

AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2016

No.62

アジア太平洋機械翻訳協会

目 次

巻頭言：	機械翻訳と言語翻訳..... 飯田 仁.....1
国際シンポジウム案内：	Coling 2016 隅田 英一郎.....3
国際シンポジウム案内：	「MT Summit XVI」のお知らせ 中岩 浩巳.....6
イベント報告：	イベント参加報告 ～LocWorld 30 Tokyo～ 目次 由美子.....8
研究会参加報告：	情報処理学会 第 122 回 情報基礎とアクセス技術研究会・ 第 101 回ドキュメントコミュニケーション研究会 合同 目次 由美子.....11
連載：	機械翻訳事始め[VII]..... 坂本 義行.....13
AAMT25 周年記念寄稿：	AAMT 設立までのこと 長尾 真.....26
レポート：	「グローバル・コミュニケーション計画」の取組状況..... 河野 弘毅.....30
製品紹介：	スマートフォン向け音声翻訳アプリ RECAIUS™ 「音声トランスレータ」 岸本 麻須美.....34
製品紹介：	多言語翻訳を支える大規模な辞書データベース 春遍 雀来.....36
会員の広場（法人会員）：	 株式会社イデア・インスティテュート.....40
委員会活動報告：	これまでの AAMT Forum メールマガジン Vol.8..... 機械翻訳課題調査委員会 WG1,2.....42
委員会活動報告：	実務翻訳における機械翻訳の利用に関するアンケート結果報告 機械翻訳課題調査委員会 WG3.....46
事務局からのお知らせ：	協会活動報告（2016 年 2 月～2016 年 4 月） AAMT 事務局.....58
編集後記：	62

CONTENT

Foreword:	Machine Translation and 'Language Translation' H.Iida 1
Symposium Information:	Coling (International Conference on Computational Linguistics) in Osaka, Japan E.Sumita 3
Symposium Information:	Machine Translation Summit XVI H.Nakaiwa 6
Research Meeting Report:	Event Report - LocWorld 30 Tokyo - Y.Metsugi 8
Research Meeting Report:	- Information Processing Society of Japan -Joint Research Meeting: the 122nd Information Y.Metsugi 11
Series:	History of Machine Translation [VII] Y.Sakamoto 13
in honor of the 25th anniversary of AAMT:	At the time of the establishment of AAMT M.Nagao 26
Report:	“Global communication project” Interview on MIC H.Kawano 30
Products:	RECAIUSTM “Speech Translator”, the speech-to-speech translation application for smartphones. M.Kishimoto 34
Products:	Large-scale dictionary databases to support multilingual translation systems J.Halpern 36
AAMT Members:	 IDEA INSTITUTE INC 40
Committee Report:	AAMT Forum mail magazine vol.8 WG1, WG2 42
Committee Report:	UTX 1.20 Beta Specification (Excerpt) WG3 46
AAMT Activities:	AAMT Activities (from February 2016 to April 2016) 58
Editor's Note:	 T.Utsuro 62

機械翻訳と言語翻訳

飯田 仁

アジア太平洋機械翻訳協会

来年2017年9月に第16回MTサミットが名古屋で開催される。機械翻訳研究開始50周年のMTサミット記念大会がサンディエゴで開催されたのが1997年であり、名古屋の会議が70周年大会となる。MT研究・開発について語られる新たなページが加えられる。さらに、4年後に迫る東京五輪で音声翻訳システムなどの活躍が強く期待されていて、国からの後押しの下、MTの歴史に一層の輝きを放つことになると思われる。21世紀に入って、機械翻訳の実用化と社会への浸透という点で、2002年の日韓共同開催のサッカーワールドカップ、08年の北京五輪、12年のロンドン五輪における多言語翻訳サービスの実証実験があった。北京においては、タクシーでの行き先指定などの音声対話などがメディアにも取り上げられた。それからの十数年後の進化に注目が集まっていると言える。

MTの歴史が暗号解読という特殊業務から派生し、米国の露英翻訳の試みであったMT黎明期（1954年のIBM&Georgetown大）、その後60年代の同大を基点として開発された露英システムMARKIIや英仏など多言語システムに展開するSYSTRAN、さらにベトナム戦争向けの英越システムLOGOSなどが開発され、MT研究が米国の対外戦略に深く関わってきた面が感じ取れる。同時にそれらシステムの性能評価がコスト計算を含めて人手翻訳との対比で進められることになり、米ソ雪解け後の66年のALPACレポートに至る。そして、67年からの米国MT予算の縮小と科学技術研究のAI（人工知能）シフトが始まった。推論手法やプロダクション・システム、それから80年代に入ってエキスパートシステムなどに注目が集まり、56年のダートマス会議で産声を上げたAIが米国において軍事利用を含め実用化されていった。ただ、当時から自己学習能力を持たないことが課題とされていた。60年代

のMTシステムも経験主義的なものであると指摘されていた。特定のテキストに適合する文法規則を随時用意する枠組みから脱却するために、普遍的で最小の文法規則（ $X \cdot X \rightarrow X$ ）により句構造解析する単一化操作を使った手法が試みられたこともあったが、統語的、かつ意味的な制約を語彙に書き込む作業（語彙記述）は煩雑を極め、言語活動の多様さを扱うには無理があった。この作業は正に手動設計の典型である。

ところが、皆さんご存知のように、今春‘アルファ碁’が世界トップ級囲碁棋士を圧倒した。メディアは一斉にAIが人間の知的ゲーム能力を超えたと報道し、人間とAIの勝負の歴史と題した年表が掲載されるほど注目を集めた。2012年にはグーグルキャットが登場していて、AIが猫の特徴を自己学習し、自ら猫を表現して見せた。これらの技術的快挙はディープラーニング（深層学習）による成果である。ただ、前者の勝者判定は全面的に納得できるものではない。例えば相撲四十八手を身につけた大型ロボットが力自慢の力士に勝つことは見るまでもなく明らかだろう。一方、囲碁の勝負は知的鍛錬と精神力、集中力を生身の人間同士が競うもので、膨大な対局をこなし、すべての経験を肥やしにして疲労しない脳をもつAIマシンが棋士と競うのは対等ではないでしょう。アルファ碁は深層学習により囲碁ゲームで勝つということを自動的に獲得し、さらに強化学習により対局を重ねて、勝ち手順を身につけている。囲碁における勝ちがどのように表現されるか、膨大な写真の中から猫がどのように表現されるのか、自動的に獲得する学習能力を実現したのが深層学習と言える。つまり、深層学習の価値は、表現すべき何者かがあることが前提であって、当たり前であるが学習する対象があることである。何百万人分の顔画像を学習対象として、数億枚の顔画像を99%以上

の精度で自動的にクラスタリングできることも示されている。そして、深層学習は言語理解や‘自動翻訳’への適用が期待されている。

機械翻訳に話を戻すと、1990年に統計的機械翻訳が登場し、その発展の中で、自動的な評価手法が様々提唱され、BLEU が中心的な位置を占めている。現状、一言で述べれば、通信路モデルを基礎として、統計的機械翻訳の研究は言語モデル、翻訳モデル、アライメントの表現を如何に最適なものとするかという翻訳問題を規定し、基礎的な研究が進展し続けていると言える。そこでは、深層学習に関して前述したように、まず原言語と目的言語の対のデータがあり、そのコーパスがあることが前提となる。90年代前半のユーゴ紛争の最中、英語－セルビア語の機械翻訳を一緒にやって欲しいと声を掛けられたことがあったが、一定の対訳コーパスがあれば、現状の統計的機械翻訳のレベルであればすぐにでも学習し、機械翻訳を実現できたのかと思う。つまり、統計的機械翻訳が文法規則や用例に縛られることなく学習していく能力を備えていて、一語単位かフレーズ単位かによるアライメントを前提とする対訳コーパスに依存するものの、システム開発の立ち上がりが早いだろう。併せて、翻訳精度がコーパスの量に依存し、十分な量を満たせば、言語モデルも翻訳モデルも言語活動の実態を反映するものとなるのだろう。

しかし、人間の言語活動の多様さを捉えるためには、所謂大規模な対訳コーパスと言語コーパスとで十分と言えるであろうか。今年2月に集英社インターナショナルから「日本語を作った男」(山口諤司著)という大著が出版された。明治期終盤にやっと標準日本語をまとめた言語学者の努力と世の中の言語活動についてたっぷりと書き込まれている。当時は、書き言葉と話し言葉の区別もなく、日本中で様々な日本語が飛び交っていたということである。そして、戦後になって話し言葉の収集なども進められてきた。今年2月末には第7回となる産業日本語研究会のシンポジウムが開催され、日本語マニュアル(試作版)が配布された。その

マニュアルは日本語という言語装置の取り扱い説明書であり、その言語装置はビジネス活動や産業活動で使われるものに限ると明記してある。解釈に誤解が生じにくく、曖昧性が少なく機械処理ができ、機械翻訳にも適した日本語の使い方がまとめられている。これは、ビジネス活動で使用する日本語を明確に規定しようという試みである。話し言葉コーパス構築の国レベルのプロジェクトが立ち上がっているが、話し言葉に常に現れると言える終助詞について、音響的な情報と同列にパラ言語情報として扱い、コーパス情報として扱わないような方針もあり、言語モデル作成に十分な情報を提供し得るのかという意見もある。さらに、日常の言語活動は、従来の書き言葉と話し言葉の範疇には納まらない、所謂スマホ間の‘Line ことば’が大きな存在になっている。そこでは、終助詞に代わる絵文字などが使われ、終助詞が担っている話し手の感情や気持ちを一層強調している。また、言語表現自体も新たな用法によって作り出されている。21世紀に入ってから20年にならんとするとき、言語コーパスも書き言葉と話し言葉という大区分ではなく、産業日本語やその対極にある‘モバイル間対話言語’、さらにスマホのアプリに応じた(あるいは、アプリを想定した)QA用言語などを捉える公的な言語コーパス構築が強く求められる。それにより、言語モデルや適切な対訳コーパスに基づく翻訳モデルをより十分なものとして、統計的機械翻訳の本来の能力を発揮させることができると考える。MT70周年以後の展開として、このような点を捉えておきたいと思う。

そして、言語コーパスの構築の下、言語使用が定まっている言語表現を適確に翻訳することを「機械翻訳」として改めて位置づけ、その一方で、小説や詩にまでに及ぶ言語表現を翻訳する内的行為を仮に称して『言語翻訳』と位置付けることが望ましいかと考える。そうすることにより、従来からの機械翻訳という問題の規定に向け一歩前進すると考える。



<http://coling2016.anlp.jp/>

隅田 英一郎

国立研究開発法人情報通信研究機構 (NICT)

1. はじめに

人間の言葉をコンピューターが理解し、人間とコンピューターがコミュニケーションする近未来に期待が高まっています。基盤になる技術は自然言語処理と呼ばれ、現在、熱い注目が集まっている人工知能、ロボット、ビッグデータの分野に不可欠な技術です。その可能性に鑑み、多くの国の、産、学、官で活発な研究開発が進められております。

そんな中、自然言語処理の最高峰の国際会議の一つである COLING が 2016 年 12 月に大阪¹で開催されることになりました。会期は、チュートリアルとワークショップが 11~12 日、本会議が 13~16 日となっております。

1980 年に東京で開催され、1994 年に京都で開催され、22 年ぶりに日本へ戻ってきました。大阪では初開催となります。COLING は 1965 年にニューヨークで始まり、隔年²で開催され、半世紀の歴史があり、今回の大阪開催は 26 回目に当たります。

世界のトップ研究者 700 人前後が COLING に集い、1 週間ほどの開催期間で、最先端の研究発表と白熱した議論が行われます。

また、人工知能のビジネス化が広く関心を集める中、

¹<http://coling2016.anlp.jp/>が本会議のサイトになります。大阪にかかわる情報も掲載しております。大阪の魅力を再確認いただき、是非参加をご検討ください。

²第 6 回は例外で 3 年目に開催。

日本の「商業」と「ものづくり」の中心地である大阪で開催する意義は高いといえます。

2. 経緯

自然言語処理の難関国際会議は長い間日本で開催されていませんでしたが、日本のプレゼンスを高めることを目的として、日本の言語処理学会³ではワーキンググループを設置し、これらの国際会議の誘致活動を 2 年前から行ってきたところです。2014 年の 8 月の開かれた COLING の開催地を決めるコンペで、ライバルを抑えて、日本開催が決定されました。

3. 組織

COLING2016 は、計算言語学国際委員会⁴ (International Committee on Computational Linguistics (ICCL)、現会長は産総研人工知能センター(AIRC) センター長辻井潤一博士) の後援の下に、言語処理学会が主催します。

General Chair は Nicoletta Calzolari (CNR、イタリア)、Program Chair は松本祐治(奈良先端大)、Rashmi Prasad (University of Wisconsin-Milwaukee、米国)、Local Chair は隅田英一郎(NICT)、徳永健伸(東工大)、黒橋貞夫(京大) で進めています。

³<http://www.anlp.jp/index.html>

⁴ <http://nlp.shef.ac.uk/iccl/>

4. 会議の構成

全てのセッションは大阪府立国際会議場（グランキューブ大阪）で行われます。大阪の中心（〒530-0005 大阪府大阪市北区中之島 5 丁目 3-51）にあり、アクセスの良い会議場です。

テーマは、言語の文法・意味のような基礎から、自動翻訳やロボットとの対話などの応用システムまで幅広い構成です。また、近年のその有用性が多方面で確認されつつある Deep Learning などの新技術の研究も大幅に増えると予想されます。

論文は募集中で、詳細の確定はこれからになります。が、これまでの実績を踏まえると、少なくとも、200 本前後の論文、4 件前後の招待講演、多数のチュートリアル、ワークショップ、デモ展示が想定されます。

また、古都奈良の興福寺と東大寺を巡る英語ガイド付きのバスツアーを用意し、普段、国宝に触れる機会の少ない日本の参加者も世界の参加者と一緒に、日本の歴史に触れていただきたいと考えています。また、和食を中心としたバンケットも用意していますので、交流の機会としてご活用いただけます。

5. アジアや若手への配慮

自然言語処理は英語を中心に始まり、日本語、中国語などで盛んになってきました。近年アジアの言語に対する研究が盛り上がっていますが、アジア諸国の研究機関は予算的に恵まれないことが多いという状況です。ですから、移動が容易な日本で国際会議を開催することは、アジア諸国の研究者の集結を容易にする効果もあります。

また、次代を担う若手研究者の参加を促進するために、学生参加者の参加費用を優遇する施策、学生参加者の渡航費の一部を支援する施策を検討しています。

6. チュートリアルとワークショップ

基礎から応用にわたって多様なトピックについて、7 件のチュートリアルと 19 件のワークショップを準備しています。翻訳やアジア言語にかかわるものもいくつかあります。

6.1. チュートリアル

- ① Compositional Distributional Models of Meaning
- ② Translationese: Between Human and Machine Translation
- ③ Natural Language Processing for Intelligent Access to Scientific Information
- ④ Translation Quality Estimation
- ⑤ The Role of Wikipedia in Text Analysis and Retrieval
- ⑥ Chinese Textual Sentiment Analysis: Datasets, Resources and Tools
- ⑦ Succinct Data Structures for NLP-at-Scale

6.2. ワークショップ

- ① The 2nd Workshop on Noisy User-generated Text (W-NUT 2016)
- ② Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 3)
- ③ The 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA-3)
- ④ The 3rd Workshop on Asian Translation (WAT2016)
- ⑤ Language Technology Resources and Tools for Digital Humanities (LT for DH)
- ⑥ Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM)
- ⑦ Cognitive Aspects of the Lexicon (CogALex-V)
- ⑧ 6th Workshop on South and Southeast Asian

Natural Language Processing (WSSANLP)

- ⑨ 5th International Workshop on Computational Terminology (CompuTerm 2016)
- ⑩ Computational Linguistics for Linguistic Complexity (CL4LC)
- ⑪ Workshop for Clinical Natural Language Processing
- ⑫ Third International Workshop on Worldwide Language Service Infrastructure (WLSI-3) and Second Workshop on Open Infrastructures and Analysis Frameworks for Human Language Technologies (OIAF4HLT-2)
- ⑬ Open Knowledge Base and Question Answering Workshop (OKBQA 2016)
- ⑭ The 12th Workshop on Asian Language Resources
- ⑮ Grammar and lexicon: interactions and interfaces
- ⑯ Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016)
- ⑰ Fifth Workshop on Hybrid Approaches to Translation (HyTra-5)
- ⑱ Workshop on Speech-centric Natural Language Processing
- ⑲ Workshop on Computational Modeling of People's Opinions, Personality, And Emotions in Social Media

7. 重要な日程

締切などの今後の主要な日程（予定）は以下の通りです。

- （会議登録開始）2016 年 7 月 1 日
- （論文投稿締切）2016 年 7 月 15 日
- （デモ論文投稿締切）2016 年 8 月 26 日
- （ワークショップの論文投稿締切）各ワークショップの WEB サイトでご確認ください。

- （展示&スポンサー締切）2016 年 11 月初旬

8. 参加申込や照会

参加申し込みは COLING 2016 のサイト⁵で行います。
また、不明の点は、メール⁶でご照会ください。

9. 終わりに

ビッグデータの利活用や人工知能の研究・開発・事業化において、自然言語処理への期待が近年、急速に高まっており、また、2020 年の東京でのオリンピック・パラリンピックの開催に向けて多言語翻訳システムの性能向上も求められています。

このような状況の中で、2016 年に自然言語処理の国際会議 Coling を日本で開催することは、2020 年をゴールの一つにして磨いている日本の技術を世界に向けてアピールする絶好のステージを提供していると言えます。

特に、AAMT の推進している多言語自動翻訳は、言葉の壁を無くしてコミュニケーションを実現するものがありますから、Coling としては、AAMT 会員に、積極的な参加（論文投稿、聴講、デモ、展示&スポンサー⁷）を期待しています。

⁵<http://coling2016.anlp.jp/>

⁶inquiry@coling2016.anlp.jp

⁷会議のサイトやコンファレンスバッグへのロゴの掲載等の宣伝の機会に加え、レセプションでの 1 分トークや展示スペースなど、会議参加者とスポンサーとの交流機会を多数用意しています。詳細は、COLING 2016 のサイト (<http://coling2016.anlp.jp/>) 内の Invitation to Sponsorship をご覧ください。

「MT Summit XVI」のお知らせ

アジア太平洋機械翻訳協会(AAMT)

