

英語コーパス学会 第 37 回大会資料

日時：2011 年 10 月 1 日（土）－ 2 日（日）

会場：京都外国語大学

(<http://www.kufs.ac.jp/>)

〒615-8558 京都市右京区西院笠目町 6

■第 1 日目

ワークショップ 1 《正規表現による英語表現の検索》

会 場：6 号館 5 階 652 教室

日 時：10 月 1 日（土）10:00-12:00（9:30 受付開始）

講 師：大名 力（名古屋大学）

定 員：定員 40 名（先着順・要予約）

参加費：会員無料。非会員 1,000 円。

※予約申し込みは、ワークショップ 1・2 の別、氏名、所属、会員・非会員の別を記入の上、郵便・電子メールで事務局（inoue@v.email.nu.jp）まで。

日 時 2011 年 10 月 1 日（土）
 受付開始 12:20（1 号館 7 階エレベーター前）
 開 会 式 13:20（1 号館 7 階 171 教室）

1. 会長挨拶
2. 開催校挨拶
3. 総会
4. 学会賞審査報告
5. 事務局からの連絡

司 会 井上永幸（広島大学）
 赤野一郎（京都外国語大学）
 久保哲男（京都外国語大学副学長）

〈研究発表第 1 室（1 号館 7 階 171 教室）〉

研究発表 1 14:20-14:50
 certainly と definitely の機能的差異について

司 会 石井康毅（成城大学）
 鈴木大介（京都大学大学院生）

研究発表 2 14:55-15:25
 前置詞の多義性とコロケーション — 認知言語学の理論に基づく句構成の分析—

鎌倉義士（愛知大学）

研究発表 3 15:30-16:00
 MI-score, t-score と“コロケーション”

大名 力（名古屋大学）

〈研究発表第 2 室（1 号館 6 階 161 教室）〉

研究発表 1 14:20-14:50
 学校文法に基づいた英文解析による言語データの頻度分析

司 会 能登原祥之（比治山大学）

田中省作（立命館大学）
 富浦洋一（九州大学）
 徳見道夫（九州大学）

研究発表 2 14:55-15:25
 英和辞典のシノニム記述に関する考察 — Sketch Engine を用いて—

吉村由佳（東京理科大学非常勤講師）

研究発表 3 15:30-16:00
 CEFR Level Marker の開発：英語到達度指標研究のコーパス言語学的展開

投野由紀夫（東京外国語大学）
 金田拓（東京外国語大学大学院生）
 土肥康輔（東京外国語大学大学院生）

〈休 憩 16:00-16:20〉

講演会 16:20-17:50（1 号館 7 階 171 教室）
 認知言語学からみたコーパス言語学の展望

司 会 堀 正広（熊本学園大学）
 講 師 山梨正明（京都大学）

《懇親会 時間：18:00-20:00 場所：11 号館 2 階，ラウンジ；会費：5,000 円》

■第2日目

ワークショップ2 《テキストエディターによる検索と加工》

会場：6号館 652教室

日時：10月2日（日）9:30-11:00（9:00受付開始）

講師：大名 力（名古屋大学）

定員：定員40名（先着順・要予約）

参加費：会員無料。非会員1,000円。

※予約申し込みは、ワークショップ1・2の別、氏名、所属、会員・非会員の別を記入の上、郵便・電子メールで事務局（inoue@v.email.nu.jp）まで。

日時 2011年10月2日（日）
受付開始 10:40（1号館7階エレベーター前）

〈研究発表第1室（1号館7階171教室）〉

司会 川崎修一（日本赤十字看護大学）

研究発表1 11:15-11:45

英語補文における“to-be 削除”の大規模コーパス分析 — 願望動詞の補文に生起する形容詞の意味パターン比較を中心に —

澁谷竜昇（大阪大学大学院生）

研究発表2 11:50-12:20

英語形容詞の使用：構文との関係

渋谷良方（京都外国語大学）

〈研究発表第2室（1号館6階161教室）〉

司会 神谷健一（大阪工業大学）

研究発表1 11:15-11:45

OED, インフォーマント, 頻度に基づく英語オノマトペ語彙の抽出

菅原 崇（岐阜工業高等専門学校）

研究発表2 11:50-12:20

統計的対訳表現抽出を利用した英日バイリンガルコロケーションリストの構築

後藤一章（摂南大学）

〈昼休憩 12:20-13:25〉

シンポジウム 13:25-15:25（1号館7階171教室）

《Current Trends in Corpus Linguistics: Voices from Britain》司会 Laurence Anthony（早稲田大学）

Introduction 講師 Laurence Anthony（早稲田大学）

Current trends in corpus linguistics at the University of Birmingham

講師 仁科恭徳（立命館大学）

Current trends in corpus linguistics at Lancaster University

講師 高橋 薫（豊田工業高等専門学校）

Current trends in corpus linguistics at the University of Nottingham

講師 Michael Handford（東京大学）

閉会式 15:25（1号館7階171教室）

閉会の辞

赤野一郎（京都外国語大学）

■1日目

【ワークショップ 1】

正規表現による英語表現の検索

講師 大名 力 (名古屋大学)

テキスト形式のファイルであれば、専用の検索ツールがなくとも、正規表現を利用してテキストエディターなどのツールでかなり自由に検索や加工ができる。また、WordBanksOnline や AntConc のように、コーパス (検索サイト)、コンコーダンスーにも正規表現が利用できるものがあり、専用のツールであっても正規表現を利用することでより複雑な検索条件の指定が可能になる。このように正規表現はコーパスを利用する上で威力を発揮するが、正規表現を扱った書籍やウェブページはあっても、具体的な言語表現の検索方法を解説したものが少なく、語法・文法研究等の分野では十分には活用されていない。このような状況を踏まえ、本ワークショップでは、これまでに正規表現を使ったことがない人を対象に、具体的な英語の表現の検索を例に、基本的な正規表現の使い方、間違いやすいポイントについて説明する。ワークショップでは、参加者が実際に自分で正規表現を書き、正規表現のどの部分が対象文字列のどの部分にマッチするのかを確認しながら進めていく予定である。基本的なコンピューターの操作に関する知識は必要であるが、特別な知識は必要としない。

【研究発表第 1 室】

【研究発表 1】

法副詞 *certainly* と *definitely* の機能的差異について

鈴木大介 (京都大学大学院生)

