

国際学習者コーパス・プロジェクト

International Corpus of Crosslinguistic Interlanguage (ICCI)

投野由紀夫

国際学習者コーパス・プロジェクト

International Corpus of Crosslinguistic Interlanguage (ICCI)

投野 由紀夫

(東京外国語大学大学院総合国際学研究院)

要 旨

本論はグローバル COE の一環で行われている国際プロジェクト International Corpus of Crosslinguistic Interlanguage (ICCI) の中間報告である。現在、8 カ国（地域）の研究者が参加して、初級・中級英語学習者の自由英作文データをコーパス化する計画を進めている。本稿ではその研究目的、方法について述べるとともに、2008 年度に中心で進めてきたパイロット・コーパス作成、およびエラー・アノテーション・ツールの開発、Xaira ベースの XML 形式コーパス検索環境の整備などについて概説し、今後の展望を述べたい。

1. はじめに

学習者コーパスは、外国語学習者の話し言葉・書き言葉の産出データを目的をもって収集、電子化し、コーパスとして整備したものである。学習者コーパスは 1990 年代の初めに、ロンドン大学を中心に構築が開始された International Corpus of English (ICE) のサブコーパスの 1 つとして学習者英語のコーパスを構築しようという試みとして Louvain 大学の Sylviane Granger が International Corpus of Learner English (ICLE) という国際学習者コーパスプロジェクトを立ち上げたのがその始まりといわれている (Granger 1998)。しかし、学習者データの電子化は日本でもかなり早くから開始されており、投野も 1980 年代後半には現在の JEFLL Corpus (投野, 2007) の原型となる英作文データの電子化を始めている。現在の学習者コーパス研究の概要に関しては、Granger, Hung, & Petch-Tyson (2002), Pravec (2002) 他を参照されたい。

本稿では、東京外国語大学グローバル COE プロジェクトの一環で行われている、国際学習者コーパス・プロジェクトの 2 年目の中間報告を行う。まず、プロジェクトの目的と研究チームの編成を説明し、次いでパイロット・コーパスの設計と収集、関連するコーパスのフォーマットやタグ付け、アノテーション・ツールおよびインタフェースの開発に関して進捗を説明し、最後に今後の計画に関して述べる。

2. 目的

本プロジェクトは「国際中間言語比較コーパス (International Corpus of Crosslinguistic Interlanguage, 以下 ICCI) プロジェクト」と称する。ICCI の主たる目的は、投野が構築している英語学習者の初級・中級の日本人中高生 1 万人の英作文コーパス JEFLL Corpus と比較対照可能かつ母語を異とする学習者データのコーパス化である。学習者コーパス研究は ICLE に代表される主要コーパスのほとんどが中上級 (主に大学生) の英語学習者データを収集したものが多く、よって、そこからわかる母語話者との言語特徴の差異はある程度出来上がった状態の中間言語の比較になり、基本的な文法項目ではなく、談話や語用などの発展的な側面が中心になることが多かった。しかし、日本のような外国語教育の環境での英語習得にはもっと基本的な習得過程でのつまずきが多く、上級レベルでの中間言語の比較はあまり教育的な示唆を持たないことが多い。英語の初学者レベル、特に学校教育で英語を学習した中高生の自由英作文データを中心にコーパスを構築することを目的とした、JEFLL Corpus は現在学習段階の異なる初学者～中級学習者のデータとしては世界で最大規模のコーパスとして知られている。

JEFLL Corpus はすでに、第 1 収集段階を終えて、一般公開されている¹。また、これらをもとにした各文法項目の記述研究 (投野, 2007), 分析結果の英語教育への応用と提言 (投野, 2008-2009) などが行われ始めている。またコーパス言語学の分野で発展してきた学習者コーパスが、e-learning や Intelligent CALL などの研究者からも興味をもたれるようになってきており、米国の大規模な CALL 学会 CALICO でも昨年から「学習者言語の自動分析 (Automatic Analysis of Learner Language: AALL)」というワークショップが開催されるようになってきた。

本プロジェクトはこういった流れの中でいくつかの重要な目的を持っている：

- (1) a. JEFLL Corpus と比較可能な日本語以外の母語を持つ初級・中級の英語学習者による自由英作文データを収集し、コーパス化する
- b. JEFLL Corpus で明らかになりつつある、英語習得過程の記述研究を、他の学習者グループのデータと対照研究を行うことで、母語の異なる学習者グループでも共通に見られる現象であるか否かを検証する
- c. 異なる母語の英語学習者グループを相互比較するための中間言語用途に特化したアノテーションやツールの開発整備を行う。

これらが進めば、第 2 言語習得研究としての成果のみならず、コーパス基盤による記述的な第 2 言語習得研究の結果を用いて、英語教育に対する新たな提言や、新たなシラバス開発・教材開発の可能性などが開ける。それらのための基礎研究としての重要な位置づけを ICCI は持っている。

