

VALIS

－ 発話データの制限的共有と分散処理に向けて －

河村まゆみ(kawamuras@pat.hi-ho.ne.jp) : 言語アノテータ
前坊香菜子(xiangcai_2@suou.waseda.jp) : 早稲田大学日本語教育研究センター
楠元範明(moto@waseda.jp) : 早稲田大学教育・総合科学学術院・総合研究機構情報教育研究所
前野譲二(joji@mnc.waseda.ac.jp) : 早稲田大学 MNC・総合研究機構情報教育研究所
原田康也 (harada@waseda.jp) : 早稲田大学法学学術院・総合研究機構情報教育研究所

第五著者を研究代表者とする一連の研究プロジェクトでは、コンピュータと教育研究会 CE-69 (3) 『エーワンのマルチカードを用いた英語応答練習』(2003年5月16日)にて報告した応答練習を中心とする学習活動をコンピュータと教育研究会 CE-80 (4) 『対面での応答を重視した英語学習活動と発話収録装置の試作と試用』(2005年6月18日)にて報告したマルチトラック・ハードディスク・ならびにコンピュータと教育研究会 CE-88 (24) 『VALIS : 学習者プロファイルに基づく学習者音声コーパス構築を目指して』(2007年2月17日)にて報告したブルートゥース・ワイヤレス・マイクとハードディスク・ビデオカメラを用いて収録している。収録した発話データはコンピュータと教育研究会 CE-90 (1) 『VALIS : 英語学習者発話データの書き起こし』(2007年7月7日)にて報告したようにきわめて多量であり、書き起こしそのほかのアノテーション作業を実行する上では分散環境の構築が必須である。また、プロジェクト管理者・システム管理者・音声ファイル管理者・書き起こし作業など、さまざまな役割に応じてファイルへのアクセス制限に階層的な制約を組み込むことも必要である。本稿では、CMS (content management system) と RDBMS (relational database management system) を利用した web インタフェース経由でのアクセス管理とネットワークを利用した分散環境構築に向けての試みを紹介する。

VALIS: Distributed Network Support for Annotation and Constrained Sharing of Utterance Data Obtained from Japanese College Learners of English

Mayumi KAWAMURA (kawamuras@pat.hi-ho.ne.jp): Language Annotator
Kanao MAEBO (xiangcai_2@suou.waseda.jp) : Center for Japanese Language, Waseda University
Noriaki KUSUMOTO (moto@waseda.jp): Faculty of Education and Integrated Science & Institute for DECODE, Waseda University
Joji MAENO (joji@mnc.waseda.ac.jp): Media Network Center & Institute for DECODE, Waseda University
Yasunari HARADA (harada@waseda.jp): Faculty of Law & Institute for DECODE, Waseda University

In this paper, we report on the hardware and software architecture for a network-based distributed support system for transcribing and otherwise annotating multi-media language data obtained from learners of English during activities in Japanese college English classes. A typical audio file recording students' interactions in English and Japanese with a sampling rate of 96 kHz in 24 bit sampling running from 20 to 80 minutes constitutes a file of size 400 to 600 mega bytes. Using annotation tools such as TableTrans, those files are manually sliced into segments and down-converted into smaller audio files. Annotators, with proper authentication, can then access particular files via web browsers and either transcribe or otherwise annotate the target audio files. In this paper, we will mainly outline our software architecture and how we process those audio files.

1. はじめに

著者たちの研究グループでは、アカデミック・リテラシの習得において学習者の相互作用が重要な位置を占めていること¹、最も広い意味での学習履歴の蓄積と分析が学習活動の有効化と学習資源の改善のために必要であること²、外国語の運用経験が外国語の学習に不可欠であること³などを前提にこれまでの研究を進めてきた。2004年度に早稲田大学特定課題研究助成費（一般助成）課題番号 2004A-033『大学英语教育高度化のための外部試験を活用した学習者プロファイリングの研究』（研究代表者：原田康也）を受け、学習者の発話を比較的高度な品質でデジタル録音する装置⁴を試作した。続く2005年度に、同じく早稲田大学特定課題研究助成費（一般助成）課題番号 2005B-022『英語教育高度化に向けた学習者プロファイリングとマルチモーダル学習者コーパスの研究』（研究代表者：原田康也）の助成を受け、この装置を実際に使用して授業中の学習活動の音声収録を開始した。2006年度には科学研究費補助金（2006年4月 - 2009年3月）基盤研究（B）：課題番号 18320093『学習者プロファイリングに基づく日本人英語学習者音声コーパスの構築と分析』の交付を受けて、上記のデジタル録音装置に加え、13台のビデオカメラとBluetooth・ワイヤレス・マイクを用いて学生の英語による発話の収録を進め、2007年度には収録したデータの書き起こし手順について具体的な検討を始めている。

