

マルチメディア学習者英語プレゼンテーションコーパスの構築

橋本 喜代太

大阪府立大学人間社会学部
hash@lc.osakafu-u.ac.jp

竹内 和広

大阪電気通信大学情報通信工学部
takeuchi@isc.osakac.ac.jp

1 はじめに

言語を中心とするコミュニケーションの中で、スライド等を使い、身振り等も含めて複数人を相手に行なうプレゼンテーションはもっとも高度な総合的技芸の一つである一方、近年では誰しもがその必要性に迫られている。とりわけ英語によるプレゼンテーションは学術、ビジネス双方の現場でますます必要となりつつある。

筆者らは英語によるプレゼンテーションに関する包括的な学習・準備支援システムを構築するプロジェクトIIを進行中であるが、その際、次の3つの点で現実の学習者のプレゼンテーション例を多数収集する必要性に直面した。

- ・学習者のプレゼンテーションそのものの分析の必要性
- ・トピックなどが類似する他学習者のプレゼンテーションを閲覧したり、プレゼンテーションの巧拙を実例閲覧により学習したりする機能の必要性

前者については改めて指摘するまでもないことではあるが、実際、学習者の英語プレゼンテーションを多角的に多くの実例に基づいて分析をした研究は筆者らの知る限り存在しない。このため、今後の研究基盤データとしての収集は喫緊の課題である。一方、後者については收拾したデータをすべて学習者の閲覧に供する必要性はなく、相当限の取捨選択ならびに模範的プレゼンテーションの追加などが考えられるが、そのためには、そもそも巧拙がどのように判定されるかについてもデータを蓄積・分析する必要がある。

以上のような問題意識から筆者らは学習者の英語プレゼンテーションについて複数メディアのデータを収集するコーパスを設計し、第1目標とした100件の収集を終えたため、本論文において報告するものである。

以下の節でそれぞれの点について報告するが、筆者らの構築したコーパスの特徴は次のようにまとめることができる。

- ・形式を統一し、トピックを自由としたプレゼンテーションの学習者英語コーパス

- ・マルチメディアのコーパスとして、映像、スライドを含むと同時に、草稿段階のデータも含み、推敲履歴も観察可能
- ・教授者、ピア学習者による全体評価、部分評価を大規模に収録

本論文の構成は次の通りである。まず2節で収録されるプレゼンテーションの形式が情報提供型かつ5段落エッセイ構造型であることを示した後、3節で事前指導の内容、4節でプレゼンテーション準備の詳細、5節でコーパスデータの収集方法並びにデータの詳細について紹介する。6節はまとめである。

2 収録英語プレゼンテーションの形式

言うまでもなくプレゼンテーションは目的や内容によってさまざまな形態があり得る。プレゼンテーションは一般に聴衆の説得を狙うが、その際、聴衆の感覚・感情に訴えるかどうかで、新商品発表に代表されるような感情訴求型プレゼンテーション(emotive presentation)と、研究会発表に代表されるような情報提供型プレゼンテーション(informative presentation)におおむね分けることができ、ここでは後者を対象とする。

情報提供型プレゼンテーションでは、発表場所、聴衆にもよるが、説明・解説を目的とするもの、主張・説得を目的とするものに分けることができる。ただ、いずれにせよ、客観的証拠を提示しつつ、破綻のない論理的な展開で構成することが重視される。この点においては、書き物としての英語のエッセイ(小論文)とその特徴は並行的である。

エッセイは日本語における随筆(これもエッセイと表記されることが多いが、本稿で述べるエッセイとはこれとまったく別種のものである)と異なり、筆者の主張等について読者を説得し納得させることを目的とし、その内部構成は論理性が重視される。教育的にはそのもっとも基本的な構成として5パラグラフ・エッセイを取り上げることが多い。5パラグラフエッセイは基本的なパラグラフの「導入から主張支持、結語へ」といたる内部構成を素直に拡張したものとして、Introduction, 3 Supporting Paragraphs, Conclusion