¹ http://scn02.corpora.jp/~jefll03/jefll_top.html

3. プロジェクト・チームの概要

現在、ICCIは投野が中心となり、日本・スペイン・オーストリア・イスラエル・ポーランド・台湾・香港・シンガポールの8カ国（地域）の研究者が共同研究を行っている。各国の研究者は、投野が国際会議等で知り合った学習者コーパスまたはコーパスの言語教育への応用に関心のある中堅・若手の研究者で、それぞれが今回のプロジェクトの趣旨を理解して、各国や地域でのデータ収集の中心的役割を果たしてくれている。代表は以下のようなメンバーである：

日本	:	投野由紀夫（東京外国語大学）
スペイン	:	Pascual Pérez-Paredes（Universidad de Murcia） Maria Belen Diaz Betmar（Universidad de Jaen）
オーストリア	:	Tom Rankin （Vienna University of Economics and Business Administration）
イスラエル	:	Tami Aviad（University of Haifa）
ポーランド	:	Agnieszka Lenko（University of Warsaw）
台湾	:	May Ma（Language Testing & Training Center）
香港	:	David Lee（Chinese University of Hong Kong）
シンガポール	:	Huaqing Hong（Center for Research on Pedagogy & Practice）

2008年2月に行われた第1回のメンバー会議で、JEFLL Corpusのコーパス・デザインと比較可能で、かつより汎用性のある学習者コーパス収集のデザインが検討された。その後、2008年4月～10月にかけて、各チームで現地の小学校～中学校・高等学校レベルのデータ収集のための協力体制を組み、パイロット・コーパスのためのデータ収集を行った。10月～2009年1月まで作文データの書き起こしを東京外国語大学の投野研究室で行い、それらをもとにXMLフォーマットのデータ形式に変換して、全員でデータを共有化し、かつエラー・アノテーションのためのツール改良および共通利用のためのwebインターフェースの開発を行った。

4. ICCI コーパス・データ採取の方法

ICCIのコーパス・データは、JEFLL Corpusのデータ収集方法に可能な限り近くということを目指して話し合いが行われた。JEFLL Corpusの収集方法の基本方針は以下のとおりである：

- (2) a. 授業時間内の20分間で行う自由英作文課題
- b. できるだけたくさん書くように指示
- c. 英語がどうしても書けない部分は日本語を使用してよい

- d. テーマは下記の 6 種類 :
- i) 朝食はパンがいいか御飯がいいか
 - ii) 今までに見た怖い夢について
 - iii) 学園祭について
 - iv) 浦島太郎のその後
 - v) お年玉をもらったらどう使うか
 - vi) 大地震が来たら何をもって逃げるか
- e. 辞書の使用は禁止
- f. 宿題などにはしない
- g. 書き直し(revision) は行わない
- h. 指示文と初学者のために簡単なサンプルを示す

これらのうち, ICCI のパイロットでは, 辞書なしで授業中 20 分のタスクということは一致が得られたが, 作文のトピックと指示文やサンプルの提示に関してはさまざまな疑問点が出され, 結局以下のような枠組みでパイロットデータの収集をすることに決定した:

- (3) a. 辞書なし
- b. 授業中の 20 分間
- c. タスクは下記の中から, 各国の事情により一番やりやすいものを選ぶ:
- i) What is your favorite food?
 - ii) How do you spend the money you got for your birthday?
 - iii) What do you do after school?
 - iv) Tell us about the funny (strange) thing that recently happened to you.

学習者データに関してはプロフィールを取得するための簡単な質問肢が用意された。共通の情報として以下のものを最低限聞くこととし, それ以外は各国の状況に合わせて, 研究上関心のある事項を各自で付随的にアンケートで聞いてもよいこととした:

- (4) a. 学年 (各国の学校教育制度に合わせたもの)
- b. 性別
- c. 母語
- d. その他 (英語力指標など)

2008 年 4~10 月の約半年間で, 上記課題に関してパイロットのデータ収集を行い, 無記名のランダム調査を条件に現地の小学校~高等学校の許可を得て自由英作文課題を実施し, 手書きデータをすべて東京外国語大学投野研究室に送付し, 院生を中心としたチームで作文の書き起こしを 7~12 月にかけて実施。2009 年 1 月にほぼ第 1 次のパイロット・データの作成を完了した。書き起こしデータは研究チームのメンバーに配布されると同時に, シ

ンガポールの Huaqing Hong 氏により、XML フォーマットに変換され、かつ、次節で解説するような基本的なアノテーションを付与されて、検索可能なデータとして加工された。

5. 書き起こしデータのフォーマット

パイロット調査の書き起こしデータ 524 ファイルは、JEFLL Corpus のオンライン版（小学館コーパス・ネットワークで無償公開中）のフォーマットに合わせて下記のように整形された：

```
<file id="00001">
<head>
<cat id="00001">
country='Austria'
school=""
schoolname=""
schoollevel=""
region=""
textbook='You & Me'
year='7'
class=""
studentid="" name=""
sex=""
mothertongue='German'
otherscore=""
medium='written'
genre='descriptive'
topic='money'
dicuse='no'
preparation='no'
time='20min'
feedback='none'
date=""
transcriber='xxxx'
</cat>
</head>
<body>
<s>I would spend the money to Caritas, because in other lands many children haven't any money.</s>
<s>Or I buy clothes with the money, then I can spend my old clothes to <?>Carias</?> too.</s>
<s>I would probably play Lotto too.</s>
</body>
</file>
```

