

神戸大学 国際コミュニケーションセンター

石川 慎一郎 研究室

Dr. Shin Ishikawa, Kobe University



ICNALE-Spoken

科学研究費ICNALE-Sプロジェクト進捗報告ページ

Updated 2015/12/04

最新の進捗状況は[こちら](#)。

最新のデータセット（ICNALE-Spoken 1.0）および音声変換ソフトASMSのダウンロードは[こちら](#)。

→ [Outline of the ICNALE-S Project and Speech Data Collection Scheme](#)

→ [The Quick Guide of the ICNALE-S Project](#)

●関連する科学研究費補助金

NEW

（2015年度～2017年度：挑戦的萌芽）

アジア圏英語学習者の作文・発話の体系的修訂に基づく大規模校間コーパスの開発と分析（15K12909）

（2013年度～2015年度：基盤B）

[アジア圏英語学習者発話コーパスICNALE-Sの構築及び学習者発話の固有特性解明\(25284104\)](#)

（2012年度～2014年度：挑戦的萌芽）

[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（24652120）](#) [【終了（報告書）】](#)

2012年度に開始された挑戦的萌芽プロジェクト（～2014年度）の目標は、将来の発話コーパス開発を念頭に、音声変換システムや音声とテキストの連動システムの開発にかかる問題点を洗い出すことでした。同プロジェクトの初年度にあたる2012年度については、並走していた基盤B（ICNALE-W構築）と一体的に研究がすすめられました。この期間の進捗については、[ICNALE-Wの進捗報告ページ](#)を参照してください。

2013年度に、新規の基盤B（ICNALE-S構築）が採択されたことを受け、2013～2014年度においては、萌芽と新規の基盤Bを一体的に推進する体制を取りました。

2015年度からは校間に焦点を当てた新しいプロジェクトが開始されました。2015年度においては基盤Bと新規の萌芽を並走させる予定です。

※3つの科研の関連性

	2012年度	2013年度	2014年度	2015年度	2016年度	2017年度
2012萌芽	発話音声処理に係る基礎調査			***	***	***
2013基盤B	***	発話データ収集及びコーパス構築			***	***
2015萌芽	***	***	***	校間データの情報付与		

●石川研究室における学習者コーパス構築

石川研究室では、過去の科研プロジェクトにおいて、日中韓の英語学習者作文コーパスCorpus of English Essays Written by Asian University Students ([CEEAAUS](#)) (2007～2009年度)と、同コーパスを拡張した、アジア圏10か国の国際英語学習者作文コーパスInternational Corpus Network of Asian Learners of English ([ICNALE](#)) (2010～2012年度)を構築してきました。

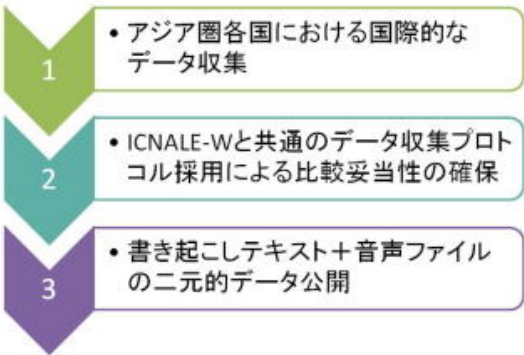
その後、2012年度より新たに発話コーパスの構築にかかる基礎研究を開始し、2013年度～2015年度の予定で、ICNALEの発話モジュールとなるICNALE-Sの開発プロジェクトを進行しています。現在、一般に公開されているアジア圏学習者の発話コーパスは存在せず、ICNALE-Sの公開は、アジア圏学習者の中間言語研究のいっそうの多元化に貢献すると考えられます。

2015年度に始まった新プロジェクトでは、収集済みのICNALE-SpokenおよびICNALE-Writtenのデータに対して、母語話者による校閲データを付与します。これにより、各国学習者の書き言葉・話し言葉産出におけるエラー研究が可能になります。



●ICNALE-Spokenの特徴

学習者発話コーパスICNALE-Spokenの基本特徴は下記の3点です。



開発済みのICNALE-Writtenとトピックをそろえることで、厳密なデザインのもとで、書き言葉と話し言葉の語彙や構文を比較することが可能になります。

●開発・公開されるリソース

本プロジェクトでは直接的にはコーパスの構築・公開を目的としますが、その準備段階として、発話データ自動収集システム (Automatic Speech Collection System : ASCS) と発話音声変換システム (Automatic Speech Morphing System : ASMS) の開発を行いました。ASMSは自由にご利用いただけます。

ICNALE-S 発話データ自動収集システムASCS 複数の国で同一環境（プロンプトの与え方、タイミング等）でデータを収集するため、プロジェクトでは、国際電話サービス、留守番電話機能、SQLデータベースを組み合わせた独自のシステム（Automatic Speech Collection	
---	--

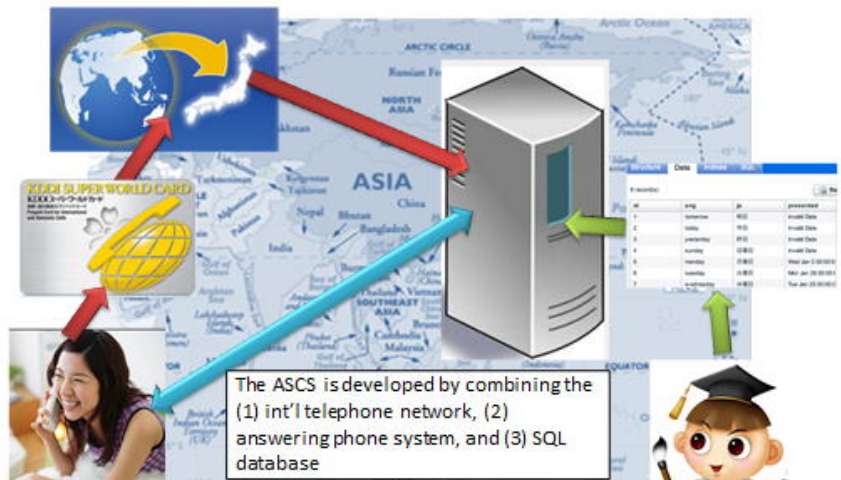
System: ASCS)を開発し、2013年10月より運用を開始しました。

システムは商用サービスを含むため、全面公開はできませんが、インストラクションやプロンプトについては申請いただければ提供させていただきます。

音声インストラクションシートは[こちら](#)

Automatic Speech Data Collection System

Anytime, anywhere, under the same condition



ICNALE-S 発話音声自動変換システムASMS

プロジェクトでは、聴覚印象を変化させ、発話者の匿名性を守る音声変換システム (Automatic Speech Morphing System)を開発しました。

ASMSの使用により、たとえば男性の発話を女性の発話のように、あるいは女性の発話を男性の発話のように変換することができます。

ASMS解説文書は[こちら](#)
[公開開始](#) (無償。要PW申請)しました。
asms_straight 低音質用
asms_world 高音質用

※発話者の匿名性の保護のため、ICNALE-Spokenの音声変換は、ASMSとは別のアルゴリズムを利用しており、詳細は非公開としています。

Parameter Setting

Noise Cut: ☒ On ☐ Off

Noise Recognition: Lower than dB

Morphing: ☐ Male to Female ☒ Female to Male ☐ Off

Pitch: %

Formant: %

Output Files

Save in: Browse

File name attachment:

Save as type:

ICNALE-Spoken V1.1

2015.11.01公開

発話音声ファイル (変換済) と書き起こしテキスト、発話者属性ファイルが含まれています。
無償公開していますがPWが必要です。

[公開中](#) (無償。要PW申請) です。

CHN_PTJ1_001_B1_1.mp3	CHN_PTJ1_001_B1_2.txt
CHN_PTJ1_002_B1_2.mp3	CHN_PTJ1_002_B1_2.txt
CHN_PTJ1_003_B1_1.mp3	CHN_PTJ1_003_B1_1.txt
CHN_PTJ1_004_B1_2.mp3	CHN_PTJ1_004_B1_2.txt
CHN_PTJ1_005_B1_1.mp3	CHN_PTJ1_005_B1_1.txt
CHN_PTJ1_006_B1_2.mp3	CHN_PTJ1_006_B1_2.txt
CHN_PTJ1_007_B1_1.mp3	CHN_PTJ1_007_B1_1.txt
CHN_PTJ1_008_B1_1.mp3	CHN_PTJ1_008_B1_1.txt
CHN_PTJ1_009_B1_2.mp3	CHN_PTJ1_009_B1_2.txt
CHN_PTJ1_010_B1_2.mp3	CHN_PTJ1_010_B1_2.txt

作業日誌 (2012年度～2015年度)

2015.12.4

●データの整理と公開

・PAKの追加データの処理を終了し、SW1.1をリリースしました。Online のデータも更新しました。Writtenとあわせ、3900人、1万件のデータが公開できました。プロジェク

トの主たる目的が無事に達成できたことになります。

- ICHELIに採択されました
- ・年末にマレーシアで開催される学会で発表します。

2015.11.12

(2015年10～11月の主な動き)

- データ整理と公開
 - ・ENS追加分, PAK追加分をデータに繰り込みました
 - ・Writtenを含めてICNALE-SW V1.0としてリリースしました
 - ・WrittenとSpokenでファイルコードを新規に統一しました
 - ・ENSデータは, 学生, 教師, その他の3元分類を導入しました
- データ収集の終了
 - ・PAKでの継続データ収集が10月末で終了しました。3年間のデータ収集はこれでいちおう終了したことになります。当初の予定を超えてデータを集めることができました。
- 新オンラインシステムの開発とリリース
旧システムより3年ぶりに大幅なアップデートを行いました。
 - ・Spokenデータの追加
 - ・オンラインでの音声ファイル再生システムの導入
 - ・Freq Graph機能の新規追加
 - ・He [v*] a [n*]のような柔軟な検索式の入力に対応しました
 - ・registrationシステムを簡略化しました
 - ・カウンタを付与しました
- 国際版ウェブサイトの修訂
 - ・SW1-0のリリースにあわせて内容を修訂しました
 - ・新オンラインシステムの導入にあわせて内容を修訂しました。

2015.09.03

(2015年9月の主な動き)

- データ整理
 - ・ICNALE-Spoken Version 2.0に向けたデータの整形処理を行っています。
 - ・ENSのデータ分類新基準の採用を検討（学生, 教師, 一般の3分割を新たに導入予定）
 - 検索システム改修
 - ・Spokenデータのアップロードに伴い, 大幅に検索システムを改修する予定で, 業者と作業に着手しました。
- <主な変更点（予定）>
- ・タグ検索への対応
- be・・・表記形beのみ
[be]・・・beの活用形すべて
play.[n*] 名詞のplayの表記形
[play].[n*] 名詞のplayの活用形全部
I [v*] a [n*]・・・ I have a pen/ I sing a son/ I played a gameなど全部

KWIC	Collocation	Wordlist	Freq Gra
Word(s)			
Case	POS HINT		
Participants	ADJECTIVE ADVERB NOUN MODAL DO HAVE ALL JJ JR JJS [ESL] : ☐ HKG ☐ PAK ☐ PHL ☐ SIN [ENL] : ☐ ENS_students ☐ ENS_others		

