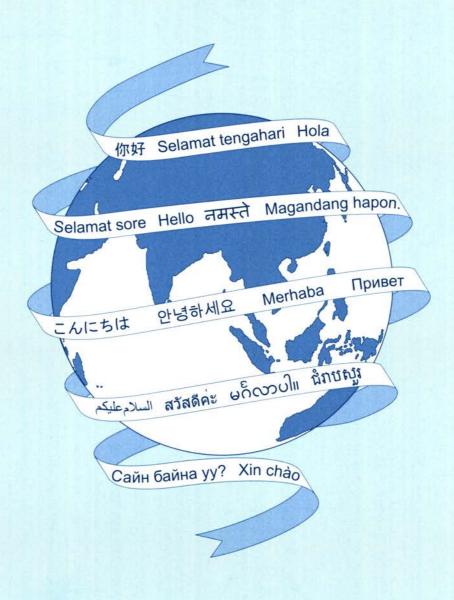
Asia-Pacific Association for Machine Translation

# 



October 2014 **No.57** アジア太平洋機械翻訳協会

# 目 次

巻頭言:	機械翻訳への今後の期待	
総会講演:	外国人旅行者の受入体制整備と機械翻訳の可能性	
総会講演:	機械翻訳を使いこなす ~品質と生産性の向上のために~	
プロジェクト報告:	豊橋技術科学大学における機械翻訳の研究開発	井佐原 均14
連載:	機械翻訳事始め[IV]	坂本 義行17
研究報告:	機械翻訳を支援ツールとして使う英語学習法	平田 周27
シンポジウム報告:	法令工学の言語処理	
シンポジウム報告:	TAUS Tokyo Executive Forum 2014 - A Sway of Innovation	Jaap van der Meer37
シンポジウム参加報告:	TKE(Terminology and Knowledge Engineering) 2014 報告	影浦 峡41
シンポジウム参加報告:	機械翻訳とデザイン思考	鈴木 博和44
パネル参加報告:	「ISO17100 と日本における翻訳職業教育の将来」	
	パネルディスカッション参加報告機	械翻訳課題調查委員会47
シンポジウム案内:	第 24 回 JTF 翻訳祭のお知らせ	中尾 勝48
AAMT 長尾賞受賞講演:	言語構造が大きく異なる言語間の長文を対象とする統計的機械翻訳の実	用化
	h	山 将夫, 隅田英一郎50
AAMT 長尾賞学生奨励賞:	自治体文書の多言語化を支援する枠組みとシステム環境の研究	宫田 玲54
AAMT MT フェア 2014 展示	示報告:AAMT 機械翻訳フェア(MT フェア) 2014 展示ブースの報告	
AAMT-MT フェア 2014 展示	示報告 : 翻訳支援ツール Transit NXT	目次 由美子69
AAMT-MT フェア 2014 展示	、報告:特許翻訳などの専門的な外国語文書も自然な日本語に!	
	~多言語統計翻訳プラットフォーム~	我妻光洋,荒井和博70
委員会活動報告	これまでの AAMT Forum メールマガジン Vol.3機械翻訳記	果題調査委員会 WG1,2 72
委員会活動報告:	AAMT 日中/中日テストセットに基づく翻訳自動評価サイトの公開	
	機械翻訳:	果題調査委員会 WG1 77
AAMT 会員のひろば:		翻訳センター84
AAMT 会員のひろば:		新田 義彦86
事務局からのお知らせ:	第 24 回通常総会および関連行事の報告	AAMT 事務局89
事務局からのお知らせ:	協会活動報告(2014年6月~2014年8月)	AAMT 事務局94
編集後記		99

# CONTENT

Foreword:	Future expectation to machine translation
General Meeting:	On the Possibility of Utilizing Machine Translation
	in the Field of Inbound Tourism
General Meeting:	Effectively Utilizing Machine Translation to Improve Quality and ProductivityM. Tokuda9
Project Report:	R&D of Machine Translation at Toyohashi University of Technology
Series:	History of Machine Translation [IV]
Report:	An application of MT as an Assistant Tool for English Learning
Symposium Report:	Language Processing of Legal Engineering
Symposium Report:	TAUS Tokyo Executive Forum 2014 - A Sway of Innovation
Symposium Report:	On TKE (Terminology and Knowledge Engineering) 2014
Symposium Report:	Machine Translation and Design Thinking
Panel Report:	Report on participation in the panel discussion
	"ISO17100 and professional translators training in Japan and the future" MT
Symposium Information:	Information for 24th JTF TRANSLATION FESTIVAL TOKYO 2014
AAMT Nagao Award	The Deployment of Statistical Machine Translation System
	for Long Sentences in Distant Languages
AAMT Nagao Student Av	ward An Authoring Environment to Support Multilingualization of Municipal Documents:
	Principle and Implementation
Exhibition report:	Report on AAMT MT fair 2014
Exhibition report:	Transit ^NXT - Computer-Assisted Translation with Translation MemoryY.Metsugi
Exhibition report:	Even Technical Documents Such as Patents Can Be Translated in Natural Japanese!
	- NTT Platform for Building Statistical Machine Translation Systems
	among Multiple Languages
Committee Report:	AAMT Forum mail magazine vol.3
Committee Report :	Releasing a New Website of Automatic MT Evaluation Based
	On AAMT C2J/J2C Test Sets
AAMT Members:	
AAMT Members	
AAMT Activities:	General Meeting
AAMT Activities:	AAMT Activities (from June 2014 to August 2014)94
Editor's Note:	

# 機械翻訳への今後の期待

### 株式会社クロスランゲージアールアンドディ

### 代表取締役 古賀 勝夫

#### 1. 商用機械翻訳の歩み

機械翻訳の研究は、1950 年代に米国ジョージタウン大学での研究発表から本格化したが、我が国の商用機械翻訳システムは、1984 年に富士通から汎用機上で動くシステムが発売されて始まった。以降、80 年代後半はワークステーション版が主流となり、弊社(当時は株式会社ノヴァ)においても、1989 年にワークステーション上の英日機械翻訳システムを発売開始した。当時はまだ、年間を通してせいぜい数千台のマーケット規模であった。

1990 年代はハードの進展とともにダウンサイジ ングされて PC 版が主流となり、一気に数十万台の 市場規模に膨れ上がった。

1990 年代半ばから 2000 年代には、インターネットの普及と共に、ブラウザ上で動作する機械翻訳が普及に拍車をかけ、さらに 2000 年初頭からヤフー、エキサイト、グーグルといった大手ポータルサイトによる無料翻訳サービスが普及しはじめ、2010 年までには、数百万の利用者に膨れ上がった。さらに2010 年代には携帯上でも利用できるようになって、今や機械翻訳の利用者は 1 千万人をも超える勢いである。

商用の翻訳ソフトが登場して約30年間、翻訳精度はかなり向上してきたが、ユーザ満足度という視点では、まだまだ多くの改善が望まれている。

技術的に見れば、機械翻訳普及の当初から今日に 至るまで、ルールベース翻訳(RBMT)が主導してきた が、今日では統計手法による翻訳技術がグーグル翻 訳と共に着目され、今後の機械翻訳の精度向上に対 し、統計ベース翻訳(SMT)の技術は大いに期待され ており、今後、RBMT、EBMT(例文ベース翻訳)、SMT は段々と融合化されていくのではないだろうか。

#### 2. 現在のサービス提供形態

現在、機械翻訳サービスは、多いものでは 50 か国語 以上の言語対に対し提供されている。その代表的な提 供形態を、以下に示す。今後はさらに新たなサービス が提供されるであろう。

提供形態	概要
インターネッ ト無料翻訳サー ビス	ブラウザ上で翻訳する仕組みで、ヤフ ー、エキサイト、グーグル、楽天、OCN 等々のポータルサイトから提供されて いる。
ホームページ 翻訳サービス	ホームページ上に翻訳ボタン (言語指定) をつけて、ユーザがそれをクリック すれば、そのページを自動翻訳する。
Web 翻訳	サーバーから、ブラウザインタフェースで利用。Webページの他、一般テキスト、MS-OFFICE文書等々を翻訳できる。 企業内サーバーで利用する他、クラウド型でも利用できる。
パッケージ翻 訳サービス	パソコン等にインストールして利用 するスタンドアローン型の翻訳アプリ で、Web 翻訳と違って、様々な編集機能 を持っている。
携帯翻訳	携帯電話から翻訳を利用する。音声で入力でき、翻訳結果は音声で出力できる。iPhone 等のスマートフォン上にも多数の翻訳アプリが提供されている。

#### 3. ユーザは機械翻訳に何を求めるか

ユーザの利用目的によっても、機械翻訳に求めるも のは変わってくる。以下、いくつかの代表的な利用 目的にそって述べる。

① 情報が早くかつ参考程度にわかればいい場合

このケースは、当初より現在まで機械翻訳が最も広く利用されてきた場面である。たとえば、海外ホテルサイトの概要を見る場合や、海外企業、海外ニュースをざっと把握する場合等では、現状の翻訳精度でもかなり利用されている。

#### ② 情報の正確な理解が必要な場合

本来、ユーザにとっては、これが機械翻訳に求める究極の目標であろう。この目的では、翻訳精度として少なくとも80%程度は望まれる。機械翻訳にとって、この目標に近づくにはまだまだ幾多の改良を必要とする。

#### ③ 翻訳の生産効率をあげる目的の場合

これは、毎年大量にマニュアル翻訳が必要な企業 等においては、既に一部では実用的に機械翻訳が利用 されている。特に多言語翻訳を必要とされる場合は、 なおさら効果的である。このケースでは、大量の翻訳 を行っている企業にとって、翻訳結果を蓄積していく ことにより機械翻訳の精度は向上し、それによって翻 訳の生産性が増し、またその成果を翻訳メモリーとし て蓄積することでさらに機械翻訳の精度は向上して いくという好循環も生まれる可能性は高い。

#### ④ コミュニケーションの場合

最近話題の旅行会話を例にとれば、何より分かり 易いことが重要である。意味的には正解であっても、 滑らかでない表現は好まれない。また、旅行会話では、 場面に応じて必ず出てきそうな定型パターンが豊富 にあり、ある程度の会話例(たとえば50~100万文) を収集すれば、定型パターンは正確にカバーできるだ ろう。このケースでは、SMT は非常に適合する。

#### ⑤ 情報を発信する場合

たとえばメールやチャット等で情報発信する場合などである。この場合は、利用者は比較的分かり易い単純な文を入力すれば、高精度の機械翻訳が期待できるので、ターゲット言語による情報発信にも十分役立つ。ただし、機械翻訳ではしばしば誤った翻訳も出すので、利用者はある程度ターゲット言語を理解でき

ることが求められる。

#### 4. 現状の技術的課題

まず、RBMT、EBMT、SMT の方式と、その長所およ び短所を下記に記す。

び短所を下記に記す。			
	RBMT	SMT	ЕВМТ
方	原文の構造	過去の大量翻訳	定型(パター
式	を文法的に	文を蓄積し、そ	ン) 文を訳文
	解析して、目	れから確率的に	と共に蓄積
	的言語に翻	最適の組み合せ	し、一致すれ
	訳する。	を探して翻訳文	ばこの対訳を
		を合成する。	採用。
長	・短文から長	・頻出する表現	・あてはまれ
所	文まで比較	の訳が滑らか。	ば、翻訳精度
	的安定して	従って誤りの少	は高い。
	訳出。	ない短文等では	・特に定型表
	・訳語がぶれ	上質訳が出る。	現が多ければ
	にくい。	・複合語の訳語	コストパフォ
	<ul><li>一定規模の</li></ul>	が的確になる。	ーマンスは高
	辞書で対応	・言語対に寄ら	٧٠ <u>,</u>
	可能。	ず基本プログラ	
	・ユーザごと	ムは同じため、	
	のカスタマ	プログラム開発	
	イズ容易。	コスト・工数が	
		抑えられる。	
短	・意味は分か	・文法的乱れ。	・作成に訓練
所	るが表現が	・訳抜けや、想像	が必要。
	固い。(会話	できない誤訳等	・パターンー
	文に向きに	発生。	致文の変数部
	< v v)	・大量の翻訳事例	分が長けれ
	・長文になる	のデータ作成費	ば、そこに翻
	ほど誤訳は	用が膨大。	訳誤りが生じ
	増す。	・ユーザ主導の	る確率が高く
	・言語対毎に	カスタマイズ困	なる。
	開発が必要。	難。	

また、翻訳の正確さという観点から見た場合、流暢 さと訳語の正確性が求められるが、上記の3方式に 対し、この観点から技術的課題を下記にあげる。

	RBMT	SMT	EBMT
流暢さ	<ul> <li>・訳文の表現が固い。</li> <li>・解析で誤った場合、原文とかけ離れた訳文になる。</li> <li>特に長文になるほどこの傾</li> </ul>	全体的な滑らか さが増える一 方、想定できない誤りが起こり うる。例えば構 文の乱れ以外、 文法誤り、訳抜 け、原語にない	・パター ンの可 変 数 長 く ば 、 同 間 様 の 問 り が に た に に り に り に り に り に り に り に り に り に
訳語	向は強い。 ・複数の訳語をもし、複数の語、語に対し、那部の音音をはし、が多性を収集が多性を収集がある。 ・複合を変える。 ・複合を変える。 ・複合を変える。 ・さいたいさい。	訳語の登場等々 ・訳抜けがある。 ・逆に言語に対 応しない訳語が 登場する。 ・数字や記号の 様に変わるはず のない語に余計 な語がついたり 消えたりする。	・上記同様

#### 5. 今後に向けて

機械翻訳は、RBMT に始まり、EBMT、そして SMT と新たな手法が試みられている。しかし、類似言語間や特定の分野を除けば、いずれも 80%以上の翻訳精度を獲得するには至っていない。そうした中で、それぞれの方式の長所を生かし、短所を補う試みが行われ、少しずつ3つの方式の距離がせばまってきている。

RBMT による的確なソース言語の構造把握、EBMT による定型パターンの的確な把握、そして SMT による滑らかな訳文生成および的確な訳語生成という それぞれの長所を生かしたハイブリッド翻訳が今

後期待される。その結果、近い将来に80%を超える 機械翻訳の登場が期待される。

ハイブリッド翻訳を実現するために、何より重要なことは、翻訳対象分野に対し、それぞれに大量の対訳コーパス(人手により翻訳された原文と訳文の正しい翻訳対)が必要なことである。

通常、SMTでは少なくとも100万文以上の対訳コーパスが求められる。SMTは、その対訳コーパスから翻訳時に必要な翻訳モデルと言語モデルを構築し、それを翻訳対象文と照合して確率的に最適解を求める手法を取るため、翻訳対象分野を十分カバーできるだけの量が求められる。しかし、対訳コーパスを作成するには莫大な時間とコストがかかり、単一企業の力では困難な状況である。

現在、特許分野に対しては、特許庁により対訳コーパスが作成され、一部機関に提供され出しており、これを利用して今後特許分野では、飛躍的な翻訳精度の向上も期待できる。政府には、この動きをさらにもっと幅広く多分野に広げて欲しいと望むものである。たとえば、2020年のオリンピック開催に向け、外国人観光客の増大を考えても、観光分野での飛躍的な翻訳精度の向上が望まれ、そのためには一刻も早く観光分野での大量の対訳コーパス作成が必須である。この成果を研究機関だけではなく広く一般企業にも提供して、商品開発に生かされることが望まれる。

我が国の機械翻訳の発展、産業の発展、それを通 しての対外的競争力の強化を考えれば、一刻も早く 幅広い分野での対訳コーパス作成を国策として取 り組まれんことを期待するものである。

#### 総会講演

# 外国人旅行者の受入体制整備と機械翻訳の可能性

日本政府観光局(JNTO) 事業連携推進部 観光情報戦略室長 山田 洋

#### 1. はじめに

インバウンド旅行(外国人の訪日旅行)が注目を 集めるようになっている。訪日外国人旅行者数は 2013年に初めて1千万人を超え、2020年オリンピック・パラリンピックの東京開催決定も受けて、2 千万人の高みを目指す新たなステージに入った。受 入体制を整備し、訪日旅行の満足度を高めてリピーター増加を図ることが今後益々重要となるが、その一環として機械翻訳の活用が有望視される。本稿では、わが国のインバウンド旅行の状況等を概説するとともに、この分野における機械翻訳の可能性を展望したい。

まず、インバウンド旅行振興を専門に行う公的機関「日本政府観光局(JNTO)」について、簡単に説明させていただきたい。JNTO は外国人観光旅客の来訪促進を目的とする独立行政法人で、前回の東京オリンピックが開催された1964年に設立されてから今年で50周年を迎えた。海外に14の事務所を設置し、わが国の国策としての「ビジット・ジャパン事業」の展開に主要な役割を果たしている。主な取組みの分野は、「日本ブランドの売り込み」、「MICE(国際会議や企業報奨旅行等)の拡充強化」、「受入体制の整備」の三つである。

#### 2. 外国人の訪日旅行の状況

JNTO が設立された 1964 年当時、年間の訪日外国人旅行者は僅か 35 万人程度であったが、その後、20 年間で 200 万人から 300 万人程度のペースで増加を続けた。インバウンド旅行振興のための国策として「ビジット・ジャパン・キャンペーン (VJC)」が開始された当時、訪日外国人旅行者は約 520 万人であったが、国を挙げた取組みが功を奏して、VJC

開始前と比べて訪日外国人旅行者数の増加に大き く弾みがついた。そして、訪日外国人旅行者数は 2013年に約1,036万人となり、VJC開始後の約10 年間で倍増となる顕著な伸びを示した<sup>(1)</sup>。

この顕著な伸びを牽引しているのは東アジアや 東南アジアの市場である。2013 年に最も高い伸び 率を示したのはタイからの訪日旅行者数で、対前年 比74.0%増となる453,600人を記録した。また、香 港(745,800人、対前年比54.8%増)、ベトナム (84,400人、同53.0%増)、台湾(2,210,800人、同50.8%増)、マレーシア(176,500人、同35.6%増)、インドネシア(136,800人、同34.8%増)の 各市場も、対前年比で大幅に増加した。ちなみに、 VJCが開始された2003年の訪日旅行者数と比べる と、タイは5倍強の著しい増加となっている。

訪日外国人旅行者の市場別構成比の推移を見ると、アジア市場の増加傾向が明瞭である。訪日外国人旅行者数に占める主要アジア市場(韓国、台湾、中国、香港、タイ、シンガポール、マレーシア、インドネシア、フィリピン、インド、ベトナム)からの旅行者の割合は、2003年には66.1%であったが、2013年には76.7%となり、約10%増加している。一方、主要欧米豪市場(英国、フランス、ドイツ、ロシア、米国、カナダ、豪州)の割合は、2003年の26.4%から2013年の16.7%へ、約10%減少している。現在、訪日外国人旅行者の4人に3人は主要アジア市場からの旅行者となっており、その割合は今後さらに高まると見込まれる。

訪日外国人旅行者数の顕著な増加傾向(2013年 は前年に比べ24.0%増)は継続しており、2014年 1月から7月まで、月間の訪日外国人旅行者数は毎 月過去最高を更新した。この7ヶ月間に日本を訪れ た外国人旅行者は約753万人で、対前年同期比26.4%を記録しており、わが国のインバウンド旅行振興は順調な状況にあると言える。しかし、諸外国と比べれば必ずしも楽観視はできない。外国人訪問者数を世界各国・地域と比較すると、2012年時点で日本は世界第33位(約835万)で、フランス(約8,300万人)や米国(約6,700万人)等のインバウンド旅行の「先進国」はもとより、中国、マレーシア、香港、タイ、シンガポール等のアジア諸国・地域の後塵を拝している。残念ながら、日本はインバウンド旅行に関しては依然として「発展途上国」と言うべき状況にある。

訪日外国人旅行者の特徴としては、全体の半分が 観光目的で来訪していること、団体旅行でなく個人 旅行の形態で訪日する旅行者が増えていること、情 報収集等に関する ICT 化が進んでいること等があ げられる。ちなみに、観光庁の調査によれば、日本 滞在中に得た旅行情報で役に立ったものとして「イ ンターネット (パソコン)」の選択率が 41.5%で最 も高くなっている <sup>(2)</sup>。

#### 3. 訪日外国人旅行者の受入体制の整備

冒頭で述べたとおり、訪日外国人旅行者の更なる増加を図っていく上で、受入体制整備を進めることが重要であり、官民による様々な取組みが進められている。政府レベルでは、平成26年6月に観光立国推進閣僚会議が「観光立国実現に向けたアクション・プログラム2014」(以下、「アクション・プログラム」)を策定し、「外国人旅行者の受入環境整備」について明記している。具体的な施策として、「多言語対応の改善・強化」や「多言語アプリの活用」等が示されているので、抜粋して紹介させていただきたい(3)。

#### <多言語対応の改善・強化>

外国人旅行者が出来るだけ「言葉の壁」を感じることなく訪日旅行を楽しむことが出来るよう、 美術館・博物館、自然公園、観光地、道路、公共 交通機関等における多言語対応について、「観光 立国実現に向けた多言語対応の改善・強化のため のガイドライン」(平成 26 年 3 月)に従って、 全国各地で多言語対応の改善・強化を図るととも に、多言語通訳・翻訳アプリや多言語対応観光ア プリ等の活用によって、外国人旅行者のスムーズ な情報取得を促進する。

#### <多言語アプリの活用>

豊富な観光情報や地図情報等を備えた多言語 対応観光アプリの活用により、外国人旅行者のスムーズな情報取得を促進するとともに、総務省「グローバルコミュニケーション計画」に基づいて多言語通訳・翻訳アプリ技術の研究開発の強化等を行い、精度向上を図ることにより、様々な地域・場面での多言語対応への活用を促進する。

訪日外国人旅行者を取り巻く言語障壁の緩和は、 従来から受入体制整備の中心的課題の一つであり、 多言語翻訳アプリなど ICT の活用による受入環境 の向上が期待される。

受入体制整備に関する国の取組みとしては、観光 庁が「外国人観光案内所の設置 運営のあり方 指 針」(平成 24 年 1 月制定、平成 26 年 8 月改定)を 定め、外国人観光案内所網の整備の方向性を示して いる。同指針に基づき、JNTO では平成 24 年度か ら外国人観光案内所の認定制度を運用している。ま た、前述のとおり、観光庁は「観光立国実現に向け た多言語対応の改善・強化のためのガイドライン」 (平成 26 年 3 月)を策定しており、その周知が現 在進められている。

ここで、受入体制整備に関する課題を踏まえておきたい。観光庁が平成 23 年に行ったアンケート調査によれば、訪日外国人旅行者が「旅行中困ったこと」または「旅行中最も困ったこと」として、無料公衆無線 LAN 環境 (Wi-Fi 環境) が不便であるという声が最も多くなっている (「旅行中困ったこと」 36.7%; 「旅行中最も困ったこと」 23.9%)。環境整

備は徐々に進んでいるが、無料で Wi-Fi を利用できる場所が少ない、会員限定のものが多く外国人旅行者は殆ど利用できない (携帯電話回線で接続すると多額の通信料が必要)等の声が少なくない。次に多く挙げられたのは、「コミュニケーション」に関して困ったという声で、困った場所・場面としては、言語一般に関して困ったという声が回答者の約5割、交通機関や飲食施設での意思疎通に困ったとする声がそれぞれ約1割となっている。

#### 4. 機械翻訳の可能性

では、観光分野における機械翻訳の活用の可能性について考察したい。前述したとおり、「アクション・プログラム」の中で多言語通訳・翻訳アプリ等の活用や多言語通訳・翻訳アプリ技術の研究開発の強化等が謳われているが、「ジャパン・クラウド・コンソーシアム」という枠組みによる提言の中で、次のとおり機械翻訳の必要性が強調されている(4)。

機械翻訳技術はアジア諸国の観光戦略においても必要不可欠なものであり、グローバルに展開すべき日本のICT技術の一つである。

今後、こうした提言等を踏まえて、観光分野において多言語通訳・翻訳アプリ等を巡る取組みが進む見込みである。観光分野における機械翻訳技術の活用に関する国の取組みとしては、総務省が平成21年度に「地域の観光振興に貢献する自動音声翻訳技術の実証実験」を実施している。これは、自動音声翻訳技術の早期実用化を加速するとともに、外国人観光客の誘致促進による観光産業振興・地域経済活性化に貢献することを目的に、北海道、関東、中部、関西、九州の5地域で実施したものである。本実証実験で得られた大規模実験データにより旅行会話の音声認識・翻訳精度が向上し、多言語音声翻訳の実用化がこれによって加速した(5)。

本実証実験を踏まえて、2010 年に独立行政法人 情報通信研究機構 (NICT) がスマートフォン向け 音声翻訳アプリ "VoiceTra"を公開した。NICTによれば、音声翻訳技術は 1986 年に基礎研究が開始されたもので、「VoiceTra は、同技術がついに実用化に至った、大きな成果の1つ」であるという <sup>(6)</sup>。 2012 年 7 月には、"VoiceTra"で培われた技術を世界標準システムとして発展させるため、ユニバーサル音声翻訳先端研究コンソーシアム (U-STAR) が各国研究機関と連携して"VoiceTra4U"の提供を開始した <sup>(7)</sup>。日本語、英語、中国語、韓国語等 27言語間の翻訳が可能で、5 台(5 言語)までの音声チャットも利用できるという。このような多言語音声翻訳システムの有望性に関して、TOEIC 600 点レベルの翻訳能力がある点に加えて、多言語対応が容易な点等が指摘されている <sup>(8)</sup>。

こうした点から、インバウンド旅行の分野で
"VoiceTra"の活用が有望視される。前述したとお
り、東南アジアをはじめとするアジア市場からの訪
日外国人旅行者の増加が引き続き見込まれるため、
タイ語等を含む多言語対応が今後重要な課題とな
る。"VoiceTra"の翻訳率(翻訳者が評価した意味
が通じる率)は、広く利用されている多言語ソフト
ウェアに比べて概して高く、タイ語、マレー語、インドネシア語等の場合でも 80%を超えているとの
ことであり (9)、十分実用に耐えるレベルではない
かと思われる。

実際、多くの訪日外国人旅行者に対して情報提供を行っている観光案内所では、多言語対応のニーズが高まっている。JNTOが認定した外国人観光案内所(以下、「認定案内所」)を訪れた外国人旅行者の約4割は、主要東アジア(台湾、韓国、中国、香港)からの来訪者であり、来訪者数上位10カ国・地域の言語別構成比を見ると、中国語系:39.5%、英語:28.5%、韓国語:18.3%、タイ語:8.2%、フランス語:5.5%となっている(10)。 多言語対応の強化が望まれる状況であり、「タイ語、フランス語、スペイン語等のサポートをしてほしい」、「翻訳サービスを提供してほしい」、等の声が実際に寄せられている

なお、認定案内所の中には、既に機械翻訳を補助的に利用しているところが複数あり、また、認定案内所を訪れる外国人旅行者の中にも機械翻訳を活用している事例が見られる。「日本語が全く話せない外国人の来訪が多く意思疎通に苦労したものの、スマートフォンの翻訳アプリ等の利用により対応した」、「IT 系のエンジニアに翻訳アプリで対応した際、こういったアプリがあることに興味津々のようだった」等の声がきかれ、翻訳アプリ等の利便性が認知されつつある模様である。ただし、使い勝手や翻訳精度に関する問題点等を指摘する向きも見られ、観光案内の現場でより便利に利用できるよう、更なる機能向上が望まれる。

あなみに、"VoiceTra"等に用いられている統計翻訳に関する今後の課題として、1. より長文の正確な翻訳を実現するための技術、2. 文章全体での訳語の一貫性を実現する文脈処理技術の創出、の2点が指摘されている(II)。また、機械翻訳自体の機能向上とは別の問題であるが、訪日外国人旅行者が簡便に利用できるWi-Fi環境の整備も重要な課題と言える。なお、翻訳の高精度化等の課題への対応がNICTの第三期中期計画に示されており、「観光分野において実利用に供することを可能とすることを目標」にする旨が明記されている(I2)。より長い話し言葉や文章の翻訳、周辺の文や段落も考慮して翻訳する技術が実現すれば、観光分野におけるさらなる利用促進が見込まれるため、そうした改善を大いに期待したい。

#### 5. おわりに

以上、本稿で述べたとおり、訪日外国人旅行者の 受入環境の整備、とりわけ多言語対応の改善・強化 の課題に取組んでいく上で、機械翻訳技術の活用は 大変有効な手段の一つと考えられる。訪日外国人旅 行者が出発前に多言語翻訳アプリをダウンロード しておけば、旅行中のコミュニケーションに関する 不安が大きく軽減する。また、受け入れる側でも、 音声翻訳アプリや自動翻訳ツールがあれば、コミュ ニケーションに関する不安が軽減する。外国人旅行者とのコミュニケーションに関して、最近、次のような指摘が行われている<sup>(13)</sup>。

自動翻訳ツール等の開発・普及が進めば、将来 的には外国人旅行者とのコミュニケーションも 大きく改善するが、訪問国の印象はそこで出会っ た人々との生の会話により形成される部分が極 めて大きい。我々国民の一人一人が外国人旅行者 をもてなす気持ちを持ち、それを一言でも相手に 通じる言語で表現できれば、世界各国からの訪問 者のわが国に対する印象も変わるはずである。

音声翻訳アプリや自動翻訳ツールを利用することで、訪日外国人旅行者へおもてなしの気持ちを伝えることに対する心理的ブレーキがかかり難くなると思われる。インバウンド旅行の分野における機械翻訳の活用に関して、関係各位と連携して取組んでいきたい。

#### (注)

(1) 1964 年以降の訪日外国人旅行者数の推移、VJC 開始以降の主要市場別訪日外国人旅行者数の推移 等については、JNTO の統計資料をご参照いただき たい(URL は以下のとおり)。

http://www.jnto.go.jp/jpn/reference/tourism\_data/ pdf/marketingdata\_outbound.pdf

http://www.jnto.go.jp/jpn/reference/tourism\_data/ pdf/pdf/marketingdata\_tourists\_after\_vj.pdf

- (2) 観光庁『訪日外国人の消費動向 訪日外国人消費動向調査結果及び分析 平成 24 年年次報告書 (平成 25 年 3 月)』、19 ページ。
- (3) 観光立国推進閣僚会議『観光立国実現に向けた アクション・プログラム 2014 - 「訪日外国人 2000 万人時代」に向けて- (平成 26 年 6 月 17 日)』、 23-24 ページ。

- (4) 観光クラウド・ワーキング・グループ『観光立 国日本を推進するための情報通信技術利活用に関 する中間提言 (2012 年 6 月 25 日)』、12 ページ。
- (5) 総務省「多言語音声翻訳システムの紹介(平成 26年4月18日)」(総務省 情報通信国際戦略局 技 術政策課 研究推進室 資料)、4ページ。
- (6) 隅田 英一郎・柏岡 秀紀「世界中の人々が母国 語で外国人と対話できる多言語音声翻訳技術」独立 行政法人情報通信研究機構『情報通信の未来をつく る研究者たち』、143ページ。
- (7) 総務省、前掲資料、7ページ。
- (8) 隅田英一郎「機械翻訳のいま 統計的手法を中 心に」『情報管理』第 57 巻第 1 号 (2014 年)、15 ページ。
- (9) 隅田、前掲論文、15ページ(図4)。
- (10) ただし、実績を把握できていない案内所もあるため、この数値はあくまでも参考値である。また、複数言語が用いられているシンガポールに関しては、便宜上、使用言語を英語に統一した。
- (11) 隅田、前掲論文、19ページ。
- (12) 独立行政法人情報通信研究機構『独立行政法 人情報通信研究機構が達成すべき業務運営に関す る目標を達成するための計画(第3期)』、32 ページ。
- (13) 日本経済団体連合会『高いレベルの観光立国 実現に向けた提言 (2014年6月11日)』、13ページ。

