

特許翻訳におけるドメイン 適応型機械翻訳



日本特許翻訳株式会社，代表取締役社長 **本間 奨**

要 約

AI 翻訳（ニューラル機械翻訳）技術の進展が著しい。翻訳業界でも特許出願明細書の翻訳に AI 翻訳を取り入れ翻訳の迅速化並びに生産性を向上する動きが加速してきている。本報告では、最初に機械翻訳システムの翻訳精度の自動評価法を紹介し、この評価法をもとに、特に特許にフォーカスした NICT の新世代翻訳エンジンの実力を中心に Google 翻訳と比較レビューする。これらの汎用型翻訳エンジンに対して、特許や医薬などの産業翻訳で要求される特定分野の特定文型や用語に適切に対応可能なドメイン適応型機械翻訳について解説を行う。ドメイン適応型機械翻訳エンジンと CAT ツールを統合化した統合翻訳環境では、簡単なマウス操作で大量の対訳や翻訳メモリなどからドメイン適応型機械翻訳エンジンを構築し、これを CAT ツールに組み込むことが可能となる。最後に、このような統合翻訳環境を特許や医薬の翻訳に活用した事例などを紹介する。

目次

1. はじめに
2. 対訳コーパスによる機械翻訳技術の発展
3. 機械翻訳の評価法
4. 翻訳システム間の精度比較
 4. 1 翻訳システム
 4. 2 特許公報の精度比較
5. ドメイン適応型機械翻訳の詳細
 5. 1 汎用型とドメイン適応型の翻訳例
 5. 2 汎用型、分野特化型、ドメイン適応型機械翻訳システムの精度比較
 5. 3 統合型翻訳環境でのドメイン適応型翻訳エンジンの活用
6. おわりに

1. はじめに

国立研究開発法人) 情報通信研究機構（以下 NICT）は、特許庁 - NICT 協力合意により特許庁の保有する大量の訓練用翻訳データをベースに特許専用の統計的機械翻訳（SMT）およびニューラル機械翻訳（NMT）エンジンを開発、特許庁では NICT 翻訳エンジンを庁内の審査官用の検索システムにとどまらず J-PlatPat の中国特許検索用およびワンポータルドシエの日英翻訳サービスに用いている⁽¹⁾。

NICT は、さらに特許庁で用いられている現世代型特許 NMT エンジンから、より高精度な Transformer

アルゴリズム（図 1 参照）の新世代エンジン、汎用 NT および特許 NT を一昨年 11 月に公開した。

NICT 以外にも国内ではみらい翻訳 Mirai Translator（登録商標）やロゼッタ T-400 及び T-3MT、国外では Google 翻訳 や Microsoft Translator に続き DeepL Translator などのあらたな NMT 翻訳サービスが提供されている。

これらの翻訳技術の進展を背景に、日本出願を海外の言語に翻訳する際の翻訳文作成方法も大きな変革が起きようとしている。人手から AI による翻訳へ移行しようとする変革期にある現状で、AI の最先端の翻訳技術を比較検討する場合、どのような点に注目して評価すればよいかを 3. で詳細を述べ、最新の NMT 翻訳システムの特許公報などに関する評価結果を 4. に、また本報告の主題であるドメイン適応機械翻訳に関する詳細を 5. にまとめた。(1~4 については筆者による（一社）日本知的財産協会「知財管理」Vol.70 No.9 2020 pp.1253~1264 の原稿を加筆・修正した)

2. 対訳コーパスによる機械翻訳技術の発展

対訳コーパスベースの機械翻訳システムの歴史は長く、翻訳しようとする原文に似た対訳を流用して組み合わせることで翻訳する用例ベース機械翻訳 (example-based machine translation, EBMT)⁽²⁾ が最も古く、次に SMT（統計的機械翻訳）が IBM によっ

