『中国語動画音声コーパス』の構築：複数モダリティによる正確な書き起こしを目指して

『中国語動画音声コーパス』の構築について
― 複数モダリティによる正確な書き起こしを目指して ―

篠崎 秀紀（大阪大学大学院人文学研究科）
于 拙（大阪大学大学院人文学研究科）
陳 宇鍇（大阪大学大学院人文学研究科）

Construction of a Chinese Audio-Visual Corpus
Aiming for Accurate Transcription through Multiple Modalities

SHINOZAKI Hidenori（Graduate School of Humanities, Osaka University）
YO Cjyet（Graduate School of Humanities, Osaka University）
CHEN Yukai（Graduate School of Humanities, Osaka University）

要旨

近年、インターネットの普及に伴い、母語話者の言語活動を観察できる機会が増加している。動画共有サイトにアップロードされている中国語の動画は、字幕が画像データとして動画のフレーム内に埋め込まれていることが一般的である。中国語コーパスの作成に際し、より広範なテキストの収集を可能にするためには、動画に対し文字認識あるいは音声認識の手法を用いる必要がある。本研究では、埋め込み字幕に対するOCR、音声に対する音声認識、動画制作者が用意した字幕など、複数のリソースから得られる、テキストを同時に表示・検索できるようなアプリケーションを実装する。また試験的にいくつかのジャンルを試行的に収集した。研究では、言語データの自動更新と動画の正確な書き起こしを実現できる中国語動画音声コーパスの構築を目指す。

Abstract

Recently, there have been increasing chance to observe native speech, thanks to the spread of Internet. Chinese videos uploaded to video-sharing platforms typically have subtitles embedded as rasterized graphical data in the video frames. To create a more comprehensive Chinese text corpus, it is essential to apply text or speech recognition methods to these videos. This study implements an application that can simultaneously display and search text obtained from multiple resources, including OCR for embedded subtitles, speech recognition for audio, and subtitles prepared by video creators. Furthermore, we conducted a preliminary experiment where we collected samples from several genres. We are planning to achieve both data auto-update and video transcription as a Chinese Audio-Visual Corpus.

1. はじめに

近年インターネットの普及により、母語話者の言語活動を観察できる機会が飛躍的に増加している。特に動画共有サイトでは、実際に母語話者が話している様子を視聴することができ、日常的な語彙や表現を豊富に含んでいる点で言語研究にとって非常に有益な資源となっている。この背景の下では、動画投稿サイトの中から画像と音声データを収集し、動画音声コーパスを構築することができるようになった。

ところが、ユーザの生成した新しい動画内容（User Generated Content, UGC）が随時にアップロードされている時代では、従来に構築したコーパスの多くは時間が経つと、収録した言

語データが古くなったり、新しい言葉遣いの用例をカバーできなくなったりする問題が存在している。そこで、収録した言語データが自動的に更新できる動画音声コーパスの構築に取り組む必要がある。

一方で、動画共有サイトに付随する字幕データの有無は動画投稿者など人手での作業に依存しており、動画音声コーパスの構築に当たって常に用意された字幕データが得られるとは限らない。したがって、字幕のない動画データも動画音声コーパスに入れるように、より正確な書き起こしの実現にはまだ工夫する必要がある。

このような背景を踏まえて、本稿では、言語データの自動的な増加、および動画における言葉の正確な書き起こしが実現できる中国語動画音声コーパスを構築する。

2. 先行研究

2.1 動画音声コーパスについて

本稿では、マルチモーダルコーパスの一種として動画上の視覚と聴覚情報を取り入れたコーパスを動画音声コーパスと呼んでいる。

人々が普段コミュニケーションをするときに、音声の聴覚情報だけでなく仕草、表情などの視覚情報も情報伝達の一環と言える。このような聴覚、視覚、ないしは触覚、嗅覚など異なるモダリティ（すなわち、異なる感覚方式）上の情報を取り入れたコーパスがマルチモーダルコーパスであると考えられている(Allwood, 2010)。マルチモーダルコーパスの中で、動画上の視覚情報と聴覚情報を収集したコーパスは本稿の中で動画音声コーパスと呼んでいる。特に、YouGlish[1]のように、動画共有サイトに掲載された映像と音声を収集する動画音声コーパスがこれに該当する。

