



高汎用性教室英語の発話コーパス構築の課題と蓄積の方向性

メタデータ	言語: Japanese 出版者: 北海道教育大学 公開日: 2016-09-30 キーワード: 作成者: 片桐, 徳昭, 大橋, 由紀子 メールアドレス: 所属:
URL	https://doi.org/10.32150/00006492

高汎用性教室英語の発話コーパス構築の課題と蓄積の方向性

片桐 徳昭・大橋由紀子*

北海道教育大学旭川校英語コミュニケーション学研究室

*ヤマザキ学園大学

Issues and Prospects in Designing and Accumulating Multi-Purpose Spoken Corpora of Classroom English

KATAGIRI Noriaki and OHASHI Yukiko*

Department of English Communication Studies, Asahikawa Campus, Hokkaido University of Education

*Yamazaki Gakuen University

ABSTRACT

This short paper summarizes issues and prospects in developing spoken corpora of Japanese English classroom conducted mainly by Japanese teachers of English. Classroom spoken corpora consisting of transcriptions from English lessons recorded in Japanese classroom setting are rarely obtained due at least four issues to clear: (1) obtaining teachers' or schools' consent to record the English lessons due to confidentiality of the students' identity, (2) length of time to transcribe the recorded speech, (3) designing tags to annotate the transcripts, and (4) classifying English and Japanese spoken tokens used in English lessons. This paper mainly discusses the latter two issues that entail assembling spoken classroom corpora in English as a foreign language (EFL) setting in order to aim at accumulating multi-purpose spoken corpora of classroom English for the betterment of English teacher education.

1. はじめに

コーパス言語学は機械可読性 (machine readability) を持った言語資料を分析する作業である。コンピュータの使用により、人間の手作業では膨大な時間のかかる作業を短時間で処理し、その結果を分析し、知見を得て積み上げて行く。

その定量的・定性的分析局面において機械可読となった言語資料 (コーパス) の大きな有用性が存在する。この資料処理において、コーパス言語学は検索技術を駆使して研究を進める (石川, 2012, p. 85)。ここに、処理技術的な側面が存在し、WordSmith Tools¹やAntConc²のようなコンピュータソフトウェアを用いて、コーパス検索を

し、コーパスから必要な情報を抽出し処理する。この場合、コーパスは研究の「情報資源」となる(田中, 2015, p. 48)。田中(2015)は情報学研究においては「情報処理技術」と「情報資源」の主従関係について、「情報資源」であるコーパスは従の位置付けとなると説明する。ここでコーパス言語学には、石川(2012)の言うように検索「処理技術」と「言語学としての分野」つまり「コーパス駆動型研究」という両者の不可分性が存在し(石川, 2012, pp. 16-17, 85)、この点において、コーパス言語学は田中(2015, p. 48)の指摘する研究に関するデータを研究者自らが調達するという側面が存在することが理解できる。

本論文は、この情報資源調達という視点から、教室発話コーパス構築に関係する諸課題について、研究目的と活用(第2節)、発話ローデータ収集方法(第3節)、発話の書き起こし(第4節)、付与する情報(タグ付け)のカテゴリー設定(第5節)について考察する。これらの考察をもとに、教室での発話コーパス資料収集・整理、そして活用に対する道筋を考察し、教室発話コーパス構築の礎としたい。以下、順に課題を列挙し、論を進める。

2. 教室発話コーパス開発の目的と活用

2.1 教室発話コーパス蓄積の意義

日本では明治維新以降様々な英語教授法による指導の効果に対する研究が行われてきたが(伊村, 2006, pp. 51-79)、1989年(平成元年)以降コミュニケーション型な授業が学習指導要領で求められ、2013年からは高校では原則英語を用いた英語教授法に転換が図られた。そしてこの流れは中学校に降りようとしている。将来、その流れは小学校まで到達するかもしれない。ここに、教室発話データ収集の必要性が存在し、教室発話コーパス蓄積の意味がある。

コミュニケーション型な授業に関する分析は、Spada & Fröhlich(1995)のCOLT Part AとPart Bを用いたものが国内にあるが(青木他, 2008; 石塚

他, 2005; 河合他, 2007など)、これらは主にPart Aを用いている。Part Bでは教師と生徒の発話の分析ができるように書き起こしを活用できる分析枠が提供されている。しかし、授業の一部分を数分単位で抽出して分析を行うtime-sampling procedure (Spada & Fröhlich, 1995, p. 66)という手法が推奨されており、full transcript 入手の困難さが想像出来る。もし、授業内の全発話が入手可能であれば、より多くの分析結果が得られることは想像に難くない。このPart Bの分析を数的に処理しやすくした方法が提案され、定量化された教師や生徒の談話特性を研究した一例が紹介されている(Katagiri & Kawai, 2015)。このデータ蓄積手法は、コーディング時にMicrosoft社のExcelのような表計算ソフトを使用するものである。もし、テキストデータの処理やテキストに付与された品詞タグなどのアノテーション処理との融合ができれば、Part AとPart Bの分析の融合によりprocess-product(教師のインプットと生徒の学習成果の関係)の分析へ貢献できる(Spada & Fröhlich, 1995, pp. 140-147)。