本発表では、英語表現の生起環境に着目し、コーパスデータに対して意味的・語用論的な情報の数量化を行うことで、法副詞 *certainly* と *definitely* の 2 つの類義語間の意味・機能の違いを論じる。先行研究では、*certainly* と *definitely* と共起する法助動詞、程度副詞、接続詞、代名詞のパターンが指摘されており、それらの意味、とりわけ蓋然性の程度の差異について言及されている (Simon-Vandenberghe and Aijmer 2007, 小西 2006, Downing 2001, Hoye 1997)。しかし、これらの研究では、法副詞がどのような表現とどの程度共起するのかについて、網羅的・体系的な分析がなされていない。さらには依然としてコミュニケーションや談話といった語用論的な観点からの分析が不十分と言える状況にある。そこで、コーパスを用いて、*certainly* と *definitely* の実際の使用パターンを記述し、この 2 表現間の意味・機能についての相違を実証する。

BYU-BNC から、*certainly* と *definitely* について、それぞれ全 18118, 3056 例が検索でき、文副詞として生起しているものを手作業で抽出していくと、各 15718, 2350 例が得られた。本発表では、これをサンプルとして用い、従来よりも広い複眼的視座から分析を行う。具体的には、法副詞を修飾する「程度副詞」、同じ節内で共起する「法助動詞や否定語」、さらには節内における法副詞の「生起位置」を変数として設定した。これらの視点から、どのような要素とどの程度共起するのか、どの位置にどのような割合で生起するのか、調査を行った。

結果として、まず、程度副詞との共起については、*certainly* が蓋然性を弱める *almost* と共起割合が高く、一方で *definitely* は蓋然性を強める *most*, *very* と共起頻度が高かった。次に、法助動詞や否定語との共起割合であるが、*certainly* が *definitely* と比べて、それらの表現と共起する割合が統計的に高いとわかり、*certainly* は生起位置に関わらず、*may* や *might* まで、様々な蓋然性を表す法助動詞と幅広く共起していた。最後に、生起位置であるが、*definitely* の場合と比べて、*certainly* は、「節頭」の割合が高く、語順と機能の関係から、談話機能を担いやすい傾向が統計的に明らかとなった。

以上のように、修飾語である程度副詞の性質から *certainly* は意味・機能が広いものと共起しやすく、さらに、*definitely* は共起する法助動詞の蓋然性や生起位置が固定・一定傾向であるのに対し、*certainly* は意味・機能がより拡大している傾向にあると言える。すなわち、この 2 表現は、意味論、さらには談話・語用論レベルにおいて相違を示しており、*certainly* の方に意味・機能の拡張が見られる。

【研究発表 2】

前置詞の多義性とコロケーション — 認知言語学の理論に基づく句構成の分析 —

鎌倉義士（愛知大学）

前置詞の意味を完全に理解するのは学習者にとって困難である。前置詞の多様性によって、学習者は習得が難しいと感じるのかもしれない。対して、英語母語話者は難なく、その意味を理解している。

本発表では、共起する名詞句が前置詞の意味を特定すると仮定し、その仮説を ICE GB コーパス（サイズ 100 万語）のデータ（*over*: 402 例, *through*: 316 例, *into*: 648 例）を用いて検証した。結果として、前置詞の意味とその名詞句のコロケーションは統計的に有意な相関が見られた。さらに、名詞句と前置詞のパターンの頻度から独自の *through* と *into* の意味ネットワーク（semantic network）を構築した。

仮説として、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意味と関係していると考えた。その仮説は前置詞の空間的意味を二つの事物の関係から説明する認知言語学の理論に基づいている（Langacker 1987）。人間の認識を反映すると仮定した上で、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。名詞句の組み合わせの頻度が、*over*, *into*, *through* の組み合わせにおいて類似したパターンを示した。この結果は、それぞれの前置詞が path スキーマを共有している（Dewell 1994, Dirven 1989, Lindstromberg 1998）という認知言語学的主張と一致する。

次の分析は、Tyler and Evans (2003) の *over* の意味ネットワークを本発表の仮説を基に再検討したものである。前置詞と名詞句の間に有意な相関が見られなかったが、*covering* の多義性において共起する動詞の特徴を明らかにした。

最後の分析は、仮説に基づき *into* と *through* の多義性を示し、その意味ネットワークを構築した。認知言語学者が示す意味ネットワークは意味の類似性を主観的に判断して作られたものであるが、本発表は名詞句のコロケーションの頻度を反映した意味ネットワークを提示している。

本発表の結果は、form and meaning の問題に示唆を与える（Evans 2009）。前置詞と名詞句の句構造が意味をどの程度まで限定しているのかをパターンの頻度とコーパスからの実例で本発表は説明している。

【研究発表 3】

MI-score, t-score と “コロケーション”

大名 力（名古屋大学）

所謂“コロケーション性の指標”としての z-score の利用は Berry-Rogghe (1973) まで遡り、また、Church & Hanks (1990), Church et al. (1991) による MI-score, t-score の利用の提案から数えても約 20 年が経過しており、既に古典となっている観もあるが、同じ名前と呼ばれていても論文やプログラムによって使用されている計算式が異なることがあり、したがって“コロケーション性”判定の基準値として示される値も意味が異なることは、意外と理解されていない（例えば、Church et al. (1991) と Hunston (2002) が目安/基準値とする MI-score 3 は、期待値の計算方法の違いのため意味が異なる）。開発者・研究者の間で共通の認識がないと、プログラム作成において異なる実装がされていてもそれに気付かず、異なるプログラムによる数値を同一の指標として利用してしまいかねない。問題はスコアの値、基準値の違いだけにとどまらず、“コロケーション性の指標”の意味やスコアの利用法にも及び、問題解決のためには、概念・用語を整理し、学界関係者間で共通理解を進めていく必要がある。

本発表では、共通理解を促進し共通の基盤に基づいた議論を可能にするために、MI-score, t-score に関し次の 8 点を中心に問題の検討・整理を行う：(a) 何に関するものか（連想関係/結びつき/共起（性、関係）/コロケーション（性））；(b) (a) の何を計るものか（強度か確信度か）；(c) 比較の対象は何か（単語の分布のモデルは何か）；(d) 計算式；(e) 期待値の算出方法（スパンの大きさに応じたものか）；