図 1 : ICCI 書き起こし 1 次フォーマット

これらのフォーマットから、XML 形式への変換を行い、ヘッダ部分で現状の分析に必要なものみに絞って図 2 のような形式になった：

```

<?xml version="1.0" encoding="utf-8"?>
<Doc>
<Header>
<Country>Austria</Country>
<Gender>Nil</Gender>
<MotherTongue>German</MotherTongue>
<Genre>descriptive</Genre>
<Topic>money</Topic>
<Year>7</Year>
<Time>20Min</Time>
<FileName>astr00001</FileName>
</Header>
<Text>
<T id="t_1"><w pos="ZZ1">I</w> <w pos="VM">would</w> <w pos="VVI">spend</w> <w
pos="AT">the</w> <w pos="NN1">money</w> <w pos="II">to</w> <w pos="NP2">Caritas</w> <w
pos="PUNC">,</w> <w pos="CS">because</w> <w pos="II">in</w> <w pos="JJ">other</w> <w
pos="NN2">lands</w> <w pos="DA2">many</w> <w pos="NN2">children</w> <w
pos="VH0">have</w> <w pos="XX">n't</w> <w pos="DD">any</w> <w pos="NN1">money</w> <w
pos="PUNC">.</w> </T>
<T id="t_2"><w pos="CC">Or</w> <w pos="PPIS1">I</w> <w pos="VV0">buy</w> <w
pos="NN2">clothes</w> <w pos="IW">with</w> <w pos="AT">the</w> <w pos="NN1">money</w> <w
pos="PUNC">,</w> <w pos="RT">then</w> <w pos="PPIS1">I</w> <w pos="VM">can</w> <w
pos="VVI">spend</w> <w pos="APPGE">my</w> <w pos="JJ">old</w> <w pos="NN2">clothes</w> <w
pos="II">to</w> <w pos="PUNC">ENT001</w> <w pos="NN2">Carias</w> <w
pos="PUNC">ENT002</w> <w pos="RR">too</w> <w pos="PUNC">.</w> </T>
<T id="t_3"><w pos="ZZ1">I</w> <w pos="VM">would</w> <w pos="RR">probably</w> <w
pos="VVI">play</w> <w pos="NP1">Lotto</w> <w pos="RR">too</w> <w pos="PUNC">.</w> </T>
</Text>
</Doc>

```

図 2 : ICCI 書き起こし XML フォーマット

テキスト部分は XML フォーマットで品詞タグを付与している。これはランカスター大学の CLAWS で自動品詞タグ付与を行っている。自動なのでエラー率が数パーセントあるが、人手による修正は施していない。

6. エラー・アノテーション・ツールの開発

ICCI のもう 1 つの重要なテーマは、各国の研究者のノウハウを結集して、学習者コーパスの研究環境の整備を行うことである (cf. 目的の (1c))。特に学習者データのエラー情報の効果的なアノテーション・ツールの開発が待たれており、我々もデータ収集と同時にエラー情報のより高度な解析に資するツールの開発を行おうとしている。

共同研究者であるシンガポール CRPP の Huaqing Hong 氏が取り組んでいるオープン・ソースのアノテーション・ツール MMAX2 を利用したエラータグ付与ツールの概要をここで紹介する。MMAX2 本体の情報は SourceForge のページを参照願いたい²。このツールの最大の特徴は自由にアノテーション項目を編集できること、すべてが XML 形式で処理されること、アノテーション操作は GUI ですべてマウスなどのシンプル操作で行えること、

² <http://mmax2.sourceforge.net/>

そしてアノテーション情報は元のファイルに埋め込みタグではなく、stand off なので通常困難な階層化したアノテーションやバウンダリーを複数設けた交差タグの処理などに柔軟に対応できる点にある。Huaqing Hong 氏と協議のうえ、ICCI 用にエラータグセットを独自に開発し、そのデモを 2008 年 2 月の第 1 回の国際会議で行った。また同様の発表を 2008 年 3 月に AACL2008 で行っている (Tono & Hong, 2008)。

6.1. MMAX2 の基本操作

MMAX2 の使用方法の概要を説明する。タグ付与をするテキストはあらかじめ、MMAX2 で使用可能なファイル形式 (.mmax 形式) に前処理で変換しておく。ファイルを開くと、図 3 のようなファイル操作画面とコントロールパネル (右上)、およびエラータグのテンプレート (左上) が表示される：

