

AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2014

No.56

アジア太平洋機械翻訳協会

目 次

巻頭言：	翻訳のイノベーション.....隅田 英一郎.....1
特別寄稿：	自然言語処理の過去、現在、将来.....長尾 真.....3
研究報告：	情報通信研究機構における長文機械翻訳の実用化.....内山 将夫.....5
言語資源紹介：	アジア学術論文抜粋コーパス ASPEC.....中澤 敏明.....9
連載：	機械翻訳事始め[Ⅲ].....坂本 義行.....13
シンポジウム報告：	機械翻訳の現状と展望.....内田 真弓.....23
シンポジウム報告：	取説新時代に向けてのテクニカルコミュニケーターの役割.....山崎 敏正.....28
シンポジウム報告：	自然言語処理の発展に向けた情報共有・討論.....岩倉 友哉.....36
製品紹介：	外国語特許調査を日本語で行える MT Plus II /NEF.....本間 奨.....39
研究会活動報告：	AAMT/Japio 特許翻訳研究会拡大評価部会の活動について.....江原 暉将.....43
委員会活動報告	これまでの AAMT Forum メールマガジン.....機械翻訳課題調査委員会 WG1,2.....45
委員会活動報告：	言語処理学会第 20 回年次大会での発表報告.....機械翻訳課題調査委員会 WG1,2.....50
委員会活動報告：	UTX Converter (UTX 変換ツール) 仕様書.....機械翻訳課題調査委員会 WG3.....55
AAMT 会員のひろば：	春遍 雀來.....59
事務局からのお知らせ：	協会活動報告（2014 年 4 月～2014 年 5 月）.....AAMT 事務局.....61
編集後記	63

C O N T E N T

Foreword:	Translation Innovation	<i>E. Sumita</i>	1
Special report:	Past, Present and Future of Natural Language Processing	<i>M. Nagao</i>	3
Report:	Deployment of NICT Machine Translation for Long Sentences	<i>M.Utiyama</i>	5
language resources:	Asian Scientific Paper Excerpt Corpus, ASPEC	<i>T. Nakazawa</i>	9
Series:	History of Machine Translation [III]	<i>Y.Sakamoto</i>	13
Symposium Report:	Present situation and future prospects of Machine Translation	<i>M. Uchida</i>	23
Symposium Report:	The role of Technical Communicator for the manual new era	<i>T. Yamazaki</i>	28
Symposium Report:	Introduction to the workshop of the 20th annual meetings of the Association for Natural Language Processing.....	<i>T.Iwakura</i>	36
Products & Services:	MT PlusII/NEF spread which can conduct foreign patent research in Japanese ..	<i>S.Honma</i>	39
Committee Report:	Research Activities of the AAMT/Japio Special Interest Group on Patent Translation, MT Evaluation Working Group	<i>T.Ehara</i>	43
Committee Report:	AAMT Forum mail magazine vol.2	<i>WG1, WG2</i>	45
Committee Report :	Report on questionnaire results of using machine translation for business translations	<i>WG1, WG2</i>	50
Committee Report :	UTX Converter Specification	<i>WG3</i>	55
AAMT Members:		<i>J.Halpern</i>	59
AAMT Activities:	AAMT Activities (from March r 2014 to May 2014)		61
Editor's Note:		<i>T. Utsuro</i>	63

翻訳のイノベーション

隅田 英一郎

独立行政法人情報通信研究機構(NICT)

eiichiro.sumita@nict.go.jp

五輪の東京開催という追い風

昨年の9月8日に2020年五輪の東京開催が決定しました。これを日本の再浮上に活用しようという活動が日本中で盛り上がっています。

今年1月に、今回の東京五輪の経済効果は19.4兆円と竹中平蔵氏が試算したと報道されており、同試算の中では訪日外国人の増加が重要な要素と想定されています。

日本語という言葉の壁とこの壁を越えるための翻訳産業と自動翻訳技術がクローズアップされることは必定です。例えば、日本政府は、東京五輪に間に合うように先端技術の研究開発を前倒し、未来を先取りすることを狙うことが報道されており、その中では、文化や背景なども理解する自動翻訳技術の開発を9年前倒しするとうたわれています。

翻訳産業と自動翻訳技術に強い追い風が吹いています。

JTF と AAMT の昨年来の接近

アジア太平洋機械翻訳協会(AAMT)の総会では、翻訳会社や翻訳者の方々の参加が全体の6割にのびました。翻訳産業側に自動翻訳技術の現状はどうなっているのか、どこまで使えるものなのか、と関心が高まっていることの表れでしょう。

日本翻訳連盟(JTF)総会後の講演会と懇親会に参加された中岩 AAMT 会長と翻訳産業の方々との間で活発な意見交換がなされたようです。また、JTFの翻訳祭ではAAMTのセッ

ションが設けられました。これは自動翻訳技術の研究開発側からのアプローチと言えます。

果たして、翻訳産業側と自動翻訳技術の研究開発側の双方の相手側への接近は、両者の間の壁を壊して、果実を実らせるのでしょうか？

自動翻訳の翻訳品質の再評価

自動翻訳システムの翻訳品質に関して相反する意見があります。翻訳対象となる言語対が何か、対象が専門的か一般的か、内容をざっと把握するために利用するのか、人間による翻訳の下訳として利用するのか、と様々な条件の影響を受けるからでしょう。

翻訳産業側では自動翻訳システムの翻訳品質について今まで否定的な意見が目立つことが多かったのですが、一方で、翻訳実務では、専門用語の辞書引きや訳語の統一のために自動翻訳システムが活用されてきました。また、先日、第一級の翻訳者の方から、英日の自動翻訳システムの翻訳品質は、かつては箸にも棒にもかからなかったけれど、最近、目に見えて良くなったと伺いました。

大幅な品質向上の理由

この背景には技術革新があります。翻訳例を集積したデータベース(対訳コーパスと呼びます)に基づく翻訳手法が創出された後、熾烈な研究競争が行われ、アルゴリズムが日進月歩しているからです。

仏英や韓日などの似た言語対では、早期から高い翻訳品質を実現していましたが、英日のように語順が大きく異なる言語対では翻訳品質が低く、年々、改善を示しつつも長い間従来技術の翻訳品質を上回ることができませんでした。

2010 年以降に提唱された一連の語順変換と訳語選択を分離する技術が奏功し、英日の特許文の場合、2013 年 6 月開催の第 10 回 NTCIR の PatentMT ワークショップで、対訳コーパスに基づく翻訳手法の翻訳品質が従来法の品質を上回ることが確認されました。今後も差が広がっていくでしょう。

ALL JAPAN 対訳コーパスの構築

対訳コーパスに基づく翻訳手法の翻訳品質の向上には二つの柱があります。①良いアルゴリズムの創出と②対訳コーパスの増量です。アルゴリズムは同じでも、対訳コーパスの増量で翻訳品質を上げることが出来ます。

この翻訳品質への貢献が大きい対訳コーパスを作成する方法はいろいろあります。新規に翻訳する、既存データの権利者に利用許諾をいただく、WEB からクロールする。あらゆる方法を駆使して集められています。

ここで、日本の翻訳産業全体の生産している対訳コーパスの量を考えてみましょう。翻訳産業の年間総売り上げは 2000 億円とされています。1 文字 10 円とすると 200 億文字になります。1 単語 2 文字として 100 億語。1 文 20 単語とすると 5 億文です。これは極めて膨大な量です。これ使って、ALL JAPAN の自動翻訳システムを構築すれば相当の高品質が期待できます。

残念ですが、現時点では、「夢物語」です。

翻訳会社や翻訳発注元の会社は、翻訳結果を自動翻訳システムの訓練のために提出するこ

とに消極的で、毎年膨大な対訳コーパスが死蔵されていくか、翻訳メモリーとして自社で局所的に活用されるに留まっています。このような小規模対訳コーパスの有用性は限定的です。ネットにつながらないスマフォと同じです。この対訳コーパスを日本全体でネット化して活用すれば、日本は高品質の自動翻訳システムを手に出、翻訳産業の高効率化、翻訳の低価格化、新分野の翻訳市場の開拓が実現するでしょう。

このビジョンに敗者はいません。

自動翻訳システムの高品質化ため対訳コーパスを再利用するポジティブ・スパイラルが実現して、5 億文/年のペースで進めば、将来は、想像を絶する高品質な自動翻訳システムが出現するでしょう。小さな対訳コーパスのネット化による共有が大きな市場を生むのです。

ステークホルダーが決断するだけで、「正夢」になります。

イノベーションが実現する

日本の翻訳産業と自動翻訳技術の研究開発に、このイノベーションは本当に起こるのでしょうか？イノベーションのジレンマに囚われて、縮小しつつ何もしないのでしょうか？

私には、イノベーションの可能性に気付いた人々が発している色々な兆しが見えています。オセロゲームのように一気に局面は変わり、この夢が 2020 年以前に実現するでしょう。異業種からも参入して翻訳産業は活況を呈するでしょう。私はこのイノベーションを楽しむ予定です☺

自然言語処理の過去、現在、将来

長尾 真

京都大学名誉教授

1. 自然言語処理の歴史

コンピュータは第二次大戦の末期に暗号解読のために作られたが、これが計算のみならず言語の翻訳に使えるだろうと気づかれた。例えば英語とフランス語の間では単語を入れ替えるだけでかなり分かるというわけである。したがって最初は単語の形態素解析が中心であったが、すぐに語順、つまり文法構造を解明しなければだめだということになっていった。

こうして機械翻訳 (MT) への関心は米英の研究者の間で高まり、1957年にソ連が世界で初めて人工衛星スプートニクを打ち上げたショックもあって、米国は露英 MT の研究にかなりの研究費を投入した。しかし1965年になっても成果はあまり上がらず、MT は時期尚早という結論となり、もっと基礎的に計算言語学 (NLP) の研究をすべきであるということになった (ALPAC 報告書、1966)。

それ以来 NLP の研究は世界各国で始められるとともに、その世界会議 (計算言語学国際会議、COLING) が1965年に始められ、2年ごとに世界各国で開催されることになった。日本では1980年に東京で開催され、日本の研究者の増加につながった。1994年には京都で開催されている。自然言語に関する国際会議はCOLING以外に TMI (MT の理論と方法論国際会議)、MT サミット、TAUS (Translation Automation Users' Society)、その他がある。MT サミットは科学技術庁の日英 MT

プロジェクトの終了に伴って、その成果を世界に示すために始めたもので、第一回を1987年に箱根で行った。

日本における MT は1959年に電気試験所の「やまと」がパリでの IFIP (情報処理国際会議) で発表されたほかに、九州大学の日英独3か国語相互翻訳のシステムがその頃に作られた。関連する活動としては1960年に情報処理学会、1970年に電子工業振興協会に MT 研究会、1982年から4年間の科学技術庁の日英 MT プロジェクト、1990年に AAMT、1991年に IAMT、1994年に言語処理学会の発足などがあって、日本の NLP の研究が世界と肩を並べて進展してきている。

MT を実用化する努力はヨーロッパ共同体で1960年代から行われてきていて、英語とフランス語間の翻訳はほぼ実用されていると言ってよいだろう。フランス語とイタリア語間の翻訳も同様である。しかし多言語間の翻訳システムの開発は努力がなされているとはいえ、困難が多い。日本の場合も日韓 MT はかなり良質であるが、その他の言語については一層の努力が必要である。これから挑戦すべきは日中 MT であろう。

NLP のもう一つの分野である情報検索関係では、H. Luhn の KWIC と重要語の抽出、G. Salton の SMART システムなどの成果があり、医学論文の検索システム MEDLARS が1964年に作られたほかに、1965年には T. Nelson のハイパーテキスト、

1967年にブラウンコーパス、1970年には E. Codd のリレーショナルデータベースといった成果が米国であったが、日本ではこの分野に見るべきものがなかった。

2. 言語のモデル研究の種々

言語のモデルとしては1950年代からオートマトン理論が発展し、1957年には N. Chomsky の有名な文法の3階層理論が出され、その後種々の文法理論が提案されてきた。これらの多くは今日実際の言語の解析に有効に使われているとは言い難いが、この分野の研究者にとっては知っておくべきものである。

NLP の研究分野は基礎から応用まで非常に広い。単に文法理論だけに興味を示すのではなく、言語学一般、形態素、文法、意味、文脈、音声など、全ての分野において何が問題となっているか、どこまで進んでいるかといったことに関心を持つ必要がある。ある分野の成果と他の分野の成果を組み合わせれば新しい成果が得られるということもあるから、常に広い視野で解決すべき課題について考えるべきであろう。NLP の種々の研究分野については文献1, 2を参照されたい。

NLP 研究を行う環境は今日かなり整備されてきている。処理のためのツールとしては、多くのソフトウェアがオープンソフトとして入手できるし、各種の電子辞書のほかに、種々のテキストコーパスが利用できるようになってきている。中でもワードネットや EDR の辞書とシソーラスなどは重要で役に立つ。

3. これからの課題

これからの課題としては、大規模テキストコーパスの解析によって、言葉の意味とは何かについてコンピュータ処理の立場から考え役に立つものを作ることや文法規則の再構築（これは確率的に示すべきだろう）、信頼性の高いシソーラスを作ること、豊かな用例を持った辞書、対訳辞書などを整備すること、さらには知識システムの作成にまで発展することが望まれる。これはこれから重要となる対話システムの作成に欠くことのできないものである。

MT は用例主導 MT と統計 MT とが大規模テキストデータの利用によって飛躍的な進展をみせているので、あと10年もすれば最も難しいと言われている日中 MT においても質の良いシステムが作られるだろう。2020年に東京で開催されるオリンピック・パラリンピックでは東南アジアを含む世界の多くの言語間の音声 MT を実現させなければならないだろう。今度こそは期待を裏切らないようにしなければならない。

（以上は2014年3月18日札幌で開催された言語処理学会大会の招待講演「NLP の過去、現在、未来」（長尾真）を基として少し付け加えた）

文献

- 1) 言語処理学事典、共立出版、2009
- 2) Oxford Handbook of Computational Linguistics, Oxford University Press, 2003

情報通信研究機構における長文機械翻訳の実用化

内山将夫

(独) 情報通信研究機構

1. はじめに

統計的機械翻訳 (SMT) は、1990年代の初頭に単語ベースのモデルが提案[1]され、その後、フレーズベースのSMT、構文構造を利用したSMTに発展した。

当初のSMTでは、日本語から英語への翻訳や、英語から日本語への翻訳という、日本人にとって非常に重要な言語間の翻訳を精度良くすることは不可能だった。その理由は、次の2つである。

まず、SMTには、大規模な対訳データが不可欠だが、一般に研究利用可能な大規模日英対訳データは2008年のNTCIR-7の特許翻訳タスク[2]まで、存在しなかった。次に、日本語と英語は、文の構造が非常に異なるため、たとえ、数百万文の対訳データがあったとしても、単純なSMTの手法を適用するだけでは、精度の良い翻訳は不可能である。

これら2つの問題は、旅行会話翻訳[3]等の短文の翻訳ではそれほど問題にならないが、長文の翻訳に対しては、大きな問題である。以下では、(独) 情報通信研究機構 (NICT) における長文のSMTの実用化について述べる。

2. 長文翻訳における課題

長文翻訳の実用化には、対訳の収集、構文構造の違いへの対処に加えて、翻訳モデルのサイズと翻訳速度も重要である。

具体的には、現状、128GBメモリ程度の計算機であれば比較的簡単にレンタルできることから、128GB程度のメモリでストレスなく動くような翻訳モデルのサイズとする必要がある。また、平均23単語程度の文について、最低限、1文あたり1秒程度で翻訳できる必要がある。

これらの課題を満たすことは、後述する特許翻訳においては、1000万文を超える対訳文をSMTの訓練に利用するため、それほど簡単ではないが、現状のSMTの方式のなかでは、フレーズベースのSMTは問題なくこれを満たすことから、NICTではフレーズベースのSMTを長文翻訳のエンジンとして採用している。

3. 長文翻訳方式

もっとも、従来のフレーズベースのSMTだけでは、前述のように、長文については、日英の言語間の構造の違いには対処できない。この問題に対して、最初の解決が、文献[4]で提案された。この方法は、英語を日本語に翻訳するときに適用されたが、そのときに、日本語では主辞が後ろに来ることから、入力 of 英語を構文解析して、主辞を後ろに移した後で、フレーズベースのSMTを適用するという手法である。

この方法を本稿では語順変換と呼ぶが、語順変換のあとでSMTを適用することにより、英日特許翻訳においては、規則ベースの機械翻訳を上回る性能が達成可能なことがNTCIR-10で確認されている[5]。

NICTでは、語順変換+フレーズベースSMTが、精度と速度とサイズの観点から、現状では実用化に最適と考えて、この手法を採用している。ただし、NICTは、英日翻訳に限定することなく、どのような言語対であっても、原言語の構文解析が利用できる場合には適用可能な手法を開発している。

4. 長文翻訳の実用化

SMTでは、同一アルゴリズムを、異なる対訳データに適用することにより、異なる分野に対応した翻訳エンジンを作成可能である。

NICTでは、中日特許翻訳[6]、英日特許翻訳[7]、英日ITマニュアル翻訳[8]について、それぞれ、日本特許情報機構(JAPIO)、日本発明資料株式会社(ニッパツ)、川村インターナショナル株式会社(KI)と共同開発した。

JAPIOと共同開発した中日特許翻訳エンジンの翻訳率(翻訳者の判定により意味が通じると判定された文の割合)は、従来技術の3倍以上である。JAPIOは、このSMTエンジンを利用して、中国特許を日本語に翻訳し、データベース化し、特許検索事業としてサービスを開始した[6]。

ニッパツと共同開発した英日特許翻訳エンジンは、特許要約1件あたりの訳語誤り数を、従来技術と比べて、約12分の1にした。ニッパツは、この技術を利用して、英語特許を対象にして自動翻訳サービスを事業化した[7]。

KIと共同開発した英日ITマニュアル翻訳エンジンは、機械翻訳の結果を人手で後編集することにより、人手翻訳のコストを30%削減可能である[8]。

これらの共同開発は、NICTのコアのSMT技術に、JAPIO・ニッパツ・KIの翻訳資源やノウハウが加わることにより、可能となったものである。

5. 残された課題

英日翻訳に関するSMTの翻訳精度は、ここ数年で飛躍的に改善したが、SMT全般に関する課題は多い。逆に言えば、SMTの翻訳精度の改善の伸びしろは大きい。

まず、SMTの大きな利点は、多言語についての横展開が比較的容易なことである。NICTでは、多言語コーパスの収集・作成のために、文アライメント技術[9]を多言語展開している。また、英語等を仲介言語とするピボット翻訳[10]についても研究している。

次に、翻訳品質の観点からは、訳抜け・湧き出しが大きな問題である。関連する話題としては、日英翻訳において、日本語におけるゼロ代名詞を、英語側で訳出しないといけないというものがある。

現状、入力文に表出されている情報については、それを訳出することはかなりできるようになっているが、ゼロ代名詞のように表出されていない情報の訳出は、今後の課題である。

最後に、SMTエンジンのカスタマイズも重要である。現状、SMTエンジンのカスタマイズには、ある程度の専門知識が必要であり、そのことがSMTエンジンの普及を妨げている一因と考えている。

7. おわりに

SMTによる自動翻訳サービスは、用途を限定すればすでに高精度なものが実用化されている。しかし、汎用的で高精度な多言語自動翻訳は、現時点では存在しない。この実現は大きなチャレンジだが、NICTでは、SMTの基礎研究と実用化への開発とパートナーとの協力により、高精度翻訳ができる言語対と分野を増やしていく予定である。

謝辞

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

参考文献

[1] Peter E. Brown; Stephen A. Della Pietra; Vincent J. Della Pietra; Robert L. Mercer. *The Mathematics of Statistical Machine Translation: Parameter Estimation*. Computational Linguistics, Vol. 19, No. 2, 1993.

[2] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, and Takehito Utsuro. *Overview of the Patent Translation Task at the NTCIR-7 Workshop*. NTCIR-7, 2008.

[3] 多言語旅行会話音声翻訳は、NICTからFEAT社にライセンスされ、VoiceTra+という名前で運用されている。詳細は次を参照

<http://voicetra-plus.jp/voicetra.html>

[4] Hideki Isozaki; Katsuhito Sudoh; Hajime Tsukada; Kevin Duh. *Head Finalization: A Simple Reordering Rule for SOV Languages*. WMT-10, 2010.

[5] Isao Goto, Ka Po Chow, Bin Lu, Eiichiro Sumita and Benjamin K. Tsou. *Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop*. NTCIR-10, 2013.

[6] 独立行政法人情報通信研究機構、一般財団法人日本特許情報機構. NICT の高精度な中日自動翻訳ソフトウェアが Japio のサービスに (2013/3/28)

<http://www.nict.go.jp/press/2013/03/28-1.html>

[7] 独立行政法人情報通信研究機構、日本発明資料株式会社. “英語特許文”の高精度「自動翻訳ソフトウェア」を開発 (2013/3/21)

<http://www.nict.go.jp/press/2013/03/21-1.html>

[8] 独立行政法人情報通信研究機構、川村インターナショナル株式会社. “IT 英文マニュアル” 専用の「自動翻訳ソフトウェア」を開発 (2013/10/1)

<http://www.nict.go.jp/press/2013/10/01-1.html>

[9] Masao Utiyama; Hitoshi Isahara. *Reliable Measures for Aligning Japanese-English News Articles and Sentences*. ACL-03, 2003.

[10] Masao Utiyama; Hitoshi Isahara. *A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation*. NAACL-07, 2007.

