ISSN 1883-1818

AAMT Journal

Asia-Pacific Association for Machine Translation

February 2015 No.58

アジア太平洋機械翻訳協会

目　次

CONTENT

巻頭言

歴史は繰り返す：ニューラルネットと言語学

永田　昌明

NTT コミュニケーション科学基礎研究所

nagata.masaaki@lab.ntt.co.jp

ニューラルネットの再興

長生きはするものである。歴史は繰り返すとよく聞くが、本当に繰り返すところを見ることができた。何かというとニューラルネットのことである。ここ数年、私は「統計翻訳の研究が始まって 20 年、ついに英日翻訳において統計翻訳の精度がルールベース翻訳を上回った」と呪文のように繰り返し述べていたが、今や少なくともアカデミックな世界では、ニューラル翻訳(Neural Machine Translation, NMT)の精度が統計翻訳(Statistical Machine Translation, SMT)を上回りそうな勢いである。

私がまだ駆け出しの研究者だった 1990 年頃、音声認識の分野は、ニューラルネット、特にリカレントニューラルネットワーク(Recurrent Neural Network, RNN)で盛り上がっていた。しかしブームはやがて下火になり、音声認識は隠れマルコフモデル、機械学習はサポートベクトルマシンなどのシンプルな学習器が主流となり、ニューラルネットは過去の遺物になったかと思われた。

しかし 2010 年代に入り、深層学習(Deep Learning)によって音声認識や画像認識の誤り率を大幅に削減できることが分かると、2013 年頃から自然言語処理へのニューラルネットの適用が盛んになった。2014 年 12 月に発表された論文では、リカレントニューラルネットの改良版である LSTM (Long Short-Term Memory)を用いた英仏翻訳の精度は、最先端の統計翻訳とほぼ同じと報告されている。

あきらめずにニューラルネットの改良を続けた研究者達に心から敬意を表するとともに、統計翻訳の誕生から 20 年ぶりの技術革新かもしれないニューラル翻訳の今後の発展が非常に楽しみである。

自然言語処理における言語学の復権

ニューラルネットが流行っていた 1990 年頃、自然言語処理の分野では、GPSG, HPSG, LFG といった文法枠組の提案が花盛りであった。今から思えば当時は自然言語処理と言語学の蜜月期だった。その中で私はひたすら主辞駆動句構造文法(Head-driven Phrase Structure Grammar, HPSG)に基づく日本語の文法規則を書いていた。約 2 年間、人手で文法規則を書き続けて心底から嫌気がさし、言語データから規則に相当する統計モデルを自動的に学習する統計的自然言語処理の研究に着手し、現在の統計翻訳に至っている。そんな私が最近思うことは、自然言語処理における言語学の重要性である。

日本語と英語のように語順が大きく異なる言語対の統計翻訳では、翻訳元言語の文の単語を翻訳先言語の語順に並べ替えてから翻訳を実行する事前並べ替え(preordering)と呼ばれる方式が主流になっている。この事前並べ替えの正解を定義しようとすると、主語の省略(pro-drop)や主辞後置(head-final)といった言語類型論から見た日本語の特徴や、IP(時制句),CP(補文),VP(動詞句)といった生成文法(普遍文法)による分析が理論的な拠り所になる。

ところが、最近の若い自然言語処理研究者は、機械学習には詳しいが言語学を知らないので、こういう話が通じない。私は統計的自然言語処理と呼ばれる技術を、その初期からずっと支持してきたが、最近はちょっと行き過ぎではないかと感じている。

言語学の入門書を読むのが最近のマイブームなのだが、20 年経つと言語学も地道に進歩していて新鮮である。言語学の知見を活かした機械翻訳に、ニューラルネットとは別の技術革新の芽を見出したいというのが最近の私の目標である。

研究報告

統計的機械翻訳とは?

渡辺　太郎
情報通信研究機構

webでの多言語翻訳サービスや、特許など分野を限定し、ビジネスに特化した機械翻訳システムなど、最近の機械翻訳システムの性能向上は著しい。機械翻訳システムの基本的な考え方は、1947年3月4日、Warren WeaverがNobert Wienerに送った手紙に集約されている[1]:

> ロシア語の文章を見て、思ったのは「本当は英語で書かれているが、変なシンボルで暗号化されている。では今から解読しよう。」

機械翻訳の問題を暗号解読の問題として捉えるという考え方は、初期の計算機の開発において最初のアプリケーションの一つとして試みられたが失敗に終わっている。この失敗を教訓に、構文解析などの言語処理の基礎技術への研究開発を呼びかけるALPACレポート[2]につながっており、近年の自然言語処理の研究開発の礎となっている。

機械翻訳は自然言語処理の要素技術を組み合わせ、言語学者による知識を記述することにより実現されてきたが、翻訳に必要とされる知識を人手で網羅的に表現することは不可能であり、次第に限界に達してきた。この結果、1990年代のIBM Watson研究所による研究成果の成功により、再び暗号解読に基づく機械翻訳技術が注目をあびることになった[3]。特に、計算機の発達および自然言語処理の基礎技術の進化により複雑な統計モデルや大量のデータ等が容易に扱えるようになった事が大きい。本稿では、機械翻訳を統計モデルで実現した統計的機械翻訳について、その鍵となる考え方について解説する。また、詳細な内容は文献[1]を参照していただけると幸いである。

翻訳をモデル化

統計的機械翻訳では、原言語の入力文$\boldsymbol{f} = f_1, f_2, ...$が与えられた時、対応する目的言語の文$\boldsymbol{e} = e_1, e_2, ...$へと翻訳される確率$Pr(\boldsymbol{e}|\boldsymbol{f})$を最大化する翻訳$\hat{e}$を求める[3]。

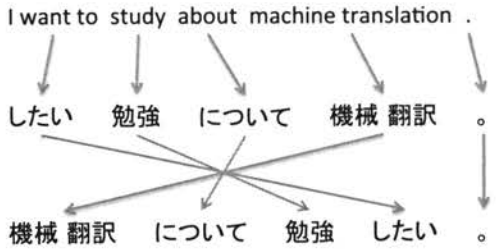
$$\begin{aligned}\hat{\boldsymbol{e}} &= \arg\max_{\boldsymbol{e}} \Pr(\boldsymbol{e}|\boldsymbol{f}) \\ &= \arg\max_{\boldsymbol{e}} Pr(\boldsymbol{f}|\boldsymbol{e})Pr(\boldsymbol{e})\end{aligned}$$

ベイズの定理により、二つの項$Pr(\boldsymbol{f}|\boldsymbol{e})$および$Pr(\boldsymbol{e})$の積を最大化する問題と置き換えられた。これは、まさにW. Weaverの手紙を表現した式であり、「本当は英語($Pr(\boldsymbol{e})$)で書かれているが、変なシンボルで暗号化されている($Pr(\boldsymbol{f}|\boldsymbol{e})$)。では今から解読しよう(arg max)。」と解釈可能である。前者の項は翻訳モデルと呼ばれ二つの言語の文が対訳として正しければ正しいほど高い確率を割り当てる。後者の項は、言語モデルと呼ばれ、生成された翻訳が目的言語の文として正しければ高い確率を割り当てる。正しさは必ずしも構文的な正しさを意味しておらず、例えば、特許や新聞、SNSでは文体が全く異なっており、適切な言語モデルを選択することによって各分野に適した翻訳を生成可能としている。逆に従来の機械翻訳では翻訳モデルのみに着目しており、分野に適した文を生成することは困難であった。

翻訳モデルは原言語および目的言語の文のペアに対して確率を割り当てるが、全ての文を列挙して記憶することはほぼ不可能である。そこで、翻訳が生成される過程を導出と呼ばれる複数のステップ$\boldsymbol{d} = d_1, d_2, ...$に分ける。

$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{e}} \sum_{\boldsymbol{d}} Pr(\boldsymbol{f}, \boldsymbol{d}|\boldsymbol{e})Pr(\boldsymbol{e})$$

例えば、知識に基づく機械翻訳では、まず構文解析により入力文の構文木を生成、その後変換ルールにより目的言語の構文木へと変換し、最後に目的言語の文を生成する、といったステップへと分割される。本稿では、具体例として非常に簡単な翻訳モデル、句に基づく機械翻訳[5]を導入する。このモデルでは、目的言語をK個で構成される句$\bar{\boldsymbol{e}} = \bar{e}_1, \bar{e}_2, \ldots, \bar{e}_K$へと分割し、句単位に原言語へと変換$\bar{\boldsymbol{f}} = \bar{f}_1, \bar{f}_2, \ldots, \bar{f}_K$した後で句単位に並び替える、といったステップで構成される。「解読」という考え方なので、入力と出力が入れ替わっている点に注意してほしい。例えば、"I want to study about machine translation."が「機械翻訳について勉強したい。」へと変換される過程は、



のように表され、例えば"machine translation"という句が「機械翻訳」へと変換され、最後から２番目の位置から文頭へと並び替えられる。$\boldsymbol{\alpha} = \alpha_1, \alpha_2, \ldots, \alpha_K$が原言語の並び替えを表現したインデックスとすると、この例の場合は、$\boldsymbol{\alpha} = 4,3,2,1,5$となる。この時$\bar{f}_{\alpha_1}, \bar{f}_{\alpha_2}, \ldots, \bar{f}_{\alpha_K}$を結合すると、$f$と同じになる。句に基づく機械翻訳の導出$\boldsymbol{d}$は、$\langle \boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}} \rangle$といった三組で表され、

$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{e}} \sum_{\boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}}} Pr(\boldsymbol{f}, \boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}} \,|\boldsymbol{e}) Pr(\boldsymbol{e})$$

が得られる。翻訳が生成される過程が$\boldsymbol{\alpha}$や$\bar{\boldsymbol{f}}$など複数のステップへと分割されたが、たった一つのモデルで表現することは困難である。従って、以下のように各ステップに毎にサブモデルを導入する。

$$\begin{aligned} Pr(\boldsymbol{f}, \boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}} | \boldsymbol{e}) &= Pr(\boldsymbol{f}, \boldsymbol{\alpha} | \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}}, \boldsymbol{e}) Pr(\bar{\boldsymbol{f}} | \bar{\boldsymbol{e}}, \boldsymbol{e}) Pr(\bar{\boldsymbol{e}} | \boldsymbol{e}) \\ &\approx \prod_{i=1}^{K} p_d(\beta_i | \beta_{i-1}) p_\phi(\bar{f}_i | \bar{e}_i) \times p_l(K | \boldsymbol{e}) \end{aligned}$$

ここで$Pr(\boldsymbol{f}, \boldsymbol{\alpha} | \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}}, \boldsymbol{e})$および$Pr(\bar{\boldsymbol{f}} | \bar{\boldsymbol{e}}, \boldsymbol{e})$、$Pr(\bar{\boldsymbol{e}} | \boldsymbol{e})$はそれぞれ、並び替え確率、句翻訳確率、分割確率と捉えることができる。また、各モデルは以前のステップに依存した複雑なものであるために、近似を入れている。具体的には、分割確率は句の数で決定し$(p_l(K|\boldsymbol{e}))$、句翻訳確率は$p_\phi(\bar{f}_i | \bar{e}_i)$により、句を単位としてそれぞれ独立で決定される。このような句単位に決定される翻訳をフレーズペアと呼ぶ。また、並び替えは、$\alpha_j = i$が成立するとき$\beta_i = j$となるような逆アライメント$\boldsymbol{\beta}$により表現され、前のアライメントβ_{i-1}に依存して決定される$(p_d(\beta_i | \beta_{i-1}))$。また、言語モデルとしては一般的に ngram 言語モデル$p_n(\boldsymbol{e})$を用いる。まとめると、翻訳を生成する問題は、以下の最大化の問題として表される。

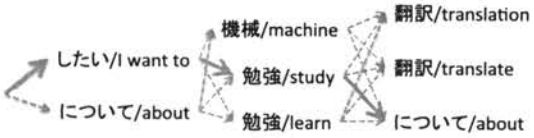
$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{e}} \sum_{\boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}}} \prod_{i=1}^{K} p_d(\beta_i | \beta_{i-1}) p_\phi(\bar{f}_i | \bar{e}_i) \, p_l(K | \boldsymbol{e}) p_n(\boldsymbol{e})$$

最適な翻訳を出力するためには全ての$\boldsymbol{e}$について可能な$\langle \boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}} \rangle$を列挙する必要があり、計算量の観点からほぼ不可能である。そこで、ビタビ近似と呼ばれる、導出を最大化する翻訳を選択する。

$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{\alpha}, \bar{\boldsymbol{f}}, \bar{\boldsymbol{e}}} \prod_{i=1}^{K} p_d(\beta_i | \beta_{i-1}) p_\phi(\bar{f}_i | \bar{e}_i) \, p_l(K | \boldsymbol{e}) p_n(\boldsymbol{e})$$

知識に基づく機械翻訳などの従来法では、構文解析や変換など各ステップで確率あるいはスコアが最大の導出を逐次選択している。統計的機械翻訳では、各ステップで確率が最大の導出を選択するのではなく、全てのサブモデルの積が最大の導出を求めている。さらに、フレーズペアなど基本単位を規定し、その組み合わせにより導出が構成される。

句に基づく機械翻訳では、この基本単位を利用し、目的言語が生成される順番に網羅的に句を組み合わせたグラフを生成する[5]。例えば、以下の図はグラフの一部を図示しており、各フレーズペアがノードで表現され、エッジで結ばれている。ある任意の経路は翻訳結果を示しており、この例の場合、



実線で示されている経路が正解の翻訳結果を表す。例えば、「したい 勉強 について」という順番で原言語(日本語)が組み合わせており、対応する目的言語(英語)は"I want to study about"が生成されている。このように翻訳の生成の問題は探索の問題に置き換わっており、確率が最大の経路を探索することにより、最適な翻訳を出力する。

以上、句に基づく機械翻訳を例として、統計的機械翻訳がどのような考え方により実現されているかを示した。まとめると、1. 翻訳が生成される過程を各ステップへと分解し、2. 各ステップを表現するサブモデルを実現、3. 近似により句などの最小単位でのモデルへと表現する。さらに、4. ビタビ近似により確率を最大にする導出を生成するが、この時、5. 複数の導出をコンパクトに表現した空間から探索する。

モデルを学習

統計的機械翻訳では、翻訳の生成過程を表現したサブモデルへと分解された。続いて、各サブモデルの確率値などを得る必要があるが、これを全て人手で記述するのは不可能であろう。そこで、対訳データと呼ばれる、原言語の文と対応した目的言語の文とのペアから構成されるデータを用いて、パラメータを自動的に学習する。

例えば、以下の様な日本語および英語の対訳文があったとする。

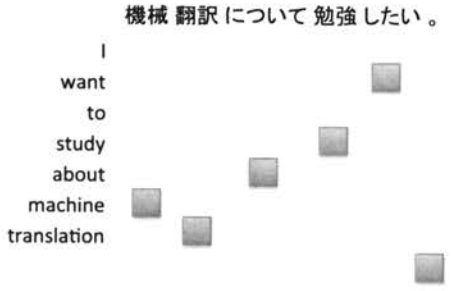
機械翻訳について勉強したい。

I want to study about machine translation.