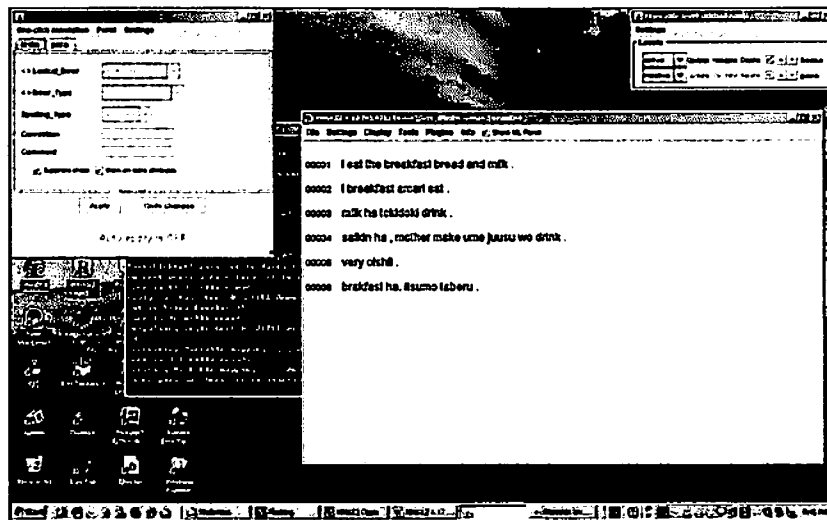


図 3 : MMAX2 起動時の画面

エラータグ付与を行いたい部分をマウスでハイライトする (図 4)：

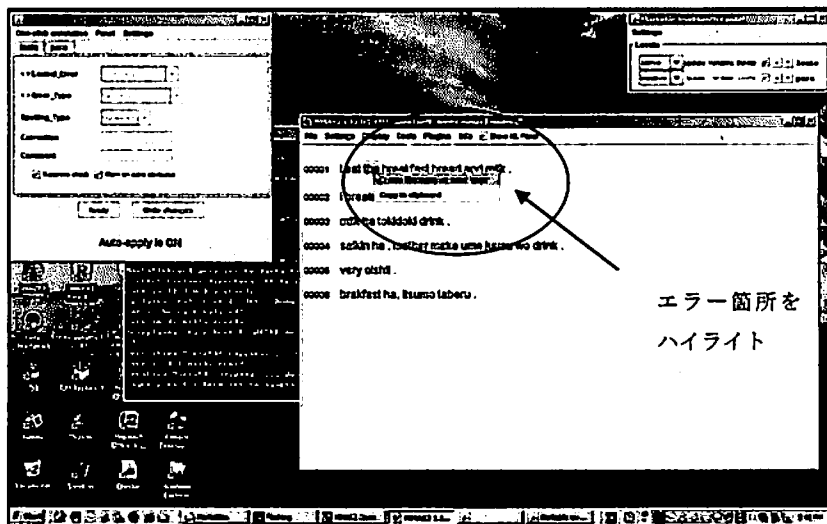


図 4 : MMAX2 エラータグ付与の手順 (1)

次いでエラータグ・テンプレートから [replacing]-[article] , correction は空欄 (余剰エラーなので訂正候補は 0 (null) になる (図 5) :

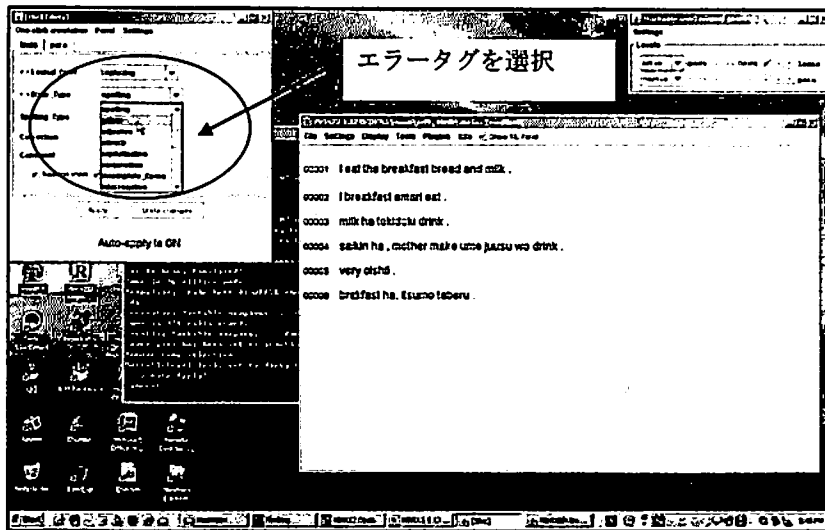


図 5 : MMAX2 エラータグ付与の手順 (2)

