

英語コーパス学会 第 40 回大会資料

日時：2014 年 10 月 4 日（土）－ 5 日（日）

会場：熊本学園大学

(<http://www.kumagaku.ac.jp/daigaku/map/access>)

〒862-8680 熊本市中央区大江 2 丁目 5 番 1 号

■ 第1日目

ワークショップ1【小中高大の連携のための教材コーパスの構築と利用：

小学校英語ウェブコンソーダンスの利用と教材のコーパス化】

会場：熊本学園大学1号館121PC教室

日時：10月4日（土）10:00-12:00（9:30受付開始）

講師：藤原康弘（愛知教育大学）・石井康毅（成城大学）

定員：定員50名（先着順・要予約）

参加費：会員無料。非会員2,000円（当日会員としての大会参加費二日間共通）。

※予約申し込みは、電子メールで jaecs.workshop@gmail.com まで空メールを送信して下さい。返信にて参加申込WebフォームのURLをお知らせいたします。（申込締切日：9月30日）。

日時 2014年10月4日（土）
 受付開始 12:00（14号館1階「高橋守雄記念ホール」）*13:50以降の受付は11号館115B教室
 開会式 13:00（14号館1階「高橋守雄記念ホール」）

司会 田畑智司（大阪大学）

オープニングセレモニー 熊本市立必由館高校和太鼓部による和太鼓の演奏

1. 会長挨拶 堀 正広（熊本学園大学）
2. 開催校挨拶 幸田亮一（熊本学園大学学長）
3. 総会
4. 学会賞審査報告
5. 事務局からの連絡

〈研究発表第1室（11号館1151教室）〉 司会 安浪誠祐（熊本大学）

研究発表1 14:20-14:50

語彙の洗練性指標に対するテキスト長の影響：

L2学習者の課題英作文と Parallel sampling method を用いて 石井卓巳（筑波大学大学院生）

研究発表2 14:55-15:25

中学校におけるコーパスを利用したデータ駆動型英語学習の実践

—ペーパー版DDLからタブレット端末DDLまで—

西垣知佳子（千葉大学）

小山義徳（千葉大学）

神谷昇（千葉大学）

中條清美（日本大学）

研究発表3 15:30-16:00

英語CEFRレベルを規定する基準特性の抽出—文法項目の自動抽出とその評価—

投野由紀夫（東京外国語大学）

石井康毅（成城大学）

〈研究発表第2室（11号館1152教室）〉 司会 福田 稔（宮崎公立大学）

研究発表1 14:20-14:50

動詞 order に後続する要素の特徴について

西原俊明（長崎大学）

研究発表2 14:55-15:25

語形成パターンの生産性：BYU-BNC hapax による検証

森田順也（金城学院大学）

〈研究発表第3室（11号館1163教室）〉

パネルセッション 14:20-16:00

次世代 *Dickens Lexicon Digital* の開発とそれに基づく後期近代英語研究

司会 堀 正広（熊本学園大学）

講師 堀 正広（熊本学園大学）

講師 今林 修（広島大学）

講師 永崎研宣（人文情報学研究所）

〈休憩 16:00-16:20〉

シンポジウム 16:20-18:20 (11号館 1163教室)
《英語教育・研究のための教材コーパスの構築と利用：実践例と課題》

司 会 石井康毅 (成城大学)
藤原康弘 (愛知教育大学)

小学校英語ウェブコンコーダンサーの構築と利用：
教科化を見据えて 藤原康弘 (愛知教育大学)

中高英語教科書コーパスの構築と利用例 中條清美 (日本大学)

CEFR レベルに基づいた教材コーパス：
レベル別基準特性の抽出に向けて 内田 諭 (九州大学)

英英辞書の定義・用例コーパスの構築と利用例 石井康毅 (成城大学)

《懇親会 時間：18:30-20:00 場所：熊本学園大学本館グリル；会費：5,000円》

■第2日目

ワークショップ 2【コーパス研究のためのXML活用手法】

会 場：熊本学園大学 1号館 121PC 教室

日 時：10月5日(日) 9:30-11:00 (9:10受付開始)

講 師：永崎研宣 (人文情報学研究所)

定 員：定員 50名 (先着順・要予約)

参加費：会員無料。非会員 2,000円 (当日会員としての大会参加費：4日の参加費納入者は不要)。

※予約申し込みは、電子メールで jaecs.workshop@gmail.com まで空メールを送信して下さい。返信にて参加申込 Web フォームの URL をお知らせいたします。(申込締切日：9月30日)。

日 時 2014年10月5日(日)
受付開始 10:40 (11号館 115B 教室)

講演 11:10-12:10 (11号館 1163教室)
《タスク駆動型のコーパス構築と情報処理技術》

司 会 投野由紀夫 (東京外国語大学)
講 師 田中省作 (立命館大学)

〈昼休憩 12:10-13:10〉

〈研究発表第1室 (11号館 1151教室)〉

司 会 長 加奈子 (北九州市立大学)

研究発表 1 13:10-13:40
Random Forests による英語理療論文特徴語分析—Corpus of Contemporary
American English を参照コーパスとして— 八野幸子 (大阪大学大学院生)

研究発表 2 13:45-14:15
学習支援用日英例文パラレルコーパス SCoRE の構築における課題：
例文作成と訳出に焦点を当てて 若松弘子 (筑波大学大学院生)
石井卓巳 (筑波大学大学院生)
中條清美 (日本大学)

研究発表 3 14:20-14:50
日本語を母語とする上級英語学習者の誤用にみられる時制スキーマ
荒川和仁 (東京外国語大学大学院生)

研究発表 4 14:55-15:25
多読学習において学習者が感じる「難しさ」の解明：リーダー・コーパス作成と分析
加野まきみ（京都産業大学）

〈研究発表第2室（11号館 1152 教室）〉

司 会 脇本恭子（岡山大学）

研究発表 1 13:10-13:40
*Days Without End*における乖離相克する人物描写：コーパスデータから見てとれる
内的／外的ダイアローグの言語的特徴からの考察
能勢 卓（京都聖母女学院短期大学）

研究発表 2 13:45-14:15
アメリカ英語における(*the*) *chances are* (*that*)について－談話機能、伝播、文法化
柴崎礼士郎（明治大学）

研究発表 3 14:20-14:50
コーパスを使った歴史社会語用論研究の試み
椎名美智（法政大学）

〈研究発表第3室（11号館 1163 教室）〉

司 会 柳 朋宏（中部大学）

研究発表 1 13:10-13:40
学習者誤用コーパスからみる *think about* と *think of* の違い
大熊洋祐（東京外国語大学大学院生）

研究発表 2 13:45-14:15
絶対形容詞の意味シフト可能性
木山直毅（大阪大学大学院生）

研究発表 3 14:20-14:50
句構成パターンに基づく前置詞のクラスター分析－図と地の理論を用いた
前置詞句とその多義性の分析－
鎌倉義士（愛知大学）

研究発表 4 14:55-15:25
統語依存関係コーパスからの構造特性特徴量抽出
大矢政徳（目白大学）

閉 会 式 15:30（11号館 1163 教室）
閉会の辞

堀 正広（熊本学園大学）

■10月4日(土)
【ワークショップ1】

小中高大の連携のための教材コーパスの構築と利用：
小学校英語ウェブコンコーダンサーの利用と教材のコーパス化

藤原康弘(愛知教育大学)・石井康毅(成城大学)

昨年12月、文部科学省より「グローバル化に対応した英語教育改革実施計画」が発表された。その計画から、近い将来、小中高の各段階で扱ってきた語彙、表現や文法など、教育上のインプットの質と量が大きく変わることが予想されている。その英語教育の過渡期において、教科書、副教材や入試問題など、さまざまな教材をコーパス化しておくことで、そのような変化を的確に分析して柔軟に対応できることが期待される。

そこで本ワークショップでは、1) まずは現状の小学校英語を把握するために、小学校英語ウェブコンコーダンサーのハンズ・オン・セッションを実施する。現状の小学校英語を把握することは、今「小中連携」を図り、今後の小中高の英語教育を模索する上での一助となるだろう。次に2) 中学校、高等学校等の教科書、副教材、入試問題などを電子データ化する手法を体験する。どのような対象にでも対応できるように、紙媒体をスキャンした電子ファイルよりテキストデータを抽出する過程を実際に経験する。本ワークショップは、コーパスを使用したことがない方の参加を想定しており、参加に当たってデータ処理に関する高度な知識や技能は必要ない。

【研究発表第1室】

【研究発表1】

語彙の洗練性指標に対するテキスト長の影響：
L2学習者の課題英作文と Parallel sampling method を用いて

石井卓巳(筑波大学大学院生)

語彙の洗練性指標は、学習者の産出語彙の測定・評価の一観点を成す語彙の豊かさ指標の一種である。語彙の豊かさ指標は、学習者の産出語彙の多様性・洗練性を量的に示す指標であり、第二言語習得研究の分野では、学習者の熟達度、アウトプットの質、産出語彙の語彙的発達等を示す指標として用いられてきた(e.g., Daller et al, 2007; Malvern et al., 2004; Nation & Webb, 2011)。

語彙の豊かさ指標における多様性・洗練性の観点の内、テキスト内の異なり語の割合に基づく語彙の多様性指標は、テキスト長の影響を受けやすいことが予てより指摘されており(e.g. Johnson, 1944), 第二言語習得研究の分野でもテキスト長の影響の検証・検証結果に基づく指標の改良が重ねられてきた(e.g. Malvern et al., 2004; McCarthy & Jarvis, 2010)。