アジア学術論文抜粋コーパス ASPEC

中澤 敏明

科学技術振興機構

1. はじめに

本稿では、科学技術振興機構（JST）より公開されたアジア学術論文抜粋コーパス（Asian Scientific Paper Excerpt Corpus, ASPEC）について、その作成・公開の経緯や特徴などについて述べる。ASPEC は、約 300 万対訳文からなる日英論文抄録コーパス（ASPEC-JE）と約 68 万対訳文からなる日中論文抜粋コーパス（ASPEC-JC）とで構成される、世界初の大規模な論文対訳コーパスである。本コーパスを広く利用してもらうことで、機械翻訳に限らず、科学技術論文を対象とした様々な自然言語処理技術の向上に寄与し、延いてはアジア全体での科学技術交流および発展を促進することが、本コーパス公開の目的である。

現在は日英中の 3 言語が含まれるのみであるが、今後様々なアジア言語（例えば韓国語、ベトナム語、タイ語、インドネシア語、ミャンマー語など）のコーパスも追加していきたいと考えて、名前に Asian が入っている。

本コーパスを利用する際には、ASPEC の Web サイト (<http://orchid.kuee.kyoto-u.ac.jp/ASPEC/>) にあるデータ利用規約同意書の内容に同意する必要がある。なお商用での利用は不可となっている。

以降の章ではまずコーパス作成の経緯を簡単に述べ（2 章）、日英・日中各コーパスの特徴について述べる（3 章）。4 章では ASPEC を利用して行われる評価型翻訳ワークショップ WAT について簡単に紹介し、最後に本稿をまとめる。

2. コーパス作成の経緯

ASPEC は、2006 年度から 2010 年度に実施された、科学技術振興調整費による重点課題解決型研究

「日中・中日言語処理技術の開発研究」プロジェクトの成果の一部である（最終成果の全容は「機械翻訳技術のイノベーション」シンポジウムにて報告された）。このプロジェクトの目的は、専門用語などの大規模対訳辞書の作成と科学技術論文に対応した機械翻訳システムの構築により、アジア諸国と日本の科学技術交流を促進・発展させることであり、その研究課題の一つとして翻訳システムでの利用を想定した対訳コーパスの作成があった。

プロジェクトの開始当初は比較的リソースを集めやすい日英の対訳コーパスを作成し、これを用いた辞書構築手法や翻訳システムの有効性の検証を行っていた。これと平行して日中の対訳コーパスを徐々に作成し、最終的に日中・中日の大規模辞書や機械翻訳システムの構築を行った。

日英対訳コーパスに関しては既に様々な科学技術分野の対訳抄録が存在し、そこから対訳文を抽出することで作成したため、コーパスにはあらゆる科学技術分野の文が含まれている。なお、元となる対訳抄録は日本語から英語に翻訳されたものを使用している。一方、日中対訳コーパスに関してはそのような既存の資源がなかったため、日本語文献を人手で中国語に翻訳することで作成した。あらゆる分野を網羅的に翻訳することが困難であったため、中国における科学技術重要分野などの調査と、日本国内の研究者・技術者などの文献データベースの利用傾向から「医学」「情報」「生物」「環境」「化学」「材料」「農学」「エネルギー」分野に絞って翻訳を行い、2,700 万字の日中対訳コーパスを作成した。

3. コーパスの特徴

本コーパスは機械翻訳での利用を想定し、表 1 に

示すように「訓練データ」「開発データ」「開発試験データ」「試験データ」の4つの部分に分けられているが、本コーパスの利用目的は機械翻訳に限らない。以下、日英・日中の各コーパスの特徴を述べる。

表 1：ASPEC に含まれる対訳文数

データ種別	ASPEC-JE	ASPEC-JC
訓練データ	3,008,500	672,315
開発データ	1,790	2,090
開発試験データ	1,784	2,148
試験データ	1,812	2,107

3.1 日英論文抄録コーパス (ASPEC-JE)

訓練データは、JST 所有の約 200 万件の学術論文日英抄録対から、内山・井佐原の方法[1]により、情報通信研究機構 (NICT) が作成したものである。抄録対から文単位の対応を自動抽出することで作成されたコーパスであるため、各対訳文は必ずしも完全な対訳になっているとは限らない。各対訳文には内山・井佐原の方法により計算された類似度が付与されており、類似度の高い順に並べられている。つまり最初の方は対訳文としての質は高いが、後ろの方に行くにつれ質が低下するため、訓練データの使用には注意が必要である。

開発データ・開発試験データ・試験データは、JST 所有の学術論文日英抄録のうち、訓練データに含まれない抄録から対訳文を抽出したものであり、それぞれ 400 抄録ずつからなる。これらのデータに関しては、文対応を自動で付け、全ての文が 1 対 1 で対応づいたもののみを利用しており、訓練データとは異なり、元の抄録全体を復元可能である。

また各対訳文には、アルファベット 1 文字からなる分野記号が付与されている。これは元の抄録がどの学術分野のものかを表すものであり、分類の詳細は JST 分類コード (<http://opac.jst.go.jp/bunrui/>) に記載されている。アルファベットと分野の対応、および訓練データ中の各分野の割合を表 2 に示す。

表 2：ASPEC-JE 訓練データ内の分野分布

記号 分野名	Top1M	ALL	ALL(%)
A 科学技術一般領域	9,863	28,608	0.95
B 物理学	149,406	297,292	9.88
C 基礎化学	36,233	116,400	3.87
D 宇宙・地球の科学	19,949	59,083	1.96
E 生物科学	80,282	266,209	8.85
F 農林水産	56,722	186,265	6.19
G 医学	284,277	1,016,024	33.77
H 工学一般領域	8,274	39,279	1.31
I システム・制御工学	11,193	47,856	1.59
J 情報工学	38,659	131,298	4.36
K 経営工学	10,112	22,583	0.75
L エネルギー工学	4,475	7,006	0.23
M 原子力工学	11,187	27,480	0.91
N 電気工学	89,232	197,269	6.56
P 熱工学、応用熱力学	13,273	47,178	1.57
Q 機械工学	33,776	102,157	3.40
R 建設工学	37,817	128,710	4.28
S 環境工学	24,791	51,765	1.72
T 運輸交通工学	6,706	16,967	0.56
U 鉱山工学	1,844	4,635	0.15
W 金属工学	26,939	86,789	2.88
X 化学工学	8,997	21,419	0.71
Y 化学工業	33,749	98,835	3.29
Z その他の工業	2,244	7,393	0.25

Top1M は、類似度順上位 100 万文内の各分野の文数であり、ALL は訓練データ全体での文数、ALL(%)は各分野の割合である。

図 1 に ASPEC-JE の訓練データのサンプルを示す。1 つ目はデータ内で最も類似度が高い対訳文、2 つ目は 100 万位、3 つ目は 200 万位である。各対訳文の上段は左から順に文書 ID、文書内文番号、類似度を示す。1 つ目と 2 つ目の対訳文はほぼ完全な対訳となっているが、3 つめは「22 日目頃」が「3 weeks」と訳されており、多少のずれがあることがわかる。なお、全体のうち 5 万文弱は類似度が 0

となっている。

G-03A0568930	0	0.880570409982175
現在、筋ジストロフィー患者の移動介助において文書マニュアルを使用している。		
At present, the document manual is used in transfer assistance of the muscular dystrophy patient.		
G-01A0204677	1	0.137147503045244
リドカイン使用濃度、使用量は0.5～10%, 0.1～1.0ml (10～60mg)であった。		
The use concentration and the amount of lidocaine were 0.5～10% and 0.1～1.0ml (10～60mg) respectively.		
G-93A0370292	0	0.0481254526317082
症例は43歳の女性で、心臓弁膜症手術後22日目頃より、発熱と共に全身に紅斑が出現した。		
A 43-year-old female was seen at our clinic with complaints of high fever and erythroderma like skin rashes, which have developed in 3 weeks after her heart operation.		

図 1: ASPEC-JE 訓練データのサンプル

3.2 日中論文抜粋コーパス (ASPEC-JC)

本コーパスは、文献データベース JDream II 搭載の和文抄録と、電子ジャーナルサイト J-STAGE (科学技術情報発信・流通総合システム) 搭載の情報処理学会、言語処理学会、人工知能学会論文誌の和文論文を各学協会から許諾を得て中国語に翻訳することで構築した対訳コーパスである。翻訳対象は抄録、もしくは本文の段落単位であり、名詞、動詞、格フレームを考慮した単語のカバー率が最大となるようにテキストを順次抽出し、全体として単語のカバー率が最大となるように選定されている。各対訳文に付与されている文 ID から各段落を復元することは可能であるが、前述の通り翻訳対象が段落単位で選定されているため、論文全体を復元することはできない。

開発、開発試験、試験データは、ASPEC-JC 全体で 1 段落しか含まれていない論文からランダムに抽出したものであり、それぞれ 400 段落 (論文) ずつからなる。つまり訓練データや、他の開発、開発試験、試験データには、これらのデータと同じ論文に属する文は含まれていない。この設定は、いずれ機械翻訳を利用した論文の翻訳を行うことを想定してのものである。

ASPEC-JC に含まれる論文の分野の割合は、およそ表 3 のようになっている。また図 2 に ASPEC-JC の訓練データのサンプルを示す。ASPEC-JC のデータは日本語を元に 1 文ずつ中国語に訳されているため、対訳の質はどの文でも同程度である。

表 3: ASPEC-JC 内の分野分布

分野名	割合
医学	28%
情報	28%
生物	14%
環境	12%
化学	6%
材料	5%
農学	4%
エネルギー	3%

JST_JC_AGR-jcs-74_149.txt-par27-sen1
精米タンパク質含有率は、5.69～6.60%の範囲で、主茎および第4・5節の1次分げつが第7節の1次分げつや第5節の2次分げつに比べ有意に低かった。
精米蛋白質含量在5.69～6.60%的范围内,主茎及第4・5节的1次分蘖比第7节的1次分蘖和第5节的2次分蘖明显低。

図 2: ASPEC-JC 訓練データのサンプル

4. アジア翻訳ワークショップ (WAT)

アジア翻訳ワークショップ (The Workshop on Asian Translation, WAT) はアジア言語に注目した、新しい評価型翻訳ワークショップである。様々な方法で翻訳実験や評価を行い、そこで得られた知見を共有することで、現状の機械翻訳の問題点の洗い出しやその解決法を模索し、アジア各国における実用的な機械翻訳システムの開発を目指す。第1回ワークショップでは科学技術論文をターゲットとし、データとして ASPEC を利用する。WAT の詳細については Web サイトを参照されたい (<http://orchid.kuee.kyoto-u.ac.jp/WAT/>)。

WAT の特徴は以下の4つである。

- ・オープンイノベーションプラットフォーム

評価対象となる試験データが固定であり、しかも正解翻訳も含めて全て事前に公開されている。参加者は翻訳の評価をいつでも繰り返し行うことができる。また WAT では他の翻訳評価ワークショップと異なり、自動評価のための翻訳結果提出期限は設けていない。

- ・ドメインと言語対

科学技術論文をドメインとし、また言語対として日中を対象とした翻訳評価ワークショップは WAT が世界初である。

- ・文脈の考慮

他のほとんどの翻訳評価ワークショップでは、試験データはランダムな文単位で準備されているが、WAT の試験データ (ASPEC の試験データ) は段落単位で準備されている。これは、機械翻訳において文脈を考慮することの重要性の確認や画期的な方法の発見を行うためである。

- ・評価方法

翻訳評価は自動評価と人手評価の両方を行う。また人手評価ではクラウドソーシングを利用することでコストを抑え、また複数回の評価を行うことを可能としている。

5. おわりに

本稿では科学技術論文を対象とした世界初の対訳コーパスである ASPEC について、その作成の経緯と特徴について述べ、また ASPEC を利用した評価型ワークショップについても触れた。本稿の読者が少しでもこの共有資源に興味を持って頂ければ幸いである。

近年の日本、中国の科学技術の発展は目覚ましいが、その成果の中には各国言語のみで発表されたものも少なくない。このような情報にアクセスすることは、科学技術分野において何が既に解決され、何が次の問題なのかを知るためなど、両国の科学技術交流の促進や発展に非常に重要である。これをサポートするには機械翻訳技術を始めとする様々な自然言語処理技術によって言語障壁を取り除くこと

が必要不可欠であり、ASPEC の公開がこの最終目的に向けた新たな一歩になることを願う。

参考文献

[1] Masao Utiyama and Hitoshi Isahara. (2007)

A Japanese-English Patent Parallel Corpus. MT summit XI.

機械翻訳事始め[Ⅲ]

坂本義行

アジア太平洋機械翻訳協会理事

1. 機械翻訳研究委員会の発足

日本で機械翻訳の研究が本格的に始まりました。そして、委員会の発足が1965年の「情報処理」¹⁾の「機械翻訳の現状と将来」和田弘、第5回情報処理学会大会における招待講演と題した記事の中に述べられています。

機械翻訳について話すよう交渉をうけました。近頃直接研究に従事していませんが、学会の機械翻訳研究委員会の委員長をいたしている関係からお引き受けしました。

1960年4月当学会が創立されました際、“計算しない計算機”(この記事については、前号で紹介しました)と題しまして、当時電気試験所でやっておりました機械翻訳の研究内容をお話したことがあります。

翌年の秋、学会に研究委員会の規則が制定され、“機械翻訳研究委員会”が設置されて以来、隔月に1回ずつひらかれて現在に至っております。どんな人がどんな項目の発表をなさっておられるかは、会誌の会告欄をご覧くださいればわかるはずです。

この委員会はopenではありますが、関係の方には特に通知しているので、参集される方も自から定まっております。言語学を専攻される方は折角の熱意は持っていますが機械の使い方がよくわからないので困っているような印象をうけましたので、希望者には電気試験所で計算機の使い方の講習をしたこともあります。1962年には総説としてご参考になることをまとめたこともあ

ります。(この文献も前号で紹介しました)日米科学協力委員会で情報交換を担当しておられる丹羽保次郎先生は、科学技術の論文だけでも機械翻訳(以下略してMTとする)ができるようになればと念願されて、文部省の助成金によって俗称“丹羽委員会”を結成されました。学会のMT委員会には地方の方が旅費の関係で出席されないことと、理学部関係の方がかけていたので、これを補われたものでした。数回の会合で同委員会も軌道に乗った頃、MTのセミナーを開催することをアメリカ側と交渉され、1964年4月東京でアメリカから9名の来訪を得て計20名が論文を発表し、意見を交換し、日本のおもな研究の実地を訪問しました。これらの論文の要旨は学会誌に紹介されております。(この件も、前号で紹介しました)

その後、丹羽委員会は学会のMT委員会の強化改組を条件としてこれに合流の上、発展的解散をいたしました。

日米科学協力を担当する日本学術振興会には、MTの研究を推進するために、MT部会が設けられています。

2. 世界の情勢

わが国では機械にかけた実験をしている研究機関としては、電気試験所、九大工学部、京大工学部、東大理学部、名大工学部、防衛庁技研、および電機大学などです。いずれも英語を和訳することを主としていますが、東大、電試では和文英訳の実験もや

っています。言語学者も文法についていろいろ建設的な提案をしておられますが、それを骨子とした実験はまだ見当たりません。研究グループとしては、国語研究所、新潟大、京大、名大、教育大、紅露研などです。

一方アメリカでの研究状況はかなり変わって来ております。われわれより一步先んじて研究を始めてみたがすでに30億円の研究費を投じたにかかわらず、実用になったとはいえないです。研究者自身が反省して、翻訳よりもまづ自国語でもよいから言葉を研究してみる必要がある。それには計算機を使う必要があるということから、Computational linguistics なる標題のもとに研究するように転じています。そして成功すればMTはもちろん、文献のabstract、情報検索などにも共通に貢献できると考えています。

こんな理由から、現在MTを中心に研究をしているところとしては、大学では Texas, Calif., Wayne State, Ohio State, Washington, Indiana, Georgetown, その他の機関としては、National Bureau of Stanard, Rand Corp., Bunker-Ramo などです。Harvard 大学、M.I.T., I.B.M.などはさきに述べた転向の色が濃いようです。

(MTの解説：中略)

さてMT研究の現状であるが、どの研究者もその点に困っている。多くの説が提案されていること自体がその間の状況を物語っているわけである。いずれもこの説明しにくいと申した内容について、ある一面に重点を置いたものともいえる。

これらを念頭に研究グループの態度を整理してみると、まづ次の3派に大別できる。

(1) 経験派

従来の文法の中で骨格と思われるいくつかをまずプログラムして実験し、その結果から、追加改良して行こうというもので、Georgetown 大学など古くからやっているものはこれに属する。

(2) 伝統派

既存の文法など言語についての知識をそのままプログラム化してゆく、集大成されている知識を数学的に記述すると、膨大なものになって、到底完成の見込みがあるまいとの見解に立ったものである。

Harvard, National Bureau of Standards, Wayne State, Bunker-Ramo などがこれに属する。

(3) 理論派

文章の構造をきめる態度として認められている説に Constituent structure 論と dependency structure 論とがある。前者は文を主部と述部とに2分し、主部をさらに名詞句とその他とに2分するなど、次々に2分して tree 状の解析をするものである。一方は、動詞に着目して、主部、目的部、副詞修飾部がこれに依存的構成であるとして分析を進めるのである。また数学から出発した N. Chomsky は Z.S. Harris の句についての研究を再構成して phase structure についての文法を確立した、これには変換理論(transformation rule)が併せて発表され、信者がふえている。また S. Lamb は言語を単語より小さい単位のもの集合であるとし、見做す層化理論(stratification rule)を主唱して異彩を放っている。それぞれの哲学に従っているが、共通していることは、数式化、形式化ということが根底にあることで

あろう。

電気試験所での MT の研究で考えたことをこれらと較べてみると、これらの諸説はみんな含まれているような気がする。つまり諸派諸説と雖も、実体は紙一重の差しかないのではなかろうか。優劣を論ずるには実験結果に待つ以外にはあるまい。

こんなわけで、さきにも述べたとおり研究の中心は MT から英語そのものの研究へと動いている。その典型的な例は Yngve が始めた文章の創作であろう。これは核となる文章をまず phrase structure の規則で作り、これに変換の法則を適用して変形して文章を打ち出させるものである。このためには辞典での単語に意味のほかにその語の持つ機能を含ませておく必要がある。京大の長尾君もこの実験を進めている。

このようなことを念頭に、日本語を見ると、まず和文英訳の実験が始まったことは喜ばしい。東大の野崎君によれば、

I think that reader know well that sound is vibration of air but sound by which we do talk and that enjoy music is only a part of air vibration which we utilize. Supersonic wave about which I talk here from now on is vibration that is inaudible to our ear and of which frequency is high. I think that there is man that wonder then if such inaudible sound is useful therefore I shall describe about utility value of supersonic wave.

という訳文がだされています。原文は読まなくてもよろしいでしょう。原文に pre-edit をして、input したものです。それにしても、日本語については、次のような諸問題がまだ手を付けられないことを申し上げねばなりません。

(1) 文字：現在では、ローマ字あるいはカナ文字が使われているが、実際の文章は漢字が含まれているのみならず、英語のように分かち書きになっていない。単語辞典を作ることは MT の第一歩ですが、この辞典の見出し語を漢字にするか、カナ文字で行くか、分かち書きになっていない単語の両端をどうして求めるか大変なことでしょう。

(2) 構文：日本語と英語との間には大きな違いがある。英語は客観的な思想で書かれているが、日本語は人主思想で貫かれているようです。そこで主語となる人が明らかな場合は省略される。このようなものを追加したり、逆に削除したり、あるいは別途の形式をとると翻訳が容易になるような事例が乏しくない。

それにも増して、MT に適した文法が見当たりません。言語学者と計算機側の学者との協力がますます必要であると考えます。幸いそういう気運は盛り上がりかけていると見ています。

なお、1965 年 5 月には、I.F.I.P の大会が New York で開かれますが、その前の週に、この方面の国際会議が開催されることになっています。

新しい進歩が生まれることを期待したいと思います。

3. 第 2 回日米機械翻訳セミナー

「情報処理」²⁾ に第 2 回日米機械翻訳セミナーの報告会が行われ、その時の様子が、

前述の和田氏からなされています。以下にこれを掲載します。

報告 和田 弘

1. セミナー参加者

日本側

和田 弘 (成蹊大)、坂井 利之 (京大)、
田町 常夫 (九大)、野崎 昭弘 (東大)、
長尾 真 (京大)、平松 啓二 (電機大)、
山田 小枝 (立正大)。

米国側

E.D. Peabody (U.Texas), V. Yngve (M.I.T.),
D. Hays (Rand), H. Josselson (Wayne State),
S. Kuno (Harvard), C. Fillmore (Ohio State),
P. Garvin (Bunker Ramo), I. Sakai (Ohio State),

2. 見学旅行

5月5日

(i) Bunker Ramo. (Canoga Park)

(ii) RAND Corp. (Santa Monica)

5月6日

(iii) U. of Calif., Los Angeles 分校

(iv) System Development Corp.

5月10日

(v) LRC, Univ. of Texas (Austin)

5月12日

(vi) National Bureau of Standards
(Washington D.C.)

5月13日

(vii) Harvard Univ. (Cambridge)

(viii) M.I.T.