6.2. MMAX2 のファイル保存形式

裏側では MMAX2 は basedata として図 6 のような形式のファイルを持つ。これは最初にヘッダとテキストを分離し、ヘッダ部分はインデックスファイルとして別形式で保存し、テキスト部分のみを XML 形式で格納したものである :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE words SYSTEM "words.dtd">
<words>
<word id="word_1">I</word>
<word id="word_2">eat</word>
<word id="word_3">breakfast</word>
<word id="word_4">every</word>
<word id="word_5">day</word>
<word id="word_6">.</word>
<word id="word_7">Today</word>
<word id="word_8">I</word>
<word id="word_9">ate</word>
<word id="word_10">bread</word>
<word id="word_11">and</word>
<word id="word_12">milk</word>
<word id="word_13">.</word>
<word id="word_14">I</word>
<word id="word_15">like</word>
<word id="word_16">bread</word>
<word id="word_17">and</word>
<word id="word_18">milk</word>
<word id="word_19">but</word>
<word id="word_20">I</word>
<word id="word_21">do</word>
<word id="word_22">n't</word>
<word id="word_23">eat</word>
```

```

<word id="word_24">in</word>
<word id="word_25">breakfast</word>
<word id="word_26">rise</word>
<word id="word_27">and</word>
<word id="word_28">misosoup</word>
<word id="word_29">.</word>
</words>

```

図 6 : MMAX2 における basedata の形式

アノテーションが付与されると、このファイルの単語の位置情報をもとにエラー情報を格納した XML ファイルが作成される (図 7) :