一方、語彙の頻度情報に基づく語彙の洗練性指標に対するテキスト長の影響を検証した研究は、未だ非常に少ない。テキスト長に対する頑健性は指標の信頼性に直結するにもかかわらず、各指標の開発者の研究では影響をほぼ検証していない上、既存の諸指標に対する影響を検証した研究もKojima and Yamashita (2014)に限られる。しかし、同研究の分析対象のテキストは、総語数が平均495語、最長902語と比較的長く、影響を検証したテキスト長も150語、200語、406～902語の3種類のみである。L2学習者の産出は、限られた時間・熟達度等により短くなることも多いため、更に短いテキスト・テキスト長における影響の検証が必要である(Koizumi & In'nami, 2012)。

そこで本研究では、語彙の多様性指標に対するテキスト長の影響を精査する際に用いられてきた Parallel sampling method という手法と、アジア人英語学習者コーパスICNALE採録の日本人英語学習者による2テーマ計136の総語数200語～300語の課題英作文を用いて、既存の語彙の洗練性指標(Beyond 2,000, AG, ATTR, P_Lex, S)に対する50語、60語……190語、200語の16種類のテキスト長の影響を検証する。分析の結果、2テーマの課題英作文で共通して、(a) Sがテキスト長の影響を最も受け難い・(b) 100語までのテキストで使用した場合、全指標が安定した結果を産出できないということが、主として明らかになった。

発表では、本研究と先行研究の結果を踏まえ、各指標をL2学習者の産出語彙の測定・評価に利用する際の注意事項等も議論する。

【研究発表 2】

中学校におけるコーパスを利用したデータ駆動型英語学習の実践 —ペーパー版 DDL からタブレット端末 DDL まで—

石垣知佳子 (千葉大学)・小山義徳 (千葉大学)・
神谷 昇 (千葉大学)・中條清美 (日本大学)

近年、コーパスを英語教育に利用する実践が活発になっている。本研究では、公立および国立大学附属中学校でコーパスを活用したデータ駆動型学習 (Data-Driven Learning : DDL) を実施した成果を報告する。これまで、発表者らは大学生や大学院生に対して DDL を導入し、DDL が語彙・文法学習に有効であることを検証してきた。それらの実践成果を踏まえ、英語学習入門期にあたる中学校において DDL を導入し、入門期の語彙・文法学習に活用したいと考えた。具体的には、普通教室で集団指導を行うことを想定してペーパー版 DDL を 4 年前に開始し、その後、モバイル機器の普及に合わせ、タブレット端末を活用する DDL へと発展させてきた。

現在、中学校の英語授業では、言葉に対する「気づき」が重要視されている。授業では、気づきを引き出すために、ターゲットの語彙や文法項目を、教師と生徒のインタラクションや教師と ALT の対話に織り交ぜて、暗示的に導入することが多い。しかし、語彙・文法の規則を明示化する段階では、教師が解説をして指導するため、気づきが活かされているとは言えない。そこで、本研究では DDL を取り入れ、生徒自身に語彙・文法の規則を発見させて、帰納的に学べるようにした。

DDL を中学校で実践するには、入門期学習者の英語力に配慮した教材と指導法が必要となる。そこで本実践では、1) 日本、中国、韓国、台湾の英語検定教科書から自作した教科書コーパスを利用する、2) 英語に日本語訳を併記して用例を示す、3) 語彙や英文のレベルを調整し、精選した用例を示す、4) 発見へと導くヒントを載せたワークシートを利用する、5) 協同学習を取り入れ、ペアやグループで発見活動の助け合いをするなどして、中学生向けの DDL を試みた。

本発表では中学 2 年生と 3 年生を対象として DDL を実践し、同一の教材を通常の方法で指導した従来型のクラスと学習効果を比較した。授業ではコンコードンスラインを紙に印刷したペーパー版 DDL を主として使用する一方、コーパス検索ソフト AntPConc を利用して、パソコンやタブレット利用の DDL も取り入れた。実践の評価には、事前・事後テスト、ルール発見テスト、質問紙調査を行った。その結果、DDL が語彙・文法知識の獲得と保持に効果があったこと、また DDL によって学習者に言葉のルールを発見する目が養われていたことなどが確認された。

【研究発表 3】

英語 CEFR レベルを規定する基準特性の抽出—文法項目の自動抽出とその評価—

投野由紀夫 (東京外国語大学)・石井康毅 (成城大学)

本研究は、ヨーロッパ言語共通参照枠 Common European Framework of Reference for Languages (CEFR) に基づき、A1~B2 の各英語力レベルを規定する基準特性 (criterial feature) となる言語特徴を抽出する研究の一環である (Hawkins and Filipović 2012)。これらのレベル別基準特性の特定は、今後、学習指導要領の改訂を踏まえ CAN-DO ベースの英語目標設定が本格化する中で、CAN-DO と言語材料を結びつけ、シラバス・教材開発の重要な基礎資料になることが期待される (投野 2013)。

特に今回の研究では以下の研究設問を設けた：

- (1) 基準特性の抽出のために、中学・高校の文法項目の一括抽出を試みるが、その適合率 (precision)、再現率 (recall) は文法項目ごとにどの程度か？
- (2) 文法項目の一括抽出を学習者データに適用した場合、通常の誤りを含まない英文に比べてどの程度の精度になるか？
- (3) 抽出した文法項目の CEFR レベル別の全般的な傾向はどのようなものか？ 現行の中学・高校の教科書と比較して ELT 教材の特徴に違いがあるか？

コーパスは、(a) CEFR レベル別の海外の ELT 教材 (語彙・表現パターン of the 列挙部分を除いて A1~B2 で約 112 万語)、(b) CEFR レベルに再分類した JEFLL コーパス (約 67 万語)、(c) 現行の学習指導要領に基づく中 1~高 1 の教科書コーパス (約 27 万語) の 3 種に対して品詞情報を付与したものを利用した。文法項目は、東京外国語大学佐野洋研究室で作成された学校文法項目リストの CQL (コーパス検索式) を変換して利用した。

各コーパスから各文法項目に該当する用例を抽出した結果、適合率は概して高い。再現率については

網羅的な評価は難しいが、特にパターンが単純な文法項目については十分に高いと考えられる（研究設問 1）。例えば、<have (+副詞) +過去分詞>というパターンで定義される「完了相」の項目を検定教科書で検索すると、中 1・中 2 では 1 例もなく、中 3 で初めてマッチし、さらに適合率も非常に高い。学習者コーパスにも本研究の手法が利用できることが確認できたが、パターンが長い場合には、エラーが含まれる表現にマッチしないため再現率は低下する（研究設問 2）。ELT 教材での CEFR レベル別の文法項目の使用状況を見ると、あるレベルで使用が急増したり、レベルの上昇に従って使用が漸増したりするものが多く見られ、これらはレベル基準特性と考えられる可能性がある（研究設問 3）。

【研究発表第 2 室】

【研究発表 1】

動詞 *order* に後続する要素の特徴について

西原俊明（長崎大学）

動詞 *order* には、不定詞補文が後続できる。これまでの研究の中には、この不定詞補文は *persuade* 型と同じ目的語コントロール構造をもつという分析が存在する。本発表では、中右(1994)・Radford(1997)で指摘されている事実に加えてコーパスなどに見られる事実をもとに、問題の補文は ECM 補文と同じ構造をもつということを明らかにする。具体的には、*expect* と同じ統語的振る舞いをみせることを指摘する。また、現代イギリス英語では、補文標識 *for* を伴う不定詞補文が *order* に後続できるという事実、この形式は法律に関する文脈に限定されるという事実から、デイスコース・レジスターが補文形式を決定していることを併せて明らかにしたい。

- (1) ...Cambridge Magistrates Court also *ordered* for him to be supervised by the Probation Service for 12 months. (www.dailymail.co.uk)

イギリス英語では、(1)に示す例が可能であり、法的な判断がなされる場面で使用されている。ここで *order* の意味は、“order X such that” の解釈ではなく、“order that it be the case that” の意味で用いられている。後者と同じ意味で用いられる表現に *make an / the order* があり、(2)に示すように、この場合も *for* NP の形式をとる。

- (2) The made an / the *order* [for him to be jailed for five years].

イギリス英語では *V-an order for* NP の形式が、名詞と同形の動詞に置き換えられ、(1)の形式が可能になったと考えられる。また、*order* に直接 NP が後続する場合と *for* NP の形式では命令の受け手の解釈に相違点が見られ、法律に関わる部分でどのような意味合いをもつのか明らかにしたい。さらに、*request for* NP、*make a request (request) for* NP が存在することも併せて指摘し、*order for* NP *to do* 形式と比較し、類似点と相違点を明らかにしたい。

【研究発表 2】

語形成パターンの生産性：BYU-BNC hapax による検証

森田順也（金城学院大学）

生産性—語形成規則が新語を生み出す度合—は、形態論研究の重要なテーマの 1 つになっている。競合する接辞群の中から生産的な接辞を規則的に選定できれば、レキシコンをより簡潔に記述することが可能になる。本研究の目的は、大規模コーパスから抽出される hapax—頻度 1 の語—に基づき各種の語形成型の生産性を測定し、その理論的含意を示すことにある。

生産性の計測法に関する先行研究を概観した後で、大規模コーパスの hapax を活用した測定法 (Baayen and Renouf 1996) を妥当なものとして採用する。新語形成力の計測のためには、辞書に格納されないような超低頻度の語を当該規則がどの程度生み出せるかが重要だからである (Hay 2003)。Baayen and Renouf は生産値 (P) を、hapax 数÷トークン数とする。例えば *-ish* 派生語の hapax 数が 18、トークン数が 7200 の場合、P は 0.0025 となる。本研究では、*-ish* の P を「80 種の *-ish* 派生語の中で 18 種の新語を生み出す力」と捉えて、タイプ数を分母に据える測定法を提案する。さらに基体形に応じた生産性を重視し (Aronoff 1976)、精密生産性—特定の基体形の新語を生み出す率—を採用する。