また、教育分野に限らず、言語研究の分野、即ちコーパス言語学においても、マルチメディア・マルチモーダルなコーパスが注目されるようになってきている。例えば、日本のテレビ番組を対象としたコーパスの構築及びそれに利用した研究(孫, 2009; 孫 & 石井, 2016; 石井, 2009; 石井 & 孫, 2013) があり、また中国語の映像作品を対象としたコーパスの構築と公開(中文视听, 2021; 王, 2024) などもある。

さらに、手話言語学の分野においても、手話が視覚言語であり、音声言語のような聴覚言語と異なり、視覚モダリティを主に用いる言語であるため、手話のコーパスでも動画重要である。(姚 ほか, 2019)。

2.2 既存の動画音声コーパス

動画音声コーパスには、母語話者が実生活の中で話す言葉の実態を提供できるのが共通の特徴である。すでに公開されている動画音声コーパスの例には YouGlish (Topal, 2023)、VoiceTube[2] (Jin, 2017)、CVC（中国語視聴コーパス）[3] (中文视听, 2021) などが存在している。

2.3 YouGlish

YouGlish は、YouTube から手話を含めた 21 言語の動画を 100 万以上収集し、動画およびリアルタイムの字幕を youglish.com に載せた動画音声コーパスである。ユーザが YouGlish の検

[1]YouGlish の公式ウェブサイト：https://youglish.com（2024 年 8 月 15 日最終アクセス）
[2]VoiceTube の公式ウェブサイト：https://jp.voicetube.com（2024 年 8 月 15 日最終アクセス）
[3]CVC の公式ウェブサイト：https://client.chinafocus.net.cn（2024 年 8 月 15 日最終アクセス）

索欄に特定のキーワードを入力すれば、入力されたキーワードと関連する動画およびリアルタイムの字幕が表示される仕組みである。

このコーパスは、母語話者が実際のコンテキストでどのように話しているのかという情報を提供することを目的に開発されたものである[4]。YouGlish の特徴には、①大規模データに基づいたデータドリブン言語学習支援、②母語話者が実生活で使用する言葉と発音の展示、③言葉遣いの異なるコンテキストの包括という三つのポイントが強みと見なされる一方で、①断片的なイントネーションの提示、②母語話者と非母語話者動画の無区別、③上級学習者と自主学習者だけに向いている傾向が指摘されている(Topal, 2023)。実用面では、YouGlish を使用し、英語スピーキング能力向上を支援することには効果が証明されている(Fu & Yang, 2019; Kartal & Korucu-Kis, 2020)。また、YouTube 上で英語発音指導のチャンネル「Rachel's English」が YouGlish を使い、英語の発音指導を行っているのが知られている[5]。

2.4 VoiceTube

もう一つの動画音声コーパス VoiceTube は、英語にフォーカスしており、英語動画とリアルタイムの字幕を「ビジネス／経済」「英語／外国語」「映画／ドラマ」「音楽」など 10 ジャンルごとに表示する英語学習支援のプラットフォームである。このプラットフォームは A1 から C1 まで英語の難易度ラベル表示機能、シャドーイング録音機能を備えている。

VoiceTube は YouGlish と類似する動画音声コーパスであり、使い方も YouGlish と同じように VoiceTube の検索欄へのキーワード入力を通じて関連する動画とリアルタイムの字幕が表示される。ただし、このコーパスでは、ジャンルごとの動画閲覧、難易度ラベルの可視化、シャドーイング機能の搭載は YouGlish にないものである。動画音声コーパスとしての VoiceTube には、①実生活での言葉遣いの提示およびディスコースレベルのインプットの提供、②自らの声を録音するシャドーイング機能の存在、③難易度とジャンルなどカスタマイズされた言語学習支援というポイントが特徴とされている (Jin, 2017)。また、VoiceTube を英語教育へ応用する調査によると、このコーパスを使うことで視覚と聴覚による刺激の相乗効果があり、学習者の英語リスニングの能力を高められる(Li, 2018)。

2.5 CVC（中国語視聴コーパス）

CVC（中国語視聴コーパス）は、映画、ドラマ、ドキュメンタリー、シチュエーションコメディなど 12 類別の動画を収集した中国語教育に向けた動画音声コーパスである。