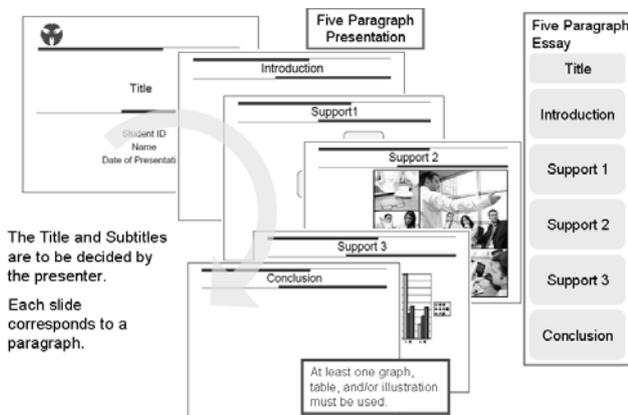


図1 本コーパスのプレゼンテーションとエッセイの対応関係

の5つのパラグラフで構成される。さらに長くなる場合も、それぞれの段落がこの構成を再帰的に持つような拡張で実現されていくべきである、というのが基本的な発想であり、ライティング教育においてはこの5パラグラフエッセイを一つの到達点とする。

そこで、筆者らは英語プレゼンテーションコーパスの構築に当たって、この論理構成に着目し、5段落構成で、各段落がスライド1枚に対応する短いプレゼンテーションを収集対象とすることとした。エッセイとの対応は図1に示すとおりである。

一方、ライティングやスピーキングの学習者コーパスなどではトピックを統一して論理構成を自由とするものが多い。これは一つには収録環境が影響する。例えば、NICTのJLEコーパスはアルクが行なっている英語スピーキングテストSSTを利用しており、イラスト描写やストーリーテリングを行なったものが収録されている。^[2] また、ライティングの学習者コーパスとしてはJEFLL Corpus^[3]などがあるが、これもゆるやかなトピックの設定がある。トピックを共通化することで、異なった学習レベルの学習者が類似テーマを扱った際の比較が行いやすい、語彙比較などが行いやすい、といった長所がある一方、特にスライドを伴ったプレゼンテーションの場合はインパクトや見栄えにこだわり論理構成に問題を多く残すものが増える恐れがある。これらのことも考慮し、図1に示す構成を持つ英語プレゼンテーションを採集し、コーパス化することとした。このため、扱うトピックについては賛否のはっきりするものを推奨するもの内容については自由としている。

3 収録英語プレゼンテーションの事前指導

現時点までで収録している英語プレゼンテーシ

ンは大学の1年生の英語の授業の一環として行なわれているものである。このため、一定の事前指導が行なわれている。

事前指導の形態は講義とオンラインで配布・閲覧できる教材と課題を組み合わせたもので、弱いブレンディドラーニングの形態を取る。オンラインで配布・閲覧できる教材は主に英語プレゼンテーションの基盤となる英語力の育成に関するものであり、発音、文法に関する教材と、そのための練習と課題が用意されている。その上で、講義においてその確認とともに、プレゼンテーションの全体像ならびに話、提示物(スライド)の構成について解説される。なお、この講義内容について別途、録画等によりオンラインで配布・閲覧される教材として作成中である。

4 収録英語プレゼンテーションの準備

既に述べたように収録される英語プレゼンテーションは授業の一環として行なわれている。その準備は次のようになる。受講者各自は自らのプレゼンテーション発表のおよそ1ヶ月以上前から準備を始め、発表予定日の2~4週間前に口頭内容の草稿とスライドを教授者に提出し、若干の添削と助言を受ける。この際、教授者は発見した間違いを直接修正するのは2割程度に留め、内容・表現についての助言と併せて各自に返却する。これを受けて当日までに最終的な口頭内容の草稿とスライドとを完成させ、発表に臨むこととなる。最終的な口頭内容の草稿とスライドは教授者による添削・修正は経ない。

このような事前の添削・助言を含むため、厳密には本コーパスのデータは100%学習者が作成したものではない点は、学習者各自の現状の英語力を忠実に反映しているといえないためにデータの信頼性に対する批判もあると考えられる。しかし、収録環境が授業に依存しているという点でやむをえないと同時に、特に大学低学年レベルでは一定の事前の添削や助言がなければ現実問題として英語の体をなさないものも残念ながら少なくない。これらの点を考慮した上で、草稿時点での添削・助言を経ることとした。