2.2 発話コーパスの活用可能性

蓄積された発話コーパスは様々な活用が考えられる。初期の教師教育においてはコーパス使用の利点がわかっており、教師が使用すべき言語的気づき(linguistic awareness)を助長することにコーパス応用領域の一つがあるとされる(O'Keeffe & Farr, 2003, p. 393)。しかし、彼らは英語母語話者であり、日本での英語教育の場では、ほとんどが英語の非母語話者となる点が異なる。藤原(2014, pp. 16-29)は「日本英語」というものを考えるにあたりEnglish as an International Language (EIL), World Englishes (WE), English as a Lingua Franca (ELF)という学際的領域の中で、「日本英語」は「学習者」と「使用者」の区分に分かれるとしている。教師は「国内外に活躍する英語使用者」の範疇には明示されていない。この点において、日本で収集された教室での発話データは「学習者コーパス」的側面を内在していると考えられ

る。

「日本英語」コーパスの利用可能性として、共時的研究、通時的研究、第二言語習得という3つの分野での利用可能性があり（藤原，2014，p. 75），その中で、教育面においてデータ駆動型学習（DDL）の応用について触れているが（同，p. 77），非母語話者教師の「学習者」という側面を認めるなら、教師教育にも応用可能であることが考えられる。

学習者コーパスを英語教育研究へ応用する場合の可能性として、和泉他（2004）は次の6点を示唆している、①英語学習プロセスの解明、②母語干渉、日本語による表現、③英語教員の養成・研修、④教授効果の予測、⑤辞書、教科書開発、⑥学習者による直接利用（和泉他，2004，pp. 17-20）。教師教育を考えた場合、教室発話コーパスには最初の4項目に関連があると考えられる。

以上のことから大別すると、教室発話コーパスは、①教師-生徒、生徒-生徒といった教室談話構造を中心とした包括的な英語授業の諸相解明、②教師教育における教室英語の指導などに活用可能性があると考えられる。

3. 教室発話データ収集

本節では、教室発話データの収集に関連する課題をデータ提供者と収集発話語数（3.1節）とデータ収集の技術的方法（3.2節）、映像情報（3.3節）の3点について考察する。

3.1 データ提供者と収集発話語数

データ提供者とは発話を収録する授業者とその生徒たちを指し、授業者とは多くの場合、英語非母語話者の日本人教師や日本人教師とともにティームティーチングを行うALT（外国人指導助手）などの英語母語話者を指す。場合によって、ALTなどは英語を第二言語として習得している場合もあるが、JETプログラム（The Japan Exchange Teaching Programme）等で雇用されていることを考えると、ここでは、英語母語話者

と同等とみなすことができる。また、生徒（小学校の場合は児童）は大学での授業を除いて未成年の場合がほとんどである。データ収集を行う校種によっては日本語（L1）の介入があることも予測できる。

授業ではCDやDVDなどの視聴覚教材を使用する場面もあり、そのデータも英語に触れているという側面からは貴重な授業「発話」データとなり得る。

発話データ収集に関しては、個人情報秘匿の観点から、同意書を交わす必要があり、場合によっては、授業が行われている学校や教育委員会の許可を取らなければならないことが考えられる。取得するデータ提供者の匿名性に関しては、APA Publicationマニュアル（第6版）には“... [researchers] are prohibited from disclosing ‘confidential, personally identifiable information concerning their ..., students, research participants, ...’”と記載されており同意書を取り交わす方法が秘匿性を守る一例として述べられている（APA, 2009, pp. 16-17）。秘匿性や同意書等がどのレベルで（教育委員会、学校長、授業者、生徒やその保護者など）処遇されるべきか当事者同士事前によく調整しておくことが必要である。

自作コーパス作成において書籍からのデータ収集で構築する場合には、収集データの母集団の問題や均衡性の確保、層化観点、サンプルサイズ等について検討を加えねばならないが（石川，2012，pp. 61-67），発話コーパスのデータ提供者の均衡性確保という点では地理的制約と同時にサンプルサイズも制約を受けざるを得なく、これらの課題に対しての対応が求められるであろう。

収集する発話語数であるが、教室発話データをコーパス化して公表されているものは多くはない。データの例として、高校の授業を収集したKatagiri & Kawai (2008, p. 1985)の研究では、日本人教師は1回の授業で平均1,969語³、ALTが2,064語の発話があり、Ohashi (2015, p. 37)が収集した中学校における日本人教師のデータから算出すると、平均854語というものがある。仮に10

