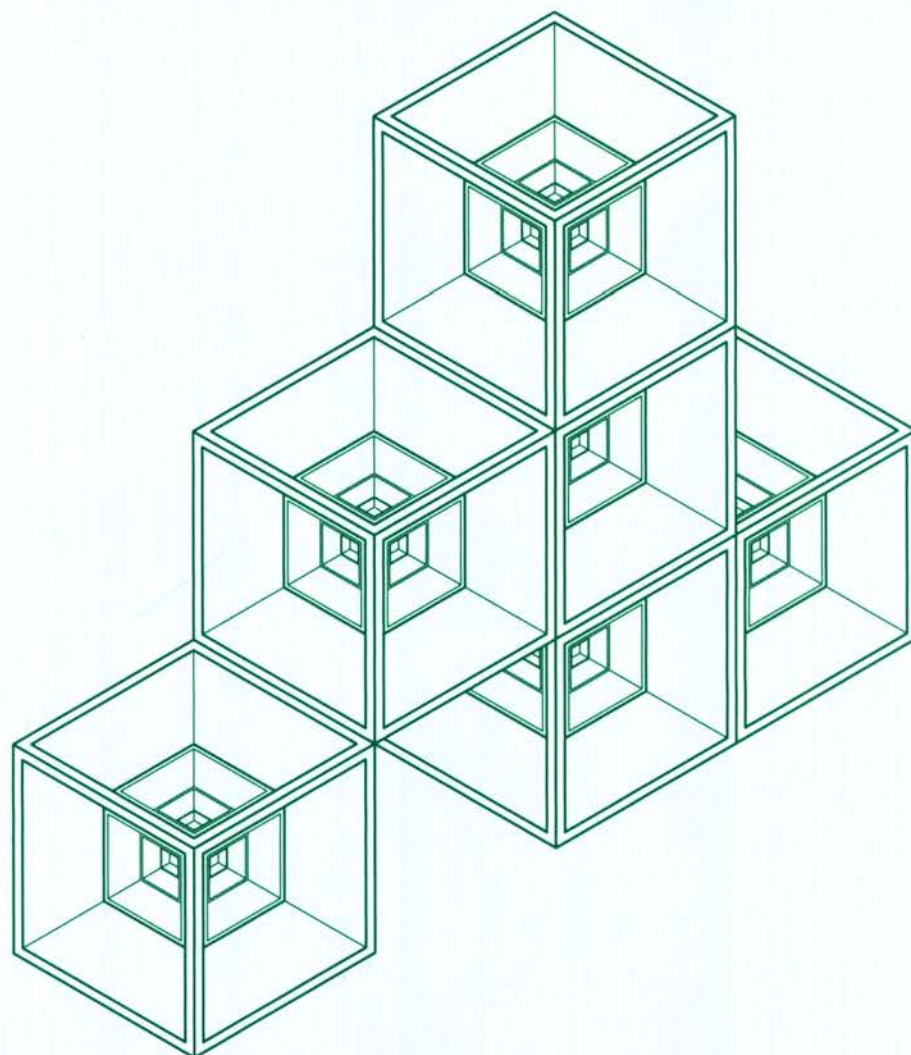


APMT

Asia-Pacific Association for Machine Translation

Journal



June 2013

No.53

アジア太平洋機械翻訳協会

目 次

巻頭言：	機械翻訳の発展と普及に向けて	中岩 浩巳	1
御挨拶：	AAMT 会長を終えて	井佐原 均	2
研究報告：	日本語の主辞後置性を利用した統計的機械翻訳技術	永田 昌明, 須藤 克二, 塚田 元	3
連載：	機械翻訳事始め [1]	坂本 義行	8
レポート：	Translation Technology from a Translator's Perspective	A. Simpson	15
シンポジウム報告：	未来の翻訳研究に関するワークショップ	隅田 英一郎	18
	第 2 回特許シンポジウム報告	横山 晶一	21
委員会活動報告：	AAMT 機械翻訳課題調査委員会		
	JTF 翻訳祭出展報告	MT 委員会	23
	用語集形式 UTX の FAQ(日本語版)	WG3	32
	翻訳祭感想と UTX 関連の記事について	WG3	37
AAMT 会員のひろば：		株式会社川村インターナショナル	40
事務局からのお知らせ：	協会活動報告 (2012 年 9 月～2013 年 5 月)	AAMT 事務局	42
編集後記			46

C O N T E N T

Foreword:	Growing Needs for Machine Translation and its Technological Development	<i>H. Nakaiwa</i>	1
The greetings	Message from ex-President	<i>H. Isahara</i>	2
Report of research	Statistical Machine Translation using Head-Finalness of Japanese <i>M. Nagata, K. Sudoh, and H. Tsukada</i>		3
Series:	History of Machine Translation	<i>Y. Sakamoto</i>	8
Report:	Translation Technology from a Translator's Perspective	<i>A. Simpson</i>	15
Symposium Report:	Workshop on Future Directions in Translation Research	<i>E. Sumita</i>	18
	Report on the Second Symposium on Patent Information Processing	<i>S. Yokoyama</i>	21
Committee Report	Committee for Seeking Future Direction of MT:		
	Exhibition Report of JFT Translation Festival 2012	<i>MT</i>	23
	UTX FAQ	<i>WG3</i>	32
	A Note on UTX articles / A Thought on JTF Translation Festival	<i>WG3</i>	37
AAMT Activities:	AAMT Activities (from September to May, 2013)		40
Editor's Note:		<i>T. Utsuro</i>	46

機械翻訳の発展と普及に向けて

アジア太平洋機械翻訳協会会長

中岩浩巳

2012 年度 AAMT 総会で会長を拝命いたしました NTT の中岩です。

私は、NTT 入社以来、機械翻訳の研究開発に携わってまいりました。また、2011 年に中国アモイで行われました AAMT 主催の MT Summit XIII ではプログラム委員長を務めさせていただきなど、AAMT の活動にも協力させていただきました。今後も、機械翻訳技術の発展と、機械翻訳の有効活用場面の拡大に向けた取組み、ひいてはアジア太平洋における機械翻訳技術を通じた翻訳産業の活性化に向けて、微力ながら取り組んでまいりたいと存じます。

20 世紀末にスタートしたインターネットは、人々のありとあらゆる生活の中に浸透した結果、一般の人が他の言語に触れるきっかけを作りました。また、従来機械翻訳で扱う言語は、英語と母国語の言語間が主流でしたが、最近の統計によると、Web ページは英語以外の言語のページと英語のページがほぼ同数になったとの報告もあるとおり、英語以外の言語との翻訳に対する需要が急速に増えております。twitter や Facebook など新たな CGM が、従来の言語によるコミュニケーションを超えた場面にも、機械翻訳利用の可能性を広げました。クラウドコンピューティングの登場により、サービスや製造など様々な営みにおける国の壁がなくなりつつあります。その結果、言語を超えたコミュニケーションやサービスの多言語化等の需要が急速に増加しております。

機械翻訳を技術的観点で考えてみますと、従来の主流であったルールベース型機械翻訳技術から、大量の対訳データを活用した統計ベース型機械翻訳に移りつつあります。また、当協会に強く関連する

日本語と英語の機械翻訳に関しましても、英日翻訳においてルールベース型に匹敵・凌駕する技術が考案されております。このように機械翻訳技術自体も大きな変革期にあると言えるのではないのでしょうか。

以上の通り、機械翻訳への期待は今まで以上にますます高まっております。

アジア太平洋地域のさらなる活性化のために貢献できる機械翻訳協会となることを目指して、国の壁を越えた活動をさらに強化していきたいと考えます。そのためには、会員の皆様の本会活動への積極的な参画と、ご支援が必要となりますので、今まで同様、AAMT をよろしくお願いいたします。

AAMT 会長を終えて

豊橋技術科学大学

井佐原 均

辻井潤一元会長から会長の任を受け、昨年の総会で中岩浩巳現会長に引き継ぐまで、6年間、アジア太平洋機械翻訳協会の会長を務めさせていただきました。

会長就任にあたり、アジア太平洋圏を中心とする海外との連携強化を目標の一つとしましたが、中国での機械翻訳サミットの開催、韓国機械翻訳協会との連携、ヨーロッパ機械翻訳協会・アメリカ機械翻訳協会の会議に参加しての交流など、一定の成果を上げることができたと考えております。会長就任前に担当しておりましたウェブやジャーナルにつきましても、新しいご担当の下、充実した内容になっておりますし、機械翻訳にかかわる諸問題を検討する課題調査委員会も委員長はじめ委員の方々のご尽力で活発な活動をしていただいております。アジア太平洋機械翻訳協会の初代会長の長尾真先生からのご寄付を得て設けた AAMT 長尾賞は機械翻訳の実用化に対する貢献に与えられる賞として、いわゆる論文賞とは異なる独自の位置づけを得ております。協会活動を支援する事務局につきましては、従来は他組織の中に事務局を設置しておりましたが、経費削減を目指し、会長の所属元での事務局運営といたしました。ただ、今後の安定した事務局運営のためには会長に依存しない体制が良いと考え、現在は外部委託による事務局運営となっております。

機械翻訳を取り巻く環境は私が会長をしていた6年の間に大きく変わりました。機械翻訳システムの性能向上や、前編集・後編集や用語辞書の整備といった支援環境の向上により、企業内文書を中心に翻訳効率向上の手段として機械翻訳が活用されるようになってきました。しかしながら、機械翻訳に

関する一般の理解はまだまだ不十分と言えましょう。内外の関連団体の行事で機械翻訳の活用についての啓蒙活動が行われております。機械翻訳システムを、どのような条件の下で、どのような対象に対して用いれば、どれだけの効率向上が図れるか、といった情報を的確に将来のユーザに向けて発信していくことが AAMT の使命の一つであると考えます。機械翻訳システムの利用者、製造販売者、研究開発者をメンバーとする本協会の特徴を活かした啓蒙活動が可能でしょう。このような活動を通して、新しい会員が増加すれば、AAMT がますます活性化することと思います。

中岩浩巳新会長の下、私自身も今後とも微力ながら AAMT の活動に協力させていただこうと考えております。皆様の変わらぬご支援をよろしくお願いいたします。

日本語の主辞後置性を利用した統計的機械翻訳技術

永田昌明, 須藤克仁, 塚田元

NTT コミュニケーション科学基礎研究所

1 はじめに

英語と日本語は語順がほぼ鏡像関係にあり、恐らく世界で最も翻訳が難しい言語対の一つである。我々は、主辞後置性という日本語の言語学的な特徴を利用して、英語の文の単語を日本語と同じ語順に並べ替えてから日本語へ翻訳する方法を考案し、英日翻訳の精度を劇的に改善した。本稿では、この日本語の主辞後置性を利用した事前並べ替え翻訳を中心に、近年の NTT における統計的機械翻訳の研究の概要を紹介する。

2 統計的機械翻訳の歩み

統計的機械翻訳(statistical machine translation, 以下、統計翻訳)は、対訳データから翻訳規則や対訳辞書に相当する統計モデルを自動的に学習し、新しい言語対や特定の分野の翻訳システムを短期間に低コストで作成することを目指す技術である。

統計翻訳は 1990 年前後に英語とフランス語の翻訳を対象として研究が始まった。2000 年前後から翻訳の基本単位を単語から句に拡張する「句に基づく翻訳」(phrase-based translation)の研究が進み、英語とフランス語のような語順が近い言語対の翻訳では実用レベルに達した。

2005 年前後からは自然言語の階層構造や構文理論を利用する「木に基づく翻訳」(tree-based translation)の研究が進み、英語と日本語ほどではないが、やはり語順が大きく異なることが問題である英語と中国語の翻訳では、ルールベース翻訳より統計翻訳の方が精度が高いことが確実になった。

NTT では、語順の違いを克服する方法として、句に基づく翻訳のための大局的並べ替えモデル[1]や階層的な句に基づく翻訳[2]の研究に取り組み、一定の精度向上を達成した。しかし、英語と中国語

よりさらに語順が大きく異なる英語と日本語の翻訳では、NTT を含むどの研究機関も人手評価でルールベース翻訳の精度を上回ることができなかった。

3 事前並べ替え翻訳

原言語(翻訳元)の文の単語を目的言語(翻訳先)の語順に並べ替えてから句に基づく翻訳を行う「事前並べ替え」(preordering)の研究は 2000 年代前半から存在したが、2010 年前後に Google, Microsoft, IBM, NTT などの研究機関がこの有効性に注目し始め、本稿執筆時点では日英間の語順の違いを克服する技術の主流になりつつある。

事前並べ替えでは、原言語の文を構文解析して得られた構文構造に対して「並べ替え規則」(reordering rule)を適用して目的言語の語順へ変換する。一般に並べ替え規則は、単語対応付きの対訳データから自動学習する方法と人手で作成する方法がある。

NTT は、主辞後置性と呼ばれる日本語の言語学的な特徴に着目し、「主辞を後置する」(head finalization, 主辞後置化)という(少なくとも原理的には)唯一の規則を使って英語の単語を日本語の語順に並べ替える方法を考案し、英日翻訳の精度を劇的に改善した[3]。

2011 年開催された評価型ワークショップである NTCIR-9 特許翻訳タスクの英日翻訳において NTT と東京大学の共同チームは高精度な英語構文解析器 Enju と主辞後置化に基づく事前並べ替えを組み合わせることによりトップの成績を収めた[4,5]。人手による英日翻訳の精度評価で、統計翻訳がルールベース翻訳を上回ったのはこれが史上初めてである。

4 主辞後置化に基づく事前並べ替え

主辞後置性(head final)とは、係り先(主辞)の単語が係り元の単語より必ず後ろに出現するという性質である。中学の国語の時間に「文節係り受け」を習う日本人は、「必ず前から後ろへ係る」ことを当然の事実として受け入れているが、世界的にみると日本語ほど首尾一貫した主辞後置性を持つ言語は珍しい。

例えば図 1 に示す“He saw a cat with a long tail”という英語の文において、動詞 saw に対して主語 he は前から係り、目的語 cat は後ろから係る。”long tail”のように形容詞は名詞に前から係り、”cat with ...”のように前置詞句は名詞に後ろから係る。

主辞後置化とは、このような英語の文の係り受け構造において「主辞を後置する」すなわち係り受けが「必ず前から後ろへ」係るように単語を並べ替えることである。

主辞後置化した英語の単語列 “He long tail with cat saw” と、対訳となる日本語の単語列「彼は長い尻尾の猫を見た」を比べると逐語訳(語順が同じでほぼ 1 対 1 対応)になっている。以下、主辞後置化した英語を主辞後置英語(Head Final English)と呼ぶことにすると、主辞後置英語から日本語への翻訳は従来の統計翻訳技術で高精度に翻訳することが可能である。

5 事後並べ替え翻訳

日本語から英語への事前並べ替え翻訳は、英語から日本語への事前並べ替え翻訳に比べて本質的に難しい。図 1 を見れば分かるように、「前から後ろからも係る」構造から「必ず前から後ろへ係る」構造への変換は主辞後置化という一つの原則で実現できるが、この逆変換を実現するには「主語ならば動詞の前に」「目的語ならば動詞の後ろに」といった場合分けが必要になる。

ところが CaboCha のような既存の日本語係り受け解析器は文節間の係り受け構造しか出力しない。述語に対する名詞の文法役割や意味役割を決定す

る述語項構造解析はまだ実用レベルにないので、上記の場合分けに必要な文法役割の情報を並べ替え規則に使えないという問題がある。一言でまとめると、日本語解析技術が未熟なために、現時点では日英翻訳における事前並び替えの実現は難しい。

そこで我々は、日本語から英語の翻訳において、日本語の文を主辞後置英語の文に翻訳し、主辞後置英語の文の単語を並び替えて英語の文を生成する「事後並べ替え」(post ordering)を考案した[6]。後段の主辞後置英語から英語への単語並べ替えは自明ではない。しかし、英語の構文解析器が出力する英語の構文構造へ主辞後置化を適用することにより、英語の文と主辞後置英語の文の対を大量かつ高精度に自動生成できるので、これを対訳コーパスとみなして統計的機械翻訳の各種の手法を適用することにより主辞後置英語から英語への単語の並び替えを実現する。この手法は従来手法の最高精度と同程度の精度を半分以下の計算量で実現できる。

図 2 に示すように、NTT が考案した英語から日本語への事前並べ替え翻訳および日本語から英語への事後並べ替え翻訳では、主辞後置性という日本語の言語学的特徴を利用して定義した「主辞後置英語」という中間言語を媒介として、言語翻訳という問題を、訳語選択(単調翻訳)と語順選択(並び替え)という二つの部分問題に分解している。その結果、英日翻訳では大幅な精度向上を実現し、日英翻訳では計算時間を大きく短縮できた。

6 順位相関に基づく機械翻訳の自動評価

機械翻訳の精度評価は非常に難しい問題であり、自動評価尺度 BLEU の誕生は機械翻訳の研究に革命的な変化をもたらした。しかし、欧米言語間の翻訳において最も標準的な機械翻訳の自動評価尺度である BLEU は、日本語と英語の間の翻訳では必ずしも信頼できない(人手による評価と一致しない)ことが知られていた。