上記の測定法に基づき、5 種の名詞形成接辞 *-ation/-al/-ment/-ity/-ness*、4 種の形容詞形成接辞 *-al/-less/-ic/-ical*、及び 3 種の動詞形成接辞 *-ize/-ify/-ate* が、計 20 種の基体形 (X-ize/X-ify など) に関してどの程度生産的かを数値化する。hapax 抽出の際は、**aization* や **bization* など特定の末尾の語を頻度順

に列挙する、BYU-BNC の語分析機能を最大限に活用する。

調査結果は 3 点にまとめられる。①総計 774 語の派生語の hapax が得られたことから当該過程の創造性が確認できる。②競合する接辞の P の比較によって、自動的に接辞が決まる部類と一方が優先される部類が判明する。(X-ify 動詞を名詞化する際の -ation の P は 0.369, -ment/-al は 0。X-ous 形容詞を名詞化する際の -ness の P は 0.341, -ity は 0.145。) ③単純語への接辞付加は生産的でない。(-ness (*sadness*) の P は 0.090。) 以上の一般化は、競合 (Competition) によって複雑語の語形が生産的に加工されるという仮説を (Embick 2010), 基本的に支持する証拠を提供する。

最後に、接辞付加に関する生産性の上記のアプローチが、複合の生産性にも適用可能であることを、4 種の項編入複合型—1) 目的語を組み込む動詞由来複合 (book possession), 2) 主語を組み込むもの (teacher collusion), 3) 補部を取り込む形容詞由来複合 (sunlight sensitivity), 4) 主語を取り込むもの (predator adaptability) —を用いて示唆する。

【研究発表第 3 室】

【パネルセッション】

次世代 *Dickens Lexicon Digital* の開発とそれに基づく後期近代英語研究

司会 堀 正広 (熊本学園大学)
講師 堀 正広 (熊本学園大学)
講師 今林 修 (広島大学)
講師 永崎研宣 (人文情報学研究所)

本プロジェクトの出発点は、山本忠雄博士(1904-91)(元広島大学・神戸大学教授)の著書 *Growth and System of the Language of Dickens: An Introduction to A Dickens Lexicon* である。本書は、1946 年東京大学より文学博士の学位を授与され、1950 年に出版された。1953 年には英語研究において最初の学士院賞を受賞された。本書は、山本博士の最終目的である Dickens Lexicon を作成するための序論にすぎず、Lexicon 作成のためにいろいろなプロジェクトを立ち上げられたが、完成を見ることなく 1991 年に他界された。その後、1997 年に山本博士の自宅から約 6 万枚の Lexicon のためのカードが発見された。そして、1998 年に Dickens Lexicon 作成の完成を目指してプロジェクトチームが結成された。

Dickens Lexicon Digital は、次のようなディケンズのイディオムに関するデータベースと様々なデータベース及び機能を搭載している。

(1) 多機能搭載型レキシコンのデータベース

- レキシコンの検索条件は、見出し語、品詞、定義、作品名、章、引用、OED の引用の有無、コメントであるが、これらは各項目に関してアルファベット順、作品順、並び替え等を行うことができる。また、コンコーダンサー、コロケーション検索機能、統計処理の機能を搭載している。

(2) 電子テキスト

- Dickens の全作品のテキストだけでなく、英国の 18 世紀、及び 19 世紀の小説の電子テキストを可能な限り網羅的に収録している。
- 検索・ソート・語彙リスト生成機能を備えたコンコーダンス、統計処理機能を有している。

本セッションでは、*Dickens Lexicon Digital* の概要、ウェブデザイン上の工夫と特徴を説明した後、Dickens のイディオム研究や英語研究においてどのような有用性があるか、通時的及び共時的な視点から実際の研究例を示しながら論じていく。

この *Dickens Lexicon Digital* は、平成 26 年中に Web 上で部分的に公開される予定である。また、本プロジェクトは平成 20 年度から科学研究費に採択され、本年度も標記の題目で平成 26 年度から 28 年度まで基盤研究 (B)で採択されている。

【シンポジウム】

英語教育・研究のための教材コーパスの構築と利用：実践例と課題

司会 石井康毅（成城大学）・藤原康弘（愛知教育大学）

学習者のアウトプットデータである学習者コーパスは近年ますます注目を集めているが、対照的に、インプットデータである教材コーパスについては構築と利用についてのノウハウは広く共有されているとは言いがたい。しかし、教材コーパスやそこから作られるデータは、教材の特徴を質的・量的に分析するなどの研究目的以外に、教育目的でも活用の幅は広い。

そこで本シンポジウムでは、教科書を中心に、広く「教材」と呼べるもののコーパスを作り、利用するという統一テーマの下で発表者が実践例を報告し、出席者による意見交換を行い、労力のかかる教材コーパス構築の意義と課題を共有することを目指したい。

なお、本シンポジウムは、教材コーパスの具体的な構築プロセスと活用を実際に体験することを目的として、同日午前開催されるワークショップと連動する。本シンポジウムとワークショップの両方に参加することで実際に教材コーパスを構築して利用するための基本的な知識と技能が身に付くことが期待される。

小学校英語ウェブコンコーダンスターの構築と利用：教科化を見据えて

講師 藤原康弘（愛知教育大学）

小学校英語ウェブコンコーダンスター（Bora & Fujiwara, 2012-）とは、小学校英語に関連する教材・指導書などのテキストをウェブ上で検索できるシステムである。賛否はあれど、小学校英語は 2002 年、選択領域の「英語活動」として導入され、2011 年、必修領域の「外国語活動」へと発展し、2020 年には教科化されると予想されている。しかしながら、今のところ週に 1 度の授業に過ぎないため、中高の教員内でさほど意識が高まっていないとの指摘もある。実際に文部科学省の「小学校における外国語活動の現状・成果・課題」（2014）には「中学校においては外国語活動を踏まえた指導が不十分である」と指摘されている。今「小中連携」を図り、今後の小中高の英語教育を模索する上で、小学校英語の現状を把握することは必要であろう。そこで本発表では、このデータベースを作成する際に遭遇した問題、とくに曖昧タグの修正についてふれ、次に構築したコーパスを具体的に教育・研究へ活用する事例を紹介する。

中高英語教科書コーパスの構築と利用例

講師 中條清美（日本大学）

18 歳人口の減少にともない「大学全入時代」を迎えた大学において、学力低下の指摘される新入生が、中学・高等学校の英語教科書でどのような語彙・文法・コロケーション、あるいは談話構造などを学習しているかを知ることは、英語教育を効果的に行う上で重要な要素と言える。実際に、中学および高校教科書の教科書本文を、入力、校正して教科書コーパスを構築し、コーパス分析ツールを使用して出現語彙、特徴語、コロケーションなどを抽出してみると、使用する教科書によって質的・量的に差があることなどが具体的に見えてくる。このような資料は、英語教科書の選定、英語教材の作成、小・中・高・大の英語教育連携の検討の際の客観的基礎データとして役立つものと考えられる。

本発表では、まず中村他（2008）、中條他（2008）、赤野他（2014）などの先行研究を概観して英語教科書をコーパス化する際の留意点を説明し、続いて、コーパスデータの分析例としてさまざまな言語資料との語彙比較などの分析結果を紹介する。

CEFR レベルに基づいた教材コーパス：レベル別基準特性の抽出に向けて

講師 内田 諭（九州大学）

近年、学習評価指標として「言語を使って何ができるか」という機能的な側面が重視される傾向がある。その流れの中で CEFR（Common European Framework of Reference for Languages）が利用されること

が多くなり、それに準拠した教材の数も増えてきている。その一方で、CEFR の指標は、CAN-DO リストが中心となっているため、語彙や文法などの形式的な言語特徴との関連については十分な研究の蓄積があるとは言い難い。本発表では、「CEFR レベル別の言語的な基準特性を明らかにする」ことを目的として現在構築中である教材のコーパスの概要と、その利用例を示すことを目的とする。学習者にとって重要なインプットである「教材」は、語彙や文法項目などの面で意図的な統制が行われており、汎用コーパスを作成する場合は異なる点に注意を要する。そのような点を可能な限り一般化して示し、レベル別の言語的な基準特性を抽出するためのケーススタディとその結果を紹介する。

英英辞書の定義・用例コーパスの構築と利用例

講師 石井康毅（成城大学）

隅から隅まで読んでいく教科書とは異なり、辞書は必要に応じて参照するものであるが、学習者にとって重要なインプットデータであるという点は教科書も辞書も同じである。また、母語話者のデータを集めた汎用コーパスの英文は学習者には難し過ぎて教育目的には適さない場合が多いが、学習者向け英英辞書の用例は難度もそれほど高くなく、適度に短く、なおかつ英語として信頼できるものであるため、教育目的で使いやすい。発表者はこのような点に着目し、通称“Big 5”と呼ばれる主要な上級学習者向け英英辞書 5 点の全ての定義と用例（約 900 万語）をコーパス化し、検索できるシステムを作製した。また、この過程で得られたデータはコーパスとして検索・閲覧するのみならず、直接分析することもできる。

本発表では、まずデータ取得とコーパス化の概要を説明する。その後、コーパスとしての利用例と、データの分析例として各辞書の定義・用例を対象とした語彙レベルの比較などの分析結果を紹介する。

■10月5日(日)
【ワークショップ2】

コーパス研究のためのXML活用手法

講師 永崎研宣(人文情報学研究所)

近年、BNCやBCCWJをはじめとして様々なコーパスがXMLによるマークアップを施された形で提供されるようになってきている。その一方で、コーパス言語処理の手法においてはXMLのタグは特に必要でない場合もあることから、どちらかという忌避されてしまってきた感がある。ともすれば、最初にタグを削除することがコーパス利用の基本であるようになってしまう場面もある。もちろん、XMLタグを前提としない分析ツールにおいては当然なことなのだが、しかし、実際には、コーパス作成者はXMLタグそれなりの手間暇をかけて様々な情報を付加してくれているのであり、それらを活用することで新しい観点を得ることも可能かもしれない。しかしながら、XMLの各種ツールは日本語圏ではあまり普及しておらず、若干敷居が高いように感じている人が少なくないように思われる。そこで、本ワークショップでは、まずはその敷居を下げるための簡単な活用手法をご紹介したい。