2. 音声データの収録と書き起こしの方向性

2.1. 書き起こし対象データ

本研究プロジェクトで集積を目指している主要なデータは大学生英語学習者の英語による発話と、それぞれの発話を行っている大学生の学習経歴や英語学習到達度の指標となるような外部試験のスコアである。発話としては、応答練習においては担当教員があらかじめ用意した質問を（2回）読み上げる音声とこの質問を聞いて（多くの場合あらかじめ用意

していない比較的自発的な）回答をする音声を中心となる。また、4名ないし6名程度のグループ内で行う少人数での相互プレゼンテーション活動における英語による発話は、発表者による（短い準備時間の後の、比較的自発性の高い）英語による発話を中心で、場合によってプレゼンテーションの途中またはプレゼンテーションのあとでその場で思いついた簡単な質問と、これに対するその場での自発的な回答が得られる場合もある。また、2年生の授業などグループで調査してきた内容を発表する場合には、あらかじめ発表原稿を用意して読み上げた内容が収録対象となる場合もある。

こうした英語での応答を除くと、音声収録を行っている学習活動中の受講生による発言の大部分は日本語によるものであり、ある意味で当初想定していた主たる分析対象のデータではないが、本研究の最終的に目指すところは単なる発話データの収集だけではなく、これを通じて学習者の自律的相互学習を根拠付ける実証的なデータ⁵を得るところにもあるため、日本語による相互交流⁶についても可能な範囲で分析可能となるようにデータを整備することが望ましい。

2.2. 音声データに対するアノテーション

一口に transcription {書き起こし|転記} といっても、対象とするデータとそのデータの想定される主要な利用方法によって作業内容・作業方針が異なってくる。音声データに対する transcription {書き起こし|転記} の annotation {注記|タグ付け} に限定しても、何が音的に発声されているかの annotation {注記|タグ付け} と何が言語的に発話されているかの annotation {注記|タグ付け} とでは、transcription {書き起こし|転記} に用いる記号も異なり、必要な事前の準備や作業員に対する訓練も異なってくる。

現在収録中のデータをどのように活用するかは今後の検討事項であるが、応答練習や少人数でのプレゼンテーションなど比較的自発的な英語による発話

⁵ 2名を超える複数話者のインタラクションの記録としては、コストや運用面の制約を度外視すれば、[1]・[3]などに見られるように、音声や画像の記録に加えて、相互作用の参加者の視線・注視点や各種バイオメトリック・データなども含めた包括的な記録が望まれるところである。

⁶ インタラクションの書き起こしならびに分析については、榎本美香・伝康晴を中心とした一連の研究報告がある。発話単位の認定基準などについては [2] の紹介がわかりやすい。

¹ 詳細については[4]などを参照。

² 詳細については[7]などを参照。

³ 詳細については[5]などを参照。

⁴ アレシス製ハードディスクレコーダ Alesis ADAT HD24 XR: 24-Track Hard Disk Recorder を中核とする機材を特注した。

の中で学生がどのような語彙や文型パターンなどを使用できるか観察し、日本語でのインタラクションから自律的相互学習の証拠を抽出しようというのが当面の目標である。英語の応答についてはあらかじめ用意して印刷・配布する質問のほかは大まかな時間枠の設定があるだけで応答内容については学生の自発的な発想に任されている上、日本語インタラクションについては教室内の英語学習活動の一貫であるという点を除いてほとんど何の制約もない。研究目的からもデータの長さから言っても、今回のデータに対してまず行うべき作業は言語的な書き起こしであり、データの内容の概観が用意できてから音声的研究資料の切り出しを考えるのが順当であろうと考える。

アノテーションツールを利用し、本稿で報告するように CMS (content management system) と RDBMS (relational database management system) を利用して web インタフェースから音声ファイルにアクセスする環境を構築すれば、音声そのものの研究・検討には音声データを直接操作することが可能となり、これを記号化した書き起こしは必ずしも必須とはいえなくなる。