BLEU (Bilingual Evaluation Understudy) は、機械による翻訳はプロの翻訳者による翻訳(参照訳、

reference)に類似しているほどよいと考え、類似度を0から1の間の数値で表す。具体的には、機械翻訳の翻訳結果と参照訳の間の長さ1から4までの単語 ngram の適合率の幾何平均に、短い文へのペナルティを掛けたものである。日本語と英語の間の翻訳では4単語を超える大きな並び替えが起きることが、この言語対を BLEU がうまく評価できない原因と考えられる。

そこで NTT では、翻訳結果と参照訳の語順の一致度を重視する自動評価尺度 RIBES (Rank-based Intuitive Bilingual Evaluation Measure) を考案した [7]。

RIBES は、機械翻訳の翻訳結果と参照訳の単語対応を求め、対応する単語の出現順序に関するケンドール順位相関を0から1に正規化したものに単語適合率を掛けたものである。RIBES は、3節で紹介した評価型ワークショップ NTCIR-9 特許翻訳タスクにおいて自動評価尺度の一つとして採用され、主催者の評価により、英日・日英・中英翻訳タスクにおいて人間による評価との相関が BLEU より高いことが示された [4]。なお RIBES を計算するプログラムは <http://www.kecl.ntt.co.jp/icl/lirg/ribes/index.html> からダウンロードできる。

7 今後の課題

英語から日本語への統計翻訳は、主辞後置化に基づく事前並べ替えによりほぼ実用レベルに到達した。しかし日本語から英語への統計翻訳は、まだ発展途上である。

日本語の解析に関して、5節で述べた述語項構造解析に加え、特に話し言葉翻訳において省略照応解析が非常に重要である [8]。また英語の生成に関して、主語や目的語を語順で表現する英語では適切な構文構造の選択が重要と考えている [9]。さらに RIBES のような語順を重視する自動評価尺度は識別学習による最適化が難しいので、最適化法が確立している BLEU と併用し、複数の目的関数を同時に最適化することを検討している [10]。

8 おわりに

本稿で紹介した日本語の主辞後置性を利用した統計翻訳技術は、原理的には英語以外の言語へも適用可能である。今後は、中日翻訳 [11] を手始めとして、日本語を中心とする多言語翻訳へ取り組む予定である。

- [1] Masaaki Nagata, Kuniko Saito, Kazuhide Yamamoto and Kazuteru Ohashi. A Clustered Global Phrase Reordering Model for Statistical Machine Translation, COLING-ACL-2006, pp.713-720, 2006.
- [2] Taro Watanabe, Hajime Tsukada and Hideki Isozaki. Left-to-Right Target Generation for Hierarchical Phrase-based Translation. COLING-ACL 2006, pp.777-784, 2006
- [3] Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. HPSG-based Preprocessing for English-to-Japanese Translation, ACM TALIP. Vol.11, no.3., 2012.
- [4] Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita, and Benjamin K. Tsou. Overview of the Patent Machine Translation Task at the NTCIR-9 Workshop. NTCIR-9, pp.559-578, 2011.
- [5] Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, Masaaki Nagata, Xianchao Wu, Takuya Matsuzaki, and Jun'ichi Tsujii. NTT-UT Statistical Machine Translation in NTCIR-9 PatentMT. NTCIR-9, pp.585-592, 2011.
- [6] Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada and Masaaki Nagata: Post-ordering in Statistical Machine Translation, MT Summit XIII, 2011
- [7] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, Hajime Tsukada: Automatic Evaluation of Translation Quality for Distant Language Pairs, EMNLP-2010, 2010
- [8] Hirotoshi Taira, Katsuhito Sudoh and Masaaki Nagata: Zero Pronoun Resolution can Improve the Quality of J-E Translation, SSST-2012, 2012
- [9] Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada and Masaaki Nagata: A Comparative Study of Target Dependency Structures for Statistical Machine Translation, ACL-2012, 2012
- [10] Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada and Masaaki Nagata: Learning to Translate with Multiple Objectives, ACL-2012, 2012
- [11] Dan Han, Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada and Masaaki Nagata: Head Finalization Reordering for Chinese-to-Japanese Machine Translation, SSST-2012, 2012

図1 日本語の主辞後置性に基づく
事前並べ替え翻訳

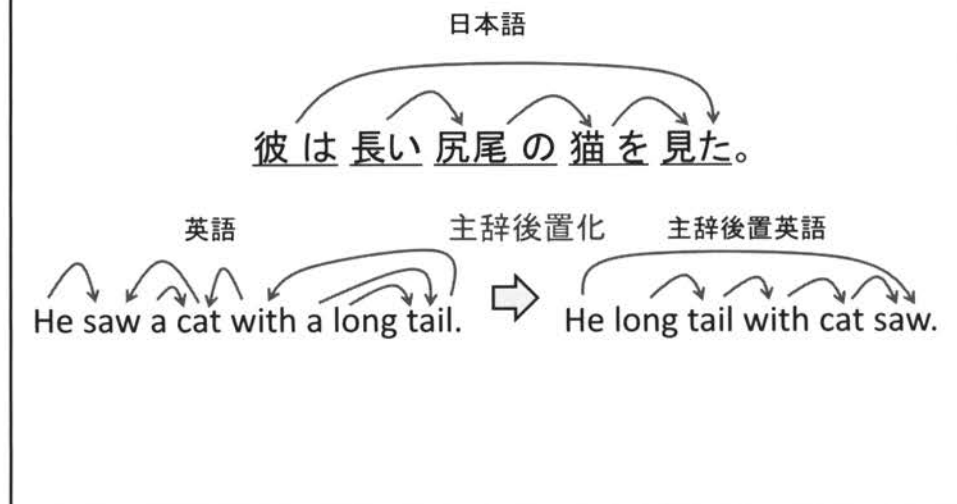
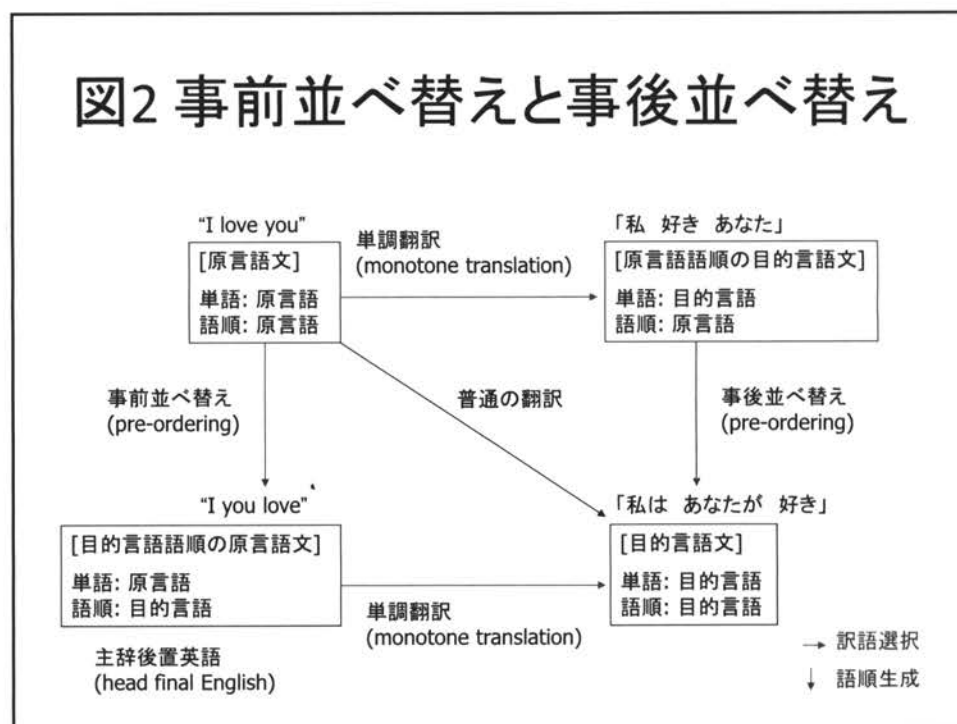


図2 事前並べ替えと事後並べ替え



機械翻訳事始め [I]

アジア太平洋機械翻訳協会理事

坂本義行

ことばの発生

ことばについて、「翻訳機械」¹⁾の訳者まえがきによると「旧約聖書によれば、創世のむかしは人類の言語は一つであった。それが現在のようにいろいろの民族が自分たちで通用する言語を話し、異民族相互間の意志の疎通をかくようになったのは、天までとどくことを意図して建設をはじめたいわゆるバベルの塔が神の怒りにふれたためであるという。

各民族が独自の言語を持って活動している現実の事態は、それぞれに固有の文化を築いていけるといふ大きな利点がある反面、相互の思想、意志の交流疎通という点で実用上多大の不便をとまなう。この障害を軽減する手段として、だれにでも思いつく方法は即応性のある組織的な大規模の“翻訳”機関を常備することであるが、人間の翻訳者に依存する限り、このようなものの実現可能性はないと考えられる。したがって、翻訳者をわずらわすことなく、翻訳を高速でやっつけのける機械というものは、われわれの空想の中に生ずる一つの大きな夢であった。

この企てを具体的に示した最初のものはペ・ペ・トロヤンスキーが 1923 年に出した翻訳機械の特許であろう。」と述べられている。

機械式計算機

計算機といえば機械式計算機があり、その代表としてパスカルの機械式計算機があり、これについては、ウィキペディアによると²⁾、ブレーズ・パスカルが 1645 年に発明した機械式計算機。Pascaline (パスカリーヌ) または Machine Arithmétique と呼ばれている。

パスカルは 1642 年、彼が 19 歳のときから計算機について研究を始めている。徴税官だった父の手伝いをしていたパスカルは、仕事を減らすための道具

を作ろうと考えた。1642 年までにパスカルは 50 台もの試作機を作ったが、売れたのは 1 ダース強である。高価であったことと複雑であったこと（加えて加減算しかできず、減算のやり方が難しかった）が原因で、それ以上売れることはなく、パスカルは 1652 年に計算機の製作を止めた。

Pascaline は十進数を使った機械である。しかし、当時のフランスの通貨は十進数ではなく、イギリスのポンド、シリング、ペンスと似ていた。従って、金額を計算するのに Pascaline を使おうとすると、計算結果を更に変換する必要があった。1979 年フランスはメートル法に切り替えた。このとき、パスカルの基本設計に触発された職人が登場したが、彼らも商業的には成功しなかった。

最初の Pascaline は 5 個のダイヤルがあり、後には 6 ダイヤルや 8 ダイヤルのものが作られている。最大のもので 9,999,999 までの数値を扱うことが出来た。各ダイヤルは数値のうちの 1 桁に対応し、計算結果は上部の窓に表示される。歯車は一方向にしか回らないため、負の値を直接計算することはできない。減算をするには 9 の補数表現にして加算する必要があった。ユーザーを助けるため、9 の補数を入力すると上部の窓には元の数値が表示されるようになっている。



図 1 Pascaline

トロヤンスキーの発明特許証書

ペ・ペ・トロヤンスキーの発明特許証書については、玉木英彦・気安善市編「自動翻訳」³⁾によると、

「以下のような特許が出願されているが、当時それを実現することはうまくいかなかった。これは、この目的に適した自動装置がまったくつくれていなかった、当然のことである。」と書かれている。

また、文献1)によると、

「この翻訳機は、同時にいくつかの言語を翻訳し、遠方まで送ることを可能にするように見えた。しかしトロヤンスキーは1939年に依頼したソビエトの言語学者たちの支持も得られなかったし、1944年には科学アカデミーの自動化と遠隔制御機構学会の支持も得られなかった。」と書かれている。

発明特許証書

発明の名称 一つの言語から他の一つ、または同時に多数の言語への翻訳に際し、単語を選別しかつ印刷する機械

特許出願人 ペ・ペ・トロヤンスキー

出願年月日 1933年9月5日



図2 発明特許証書

暗号解読から計算機の技術へ

第2次大戦中に発展した暗号解読の技術(文献1では)を基礎として、電子計算機が急速に実用化され出した1946年頃から研究が始められた。もちろん当初W.ウィーバーが考えていたような《翻訳とは、外国語という暗号で書かれた文章を母国語に解読することである》という単純な考え方だけでは通

用しないようになっている。

英国人A.D.ブースとロックフェラー財団のワレン・ウィーバーとが同時に翻訳問題に手を付けたのは1946年のことであった。第二次大戦中に用いられた暗号解読技術をこれに適用しようとウィーバーが提案したのに対し、ブースはデジタル計算機が十分な数のデータを蓄えることができるので、《辞書と首っ引きで》行えるような、単語を単語に移す型の逐語訳を実行する能力があることを強調した。文章論や語順やあるテキストのすべての単語の機械的な翻訳についてすら問題にならなかった。というのは当時の考え方は外国語のテキストを理解するために、学者の目の前に翻訳された主要な単語の系列を置けば、この題目に関する読者の深い知識により彼が手探りで自分の道を見つける役に立つであろうということであった。

プリンストンにおいてブースはブリテンとともに、計算機がその記憶装置中に記録された辞書を引き、かつ一本の穿孔テープで入力に導かれる一連の単語の逐語訳を与えることを可能にする《命令》について研究した。

1948年に英国人リチェンズは、語尾に自動文法解析の考えを主張した。このことは単に読者に単語の文法的役割を説明することで自動翻訳を改良するのみならず、電子辞書における探索の速度を高めるというのは、理論的にみて見出し語の全数を減少させるからである。

1949年ウィーバーは言語ごとに見かけ上の相違はあるが、暗号術で認められ、かつ情報理論により認められている統計的不変量や、中国語学者E.ライフらが歴史的つながりのない言語のあいだに認めた意義的不変量や、さらにライヘンバッハにより指摘された論理的不変量が存在することを強調した。こういった不変量は、人間の頭脳的基本性質の存在と、言語に共通な心理社会的な起源とに相応する、ブースとリチェンズとの仕事を参照して、ウィーバーは、単語に単語を当てる純粹の逐語訳はその性格

上科学技術の研究にとって大きな役割を果たすであろうと主張した。しかも彼はずっと先に進み、直接の文脈を調べることで、意義のあいまいさを解決することを強く主張した。

言語の論理的要素は、計算機の論理回路で処理できると断言した。すなわちシャノンの情報理論は翻訳の問題に関し、統計的組織面で光を投げかけるのであって、彼はこの理論に照らして統計的語義論の研究を企てることを特にすすめた。おわりに彼は通信の道具である言語自体の性質に関する研究の問題を生彩のあるやりかたで提出した。

研究はいくらか分散しながら多方面に分散していった。オスワルドとフレッチャーはドイツ語の文章論構成の機械的解法を研究した。マサチューセッツ工科大学 (M.I.T.) では、イスラエルの論理学者バー・ヒレルが機械翻訳の目的で言語の研究に没頭した最初の人である。1952 年の初めに、ロックフェラー財団は M.I.T. 主催の翻訳問題に専念している言語学者と電子工学者との会議に財政的援助を行った。18 人の参加者は次の二つの段階を熟考することに同意した。まず第一に、科学的題目を取り扱うテキストについて単語の頻度の研究、異なる言語間の等価単語の研究、電子記憶装置を用いる方法についての調査ならびに自動辞書のその他の技術的様相の調査を同時に行うことであった。その後で機械翻訳をするプログラムを作る目的での文章論解析に移ることができよう。文法的小および文章論的な諸問題を解決するために必要な回路の研究は、当分の間さきのこととしてすぐに手を付けなくてもよい。多言語機械翻訳をつくることは、ある一つの言語 A からもうひとつの言語 B へ行いう一方の翻訳の経験から得られる結果をみてからすればよい。しかしながら、媒介言語、いわゆる《機械語》すなわち最終的にはあらゆる言語のあいだの回転板として役立ちうるものの可能性は除外されなかったのである。

1954 年

前述の見解に刺激されて米国の研究が発展した。記憶、容量、語彙を制限することの効用、意味ならびに語尾の機械的確認などの研究がなされた。

1954 年にジョージタウン大学のドスタートとガルビンとは、I.B.M. のシェリンダンと協力して 250 個の単語からなる辞書と、6 個の文章論の法則を用い、IBM701 電子計算機を使ってロシア語から英語への自動翻訳の仕事を首尾よく成し遂げた。

電子計算機 ВЭСМ³⁾

1955 年の末に電子計算機で行った自動翻訳の最初の実験では、この実験用に専門的な数学のテキストの翻訳を考えて、あらかじめ 952 の英単語と 1073 のロシア単語を選んで辞書を作り、数学のテキストの翻訳のために使用されうるプログラムを作った。

INTRODUCTION

When a practical problem in science or technology permits mathematical formulation, the chances are rather good that it leads to one or more differential equations. This is true certainly of the vast category of problems associated with force and motion. So that whether we want to know the future path of Jupiter is the heavens or the path of an electron in an electron microscope we resort to differential equations. The same is true for the study of phenomena in continuous media, propagation of waves, flow of heat, diffusion, static or dynamic electricity, etc., except that we here deal with partial differential equations.