万語収集を目標とした場合は、Katagiri & Kawai (2008) のデータから推測すると約50回、Ohashi (2015) のデータでは約117回分の授業データ収集が必要となることが予測される。目標とする発話のトークン数をよく検討してデータ収集の校種や収集回数・期間を設定する必要がある。

3.2 音声収録の技術的問題

教室で授業観察をしている時は、普通に聞こえる音声も、マイクロフォン（いわゆるマイク）を通して拾った音を2次的に聞くと、意外と聞き取りにくいことに気がつく。これは、収録した音に臨場感が欠けることによるものと想像がつく。モノスピーカー環境により収録した音声はその典型であろう。臨場感を上げるために、チャンネル数を増やし、例えば、ステレオ環境（2チャンネルで2つのスピーカ）を用いるが、さらに収録した音の臨場感を高めるため、5.1chサラウンドや場合によっては22.2chマルチチャンネル音響を相応数のスピーカとサブウーファで音聴者を囲む（5.1chなら5つのスピーカと1つのサブウーファ）という方法がある（鈴木他, 2013, pp. 66-67）。教室での音声収録時に臨場感を上げようとしてチャンネル数を増やすと、それ相応数のマイクが必要となるので、現実的には教室の集音はステレオマイク、一般的にはビデオカメラやICレコーダーなどは2つのステレオマイクで音を収録することとなる。さらに、発話収録時には、教室の音声を雑音（椅子や、教科書・プリント類の擦れる音、騒音など）を含め全て集音するので、録音されたものを聞くと、人間の耳で聞いている音環境とは比べ物にならないくらい臨場感は乏しくなり、その聞きとりにくさは書き起こす時の障壁となる。Katagiri & Kawai (2007) はビデオカメラ1台のシングルチャンネルにより小教室の収録をしている（図1）。

Spada & Fröhlich (1995, p.119)は2つのマイクを教室の端と端に設置したり、教師の襟にマイクをつけたりすることを勧めている。マイクを複数使用することによりマルチチャンネル環境を構

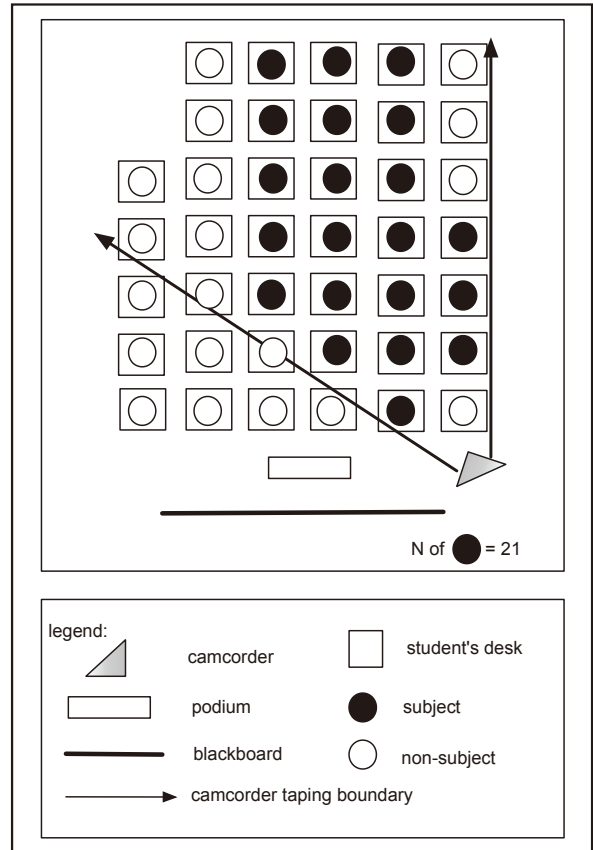


図1 シングルチャンネルの収録機材レイアウト例 (Katagiri & Kawai, 2007, p. 38より)

築できる。良質な臨場感を求めるのは発話コーパスの主目的でなく、教師や生徒や音声を収録することが発話コーパスの主目的なので、一つの方法として、教師や生徒の発話収録を個別に単一指向性マイク⁴等を使用してマルチチャンネル環境を作ることにより、書き起こし作業に耐えうる音声を収録できることが期待される。場合によっては専門家の援助も必要とされる (Spada & Fröhlich, 1995, p. 119)。これは生徒の発話が複数の組み合わせで教室内のあちこちで起こることから、2つのマイクだけでは、発話の多くまたは全てを収録することは不可能だからである。図2にマルチチャンネルの収録機材は一例を示す。

極端な場合、仮に40人のクラスで20組のペアワークを想定すると、双指向性マイク⁵を20個と教師用マイク1本、合計21本のマイクが必要となる。それらを音声と書き起こしをマルチチャンネルで保存するには、専門家の助けがどうしても必要となろう。現実問題としては、全体を収録する