開発中の画面例

2015.08.31

(2015年6～8月の主な動き)

- データ収集業務
 - ・母語話者（ENS）の追加データ収集, テキスト化が終了しました。
 - ・日本人（JLE）の追加データ収集, テキスト化が終了しました。
 - ・パキスタンのみ, 継続収集しています。
- 関連研究発表・講演など
 - 2015/7/19 大学英語教育学会（JACET）東アジア英語教育研究会特別シンポ
 - 2015/7/25 熊本大学英語教育セミナー「グローバル化の進展に対応した英語力の目標と評価の在り方」
 - 2015/8/23 全国英語教育学会第41回熊本大会
 - 2015/8/25 The 8th Int'l Conference of English as a Lingua Franca (ELF8)
 - 2015/9/14 Learner Corpus Research 2015（予定）
- 関連の出版など（ICNALEのデータを使用した分析研究です）

- [Discourse and Interaction 8\(1\)](#) [Masaryk U, Czech Republic]
- [Procedia, Social and Behavioral Sciences, 192](#) [Elsevier]
- [Life beyond Dictionaries](#) [Cambridge Scholars Pub]
- [Procedia, Social and Behavioral Sciences, 198](#) [Elsevier]

●データ収集の最新状況

国名	担当者・進捗	収集済（人）	検証・公開済（人）
ENS	一次（2013年度）：50名 ・・・W5 Incにより50名分完了。 二次（2014年7月）：50名 ・・・W5 Incにより25名分完了。 三次（2015年5月～）：80名 ・・・OSPにより80名分（学生40名，成人40名）完了。	150 《収集終了》	75
CHN	一次（2013年度）：50名 ・・・国内収集により14名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により36名分完了。 二次（～2014年10月）：100名 ・・・国内収集により3名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により10名分完了。 ・・・LIU Rui (Jinggangshan University)氏により39名分完了。 ・・・ Xinghua Liu (Shanghai Jiao Tong University) 氏により40名分完了。 ・・・Enyu Feng (Southeast University, Chengxian College, Nanjin)氏により8名分完了。	150 《収集終了》	150
HKG	一次（2014年12月～2015年4月） ・・・ WONG Suzanne Shu Shan ／黃淑珊（Chinese University of Hong Kong）氏により50名分完了。	50 《収集終了》	50
IDN	一次（～2014年6月）：34名 ・・・Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)氏により34名分完了。 二次（～2014年11月）：66名 ・・・ Prihantoro (Universitas Diponegoro) 氏により57名分完了。 ・・・ Hanung Triyoko (STAIN Salatiga) 氏により9名分完了。	100 《収集終了》	100
JPN	一次（2013年度中）：50名 ・・・国内先行収集により50名分完了（※ASCS開発前につきCALL教室収集） 二次（2014年4月）：50名 ・・・国内収集により50名分完了。 三次（2014年11月～） ・・・国内収集により19名分完了。 四次（2015年4月～8月） ・・・国内収集により31名分完了。	150 《収集終了》	100
KOR	一次（2014年10月～2015年4月） ・・・Xinghua Liu氏により2名データ提供。 ・・・ Hye-Sook Kim (Konyang University) 氏により22名分完了。 ・・・Kyong-Ae Yu (Chung-Ang University)氏により44 (27+17)分完了。 ・・・Paik/Eunbi氏により33名分完了	100 《収集終了》	100
PAK	一次（～2014年3月）：100名 ・・・Muhammad Asim Mahmood (Government College University)氏によりCALL教室収集で100名分完了するがデータ品質の問題が見つかる。 二次（2014年8月） 現地訪問による直接収集を計画したが政情不安で中止。 三次（2014年10月～2015年2月）：50名 Moazzam Ali (University of Gujrat) 氏により完了。 四次（2015年3月～） Ali氏による50名の追加収集実施中。	旧データ100 新データ 現在60/ 目標100	旧データ50
PHL	一次（～2014年3月）：100名 Mari Karen L. Gabinete (De La Salle University-Manila)氏により100名分完了。	100 《収集終了》	100
SIN	一次（～2014年8月）：50名 ・・・ Vincent Ooi (National Singapore University)氏により50名分完了。	50 《収集終了》	50
THA	一次（2014年11月～）：50名 ・・・ Penpitcha Prakaiborisuth (Nics) (King Mongkut's University of Technology)氏により現在50名分完了。	50 《収集終了》	50
TWN	一次（～2014年3月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学) 氏により50名分完了。 二次（～2014年10月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学) 氏により50名分完了。	100 《収集終了》	50

2015.5.31

（2015年5月の主な動き）

- ELF2015（2015年9月@北京），LCR2015（2015年9月@ナイメーヘン）に採択されました。

- ICNALE Spoken Version 1.0を公開しました。

●データ収集の最新状況

国名	担当者・進捗	収集済（人）	検証・公開済（人）
	一次（2013年度）：50名 ・・・W5 Incにより50名分完了。		

ENS	二次（2014年7月）：50名 ・・・W5 Incにより25名分完了。 三次（2015年5月～）：80名 ・・・学生40名，成人40名を収集予定	75／目標150	75
CHN	一次（2013年度）：50名 ・・・国内収集により14名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により36名分完了。 二次（～2014年10月）：100名 ・・・国内収集により3名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により10名分完了。 ・・・LIU Rui (Jinggangshan University)氏により39名分完了。 ・・・ Xinghua Liu (Shanghai Jiao Tong University) 氏により40名分完了。 ・・・Enyu Feng (Southeast University, Chengxian College, Nanjin)氏により8名分完了。	150 《収集終了》	150
HKG	一次（2014年12月～2015年4月） ・・・ WONG Suzanne Shu Shan ／黃淑珊（Chinese University of Hong Kong）氏により50名分完了。	50 《収集終了》	50
IDN	一次（～2014年6月）：34名 ・・・Leonardi Lucky Kurniawan (Polytechnic of Ubya, Surabaya)氏により34名分完了。 二次（～2014年11月）：66名 ・・・ Prihantoro (Universitas Diponegoro) 氏により57名分完了。 ・・・ Hanung Triyoko (STAIN Salatiga) 氏により9名分完了。	100 《収集終了》	100
JPN	一次（2013年度中）：50名 ・・・国内先行収集により50名分完了（※ASCS開発前につきCALL教室収集） 二次（2014年4月）：50名 ・・・国内収集により50名分完了。 三次（2014年11月～） ・・・国内収集により19名分完了。 四次（2015年4月～） ・・・国内収集により15名分完了。	134／目標150	100
KOR	一次（2014年10月～2015年4月） ・・・Xinghua Liu氏により2名データ提供。 ・・・ Hye-Sook Kim (Konyang University) 氏により22名分完了。 ・・・Kyong-Ae Yu (Chung-Ang University)氏により44 (27+17)分完了。 ・・・Paik/Eunbi氏により33名分完了	100 《収集終了》	100
PAK	一次（～2014年3月）：100名 ・・・Muhammad Asim Mahmood (Government College University)氏によりCALL教室収集で100名分完了。 二次（2014年8月） 現地訪問による直接収集を計画したが政情不安で中止。 三次（2014年10月～）：50名 Moazzam Ali (University of Gujrat) 氏により28名分完了。	100 ＜差替予定＞ 28／目標50	50
PHL	一次（～2014年3月）：100名 Mari Karen L. Gabinete (De La Salle University-Manila)氏により100名分完了。	100 《収集終了》	100
SIN	一次（～2014年8月）：50名 ・・・ Vincent Ooi (National Singapore University)氏により50名分完了。	50 《収集終了》	50
THA	一次（2014年11月～）：50名 ・・・ Penpitcha Prakaiborisuth (Nics) (King Mongkut's University of Technology)氏により現在50名分完了。	50 《収集終了》	50
TWN	一次（～2014年3月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学) 氏により50名分完了。 二次（～2014年10月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学) 氏により50名分完了。	100 《収集終了》	50

2015.4.25

（2015年4月の主な動き）

●萌芽に採択されました。以後，メインのプロジェクトと連携しつつ，進めていきます。

●データ収集の最新状況

国名	担当者・進捗	収集済（人）	検証・公開済（人）
ENS	一次（2013年度）：50名 ・・・W5 Incにより50名分完了。 二次（2014年7月）：50名 ・・・W5 Incにより25名分完了。 三次（2015年4月～）：25名 ・・・直接収集により25名分追加を予定。	75	75
CHN	一次（2013年度）：50名 ・・・国内収集により14名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により36名分完了。 二次（～2014年10月）：100名 ・・・国内収集により3名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により10名分完了。 ・・・LIU Rui (Jinggangshan University)氏により39名分完了。 ・・・ Xinghua Liu (Shanghai Jiao Tong University) 氏により40名分完了。 ・・・Enyu Feng (Southeast University, Chengxian College, Nanjin)氏により8名分完了。	150 《収集終了》	50

HKG	一次（2014年12月～2015年4月） ・・・ WONG Suzanne Shu Shan ／黃淑珊（Chinese University of Hong Kong）氏により50名分完了。	50 《収集終了》	0
IDN	一次（～2014年6月）：34名 ・・・ Leonardi Lucky Kurniawan（Polytechnic of Ubaya, Surabaya）氏により34名分完了。 二次（～2014年11月）：66名 ・・・ Prihantoro （Universitas Diponegoro）氏により57名分完了。 ・・・ Hanung Triyoko （STAIN Salatiga）氏により9名分完了。	100 《収集終了》	0
JPN	一次（2013年度中）：50名 ・・・ 国内先行収集により50名分完了（※ASCS開発前につきCALL教室収集） 二次（2014年4月）：50名 ・・・ 国内収集により50名分完了。 三次（2014年11月～）：50名予定 ・・・ 国内収集により現在19/50名完了。 四次（2015年4月）～：31名予定 ・・・ 募集告知済み	100+ 19 (※+31目標)	100
KOR	一次（2014年10月～）：50名 ・・・ Xinghua Liu氏により2名データ提供。 ・・・ Kyong-Ae Yu（Chung-Ang University）氏により現在27名分完了。 ・・・ Hye-Sook Kim （ Konyang University ）氏により現在22名分完了。 二次（2015年3月）：50名 ・・・ Paik氏により24名分完了	74 (※+26目標)	0
PAK	一次（～2014年3月）：100名 ・・・ Muhammad Asim Mahmood（Government College University）氏によりCALL教室収集で100名分完了。 二次（2014年8月） 現地訪問による直接収集を計画したが政情不安で中止。 三次（2014年10月～）：50名 Moazzam Ali （University of Gujrat）氏により現在24名分完了。	100+24 ※ASCSを使用していない旧データを差替え予定	50
PHL	一次（～2014年3月）：100名 Mari Karen L. Gabinete（De La Salle University-Manila）氏により100名分完了。	100 《収集終了》	100
SIN	一次（～2014年8月）：50名 ・・・ Vincent Ooi （National Singapore University）氏により50名分完了。	50 《収集終了》	50
THA	一次（2014年11月～）：50名 ・・・ Penpitcha Prakaiborisuth （Nics）（King Mongkut's University of Technology）氏により現在50名分完了。	50 《収集終了》	0
TWN	一次（～2014年3月）：50名 ・・・ Ming Huei (John) Lin （Tamkang University 淡江大学）氏により50名分完了。 二次（～2014年10月）：50名 ・・・ Ming Huei (John) Lin （Tamkang University 淡江大学）氏により50名分完了。	100 《収集終了》	50