こうして当面の達成課題を極小化しても、ひとつのトラックのデータが 20 分から 80 分超にわたるデータ長に加えて、応答が英語学習者による比較的自発的な発話であるという点がこの書き起こし作業を困難にしている。

出版物の文字テキストなど、母語話者により注意深く執筆・編集・校正された書記データであっても、学習者の言語生産物には何らかの『誤り』が含まれることが一般的であり、効率的な検索のためにはこれを何らかの形式で『正規化』する必要がある。しかし、学習者による『誤り』を含む言語生産物を『正しい』形式に修正することは、言語使用者の『意図』の推定が必要となるため、多大な困難を伴う。⁷ たとえば、There was dogs in the room. という文があった場合、書き手が単数の犬を意図して名詞の形式（とそれに伴う冠詞）を間違えたのか、複数の犬を意図して be 動詞の形式を間違えたのか、

⁷ 本プロジェクトの現段階では、『誤り』に関するタグを書き起こしの初期段階から記録することは考えていないが、allow を『アロウ』と発音するなどの典型的な誤りについては繰り返し特徴的に現れることも多く、気がついた範囲で記載するべきか、あとでまとめて処理すべきか、作業の進め方について試行錯誤している段階である。

この文を見ているだけではまったくわからない。前後関係を見ていくと、そもそも犬ではなく、何か別の動物であったというようなことがわかるかもしれない。非母語話者による発話の場合、これに加えてさまざまな発音の『くせ』や『誤り』が加わるが、人間が言語を聞いて理解する場合には文脈に依存する部分も多く、『誤り』を認識せず話し手の意図を汲んで『正しく』聞き取ってしまうことも多い。こうした作業による『偏り』に対処するためには、ひとつのデータセグメントに対して複数の作業者がアノテーションを付与して、その傾向を分析する必要もある。

2.3. 音声データアノテーションツールの選定

音声データに対する書き起こしも含めたアノテーションの方法論については[10]-[12]が基礎的な情報を整理している。

Windows で利用できる音声データ処理用ソフトウェアのうち、ライセンスや使用料も含めて利用しやすいものとしていくつかの候補を検討し、⁸2006 年度後半に、動作状況を確認のため、LDC⁹の Wavesurfer, MultiTrans, TableTrans など AGTK¹⁰ ツールキットを SourceForge.net よりダウンロードし、作業用 PC にインストールした。

インタラクション研究のための音声データ書き起こしには MultiTrans を使用する例が多いように見受けられるが、今回収録した音声データは最大 12 トラック同時収録でひとつのトラックに 2 名ないし 3 名の受講生の音声収録されている¹¹ため、複数のトラックのそれぞれに単一話者の音声を収録したデータの書き起こしを前提とする MultiTrans が最適なツールとは判断しがたかった。

TableTrans はひとつのトラックの音声を時間的な小部分に領域分割しつつ、開始点・終了点情報とあわせて、複数の属性を定義してこれを表形式で入

⁸ ビデオカメラを使用して音声・画像も収録しているが、操作を学生に任せるため、学習活動の記録として重要なインタラクションが欠落することも多く、音声データに比して補助的な記録手段と考える必要がある。

⁹ LDC は Linguistic Data Consortium
<http://www ldc upenn edu/> に詳細。

¹⁰ AGTK は Annotation Graph ToolKit
<http://sourceforge.net/projects/agtk/>
よりダウンロード可能。

¹¹ 本研究プロジェクトでは、授業に限定して音声収録の設置と撤収を完了しなければならないという運用面での制約ならびに予算上の制約などから、3 人の学習者に 1 つのトラックを割り当てている。

力・表示できるため、話者・使用言語・発話内容を最小限の情報として書き起こしする今回の作業については適切なツールと思われた。¹²

2.4. 試行段階での作業の流れ

2007 年度前半では主に河村を中心にいくつかの音声トラックについて試行的に書き起こしを行った。この試行作業の方針として音声転記でなく、音声ファイルがオリジナルデータで、それに対する注釈（アノテーション）として書き起こしを付けていると考えるとともに、ひとつの音声トラックのすべてのセグメントに対してアノテーションをつけ、英語の間違ひについては事後に処理することを期待して扱わないこととした。音声ファイルを TableTrans で読み込んで、発話単位ごとにセグメント化しつつ、若干のタグを挿入し、書き起こしを直接書き込むという前提で作業を進めてみたが、ソフトウェアの動作上の特性などもあり、円滑な進行からは程遠い状況であった。このため、ひとつのトラック（20 分程度）のおおまかな書き起こし作業に 4 時間から 10 時間程度を要した。