図 3 にミルヌの本の「序文」の最初の一節

ВВЕДЕНИЕ

ЕСЛИ ПРАКТИЧЕСКАЯ ЗАДАЧА В НАУКЕ ИЛИ ТЕХНИКЕ ДОПУСКАЕТ МАТЕМАТИЧЕСКУЮ ФОРМУЛИРОВКУ, АНСИ ДОВОЛЬНО ВЕЛИКИ ЧТО ЭТО ПРИВОДИТ К ОДНОМУ ИЛИ БОЛЕЕ ДИФФЕРЕНЦИАЛЬНЫМ УРАВНЕНИЯМ, ЭТО ВЕРНО БЕЗУСЛОВНО ДЛЯ ОБЖИРНОЙ КАТЕГОРИИ ЗАДАЧ СВЯЗАННЫХ С СИЛОЙ И ДВИЖЕНИЕМ, ТАК ЧТО ХОТИМ ЛИ МЫ ЭНАТЬ БУДУЩАЯ ПУТЬ ЮПИТЕРА В НЕБЕСАХ ИЛИ ПУТЬ ЭЛЕКТРОНА Э ЭЛЕКТРОННОМ МИКРОСКОПЕ МЫ ПРИБЕГАЕМ К ДИФФЕРЕНЦИАЛЬНЫМ УРАВНЕНИЯМ. ТО ЖЕ ВЕРНО ДЛЯ ИЗУЧЕНИЯ ЯВЛЕНИЙ В НЕПРЕРМВНОЙ СРЕДЕ, РАСПРОСТРАНЕНИЯ ВОЛН, ПОТОКА ТЕПЛА, ДИФФУЗИИ. СТАТИЧЕСКОГО ИЛИ ДИНАМИЧЕСКОГО ЭЛЕКТРИЧЕСТВА, И Т Д. ЗА ИСКЛЮЧЕНИЕМ ТОГО ЧТО ИЫ ЭДЕСЬ БУДЕМ РАССМАТРИВАТЬ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ В ЧАСТН ЪХПРОИЗВОДНЫХ.

図 4 にそのロシア語訳をタイプで印刷した形

プログラムは實際上辞書には関係なく適用できるものである。翻訳のためにミルヌの《微分方程式の数値解》から断章を選んだ。

図3に入力として「序文」の最初の一節の英文テキストを、図4に出力として語訳をタイプで印刷した形で示した。

最初の機械翻訳に関する国際会議¹⁾

1956年10月にM.I.T.は、機械翻訳に関する最初の国際会議を招集した。約30人の英国、カナダ、米国の専門家達が集まり、ソ連科学アカデミーのD.J.パノフ博士は手紙でソビエトの研究についての重要な報告を行った。三つの主要な活動の中心は、米国、英国、並びにソ連であり、これらのものより規模の小さいグループはイタリアとスカンジナビアであった。問題はもはや翻訳機械が可能であるということによって確かめるということではなく、努力を集中して研究能率を有効にし、ばらばらに行われている仕事を総合的見地で研究を組織化することにあった。

1957年5月にモスクワで科学技術会議が開かれ、56論文が提出され、それらの中には、ドイツ語、英語、ハンガリー語、中国語からの自動翻訳に関するものや、フランス語からの実験的翻訳に関するものがあり、さらに主として数学的な諸方法を利用して、この学問を精密科学にするようなやり方で言語の研究を方向付ける必要性を立証した。

1958年5月に、自動翻訳についての会議が開かれ、その結果71の提出論文の要約が刊行された¹⁾。

1958年頃の機械翻訳の研究¹⁾

1958年には米国で、1ダースばかりのグループが大小規模の活動的な研究がなされていた。

米国では、

*ハーバード大学では、エティンガーが露英自動辞書に関する研究を行っていた。

*M.I.T.では、ロック、ユングブ、コムスキー等が文章構造、ドイツ語の文章論、方法論を研究していた。

*ジョージタウン大学では、ドスタートとガルビンが

ロシア語の文章論と語彙論の研究を行っていた。

*ミシガン大学では、クートスダとコーフヘージがロシア語ならびに多意義語の問題を研究していた。

*ワシントン州立大学では、ライフラー、マイクルセン、ヒルがロスアンゼルス国際テレメータ会社と共同で、ロシア語とドイツ語の研究をしていた。

*ランド会社では、ヘイとエドマンソンとが方法論に関する興味ある研究に従事している。

ワシントンの国立科学財団は、これらの研究に財政的援助を与え、これらの研究の調整に努力を払った。

英国では、

*ブース、ブランドウッド、クリーブがスフィールド財団の援助を受けて、パークベック単科大学で機械翻訳の方法論を研究し、さらに点字やフランス語やドイツ語についての仕事、ならびに翻訳機械の特性と原価に関する仕事を追及した。彼らは1958年に‘Mechanical Resolution of Linguistic Problems’⁴⁾という著作を発行したが、これは言語データとアイデアとの豊かな著作であった。

*ケンブリッジ言語研究グループでは、マスターマンが指導して、辞書編集と普遍文法論との研究を進め、辞書と文章論とに《機械辞書的熟語辞書》すなわち理念で分類した機械辞書の考えを適用することを探究した。

ソ連では、

1956年9-10月より刊行された『言語学の諸問題』は自動翻訳の問題に定期的な題目を提供している。

*ソ連科学アカデミーの研究チームは、言語学者と電子工学者との間の協力についての指導原理と一般的な方法論との輪郭を明らかにした。

*メルチュウクとクラギナ女史とは、フランス語からロシア語への翻訳に関する非常に興味深い研究を公表した。

*コロレーフは辞書構成のための符号圧縮の問題について研究した。

*ラズモフスキーは翻訳機械の自動プログラミングについて研究した。

*モロシユナヤ女史はイエスペルセンとフリーズとの構造言語学上の仕事を種々の言語間の文章論的差異の解決に使用する方法を示した。

日本では、

1950 年代に電気試験所機械翻訳「ヤマト」や九州大学の KT-1 が開発された。

機械翻訳機「ヤマト」⁵⁾

電気試験所（現産業技術総合研究所）では、1958 年 1 月より和田弘（当時電子部・部長）の指示により、高橋茂を中心に、トランジスタ電子計算機 ETL Mark IV の基本論理回路を使用して、英日翻訳専用機「ヤマト」を試作することが計画された。翻訳のための辞書を収容できる大容量記憶を持った電子計算機が手許になかったため、専用機として 600 万円の予算で製作された。

装置の製作は 1958 年 11 月に終了した。

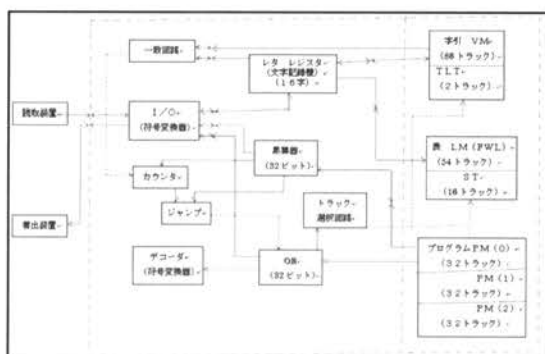


図3 「ヤマト」の計算機ブロック図

基本設計

辞書: 英単語 3000 語 150,000bit=3,000 語×6 字×8bit
 和単語 3000 語 500,000bit
 熟語 300 語 20,000bit
 その他 350,000bit

文字: 英文; ABC, abc, 記号、数字

和文; カタカナ、ABC、記号、数字

プログラム: 機械語 4,000 語

Add Numeric, Sub Numeric, Add Chr,

Read in to AC, Store Word in VM,

Write Mark. Atop

プログラムメモリ 2ブロック 8,192 語

1 語 32 ビット、1 トラック=128 語、

1 ブロック=32 トラック=4,096 語、

計算機の構成

本体: 基本回路の 3 種のプラグイン約 1,000 枚が
 用いられた

記憶装置: 磁気ドラム（電気洗濯機位の大きさ）

回転数 約 3,000r.p.m（平均待ち合わせ

時間は約 10ms）の低速型磁気ドラム 1

台、

1 周 4,096bit、

記憶容量 約 82 万 bit=820,000bit=約
 100kB

回路: トランジスタ 500 個、ダイオード 700 個
 で構成

入出力装置:

1. 機械式テープ読取機（8 単位） 10 字/秒
2. プリンタ 8 字/秒
 ABC、カタカナ、記号、数字
3. パンチャ 8 字/秒
 ABC、abc、記号、数字
4. 万能入出力装置 8 字/秒
 カタカナ、記号、数字

このうち、3 と 4 は計算機と独立で働かすことも
 できる。

蓼沼良一を中心に、英文和訳の研究に着手し、研
 究の課題として、

1. 翻訳に適した機械装置を研究すること
2. この装置を用いて翻訳させること
 であると記されている。

1959 年 2 月には「I like music」とパンチした紙テ
 ープを入力すると、「ワレガ オンガクヲ コノム」
 とプリンタに出力することができるようになった。

本研究の成果については、1959 年 6 月にパリで
 開催された第 1 回国際情報処理会議で和田氏によ
 って発表された。

入力の英文	出力できた和文
1. This is a pen.	1. コレガ ペン ダ
2. Is this a pen?	2. コレガ ペン ダカ
3. Yes, it is a pen.	3. ハイ ソレガ ペン ダ
4. No, it is not a pen.	4. イイエ ソレガ ペン デナイ
5. What is this?	5. コレガ ナニ ダカ
6. This is my watch.	6. コレガ ワシノ トケイダ
7. It is a blue pencil.	7. ソレガ アオイ エンピツ ダ
8. Is that a window, too?	8. ソノウエ アレガ マド ダカ
9. What is it, then?	9. ソレカラ ソレガ ナニ ダカ
10. I have a pen.	10. ワレガ ペンヲ モツ
11. Have I a pen?	11. ワレガ ペンヲ モツカ
12. What have I?	12. ワレガ ナニヲ モツカ
13. I have two arms.	13. ワレガ 2 ウデヲ モツ
14. I have a camera in my hand.	14. ワレガ カメラヲ ワレ テ ノナカニ モツ
15. What have you in your hand?	15. ナレガ ナニヲ ナレノ テ ナカニ モツカ
16. But I have no brother.	16. シカシ ワレガ ナイ アニヲ モツ
17. Who is that gentleman?	17. ソノ シンシガ ダレ ダカ
18. He is Mr. Smith.	18. カレガ ミスター SMITH ダ
19. He is Betty's father.	19. カレガ BETTY ノ チチ ダ
20. He has some eggs.	20. カレガ イクラカノ タマゴヲ モツ
21. Is that girl your sister?	21. アノ オンナノコガ ナレノ アネ ダカ
22. The rooster is black.	22. オンドリガ クロイ ダ
23. There is a desk in the room.	23. ソコニ ヘヤア ノナカニ ツクエガ アル
24. Is there a desk in your room?	24. ソコニ ナレノ ヘヤ ノナカニ ツクエガ アルカ
25. Where is the vase?	25. ドコニ カビンガ アルカ
26. There is a picture of my family.	26. コレガ ワレノ カゾク ノ エ ダ
27. I do not speak French.	27. ワレガ フランスゴヲ ハナサ ナイ
28. Do I speak English?	28. ワレガ エイゴヲ ハナスカ
29. Thank you.	29. アリガトウ
30. It is half past four.	30. ソレガ 4 ヲスギテ ハンブン ダ
31. It is twenty minutes past six.	31. ソレガ 6 ヲスギテ 20 フン ダ
32. We have breakfast at seven.	32. ワレガ アサゴハンヲ 7 ニオイテ モツ
33. Then I go to school.	33. ソレカラ ワレガ ガッコウ ニイク
34. School begins at half past eight.	34. ガッコウガ 8 ヲスギテ ハンブン ニオイテ ハジマル
35. What do the boys do?	35. オトコノコガ ナニヲ スルカ
36. He sometimes helps me with my lessons.	36. カレガ ワレヲ トキドキ ワレノ カ トトモニ タスケル
37. Can you swim?	37. ナレガ オヨグ コトガデキルカ
38. Yes, I can swim.	38. ハイ ワレガ オヨグ コトガデキル
39. I am walking.	39. ワレガ アルキ ツツアル
40. Am I walking?	40. ワレガ アルキ ツツアルカ
41. I am reading a story.	41. ワレガ ハナシヲ ヨミ ツツア カ
42. How do you do?	42. イカガ デスカ
43. He is the fastest runner in our class.	43. カレガ ワレヲ クミ ノナカニ モットモ ハヤイ ハシルモノ ダ
44. Come to the blackboard.	44. コクバンニ クルベシ
45. Write the names of the seven days of the week.	45. シュウ ノ 7 ヒ ノ ナマエヲ カクベシ
46. Let's have a race.	46. キョウソウヲ モツ コトラシヨウ ベシ
47. The crow forgot her cheese.	47. カラスガ サレノ CHEESE ヲ ワスレタ

機械翻訳「ヤマト」で翻訳できた例

KT-1⁶⁾

KT-1 は、我が国初の日・英・独 3 カ国語相互実験翻訳のために、九州大学が 1950 年代後半に開発した言語処理用計算機である。言語理論は直接構成要素法を採用し、処理方式は各国語特有の性質を表にまとめ、プログラムは各国語間のどの方向の翻訳にも共通にした特徴あるものであった。KT-1 ではおもに英語、ドイツ語、日本語間の操作について、原語の構文解析により原語を一度中間言語的な構文情報に変換する方法で、設計試作が行われた。KT-1 では 2 進法 6 チャンネル（直列には 64 進法）の直並列方式をとり、単語は直列 8 桁を単位とした。検索命令やパターンマッチに工夫を凝らし、文字列演算に優れていた。

論理回路にはトランジスタによるダイナミックフリップフロップ方式を採用した。記憶装置は記憶容量約 10 万ビット・回転数 12,000 回転/分の磁気ドラムを使用した。入出力には 6 単位符号を採用し、鑽孔紙テープによる方法を使った。

KT-1 の設計は九州大学、製作は三菱電機が行った。単能の翻訳機ではなく、一般技術計算用機能を持たせ、その利用も考慮されていた。

現在、KT-1 に使用された論理パッケージ、磁気ドラム、および入力データの鑽孔紙テープが、九州大学に保存されている。

なお、元九州大学、田町研究室の岡田直之氏の記憶では、次のことは事実です：

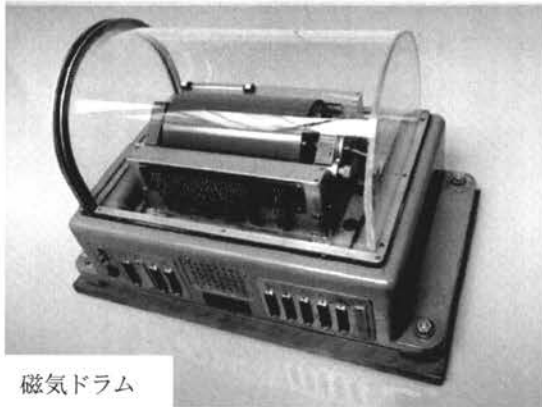
*自動翻訳機 KT-1 の設計は、大野克郎、栗原俊彦、田町常夫の3教授が行った。

*翻訳プログラムは、ほぼ田町常夫教授が担当した。

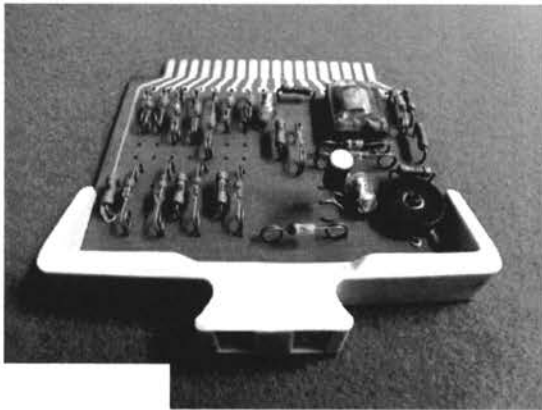
*10万ビットの主記憶に単語辞書、文法規則、それに翻訳プログラムを機械語を用いて配列した。

*栗原の意志をつぎ吉田将は、単語と単語のつながりに意味が反映されるという立場で、大規模言語データの収集やかな漢字変換に努めた。

*田町の翻訳プログラムを発展させ岡田直之は、言語や図形の持つ意味に注目し、語彙の意味分析や言語・図形翻訳などマルチモーダルな人工知能に進んだ。



磁気ドラム



製造者 九州大学

所有者 九州大学

史料所在地

〒819-0395 福岡県福岡市西区元岡 744

九州大学工学部百周年記念展示室『CENTENNIAL GALLERY』

公開情報 公開中

文献

1) エメール・ドラヴネ著、別所照彦、沢辺弘共訳：「翻訳機械」、文庫クセジュ、白水社 1964 年。

2) ウィキペディアから機械式計算機より、(<http://ja.wikipedia.org/wiki/>)

3) 玉木英彦、喜安善市編、「自動翻訳」、みすず書房、1960 年。

4) Booth (Andrew J.), Brandwood (L.) Cleave (J.P), Mechanical Resolution of Linguistic Problems, Butterworths Scientific Publications, London, 1958.

