動画字幕の翻訳および制作における効率化の検証

森口 功造　　伊澤 力
株式会社川村インターナショナル
kzmgh@k-intl.co.jp, rkizw@k-intl.co.jp

中岩 浩巳　　薄井 智貴
名古屋大学 大学院
nakaiwa@i.nagoya-u.ac.jp, tomo.usui@nagoya-u.jp

1 はじめに

YouTube をはじめとした動画共有サイトの台頭により、インターネット上での情報発信方法として動画コンテンツが勢いを得ている。ある動画広告の市場調査によると、2017 年の動画広告市場は前年比 163%で成長しており、米シスコシステムズは、「2016 ～ 2021 年の間にインターネット ビデオ トラフィックは約 4 倍に増加する」[1]と予測している。

同時に、コンテンツを外国語に翻訳して動画コンテンツを海外に発信したり、海外の動画コンテンツを日本語化して国内向けに発信するといったニーズも高まっている。

ただし、増え続ける動画コンテンツとその翻訳需要に対して、実現するための予算は必ずしも比例するわけではない。従来の動画翻訳や字幕制作の工程には、文字起こし（聴き起こし）や人手翻訳、さらに動画への字幕挿入といったマニュアル作業が多く含まれているため、増加する動画コンテンツを従来方式の作業で処理することは難しくなっている。

本稿では、2018 年に国立大学法人 名古屋大学で実施された講義の収録動画（英語）に日本語字幕を付けるプロジェクト（以下、本件。）において、効率化を目的に実施したプロセス改善の結果を検証した。

本件は、従来のプロセスによる作業ではなく、一連の工程においてさまざまな自動化・効率化手法を用いることで、可能な限り作業の効率化を図ることを試験的に導入して作業を実施した。それら効率化の概要および課題について説明する。

2 プロジェクト概要

本プロジェクトの概要は以下のとおり。

- 対象講義（動画）数： 48 本
- 対象動画再生時間： 54 時間
- 翻訳言語： 英語から日本語
- 翻訳ワード数（英語原文）： 約 256,000 ワード

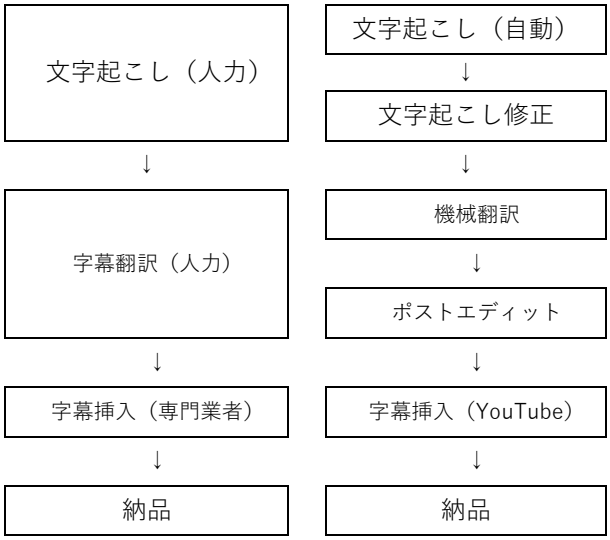
講義動画の特徴としては、講師が英語の母語話者である場合と、そうでない場合が混在している点、動画の撮影環境が様々で画質や音質にバラツキがある点、そして講義の分野が単一ではなく多岐にわたる点が挙げられる。

3 作業工程

本件では、従来の動画字幕制作および翻訳における一連の工程で、ツールやオンラインサービスを活用して自動化・効率化が可能な部分を洗い出し、可能な限りそれらを活用することを目指した。また、「日本人の受講者が講義内容を理解できること」を品質目標とした。細部の正確性やこなれた日本語表現、読みやすい日本語や字幕表示といった点は、従来の工程による成果物と比較した際に、低下するリスクがあることを考慮しつつ自動化・効率化を優先した。

自動化・効率化のアプローチは、文字起こし、翻訳、そして字幕挿入の工程において適用することとした。本件で採用した新プロセスを従来のプロセスと比較した工程表を図 1 に示す。

図 1. 従来プロセスと新プロセスの作業フロー比較



3.1 文字起こし

従来の文字起こし作業は、作業者が動画を視聴し、聞き取った音声をタイピングすることで音声をテキスト化する作業である。人間が発話を聞き取ることで高い正確性を得ることができるが、すべて人手による作業のため費用と時間

がかかってしまう。

昨今の音声認識技術の向上により、特に英語においては人間と同程度の正確性で聞き取れると言われる音声認識システムも登場している。[2]本件では、動画共有サービス YouTube の自動文字起こし機能を利用し、音声を自動書き起こしした。その結果、①YouTube の自動文字起こしは言語の検知を自動で行うため、動画によっては誤った言語を選択してしまうことがあった他、②事情ははっきりとしないが自動文字起こし自体ができないないケースもあった。こうした場合には、動画から音声のみを抽出し、その音声データを IBM 社の WATSON を用いて文字起こしして代用した（https://www.ibm.com/watson/services/speech-to-text/）。