(f) “基準値”とその根拠；(g) スコア採択の条件（総語数，各語の頻度，共起頻度，など）とその意味；(h) スコアの利用法とその妥当性。また，各研究者のコロケーションの捉え方・定義の違いとは別のレベルで「コロケーション」という語が多義に使われていることも誤解を引き起こす原因の1つと考えられるため，次の(a) - (c)の3つに分け用語の意味を整理した上で，「コロケーション」とMI-score, t-scoreとの関係を明らかにする：(a) 連語関係/具体的な個々の連語；(b) 単なる語の組み合わせ/共起性の高い語の組み合わせ/特定の関係にある語の組み合わせ；(c) 心的存在/外在化されたもの（発話，コーパスにおいて観察されるもの）。

【研究発表第2室】

【研究発表 1】

学校文法に基づいた英文解析による言語データの頻度分析

田中省作（立命館大学）・富浦洋一（九州大学）・徳見道夫（九州大学）

本発表は，コーパスなどの言語データに対する頻度分析を日本の学校文法（綿貫他，2000；小池他，2003）の観点から直接行うための枠組みを提案し，その可能性を論じる。

現在，コーパスなどの分析にも利用される言語の自動解析のほとんどは，句構造文法をはじめとした科学文法をベースとしている。しかし，コーパス分析などで最終的に得られた言語特徴を学校文法あるいは文法シラバスの項目で説明することは少なくない。たとえば，Granger (1998) や田中他 (2006) は，分析の単位を広義の n-gram としつつも，得られた特徴の一部を論じる際には，結局，学校文法の諸項目に翻訳した上でを行っている。このような翻訳過程では専門家の観察や内省が求められ，非効率で網羅性に欠ける場合がある。また，それら諸項目の頻度情報を直接取り扱えないため，項目間に潜む関係性を見出すような高度な分析法を適用することが難しい，という問題もある。

そこで，我々は任意の英文に対して，それに含まれる学校文法の諸項目（佐野他，2000；田中他，2008）とその関連箇所を自動検出するシステムの開発を進めている。たとえば，“Being tired, he sat down to rest.” に対して「分詞構文」（文頭 Being ~ sat）, 「TO 不定詞（副詞的用法）」（sat ~ rest）, Being tired, sat, rest からそれぞれ態・時制・相などの情報が得られる。本システムは各項目の検出精度（適合率・再現率）が事前に見積もられており，データの性質が合致すれば諸項目の頻度については補正が可能で，言語データの頻度分析などに活用することができる。

本発表では，本システムとあわせて，それを活用した試験研究の一部を示す。試験研究は，論文の表現という点で良質な英語科学論文（G 論文）と日本人によるそうではない論文（JP 論文）を学校文法の諸項目の頻度で直接対比し，日本人英語科学論文の特徴の一端を探っている。本発表では特に動詞に関わる情報に焦点をあて，紹介する。たとえば，G 論文には「分詞構文」「助動詞」などが多く「時制」も多様，JP 論文については従来から言われている「受身」が多いといったことに加え「助動詞を付随した受身」「未来時制の受身」が多い，などの特徴が直接観察されている。

【研究発表 2】

英和辞典のシノニム記述に関する考察 —Sketch Engine を用いて—

吉村由佳（東京理科大学非常勤講師）

英英・英和辞典は学習者が戸惑いやすいシノニムの違いについて，コラムや注釈を使って説明したり，シノニム同士の相互参照により，違いを学習者に比較させるといった方法で記述を行なっている。最近の英英辞典では CD-ROM やインターネット利用により，シソーラスやコロケーション辞典と連動させたシノニム記述も実現している。しかし，英英辞典の場合は英語話者の観点からシノニムが記述されており，かならずしも日本語を母語とする英語学習者が知りたいと思う使い分けが記されているとは限らない（南出，2001）。一方，英和辞典では日本語話者に役立つ情報を考慮してシノニムが説明されている（井

上, 2008)。また, 対象語の定義として日本語を用いるため, 訳語選択にシノニムの使い分けを反映させることもできる。こうした独自性に加え, データとしてコーパスを利用すれば, シノニム同士の共起語の比較などにより英語話者も気づかない言語事実を明らかにできる可能性がある(赤野, 2008; 井上, 2010)。

本発表の目的は, 名詞 “smell” に代表されるシノニム (smell, scent, fragrance, aroma, odor) を例として取り上げ, コーパスを用いた英和辞典独自のシノニム記述のあり方を探ることである。まず “smell” のシノニムについて, 一般的な学習英英辞典・英和辞典の記述を概観した後, コーパスデータからシノニムの特徴を観察する。コーパス検索ツールとして, 編集方法が似通った英語 (ukWac) と日本語 (jpWac) の WaC (Web as Corpus) が存在する Sketch Engine を使用し, 動詞・形容詞・名詞などとの共起関係を中心に分析を行なう。具体的には, 英語では “smell” のシノニム, 日本語では訳語として最もよく用いられる「におい」のシノニムの特徴を比較・分析する。それぞれのシノニムの共起語の特徴を分類した結果, “smell” のシノニムと訳語に用いられる日本語との意味的・統語的類似点・相違点が明らかにされる。以上の考察を踏まえ, “smell” のシノニムを英和辞典で扱う場合, (1) 相互参照の利用, (2) 共起語の差を反映した訳語の選定, (3) 日・英の統語構造の差を考慮した例文提示, などが有効であることを示し, まとめとしたい。

【研究発表 3】

CEFR Level Marker の開発：英語到達度指標研究のコーパス言語学的展開

投野由紀夫 (東京外国語大学)
金田拓 (東京外国語大学大学院生)
土肥康輔 (東京外国語大学大学院生)

近年, 言語教育の分野で Common European Framework of Reference (CEFR) が世界的な注目を集めている。これは Council of Europe が提唱する外国語能力の汎用的到達指標である。CEFR は初級・中級・上級を A1/A2 (basic user), B1/B2 (independent user), C1/C2 (proficient user) の 6 段階で表示し, それぞれ 5 技能 (spoken production, spoken interaction, listening, reading, writing) の能力記述文を示したものである。

本発表では, CEFR を我が国に導入する際の基礎研究の一環として, CEFR のレベル別言語特徴を対象テキストから検索し, それによりテキストの CEFR レベルの判定に参考になる言語統計情報を提供できるツール (通称 CEFR Level Marker) のプロトタイプ開発について発表する。

CEFR Level Marker 開発の意義は何であろうか。今後 CEFR が普及し CEFR レベルに準拠した教材 (教科書, テストを含む) が必要になった場合, 日本では CEFR 準拠の教材のイメージは極めて乏しい。そこに, このようなレベル判定の資料を客観的に提供してくれるプログラムがあれば, 国内海外の教材をさまざまに分析して, 優れた教材開発の一助とすることが可能だ。さらに, このような言語特徴データベースを作っておくことで, 学習者データの発話・作文のレベル判定などにも活用できる可能性が開けてくる。