まず新学期の授業 1 週間分（2 年生 Theme 2 クラスと 1 年生 Bridge 3 クラス）の音声データの書き起こしから試行作業を始めたが、書き起こしの詳細度や手順を検討するための準備段階として位置づけ、いくつかの方法を試みた。さまざまな組み合わせを仔細に検討する時間的余裕はなかったが、2 クラス 20 分×12 トラックの概略の書き起こしが完了し、データベースとの連携のためのアノテーション記号の見直し作業を進めた。

英語部分については、開始点と終了点の確定も比較的容易で、エラータグについて見送る限り、ある程度の時間で一通りの作業が完了する。しかし、日本語部分については話者交代が明瞭でない場合が多く、音声小さく聞き取りにくい¹³場合は波形で分割の検討をつけるということもうまくできないため、セグメント化の作業に多大な時間がかかる場合もあった。ひとつのグループが他のグループの受講生に進め方を問い合わせることもあり、そうなると話者の識別がますます困難になる。話者の交代が多くなすぎると、これまたセグメント化が難しくなる。こうしたことから、トラックによってはセグメ

¹² TableTrans を先行して活用している国内の研究グループを探しているが、いまのところははっきりしたことがわかっていない。

¹³ 学期始めの授業では学習活動ならびに録音作業に慣れていない側面もある。

ント化と書き起こしで 7 時間以上も必要となった。別のグループに関しては 2 時間から 3 時間程度で概略の書き起こしが終了するが、それにしても 1 クラス分の書き起こしに対して 36 時間かかるとすると、専従的な作業 1 週間分の作業量に近く、1 週間 5 クラスの概略の書き起こしを完了するには、専従的な作業の 1 月分の業務量を要するという概算になる。

3. 分散環境の構築と活用に向けて

上記が示唆するように収録した発話データはきわめて多量であり、書き起こしそのほかのアノテーション作業を実行する上では分散環境の構築が必須である。発話データには Versant for English¹⁴ならびに TOEIC¹⁵などの proficiency test に基づく proficiency level など発話者のさまざまな属性に関する情報を付与する予定であるが、これらの属性に基づく検索要求に対応するためには、セグメント化した音声ファイルに RDBMS (relational database management system) 経由でアクセスすることがいざいざにしても必要となる。また、プロジェクト管理者・システム管理者・音声ファイル管理者・書き起こし作業員など、さまざまな役割に応じてファイルへのアクセスに階層的な制約を組み込むことも不可欠である。こうした観点から、CMS (content management system) を活用した web インタフェースによる制限的アクセスと書き起こしも含めたタグ付与作業のための分散環境構築に向けて検討を始めた。

3.1. 音声データと書き起こしソフトの特徴

収録している音声データは大量で書き起こしに時間がかかることが予想されるほか、以下のような特徴がある。

- ・ オリジナルファイルのサイズが大きい：1 track の録音時間が約 20 分でファイルサイズが 300MB ないし 500MB。転送・コピーの所要時間が大きく real time で聞くのに 20 分かかる。大まかなセグメント化の作業に 1 時間、ある程度まで詳細なセグメント化に 3 時間を要する。
- ・ 扱うファイルの数が多いため：1 コマ 90 分の授業で最大 12 トラックを週 5 回収録し、1 学期の授業回数が 12 ないし 14 週とすると一学期で

¹⁴ <http://www.ordinate.com/versant/versant.jsp>

¹⁵ <http://www.toeic.or.jp/toeic/>

720 ないし 840 トラックとなる。(2 年生の授業では 2 ないし 6 トラックの収録もある)

- 自発発話が含まれ、書き起こしに時間がかかる要因となる。
- 書き起こし作業に当たって、教室の学習活動に関する理解が必要となる。
- 発話内容に音声など広義の個人情報が含まれるため、セキュリティへの配慮が必須。

また、採用した書き起こしソフト TableTrans の動作に不明な点があり、操作に困惑する場合がある。このため作業手順に課題があり、書き起こしの作業手順習熟に時間を要する。