```

<?xml version="1.0"?>
<!DOCTYPE markables SYSTEM "markables.dtd">
<markables xmlns="www.eml.org/NameSpaces/lexis">
<markable id="markable_12" span="word_28" error_type="spelling" spelling_type="japanese"
correction="miso soup" lexical_error="replacing" />
<markable id="markable_11" span="word_26" error_type="spelling" spelling_type="wrong"
correction="rice" lexical_error="replacing" />
<markable id="markable_13" span="word_24..word_29" error_type="others" o_type="word_order"
correction="rice and miso soup for breakfast" lexical_error="replacing" />
<markable id="markable_10" span="word_24" prp_type="lexical_choice_bound_prp"
error_type="preposition" correction="for" lexical_error="replacing" />
<markable id="markable_9" span="word_16..word_18" error_type="noun" correction="breakfast"
n_type="lexical_choice" lexical_error="replacing" />
</markables>

```

図 7 : エラーアノテーションを格納した XML ファイル

現在、このアノテーション・ツールの改良のために、(a) エラータグセットの見直し、(b) エラータグ付与のインタフェース部分の改善、(c) 語彙レベルから文・段落・談話レベルまでのアノテーション・スパンの拡張などを検討している。

今後、このツールの改善を3年間で行い、ICCI のデータとセットにして、部分 (または全体) をエラータグ付与を行ったものを一般公開する予定である。

7. コーパス・データの検索インタフェース

ICCI パイロット版の共同研究者間での利用の促進のため、以下の3つのインタフェースの利用を実験的に行っている: (1) Xaira によるスタンドアローンの検索、(2) Huaqing Hong 氏の自作 web インタフェース、(3) G-COE 専属プログラムによる ICCI 公式 web サイトでの検索インタフェース。それぞれ以下に簡単に解説する。

7.1. Xaira ペースの検索

Xaira³ はオープンソースの XML コーパス検索ツールである。ICCI データは Xaira を利用して、リモートサーバーにログオンする形式でも、スタンドアローンのマシンでも利用可能である。Xaira にテキストを格納する際には、付属の Xaira Index Toolkit を用いて格納する。本プロジェクトでは Huaqing Hong 氏が Xaira を用いた検索のカスタマイズを担当している。Xaira の強力なところは、XML 形式で格納されたファイルのメタ情報およびテキスト内言語情報を極めて柔軟に検索条件として利用できる点にある。

図 8 は Xaira でパイロットデータを検索した結果画面である：

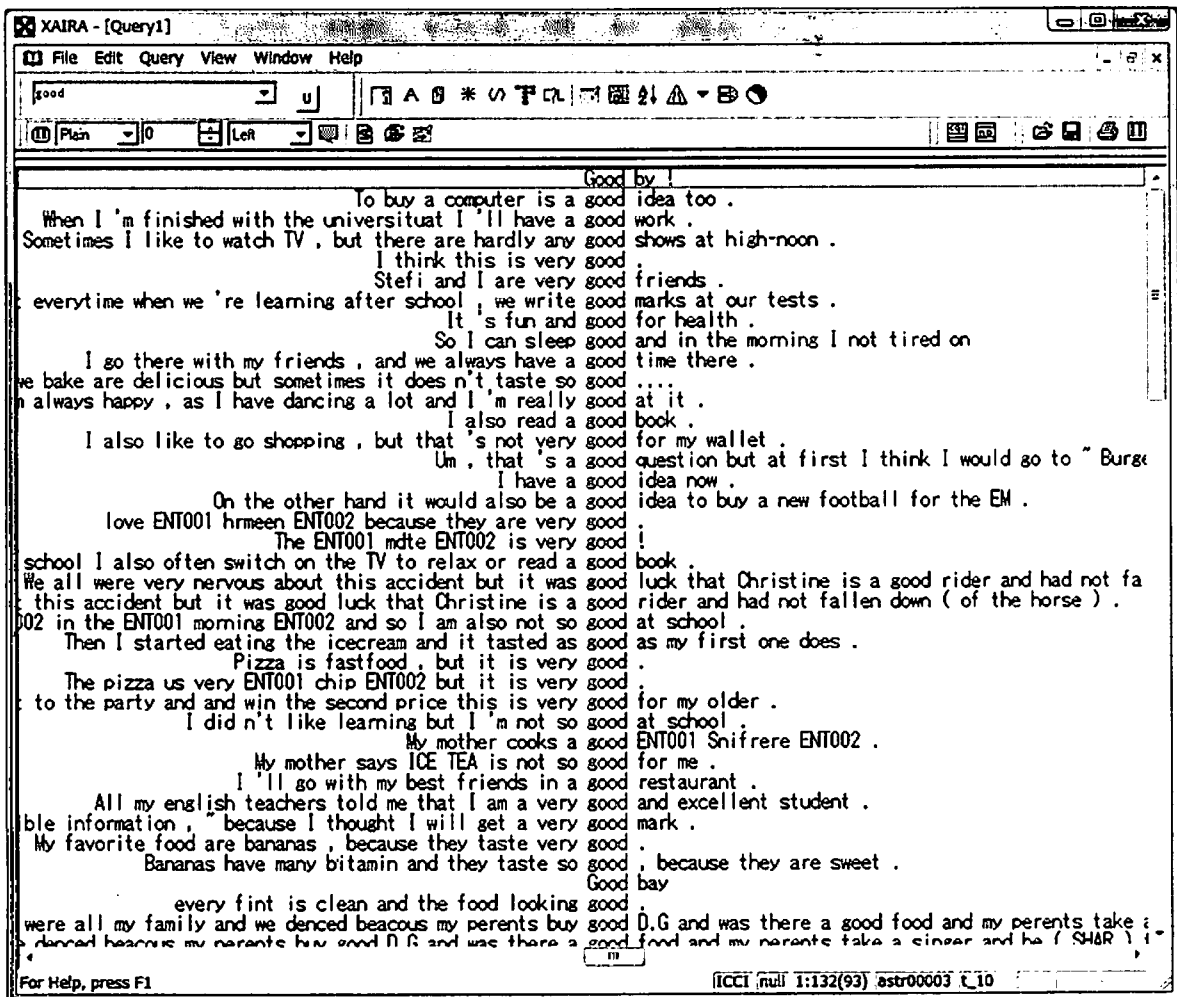


図 8 : Xaira 検索結果画面

データのヘッダに <country>, <genre>, <mother_tongue>などの情報が付与されている場合、これらを Partition として定義してやると、コンコーダンサーの検索結果を国別、ジャンル別、母語別などに区分して分布図を得ることができる (図 9, 10 参照)：

³ <http://www.oucs.ox.ac.uk/rts/xaira/>

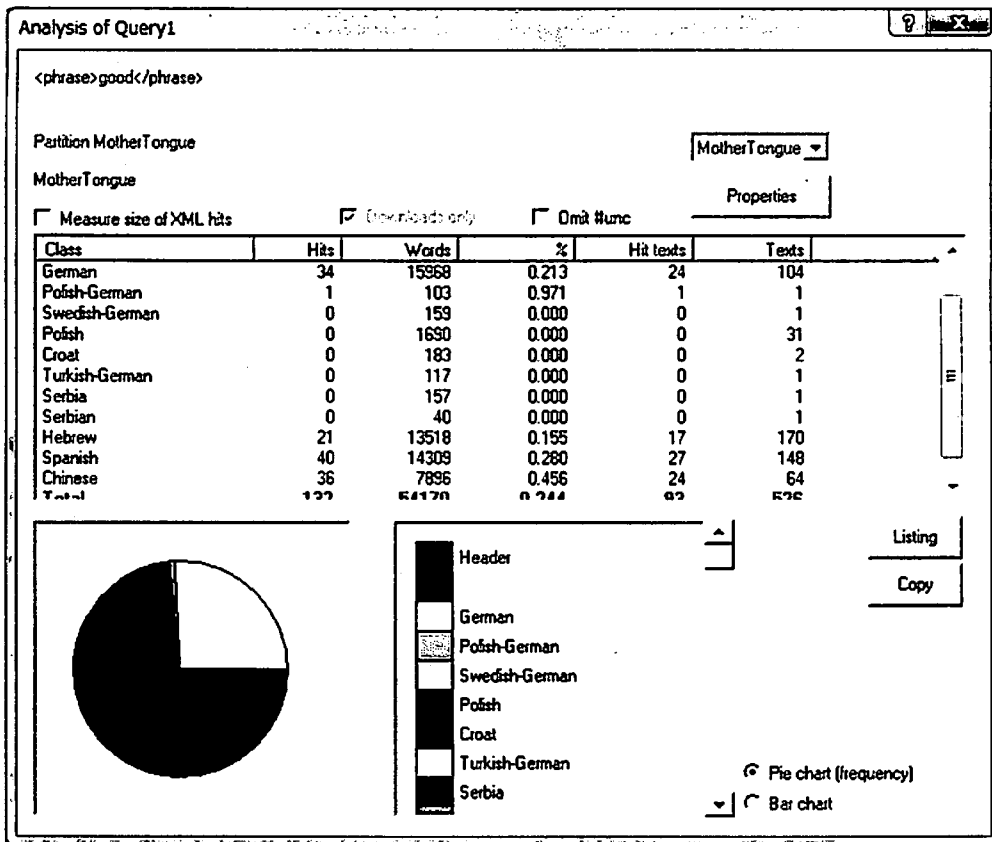


図 9 : good の検索結果 (母語別)

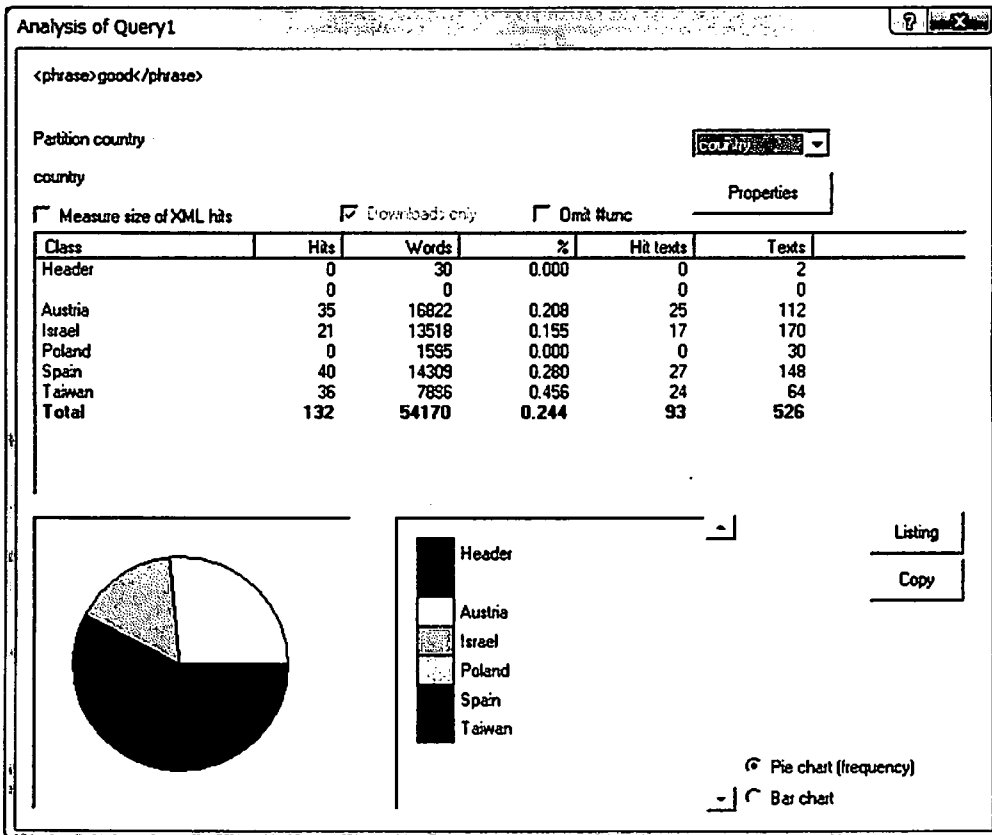


図 10 : good の検索結果 (国別)

