English

Top

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

科学研究費プロジェクト進捗報告ページ
アジア圏英語学習者コーパスネットワークの構築による多層的中間言語対照分析（22320104）

Updated 2013/06/28
（本プロジェクトは2012年度で終了し，現在，このページの更新は終了しています。）

●はじめに
・石川研究室では，2010～2012年度にかけて，アジア圏英語学習者の大型統制作文コーパスであるInternational Corpus Network of Asian Learners of English（ICNALE）の構築を行ないました。このプロジェクトでは，2009年度末に完成した日中韓３か国の学習者の作文を集めたCorpus of English Essays Written by Asian University Students (CEEAUS）を大幅に拡張し，アジア圏10カ国・地域において，国際比較研究に耐える統制度の高い英作文データ130万語を収集しました。ICNALEは世界最大級の学習者コーパスの1つであり，収集したデータはICNALE Onlineで無償公開されています。

●注意
・下記の記述は，その都度のエントリ作成時点での処理の方針を示したもので，最終版では多くの点が変更されています。作業方針の確定に至る過程の記録として公開しておりますが，最終版のICNALEについての公式情報は下記論文でご確認ください。
Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In Ishikawa, S. (Ed.), Learner Corpus Studies in Asia and the World Vol.1 (Kobe University), 91-118.

・ICNALE最終版の公式サイトはこちら。データのダウンロードおよびオンライン検索も当該ページから可能です。

・ICNALE公開にともない，ICNALEの前身であったCEEAUSの公開は終了しました。

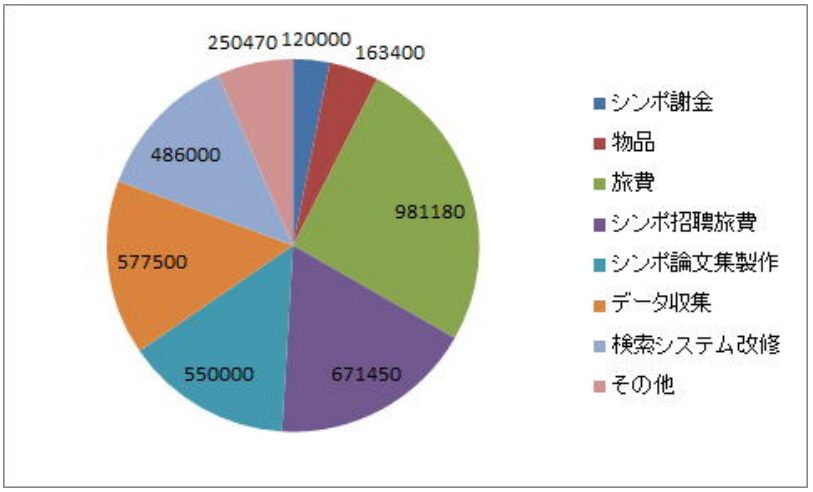
●謝辞
・多くの方のご協力を得て，130万語分のデータを収集し，世界最大級の学習者コーパスとして公開することができました。プロジェクトへのご協力に改めて御礼申し上げます。

作業日誌（2010年度～2012年度）

2013/3/31 ICNALEプロジェクト終了

3年間のプロジェクトが終了しました。各位のご支援にあつく御礼申し上げます。

平成24年度（2013年度）会計報告 直接経費380万円



2013/3/23-24 LCSAW2013

ICNALEの完成・公開を記念する国際シンポジウムを開催しました。また，論文集Learner Corpus Studeis in Asia and the Worldを刊行しました。

イベント告知
イベント実施報告
論文集目次

2013/2/28 LCSAW2013参加登録締め切り

海外参加者の登録を締め切りました。以後の申し込みは国内のみとなります。

2013/1/31 LCSAW2013フルペーパー投稿締切

国際シンポジウムのプロシーディングズ用フルペーパーの投稿を締め切りました。今後編集作業を進めていきます。

2013/1/23 ICNALE Online検索システム改修業務完了

最終版として各種調整が終了しました。

2013/1/18-20

全米コーパス学会（サンディエゴ州立大学）に出席し，ICNALE分析の結果について口頭研究発表を行いました。

2012/12/15 大井科研公開講演会@法政大学

表記講演会で，ICNALEを紹介しました。

2012/12/1 語用論学会　招待発表

コーパスと語用論の関係について，シンポジウムで招待発表を行いました。

2012/11/8 ICNALE Online検索システム総点検・改修業務発注

モニタの意見も反映し，現行版の改修項目をまとめ，業者と打ち合わせの上，発注を行いました。改修計画。

2012/10/31 LCSAW2013発表申し込み締切

ICNALE公開記念国際シンポジウムLCSAW2013の研究発表申し込みを締め切りました。世界6か国から多数の発表申し込みがありました。実行委員会で選考を行い，採択者に採択通知を送りました（11/15）。

2012/8/31 ICNALE Onlineユーザー情報自動収集システムの改修について

オンラインシステムの利用者に情報を登録させ，当該データを収集するシステムを現在の検索サイトに実装しました。

2012/08/6 ICNALE テキストデータ自動抽出システム開発打ち合わせ

フォルダ内に保存された約5000件のテキストファイルに対し，Excelに登録された個別テキストの詳細な書き手属性から自動で該当するテキストデータを一括選択する（たとえば，タイ・女性・TOEIC600点台などの複合検索条件で該当するテキストをバルクで抜き出す）システムの作成を要望し，見積もりを取りましたが，本年度の予算枠内では対応不可であることがわかり，システム開発は断念しました。