このコーパスは中国語コーパスの中では比較的にまれな動画音声コーパスであり、従来の CCL（北京大学中国語言語学研究センターの中国語コーパス）、BCC[6]、MLA[7] など単一モダリティの情報だけ含めている伝統的な中国語コーパスと比べて実生活のコミュニケーション全貌により近い状況を示せると思われている(王, 2024)。一方で、CVC の中のデータは映画、ドラマ、ドキュメンタリー由来のものなのであり、最新のネット用語、若者言葉など実際の日常生活で常に使用するデータはまだ足りていない。最新の中国使用実態には、すべてカバーしているとは言えない。なお、このコーパスはドラマなどの動画を 15 秒ごとに切り取っており、より広いコンテキストの表示には足りないと指摘できる。

[4]YouGlish を紹介するウェブページ：https://youglish.com/about（2024 年 8 月 15 日最終アクセス）
[5]YouGlish を通じて発音指導を行う動画：https://www.youtube.com/watch?v=opKPVqxE_QY（2024 年 8 月 15 日最終アクセス）
[6]北京言語大学現代中国語コーパス：https://bcc.blcu.edu.cn（2024 年 8 月 15 日最終アクセス）
[7]メディア言語の中国語コーパス：https://ling.cuc.edu.cn/RawPub（2024 年 8 月 15 日最終アクセス）

なお、このコーパスはドラマなどの動画を 15 秒ごとに切り取っており、より広いコンテキストの表示に足りないのが指摘できるところである。

2.6 既存の動画音声コーパスの価値と課題

上文で述べたように、動画音声コーパスは、一般に動画共有サイトやドラマ上の動画を収集し、動画上の画像と音声、およびその中でのリアルタイムの字幕データをコーパスのインターフェイスに表示している。視覚と聴覚という二つのモダリティから言葉遣いの実態を把握できるのがこの類のコーパスの価値である。

一方で、既存の動画コーパスに対して、以下のような課題が指摘できる。

まず、言語データを随時更新できるコーパスの構築は課題である。昨今、新しい動画が UGC（ユーザ生成コンテンツ）形式で大量にアップロードされているが、既存の動画音声コーパス中のデータ量を相応に増やすにはコーパスの開発者がコツコツ作業し、データを少しずつ増やす必要がある。効率的にコーパスのデータを更新できるとは言えない。この傾向を中国語の動画音声コーパスに当てはめると、インターネット上の最新用語をカバーするにはまだ足りていないという状況である。

また、動画音声コーパスから正確な書き起こしの実現には余地があると考えられる。動画音声コーパスにおける話し言葉の書き起こしについては、今まで伝統的な方法には人間が用意した字幕がある。一方で、用意されたテキスト字幕がない場合、画像内字幕への OCR 適用、ないしは音声認識、読唇技術など異なるモダリティによる視聴覚音声認識（Audio-Visual Speech Recognition, AVSR）技術の採用で正確な書き起こしの実現にはさらなる精度向上が必要であると考えている。

3. 『中国語動画音声コーパス』の基本設計

『中国語動画音声コーパス』は、中国語の動画・音声とテキストが同時に含まれているマルチモーダルなコーパスである。また、このコーパスのデータは、インターネットから収集した動画であるため、ウェブ供給の（web-sourced）コーパスとも言える。本章では、このコーパスの基本設計について述べる。

まず、このコーパスは、インターネット上の動画共有サイトにアップロードされている動画とリンクし、その映像・音声・字幕情報などから総合的にテキストを取得し、タイムスタンプと対応するテキストデータを格納し、動画と同時に表示され、そのテキストを検索することで、動画の中の特定の箇所を確認することが可能になっている。

中国語の動画・音声とテキストが対応されているため、このコーパスを利用することで、中国語の学習者にとってリスニングやスピーキングスキルの練習に役立つのみならず、中国語の研究者・教育者にとって、伝統的なコーパスやデータ収集の手法によって容易コンコードを見ることができない言語データも確認することができ、用例収集や動画の情報量により、より豊富なコンテキストによって、言語運用を考察することができる。

当コーパス自体の公開スケジュールは未定であるが、本発表において当コーパスのデモ版を展示する。また、本コーパスを作成するシステムを近日オープンソースにて公開する予定である。