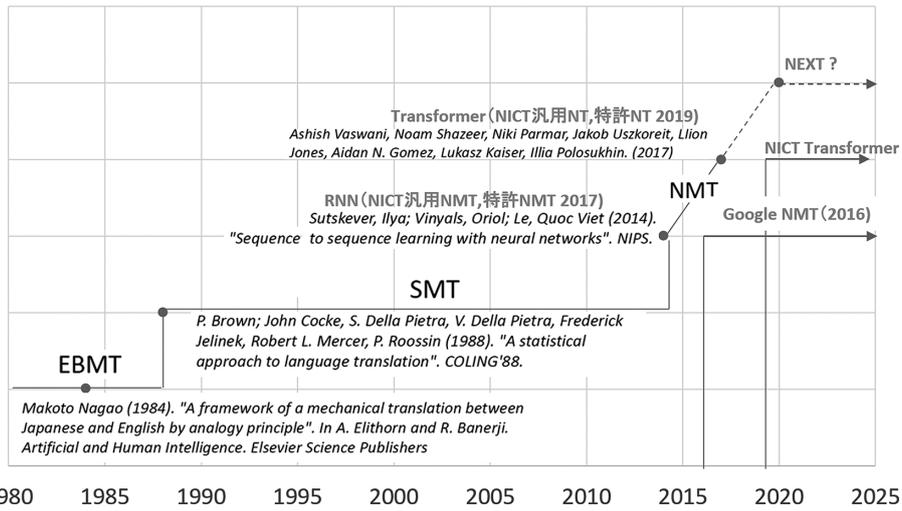


図1 コーパスベース MT アルゴリズムの進展 (「日経エレクトロニクスセミナー」P7 2019/12 NICT 内山将夫氏講演資料に基づく)

て研究が開始された⁽³⁾。SMTでは、大量の対訳文から統計モデルを作成する。これに対してNMT(ニューラル機械翻訳)では、1つのニューラルネットワークを用いて訓練-翻訳とも同じフレームワークで行う。対訳を与えるだけで、ニューラルネットワークが翻訳に必要な情報を自動的に学習することができる。AI翻訳とはこのNMTの技術に基づく機械翻訳である。NMTの実用化はGoogleが2016年に先鞭をつけた。NICTは2017年に第一世代のNMT(Sequence to Sequence Learning⁽⁴⁾)をリリース、2019年には第二世代のNMT(Transformer⁽⁵⁾)をリリースした。これらは国内の特許分野ではデファクトの翻訳エンジンとして広く利用されている。2020年に入ると、Transformerより高精度と云われているDeepLが欧州言語に加えて日本語・中国語に対応した⁽⁶⁾。NMTの解説は中澤⁽⁷⁾を、Transformer以降のNMTの技術的な進展の例としては、Sergey Edunov⁽⁸⁾を参照されたい。

3. 機械翻訳の評価法

機械翻訳されたテキストの翻訳精度を自動的に評価するための指標として、代表的な指標にBLEU(BiLingual Evaluation Understudy)⁽⁹⁾がある。BLEUは、機械翻訳されたテキストと高品質な(人手翻訳による)参照訳の類似度を表す。BLEUはコーパスベースの指標であり、一般的にはBLEUなどの自動評価指標を個々の文の評価に使用した場合はうまく機能しない。BLEUを計算する際には、コーパス全体で算出されるのが一般的である。

BLEU 解釈 (Google)⁽¹⁰⁾

- <10 ほとんど役に立たない
- 10~19 主旨を理解するのが困難である
- 20~29 主旨は明白であるが、文法上の重大なエラーがある
- 30~40 理解できる、適度な品質の翻訳
- 40~50 高品質な翻訳
- 50~60 非常に高品質で、適切かつ流暢な翻訳
- >60 人が翻訳した場合よりも高品質であることが多い

BLEUのスコア計算は、参照訳とMT訳との間のNグラム(通常Nは1~4)が一致した数を数える。欧米言語間のような語順の近い言語間では人手翻訳との相関はよいとされるが、日英・英日のように語順が大きく異なる言語対では問題がある。RIBES(Rank-based Intuitive Bilingual Evaluation Score)⁽¹¹⁾は、参照訳とMT訳との間で共通して出現する単語の出現順序に着目した自動評価法で、日英翻訳などではBLEUと並んで評価指標として用いられる。

日本特許庁では特許翻訳用の翻訳システムの要件の一つに、5000文程度のコーパスにおいて、BLEU/RIBESそれぞれが評価基準値を上回ることにしている。大雑把に言うと、BLEUは訳語適切性と訳語の過不足のなさをRIBESは語順の正確さを反映している。したがって、日本語を含む翻訳精度評価では、BLEUやRIBESなど1つの指標のみで翻訳精度評価をおこなうべきではなく、BLEUとRIBESのそれぞれを評価する必要がある。本報告ではBLEUとRIBESの2次元空間の位置で翻訳システムの精度評価を行った。またそのベクトル長に相当する次の値sqrt((3*

