

# 機械翻訳(MT)は字幕翻訳できるのか

## YouTube 字幕の記述および字幕におけるリエディットの有効性の検証

平岡裕資 山田優

関西大学

### 1 はじめに

ニューラル機械翻訳(Neural Machine Translation: NMT)が実用化し、これまでの機械翻訳に比べて飛躍的に翻訳品質が向上したと言われ、その実用化が進んでいる。この背景には、世界的な翻訳者不足が挙げられる。例えば、世界中のあらゆる言語をペアによる翻訳需要を満たすにはおよそ20億人の翻訳者が必要であるとの報告もある[1]。このような翻訳需要の増加に対する問題を機械翻訳の長所であるコストとデリバリーの短縮によって打開できるかもしれない。

機械翻訳の実践方法で代表的なものはポストエディットであるが、最近ではリエディット(Pre-edit)も注目を集めている。しかし、研究が立ち遅れるリエディットは、その費用対効果や効率性の調査は不明で、かつ具体的な書き換え手法もはっきりと定義されていない[2]。それでも、リエディットが注目されるのは、その書き換えルールに基づく手法においては単一言語話者でも機械翻訳を使用して翻訳が可能であるというポテンシャルにあるのかもしれない。

このような技術的発展が翻訳に与える影響は、間接的にも存在する。例えば、インターネット上の情報がマルチモーダル化しつつあり、映像や音声情報を含む媒体が新たな翻訳の対象となっている。代表的なものに YouTube があり、これはかつて消費者であったユーザー側がコンテンツを生成できるようになったという意味において、ユーザー生成型コンテンツと言える。このような流れは、映像翻訳、特に安価な字幕翻訳の需要が増加することを意味する。この字幕翻訳は翻訳研究において度々対象とされるが、その多くは映画字幕である。この種の翻訳には、読みやすさを考慮した文字数制限などのいくつかの制約があり、この制約を回避するために不要な内容を削除するなど、字幕翻訳に特有の規範が存在する。しかし、この規範がテクノロジーの発展によって誕生したユーザー生成型翻訳はこの規範を変えつつある[3][4]。先述のように、テクノロジーの発展によって YouTube などのユーザー生成型コンテンツが生まれたが、これは翻訳においても同様で、これまで消費者であったユーザー側が自律的で自由な翻訳をし、それが発信可能となったのである。これによって、これまで可視化されていなかった受容者側の期待がユーザー生成型翻訳によ

って顕在化された。

以上のような状況を踏まえ、本研究の目的は、今後さらなる需要拡大が見込まれる、ユーザー生成型のマルチモーダルコンテンツである YouTuber の映像に字幕翻訳を付与することを、機械翻訳が行えるのかを検証する。より具体的には、日本国内で人気のある YouTuber のコンテンツ(日本語)を英語に翻訳をしてグローバル配信することを想定し、それを機械翻訳とリエディットによって、どこまで達成できるのかを調査する。

### 2 リサーチクエスチョン

本研究では「機械翻訳は字幕翻訳ができるのか」というリサーチクエスチョンを設定し、YouTube における字幕を目標として、そしてリエディットの有効性を検証する。YouTube 字幕を比較の対象とする理由は、上述したとおり、将来の字幕翻訳規範となりうると思われるからである。そのため、リエディットを目指すに当たって YouTube の字幕がどのような特徴を持つのかを記述することは必要である。

まず、YouTube 字幕の記述に関しては、形式的な特徴と訳出方針の観点から分析を行う。具体的にはアニメファンサブの特徴[5]と比較し、その差異があるのかを見る。次に、リエディットの検証では、YouTube 字幕と自動翻訳の品質評価を行い、その差異を見ること、そして評価の低い自動翻訳文に対してアドホックなリエディットを行う。その後、再度品質評価を行い、リエディットの前後でその差異があるかどうか、人間翻訳(HT)にどこまで近づけるかを検証する。

### 3 検証デザイン

翻訳言語方向は日本語から英語である。対象とするテキストは、YouTube のユーザーが YouTuber による日本語の映像を対象に翻訳した英語字幕と、日本語の書き起こしを原文とする自動翻訳文の2つである。MT システムはグーグル翻訳を使用した。

比較の基準として、以下の二点を挙げた。

- 文字数制限
- 正確性

文字数制限は字幕翻訳のわかりやすい特徴であるが、

これは字幕の表示時間内に読み切れるかの視聴者側の読みやすさを考慮したうえの制約である。英語の文字数制限は一般的に1秒あたり12文字(レター)であるが、DVD やインターネットなどの配信では17文字まで許容されつつある。今回はこの2つの基準をもとに、YouTube 字幕がどの程度の文字数制限に収まるかを見る。

