ISSN 1883-1818

AAMT Journal

Asia-Pacific Association for Machine Translation

October 2017 No.66

アジア太平洋機械翻訳協会

目　次

巻頭言：　Will MT blow for full time on ノーサイド (no side)?……Anthony Hartley……1

招待講演：　ニューラル機械翻訳ってどうなの？……中澤　敏明……4

招待講演：　機械翻訳収益化の可能性……佐藤　弦……9

AAMT 長尾賞：　翻訳業界における「ブルー・オーシャン戦略」……湊　幹……11

AAMT 長尾賞学生奨励賞：　目的言語側の依存構文による統計的機械翻訳の改善……ジョン・リチャードソン……15

AAMT-MT フェア 2017 展示報告：翻訳バンク始めました！……内山　将夫……19

AAMT-MT フェア 2017 展示報告：医療現場向けハンズフリー多言語音声翻訳システムのご紹介……富士通株式会社……21

AAMT-MT フェア 2017 展示報告：「みんなの翻訳実習」プロジェクト翻訳時代の翻訳教育支援統合プラットフォーム
影浦　峡, 藤田　篤, 内山　将夫, Anthony Hartley, 山田　優, 阿辺川　武, Martin Thomas……22

AAMT-MT フェア 2017 展示報告：SMT を用いた NMT の改良……江原　暉将……24

AAMT-MT フェア 2017 展示報告：スマホやパソコンを用いる地域資料の多国語翻訳と通訳……吉川　潔……27

イベント報告：　2017 年度 JTF 総会「基調講演」報告……目次　由美子……29

イベント報告：　2017 年度 JAT 総会「基調講演」報告　パネルディスカッション
『Is there an elephant in translation? 機械翻訳、まだ見かけていませんか？』..冨永　陽……32

セミナー報告：　2017 年度 JTF 第 1 回翻訳支援ツール説明会 報告……目次　由美子……36

会員の広場（個人会員）……加島　広基……39

シンポジウム案内：　「第 27 回 JTF 翻訳祭」のお知らせ……高橋　聡……42

事務局からのお知らせ：　第 27 回通常総会および関連行事の報告……AAMT 事務局……44

事務局からのお知らせ：　協会活動報告（2017 年 5 月～2017 年 8 月）……AAMT 事務局……52

編集後記……宇津呂　武仁……57

CONTENT

Foreword:　Will MT blow for full time on ノーサイド (no side)?……*A. Hartley*……1

General Meeting:　How is the Neural Machine Translation?……*T. Nakazawa*……4

General Meeting:　Monetization of Machine Translation……*G. Sato*……9

AAMT Nagao Award:　Blue Ocean Stragtegy in Translation Industry……*T. Minato*……11

AAMT Nagao Student Award:　Improving Statistical Machine Translation with Target-Side Dependency Syntax……*J. Richardson*……15

Exhibition Report:　Translation Bank Begins!……*M. Utiyama*……19

Exhibition Report:　An Introduction to Hands-free Speech Translation Terminal for Medical Fields……*Fujitsu Limited*……21

Exhibition Report:　Minnano Hon'yaku for Translator Training: An Integrated Platform
K. Kageura, A. Fujita, M.Utiyama, A. Hartley, M. Yamada, T. Abekawa, M. Thomas……22

Exhibition Report:　Improvement of NMT by SMT……*T. Ehara*……24

Exhibition Report:　Multilingual Translation and Interpretation of Regional Information
of Using Smart Phone and Personal Computer……*K. Kikkawa*……27

Research Meeting Report:　Report on "Keynote Speech" at JTF General Meeting 2017……*Y. Metsugi*……29

Research Meeting Report:　Report on "Keynote Speech" at JAT's Ordinary General Meeting 2017- Panel discussion:
Is there an elephant in translation?……*A. Tominaga*……32

Research Meeting Report:　Report on First JTF Translation Support Tools Presentation Meeting 2017……*Y. Metsugi*……36

AAMT Members:……*H. Kashima*……39

Symposium Information:　Information for 27th JTF TRANSLATION FESTIVAL TOKYO 2017……*A. Takahashi*……42

AAMT Activities:　General Meeting……44

AAMT Activities:　AAMT Activities (from May 2017 to August 2017)……52

Editor's Note:……*T. Utsuro*……57

巻頭言

Will MT blow for full time on ノーサイド (no side)?

Anthony Hartley
Rikkyo University, College of Intercultural Communication

1. Sport – a universal code?

If you get the metaphor in the title, then you have probably devoted some of your leisure time to that 'hooligan's game played by gentlemen' (and, increasingly, women) called rugby football. If not, then I hope you will find by the end of this rather non-technical piece an explanation that hints at a wider influence of MT on cultural exchange.

Japan is investing significant effort into MT – well documented in the pages of this journal – in preparation for the 2020 Olympics and Paralympics. Quite rightly, priority is being given to ensuring the personal safety, hassle-free travel and enjoyable tourism of visitors from overseas who speak little or no Japanese.

But what of exchanges about the sporting disciplines themselves? After all, many spectators come especially for the gymnastics, the track and field, the swimming … Can we assume a lingua franca of sport that allows fan to talk freely with fan about a shared passion? Given that Japanese has assimilated 'ホームラン' and 'スラムダンク', while other languages have done the same with the judo terms 'ippon' and 'uki goshi', can we conclude that knowledge of the technical terms alone is enough to enable meaningful conversation?

If my experience in the linguistic field of rugby is a reliable guide, then the answer is 'no'. Just ensuring that 'ラック' translates as 'ruck' and 'モール' as 'maul' is not enough. Let me outline the reasons for this with some supporting evidence.

2. Why rugby?

In translation, the more you know about the subject domain of the texts you deal with, the better; I played my first game of rugby at age 11 and continue to play within a veterans club in Tokyo. So for me it's a familiar topic. This brings us to a second reason – Japan has more registered rugby clubs and players than any other country in the world except England. Within this, it boasts an active veterans' scene with hundreds if not thousands of players from their 40s into their 80s playing on a regular basis throughout the year. So there is a sizeable pool of knowledgeable players and ex-players.

The popular awareness of rugby soared with the Brave Blossoms' famous victory over South Africa in the last Rugby World Cup in England in September 2015. And even if this interest has dipped in the meanwhile, it is about to pick up. This brings us to the third and final reason for taking an interest in the language of rugby – Japan's hosting of Rugby World Cup 2019, which will kick off at Ajinomoto Stadium in Chofu on 20 September. So there is a potentially huge fan base.

This Autumn (2017) will see the launch of the ScrumSourcing project[1], which aims to produce an

[1] The lead participants are Dr Eiichiro Sumita and Dr

MT engine to support 'rugby talk' between Japanese- and English-speaking fans during RWC 2019. The plan is to deliver the service using the NICT engine and the VoiceTra platform,[2] and to crowdsource the post-editing of the MT output by volunteer rugby fans in order to incrementally improve quality during the training phase.

What, then, are the main challenges from a linguistic perspective?

3. Contrasting writing conventions

The reporting of rugby in Japanese in the national media is – in relative terms – rather sparse, the pre-eminent outlet being Koichi Murakami's column for J-Sports.[3] In contrast, match reports proliferate in the national and local press of Australia, New Zealand, England, Ireland, Scotland and Wales. The UK-based 'Guardian' alone has four regular rugby writers. This means that writers compete not only on perspective but also on style. As a result, there is a marked asymmetry between the writing conventions in the two languages.

Post-match reports in Japanese tend to consist of a mix of interviews with players and coaches and of factual move-by-move descriptions of the game. Here is a typical example:

その後、28 分にゴール前のスクラムから、SH 内田がサイドを突き、リターンパスを FL 長谷川に通して 2 つ目のトライを奪い、36 分にはゴール前のラックから LO ヒーナンが強引にポスト下にトライを決め、40 分にはラインアウトから WTB 北川がサインプレーでライン参加し抜け出し、サポートした CTB 森谷にパスが通り、前半 4 つめのトライを決めた。

In contrast, many reports in English juxtapose such factual descriptions with sustained figurative passages, as here:

Saracens, so expressive against Northampton last week, were lost for words in possession, muted by a furious rush defence and the retrieval qualities of Louw. They found their voice after Farrell replaced Alex Lozowski, a move that allowed Goode, whose first half was mired in mistakes, to roar.

Some of these metaphors become frozen as clichés which recur in the live match commentaries and fans' tweets to be found on sites like Ultimate Rugby.[4]

To investigate these linguistic trends more closely, at Rikkyo we created comparable corpora of texts in the rugby domain, totaling about 3.5M words in each language. We used the SketchEngine[5] software designed for lexicographers and which offers in-built functions for identifying a wide range of distributional patterns, including collocations and thesaural relations, as well as for extracting domain-specific terms. Two tendencies can be observed that distinguish the English from the Japanese: synonymy and economy.

Even the most ardent rugby fan will concede, in all honesty, that a typical match consists of many repetitions of a relatively small number of action types. For example, the ball is passed (backwards) from player to player (running forwards) from one side of the pitch to the other. But a writer who uses always the same form of words to describe this repeated event risks boring the reader. Hence the high frequency of 'spread the ball wide', 'spin the ball

---

Masao Utiyama of the National Institute of Information and Communications Technology (NICT), Professor Hitoshi Isahara of Toyohashi University of Technology, and Rikkyo University.

[2] http://voicetra.nict.go.jp/

[3] https://www.jsports.co.jp/press/column/writer/26/

[4] https://www.ultimaterugby.com/

[5] https://the.sketchengine.co.uk/

wide', 'go through the hands' or 'go wide' and other synonymous expressions. Reports in Japanese, however, draw on a more limited lexicon.

This concern for variety is accompanied by a need for concision. Journalists are constrained by space, live commentators by time, tweeters by the 140 text characters limit. Thus, the ball being central to the game, mention of it is often elided in text, hence: 'Smith spins it wide' or 'Jones carries ø.' Similarly, in our corpus 'The referee blows the whistle' is significantly less frequent than 'The referee blows.'

In this same respect, let's consider the example of 'to clear'. A fan reading 'Brown clears' can picture Brown receiving (or 'gathering') the ball in a defensive position within his/her own half of the field, probably close to the goal line, and kicking it into or towards the attackers' half, or out of play. In Japanese there is no single verb with the same semantics. Instead, the translator can opt for simply the underspecified 'キックを蹴る' or add a space- and time-consuming supplementary phrase to make the player's location on the field explicit.

4. What are the implications for PE?

In the rugby domain we have found very little parallel data to harvest for training – bad news for the MT developer. Even on an international site like World Rugby,[6] which has pages in Japanese as well as English, the volume of segment-for-segment translations is very small as a result of extensive editing, most commonly deletion.

So the role of post-editing in accumulating high quality bitexts becomes crucial. In our scenario, this task is to be performed by volunteers who present the disadvantage of being non-professional translators but the – overriding, we hope – advantage of being rugby insiders.

Given the mis-match between rugby reporting styles in Japanese and English, our Japanese volunteer post-editors will face choices. In the jargon of translation studies, they will find themselves deciding – no doubt unconsciously – between a 'foreignising' option that imports novel words and metaphors into Japanese and a 'domesticating' choice that preserves the conservative norms of the target language.

With this background, let's return to the metaphor in the title of this article. At the level of denotation 'full time' and 'no side', 'フルタイム' and 'ノーサイド' are synonyms: they all denote the end of playing time. But the differences in their distribution in our corpora are remarkable. While 'full time' occurs 18.01 times per million words (pmw), the frequency of 'no side' is only 0.71 pmw. For Japanese the picture is reversed: 'フルタイム' has a frequency of 8.2 pmw and 'ノーサイド' 38.0 pmw.

At the level of connotation, however, 'ノーサイド' encapsulates the essential spirit of rugby, which at the end of a hard-fought game sees the former opponents enjoying celebratory refreshments together.

So what will happen? Will the Japanese yield under statistical pressure from English, with 'フルタイム' supplanting 'ノーサイド'. If so, will the former take on the connotations of the latter? Or will the current camaraderie after the final whistle that sustains veteran players into their 80s start to fade? In the reverse direction could Japanese influence a convivial 'no side' to displace the prosaic 'full time'?

If MT is to become more instrumental in bringing about cultural change through contact – which it surely is – then we prefer change for the better. The outcome lies in the hands of the post-editors.

[6] http://www.worldrugby.org/

招待講演

ニューラル機械翻訳ってどうなの？

中澤　敏明
国立研究開発法人　科学技術振興機構

1. はじめに

2014 年にニューラル機械翻訳(NMT)が登場したことにより機械翻訳のパラダイムは大きく変化した。平均的な翻訳精度は従来の統計翻訳(SMT)と比較すると格段に向上したが、誤りの傾向が SMT とは大きく異なっており、NMT 特有の誤りも存在する。また 2016 年 11 月に Google 翻訳が日本語を含む 8 つの言語を対象に NMT の提供を開始したというニュースは大きな話題となり、その精度の向上ぶりには一般ユーザーだけではなく、翻訳関係者や研究者にも衝撃を与えた。Google の発表によれば、言語対によっては人間の翻訳に匹敵するほどに向上した。

本講演では SMT の簡単な復習と NMT を理解する上で必要なニューラルネットワークの知識を説明した後、SMT と比較しながら NMT の仕組みを解説し、その長所と短所を説明する。またそれぞれの翻訳結果の特徴を明らかにし、今後の翻訳業界と NMT との関わり方についても触れる。

2. ニューラルネットワーク

ニューラルネットワークは動物の神経細胞（ニューロン）における情報伝達の仕組みを模したモデルである。個々のニューロンは他の細胞から情報を受け取り、次の細胞に情報を渡すという簡単な機能しか持たないが、数多くの神経細胞が複雑に結合し、大規模なネットワークを形成することで、高度な情報処理を行うことが可能となる。なお受け取った情報に対し、次の細胞に情報を渡すか否かの基準や渡す量は細胞ごとに異なる。

ニューラルネットワークでは各ニューロンは入力信号を足し合わせ、活性化関数により足し合わせた信号に対して非線形な変換を行い、次のニューロンに情報を伝える。ニューロン間をつなぐエッジにはそれぞれ異なる重みが割り当てられており、エッジを通る信号にはこの重みが掛け合わされる。ニューラルネットワークの訓練は、この重みを最適な値に調整し、望む出力が得られるようにすることである。訓練には教師データ（正解データ）が必要であり、NMT においては対訳コーパスが教師データである。

3. ニューラル機械翻訳の仕組み

NMT エンジンの訓練には、SMT と同様、教師データとなる大量の対訳コーパスが必要である。SMT の訓練や翻訳を行うには様々なモジュールが必要であったが、NMT では 1 つのニューラルネットワークを用意するだけで訓練も翻訳も全て同じ枠組みで行うことができることが特徴の一つである(end-to-end のモデルと呼ばれる)。

NMT には様々なモデルがあるが、JST で開発した NMT は Bahdanau et al.[1]が提案したモデルをベースに改良を加えたものである。本モデルは大きく分けて 3 つのパーツで構成されている。1 つ目は入力文を実数値の集合であるベクトル表現に符号化（変換）するエンコーダー（encoder)、2 つ目は出力するべき単語を決定する際に、符号化された入力文のどこに注目すべきかをコントロールするアテンション機構（attention mechanism)、3 つ目は符号化された入力文とアテンション情報を元に出力文を復号化（生成）するデコーダー（decoder）である。図 1 にこれら 3 つ

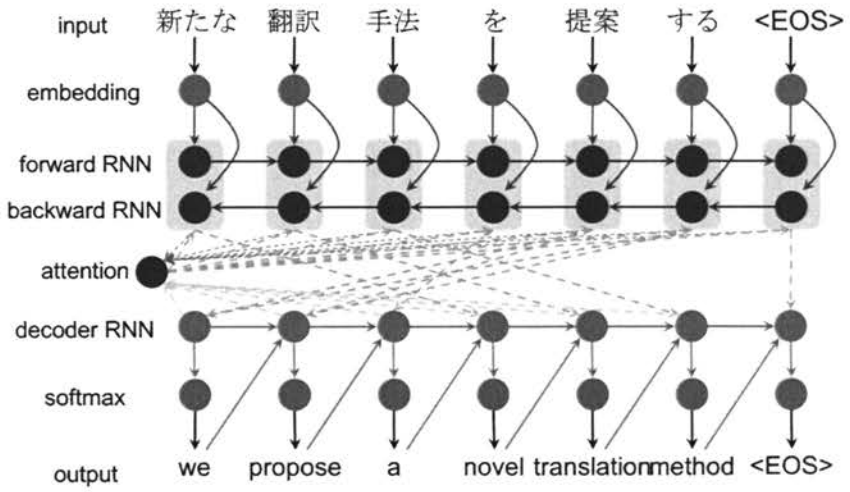


図 1：ニューラル機械翻訳の概略図

のパーツから構成される NMT の概略図を示す。図内の<EOS>は文の終わりを示す特別なタグである。なお一般的に、入力情報を符号化するエンコーダーと、エンコードされた情報を復号化し所望の出力を得るデコーダーからなるモデルをエンコーダーデコーダーモデル（encoder-decoder model）と呼び、文書要約、画像のキャプション生成、気象データからの天気予報文の生成など様々なタスクで応用されている。次に各パーツについて簡単に述べる。