●書き起こし

香港, 韓国, 日本, タイ・・・作業中（5/10までに完了予定）

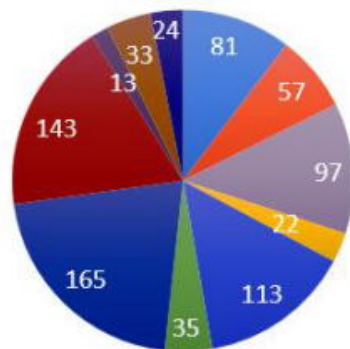
2015.3.18

（2015年1～3月の主な動き）

- 韓国, 香港でデータ収集を継続しました。
- 2月 ラウンドテーブル「学習者×コーパス」（筑波大東京キャンパス）でプロジェクトを紹介しました。
- 3月 スペインのバリャドリッド大学で開催された第7回国際コーパス学会で研究発表を行いました。
- 年度会計を締めました。

H26年度ICNALE-Sプロジェクト決算（※収支見合いの為、学内研究費もプロジェクトに合算充当）

（収入）基盤B：578万 挑戦的萌芽：94万 学内センター研究費（科研費取得支援金含）：81万、研究科研究費：50万
（科研関係支出）783万／803万



- LCSAWシンボ
- LCSAW出版
- 音声変換システム開発
- 英文校正
- 海外収集謝礼
- 書き起こし
- 国際電話代
- 学会旅費
- 学生学会旅費
- 研究図書
- 消耗品等(プリンタ他)

※ 2014年度費目別出費概況 (単位: 万円)

2014.12.25

(2014年12月の主な動き)

●ASMSシステム公開

●Baby 1.3公開 (インドネシアのデータが追加されました。)

●成果公開

2014年12月11-12日 第2回国際言語学・外国語教育学会において研究発表を行いました (アラブ首長国連邦)。また発表論文を投稿しました。

2014年12月 中部地区英語教育学会紀要投稿論文の採択が決定しました。

●現在進行中の収集・書き起こし (12月25日現在)

韓国, パキスタン, タイ, 香港で収集作業が継続中です。

●収集計画の進捗 (12月25日現在)

※収集計画: 1か国当たり50人 (200サンプル) 以上。主要国は100人 (400サンプル) を目指す。

国名	担当者・進捗	収集済 (人)	検証・公開済 (人)
ENS	一次 (2013年度) : 50名 ・ ・ ・ W5 Incにより50名分完了。 二次 (2014年7月) : 50名 ・ ・ ・ W5 Incにより25名分完了。 三次 (2015年4月〜) : 25名 ・ ・ ・ 直接収集により25名分追加を予定。	75	75
CHN	一次 (2013年度) : 50名 ・ ・ ・ 国内収集により14名分完了。 ・ ・ ・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により36名分完了。 二次 (〜2014年10月) : 100名 ・ ・ ・ 国内収集により3名分完了。 ・ ・ ・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により10名分完了。 ・ ・ ・ LIU Rui (Jinggangshan University)氏により39名分完了。 ・ ・ ・ Xinghua Liu (Shanghai Jiao Tong University) 氏により40名分完了。 ・ ・ ・ Enyu Feng (Southeast University, Chengxian College, Nanjin)氏により8名分完了。	150 《収集終了》	50
HKG	一次 (2014年12月開始予定) ・ ・ ・ WONG Suzanne Shu Shan / 黃淑珊 (Chinese University of Hong Kong) 氏により1名分完了。	1	0
IDN	一次 (〜2014年6月) : 34名 ・ ・ ・ Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)氏により34名分完了。 二次 (〜2014年11月) : 66名 ・ ・ ・ Prihantoro (Universitas Diponegoro) 氏により57名分完了。 ・ ・ ・ Hanung Triyoko (STAIN Salatiga) 氏により9名分完了。	100 《収集終了》	0
JPN	一次 (2013年度中) : 50名 ・ ・ ・ 国内先行収集により50名分完了 (※ASCS開発前につきCALL教室収集) 二次 (2014年4月) : 50名 ・ ・ ・ 国内収集により50名分完了。 三次 (2014年11月〜) : 50名予定 ・ ・ ・ 国内収集により現在13/50名完了。	100+ 13	100
KOR	一次 (2014年10月〜) : 100名 ・ ・ ・ Xinghua Liu氏により2名データ提供。 ・ ・ ・ Kyong-Ae Yu (Chung-Ang University)氏により現在22名分完了。 ・ ・ ・ Hye-Sook Kim (Konkang University) 氏により現在17名分完了。	41	0
	一次 (〜2014年3月) : 100名 ・ ・ ・ Muhammad Asim Mahmood (Government College University)氏によりCALL教室収集で100名分完了。		

PAK	二次（2014年8月） 現地訪問による直接収集を計画したが政情不安で中止。 三次（2014年10月～）：50名 Moazzam Ali （University of Gujrat）氏により現在17名分完了。	100+17	50
PHL	一次（～2014年3月）：100名 Mari Karen L. Gabinete (De La Salle University-Manila)氏により100名分完了。	100 《収集終了》	100
SIN	一次（～2014年8月）：50名 ・・・ Vincent Ooi (National Singapore University)氏により50名分完了。	50 《収集終了》	50
THA	一次（2014年11月～）：50名 ・・・ Penpicha Prakaiborisuth (Nics) (King Mongkut's University of Technology)氏により現在20名分完了。	20	0
TWN	一次（～2014年3月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学) 氏により50名分完了。 二次（～2014年10月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学) 氏により50名分完了。	100 《収集終了》	50

2014.12.5

(2014年11月の主な動き)

●ASMSシステム完成

ようやく音声変換のチューンにめどがつき、完成しました。検証後、一般公開の予定です。協力者より公開同意がとれた発話データについて、音調イメージを変化させることで、発話者プライバシーを守りながらデータを公開できます。

●成果公開

2014年11月 海外ジャーナルに論文を投稿しました（審査中）。

2014年11月 東京国際大学で学生・教員を対象にコーパス言語学の特別講義を行い、ICNALE-Spokenについて紹介を行いました。

2014年12月 第2回国際言語学・外国語教育学会において研究発表を行います（アラブ首長国連邦）。

●収集・書き起こし（12月2日現在）

台湾（50名追加分）、インドネシア（100名）の収集が完了。書き起こし開始。

●収集計画の進捗（12月2日現在）

国名	担当者・進捗	収集済（人）	検証・公開済（人）
ENS	一次（2013年度）：50名 ・・・W5 Incにより50名分完了。 二次（2014年7月）：50名 ・・・W5 Incにより25名分完了。 三次（2015年4月～）：25名 ・・・直接収集により25名分追加を予定。	75	75
CHN	一次（2013年度）：50名 ・・・国内収集により14名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により36名分完了。 二次（～2014年10月）：100名 ・・・国内収集により3名分完了。 ・・・ Yuanwen Lu (Shanghai Jiao Tong University) 氏により10名分完了。 ・・・LIU Rui (Jinggangshan University)氏により39名分完了。 ・・・ Xinghua Liu (Shanghai Jiao Tong University) 氏により40名分完了。 ・・・Enyu Feng (Southeast University, Chengxian College, Nanjin)氏により8名分完了。	150 《収集終了》	50
HKG	一次（2014年12月開始予定） ・・・ WONG Suzanne Shu Shan ／黃淑珊（Chinese University of Hong Kong）氏による収集予定。	0	0
IDN	一次（～2014年6月）：34名 ・・・Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)氏により34名分完了。 二次（～2014年11月）：66名 ・・・ Prihantoro (Universitas Diponegoro) 氏により57名分完了。 ・・・ Hanung Triyoko (STAIN Salatiga) 氏により9名分完了。	100 《収集終了》	0
JPN	一次（2013年度中）：50名 ・・・国内先行収集により50名分完了（※ASCS開発前につきCALL教室収集） 二次（2014年4月）：50名 ・・・国内収集により50名分完了。 三次（2014年11月～）：50名予定 ・・・国内収集により現在11/50名完了。	111	100
KOR	一次（2014年10月～）：100名 ・・・Kyong-Ae Yu (Chung-Ang University)氏により現在21名分完了。 ・・・ Hye-Sook Kim (Konyang University) 氏により現在17名分完了。	38	0
PAK	一次（～2014年3月）：100名 ・・・Muhammad Asim Mahmood (Government College University)氏によりCALL教室収集で100名分完了。 二次（2014年10月～）：50名 Moazzam Ali （University of Gujrat）氏により現在16名分完了。 ※2014年8月に 現地訪問による直接収集を計画したが政情不安で中止。	116	50
PHL	一次（～2014年3月）：100名 Mari Karen L. Gabinete (De La Salle University-Manila)氏により100名分完了。	100 《収集終了》	100
SIN	一次（～2014年8月）：50名	50	50

	・・・ Vincent Ooi (National Singapore University)氏により50名分完了。		
THA	一次（2014年11月～）：50名 ・・・ Penpicha Prakaiborisuth (Nics) (King Mongkut's University of Technology)氏により現在19名分完了。	19	0
TWN	一次（～2014年3月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学)氏により50名分完了。 二次（～2014年10月）：50名 ・・・ Ming Huei (John) Lin (Tamkang University 淡江大学)氏により50名分完了。	100	50

2014.0.10.20

(2014年9-10月の主な動き)

