

英語コーパス学会第34回大会

ワークショップ《オンライン版『英文語彙難易度解析プログラム』(Word Level Checker)の利用法およびその教育研究分野への応用》

会場：15号館 502 CALL 教室

時間：10:00～11:45 (9:30 受付開始)

講師：染谷泰正 (青山学院大学)

定員：定員 45名 (先着順・要予約)

参加費：会員無料・非会員 1,000円 (予約申し込みは郵送・電子メールで事務局まで)

日時 2009年10月3日(土)

会場 青山学院大学青山キャンパス

受付開始 12:00 (11号館 1173教室前)

開会式 13:00 (11号館 1173教室)

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

司会 井上永幸 (徳島大学)
赤野一郎 (京都外国語大学)
西澤文昭 (青山学院大学文学部長)

〈研究発表第1室 (11号館 1173教室)〉

研究発表 1 13:40-14:10
低頻度共起でのコロケーション強度を表す指標

研究発表 2 14:15-14:45

新しい lexical richness 指標の提案：学習者の産出語彙における頻度情報に着目して

小島ますみ (名古屋大学大学院生)

研究発表 3 14:50-15:20

Webを源とした英語科学論文コーパスの構築—技術的方法論と法的観点からの検討—

田中省作 (立命館大学)・富浦洋一 (九州大学大学院システム情報科学研究院)

安東奈穂子 (九州大学大学院法学研究院)・柴田雅博 (九州大学大学院システム情報科学研究院)

〈研究発表第2室 (11号館 1143教室)〉

司会 高見敏子 (北海道大学)

西原史暁 (東京大学大学院生)

研究発表 1 13:40-14:10

地域発信用英語コーパスの4-gram分析—事例研究：「かながわ見聞録」に学ぶ定型表現—

後藤隆昭 (熊本大学大学院生)

研究発表 2 14:15-14:45

日本人英語学習者の causality の習熟度：パラレルコーパスの分析を通して

三木 望 (大阪大学大学院生)

研究発表 3 14:50-15:20

BNC サブコーパスの統語構造上の特徴分析

高橋 薫 (豊田工業高等専門学校)

白井翔悟 (プログラマー)

〈休憩 15:20-15:40〉

シンポジウム 15:40-17:50 (11号館 1173教室)

《フリーオンラインコーパスの活用と英語学研究：BYU-Corporaを中心に》

フリーオンラインコーパス概観

ANC との比較による BYU-COCA の活用と研究事例

BYU-OED の特徴と BYU-OED を利用した研究事例

複数のオンライン版 BNC と BYU-BNC を活用した英語学研究

閉会の辞

司会 新井洋一 (中央大学)
講師 林 裕 (関東学院大学)
講師 山崎 聡 (千葉商科大学)
講師 新井洋一 (中央大学)
講師 染谷泰正 (青山学院大学)

《懇親会 時間：18:15-20:00 場所：IVY HALL 青学会館2階「シノノメ」 会費：4,000円》

英語コーパス学会 (Japan Association for English Corpus Studies)

会長 赤野一郎 事務局 〒770-8502 徳島市南常三島町1丁目1番地 徳島大学総合科学部 井上永幸研究室

電話・ファックス：088-656-7125 email: inoue@ias.tokushima-u.ac.jp 振替口座：00940-5-250586

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

◆大会当日、入会受付も致しますので、お誘い合わせの上ご参加下さい (年会費 一般 5,000円 学生 3,000円)。

◆「当日会員」としてのご参加も受け付けております (会費 1,000円)。

英語コーパス学会 第 34 回大会資料

日時：2009 年 10 月 3 日（土）午後 1 時より（正午受付開始）

会場：青山学院大学青山キャンパス

（<http://www.aoyama.ac.jp/>）

〒150-8366 東京都渋谷区渋谷 4-4-25

第 34 回大開プログラム

ワークショップ 《オンライン版『英語語彙難易度解析プログラム』(Word Level Checker) の利用法およびその教育研究分野への応用》

会場：15 号館 502 CALL 教室

時間：10:00～11:45 (9:30 受付開始)

講師：染谷泰正 (青山学院大学)

