

ニューラル MT を構造化用語データ UTX で補完して 翻訳品質を改善する簡便なポスト エディット手法

山本ゆうじ/AAMT (アジア太平洋機械翻訳協会) <http://cosmoshouse.com/mail.htm> (個人連絡先)

概要 機械翻訳では、翻訳結果が用語集に沿っているか、また用語に一貫性があるかは、これまで必ずしも重要視されてこなかった。だが、商業翻訳では用語の正確性が重要になることがある。それまでの手法と比較して大きく改善されたニューラル機械翻訳でも、用語の翻訳に関しては、一貫性がない、訳語が欠落するといった点が指摘されている。品質を必要としない概訳の場合はこれでもよいが、商業翻訳で機械翻訳を活用する場合、用語の正確さが保証されないのは致命的な問題である。本論考では、ニューラル機械翻訳の翻訳結果を、シンプルな構造化用語データ形式 UTX でチェックし修正する簡便なポスト エディット(後編集)作業により翻訳品質を改善する手法について検討する。

対訳キーワード

term:ja	用語管理	体系的翻訳	コンピューター 支援翻訳	機械翻訳	用語集	用語ベース	オープン データ
term:en	terminology management	systematic translation	computer- assisted translation	machine translation	glossary	termbase	open data

term:ja	翻訳品質評価	ポストエディット	NMT
term:en	translation quality assessment	post edit	NMT

A simple post-editing method to improve the translation quality of neural MT results by supplementing with UTX, a structured glossary data

Yamamoto Yuji/AAMT (Asia-Pacific Association for Machine Translation)

Abstract In machine translation, whether the result matches a glossary or the term usage is consistent has not always been a priority. However, the term-level accuracy can be essential in commercial translation. While neural machine translation (NMT) shows a significant improvement over previous approaches, it cannot incorporate a glossary and lacks term consistency. Although this might be acceptable for quick translation that requires less reliability, it is a fatal problem when using MT in commercial translation. This thesis examines a method for improving NMT quality by a quick post-editing, that is, checking and correcting the translation result with the simple structured term data format UTX.

1 はじめに

用語の正確さは、機械翻訳の研究では必ずしも重要視されていない。用語の正確さは、一般ユーザーよりも商業翻訳での優先事項からである。精度が高いと言われる NMT (neural machine translation:ニューラル機械翻訳)でも用語の翻訳に関しては、一貫性がない、訳語が欠落するといった点が指摘されている。品質を必要としない概訳の場合はこれでもよいが、商業翻訳で MT を活用する場

合、用語の正確さが保証されないのは致命的な問題である。本論考では、ニューラル NMT の翻訳結果を AAMT(アジア太平洋機械翻訳協会)が策定したシンプルな構造化用語データ形式 UTX でチェックし修正するポスト エディット(後編集)作業により翻訳品質を改善する手法について検討する。特に構造化用語データとはなにか、体系的翻訳の観点からの機械翻訳評価方法の課題はなにか、人間翻訳指標にどのようなものがあるか、NMT の用語に関する特性はどのよう

なものがあり、具体的な課題はなにか、用語検証はポスト エディット作業としてどのように有効か、といった点である。

2 構造化用語データ UTX

構造化用語データとは、自動的なチェックでの用途に耐えうる調整が施され、用語管理がされたデータである。単純に用語を抽出しただけで未調整の用語リストは構造化用語データではない。構造化用語データは、データとして処理できるように、各種のフィールドが定義されている必要がある。また、用語に関与するステークホルダーと承認過程が定義されている必要がある。さらに各項目が、特定目的の用語データとしてふさわしいか、用語管理者の承認を得られるようになっていなければならない。またスタイル ガイドが別途定義され、スタイル ガイドと用語データの間で齟齬がないようにする必要がある。本来、企業や組織で体系的翻訳を行う際には、このような構造化用語データが必要となる。構造化用語データなしでは、用語が分かりにくく、不統一で、読者を混乱させる。共同編集者や翻訳者が関与する場合、編集や翻訳にも、本来は避けられるはずの労力・時間・費用を要することとなる。残念ながら、企業や組織で理想的な体系的翻訳が行われているとは限らない。だが、組織を管理する立場にある者は、体系的翻訳を理解し、構造化用語データの作成をしなければならない。

UTX は、AAMT (アジア太平洋機械翻訳協会) が策定した用語集形式の標準規格である。UTX は、前述の要件を満たす構造化用語データを記述できる。AAMT は非営利団体であり、UTX 仕様を無償で提供している<<http://www.aamt.info/japanese/utx/>>。UTX は特許庁や製造業などの企業・組織での用語管理や機械翻訳に使われている。

3 体系的翻訳の観点からの機械翻訳評価方法の課題と人間翻訳指標

従来の機械翻訳評価方法では、用語の正確さという観点が欠落している。たとえば自動評価尺度 BLEU や METEOR では、機械翻訳訳文が人間翻訳にどれだけ近いかを評価する。しかし、機械翻訳評価方法では、参照訳となる人間翻訳の翻訳評価は問題にされない。翻訳対象文書と用語集との関連性は一定ではなく、用語管理者の技能、用語集の品質によって異なる。これが、用語集が機械翻訳評価から除外される理由であろうが、今後は人間翻訳指標にもっと目を向ける必要があるのではないだろうか。