5 マルチメディア学習者英語プレゼンテーションコーパス

本節ではコーパス化の具体について述べる。

5.1 プレゼンテーションの収録

4節で述べた準備を経て、提出された最終の草稿・スライドを用いてプレゼンテーションが行なわれる。プレゼンテーションは教室でピア学習者の前で



図 2 収録時のイメージ

なわれるが、これを(株)メディアサイトの ML Recorder™^[4]を用いて図 2 のように発表者映像と提示スライドとを同時に 2 系統録画する。

5.2 内容データ

プレゼンテーションの内容そのものに関わるデータとしては表 1 にあるような 6 つの項目のデータが収録されている。

このうち、初期原稿、発表原稿、スライド原稿についてはミススペリングやタイプミスについては事前に修正を行なっている。どのようなミススペリングを起こすかも興味深い観点ではあるが、提出される原稿はいずれも Microsoft Word や Microsoft PowerPoint、OpenOffice.org などを利用して作成されており、自動スペルチェック機能で一定のミスは自動的に修正され、本来の誤り状況は不明である。このため、後の主たるデータ活用目的を鑑みて、スペリングやタイプレベルのミスは修正するものとした。ただし、この訂正は時間の関係上、事後となるため、実際の発表時に録画したスライド映像については訂正が行なわれていない。

現段階ではこれらすべてを手軽に閲覧できるインタフェースは用意できていないが、録画時に使用する

データ	内容
発表者映像	320x240 の MPEG2。スライドの切り替えのキュー信号の埋め込みあり
スライド映像	640x480 の JPEG
初期原稿	発表の約 1 ヶ月前に提出された口頭発表内容草稿
発表原稿	発表の数日前に提出された口頭発表内容の最終原稿
スライド原稿	発表用に提出された PowerPoint ファイル
口頭内容転写	発表時の口頭内容を転写したもの

表 1 内容面のデータ

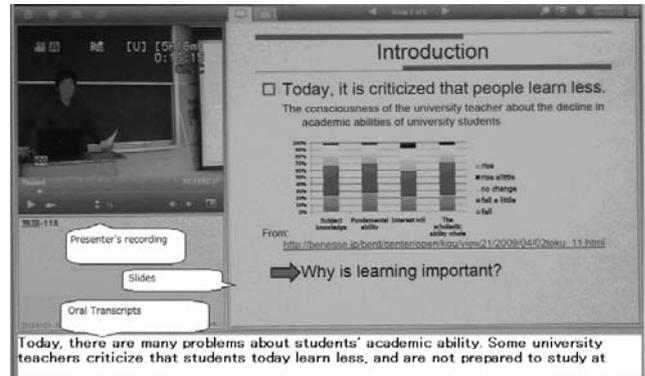


図 3 データの閲覧例

ML Recorder™が自動生成する Internet Explorer で閲覧するインタフェースを流用して、図 3 のような閲覧ができるようにしている。

一方、テキスト内容については現段階では形態素解析結果やエラーなどを示すタグを埋め込むことは行なっておらず、スライドとの対応関係を示す仕切りだけが埋め込まれている。その点で近年のライティングを中心とした大規模な学習者コーパスに比べると、内容データの構造化はなされていない。これについては膨大な人手が必要とされることから、タスクの優先順位も鑑み、当面予定しない。

なお、スライドの JPEG ならびに原稿については個人情報保護のため、タイトルスライドに示される学籍番号や名前、PowerPoint ファイルに埋め込まれるプロパティ情報は削除する等の配慮を行なっている。

5.3 メタデータ

内容面のデータに対して、その出所等を明示するためのメタデータは表 2 のようなものを用意している。

メタデータ	内容
ID	プレゼンテーション ID
UserID	発表者 ID
RecDate	録画日
Age	年齢
Grade	学年
Sex	性別
Time	発表時間
NumWords	発表で話された語数
Title	プレゼンテーションのタイトル
Topic	トピックを示すキーワード(最大 3 つ)