今後はさらにデータを整備し、エラーアノテーションを充実させることで、国・母語・学年・作文テーマ（ジャンル）などを複合的にからめたコーパス検索が可能になる。

なお、Xaira はクライアントとして検索ツールのみを端末で利用し、データ自体はサーバーに置いておくことも可能である。将来的には検索の便を考えて、データ管理はサーバー側で行い、クライアント・マシンからのアクセスを想定した環境も構築したい。

7.2. Web ベースの検索

Web 検索の試用版として、2 種類のインタフェースを開発中である。1 つは、Huaqing Hong 氏が中心で開発している web コンコーダンサーである。これは氏のシンガポールにおける SCoRE プロジェクト⁴のインタフェースを援用したものである。SCoRE のインタフェースはマルチメディア・ファイル（音声・映像ファイル）のリンクが可能なコンコーダンサーで、シンガポールの学校教育における授業改善のため、授業観察ビデオをコーパス化するというユニークな試みが行われている。

Huaqing Hong 氏が ICCI 用に試作したインタフェースを図 11 に示す：

ICCI Corpus Query

There are two types of query methods available for this corpus, one is web-based query (online) query, and the other is Xaira-based (offline) query:

Web-based Query (For Online Use)

<p>Country/Region:</p> <p><input type="checkbox"/> Austria <input type="checkbox"/> Israel <input type="checkbox"/> Japan <input type="checkbox"/> Poland <input type="checkbox"/> Spain <input type="checkbox"/> Taiwan</p> <p>Year/Level:</p> <p><input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/> Unknown</p> <p>Gender:</p> <p><input type="checkbox"/> Male <input checked="" type="checkbox"/> Female <input type="checkbox"/> Unknown</p> <p>Genre:</p> <p><input checked="" type="checkbox"/> Descriptive <input type="checkbox"/> Argumentative</p> <p>Mother Tongue:</p> <p><input type="checkbox"/> Chinese <input type="checkbox"/> German <input type="checkbox"/> Hebrew <input type="checkbox"/> Japanese <input type="checkbox"/> Polish <input type="checkbox"/> Spanish <input type="checkbox"/> Other</p>	<p>Search Word or Phrase:</p> <p>.....</p> <p>Search Part of Speech:</p> <div style="border: 1px solid black; padding: 2px; width: fit-content;"> <p>[-NONE-] ▲</p> <p>Adjective <input type="checkbox"/> E</p> <p>Adverb</p> <p>Alphabet</p> <p>Article ▼</p> </div> <p>Concordance:</p> <p>Left 0 words Right 0 words</p> <p><input type="checkbox"/> Show Results in Sentence: (check this to display the results in sentence mode.)</p>
--	---

図 11 : ICCI Pilot Corpus Web インタフェース (Huaqing Hong 氏作成)

