

【貴社は大丈夫?】対策すべき自動翻訳のリスクコントロール

Xtra

- はじめに
- 自動翻訳の誤訳事例①(観光webサイト)
- 自動翻訳の誤訳事例②(論文・開発資料)
- 自動翻訳の誤訳事例③(交通情報)
- 自動翻訳で誤訳が発生してしまう原因
- 誤訳のチェック方法と対策

翻訳トラブルを未然に防ぎましょう

Google翻訳やDeepLなど、各社サービスの向上によってさまざまな用途で自動翻訳の利用が広がっています。その一方で、訳文の品質を判断することは容易ではなく、「正しく翻訳できているのか？」という不安の声も多く聞かれます。

以前、大阪メトロの「堺筋線」を「**Sakai Muscle(サカイマッスル)**」と誤訳したまま、webページに掲載していたことが話題となっていました。また2019年の台風19号の際には、「**川の周辺に避難してください**」という誤った緊急メールが日系ブラジル人向けに配信されたことがありました。いずれも自動翻訳の「誤訳」が招いた情報事故です。

こうした社会問題にもなりつつある自動翻訳の課題について、本資料ではリスクコントロールのための傾向と対策をご紹介します。

自動翻訳の誤訳事例

実際にどんな誤訳が発生しているのか、実例のパターンを紹介



【観光webサイトの誤訳事例】

料理用語の誤り

自動翻訳では前後の文脈に合わせた翻訳が難しいため、料理用語とは無関係の訳になってしまうことがあります。特に日本料理では誤訳が発生しやすい傾向が見られました。

- 香物 → incense(お香)
- 油物 → oil(油)
- 先付 → Prelude(前奏曲、プレリュード)
- 造り → structure(構造)※
- 八寸 → eight dimensions(八次元)
- 舟盛り → boat height(ボートの高さ)

※「お造り」にすれば正しく英訳されるなど、日本語表記の改善で解決できる場合もあります。



【観光webサイトの誤訳事例】

お部屋情報や観光プランの誤訳

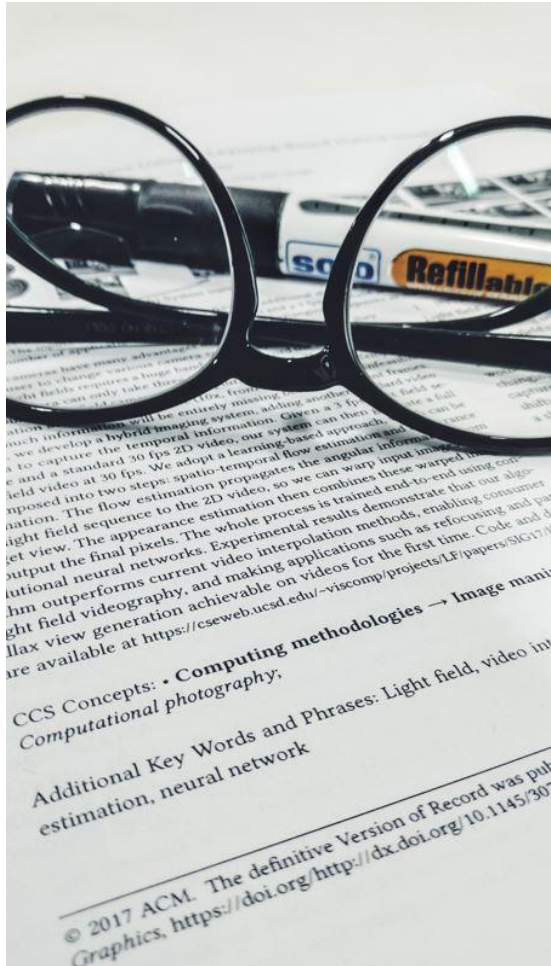
誤訳の中でも、特に訳漏れが原因になっているものが多く見られました。自動翻訳では訳漏れが発生しやすいことも特徴のひとつです。