# 機械翻訳を使いこなす ~品質と生産性向上のために~

#### 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

かつ低コストに翻訳する必要性から、産業翻訳の現場でも機械翻訳の可能性に注目が集まっている。 国内で注目されるのは、日本語と英語間の翻訳における実用性だが、言語構造などの違いから実用面での課題が多いという声が大きい。しかし文法構造や語彙に共通性のある欧州言語間などでは、機械翻訳は実用レベルに達していると言われ、海外ではすでに導入が進んでいる。こうした海外での機械翻訳の普及状況を考えると、競争の激化するグローバル市場で生き残っていくためには、日本の翻訳市場でも機械翻訳を「使うかどうか」ではなく、「いかに使いこなすか」ということが今後は重要な鍵となってくると考えられる。

近年、急速に増え続けるコンテンツをスピーディー

もちろん、機械翻訳エンジンをただ使用するだけでは、マニュアルなどの翻訳に使用できる十分な品質を得ることは困難な場合が多い。しかし、ポイントをおさえたワークフローを構築することで、マニュアル翻訳においても、人による翻訳と同じレベルの品質を保ちながら、コスト削減と翻訳期間の短縮を図ることは十分に可能である。

#### ■機械翻訳を使いこなすポイント

機械翻訳を使って、「人と同じレベル」の品質を保 ちながら、コスト削減や翻訳期間短縮を実現するた めには、次の3つのポイントをおさえてワークフロ ーを構築することが重要である。

- ① 品質基準の設定
- ② 翻訳品質向上のためのプリエディット
- ③ 最適な機械翻訳エンジンの選択 ここからは、それぞれのポイントについて詳述して いく。

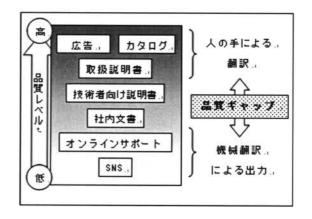
#### ① 品質基準の設定

1つ目のポイントは「品質基準の設定」である。機 械翻訳のワークフローを構築するにあたって、まず 重要となるのが「翻訳対象のコンテンツに期待する 品質」と「機械翻訳によるアウトプットの品質」と いう2点を把握しておくことである。

なぜこの 2 点を把握しておくことが重要かというと、この 2 点の品質の差が機械翻訳のワークフローで埋めなければならない「品質ギャップ」となるからである。そして、この「品質ギャップ」の大小により、機械翻訳処理の前後のワークフローでどういったタスクを行わなければならないかということが決まってくる。

もちろん、翻訳対象のコンテンツによっても目標と する品質基準は様々なため、品質ギャップの大きさ も異なってくる。

以下の図は、コンテンツの種類による目標の翻訳品質レベルと品質ギャップをまとめたものである。



上の図の場合、「オンラインサポート」は、目標とするのが「書かれている内容が正確に伝われば良い」という品質であり、機械翻訳によるアウトプットとの品質ギャップもあまりないため、機械翻訳処理前後のタスクというのはあまり必要がないと言える。

一方、「取扱説明書」は、「人と同じレベル」の「わかりやすく、正確な」品質が求められるため、品質ギャップを埋めるために「プリエディット」や「ポストエディット」といったタスクを行う必要がある。しかし、この品質ギャップを埋めるために、通常の人による翻訳よりも工数がかかってしまっては、機械翻訳を導入する意味がないということになる。

そのため、機械翻訳を導入する前に、必ず「品質ギャップ」を埋めるために、どういったタスクが必要で、どれくらい工数がかかるのかということを把握しておくことが重要である。

品質ギャップを埋めるための方法として、代表的な 作業が「ポストエディット」である。

「ポストエディット」は、機械翻訳エンジンにかけた後の文章を、求める品質に応じてリライトし、仕上げていく作業である。ポストエディットについても、目標とする品質に合わせて、以下のような三段階のレベルを使い分けることが多い。

#### 1) ラピッドエディット

原文の意味が正確に訳文に反映されているか、という観点で修正する。訳文の読みやすさは考慮せず、 読みづらくても意味さえあっていれば修正は行わない。

#### 2) ミニマムエディット

意味の正確さだけでなく、読みやすさという観点で も修正する。社内研修用のマニュアルや、現場の技 術者が参照程度に見るものであれば、このレベルで もよい可能性がある。作業負荷としては人の手による翻訳の50%~80%と想定される。

#### 3) フルエディット

意味の正確さと読みやすさに加えて、用字用語の統一やスタイルの統一という観点でも修正する。ポストエディットとしては作業負荷が一番高く、人の手による翻訳の70%~100%と想定される。

一般消費者に渡る製品のマニュアルや、操作を誤る と使用者に不利益を与えかねない製品のマニュアル については、このレベルでのポストエディットが求 められるだろう。

#### ② 翻訳品質向上のためのプリエディット

2 つ目のポイントは、「翻訳品質向上のためのプリエディット」である。「プリエディット」とは、機械翻訳にかける前に、機械翻訳エンジンが理解しやすいように原文文書をリライトする作業である。特に英語から複数言語に多言語翻訳するような場合は、原文の品質が悪いと、その品質が翻訳言語にも影響するため、翻訳する言語数の分だけ、機械翻訳処理後のポストエディットの工数が増加することになる。そのため、機械翻訳にかける前の英文を機械翻訳に適した形にリライトしておくことが工数削減に効果的である。

プリエディットの代表的なルールとしては、次のようなものが挙げられる。

- 短文化でシンプルにする(1 文は 15 ワード以下にする、など)
- 一文一義を原則とする
- 表記統一を行う

なお、プリエディットは「機械翻訳のための中間生成物を作る」という目的で行うこともあるが、より効率的に行うためには、「機械翻訳にかけやすい文体にしつつ、原文文書をブラッシュアップしていく作業」であることがより望ましい。つまり、プリエディット後の文章は、原文のマニュアルとしてそのま

ま使用される品質を目指すということである。そう することにより、そのマニュアルに将来改訂が入っ た場合にも、改めてプリエディットを行うことな く、そのまま機械翻訳にかけていくことができる。

#### ③ 最適な機械翻訳エンジンの選択

3 つ目のポイントは、「最適な機械翻訳エンジンの 選択」ある。

翻訳エンジンの種類としては、「統計ベース」、「ルールベース」、そしてその両方を組み合わせた「ハイブリッド」といったものがあるが、エンジンごとに得意な言語や分野は異なる。また、導入・運用コストや対応するファイル形式、ユーザビリティもエンジンによって異なるため、それぞれの翻訳エンジンのメリット・デメリットをよく比較した上で、プロジェクトに合った最適な翻訳エンジンを選択することが重要である。

#### ■プロジェクト事例

では、これまで述べてきた3つのポイントをおさえ て機械翻訳を使いこなすことによって、品質、コスト、納期に対してどのような影響があるのか。英語 からフランス語、イタリア語、ドイツ語、スペイン 語への翻訳プロジェクトの事例を紹介する。 プロジェクトの概要は以下の通りである。

ACCESS TO BE SEE THE

#### [翻訳対象]

エンドユーザー向け製品の取扱説明書 [ボリューム]

75,000 ワード (約370ページ)

[コーパス (統計ベース使用時)]

約45,000分節(約500,000ワード)

機械翻訳の品質の評価については、各言語のネイティブ翻訳者 2 名が、分節ごとに "Excellent"、"Good"、Medium"、"Poor"という四段階で分類した。 各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	

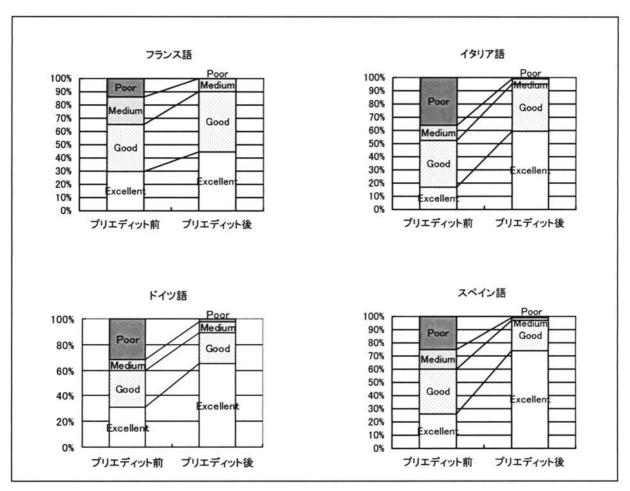
#### ① 品質基準の設定

今回のプロジェクトは翻訳対象がエンドユーザー 向けの取扱説明書のため、目標とする品質基準は 「人による翻訳と同じレベル」に設定した。ポスト エディットについても「フルエディット」で、「正 確さ」、「わかりやすさ」に加えて、「用字・用語、 表現の統一」という観点でも修正を実施した。

② 翻訳品質向上のためのプリエディット 翻訳対象の英文に、「長文が多くわかりにくい」な どの問題が見られたため、プリエディットを実施し た。

次に示すグラフは、統計ベースエンジンを使用した際に、プリエディットを行う前と行った後の翻訳品質の変化を示したものである。

全言語平均すると、"Excellent"および"Good"の全体に占めるパーセンテージは、プリエディットを行うことで「平均29ポイント」も上昇した。こうした結果からも、プリエディットにより翻訳品質を向上し、ポストエディットにかかる負荷を軽減できることがわかる。



#### ③最適な機械翻訳エンジンの選定

次に、今回のプロジェクトで使用する最適なエンジンについて、価格面や普及の度合いなどから見て、 比較的導入しやすいと思われる以下の3種類のエンジンの中から選定した。

#### エンジンA:

海外でも有名なルールベースエンジン エンジン B:

Moses ベースの統計ベースエンジン エンジン C:

海外企業が独自に開発した統計ベースエンジ ン

統計ベースのエンジンについては、コーパスを読み 込ませた上で翻訳を行った。各エンジンのアウトプットを評価した結果、"Excellent"および"Good"と評 価された分節の全体に対するパーセンテージをエ ンジン別にまとめたのが、次の表である。

	エンジャンA	エンジ*ソB 統計ペ*-ス	エンジ*ン C 統計へ*-ス②
フランス	65%	90% (11.5%)	75% (83.5%)
イタリア	76%	95.5%(13.5%)	96.7% (77%)
ドイツ	61.6%	89% (11%)	96% (62%)
スペイン	68.5%	97% (19%)	68% (84%)

※ 括弧内の数字は、統計ベースエンジンにおいてコーパスを読み込んだ後の結果

今回のプロジェクトでは、同製品群のコーパスがあったため、それを利用することで、ルールベースよりも統計ベースでのアウトプットの方が品質が良くなった。ただし、統計ベースのトレーニング前の結果と比べると、ルールベースの方が品質が良い言語もあるため、大量のコーパスが準備できないという場合はまずはルールベースで機械翻訳を導入するというのも選択肢の一つとして考えられる。

今回のプロジェクトでは、全言語で平均して品質が 最も良かった「エンジンB」の統計ベースエンジン をプロジェクトに使用することにした。

#### ●コスト削減・翻訳期間短縮効果

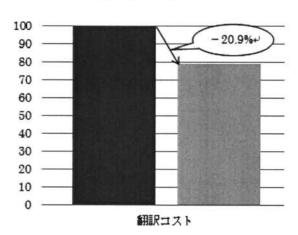
ここまで説明してきたとおり、以下の3つのポイントを適用したワークフローを構築して機械翻訳を 実施した。

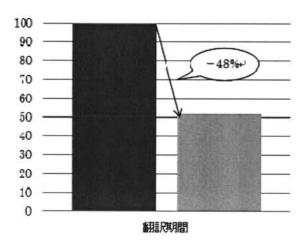
①品質基準を「人による翻訳と同じレベル」に設定②英文のプリエディットを実施

③3 つのエンジンの品質比較の上で、最適なエンジンを選定

その場合の翻訳コストと翻訳期間について、人による「従来翻訳」におけるコストと翻訳期間を基準と して比較したのが、以下のグラフである。

#### ●翻訳コスト・翻訳期間の比較





機械翻訳を使用することで、「従来翻訳」に比べて、 翻訳コストについては「20.9%」、翻訳期間につい ては「48%」も短縮することができた。

#### ■まとめ

プロジェクト事例からもわかるように、ポイントを おさえたワークフローを構築することで、機械翻訳 を使用しても、「人と同じレベル」に品質は維持し つ、コスト削減と翻訳期間の短縮を実現することは 十分に可能である。

その際に重要になるのは、翻訳エンジン、分野、言語などの特徴を適切に把握し、よく理解して、機械翻訳を使いこなすことである。機械翻訳を導入する際には、導入のコストだけでなく、様々な検討事項について分析し、知識に基づいて正しく検証を行い、最適なワークフローを構築しなければならない。

国内でも機械翻訳に注目が集まっているが、今後、 機械翻訳をより普及させるためには、機械翻訳を使 いこなすための知識やノウハウを日本市場により 広めていく必要がある。

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、日本企業に向けた 機械翻訳導入やプロセス構築のコンサルティング を行う。多言語翻訳を左右する原文品質を重要視 し、和文ライティング工程も担当し、機械翻訳に適 したマニュアル作成のコンサルティングも実施。

#### Project Report

# 豊橋技術科学大学における機械翻訳の研究開発

#### 井佐原均

#### 豊橋技術科学大学

#### 1. はじめに

国立大学法人豊橋技術科学大学では、「学内研究基盤強化による産業分野向け高度機械翻訳技術の研究開発」に対して文部科学省から運営費交付金特別経費の交付を受け、平成26年度に多言語翻訳研究本部を設置し、プロジェクトを開始した。本稿では、このプロジェクトの概要を述べる。

#### 2. 背景及び目標

Globalization の高まりの中、企業の海外展開、製品輸出、知財貿易など、様々な場面で産業文書の多言語化のニーズが高まっている。また、第2のデジタルデバイドとして、インターネット上に蓄積された大量の英語情報にアクセスできるかどうかが、生活の質に直結する環境となりつつある。翻訳プロセスの効率向上は必須であり、機械翻訳等のICT技術の活用が求められる。

我が国の企業の国際競争力強化の観点からは、産業に関する情報を表現する日本語をこれまで以上に、迅速かつ正確に翻訳することが重要である。例えばアジア諸国、アフリカ諸国へプラント輸出を行う場合は、現地の技術者にも理解可能な現地語(母語)でのマニュアルも必要となり、翻訳を迅速にかつ正確に行わなければ、製品の技術力以外の部分で国際競争力が低下してしまう。現地語でのマニュアル作成により、海外の機関等が、日本のものづくり産業がこれまで培ってきた高度な技術(日本ブラ

ンド)を正確に活用することが容易になり、 日本ブランドに対する需要が増え、生産性 の向上にも寄与することが可能となる。

本プロジェクトでは、これまでの研究成果を活かし、文理融合を促進することにより、大学のイノベーション(技術革新)基盤を強化する。本プロジェクトの目標は機械翻訳システムのエンジンの開発・改良ではなく、機械翻訳をツールとして活用し、情報発信における翻訳過程を効率化・高品質化できるようにユーザを支援する枠組みを構築することである。我が国が遅れているサービス科学の研究として、産業分野の各種マニュアル等の翻訳を短時間かつ正確に行う高度な機械翻訳機能の技術開発・運用を行うことを目指す。

#### 3. 情報発信と機械翻訳

情報伝達には、多言語情報検索から翻訳につながる Google 型の情報受信と、ビジネス文書の多言語化といった情報発信とが考えられる。情報発信型の場合、著者の母語から非母語への翻訳となり、著者には翻訳文の良否が判断できないことがある。また日本からの情報発信を考えた場合、日本語から他の言語への翻訳となるが、日本語の解析の難しさから、日本語を原文とする機械翻訳の品質は、他と比べて劣る。このように日本語からの多言語情報発信には困難が伴うが、その一方で、情報発信型翻訳の有利な点として、翻訳対象となる文章をコントロールすることが可能であることが挙

げられる。企業にとってみれば、自社内で 社員教育を行うことにより原文作成時に文 体等を制約することが可能であること、翻 訳対象が限定されているので、事前に対訳 辞書を整備できることが挙げられる。

#### 4. 情報発信を支える機械翻訳

製品の取扱説明書や企業活動のノウハウ 文書等を機械翻訳システムを活用して翻訳 する場合、訳文の精度を向上させるチャン スは翻訳前、翻訳中、翻訳後と 3 か所にあ る(図 1)。

翻訳前には、書き手の協力を得て、入力 文をコントロールすることが可能である。 入力文への制約(制限言語)は、機械翻訳 の出力を向上させるとともに、人間の理解 容易性を低下させないように設計するべき である。このような制約に基づいて文書作 成を行うことにより、国内向けに日本語で 作成した文書を同時に機械翻訳の入力文と して利用できる。筆者らは既に自動車関連 企業での実証を行い、ISO での国際標準化 を進めている。 翻訳中、つまり機械翻訳システムの出力の理解容易度を直接向上する手段として、対訳用語辞書の充実が挙げられる。重要な語句が適切に翻訳されているかどうかは、読み手の評価に直結する。ここでは重要な語句を取り出す技術、その翻訳を定める技術、人手による確認ツールが必要となろう。筆者らは自動車や楽器のマニュアルから、翻訳に有効な長い語句(単語や複合語だけではなく、意味のあるひとまとまりの語句)を取出す技術の開発を行った。

翻訳後は、後編集と呼ばれる過程となる。 プロの翻訳者による後編集はコストが高い が、翻訳者ではないが、その分野の知識が あり、出力言語を母語とする人を活用する ことにより、安価な後編集が可能となろう。 豊橋技術科学大学では大学の英語ウェブペ ージに機械翻訳システムを設置し、留学生 によるクラウドソーシング後編集の実証を 行った。

これらを踏まえて、本プロジェクトでは 人間に理解しやすく、かつコンピュータが 翻訳しやすい制限言語の開発を行い、さら に制限言語の国際標準化を進める。機械翻

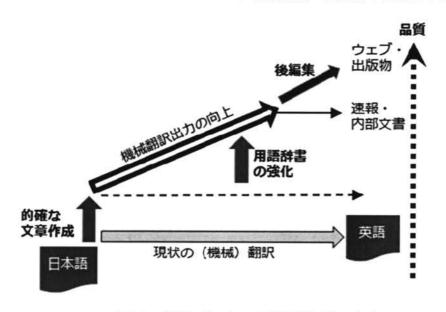


図1:翻訳プロセスと翻訳品質の向上

訳で用いる対訳語句集を比較的容易に構築できるよう、半自動の用語句抽出システムと訳語選択システムの開発を行う。そして知識と意欲を持つボランティアによる後編集の枠組みを構築する。

以上まとめると、制限言語に基づく入力 文作成と、専門用語および句の辞書を用い た機械翻訳により、高精度の翻訳を実現し、 後編集の負担を軽減する。スポーツや趣味 などの翻訳の場合にはボランティアベース の後編集も可能となる。

### 5. コンソーシアム

機械翻訳システムを活用した翻訳プロセスの構築に、その分野の知識を持っていれば誰でも寄与できるという視点は企業から世界への情報発信を促進する足掛かりとなろう。このような考えの下、多言語化したい文書を持つユーザ企業と、翻訳プロセスに関わる企業とを結ぶ多言語情報発信コンソーシアムの構築を目指している。ここでは海外の翻訳用データ共有組織との連携を予定している(図 2)。

このコンソーシアムの下、豊橋技術科学 大学は翻訳(個人あるいは会社)と文書作 成(ユーザ企業)とが連携して翻訳プロセ スの効率向上を図ることを支援する。ユー ザ企業の持つ対訳データなどの資源を共有 し、さらにはそのような資源から対訳用語 集などの翻訳用データを抽出することを検 討しており、このコンソーシアムの参加者 を募集する予定である。

この活動を通して、翻訳会社とユーザ企業との連携が双方の利益につながることを示すことが本プロジェクトの目的である。 具体的な成功事例を示すことで、より多くの翻訳会社や企業が参加してくれることになろう。データ共有の利益についての国内企業の理解が深まることを期待している。

#### 6. 国際標準化への動き

本稿で述べた翻訳支援に関連する国際標 準化の動きとしては、ISO TC37 におい て、以下の標準化が進められている。制限 言語に関しては、SC4 において、24620-1 Language resource management - Controlled natural language -- Part 1: Basic concepts and general principles が標準化 に向けて進行中である。筆者は、このプロ ジェクトリーダーを務めている。辞書に関 しては、SC3 において、ISO 30042:2008 Systems to manage terminology, knowledge and content -- TermBase eXchange (TBX) が国際標準化されている。後編集に 関しては、SC5 において、18587 Translation services -- Requirements for machine translation (MT) and post edition levels が標準化に向けて進行中である。

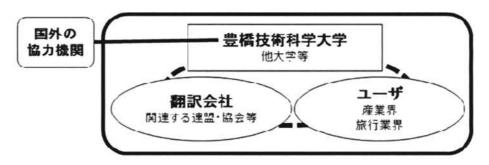


図2:多言語情報発信支援コンソーシアム

# 機械翻訳事始め[W]

#### 坂本義行

#### アジア太平洋機械翻訳協会理事

機械翻訳の研究で、最も衝撃的な出来事は、1960年代に発表された ALPAC レポートの発表でした。今回はこの内容について、報告したいと思います。

#### 1. ALPAC レポートの発表

米 国 National Research Council, National Academy of Sciences of Division of Behavioral Sciences に設置された ALPAC (Automatic Language Processing Advisory Committee)が、1966年米国政府において翻訳を機械化する目的で推進されていた機械翻訳開発への研究費援助の継続に関して審議した結果による勧告である。1)

ALPAC レポートは、米国政府機関における翻訳および機械翻訳に関して広範囲な調査・検討を行いその結果を 15 個の項目にまとめている。 さらに、各項目に関する詳細な考察およびデータを 19 個の項目にまとめて付録として付加している。

#### 2. ALPAC レポートの詳細

このレポートについて、日本電子工業振 興協会機械翻訳システム調査専門委員会が 1989年度に行った調査に基づいた報告書が 出ている。<sup>2)</sup>

この報告書によると、

#### a) 人間翻訳 (Human Translation)

<主張>

翻訳者の選択、環境の整備、翻訳の方 法の工夫により、翻訳作業の効率向上が 図れる。

#### <概要>

人間が行っている翻訳作業について 分析すると、次の点が指摘できる。

- 1) 人が翻訳する場合、よい翻訳をするには、①目標言語についての十分な(good) 知識、②翻訳対象についての理解能力、③原言語についての適当な(adequate)知識、を有する人がおこなうのがよい。
- 2) 人が翻訳する場合、よい翻訳を得る には、よい辞書と参考文献が必要で ある。
- 3) 人が翻訳する場合、ツールとしては、口述録音機 (dictating machine) やタイプライタを使うが、これらでは、図、式などを挿入した最終稿が出力できず、更に別の工程が必要である。それでも費用は翻訳に一番かかる
- 4) Rapid dictated 翻訳(意味をつかんで同時通訳的に目標言語で表現する)と full 翻訳との比較実験において、前者は品質がほぼ同程度で時間的には 1/4 で済む。

# b) 翻訳者の雇用形態(Types of Translator Employment)

<主張>

翻訳者の雇用形態は、当時存在した形態のどちらも、一長一短で雇用者側が満足できるものではない。

#### <概要>

翻訳者の雇用形態は、IN-HOUSE (社

内 で か か え る 正 社 員 翻 訳 者 ) と **CONTRACT** (フリーランサー)、の **2** 種類である。

IN-HOUSE (正社員翻訳者)
 翻訳側からみた利点:
 正社員として保証されている。
 雇用側からみた利点・欠点

#### 利点:

- ① 必要時にその場で翻訳して貰える
- ② 依頼者/翻訳者間のよりよい協力体制
- ③ 必要時に迅速なサービスが受けられる
- ④ 保持が容易

#### 欠点:

- ⑤コストがかかる
- ⑥ワークロード調整が時に困難 ⑦全ての分野に精通した翻訳家を 揃えることはできないので、一貫 した良い翻訳は得られにくい
- 2) CONTRACT (フリーランス翻訳者)翻訳側からみた利点:

自由に仕事を選ぶことができる。 希少言語で広い分野がカバーでき れば収入も高い。

発注側からみた利点・欠点 利点:

- ①その場ですぐ翻訳を依頼できない
- ②機密保持が困難
- c) 科学言語としての英語 (English as the Language of Science)

<主張>

英語母国語者にとって翻訳の需要は それほど大きくない、

#### <概要>

英語で書かれた学術論文が圧倒的に 多いため、アメリカを始めとする英語圏 科学者が母国語以外の言語で書かれた 論文を読んだり翻訳したりする必要性 は他の母国語より低い。

d) 科学者がロシア語を学習するために必要な時間 (Time Required for Scientists to learn English)

#### <主張>

ALPAC の関心事は、ソ連の科学技術 文献を効率よく理解し、翻訳することに ある。ロシア語翻訳を頻繁に行う者は、 元の文書を読めるようにするほうが簡 単で経済的である。政府職員向けには既 にロシア語科学文献を読むための教育 コースがあるので、これを活用すべきで ある。

#### <概要>

和LPAC は、ある場合には、ロシア語翻訳をよく利用する者が文書を原語のまま読めるように学習した方がより簡単で経済的であると確信する。ある論文は、科学者が2時間クラスを8ないし16回受講する間に、関心のあるロシア語文献を判別できるようになる、と指摘している。数式、図表、グラフから必要事項を抜き出すことができる場合もある。また、関心のある文献について口頭の部分訳で十分な場合もある。このように、科学技術系の読者が専門の外国語の知識しか必要としないことは、一般に認められていることである。

事実、いくつかのよく知られた研究によって、科学者は200時間以内に専門分野のロシア語資料を読むための知識を得られることが指摘されている。また、その程度の知識を持ち合わせている科学技術者は増え続けている。

ロシア語テキストを読むための政府

職員養成のための施設が既に存在しているが、これまで大々的に利用されることはなかった。国防言語研究所の西海岸分室では、この目的のために2種類(6週間・10週間)の教育コースとテキストを開発した。ALPACは既に、国防言語研究所より生徒の登録を歓迎する旨の連絡を受けている。

e) 米国政府における翻訳の現状 (Translation in the United States Government)

#### <主張>

JPRS (Joint Publications Research Service) による人間翻訳と空軍の機械翻訳システムを比較すると、コスト、所要時間、翻訳の質のいずれにおいても前者が優れている。また、JPRS は翻訳者を十分に確保しており、翻訳の増加にも十分対処できる。従って、政府機関における翻訳のために機械翻訳システムの利用を拡大することは不要である。

#### <概要>

米国政府における翻訳は、①業者への 発注、②部内での翻訳、③JPRS(共同 刊行物調査事業)の利用など種々の方法 で行われている。④一部の機関では、 Public Law 480 (Agricultural Trade Development and Assistance Act 農産 物通商開発援助法)の見返り資金による 翻訳を利用している。また、主として空 軍で、⑤機械翻訳システム(後編集付き) の利用が行われている。さらに、⑥NSF (National Science Foundation) の援

f) 米国政府における翻訳者の数 (Number of Government Translators)

助による雑誌の翻訳も行われている。

#### <概要>

政府内部の翻訳者の正確な数は不明。

政府内での職制分類の「翻訳者」の地位が低いため職種として「翻訳者」が敬遠されている。

# g) 翻訳のための予算 (Amount of Spent for Translation)

#### <主張>

政府が翻訳に使っている金額は非常に少ない。

#### <概要>

翻訳を安全にするために用いられているいろいろな方法を考えると、軍関係が翻訳にさまざまな異なった代価 (1,000 語あたり 9 ドル~66 ドル)を支払っていることは驚くべきことではない。

最初の会合で、翻訳にかかる金額の見 積もりがあれば役に立つであろうとい うことになった。以下がその見積もりで ある。

政府機関が翻訳に使った金額は、(1963 ~65 会計年度で、13.07 百万ドル)

これからわかるように、政府の研究、 開発を援助している多くの事業と比べ ると、政府が翻訳に使っている金額は非 常に少ない。

商業的翻訳は、アメリカ翻訳協会の理事によれば、約750万ドルである。他に政府以外のインハウスの翻訳者によるものが約200万ドルと見られる。従って翻訳に使われる金額の見積もりは、合計で約2,200万ドルというところであろう。

h) 翻訳者あるいは翻訳の供給量は不足して いるか (Is There a Shortage of Translator or Translations?)