- 利用法に関するドキュメントや web 上の情報が少ないため、作業結果の保存が思ったように行かない時など、原因を特定するのが困難な場合がある。
- アノテーション・ファイルの文字コードは初期設定では UTF-8 で、そのまま Excel で開くと日本語を含むテキストが読めない。プログラム付属の SHIFT-JIS の英文サンプルテキストに日本語を加えて保存してみると、その部分は文字化けしてしまうなど、さまざまな制約がある。
- 初期設定では、一般のパソコン操作とキー割り当てや挙動が異なり、操作の undo ができないため、注意が必要。
 - ▶ 音声ファイルを読み込むと、同じ名前の SHAPE ファイルができる。初期設定では(ファイル名).shape というファイルを作る。この状態で(別ディレクトリの)同じ名前の別の音声ファイルを読み込むと、以前の音声波形が表示される。¹⁶
 - ▶ セルに text をペーストすると、スペースで分かれた語は、縦方向に 1 セル 1 語単位で入力される。

こうしたことから、新たなアノテータ要員を採用しようとした場合には、準備作業や手順説明が煩雑となる。伝達すべき内容には(1)ソフトの使い方・くせ(2)データ利用の仕方・転送(3)作業で使用する文字コード(4)作業記録・文法事項・用語説明など、内容的にレベルの異なる事項が混在することとなる。

¹⁶ アレシス製ハードディスクレコーダ Alesis ADAT HD24 XR で録音した音声ファイルを Windows にエクスポートすると(デフォルトでは)ひとつのフォルダに Track01 から Track12 までのファイルを作成する。

作業に使用する PC 一台一台について必要なソフト一式をインストールし、環境設定する必要があるほか、データについても一つ一つ利用するに整形しなければならない。このような状況から、外付けハードディスクに生データ(またはそのコピー)を用意して行う書き起こしの手順はアノテーション作業者に対して scalable でないことが直ちに了解される。

新たなアノテーション作業者に対して伝達すべき事項は書き起こしなどのアノテーション作業に本質的な規則と方針のほか、作業に必要なソフトのインストールと操作方法の説明・並行的に使用する必要が生じる複数のソフトの使い方と作業手順・そのほかの詳細な注意事項にわけて考えることができる。web インタフェースを採用することで、作業手順を単純化し多人数で分散して作業を行うことが可能となるほか、特殊なソフトウェアのインストールならびに習熟が不要となり、アノテーション作業者を増やすことが比較的容易になる。音声データは最終的に何らかの形で(アクセスに制限を付与して)ネットワーク的に共有することを想定しているため、分散作業環境の構築はいずれにしても必要な準備作業であるため、こうした環境を早めに構築することがいずれにしても必要であるとの認識に至った。

3.2. 分散環境の構築

収録した音声・画像ファイルは一時的に USB2.0 外付けハードディスク¹⁷に転送し¹⁸、多重に保存している。試行的な書き起こし作業等には同様の外付けハードディスクを使用してきたが、今後の分散環境構築に向けて Lan 接続ハードディスク¹⁹とインテリジェント型高速ギガビットイーサネットスイッチ²⁰以下の機材を用意している。

web インタフェースからの音声ファイルへのアク

¹⁷ 株式会社バッファロー製 USB2.0 外付けハードディスク 2 ドライブ 1 テラバイト [HD-W1.0T1U2/R1]ならびに株式会社バッファロー製 eSATA&USB2.0 対応 外付けハードディスク 4 ドライブ RAID 対応 (2TB) [HD-QS2.0TSU2/R5]を使用している。

¹⁸ Alesis ADAT HD24 XR はリムーバブル・キャディ装着のハードディスクを使用するので、独自フォーマットの生データを残した上で Windows に転送したファイルを多重に保存している。Sony DCR-SR100 のハードディスクは Windows のファイル形式であるが、内蔵ハードディスクの容量が 30GB と限られているため、毎週の授業後にハードディスクに転送した上でファイルを削除している。

¹⁹ 株式会社アイ・オー・データ機器製 LAN 接続ハードディスク (4.0TB) HDL-GTR4.0

²⁰ デル株式会社製インテリジェント型高速ギガビットイーサネットスイッチ PowerConnect 2716

セスを可能とするため CMS (Content management system) と RDBMS (relational database management system) を組み合わせて使用することとした。