会長 中岩 浩巳

1. MT Summit とは

MT Summit は、機械翻訳に関する様々な最新の成果の発表や、招待講演などが行われる国際会議です。この会議の特徴は、機械翻訳の研究開発者に加え、翻訳者、翻訳会社、クライアント、翻訳支援ツールメーカー等の機械翻訳に関連する関係者が一堂に会し、最新の機械翻訳技術に関する発表だけではなく、機械翻訳技術を翻訳業務で実際に活用する方法等に関する発表も行われる点です。その意味では、本会議に参加すると、機械翻訳に関するあらゆる観点からの動向を把握することができるといえます。

本会議は、アジア太平洋機械翻訳協会（AAMT）の上位組織である国際機械翻訳協会（IAMT）が主催で、欧州機械翻訳協会（EAMT）、アメリカ機械翻訳協会（AMTA）、及び、AAMT が交代で、欧州、北米、アジア・太平洋と開催場所を変えて、隔年で開催しています。今まで AAMT が運営した会議は、日本・箱根（第1回；1987）、日本・神戸（第4回；1993）、シンガポール（第7回；1999）、タイ・プーケット（第10回；2005）、中国・アモイ（第13回；2011）でしたが、2017年に行われる第16回の MT Summit は、24年ぶりに日本で開催することとなりました。日本は2020年のオリンピック・パラリンピック開催に向けた外国人向けの情報提供サービスとしての機械翻訳への期待の高まり、インバウンド需要の増大に伴う機械翻訳需要の増大など、今まで以上に機械翻訳が注目されているので、2017年に日本で行われる、MT Summit は盛り上がる事が期待されます。皆様の、論文投稿、及び、ご参加をお待ちいたしております（詳細は決まり次第 AAMT ホームページ、メール等でご案内します）。

2. 開催概要

【開催日】

2017年9月18日（月）～9月22日（金）

【開催場所】

名古屋大学東山キャンパス（名古屋市千種区不老町1）
名古屋市営地下鉄「名古屋大学」下車



名古屋大学東山キャンパス

【会議概要】

・論文発表

機械翻訳に関連する投稿された論文の中から査読を経て、選ばれた論文が当日発表（口頭発表もしくはポスター発表）される。

研究トラック：機械翻訳の最先端の研究開発についての論文発表が行われる。

ユーザートラック：機械翻訳を実際に翻訳プロセスに導入する方法や、うまく活用する方法など、ユーザーの視点に立った成果の発表が行われる。

・招待講演

機械翻訳の研究開発者、機械翻訳の利用に関する専門家などをお呼びし、講演をいただく。

・チュートリアル

機械翻訳に関する様々な技術、利用法等について、
チュートリアルを実施する。

- ・ 併設ワークショップ

機械翻訳に関連するトピックについて、ワークショ
ップを開催する。開催ワークショップの選定は、公
募形式で行う。

- ・ 日本翻訳連盟（JTF）様との共同セッション

日本最大の翻訳業界団体である JTF 様と共同で機械
翻訳の利用法などを対象としたセッションを実施す
る予定。

【主催・運営】

主催：国際機械翻訳協会（IAMT）

運営：アジア太平洋機械翻訳協会（AAMT）

【お問い合わせ】

アジア太平洋機械翻訳協会 事務局

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

LocWorld 30 Tokyo

目次 由美子

機械翻訳課題調査委員会

「ヤァ！LocWorld がついに日本へやって来る!!」と心を躍らせた業界関係者も多かったのではないだろうか。「ロックワールド」とは言っても音楽の祭典ではなく、れっきとした翻訳に関する展示会である。4月13日（水）～15日（金）まで、新宿の京王プラザホテルにて「LocWorld 30 Tokyo」と称するイベントが開催された。

LocWorld 01 は 2003 年に米国シアトルで開催された。LocWorld の公式サイト (<http://locworld.com>) の略歴には、The Localization Institute の Ulrich Henes 氏と MultiLingual 誌の Donna Parrish 氏が共同主催者として、ローカリゼーション業界の発展のため、友好に満ちた交流の場の形成を目的として LocWorld が展開されていることが紹介されている。1年に1回は北米、1回は欧州で開催されてきたこの国際的なイベントは、2007年に上海で開催され、そして2013年にシンガポールで開催されて以降はアジアも開催対象地域に加えられ、本2016年によりやうく東京で開催された。

「Engaging Global Customers」というテーマが掲げられた3日間にわたるイベントの初日は「プレカンファレンス」で始まった。「ラウンド テーブル」が用意され、「Game Localization」や「Localization Sales and Marketing」など、特定のトピックについて参加者同士による協議が展開された。さらに、「オープニング レセプション」も開催され、登壇者や展示ブースの担当者を含む関係者が集った。

2日目からはセッションと展示会が開催された。45分間の講演が3つのトラックで並行してスケジュールされたプログラムでは、各セッションに対して「Advanced Localization Management」、「Global Business」、「Keynote」、「Localization for Asia」、

「Technology」というラベルが示されていた。セッション タイトルには、開催地を意識した「Tailoring Your Product to Succeed in Japan」、ローカリゼーション事業の先端を伺わせる「Journey from Project Manager to Program Manager」、国際性が感じられる「Localizing for Customers and Beyond」、最新のテクノロジーを彷彿とさせる「XML in Translation」などが見受けられた。

機械翻訳に関連するであろうセッション タイトルも見受けられた。用語集に関する「Fuji Xerox – Terminology Management Implementation」、翻訳業務における機械翻訳との直面を想像させる「The Language Secret – Partnering between Consumers, PE Vendors and MT Vendors」、また、IBM 社による機械翻訳システムに関するセッションは「Surfing the Chaos of Linguistic Noise」と題されていた。



参加者で溢れたセッション会場

ジャパン・インターカルチュラル・コンサルティング社の Rochelle Kopp 氏による同時通訳付きの基調講演「Bridging Cultural Differences with the Hybrid

Approach」では、ビジネスに影響し得る日本の文化的な背景が紹介された。3 日目の午前中に開催されたもう 1 つの基調講演「The Growing Importance of Emerging Markets」では LocWorld 共同主催者の Henes 氏がモデレータを勤め、Google 社、Lionbridge 社、NEC 社、Welocalize 社からの豪華なパネリストを擁していた。

基調講演の開催中はほかのトラックでセッションを開催しない、セッションとセッションの間には 30 分間の休憩を置くことで展示ブースを訪問する時間帯を設けるなど、運営側の配慮が感じられた。



いつもと様子が違う翻訳会社の展示ブース、準備は万端

さらに、約 1 時間のランチも設けられていた。そう、LocWorld の参加料には「昼食」も含まれていた（昼食を含まない「午後券」も用意されていた）。LocWorld 参加者専用の「食堂」には大きめの丸テーブルが配され、ランチタイムでも参加者同士が親睦を深められる機会が準備されていた。

2 日目の夜には Moravia 社により「カンファレンスディナー」が提供され、展示会場の外の空気も楽しむことができた。



和気藹々としたカンファレンス ディナー

自分自身も含め、日本からは「LocWorld 初心者」が多く見受けられた。海外からの参加者には 3 回、4 回と各地の LocWorld を経験済みの人が多いように思われた。北米や欧州を本拠地とする企業にとっては地理的な条件が日本とは大きく異なり、開催地の考慮は不要なのかもしれない。また、翻訳に関するサービスやテクノロジーの情報交換の場として、LocWorld は参加しなくてはならないイベントとして認知されつつある傾向も伺えた。



コーヒーや軽食も用意された展示エリア

LocWorld への参加経験は豊富だが、来日は初めてという参加者も少なくなかった。展示ブースにも日本

国内ではまだ馴染みの薄い海外の翻訳会社や、翻訳に関するテクノロジーベンダーが名前を連ねていた。日本での展開を検討する翻訳関連企業にとっても、東京で開催された LocWorld は良い機会となったはずだ。

今回の東京開催は小さめである、前に参加したときはトラックが5つあったなどという声も LocWorld 経験者から聞こえた。しかしながら、LocWorld というイベントを通して、翻訳に関わる人たちの豊かな国際性と、それに伴うスケールの大きさが感じられた。さらにスケールアップしたイベントとして、また日本で開催される日を楽しみにしたい。

情報処理学会
第 122 回 情報基礎とアクセス技術研究会
第 101 回 ドキュメントコミュニケーション研究会
合同研究会

目次 由美子

機械翻訳課題調査委員会

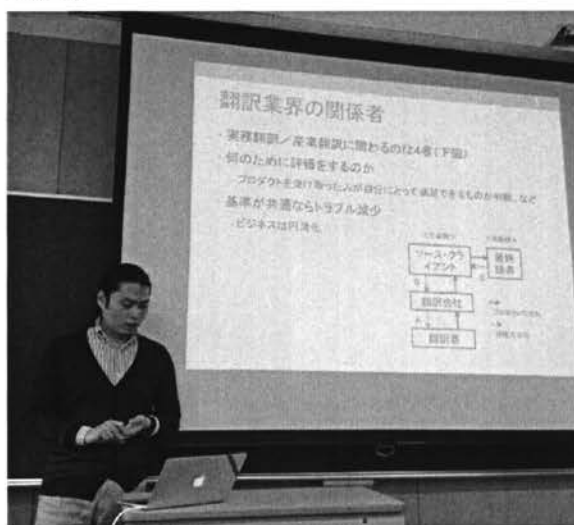
この度、情報処理学会の研究会である情報基礎とアクセス技術研究会とドキュメントコミュニケーション研究会により、『「多言語ドキュメントと情報技術」および一般』というテーマが掲げられた合同研究会が開催された。当日のプログラムは大きく 3 つ (1. 検索・推薦、2. 翻訳・多言語、3. 内容分析) に区分されていた。この合同研究会の「2. 翻訳・多言語」において、自ら講演した内容と、他の発表者による講演内容をここに報告する。

当該の発表ではドキュメントコミュニケーション研究会より、4 つの講演があった。まず、異文化コミュニケーションである大野邦夫氏より「グローバル企業の文書管理と企業文化に関する検討」と題された発表があった。この内容は、2015 年 5 月末にスペインで開催された SIETAR Europa Congress での Richard D. Lewis 氏によるワークショップ “Story Telling as a Training Tool & Research Method” に基づくものとのことだった。Lewis 氏の書籍では、6大陸の 85 の国民文化が分析され、「リニアアクティブ文化」(例: 任務指向が伺える)、「マルチアクティブ文化」(例: 感情的、多弁で衝動的)、「リアクティブ文化」(例: 集団の調和を重んじて目標に邁進する) に分類して比較検討されているようだ。

文書管理や企業文化、また英語教育についてこれら 3 つの文化の特性と照らし合わせながら、傾向や将来的な観測が述べられていた。

続いて、IT 翻訳者であり、ソフトウェア開発にも従事している西野竜太郎氏より、「翻訳プロダクトの品質評価に向けた翻訳業界の取り組み」について発表があった。「翻訳品質」の定義における曖昧さや、評価者によ

る基準の相違、過去の評価方法および問題点などが紹介された。



さらに、昨今翻訳業界で提案され、共通化が進行している MQM (Multidimensional Quality Metrics) や DQF (Dynamic Quality Framework) といった新たな評価方法についても紹介があった。

3 番目となった自らの発表は「産業翻訳実務にフリーランス翻訳者が活用するツール類」と題した内容であり、ツールベンダーにより製作・販売されているソフトウェアではなく、翻訳に携わる個人により開発・配布されており、幅広くフリーランス翻訳者に重宝されているツールを中心に紹介した。ベテラン翻訳者である井口耕二氏による「SimplyTerms」、コーディネータとして活躍している齊藤貴昭氏の「Wildlight」、そして Microsoft 社より Word 部門の MVP を 5 年連続受賞している新田順也氏による「ばらばら」をそれぞれ実演した。これらのツール類はいずれも、翻訳という業務に従事してきた個人それぞれが自らの作業の利便性を考慮して開発が開始さ

れ、営利目的ではなく他の翻訳者へも共有されている。実際にこれらのツールを利用している翻訳者も多く見受けられる。また、これらのツールの機能性には共通項も見られ、産業翻訳実務に携わっている翻訳者が必要としている機能を伺い知ることもできる。

参加者にも興味をもってもらえたようで、こういったツールに関する情報をどこで得られるのかといった質問が挙げられ、自らの文書作成業務に活用性を見出す参加者もあった。

最後の発表は AAMT を代表して参加された山本ゆうじ氏による「体系的翻訳・文書品質管理での高品質言語資産としての UTX 用語集形式」であった。体系的翻訳が日本では浸透していないという指摘に始まり、UTX は用語集データそのものではなく、良い用語集を作るための形式であることが紹介された。さらに、UTX は ISO (TBX) との連携も考慮されており、コンバータが提供されていることについても案内があった。

用語集の課題としては、そもそも組織内で用語集を作成する取り組み自体がなされていないこと、用語管理者の不明瞭さ、また、作成された用語集が使用されないことなどが指摘された。そして、UTX を使用することで解決できる課題もあることが結論として述べられ、スタイルガイドや翻訳支援ツールなどとの関連性についても今後の展望として紹介されていた。

尚、学会の関係者からは以下のコメントをいただいた。情報処理学会ドキュメントコミュニケーション研究会では、今後一層「翻訳・多言語」の分野に力を入れて、活躍されている多くの翻訳家の方々からのご参加を歓迎いたします。関連する分野は、インターナショナル化、ローカリゼーション、翻訳支援技術、翻訳品質、ポストエディット、用語集・スタイルガイドなど多様なご発表をお待ちしております。

URL: <http://sigdd.sakura.ne.jp/>

機械翻訳事始め[VII]

坂本義行

アジア太平洋機械翻訳協会理事

前回に続き、今回は、「日英科学技術文献の速報システムに関する研究」について、文献¹⁾と文献²⁾をもとに、その続きを報告します。

1. システム設計の考え方

かなり複雑な文にたいしても対応できるような機械翻訳システムを非常に短期間に4つの研究機関の協力によって作らねばならないということは非常に難問であった。現状の技術の統合によってシステムを実現するということが目標で、それ以上の研究的要素はなるべく持ち込まないということが前提とされたが、小さなモデルで少数の文を翻訳する研究のレベルとはまったく異なったシステム開発であるから、未知の部分はかなりあり、それらに対しても対処できるシステムでなければならないなどのことからシステム設計の考え方として次のような方針をとった。

- (1) 未知語や未知の構造の存在によって処理を途中で放棄することなく、不完全であっても、最後まで処理を行って翻訳結果を出すシステムであること、これは翻訳文の後修正をする場合の大切な考え方である。
- (2) 研究開発の発展途上において明らかになってくるであろう未知の言語現象にも十分対処できるように、機械翻訳用のソフトウェア・システムは強力な表現機能をもつように設計する。ソフトウェアは解析・変換・生成のすべての段階で用いることができ、種々の異なった性質の言語が扱えるように処理過程の制御がきめこまかくできるシステム作る。
- (3) 解析・変換・生成、それぞれが膨大な規則になるた

めに、多人数で平行に開発ができる必要がある。したがって、部分文法という考え方を導入し、部分文法ごとに独立に開発できるようにする。こうすることによって、翻訳の誤り原因の発見なども早くなることによって、翻訳の誤り原因の発見なども早くなり、システム全体の管理がしやすくなる。

- (4) 将来の多言語間機械翻訳の可能性を考えて、解析と変換。生成の3つの段階を明確に分ける。それらの段階をつなぐインターフェイスの構造をできるだけ明らかにする。
- (5) 言語は少数の基本的な規則によって説明されるものではなく、膨大な例外によって成り立っているものである。したがって、できるだけ広い言語現象に対して適用できる規則を探すことは当然であるが、各単語に固有の言語現象がうまく扱えるようなメカニズムを作ること考えた。
- (6) 辞書情報としては、現在の機械翻訳に直接必要なものでなく、将来必要となるであろう情報をも入れた。できるだけ中立的で理想的な形のものを作ることが必要である。一方、大量の単語に対して正確な情報を与えるためには、複雑な辞書構造をとることはよくない。さらに単語固有の言語現象を十分に記述できることも必要である。このような矛盾する要求に対処する方策を考えた。
- (7) 言語の解析の基本には格文法の考え方をとる。これは日本語の解析に特に適しているだけでなく、英語その他のヨーロッパ諸言語の解析にも広く用いられつつある方法である。格文法をとるために各要素の設定と、各格要素にどのような名詞がはいりうるか記述するための意味素性の集合

を決定しなければならない。これは日本語のほか
に英語、その他のヨーロッパ諸言語の特性を知っ
て作る必要がある。

2. ソフトウェアシステム

前章で述べた種々の言語記述に関する要請から、文
法記述のためのソフトウェアシステムとしては、木構造
を木構造に変換する規則を採用することにした。木構
造の各ノードは任意の数の属性・属性値対をもつこと
ができるようになっている。この部分に性・数・格の
一致、意味の整合性などを記述する。このように複雑
な木構造を取り扱い、しかも文法規則、辞書構造とと
もに定形的なものにはできないということから、基本
となるプログラム言語として LISP を採用すること
にした。LISP の採用については、PL/1、PASCAL など
の他の言語との比較検討をしたが、今回の研究におい
ては翻訳速度よりは言語現象を十分に記述できること
に重点を置いたこと、さらに短期間にこのソフトウェア
システムを開発しなければならないことなどの要因か
ら LISP の採用を決定した。

われわれの開発した GRADE (GrammarDescriber) と名
づけた文法記述言語の書き換え規則の部分は次のよう
な構造を持っている。文法を表現する書き換え規則と
しては、「左辺の木→右辺の木」という形に書かれる。
この規則の左辺の木は、翻訳のあらゆる中間段階にあ
る翻訳対象文の構造（この節ではこれを単にデータと
呼ぶ。これも一般には木構造をしている）との間でマ
ッチングがとれるかどうか調べられ、データのある部
分でマッチングがとれたら、その部分が規則の右辺の
木構造に書き換えられる。

木構造のマッチングにおいて次のような機能が存在
する。(1) 属性・属性値対がついていれば、それもマ
ッチングの対象となる。(2) 変数を木構造ノードに書
くことができる。(3) 木の枝の左右の順序が任意であ
ることを指定できる。(4) 規則の左辺の木のマッチン
グに関して複雑な条件が必要な場合には、それを LISP