2012/7/24 個人属性データ付与済みJPNモジュール作業終了（β版をアドバイザーに送付）

日本人モジュールにつき，性別・年齢・習熟度・学習動機などの個人属性のひも付きデータ付与済み版が完成しました。プロジェクトのアドバイザーに送付し，助言をいただき，最終確定版としていきます。なお，語数が不足ないし超過しているものや，属性データの収集に一部不備があるものは最終データセットから除去することにしました。

2012/07/19 各種習熟度データのCEFR変換テーブル素案（V1）の決定
2012/07/20 Version2に変更（VST基準変更）
2012/07/22 Version3に変更（TEPSの変換基準をTOEICからTOEFLに変更）

	TOEFL(iBT)	TOEFL(PBT)	TOEIC	IELTS	Cambridge	STEP	TEPS	CET, TEM	VST
A2	*	*	225	*	KET	*	*	*	20
B1.1	57	487	550	4	PET	2	417	CET4	29
B1.2	72	533	650	4.5	*	*	502	*	32
B2.1	87	567	785	5	FCE	Pre1	608	TEM4/CET6	35
B2.2	99	597	860	6	*	*	700	*	38
C1	110	637	935	7	CAE	1	828	TEM8	41

■CEFRについて
・CEFR（Common European Framework of Reference）ではAはbasic，Bはindepent，Cはproficient usersと規定されます。
・B1とB2のサブカテゴリー基準は，主として日本人学習者の詳細な分析のために設定しています。

■TOEFL, TOEICからの関連付けについて
・TOEIC/TOEFLのCEFRへの関連付けは，"Mapping TOEFL® iBT, TOEIC® and TOEIC Bridge™ on the Common European Framework Reference"に準拠します。
・TOEFLのｉＢＴとＰＢＴの変換はETS TOEFL Internet-based Test Score Comparison Tableに準拠します。

■IELTS，ケンブリッジテストからの関連付けについて
・Cambridge ESOLによる"Cambridge ESOL and CEFR"に記載された公式対応付けに基づきます。
・上記のデータを引証する文献の中にはC1=7.5のように対応レベルの中央値で記載するものがありますが，ここでは下限値を基準とします。

■STEP（英検）からの関連付けについて
・日本英語検定協会　主任研究員　Jamie Dunlea氏による「英検とCEFRの関係性について」という公式レポートの結果に準拠しています。
・今回の調査では，日本人学習者は全員がTOEICスコアを持っているため，英検の級はレベル判定に使用していません。ただし，後述するCET，TEMによるレベル分けの際に，CET，TEMと，CEFRを媒介する基準として間接的に利用しています。

■TEPSからの関連付けについて
・TEPS（Test of English Proficiency Developed by Lanaguage Education Institute at Seoul National University）はソウル大学が開発した試験で，1999年以降，韓国の一部の大学生が受験しています。全200問で，語彙・文法・リーディング・リスニングの４セクションから構成されています。TOEICより難度の高い試験です。
・TEPS運営委員会は，TOEFL，TOEIC，TEPSの公式相関表を発表しています。TOEICとTOEFLはそれぞれCEFRと関連付けられているので，ここでは，各レベルに相当するTOFEL（iBT）の下限値をTEPSのスコアに変換したものを基準とします。なお，TOEICを根拠とした場合とTOEFLを根拠とした場合，後者のほうが基準が若干緩くなるようです。

■CET他からの関連付けについて
・CET（College English Test），TEM（Test for English Majors）は中国の大学生が受験する全国共通英語試験です。ともにスピーキング以外の３技能テストが含まれます（スピーキングが別途追加される場合もあります）。
・100点満点で60点以上が合格，85点以上が優等とされていますが，成績の差は今回の分析には含めません。
・CET4（120分）は一般学生（英語専攻者以外）が卒業時までに合格するよう義務付けられているテストです。
・CET6（120分）は国際関係学部などの学生が卒業時までに合格を義務付けられ，また，一般学生が大学院進学までにそれぞれ合格を義務付けられているテストです。
・TEM4（145分）は英語専攻学生が２年生の４月のときに受験し，卒業までに合格が義務付けられているテストです。
・TEM8（215分）は英語専攻学生が4年生の３月に受験し，合格が奨励され（一部では義務付け）られているテストです。
・これらの試験のTOEFL等のスコアへの換算，または，CEFRレベルの関連付けについて信頼できる指針は存在しませんが，井上裕子（2002）「大学生・大学院生対象英語検定試験：中国の場合」および宮内敦夫（2005）「中国における英語教育の現状：日本の英語教育を再考するために」における記述等を根拠とし，CET4を英検２級相当，CET6を準１級相当，TEM8を１級相当とみなし，英検側が行わっているCEFRへの関連付けを準用してレベル配当を行いました。
・各試験の要求語数は CET-4（4500語），CET-6（5500語），TEM4（6000語），TEM8（13000語）となっています。
・TEM4とCET6の等価性については，Shi Hui(2010)Pragmatic Transfer in English Emails Produced by Chinese L2 English speakers: A Study of the Underlying Cultural Ethos, and the Effect of Speakers' English Proficiency and Exposure to Englishにおける下記の記述を根拠とします（p.29）。

There is no official way to convert results between two tests; however it is widely believed in China that CET6 and TEM4 have the same level of difficulty. To conclude, the ranking of these four Chinese Englishproficiency tests on the difficulty level would be CET4<CET6=TEM4<TEM8.