BLEU² + RIBES²) / 4) で「スコア」を算出し、精度の目安とすることも可能である。BLEU, RIBESとも0~1の実数であるが、以下ではそれらを100倍した0~100の実数とした。

4. 翻訳システム間の精度比較

4.1 翻訳システム

(1) 汎用型翻訳システム

適用分野が広く、また扱える言語も広く選択できる。代表的な例として以下の4つの翻訳システムがある。

① Google, Microsoft, Amazon など：訓練データはインターネットをはじめあらゆる分野のデータを使用しているため、適用範囲は最も広い。

② NICT：汎用 NT (NICT の第二世代：Transformer) は、総務省翻訳バンクの仕組みで収集された特許や医薬、IT、など複数分野にまたがる良質な訓練データにより構築された NICT 新世代 NMT である。

③ DeepL：DeepL translator は 2020 年 3 月より欧州言語に加え日本語・中国語に対応した。同社のブログによれば、『Google 等の汎用翻訳システムは、ニューラル機械翻訳のトレーニングを行うための良いデータを豊富に持っているが、DeepL では、数学とニューラルネットワークの方法論において数々のイノベーションを達成したことで、他よりも一層優れた結果を出すことができるようになった。』とのことであり、NMT アルゴリズムに特徴がある。

④ みらい翻訳 (登録商標)：Mirai Translator (登録商標) は NICT との共同研究成果を利用し、みらい翻訳 (登録商標) にて製品化した NMT エンジンを搭載する機械翻訳サービスである。同社によれば、『和文英訳がプロ翻訳者レベル、英文和訳は TOEIC960 点レベル、日中翻訳は翻訳者による翻訳と同等レベルの翻訳が可能』としている。

(2) 分野特化型翻訳システム

特許や医薬、法務などの各分野に特化したタイプで、扱える言語は限定され、精度は汎用型とドメイン適応型機械翻訳システムの中間である。ドメイン適用型翻訳エンジンのベースエンジンとしても用いられる。

① NICT 特許 NT (NICT の第二世代：Transformer)：特許公報から作成した対訳のみで訓練した特許用エンジン。NICT 英日・日英は 2 億文程度、NICT 中日・日中は 1 億文程度の文対を用いてエンジンを構築した

と推定される。

② ロゼッタ T-3MT：ロゼッタ社では、医薬、特許、会話、メールなどの分野別にあらかじめ翻訳エンジンを用意して、ユーザーが選択可能としている⁽¹²⁾。

③ Mirai Translator 「契約書・法務モデル」：みらい翻訳では、大手法律事務所の協力を得て開発した「契約書・法務モデル」特化型エンジンが用意されている。

(3) ドメイン適応型機械翻訳システム

訓練データと同じ言語および同種の原文に限定されるが最も高精度である。顧客の保有する対訳や翻訳メモリをもとに独自翻訳システムを構築する。ベースエンジンに対して、対訳による追加学習 (アダプテーション) したエンジンに基づく翻訳システム。対訳文の精度が高い場合、この訓練データと同種の原文については最高精度の翻訳結果を得ることが可能であるが適用範囲は限定される。

① NICT アダプテーション+EBMT (NICT の第三世代エンジン)：npat ProTranslator などでは NICT 特許 NT エンジンをベースエンジンとして、顧客対訳データや翻訳メモリなどから NICT アダプテーション+EBMT エンジンを構築し、Trados/Memsource/MemoQ などの CAT ツールと連携した統合型翻訳環境を提供している。

② ロゼッタ T-3MT：ロゼッタでは分野特化型エンジンをベースに、顧客対訳データや翻訳メモリなどからドメイン適応型機械翻訳構築を可能としている。

③ RWS Language Weaver：Trados と連携して、Trados 翻訳メモリからドメイン適応型機械翻訳構築を可能としている。Trados と連携した統合型翻訳環境を提供している。

(4) マルチ NMT 翻訳システム⁽¹²⁾

マルチ NMT 翻訳では、翻訳セグメント単位に、汎用翻訳システムと分野特化型翻訳システム、アダプテーションエンジンを同時翻訳して、さらに逆翻訳をおこない、各エンジンのスコアを求め、最高スコアの翻訳結果を出力することで特許文書から科学技術文献まで幅広い分野で高精度な翻訳を提供することが可能である。自動翻訳に適した翻訳システムといえる。