ここから任意のフレーズペア、例えば「機械翻訳」と"machine translation"だけでなく、"about machine"など網羅的に対応付けを計算し、対応付けられたフレーズペアの頻度にもとづいて最尤推定で翻訳確率を求めることができる。具体的には、句翻訳確率$p_\phi(\bar{f}|\bar{e})$は、フレーズペア$\langle \bar{f}, \bar{e} \rangle$の原言語側および目的言語側が同じ対訳文で共起する頻度$N(\bar{f}, \bar{e})$を元にして計算される。

$$p_\phi(\bar{f}|\bar{e}) = \frac{N(\bar{f}, \bar{e})}{\sum_{\bar{f}} N(\bar{f}, \bar{e})}$$

実際には、網羅的な対応付けは計算量が大きく、また精度も良くないため、単語アライメントと呼ぶ、単語単位の対応付けによる制約を加え、尤もらしいフレーズペアのみを列挙して計算される[5]。例えば、



のような四角で表された単語アライメントが与えられているとする。「機械翻訳について」は"about machine translation"との対応付けが可能であるが"about machine"との対応付けは「翻訳」と"translation"との単語対応を無視しているため、不可能である。

では、単語の対応付けはどのように求められるであろうか。言語の専門家により百万対訳文など大規模な対訳データに対して単語アライメント付与するのは不可能であろう。辞書などを使うことによって、「機械」と"machine"、「翻訳」と"translation"が対応付けられるであろうし、"I"は対応する単語が日本語に存在しない、ことが自動的にわかる。では、例えば、ミャンマー語・日本語など辞書が存在しない言語対の場合、どのように単語アライメント求められるであろうか。

IBM 翻訳モデルと呼ばれる単語アライメントモデルは、対訳データから単語の対応付けに必要とされる知識を自動的に獲得し、その知識を元にして対訳

データに単語アライメントを付与するモデルである[3]。具体的には、単語の対応付けを導出として用いてモデル化され、対訳データに対して EM アルゴリズムを適用することで教師なしでモデルのパラメータの学習が行なわれる。ビタビ近似により、任意の対訳文に対して単語アライメントが推定される。

まとめ

統計的機械翻訳について、その基本的な考え方を解説した。最も特徴的なのは、翻訳の問題を分割および近似により計算可能な問題へと置換え、探索により最適な翻訳を生成する問題へと落としている点にある。また、対訳データに基づいて自動的にモデルのパラメータを学習する。

機械翻訳の大幅な性能向上には BLEU[6]などの客観評価尺度およびその評価尺度に基づいて計算される翻訳誤りを直接最小化する最適化手法[7]が挙げられる。また、句をより一般化した、同期文法に基づく手法[8]や、特に日英など文法的に全く異なる言語対に関しては、事前並び替えなどの手法であらかじめ原言語を目的言語の語順に並び替える手法が成功している[9]。

統計的機械翻訳が始まったのは今から 20 年前であり、10 年前には、その枠組がほぼ固まった。今後は同期文法などの形式言語論の進化および機械学習技術の発展とともに機械翻訳のさらなる性能向上が果たされると思われる。

参考文献

[1] W. Weaver, "Translation," Rockefeller Foundation Archives, a memorandum written on July 15, 1949.

[2] Automatic Language Processing Advisory Committee, "Language and Machines: Computers in Translation and Linguistics," National Academy of Sciences, National Research Council, Washington D.C., 1966.

[3] P.F. Brown, V.J.D. Pietra, S.A.D. Pietra, and R.L. Mercer, "The Mathematics of Statistical Machine Translation: Parameter Estimation," Computational Linguistics, Vol. 19, No. 2, pp. 263-311, 1993.

[4] 渡辺太郎, 今村賢治, 賀沢秀人, G. Neubig, 中澤敏明, 奥村学, 機械翻訳, 自然言語処理シリーズ, No. 4, コロナ社, 2014.

[5] P. Koehn, F. Och, and D. Marcu, "Statistical Phrase-based Translation," Proc. of NAACL-HLT, pp. 48-54, Edmonton, Canada, 2003.

[6] K. Papineni, S Roukos, T. Ward and W.J. Zhu, "Bleu: a Method for Automatic Evaluation of Machine Translation," Proc. of ACL, pp. 311-318, Philadelpha, Pennsylvania, 2002.

[7] F.J. Och, "Minimum Error Rate Training in Statistical Machine Translation," Proc. of ACL, pp. 160-167, Sapporo, Japan, 2003.

[8] D. Chiang, "Hierarchical Phrase-based Translation," Computational Linguistics, Vol. 33, No. 2, pp. 201-228, 2007.

[9] H. Isozaki, K. Sudoh, H. Tsukada and K. Duh, "Head Finalization: A Simple Reordering Rule for SOV Languages," Proc. of WMT, pp. 244-251, Uppsala, Sweden, 2010.

Project Report

科学技術論文の日英・日中間の機械翻訳の評価 -第1回アジア翻訳ワークショップ概要-

中澤敏明（科学技術振興機構）美野秀弥（情報通信研究機構）後藤功雄（NHK）
黒橋禎夫（京都大学）隅田英一郎（情報通信研究機構）

1．はじめに

アジア翻訳ワークショップ（The Workshop on Asian Translation, WAT)はアジア言語を対象とした、新しい評価型機械翻訳ワークショップである。本ワークショップを通じて得られた知見を共有することで、機械翻訳研究において今必要なことが明らかとなり、アジア各国間の機械翻訳が実用的なものになることが期待される。WATのキーワードとして「オープンイノベーションプラットホーム」がある。テストデータを含む全てのデータがあらかじめ公開されており、定められたテストデータでの翻訳評価を繰り返し行うことで、1システムでの翻訳精度の経年変化を見ることや、翻訳システムごとの本質的な翻訳精度の違いを見ることを可能にする。

第1回目のワークショップ(WAT2014)[1]では科学技術論文をドメインとして選び、日英（JE）・英日（EJ)、日中（JC）・中日（CJ）翻訳の評価を行った。報告会は2014年10月4日に行われ、50名を超える参加者が集まり盛況であった。また Agency for the Assessment and Application of Technology (BPPT)の理事である Dr. Ir. Hammam Riza 氏の招待講演も行われ、ASEAN諸国における機械翻訳の現状について知ることができた。

本稿ではWAT2014の概要や結果などをまとめて報告する。なお評価結果の詳細や各参加チームの翻訳システムの説明などは、全てWAT2014のウェブサイトで確認することができる(http://lotus.kuee.kyoto-u.ac.jp/WAT/)。

2．データセット

データはJSTより研究利用目的で一般に公開されている「アジア学術論文抜粋コーパス(ASPEC)」を利用した。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）とで構成される。ASPECの詳細についてはAAMTジャーナル56号（2014年6月）ですでに解説されているので、ここでは割愛する。

3．ベースラインシステム

人手評価は特定のベースラインシステムとの比較に基づいて行った。この比較基準となる特定のベースラインシステムとして、フレーズベース統計的機械翻訳システムを選択した。

フレーズベース統計的機械翻訳システムに加えて、3種類の他の統計的機械翻訳システム、5つの商用ルールベース機械翻訳システム、2つのオンライン機械翻訳システムもベースラインシステムとして利用した。ベースラインシステムの統計的機械翻訳システムは、公開されているソフトウェアで構成し、システムの構築方法と翻訳方法の手順はWAT 2014のウェブサイト[1]で公開している。ベースラインシステムの統計的機

[1] http://lotus.kuee.kyoto-u.ac.jp/WAT/baseline/baselineSystems.html

械翻訳システムにはMosesを利用し、英語と中国語の構文解析器にはBerkeley parserを利用した。ベースラインシステムと適用したサブタスクを表1に示す。表1では商用システムおよびオンラインシステムのシステムIDは匿名にしている。

以下、ベースラインシステムの統計的機械翻訳システムの設定について示す。

- フレーズベース統計的機械翻訳
  - distortion-limit = 20
  - msd-bidirectional-fe lexicalized reordering
  - Phrase score option: GoodTuring
- 階層フレーズベース統計的機械翻訳
  - max-chart-span = 1000
  - Phrase score option: GoodTuring
- String-to-Tree 統計的機械翻訳
  - max-chart-span = 1000
  - Phrase score option: GoodTuring
  - Phrase extraction options:
    MaxSpan = 1000, MinHoleSource = 1, NonTermConsecSource
- Tree-to-String 統計的機械翻訳
  - max-chart-span = 1000
  - Phrase score option: GoodTuring
  - Phrase extraction options:
    MaxSpan = 1000, MinHoleSource = 1, MinWords = 0, NonTermConsecSource, AllowOnlyUnalignedWords

他のシステムパラメータの設定には、標準設定を用いた。

4. 自動評価

4. 1 自動評価スコアの計算手法

機械翻訳の自動評価は、機械翻訳結果と

表1：ベースラインシステム

システムID	システム	種類	JE	EJ	JC	CJ
SMT Phrase	Moses フレーズベース統計的機械翻訳	統計ベース	✓	✓	✓	✓
SMT Hiero	Moses 階層フレーズベース統計的機械翻訳	統計ベース	✓	✓	✓	✓
SMT S2T	Moses String-to-Tree 統計的機械翻訳および Berkeley parser	統計ベース	✓		✓	
SMT T2S	Moses Tree-to-String 統計的機械翻訳および Berkeley parser	統計ベース		✓		✓
RBMT X	The 翻訳 V15（商用システム）	ルールベース	✓	✓		
RBMT X	ATLAS V14（商用システム）	ルールベース	✓	✓		
RBMT X	PAT-Transer 2009（商用システム）	ルールベース	✓	✓		
RBMT X	J 北京 7（商用システム）	ルールベース			✓	✓
RBMT X	蓬莱 2011（商用システム）	ルールベース			✓	✓
Online X	Google translate (July, 2014)	（統計ベース）	✓	✓	✓	✓
Online X	Bing translator (July, 2014)	（統計ベース）	✓	✓	✓	✓

参照訳(翻訳の正解となる訳)とを比較して翻訳結果の精度を数値化して行う。WAT2014では、2種類の異なる自動評価尺度BLEU[2], RIBES[3]を用いた。BLEUの自動評価スコア計算には、Moses toolkit内の *multi-bleu.perl* を用いた。RIBESの自動評価スコア計算には、*RIBES.py version 1.02.4* を用いた。参照訳は各タスクで1つずつ用意した。また、スコアの計算時に必要となる単語分割処理には、日本語、中国語で複数の単語分割ツールを利用した。各言語で利用した単語分割ツールは次の通りである。

- 日本語単語分割ツール:Juman, KyTea, MeCab (IPA辞書)
- 中国語単語分割ツール:KyTea, Stanford Word Segmenter (CTBモデルおよびPKUモデル)
- 英語単語分割ツール:Moses toolkitの *tokenizer.perl*

自動評価の詳細な手順は、WAT2014のウェブサイト[2]にて公開している。

4．2 自動評価システム

WAT2014では、自動評価システムを用意し、参加チームが機械翻訳結果の自動評価結果をいつでも確認できるようにした。翻訳結果はWAT2014のウェブサイトからいつでも提出することができる。提出された翻訳結果は即時に自動評価サーバーによって自動評価が行われ、スコアが出力される。翻訳結果の提出の際には下記の情報を入力してもらい、提出時にスコアの公開を許可した(ivの項目を可とした)場合は、自動評価後にWAT2014のウェブサイト上にて提出ファイルの自動評価スコアがランキング形式で公開される。

i) タスク:日本語⇔英語, 日本語⇔中国語
ii) 手法:統計的機械翻訳, ルールベース翻訳, 統計的機械翻訳とルールベースの両方を用いた手法, その他の手法
iii) ASPEC以外のデータ(対訳データや単言語データなど)の利用の有無
iv) 自動評価スコアのウェブサイト上での公開の可否

自動評価サーバーは、提出された全ての情報を保持、保管しており、提出されたデータの確認、修正は提出チームがいつでも可能である。自動評価システムは本ワークショップ終了後も利用可能であり、引き続き翻訳結果の提出・修正を受け付けている。

5．人手評価

機械翻訳の人手評価には、1.非常に多くの時間とお金がかかる、2.様々な基準が存在する、3.評価者間の一致度が低いなど、様々な解決すべき問題が存在する。WAT2014ではクラウドソーシングを利用することで問題1を解決した。クラウドソーシングを利用した翻訳の評価は、他のワークショップにおいても採用されている(IWSLT2011, 2012やWMT2012, 2013など)。今回は様々存在するクラウドソーシングサービスの中からランサーズを利用した。ランサーズを利用した理由は二つあり、一つは依頼する作業のカテゴリーを指定できる点、もう一つは、「本人確認済」の作業者を指定できる点である。これらの機能を使うことで、より適切な作業者が作業を行うことが期待できる。

問題2については、ベースラインとなる機械翻訳結果を用意しておき、これと各システムの翻訳結果を1文ずつ比較し、その勝敗数をスコア化(HUMANスコア)することで各システムを評価するという方法を採