3.1 エンコーダー

エンコーダーではまず各単語を分散表現（distributed representation）と呼ばれる数百次元からなる実数値ベクトルに変換する作業を行う。この操作は embedding と呼ばれる。この分散表現では単語の意味のようなものが表現されていると言われており、例えば「王様」を表すベクトルから「男」を表すベクトルを引き算し、「女」を表すベクトルを足し算すると、「女王」を表すベクトルに近いベクトルになったりすることが知られている。

embedding で表現されるのは各単語独立の意味表現であり、助詞や前置詞などの機能語や複合名詞など、他の語との結びつきにより意味が決定される語がうまく表現できない。そこで次に各単語を先頭から（forward）と末尾から（backward）1 つずつ順に読み込んで行き、前または後ろの単語を考慮したベクトル表現を作り出す。この操作を実現するのがリカレントニューラルネットワーク（Recurrent Neural Network: RNN）である。RNN では自身の出力を再度自身の入力としてフィードバックする機構を持っており、これにより過去の情報と現在の情報を同時に利用することが可能となっている。forward RNN のみでは前の単語しか考慮することができないため、backward RNN で後ろの単語を考慮し、2 つの RNN で生成された 2 つのベクトルを単純につなげたものを、各単語の分散表現として翻訳で利用する。以上がエンコーダーの役割である。

3.2 アテンション機構

アテンション機構は、エンコードされた入力文と次節で説明するデコーダーの内部状態を判断材料として、次の単語を訳出する際に注目すべき箇所を判断する役割を担っている。注目する単語は決定的に 1 つに決めるのではなく、入力文のすべての単語に対して確率的に注目する。1 単語訳出するごとにこの確率（＝アテンション）が再計算され、注目すべき箇所が毎回変化する。

最後に、入力文の各単語の分散表現とアテンション

確率を掛け算し、掛けた結果を足し合わせることで、入力分散表現の重み付き和を計算する。これをコンテキストベクトル（context vector）と呼び、デコーダーが次に訳出する単語を決定する材料として用いる。

3.3 デコーダー

デコーダーは1つのRNNで構成されており、コンテキストベクトルと1つ前に出力した単語の情報を入力として受け取り、次の単語を出力する。このため、訳文は先頭から順に1単語ずつ生成される。RNNからの出力は目的言語で使われる単語の種類数と同じ次元数を持つベクトルである。これを確率のように正規化するのが、ソフトマックス関数（softmax function）であり、最も確率の高い単語を出力する。出力する単語候補には<EOS>も含まれており、これが出力された時点で翻訳が完了する。以上がNMTの仕組みの概略である。

3.4 ニューラル機械翻訳の特徴

ここでNMTの特徴をまとめておく。まずは前述の通りend-to-endのモデルであり、単語アライメントやフレーズテーブルの構築は不要である。またSMTでは翻訳の過程は全て記号（原言語もしくは目的言語の文字や単語）の上で行われるのに対し、NMTでは最初の入力文と最後の出力文以外の部分は全てベクトルのような数値で表現されており、翻訳の過程は数値の足し算や掛け算で実現されている。モデル自体は非常にシンプルになり使いやすくなったのだが、翻訳の過程の解釈は困難、もしくは不可能である。利点としては単語をベクトルで表現するembeddingのおかげで、同じ意味だが別の単語であっても、同じような翻訳を出力することができる点にある。これをSMTで実現するのは困難であった。

翻訳文の生成においてはSMTが入力文の部分的な置き換えの組み合わせで実現され、入力文を全て置き換えることで翻訳が終了したが、NMTでは符号化された入力文とアテンションの情報からデコーダーが次に出力する単語を1つずつ決定し、<EOS>が出力された時点で翻訳が完了する。つまりデコーダーが<EOS>を出力するまで翻訳は継続し、いつ<EOS>が出力されるかは今のところ知る術がない。このため入力文の全てが過不足なく訳出されるという保証がない。アテンション機構が次に訳出するべき箇所を決定するのだが、過去に訳出した箇所の情報を持っていないため、同じところを複数回訳出したり（重複訳、over translation）、まだ訳出していない部分があるにもかかわらず、<EOS>にアテンションして翻訳を終了させてしまったり（訳抜け、under translation）ということが起こる。この点に関しては解決に向けて研究しているところである。

4. ニューラル機械翻訳の質と課題

図2はJSTが情報通信研究機構（NICT）及び京都大学と共催している国際的な機械翻訳の評価ワークショップであるアジア翻訳ワークショップ（Workshop on Asian Translation, WAT）における科学技術論文翻訳タスクの人手評価結果を、2015年度のSMTと2016年度のNMTで比較したものである。評価基準は特許庁が公開している特許文献機械翻訳の品質評価手順[1]のうち、「内容の伝達レベルの評価」に即しており、5段階の絶対評価で評価5が最も良く、評価1が最も悪い。グラフでは各評価値の割合をパーセントで示している。

結果を見ると日⇔中ではSMTの精度がそれほど高くなかったため、NMTに変わったことで大きく精度が向上していることがわかる。一方で日⇔英に関しては評価の低い文の割合は減っているものの、評価5の

---

1 https://www.jpo.go.jp/shiryou/toushin/chousa/tokkyo honyaku_hyouka.htm

割合も減っていることがわかる。これは現在の NMT の最大の問題である「訳抜け」によるものである。

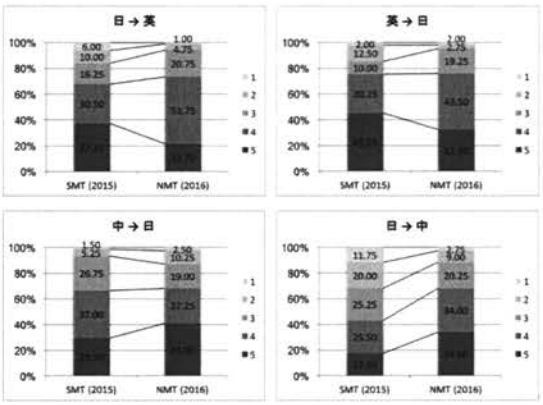


図 2: 日⇔英、日⇔中科学技術論文翻訳タスクにおける NMT と SMT の精度の比較

図 3 に SMT と NMT の英→日翻訳例とその人手評価結果を示す。NMT による訳では入力文中の下線で示した箇所が訳出されておらず、このため評価が 4 以下になってしまっている。3 つ目の例に至っては「標記の方法」などとごまかして訳しており、変に人間味が感じられる。ある研究では、このような訳抜けが全体のおよそ 18%存在すると報告されている[2]。

入力	By this approach, monodispersed swollen styrene particles were obtained and the fine particles in the title were synthesized.
正解	この方法で，単分散なスチレン膨潤粒子を得，標記微粒子を合成した。
SMT	このアプローチにより，単分散膨潤スチレン粒子が得られ，標記微粒子を合成した。(評価5)
NMT	この方法により，単分散のスチレン粒子を得，標記微粒子を合成した。(評価4)
入力	No side effect was noted during treatment.
正解	治療中，副作用は認めなかった。
SMT	治療中に副作用はなかった。(評価5)
NMT	副作用は認めなかった。(評価4)
入力	Using Hartree-Fock-Bogoliubov method, pairing density distribution of the deformed nuclei was investigated.
正解	Hartree-Fock-Bogoliubov法を用いて変形した原子核の対密度分布を調べた。
SMT	Hartree - Fock - Bogoliubov法を用いて，変形した核の対密度分布を調べた。(評価5)
NMT	標記の方法を用いて，変形核の対密度分布を調べた。(評価3)

図 3：NMT における訳抜けの例

ここで注意したいのは、NMT の翻訳は訳文だけを見るとどれも完璧な日本語の文として成立しているという点である。原文と照らし合わせて初めて翻訳誤りが含まれていることが判断できるが、訳文だけを見ても気付くことはできない。SMT では翻訳に失敗した場合は明らかにおかしな訳文が生成されていることが多かったため、翻訳誤りに気づくことも容易であったが、NMT ではその判断が難しくなっていると言える。このように、NMT による翻訳は流暢さ（fluency）は完璧に近いが、訳抜けが起こりやすく正確性（adequacy）が完璧ではない傾向がある。

一方割合としては少ない（約 4%）が、重複訳が生じることもある。図 4 の例では" assistance" や" ski" が何度も出力されていることがわかる。このように一目見ておかしな訳は SMT ではほとんど出力されなかった。

入力	食事については，全介助から半介助となった。
正解	On the meal, changing from all support to half support was realized.
SMT	The meal became half assistance from the assistance.
NMT	On the meal, it became the assistance from all assistance assistance.
入力	スキー滑走中に転倒受傷し右膝変形ととう痛にて受診。
正解	A man tumbled over in a ski slide and was injured and then had right knee deformation and pain and so received medical examination.
SMT	He had medical examination by pain and right knee deformation injured overturning during ski.
NMT	The ski was injured in ski ski sliding, and it had medical examination by right knee deformation and pain.

図 4：NMT における重複役の例

これらの問題は基本的には全て入力文を過不足なく翻訳することができないことが原因である。この点を解決することが今後の NMT 研究の最大の課題と言える。

5. おわりに

機械翻訳の手法には大きく分けてルールベース翻訳、統計的機械翻訳、そしてニューラル機械翻訳の 3 つがある。ルールベース翻訳は全体的な精度や流暢さは低いが、訳抜けを起こすことはなく、また翻訳の根拠が説明できるという利点がある。統計的機械翻訳ではルールの作成が不要になり、容易に多言語化が可能となり、また翻訳の質も大きく向上した。その反面、翻訳の根拠を説明することが困難になり、マシンパワーが必要になるなどのデメリットもある。ニューラル機械翻訳では統計的機械翻訳よりも一層翻訳精度が向上したが、訳抜けや重複訳が発生するといった新たな問題を抱えており、また翻訳の根拠の説明はほぼ不可能となった。

機械翻訳に限らず、ニューラルネットワークに基づく方法はプログラム的な実装が非常に容易である。このため、機械翻訳エンジン自体にはそれほど価値がなくなりつつある。そこで重要となるのはやはりデータ

であり、いかにデータを蓄えるか、いかにデータを共有し、日本全体として機械翻訳に取り組むかといったことを真剣に考える必要がある。

NMT はここ数年で急激に発展し、平均的な翻訳精度は SMT を追い越している。NMT の研究はまだ発展する可能性が高いが、現状の NMT はほとんど制御不能といって良く、解決すべき課題が多く残されており、実用面でいうと不安が残ることは事実である。しかし他言語で書かれた文書の内容をざっと把握する目的ならば現状でも十分な精度であり、人間の翻訳の下訳として使うのも可能であろう。ただし、誤り傾向が SMT と異なるので、新たなノウハウの蓄積が必要となる。今後は NMT をどのように実務で使うかという議論も行なっていく必要がある。

[1] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate.
https://arxiv.org/abs/1409.0473

[2] Zhaopeng Tu, Yang Liu, Lifeng Shang, Xiaohua Liu, Hang Li. Neural Machine Translation with Reconstruction.
https://arxiv.org/abs/1611.01874

招待講演

機械翻訳収益化の可能性

佐藤　弦

SDL ジャパン株式会社

1. はじめに

世間の注目がAIやディープラーニングに集まる中、Google ニューラル機械翻訳がリリースされ、これまで機械翻訳が産業翻訳（※）にあまり援用されてこなかった日本語についても状況が変わるのではないかと翻訳業界が色めき立った。これに対して、現場で機械翻訳ソフトウェアを販売している立場としては、当面、機械翻訳の導入範囲は限定的であると感じている。

本稿では機械翻訳収益化の現状を省みながら、世間の期待と現状のギャップを考察したい。

※本稿においては「専門の翻訳者が企業のために行う翻訳」という意味。

2. 機械翻訳市場の区分け

まず産業翻訳とそれ以外の市場を区分けする必要がある。Google や Microsoft などの機械翻訳の主要プレイヤーが対象としているのは、主に産業翻訳外での使用となる。多くは Web サービスとして無料で提供され、他のシステムのバックエンドで使用するような場合にのみ、接続に必要に API を提供することに対して課金している。

また企業内であっても、社内翻訳や参照用の翻訳などに使われる場合には、翻訳を専門に従事している職員が翻訳していないケースも多いので、やはり産業翻訳とはやや異なる。

これらの産業翻訳以外のケースでの機械翻訳の使用については、加速していくことが見込まれる。個人や企業にとって金銭的な負担が少なく、柔軟性が高い環境で翻訳が行われているからである。しかし、このようなケースであっても Web サービスに対するセキュリティの心配など、障壁がまったくないとは言えない。

3. 産業翻訳における機械翻訳の使用

産業翻訳界においては、Google や Microsoft ではなく、産業翻訳を対象とした機械翻訳ベンダーのソフトウェアが使用されてきた経緯がある。前項に挙げたセキュリティの問題以外に、専門用語の辞書を登録する、XML のタグを適切に処理するなど、カスタマイズの必要性があるためだ。

産業翻訳の現場では機械翻訳の結果がそのまま使用されることはなく、人間の翻訳者がそれを編集するポストエディットという手法が用いられる。機械翻訳の質が向上することで、ポストエディットに移行するプロジェクトの数が増えることは十分に考えることができる。

4. 欧州言語における機械翻訳利用の状況

欧州言語については日本語とは事情がことなり、統計解析型の機械翻訳で既に一定の品質に達していた。そのため、欧州言語おける機械翻訳利用の状況を見ることで、今後の日本語でのあり方が予測できる部分がある。

欧州言語についても、産業翻訳ではポストエディットが中心となっており、その範囲については年々広がっているようだ。これは Common Sense Advisory の発行しているレポートなどで確認ができる。また、一部の翻訳ジャンル（Web ページのユーザーレビュー、

サポート関連ページ）などでは機械翻訳がそのまま使われているケースもあり、こちらはベンチャー企業による新しいWebサービスが増えるのに従い、機械翻訳の導入も広がることが予測できる。

ただし、一部の翻訳会社や翻訳者が想像しているように、人間の翻訳者が機械に取って代わるという兆候は現時点ではまだ確認できていない。

5. まとめ

長い目で見れば、技術による社会変化は必ず起きる。また産業翻訳業界に対する機械翻訳の影響も大きくなっていくだろう。ただし、現状ではまだ将来をはっきりと予測できるような兆候は表れていないというのが私の見立てである。

機械翻訳の品質が向上したから翻訳者は不要になるというのは、大きく捉えれば正しいのかもしれないが、より細かく状況を見ていくと、産業翻訳が翻訳者から機械翻訳に切り替わっていくには、いくつもの障壁があることを理解いただければ幸いである。

最後に、さまざまな経緯から、講演内容がタイトルとややずれたものになってしまったことをお詫びしたい。

AAMT 長尾賞受賞講演

翻訳業界における「ブルー・オーシャン戦略」

湊　幹
八楽株式会社

1. はじめに

皆さんは「ブルー・オーシャン戦略」という言葉をご存知だろうか。

既に広くビジネスタームとして定着しているので、耳にしたことがある方も多いかもしれない。新規事業立案時に用いられる近年人気の高い経営戦略論である。

本稿では、弊社が 2017 年度長尾賞を受賞するにあたり、その大きな要因となった CAT ツール「ヤラクゼン」の目指す世界観と独自のポジショニングについて、ブルー・オーシャン戦略の枠組みから述べていきたいと思う。なお、本稿はあくまでビジネス的な視点に基づく内容であり、機械翻訳技術に関する説明や解説ではないことを予めご了承願いたい。

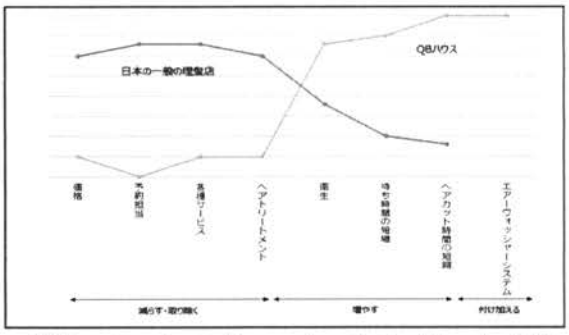
2. ブルー・オーシャン戦略とは

ブルー・オーシャン戦略とは、競合がひしめく血みどろの戦いの場である既存市場（レッド・オーシャン）でしのぎを削るのではなく、新しい価値体系を元に新しい市場を切り拓き、競争のない世界（ブルー・オーシャン）を創り出そうという理論である。価格競争や機能・サービスの高度化の追求により競合に打ち勝つという、マイケル・ポーター的競争戦略論の世界観とある意味対局となる考え方である。

個人的な話で恐縮だが、筆者はブルー・オーシャン戦略が発表された 2005 年当時、同戦略の提唱者であるチャン・キム教授が所属する INSEAD というビジネススクールに在籍しており、直接講義を受ける機会に恵まれた。当時は「シルク・ドゥ・ソレイユ」や「QB ハウス」などを題材に、古くからの業界慣習に捉われずにユニークな立ち位置を築き大成功していく企業を分析し、新しい戦略立案に結びつけるアプローチが斬新で強烈なインパクトを受けたことを覚えている。

では、どうやって競争のない世界を創り出し、新しい需要を掘り起こすのか。基本となるアプローチの 1 つに「非顧客層」に着目するという方法が挙げられる。任天堂の「Wii」が「母親も含めた家族全員」をターゲットにしたように、従来の市場における「非顧客層」に着目することで、今まで未開拓の市場にリーチするチャンスが生まれるのである。

また、ブルー・オーシャン戦略の成功の要諦は、顧客に与える価値を幾つかの要素に分解した上で、その価値体系（「バリュー・カーブ」と言う）が既存サービスと明確な違いを描けているかがポイントであるとしている。その際、業界では当たり前だと思われていた価値に対して「取り除く」「減らす」「増やす」「付け加える」という 4 つのアクションを行うことで、既存サービスとの違いが明確なバリュー・カーブを設計することが重要である。ここは文章だけでは伝わりにくいと思うので以下の QB ハウスの例（図表 1）によって視覚的にご理解頂ければと思う。