CMS をインタフェースとすることで、書き起こし等のアノテーション作業をネットワーク的に管理することが可能となり、パスワードの変更も各自で可能となる。²¹

音声ファイルについては、一時的なセグメントファイルは別として、サーバで集中的に管理することになり、データの保護の観点からも作業に生データをコピーするよりは望ましい管理となる。標準的な web ブラウザをインタフェースとするため、特定の書き起こしソフトをインストールし、環境設定し、使い方に習熟するという作業訓練過程を簡略化することが期待できる。必要なソフトをインストールしたパソコンに限定されず、いつでも作業を進めることができる。

RDBMS (relational database management system) を利用することで²² web インタフェースと合わせて、たとえば特定の学習者の音声ファイルをまとめて聴取するとか、特定の TOEIC スコア範囲の学生の音声ファイルをまとめて視聴するとか、特定の質問に対する回答をまとめて視聴するとか、というようなアクセスが（書き起こし作業の完了を待つことなく）可能となることが見込まれる。

前項で紹介したように、20 分程度のトラック単位で一学期に 800 トラック前後収録することになる。1 トラックを大まかに 4 種類の発話タイプに分類しながらセグメント化した場合 40 ないし 80 レコードとなるため、ひとつの学期のデータとして扱うレコードは 60,000 程度となる。1 トラックをさらに詳細に分割をした場合には 136,000 レコード以上となり、これだけの規模のレコードを管理するには RDBMS が必須となる。さらに RDBMS を使用することで、データの保全・保守・利用についての自由度が増すことが期待できる。

3.3. 分散環境でのタグ付与作業

これまでの試行的作業で行っていたタグ付与のうち、今後とも継承されるものとして以下が想定され

る。

・ アノテーションタグの種類

- TableTrans が自動的に付与するタグ
 - ✧ ID : 音声ファイルをセグメントに切り分けたファイルの識別番号
 - ✧ Start : 区切ったセグメントの開始時間 (秒表示²³)
 - ✧ End : 区切ったセグメントの終了時間 (秒表示)
- 独自設定タグ²⁴
 - ✧ SP (Speaker) : 発話者个体識別番号
 - ✧ PrimaryLanguage : インタラクションで主となる言語
 - ✧ UtteranceType : 発話タイプ
 - ✧ Text : 発話内容の書き起こし

- ・ UtteranceType (発話タイプ) の大まかな分類
 - Q : カードの質問を読む部分
 - R : カードに対する答えの部分
 - D : 発話者が自分の名前などを言う部分
 - P : 上記の Q, R, D 分類以外の部分

アノテーション text には英語も含まれるため、他のソフトでの利用も考え (英文中では使用しない) タブでの区切りに直した。

新しい分散環境を前提とするとアノテーション作業の流れは概略以下ようになる。

- ・ 音声ファイルのセグメント化・最適化
 - 音声ファイルをセグメント化・アノテーション記号 (発話タイプ) 挿入 : TableTrans を使って、読み込んだ音声ファイルを発話タイプ別にセグメント化し、独自設定項目も書き込んでおく。
 - アノテーション整形 : 記号のみ付いたアノテーションをサーバに用意されたデータベースの表で読み込める形に整形する。web 上では音声データリンク用のフィールドなども付け加えるので、双方の項目とデータ型を一致させる。
 - 音声ファイルの分割 : 音声ファイルの sampling rate などを変更し、発話タイ

²¹ CMS としてはオープンソースの Zope を採用した。Zope については次を参照。
<http://www.zope.org/>

²² RDBMS (データベース管理ソフト) としてはオープンソースの MySQL を採用した。MySQL については以下を参照。
<http://www.mysql.com/>

²³ 秒表示 : 1228.475 は Track の開始から 20 分 28 秒。

²⁴ web 上の作業のためには、アノテーション付与作業者の ID などいくつかつかわれる予定。

ブ単位に一括して分割する。²⁵

- 分割したファイルのアップロード
 - 整形したアノテーション・ファイルと分割した音声ファイルをサーバに転送²⁶
 - 発話タイプ D のアクセス管理確認（発話者の氏名等を述べているため）
 - ファイル名の付与
 - サーバに置かれたデータベース管理ソフトでアノテーション・ファイルの一括読み込み
- 分担して書き起こし等のアノテーション作業
 - 作業者の認証
 - 音声ファイルの一覧取得：Relational database のアノテーションデータを CMS で一覧表示する。アノテーションテキスト部分はまだ空欄。音声ファイル名に音声へのリンクを設定する。
 - アノテーションの入力・編集：作業者は web ブラウザでサーバにアクセスし音声を聞き、書き起こした結果も web 上で入力する。編集・訂正もできる。