具体的な研究手順であるが, まず我々は CEFR の基礎となった Threshold Level (1975) を参考に各 CEFR レベルに必要な言語特徴のインベントリー作成を試みた。Threshold Level は B1 レベルの機能・言語特徴のリストであり, その後 Breakthrough (=A1), Waystage (=A2), Vantage (=B2) の 4 レベルが公開されている (これを総称して T-series と呼ぶ)。この 4 レベルの言語特徴, 機能, 概念のデータベースを構築, レベル間の重複などを調整して T-series Database と命名した。

次に CEFR Level Marker の設計を以下のように行った。

- (1) T-series Database にリストされている言語特徴を抽出する正規表現一覧を用意
- (2) データベースの ① Language Function (LF), ② General Notion (GN), ③ Specific Notion (SN) の 3 つのカテゴリーで絞り込む機能を利用し, フィルターした言語特徴の (1) の正規表現検索式リストを出力できるようにした
- (3) (2) の検索式一覧を perl などのスクリプトに渡し, テキストを検索, 該当言語特徴の頻度を CSV

形式で出力

今回の発表ではこの T-series DB の概要, MySQL + PHP による web インタフェースへの試験的実装, ruby による抽出プログラム試作, および機能評価として CEFR Level Marker を Cambridge Exam の CEFR レベル別リーディング・セクションにかけてみた結果を報告する。

【講演会】

認知言語学からみたコーパス言語学の展望

山梨正明 (京都大学)

理論言語学の研究では、生成文法を中心とする形式文法の影響により、コーパス言語学と理論的な視点からみた言語学の交流は積極的にはなされてきていない。しかし、近年、認知言語学の研究の進展に伴い、コーパス言語学と理論的な視点からみた言語分析の関係が積極的に見直されつつある。認知言語学の研究プログラムでは、特に用法基盤モデル (Usage-Based Model) の枠組みに基づく言語分析が進展している。用法基盤モデルのアプローチの基本的な特徴は次の点にある：(i) 語彙レベル、文レベルから談話レベルの自然発話を分析対象として重視する、(ii) 質的分析だけでなく、事例のトークン頻度、タイプ頻度、等の量的分析から言語現象の一般化を試みる、(iii) 典型事例と周辺事例を、個々の言語使用者の変異、ゆらぎ、等を含めて相対的に分析する、(iv) 形式文法のトップダウン的アプローチではなく、言語使用を重視するボトムアップ的アプローチを前提とする、(v) 経験基盤主義の観点から、言語的知識を、ダイナミックな言語運用の発現として創発的に規定していく。本講演では、認知言語学の以上の用法基盤モデルの研究方向を明らかにするために、実証面と理論面に関する次のような問題を考察する。まず実証面に関しては、(i) 用法基盤モデルの研究の中心的なターゲットであるイディオム表現、メタファー表現、メトニミー表現、等の連語パターン、(ii) 文字通りの意味と修辭的な意味を表現する語彙、句、構文の分布関係のコロケーションの制約と語用論的な制約、(iii) トークン頻度、タイプ頻度の観点からみたプロトタイプの言語表現とその拡張表現の分布関係、(iv) 理論/仮説の実証と反証に用いられる作例、文学的テキストの引用例、自然発話のコーパス、等の各種の言語事実の意味解釈、(v) 文脈と言語外的な知識による各種事例の容認性のゆらぎ、等の問題を考察していく。また、理論面に関しては、以上の考察を背景に、言語能力と言語運用の二分法に基づく言語学のアプローチの問題、辞書と文法のモジュール的な区分に基づく言語学のアプローチの問題、意味的要因と言用論的な要因を捨象した形式文法の自律性に基づく言語学のアプローチの問題、等を批判的に検討していく。また、以上の考察を通して、今後の理論言語学とコーパス言語学の相互のリサーチ・パラダイムの新たな研究の方向を探っていく。

■第2日目

【ワークショップ2】

テキストエディターによる検索と加工

講師 大名 力 (名古屋大学)

最近ではユーザーフレンドリーな環境も整って来ており、テキストエディターでファイルを直接検索・加工しなくてもコーパスは利用できるが、テキストエディタが使えれば、コンコーダンサーで扱いやすいよう前処理をしたり、専用ツールによる処理結果からさらに条件を指定して該当例を抽出したり、表計算ソフトで処理しやすいように加工したりするなど、可能な処理の幅が広がる。コーパスの中にはデータのみが提供されているものもあり、また、狭義のコーパスでなくとも言語研究に利用可能な電子テキストは多量にあるので、専用ツールがなくても扱えるとなれば、研究に利用できるデータの範囲も広がる。

エディターにより使える機能も異なるが、今回のワークショップでは、多くのエディターが共通して持っている基本的な機能を利用して、次のような処理の方法について学ぶ。

処理の例

- ・ 正規表現による検索
- ・ ファイル一括検索 (grep 検索, ワイルドカードによるファイル指定)
- ・ HTML などのタグの削除
- ・ Brown Corpus などの行頭の ID の削除
- ・ 行の連結

基本的なコンピューターの操作に関する知識は必要であるが、正規表現やテキストエディターに関する知識は前提とはしない。正規表現については簡単なものしか扱わないので、詳しいことが知りたい方は前日のワークショップにご参加いただきたい。

【研究発表第 1 室】

【研究発表 1】

英語補文における“to-be 削除”の大規模コーパス分析 — 願望動詞の補文に生起する形容詞の意味パターン比較を中心に —

しづたに
澁谷 竜昇 (大阪大学大学院生)

本発表は、英語の願望動詞 (e.g. *want, wish*) の補文に生起する形容詞の意味パターンの大規模コーパス分析を通して、いわゆる“to-be 削除” (*to-be* deletion) と称される現象が、統語基盤ではなく意味基盤で動機付けられていることを実証するものである。

生成文法は、*to* 不定詞節補文から *to be* が削除された結果、*to be* を含まない小節補文の構造が生成されると論じている (cf. Rosenbaum 1967)。これは純粋な統語的要請に拠る変形によって実現されるものであり、そこに意味的な動機付けは想定されていない。一方、認知文法の観点から Langacker (1991) は、両補文が異なる概念化によって動機付けられたものであると主張している。つまり、両補文の間には概念化主体による時間幅の概念化の有無があるとした上で、*to* 不定詞節補文は直接的経験を越えた包括的判断によって、一方、小節補文は直接的知覚体験によって、それぞれ概念化されたものであるとしている。上記をはじめとした既存理論の考察より、以下 3 点の問題提起を行う。

