受賞内容についての貴重な御寄稿を頂きました。

また、AAMT 長尾賞学生奨励賞の選考結果を受けまして、受賞者である奈良先端科学技術大学院大学 三浦明波様より、受賞理由となった「中間言語モデルを用いた多言語機械翻訳の精度向上」についての紹介記事を御寄稿頂きました。

さらに、総会会場におきまして、「AAMT 機械翻訳フェア (MT フェア) 2016」と題しまして、昨年に続き、機械翻訳技術提供各社、ならびに、機械翻訳課題調査委員会からの展示の試みを実施したのを受けまして、展示コーナーへの参加団体より、情報通信研究機構 内山将夫様、NHK 放送技術研究所 田中英輝様、株式会社富士通研究所 片江伸之様から、それぞれ、展示参加報告を御寄稿いただきました。

また、理事の坂本義行先生からは、機械翻訳の歴史に関する貴重な御寄稿の第八弾を頂きました。

今号が会員の皆様のお手元に届くその少し後、来る 11 月 29 日(火)に、AAMT が後援をします第 26 回 JTF 翻訳祭が開催されます。日本翻訳連盟(JTF)理事・第 26 回翻訳祭企画実行委員長 古谷祐一様からは、この JTF 翻訳祭のお知らせを御寄稿いただきました。

その他、2016 年 3 月に開催されました、言語処理学会第 22 回年次大会テーマセッション「文理・産学を越えた翻訳関連研究」の報告を、関西大学 山田優先生、情報通信研究機構 藤田篤様、東京大学 影浦峯先生より御寄稿頂

きました。秋桜舎 山本ゆうじ様からは、2016年8月に開催されましたTCシンポジウム2016の参加報告を御寄稿頂きました。

さらに、豊橋技術科学大学 井佐原均先生から、プロジェクト報告「多言語情報発信支援の社会実装に関する研究」を御寄稿頂きました。

株式会社ヒューマンサイエンス 徳田愛様からは、「産業翻訳分野における日英機械翻訳の実用化に向けて」を、富士通株式会社 上柿稔様からは、「FUJITSU Business Application Operational Data Management & Analytics 自動翻訳」を、それぞれ御寄稿頂きました。

AAMT内の活動報告として、機械翻訳課題調査委員会から、「AAMTフォーラムメールマガジンバックナンバー」を掲載いたしました。

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）

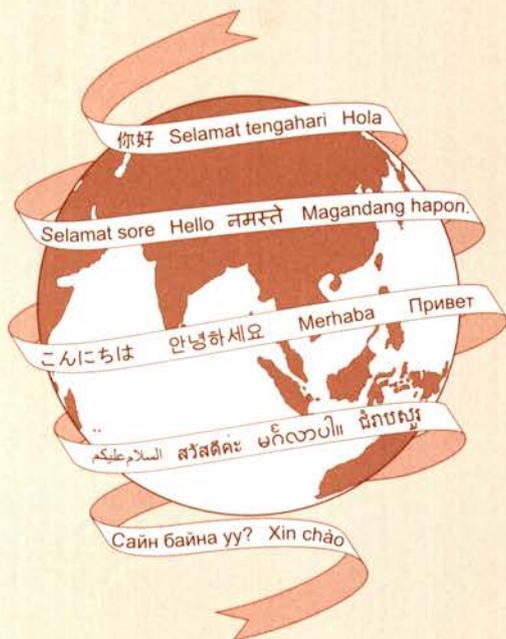
ご関心のある方は、事務局までお問い合わせください。

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

AAMT



AAMTジャーナル No.63

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

阿部 さつき 釜谷 聡史 河野 弘毅 目次 由美子

表紙(図部分)デザイン：阿部 さつき

事務局：神崎 享子 荻野 孝野

印刷所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)

c/o Japan System Application Co., Ltd.

Suzuki Building 3F 2-55-2, Ikebukuro, Toshima-ku Tokyo 171-0014, JAPAN

aamt-info@aamt.info