また、一方で、自分のデータを蓄積していくにあたって、XMLは非常に有益なものである。XMLは多層的な検索を可能とするところに大きな特徴があり、これをうまく活用することで、蓄積したデータも様々な形で掘り起こしていくことも可能となる。XMLでデータを作成する場合の簡単な手順についても、時間が許せばご紹介したい。

ただ、講師は英語コーパス言語学の研究者ではなく仏教学のテキストデータベース構築を主な仕事としているため、残念ながら、英語コーパスそのものに深く入り込んでのご紹介というわけにはいかない。やや通り一遍のご紹介になってしまうかもしれないことは御承知置きいただきたい。それを踏まえた上で、可能であれば、BNCを使った事例をできる範囲でご紹介してみたいと考えている。

【講演】

タスク駆動型のコーパス構築と情報処理技術

講師 田中省作(立命館大学)

コーパス研究において、言語処理などの情報処理技術は、一部の言語情報の付与をのぞき、コーパスの分析手段として位置づけられることが多い。しかし、このような情報処理技術は、近年、コーパス構築においても積極的に活用され始めている。コーパス研究や情報学研究で、研究対象となる言語現象や取り扱うべき言語素性が多様化、複雑化する現在、情報処理技術を活用したコーパス構築の効率化と高度化は、ますます重要となることは間違いない。

情報学研究では、研究課題を解決する一過程として、広義のコーパスのような電子化言語データの構築が要請されることも少なくない。本講演では、このようなタスク駆動型の、比較的小規模な研究グループによるコーパス構築に焦点をあてる。情報学・言語処理を学術的基盤とする講演者とその研究グループが取り組んできた研究事例を具体的に示し、両研究からみたこのようなコーパス構築の意義、その技術的・社会的課題について考える。研究事例として、日本人の英語科学論文に含まれる言語特徴の探究のためのWebを源とした質情報付き科学論文コーパスの構築、学校文法に基づいた英文解析のための文法情報付き学参例文の収集などを取り上げる予定である。

【研究発表第1室】

【研究発表1】

Random Forestsによる英語理学療法論文特徴語分析 —Corpus of Contemporary American Englishを参照コーパスとして—

八野幸子(大阪大学大学院生)

近年、大学における実学系学部・研究科の増設が目覚ましいが、その傾向は理学療法士養成教育においても同様で、理学療法分野の英語教育へのニーズの高まりが予測される。しかし、当該分野の英語教育関連の研究は少なく、語彙に関する研究では、宮本ほか(2007)、Mitsuda(2009)、宮本ほか(2011)、宮本ほか(2012)が存在するのみであり、さらなる研究の発展が期待される。

本研究と特に関連の深い宮本ほか（2011）では、ESP語彙表の作成が行われ、一般英語コーパスと理学療法英語コーパスの特徴の調査として、BNCのLemmatized Frequency Listとの語彙の品詞別分布状況の比較が行われた。しかしこの研究では、著者ら自身も指摘するように、分析対象コーパスの元データの出版国（米国）およびコーパスを構成する個々の論文の下位分野（理学療法分野をさらに詳細に分類した、関節疾患系、神経疾患系など）に偏りがある。

本研究では、データの地域変種やジャンルの影響による文体的特性に注目した参照コーパスおよび研究手法の選定を行い、宮本ほか（2011）とは異なる角度から、英語理学療法論文の語彙の特徴について調査を行う。

分析対象コーパスとして、宮本ほか（2011）と同様の学術雑誌2誌を元データとする、発表者編纂の英語理学療法論文コーパスを用いる。参照コーパスは、分析対象コーパスの地域変種の特徴、学術文章というジャンル、さらに医療系分野である点に配慮し、アメリカ英語の大規模汎用コーパス、Corpus of Contemporary American English (COCA)のfull text版のうち、Academic Proseの集積である、サブコーパスAcademic、とその下位分類である、医学・医療系ジャーナルを元データとする、Medicineを参照する。これにより、英語理学療法論文と他の学術分野、とりわけ他の医療分野の論文との語彙の違いについて明らかにする。

また、主たる特徴語抽出方法にはRandom Forests (RF)を用いる。RFは近年の英語研究では、金・村上(2007)、小林・田中・富浦(2011)、田畑(2012)等で文章の著者の判別、科学論文の分類などに用いられた手法であるが、もとはBreiman(2001)により開発された集団学習法の一種であり(金・村上 2007)、高い分類精度を誇り、分類に加え、データ分類に貢献度の高い変数を出力できる(田畑 2012)とされている。本研究では、田畑(2012)の文学作品の分類における、RFによる判別指標語彙の抽出法を参照した。同研究は、カイ2乗検定および対数尤度比を用いた、従来の特徴語抽出法では、ある特定の作品に特徴的に生起する語を全体の特徴語としてしまうという問題点を指摘しており、この問題点をカバーするためにRFを用いている。

これらを踏まえて行った、ミニコーパス（使用データの一部を無作為抽出し作成）によるパイロット研究では、RFにより*limb*, *rehabilitation*, *motor*など、理学療法分野に関連の深い語が抽出された。また、カイ2乗検定および対数尤度比により抽出された特徴語との比較では、これら2つのいずれの方法でも1つの論文においてのみ、高頻度で出現する*kneeling*が理学療法論文の特徴語として抽出されたが、RFでは抽出されず、田畑(2012)の指摘する問題点と同様の傾向も確認された。

本発表では、英語理学療法論文と他の学術分野、とりわけ他の医療分野の論文の特徴語の違いに関して、RFの結果をもとに、質的、量的考察を行った結果について報告する。また、カイ2乗検定および対数尤度比による特徴語抽出との違いについても報告する。

【研究発表 2】

学習支援用日英例文パラレルコーパス SCoRE の構築における課題： 例文作成と訳出に焦点を当てて

若松弘子（筑波大学大学院生）・石井卓巳（筑波大学大学院生）・中條清美（日本大学）

日英例文パラレルコーパス SCoRE (Sentence Corpus of Remedial English) は、英語学習支援のため、簡潔で自然な英文とその日本語対訳を集積したものである (Chujo, Oghigian, & Akasegawa, in press)。学習者が言語使用の例に多量に触れることで帰納的学習が可能となるデータ駆動型学習 (data-driven learning; DDL) が、語彙や文法の学習において効果があることは広く知られている (e.g. Boulton, et al., 2012)。しかし、既存のコーパスに含まれる英文は、日本の教育現場で利用するには難易度が高過ぎて不向きであるため、DDL 学習に適したコーパスとして SCoRE が開発された。本発表では、SCoRE 構築における課題のうち、例文作成ならびに訳出に焦点を当てて報告する。

SCoRE は、プロファイリング例文表示システム (Chujo et al., in press) に搭載され、文法項目ごとに英語例文を提示できる特長を持つ。英語例文については、学習者の不得意とする文法項目とそれに対応する高頻度動詞等のキーワードを選定したうえで、L2 教師の経験を持つ英語母語話者が米国の低学年用リーディング教科書等に含まれる英文を情報源にして英文を作成するという手順を踏んだ。そして、初級・中級・上級者用にレベル分けをした英語例文に、日本語対訳を付与した。

簡潔で自然な英語例文の作成および日本語対訳の付与は、どのような文が学習者に利するのかを追求し続ける作業であったと同時に、学校文法の範疇を超えて、言語運用上の語用・文法・語彙等の認知的な日英の差に関する気づきや考察を伴うものであった。例えば、口語的な表現や指示詞の後方照応を用いた文を例文に含めるか否かの判断；*come/go* が「来る」・「行く」に対応しないことに見られるような視座を置く場所の差異；*man/woman* が「男」・「女」だけでなく「男性」/「女性」にも訳出できることに

見られるような語レベルでの意味範疇の差異；簡潔で自然な表現を優先すると対応する品詞が異なってしまうこと（「彼女は登山を趣味に持つ女性です」vs.「彼女は趣味が登山の女性です」）などである。これらを考慮し、例文作成者に英語例文の変更を求めることがしばしばあった。日本語対訳については、方針をなるべく統一しつつ、英語と日本語の品詞が対応するよう心掛けた。

SCoREは、すでに授業で使用された実績もあり、教材等の作成や作問への活用も見込まれる。SCoREには受動態や仮定法等の文法項目7つに分類される著作権フリーの例文とその対訳が収集されており、今後さらに拡充予定である。

【研究発表 3】

日本語を母語とする上級英語学習者の誤用にみられる時制スキーマ

荒川和仁（東京外国語大学大学院生）

本研究では、東京外国語大学国際日本研究センターのプロジェクト、「日本語学習者の母語・地域性を踏まえた日本語教育研究」において収集された英語エッセイデータを活用し、日本語を母語とする上級英語学習者の持つ英語時制スキーマについて分析した。このエッセイデータを使用した研究に望月・狩野（2012）があるが、本研究ではその内 2011 年度のデータ（執筆者は学部 1 年生 70 名、338 提出エッセイ、121,999 語）を使用した。エッセイのテーマは学生生活や大学祭といった学生に身近な事柄や環境問題等で、特定の文法項目の産出を促すものではない。

研究対象は、Celce-Murcia ら（1983）の分類（現在、過去、未来の 3 テンスと、単純、完了、進行、完了進行の 4 アスペクトの組み合わせ）に基づく時制表現である。ファイル比較ソフトを使用し、提出されたエッセイと添削後のエッセイを比べ、記述の異なる箇所を誤用としたところ、295 箇所の時制表現の誤用が見つかった。

誤用の分析では、フランスの大学生のエッセイにおける時制の誤用を分析した Granger (1999)を参考に、まず添削前後の文法項目を基にしたカテゴリー分類を行った。これにより、見つかった誤用は 22 のカテゴリーに分類され、さらに、誤用数の多い上位 5 つのカテゴリーに 7 割以上の誤用が含まれると分かった。すなわち、単純現在—単純過去（約 21%）、単純過去—単純現在（約 17%）、単純過去—現在完了（約 13%）、単純現在—現在完了（約 13%）、単純現在—未来（約 9%）の各カテゴリーである。Granger (同上)の指摘のように、単純現在、あるいは単純過去の誤用が大半を占めるが、使用頻度の高さの影響も考えられる。

次に、日英語の時制表現を対照し誤用の要因を検証した。例(1)や(2)のような単純過去—単純現在の誤用は、行為・出来事の結果の残存を表せる日本語の過去時制（動詞のタ形）の影響により起きると考えられる。

- (1) They *were* (→*are*) also deprived of their habitats.
- (2) Since then, I *came* (→*have come*) to be more positive to everything.