定員：定員 45 名 (先着順・要予約)

参加費：会員無料・非会員 1,000 円 (予約申し込みは郵送・電子メールで事務局まで)

日時 2009 年 10 月 3 日 (土)
会場 青山学院大学青山キャンパス
受付開始 12:00 (11 号館 1173 教室前)
開会式 13:00 (11 号館 1173 教室)

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

司会 井上永幸 (徳島大学)
赤野一郎 (京都外国語大学)
西澤文昭 (青山学院大学文学部長)

〈研究発表第 1 室 (11 号館 1173 教室)〉

研究発表 1 13:40-14:10
低頻度共起でのコロケーション強度を表す指標

研究発表 2 14:15-14:45

新しい lexical richness 指標の提案：学習者の産出語彙における頻度情報に着目して
小島ますみ (名古屋大学大学院生)

研究発表 3 14:50-15:20

Web を源とした英語科学論文コーパスの構築 —技術的方法論と法的観点からの検討—
田中省作 (立命館大学)・富浦洋一 (九州大学大学院システム情報科学研究院)
安東奈穂子 (九州大学大学院法学研究院)・柴田雅博 (九州大学大学院システム情報科学研究院)

〈研究発表第 2 室 (11 号館 1143 教室)〉

司会 中條清美 (日本大学)

研究発表 1 13:40-14:10

地域発信用英語コーパスの 4-gram 分析—事例研究：「かながわ見聞録」に学ぶ定型表現—
後藤隆昭 (熊本大学大学院生)

研究発表 2 14:15-14:45

日本人英語学習者の causality の習熟度：パラレルコーパスの分析を通して
三木 望 (大阪大学大学院生)

研究発表 3 14:50-15:20

BNC サブコーパスの統語構造上の特徴分析
高橋 薫 (豊田工業高等専門学校)
白井翔悟 (プログラマー)

〈休憩 15:20-15:40〉

シンポジウム 15:40-17:50 (11 号館 1173 教室)

《フリーオンラインコーパスの活用と英語学研究：BYU-Corpora を中心に》

フリーオンラインコーパス概観

司会 新井洋一 (中央大学)

ANC との比較による BYU-COCA の活用と研究事例

講師 林 裕 (関東学院大学)

BYU-OED の特徴と BYU-OED を利用した研究事例

講師 山崎 聡 (千葉商科大学)

複数のオンライン版 BNC と BYU-BNC を活用した英語学研究

講師 新井洋一 (中央大学)

閉会の辞

染谷泰正 (青山学院大学)

〈懇親会 時間：18:15-20:00 場所：IVY HALL 青学会館 2 階「シノノメ」 会費：4,000 円〉

【ワークショップ】

オンライン版「英文語彙難易度解析プログラム」(Word Level Checker) の 利用法およびその教育研究分野への応用

染谷泰正 (青山学院大学)

講師の所属する青山学院大学文学部英米文学科では、主として同文学部で開講されている「英文講読」の授業に使う教材の難易度を客観的に測定するための支援ツールとして表題のオンラインプログラム(以下、WLC)を開発し、2006年度からこれを同学科のウェブサイト上で試験的に公開している。2008年度にはかねてからの懸案事項のひとつであった「リーダビリティ」(テキストの「読みやすさ」の指標)の自動判定機能を追加した。本ワークショップでは、まず前半部(10:00-10:45)でWLCの各種機能および具体的な使用方法について解説する。また、新たに追加されたリーダビリティ評価機能について、その信頼性を検証するために行った評価実験の結果についても報告する。後半部(10:45-11:45)では、WLCの教育研究のさまざまな場面への応用—たとえば「英文講読」授業への応用、「英文ライティング」分野への応用、あるいは大学入試問題の作成と品質管理への応用等—について、参加者の関心に応じた具体的な事例を取り上げ、テキスト入力から解析、結果の評価、対象テキストのプロファイリング、あるいは(目標とする評価レベルに合わせるための)テキストの改変・編集等の一連の作業を行う予定である。

【研究発表第1室】

【研究発表1】

低頻度共起でのコロケーション強度を表す指標

西原史暁 (東京大学大学院生)