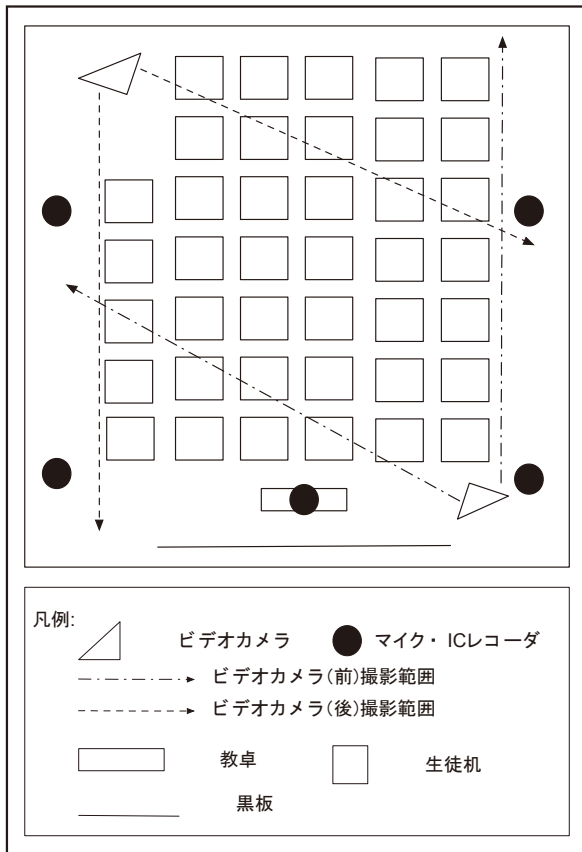


図2 マルチチャンネルの収録機材レイアウト例

カメラやICレコーダー内蔵のステレオマイクに加えて、教師用マイク1本と生徒用マイク数本くらいが妥当な数と考えられる。そしてそれらをPC上で走るソフト⁶を利用して、ミキシングをし、1つの発話ファイルを作るのである。そして、第4節で述べる書き起こし作業にかかるのである。

3.3 映像情報

音声を収録する時にビデオカメラを使用して収録する場合がある。また、ビデオカメラで授業収録をして、その音声を書き起こして研究に使用する場合もあるかもしれない。映像情報を使用して提供されている教室発話コーパスは極端に少ない。おそらく授業者や生徒たちの個人情報の秘匿性の問題があるためであると考えられる。しかし、マルチモーダルコーパス構築の分野においては、映像情報をもとに、身体動作セグメンテーション（音声波形などを元にした音素のセグメンテーションのようなもの）による身体動作アノテ

ション開発が始まっている（城他，2015）。映像情報は発話のテキストやメタ情報活用時に有益な情報となりうるので、映像とテキストなどを同期させて活用する方法が期待される。現段階ではIshizuka & Yorozuya (2014) のVideo on Demand system (VODCST)などが参考となる。このシステムでは、授業のビデオをCOLT Part A (Spada & Fröhlich, 1995)に基づいて、時間ごとにチャプターを区切り（映像のセグメンテーション）、コーディングするシステムを開発している。教室の映像情報にアノテーションが施されると、例えば、発話間に存在する休止（不自然な「間」）の意味などの分析が可能となるかもしれない。

4. 書き起こしと課題

音声ファイルができれば、それを聞いて書き起こすこととなる。英語コーパスを作成するのであれば、英語のみを入力することとなるが、中には、3.1節で述べたように日本語が混ざったり、発話そのものが、日本語のみだったりというものも存在する。次節で取り扱うタグに<japanese/>などの空タグを使用して、日本語の発話を無視するという書き起こし方も考えられるが、パラレルコーパス等への発展性を考えると、日本語もきちんと書き起こして記録の方が望ましいと考えられる。

日本語発話を書き起こす場合、ローマ字で日本語を書き起こすか、漢字仮名交じりの文字を使うかで汎用性に違いが出る。ローマ字表記にすると、文字コードで頭を悩ませることはないが、漢字仮名交じりとなると文字化けの問題などが出てきて、ファイルそのものの読み取りの信頼性が欠けることがある。著者の経験では、英語で入力していた時に、日本語を漢字仮名交じりで入力したためにそのファイルが日本語のキャラクターの文字コードが自動的に設定されてしまい、入力した英文全てが違う文字コードになり、文字化けを起こすキャラクターも出て、Perlスクリプトに記述した置換処理（s// /）などが、上手くいかずに苦

労したことがあった。石川（2012, pp. 68-72）はこの文字化けの問題を詳しく解説しており、自作コーパスで使用するエンコード形式の事前設定を勧めており、Unicode系を推奨している。また、坂上・仁科（2014, p. 133）は、文字コードはWindowsでもMac OS XでもUTF-8が望ましいとされているとしている。そして、改行コードはWindows上ではCR+LF, Mac OS X上ではLFで保存するのが望ましいとしている（同, p. 133）。