正確性は機械翻訳の品質評価で一般的に用いられる5段階のものを用いた[2]。しかし、最も評価の高いものに非母語話者には判断しにくいものがあつたことや、曖昧性を抑えるため、これを3段階に簡易化させた(以下の表を参照)。

3	Perfect	原文の情報が完全に翻訳されている。訳文は文法誤りを含まない。母語話者から見ても、語句の選択が自然である。
	Good	語句の選択がやや不自然であるが、原文の情報が完全に翻訳され、訳文は文法誤りを含まない。
2	Fair	原文のあまり重要でない情報の翻訳に些細な誤りがあるが、原文の内容は容易に理解できる。
	Acceptable	重要な情報の一部が欠落したり、誤訳されたりしているが、原文の中核となる情報は何とか理解できる。
1	Nonsense	原文の意味が全く理解できない。

表 1: 正確性の評価基準

プリエディットはバイリンガルによるアドホックな(試行錯誤的な)手法で行った。作業者は編集者の要件[2]を満たしていると仮定し、以下のサイクルでプリエディットを行った。

**Step 1.** 評価1に分類された字幕において、品質を低下させていると想定される原文の要素に対して書き換えを行い、**Step 2** に進む。

**Step 2.** 書き換え後の原文を自動翻訳する。**Step 3** に進む。

**Step 3.** 書き換え後の自動翻訳が評価2以上に達したと判断される場合は書き換えを終了する。達していない場合は **Step. 1** に戻る。

このサイクルを繰り返し行い、これ以上の品質向上は見られないと判断できれば、作業を終了する。

## 4 検証結果

### 4.1. YouTube における字幕

YouTuber の発話に関してはそれぞれの特徴はありつつも、一貫して即興に近い発話であった。これには感嘆詞などのフィラーが多く見られたことや動画の更新が頻繁にあることなどから仮定できる。また、YouTuber 同士の会話はありながらも、視聴者を意識した語りが伺えることも注目

したい。

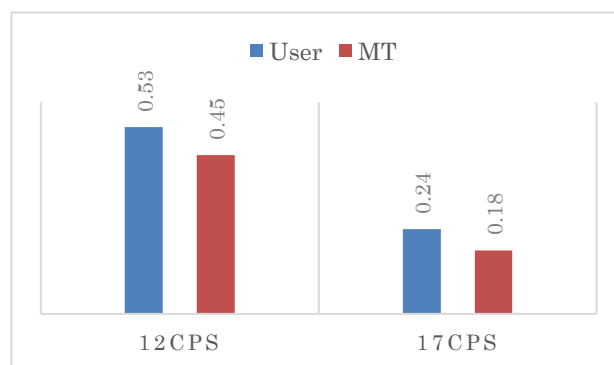
ユーザーによる英語字幕はアニメファンサブと類似するものであつた。アニメファンサブの大きな特徴として、翻訳者による注釈が字幕として表示されることや、異文化的要素を残すような原文寄りの訳出であること(つまり映画字幕などに見られる情報の削除がない)などが挙げられる[5][6]。これらの特徴は YouTube 字幕でも同様に観察された。異文化要素の訳出に関しても、英語字幕のための翻訳方略の分類をもとに分析を行った結果、起点志向の方略がほとんどであった(15 本中14本)。これらのことから、アニメファンサブと YouTube 字幕は類似していると仮定できる。通常映画字幕と類似したフローで作成されるアニメファンサブと異なり、YouTube 字幕はクラウドソーシング翻訳であり、不特定多数のユーザーが翻訳・編集・審査できる。そのため、訳出のプロセスが大きく異なるが、類似した訳文となっていることは、観察された字幕が一般的にもユーザーの期待に沿うものであつたと言えるかもしれない。

### 4.2. 文字数制限

ユーザーによる字幕と自動翻訳をそれぞれ 12 文字と17文字にどの程度収まっているかを調査した。結果は以下の表に示す。

	12 cps	17 cps
User	495 (0.53)	225 (0.24)
MT	348 (0.45)	144 (0.18)

表 2: 文字数制限を超過する字幕



ユーザーによる字幕と自動翻訳の文字数の平均に有意差は見られなかったため(Wilcoxon rank,  $W = 381290$ ,  $p\text{-value} = 0.1387$ ), 全体的に同程度の文字数であった。また、文字数制限から見ても、同程度の超過率であった。これまでの制限の基準であった 12 文字は両者とも半数の超過率であり、従来の規範に従った訳出であるとはいえない。このことから、YouTube のユーザ字幕はアニメファン

