別添５

機械翻訳の利用及び将来性に係る調査結果

機械翻訳の利用及び将来性に係る調査結果　目次

文献等調査及びヒアリング調査

文献、書籍、Ｗｅｂサイト等の調査及びヒアリング調査により機械翻訳システムの利用及び将来性について参考となる情報を収集、整理、分析した。

1. 民間企業によるＷｅｂ翻訳サイトにおける機械翻訳エンジンの利用状況

Ｗｅｂ翻訳サイトにおいて使用されている機械翻訳エンジンについて調査した。調査結果は表１のとおりである。最終確認日は 2017 年２月 15 日である。

表１　Ｗｅｂ翻訳サイトで使用されている機械翻訳エンジン

#	Web 翻訳サイト	機械翻訳エンジン
1	Google 翻訳[1]	Google 独自のもの
2	Yahoo!翻訳[2]	クロスランゲージ製及び PROMT 製
3	みんなの自動翻訳[3]	みんなの自動翻訳@TexTra®
4	Bing 翻訳[4]	Microsoft Translator
5	Infoseek マルチ翻訳[5]	クロスランゲージ製
6	So-net 翻訳[6]	クロスランゲージ製
7	エキサイト翻訳[7]	高電社製

2. ヒアリング調査

機械翻訳システムの利用（提供）状況について、機械翻訳の機能や性能等詳細を確認するため、ヒアリング調査を実施した。調査対象は、前項１で調査した各翻訳サイトで使用されている翻訳エンジンの提供元、及びその他機械翻訳システム（サービス）を提供している者、計６者である。調査は 2016 年 12 月から 2017 年２月にかけて実施した。調査結果は以下のとおりである。

1 https://translate.google.co.jp/
2 http://honyaku.yahoo.co.jp/
3 https://mt-auto-minhon-mlt.ucri.jgn-x.jp/
4 https://www.bing.com/translator/
5 http://translation.infoseek.ne.jp/
6 http://www.so-net.ne.jp/translation/
7 http://www.excite.co.jp/world/

2.1　回答者のプロフィール

回答結果を図２．１に示す。部長級が一番多く 33%となっている。次に課長級、一般職、経営層が 17%程度、係長級、その他が０%となっている。また、全回答者のうち、17%はプロフィールについての回答が無かった。

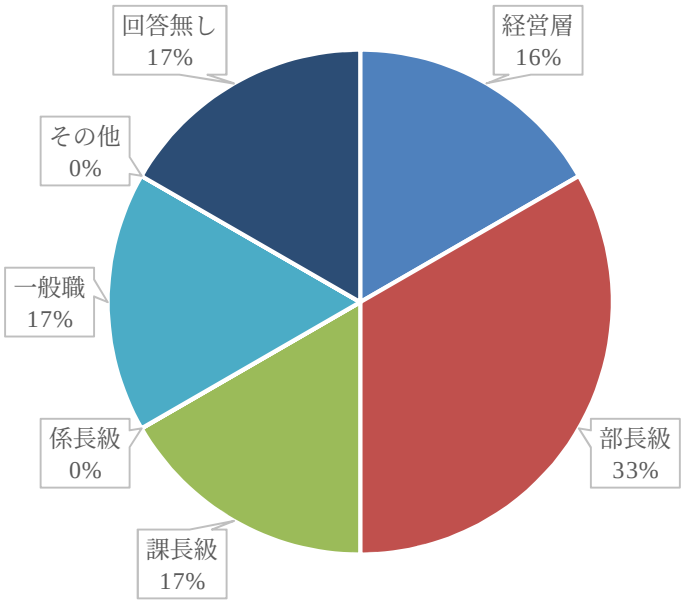

図２．１　回答者のプロフィール

2.2　機械翻訳システムの基本構成について

2.2.1　翻訳方式

回答結果を図２．２．１に示す。ルールベース翻訳を採用している者が一番多く 67%となっている。次に統計翻訳の 50%、用例ベース翻訳の 33%、ニューラル翻訳、統計的後編集の 17%となっている。ハイブリット翻訳、その他が０％となっている。

従来のルールベース翻訳、用例ベース翻訳、統計翻訳に加え、ニューラル翻訳を採用している事業者がいることも確認できた。

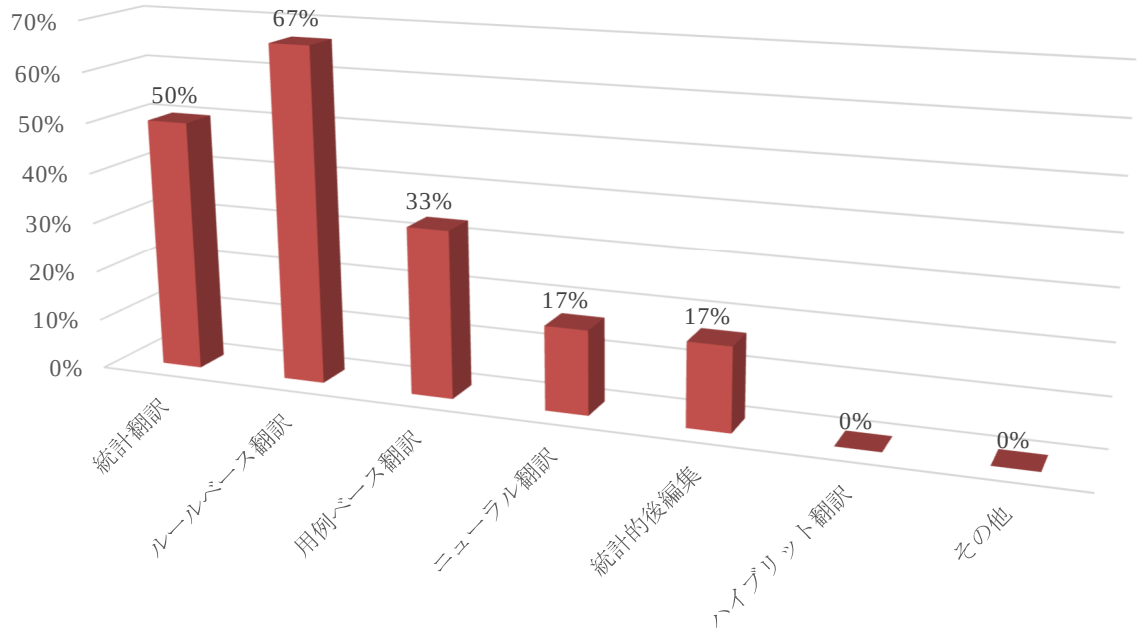

図２．２．１　翻訳方式

※複数回答あり

※統計的後編集とは、ルールベース翻訳結果を統計翻訳で修正する方式のこと

2.2.2　翻訳言語

回答結果を図２．２．２に示す。英日（英語から日本語への翻訳。以下同様。）、日英、中日、日中、韓日、日韓が一番多く 100%となっている。次に日葡の 83%、独日、日独、仏日、日仏、西日、日西、葡日、泰日、日泰が 67%、伊日、日伊、露日、日露、越日、日越が 50%、尼日、日尼が 33%、その他が 17%となっている。

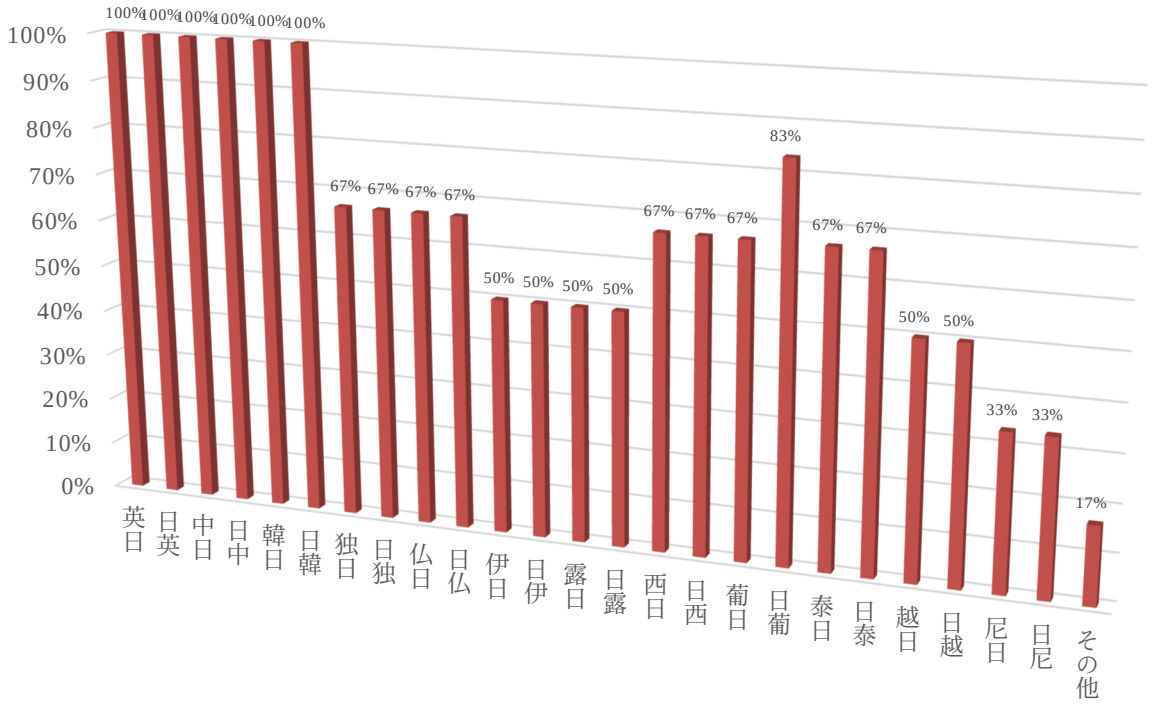

図２．２．２　翻訳言語

※複数回答あり

※西：スペイン

葡：ポルトガル

泰：タイ

越：ベトナム

尼：インドネシア

2.2.3　各翻訳方式における対応言語

（1）統計翻訳の対応状況

統計翻訳を行っている者について、各言語の対応状況を図２．２．３－１に示す。英日、日英、中日、日中、韓日、日韓、独日、日独、仏日、日仏が一番多く 100%となっている。次に伊日、日伊、露日、日露、西日、日西、葡日、日葡、泰日、日泰、越日、日越、尼日、日尼が 67%、その他が 33%となっている。