（図表 1：QB ハウスのバリュー・カーブ：ランダムハウス講談社「ブルー・オーシャン戦略」を元に筆者加筆）

3. 「非顧客層」＝「非翻訳者」

さて、能書きはこのくらいにして本題に移りたい。これらを CAT ツール「ヤラクゼン」について当てはめて考えてみよう。

まず、我々が着目しているのは、一般のビジネスパーソン、いわゆる「非翻訳者」である。他の CAT ツールがターゲットとしている「専業の翻訳者」ではなく、ビジネスの現場で日々外国語対応に追われる人たちである。語学が得意な人もいれば苦手な人もいる。外国語対応に多くの時間を取られる人もいれば、たまに必要なだけの人もいる。所属部門も様々である。そのような人たちをメインのユーザーとして想定している。

それではこの「非翻訳者」市場はどう評価するべきか。少なくとも我々はポテンシャルのある市場と見ている。なぜなら、単純に「専業の翻訳者」よりも「外国語対応が必要な一般のビジネスパーソン」の方が人数が桁違いに多く、且つ今後も高い増加率が見込まれるからである。また、組織内外のグローバル化が加速する中、単純に「数」だけでなくユーザーの「痛み」も確実に高まることも容易に予想できる。

もちろん、現時点ではマーケットとしてほぼ存在しない市場であり、それゆえ誰も証明することはできないが、やり方次第では新しい需要を創り出すことができるポテンシャルの大きな市場である、と考えている。ベンチャー企業である我々にとっては非常に適した市場であると言えるだろう。

4. ヤラクゼンのバリュー・カーブ

では次に、どのようなバリュー・カーブを描くべきなのか。言い換えれば、どこに組織としての力点を置き、従来の CAT ツールとの違いを生み出すべきか。ここでも発想の起点は想定顧客からとなる。現場のビジネスユーザーにとって既存の CAT ツールで使える部分、不要な部分はどこか。何を「取り除き」、「減らし」、「増やし」、「付け加える」べきなのか。それらを徹底的に考える。この点における弊社の強みは、社員の大半の人間が翻訳者ではなく、翻訳業界出身でもない点だ。あくまで先入観に捉われず、自分自身の経験も踏まえながら現場のビジネスユーザーの視点から考えることができる。

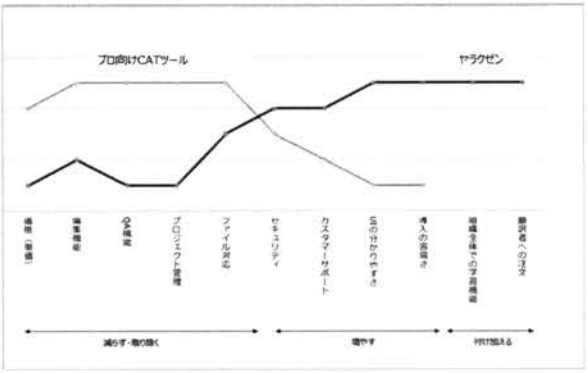
さて、具体的に要素ごとに見ていこう。まずは「取り除く/減らす」だが、一般のビジネスユーザーにとって必要となる翻訳は主としてビジネス文書やメール、ウェブ上での情報収集等であり、数十ページ、数百ページのマニュアルや大量の Web サイトの翻訳等を複数人で分担してやることはレアケースであろう。そう考えると、従来のプロ向けの CAT ツールが備える充実したプロジェクト管理機能をフルに搭載する必要はない。また日々利用するファイルの種類も限られるため、対応ファイルの種類も思い切って絞り込むことができる。プロの翻訳者には欠かせない多様な編集機能や QA 機能も基本的なものだけで日常使いには十分である。

そのように機能を思い切って絞り込むことにより、ツールそのものの分かりやすさ、シンプルさを実現する（「増やす」）。ビジネスユーザーにとって、翻訳業務に費やす時間はあくまで業務時間の数十分の一そのために難しい操作方法を覚えたり、マニュアルを読んだり、トレーニングを受けるというのはナンセンスである。ぱっと見て、ぱっと使える。社内のスケジューラーやメーラーなどと同じレベルの手軽さ、分かりやすさで使いこなせるようなツールになることがこの層には重要になってくるのだ。（図 2）

（図 2：シンプルな画面設計にこだわったヤラクゼンのユーザーインターフェース）

また、幅広い部門のユーザーに使ってもらうとすると、組織内での利用人数が一気に増えるため、カスタマーサポートやセキュリティ対策が相対的に重要になってくる。そのため、ヤラクゼンでは法人顧客向けに個別の専任担当者を配置し電話・メールで直接サポートできる体制を整えている。セキュリティに関してもパブリッククラウド方式だけでなくプライベートクラウドやオンプレミス方式での導入やネットワークレイヤーでの様々なオプションも準備している。

加えて、非翻訳者ということで、ユーザーの語学レベルには当然ばらつきが出てくる。そのため、組織内でナレッジを蓄積・再利用する仕組みを新しいコンセプトで作り出したり、プロの翻訳者に画面上から簡単に翻訳を注文できるシステムを追加したりすることで、語学レベルが低いユーザーでも一定のレベルでの業務ができるような仕組みを「付け加え」ている。



（図表３：ヤラクゼンのバリュー・カーブ：著者作成）

これらを反映させた「ヤラクゼン」のバリュー・カーブが図表３である。CAT ツールという製品カテゴリの中で既存のプロ向け製品と比べて明確な違いを創り出すことに成功しており、それゆえ現場のユーザーからの強い支持が得られる製品となっている。

（なお、前述の通り本稿はヤラクゼンの立ち位置をお伝えするのが目的であり、図表中の具体的な値についての厳密な議論はここではご容赦頂きたい。）

5. 組織能力とパートナーシップ

このように我々は新しい価値体系を設計し、それに沿った形で製品開発の優先順位の意思決定を行っている。但し、設計するだけでは絵に描いた餅であり、実現するには組織としてのケイパビリティが必要となる。

弊社は元々機械翻訳を活用して人々のグローバルコミュニケーションを広げていこうという理念のもとに設立された会社である。当初の事業である多言語掲示板の開発・運営の時代より機械翻訳/自然言語処理の研究者とソフトウェア開発・UI デザインの専門家を世界中から集め、ノウハウを継続的に蓄積してきている。限られた人数ではあるが１つのことを一丸となってコツコツと続けて組織能力を高めてきている。

とはいえ、新しい市場を創り出すには認知の拡大と顧客へのリーチが欠かせない。弊社だけで進めるには営業力・信用力・カバレッジが圧倒的に不足することをここ数年痛感してきた。そこで本年より「ヤラクゼン」の OEM ビジネスを開始することにした。顧客との信頼関係が既にあり、質の高い翻訳サービスを提供している翻訳会社と提携し、共同で顧客に価値を届ける枠組みにより想定以上の成果を上げつつある。

ニューラルネットワークによる翻訳技術の革新は目覚ましい。但し、現状はビジネスの現場で多くの人が利用するには適切な形になっているとは言い難い。どんなに優れた技術でも実際に利用されず、フィードバックが得られなければ進化に繋がらない。その実用化の取り組みを業界各社と力を合わせて実現していきたい。パートナーシップは前述の翻訳会社との取り組みだけでなく、機械翻訳エンジン開発企業や辞書提供会社など様々な会社と開始をしている。一緒に取り組んで頂ける企業があれば是非お声がけ頂きたい。

6. 終わりに

長尾賞の副賞として頂いた図書カード。長尾先生に

御揮毫頂いた今年の言葉は「快哉」であった。(写真4)機械翻訳の実用化により人々のグローバルコミュニケーションを支え、日本産業界の競争力向上へ貢献する。ひいては世界中の様々なバックグラウンドの人々の相互理解促進のプラットフォームになっていく。その時には関係各位で心より快哉を叫びたい。そんな未来を見据え、長尾賞受賞企業の名に恥じぬよう社員一丸となって頑張っていきたい。

(写真4：長尾賞副賞の図書カード)

(参考文献)

- W・チャン・キム、レネ・モボルニュ(2005)『ブルー・オーシャン戦略』ランダムハウス講談社.
- 三谷宏治(2013)『経営戦略全史』ディスカヴァー・トゥエンティワン

AAMT 長尾賞学生奨励賞受賞講演

Improving Statistical Machine Translation with Target-Side Dependency Syntax

John Richardson

Google Japan

1. Introduction

Translation is particularly challenging for language pairs with widely different grammatical structures, such as English and Japanese. One of the more popular approaches for successful handling of such language pairs is syntax-based MT, a translation paradigm based on the principle of abstracting language with grammatical rules.

The majority of previous approaches to syntax-based MT have employed only source-side grammar (known as 'tree-to-string' MT). This is mainly because syntactic analysis must be performed on both source and target sides for training tree-to-tree systems, and the resulting translation rules are in general sparser than their string-based counterparts (as the same string can be represented by multiple trees). While there have been previous studies on exploiting target-side syntax (or 'tree-to-tree' MT [1]), results have not been promising. There are however a number of potential benefits of using target-side syntax, in particular assisting reordering and language modeling, two important areas for improving the fluency of machine translation output.

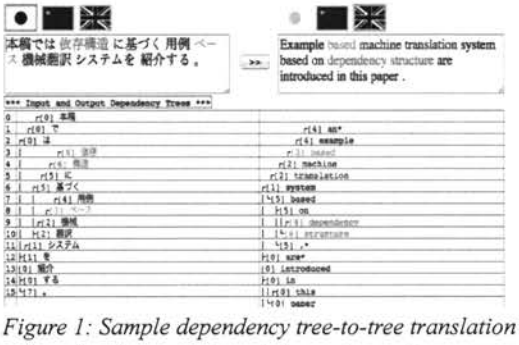


Figure 1: Sample dependency tree-to-tree translation with colors showing each translation rule used to generate the final translation.

Figure 1 shows an example translation with such a tree-to-tree translation system, KyotoEBMT [2], demonstrating how syntax can be useful in building large translation rules covering distant parts of an input sentence.

Our aim is to analyze the effectiveness of target-side syntax, in particular dependency grammar, and ask whether the potential improvement in translation quality is able to outweigh the increased complexity of employing a structured target-side representation.

2. Dependency Tree Language Model

We begin by considering the task of language modeling and how it can be improved by exploiting target-side syntax. Long-distance agreement is a tricky issue for string-based machine translation, which is susceptible to errors such as incorrect noun/verb agreements, and we hope to solve this problem by considering target-side syntax.

Classic *n*-gram language models are effective at capturing translation fluency at the word level, however such approaches often fail at the syntactic and semantic level. We abstract the traditional definition of a classic *n*-gram to dependency trees and show how our approach is able to improve more challenging issues such as long-distance word agreement. The primary motivation for using structured language models is that we can reduce the 'distance' between words that are interdependent on a syntactic (and often semantic) level. While tree-based models have more complicated structure than their string-based counterparts, the sparsity of the most important information is reduced, giving a more compact and relevant representation of context.

Our first approach extends the linear context described in the classic *n*-gram language model by adding dependency tree structure [3]. We define '*t*-treelets' as all possible dependency tree structures with *t* nodes (words) and the '*t*-treelet context' of a node in a dependency tree as the set of all *t*-treelets containing the node that are strictly to the 'left' of the node. The notion of 'leftness' is the tree-based parallel of only considering previous words as context in a traditional *n*-gram, and we define it a node as 'left', or 'context', of another if it would be reached first in a depth-first search starting from the root node. Figure 2 shows an example of the *t*-treelets extracted as context for an example sentence.

We now construct a language model similar to a standard n-gram model by counting the frequency of all t-treelets (within some maximum size) in the parsed training corpus. The t-treelet language model probability is similar to the n-gram case, however we must consider the probability of all t-treelets (as there are multiple t-treelets of the same size for any node). We do this by taking their geometric mean.

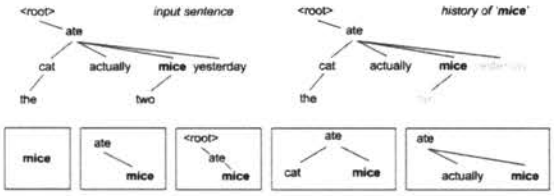


Figure 2: Example of t-treelet extraction. The left figure shows a parsed example sentence and the right figure highlights the context for the word 'mice'. The five possible t-treelets of order < 3 are enumerated below.

To evaluate the potential of this method, we assessed the improvement in translation quality attainable by using our language model as a feature for reranking parsed n-best translations. Our baseline was a strong phrase-based baseline system (Google Translate, as of late 2014) and we applied the language model to rerank the 1000-best translations of English to a variety of target languages. Table 1 shows the results of our human evaluation, in which raters were asked to give a score from 1-6 (higher is better) to the translation quality of the baseline and proposed (reranked) systems.

	Slavic		Uralic	Germanic	
Language	Czech	Russian	Hungarian	Dutch	German
CR	8.51	2.74	5.57	7.04	7.57
mean-diff	**0.08**	**0.10**	**0.06**	-0.02	**0.06**
mean-sign	**0.08**	**0.08**	**0.06**	-0.02	**0.06**

	Romance			Semitic
Language	French	Portuguese	Spanish	Hebrew
CR	4.71	10.21	4.98	6.18
mean-diff	**0.11**	0.01	0.01	0.01
mean-sign	**0.11**	0.01	0.03	0.02

Table 1: Results of human evaluation. 'mean-diff' is the mean rating difference (0-6) between baseline and proposed systems, 'mean-sign' is the mean sign difference (diffs are normalized to -1, 0 or 1), and 'CR' (change rate) is the percentage of sentences with different translations for baseline and proposed systems.

The results showed that the proposed model is most effective for morphologically rich languages with free word order, such as Slavic and Uralic families, because they contain more examples of non-local dependencies that affect fluency, and we saw no significant improvement over a classic n-gram language model for groups such as Romance languages with high baseline quality and few non-local word agreements. On the positive side, we demonstrated that a syntax-based model is capable of performing positively for some language pairs by directly parsing string MT output.

3. Exploiting Projected Source Dependencies

There remain a number of drawbacks to the first approach. The first is the assumption of the availability of a high-quality target-side parser, especially since we are parsing potentially malformed translations. The second disadvantage is that we do not consider source-side context, which can be important for preserving the intended meaning.

We now present a practical solution to solve these two issues. The idea is to exploit relatively high quality source-side English parses and to build a projected language model that is independent of the target language.

We trained the projected language model by parsing the source side of the training corpus and projecting the dependencies to the target side using word alignment, then using the projected target side parses to learn a maximum likelihood model. For inference we parsed the source sentence and created projected dependencies on the target side for all n-best translations. Note that this method only requires one parse, as opposed to n. See Figure 3 for an example of how the projected language model was used for post-editing at inference time.

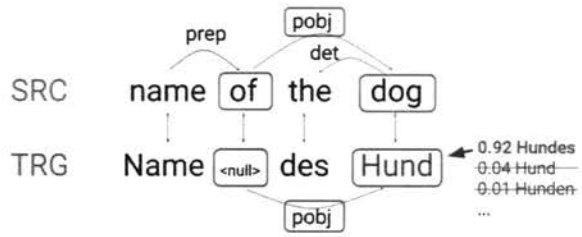


Figure 3: Post-editing an incorrect word form. Firstly the context is extracted for the word in question ('Hund') on both source (green) and target (blue) sides using mapping based on the word alignment (shown with double-headed arrows). 'Hund' is replaced by 'Hundes', the word with the highest probability of having the marked bilingual context according to the projected dependency language model.

A potential downside of such a mapping approach is that we assume the syntax of the source language is sufficiently similar to the target language. We therefore attempted to mitigate this issue by considering only context of a limited size. We restricted the language model to consider only a hand-crafted subset of the t-treelets defined in the previous section, designed specifically to capture

common word agreements such as noun-verb and adjective-noun bigrams.

The improved method was evaluated with a similar experimental setup and human evaluation to the previous model. Table 2 shows the human evaluation results.

	CR	mean-diff	mean-sign
bg	0.63	0.34 [0.25, 0.42]	0.28 [0.21, 0.35]
ca	0.76	0.15 [0.08, 0.22]	0.14 [0.08, 0.20]
cs	4.34	0.12 [0.05, 0.18]	0.10 [0.04, 0.17]
da	1.70	0.45 [0.38, 0.52]	0.42 [0.35, 0.48]
de	2.82	0.29 [0.22, 0.37]	0.24 [0.19, 0.30]
el	1.10	0.07 [0.01, 0.13]	0.06 [0.01, 0.11]
es	1.09	0.27 [0.18, 0.37]	0.23 [0.17, 0.30]
fr	1.51	0.10 [0.02, 0.17]	0.10 [0.04, 0.16]
hr	2.15	0.19 [0.12, 0.25]	0.17 [0.11, 0.23]
it	1.07	0.27 [0.17, 0.36]	0.21 [0.15, 0.27]
lt	3.72	0.24 [0.18, 0.31]	0.21 [0.15, 0.27]
lv	3.76	0.11 [0.06, 0.17]	0.11 [0.06, 0.16]
nl	2.58	0.46 [0.38, 0.54]	0.42 [0.34, 0.49]
no	1.34	0.23 [0.16, 0.31]	0.22 [0.16, 0.29]
pl	2.97	0.11 [0.03, 0.19]	0.11 [0.04, 0.18]
pt	2.02	0.29 [0.20, 0.39]	0.24 [0.18, 0.31]
ru	2.73	0.09 [0.02, 0.15]	0.06 [0.01, 0.12]
sl	3.42	0.10 [0.05, 0.15]	0.08 [0.04, 0.13]
sv	1.70	0.34 [0.26, 0.42]	0.28 [0.22, 0.35]
uk	2.38	0.07 [0.01, 0.14]	0.06 [0.01, 0.11]

Table 2: Human evaluation for projected language model.