### <主張>

翻訳者あるいは翻訳の供給量は不足 していない。したがって、機械翻訳の研 究が必要であるという議論には結びつ かない。

#### <概要>

ALPAC 委員会は、翻訳者あるいは翻訳の供給量は、多言語から英語への翻訳で見る限り、不足していないという結論に達した。これは以下のデータによる。

- 1) 翻訳者への報酬は 1,000 語あたり 6 ドルという安い金額であること、および翻訳者たちへのインタビューの結果。
- JPRS (Joint Publications Research Service) は、約 4,000 人の翻訳者と 契約しているが、月平均ではわずか 300 人たらずが仕事をしているにす ぎない。
- 3) NSF (National Science Foundation) で翻訳している雑誌の種類と量。

1964 年度では 30 種類の雑誌を、一誌あたりたった 200 コピー配布しただけであった。また、米国議会の下院に設けられた委員会により、主に政府系の 45 機関に対する調査が行われた。回答のあった 25 件のうち、翻訳者は不足していないとしているのは 7 件で、不足しているとしているのは 9 件であった。しかし、その 9 件も原因は翻訳者の待遇が良くないせいだという意見が大部分であった。

i) 翻訳された文書へのアクセス (Regarding a Possible Excess of Translation)

#### <主張>

現在、翻訳は、ユーザーの声に関係なく行われており、ユーザーの実際に必要とする文献の翻訳に、支障をきたしている。これに対する解決策としては、現在のところ、ユーザーの要求する文献を如

何に素早く、高品質で、経済的に、供給するかを検討すべきである。

#### <概要>

委員会の見解として、翻訳の不足の心配はなかったが、翻訳の過多の可能性の心配は、確かに存在したと述べている。ハッキリとした読者の存在しない文献を翻訳するのは、無駄であるばかりでなく、翻訳や情報の流れを詰まらせる。片端から翻訳を行うのは、確実に観覧される文献に限り、更に翻訳を追加する場合は、特に要求のあった場合に限るべきである。ユーザーの声として、機械的に翻訳を行うよりも、実際にユーザーの要求のある追加の翻訳を急いで欲しいと云う声もあった。

また、機械的に翻訳を行った場合、原稿の質や、新しさから、必ずしも、その有用度が高いとは言えない。具体的には、ソ連の雑誌から取った論文を翻訳した場合、アメリカの雑誌に採用できるものが、3分の1以下になる場合もあった。

更に、この不要な翻訳量の多さより、 本当に必要と読者が思い、要求している 論文は、かなりの時間を経過しないと、 翻訳が手に入らないのが現状である。

これに対する対策として、

- 1) 文献の要求のみの翻訳の作成
- 2) ある分野全体の現状の報告の作成
- 3) 個々のユーザーの要求に敏速に対 応する仕組み作り

といった対応策を挙げる者もいるが、委 員会としては、

- 1) 全ての技術論文を翻訳するのは価値がないが、
- 2) どの文献が字際に必要かを予測しようとしても無駄であるので、

現在は、要求される文献を如何に素早く、

高品質で、経済的に供給するかに集中する べきであるというのが、委員会の結論であ る。

# j) 翻訳に関する重要な問題点 (The Crucial Problems of Translation)

#### <概要>

- 1) 品質
  - ・翻訳の品質は、requester の要求に 十分耐えうるものでなければなら ないが、多くの場合十分な品質の翻 訳がなされていない。
  - ・翻訳の品質の評価実験
  - · 評 価 尺 度 intelligibility informativeness
  - ・人間訳すべてが、機械訳より優れて いた。
  - ・概して、機械訳は、最も良い訳から 最も悪い訳までの評価点のほぼ中 間に位置していた。
  - ・翻訳の品質とコストの関係は、精密 には分からない。

#### 2) 速度

- ・翻訳の lag time はまだ改良の余地が あるが、現状は以下のようになって いる。
- ・平均的な文書に対する lag time FTD の Automatic Language Translator 109 日 (AN/GSQ-16(XW-2))

外部のサービス会社 (JPRS) 65 日

・lag time の中には、当然、翻訳その ものだけでなく、ポストエディトや 出版等にかかる日数も含まれてい るが、FTD の機械翻訳にはさらにマ シーンにかけるための時間がかか っている。

#### 3) コスト

・機械翻訳は、高品質を要求される外

交文書や文学の翻訳には向かない が、科学文献等の翻訳を支援できる であろう。

- ・政府がどのような翻訳形態を採用するかは、結局はコストによる。
- ・翻訳コストの算出が示されている。
- ・米国においては、露英翻訳が中心で あり、この先もそれが変わる見通し はない。
- ・露語の文献を現状の頻度で用いるならが、人材の露語教育に力を入れた 方が良い。

# k) 機械翻訳の現状 (The present State of Machine Translation)

#### <主張>

機械翻訳は、後編集なしでは役に立た ず、後編集をすると人手による翻訳より 非経済的になる。

#### <概要>

1) 翻訳が非常に限られている。

簡単でかつ選別された文章の初期 の機械翻訳 (10 年前のジョージタウン・IBM の実験) は励ましになった が、一般科学技術文献の機械翻訳 (宇宙生物学に関するロシア語記 事の4つの異なる MT システムの翻 訳例) は落胆的である。

2) 品質が悪く、後編集なしでは役に立たない。

科学技術文献の生の機械出力は、 ほぼ判読できるが、時々誤解をまね き、また間違いもある。そして、読 むのが遅く苦痛になる。

FDT 機械翻訳システムと人間が翻訳した、物理学及び地学のロシア語文書の読む速さ、正確さ、理解度の比較で、MTの生出力は10~29%劣り、後編集された出力でも3~23%

劣っていた。

問題の多くは、構文と語順の不自 然さにある。

3) 後編集をすると人手による翻訳より非経済的になる。

現在の機械翻訳は後編集なしでは 役に立たず、後編集をすると人手に よる翻訳より非経済的となる。

- 4)機械翻訳のレベルを高めるには機械が「理解」することが必要であるが、技術的な見通がない。
- 5) 機械翻訳は現在それほど必要でないので、そのような研究の必要性を 感じない。
- 1) Mannheim および Luxembourg における機械援助の翻訳 (Machine-Aided Translation at Mannheim and Luxembourg)

#### <主張>

専門用語集などの翻訳支援システムは翻訳の生産性向上に役立っている。一方、後編集のコストなどの評価はできなかったが、後編集付の機械翻訳の質は通常の人間翻訳よりも質が悪いという結果になった。

#### <概要>

当面は完全な自動翻訳の実現が難しいことから、翻訳支援システムへの関心が高まっている。ここでは翻訳支援システムに関係する事項を紹介している。

1) 翻訳支援システム例

The Federal Armed Forces
Translation Agency
(MANNHEIM, GERMANY) での
実験では、適切な専門用語集を用い
ることにより翻訳の生産性が向上
することが示された。

英語の技術文献をドイツ語に翻訳

するのに、以下の2つの方法、

- ① 一般の辞書、参考文献を参照した り専門家の意見を聞いて翻訳を する従来の方法
- ② テキストに出現する専門用語集 (対訳付きで出現順にリストさ れたもの)だけを参照して翻訳を する方法

を比較しており、後者の方が5割 以上も効率が良く、翻訳ミスが3分 の1になっている。

#### 2) 後編集

通常の(人による)翻訳と、機械翻訳の出力を後編集する方法との比較実験を行っている(ロシアがから英語へ翻訳/被験者23名)。

後編集の難しさやコストについては 結論を出せなかったとしているが、実験 結果は以下のようになる。

- ① 後編集の難しさに関しては意見が分かれた。興味のある点は、翻訳速度の速い翻訳者の方が、後編集は難しいと感じている点である。
- ② 翻訳者の3分の1は、後編集する ために原文が必要であるとし、他 の翻訳者の多くはどちらとも云 えないとしている。
- ③ 機械翻訳の後編集は、翻訳速度が 速い翻訳者にとっては妨げとな り、翻訳速度が遅い翻訳者にとっ ては手助けになる傾向がある。
- ④ 通常の翻訳の速度と後編集の速度は、平均してほぼ同じである。
- ⑤ 翻訳者にとっては後編集作業は 退屈だが、翻訳の際に役立ってい ることもあるようだ(特に専門用 語に関して)。

- ⑥ 翻訳者は修正を容易に行える環境が必要と感じている。
- 3) ポストエディト付き機械翻訳と通 常翻訳の品質比較

JPRS ( Joint Publications Research Service) と FTD (Foreign Technology Division Translations) による翻訳の質の比較 (FTD の翻訳 は後編集付きの機械翻訳) した結果、JPRS の翻訳の方が FTD のものよりも質が良かった。

# m) 自然言語処理と計算言語学 (Automatic Language Processing and Computational Linguistics)

#### <主張>

機械翻訳の開発に資金を投資したが 成果が出ない。これは機械翻訳の基礎技 術である言語学が未熟であることが原 因である。そこで言語学者用に、言語デ ータの整備、言語理論検証用のコンピュ ータ支援ツールを開発することが重要 である。

#### <概要>

政府は過去 10 年間に 200 万ドルを機 械翻訳の研究に投資した。これは 1 年間 に政府が翻訳に支払う金額よりも多か った。

一方政府援助は少なかったけど、コンピュータ産業界はこの間マシーンの開発に重点をおき、電子化出版技術も発展させた。これは、マーケットが明白なためであった。

機械翻訳の研究は、コンピュータハードウェアの進展にはあまり寄与しなかったが、コンピュータソフトウェアには 多大な影響をあたえた。何にもまして重要な機械翻訳研究の成果は、言語学に与えた影響である。 計算言語学を研究すれば、自然言語の研究に革命を与えるであろう。現在、理論をデータと付きあわせるためには計算機のサポートを要求している言語学者が急増している。一世代まえに一生かかっていた作業は今では数週間できる。この言語学の革命は、単に機械翻訳や解析技術の研究によるものではないが、これなくしては考えられなかった。

コンピュータは言語学者に多くの挑戦をつきつけたが、この挑戦に立ち向か うことにより、多くの分野に重要な貢献 をもたらすことができるだろう。例え ば、

- 1) 外国語を教えたり、
- 2) 言語の本質を教えたり、
- 3) 令や交信に自然言語を使ったり、
- 4) 特殊目的の人工言語を使ったり、
- 5) 人間の言語能力、思考等の心理学実 験をしたり、
- 6) 翻訳とか検索の支援にコンピュータを利用したり、

#### に貢献できる。

これらのための言語学の研究には、コンピュータが不可欠である。しかし、コンピュータが言語データを容易に扱える一般的な方法はまだ開発されていない。従って今後重点的に行い、サポートしていかなければならない研究分野は、

- 1) 言語データを扱うための基礎研究
- 2) 言語学者の言語理論を記述し、それ を検証するためのコンピュータ利 用法

の2点である。

# プログラミング言語と言語学との関係 <主張>

プログラミングは言語学にとって実

際的なモデルであり、言語学はプログラミング言語の設計に理論的な方策を与えており、将来も互いの影響が予想される。

#### <概要>

- 1) 言語学からプログラミングへ
- ①FORTRAN 以前

プログラミングは機械語で行われ ており、言語学の影響を受ける余地 はなかった。

②FORTRAN ~ALGOL(1960) 言語学から文法を構文を借りて設 計された。

③ALGOL60~ALPAC 発表当時 数理言語学理論とその記法および 技法を借りている。

プログラミング言語の設計は、数 理言語学のために以前より理論的 な手続きでおこなわれるようにな った。

2) プログラミングから言語学への影響

プログラミングは、言語学に"実際的な"モデルを提供し、言語理論に対し計算に関する課題(表現の効率性や計算速度)を与える。

プログラミングと言語学の将来への影響

あるプログラミングは、言語で書かれたプログラミングの実行は何を意味するか、というプログラミング言語の意味論に注目が集まるだろう。また、自然言語理論における意味論の研究からの数理言語学的モデルへの影響も生じよう。それは、プログラミングに影響を与えることになる。

#### 機械翻訳と言語学

#### <主張>

コンピュータの出現により、大量データによる言語理論の検証が可能になった。機械翻訳による検証は、従来の言語理論が不十分であることを指摘した。しかし、コンピュータを得たことにより、今後、言語学の大きな進歩が期待できる。

#### <概要>

ALPAC レポート提出の10年前にはほとんどの言語学者は、「syntax とは語順、活用、機能語、イントネーション、句読法のような表層構造だけを考察するものである」と考えていた。コンピュータの出現により、大量データを使って従来の理論を検証することができるようになり、それまでに得られていた文法で構文解析実験を行った。これにより、多くの言語学者は初期の構文に対する考え方が不十分であることに気がついた。

このように、コンピュータは理論を確認する上で不可欠なものである。言語学の最近の多くの変化は、理論的進歩とともにコンピュータによる実証実験によるところが大きい。まさに、「コンピュータは自然言語の解析の水準を変ええることができる」と云えよう。事実、以前より多くの言語学者がコンピュータの支援を求めている。

機械翻訳は上記 2 つの側面を持つものである。一つは、言語学的研究の側面であり、一つは研究によって得られた方法を実験によって検証する側面である。この検証により機械翻訳はいろいろな否定的結果を持ったが、このことが従来の理論の不十分さを言語学者に指摘した。

# n) 翻訳改善への旅程 (Avenues to Improvement of Translation)

#### <主張>

有益な機械翻訳を実現するには、まだ 長期的な研究開発を必要とする。機械の 助けは、人間翻訳にも、マシンエイデッ ドの翻訳にも重要な要素である。翻訳作 業そのものだけでなく、文書処理、印刷 等の、付随的業務にも改善の余地があ り、それらの研究開発も重要である。目 標とする品質により、翻訳作業の内容 も、コストも大幅に変わる。研究開発投 資の方針としては、現実的で、比較的短 期的な目標に対して、引き締めた投資を すべきであると考える。

#### <概要>

一般の科学技術文献についてのマシ ンエイデッドな翻訳はできるが、有用な 機械翻訳は実現されていない。

機械翻訳が貢献したのは、第一に言語 学に対してであり、次いでコンピュータ・プログラムに対してである。翻訳は 重要であるが、その需要は、小規模ながら有能な人々の活動によって満たされている。しかし、翻訳の業務には改善の余地があり、それに向けての活動を提唱する。翻訳の品質が重要であり、目標とする品質によりコストは大幅に変動する。したがって、正確さと品質についての客観的な評価を達成することが重要である。

機械の助けは、人間翻訳にも、マシン エイデッドの翻訳にも重要な要素であ る。FTD (USAF Foreign Technology Division) のデータによると、製作費(最 終原稿の複製や製本の費用) が非常に大 きい。翻訳ジャーナルにおける遅れは、 翻訳よりも製作の方に起因しているよ うだ。

マシンエイデッドの翻訳はより良く、早く、安い翻訳への重要な一つの道かもしてない。人間の助けとして何が最も良いであろうか?——専門用語の選択、単語の辞書引き、あるいは租訳においてか、どのようにして、多段階の工程の遅れを無くすことができるのか?どのようにして製作費を切り下げられるのか?

マシンエイデッドの翻訳にとって自動文字認識が重要だと云われるが、FTD のデータではその効果は僅かであり、情報検索や郵便業務など、その技術がもっとも有効な分野において、その研究開発は、支持されればよい。

翻訳の改善のための研究開発に、いかほどの投資をすべきか?うまく運んでいる比較的小さなビジネスに、法外な投資をする理由はない。重要で、現実的で、比較的短期的な目標に対して、引き締めた投資をすべきであると考える。

#### o) 勧告 (Recommendations)

#### <主張>

第一に計算言語学を自然科学の分野 の一つと見做し、科学的に研究すべきで ある。

第二に翻訳業務の改善を抜本的に行 うべきである。

(直接言及していないが、工学的アプロ ーチでの機械翻訳の研究には資金援 助はしない。)

#### <概要>

本委員会は以下に述べる二つの異なった領域に資金を投入することを勧告する。まず第一に言語学の一分野としての計算言語学である。これは、機械の支援を受ける受けないにかかわらず翻訳

の実験を含め、パージング、文の生成、 構造、意味論、統計、計量的言語データ の研究である。言語学は自然科学と見做 すべきであり、実用的な翻訳にたいして すぐ目に見えるような貢献をするもの と判断してはいけない。また具体的な提 案事項は現代の言語学的研究の成果を 正しく見極められ、しかも科学的な価値 観に従ってこれを評価できるような人 材によって検討されることが重要であ る。

もう一つの領域は翻訳業務の改善で ある。その研究対象として以下のような ものが挙げられる。

- 1) 翻訳を評価するための実用的な方法
- 2) 人手による翻訳作業の能率アップ を計る手段
- 3) 様々の方法による翻訳の品質とコストの評価
- 4) 一度も目を通されることのないような不要な翻訳を阻止することを 目的とした翻訳の利用実態調査
- 5) 雑誌と個別の文書の両方の翻訳対象として、全翻訳過程における遅延の研究と、その改善策の研究
- 6) 様々な種類の機械支援翻訳の相対 的な速度とコストの評価
- 7) 翻訳の実施過程および編集過程の 機械化促進
- 8) 翻訳過程全般に関する研究
- 9) 今、主に機械翻訳中の自動辞書引き のために用意されている用語集の 改良作成を含め、翻訳者のための適 切な手引き書の作成

上記の研究はすべて容認できる品質の程 度の仕様化、および翻訳の迅速化、低価格 化を目指すべきである。 この勧告を受けて、1960年代後半から、 1970年代にかけて、我が国では、政府が「機 械翻訳」に関するプロジェクトに予算をつ けることはなかった。それについては、次 号に。

#### 参考文献

- Language and Machines: Computers in Translation and Linguistics, 1966.
- 2) 日本電子工業振興会編:「機械翻訳の開発 と実用化に関する実態調査」、p21~31、 1989,7.

# 機械翻訳を支援ツールとして使う英語学習法

# 平田 周 三田教育研究所LLP

Unfortunately people's credit to MT is very poor in Japan. It is necessary to show that even state of the art MT is practically useful. We studied an application of MT to English learning. This is based on our earlier finding that conversion of natural Japanese sentences to English-like Japanese (EJ) reduces errors of MT as low as five percent or less. We found that business people could correctly write e-mails in English by assistance of MT. The practice of EJ editing as well as proof-reading of English translated by MT can help business people's English skill training. We expect that this will never fail to enhance creditability toward MT among Japanese people.

#### 1. はじめに

残念ながら、MTに対する人々の信頼はきわめて低い。state of the art のMTであっても、その実用性が人々のあいだに広く認識されるならば、MTの開発や利用に拍車がかかるだろうと考える。

一般に、研究開発には、基礎研究のごとき無限の 目標をめざす exploratory のアプローチと、一定の 目標を実現する disciplinary な研究アプローチが ある。筆者は、後者の立場から、インターネット上 に提供されているMTでも十分に実用性を発揮で きるアプリケーションの開発を進めてきた。

#### 2. 問題の所在

まだわが国のMT開発が黎明期だった 1984 年に 筆者は1冊の本を出した。『日本語で英語を書く』 (東洋経済新報社) [5]というこの本で、思考実験 から、もし和文英訳のMTが完成し(若干の誤訳が あり得る)、大量にコンピュータが翻訳するように なった場合、誰がその誤訳を訂正するのだろうか という疑問を提示した。当時からプリエディティ ングとポストエディティングはMT翻訳モデルと して必須とされていた。しかし、コンピュータが 訳す英文でもって完結としなければ、商業的にな り立たないというのが筆者の主張だった。

同時に1つの実験を行った。女子大英文科の生徒の手を借り、1つのグループには私のところに届いていた海外からの英文のレターを高校生になったつもりで日本語に直訳してもらい、もう1つのグループにはその日本語を英訳させた。その結果、原文に使われていた重要な英単語を参考に付しておけば、ほぼ元通りの英語になることがわかった。

翻訳すべき日本語を英語的な構造の日本語に直せば、コンピュータは正しく英語にできる。そう確信して、当時のMT開発者に提案したが、聞き入れられなかった。わが国の翻訳業界では、翻訳依頼された日本語に手を加えることはタブーとされており(今日も続いている)、一方開発者は困難な日本語の自然言語処理の夢を追っていた。以後、MTから手を引いた。

しかし、2000 年代に入って文法解析をベースとするMTの困難さが認識され、コーパスの時代へと移っていった。コーパスは、観光や医師と患者との会話など特定の場面に絞ればかなりに実用性がある。しかし、一般的な文書の翻訳にはあまり適さな

い。対訳文例を限りなく増やしていけば翻訳の精度 は上がるであろうが、はるか先のことである。また、 一般にテンプレートで書いた文章に心が表れない ように、コーパスでどこまで微妙な心情が表現でき るか疑問でもある。

再び、日本語を英語的構造のものに直してから MTで翻訳するモデルの可能性を追った。その結 果、一定の成果を得て、EJ-TRANS SYSTEM と名 づけた。

図1に示すように、コンピュータを中心に、英語的日本語(EJ)に直す前処理工程と、訳された英語を自然な英語に直す後処理工程に分けた。前処理工程(EJエディティングと呼ぶ)は、中学高校程度の英語力で可能であり、後処理工程の英文エディティングはバイリンガルのネイティブを必要としないため、コスト面での有利さがある。このモデルは、翻訳者が脳内で行っている思考を工程化したものである。

#### TRADITIONAL TRANSLATION English Japanese EJ-TRANS SYSTEM Original English Japanese English EJ editing editina Controlled Mechanical Jpanese translation

図1 EJ-TRANS SYSTEM の概念図

EJ-TRANS SYSTEM を使って 20 冊の単行本を 1 人で約3カ月間で英語にした(全訳ではなく、 1/5 縮訳)。この結果、十分にこれが翻訳に実用可能なことを確認した。

#### 3. E J エディティング

自然な日本語の文章を英語的構造の日本語EJ に直す作業「EJエディティング」が、どれほど効 果的なものか、次に例文をもって示す。

#### 【日本語原文】

メールしてくださったニュース、大変興味深拝 見しましたが、さらに詳しいことを知りたいと思 いますので、詳細を知るにはどうすればよいか教 えていただけませんでしょうか。

これをそのままコンピュータに訳させると次の ような間違いだらけの英語になる(Yahoo)。

I saw it, but I should do what, or, the news that I e-mailed, would you tell me a more detailed thing very with fascination because I think that I want to know it to know the details?

そこで、元の日本語を次のようにEJエディティングする。

#### [EJ]

あなたのメールの中のニュースに私は興味を持ちました。このニュースについて私はより詳細を知りたい。その詳細を知るための方法を私に助言して下さい。

これを自動翻訳させると、次のように文法的に ほぼ正しい英語になる。

(Yahoo) I was interested in the news in your email. I want to know the details about this news more. Please advise me by a method to know the details.

(Excite) I got interested in the news in your mail. I would like to know details more about this news. Please advise me on the method for getting to know the details.

(Google) I was interested in news in your mail. I want to know more about this news. Please advise me how to get to know the details.

英語的構造の日本語というだけではいかにも曖昧である。その変換のルールが求められる。英語に翻訳しやすい日本語のルールづくりについてはすでにいくつかの研究があり、中でも Anthony Hartley らの研究 [1]は、同様の考えに基づくcontrolled language のルールを体系化するという試みとして注目される。また、Junko Matsui ら [2]は、88 人の英語学習者を対象に、6つの注意事項を参考にしながら、MTを使って各自が書いた日本語を英語に訳させる実験を行い、成果を報告している。

MTの翻訳精度を上げるために必要、あるいは 有効として、前者で 10、後者では6つのルールが 提示されている。これらは概ね正しいが、これだ けを守れば、MTが間違いなく翻訳することを保 証するものではない。多くの文例から帰納的にま とめられた原則にすぎず、これを使ってもなお正 しく訳されないケースは数多い。

筆者は、これ以外にも多くのルールを用意したが、その中には、上述の研究にはあげられていない「1つの文中で使う動詞は出来るだけ少なくする」といったものがある。これらのルールはノウハウに類するもので詳細の公表は控えるが、これらのルールへの過度の依存を避け、セマンティックな視点から日本語文をEJにする方法を重視するのが筆者の立場である。ロジカルな日本語の文章にすることがMT翻訳を容易にする。しかし、正しく訳されるようEJエディティングを行うには、変換に熟練を要することがわかった。

上述の研究論文に用いられている文例は、紙面の都合もあろうが、いずれも短く、比較的単純な内容のものであることが気になった。少々長い引用になるが、筆者が行った長文のEJエディティングの例を表1に示す。

#### 4. 英語学習への応用

EJ-TRANS SYSTEM を翻訳サービスに応用することは十分可能であるが、一般の人たちにMT

の実用性を実感させるには、別の方法が必要と考えた。これを英語学習法として利用することに着手したのは 2012 の 3 月であった。

ビジネス界では、グローバル人材育成の急務が叫ばれている。グローバル人材に必要な資質のトップにあげられるのは常に外国語力である。 しかるに、どの研修コースを見ても、英語抜きである。 一方、英語指導は、依然として英会話中心であり、基本的には学校における英語教育メソッドの延長でしかない。

企業あるいは公的機関において、英語ができる 従業員はきわめて数少ないのが現実である。わが 国の世界からの孤立を避けるには、海外部門だけ でなく、ビジネスに携わる人たちの3分の1以上 が、英語が "できる" 必要がある。また、英会話 が重視されているが、海外とのコミュニケーショ ンの主体は、とりわけビジネス界にあってはeメ ールに移行している。日本人の発言力が求められ ているが、ビジネスの実務においては日常英会話 では役に立たない。仕事上で自分の考えを英語で 表現するには、高度の内容が必要であり、英語を 「書く力」が要求される。

しかし、多忙なビジネスマン、とりわけ管理職には英語学習にあてる時間もなく、実務に直結する英語力を身につけることは絶望的である。そこで、eメールをMTを使って書く技術を教えることを考えた。

文法的に正しい翻訳をMTにさせるには、EJエディティングの技法を習得せねばならない。それは、日本語から日本語への変換であるから、さしたる英語的知識は要らない。基本的には、英語の基本5文型がわかっていればよい。そのことは、EJ-TRANS SYSTEM ですでに実証済みだった。

#### 5. MTを用いる英語学習の効果

日常、社内に出すeメールをMTで英語にする 演習を行った。24 回の添削指導により、受講した

#### 表1 長文のEJエディティングの文例

#### 【オリジナルの日本語】

今回の研究では、2012年の研究で見出した自閉症スペクトラム障害群の行動および脳活動の特徴が、オキシトシンの点鼻スプレー投与によって改善するか否かを検証しました。東京大学医学部附属病院において40名の自閉症スペクトラム障害の成年男性を対象として二重盲検など最も客観性の高い厳密な方法で医師主導臨床試験を行った結果、オキシトシン投与が自閉症スペクトラム障害群において元来低下していた脳活動を有意に上昇させ、それと共に対人コミュニケーションの障害が有意に改善されることを世界で初めて明らかにしました。

#### [EJ]

われわれは、2012年に、自閉症スペクトラム障害群の行動および脳活動の特徴を見出した。今回のわれわれの研究目的は、これらの特徴がオキシトシンによって改善されるかどうかを検証することであった。われわれは、オキシトシンを患者に点鼻スプレーによって投与した。実験は、東京大学医学部附属病院において行われた。40名の自閉症スペクトラム障害の成年男性が臨床試験の対象に選ばれた。二重盲検を含む、最も客観性の高い正確な方法をわれわれは採用した。この臨床試験は医師たちが主導した。その結果、われわれはオキシトシン投与が、自閉症スペクトラム障害群の低下していた脳活動を上昇させたことを確認した。それから、対人コミュニケーションの障害が有意に改善された。これは世界で初めての発見である。

#### 【EJのMT英訳】

We found characteristics of an action of the autistic spectrum disorder group and the brain activity in 2012. Our study purpose of this time was to inspect whether these characteristics were improved by oxytocin. We gave oxytocin to a patient by nasal drop spray. The experiment was conducted in University Tokyo Hospital. 40 adulthood men of autistic spectrum disorder were chosen as an object of the clinical trial. We adopted the exact method that had highest objectivity to include double-blind. Doctors led this clinical trial. As a result, we confirmed that the oxytocin dosage significantly raised the brain activity that decreased of the autistic spectrum disorder group. Then, an obstacle of the interpersonal communication was significantly improved. This is the first discovery in the world.

10 人全員がほぼ意味が通じる英文にすることができるようになった。英語的構造の日本語E J を書くこと自体が、英語の学習になることが確認された。さらに、MTが訳した文法的に誤りのある文章を受講者自身が正しい英語に直すことも英語力を高める。とくに、自分が書いた英語を教師から直されるのではなく、MTが訳した英語を受講者が添削するという立場は心理的優越感を与えるため、英語学習への興味を高める効果がある。

MTの翻訳能力が完全でないがゆえに、英語学 習にこれを役立てることができるというのはいさ さか皮肉だが、利用者にMTの利用価値を認識さ せ、何よりも英語学習に大きく貢献することの意 義は大である。

この英語学習法を EdeM と名づけた。英語教育の専門家の中には、これを批判する向きもある。しかし、明治以来の辞書と文法をツールとする英語学習法に、MTという新たな支援ツールが加わるのは画期的なことである。電子辞書が許されるのだから、MTもそれに準じるものとしてよいのではなかろうか。

#### 6. 英語脳をMTでつくる

さらに大きな発見があった。英語的構造の日本

表 2 英語力評価基準 (平田周考案)

分類	レベル	英語ができる内容
英語に自	A 1	英文を読むこともないし、話すこともない。外国人の話していることはほとんど理解できない。英語 の文章を読めと言われても、たどたどしい読み方になる。
信はない	A 2	外国人から話しかけられたら、片言英語で答える。英語の文章を声に出して読めと言われれば、 発音はかなりひどいがなんとか読める。いくつかの会話の表現を知っている。
ある程度 B1 英語がで きる B2	В 1	街で、外国人から英語で道を聞かれたら、簡単に教えてあげる。辞書を片手なら、英語の文書を 読んである程度理解できるが、英語でメールは書けない。パーティ等で簡単な会話ならできる。
	В 2	ほぼ正しく英文メールが書ける。外国人と日常会話ができる。ビジネス文書やレポートを、辞書を 引きながらであれば、読んでほぼ理解できる。
英語に自	C 1	英語で書かれた雑誌記事やビジネス文書をさっと読んで理解できる。報告書を英語で書ける。外 国人との会合で話し合いができる。講演を聞いておおよそ理解できる。
信がある	C 2	聞いたり読んだりしたほぼ全てのものを、理解できる。自然で流暢かつ正確に自己表現ができる。 国際会議で、質疑応答、ディスカッションができる。

語E J は、英語脳に相当するものではないか。中 津燎子および国広哲弥が「中間日本語」[6]と呼ん だものに近い。英語熟達者は、英語を話すとき、 自然な日本語ではなく、英語的構造の日本語で思 考する。その英語脳がなければ、MTが誤訳する ように、いくら英語の文法や語彙を覚えても、ネ イティブに通じる英語にはならない。

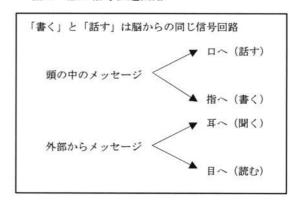
この英語脳は、普通には、英語の上達とともにつくられていく。学校では教えない。だが、EJエディティングによって、日本語で英語脳をつくることが可能になる。英語的センスの文章になっているかどうかをMTが実証する。その意味では、MTはイングリッシュ・ネイティブなのである。

一般に英語教育界では、oral English と literal English にわけて英語教育法を考える。しかし、脳のはたらきからいえば、図2に示すように、「書く」と「話す」は脳からの信号伝達回路として1つのものと考えるべきである。

いまだに、英会話信仰は強い。世間では、英語を話せるようになりたいという願望が圧倒的に多い。しかし、書くことは、正確な英語表現力を養うとともに、英語学習者の英語の間違いを訂正す

る場合、はるかに容易である。

図2 脳の信号伝達回路



#### 7. 英語力評価基準

「英語ができますか?」と聞かれて、yesかnoと答える基準はまったく明確でない。TOEICの点数が人気だが、300点でも英語ができるとしてよいのかどうか判断できない。このテストには「書く」「話す」は含まれていない。EUでは、CEFR[11]という「英語で何ができるか」を基準とした6段階の評価基準がある。わが国でも、これと同様のものをつくろうと、文部科学省と大学が研究を進めている。しかし、現在、試案として発表されているものは学校教育向けのものであるのと、非常

に複雑である。組織における全員の英語力を調べるにはこれは使えないので、独自の案をつくった(表2)。この中で、「ある程度英語ができる」のB2の条件に「ほぼ正しく英文メールが書ける」を入れた。この条件を満たせる人は少なく、大部分は、A2かB1にランクされるであろう。しかし、MTを使ってもよいということにすれば、現在A2の人でも、一気にB2に上がる。

まだ調査準備中だが、一般的な会社では、A1かA2が70%程度と想像される。しかし、MTをツールとして使うことを教えれば、そのうちの30%が一気にB2に上がる可能性がある。英語ができない″国民を、短期間に英語ができる″国民に変えることができれば、わが国のグローバリゼーションへの貢献は大である。さらに英語ができる人であっても、海外に向けての情報発信力を高め、英語で積極的に意見を述べることに役立つであろうと思う。

#### 8. 結語

MTを英語教育に役立てるという発想は、寡聞にして、これまでに聞くことはなかったように思う。英語的構造の日本語を書くことによって、英語脳がつくられ、さらに英語学習にはずみがつく。MTによって、EJエディティングした日本語の変換の妥当性を評価できることは、学習者のみならず、英語学習指導者にも、またロジカルな日本語を書く訓練にも使えるであろう。

英文和訳についても、日本人の場合、controlled English とは異なる考え方が必要だと考える。 とくに問題とすべきは、英語の語彙の多義性である。

これについての解決法はすでに考えてあるが、 MTのプログラムに手を加える必要があり、まだ 開発に着手していない。MTプログラム開発者の ご協力を仰ぎたい。

#### 参考資料

- Anthony Hartley, Midori Tatsumi et al. 2012.
   Reliability and Translatability Judgment for
   'Controlled Japanese'. Proceedings of the 16th
   EAMT Conference.
- [2] Junko Matsui, David Magnusson. 2011. Six Pre-edit Techniques for Enhancing Japanese to English Machine Translation. *Interpreting and Translation Studies*, No.11, 2011.
- [3] Sharon O'Brien. Controlling Controlled English – An Analysis of Several Controlled Language Rule Sets.
- [4] Hiroyuki Kaji. 1999. Controlled Languages for Machine Translation: State of the Art. MT Summit VII
- [5] 平田周『日本語で英語を書く』(東洋経済新報社 1984)
- [6] 中津燎子『再びなんで英語をやるの?』(文 藝春秋 1983)
- [7] 三森ゆりか『外国語を身につけるための日本 語レッスン』(白水社 2003)
- [8] 大出晃『日本語と論理』 (講談社 1965)
- [9] 荒木博之『日本語が見えると英語も見える』 (中央公論新社 1994)
- [10] 金谷武洋『日本語は敬語があって主語がない』 (光文社 2010)
- [11] 小池生夫ほか『企業が求める英語力』 (朝日 出版社 2010)

# 法令工学の言語処理

#### 島津 明

#### 北陸先端科学技術大学院大学 (JAIST)

#### 1. はじめに

私たちが住む社会は、税、年金などの多数の社会制度により支えられています。税 や年金などの制度は、法令として定められるとともに、法令に基づいて法令施行情報システムとして実現され運用されています。メディアなどにおいて話題となる企業の法令順守や内部統制、インターネットにおける情報倫理なども、組織や個人の行いを規定する法令や社会規範の問題です。近年、工業技術なども世界的な標準化が進んでいますが、これも社会を規定するものと言えます。

変化が速く、グローバル化が進む社会に対応するためには、適切な制度が設計され、タイムリーに法令が制定や改定されなければなりません。法令に基づく電子社会は安心安全でなければなりません。住み良い社会であるためには、法令、情報倫理、技術標準などが適切に作られ運用されなければなりません。

法令工学は、「法令は社会の仕様書」、「法令は社会を動かすソフトウェア」という立場から上記の課題に対応する新しい学問です[3,4]. 法令が言語により記述されていることから、言語処理の役割が大きくなります。法令工学やその言語処理について、第5回産業日本語研究会・シンポジウム(平成26年2月27日開催)において紹介しました。発表資料は産業日本語研究会のホームページ(http://www.tech-jpn.jp/)にあります。以下では、シンポジウムにおいて発表

しました内容の概要を報告します.