4. 2 特許公報の精度比較

Google 翻訳, NICT 汎用 NT, NICT 第一世代の特許 NMT および第二世代の特許 NT の 4 システムについて精度比較した。要約+請求項+詳細な説明各 5,000 文の BLEU/RIBES 平均値により評価した。NICT 特許 NT が最高スコアを示し, 日英では BLEU47.0, RIBES88.2 と英日の評価値 BLEU45.1, RIBES85.5 よりも高いスコアを示した。

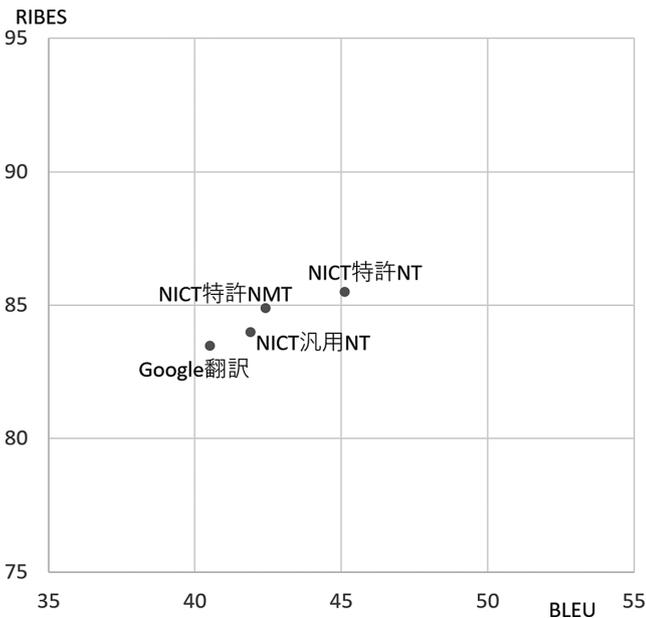


図 2 特許公報の英日翻訳評価

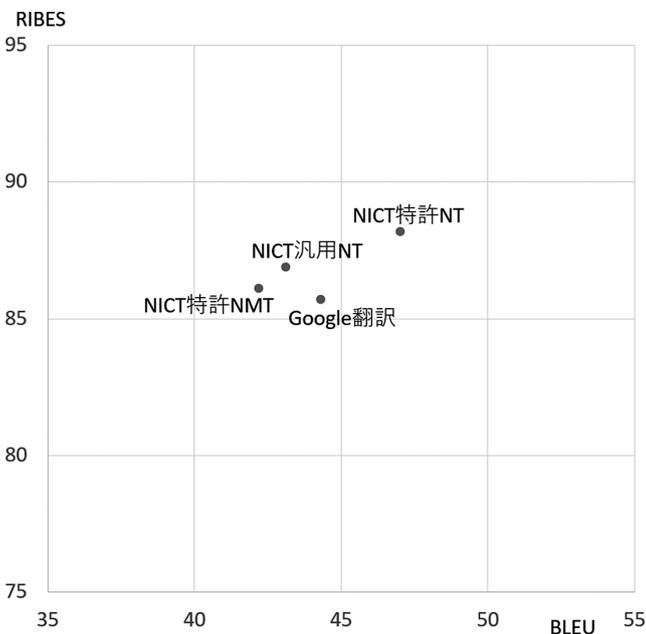


図 3 特許公報の日英評価

5. ドメイン適応型機械翻訳システムの詳細

5. 1 汎用型とドメイン適応型の翻訳例

本報告では, 分野は, 特許 (全分野) や医薬 (全医薬分野) などを指すのに対して, ドメインは分野より狭い領域を指している。特許でいうと, 特定出願人の

特定 IPC サブクラス程度をイメージしている。例としては, 出願人が富士通で, IPC サブクラス G06F などを想定しているが, 翻訳メモリなどの非常に狭い領域をドメインとしてもかまわない。ドメインが狭い場合は訓練文対数も少数でよく, NICT アダプテーション+ EBMT エンジンでは, 1,000 文対程度あれば訓練は可能である。一方, ドメインが狭い場合は, 適用範囲はより限定されてくる。

Google 翻訳, マイクロソフト翻訳, DeepL, NICT 汎用 NT などの汎用型翻訳システムは, 学習量は最も大きいエンジンであるが, 訳文は平均化された表現になる。特許や医薬分野のような産業翻訳では, 平均的な表現では翻訳で要求される特定表現に十分対応できないケースが少なくない。特定分野, 統制された特定表現に限定された文対でベースとなる機械翻訳エンジンを訓練すると, 訳文はドメイン固有の表現で翻訳されるようになる。汎用型翻訳エンジンとドメイン適応型機械翻訳の違いを以下の簡単な例で説明する。

【原文】

Pharmacist/designee (i.e. delegated site staff authorized to perform study specific tasks per the “site signature/delegation log”) will inspect the content of IP shipment immediately after the receipt.