We can see that the projection approach gives a significant improvement in translation quality compared to the baseline for all of the 20 languages we tested. This reranking approach is simple, efficient and decoder-independent, and unlike much post-editing work does not require human post-edit logs as we can train directly on the parallel corpus.

The projection approach was successful in terms of precision, however the change rate was low for out-of-vocabulary words and unseen n-grams, in particular for languages that display rich case endings and agglutination. To improve the coverage of our approach for unseen n-grams we also considered an extension to a more general delexicalized language model. We mined word inflections from Wiktionary for six languages (Czech, German, Spanish, French, Portuguese and Russian), creating a mapping of word forms to grammatical attributes including gender, number and person. The language model was retrained, replacing all context words with their corresponding list of morphological tags from the mined lexicon. At inference time, we backed off to the original language model if no tags were found in the lexicon. This approach improved the coverage but also reduced the translation quality by a comparable amount, so it not clear this approach is better.

4. Applying Target-Side Syntax to Reordering

Word order is one of the major challenges for translating syntactically distant language pairs such as English-Japanese. It is therefore also one of the potentially most fruitful applications of systems exploiting target-side syntax, and so we attempted to apply the dependency language model described above to this task.

Syntax-based reordering is relatively simple for binarized constituency parses, as it is sufficient to decide either to retain the order of children or to swap them, however there has been little work on reordering with dependency parses since standard dependency nodes can have arbitrarily many children, giving $n!$ possible reorderings for n children. We propose a solution for this more general setting.

Our baseline system was KyotoEBMT [2], a state-of-the-art dependency tree-to-tree syntax-based translation framework. KyotoEBMT by default combines translation hypotheses based on the weighted sum of a number of hand-crafted features including a classic n-gram language model score.

We proposed the addition of a special type of non-terminal symbol (known as a 'flexible non-terminal' [4]) that allows insertion of hypotheses at an arbitrary number of positions in the parent hypothesis based on the tree structure of the parent hypothesis, and extended the hypothesis combination model by adding three tree-based language model features. These are shown in Table 3 as 'Pref', the preference (probability) for joining a subtree as a left or right child of the head of the parent hypothesis; 'Ins', the probability of inserting a subtree into the parent hypothesis; and 'Rel', the probability of inserting the subtree at some position amongst the other child nodes of the parent hypothesis. These probabilities were also conditioned on the source token being translated, and were trained with a simple maximum likelihood model on aligned trees from the parsed training corpus.

The following is an example of a translation improved with the proposed method, as the hypotheses are combined in a more natural order considering the target-side structure. The constituent translation hypotheses are shown in brackets.

Baseline: *[U.S.A. promotes the superiority] [of the domestic] [of international competitive power] [as a strategic tool] [by the close cooperation of the government and industry and university] [and] [.]*

Flexible: *[U.S.A. promotes the superiority] [of international competitive power] [of each country] [as a strategic tool] [and] [by the close cooperation of the government and industry and university] [.]*

	JA–EN	EN–JA	JA–ZH	ZH–JA
Moses	18.09	27.48	27.96	34.65
Baseline	19.97	28.41	28.13	33.51
Flexible	21.23†	**30.11**†	29.42†	35.37†
+Pref	**21.66**‡	29.90†	29.48†	35.57‡
+Pref+Ins	21.47‡	30.03†	29.64†	35.71‡
+Pref+Ins+Rel	21.34†	29.99†	**29.78**‡	**35.81**‡

Table 3: BLEU scores for our reordering experiment. 'Flexible' is the proposed system with no LM features. Results marked †/‡ are significantly higher than baseline/'Flexible' at $p < 0.05$.

Table 3 shows the results of our evaluation experiment. The proposed method gives an improvement of the baseline by around 1-2 BLEU points at the cost of only a 30% increase in decoding time.

5. Conclusion and Future Work

In this article we have explored the potential of exploiting target-side dependencies to improve statistical machine translation. This has led to the development of a number of new approaches to important machine translation tasks including language modeling, reordering, and post-editing.

There remain however a number of challenges. Parsing quality is still a serious issue and remains the major drawback to syntax-based translation approaches. Many of our proposed methods are also strongly influenced by the baseline translation quality, as they mainly consider the target-side, and word alignment quality affects mapping-based approaches.

Recently many of these issues have been mitigated by neural network translation methods, which do not explicitly use (or significantly improve the quality of) parsing and alignment. It is therefore natural that future work must consider whether it is possible to integrate syntax-based approaches into neural translation frameworks. While there is certainly potential for hybrid models incorporating syntax, the question is to what extent this information can be useful. Recent approaches such as the 'Transformers' attention-based model [5] have an arguably similar motivation to tree-to-tree translation, selecting the most important context for each word, however do not explicitly use parsing as a preprocessing step, instead learning similar representations jointly with the translation model.

It remains to be seen whether our current linguistic theories can express the optimal abstractions for translation, and indeed recent advancements in neural network research have hinted that linguistic theories are already being learned and not taught. Once we are able to understand abstractions learned using unsupervised methods, we can aim improve existing linguistic theories and their applications to MT.

Acknowledgements

This work was conducted at Kyoto University under the supervision of Prof Sadao Kurohashi, and at Google Japan with support from Taku Kudo and Hideto Kazawa. I am grateful to Toshiaki Nakazawa and Fabien Cromières for their collaboration on the development of KyotoEBMT.

References

[1] Min Zhang, Hongfei Jiang, Aiti Aw, Jun Sun, Sheng Li, and Chew Lim Tan. *A Tree-to-Tree Alignment-Based Model for Statistical Machine Translation*. In Proceedings of MT Summit (2007).

[2] John Richardson, Fabien Cromieres, Toshiaki Nakazawa and Sadao Kurohashi. *KyotoEBMT: An Example-Based Dependency-to-Dependency Translation Framework* [System Demonstration]. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (2014).

[3] John Richardson, Taku Kudo, Hideto Kazawa and Sadao Kurohashi. *A Generalized Dependency Tree Language Model for SMT.* Journal of Natural Language Processing, Vol. 23, No. 3. (2016).

[4] John Richardson, Fabien Cromieres, Toshiaki Nakazawa and Sadao Kurohashi. *Flexible Non-Terminals for Dependency Tree-to-Tree Reordering*. In Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics (2016).

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. *Attention Is All You Need*. arXiv preprint arXiv:1706.03762 (2017).

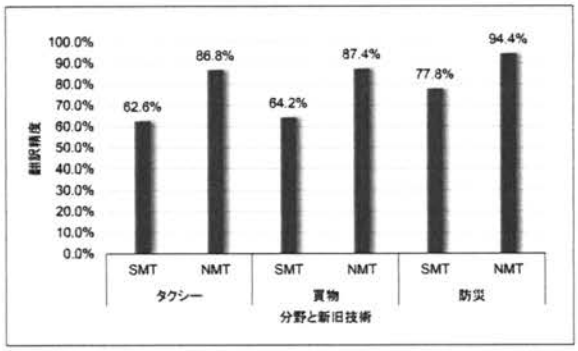
翻訳バンク始めました！

内山　将夫
情報通信研究機構

1. ニューラル機械翻訳で翻訳精度が向上

近年のニューラル機械翻訳（NMT）の出現により、自動翻訳精度が格段に向上しました。情報通信研究機構（NICT）においても、音声翻訳（VoiceTra）とテキスト翻訳（TexTra）でNMTを導入しています。

たとえば、VoiceTraにおいては、下の図に示すように、従来技術である統計的機械翻訳（SMT）と比べて、英日自動翻訳において、各分野において顕著に翻訳精度が向上しています。[1]

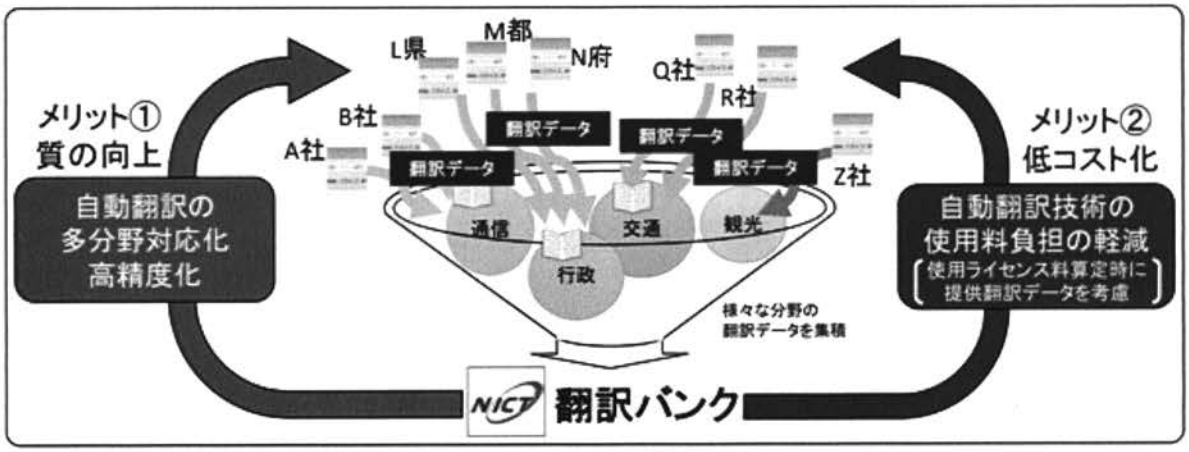


また、みんなの自動翻訳@ＴｅｘＴｒａにおいても、日英・英日の汎用NMTを利用可能です。このNMTを利用しますと、たとえば、「近年のNMTの進展により、従来は自動翻訳が非常に困難だった日本語文章の英語への自動翻訳精度が顕著に向上してきた。」が「With the recent development of NMT, the automatic translation accuracy of Japanese sentences that had previously been very difficult to translate has improved significantly.」のように、「into English」という訳抜けはあるものの、かなり流ちょうな英文へと自動翻訳されます。

このように、NMTにより翻訳精度が顕著に向上しましたが、更なる翻訳精度の向上のためには、対訳データの更なる蓄積が必要です。

2. 翻訳バンク始めました！

総務省とNICTは、自動翻訳システムの様々な分野への対応や高精度化を進めるため、『翻訳バンク』の運用を開始しました。[2]



[1] 2017年6月28日 ニューラル機械翻訳で音声翻訳アプリVoiceTraが更なる高精度化を実現
https://www.nict.go.jp/press/2017/06/28-1.html

[2] 2017年9月8日 『翻訳バンク』の運用開始
http://www.soumu.go.jp/menu_news/s-news/01tsushin03_02000220.html

一般に、自動翻訳エンジンの性能は、翻訳アルゴリズムと、エンジンを訓練する対訳データの量に大きくかかってきます。特に、NMT は、対訳データの量が大きい場合の性能の向上が顕著です。

そのため、翻訳バンクの当面の目標として、1 億文の翻訳データの集積を目指します。この目標の達成のためには、すでに対訳データを保有している方々にご協力いただくことが不可欠です。そこで、翻訳バンクの開始にあたり、データを提供して頂く方々のメリットを明確化するため、NICT の自動翻訳技術の使用ライセンス料の算定の際に、提供が見込まれる翻訳データを勘案して負担を軽減する仕組みを用意しました。

自動翻訳技術の性能向上のためには、NICT が研究開発において取り組んでいる翻訳アルゴリズムの改良に加え、翻訳データの質と量の確保が重要となっています。翻訳バンクでは翻訳データを集積して自動翻訳技術に活用することで、自動翻訳技術で対応できる分野を広げるとともに、さらなる高精度化を実現します。

皆さま方のご支援・ご協力をよろしくお願いいたします。

医療現場向けハンズフリー多言語音声翻訳システムのご紹介

富士通株式会社

2017 年 6 月 14 日に、ホテルアジュール竹芝にて開催された「AAMT 機械翻訳フェア（MT フェア）2017」にて、株式会社富士通研究所、富士通株式会社で研究開発中の医療現場向けハンズフリー多言語音声翻訳システムのご紹介と、それを用いたデモの展示を行いました。

1. 医療現場向けハンズフリー多言語音声翻訳システム

医療現場向けハンズフリー多言語音声翻訳システムは、タブレットを端末として用い、ネットワークを介してサーバで翻訳する日本語／多言語の音声翻訳システムです。タブレットに格納されたアプリケーションから、無線 Wi-Fi 経由で、サーバ内の翻訳エンジンにアクセスします。翻訳エンジンは、情報通信研究機構（ＮＩＣＴ）のエンジンを使用しています。端末では、雑音環境の中でも発話の始めと終わりを正しく認識する発話検出機能と、指向性ステレオマイクによる発話方向の特定によって話者を見分ける話者特定技術とを組み合わせることによって、端末に触れずに他言語の２者間でのハンズフリー会話が可能になります。

2. ハンズフリーのメリット

●会話や作業に専念できる

音声を認識して自動的に翻訳を行ってくれるため、話者により端末の操作が不要になります。手を使う作業をしているときでも、作業を中断することなく利用できます。

本システムを利用した模擬実験の例

●コミュニケーションの質が向上

相手の目を見て会話することができまので、より自然で人間らしい会話をすることができます。

●医療現場での活用

手がふさがりやすい医療・看護・介護の現場や、受付、窓口などの対面業務での活用ができます。話者が端末操作のために触れる必要が無いため、医療現場における感染リスクの減少も期待できます。

最後に

当日はたくさんの方々にお越しいただき、意見交換をさせていただくことができました。誠にありがとうございました。

お問い合わせ先：

・お問い合わせフォーム
https://jp.fujitsu.com/cgi-bin/fjid/formoutput_cs.cgi?FMT=/contact/csform/csque11701/form0010/1

・富士通コンタクトライン（総合窓口）
TEL: 0120-933-200

「みんなの翻訳実習」
プロジェクト翻訳時代の翻訳教育支援統合プラットフォーム

影浦峡・藤田篤・内山将夫・Anthony Hartley・山田優・阿辺川武・Martin Thomas
東京大学・NICT・NICT・立教大学・関西大学・NII・リーズ大学

2017 年 6 月 14 日に、ホテルアジュール竹芝にて開催された「AAMT 機械翻訳フェア（MT フェア）2017」にて、「みんなの翻訳実習」の紹介を行いました。以下ではシステムを簡単に紹介致します。

1. 翻訳産業と翻訳教育の動向

翻訳産業界ではプロジェクトで翻訳を行うことが基本となっており、それは翻訳サービス国際規格 ISO 17100 でも示されています。また、明示的な品質管理の必要性もますます強く認識され、それに伴い、翻訳納品物に対する説明も求められるようになっています。そうした動向を受け翻訳教育では、狭義の翻訳力だけでなく、プロジェクト管理・コミュニケーション・説明力を含む翻訳者力の養成が重視されるようになってきています。さらに、こうした力をつけるために現実の文脈で翻訳実習を行うことの有効性が認められています。MT への対応力も重要になっています。これまで、こうした環境を考慮した翻訳教育用プラットフォームは存在しませんでした。

2. 「みんなの翻訳実習」

こうした背景から、我々のチームはプロジェクト翻訳時代の実践的翻訳力・翻訳者力養成をサポートするプラットフォーム「みんなの翻訳実習」を開発しています。主な特徴は以下の通りです。

●標準的プロジェクトモデル
翻訳産業界における標準的な翻訳プロジェクトの中核的な機能を提供しています。

●精選された翻訳支援ツール（CAT）機能
「ツールを学ぶ」ではなく「ツールで学ぶ」ことを可能にするため、必要な CAT 機能を絞り込み、論理的に推測可能な機能ステップとインタフェースでそうした機能を提供します。翻訳文書作成のために、組込み翻訳支援エディタ「QRedit」を提供しています。

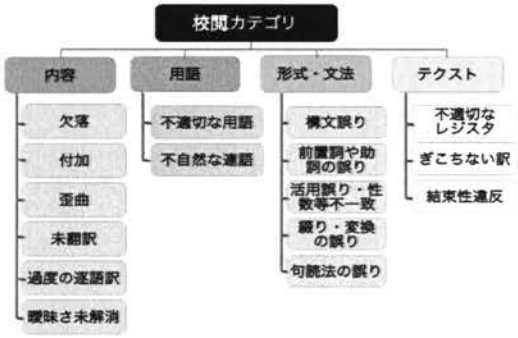
●学習の足場（メタランゲージ）
訳文の校閲に際し、どうして校閲を加えるのかを意識し言語化を促す校閲カテゴリ、プロジェクト参加者間でコミュニケーションを取る際に目的や役割の意識化を促す対話カテゴリ、翻訳に際しどのような情報参照を行うかを意識化させる階層的レファレンス参照機能等を通して、反省的な学習を促します。それにより、「翻訳できる」「対話できる」「情報参照できる」力と同時に「翻訳することを説明できる」「対話することを説明できる」「情報参照することを説明できる」力を身に付けることができます。

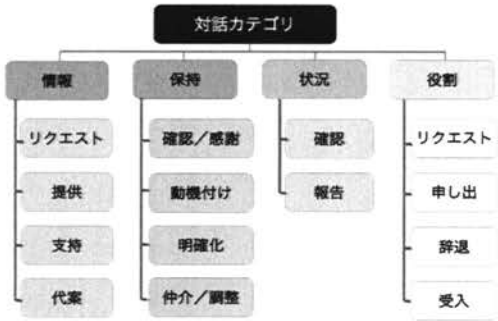
3. お問合せ・アクセス

「みんなの翻訳実習」は https://edu.trans-aid.jp/ で運用中です。利用にあたっては登録が必要です。お問合せ・試験利用のお申し込みは mnhtt-info@panflute.p.u-tokyo.ac.jp へどうぞ。