表 2 メタデータ

5.4 評価データ

本コーパスでは、教授者レベルの者、ピア学習者による評価データを収集する点が大きな特徴の一つで

ある。英語プレゼンテーションの評価についての考察回に基づき、全体評価と部分項目別の両方の評価を評価者各自が行なうものとした。ピア学習者についてはプレゼンテーション実施後、即座に評価を行なうものとし、教授者レベルの者については収録されたものを15"以上のモニタ上で全画面表示で閲覧して評価を行なうものとした。現時点では、授業の一環として行なわれたプレゼンテーションであるため、ピア学習者評価は1件のプレゼンテーションについて30~40件得られる。教授者レベルの者による評価については現在は4件ずつであるが、今後10件ずつ程度に拡充することを予定している。

この評価については部分項目別評価の評価観点は5側面12項目が用意されており、表3の通りである。一方、全体評価、部分項目別評価のそれぞれの評価ガイドラインは表4に示すとおりである。

この評価データの内容に関する分析は別の機会に譲るが、各プレゼンテーションに対して同一評価者による全体評価、部分項目別評価が各数十件ずつ採集されているのはプレゼンテーションに限らず本コーパスが初めてである。

5.5 個人情報保護とコーパスの公開予定

学習者の映像を含むデータはその性質上、個人情報保護について高い配慮が必要となり、コーパスの公開と合わせて検討を要する。

まず、学習者の名前や学籍番号等、個人を特定できる映像以外の情報についてはコーパスデータそのものには一切含まれない。各プレゼンテーションのUserIDは乱数データを活用して作成され、UserID

身体面	1 姿勢、ジェスチャー、アイコンタクトは適切か
	2 声の大きさや明瞭さが効果的か
全体	3 タイトルは適切か
	4 スライドも含め、全体構成は適切か
スライド	5 全体として説得的であったか
	6 各スライドの構成は適切か
	7 スライドは分かりやすく効果的か
発音面	8 グラフ、写真、イラストが適切・効果的か
	9 発音は適切であったか（正しさ）
文法面	10 流暢な発音であったか（スムーズさ）
	11 文法的な間違いはなかったか
	12 多様な表現、文法を使いこなしていたか

表3 本研究での部分項目別評価の観点

全体評価	30点：英語を母国語としない日本人のような大学生として理想的に上手である 20点：大学生として比較的上手である 10点：言いたいことは通じるがいろいろ問題がたくさんある (点数そのものは1点刻みでつけられる)
部分項目別評価	6：大学生としてきわめて優れている 5：大学生としてよくできている方だ 4：どちらかというによくできている（まんなかよりちょっと上） 3：もうちょっと努力が必要（まんなかよりちょっと下） 2：問題が多く、かなり改善が必要 1：きわめて問題が多い

表4 評価尺度のガイドライン

と実際の学籍番号や名前との対応については、コーパスとは別のネットワークからも切り離された保管場所に保管されている。

その上で、収録とその分析については学生各自に許諾を得ているが、後に公開を予定するプレゼンテーションについてはさらに名前等を特定しない形での公開について同意許諾を得るものとしている。この公開許諾を得たデータについては2012年度中をめどに部分コーパスとして公開を予定している。

6. おわりに

本稿では筆者らが現在構築中のマルチメディア学習者英語プレゼンテーションコーパスについて報告した。

謝辞 本研究は科研費(22520576)の助成を受けたものである。

参考文献

- [1] Hashimoto, K. and K. Takeuchi (2010) Prototypical Development of Awareness Promoting Learning Support System of Basic Presentation, *Proc. of 2nd International Symposium on Aware Computing (ISAC2)*, 304-311
- [2] 石田 亨ら (2003) 適合型コミュニケーションの研究開発, 『平成15年度通信・放送機構研究発表会予稿集』
- [3] JEFFLL プロジェクト
<http://jeffll.corpuscobo.net/>
- [4] ML Recorder™ <http://www.mediasite.co.jp/product/mlrecorder.html>
- [5] 橋本喜代太・竹内和広 (2010) 英語プレゼンテーションの評価尺度の提案と検証, 『電気学会研究会資料 IS-10-050~072』 99-104