Aプラン → 一个计划
(※「A」が抜け落ち、「計画」を示す表現のみ)

1泊朝食×2泊 → 1晩x 2晩
(※「朝食」が抜け落ちている)

楓の間(和) → Kaeda (Japanese)
(※「間」が抜け落ち、室名であることがわからない)

セミダブル 食事: 朝・夕 → 半双餐: 早晚
(※ベッドではなく、セミダブルが食事に係っている)



【論文・開発資料の誤訳事例】

過剰翻訳による誤り

原文には表記されていない番号が訳出されるなど、誤った文章が訳出されることがあります。

【原文】The media may be assembled similarly to the invention disclosed in U.S. patent application Ser. No. 15/64,72, filed Jun. 12, 2020 and entitled “THE DEVICE,” which is herein incorporated by reference.

【訳文】媒体は、米国特許出願第07 / 079, 064号に開示された発明と同様に組み立てられてもよい。2020年6月12日に出願され、「THE DEVICE」と題された米国特許第15 / 64, 72号は、参照により本明細書に組み込まれる。



【論文・開発資料の誤訳事例】

数字の抜け落ち・誤訳

数字は特に正確性が求められる部分だが、訳文で抜け落ちてしまったり、誤った表記になる場合があります。

【原文】この実験の効果について、**図3**を参照して詳細に説明する。

【訳文】The effect of this experiment will be described in detail with reference to **FIG.**

【備考】**図3**の数字だけが抜け落ちてしまっています。

【原文】第一三〇八・二〇一三号

【訳文】1st No. 308.213

【備考】「第一」が「1st」となり、2013の「0」が抜け落ちている。



【交通情報の誤訳事例】

地名や駅名の誤り

地名や駅名など、固有名詞は自動翻訳で誤訳・誤表記になりやすい代表例です。人名や企業名、製品名なども同様です。

- 富山きときと空港 → Toyama **Kitoki and** Airport
- 総曲輪 → Total track(本来は「そうがわ」という地名)
- 愛冠 → Love Crown(本来は「あいかっぷ」という地名)
- 音威子府 → Otokofu(本来は「おといねっぷ」という地名)
- 押帯 → Obi belt(本来は「おしょっぷ」という地名)
- 足羽 → Foot feather(本来は「あすわ」という地名)
- 西武門 → Seibumon(本来は「にしんじょう」という地名)
- 手登根 → Tetone(本来は「てどこん」という地名)
- 軍畑駅 → Gunbata Station(本来は「いくさばた」駅)
- 川内駅 → Kawauchi Station(本来は「せんだい」駅)