●シンガポール200本、母語話者100本のデータを追加したICNALE-Spoken Baby Version 1.2をリリース。

●ASMSシステム開発

音声変換の調整に難航しており、契約を切り替えてさらに1か月程度延長を行うことになりました。

●成果公開

2014年9月 チェコ共和国brno第6回言語学会で日本人学習者データ分析について中間成果の口頭発表 ※11月投稿予定。

2014年12月 海外学会で海外学習者データの分析成果を発表予定

●収集計画の見直しと修正（10月20日現在）

国名	担当者	公開人数 収集人数 目標人数	完了（予定）	進捗
ENS	W5 Inc	75 75 100	2014.08	一次収集分（50名分）・・・完了。 二次収集分（25名分）・・・完了。 ※英米でASCSシステムで直接に追加収集を予定。
CHN	【第1次：50】 ～2014.3 国内収集・・・14 Yuanwen Lu (Shanghai Jiao Tong University) ・・・36 【第2次：現在65】 2014.9～ 国内収集・・・3 Yuanwen Lu (Shanghai Jiao Tong University) ・・・10 LIU Rui (Jinggangshan University) ・・・10 Xinghua Liu (Shanghai Jiao Tong University) ・・・42 Enyu Feng (Southeast University, Chengxian College, Nanjin) ・・・0	50 115 150	2014.12	一次収集50人分（上海36名+国内14名）・・・収集・公開完了。 二次分・・・現在65人分 ※Liu Rui氏がプラス30, Xinghua Liu氏もしくは国内収集でプラス5→全体で150人データ为目标
HKG	—	0 0 50	—	2名に依頼。回答待ち。
IDN	【第1次：34】～2014.6 Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya) ・・・34 【第2次：現在56】 ～2014.9～ Prihantoro (Universitas Diponegoro) ・・・46 Hanung Triyoko (STAIN Salatiga) ・・・10	0 90 100	2013.12 2014.12	一次収集分・・・35名分収集完了。 7月より新リエゾン2名を指名。9月以降残り65名分の収集開始。 10/20現在あと10名不足
JPN	【第1次：50】（ASCS開発前試行収集分） Shin'ichiro Ishikawa (Kobe University) 【第2次：50】 Shin'ichiro Ishikawa (Kobe University) 【第3次：50】2014.11～ Shin'ichiro Ishikawa (Kobe University)	100 100 150	2014.12	一次・二次収集分（100名）は収集・検証・テキスト化・公開が完了。 新規にプラス50名の収集を開始。
KOR	2014.9～ Kyong-Ae Yu (Da Vinci College of General Education Chung-Ang University) Hye-Sook Kim (Dept. of English Language and Culture, Konyang University)	0/7/50 0/2/50	—	7/14 リエゾン決定。10月収集開始。
	（無償国際電話回線使用不可） 【第1次】・・・CALL教室で100 ～2014.3 Muhammad Asim Mahmood (Government College			（一次）2014.3までにCALL教室で100+α名分の取得が完了。データ

PAK	University, Faisalabad, Pakistan) 【第2次】 2014.7～ Moazzam Ali (University of Gujrat) ※当初現地収集を計画したが政情不安で中止。 ※その後有償国際電話での収集を計画したが現在頓挫中。	50 100? 100	2014.3	品質に問題が多いことから、使用に耐える50人分のみ選び、テキスト化完了。 (二次) 7月に新規リエゾンを委嘱。ASCSでの収集を目指す。
PHL	Mari Karen L. Gabinete (De La Salle University-Manila)	100 100 100	2014.1	100名分収集・検証・テキスト化・公開が完了。
SIN	Vincent Ooi (National Singapore University)	50 50 100?	2014.7	50名分の公開終了。サイズ拡張を計画。
THA	Sonthida Keyuravong (King Mongkut's University of Technology Thonburi)	0 0 100	2014.5	2014.5開始予定だったが、まだ開始されず。リエゾンに再確認(6/16)。秋に再開を確認(6/20)。10月現在、リエゾン未確定。
TWN	Ming Huei (John) Lin (Tamkang University 淡江大学)	50 70 100	2014.12	一次収集分(50名)収集・検証・テキスト化・公開完了。追加収集の可否をリエゾンに確認(6/16)。10～12月に追加収集の予定で合意(6.21)。現在20名分終了。

2014.8.31

(2014.8月の主な動き)

●シンガポールデータ収集完了

50名、200本のデータが収集されました。

●母語話者追加データ収集

25名、100本のデータが収集されました。

●シンガポール、母語話者追加データ書き起こし

75名、300本のデータの書き起こしが終了しました。今後検証を経てBabyコーパスに追加し、公開予定です。

●ASMSシステム開発

5/4、7/19、7/25、7/29、8/6、8/14、8/19、8/29と開発版のバージョンアップを進めてきました。8月中旬には、Avastというセキュリティソフトとのバッティングが見つかり、システムの基盤言語を修正するなど、開発は難航しました。8月30日現在、音声変換・インタフェースは完成し、ノイズ除去のチューニングを行っている段階です。9月上旬完成予定です。

●パキスタンでのデータ収集計画とその中止について（ご説明・ご報告）

ICNALEの収集対象国中、パキスタンのみ、KDDIが着払い式の国際電話サービスを提供しておらず、ASCSシステムでのデータ収集ができません。そこで、2013年度には、現地協力教員に依頼してCALL教室でデータを収集してもらいましたが、後日、その内容を検証すると、複数人がきわめて類似した内容で話しているなど、事前にプロンプトを知ったうえで原稿を書くなどの対策を行ったのではないかとこの疑念が生じました。

ゆえに、新しい協力者を委嘱し、プロジェクトリーダーである石川が現地を訪問し、直接の監督の下でデータ収集を行う計画を立てました。渡航計画の概要は下記の通りです。

日程：2014年8月11日～16日

引受先：パキスタン、グジュラート大学

活動：データ収集、特別講演会LCSAW2014 Summar Seminarの実施

渡航期間中には、2日間にわたってデータ収集を行うほか、現地研究者を対象としたプロジェクトのワークショップ講演も予定されていました。そこで、準備として、ICレコーダ30台を購入し、飛行機・現地でのガイドの手配なども整えていましたが、8月上旬より、同国においては、反政府派による大規模デモと、路面を閉鎖してデモ鎮圧を図る政権派の対立が高まり、グジュラート近郊のラホールでは、8名以上が殺害されることとなりました(8月9日付BBC News記事参照)。8月9日付で、現地協力教員より、学生の安全確保のために渡航期間中の大学閉鎖が決まったことが伝えられ、さらに日本の外務省から「パキスタン：国内政治団体による大規模抗議行進に関する注意喚起」が出されるに至りました。このため、やむなく、渡航とパキスタン現地でのデータ収集を中止することとなりました。以上、関係経費(航空券・地上手配費のキャンセル手数料)の執行の背景としてご報告申し上げます。

なお、同国では引き続き政情の不安定さが予測されることから、無償電話以外の地上電話回線を開く可能性について引き続き検討したいと考えています。

●関連研究発表

8月9-10日 全国英語教育学会：ICNALE-SpokenとICNALE-Writtenの語彙比較について

8月25-26日 統計数理研究所言語系共同研究発表会 S/W語彙比較の統計的モデリングの可能性について

●データ収集状況

インドネシア (Hanung Toriyoko氏担当)・・・20名分終了

2014.7.15

(2014.7月の主な動き)

●新規リエゾン確保

複数の国で9月～データ収取が新規に始まります。

2014.6.30

(2014.6月の主な動き)

●国際シンボ終了報告など

参加者への通知発送、[実施報告](#)の作成、国会図書館への蔵本、[神戸大学附属図書館電子アーカイブKERNELへの登録](#)ほかを行いました。また、会計の事後処理を行いました。

●音声変換システムASMSの開発計画の見直しについて

ベータ版開発までご協力していましたが、当初予想していた変換品質が達成できず、熟慮の結果、現在のシステム開発を中止し、新規のアルゴリズムでの開発に切り替えました。今後、仕様書の作成から再度やりなおすことになります。予想に対してくれぐれもですが、音声変換処理の問題点が明確化した点では成果があったと思います。→その後、請負社のほうで技術的な課題が解消され、仕様書を切りなおして再見積もりを依頼することになりました。夏までに完成予定です。

●収集計画の見直しと修正（7月3日現在）

国名	依頼	回答	担当者	人数	完了（予定）	進捗
ENS	済	○	W5 Inc	100	2014.08?	一次収集分（50名分）・・・収集・公開完了。 二次収集分・・・現在21名分収集のみ完了。
CHN	済	○	～2014.3 Yuanwen Lu (Shanghai Jiao Tong University) ／国内収集 2014.9～ LIU Rui (Jinggangshan University) Enyu Feng (Southeast University, Chengxian College, Nanjin) Xinghua Liu (Shanghai Jiao Tong University)	50	2014.03	一次収集50人分（上海36名＋国内14名）・・・収集・公開完了。 二次収集分・・・現在13名分収集のみ完了（上海10名＋国内3名）。 三次収集分・・・新規エゾンとして3名を委嘱。1人25人分。 ※必要数＝37 or 87人分
HKG	未	—	—	—	—	2名に依頼。回答待ち。
IDN	済	○	～2014.6 Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya) ～2014.9～ Prihantoro (Universitas Diponegoro) Hanung Triyoko (STAIN Salatiga)	100	2013.12 2014.12	一次収集分・・・35名分収集完了。 7月より新規エゾン2名を指名。9月以降残り65名分の収集開始。
JPN	済	○	Shin'ichiro Ishikawa (Kobe University)	100	2014.3	一次収集分（50名）は収集・検証・テキスト化・公開が完了。 二次収集分（＋50名）は収集・検証・テキスト化まで完了。現在公開準備中。
KOR	済	○	2014.9～ Hye-Sook Kim (Dept. of English Language and Culture, Konyang University)	—	—	7/14 リエゾン決定。
PAK	済	○	～2014.3 Muhammad Asim Mahmood (Government College University, Faisalabad, Pakistan) 2014.7～ Moazzam Ali (University of Gujrat)	100	2014.3	（一次）2014.3までにCALL教室で100＋α名分の取得が完了。データ品質に問題が多いことから、使用に耐える50人分のみ選び、テキスト化完了。 （二次）7月に新規リエゾンを委嘱。ASCSでの収集を目指す。
PHL	済	○	Mari Karen L. Gabinete (De La Salle University-Manila)	100	2014.1	100名分収集・検証・テキスト化・公開が完了。
SIN	済	○	Vincent Ooi (National Singapore University)	50	2014.7	2014.4より作業開始。現在、40名分の音声データを収集。
THA	済	○	Sonthida Keyuravong (King Mongkut's University of Technology Thonburi)	50	2014.5	2014.5開始予定だったが、まだ開始されず。リエゾンに再確認（6/16）。秋に再開を確認（6/20）。★新規リエゾンを依頼予定。
TWN	済	○	Ming Huei (John) Lin (Tamkang University 淡江大学)	50+50	2014.3	一次収集分（50名）収集・検証・テキスト化・公開完了。追加収集の可否をリエゾンに確認（6/16）。10～12月に追加収集の予定で合意（6.21）。

●ICNALE Baby 1.1公開（2014 June）

データを増量。350名、1400件になりました。ダウンロードは[こちら](#)。詳細は[こちら](#)。

2014.6.16

(2014.5月の主な動き)

●ICNALE Spoken Baby 1.0公開

250名、1000件のスピーチを収録したデータセットBaby 1.0が公開されました。総語数は11万語です。ダウンロードは[こちら](#)。



●邦訳書『概説コーパス言語学』（トニーマケナリー，アンドリューハーディー著）が刊行されました。[マケナリー氏と刊行書](#)（LCSAW2014会場にて・）。撮影：ひつじ書房松本社長）。神戸大学サイト内の[紹介ページ](#)。



●国際シンポLCSAW2014を開催しました。
当日は、ICNALEプロジェクトの統括責任者である石川による基調報告のほか、ランカスター大学(英国)のTony McEnery博士、Andrew Hardie博士、および、東京女子大学の金子朝子博士による講演、また、英国・オーストラリア・タイ・台湾・中国・トルコ・フィリピン(50音順)を含む内外の研究者による20本以上の研究発表が行われました。