プログラムで書くことができる。(5) この規則の部分
で単語辞書内に書かれている単語固有の規則を呼び出
し動作させることができる。

文法規則の数は、たとえば日本語の解析文法を考え
ても数 100 から 1000 近くになるから、解析のあらゆる
中間段階でこれらすべての規則の適用可能性を調べる
というは大変な手間になる。また、名詞句を解析する
ための規則、文の構造を決定するための規則などは自
ずから適用される時点が違っていて、言語学的に自然
な順序で解析を進めるがよい。したがって、文法規則
すべてを一様に扱うのではなく、異なった文法現象に
関する規則をそれぞれ別々に集めて、これらを部分文
法と名付けることにした。単純名詞句を解析する部分
文法、並列名詞句を解析する部分文法、埋め込み文構
造を決定する部分文法などである。そして、1つの部
分文法を1つのノードに対応させ、ノードを有向グラ
フ構造につないで部分文法の適用順序を規定すること
にした。これを部分文法ネットワークと呼んでいる。
上記の名詞句、並列句、埋め込み文などの処理部分は、
実際はすべてこの部分文法ネットワークとなっている。
文法全体は部分文法の組み合わせから構成される。

そこで、GRADE にはいくつかの書き換え規則をまと
めて1つの部分文法をとる機能、その部分文法の実行
後、次に適用すべき部分文法を条件つき指定する機能
をもっている。さらに1つの部分文法の中の書き換え
規則の適用の仕方を指定することもできる。すなわち、
書き換え規則をデータの左から適用していくか、右か
ら適用していくかの指定、さらにマッチングが成功し
て書き換えが行われたあと、さらに同一規則の適用の
可能性を調べてみるか、あるいは成功した規則を用い
ず、次の規則を試みるかという違いを指定できる。書
き換え規則の表現の仕方と規則適用のモードとの間
には関係があつて、うまくこれらを調整すれば能率のよ
い解析ができるのである。

文の解析を行うと通常非常に多くのあいまいさが生
じる。すなわち、1つのデータに異なったいくつかの
構造を対応させることになる。われわれの場合には、

できる限り広い範囲を一度にみて意味のあいまいさの生じることを防いでいるが、それでもあいまいな構造は生じる。文法規則を書くときに、ある構造と別のある構造が同じ部分に生じた場合に、それらは同じ句として解釈されることを積極的に表現することができるように PARA ノードと称する特別なノードを作り、システム側がその解釈を自動的に行うよう GRADE を作っている。

GRADE の制御部は 2 つの部分からなる。1 つは人間にわかりやすい形で書かれた上記の書き換え規則をパターンマッチングしやすい内部表現に変換する変換部であり、もう 1 つは内部表現に変換された書き換え規則をデータに適用して処理を進める実行部である。これらはすべて LISP で書かれていて種々の支援プログラムも入れて 1 万行をはるかに超える複雑なプログラムとなっている。

GRADE には、この他に辞書システムとの結合をする部分がある。辞書データは膨大であるので現在のところ大型計算機の標準ランダムアクセス・ファイルシステムを利用して記憶されている。したがって、必要な辞書データはファイルから読み出して LISP 形式で主記憶の上に展開し GRADE 文法の使用に供するようになっている。

GRADE に付属した支援プログラムには種々のものがある。辞書データの誤りをできるだけ少なくするために、パンチされた辞書データの誤りをデータの初期入力のとくにできるだけ発見するプログラムが作られており、また、計算機内の辞書データの変更、追加などのためのプログラムなどが代表的なものである。4 つの研究機関が文法、辞書データ作成を分担しているため、計算機間のコード変換をはじめとして、お互いのインターフェイスをうまくまとめるためのプログラムがいろいろ必要で、こういったなんでもないところに予想以上の手間がかかった。GRADE の詳細は他にゆずる。

3. 辞書情報とそのシステム化

機械翻訳において辞書が大切なものであることはますます認識されてきつつある。われわれの場合には、短期間の間に大量の辞書データを整備蓄積しなければならなかったために、いくつかの工夫を必要とした。それは、辞書データをできるだけ一目でわかる形の形式で表し、むだな情報の記入をせず、最小限度の記入ですむようにすること、計算機のことを知らない一般の人でも記入ができることであり、さらに多くの人に平行に作業をしてもらうときにできるだけ質をそろえるために、辞書情報の記入のための作業手続きを詳しく書いたガイドラインを作成したことにある。

辞書は使用の上から入力文解析のための辞書、言語間変換のための辞書、出力文生成のための辞書、および形態素解析のための辞書、形態素生成のための辞書の 5 つに分けられる。これらの辞書のうち、言語間変換の辞書の部分を除いて、他の 1 つの中立的な形の辞書として作ることができ、そこから使用目的に応じた辞書を生成して使用する。したがって、単一言語の辞書としては、この中立的な辞書を作ることになり、辞書管理、すなわち辞書データの追加、変更、削除などはすべてこの中立的辞書において行うことにしている。

辞書は使用上から入力文解析のための辞書、言語間変換のための辞書、出力文生成のための辞書、および形態素解析のための辞書、形態素生成のための辞書の 5 つに分けられる。これらの辞書のうち、言語間変換の辞書の部分を除いて、他は 1 つの中立的な形の辞書として作ることができ、そこから使用目的に応じた辞書を生成して使用する。したがって、単一言語の辞書としては、この中立的な辞書を作ることになり、辞書管理、すなわち辞書データの追加、変更、削除などはすべてこの中立的辞書において行うことにしている。

辞書データは現在約 55,000 語計算機に入れてあるが、その内容は動詞約 3,000 語、形容詞、形容動詞約 1,000 語、普通名詞および電気工学を主とする科学技術用語約 5 万語、連体詞、副詞その他の付属語約 800

といったところである。動詞、名詞の主なもの翻訳を対象した JICST の抄録の最初の 3,000 文に現れるものを取り、それに JICST で別途抽出した 11 万語の用語辞書から選んで作成した。

辞書に記入すべき情報はどのような文法概念を用いてどのような解析を行うかということに依存することはやむをえないが、できるだけ中立的、一般的に作ることを心がけた。文法としては格文法をとることにし、格の種類として表 1 に示すものを採用した。ある動詞の種々の格スロットにどのような名詞が入りうるかを規定する方法としては、意味素性としては図 1-1、図 1-2 に示すものを採った。これらの意味素性のいくつかのものの AND, OR 結合によって、動詞の格スロットに入りうる名詞を規定する。

辞書情報は記入フォーマットをきめ、記入のための作業基準書を作って多人数での作業が可能なよう工夫をした。記入フォーマットは名詞、動詞、形容詞、副詞その他の品詞について作られている。辞書は日本語（解析、生成に供用）辞書、日英変換のための辞書、英語（解析、生成）辞書などが用意されているが、その詳細は他にゆずる。

4. 機械翻訳システム・マンマシンインターフェースの作成

現在の機械翻訳システムは、今回のプロジェクトで作成したシステムを含めて、何らかの形式で人間の援助を必要とする。また、辞書等の翻訳に必要な言語データも、対象分野に応じて適宜用意する必要がある、この点でも柔軟なマン・マシンインターフェースを準備する必要がある。このようなことから、今回のプロジェクトでは、機械翻訳システムを中核に持つ総合システムとして、次のようなシステムを開発した。これを図 2 に示す。

- ①テキストの準備のためのエディタ（前編集エディタ）
- ②入力テキストと出力テキストを相互に参照しながら、

出力テキストを編集するためのエディタ（後編集エディタ）

- ③汎用辞書とユーザ用の個人辞書とを分離して管理する階層的な辞書管理システム
- ④辞書フォーマットの記述を直接編集する画面エディタ

特に、④の辞書フォーマット・エディタは、辞書作成者が辞書データの内部的な管理形式とは独立に、辞書の記入フォーマット用紙のイメージを直接走査できるようにしたものであり、一般用辞書・ユーザの個人辞書の双方を編集することができる。

また、この総合システムは、京大・ETL・JICST の各研究組織で開発されたソフトウェア、及び言語データを相互に結合し、形態素解析・構文解析・変換・構文生成・形態素生成を、バッチ運用・TSS 運用の各モードで実行できるようにしたものであり、翻訳システム開発時の各種テストも可能なように工夫されている。

5. 翻訳結果の評価

翻訳結果の訳文の質の評価法は、できるだけ主観的な判断を避け、評価者個人による評価のバラツキを少なくする必要がある。今回のプロジェクトでは、

- ①理解容易性：訳文だけを読んだときの理解の容易さ
- ②忠実度：訳文と原文との意味のずれ

という 2 つの尺度を設け、各尺度ともに訳文の修正作業と関連づけること、評価法を詳細に規定したマニュアルを用意することにより、評価値が個人の主観に依存する度合を極力低くすることとした。この評価法は、本プロジェクトの翻訳結果の評価に用いただけでなく、いくつかの他の翻訳システムにおいても採用されるようになっている。

理解容易性の基準を表 2 に、忠実度の基準を表 3 にしめた。

日英翻訳システムの性能評価は、翻訳実験の進行とともに2回に分けておこなわれ、それぞれ3,000文と10,000文の翻訳結果が評価された。10,000文の翻訳結果は(表4に示す)、3,000文の評価に比べ、未知語(未登録語)が頻出するなどの理由から多少悪くなっているが、いずれの場合も、理解容易性、忠実度ともに80%前後が翻訳結果として許容範囲に入っていると判断され、両尺度ともに許容範囲となったのは60%を超えている。

この結果は、原文の前編集を全く行なわなかった場合であり、論文抄録文の原文自体が長く、複雑で理解困難なことを考えると、前編集、後編集を前提とする現在の機械翻訳システムの性能としては十分であると考えられる。

英日翻訳システムの性能評価も、日英の性能評価法とほぼ同じ方式で行なわれ、翻訳結果3,000文の評価結果(表5に示す)もほぼ同じであった。

なお、日英翻訳システムの対象テキストは、JICST発行の論文抄録誌から、また、英日翻訳システムのテキストは、INSPEC抄録誌から、いずれも、無作為に採られ、原文の前編集は全く行なっていない。

言語間についての機械翻訳用辞書のあるべき姿を言語理論的、計算言語理論的に研究する必要がある。

6. 研究の波及効果及び問題点

6.1 研究の位置付け及び波及効果

本研究が開始されるまで、機械翻訳の研究は小規模で限定されたモデルがいくつかの大学や研究所で行なわれていた程度であった。

外国においては、少数の商用システムがあったが、いずれも1960年代の考え方の基づいた古い形式のもので十分な能力をもつとは言いがたいものであった。研究としてはカナダで英仏機械翻訳システムの研究が行なわれていたが行き詰まりを見せていた。一方ヨーロッパ共同体ではEUROTRA計画という多言語間機械翻訳の研究がようやく開始されようとしていた。

いずれにしても日本語とヨーロッパ諸言語、特に英語との間の機械翻訳は全く行なわれていなかった。一方欧米からは日本の目覚ましい科学技術の進展の基礎をなす科学技術情報の入手が盛んに要求されるようになり、日英機械翻訳システムの開発が国として必要であるとされた。

このような日本語と英語との間の機械翻訳システムの開発の必要性は特に電子計算機関係の企業で必要と感じられて来ていたが、従来の技術開発とは異なり、言葉という複雑なものをどのように処理していったらよいかについて明確な方針がたたないという状況にあった。

このような状況の中で、本研究が果たした役割りは次のように要約される。

- (1) 科学技術文献の抄録の日英、英日機械翻訳によって文献内容の情報の伝達の可能性を明らかにした。
- (2) この機械翻訳システムを科学技術文献抄録のデータベースに接続することによって、翻訳機能を持ったデータベースサービスの可能性を示した。
- (3) 日本が国として文献速報のための機械翻訳システムの研究開発を開始し具体的成果をあげつつあるということは、ヨーロッパ共同体、カナダ政府等における機械翻訳システム開発に大きな刺激をあたえた。
- (4) 本研究は世界各国の計算言語業界に影響を与え、最近各国でこの方向の研究が活発におこなわれるようになって来た。これは計算言語学国際会議論文集、いくつかの学術雑誌にも現れてきている。米国カーネギーメロン大学にも機械翻訳研究所が最近作られた。

[参考]

JICSTでは、科学技術振興調整費による研究の成果を活用し、実用機械翻訳システムを開発する計画である。

これは、JICSTのオンライン情報システム(JOIS)で、

1986 年 10 月より国内外にサービスする予定の JICST 英文データベース(日本文献の抄録、キーワード等を英文化したデータベース)作成を支援するための日英翻訳システムと、同じく JOIS でサービスしている外国のデータベースの和訳のための英日翻訳システムより成る。

開発期間は、日英翻訳システムが 1986 年度より 3 年間、英日翻訳システムが 1988 年度より 2 年間の予定である。開発完了時の辞書の語数は日英、英日翻訳システムそれぞれ 30 万語の見込みである。

なお、JICST 英文データベースは現在手作業で英訳し、かつ当面タイトルのみの英訳が主であるので、少しでも早く機械翻訳システムを開発し、抄録文まで含めた英文データベースを作成することが望まれる。

6.2 社会、経済に対する貢献の状況及び可能性

本研究は文献抄録のもつ情報の機械翻訳を通じての速報システムの可能性を示すためのシステムを作り、その可能性を実証することにあつたため、現在の段階では学問世界、企業等における製品化等に対して貢献はしたが、社会一般、経済界に対して具体的に貢献するところまで行っていない。

しかし、この研究を通じて自然言語処理技術がこれからの情報社会の中核的技術であり、その中でも機械翻訳システムは日本語ワードプロセッサの次に来るべき大型商品であり、この技術を用いた人間と機械との対話システムの作成もすみやかに開発すべきものであることが広く認識されて来ており、そのうちに「言語産業」と称すべき産業分野が形成されてゆくものと期待されるようになって来た。

6.3 今後の問題点

- (1) 言語は膨大な例外の集積であるとも言えるべきものであるから、文法、辞書の不断の改良、精密化が不可欠であり、またこれを行なうことによって、

質の改善された翻訳結果を得ることが期待されている。

- (2) 本研究で開発したシステムはあくまでも機械翻訳の実用への可能性を示すためのプロトタイプシステムであつたから、実用システムとして用いるためにはそれなりの改良を行い、システムの作り直しをする必要がある。
- (3) 本研究で開発したシステムは科学技術論文の抄録を対象としているため、そこにはほとんど現れない疑問文、命令文、感嘆文などの翻訳のための文法、さらに一般の文章の翻訳のための辞書データの蓄積など、使用目的に応じた改良・拡充を行う必要がある。
- (4) 上記(3)と同様な観点から、現在対象としている電気工学関係のテキストから、さらにその他の科学技術分野の文章が翻訳できるよう、専門用語辞書の充実を行なう必要がある。
- (5) 日英・英日翻訳システムにおける解析文法、生成文法、変換文法、変換辞書はそれぞれ独立であるが、これが両システムで共通に使えるかどうか、研究を行なう必要がある。
- (6) 代名詞の指すものの推定、省略句の推定、さらに文脈処理、文章理解の方法など機械翻訳のための基礎研究を今後強力に進める必要がある。単一言語、多言語間についての機械翻訳用辞書のあるべき姿を言語理論的、計算言語理論的に研究する必要がある。

【参考文献】

- 1) 長尾真、辻井潤一、中村順一、坂本義行、鳥海剛、佐藤雅之、「科学技術庁機械翻訳プロジェクトの概要」、「情報処理」、Vol. 26, No. 10, 1985.
- 2) 科学技術庁科学技術振興局、昭和 57 年度～60 年度 科学技術振興調整費「日英科学技術文献速報システムに関する研究」。

表1 日本語格ラベル一覧表

(1) 主体	SUBject	－が
(2) 対象	OBJect	－を
(3) 受け手	RECIpient	－に与える
(4) 与え手	ORIgn	－から受ける, 奪う
(5) 相手 1	PARtner	－と協議する、異なる、－に関連する
(6) 相手 2	OPPonent	－から保護する, 独立する
(7) 時	TIMe	1980 年に
(8) 時・始点	Time-From	5 月から
(9) 時・終点	Time-TO	来年まで
(10) 時間	DURation	5 分間加熱する
(11) 場所	SPAcce	－に位置する、－で発生する
(12) 場所・始点	Space-From	－から帰る
(13) 場所・終点	Space-TO	－へ送る、－に到達する
(14) 場所・経過	Space-Through	－を通る、上空へ飛ぶ
(15) 始状態	SOUrce	5.5%から6%へ引き上げる
(16) 終状態	GOAl	英語から日本語に翻訳する
(17) 属性	ATTribute	適応性に富む、欠ける、乏しい
(18) 原因・理由	CAUse	事故で死ぬ、－から分かる
(19) 手段・道具	TOOl	イオン法で、ドリルで
(20) 材料	MATerial	ペーストで作る
(21) 構成要素	COMponent	－から成る、－で構成する
(22) 方式	MANner	並列に、10m/sec で
(23) 条件	CONdition	焦点深度で決まる
(24) 目的	PURpose	－に適する、備える、必要な
(25) 役割	ROLe	議長に選ぶ、－として用いる
(26) 内容規定	ConTent	－と呼ぶ、述べる、みなす
(27) 範囲既定	RANge	－について、－に関して
(28) 提題	TOPic	－より大きい、－に劣る、－を上回る
(29) 観点	VIEWpoint	立場から、－の点で
(30) 比較	COmpaRison	－より大きい、－に劣る、－を上回る
(31) 随伴	ACCompaniment	－とともに、－に伴って
(32) 度合	DEGree	5%増加する、3 キロやせる
(33) 陳述	PREdicative	－である

注) 英語名中、大文字の部分(3 字)を略称とする。

表2 理解容易性の基準

1	文意は明瞭で疑う余地はない。文法、用語用法、文体ともに適切で、なんら修正の必要はない
2	文意は明瞭で理解できるが、文法、用語用法、文体上に多少の問題が見られ、1には至らない。ただし、こうした欠陥を改善するにあたっては出力文だけ見て外人評価者単独で修正を施すことができ、日本人評価者の援助を必要としない。
3	全体的文意は把握できるが、文法、用語用法上の問題のために細部理解に自信が持てない。こうした問題を取り除き、さらに文体面の改善を施そうとすると、修正方法がひとつに定まらず、入力文の意味するところを日本人評価者に確認しなければならない。
4	文法、用語用法上の問題が多いので、相当な思考の後わずかに文意を仮定できるか、もしくはほとんど把握できない。こうした問題を解決して文体を改良するよう手を加えるより、新たに翻訳をやり直した方が時間も早く、質も向上すると思われる。
5	まったく理解できない。どんなに熟考、検討しても文意をとることは不可能である。