最後に
当日はたくさんの方々にお越しいただき、意見交換をさせていただくことができました。誠にありがとうございました。

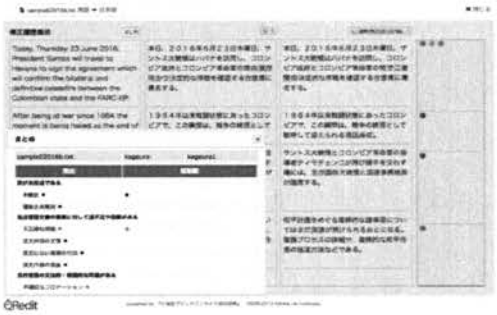
謝辞
本システムは日本学術振興会科学研究費補助金基盤研究(A)「翻訳知のアーカイヴ化を利用した協調・学習促進型翻訳支援プラットフォームの構築」の枠組みで開発しています。





翻訳エディタ（修正モード）

校閲統計（詳細表示）



SMT を用いた NMT の改良

江原　暉将
江原自然言語処理研究室

1. はじめに

ニューラル機械翻訳(NMT)の出現で、翻訳精度が格段に向上した機械翻訳であるが、NMT には訳抜けや湧き出しという問題があることが指摘されている[1]。一方、従来の統計的機械翻訳(SMT)では、訳抜けや湧き出しが比較的に少ない。そこで、本文では、SMT を用いて NMT の上記の欠点を軽減する方法を提案する。

2. システム構成

本システムの構成図を図 1 に示す。

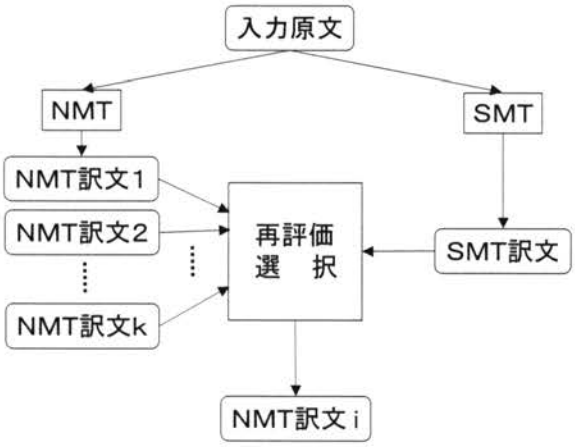


図 1 システム構成

入力原文は NMT 部と SMT 部に入力される。NMT 部からは、訳文の候補が k 文得られ、SMT 部からは、訳文が 1 文得られる。通常の NMT であれば、第 1 候補の訳文である「NMT 訳文 1」が出力されるが、本システムでは SMT 訳文を用いて NMT 訳文候補を再評価し、必要に応じて候補の順序を入れ替えるリランキングを行う。そして、再評価の結果、最も評価値の高かった訳文を出力する。図 1 では、「NMT 訳文 i」が出力されている。

3. 再評価の方法

再評価には、機械翻訳の自動評価に使われる指標である IMPACT [1]を用いる。機械翻訳の自動評価は、理想的な翻訳結果であると仮定した参照訳文と機械翻訳文を比較して行うのが一般的であり、IMPACT もそうである。ここでは、SMT 訳文を参照訳文として、k 個の NMT 訳文の IMPACT 値を計算することで再評価を行う。

4. 実験設定

NMT には OpenNMT [2]を用いる。語彙サイズは 5 万語であり、未知語辞書として、SMT の訓練過程で得られる phrase-table から作成した辞書を利用する。NMT の出力文数(k)は 50 に設定した。

SMT には Moses [3]を用いる。句ベースの SMT である。Moses に入力する前に、原文の語順を訳文のそれに近づけるリオーダリングを行っている。リオーダリングのための構文解析は Berkeley parser [4]を用い、リオーダリングは自作の規則方式ツール[5][6]を用いている。アライメントは mGIZA を用い、言語モデルの次数は 6、distortion limit も 6 である。

実験データとしては WAT2016 (Workshop on Asian Translation 2016)の中日特許翻訳タスクのデータ[7]を用いる。訓練データは 100 万文対であり、開発データ、開発試験データ、試験データはいずれも

[1] 訳抜けや湧き出しの具体例は表 2 を参照。

2000 文対である。

5. 実験結果

開発試験データに対する、SMT の訳文、NMT の第 1 訳文、提案法の訳文を標準的な機械翻訳自動評価基準である BLEU [8]、RIBES [9]、IMPACT の 3 者で自動評価した結果を表 1 に示す。いずれの評価基準でも提案法が最も高い評価値を得ている。

システム	BLEU	RIBES	IMPACT
SMT訳文	42.10	0.8296	0.5072
NMT訳文1	46.52	0.8570	0.5478
提案法	47.44	0.8583	0.5495

表 1 自動評価結果

翻訳例を表2に示す。例文1から3は改善例である。例文 1 では、NMT 訳文 1 に湧き出しがあるが、SMT 訳文と提案法にはない。例文 2 では、NMT 訳文 1 に訳抜け(「望ましい」または「望まれる」)があるが、SMT 訳文と提案法にはない。例文 3 でも、NMT 訳文 1 において、「ロック解除ピン」となるべきところ「ロックピン」となり、訳抜けがある。

一方、例文 4 と 5 は改悪例である。例文 4 では、「孔径」とすべきところ、提案法では「開口」となっている。例文 5 では、名詞句として訳されるべきところ、文として訳されている。

6. おわりに

NMT の欠点の一つである訳抜けや湧き出しを防ぐ方法として、SMT の出力を用いて、NMT の出力候補の再評価を行う手法を提案した。WAT2016 のデータを用いた実験で、訳抜けや湧き出しを防ぐ効果があることが分かった。本手法は WAT2017 への筆者からの参加データにも適用しており、文献[10]で発表予定である。

NMT によって機械翻訳の精度が大幅に向上したことは事実である。特に流暢性の高い訳文が出力されるようになった。しかしながら、今回検討した訳抜けや湧き出しの問題など欠点もある。従来からある RBMT(Rule-based MT, 規則方式機械翻訳)、EBMT(Example-based MT, 用例方式機械翻訳)、SMT を NMT と組み合わせてハイブリッド方式とすることで NMT の残された欠点を除ける可能性がある。今回の提案もこのような方式の一つであると位置付けられる。

1	原文	在 再 另 一 个 实施 方案 中 , R 8 和 R 9 是
	参照訳文	更に 別の 実施 形態 では 、R 8 及び R 9 は で ある 。
	NMT訳文1	さらに もう 一 つ の 実施 形態 に おいて 、R 8 および R 9 は 、さらに 、もう 1 つ の 実施 形態 では 、R 9 および R 9 は 、さらに もう 1 つ の 実施 形態 に おいて 、さらに もう 1 つ である 。
	SMT訳文	さらに もう 一 つ の 実施 態様 に おいて 、R 8 および R 9 である
	提案法	さらに もう 一 つ の 態様 に おいて 、R 8 および R 9 は 、
2	原文	然而 , 可能 期望 在 非操作员 允许 的 情况 下 不要 发生 这样 的 更新 。
	参照訳文	しかしながら 、そのような 更新 が 、人間 の オペレータ が 許可 する まで 行わ れ ない こと が 望ま れる 場合 も ありうる 。
	NMT訳文1	しかし 、非 操作 员 が 許容 され る 場合 に は 、そのような 更新 が 不要 と なる 場合 が ある 。
	SMT訳文	しかしながら 、非 操作 员 許可 の 場合 に は 、このような 更新 が 発生 し ない こと が 望ましい 場合 が ある 。
	提案法	しかし 、非 操作 员 が 許容 され る 場合 に は 、そのような 更新 が 発生 し ない こと が 望ましい 場合 が ある 。
3	原文	在 滑动 板 35 的 一 端 形成 有 解锁 销 35 b 。
	参照訳文	スライド 板 35 の 一端 に は ロック 解除 ピン 35 b が 形成 され て いる 。
	NMT訳文1	スライド プレート 35 の 一端 に は ロック ピン 35 b が 形成 され て いる 。
	SMT訳文	スライド 板 35 の 一端 に ロック 解除 ピン 35 b が 形成 され て いる 。
	提案法	スライド 板 35 の 一端 に は ロック 解除 ピン 35 b が 形成 され て いる 。
4	原文	若 分别 浇铸 超滤层 则 会 致使 阻止 带 表皮 表面 的 形成 , 并 改变 孔径 。
	参照訳文	これ に よって スキン 面 の 形成 が 妨害 され 、UF 層 が 独立 して 流 延 され る 場合 に 得 られ る 孔 径 と は 異なる 孔 径 に なる 。
	NMT訳文1	これ に より 、超 濾 層 を 別々に キャスト する と 、表皮 表面 の 形成 が 阻止 され 、孔 径 が 変化 する 。
	SMT訳文	それぞれ キャスト 超 濾 層 と 表皮 、ベルト 表面 の 形成 を 阻止 する こと に よって 、ポアサイズ が 変更 され る こと は ない 。
	提案法	超 濾 層 を 別々に キャスト する と 、表皮 表面 の 形成 が 阻止 され 、開口 が 変更 され る 。
5	原文	微生物 检测 试剂 , 其 构造 为 检测 微生物 的 存在 ; 和
	参照訳文	微 生物 の 存在 を 検出 する よう に 構成 され た 微 生物 検出 試薬 。
	NMT訳文1	微 生物 の 存在 を 検出 する よう に 構成 され た 微 生物 検出 試薬 。
	SMT訳文	微 生物 試薬 を 検出 し 、その 微 生物 の 存在 を 検出 する よう に 構成 され る 。
	提案法	微 生物 検出 試薬 は 、微 生物 の 存在 を 検出 する よう に 構成 され る 。

表 2 翻訳例

参考文献

[1] Hiroshi Echizen-ya and Kenji Araki. 2007. Automatic Evaluation of Machine Translation based on Recursive Acquisition of an Intuitive Common Parts Continuum, *Proceedings of the Eleventh Machine Translation Summit (MT SUMMIT XI)*, Page.151-158.

[2] Minh-Thang Luong, Hieu Pham and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421.

[3] Philipp Koehn, Franz J. Och and Daniel Marcu. 2003. Statistical Phrase-Based Translation. *Proceedings of HLTNAACL 2003*, pages 48-54.

[4] Slav Petrov, Leon Barrett, Romain Thibaux and Dan Klein. 2006. Learning Accurate, Compact, and Interpretable Tree Annotation, *Proceedings of the 21st COLING and the 44th ACL*, pages 433-440.

[5] Terumasa Ehara. 2015. System Combination of RBMT plus SPE and Preordering plus SMT. *Proceedings of the 2nd Workshop on Asian Translation (WAT2015)*, pages 29–34.

[6] Terumasa Ehara. 2016. Translation systems and experimental results of the EHR group for WAT2016 tasks. *Proceedings of the 3rd Workshop on Asian Translation (WAT2016)*, pages 111-118.

[7] Toshiaki Nakazawa, Hideya Mino, Chenchen Ding, Isao Goto, Graham Neubig, Sadao Kurohashi. 2016. Overview of the 3rd Workshop on Asian Translation, *Proceedings of the 3rd Workshop on Asian Translation*, pages 1–46.

[8] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu：BLEU: a Method for Automatic Evaluation of Machine Translation, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2002, pages 311-318.

[9] 平尾努、磯崎秀樹、Kevin Duh、須藤克仁、塚田元、永田昌明：RIBES: 順位相関に基づく翻訳の自動評価法、*言語処理学会第 17 回年次大会発表論文集*、2011 年 3 月、pages 1115-1118.

[10] Terumasa Ehara. 2017. SMT reranked NMT, *Proceedings of the 4th Workshop on Asian Translation*, (to be published).

スマホやパソコンを用いる地域資料の多国語翻訳と通訳

吉川　潔
翻訳業

AAMT-MT フェア 2017 で、30 数年の翻訳業の経験と 15 年の MT 試訳結果に基づいて、外国人を防災や医療や観光のために支援する多国語翻訳と通訳を展示し実演しました。

MT の誤訳に注意して、長所を活用すると、好結果になることを説明しました。

①　雨災害対応マニュアルの多国語翻訳（全２７頁）

13 年前の水害の対策として、行政が作成した防災マニュアルを英語、韓国、中国、ロシア、スペイン、ポルトガル、インドネシア、ベトナム、タイの各国語に、ＭＴを用いて翻訳し、各々の native にチェックを依頼し OK の翻訳文が記載してあります。

例

避難勧告が発令 →

英語　An evacuation preparation was issued.

中国語　　发出疏散准备

②地震や原子力発電所の事故に備えたマニュアルの多国語翻訳

例

原子力発電所の近くで地震が発生しました　→

韓国語　원자력　발전소　근처에서　지진이　○○시　발생했습니다

③スマホの通訳機能を用いて、外国人の患者の診療を支援

下記を、スマホやパソコンの近くで言うと、両方向に通訳します。

例

体の具合が悪いですか　→

ロ シ ア 語　　Ваше　плохое　состояние

例

I am sick　　→　私は病気です

④観光用の案内表示の多国語翻訳

スキー場のコース外を滑るバックカントリーの遭難が発生しています。

対策として、警告文「バックカントリー・スキーの希望者は届書に記入してください」の翻訳を試みました。バックカントリーの英訳は backcountry ですが、back に帰るの意味があり、country に国の意味もあります。そのためにバックカントリー希望者を（帰国希望者）と珍訳する、多国語翻訳の実例もあります。

（スキー場の施設外を滑りたいスキーヤーは届書に記入してください）の趣旨で日本語の原文を書き直してからの多国語翻訳が好ましいかも知れません。

⑤観光用のカタログを、パソコンの単語登録機能を用いて多国語翻訳

MT の翻訳精度は完璧でないので、パソコンの単語登録機能の応用を提案します。

例えば、（この人形は１００円）という native チェック済みの多国語の翻訳文をパソコンに保存し、人形を他の商品、単価を他の値段に容易に変更可能に設定します。

⑥AI（人工知能）と深層学習に基づく翻訳の精度の調査

AI や深層学習により音声・画像認識技術が発展し、翻訳・通訳機能も向上しました。

それでも誤訳があるので、MT の更なる進展に誤訳調査が必須です。インターネットから利用できる翻訳や通訳の誤訳について１１０以上の文例を用いて試訳しました。

上記の資料や試訳結果を希望されるかたに、郵送や電子メールで送ります。

⑦追記

９月８日に新潟大学で ICT の大会があり、知事と市長も講演されました。AAMT 総会と同様に、上記を会場で展示し実演しました。参加者の大半が、インターネットから利用できる多国語翻訳機能を知っていましたが、スマホ通訳は初耳のようでした。中国と韓国の留学生は、日中と日韓を正確に通訳するのに驚いていました。

上記を医師や外国人支援者に説明し、次に会うと、その機能を既にインストールしています。MT 必要としている人達が地域社会にいることを痛感しました。

私は 74 才ですが、特許関連の翻訳で働いています。そこから MT に関連したデータを得て AAMT の皆様に報告しています。そして、スマホやパソコンの翻訳・通訳の欠点に注意しながら、長所を最大限に活用すると大きな成果を得る可能性のあることを地方の末端で広めることを、ライフワークとして頑張る意欲に燃えています。

イベント報告

2017年度JTF総会「基調講演」報告

目次　由美子

LOGOStar

【日時】2017年6月7日（水）

【会場】アルカディア市ヶ谷

【主催】日本翻訳連盟

【テーマ】機械翻訳技術を取り巻く最近の動向について

【登壇者】中岩 浩巳　Nakaiwa Hiromi　JTF 監事、名古屋大学特任教授、アジア太平洋機械翻訳協会 会長

※　本稿は、日本翻訳連盟が発行する『日本翻訳ジャーナル』（Web版）にも同時掲載いただいています。

つつがなく執り行われた日本翻訳連盟の総会の基調講演では、昨今目覚ましい進化を遂げている機械翻訳（MT：Machine Translation）がテーマに取り上げられた。会場が満席であったという事実からも、注目度の高さが伺えた。

2020年に東京で開催されるオリンピック・パラリンピックへ向けてさらなるグローバル化が急ピッチで推進される最中、ニューラル機械翻訳（NMT：Neural Machine Translation）技術は突如として登場し、日本語に関する品質改善がもたらされ、機械翻訳を取り巻く環境にも変化が見られている。本講演では単に機械翻訳に関する最新状況のみを紹介するのではなく、機械翻訳に馴染みがない参加者のことも考慮してシンプルに構成されていた。

まず、国際ビジネスの拡大による翻訳の需要増加や、円安に伴う訪日外国人の急増による外国人とコミュニケーションする機会の増加が紹介され、「言語バリアー」という障壁を取り除くためのMTの必要性が指摘された。そして、機械翻訳の技術について以下のように分類して紹介された。