この原文は, 医薬分野の特定ドメインであり, この原文を用いて, 汎用型とドメイン適応型の簡単な比較を行った。汎用型翻訳エンジンとしては, Google 翻訳, DeepL, NICT 汎用 NT, みらい翻訳の 4 つを取り上げそれぞれの翻訳結果を示す。

【汎用型 Google】

薬剤師/被指名人 (つまり, 「サイト署名/委任ログ」に従って調査固有のタスクを実行する権限を与えられた委任サイトスタッフ) は, 受領後すぐに IP 出荷の内容を検査します。

【汎用型 DeepL】

薬剤師/被任命者 (「施設の署名/委任記録」に基づいて試験特有の作業を行うことを許可された委任された施設スタッフ) は, IP 貨物の受領後直ちに内容物を検査します。

【汎用型 NICT 汎用 NT】

薬剤師/被指名人 (すなわち, 「サイト署名/委任ログ」に従って試験固有の業務を実施する権限を与えられた委任されたサイトスタッフ) は, 受領後直ちに

IP 発送の内容を検査する。

【汎用型みらい翻訳】

薬剤師／被指名人（つまり、「サイトの署名／委任ログ」ごとに調査固有のタスクを実行する権限を与えられた、委任されたサイトスタッフ）は、受領後速やかに IP 出荷の内容を検査する。

上記汎用型エンジン対して、NICT 汎用 NT をベースエンジンとし、NICT アダプテーション + EBMT を用いて医薬のあるドメインで訓練したドメイン適応型では、以下のような翻訳結果が得られた。

【ドメイン適応型】

薬剤師／被指名者（すなわち、「実施医療機関の署名／委任記録」に従って治験特有の業務を遂行する権限を与えられた委任された実施医療機関の職員）は、受領後直ちに治験薬の発送内容を確認する。

汎用型では、医薬の特定ドメインの翻訳に対して、多くの誤訳があるため、機械翻訳を使っても生産性はさほど向上しない。汎用型エンジンに用語辞書で対応することも一案であるが、例えば誤訳の原因の一つとなっている「study」を治験と名詞で辞書登録した場合、別の原文で「study」が動詞で用いられた場合や別の意味の名詞として用いられた場合には、あらたな誤訳となってしまいます。さらに 1 文に複数用語がある場合の辞書適用では、AI 翻訳精度が低下してしまう。用語辞書を AI 翻訳に用いる場合は、このような点に留意する必要がある。一方、ドメイン適応型翻訳においてはこのような問題は起こり得ない。

5. 2 汎用型、分野特化型、ドメイン適応型機械翻訳システムの精度比較

特許分野における特定ドメインとして、出願人が富士通、IPC 分類で G06F の公報からドメイン適応型翻訳エンジンを構築した。具体的には、前記ドメインの日本公報と対応米国公報の請求項の集合のうち、4,037 文対を訓練データとして用いた。同じ集合から 504 文対を評価用文対として用いた。ベースエンジンを NICT 特許 NT とし、アダプテーション + EBMT という少量文対に対応した NICT の訓練システムを用いて前記 4,037 文対で訓練して、ドメイン適応型翻訳エンジンを構築した。訓練時間はおよそ 30 分程度であった。