(1)は環境問題の現状を述べるという文脈上、英語では単純現在が適切だが、日本語の捉え方の影響で単純過去を使用してしまう。過去から現在までを含んだ事態を表すときも単純過去を使用してしまう(2)のような誤用も見られた。

現在及び未来時制の区別のない（どちらも動詞のル形で表せる）日本語と、主動詞の時制との相対的な関係で従属する動詞の時制が変わりうる英語の違いによる誤用も見られた。

- (3) Some creators hope that more people *get* (→*will get*) to know their works, so they provide trial versions on the Internet.

本研究から得られる英語時制表現の教育的示唆として、5つのカテゴリーにおける時制ペアの違いを認識させる指導の必要性が挙げられる。また、日英語の文法の違いにも留意し、特に誤用の多い単純現在及び単純過去に対する教材・教授法を開発することが考えられる。

【研究発表 4】

多読学習において学習者が感じる「難しさ」の解明：リーダー・コーパス作成と分析

加野まきみ（京都産業大学）

本発表では、京都産業大学で学生の英語力向上のための取り組みの一つとして実施している「多読学

習プログラム」で、学生がリーダーを難しいと感じるのはテキストのどのような特徴によるものなのか、コーパス言語学的アプローチにより特定することを目指す。本プログラムで学生が読むのは、レベル分けされたリーダーで、英語学習者向けの Graded Readers (GR) とネイティブスピーカー用学習書 (児童書) である Leveled Readers (LR) の 2 種類がある。Claridge (2005) によると、GR はオリジナルのストーリーを保ちつつ、低頻度語やコロケーションの置き換え、文の長さの調整、文体の均質化などのリライト作業が行われた後出版される。一方、LR も段階的に難易度が増すよう、語数、低頻度語、文の長さやストーリーの複雑さ、テーマ・ジャンルなどが調整されている。学生は自分に合ったレベルの本をあらかじめ設定された語数まで読むことが課される。学生はどちらのリーダー・シリーズからでも読む本を選択できるが、読書後のこの二種類のリーダーに対する学生の反応は大きく異なる。多くの学生が LR を読むことに「難しさ」を感じると報告するのである。学生が感じる「難しさ」の要因には、Claridge (2005) が GR の書き換えの対象としてあげている低頻度語や文の長さ・複雑さ、文法的な難しさなどが LR には含まれると推測される。また、文化的背景知識の不足や、含まれるスラングの多さなども要因となり得るであろう。GR と LR におけるこれらの違いを客観的に示した研究はない。

そこで本研究では、英語学習者の視点から LR の言語的特徴を GR と対照しながら明らかにすることにより、多くの学生が感じる LR の「難しさ」を解明することを目的とする。まず、実際に学生によく読まれている GR と、それと同等のレベルに指定されている LR を選び、2 種類のコーパス (Graded Reader Corpus と Leveled Reader Corpus) を作成し、Compleat Lexical Tutor サイト上で利用できる様々なコーパスツールを用いてこれらと比較し、この二つのコーパスの間にある差を統計的に示す。例えば、VocabProfilers ツールを用いて、レベル別の語彙の割合や語彙密度を示す、Range ツールを用いて両コーパスで使用される語彙の頻度を比較する、n-gram ツールを用いて、頻度の高いコロケーションの違いを分析するなどがある。その他、文の長さ、文の構造の複雑さなどの側面を比較することで、同程度のレベルとされる 2 種類のリーダーの間にはどのような違いがあるのかを明らかにし、どのような要素が「難しさ」に繋がっているのか検証する。

【研究発表第 2 室】

【研究発表 1】

Days Without End における乖離相克する人物描写：

コーパスデータから見てとれる内的／外的ダイアログの言語的特徴からの考察

能勢 卓 (京都聖母女学院短期大学)

Days Without End (1934 年初演) において Eugene O'Neill は、仮面と二重俳優の手法を用いることによって主人公 John Loving を素顔の John と仮面の Loving に分裂させて舞台上に登場させた。*Days Without End* はその内容に関しては極めて厳しい批評がなされた一方で、この作品で取り入れられた実験的演劇手法に関しては多くの研究者の関心を引き、特に仮面と二重俳優の手法を導入したことにより登場人物の内面が巧みに描き出されている点や、素顔と仮面の乖離が最も鮮明に舞台上に描き出されている点が先行研究において評価されてきた (Tiusanen, 1968; Waincott, 1988; Eisen, 1994)。

この *Days Without End* は作品としての問題点や演劇手法に関して様々な議論がなされてきたのではあるが、その台詞の文体面に関してはこれまで十分な研究がなされてきたとは言いがたい。しかし主人公が素顔と仮面の二つに分裂して舞台上に登場した結果、素顔の John と仮面の Loving の間で取り交わされる、独白とは異なる内的ダイアログと、John (または Loving) とその他の登場人物の間でなされる外的ダイアログがこの作品には存在する。そしてこの二種類のダイアログの展開の中で主人公の乖離相克する人物像が鮮明に描き出されたという点で、両ダイアログを構成する台詞の言語的特徴には興味深いものがある。John と Loving の間でのみやり取りされる内的ダイアログでは両者の直接の対話を通して本心の葛藤する様子が台詞の展開の中で明示的に描き出され、他方対話の第三者が参与している外的ダイアログでは、John と Loving の発言における表現の差異が対照的に浮かび上がるように工夫されている (例えば両親の死という出来事に関して、John は“they died”と表現する一方で、Loving の台詞では、“parents were killed”と否定的意味合いが付加されるように語彙の選択がなされている)。

そこで本発表において、文体分析の為に作成した *Days Without End* のコーパスから得られる語彙や文構造のデータを用いて、内的／外的ダイアログを構成する台詞の言語的特徴の分析を行い、相反する個性を体現する主人公 John と Loving が発話する台詞にどのような言語的工夫が施され、その結果両者のト書きに描写されている登場人物の様々な側面がダイアログの展開の中で台詞にどのように反映されているかに関して考察を加えていく。

【研究発表 2】

アメリカ英語における *(the) chances are (that)* について—談話機能, 伝播, 文法化

柴崎礼士郎 (明治大学)

本発表では, 後期近代英語から現代英語にかけて発達してきた *(the) chances are (that)* の談話機能を文法化の視点から考察する。

The Corpus of Contemporary American English 1990-2012 (COCA), The Corpus of Historical American English 1810-2009 (COHA) および The Santa Barbara Corpus of Spoken American English (SBCSAE) を中心的に用い, 直近のアメリカ英語史における *(the) chances are (that)* を調査すると, (1) および (2) のような例が多数存在することが分かる。該当箇所には下線を施してある。

(1) UNIDENTIFIED-FEMAL: Well, they said five to eight days is usually the case. But we don't really know.

Hopefully, we will find out today. Maybe this will be all over today for us.

(END-VIDEOTAPE) COOPER: Well, chances are, though, it will be weeks or months before the world can say the same. (2009: COCA, SPOK, *CNN Cooper*)

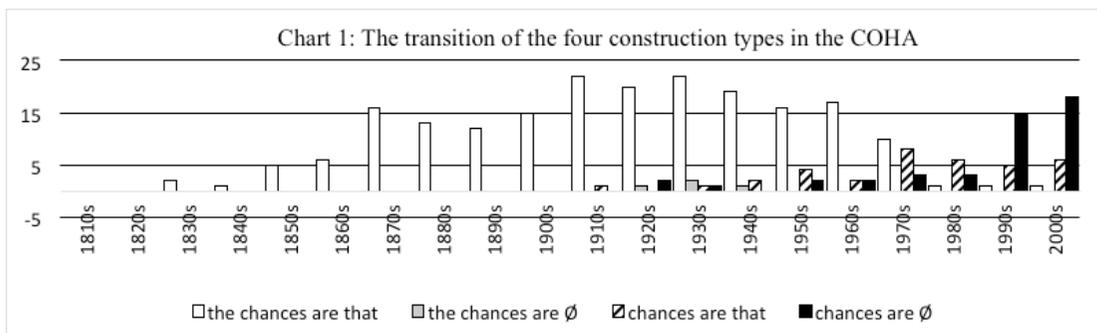
(2) CHEN: But there are definitely different standards for different types of industries. For example, if a broadcaster like myself were to curse, legally, I would—I could—this network could be fined. But if you're in a two-person home office and there's cursing going on, chances are that's not going to go anywhere legally. Right? (2007: COCA, SPOK, *CBS Early*)

(1) では, 対話者の意見に対して, キャスター (Cooper) 自身の意見を明示する直前に *chances are* が用いられている。(2) ではトピック提示とも言える条件節の後に発話者 (Chen) の見解が続く情報構造であるが, *chances are* は発話者の見解である帰結節の直前に使用されている。

本稿で用いたコーパスに基づく調査結果では, 全ての事例で (3) の情報構造が確認できた。換言すると, *(the) chances are (that)* は発話者の見解 (speaker statement) の直前に生起し, 話題導入に特化した機能を果たしている (cf. Aijmer 2007; Hopper and Thompson 2008; Curzan 2012)。