[2] http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html

用した。評価者には入力文とベースラインおよび評価対象システムの翻訳の 3 つが提示され、どちらの翻訳がより良いか、または同程度かの 3 択で評価を行う。ベースラインに対する勝ち数を W、負け数を L、引き分け数を T とすると、HUMAN スコアは以下の式で計算できる：

$$HUMAN = 100 \times \frac{W - L}{W + L + T}$$

HUMAN スコアは-100 から 100 の値を取り、正の値は全体としてベースラインより良い翻訳結果であり、負の値は逆に悪い翻訳結果であるという傾向を示す。

クラウドソーシングの性質上、各文ペアの評価は異なる評価者が行うことになる。ここで問題 3 の影響を軽減するために、各文ペアの評価を複数の異なる評価者に行ってもらい、意見を集約することで評価を安定させた。評価対象システムの翻訳がベースラインよりも良いという判断を+1、悪いという判断を-1、同程度を 0 としたとき、全ての評価者の判断を足し合わせて正の値となれば最終判断を勝ち、負の値ならば負け、0 ならば同程度とした。

WAT2014 では人手評価対象文として、テストデータからランダムに 400 文を選択した。また各文の勝敗は異なる 3 人の評価者の判断を集約することで決定した。なお評価者による 1 つの文ペアの評価費用は 5 円と設定した。1 システムの評価には異なる 3 人ずつに 400 文を評価してもらう必要があるため、1 システムの 1 つの翻訳結果の評価にかかる費用は 3 人×400 文×5 円で 6000 円となる。

6．参加チームリスト

表 2 に WAT2014 に翻訳結果を提出したチームと、各チームが参加したサブタスク

表 2：WAT2014 に翻訳結果を提出したチームと参加タスク一覧

チーム ID	組織名	JE	EJ	JC	CJ
NAIST	奈良先端科学技術大学院大学	✓	✓	✓	✓
EIWA	山梨英和大学	✓			✓
Kyoto-U	京都大学	✓	✓	✓	✓
WEBLIO-EJ1	ウェブリオ		✓		
TMU	首都大学東京	✓			
BJTUNLP	北京交通大学			✓	
NII	国立情報学研究所	✓			
SAS_MT	SAS Research and Development		✓		✓
Sense	ザールラント大学 & 南洋理工大学	✓	✓	✓	✓
NICT	情報通信研究機構			✓	
TOSHIBA	東芝	✓		✓	
WASUIPS	早稲田大学			✓	✓

の一覧を示す。全 12 チームが自動評価サーバーに翻訳結果を提出し、WASUIPS を除く 11 チームが人手評価も行った。国外から 3 チームが参加し、企業から 3 チームが参加した。

7．評価結果

図 1 に自動評価結果、図 2 に人手評価結果を示す。また人手評価の結果から、次のことが確認された。

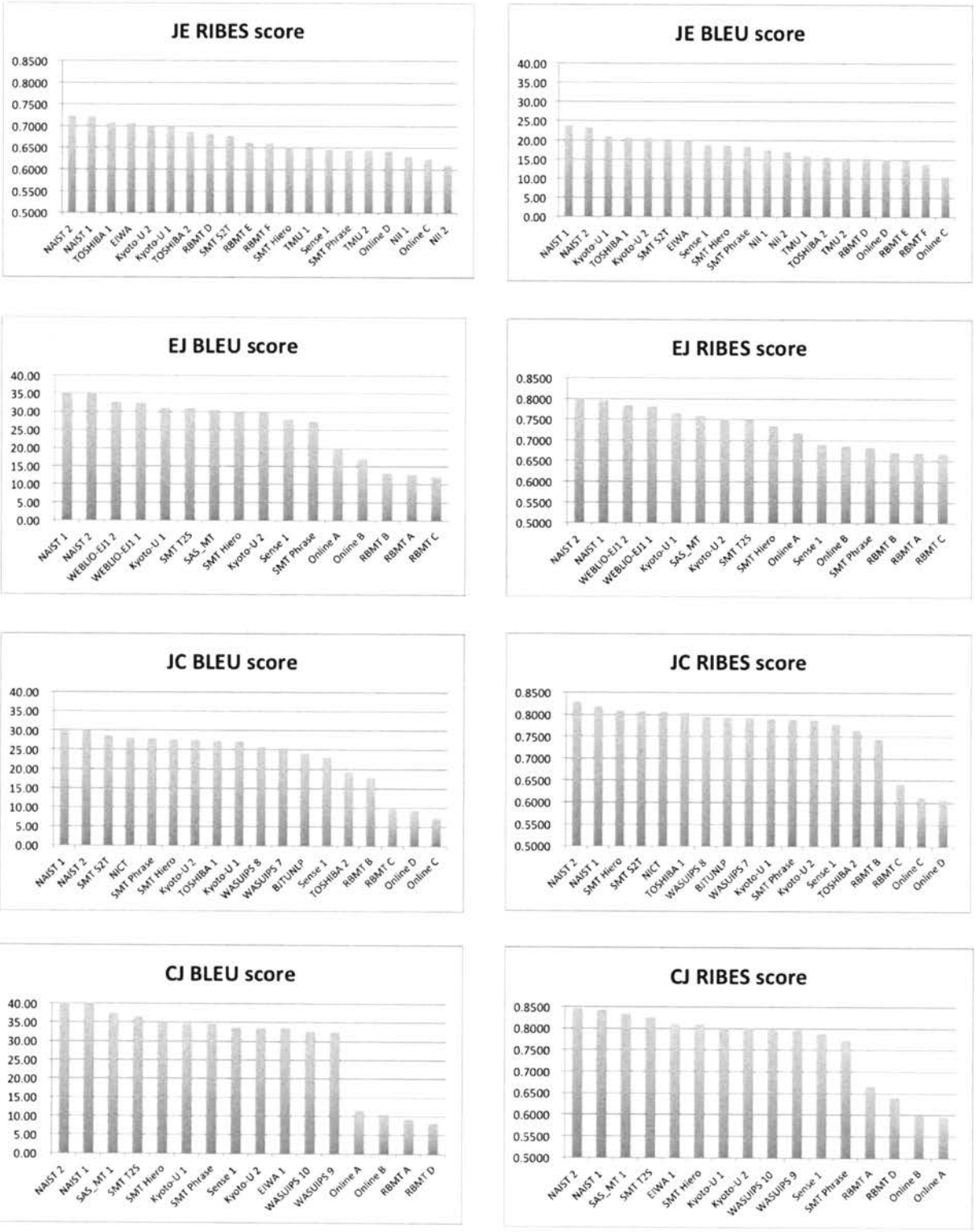


図 1 ：自動評価結果

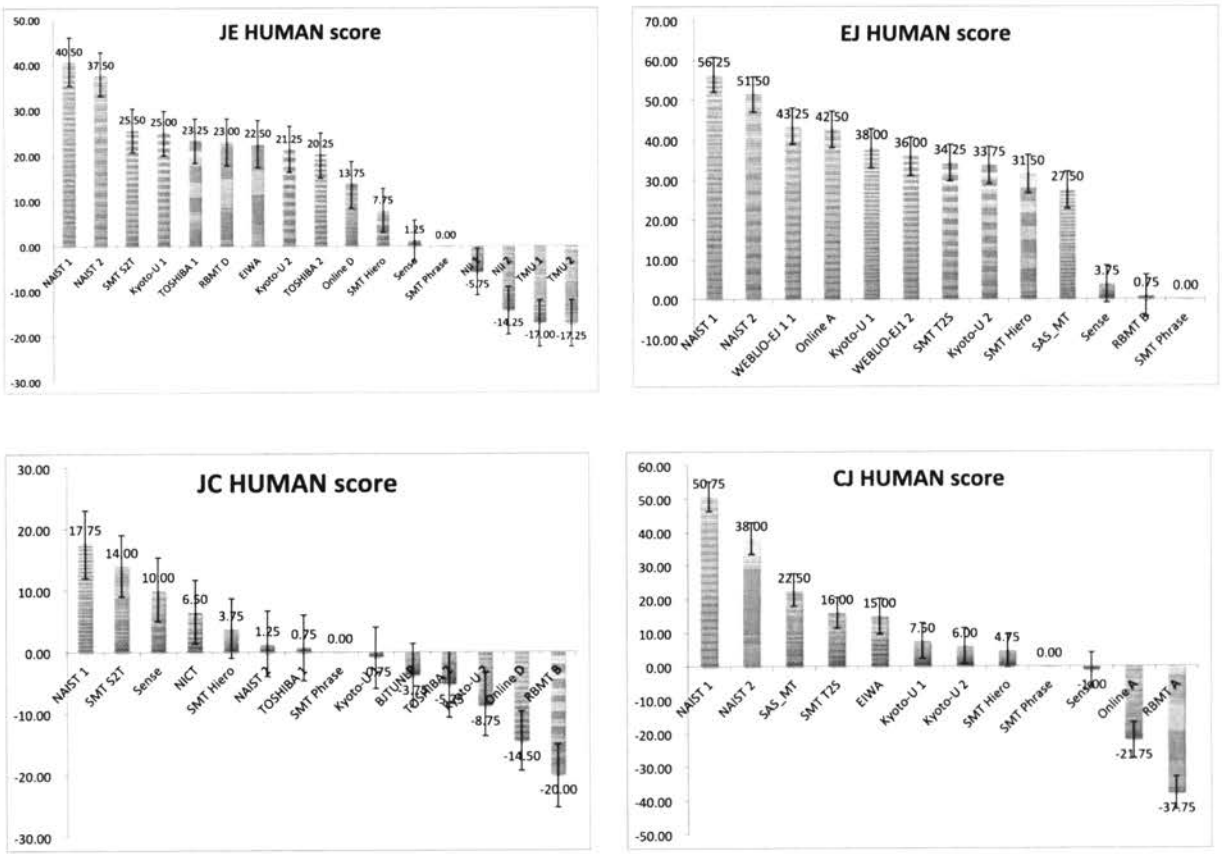


図２：人手評価結果

- 最高性能の統計的機械翻訳システムはルールベースシステムより良い評価であった。
- ベースラインシステム間の比較による訳質の順は次のようであった。フレーズベース統計的機械翻訳＜階層フレーズベース統計的機械翻訳＜Tree-to-String/String-to-Tree 統計的機械翻訳
- Forest-to-String 統計的機械翻訳システム[4]が全ての翻訳方向で最高評価を達成した。

なお、より詳細な評価結果や分析結果などは、WAT2014 のオーガナイザーによる論文に詳しく記載されている[1]。

８．提出データ

本ワークショップには 12 チームが参加した。自動評価用に提出された翻訳結果数は 100 以上あり、人手評価用に提出された翻訳結果数は 37 であった（ワークショップ開催後も自動評価サーバーへの提出は可能なため、自動評価用に提出された翻訳結果数はワークショップ直前のものである）。

９．まとめと今後の展望

本稿では WAT2014 の概要や結果などについて概説した。初の試みであったが国内外から 12 チームが参加し、様々な手法での翻訳結果が集まり、これらを分析することで様々な知見が得られた。なお提出された翻訳結果は近日中に一般公開する予定である。

WAT は今後も毎年開催する予定である。現在のところ日中の特許文書を新たなデータセットとして利用できる見通しがたっており、新たな言語ペアを追加することも検

討中である。なお今回は行えなかったが、文脈を考慮した翻訳評価も今後行っていきたいと考えている。

参考文献

[1] Toshiaki Nakazawa, Hideya Mino, Isao Goto, Sadao Kurohashi, and Eiichiro Sumita. 2014. Overview of the 1st Workshop on Asian Translation. In Proceedings of the 1st Workshop on Asian Translation (WAT2014), pages 1-19.

[2] Kishore Papineni, Salim Roukos, ToddWard, andWeiJing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of ACL, pages 311–318.

[3] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. Automatic evaluation of translation quality for distant language pairs. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 944–952.

[4] Graham Neubig. 2014. Forest-to-String SMT for Asian Language Translation: NAIST at WAT 2014. In Proceedings of the 1st Workshop on Asian Translation (WAT2014), pages 20-25.

シンポジウム参加報告

COLING 2014 参加報告

ヤフー株式会社
野口 正樹，中島 泰，小林 隼人

1. はじめに

COLING は自然言語処理分野の国際会議の中でも最も歴史が古く、隔年開催と丸一日の遠足（Excursion）で有名な会議です。今年はアイルランドの首都ダブリンで開催され、全体で 700 人以上の参加者が集まりました。採択率は 31.55%（オーラル 139 件＋ポスター79 件／投稿 691 件）と比較的低く、難関会議の一つとなっています。

全体観としては、やはり最近流行りの分散ベクトル（word embedding）や深層学習（deep learning）に関する話が多い印象を受けました。これらは（温故知新な部分があるとはいえ）研究の最前線が実応用に使われている稀有な例ですので、企業研究者の一人としては好意的に見ています。一過性のものなのか、次のブレイクスルーが起こるのか、注意して見守っていく必要がありそうです。

本報告では、著者らが参加したワークショップ、チュートリアル、本会議の中からそれぞれ印象に残った発表をご紹介します。

2. ワークショップ

ワークショップからは、セマンティクスや情報抽出に関する「Semantic Web and Information Extraction (SWAIE)」(http://swaie2014.wordpress.com/)をご紹介します。既存手法を様々な分野のコーパスに対して適用する取り組みや、その際に得られた知見を発表する論文が多かったように思います。特に、電子化されて機械処理が可能になった電子カルテや、利用者が拡大しているソーシャルメディアの投稿をコーパスとして扱うための取り組みが印象的でした。また、どうすればその課題が解決できそうかという実応用視点での質疑応答が活発に行われていました。

本ワークショップでは招待講演が 2 件あり、これらも興味深かったので少し詳しくご紹介します。

1 件目は「How can IE and Semantics Improve Medical Practice?」というタイトルで、開催地である Dublin City University のポスドク研究者である Lorraine Goeuriot 氏の医療分野における情報抽出に関する講演です。

この講演では、情報処理技術・検索技術を医療分野の文書に適用した際の精度と解析ミスの分析を、実例を交えて説明されていました。特に、カルテの文書については、医師がメモとしても使うため文の構造が不十分であったり、ある物質を表現するのに各医師で（同じ医師でもしばしば）異なる表記を用いたりするため、抽出ルールを決めたり学習させることが困難であることを強調していました。講演の最後には、医療データを使ったマイニングは未だ発展途上であり、現在使えている技術と現実とのギャップの大きさを理解して、医師とのコミュニケーションをとりながらシステム化すること、またシステムをよりパーソナライズし使いやすくしていくことが必要だとまとめていました。