■VSTからの関連付けについて
・プロジェクトで使用したのはNation, I.S.P. and Beglar, D. (2007) "A vocabulary size test" （The Language Teacher 31(7), pp. 9-13）によるVocabulary Size Test（monolingual version）（5000語＝50問バージョン）の正答数です。1000語刻みで１レベルにつき10問が出題されています。一般に，100倍することで学習者の推定語数となります（最大5000語）。
・CEFRと語数の関係について，信頼できる公式の見解は（知る限り）存在しません。
・推定法として，たとえばCEFRの各レベルに関連付けられた習熟度テストにおける要求語数を考えることも可能ですが，この場合，語数は非常に大きなサイズとなりがちです（たとえばTEM8は13,000語）。
・一方，今回の国際調査では，ESL圏の香港などの学生でも50点満点になることはなく，実際のコミュニケーションで必要となる語数はより小さいという考え方も存在します。
・J. Milton（2009），Measuring second language vocabulary acquisition（Bristol : Multilingual Matters）は，著者自身のギリシアとハンガリーにおける調査の結果，および，Meara & Milton（2003）の調査結果などをふまえ，C2レベルの語数上限を5000語としたうえで，下記の区分がrobustと思われるとしています（pp.185-186）。今回はこの枠組みに準拠し，各レベルについて示されているX-Lex（Meara開発の語彙テスト）の語数バンドの平均値を各レベルの判定基準とします。また，B1とB2の内部区分は単純に必要語数を等分する方式を採用しました。

Table 8.8 Mean EFL vocabulary size scores and the CEFR

CEFR level	Wordlist size	X-Lex	EFL Greece	EFL Hungary
A1		< 1500	1477	
A2	1000	1500–2500	2156	
B1	2000	2500–3250	3264	3136
B2		3250–3750	3305	3668
C1		3750–4500	3691	4340
C2		4500–5000	4068	

・しかし，X-LexとNation & Beglerのテストでは測定している語彙の性質も異なり，結果をそのまま援用することには危険性も存在します。VSTにおけるスコアで学習者のレベル区分を行った場合，実際の能力より過剰評価が起こりがちなことがわかりました。そこで今回はX-Lexのスコアレンジの中央値を当該レベルの下限値とみなすことにしました。たとえばC1の評価基準はX-Lex下限値の3750語ではなく，3750語～4500語の中央値である4125語（VSTスコア41）とします。
・今回の分析では，学習者が標準的な習熟度テストのスコアを持っている場合は（VSTの結果に関わらず）それによってレベルを判定し，テストのスコアがない場合に限って，便宜的にVSTのスコアに基づいてレベル区分を行います。
・TOEFLなどの本格的な習熟度テストに基づく能力区分と，簡易なVSTによる区分では信頼度に差があることから，データベースでは両者に異なるマークを付与しています（習熟度テストに基づく判定レベルはA1，B1など，VSTに基づく判定レベルはA1"，B1"などのマークが記載されています）。ただし，オンライン検索システムでは両者は一体的に扱われます。

2012/07/10 . ICNALE海外協力研究者・データ収集状況

香港：John Milton (Hong Kong University of Science & Technology) 《2011年2月完了》
台湾：Siaw-Fong Chung (National Chengchi University) 《2010年9月完了》
タイ（1）：Sonthida Keyuravong/ Punjaporn Pojanapunya （King Mongkut's University of Technology, Thonburi）《2010年9月完了》
タイ（2）：Sonthida Keyuravong/ Punjaporn Pojanapunya （King Mongkut's University of Technology, Thonburi）《2011年2月完了》
中国（1）：Katsuki　Mayumi (大連理工大)《2006年月完了》
中国（2）：Fang Li (Wuhan University) 《2010年12月完了》
中国（3）：Lu Yuanwen （School of Foreign Languages, Shanghai Jiaotong University)《2011年2月完了》
パキスタン：Asim Mahmood (Government College University (GCU) Faisalabad)《2011年2月完了》
韓国（1）：Sook Kyung Jung（Daejeon University）《2011年8月完了》
韓国（2）：Oryang Kwon (Seoul National University) 《2011年10月完了》
インドネシア：Leonardi Lucky Kurniawan（Polytechnic of Ubaya, Surabaya）《2011年2月完了》
フィリピン：Karen L. Gabinete（De La Salle University-Manila）《データ収集中 2012年7月完了予定》
シンガポール：Vincent Ooi（National Singapore University）《データ収集中　2012年9月完了予定》
英米圏（1）：Arc Inc.《2009年3月完了》
英米圏（2）：My Language Japan《2010年6月完了》
英米圏（3）：My Language Japan《2011年8月完了》

2012/07/01 個人属性データ付与作業およびテキストデータ整形作業開始

ICNALE Onlineでは，国別のすべてのデータを１本化して分析を行っていますが，プロジェクトでは書き手個人について詳細な属性データを収拾しており，この活用が課題となっていました。最終年度に入り，これから３か月の予定で，収集済みのすべてのデータについて，個人データのひも付け作業を行っていきます。ゼミ生の松下氏と陳 曄氏をこの業務に雇用することにしました。

また，これと並行して，最終公開のために，テキストデータの整形を行うことになりました。厳しい基準を決めて作文をしてもらいましたが，内容面の統制は取れているものの，書式や表記のレベルでは細かいずれがあり，それらが細かい分析の際に問題を引き起こす可能性があります。したがって，主要な書式のずれを統一的に修正する処理のデザインを検討中です。下記は現時点での作業プロトコル（201207現在）です。作業に使用しているエディタは「秀丸」です。以下，□は半角スペース，■は全角スペースを意味します。