(3) [*(the) chances are (that)* + speaker statement] discourse

前掲の COHA, A Corpus of Late Modern English Texts extended version 1710-1920 (CLMETEV) および *The Oxford English Dictionary (OED)* に基づく調査分析では, *(the) chances are (that)* の本格的発達は 19 世紀初頭以降である。CLMETEV および *OED* での用例は少ないため, COHA に基づく史的変遷を Chart 1 に示す。1960 年代までは *the chances are that* が優勢であるが, その後急速に使用頻度が下がっている (縦軸は素頻度)。一方, 無冠詞の *chances are that* が 1970 年代と 1980 年代に優勢となるが, 1990 年代以降には, 冠詞および補文化辞を伴わない *chances are* が他を圧倒している (COCA の場合にも *chances are* の使用頻度が最も高い)。使用頻度が高まるにつれて *the chances are that* > *(the) chances are (that)* > *chances are* と統語的縮約を提示してゆく過程は文法化の観点からも自然である。また, (1) のように *well* や *though* などの談話標識と伴に使用される場合も多く, 発話者の見解の直前に生起する談話機能を鑑みると, 語用論標識 (pragmatic marker) として慣習化されつつあると判断できる。



アメリカ英語における *(the) chances are (that)* の諸特徴を, British National Corpus (BYU-BNC) および The Corpus of Global Web-Based English (GloWbE) を用いて比較すると興味深い分析結果が見えてくる。アメリカ英語では, 無冠詞の *chances are (that)* が定冠詞付き *the chances are (that)* よりも強く好まれる傾向にあ

り、同傾向は世界各地の英語でも確認できる。しかし、イギリス英語は特徴的であり、*chances are (that)* と *the chances are (that)* が大差なく使用されており、特に定冠詞と補文化辞を伴う *the chances are that* の用法が高頻度で堅持されている点は、世界の英語の中でも特徴的である (Mair 2009; Aijmer 2013; cf. 藤井 2006)。

【研究発表 3】

コーパスを使った歴史社会語用論研究の試み

椎名美智 (法政大学)

本研究は、17,18 世紀の初期近代英語期の口語表現を集めた Socio-Pragmatic Corpus を量的・質的に分析し、過去の言語状況を観察しようとする試みである。このコーパスは裁判調書と裁判記録、そして喜劇が集められた約 24 万語の小さなコーパスだが、各発話における会話者の社会的・語用論的な属性がアノテーションとして付与されている点に特徴がある。特定の言語的な事象の発現状況に注目し、量的分析ではコーパス内での全体的な使用傾向を、質的分析ではその典型例と例外的用法を観察していく。そこで出てきた言語現象を、話者の社会言語学的属性や会話状況の語用論的特徴との関連で説明しているとする。

今回は発表者自身が作成に協力したコーパスに独自にアノテーションを付与したオリジナルコーパスにおける「呼びかけ語」の量的・質的分析を研究例として紹介するが、こうした研究は歴史言語学、語用論、社会言語学という学問分野が重なる領域で、近年は「歴史社会語用論」という新しい分野として認知されつつある。歴史社会語用論は歴史語用論から派生したもので、過去の発話における言語的規則性を語用論的視点から調べると同時に、そうした慣習や規則がどの程度、当時の社会に浸透していたのか、その社会的・時代的な文脈を社会言語学的視点から探ろうとする研究分野である。

最近の傾向としては、研究者自身が自分の研究テーマに合った独自のコーパスデータを作ったり、既存のコーパスに独自のアノテーションを付与したり、独自の検索ソフトを開発したりすることが行われている。歴史的データを扱う言語学者がコーパス言語学に興味をもち、応用しようとしているという意味で、当該研究分野におけるコーパス言語学の貢献度は高い。しかし、コーパス構築に詳しい言語学者が歴史的データに興味をもって、研究分野を進展させるという逆方向の動きは、それほど大きいとはいえない。コーパス言語学の知見や応用の度合いが、分野によって格差があるからではないだろうか。

そこで、本発表ではコーパスを使っての言語分析をなさっている研究者の方々に、こうした歴史的データを扱う研究における動向を知ってもらい、さらなるコーパス利用の提案を頂いたり、共同研究にお誘いできればと考えている。領域横断的な共同研究によって、相互の分野の問題やメリットを共有し、ともに発展する可能性を探るきっかけとしたい。

【研究発表第 3 室】

【研究発表 1】

学習者誤用コーパスからみる *think about* と *think of* の違い

大熊洋祐 (東京外国語大学大学院生)

本発表は 2011 年度、2012 年度の 2 年間東京外国語大学英語科 1 年生によって書かれたエッセーを集め構築されたコーパス(国際日本研究センター支援:学生数 153, ファイル数 1,656, 添削前総語数 279,001, 添削後総語数 283,186)の中に存在する誤用のうち、*think about* と *think of* の誤用について分析したものである。両者の違いについては Vendler (1967) や小西 (1974) も述べているが、本コーパスにおいて *think of* とすべきところを *think about* を用いていた例が 5 件、*think about* とすべきところを *think of* にしている誤用例が 9 件見られた。両表現とも英語学習の初期の段階で学習するにもかかわらず、日本語母語話者にとっては両者の意味の違いが必ずしも明確に理解されていないということが明らかになった。

エッセーを添削した英語母語話者によると *think about* は「ある一定期間考える」ことを意味する一方、*think of* は「何かを思い出す」ことや「何かを突然思いつく」ことを意味するという違いがあるという。これらの違いの重要な点は、*think about* は「一定期間」という時間の幅が含まれているのに対し、*think of* は時間の幅という概念が含まれていないということである。

think about が時間の幅を有するのは、*about* が本来は空間(空間移動)を表す前置詞であるからだと考えられる(Radden 1989, Dirven 1993, Lindstromberg 1998)。つまり *think* の主体(動作主)が *think* という動作を行い、その考えるという意識が対象に向かって移動するように働きかけると認知されるのである。ただし頭に思い描く対象は必ずしも明確に定まってイメージされるわけではない。その不定性を表すの

が、対象を中心に「さまざまな方向への不定の移動: indeterminate movement taking any direction (Radden 1989: 570)」という空間的イメージを持つ *about* なのである。そして対象に向かって移動するには一定の時間を要するため、「一定期間考える」という時間の長さを含むと理解されるのである。また時間をかけて考えるため、「(物事を)深く考える=*consider*」の意にも使用することができる。一方 *think of* に時間の幅という概念が含まれていないことに関しては、*of* が元来は分離を表す前置詞に由来しているからであり (Dill 1989, 宗宮 2012), そこには時間的な幅が生じず、瞬間的・直観的な認識として捉えられているのである。それが英語母語話者の述べた「突然思いつく」といった表現に結び付くと考えられる。Vendler (1967:111) も *think of* を到達動詞(achievement verb)に分類している。

BNC でも *think about* は *carefully* (11 例), *too much* (7 例), *a lot* (4 例), *seriously* (3 例) のように時間の幅や程度の深さを表す副詞や *for* 前置詞句(21 例) と共起する傾向が高かった一方, *think of* にその傾向はみられなかった。

以上のように *think about* と *think of* の意味の違いは *about* や *of* という前置詞の持つ空間性及び時間性の有無の違いに由来するということが明らかになった。そして学習者はこれらの違いを正しく理解しておらず, そのため誤用が生じているということが明らかになった。

【研究発表 2】

絶対形容詞の意味シフト可能性

木山直毅 (大阪大学大学院生)

本研究は、英語の絶対形容詞 (absolute adjectives) が *very* などの修飾語と共起しやすい語と共起しにくい語のパターンを明らかにするものである。具体的には、例えば *very possible* と *very impossible* という対義語において、前者は高頻度に生じる一方で、後者は低頻度である。本研究は、なぜそのような分布を見せるのかを明らかにする。

Kennedy and McNally (2005) と Kennedy (2007) は、段階的形容詞 (Gradable adjectives) を相対形容詞 (=1) と絶対形容詞 (=2) の 2 つに分類した。この差は、形容詞が *completely* のような最大を含意する副詞と共起するかどうかで捉えられる。即ち、形容詞が終点 (endpoint) を持つならば絶対、持たないならば相対となる。

- (1) 相対形容詞 (relative adjectives)
 - a. Her brother is completely ??tall/??short.
 - b. The pond is 100% ??deep/??shallow.
- (2) 絶対形容詞 (absolute adjectives)
 - a. The room was 100% full/empty.
 - b. The figure was completely visible/invisible. (Kennedy and McNally, 2005, 355)

終点という考え方は、Kennedy and McNally (2005) や Kennedy (2007) によると、相対形容詞は常に比較の基準 (e.g. 年齢や深さ) を必要とする一方で、絶対形容詞の解釈は、終点を軸とするため、絶対形容詞は文脈に依存しない形容詞であると述べた。絶対、相対形容詞の差は表 1 の通りである。

	相対形容詞	絶対形容詞
終点の有無	無	有
語が指すスケール上の基準	スケール上の中間地点	スケール上の終点
比較の基準	文脈依存型	文脈独立型

表 1 相対的形容詞と絶対的形容詞の特徴と違い

Kennedy らの分析は、今日のスケール研究において標準的な理論となっている。しかし、近年、絶対形容詞の解釈は、文脈独立型ではなく、文脈依存型であることを提案する議論も見られる (Rotstein and Winter, 2004; McNally, 2011; Toledo and Sassoon, 2011)。

これまで、絶対形容詞が文脈に依存するのか、文脈から独立するのかというパラダイムを概観した。では、どちらの立場が正しいのだろうか。以下では、Kennedy らの立場を「**文脈独立説**」、新しい考え方を「**文脈依存説**」と呼ぶ。