#### 2. 法令工学とは

法令は, 各条文が矛盾なく全体として整 合性を持ち、制度や政策を過不足なく正確 に記述していることが求められます. 昨年, 所得税に関する法律に記述ミスがあり、日 本政府はかなりの税収を失ったと報道され ました、大きな法令文書であれば、その作 成修正作業は多大なものとなります。 法令 は他の法令とも矛盾や不整合があっては困 ります. 日本政府の法令は、7000程度ある とされますから, 既存法令との関連を調べ る労力も大変大きなものとなります. この ために、法務担当者の労力は大変なもので あると言われていますが, 社会の変化は速 く,変化に対応した制度や政策の実現のた めに, タイムリーな法令の制定や改定が期 待されています.

法令は、このような状況にあると言えますが、情報システムのソフトウェアの設計開発におけるような方法論やシステムが欠けています。法令工学は、このような問題認識と立場から、法令のタイムリーな制定や改定を支援する系統的方法論、安心安全な法令施行情報システムの設計開発のための方法論、組織の規則や技術標準などの作成、利用、運用などを支援する方法論などを明らかにすることを目的にします[3, 4].

法令工学は,21世紀 COE プログラム「検 証進化可能電子社会」(2005~2009)[11] において提案されました. 当初の研究にお

いて、法令の矛盾検出・除去の基本原理が示 されました[2, 3]. それは、法令文書を論 理・形式表現に変換し、形式推論により矛 盾や不完全性の検出を行い, 問題箇所を除 去あるいは変更し、内容を確認し、新たな 法令文書を求めるというサイクルからなる というものです. この考え方に基づいて, 法令を1階述語論理表現で記述し、矛盾を 定義して, 定理証明技術に基づいて矛盾検 出を行うシステムが提案されました[2]. 実 際, このシステムが実装され, 富山県庁の 協力のもと、富山県の電子化条例など関連3 条例を対象に、矛盾検出が試みられました. この他、条文の論理・形式表現への変換法、 法令施行情報システムのアカウンタビリテ ィ機構の提案と実証, 法令ドメインの形式 記述と検証、企業の内部統制の形式記述と 検証などの研究がこれまでに行われていま す[4, 13].

法令は民主社会の規則ですから、誰にでも理解できることが望まれますが、法令文書には長文が多く、複雑で理解しにくいところがあります.この点については、法令文書の可読性の向上が課題となります.これについては、次節に紹介します.

## 3. 法令工学の言語処理

法令工学により,矛盾・不完全性の検出 除去,法令文書の作成支援,法令施行情報 システムの設計開発の支援,組織などの活 動の適法性の判定,法令や判例に特化した 高度検索,法令や判例についての質問応答, 法令文書の機械翻訳,法令や判例の要約な どが可能となります。これらの処理は法令 文書の法令構造に基づいて行われます。法 令構造は,当該法令における法律行為,法 律行為に係わる要件効果構造,法律行為間 の関係,当該法令における対象や法律行為 に関するオントロジーなどにより定義されます.そこで、法令工学の言語処理は、法令文書から法令構造やその一部を求めること、法令構造から論理・形式表現など応用に向けた表現に変換すること、法令構造に基づいて、文書作成支援、高度検索、質問応答、翻訳、要約等を行うこととなります.

法令文書は一定の様式で書かれ、十分に 吟味されていることから一般のテキストよ りも扱い易いと期待されますが、それでも 自然言語表現として多様であり、省略や曖昧性もあり、既存技術がそのまま適用でき るわけではありません. 北陸先端科学技術 大学院大学では、これまでに条例、年金法 などを対象に、要件効果構造の解析[8,14]、 条文等の論理式への変換[7,9]、条文間の照 応関係の解析[15]、号の条文への埋込み[5]、 法令の検索や修正のための該当箇所探索 [10]、条文の翻訳[1]、質問応答[16]、条項の 構造的書換え[12]、オントロジーの作成、判 例の構造化[6]などを研究してきました.

上記の中で要件効果構造の解析は様々な 処理の基盤となるものです. 要件効果構造 は条文が適用される要件と条文が適用され たときに働く効果とからなる構造です. 条 項は一般に複数の条文からなり、それらが 複数の要件効果構造を表現しますが、条文 と要件効果構造とは複雑な対応をしていま す.このような現象に対して、規則ベース の方法[14]と機械学習ベースの方法[8]を研 究しています. 国民年金法を対象にした実 験では、規則ベースは 100 ほどの規則数と なり,解析結果の曖昧性が課題です.機械 学習ベースでは, 要件効果構造の要素から なる重み付きグラフを作り、線形計画法で 構造を認識する方法を提案しました. 国民 年金法を対象にした実験では、要素の認識 が 74%、構造の認識が 83%, 全体で 62%の

結果を得ています. データ量や素性の考慮により精度は向上すると期待されます. このような解析に、号の埋込み、照応解析、格構造解析、条項の役割や関係の解析などを加えることにより、法令構造の内容がより豊富なものにできます.

要件効果構造は、これを照応解析、質問 応答、翻訳などに適用することにより、そ れらの性能が向上することも期待されます [1,15,16].

要件効果構造は、条項の可読性向上にも 利用できます. 所得税法や年金法の条項に は長文が多く, 中には 5000 字を超えるよう なものもあります. また, 長い埋込み文や 括弧による挿入文も数多くあり, 可読性が 大変低くなっています. そのような条文に ついて, 要件としてどのような条件が並ん でおり, 効果は何かが明示的になるだけで も読み易さが違います. 要件や名詞句の並 列も多く, それらを分かり易く列挙するこ とでも読み易さが変ります. このような点 を考慮した規則を作り, 実際に国民年金法 の被保険者に関する条項(年金法全体の3 分の2程度)を構造的に書換えました[12]. 社労士に検証してもらい, 誤った解釈のと ころを修正した上で, 非専門家を被験者と して読解実験を行いました. 条項を8つ選 び、15 人の被験者には元文を、別の15 人 には構造的書換え文を読んでもらい、それ ぞれの条項に関する質問を正解するまで答 えてもらう実験をしてもらいました. 結果 は、構造的書換えのグループがより早く正 答するというものでした.

## 4. おわりに

情報処理の歴史において, テキストが電 子化されるとともに, テキスト作成を支援 するソフトウェアが作られ, 多くの人が電 子テキストをたやすく作り利用する時代となりました. ソフトウェアを利用しないオフィス業務は考えられない時代となっています. 法令が民主社会を支える基盤であるにもかかわらず, そこにはソフトウェア科学の恩恵が驚くほど少ないと言わざるをえません. 法令工学は, 法務に従事する人々はもとより, 一般の人々にも恩恵をもたらし, 安心安全な社会の構築に貢献するものと思います.

### 参考文献

- O. H. Bui, M. L. Nguyen and A. Shimazu. Divide and Translate Legal Text Sentence Using Logical by its Structure. International Conference on Knowledge, Information and Creativity Support Systems (KICSS2012), pp.18-23, Melbourne, 2012.
- 萩原信吾, 東条 敏. 対立を用いた法的 知識の整合性検証. 人工知能学会誌, 24,5,p.405-416,2009.
- 片山卓也(編). 法令工学の提案. JAIST Press, ISBN4-903092-04-6, 2007.
- 片山卓也,島津明,東条敏,二木厚吉,落水浩一郎.電子社会と法令工学. 人工知能学会誌, vol.23, no. 4, pp.529-536, 2008.
- 4. Y. Kimura, M. Nakamura and A. Shimazu. Treatment of Legal Sentences Including Itemized and Referential Expressions Towards Translation into Logical Forms-, New Frontiers in Artificial Intelligence: LNAI, Vol.5447, pp.242-253, 2009.
- H. Igari and A. Shimazu. Syntactic Document Structure of Legal Documents. International Workshop

- on Juris-informatics (JURISIN 2012), pp.49-62, Miyazaki, 2012.
- M. Nakamura and A. Shimazu. Semantic Analysis of National Pension Law, International Workshop on Juris-informatics (JURISIN 2009), pp.83-94, 2009.
- 8. B. X. Ngo, M. L. Nguyen, O. T. Tran, and A. Shimazu. A Two-Phase Framework for Learning Logical Structures of Paragraphs in Legal Articles. ACM TALIP, Vol.12, No.1, pp.3:1-3:32, doi>10.1145/2425327. 2425330, 2013.
- M. L. Nguyen and A. Shimazu. A Semi-Supervised Learning Model for Mapping Sentences to Logical form with Ambiguous Supervision. International conference on Applications of Natural Language Processing to Information Systems (NLDB2012), pp.116-127, Groningen, 2012.
- 10. M. Q. N. Pham, M. L. Nguyen, B. X.
  Ngo and A. Shimazu. A Learning to
  Rank Method for Information
  Updating Task. Applied
  Intelligence,
  DOI={10.1007/s10489-012-0343-2},
  pp. 1-12, 2012.
- 11. 島津 明(編). 21 世紀 COE プログラム「検証進化可能電子社会」最終成果報告書, 北陸先端科学技術大学院大学, 2009.
- 12. 島津 明. 国民年金法の構造的書き換え・法令工学の立場から・. JAIST Press, ISBN978-4-903092-19-5, 2009.

- 13. 島津 明. 法令工学:安心な社会システム設計のための方法論-法令文書の解析を中心に-(解説論文). IEICE, Fundamental Review, Vol.5, No.4, pp.320-328, 2012.
- 14. K. Takano, M. Nakamura, Ohyama and A. Shimazu. Semantic Analysis of Paragraphs Consisting of Multiple Sentences - Towards Development of Logical Formulation System Legal Knowledge and Information Systems (JURIX2010), pp.117-126, 2010.
- 15. O. T. Tran, M. L. Nguyen and A. Shimazu. Reference Resolution in Legal Texts. International Conference on Artificial Intelligence and Law (ICAIL), pp.101-110, Rome, 2013. (Award for Best Student Paper)
- 1 6. O. T. Tran, B. X. Ngo, M. L. Nguyen and A. Shimazu. Answering Legal Questions by Mining References Information. International Workshop on Juris-informatics (JURISIN 2013), Yokohama, 2013.

## TAUS Symposium Report

# TAUS Tokyo Executive Forum 2014 - A Sway of Innovation (www.taus.net)

Jaap van der Meer

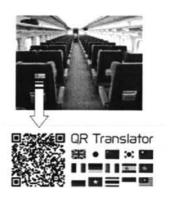
TAUS (Translation Automaton User Society)

The fourth TAUS Executive Forum was held on April 10 and 11, again hosted by Oracle Japan. We are delighted to share here a snapshot of the sway of innovation entering the global translation from Japan. It takes just a day of wandering in Tokyo as a non-Japanese speaker to understand why translation innovation starts at this side of the world, and much less so in the western part of the world. Both visitor and host are totally helpless and hopeless when trying to talk to each other. That is much less the case in Europe and North America where English often becomes the 'langue véhicule' for tourists and business men.

In January I called the acquisition of Systran by the Korean company CLSi a "hazard troublant" (a troubling coincidence), happening at the 60th anniversary of MT, but now I must admit that I fully understand. The needs have shifted to a new economic epicenter of the world, where the language barrier is challenging the global expansion perhaps more than anywhere else. Jean Senellart and Dong-pil Kim presented the newly created translation technology company with an unequalled energy and confidence.

It is not official yet, but the new company is likely to get the name Systran International. Systran International is probably the largest independent translation technology company in the world now with ambitions to serve the needs of consumers (through apps on among others Samsung phones) and businesses. Plans were announced at the TAUS Tokyo Forum for the extension of the product portfolio into translation management systems. We were proud to have the premiere.

In the Innovation section of the TAUS Tokyo Forum we were thrilled to hear also from QR Translator, a Japanese start-up with an idea so simple and yet so powerful, making you think: "why did nobody else come up with this yet?" QR Translator issues a special QR code that bridges the language barrier. Scanning the QRT code — a combination of a QR code and icons representing available languages — leads the smart phone user to a translation. The company obviously has registered patents now.





You will find these QR codes with language flags in the Shinkansen (the Japanese trains), taxis, museums and who knows how many other public places. Translations are crowdsourced or machine translated. It is very simple to create a QRT code. The benefits are obvious. The user is able to separate creative works from the translation task, save space and there is no need to change the signboard after you found some mistakes in the translation. The company has registered patents now.

Another example of true innovation came from NTT DOCOMO. The audience voted for the NTT DOCOMO presentation of the Translation Glass as the most innovative presentation at the TAUS Tokyo Forum, making NTT DOCOMO the winner of the 2014 TAUS Excellence Award. Translation Glass is a simple eyewear device that lets users read a restaurant menu in a language of choice.

Wearing a Translation Glass an English speaking visitor in Tokyo can for instance glance over a Japanese menu the English read translation displayed on the inner screen of the glasses. Translation Glass is already available in various combinations of Japanese, Korean, Chinese Simplified Traditional, Indonesian, and Thai, Vietnamese and English.

Translation Glass is the result of a major engineering effort combining character recognition with language processing. The character recognition had to work for text printed vertically and horizontally and written in both directions, and of course the character recognition should not be hindered by colors and backgrounds. Making these applications available through a web API to the developers community, NTT DOCOMO is driving innovation in translation.

#### Social Translation

The Asian translation industry has generated a unique new translation platform that we refer to as 'social translation', a form of innovation born out of necessity. When Naoki Yamada returned to Tokyo from his studies in Los Angeles, colleagues and friends kept asking them to translate something, a sentence, a letter, an email, anything. He saw the business opportunity and started developing a social translation platform. What he had in mind was to find the hundreds or perhaps thousands Japanese students who had studied abroad or at least people like himself had enough knowledge of English to help others. Not a professional translation B2B service, but more like a community where volunteers get rewarded 1 or 2 cents per word for helping someone else with the translation of a love letter or an eBay add for the sale of a bicycle.

Finding these volunteer translators was the biggest challenge when he launched the business in 2009. He printed flyers and handed them out at metro stations in Tokyo. Now Conyac – that's the name of the platform – brings together

15,000 translation volunteers and it has received about half a million translation requests varying from one phrase to a page or a bit more.

Conyac does not stand on its own as an innovative social or community-based translation platform. There is also Gengo with a similar history and background. Gengo was founded by Matt Romaine and Robert Laing who as bilinguals living in Tokyo had the same experience as Yamada. Different from Conyac, Gengo is also targeting business customers. Both Gengo and Conyac are funded by investors. Gengo has attracted investments of \$14 million, Canyon a much more modest amount of \$1 million.

Laing sees the translation industry as an industry just waiting to be disrupted. The traditional LSP entrenched in bureaucracy, complex workers roles, fragmentation and working with outdated technology. They know it the traditional LSPs - but they can't help it. It's like a car accident in slow motion. You see it's happening, but you can't stop it. He compares Gengo's disruptive potential for the translation industry to what Amazon has done to book distribution, to what DuoLingo is doing to RosettaStone in the language training sector and to what Uber is doing to the taxi sector. Not every business may be inclined to trust their translations to volunteer translators so quickly, but socially oriented companies will find themselves more naturally attracted to the social translation platform.

Eventbrite TripAdvisor and for instance both use Gengo for their translations. The investors behind Gengo and Conyac wish that the car accident wasn't that slow-motion, but if they lose their patience there are a bunch of others still, and all from Japan and Korea again, lined up to carry the innovation forward. We have had Miori Sagara from Baobab from Tokyo - presenting at TAUS events in 2012 and we look forward to also meeting One-Hour Translation, OneSky and Flitto. Who knows, by the time we are back in Tokyo in April 2015, there will be new Japanese innovators with a promise to disrupt the old-fashioned translation industry.

The TAUS Tokyo Forum provided for a good insight in this new emerging social translation model with presentations by Conyac, Gengo, Eventbrite and TripAdvisor.

Mark Seligman from Spoken Translation presented at the TAUS Tokyo Forum the latest developments speech-to-translation and he demonstrated its actual workings through his mobile phone and .... his wristwatch. A great prelude to the TAUS Annual Conference on October 27 and 28 in Vancouver where speech translation will be one of our key focus areas and where Alex Waibel will be our keynote speaker.

Exciting was also to hear from Toppan Printing (the largest printing corporation in the world) that automated translation is now becoming a focal point for strategic growth in the company. Mr. Masanobu Ogata presented a wide scope of use cases where Toppan in partnership with the Kyoto-based research institute NICT is developing customized MT solutions. Human Science gave the audience a deep insight in their experience with applying pre-editing on source texts before starting the MT process.

Three European based vendor organizations Yamagata Europe, Hunnect and Pangeanic - all three with a strong service presence in Japan joined the TAUS Tokyo Forum to share their experiences with machine translation. We were thrilled also to hear the presentation from Mike Tian-Jian Jiang from Yaraku, another translation technology company born in Japan with global potential. We will cover this company in another article later.

The host of the TAUS Executive Forum

Oracle — was represented by their Vice

President Aiman Copty who is responsible
for the Worldwide Translation Group that
manages all of Oracles global translation
activities. Aiman shared insights with the
audience on how the world of 'big data' is
starting to influence translation decisions.

## Acknowledgment from the editorial board

This article is on TAUS web page (https://www.taus.net/taus-tokyo-executiv e-forum-2014-a-sway-of-innovation).

Special thanks are due to Mr. Jaap van der Meer (TAUS Director) for his courtesy to allow us to cite the web page.

## シンポジウム参加報告

## TKE (Terminology and Knowledge Engineering) 2014 報告

## 影浦峡

(東京大学·大学院教育学研究科)

TKE (Terminology and Knowledge Engineering) 2014 は、2014 年 6 月 19 日と 20 日、ドイツ標準化機構で開催された。参加者は約 70 人。大学関係者と産業関係者、標準化機構や EU 関係機関の関係者が参加していた。会議は 2 セッション並列で進められ、発表は招待講演を除いて 23 件。Proceedings は http://tke2014.sciencesconf.org/conference/tke2014/eda\_en.pdf で公開されている。参加者は 8 割以上が EU 圏。報告者は TKE 1996 に参加したことがあるが、あるときから査読を強化したこともあってか、参加者の数は大幅に減っている。

私が参加したセッションの発表は、大きく、状況・実践報告系、資源構築系、枠組み(システム・基準)提案系に分けることができる。以下、それぞれを簡単に(一部は報知的に、一部は指示的に)紹介しよう。

#### 状況・実践報告系

"Combine EU Terminology with Communication and Ontology Research." EU ターミノロジーIATE は EU の法律や関連文書を翻訳するプロセスで構築され、そのプロセスに組み込まれたかたちで管理されており、EU の 24 言語を対象としている。確定された用語約 800 万は一般からアクセスできるが、検討中・討論中のものは別扱いである。構築されたターミノロジーをどのように公開し、利用に供し、そこからフィードバックを得て活用していくかについて、(a) 組織内の方針、(b) 関連組織間での共有、(c) 国際的な機関との協力、(d) 大学や一般との関係構

築、という4つのレベルにわけて説明がなされた。 技術的な側面では、現在のターミノロジーをオントロジー体系と統合することが重要な課題として示された。東アジア言語圏でも同様の組織と実践が望まれる。

"Terminology Management Systems and its Role in the Development of the Language Industry in Colombia." コロンビアの言語産業 (翻訳・通訳・技術文書執筆) において、用語管理システムがどの程度活用されているかの調査報告である。コロンビアの言語産業は中小企業及びフリーランスがほとんどで、106 の調査対象のうち用語管理システムを利用しているのは 8 カ所のみとのことである。27%は国際的な協力関係を持っていない。世界的なマルチランゲージベンダーが各国でシェアを延ばしているが、コロンビアの産業は今のところ地元の言語産業にかなり依存しているとのことであった。

翻訳関係の実践報告が 2 件あった。"Assisting the Use and Usefulness of KRCs in Translation: Towards a Protocol"は、KRC (Knowledge Rich Context)を翻訳者に示すことの有用性をめぐる実践で、既往の翻訳支援環境に近い環境を構築して、実験を進めている。参照行動のログを取るだけでなく、研究協力者(学生)への聞き取り等を含めて KRC の有効性を検証しようとしている。教育への応用は次のステップとのこと。"TermWise: Leveraging Big Data for Terminological Support for Legal Translation"は、法律分野を対象に、有効な翻訳データを翻訳者に提示するための知識抽出アルゴリズムとサーバベースのツールを開発したとの報告。学生およびベルギー法

務省の翻訳者が評価を行っている。

#### 資源構築系

"Concept System Analysis of the ThirdPlaceLearning Theory." 認知言語学的なアプローチを使って対象領域の概念関係を反映した用語集を構築していく方法の報告で、方法論よりもむしろ、教育プロセスで領域の全体像を示すためにこうして構築された用語集を利用しているとの指摘が興味深い。用語集の規模は120 語程度。

"A Combined Taxonomic-Frame Approach to Terminological Verb Collocations." FrameNet の枠組みを用いた専門動詞の記述に関する報告。最近、専門用語として形容詞や動詞など、名詞以外の品詞を扱う研究が増えており、その一つである。

#### 枠組み (システム・基準) 提案系

TBX (TermBase eXchange)関係の報告は3つあった。

"TBX-Min: A Simplified TBX-Based Approach to Representin Bilingual Glossaries." 大多数の用語データは Excel などのスプレッドシートで管理されているが、csv はやはり扱える情報の範囲に限界がある。一方、ほとんどの翻訳者にとっては、完全な TBX (TermBase eXchange) 仕様(117 データカテゴリ)は必要ではない。実際、ある用語ベースで、TBX のどのデータカテゴリーが使われているのかはわかりにくい。そうした状況から 24 カテゴリーからなる TBX Basic が 2009 年に提案されたが、これでもまだ複雑すぎる。報告する TBX-Min は、TBX のコア構造を維持しながら、簡単に利用できるよう、さらに簡略化した枠組みで、エンコーディングはデフォールトで UTF8、スプレッドシート形式や UTX

への変換も容易である。RDFとの互換性も検討されている。TBX-Min は二言語を扱うもので、複数言語を扱うときは複数のファイルが必要である。コンバータは http://tbxinfo.net にある。実際の利用がどのように分布しているのか、その中で TBX-Min がどのように位置づけられるかをめぐって多くの質疑がなされた。一つの意見は、翻訳支援システムはTBX を提供しているので、その一部を使えばそれでよいではないかというものであった。また逆に、個人翻訳者は Excel で情報を管理しており、それが実態であるというものである。二言語対に限る用語管理か、多言語の用語管理が必要かについても、議論があった。

TBX 関係の他の二つの発表は、データフォーマットの変換と互換性を中心としたもので、"TBX Goes TEI - Implementing a TBX Basic Extension for the Text Encoding Initiative Guidelines."は、人文学のテキストにおける用語管理という課題を通して、TEI (Text Encoding Initiative)と TBX の関係を論じたもの。"TBX between Termbases and Ontologies"は、TBX と RDF との完全な変換を中心に論じたものである。

TBX をめぐる発表は、それぞれの視点が微妙に 異なるようだが、発表の中でその部分は明らかにさ れない。視点の際をもたらす、想定されている利用 の場やニーズといった重要な点については主に質 疑でカバーされている。

"A Model and Method to Terminologize Existing Domain Ontologies."は、ISO TC46 の用語記述の枠組みと WOL で記述されたオントロジーとをメタモデルを使って接続する提案で、枠組みと手続きを示している。TBX そのものについてではないが、関連は深い。異なる記述枠組みを摺り合わせる課題は、情報の統合的利用やリンクデータが注目を浴びて

いる中で、重要なものと位置づけられているのだろう。

"Tangled Web of Concept Relations. Concept Relations for ISO 1087-1 and ISO 704."は ISO の専門 用語に関係する基準を、特に概念間の関係に関して論じたもので、専門用語管理に関する二つの ISO 基準における概念間の関係を検討し、改善を提案している。

2 日目(最終日) セッションの最後に、2 件の Industrial Talk が行われた。全体に、言語処理技術の 点からはそれほど最新のものはないが、発表の多く が、技術そのものでも、言語学者が言う抽象化された言語でもなく、人間の言語利用・言語実践の場に 見られる課題を扱うことから発生した専門用語を めぐる問題に関するものであり、言語処理系の学会よりも応用・実践への関心は高い。EU 諸国では全 般に言語実務や言語産業と技術、標準化との関係が 密接である。

色々な人と興味深い話をしたが、WIPO のGeoffrey Westgate 氏との話は一般にも関心が高いと思われるので紹介しよう。WIPOでは、日本語・中国語・韓国語の専門用語を強化するために、毎年、修士レベルでのインターンを募集している。中国と韓国からは多数の応募があるが、日本からは昨年度3名、今年度はゼロで、インターンの手続き(オンラインの試験およびそれに合格した場合インタビューがある)を送るから、ぜひ広めて欲しいというものである。英語力の壁に加えて、技術翻訳及び関連領域について大学レベルでの教育体制が確立していないことが、日本からの応募が少ない背景にあると思われる。

会議の最後に Eugen Wuster Terminology Prize が発表された。3 年に1度、ターミノロジーの領域で顕

著な貢献をした個人に与えられる賞で、今回の受賞者は韓国 KAIST の Key-Sun Choi 教授であった。言語処理の領域で広く名前を知られているが、ターミノロジーの領域でも、韓国専門用語協会(Korterm)を設立し、韓日中の用語管理や国際的な専門用語調整活動に精力的に関わってきた。今回、Choi 教授のこうした功績に対してこの賞が与えられたことは大変喜ばしい。

## TCシンポジウム参加報告

## 機械翻訳とデザイン思考

(株) 東芝 研究開発センター 鈴木 博和

2014年8月26日、27日に工学院大学新宿キャンパスにてテクニカルコミュニケーションシンポジウム2014(以下TCシンポジウム)が開催された。本年度のテーマは「TCが起こすイノベーション デザイン・シンキング」。モノづくりの現場においてTCを考える際に、デザイン思考によりユーザの視点を見直し、TCにイノベーションを起こそうというテーマであった。本レポートでは、聴講したパネルディスカッションの内容を簡単に紹介するとともに、機械翻訳におけるデザイン思考の可能性を考察する。

# 基調講演「デザイン・シンキングにより TC が起こすイノベーション」

慶應義塾大学大学院メディアデザイン研 究科 奥出直人教授による基調講演は、デ ザイン思考とはなにか、それをマニュアル 作成にどのように活用するかなどに関し、 実際のデザイン思考のプロセスを詳細に解 説されながらの非常に分かりやすい講演で あった。デザイン思考においてはエキスパ ート(TCではマニュアル作成担当者)と素 人(ユーザ)とのギャップを考える必要が ある。そのギャップを捉えるためには、① どうやって使っているのか観察し、②そこ からパターンを抽出、③コンセプトに落と しこんで、④試作、という流れでプロトタ イプを作成し検証を行っていく。コンセプ トをユーザの「メンタルモデル」と如何に 一致させるかが非常に重要なポイントであ るが、これはアンケート等で明文化される

わけではないので、プロトタイプによる検 証が必須ということであった。

# パネルディスカッション「コミュニケーションを大きく変える機械翻訳」

(株) 川村インターナショナルの森口功 造氏がコーディネーターを務められた同パ ネルディスカッションを聴講した。パネリ ストには凸版印刷(株)緒方雅信氏からは 電子版コミックの翻訳における機械翻訳の 活用に関して、(株) PIJIN の高岡謙二氏か らは QR コードを活用した翻訳ビジネスに 関して、(株) NTT ドコモの三井力氏からは Docomo が展開する機械翻訳サービス戦略 に関してそれぞれ紹介があり、将来的な機 械翻訳の活用方法などについてのディスカ ッションが行われた。印象深かったのは新 しい競争軸として提案された新しい概念で ある。従来の翻訳ビジネスの競争軸は品質、 コストの二軸で表されることが多かった。 今後はユーザのビジネスに如何に統合しや すいかという integratability も重要になる とのことで、これまでシーズドリブンであ った機械翻訳ツールの開発も今後はニーズ ドリブンに行う必要があるという示唆が得 られた。