- a. 純粋な統語的要請によって、*to* 不定詞節補文から小節補文が“生成”されるならば、単一の構造ではなく、複数の構造が“生成”されなければならない根拠は何か。
- b. *to* 不定詞節補文と小節補文が独立した意味を持つというのであれば、その理論がアドホックではないことをどのように保証するか。
- c. 既存理論が実際の言語使用の実態を反映していることをどのように保証するのか。

これらの問題提起に対して、本発表では大規模コーパス調査によって回答を求める。具体的には、願望動詞を主動詞とする事例に着目し、*to* 不定詞節補文と小節補文に生起する形容詞のパターンを統計的に記述する。大規模コーパス Corpus of Contemporary American English (COCA) よりサンプルを採取し、因子分析の手法を用いて分析を試みた。分析の結果、2 因子が抽出され、因子パターンは *to* 不定詞補文と小節補文を二分する結果となった。このことは、両補文の生起環境下においては、生起する形容詞の質が異なるということを示唆している。具体的に、*to* 不定詞節補文においては、段階的形容詞 (e.g. *happy, safe, successful*) が有意に観察された。一方、小節補文においては、非段階的形容詞 (e.g. *dead, alive*) が有意に観察された。この結果は、時間性的概念がそれぞれの構文を生起させる動機付けとなっていることを示唆するものであり、Langacker の主張を支持するものとなった。本発表の研究結果は、さまざまな主動詞のタイプにおける補文形式の意味的動機付けを解明する上でのひとつの指針を提示するものであると思われる。

【研究発表 2】

英語形容詞の使用 — 構文との関係 —

しづや
渋谷良方 (京都外国語大学)

本発表は、(1) ~ (3) の用法 (構文) における英語形容詞の使用について、量的研究の手法に基づいた分析を展開する。

(1) 限定用法：前置修飾形容詞構文

His report, published yesterday, demands *fundamental* changes in the way safety is regulated in the North Sea. (ICE-GB: W2C-007)

(2) 叙述用法 I：複合自動詞形容詞構文

And, uhm their response was, uhm very *positive* uhm, (ICE-GB: S1A-001)

(3) 叙述用法 II：複合他動詞形容詞構文

And clearly I've got what I need to make my life *happy* (ICE-GB: S1A-031)

英語形容詞の使用に関する興味深い事実は、用法 (構文) によって形容詞の生起分布が異なることである。例えば、無作為抽出した一部の形容詞 (*good, difficult, clear, right, warm, dead*) の限定用法と叙述用法 (I と II) について、ICE-GB (the British component of the International Corpus of English) のタギング&パーシング・スキーム (Tagging & Parsing Scheme) を用いて (半) 機械的に生起頻度をカウントして分割表を作成し、用法 (構文) 毎の重み付けや、フィッシャーの正確確率を求めると、限定用法と叙述用法 (I と II) の中で最も特徴がある形容詞は *good* であり (1.06E-76)、これが限定用法での使用が特に著しいことが分かる。次に特徴的な形容詞は順に *difficult* (1.23E-32) と *clear* (8.22E-20) であり、これらは頻度では叙述用法 I での使用が顕著である。*dead* の p 値も低い (2.28E-03)、特徴語としての地位は相対的には低い。重み付けの観点で見れば、限定用法では *good* (0.941995)、叙述用法 I では *right* (0.62452) と *good* (0.62077)、叙述用法 II では *clear* (0.660115) と *difficult* (0.608341) などが、相対的重要度が高いことが分かる。

このように、形容詞は同一のものであっても限定用法と叙述用法 (I と II) とで生起頻度 (相対的重要度) は異なる。統計的手法に基づく特徴語の抽出結果は、用法 (構文) 間における形容詞の使用 (分布) の違いがランダムでないことを、つまり、形容詞の使用 (分布) は生起する構文 (用法) と強く関係付けられていることを示唆する。本発表は、「構文とそれに生起する語との関係」を量的に計算することによって、形容詞の使用と構文 (用法) の関係の特徴付けることを目指す (e.g. Gries, Hampe, and Schönefeld 2005, Shibuya 2011, Stefanowitsch and Gries 2003)。

【研究発表第 2 室】

【研究発表 1】

OED, インフォーマント, 頻度に基づく英語オノマトペ語彙の抽出

菅原 崇 (岐阜工業高等専門学校)

英語は日本人にとって馴染みのある外国語の一つにもかかわらず、日本国内における英語オノマトペ研究はそれほど盛んではない。その大きな理由として、信頼度の高い英語オノマトペリストがないという点が挙げられる。一般的に、英語オノマトペは日本語オノマトペに比べ音韻の特徴がないとされている。そのため、特に英語を第一言語としない研究者にとっては何を英語オノマトペ (の代表) としてよいのか特定し難い。また、諸外国 (とくに英語圏) に目を向けても伝統的にオノマトペを周辺的で幼稚な現象とみなす傾向があるためか、収集規準を明確にした大規模な英語オノマトペリストはない。

そこで本発表では、1) より信頼度が高く、多くのものに検証可能な英語オノマトペリストの作成、2) 上記リストのなかで英語話者が「よりオノマトペらしい」と認識している語彙リストの作成、3) 上記の「よりオノマトペらしい語彙」リストから高頻度の 5 語を話し言葉、書き言葉それぞれから抽出、の 3 つを目的とする。

英語オノマトペのリストの作成に当たっては電子版の OED を用い、語源の項目に「オノマトペであ

る（可能性がある、起源である、etc.）」と記載のあるものを抜き出した。結果、OEDを用いて英語オノマトペ（もしくはその可能性のある語）287語をリスト化できた。

第二の目的に当たっては、完成した英語オノマトペリストの各語彙の「オノマトペらしさ」を5名のネイティブスピーカーに4つのパラメーター（0 = This word is not onomatopoeic / 1 = This word is a little (= slightly, maybe etc.) onomatopoeic / 2 = This word is totally (= very, most etc.) onomatopoeic / x = I don't know this word）で判断してもらい、5名の評価の合計（満10点）が6点以上のもの「よりオノマトペらしい語」とした。結果、287語中55語が6点以上獲得した「よりオノマトペらしい語」であった。