自動翻訳で誤訳が発生してしまう原因

ニューラル機械翻訳の仕組みと、特徴的な誤訳のパターンについて

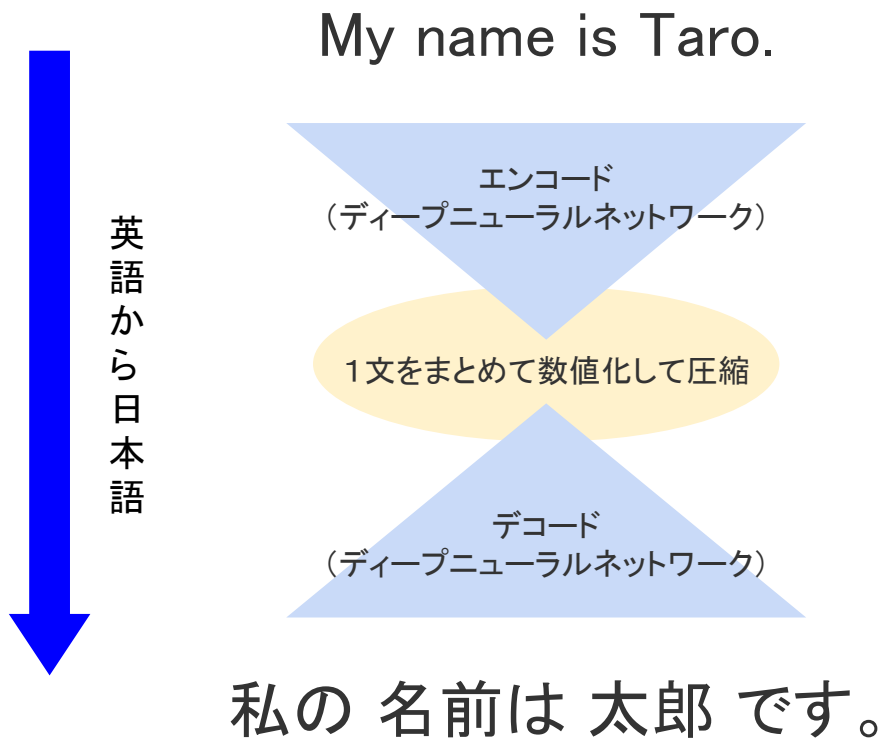
ニューラル機械翻訳の特性

Google翻訳をはじめ、近年の自動翻訳は「ニューラル機械翻訳(NMT)」という手法が採用されています。これは**人間の脳神経回路が情報伝達を行う仕組み**を取り入れ、コンピュータに翻訳を学習(※)させることで、精度の高い翻訳を実現しています。

このニューラル機械翻訳では、文全体をひとつの固まりとして翻訳するのが特徴です。例えば「My name is Taro.」という文章の場合、「My = 私の」「name = 名前」のように**ひとつひとつの単語を訳していくのではなく、いったん一文全体をまとめて数値化(ベクトル化)してしまいます**。単語それぞれの意味を対応させると自然な表現にすることが難しいのですが、あらかじめ学習した結果を生かして訳文を生成することで、**ひとつの文全体として自然で流ちょうな表現**を実現することができます。

※ニューラル機械翻訳では、たくさんの対訳コーパス(原文と訳文がセットになったデータ)を使って機械学習を行います。幅広い分野の対訳コーパスを大量に学習することによって、さまざまな翻訳を高精度で実現できるようになります。

ニューラル機械翻訳の仕組み



ニューラル機械翻訳では、入力された文全体をエンコードして数値化・圧縮し、それを基に翻訳先の言語で再構成しています。DNN(ディープニューラルネットワーク)の最適な数学的モデルは、機械学習によってコンピュータ自身がつくりだしたものです。またエンコードによって圧縮された情報は「中間言語(Interlingua)」などと呼ばれ、単語それぞれの対応情報を保持しないため、**訳漏れ(一部の情報が消失)が起こる場合があります。**

ニューラル機械翻訳の特徴的な誤訳

①訳漏れ、訳語の重複

単語ひとつひとつを逐一翻訳する仕組みではないので、翻訳の際に原文の一部が漏れてしまったり、また逆に訳語の一部を繰り返し表記してしまったりすることがあります。精度向上のためにアテンション機構(訳出すべき語順を推定する仕組み)などが備えられたりもしていますが、まだ完全にミスを防ぐには至っていません。

②文脈を考慮した意訳ができない

一文全体を自然な表現で訳出することができますが、一方で文章同士のつながりや文脈を考慮して翻訳するような仕組みにはなっていません。そのため、文脈に合わせた意訳や文脈の背景に沿った訳出、また文章全体での用語の統一などの対応が難しくなっています。