2 件目は「Watson Today. Cognitive Computing's road to Reality.」というタイトルで、D.J. McCloskey 氏（IBM Watson Group）によるクイズ番組 The Jeopardy!で優勝したワトソンについての講演です。

この講演では、ワトソンが Deep Blue の後継として開発がスタートしたという話から始まり、The Jeopardy!に挑戦するにあたっての改善の歴史などを交えつつ、どういう手順で現在に至っているかを紹介していました。現在のワトソンは、医療やファイナンスなど様々な分野で意思決定をサポートできるよう開発が進められているそうです。その中

で、目標とすることが一見不可能に見えても、実現をするためにはどんな技術が必要か、現状でどこまでできているのかを整理することが大切で、それが成功への第一歩だという話が印象的でした。

3. チュートリアル

チュートリアルからは、分散ベクトルの学習ツール word2vec の開発者として有名な Thomas Mikolov 氏（Facebook）によるニューラルネットワークの講義「Using Neural Networks for Modelling and Representing Natural Languages」をご紹介します。

この講義では、基本的な N-gram の紹介から始まり、分散ベクトルや Recurrent Neural Network までの解説が行われました。用意されていた席のほとんどが埋まっていたことからも、ニューラルネットワークの自然言語処理応用への関心の高さがうかがえました。講義内容のスライドは COLING 公式ページで公開されていますので、興味のある方はご確認ください（http://www.coling-2014.org/tutorials.php）。

質疑応答の時間には、研究者としてのあり方や word2vec に対する見解など、Mikolov 氏の頭の中の一端を覗き見ることが出来ました。その中でも特に印象に残ったのは、採択される論文のあるべき姿についてでした。学会で採択される論文のほとんどは既存手法に対して精度を上げたことを成果としています。それに対して Mikolov 氏は、良い精度が出ていなくてもパラダイムシフトを起こしそうな研究を学会側がもっと採用するべきだと話していました。また、研究者は自分が作ったものを公開しないケースが多く、もっと GitHub などを活用して第三者が利用可能な状態にするべきだと話していました。

4. 本会議

本会議からは、IBM Best Papers Awards に選ばれた論文の 1 つを紹介します。深層学習手法の一つである Convolutional Deep Neural Network (CDNN)を名詞間の関係抽出に利用した論文です。例えば、「The machine has two units.」という文の中で「units」は「machine」に含まれる関係（Component-Whole）にあり、「The fire inside WTC was caused by exploding fuel.」の中では「fire」を「fuel」が引き起こす関係（Cause-Effect）にあるといった関係性を抽出することが目標になっています。

論文中では、CDNN の入力として単語レベル特徴と文レベル特徴を用いています。前者は注目している名詞とその周辺単語や上位語単語の分散ベクトルを連結したもので、後者は文中の N-gram の分散ベクトル集合と単語の位置情報を表すベクトルを連結したものです。実験結果として、人手で作成した特徴を用いた SVM や、Recurrent Neural Network を用いた State-of-the-art 手法と比較して、提案手法の認識精度が一番高かったことが報告されています。また、文レベル特徴の方が単語レベル特徴よりも精度への影響が大きく、両方の特徴を合わせて使うことで大幅に精度改善ができるとのことです。

これらの結果は、画像・音声処理で猛威を振るっている CDNN が、自然言語処理でも活用していける可能性を示唆しています。COLING は ACL 等と比べて伝統的な言語処理を扱うイメージがあるので、この論文が COLING のベストペーパーを取るというのは、いよいよパラダイムシフトが起こりつつある（起こそうとしている）のかなという印象を受けました。

5. おわりに

今年の COLING における国別の採択数によると、日本のプレゼンスは、中国、アメリカ、ドイツについで四番目だそうです。しかし、国別の採択率を見てみると、実は日本の採択率は全体の平均よりも低いのだそうです（もちろん中国、アメリカ、ドイツよりも低い！）。中国に負けているということは、対象言語だけの問題ではないように思います。国内の Web サービス等で直接恩恵を受けうる日本語の研究を推進するためにも、我々はもっと頑張る必要がありそうです。

次回の COLING 2016 は大阪で開催されます。本報告が、COLING 2016 で発表されるような新しい研究ネタを探す一助になれば幸いです。

シンポジウム報告

第3回特許情報シンポジウム開催報告

梶　博行

AAMT/Japio 特許翻訳研究会副委員長・静岡大学

1．はじめに

AAMT/Japio 特許翻訳研究会は、2003 年に設置されて以来、特許の機械翻訳に関わる内外の技術やシステムの調査研究を行っている。その活動の一環として、2010 年と 2012 年に特許情報シンポジウムを開催し、特許情報処理に関心をもつ研究者、開発者、利用者、政策担当者が一堂に会して議論する場を提供してきた。今回、第 3 回のシンポジウムを 2014 年 11 月 28 日（金）にキャンパスイノベーションセンター東京で開催したので、その概要を報告する。

AAMT/Japio 特許翻訳研究会の辻井潤一委員長の開会挨拶のあと、3 件の招待講演、2 件の研究会報告、1 件の特別講演、6 件の一般講演が行われ、Japio の守屋敏道専務理事の閉会挨拶でシンポジウムを締めくくった。過去 2 回と異なり、外国の特許庁関係者による招待講演はなく、すべて日本語による発表であったが、多数の参加者を得て、活発な討論が行われた。参加者の内訳は AAMT 会員 2 名、大学等教育機関 13 名、団体・研究機関 10 名、翻訳会社（システム提供を含む）11 名、調査会社（DB 提供を含む）5 名、一般事業会社・個人等 13 名で、講演者ならびに AAMT/Japio 特許翻訳研究会メンバーとあわせて計 84 名であった。

2．招待講演

最初に、『特許庁における機械翻訳への取組』と題して、特許庁総務部特許情報企画室の櫻井健太様にご講演いただいた。特許審査業務の重要なステップである先行技術調査では外国文献を含めて調査しなければならないが、英語だけでなく中国語や韓国語の文献が重要になっている。審査官には中国語や韓国語の読解力がある人が少ないので、機械翻訳が必須である。中国語・韓国語の特許公報を日本語に機械翻訳し、日本語で検索できるシステムを活用していること、今後は ASEAN 言語についても同様なサービスが必要になることが述べられた。また、各国特許庁が審査情報を共有する AIPN（Advanced Industrial Property Network）が紹介され、日本の審査結果を英語で提供するために日英機械翻訳を利用していること、利用者である海外特許庁審査官からのフィードバックや未知語の収集・翻訳メモリの構築など特許庁における取り組みが述べられた。

次に、『多言語機械翻訳の研究開発動向』と題して、（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の隅田英一郎様にご講演いただいた。文法規則を人間が作成する旧方式「規則翻訳（RBMT）」と翻訳例を集めたパラレルコーパスから翻訳知識を自動的に学習する新方式「統計翻訳（SMT）」を対比し、後者の翻訳精度が優ってきていることが強調された。旅行分野のような限られたドメインだけでなく、特許翻訳にも新方式が有望であるという見方が示された。また、従来は原言語と目的言語の構文解析が必要であったが、目的言語の構文解析のみを必要とし、原言語の文

法を自動獲得する新しいアイデアも紹介され、新方式が多言語翻訳に適した方法であると述べられた。

さらに、『特許情報検索サービスにおける機械翻訳の活用』と題して、日本パテントデータサービス（株）企画室の早川浩平様にご講演いただいた。特許情報と機械翻訳は切り離せない関係になっており、同社の特許情報検索サービスでは各国の特許情報をすべて英語に翻訳し串刺し検索を可能にしている。対訳比較表示などのインタフェースや活用方法を含めてこのサービスが紹介された。また、翻訳精度に関しては改善が必要であることが指摘された。特許文書特有の言い回しで正しく翻訳されない例、文脈に沿って訳出できない多義語の例、アメリカ英語とイギリス英語の取り扱いの問題などが具体的に示された。

３．研究会報告・特別講演

AAMT/Japio 特許翻訳研究会の重要なテーマの一つである辞書の自動構築に関し、宇津呂武仁委員（筑波大学）が『パテントファミリーからの専門用語対訳辞書の構築』を報告した。パテントファミリーとは同一発明を各国に出願することによって得られるコンパラブルな明細書の組である。そのようなコーパスから対訳専門用語を抽出する手法に関する宇津呂委員の研究グループの研究成果が紹介された。文レベルで対訳になっている部分については、統計的機械翻訳の標準的な手法であるフレーズテーブルを作成することによって、また文レベルで対訳になっていない部分については、要素合成法を用いることによって、それぞれ高精度で対訳が抽出できることを明らかにした。また、一つの語の訳語として抽出された複数の語が同義語であるかどうかを判定する試みにも言及した。

特許翻訳研究会が力を入れているもう一つのテーマが翻訳システム／翻訳文の自動評価である。これに関連し、越前谷博委員（北海学園大学）が『自動評価法を用いた機械翻訳の定量的評価』と題して、本年 6 月にボルチモアで開催された WMT 2014 (The 9th Workshop on Statistical Machine Translation) の自動評価タスクについて報告するとともに、日本発の自動評価法である APAC と RIBES を紹介した。さまざまな自動評価法が提案されているが、ほとんどが機械翻訳システムによる翻訳文と人間が作成した参照訳との類似度を計算する方法である。自動評価法自体の評価は、自動評価法によるスコアと人手による翻訳文評価の相関を求めることによって行われる。WMT 2014 の自動評価タスクでも、参加した自動評価法がそのような方法で競い合った。機械翻訳システムに対して一つのスコアを出力する“システムレベルの評価”では人手評価との相関が 0.8 を超えるのに対し、翻訳文ごとにスコアを出力する“セグメントレベルの評価”では人手評価との相関が 0.4 程度と低かったこと、どの言語対に対しても一貫して優位な自動評価法はなかったことなどが報告された。自動評価法も発展途上にあるといえる。APAC は越前谷委員が考案した方法で、参照訳との一致単語列（チャンク）に着目した方法である。また、RIBES は磯崎秀樹委員（岡山県立大学）の提案によるもので、参照訳との語順の近さを測定する方法である。

特許翻訳においても中国語や韓国語などの重要性が増していることから、アジア言語の機械翻訳システムの研究開発に携わっている中澤敏明委員（科学技術振興機構／京都大学）に特別講演『アジア言語を中心

とした機械翻訳研究』をお願いした。講演の前半は、科学技術振興機構と京都大学が中国科技技术信息研究所などと協力して推進中の日中・中日機械翻訳実用化プロジェクトの紹介であった。言語構造や語順が大きく異なる言語対であることを考慮した依存構造木のアライメント・依存構造木間の翻訳方法を開発し、従来方法より高い翻訳精度を達成している。後半は、中澤委員がオーガナイザとして本年10月に東京で開催した WAT 2014 (The 1st Workshop on Asian Translation) の報告であった。今回は日・中・英 3 言語間の科学技術論文翻訳タスクを実施したが、来年はインドネシア語 - 英語の新聞記事翻訳や日本語 - 中国語の特許文献翻訳を含めることが検討されている。

4．一般講演

投稿ベースの一般講演では 6 編の論文が発表された。機械翻訳に関する論文が 3 編、その他のトピックの論文が 3 編であった。

機械翻訳関連では、ルールベースの韓日機械翻訳に統計的な訳語選択技術を組み込むことにより、BLEU スコアを 8.5 ポイント向上させた研究が発表された。また、特許事務所における機械翻訳システムや翻訳メモリの活用事例の発表、インターネットから利用できる翻訳ソフトが新語や造語の訳語を調べるのにたいへん有用であることを指摘する発表があった。

翻訳以外の特許情報処理に関しても興味深い研究が発表された。一つ目は、特許分類コード体系である F タームをベースとしてブートストラッピングにより、上位－下位、全体－部分といった用語間の関係を獲得する方法の論文であった。二つ目は、特許文書から化学物質名と化学式の対応関係を抽出し、化学物質名の命名規則を用いてそれらを部品化し、部品を組み合わせることによって新しい化学物質名に対する化学式を生成するという研究であった。三つ目は、特許情報のマイニング手法と分析ツールに関する発表で、特徴技術ネットワーク、課題－解決手段のクロスマップなどの機能が紹介された。

5．おわりに

以上のように、さまざまな立場からさまざまな内容が発表され、会場からも多数の質問と意見が出た。研究開発者と利用者の相互理解が深まったものと思われる。Japio の守屋専務理事の閉会挨拶では、非常に有意義なシンポジウムであったとして、発表者と参加者への謝意とともに 2016 年の第 4 回シンポジウム開催への期待が表明された。

本シンポジウムの発表資料は AAMT/Japio 特許翻訳研究会ホームページ（http://aamtjapio.com/）での公開を予定しています。AAMT/Japio 研究会のその他の活動についてもホームページからご覧いただけますので、興味のある方にアクセスしていただければ幸いです。

シンポジウム報告

第 24 回 JTF 翻訳祭セッション開催報告

秋元圭
AAMT 機械翻訳課題調査委員会

1. はじめに

2014 年 11 月 26 日（水）にアルカディア市ヶ谷において、日本翻訳連盟（JTF）主催の第 24 回 JTF 翻訳祭が開催された。AAMT 機械翻訳課題調査委員会では、この翻訳祭のセッション運営に企画協力し、「いまさら聞けない機械翻訳～動作原理から上手な活用法まで～」と題したパネルディスカッションを実施した。

2. JTF 翻訳祭

JTF 翻訳祭は毎年開催される国内最大規模の翻訳業界のイベントであり、24 の講演&パネルディスカッション、6 つのプレゼン・製品説明コーナー、30 社あまりが出展する翻訳プラザ（展示会）が用意された。本年は過去最高の 884 名（前年 824 名）の参加者があった。登壇者・出展者・受付不要の説明コーナーと展示会のみの参加者を加えると参加者は 1000 人を超えていたようである。

