

科学研究費助成事業 研究成果報告書

平成 27 年 4 月 24 日現在

機関番号：14501

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24652120

研究課題名(和文) アジア圏学習者を対象とした音声・テキスト連動型英語発話コーパスの構築と分析

研究課題名(英文) Investigation of Problems in Development of Asian English Learners' Speech Corpus

研究代表者

石川 慎一郎 (Ishikawa, Shin'ichiro)

神戸大学・大学教育推進機構・教授

研究者番号：90320994

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：従来の学習者発話コーパス開発では、発話を書き起こしたトランスクリプトは公開されるものの、学習者の匿名性の確保の観点から、音声データそのものの公開はほとんどなされていなかった。音声データの公開を可能にするには、発話者の特定を回避するための音声変換が必要になる。そこで、様々な変換パラメータを実験し、発話者の性別ごとに、発話者が特定されず、かつ、音声として懲戒可能となるパラメータ候補を特定した。あわせて、当該処理を施した試行データを公開した。また、音声データとテキストデータをリンクして検索できるシステムのありかたについて基礎的な調査と分析を行った。

研究成果の概要(英文)：Many of the existing learner speech corpora include speech transcripts, but not audio files. In order to make the audio files publicly available, we need to consider how to prevent speakers' being identified by corpus users. In this research project, we investigated the possibility to apply the sound morphing technology for protection of speakers's anonymity and clarified how pitch and frequency of the original speech should be altered. We also investigated what kind of data retrieval system is needed for simultaneous analysis of transcripts and speeches.

研究分野：応用言語学

キーワード：学習者コーパス 発話 音声変換

1. 研究開始当初の背景

学習者コーパス研究は 1990 年代ごろから世界的に活発に行われており、その成果として、各種のコーパスが構築・公開されている。しかし、その大半は、書き言葉コーパスである。これに対し、話し言葉コーパスは非常に数が少ない。また、話し言葉コーパスを構築・公開する場合も、書き起こしたトランスクリプトの公開に留まる場合が一般的で、発話音声そのものはほとんど公開されない。今後、学習者の発話データの公開を考える場合、学習者の匿名性を保護するシステムと、発話とトランスクリプトを連動させた検索システムが必要になるとと思われる。

2. 研究の目的

将来的に大規模な学習者発話コーパスを公開することを念頭に置き、学習者の匿名性を保護する音声加工システムと、発話とトランスクリプトを連動させた検索システムについて、それらの開発に関して想定される基本的な問題点を整理し、実際の開発への道筋をつけることが本研究プロジェクトの狙いである。

3. 研究の方法

まず、音声加工については、既存の sound morphing 技術などを活用し、元の音声の周波数やピッチを変更することで、音声データとしての基本的特性を保持しつつ、発話者本人が特定できなくなるよう音調イメージを変更させる具体的な方法論、また、パラメタについて検討を行う。

次に、音声とテキストを連動させた検索システムのありようについて基礎的な調査と検討を行う。

4. 研究成果

音声加工については、既存の sound morphing 技術などを活用し、元の音声データについて、周波数やピッチをさまざまに変更する実験を行い、音声データとしての基本的特性を保持しつつ、発話者本人が特定できなくなるよう音調イメージを変更させるパラメタの候補を確定した。あわせて、こうした音声変換処理を自動的に実行するソフトウェアの設計を行った（実際の開発は、進行中の別の科研費資金により行った）。