本研究は、コロケーションの強さを示す指標について議論し、コロケーションの生頻度(raw frequency)自体が低い(だいたい頻度が20以下)とき、相互情報量を用いるのは適切でなく、コーパスサイズを大きくするか、ダイス係数のような他の指標を用いるべきであると主張する。

相互情報量もダイス係数も語と語の結びつきの強さを表す指標である。相互情報量は、片方の語が出てきたとき次の語がある語となる確率に基づく指標であり、情報理論上の意味づけがしっかりとされている。ダイス係数は、2つの語のそれぞれの頻度の合計に対する2つの語の共起する比率を示しており、あるコロケーションがその構成される2つの語に対してどれだけよく用いられるかを示す。

Hunston (2002: 73) では、相互情報量はコーパスサイズに特に依存しないとされている。しかし、Hall & Morton (1993) や Paninski (2003) で示されているように、相互情報量は標本数が少ない場合、母集団での値から外れた値をとる。このため、小規模のコーパスでは、コーパス上の値と母集団の値が異なると予測できる。また、久光・丹羽(1997)では、「低頻度要素の過大評価が生じる」ことが指摘されているが、実際どの程度の低頻度の時に問題が生じるかについてはよく研究されているとは言えない。

これらを検証するため、事前に与えられた単語・コロケーションの生成確率に基づき架空のコーパスを作成するシミュレーション実験と、英語版 Wikipedia の記事データという実際の大規模言語データを用いた実験を行った。

実験の結果、1つのコーパスの中で、あるコロケーションの生頻度が20以下の場合、そこから得られる相互情報量は、理論上の母集団における相互情報量に比して、高めの値が出る傾向があることが明らかになった。標本での最頻値は、母集団の相互情報量より高い値になる。これに対して、生頻度やダイス係数は最頻値と母集団における理論値とが合致し、しか

も低い頻度でも、標本における数値が正規分布する特徴を持つ。

生頻度が 20 以下というのは、いかにもデータ数が少なく、そこから推論をするのは危険であると考えられる人も少なくないだろう。しかし、例えば、100 万語程度で生頻度が 20 以下となる例として、“United Nations”, “familiar with” (BNC ではそれぞれ 100 万語あたり 13.6 回, 11.9 回) のようなコロケーションが存在する。これらを抽出するには、生頻度が低いときに不適切な相互情報量は用いるべきでなく、ダイス係数を用いた方がよい。

【研究発表 2】

新しい lexical richness 指標の提案：学習者の産出語彙における頻度情報に着目して

小島ますみ (名古屋大学大学院生)

子どもや L2 学習者の産出する語彙を量的に測定する試みがなされてきた (e.g. TTR, Guiraud Index)。しかし、Type と Token のみに基づく指標では、例えば、“The bishop observed the actress” と描写する人と、“The man saw the woman” と描写する人は、区別されない。近年、語彙の一般的な使用頻度に基づき、学習者の産出する語彙の豊かさ (lexical richness) を分析する指標が提案されてきている (e.g. LFP, Advanced Guiraud, P_Lex)。語彙の頻度は、語彙の習得順序を決める要因の一つに過ぎないが、最も重要な要因の一つに挙げられる (Nation & Beglar, 2007)。しかしながら、現在提案されている指標は、スコアがテキストの長さの影響を受ける、明らかに習熟度の異なる学習者を区別しないなど、問題点も多い。レベルの異なる学習者を区別しない原因の一つは、これらの指標の多くが、語彙を basic と advanced に 2 分するのみである点にある可能性がある。そこで本研究は、言語産出者の語彙を頻度レベルで集計し、累積使用率をグラフにしたときのカーブから得られるパラメータを lexical richness の指標とすることを提案する。累積使用率カーブのモデル式には、関数式 $\ln(x)/\ln(F) \times 100$ を用いる。語彙の使用頻度を頻度の高い順に累積する場合、対数グラフに近似できると考えられることから (Meara, 2005)、この式が導き出される。x は語彙の頻度順位、F は使用語彙の頻度レベルパラメータを表す。例えば、F 値が 1,230 と推定される場合、その学習者が産出する語彙は、そのデータにおける平均的なパフォーマンスにおいて、頻度レベル 1,230 ワードファミリー以内であることを示す。