5) 蓼沼良一：「電子計算機による英文和訳の研究、電気試験所研究報告 第 624 号」、1961 年。

6) 情報処理学会、コンピュータ博物館より、(<http://museum.ipsj.or.jp/heritage/KT-1.html>)

Translation Technology from a Translator's Perspective

Alun Simpson, Freelance Translator

<http://www.linkedin.com/pub/alun-simpson/18/24/365>

I started working as a translator in 2002. My specialist field is technical documentation, user manuals in particular. In a way, I am a part of a new generation of translators, because I have been using translation memory (TM) since the very beginning of my career. When I started as an in-house translator, Trados was already the dominant TM product in my field, and one of my first tasks was to learn how to use it. TM is very well suited for technical documentation and user manuals, so I have always had a positive image of the role of technology in translation workflows. I know how technology can help us as translators be more efficient and deliver work of better quality and consistency.

Although I am an enthusiastic user of TM, I do not yet use machine translation (MT) in my day-to-day work. I might occasionally use Google Translate as a dictionary, but in my normal workflow it is still quicker for me to translate from scratch or from a low TM fuzzy match than to post-edit the results of online MT services. This has a lot to do with my language combination; Japanese to English translation is especially difficult for MT engines because nouns and other words are frequently omitted from the Japanese, and the simplicity of Japanese grammar means that the appropriate English tense is difficult for the engine to select. But I hear that in other, more similar, language combinations, the quality even of Google Translate is already good enough in some domains to be used by professional translators as a basis for post-editing, similar to a 60 to 75% fuzzy match from a TM.

I could use one of the many commercially available MT products to build my own engine, but as a freelance translator who works for many different clients, I think that the time needed to train the engine for a particular job would be longer than the time I would save by using it.

Because of my interest in translation technology, in April 2012, I attended the TAUS Executive Forum in Tokyo, which focused on the theme of Translation in the 21st Century. In December, I went to the Workshop on Future Directions in Translation Research in Nara, which was organized jointly by the National Institute of Information and Communications Technology (NICT) and the Center for Research and Innovation in Translation and Translation Technology (CRITT) at the Copenhagen Business School. Both these events gave me an insight into the current state of translation technology and research, with speakers giving presentations about the latest trends in linguistic processing, translation technology and language services.

I was surprised at the extent to which companies are already making use of machine translation in their actual operations. Since the advent of the internet, there has been an explosion in the amount of data with important information, such as websites, online Help, forums and social media. Listening to the different speakers made me realize just how much data out there needs to be translated – there are 2 billion words in the technical library of Cisco alone. Translating everything through the traditional channels of agencies and professional translators would be prohibitively expensive, even for the biggest companies. As the amount of data increases, so too must new ways of efficiently making this information understandable to people in other

languages.

Chris Wendt of Microsoft gave a presentation at TAUS describing how his company has tried to address this issue. Microsoft uses an innovative fusion of human translation and machine translation for its online knowledge base. Their support pages contain an enormous amount of information, which cannot be realistically translated by humans alone. Instead, they use MT for about 90% of the articles, and the most frequently viewed MT pages are flagged up for translation by humans. I think that this is an interesting example of how MT and human translators can support and complement each other in a company's translation workflow.

When evaluating the MT results, Microsoft ask whether the user's problem was solved, and deliberately do not ask whether the translation was correct or good. As the use of MT increases, I think that this kind of utilitarian evaluation of translations will become more important, in particular for large databases where the choice is not between human translation and machine translation, but between machine translation and no translation at all. Microsoft's system also incorporates feedback loops consisting of user ratings and user-generated corrections of the MT results. I think that this combination of machine translation, professional translation and user-generated content is a model that many companies and startups will develop in the future.

A part of this new paradigm of translation is crowd sourcing. At TAUS, Miori Sagara of Baobab talked about her company's translation platform, Ryugakusei Network. The concept of this service is for language students to gain translation experience and make a little money while translating Japanese products into English or Chinese. The translators revise the output of machine translation, and the site uses techniques such as rankings and gamification to improve quality and motivation. Although the site aims to attract part-time translators, professional translators from China have also joined. I think this is an interesting trend; in establishing platforms such as these, the translation is moved closer to the customer, and the process is democratized by allowing access for more people to make money from translation. Gengo performs a similar role; 90% of their jobs are less than 100 words, and these can be translated anywhere, even on a smartphone or tablet. I think that these jobs are not the kind that traditional, highly experienced translators would take, but rather open up a new market for part-time translators, such as busy mothers who can work a little on the side, or novice translators who want to gain more experience. The internet breaks down barriers and allows customers and suppliers to make contact directly, bypassing traditional intermediaries. This trend can be seen not only in translation, but also in many other industries.

The NICT and CRITT workshop in Nara was rather more technical in nature, with researchers describing how they are trying to solve problems in language processing. Daniel Ortiz of Universitat Politècnica de València gave an interesting presentation about incremental learning for MT systems. Here, human interactions and corrections are fed back into a statistical MT system to improve the output. This incremental learning is performed sentence by sentence, and so fits in naturally with post-editing and interactive MT systems. I can see how this could be effectively incorporated into a translation workflow.

I also found the last session at Nara about human translation processes very interesting. Researchers had analyzed the behavior of translators with varying skill levels as they worked, through keystrokes, mouse

movements, and even heat maps obtained from eye-tracking analysis. As an actual translator I found it fascinating that their research reflected how I have developed with experience – nowadays I spend only a brief time on orientation before starting, I perform my draft translation rapidly with minimal backtracking as I go along, and then spend more time on a detailed revision at the end. I am not sure whether I should be happy or disturbed that I conform so closely to the results of their research!

In conclusion, I think that now is an exciting time for machine translation. It is only recently that advances in computational power and data analysis have resulted in significant improvements in output. This has opened up new opportunities in the translation industry, including the post-editing of machine translation and crowd-sourcing. Although some translators may be uneasy about these trends, I firmly believe that ever-increasing volumes of data mean that highly skilled, professional linguists will always be in demand, at least during my lifetime. MT systems will start performing more varied functions and the nature of our jobs may change, but as globalization drives forward, so too will opportunities grow. I think that those language professionals who embrace new technology in innovative ways will be the people who shape our industry in the 21st century.

Workshop on Future Directions in Translation Research (未来の翻訳研究に関するワークショップ)

隅田 英一郎 (NICT)

1 概要

掲題のワークショップ¹⁾が2012年12月3日(月)～4日(火)に、東大寺総合文化センターで開かれた。CBS (Copenhagen Business School²⁾、デンマーク)とNICT(情報通信研究機構³⁾)の共催である。二日間にわたったワークショップのプログラムと発表のスライドはワークショップのホームページから入手できる。本稿では、主催者の一人の隅田が概要と一部の講演のサワリを紹介する。

会場の東大寺総合文化センターは世界遺産である東大寺の境内にある異色の会議施設である(図1)。



図1 東大寺総合文化センター

センターのホームページ⁴⁾には、「東大寺の意義はなんといっても創建時代にあります。聖武天皇は動物も植物も共に栄えることを願って大仏さまを造られ、その思いを学ぶために、大勢の若い僧侶が全国から東大寺に集まりました。今でいう総合大学です。」とあるように、アカデミックな活動に理解があり今回のような国際会議にも門戸を開いている。

設備も良く、また、東大寺の境内にあるので、会議の前後に参加者が観光するにも大変都合が良い

(図2、図3)。リラックスして会議することが出来る会場である。



図2 東大寺大仏殿の鴟尾



図3 東大寺境内の紅葉

この会議は今回が初開催で、「翻訳」を取り巻く重要なテーマを招待講演者に語っていただき、「翻訳の未来」を少しゆっくりと考えるヨスガにしようというのが趣旨になっている。また、BRICS⁵⁾のブラジルやインドとのコミュニケーション需要が急速に高まっている中で、ブラジルやインドの言語について両国の研究者と連携していくことも目的の一つになっていて、インドの CDAC/ Noida の Rai Mahesh K. Sinha、ブラジル UFMG /LETRA の Fabio Alves らとの議論を通して企画した。

同時通訳の自動化目指した話、アイトラッカーを

利用した翻訳メカニズムの解明などの多岐にわたる研究テーマで、松原 茂樹（名古屋大学）、Srinivas Bangalore（AT&T）、Daniel Ortiz（UPV）、Andrew FINCH（NICT）、Peter Juel Henriksen（CBS）、Ronaldo Martins（NILC）、相澤 彰子（NII&東大）、Michael Carl（CBS）、Fabio Alves（FALE-UFMG）という多様な研究者でプログラムを構成した。

また、参加者も ICT 系の企業の研究者、大学の研究者、翻訳者や翻訳会社関係者、など、総勢 50 名に上った（図 4）。



図 4 会場の様子

2 講演のピックアップ

以下、講演のいくつかをピックアップしてご紹介する。

2. 1 松原は“Incremental Parsing and its Applications”と題して同時通訳の自動化という夢の技術の文脈の中で Incremental（漸次的又は漸進的）構文解析について紹介した。

まず、松原は、講演とその同時通訳を集めたコーパス（図 5）を作成し、これを通訳パターンや通訳タイミングなどの研究素材とした（このコーパスは公開されている）。

通常の自動翻訳では、一文をまとめて構文解析するが、このコーパスを調べると、人間の同時通訳ではもっと短い単位（句や分節に近い短い単位）で訳していることがわかる。同時通訳の場合は話者が一文話し終わる前に翻訳を開始しなくては同時性が成立しないからである。従って、

同時通訳の要素技術である構文解析も、話の途中で開始しなくてはならない（これが「incremental」の本質である）。

このための工学的な手法の現状と課題を紹介した。

Bangalore が言及したように、全人類 68 億人のうち 38 億人は中国語、スペイン語、英語、アラビア語、ヒンディ語 のいずれかの言語を話しており、そのうちの 70% は携帯電話を所有している。これらの言語間の携帯電話を介した通訳システムを開発できれば非常に大きな需要に応えることになる。

0001 - 00:05:204 00:09:399 N: The theme for this speech is going to be the American	0001 - 00:06:440 00:08:207 I: (F え) 次のテーマですが
0002 - 00:09:840 00:11:032 N: presidential debate	0002 - 00:08:044 00:09:783 I: アメリカの
0003 - 00:11:424 00:13:391 N: and who would be the	0003 - 00:10:296 00:12:775 I: (F え) 大統領に関するディベート
0004 - 00:13:640 00:15:215 N: better president for America <SB>	0004 - 00:13:006 00:14:424 I: そして誰が
0005 - 00:16:272 00:18:327 N: (F um) Let's see, today is	0005 - 00:14:648 00:18:255 I: より良い大統領とアメリカのためになり 得るかということですが

図 5 同時通訳のコーパス
(スライドからの引用)

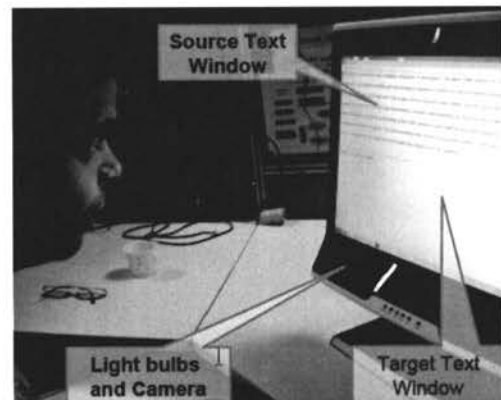


図 6 アイトラッカーの利用（スライドの引用）

2. 2 相澤やCarlの用いたアイトラッカー（図 6）は、眼の動きを追跡し、画面の中の、どの部分からどの部分へと視線が移ったか、その部分を何秒間見たか（滞留時間）といったデータを記録する。

テキストを読む際や翻訳する際に、このデータを分析すれば、ベテラン翻訳者と新人翻訳者の相違がわかったり、これをトレーニングに活用したりな

ど、今までにない様々な展開が可能になる。

3 NICTのデモ

NICTはその研究成果4件を詳しく紹介した。

AAMTジャーナルなので、このうち翻訳と直接関係する2件について紹介する。

3.1 NICTでは、統計翻訳(SMT)のパラダイムで、特許分野(英⇔日、中⇔日⁵⁾)・Eコマース分野(英中韓⇔日)・旅行会話分野(21言語の全組合せ)など、分野を特化した高精度自動翻訳の研究している。当日は英日特許翻訳について詳しく紹介した。

1800万文の特許対訳コーパスを、過去17年間におよぶ特許ファミリーの文書から自動構築した。優先権番号にもとづいて特許ファミリーを特定した後、請求項・要約・説明の各パートを抽出し、パートごとに文のアライメントをとる。このときに、文の類似性を辞書の対訳とのオーバーラップを使って判定している。自動アライメントの精度は、1000文のうちA判定が899文(ほぼ問題なく原文と訳文が対応)、B判定が72文(80%の内容が対応)C判定が26文(対応度が80%以下)、その他3文という高い精度を達成した。

英語と日本語は語順が大きく違うので、標準的なSMTの語順調整の仕組みを使って、文を翻訳してもうまく翻訳できない。語順の事前変更(Pre-orderingと呼ぶ)を行ってからSMTにかけると精度がよくなることが知られているので、翻訳方向に依存しない新たなアルゴリズムでこれを実現した。

翻訳精度は平均文長が約25語の文であれば、70%~80%と高い。より長い文での精度を上げることが今後の課題になっている。

翻訳速度は1秒あたり100ワード(おおむね3文)と特に遅くはないが、まだまだ高速化の余地が残っている。また、NICTのSMTでは対訳辞書に登録した訳語を優先する機能も実装してある。

3.2 多言語音声翻訳システムは、多国籍の研究共同体で研究開発を進めると効率が良い。

1993-1999年のC-STAR、2006-2009年のA-STARという前身があって、2010年以降U-STAR⁶⁾へと進化している。世界中の言語を音声翻訳する枠組を初めて提案し、同枠組の核心であるモジュール間通信技術をITUにおいて国際標準化した。これによって、U-STARへの参加は23ヶ国の26機関に急増した。

2012年7月より、スマートフォン用の音声翻訳のアプリケーションVoiceTra4U-Mを無償公開し、全世界で同技術を用いた実証実験を実施している。VoiceTra4U-Mでは、27言語の自動翻訳、17言語の音声認識、14言語の音声合成に対応している。

また、来年の会議について、現在、企画中である。詳細が決定し次第、AAMTの皆様にもご案内する予定である。

参考資料

- 1) <http://www.mastar.jp/wfdtr/index.html>
- 2) http://www.cbs.dk/Forskning/Institut/centre/Institut/ISV/Menu/Medarbejdere/Menu/Videnskabelige/Videnskabelige/Lektorer/michael_carl
- 3) http://www2.nict.go.jp/univ-com/multi_trans/index.html
- 4) <http://culturecenter.todaiji.or.jp/accessmap.html>
- 5) <http://www.nict.go.jp/press/2012/11/05-1.html>
- 6) <http://www.ustar-consortium.com/>

第2回 特許情報シンポジウム報告

横山晶一（山形大学）

1 はじめに

AAMT/Japio 特許翻訳研究会（委員長：辻井潤一東大名誉教授）では、機械翻訳国際会議(MT Summit)のない年に、国内でさまざまなイベントを行っている。

特許情報シンポジウムは、2010年（平成22年）に第1回が東大・本郷キャンパスで行われ、今回2012年（平成24年）に第2回が行われた。ここでは、その内容等について簡単に報告する。

2 シンポジウムの開催趣旨

論文募集には、次のような開催趣旨が書かれている。

「特許情報は情報処理技術の応用分野の一つとして近年世界的に関心が高まっている。本シンポジウムは、特許情報処理技術の研究開発を促進することを目的として、2010年に第1回が開催された。研究者、実務家、政府関係者が集まって、構想、方法論、将来展望、実務経験、政策などを議論する場として企画され、さまざまな発表が行われた。今回第2回のシンポジウムも、第1回の成果と趣旨を踏まえて、招待講演（3件）、検討会報告（1件）、一般講演（5ないし10件）で構成される。」

この研究会がふだん取り扱っている特許翻訳の問題よりはやや幅広く、特許情報全般について取り扱う意志のもとに企画されている。

上記のように、このシンポジウムは、招待講演や報告とともに、一般に公募した論文発表も行われる。一般公募の論文の領域は、上記趣旨では次のようになっている。

論文の分野

下記の分野を含む特許情報処理の諸分野:

- Analysis and classification for patent documents,

- Machine translation and translation aids for patent documents,

- Contrastive studies for multilingual patent documents,

- Language resources for patent documents,

- Dictionaries and terminology databases for patent documents,

- Parallel, comparable or monolingual corpora for patent documents,

- Information extraction and information mining from patent documents,

- Patent map development,

- Evaluation techniques for patent translation,

- Patent information retrieval.