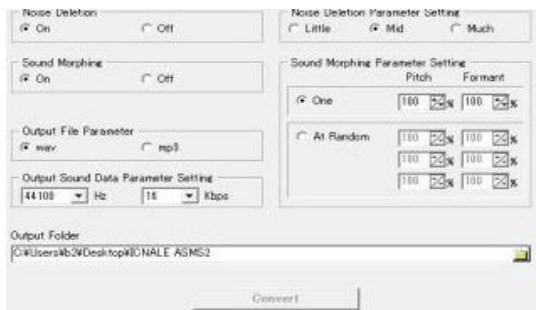


図 1：音声変換システム（デモ版画面）

さらに、試行として小規模のデータを収集し、実際に音声変換を加え、サンプルとして公開した。

また、音声とテキストを連動させた検索システムについて、既存の類似の検索インタフェースを調査し、求められるスペック、仕様について検討を行った。

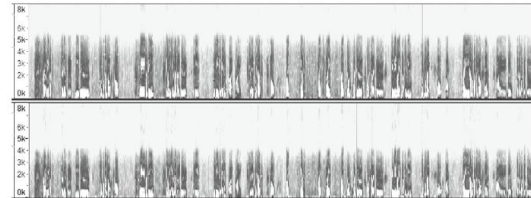


図 5-6. ASMS による加工前・加工後のスペクトログラム（周波数を可視化したもの）

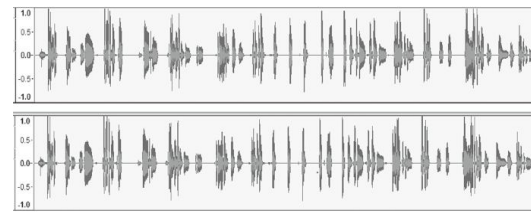


図 7-8. ASMS による加工前・加工後の音声波形

図 2：パラメタによる音声変換の様子（石川 2014「マルチモーダル型中間言語対照分析資料としての英語学習者話し言葉コーパスの開発」より転載）

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 16 件)

《査読あり》

Ishikawa Shin'ichiro / 2012/12/30 / POS Identification by L2 English Learners: A Study on Brain Activation / 単 / International Journal of Humanities and Management Sciences (IJHMS), 1(1) / 39-44

Ishikawa Shin'ichiro / 2013/8/20 / How to Incorporate Findings from Learner Corpus Studies in EFL Dictionaries From Misuse to Over/Underuse / 単 / Deny A. Kwary, Nur Wulan, & Lilla Musyahda (Eds.) Lexicography and Dictionaries in the Information Age: Selected Papers from the 8th ASIALEX International Conference / 138-144 /

Ishikawa Shin'ichiro / 2015/1/31 / A New Corpus-Based Methodology for Pedagogical Vocabulary Selection: Compilation of HEV1800 for Japanese High School Students / 単 / 中部地区英語教育学会紀要, 44 / 41-48 /

石川慎一郎 / 2015/3/25 / 学習者のグローバル意識の変化を観察する測定手法の開発と検証：コーパス言語学を応用した自由記述

型回答データの分析 / 単 / グローバル教育,
17 / 2-16 /

《査読なし》

Ishikawa Shin'ichiro / Qiang Wei /
2012/7/21 / The Quicker, the Better? L2
Proficiency, Reaction Time, and Brain
Activation / 共 / 電子情報通信学会技術研究
報告: 信学技報, 112(145) / 79-84 /

石川慎一郎 / 2012/11/1 / 英語教育におけ
る辞書 コーパス準拠と学習者志向 / 単
/ 外国語教育学会紀要, 15 / 87-94 /

増見 敦 / 石川慎一郎 / 2013/3/25 / 英語
多読活動に対する学習者の習熟度適性: 高校
における指導実践をふまえて / 共 / 神戸大
学国際コミュニケーションセンター論集, 9 /
19-30

石川慎一郎 / 2013/3/1 / ICNALE を用い
た中間言語対照分析研究入門: 日本人学習者
の「特徴語」を再考する / 単 / 英語教育 (大
修館書店), 61(13) / 64-66 /

石川慎一郎 / 2013/3/27 / 語彙多様性・文構
成度: 母語話者と学習者の区分基準は何か
決定木を用いた学習者コーパス分析 / 単
/ 統計数理研究所共同研究レポート 290: 計
量的言語データ処理法研究 / 107-124 /

石川慎一郎 / 2013/9/5 / テキスト関連属性
と助詞選択: 計量的アプローチに基づく探索
的研究 主語・主題を導く「は」と「が」を
めぐって / 単 / 第4回コーパス日本語学
ワークショップ予稿集 (国立国語研究所) /
213-222 /

石川慎一郎 / 2014/2/28 / 英語教育におけ
る異文化理解教育の課題と展望: 文化の定義
の再考と異文化理解教育の発達段階別多層
モデルの提案 / 単 / 東京外国語大学 世界言
語社会教育センター国際シンポジウム 2013 報
告書 / 105-119 /

石川慎一郎 / 2014/3/1 / マルチモーダル型
中間言語対照分析資料としての英語学習者
話し言葉コーパスの開発 ICNALE Spoken
プロジェクトのねらいと概要 - / 単 / 外国
語教育メディア学会 (LET) 九州沖縄支部紀
要, 14 / 1-20 /

石川慎一郎 / 2014/3/25 / コーパス文体論
の可能性: プロンテ姉妹の文体位相を例に /
単 / 文体論研究 (日本文体論学会), 60 /
121-143 /