2010.07.22 修正（文頭・文末スペース処理追加）
2012.07.20 修正（連続スペース処理方法変更）
2012.07.19 修正（タブ処理追加）

<分量チェック>
0）作業に先立ち，すべてのファイルの語数を一覧で検証し，200語～300語（±10%を許容しますので180語～330語）の範囲外のデータを除去します。

<改行削除>
1）改行記号の削除　¥n→□　（１行ずつ改行したりしている特殊なデータが修正されます）
2）タブ記号の削除　￥t→□（タブの混入が削除されます）

<全角記号置換>
3）「．」→「.」（全角ピリオドの混入を半角に修正します）
4）「，」→「,」（全角コンマの混入を半角に修正します）
5）「’」→「' 」（全角アポストロフィの混入を半角に修正します）
6）「‘」→「'」（全角シングル開始引用符を半角のアポストロフィに修正します）
7）「' '」→「"」（半角アポストロフィの二重併記を半角ダブル引用符に修正します）
8）「`」→「'」（逆向き半角アポストロフィーを通常の半角アポストロフィに修正します）
9）「—」→「-」（全角ダッシュを半角ダッシュに修正します）

<全角記号等削除>
１０）[、-○]→□（「（」や「<」など，全角記号をすべて削除し，半角スペースに修正します）
１１）■→□（全角スペースを半角スペースに修正します）

<句読点前後スペース調整>
１２）「. 」→「 .□」（ピリオドの後にスペースなしで文字が書かれているものを修正します　※すでに□がある場合はピリオド後が□□となります）
１３）「,　」→「,□」（同じくカンマの後にスペースなしで文字が書かれているものを修正します　※同上）
１４）「□. 」→「. 」（語の直後にピリオドが入るべきところ，スペースを介在させている例を修正します）
１５）「□, 」→「, 」（同じくスペース+カンマを修正します）
１６）「□'」→「' 」（前の語につながるべきアポストロフィーの前のスペースを修正します。例：isn 't）

<余剰スペース調整等>
１７）（正規表現で）□□+→□（これにより語間の余剰のスペース混入がスペース１つに修正されます。）
１８）（検索）^"（文字化けで出現しやすい文頭の引用符の混入を確認。あれば手作業で削除します。）
１９）（検索）"$（文字化けで出現しやすい文末の引用符の混入を確認。あれば手作業で削除します。）
２０）^□+→なにもなし（文頭のn個のスペースを削除します。）
２１）□+$ →なにもなし（文末ESO記号前のn個のスペースを削除します。）

<データ内容検証>
２２）AntConc等のコンコーダンサでアルバイトの全データセットに対して，smoke/smokingなどのキーワードで検索を行い，別テーマのエッセイの混入がないか確認します。
２３）同じく，禁煙のデータセットに対して，part time jobなどのキーワードで検索を行い，別テーマのエッセイの混入がないか確認します。

2012/06/20　ICNALE完成記念国際研究シンポジウム開催決定（2013.3.23-24予定）

Sylvian Granger氏，投野由紀夫氏，Vincent Ooi氏から基調講演の受託をいただきました。今後，企画の詳細を詰めていきます。

2012/6/30 中部地区英語教育学会学会賞受賞

ICNALEの改訂モジュールの分析結果をまとめた論文に対し，中部地区英語教育学会学会賞が授与されることとなり，本年度の岐阜大会において授賞式が執り行われました。

2012/5/10

フィリピンとシンガポールで最終年度のデータ収集が開始されました。

ICNALEはアジア圏のInner Circle，Outer Circle，Expanding Circleのすべてのデータ収集を目指しています。上記の2ヶ国のデータ収集により，国際英語研究資料としてのICNALEの価値がさらに高まることが期待されます。

2012/3/31
平成23年度の科研プロジェクト業務が終了しました。本年度のプロジェクト進捗状況について下記に略記します。

１．プロジェクトの目標について
本プロジェクトでは，アジア圏英語学習者の統制英作文を収集した世界最大級の英語学習者コーパス（International Corpus Network of Asian Learners of English : ICNALE）を構築し，言語教育学・英語学・日本語学・心理言語学・統計学をはじめとする関連分野の知見を統合しつつ，アジア圏英語学習者の中間言語使用特性を多角的に分析・解明することを目標としています。

２．平成23年度の進捗
1）データ収集
・昨年度末の80万語から，さらにデータ収集を進め，コーパス全体のデータ量は122万語となりました。公開されている統制学習者コーパスとしては世界最大級の規模となっています。

2）オンラインインタフェースの改修
・上記データの検索のためにICNALE Online Interfaceを開発し，運用を開始しています。フィードバックをふまえ，本年度は２回にわたり，システムの改修を行いました。

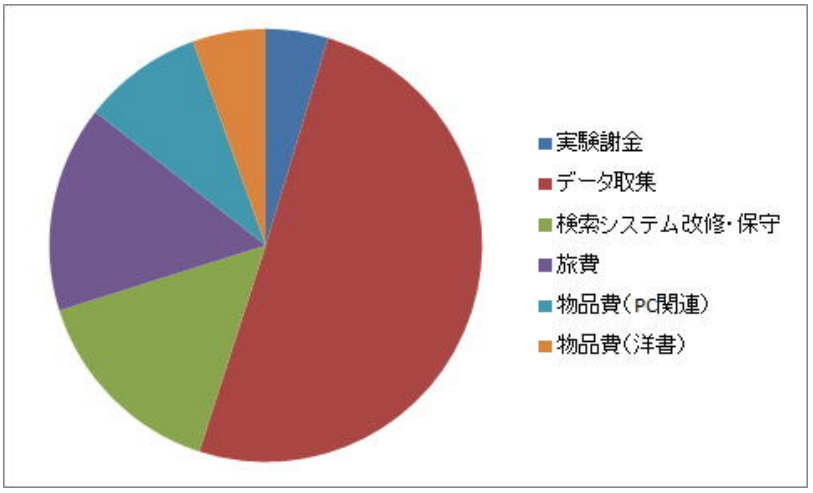
これにより，データの加除の自由度がまし，また，結果表示の言語学的妥当性が向上しました。