パラメータの具体的な算出方法を述べる。ある言語産出者のデータから 50 語を無作為に抽出し、4 つの頻度レベル (500, 1000, 1500, 2000) での累積使用率を求めることを 100 回繰返す。この 4×100 のデータをもとに、パラメータ F を推測する。頻度表は、BNC に基づくワードファミリーの頻度表 (Nation, 2006) を用いた。

学習者コーパス NICE を対象に、TOEIC700 で分けた中級者と上級者、及び母語話者 (全グループ N=17) のスコア (F 値) を比較したところ、中級者の平均が 2,007 (SD=603.3)、上級者の平均が 2,522 (SD=803.8)、母語話者では 3,072 (SD=722.6) であった。分散分析と多重比較の結果、全てのグループ間に有意差が見られた。また、データのモデルへの当てはまりは大変良く、対応確率 p 値は、全てのデータでほぼゼロであった (ゼロに近いほど良い)。今後は、指標の信頼性について調査を進める予定である。

【研究発表 3】

Web を源とした英語科学論文コーパスの構築
—技術的方法論と法的観点からの検討—

田中省作 (立命館大学)

富浦洋一 (九州大学大学院システム情報科学研究院)

安東奈穂子 (九州大学大学院法学研究院)

柴田雅博 (九州大学大学院システム情報科学研究院)

近年、英語科学論文の言語的特徴の解明に、コーパス・ベースのアプローチが成果を挙げつつある。例えば田中他（2006）は著者の母語を問わず比較的英文の質が高い論文と日本人の論文を収集し、それらを品詞列レベルで対比することで、日本人が書く論文の特徴の一端を明らかにしている。このようなアプローチをさらに語彙、そして構文といった高次なレベルで網羅的に進めようとする、英文の質的情報が付与された、より大規模な英語科学論文コーパスが必要となる。しかし、このようなデータは現時点では十分に整備されていない。その一方で科学論文は学会等が有する著作権に反しない場合は、その著者らが自主的に Web で公開していることも少なくない。本研究はこのような Web 上の論文を自動的に収集し、質的情報を付与、コーパス化する試みである。本発表では、前半にその技術的な方法論を概括し、後半に実作業に伴う、特に法的な問題を考察する。

本研究は、まず電気情報関係の国際会議録をもとに、論文の質的情報（論文中の誤りの種類と回数に基づき英文校正者が 5 段階で評価）を付与した 1,281 編の論文データを人手で整備した。そして、品詞可変 n-gram に基づいた英語論文の質的情報（ただし、Good/Poor の二分で推定）を判別するシステムを開発した。評価実験で 92.5% の精度を得ている。次に、予め設定したキーワードに基づいて Web 上の英語科学論文を研究業績に関するページなどから自動収集するシステムを開発した。これに前述のシステムによる質的評価を付与し、コーパス化する。予備実験では、1 万アクセス（約半日で無関係なページのアクセスも含）で 200 編程度の電子化論文が収集されることが確認された。

また、Web 上で自主的に公開されている論文であっても、それらをコーパスとして取り込むとなると、知的財産や著者らの感情にも配慮する必要がある。本年 6 月に著作権法の一部改正（文部科学省、2009）が行われ、来年より施行される。その一つである第 47 条の 7（情報解析のための複製等）は、情報解析を「多数の著作物その他の大量の情報から＜中略＞情報を抽出し、比較、分類その他の統計的な解析を行う」と規定し、必要と認められる限度において著作物の記録・翻案が可能であるとしている。本研究はこの条項により説明されるものと考えており、来年以降に本格的なコーパス構築を予定している。このような法的説明がなされたとしてもなお、著作者人格権や著者らの感情には留意しなければならない。したがって、本研究は個々人ではなく日本人に共通する科学論文の特徴を追及し、コーパス構築の目的・分析過程の十分な説明に努め、論文の特徴などの研究成果については著者らの不利益とならないよう、個人が特定されない形での公開を検討する。