石川慎一郎 / 2014/3/30 / 共起ネットワ
ーク分析をふまえた類義語定義の精緻化 コ
ーパス準拠日本語辞書の可能性 / 単 / 統
計数理研究所共同研究レポート 308: 言語変
種調査における統計処理 / 1-21 /

石川慎一郎 / 2015/3/23 / FROWN/FLOB
Corpus および BCCWJ データの再構成に基
づく英日対照言語研究用小説テキストデー
タセットの構築の試み English-Japanese
Modern Fiction Corpus (EJ-MoFiC) の概要
/ 単 / 統計数理研究所共同研究レポート
340: コーパス頻度データの統計的加工 /

1-18 /

増見 敦 / 石川慎一郎 / 2015/3/31 / 高校
生の英語力を推定する語彙テストの検討
サイズテストと速度テストの比較 / 共 /
神戸大学国際コミュニケーションセンター
論集, 11 /

〔学会発表〕(計 61 件 招待講演含)

《招待講演 (学会等)》

石川慎一郎 / 2012/7/8 / CALL から CALLS
へ: アジア圏英語学習者コーパス ICNALE プ
ロジェクトの目指すもの / 大学英語教育学
会 (JACET) 九州沖縄支部 第 25 回研究大会
記念 CALL 教育実践ワークショップ「クラウ
ドコンピューティング時代の英語教育の実
践事例」

石川慎一郎 / 2012/9/19 / アジア圏英語学
習者国際コーパスネットワーク ICNALE を用
いた域内学習者別言語使用傾向の抽出とそ
の教育的応用 / 北海道大学大学院メディ
ア・コミュニケーション研究院講演会 「計
量的言語研究の諸相」

石川慎一郎 / 2012/12/8 / Strategies for
Effective Academic Presentations / 神戸
大学女性研究者養成システム改革加速事業
「キャリアアップカフェ: 英語プレゼンテー
ションスキルアップセミナー」

石川慎一郎 / 2012/12/15 / 国際英語学習
者コーパス ICNALE に見る日本人英語学習者
の言語使用: 多元的 CIA による特性分析の
精緻化 / 第二言語ライティング研究の現代
的課題と解決のための将来構想 東アジア
からの発信 講演会

石川慎一郎 / 2013/3/5 / アジア圏国際英
語学習者コーパス ICNALE の設計と構築: 中
間言語対照分析の精緻化に向けて / シンポ
ジウム: 「第 2 言語習得研究のための学習者
コーパス」

石川慎一郎 / 2013/3/8 / ことばのなかの
文化・教室のなかの文化 / 国際シンポジウ
ム: 「外国語教育と異文化教育」

Ishikawa Shin'ichiro / 2013/3/23 / The
ICNALE and Sophisticated Contrastive
Interlanguage Analysis of Asian Learners
of English / Learner Corpus Studies in Asia
and the World (LCSAW) 2013 Conference

石川慎一郎 / 2013/6/1 / 中間言語研究へ
の新しいアプローチ: SLA とコーパスの架橋
/ 外国語教育メディア学会 (LET) 第 43 回
(2013 年度) 九州・沖縄支部研究大会

石川慎一郎 / 2013/9/24 / コーパスが明ら
かにすること: 言語教育研究への新しいア
プローチ ~ 語彙指導を中心に ~ / 津田塾大
学文学部講演会

石川慎一郎 / 2013/11/23 / データとして
の文体: コーパス文体論の可能性 / 第 104 回
日本文体論学会大会

石川慎一郎 / 2014/5/10 / 日本人英語学習
者を科学する 国際学習者コーパス ICNALE

の挑戦 / 日英言語文化学会第 44 回例会
Ishikawa Shin'ichiro / 2014/5/31 /
Learner Corpus Studies in Asia
-Challenge of the ICNALE- / Learner Corpus
Studies in Asia and the World (LCSAW) 2014
Conference

石川慎一郎 / 2014/10/18 / グローバル人
材教育の実化 コーパス言語学からの示唆
- / 英語教育セミナー「小・中・高・大の連
携によるグローバル人材育成と英語教育
の在り方」

石川慎一郎 / 2014/11/20 / コーパス言語
学入門: コーパスは小中高の英語教育をどう
変えるか / 東京国際大学言語コミュニケー
ション学部「コーパス言語学」特別講義