●音声モーフィングシステム開発作業
パラメターの調整に難航しており，業者と数次にわたって相談を行いました。引き続き検証を進めていきます。



●日本とパキスタンの追加分（各50名）についてデータ整理と書き起こしが終了しました。これらはBaby2.0に入る予定です。

●日英言語文化学会において，「ICNALEプロジェクトの挑戦」のテーマで講演を行いました。



2014.4.8

(2014.4月の主な動き)

●自動音声モーフィングシステム仕様書確定・発注

本研究プロジェクトでは、データ収集システムASCS, 公開準備のためのデータ加工システムASMS, 公開用検索システムICNALE Online（既存のものを拡張）の3つのシステム開発を行います。このうち、ASMSの仕様については昨年度末から業者と意見交換をしていました。このたび正式に開発仕様が確定し、発注を行いました。

●音声収集システム修正

業者とミーティングを行い、1) 音声ファイルの差し替え（音量統制）、2) 指定タスク（2種類のタスクの1回目）以外のスキップ&再聴機能の除去、3) 検索サイトのデフォルト表示変更を行いました。

●タイでのデータ収集開始

4月第3週より収集を開始します。電話回線敷設、先方コーディネイター教員への連絡、トレーニングが終わりました。

●シンガポールでのデータ収集開始。

4月から5月にかけてデータ収集を行っています。

2014.3.31

(2014.1.14～年度末までの主な動き)

●LCSAW2014準備

- ・海外発表者へのビザ書類の作成・送付
- ・ウェブサイトの更新（発表予定者リスト公開）
- ・参加申し込み受け付け用フォーム作成

●関連学会発表

2月20日：筑波大学日本語辞書科で講演

3月7～9日：[Second Asia Pacific Corpus Linguistics Conference](#)で研究発表（題目：The ICNALE-Spoken: A new speech corpus of Asian learners of English: Possibility of spoken-written cross contrastive interlanguage analysis）



3月23日：関西大学英語学会で講演（題目：コーパス研究のこれまでとこれから）

3月29～30日：統計数理研究所言語系合同研究発表会で発表

●データ収集

各国50名（200サンプル）を先行して収集・処理し、5月のLCSAW2014で評価版として提供することとしました。ICNALE-Wとの整合性確保のため、マレーシアでの収集は対象外としました。

国名	依頼	回答	担当者	人数	完了（予定）	進捗
ENS	済	○	W5 Inc	50	2014.03	一次収集完了。
CHN	済	○	Yuanwen Lu (Shanghai Jiao Tong University)／国内収集	50	2014.03	上海分として36名、国内分として14名あわせて一次収集完了。
HKG	未	—	—	—	—	翌年度対応
IDN	済	○	Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)	100	2013.12	24名分終了、次年度繰越。
JPN	済	○	Shin'ichiro Ishikawa (Kobe University)	100	2014.3	一次分50名完了。 二次分は40名完了。次年度繰越。
KOR	未	—	—	—	—	検討中。
PAK	済	○	Muhammad Asim Mahmood (Government College University, Faisalabad, Pakistan)	100	2014.3	CALL教室で100+α名取得完。データ検証中。
PHL	済	○	Mari Karen L. Gabinete (De La Salle University-Manila)	100	2014.1	100名完了。書き起こし完了。
SIN	済	○	Vincent Ooi (National Singapore University)	50	2014.7	2014.4開始予定
THA	済	○	Sonthida Keyuravong (King Mongkut's University of Technology Thonburi)	50	2014.5	2014.5開始予定
TWN	済	○	Ming Huei (John) Lin (Tamkang University 淡江大学)	50	2014.3	一次分完了。処理済み。

●H25年度ICNALE-Sプロジェクト決算（※収支見合いの為、学内研究費もプロジェクトに合算充当）
（収入）基盤B：370万 挑戦的萌芽：110万 学内センター研究費（科研費取得支援金含）：66万
（繰越）22万
（支出）524万
（内訳）



「データ収集」には、参加者謝金と、データ収集システムの開発維持管理経費他を含みます。
科研費の出所が税金であることを常に認識し、今後とも、資金の効率的運用と研究成果の社会還元に努めます。

2014.1.14
(2013.11.20～2014.1.14までの主な動き)

●LCSAW201のCFP作成

科研費による国際シンポジウム [Learner Corpus Studies in Asia and the World2014](#) の概要がかたまり、CFPをウェブ上にアップし、linguistics, corpora, JSLA, コーパス学会等のMLで告知いたしました。

●オンライン検索サイト改修のための業者打ち合わせ

- ・品詞のみ指定検索対応（Modal+Vなど）
- ・書き手、話し手同時検索のための入力画面設定
- ・連動音声ファイル再生機能
- ・音声DL援助システム

※今年度の予算残金と調整しながら回収に着手していきます。

●インストラクション再聴プロトコル追加

Q6及びQ8の質問wavの後に新規wavを挿入（Do you like to listen to the instruction again? Yes/No）

- ・No="0"を押すと次へ進み20秒の考慮時間が始まる。
- ・Yes="1"を押すと再び質問が繰り返されます。
- ・"その他のボタン"を押す、または"何も押さない"だと、新規wav（もう一度聞か、次へ行くか）を繰り返す。

※2回この操作をすると切。

※これにより、1回目の試行で指示を正確に聞き取れないまま発話を始めてしまうエラーの発生を抑制できます。

●データ収集現状

2014.1.14現在

国名	依頼	回答	担当者	人数	完了予定	進捗
ENS	済	○	W5 Inc	100	2014.02	約20名分終了
CHN	済	○	Yuanwen Lu (Shanghai Jiao Tong University)	50	2014.03	約40名分終了
HKG	未	—	—	—	—	3月の香港理工大学での学会で募集
IDN	済	○	Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)	100	2013.12	24名分終了
JPN	済	○	Shin'ichiro Ishikawa (Kobe University)	100	2014.3	80名分終了
KOR	未	—	—	—	—	依頼者よりキャンセル
MYS	済	○	Rafidah Kamarudin (UNIVERSITI TEKNOLOGI MARA (UiTM) MALAYSIA)	50	2013.12	0名。未着手？
PAK	済	○	Muhammad Asim Mahmood (Government College University, Faisalabad, Pakistan)	100	2013.11	0名？国際電話不可。指定プロトコルにより、CALL教室で取得。詳細確認中。
PHL	済	○	Mari Karen L. Gabinete (De La Salle University-Manila)	100	2014.1	100名完了。書き起こし中
SIN	済	○	Vincent Ooi (National Singapore University)	50	?	0名 2013.11 開始？
THA	済	△	同僚と討議中	—	—	—
TWN	済	×	Ming Huei (John) Lin (Tamkang University 淡江大学)	—	—	—

2013.11.16～20 システム改修補正・修正シナリオ公開・韓国他協力者依頼・母語話者データ収集仕様書作成他

●システム改修補正

「0」でインストラクションをスキップ、「1」でインストラクションを再度聞く、改修時に機能が誤って削除されていたため、復活させました。

●修正シナリオ公開

修正後のシナリオを公開し、あわせて、作業用エクセルにも関連記述を追加しました。

[シナリオ](#) [エクセル](#)

●韓国他協力者依頼

前回のコーディネイターから協力が得られなかったため、ICNALEユーザーを中心に照会をかけ、韓国側のコーディネイターが新たに決まりました。また、タイからは調整中との返事です。

2013.11.20現在

国名	依頼	回答	担当者	人数	完了予定	進捗
ENS	未	○	相見積もりを済ませて発注指示待ち	50	2014.02	発注待ち
CHN	済	○	Yuanwen Lu (Shanghai Jiao Tong University)	50	2013.12	約20名分終了
HKG	未	—	—	—	—	—
IDN	済	○	Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)	100	2013.12	未着手？
JPN	済	○	Shin'ichiro Ishikawa (Kobe University)	100	2014.3	50人分収集・書き起こし完了。追加分50名を収集中。
KOR	済	○	Lee, Soyoung (Inha University)	—	2013.12	電話番号発給
MYS	済	○	Rafidah Kamarudin (UNIVERSITI TEKNOLOGI MARA (UiTM) MALAYSIA)	50	2013.12	未着手？
PAK	済	○	Muhammad Asim Mahmood (Government College University, Faisalabad, Pakistan)	100	2013.11	ワールドカード使用できないとの報告。対応検討中。
PHL	済	○	Mari Karen L. Gabinete (De La Salle University-Manila)	100	2013.10	現在60名完了
SIN	済	○	Vincent Ooi (National Singapore University)	50	?	2013.11 開始
THA	済	△	同僚と討議中	—	—	—
TWN	済	×	—	—	—	—

●母語話者データ収集開始

5社に見積もりを依頼。1社より見積もり有、1社見積もり準備中。3社返答無。

2013.11.15 システム改修・PHL/CHN/JPNデータ検証

●システム改修

収集情報一覧

全件A連番 完了A連番 全件C連番 記録#1 記録#2 記録#3 記録#4 履歴番号を表示 CSV形式で一覧をダウンロード

webDL	A連番	C連番	学生番号	記録番号	発信電話番号	着信時間	通話時間	Q2名	Q3国	Q4大学	Q5紹介	Q6P1J1
	A0005	C0005	11351136	001	非通知	2013/11/15 2:06:57	017					
	A0003	C0003	11321065	001	非通知	2013/11/15 0:44:43	537					
	A0002	C0002	11344903	001	非通知	2013/11/14 23:41:02	014					

- ・A番号（被験者連番）、C番号（最後までタスクを完了した被験者の連番）の2重管理を行いました。
- ・シナリオ微修正により、2回目の取り直し被験者が自己紹介セッションをスキップできるようにしました。
- ・シナリオ微修正により、大学名だけでなく国名も述べてもらうようになりました。
- ・全体のソート機能を大幅に強化しました。連番管理の原則は時間降順としました。

●データ検証

- ・3か国で収集済みデータの検証を行ない、フィリピンと中国のコーディネイターに現状報告を行いました。
- ・rejectデータの基準を明確にし、wrong（トピックを誤解した発話がある）とshort（60秒のタスク中、45秒以上話せていない発話がある）に整理しました。
- ・国内は、quoカードによる新しい謝金管理システムで収集を実施中です。
- ・フィリピンは50件の標準単位に到達しました。約100件の着信があり、内容を精査した結果最終の採択率は50%です。この割合を高める工夫を今後考えて行きます。

2013.11.03 新規データ収集国へのコンタクト開始、LCSAW2014の準備

前回のプロジェクトに参加したシンガポール、台湾、韓国、タイの教員に連絡を取り、今回のプロジェクトへの参加意向を尋ねました。協力が得られない場合、ICNALE-Wの利用者ネットワークを使ってコーディネイターを募集する予定です。

2013年3月のLCSAW2013に続き、LCSAW2014の開催を企画しています。次回の大会は、McEnergy & Hardie（2012）[Corpus Linguistics](#)（CUP）の日本語版（石川訳、ひつじ書房）の刊行記念を兼ねており、5月ごろの開催予定で、基調講演を依頼する著者と日程調整中です。