AAMT では、翻訳者や翻訳会社などへの機械翻訳普及のため、JTF 翻訳祭の後援を行っている。AAMT 課題調査委員会では、AAMT の広報・調査活動のため毎年翻訳プラザに出展し、活動紹介・機械翻訳の利用についてのアンケート・UTX（Universal Terminology eXchange）の紹介等を行っており、また本年は昨年に引き続きパネルディスカッションを企画・実施した。

3. パネルディスカッションの様子

パネルディスカッションでは、山田優氏（株式会社翻訳ラボ代表、麗澤大学、青山学院大学など講師）、内山将夫氏（独立行政法人情報通信研究機構(NICT) 主任研究員）、秋元圭（元ルールベース翻訳ソフト開発会社 エンジン開発責任者）の 3 名がパネリストを務め、井佐原均氏（豊橋技術科学大学情報メディア基盤センター センター長・教授、AAMT 理事）がモデレータとして議論を進めた。

定員 120 人のセッション会場は、昨年同様超満員で立ち見が出る盛況ぶりで、機械翻訳に対する高い関心がうかがえた。

4. 議論の趣旨

今回のパネルディスカッションでは、これから本格的な機械翻訳システムの活用を考えている翻訳会社・翻訳者のために、機械翻訳を上手に使うためのエッセンスを提供することを目的に企画を行った。具体的には、「ルールベース翻訳」（以降ルールベースと呼ぶ）と「統計的翻訳」（以降統計ベースと呼ぶ）という機械翻訳の二つの方式について処理概要と訳文の特徴、問題点について豊富な事例を用いて説明するとともに、「それぞれの機械翻訳を運用するために必要となる作業」について紹介することとした。議題は、秋元が「ルールベース」、内山氏が「統計ベース」、山田氏が「運用のための作業」を担当した。90 分の講演時間のうち、各担当者の持ち時間を 20 分ずつ、会場との議論を 30 分とした。

5. ポストエディット

まず初めに、山田氏が運用のための作業として、ポストエディットの研究について紹介し、特徴と効率的な方法を紹介した。要旨は以下のとおり。

- ポストエディットの負荷は翻訳メモリの 85-90% マッチと同等であるという研究結果がある。
- 品質とコストと納期を考慮すると、人手翻訳を選ぶ人は 32.2%、機械翻訳＋ポストエディットは

67.7%、機械翻訳のみは0%。

- ポストエディットによるプロの負荷軽減はほぼ0%で修正量は64.7%。学生の負荷軽減は10～25%で修正量は50.4%。
- 誰がポストエディターになるのかという観点から、学生がなれるのかの実験結果を紹介。プロ3名がポストエディットを行った場合、マイナーエラーがひとつ残った。学生43名では致命的な意味エラーが2つ以上残った。学生43名中、プロの品質レベルに達したのは0名。
- ポストエディットの問題として、エラーの伝播（機械翻訳の間違いが修正されずに残されること）がある。統計ベースの訳は一見自然に見えるので、エラーを見落としやすい。フル・ポストエディットでは訳文の「文体」に注意が向いてしまう。
- 効果的なポストエディットについて、スラッシュリーディングの要領で文を分割して機械翻訳にかけてから人手でつなぎつつポストエディットを行う方法を提案。

6. ルールベース機械翻訳

次に秋元がルールベースの仕組みと、誤訳の原因と対処方法、特徴と課題について紹介した。要旨は以下のとおり。

- ルールベースの仕組みを簡単に説明。
- パッケージ翻訳ソフトの画面、翻訳機能、翻訳メモリ機能、その他の機能を紹介。
- 翻訳ソフトの使用フローとして、(1)用語集設定→(2)翻訳メモリ設定→(3)コーパス設定→(4)翻訳環境/専門語辞書選択→(5)未知語抽出・登録→(6)機械翻訳実行→(7)各文を処理（人手修正・訳語選択・フレーズ翻訳等）→(8)用語抽出・用語確認というフローを提案。
- 誤訳の原因と対処方法について、文分割、係り受け解析ミス、句構造解析ミス、訳語選択ミス（訳語不正）、未知語の原因と対処方法を解説。係り受け解析や句構造解析のミスへの対応には句にフレーズをかける方法、訳語選択ミスや未知語への対応には詳細な辞書登録を行う方法を紹介。
- 訳し分け方法のうち、ユーザーサイドで可能なものとして、意味素性と構文パターンを解説。
- 辞書登録に関連し、UTXとUTX Converterを紹介。
- ルールベースの特徴として、確率的な解釈より文法的な解釈を選ぶこと、訳揺れがないこと、訳抜けがないこと、コーパスが無い異なる分野でもそれなりに訳すことなどを紹介。
- 上記誤訳の原因以外の課題をいくつか紹介し、いずれも文脈解析ができてないという点を指摘。また訳語不足には辞書の増強が必要であり、さらに係り受け解析の精度向上も必要であることを指摘。

7. 統計ベース機械翻訳

内山氏は統計ベースの仕組みと、翻訳結果例と、「みんなの自動翻訳@TexTra®」を紹介した。要旨は以下のとおり。

- NICTの活動として、Japio、日本発明資料、川村インターナショナル、TOPPANへの技術移転、特許庁との協力合意を紹介。
- 統計ベースの仕組みの紹介と、長文解析で2分木の入れ替えによる語順変換（動詞句の動詞と名詞の順番の入れ替え）を使用した統計ベース翻訳で訳質が向上することを紹介。
- 統計ベースの利点として、対訳があれば精度が向上すること、ニューラルネットワーク等の最新アルゴリズムで精度が年々改善していることを指摘。
- みんなの自動翻訳@TexTra®を紹介。汎用は日英・英日・日中・中日・日韓・韓日、特許用は日英・英日・日中・中日が使用可能で、非商用は無料。商用は要連絡。対訳文や対訳用語を登録するとNICTがそれと既存の対訳を合わせて精度向上するので、登録を呼びかけた。
- 特許翻訳については日英・日中とも1000万文以上の対訳を使用しており、訳語選択が得意で、分野

により異なる訳語も周辺文脈から的確に訳出することを指摘し、具体的な翻訳例を紹介。

- みんなの自動翻訳@TexTra®のデモを実施。シンプルなテキストの翻訳機能に加え、PowerPoint ファイルを読み込み、原文だけ抽出し、翻訳を実行して保存すると、レイアウトを保ったまま翻訳ができることを紹介。

8. 議論

最後に会場からの質疑にこたえる形で議論を行った。時間の都合上、議論は4件にとどまったものの、内容の濃い議論が行えた。それぞれの議論の内容について以下に示す。

（1）機械翻訳の未来について

翻訳者 K 様：

工学や IT のような繰り返しの多い仕事をしている。ポストエディットを引き受けてから自分の日本語がおかしくなった。金額的にポストエディットしたくない。IT や特許は、今後 5 年くらいでどれくらい機械翻訳が使えるようになるか？（人間翻訳の仕事が無くなるか）

山田氏：

実験してから 3 年が経過し、かなり使い物になるようになるかと思ったがそうでもない。1 年前の Google の誤訳が今は直っているものがあることから、日々進化はしている。

内山氏：

総務省のグローバルコミュニケーション計画で旅行会話などの訳質向上を行っている。

秋元：

ジャンルを絞れば使い物になるものが出てくると思う。会話文などはオリンピックまでにできるようになるのではないか？

（2）ツールの選び方

CG 制作会社 O 様：

どのツールが良いか。どうやって選べばよいか。

秋元：

さきほどの発表で紹介した機能のうち、どれが必要かを考え、各社のホームページを見て必要機能のある製品を選ぶと良い。なお、翻訳ソフトの使用フローを紹介したが、そのフローの全機能を搭載した製品は存在しない。パッケージ製品は翻訳支援機能が強いが、サーバー系は弱い。

O 様：

パッケージ系はルールベース、サーバー系は統計ベースということか？

秋元：

ルールベースにはサーバー版もある。

O 様：

NICT の「みんなの自動翻訳@TexTra®」は守秘性が心配。

内山氏：

別途契約してもらえれば、技術移転し、NICT にデータが残らないようにもできるので守秘性も問題ない。

O 様：

VoiceTra の音声翻訳はどう実現しているか？

内山氏：

音声認識し、機械翻訳し、音声合成している。

（3）定型外の翻訳に適した方式

翻訳者 H 様：

特許やマニュアルは定型の翻訳だが、例えば、会社の人事・ファイナンス・営業・リーガルのような業務活動一般といった非定型の翻訳については、ルールベースと統計ベースはどちらを選べばよいか？

秋元：

人事などのジャンルは、コーパスを用意するのが難しいと思うので、ルールベースが良いと思う。リーガルについては、契約書などの翻訳には専用の翻訳

ソフトがある。

内山氏：

コーパスをためてほしい。ためしに「みんなの自動翻訳@TexTra®」を使ってみてほしい。

（4）コーパスの収集について

翻訳会社の方：

統計ベースでは訳語（秋元注：訳語ではなくコーパスのことだと思われる）がたくさんある方が訳質向上するという話だったが、日本でどこかにたくさん集めているところはないのか？TAUS（Translation Automation User Society）のように、どこかが統合して集めていないのか？

井佐原氏：

欧米の大企業は TAUS に加入しているが、日本は少ない。TAUS の仕組みを使って日本企業にも参加してもらう活動を始めた。JTF ジャーナルにも書いたので見てほしい。

9. 反響

JTF が実施したアンケートの回答における本パネルディスカッションの反響は以下の通りであった。

全アンケート回答数：64

本セッションについての回答数：9

本セッションについての評価内訳：

とても参考になった	2
参考になった	5
普通	1
あまり参考にならなかった	0
全く参考にならなかった	0
無回答	1

また、自由記述では以下のような意見をいただいた。

- NICT の新しい機能を知ることができて良かったです。ポストエディターの適性の話しも参考になりました。（翻訳会社の翻訳者）
- MT の現状を知ることができました。（翻訳会社の翻訳者）

10. 反省と課題

昨年はパネリストを課題調査委員だけで担当したためテーマに沿うのが難しい部分があり、「聞きたいことが聞けなかった」等の厳しい意見をいただいた。本年はその反省を踏まえ、課題調査委員以外にも各分野の専門家にパネリストになっていただき、内容としては具体的な実例を盛り込み、会場との質疑の時間も多くとった。その結果、アンケートでは好意的な意見をいただき、来場者の方々にはおおむねご満足いただけたようである。

一方、質疑ではたくさんの方が挙手してくれたが、議論できたのが 4 件のみとなってしまったのが残念である。時間があればパネリスト同士の議論も行いたかった。また、テーマの決定にあたり無難な線でまとめたという印象も若干あり、チャレンジが足りないとも思われる。来場者が今何を聞きたいかを事前に明確にし、もう少し議題を絞って深い議論ができた方がさらに面白かったかもしれない。ただし、あまり議題を絞りすぎると、来場者が聞きたいことと合致しなくなる危険性が高くなるのも難しいところである。

今後も JTF との連携や AAMT の活動を活発化し、翻訳者や翻訳会社、翻訳ソフトを使用して事業展開する企業などの要望や疑問などを把握し、的確にこたえられるような質の高い講演を提供し、機械翻訳の利点や最新動向を伝えていきたい。

製品紹介

多言語機械翻訳エンジン Globalese 製品概要

森口功造
株式会社川村インターナショナル

はじめに

Globaleseは、機械翻訳とポストエディットを想定したプロジェクトのワークフローの最適化を目的としたMosesベースの多言語機械翻訳エンジンです。川村インターナショナルは、開発元のMorphoLogic Localisation Ltd.(ハンガリー)との間で、日本国内での独占的なリセラー契約を締結し、2014年4月から本格的に日本市場での販売を開始しています。

多機能な機械翻訳エンジン

Globaleseのインタフェースはとても使いやすく、エンジンの管理者やポストエディターからの要望を取り入れた様々な機能をもち、この機能を簡単に使用できます。用語集のアップロード機能はもちろん、機械翻訳の品質を評価する機能も実装されています。英語と日本語のコンビネーションだけではなく、約40言語の組み合わせに対応可能なのも、Globaleseの強みです。

各種CATツールともシームレスに連携ができるため、既存の翻訳ワークフローの中にも容易に導入できます。

（エンジンのステータスをわかりやすく表示）

費用対効果の高さ

翻訳のニーズを抱えた企業ユーザーのみならず、翻訳会社でも利用できる価格体系を設定しました。これにより、機械翻訳エンジンの導入をためらっていたユーザー層にもソリューションをご利用いただいています。

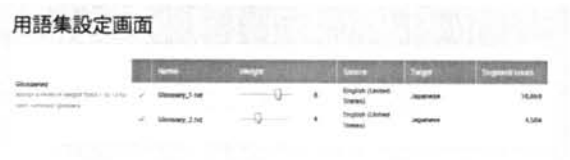
2 種類のライセンス提供モデル

Globaleseには、ユーザーのニーズに応じて選択が可能な、ライセンス契約モデルとレンタルモデルの2 種類のライセンス提供モデルがあります。
ライセンス契約モデルでは、サーバーをご用意いただく必要がありますが、対訳データの外部提供に慎重なユーザーには最適です。また、クラウド型を含む仮想サーバー上で運用すれば、使用していない時間帯はサーバーを休止しておくことで、運用費用を節約することも可能です。

一方、レンタルモデルでは、初期費用と保守料金を含めた月額料金での提供となるため、短期の利用でもご活用いただけます（最低三ヶ月の契約が必要となります）。

また、Globaleseを一ヶ月間、無償でお試しいただくことも可能です。機械翻訳を既に導入している企業も、導入を検討している企業も、まずはトライアルでその費用対効果を体験してください。

（複数の用語集に優先順位を設定して参照可能）



独自の機能

Globalese には、多くの独自機能が実装されています。これらの機能を最大限に活用することにより、機械翻訳ワークフローを効率化できます。

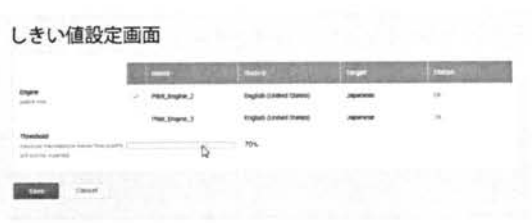
■不明用語検出機能「Unknown Words Report」

「Unknown Words Report」機能により、コーパスに含まれていない不明用語を事前に検出することができます。機械翻訳を適用する前に、不明用語だけを訳してコーパスや用語集に追加すれば、より高い品質の機械翻訳結果を得ることが可能です。