将来的には、中国語に限らず、他の言語についても、データ収集のパイプラインおよびデータ分析のプロセスを一般化することにより、OCR や音声認識モデルなどが備えられている他言語のデータを扱うことも考えられる。

図 1: 『中国語動画音声コーパス』のアプリケーション [8]

3.1 システムの構成

まず、当コーパスのシステムは、2つのサブシステムに大別される。1つはデータの収集・処理・変換を行うバックエンドであり、もう1つは検索・表示を行うフロントエンドである。バックエンドでは、データ処理パイプラインとデータ提供パイプラインに分かれている。

データ処理パイプラインであり、2つ目はデータの提供を行うデータ提供パイプラインである。

3.2 データの構造

本コーパスのデータは、主に動画のメタデータ（タイトル・チャンネル名・URL・動画の識別子など）および動画から得られたテキストデータによって構成される。これらのデータは、データベースに格納され、表示・検索可能である。

現在は試行的にYouTubeから一定量の動画のみを収集しているが、今後は、同じパイプラインを適用することにより、自動的にデータを収集することが可能であり、また、ユーザーによる対象動画の指定を受け付けることも可能であり、動的にコーパスを拡張することが容易である。

4. データの処理

4.1 データの選定と収集

前述のように、中国語の動画は字幕があるものが多いが、その場合でも一般的には字幕（サブタイトル）が画像データの一部として、レンダリングされた文字が埋め込まれている。一方、一般的に動画データとは別に[9]事前に用意された字幕の文字データが動画サイトにアップロードされ、いわゆるCC字幕と類似した形式になっているケースもあるが、中国語でこの形式は稀である[10]。

[8]取り上げられているサンプル動画：https://www.youtube.com/watch?v=JFHc3qynoEs（2024年8月15日最終アクセス）

[9]動画データと同じファイルに格納されているが、内部的には別データであるケースもある。

[10]この形式は中国語では「外挂字幕（外掛字幕）」という。

このため、中国語の動画をコーパスとして利用することは、英語などと比べて字幕の抽出が容易ではない。この問題を解決するため、本コーパスでは、例示的にこのような中国語の字幕付きの動画を収集し、そのデータを元にコーパスを構築する方法について模索する。

あくまで字幕が画像データとして生みこまれているものを対象とするため、本稿時点では特にそのような動画を収集することに焦点を当てている。また、基本的に、VLOG 類に比べ、料理動画は字幕が備えられているケースが多いため、例示的に料理動画をいくつかサンプリングした。

データの収集元は、動画共有サイトである YouTube 及び Bilibili[11]である。まず、手動で選択した動画の URL の一覧から、`yt-dlp`[12]を用いて動画および備え付けの字幕、メタデータなどのデータを取得する。取得されたデータに対し、音声データと動画の画像データを分けて、後述の処理を行う。

4.2 データ処理パイプライン

まず、動画の動画データ、音声データ、字幕データを収集し、そのデータに対してそれぞれ処理を行う。動画データには、字幕抽出のサブパイプラインに、音声データには音声認識のサブパイプラインに、字幕データには字幕変換のサブパイプラインにそれぞれ送る。これらのサブパイプラインは、それぞれのデータに対して[https://github.com/google/jax]処理を行い、最終的に複数のソースによるテキストデータを生成する。その後、これらのテキストデータを統合し、コーパスのデータとして提供する。

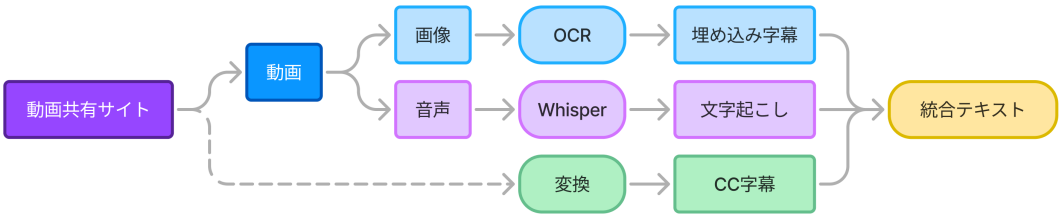

図 2: データ処理パイプラインのダイアグラム

4.2.1 音声データの処理

OpenAI 社の Whisper[13]によって、タイムスタンプ付きの音声書き起こしを行う。具体的には `faster-whisper`[14] および `JAX`[15] ベースの `whisper-jax`[16] という実装を利用し、処理の高速化を図った。また、オンデマンドなデータ処理のため、Docker コンテナを用いて、データ処理用の API を開発した。音声から得られた文字起こしデータを文字データとして、データベースに格納した。