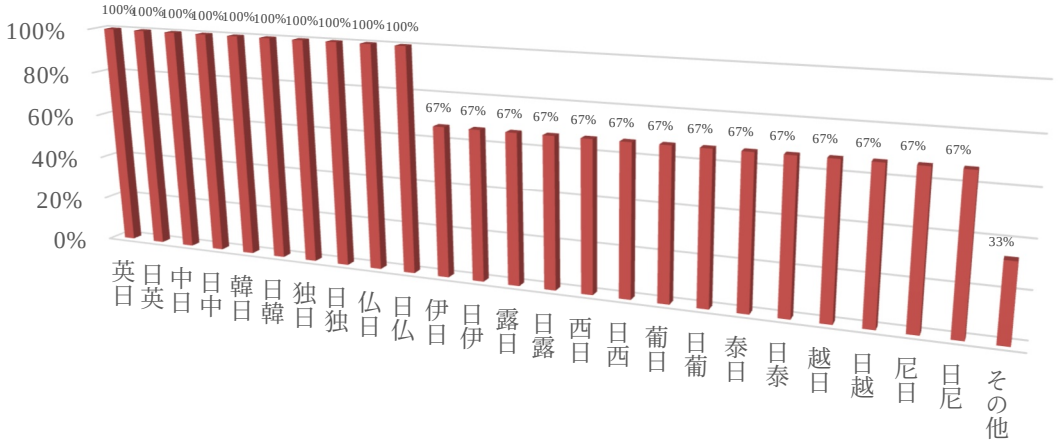

図２．２．３－１　統計翻訳の対応状況

（2）ルールベース翻訳の対応状況

ルールベース翻訳を行っている者について、各言語の対応状況を図２．２．３－２に示す。英日、日英、韓日、日韓が一番多く 100%となっている。次に中日、日中が 75%、日葡が 50%、独日、日独、仏日、日仏、伊日、日伊、露日、日露、西日、日西、葡日、泰日、日泰が 25%、越日、日越、尼日、日尼、その他が０%となっている。

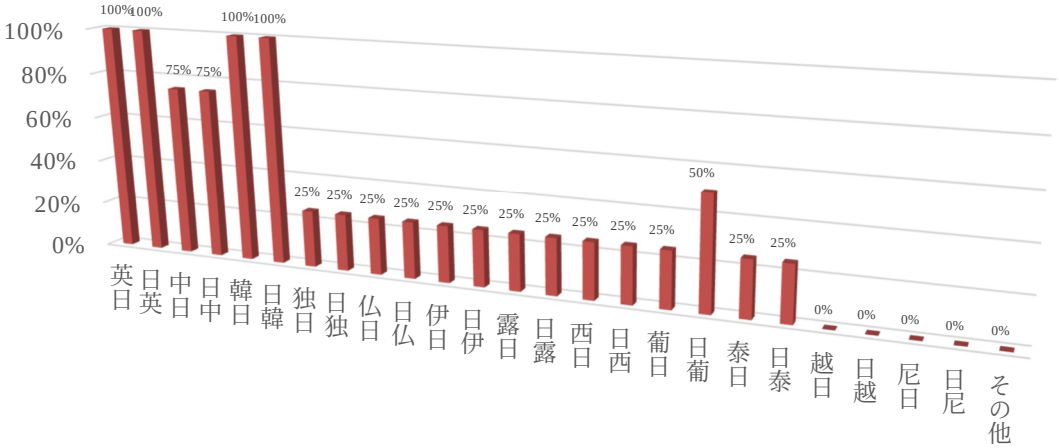

図２．２．３－２　ルールベース翻訳の対応状況

（３）用例ベース翻訳の対応状況

用例ベース翻訳を行っている者について、各言語の対応状況を図２．２．３－３に示す。英日、日英、中日、日中、韓日、日韓が一番多く 100%となっている。次に独日、日独、仏日、日仏、伊日、日伊、露日、日露、西日、日西、葡日、日葡、泰日、日泰が 50%、越日、日越、尼日、日尼、その他が０％となっている。

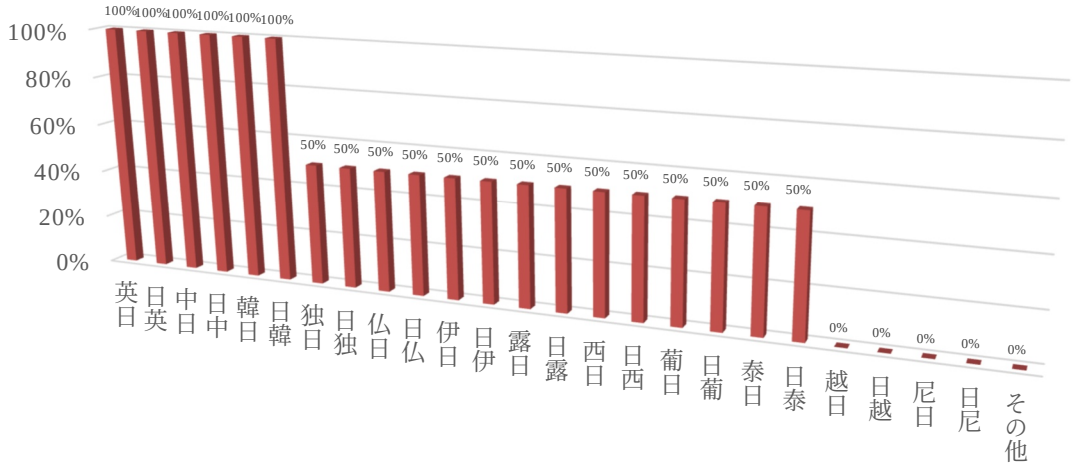

図２．２．３－３　用例ベース翻訳の対応状況

（４）ニューラル翻訳の対応状況

ニューラル翻訳を行っている者について、各言語の対応状況を図２．２．３－４に示す。英日、日英、中日、日中、独日、日独、仏日、日仏、伊日、日伊、露日、日露、西日、日西、葡日、日葡、その他が 100%となっている。韓日、日韓、泰日、日泰、越日、日越、尼日、日尼が０％となっている。

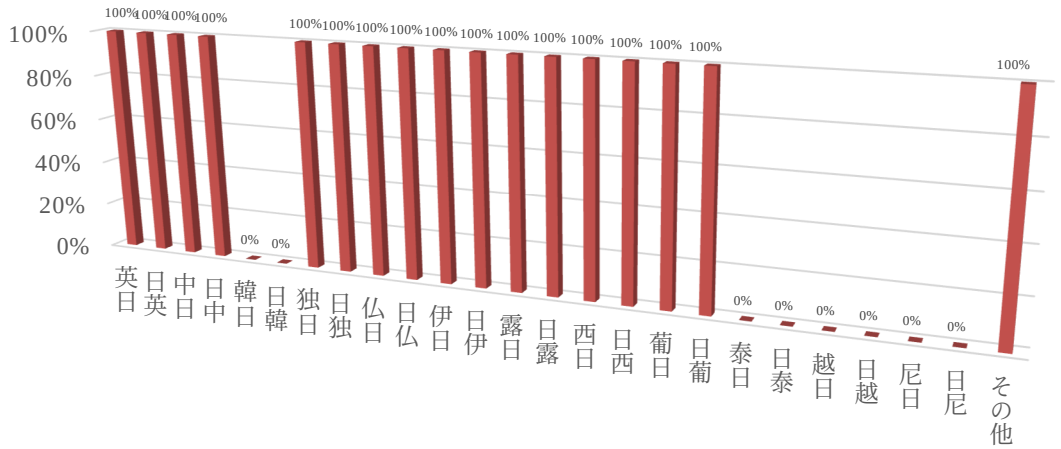

図２．２．３－４　ニューラル翻訳の対応状況

（５）統計的後編集の対応状況

統計的後編集を行っている者について、各言語の対応状況を図２．２．３－５に示す。英日、日英、韓日、日韓が 100%となっている。中日、日中、独日、日独、仏日、日仏、伊日、日伊、露日、日露、西日、日西、葡日、日葡、泰日、日泰、越日、日越、尼日、日尼、その他が０%となっている。

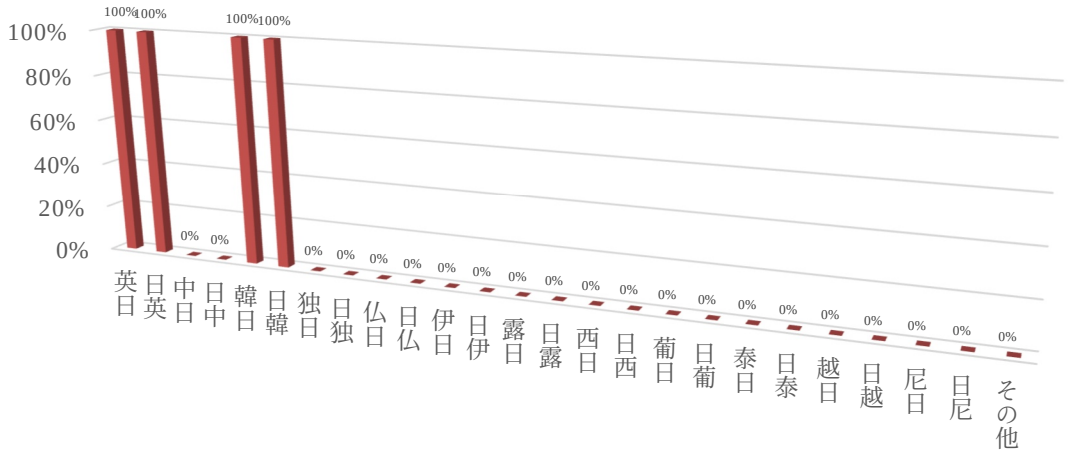

図２．２．３－５　統計的後編集の対応状況

2.2.4　翻訳前処理

回答結果を図２．２．４に示す。語順並べ替え、長文分割が一番多く 67%となっている。改行整形、引用括弧の別処理、その他が 17%となっている。また、回答無しが 17%となっている。

その他の具体的な内容は、特許請求項への対応等である。

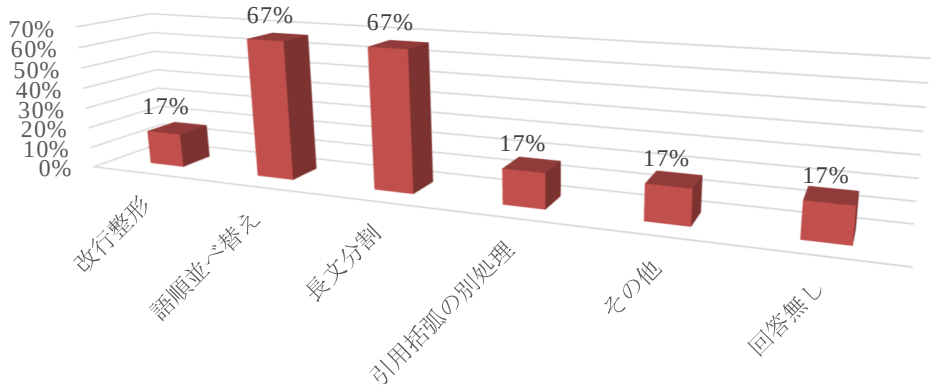

図２．２．４　翻訳前処理

※複数回答あり

2.2.5　使用している辞書の種類

回答結果を図２．２．５に示す。技術文書が一番多く 83%となっている。次に特許文書が 67%、その他が 33%、新聞が 17%となっている。また、回答無しが 17%となっている。