【研究発表第 2 室】

【研究発表 1】

地域発信用英語コーパスの 4-gram 分析 —事例研究：「かながわ見聞録」に学ぶ定型表現—

後藤隆昭（熊本大学大学院生）

大学などの英語教育において、身近な地域社会を取材し、メディアを通して英語で地域情報を発信する取り組みが行われている（「地域丸ごと翻訳力（東京国際大学）」「Decoding Kyoto Project（立命館大学）」）。こうした実践は、地域社会と連携しながら、学習者の英語発信能力を高めようとする点で興味深いプログラムでもあると思われる。

今回はそうした取り組みのモデルとなりうるテキストとして「かながわ見聞録」（バーリット・セービン著、朝日新聞・朝刊・神奈川版、約 2 万語、80 トピック）を取り上げる。この中で著者は神奈川内で見聞きした身近な出来事を英語でレポートしており、日本人読者が対象のため、英語も分かりやすく、語数も多くて 400 字前後と、適度な長さに収まっている。

研究の目的は、N-gram を利用し、書く際に参考にできるような定型表現を抽出することであり、リサーチ・クエスチョンとして、1) どのような定型表現がよく使われているのか、2) 他のジャンルと比べると、どのような言語学的特徴があるのか、の 2 点を明らかにしてみたい。

研究方法としては、テキストをコーパス化し、分析ソフト Antconc (Anthony, 2004; v 3.2.1) で 4-gram 分析を行う。4-gram に絞ったのは、一般的に 4-gram が適している (Cortes, 2004; Hyland, 2008) との指摘があることや、先行研究の枠組みを利用したり、分析結果を比較することを考慮したためである。焦点を 4 語の定型表現の機能面に絞り、Biber et al. (2004), Cortes (2004), Biber & Barbieri (2007), Hyland (2008) の分析枠組みを参考にし、トピックの影響によるものを除いて、新たに、時間表現、場所表現、理由表現、伝聞表現、知覚・認識表現、その他、などに分類を試みる。

結果として、以下のことが明らかにされる。

- 1) 場所 (site, end, foot, front, across) や時間 (first, time, end, last, age, opening) に関する語を含む表現 (the former site of, in front of the, at the end of, first time in ages, of the history of, etc) がよく使われている。
- 2) the reason is (was) that, that is perhaps because など理由を表す構文が見られる。
- 3) 伝聞表現として、is said to be (or have), it is said the, it is though that や reportedly を含むもの (reportedly used as, etc) が確認できる。
- 4) 知覚・認識表現では、驚き (was surprised by the) や、思われる、感じられる等 (seems to have been, feel as if I, fell to thinking why) や、「…したかった」(I would like to, would have liked to, etc)などを組み込んだものが見られる。
- 5) 伝聞や知覚・認識表現には、過去であることを示すため、to 不定詞や助動詞の後に過去分詞の have が使われている (are said to have, seems to have been, would have like to, etc)。

最後に、4-gram による、Nishina (2007) の Newspaper と Literature のジャンル分析の結果と比較する。Nishina は、前者で政府関連や経済に関する語句が多く使われ、特徴として the fact that the の使用を挙げており、後者では、代名詞を伴って、人物の思考、感情、願いなどを描写する口語的表現を使用する傾向を指摘している。本発表では 2 つのジャンルを比較した 4-gram の頻度表から、両ジャンルに共通のものを取り除き、再度、各ジャンルの特徴を解釈し直す。結果として、今回のテキストが両ジャンルの中間に位置するような特徴があることを示したい。

【研究発表 2】

日本人英語学習者の causality の習熟度：パラレルコーパスの分析を通して

三木 望 (大阪大学大学院生)

