

論文の和文要旨	
論文題目	英会話教材として最適なインフォーマルな YouTube 動画の識別と評価
氏名	クーパー・クリストファー・ロバート

本研究の目的は、英語学習者（EFL 学習者）が英会話の特徴を認識するために活用できる YouTube 動画を特定することでした。この研究の必要性は、論文の最初の 2 章でレビューした文献に基づいています。第一言語の習得は音声から始まり、会話は言語の最も基本的な形態とされています。しかし、先行研究によると、EFL 環境における口語の教育と学習は、多くの場合、文字言語に基づいて行われています。これまでの研究では、様々なテレビ番組や映画と実際の会話を比較し、それらがインフォーマルな口語英語の学習モデルとして有効である可能性が示されてきました。しかし、YouTube などのオンライン動画プラットフォームの近年の人気や文化的重要性にもかかわらず、これらの研究は YouTube 動画が EFL 学習者の口語モデルとして適切かどうかを調査するまでには至っていませんでした。

本研究では、以下の研究課題に取り組みました：

RQ1: 日常的なインフォーマル英語の学習に適した YouTube 動画とはどのようなものか

RQ2: 異なる CEFR レベルの学習者に適した YouTube 動画とはどのようなものか

RQ3: インフォーマルな YouTube 動画と本物の英会話の言語使用にはどのような類似点と相違点があるか

RQ4: 英会話の指導に YouTube 動画の言語をどのように活用できるか

第一の研究課題に答えるため、2022 年にアップロードされた YouTube 動画から準ランダムサンプリングした字幕を収集し、会話的な言語を特定するために口語参照コーパスを使用しました。Biber (1988) のスピーチとライティングの変異モデルを活用するマルチディメンショナル分析タガー (Nini, 2019) を使用して、67 の言語特徴でテキストをタグ付けしました。各テキストのディメンションスコアを抽出し、クラスター分析を用いて、どの YouTube テキストが口語参照コーパスの会話に類似しているかを検討しました。k-means クラスタリングにより 2 クラスタ解を選択し、2 つのクラスタを比較するキーワード分析により、インフォーマルクラスタにおける会話的言語に関連する語彙項目の明確な傾向が確認されました。コーパスの内容の一般的傾向を記述するために、最新のトピックモデリングアルゴリズムである Top2Vec を使用しました。

第二の研究課題に対しては、B1 および B2 レベルの単語の割合などの言語特徴と、大規模言語モデルからのテキスト埋め込みを使用して、様々な機械学習手法を試みました。機械学習モデルは、英語学習ウェブサイトとリスニングテストから CEFR

に準拠したテキストを収集して作成した CEFR リスニングコーパスを使用してトレーニングされました。最も精度の高い方法は、OpenAI 埋め込みをサポートベクターマシンと組み合わせたもので、動画を A1、A2、B1、または B2+レベルに分類しました。

結果として、YouTube 動画の 69.7%が会話的な言語を含んでいることが判明しました。これは、大多数のトピックにおいて、他のクラスと比較してインフォーマルな動画の割合が高いことを意味しています。特に会話的な動画の割合が高かったトピックには、ゲーム実況、ヘアメイク、タロット占いの動画が含まれていました。これらの動画は一人の YouTuber が視聴者と対話する形式が多く、会話的特徴の使用により視聴者とのラポール形成を図っていると考えられます。ニュースやアニメ・映画のまとめなどのトピックでは、会話的な動画の割合は低く、これらのトピックの動画は会話の習得に有用な言語を含む可能性が低いことが示唆されました。

CEFR 分類モデルの結果によると、ほとんどの動画が B2 レベル以上 (95.1%) の学習者に適していることが判明し、これは Council of Europe (2020) の記述子で各 CEFR レベルで期待される能力と一致しています。B1 レベルの動画は約 1,000 本 (約 4.2%) 特定され、A1 および A2 レベルはごく少数でした。B1 レベルではトピックの範囲が広がり始め、A1 と A2 レベルの動画は主に子供向けまたは言語学習用でした。B1 レベルでは、非会話的な動画を含むトピックの数も増加しました。

会話的な YouTube 動画における会話ストラテジーの使用は、口語参照コーパ

スとおおむね類似していることが示されました。一部のケースでは、YouTube 動画においてモノローグ的な会話ストラテジーの使用が確認されました。これは YouTuber と視聴者との間に存在する擬似社会的関係に関連していると考えられます。モノローグ的な会話ストラテジーは、動画をよりインタラクティブにし、視聴者が YouTuber とのコミュニケーション行為に参加しているような感覚を持てるようにするために使用されていると推測されます。

本研究は学術的に複数の貢献をしています。方法論的には、これまでリーディング研究で用いられてきた CEFR 分類機械学習をリスニングに拡張し、小規模データセットから有望な結果を得ました。さらに、マルチディメンショナル分析の新しいアプローチとして、まずディメンションスコアを抽出し、それらを口語参照コーパスとのクラスター分析に使用して適切なテキストを特定する手法を提案しました。また、最新のトピックモデリングアルゴリズムを使用してコーパスの一般的傾向を記述しました。理論的には、このバージョンのトピックモデリングは、ストップワードの除外、言語をバッグ・オブ・ワーズとして扱うこと、トピック数の恣意的な決定といった、これまでの手法への批判の多くを克服しています。

教育的観点からも、本研究は重要な貢献をしています。会話言語学習に活用できる動画タイプについての一般的な記述を提供し、学習者の会話ストラテジーへの認識を高めるための顕著な例を示しました。これらのテキストは、第一言語との比較、

会話的発話の特徴の認識、モノローグ的な例を対話的な会話テキストに展開するための出発点として活用できます。本研究は、将来的な教育用ウェブサイトの作成を通じてさらに発展させることができる、YouTube 基盤の学習教材作成の基礎を提供しています。