3.4. 将来的な展望

書き起こしを含めてデータのアノテーションには誤りや不統一が混入する。同一の音声ファイルに対して複数のアノテーション作業（または複数回のアノテーション作業）によりどのような異同が生じるかを確認することも本来は必要である。また、それぞれの作業者の信頼性などを評価することも必要となってくる。このほか、アノテーション作業者の属性（日本語母語話者・英語母語話者・年齢・日本の大学生への接触度）による作業結果の違いが書き起こしに何らかの系統的異なりをもたらす可能性も検証が必要である。こうしたことから、今回報告するような relational database の構築が求められている。さらに、効率的な作業進行のためには、作業手順を簡略化し、アノテーション担当者の人数を増やすことを可能とする必要がある。担当箇所を決め、作業を単純化するほか、音声認識による自動 speech-to-text 変換ソフトなどを利用した自動処理なども視野に置く必要がある。

書き起こしも含めたアノテーションに関して、現在のプロジェクトでは予算・人手が決定的に不足し

ている。本格的なアノテーション作業については別途予算の確保が必要である。外部資金の獲得と並行して、本プロジェクトが収集するデータに関心を寄せる組織・企業との共同作業を可能とするような枠組みづくりも視野に入れる必要がある。

上記のような労働集約的手作業での書き起こしを現在収録中のデータすべてに対して行うことは現実的でもないし、現在のプロジェクトの予算で実行可能でもない。また、そのような作業が必要とされるかどうかは必ずしも明らかでない。RDB を利用した web インタフェースからの音声ファイルへのアクセスが実現すると、書き起こしがなくとも同一の質問に対して proficiency level の異なる学生がどのように回答しているかの比較を（音声として）確認することができる。あるいは、同一の学生が質問を読み上げるときと質問に回答するときで（発音や流暢さに関して）同じような特徴を示すのか異なる特徴を示すのか（音声として）比較することが可能となる。こうしたアクセス環境の整備が進めば、全体としての書き起こしを待つことなく、さまざまな問題意識に基づいて各種の音声セグメントを視聴しながら、（必要があればこれを書き起こしつつ）学習者の発話コーパスから重要な知見を得ることが可能となることが期待される。このことが翻って長期的に書き起こし作業の充実につながることも期待できる。

こうした方向性とは別に、人手を介するとしても、音声認識エンジンにデータをフィードして、その出力をアノテーションに利用するような手法の検討も必要であろう。²⁷

4. 付記

本稿で報告する共同研究においては、楠元範明と前野譲二がデータ処理環境の構築と機材の選定に関わる情報提供と意思決定を行っている。音声書き起こしならびに分析用各種ソフトウェアの検討・試用については前坊香菜子が、デジタル音声収録装置の運用支援・試行的アノテーションならびに本稿で報告する CMS を活用した web 環境の構築については河村まゆみが主に担当し、早稲田大学法学部 4 年田中隆裕がアドバイスをを行った。

²⁵ TableTrans を使用する。

²⁶ WinSCP3 を使用する。

²⁷ 株式会社アドバンスト・メディア製 AmiVoice を試用してみたが、市販ソフトの活用は難しそうである。

5. 謝辞

本稿の著者たちを中心とする共同研究は科学研究費補助金(2006年4月-2009年3月)基盤研究(B):課題番号18320093『学習者プロファイリングに基づく日本人英語学習者音声コーパスの構築と分析』の助成を受けている。本稿で報告した発話収録装置の試作と試用にあたって早稲田大学特定課題研究助成費(一般助成)課題番号2004A-033『大学英語教育高度化のための外部試験を活用した学習者プロファイリングの研究』(研究代表者:原田康也)ならびに課題番号2005B-022『英語教育高度化に向けた学習者プロファイリングとマルチモーダル学習者コーパスの研究』(研究代表者:原田康也)による助成を受けている。

本稿で紹介するweb環境の仕様検討と構築ならびに本稿の起草にあたって、河村まゆみは(有)石川オープンシステム・ソリューションズの石川泰光氏・三菱電機システムサービス(株)吉井兼治氏のお二方から貴重なご教示・ご示唆を受けている。ここに記して謝意を表したい。

6. 参考文献

[1] 伊藤禎宣・岩澤昭一郎・土川仁・角康之・間瀬健二・片桐恭弘・小暮潔・萩田紀博,「装着型体験記録装置による対話インタラクションの判別機能実装と評価」,ヒューマンインタフェース学会論文誌, Vol. 7, No. 1, pp. 167-178, 2005/2.