現状では単純に国・地域、学年、性別、作文ジャンル、母語で学習者データを選別し、検索語を入れてコンコーダンスや頻度表を出すようなシンプルな機能だけであるが、今後、これを品詞分布やコロケーション分布、またエラー分布などを出せるような機能を実装し

⁴ <http://score.cripp.nie.edu.sg/score/>

ていきたいと考えている。

現在、これ以外に G-COE 公式 web 上で ICCI 紹介ページを作成している。これが公開されると、パイロット・データの一部を検索できるコンコーダンスーが試用できる。現状では、まだ機能制限があり次回の報告書に詳細を紹介できる予定である。

8. 大学院生教育の実践

東京外国語大学グローバル COE の研究計画の一環として実施されている ICCI プロジェクトでは、その趣旨に鑑みて、大学院生教育との連携を強化している。すなわち、上記プロジェクト・チームに東京外国語大学の大学院博士後期・前期課程の院生たちを有機的に動員しており、2008 年 2 月に行われた最初の立ち上げのための国際シンポジウムでは、院生にツール講習会への参加やシンポジウムでの研究発表を体験させるだけでなく、上記の研究チームの前で院生が自分たちの研究の関心などを発表してアドバイスを受けるなどの機会を設けた。現在、投野研究室の主要メンバーがこのプロジェクトをサポートする形でさまざまな訓練を受けている。コーパスの設計、データ収集、電子化、フォーマット、アノテーション、検索システムのデザインと制作、さらに実際のデータ分析方法と統計処理、こういった一連のコーパス基盤の研究方法のプロセスを実地に学ぶ機会として ICCI は学生に対する実践的な研究の場を提供してきている。

9. 今後の計画

2 年目を終えようとする ICCI は現在 5 カ国(地域)、500 名余りの初級・中級学習者のデータ収集をパイロットで終えている。今後、研究チームによりパイロット・コーパスの検討が行われ、ついで本実験のデータ収集が行われる(2009 年～2010 年)。これらの計画では約 10 カ国(地域)で 4000 名程度のサンプル(各国最低 400 名、40,000～50,000 語規模)のデータを目標としている。最終的には、これらのデータを国別、学習者の学年別(英語力レベル別)に比較分析できることが期待されている。そのための最低コーパス・サイズなどを 2009 年前半に確定したいと考えている。

同時に、本稿で紹介した MMAX2 によるエラー・アノテーション・システムの整備、それらによるエラー・タグ付与を施したデータの作成、Xaira, Web 検索システム等を用いて学習者メタ情報とテキスト内情報を複合的に条件検索できるシステムを構築する。最終年度(2011 年)の初めに国際シンポジウムを開催し、最終データで可能な研究テーマを検討し、John Benjamins からの論文集の刊行を計画中である。

引用文献

Granger, S. (1998) *Learner English on Computer*. London & New York: Addyson Wesley Longman.

- Granger, S., Hung, J., Petch-Tyson, S. (2002) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. (Language Learning and Language Teaching 6). Amsterdam: John Benjamins.
- Pravec, A.N. (2002) Survey of learner corpora. *ICAME Journal*, 81-114.
- Tono, Y. & Hong, H. (2008) Towards a multi-layered & multimodal annotation model of learner corpora. Paper presented at the American Association for Applied Corpus Linguistics 2008, Brigham Young University, Provo, Utah, March 13, 2008.
- 投野由紀夫. (2008-2009). 「進化する学習者コーパス」『英語教育』, (大修館書店, 2008年4月号～2009年3月号連載) .
- 投野由紀夫. (2007). 『日本人中高生1万人の英語コーパス JEFLL Corpus: 中高生が書く英文の実態とその分析』. 東京: 小学館.