2013.11.02 第1期収集データの検証

システムの稼働効率を検証するため、先行してデータ収集を開始したフィリピンのデータを調査しました。

- ・調査期間内の総アクセス回数 N=87

（内訳）

- ・ドロップアウトA（10秒以内） N=34（電話番号がつかぬかどうかの確認等）
- ・ドロップアウトB（1～5分） N=10（個人情報登録まで、スピーチタスク未了）
- ・ドロップアウトC（6～9分） N=7（4つのスピーチタスクのどれかを未了）
- ・最後まで終了（10分～15分） N=36

（完了データのうち適格データの比率）

- ・データ検証の結果、アクセプトされたデータ N=22

※リジェクトデータの特徴・・・無音、60秒前に発話を終了、プロンプトを聞きもらす（とくに第1課題）

（経費）

- ・ワールドフォン（回線）経由 約2000円/1通話（10分の場合）
- ・ワールドカード経由 約500円/1通話

※およそ、回線とカードの使用率は50%ずつでした。カード使用が増えると経費を抑えることができます。

（対応）・コーディネイター教員に連絡し、下記の注意事項を伝達しました。

We really appreciate if you could tell new participants (or those who retry) to

- 1) record their speeches in a quiet room
- 2) speak sufficiently loudly
- 3) understand the topics correctly and talk about them
- 4) continue to talk for 60 seconds till they hear the beep sound

2013.11.01 システム業者への改修作業発注・国内データ収集開始

システム改修は納期は11月15日を予定しています。

また、国内での2次データの募集を開始しました。2次以降はASCSで収集します。

2013.10 18 システム業者との打ち合わせ@神戸大

システム改修に向け、ディスカッションを行いました。

主な改修要望点は下記です。

- ・発信者IDと識別番号の差を明確にする

- ・発信者電話番号の非表示に対応するため、シナリオの途中に国名入力質問を追加。
 - ・一括ダウンロード機能の実装
 - ・一括無音チェック機能の実装
 - ・データのソート、不要データの削除、ないし別画面表示
 - ・タイ、韓国、台湾でのデータ収集に向けた回線設置作業
 - ・busyを減らすため、受け元の電話番号の二重化（国によっていずれかの電話番号に誘導する）
-

2013.10. 5-6 英語コーパス学会第39回大会@東北大学で研究発表

発表題目：Contrastive Lexical Analysis of Spoken and Written Interlanguage Productions by JLE: A Study Based on the Newly Added Spoken Module of the ICNALE

ICNALE-Sの全体構想と試行データの分析結果について報告しました。当日のレジュメの修正版は[こちら](#)にあります。

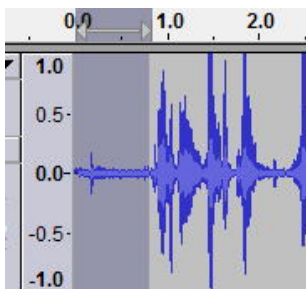


2013.9.27-10.3 データ収集システムの回線二重化と音声指示の修訂

KDDIには直接回線を引くサービスと、プリペイドカードのサービスがあり、後者が圧倒的に安価なため、後者でシステムを構築して収集を開始しましたが、国によっては学生の携帯電話からアクセスできない事案が発生し、回線との二重システムに更新することになりました。これに合わせて、指示文を短くして、全体の所要時間の若干の圧縮を行いました。

なお、（前回と同様の作業ですが）指示文はICレコーダで録音しますが、そのままではノイズが多く入ります。そこでソフトウェア（Audacity）を使ってアップロード前に音声ファイルの音質向上を行います。手順は下記。

1)無音部を範囲指定（下記の例では冒頭部）



2)ノイズプロファイルの取得

ノイズの除去

Dominic Mazzoni によるノイズの除去

ステップ 1

Audacity が除去すべき対象を判断するために、ノイズのみの区間(数秒間)を選択し、次にクリックすることでノイズプロファイルを獲得します:

ノイズプロファイルの取得(G)

ステップ 2

フィルタリングしたいオーディオをすべて選択し、除去の程度を選択し、次に'OK'をクリックすることでノイズを除去します。

ノイズ除去(dB)(d): 24

感度(dB)(S) 0.00

周波数平滑化(Hz)(e): 150

アタック/ディケイ時間(秒)(k): 0.15

ノイズ: ☒ 消去(m) ☐ 分離(I)

3) データ全体を範囲指定し、ノイズ除去を実施

4) その後、冒頭部と末尾部の無音部を削除

2013.9.25 第1次海外データ収集依頼 (toll-free call番号を伝達)

(フィリピン)

Mari Karen L. Gabinete (De La Salle University-Manila)

100 students

October 2013

(中国)

Yuanwen Lu (Shanghai Jiao Tong University)

50 students

Dec. 2013

(パキスタン)

Muhammad Asim Mahmood (Government College University,Faisalabad, Pakistan)

100 students

Nov. 2013

(マレーシア)

RAFIDAH KAMARUDIN (UNIVERSITI TEKNOLOGI MARA (UiTM) MALAYSIA)

50 students

Dec.2013

(インドネシア)

Leonardi Lucky KURNIAWAN (Polytechnic of Ubaya, University of Surabaya)

100 students

Dec.2013

なお新システムへの海外からの無料電話の掛け方は下記ようになります。

- 1) 各国から[KDDIへの専用アクセス番号](#)入力
- 2) 日本語のガイダンスが流れます
- 3) 国別カード番号の入力 (各国責任者に伝達済み)
- 4) カード番号の入力を間違えた場合は*を入力し、再度入力
- 5) 日本語のガイダンスが流れます (残金告知)
- 6) 国内のシステム専用電話番号入力 (※伝達済み)
- 7) 発話収集システム開始

2013.9.24 海外データ収集用[Excel Task Sheet](#)修正

- 1) 学部名称の記入をわかりやすく
- 2) 専門名の自由記入欄を追加

- 3) 個人情報→電話インタビュー→アンケート→単語テストの順番を、個人情報→アンケート→単語テスト→電話インタビューに変更（※電話インタビュー部分を後まわしできるように）
- 4) Excelのページ番号と連動させてprocedureの説明をよりわかりやすく変更

教員用データ収集ガイド修正（[修正版](#)）

- 1) 学生用ガイドと教員用ガイドを統合
- 2) 電話番号のアサイン依頼を追加（※busyを回避するため）
- 3) 集めるのはエクセルのみで音声データはサーバー自動蓄積されることを明記

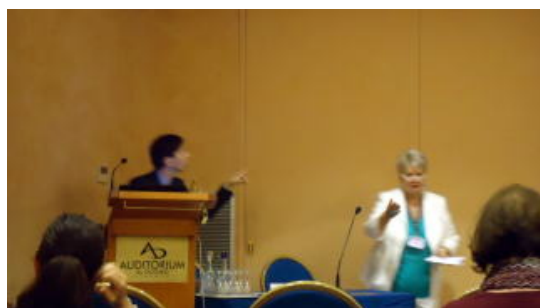
2013.9.24 海外電話番号発給完了（5か国）

スーパーワールドカードによる各国電話番号の発給が完了
あわせて業者に支持し、いたずら対策の自動回線強制切断機能を追加。

2013.9.20 業者との面会打ち合わせ

パキスタン以外の開通を確認。
しかし、料金シミュレーションの結果、莫大な費用が掛かることが判明し、全ての国をパキスタン同様のカード方式に切り替えることにする。

2013.9.11-18X INTERNATIONAL SCHOOL ON LEXICOGRAPHY: Life beyond Dictionaries @Auditorium Duomo Conference Centre, Florence, Italyで研究発表



ICNALE-Wの辞書開発への応用プロジェクトについて報告しました。

2013.9.5 業者との打ち合わせ

下記3点の機能をシステムに実装

1. ダウンロードファイル名の対応（連絡済）
発信者電話番号＋発行シーケンス番号＋質問番号. wav

2. 収集情報一覧表示 表示順の変更
一覧は発信者電話番号（昇順）、シーケンス番号（降順）に表示

3. 無音チェックボタン追加
ボタン押下により、収集wavファイルの無音チェックを実行。
結果は、再生ボタンの色替えにより表示します。
黒：チェック未実行 赤：無音 青：無音でない

ICNALE-Sデータ管理システム管理画面（2013.9現在）

[ホーム](#)
[収集情報一覧](#)
[パスワード変更](#)
[応答設定](#)
[お知らせメール設定](#)
[電話帳](#)
[ログアウト](#)

収集情報一覧

発信者ID	発信者電話番号	学生番号	自己評価	識別番号	着信時間	完了時間	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	無音
8	0783258855	0	0	0005	2013/09/20 17:18:57	2013/09/20 17:18:57									
4	0783258855	632	0	0004	2013/09/04 16:47:49	2013/09/04 16:49:01									
3	0783258855	665	2	0003	2013/09/04 14:01:12	2013/09/04 14:05:32									

自動でセッションごとの回答音声サーバに蓄積されていき、無音検出で自動で無音を見つけます。

2013.9.2 業者との打ち合わせ

インストファイル差替えプロトコルの決定

EcoDeco tool (wavを8000Hz 8bitに変換)

audacity-win2.0.3 (wavファイルをμ-Law形式に変換)※ITU-Tで規格化されている音声符号化における圧縮方式の1つ。このμ-law方式では、14bitのリニアPCMデータを8ビットに圧縮する。『ASCIIデジタル用語辞典』

2013.8.30 業者との打ち合わせ

パキスタンについて、100人からデータを集める場合、スーパーワールドカードによるプリペイド方式だと77000円、クレジットカードによるコレクトコールの場合は353,000円程度が想定されることから前者を採用。

2013.8.27 業者との打ち合わせ会議

無音検知システム他について意見交換を行う

ワールドフリーフォンについて、使用実態がない限り経費が発生しないことを確認し、早急に手配することを決定。パキスタンのみスーパーワールドカードシステムを併用。インスト用のwaveファイル差替えマニュアル作成

2013.8.16 仮システムテスト

途中でシステムがハングするエラーが発生し、再度検証にかけます。

2013.8.6 パキスタン問題への対応会議

使用予定のKDDIのワールドフリーフォンサービスがパキスタンに対応していないことがわかり、海外間での転送システムを組み込む必要が生じ、対応中です。

2013.8.5 音声収集システム用差替えwaveファイル作成・送付

全体のシナリオ修正に伴い新規に必要な音声ファイルを作成しました。

2013.7.31 音声収集システム開発打ち合わせ会議@神戸大石川研究室

業者との3回目の打ち合わせ会議です。仮稼働システムについて、詳細な動作チェックを行いました。

変更点

1)冒頭のプッシュ回線識別コード入力の削除。当初案のinstの1番と2番を統合しました。

Welcome to the ICNALE speech data collection system.

This recording will last for approximately six minutes. If you stop performing the tasks halfway, you will have to do them all over again from the beginning. You are advised to complete all the tasks within a single session. By pressing zero [0] on your phone, you can listen to the instructions again. And by pressing one [1] you can skip the instruction. Now, kindly respond to the following ten questions.