3）関連実験
・日本人英語学習者の中間言語運用の実態を探るため，語彙処理過程に関する心理言語学的実験，および，明示的・暗示的ライティング指導に基づく作文リライト実験を実施しました。

4）成果発表
・プロジェクトに関連して，年度内に，共著書1冊，論文3本（うち査読つき2本）が公刊されました。また，韓国・ベルギー・タイ（バンコク）・タイ（コンケーン）・アラブ首長国連邦・オランダを含め，内外の学会で17本の関連発表・講演を行いました（うち招待講演6本）

３．　決算報告について
・本年度は直接経費として400万円，間接経費の部分再交付分として30万円の助成を受けました。
・平成23年度の支出については，コーパスデータ収集（各国での作文データ収集）事業および検索システムの保守にかかる経費が全体の60%を占めています。
・科研費の原資が税金であることを常に意識しつつ，引き続き，研究目的の達成をめざし，適正な資金執行を行ってまいります。

支出内訳（直接経費分のみ）



平成23年度（2011年度）報告書

2012/3/27-29
第5回慣用連語国際学会(Formulaic Language Research Network)（Tilburh, オランダ）において研究発表を行いました。

当日はポスターセッションで発表を行い，アジア圏学習者のnグラム使用における共通性と特殊性について報告を行いました。

2012/1/27-29
第32回Thai TESOL International ConferenceにおいてWorkshopを行いました。

当日は45分のスロットをいただき，ICNALEの内容を紹介しました。先般のコンケーンでの講演の聴衆の方にも再訪いただき，充実したワークショップとなりました。また，ワークショップに参加されたカンボジアの研究者とも知己を得て，ICNALEのカンボジア展開の可能性について討議しました。

2011/12/19
ICNALE Onlineシステムの改修を行いました。

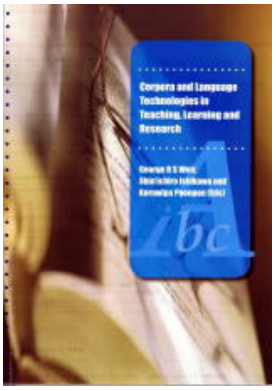
・KWICの検索条件設定画面のNODE Word Unit Case等の各行の部分の左寄せ
・KWICKの検索結果の「Sorting」各ブロックへの項目文言表記
・縮約形の表示アルゴリズムの変更
　isn't ⇒　isとn'tという2語として独立して処理，n'tはnotと別語扱い
　can't ⇒　同様にcan+n't
　won't ⇒　同様にwill+ n't
　couldn't ⇒同様にcould + n't
　he's ⇒he+'s
　they're ⇒　they+'re
　I'm -> I +'m）

2011/10/10
中国人英語学習者の第3次データの納品発注しました。
韓国人英語学習者の第2次データの納品入札の準備を行いました。

2011/9/27
ICNALE Onlineのデザイン修正を行いました。

また，中国人英語学習者の第3次データの納品入札の準備を行いました。

2011/9/20
ICTATLL 7th International Conference（タイ，チャロエンタニホテル@コンケーン）において基調講演およびテーマシンポジウムを行いました。



題目：A New Horizon in Learner Corpus Studies--What CL can do for studies of Englishes in Asia--
講演の中でICNALEを紹介しました。また，学会発表論文を中心として，George Weir, Shin'ichiro Ishikawa, Korwnpia Poonpon (Eds.) Corpora & Language Technologies in Teaching, Learning and Research (University of Strathclyde Press, UK)が出版されました。基調講演論文として，A New Horizon in Learner Corpus Studies: The Aim of the ICNALE Project (pp. 3-11)が収録されています。同論文には，現時点のプロジェクトの概要をまとめました。

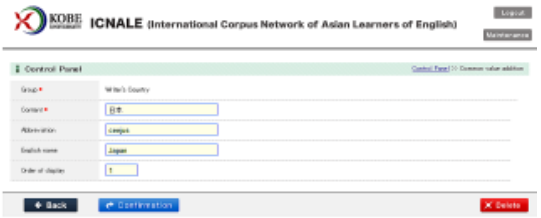
2011/9/17
Learner Corpus Research 2011（ベルギー，メルキュール・ルヴァンラヌーヴホテル）でシンポジウム発表を行いました。

題目：Uniqueness and Commonality in the World Englishes in Asia:The Aim of the ICNALE Learner Corpus Project
World Englishesのシンポジウムの中で，ICNALEを紹介し，アジア圏のInner Circle，Outer Circle，Expanding Circleの研究資料としてのICNALEの可能性を論じました。

2011/9/1
大学英語教育学会全国大会（@西南学院大学）で研究発表を行いました。

題目："Changing Grammar in World Englishes: A Corpus-Based Study "