## パネルディスカッション「ユーザのストー リーをつかむ」

ヤマハ (株) の大石浩史氏がコーディネーターを務められた同セッションでは、ユーザのニーズを突き止めるために如何に行

動観察を活用するかがテーマであった。例 えば近年のスマホの普及であるが、ネット でいつでも簡単に調べる方法が提供された ことにより、ユーザは深く探すことから今 すぐできることに価値観がシフトしている 点が指摘された。「今すぐ」できるというこ とはユーザの置かれているコンテキストが 重要になっていることを意味し、ユーザの コンテキストに合わないアプリは使われな くなっている。従って今までの UI は構造を 重視して作られてきたが、今後は「手順」 を重視する必要があるということである。 この「手順」について、開発者は「自らの 思考プロセスと行動を容易に説明できる」 と思っているがそれは思い込みであること が多く、実際には行動をよく観察しないと わからないことがある。そのため行動観察 が注目を集めているということである。

我々機械翻訳の開発者はこれまで、ユーザのニーズを調査する手法としてアンケートを多用していた。しかし真のニーズを明らかにするために行動観察を行う必要があるのかもしれない。実際にユーザが機械翻訳システムを使用している場面を観察する機会はほとんどないという現状を打破し、機械翻訳にイノベーションを起こすにはどうすればよいのかいろいろと考えさせられたセッションだった。

パネルディスカッション「異文化理解と発信のストラテジー ~TC や広告における 日本と海外の違い~」

(株) エレクトロスイスジャパンの中村 哲三氏がコーディネーターを務める同セッ ションでは文化の違いがどのようなコミュ ニケーションの違いを生み、どのような注 意点があるのか、どのように情報発信すれ ばよいのかについて議論が行われた。日本 はハイ・コンテキスト文化の言語として知られている。情報をすべて伝えなくてもあうんの呼吸でコミュニケーションがとれてしまうとか、自分の意見をはっきりと伝えるよりも周りに合わせた発言をすることが求められる、などと言われる。それに対して英語はロー・コンテキスト文化であり、言いたいことをはっきりと明示する文化であるとされる。

さて、「あなたの同僚が遅くまで仕事をし ていることに気づいた時、あなたはどのよ うにアドバイスするか?」というアンケー トに対して、国に依る違いはあるのだろう か?もちろんある。しかしそれは私が思っ ているのとはだいぶ違っているようだ。例 えば日本人であれば、欧米人はこのような 場面では「早く帰ったほうがいい」と直接 的なアドバイスをするだろうと思ってしま うのではないだろうか。実際には全く逆で やんわりと間接的な表現でアドバイスをす る。逆に同じ質問を日本人にした場合、婉 曲的にアドバイスするものだと思ってしま うがこれも回答は逆で、直接的なアドバイ スが多いという。意外にも思えるこの結果 は相手と自分の「領域」を考えると説明し やすい。即ち欧米人は自分の「領域」に関 わることはしっかり主張するが、相手の「領 域」に関することになるとそれに敬意を払 い踏み込まないようにするということだ。 先ほどの例で言うと、同僚が仕事をしてい るのは同僚がそう選択したという意志であ り相手の「領域」であるため、それに意見 を述べるときには間接的な物言いになると いうことである。欧米人に対するステレオ タイプな考え方しか持っていないと、異文 化コミュニケーションにおいて失敗してし まう可能性があるという警鐘であった。

このように異文化コミュニケーションに

おける注意点を、日本翻訳連盟理事の田中 千鶴香氏、慶應義塾大学井上逸兵教授、(財) ニッポンドットコムの Peter Durfee 氏のディスカッションにより提示された。

## 機械翻訳とデザイン思考

TC シンポジウムへはパネリストとして 2回参加しており、今回3回目の参加であ ったが、機械翻訳業界の盛り上がりを感じ たシンポジウムであった。前回2回の参加 の際には機械翻訳に関する基本的な知識や 注意点など、主に機械翻訳初心者に向けた 内容を意識していた。しかし今回はより実 務的・ビジネス的な観点での発表となって おり、同じ業界に属するものとして非常に 興味深く聞かせて頂いた。特に「デザイン 思考」に関する考え方は注目に値する。例 えば大阪ガス行動観察研究所(株)の越野 孝史氏の話は興味深かった。同氏がある高 齢者の歯磨きについてインタビューを行っ た時の話であるが、その高齢者は先端にラ イトが付いている歯ブラシをしきりに薦め ていたという。磨いている歯がよく見える ので非常に使いやすいということであった が、実際に歯磨きをしているところを観察 すると、何とその方は目を「瞑って」歯磨 きをしていたのだ。本人もそのことに気づ いていなかったという。この時の越野氏の 言葉が印象的である。「ユーザは嘘をついて いない。思い込んでいるだけである。」

我々機械翻訳の研究・開発者も思い当たる点はないだろうか?例えばアンケートやインタビューなどで「こういう機能が欲しい」と言われ、その通りの機能を作ることが果たしてユーザの真のニーズに応えていることになるのであろうか。ニーズにはアンケート等で分かる明示的なニーズと、ユーザは意識はしているが言葉になっていな

い潜在的ニーズ、ユーザ自身も気づいていない未知のニーズが存在する。これまでの機械翻訳ツールの開発において、上記の3つのニーズの中の「明示的なニーズ」のみを追い求めすぎていた気がしてならない。しかしこれからのグローバル競争の時代では、「潜在的ニーズ」「未知のニーズ」を追求することが重要である。それには行動観察などから得られる知見を使い、プロトタイピングを通して検証を行う以外に方法はない。

どの業界にもイノベーションが求められる昨今、このような視点を持ち古い「機械翻訳業界」に新しい「イノベーション」を起こさなくてはならないと痛切に感じた、今回のシンポジウムであった。

## パネル参加報告

# 「ISO17100 と日本における翻訳職業教育の将来」 パネルディスカッション参加報告

機械翻訳課題調査委員会

ISO17100 Translation Services - Requirements for translation services は本年には制定されるべく検討が進められている。この規格には職業翻訳者の認定基準の一つに翻訳の学位保持者であることが盛り込まれる予定であるが、日本国内にはこの基準を満たす学位を授与できる教育機関が存在していない。2014年6月にドイツのベルリンで開催された TC37: Terminology and other language and content resources 総会の参加者による最新情報を共有するのみならず、日本における職業翻訳者育成の仕組みを考えるために日本翻訳連盟(JTF)により掲題のパネルが開催された。JTF 理事会、JTF内 ISO 検討会を中心に、AAMT 課題調査委員会も参加の招待を受けた。

まず株式会社翻訳センター社員であり、ISO TC37 国内委員である田嶌奈々氏による「TC37 ベルリンでの議論」と題した出張報告が実施された。この報告においてベルリン総会でもたらされた大きな変更点が紹介された。「第3章:翻訳者の要件」において挙げられていた4項の内の1つ「政府認定の資格」が削除されたとのことだった(その他3項:「翻訳の学位」、「翻訳以外の学位+翻訳経験2年」、「翻訳経験5年」)。日本には政府認定の資格がないため政府以外の機関を認めるよう以前から要請していたが、他国から受け入れられなかった。しかしベルリン総会ではISO適合性評価委員会が介入し、該当項を完全に削除することで最終的に決定したとのことであった。

そして、以下のメンバーによるパネルディスカッションが開催された。

- · Tony Hartley (英リーズ大学翻訳研究所教授)
- ・武田珂代子(立教大学異文化コミュニケーション学部・研究科教授)
- ・田嶌奈々 (㈱翻訳センター、ISO TC37 国内委員)

モデレータ: 影浦峡 (東京大学大学院教育学研究 教授)

先生方の大学または大学院での「翻訳」に関連する課題および今後取り組むべき多くの事項について紹介された。

- 有資格の指導者不足
- 翻訳学位の形骸化に対する危惧など日本の大学または大学院で翻訳学位プログラムを創設する上での課題
- 形だけになりつつある ISO についてなどの制度的・枠組み的側面
- ・ 翻訳教育における TM (翻訳支援ツール) / MT (機械翻訳システム) 導入の必要性と問題
- ・ 専門分野における知識の重要性

この他にも参加者からパートナー企業の必要性、単価 についてなど翻訳者が直面する現実的な質問が挙げら れ、熱のこもったパネルとなった。

## シンポジウム案内

## 「第24回 JTF 翻訳祭」のお知らせ

### 一般社団法人日本翻訳連盟

理事·翻訳祭企画実行委員長 中尾 勝

## < 「第24回 JTF 翻訳祭」開催のご案内>

日本翻訳連盟では、来る 11 月 26 日 (水)、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、「2020年へ 新翻訳時代の 幕開け〜Break The Paradigm, Shape The Future〜」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」(全30 セッション)、「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。交流パーティーで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機会が生 まれることを願っております。

翻訳祭に是非ともご出席賜りますよう、よろしく お願い申し上げます。

# ★☆★ 第 24 回 JTF 翻訳祭 ★☆★【日時】

2014年11月26日(水)9:30~20:30 (開場9:00)

#### 【場所】

「アルカディア市ヶ谷(私学会館)」 ※JR・地下鉄「市ケ谷駅」より徒歩2分

## 【主催・運営】

一般社団法人日本翻訳連盟 第 24 回 JTF 翻訳祭企画実行委員会

#### 【後援・特別協力】

日本翻訳者協会 (JAT)

#### 【後援】

経済産業省・アジア太平洋機械翻訳協会(AAMT) 他

#### 【概要】

今年翻訳祭は24回目を迎えます。1991年に産声をあげたこのイベントも、まもなく誕生から4半世紀が経とうとしています。年々規模は拡大、昨年は824名の参加を数えるに至りました。会場も初期のマツダホールからアルカディア市ヶ谷に移し、マルチセッション式で7会場を同時進行させるまでの規模へと成長しています。

来る 2020 年に開催が決定した東京オリンピック。この大きなイベントに向けて日本中が激しく揺れ動いています。言葉の世界もまた揺れ動いています。今年の翻訳祭のテーマは「2020 年へ 新翻訳時代の幕開け」。昨年は「大翻訳時代」と位置づけ、いまだ参入者のいない未開拓の市場 Blue Ocean を探りましたが、今年は時代の胎動を感じさせる「新翻訳時代の幕開け」と名付けました。そしてサブテーマを「Break The Paradigm, Shape The Future」とし、「既成概念を壊して、新しい時代を創っていこう」と高らかに目標を掲げました。

このテーマの元に 24 の講演&パネル・ディカッション、6 つのプレゼン・製品説明コーナー、30 社 あまりが出展する翻訳プラザ (展示会) が用意されます。セッション終了後の交流パーティーには 300 名の翻訳者、翻訳会社、業界関係者も集います。情報交換をし、ビジネスチャンスの開拓にもとても有

効です。

飲み物コーナーも充実させ、より寛げる雰囲気 も。今年の翻訳祭にぜひご参加いただければと思い ます。

今年の翻訳祭では、AAMT様にパネルディスカッションを企画していただきました。

16:30~18:00 (90分)

「いまさら聞けない機械翻訳

~動作原理から上手な活用法まで~」

<パネリスト>

秋元圭氏(AAMT機械翻訳課題調査委員会委員) 内山将夫氏((独)情報通信研究機構主任研究員) 山田優氏(㈱翻訳ラボ代表、麗澤大学・青山学院大 学ほか兼任講師)

<モデレーター>

井佐原均氏(豊橋技術科学大学教授、AAMT 理事)

#### 【総来場者数 (見込み)】

850名 (前年度実績 824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 18:30~20:30 (120 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 33 社 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
- ・翻訳プラザに隣接した特設会場にて、各社のプレゼン発表を受講!
- ・プレゼンコーナーを増設し、65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)。

## 【お問い合わせ】

一般社団法人 日本翻訳連盟 事務局 http://www.jtf.jpE-mail:info@jtf.jp

## AAMT 長尾賞受賞講演

## 言語構造が大きく異なる言語間の長文を対象とする統計的機械翻訳の実用化

(独) 情報通信研究機構多言語翻訳研究室 内山将夫 隅田英一郎

## 1. はじめに

本稿では、第9回 AAMT 長尾賞受賞を記念して、(独)情報通信研究機構 (NICT)における「語順変換と訳語選択の独立実行による統計的機械翻訳 (SMT)技術の研究ならびに複数企業への技術移転[1][2][3][4]による高性能機械翻訳システムの実用化」に関して解説する。

## 2. NICT における長文対象の SMT の開発

SMTの開発には、対訳コーパスとアルゴリズムの双方が重要である。本受賞に関する SMT の最初の開発は10年以上前に遡るため、その開発に影響したことを順に紹介する。

## 3. 対訳コーパスの開発

まず、2000年には、日英大規模対訳コーパスは存在しなかった。2000年の頃は、フレーズベースSMTが世界的に研究されだした頃であった。当時は、英仏の大規模対訳コーパスはあったが、日英の大規模対訳コーパスはなかった。この背景の中で、2001-2003年度科学技術振興調整費若手任期付研究員支援「言語横断文検索」(内山将夫)の成果として、「日英新聞記事対応付けデータ(JENAAD)」を作成し、一般公開した[5]。

5年後の、2005年になっても、まだ、大規模日英対訳コーパスは存在しなかった。しかし、2006-2010年度科学技術振興調整費重要解決型研究等の推進「日中・中日言語処理技術の開発研究」(代表者井佐原均)において、前半部で300万文規模の英日対訳コーパスを作成し、全体を通じて、100万文規模の日中対訳コーパスの構築を進めた。なお、本プロジェクトで作成された対訳コーパスは、現在、ASPECコーパス[6]として公開されていて、それを利用したワークショップ WAT が開催された[7]。

2007年頃になって、日米特許対訳コーパスが作成された。そのきっかけは、NTCIR-6において、日米の特許10年間分が研究利用可能になったことである[8]。これを利用することにより、日米同時出願特許から180万文規模の対訳コーパスを作成した[9]。これにより、初めて、英仏等と同等規模の日英対訳コーパスが SMT の研究に利用可能となった。

NTCIR-7では、この対訳コーパスを活用して、特許翻訳タスクが開催された[10]。その結果、規則ベースの MT の優位性が確認された[10]。

## 4. 長文翻訳精度の向上

SMTでは、英日・日英翻訳など構造の大きく違う言語間の翻訳精度は、低かった。

NICTでは、旅行会話用の音声自動翻訳エンジンの開発を進めていて、2008年には実用化していたが、そのエンジンはフレーズベース SMTを改良したもので、日英・英日の長文翻訳の精度は低かった。また、NICTでは、2010年ごろに、アパレル等の特定分野向けの SMTエンジンを開発・実用化していたが、これは分野の特徴を利用した語順の並べ替えにより高精度化を図っていたため、一般的な長文の翻訳精度は低かった。

更に、NICT だけでなく、SMT 全体においても、日英・英日翻訳における長文翻訳 精度は低かった。

このような背景の中、NTCIR-9,10 で、英 日 SMT の精度向上が確認された[11][12]。

まず、NTCIR-9で、日本語が Head Final であることを利用して、前もって英語を日本語に近い語順に変換し、それをフレーズベース SMT で翻訳する方式[13]の精度が、ルールベースの英日翻訳と匹敵か少し上回る性能を達成した。

次に、NTCIR-10 で、同手法を採用した 英日 SMT の精度が、ルールベースの手法の 翻訳精度を明確に上回った。

NICTにおいては、同手法にインスパイア されて、多言語に適用可能な語順変換の手 法を開発した[14]。

その方法は、

- (1) 入力文を構文解析して2分木にする
- (2) 2分木の各ノードにおいて子供のノードを入れかえるべきノードを同定する というものである。

上記の方法は単純なものだが、複雑な語順 変換もサポート可能である。たとえば、 FIG. 3C is a graph illustrating a simulation that includes the effects of resonance, cyclic clocks, and a change in logic current.」について、上記の手法を 適用すると、「FIG. 3C \_val\_ resonance of effects, cyclic clocks, and logic current in change \_va2\_ includes that simulation \_va2\_ illustrating graph is.」のように日本語風の語順に変換され るので、それを SMT への入力とすると、「図 3 Cは、共振による効果、環状のクロック、 および論理電流の変化を含むシミュレーシ ョンを示すグラフである。」のように機械翻 訳される。なお、「\_va1\_」や「\_va2\_」は、 日本語の助詞に相当するものを語順変換の 後で挿入したものである。

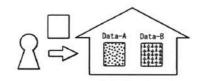
この手法を入力に適用したあとで、フレーズベースの SMT 手法を適用することで、翻訳性能が質的に向上した。具体的には、英日特許翻訳について、2010年以前の翻訳精度と現在の翻訳精度を比べると、以前は、端的に言って、意味不明な翻訳が多かった。一方、現在は、対訳コーパスが十分にあれば、意味が分かる翻訳文を出力することが可能になった。

#### 5. みんなの自動翻訳@TexTra®

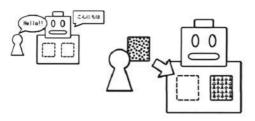
NICT で開発した SMT エンジンは「みんなの 自動翻訳@TexTra®」サイトにて一般公開さ れている。

(<a href="https://mt-auto-minhon-mlt.ucri.jgn-x">https://mt-auto-minhon-mlt.ucri.jgn-x</a>
<a href="mailto:jgn-x">. jp/</a>)

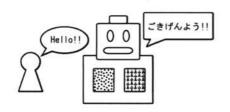
本サイトでは次のようにして NICT の SMT エ ンジンをカスタマイズできる。 ①自分で用意した用語や対訳データを登録



②登録したデータを用いて SMT をカスタ マイズ



③作成した SMT エンジンを利用



また、登録した対訳データからオリジナルの SMT エンジンを作成可能である。

### 6. 今後の展開

SMT の技術は、頭打ちには程遠い状況に あり、今後、ますますの精度向上が望める と確信している。

英日 SMT の精度向上は、言語構造の違い を構文解析により克服することにより達成 された。一方、日英 SMT は、改善の余地が 大きい。

その理由としては、日本語には、文脈からわかることは表現されないという特徴があることと思われる。今後は、このように必ずしも表出されない単語等の翻訳をどうするかが重要であると考える。

また、英日に限らず、SMT 一般の問題の解決も重要である。

### 7. 謝辞

NICTの技術を評価し実用化してくださった ATR-Trek、FEAT、KDDI、株式会社バオバブ、日本特許情報機構、日本発明資料株式会社、川村インターナショナル株式会社、凸版印刷株式会社、株式会社ロゼッタの関係者に深く感謝する。

## 8. 参考文献

[1]NICT の高精度な中日自動翻訳ソフトウェアが Japio のサービスに〜中国特許を日本語で検索・表示できる自動翻訳技術が、ついに実用化! ~、2013 年 3 月 28 日

http://www.nict.go.jp/press/2013/03/28-1.h tml

[2] "英語特許文"の高精度「自動翻訳ソフトウェア」を開発、~NICT とニッパツが特許向け英日翻訳技術を共同開発し、5月にサービス開始~、2013 年 3 月 21 日 http://www.nict.go.jp/press/2013/03/21-1.html

[3] "IT 英文マニュアル"専用の「自動翻訳 ソフトウェア」を開発 ~川村インターナショナルと NICT が英日翻訳技術を共同開発 し、10月からサービス開始~、

2013年10月1日

http://www.nict.go.jp/press/2013/03/28-1.h tml

[4]凸版印刷、低コストで高品質な翻訳受託 サービスを開始、~NICT と共同開発した翻 訳支援システムを活用、企業のグローバル 対応を支援~、2013 年 11 月 14 日 http://www.toppan.co.jp/news/2013/11/ne wsrelease131114.html [5] Masao Utiyama and Hitoshi Isahara.
(2003) Reliable Measures for Aligning Japanese-English News Articles and Sentences. ACL-2003, pp. 72--79.

[6] 科学技術振興調整機構、情報通信研究機構 (2014) Asian Scientific Paper Excerpt Corpus.

## http://lotus.kuee.kyoto-u.ac.jp/ASPEC/

[7] Toshiaki Nakazawa, Hideya Mino, Isao Goto, Eiichiro Sumita, and Sadao Kurohasi. (2014)

The 1st Workshop on Asian Translation. http://orchid.kuee.kyoto-u.ac.jp/WAT/

[8] Atsushi Fujii, Makoto Iwayama, and Noriko Kando. Overview of the Patent Retrieval Task at the NTCIR-6 Workshop. Proceedings of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access, pp.359-365, May 2007.

[9] Masao Utiyama and Hitoshi Isahara.
(2007) A Japanese-English Patent
Parallel Corpus. MT summit XI, pp.
475--482.

[10] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto. and Takehito Overview of the Patent Translation Task at the NTCIR-7 Workshop. Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval. Question Answering and Cross-lingual Information Access, pp.389-400, Dec. 2008.

[11] Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita, Benjamin K. Tsou. (2011) Overview of the Patent Translation Task at the NTCIR-9 Workshop.

[12] Isao Goto, Bin Lu, Ka Po Chow,Eiichiro Sumita, Benjamin K. Tsou.(2013) Overview of the Patent TranslationTask at the NTCIR-10 Workshop.

[13] Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2010. Head finalization: A simple reordering fule for SOV languages. In Proc. of the Joint Fifth Workshop on Statistical Machine Translation.

[14] 特開2013-250605: 機械翻訳装置、機械翻訳方法、およびプログラム

## AAMT長尾賞学生奨励賞

## 自治体文書の多言語化を支援する枠組みとシステム環境の研究

東京大学大学院 教育学研究科 図書館情報学研究室 宮田 玲

この度は、2014年度 AAMT 長尾賞学生奨励賞を 頂き、大変光栄に存じます。私の専攻している「図 書館情報学」という、機械翻訳から一見離れた分野 でなされた研究をこのように評価していただいた ことを大変嬉しく思うと同時に、本受賞を励みにこ れからも地道に研究に邁進していきたいと思いま す。また、今回受賞対象となった修士論文「自治体 ウェブサイト文書の多言語化を支援する枠組みと システム環境の研究」を執筆するにあたり、指導教 員の影浦峡先生、本研究の出発点となったアイディ アを下さった Anthony Hartley 先生をはじめ、多く の方にお世話になりました。この場をお借りして、 お礼を申し上げます。

ここで、図書館情報学という分野からなされた本 研究の立場について少しだけ説明します。図書館情 報学は、伝統的には図書館という社会機関や、図書 (本)・文書 (ドキュメント) といった物理的にパ ッケージされた媒体を主な研究対象としてきまし た。近年はテキストの電子化を背景に、章・節・段 落・単語・形態素・文字と分析の粒度を上げ、ゆる やかに自然言語処理・計算言語学と関係を保ちなが らも、テキストの存在様態をあくまで現実的な社会 に位置づけられたものとして捉える視点を、意識的 にせよ、無意識的にせよ、設定してきたといえます。 そして、そのような視点に立つと、これまで自然言 語処理・計算言語学が主に進めてきたような文法規 則と語彙集合の公理的な組み合わせによる、言語表 現の強力な生成・変換処理とは別に、図書・文書・ テキストの側から言語表現を定位していくことは、 極めて自然な流れであり、それが本研究に通底する 視座となっています。あえて単純化し図式的に言い 換えると、下から言語を組み上げつつ、上からその 可動域をおさえることで、「現実的にありうる」表 現範囲を捕捉しようとするものです。

とはいうものの、図書(あるいはさらに視点を外にずらして社会・文化)というレイヤーからいきなり個々の言語表現に接近することは容易ではありません。そこであくまで形式的に操作可能な文字テキストに踏みとどまり、「文書」という情報ユニットに着目し、そこに見られる言語表現を観察し、その振れ幅を記述していくことにしました。そして、いかにして機械翻訳という強力なテクノロジーをその中に位置づけ、制御し、活用していけるのかを見定めたい、と無謀にも意気込んでいました。

この修士論文を通じて、社会に織り込まれたテキストの複雑な布置の中では、「私たちは驚くほど、自由には書くことができない」ということを、まざまざと思い知らされました。そして、この「不自由さ」を少しでも言葉にかえて解きほぐしていきたいという問題関心のもとで、様々な研究分野に手を出し、視界不良の中でじたばたしながらも、何かを掴んだ手触りとその切れ端は確かに残っており、まとまりがなく、得体のしれない修士論文の中から、その断片を丁寧に拾い上げ、評価していただけたことは、この上なく感謝すべきことだと感じています。

前置きが長くなってしまいましたが、本稿ではまず受賞対象となった修士論文の概略を述べます。限られた紙面の中では、指示的な要約に留まる箇所も多くなってしまうため、後半では、改めていくつかの要点を取り出し、背後にある問題関心に立ち戻りつつ、本研究の位置づけ・意義を説明します。そして、最後に本研究を超えた先の、長期的な課題と展望を示したいと思います。

### 1. はじめに

#### 1.1. 研究の背景

自治体は、日本語を母語としない外国人住民に対しても、生活や命に関わる情報を、漏れなく正確にかつ効率的に提供することが求められる。正確な情報伝達を担保する上では、まず文書の構造化と効果的な情報の配置といった文書構造レベルの設計が必要である。「どのような情報をどのような順序・構造で配置するか」といった文書設計に関する問題は、一般的にはテクニカルライティング、情報デザイン、情報アーキテクチャ、ドキュメントデザインといった領域で扱われてきたが、自治体ウェブサイト文書といった個別具体的な文書クラスにおける文書設計の技術は、現場レベルでの暗黙知として留まりがちである。

また自治体は、これらの情報を多言語で展開することが要求される。コスト・時間的な制約から、人手による高品質の翻訳を網羅的に用意することは難しく、機械翻訳(Machine Translation: MT)を導入する自治体ウェブサイトが増えている。近年、機械翻訳の精度は向上しつつあるが、日本語の構造が他の多くの言語と異なるために、依然として実用レベルには達しているとは言いがたい。

これまで機械翻訳の効率改善手法の1つとして、起点テキストの語彙・文法・スタイルに一定の制限をかける「制限言語(Controlled Language: CL)」のアプローチが有効であるとされてきた(Roturier、2004)。制限言語の研究は、マニュアル等の技術文書を多言語に翻訳するニーズに応じて、1970年代頃から特に欧米圏の産業界の主導で進められてきた。日本語でも近年、「産業日本語プロジェクト」など、主に産業文書(特に特許文書)の翻訳における日本語の規格化が進められている(渡邊,2010)。しかし、日本語に関しては、具体的な研究としての知見が十分に蓄積されておらず、また評価手法も定まっていないため、実用レベルでの運用事例も限られている。

また関連して、翻訳のクオリティコントロールの 観点からは、適切な情報資源と自治体固有の用語の 管理を相乗的に考える必要がある(Wright & Budin, 2001)。自治体文書に頻出する固有表現や専門用語 は、原文の理解のみならず、翻訳品質に決定的な影 響を与える。そのため一貫した用語の使用を支援す る統制された対訳用語集の整備が急務である。

一方、実運用上の場面では、執筆者が要件に沿った形でスムーズに文書を作成できるような機械的な支援システムが求められる(長尾ほか、1984; Hartley & Paris、1997)。実際に文書を作成するのは現場の人であり、円滑な文書多言語化の理念的な枠組みを構想するだけでなく、それを効率的・効果的に運用できるような仕組み・ツールを提供することが重要である。

#### 1.2. 研究の問い

以上の背景を踏まえて、本研究では大きく3つの 問いを提示した。

- 1) 自治体ウェブサイト文書の多言語化における 課題は何か。またその状況を記述、分析、解 決する上で、有効な枠組みは何であるか(枠 組みの検討)
- 2) 自治体手続きに関する文書を執筆・多言語化 する上で、文書構造と言語表現をいかにコン トロールすべきか(枠組みの適用)
- 3) そのような文書構造・言語表現のコントロールを執筆時に支援するシステム環境はいかなるものか(工学的な応用)

また、2)の問いを、さらに書き下すと以下の2点にまとめられる。

- 2-a) 読み手(住民)が特定の自治体手続きを確実 に遂行するために必要とされる情報とその配 置をいかに定義し、文書上で表現できるか
- 2-b) 上記を満たした上で、そのような情報を起点 言語でも目標言語でも、正確かつ理解可能な 形で読み手に提示するためには、いかなる言 語表現上のコントロールが有効か

#### 2. 論文の概要

#### 2.1. 枠組みの検討

問いの1) に対応して、多言語化のニーズの拡大や 機械翻訳の導入が進んでいる社会的状況を踏まえ つつ、自治体ウェブサイト文書においては、文書の 内容構成の品質、起点言語・目標言語における文章 表現の品質が、大きな課題であることを整理した。 続いてこれらの課題に取り組むための足場とし て、図書館情報学の立場を確認した。文書を把握す るときに前提となっている認識的な基盤を段階的 に点検しながら、また知識を捉える見方を図書分類 法の考察から整理しながら、図書館情報学は、認知 でも、言語でも、論理でもなく、人間の知識コミュ ニケーションを扱おうとしてきたことを指摘した。 またそのためには構造規定的に「現実的に存在可能 な」知識 (メディア) を捉えつつ、部分的に要素構 成的な形式化にも踏み込んでいくことが必要であ る、という点を確認した。

この理論的な立場を前提として、機能的側面に着 目しながら文書ユニットを捉えていくことを宣言 しつつ、自治体ウェブサイト文書の課題に取り組む 上で参照すべき先行研究を検討した。

最終的には、文書構造レベルの枠組みとして神門 (1992) の提唱する機能構造と技術文書の XML 標準規格である DITA (Darwin Information Typing Architecture) (Hackos, 2010) を主に参照し、言語表現レベルの枠組みとして制限言語とテクニカルライティングを参照することを定めた。

### 2.2. 文書構造レベルの研究

問いの 2-a) に対応して、文書構造レベルの研究として、自治体手続き型文書(「転出届の出し方」や「印鑑登録の方法」など、行政手続きに関する文書のこと)の機能構造を探索的に分析し、暗黙知に留まりがちであった自治体手続き型文書の構造を、文書の社会的・法的な位置づけと関連させながら、明らかにした(図1)。

具体的には自治体国際化協会の「多言語生活情

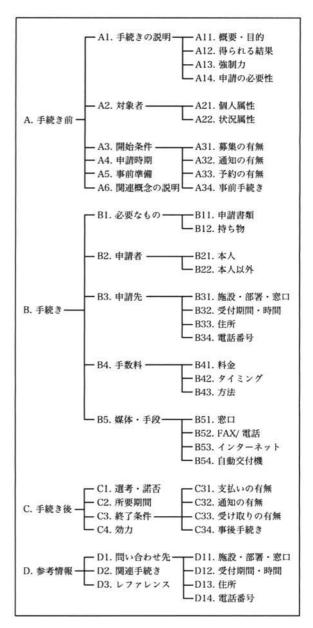


図1: 自治体手続き型文書の構成要素カテゴリー

報」、新宿区の「生活情報」、浜松市の「カナル・ハママツ」のウェブサイトから、合計 123 の自治体手続き型文書を選定し、修辞的・機能的な構成カテゴリー(「対象者」「開始条件」「必要なもの」「申請者」「手数料」「所要期間」等々)を洗い出し、時系列に沿って整理しなおした。これにより、自治体手続き型文書には、「どのような内容がどのような配置で書かれてきたか」について明示的に取り出すことができた。そもそも自治体文書を対象とした機能構造の分析はこれまでなされておらず、本研究を足が