- 1)学生番号入力カインストの変更。数字以外のコードを飛ばす点がよりわかりやすくなるよう指示文を修正。
Question No. 1: After the beep, please key in your student number provided by your college using the keys on your phone. If your student number includes the alphabets or symbols, please ignore them and key in just the numbers.
- 3)学生番号が正しく入力されたことを確認させるプロトコルを追加。
Your ID number is XXXX...
If this is correct, press 8 on your phone. If not correct, press 9.
- 4)作業完了コードをメモ書きさせたことを確認させるプロトコルを追加。
Did you make a note of the number? If yes, press 8 on your phone. If no, press 9.
- 5)システムエラーにより、再録音を依頼する場合のメッセージを追加。
There was an error in recording. Sorry, but please respond once more again after the beep.
- ※8/1～5にかけて再録音を行い、音声ファイル処理（前後無音部カット）して業者に送付しました。なお、このプロジェクトでは、国際英語という観点をふまえ、インスト音声は母語話者に録音依頼するのではなく、あえてプロジェクトリーダー自身が担当しています。
※あわせて、[プリントアウト版のインストガイド](#)をupdateしました。被験者はインストガイドを目視確認しながら音声応答します。

2013.7.31 海外データ収集進捗状況まとめ（2013年7月末現在）

前回プロジェクトの協力者を中心に、7月上旬に、新プロジェクトへの参加を呼びかけました。本日現在の依頼・回答状況は下記の通りです。年度ごとの予算管理の関係で（※プロジェクト総額内であっても1年ごとに使える上限が決まっている）、一次依頼国、二次依頼国・・・のように逐次的に依頼を進めています。
データ件数については、1人の被験者が2種4ヶの発話産出を行いますので、100人の場合は200種、400件のデータとなります。
ICNALE-Wにはなかった国として、マレーシアを新規に追加しました。アジア圏の重要なESL国であるマレーシアのデータは、コーパスの地域的な均衡性をさらに高めるものと期待されます。

国名	依頼	回答	担当者	人数	完了予定	進捗
ENS	未	—	—	—	—	—
CHN	済	○	Yuanwen Lu (Shanghai Jiao Tong University)	50	2013.12	—
HKG	未	—	—	—	—	—
IDN	済	○	Leonardi Lucky Kurniawan (Polytechnic of Ubaya, Surabaya)	100	2013.12	—
JPN	済	○	Shin'ichiro Ishikawa (Kobe University)	100	2014.3	50人分の収集・書き起こし完了
KOR	済	×	—	—	—	—
MYS	済	○	Rafidah Kamarudin (UNIVERSITI TEKNOLOGI MARA (UiTM) MALAYSIA)	50	2013.12	—
PAK	済	○	Muhammad Asim Mahmood (Government College University, Faisalabad, Pakistan)	100	2013.11	—
PHL	済	○	Mari Karen L. Gabinete (De La Salle University-Manila)	100	2013.10	—
SIN	未	—	—	—	—	—
THA	済	△	同僚と討議中	—	—	—
TWN	済	×	—	—	—	—

2013.7.20 JACET東アジア英語教育研究会特別シンポジウム@西南学院大学

表記シンポを企画し研究室より3名が出席して発表を行いました。

第133回JACET東アジア英語教育研究会
日 時：平成25年7月20日（土）14:30～17:30
会 場：西南学院大学 1号館 205教室
発表題目：シンポジウム「アジアの言語教育研究へのコーパスの貢献—英語教育と日本語教育を例に—」

- 言語教育研究とコーパス（神戸国際文化学研究所・石川慎一郎）
言語教育研究へのコーパスの可能性について導入的議論を行い、3発表の概要を紹介する。
- アジア圏英語学習者のly副詞使用（神戸国際文化学研究所博士前期課程・川村晃市）
各種の母語話者コーパスを比較検討することにより、多様なly副詞群の中から、汎用的で教育的に重要なものを特定する。そのうえで、アジア圏学習者の作文コーパスを検証し、学習者のly副詞使用の課題を特定する。
- アジア圏日本語教育の観点から見たスル型漢語動詞—内部構成に着目して—（神戸国際文化学研究所博士後期課程・李楓）
スル型漢語動詞（例：努力する、破壊する、工夫する）は、日本語の語彙の中で重要な位置を占めるが、その使用の実情や意味機能の詳細は先行研究においても完全には明らかにされておらず、漢字文化を共有するアジア圏の日本語教育の現場でもほとんど扱われていない。本発表では、現代日本語の多面的変種において汎用的に出現するスル型漢語動詞を特定した上で、その内部構成パターンを計量的に特定する。
- 「校閲量単位」から見たアジア圏学習者の中間言語使用（石川慎一郎）
アジア圏国際英語学習者コーパスICNALEに新規に付加された「ブルーフモジュール」の概要を紹介し、ブルーフリーダーによる「校閲量単位」が、書き手の習熟度尺度になりうるかどうかを検証する。



2013.7.20 ICNALE-P第1次データ公開

本科研では、既存のICNALE-Wと対比可能な発話データベースとしてICNALE-Sを構築することに加え、公表済みのICNALE-Wの各種アドオンモジュールの開発も並行的に行っています。その1つにICNALE-P（Proofreading Module）があります。
従来の学習者コーパス研究ではエラータグを付与することがありましたが、明白なエラーを除くと、判断に迷うものが無数に存在します。一定以上の習熟度の学習者について言
うと、コミュニケーション上より大きな問題になるのは、個別的にタグ付けできるような局所的で自明のエラーよりも、間違いではないが不自然さが感じられる数々のL2運用で
しょう。

そこで、ICNALE-Pでは、海外一流学術誌への投稿支援を専門とする職業校閲者（10年以上の経験を積んだ母語話者）に学生データの全面ブルーフを依頼する形で、本格的な校
閲モジュールの開発を行っています。

第1次公開版（ICNALE-P_V201307）に含まれるのは、下記のデータです。なお、目下のところ、トピックはすべてアルバイト（PTJ）となっています。

- ◎日本人A2レベル学習者20名
 - ◎日本人B1_2レベル学習者20名
 - ◎中国人A2レベル学習者20名
 - ◎中国人B1_2レベル学習者20名
 - ◎韓国人A2レベル学習者20名
 - ◎韓国人B1_2レベル学習者20名
- 以上合計120名

I ~~believe~~think it is important for college students to have ~~a~~ part-time jobs. I
can offer ~~have~~ three reasons to support my opinionwhy it is so important. First-
of all, part-time jobs can help studentsin order to earnmaintain their
livelihoods. Many students are struggling financially. There is many students
who under difficult. It is impossible to stay energetic and remain alive without
foodWe can't live if you do not eat something. If there is no energy, no life
can live. Second, students can improve their reason is conversation skills
their part time jobs. Social skills are many. To is many

校閲データ実物例

たとえば、同習熟度で比較すると、日本人学習者の作文に対する校閲量は中韓に比べて有意に少なくなっています。しかし、データを仔細に観察すると、日本人学習者は典型的
なテンプレートの枠組みの中に無難な内容を盛り込んで作文する傾向が見えてきます。ICNALE-Pは「文法力と表現力の乖離」といった従来あまり触れられてこなかった観点か
らの分析を可能にするものです。第1次データの分析結果は、7月20日のJACETシンポジウムで報告予定です。石川研究室では、今後も、ブルーフデータの拡充と公開に務めて
いきます。

第1次公開版データ（修正前オリジナルテキストファイル、修正後テキストファイル、修正過程記録ワードの3種が同梱されています）のダウンロードは[こちら](#)（無償ですがパ
スワード取得のために登録が必要です）。

2013.7.13 海外データ取得用プロトコル・海外協力研究者用プロジェクトガイド作成

海外での作業プロトコル第1次案と協力研究者用のプロジェクトガイドを策定しました。
[こちら](#)

2013.7.12 音声データ収集システム開発業者打ち合わせ（3回目）

シナリオフローとインストラクション音声を当方より提供しました。
業者側からは、無音部を検出し、削除するシステムの提案がありました。
また、タスク終了後に、乱数で4桁の数字を発生させ、タスクの完了コードとして音声発給するシステムの組み込みを依頼しました。これにより、被験者の作業完了が確認しや
すくなると考えられます。

2013.7.11 音声データの変換実験

このプロジェクトでは、個人情報の匿名性を保持しながら音声公開を行なうためのプロトコルならびに自動処理システムの作成を行います。下記は処理実験の資料です。

■オリジナル

↓

■ピッチ・フォルマント変換（ピッチ+20.0%, フォルマント+12.6%）

※性別印象などが変更され、話者特定が困難になります

↓

■ノイズ自動除去（1000Hz閾値）

※各種のフィルタがありますが、自動化ではすべてのノイズは除去されません。

↓

■手作業によるノイズ削除

※手作業を加えるとかなり音声ファイルのクオリティを上げることができます。

プロジェクトでは、個々のファイルを解析して処理パラメタをアドホックに決定し、一連の処理を自動化するシステム開発を計画しており、実現すれば、学習者音声コーパス研究における音声データの公開と共有がさらに進むきっかけになると思います。また、ノイズフィルタリングを自動化できれば、書き起こしの一次処理を自動化する可能性も広がります。

2013.7.1～12 第2回国内発話データ収集実験

2回目の実験を行いました。

2013.7.8 音声収集システム用インストラクション英文チェック完了・録音

インストラクション用の英文をブルーフにかけて修正し、録音作業を行いました。

2013.7.5 音声収集システム開発業者打ち合わせ（2回目）

国別の電話番号取得、無音検知システム（ないし無音率の自動記録）、不正回答に対するストーリーの分岐システムなど、当方の要望を伝え、技術的に検討していただくことになりました。8月仮運用・試行、9月本運用のスケジュールで動いています。

2013.06.29 音声収集システムの仕様確定（国際留守番電話サービスを軸にしたシステム）・業者選定・契約。

見積り結果、業者が決定し、発注を行いました。8月運用開始を目標にしています。

2013.06.28 第1次試行実験（2013年2月実施）の書き起こし作業完了・プロトコルの再検討。

1人の被験者あたりの1分間平均発話量は、60語～120語程度となることがわかりました。B1_2はTOEIC675+, B2+はTOEIC765+に相当します。

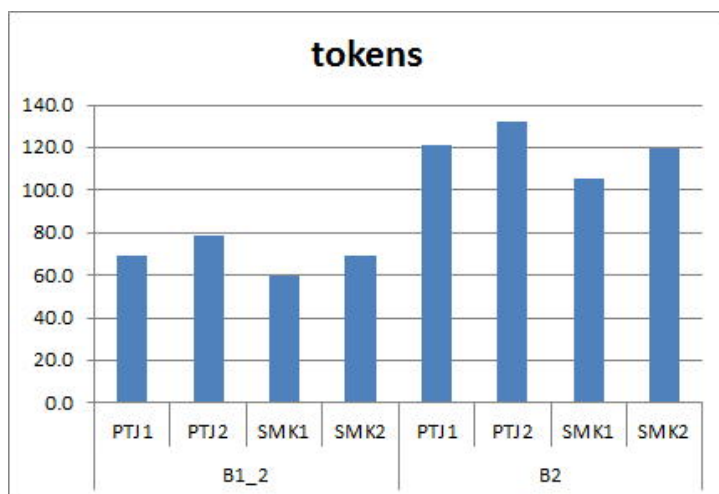


図1 日本人大学生のタスク（1分間）内発話量平均値

英語発話量は内容や状況によって可変的ですが、1分間の発話語数については、たとえば、[中学生で12～25語](#)、[中学生で18～23語](#)（資料内実践1）、[中学校英語授業内でのALTで26～68語](#)、[高校生で51～93語](#)、[TOEICおよびセンター試験のリスニングで150～170語](#)、[TEDで100～270語](#)、[政治家の演説で120～140語](#)、[米国のニュースで180～200語](#)、[一般のネイティブの会話で150～180語](#)、[語学学校における講師で200～230語](#)、などの報告があります（出典はリンク元参照）。