この結果、自動書き起こしにより音声をテキスト化することに成功したが、その精度は依然として完璧ではなく、そのまま翻訳することはできない品質であったため、自動書き起こし後に人手による確認・修正作業を行った。

YouTube では、動画の投稿者が同サイト上で動画を編集できる「クリエイターツール」が提供されており、字幕編集の機能も提供されている。本件では、クリエイターツールの字幕編集機能を用いて動画字幕を修正した。

自動文字起こし後の修正では、文法的に不要な語の削除（下記例の下線①）、誤認識の修正（下記例の下線②）、キャピタライゼーションの修正（下記例の下線③）を中心に修正を行った。修正の例を以下のサンプルに示す。

【自動文字起こしテキスト】

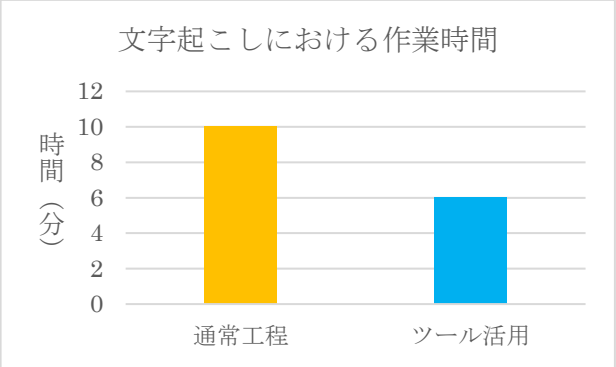
the this① robot will provide you provide you the secure and efficient way of meta vibration② because so the neither one may disclose the secure data during the update

【修正後】

This③ will provide you the secure and efficient way of data manipulation, because the naive one may disclose the secure data during the update

一般的な基準として、文字起こしの一般的な生産性は動画再生時間 1 分あたり 10 分と言われている。本件で、自動書き起こしを修正する作業を行った結果、動画再生時間 1 分あたりの作業時間は平均 6 分となり、約 40%作業時間を削減することができた。

YouTube でテキスト化した字幕は、タイムコード付きの字幕ファイル（.sbv、.srt）としてダウンロードすることが可能である。そのため、YouTube で認識・修正された字幕テキストをダウンロードし、外部のソフトウェアを用いて翻訳することができる。

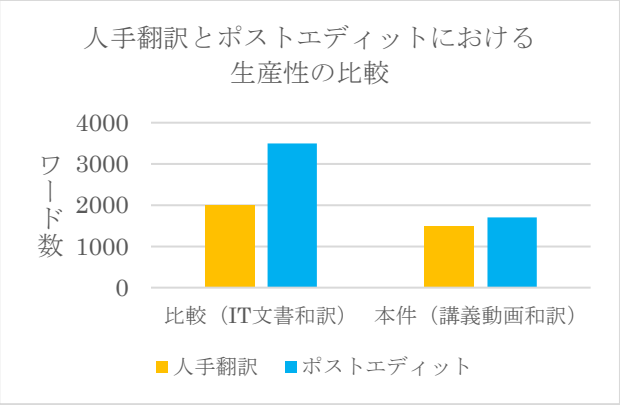


3.2 翻訳（ポストエディット）

文字起こしと同様に、翻訳作業も大部分が手作業である。本件では自動化ツールとして機械翻訳を活用し、その翻訳結果を人手で編集する「ポストエディット」を行うことで、翻訳作業の効率化を図った。機械翻訳エンジンは汎用性および翻訳品質の観点から総合的に判断し、Google 翻訳を採用した。

通常、当社内における同種プロジェクトで適用される人手翻訳の生産性は、英語から日本語への翻訳を行う際には 1 日（8 時間）あたり 1,500 ワードを見込むが、本件においてポストエディットを行った結果、1 日あたりの生産性の平均値は 1,700 ワードであった。

当社別プロジェクトにおいて、IT 関連の技術文書を翻訳する際は、人手翻訳の場合には 2,000 ワードを見込むが、ポストエディットでは 3,500 ワードを平均的に処理している。それと比較した場合、本件における生産性の向上幅は小さかった。



生産性の向上度合いが上述の IT 文書の翻訳と比較して低かった理由は、①本件は技術文書のような文章ではなく話し言葉を主とする口語調のテキストが多かったこと、②高度な学術的内容を含むものであったため、ポストエディターが必ずしも担当する文書のトピックについて適切なバックグラウンドを持っていなかったこと、も原因と考えられる。