ルールベース型機械翻訳 Rule-Based Machine Translation (RBMT)	単語直接方式	計算機で自然言語処理を行う文法論理がまだ提案されていなかった初期の方法	・表層的単語レベルの処理のみであり、構造的情報や意味情報を訳文に反映できない ・語の並び替えの規則の記述が容易でない
	変換方式	原言語を解析して構文構造を決定し、これを目的言語の構文構造に変換し、目的言語を生成	原言語の構造から目的言語の構造に直接変換するため、原言語での語順など表層的な情報や構文構造が目的言語に反映されやすい
	中間言語方式	特定の言語に依存しない中間言語へ原言語から変換し、生成部で目的言語に変換	言語対ごとに変換規則を作成する必要がある
用例を活用した機械翻訳	用例翻訳 Example-based MT (EBMT)	翻訳用例を直接活用して適切な訳語・表現を選択	
	統計的機械翻訳 Statistics-based MT (SBMT)	翻訳用例を統計的に分析し、翻訳ルールに統計情報を付与して適切な訳語・表現を選択	
	ニューラル機械翻訳 Neural MT (NMT)	翻訳用例を深層学習であるニューラルネットで学習し、学習したネットワークに基づき訳語・表現を生成	

RBMTの限界として、深い解析が翻訳精度の向上に必ずしも結びつかない、辞書やルールの構築にコストを要するため分野を特定してのチューニングが難しいといったことも言及された。また、用例を活用した翻訳システムについては、参考とする訳例がインターネットなどで数多く入手可能になったことや、計算機パワーが急増したことによって大規模な統計処理が一般的なパソコンでも実行できるようになったことなど、翻訳をめぐる環境の変化について指摘があった。

SBMTについては、原文1単語がいくつのターゲット言語に訳されるか（Fertility Probability）、原文で何番目の単語が訳文では何番目に置かれるか（Distortion Probability）といった単語の訳し方や位置、何個の単語に訳されるかといった具合に統計的に処理されるという解説があり、言語対の距離が長い日本語と英語では性能が発揮されにくいと機械翻訳業界では認識されてきたそうだ。しかしながら、翻訳前に目的言語の語順に入れ替えることで改善が図られてはいた。

NMTは1980年代にブームとなったが、計算機パワーの不足などもあり効果が限定されていた。2010年以降、GPU（Graphics Processing Unit、グラフィックス プロセッシング ユニット）の普及などにより計算機能力が急激に普及したことや、アルゴリズムの改良によって改めて脚光を浴びている。NMTは自然言語処理でも効果を発揮し始め、機械翻訳にも導入され、従来主流であったSBMTをも凌駕する性能を達成し、2016年11月にリリースされたGoogle翻訳では翻訳業界でも周知されるほどの効力を示している。中岩氏は、現在の主流である、どの単語が注目されるべきかが計算される「アテンションに基づくニューラル機械翻訳」などのNMT方式も解説した。

さらに、本年5月28日～31日にチェコ・プラハで開催されたばかりの2017 EAMT conference※で入手した最新情報として、NMTの評価が高かったことも紹介された。

※ 機械翻訳国際連盟 International Association for Machine Translation（IAMT）は、以下の3つの地域組織を擁する。

- アメリカ機械翻訳協会 AMTA: Association for Machine Translation in the Americas
- ヨーロッパ機械翻訳協会 EAMT: European Association for Machine Translation
- アジア太平洋機械翻訳協会 AAMT: Asia-Pacific Association for Machine Translation

中岩氏は機械翻訳の評価についても解説した。

目的	利用者側	・翻訳対象に適したシステムを選定 ・機械翻訳導入による効果を検証
	開発者側	・機械翻訳システムの性能検証 ・機械翻訳システムの性能向上のためのパラメータ調整用
方法	人手評価	・訳文品質評価（Fluency 流暢さ、Adequacy 適切性、など） ・翻訳効率評価（翻訳過程へのMT導入による時間短縮・人件費節減など）
	自動評価	・BLEU、METEO、RIBES

人手評価と自動評価の基準や、実際の性能比較例などもいくつか紹介された。多くの例でNMTは概して流暢さでは高い評価を得ているが、訳抜けが多く検出されていたり、BLEUでの評価が低めであることが指摘された。

NMTの利点としては、生成される文章が自然であることと紹介された。また、欠点としては、過剰訳出や訳抜けの頻出、エンジンを学習させるための時間や計算などリソースの必要性、RBMT同様に自動評価と人手評価に見られる矛盾、単語アライメントの情報の欠如が上げられた。

さらに、日本で唯一の MT に関する組織である AAMT では、研究者、製造者、ユーザを結びつける活動を展開していること、AAMT が中心となって本年 9 月 18 日〜22 日には名古屋大学東山キャンパスの豊田講堂にて MT Summit XVI が開催されることも紹介された。

イベント報告

2017年度 JAT総会「基調講演」報告

パネルディスカッション 『Is there an elephant in translation? 機械翻訳、まだ見かけていませんか?』

冨永 陽

＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊

2017年度JAT総会「基調講演（パネルディスカッション）」報告

日時：2017年6月24日（土）13：15～15：15

開催場所：国立オリンピック記念青少年総合センター

タイトル：『Is there an elephant in translation? 機械翻訳、まだ見かけていませんか?』

パネリスト：

隅田英一郎　（国研）情報通信研究機構(NICT) フェロー。先進的音声翻訳研究開発推進センター副センター長。JTF理事、アジア太平洋機械翻訳協会（AAMT）理事

河野弘毅　ポストエディット東京、JTF理事

マシュウ・ヒートン　TBSJ 最高経営責任者

トム・ガリー　東京大学大学院総合文化研究科・教養学部附属グローバルコミュニケーション研究センター教授・センター長

モデレータ：目次 由美子　LOGOStar

主催：日本翻訳者協会（Japan Association of Translators：JAT）

報告者：冨永 陽　翻訳者/翻訳チェッカー/翻訳コーディネータ

＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊

夏の始まりを感じさせる好天にも恵まれた6月の終わり、日本翻訳者協会の通常総会が執り行われた。基調講演の1つとして開催されたパネルディスカッションでは機械翻訳をトピックとし、産業翻訳においてもその利用価値が認められつつある最中、機械翻訳を多様な視点から最新状況を確認し、今後の傾向を把握するべく展開された。

各パネリストによるプレゼンテーション

機械翻訳の現況（隅田氏）

昨秋、Google翻訳の性能が飛躍したと話題になった。この流れが進めば、これまで人間翻訳（HT）に頼っていた分野でも、部分的にも機械翻訳（MT）の利用が加速するだろう。個人的には、特許、マニュアル、法務、医療、ウェブ、SNS、マーケティングなどで使われていくとみている。逆に、文芸翻訳やtranscreation、thick translation（文化的な違いなどで補足説明を要するもの）は、MTで太刀打ちできない。

これまでのMTは、第1世代のルールベース型（RBMT）の後、第2世代の統計ベース型（SMT）が登場し、昨年登場したのが第3世代のニューラル翻訳（NMT）である。これは、ヒトの神経細胞（ニューロン）をたくさんのコンピュータを使って再現したもので、神経細胞に信号が流れて複雑な計算をするように、ニューロンを模したコンピュータに信号を流し、多量の翻訳を読ませて学ばせる仕組みである。現在話題のGoogle翻訳もこのNMTであり（GNMT）、Microsoft Translatorなども同様である。今後NMTはさらに伸びると思われる。

昨年9月にGoogleがSMTとNMTを比較したところ、NMTはSMTよりも一段とHTに近い訳文だったという。ただし、分野や原文の質、言語的な距離、採点方法など諸条件によってMTの評価結果は変わるので、一概にNMTがSMTより優れているというわけでもない。NMTはHTのような流暢な訳文である反面、部分的な訳漏れが発生する欠点があり、学習にも非常

に時間がかかるという一面もある。

翻訳業界における MT 導入（河野氏）

1. ソースクライアント

今年、ある団体で MT の研修が開催された折、開催告知から 2 時間で定員に達したそうだ。MT への関心は高いようだが、日本の市場全体を見るとクライアントにおける MT 導入はさほど進んでいない。翻訳の仕様を自分で決めたり、自分でも翻訳したりする発注担当者であれば、MT や新技術の導入も比較的早いが、翻訳を外注に任せ切りの場合は、新技術導入までに時間がかかる傾向にある（CAT ツールを自分で使わないクライアントで顕著）。

2. 翻訳会社

昨日、某求人サイトを調べたところ、翻訳関係の求人 300 件余のうちポストエディット（後処理）業務の募集は 5 件。翻訳会社でも MT がまだ十分に普及していないと想像できる。GNMT の場合、セキュリティ面を心配するクライアントが多く、翻訳会社の営業担当も売上減少や社内の反対を危惧し、翻訳会社の経営陣も MT を見据えた議論ができる状況にないといった様々な理由が考えられる。私が現在知る限り、MT 導入に前向きに取り組む姿勢を示している翻訳会社は 5 社ほどだ。

3. 翻訳者

最終的な訳文がきちんと仕上がっていれば基本的に問題はないので、作業の中で MT をコッソリ使う翻訳者はいるかもしれない。だが、MT のアウトプットはそのまま納品できないし、ポストエディットで翻訳と同等以上の時間と労力を要する場合もあるので、実際のところ翻訳者も得しているとは考えにくい。

重要なのは、MT を適用する分野・範囲を見極めること。自分の請けている範囲で MT の適用範囲をマッピングできるのが望ましい。また、プリエディット（前処理）を工夫し、チェックを可能な限り自動化できれば、ブレイクスルーは起こり得る。個人レベルで上手く機能する MT プロセスを組織レベルに拡張できれば次世代の勝者になれるだろう。

翻訳会社の立場から（ヒートン氏）

TBSJ では MT ではなく TM（翻訳メモリ）を使用する案件が非常に多い。MT で HT と同等の質の高い翻訳はまだできないが、決してデメリットだけだとは捉えていない。

ここで 1 つ問題。自由の女神像の写真があり、その下に “What state is this statue in?” と書いてあったとする。この 1 文は GNMT でどんな訳になるだろう？ どの州にあるか記憶を辿る人もいると思うが、GNMT では「この像はどのような状態ですか?」と訳される。MT は原文の文字情報に縛られるが、人間なら、原文の周辺情報、自分の経験や知識も動員しながら訳すはずだ。MT は文字を訳せても内容の理解や要約はできない。

しかし、案件によっては MT で作業時間が 15~50%短縮できるとの声もある。翻訳市場は世界全体で年 5%の成長率を遂げており、翻訳するコンテンツも増えている。MT は「原文の構造が単純で質の高いポストエディットが楽にできる」「大まかな意味が分かれば良い」「大量の文書から翻訳すべき対象を絞りたい」といった案件なら使える。一方、前提情報が省略された中間報告書やメール、一文が長い法務文、原文の質が低い文書などには向かない。

大切なのは、テクノロジーの進歩を無闇に恐れず、動向を見守りながら柔軟に対応すること。そうすれば、AI（人工知能）や MT が発達するにつれて翻訳が手堅い仕事になると思う。

独立5年目ぐらいの翻訳者へ（ガリー氏）

フリーランスで翻訳を主な生業としていた時の経験を踏まえて、現在の自分から30年前の自分に伝えたい話をする。

翻訳は言語を扱う作業であるが、言語とはどのように捉えるべきだろうか? 研究者による言語の定義を歴史的に辿ってみると、単語・語句を扱うレベルから出発し、文法やルールの分析、コーパスなど実例の分析へと変化し、さらにCEFR[i]では、言語が行為や社会的文脈に根差しているからこそ意味を持つことに言及している。こうした言語観の変遷は、MTの変遷と似ているが、現在のNMTでも行為や社会的文脈が絡むことには、まだ対応できない。つまり、人間同士のやりとりが深く関わる形で働く翻訳者は生き残れるだろう。

とはいえ、AIやMTは日進月歩。開発に携わる人も優れた洞察力の持ち主ばかり。翻訳には意識や感情が絡むことやMTが文字だけに縛られる話も先ほど出たが、今後は文章だけでなく、関連する画像や動画を認識し、文脈や状況を把握した上で翻訳するMTやAIがGoogleなどから出る可能性もある。

なお、Googleにとっての翻訳は数あるケーススタディの1つに過ぎない。同社では、もっと大きなプロジェクトが進んでいるので、今後の状況によっては、アルファ碁（Google DeepMindによって開発されたコンピュータ囲碁プログラム）のようにGNMTが突然終了する、または有料化することも考えられる。

ディスカッション

（隅田英一郎＝隅、河野弘毅＝河、マシュウ・ヒートン＝ヒ、トム・ガリー＝ガ）

Q.: ポストエディットは割に合わないと言われるが本当か?

ヒ: ポストエディットは翻訳に比べて単価が低く時間もかかるが、MTの普及で新たな需要が出てくるかもしれないし、新しい仕事が生まれる可能性もある。

河: かな漢字変換のように学習してくれる機械なら、使い込むうちに自分に合ったツールになる。また、原文のプリエディットを工夫すれば、ポストエディットも最小限に抑えられる。コンマ1つで訳文がガラッと変わるのは面白い。

Q.: コンテキストを考慮するMTは登場するか?

隅: 現時点では文単位の処理だが、文章単位での処理

や原文以外の情報を取り入れるようにすれば、可能性はある。現に、日⇔英間の MT では主語や目的語を補う機能の研究が進み始めた。

Q.: MT に翻訳の仕事を奪われるのでないかと心配する人へのアドバイスは?

ヒ: 翻訳の需要やボリュームは確実に増えているが、MT 云々よりも、翻訳が楽しいと感じられなければ、仕事として続かない。

ガ: あくまで個人の選択だが、人と人との繋がりや橋渡しなどに深く関わりつつ、翻訳を通して自分を活かせる道を私は勧める。自分で訳せることに誇りを感じてほしい。

河: MT が日常的な場面で普及すれば、翻訳者は訳文の良し悪しを判断する専門的な語学メンターの役目を担うのではないか。

隅: たくさん翻訳してほしい。MT もツールとして使いこなす気概があれば、工夫次第でうまく使える。

---

[i] Common European Framework of Reference: **ヨーロッパ言語共通参照枠**

セミナー報告

2017年度ＪＴＦ第１回翻訳支援ツール説明会 報告

クラウドを活用したコミュニティ翻訳
目次　由美子
LOGOStar

＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊
2017年度ＪＴＦ（日本翻訳連盟）第１回翻訳支援ツール説明会
日時：2017年8月21日（月）14:00 ～ 16:40
開催場所：GMOスピード翻訳株式会社　大2会議室
共通テーマ：「クラウドを活用したコミュニティ翻訳」
第一部:「大学向けオープンソースソフトウェアの翻訳に適用する共通翻訳メモリの開発と実践」
第二部：「Transifexを利用したコミュニティ翻訳 ～オンライン翻訳ツールを活用した翻訳ワークフローの自動化～」
登壇者：
第一部：常盤 祐司（法政大学　情報メディア教育研究センター教授）
第二部：出口 大輔（名古屋大学　情報連携統括本部情報戦略室准教授）
報告者：目次 由美子（LOGOStar）
＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊
※　本稿は、日本翻訳連盟が発行する『日本翻訳ジャーナル』（Web版）にも同時掲載いただいています。

今回の翻訳支援ツール説明会は「クラウドを活用したコミュニティ翻訳」をテーマに二部構成とし、教育関連ソフトウェアの開発プロジェクトへ共に参加されている法政大学の常盤　祐司先生と名古屋大学の出口　大輔先生に登壇いただいた。リソースファイル翻訳のデモンストレーションもあり、スタイルガイドなど翻訳作業が抱える課題、翻訳コミュニティを形成する上での難しさなど、多岐に亘る内容が展開された。

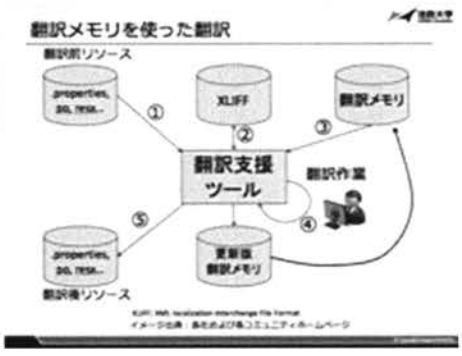
第一部　「大学向けオープンソースソフトウェアの翻訳に適用する共通翻訳メモリの開発と実践」　（常盤 祐司）

常盤先生は講演の冒頭で、今回の講演ではオープンソース ソフトウェア（Open Source Software：OSS）を翻訳することがトピックであると紹介された。
Linux に代表されるOSSは2000年以降、世界的に普及し、数多くの言語への翻訳が展開されている。これに伴い、翻訳のためのツールも数多く利用できるようになってきた。OSSの翻訳においては、ほとんどが無償であり、誰もがそれらのツールを無料で利用できることによって、基盤が作成されていると指摘があった。昨今、多くの大学で導入されている「授業支援システム」は、名前のとおり授業を支援する目的で、学生がテストを受ける、レポートを提出する、連絡事項を通知するなどの機能を有する。ウェブ アプリケーションの多くは英語で開発されており、これを日本語にするには単語レベルでの訳が数多く求められ、和訳するには多様な選択肢のなかから適切な訳を選ぶことの難しさが指摘された。例として、「exam」を「試験」とするのか、「テスト」とするのかといったことが挙げられた。こういった翻訳に取り組む上でも、i18n（Internationalization、国際化：地域化することのフレームワークのこと）とl10n（Localization、地域化：その国あるいは地域の言語および文化にあうようにシステムを実装すること）を明確に区別することが重要とのこと。