3. セミナー

5月17、18日の両日、New York の
Park-Sheraton Hotel で開催されたアメリカ側
の出席者はさきに示した方々のほかに、
N.S.F.の Dr. See, Mr. Pronko, RAND
Corp.の Mr. Ziehe と Mr. Kay, Bunker

Ramoの Mr. Penn.さらに通訳として日本人
が1名加わりました。

われわれは論文を用意して行って発表し
ました。

すなわち

Input and Data Format of Japanese
Texts:

和田 弘

Procedure for the Analysis of Japanese
Text:

坂井 利之

Syntactic Description of Japanese
Grammar:

田町 常夫

On the Dictionary Preparation:

野崎 昭弘

Japanese-English Translation
regarded as Sentence Generation:

長尾 真

Syntactic analysis of Japanese, by
Using the Property of Dependency:

田町常夫、石原好宏

の6論文です。

アメリカ側は見学旅行の際に資料を提供
してくれたため、セミナーではまとまった
論文を中心に自由討論が大部分であったと
申せます。

第2日目に N.S.F.と学振とを經由して日
米科学協力委員会に提出する勧告書の草案
を審議しました。議論が活発に出て、原案
が相当修正されました。

勧告書の訳を記します。

勧告

昨年のセミナーでの勧告を再確認すると
次の5項目を一層に明確にする好機である
との前置きにつづいて、

(1) 研究者の交換は日米協力の中で大切

なものである。現在の計画は 1 年間で成果を上げるには短すぎるから、延長するつもりで続けていこう。別の計画も奨励しよう。

(2) 以下の勧告を実行するには日米の学会(情報処理学会と AMTCL)との支持が要る。各学会は個人または委員会を任命して、それぞれのスポンサー機関との密接に連絡しよう。

(3) 文献などの交流をもっと促進しよう。日本の文献の題目をアメリカの The Finite String のリストに掲載しよう。

(4) テキストの交換を便利にするために共通の code と format を採用できるように考えよう。両国の学者はそれぞれの国語についての提案を出してもらいたい。

(5) 昨年と今年との 2 回のセミナーの結果、専門家による小規模の会合が有効なことがわかった。上記(3)および(4)の実体について seminar, planning meeting あるいは working group が時々必要だ。少なくとも 2 年以内には次回の集合をしよう。日米の両学会はテーマと会合場所とを選定して、スポンサーに進言しよう。

4. 国際会議

5 月 19 ~ 21 日の 3 日間、同じ Park-Sheraton Hotel で International Conference on Computational Linguistics が開催されました。日本はさきに 6 個の論文を送ったところ、京大の坂井、長尾両君による sentence generation の論文が発表されることになっていました。論文は約 40 ぐらい参加者に配布されました。2 日目の夕、晩餐会があって主催者たる ATMCL の会長 Prof. Lehman の講演がありました。そのあとで同氏の部屋に各国から 1 名ずつ代表が招かれまして、Whisky をご馳走になりながら非公式に国際的な組織を作ることに

よって、こういう会議を開催することはどうだろうかと提案がありました。

ヨーロッパの人々はいずれも賛成しましたが、私は反対しておきました。本日ここにおられる久野さんも反対しておられたようです。しかし多数の意見によって推進することになり、その結果すでに Prof. Lehman から手紙が来しました。

それによると、Federation と Central Committee をつくることについて、各国の意見をまとめて 7 月中に返事が欲しいとのこと。もし大多数が賛成ならば、1967 年には会議を開くことに予定したい、との趣旨です。返事を出さねばなりません。

質問 反対の理由は、何か義務でもできますか。

和田 できます。できれば、大した額ではないが費用を負担しなくてはなりません。さらに、会合の都度誰か一人は出ねばならず、できれば同一人でありたいものですが、無理でしょう。

国際的な会合はヨーロッパで開かれると思わねばなりません。旅費が大変です。それにヨーロッパの諸国では、国語がいずれも違うのですからどうしてもその辺が議論の中心になって、日本語はやはりツンボ栈敷におかれることになりましょう。

さらにこの動きを注目しますと、言語学者がその研究に計算機を利用しようといういわゆる computational linguistics の人々が中心になっています。わが国でいえば計量国語学会に相当するものです。理工科系の人はずでにある IFIP の 3 年ごとに開かれる Congress の中で MT のための場をもっています。

また仮に反対しても、設立されるでしょう。加入してなくとも 67 年の会議には参加することはもちろんできると思います。久野さんも理由を一つ、

久野 アメリカの連中は肩書きを欲しがっている。またヨーロッパに行く理由を求めています。委員会ばかり沢山作っていて、意味がない。

座談会

出席者 和田 弘、坂井利之、久野 瞭、
野崎昭弘、長尾 真、小笠原林樹、
伊藤正安、 ベーツホフファー
司 会 中野道夫

1. M.I.T.

中野 はじめに、とくに印象に残ったことを伺いたいのですが

和田 アメリカでいま MT を一生懸命やっているところでは、テキサス大学だけです、たいていのものはもうあきらめて、N.S.F.から金をもらって、好きなことをしていると見ました。ただ日本人がぼやぼやしていると、アメリカの方が日本語の機械翻訳を先にやってしまうのではないかという心配があります。

長尾 M.I.T. は Yngve が彼の depth hypothesis が日本語でどうなっているかと興味をもって尋ねていた。彼の部屋には、マルチアクセスのステーションがあって、ここで彼がつくった COMIT を呼び出して日本語のジェネレーションの実験をしてくれた。田町さん、山田さん、彼はここで日本語をやったのは初めてだと言っていたが、やはりプログラミングシステムは完

備している。だから、言語学者は言語のことを専心して考えればよい。つまり計算機とかプログラミングを使う方に関しては、非常に自由に使える。そういう環境にあるということはやはり良いことですね。

中野 マルティアクセスというと、実際使う人はどうするのですか、input は何ですか？

野崎 簡単にいったら、オンラインのフレクソです。

和田 長尾さん、あなたもそれに似たようなことをおやりになっているようだけれども、能率についてくらべて頂けると…

長尾 それは比較にならないですね。彼等の場合だと、現に黒板に書いたものを、その場でポンポンとルールに打ち込めば、すぐ何百というようなセンテンスがパッパッとでてくるわけです。

小笠原 その出てくるセンテンスは立派なものですか。

長尾 MT をやっているのではなくて、文法の検定ですよ。ゼネレートされた文章は全部われわれが見て、文章の意味を持っていなければまだ作られた文法が完全でないということです。

中野 ゼネレートしてみれば、文法のどこがおかしか、すぐわかるわけですね。

花もあるきたい、みます。わたくしでも わたくしでも わたくしは花でも 食べます。みます。あるきます。犬が花でも花をわたくしはみたい。

わたくしがたべたい。たべたい。花は私でも犬もたべたい。犬も犬がわたくしでも犬はわたくしは わたくしはでも みます。わたくしでもあるきたい。犬がたべたい。etc

小笠原 M.I.T.はいわゆる chomsky 言語学が中心になっていると思うのですが、彼の文法はやはりあすこて使われているか？それと、チョムスキーは MT に関心を持っているのですか。

和田 使っている人は多いでしょう。しかし、チョムスキー自身は、MT についての関心は、今は全然もっていない。興味はもっているが、とっくにやめた。

2. 彼我の差

中野 ハードウェアの面がうらやましいという声があったでしょう。

野崎 ハード、ソフト、データ伝送も含めて、要するに計算機については差が大きいですね。

和田 そうだ。彼等とはとにかく、機械翻訳をやりたい時はいつでも機械をつかっている。日本人は何か頭が良さそうな顔をして、キョロキョロしているけれども、機械をいじりもしないのが多い。

坂井 MT の見学に行く前に、アメリカ帰りの MT の専門家でない人がいったことに、1950 年代の初めに、アメリカで非常にさかんだったことを、今

日日本でヤイヤイやっている。ところがアメリカ自体は、エンジニアが機械にかけた、これは面白いとやってみたものは、言語学的にみて何の翻訳にもなっていなかった。それで今度は言語学者がそれはいけないというので、非常にもとの方へもどりすぎているのではないかと、すなわち、学問の研究というのは確かにオッシレーションはあるけれども、少しその反動が大きすぎるのではないかというような印象をもったわけです。もう一つは、アメリカは機械と送付とソフトウェアが進んでいて、確かに環境は十分なのだけれども、とくに新しいものが出たというのではなしに、日本でやっても別に今からでも遅くはない。わたくし自身はそういう反省のもとで、初めから機械を使ってやっているけれども、機械を使った MT も、日本ではやっていける余地は十分あると感じて帰ったわけです。わたしの感じでは、ランドにしる、テキサスにしる、辞書なんかだとやはり何万という単語を入れている、久野さんの文法にしる、われわれがいつているようなものより、相当ち密な内容をもったもんをもっている。それでいろいろ議論をしているわけで、だからいろいろな面もわかってきている。抽象的なことだけでないということはあるですね。もちろん周囲環境は日本にくらべて恵まれている。考えようによっては、ああいう環境があれば、われわれでもある程度おもしろいことができるのではないかという気がする。そういうことです。

野崎 少し長い目でみた話だと、アメリカの

方が計算機そのものばかりでなく、計算機についての教育が普及していることの差を感じました。ほんとによい MT を長期的な見透しで実現しようとしたら、だいたい大学における教育から考えないと結局進まないんじゃないかな。

中野 皆さん、だいたい量的な人員、設備、環境、それは確かに良いが、質的には日本と比べてどういうことはない。

坂井 パーソナリティーがないという感じはわたくしはした。

和田 それをいうと危ないんだよ。一面の真理だが、坂井さんがさっきいわれたことですが、アメリカは「アツモノにこりてナマスを吹く」という感じがする。要するに MT をしていないのだから、MT がそんなに進歩しない。アメリカが CL をやっているから、日本もすべきだとは思わない。やはり MT を頑張れば良いと思います。

田町 小生の印象ではアメリカは計算機は進んでいる。Vocabulary もある。しかし MT や MT に関係のある CL の面では特にこれはと思うものにお目につかなかったこと、率直にいうと、すぐれたアイデアでとられたデータがない。あるいはデータはあってもどうしてよいかわからないというのが現状で、特に意味に関しては役に立ちそうな具体論もデータもないようだということ…などです。これは日本に比べて劣っているというのではありませんが、その点役に立つデータをとることが必要で、数の上だけで“日本はよほど頑張らなければ”というのは当たらないと思います。

4. セミナーの印象

中野 今度はセミナーの印象をおきかせください。

和田 先に申しあげたとおりで、セミナーでアメリカ人は良い話をしませんでした。自分の方が経験者のような顔をしていた。日本人がどのくらいのレベルの話をするか、まあ聞いてみてやろうということだった。日本は1年間にたいして進歩してないが、彼等が思っていたよりやっていたという印象を受けたのではないかな。

坂井 わたしはレゾリューション・コミティにいて、項目としてどういうことをうたうかを協議したのですけれども、その始め日からハッキリしたことは、俺のところにも人物交流のチャンスを与えろということで、これが文章にも強く出たことです。ところで日本人をどの程度欲しがっているかという、自分らが支払うものに対して有利であればよいということで、よい言葉でいえば、cooperation でやりたいことが、ハッキリでていた、もう一つはテキサスの話でペンダクラフトが非常に大きな組織をつかってやっていると、それに対してガーヴィンが非常に違った反対の意見を出している。アメリカの中にもいわゆるフランス流の考え方が分かれていた。

中野 カンファレンスの方では、日本の講演に対して、良い質問は出ましたか。

長尾 その時は30分のもち時間をほとんど使ったので、その時は出なかった。しかし、たとえばランドのハーパーという人などは、自分のと非常によく似ているといっていました。ドイツのボンの MT の研究所のホッペが“アメリカ

人は意味のことについてはほとんど関心をもっていないと、あれはけしからんことである”と書いてました。彼等も私達と非常によく似た考えをもってやっているのだそうです。

5. トランスフォーメーション・グラマーの評価

小笠原 久野さんはトランスフォーメーション・グラマーのその位置というのはどういうふうにお考えですか。

久野 やはりフレーズ・ストラクチャ・グラマーの時代は過ぎて、どのグループもそれではもう十分ではないということが分かって来た。型はいろいろ違うが、なんらかのかたちで Chomsky 流のトランスフォーメーションを組み入れている。

和田 Chomsky のグラマーは、日本語に直接適用できないから、つまらない、チョムスキーの本の付録に英語についてだ。

野崎 ある付録を例に挙げるのは拙いでしょう。

和田 それは知っている。シンタクテック・ストラクチャの初めの方に書いてあるマセマティカルな所はいいと思う。あそこなら日本語にアプライできる。

長尾 トランスフォーメーション・グラマーは非常に有益ですネ。いろいろなことを説明するのに、だけれどもその…どういうふうに一般的なレベルで利用できるかということになると非常にむずかしいと思う。皆はトランスフォーメーション・グラマーの云々とかいういろんなことを

意識していつてるけれども實際上それを有益に使った人は未だにないのではないか。否定をとるとかいうくらいのものでネ。一見よくて皆とびついたらけれども、にっちもさっちもいなくなつたという危険性をいつも考えるべきだといいたい。

久野 Chomsky がいいだしたから、トランスフォーメーションをいうのではなくて、要するに言語にそういう面があるので、それを使わないと、いろいろな面を描き出せない。結局説明が不足になるわけですよ。

6. IFIP での MT

長尾 IFIP の MT が少ないようですけれども、あれが何故そんなつまらぬのが集まつたか不思議ですね。

中野 それは日本だけですか。全体ですか？

長尾 全体ですね。

小笠原 それは、その前に MTCL の方にこちらの人達が主力をおいたということですか？

久野 MT の良いペーパーがないのです。

和田 あまりやってないのに、どうしてあるのだ。

中野 それが世界的傾向ですか？

野崎 MT という意味をせまくとれば、良いペーパーはそうはないんじゃないですか？

久野 一度フレームワークを出したら、それから 1 年後に文法がこれだけよくなつたって書いてもしょうがないでしょう。

中野 おしまいにしてしまおう。ありがとうございました。

私事ですが、

1965 年 4 月といえば、私が電気試験所に入所し、電子計算機部情報処理特別研究室に配属された年です。その時の電子計算機部長が和田弘氏でした。

私は、1960 年 3 月に大学の通信工学科を卒業し、4 月から東京大学生産技術研究所、第 3 部（電気工学）に入所しました。研究所では、東大のカッパー、ラムダーのロケット打ち上げが行われていました。弾道を描いて飛ぶロケットです。

1961 年 9 月から、1 年間の予定で、フルブライト交換留学生として、この時に前出の丹羽保次郎先生に推薦状を書いて頂きました。米国、プリンストン大学大学院電気工学科に留学しました。ここで、ある日、大学にソ連の宇宙飛行士、チトフ氏が講演に来て、これを拝聴しました。また、大学院では、プログラムの講義を受けました。

その影響を受け、将来を考え、ロケットの仕事とソフトウェアの仕事について、検討しました。ロケットといえば、軍事目的、計算機を利用した平和目的はないかと調査をすすめ、私の英語の力を使えば、機械翻訳の仕事ができるのではないかと考えました。

帰国後、調査し、電気試験所で研究が行われていることを知り、公務員試験を受け、1965 年から機械翻訳の仕事に従事することができました。

当時、研究室には、「ヤマト」が置かれて、電気洗濯機のような磁気ドラムが回っていました。しかし、友達に研究室の名前を話すと、情報処理とは、下水処理とか污水处理みたいな研究か、計算機で言語を研究するのだと答えると、外務省の下請けで、暗号解読の研究をするのかと質問され、当時の日本人の情報処理にたいする認識はこん

な程度でした。

ただ、昨年、「はやぶさの宇宙からの帰還」のニュースが流れたときには、ロケットの使い道について、浅学であったなあという残念な気持ちにもなりました。

参考文献

- 1) 和田弘：情報処理、Vol. 6, No. 3, 1965 pp. 126~128.
- 2) 和田弘：情報処理、Vol. 7, No. 1, 1966, pp. 37~41.

機械翻訳の現状と展望

内田真弓

関西外国語大学

はじめに

日本通訳翻訳学会の分科会である「通訳翻訳テクノロジー研究プロジェクト」の第1回会合が、2014年2月22日（土）、立教大学のマキムホールで開催されました。本プロジェクトは発足してまだ日が浅いにも関わらず、参加者はおおよそ20名程度あり、今後の拡大が期待されます。

記念すべき第1回目の発表者は、「The 翻訳」の開発に携わる熊野明氏（東芝ソリューション株式会社）、そして長年にわたって翻訳テクノロジー教育を率先してきた Tony Hartley 氏（東京外国語大学）の2名でした。

筆者は産業翻訳に従事しているわけではなく、機械翻訳（以下、MT）を必要とする大量の翻訳作業とは無縁の生活を送っていますが、翻訳者として MT の存在は常に気になるところです。そして、現時点で MT はいったい何をどこまでできるのかなど、MT の現状が気にはなるものの、それを知る機会に恵まれませんでした。また、自ら MT の存在を脅威に感じ、あえて見ないふりをしてきたのではと問いかけ、今回は勇気を出して、自分の視野を広げるべく、この研究会に参加しました。

今回、この研究会を通じて、MT の実態について翻訳者が「へえ～、なるほど」と思った点を中心にご報告させていただきます。もちろん、これらの内容は本誌読者にとっては常識的なことだと思いますが、MT の素人の認識を汲み取っていただければ幸いです。

機械翻訳の仕組みと課題

熊野氏は、コンピュータ上でどのようなステップを踏んで翻訳がなされるのか、また、どのような課題があるのかを丁寧に説明してくださいました。MT については、遊びでネット上のものを使ったことはあったものの、その仕組みについてはブラックボックス状態で気にもしたことがありませんでした。しかし、実際には翻訳者と同様の作業過程を経ているため、課題も同じようだという印象を受けました。

まず、MT の仕組みは一通りではなく、三種類もあることに驚きました。直感的に MT の仕組みとして考えていたのは、辞書や言語規則、知識を積み上げて翻訳する Rule-based タイプでした。それ以外に、例文と対訳文を集積し、そこから最適な訳を探す Example-based タイプ、Google 翻訳などで使われているような、原文とその訳文を用意し、それを統計的に処理していく Statistical Base タイプがあることが分かりました。ちなみに、発表者熊野氏が携わる「The 翻訳」は Rule-based タイプとのことでした。

また、Rule-based タイプの詳細も説明がありました。翻訳の過程は、1. 辞書の情報に基づいて単語の品詞を決める（形態素解析）、2. 多品詞語に対しては、品詞候補の優先順位をつける、3. 単語と単語のつながりを判断する（構文解析・係り受け解析）、4. 共起している語との関係から訳語を決める（英語の場合はここまで）、5. 語順を決

定し（構文生成）、格助詞を補い、活用を決めるという手順を（日本語の場合は）踏んでいるとのことでした。

このような明確に規定された過程を経ると正しい翻訳ができそうですが、実際には、翻訳者も頭を悩ます文の曖昧性がここでも問題になると聞き、MT に少し親近感を覚えました。MT が失敗する例として次のような場合が紹介されていました。

- (1) 構文が同じでも目的によって解釈が異なる場合

The failure which he performed.

The store which he visited.

- (2) 複数の解釈がある場合

I found the hospital.

found が find の過去形で「見つけた」となるのか、found の現在形で「設立する」となるのか。

- (3) 係り受け候補が複数ある場合

I saw a girl with a telescope.

saw に係り「望遠鏡で」となるのか、girl に係り「望遠鏡を持っている」となるのか。

- (4) 共起する語により訳し分けをする必要がある場合

I take a bath/bus.

I take a bus.

take は bath と共起すると「入る」となり、bus となら「乗る」となる。これは翻訳者にとっても MT にとっても翻訳精度の命ともいえる点です。

日→英の場合も同様に、
「彼は歯医者にかかった」

「歯の治療費が 30,000 円かかった」
では、同じ「かかった」でも、“He consulted the doctor.”、“The dental treatment cost ¥30,000.”と訳し分けなければなりません。

- (5) 専門用語として訳語を選定する必要がある場合

200m butterfly heats（予選）

ある語が、専門用語として用いられているのか否かを峻別する必要があります。

これらの課題に対して「The 翻訳」では、先行文脈を利用して、分野推定による訳語選択を行うなどの工夫をしてきたとのことでした。ただし、ここでいう「文脈」とは、MT においてはあくまでも言語化されている情報でなければなりません。当然のことですが、人間のように想像力や機転を働かせて、原文に現れていない情報や常識的な知識を補って解釈することはできません。この点が、MT と人間の手による翻訳との大きな違いのひとつ、と言えるのではないのでしょうか。

一瞬にしていくつかのプロセスを経て訳文をこしらえてしまう MT も、単体での取り組みでは限界があるようでした。実際、翻訳者自身も頭を抱える問題ですが、言語の変化、新出用語への対応などが課題として挙げられました。この問題への取り組みとして、利用者のサポート体制の整備が必要と聞き、少し意外に思いました。

利用者が MT を自分用にカスタマイズする機能として、具体的には次のような機能があるとのことでした。利用者は、自分が有する翻訳知識を MT と共有できます。また、翻訳知識は単語レベル、句レベル、文

レベルで書き込めるといいます。利用者は、専門用語・固有名詞といった単語レベルの知識を書き込む辞書を構築できるだけでなく、句レベルの知識を訳し分け辞書、文レベルの知識を翻訳メモリ・パターン辞書、分野や文脈によって制御された単語の意味をセレクトコーパス辞書に書き込むことができます。そして、翻訳対象の分野に応じて MT を調整することで、翻訳精度の向上が期待できそうです。

「The 翻訳」では以上のように、システム（基本的翻訳エンジン・大規模言語知識）と利用者（文書に応じた言語知識・低コストで知識の追加）の分業によってさらなる性能アップを実現していると知り、人間と同じように努力を積み重ねている MT の姿に、なんとなく親近感が沸きました。

MT を利用する際、利用者のサポートが重要であるということが印象的だったのは、翻訳精度の話をしている時も同様です。MT が目指す翻訳精度として、我々が普段目にするような、正確な、そして、滑らかな文ではなく、あくまでも「人間が読んで内容が理解できるもの」との位置づけもあるとのことでした。

同様に、原文を作成する側も注意が必要と分かりました。MT の品質低下の原因の一つとして挙げられていたのが、原文の不正確さでした。熊野氏らのチームは、原文の誤りをチェックするシステムを開発中とのことでした。例えば、Word のように変な日本語を書くと warning が出るようなシステムがあると、正確な原文が作成できます。このシステムは、MT だけでなく、翻訳者にとっての支援ツールとしても期待しています。

最後に、MT の今後の展望などについてもお話がありました。MT の目指す翻訳精度を

「人間が読んで内容が理解できるもの」から「滑らかさ」を高める手立てとして期待できるのが、Statistical Base タイプの翻訳とのことでした。統計的な技術を取り込み、Rule-based タイプの MT による訳文の滑らかさをどこまで高められるか、今後の MT の発展に目が離せません。

翻訳テクノロジー教育

Hartley 氏は、20 年ほど前から MT や翻訳メモリ（以下、TM）といった翻訳支援ツールを使い、翻訳テクノロジー教育に携わってこられたそうです。20 年前といえば、ようやくパソコンが普及し始めてきたころ。そのころからすでに「翻訳テクノロジー」という概念があったこと自体が、新しい発見でした。

近年、日本の各大学で通訳・翻訳の授業が積極的に取り入れられてきていますが、MT を授業に取り入れている大学はごくまれと言えるでしょう。大学における翻訳の授業は、ともすればリーディングの授業の延長ともなりかねません。一体どのような授業を展開されているのか、自身も大学で翻訳の授業を担当している筆者としては、非常に興味深く発表を聞かせていただきました。

現在、東京外国語大学で教鞭をとられている Hartley 氏の授業では、受講生が実際に MT や TM を利用して翻訳作業を行っているとのことでした。しかし、その目的は MT の使い方を教授することではなく、MT を介して翻訳とは何かを問い直すことだといいます。

まず、ある翻訳をする際に、なぜ翻訳者ではなく MT にさせるのか、というところから考える必要があります。人間が唯一コンピュータを超えられないといわれている

のが、情報の処理スピードです。人間の手による翻訳は、1日に2,000 wordsが限度だといえます（1時間に664語という、超高速の翻訳家もいらっしゃるそうです）。それに対してMTは、1日に130,000 wordsの処理が可能とのこと。この差は歴然としています。

単一民族国家であるこの日本にもグローバル化の波が押し寄せ、世界各国からの訪問者や移住者が増加してきています。このような状況において、人間の手だけであらゆる情報を翻訳しては、とてもその需要を満たすことができません。そのために、翻訳テクノロジーが必要になってくるのです。

だからといって、私たちが小さくなる必要はまったくありません。TQA（精度保証）を見てみると、人間は5000、MTは3500という数値がでてくるそうです。つまり、翻訳の精度に関しては人間のほうがすぐれている、ということになるのです。こうしたテクノロジーを使用することにより、学生は翻訳の「質」という問題に正面からぶつかることになります。

大量の翻訳を短時間で仕上げるには、どうしても翻訳支援ツールの助けが必要になります。MTやTMを使用すると、同じような内容のものを繰り返して翻訳する場合、過去の翻訳の軌跡をそのまま活かすことができる、という利点があります。その一方で、テクノロジーには翻訳の「質」という問題があります。

翻訳の質は何を基準にして評価することができるのでしょうか。実は、この問題に対する答えはまだありません。だからこそ、翻訳の質を上げる技術も開発できないといえるでしょう。

Hartley氏は数年にわたり、翻訳の授業の

一環として、藝大のアートイベント「藝大アート イン東京丸の内」のパンフレットを翻訳するプロジェクトを行っているそうです。今日、翻訳テクノロジーは多機能なものが多く、翻訳のワークフローをカバーしてくれるツールは、ネット上で無料、あるいは安価で手に入れることができます。これらを利用して、実際に受注したパンフレットを一般の人に向けて翻訳する。このように翻訳の仕事を疑似体験することで、翻訳の質（Q）と時間的制約（D）と価格の制約（P）というQ-D-Pの関係を自ずと考える必要がでてきます。翻訳支援ツールによる翻訳と人間の手による翻訳をどのように組み合わせたらQ-D-Pの関係が満たされるのか。「翻訳テクノロジーは翻訳者の代わりとなるものではなく、その延長線上にあるもの」というHartley氏の言葉が印象的でした。