	絶対	相対	相対が生じる割合 (%)
impossible	2	305	6.51
possible	18	94	160.71
necessary	38	81	319.33
unnecessary	1	224	4.44
wrong	33	97	253.85
right	8	211	36.53

表2 絶対が相対に変化する割合

以上のことを検証するために、(1)-(2)のような、修飾語を伴う叙述用法と、*a very tall man*のように、不定冠詞 *a(n)* を伴う限定用法を BNC で検索した。次にどの形容詞が絶対形容詞かを調べるために、Collostructional Analysis を用い (Gries and Stefanowitsch, 2004), 形容詞が相対と絶対でどちらに多く共起するのかを計算した。その結果の一部が表2である。表2では、*possible* や *necessary*, *wrong* は絶対形容詞であることは認められるが、相対形容詞への意味シフトを比較的頻繁に受ける。一方で、それらの対義語が意味シフトを受けることは稀、または割合が低いようである。語によっては頻繁に意味シフトを受けるという事実に基づくと、文脈依存絶対説を採用するべきであることが分かる。

しかし、次に、(a) なぜ表2のような分布の差を示すのか、(b) 意味シフトを受けにくい語が意味シフトを受ける場合、副詞は何を強調しているのか、という2つを考慮する必要がある。(a) に関して、Sassoon and Toledo (2011) は、*full* を例に、過去の経験に基づくことを述べている。しかし、*impossible* と *possible* の場合、同様の説明は困難である。本研究は、*very* の意味を再考し、なぜ表2のような差が生じるのかという要因を探るものである。

【研究発表3】

句構成パターンに基づく前置詞のクラスター分析 — 図と地の理論を用いた前置詞句とその多義性の分析 —

鎌倉義士 (愛知大学)

前置詞の語義には言語使用者の心的表象に表れる概念となるイメージスキーマが深く関連する。例えば、容器のスキーマは *in* の概念を示し、起点・経路・着点のスキーマは *over* の概念を表すと言われる (辻 2002)。本研究では、前置詞の場所の意味から比喩的な意味までの多義を前後する名詞句の構成パターンで分類し、さらに対象となる7つの前置詞句の構成パターンをクラスター分析にて表すことが可能かを試みる。

先行研究にて、前置詞の意味は認知言語学の理論である「図と地 (trajector and landmark 以下 TR, LM)」の二つの事物の関係から説明がされている (Langacker 1987, Tyler and Evans 2003)。それに基づき、前置詞の前後に共起する名詞句の組み合わせが前置詞のそれぞれ異なる意味と関係するという仮説をたてた。まず、前置詞 *over* と名詞句からなる前置詞句が人間の認識を反映すると仮定した上で、前置詞と共起する名詞句を「有生物 (人・動物など)」、「無生物 (車・場所など)」、「抽象物 (問題・時間など)」の三種に分類した (Schönefeld 2006)。その分類例は以下ようになる。

- (1) ...and FOR SALE signs hanging dispiritedly over sagging wooden doors.
(TR 無生物) (LM 無生物) <W2F-006#025>
- (2) Now, Tony, we'll just go over the basic procedures to start your bike before you sit on it.
(TR 有生物) (LM 抽象物) <S2A-054#053>
- (3) The only remaining problem will be that it may take several years before the logic of this situation triumphs over the desire to make a quick profit.
(TR 抽象物) (LM 抽象物) <W2E-008 #041>

続いて *over* と名詞句から成るパターンが場所の意味と比喩の意味で異なるのか、もしくは場所・時間・比喩などの意味で特徴的な名詞句のパターンを示すのかという分析を行った。その結果として、前置詞と共起する名詞句のコロケーションによって前置詞の意味が特定可能であることを示した (Kamakura 2011, 2012)。

本研究では、前置詞や名詞句にタグ付けがされた ICE GB コーパス (サイズ 100 万語) より *in*, *on*, *at*, *over*, *through*, *into*, *against* それぞれ7つの前置詞と共起する名詞句を含む 100 例を抽出し、その句構成パ

ターンを検証した。前述した研究と同様に、共起する名詞句の組み合わせを三種に分類し、その組み合わせの多寡が残差分析による数値にてどのような違いを示すかに着目した。クラスター分析の結果、それぞれの前置詞の名詞コロケーションが示すパターンは前置詞のイメージスキーマと合致した。場所や時間を含め、場所的な意味から抽象的な意味まで使用される *in* と *at* は同じ句構成パターンを持つ前置詞と分類された。さらに、同結果にて *over*, *into*, *through* は同じクラスターに分類された。この結果は、それら3つの前置詞が起点・経路・着点のスキーマを共有する (Dewell 1994, Dirven 1989, Lindstromberg, 1998) という認知言語学の主張と一致する。

本研究の結果は、*form and meaning* の問題に示唆を与える (Evans 2009)。前置詞のみに注目するとその多義性は顕著であるが、その複数の語義から一つの意味に限定するのは前後に共起する名詞句とも考えられる。前置詞は動詞や形容詞に付随するものか、もしくは前置詞に続く名詞とでその意味が説明されることが多い。それに対し、図と地の関係が示すように前置詞の前後に共起する名詞句とその句構造が、前置詞の意味をどの程度まで限定するのかを本研究は説明している。

【研究発表 4】

統語依存関係コーパスからの構造特性特徴量抽出

大矢政徳 (目白大学)

本研究では、The English Web Treebank corpus (Linguistic Data Consortium 2012) に対して Stanford Dependency standard (de Marneffe, MacCartney, & Manning 2006) に基づき単語間の統語依存関係に関するタグを付与した英語コーパス (a Gold Standard Dependency Corpus for English, 以下 GSDC) (Silveira, Dozat, de Marneffe, Bowman, Connor, Bauer & Manning 2014) を用い、その単語間依存関係を客観的に把握する手法を紹介する。GSDC は、様々なジャンルの文章中の単語間の依存関係を解析するための構文解析器 (parser) の正確さの指標として構築された。GSDC 内部では、各単文中のどの単語がどの単語に依存している、どの依存タイプに属するかが人力でタグ付けされている。GSDC に含まれる依存タイプは、Stanford Dependency (de Marneffe et al. 2006) として標準化されており、構文解析器である Stanford Parser (Klein & Manning 2003) に実装されている。Stanford Parser による構文解析結果は、各単文中のどの単語がどの単語に依存している、どの依存タイプに属するかについての情報であり、これは GSDC のタグが表現する情報と本質的に同一である。GSDC に含まれるジャンルは、ブログ記事、ニュースグループのスレッド記事、電子メール、製品のレビュー、そして質問サイトの解答文である。

本研究では、コーパス中の各単文を単語のネットワークと見立て、その単語間の統語依存関係の構造特性を数値化する手法 (Oya 2013, 2014, etc.) を適用し、GSDC 中の異なるジャンルに含まれる単文の構造特性に異なる傾向がみられるかどうかを検証する。ここで構造特性値として取り上げるのは、度数中心性 (degree centrality) と近接中心性 (closeness centrality) であり、前者は単文の統語依存構造の平らさ (どの程度一つの単語に他の単語が依存しているか) を示し、後者は統語依存構造の深さ (どの程度一つの単語から他の単語まで埋め込まれているか) を示す。例えば、三つの単語からなる二つの単文 “Sarah read it” と “Read this book” とでは、前者は動詞 *read* に他の二つの単語が依存しているが、後者では動詞 *read* に名詞 *book* が依存し、さらにこの *book* に限定詞 *this* が依存している。前者は後者よりも平らな統語依存構造を持ち、後者は前者よりも深い統語依存構造を持つ。Oya (2014) などでは、単文を Stanford Parser で構文解析した結果からこれらの構造特性値を抽出していたが、構文解析器からの出力ではなく、コーパスに付与されたタグを使って構造特性値を抽出することによって、正確な値が得られることが予想される。本研究では、GSDC 内のタグをいったんすべて取り払った文を Stanford Parser で解析した結果から計算された構造特性値と、GSDC のタグに基づいて計算された構造特性値とを比較し、構文解析器による解析出力から得られた構造特性値がどの程度正確な値から乖離しているかも検証する。

《大会参加者へのご案内》

- ・ ワークショップの受付：会場の熊本学園大学 1 号館 121PC 教室前で、第 1 日は午前 9 時 30 分から、第 2 日は午前 9 時 10 分から受付を行います。
- ・ 大会受付：第 1 日（10 月 4 日）は熊本学園大学 14 号館 1 階「高橋守雄記念ホール」で正午から行います。（※なお、10 月 4 日午後 1 時 50 分以降の受付は、研究発表が行われる 11 号館 115B 教室にて行いますのでご注意ください。）第 2 日（10 月 5 日）は 11 号館 115B 教室で午前 10 時 40 分から受付いたします。
- ・ 構内での喫煙は原則できません。構内禁煙にご協力いただきますようお願いいたします。
- ・ 第一日（10 月 4 日）の昼食は、学内では 7 号館レストランとローソンが開いています。第 2 日（10 月 5 日）は 7 号館レストランは閉まっていますが、ローソンは開いています。ローソンの横には食事ができるスペースがあり、テーブルと椅子が設置されています。数十人は食事をする事ができます。また、通りを隔ててショッピングセンター「ゆめタウン大江」（24 ページに地図掲載）があり、両日とも営業しています。店内には次のような飲食店があります。大阪王将（中華レストラン）、グレンドール（ベーカリー）、ミスタードーナツ、ケンタッキーフライドチキン、たこ一番（たこ焼き）、イタリアントマト（洋菓子）、スティックスイーツファクトリー（カフェ）、モミ&トイズ（クレープ）。また、同じ敷地内に、焼肉彩畑とスターバックス（カフェ）があります。
- ・ 当日会員について：会員ではない方も、「当日会員」としてご参加いただけます（会費 2,000 円、二日間共通）。懇親会へもぜひご参加下さい。
- ・ 大会第 1 日の学術プログラム終了後の懇親会では、インフォーマルな雰囲気の中で、参加者同士さまざまな意見交換、情報収集ができる場です。大会ご出席の方々には、ぜひ奮ってご参加いただけましたら幸いです。熊本ならではの料理をはじめ、ソフトドリンクやビールに加え、地元の美味しい焼酎、日本酒、ワインも用意していただいているとのことです。どうぞご期待ください。なお、会場準備の都合で、ご参加予定の方には事前の予約をお願いしております。ご協力のほどよろしくお願い申し上げます。