その他の辞書として、一般文書や会話等から作成したものがある。

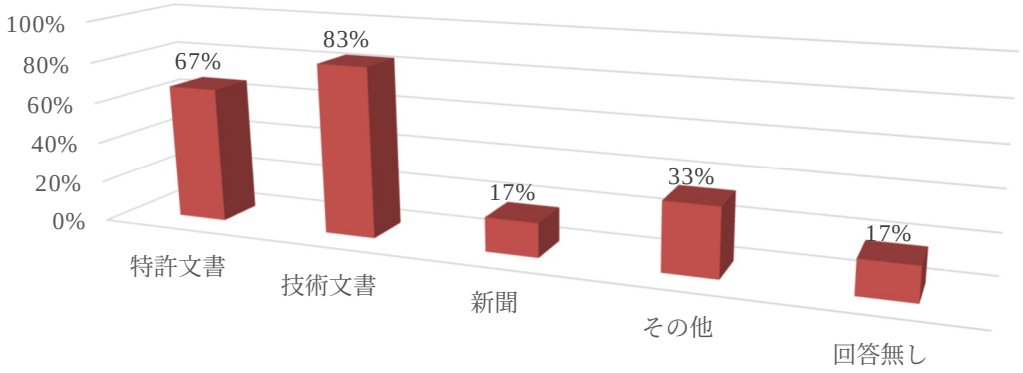

図２．２．５　使用している辞書の種類

※複数回答あり

2.2.6　使用している辞書の語数

回答結果を図２．２．６に示す。辞書の語数の観点から見ると、語数 100 万語未満の辞書は利用されておらず、1,000 万語以上のものと 100 万語以上 1,000 万語未満のものが同程度（約 33%）利用されている。図中のその他は、語数の詳細を公開していない場合や、言語により辞書の語数が異なる場合等を含む。

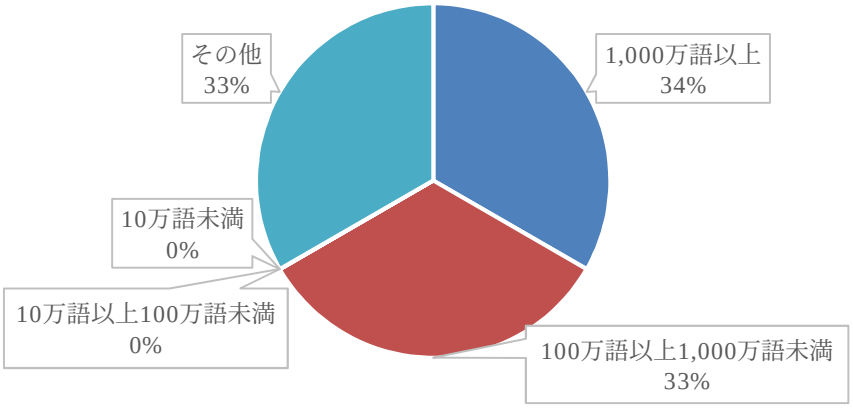

図２．２．６　使用している辞書の語数

2.2.7　辞書に含まれる情報

回答結果を図２．２．７に示す。品詞情報が一番多く 50%となっている。次に頻度情報が 33%、変数情報が 17%、その他が０%となっている。また、回答無しが 50%となっている。

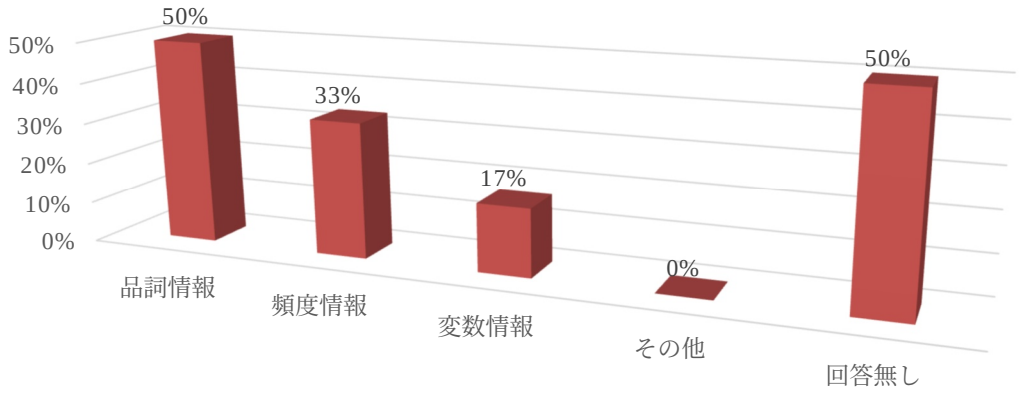

図２．２．７　辞書に含まれる情報

※複数回答あり

2.2.8　対訳コーパスの種類

回答結果を図２．２．８に示す。技術文書をもとに作成した対訳コーパスが一番多く、83%となっている。次に特許文書が 67%、新聞が 50%、その他が 33%となっている。また、回答無しが 17%となっている。

なお、その他は、専門分野、Ｗｅｂ、会話等をもとに作成した対訳コーパスである。

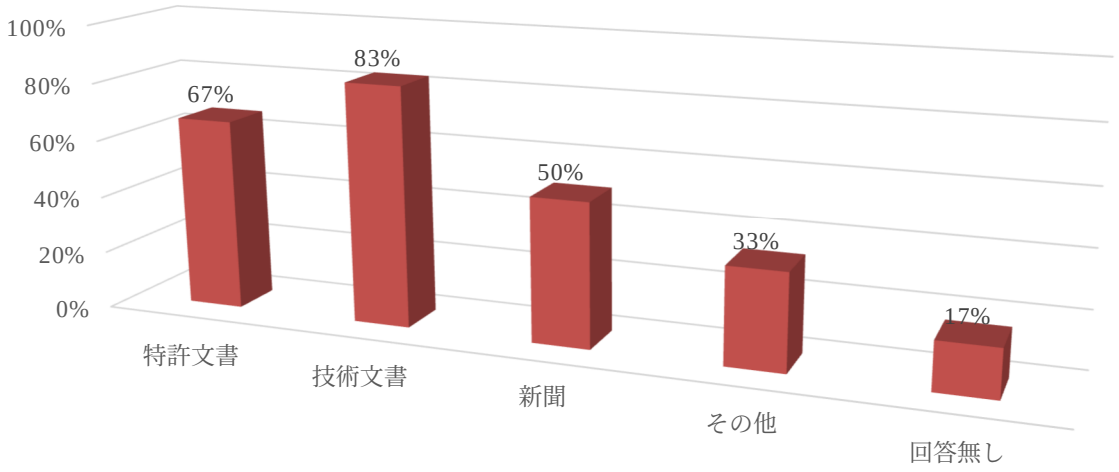

図２．２．８　対訳コーパスの種類

※複数回答あり

2.2.9　対訳コーパス（データ数）

回答結果を図２．２．９に示す。１億文対以上、1,000 万文対以上１億文対未満、100 万文対以上 1,000 万文対未満、100 万文対未満、その他、回答無しが約 16%となっている。

その他の具体的内容は、データ数未公開等である。

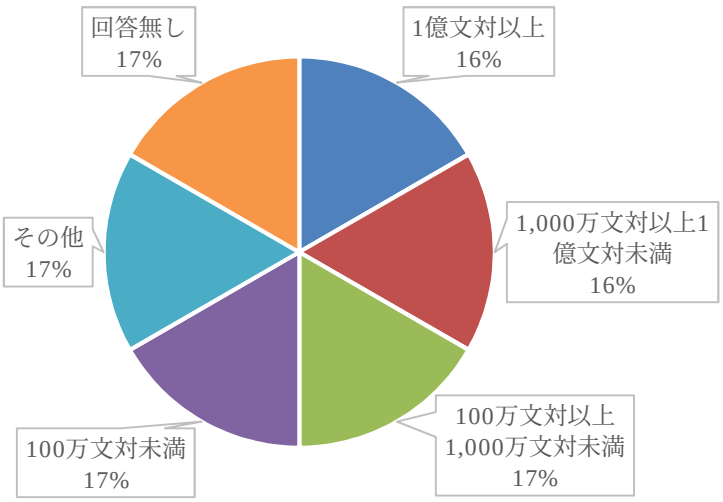

図２．２．９　対訳コーパス（データ数）

2.2.10　翻訳文字数（１秒当たり）

回答結果を図２．２．１０に示す。1,000 文字（統計翻訳）、300～1,500 文字（ルールベース・用例ベース翻訳）が 17%となっている。また、回答無しが 66%となっている。

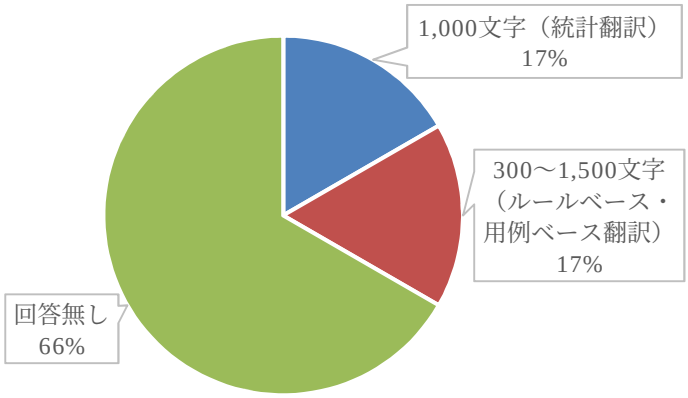

図２．２．１０　翻訳文字数（１秒当たり）

2.3　システム提供の状況について

2.3.1　提供形態

回答結果を図２．３．１に示す。Ｗｅｂサービスが一番多く 100%となっている。次にソフトウェアライセンスが 83%、個別システムへの組み込み、市販ソフトウェアが 50%、その他が 17%となっている。