発話されている音声を書き起こしていると、分析の時に語（token）として認識させる必要があるため、以下の点などに気をつけなければならない。例えば、“fifty-four”などの数が発話された場合、“54”なのか“fifty-four”なのか、といった問題である。また、教科書の設問などを明示する発話に“Question number one.”などが考えられるが、これは“No. One.”などと入力しても良いのかという問題が生ずる。さらに、“I’m”などの縮約形は“I am”とすべきなのか、そのままが良いのかという問題も存在する。“He’s”は“He is”なのか“He has”なのかという問題も生ずる。このような「語認定」に対しては基準が存在せず、研究者自らが語認定基準を決定する必要がある（石川, 2012, p. 138）。表1に語認定基準の例を示す。

表1 標記語と認定語の例

カテゴリ	標記語例	語認定例
数 字	54	fifty-four
年 号	1984	nineteen eighty-four
数字羅列	3, 7, 1	three, seven, one
縮 約 形	I'm	I am
	He's	He is/has
度 量 衡	km	kilometer
	g	gram
	℃	degree Celsius

坂上・仁科（2014, pp. 137-141）は「テキストの整形」という問題について論じている。彼らはOCR（optical character recognition reader :

光学式文字読み取り装置）で読み取ったデータの語変換を確認することを推奨しているが、発話の書き起こしを考えると、書き起こした発話は他者による校正読みをする必要がある。

最後に書き起こしを保存するファイル名について言及する。坂上・仁科（2014, p. 135）は検索時のデータ情報取得という点から「ファイル名の統一」を勧めている。書き起こしは、時間のかかる作業ではあるが、継続的に続けていくとファイルが蓄積していくことは容易に想像がつく。3.1節で述べた10万語を想定した場合、Katagiri & Kawai（2008）の研究例のデータで50ファイル、Ohashi（2015）のデータで117ファイル出来上がることとなる。このため一定の書式でファイルを保存しておく、後に検索が必要となった場合や、AntConcなどでKWIC（key word in context）検索をする時には、参照しているファイルの情報が容易に分かることとなる（坂上・仁科, 2014, p. 135）。

5. タグ付与について

本節ではタグ付与について、形式（5.1節）と分類（5.2節）に分けて論ずる。

5.1 タグ形式

コーパスを構築するときタグ付与を行うことが一般的である。必要な情報を抽出する時のマーカー（標識）となるからである。タグ付与の形式としてはExtensible Mark-up Language（XML）を用いる例が多いようである。英語学習者コーパスのうち話し言葉コーパスを概観すると（投野他, 2013, p. 17）、The National Institute of Information and Communication Technology（NICT）Japanese Learners of English（JLE）Corpus（NICT）やThe Louvain International Database of Spoken English Interlanguage（LINDSEI）などもXML形式で表記している。よって、XML形式によるタグ付けは汎用性が高いと考えられる。

XML形式で表記されたデータは、「自己説明型データ(self-describing data)」(日本ユニテック, 2005, p. 19)なので、発話データがXMLタグ内の要素としてテキストデータを含むものであるのと同様に、タグにより示されるデータの意味する情報が重要となる。次節で述べる研究対象とする標識(タグ)が付与されていることにより研究が進められる。また、発話データの書き起こしは複数のファイルを複数の手で書き起こすので、共通のフォーマットで行われなければ、データとして活用は不可能となる。

5.2 タグの分類

LINDSEIではTranscription guidelinesが20項目に渡り示されており(Center for English Corpus Linguistics, 1995)、約16.6万語の母語話者と非母語話者の教師の教室発話を書き起こしたCorpus of Native and Non-native EFL Classroom Teacher Talk (CoCONNECT)などもこの指針に従っている。さらに、NICTの書き起こし・基本談話タグ付与ガイドラインではタグ付与形式に関し8分野23項目にわたり詳細に流れが示されており(和泉他, 2004, pp. 156-175)、エラータグ(47種類)や英語発話を日本語訳にしたback-translation(23種類)まで含めると相当数のタグを使用してコーパスが構築されている。

テキストベースの自作コーパスでは、品詞(POS)タグなどをつけた場合、プレーンテキストと分けて保存しておくべきとされ(石川, 2012, pp. 72-73)、NICTのように大規模な発話コーパス構築を将来的に視野に入れた場合、タグセットの設定は極めて重要な課題となる。Tono(2003)は学習者コーパス構築の際に、(a)言語(language-related criteria)、(b)タスク関連(task-related criteria)、(c)学習者関連(learner-related criteria)を考慮しなければならないとしている(Tono, 2003, p. 800)。本節では教室発話コーパス構築におけるタグセットの具体例を示すことが重要な課題の一つではあるが、前段で述べたように相当数のタグの存在を考慮すると、まずは