■しきい値設定機能「Threshold Value」

「Threshold Value」機能により、しきい値を調整して一定以上の品質レベルが保たれた機械翻訳（MT）のみを挿入することが可能です。ポストエディットの担当者は、低品質な機械翻訳の結果を確認する必要がなくなり、最初から翻訳を入れるだけになるため、ポストエディット全体の効率が上がります。

（低品質なアウトプットを事前に除外）



■プロジェクトレベルで品質を向上させる追加調整機能「Tuning」

翻訳対象の最初の1,000 セグメント程度を先にポストエディットし、そのデータを使って「Tuning」を実行することで、エンジンを個別のプロジェクト用に最適化し、さらに品質を上げる事が可能です。

■ポストエディット後の負荷評価機能「Evaluation」

「Evaluation」機能により、機械翻訳の訳文とポストエディット後の訳文を比較することが可能です。比較の結果は、翻訳支援（CAT）ツールのような解析結果で示され、ポストエディットにかかった負担の客観的計測値として活用することができます。

（三種類の解析結果を並べて表示）

■ポストエディット前の負荷予測機能「Quality Estimation」

「Evaluation」機能を用いて、ポストエディット後の負荷を評価できるのに対し、「Quality Estimation」機能では、事前に概算のポストエディット負荷を予測することができます。実際のポストエディット作業と完全に一致することは不可能ですが、概算値を算出することができるため顧客への見積もり提出時や、作業者への発注作業量の把握に役立ちます。

（事前に概算のポストエディット負荷を予測）

■高度なタグ処理機能

トレーニング前に対訳リソースをアップロードする際、自動的にタグが取り除かれるため、トレーニングデータのクリーニングが不要です。翻訳時のタグ処理についても、翻訳が完了すると、訳文にタグが自動的に再挿入されます。

■サポートしている翻訳支援（CAT）ツール

SDL Trados 2007（ .ttx）

SDL Trados Studio 2011/2014 (.sdlxliff)

memoQ (.mqxliff/mqxlz)

MemSource (.mxliff)

XTM (.xlf)

Wordfast (.txml)

Wordbee (XLIFF)

Standard XLIFF 1.2

■API を介して連携可能なツール

RESTful API を提供しているため、自社内のワークフローツールとの連携が可能です。また、下記の CAT ツールとは、すでに API を介して連携が可能です。これ以外の翻訳支援ツールとの連携についても API 文書を公開いただければ原則無償で連携をさせることができます。

memoQ server

MemSource Cloud

XTM Cloud

■サポートしている言語

サポートしている言語は、下記の 40 言語の組み合わせになります。

対応言語			
アイスランド語	クロアチア語	トルコ語	ポルトガル語
アフリカーンス語	スウェーデン語	ノルウェー語	マレーシア語
アラビア語	スペイン語	ハンガリー語	ラトビア語
イタリア語	スロバキア語	ヒンディー語	リトアニア語
インドネシア語	スロベニア語	フィンランド語	ルーマニア語
ウクライナ語	セルビア語	フランス語	ロシア語
エストニア語	タイ語	ブルガリア語	英語
オランダ語	チェコ語	ベトナム語	韓国語
カタロニア語	デンマーク語	ヘブライ語	中国語（簡体字）
ギリシャ語	ドイツ語	ポーランド語	日本語
			他

想定されるワークフロー

① TM（Translation memory）による翻訳を事前にバッチ処理する。

例：

75%-85%まで TM を適用

74%以下を機械翻訳

② 一定以下の品質の MT を事前に除外する。（Threshold Value 機能）

③ 機械翻訳後のポストエディット作業の負荷を予測する。（Quality Estimation 機能）

④ 翻訳支援ツール上で、ポストエディットを実施する*。

*TM と MT のそれぞれの適用箇所は、CAT ツール上で明確に区別されています。

⑤ 完了したポストエディットと機械翻訳との編集距離を計測し、解析した結果を表示する。（Quality Evaluation 機能）

⑥ Quality Evaluation 機能で算出された解析結果に基づき、顧客への請求書を発行する。

⑦ 最終成果物を、Globalese の言語モデルに対してアップロードし、定期的にトレーニングを行う。

以上

これまでの AAMT Forum メールマガジン Vol.4

機械翻訳課題調査委員会 WG1、WG2

はじめに

AAMT Forum を通じて配信したメールマガジンのバックナンバーをお届します。今回は第 29 号から第 37 号です。

今回は、新たな翻訳サービス，翻訳関連事業のニュースがありました。また、翻訳対象言語の増加に関するニュースもありました。また、特別号として長尾真先生、AAMT 初代会長の学士院会員選出、豊橋技科大と TAUS との提携のニュースも配信しました。今後もメールマガジンを通じて様々な情報を皆様に配信させていただきます。

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

では、メールマガジン第 29 号から第 37 号、そして特別号を再びお届けします。

AAMT Forum　メールマガジン 29 号（2014/09/18 配信）

こんにちは。AAMT Forum メルマガ担当です。秋風さわやかな過ごしやすい季節となりました。今年はまだ例年よりも少ないですが、台風シーズンでもありますので、空模様には気をつけましょう。では、メールマガジン第 29 号をお届けします。

■ロゴヴィスタ、ビジネス向けの英日／日英翻訳ソフト「LogoVista PRO 2015」

ロゴヴィスタは 9 月 9 日、同社製翻訳ソフト「LogoVista PRO」シリーズの最新版「LogoVista PRO 2015」シリーズを発表、10 月 3 日に販売を開始します。
http://www.itmedia.co.jp/pcuser/articles/1409/09/news099.html

■訪日外国人観光客を取り込め「ITPARK ホームページサービス」英語の自動翻訳サービス提供開始

ホームページ作成サービス「ITPARK」を運営する ITX 株式会社は、自動翻訳株式会社クロスランゲージと共同で、ホームページの英語切り替えができる自動翻訳サービスを、2014 年 9 月 1 日より無料で提供を開始しました。
http://www.itx-corp.co.jp/jp/news/pdf/2014090201.pdf

AAMT Forum　メールマガジン 30 号（2014/10/01 配信）

こんにちは。AAMT Forum メルマガ担当です。涼しさの中に寒さも感じ始めるこの頃、いかがお過ごしでしょうか。近頃は豪雨による土砂災害や噴火など自然災害のニュースが目に付きます。非常用品の準備や気象情報の確認など、日ごろから防災意識を高めておきましょう。では、メールマ

ガジン第 30 号をお届けします。

■翻訳事業会社の設立に向けた合弁契約の締結

株式会社 NTT ドコモ、SYSTRAN INTERNATIONAL Co., Ltd、株式会社フュートレックは、世界最高レベルの機械翻訳精度をもつ翻訳技術の開発及びサービス提供を行うための合弁会社「株式会社みらい翻訳」の設立について、合弁契約を締結しました。
https://www.nttdocomo.co.jp/info/news_release/2014/09/29_01. html

■日本初！本格的な PC 版アラビア語機械翻訳ソフトウェアを発売

ロゴヴィスタ株式会社が、アラビア語を日本語に翻訳する Windows PC 上で動く本格的な機械翻訳ソフトウェアを日本で初めて発売しました。
http://www.dreamnews.jp/press/0000099700/

AAMT Forum　メールマガジン
31 号（2014/10/17 配信）

こんにちは。AAMT Forum メルマガ担当です。8 月、9 月が少なかった反動でしょうか、台風接近のニュースが続くこの頃、いかがお過ごしでしょうか。日中も寒くなってくる時期ですので、その時々に合った装いで快適に過ごしましょう。では、メールマガジン第 31 号をお届けします。

■外国人観光客にリアルタイムな案内・翻訳サービス、沖縄で実証実験

株式会社琉球ネットワークサービス、株式会社レキサス、株式会社 WAKON は沖縄の外国人観光客向けに、多言語化した観光情報データベースを GPS ナビゲーション機能で案内し、かつ音声翻訳機能も提供する ICT サービスの実証実験を開始しました。

http://cloud.watch.impress.co.jp/docs/news/20141014_670883.html

■鈴鹿市（三重県）の HP、海外 5 カ国語の自動翻訳機能を付加

鈴鹿市は１０月１日から、市のホームページに自動翻訳の機能をつけ、英語や中国語、韓国語、ポルトガル語、スペイン語で読めるようにしました。
http://www.asahi.com/articles/ASG9V52N8G9VONFB010.html

AAMT Forum　メールマガジン
32 号（2014/10/29 配信）

こんにちは。AAMT Forum メルマガ担当です。清涼の秋気が身にしみるこの頃、秋の深まりを感じます。読書、スポーツなど、いろいろと楽しみの増えるこの季節、何をして過ごすかはもうお決まりでしょうか。では、メールマガジン第 32 号をお届けします。

■ミャンマー語の日・英自動翻訳システムの実用化に向けて

NICT ユニバーサルコミュニケーション研究所は、ミャンマー語を対象にした日・英自動翻訳システムを世界で初めて開発しました。
http://www.nict.go.jp/press/2014/10/15-1. html

■手話と音声による話の輪を実現する手話音声翻訳デバイス

カメラだけで手話言語を解読し、音声言語へ翻訳することで手話言語話者と音声言語話者の双方向のコミュニケーションを実現するツールUNIの紹介です。
http://nge.jp/2014/10/27/post-86936

AAMT Forum　メールマガジン
33号（2014/11/05 配信）

こんにちは。AAMT Forum メルマガ担当です。秋色いよいよ深まり、夜長の頃となりました。朝晩冷え込み、未だ寒さに慣れないこの時期、体調管理には気をつけましょう。では、メールマガジン第33号をお届けします。

■「Skype」の同時通訳機能の早期プレビュー版の受け付けを開始

米Microsoft Corporationは3日（現地時間）、「Skype Translator」の早期プレビュー版の利用申し込み受け付けを開始しました。
http://www.forest.impress.co.jp/docs/news/20141104_674357. html

■ベストセラー翻訳ソフトの最新版！「コリャ英和！一発翻訳 2015 for Mac」（DVD-ROM）シリーズを新発売

ロゴヴィスタ株式会社の最新のMac OS環境に対応した英日・日英翻訳ソフト『コリャ英和！一発翻訳 2015 for Mac』が、11月28日（金）から販売されます。
http://www.dreamnews.jp/press/0000102000/

AAMT Forum　メールマガジン
34号（2014/11/21 配信）

こんにちは。AAMT Forum メルマガ担当です。しばしば寒さに身を震わるこの頃、いかがお過ごしでしょうか。コタツ、ストーブなど、既に暖房器具のお世話になっている方もいるかと思います。暖かい部屋は快適ですが、換気も忘れずにしましょう。では、メールマガジン第34号をお届けします。

■「はなして翻訳」の海外向けサービス「Jspeak」を提供開始

株式会社NTTドコモは、AndroidTM搭載のスマートフォンやタブレットを利用して、日本語と外国語の間での会話を可能とする翻訳サービス「はなして翻訳 R」の海外向けサービス「JspeakTM」の提供を開始しました。
https://www.nttdocomo.co.jp/info/news_release/2014/11/06_00.html

■中国・韓国語の特許文献を日本語で検索可能なシステムの試行版が利用可能になります

「中韓文献翻訳・検索システム」の試行版の提供が11月13日に開始されました。本格稼動は平成27年1月の予定です。
http://www.meti.go.jp/press/2014/11/20141112003/20141112003.html

AAMT Forum　メールマガジン
35号（2014/12/05 配信）

こんにちは。AAMT Forum メルマガ担当です。師走となり、今年も残すところあとわずかになりました。なにかと忙しくなる年の瀬ですが、体調管理にも気をつけましょう。では、メールマガジン第35号をお届けします。

■リコー、複合機でスキャンして文書翻訳可能な「RICOH ドキュメント翻訳サービス」開始

リコーはリコー複合機とクラウドを連携した機械翻訳サービス「RICOH ドキュメント翻訳サービス」を開始しました。対応している言語は、日本語・英語・中国語・韓国語・ドイツ語・フランス語・イタリア語・スペイン語・ポルトガル語になります。
http://www.sbbit.jp/article/cont1/28845

■モリサワ 電子情報配信サービス「MCCatalog+」をリニューアル

株式会社モリサワは、電子情報配信サービス「MCCatalog+」を多言語対応電子配信ツールとしてリニューアルし、2015年2月1日より販売開始します。
http://www.sankei.com/economy/news/141201/prl1412010140-n1.html

AAMT Forum メールマガジン
36号（2014/12/17 配信）

こんにちは。AAMT Forum メルマガ担当です。全国的に冷え込むこの頃、いかがお過ごしでしょうか。こう寒いと動きたくなくなってしまいますが、今年も残りあと僅かです。気持ちよく新年を迎えるため、やり残したことはしっかりと済ませておきましょう。では、メールマガジン第36号をお届けします。

■自分の全く知らない言語で会話が可能に!?自動通訳「Skype Translator」開始

米 Microsoft 傘下の Skype Communications SARL は15日、音声通話の自動通訳技術「Skype Translator」のプレビュー提供を開始したと発表しました。音声通訳はまず英語とスペイン語からになります。
http://internet.watch.impress.co.jp/docs/news/20141216_680449. html

■Google 翻訳に新たに10の言語が加わる、ビルマ語やインドのローカル言語など

中でもビルマ語はミャンマー（ビルマ）の公用語であり、マラヤーラム語はインドの6つの古典言語の一つで、3800万人が使っています。
http://jp.techcrunch.com/2014/12/12/20141211google-translate-adds-10-more-languages/

■「はなして翻訳－Jspeak」のキャリアフリー対応について

日本語と外国語の間での会話を可能とする対面型翻訳サービス「はなして翻訳－Jspeak」が、15日からキャリアフリー対応となりました。
https://www.nttdocomo.co.jp/info/news_release/2014/12/11_00.html