2011/8/25
ICNALE Onlineの管理画面の開発が完了しました。



これにより，新しいデータの整理区分，表示方法などが自由に設定できることとなります。また，データの追加・変更の情報がリアルタイムに表示される仕様となりました。

2011/8/22
Asialex 2011 Kyoto Conferenceで研究発表を行いました。

題目：Learner corpus and lexicography: "Help-boxes" in EFL dictionaries for Asian learners—A study based on The International Corpus Network of Asian Learners of English—
また，上記題目の論文がAkasu, K., & Uchida, S. (Eds.) ASIALEX2011 Proceedings: LEXICOGRAPHY Theoretical and Practical Perspectives（pp.190-199）に収録されました。

2011/8/21
全国英語教育学会山形研究大会で研究発表を行いました。

題目：Writing, Rewriting, Proof Writing：量的言語指標に見る可変性と不可変性
また，上記発表に基づく論文を学会誌に投稿しました（審査中）。同発表・論文では，ICNALEのアドオンモジュールとして別途開発を進めているプルーフライティングモジュール，および，２条件に基づくリライティングモジュールをオリジナルモジュールと比較した結果について分析しています。

2011/7/31
ICNALE Online（修訂版）が納品完了しました。

処理速度が40%以上早くなったほか，細かい検索設定を変更しています。また，パスワード制を廃止しました。

2011/7/29
AsiaTEFL 2011 Conference（ソウル文化教育会館）において研究発表を行いました。

題目：Use of tense in writing by Asian learners of English: A contrastive study based on the ICNALE learner corpus
また，韓国滞在中に，国際交流基金ソウル事務所を表敬訪問し，事務所長ならびに事業部長氏とアジア圏日本語コーパス開発の可能性について意見交換を行いました。

2011/7/3
国立国語研究所日本語教育研究・情報センタープロジェクト 「多文化共生社会における日本語教育研究」合同研究発表会において講演を行いました。

題目：学習者コーパス研究の可能性 ―日本語教育とコーパス―
ICNALEの構築過程を紹介し，日本語学習者コーパス開発への応用について意見を述べました。

2011/7/2
日本アジア英語学会第28回全国大会（京都外国語短期大学）で大会基調講演を行いました。

題目：Englishes in Asia: What ICNALE Corpus tells us
本発表では，Kachruの言う3つの同心円モデルを実証するコーパスとしてのICNALEの可能性を考察しました。学会ニューズレターはこちら。

2011/5/27

語彙習得・処理に係る心理実験を実施しました。

被験者5名
於：情報通信研究機構

また新規に実験用の携帯型心理機能測定機器を購入いたしました。

2011/5/10

ICNALE Onlineの機能追加に係る仕様検討会議を開きました。

主な改造予定点
・データ更新管理画面の新設
・コンテキスト検索への対応（Xを含み，前後n語以内にYを含む／含まない）
・POSタグ検索の完全対応
・高校生データ追加に備えた仕様変更
・利用者ログ管理画面の新設

2011/4/1

平成23年度のプロジェクトがスタートしました。
本年度はICNALE onlineの正式公開を予定しています。

2011/3/29

平成22年度の科研プロジェクト業務が終了しました。本年度のプロジェクト進捗状況について下記に略記します。

１．プロジェクトの目標について
本プロジェクトでは，アジア圏英語学習者の統制英作文を収集した世界最大級の英語学習者コーパス（International Corpus Network of Asian Learners of English : ICNALE）を構築し，言語教育学・英語学・日本語学・心理言語学・統計学をはじめとする関連分野の知見を統合しつつ，アジア圏英語学習者の中間言語使用特性を多角的に分析・解明することを目標としています。

２．平成22年度の進捗
1）データ収集
・年度内に，英米8万語，日本24万語，タイ18万語，中国11万語，パキスタン9万語，台湾9万語，香港4万語のデータ収集が完了しました。
・現時点でコーパスのサイズは約80万語となり，公開されている統制学習者コーパスとして最大級の規模となりました。
全体：792,613 tokens
US/UK： 81,229 tokens　Japan： 242,431 tokens　China： 108,602 tokens　Pakistan： 91,232 tokens
Taiwan： 91,680 tokens　Thailand： 177,439 tokens　Hong Kong： 40,000 tokens
・新年度は，既存データの量的拡充を目指すとともに，韓国・フィリピン・シンガポールを重点国に定め，データ収集を進めていく予定です。
・ICNALEは，ベルギーのル=ヴァンカトリック大学で構築する世界の学習者コーパスデータベースにも登録されました。

2）インタフェース開発
・上記データの検索のためにICNALE Online Interfaceを開発し，運用を開始しました。
・インタフェースには，語彙分析・語彙表分析・コロケーション分析のほか，統計値に基づく特徴語自動抽出機能を実装しました。

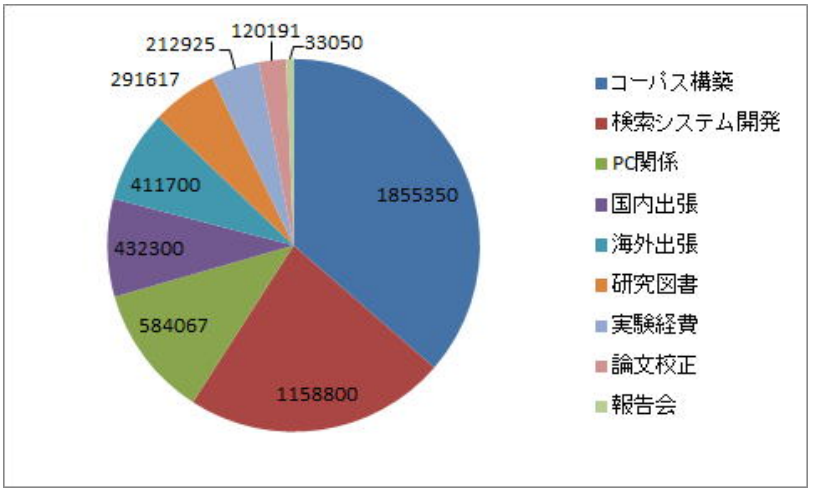
3）関連実験
・日本人英語学習者の中間言語運用の実態を探るため，語彙処理過程に関する心理言語学的実験，および，明示的・暗示的ライティング指導に基づく作文リライト実験を実施しました。