[2] 榎本美香・石崎雅人・小磯花絵・伝康晴・水上悦雄・矢野博之,「相互行為分析のための単位に関する検討」,人工知能学会研究会資料, SIG-SLUD-A402, pp. 45-50, 2004.

[3] 角康之・伊藤禎宣・松口哲也・シドニーフェルス・間瀬健二,「協調的なインタラクションの記録と解釈」,情報処理学会論文誌, Vol.44, No.11, pp.2628-2637, 2003年11月.

[4] 原田康也・辰巳丈夫・楠元範明,「『情報教育』の情報化」,情報処理学会研究報告, Vol.2000, No.20, コンピュータと教育 55-6, pp.41-48, 情報処理学会, 2000年2月18日.

[5] 原田康也,「外国語学習における知的情報処理と言語処理技術の応用」,2001年情報学シンポジウム講演論文集, pp.25-32, 社団法人情報処理学会発行, 2001年1月18日.

[6] 原田康也,「エーワンのマルチカードを用いた英語応答練習」,情報処理学会研究報告 CE-69-3 pp.17-22, 情報処理学会, 2003年5月16日.

[7] 原田康也・前野譲二・楠元範明・辰巳丈夫,「学習履歴の双対性:学習履歴を活用した e-learning 高度化の数理的基礎を目指して」,情報処理学会研究報告 IPSJ SIG Technical Reports 2003-CE-70 (1), 学術刊行物 情処研報 Vol. 2003, pp.1-8, 社団法人 情報処理学会, 2003年7月12日, ISSN 0919-6072.

[8] 原田康也・辰巳丈夫・前野譲二・楠元範明・鈴木陽一郎,「対面での応答を重視した英語学習活動と発話収録装置の試作と試用」,情報処理学会研究報告 IPSJ SIG Technical Reports 2005-CE-80 (4), 学術刊行物 情処研報 Vol. 2005, pp.25-32, 社団法人 情報処理学会, 2005年6月18日, ISSN 0919-6072.

[9] 原田康也・前坊香菜子・河村まゆみ・前野譲二・楠元範明・鈴木陽一郎・鈴木正紀,「VALIS: 学習者プロフィールに基づく学習者音声コーパス構築を目指して」,情報処理学会研究報告 IPSJ SIG Technical Reports 2006-CE-88 (24), 学術刊行物 情処研報 Vol. 2007, No. 12, pp.169-176, 社団法人 情報処理学会, 2007年2月16日, ISSN 0919-6072.

[10] 原田康也・前坊香菜子・河村まゆみ,「VALIS: 英語学習者発話データの書き起こし」/「VALIS 2.0: Transcription of What was (not) Uttered」,情報処理学会研究報告 IPSJ SIG Technical Reports 2007-CE-90 (1), 学術刊行物 情処研報 Vol. 2007, No. 69, pp.1-8, 社団法人 情報処理学会, 2007年7月7日, ISSN 0919-6072.

[11] 原田康也・前坊香菜子・河村まゆみ・前野譲二・楠元範明・鈴木陽一郎・鈴木正紀,「学習者主体の英語学習環境の構築と学習者プロフィール・発話データの収集」,平成19年度情報教育研究集会論文集, pp.486-489, 2007年11月9日.

[12] 原田康也・前坊香菜子・河村まゆみ・鈴木正紀,「VALIS: 英語学習者のプロフィールと発話データの収集」, pp. 25-30, 信学技報 Technical Report of IEICE TL2007-11, (2007-12), 社団法人 電子情報通信学会, 2007年11月16日.

[10] 自然言語処理システムに関する調査報告書,社団法人日本電子工業振興協会, 2000年5月.

[11] 社団法人日本電子工業振興協会対話コンテンツ技術専門委員会 web page:

<http://it.jeita.or.jp/eltech/committee/knowledge/mmcampus/index.html>

[12] 音声対話技術コンソーシアム web page

<http://www.astem.or.jp/fistc/>