NICTのガイドラインにあるように、授業者のプロフィール(和泉他, 2004, p. 159)や談話情報(同, pp. 163-173)などタグのカテゴリ(分類)を決めてからタグセットを構築する必要がある。

以上のことから、汎用性のある教室発話コーパスを構築する上でファイル内に記載が必要と考えられるタグのカテゴリーは、①ファイルネーム(収録日、収録クラス校種などの属性を含む)、②授業者プロフィール(年齢、性別、TOEICなどの語学能力情報)、③話者(教師、生徒など)、④発話言語(L1, L2, mixなど)、⑤授業段階、⑥談話構造、⑦活動内容などがあげられる。また、「学習者コーパス」として考えた場合には、エラータグやPOS(品詞)タグなども汎用性向上に役立つと考えられる。表2に発話コーパス構築に用いるタグのカテゴリーの例を示す。

表2 発話コーパスタグ分類例

分類	属性	例
ファイル情報	形式、テキストエンコーディングなど	
授業者情報	校種、性別、経験年数、など	
話者情報	母語話者、非母語話者	
発話言語	母語、目標言語、混合	
授業段階	導入、活動、まとめ	
談話構造	発問、反応	
活動内容	個人、一斉、ペア、グループ	

6. 教室英語の発話コーパス設計例

本節では、タグセット例を述べ(第6.1節)、書き起こしのサンプルについて示す(第6.2節)。

6.1 発話コーパスタグセット例

表3に書き起こしに使用するタグセット例を示す。

表3 発話コーパスタグセット例 (テキスト部分)

分類	タグ例 ¹	意味
話者	<instructor>	教師
	<st>	生徒 (一人)
	<sts>	生徒 (複数人)
使用言語	<j>	日本語 (母語)
	<eng>	英語 (目標言語)
	<mix> ²	言語混合発話
音声教材	<cd>	音声CD
	<dvd>	DVD教材

注. ¹ タグは開始タグのみを表記した。実際は終了タグを併用する。終了タグは開始タグのタグネームの前にスラッシュ"/"を記載する。例えば、<instructor>という開始タグを使用する場合は</instructor>という終了タグを用いる。
² <mix>はその性質上、子要素に<j>と<eng>を含有する構造となる。

表4に書き起こし部分以外のタグセット例 (メタ情報) を示す。

表4 発話コーパスタグセット例 (メタ情報)

分類	タグ例 ¹	意味
授業展開 ²	<pre>	導入
	<main>	展開
	<consolidation>	まとめ
活動内容 ²	<pair_work>	ペア活動
	<group_work>	グループ活動
	<listening>	聞き取り活動
	<writing>	作文活動
非言語情報	<overlap>	発話の重なり
	<laughs>	笑い
	<pause>	沈黙

注. ¹ タグは開始タグのみを表記した。実際は終了タグを併用する。
² 授業展開や、活動内容は他にも様々な分類が考えられる。

この中では、活動内容や授業構成に関するタグ

付与が特に重要な情報となる。授業展開に関してはSinclair & Coulthard (1975) に教室の談話構造に関する詳細があるが、その中から大枠の考えを用いた。また、活動内容に関してはSpada & Fröhlich (1995) に詳しい、タスクについての分類があるが、最も単純な例と思われるものをあげた。

6.2 発話コーパスの書き起こし

本節では、表3で用いたタグセットの一部を用いて、書き起こしをした例を示す。図3にXML形式version="1.0"を用いたXML整形文書として書き起こした発話コーパスの見本を示す。紙面の大きさから表3と4で示した全てのタグを示すことは不可能ではあるが、表4にある授業の「導入 (pre)」、「展開(main)」、「まとめ(consolidation)」という展開ごとにより書き起こしが、入れ子式に表現されており、発話タグの中に、実際の発話部分がテキストで記載されている。