MTを使って日本語を英語に翻訳する際の気づきとして、ヨーロッパ言語からの翻訳と比べて、日本語のMTは信頼性が低いことが挙げられます。これは、英語と日本語の構造が違うのでいたしかたないことであり、だからこそ熊野氏のグループのように、あらゆる角度からのアプローチにより、正確な訳を作成できるよう、努力を重ねておられるわけです。この点、MTも人間も似かよっていて、親近感を覚えます。

最後に、大学における翻訳教育の課題として、大学と翻訳会社とのつながりが弱いことを指摘されていました。ヨーロッパ諸国では、翻訳教育を行う大学と企業とが密に関係しているため、大学で翻訳教育を受けた学生が翻訳の仕事につくチャンスが多くあります。一方、日本では、そのような役割を担っているのは大学ではなく、いわゆる「翻訳学校」であり、大学で翻訳を学

んだからといって、翻訳の仕事に近づけるとは限りません。

日々、大学生の英語力を目の当たりにしている筆者としては、これも当然かと思わざるを得ないのですが、近年、関西大学（大学院）のように、本格的な通訳・翻訳プログラムを設置する大学もでてきています。このような取り組みが増えれば、またこの状況も変わってくるのかもしれません。

おわりに

今回、この「通訳翻訳テクノロジー研究プロジェクト」の第1回会合に参加し、MTとはどのようなものか、基本的な知識を得ることができたと同時に、「翻訳とは何か」を改めて考え直すきっかけとなりました。そして何より、MTは決して恐れるべき存在ではなく、翻訳家の手の届かないところ（追いつかないところ）を補ってくれる、とても有能なアシスタントであることがわかりました。

MTの実情をまったく知らなかったときには、MTの研究が進んでしまつては翻訳家の出る幕がなくなる、と思っていましたが、今では、これがどこまで有能なものに成長していくのか、今後の活躍が楽しみでなりません。

取説新時代に向けてのテクニカルコミュニケーターの役割

山崎 敏正

一般財団法人テクニカルコミュニケーター協会

第5回産業日本語研究会・シンポジウム（2014年2月27日開催、於：東京大学 情報学環・福武ホール）で、一般財団法人テクニカルコミュニケーター協会（通称 TC 協会）の取り組み紹介を発表する機会をいただきました。タイトルは「取説新時代に向けてのテクニカルコミュニケーターの役割」として最新のテクニカルコミュニケーション（以下、TC という）の動向を報告しました。今回、AAMT ジャーナル誌への投稿依頼を受け、今後の TC を取り巻く環境変化に対応するために必要な取り組みポイントをまとめてレポートする。

1. 日本における使用説明書とは

近年、日本市場において製品に起因しない事故の原因で一番多いのは誤使用と不注意であることが調査で分かっており、製造者は使用者に、適切な使用や安全な使用についてよりわかりやすく伝えることが求められている。そうした社会環境の中で使用説明の重要性も一層高まってきている。このような中、2012年8月、さまざまな製品を対象とした使用説明の作成にあたっての基準となる国際規格 IEC 82079-1 が制定された。この規格は、日本市場における使用説明の唯一のグローバルスタンダードとして多くのメーカーや使用説明制作会社にインパクトを与えている。

◆日本市場向け使用説明書制作の法的要求

日本で使用説明に関わる法的要求は複数ある。それらは、4つの側面から考えられる要求事項や対応によって違い、以下を基本的な要求事項として使用説明書は制作されている。

①安心面の要求：

景品表示法（正式名称：不当景品類および不当表示防止法）、表示に関する公正競争規約、家庭用品品質表示法 など

②安全面の要求：

消費生活用製品安全法、電気用品安全法、食品衛生法 など

③権利責任の要求：

製造物責任法（PL法）、著作権法、商標法、意匠法、不正競争防止法 など

④国際、業界要求対応：

各種国際標準への対応、各種業界ガイドラインへの対応 など

この中で、景品表示法は使用説明の基本原則に大きく影響を与えている。製品の表示に対して「やってはいけないこと」と「違反した場合の罰則」が定められている。そこで「どう作ればいいのか」という問いに対しては、具体的な制作ルールとして業界団体や企業などで構成される公正取引協議会で策定される公正競争規約（1962年に制定）や業界別の安全確保のための安全表示ガイドラインなどによって定められている。

◆日本人の消費者嗜好やニーズ

業界の違いに応じて、消費者、利用者の使用説明書に関する嗜好、ニーズは異なる。しかし、一般的に次のことが消費者から企業に求められる傾向が強い。

「規約を守る」に加えて「消費者のより高い満足度を目指す」取り組みを施した使用説明書としてまとめられている。規約を守ることは“当たり前品質”である。日本人はよく使用説明書を読むことも各種

調査で判明しており、“わかりやすさ”への要求が高いことも分かっている。日本において、TC 技術が注目を集める契機となったのは、情報処理技術が発展し、1980 年代から登場したパーソナルコンピュータを始めとするコンシューマー市場向け情報機器の普及であった。コンシューマー向け製品の使用説明情報では、利用者が文字中心の文章を読むことに負担を感じる傾向が強く、イラストの多用や、誌面レイアウトの工夫によって、よりビジュアルに情報を提供する手法が重視されてきた。このような背景から、欧米人がいうわかりやすさと日本人がいうわかりやすさに違いが生じるのである。

・欧米人のいうわかりやすさ：

文章 + ロジカル

・日本人のいうわかりやすさ：

ビジュアル + シンプル + ユーザー視点

企業側はこの日本人消費者の特徴をよく理解し、使用説明書をマーケティングに利用することまで考えて対応している。

2. IEC82079-1 制定までの経緯

◆日本における取扱説明書制作の実態

日本市場で使用説明に関連する国際規格の浸透は 2000 年以降になってからである。多くの日本企業が PL 法への対応を開始した 1995 年から、使用説明や安全表記の国際規格の存在を気づき、業界ガイドライン制定のための資料として活用してきた。取扱説明書においては、2000 年に日本工業規格として制定された ISO/IEC Guide37:1995 の翻訳版の JIS S 0137:2000「消費生活用製品の取扱説明書に関する指針」として発行されました。タイトル名に指針と明記されているにもかかわらず取扱説明書の唯一の国際規格と考えている人がほとんどでした。また、日本では JIS 規格は任意規格の位置づけで、法的に義務づけられていない、すなわち強制力がないものをいう。

次に、安全表示に関する業界ガイドラインは 1993 年に発行された「家電製品の安全確保のための表示

に関するガイドライン（制定・発行：家電製品協会）」を拠り所としてほとんどの業界が、日本における警告表示方法の標準とし運用している。現行の第 4 版は 2009 年に改訂したものである。このガイドラインは、ISO 3864 シリーズや ANSI Z 535 シリーズおよび ISO/IEC Guide51 を意識してまとめられている。警告図記号においては、ISO 3864 や ISO 7010 で規定された図記号は採用しているが、一部の図記号は日本独自のマークとして存在し、合計 19 種類を規定している。また、警告ラベルや取扱説明書の警告表示方法などを、各種媒体別に具体事例を掲載し説明している。

さらに日本市場を考えた場合、1999 年に発行された ISO13407（人間中心設計プロセス）が、多くの企業にインパクトを与え、アクセシビリティ、ユニバーサルデザインやユーザビリティ対応のための取り組みが強化された。取扱説明書の改善活動も多くの家電製品メーカーを中心に推進されてきた。これは、国際規格準拠の活動ではなく、操作手順や操作方法を使用者に“分りやすく”、“見て分かる”をコンセプトとした活動である。

◆IEC 82079-1 発行までの TC 協会の取り組み内容

IEC 62079 は 2006 年に日本語翻訳版 JIS C 0457（電気および関連分野—取扱説明書の作成—構成、内容および表示方法）として発行された。しかし、内容がプラントなどの大規模設備向けに偏った内容が多くあり、コンシューマー製品のマニュアルに対して適用するのが難しい部分があった。この規格に対する適合性を評価すると、コンシューマー製品のマニュアルには合格しない可能性すら考えられる。TC 協会ではこのようなことに危機感を持ち、新しい国際規格を作成するために 2005 年 11 月にワーキンググループを立ち上げた。この WG は 2007 年 1 月にコンシューマー向けの製品のための製品情報規格 IEC62079 Part2 のドラフトを完成させた。その後数回の IEC 会議を経て、2008 年 1 月の IEC 会議において、改訂し IEC 62079 Ed.2 を策定するという案が承認された。この規格は

ISO/IEC Guide37 に準拠して作られている。この Guide37 は ISO の COPOLCO (Committee on Consumer Policy:消費者政策委員会) が策定した使用説明に関する規格を策定する人達のためのガイドラインです。つまり、IEC62079 と ISO/IEC Guide37 とは密接な関係を持っており、2008 年から COPOLCO メンバーが IEC62079 のメンテナンスチーム IEC/MT21 に参加している。ISO/IEC Guide37 の改訂と IEC 62079 Ed.2 の改訂は、同期をとりながら作業を進めてきた。その後、会議や 3 回にわたる CDV(Committee Draft for Vote)などの手続きを経て、2012 年に FDIS(Final Draft International standard)となり、国際投票の結果、2012 年 8 月に正式に IEC 82079-1:2012 として制定された。その後、2012 年 12 月に ISO/IEC Guide37:2012 が正式改訂された。TC 協会の代表者は 2008 年から MT21 のメンバーに参画した日本からの 3 人の一人で、MT21 のコンベナーに任命され改訂作業をリードしてきた。

3. TC 協会が取り組む国際規格準拠活動

2012 年 8 月に発行された IEC 82079-1 の解説セミナーを TC 協会主催のテクニカルコミュニケーションシンポジウム 2012 東京開催 (8 月) に特別イベントとして実施しました。この規格の影響力の大きさは前規格 IEC 62079 とは比較にならないものがある。日本において、本規格の影響で使用説明の作成にあたって考慮すべき前提が大きく変わった点は、次の 4 項目である。

①対象とする製品や提供する媒体物の範囲が拡大した。すべての種類の製品が対象となる。例えば、塗料用のブリキ缶、大型産業機械、ターンキーベースのプラントや建造物、消費者製品、非消費者製品、電気製品、電子製品、機械製品、ソフトウェア、各種サービスなどである。対象とする媒体は、従来の使用説明書に加えて、包装箱の使用上の注意、カタログの仕様情報、保証書、製品上の識別情報や警告ラベルおよび組み込み表示情報、Web 上の製品情

報などである。

②情報の一貫性が要求される

使用説明中のすべての情報はその中において整合していなければならない。製品、広告、包装、保証書、Web 情報など、供給者が発行する同じ製品に関連するすべての情報は整合していなければならない。

③ターゲットグループや意図した使用を明確にする。製品を使用する可能性の高い人々のニーズおよび能力に対応し、対象製品の意図した使用を明確にしないといけない。最初に、製品の供給者が意図する使い方を積極的に知らせる内容とユーザーが製品を使用する前に知っておくべき情報の双方を含めた使用説明を必ず製品に添付しなければならない。

④リスクの最小限化

製品のリスクアセスメントを実施し、許容可能なリスクにまでリスクが低減されていることが前提となる。そのために合理的に予見可能な誤用に起因するリスクがあれば、使用説明により残留リスクを説明しなければならない。

その後、TC 協会は日本規格協会と共同で英和対訳版の冊子の完成に向けた作業を行なった。2013 年 2 月に冊子が完成し、IEC 82079-1 規格説明会～使用説明のグローバルスタンダードを学ぶ～というタイトルで 2013 年 2 月 22 日にセミナーを日本規格協会と共催で開催し、100 名を超える参加者が集まった。このセミナーの参加者の多くは、違う業種の製造業者の方で使用説明の国際規格に強い関心を寄せた。また、5 月に東京、6 月に京都で TC 協会の会員を対象とした規格の説明会を TC 協会主催で行なった。IEC 82079-1 を理解し、使用説明の制作プロセスの上流から対応しなければいけないことを訴えている。このことは、日本市場の特長である、業界標準や関係団体のガイドラインに従うこと (ガラパゴス化した対応) では、国際規格に準拠したことにならない。

4. TC に求められる創造性とは

多様な使用説明を作成するために、多彩なコンテンツを複数の媒体に展開し、最適な文書の設計を実現するプロデュース力が必要になった現在では、創造性豊かな情報発信が求められる。

情報の読み手である利用者から「わかった」という納得や「よかった」という満足を引き出すためには、利用者の主観に働きかける伝達が必要だ。利用者の関心を引き付け、よい結果につながる体験をもたらすように、ポイントを絞り込み、利用者の視点を重視した情報伝達が大切になる。そこから、利用者の心に響く使用説明が生まれるのである。

テクニカルコミュニケーターに求められる新しい技能として、豊かな創造性を整理すると次のようになる。

① 利用者の動機や反応を常に意識すること

デジタルカメラの利用者を例に考えてみよう。いつ、どのようにして「欲しい」と感じるようになるか、購入するまでにどんな点を何と比較するか、購入して家に持ち帰りパッケージを開くときの期待や不安は何か、最初にどんな撮影を試したいと思うか、撮影した画像をどのように楽しむか、家族や友人とのコミュニケーションで写真をどんなふうにご利用かなど、利用者が製品に出会い、使っていく場面が数多く想像できる。

使用説明を作り出す者は、それらの場面ごとに、利用者がどのような動機で説明を読み、その結果どんな反応が生まれるかを、自分の頭の中に生き生きと描き出すことが大切である。それには、製品や利用者について深く理解すると共に、自分の作り出す情報がどのように役立つのかを具体的に想像する習慣を身に付ける必要がある。利用者の動機や反応を常に意識することが創造性の土台となる。これは、使用説明の制作においても、ユーザーエクスペリエンス（UX）を重要視する傾向を反映したものだ。

② 市場の特性に対する豊かな想像力を持つ

製品やサービスを誰が、どこで、どのように使うのか、さまざまな市場について、利用場面を具体的に描き出すためには、市場や利用者に対するさまざまな調査・分析と、それに基づく想像力が重要になる。海外に輸出される製品の使用説明では、日本の習慣や社会ルールが適用できない場合が少なくない。仕向け地の市場や利用者の生活について、できるだけ多くの情報を入手するようにしたい。特に仕向け地が新興国である場合は、先進国の常識が通用しないことがある。たとえば、新興国の平均的な家屋では、先進国と比べて、コンセントの数が少ない。常に電源ケーブルをコンセントにつなげた状態ではなく、家電製品を使用するたびに同じコンセントに接続しなおすことも珍しくない。そうした状態を前提に、使用説明の内容も見直さなければならない。

③ 利用者と製品、サービスの関係を作り上げること
使用説明は、利用者と製品やサービスをつなぐ橋渡しとなるものである。また、製品やサービスを初めて使う利用者にとっては、最も効率的かつスムーズに、基本的な操作を習得できる筋道を示すことも使用説明の重要な役目だ。

現在の製品やサービスでは、操作のきっかけとなる情報が、製品自体の各部に表示されることが増えている。それらの情報には、ボタンやスイッチの形状や配置、アイコンや機能名の表示、ディスプレイに表示される情報など、多種多様なものがある。それでも、表示の意味が設計の意図どおりに利用者に伝わらないことや、利用者が正しく操作を行ったつもりでも思いどおりの結果が得られないこともある。そのような場合に、どうすればよいかという解決方法を示すことも使用説明の役割である。

テクニカルコミュニケーターは、利用者と製品やサービスの間にどんな関係を作り上げればよいかを問い続け、最適な答を導き出す必要がある。

④ 「よかった！」を引き出すこと

現在の製品やサービスは、機能や性能面での差異化要因が減り、コモディティー（日用品）化が進んでいると言われる。その中で、利用者に独自性をアピールし、積極的に選択してもらおうとするために、多くの企業がさまざまな努力を重ねている。そのひとつが、利用者から見た「経験価値」の重視だ。

利用者が製品やサービスと出会って使う時間や空間の全体を「特別な経験」として印象づけ、買ってよかった、使ってよかった、来てよかったなど、個々の利用者に喜びや感動をもたらすことが大切になってきた。この考え方は「ユーザーエクスペリエンス（UX）の重視」と表現されることもある。

使用説明も、経験価値を創出するための重要な手段とされている。利用者によりよい経験をもたらすために、使用説明ができることを具体的にイメージし、表現につなげることが求められる。

⑤ 利用者の感情の動きに鋭敏であること

利用者によりよい体験をもたらすためには、個々の使用説明について、利用者がどんな感情を抱くかを無視することはできない。「わかった！」や「できた！」という喜び、「そうだったのか」や「なるほど」という納得などの感情は、製品やサービスについてもプラスのイメージに通じるものだ。反対に、利用者が不快に感じる説明や不安を招くような説明は、正確な理解の妨げになるだけでなく、製品やサービス、メーカーのブランドイメージにもマイナスの影響を及ぼしてしまう。

自分の作る使用説明によって、利用者がどのような感情の動きを見せるかについて、常に敏感であることがテクニカルコミュニケーターには求められる。

⑥ 慣例や既成概念にこだわらないこと

技術進歩と市場変化のスピードが速い現在、1年前の利用者に通じた説明がそのまま通じるとは限らない。

最も好ましい反応を利用者から引き出すことができる使用説明は、時と共に変わり続けている。

このため、使用説明を作るうえで、慣例や既成概念に縛られないようにすることが大切だ。使用説明の制作現場では、過去の情報資産が徹底的に再利用されることが普通である。過去の情報資産（類似製品の取扱説明書のデータなど）の再利用の徹底は、独創性や新鮮さが最優先されることの多いほかのコンテンツにはない、使用説明というコンテンツに際だつ特徴とも言える。しかし、利用者との質の高いコミュニケーションの実現のために、あえて過去の情報資産の一部を切り捨てる決断も求められる時代になっている。

◆ 「使用説明」と「取扱情報」

ここまでに述べてきたような使用説明の多様化は、PCをはじめとするIT分野の製品では、1990年前後から始まっていた。その後、情報家電分野に多彩な製品が登場した2000年以降に急速に進展し、多くの製品分野で見られるようになった。TC協会では、このような変化を受け、2005年ごろから「製品取扱情報」や「取扱情報」という表現で、「使用説明」に含まれる多様な情報を総称してきた。同じころから、使用説明に関連する国際標準の制定や、中国や韓国などアジア市場での使用説明に関する情報収集にも、TC協会は積極的に関与してきた。

「使用説明」は、国際標準においては、「instructions for use」という英語で表記されることが定着している。また、中国では、「instructions for use」を翻訳した「使用説明」という用語が、関連法規や国家標準で明確に定義され、その使用が徹底されている。

このような国際環境の変化に対応するうえで、「取扱情報」というTC協会独自の表記に固執することは、表記の混乱や誤解を招く一因となる恐れがある。そのため、TC協会では、「使用説明」という表記への統一を図っている。

◆テクニカルコミュニケーターに求められる意識と知識

ここでは、制作担当者に求められる資質について、意識と知識の両面から説明する。

① 企画意図に対する理解

制作担当者は、自分が担当する使用説明の企画意図を深く理解し、その実現に務めなければならない。企画意図を理解するためには、企画段階で検討された製品の特徴や仕様、利用者の特性や使用目的、使用することで得られるメリットなどについて考え、深く洞察することが欠かせない。また、企画意図を理解するうえで、さらに重要なのは、使用説明の全体を把握して、自分の担当する部分が果たすべき役割や、読まれるタイミング、状況を理解することだ。たとえば印刷媒体の取扱説明書を担当するときには、組込型使用説明やカタログ、Web など、ほかの使用説明が提供する情報を利用者がどのように役立てるかを検討し、全体として無駄のない情報提供を行うことが大切になる。

② 品質に対する意識

使用説明の品質は、ユーザー品質と製造品質の両面でとらえる必要がある。ユーザー品質は、利用者から見て使用説明がどれほど役に立っているか、良い利用経験をもたらすか、安全の確保と安心の提供に有益であったかなど、利用者にとっての価値を測る尺度で決められるものだ。製造品質は、それぞれの使用説明を作り上げるうえで、正確で効率的に、かつ適正なコストで制作できたかを測る尺度によって決まる。

一般にユーザー品質を高めることだけを目指せば、製造品質は低下するし、逆もまた真である。制作担当者は自分の担当する職務について、ユーザー品質と製造品質が最適のバランスを保つように配慮しなければならない。ユーザー品質と製造品質の双方について、ディレクターから達成すべき目標が示されているときは、与えられた目標を上回ることができるように努める。

③ コストに対する意識

使用説明の制作実務においては、成果物を適正なコストで作り上げる必要がある。原稿の執筆やレイアウトなどの作業工程では、所要時間がコストに直結する。制作業務がスタートする時点で、ディレクターから提示されるコストの目標を制作担当者は常に念頭に置き、できるだけ無駄のないように作業を進める。

④ 納期に対する意識

使用説明の納期は、多くの場合、製品やサービスの出荷スケジュールによって決定される。納期の遅れは、製品やサービスの出荷の遅れを招く原因となるので、決められた納期を守ることは、制作担当者に課せられた大きな責任のひとつである。

⑤ 自己啓発に対する継続的努力

利用者に素晴らしい利用経験をもたらすための使用説明を作り出すためには、制作担当者は市販されている書籍の書き手に負けない、魅力的な文章を表現する力を身に付ける必要がある。日頃から、分野を問わず魅力的な表現に接し、読み手の心に響くように表現するにはどのような工夫や配慮があるのかを知ろうとする探求心を持つことが大切だ。また、現在の使用説明の制作技術は、速いスピードで変化と発展を続けている。制作基盤を支えるテクノロジーや制作ツールの分野でも改良と革新が次々に行われる。これらの動向について関心を抱き、新たなテクノロジーや制作ツールを取り入れようとする意欲を持ち続けなければならない。

⑥ 知識

制作担当者には、次のような業務知識が求められる。

【使用説明に関する知識】

使用説明と制作技術、「使用説明」とは何か、使用説明の制作実務、伝達の要となるコンテキスト

【各制作工程で求められる知識と技能】

使用説明の制作工程、情報アーキテクチャーの設計、表現設計、執筆、作図、印刷媒体の版下データの作成、印刷・製本、画面表示データの作成・組込・配布

【複数の工程にかかわる知識と技術】

カラー表現、翻訳と多言語展開、品質管理と保守

【制作基盤技術とツール活用】

制作基盤とは何か、フォントと汎用データフォーマット、執筆ツール、作図ツール、レイアウトツール・オーサリングツール、翻訳ツール、校正支援ツール

【付帯技術】

コンプライアンス、社会的配慮、標準化と規格、認知科学、使用説明に関連する国際規格や国家標準、安全表記に関する国際規格や国家標準

5. 取説新時代 ～テクニカルコミュニケーターはどう変わるべきか！～

スマートフォンやタブレット端末の爆発的普及は、紙メディア依存からの本格的脱却を可能にした。域内自由貿易協定が政治の中心テーマとなる時代は、ガラパゴス化している国内制度依存から国際標準準拠への転換を私たちに強く促すであろう。2012年に制定された使用説明の国際規格 IEC 82079-1は、Web など電子的手段での使用説明の提供を積極的に奨励している。同時に、使用説明を取扱説明書に限定せず、カタログや Web サイト、ラベルや包装、機器内蔵の操作ガイダンスや UI をも含むものとして再定義された。