AAMT Forum メールマガジン
37号（2015/01/09 配信）

こんにちは。AAMT Forum メルマガ担当です。新年あけましておめでとうございます。この新しい年がより佳き年になるよう心より祈念いたします。では、メールマガジン第37号をお届けします。

■特許庁、中韓文献翻訳・検索システム本格版の提供開始

特許庁は1月5日、中国・韓国語の特許文献を日本語で検索可能なシステムである「中韓文献翻訳・検索システム」の本格版を、同日から提供すると発表しました。
http://news.braina.com/2015/0106/rule_20150106_002____.html

■外国人おもてなしに一役　２７言語を自動翻訳　板橋区でアプリ説明会

板橋区は12月24日、スマートフォンや携帯型端末用の多言語の音声を自動翻訳できる専用アプリ（VoiceTra4U）の説明会を開き、区幹部職員ら百人が使い方を学びました。
http://www.tokyo-np.co.jp/article/tokyo/20141225/CK2014122502000123.html

---

AAMT Forum　メールマガジン
特別号その１（2014/12/06 配信）

****
長尾先生おめでとうございます。
****
長尾真先生が学士院会員に選出

日本学士院は12月12日、情報処理の研究、電子図書館の構築への功績から、初代AAMT会長である長尾真先生を新会員として選出しました。学士院会員とは学術上顕著な功績のある科学者から選ばれる国家公務員です。
http://www.japan-acad.go.jp/japanese/news/2014/121201.html
AAMTからもお祝い申し上げます。

AAMT Forum　メールマガジン
特別号その２（2014/12/06 配信）

****
豊橋技術科学大学、TAUS*と多言語情報発信コンソーシアムの構築に向けて提携！！
****

国立大学法人豊橋技術科学大学は多言語翻訳研究本部を設置し、「学内研究基盤強化による産業分野向け高度機械翻訳技術の研究開発」プロジェクトを行ってきました。このプロジェクトの一環として、豊橋技術科学大学はTAUSと提携関係を結びました。この提携関係により、多言語化したい文書を持つユーザ企業と翻訳プロセスに関わる企業とを結ぶ多言語情報発信コンソーシアムの実現へ大きな一歩を踏み出したと言えます。また、来年4月7日に東京で関連シンポジウムが開催される予定です。
*TAUS（https://www.taus.net/）
https://www.taus.net/think-tank/news/press-release/taus-and-toyohashi-university-of-technology-partner-to-support-japanese-industry-with-machine-translation

---

委員会活動報告

実務翻訳における機械翻訳の利用に関するアンケート結果報告

AAMT 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 WG2（調査・広報・啓蒙）では、産業翻訳業界における機械翻訳システム活用の実態を把握することを目的として、2012 年以降、JTF 翻訳祭においてアンケートを実施してきた。JTF 翻訳祭は、翻訳者、翻訳会社、翻訳発注企業などの翻訳業界団体である日本翻訳連盟（JTF）が主催するイベントであり、講演会、展示会により構成される。

本稿はそのアンケートによる調査結果を報告する。

2. アンケートの概要

- 調査実施日

  第 24 回 JTF 翻訳祭当日（2014 年 11 月 26 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。

  翻訳祭の展示コーナーに AAMT がブースを確保し、ブースの前を通りかかる人に AAMT の委員から直接協力を依頼した。AAMT ブースにアンケート用紙と筆記用具を用意しておき、その場でアンケートに回答いただき、回収した。なお、回答者に抽選で当たる景品として、500 円分クオカード 10 枚と、翻訳ソフト「PC-Transer」、「J 北京 7 プレミアム 3」（株式会社クロスランゲージ様、株式会社高電社様のご厚意による供出）を用意した。

- 回答状況

  回答者数は、62 名であった。2012 年度が 60 名、2013 年度が 64 名であり、翻訳祭参加者（884 名）全体に対する回答率はいずれの年もおよそ 8%程度である。

3. 回答者

- 回答者の職種

  翻訳業界の従事者といっても様々な仕事がある。以下は回答者 62 名の職種の内訳である。

  [1] フリーランスの翻訳者　18 人

  [2] 翻訳会社の社内翻訳者　6 人

  [3] 企業（翻訳会社以外）・官公庁等の翻訳者　11 人

  [4] 翻訳会社の翻訳業務管理者、コーディネーター　15 人

  [5]企業（翻訳会社以外）・官公庁等の翻訳業務管理者、コーディネーター　4 人

  [6]その他・無回答　8 人

  回答者 62 名のうち 35 名（56%）が翻訳を生業としている人であり、20 名（31%）が翻訳の管理を仕事にしている人であった。

- 性別・年齢

  年齢は 10 代から 70 代まで広く分布しているが、30 代と 40 代が半数以上（20 人、18 人）であった。以下、10 代（2 人）、20 代（10 人）、50 代（7 人）、60 代（3 人）、70 代（1 人）と続く（無回答含む）。

  性別は女性 38 人に対して男性 19 人であり、女性の回答者の割合が多かった（無回答含む）。

4. 調査結果

本節では、回答者のうち翻訳作業そのものを生業とする 35 人のみを対象とした集計結果を紹介する。なお、内訳の合計が 35 人に満たない項目には無回答が含まれる。

4.1. 機械翻訳の利用頻度

翻訳者が普段どのくらいの頻度で機械翻訳を使っているかについて、翻訳者のタイプ別に集計した結果を表 1 に示す。

表 1　機械翻訳の利用頻度

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] ほぼ毎日	5	3	4
[2] 週 1～3 回	6	1	2
[3] 月に数回	2	1	1
[4] 年に数回	2	0	2
[5] 使っていない	0	0	1

質問：翻訳ソフト/サイトの利用頻度は？

翻訳者の 6 割以上（21 人）の人が機械翻訳を日常的に使っていると回答している。「使っていない」と答えた人は 1 人にすぎず、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。

4.2.　利用している機械翻訳の種類

機械翻訳には、インターネット上で無料で使えるサービスから、一本が 10 万円以上する高価なソフトまで、様々な形態で提供されている。翻訳者は業務においてどのようなソフトあるいはサービスを使っているのだろうか。表 2 は翻訳者が良く利用する機械翻訳の種類についてまとめたものである。

表 2　よく利用する機械翻訳の種類

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] 無料の翻訳サイト	8	5	4
[2] 有料の翻訳サイト	1	0	1
[3] 自社または翻訳会社などで提供される翻訳システム（サービス）	2	1	1
[4] 市販の PC 翻訳ソフト	3	1	1

質問：現在最もよく使う翻訳ソフトや翻訳サイトは？

翻訳者がよく使っている機械翻訳のトップは、無料の翻訳サイトという結果となった。会社から翻訳ソフトが提供される例は少数に留まっていることから、翻訳者が個人的に利用するケースが大半を占めるようである。翻訳者をタイプ別に見ると、フリーランスの翻訳者は有料の翻訳サイトや市販のソフトも利用しているのに対し、その他の翻訳者は無料の翻訳サイトを中心に利用している。

4.3.　機械翻訳への期待

では、翻訳者たちは機械翻訳にどのくらいの品質を期待しているのだろうか。表 3 に「翻訳サイトにどのような品質を求めますか？」という質問の回答をまとめる。

表 3　機械翻訳が期待される品質

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] 辞書代わり	5	3	4
[2] おおよその意味が分かること	5	2	1
[3] 翻訳の下訳として使えること	5	1	5

質問：翻訳ソフト/サイトにどのような品質を求めますか？

機械翻訳を翻訳の下訳として使いたいという翻訳者の期待が読みとれる一方で、機械翻訳が辞書引きレベルで使えれば十分と考えている翻訳者もかなりの割合でいることがわかる。

4.4.　機械翻訳の満足度

表 4 は機械翻訳の満足度を 5 段階で評価した結果をまとめたものである。満足と回答した人は 6 人（17%）だった。満足とも不満ともいえない人が 16 人（46%）であり、現在の翻訳システムでも使い方しだいで翻訳者の役に立つ場面があるようである。

表 4　機械翻訳全般の満足度

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] とても不満	0	0	1
[2] やや不満	4	1	2
[3] どちらでもない	6	3	7
[4] やや満足	3	1	0
[5] とても満足	2	0	0

質問：翻訳ソフト/サイトの満足度は？（5 段階で評価）

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」を別々に満足度を調べており、以下ではその結果について紹介する。

● 翻訳品質

表５の数字を見ると、翻訳品質に満足していると答えた人（4 人）の割合は不満であると答えた人（17 人）の四分の一程度であり、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。

表５　翻訳品質の満足度

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] とても不満	1	0	3
[2] やや不満	6	3	4
[3] どちらでもない	6	1	3
[4] やや満足	3	1	0
[5] とても満足	0	0	0

質問：翻訳ソフト/サイトの品質の満足度は？（5 段階で評価）

● 翻訳機能

表６から、機械翻訳は機能としては「満足」と答えた人（8 人）と翻訳品質の二倍である。したがって、機能については及第点といえそうである。

表６　翻訳機能の満足度

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] とても不満	0	1	0
[2] やや不満	4	1	5
[3] どちらでもない	5	2	3
[4] やや満足	5	1	1
[5] とても満足	1	0	0

質問：翻訳ソフト/サイトの機能の満足度は？（5 段階で評価）

4.5.　翻訳メモリの利用状況

翻訳の現場では、機械翻訳よりも先に翻訳メモリの活用が進んでいると考えられる。表７は翻訳者による翻訳メモリの利用状況である。

翻訳メモリを使っている人が 18 人（51%）と、半数が使っている。また、翻訳メモリと機械翻訳を比較した場合、翻訳メモリを主として使う傾向がある。このことから、翻訳メモリの活用が通常の翻訳業務プロセスに取り込まれていることが考えられる。

表７　翻訳メモリの利用状況

	ﾌﾘｰﾗﾝｽ	翻訳会社	企業・官公庁
[1] 知らない	1	1	0
[2] 使っていない	8	2	3
[3] 主として翻訳メモリ、補助的に機械翻訳を使っている	6	1	5
[4] 主として機械翻訳、補助的に翻訳メモリを使っている	3	1	2

質問：翻訳メモリを使っていますか？

表８は、機械翻訳の種類と翻訳メモリの利用状況をクロス集計した結果である。この結果から、二つの翻訳ツールの利用の相関関係をみることができる。

表８　翻訳メモリ利用状況と利用中の機械翻訳の種類

	未使用	主として使用	補助的に使用
[1] 無料の翻訳サイト	10	5	1
[2] 有料の翻訳サイト	1	1	0
[3] 自社または翻訳会社などで提供される翻訳システム（サービス）	0	2	1
[4] 市販の PC 翻訳ソフト	1	1	3
[5] 業務では使っていない	4	3	0

翻訳メモリを使っていない翻訳者の多く（63%）は無料の翻訳サイトを利用する傾向にある。また、翻訳メモリと機械翻訳のどちらも使っていない翻訳者は 4 人（11%）のみで、14 人（40%）は両方のツールを使っていると回答している。

4.6.　機械翻訳を利用するうえでの翻訳者の工夫

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？」という設問を設け、回答は自由に文書を記述してもらう形式とした。18 人（51%）から回答があった。

18 人中 9 人が、用語集の整備の重要性を指摘していた。そして、6 人が翻訳メモリなどと並行しての使用を

挙げていた。この２つが群を抜いて数の多い回答だった。このほかの回答としては、

・「見直しをよくする」

・「単語訳のみをする」

というように機械翻訳の使い方を指摘するものがあった。

4.7. 用語集の利用状況

先の機械翻訳の使い方に関する工夫でも指摘されたように、用語集は翻訳業務において重要なツールと位置づけられる。表９は翻訳者による用語集の利用状況である。

表９　用語集の利用状況

	フリーランス	翻訳会社	企業・官公庁
[1] 会社作成の用語集の利用	4	4	5
[2] 個人作成の用語集の利用	7	0	1
[3] いずれも利用	4	1	3
[4] 利用していない	3	0	1

質問：翻訳業務で用語集を作成していますか？

用語集を使っていない人が４人（11%）であり、用語集の利用頻度は高い。また、用語集の利用者 29 人のうち、16 人（56%）が用語集を個人で作成している。このことから翻訳業務の効率化には用語集が重要であることがうかがえる。

AAMT ではさまざまな翻訳ツールで利用可能な用語集形式の共通規格（UTX (Universal Terminology eXchange)）を作成している。表 10 はこの UTX がどの程度知られているかを示す。

UTX を知っている人が 21 人（60%）であったが、その中身までを知っている人は 13 人（37%）であった。今後、UTX の広報、普及活動をより一層進める必要性がうかがえる。特に、今回の調査結果（4.3 節）からも明らかなように「単語訳のみ」を求める翻訳者が多いため、UTX を利用すると単語訳の適正度が高まることなどを広報すべきと考えられる。

表 10　UTX の認知度

	フリーランス	翻訳会社	企業・官公庁
[1] 名称、使用法などを知っている	4	4	5
[2] 名称のみ知っている	7	0	1
[3] 知らない	4	1	3

質問：用語集形式 UTX のことを知っていますか？

5. まとめ

本稿では 2014 年度の第 24 回 JTF 翻訳祭で実施したアンケート結果を中心に、実務翻訳における機械翻訳の利用実態についてまとめた。この調査を通じて、機械翻訳のユーザーである翻訳者の方々にとって、現在の機械翻訳の何が問題点であり、将来、どのような機械翻訳が必要とされているかを知ることができるため、このようなアンケート調査は継続すべきと考えている。

機械翻訳は翻訳の品質などにおいて多くの問題が山積するが、実務翻訳の現場で日常的に機械翻訳が活用されていることが判明した。しかしながら、利用される機械翻訳の多くはインターネットで無償提供されている翻訳サイトであり、翻訳ベンダが長年取り組んできた翻訳パッケージの利用はまだまだ浸透していないのが現実である。

用語集の整備と翻訳メモリの活用については、すでに実践して効果をあげている翻訳者もいるようである。昨年、機械翻訳を用語集や翻訳メモリと連携させて効果を引き出すことで、実務翻訳における機械翻訳普及の鍵になると考えた。そこで今年度は用語集の利用状況や用語集形式についての調査も行った。用語集の利用は高いが、UTX の認知度をさらに高める必要があることが分かった。

最後に、この調査を通じて初めて「AAMT」という団体を知ったという方もいたため、AAMT 自体の知名度を高める必要があることが分かった。AAMT 機械翻訳課題調査委員会では、今後も日本翻訳連盟との連携を深めてイベント共催などを進めながら、実務翻訳者の実態調査を継続するとともに、機械翻訳の効率的な活用方について翻訳者に向けた情報発信を行っていきたいと考えている。

AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第１５回を迎えることができました。今号では、個人会員１名からのご寄稿をいただいております。

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

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

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

個人会員（敬称略・50 音順）

会員名

Tony Hartley

Emeritus Professor of Translation Studies, University of Leeds, UK

Professor of Translation, Rikkyo University, Japan (from April 2015)

自己紹介

Born in London, I escaped to Manchester for my university studies and in 1970 graduated from Salford University with a degree in French and Russian and one season of playing rugby in France under my belt. That took me to France for another three years, playing more rugby while teaching translation and linguistics at universities in Montpellier, Besançon and, finally, Ecole Normale Supérieure de Saint Cloud, in Paris. My academic itinerary has meandered via Bradford, Sussex (where I also acquired an MSc in Cognitive Science) and Brighton to Leeds in UK, with lengthy detours to Quebec and Sydney. What was to have been a one-year sabbatical at the University of Tokyo, starting September 2009, has turned into a protracted stay as research and teaching opportunities opened first at Toyohashi University of Technology, then Tokyo University of Foreign Studies and, on the near horizon, Rikkyo. Japan is a final destination I am happy with – the air is fresh with ideas, enthusiasm and commitment. And what's more, it boasts the world's largest community of active veteran rugby players. So I am grateful to Tokyo Fuwaku Rugby Club for allowing me to resurrect a rugby career which, I hope, will outrun the academic one.