3.2 字幕挿入

字幕テキストの翻訳が完了した後、動画に字幕を挿入し、適切なタイミングで表示させるよう調整する作業が必要となる。従来の字幕挿入作業は、専用の動画編集ソフトウェアとエンジニアを持つ業者などに作業を依頼する必要があった。その場合、プロによる作業のため仕上がり品質は概して高いが、その反面高額かつ時間もかかる。本件では、この工程を動画編集に関して特別な訓練を受けたことのない担当が YouTube 上で行うにより、安価かつスピーディーに作業を行うことを目指した。

作業は、文字起こしと同様に YouTube が無料で提供する「クリエイターツール」の字幕編集機能を活用した。プロ向けの動画編集ソフトと比較して機能は限定的であるものの、動画編集用のワークステーションといった特殊なハードウェアを必要とせず、動画編集未経験の作業者が一般的なオフィス業務で使用するノート PC で一連の作業を行うことができた。

その結果、通常当社にて同様の作業を専門の業者に委託する場合の費用と当社の作業者が稼働した時間分の人件費を比較したところ、約 2 割の費用で字幕挿入作業を行うことができた。

4 課題

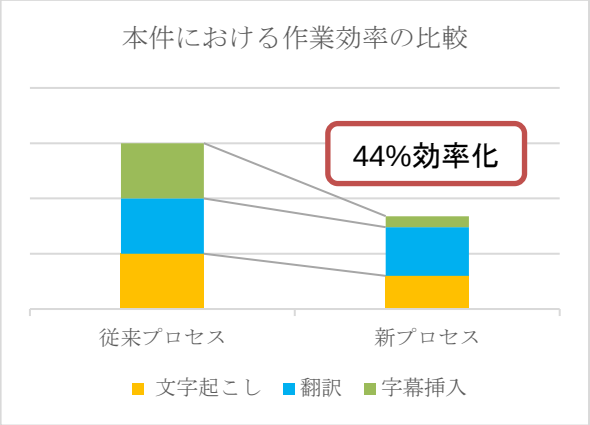
今回効率化を行ったいずれの工程においても、自動化ツールやオンラインサービスを使用したことにより従来のプロセスと比較して効率化を達成することができた。ただし、前述のとおり本件で対応した講義の内容は話し言葉や専門的な内容を多分に含んでおり、本件で使用した汎用的な音声自動認識や機械翻訳エンジンではそれらの点において必ずしも最適ではなく、結果として予想された生産性の向上が得られなかった可能性がある。

今後の研究テーマとしては日語に特化した自動認識システムや講義内容の分野に特化したドメイン特化型の機械翻訳エンジンを使用することで生産性がどのように変化するかを検証するといったことが考えられる。

5 おわりに

本件では、字幕の文字起こし、翻訳、そして字幕挿入の工程において、従来の作業を効率化すべく自動化ツールやオンラインサービスを積極的に活用し、その結果全工程に置いて程度の大小はあるものの、従来のプロセスと比較して効率化を達成することができた。

各工程の効率化を総合すると、下図に示すとおり全体として 44%の効率化を実現することに成功した。



文字起こしでは、従来の人手による作業を YouTube の自動文字起こし機能を活用することによって効率化し、翻訳は機械翻訳を用いることで従来の人手による作業を効率化した。いずれの効率化も、いわゆる AI の活用により可能となったものであるが、その精度は人手の作業をすべて置き換えるものではなく、人手による確認・修正を要することも確認できた。

字幕挿入の作業に関しては、いわゆるクラウドサービスを活用することにより、作業者の側に専用のハードウェアを用意することなく、特別な知識や経験の無い者でも必要十分な成果物を得ることができた。

ただし、課題の点において論じられたとおり、AI 等のツールを活用したことによる一定の効率化はみられたものの、本件のもつ専門性の部分においては今回使用した YouTube や Google 翻訳といった汎用的サービスでは十分に対応できていない可能性があった。この点に関してはより最適化された AI による自動処理によって人手作業をさらに軽減することができるのかという興味深い研究課題を得ることができた。

本件で行った工程はすべての動画翻訳に利用できるものではないが、従来のプロセスと比較してコスト低減と納期短縮を実現することが可能なため、増え続けるニーズに対応し、今までに無く早く消費されていくコンテンツのライフサイクルに対応するための新たな動画翻訳の手法として検討する価値はあると考えられる。

参考文献

[1]. Cisco Visual Networking Index：予測と方法論、2016 ～ 2021 年 ホワイト ペーパー 2017 年 6 月 6 日（https://www.cisco.com/c/ja_jp/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html）

[2]. マイクロソフトブログ 2016 年 10 月 18 日（Historic Achievement: Microsoft researchers reach human parity in conversational speech recognition）(https://blogs.microsoft.com/ai/historic-achievement-microsoft-researchers-reach-human-parity-conversational-speech-recognition/)