かりにしたさらなる研究が期待される。

さらに、この機能構造を、より一般的な「手続き型」文書として定義されている DITA の「タスク・トピック」にマッピングすることで、機能構造分析結果の妥当性を検証すると同時に、文書作成工程への適用可能性を議論した。一定の配置換えをすることで、DITA 構造に十分適合することが明らかになった(表 1)。

表 1: DITA マッピングの結果 (一部省略)

DITAタスク本文	機能構造
事前条件 (prereq)	「A2. 対象者」「A3. 開始条
	件」「A4. 申請時期」「A5. 事
	前準備」
背景情報 (context)	「A1. 手続きの説明」「A6.
	関連概念の説明」「D1. 問い
	合わせ先」
手順 (steps)	
操作 1 (step)	「B12. 持ち物」を用意する
操作2(step)	「B2. 申請者」を指定する
操作3(step)	「B3. 申請先」に行く
操作 4 (step)	「B11. 申請書類」を書く
操作 5 (step)	「B12. 持ち物」と「B11. 申
	請書類」を提出する(必要
	に応じて「B4. 手数料」を
	指示する)
期待結果(result)	「C1. 選考・諾否」「C2. 所
	要期間」「C31. 支払いの有
	無」「C32. 通知の有無」
	「C33. 受け取りの有無」
実行例 (example)	該当なし
タスク完了後の操	「C34. 事後手続き」「C4.
作(postreq)	効力」「D2. 関連手続き」

## 2.3. 言語表現レベルの研究

問いの 2-b) に対応して、言語表現レベルの研究として、機械翻訳を前提とした自治体ウェブサイト文書の多言語化の課題に、制限言語の枠組みから取り

組んだ。

経験的に集積されてきたテクニカルライティング の知見を整理・再構成することで、起点言語の読み やすさと機械翻訳の品質を両立させる制限言語ル ールの構築を試みた。

まず文章技術に関する書籍 17 冊から文章作成に 関するルールを 47 種類収集し、それらの仮評価を 踏まえて、最終的に 22 種類の制限言語ルールを定 式化した (表 2)。

続いて、各ルールに対して、次の3つの観点から 評価した。

- ① 翻訳英文の理解容易度・正確性(目標テキストの品質)
- ② 日本語文の読みやすさ(起点テキストの品質)
- ③ ルールの実行可能性(運用場面における使いやすさ)

人手による評価実験の結果、日本語の読みやすさに対しては大幅な効果が見られ、翻訳品質も部分的に向上することが明らかになった。例えば、ルールq(サ変名詞にはなるべく「行う」を付けないでください。)により、「行う」が受動態を使って訳された「is performed」という不自然な表現が、シンプルな能動態に置き換わることで、翻訳精度が向上している。また機械翻訳の方式(統計的機械翻訳かルールベース機械翻訳か)によって、制限言語の効果に違いが見られ、両方式において翻訳精度の向上に貢献したルールは b, e, p, q の 4 つのみであった。本実験により、機械翻訳の種類や機能構造に応じたルールの選択とさらなる精緻化が必要であると診断し、改善の方向を示した。

本評価の遂行にあたり、現実の翻訳利用者を見据 えた機械翻訳(制限言語)の評価手法を考案した。 参照翻訳文を用いながら、2 段階で質問し、その評 価結果を 4 つの結果カテゴリー、すなわち(1)機 械翻訳文が理解でき、かつその内容も正しい、(2) 機械翻訳文は理解できるが、その内容が誤ってい る、(3)参照翻訳文は理解できるが、機械翻訳文は

表 2: 制限言語ルールとサンプル文

No	書き換えルール	書き換え前・後の日本語文サンブル
a	一文はできる限り 70 文字以内におさめ	[前] 豊橋の民営クラブの多くは、昭和 50 年代より保護者同士の助け合いの中で先駆的に設置
	てください。それ以上になる場合でも、	され、平成5年度に最初の公営クラブを開設する以前から運営が続けられています。
1	100 文字以内にはおさめてください。	[後] 豊橋の民営クラブの多くは、昭和 50 年代より保護者同士の助け合いの中で先駆的に設置 されました。平成 5 年度に最初の公営クラブを開設する以前から運営が続けられています。
b	箇条書きで書くときは、列挙項目の前後	[前] その他では、
	の文を完結させてください。	・総合動植物公園での動物飼育の仕事
		・保健所で迷子になった犬、ネコの保護や狂犬病予防に取り組む仕事
		があります。
		[後] その他では、以下の仕事があります。 ・総合動植物公園での動物飼育の仕事
		・保健所で迷子になった犬、ネコの保護や狂犬病予防に取り組む仕事
С	文の中に、括弧書きで長い説明を入れな	[前] ゴミ袋(40ℓ 程度)を二重にして、中に半分程度の水(風呂の残り水を使うと便利)を入
	いでください。	れ、玄関などにすき間なく並べる。
		[後] ゴミ袋(40ℓ 程度)を二重にして、中に半分程度の水を入れ、玄関などにすき間なく並べる。水は、風呂の残り水を使うと便利です。
d	主語と述語の関係を明確にしてくださ	「前」音楽科の 1、2 年生により演奏され、プログラムから演出、アナウンスまですべて生徒が
	V).	作ったものです。
	32.79	[後] 音楽科の 1、2 年生により演奏されました。プログラムから演出、アナウンスまですべて
		生徒が作ったものです。
e	修飾語と被修飾語の関係を明確にしてく	[前] 豊橋・田原地域の地産地消の拠点として、安全安心で新鮮な地場農産物を販売する東三河で最大級の農産物直売施設です。
	ださい。	で戦入級の襲産物直元施設です。 「後」豊橋・田原地域の地産地消の拠点として、安全安心で新鮮な地場農産物を販売する、東三
		河で最大級の農産物直売施設です。
f	「が」を使って文をつなげるのは、「しか	[前] 朝市ですが、豊橋のどこでいつ開催されているか詳しく教えて下さい。
	し」の意味を持つ場合だけにしてくださ	[後] 朝市は、豊橋のどこでいつ開催されているか詳しく教えて下さい。
-	い。 「ので」の意味で「ため」を使わないで	[前] 地震で地盤が揺れると土砂が水とともに液体のように流れ動くため、地盤の液状化現象と
g	ください。「ので」を使ってください。	いわれます。
	viceva, orej elebericeva	[後] 地震で地盤が揺れると土砂が水とともに液体のように流れ動くので、地盤の液状化現象と
		いわれます。
h	「from」を意味するときは「~から」を	[前] 平成21年度より整備を進めてまいりました豊橋市南消防署西分署庁舎が、このたび竣工
	使ってください。「より」は比較のとき	いたしました。 [後] 平成 2 1 年度から整備を進めてまいりました豊橋市南消防署西分署庁舎が、このたび竣工
	だけ使用します。	[後] 下成と1年及から返溯を進めてよいりよした豆飼巾用作奶看四刀看月音か、このたい液上いたしました。
i	1 つの文の中で複数の否定形を使わない	[前] なお、期限内に手続きをしないと、受給資格があっても手当を受けることができなくなる
	でください。	場合がありますので、ご注意ください。
		[後] なお、受給資格がある場合でも、手当を受けるには、期限内に手続きをする必要がありま
-1	可能や尊敬の意味で「~れる」「~られ	すので、ご注意ください。 [前] 立っていられず、ブロック塀が壊れる。
j	る」を使わないでください。	[制] 立うていられり、プロック州が破れる。 [後] 立っていることができず、ブロック塀が壊れる。
k	複数の意味に解釈できる言葉ではなく、	[前] 木の枝は、60 センチ以下に束ねて出しましょう。
	なるべく明確な意味を持つ言葉を使って	[後] 木の枝は、60 センチ以下に束ねて捨てましょう。
	ください。	(20) 卢林市被告的《同志传》中《日本》同志传》中,为本土
1	口語表現の「~になります」表現を避け	[前] 自然史博物館の展示物の中で最大の展示物になります。 [後] 自然史博物館の展示物の中で最大の展示物です。
m	てください。 「~という」表現はなるべく省いてくだ	[前] 1ヶ月単位で入院費が安くなるという制度があると聞きました。
	さい。	[後] 1ヶ月単位で入院費が安くなる制度があると聞きました。
n	「ような」、「こと」、「もの」はなるべく	[前] 近くの職場同士で協力し合うものとする。
-	省いてください。	[後] 近くの職場同士で協力し合う。 [前] 応募の際には各試験ごとの募集要綱で確認してください。
0	1 つの文の中で、同じ語句や重複した意味を持つ語句を使わないでください。	[制] 心券の際には各試験の募集要綱で確認してください。 [後] 応募の際には各試験の募集要綱で確認してください。
р	「思われる」「考えられる」は必要なとき	[前] 民間企業も含め、今後駅前の開発に様々な事業者が関与する可能性も考えられます。
	以外は省いてください。	[後] 民間企業も含め、今後駅前の開発に様々な事業者が関与する可能性もあります。
q	サ変名詞にはなるべく「行う」を付けな	[前]また、願書は、消防本部予防課、中消防署及び南消防署にて配布を行っています。
r	いでください。 「~したり、」を使うときは列挙項目す	[後]また、願書は、消防本部予防課、中消防署及び南消防署にて配布しています。 [前] 屋根瓦やトタンがめくれたり壊れてないか
RTS	べてに「したり」を付けてください。	[後] 屋根瓦やトタンがめくれたり壊れたりしていないか
s	項目を並べるときは、品詞や表現をそろ	[前] 生徒・児童等はあらかじめ学校等で定められた方法によって帰宅、又は保護者に引き渡
	えてください。	<b>†</b> .
		[後] 生徒・児童等はあらかじめ学校等で定められた方法によって帰宅させるか、又は保護者に
t	なるべく標準的な和英辞典に載っている	引き渡す。 [前] ブラジル・パラナ州経済視察団一行が来豊し、市長表敬を行います
	語を使ってください。	[制] ファンル・パファ州経済視察団一行が豊橋市に来訪し、市長表敬を行います
u	サ変名詞をつなげた複合語を避けてくだ	[前] 愛知県内の同システムに掲載参加している全自治体分を検索機能により見られます。
	さい。	[後] 愛知県内の同システムに掲載・参加している全自治体分を検索機能により見られます。
v	誤字、脱字がないように注意してくださ	[前] 毎年8月に中央図書館にて「平和を求めて」と題して、パネル・写真展行っています。 「終」毎年8月に中央図書館にて「平和を求めて」と題して、パネル・写真展を行っています。
	い。また、同音異義語や助詞の抜けにも	[後] 毎年8月に中央図書館にて「平和を求めて」と題して、パネル・写真展を行っています。
$\overline{}$	注意してください。	

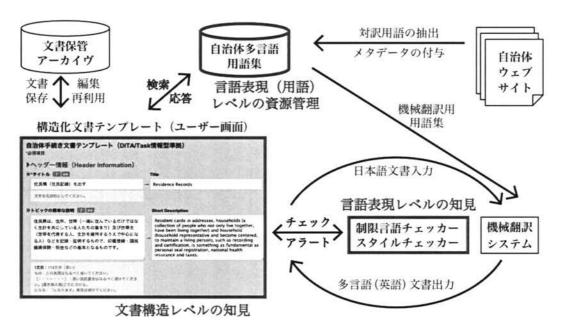


図 2: 文書作成支援システムの全体像

理解できない、(4)参照翻訳文自体が理解できない、の4つに分類することで、機械翻訳に関する専門的な知識を前提とせずに、機械翻訳の有用性を査定することが可能となった。とりわけ(2)機械翻訳文は理解できるが、その内容が誤っている、という場合は、誤り自体への認識を隠蔽し、正しい理解への回復の契機が閉ざされている点で、より弊害が多いといえよう。

さらに、第三者による原文書き換え評価実験により、用意した 120 の日本語文のうち約 90%において、筆者らが想定した書き換えが遂行されていたことからも、今回定めた 22 種類のルールは一般の執筆者でも概ね運用可能であることが実証された(ただし、本実験は 22 種類のルールを個別に適用する限定的な統制条件下で実施された点は留意する必要がある)。しかし一方で、ルールに違反した箇所の同定に失敗した事例や、そもそもルールの指示する内容を理解していない事例も見受けられた。また、ルールの説明に用いた「サ変名詞」といった文法用語が、執筆者の混乱を招きかねないことも、実験協力者からのフィードバックから明らかになった。

以上の考察を踏まえ、制限言語ルールの一層効果 的な運用のためには、システム面での支援が必要で あると指摘した。

## 2.4. 文書作成支援システムの開発

最後に問いの3) に対応して、システム環境の側から自治体ウェブサイト文書(特に自治体手続き型文書) の作成を支援する方略を考察した。

既存のシステムの機能・インタフェースを整理した上で、「日本語を母語とし目標言語の知識を前提としない一般的な自治体職員」を想定しながら、多言語文書作成を支援する統合的なシステム環境を設計した(図 2)。具体的には「2.2. 文書構造レベルの研究」の成果を構造化テンプレートと DITA 出力機能として、「2.3. 言語表現レベルの研究」の成果を制限言語チェッカーとして部分的に実装した。また加えて、日英対訳の自治体用語集(2027 語)を試行的に構築する中で、可用性の高い言語資源のあり方を提案した。

これらの仕組みはどれもプロトタイプであり、今後の開発・評価・改善が必要であるが、正確で漏れのない多言語文書を作成する上で有効であると考えられる。

#### 3. 本研究の要点

#### 3.1. 翻訳の上流工程のコントロール

原文 (起点テキスト) の執筆から多言語の目標テ キストの作成までを、翻訳工程として含めたとき、 主な介入点としては、原文執筆(制限言語)・前編 集・機械翻訳・後編集を便宜的に挙げることができ、 それぞれ比較的独立した研究として取り組まれて きた。このうち、言うまでもなく、機械翻訳自体は、 継続的に改良され、日々進歩している。また後編集 の研究も盛んであり、機械翻訳系・言語処理系の学 会でも多くの研究が発表されている。最終的に読み 手に届く直前のテキストの品質を高める後編集は、 成果物の品質に直結する工程であり、現実的なニー ズも高い。それに対して、前編集や制限言語は、そ の後の工程に機械翻訳による処理(さらには後編 集)が介在するため、その効果を明確に示すことが 難しい。また制限言語のように原文執筆工程に手を 加えることは、文書作成のワークフロー全体の見直 しを迫るものであり、導入コストも比較的高いとい えよう。

それにもかかわらず、本研究で制限言語などの上流工程のコントロールに着目したのは、(1) そもそも原文段階での情報提示やテキスト品質に問題があり、改善の必要があること、(2) 多言語化の効率をトータルで改善するには元となる原文を統制することが有効であること、を理由として挙げることができる。

引き続き、翻訳の上流工程でのコントロールをより効果的に実現する上で、とりわけ重要だと考えられるポイントを3つ提示する。

#### 3.2. 文書構造と言語表現

井佐原ら (2012) は、機械翻訳を用いた産業文書の多言語展開を支援するために、用語レベル、文章レベル、文書レベルでのアプローチが重要であるとし、翻訳に適した文書構造に関する議論が不足していることを指摘している。現実の運用場面を見据えて機械翻訳などの言語処理技術を活用するために

は、テキスト内の言語表現をフラットに扱うのでは なく、文書構造上の位置と対応づけて処理すること が有効である、と本研究でも強調してきた。

しかし、文書構造と一言に言っても、章・節・段 落や見出しといったテキストの論理的構造や HTML で定義されるようなマークアップ構造だけ でなく、センテンス単位の局所的な関係を捉える修 辞構造理論(Rhetorical Structure Theory: RST)など の様々なタイプが挙がるだろう。さらには、太字、 フォント、一列の文字数、段組といった文書の表示 に関わる構造も、言語表現に影響を少なからず与え ることは論をまたない。

本研究では、学術論文に広く見られる IMRAD (Introduction, Methods, Results And Discussion) 型式のような、特定の文書ジャンルにおいて広く観察される修辞的・機能的な構造(機能構造)にまず焦点を合わせ、ジャンル研究におけるテキストの修辞的構成の分析 (Biber & Conrad, 2009) や神門の提唱する機能構造分析 (神門, 1992) を参考に、文書の構成要素を整理した。その上で、DITA で定義される文書構造に重ね合わせることにした。

これにより、例えば、DITAの「手順(steps)」要素 (表 1 参照)では、日本語文の文末を動詞形に揃えるように制約をかけることで、原文のスタイルを統一できるだけでなく、命令形を使って訳す、といった指定が可能となる。この他にも、機能構造の「A3. 開始条件」「A4. 申請時期」要素のように、「~してから~日以内に~しなければなりません」といった形で構文パターンを明確に定義しやすい文書構造では、あらかじめ厳密な制限言語ルールを規定することが有効かつ可能である。

修士論文では着手する余裕がなかったが、先に述べた見出し、段落、箇条書きといったテキストの論理的な構造も、機械翻訳の高度な活用には不可欠であると考えている。箇条書きを例にとってみる。

6) 市役所、消防、警察の指示に従います。
というセンテンスを Google 翻訳で訳してみると、

日本語文(Japanese Sentence) ② ①	自動翻訳文 (Machine Translation) MNJH MS
燃えるゴミになります。	It's been combustible garbage. →
折り返し棚駅文(Back Translation) ②	一致度 (Agreement Score)
燃えるゴミでした。	C ==== 55 %
1文目:10文字	
	▲ 書き換え例: 「こちらが申請場所になります。」 ・「でいるのではです。」 「最大の展示物になります。」 ・「最大の展示物です。」

図 3: 制限言語チェッカーのプロトタイプ

I will follow the instructions 6) City Hall, fire, police. と先頭の番号「6)」が、箇条書きの番号と識別されずに、「6) 市役所」のまとまりで訳出されてしまう。 箇条書きというテキストのメカニズムを考慮すれば、前処理の段階でこのような誤訳を回避することが可能だろう。また、「従います」という部分が「I will follow」とI (私)を主語とした平叙文で訳されおり、ぎこちない英文になっている。

ここで、例えば、便宜的に箇条書きの数字「6)」 を削除し、語尾を「~てください」の形に変え、 市役所、消防、警察の指示に従ってください。 と書き換えることで、

Please follow the instructions city hall, fire department, police.

と訳される。この結果は、参照翻訳文

6) Follow the instructions of City Hall, the fire department and the police.

にかなり近づいていることが分かる。ここで、箇条書きの番号「6)」は後から付ければよいし、必要に応じて箇条書きの文頭の「Please」は自動的に削除する、等の柔軟な操作が可能となる。

## 3.3. 制限言語とテクニカルライティング

翻訳の上流工程を原文執筆にまで遡れば、テクニカルライティングの領域は自然と視野に入る。起点テキスト読者を指向したテクニカルライティング

は、機械翻訳向けの制限言語と、対象・目的こそ異なるものの、テキスト表層に現れる言語表現への操作を介して、曖昧性が少なく一意に理解できる文章を書くことを目指している点では、共通している。例えば、多くのテクニカルライティングの教本・指南書には「1 文をなるべく短くする」「修飾関係を明確にする」といった執筆指針やテクニックが記載されており、これは制限言語のルールとして援用できる。

以前より制限言語とテクニカルライティングとの接点は意識されており、吉田 (1983) は、制限言語の構築に際して、既存の作文技術に関する書籍を参考にしている。また近年大規模に推進されている産業日本語も、人に理解しやすく、コンピュータにも処理しやすい日本語を目指している(渡邊,2010)。 長尾・牧野 (1997) も著書『コンピュータで翻訳する』の中で、以下のように示唆している。

人間にとって読みやすい文と、コンピュータにとって解析しやすい文とでは、多少のちがいがあります。そのちがいは、現在の機械翻訳の性能の限界に起因しています。機械翻訳の性能の向上によって、このちがいはしだいに少なくなるはずです。(p.123)

機械翻訳の進歩に相まって、制限言語とテクニカ ルライティングの重なりは広がりつつありが、その 両立可能性について、原文と翻訳文の品質を定量的 に評価しながら検証した研究は、筆者らの研究の他 にほとんどない。なお本研究の結果だけ見ると、テ クニカルライティングのルールが、機械翻訳の精度 向上に直結しなかったが、これはテクニカルライティングの比較的柔軟なルールに引きずられすぎた ことが理由であろう。今後は機械翻訳の方式による 差異を考慮しながら、テクニカルライティングの知 見をより詳細にブレークダウンして統合したい。

## 3.4. 文書作成支援システムの考案

本研究は多言語での文書作成に関する枠組み・モデルを提案するに留まらず、文書作成支援システムの提案とプロトタイプの制作まで進めた。

これまで、日本語の制限言語の研究が一定量なされながらも、実用的な成果を出しきれていないのは、この応用面でのつなぎが不十分であったためと考えられる。英語の制限言語に関しては、制限言語のルールセットと同時に、それを実装したシステム環境まで開発されていることが少なくない。

理念的・規範的に「このように書けば翻訳しやすい・読みやすい」ということが定式化されたとしても、「実際に人間がそれに従って執筆することができるか」が保証されるわけではない。何十・何百ものルールを駆使して、統制された文書を執筆することは容易ではなく、また負荷もかかる。

特に社会的な背景を踏まえると、自治体においては、文書執筆のノウハウがうまく継承されていないことが課題とされており(安井,2009)、専門的なスキルを持たない一般的な執筆者を想定しながらシステム環境を提案することは、極めて有用であろう(仮に熟練の執筆者であっても、見落としやミスの問題は、多かれ少なかれ不可避であり、それをいかに機械的に解決していくかは重要である)。

現在筆者らは、制限言語ルールを実装したチェッカーの開発を進めており(図 3)、これを改良しつつ、各種用語集と連携させた上で、文書オーサリングシステムに組み込んでいく予定である。

## 4. 長期的な課題と展望

#### 4.1. 枠組みの拡張

本研究は、大きく3つの点で限界があり、今後拡張が求められる。

一つ目は言語対である。今回、多言語化の問題を対象にしながら、基本的には日英の言語ペアで研究を進めてきた。一通りシステム環境の開発まで進めることを優先してきたが、研究の方向性が見えた今、徐々に多言語に拡張していくことが求められる。特に、起点テキストに介入する制限言語のアプローチは、多言語での展開において一層効果を発揮すると考えられる。また DITA を組み合わせることで、多言語文書の管理や再利用も期待できる。

二つ目は対象文書である。今回は自治体ウェブサイト文書を対象としながら、自治体手続き型文書まで絞り込んだ。この文書クラスまで焦点化することで、文書一般に成り立つ議論にまで抽象化せずに、個別具体的な機能要素をおさえることができた一方で、自治体文書の一部しかカバーできなかった。特に、「手続き」以外の「概念」「ルール」「参照」といったタイプの文書群に対象を広げていく必要があるだろう。また自治体文書だけでなく、製品マニュアルなどの技術分野に拡張していくことも十分可能である。

三つ目は非言語情報である。今回はあくまで文書 要素として、言語情報 (文字列) のみを対象とした。 テクニカルコミュニケーションやドキュメントデ ザイン (Schriver, 1996) の分野でも、非言語情報は 重要な位置づけを占めている。また多言語サイトの オーサリング全体を射程に入れると、アイコンや画 像の入れ替えといった非言語要素の処理は無視で きない。優先順位としては、機械翻訳が直接対象と する言語情報に取り組む方が先だが、将来的な課題 として指摘しておく。

## 4.2. 関連領域との接点

断片的にではあるが、本研究の関連領域に言及し て締め括りたい。 まず翻訳学から「機能主義的アプローチ」を挙げる。Reiß の提案した、テクストタイプ別翻訳理論では、テキストをその機能に応じて「情報型」「表現型」「効力型」に類型化し、訳出法を選ぶもので(藤濤, 2007)、文書構造を定式化する上で有用な示唆を与えてくれる。さらにこの理論を発展的に継承したスコポス理論は、目標テキストで設定されるコミュニケーションの目的(スコポス: Skopos)によって、翻訳方略が変わることを理論化したもので、機械翻訳を使いつつ、適切に・効率的に文書を多言語で提供する状況にも適用できる。例えば、「事足りる翻訳(good enough translation)」という視点から機械翻訳を捉えることが有効ではないかと指摘されており(武田, 2013)、機械翻訳の言語学的な精度を超えた議論がますます期待される。

さらに、文書作成の上流工程に着目して、包括的な翻訳効率の改善を目指す手法は、翻訳学における「ローカリゼーション」パラダイムの中の「国際化」の文脈で取り組まれている(Pym, 2010)。

最後に元に戻って、図書館情報学について、少しだけ言葉を付け足したい。これまで見てきたように、本研究では「文書」という情報ユニットを介入点として、機械翻訳への応用を試みたが、図書館情報学の射程の広さと奥行きはこれに留まるものではない。「図書館」というマクロな社会的概念の側にも、「情報」というミクロな数理操作を喚起させる概念の側にも開拓の余地は残っており、図書・文書・テキストを起点に上下左右に切れ込みを入れながら、「機械翻訳が本当にどこまで使えるのか」という問いに答えるべく、研究を進めていきたい。

#### 5. 謝辞

制限言語ルールの構築と評価実験は、立見みどり 博士、Anthony Hartley 特任講師(東京外国語大学)、 影浦峡教授(東京大学)、井佐原均教授(豊橋技術 科学大学)と共同で進めました。深く感謝申し上げ ます。

### 6. 参考文献

- Biber, D., Conrad, S., *Register, Genre, and Style*, New York: Cambridge University Press, 2009.
- Hackos, J. T. 『DITA 概説書』 DITA コンソーシアム ジャパン 訳, エスアイビーアクセス, 2010.
- Hartley A., Paris C. "Multilingual Document Production From Support for Translating to Support for Authoring," *Machine Translation*, 12(1-2): 109-129, 1997.
- Pym, A. 『翻訳理論の探求』武田珂代子 訳, みすず 書房, 2010.
- Roturier, J., "Assessing a Set of Controlled Language Rules: Can They Improve the Performance of Commercial Machine Translation Systems?" The 26th International Conference on Translating and the Computer, 1-14, 2004.
- Schriver, K. A., Dynamics in Document Design: Creating Text for Readers, New York: Wiley, 1996.
- Wright, S. E., Budin, G. (Comp.), Handbook of Terminology Management, Amsterdam: John Benjamins, Vol.2, 2001.
- 井佐原均ほか「企業の多言語情報発信を支援する取り組み:国際化をにらんだ産業文書の効率的作成へ向けて」『言語処理学会第18回年次大会発表論文集』369-372,2012.
- 一般財団法人テクニカルコミュニケーター協会『日本語スタイルガイド第2版』テクニカルコミュニケーター協会出版事業部,2011.
- 神門典子「構成要素カテゴリを用いた原著論文の内 部構造分析」『情報処理学会研究報告』1992(32): 39-46,1992.
- 武田珂代子「機能主義的アプローチ(スコポス理論)」鳥飼玖美子 編著『よくわかる翻訳通訳学』 ミネルヴァ書房、2013.
- 藤濤文子『翻訳行為と異文化間コミュニケーション 一 機能主義的翻訳理論の諸相』松籟社,2007.
- 長尾真,田中伸佳,辻井潤一「制限文法にもとづく 文章作成援助システム」『情報処理学会研究報告』 1984(27): 1-8, 1984.
- 長尾真, 牧野武則『コンピュータで翻訳する』共立 出版,1997.
- 安井秀行『自治体 Web サイトはなぜ使いにくいのか? "ユニバーサルメニュー"による電子自治体・電子政府の新しい情報発信』時事通信出版局,2009.
- 吉田将「科学技術文書を記述するための日本語の規格化 係り受け関係の制限について」『九州大学工学集報』56(3): 205-211, 1983.
- 渡邊豊英「産業日本語プロジェクトの概要 特許・ 技術情報の利用性向上のために」『情報管理』 53(9): 480-491, 2010.