今回は締め切りまでに4編の論文投稿があった。

3 開催概要

開催日時：平成24年11月30日（金）

13:00～18:00

場所：京都大学東京オフィス（東京都港区）

参加者：事前登録者98名、当日登録者9名、欠席者20名、当日参加者87名

会場の関係で、100名限定で参加を受け付け、ほぼ満席となる盛況であった。また、昨今の国際情勢から、海外からの参加がやや懸念されたが、特に問題なく参加していただいた。

次節で、実際の内容について述べる。

4 開催内容

今回は、海外からの参加者のことを考慮して、一般講演以外は英語で行われ、一般講演は、英語、日本語どちらでもよいことにしたが、4件の発表すべて日本語で行われた。

最初に、主催者側として、守屋敏道氏（Japio

専務理事、特許情報研究所所長）と辻井委員長から開会のあいさつがあった。

招待講演は3件行われた。

最初は、China Patent Information Center (CPIC) の王丹氏による“Making Effective Use of Machine Translation for Patent Documents: Practice of CPIC”という講演であった。ここで氏は、CPIC で特許文書を扱う際のMT使用ロードマップを示すとともに、人間の翻訳とMTとの使い分けなどにも言及した。ロードマップでは、最近4～5年間で、中英、英中、日中、中日などのMT使用や開発・研究が行われていることを述べた。

2番目の招待講演は、Paul Schwander 氏 (Director External Products and Services European Patent Office, EPO)による“Machine Translation at the EPO Removing language barriers from patent documentation”である。Schwender氏は、EU内での23言語に関する取り組みと、最近EU内でも需要が高まっているアジア系言語による特許への調査にMTがどのように寄与しているかについての現状を述べた。

3番目の講演は稲葉崇氏（特許庁総務部普及支援課情報企画室調査第二係長）による「特許庁における機械翻訳に関する取組」（英文タイトルはJPO's Approach for Machine Translation - To establish productive utilization method and create appropriate policies）である。この講演では、特許庁におけるMT利用の現状、特に、MTの精度向上や、MTを活用した外国文献へのアクセス向上の取組、MTの品質評価について述べた。

次に、9月7日に開かれた特許文書の機械翻訳結果評価方法検討会（東大本郷キャンパス、このジャーナルで概要を既報）のまとめを、このシンポジウムのまとめ役となった当研究会の江原暉将副委員長（山梨英和大教授）、越前谷博委員（北海学園大准教授）が行った。MT評価に用いられるBLEU, NIST, RIBES, IMPACTといった自動評

価や、人間による評価について議論をしたシンポジウムを総括し、今後の指針を示した。

一般発表は次の4件である。

「依存関係を用いた特許分野のための日英中対訳フレーズの切り出しアルゴリズム」池田秀人氏ら立命館大グループによる発表。依存関係階層木の間節点を発展させた対訳フレーズを用いてMTを行おうという試みである。

「特許明細書の翻訳者から基本的な誤訳の実例を示して対策を提案」吉川潔氏（フリー翻訳者）。氏が長年にわたる特許翻訳の経験から、MTにおける種々の誤謬について持論を述べたものである。

「特許翻訳の品質を向上させるための形態素解析結果を利用した文書比較・日本語精査ツール-歌詠と鶯-の試作」楠本浩二氏（（株）クレステック）らのグループによる発表。特許翻訳品質向上のためのパーソナルツールとして、類似ファイルの差分などを提示するシステムについて述べた。

「技術調査のための特許情報抽出」太田貴久氏（豊橋技科大）らのグループによる発表。技術調査に必要な技術の総称、発明の作用などを高精度に取り出すことで、ユーザに有用な情報抽出を行うシステムについて述べた。

最後に簡単にまとめが行われた。

5 終わりに

以上が第2回特許情報シンポジウム報告である。主催者側で活躍された委員の皆さん、講演者に対する種々の交渉に当たられたJapioの方々、会場をお貸し下さった京都大学、事務を担当された（株）ナビックスの担当者に感謝します。

今後もこのような催しを積極的に実施していきたいと考えております。

なお、シンポジウムの内容については、当研究会のHP (<http://aantjapio.com/>)上で公開の予定です。

JTF 翻訳祭出展報告

機械翻訳課題調査委員会

1. はじめに

AAMT 機械翻訳課題調査委員会は、昨年度に引き続き JTF 翻訳祭の展示コーナーに出展し AAMT の活動紹介、新会員募集、課題調査委員会で仕様策定している UTX の PR、機械翻訳に関するアンケート等を行った。本稿では、AAMT ブース概要と、出展者の立場で見た翻訳祭全体の印象、また、出展を通して得られた翻訳家の現状など紹介する。

2. 国内最大の翻訳業界のイベント

JTF 翻訳祭は、JTF（日本翻訳連盟）によって年に 1 回開催され、700 名以上が参加する翻訳業界最大のイベントである。今年度は 11 月 28 日にアルカディア市ヶ谷において、8 つの講演会場で計 30 の講演と、「翻訳プラザ」と呼ばれる展示コーナーにて 35 の会社や団体からデモ展示があった。講演の内容は、映像翻訳ライブレッスンといった翻訳家向けのものから、翻訳ワークフローや IT 利用といった翻訳会社向けのものまで多岐にわたる。参加者を見ると男女比はおおよそ半々、就職したばかりの若手から翻訳経験 30 年のベテランまで、翻訳業界から様々な立場の人が参加されていた。

3. 機械翻訳は関心の的

翻訳祭に参加する翻訳家（翻訳を生業としている人たち）が機械翻訳にどのくらい関心を持っているかは、AAMT メンバにとっては興味深い点であろう。筆者は、現状の機械翻訳は翻訳家が満足するレベルに到達していないことから、翻訳家の機械翻訳への関心はかなり低いだろうと予想した。しかし、「機械翻訳を組み込んだ翻訳工程の最適化」というテーマの講演が盛況だったことや、出展ブースでのアンケートで機械翻訳の利用頻度で「毎日」と答えた人が 36% もいた（詳細は付録参照）ことを考えると、機械翻訳はプロの翻訳家にもかなり浸透していると言ってもよさそうである。

4. AAMT ブースは大盛況

AAMT のブースには課題調査委員会のメンバが交代で詰めて、前を通りかかる人たちに各種資料を手渡し、アンケート回答を呼びかけた。

・資料配布

AAMT の活動を紹介するため AAMT ジャーナルのバックナンバー 4 種類の配布を行った。鞆が重くなるにもかかわらず 4 冊全部持ち帰られる人がいた一方で、パラパラと立ち読みしただけの人もいた。AAMT ジャーナルについての感想を伺うと、「機械翻訳製品のことが良くわかって参考になる」、「専門的な印象が強くてとっつきにくい」などの意見が聞かれた。

ブースは幅 180cm の机 1 個という狭いスペースだったが、AAMT ジャーナルのほか AAMT の活動を紹介するチラシ、新入会員申込み用紙、UTX のパンフレットを所狭しと並べ、最後には持ち込んだ資料（各々 50~100 部）の大半を配布することができた。AAMT ブースは他のブースに比べて立ち止まる人が多く（懸賞の効用だろうか？）対応メンバは大忙しだった。このように、AAMT の知名度アップとアクティビティの PR という今回の出展目的は十分に果たせたのではないかと思われる。



・アンケート

昨年に続いて翻訳家向けに機械翻訳に関するアンケートを実施した。今回は特許翻訳従事者向けとそれ以外の翻訳家向けにアンケートを2種類用意し、どちらかの該当するアンケートに回答してもらった。アンケートはWebシステムとして開発し、ノートPCのブラウザ上でマウス選択によって回答してもらう形式とした。こうすれば回答と同時に自動でDB入力できるため、後の集計作業が楽である。ブースの机の上にノートPCを2台設置し、レンタルルータを利用してインターネット接続を行った。回答数は特許翻訳従事者が15名、その他の翻訳家が42名の計57名であった。(アンケート結果の詳細については付録を参照)なお、回答者集めの索として、回答者にはその場でくじを引いてもらい10名の当選者に1000円分のQUOカードを進呈した。また後日、抽選で回答者全体から1名に、クロスランゲージ様が景品として提供してくださった翻訳ソフト「PAT-Transer V11」を進呈した。

側としてのAAMTと利用側としてJTFが連携して、様々な情報交換をする場を作ることが必要であると感じた。機械翻訳課題調査委員会としても今後JTFとの連携強化に向けた取り組みを具体化させていきたいと考えている。

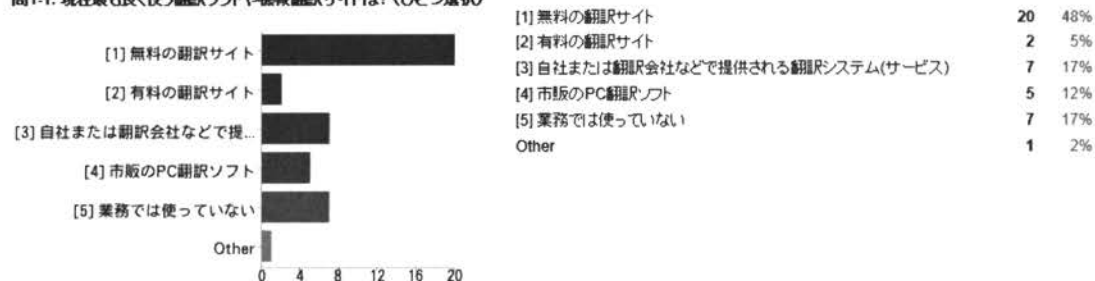
5. AAMT-JTFの連携強化の必要性を実感

翻訳家は機械翻訳の利用者として本命でありながら、これまで本格的に使う人は少数に止まっていた。これにはいろいろな理由があるだろうが、機械翻訳を敵視する(あるいはライバルとみなす)翻訳家が多いことが問題の一つだったように思う。近年、翻訳家の世代交代が進み、また、機械翻訳がより身近で簡単に使える存在になるにつれ、翻訳家の意識も変わってきたのではないだろうか。つまり、従来のように生理的に機械翻訳を嫌う人が減り、機械翻訳をどうやって使えば便利だろうかという発想ができる人が増えてきたように思う。今回の翻訳祭出展を通じていちばん印象深かったことは、上記のような機械翻訳に対する翻訳家の意識の変化である。

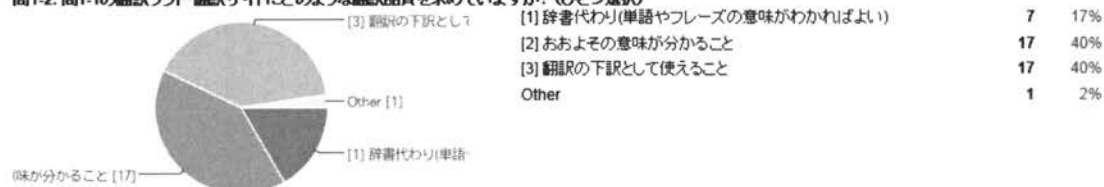
翻訳家の意識がこのように変化したことは機械翻訳を提供する側として喜ばしい限りである。ようやく、機械翻訳の開発者と利用者が手を結んで機械翻訳の活用を啓蒙できる時期が訪れたと言ってもよいかもしれない。この意味からも、機械翻訳の発展のためには、開発

付録1： 機械翻訳についてのアンケート結果(特許以外の翻訳者向け)回答者数:42

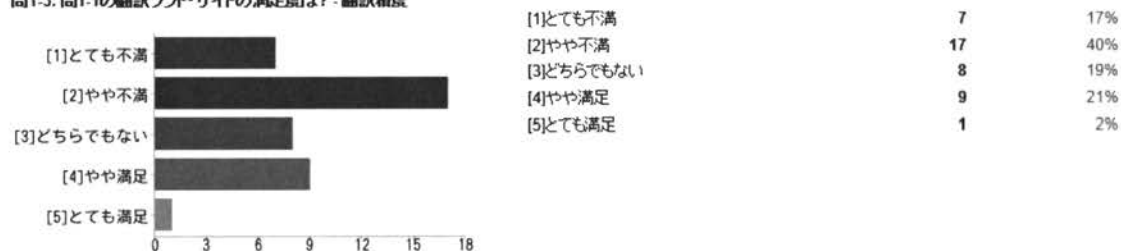
問1-1. 現在最も良く使う翻訳ソフトや機械翻訳サイトは? (ひとつ選択)



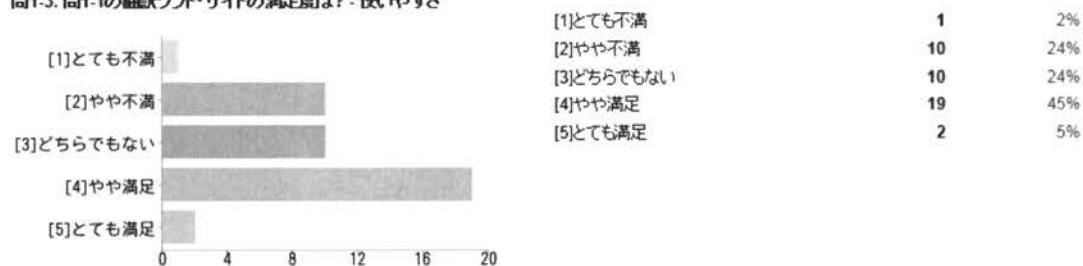
問1-2. 問1-1の翻訳ソフト・翻訳サイトにどのような翻訳品質を求めていますか? (ひとつ選択)



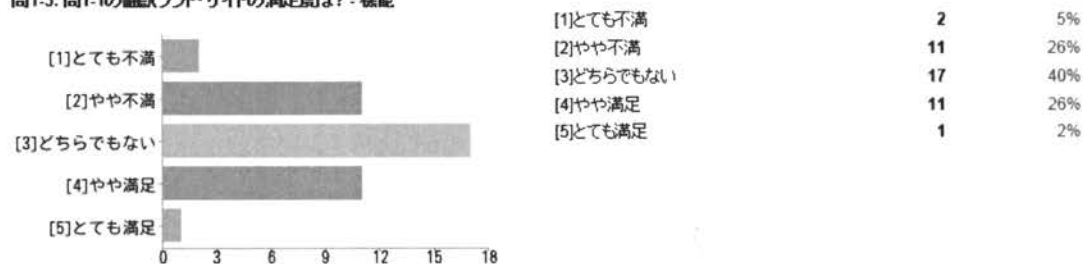
問1-3. 問1-1の翻訳ソフト・サイトの満足度は? - 翻訳精度



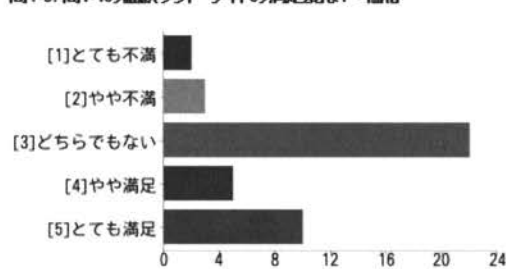
問1-3. 問1-1の翻訳ソフト・サイトの満足度は? - 使いやすさ



問1-3. 問1-1の翻訳ソフト・サイトの満足度は? - 機能

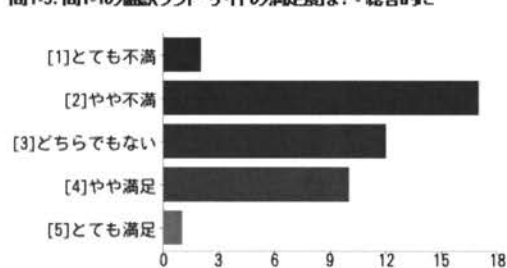


問1-3. 問1-1の翻訳ソフト・サイトの満足度は? - 価格



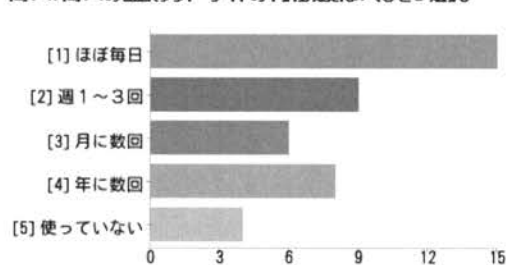
[1]とても不満	2	5%
[2]やや不満	3	7%
[3]どちらでもない	22	52%
[4]やや満足	5	12%
[5]とても満足	10	24%

問1-3. 問1-1の翻訳ソフト・サイトの満足度は? - 総合的に



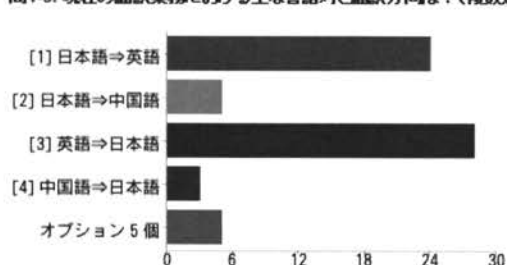
[1]とても不満	2	5%
[2]やや不満	17	40%
[3]どちらでもない	12	29%
[4]やや満足	10	24%
[5]とても満足	1	2%

問1-4. 問1-1の翻訳ソフト・サイトの利用頻度は? (くどつ選択)