- ・ 英語コーパス学会第 40 回大会・懇親会
- ・ 日時：10 月 4 日（土）18:30-20:00
- ・ 場所：熊本学園大学本館グリル
- ・ 会費：5,000 円

※懇親会参加ご希望の方は、**9月21日（日）までに**電子メールで jaecs.reception@gmail.com まで空メールを送信して下さい。自動返信にて参加申込 Web フォームの URL をお知らせいたします。（申込締切日：9 月 30 日）。

- ・ 懇親会の後、参加者同士でさらに情報交換をしたいという方々のために、大学の直ぐ近くにあるスペイン料理店「papa's & mama's」を用意しています。オーナーはフラメンコ教室も開いておりますので、当日はオーナーの生徒さんたちによるフラメンコの踊りも楽しんでいただけます。会費は 2,500 円です。スペインのワインとタパス料理を味わっていただけます。ご希望の方は、懇親会の申し込みの際に 2 次会参加とご記入ください。

◆熊本学園大学までのアクセス◆



JR鹿児島本線 熊本駅より

【車利用】 約15分 【熊本都市バス利用】 3番のりば（白川口）約20分 ◎第一環状線（大学病院回り）バス停「大江渡鹿」下車 徒歩約5分 ◎中央環状線（大学病院回り）バス停「学園大前」下車すぐ

熊本交通センターより

【車利用】 約15分 【熊本都市バス利用】 約20分 ◎子飼渡瀬線（こかいわたるぜせん）バス停「学園大前」下車 ◎大江城西線（おおえじょうせいせん）バス停「学園大前」下車 ◎渡鹿長嶺線（とろくながみねせん）バス停「大江渡鹿」下車 徒歩約5分 【産交バス利用】 約20分 ◎森都（しんと）病院経由バス停「大江渡鹿」下車 徒歩約5分 ◎託麻原本通り経由 バス停「大江渡鹿」下車 徒歩約5分

JR豊肥本線 水前寺駅より

【同駅北口より徒歩】 約10分 【同駅北口より熊本都市バス利用】 約3分 ◎（大江城西線）バス停「学園大前」下車

熊本市電

◎電停「味噌天神前」より 徒歩約15分

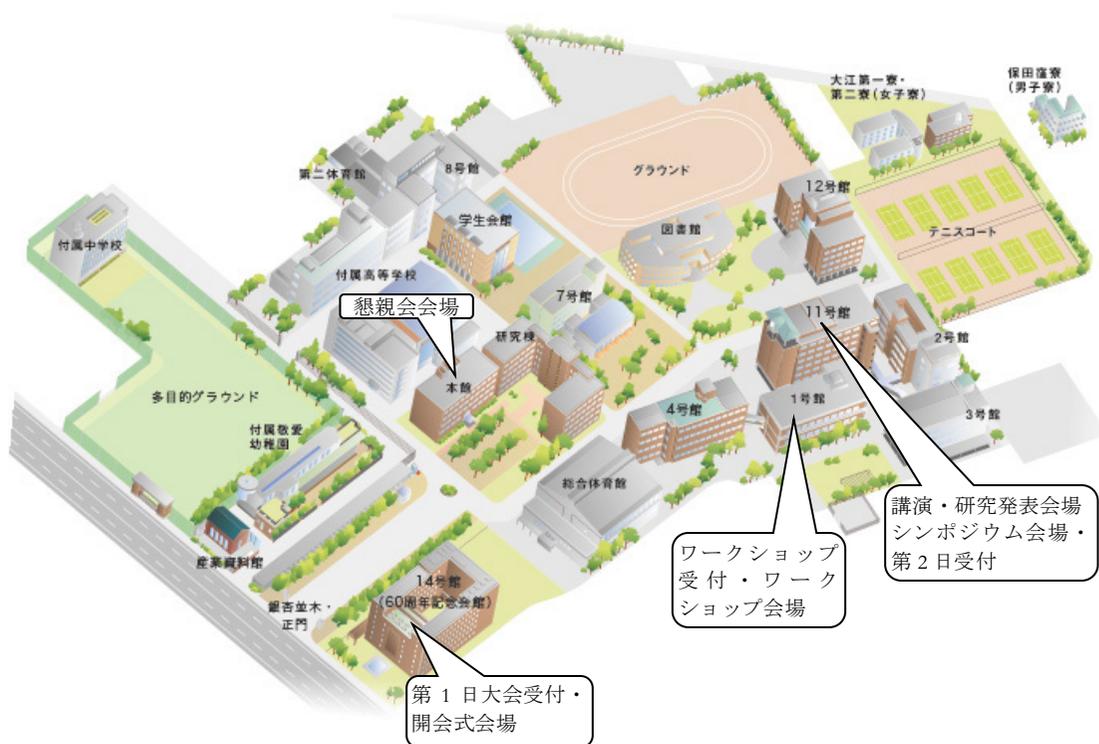
九州自動車道 熊本インターより

【車利用】 約15分

熊本空港より

【車利用】 約 40 分 【空港バス利用】 約 25 分 ◎バス停「味噌天神」下車 徒歩約 15 分

◆会場案内図◆



※なお、10月4日午後1時50分以降の受付は、研究発表が行われる11号館115B教室にて行いますのでご注意ください。

◆熊本学園大学キャンパス内の売店・食堂など◆

※ 日曜日は、ローソン以外の学内売店等は開いていませんのでご注意ください。



7号館レストラン

栄養バランスの良いヘルシーでバラエティー豊かなメニューがずらり。地域色たっぷりのメニューも大人気！



ローソン

お弁当や、飲み物、日用品、雑誌など幅広い品ぞろえ。8時～22時の営業で、年中無休（一斉休業中は除く）なのもうれしいところ。



Hairmake Amore

カット、カラー、パーマがお手頃価格で！卒業式の着付・メイク・セットもおまかせあれ。



写真のハマヤ

高品質の10分仕上げ現像、各種証明写真の撮影も。スマートフォンからのデジタルプリントもできます！



トラベルサロン西鉄旅行

切符1枚から団体旅行に至るまで、チケット購入や旅行の申込みはこちらまで。学割も利用できます。まずは気軽にご相談を。



グリーンキャンパス カウンター

損保代理店、不動産部のアパート紹介、自動車学校や引越し、宅配便など各種業者の取次など。暮らしの相談はまずカウンターへ！



ブックストア丸善

書籍・雑誌はもちろん、文具、雑貨、お菓子、弁当、飲料などを学割で買えます。工場直送の「学園大珈琲」はギフトにも大好評。



熊本学園大学
オリジナルコーヒー
も売ってます



12号館ベーカリー

リーズナブルな値段で焼きたてパンを提供しています。毎月発売される新商品にもご注目を。



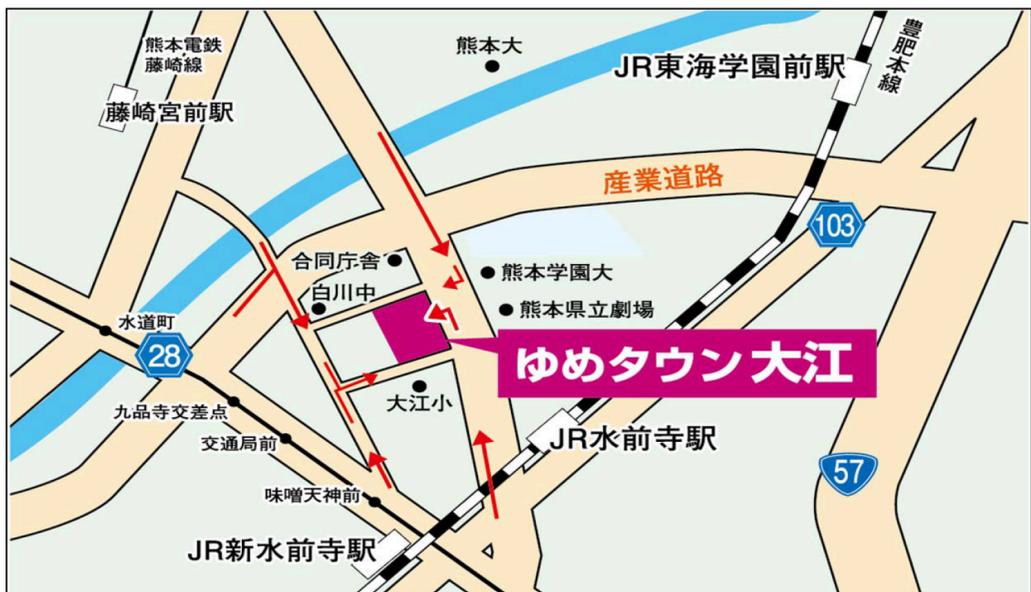
メロンパン110円



3号館レストラン ユメシヨク

安くておいしい学生の味方。田崎市場直送の厳選した食材を使ったこだわりのメニューを用意してお待ちしています。

◆熊本学園大学キャンパスそばのショッピングモール（ゆめタウン大江）◆



2014年8月25日発行

編集・発行 英語コーパス学会

会長 堀 正広

事務局 〒560-0043 大阪府豊中市待兼山町 1-8

大阪大学大学院言語文化研究科

田畑 智司研究室気付

電話：06-6850-5866

e-mail: jaecs.hq@gmail.com twitter: @JAECs2012

URL: <http://english.chs.nihon-u.ac.jp/jaecs/>