4）講演会実施
・Hong Kong University of Science & Technology（HKUST）よりJohn Milton博士を招聘し，公開講演会を実施しました。

5）成果発表
・プロジェクトに関連して，年度内に，共著書4冊（『言語研究のための統計入門』，『Corpus, ICT, and Language Learning』，『英語研究と英語教育』，『Phraseology，Corpus Linguistics and Lexicography』），論文6本（うち査読つき4本）が公刊されました。また，シンガポール・ポーランド・マレーシアを含め，内外で16本の関連発表・講演を行いました（うち招待講演7本）。

6）その他
・コーパスデータに特化した統計処理のため，統計分析マクロSeaull-Statの改造を行いました（『言語研究のための統計入門』付属CD-ROMに所収）。
・複数辞書に対応した日本語形態素解析システムを開発しました（非公開）。

３．　決算報告について
・本年度は直接経費として510万円の助成を受け，研究目的に合わせ，下記の通り，支出を行いました。
・平成22年度の支出については，コーパス構築（各国での作文データ収集）事業および検索システム開発（検索インタフェース・統計処理システム構築）事業に掛かる支出が全体の60%を占めています。
・科研費の原資が税金であることを常に意識しつつ，引き続き，研究目的の達成をめざし，適正な資金執行を行ってまいります。



平成22年度（2010年度）報告書

2011/3/10
パキスタン，香港，タイ（2次）のデータ収集が完了しました。
このうち，パキスタンとタイをシステムに実装しました。

2011/1/10
オンライン検索システムの開発（β版）が完了しました。
http://icnale.por.jp/login.php
デモ検索を体験なさりたい方は，仮パスワードを発行しますのでメイルにてご連絡ください。

システムには，現在，下記のデータセット（POSタグ付与済）がロードされています。
Native Speakers: 81,229 tokens
Japan: 242,431 tokens
China: 20,017 tokens
Taiwan: 91,800 tokens
Thailand: 93,909 tokens

今後，韓国・中国・パキスタン・インドネシア他のデータが順次追加予定です。

ICNALE（International Corpus Network of Asian Learners of English）プロジェクトでは，アジア圏EFL／ESL諸国を対象に，世界最大級の統制英作文国際学習者コーパス（※トピックや執筆条件を統一してデータをアジア圏各国で取得する）の構築を目指しています。

2010.12/10

石川慎一郎他編著『言語研究のための統計入門』（くろしお出版）刊行。本書には，統計分析のサンプルデータとして，ICNALEの前身であるCEEAUSのデータと，本科研プロジェクトで開発委託を行った統計ソフトウェアがCD-ROMに同梱されています。

2010.12/5

英米圏母語話者データの収集が完了しました。

2010.12/5

英米圏母語話者データの収集が完了しました。

2010.12/1-2

マレーシア国際外国語教育学会（プトラジャヤ大学）において研究発表を行いました。
発表では，ICNALEのデータの一部を用い，学習者の属性データがライティングに及ぼす影響についての分析結果を報告しました。

2010.12/1

香港のJohn Milton教授より，香港モジュールデータのサンプルが送付されました。今後，年度内に所定数の収集を完了する予定です。

2010.10/18-22

ポーランド・ワルシャワ国際言語学会において研究発表を行いました。
発表では，国際学習者作文比較研究の基礎となるEFL/ESL圏のEnglishesの諸相についての計量分析結果を報告しました。

2010.9/30

台湾およびタイでのデータ収集（1次）が完了しました。今後，タグ付処理の上，オンラインインタフェースに登録される予定です。

2010/9/21-23

ICTATLL2010 The 6th International Conferenceにおいて，下記の研究発表を行いました。
Shin Ishikawa "Modality expression in interlanguage: A study based on learner corpus"

また，同論文が下記の研究書に収録されました。
Weir, G., & Ishikawa, S. (Eds.),　Corpus, ICT, and language education (Glasgow, UK: University of Strathclyde Publishing)

2010.9/20

科研費による講師招聘講演会を実施しました。

「学習者コーパス研究の新しい展開」
日時：平成22 年（2010 年）9 月20 日（月・祝）2 時30 分～5 時（受付2:15 分～）
会場：メルパルク京都6 階第４会議室（JR 京都駅下車徒歩１分）
招聘講師：John Milton 博士（香港科技大學）
主催：神戸大学石川研究室（ICNALE 科研プロジェクト）
参加者：18 名（うち，海外参加者は，香港１，シンガポール１，英国１）

実施報告書はこちら。

同講演会では，招聘講師の講演に先立ち，下記の小講演を行いました。
Shin Ishikawa, "The outline of ICNALE learner corpus project"

2010.9/15

第3次台湾データ収集業務の委託を完了しました。
ICNALEオンライン検索システムが完成し，運用サーバーへの移転準備に入りました。

2010.9.7-10

オンライン検索システム　改修
・用語などの英語化（ユニバーサル化）
・トップページにnotice boardを追加
・頻度計量時に，type/token表示，およびPMW表示を新設
・Keyword表示時の計算アルゴリズムのミスを修正
・検索ボックスの単一化とインタフェースの改良