第三の目的に当たっては、先の55語のうち、話し言葉コーパスLLCと書き言葉コーパスLOBでそれぞれもつとも頻度の高い5語を抽出した。結果、LLCにおけるトップ5は *pop, bash, bounce, tick, clash*, LOBにおけるトップ5は *murmur, flap, mutter, crash, dash*, となり、それぞれのトップ5が互いに重ならなかった。さらにそれら語のトークンを観察した結果、LLCのトップ5のうち4語 (*pop, bash, bounce, dash*) は「位置変化 (change of location)」を表す頻度が高かったのに比べ、LOBのトップ5においてはそのような語は1語のみ (*dash*) だった。さらにLOBのトップ5の *murmur* と *mutter* の2語は「引用 (quotations of speech)」「that 節」「会話の内容を表す目的語 (objectives as contents of speech)」を伴い動詞で用いられることが多い語であった。

【研究発表2】

統計的対訳表現抽出を利用した英日バイリンガルコロケーションリストの構築

後藤一章（摂南大学）

近年、英語学習におけるコロケーションの重要性が認識されるようになり、有用なコロケーション学習リソースの開発がコーパス言語学分野を中心に進められている。特に、言語解析技術を駆使したコーパスからの大規模なコロケーション抽出の実現は、当該分野に重要な進展をもたらした。こうした多量のコロケーションデータを比較的容易に収集可能な現況において、次に喫緊の課題として考えられるのは、それらのコロケーションにいかに関与するかという問題である。英語コロケーションの中には字義通りの意味ではない項目も多く、また、語彙量の少ない初・中級の学習者の場合、たとえ字義通りの意味であっても、理解できないケースも起こり得る。従って、日本語訳の付与は、有用な学習リソースを開発する上で有力な手段となる。しかしながら、手作業で訳語を付与するには膨大なコストがかかり、コーパスから多量のコロケーションデータが得られればそれだけコストが増大するというジレンマも存在する。そこで本発表では、対訳表現をパラレルコーパスから機械的に抽出する手法を活用し、日本語訳の付与された英語コロケーションリスト、すなわち「英日バイリンガルコロケーションリスト」の構築を試みる。

コーパスデータには、行単位で日英対応が取られた「Wikipedia 日英京都関連文書対訳コーパス」を使用する。これは現時点で公開されている日英パラレルコーパスの中で最大規模のものであり、英語の総語数は約1,000万語に及ぶ。英語及び、日本語セクションにそれぞれ統語解析を行い、英語からは「述語-目的語」表現を、日本語からは用言中心の「格構造」を抽出する。詳細は発表時に報告するが、各英日表現の抽出時には、後述する類似度計算のため、出現番号を保持させておく。

日英表現の意味的な対応関係を、本発表では、候補表現間の出現パターンの類似度問題として定義する。すなわち、英日の各表現パターンが各言語上で独立して発生する確率と、それが対訳文上に同時に発生する確率に基づいて類似度を算出し、ある英語の「述語-目的語」コロケーションと、最も類似度の高い日本語表現を対訳表現として特定する。例として、*commit a crime* というコロケーションが生起する英文の対訳文の集合上に、「罪を犯す」、「犯人を逮捕する」、「話をする」といった日本語表現がそれぞれ50回ずつ生起していると仮定する。この時、「罪を犯す」のコーパス全体における単独頻度が60回、「犯人を逮捕する」が120回、「話をする」が180回とすれば、単独頻度と同時生起頻度（共起頻度）の割合から「罪を犯す」の類似度が最も高くなり、最良の対訳表現候補として判定されることになる。なお、本発表では類似度計算に「重み付きDice係数」を使用した。

実験の結果、高い適合率で日本語訳が付与され、効率的な英日バイリンガルコロケーションリスト開

発の可能性が示された。本抽出手法は、行対応さえ取られていれば容易に他のパラレルコーパスにも適用可能であり、分析対象も様々な統語構造に拡張することで、より充実したバイリンガルコロケーションリストの開発が期待される。一方、今後の課題として、今回使用したパラレルコーパスの翻訳方向が日→英であったことが英語コロケーションの多様性を制限していた可能性があり、様々な条件下でさらに検証を行っていきたいと考える。

【シンポジウム】

Current Trends in Corpus Linguistics: Voices from Britain

Since the birth of modern corpus linguistics in the 1960s, innovative research has continued to emerge from leading British institutions, in particular the University of Birmingham, Lancaster University, and the University of Nottingham. At the University of Birmingham, for example, the COBUILD project, led by John Sinclair, resulted in the creation of the first corpus-informed learner dictionary, *Collins COBUILD English Language Dictionary* (1987). The influential work of Tim Johns on data driven learning also emerged from the University of Birmingham, as has the pattern grammar work of Susan Hunston, and a host of other hugely important projects. Similarly, Lancaster University has been home to many important developments in corpus linguistics. For example, much of the work on the British National Corpus was carried out under the leadership of Geoffrey Leech at the Lancaster University Centre for Computer Corpus Research on Language (UCREL). Lancaster University is also home to leading researchers on language change, the language of minorities, and critical discourse studies. It is also home to the state-of-the-art CLAWS tagger. In contrast, many important findings on spoken language have emerged from the CANCODE project at the University of Nottingham, under the guidance of Michael McCarthy and Ronald Carter. The University of Nottingham is also unique in its attempt to apply corpus linguistics techniques in the field of medicine.

In this symposium, three representatives from the University of Birmingham, Lancaster University, and the University of Nottingham will provide a unique insight into the work currently being carried out at these institutions. First, the panelists will review some of the major on-going projects at each institution. Then, they will explain how their own work fits into this picture. Finally, they will offer their opinions on the future of corpus research in Britain. Following these presentations, the moderator will briefly review work carried out at other British institutions. Then, the audience will be invited to join the panelists in a discussion of some of the past and present issues and controversies that have been discussed at British corpus research institutions, such as the importance of large corpora, the need for annotation, the importance of statistical measures, and the differences between corpus-based and corpus-driven approaches.

Introduction

講師 Laurence Anthony (早稲田大学)

Many of the major corpus linguistics research projects known throughout the world have been initiated at British university institutions. First, I will briefly review some of the work carried out at institutions outside of Birmingham, Lancaster, and Nottingham. Then, I will discuss the impact of recent changes in British university education policy on corpus linguistics departments and the field as a whole.