[1] ほぼ毎日	15	36%
[2] 週1~3回	9	21%
[3] 月に数回	6	14%
[4] 年に数回	8	19%
[5] 使っていない	4	10%

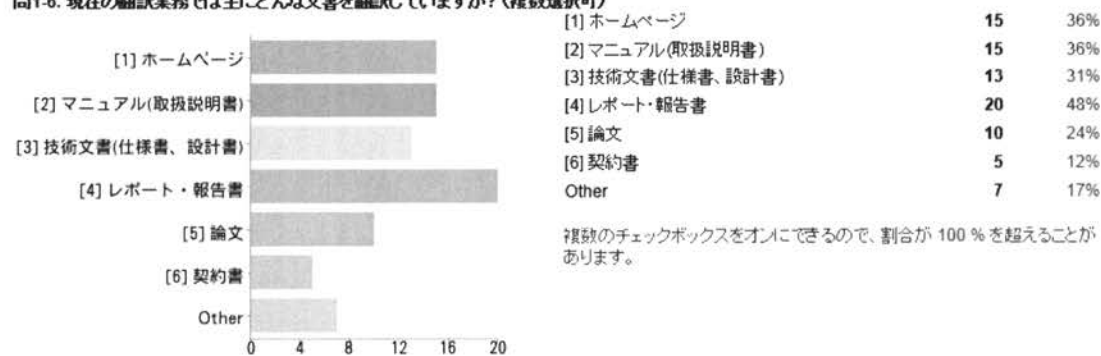
問1-5. 現在の翻訳業務における主な言語対と翻訳方向は? (複数選択可)



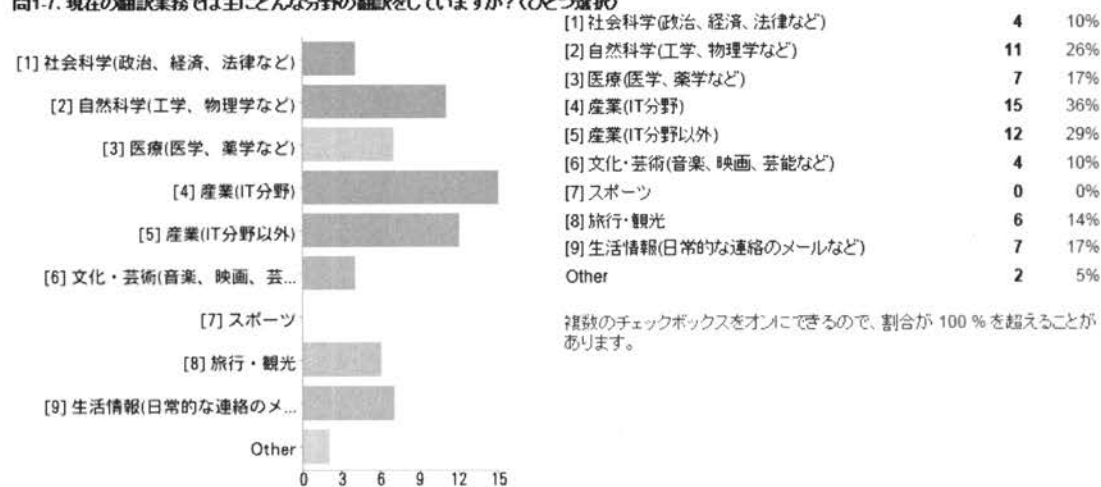
[1] 日本語⇒英語	24	57%
[2] 日本語⇒中国語	5	12%
[3] 英語⇒日本語	28	67%
[4] 中国語⇒日本語	3	7%
オプション5個	5	12%

複数のチェックボックスをオンにできるので、割合が 100 % を超えることがあります。

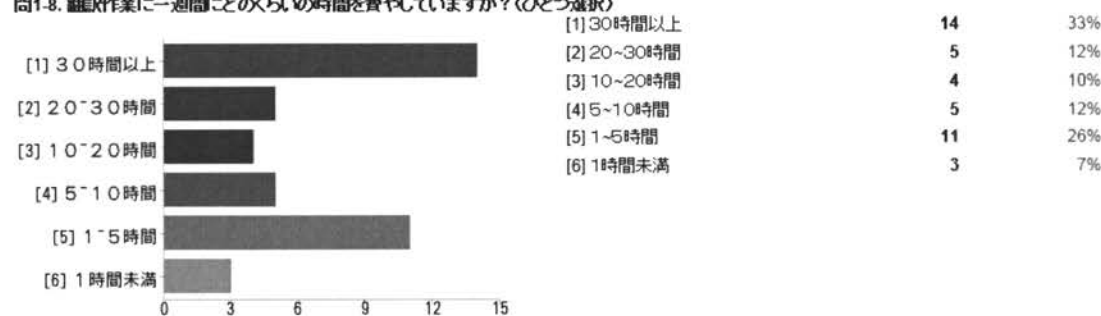
問1-6. 現在の翻訳業務では主にどんな文書を翻訳していますか？(複数選択可)



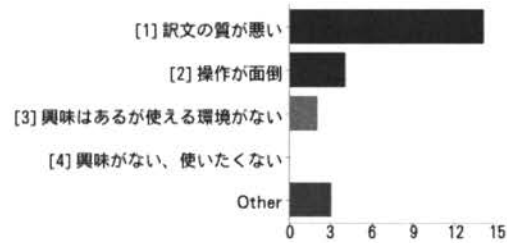
問1-7. 現在の翻訳業務では主にどんな分野の翻訳をしていますか？(ひとつ選択)



問1-8. 翻訳作業に1週間どのくらいの時間を費やしていますか？(ひとつ選択)



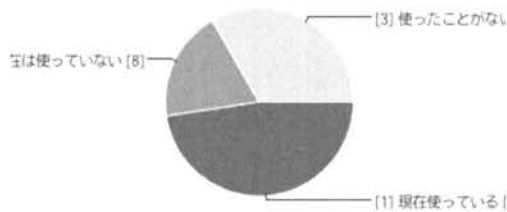
問1-10. 機械翻訳を使っていない理由をお聞かせください(複数選択可)



[1] 訳文の質が悪い	14	74%
[2] 操作が面倒	4	21%
[3] 興味はあるが使える環境がない	2	11%
[4] 興味がない、使いたくない	0	0%
Other	3	16%

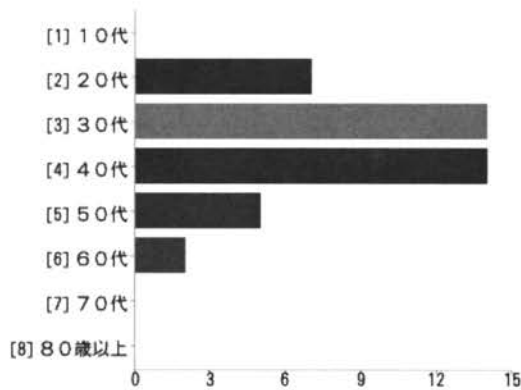
複数のチェックボックスをオンにできるので、割合が 100 % を超えることがあります。

問1-11. 翻訳メモリをお使いですか？(ひとつ選択)

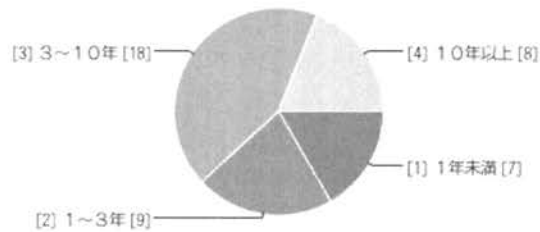


[1] 現在使っている	20	48%
[2] 以前使っていたことがあるが現在は使っていない	8	19%
[3] 使ったことがない、翻訳メモリを知らない	14	33%

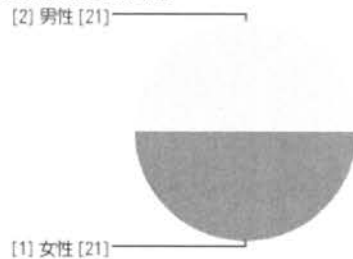
問2-1. 年齢(ひとつ選択)



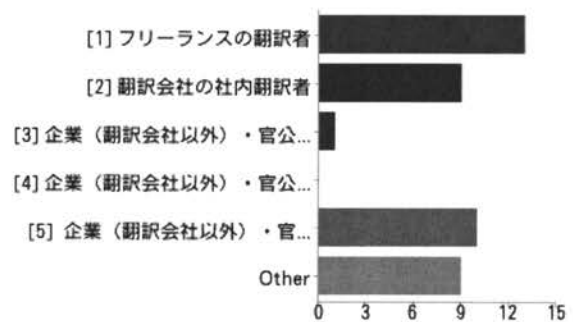
問2-3. 翻訳業務につかれてからの期間ほどのくらいですか？(ひとつ選択)



問2-2. 性別(ひとつ選択)



問2-4. 現在どのような立場でお仕事されていますか？(ひとつ選択)



付録2：機械翻訳についてのアンケート結果(特許翻訳者向け) 回答者数:15

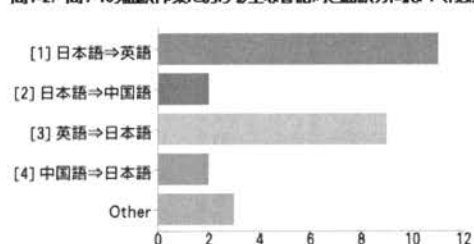
問1-1. おこどのような目的で特許翻訳を行っていますか？(ひとつ選択)

海外の技術動向調査 [0]
特許の公知例調査 [0]



[1] 外国への出願明細書の翻訳/作成	11	73%
[2] 海外の技術動向調査	0	0%
[3] 特許出願時の公知例調査	0	0%
Other	4	27%

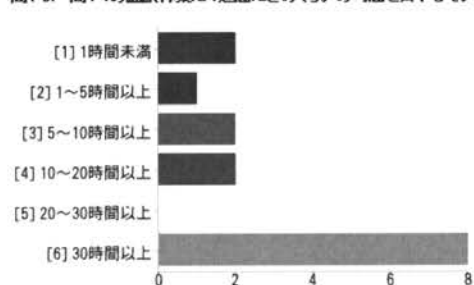
問1-2. 問1-1の翻訳作業における主な言語対と翻訳方向は？(複数選択可)



[1] 日本語⇒英語	11	73%
[2] 日本語⇒中国語	2	13%
[3] 英語⇒日本語	9	60%
[4] 中国語⇒日本語	2	13%
Other	3	20%

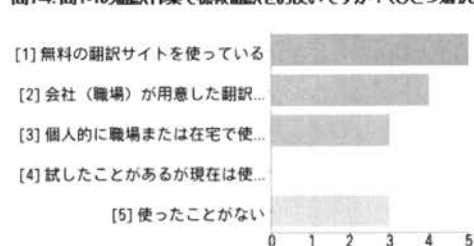
複数のチェックボックスをオンにできるので、割合が 100 % を超えることがあります。

問1-3. 問1-1の翻訳作業に1週間にどのくらいの時間を費やしていますか？(ひとつ選択)



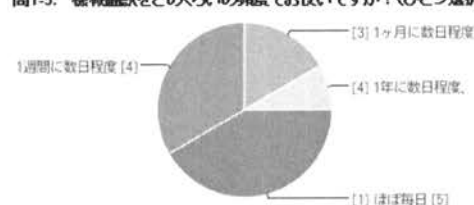
[1] 1時間未満	2	13%
[2] 1～5時間以上	1	7%
[3] 5～10時間以上	2	13%
[4] 10～20時間以上	2	13%
[5] 20～30時間以上	0	0%
[6] 30時間以上	8	53%

問1-4. 問1-1の翻訳作業で機械翻訳をお使いですか？(ひとつ選択)



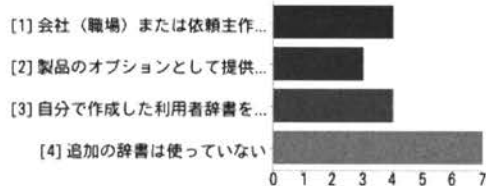
[1] 無料の翻訳サイトを使っている	5	33%
[2] 会社(職場)が用意した翻訳ソフト(サービス)を使っている	4	27%
[3] 個人的に職場または在宅で使っている	3	20%
[4] 試したことがあるが現在は使っていない	0	0%
[5] 使ったことがない	3	20%

問1-5. 機械翻訳をどのくらいの頻度でお使いですか？(ひとつ選択)



[1] ほぼ毎日	5	33%
[2] 1週間に数日程度	4	27%
[3] 1ヶ月に数日程度	2	13%
[4] 1年に数日程度、またはそれ以下	1	7%

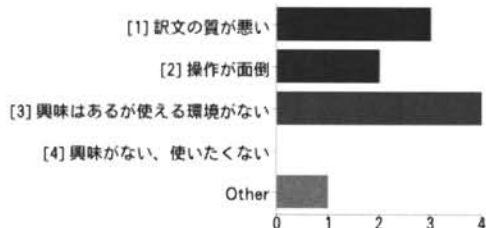
問1-6. 機械翻訳の辞書についてお聞かせください(複数選択可)



[1] 会社（職場）または依頼主作成の専用辞書を適用している	4	33%
[2] 製品のオプションとして提供されている専門用語辞書を適用している	3	25%
[3] 自分で作成した利用者辞書を適用している	4	33%
[4] 追加の辞書は使っていない	7	58%

複数のチェックボックスをオンにできるので、割合が 100 % を超えることがあります。

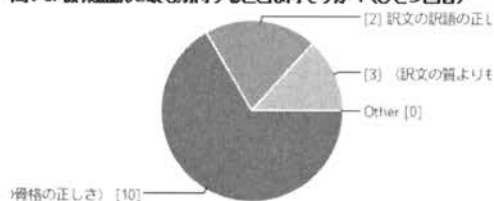
問1-7. 機械翻訳を使っていない理由をお聞かせください(複数選択可)



[1] 訳文の質が悪い	3	43%
[2] 操作が面倒	2	29%
[3] 興味はあるが使える環境がない	4	57%
[4] 興味がない、使いたくない	0	0%
Other	1	14%

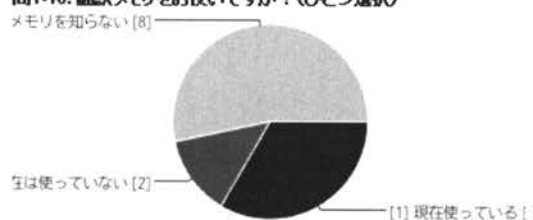
複数のチェックボックスをオンにできるので、割合が 100 % を超えることがあります。

問1-8. 機械翻訳に最も期待することは何ですか？(ひとつ回答)



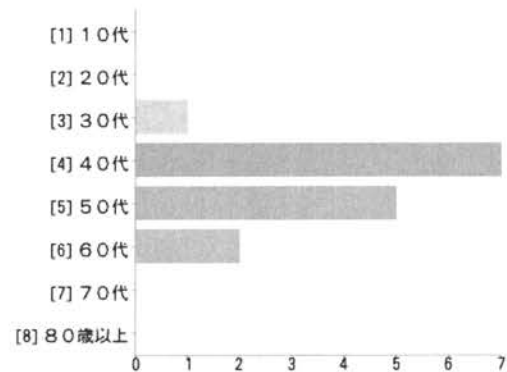
[1] 訳文の構文の正しさ(文章の骨格の正しさ)	10	67%
[2] 訳文の訳語の正しさ、または、用語辞書機能の充実	3	20%
[3] (訳文の質よりも)不適切な表現やスタイル違反チェックなどの付加機能	2	13%
Other	0	0%

問1-10. 翻訳メモリをお使いですか？(ひとつ選択)

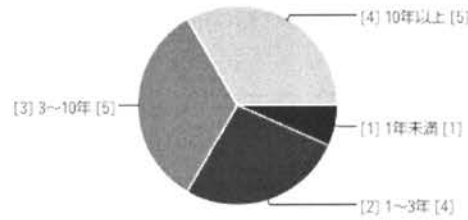


[1] 現在使っている	5	33%
[2] 以前使っていたことがあるが現在は使っていない	2	13%
[3] 使ったことがない、翻訳メモリを知らない	8	53%

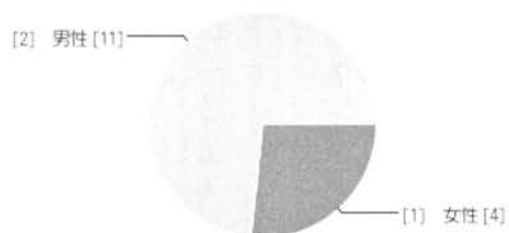
問2-1. 年齢（ひとつ選択）



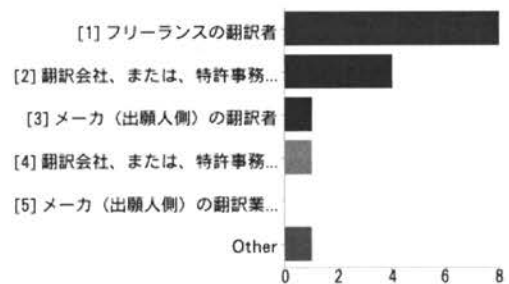
問2-3. 特許翻訳業務に従事している期間はどのくらいですか？（ひとつ選択）



問2-2. 性別（ひとつ選択）



問2-4. どのような立場でお仕事されていますか？（ひとつ選択）



用語集形式 UTX の FAQ(よくある質問)

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

(英語版 : <http://www.aamt.info/english/utx/faq.htm>)

機械翻訳課題調査委員会ワーキンググループ 3 (UTX チーム)