[11]中国の主流動画共有プラットフォーム Bilibili のウェブサイト：https://bilibili.com（2024 年 8 月 15 日最終アクセス）
[12]yt-dlp のリポジトリ：https://github.com/yt-dlp/yt-dlp（2024 年 8 月 15 日最終アクセス）
[13]whisper のリポジトリ：https://github.com/openai/whisper（2024 年 8 月 15 日最終アクセス）
[14]faster-whisper のリポジトリ：https://github.com/SYSTRAN/faster-whisper（2024 年 8 月 15 日最終アクセス）
[15]jax のリポジトリ：https://github.com/google/jax（2024 年 8 月 15 日最終アクセス）
[16]whisper-jax のリポジトリ：https://github.com/sanchit-gandhi/whisper-jax（2024 年 8 月 15 日最終アクセス）

4.2.2 動画データの処理

まず、字幕は音声と同時に表示されるものであるため、音声に対応する字幕を取得したい場合は、発言している箇所のタイムスタンプの範囲を取得し、その発言に対応する字幕を取得する必要がある。そのためには、動画データと音声データの両方に対して処理を行う必要がある。よって、まず、動画データを画像フレームデータと音声データにわける。これには `moviepy`[17] を利用した。

それから、Whisper のように End-To-End の操作ではなく、発言している部分の音声に区切っておく必要がある。これを実現するためには、音声区間検出（Voice Activity Detection、VAD）という技術を用いることができる。それには `webrtcvad`[18] に独自実装の `vad_collector()`を用い、`scipy`[19] を利用し、音声データから発言している部分を検出し、セグメントの一覧を保存した。

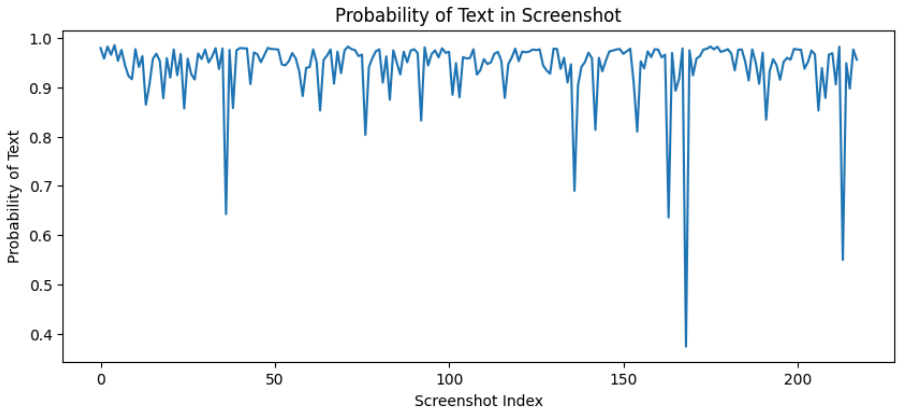

図 3: 全く異なる動画から訓練された字幕画像有無の判定結果

その後、動画データをフレームに分け、音声のある部分に対し、一定のサンプルレートにて動画フレームを画像データとして抽出し、抽出された動画を下 1/5 ほどに裁断する。また、SSIM などの類似度値を基準に、前後のフレームと比較することで、極端に似ているフレームを削除することにより、重複を減らし処理を高速化した。その後、予め訓練された SVM モデルにより、字幕の有無を判定し、字幕が含まれているフレームのみを保存した。これは、Hongyu Yan & Xin Xu (2020)と類似した手法により字幕の前処理を行った。

図 4: 大津の二値化法による字幕の処理

[17]moviepy のリポジトリ：https://github.com/Zulko/moviepy（2024 年 8 月 15 日最終アクセス）
[18]webrtcvad のリポジトリ：https://github.com/wiseman/py-webrtcvad（2024 年 8 月 15 日最終アクセス）
[19]scipy の公式ウェブサイト：https://scipy.org（2024 年 8 月 15 日最終アクセス）