今回のタスクでは、一定の難度のある課題に対して（短いplanningの後）その場で考えながら論理的に立論を行なうことが要求されていますので、現在のプロトコルにおけ

る大学生の発話速はおおよそ妥当なものと言えるでしょう。

ただ、ICNALE-Wの要求語数が300～400語（制限時間20～30分）であったことと比べると、現在のプロトコルで得られる発話データの量はきわめて制約的なものになっています。そこで追実験で、制限時間を1分から2分に伸ばしたレコーディングを行いました。無発話時間（何を話していいかわからず長い時間黙り込んでしまう）が増え、タスクごとの条件統制が甘くなるというデメリットが観察されました。被験者により話せないことによる空白部が多く出現する2分データよりも、空白部が少なく、発話が詰まった1分間のデータのほうが比較資料としては有効性が高いと考えられます。

これらを総合的に判断すると、現在のプロトコルは一定の現実的妥当性を持つものと考えられます。また、個々の被験者による語数の不足は、連続して行われる2タスクを1タスクとみなすことで、ある程度解消できると予想されます。

2013.05 Learner Corpus 2013 Conference(イタリア・パドヴァ大学)において研究発表。

ICNALE-Wデータを分析し、学習者属性が言語使用に及ぼす影響について研究発表を行いました。



2013.05.01-10 音声収集システムの仕様をオンライン版と電話版について作成し、複数業者より見積もりを取得。

国により状況が多様であるため、現在、4種のデータ取得システムを構想しています。

- (A) 国内用：CALL教室でパソコンを使ってデータ取得 ※2013.2に試行済み
- (B) 海外用：留守番電話を用いたシステム開発
- (C) 海外用：Skypeを用いたシステム開発
- (D) 海外用：ICレコーダを海外に搬送して協力教員がデータ取得

今回は（B）と（C）について見積もりを集めました。費用対効果の関係で、(B)を先行開発することとなりました。

2013.04.01 新規科研プロジェクトの採択。

種別：基盤B（2013～2015年度）

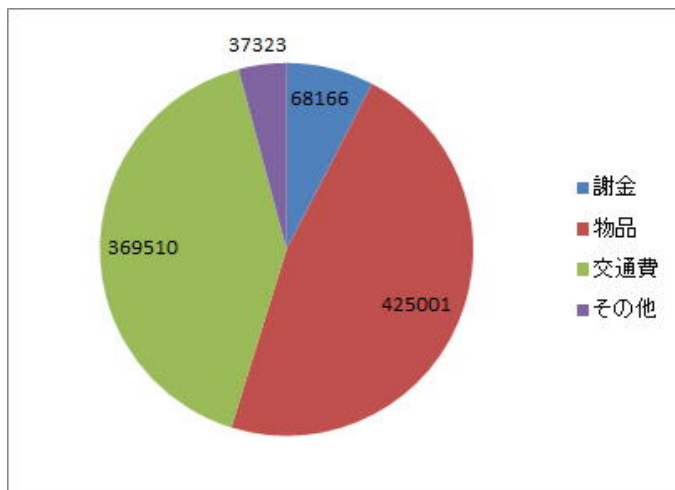
題目：[アジア圏英語学習者発話コーパスICNALE-Sの構築及び学習者発話の固有特性解明\(25284104\)](#)

2013～2014年度については、萌芽と基盤Bが並走しますので、2つのプロジェクトを一体運用していく体制を取ります。

2013.3.31萌芽プロジェクト初年度終了

本年度は、並走中の基盤Bプロジェクト（ICNALE-W構築）が最終年度であったため、2つのプロジェクトを融合させる形で研究を進めました。各種の文献研究などをふまえて発話データの収集にかかる技術的な問題を整理した上で、新しい発話収集のプロトコルの考案と試行実験を行い、新規の科研費に応募しました。

本年度の予算（90万）執行状況は下記の通りです。



2013.2.5/2013.2.6 第1回音声データ収集実験準備

今回はTOEICで700点以上の学生に絞って募集を行い、n=17のデータを取得しました。作業はCALL教室において、下記のフローで行いました。

事前単語テスト・学習アンケート回答（※ICNALE-Wと共通）

作業説明（5分程度）

自己紹介録音（1分） ※録音システムと1分間という時間に慣れるために行なう。本データは収集対象外。

タスク説明（1分）

タスク1（アルバイト or 禁煙）準備（1分）

タスク1第1録音（1分）

タスク1第2録音（1分）

タスク2（禁煙 or アルバイト）準備（1分）

タスク2第1録音（1分）

タスク2第2録音（1分）

音声データ送信

本プロジェクトでは、2種類のタスクにつき、それぞれ、第1録音と第2録音を連続して行わせることとしています（同じ内容を話させる）。一般に、初回のデータは緊張などのために十分な発話量が得られない可能性があります。連続して2回行うことで、学習者の発話をより多く誘引できると考えました。

なお、実験は複数スロット（時間帯）にわけて行い、スロットごとに、タスク1とタスク2の順番を逆にし、データ全体でタスク順序のカウンターバランスが取れるように配慮しました（※たとえば、すべてのデータ収集で禁煙が後になると、「馴れ」というファクターが禁煙テーマでの発話産出に影響を及ぼす可能性を統制するためです）

2013.1 第1回音声データ収集実験準備

2013年2月に予定されている音声データ実験の準備として、下記の作業を行いました。

- 1) CALL用データ収集プロンプトシートの作成
- 2) 被験者承諾書の内容検討
- 3) 謝金支払条件の検討

2012.12.28-30 International Conference on Education Systems (ICES'12)（クアラルンプール）

日本人学習者の音声聴解時の品詞性判断にかかる心理実験の結果について、研究発表を行いました。

2012.8.13-16 国際社会科学学会参加（上海）

表記学会に参加し、音声処理、学習者分析関係の発表を聴講しました。

2012.4.1

科学研究費に採択されました。

種別：挑戦的萌芽

題目：[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（24652120）](#)

このプロジェクトでは、将来の学習者発話コーパスの開発にかかる技術上の問題の洗い出しを目標としています。

[Back to top](#)

神戸大学石川慎一郎研究室 (C)2006-2016

神戸大学 国際コミュニケーションセンター

石川 慎一郎 研究室

Dr. Shin Ishikawa, Kobe University

The ICNALE-Spoken: A Teacher Guide

Updated 2014/07/05

Information about the project

We released a new learner corpus, The ICNALE-Written, in 2013. And now we are compiling its spoken counterpart, The ICNALE-Spoken,

- The project outline ([Ishikawa, 2013](#), [Ishikawa, 2014](#))
- [The ICNALE website](#)
- [The ICNALE-Written Online](#)

A Teacher Guide for Data Collection

1

Recruit local university students who volunteer to participate in the project.
The participants NEED TO...

- [A] be born and brought up in your country/region (Overseas, exchange, and returnee students, and the students who have stayed for more than one year in English-speaking countries, cannot join the project)
- [B] have the same L1s
- [C] be as varied as possible (For instance, the ratio of English major students should be smaller than approx. 20%. Also, collecting data at several universities of different academic levels would be desirable, though not compulsory).

And it is desirable for you to recruit the students who can report their scores in the standard English proficiency test such as TOEFL, TOEIC, or IELTS, though not compulsory, especially in ESL countries.

2

We will give you a toll-free phone call number. Before you begin data collection, please make a phone call by yourself both from your cell-phone and home/office phone and confirm that there are no connection problems.

3

First, tell **the first 10 students** (Not ALL!) to download both the [Excel Task Sheet](#) and [PDF Instruction Sheet](#). Or you can also distribute the files to your students by yourself.

4

Explain the procedure below to your students. Although the procedure is illustrated in the Excel file, you need to explain it to your students by yourself before they start the tasks. This is important to ensure that all the participants wholly understand the task structure and do the tasks in the same way. You also need to make participants understand and agree that **their data is included in the corpus after anonymization, which will be made public to the research community in the future.**

To Participants →

To the participants of the ICNALE-S Project

Thank you for your participation in the project.

Based on the instructions from your teacher, now you have to do seven tasks below. The needed time is approx. 30 minutes in total.

[0] Download the [Excel Task Sheet](#) and [PDF Instruction Sheet](#) from the web.

[1] Open your Excel Task Sheet and respond to the questions about your background (Excel Page 0)

[2] Respond to the 34 questions about your background of learning English (Excel Page 1)

[3] Answer the 50 vocabulary quizzes (Excel Page 2-5)

[4] Make a phone call to have a phone interview. You are given a task completion code at the end of the interview.

[5] Please fill in your task completion code in your Excel Task Sheet.

[6] Save your Excel file and give it a new name "Task completion code + your country + your family name" (Eg. 0105_Japan_ishikawa, 0123_China_Li)

[7] Send it to your teacher

In speech recordings, all of you NEED TO

1) record your speeches by yourself (without any friends around you) in a quiet room without noises

2) correctly understand the topics given in the phone interview and talk about them

3) continue to talk for 60 seconds in all the five speeches. Never stop till you hear the beep sound signaling the time-up (IMPORTANT)

4) make impromptu unprepared speeches (You should NOT speak with a script or anything prepared in advance)

5) keep the topics secret to your friends

Otherwise, your data will be rejected and you will be asked to do the same task again and again. In order to avoid wasted efforts, follow the rules above rigidly. Thank you very much for your participation and cooperation.

5

Collect the Excel files from the first 10 students and send them together to a project team. We will check their speech data and report the number of accepted files within a day.

6

Then, tell the next 10 students to do the same thing.

7

By repeating this process a couple of times, we can collect the good data without too many rejections.

8

The ICNALE-S will be made public after the completion of the project. As a joint researcher, however, you can use it freely for your own research even before the release of the corpus. If you are interested not only in the data collection but also in the data analysis and writing some academic papers using the ICNALE-S, let me know it. After checking the data, my grant agent will remit a commission to you. The amount changes according to the number of finally accepted files. The honorarium will be remitted directly to you not to your institute.

If you are interested in the project, please contact the ICNALE-S Research and Development Team.

The ICNALE-S Research & Development Team

Prof./Dr. Shin ISHIKAWA

School of Languages & Communication, Kobe University, Japan

1-2-1, Tsurukabuto, Nada-ku, Kobe, Japan

iskwshin@gmail.com

<http://language.sakura.ne.jp/s/eng.html>

[Back to top](#)