## AAMT-MT フェア2014展示報告

## **AAMT 機械翻訳フェア(MT フェア)2014 展示ブースの報告**

小谷克則

関西外国語大学

## はじめに

去る 6 月 17 日、AAMT 機械翻訳フェア (MT フェア) 2014 が開催されました。こうした機械翻訳の催しは、1989 年「翻訳技 術国際フォーラム (日本電子工業振興協 会)」、1990 年「自動翻訳フェア (日本翻訳協会主催)」、1992 年「機械翻訳フェア (日本科学技術情報センター主催)」、1993 年「機械翻訳フェア (日本科学技術情報センター主催)」と 20 有余年前に溯ります。

昨年度までは「報告会・講演会・AAMT 長尾賞授与式・懇親会」という名称でした が、今年度は、AAMT機械翻訳フェア 2014 と一新し、総会、報告会、講演会、AAMT 長尾賞授与式、懇親会などの恒例行事に加 え、新たに展示ブースが設けられました。

この展示ブースの運営を担当した委員 (長瀬友樹委員長他)より、展示ブースの 概要と来年度以降の課題を報告いたしま す。御一読いただき、機械翻訳フェアのよ り一層の盛況に向け、お気付きの点などを 是非当協会事務局までお寄せください。

#### 展示ブース開催まで

これまで当協会では、本会をより一層盛況にするために検討を重ねてきました。従来はご参加いただいた方々に有意義な場を提供することが検討の中心的課題でしたが、今回はまだ参加されたことの無い方々にご来場いただくということが新たな課題となりました。この課題への取組みの一環として、展示ブースを設けることが提案さ

れました。

この展示ブースの開催は、期待と不安が 入り混じる中、数多くの関係者の中で検討、 議論が進められてきました。名称を決める 際も「機械翻訳コレクション」を始めとす る様々な案があった中、議論のうえ「機械 翻訳フェア」に決定しました。

開催に際しましては、関係者並びに当協会法人会員の方々のご支援を賜り、第一回の展示ブースには6団体(当協会法人会員5社、当協会委員会)にご参加いただきました。結果、当日は場内大変盛況で、無事、閉会することが出来ました。

## 展示ブースの概要

展示ブースへの参加企業は、日本電信電話株式会社、独立行政法人情報通信研究機構、株式会社シュタール ジャパン、株式会社高電社、株式会社クロスランゲージ(順不同)でした。これらの企業に加え、当協会の機械翻訳課題調査委員会も参加しました。



開催時間は、報告会と招待講演の間(14時 40分から15時 20分まで)、AAMT 長尾賞・学生奨励賞授与式と懇親会の間(17時 20分から18時まで)と二回ありました。準備中(上図)は会場スペースが広すぎたのではないかと不安になりましたが、実際には、下図のように、どちらの回にも多くの方々が来られ、心配は杞憂に終わりました。



各団体の展示ブースの表題とその要旨は 次のとおりです。

## 特許翻訳などの専門的な外国語文書も 自然な日本語に!

# ~多言語統計翻訳プラットフォーム~ (日本電信電話株式会社)

- ・NTT が開発した、世界最高性能の構文解析技術と日本語の特徴を利用した事前並べ替え技術により、翻訳精度を飛躍的に向上させた機械翻訳を用いて、英中韓から日本語への特許文向け機械翻訳システムをデモします。
- ・対訳データ(過去の翻訳事例の蓄積)から翻 訳規則や翻訳辞書を自動学習するため、 対訳データを入れ替えれば、特定分野向 けの高品質な翻訳システムを自動で構築 する事が可能です。
- ・特許やマニュアルなどの書き言葉の翻訳

における下訳や外国語文書からの情報収 集を支援します。



# 情報通信研究機構における自動翻訳研究 (独立行政法人情報通信研究機構)

・情報通信研究機構において開発している 自動翻訳エンジンについて技術的な解説 とデモをします。出展する翻訳エンジン は、特許翻訳エンジンと汎用翻訳エンジ ンです。



# 翻訳支援ツール Transit NXT (株式会社シュタール ジャパン)

・翻訳支援ツール Transit NXT には創立 30 周年を迎えるシュタールグループの翻訳 会社としてのスキルとノウハウが凝縮さ れています。翻訳現場のニーズに則した実践的なツールとして定評のあるTransitの最新版にはプリトランスレーションに代表される既存機能の強化に加え、字幕翻訳のための機能なども実装されており、機械翻訳へのアプローチにおいても飛躍的な進化を遂げています。



## 企業向け翻訳システム (株式会社クロスランゲージ)

・WEB-Transer@Enterprise は、翻訳の新たな仕組みにより、社内の情報漏洩防止、翻訳コスト削減、翻訳資産の蓄積と再利用ならびに共有化、などの課題を解決する翻訳プラットフォームです。



# 未来を拓く! グローバル翻訳ソリューション (株式会社高電社)

・ルールベース翻訳に加えて統計翻訳や用 例翻訳の技術を取り入れた「トリプル MT メソッド」により、高電社の機械翻訳ソ リューションは日々進化を遂げていま す。特許翻訳ソフト、多言語音声合成シ ステム、スマホカメラ翻訳アプリなど、 多種多様のシステム・サービスの中で使 われている機械翻訳の活用事例をご紹介 いたします。



# AAMT 機械翻訳課題調查委員会活動報告 (AAMT 機械翻訳課題調査委員会)

- ・WG1:中日基本文テストセットを用いた 評価をデモします。
- ・WG2: 翻訳祭 2013 で実施した機械翻訳 の利用実態に関するアンケート結果を報 告します。
- ・WG3:用語集形式 UTX の紹介とデモを 行います。具体的には、UTX スライドを ご覧いただき、変換ツールのデモを行い ます。



## 次回に向けて

今回は会場を狭く感じるほどの盛況となり、嬉しい悲鳴をあげましたが、同時に、 今回の課題を洗い出す必要性もヒシヒシと 感じました。そこで、本会終了後、出展者 の方々を対象に、アンケートを通じてご意 見などを頂戴しました。このアンケート調 査の質問項目と回答例は次の通りでした。

質問は次表の 8 つの項目からなり、回答は選択式とコメントによる記述式の両方式です。選択式の回答のうち最多であったものにマーク(✔)を付けました。

- 1. 今回の出展に満足されましたか?
  - /満足

ふつう

改善が必要

- ブースへの来訪者数はいかがでしたか?
   ちょうどよかった
  - ✓多すぎた 少なすぎた
- 開催時間はいかがでしたか?
   ちょうどよかった
   長すぎた
  - ✓短すぎた
- 4. 展示回数はいかがでしたか?
  - √ちょうどよかった
    - 一回目だけでよかった
    - 二回目だけでよかった
- 次回の MT フェアでも展示を希望されますか?
  - ✓参加する 参加しない わからない

6. 出展による成果はいかがでしたか?大いにあったまぁまぁあった

✓なかった

7. 全体的な運営はいかがでしたか?

良い

**V**ふつう

悪い

8. 今後の MT フェア展示について改善すべ き点やご意見をお伺いします。

(自由記述)

全般について

展示ブースについて (部屋サイズ、 ブースサイズ、配置など)

搬入・搬出について (時間や荷物の 授受の方法など)

その他(どのようなご意見でも結構です)

また、コメントを通じて多くのご感想と ご意見を頂戴しました。概略は以下の通り です。

質問1:来場者数が予想以上に多く、展示ブ ースへの出展は概ね満足した。

質問 2: 会場スペースや展示時間では対応が 困難なほど多かった。

問題 3: 来場者全員に紹介しきれなかったため、もう少し時間は長くても良かった。

問題 4: 一回だけではなく、二回あってよかった。さらに、受付終了後から総会開始までの時間も活用するなど回数は増やすべきである。

問題5:次回も展示を希望する。

問題 6:直接的な成果は多くないにしても、 今後の成果へ契機となった。

問題 7:展示ブース担当者が総会や懇親会に

参加が困難な場面があった。

その他、たくさん有意義なコメントを頂 戴しました。紙面の関係上、その一部をご 紹介させていただきます。

- ・来年も出展できるのであれば対応人員を 増やすべく検討します。
- ・総会終了から懇親会開始まで、撤収作業 の時間を考慮いただきたかったです。
- ・デスクはもう 1 周り大きいとありがたいです。集客が多かったため、部屋が小さく感じられました。人が一杯で後部からは見えなかったとのご指摘がありました。
- ・今回のように二回か、あるいは、三回く らいが適当と思います。
- フェア展示については、もう少しスペースがあった方が良いと思います。
- ・お昼の時間帯などを利用して展示ブース を見て頂ける時間を、増やして頂ければ と思います。
- ご来場される方の特徴などが事前にわかっていると、ご説明しやすいと思います。
- ・たくさんの方が集中されたため、満足に 名刺交換も行う事が出来ずにおりました ので、展示ブースでゆっくりご覧頂ける 環境を作って頂けると、より良い展示会 になると思います。
- ・お客様が集中していらっしゃるため、部 屋サイズはもう少し広い方がよかった。
- ブースサイズは妥当だと思います。

## おわりに

展示担当者の方々から頂戴したご意見により、次回の課題を確認することができました。本紙面を借り、ご回答いただきました担当者の方々に御礼申し上げます。また、展示を担当されました秋葉泰弘氏ら(日本

電信電話株式会社NTTメディアインテリジェンス研究所)、目次由美子氏(株式会社シュタール ジャパン)から報告記事も御寄稿いただき、重ねてお礼申し上げます。

来年度はより盛況な展示ブースを開催できるよう、課題に取り組む所存です。より 一層の皆さまのご協力をお願い申し上げます。

## AAMT-MT フェア2014展示報告

# 翻訳支援ツール Transit NXT

株式会社 シュタール ジャパン 目次 由美子

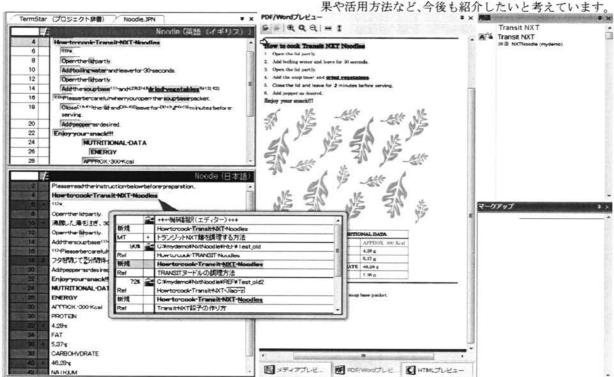
2014年6月17日に「AAMT機械翻訳フェア 2014」が東京都港区のホテルアジュール竹芝にて開催されました。「MT フェア」と称された同会では AAMT 初の試みとして法人会員企業による展示のための時間も設けられ、当社も自社開発の翻訳支援ツール Transit NXTを紹介させていただきました。

Transit はシュタールグループのスイス本社とドイツの専門チームによって開発された翻訳支援ツールであり、シュタール ジャパンにて日本語化し、販売とサポートを行っています。創立30周年を迎えるシュタールグループの翻訳会社としてのスキルとノウハウが凝縮され、翻訳現場のニーズに則した実践的なツールとして定評のある Transit の最新版には機械翻訳機能も搭載されています。今回のMTフェアでは、この機械翻訳機能を中心に展示させていただきました。

展示場には多くの参加者が訪れ、翻訳支援ツールの概要、翻訳メモリの活用、用語集の利用、機械翻訳機能とのインターフェースやオプション、オンライン翻訳サービスとの連携についてなど、多様な質問が熱心に寄せられました。MTフェア参加者の翻訳テクノロジーに対する感心の高さが伺える結果となったと言っても過言ではないと思います。

参加者からは貴重なご意見も頂戴し、誠に得難い体験となりました。来年も展示の機会をいただけるのであれば、改善事項として是非取り組みたいと考えております。

特にこの 10 年で日本の翻訳業界に浸透してきた翻訳支援ツールには、今や機械翻訳機能との連携が求められています。Transit に装備された機械翻訳機能の効



#### AAMT-MT フェア2014展示報告

# 特許翻訳などの専門的な外国語文書も自然な日本語に! ~多言語統計翻訳プラットフォーム~

NTT メディアインテリジェンス研究所

秋葉泰弘 我妻光洋 荒井和博

セキュアな翻訳クラウドを必要とする翻訳会社や企業内翻訳部門向きにNTTグループの事業会社等から提供する翻訳サービスの実現を目指し、当研究所では、弊社で研究を推進している翻訳技術[1,2,3,4]をコアに多言語統計翻訳プラットフォーム(以下、翻訳PFと呼ぶ)の開発を進めています。

現在開発が完了している翻訳方向はX日翻訳(英語、中国語、韓国語から日本語への翻訳)で、統計翻訳では珍しくユーザー辞書も利用可能です。翻訳PFにより、従来難しいとされて来た外国語特許などの専門的な外国語文書も自然な日本語に翻訳が可能となりました。特許やマニュアルなどの書き言葉の翻訳における下訳や外国語文書からの情報収集支援が可能となります。

AAMT機械翻訳フェア(MTフェア)2014 (2014年6月17日、ホテルアジュール竹芝にて開催)では、英中韓から日本語への特許翻訳を題材に、翻訳PFを用いた翻訳デモをご覧に入れました。以下本稿では、その概要をご紹介します。

統計翻訳では、過去の翻訳例(原文とその訳文の対)の蓄積(対訳データ、対訳コーパスと呼ぶ)を準備し、どのような単語列をどのような単語列をとのような単語列に翻訳するかと言う「翻訳規則」や「対訳辞書」にあたる翻訳の統計モデルを対訳データから全自動で学習し、学習したモデルを活用して新たな原文を翻訳先言語に翻訳します。

統計翻訳は、文構造が比較的類似している欧米系の言語同士では、実用レベルの翻訳品質に達し議事録の翻訳などで実用されるほどですが、英語と日本語のような文構造があまり類似しない言語同士(例えば、SVO言語とSOV言語)では、訳文の文構造が原文の文構造に引きずられるなど読み難く問題でした。従来の統計翻訳では局所的な語順並べ替えに留まり、大局的な並べ替えが必要な翻訳は未解決でした。

この語順の違いと言う課題を克服するために、翻訳PFでは、

- 1)まず、原文の単語を目的言語の語順に並べ替え、
- 2) その後、各単語の逐語訳を行うと言う2段階の処理で翻訳を進めます。英日翻訳の場合、例えば、「He saw a cat with a long tail.」と言う原文が入力されると、「He long tail with cat saw.」と単語は英語のままで語順だけを日本語風の語順に並べ替え、その後に、「彼 は 長い 尻尾 の 猫を 見た。」と逐語訳を行います。

日本語では非修飾語は修飾語よりも後ろに必ず位置しますが(図 1)、英語や中国語などでは前にも後ろにも位置し得ます(図 2 の上側)。日本語風の語順にするとは、非修飾語が修飾語よりも必ず後ろに位置するように原文を並べ替える事です(図 2 の下側)。この日本語の特徴を利用した事前並べ替え技術は弊社独自の技術です。



図 1:日本語の係り受けの例 (→:修飾語から非修飾語への矢印)



図 2:並べ替えの例 (上:原文英語の係り受け、

下:日本語風に並べ替えた結果の係り受け)

この並べ替えを上手く機能させるためには、文を構成する単語同士の修飾・非修飾の関係を高精度に捉える必要がありますが、弊社で研究開発した世界最高性能の構文解析技術(英語のテストベットで評価)によりこの並べ替えを可能としています。

今回のデモでは、構文的に複雑な文書の 代表例として特許文を題材に、英中韓から 日本語へ特許翻訳のデモをご覧に入れまし た。翻訳の統計モデルを学習する際に用い る対訳データを入れ替えれば、特定分野向 けの高品質な翻訳システムを自動で構築す る事が可能です。例えば、御社で眠ってい る過去の翻訳資産を活用し、翻訳規則に当 たる翻訳の統計モデルを御社向けにカスタ マイズした高品質な翻訳システムが実現可 能です。

【お問合せ先】megen-pro-liaison (at) lab (dot) ntt (dot) co (dot) jp

- [1] Isozaki et.al.: Head Finalization: A Simple Reordering Rule for SOV Languages. In Proc. WMT Metrics MATR,, pp. 244–251, 2010.
- [2] 須藤ほか: 英中韓から日本語への特許 文向け統計翻訳システム, 言語処理学会 年次大会, 2014.
- [3] 須藤ほか: 英中韓から日本語への特許 文向け統計翻訳 -高精度言語解析と事前 並べ替えによる高精度な特許翻訳システ ムの構築・, Japio YEAR BOOK 2014, (掲載予定).
- [4] 永田ほか: 革新的発展期を迎えた統計 翻訳, 技術ジャーナル, Vol. 25, No. 9, pp. 14-17 2013. [URL: http://www.ntt. co.jp/journal/1309/files/jn201309014.pd f]

#### 委員会活動報告

#### これまでの AAMT Forum メールマガジン Vol.3

#### 機械翻訳課題調査委員会 WG1、WG2

#### はじめに

第3回となりましたメールマガジンのバックナンバーをお届します。今回は第21号から第28号です。

今回は、AAMT機械翻訳フェア 2014 において展示ブースをご担当いただいた当協会の法人会員の方々(独立行政法人情報通信研究機構、株式会社高電社)関連のニュースもありました。今後もメールマガジンを通じて様々な情報を皆様に配信させていただきます。

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、その他、お気付きの点などもAAMT事務局までお寄せください。よろしくお願いいたします。

では、メールマガジン第 21 号から第 28 号までを再びお届けします。

# AAMT Forum メールマガジン 21号 (2014/05/23 配信)

こんにちは。AAMT Forum メルマガ担 当です。若葉が初夏の日ざしにまぶしくか がやく季節となりました。また、夏を前に 梅雨入りも迫ってきましたので、大雨、湿 気への備えを忘れないようしましょう。で は、メールマガジン第21号をお届けします。

■Google がカメラをかざすだけの簡単 翻訳アプリ「Word Lens」の開発元を買収 カメラをかざすだけでさまざまな言語の 翻訳をしてくれるアプリ「Word Lens」の 開発元である Quest Visual が新たに Google の参加となりました。

http://gigazine.net/news/20140519-google-quest-visual/

■ 運 賃 規 則 自 動 翻 訳 「 INFINI FareLingual」をリリース

株式会社インフィニ トラベル インフォ メーションは、 運賃規則自動翻訳ソリュー ション「INFINI FareLingual(インフィニ フェアリンガル)」を 2014 年 5 月 14 日(水) にリリース致しました。

http://www.infini-trvl.co.jp/pressrelease/?i d=1399948799-587138

■翻訳ソフト「コリャ英和! 一発翻訳 2015 for Win」最新版

ロゴヴィスタは8日、英日・日英翻訳ソフトの最新版となる「コリャ英和! 一発翻訳 2015 for Win」シリーズを発表しました。 通常版のほか専門辞書パックにより5モデルを用意し、5月30日より発売します。

http://news.mynavi.jp/news/2014/05/09/31 1/

# AAMT Forum メールマガジン 22 号 (2014/06/04 配信)

こんにちは。AAMT Forum メルマガ担 当です。日本各地で真夏日を迎え、一部で は梅雨入りしたこの頃、いかがお過ごしで しょうか。梅雨入りすれば雨で気温は下が りますが、温度差で体調を崩さないように 気をつけましょう。では、メールマガジン 第22号をお届けします。

■Microsoft、Skype の通訳アプリを年内に β 提供へ

米 Microsoft は 5 月 27 日 (現地時間)、 Skype の通話を通訳する音声翻訳アプリ 「Skype Translator」を発表しました。

http://www.itmedia.co.jp/enterprise/articles/1405/28/news063.html

■CSLi、フランスの自動翻訳ソリューション世界最大手「シストラン」を買収

韓国最大手の自動翻訳ソリューション専門企業(株)シーエスエルアイが自動翻訳ソリューション開発の世界最大手企業シストランを買収しました。

http://www.zaikei.co.jp/releases/171287/

## AAMT Forum メールマガジン 23号 (2014/06/20 配信)

こんにちは。AAMT Forum メルマガ担 当です。雨、雨のうっとうしい日が続き、 心もしめりがちですが、お元気でしょうか。 ところによっては豪雨にもなるようですの で、川の増水や地盤の緩みなどには注意しましょう。では、メールマガジン第 23 号をお届けします。

■特許庁、特許文献の翻訳ソフト"精度" を数値化-指針公表

特許庁が外国語で著された特許文献を翻訳ソフトなどで機械翻訳した際の精度を評価するための指針をまとめ、公表しました。

http://www.nikkan.co.jp/news/nkx022014 0605abbh.html

■訪日外国人向けICT整備で行動計画 総務省、自動翻訳システムも検討

総務省が訪日外国人のためのICT(情報通信技術)の利用環境整備に向けたアクションプランを取りまとめたと発表しました。

http://sankei.jp.msn.com/politics/news/14 0612/plc14061219390021-n1.htm

■ロゴヴィスタ/10 言語に対応した多言語 翻訳ソフト

ロゴヴィスタは Windows 用の英日・日 英翻訳ソフトの最新版「コリャ英和!一発 翻訳 2015 for Win」シリーズの多言語対応 製品「コリャ英和!一発翻訳 2015 for Win マルチリンガル」を、27 日から発売すると 発表しました。

http://ict-enews.net/2014/06/04logovista/

# AAMT Forum メールマガジン 24号(2014/07/04 配信)

こんにちは。AAMT Forum メルマガ担 当です。沖縄はすでに梅雨明けし、だんだ んと暑くなってきたこの頃、いかがお過ご しでしょうか。気温や湿度の上昇に体が慣 れていない時期もまた、熱中症には注意が 必要です。では、メールマガジン第24号を お届けします。

■「みんなの自動翻訳@TexTra」を一般公 開

NICT は、ユニバーサルコミュニケーション研究所の多言語翻訳研究室で開発した自動翻訳エンジンを「みんなの自動翻訳@TexTraR」一般公開しました。

http://www.nict.go.jp/info/topics/2014/06/1 40619-1.html

■しゃべるだけでOK 17言語おまかせ 最新の音声翻訳システム 「ボイストラ」

NICT が開発した翻訳システムが注目を 集めています。スマートフォン用アプリの ダウンロード数は100万を突破し、今秋 からは、日本の病院で外国人の患者の問診 に使えるかなどをみる社会実験も計画され ています。

http://sankei.jp.msn.com/west/west\_economy/news/140623/wec14062317000005-n1.htm

■手話が聞こえる! Google Gesture が拓く 次世代の扉

手話の動きを特殊なリストバンドでデータ化し、音声に翻訳する「GoogleGesture」の紹介です。残念ながらまだ実現はできていませんが、その仕組みを説明しています。

http://nge.jp/2014/06/26/post-1355

#### AAMT Forum メールマガジン 25 号 (2014/07/18 配信)

こんにちは。AAMT Forum メルマガ担 当です。夏の日盛りに木陰の恋しい季節と なりました。空調の効いた室内は過ごしや すいですが、外で涼しさを感じてみるのも いかがでしょうか。では、メールマガジン 第 25 号をお届けします。

■総務省、東京五輪に向けテレビの外国語 字幕充実へ

総務省の有識者検討会は 2020 年の東京 五輪・パラリンピックに向け、テレビ放送 の外国語字幕サービスを本格的に始める計 画を発表しました。

http://www.sanspo.com/geino/news/20140 703/sot14070321100011-n1.html

■高電社、『翻訳クラウド』を 39 言語対応 に大幅拡充!

株式会社高電社は、様々な自動翻訳を実 現するクラウドサービスである『翻訳クラ ウド』に、11 日(金)より、タイ語、ベトナ ム語、インドネシア語などのアジア言語を 含む 29 言語を新たに追加しました。

http://www.sankeibiz.jp/business/news/14 0711/prl1407111008008-n1.htm

# AAMT Forum メールマガジン 26号 (2014/08/01 配信)

こんにちは。AAMT Forum メルマガ担 当です。全国的に梅雨も明け、夏本番のこ の頃、いかがお過ごしでしょうか。強い陽 射しには注意が必要ですが、今年も夏を楽 しみましょう。では、メールマガジン第 26 号をお届けします。

■NICT と特許庁が多言語特許文献の高精 度自動翻訳の実現に向けて協力合意

特許文献の自動翻訳に関しては、特許庁が主に原語文献の収集を行い、NICTがそれを基に主に自動翻訳の精度向上策を検討します。

http://www.nict.go.jp/press/2014/07/28-1.h tml

■Google 翻訳に誰でも参加できる品質向上 コミュニティー 「Translate Community」

米 Google はユーザーの言語知識をクラウドソースとして役立てる Google 翻訳の品質向上を目的とした新コミュニティー「Translate Community」を立ち上げました。

http://www.itmedia.co.jp/enterprise/articl

#### es/1407/28/news042.html

■人気の自動翻訳アプリ「iTranslate」が Mac でも使えるように!

iOS、Android、Windows で利用可能な 自動翻訳「iTranslate」から、Mac 版がリリ ースされました。

http://news.mynavi.jp/news/2014/07/17/47

# AAMT Forum メールマガジン 27号(2014/08/20 配信)

こんにちは。AAMT Forum メルマガ担 当です。お盆の頃は台風の影響で涼しかっ たですが、また暑い日々が戻ってきました。 何をしなくても汗が流れるこの頃、こまめ な水分補給は大切です。では、メールマガ ジン第 27 号をお届けします。

■SDL が翻訳メモリのトレンドを牽引

SDL Trados Studio ライセンスの販売数がわずか 5 年で 100,000 に達したことが発表され、人間による翻訳、翻訳メモリ、機械翻訳が市場で共存していることが示されました。

http://japan.cnet.com/release/30078257/

■男女の声でエンジンの「気持ち」翻訳ア プリ配信

ヤマハ発動機は、オートバイなどのエン ジン音を言葉に「翻訳」するスマートフォ ン向けアプリ「レヴ トランスレーター」 の無料配信を始めました。

http://www.yomiuri.co.jp/it/20140803-OY T1T50087.html?from=ytop\_main9

# AAMT Forum メールマガジン 28号 (2014/09/03 配信)

こんにちは。AAMT Forum メルマガ担 当です。初秋の候、いかがお過ごしでしょ うか。日中は暑い日がまだまだ続きますが、 朝夕の涼しさと虫の音に、夏の終わりと秋 が近づいているのを感じます。では、メー ルマガジン第 28 号をお届けします。

■ Android 版 多言語音声翻訳アプリ VoiceTra4U を公開 さらに便利! 無料 Wi-Fi 利用アプリと連携

独立行政法人情報通信研究機構は、多言語音声翻訳アプリ VoiceTra4U について、Android 版を無償公開し、日本全国のフリーWi-Fi に簡易に接続できる無償アプリJapan Connected-free Wi-Fi の標準の翻訳アプリとしました。

http://www.nict.go.jp/press/2014/09/01-1.h tml

■世界最大口コミサイト「Yelp」、アプリで レビューを自動翻訳

世界最大規模の口コミサイト「Yelp (イェルプ)」を運営する Yelp は 8 月 20 日、 iPhone 向けアプリで日本語や英語、フランス語を含む全 15 言語に対応した「レビュー 自動翻訳」サービスを開始しました。 http://japan.cnet.com/news/service/350526 19/

■飲食店向け"無料"ホームページ作成サービス「favy (ファビー)」が多言語翻訳に対応!

株式会社マネタイズは、favy の新機能である『多言語翻訳機能』の提供を8月25日より開始いたしました。現在翻訳可能となっている言語は日本語、英語の2つで、今後中国語や韓国語といった言語を追加していく予定です。

https://www.dreamnews.jp/press/0000098 314/

#### 委員会活動報告

#### AAMT 日中/中日テストセットに基づく翻訳自動評価サイト公開

機械翻訳課題調査委員会ワーキンググループ1

#### 1. はじめに

AAMT 機械翻訳課題調査委員会(以下、課題調査委員会)では、機械翻訳システムの翻訳品質を評価する方法として「テストセット評価」に着目し、機械翻訳エンジンの評価のためのテストセットの作成を行い、2011年に中日翻訳用、2013年に日中翻訳用のテストセットが完成した。また、2011年から日中翻訳用テストセットを対象に設問回答の自動化についての検討を進め、2013年にWebブラウザ上で日中翻訳エンジンの評価を実行できる自動評価サイトを作成した。今回、中日翻訳用テストセットについても自動評価に対応することにより、日中/中日の双方向で翻訳エンジンの自動評価を可能とした。本稿では、近日 AAMT ホームページで公開される予定の日中/中日自動翻訳サイトについてその概要と使い方について説明を行う。

#### 2. AAMT 機械翻訳文テストセット

「AAMT 機械翻訳文テストセット」は、原文と参照訳、文法カテゴリ、設問の4つを基本レコードとするデータの集合である。テストセットに基づく翻訳エンジン評価の特徴や、他の評価手法との比較についてはAAMT Journal No.45(2009)の解説を参照されたい。

自動評価サイトは、2009~2013年に課題調査委員会で作成してきた日中翻訳評価用テストセット、 および、中日翻訳評価用テストセットを用いて作成した。それぞれのテストセットの概要と自動評価 に用いた例文数について触れておく。

#### 1) 日中翻訳評価用テストセット

JEIDA で開発された既存の日英翻訳用テストセットをベースに整備した第1版(評価文数 325 文)と、その後中国語特有の文法項目とテスト文の増強を行った第2版(評価文数 576 文)がある。自動評価に対応したテスト文は第2版作成時に増強した39項目251文のみである。

#### 2) 中日翻訳評価用テストセット

中日方向では流用できる既存テストセットが存在しないため、新規でテストセットを整備した。 複数の中国語文法書を参考にして文法項目について検討し、合計 31 項目 502 文から成るテスト セットを作成した。中日翻訳についてはテストセット全体を自動評価に利用した。

#### 3. 自動評価サイト

前節で述べた日中翻訳用テストセットは、当初は人間が評価(翻訳結果を見て設問回答)することを想定したものだが、設問評価はYesとNoを判断するだけの簡単な評価であるため、設問評価の自動化するためのプログラムを作成した。ただし、人間の文法知識に基づく判断が必要とされるものなど自動化が難しい設問については、特定の文字列候補が訳文に含まれるか(含まれないか)否かで回答できるように設問の簡略化を行った。このため、自動評価用の設問はオリジナルの人間評価用の設

問と同一ではなくなってしまったが、日中翻訳用テストセットを用いた実験では、自動評価と人間評価の回答の分布が統計的に十分に一致していることが確認されている。<sup>1</sup>

今回、この自動動評価プログラムを Web サイト上で公開し、テストセットに基づく日中/中日翻訳エンジンの評価が全世界からネット経由で試すことができる環境を整えた。

図1に自動翻訳サイトの構成を示す。Web アプリケーションとして実現されており、Web ブラウザ のみで翻訳結果評価の実行が可能である。評価のためのテスト文(原文)と評価用のプログラムは、評価サーバー(AAMT サーバー)内で管理されている。利用者は原文をダウンロードし、翻訳結果を評価サーバーにアップロードするだけで評価結果を見ることができる。

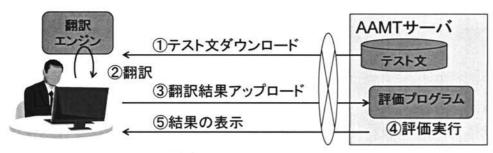


図1 システムの構成

#### 4. 実行例

評価サイトの利用手順および評価結果の見方について、実行例をもとに説明する。

#### ① 翻訳方向の選択

サイトへのログイン<sup>2</sup>に成功すると、図2のような初期メニューが表示される。評価するエンジンに合わせて[中日翻訳評価を行う]または[日中翻訳評価を行う]のボタンをクリックする。(翻訳方向がどちらであっても基本的手順は同じであるため、以下では[中日翻訳評価を行う]を選択した場合について説明する。)

<sup>1</sup> カッパ係数をもとに有意水準 1%の検定を実施

<sup>&</sup>lt;sup>2</sup> ログイン ID の取得については AAMT のホームページの説明を参照

		- Annahum man	0.0	X
∃ littp://chinese-mt.ir-alt.info/			D-0 0	7 6
HIMBERTHINE CONSTR	And there's the	TOOT ENTER		
中国語・日本語翻訳評価				
中国語→日本語の翻訳評価	日本語→中国語の翻	款評価		
▶中日翻記時価を行う	<b>》</b> 日中翻訳平值5行5			
<b>》</b> 中日\$P\$55.64周末至2条	ノ日中語をお記される			
		Supported by IR-ALT		
		- ALT		
ロヴァウト				
			€ 100%	

図2 初期メニュー

#### ② テスト文の取得

評価対象エンジンで翻訳を実行する原文をサーバーから得する。

「原文ページ」(図3) の「入力テキスト」領域に原文が入った状態で表示される。中国語テキスト全体をクリップボードにコピーし、テキストファイルとしてローカル PC に保存する。(このとき 1 原文/1 文の状態を崩さないように注意。)原文は全部で 1420 文あるが、実際に評価に使われる文はこのうち 502 文である。原文を取り込んだら[翻訳評価ページへ]をクリックする。

なお、評価に用いない原文を大量に混ぜているのは、エンジンベンダーによる原文へのチューニングを難しくするためである。

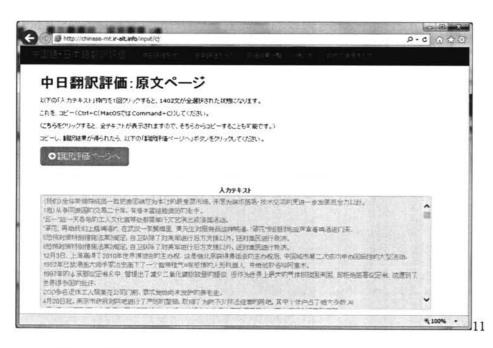


図3 原文ページ

#### ③ 翻訳の実行と翻訳結果のアップロード

評価対象翻訳エンジンによる取り込んだ原文全体を評価対象エンジンで機械翻訳し、翻訳結果を評価サーバーに登録する。

原文 1 行を 1 文とみなして翻訳を実行し、翻訳結果の行数が原文行数と等しいことを確認する。 翻訳結果全体を、図 4 の「翻訳結果画面」のテキスト領域にクリップボード経由で貼り付け、[送信する]ボタンをクリックする。すべての翻訳結果の評価サーバーへの送信が終わると、サーバー内で自動的に評価プロセスが実行され、図 2 の初期メニューに推移する。