- 基本
- UTX 用語集の作成に関して
- 翻訳発注者／翻訳会社向け
- 機械翻訳に関して
- 開発者向け

基本

用語集の作成には多くの手間がかかるので、したくありません。

時間と金の無駄ではないでしょうか。

実際のところ、用語集を作らないことこそ、時間の浪費です。用語集なしで文書の作成や翻訳をすると、非常に時間がかかり、苦勞することがあります。用語集を作成すれば、時間を節約できます。その他の人々が参照できる用語集という形式で、信頼性が高い情報を持っていれば、間違いを減らせます。他のだれかがすでに知っている用語を、(ご自身を含め)多くの人がそのつど確認して時間を無駄に費やす必要がなくなります。複数の用語の候補がある場合に、どの用語が最良かと思悩むことがなくなります。

また、用語集なしで、文書作成やソフトウェア開発をすると、費用の浪費につながります。アプリケーションにすばらしい機能を付けようとして多くの費用を費やしても、機能の命名が不適切で一貫性がなければ、ユーザーはその存在にすら気づきません。一貫した用語集を使用すれば、このようなことを防げます。

用語集の作成は難しくないですか？

その点は、まさに UTX がお手伝いできることです。UTX は、最小限のシンプルなルールを使うこ

とで、用語集の作成と保守を大幅に単純化します。

AAMT は、UTX から利益を得ているのですか？

いいえ。UTX チームのメンバー(すなわち AAMT メンバー)は献身的なボランティアです。UTX チームの活動は、AAMT から資金を受けています。

UTX の詳細はどこで見られますか？

パンフレット、仕様、サンプル辞書をダウンロードできます。

他人と用語集を共有したくありません。

それは残念です。ただ、UTX を使用すれば、自分の辞書を共有しなくても、その他の UTX 用語集を自分の用語集に簡単に結合できるという利点 もあります。

UTX 用語集の作成に関して

役立つ用語集を作成するには何千も項目が必要でしょうか？

とんでもありません！ UTX チームの調査結果では、わずか 50 項目の用語集が、4000 語の文書の機械翻訳の品質を向上することが示されました（2011 年にアモイの MT Summit で発表）。UTX 用語集が、人間翻訳者の全体的な作業効率をどのように向上できるかの測定はより困難です。しかし、用語集の利点には、読み手にとって文書をより読みやすくし、より理解を深めることも含まれます。

UTX 用語集には、どのような種類の用語を含める必要がありますか？

基本的に、UTX 用語集には、特定分野の専門用語のみを含めます。このような用語の大部分は複合名詞です。詳細については、パンフレットと仕様を参照してください。

UTX 用語集はどうやって編集できますか？

UTX は、(Microsoft Excel または LibreOffice のような) 表計算アプリケーション、あるいは UTF-8 に対応したテキスト エディター（たとえば Windows に付属する「メモ帳」）で編集できます。

UTX 用語集に文を含めることはできますか？

UTX 用語集に文を含めることはできますが、文は、TMX のような、翻訳メモリー形式のほうが適切に扱えます。絶対に必要でないかぎり、UTX 用語集からは文を除外することを推奨します。一般的に、あまりにも長すぎる用語は、UTX 用語集には含めないようにすべきです。また、用語を一定の長さに収めれば、表形式での列を読みやすくなります。

UTX 用語集は高品質ですか？

UTX 用語集は、項目を手作業で精選し、辞書管理者が検査していることから、高品質であるはずですが、これに対して、自動的に生成された生の用語集データには、翻訳品質の質を落とす多くの不適切な項目が含まれます。この状況は「ビッグデータ、ビッグ ノイズ」とも言えます（ここでのノイズとは「不要な情報」という意味です）。辞書管理

者は、UTX の用語ステータスの属性により、さまざまな用語提出者から収集した用語を承認、または却下できます。

UTX を使用して用語を正規化(ばらつきをなくすこと)ができますか？

はい。詳細な手順は、今後ご紹介する予定です。

自分で UTX 用語集を作成するために AAMT にお金を払う必要がありますか？

いいえ。AAMT は、UTX 仕様の使用に対して対価を求めません。

既存の UTX 用語集を変更、または販売できますか？

用語集のヘッダーに含まれる使用許諾に依存します。UTX 仕様は、用語集の使用許諾を明記することを推奨しています。クリエイティブ コモンズ <<http://creativecommons.org/>> はお勧めです。法的に妥当であれば、辞書にどのような使用許諾を宣言することもできます。自社での内部使用に限定することもできます。しかし、UTX 用語集は、共有すれば、より役立ち、充実したものになります。

1 つの用語に対して、ただ 1 つの訳語を選択することは不可能です。どうしたら「一語一義」を実践できますか？

この原則に抵抗を感じる場合、以下のような問題があるかもしれません。

1. あなたが文書の作成者である場合、特定の用語に対して、複数の意味を持たせていませんか？

技術文書では、あいまいな用語を使わないようにする必要があります。単一の意味を表すのに複数の用語を使用することや、複数の意味を単一の用語で表そうとすることは避けるべきです。

たとえば、"terms"には、「用語」と「合意」の他、多くの意味があります。特に、文章の主要なトピックが「用語」の場合は、「合意」の意味で"terms"という語は使わないようにします。あいまいな用語を使用する必要がある場合、そのような各用語は明確に定義し、それぞれの異なる適切な用法を示すよう区別する必要があります。

2. 1 つの用語集に複数の分野を混在させていませんか？

原則として、1 つの分野には、1 つの用語集を使用します。翻訳プロジェクトで、複数分野、たとえば、医療機器を扱う場合、医学用語集、機械用語集、医療機器自体の用語集、またはその他の用語集を必要とすることがあります。このようなプロジェクト全体で単一の用語集を使用すると、分野の区切りがなく、再利用が困難になります。

複数の分野からの項目が、適切な理由なく1つの用語集に含まれる状況は「分野汚染」と呼ぶことができます。"異なる分野では、異なる用語を必要とします。たとえば、"file"と"window"は、建築分野では「やすり」と「窓」ですが、IT 分野では「ファイル」と「ウィンドウ」です。1 つの分野で 1 つの用語集を管理している場合、1 つの原語に対して 1 つの訳語があれば十分です。

翻訳発注者／翻訳会社向け

(翻訳発注者／翻訳会社として)UTXを使用すべき理由はなんですか？

専門用語が適切に翻訳されずに、失望したことはありませんか？ 用語集を作成するだけでその失望は避けられたかもしれません。

機械翻訳で UTX を使用する計画がなくても、UTX はシンプルな用語管理として活用できます。適切な理解、合意、取り決めを経て、個人翻訳者から新しい訳語候補を収集できます。そこから、自社の用語集を作成して、使用できます。

RBMT(こちらも参照)を使用する計画がある場合、UTX 用語集に基づいて、高品質のユーザー辞書を簡単に作成できます。

書籍やゲームを翻訳しています。どのように UTX を使用できますか？

書籍またはゲーム ソフトウェアでは大量の用語があり、その多くは固有名詞です。人名、地名、技能やアイテムの名前などでしょう。これらは、実際にはすべて専門用語です。" 用語集なしでは、数千語の用語を、数か月にわたる翻訳中、適切に管理できません。書籍の読者なら、バラバラの訳語を見て混乱するでしょうし、ゲームのユーザーならまともにプレイできずに怒るでしょう。また、翻訳には多くの翻訳者とチェッカーが関わるはずです。UTX は、用語ツールを使用しなくても、複数の翻訳者およびチェッカー間での、用語の使用を正規化するために非常に役立ちます。

UTX を使用する翻訳プロジェクトを提案したいのですが？

喜んで！ 連絡フォーム
<<http://www.aamt.info/english/utx/index.htm#contact>>でご連絡ください。

UTX でなぜ翻訳生産性が向上できるのか理解できません。

おそらく、UTX 用語集を再利用していないためです。UTX は、さまざまなユーザー、ツール、および環境間で再利用され共有されるとき、最大の効果を発揮します。

良い UTX 用語集を作成するにはスタイル ガイドを持つ必要がありますか？

一貫性を維持する目的で、スタイル ガイドを持つことを強くお勧めします。英語とその他の言語では、確立された、さまざまなスタイル ガイドを使用できます。日本語では、JTF 標準スタイル ガイド
<http://www.jtf.jp/jp/style_guide/styleguide_top.html>を使用できます。

機械翻訳に関して

なぜ UTX は XML ではなくタブ区切り形式なのですか？

UTX はシンプルであるように設計されています。UTX 用語集は極めてシンプルなので、わずか 3

つの必須指定の列(原語、訳語、品詞)だけで成り立ちます。これらの列を扱うには、XML を使用する必要はありません。

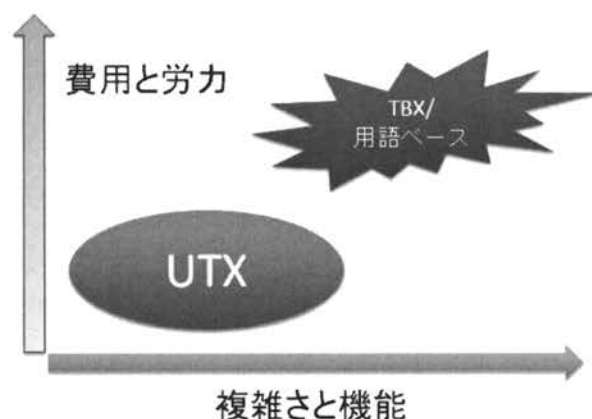
そんなにシンプルなら、なぜファイル形式とする必要があるのでしょうか？

多くのオンライン用語集がウェブで公開されていますが、それらの多くは使いにくいものとなっています。これらの用語集は、最善慣行に従っていません。このような用語集には、一般的に、さまざまな使用状況での優先順位や説明がないまま、類似の項目が含まれています。各項目は整形化されておらず、語の基本形(単数形または原形)を示していません。UTX はシンプルに見えますが、特定の仕様に従うことで用語集として目的を確実に果たせます。

UTX は、TBX、TBX-Basic、またはその他の既存の用語集形式を置き換えるものですか？

いいえ。UTX 用語集は、翻訳者が手作業で選ぶ技術用語の集合として、ゼロから作成できます。UTX は、非常にわずかの作業で作成できます(以下の図を参照)。より大きい翻訳プロジェクトで使う、大規模で複雑な用語ベースの基盤となります。しかし、小・中規模の翻訳プロジェクトでは非常に役立ちます。

UTX と TBX の位置付け



TBX、TBX-Basic、またはその他の既存の用語集形式のなにが問題なのでしょうか？

問題というわけではありません。ただこれらは、多様な用語提出者にとっては複雑過ぎます。用語提出者は、XML やさまざまな用語集形式の詳細に慣れていないかもしれません。ただ適切な訳語を知っているだけのプロ翻訳者であることもあります。

このような知識を実際に使用できる用語集の形で使用できれば、素晴らしいでしょう。

(翻訳ソフトの)システム辞書とユーザー辞書の違いはなんですか？

(ルールベースの)翻訳ソフトは、システム辞書とユーザー辞書の2種類の辞書を使用します。システム辞書は、最良の翻訳結果を得る目的で細かく調整された、定義済みの用語の集合です。ユーザー辞書とは、特定の翻訳プロジェクトでの翻訳品質をさらに向上させる目的で、ユーザーが定義し、追加した用語の集合です。この目的のため、ユーザー辞書の項目は、通常、システム辞書の項目より優先的に使用されます。通常、システム辞書にすでに含まれる項目は、ユーザー辞書からは除外する必要があります。しかし、ユーザーは、ユーザー辞書にこのような用語を意図的に追加することにより、システム辞書の訳語に優先して、より適切な訳語を選択できます。

用語集と(翻訳ソフトの)ユーザー辞書の違いはなんですか？

「用語集」とは、ユーザーあるいはソフトウェアが使用する専門用語の集合です。用語集には、翻訳ソフトでは使用されない用語の「定義」や「詳細」が含まれることがあります(用語集を翻訳ソフトが使用するには、翻訳ソフトに使える形式にする必要があります)。これに対して、ユーザー辞書は、翻訳ソフトで使う目的で作成され使用されます。用語集はユーザー辞書に変換できます。変換直後では、用語集とユーザー辞書の内容はよく似ています。しかし、ユーザー辞書には、人間が使用しない追加の属性または項目が含まれることがあります。一般的に、充実した用語集は、

高品質なユーザー辞書の非常によい元データとなります。

開発者向け（機械翻訳他）

UTX 仕様は、RBMT（ルールベース機械翻訳）を念頭に策定されたのですか？

はい。しかし、UTX はほとんどすべての翻訳ツールと用語ツールで使用できます。

AAMT はなぜ UTX 形式を作成したのですか？ 背景はなんですか？

SYSTRAN のような商用翻訳ソフトパッケージは世界的に知られていますが、AAMT がある日本の翻訳ソフトには、日本人以外のユーザーはなじみがないかもしれません。UTX 仕様は、日本製、または日本語のソフトウェアに、限定されていませんが、UTX が日本で作られた背景を知ると、より深く理解できるはずです。日本には、さまざまな商用 RBMT 翻訳ソフト パッケージがあります。これらのハイエンドのソフトには、出荷時に、7～800 万の基本および技術用語が付属しています。これらのソフトは極めて高機能で、翻訳のさまざまな側面を制御できるように 30 以上のオプションを持っています（SYSTRAN のハイエンドバージョンには、日本語に関して 2 つのオプションしかありません）。これらのソフトは、ユーザー辞書の活用を推測できるため、各用語項目の詳細な属性を与える必要はありません。

それでも、翻訳ソフトでよい翻訳結果を得るには、優れた用語集を必要とします。大規模な辞書群は翻訳品質を向上できることもあります。しかし、辞書品質が適切に維持されなければ、翻訳品質を低下させることもあります。私たちの研究では、UTX 用語集内の用語は、適切に選択されていれば、項目数が少なくても、大きく翻訳品質を向上できることが証明されました。これが、適切な技術用語を翻訳に反映する目的で、AAMT がシンプルな用語集形式を作成した理由です。

当社では **SMT**（統計機械翻訳）を使用しています。用語集は不要です。

はたしてそうでしょうか。SMT を使用しており、翻訳品質を保証する必要がある場合、そのプロジェクトでは用語検証用に個別の過程が必要になります（用語検証は、RBMT ではシステムに統合されています）。翻訳するときに用語集を使用しなくても、品質保証の目的で使用する必要があります。個別の用語検証過程が必要ということは、余分な時間と作業も必要ということになります。

UTX を変換するとき、情報損失のある変換過程になりますか？

場合によります。UTX では、追加の列を定義することにより、どのような情報でも含めることができますが、そうすることは最善とは限りません。多数の追加の属性を保持する必要がある場合、その他の XML ベースの形式の使用を検討したほうがよいことがあります。しかし、ある形式を別の形式に変換するとき、実際には、ごく一部の、特定の属性のみが重要であるはずで

なぜ UTX にはわずかな用語属性しか事前定義されていないのですか？

このような属性は、翻訳の正確さや適切さの向上にはわずかしか貢献しません。複雑さを減らすことのほうがより重要です。

用語集を提供したいです。／変換ツールを作成したいです。

ありがとうございます。連絡フォーム<<http://www.aamt.info/english/utx/index.htm#contact>>でご連絡ください。

UTX 仕様について提案したいことがあるのですが。

喜んでお聞きします！ 連絡フォーム<<http://www.aamt.info/english/utx/index.htm#contact>>でご連絡ください。

翻訳祭感想と UTX 関連の記事について

機械翻訳課題調査委員会ワーキンググループ 3 (UTX チーム)

秋桜舎・山本ゆうじ

UTX チームから

今回、本ジャーナルには UTX チームから以下の 3 点の記事を掲載しています。

- 翻訳祭感想(本記事)
- UTX クイック ガイド(英語版)
- UTX のよくある質問(日本語版)

UTX クイック ガイド(英語版)は、以前に作成したクイック ガイドを全面的に改訂したものです。UTX 用語集はどのようなものか、どのように作ればいいのかを、大きな図を使って、より具体的に解説しています。UTX パンフレットを補完するものです。UTX サイト

<http://www.aamt.info/japanese/utx/>からもダウンロードできます。企業、大学の会員の皆様は UTX の普及にぜひご協力いただき、クイック ガイドを配布していただければ幸いです。

「UTX のよくある質問」(日本語版)は、UTX 仕様には掲載していない、具体的な疑問に答えています。UTX は、特許庁ほか、各方面で浸透しつつあります。残念ながら UTX の特性についてはまだ誤解されていることも多いようです。この「よくある質問」で誤解を解ければと思います。ウェブでは <http://www.aamt.info/japanese/utx/faq.htm> で参照できる他、「UTX FAQ」というキーワードでウェブ検索して見つけることもできます。

翻訳祭感想—プロ翻訳の現場の観点から

2012 年、AAMT 機械翻訳調査委員会は、JTF (日本翻訳連盟) 翻訳祭にブース出展をしました。また、UTX パンフレットとクイック ガイドを配布しました。昨年は私が UTX を紹介しましたが、今回は、ブース出展のみでした。次回は JTF と連携の上、企業向けの UTX のデモができればと考えています。

翻訳業界の情報交換はまだ貧弱

翻訳祭では、翻訳業界の情報交換が不十分で、立場により大きな差があると感じました。翻訳祭の参加者の立場は、翻訳者、翻訳会社経営者、営業担当者などさまざまです。翻訳者だけでも、分野、顧客、単価に幅があります。翻訳者の中で翻訳ツールに取り組む人もいますが、必要性を感じていない人も多いようです。違う立場の人たちで問題意識がしっかりと共有されていないので、何年も同じ問題に苦しんでいます。

用語集がないと用語ツールの意味がないのでは？

また、日本語や訳文の客観評価が必要と感じました。残念ながら誤訳が放置されている例は、相変わらず多いようです。素人の読みにくく、意味が通らない翻訳が、平然と製品に使われています。発注企業が、用語集や表記などの日本語品質にもっと関心を持たなければ、いくら用語ツールがあっても無意味です。ブースでは多数の用語ツールが展示されていましたが、すぐに用語集が作れないとそのようなツールが活用できないのではと感じました。それこそ、シンプルで管理しやすい UTX の出番です。

翻訳品質向上のために発注企業の意識を変えよう

UTX をより活用してもらうには、問題意識を共有して、翻訳品質を向上できるように、発注企業向けイベントの実施が必要と考えます。

AAMT 公式ではありませんが、Facebook の翻訳工学ページでも意見交換していますのでぜひご参加ください。

<https://www.facebook.com/translation.engineering>

UTX Glossary Quick Guide

Let's create your own UTX glossary!