個別システムへの組み込みとは、他者に機械翻訳システム（翻訳エンジン）を提供して、他者がサービス（Ｗｅｂサービス等）を行うことである。

その他の具体的な内容は、スマートフォンへの組み込み等である。

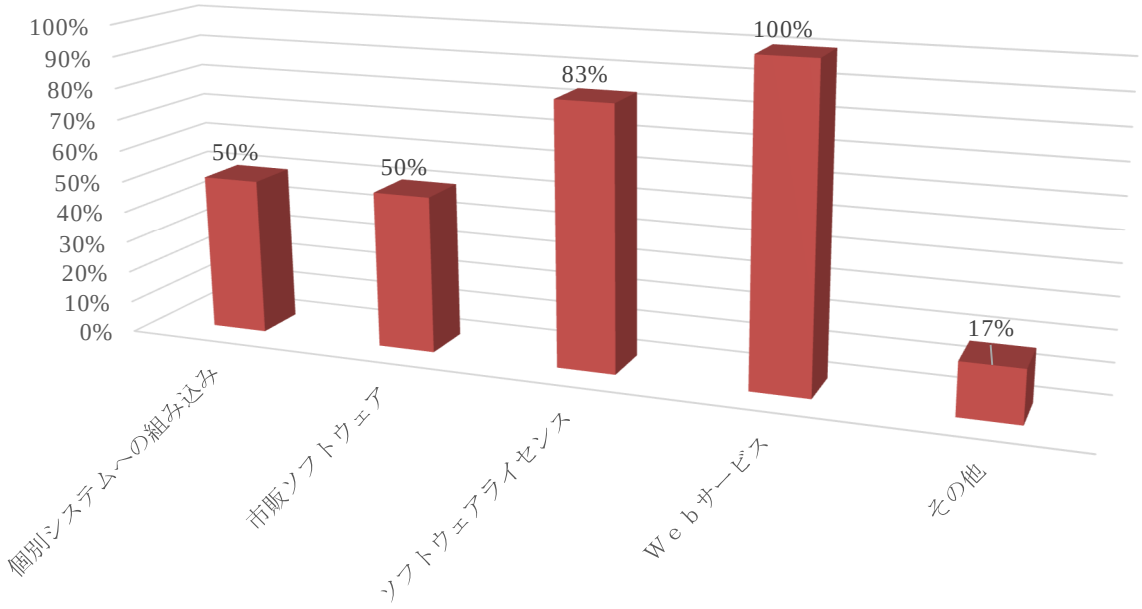

図２．３．１　提供形態

※複数回答あり

2.3.2　提供先業種

回答結果を図２．３．２に示す。情報・通信システムが一番多く 83%となっている。次に製造、金融・銀行・証券・保険、医療、官公庁・自治体が 67%、電気・石油・ガス・鉱業、サービス（ＩＴ除く）、教育、交通が 50%、建設・土木、流通・小売が 33%、その他が０％となっている。また、回答無しが 17%となっている。

近年の傾向として、観光関連サービス、官公庁・自治体への提供が増加しているとの回答があった。

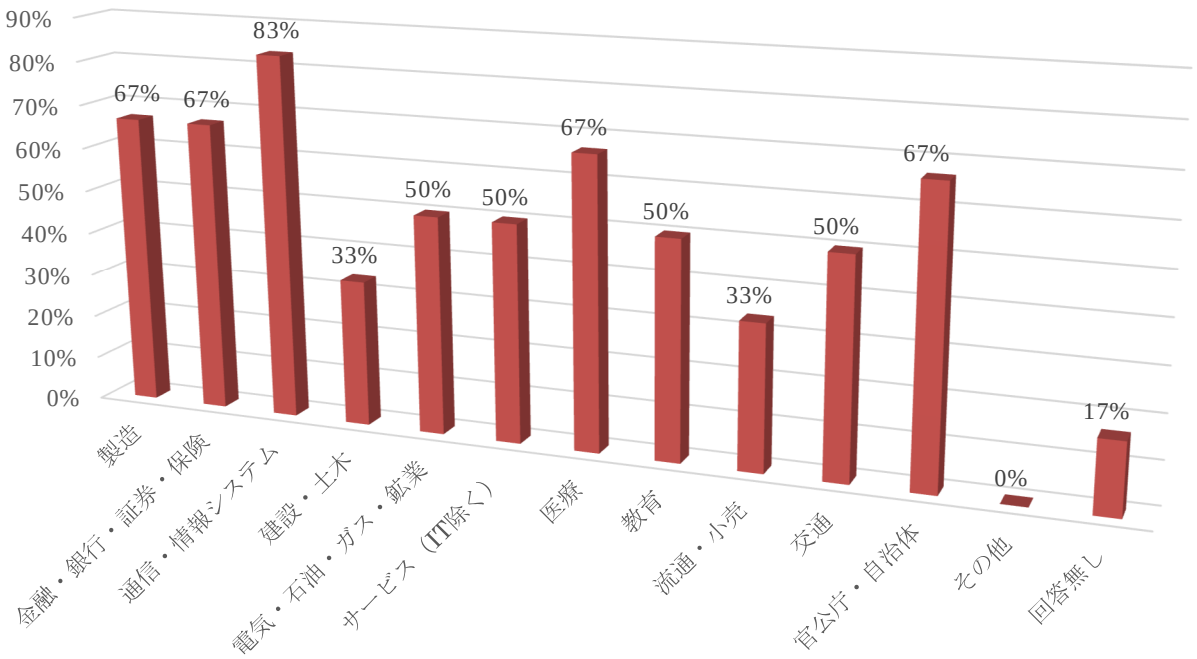

図２．３．２　提供先業種

※複数回答あり

2.3.3　ユーザ数

回答結果を図２．３．３に示す。10 万以上が一番多く 33%となっている。次に 10 以上 100 未満、その他が 17%、10 未満、100 以上 10 万未満が０％となっている。また、回答無しが 33%となっている。

その他の具体的内容は、ユーザ数非公開等である。

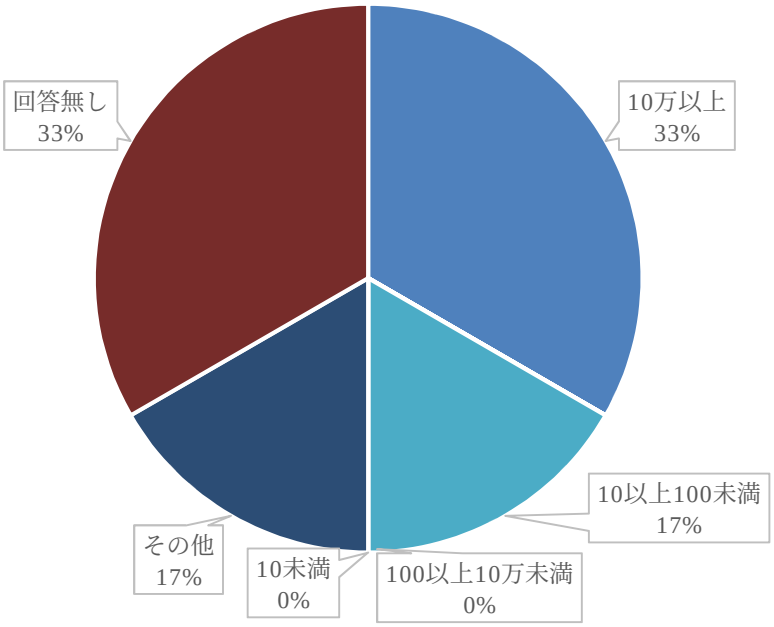

図２．３．３　ユーザ数

2.4　翻訳精度向上の取組について

2.4.1　精度向上の重点項目

回答結果を図２．４．１に示す。未知語を減らす、自然な訳文・読みやすさ、単語の誤訳を減らす、構文・係り受けの誤りを減らすが一番多く 83%となっている。次にその他が 17%となっている。

その他の具体的内容は、有識者の評価の実施等である。

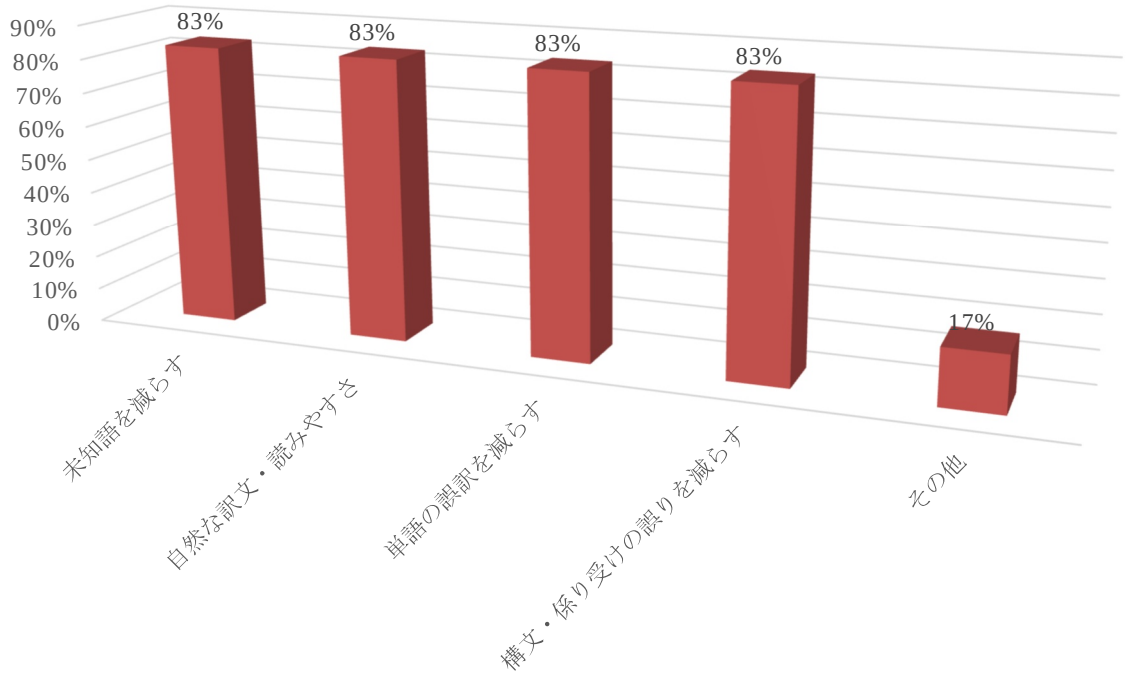

図２．４．１　精度向上の重点項目

※複数回答あり

2.4.2　精度向上に関する取組

回答結果を図２．４．２に示す。コーパス量増加が一番多く 83%となっている。次に分野別コーパスの使い分けが 67%、辞書語数増加、分野別辞書の使い分けが 50%、その他が 33%となっている。