これまで、テクニカルコミュニケーションは PC や情報家電を主な適用対象として発展してきたが、これからは輸送機器、産業機器、医療機器なども強く意識したテクニカルコミュニケーション技術へと昇華していくであろう。

このような変化を前にして、私たちテクニカルコミュニケーターは、どのようなテクニカルビジョンを描くべきだろうか。私たちはどう変わるべきなのだろうか。

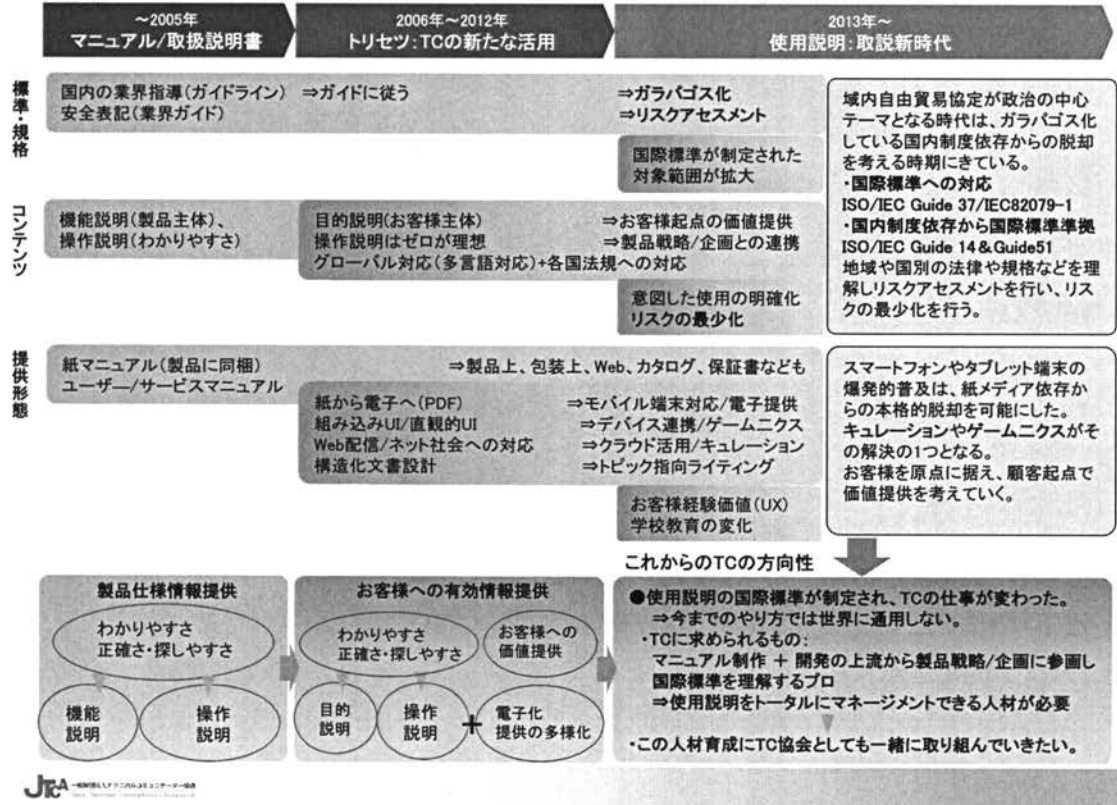
この命題について、2013 年 8 月に開催されたテクニカルコミュニケーションシンポジウム 2013 東京開催 25 周年記念イベントにおいて、次のようにまとめた。

・使用説明の国際標準が制定され、TC の仕事が変わった。今までのやり方では世界に通用しないことを、理解することである。これからの TC に求められるものは、マニュアル制作業務に加え、開発の上流から製品戦略/企画に参画し、国際標準を理解するプロになることが第一歩である。つまり、使用説明をトータルにマネジメントできる人材が必要である。

この人材育成に一般財団法人テクニカルコミュニケーター協会としても、企業と一緒に取組んで行きたいと考えている。

より詳しくは、次の図 取説新時代：テクニカルコミュニケーターはどう変わるべきか で確認ください。

取説新時代：テクニカルコミュニケーター（TC）はどう変わるべきか



自然言語処理の発展に向けた情報共有・討論

開催報告

株式会社富士通研究所

岩倉友哉

iwakura.tomoya@jp.fujitsu.com

1. はじめに

本稿では、著者が主催者の一人として関わった、2014年3月開催の言語処理学会第20回年次大会ワークショップについて紹介する。このワークショップは、タイトルの通り、自然言語処理の発展に向けた情報共有・討論を行うことを目的とした。また、普段の研究発表を中心としたワークショップとは異なる趣旨であるため、会場での議論を促進することに重きを置き、次のような工夫を凝らした。

- 長めの質疑時間を割り当てるようにした。今回は、発表と質疑の時間をそれぞれ10分に設定した。
- 全てのセッションを、全発表が終了した後に、まとめて、質疑の時間を取るパネル形式とした。

当日は、130名以上の参加者を迎え、活発な議論が行われた。また、今後につながる多くの有益な意見も得られた。以降、本ワークショップの発表概要を紹介し、著者の所感を述べる。最後、本報告書をまとめる。

2. 発表概要

本ワークショップは、6つのセッションで構成され、自然言語処理の研究者15名に発表いただいた。ここでは、それぞれのセッションの発表の概要について紹介する。発表内容については、予稿集およびワークショップホームページ

にある発表資料もご参照願いたい。¹ なお、本稿の発表者の所属は、2014年3月時点のものとし、以降、全ての方の敬称は「氏」で統一させていただく。

最初のセッション「言語処理と機械学習」では、鈴木潤氏（NTT）と工藤拓氏（Google）に発表いただいた。鈴木氏からは、自然言語処理分野における機械学習手法の解釈の整理および位置づけについて、工藤氏には、機械学習を用いた自然言語処理アプリケーション開発に必要な心得について発表いただいた。

2つ目の「若手の主張」セッションは、金子貴美氏（お茶大）と吉田光男氏（筑波大）に発表いただいた。金子氏からは、“経験主義”と“合理主義”に関する疑問点などについて研究者にインタビューを行った結果を発表いただき、吉田氏には、スタートアップ企業との産学連携の可能性について発表いただいた。

3セッション目「現実的な自然言語処理へ向けて」では、徳永健伸氏（東工大）、飯田龍氏（東工大）、松林優一郎氏（東北大）に発表いただいた。徳永氏からは、計算言語学分野における研究者の関心の推移調査結果の紹介および、“そもそもの言語処理研究”に対する問題

¹ 予稿集と一部の発表資料はワークショップのページから入手可能である。
<https://sites.google.com/site/nlp2014ws/home/wakushoppupuroguramu/>

提起をいただいた。飯田氏は、現状の研究におけるエラー分析の不十分さを指摘され、“ポストアノテーション”という、実験結果に対するアノテーションを提案された。松林氏からは、自然言語処理の解析技術の発展に必要となる”言語現象分析”と”言語資源構築・共有のあり方”について発表をいただいた。

4 セッション目では、「現実的な自然言語処理へ向けて」として、横野光氏（NII）と 西川仁氏（NTT）に発表いただいた。東ロボプロジェクトの一員である横野氏からは、“書かれていることを理解する”ということに対する問題提起をいただいた。また、西川氏からは、企業での要約技術の研究という観点で、現実的な要約実現のために解決すべき課題について発表いただいた。

5 つ目のセッション「大学・企業における今後の言語処理研究テーマ」では、関根聡氏（楽天技研 NY&NYU）、乾健太郎氏（東北大）、松井くにお氏（ニフティ）の 3 名に発表いただいた。関根氏は、言語処理技術の改善のためのエラー分析の重要性を述べられ、エラー分析に関するワークショップ開催の提案をされた。乾氏は、言語の意味の問題に踏み込んでいくことの重要性を発表され、また、幅広い内容の論文を集めるために”査読を大らかに”との提言をされた。松井氏からは、ビックデータ活用における個人情報保護のための自然言語処理技術の研究について提案された。

最後の 6 セッション目では、「これから 10 年の自然言語処理」という題目にて、鶴岡慶雅氏（東大）、持橋大地氏（統数研）、松本裕治氏（奈良先端大）に研究の方向性について発表いただいた。鶴岡氏からは”世界モデル”と対比させながらの自然言語理解に向けた研究の方向性、持橋氏からは自然言語処理研究における広い・深い統計の利用について、松本氏からは

文の意味理解に向けた研究について発表いただいた。

3. 所感

本ワークショップは、通常の学会発表ではあまり耳にすることがない、研究への想いを聞くことができる場であった。また、質疑においては、計算言語学だけでなく、言語学、機械学習など、様々な分野の研究者を交えた議論が行われた。これら発表や議論を通して、多くの研究者が共通に持つ意見や課題がいくつか明らかになった。ここでは、その中で、著者の印象に残った 3 点を紹介させていただく。

1 点目は、「エラー分析」の重要性である。現在までに、形態素解析や構文解析、固有表現抽出など多くの自然言語処理タスクにおいて高い精度が得られるようになった。しかし、これらのタスクにおいても、未解決の問題が多く残されているのも事実である。そこで、今後、意味に踏み込んだ研究を進めるためにも、既存の技術の「誤り分析」をきちんと行った上での研究が重要であるということが多くの研究者の問題意識であると判明した。このワークショップを受けて、様々な言語処理技術においてエラー分析を行い、次に対処すべき課題を明確にすることを目的にしたプロジェクト「Project Next NLP」² が開始された。

2 点目は、当然のことではあるが、より良い研究を行うための基礎の重要性である。何名かの発表者から、基礎として学ぶべき内容や、推薦図書が紹介された。

3 点目は、「自分が好きなテーマを見つけることが重要」ということである。実際、今回の発表者からは自然言語処理に対する情熱が感じられ、若手研究者にとっては良い刺激となっ

² Project Next NLP ホームページ：
<https://sites.google.com/site/projectnextnlp/>

たのではないかと思われる。

また、今回の論文や発表資料には、通常の論文にはなかなか記載されることがない、自然言語処理の現状に対する意見や見解、今後への提言が数多く含まれている。今回の発表者以外の研究者の意見は、金子氏が実施したインタビュー結果にて、知ることができる。是非、そちらの資料も、御一読願いたい。³

4. おわりに

今回のワークショップは、普段の研究発表や会話の端々から垣間見えるものの、直接には聞くことが少ない研究者の想いや現状への見解を率直に聞くことができる大変有意義な場になった。また、他分野の研究者との交流を深める場としても非常に良いワークショップであったと思われる。これもひとえに、発表者および参加者、本ワークショップの開催にご協力いただいた皆様のおかげである。この場を借りて、厚くお礼を申し上げる。また、本ワークショップが、若手研究者の動機づけや、今後の研究テーマの創出に繋がり、自然言語処理分野の発展の一助となれば幸いである。

³金子氏のインタビューのまとめ：
<http://www.kimi-kaneko.miraiserver.com/NLP2014WS.pdf>

外国特許調査を日本語で行える MT Plus II /NEF スプレッド

本間 奨

日本発明資料株式会社

1.はじめに

特許庁データ（図1）から明らかなように、日本特許のみならず米国特許・中国特許・欧州特許・韓国特許など外国特許調査が必須になってきている。特許庁 IPDL や日本特許情報機構 JGPG 検索サービスでは、いち早く外国特許公報を日本語で串刺し検索して、公報テキストを日本語で表示する検索サービスを提供している。日本発明資料株式会社では、昨年秋より特許庁向けに中国実案の機械翻訳事業を請け負い、この一旦を担っている。ここに用いられている翻訳システムが MT Plus である。

既存のルールベース翻訳ソフトでは重要技術キーワードとなる英単語の訳し分けの誤りが多い^{注1)}ことから、MT Plus では統計的機械翻訳エンジンを採用して訳語精度向上を図った。あわせて、非テキスト情報処理機能向上、発明の主題、構成要素分離性能を向上させた。この結果、外国公報の発明特定事項をより迅速に理解することができ、外国特許調査生産性の向上に寄与することが期待される。

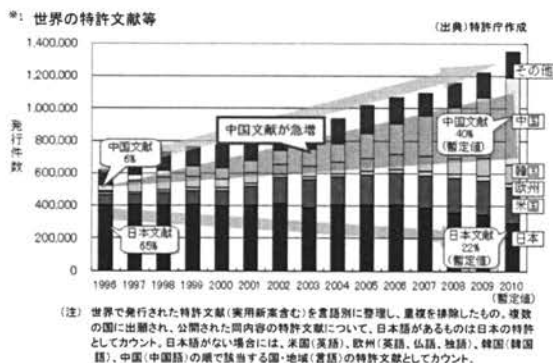


図1 世界の特許出願件数動向（特許庁データより）

注1) 英文公報では、公報中に繰り返し出現する発明のキーワードで、同じ単語が別の公報では異なる意味をもつことが少なくない。例えば、communicate は「連通する」、「通信する」、develop は「開発する」、「現像する」、application は同一公報内でも「出願」、「塗布」、「印加」、「応用」、「アプリケーション」などの訳し分けが必要となる。統計的機械翻訳以外では訳し分けが難しい。

2. MT Plus II とは

MT Plus は英文特許翻訳に特化した特許専用翻訳システムであり、訳語と語順の正確性を向上させるため、(独)情報通信研究機構(NICT)と共同開発した統計的機械翻訳エンジンをシステムのコアとして採用した。この翻訳エンジンは、日米対応特許の訳文対から抽出した約3億のエントリーフレーズを有している。

MT Plus II は、MT Plus の精度向上とあわせて多言語対応を可能とした。現時点では、英語・中国語・韓国語から日本語に翻訳可能な多言語翻訳システムであり、社内評価結果では他社比較で高精度であることが確認されている。

3. MT Plus II の翻訳精度

中日、韓日の翻訳例を以下に示す。MT Plus II では、手がかり句分析により、公報特有の文体に従った翻訳文となるよう工夫されている。

中国実案翻訳結果の比較例

CN202577804 請求項1

一种太阳能建筑结构，其特征在于，包括混凝土构件和光伏板，所述混凝土构件与所述光伏板贴合在一起，且所述光伏板位于所述混凝土构件的外侧

MT Plus II 中日翻訳

ソーラー建物構造であって、以下のものを含む。コンクリートとソーラー電池パネル、前記コンクリート部材と嵌合しソーラー電池パネルと前記ソーラー電池パネルは、前記コンクリート部材の外側に配置されることを特徴とする、ソーラー建物構造。

A 社中日翻訳

太陽電池パネル部材とコンクリートを含むが、具体的な部材や光発電パネルを貼り合わせて、太陽光発電パネルが外に配置され、前記太陽建物の構造は、コンクリート部材と述べた。

B 社中日翻訳

太陽建築構造，その特徴は，コンクリート部材と太陽光板，前記コンクリート部材と前記太陽光パネル貼り合わせ，且つ前記太陽光板は前記コンクリート部材の外側に位置している。

韓日翻訳結果の比較例

KR2007-0097923 発明の名称

리튬 전지용 양극 활물질, 그 제조 방법 및 그를 포함하는 리튬 이차 전지

MT Plus II 韓日翻訳

リチウム電池用正極活物質、その製造方法及びそれを用いたリチウム二次電池

A 社韓日翻訳

リチウム電池用正極活物質、その製造方法および彼含むリチウム二次電池

C 社韓日翻訳

リチウ電池用正極活物質、その製造方法及びそれを含むリチウム二次電池

精度比較は以下の中国実案請求項（中日）、米国特許・中国実案 DOCDB（英日）について実施した。

技術分野と公報原文

- | | |
|---------------|-----------|
| ①化学（化学式を含む） | 中国実案、米国特許 |
| ②物理（数式含む） | 中国実案、米国特許 |
| ③自動車（電気自動車等） | 中国実案 |
| ④マーキング（プリンタ等） | 中国実案 |
| ⑤医薬（遺伝子組み換え等） | 中国実案 |
| ⑥電気通信 | 中国実案 |
| ⑦機械 | 中国実案 |
| ⑧建築 | 中国実案 |

前記公報原文をもとに、現時点で、最も高精度と考えられる 3 社の特許検索サービス/翻訳システムと MT Plus II を以下の (1) ～ (3) の観点で比較分析をおこない、その結果を図 2 にまとめた。

(1) 評価対象の段落

- ①発明の名称
- ②要約
- ③請求の範囲 第一請求項、第二請求項

(2) 評価方法

- ①(3) で定義される段落単位あたりの誤り数
- ②段落中に同じ語で複数誤りがあっても 1 とする

(3) 誤りの定義

- ①訳語が付与されない未訳語があれば未訳語 1 につき誤り数 1 とする
- ②訳し分けができていない語（例：light を光と訳すべきところを軽いと翻訳するなど重要な用語での誤り）があれば誤り数 1 とする。
- ③関係代名詞や表現の誤訳（例：前段とのつながりで、where を「どこ」と訳すなどの誤り
- ④数式の誤り 原文比較での誤り
- ⑤化学式の誤り 構造式の誤り（語順含む）
- ⑥要素記号（構成要素となる名詞の単語と対になる語尾の数字やアルファベット）の対応関係の誤り
- ⑦数値範囲指定の誤り
- ⑧日本特許公報では用いられない用語

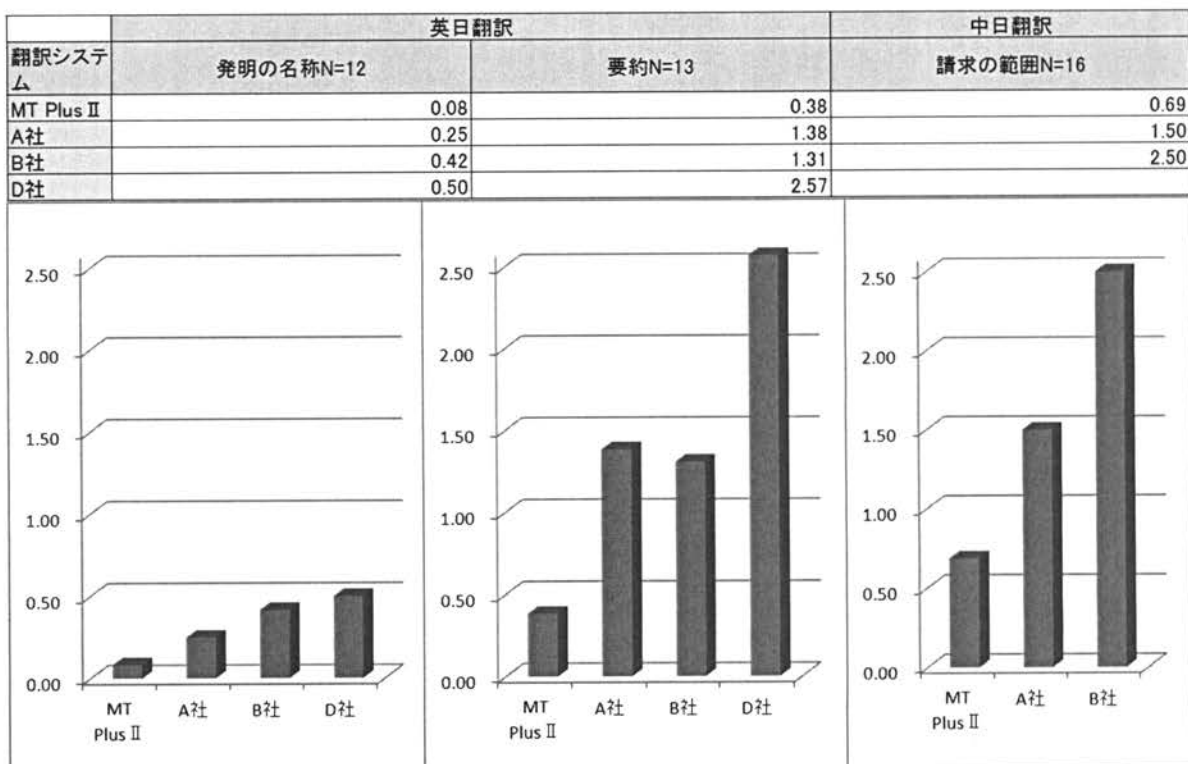


図 2 精度比較 表の値とバーの縦軸は段落 1 件あたりの誤り数で、値が小さいほど高精度である。

MT Plus II では、NICT 新翻訳エンジンと独自辞書により、テキスト部の翻訳精度が高いこと、さらに、数式や化学式(上付き文字や下付き文字に隣接する化合物名)などの非テキスト部についても正確に翻訳できるよう改善されており、総合評価では図2にあるように他社翻訳結果に比べて高い翻訳精度を有している。

4. NEF スプレッドとは

NEF スプレッドは、米国特許、中国特許・実案、韓国特許、ヨーロッパ特許、WO 特許の五大特許庁の外国公報を多言語翻訳システム MT Plus II により日本語化された PDF 明細書である。公報原文は XML 形式で各国特許庁より提供されており、NEF スプレッドは数式や化学式、図、改行などの XML タグ情報を翻訳結果に反映可能とした。あわせて、NEF スプレッドは以下に述べるメリットを有する 2 頁見開き形式(スプレッド)で提供される。

一般に、公報種別(公開・登録)や国により公報の段落の順序が異なるが、NEF スプレッドでは、書誌情報、発明の名称、要約+代表図、請求の範囲、詳細な説明、図面という順番に規格化されている。

その結果、NEF スプレッドでは、最初の頁に発明特定事項に関する重要情報が、次ページ以降に詳細な説明、最後に図面という順番でどの公報でも同じように表示される。そのために、翻訳結果を独自の XML 形式に変換してから XML⇒PDF 出力を行っている。

NEF スプレッドの大きな特徴は、図3のように、2 頁分の情報が見開き状態で横長表示され、PC ディスレイの縦横比率に適合するため PC 画面上で見易いというメリットを持つ。2 頁を同時に表示するもうひとつのメリットは、発明の名称、要約、代表図、請求の範囲が 1 画面に集約されるため、発明特定事項の理解がしやすくなり、調査する上で使いやすというメリットがある。

図3に数式が発明の構成要素になっている中国実案の NEF スプレッドの例を示した。図3では、DOCDB 英文タイトルと要約を併記し、英文をもとに日本語を生成した。請求の範囲以下は中国語原文から日本語を生成した。この例では、請求の範囲に数式が含まれるケースであるが、原文 XML の数式タグに NEF スプレッドが対応しているため、フォントの数式が表示可能となっている。