石川慎一郎 / 2014/11/23 / 日本語文法研
究のための数量データの扱い 統計検定と
はなにか / 日本語文法学会

石川慎一郎 / 2015/2/22 / 対照中間言語分
析の精緻化の可能性: ICNALE を用いた国際
比較・習熟度比較・産出モジュール比較
/ ラウンドテーブル「コーパス×学習者」

《その他研究発表等》
略

〔図書〕(計 10 件)

石川慎一郎 / 2012/4/5 / 『ベーシックコー
パス言語学』 / 単 / ひつじ書房 / 全 275p /

Joanna Szerszunowicz / Boguslaw
Nowowiejski / Yagi Katsumasa / Kanzaki
Takaaki / Kanzaki (Eds.) Ishikawa
Shin'ichiro / Wojciech Chlebda / 他 / 全 22
名 / 2012/6/15 / 『Research on Phraseology
in Europe and Asia: Focal Issues of
Phraseological Studies Vol.1』 / 共 /
University of Bialystok Publishing House,
Poland / 全 437p / 担当部 Ch. -2 L2
Learners' Use of English Words and
Phraseologies: Corpus-based Identification
of Lexical Proficiency Markers (pp.
389-410)

李 在鎬 / 石川慎一郎 / 砂川 有里子 /
2012/8/29 / 『日本語教育のためのコーパス調
査入門』 / 共 / くろしお出版 / 全 233p / 担
当部 Ch. 1,2,3,4,10(pp.1-62/ 139-152)

井上 亜依 / 神崎 高明 (編), 石川慎一
郎 / 澤田 治美 / 他 / 全 30 名 / 2012/9/15 /
『21 世紀英語研究の諸相: 言語と文化からの
視点』 / 共 / 開拓社 / 全 568p / 担当部 Ch.
9 英語教育: 「アジア圏同心円における英語
語彙使用」(pp.450-464)

Ishikawa Shin'ichiro (Ed.) , Sylviane
Granger / Ishikawa Shin'ichiro / 他 / 全 17
名 (第 1 著者) / 2013/3/23 / 『Learner
Corpus Studies in Asia and the World
Vol.1』 / 共 / Kobe University / 全 300p / 担
当部 Ch. 1-8 The ICNALE and
Sophisticated Contrastive Interlanguage
Analysis of Asian Learners of English

(pp.91-118)

Joanna Szerszunowicz / Boguslaw
Nowowiejski / Yagi Katsumasa / Kanzaki
Takaaki (Eds.), Ishikawa Shin'ichiro /
Wojciech Chlebda / 他 / 全 26 名 /
2013/12/31 / 『Research on Phraseology
Across Continents Vol.2』 / 共 / University
of Bialystok Publishing House, Poland / 全
441p / 担当部 Ch. 1-4 Phraseological
Discourse Markers of Reliability: From a
Viewpoint of World Englishes (pp. 97-114)

塩澤 正 (編), 赤野 一郎 / 井刈 幸
男 / 石川慎一郎 / 他 / 全 34 名 / 2014/3/30 /
『現代社会と英語 英語の多様性をみつめ
て』 / 共 / 金星堂 / 全 398p / 担当部
Ch.1-2 「テキストマイニングによる学校教
育目標としての『グローバル人材』の再定
義」(pp.15-27)

T. McEnery & A. Hardie (著), 石川慎一
郎 (訳) / 2014/5/30 / 『コーパス言語学: 手
法・理論・実践』 / 単 / ひつじ書房 / 全 412p
/

Ishikawa Shin'ichiro (Ed.), Tony
McEnery / Ishikawa Shin'ichiro / 他 / 全
32 名 (第 1 著者) / 2014/5/31 / 『Learner
Corpus Studies in Asia and the World
Vol.2』 / 共 / Kobe University / 全 451p / 担
当部 Ch. 2-2 Design of the
ICNALE-Spoken: A New Database for
Multi-modal Contrastive Interlanguage
Analysis (pp.63-76)

齊田智里 (編), 高田 智子 / 泉 恵美子
/ 石川慎一郎 / 他 / 全 105 名 / 2014/8/9 /
『英語教育学の今: 理論と実践の融合』 / 共
/ 全国英語教育学会 / 全 436p / 担当部 Ch.
6-4 「コーパス研究の動向と活用」(pp.
164-167)

〔その他〕

ホームページ等

http://language.sakura.ne.jp/s/kaken_icales.html

6 . 研究組織

(1) 研究代表者

石川慎一郎 (ISHIKAWA, Shin ' ichiro)
神戸大学・大学教育推進機構国際コミュニケー
ションセンター・教授
研究者番号: 90320994