You can use it every day in various environments at no cost.

- UTX is a simple tab-delimited format. You can make it instantly.
- Create, edit, and use with a familiar spreadsheet application, such as Excel (save as text (tab delimited)).
- You can quickly make a practical glossary with minimum information.
- By formatting as terminological data, you can use a glossary in various terminology tools and translation software.
- By accumulating, sharing, and reusing terminological knowledge, you can increase the translation accuracy and reduce the cost.
- The detailed UTX specification and samples are available from <http://www.aamt.info/english/utx/> for free.

Basic Components of UTX Glossary (AAMT/UTX Glossary)

UTX format unlocks your glossary
to be shared and reused in various environments

Information about the glossary (creation date, license etc.)

Glossary information	#UTX 1.20; en-US/ja-JP; 2012-09-25T19:00:00Z+09:00; copyright: AAMT (2012); license: CC-by 3.0			
Column definitions	#src	tgt	src:pos	term status
Entry	Asia-Pacific Association for Machine Translation	アジア太平洋機械翻訳協会	properNoun	approved
	dictionary administrator	辞書管理者	noun	approved
	contributor	用語提出者	noun	provisional
	domain	分野	noun	
	glossary	用語集	noun	
	bidirectional	双方向	adjective	approved
	merge	統合する	verb	approved
	Source term	Target term	Part of speech	Term status (optional)
Singular, basic form, starts with lower case		Single best translation	noun, properNoun, verb, adjective	approved, provisional, non-standard, forbidden

* The current version is UTX 1.11. We are preparing UTX 1.20 for its final release.

Common Errors in a UTX Glossary

Use singular form. Create a separate entry if the plural form has a different meaning.

A term not yet approved is indicated as "provisional."

1 source term for 1 entry. Terms included for compatibility reasons should be indicated with term status as "non-standard."

Lower case if not a proper noun

Single-byte alphanumeric only

Comments on a term should be in a dedicated column

Verbs should be in its basic form ("する" in this case)

If you have multiple candidates for translation, create a separate entry and explain the reason in a comment column. A term from a different domain should be kept in a separate glossary.

#src	tgt	src:pos	term status
Asia-Pacific Association of Machine Translators	アジア太平洋機械翻訳協会	properNoun	approved
dictionary administrators	辞書管理者	noun	approved
contributor, term contributor	用語提出者	noun	provisional
Domain	分野	noun	
glossary	用語集、グロッサリー、語彙集	noun	
bidirectional	双方向	adjective	approved
merge (of glossary)	統合	verb	approved

Source term **Target term** **Part of speech** **Term status**

UTX Glossary Information

The first line of a UTX glossary is called "glossary information." The following is a simple example of how to describe glossary information.

#UTX 1.20; en-US/ja-JP; 2012-09-24T19:00:00Z+09:00; copyright:AAMT (2012); license:CC-BY 3.0



#UTX 1.20; UTX version

en-US/ja-JP; Translation from American English (en-US) to Japanese (ja-JP).

2012-09-24T19:00:00Z+09:00;

2012-09-24 is the creation date. 19:00:00 is Coordinated Universal Time (UTC). If the glossary is created in Japan, the time is Japan Standard Time (JST) minus 9 hours. By adding 9 hours ("+09:00"), the time will be equivalent to the local time (in this case, Japan Standard Time). You can distinguish the latest edition even when you exchange the glossary across the time zones.

copyright:AAMT (2012); Copyright. Optional, but recommended.

license:CC-BY 3.0 License. Optional.

In addition to the above basic information, you can define and add any property. An annotation regarding the entire glossary, such as terms of use should be added to the second line. Comment it out by "#". Use two or more lines to avoid a line becoming too long.

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

株式会社川村インターナショナル/Kawamura International Co., Ltd.

自己紹介

株式会社川村インターナショナルは、創業 26 年目を迎えた翻訳会社です。2007 年からは、多言語の翻訳事業も展開し、現在では 40 カ国語以上の翻訳およびローカリゼーションを提供しています。川村インターナショナルは、「翻訳に関連するビジネスによってお客様の業務が円滑に進むようにサポートする」という理念 (KI Philosophy 1 項) を掲げており、現在では翻訳事業のほか、通訳事業、派遣事業へも領域を広げ、マニュアル、Web、デザイン、マルチメディアなどの制作も提供しています。また、グローバルに展開する大企業のみならず、海外進出を志向している中小企業のサポートを強化するため、オンライン翻訳サービス「メール君」や「貿易本舗」など、新たな市場への挑戦も開始しています。

私たちのもう一つの特徴は、経営理念でも「翻訳およびサービスの品質を重視し、常に改善を怠りません」(KI Philosophy 2 項)と宣言している通り、翻訳の品質改善のほか、お客様の「好み」の文体についての解析、プロジェクトマネジメントの品質、翻訳者採用時のパターン分析など、独自の科学的アプローチを用いて改善を図っていることです。

今年、川村インターナショナルは法人会員として AAMT に参画し、代表取締役川村みどりが監事に就任いたしました。どうぞよろしくお願い申し上げます。

MT/翻訳とのかかわり

MTおよび翻訳業界に期待すること

翻訳というしがらみから解放されて、一人のユーザとして機械翻訳に対峙すると「こういう使い方があったのか!」と驚くことが多々あります。翻訳を事業として展開する弊社においては、今まで我々にとって機械翻訳が翻訳として「使えるのか、使えないのか」の話題の対象にしかなくておらず、「どこなら使えるのか」という目線を欠いていたと感じています。精度はどうあれ、最近では機械で翻訳された日本語や英語に触れる機会が増えています。Web サイトを閲覧する、旅行サイトや EC サイト上でユーザ評価を情報提供する、あるいは Facebook、twitter などの SNS でメッセージを翻訳するなど、本当にさまざまな場所で、機械翻訳に遭遇することが多くなりました。

これらの翻訳の中には、すでにユーザの一定の目的を達成するレベルにまで到達しているものもあります。

機械翻訳を活用して翻訳生産性を向上させ、コストの低減と納期の短縮を実現するという市場からの期待は、翻訳会社にとっては避けては通れない命題になってきていると感じています。川村インターナショナルが翻訳会社として今後積極的に関わってゆくのは、機械翻訳を活用してマニュアル翻訳でポストエディットを実施する領域だと考えていますが、SNS の翻訳にせよ、マニュアルの翻訳にせよ、アウトプットの翻訳品質と同様に重要視されるのは、ワークフローの（半）自動化だと考えています。コーパス/用語集の管理、作成、クリーニングとトレーニングから、機械翻訳、ポストエディットの実施、そしてコーパス/用語集へのフィードバックが一連のワークフローとして提供するサービスに組み込まれ、品質向上のサイクルに乗っていることで、初めてユーザからは「使える」との評価を得ることができ、投資してもらえる対象になると考えます。そこにはエンジンのみではなく、ポストエディットに使用するエディタや、用語集/TM の形式、そして対応する人材など、翻訳会社として提案できることがたくさん隠されていると日々思いを強くしています。

機械翻訳についてより理解を深め、強み/弱みを把握した上で、お客様の課題の中で「どこなら MT を使ってもらえるか」を正しく把握し、機械翻訳、MT+PE、そして人力翻訳など、課題に応じた最適なソリューションを提供できるようになることが、今後しばらくの弊社と機械翻訳との付き合い方になると信じています。

株式会社川村インターナショナル 執行役員・ゼネラルマネージャ 森口功造

AAMTへの要望

弊社は本年より本協会に監事として参画することになりました。よろしくお願い申し上げます。協会の皆様との交流を経て、機械翻訳エンジンの性能のみならず、翻訳の生産性を向上させるあらゆる要素について情報の共有ができる機会が得られればと考えています。

協会活動報告

(2012 年 9 月～2013 年 3 月)

機械翻訳課題調査委員会

2012 年 10 月 5 日 (2012 年度 第 5 回)

- ① 前回委員会の議事録の確認
- ② 各 WG の活動について (各 WG に分かれて議論)
(WG3)
 - ・今年度の学会発表の取り組みについて
 - ・ UTXTool の提案
 - ・ ゲスト講演者のついて
- ③ 活動内容の報告 (各 WG から)
- ④ 活動内容についての議論
 - ・ JTF 翻訳祭 2012 の参加について
 - ・ UTXTool の提案
 - ・ ゲスト講演者のついて
- ⑤ まとめと次回委員会について

2012 年 11 月 9 日 (12 年度 第 6 回)

- ① 前回委員会の議事録の確認
- ② 各 WG の活動について (各 WG に分かれて議論)
(WG1)
 - ・ AAMT/Japio 研究会との協力について
 - ・ 2012 年度活動(学会参加の可能性、WEB 上に自動評価サイトを設定することについて)
(WG2,WG3)
 - ・ JTF 翻訳祭に出展に関する検討
 - ・ 配布物の確定 (AAMT 紹介資料、UTX 資料など)
 - ・ アンケート関連の検討 (アンケート形式&景品など)
- ③ 活動内容の報告 (各 WG から)
- ④ 活動内容についての議論
- ⑤ まとめと次回委員会について

2012 年 12 月 14 日 (2012 年度 第 7 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)
 - ・ UTX 用語集作成クイックガイド

- ・ UTX 基本方針の確認
- ・ UTX 1.20 仕様検討
- ③Acrolinx 柳英夫氏によるご講演
 - ・「Acrolinx IQ の紹介」
- ④合同討議
 - ・ 翻訳祭で行ったアンケート速報
 - ・ 翻訳祭における AAMT 広報に関するコメント

2013 年 1 月 25 日 (2012 年度 第 8 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)
 - ・ AAMT ジャーナル掲載内容の確認
 - ・ UTX 1.20 仕様検討
 - ・ 学会発表検討
- ③講演:「Transit の紹介」(シュタールジャパン目次氏)

2013 年 3 月 8 日 (2012 年度 第 9 回)

- ① 前回委員会の議事録の確認
- ② 機械翻訳アンケートの実施についての検討
 - ・ アンケート実施サイトの検討
 - ・ アンケート項目の検討
 - ・ アンケート収集・集計の検討
- ③ まとめと次回委員会について

編集委員会

2012 年 9 月 24 日 (2012 年度 第 2 回)

- ① AAMT Journal No.52 の進捗状況について
- ② AAMT Journal No.53 の記事の執筆依頼計画について
- ③ AAMT Journal No.53 のスケジュールについて

2013 年 2 月 19 日 (2012 年度 第 3 回)

- ① AAMT Journal No.52 の 進捗状況について
- ② AAMT Journal No.53 の 記事の執筆依頼計画について
- ③ AAMT Journal No.53 のスケジュールについて

インターネット WG

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

AAMT/Japio 特許翻訳研究会

2012 年 10 月 12 日（金）（2012 年度 第 4 回）

1. 前回議事録の確認
2. 『特許文書の機械翻訳結果評価方法検討会』結果報告（越前谷委員、事務局）
3. 第 2 回特許情報シンポジウム関連（横山副委員長、Japio）
4. 特許・情報フェア & コンファレンス関連（Japio）
5. 次回の開催について

2012 年 12 月 7 日（金）（2012 年度 第 5 回）

1. 前回議事録の確認
2. 第 2 回特許情報シンポジウム関連
3. 機械翻訳結果評価関連（江原副委員長）
4. MTサミット関連（横山副委員長）
5. 中日特許翻訳プレスリリース関連（Japio）
6. 平成 24 年度活動報告書関連
7. 事務局連絡
8. 次回の開催について

2013 年 3 月 8 日（金）（2012 年度 第 6 回）

1. 前回議事録の確認
2. 平成 24 年度活動報告書関連
3. 拡大評価部会報告（江原副委員長）
4. 研究報告「日英パテントファミリーにおける対訳文対非抽出部分を利用した専門用語訳語推定」（宇津呂委員）
5. MT サミット関連（横山副委員長）
6. 次回の開催について

2012 年 9 月 7 日（金）特許文書の機械翻訳結果評価方法検討会

2012 年 11 月 30 日（金）第 2 回特許情報シンポジウム

2013 年 1 月 11 日（金）（2012 年度 拡大評価部会）

1. 自動評価の整理
2. 人手評価の整理 — intrinsic な評価と extrinsic な評価
3. 人手評価と自動評価の組み合わせ
4. 人手評価作業の補助ツール
5. テストセット構築の方法（特に中日）
6. その他 I
7. その他 II

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 53 号をお送りします。

ちょうど 1 年前の 2012 年度 AAMT 総会におきまして、前会長の井佐原均先生から現会長の中岩浩巳氏へと会長が引き継がれました。今号の巻頭言は、「機械翻訳の発展と普及に向けて」と題しまして、中岩会長より抱負を語って頂きました。また、前会長の井佐原均先生には、会長を終えられた感想とともに、今後の AAMT のあり方について考え方を語って頂きました。

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ご寄稿を頂きました。一方、フリーランスの翻訳者 Alun Simpson 氏からは、翻訳作業における機械翻訳の利用についての考えや願望をご寄稿頂きました。

さらに、プロジェクト報告として、NTT コミュニケーション科学基礎研究所の永田昌明氏、須藤克仁氏、塚田元氏より、「日本語の主辞後置性を利用した統計的機械翻訳技術」の最新の研究動向についてご寄稿頂きました。

その他、国際・国内会議の参加報告を、3 件ご寄稿頂きました。Workshop on Future Directions in Translation Research (未来の翻訳研究に関するワークショップ)、および、第 2 回特許情報シンポジウムへの参加・出展報告をご寄稿いただきました。また、機械翻訳課題調査委員からは、JTF 翻訳祭へ出展した際の様子についてご寄稿頂きました。

一方、AAMT 内の活動報告として、機械翻訳課題調査委員会から、機械翻訳用ユーザー辞書共通フォーマット UTX (universal terminology exchange)用語集の FAQ をご寄稿頂きました。

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

AAMT

Asia-Pacific Association for Machine Translation

AAMT 入会のご案内

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

機械翻訳にご関心のあるすべての方にご入会をお勧めします。

* * AAMT 会員の特典 * *

1.AAMT Journal の購読ができます。

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

2.機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機械翻訳関連の情報をお届けします。

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

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

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

4.関連機関の主催する国際会議に参加できます。

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

AMTA* : Association for Machine Translation in the Americas

EAMT** : European Association for Machine Translation

年会費は以下の通りです。

法人会員：入会金	1 口	10,000 円	年会費	1 口	50,000 円
個人会員：入会金		1,000 円	年会費		5,000 円（学生は学生会費 1,000 円）

ご関心のある方は、事務局までお問い合わせください。

アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

電子メール：aamt-info@aamt.info



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。

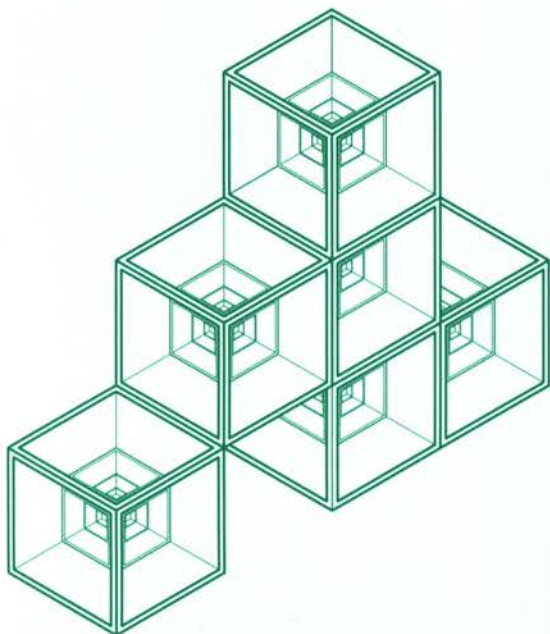
申込日 20 年 月 日

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.53

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

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

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