その他の具体的内容は、前処理の高精度化、ニューラル翻訳の採用等である。

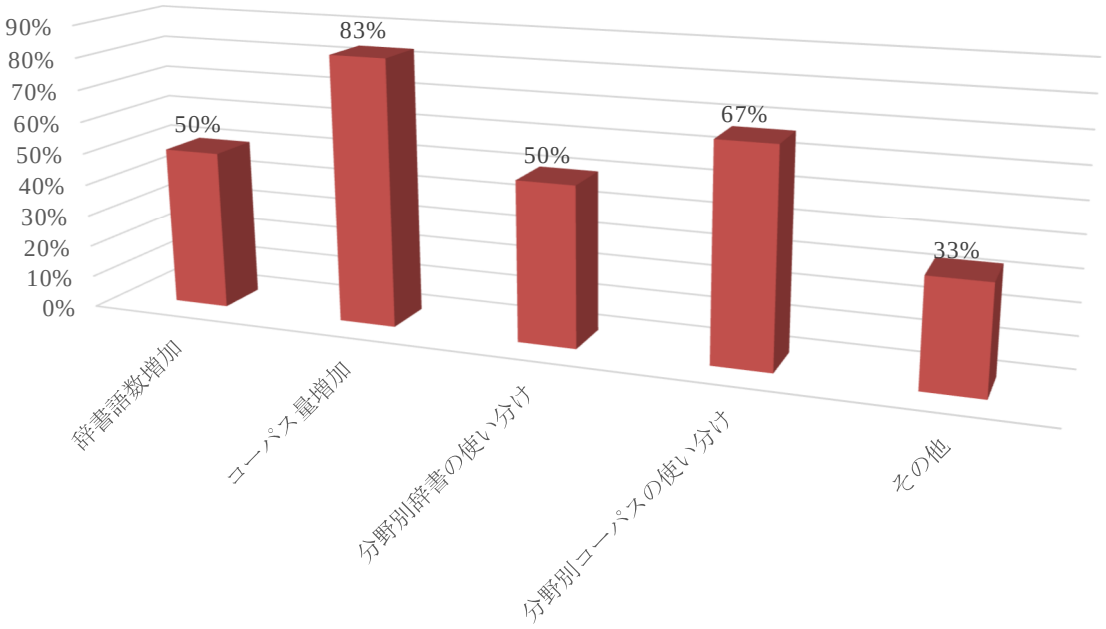

図２．４．２　精度向上に関する取組

※複数回答あり

2.4.3　辞書・コーパス更新頻度

回答結果を図２．４．３に示す。年１回以上が一番多く 50%となっている。次に月１回以上が 17%、毎日が 16%となっている。また、回答無しが 17%となっている。

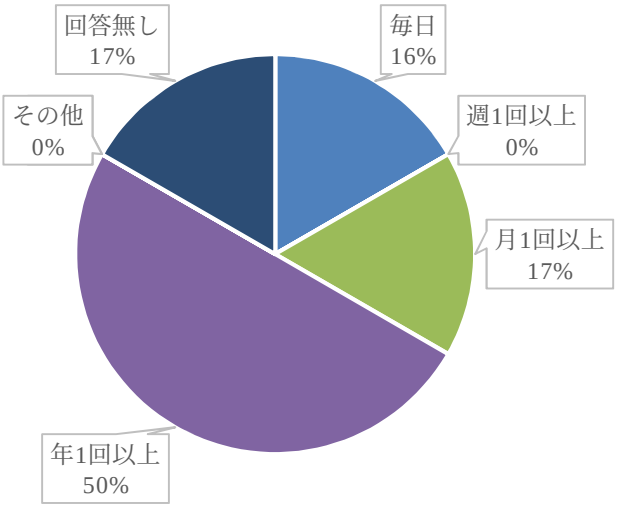

図２．４．３　辞書・コーパス更新頻度

2.4.4　誤訳事例の抽出・収集方法

回答結果を図２．４．４に示す。専門チーム、ユーザの指摘が一番多く 83%となっている。次にその他が 17%となっている。

その他の具体的内容は、指標の自動取得等である。

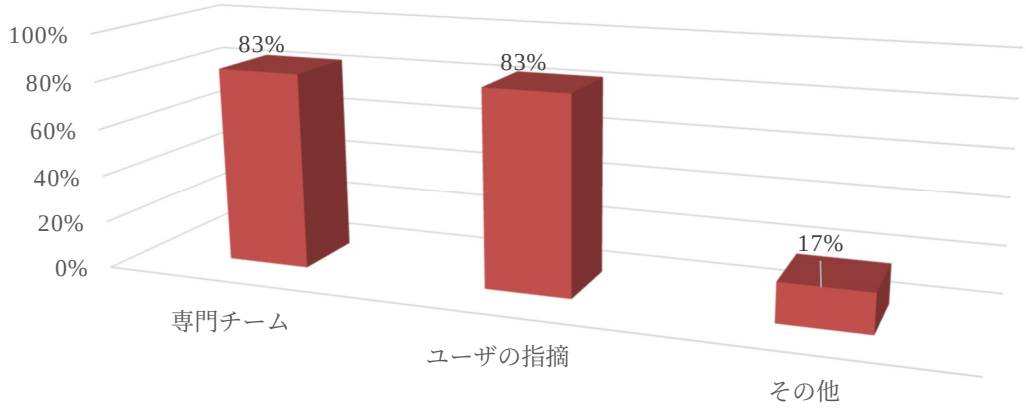

図２．４．４　誤訳事例の抽出・収集方法

※複数回答あり

2.4.5　精度評価方法

回答結果は人手評価、機械評価が 100%となっている。その他が０%となっている。

※複数回答あり

2.4.6　評価文数規模

回答結果を図２．４．６に示す。10 万文以上 100 万文未満、１万文以上 10 万文未満、1,000 文以上１万文未満、その他が約 16%となっている。次に 100 文以上 1,000 文未満、100 文未満が０%となっている。また、回答無しが 33%となっている。

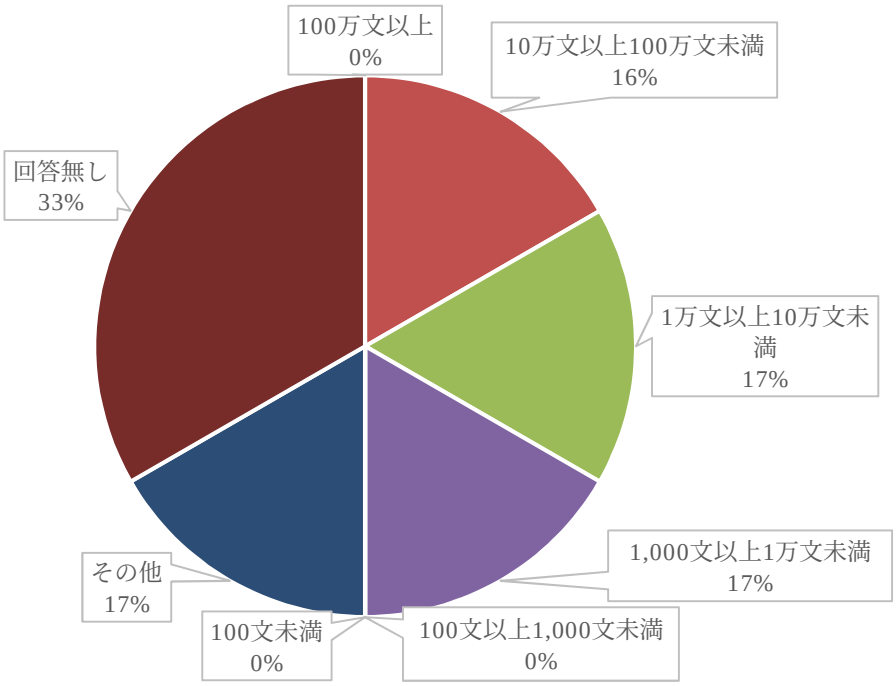

図２．４．６　評価文数規模

2.5　今後の展開について

2.5.1　翻訳言語の拡張

回答結果を図２．５．１に示す。有りが一番多く 83%となっている。次に未定が 17%、無しが０%となっている。

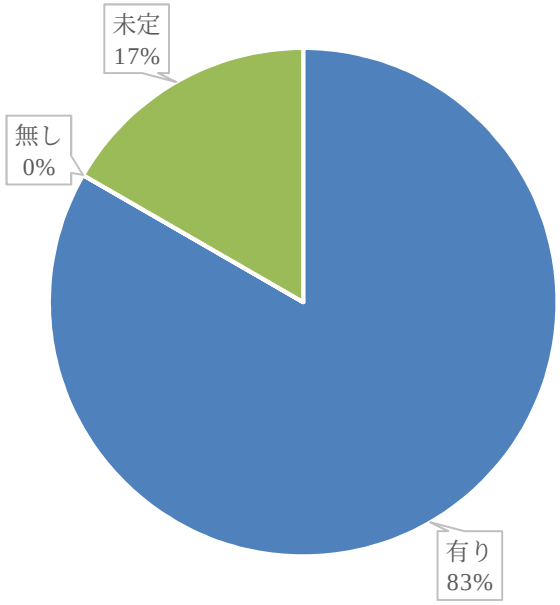

図２．５．１　翻訳言語の拡張

2.5.2　翻訳精度向上のための改善

回答結果は有りが 100%となっており、未定、無しが０%となっている。

有りと回答があった中で具体的な内容については 66%が非公開・回答無しであるが、回答を得られた中にはニューラル翻訳の対応を行う予定等の回答があった。

2.5.3　提供形態の変更

回答結果を図２．５．３に示す。未定が一番多く 50%となっている。次に無しが 33%、有りが０%となっている。また、回答無しが 17%となっている。

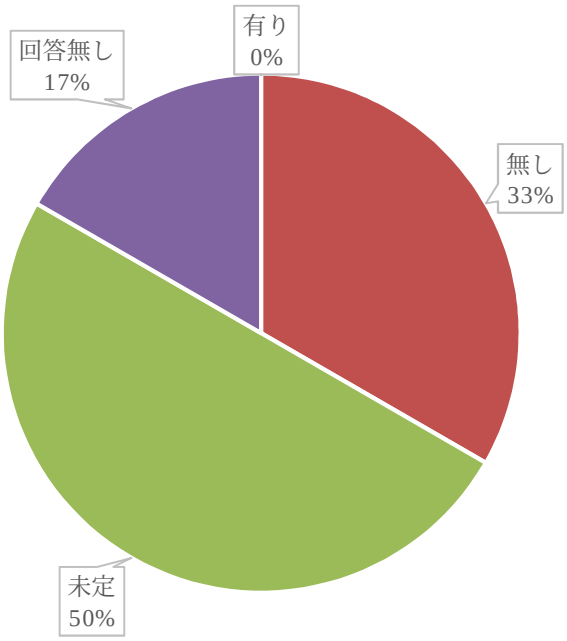

図２．５．３　提供形態の変更

3. 翻訳精度等の技術的な観点からの発展の見通し

3.1 対訳辞書を利用する仕組み

語学学習や翻訳作業を支援するため、主要な言語間では対訳辞書が整備されることが多い。また、出現頻度の高い基本的な用語を網羅した辞書とは別に、特定の専門分野で使用される専門用語を収集した専門用語辞書が作られることも少なくない。ルールベース翻訳は、このような対訳辞書の利用することを基本としている。