Current trends in corpus linguistics at the University of Birmingham

仁科恭徳 (立命館大学)

In this presentation, I will first summarize several historical and recent works by past and present researchers at the University of Birmingham, including John Sinclair, Susan Hunston, Wolfgang Teubert, Rosamund Moon, Paul Thompson, Maggie Charles (Oxford University), Nicholas Groom, and Caroline Tagg. Few of these researchers would call themselves corpus linguists. Rather, they would describe themselves as discourse analysts, lexicographers, or ESP

practitioners, in view of the fact that they more commonly utilize corpus linguistic approaches in the identification and application of qualitative aspects of authentic language use. As such, I will explain how the purpose of their research is not always quantitative in nature. For example, Hunston has written about the “Corpus Approach to Evaluation,” Teubert has examined “Lexicology and Critical Discourse Analysis,” and Moon has focused on “Fixed Expressions and Idioms, and Monolingual Lexicography.”

The historical tradition in the Birmingham school follows an idea, first suggested by J.R. Firth, that “a word is always contextual.” In the second half of the presentation, I will show how this essence still lives in much of the current research at the University of Birmingham, linking it with my own work on the interface between corpus linguistics and academic discourse analysis. In particular, I will demonstrate how corpus linguistics and discourse analysis can harmonize with one another. I will then argue that this combination can be an effective approach for identifying the link between language and culture and particularly revealing the detailed cultural values of a particular discipline.

Current trends in corpus linguistics at Lancaster University

高橋 薫（豊田工業高等専門学校）

There is no doubt that the marriage of corpora with computer technology fascinates many linguists in corpus methodology. In this respect, no review of the work at Lancaster University can avoid focusing on the Centre for Computer Corpus Research on Language (UCREL), which is home to researchers from both the Department of Linguistics and English Language and the Department of Computing. UCREL has been a leading partner in the BNC consortium, playing a central role corpus tagging and driving more recent improvements, such as fast and easy web-based access. In the first half of the presentation, I will shed light on the current work at UCREL. I will also talk about some of the controversial issues surrounding the BNC that relate directly to my own research work.

In the second half of the presentation, I will discuss some of the important trends in research at Lancaster University that have emerged from the Summer School in Corpus Linguistics, organized by UCREL staff. The summer school is intended primarily for postgraduate research students. In each session, a moderator leads discussions and encourages participants to consider various claims in areas such as critical discourse analysis, sociolinguistics, statistics, corpus design and documentation, ethics, and pragmatics. Through this explanation, I hope to explain how these sessions have shaped both mine and other Lancaster researchers’ understanding of the field.

Current trends in corpus linguistics at the University of Nottingham

Michael Handford（東京大学）

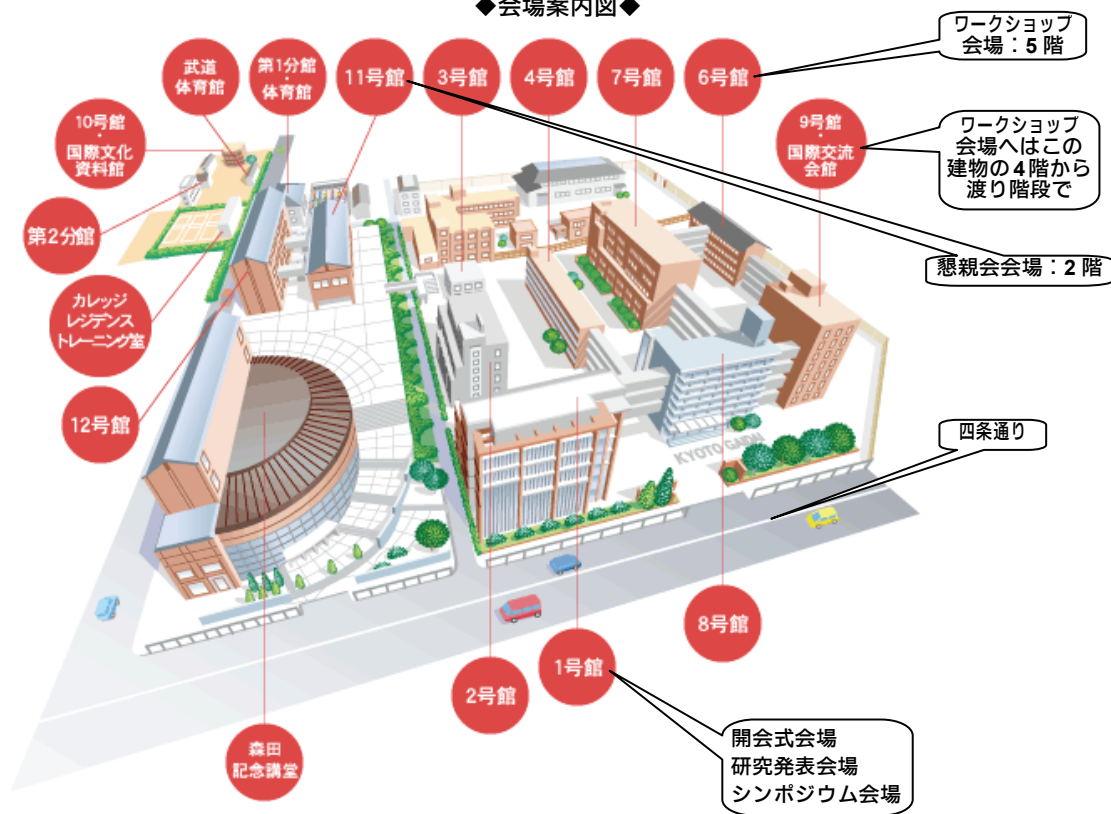
Corpus linguistics at the University of Nottingham has been directed and inspired by Mike McCarthy and Ron Carter, both of whom spent time at the University of Birmingham working with John Sinclair in the early part of their careers. Nottingham researchers at the Centre for Research in Applied Linguistics (CRAL) have focused on spoken interactions, email interactions, differences between spoken and written registers, multimodal communication and professional interactions. They also have a strong interest in the application of corpus research in language teaching and training materials, and have played a central role in the development of major corpora, such as the CANCODE corpus of spoken discourse, the CANBEC corpus of business English interactions, and more recently, the interdisciplinary, multimodal Nottingham Health Communication Corpus (NHCC).