約30年前は「国際化」が実施されていなかったため、

基本ソフトやアプリケーション自体の各国語版を製作しているという状況があった。ある時期を境に、ソフトウェア製品は1つのみで言語を切り替えることができるような「国際化」が展開されてきた。たとえばInternet Explorerでは言語のシステム設定を読み取り、画面上に表示させるコンテンツの言語を選択できる。国際化の難点として、単純な翻訳ではなく「氏名」のように言語によって順番を入れ替える、「日付」のように言語ごとに表記が異なる項目についても紹介された。これらについては、各言語版の担当者によるルール作りが必要とされる。また、ダブルバイト言語である日本語は、テキストをコードとして記述する必要があり、文字化けを回避するためにプログラムの修正を要する場合があることも紹介された。さらに、これを簡易にするためのテキスト編集ツールも紹介された。

世界中の大学や営利団体などの機関が参加する教育関連ソフトウェアの開発プロジェクトである「Sakai」について、法政大学、名古屋大学を含むユーザによる会合が2012年に開催された。このとき、すでに機能として翻訳メモリも含まれており、TMXやXLIFFを取り扱っていた。当時は個人の翻訳者がスタンドアロンのパソコンでSakaiのみを翻訳していた。今後は、クラウドベースですべてのソフトウェアを対象に翻訳データを共有し、大学共通の翻訳メモリを作成するという大きな目標が掲げられた。2013年には3つのシステム（Sakai、Moodle、Mahara）を選択し、翻訳メモリの出力が可能なTransifexでの翻訳が開始された。



あるとき、翻訳メモリの効果を測定するためにSakaiの新版に対して共通の翻訳メモリを適用したところ、79%の翻訳率が見られたそうだ。共通の翻訳メモリを作る上で翻訳に「ゆらぎ」が見られる場合、これを統一させるためのツールも別途に作成した。同じソース言語に対する異なる翻訳、体言止め・用言止めなどスタイルの違いをリストアップするそうだ。もっとも違和感が感じられる翻訳としては「カタカナ」が多いものとのこと。カタカナ新語辞典での使用率に準ずる、日本翻訳連盟の「日本語標準スタイルガイド」を参考にするなどの対策が講じられている。

Sakaiを翻訳するためのコミュニティは、アメリカでの会議の発表を経由し、スペインの大学によるスペイン語への翻訳も展開している。国際コミュニティへと発展している中、SakaiのソースデータはGitHubからダウンロードできるようになり、最新データを自動的に毎日ダウンロードできる仕組みを作成するなど、さらなる発展を遂げている。

第二部 「Transifexを利用したコミュニティ翻訳 ～オンライン翻訳ツールを活用した翻訳ワークフローの自動化～」 （出口 大輔）

出口先生には、第一部で常盤先生が紹介された仕組みに対する翻訳ワークフローの自動化について講演いただいた。

出口先生は、バージョン管理ソフト「TortoiseSVN」の翻訳への取り組みが、初めてのOSS翻訳だったとのこと。2003年頃はVisual Studioのリソース エディタを使用して、翻訳対象となるリソースファイルを直接編集していた。当時は、ソフトウェアのリリースごとにすべて翻訳し直す必要があったそうだ。2004年にかけてはTortoiseSVNのi18n化が進み、「Smart Translator」というリソース編集ツールを利用して言語リソースDLLの翻訳を行う枠組みができあがった。

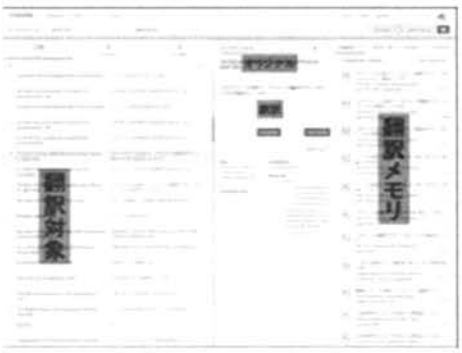
その後、独自の変換ツールによってGNU GettextのPOファイル（翻訳リソース ファイル）を生成し，より一般的な翻訳ツールが利用できる枠組みへと見直され、2010年頃までその枠組が続いた。しかしながら、複数の翻訳者の協力を継続的に得ることが難しかったとのこと。現在ではTransifexを利用してクラウド ベースでの翻訳を展開しているとのこと。

現状としては、翻訳とソフトウェア開発が分離され、翻訳作業には複数の担当者やレビュー プロセスも組み込まれている。翻訳メモリの効果として、既訳の流用や「ゆらぎ」の回避が挙げられた。しかしながら、実際のソフトウェアを見ずに翻訳作業が進むことからコンテキストに応じた翻訳が難しいこと、翻訳対象がいつ・どのように使用されるか不明であること、といった問題点が存在することが紹介された。

クラウド型翻訳基盤の特徴として、翻訳対象（リソース）をクラウド上で管理し、複数人の翻訳者がWebインターフェースなどを利用して、同時並行翻訳することが可能なプラットフォームである点が紹介された。便利な翻訳ツールも多数備わっており、翻訳メモリ、用語集、ライブ翻訳、さらに翻訳ワークフローとしてWeb APIを利用した同期システムが可能であることが紹介された。

具体例として、クラウド型翻訳基盤のパイオニア的存在のTransifex、そのほかにCrowdin、Get Localization、PO Editorというシステムの詳細も紹介された。

Transifexの特長としては、オフライン翻訳や翻訳ワークフローの自動化が可能であること、POやJava プロパティ ファイルなど複数のファイル形式がサポートされているほか、オープンソース プロジェクトについては無料で使用できること（翻訳者としてログインする場合は無料）が紹介された。担当者に対するロール（翻訳者、レビュアー、コーディネータなど）が複数用意されていること、翻訳のバージョン管理が可能であることなどにも言及があった。



Sakaiの翻訳プロジェクトでは10万以上のファイルがあり、リビジョンも30万回を超えている。翻訳対象はJavaのプロパティ ファイルに記述されており、たとえば「chat.properties」というファイルに対しては「chat_ja.properties」を作成し、内部リソースを日本語に翻訳する。

翻訳を開始するまでのフローは、GitHubから更新箇所をダウンロード → データ変換（リソースファイルからPOへ）→Transifexツールでのデータ同期→Transifexでの翻訳、の4段階である。Transifexでの翻訳以外のプロセスは自動的に実行されるため、翻訳者自身による煩雑な手作業は不要とされている。

なお、データ変換時には翻訳対象のプロパティ ファイルがモジュール内で束ねられ、単一のPOファイルへ変換される。翻訳には、必ずしもTransifexを使用する必要はないそうだ。

これに続くデータ同期のプロセスでは、サーバとクライアント間のリソースが同期される。翻訳プロジェクトの初期化および認証情報が設定されており、言語jaの翻訳済みリソースのみをダウンロードするなどを実行するための独自のプログラムとなっている。

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」は、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第二十一回を迎えることができました。今号では、個人会員一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

会員名

加島　広基

自己紹介

私は２００２年より特許事務所に勤務し、機械分野全般の国内特許出願、内外特許出願、外内特許出願の業務に携わっております。２０１２年には現事務所のマクスウェル国際特許事務所を化学分野の弁理士と共に開設し、現在は特許事務所の共同経営を行っております。

現在の業務としては、国内特許出願に係る明細書作成業務とともに、アメリカ、ヨーロッパ、中国等への外国特許出願業務を数多く手がけております。

２０１７年６月に行われた JTF 定時総会における中岩浩巳先生の「機械翻訳技術を取り巻く最近の動向について」という講演にて機械翻訳に本格的に興味を持ち、この度ＡＡＭＴの新規会員になりました。どうぞよろしくお願いいたします。

MT/翻訳とのかかわり

外国特許出願業務を行うにあたり、日本語明細書を翻訳して英文明細書を作成する必要がありますが、従前ではこのような日英翻訳作業は翻訳会社や個人翻訳者に依頼しており、納品された英文明細書をチェックするのが特許事務所の役割でした。

これに対し、２０１６年秋頃よりGoogle社がニューラル機械翻訳のサイトを公開してから私も試しに日本語明細書の一部を機械翻訳で英訳してみましたが、昔の機械翻訳と比べてその翻訳精度の高さに驚かされました。日本語の一文が５０語から１００語程度と比較的短いものであれば、ニューラル機械翻訳された英文は多少の修正を加えるだけでほぼ外国特許出願に係る英文明細書としてもそのまま使えるほど、その精度は向上していると思います。なお、それよりも長い日本語の文章であれば複数に区切るといった作業が必要になりますが、それでもこのような「文章を上手に区切る」スキルを向上させると、日本語の長文に対して驚くべき短時間で高品質の翻訳を行うことができると考えております。

このように、特許業界における明細書の翻訳のあり方も、ニューラル機械翻訳技術の近年の驚異的なレベルアップとともに、大きく変わってくるのではないかと考えております。翻訳エンジンを活用した作業となると、「日本語を見ながら英文をタイプする」という従来の方法から「日本語を何度もいじることにより最適な英文を見いだす」という方法に変わり、下手をしたら将来的にはニューラル機械翻訳に適した日本語の文章を作ることが翻訳業にとって一番大事になってくるかもしれません。

特許事務所側の立場としても、このようなニューラル機械翻訳の技術の進展に少しでもキャッチアップできればと考えております。幸いにも「日本語明細書を作成する」ことと「日英翻訳を行う」ことの両方の立場を兼ね備えておりますので、ニューラル機械翻訳の技術そのものもさることながら、このような機械翻訳に適した元の文章をどう作っていけばいいかということも追求していければと思います。

ＭＴおよび翻訳業界に期待すること

今のニューラル機械翻訳は非常にレベルが上がっており実用化に耐えうる品質に限りなく近づいてきているかと思いますが、それでも不便に感じることは多々あります。特許明細書の翻訳という観点からは、「表現のゆらぎ」が残るということでしょうか。特許明細書では同じ名詞、同じ動詞はできるだけ同じ用語で翻訳が行われるべきですが、ニューラル機械翻訳はその特性からか、この翻訳のゆらぎ防止には対応できていないように見受けられます。特許明細書では使われる用語が比較的限定されておりますので名詞については元の文章の単語を予め一括変換しておき日本語と英単語が混在する文章をニューラル機械翻訳にかけるという手法が適切かと思われますが、動詞のゆらぎについてはなかなか対処が難しいと実感しております。このような表現のゆらぎが解消されるような翻訳エンジンがあれば、特許業界にとっては非常にありがたいことではないかと思います。

このように、ニューラル機械翻訳や翻訳業界に期待することといえば、より実用に即した翻訳エンジンの開発や情報の提供にありますが、私もユーザーとしての立場から要望を出していきニューラル機械翻訳の技術の進歩に少しでも貢献していくことができればと思っております。とりわけ、翻訳エンジンの開発を行う企業や個人の

方に対して、ユーザーとしてフィードバックを行うことができると思いますので、何らかの形でジョイントしていければと考えております。もしご興味があればいつでもご連絡をお待ちしております。

ＡＡＭＴへの要望

上述したようにニューラル機械翻訳の技術の進展に少しでもキャッチアップできればと考えておりますので、最新の情報を定期的に提供していただければと思っています。また、私にとってはニューラル機械翻訳の分野はまだまだ未踏の領域ですが、様々な会員様と交流を行うことができればありがたいです。とりわけ特許業界ではこれからの数年で明細書の翻訳のあり方も大きく変わってくると思われますので、ＡＡＭＴが主導的な立場で引っ張っていくことを望みます。

シンポジウム案内

「第27回JTF翻訳祭」のお知らせ

高橋 聡
一般社団法人日本翻訳連盟　JTF専務理事・第27回翻訳祭企画実行委員長

<第27回JTF翻訳祭開催のご案内>

日本翻訳連盟は、来る11月29日（水）、恒例の「JTF翻訳祭」を開催いたします。日時、場所、概要は後述するとおりです。四半世紀を超えて毎年、翻訳祭を開催してこられたのは、AAMT様をはじめとする各関連団体様のご協力のたまものです。今年も、昨年どおり24セッションの「講演・パネルディスカッション」を用意するとともに、「翻訳プラザ（展示会）」を終日開催し、並行して「プレゼン・製品説明コーナー」をお届けします。夕刻からは、参加者どうしの交流および情報交換の場として、「交流パーティー」に皆さまをお迎えいたします。

翻訳祭は、翻訳者や翻訳会社のみならず、クライアント、翻訳支援ツールメーカー、翻訳スクールなどの業界関係者、そしてAAMT様など協力団体が一堂に会するイベントです。翻訳祭を通して、参加者の皆さまに新たな人脈形成やビジネスの機会が生まれることを願っております。

JTF翻訳祭にぜひともご来場くださいますよう、よろしくお願い申し上げます。

【日時】
2017年11月29日（水）9:30～20:30（開場9:00）

【場所】
「アルカディア市ヶ谷（私学会館）」
※JR・地下鉄「市ヶ谷」駅より徒歩2分

【主催・運営】
一般社団法人 日本翻訳連盟
第27回JTF翻訳祭企画実行委員会

【後援】
アジア太平洋機械翻訳協会（AAMT）、経済産業省、他

【概要】
今年の翻訳祭では、「きわめよう、それぞれの道～つなげよう、言葉の世界」というテーマを掲げました。翻訳を取りまく環境は、社会的にも技術的にも大きく変わりつつあります。その中で、環境翻訳者、翻訳会社、クライアント、翻訳支援ツールメーカーなど業界の全員が、それぞれの技能を極めていく必要があることは言うまでもありません。しかし、言葉を扱う者という同じ立場で、それぞれが極めた道をひとつにつなげていくことが、業界全体の発展につながっていくのではないか――そんな思いを込めたテーマです。

昨年は、「翻訳の原点」を考えるということを主眼として、古谷祐一委員長のもとに個人翻訳者が集まり、「何か違う翻訳祭」をめざしました。今年めざすのは、「もっと違う翻訳祭」です。

マルチトラックの全24セッション、展示会、製品説明コーナーという構成は変わりませんが、セッションでは初めての試みとして「1人10分のミニ講演会」を2コマ設定しました。少しでも多くの業界関係者が発信する場を設けようという意図です。その他の22セッションについては、昨年と同じく個人翻訳者向けコンテンツを充実させる一方、翻訳会社やソースクライアント向け、また通訳関連のセッションなど、広く業界関係者のお役に立てるような内容をそろえました。AAMT様からは、中岩会長と長瀬様、理事の隅田様がご登壇くださいます。

また、業界の製品・ソリューションに関する最新情報をゆっくりお聞きいただくために、プレゼン・製品

説明コーナーを別会場に移しました。あわせて、展示会の出展企業も増える予定です（AAMT 様のブースもございます）。

昨年以上に「何か違う」翻訳祭、参加者全員が業界のこれからを一緒に考える翻訳祭に、ぜひとも足をお運びください。

【総来場者数（見込み）】
1,000 名（前年実績：1,000 名）

●プログラム
・開場／受付開始 9:00～
・トラック 1～6（6 会場）
9:30～18:00、全 24 セッション
・プレゼン・製品説明コーナー（1 会場）
9:30～17:45、全 8 セッション
・翻訳プラザ（展示会）9:00～17:30
・交流パーティー 18:30～20:30

●お申し込み
下記サイトからお申し込みください。
https://www.jtf.jp/festival/festival_top.do

●翻訳プラザ（展示会）
出展 44 社（2017 年 9 月現在）
翻訳プラザは、翻訳業界の企業が参加する展示会です。来場者と出展企業が交流し、情報交換する場としてご利用いただけます。AAMT 様も出展されます。展示会会場（3 階）に、ぜひお立ち寄りください。
＜展示内容＞
・翻訳支援ツールメーカーの製品デモ
・翻訳会社との翻訳相談（翻訳発注・翻訳者登録等）
・翻訳学校の翻訳講座紹介
・出版社の翻訳書籍販売

●プレゼン・製品説明コーナー
今年は「プレゼン・製品説明コーナー」として専用の会場を用意しました。落ち着いた環境で、業界の最新情報をお聞きください。セッション数も 8 企業に増やしました。
・1 セッション 60 分／8 セッション
※翻訳プラザ、プレゼン・製品説明コーナーへの参加は無料です（事前登録不要）。

●交流パーティー
情報を交換し、これからの業界の展望などを語り合える絶好の機会です。「ほんやく検定 1 級合格者の表彰式」も行います。お仕事につながる人脈作りのために、また最新の業界動向をつかむために、ぜひご参加ください。

＜ボランティアスタッフ募集中！＞
募集内容は、「受付」「会場司会」「会場アシスタント」「書記」です。ボランティアスタッフとしてご協力くださる方の「講演・パネルディスカッション」参加料は無料となります（パーティーは有料）。詳細は JTF ウェブサイトに随時掲載いたします。

【お問い合わせ】
一般社団法人　日本翻訳連盟　事務局
http://www.jtf.jp　E-mail:info@jtf.jp

第 27 回通常総会および関連行事の報告

AAMT 事務局

当協会の第 27 回通常総会が 2017 年 6 月 14 日（水）13 時より・ホテルアジュール竹芝にて開催された。総会後、招待講演会、さらにモデレーテッド　セッションが行われた。そして第 12 回 AAMT 長尾賞授与式と第 4 回 AAMT 長尾賞学生奨励賞授与式と、引き続いての AAMT 長尾賞学生奨励賞および長尾賞受賞者による記念講演会が盛況のうちに行われた。各参加者数は下記に示す通りである。

総会の参加者：個人会員：出席 22 名、委任状参加者 19 名、
　　法人会員：出席 23 法人（法人会員参加者 名）、委任状参加 3 法人

報告会・展示会・講演会・AAMT 長尾賞関連の式及び講演会：
　　個人会員：　28 名
　　法人会員：　58 名
　　会員外一般：20 名
　　合計：　　106 名