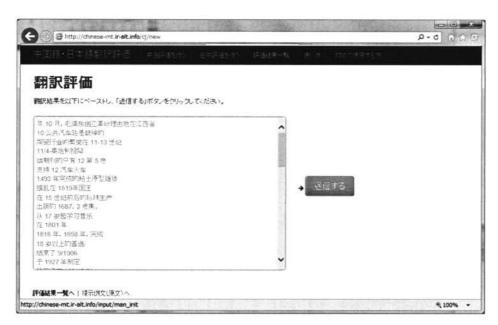


図4 翻訳評価ページ

#### ④ 評価結果の確認

初期メニュー (図2) から[中日評価結果を見る]ボタンをクリックして、図5に示す「評価結果一覧」画面を表示させる。一覧から実行時間が最も新しい行の[結果を見る]ボタンをクリックすると評価結果が表示される。なお、評価サーバーにはこれまでに実行された評価履歴が保存されており、「評価結果一覧」画面のボタンをクリックするだけで、いつでも過去の評価結果を確認することができる。

評価結果は文法項目別にレーダーチャートの形で表示されるが、中日 33 項目、日中 39 項目のままでは表示が煩雑になるため、中日 13 項目、日中 10 項目の大項目にまとめ、大項目単位でも結果を表示することができるようにしている。中日は元の 33 項目でもレーダーチャート表示が可能だが、日中は大項目単位でしか表示することができない。(日中についても年内に 39 項目単位の表示に対応予定。)



図5 評価結果一覧

#### <中日評価結果の見方>

画面上半分では 13 種の大項目単位で、下半分には全 31 項目の単位で、文法項目別の評価値をレーダーチャートで可視化して表示している (図 6、図 7)。項目別の評価値は画面右側の表で確認できる。評価値は項目別の正解(設問回答が可)となった例文の比率(%)である。

#### <日中評価結果の見方>

10種の大項目単位で、文応項目別の評価値をレーダーチャートで表示している(図 8)。レーダーチャート中の薄い線(実際には黄色)は、主要なフリー翻訳サイトのエンジン(6種類)を対象に評価したときの平均値を示している。薄い色(実際には黄色)の線との比較を行うことで、評価対象エンジンの相対的な長短所が明確になる。(中日評価結果のレーダーチャートにおける平均値表示については年内に対応予定。)

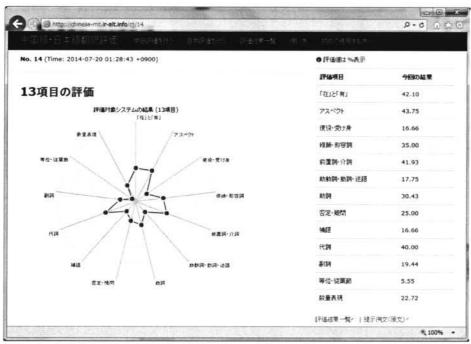


図6 中日評価結果(大分類)

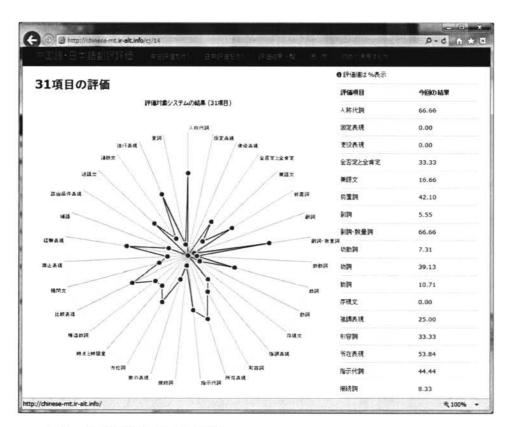


図7 中日評価結果(小分類)

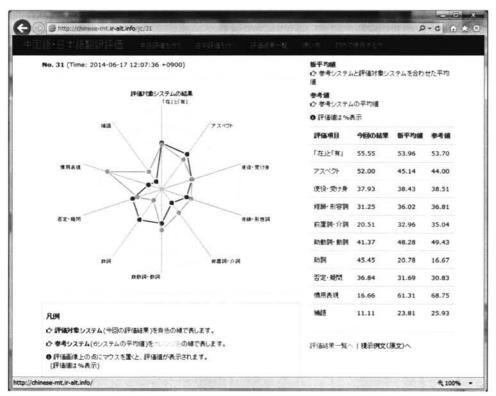


図8 日中評価結果

#### 5. おわりに

テストセットに基づく翻訳エンジン評価は、BLUE や NIST といったメジャーな自動評価にはない 特徴として、評価エンジンの成績が文法項目別に明らかになることが挙げられる。今回作成した自動 評価サイトでは、項目別評価値をレーダーチャートで可視化することで、評価エンジンが他のエンジンと比べてどの項目が優れて(劣って)いるかが一目でわかるというメリットがある。

今後、中日翻訳エンジン評価におけるテストセット自動評価と人間主観評価や他の自動評価指標との相関を調べ、中日翻訳においてテストセット自動評価がエンジンの品質を表す指標として使えるかどうかという観点での評価を行う予定である。また、今回作成した日中/中日エンジンの自動評価サイトを広く公開することで、多くの評価データを収集できることが期待される。日中翻訳については、テストセット自動評価と BLUE 等の自動評価が高い相関を示すことを確認しているが、サイト運用で集まるデータを使って、より多くのエンジンでの追加検証を行っていきたい。

将来的には、特許文テストセットやニュース文テストセットというように、テストセットを文種別に整備し、AAMTの機械翻訳評価サイトを文種別に翻訳エンジンの自動評価ができるサイトに成長させたいと考えている。

# AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから 第七回を迎えることができました。今号では、法人会員一社、個人会員二名の皆さまからのご寄稿をいただいております。

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

AAMT Journal では今後も引き続き、会員の皆さまからのご寄稿を心よりお待ちしております。 ご寄稿・お問い合わせは AAMT 事務局(E-mail: AAMT-info@AAMT.info)まで宜しくお願いいたします。

#### 法人会員(敬称略·50音順)

#### 会員名

株式会社翻訳センター

http://www.honyakuctr.com/

【大阪本社】〒541-0056 大阪府大阪市中央区久太郎町 4-1-3 大阪御堂筋ビル 13 階

【東京】〒108-0073 東京都港区三田 3-13-12 三田 MT ビル 7・8 階

【名古屋】〒450-0002 愛知県名古屋市中村区名駅 3-16-4 太陽生命名駅ビル5階

【福岡】〒812-0011 福岡県福岡市博多区博多駅前 1-4-4 JPR 博多ビル 10 階

#### 自己紹介

(株)翻訳センターは、特許、医薬、工業、金融・法務の主要 4 分野を軸として企業向けに専門性の高い翻訳を提供している翻訳会社であり、国内最大の売上高を計上する業界唯一の上場企業です。 2012 年には通訳や人材派遣、コンベンション事業を行う(株)アイ・エス・エスをグループ会社化。「すべての企業を世界につなぐ言葉のコンシェルジュ」を経営ビジョンに掲げ、言語サービスの総合プロバイダを目指しています。

翻訳会社が多く所属する一般社団法人日本翻訳連盟にも長年にわたり参加し、業界の発展にも努めています。

#### MT/翻訳とのかかわり

#### MTおよび翻訳業界に期待すること

機械翻訳(以下、MT)については現在、海外および日本の翻訳会社による進んだ取り組みについて研究している段階です。

海外、特に欧米では、ここ数年でMTが産業翻訳の現場にも定着してきたと認識しています。MTを組み込んだ翻訳業務プロセス、MTの出力を修正するポストエディタの育成方法など、海外での取り組みから学ぶ点は多くありますが、英日・日英のMTエンジンの性能部分においては、現状、国産のエンジンに品質上の優位があると考えています。

日本国内の産業翻訳においても今後数年の間に MT が普及する可能性は非常に高いと認識しており、市場の動向を注視しつつ、情報の収集と業界内での協力に、これまで以上に熱心に取り組んでいきたいと考えています。

(株) 翻訳センター 業務推進部 部長代理 河野弘毅

#### AAMTへの要望

MT というテーマのもとに研究者と翻訳会社が一同に集まることができる場としての AAMT の存在意義はとても大きいと考えております。引き続き、MT についてのセミナー開催、議論や情報交換の場として、また、情報発信の媒体として、AAMT には大きな期待を寄せております。

# AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第八回を迎えることができました。今号では、法人会員一社、個人会員二名の皆さまからのご寄稿をいただいております。

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

AAMT Journal では今後も引き続き、会員の皆さまからのご寄稿を心よりお待ちしております。 ご寄稿・お問い合わせは AAMT 事務局(E-mail: AAMT-info@AAMT.info)まで宜しくお願いいたします。

\_\_\_\_\_

## 個人会員(敬称略・50音順)

------

#### 会員名

新田義彦

日本大学経済学部産業経営学科·大学院総合社会情報研究科 教授

-----

#### 自己紹介

日立製作所の中央研究所・システム開発研究所・基礎研究所において、25 年間、形式言語および自然言語の研究開発に従事しました。大型電子計算機の論理回路設計・基本ソフトウェアと言語、そして情報検索システム、機械翻訳システムなどの研究開発に携わり貴重な経験を積むことができたと振り返っております。その後 1994 年に、日本大学の理工学部および経済学部に移り、研究を続行しつつ教育業務に関与して現在に至っております。柔軟な発想・連想をする若い人々と接する機会が増えて、企業の研究開発では得られぬ異次元の"何か"が見つかるかもしれないと密かに期待しつつ凡愚の努力を継続しています。

------

#### MT/翻訳とのかかわり

#### MTおよび翻訳業界に期待すること

機械翻訳の研究は、電子計算機の誕生と同時に始まったことは広く知られています。電子 計算機はその構成要素を、真空管、半導体、集積回路、大規模集積回路、・・・と進化発展 させながら、計算能力や記憶容量を指数のスケールで計測せねばならぬほどに急速発展させてきました。このようなハードウェアの目覚ましい性能向上に随伴して、機械翻訳の性能も向上したと言えます。一方また、機械翻訳の性能改善は遅々として、誕生期の片言翻訳の水準からあまり進歩していない、という見方もできます。

機械翻訳の出力訳文をそのまま翻訳書に採用して物議を醸した事件(註:アインシュタイン その生涯と宇宙(下)初版における誤訳。その後訂正版が出版されました)は記憶に新しいでしょう。この事件は、「肉体は滅ぶれども霊魂は常しえなり」と訳すべきところを、「肉は腐ったが酒は大丈夫だ」と機械翻訳した昔の誤訳の現代版と言えるかもしれません。

要するに、機械翻訳は未だ人間(翻訳者による)翻訳には、到底及ばない水準にあり、 この隔たりを埋めることは著しく困難であります。

以上述べたことは、機械翻訳を嫌う立場からの御託ということもできましょう。現代社会は、多言語による大量の文書(あるいは音声)情報が飛び交い、これを翻訳して理解せねば成り立ち行かぬ状況にあります。限られた人手による翻訳では処理不可能です。たとえば、来る 2020 年の東京五輪においては、多数の外国人の対応接待のために機械翻訳を導入することが不可避と予想されます。

AAMTの会員諸氏にとっては、釈迦に説法ですが話の流れのために付言します。現代の機械翻訳(MT)は、下記の2つのタイプあるいは3つのタイプに大別できます。

- 1) GBMT (Grammar Based MT,):RBMT(Rule Based MT)と呼ぶこともあります。古典的な人工知能型機械翻訳です。人手で翻訳文法や翻訳辞書などの言語知識ベースを構築する方式です。
- 2) SBMT (Statistics Based MT):言語コーパス、特に大量の訳文 原文の対データを 統計処理しつつ言語知識ベースを構築する方式です。インターネット時代に対応した近代 的MTと言えるかもしれません。
- 3) Hybrid MT: GBMT と SBMT の組み合わせ方式です。様々な組み合わせ方がありますが、文法や辞書を統計的に下拵えしてから人手を加える、などの方式があります。

現代の機械翻訳、特にインターネット上の機械翻訳では、2)のSBMTが圧倒的な強みを見せています。2011年頃を境に、SBMTの性能はGBMTのそれを凌駕したという感触を持つ研究者が少なくありません。日々生産される新しい術語・用語・言い回し(語句)の翻訳(対訳)への対応は、人手(GBMT)では、明らかに不利・困難であります。機械翻訳の本流は、SBMTおよびHybridに転進したと考えるのが妥当でありましょう。

しかしSBMTは、常にデータ・スパースネスの問題に苦しめられています。2-gram や3-gram のレベルならば、対訳コーパスの稠密性はほぼ確保できますが、4-gram となると対訳コーパスは希薄な言語空間に発散します。対訳コーパスの品質管理も難しい課題です。

紙数が尽きかかってきましたので、小生の現在の"MT/翻訳とのかかわり"について一言します。

機械翻訳と人手翻訳の狭間に坐して、両者の中間にあるものについて漠然と考えています。原文 (テキスト) の持つ空気 (雰囲気) を訳文に持ち込む工夫、翻訳理論でいう文化翻訳、モノクロ写真のような翻訳ではなくフルカラーの翻訳、などと言われるものです。このような目的のために、俳句などの詩文の翻訳を人手と機械 (GBMT) で小規模実験しています。GBMT信奉者のように見なされているようですが、SBMTの将来性と威力は十二分に理解しているつもりであります。

#### AAMTへの要望

AAMTは機械翻訳そして一般の自然言語処理の研究開発に従事する個人や組織が、情報 交換するための、開かれた場であると思います。同時にまた自然言語処理(つまりコンピュータ・プログラムによる処理)とは縁の薄い、翻訳業務、同時通訳、言語教育、文学、芸術、数学、哲学、美学、宗教など「何らかの様式で言葉を扱う分野」に属する方々のための場でもあるように思います。

前者の観点から、これまでAAMTは機械翻訳研究の進展に多大なる貢献をしてきました。今後は後者の観点から、コンピュータ利用技術から離れた分野の方々を招き、発表や議論をしていただき交流(意見交換)する努力や工夫をされると、さらなる進展があるように感じております。

以上

#### 第24回通常総会および関連行事の報告

AAMT 事務局

当協会の第 24 回通常総会が 2014 年 6 月 17 日 (火) 13 時より・ホテルアジュール竹芝に て開催されました。総会後、各委員会からの報告会、講演会、さらに今年度が行われた。そし て第 9 回 AAMT 長尾賞授与式と今年度から新たに設けられた第 1 回 AAMT 長尾賞学生奨励賞 授与式と、AAMT 長尾賞受賞者による記念講演会が盛況のうちに行われました。各参加者数 は下記に示す通りです。

総会の参加者:個人会員:出席26名、委任状参加者25名、

法人会員:出席17法人(27名)、委任状参加4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: 34名 法人会員: 31名 会員外一般:23名

合計:88名

懇親会:

個人会員:23名 法人会員:32名 会員外一般:21名 合計:80名

#### 第24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 2013年度事業報告(案) 第2号議案 2013年度決算報告(案)

第3号議案 2014年度事業計画 (案) 第4号議案 2014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

第5号議案として、以下の方々が平成26年度、27年度の役員として提案され、 当総会議決により、役員が決まりました。

#### 理事(50音順)

飯田 仁 (東京工科大学名誉教授)

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)

小谷 泰造 (株式会社インターグループ)

坂本 義行

辻井 潤一 (マイクロソフトリサーチアジア)

中岩 浩巳 (名古屋大学)

長尾 真 (京都大学名誉教授)

永田 昌明 (日本電信電話株式会社)

長谷川 英一(一般社団法人電子情報技術産業協会)

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)

Key-Sun Choi (韓国 KAIST)

Virach Sornlertlamvanich (タイ TPA)

#### · 監事 (50音順)

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) 種子田 暁夫 (日本電気株式会社)

#### 報告会

開会挨拶

会長・名古屋大学・中岩浩巳

1. 機械翻訳課題調査委員会 委員長 長瀬 友樹 ((株)富士通研究所)

2. AAMT/Japio 特許翻訳研究会 副委員長 横山 晶一 (山形大学)

3. インターネットワーキンググループ リーダー 富士 秀 ((株)富士通研究所)

4. 編集委員会 委員長 宇津呂 武仁(筑波大学)

#### 講演会

・外国人旅行者の受入体制整備と機械翻訳の可能性

山田 洋氏(日本政府観光局事業連携推進部 観光情報戦略室長)

・機械翻訳を使いこなす~品質と生産性の向上のために~

徳田 愛、エイリー・エイドリアン氏 (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)

#### • 展示会

- ・特許翻訳などの専門的な外国語文書も自然な日本語に!
  - ~多言語統計翻訳プラットフォーム~ 日本電信電話株式会社
- 情報通信研究機構における自動翻訳研究 独立行政法人 情報通信研究機構
- ・翻訳支援ツール Transit NXT 株式会社 シュタール ジャパン
- ・未来を拓く! グローバル翻訳ソリューション 株式会社高電社
- ・企業向け翻訳システム株式会社クロスランゲージ
- · AAMT 機械翻訳課題調査委員会活動報告 AAMT 機械翻訳課題調査委員会

# 第9回 AAMT 長尾賞・第1回 AAMT 長尾賞学生奨励賞授与式 及び第9回 AAMT 長尾賞記念講演会

第9回 AAMT 長尾賞

第9回 AAMT 長尾賞は以下の方が受賞しました。

受賞者: (独)情報通信研究機構

ユニバーサルコミュニケーション研究所多言語翻訳研究室

内山将夫、隅田英一郎

#### 受賞理由:

語順変換と訳語選択との独立実行による統計翻訳技術の研究実用化ならびに複数企業への技術移転による高性能機械翻訳システムのサービス実現による功績が顕著なためである。これは、長尾賞の設立趣旨である、「高性能の機械翻訳システムを商品化した」、あるいは「機械翻訳システムを使った新しいサービスを開始した」という観点を鑑み、長尾賞にふさわしいと認めるものである。

選考委員長:飯田仁(東京工科大学)

選考委員 :Key-Sun Choi (韓国 KAIST) 宇津呂 武仁(筑波大学)

横山 晶一(山形大学)

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)



#### 第1回 AAMT 長尾賞学生奨励賞

本年度から新たに AAMT 長尾賞学生奨励賞が設定され、第1回 AAMT 長尾賞学生奨励賞を以下の方が受賞しました。

受賞者:東京大学 大学院教育学研究科・博士課程1年

宮田 玲

#### 受賞対象論文:

影浦峡先生の指導のものに、2013年度東京大学大学院研究科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。

「自治体ウェブサイト文書の多言語化を支援する枠組みと システム環境の研究」

#### 受賞理由:

本論文では、自治体の手続き型文書を対象として、機械翻訳技術を活用した多言語翻訳に必要な制限言語ルールやオーサリング環境を提案している。制限言語とその機械翻訳への効果を検証する研究は多いが、本論文は図書館情報学的な立場から文書構造や文書の作成過程にまで踏み込んだ独自の方法論を提示している点が高く評価できる。本論文は、機械翻訳に関連する新しいシステムやサービスに関する、理論的かつ実証的な優れた研究報告であり、AAMT長尾学生奨励賞にふさわしいと考える。

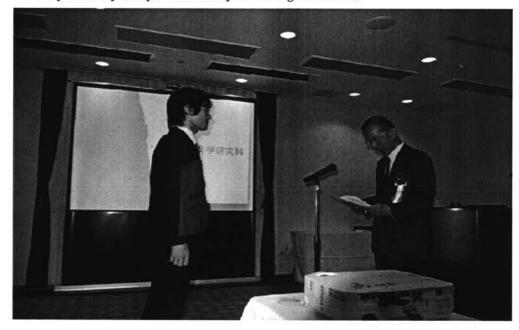
選考委員長:永田昌明(日本電信電話株式会社)

選考委員 : 鈴木博和 ( (株)東芝)

富士 秀 (富士通株式会社)

推薦者:影浦峡(東京大学・大学院教育学研究科)

Tony Hartley(Tokyo University of Foreign Studies)



# 協会活動報告

(2014年6月~2014年8月)

#### 第24回通常総会

2014年6月17日

第 1 号議案 2013 年度事業報告 (案) 第2号議案 2013年度決算報告(案)

第3号議案 2014年度事業計画(案) 第4号議案 2014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

2014年6月17日

①機械翻訳課題調查委員会 ②AAMT/Japio 特許翻訳研究会

③インターネット WG

④編集委員会

#### 讚演会

#### ○講演:

・外国人旅行者の受入体制整備と機械翻訳の可能性

山田 洋氏(日本政府観光局事業連携推進部 観光情報戦略室長)

・機械翻訳を使いこなす~品質と生産性の向上のために~

徳田 愛、エイリー・エイドリアン氏

(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)

#### 第9回 AAMT 長尾賞授与式·記念讚演会

受賞者:(独)情報通信研究機構

ユニバーサルコミュニケーション研究所多言語翻訳研究室

内山将夫、隅田英一郎

#### 受賞理由:

語順変換と訳語選択との独立実行による統計翻訳技術の研究実用化ならび に複数企業への技術移転による高性能機械翻訳システムのサービス実現による 功績が顕著なためである。これは、長尾賞の設立趣旨である、「高性能の機械翻 訳システムを商品化した」、あるいは「機械翻訳システムを使った新しいサービ スを開始した」という観点を鑑み、長尾賞にふさわしいと認めるものである。

#### 第1回 AAMT 長尾賞学生奨励賞

本年度から新たに AAMT 長尾賞学生奨励賞が設定され、第1回 AAMT 長尾賞学生奨励賞を以下 の方が受賞しました。

受賞者:東京大学 大学院教育学研究科・博士課程1年

宮田 玲

#### 受賞対象論文:

「自治体ウェブサイト文書の多言語化を支援する枠組みとシステム環境の研究」 受賞理由:

> 本論文では、自治体の手続き型文書を対象として、機械翻訳技術を活用した 多言語翻訳に必要な制限言語ルールやオーサリング環境を提案している。制限 言語とその機械翻訳への効果を検証する研究は多いが、本論文は図書館情報学 的な立場から文書構造や文書の作成過程にまで踏み込んだ独自の方法論を提示 している点が高く評価できる。本論文は、機械翻訳に関連する新しいシステム やサービスに関する、理論的かつ実証的な優れた研究報告であり、AAMT 長尾学 生奨励賞にふさわしいと考える。

#### 懇親会

2014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』

#### 決算理事会

2014年6月17日

第1号議案 2013年度事業報告(案) 第2号議案 2013年度決算報告(案)

第3号議案 2014年度事業計画(案) 第4号議案 2014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

2014年5月23日 (2014年度 第2回)

- ① 前回委員会の議事録の確認
- ② 新委員の着任と各自の自己紹介
- ③ 各 WG の活動について(各 WG に分かれて議論)(WG3)
  - · AAMT 総会(2014/6/17)の準備
  - · UTX の新しいウェブサイトの確認
  - ·UTX 変換ツールのライセンス
  - ・UTX 仕様(論点表)の確認
  - その他
- ④ 活動内容の報告(各 WG から)
- ⑤ 活動内容についての議論
  - 総会のネーミングに関する検討
  - ・展示ブースの申し込み状況の確認
  - · AAMT ホームページ進展状況の報告
    - ・総会準備状況の確認
  - ・まとめと次回委員会について

#### 2014年6月27日 (2014年度 第3回)

- ① 前回委員会の議事録の確認
- ② 各 WG の活動について(各 WG に分かれて議論)(WG1,WG2)
  - ・評価サイト公開に関する検討
  - ・サイトID、パスワードの発行管理について
  - 利用規約、個人利用の扱い
  - ・無料翻訳サイト(日中、中日)の評価

#### (WG3)

・AAMT MT フェア(2014/6/17)の反省

- ・TBX-Min 関係者からの UTX に対するコメント
- 人工知能研究振興財団 人工知能研究助成応募
- ・AAMT 翻訳フェア展示ブースの内容確認:質問への応対、UTX パンフ配布、 UTX の自動デモ
- ③ 活動内容の報告(各 WG から)
- ④ 全体会議
  - ・翻訳フェア(6/17)の反省
  - · AAMT ホームページのトラブルについて
  - ・本年度の JTF 翻訳祭の参加内容検討
- ⑤ まとめと次回委員会について

#### 2014年7月25日 (2014年度 第4回)

- ① 前回委員会の議事録の確認
- ② 各 WG の活動について (各 WG に分かれて議論) (WG3)
  - ・TBX, UTX (loose), TBX (strict)の関係の検討
  - ・TBX-Min 側からの UTX に対するコメントについて
  - 論点とその結論のまとめ
  - ・UTX Specification Version 1.20(2014/4/28 版)の仕様の再検討
- ③ 活動内容の報告(各 WG から)
- ④ 全体会議
  - ・AAMT JTF 共同機械翻訳勉強会の報告
  - ・工藤による ISO17100 規格検討会 (翻訳) および ISO18587 規格検討会 (ポストエディット) および通訳 ISO の規格検討会の報告
  - ・翻訳祭における翻訳プラザへの出展およびセッションへの参加について
- ⑤ まとめと次回委員会について

#### インターネット WG

- ①新 AAMT ホームページ公開に向けた体制の立ち上げ
- ②AAMT MT フェアでの新 AAMT ホームページの紹介
- ③AAMT サイトのセキュリティー対応
- ④総会・MT フェアに対応した AAMT ホームページの更新

#### 編集委員会

2014年6月17日 (2014年度 第1回)

- ① AAMT Journal No.57 の記事の執筆依頼候補者の検討 2014 年 6 月 25 日(2014 年度 第 2 回)
  - ① AAMT Journal No.57 の記事の執筆依頼計画について
  - ② AAMT Journal No.57 のスケジュールについて

#### AAMT/Japio 特許翻訳研究会

2014年7月11日(金)(2014年度 第2回)

- 1. 研究会活動計画について
- 2. クラウドソーシングについて
- 3. JST での日中翻訳プロジェクトの現状 (中澤先生)
- 4. シンポジウム関連
- 5. その他
- 6. 次回の開催について
  - ・開催の日時(場所)
  - ・主な議題

編集後記

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 57 号をお送りします。

今号の巻頭言は、株式会社クロスランゲージアールアンドディ代表取締役 古賀勝夫様より、御寄稿を頂きました。

また、今号が会員の皆様のお手元に届くその少し後、来る 11 月 26 日(水)に、AAMT が後援をします第 24 回 JTF 翻訳祭が開催されます。日本翻訳連盟(JTF) 翻訳祭企画実行委員会の担当の方からは、この第 24 回 JTF 翻訳祭のお知らせをご寄稿いただきました。

一方、AAMTの活動として、今号に先立ちまして、2014年6月に開催されました総会におきまして、日本政府 観光局事業連携推進部 観光情報戦略室長山田洋氏、ならびに、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 徳田愛氏、エイリー・エイドリアン氏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詳細な御寄稿を頂きました。

あわせて、第9回の AAMT 長尾賞の選考結果を受けまして、受賞者である (独)情報通信研究機構 ユニバーサルコミュニケーション研究所多言語翻訳研究室内山将夫氏、隅田英一郎氏より、受賞理由となった「言語構造が大きく異なる言語間の長文を対象とする統計的機械翻訳の実用化」についての紹介記事をご寄稿していただきました。また、本年より新設されました AAMT 長尾賞学生奨励賞の第1回の選考結果を受けまして、受賞者である東京大学 大学院教育学研究科・博士課程1年宮田玲氏より、受賞理由となった「自治体文書の多言語化を支援する枠組みとシステム環境の研究」についての紹介記事をご寄稿していただきました。

また、総会会場におきまして、「AAMT 機械翻訳フェア(MT フェア)2014」と題しまして、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初めて実施したのを受けまして、機械翻訳課題調 査委員会による、MT フェア開催報告を掲載いたしました。

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四弾を頂きました。

プロジェクト報告として、前会長、豊橋技術科学大学 井佐原均氏より、「豊橋技術科学大学における機械翻訳の研究開発」の御寄稿を頂きました。

また、研究関係に関する御寄稿として、三田教育研究所 LLP 平田周氏より、「機械翻訳を支援ツールとして 使う英語学習法」に関する御寄稿を頂きました。 さらに、シンポジウム報告として、北陸先端科学技術大学院大学 島津明氏、TAUS (Translation Automaton User Society) Jaap van der Meer 氏、東京大学・大学院教育学研究科 影浦峡氏、(株)東芝 研究開発センター 鈴木博和氏より、第 5 回産業日本語研究会・シンポジウムにおける「法令工学の言語処理」、TAUS Tokyo Executive Forum 2014、TKE (Terminology and KnowledgeEngineering) 2014、テクニカルコミュニケーションシンポジウム 2014 について、それぞれ御報告頂きました。

その他、2014 年 6 月にドイツのベルリンで開催された TC37: Terminology and other language and content resources 総会の報告を兼ねて開催されたパネルディスカッション「ISO17100 と 日本における翻訳職業教育の将来」に AAMT-MT 委員会委員が参加させていただき、主催者(日本翻訳連盟様)およびパネラーの方々のご了解をいた だき、そこで行われた議論について、掲載させて頂きました。

AAMT 内の活動報告として、機械翻訳課題調査委員会から、「テストセットに基づく日中/中日翻訳自動評価サイト公開について」の報告、および、「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。

その他、「AAMT 会員のひろば」の企画におきましては、法人会員、個人会員、各 1 件の紹介文を掲載しました。

# AAMT

#### Asia-Pacific Association for Machine Translation

### AAMT 入会のご案内

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

機械翻訳にご関心のあるすべての方にご入会をお勧めします。

#### \* \* AAMT 会員の特典 \* \*

#### 1.AAMT Journal の購読ができます。

会員には、機関誌である AAMT Journal (年  $2\sim3$  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、最新の機会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

Asia-Pacific Association for Machine Translation

# 入会申込書

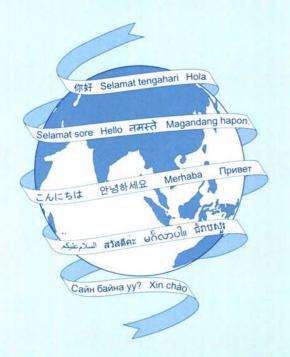
以下の通り、アジア太平洋機械翻訳協会の会員申し込みを致します。

申込日 20 年 月 日

氏名(ローマ字)			
氏名(漢字)			
電話番号			
メールアドレス			
所属先			
所属先住所	₸		
種別	ユーザ	研究開発者	その他
機械翻訳に関するお知らせメールの配信	希望す	でる 希望	見しない
コメント			



# AAMT



# AAMTジャーナル No.57

発行: アジア太平洋機械翻訳協会 (AAMT)

ホームページ: http://www.aamt.info

住所: 〒171-0014 東京都豊島区池袋2-55-2鈴木ビル3階

(株)日本システムアプリケーション内

phone: 03-5951-3961 fax: 03-5951-3966

編集委員会:宇津呂 武仁 小谷 克則 大倉 清司

鈴木 博和 阿部 さつき

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