In this presentation, I will discuss some of the Nottingham corpora and related findings. I will then show how researching and lecturing at Nottingham has informed my own work in the area of professional communication. Specifically, I will outline some of the key findings from the CANBEC corpus, which I was involved in designing. These include the keyness of several interpersonal features, e.g., the pronoun *we* and certain modal verbs, such as *need* and *have to*, as well as a high degree of metaphorical and idiomatic language use. I will also discuss the application of these findings in the forthcoming business course book series *Business Advantage* (Handford et al., 2011, CUP).

《大会参加者へのご案内》

- ・ **ワークショップの受付**：会場となる6号館 652教室の前で、1日目の「ワークショップ1」は午前9時30分から、2日目の「ワークショップ2」は午前9時から行います。
- ・ **大会の受付**：1号館7階エレベーター前で、1日目は12:20から、2日目は10:40から行います。
- ・ **昼食について**：土曜日は12号館1F食堂「リブレ」と1号館地下1Fファミリーマートが20:00まで営業しています（いずれも日曜日は営業していません）。また、大学キャンパス東側の葛野大路沿いにはマクドナルド、回転寿司のくら寿司、いくつかのレストランが入っている複合商業施設京都ファミリーが、大学キャンパス南側の四条通り沿いには、コンビニ、ラーメン店、定食屋などいろいろな飲食店があります。
- ・ **校内での喫煙について**：敷地内では、「指定喫煙場所」（8号館1Fホール裏と11号館裏）においてのみ喫煙が許可されています。「指定喫煙場所」での喫煙にご協力いただきますようお願いいたします。
- ・ **当日会員について**：会員でない方も、「当日会員」としてご参加いただけます（会費1,000円）。

◆会場案内図◆

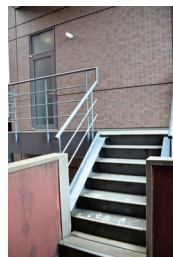


ワークショップ会場へのアクセス

写真 1: 正門を入れて右手に見える 9 号館・国際交流会館。この建物を入れて奥のエレベーターで 4 階へ。エレベーターを降りて右手にある渡り階段で 6 号館 5 階へ。右写真参照。



写真 2: 9 号館 4 階とワークショップ会場となる 6 号館 5 階 652 教室前の廊下へとつながる渡り階段。



大会会場へのアクセス

写真 3: 正門から正面に見える 8 号館。



写真 4: 正門を入れて左手に見える大会会場となる 1 号館入り口。



図 1: 1 号館 7 階平面図。

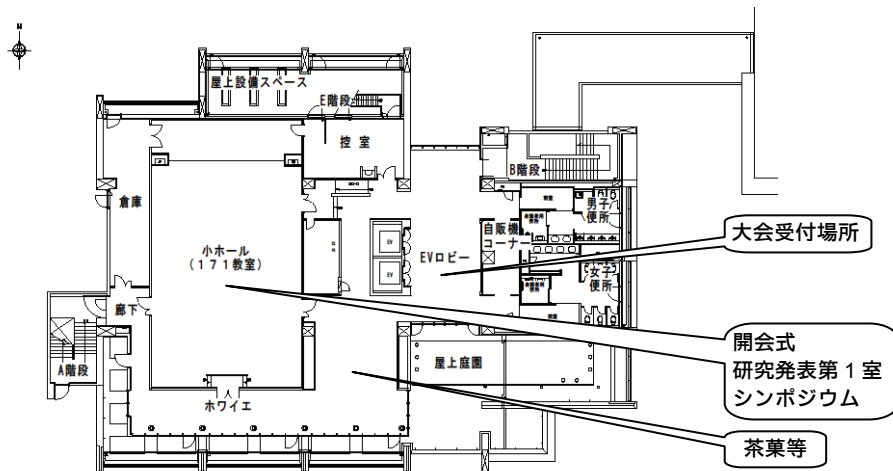
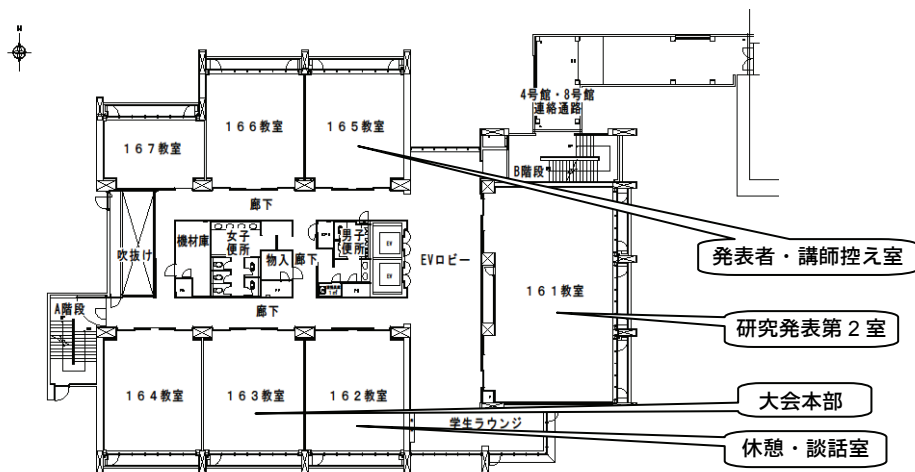


図 2: 1 号館 6 階平面図。



※7 階と 6 階の移動には、エレベーターの他、A 階段と B 階段が利用できます。

懇親会場へのアクセス

写真 5：懇親会場となるラウンジ（2階）がある11号館。



写真 6：2階のラウンジには建物の西側にある階段を利用。



京都駅からの公共交通機関アクセスマップ



※経路

- (1) 阪急電車「西院駅」から、西へ徒歩約15分、またはバス停「西大路四条（西院）」から市バス3・8・28・29・67・69・71に乗り、「京都外大前」で下車（乗車時間約5分）
- (2) JR「京都駅」烏丸口から、市バス28に乗り、「京都外大前」で下車（乗車時間約30分）、京都バス81・83に乗り、「京都外大前」で下車（乗車時間約30分）
JR「京都駅」八条口から、市バス71に乗り、「京都外大前」で下車（乗車時間約30分）
- (3) 地下鉄東西線「太秦天神川駅」から、南へ徒歩約13分

2011年9月1日発行

編集・発行 英語コーパス学会

会長 赤野一郎

事務局 〒739-8521 東広島市鏡山 1-7-1

広島大学大学院総合科学研究科

井上永幸研究室内

電話・ファックス：082-424-6431

email: inoue@v.email.ne.jp

URL: <http://muse.doshisha.ac.jp/JAECS/index.html>