懇親会：
　　個人会員：　18 名
　　法人会員：　34 名
　　会員外一般：　4 名
　　合計：　　56 名

第 27 回通常総会

1. 開会の辞
2. 会長挨拶　　名古屋大学・中岩浩巳
3. ご来賓挨拶
4. 出席会員の確認
5. 議案
   第 1 号議案　2016 年度事業報告（案）　第 2 号議案　2016 年度決算報告（案）
   第 3 号議案　2017 年度事業計画（案）　第 4 号議案　2017 年度収支予算（案）
   第 5 号議案　監事人事(案)　　その他・会員提案事項
6. 閉会の辞

招待講演

・「多言語音声翻訳システムの利活用実証について～現状と活用～」

吉田麻衣子

（株式会社リクルートコミュニケーションズ　ソリューション開発局

ソリューション開発部　グローバルコミュニケーション推進グループ）

・「ニューラル機械翻訳ってどうなの？」

中澤敏明

（国研　科学技術振興機構）

・「機械翻訳の収益化の可能性」

佐藤弦

（SDL ジャパン株式会社）

ＡＡＭＴからのお知らせ

・「MT Summit 紹介」　中岩　浩巳（AAMT 会長）

展示会

【法人会員による展示】

みんなの自動翻訳@TexTra のご紹介

国立研究開発法人 情報通信研究機構

機械翻訳エンジン Globalese NMT の紹介

株式会社川村インターナショナル

自社製タイ語翻訳エンジンのデモ展示

株式会社 クロスランゲージ

機械翻訳と十印の関わりの変遷、機械翻訳ソリューション「T-tact」について

株式会社 十印

クラウド型翻訳支援ツール「ヤラクゼン」

八楽株式会社

インバウンド翻訳ソリューション

株式会社 高電社

日本語が話せない患者とのコミュニケーションを支援するハンズフリー音声翻訳端末

富士通株式会社

【個人会員による展示】

「みんなの翻訳実習」

プロジェクト翻訳時代の翻訳教育支援プラットフォーム

東京大学・NICT・立教大学・リーズ大学・関西大学・NII

影浦峡・藤田篤・内山将夫

SMT を用いた NMT の改良

江原自然言語処理研究室

江原暉将

機械翻訳の有効な活用としての防災マニュアルの多国語翻訳

吉川 潔

AAMT 各委員会活動報告会

（本年度の委員会活動報告は展示形式にて行われました。）

1. 機械翻訳課題調査委員会	委員長	長瀬 友樹	（富士通研究所）
2. AAMT/Japio 特許翻訳研究会	副委員長	宇津呂 武仁	（筑波大学）
3. インターネットワーキンググループ	リーダー	富士 秀	（富士通研究所）
4. 編集委員会	委員長	宇津呂 武仁	（筑波大学）

2017 年度総会後の組織・人事について

人事

- 会長、副会長、委員長、ワーキンググループリーダー、事務局長は下記の通りとする。
- （敬称略：50 音順）

会長	中岩　浩巳	（名古屋大学）
副会長	飯田　仁	（東京工科大学名誉教授）
	小谷　泰造	（株式会社インターグループ）
理事	安達　久博	（株式会社サン・フレア）
	井佐原　均	（豊橋技術科学大学）
	石崎　俊	（慶應義塾大学名誉教授）
	宇津呂　武仁	（筑波大学）
	熊野　明	（東芝ソリューション株式会社）
	黒橋　禎夫	（京都大学）
	坂本　義行	
	隅田　英一郎	（国立研究開発法人　情報通信研究機構）
	辻井　潤一	（国立研究開発法人 産業技術総合研究所）
	田中　英輝	（ＮＨＫ放送技術研究所）
	長尾　真	（国際高等研究所）
	永田　昌明	（日本電信電話株式会社）
	中村　哲	（奈良先端科学技術大学院大学）
	樋口　隆一	（富士通株式会社）
	守屋　敏道	（一般財団法人日本特許情報機構）
	横山　晶一	（山形大学名誉教授）
	Key-Sun Choi	（韓国 KAIST）
	Virach Sornlertlamvanich	（タイ SIIT）

退任

星野　正広　（富士通株式会社）

（所属元の異動のため、2016 年度を持って本協会理事を退任）

監事	古谷　祐一	（一般社団法人日本翻訳連盟・ＧＭＯスピード翻訳株式会社）
	福嶋　健一郎	（株式会社クロスランゲージ）

退任　種子田　暁夫　（日本電気株式会社）

（所属元を 2017 年 4 月末日にて退任のため本協会監事を退任）

機械翻訳課題調査委員長	長瀬　友樹	（富士通研究所）
AAMT/Japio 特許翻訳研究会委員長	辻井　潤一	（産業技術総合研究所）
編集委員長	宇津呂　武仁	（筑波大学）
インターネットワーキンググループリーダー	富士　秀	（富士通研究所）
事務局長	神崎　享子	（豊橋技術科学大学）

組織

ＡＡＭＴ機械翻訳課題調査委員会

退任：　秋葉 泰弘（NTTコミュニケーション科学基礎研究所）
　　坂本　義行

新任:　半田 峻一（株式会社インターグループ）
　　永田　昌明（NTTコミュニケーション科学基礎研究所）

その他委員会、ワーキンググループとも現状どおりとする。

以上

第 12 回 AAMT 長尾賞・第 4 回 AAMT 長尾賞学生奨励賞授与式
及び記念講演会

第 12 回 AAMT 長尾賞
受賞者：八楽(やらく)株式会社
受賞理由：

機械翻訳オンライン CAT ツール「ヤラクゼン」を開発し、翻訳の専業者でない一般の企業人が英文メールや社内ドキュメント、各種マニュアルなどを扱う際に有用となるツールを提供した。大企業を含む 100 社以上に数万人に上る利用者登録があり、同ツール提供による機械翻訳サービスの利用実績はビジネスでの有用性を示すものであり、機械翻訳の実用化を更に発展させる点で長尾賞にふさわしい。

選考委員長：飯田 仁（東京工科大学名誉教授）
選考委員　：Key-Sun Choi(韓国 KAIST)　宇津呂 武仁(筑波大学)
黒橋　禎夫(京都大学)　隅田 英一郎（情報通信研究機構）
推薦者　：河野弘毅（ポストエディット東京）
同意人　：谷口信也（株式会社コングレ・グローバル・コミュニケーションズ）

授賞式

第4回 AAMT 長尾賞学生奨励賞

受賞者：John Richardson 京都大学（現在 Google）

受賞対象論文：

：Improving Statistical Machine Translation with Target-Side Dependency Syntax（京都大学 博士論文）

本論文は、京都大学情報学研究科 黒橋禎夫教授の指導のものに、博士論文としてまとめられ、2016 年 9 月に京都大学から博士（情報学）の学位を受けたものである。

受賞理由：

本論文は、木構造から木構造への変換に基づく翻訳(Tree-to-Treetranslation, T2T)において、目的言語の依存構造の情報を用いて、依存構造木言語モデルによる翻訳候補のリランキングや、英語の副詞のような挿入位置の自由度が大きい要素の挿入位置を決定する方法を提案し、複数の言語対において翻訳精度を改善できることを示した。また翻訳候補のリランキングや自由度が高い要素の位置決めの問題に RNN 言語モデル(Recurrent Neural Network language model)を用いることにより翻訳精度をさらに改善できることを示した。本研究の成果は KyotoEBMT として公開されている。

このように本論文は、木構造から木構造への変換に基づく翻訳および用例翻訳(Example-Based Machine Translation, EBMT)に関する長年の研究の到達点を示し、かつ、ニューラル機械翻訳の技術を取り入れた T2T/EBMT の新しい研究の方向性を示す優れた論文であり、AAMT 長尾学生奨励賞にふさわしいと考える。

選考委員長：永田昌明（日本電信電話株式会社）
選考委員　：
鄭 育昌　　（富士通研究所）
藤田　篤　　（情報通信研究機構）
釜谷 聡史　（東芝 インダストリアルＩＣＴソリューション社）

推薦者　　：黒橋禎夫教授（京都大学）

授賞記念講演

協会活動報告

(2017 年 5 月～2017 年 8 月)

機械翻訳課題調査委員会

2017 年 5 月 20 日（2017 年度　第 2 回）

① 各 WG の活動について（各 WG に分かれて議論）

(WG1、WG2)

・MT フェア 2017 の準備

展示コーナー応募状況

WG1&3 の展示内容の候補

部屋割り

時間割り

やるべきことのリスト

・委員会議事録共有

個人用と共有用と 2 つ並行してもつのは煩雑。

サイトにアクセスして、ローカルにダウンロードし、必要があれば編集してアップロードして使うこととする。

・MT Summit

中岩会長より

来週の名古屋での人工知能学会、その次の週のプラハでの EAMT で MT Summit のパンフレットを配付し宣伝する。

② 全体会議

・WG1&2, WG3 の議事内容の紹介

・MT フェア

３人目の講演者決定

当日必要なもののリスト

・MT Summit

論文募集

Workshop と tutorial は決定。

Technical showcase のお知らせ 文面完成、これから送付

Sponsor および後援も増えつつある。

・JTF 総会(6 月 7 日)

中岩会長講演

「機械翻訳技術を取り巻く最近の動向について」

2017年6月16日（2017年度　第3回）

① 各WGの活動について（各WGに分かれて議論）

(WG1、WG2)

・前回議事録確認

畠中先生によるテストセットの分析→進捗報告なし。

委員会議事録共有→進展なし。

サイトにアクセスして、ローカルにダウンロードし、必要があれば編集してアップロードして使うこととする。

・委員会によるアンケート

昨年実施のアンケート集計結果報告。

今年度のアンケート、MT Summitで配布するアンケート

今年度活動計画にあった詳細Webアンケート実施等について検討した。

・AAMT MTフェアの振り返り

ブースの配置に工夫が必要。

研究者による発表機会を設けたのは良かった。

・2017JTF翻訳祭企画

今年は11月29日(水)に決定。

・今後の委員会における定期的な勉強会

有識者から2回に1回程度の頻度で再開する。

② 全体会議

・WG1&2, WG3の議事内容の紹介

・MT Summit準備状況

スポンサーが次々と増えている。

招待講演決定した

研究論文(research track)は39本。

user枠は21の応募→全部採用。

Workshop 一部はすでにCFPが出ている。

・6月7日のJTF総会での基調講演報告

EAMTの報告を中心に話した。

2017年7月28日（2017年度　第4回）

① 全体会議

・JTF翻訳祭（2017年11月）の準備

展示（翻訳プラザ）は申込みが受理された。

セッションへ中岩会長から申し込みを行い受理された。

最近話題のNMTに触れるが、JTF向けにはMTの使い方をメインに検討する。

・アンケート結果の集計報告

翻訳メモリはTradosの次はmemsourceの利用が多い。

回答者の過半がGoogle翻訳を使っている。

回答数が40弱といつもより少なかった。

・Webでのアンケート実施

過去のアンケート回答者へ送る。

ＪＴＦのアンケート調査にMT調査を入れてもらうことができるとよい。

他のアンケートチャネルも検討する。

参考になりそうな海外のアンケートの紹介。

・MT Summitの準備

AAMTブース

受付で渡す配布物について

Ｏｎｅｄａｙ料金の設定について

プロモーションについて

・ウィーンISO/TC37国際会議の報告

2017年8月25日（2017年度　第5回）

全体会議

・MT Summitの準備状況、課題の共有、アクション

・JTF翻訳祭（2017年11月）のセッション、展示について

インターネットWG

① 総会後のAAMTホームページの更新

② MT SummitにかんするＷＥＢ対応

編集委員会

No.66(2017年10月発行)の原稿依頼、編集作業を行った。

AAMT/Japio 特許翻訳研究会

2017 年 5 月 12 日(金)(2017 年度　第 1 回)

1. 議事録の確認
2. 愛媛大学　二宮崇先生による「愛媛大学におけるニューラル機械翻訳の取り組み」のご発表
3. MT サミットについて
4. 昨年度の報告書の内容紹介
5. 第二回研究会での「新年度の活動計画紹介」の依頼について
6. 次回の開催について
   - 今後の研究会開催の日時予定（場所）
   - 次回研究会予定　(平成 29 年 6 月 16 日　開催予定)
   - 主な議題
   - 新年度の活動計画書

AAMT/Japio 特許翻訳研究会　拡大評価部会

2017 年 5 月 12 日(金)(2017 年度　第 1 回)

1. 各グループの活動発表
   - 自動評価グループ
   - 人手評価グループ
   - テストセットグループ

2.その他

3. 次回開催日予定　10 月 13 日開催予定

AAMT/Japio 特許翻訳研究会

2017 年 6 月 16 日(金)(2017 年度　第 2 回)

1. 議事録の確認
2. NHK 放送技術研究所　後藤功雄様　「ニューラル機械翻訳での訳抜けした内容の検出」についてのご発表
3. 各委員、今年度研究計画のご報告
4. PLST2017 関連
5. 次回の開催について
   - 今後の研究会・拡大評価部会開催の日時予定（場所）
   - 次回研究会予定　(平成 29 年 7 月 21 日開催予定)
   - 主な議題

AAMT/Japio 特許翻訳研究会

2017 年 7 月 21 日(金)(2017 年度　第 3 回)

1. 議事録の確認
2. 日本特許情報機構（Japio）による疑問提供
3. PLST2017 関連
4. 次回の開催について
    - 今後の研究会・拡大評価部会開催の日時予定（場所）
    - 次回研究会予定　（平成 29 年 10 月 13 日開催予定）
    - 主な議題

編集後記

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

AAMT ジャーナル 66 号をお送りします。

今号の巻頭言は、立教大学の Anthony Hartley 先生より，御寄稿を頂きました。

AAMT の活動として、今号に先立ちまして、2017 年 6 月に開催されました総会におきまして、科学技術振興機構の中澤敏明様、SDL ジャパン株式会社の佐藤弦様より貴重な御講演を賜りましたが、今号におきましては、御講演内容についての貴重な御寄稿を頂きました。

あわせて、AAMT 長尾賞の選考結果を受けまして、受賞者である八楽株式会社湊幹様より受賞内容についての貴重な御寄稿を頂きました。また、AAMT 長尾賞学生奨励賞の選考結果を受けまして、受賞者である京都大学(現 Google Japan) John Richardson 様より、受賞理由となった"Improving Statistical Machine Translation with Target-Side Dependency Syntax"についての紹介記事を御寄稿頂きました。

さらに、総会会場におきまして、「AAMT 機械翻訳フェア（MT フェア）2017」と題しまして、昨年に続き、機械翻訳技術提供各社、ならびに、機械翻訳課題調査委員会からの展示の試みを実施したのを受けまして、展示コーナーへの参加団体より、情報通信研究機構 内山将夫様、富士通株式会社様、「みんなの翻訳」プロジェクトの東京大学 影浦峡先生他、江原暉将様、吉川潔様から、それぞれ、展示参加報告を御寄稿いただきました。

また、6 月 7 日(水)に開催されました日本翻訳連盟(JTF)総会にて中岩浩巳 AAMT 会長が行いました基調講演の報告を LOGOStar 目次由美子様より御寄稿頂きました。

今号が会員の皆様のお手元に届くその少し後、来る 11 月 29 日(水)に、AAMT が後援をします第 27 回 JTF 翻訳祭が開催されます。日本翻訳連盟(JTF)理事・第 27 回翻訳祭企画実行委員長 高橋聡様からは、この JTF 翻訳祭のお知らせを御寄稿いただきました。

その他、6 月 24 日(土)に開催されました 2017 年度 JAT 総会「基調講演」の報告を冨永陽様より御寄稿頂きました。LOGOStar 目次由美子様からは、8 月 21 日(月)に開催されました 2017 年度 JTF 第 1 回翻訳支援ツール説明会の参加報告を御寄稿頂きました。

また、「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、シンポジウムの開催など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

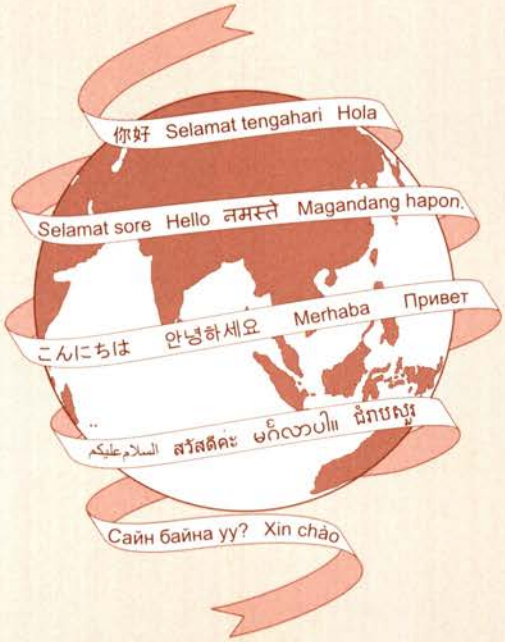
以下の通り、アジア太平洋機械翻訳協会の会員申し込みを致します。

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

AAMT



AAMTジャーナル No.66

発行：アジア太平洋機械翻訳協会（AAMT）
ホームページ：http://www.aamt.info
住所：〒171-0014　東京都豊島区池袋2−55−2鈴木ビル3階
(株)日本システムアプリケーション内
phone：03-5951-3961　fax：03-5951-3966
編集委員会：宇津呂 武仁　小谷 克則　大倉 清司
阿部 さつき　釜谷 聡史　河野 弘毅　目次 由美子
表紙(図部分)デザイン：阿部 さつき
事　務　局：神崎 享子　荻野 孝野
印　刷　所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)
c/o Japan System Application Co., Ltd.
Suzuki Building 3F 2-55-2, Ikebukuro, Toshima-ku Tokyo 171-0014, JAPAN
aamt-info@aamt.info