表3 忠実度性の基準

0	入力文の構造は出力文に忠実に再現されていて、外人がみても明瞭で、修正する必要がある。
1	入力文の構造は出力文に忠実に再現されていて、また外人にも明瞭に理解できるが、多少の修正は必要である。ただし、その修正は与えられた修正文だけにもとづき、外人が単独で行える。
2	入力文の構造は出力文に忠実に再現されているが、語の置き換えを必要とする。
3	入力文の構造は出力文にほぼ忠実に再現されているが、句と句の関係、時制、数、態、副詞の位置などに誤りがある。また、単語が重複して出力されているなどの誤りがある。
4	入力文の構造は出力文にあまりよく再現されていない。句の要素の欠落、節と節、節と句、あるいは節内の関係に誤りがある。
5	入力文の構造は出力文に良く再現されていない。節や句の欠落があるが、一応文として成立している。
6	入力文の構造はまったく出力文に再現されていない。主語や述部の欠落があるため、文として成立していない。ただし、表題のような名詞句にあつては、その主名詞（日本語で最後に位置する名詞）、またはこれを直接修飾する節・句の動詞的要素が欠落している。

表 4 日英翻訳文の評価結果 (10,000 文)

理解 容易性	忠実度									
	0	1	2	3	4	5	6	不備	合計	割合
1	2058	79	618	376	303	144	63	8	3649	32.1
2	52	863	667	503	336	200	99	5	2725	24.0
3	36	111	688	839	1012	346	407	22	3461	30.4
4	2	3	94	181	394	140	248	9	1071	9.7
5	1	0	34	38	118	49	227	2	469	4.1
合計	2149	1056	2101	1937	2163	879	1044	46	11375	
割合	18.9	9.3	18.5	17.0	19.0	7.7	9.2	0.4		

表 5 英日翻訳文の評価結果 (3,000 文)

理解 容易性	忠実度									
	0	1	2	3	4	5	6	不備	合計	割合
1	592	59	85	47	54	8	12	1	860	29.7
2	10	285	53	54	52	4	8	0	466	16.1
3	6	4	437	106	144	11	47	1	756	26.1
4	1	0	144	59	121	20	119	2	466	16.1
5	1	0	38	17	44	4	236	4	346	11.9
合計	612	348	757	183	415	47	424	8	2894	
割合	21.1	12.0	26.1	6.3	14.3	1.6	14.6	0.2		

図 1-1 名詞意味マーカ体系表 (1)

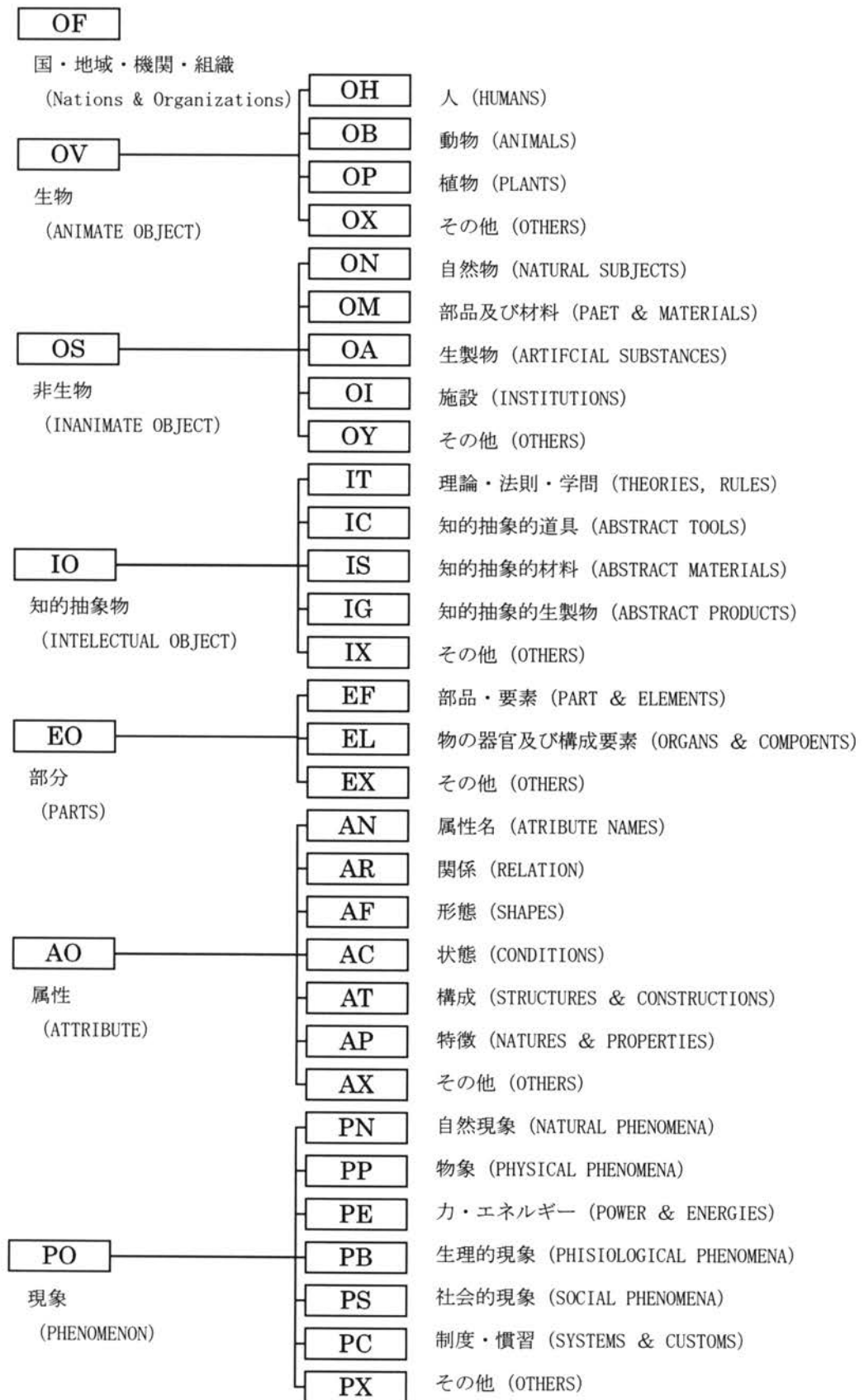


図 1-2 名詞意味マーカ体系表 (2)



ZZ
全体のその他 (OTHERS)
この他、特定の語につける意味スロットとして以下のものがある

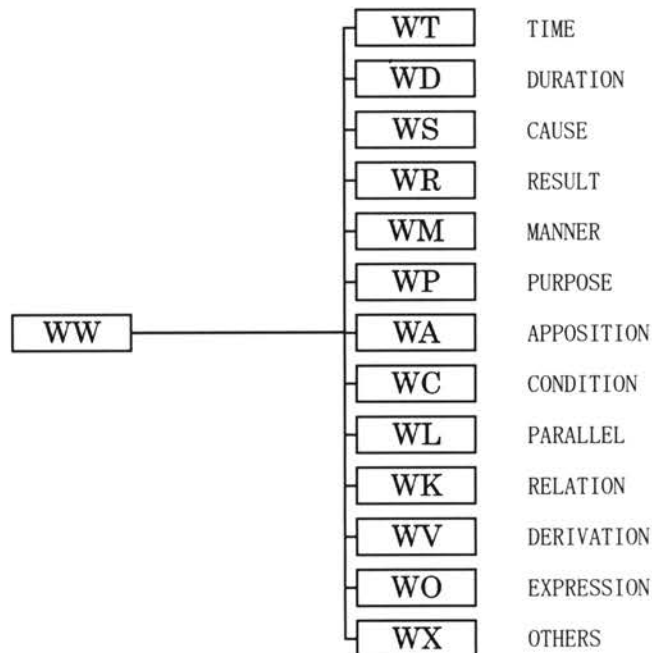


図3 辞書システムの概念図

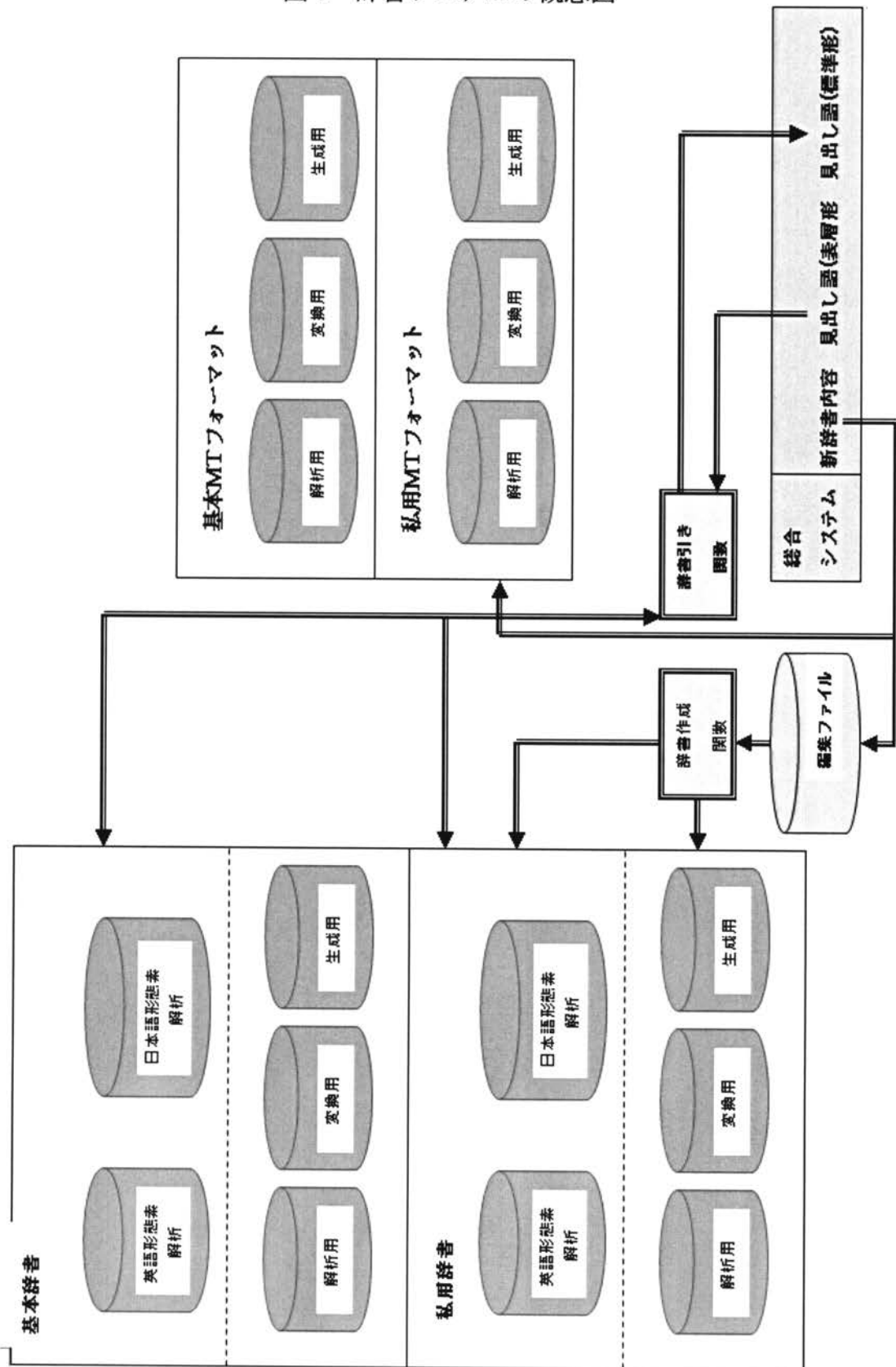
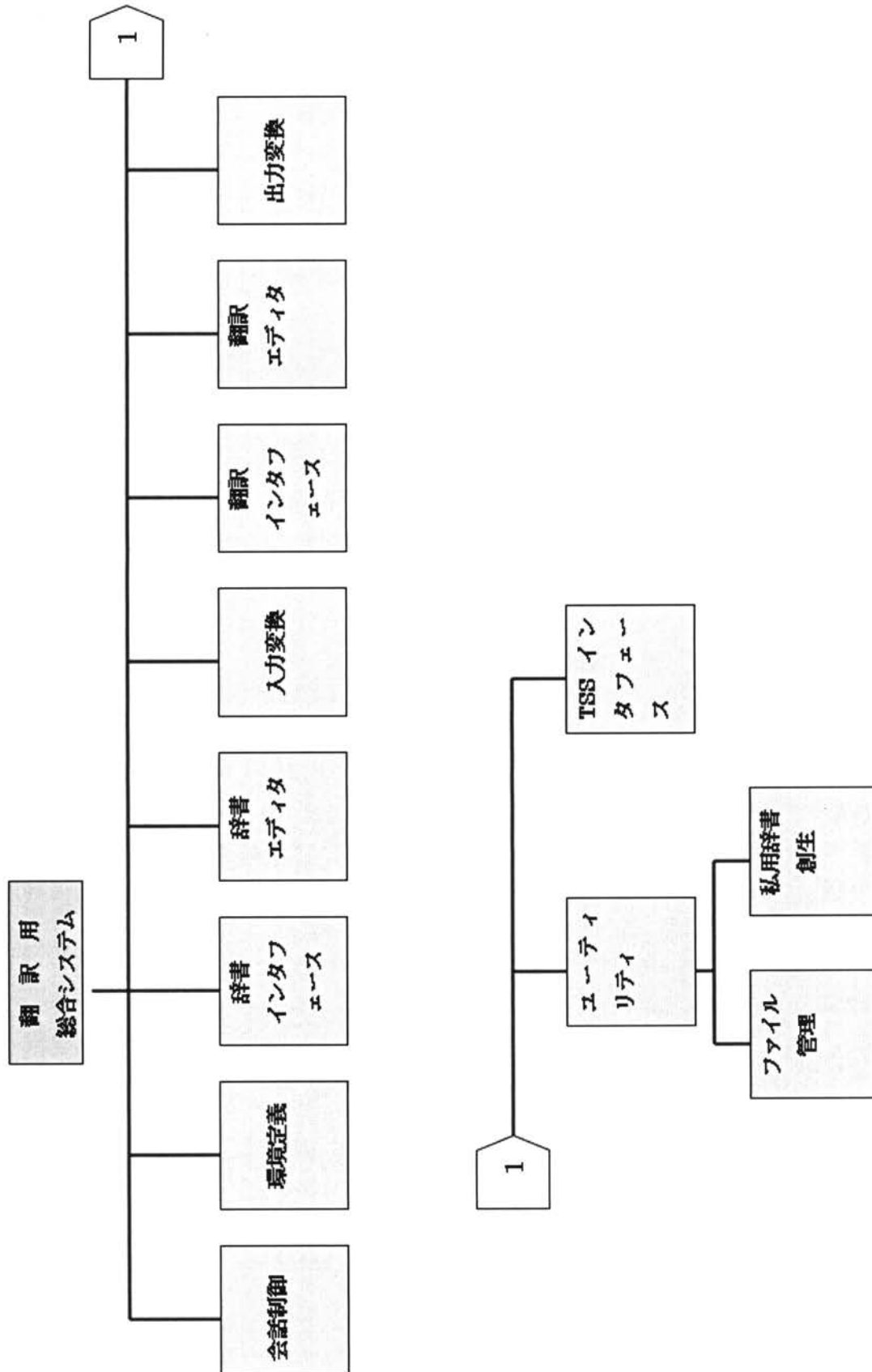


図 4 翻訳総合システム構成図



AAMT設立までのこと

長尾 真

AAMT 初代会長、国際高等研究所所長

1. MT研究開発の始まり

コンピュータが暗号解読のために第2次世界大戦の終盤に作られたが、これが言語の翻訳にも使えそうということから、1950年代中ごろに機械翻訳(MT)の研究が米国と英国を中心に始まった。1957年10月にソ連が人工衛星を世界で初めて打ち上げたが、これは米国にとっては大きなショックであった。ソ連の科学技術が非常に進んでいるとすれば、彼らの科学技術論文を英語に翻訳して知る必要があるということになったが、ロシア語の文献を翻訳できる翻訳者の数は圧倒的に足りない。そこでMTに頼らざるを得ないということになり、米国政府は1959年から全国の主要大学・研究所に研究費を投入し、MT研究を本格化した。

電気試験所(電子技術総合研究所を経て、現在の産業技術総合研究所)の和田弘氏は1950年代中ごろMITに留学していて、米国のMT研究の動きを知り、帰国後1956年ころから日英MTの研究を始めた。1958年ころに一応の完成を見、簡単な短文を翻訳した。これは1959年にパリで開催された第1回情報処理国際会議(ICIP: International Conference on Information Processing)で展示・デモされ好評を博した。この詳細は坂本義行氏によってAAMTジャーナルに報告されている。

2. 日米セミナー

和田弘氏は日米のMT研究者の交流のために日米セミナーを1964年東京で開催した。米国の主要なMT研究者が来日したが、彼らはこの機会に東京、京都、

九州のMT研究を視察した。第2回のMT日米セミナーは次の年にニューヨークで開催された。その後数回行われ、両国のMT、自然言語処理研究の発展に寄与した。

3. 計算言語学国際会議(Coling)

1965年に第2回のMT日米セミナーがニューヨークで開催されたが、その前後に日本からの研究者は米国の主要MT研究機関を視察した。またこの時に合わせて計算言語学の国際会議、およびIFIP World Computer Congress(IFIP: International Federation for Information Processing)がニューヨークで開かれ、MTは大きな話題となった。この計算言語学国際会議(第1回Coling: International Conference on Computational Linguistics)において、この会議を今後2年ごとに開催するために計算言語学国際委員会(ICCL: International Committee on Computational Linguistics)が結成され、日本からは和田弘氏が委員として参加した。筆者は第3回のColingからICCL委員になっている。

なお今年(2016年)の12月に大阪でColingが開かれる。日本では東京、京都に次いで、3回目の開催となる。

4. 電子協のMT調査研究会

日本電子工業振興協会(JEIDA。現在の電子情報技術産業協会(JEITA))は1958年に設立され、主として日本のコンピュータメーカーを会員として、日本のコンピュータ産業の発展に寄与してきたが、

MTが将来大切な技術になるという認識から、1970年代に入ってMTの調査研究会を作った。筆者がその委員長となったが、大学の研究者、主要コンピュータメーカーのソフトウェア部門の責任者などが構成員となり、欧米のMTの状況の調査と英日言語間のMTの研究開発に関する検討を中心に活動した。