中国実用新案公報 (Y) 【公報番号】CN 201344015Y 【公報発行日】20091111 【出願番号】200820124583.9 【公開日】20081225 【IPC 分類】E 01B 9/087 【優先権情報】CN20082124583U 2008.12.25 【出願人情報】North China Electric Power University 102206 Beijing Germany wen Zhongzhong North China Electric Power University at the door 【発明の名称】Transitional disk cutter ring with asymmetrical inner side and outer side cutting edge angles 非対称な内側と外側切刃エッジ角度を有する過渡ディスクカッターリング 【要約】 The utility model belongs to disk cutter equipment, and particularly relates to a transitional disk cutter ring with asymmetrical inner side and outer side cutting edge angles, which is installed in the transition region on a cutter head of a full face tunnel boring machine (a full-section rock tunneling machine and shield). Two side surfaces of the cutter ring edge are asymmetrical with the respect to the plane on which the cutting edges are positioned, the cutting edge angle on the inner side is smaller than the cutting edge angle on the outer side by 2 Theta, whereas, Theta equals arctan (sin Phi cos alpha) / (cos Theta sin theta minus sin Phi sin alpha cos theta), alpha refers to a transitional disk cutter installation angle, and Theta refers to a rock contact angle at one third of the penetration. The asymmetrical design of the inner side and the outer side cutting edge angles results in not only a reduction in the slip volume of the transitional disk cutter ring, but also an increase in the rock breaking live area of the transitional disk cutter ring, thereby effectively reducing the consumption of the transitional disk cutter ring during rock breaking process, prolonging the service life of the transitional disk cutter ring, and effectively reducing the energy consumption of the transitional disk cutter ring during the rock breaking process. The preliminary estimate shows that, where the utility model is used, the consumption of the transitional disk cutter ring is reduced by about 20 to 30 percent, the service life of the cutter ring is prolonged nearly one year, project duration is effectively shortened, and construction cost is decreased markedly. 実用新案は、ディスクカッター装置に属しており、非対称な内側と外側切刃角度、フルフェーストンネルボーリングマシン(全断面岩石掘削機とシールド)の刃ヘッド部の過渡領域に設置された過渡的なディスクカッターリングに関するものである。カッター環状エッジの二側面は、切刃縁部が位置する平面に対して非対称である。内側切刃の切刃角度は、2θ、そこにおいて、θは、だけ外側切刃角より小さい値になっている。θは、arctan (sin φ cos α) / (cos θ sin θ - sin φ sin α cos θ) に等しく、αは過渡的なディスクカッター設置角度を参照し、θは、貫通(perforation)の1/3で接触角である。内側切刃と外側切刃の非対称設計は、過渡的なディスクカッターリングのスリップ量の減少をもたらすだけでなく、可動ディスクカッターリングの運動断続立き領域が増加し、ロックアップ過渡プロセス中に過渡的なディスクカッターリングの消耗を効果的に低減する。可動ディスクカッターリングの寿命を、及び岩石破壊プロセス中に過渡的なディスクカッターリングのエネルギ消費の削減に有効である。予備的評価は、実用新案は用いられる場合には、過渡的なディスクカッターリングの消耗は約20〜30%だけ低減されること、カッターリングの寿命は長く倍増、プロジェクト継続時間が効果的に短縮されること、建設コストを大幅に減少させる。	【請求の範囲】 【請求項1】 内側及び外側コーナエッジ過渡アシメトリディスクカッターナイフリング、非対称の平面の両側にリングブレードナイフブレードが、外側コーナエッジ角度の内側エッジがθの角度より小さい場合であって、以下のものを含む。 $\theta = \arctan \frac{\sin \phi \cos \alpha}{\cos \theta \sin \theta - \sin \phi \sin \alpha \cos \theta}$ θは、ディスクカッター取付角度、αは角度第3ピク岩石浸透と、αはトラッキング過渡ディスクカッター環状エッジであり、過渡ディスクカッター切刃最大深さの中心を通過するトラックの半径とθの角度を指すことを特徴とする。内側及び外側コーナエッジ過渡の非対称性ディスクカッターナイフリング。 【請求項2】 請求項1であって、以下のものを含む。内側と外側角非対称過渡ディスクカッターナイフリングは次のように特徴付けられる。 $\theta = \arctan \frac{2\theta}{2\theta + \theta_{max}}$ $\phi = \arctan \frac{2\theta - (2\theta_{max} - \theta)}{R^2 - 2R_{max}(R - 1) - 2\theta - (\theta_{max})^2}$ θは浸透と過渡ディスクカッターナイフリングの半径、R、過渡ディスクカッターナイフリングの半径であることを特徴とする。請求項1に記載内側及び外側コーナエッジ過渡アシメトリディスクカッターナイフリング。 【技術分野】 実用新案はディスクカッター装置、であって、特に、カッター刃角非対称過渡ゾーンインストールディスクカッターナイフリングの外側に位置するフルフェーストンネルボーリングマシン(TBM)機やシールドに関するものである。 【背景技術】 フルフェーストンネルボーリングマシンカッターは中心ホブ、ホブ、ホブがインストールされているのが一般的であって、ナイフエッジ過渡は、半面の幾何学的形状の側に存在する過渡的なフルフェーストンネルボーリングマシンカッターナイフロードでは線対称性とし、この過渡的なツールは、使用時に磨耗、消耗が大きい、複雑な交換を必要と、建設コストを増加させる。過渡ホブ寿命を向上させる過渡の過渡ホブの磨耗をいかに小さくするかという現在の研究の重要な課題である。 【発明の概要】 内側と外側角非対称過渡ディスクカッターナイフリングに使用される実用新案は、寿命が短い欠点を排除、内側と外側エッジ過渡ディスクカッターナイフリングの両側非対称性は、両側の切刃ナイフブレードは外側エッジ角が小さいθ角よりその非対称性があるエッジ角度内側を参照し、以下のように特徴付けられる。 $\theta = \arctan \frac{\sin \phi \cos \alpha}{\cos \theta \sin \theta - \sin \phi \sin \alpha \cos \theta}$ する。また、αは過渡ディスクカッター取付角度が浸透する際の角度を第3ピク岩石と、カット点αの過渡ディスクカッター最大深さの中心を通過するトラックとθの角度の半径半径ディスクカッターがトラッキング過渡領域に入る。 $\theta = \arctan \frac{2\theta}{2\theta + \theta_{max}}$ $\phi = \arctan \frac{2\theta - (2\theta_{max} - \theta)}{R^2 - 2R_{max}(R - 1) - 2\theta - (\theta_{max})^2}$ θは、浸透と過渡ディスクカッターナイフリングの半径、R、過渡ディスクカッターナイフリングの半径を参照する。 θは2以上の値、θは2以上、またはすべてのθの過渡的なディスクカッターナイフリングの異なる仕様の設計は平均化されること、過渡的なディスクカッターナイフリングだけ作成された同一仕様は、製造および組み立てを容易にする。 実用新案は、以下の利点を有する。内側と外側角非対称設計は、スピン過渡ディスクカッターナイフ内の量を減少させる。だけでなく、過渡の過渡ディスクカッターナイフリングロック過渡領域、磨刃切れ時過渡岩石破壊ディスクカッターナイフホブ寿命の削減と過渡ディスクカッターナイフリングの寿命を向上させると同時に、消費エネルギーを過渡ディスクカッターナイフリングを効果的に低減する結果が増加する。予備的評価は、約20〜30%、2倍近いリングの寿命のために使用される過渡
--	---

図3 NEF スプレッド中国実案例。タイトル要約は DOCDB 英文から、請求項以降は中国語から翻訳を行った。

化合物の構造式が発明の主題になっている米国特許の NEF スプレッドの例を図 4 に示した。図 4 の例では、大量の数値実体参照コードや下付き記号タグが化合物名に接しているが、化合物名が未訳にならないような対応をしており、「C₅-C₁₀ アリーレン」のように化合物名が正しく翻訳される。また、発明の主題そのものが化学構造式になっているため、構造式が公報 XML の指示通り、正確にテキスト中に埋め込まれて表示される。

NEF スプレッドでは、原文が中国語、韓国語の場合、原文表示は行わないが、英語の原文が存在する場合は、図 3、図 4 に示すように段落単位に英文と翻訳日本語を上下に配置した。

5. おわりに

これまでの特許機械翻訳では、訳語精度の課題とあわせて、数式、化学式、文中への図挿入などの原文 XML 情報が欠落していたり、翻訳で破損することがあり、翻訳文と公報原文イメージを相互参照する必要があった。NEF スプレッドでは、テキストの高精度化とあわせてこれまであまり問題視されてこなかった非テキスト部の情報が損なわれ^{注2)} ないような改善を行った。

その結果、主要五大特許庁の外国公報を日本語で、公報原文を参照しなくとも素早く内容把握することが可能となり、調査生産性向上が期待されるものである。

特許検索サービス分野では、「はじめに」述べたように中国・韓国を含む五大特許庁から発行される外国特許の日本語対応が急務の課題であり、膨大な特許文献を精査するためには、正確で見やすい特許翻訳が強く望まれている。XML 対応の高精度な機械翻訳以外にその解決手段はなく、MT Plus II/NEF スプレッドが外国特許調査の生産性向上、ひいては研究開発の一助となることを願ってやまない。^{注3)}

最後に、(独)情報通信研究機構ユニバーサルコミュニケーション研究所多言語翻訳研究室 隅田英一郎室長、内山将夫主任研究員には統計的機械翻訳エンジンの技術移転にあたり多大なご協力をいただきました。ここに厚く御礼申し上げます。

注 2) その多くの要因は、公報原文 XML に翻訳システムが非対応であるために生じていた。

注 3) 2014 年 9 月リリース予定。本製品紹介の画面レイアウトは予告なく仕様変更されることがあります。

 米国公開特許公報 (A1)	【請求項1】 1. Compounds of formula (1)
【発明の名称】 THIOL DERIVATIVE DYES チオール誘導体染料 【要約】 Disclosed are thiol dyes of formula (1) wherein L₁ is hydrogen, C₁-C₁₂alkyl, or phenyl-C₁-C₄alkyl; X is C₁-C₁₂alkylene, C₂-C₁₂alkenylene, C₃-C₁₀cycloalkylene, C₃-C₁₀arylene, or C₃-C₁₀arylene-C₁-C₁₀alkylene, which may be interrupted by —O—, —NH—, —S—, —CO—, —SO₂—; Y is the residue of an organic dye Z is a group of formula (1b) or —C≡N, wherein A is O, S, or N; L₂, B is L₃, —OL₃, —NL₃L₄, or —SL₃; and L₂, L₃ and L₄, independently from other are hydrogen, C₁-C₁₂alkyl, C₃-C₁₂aryl-C₁-C₁₂alkyl. The compounds are useful for the dyeing of organic materials, such as keratin fibers, preferably human hair. 本発明は、以下のように開示されている。式(1)のチオール染料であって、以下のものを含む。L₁は水素、C₁-C₁₂アルキルであり、またはフェニル-C₁-C₄アルキルであること、XはC₁-C₁₂アルキレン、C₂-C₁₂アルケンレン、C₃-C₁₀シクロアルケンレン、C₃-C₁₀アリーレンまたはC₃-C₁₀アリーレン-C₁-C₁₀アルキレン、—O—、—NH—、—S—、—CO—、—SO₂—、と、Zは、式(1b)または—C≡Nの基である。Yは有機色素の残基であること、式、は、Oであること、と、S、と、あるいはN-L₂、と、BはL₃、と、—OL₃、と、—NL₃L₄、と、あるいは—SL₃、と、L₂、L₃、およびL₄、他から独立して水素と、C₁-C₁₂アルキル、C₃-C₁₂アリール基を含むC₁-C₁₂アルキルである。この化合物は、有機材料の染色、ケラチン繊維、好ましくはヒトの毛髪等に有用である。  (1)  (1b) 、からなることを特徴とする。式(1)のチオール染料。 【請求の範囲】	wherein L₁ is hydrogen, C₁-C₁₂alkyl, or phenyl-C₁-C₄alkyl; X is C₁-C₁₂alkylene, C₂-C₁₂alkenylene, C₃-C₁₀cycloalkylene, C₃-C₁₀arylene, or C₃-C₁₀arylene-C₁-C₁₀alkylene, which may be interrupted by —O—, —NH—, —S—, —CO—, or —SO₂—; Y is the residue of an organic dye Z is a group of formula (1b) or —C≡N, wherein A is O, S, or N; L₂, B is L₃, —OL₃, —NL₃L₄, or —SL₃; and L₂, L₃ and L₄, independently from each other are hydrogen, C₁-C₁₂alkyl, C₃-C₁₂aryl, or C₃-C₁₂aryl-C₁-C₁₂alkyl. 式(1)の化合物  (1) であって、以下のものを含む。L₁ 水素であること、と、C₁-C₁₂アルキルであり、あるいはフェニル基を含むC₁-C₄アルキルと、 XはC₁-C₁₂アルキレン、C₂-C₁₂アルケンレン、C₃-C₁₀シクロアルケンレン、C₃-C₁₀アリーレンまたはC₃-C₁₀アリーレン-C₁-C₁₀アルキレン、—O—、—NH—、—S—、—CO—、—SO₂—による遮断であってもよいこと、と、 Yは、有機色素の残基である。 Zは、式  (1b) あるいは—C≡Nと、式中ここで、はOであること、と、S、と、あるいはN-L₂と、 BはL₃、と、—OL₃と、—NL₃L₄と、あるいは—SL₃と、そして、 L₂、L₃ およびL₄、互いに独立する。水素であること、と、C₁-C₁₂アルキルであること、と、C₃-C₁₂アリールであること、と、あるいはC₃-C₁₂アリール基を含むC₁-C₁₂アルキルである。  (1) 、からなることを特徴とする。式(1)の化合物 【請求項2】 2. Compounds according to claim 1, wherein is formula (1) X is C₁-C₁₂alkylene 請求項1の化合物であって、以下のものを含む。式(1)に式中、XはC₁-C₁₂アルキレンである。 上記からなることを特徴とする。化合物。 【請求項3】 3. Compounds according to claim 2, wherein

図 4 NEF スプレッド米国特許例。化合物・構造式が発明の主題、構成要素となっている。

AAMT/Japio 特許翻訳研究会 拡大評価部会の活動について

江原暉将

山梨英和大学(AAMT/Japio 特許翻訳研究会 拡大評価部会 部会長)

1. はじめに

翻訳の評価は難しい。人間の翻訳でも、いろいろな議論がある。機械翻訳(MT)においては、AAMT 機械翻訳課題調査委員会が、利用者に対するアンケートの形で、広範な観点から MT の評価を実施している[1][2]。

AAMT/Japio 特許翻訳研究会では、特許文書の機械翻訳に関する様々な研究を行ってきた。その中で MT の評価に焦点を合わせた「特許文書の機械翻訳結果評価方法検討会」を 2012 年に開催した[3]。この検討会では、研究者、開発者、利用者がそれぞれの立場で評価についての見解を述べ、総合討論を行った。当研究会としては、この検討会を受けて更に議論を深める必要性を認識し、拡大評価部会を設置した。本部会の活動の詳細については[4][5]で報告されており、本文では概要を紹介する。

2. 拡大評価部会の組織と目的

拡大評価部会の組織は、親委員会である AAMT/Japio 特許翻訳研究会の委員を中心として、外部専門家にも加わっていただき構成されている。現時点での構成員は 11 名である。

本部会の活動目的は以下のとおりである。まず、MT の評価を行うには、「評価の目的」を定めないと議論が発散する。特許文書の MT の利用目的としては、(1) 技術調査、(2) 外国への出願、(3) 特許庁での審査業務での利用、が考えられる。その中で、本部会では、「技術調査目的」のために特許文書を機械翻訳する場合に目的を絞って議論を進めている。そして、評価手法の理想形、理想を実現するための課題、課題克服への道程を明らかにすることを活動目的としている。

具体的には、次の 3 つの項目に分けて活動を行っ

ている。

- ・ 人手評価
- ・ 自動評価
- ・ 評価用テストセット

3. 人手評価

人手評価は MT の結果を人間の評価者が評価するものであり、手法自体は ALPAC レポートでも用いられている歴史の古いものである。人手評価は、コストはかかるものの、信頼性が高く、実用システムの評価で多く用いられている。

本部会では NTCIR-9 と NTCIR-10 での人手評価手法を中心に検討した。NTCIR は多くの参加者のシステムを評価する評価型ワークショップであり、NTCIR-7 から特許の機械翻訳タスクが行われてきた(PATMT または PatentMT)。NTCIR-9 と 10 では人手評価基準として Adequacy と Acceptability を用いている。Adequacy は従来からある人手評価基準であり、翻訳の正確性を評価している。一方、Acceptability は NTCIR-9 で独自に考案された評価基準であり、特許文書の MT 結果を技術調査目的に利用したときの有効性について評価するものである。さらに、NTCIR-10 では特許審査時に MT を利用することを想定した評価も実施している。現在進行中の NTCIR-11 では、特許機械翻訳タスクを実施していないため、今後は他のワークショップと連携して人手評価基準の充実に努めてゆきたい。

4. 自動評価

自動評価は、MT 結果と理想訳(基準訳ともいう)との近さを何らかの方法で自動的に計算するものであり、人手評価と比較してコストが低い。自動評価基準は、BLUE から始まり、その後、多くの手法が提案されている。精度も徐々に高くなってきてい

るが、未だ十分であるとはいえない。

本部会では、まず各種自動評価基準の性能や得失を確認することから始めている。具体的には、WMT や NTCIR のデータを用いて各種の自動評価結果と人手評価結果との相関を比較することで調査している。今後は、より良い自動評価基準を考察するとともに、人手評価を補完するために自動評価を利用する手法についても検討したい。

5. テストセット

AAMT では MT 評価のためのテストセットを提供している[6]。これらは、基本文を対象としたものであり、文型や語彙が特徴的な特許文書に適用するには内容が不足する。そこで、本部会では特許文書を対象にした中日テストセットを構築している。そのために文献[7]やパテントファミリーを利用して中日の平行コーパスを整備し、それらに、MT 評価のための設問を設定している。2013 年度末現在で、32 文のテスト文を整備し、104 の設問を設定した。しかし、まだ不足しており 2014 年度も引き続き整備を継続する予定である。

6. おわりに

AAMT/Japio 特許翻訳研究会 拡大評価部会の活動について概要を紹介した。特許分野では MT が実用されており、MT 結果の正確な評価の重要性はますます高まっている。しかし本評価部会の活動目的である「評価手法の理想形、理想を実現するための課題、課題克服への道程を明らかにする」ことは、緒についたばかりであり、やるべきことが多く残されている。

人間の翻訳に対してではあるが、文献[8]に次のような記述がある。「たとえ文学ではなくても、翻訳をするからには、これだけはおろそかにしていただきたくない。それは、美しい日本語を書くということ。ケーキの焼き方でも、カタログの説明でも、意味さえとおればよいという、機能一点ばりの考えは捨ててください。」

MT の結果に対してこのような評価が可能になる日までには、未だ相当な時間がかかるであろう。

「実用一点張り」かもしれないが、より良い評価手法の実現を目指して今後とも努力してゆきたい。

本部会の活動に対する読者のみなさまのご批判やご提案を歓迎いたします。ご意見は AAMT/Japio 特許翻訳研究会のホームページ (<http://aamtjapio.com/>)にある "Contact us" のボタンからお願いいたします。

参考文献

- [1] AAMT 機械翻訳課題調査委員会 WG2：英日・日英機械翻訳に関するアンケート結果報告，AAMT ジャーナル，No.45，Jun. 2009.
- [2] AAMT 機械翻訳課題調査委員会 WG2：英日・日英機械翻訳に関するアンケート結果報告，AAMT ジャーナル，No.50，Oct. 2011.
- [3] AAMT/Japio 特許翻訳研究会：特許文書の機械翻訳結果評価方法検討会資料，Sept. 2012.
<http://aamtjapio.com/kenkyu/discussion01-01.html>
- [4] 日本特許情報機構：平成 24 年度 AAMT/Japio 特許翻訳研究会報告書，Mar. 2013.
http://aamtjapio.com/kenkyu/files/kenkyu03/AAMT_Japio_20130331s.pdf
- [5] 日本特許情報機構：平成 25 年度 AAMT/Japio 特許翻訳研究会報告書，Mar. 2014.
http://aamtjapio.com/kenkyu/files/kenkyu04/AAMT_Japio_20140331.pdf
- [6] AAMT 機械翻訳課題調査委員会 WG1：AAMT 機械翻訳文テストセット（第 1 版）公開に向けて，AAMT ジャーナル，No.45，Jun. 2009.
- [7] 特許庁：日本語特許出願書類の中国語への機械翻訳に関する調査報告書，Feb. 2011.
http://www.jpo.go.jp/shiryou/toushin/chousa/pdf/tokkyo_kikai_honyaku_houkoku/chn_all.pdf
- [8] 別宮貞徳：翻訳読本，講談社現代新書，p.19，April, 1979.