一方、統計翻訳や後述のニューラルネットワークを使った機械翻訳（NMT）では、対訳コーパスをもとに対訳辞書（フレーズテーブル）を自動的に学習するという設計になっている。未知語を解決するためにこの仕組みにしたがって、外付けの専門用語辞書を対訳コーパスの一部として対訳コーパスに追加して学習させることが考えられるが、このように追加学習させても翻訳品質はあまり向上しないことが経験的にわかっている。この理由は、辞書中の語は文脈がなく単独で出現するため、フレーズテーブルに採用されたとしても N-gram が短めになってしまうこと、そして、用語の出現回数が多くないので、文の翻訳において辞書中の訳語が訳文に採用される可能性が高くないためである。

このように、従来は統計翻訳に対訳辞書は適用することは難しかったが、近年では以下の手法が提案され、統計翻訳の中で対訳辞書を使うようになってきている。

① 意味に基づいて単語を分類し、対訳辞書に分類を付与する。
② 同分類に従って、対訳コーパスを加工した上で対訳コーパスから翻訳に必要なモデルを学習する。
③ 単語の分類を利用して翻訳プログラムが動作し、対訳辞書を参照して訳語を決定する。

先行研究[1]では、単語「品川」や「五反田」などを分類「地名」に、単語「鈴木」や「田中」などを分類「人名」に割り当て、この分類に従った対訳コーパスからフレーズベース統計翻訳システムを構築している。これにより、翻訳精度が向上したことが報告されているが、この手法には、以下の課題が残されている。

・分類の決め方には人手付与と自動付与[2]があり、それぞれに欠点がある。
・人手付与は作業のコストが嵩む。
・自動付与は低頻度語をうまく分類できず、また、分類の意味を説明できない。
・分類を変更するとシステムを再構築しなくてはならない。
・NMT で統計翻訳と同じ手法がうまく働くか、わかっていない。

---

1 Hideo Okuma,Hirofumi Yamamoto,Eiichiro Sumita. 2008. Introducing a Translation Dictionary into Phrase-Based SMT. IEICE TRANSACTIONS on Information and Systems, Vol.E91-D, No.7, pp.2051-2057, July 2008.

2 http://textprocessing.org/open-source-text-processing-project-mkcls

4. 将来有力な翻訳方式や技術

4.1 ニューラルネットワークを使った機械翻訳

世紀の変わり目に、ルールベース翻訳から統計翻訳（SMT）への研究のパラダイムシフトが起こった。ここ数年のニューラルネットワーク（NN）を利用したいわゆる深層学習（deep learning）と呼ばれる機械学習技術の進展により、他分野と同様に自然言語処理においても一気呵成に研究のパラダイムがNNを使った手法へシフトしている。例えば、2016年の7月9日～15日にNew Yorkで開催された第25回人工知能国際会議[1]（International Joint Conference on Artificial Intelligence(IJCAI)）におけるNNの動向をみると、従来のIJCAIと異なり、NNに関連する論文であふれていた。特に、自然言語処理のセッションは7つと多く、約75%はNNを含んでいた。人工知能に関する他の会議（例えば、AAAI）や自然言語処理の第一級の会議（例えば、NAACL、ACL）も同様の傾向にある。また、多数の国際会議等に付随する形でNNに関する多数のチュートリアル[2]が行われている。

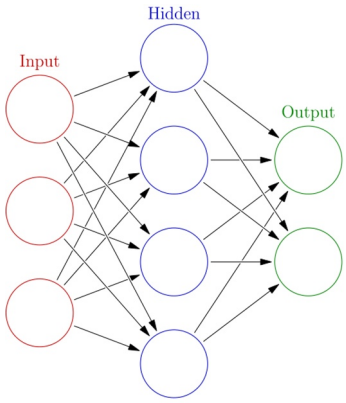

図４．１　ニューロンの相互結合の模式図[3]

図４．１は神経細胞（ニューロン）が相互結合した神経回路網からなる脳を模倣して、人工物のノードの相互結合として構成されることを示す。ここで、「○」は人工的なニューロンを表し、「→」は一つのニューロンの出力から別のニューロンの入力への結合を表す。

---

1 人工知能に関する老舗の国際会議である。

2 重要なスライドを下にまとめる。
- https://drive.google.com/file/d/0B525s-oGR9wpN1hWOGMyaE5pcXc/view
- http://phontron.com/slides/neubig15kyoto-slides.pdf
- http://cl.naist.jp/~kevinduh/notes/cwmt14tutorial.pdf
- http://www.hangli-hl.com/uploads/3/4/4/6/34465961/naacl_tutorial_version2.2.pdf
- http://eecs.wsu.edu/~jana/SP-tutorial-IJCAI2016.html
- http://research.microsoft.com/en-us/um/people/scottyih/IJCAI-2016_tutorial.pdf
- http://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf
- http://nlp.stanford.edu/projects/nmt/Luong-Cho-Manning-NMT-ACL2016-v4.pdf

3 https://commons.wikimedia.org/wiki/File:Colored_neural_network.svg

4.2　NN の概要

脳において、単純な機能しか持たない神経細胞（ニューロン）の組み合わせからなる神経回路網が高度な情報処理を実現していることをヒントに、ニューロンをハードウェアやソフトウェアのような単純な計算のみ行うユニットに置き換えれば、その組み合わせからなるネットワークで高度な情報処理を実現できるであろうと期待され、ニューラルネットワーク(Neural Network,NN)[4]は提案された（図４．１）。

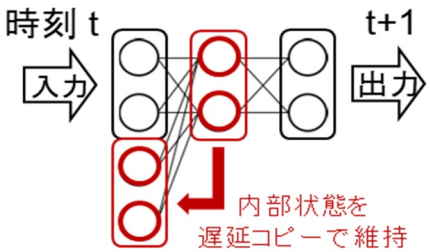

図４．２　Recurrent Neural Network(RNN)

NN で入力から出力を得るには、入力があった各ユニットで同時並行して重みと入力の積和という単純な計算を行う。つまり、同時並行で計算できる能力を高めることで処理速度が速くなることが NN に General purpose computing on GPU(GPGPU)[5]が多用される理由である。GPGPU とは、元来はゲームやバーチャルリアリティの描画のような画像処理（Graphics Processing）で必要だった、多数のユニットで同時並行的に実行される（条件分岐がない）単純な演算を高速化するために開発された演算機能が、画像処理以外の用途、特に、NN で転用されたものである。NN は大量に（ユニット数に比例した個数だけ）GPGPU を使うことで実行時間を短縮できるため、大規模な NN は大量の GPGPU にアクセスできない組織では追試・改良はできないという規模依存性が目立つ技術になっている。

代表的な NN に、Feedforward Neural Network (FFNN)と Recurrent Neural Network (RNN)がある。

- FFNN は情報を常に一方向に流す。図４．１は FFNN の例になっている。
- RNN は図４．２にあるように、ある時刻 t における内部状態を、次の時刻 t+1 の入力に使う。そのため、時刻 t+1 では、その時刻における入力+前回の履歴を利用する。

FFNN がある一点での識別問題を解く構成になっているのに対して、RNN は識別問題の系列を、それ以前の回答内容も考慮して解く構成となっている。従って、長距離の依存関係を持ち得る記号の系列を扱うことの多い自然言語処理において、RNN は非常に有効であると考えられる。

[4] 又は、人工ニューラルネットワーク(artificial neural network, ANN)とも呼ばれる。
[5] 最近は発音しにくいので省略して GPU と呼ばれることも多い。

4.3 NN を利用した SMT の強化

NN の機械翻訳への応用は、SMT を構成する部品である各種の統計モデルを NN によって代替させるところから始まった。

4.4 NN による言語モデル

SMT では、翻訳モデルを用いて生成される多様な訳文の候補の中から、目標言語における尤もらしい候補を選択するのに、言語モデルが用いられている。言語モデルには従来は N-gram モデルが主流であったが、NN、特に RNN による言語モデル[6]が音声認識や機械翻訳の性能向上に寄与している。N-gram 言語モデルにおいては、直前 N-1 単語から次の単語の確率が決定されるが、RNN 言語モデルでは、文頭からすべての単語の履歴を文脈ベクトルの形で保持して次の単語の確率の決定に利用することができ、長距離の依存関係があるような単語の予測精度が向上することがその理由である。

4.5 NN による単語アラインメントモデル

SMT における単語アラインメントモデルは IBM モデルや HMM が主流であったが、RNN を用いることで単語アラインメント精度、及びその結果得られる翻訳モデルの性能を向上させる手法が提案されている[7]。HMM では直前の単語のアラインメントのみを参照していたものが、RNN によってそれ以前の単語のアラインメント履歴を考慮してアラインメントを決定できることが理由であると考えられる。

4.6 NN によるフレーズ並べ替えモデル

フレーズに基づく SMT においては、フレーズの順序がどう入れ替わるかを表現する統計モデルが利用される。従来の頻度に基づく統計的なフレーズ並べ替えモデルを、FFNN による識別モデルとして高精度化を図ったモデルが提案されている[8]。

4.7 NN による翻訳モデル

NN を大規模な SMT で利用し大きく成功した方法の一つが Neural Network Joint Model(NNJM)[9]である。NNJM は、原言語の入力単語列と目的言語の出力単語の履歴を特徴量とする FFNN を用いて確率付きで訳語を出力するようになっており、フレーズに基づく SMT におけるフレーズテーブルの代替となるのみならず、言語モデルの機能をも含んだ同時（joint）モデルとなっている。FFNN を利用することで、既存のフレーズベース翻訳の解

---

6 Thomas Mikolov et al., Recurrent neural network based language model, In Proc. INTERSPEECH 2010.

7 田村,渡辺,隅田,リカレントニューラルネットワークによる単語アラインメント,自然言語処理 第 22 巻 4 号, 2015.

8 Peng Li et al., A Neural Reordering Model for Phrase-based Translation, In Proc. Coling 2014.

9 Jacob Devlin et al, Fast and Robust Neural Network Joint Models for Statistical Machine Translation, In Proc. ACL 2014.