2010.8.16

オンライン検索システム　パイロットV2班完成

KWICの結果表示などが完成。今後，インタフェースの改善と，動作のスピードアップを行います。

2010.8.8

全国英語教育学会大阪研究大会

Pedagogical application of the findings from learner corpus studies: What changes the learners' essays?のタイトルで研究発表を行い，ICNALEの新しい日本人モジュールであるProficiency モジュールとRewriteモジュールを使った分析結果を報告しました。

2010.7.29

オンライン検索システム開発会議

受託業者と面談し，パイロット版をもとに，今後の開発の方向性を打ち合わせました。

2010.7.27

オンライン検索システム　パイロットV1版完成

受託業者と面談し，仕様の詳細を打ち合わせました。

2010.7.21, 23

ICNALE 日本人新モジュール　Rewriteモジュールデータ収集（第１次）

ICNALEの日本人モジュールのコア部分はデータ収集がほぼ完了し，現在は，特殊環境下でのデータを含んだ小規模モジュールの開発を進めています。Rewriteモジュールは，インプリシット，エクスプリシットなインストラクションを受けることで，言語産出がどう変化するかを調べる新しいタイプの学習者データです。今後，データ収集を強

化していきます。

2010.7.27

オンライン検索システム　パイロットV1版完成

受託業者と面談し，仕様の詳細を打ち合わせました。

2010.7.13

オンライン検索システム開発会議

受託業者と面談し，仕様の詳細を打ち合わせました。

2010.6.13

John Milton博士講演会を実施します

当研究室がホストしているICT・コーパスの国際学会ICTATLL2010 Kyoto Conference（2010.9.21-23）に博士が来日される機会を活用し，学会前日の2010年9月20日に研究室主催による科研費講演会を開催することが決まりました。詳細は後日アップします。

2010.6.12

海外協力研究者の追加

学習者コーパス研究の第一人者であり，HKUST Learner Corpusの開発者でもあるJohn Milton博士（Hong Kong University of Science & Technology）がICNALEプロジェクトの海外協力研究者に就任されました。香港でのデータ収集は2010年11月以降に開始予定です。

2010.6.7

海外協力研究者の追加

インドネシアのWahju Bandjarjani氏（Lecturer of Universitas PGRI Adibuana Surabaya）がICNALEプロジェクトの海外協力研究者に就任されました。
Bandjarjani氏には先般のRELCセミナーでプロジェクトを紹介し，就任の快諾を得ました。
インドネシアでのデータ収集は本年7月以降，順次開始の予定です。

また，現在，台湾・中国・香港の研究者にプロジェクトへの協力を要請中です。

2010.6.6

JACET中部支部大会（中京大学）において招待シンポジウム発表

シンポジウム「多文化共生時代の英語教育」
石川発表題目：Grammar in Englishes：コーパスで見る文法規範の揺らぎ

アジア圏の英語学習者データの分析にあたっては，当該地域で使用されている英語の諸相を知ることが必要になります。本発表では，Kachru（1992）の言う内円英語変種コーパスと，外円英語変種コーパスを構築し，語彙文法の異同性を観察しました。

2010.5.15

業務委託

データ収集（国外），データ収集（国内），ウェブ検索システム開発の3業務について業者選定・委託を行いました。 ウェブ検索システムは，2010年12月の仮運用開始を予定しています。

2010.5.10

海外協力研究者の追加

タイのSonthida Keyuravong氏（Professor of King Mongkut's University of Technology, Thonburi）がICNALEプロジェクトの海外協力者に就任されました。
同教授は，沖原勝昭神戸大名誉教授の共同研究者として神戸大学にも何度かおいでいただいている方です。
先般のRELCの学会で就任を依頼し，快諾を得ました。
タイでのデータ収集は本年6月にスタートする予定です。

2010.5.1

海外データ収集フォーマットの作成

アジア圏各国で同一の統制条件下でデータ収集を行うため，新規にICNALEのためのデータ収集フォーマットとデータ収集ガイドライン（英語版）を作成しました。
新フォーマットはExcelのテンプレートとなっており，被験者の属性情報や語彙力など，必要な情報が悉皆的に収集できるようになっています。
これは，先行して実施したCEEAUSコーパスのデータ収集フレームに改良を加えたものです。

2010.4.19-20

東南アジア諸国教育大臣機構(South-East Asian Ministries of Education Organization / SEAMEO)外国教師養成研修センター年次会議（RELC Seminar2010）において，大学英語教育学会（JACET）代表として招待発表を行いました。

発表題目Interlanguage of the Asian Learners of English: Redefinition of Proficiency

発表では，ICNALEの前身であるCEEAUSの紹介と，CEEAUSから得られた知見について報告しました。会場で共同研究の呼びかけを行ったところ，インドネシア・タイ・韓国・フィリピンの研究者より研究参加の申し出を受けました。

2010.4.6-30

プロジェクトの連携研究者に本年度の計画についてご説明を行いました。
本年度は，１）アジア圏主要国におけるデータ収集，２）母語話者データの拡充，３）検索システムの開発，４）先行取得データを用いた分析を軸にプロジェクトを展開していきます。

2010.4.5

科学研究費に採択されました。

種別：基盤B
課題番号：　22320104
課題名：アジア圏英語学習者コーパスネットワーク構築による多層的中間言語対照分析

Back to top

神戸大学石川慎一郎研究室 (C)2006-2016