これまでの AAMT Forum メールマガジン Vol.2

機械翻訳課題調査委員会 WG1、WG2

はじめに

前号に引き続き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 13 号から第 20 号です。

今回も様々な情報がありました。14 号では AAMT 会員の方から提供された情報を配信しました。今後も引き続き、可能な限りご提供いただいた情報には対応させていただきたいと思います。メールマガジンの趣旨に適した情報がありましたら、お知らせいただけましたら幸いです。また、その他、お気づきの点などがありましたら AAMT 事務局までご意見をお寄せいただけましたら幸いです。

では、メールマガジン第 13 号から第 20 号までを再びお届けします。

AAMT Forum メールマガジン

13 号 (2014/01/24 配信)

こんにちは。AAMT Forum メールマガ担当です。厳寒の候、寒さはまだまだ厳しく、各地で雪も降りましたが、いかがお過ごしでしたでしょうか。路面も凍結しやすいです。足元にはお気をつけください。では、メールマガジン第 13 号をお届けします。

■地図上で言語を確認できる!? ビジュアル型翻訳サイト『European word translator』登場

ヨーロッパの言語や地理に興味がある人

にうってつけの翻訳サイトが公開されました。この翻訳サイトでは、英単語を入力するとその結果が 35 パターン、ヨーロッパ地図上で表示されます。

http://www.excite.co.jp/News/net_clm/20140117/Getnews_496035.html

■ファーストデザインが全クリエイターのプロフィールを英語対応に

エフ・プラット株式会社が運営するデザイン・イラストを WEB で簡単発注できる「ファーストデザイン」(<http://www.fast-d.com/>)は、クリエイター(デザイナーやイラストレーター等)のプロフィールを登録完了と同時に英語化し、海外のクライアント(顧客)が日本のクリエイターに漫画やイラスト、デザインの制作を依頼できる環境を整えました。

<http://sankei.jp.msn.com/economy/news/140116/prl14011609500020-n1.htm>

■40 カ国語以上に対応する翻訳アプリ「Bing 翻訳」が音声入力をサポート

入力したテキストの翻訳だけでなく、カメラで撮影したテキストの翻訳もサポートしている「Bing Translator」が、言葉を聞き取ってテキスト化し、翻訳してくれるようになりました。

<http://www.forest.impress.co.jp/docs/news/>

AAMT Forum メールマガジン

14号 (2014/02/07 配信)

こんにちは。AAMT Forum メールマガ担当です。暦の上では春となりましたが、まだまだ厳しい寒が続いています。関東や日本海側では雪も降りましたが、体調などは崩しておられないでしょうか。では、メールマガジン第14号をお届けします。

■TAUS がテクニカルコミュニケーター協会とパートナー提携

TAUS とテクニカルコミュニケーター協会(TC 協会)がパートナーとして提携しました。この提携により、TC 協会の会員である多くの日本企業が、TAUS やそのメンバーが作成したレポート、オンライン上のリソース、ツール、データ、マトリックスにアクセスできるようになりました。

<https://www.taus.net/taus-and-japan-technical-communicators-association-announce-partnership-japanese>

■ASPEC (アジア学術論文抜粋コーパス) 公開

独立行政法人科学技術振興機構(JST)と独立行政法人情報通信研究機構(NICT)が共同で作成したアジア学術論文抜粋コーパス(Asian Scientific Paper Excerpt Corpus: ASPEC)が公開されました。ASPEC は、約300 万対訳文からなる日英論文抄録コーパス(ASPEC-JE)と約68 万対訳文からなる日

中論文抜粋コーパス(ASPEC-JC)から成る、世界初の大規模な論文対訳コーパスです。

<http://orchid.kuee.kyoto-u.ac.jp/ASPEC/>

■CIEC/第101回研究会「機械翻訳との上手なつきあい方」を3月に開催

コンピュータ利用教育学会(CIEC)は、第101回研究会「機械翻訳との上手なつきあい方」を3月8日に開催します。現時点での機械翻訳ソフトの実力を知り、その可能性と限界、問題点などを把握し機械翻訳とのつきあい方を考えます。

<http://ict-enews.net/2014/01/28ciec-2/>

■日常日本語を平易でわかりやすい英語に翻訳する格安なネットサービスを開始
(以下、AAMT 会員様からの情報です)

ラングテックは、ネットにつながったパソコン、タブレット端末、スマホで利用できる、英語での情報発信(パーソナル/ビジネスEメール・ブログ・日常会話)や英語学習の手軽なツールとなる Laurel-II 日常語日英翻訳サービスを開始しました。

<http://www.languetech.co.jp/>

AAMT Forum メールマガジン

15号 (2014/02/21 配信)

こんにちは。AAMT Forum メールマガ担当です。関東や東北では記録的な大雪が降るなど、未だ冬を強く感じるこの頃、いかがお過ごしでしょうか。寒い日が続きましたが、週明けには気温も上がり、ようやく春の気配を感じられるそうです。では、メールマガジン第15号をお届けします。

■外国人旅行者向けに翻訳機能付きの「スマートメニュー」、エクスウェア

エクスウェアは、飲食店向けに iPad などのタブレット端末を利用し、外国人への接客を支援する「スマートメニュー」を発表しました。

<http://www.sbbit.jp/article/cont1/27555>

■wWeb: 英語の原文と日本語翻訳が同時に読めるブラウザ！海外ニュース調査が捗ります。

2 画面でウェブが見られるブラウザアプリの紹介です。自動翻訳機能を使えば画面の上半分に原文、下半分に翻訳ページと表示でき、原文と翻訳文を簡単に比較できます。

<http://www.appbank.net/2014/02/15/iphone-application/750429.php>

■「猪八戒」→「Pig」: Google 翻訳でいろんなキャラクターの名前を英訳してみた

話題になっている Twitter 上のとあるツイートでは、他のキャラクターの表現がどのようになるのか試しています。

<http://news.livedoor.com/article/detail/8545272/>

AAMT Forum メールマガジン
16 号 (2014/03/07 配信)

こんにちは。AAMT Forum メールマガジン担当です。春とはいえまだ浅く、寒さの名残が感じられる日が続きます。かと思えば大変暖かくなる日もあり、寒暖入り混じるこ

の頃、気温の変化に体調を崩されませんように。では、メールマガジン第 16 号をお届けします。

■機械翻訳との上手な付き合い方 こんな時こそ、機械翻訳！

機械翻訳を利用する際の注意点や、機械翻訳を使ったことによる悪い事例などを解説しています。

<http://www.worldjumper.com/machine-translationtips/>

■富士通、リアルタイムのメッセージ翻訳や触感を再現するディスプレイなどスマートデバイスの未来像を提示

スペイン・バルセロナで開催された Mobile World Congress 2014 の富士通ブースでは、富士通の翻訳エンジン「ATLAS」をベースにして、英仏日の 3 カ国語によるリアルタイム同時変換のデモが行われました。

<http://wirelesswire.jp/special/201402/01/report/16.html>

AAMT Forum メールマガジン
17 号 (2014/03/20 配信)

こんにちは。AAMT Forum メールマガジン担当です。春一番も吹き、ようやく暖かくなってきたこの頃、いかがお過ごしでしょうか。本年度もそろそろ終わりが近づいてきましたが、新年度を迎える用意はよろしいでしょうか。では、メールマガジン第 17 号をお届けします。

■これって「ほんやくコンニャク」じゃん。
会話も翻訳できる『Translator Voice
Translate』で外国語を楽しもう

パッと起動してちゃちゃっと検索・即表
示じゃないと、いざというときに使えない。
そんな思いを具現化したような翻訳アプリ
『Translator Voice Translate』の紹介です。
[http://ch.nicovideo.jp/tabroid/blomaga/ar4
78978](http://ch.nicovideo.jp/tabroid/blomaga/ar478978)

■機械翻訳のノイズの中にも何かを見つける

機械翻訳で生成されたオモシロ日本語で
は、まれに、あまりに唐突すぎて原文がま
ったく想像できない場合もあります。その
ような翻訳への遭遇を新しい知識を得るチ
ャンスとし、その一例を紹介しています。
[http://cb-l10n.blogspot.jp/2014/03/blog-pos
t_6.html](http://cb-l10n.blogspot.jp/2014/03/blog-post_6.html)

■Android 版「Google 翻訳」が v3.0.5 にア
ップデート、手書き入力で 13 言を新たに
サポート

アラビア語、ボスニア語、セブ語、グジ
ャラート語、ミャオ語、カンナダ語、マル
タ語、モンゴル語、ペルシャ語、パンジャ
ブ語、ソマリ語、タミル語、テルグ語の手
書き入力がサポートされました。
<http://getnews.jp/archives/532519>

AAMT Forum メールマガジン

18 号 (2014/04/04 配信)

こんにちは。AAMT Forum メルマガ担

当です。温暖の候となり、桜も咲き始め、
過ごし易くなりました。春眠暁を覚えず、
とは言いますが、生活のリズムを崩さない
ように気をつけましょう。では、メールマ
ガジン第 18 号をお届けします。

■大手 IT 企業が殺到、音声翻訳アプリが
台頭

マイクロソフトは音声認識を含む自律的
インタラクションの開発に取り組み、グー
グルはモバイル音声翻訳技術の研究、フェ
イスブックはモバイル音声翻訳アプリの開
発企業を買収しました。

<http://www.xinhuaia.jp/social/29531>

■スマホが翻訳機になる Android アプリが
イカス！

36 カ国語の相互翻訳に対応した（うち
12 ケ国語は音声入力に対応）アプリ、『旅行
翻訳』の紹介です。

[http://weekly.ascii.jp/elem/000/000/206/20
6492/?rankaad](http://weekly.ascii.jp/elem/000/000/206/206492/?rankaad)

AAMT Forum メールマガジン

19 号 (2014/04/18 配信)

こんにちは。AAMT Forum メルマガ担
当です。ぽかぽか陽気でうたた寝の心地よ
いこの頃、いかがお過ごしでしょうか。仕
事中は眠気に負けないように気をつけま
しょう。では、メールマガジン第 19 号をお届
けします。

■翻訳システムで“おもてなし” 総務省、
東京五輪へ向け社会実験

2020年の東京五輪に向け、総務省は11日、日本語や英語などでの会話を即時に翻訳するシステムの社会実験を15年度にも始める計画を明らかにしました。

<http://www.sankeibiz.jp/macro/news/140411/mca1404110703022-n1.htm>

■翻訳費用・納期を大幅に削減・短縮！機械翻訳エンジン「Globalese」を4月8日より販売開始

4月8日より、株式会社川村インターナショナルは、ハンガリーの MorphoLogic Localisation Ltd.と国内販売代理店契約を締結し、同社の機械翻訳システム「Globalese(R) (グローバルーズ)」の販売を開始しました。

<http://japan.cnet.com/release/30068204/>

■Google 翻訳にサイトの日本語テキストを最適化させるという考え方とその方法

Google 翻訳を使って日本語を英語に翻訳する際に気をつけるべき日本語の記述方法を記しています。また、それに基づき、日本語の記述を書き換えた場合とそうでない場合の翻訳結果の比較を行っています。

<http://kybrain.com/it/google-translate-opimization/>

AAMT Forum メールマガジン

20号 (2014/05/09 配信)

こんにちは。AAMT Forum メールマガ担当です。目には青葉に初がつお、さわやかな季節がやってまいりました。さて、今年

のゴールデンウィークはいかがすごされたでしょうか。行楽に出かけた方も多いかと思います。リフレッシュしたら気持ちをしっかりと切り替えましょう。では、メールマガジン第20号をお届けします。

■Google 翻訳には秘密の特別版があるのか？

公に公開されている Google 翻訳と、Android や Chrome 用のソフトウェア流通サイトに組み込まれている Google 翻訳機能の翻訳精度の違いについて述べています。

<http://cpplover.blogspot.jp/2014/05/google.html>

■eBay 自動翻訳システムの開発に力を注ぐ！

eBay が BRIC (ブラジル、ロシア、インド、中国) といった新興国への更なる投資を急ぐ中で、「システム翻訳」のシステムの開発への取り組みについて述べています。

<http://ameblo.jp/saats/entry-11817394632.html>

言語処理学会第 20 回年次大会での発表報告

AAMT 機械翻訳課題調査委員会 WG1、WG2

はじめに

これまで AAMT 機械翻訳課題調査委員会（以下、課題調査委員会）は機械翻訳に関する調査活動を行ってきた。また、その調査結果を AAMT 総会や本誌を通じて報告してきた。この調査結果を自然言語処理の研究者に向けて発信するために、言語処理学会第 20 回年次大会（NLP2014、北海道大学、3 月 17-20 日）に於いて発表を行った。本稿は、言語処理学会での発表内容の一部を報告する。

調査の概要

課題調査委員会が行った機械翻訳に関するアンケート調査の目的は、機械翻訳システム活用の実態を把握することであった。この調査は、日本翻訳連盟（JTF）主催の第 23 回 JTF 翻訳祭（2013 年 11 月 27 日）の参加者を対象として行われた。ちなみに、JTF 翻訳祭とは、翻訳者、翻訳会社、翻訳発注企業などを対象として、講演会、展示会が行われる 1 年に 1 度のイベントである。

この翻訳祭では、過去に実施したアンケート[1]と同様に、実務翻訳においてどのように機械翻訳が活用されているかを調査した。調査方法や結果は以下の通りであった。

・調査実施日

第 23 回 JTF 翻訳祭（2013 年 11 月 27 日）

・調査方法

アンケート用紙（A4 表裏）配布による無記名の調査とした。具体的には翻訳祭の展

示コーナーの AAMT ブース前を通りかかる人に直接アンケート用紙を手渡し、調査協力者を募った。

・回答状況

回答者数は 64 名であり、翻訳祭参加者（約 800 名）全体に対する回答率は 8%程度であった。

・性別、年齢

年齢は 20 代から 70 代まで分布しているが、40 代が 29 名と突出して多かった。以下、30 代（17 名）、20 代及び 50 代（ともに 6 名）と続く。性別は女性 42 名に対して男性 22 名であり、女性の回答者の割合が高かった。

・回答者の職種

回答者 64 名のうち、35 名（55%）が翻訳者であり、21 名（33%）が翻訳業務の管理者であった。このうち 35 名の翻訳者の回答を抽出し、調査結果として報告した。

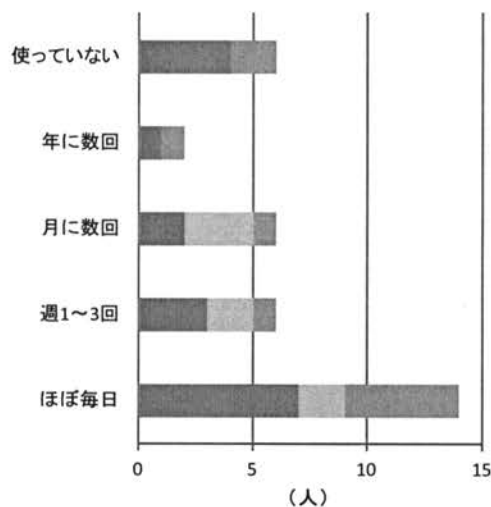
- (1) フリーランスの翻訳者：17 名
- (2) 翻訳会社の社内翻訳者：7 名
- (3) （翻訳会社以外の）企業や官公庁の翻訳者：11 名
- (4) 翻訳会社の翻訳業務管理者、コーディネーター：15 名
- (5) （翻訳会社以外の）企業や官公庁の翻訳業務管理者、コーディネーター：6 名
- (6) その他・無回答：8 名

・調査結果その1：利用頻度

まず、機械翻訳の利用頻度に関する調査結果を報告した。

回答者の4割以上（14名）が機械翻訳を「ほぼ毎日」使っていると回答している。「使っていない」と答えた人は17%（6名）のみで、機械翻訳は翻訳者に幅広く浸透しているといえそうである。

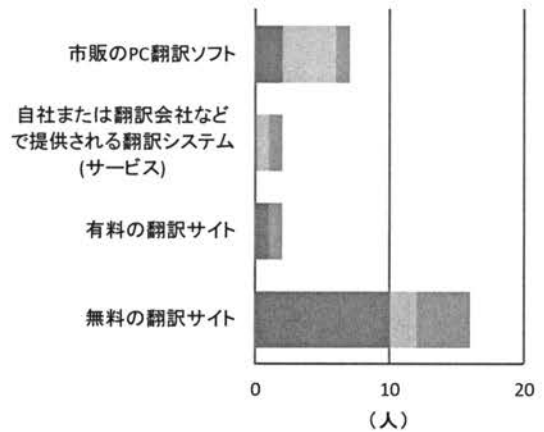
■ フリーランス
■ 翻訳会社
■ 企業・官公庁



・調査結果その2：利用する種類

次に、翻訳者が利用する機械翻訳の種類を報告した。

翻訳者がよく使う機械翻訳のトップは無料の翻訳サイトであった。翻訳会社などから翻訳ソフトが提供される例は少数であり、翻訳者が個人的に機械翻訳を利用するケースが大半を占めるようである。翻訳者タイプ別に見ると、フリーランスの翻訳者では圧倒的に無料の翻訳サイトが利用されている一方、翻訳会社内の翻訳者では市販のPC翻訳ソフトを利用する割合が高くなっている。

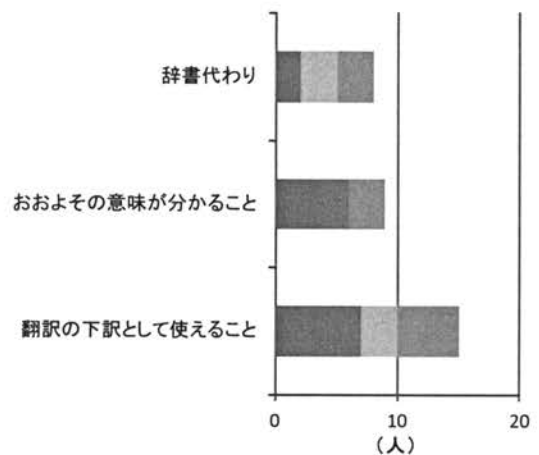


無料の翻訳サイトがトップであることから、機械翻訳にとって手軽に使えることが重要な条件といえそうである。このことは、現在は利用が少ない市販のPC翻訳ソフトが、その昔パソコンの付属ソフトとして流布していた当時は利用が多かったことからもうかがえる。

・調査結果その3：期待する品質

そして、機械翻訳に対して期待する品質を報告した。

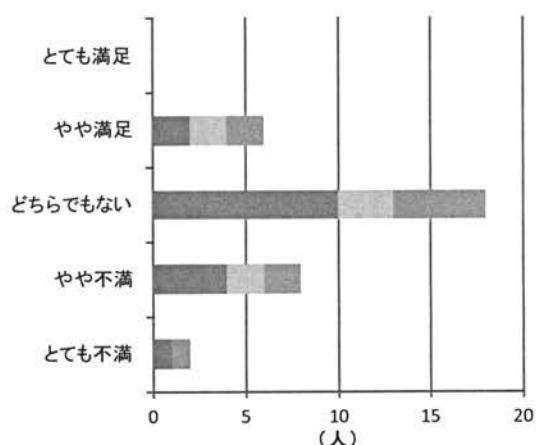
機械翻訳を翻訳の下訳として使いたいと考える翻訳者が多いようだが、その一方で、機械翻訳は辞書引きレベルでの使用と割り切っている翻訳者もかなりの割合でいることがわかる。



・調査結果その4：満足度

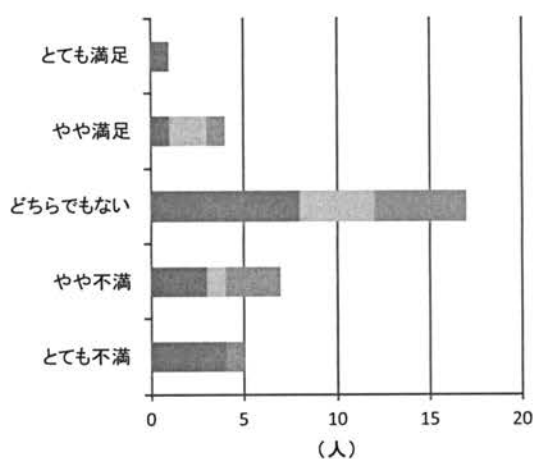
また、実際に機械翻訳を利用したことがある翻訳者の満足度を報告した。

機械翻訳の満足度を5段階で評価した結果、「とても満足」と回答した翻訳者は皆無だったものの、18%（6名）が「やや満足」と答えており、現在の翻訳システムでも使い方次第で役に立つ場面があることがうかがえる。



・調査結果その4a：翻訳品質の満足度

満足度の根拠を具体的に知るために、「翻訳品質」と「翻訳機能」に分けて満足度を報告した。



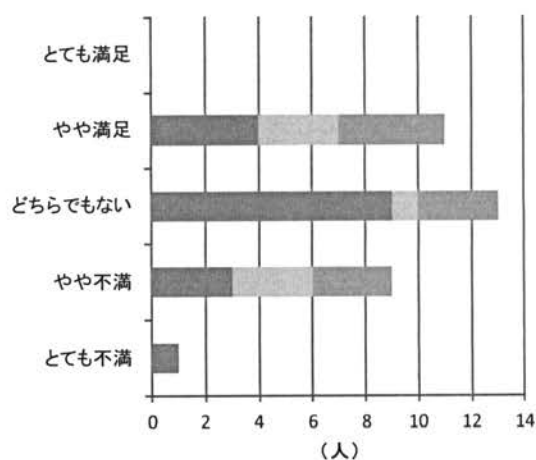
翻訳品質の満足度

翻訳品質に満足していると答えた人（5名）の割合は不満であると答えた人（12名）

の半分以上であり、翻訳者を満足させるためには、さらなる翻訳品質の向上が不可欠といえる。

・調査結果その4b：翻訳機能の満足度

機械翻訳について、機能としては「満足」と答えた人（11名）と「不満」と答えた人（10名）の割合が拮抗していることがわかる。翻訳ソフト／サイトの使い勝手については、品質ほど不満に感じている翻訳者はいないようである。



翻訳機能の満足度

・自由記述

今回のアンケートの中に、「翻訳業務の効率化のために工夫していることはありますか？」という自由記述形式の項目を設けたが、回答があったのは27名（42%）にとどまった。立ったままで回答をお願いしたこともあり、回答も1行以内の短いものが大半だった。

他方、AAMTでは2012年にWebによるアンケートを実施している。このアンケートに含まれていた「あなたが実践している、あるいはやってみたい翻訳ソフトの効果的な使用方法を教えてください」という設問に対しては、124名から平均100文字程度の回答を収集することができた。

そこで、今回の翻訳祭でのアンケート結果と2012年のWeb調査による自由記述欄の双方をもとに、翻訳者が普段の業務で実践している機械翻訳の具体的な活用方法を調査し、その結果を報告した。以下、ユーザコメント（自由記述）の引用は、2012年の調査で収集したものからの抜粋である。

2013年アンケート調査で翻訳業務効率化のための工夫について回答のあった27名中11名が、自分用もしくは部門用の用語集を整備していると回答していた。翻訳者の多くが、機械翻訳活用において辞書整備を実践していることが裏付けられた。

共有辞書については、次のような要望を紹介した。

- ・分野別の用語集が共有できれば助かる。

また、その一方で次のような興味深い指摘も紹介した。

- ・原文で専門用語の統一がされていない場合には、用語集（共有辞書）は利用価値が少ない。
- ・同業でも企業に依って訳語が異なるので、共有辞書は危険。

2012年度調査の自由記述欄では、折り返し翻訳についての記述が目についた。折り返し翻訳とは機械翻訳の英文を修正し、その英文をもう一度機械翻訳で日本語にし、原文の内容とあっているかチェックするという使い方である。折り返し翻訳は、チャット翻訳などを使うときに一般利用者が使うものと予想したが、一部の翻訳者にも積極的に活用されているようである。

「利用する種類」において、翻訳者に最も多く使われているのは無料の翻訳サイト

という結果となった。具体的な使い方としては原文の概要把握に使われるケースが多くあり、その一部を紹介した。

- ・専門分野以外の英文を読んでよくわからないときに翻訳ソフトにかける。

少し特殊な使い方としては、機械翻訳が得意とする言語対に限定して使うケースやユニークな使い方を紹介した。

- ・ソース言語がゲルマン語系のとき、一旦Google翻訳で英語に自動翻訳させてから、訳文（和文）と照合して校正を行う。
- ・韓国語の場合は日本語へ直接翻訳させている。
- ・複数の翻訳サイトを横断的に利用して、最も適切と思われる翻訳結果を採用している。

有償ソフト／サイトについては、「数万円は（高くて）払えない」という指摘があり、無料サイトとの対比で有償ソフト／サイトのコスト対効果が年々下がっていることが、結果として無料サイトの利用を増やしている可能性を示唆した。

Webアンケートの自由記述に含まれる翻訳者からの機械翻訳への要望のなかで、代表的な記述も紹介した。

- ・入力した文に関してどのような統語解析が行われ、どの語がどう訳されたか、という処理過程も訳文といっしょに出力してほしい。

これは完全な翻訳が出せない機械翻訳を実務翻訳に普及させるために憂慮すべき課題として位置づけた。また、実務翻訳ならで

はの要望も紹介した。

- ・構文上可能なすべての訳を提示してほしい。
- ・くだけた表現、フォーマルな表現を訳し分けたい。
- ・原稿の日本語を機械翻訳に適した日本語に翻訳する『日日翻訳』を翻訳ソフトで使いたい。

さいごに

言語処理学会での発表を通して、機械翻訳に関心のある研究者や学生諸氏と機械翻訳について様々な議論を行うことができた。また、課題調査委員会の調査活動の今後の課題や AAMT として取り組むべき課題も認識できた。さらに、この発表を通じて AAMT がどのような活動を行っている団体かということも認知してもらえた。これらのことから、一定の成果は得られたといえる。

参考文献

AAMT 課題調査委員会：機械翻訳・翻訳メモリに関するアンケート結果報告, 『AAMT ジャーナル, No.51, pp.76-80, 2012.