探索を大きく変更することなく導入可能であることも大きな特徴となっている。それ以前に報告された、RNN 言語モデルに原言語の入力単語列の情報も加える形で拡張して翻訳モデルとして利用した研究[10]では、フレーズベース翻訳において複数の解候補をいったん生成したのち、RNN モデルによるリスコアリングを行っていた。

4.8 NN による PRE-ORDERING

日英間のように語順の違いの大きい言語対での翻訳において、SMT による語順の誤りを軽減するための PRE-ORDERING（事前並べ替え）がよく用いられている。PRE-ORDERING はルールに基づく構文木の変換やサポートベクタマシン（SVM）を利用した識別学習によって実現されてきたが、この箇所の識別学習に NN を利用した手法が提案されている[11]。

4.9 ニューラル機械翻訳 (Neural Machine Translation;NMT)

前節で紹介した各手法が、SMT の様々な統計モデルの性能を向上させるために NN を適用したものであったのに対し、NN を利用して翻訳の過程をまとめて学習・実行する、ニューラル機械翻訳（Neural Machine Translation;NMT）と呼ばれる方式が急速に発展している。

4.10 sequence-to-sequence の翻訳

原言語の単語列を入れて目的言語の単語列へのマッピング全体を学習する「NN による end-to-end の翻訳」が、Google から提案された[12]。この方法は、RNN の一種である Long Short-Term Memory(LSTM)[13]を用いて、図４．１０に示すように、原言語の単語列「A」「B」「C」「<EOS>」に続く形で（原言語の単語列と長さが異なる）目的言語の単語列「W」「X」「Y」「Z」「<EOS>」を対にして RNN（LSTM）に与えて学習する。この入力をベクトルの形で記憶（符号化）し、ベクトルから順次単語を出力（復号化）する仕組みから、encoder-decoder モデルと呼ばれることも多い。

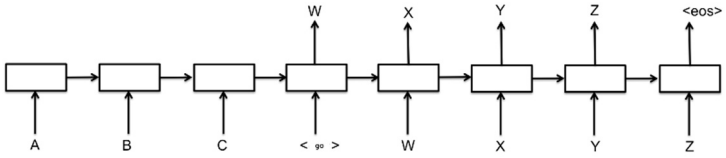

図４．１０ sequence-to-sequence の翻訳

これは、NN を単語列の対だけで学習するもので、SMT で行われていたような単語アライ

[10] Michael Auli et al., Joint Language and Translation Modeling with Recurrent Neural Networks, In Proc. EMNLP 2013.
[11] Antonio Valerio Miceli-Barone and Giuseppe Attardi, Non-projective Dependency-based Pre-reordering with Recurrent Neural Network for Machine Translation, In Proc. ACL-IJCNLP 2015.
[12] Ilya Sutskever et al., Sequence to Sequence Learning with Neural Networks, In Proc. NIPS 2014.
[13] LSTM はフィードバックをもった RNN であって、履歴の更新(記憶と忘却)を細かく制御することを可能にするためスイッチを持たせたものである。

ンメントはもちろんのこと、構文解析や言語依存のヒューリスティクスを一切用いていないにも関わらず、従来の最高精度の SMT に迫る翻訳精度を達成したことで大きな注目を集めた。

しかし、このモデルでは入力単語列を読み込み終わった時点で、そのすべての内容を一つの固定次元数の実数値ベクトルで保存しているため、特に文が長い場合に情報の符号化や復号化が正しく行えず、翻訳の抜けや不必要な翻訳の繰り返しが生じやすいという問題があった。

4.11 注意機構を導入した sequence-to-sequence の翻訳

最初の sequence-to-sequence 翻訳のモデルにおいて、入力文を一つの実数値ベクトルに保存していたが故に特に長い文の翻訳の精度が著しく低下する問題に対処すべく、注意（attention）の機構を持つ NN を利用した sequence-to-sequence 翻訳手法が提案された[14]。

注意機構とは、ある単語を翻訳結果の一部として出力するときに入力文のどの単語に「注意を払って」翻訳するかを、入力文の各単語を符号化した情報を重み付きで利用することで重みの大きい単語の情報に注意を払って（重んじて）訳語を選択する仕組みのことである。この手法ではまず入力文の符号化を行う際に、各入力単語を読み込んだ後の符号化結果をすべて履歴として保存しておき、復号化の開始時から訳語を出力する毎に注意情報（各入力文符号化履歴の読み出し重み）を更新する。そのため、文が長くなったとしても、注意情報が正しく更新される限りは注意箇所に対応する訳語が得られ、翻訳の抜けや不必要な繰り返しが起こりづらい仕組みになっている。

注意機構の導入によって、長い文に対する翻訳精度は大きく改善されることが示されている[15]が、注意情報の更新誤りによって依然として翻訳の抜けや繰り返しが起こることがあり、これを解消するために注意情報の履歴を管理し、すべての入力単語に満遍なく注意が向けられるようにする方法も提案されている[16]。

4.12 ニューラル機械翻訳の実装

2015 年 5 月に Baidu が NMT のプロダクト化（中⇔英）[17]を行ってから、SYSTRAN（2016 年 8 月、12 言語）[18]、WIPO（同 10 月、日中韓→英）[19]、Microsoft（同 11 月、10 言語）[20]、Google（同 11 月、 8 言語）[21]などが次々と NMT をプロダクト化していった。また、数

14 Dzmitry Bahdanau et al., Neural Machine Translation by Jointly Learning to Align and Translate, In Proc. ICLR 2015.
15 Minh-Thang Luong et al., Effective Approaches to Attention-based Neural Machine Translation, In Proc. EMNLP 2015.
16 Zhaopeng Tu et al., Modeling Coverage in Neural Machine Translation, In Proc. ACL 2016.
17 http://www.aclweb.org/anthology/W15-4110
18 https://arxiv.org/abs/1610.05540
19 http://www.wipo.int/pressroom/en/articles/2016/article_0014.html
20 https://blogs.msdn.microsoft.com/translation/
21 https://japan.googleblog.com/2016/11/google.html

多くの NMT の実装がオープンソースとして GitHub 上で公開されている。表４．１２にその一部をリストアップする。

表４．１２　オープンソースの NMT の例

	Interface	Framework	https://github.com	Note
GroundHog	Python	Theano	lisa-groundhog/GroundHog	開発終了 (Blocks に移行)
Blocks	Python	Theano	mila-udem/blocks	モントリオール大
DL4MT	Python	Blocks,Theano	nyu-dl/dl4mt-tutorial	NYU の Cho ら
NEMATUS	Python	Theano	rsennrich/nematus	WMT2016 best
KyotoNMT	Python	Chainer	fabiencro/knmt	京大黒橋研
OpenNMT	Lua	Torch	OpenNMT/OpenNMT	元は harvard-nlpSYSTRAN が利用
lamtram	C++	DyNet	neubig/lamtram	CMU の Graham
NMTKit	C++	DyNet	odashi/nmtkit	NAIST 中村研
N3LP	C++	なし	hassyGo/N3LP	東大鶴岡研

注目は OpenNMT で、これは SYSTRAN が公式に利用している実装である。つまり SYSTRAN は、翻訳エンジン自体を無料で公開していることになる。また、Google も実装こそ公開はしていないものの、内部で動いているエンジンのアーキテクチャーの詳細を論文で公開している[22]。公開されている情報を見る限り、ほとんどのシステムは注意機構を導入した NMT をベースとしており、これにそれぞれ独自の工夫を入れたものになっている。これはちょうどフレーズベース SMT を基礎として翻訳精度の向上を図っていたのと似た状況であると言える。

SMT の翻訳エンジンを一から実装するとなると非常に骨の折れる作業であったが、NMT ではモデルがシンプルな上に、深層学習用のフレームワークを利用することで誰でも簡単に翻訳エンジンを作ることができる。例えば、博士課程に入りたての学生が、数ヶ月で新たな NMT モデルの提案、実装、論文執筆を行い、国際的なトップカンファレンスに採択されたという例もある。つまり、ソフトウェアとしての翻訳エンジン自体の価値は以前よりも低くなっており、より一層データの重要性が増していると言える。

4.13　現状のニューラル機械翻訳の課題

現状の NMT の精度は、すでに多くの状況において SMT の精度を大きく上回っているが、解決すべき課題はまだまだ残されている。大雑把にいうと、SMT は単語レベルでの必要情

[22] https://arxiv.org/abs/1609.08144

報の訳出精度（adequacy）は高いが、それらを組み合わせて滑らかな訳を作り出すこと（fluency）は不得意であった。NMT はこれとは真逆で、非常に滑らかだがよくみると必要な情報が欠けていることが多い。また、同じ内容を何度も繰り返すような翻訳を生成することもある。本章では、NMT の課題について整理をする。

（１）扱える語彙数が少ない

SMT では、対訳コーパスの量を増やしたり、対訳コーパスとは別に対訳辞書を整備したりすることで、フレーズテーブルのサイズはどんどん大きくはなるが、未知語の数を減らすことができた。一方で、NMT ではモデルの構造と計算時間の問題から、モデル内で扱える語彙数が非常に小さい。多くの論文では、語彙数を３万や５万、多くても 10 万程度に制限して実験を行っている。この語彙内に含まれない単語は、トレーニング時もテスト時も全て[UNK]のような特別な記号に置き換えられる。テスト時の翻訳文に出力された[UNK]は、注意機構の情報を使うことでどの入力単語を参照したものかが判断できるため、元の入力文に遡って単語をコピーしたり、辞書を使って翻訳したりすることで翻訳を行う[23][24]。

これとは別の方法で語彙数の問題を解決する方法も存在する。全てを文字単位で翻訳する方法[25]や、UNK の単語のみ文字単位のモデルにバックオフするハイブリッドモデル[26]などである。さらに、単語と文字の中間に位置するような単位（sub-word と呼ばれる）を使う方法も存在する。代表的なものは Byte Pair Encoding (BPE)[27]と呼ばれるもので、WMT2016 において最も高精度な翻訳を達成している。BPE では語彙は文字単位からスタートし、高頻度で現れるシンボル列をひとまとめにして新たな語彙として登録することを、語彙サイズの上限まで繰り返す。これにより単語単位での翻訳で UNK になるものも、より小さな単位に分割して翻訳することで、理論的には全ての単語を扱えるようになる。BPE は非常に強力だが、例えば日本語の未知のアルファベット列を英語に翻訳する際に、BPE により細く分割された後に翻訳されるため、必ずしも元のアルファベット列が復元されるという保証がないという問題がある。なお、Google の NMT も BPE に似た word-piece という手法で翻訳を行っている。

---

23 Minh-Thang Luong, Ilya Sutskever, Quoc V. Le, Oriol Vinyals, andWojciech Zaremba. 2015. Addressing the rare word problem in neural machine translation. In Proceedings of ACL.Franz

24 Sebastien, J., Kyunghyun, C., Memisevic, R., and Bengio, Y. On using very large target vocabulary for neural machine translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (2015)

25 Chung, J.; Cho, K.; and Bengio, Y. 2016. A character-level decoder without explicit segmentation for neural machine translation. In ACL 2016.