7. 結論

本節では、高汎用性教室英語の発話コーパス構築の課題と蓄積の方向性についてまとめる。

教室発話コーパス構築について、発話コーパス蓄積の意義と方法 (第1, 2節),そして諸課題 (第3, 4, 5節) について論じてきた。発話コーパスが書き起こしを必要とする性質上、音声収録の機会の確保とタグ付けは避けては通れない問題である。また、蓄積ということ考えると、書き起こしたテキストに標識付けするタグセット設定の方向性は大きな意味を持つ。第6節で述べたようなタグセットの設定が、話者レベルの発話抽出、使用言語の調査 (目標言語使用割合、発話数、語彙リストなど)、授業展開や、活動内容ごとの発話調査、非言語情報の描写などに結びつく。これらの情報は第2.2節で述べたように大きく2つの活用法がある。第一は研究者に対しての言語資料の提供である。書き起こされたデータは希少で特に国内の英語母語話者のデータは定量化され

```

<?xml version="1.0" encoding="utf-8"?>
<body>
  <pre>
    <instructor> <eng>Hello, everyone.</eng> </instructor>
    <sts> <eng>Hello, Mr. Shigeru.</eng> </sts>
    . . . . .
  </pre>
  <main>
    . . . . .
    <instructor> <eng>Okay, subject. What's this?</eng> </instructor>
    <sts> <eng>Music.</eng></sts>
    <instructor> <eng>How do you say music in English, In Japanese?</eng> </instructor>
    <sts> <j>音楽。 </j></sts>
    <sts> <eng>Science.</eng></sts>
    <instructor> <eng>What's this?</eng></instructor>
    <sts> <j>理科。 </j></sts>
    <sts> <mix><eng>PE, </eng> <j>体育。 </j></mix></sts>
    <sts> <mix><eng>Japanese, </eng> <j>国語。 </j> </mix></sts>
    <instructor> <eng>What's this?</eng></instructor>
    <sts> <mix> <eng>Art and Craft, </eng><j>図工。 </j> </mix></sts>
    <instructor> <eng>What's this?</eng></instructor>
    <sts> <mix> <eng>English, </eng> <j>英語。 </j> </mix></sts>
    . . . . .
    <instructor> <eng>What's this?</eng></instructor>
    <sts> <eng>Home economics.</eng></sts>
    <instructor> <eng>Oh, good. What's this?</eng></instructor>
    <sts> <j>家庭科。 </j></sts>
    <instructor> <mix><j>過去に勉強したね。 </j> <eng>subject</eng> <j>ね。 じゃあ、発音の
      練習だけしよう。 はい。 </j></mix></instructor>
    <sts> <eng>Music, Science, PE, Japanese.</eng></sts>
    <instructor> <j>もっとしっかり, はい。 </j></instructor>
    <sts> <eng>Art and Craft, English, Math, Social study.</eng></sts>
    <alt> <eng>Math.</eng></alt>
    <sts> <eng>Math, Social study, Home economics. </eng></sts>
    <sts> <eng>Social study, Math, English, Music, Science, PE, Japanese, Art and Craft. </eng>
      </sts>
    <instructor> <eng>Good, Okay.</eng></instructor>
    . . . . .
  </main>
  <consolidation>
    <instructor> <eng>Today, we learned names of the subjects. </eng></instructor>
    <instructor> <mix><j>次の</j> <eng>subject</eng> <j>は何。 </j></mix></instructor>
    <sts> <eng>Math. </eng></sts>
    <instructor> <eng>That's right. Very good. </eng></instructor>
    . . . . .
  </consolidation>
</body>

```

図3 XML整形文書にした教室発話書き起こし例

この例においては、要素の内容となっている発話部分を見やすくするために、ボールドとした。

た資料としての活用が期待される。ここから新たな知見が生み出されることが期待される。第二は現場の教師に対する示唆である。現場の教師たちが実際に自らの授業のデータを目の当たりにしたら、自らの授業を振り返り、改善していく上での様々な情報を得ることとなるであろう。また、いわゆる授業の達人たちの発話コーパス資料があれば、発話上の英語の使い方や、生徒とのインターアクションの技術の吸収などが可能となろう。授業の展開方法というメタ情報をタグセットの開発により、分析をすることもできるようになることも考えられる。

本論文で展開したことは教室発話コーパス構築の一方法にすぎず、教室発話データ収集から作業を始め、コーパス構築作業を進めていくうちに、修正が必要となる場合も生ずるであろう。今後は本論文で述べた発話コーパスの構築に取り組み、研究者や現場での教師たちに公開され、得られた知見が和泉（2013）の指摘する「英語教育の改善に貢献」（p. 187）できることを目標としたい。

注

1. http://www.lexically.net/downloads/version5/HTML/index.html?getting_started.htmに概要あり。
2. <http://www.laurenceanthony.net/software/antconc/>よりWindows版とMac OSXともに無償ダウンロード可能。使用方法に関しては著者自らが作成した解説ビデオがYouTubeで視聴できる (https://www.youtube.com/watch?v=O3ukHC3fyuc&index=1&list=PLiRIDpYmiC0Ta0-Hdvc1D7hG6dmiS_TZj)。
3. 本論文の中では便宜的に「語」という表記をしているが、厳密にはtoken（トークン：表記語）の意味で、type（タイプ：異なり後）とは違う意味である。
4. 正面方向からの感度が高い（鈴木他，2013，p. 54）。
5. 左右に感度を持たず、前後に感度を持ち、対談などに用いられる（鈴木他，2013，pp. 55-56）。
6. 無料のソフトでは“Audacity” (<http://sourceforge.net/projects/audacity/>) などがWindows, Mac, Linux版で使用できる。