さらに、図 4 のように OpenCV[20] にて得られた画像ファイルに対し大津の二値化法（Otsu binarization）を用いて、字幕の文字を際立たせてから、Paddle OCR を用いて文字認識を行った。しかし、縁取りや陰影などの処理が行われている動画に対しては、効果的に働くが、それ以外の動画に対しては、精度が低いことがわかった。

4.2.3 字幕データの処理

字幕データをそのまま利用し、テキストデータを変換して提供する。一般的には SRT や ASS、VTT などの形式によって動画共有サイトに公開されており、`yt-dlp` などのツールで簡単にダウンロードすることができる。ダウンロードされた字幕データは、タイムスタンプとテキストが対応した内部形式に変換し、以上の処理で得られた音声データと動画データと同じように保存される。

4.2.4 データの統合

まず、映像から得られた埋め込み字幕の OCR データ、音声から得られた文字起こしデータ、提供されている字幕データは、同じくタイムスタンプとテキストが対応した形式になっているが、完全に一致することはほとんどない。多くの場合、音声・文字認識の誤りや、区切り方などによって、異なるテキストが得られる。

このため、これらのデータをまず揃える（align）ことが重要である。時間の情報を予め取得されているため、幾何学的な計算によって、音声と字幕が多対多関係がある程度取得できる。その後、文字列比較の手法を用いて、差分を取り、文の単位で統合する。さらに、差分結果をグラフの形式で保存し、アプリケーションにおいては、同時に表示する。最終的にはコーパスデータとして提供する。

データ統合においては、データの蓋然性について、明確な解決法がまだ見つかっていないが、n-gram の計算や、言語モデルを用いる手法が考えられため、今後の課題としたい。

4.3 カスタムデータの提供

ユーザーによって提供されたカスタムデータから、同じ手順を経て、前述の処理パイプラインを通すことで、データを追加したり、ユーザー専用のコーパスデータを提供できる。そうするためのデータ自動生成 API やインタフェースおよびプロトコルを提供する。

5. 本コーパスの応用可能性

5.1 実際の場面と結びついた言語習得

本コーパスは、人工的に編纂した教科書や設計した語学授業だけにとらわれず、母語話者が実生活の中で自然に使っている言葉およびその発音の習得をサポートできる。特に、学校教育で学んだ中国語は母語話者がどのように使っておりどのように発音しているのか、または学校教育では教えられず中国語の母語話者が日常的に使う話し言葉はどのようなものかを、このコーパスを通じて提示できる。このように、実際の言葉の使用場面と結びついた効果的な言語習得の実現に本コーパスの応用価値がある。

[20]OpenCV の公式ウェブサイト：https://opencv.org（2024 年 8 月 15 日最終アクセス）

5.2 話し言葉の書き起こし改善に活用するリソース

本コーパスは、異なるリソースからの書き起こしの特徴を比較することができており、話し言葉の書き起こし改善に活用するリソースになる。現時点で、音声認識、埋め込み字幕に対する OCR、動画製作者の用意した字幕という複数モダリティとリソースから話し言葉の書き起こしを同時に表示している。将来は、音声学分析での方言識別と機械読唇技術を導入すると考えている。それぞれのリソースでの書き起こしの強みを見比べ、より正確な書き起こしの実装および自然言語処理の進歩に貢献できる。

6. おわりに

本稿では、『中国語動画音声コーパス』の構築について述べた。本研究では、中国語動画の字幕が一般的に画像データとして、動画に埋め込まれていることから、一例として中国語の動画に限定して、ウェブからデータを収集し、マルチモーダルなソースからテキストを抽出し、自動的に処理する方法を提案した。しかし、この方法は、他の言語にも適用可能であると考えられる。特に、収集から処理、表示するまで自動の手順は言語に依存せず、言語資源が十分に用意されている任意の言語に対し適用可能である。少数言語に関しては、音声認識や文字認識のインフラストラクチャが整っていないことが多いため、そのまま利用することは容易ではないが、工夫によってある程度可能と考えられる。実際に筆者の一人はアイヌ語に対し、類似した手法による処理の試みについて言及している(于, 2024)。