③固有名詞の誤訳

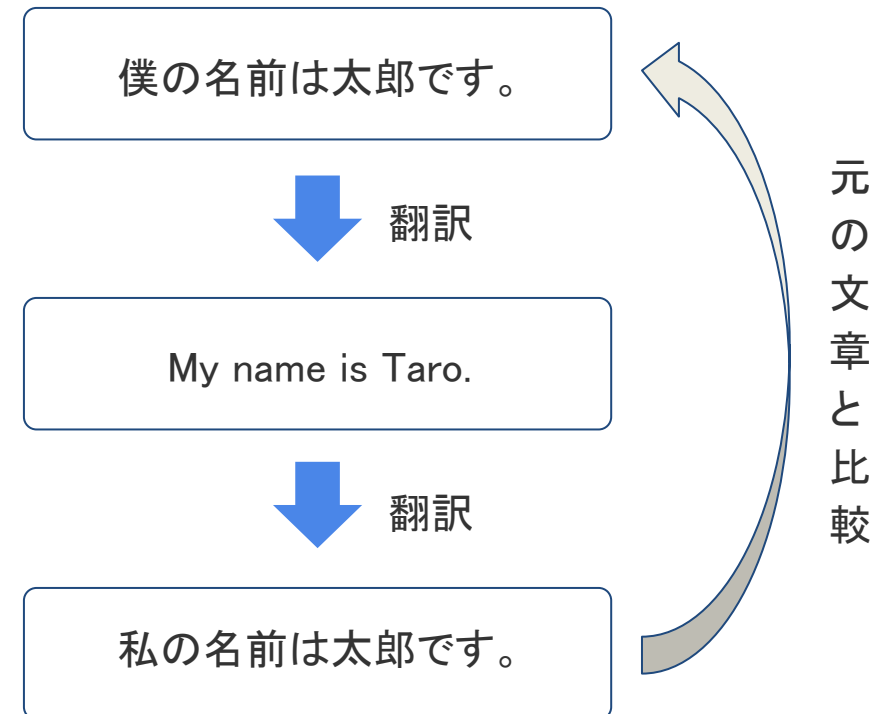
対訳コーパスの機械学習によって精度が向上しますが、全ての固有名詞を学習させるのは非常に困難です。そのため学習経験のない固有の用語は誤訳の可能性が高くなります。

誤訳のチェック方法と対策

誤訳リスクをコントロールするためのチェック方法と対策について

チェック方法①:リバーズ翻訳

自動翻訳の訳文が正しいかどうか、最も簡単にチェックできるのが「リバーズ翻訳」です。これは一度翻訳した文章を元の言語に再度翻訳し、原文との比較を行います。翻訳スキルがない場合でも、訳漏れや訳語の重複をある程度見つけることができるだけでなく、重大な誤訳の発見にもつながります。緊急避難メールのような、チェックに時間をかけられない場合にも簡易的なチェックとして有効です。



チェック方法②: 翻訳者による品質評価

原文の情報が正確に翻訳されているか、**5段階で1文ずつ**評価する方法です。プロの翻訳者が確認することで、より自動翻訳の品質を正確にチェックすることができます。訳文の正確性だけでなく、誤訳のパターン(訳漏れ、訳語の重複、不自然な表現、等)もチェックすることができるので、**どの言語で、どんな内容のときに、どんな誤訳が発生しやすいか**、全体の傾向を把握し、誤訳のリスクに対して改善策を検討することも可能です。

正確性を5段階で評価

原文の情報をどのくらい伝達できているか？



誤訳のパターン分析

誤訳の原因は、訳漏れ？ 訳語の重複？

自動翻訳の誤訳リスクを防ぐ対策

①最適な自動翻訳サービスの選定

自動翻訳のサービスごとに地名の翻訳が正確だったり、専門用語を訳出できたり、得意分野や翻訳の傾向が異なります。用途に合わせた最適な自動翻訳サービスを選定しましょう。

②用語集の作成

用途に合わせて適切な用語集を作成することで、それを訳文に反映させることも可能です。固有名詞や専門用語などの訳出に課題がある場合には、とても有効な対策です。

③原文表現の改善

長い文章や主語が不明な場合、余計な改行が入っている場合など、自動翻訳が苦手なパターンが多くみられる場合は、原文の表記ルールを変更することで改善されることがあります。