謝 辞

本研究は科学研究費助成事業の助成を受けて行っている（研究種目：基盤研究（C）課題番号：15K02778）。

参考文献

- American Psychological Association. (2009). *Publication Manual of the American Psychological Association, 6th Edition*. Author.
- Center for English Corpus Linguistics. (1995). Louvain International Database of Spoken English Interlanguage (LINDSEI). [Online].
URL: <https://www.uclouvain.be/en-cecl-lindsei.html>.
- Ishizuka, H., & Yorozuya, R. (2014). “Collaborative VOD Platform for Classroom Observation.” *World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2014*(1): 2427-2432.
- Katagiri, N., & Kawai, G. (2007). Students react confidently to consistent oral classroom commands using controlled vocabulary. *Proceedings of JASELE, 2007*, 37-40.
- Katagiri, N., & Kawai, G. (2008). Lexical analyses of native and non-native English language instructor speech based on a six-month co-taught classroom video corpus. *Proceedings of Interspeech, 2008*, 1984-1987.
- Katagiri, N., & Kawai, G. (2015). Tabulating Transcripts and Coding on COLT Part B Scheme to Quantify Classroom Interaction Analysis Categories. *HELES JOURNAL*, 14: 23-41.
- Ohashi, Y. (2015). A Corpus-Based Study on the Relationship Between the Languages Used in Junior High School Classrooms and Learners' Uptake. *KATE Journal* 29. 29-42.
- O'Keeffe, A., & Farr, F. (2003). Using Language Corpora in Initial Teacher Education: Pedagogic Issues and Practical Applications. *TESOL Quarterly*, 37(3): 389-418. doi: 10.2307/3588397
- O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge Language Teaching Library.
- Spada, N., & Fröhlich, M. (1995). *COLT Communicative Orientation of Language Teaching Observation Scheme, Coding Conventions and Applications*. Sydney: Macquarie University.
- Tono, Y. (2003). Learner corpora: design, development

- and applications. *Proceedings of Corpus Linguistics 2003*. 800-809. Lancaster University.
- 青木千加子・石塚博規・横山吉樹・酒井優子・河合靖 (2008). 「COLT Part Bによるコミュニケーションを指向した英語プログラムの授業分析」『*Research bulletin of English teaching*』5 : 1-25.
- 赤野一郎・堀正広・投野由紀夫編著 (2014). 『英語教師のためのコーパス活用ガイド』大修館書店.
- 石川慎一郎 (2012). 『ベーシックコーパス言語学』ひつじ書房.
- 和泉絵美 (2013). 「学習者英会話データの分析」投野由紀夫・金子朝子・杉浦正利・和泉絵美編著 (2013). 『英語学習者コーパス活用ハンドブック』大修館書店 : 165-188.
- 石塚博規・横山吉樹・平田洋子・青木千加子・伊東優子・河合靖・高井収・新井良夫 (2005). 「COLT Part Aによるコミュニケーションを指向した英語プログラムの授業分析」『*Research bulletin of English teaching*』2 : 41-63.
- 和泉絵美・内元清貴・井佐原均編著 (2004). 『日本人1200人の英語スピーキングコーパス』アルク.
- 伊村元道 (2006). 『日本の英語教育200年』大修館書店.
- 河合靖・酒井優子・横山吉樹・石塚博規・青木千加子 (2007). 「COLT Part Bによる観察方法とその問題点」『*メディア・コミュニケーション研究 = Media and Communication Studies*』53 : 99-113.
- 阪上辰也・仁科恭徳 (2014). 「コーパスの作成」赤野一郎・堀正広・投野由紀夫編 (2014). 『英語教師のためのコーパス活用ガイド』大修館書店 : 130-143.
- 城綾実・牧野遼作・坊農真弓・高梨克哉・佐藤真一・宮尾祐介 (2015). 「異分野融合によるマルチモーダルコーパス設計 - 各種アノテーション方法と利用可能性について -」『*言語処理学会 第21回年次大会 発表論文集*』 : 561-564.
- 鈴木洋一・赤木正人・伊藤彰則・佐藤洋・荳木禎史・中村健太郎 (2013). 『音響学入門』コロナ社.
- 田中省作 (2015). 「タスク駆動型のコーパス構築と情報処理技術 - 「分類」を活用した新技術開発」『*英語コーパス研究*』第22号 : 47-61.
- 投野由紀夫・金子朝子・杉浦正利・和泉絵美編著 (2013). 『英語学習者コーパス活用ハンドブック』大修館書店.
- 日本ユニテックWeb技術研究グループ (2005). 『はじめて読むXML - 標準データ記述言語入門 -』アスキー.
- 藤原康弘 (2014). 『国際英語としての「日本英語」のコーパス研究』ひつじ書房.

(片桐 徳昭 旭川校准教授)

(大橋由紀子 ヤマザキ学園大学講師)