サブと同様に起点志向であり、そのため自然と文字数制限も超過することがわかる。一方、17文字に関しては、両者がおおよそ2割の超過であった。この超過率は映画字幕においても一般的であることから、ユーザー字幕と自動翻訳のどちらも17文字の新しい規範に収まっていると言える。

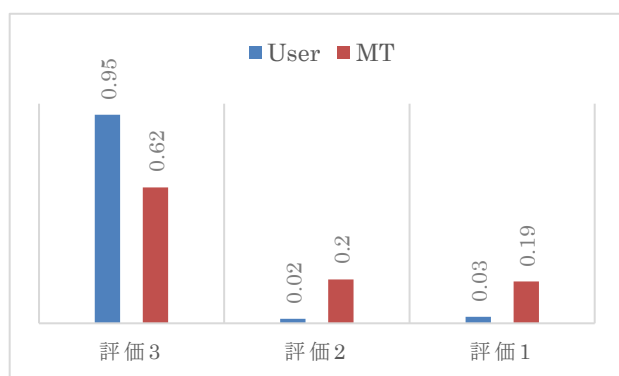
これに加えて、17文字を2倍以上超過する字幕を分析した。これはユーザー字幕においては45本、自動翻訳では7本の字幕で見られた。しかし、その大半がスポッティング(字幕の表示時間を設定する作業)のミスや、それぞれの発話に対する字幕にまたがってテロップの訳出をしている場合であった。つまり、2倍超過する字幕のほとんどが、YouTube字幕の機能に制限されたこと(一度に複数の字幕を表示できない)に起因する。むしろ、ユーザーについて言えば、テロップをある程度長く表示されるなど、「読みやすさ」に配慮していることが伺える。自動翻訳においても、これらの問題は書き起こし字幕に起因し、この書き起こしも同様に機能上の制限にしばられている。

### 4.3. 正確性

前述の評価基準を用いて、ユーザー字幕と自動翻訳の正確性を評価した。結果は以下の通りである。

	評価3	評価2	評価1
User	891 (0.95)	22 (0.02)	27 (0.03)
MT	468 (0.62)	149 (0.20)	143 (0.19)

表3: 正確性の評価結果



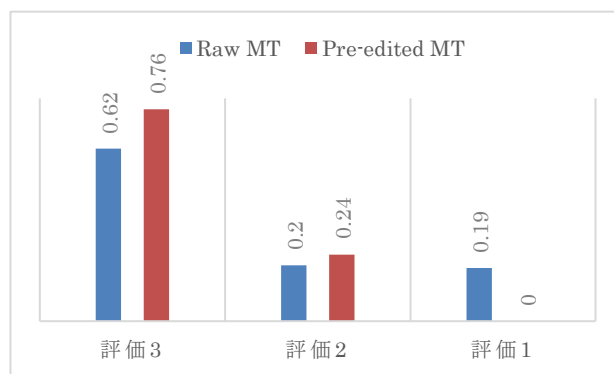
やはり、HTであるユーザー字幕よりも自動翻訳は正確性の面で劣ることがわかった。つまり、ユーザー生成型の字幕においても機械翻訳+プリエディットの必要性が示唆される。ユーザー字幕においては、評価1に分類される字幕がサンプルによって偏りがあった。これは、原文の映像における発話が理解しづらいものであったことが理由として考えられる(実際に分類された字幕は原文を理解できていないことが伺えるものであった)。

### 4.4. プリエディット後の正確性

正確性の評価で最も低い評価1に分類された自動翻訳143本を対象にプリエディットを行なった。プリエディット後の評価結果は以下の通りである。それぞれの評価を3から1の点数に置き換え、合計の平均を見た結果、Raw MTとプリエディット後のMT結果の両者に有意差があった(Wilcoxon rank,  $W = 227370$ ,  $p\text{-value} = 4.845e-15$ )。また、評価1の割合が人間翻訳であるユーザー字幕に近いものとなった。

	評価3	評価2	評価1
Raw MT	468 (0.62)	149 (0.20)	143 (0.19)
Pre-edited MT	563 (0.76)	175 (0.24)	1 (0.00)

表4: プリエディット後の正確性



書き換えの目標は評価2に設定したが、書き換え後に評価3に分類できるものも多くあった。結果的に、ほとんどの字幕をプリエディットで評価2以上に向上することができた。書き換えのステップ数は最多でも10回ほどであったため、先行研究[2]と比較して容易に書き換えが可能であった。

## 5 ここまでのまとめ

これらの結果から、ユーザー字幕と自動翻訳は文字数制限の観点からは同程度の品質である一方、やはり正確性の面では自動翻訳が劣る結果となった。そのため、正確性の改善を求めて、評価の低い字幕に対してアドホックなプリエディットを行なった結果、そのほとんどに正確性の向上が見られた。