MT/翻訳とのかかわり

ＭＴおよび翻訳業界に期待すること

My first close encounter with MT was in 1985, on introducing the Weidner MicroCAT system (EN/DE/ES/FR) into the Masters translation curriculum at Bradford, a UK first. In 1995 I was using SDL`s TM product in training translators at Brighton. The MicroCAT experience led me, even then, to some very successful results with controlled authoring, MT and post-editing as a consultant for Perkins Engines. And now SDL is a global brand with the same aura as Microsoft or Google. Over the past 30 years we seem to have seen many false dawns, but now, again, could we be hovering on the verge of a breakthrough, with the massive adoption of MT technologies by the biggest Language Service Providers?

It is, for me, a paradox that Japan, the country which has made probably the most sustained and innovative contribution to MT research, has been so cautious in adopting these technologies in its translation business and in training. There must be few translation departments in European universities that do not offer some exposure to TM, MT and their underlying principles. MA students at Leeds, for example, routinely use Systran and about seven of the leading TM and software localization applications. With the current draft standard ISO17100 on requirements for translation services poised to come into force in 2015, universities and translation providers in Japan can hardly afford to neglect this opportunity, if it is not to become a threat to their competitiveness.

Another opportunity that excites me is Japan`s hosting of the Rugby World Cup in 2019, just one year before the Olympics. Surely there are prospects there for record-breaking advances in the deployment of Japanese MT ...

ＡＡＭＴへの要望

It`s a privilege to be admitted to AAMT. My hope is that the researchers, developers, translation vendors and educators who make up the Association can join forces and rise to the challenge of ISO17100. Certainly, I look forward to discussing these issues with you all, and to welcoming many of you into the lecture halls and labs of Rikkyo`s translation program to share your expertise with tomorrow`s students.

協会活動報告

(2014 年 9 月～2014 年 12 月)

機械翻訳課題調査委員会

2014 年 9 月 5 日（2014 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）

(WG1,2)

・評価サイトに関する検討

規約、公開スケージュールなどの検討

公開する内容や形式などの検討

(WG3)

・JTF 翻訳祭配布のパンフレットなどの検討

・UTX 広報

・人工知能研究振興財団　研究助成の応募の検討活動内容の報告

2014 年 10 月 2 日（2014 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,WG2)

・日中テストメタ評価に関する検討

日中テスト評価の分析、

人手評価の検討、

次週の日中会議にて成果発表に関する検討など

・翻訳祭準備について

アンケートと展示内容の原稿準備などについて

・サイト公開について

規約の確認（終了）

サイトの名称決定：AAMT 機械翻訳評価サイト

(WG3)

・UTX コンバータ

開発にかかわっている委員から現状と今後の計画について報告

・人工知能研究振興財団　研究助成の応募の件

申請内容（使途など）について、委員より説明。

採択結果の発表は 11 月下旬の予定。

・学会発表の検討

第 3 回特許情報シンポジウム(2014/11/28)への発表を検討。

・UTX 論点シートの検討

本件に関する議論の仕方、スケージュール管理などについて検討。

3. 全体会議

・中岩会長の報告

多言語情報発信支援コンソーシアム

（井佐原先生（豊橋技術科学大学）主導）

グローバルコミュニケーション計画の紹介

（多言語音声翻訳システムなどを主とする）

JTF 祭で配布のちらしなどについて

4. まとめと次回委員会について

2014 年 11 月 6 日（2014 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）

(WG3)

・ISO/TC 37　2015 年の松江会議参加の検討

・学会発表の検討

言語処理学会第 21 回年次大会（2015 年 3 月 16 日～20 日，京都大学）の参加についての検討

・UTX 形式辞書の作成に関する検討

「用語集パートナー募集」の記事原稿（山本委員）の検討

作成する辞書の詳細の検討など

・UTX 論点シートの検討

多言語用語集に関する用語ステータスの検討など

3. まとめと次回委員会について

2014 年 12 月 12 日（2014 年度　第 8 回）

1. 前回委員会の議事録の確認
2. 全体会議

・JTF 翻訳祭に関して

アンケート結果速報の報告、反省点の検討など

・AAMT の知名度向上策、会員増強策の検討

・AAMT 関連の今後の主なイベントについて

（2015/1 の JTF 主催のワークショップ、

次回およびその次の MT サミット、AMTA からの要望）

3. 各 WG の活動について（各 WG に分かれて議論）

(WG3)

・言語処理学会での発表の検討

・UTX 論点シート検討

4. まとめと次回委員会について

インターネット WG

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

編集委員会

2014 年 10 月 7 日（2014 年度　第 3 回）

1. AAMT Journal No. 58 の記事の執筆依頼候補者の検討

AAMT/Japio 特許翻訳研究会

2014 年 9 月 5 日(金)(2014 年度　第 3 回)

1. 議事録の確認
2. WMT　2014 報告　越前谷先生からのご発表
3. WMT　2014 報告　磯崎先生からのご発表
4. シンポジウム関連
5. その他
6. 次回の開催について
   - 開催の日時（10 月 24 日（金曜日）18：00~20：00）
   - 主な議題

2014 年 10 月 24 日(金)(2014 年度　第 4 回)

1. 議事録の確認
2. 「日中パテントファミリーから抽出した対訳文を用いた専門用語の訳語推定」
   宇津呂生からのご発表（董さん発表）
3. COLING　2014 報告　綱川先生からのご発表
4. シンポジウム関連
5. その他
6. 次回の開催について
   - 開催の日時（12 月 12 日（金曜日）18：00~20：00）
   - 主な議題

2014 年 12 月 12 日(金)(2014 年度　第 5 回)

1. 議事録の確認
2. 二宮先生のご発表
   リサンプリングを用いた統計機械翻訳のドメイン適用
3. AMTA2014 参加報告（須藤委員）
4. MT サミット 2015 の Call for Workshop Proposals への対応
   第 6 回特許翻訳ワークショップを提案するかどうか
5. 26 年度 AAMT/Japio 特許翻訳研究会報告書
   作成スケジュールなど

6. 27年度委員会の体制
   - 副委員長（横山→宇津呂）
   - 拡大評価部会長（江原→磯崎）
   - その他、委員会メンバー、拡大評価部会メンバーの補充と交替
7. その他
8. 次回の開催について
   - 開催の日時（1月12日（金曜日）18：00~20：00）
   - 主な議題

AAMT/Japio特許翻訳研究会　拡大評価部会

2014年10月24日(金)(2014年度　第2回)

1. 前回議事録の確認
2. 報告書内容概要(各グループより)
3. その他

第 3 回特許情報シンポジウム

2014 年 11 月 28 日(金)

開会挨拶

辻井潤一（AAMT/Japio 特許翻訳研究会委員長，
マイクロソフトリサーチアジア首席研究員，東京大学名誉教授）

セッション１（招待講演）

・招待講演 1「特許庁における機械翻訳への取組」
櫻井健太（特許庁）
・招待講演 2「多言語機械翻訳の研究開発動向」
隅田英一郎（情報通信研究機構）
・招待講演 3「特許情報検索サービスにおける機械翻訳の活用」
早川浩平（日本パテントデータサービス）

セッション２（研究会報告・特別講演）

・研究会報告 1「パテントファミリーからの専門用語対訳辞書の構築」
宇津呂武仁（筑波大学）
・研究会報告 2「自動評価法を用いた機械翻訳の定量的評価」
越前谷博（北海学園大学），磯崎秀樹（岡山県立大学）
・特別講演「アジア言語を中心とした機械翻訳研究」
中澤敏明（科学技術振興機構／京都大学）

セッション３（一般発表(1)）

・統計的訳語選択技術による韓日機械翻訳の高精度化
田中浩之，園尾聡，木下聡，釜谷聡史（東芝研究開発センター）
・特許事務所における機械翻訳と人手による翻訳の Mix 事例
正林真之，杉浦伸夫（正林国際特許商標事務所）
・インターネットから利用できる翻訳ソフトを優れた辞書として活用する方法
吉川潔（翻訳業）

セッション４（一般発表(2)）

・F タームに基づいたオントロジーの構築
福田悟志，難波英嗣，竹澤寿幸（広島市立大学），乾孝司（筑波大学），
岩山真（日立製作所），橋田浩一（東京大学），藤井敦（東京工業大学）
・特許文書からの化学物質情報の抽出
池田紀子，田中一成（富士通研究所）
・特許情報分析のためのマイニング手法と分析ツール Patent Mining eXpress
岩本圭介（NTT データ数理システム）

閉会挨拶

守屋敏道（日本特許情報機構専務理事特許情報研究所所長）

・懇親会

編集後記

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

今号の巻頭言は、NTT コミュニケーション科学基礎研究所 永田昌明様より、御寄稿を頂きました。

また、近年の機械翻訳技術動向に関する解説として、情報通信研究機構 渡辺太郎様より、「統計的機械翻訳とは?」の御寄稿を頂きました。

一方、今号に先立ちまして、昨年 11 月 26 日(水)に、AAMT が後援をします第 24 回 JTF 翻訳祭が開催され、AAMT 機械翻訳課題調査委員会では、パネルディスカッション「いまさら聞けない機械翻訳～動作原理から上手な活用法まで～」を開催しました。同委員会 秋元圭様には、このセッションの開催報告を御寄稿頂きました。

また、昨年 11 月 28 日(金)には、AAMT/Japio 特許翻訳研究会の主催で、第 3 回特許情報シンポジウムが開催されました。同委員会副委員長・静岡大学 梶博行先生には、同シンポジウム開催報告を御寄稿頂きました。

さらに、昨年 10 月 4 日(土)に開催されました、第 1 回アジア翻訳ワークショップの報告を、科学技術振興機構 中澤敏明様、情報通信研究機構 美野秀弥様、隅田英一郎様、NHK 後藤功雄様、京都大学 黒橋禎夫先生より御寄稿頂きました。

その他、ヤフー株式会社 野口正樹様、中島泰様、小林隼人様より、昨年 8 月にアイルランドのダブリンで開催されました COLING2014(計算言語学国際会議 2014)の参加報告を御寄稿頂きました。

新たな製品の紹介として、株式会社川村インターナショナル様より、多言語機械翻訳エンジン Globalese の紹介を御寄稿いただきました。

AAMT 内の活動報告として、機械翻訳課題調査委員会から、「実務翻訳における機械翻訳の利用に関するアンケート結果報告」、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、1 件の紹介文を掲載しました。

AAMT

Asia-Pacific Association for Machine Translation

AAMT 入会のご案内

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

＊＊AAMT 会員の特典＊＊

1.AAMT Journal の購読ができます。

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

2.機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機会翻訳関連の情報をお届けします。
MT 新製品、新サービスの紹介、国際会議、シンポジウムのお知らせ、WEB での MT 関連記事の紹介など盛りだくさんです。

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

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

4.関連機関の主催する国際会議に参加できます。

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

AMTA* ：Association for Machine Translation in the Americas
EAMT** ：European Association for Machine Translation

年会費は以下の通りです。

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

ご関心のある方は、事務局までお問い合わせください。

アジア太平洋機械翻訳協会（AAMT）
ホームページ：http://www.aamt.info
電子メール：aamt-info@aamt.info

AAMT

Asia-Pacific Association for Machine Translation

入会申込書

以下の通り、アジア太平洋機械翻訳協会の会員申し込みを致します。

申込日　２０　　年　　月　　日

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

コメント

アジア太平洋機械翻訳協会(AAMT)

総会のご案内

アジア太平洋機械翻訳協会
会 長　中岩　浩巳

アジア太平洋機械翻訳協会では、総会および報告会・懇親会などの関連行事を下記の通り開催いたします。

会員の皆様には、以下の日時を予定してぜひご参加いただければ幸いです。

また関連行事は会員・非会員の方々のいずれも参加できます。
ぜひ多数ご参加くださいますよう、ご案内申し上げます。
プログラムなど詳細につきましては、後日ＭＬおよび書面にてご案内いたします。

****

1. 日　時　2015 年 6 月 16 日（火）　午後

2. 会　場:ホテルアジュール竹芝

アクセス:　http://www.hotel-azur.com

2.1)総会および関連行事

同ホテル内 16 階　曙、藤、橘

2.2)懇親会

同ホテル内 21 階フレンチレストラン『ベイサイド』

※懇親会会場は今年も東京湾の夜景が見える最上階です。

AAMT

AAMTジャーナル No.58

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
aamt-info@aamt.info