機械翻訳評価が正しく行われるには、参照訳の品質、参照訳が用

語集に沿っているか、原文の品質などの観点も必要に応じて取り入れるべきと考えられる。人間翻訳の評価指標としては、LISA QA 指標、SAE J2450 QA モデル、TAUS DQF などがある。LISA QA (quality assurance) 指標は、LISA (Localization Industry Standards Association) によって 1990 年代に策定された。SAE J2450 QA モデルは、航空・自動車業界のエンジニア組織である SAE International によって 2001 年に公開され、評価項目は 7 つと少ない。TAUS DQF (Dynamic Quality Framework) は、TAUS (Translation Automation User Society) によって 2012 年に策定された。DQF では、古い指標には欠落している、翻訳支援ツールや用語ベースなど新しい技術的観点が考慮されている。さらに EU により策定された MQM (Multidimensional Quality Metrics) がある。MQM は単独の評価指標というより、評価指標のフレームワークであるとされている。DQF と MQM は調整が進められ、DQF は MQM のサブセットとなっている。この他に、品質評価指標そのものではないが、翻訳サービス提供者の品質を規定する規格である ISO 17100、そしてそのベースとなった EN 15038 がある。EN 15038 は、CEN (European Committee for Standardization) により策定された。これらすべての人間翻訳指標で、用語集が評価項目として挙げられており、用語集の重要性がうかがえる。ただ、これらすべての人間翻訳評価は自動化されておらず、手作業で行う必要があり、手間がかかる。そのため、訳文を抜き出してチェックさせるを得ないこともあった。しかし、用語のチェックは、構造化用語データを使用することにより自動化できるため、全数チェックできる。

原文の品質も翻訳に関係する重要な要素だが、これには機械翻訳評価も人間翻訳指標も適用できない。日本語文書、特に技術文書では、日本語の体裁として極めて不完全であり、著者以外だれが読んでも理解できない低品質な文章がしばしば見受けられる。またテキストの品質とは別に、書式上の問題で、ハード リターンにより途中で文が分割されていると、文としてそもそも不完全な形で機械翻訳処理される。日本語文章、特に特許文書の品質については、産業日本語研究会 <<https://www.tech-jpn.jp/>> で指標が検討されている。

このように、現状では機械翻訳の自動訳文評価には用語の観点は無いが、今後は、人間翻訳指標、原文品質、用語の観点を考慮した指標が必要とされるであろう。

4 NMT の問題点と効果的なポスト エディット手法の仮説

NMT の精度はそれ以前の機械翻訳と比較してかなり高精度ではあるものの、問題点として、訳語が類似の言葉にすり替わることがある。また、訳語の使い方に一貫性がない。たとえば NMT に基づく Google Translate <<https://translate.google.com/>> で UTX 仕様書を英日翻訳したところ、誤訳はあるものの、大体の意味を把握する概訳としてはかなり理解できるものであった。しかし、用語レベルで見ると、"glossary contributor" という同一の語が「用語集の寄稿者」、「用語集の貢献者」などと複数の訳語に訳された。"term status" は、「用語ステータス」、「期間ステータス」、「タームステータス」などと訳された。さらには「語学ステータス」というまったく関係ない訳語もあった。"target language" は、「ターゲット言語」、「目標原語」、「対象言語」などと訳された。科学分野の文章の訳文でよく似た名前の薬物や化学物質が他のものにも変わるのとは重大な問題である。また現在の Google Translate は文単位では理解できるレベルに訳せていても、「訳語が一貫していないと読者が混乱する」という人間にとっての常識に配慮されるように作られてはいない。人間の読者にとっては、一つの文の中のいわば「横系」だけでなく、文どうしをつなぐ「縦系」の一貫性が必要である。

機械翻訳でのポスト エディットの問題として、機械翻訳の問題を修正する作業は全文を読む必要があり、ポスト エディット者の負担が増えるということがある。人間の補助として役立つどころか、翻訳料金が引き下げられたにも関わらず必要な労力に大差ないのは不条理と言わざるを得ない。

だが構造化用語データがある場合、用語の問題に限定してチェックを行えば、問題は比較的容易に修正可能であるという仮説が立てられる。筆者は、簡便なポスト エディットの手法として、NMT の結果を用語データで確認し、修正するという手法を提案する。この手法の利点は、作業量の見積もりが困難なポスト エディットで、なにを修行して修正すべきかを明確にし、最小限のポスト エディット作業で、用語レベルのみの正確さを保証できる点にある。