5. Muシステム開発

1970年代後半からは日本の技術力による各種製品が世界市場、特に米国の市場を席捲する勢いになって、米国は危機感を募らせた。「日本は米国の科学技術論文を読み大きなメリットを得ているが、日本の科学技術論文はほとんどが日本語で書かれていて米国の研究者は日本の実情を把握できない。日本は言語障壁に守られ、情報の流れは片方向である。これを改善するために、日本は科学技術情報を英語で発信すべきである」という強い圧力が日本にかかってきた。

このような状況を改善するために科学技術庁は日本語の科学技術論文の抄録の部分だけでも英語に翻訳することを目的としたMTシステムの研究開発を行うことになった。これは1982年から4年間行われた。Muシステムがそれである。このMuシステムの開発には京都大学が中心となり、電子技術総合研究所、日本科学技術情報センター、さらに電子協のMT調査研究会のメンバー企業、そのほかに翻訳会社数社などが参加した。

6. MTサミットとIAMTの設立

このMuシステムの成果を世界に発信するとともに世界のMT研究の状況についての国際的な情報交換を目的として、MTサミット会議を1987年に箱根で開催した。その成功によって2年後にヨーロッパで第2回を開催することになった。ミュンヘンで開催した第2回のMTサミットにおいて、この会議を定期的で開催するための母体としてIAMT (International

Association for Machine Translation) を作ることに、その下にアジア・太平洋地区(AAMT: Asia-Pacific Association for Machine Translation)、ヨーロッパ地区(EAMT: European Association for Machine Translation)、南北アメリカ地区(AMTA: Association for Machine Translation in the Americas) という3つの組織を作り、それぞれに活動するとともに、2年ごとにこれらすべてが参加するMTサミットを3者持ち回りで開催すること、という組織の骨組みの議論をし、1991年ワシントンで開催した第3回MTサミットでこの組織を正式に発足させた。筆者はその初代のIAMT会長に就任した。

7. AAMT

この組織を作るために、日本ではまず日本機械翻訳協会を1991年に作ったが、IAMTの発足に従って、これをアジア・太平洋に広げ、AAMTに改組した(1992年、初代会長長尾真)。その構成員としては韓国、中国、タイ、その他の国の研究者、企業、翻訳者が対象であるが、残念ながら十分に広がっていないのが現状である。今年(2019年)は日本機械翻訳協会発足から25年になる。発足以来ジャーナルの発行、総会での諸行事、日本翻訳連盟との協賛会議などの活動を行っているが、これらは歴代会長(田中穂積氏、辻井潤一氏、井佐原均氏、中岩浩巳氏)と事務局(特に熱心に世話をしてくださった荻野孝野さん)の努力のお陰である。

8. MT研究開発の今後

MTシステムは曲がりなりにも使える質にまでなってきた、社会のいろんなところで利用され始めている。これからはさらなる質の向上と多言語への対応、種々の専門分野での利用のための専門用語辞書の充実、多言語の音声対話システムのための音声データベースの整備、さらに重要なことは翻訳専門家の持つノウハウ

のシステムへの取り込みと利用をうまくやることであろう。人工知能的手法でこれらのことをいかに自動的に
に行い、システムを賢くしてゆくかも大きな課題である。

9. バベルの塔の建設

30年先には世界の言語障壁を十分低くし、世界中の人たちがお互いに心を通じ合えるように努力をしたいものである。これまで人類が蓄積してきた知識によってバベルの塔を本当に築きあげる、そのために世界中の人たちが相互協力をするようにしましょう！！！！

長尾先生にはご寄稿に合わせて、25周年を祝う書（次ページ）もお願い致しました。心より感謝申し上げます。（編集委員）



言霊（ことだま）

万葉の時代から「大和の国は言霊の幸ふ国」といわれてきた。言葉には力が宿っていると考えられてきたが、それがどこまでコンピュータで扱えるだろうか。

「グローバル・コミュニケーション計画」の取組状況

総務省情報通信国際戦略局技術政策課研究推進室長荻原直彦氏インタビュー

聞き手 河野弘毅（日本翻訳ジャーナル編集長）

総務省が推進するグローバル・コミュニケーション計画の取組状況を監督官庁である総務省の荻原直彦氏にインタビューしましたので紹介します。¹インタビューを受けてくださった荻原直彦氏のプロフィールは以下のとおりです。

荻原直彦（おぎはらなおひこ）氏

総務省情報通信国際戦略局²技術政策課研究推進室長

平成4年郵政省（現総務省）入省。以来、情報通信行政に携わり、平成21年より情報流通行政局衛星・地域放送課地域放送推進室技術企画官、平成22年より総合通信基盤局電波部電波政策課電波利用料企画室長。平成25年より現職となり、多言語音声翻訳システムの研究開発と社会実装を目標とする「グローバルコミュニケーション計画」を推進している。

取材：2015年11月

記事執筆：目次由美子（LOGOSStar）

2020年に東京で開催されることとなったオリンピックとパラリンピックへ向けて、訪日観光客のみならず、日本国内の地域社会に居住する外国人の増加に伴い、グローバルなコミュニケーションの支援が高まっている。このような状況において、総務省は2014年4月に世界の「言葉の壁」をなくし、グローバルで自由な交流を実現することをミッションとした「グローバ

ルコミュニケーション計画」を発表した。機械翻訳の波が押し寄せる最中、この動きを牽引するともいえる計画の最新状況について、総務省の担当者に直接インタビューをする機会を得ることができた。

1. はじめに

荻原氏は音声翻訳の研究が始められたのは1980年第後半であり、世に知られるようになったのは2007年の後半であると紹介してくれた。2007年に総合科学技術会議（現総合科学技術・イノベーション会議）で音声翻訳のデモをしたことはニュースなどにも取り上げられた。スマートフォンが普及しておらず、パソコンで実演されたシステムは、使い勝手の観点から良い評価にはつながらなかった。

スマートフォンの普及、通信事業者による音声対話や翻訳アプリの実現なども経て、2020年オリンピックへ向けて「外国人が日本に来たときに困らないために」と音声翻訳技術の社会実装の取り組みが進められている。

2. グローバルコミュニケーション計画の概要

グローバルコミュニケーション計画には、「世界の『言葉の壁』をなくす」というミッションが掲げられている。

このミッションには、以下の三つのビジョンが伴う。(1) 多言語音声翻訳システムの利用による「グローバルで自由な交流の実現」、(2) 多言語音声翻訳シ

¹この記事は一般社団法人日本翻訳連盟発行『JTFジャーナル』2016年3/4月号との同報掲載です。

²総務省情報通信国際戦略局は、情報通信技術（Information & Communications Technology: ICT）を経済の成長・競争における重要な源ととらえ、グローバルな視点での総合的および戦略的な政策を展開する上で、研究開発、標準化活動、国際展開活動などを一体的に推進しています。

システムの社会実装および活用をもたらすパイオニアとしての「日本のプレゼンス向上」、(3) 世界から集まる選手や観客を対象とし、言葉の壁を感じさせることのない「東京オリンピック・パラリンピックでの『おもてなし』」。

この計画の核を成す「多言語音声翻訳システム」は、国立研究開発法人である情報通信研究機構（National Institute of Information and Communications Technology: NICT）で開発されている。たとえば、NICTは「VoiceTra」（ボイストラ）という名前のスマートフォン向けアプリを無償で提供している。多言語音声翻訳システムは、(1)音声認識、(2)多言語翻訳、(3)音声合成という三つの技術で構成されており、日本語を音声入力すると、直ちに外国語に翻訳し、音声出力される。

2015年10月にはグローバルコミュニケーション計画に基づき、NICTの従来の研究成果を反映させた最新版のVoiceTraが公開された。ユーザビリティが考慮されユーザーインターフェースが刷新されたばかりではなく、翻訳精度も著しい向上を遂げ、また、世界で初めてミャンマー語の音声認識・音声合成にも対応した。29言語に対応し、10言語における旅行会話は実用レベルに達しているようだ。

荻原：翻訳の精度を高めるためにはコーパスが必要とされています。要するにデータベースです。たとえば、「日本語の音声認識コーパス」については、聞こえ方の異なる「車」というデータをどれだけ多く集められるか、「対訳コーパス」については、日英の対訳データをどれだけ多く集められるか、これが翻訳精度の決定に関わってきます。

そして「グローバルコミュニケーション計画」が策定され、2020年オリンピックまでに社会実装するという目標の下、取り組みが進められている。

3. グローバルコミュニケーション計画の推進

多言語音声翻訳技術では、旅行分野の会話だけでなく、特に、医療やショッピング分野における会話の翻訳精度の向上に今後も取り組む予定であるが、対応言語数や雑音対策、長文翻訳などの課題も抱えている。病院、商業施設、観光地などでの社会実証を展開していく予定である。

河野：日常会話全般を最終目的にしていますか？

荻原：旅行や医療の会話を中心に取り組んでいる中で、一般的な日常会話も自然な展開として翻訳精度が上がっていくと考えています。

たとえば、成田空港ではターミナル内の巡回案内スタッフがタブレットを利用してフライト情報や施設情報などを案内している。これにはNICTの技術が活用された多言語音声翻訳アプリ「NariTra」が活用され、アジア地域からの来客に対応している。

2015年2月に開催された「東京マラソン2015」では、約3万6千人の参加ランナーのうち、約5千人は外国人ランナーだったとのこと。ランナーやその家族に対する給水・給食、手荷物管理などを行う約1万人のボランティアスタッフに「VoiceTra」を活用してもらったようだ。このような実験的利用において、周囲の雑音により音声認識が芳しくなかったなどの課題が示されたと同時に、外国人との会話における利便性も挙げられ、期待の高さも明らかとなった。東京都や関係機関とも連携し、さらに多言語音声翻訳システムの社会実装へ向けての取り組みが続けられていくこととなっている。

この他にも、「2015 東京国際ユース（U・14）サッカー大会」の選手交流会や、「2015 ジュニアスポーツアジア交流大会（バドミントン・卓球）」に参加する海外チームと都内学校との交流会でも実験的利用により、専門用語や年齢層に依存する話題の拡充、スマートフ

オンを操作するために作業を中断する必要性など、取り組むべき新たな課題も認識された。

2015 年 8 月、東京メトロでは訪日外国人との円滑なコミュニケーションを支援するツールとして導入され、岡山県警では外国人旅行者に対する道案内や遺失物に関する対応の支援として「VoiceTra」の試験利用が開始されたなど、音声技術翻訳の活用の幅広さが伺える。今後、フィードバックを受けてさらなる改善へとつなげる予定であるとのこと。

また、民間企業との共同研究も進んでおり、アナウンスにおける音声技術やロボットなどの先進技術との融合により、さらなる効果が期待されている。

荻原：グローバルコミュニケーション計画において大きく掲げてはいませんが、「防災分野」も重要であることを認識しています。

音声認識技術においては、人の声は認識されても、テレビの音声やスピーカーを介したアナウンスの声は認識されないことがある。アナウンスにおける技術の発展は、防災分野での技術展開が期待されている。

2015 年 4 月には首相官邸で開催された総合科学技術・イノベーション会議において進化した「VoiceTra」のデモを実施し、2020 年オリンピック・パラリンピック東京大会へ向けて端末の形状や翻訳精度のさらなる向上について言及した。

また、内閣府による「2020 年オリンピック・パラリンピック東京大会に向けた科学技術イノベーションの取組に関するタスクフォース」と連携した総務省の研究開発も進められている。このタスクフォースでは、「海外からの来訪者などに多様なサービスを提供するための意志・情報伝達サポートの実現」や、『『サイバーフィジカルシステム』による安全・安心の実現および快適な『おもてなし』の実現』など、9 つのプロジェクトが検討されている。

4. 研究体制の強化

グローバルコミュニケーション計画を推進するにあたり、NICT 内には「先進的音声翻訳研究開発推進センター」(Advanced Speech Translation Research and Development Promotion Center: ASTREC) が設立され、民間企業から約 20 名の研究者を迎えた。

さらに、グローバルコミュニケーション開発推進協議会も設立され、NICT を中心に産学官の力を結集して、多言語音声翻訳技術の精度を高め、さまざまなアプリケーションに成果を適用して社会展開していくための検討を行っている。協議会には音声翻訳に関連する大学・研究機関ばかりではなく、医療、交通、観光など関連性が強く伺える分野、その他の各種メーカーや翻訳に携わる企業など、2015 年 10 月現在で 128 機関がすでに参加している。

5. 将来の社会実装に向けて

音声翻訳で今日抱えている課題の克服のみならず、現実的な理想イメージが掲げられている。たとえばショッピングの分野では、見栄えも考慮し、店員が所持する端末のデザイン性への配慮なども挙げられている。

荻原：たとえば緊急医療でも、機械翻訳を使うのに適した場面としては想像しにくいかもしれませんが、頭が痛いのかお腹が痛いのかを把握するだけでも大きな違いをもたらすことがあると考えられます。

クリティカルな場面では通訳者に接続するための HELP ボタンの搭載など、実社会での現実的な利用を見据えた目標が示されており、音声翻訳システムと人による翻訳・通訳との棲み分けも考慮されていることが伺える。

タクシー業界などでは運転中にヘッドセットを装着したり、スマートフォンの操作は不可能であり、カーナビと組み合わせた端末の必要性など、協議会でさら

に改善へつながる課題を集積する予定とのこと。

6. 翻訳業界に及ぼす影響

河野：翻訳業界では機械翻訳の発展が目覚ましいという現状があります。グローバルコミュニケーション計画の一部として、音声翻訳への取組以外に、テキスト翻訳へ発展させる予定はありますか？

荻原：NICT では特許文書なども含め、テキスト翻訳にも取り組んでいます。ただし、旅行会話のためなどとは異なるコーパスを収集する必要が発生します。音声翻訳への取組と並行して進めていく予定です。

河野：私企業の事業とは異なると思いますが、世界戦略についてお考えはお持ちですか？たとえば、2015年10月末のアプリのダウンロード数を見るとVoiceTraは約14万、一方、Google翻訳は約400万ですが？

荻原：日本語を取り扱う翻訳という点では優れた技術であると思っています。表現豊かな日本語特有の言い回しなども含め、日本国内の利用者からの要請にはさらに迅速に対応する枠組みをつくっていきたいと考えています。

河野：グローバルコミュニケーション計画が翻訳業界に影響を及ぼすようなことは考えられますか？たとえば、シンポジウムやセミナーを通じて、協議会の参加企業が機械翻訳の利用を促進するような機会がもたらされるでしょうか？NICTに限定せず、機械翻訳の技術を活用している翻訳支援ツールはありますが。

荻原：協議会の参加企業にも、同様のシステムを活用している企業はあります。また、協議会の次の活動では、会員企業の情報交換会を計画しています。翻訳技術を活用する様々なテクノロジーを見せ合うというところでしょうか。グローバルコミュニケーション計画を推進することにより、NICTの翻訳技術が多方面で活用されることを望んでいます。

スマートフォン向け音声翻訳アプリ
RECAIUS™ 「音声トランスレータ」

株式会社東芝 インダストリアル ICT ソリューション社

株式会社東芝は、意思疎通・コミュニケーションをサポートするスマートフォン向け音声翻訳アプリ「RECAIUS™ 音声トランスレータ」を発表しました。

1. RECAIUS (リカياس) とは

音声や映像から人の意図や状況を理解し、人にわかりやすく伝え、人々の様々な活動をサポートするクラウドサービスの総称です。東芝は、人の「見る・聴く・話す」といった能力をRECAIUSで補完・支援し、多くの人が安心、安全、快適に過ごせる社会の実現を目指します。

2. 「音声トランスレータ」とは

「音声トランスレータ」は音声翻訳アプリです。スマートフォンに向かって発話した内容を、テキストと音声でリアルタイムに翻訳します。

★逐次翻訳

長い文や複雑な文章でもシステムが自動判別し、話すと同時に翻訳。自然な会話が可能。

★専門用語の辞書登録

専門用語や外国人観光客に分かりやすい対訳を付けてお客様専用辞書に登録可能。

※別途登録作業が必要です。

★音声認識・口語翻訳・音声合成の連動システム

同時通訳に必要な音声認識・口語翻訳・音声合成の技術を自社開発。これらの連動システムをクラウドで実現。

＜翻訳対象言語＞

- ・日本語⇄英語
- ・日本語⇄中国語（簡体字）
- ・日本語⇄中国語（繁体字）
- ・日本語⇄韓国語

＜主な用途＞

商品や施設のご案内、イベント・ツアー・店舗での接客において外国の方とのコミュニケーションを助けます。交通機関・商業施設など様々なシーンでご利用可能です。

【例】一般的な翻訳サービスでは、「トレーナー」などの和製英語は、本来の意味「スポーツのトレーナー」として翻訳されてしまいますが、和製英語と対訳の事前登録により、正確な翻訳でショッピングをサポートできます。



【画面例】



<主な機能>

★同時通訳機能

日本語・英語・中国語の音声翻訳を実現

★多言語同時通訳機能

日本語から英語、中国語、韓国語への翻訳結果を同時に表示することが可能。翻訳言語がわからない際にも対応が可能。

★フレーズ集（定型文）登録機能

説明で良く使うフレーズを、カテゴリーごとに分けて登録。正確な対訳を用いることで、丁寧な対応が可能。

<動作環境>

OS : Android5.1

※iOS 次バージョン以降リリース予定

<検証端末>

Nexus6 (AndroidOS5.1.1)

3. その他の製品ラインナップ

製品ラインナップとして、WebAPI タイプの「同時通訳サービス」もご用意しています。

★同時通訳サービス

「同時通訳サービス」は、音声翻訳に必要な音声認識、口語翻訳、音声合成エンジンを WebAPI で利用できるサービスです。

<管理機能>

アカウント認証

ログ管理

<音声認識機能>

対応言語：日本語、英語、中国語

辞書カスタマイズ：日本語のみ単語の辞書登録
(最大1万語)

<口語翻訳機能>

対応言語：日本語⇄英語、日本語⇄中国語、日本語⇄韓国語

辞書カスタマイズ：対応言語の単語の辞書登録
(最大1万語/言語)