図 4 に、汎用型として Google 翻訳と DeepL を、分野特化型として NICT 特許 NT を、ドメイン適応型

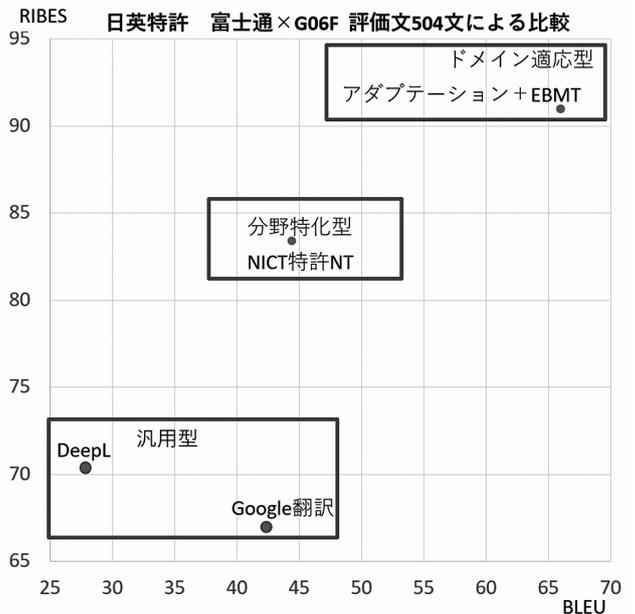


図 4 富士通×G06F 請求項ドメインの公報の対訳 504 文対による汎用型、分野特化型、ドメイン適応型機械翻訳システムの精度比較

として NICT アダプテーション + EBMT の BLEU と RIBES スコアをまとめた。図 4 から明らかなように特定ドメインに関する自動評価スコアでは、

汎用型 < 分野特化型 < ドメイン適応型
のように精度が向上していくことがわかる。

5. 3 統合型翻訳環境でのドメイン適応型翻訳エンジンの活用

上記のようなドメイン適応型エンジンの構築は、翻訳メモリやこれまでの翻訳資産の対訳を訓練データとして用いることになる。従って、だれでも簡単にエンジンを構築するためには、①翻訳メモリ tmx や対訳エクセルから簡単にエンジン構築ができ、②それを CAT ツールの翻訳メモリのマッチ率の閾値をユーザが設定可能とし、閾値以上のセグメントは翻訳メモリの結果を、閾値未満についてはドメイン適応型エンジンによる翻訳結果を統合する仕組みが必要となる。

このようなドメイン適応型エンジンと CAT ツールとの統合化については、すでに RWS Language Weaver (Trados との連携) や npat ProTranslator (Memsource API による統合環境に加え Trados/MemoQ との連携も可能) などの統合型翻訳支援サービスが発表されている。後者は NICT 特許 NT をベースとするドメイン適応型翻訳エンジンに特化した翻訳サービスである。

事務機の業界団体である JBMIA では、複合機の技術ドメイン単位のエンジン構築を行い、JBMIA のメ

ンバー企業（キヤノン，京セラドキュメントソリューションズ，コニカミノルタ，ScienBiziP Japan，セイコーエプソン，東芝テック，富士フイルム，富士フイルムビジネスイノベーション，ブラザー工業，リコーなど14社）において外国出願案件業務への適用へ向けたフィービリティ・スタディが進められている。また医薬翻訳大手のアスカコーポレーション並びに特許翻訳大手の知財コーポレーションなどで自社で構築した大量の対訳や翻訳メモリを用いてドメイン適応型翻訳エンジンを構築し，翻訳業務に活用されている。

6. おわりに

医薬や特許などの産業翻訳では，Google 翻訳や DeepL などの汎用型翻訳エンジンから，ドメイン適応型翻訳エンジンへと移行しつつあり，本報告が特許翻訳にかかわる読者の参考になれば筆者としては望外の喜びである。

(参考文献)

(1) 東芝デジタルソリューションズ株式会社 ニュースリリース 2018年7月10日

- (2) Makoto Nagao, "ARTIFICIAL AND HUMAN INTELLIGENCE" chap. 11 1984
- (3) P. Brown et. al. "A Statistical Approach to French/English Translation"
- (4) Ilya Sutskever, Oriol Vinyals, Quoc V. Le, "Sequence to Sequence Learning with Neural Networks" 2014
- (5) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention Is All You Need" 2017
- (6) <https://www.deepl.com/blog/20200319.html>
- (7) 中澤敏明, 「機械翻訳の新しいパラダイム：ニューラル機械翻訳の原理」情報管理 2017年60巻5号 p.299-306
- (8) Sergey Edunov, Myle Ott, Michael Auli, David Grangier, "Understanding Back-Translation at Scale" 2018
- (9) Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "a Method for Automatic Evaluation of Machine Translation" 2002
- (10) <https://cloud.google.com/translate/automl/docs/evaluate?hl=ja> (参照日：2020.5.4)
- (11) 平尾努 磯崎秀樹 Kevin Duh 須藤克仁 塚田元, 「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)
- (12) JTF JOURNAL #306 2020 03/04 p.10-11, 18-19

(原稿受領 2021.11.5)