5 検証の手順

以下の手順で前項の手法の検証を行った。

1. UTX 仕様書 (Microsoft Word 形式) の英語版を、NMT を採用する Google Translate で日本語に翻訳する。

2. この結果を対訳形式に変換して、翻訳メモリーにインポートする。
3. 翻訳支援ツール SDL Trados Studio で一括翻訳する(翻訳メモリーの内容を訳文文書に適用するだけで人間翻訳ではない)。
4. 機械翻訳の結果に対して、SDL Trados Studio で用語検証を行う。

用語検証には事前に作成した用語ベースを使用した。なお UTX 仕様書は、英文で作成されたのち、日本語に人間翻訳された。用語ベースはこの人間翻訳で用語統一をする必要から作成されたものであり、機械翻訳の修正用目的のみで作成されたものではない(換言すれば、適切に作成された UTX が多目的に柔軟に活用できることが示されているといえる)。

Google Translate で原語部分の出力が正しくされなかった文を除く 550 の翻訳単位のうち、書式が混在している分節をさらに除外した。結果的には、文書上にある 560 の分節が翻訳された。

用語の評価観点として、(1)用語が一貫して訳されているか、(2)用語データ(あるいは用語のデータベースである用語ベース)に沿っているか、の 2 点がある。観点(2)が満たされれば観点(1)も満たされることが多いが、用語データとして定義されていない用語については観点(1)も必要となる。ここでは主に観点(2)のみに注目した。

6 検証結果とその考察

検証の結果、SDL Trados では 23 件の禁止用語が検出され、115 件の不適切な訳語が検出された(次ページ図 1)。訳語同様に、禁止用語もまた事前に作成された構造化用語データである。人間作業、つまりポスト エディターの作業は、これらの用語をチェックして、修正することのみである。ポスト エディターはすべての文章を読む必要がなく、最低限の修正、すなわちライト エディットのみが求められる案件でも、責任範囲が明確になる。

これらの用語は、あくまで SDL Trados の設定に基づいて問題があると推測された用語である。修正が必要かどうかは、ポスト エディターが判断する。たとえば、UTX 仕様書の特徴として、一つの用語が原語のままにすべき箇所と訳すべき箇所が混在している箇所がいくつかある。たとえば、term という原語は、訳文でも term のままにすべき場合と、「用語」と訳すべき場合がある。これは訳し分けが必要になるが、実際の翻訳案件では比較的多まらなケースと思われる。

また SDL Trados 以外に、無償の ApSIC Xbench 2.9 などのツールを使用して用語検証をした。ただし、原文であいまい一致(名詞複数形や動詞活用形の同一視)がされないため、検出できた問題は(重複を除外後)11 項目のみとかなり限定的であった。

7 結論と今後の展望

検証の結果、NMT である Google Translate の翻訳結果を、UTX 構造化用語データを事前に準備し、それを使用して用語検証をすることにより、極めて効率的にポスト エディットできることが示された。現状では、Google Translate の結果は手作業で原文訳文の対訳形式 TMX に変換する必要があるが、直接 TMX で結果出力できるようになれば、この作業はさらに容易になる。

NMT そのものに、用語データをなんらかの形で補充する方法を組み込むことは技術的には解決不可能な問題ではない。この処理には、用語データが必要にはなる。だが用語データの抽出と選定の支援には、自然言語処理技術を十分に活用することができる。

8 参考文献

AAMT. (2016). UTX 1.20 Specification. Retrieved January 12, 2017, from <http://www.aamt.info/english/utx/utx1.20-specification-e.pdf>

AAMT. (2016). 「UTX 仕様バージョン 1.20」 Retrieved January 12, 2017, from <http://www.aamt.info/japanese/utx/utx1.20-specification-j.pdf>

Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* (Vol. 29, pp. 65-72).

Burchardt, A., & Lommel, A. (2014, November 19). Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality. Retrieved January 12, 2017, from <http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf>

German Research Center for Artificial Intelligence (DFKI) & QTLanchPad. (2015, December 15). Multidimensional Quality Metrics (MQM) Definition. Retrieved January 12, 2017, from <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 311-318). Association for Computational Linguistics.

The screenshot shows the SDL Trados Studio interface. The 'Messages (138)' window is open, displaying a list of error messages. The messages are categorized by severity: 3 'Wrong usage of the term' errors (marked with a red 'x') and 15 'Target segment does not contain the translated term' warnings (marked with a yellow triangle). The messages are sorted by origin, with 'Terminology Verifier' being the source for all. The messages include errors like 'Wrong usage of the term "エントリー" - this term is defined as forbidden.' and warnings like 'Target segment does not contain the translated term for "user role" - "ユーザー ロール" expected.' The 'Term Recognition' window on the right shows a list of terms: 'body', '本文', 'glossary', '用語集', and 'グロッサリー'. The bottom status bar shows the document name 'UTX仕様 (Google Translated).stripped till almost plain EN.docx.sdlx.liff [Translation]' and a message: 'The body of a UTX glossary consists of entries in each line. 631'. A progress indicator shows '100%' and another message: 'UTX用語集の本文は、各行のエントリーで構成されています。'.

図 1 UTX 構造化用語データに基づく SDL Trados Studio での用語検証の結果