<音声合成機能>

対応言語：日本語、米語、北京語、韓国語

<動作環境>

管理ポータル Web ブラウザ

- InternetExplorer11
- Chrome 最新版
- Firefox35 以上
- Safari 8.0.3

■製品情報

「RECAIUS™ホームページ」

<http://www.toshiba.co.jp/cl/pro/recaius/>

- ・本内容は予告なく変更する場合があります。
- ・RECAIUSは、株式会社東芝の商標です。
- ・Microsoft、Windows、InternetExplorerは、米国Microsoft Corporationの米国またはその他の国における登録商標または商標です。
- ・本記載の社名及び商品名はそれぞれ各社が商標または登録商標として使用している場合があります。

多言語翻訳を支える大規模な辞書データベース

春遍 雀来(ハルペン ジャック)

株式会社 日中韓辞典研究所(CJKI) <http://www.cjk.org>

1. はじめに

近年、技術・学術・文化等の多方面に於いて諸外国との相互理解・交流の重要性が再認識されている。2020年の東京オリンピック開催に向け、多言語情報の需要は今後ますます高まるであろう。

IT技術の発達に伴い、多言語情報は企業から一般ユーザーまで広く活用されるようになったが、そのような技術に不可欠なのが豊富な情報を包括した大規模な辞書データベースである。

日中韓辞典研究所(CJKI)は、日中韓英を中心とする辞書データベースの編纂を行っている。

日中韓英辞書データベースには、固有名詞・専門用語の他、日本語の語彙・異表記等も含めて約2400万項目が収録されている。またIT関連の大手企業に広く採用されている中日・日中専門用語データベースは、20分野に亘る専門用語を網羅した日中対訳辞書である。

これらの辞書データベースは、特許翻訳を含む翻訳業務の分野で人間の手による翻訳及び機械翻訳の作業効率と精度を格段に向上させてきた他、形態素解析・固有名認識・用語抽出等、自然言語処理に於ける幅広い分野で応用が可能である。

2. 多言語辞書の重要性 (固有表現抽出と処理)

辞書データベースは、翻訳のみならず各種データ処理の場面でも活用される。

例えば自然言語処理において特に扱いが難しい固有名詞では、異表記のバリエーション(アラブ人名「アブドゥル・ラフマーン」には千通り以上のアルファベット表記がある)、平仮名表記の中国語訳(市町村名等)に対応しなければならない。また、一般語彙に於ける同義語(「software」は簡体字では「软件」、繁体字では「軟體」と表記)を処理する際にも辞書データベースは有用である。

これらの問題点を踏まえ、当研究所では専門用語を含む

膨大な辞書データベースの構築・拡張を続けている。

3. 機械翻訳と人間翻訳、および作業効率と精度の向上

地名やPOI(points of interests = ホテル、公園、大学等)は数が膨大である上、名称が変更される場合もあり、人間が全て翻訳するのは難しい。各言語体系に基づく正しい表記が必要であるため全面機械化も不可能で、辞書データベースが必須となる。

全面的な自動翻訳が不可能であるため、求められる精度と費用を勘案して機械翻訳と人間翻訳の割合を決定する。固有名詞翻訳作業に当たっては字訳・音訳・意訳・意音訳・翻訳の5種類と、それらの組み合わせによる変換方法がある。機械化可能な割合が高く、安価で速いが精度が上がりにくいものから、人間翻訳で高価だが翻訳としての正確さを期す(定訳を選択する)ものまで様々である。

日本の地名・公共施設名

Japanese	成田国際空港	京都府庁
Chinese (SC)	成田国际机场	京都府庁
Chinese (TC)	成田國際機場	京都府廳
Korean	나리타국제공항	교토부청
English	Narita International Airport	Kyoto Prefectural Office
Arabic	الدولي ناري تا مطار	ك يوطو محافظة مكن تب
Indonesian	Bandar Udara Internasional Narita	Kantor Pemerintahan Kyoto
Vietnamese	Sân bay quốc tế Narita	Tòa nhà chính quyền tỉnh Kyoto
Thai	สนามบินนานาชาตินาริตะ	ที่ว่าการจังหวัดเกียวโต
Hindi	नारिता अंतर्राष्ट्रीय हवाई अड्डा	क्योटो प्रीफेक्चर मुख्यालय
Russian	Международный аэропорт Нарита	администрация префектуры Киото
German	Internationaler Flughafen Narita	Prä fekturverwaltung g Kyoto
Portuguese	Aeroporto Internacional de Narita	Sede do Governo de Quioto
Spanish	Aeropuerto Internacional de Narita	Oficina Prefectural de Kyoto
French	Aéroport international de Narita	Préfecture de Kyoto
Italian	Aeroporto Internazionale di Narita	Sede del Governo prefettizio di Kyoto

4. 自然言語処理への幅広い応用

自然言語処理で課題となる以下の異表記をデータベースに全て包括することで、対応が可能になる。

異表記：日本語は世界でも表記の幅が広い言語であり、漢字表記・平仮名表記・片仮名表記・交ぜ書き等がある。更に片仮名語の異表記（コンピュータとコンピューター、メイドとメード等）も多数出現する。

同音異形異義語：うまい＝美味しい、上手い、巧い等。

5. 日中韓英大規模辞書データベースのご提供

当研究所では、こうした異表記を網羅する日中韓英各語とアラビア語の大規模な辞書データベースを提供し、世界の大手企業がこれを採用している。中国語のデータベースには検証済みの正確なピンインも収録されている。先進的な計算辞書学の手法によって構築・維持された当研究所のデータベースは、固有名詞・専門用語の他、日本語の語彙・異表記・音韻等も含め、約2400万項目に上る。

日中韓英固有名詞データベース

	日英	日中	日韓
中国人名	1,000,000	1,000,000	1,000,000
中国地名	2,400	5,600	3,000
韓国人名	13,000	2,100	13,000
韓国地名	5,900	2,000	5,900
日本人名	390,000	281,000	390,000
日本人姓	150,000	91,000	150,000
日本地名	77,000	74,000	77,000
西洋人名	31,000	38,000	10,000
西洋地名	1,100	2,500	1,800
合 計	1,670,400	1,496,200	1,650,700

「日中韓英固有名詞データベース」は日中韓英語の各種固有名詞辞典を含み、総計1000万項目に及ぶ、他に類を見ない大規模なデータベースである。

用途

- ◇ 機械翻訳
- ◇ 情報検索
- ◇ 形態素解析
- ◇ 電子辞書
- ◇ 入力システム
- ◇ 固有名認識

中日日中専門用語データベース

分野	中国語	日本語
医学	腎上腺素能受体	アドレナリン受容体
生物	亲和性	親和性
生物	亲和层析法	アフィニティークロマトグラフィ
生物	琼脂扩散法	寒天拡散法
生物	琼脂糖	アガロース
生物	琼脂胶	アガロペクチン
生物	类蛋白	アルブミノイド
医学	类天花	アラストリム
医学	变应性试验	アレルギー試験
医学	变应性肉芽肿	アレルギー性肉芽腫

「中日日中専門用語データベース」は日中二ヶ国語の双方向対訳辞書である。コンピュータ科学からバイオテクノロジーに至る20分野に亘る幅広い専門用語を収録しており、収録語は中日・日中それぞれ約80万語、総計約160万語に及ぶ。

用途

- ◇ 特許翻訳を含む各種翻訳業務
- ◇ 用語の抽出やインデックス作成に役立つ情報検索アプリケーション
- ◇ 形態素解析や分節システム等、各種の自然言語処理アプリケーション
- ◇ スマートフォンアプリケーションや電子辞書・CD-ROM等

6. まとめ

統計型機械翻訳だけでは翻訳及びデータ処理上の全ての問題を解決することはできない。統計処理、辞書検索、ルール処理のハイブリッド型機械翻訳が最も有効である。CJKI は人間翻訳に役立つ辞書データのみならず、このような MT システムをサポートする各種の大規模な辞書データベースを提供している。

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 のスタートから第十八回を迎えることができました。今号では、法人会員一社からのご寄稿をいただいております。

独自のお取り組みのご紹介、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

株式会社アイデア・インスティテュート / IDEA INSTITUTE INC.

<http://www.idea institute.co.jp/>

ブログサイト <http://www.idea inst.com/blog/>

【本社】〒150-0013 東京都渋谷区恵比寿 3-1-3 アイデアビル

【大阪営業所】〒530-0055 大阪市北区野崎町 9-8 永楽ニッセイビル 8F

自己紹介

弊社は 1973 年に創業させていただき、今期は 44 期目となりますが“所謂”翻訳会社であります。“所謂”を殊更につけた理由は「翻訳会社」という自己紹介につき、多様な場面で、「翻訳会社」を知っていらっしゃる方と、そうではない方とお話を始める時のギャップを感じる時が多いからであります。従って筆者は最近「翻訳会社」との係りがあまりない方への自己紹介は、横文字を安易に使うことは大変失礼なことで十分に承知しつつも、＜所謂“LSP(Language Service Provider)”と呼ばれていて、言語周りの翻訳を含めたライティング、制作、DTP、印刷、通訳、ローカライゼーションなど含め、お客様のご要望に沿った、80 言語以上の多言語展開のサービスをさせていただいている会社です＞とお答えするようにしています。これでなんとなく“ふーん、そうですか”と言って頂いてほっとしているところもございます。

弊社は過去 AAMT に所属させて頂いていた時期もございますが、昨年 2015 年に復帰させて頂くことができました。事務局の皆様、先輩会員の皆様のお世話になっているところでございます。

筆者自身は 1978 年に初めて大学の教養課程で、工学系学生は必修であった確か機械計算演習とかいう名前の実習で、UNIVAC 1106 という“当時”の大型コンピューターに触れ、今ではパソコン一台で難なくできることをパンチカードを打つところから始め、今では古典であろう FORTRAN や COBOL の入門編を恐る恐るさわっていた程度であります。誤解がないようにですが、筆者は MT エンジン作成に必要であろうところの高等数学とは縁のないバケガクの世界へ進むことになり、その後企業からフランスやドイツへ留学させていただいたのは有機化学であり、30 代中頃には既に“文転”し、今日でも MT のアルゴリズムはブラックボックスであります。

MT/翻訳とのかかわり

MTおよび翻訳業界に期待すること

今では自分の学生時代であったのか、企業へ入ってからのことか記憶も定かではありませんが、「インベーダーゲーム」というゲームを覚えている方もいらっしゃるかと思います。これを当時、自分の浅学も顧みずに機械語（所謂二進法の世界そのもの）のプログラムを雑誌に公開されているそのままを、わけもわからずコンピューターへ入力し、ピコピコとゲームセンターそのままの世界が出現したときは非常に驚いた経験があります。つまりこれはブラックボックスで遊んだ大昔の経験であります。

MTにはバックナンバーにおいて、先輩会員さまがこの広場でも書かれていますように、大きく分けて、RBMTとSMTと、ついでに言えばそれらの複合したハイブリッド型という分類があるかと思います。弊社では欧米よりおくれること10年余で翻訳支援ツールを、それでも本邦では決しておくれずに、20年ほど前の1996年に導入をしております。しかし、上記MTツールについてはやっと実用化の段階へ踏み込んできており、欧州からの遅れは十二分に承知しつつも、やっとならんと立ちあがった新生児の状況です。一方、SMTへ適用できうる、44年の翻訳業務を営んだ中で蓄積したメモリー(TM)は弊社のアセットといつて宜しいかと思います。

RBMTにせよ、SMTにせよ、弊社にとってはブラックボックスではありますが、そのアウトプットの品質の評価、そして、それをHT(所謂人間が翻訳した完成品？レベル)のレベルまで高めるPE(Post Edit; ポストエディット)、MTでデータ処理をする前の元言語を機械に易しい言語表記・配列へ修正するプリエディットや、機械に易しいライティングなど、MTはブラックボックスであっても大昔にインベーダーゲームで（知的に？）遊んだように、弊社のようなユーザーはそれらを実用的に使うことはできるのではないかと近未来に期待をしているところです。そしてそれはさまざまなお客様と外部の協働者の皆様ともウィンウィン、いえ、近江商人の「三方よし」の世界の実現を期待しています。

最近の文献類によりますと、RBMT単独は決してMTの世界のトップランナーではないようですが、そのデータ処理の速さには驚かされるばかりです。これもブラックボックスであってもそのブラックボックスは所詮機械ですので、その性格、機能、特徴をつかんで優しくそのデータを人間が後処理をする、またはそのアウトプットの特徴を踏まえうえて活かしてやる、といった新入社員を教育するような優しさも業務に展開できるのではないかと筆者のような現役卒業世代は感じることの多いこの頃です。

株式会社アイデア・インスティテュート

相談役

古畑 欣也

AAMTへの要望

AAMT内でのさまざまな会員企業さま、個人さまとのふれあいの場の提供を感謝しています。さらにAMTAやEAMTとの情報の共有化のプラットフォームのような場の深化も期待しております。

これまでの AAMT Forum メールマガジン Vol.8

機械翻訳課題調査委員会 WG1、WG2

はじめに

AAMT Forum を通じて配信したメールマガジンのバックナンバーをお届けします。今回は第 66 号から第 74 号です。

今回は、昨今の訪日観光客の増加を背景とした様々な MT アプリケーション・サービスに関する記事に加え、Google 翻訳サービスが 10 周年を迎えたという記事もありました。この 10 年で対応言語が 3 言語から 103 言語に増加したそうです。これから 10 年後、20 年後、どのように変化を迎えるのか楽しみです。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 66 号から第 74 号、そして号外を再びお届けします。

AAMT Forum メールマガジン

2016/02/04 配信 vol.066

こんにちは。AAMT Forum メールマガ担当です。立春が過ぎましたが、まだまだ寒さは厳しいですね。例年よりも遅いそうですが、今年もインフルエンザの流行シーズンに入りました。うがい、手洗いを忘れないようにしましょう。では、メールマガジン第 66 号をお届けします。

■看板を撮るだけで外国語表示に！訪日外国人向け案内サービス

NTT サービスエボリューション研究所は、1 月 22 日までパシフィコ横浜で開催された「SC ビジネスフェア 2016」にて、丸紅テクノシステムと pdc の合同ブース内で「かざして案内サービス」の参考展示を行いました。

<http://www.rbbtoday.com/article/2016/01/29/139208.html>

■進化する医学専門翻訳ソフト！「LogoVista メディカル」シリーズ

ロゴヴィスタ株式会社は、医学文献翻訳ソフト「LogoVista メディカル 2016 ベーシック / フルパック」Windows 版、Macintosh 版を、2016 年 2 月 12 日（金）より、全国パソコンショップ、カメラ系量販店などで発売します。

<https://www.logovista.co.jp/LVERP/information/news/2016-0120-med2016.html>

AAMT Forum メールマガジン

2016/02/18 配信 vol.067

こんにちは。AAMT Forum メールマガ担当です。春寒の候、いかがお過ごしでしょうか。暖かくなってきたと思ったら真冬の寒さに逆戻りなど、春を前にして

天気が変わりやすい時期になりました。体調を崩さないようにその時々 of 服装に気をつけましょう。

■フィート、スマートフォン向けサーバーレス型音声翻訳アプリ

情報通信研究機構（NICT）が開発した翻訳技術をベースに、フィートはインターネットに接続できない環境でも利用できるスマートフォン向けサーバーレス型音声翻訳アプリ「どこでも翻訳」を 2016 年 2 月 4 日から提供開始しました。

<http://itpro.nikkeibp.co.jp/atcl/news/16/020800399/?rt=nocnt>

■三菱地所が接客向け翻訳アプリを開発、まず 1000 店舗で活用開始

情報通信研究機構（NICT）が開発した多言語音声翻訳アプリ「VoiceTra」をベースに、三菱地所と三菱地所プロパティマネジメントは接客用途に特化した多言語翻訳アプリを開発すると発表しました。

<http://itpro.nikkeibp.co.jp/atcl/news/16/021000422/>

AAMT Forum メールマガジン

2016/03/03 配信 vol.068

こんにちは。AAMT Forum メールマガ担当です。ひと雨ごとに春の色が濃さを増すこの頃、いかがお過ごしでしょうか。真冬に戻ったような大荒れの天気から暖くなり、春の到来を感じますね。では、メールマガジン第 68 号をお届けします。

■Google、Google 翻訳の対応言語に 13 言語を追加

Google が翻訳サービス「Google 翻訳」の対応言語に海外 13 言語を追加し、合計で 103 の言語に対応したと発表しました。

<http://getnews.jp/archives/1403221>

■Microsoft Translator で 11 言語のオフライン翻訳が可能

最新バージョンの Microsoft Translator アプリにオフラインでのサポートが搭載されました。

http://www.lifehacker.jp/2016/02/160227microsoft_translator11.html

AAMT Forum メールマガジン

2016/03/18 配信 vol.069

こんにちは。AAMT Forum メールマガ担当です。この頃はすっかり春らしくなってきましたね。本年度も終わりが近づいてきましたが、新年度の準備はよろしいでしょうか。では、メールマガジン第 69 号をお届けします。

■翻訳能力を TOEIC に換算してよいか

機械翻訳の性能をどのように TOEIC スコアへ換算しているのかと、その方法に対して、翻訳者として感じる疑問についてのブログ記事です。

<http://blog.nishinos.com/archives/5019627.html>

■音声翻訳アプリ搭載の駅案内用タブレットを全駅に導入～“訪日外国人旅客のおもてなし”を強化！～

京浜急行電鉄株式会社は、一部の駅のみに導入されていた音声翻訳アプリ搭載の駅案内用タブレット端末を、2016年2月18日から泉岳寺駅を除く京急線全72駅に導入しました。

http://www.keikyu.co.jp/company/news/2015/20160217HP_15201TK.html

AAMT Forum メールマガジン

2016/04/01 配信 vol.070

こんにちは。AAMT Forum メールマガ担当です。陽春の候、いかがお過ごしでしょうか。この時期は春眠を覚えずといいますが、寝坊しないように気をつけましょう。では、メールマガジン第70号をお届けします。

■凸版印刷、自治体窓口業務向け音声翻訳システムの研究開発に着手

凸版印刷はフィートと共同で、自治体窓口業務に対応した国内で初めての音声翻訳システムの研究開発に着手したと発表しました。

<http://news.mynavi.jp/news/2016/03/23/089/>

■「瀬戸内国際芸術祭 2016」で多言語音声翻訳システムを活用した訪日外国人の「おもてなし」を実現

KDDIは、多言語音声翻訳システムを「瀬戸芸」の案内スタッフにご利用いただくことで、従来課題となっていた「瀬戸芸」の案内スタッフと訪日外国人の円

滑なコミュニケーションを実現しました。

<http://news.kddi.com/kddi/corporate/csr-topic/2016/03/20/1684.html>

AAMT Forum メールマガジン

2016/04/15 配信 vol.071

こんにちは。AAMT Forum メールマガ担当です。吹く風もやわらかく、過ごしやすい日々が続く季節となりましたね。新生活にも慣れた頃でしょうか。ずっと張り詰めていると疲れますが、慣れ始めが危険ともいいますので、適度な緊張感を保ちましょう。では、メールマガジン第71号をお届けします。