本研究の方法を利用することで、ドラマや講演などセッティングされている場の言語運用のみならず、料理や日常会話などインフォーマルな場面の言語 運用に対する調査が比較的に簡単に行えるようになることを期待する。また、動画データに対し、特に中国語のような埋め込み字幕が主流のデータに対して、単一なモダリティでは不十分の問題をある程度解消できると思われる。今後は、データを拡張し、より有効的なツールを利用し、より多くの言語に対して同様のコーパスを構築することを目指したい。

文　献

Jens Allwood (2010). “Multimodal Corpora”, *Lüdeling, A. & Kytö, M. Corpus Linguistics. An International Handbook*, pp.207-225.

Hossam Elshahaby and Mohsen Rashwan (2022). “An end to end system for subtitle text extraction from movie videos”, *Journal of Ambient Intelligence and Humanized Computing*, 13, https://doi.org/10.1007/s12652-021-02951-1.

Jo Shan Fu and Shih-Hsien Yang (2019). “Exploring How YouGlish Facilitates EFL Learners’ Speaking Competence”, *Journal of Educational Technology & Society*, 22:4, pp.47-58.

Haeyun Jin (2017). “VoiceTube Review”, *Pronunciation in Second Language Learning and Teaching Proceedings*, 8:1, pp.248-253.

Galip Kartal and Saadet Korucu-Kis (2020). “The use of Twitter and Youglish for the learning and retention of commonly mispronounced English words”, *Education and Information Technologies*, 25:1, pp.193-211.

Chia-Yi Li (2018). “The Use of VoiceTube for TEFL Listening Fluency”, *The 26th Korea TESOL International Conference-2018. Extended Summaries. Seoul: Korea TESOL*, pp.51–53.

Ibrahim Halil Topal (2023). “YouGlish: A web-sourced corpus for bolstering L2 pronunciation in language education”, *Journal of Digital Educational Technology*, 3:2, pp.1-8.

Hongyu Yan and Xin Xu (2020). “End-to-end video subtitle recognition via a deep Residual Neural Network”, *Pattern Recognition Letters*, 131, pp.368-375, https://doi.org/10.1016/j.patrec.2020.01.019.

中文语言资源联盟 (2024). - ChineseLDC.Org -, http://www.chineseldc.org/ (2024 年 8 月 14 日確認).

中文视听 (2021). CVC 中文视听语料库, HanLP.自然语义, https://client.chinafocus.net.cn/ (2024 年 8 月 14 日確認).

于拙 (2024).「アイヌ語テキストに対するルールベース処理の限界」『情報処理学会研究報告』2024-CH-135:No.9.

姚登峰, 江铭虎, 鲍泓, 李晗静, 阿布都克力木・阿布力孜 (2019).「手语计算 30 年：回顾与展望」『计算机学报』42.

孫栄奭 (2009).「終助詞ネの用法と視線行動—テレビ放送のマルチメディア・コーパスによる計量的分析—」『日本語学会 2009 年度春季大会予稿集』.

孫栄奭, 石井正彦 (2016).「映像 KWIC による言語行動の直観的探索—対談番組のマルチメディア・コーパスを例に—」『日本語学会 2016 年度春季大会予稿集』.

王涛 (2024).「基于 AI 技术的 CVC 中文视听语料库设计与应用」『科技与中文教学』15:1, pp.70-81.

石井正彦 (2009).「テレビ放送のマルチメディア・コーパス—映像・音声を利用した計量的言語使用研究の可能性—」『阪大日本語研究』21.

石井正彦, 孫栄奭 (2013).『マルチメディア・コーパス言語学—テレビ放送の計量的表現行動研究—』, 大阪大学出版会.

謝韞 (2005).「公開中の現代中国語話し言葉コーパスの紹介」『言語運用を基盤とする言語情報学拠点』, pp.157-162.

音声資源コンソーシアム (2024). コーパスリスト, https://research.nii.ac.jp/src/list.html (2024 年 8 月 15 日確認).

謝　辞

本研究を進めるにあたり、大阪大学人文学研究科准教授のBor Hodošček氏に多くご助言をいただき、この場を借りて深く感謝の意を表す。尚、本稿の内容の不備や誤りに関する一切の責任は筆者らに帰属する。