26 Minh-Thang Luong and Christopher D. Manning. Achieving open vocabulary neural machine translation with hybrid word-character models. 2016.

27 Rico Sennrich; Barry Haddow; Alexandra Birch. 2016. Edinburgh Neural Machine Translation Systems for WMT 16. In Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers. Berlin, Germany.

（２）訳抜けと重複

SMT は入力文の各部分について、フレーズテーブルのエントリーを使って「置き換える」ことで翻訳文を生成するため、フレーズテーブルが高精度であれば訳抜けや重複は起こりにくかった。一方で NMT は言語モデルなどと同様に、入力文とそれまでの出力が与えられた上で次に来るべき単語を予想することで、先頭から順に翻訳文を生成している。このとき、入力文の各単語を過不足なく出力する、といった指標は入っていないため、訳抜けや重複が SMT よりも起こりやすくなっている。具体的には、翻訳文のうち約 18%に訳抜けがあり、約４%に重複があるという報告がある[28]。

この問題を解決するために、注意機構においてこれまでに注意された入力単語を記憶しておき、この情報を次に注意する単語を決定する際に利用する方法[29]や、出力を逆翻訳し、なるべく入力文と同じになるようにトレーニングを行う方法[27]、NMT とは別に用意された単語アライメント情報を教師データとして使う方法[30]などが提案されているが、決定的な解決方法はまだ提案されていない。

（３）計算量が多い

前述のとおり、NMT を含むニューラルネットワークの訓練には GPGPU が必要不可欠である。GPU が１つ（１枚）でもあれば、CPU しかない場合に比べて 10 倍程度高速に学習を行うことができる。これをさらに高速化するために、複数の GPU を使いそれぞれが異なる訓練データを使って学習し、学習の中間状態を逐次共有する方法[31]や、モデルのパラメータを量子化（浮動小数点数ではなく、整数等に丸め込む方法）して計算を高速化する方法[32][33]などが提案されている。

一方で、テスト時（訓練したパラメータを使って翻訳のみ行う場合）は CPU であっても GPU に比べて極端に遅くなるということはあまりない。また、NMT は SMT のように巨大なフレーズテーブルが必要なわけではなく、ネットワークの接続ごとに学習されたパラメータ（=行列）を保存しさえすれば良いため、モバイルなどの計算リソースが限られた場所であっても容易に搭載することができる。ここで、この翻訳モデルを翻訳精度を大きく落とさずに圧縮（pruning）する方法[34]や、パラメータ数の多い教師となるモデルを真似るように、パラメータ数の小さい生徒となるモデルを学習することでモデルを小さくする蒸留（distillation）という方法[35]が提案されており、これらを用いることで計算資源の少な

---

28 Neural Machine Translation with Reconstruction. Zhaopeng Tu, Yang Liu, Lifeng Shang, Xiaohua Liu, and Hang Li. AAAI 2017

29 Tu, Z.; Lu, Z.; Liu, Y.; Liu, X.; and Li, H. 2016b. Modeling Coverage for Neural Machine Translation. In ACL 2016.

30 Neural Machine Translation with Supervised Attention
Lemao Liu, Masao Utiyama, Andrew Finch and Eiichiro Sumita Coling2016

31 http://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks

32 https://arxiv.org/abs/1605.04711

33 https://arxiv.org/abs/1512.06473

34 https://arxiv.org/abs/1606.09274

35 Y Kim and Alexander M.Rush.2016. Sequence-level knowledge distillation. Pro- ceedings of the 2016 Conference on EmpiricalMeth- ods in Natural Language Processing,pages 1317-1327,November.

い状況であってもNMTを動作させることができるようになっている。

（4）モデルの解釈や操作の困難さ

SMTが単語や句、構文木のように記号を直接モデル化していたのに対し、NMTではモデルの内部表現はすべて数百〜数千次元の実数値ベクトルである。NNはその内部表現である実数値ベクトルによって抽象化を行うことで、単語の類似性を暗黙的にモデル化することができる。そのことにより、N-gramモデルに代表される記号に基づくモデルと比較して格段に高い言語生成能力を有し、非常に流暢な文の生成が可能になった。

一方で、モデルや動作過程もすべて実数値ベクトルの形で表現されるため、その動作や誤りの分析・解釈がほぼ不可能であることが、NMTひいてはNNを用いた自然言語処理全般において、誤り分析に基づくシステムの改良を非常に困難なものにしている。SMTは単語アラインメントの結果やフレーズペア等のモデル、さらに翻訳過程を分析することで誤りが特定できることも多いため、特定の誤りに対する対処・修正を行いやすかった。しかしながら、NMTはその高い抽象化能力と引き替えに、そうした対処が困難になるというトレードオフを本質的に抱えている。

4.14 ニューラル機械翻訳の多言語化

SMTにおいては、複数の言語ペアの対訳コーパスを総合的に用いて翻訳精度を向上させることは容易ではなかった。これはSMTが基本的には記号の置き換えによる翻訳であったからである。古くから中間言語という考えがあり、中間言語にさえ変換できれば、どの言語からどの言語へも翻訳ができると仮定されてきたが、有効な中間言語は未だに定義されていない。これに対し、NMTでは翻訳の過程は行列計算であり、必要な情報はすべて数値で表現されている。つまり、どの言語であっても、同じ内容を表している単語や文を同じような値を持つベクトルに変換できれば、これが中間言語的な役割を果たし、どの言語間でも翻訳が行えるし、さらには直接の対訳コーパスがない言語対間の翻訳も行えるようになると考えられる。

実際にこのような問題設定に取り組んだ研究はいくつか発表されている。１つの言語から複数の言語に翻訳するもの（one-to-many）[36]や、逆に複数の言語から１つの言語に翻訳するもの（many-to-one）[37]、複数の言語から複数の言語に翻訳するもの（many-to-many）[38]、そして直接の対訳コーパスがない言語対でも、複数の対訳コーパスを同時に利用することで翻訳を可能とするもの（zero-shot）[39] [40]などがすでに提案されている。

[36] Dong, Daxiang, Wu, Hua, He, Wei, Yu, Dianhai, and Wang, Haifeng. Multi-task learning for multiple language translation. In ACL, 2015.

[37] Barret Zoph and Kevin Knight. 2016. Multi-source neural translation. In NAACL.

[38] Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. Multi-way, multilingual neural machine translation with a shared attention mechanism. In NAACL.

[39] Orhan Firat, Baskaran Sankaran, Yaser Al-onaizan Fatos T. Yarman Vural and Kyunghyun Cho. 2016. Zero-Resource Translation with Multi-Lingual Neural Machine Translation. In ACL.

[40] https://arxiv.org/abs/1611.04558

また、大規模な対訳コーパスが利用可能な言語対で一旦 NMT を学習し、少量の対訳コーパスしかない他の言語対の翻訳モデルの初期パラメータとして流用する転移学習（transfer learning）という方法も提案されている[41]。一般的に NMT は少量の対訳コーパスしか利用できない場合には翻訳精度が著しく低下するが、転移学習を行うことで Phrase-based SMT と同等程度の翻訳精度を達成している。

4.15 まとめ

SMT と NMT の研究状況を簡単にまとめる。NN の翻訳への応用は学術的に目覚ましい成果をあげ、アカデミアにおける研究のパラダイムは大きく変わりつつある。既に SMT で利用した対訳データの蓄積があるので、SMT の出現当時に比べて、NMT はアルゴリズムの進化が速い。この１年余りで、NMT は対訳データが著しく不足するような状況を除き、平均的には SMT の翻訳精度を上回った。学術論文や特許の翻訳を対象とした共通タスクの結果[42]でも、上位のシステムはほぼ NMT という状況である。そのため、機械翻訳の研究者の興味は NMT に移っており、SMT はマイナーな研究分野になってしまった。

次に SMT と NMT の実用面を簡単にまとめる。NMT における未解決の問題として、特許のように大規模な用語対訳が存在するような場合にそれを活用する術が明らかでないことが挙げられる。しかしながら、その他の面、翻訳結果の流暢さや、従前の SMT で大きな問題となっていた言語による語順の変化に対しては明らかに NMT が SMT より優位に立っており、この状況は今後も変わらないものと考えられる。実用的にも、少ない主記憶容量で翻訳処理が可能な NMT は有利である。

一方、精度向上のための工夫の観点から見ると SMT から NMT への移行に伴い、これまで SMT で培われてきた様々な訳質改善のためのノウハウが NMT では利用できない可能性がある。例えば、PRE-ORDERING は NMT においては精度を低下させる可能性があることが示唆されている[43]。また、SMT のようにフレーズテーブルから誤訳の原因を取り除いて翻訳結果を調整することも NMT では不可能である。その他、仮想単語の導入、プレースホルダを利用した用語置換等、今後 NMT における有効性について検証が必要であろう。

短期的には、実用面のノウハウがあるため SMT が精度面で満足できるレベルにあるのであれば、あえて発展途上の NMT に移行しなくても良いかもしれない。一方、数年単位のシステムである場合は、その間に NMT の研究が進み、SMT が一挙に陳腐化してしまう可能性がある。

最後に自動評価について触れる。これまで N-gram に基づく BLEU 等の自動評価尺度は

41 https://arxiv.org/abs/1604.02201

42 Toshiaki Nakazawa et al., Overview of 3rd Workshop on Asian Translation. In Proc. 3rd Workshop on Asian Translation, 2016.

43 Katsuhito Sudoh and Masaaki Nagata, Chinese-to-Japanese Patent Machine Translation based on Syntactic Pre-ordering for WAT 2016. In Proc. 3rd Workshop on Asian Translation, 2016.

SMT 学習時のチューニングに使われて SMT の発展を支えてきた。しかし、SMT の進歩や NMT の出現により、機械翻訳全体の訳質が向上していること、人手評価の代替とするため自動評価に対して以前より深いレベルでの訳質評価が期待されていることから、参照訳との N-gram の一致度に基づく自動評価手法は限界に達している。このような状況により、自動評価に NMT の考え方を取り入れてゆくなど、自動評価の技術が大きく変化していくことが求められている。