■Microsoft のディープラーニング採用翻訳アプリ、iOS版でもオフライン対応

米 Microsoft は iOS 版翻訳／通訳アプリ「Microsoft Translator」をアップデートし、オフラインでの利用を可能にしたと発表しました。利用できる言語パッケージは Android 版より多い 40 カ国語以上です。

<http://www.itmedia.co.jp/news/articles/1604/11/news079.html>

■特許庁と国立研究開発法人情報通信研究機構（NICT）の特許文献の機械翻訳に関する協力の継続について

平成 26 年 7 月、特許庁と NICT は平成 27 年度末まで特許文献の機械翻訳の精度向上及び活用促進の協力を進めることに合意しており、平成 28 年度はその協力の範囲を拡大した上で、協力関係を継続することとなりました。

https://www.jpo.go.jp/shiryou/toushin/chousa/nict_kikai_honyaku.htm

AAMT Forum メールマガジン

2016/04/27 配信 vol.072

こんにちは。AAMT Forum メールマガジン担当です。今年もゴールデンウィークがやってきましたね。旅行に出かけたり、家でゆっくり過ごしたりと予定は人それぞれだと思いますが、大型連休を満喫しましょう。では、メールマガジン第72号をお届けします。

■翻訳ピカイチシリーズ3製品の新バージョンを5年ぶりに発売！

株式会社クロスランゲージは、翻訳ピカイチシリーズの中日・日中翻訳、韓日・日韓翻訳、英語+欧州5か国語⇄日本語翻訳の3製品を、2016年5月27日より販売します。

<http://prtimes.jp/main/html/rd/p/000000080.000002804.html>

■Android アプリ「Microsoft Translator」に画像認識での自動翻訳機能が追加

撮影した画像に含まれる言語は、自動検出のほか英語や中国語など21か国語を指定可能です。そこからクリンゴン語などを含む48か国語への翻訳が可能です。

http://internet.watch.impress.co.jp/docs/news/20160422_754668.html

AAMT Forum メールマガジン

2016/05/13 配信 vol.073

こんにちは。AAMT Forum メールマガジン担当です。立夏の候、いかにお過ごしでしょうか。大型連休を終え、リフレッシュはできたでしょうか。心身を休めたら、五月病などに罹らぬように気を引き締めていきましょう。では、メールマガジン第73号をお届けします。

■Google 翻訳が10歳になりました。

3言語から始まった対応言語は103言語に拡大し、数百人規模だったユーザー数は、数百万人を超えるまでになりました。そんなGoogle 翻訳にまつわる10のことの紹介記事です。

<http://googlejapan.blogspot.jp/2016/04/google-10.html>

■ビジネス向け翻訳ソフト「YarakuZen（ヤラクゼン）」 旅館向けの翻訳データセットが追加可能に

八楽は、ビジネス向けオンライン翻訳ソフト「ヤラクゼン」において使用することができる業種、業界別翻訳データセットの販売を開始しました。その第1弾では旅館向けのデータベース（旅館パッケージ）が用意されています。

<http://japan.cnet.com/release/30142491/>

UTX1.20 β 仕様書（抜粋）

課題調査委員会ワーキンググループ3

※以下は UTX 1.20 β 仕様書の抜粋であり、一部の項目、表、例が省略されています。正式な仕様書は <http://www.aamt.info/japanese/utx/> からダウンロードできます。

1. 用語と定義

このセクションでは、UTX に関する重要な用語と定義を説明する。

1.1. UTX 用語集の種類

UTX 用語集は、特定の専門分野内での用語の集合である。

UTX 用語集は、翻訳ソフトの正確さと流暢さを向上するためのデータ辞書として使用できる。「ユーザー辞書」とはユーザーによって作成された辞書であり、「システム辞書」とは、MT システム内に組み込まれた辞書である。

含まれる言語数の点では、単一言語、二言語、多言語の UTX 用語集が作成できる。

単一言語用語集は、1 つの言語のみの用語を含む。単一言語用語集は、用語チェックや、表現・技術用語の変化形を統一するために使用できる。

二言語用語集は、2 つの言語の用語を含む。UTX 用語集は、多くの場合、二言語である。別記ない限り、この仕様書では、UTX 用語集が二言語であると仮定する。

多言語用語集は、3 つ以上の言語の用語を含む。

注：UTX 1.11 以前は、二言語および単一言語用語集のみを認めていた。UTX 1.20 以後で、多言語用語集を作成できるようになった（「7 多言語用語集」参照）。

1.2. 用語

用語とは、起点言語あるいは目標言語の見出し語である。UTX 用語集内の用語は、辞書の見出し語のような単語の基本形（レンマ）である必要がある。「8 付録 A：UTX コンテンツガイドライン」も参照。

1.3. 項目

項目とは、UTX 用語集内の、1 つ以上の言語の用語および追加情報から構成される単位である。

注：UTX 用語集がルールベース機械翻訳で使用される場合、「一語一義」の原則に従う必要がある。すなわち、1 つの用語は、その辞書の特定分野においてはただ 1 つの意味（概念）しか持たないということである。

1.4. ユーザーロール

UTX 用語集には、3 つのユーザーロールがある。

用語集ユーザーは、用語集を使用する個人である。用語集ユーザーは、用語集の分野の知識と、UTX についての最低限の知識を持つ。用語集ユーザーは、コメントや他の手段により、用語の優先度や用語に関するその他の

事項についての意見を表明できる。

用語提出者とは、用語集への新しい追加項目を提案する個人である。用語提出者は、用語集の分野の知識を十分に持ち、UTX の基本的な知識を持つ必要がある。

用語集管理者とは、ある用語集に対して責任を持つ個人である。用語集管理者は、用語集の分野の専門的知識を持つ同時に、UTX についても深い知識を持つ必要がある。用語集管理者は、ある項目が用語集での適切な項目として承認されるか、(必要なら) 削除されるべきかを決定する (参照「5.4.2 approved (承認)」)。また用語集管理者は、各種の用語ステータスの割り当てや変更ができる。

1.5. 翻訳方向

ある言語から別の言語への翻訳の方向 (「翻訳方向」) は、一方向、双方向、多方向のいずれかである。この情報は、「4.3.10 directionality プロパティ」で指定できる。

一方向用語集は、翻訳方向が主として起点言語から目標言語への一方向となる用語集である。

例：一方向二言語日英 UTX 用語集

主な翻訳方向：日本語から英語

注：一方向用語集内のいくつかの用語をエクスポートして逆方向で 사용할 ことができることがある。この場合、起点言語が目標言語になり、目標言語が起点言語になる。反対の翻訳方向で使うように用語集をエクスポートする操作は、逆方向エクスポートと呼ばれる。

双方向用語集とは、二方向の翻訳でできるように作られた用語集である。一方の言語の用語はもう一方の言語に翻訳でき、その逆も可能である。

例：双方向多言語の英語・フランス語・ドイツ語用語集

翻訳方向：英語からフランス語、フランス語から英語、英語からドイツ語、ドイツ語から英語 (しかしフランス語からドイツ語、ドイツ語からフランス語には不可)

多方向用語集とは、任意の翻訳方向の翻訳でできるように作られた多言語用語集である。

1.6. UTX 対応アプリケーションと UTX 変換ツール

UTX 対応アプリケーションとは、用語集を UTX 形式で読み込み、または書き出しできるアプリケーションである。そのようなアプリケーションの例として、翻訳ソフト、用語集ツール、および UTX エディターがある。

UTX 対応アプリケーションのうち、UTX 変換ツールとは、UTX 用語集をその他のファイル形式に変換し、またその他のファイル形式を UTX 用語集に変換するツールの総称である。

2. UTX ファイルの構造

UTX ファイルのファイル拡張子は ".utx" である。

UTX ファイルは以下の構造を持つ。

ヘッダー

ヘッダーは、以下の要素から構成される。

1. 用語集プロパティ。「4.3 用語集プロパティ」を参照。
2. 用語集詳細 (省略可能)。「4.4 用語集詳細」を参照。

3. フィールド定義。「4.5 フィールド定義」を参照。

本文

項目から構成される本文（タブ区切りテキスト）。「5 フィールド定義と本文」を参照。

以下の図は、UTX 用語集の構造を示す。構成要素は後節で詳述する。

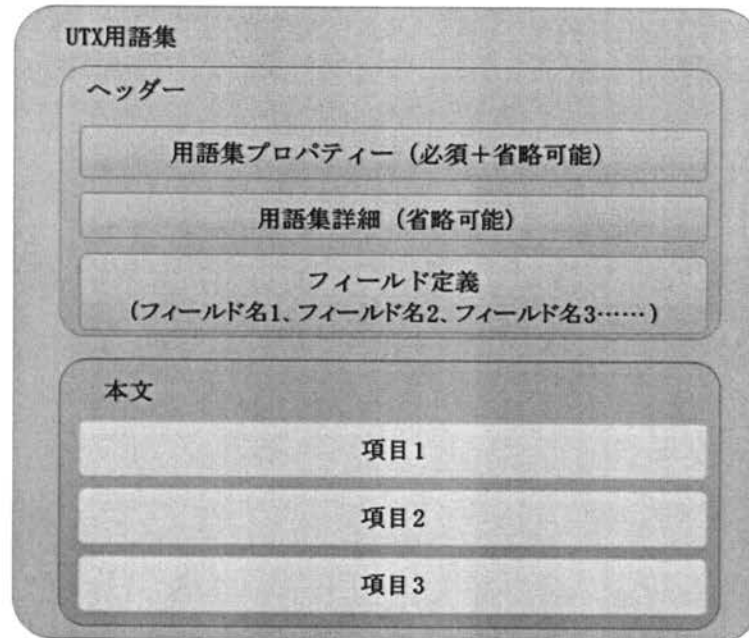


図 1 : UTX 用語集の構造

2.1. 文字コード

UTX ファイルの文字コードは、UTF-8（BOM 付き）である。

注：「BOM なし」であった UTX 1.11 から変更されていることに注意。BOM は、Microsoft Windows などの OS で UTF-8 エンコードを識別するために使用される特殊文字である。

2.2. 行

UTX ファイルで使用される改行コードは、「¥r¥n」（CR+LF）である。改行コードは、行の終端を表す特殊文字である。空行は認められない。

2.3. 行コメント

#（シャープ記号）で始まるすべての行は、行コメントとして扱われる。

3. 要素の概要

以下は、UTX の要素の概要である。

3.1. ヘッダー

3.1.1. 必須の用語集プロパティ

名前	構文	例
UTX version	#UTX <version>	#UTX 1.20

3.1.2. 省略可能の用語集プロパティ

名前	構文	例
lang/term/src/tgt(言語宣言)	lang: <起点言語>/<目標言語>あるいは <term/src/tgt>:<言語 1>/<term/src/tgt>:<言語 2>.....	lang: en/ja あるいは term:en/term:ja あるいは src:en/tgt:ja/tgt:fr
creation date	creation date: YYYY-MM-DD	creation date: 2015-04-10
	creation date: YYYY-MM-DDThh:mm:ssTZD	creation date: 2015-04-10T12:34:56Z
last modified date	last modified date: YYYY-MM-DD	last modified date: 2015-05-10
	last modified date: YYYY-MM-DDThh:mm:ssTZD	last modified date: 2015-05-10T12:34:56Z
glossary ID	glossary ID: <文字列>	glossary ID: Brain surgery
creator	creator: <文字列>	creator: Yamada Tarou
glossary administrator	glossary administrator: <文字列>	glossary administrator: Yamada Hanako
copyright	copyright: <文字列>(<年>)	copyright: AAMT (2013)
license	license: <文字列>	license: CC-by 4.0
directionality	directionality: <uni/bi/multi>	directionality: bi
sortable	sortable: <true/false>	sortable: true
glossary version	glossary version: <数値>	glossary version: 1.00

3.2. フィールド (本文)

フィールド名	構文	例
term/src/tgt	term:<言語> src:<言語> tgt:<言語>など	src:en tgt:ja (cranium 頭骨)

part of speech	noun/properNoun/verb/vt/vi/prenomina l adjective/adverb/sentence	noun
concept ID	<数値>	45
term status	blank/provisional/approved/non-stand ard/forbidden/rejected/obsolete	approved
glossary ID	<文字列>	Brain surgery

3.3. 最もシンプルな UTX の例

以下は、最小の必要情報を備えた完全な UTX 用語集の例である。各要素は、以下の章で後述する。

#UTX 1.20	
#term:en	term:ja
Test	テスト

4. ヘッダー

4.1. UTX ヘッダーの構造

UTX ヘッダーは、「#」から始まる、用語集プロパティおよびフィールド定義の、少なくとも 2 つの必須行から成る。用語集の詳細を説明する目的で、2 つの必須行の間に、1 行以上のコメントアウト行を追加できる。この場合、2 つの「#」文字を先頭に置く。

```
#<用語集プロパティ>      必須+省略可能
#<追加の用語集プロパティ>      省略可能
(##<用語集詳細>省略可能)
(#<追加の用語集プロパティ>      省略可能)
.....
#<フィールド定義>          必須
```

4.2. 言語タグ

言語宣言 (language declaration) を含む UTX 用語集の任意の部分で、言語を示すために IETF BCP 47 言語タグを使用する。<<http://www.w3.org/International/articles/language-tags/Overview.en.php>>を参照。

言語タグは、言語宣言以外に、あるフィールドが特定の言語に属することを示すために使用される。「5.3 品詞フィールド」を参照。

言語タグの例

言語	言語タグ	地域サブタグが付いた言語タグ	文字種サブタグが付いた言語タグ
中国語（簡体字、中華人民共和	zh	zh-CN	zh-Hans

国)			
中国語（繁体字、台湾）	zh	zh-TW	zh-Hant
英語（アメリカ）	en	en-US	
英語（イギリス）	en	en-GB	
日本語	ja	ja-JP	
韓国語	ko	ko-KR	

4.3. 用語集プロパティ

用語集プロパティは2行以上にわたることがある。追加の行は、「#」、半角スペースに続けて、用語集プロパティを記載する。

用語集プロパティは、プロパティ名、コロン、半角スペース、およびプロパティ値から成る。<プロパティ名>:<プロパティ値>

注：UTX バージョンおよび言語宣言は、プロパティ値のみで示される。

省略可能の用語集プロパティを含む用語集プロパティ

```
#UTX <バージョン>;<(起点)言語>/<(目標)言語 1>/<(目標)言語 2>;<creation date>;<last modified date>;<glossary ID>;<creator>;<copyright>;␣# <license>; <directionality>; <sortable>; <glossary version>; <other optional properties>␣
```

注：「␣」記号は改行を意味する。

二言語 UTX ファイル用のヘッダーの例

```
#UTX 1.20; lang: en/ja; creation date: 2012-04-15; copyright: AAMT (2012)
```

4.3.1. UTX version プロパティ

ヘッダーの最初の行は用語集プロパティを含む。これは、UTX ファイルに関する基本的な情報であり、各項目はセミコロンと半角スペースによって区切られる。UTX version は、唯一の必須の用語集プロパティである。

UTX version は、「UTX」および半角スペースの後に指定する。バージョン番号は 1.20 となる。

4.3.2. lang プロパティ（言語宣言）

lang プロパティ（言語宣言）は、用語集に含まれる用語の言語を示す。二言語用語集の言語宣言は以下の形式で指定する。

```
lang: <起点言語>/<目標言語>
```

各言語の役割は、term（中立）、src（起点言語）、および tgt（目標言語）を使用して明確にすることができる。この場合、「lang:」は記述しない。

```
<term/src/tgt>:<言語 1>/<term/src/tgt>:<言語 2>……
```

注：言語宣言は省略可能である。用語集に含まれる言語は、term/src/tgt フィールドでも示される（「5.2 term/src/tgt フィールド」を参照）。

4.3.3. **creation date** プロパティ

creation date プロパティは、用語集が最初に作成された日付（および時刻）を示す。ISO 8601 形式を使用する。

4.3.4. **last modified date** プロパティ

last modified date プロパティは、用語集が最後に修正された日付（および時刻）を示す。ISO 8601 形式を使用する。現地時間は、UTC に時差を追記することにより示すことができる。

4.3.5. **glossary ID** プロパティ

glossary ID プロパティは、用語集に対して一意の識別子として使用されるテキスト文字列である。これは、通常、用語集の分野を表す、用語集の「名前」である。

glossary ID プロパティは、用語集が単一分野からの用語を含んでいる場合に限り、使用される。用語集が複数の分野からの用語を含んでいる場合は、代わりに用語集 ID「フィールド」を使用する必要がある。「6.4 用語集 ID (glossary ID) フィールド」を参照。

4.3.6. **creator** プロパティ

creator プロパティは、用語集を作成した人の名前を示す。

4.3.7. **glossary administrator** プロパティ

glossary administrator プロパティは、用語の承認の責任を負う個人の名前を示す。

4.3.8. **copyright** プロパティ

copyright プロパティは、用語集のコンテンツ全体に対して著作権を持つ個人や組織、および著作権が効果を現した年を示す。

4.3.9. **license** プロパティ

license プロパティは、用語集の使用許諾、つまりユーザーがこの用語集をどのように使用できるかを示す。使用許諾は、クリエイティブコモンズ、パブリックドメイン、あるいはその他の形式で指定できる。未定の場合は、「undetermined」と指定できる。

例 : `license: undetermined`

4.3.10. **directionality** プロパティ

directionality は、用語集の翻訳方向を示すプロパティである（「1.5 翻訳方向」を参照）。**uni** は、用語集の項目が 1 つの翻訳方向で使用されるよう意図されていることを示す。**bi** は、用語集内のすべての項目が、双方向の翻訳で使用できることを示す。**multi** は、用語ステータスで許可されている範囲で、用語集内のすべての項目が、どのような方向の翻訳でも使用できることを示す。プロパティ値 **multi** は、多言語用語集でのみ指定できる。

用語集の種類	directionalityプロパティの可能な値
一言語用語集	N/A
二言語用語集	uniまたはbi
多言語用語集	uni、bi、またはmulti

注：UTX 1.11 では、翻訳方向性を示す目的で `bidirectional` フラグが使用された。これは、UTX 1.20 で多方向の翻訳方向性が導入されたことに伴い使用されなくなった。

4.3.11. sortable プロパティ

`sortable` プロパティは、いずれかのフィールドを使用して、用語集の項目を並べ替えであることを示す。プロパティ値はブール値 (`true/false`) である。項目の順序を維持するには、このプロパティ値を `false` に指定する。

注：たとえば、`sortable` プロパティにより、ユーザーは用語集に用語を追加した順番を維持したうえで、新しい用語を追加できる。また、このプロパティを指定することで、ある項目がコメントアウトされた場合でも、用語集内でのその項目の位置を維持できる。

4.3.12. glossary version プロパティ

`glossary version` プロパティは、用語集のバージョンを示す。これにより、修正の進行状況を記録できる。UTX 形式のバージョンと混同しないように注意すること。

4.4. 用語集詳細

用語集詳細は、用語集の詳細情報、追加の法的通知、および免責条項などの情報提供を行う、1 行以上のテキスト行である。用語集詳細の先頭には「`##`」を置く。

例

`##` 用語集の詳細情報をここに追加できる。

.....