以下では、これからの字幕翻訳と、字幕におけるプリエディットの有効性について考察を行う。

## 6 考察

### 6.1. これから求められる字幕

文字数と正確性の観点から YouTube 字幕を分析した結果、正確性は個々の映像においてばらつきはありつつも、全体的には視聴に十分な正確性を持つ字幕であったと言える。また、文字数は 17 文字に収まる字幕が大半であった。このことから、YouTube 字幕においても文字数の規範が意識されていると言えるかもしれない。しかし、ユーチューブの翻訳形態がクラウドソーシング型であり、不特定多数の視聴者が編集できることを考慮すると、その参加者の多くが文字数制限の規範を持っていたとは考えにくい。このことから、文字数に特別な考慮をしなくても訳文は 1 秒 17 文字に自然と収まると予想できるかもしれない。

同時に、映像を視聴する環境も変わりつつある。映画以外の映像、例えば DVD やブルーレイ、インターネット上で視聴できる映像は何度でも巻き戻しができることから、今後は、そもそも文字数制限を厳密には考慮する必要もなくなると予想できる。つまり、ファンサブのように、文字数制限による情報の取捨選択がなされず、特定の視聴者層に絞った訳出方法が字幕翻訳の主流になっていくだろう。

### 6.2. 書き換え対象として字幕

前述のように、字幕翻訳の規範である文字数制限から見ても、従来のものから大きく外れるような特徴がアニメファンサブだけでなく、YouTube 字幕でも見られた。このことから、一般的にも、台詞と字幕の不一致があるような訳出よりも、原文に忠実な訳出が視聴者から好まれ、それが新たな規範になりつつあることが伺える。この流れは、基本的に忠実な訳出のみが可能な機械翻訳に好ましいものといえる。しかし、文字数制限が許容されたといっても、YouTube 字幕の記述から見られる通り、視聴者もある程度の読みやすさを意識していることや、直感的にも、1 秒 17 文字を大幅に超える字幕ばかりでは、繰り返し再生可能であるとはいえ、視聴に影響が出ることが考えられる。これらのことから、プリエディットの目標として、正確性を向上させることに加え、最低限の文字数制限(例えば 17 文字を 2 倍以上超過しないなど)、意識する必要があるだろう。

### 6.3. プリエディット手法の有効性

正確性評価の結果の通り、ユーザー生成型の字幕翻訳においても、機械翻訳は人間翻訳よりも劣る結果となった。ゆえに、ユーザー生成型かつボランティア翻訳の字幕においても、人間翻訳に近い品質を目指すのであれば、人間の介入がまだ必要であるといえるだろう。バイリンガルプリエディット作業を行ったところ、ある程度のプリエディット能力がある作業者がアドホックに書き換えを行えば、ユーザ

一字幕、つまり人間翻訳に近い品質に達することが検証された。その書き換え例は語順変更や主語の追加・削除など、先行研究のもの(異なる MT システムと文書ドメインにおけるプリエディット研究)と類似するものがあった。これらは、MT システム間、文書ドメイン間で共通した普遍的書き換えであると予想できる。今後はこれらの書き換えに焦点を当て、詳細に記述する必要があるだろう。

## 参考文献

- [1] Common Sense Advisory. (2018). Machine Translation for Human Innovation. Retrieved on November 17, 2018: [http://www.common senseadvisory.com/machine\\_translation.aspx](http://www.common senseadvisory.com/machine_translation.aspx).
- [2] 宮田玲, 藤田篤. (2017). 機械翻訳向けプリエディットの有効性と多様性の調査. *通訳翻訳研究への招待*, Vol. 18, pp. 53-72.
- [3] 篠原有子. (2013). 映画字幕は視聴者の期待にどう応えるか. *通訳翻訳研究*, Vol 12, pp. 209-228.
- [4] 篠原有子. (2016). 日本映画の英語字幕における標準化 制作プロセスの観点から. Retrieved on June 4, 2018: [https://rikkyo.repo.nii.ac.jp/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=15013&item\\_no=1&page\\_id=13&block\\_id=49](https://rikkyo.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=15013&item_no=1&page_id=13&block_id=49)
- [5] Perez Gonzalez, L. (2007). Intervention in new amateur subtitling cultures: a multimodal account. *Linguistica Anterpenensia*, Vol 6, pp. 67-79.
- [6] Cintas, J.D. & Sánchez P.M. (2006). Fansubs: Audiovisual Translation in an Amateur Environment. *The Journal of Specialised Translation*, Vol 6, pp. 37-52.