④ポストエディットの導入

どの自動翻訳も100%正しく翻訳を実行するのは不可能です。正確性が求められる文章の場合にはプロの修正を加えることで、誤訳リスクを防ぐのがオススメです。

【Xtra】翻訳品質チェックサービス

Xtra株式会社では**プロ翻訳者が訳文の正確性を評価し、誤訳のパターン分析を行**翻訳品質チェックサービスを提供しています。分析結果をもとに対策方法を検討し、貴社の翻訳リスクコントロールをサポートさせていただきます。



翻訳品質チェックサービス

プロ翻訳者が自動翻訳の品質を評価・分析します。
評価対象5,000文・50,000円～

無料でのお見積り可能です。お気軽にお問い合わせください。
Xtra株式会社・翻訳品質チェック担当 (client@x-tra.jp)

【Xtra】誤訳対策サービスの例

自動翻訳コンサルティング

弊社では自社の自動翻訳サービスも提供していますが、他社の自動翻訳サービスも含めて選定のコンサルティングを行っています。誤訳の分析結果をもとに、最適な自動翻訳の選定・運用をサポートいたします。

用語集の作成

さまざまな言語・分野の用語集は、多言語でデータ作成実績の豊富な弊社にお任せください。特に専門用語の誤訳が多い場合におすすめのソリューションです。

【Xtra】誤訳対策サービスの例

プリエディット(原文の編集)

原文を修正することで、訳文の品質が大きく改善する場合があります。主語を明確にしたり、**長文を分割・整理したり、慣用的な表現を避けたり**等の編集を、各種言語で対応します。

ポストエディット(訳文の修正)

自動翻訳の誤訳が発生しやすい分野・形式の傾向がわかれば、そこだけプロが訳文を修正するという対策が可能です。**外部の人が目にするwebや資料など、正確性が求められる場合**におすすめのソリューションです。

【参考】内容の伝達レベルに関する評価基準

以下のような5段階の基準で、1文ずつ評価を実施します。

評価 5	すべての重要情報が正確に伝達されている。(100%)
評価 4	ほとんどの重要情報は正確に伝達されている。(80%~)
評価 3	半分以上の重要情報は正確に伝達されている。(50%~)
評価 2	いくつかの重要情報は正確に伝達されている。(20%~) もしくは情報を大きく誤認させる可能性がある内容である。
評価 1	正確に伝達されている重要情報はほとんどない。(~20%) もしくは情報を著しく誤認させる内容である。

【参考】誤訳パターンの分類

以下のような6つのカテゴリで、どのパターンに当てはまるかを1文ずつ分類します。

訳漏れ	原文の情報の一部が漏れてしまい、情報が不足しているもの
過剰翻訳	訳語が重複していたり、原文にはない情報が付加されていたりするもの
不自然な表現	訳文の情報に誤りはないが、表現や言い回しが不自然なもの
固有名詞の誤訳	地名や人名、企業名、商品名などの固有名詞に誤りがあるもの
専門用語の誤訳	業界や分野特有の専門用語に誤りがあるもの
その他の誤訳	その他、訳文の内容になんらかの誤りがあるもの

ニューラル機械翻訳は劇的に精度が向上しているため、近年ではプロの翻訳者に匹敵するような高品質のサービスも生まれています。しかし一方で、その仕組みの特性から起こり得る**誤訳のリスク**も無視できません。

大切なのは、**どのような誤訳がどのくらい発生しているかを把握すること**です。リスクを把握することができれば、それらをコントロールすることができます。

貴社がどのように自動翻訳を活用し、翻訳リスクをどうコントロールしていくかを検討する上で、本資料がそのヒントになりましたら幸いです。もしご不明な点などございましたら、お気軽にお問い合わせ・ご相談ください。

Xtra株式会社

101-0047 東京都千代田区内神田2-15-2 内神田DNKビル4F

URL: <https://x-tra.jp/> Mail: client@x-tra.jp (翻訳品質チェック担当)