4.5. フィールド定義

UTX ヘッダーの最後の行は、「`#`」で始まり、一連のフィールド定義（あるいは列定義）を含む。フィールド定義と本文は密接に関連するため、詳細は「5 フィールド定義と本文」で説明する。

5. フィールド定義と本文

UTX 用語集の本文は、1 行 1 項目の集合から構成される。行内では、要素（表計算での「セル」）はタブ文字によって区切られる。

5.1. フィールドのための言語タグ

言語固有のフィールドは、言語タグによって示される。言語タグのコロンの後にはスペースを入れない。

例

言語固有のフィールド	表記法
英語の用語（起点言語）	#src:en
日本語の用語（目標言語）	tgt:ja
日本語の用語ステータス（目標言語）	term status:ja
中国語（簡体字）用語（目標言語）	tgt:zh-CN
中国語（簡体字）用語ステータス（目標言語）	term status:zh-CN
日本語の用語に関するコメント（日本語の項目のためのユーザー定義フィールド）	comment:ja

注：「comment」のようなユーザー定義フィールドが項目全体に関わる場合は、言語タグを必要としない。しかし、特定の言語の用語に関係するユーザー定義フィールドは、言語固有のフィールドを必要とすることがある。たとえば、日本語の訳語のみに関するコメントがある場合、Comment フィールドではなく Comment:ja とする。

UTX を TBX に変換する場合、言語タグを持つフィールドは「用語」レベルとして扱われる。言語タグのないフィールドは、（概念 ID を除いて）「項目」レベルとして扱われる。これらのレベルの詳細については、TBX 仕様を参照のこと。<http://www.ttt.org/oscarstandards/tbx/>

5.2. term/src/tgt フィールド

用語のフィールドを示すために、term、src、あるいは tgt フィールドを使用する。これらのフィールドの言語は言語タグによって示される。これらのフィールドは用語集の言語宣言と一致する必要がある。

term フィールドは、一言語用語集や、起点言語と目標言語の役割を区別しなくてよい場合に使用する。

src フィールドは、1 つ以上の起点言語を指定する場合に使用する。

tgt フィールドは、1 つ以上の目標言語を指定する場合に使用する。

例については、以下の表を参照。コロンの後に半角スペースは入れない。

用語ステータスは「6.1.1 行用語ステータスと言語ごと用語ステータス」で詳しく説明している。

例:用語集の種類と翻訳方向の表記法

用語集の種類	翻訳方向表記法の例	詳細
一言語	term:en	英語がこの用語集で使われる唯一の言語である。
二言語	src:en tgt:ja	起点言語が英語で、目標言語が日本語である。
二言語	term:en term:ja	起点／目標の区別が不要の場合。

5.3. 品詞フィールド

5.3.1. 品詞フィールドとそのフィールド値

品詞フィールドは品詞を示す。このフィールドが言語タグを伴わず `pos` のみである場合、品詞は用語集内のすべての言語の用語に適用されると仮定される。さらに、`pos:<言語タグ>` の形式で、特定の言語の品詞を示すこともできる（コロンの後にスペースを入れない）。品詞は省略可能である（UTX 1.11 からの変更点）。

UTX では、以下の品詞フィールド値が定義されている。

品詞フィールド	詳細
noun	名詞
properNoun	固有名詞
verb	動詞
Vt	他動詞
Vi	自動詞
adjective	形容詞
prenominal	名詞前置修飾詞
adverb	副詞
sentence	文

名詞前置修飾詞は、形容詞に似ているが扱いが異なる。たとえば、日本語の連体詞は、名詞前置修飾詞である。いくつかの MT システムでは、形容詞ではなくこの品詞を使用する必要がある。

`adverb` は副詞を指定する。副詞句も含まれる。

5.3.2. sentence と特殊文字

`sentence` フィールド値は、「用語」が文であることを示す特別のフィールド値である。`sentence` は、たとえば、文として現れるユーザーインターフェイスのメッセージに使用する。UTX 用語集を特定システムのためにエクスポートする場合で、そのシステムに TMX 形式を使用する手段がないときは、`sentence` 項目を名詞として扱うことができる。

`sentence` 項目では、いくつかの特殊文字を使用できる。これらは `sentence` 項目でのみ使用される。

文字	表記法
タブ文字	¥t
改行文字	¥n
リテラルエスケープ文字（バックスラッシュ（円記号）そのものを表す場合）	¥¥

5.4. 用語ステータスフィールド

用語ステータスは、空白、provisional、approved、non-standard、forbidden、rejected、あるいは obsolete の 7 つの用語ステータスのうちの一つを示すフィールドである。

用語集管理者およびその代理者のみが、用語ステータスの値を変更できる。

注：用語集に用語ステータスフィールドがない場合、すべての項目は承認済み（approved）とみなされる。

同義語とは、同じ概念（または意味）を持つ用語である。用語ステータスは、同義語や異表記の異なるステータスを示すために使用できる。詳細は、「6.3 概念 ID と用語ステータス」を参照のこと。

5.4.1. provisional（暫定）

用語ステータス「provisional（暫定）」は、用語提出者によりその項目が提案されたが、用語集管理者によってまだ承認されていないことを意味する。暫定ステータスは一時的なものであり、用語管理者は、「承認」などの用語ステータスをすみやかに決定する必要がある。

注：用語集管理者は、その用語を用語集から除外（削除）するか、別の用語集に移動することもできる。

5.4.2. approved（承認）

用語ステータス「approved（承認）」は、その項目が特定用語集（分野）に属することが用語集管理者により承認済みであることを示す。承認ステータスは、文脈が適切な場合は、その用語（訳語）を最優先で使用する必要があることを示す。用語に複数の同義語や異表記がある場合（たとえば"plug-in"と"plugin"）、これらのうちの一つのみを承認ステータスとする必要がある。

承認語は別の承認語とペアになる。複数項目であっても品詞が異なる場合は、異なる用語とみなされる。たとえば、"plot"という語には名詞と動詞があるため、それぞれに承認ステータスを付けられる。

5.4.3. 空白用語ステータス

用語ステータスが空白の場合、承認と同じとみなされる（UTX 1.11 からの変更点）。非標準、禁止、却下、および廃止用語（後述）とペアになる用語の用語ステータスが、空白となっている場合、暗黙の承認ステータスを意味する。

5.4.4. non-standard（非標準）

non-standard（非標準）用語ステータスは、一群の同義語や異表記の中で、優先度が低い用語を示す。

非標準訳語の例：

下記の例では、"plug-in"が"plugin"より優先される。

#src:ja	tgt:en	term status:ja	term status:en
プラグイン	plug-in	approved	approved
プラグイン	plugin		non-standard

5.4.5. forbidden（禁止）

forbidden（禁止）用語ステータスは、その用語（訳語）が使用されてはならないことを示す。用語が禁止され

るのは、訳として不適切というだけでなく、最終結果の文書の文脈でその語が使われることが不適切とみなされるからである。

禁止語は、非標準語と異なり、訳語候補として示されてはならない。

注：UTX 1.11 までは、訳語のみを禁止と示すことができた。UTX 1.20 では、（原語を含め）任意の用語を禁止できる。

5.4.6. rejected (却下)

rejected (却下) 用語ステータスは、その用語を特定の用語集に含めるべきではないことを示す。却下語は、UTX 変換ツールおよび UTX 対応アプリケーションからは無視される。却下語は、記録のために用語集に維持するか、別のリストへ移動するか、後で削除することができる。

5.4.7. obsolete (廃止)

obsolete (廃止) 用語ステータスは、その用語（訳語）が、以前は使用されていたが、現在は使用してはならないことを示す。

5.5. ユーザー定義フィールド

UTX 用語集には、任意の数のユーザー定義フィールドおよびそれらのフィールド項目を追加できる。

言語固有のフィールドについては、その言語を示すために言語フィールドを使用する。

ユーザー定義フィールド:<言語タグ>（コロンの後にスペースを入れない）

例

x-termUsage:en

協会活動報告

(2016 年 2 月～2016 年 4 月)

機械翻訳課題調査委員会

2016 年 2 月 19 日（金）（2015 年度 第 11 回）

① 前回委員会の議事録の確認

② WG 検討

(WG1,2)

- ・ 中日テストセット評価について
- ・ AAMT/Japio 特許翻訳研究会評価部会との連携について
- ・ 言語処理学会の発表について
- ・ その他

(WG3)

- ・ UTX 1.20 と TBX との互換性

UTX-TBX 変換仕様書を作成する目的として何を変換対象としたいのかを明確にする必要がある。

- ・ その他山本委員より

- ・ 学会発表予定 2 件：

3/10 NLP2016 言語処理学会年次大会 於：東北大学

3/24 情報処理学会/ドキュメントコミュニケーション研究会

第 101 回研究会 於：東工大

- ・ ISO/TC37 国内委員会で、UTX を使って ISO/TC37 用語集の日本語訳を付けることを提案。コペンハーゲン国際会議で議題として取り上げられる可能性あり。

- ・ 見学者紹介

日立オートモティブシステムズより 2 人参加。

- ・ 宿題

2016 年 3 月 18 日（金）（2015 年度 第 12 回）

① 前回委員会の議事録の確認

② WG 検討

(WG1,2)

- ・ 言語処理学会発表の報告
- ・ AAMT/Japio 特許翻訳研究会評価部会のテストセット評価の報告について
- ・ 翻訳フェアの企画について

- ・その他

(WG3)

- ・言語処理学会発表の報告

山本委員が NLP2016 仙台で発表した報告を行った。

- ・UTX 1.20 正式版リリースに向けての調整

修正事項の確認と議論を行った。

- ・ISO/TC37 コペンハーゲン会議

山本委員が ISO/TC37 コペンハーゲン会議で、日本から国内委員会に、「既存の i-term 用語集から重要な 300 語の TC37 用語を選定し、各国委員会が各国語での既定訳語を定める」ことを提案。実施について 5 月の WG までに検討し決定する。

- ・その他

- ・宿題

UTX 変換ツールで、UTX の言語タグが小文字でないとエラーになる件の調査。

(秋元委員)

③ 全体会議

- ・各 WG の検討事項共有
- ・言語処理学会の発表報告
- ・委員の交代について
- ・MT Summit の準備
- ・その他

2016 年 4 月 15 日 (金) (2016 年度 第 1 回)

① 前回委員会の議事録の確認

② WG 検討

(WG1,2)

- ・AAMT/Japio テストセットのレビュー

テストセット最終版のレビュー。特許パテントファミリーからテストセット原文を抽出し、著作権について確認することとした。

- ・翻訳フェアの企画について

時間割を決定。講演会講演者候補、招待後援の講演者候補を選定した。

- ・MT Summit

各コミッティメンバーの確認、およびスポンサー候補の列挙を行った。

(WG3)

- ・第 101 回ドキュメントコミュニケーション研究会(情報処理学会)における発表に関して報告があった。
- ・UTX 変換ツールの開発に関して進捗報告があった。

- ・ UTX1.20 と TBX との互換性について議論した。
- ・ ISO 国内・各国委員会への提案を説明した。

③ 全体会議

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

・ インターネットWG

以下の活動をオンラインで実施した。

- ① 総会・AAMT 機械翻訳フェアに向けての情報掲載の準備
- ② AAMT ジャーナル目次や法人会員情報等の通年コンテンツの掲載・更新
- ③ UTX や AMTA 等の協会関連情報の掲載
- ④ 会員専用ページコンテンツの更新体制の立ち上げ
- ⑤ 来年開催の MT Summit のホームページ運用体制の検討
- ⑥ 旧ページから新ページへの移行作業の残務処理

編集委員会

2016 年 3 月 15 日 (2015 年度 第 3 回)

- ① AAMT Journal No.62 の執筆依頼計画について

AAMT/Japio 特許翻訳研究会

2016 年 3 月 11 日(金) (2016 年度 第 7 回)

- ① 議事録の確認
- ② 奈良先端科学技術大学院大学、ニュービック先生
「ニューラル機械翻訳(仮)」講演ご発表
- ③ 27 年度 AAMT/Japio 特許翻訳研究会報告書
- ④ その他
- ⑤ 次回の開催について
 - ・ 開催の日時(場所) 2016 年 5 月 13 日、金曜日 決定
 - ・ 予定
 - ・ 主な議題

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

AAMT ジャーナル 62 号をお送りします。

今号の巻頭言は、AAMT 副会長の飯田仁先生より、御寄稿を頂きました。

AAMT が設立されて今年で 25 周年になります。そこで、今号では、AAMT 設立 25 周年を祝して、AAMT 初代会長の長尾真先生に、「AAMT 設立までのこと」と題しました記念寄稿を頂きました。また、長尾先生からの御寄稿には、記念寄稿とともに、AAMT 設立 25 周年を記念した書も併せて掲載させていただきました。

また、理事の坂本義行先生からは、機械翻訳の歴史に関する貴重な御寄稿の第七弾を頂きました。

今年 2016 年、および、来年 2017 年は、自然言語処理分野、機械翻訳分野、ローカライゼーション分野等、翻訳に関係する学術・産業分野の代表的国際会議・イベントが、相次いで日本で開催される年であります。

まず、今号が会員の皆様のお手元に届く時期に先立ちまして、4 月 13-15 日には、ローカライゼーションの代表的国際イベント LowWorld 30 Tokyo が東京において開催されました。本イベントは、今年が初めての日本開催になります。そこで、機械翻訳課題調査委員会・AAMT ジャーナル編集委員会の目次由美子様より、同イベントの参加報告を御寄稿頂きました。

一方、今年 2016 年 12 月には、計算言語学国際会議(COLING)が、大阪で開催されます(日本開催は 22 年ぶり)。そこで、Local Chair の情報通信研究機構 隅田英一郎先生に、会議開催案内を御寄稿頂きました。また、来年、2017 年 9 月には、第 16 回 MT サミットが、名古屋で開催されます(日本開催は 24 年ぶり)。現在、AAMT を中心としてその準備が進められているところですので、AAMT 会長、名古屋大学 中岩浩巳先生より、会議開催案内を御寄稿頂きました。

また、AAMT ジャーナル編集委員であり、日本翻訳ジャーナル編集長の河野弘毅様からは、『グローバル・コミュニケーション計画』の取組状況』に関して、総務省情報通信国際戦略局 技術政策課 研究推進室長 荻原直彦氏へのインタビュー内容を御寄稿頂きました。

その他、前掲、機械翻訳課題調査委員会・AAMT ジャーナル編集委員会の目次由美子様からは、3 月 24 日に東工大において開催されました、情報処理学会 第 122 回 情報基礎とアクセス技術研究会・第 101 回ドキュメントコミュニケーション研究会合同研究会(テーマ『「多言語ドキュメントと情報技術」および一般』)の参加報告を御寄稿頂きました。

また、製品・サービス紹介として、東芝 インダストリアル ICT ソリューション社のスマートフォン向け音声翻訳アプリ『RECAIUS? 「音声トランスレータ」』、および、日中韓辞典研究所(CJKI)の「多言語翻訳を支える大規模な辞書データベース」をそれぞれ掲載しました。

AAMT 内の活動報告として、機械翻訳課題調査委員会から、「UTX 1.20 β 仕様書（抜粋）」、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 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 2～3 回発刊予定）が送付されます。購読料は年会費に含まれています。

2.機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機械翻訳関連の情報をお届けします。

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

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

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

4.関連機関の主催する国際会議に参加できます。

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

AMTA* : Association for Machine Translation in the Americas

EAMT** : European Association for Machine Translation

年会費は以下の通りです。

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

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

ご関心のある方は、事務局までお問い合わせください。

アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

電子メール：aamt-info@aamt.info

AAMT

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。

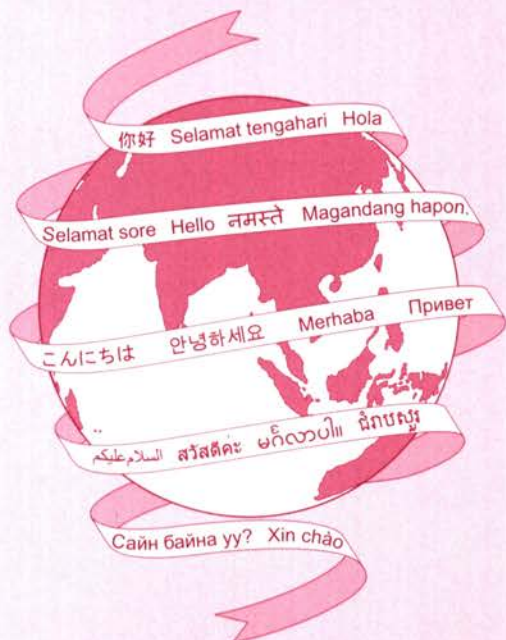
申込日 20 年 月 日

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	希望する 希望しない

コメント

--

AAMT



AAMTジャーナル No.62

発行：アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

住所：〒171-0014 東京都豊島区池袋2-55-2鈴木ビル3階
（株）日本システムアプリケーション内

phone：03-5951-3961 fax：03-5951-3966

編集委員会：宇津呂 武仁 小谷 克則 大倉 清司

阿部 さつき 釜谷 聡史 河野 弘毅 目次 由美子

表紙(図部分)デザイン：阿部 さつき

事務局：神崎 享子 荻野 孝野

印刷所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)

c/o Japan System Application Co., Ltd.

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

aamt-info@aamt.info