UTX Converter (UTX 変換ツール) 仕様書

AAMT 機械翻訳課題調査委員会 WG3

UTX チーム (AAMT) 2014/05/06 版

UTX Converter (UTX 変換ツール) は、主に以下の機能を持つ。

- UTX 形式ファイルが UTX 仕様に沿っていることを検証する。
- UTX 形式とその他のファイル形式との相互変換をする。

ツール開発の基本方針

- 開発コストを抑えつつ、当初は単純だが役立つ機能に厳選して実装する。その後、優先度が高い機能を順次実装していく。対応するファイルの数など、将来的に拡張が予想される場合は、その点を配慮して設計する。
- 言語タグ、文字コードなどの知識が少ないユーザーでも使いやすいように配慮する。
- UI 文字列では専門用語を避ける。
- 操作をシンプルにする。
- オープン ソースの開発者の協力が得やすい配慮をする。ソースコードを公開する。

使用シナリオの例

- Excel 上で UTX 用語集を手作業で作成した場合に、UTX 仕様に沿っているかチェックする。
- UTX を用語集管理のハブとして、各形式への用語追加をする (1 週間に 1 度など)。翻訳ソフト A、用語ツール B などにそれぞれ独自形式で用語追加をしてしまうと、集中管理できず手間が増える。UTX を中心にして、UTX 形式で新規用語をまとめるようにし、その後、UTX Converter を使って、各形式にワンステップで変換できる。

基本要件

- 作成したコードに関する著作権は、AAMT (アジア太平洋機械翻訳協会) が保持する。
- コードは、著作権を明示したうえで改変/商用利用を許諾し、一般に公開される (クリエイティブ コモンズ「表示」に類似のライセンスを採用する予定)。
- 他の開発者がコードを改変・追加しやすいように、英語で動作の説明を簡単に記述する。変数名も省略しすぎず、分かりやすいものとする。
- ツールの対応 OS は、Windows 7 および 8.1 (32 ビット版、64 ビット版) とする。
- ツールの動作をテストする OS は、Windows 7 (64 ビット版) とする (Windows 8.1 の動作のテストは AAMT で行う)。
- 検証・変換用の部分は、C++ で DLL として作成する。マルチ プラットフォームにするため、標準関数と STL のみ使用することとする。
- UI 部分は、Visual Basic 2008 以降、Visual C++ 2008 以降、Visual C# 2008 以降などで作成する。

バージョン 0.1 機能要件

1. バージョン 0.1 (テスト用、非公開) の段階では、試験的な機能のみとする。
2. インターフェイス (エラー メッセージなど) とコードの説明は英語のみとする (将来的には日本語化する)。
3. 入力ファイルは、*.utx ファイル (タブ区切りテキスト) とする。

4. 入力ファイルは、複数フォルダーに存在するファイルを複数指定可能とする(最低 100 個まで処理できることを確認する)。
5. メイン ウィンドウには、以下の UI 要素を含める。
 - A) ドラッグ アンド ドロップ 領域 (ラベル [(1) Drag and drop file(s) here to validate or convert.])
 - B) 入力ファイルのリスト ボックス (ラベル [(2) File(s) to validate or convert:])。表形式で、2 列目にファイル名、3 列目にパス名を示す。1 列目には処理後の処理結果 (成功 = OK、失敗 = Failed) を示す。
 - C) [Clear all items] ボタン。入力ファイルのリスト ボックスの下に配置し、入力ファイルのリストを全消去する。
 - D) 処理元ファイル形式リスト ボックスと変換先ファイル形式リスト ボックス (ラベル [(3) Process Type:])。
 - i. 処理元ファイル形式リスト ボックス
入力ファイルがドロップあるいは指定されたら、拡張子から判定するが、csv などは複数の可能性があるので、正確な処理元形式は変更できるようにする。
 - ii. 変換先ファイル形式リスト ボックス
以下の項目を持つリストボックスを作成する。
 1. [(Validation only)] リストボックス項目 (以下の検証機能 1・2 に該当)
 2. [Forbidden term list file] リスト ボックス項目 (以下の変換機能 1 に該当)
 - [MultiTerm text file] リストボックス項目 (以下の変換機能 2 に該当)
 - E) [Execute] ボタン (上記処理のいずれかを実行する)。
 - F) [File] メニュー
 - i. [Add] コマンド: 「ファイルを開く」ダイアログを開く。標準的な [Open] ダイアログ ボックス。ファイル名指定欄には、変換可能なファイル形式の拡張子を含める。選択されたファイルを入力ファイルのリストボックスに追加する。
 - ii. [Exit] コマンド: プログラムを終了する (確認は不要)。
 - G) [Help] メニュー
 - i. [Help] コマンド: マニュアル URL を開く。現状では AAMT サイトの UTX ページを表示する。
 - ii. [About] コマンド: バージョン、著作権情報、UTX ページの URL を表示する。
6. 変換中は以下のウィンドウを表示する。
 - (ア) [Processing...] テキスト ラベルと [Cancel] ボタンを含むウィンドウ。
7. 検証機能 1: ヘッダーと列定義の構造をチェックして、仕様どおりかを確認する。
8. 検証機能 2: 品詞と用語ステータス (ある場合のみ) が仕様どおりかを確認する (properNoun のような、UTX 仕様書で規定されている要素名の綴りが正しいか、半角スペースが規定どおりか、など)。
9. 検証後の処理: 検証後の結果は、入力ファイルのリストボックスの 1 列目に表示する。正常に処理されたファイルは入力ファイルリストの 1 列目に「OK」と表示し、異常があるファイルは「Failed」と表示する。また、すべてのファイルが正常に処理された場合はメッセージボックスで「No error is found.」と表示する。異常がある場合は、メッセージボックスに「<N> error(s) were found.<改行><改行>Do you wish to save the detailed error information in a log file?」 ([Yes/はい] /

[No/いいえ] ボタン) を提示する。[はい] の場合は、ファイルの保存ダイアログ ボックスを表示する。元ファイルのフォルダーを既定で表示する。複数のファイルを指定しており、フォルダーが複数ある場合は最後に開いたフォルダーを表示する。

10. ログファイルの各行は「ファイルのフルパス (行番号): エラー メッセージ」とする。
11. 変換機能 (共通仕様): #で始まるコメント行を無視する。
12. 変換機能 1: 禁止語リストの抽出と出力を行う。禁止語リストとは、用語ステータスが「禁止」となっている語、つまり「禁止語」のみを抽出したリストである。禁止語リストは、用語チェック ツールで、禁止語のチェックに使用できる。出力されるリストは訳語と品詞のタブ区切りテキストとする。品詞は UTX の品詞コードのままとする。

変換結果は以下のファイル名で出力する。

<元のファイルパス>¥<元のファイル名>_forbidden_terms.txt

13. 変換機能 2: SDL Trados の用語管理ツール MultiTerm 用に単純な変換を行う。具体的には、UTX ヘッダーを削除し、1 列目に列の定義を記述する。列の定義は UTX と同様だが、HTML 言語タグ jaなどを、言語タグの知識のないユーザーにも分かりやすい Japanese のような語に置換する。置換する主要な十数カ国語を別途指定する。

- HTML 言語タグ:

<<http://www.w3.org/International/articles/language-tags/Overview.en.php>>

- 主要言語タグ:

<<http://www.microsoft.com/resources/msdn/goglobal/default.mspx>> の Windows7

変換結果は以下のファイル名で出力する。

<元のファイルパス>¥<元のファイル名>_MultiTerm.txt

14. 変換後の処理: 変換時には検証も同時に行う。したがって、変換後には「検証後の処理」と同じ処理を行う。

作業シナリオの例

1. 入力ファイルをダイアログ ボックスまたはドラッグ アンド ドロップ ボックスで指定する (複数指定可能)。
2. 入力ファイルが入力ファイルリストに表示される。
3. 変換処理を実行する前や後で、[Clear all items] ボタンを押すと、リストが消去される。処理元のファイル形式と変換先のファイル形式を指定し、[Execute] ボタンを押す。

今後の予定

今後は実用性のある基本的な機能を実装する。以下のような機能を検討している。

1. ファイルのドロップ領域を画面全体に変更。
2. UTX 1.20 への対応。
3. 英語に加え、日本語インターフェイスを作成。
4. ファイル形式の変換時に失われる情報をログに出力。
5. 検証機能の強化。
 - (ア) 異常個所の修正方法の提示。
 - (イ) 重複項目の指摘。
6. 他の用語集形式との変換。
 - TBX 形式、MultiTerm テキスト形式、UPF 形式、MultiTerm XML 形式等。
7. 各種翻訳ソフトのユーザー辞書形式との変換。
 - ATLAS (富士通)、The 翻訳 (東芝)、PC-Transer (クロスランゲージ)、翻訳ピカイチ中国語 (クロスランゲージ)、j・北京 (高電社)、LogoVista (ロゴヴィスタ) 等。

8. その他の形式への変換。
 - (ア) 非標準語置換リスト。非標準語を承認語で置換するためのリスト。
 - (イ) 禁止語置換リスト。禁止語を承認語で置換するためのリスト。
9. TMX 形式への変換。
品詞が sentence のものを抜き出し、TMX 形式で保存する。
10. 複数の UTX 形式のマージ機能。
11. マニュアルの作成。
12. インストーラの作成。

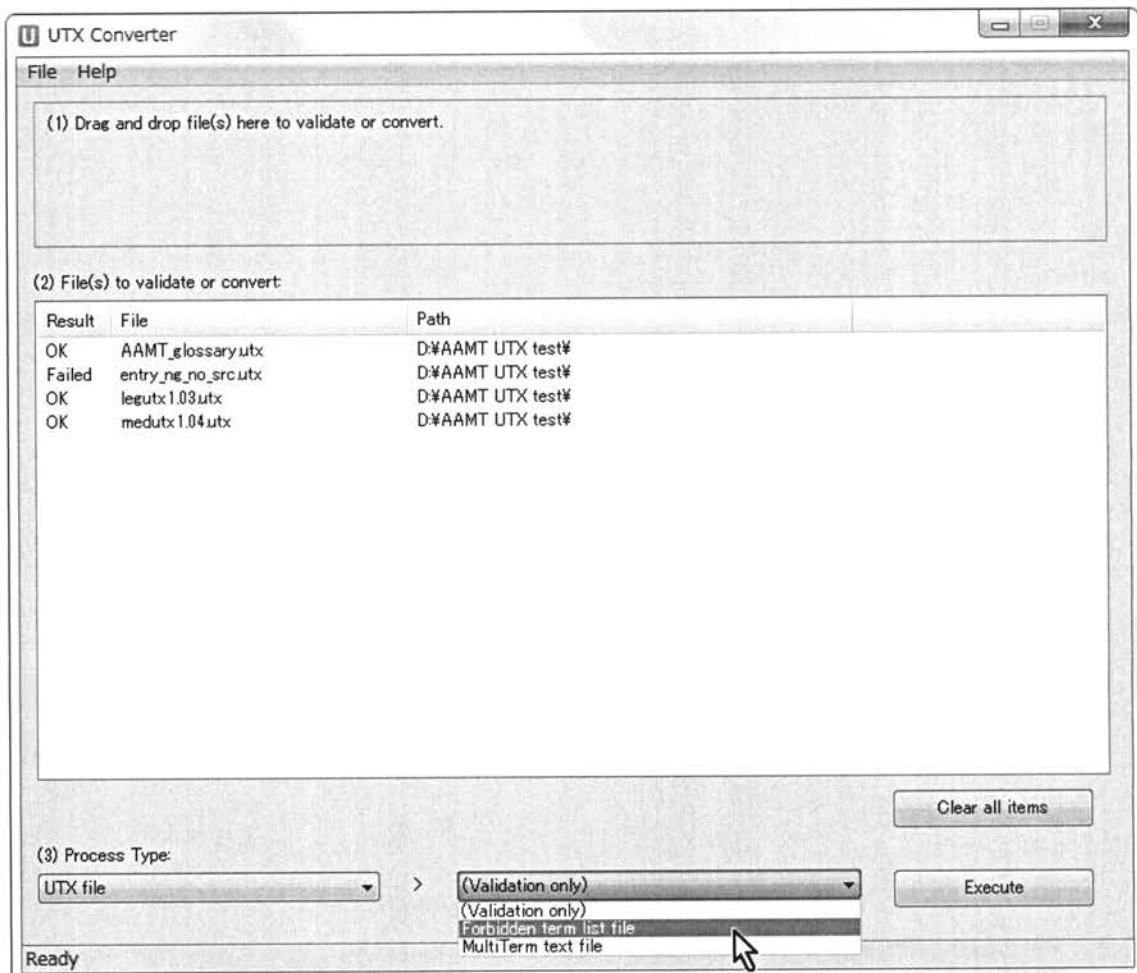


図 1. UTX Converter (UTX 変換ツール) の画面例

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 のスタートから第十三回を迎えることができました。今号では、個人会員一名さま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など、今回も貴重なご意見をお寄せいただきました。

AAMT Journal では今後も引き続き、会員の皆さまからのご寄稿を心よりお待ちしております。

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

個人会員（敬称略・50 音順）

会員名

春遍 雀來（ハルペン・ジャック）

Jack Halpern

日中韓辞典研究所取締役社長

漢英字典刊行会編集長

自己紹介

ドイツ生まれ。6 カ国に移り住み、15 カ国語を習得。イスラエルのキブツで漢字に魅せられ、1973 年 5 月家族と共に来日。16 年を費やして漢字を体系的に学ぶ方法を考案し、1990 年に外国人向け漢字学習書『新漢英字典』（研究社）、1999 年に『漢英学習字典』（講談社）を編纂した他、多数の漢字学習辞典を作成してきました。

その後は日中韓各語の自然言語処理用大規模辞書編纂の傍ら、著作業、ラジオやテレビ出演、新聞・雑誌の取材や講演会を数多く手掛けてきました。

現在は日中韓辞典研究所代表取締役として各種辞書データベースの構築を続ける他、10 か国語による『漢字学習辞典シリーズ』の編纂に取り組んでいます。

また、国際一輪車連盟の創立者、日本一輪車協会の常務理事として、日本と世界各国で一輪車スポーツの普及を進めています。

MT/翻訳とのかかわり

MT および翻訳業界に期待すること

日本では機械翻訳の分野に於いて、長年ルールベース型の機械翻訳（RBMT）を中心とする傾向がありました。ルールベース型翻訳は文法や辞書データを手掛かりとする方式ですが、開発に膨大な時間と労力を要するため、現在は壁にぶつかっていると言えるでしょう。

これに対して、海外や一部の日本企業で注目を集めているのが統計型の機械翻訳（SMT）です。これは辞書や文法ルールを出発点とせず、統計と数学的アルゴリズムに基いて翻訳を行うもので、ルールベース型翻訳の土台となっている様々な「言語知識」を必要としない反面、膨大な量の対訳データ（コーパス）の収集を必要とし、翻訳精度の向上にも限界があります。

これらの状況を総合的に鑑み、現在注目を集めているのが両者の混合型機械翻訳（Hybrid）です。有名な Dr. Alan Kay は「機械翻訳には辞書情報・文法情報等の言語知識が不可欠である」と述べていますが、ハイブリッド型機械翻訳はルールベース型の基礎となる辞書や文法ルールと統計型で利用されるコーパスを複合的に活用する方式で、大幅な翻訳精度向上が期待できます。

日中韓辞典研究所（CJKI）では、日中韓各語のあらゆる側面を網羅した包括的辞書データベースの開発・拡張を行ってきました。先進的な計算辞書学の手法を用いて構築・維持が行われた CJKI のデータベースは、下記のような場面で活用されます。

1. 固有表現抽出 (NER) 及び処理
2. 機械翻訳、オンライン翻訳ツール及び音声処理技術
3. 情報検索技術
4. 書籍版・電子版を含む各種辞書
5. 日中韓入力システム (IME) や形態素解析

長年に亘って蓄積された約 2400 万項目に上る言語データベースには、一般語彙の意味属性や品詞等の詳細な文法情報に加え、専門用語及び固有名詞が収録されています。とりわけ 1000 万項目に上る日中韓固有名詞データは、貴重な辞書データとしてハイブリッド型機械翻訳は勿論、NER 等を支える有効な情報となり得るものです。

私たちは世界有数の日中韓並びにアラビア語辞書データの提供者という役割を担い、世界を代表する数々の企業やソフトウェア開発者に高品質の語彙資源を提供してきました。これからも機械翻訳を始めとする自然言語処理の分野に大いに貢献していきたいと願っています。

協会活動報告

(2014 年 4 月～2014 年 5 月)

機械翻訳課題調査委員会

2014 年 4 月 18 日 (2014 年度 第 1 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)
 - ・ UTX 変換ツールのライセンスについて
 - ・ UTX の新しいウェブサイトの作成に関する検討
 - ・ ISO Berlin の参加の可否についての検討
 - ・ UTX 仕様検討
- ③ 全体会議
 - ・ 2014 年度 AAMT 総会・講演会・展示会について
 - ・ AAMT 長尾賞学生奨励賞の新設と応募よびかけについて
 - ・ AAMT 総会・報告会における展示について
 - ・ AAMT 新ウェブサイトに対するコメント交換など

編集委員会

- ・ 総会号発行に向けての編集作業

インターネット WG

- ① 新 AAMT ホームページ公開に向けた準備
 - ・ 新ホームページの更新・チェック体制の確立 (完了)
 - ・ 専門作業によるコンテンツ作成体制の確立 (完了)
 - ・ 旧ホームページコンテンツの移行 (作業中)
 - ・ 旧ホームページ以降の差分コンテンツの作成 (作業中)
 - ・ 総会時期の新ホームページ公開に向けた手順検討 (検討中)
- ② 既存 AAMT ホームページの更新
- ③ AAMT Forum メールマガジンのコンテンツ作成・配信業務の、課題調査委員会への移管作業 (完了)
- ④ 他委員会の AAMT ホームページ利用に関するサポート
- ⑤ 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

2014 年 5 月 9 日(金) (2014 年度 第 1 回)

1. 新年度のご挨拶・新メンバーご紹介 (委員長)
2. 報告書寄稿内容説明 (各著者)
3. 拡大評価部会関連 (江原先生)
4. シンポジウム関連
5. 次回の開催について
 - 開催の日時
 - 主な議題

AAMT/Japio 特許翻訳研究会 拡大評価部会

2014 年 5 月 9 日(金) (2014 年度 第 1 回)

1. 今年度の活動計画
 - 1.-1 自動評価グループ
 - 1.-2 人手評価グループ
 - 1.-3 テストセットグループ
2. 基本文テストセットの現状
3. 中日ツリーバンクとその応用

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

今号の巻頭言は、情報通信研究機構 隅田 英一郎様より、御寄稿を頂きました。

先日、言語処理学会の第20回年次大会が北海道大学にて開催され、理事の長尾真先生が「NLPの過去、現在、将来」と題して特別講演をされましたが、今号におきましては、この御講演の内容に基づきまして、貴重な特別寄稿を頂きました。

また、研究関係ならびに言語資源に関する御寄稿として、情報通信研究機構内山 将夫様より、同機構における機械翻訳技術の研究動向に関する御寄稿を、また、科学技術振興機構 中澤敏明様より、アジア学術論文抜粋コーパスに関する御寄稿を、それぞれ頂きました。

理事の坂本義行先生からは、機械翻訳の歴史に関する貴重な御寄稿の第三弾を頂きました。

さらに、シンポジウム報告として、関西外国語大学 内田真弓先生、一般財団法人テクニカルコミュニケーター協会 山崎 敏正様、富士通研究所 岩倉友哉様より、「通訳翻訳テクノロジー研究プロジェクト」第一回会合、第5回産業日本語研究会・シンポジウムにおける最新のテクニカルコミュニケーション動向報告、言語処理学会第20回年次大会ワークショップ「自然言語処理の発展に向けた情報共有・討論」について、それぞれ御報告頂きました。

新たなサービス・製品の紹介として、日本発明資料株式会社 本間奨様より、特許翻訳用システム MT Plus II / NEF スプレッドの紹介を御寄稿頂きました。

AAMT 内の活動報告として、AAMT/Japio 特許翻訳研究会からは、拡大評価部会の活動について御寄稿頂きました。機械翻訳課題調査委員会からは、これまで行ってきた機械翻訳に関する調査活動の報告を言語処理学会第20回年次大会にて行った様子、用語集形式 UTX 変換ツール 仕様書、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、個人会員の紹介文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

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。

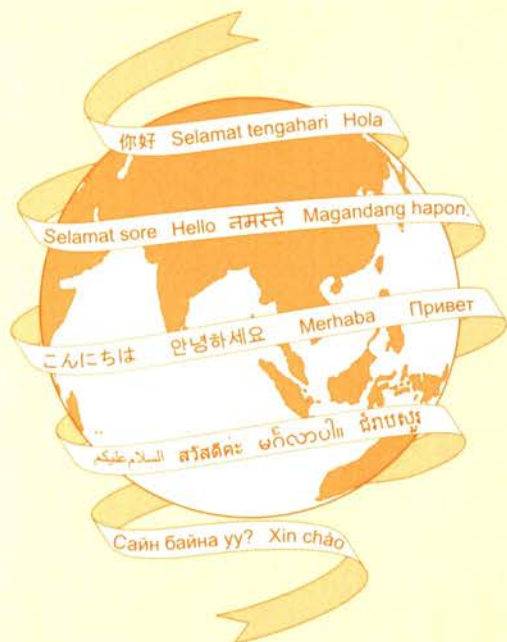
申込日 20 年 月 日

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.56

発行：アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

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

phone：03-5951-3961 fax：03-5951-3966

編集委員会：宇津呂 武仁 小谷 克則 大倉 清司

鈴木 博和 阿部 さつき

表紙(図部分)デザイン：阿部 さつき

事務局：神崎 享子 荻野 孝野

印刷所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)

c/o Japan System Application Co., Ltd.

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