理由・原因と結果・効果を表す causality (例: *because, reason, so*) は、議論文の論理の中核を形成する。Trimble (1985) は科学系論文における causality の重要性を、Xuelan and Kennedy (1992) は notional syllabus で causation の概念と機能、語彙に注目が集まっていることを指摘している (以下、causation を causality に統一)。しかし、学習者英語に関しては、接続詞あるいは、談話標識の一部としての研究があるが、causality 全体の解明はまだ始まったばかりである (Flowerdew, 1998)。例えば、日本人英語学習者は英語母語話者と比較して、主張や根拠を述べるのが得意でないという指摘がある一方で (大井, 2002)、日本人学習者が初期の段階で *because* の使用を始めるという報告がある (投野, 2007)。本研究発表では、日本人英語学習者がどのように causality の語彙を使用しているのかについて、NICE (Nagoya Interlanguage Corpus of English) (杉浦, 2007) を使用して、分析を行う。

NICE から英語母国語話者による議論文 (NICE-NS) と日本人大学生の議論文 (NICE-NNS)、そして NS による添削文 (NTV) を抽出して、それぞれコーパスを編成した。NICE-NNS は、TOEIC の得点と最終受験期間に基づいて 20 ファイルに絞り込み、High (9,030 語)、Middle (7,898 語)、Low (5,137 語) のサブコーパスとして習熟度別に分類した (Cf. 小島, 2007)。NICE-NS からは 20 ファイルを無作為抽出した (11,267 語)。NICE-NNS と NICE-NTV は、コメント文を参照しながら、行ごとに alignment を行いパラレルコーパス用に再編集した。

初めに NICE-NS と習熟度別の NICE-NNS の causality の語彙の使用について、対応分析で視覚的に要約する (causality の語彙は Flowerdew (1998) の一覧表を一部修正して使用した)。結果は、NS に付帯状況の *with*, *as*, *since* が、NNS の High グループで *reason(s)*, Middle グループで *because*, Low グループで *so* がそれぞれ対応していて、他の統計的手法 (尤度比検定) の結果とほぼ同じであった。発表ではこれらの NNS の causality の特徴語の一部を取り上げて、実際にどのように使用されて、添削されるのかを、NICE の NS のコメントの結果を踏まえながら、分析する。

【研究発表 3】

BNC サブコーパスの統語構造上の特徴分析

高橋 薫 (豊田工業高等専門学校)
白井翔悟 (フリーランスプログラマー)

本発表では、これまでの筆者の BNC の語彙分析に関する研究を踏まえ、特定のサブコーパスで生起率が高い動詞を中心とする統語分析のプログラミングの開発状況を紹介する。また、言語スタイルとの関連から統語構造を分析して、その特徴について言及する。なお、これらの考察と結論については現在検討中で、この紙面にて公表するには至っていないことを申し添えたい。

さて、前回の本学会発表にあったように、単語レベルに留まっていたコーパスの統計分析が、句や構文などのより高次の統語レベルにおいて適応され始め、Link Grammar Parser などの先進性のあるツールの紹介は大変有益であった。また、同発表で「語彙選択と統語構造の選択には密接な関連性」があるとしたことに関して、筆者の予測として、語彙選択がサブコーパスの特徴を反映したものであれば、当然のことながらサブコーパスの統語構造の様相にも差異が生じるものと考え。筆者がこれまで実施した BNC のサブコーパスにおけるタグの頻度分析では、それを特徴的なタグの語彙分析へと進めることで、特定の動詞タグと関連する語彙を見いだすことができた (Takahashi, 2007)。動詞がどの品詞よりも統語構造との関連が強いとの想定に立てば、統語分析の手始めとして、これらの動詞の統語分析に注目したいところである。

具体的には、BNC において、「age 0-14 の少女」は、タグ VVZ (-s 動詞) の使用が他のサブコーパスよりも頻度が高い。動作動詞としては、*goes*, *says*, *gets*, *comes*, *looks*, *gives*, *takes*, *puts* など、また状態動詞では、*wants*, *thinks*, *knows*, *likes* などが挙げられる。ここで Hornby (2000) の動詞パターンを用いた基本文型についてこれらの動詞の文型を当てはめると、*goes*, *comes* などは、V+adv./prep (verb + adverb or prepositional phrase) の型、*gets*, *gives*, *takes*, *puts* は VN (verb + noun phrase) の型であると容易に予想できる。では、*goes* が V-ADJ (verb + adjective) の型はどの程度の割合で含まれるのか、また、このサブコーパスと対極に位置する「age 60+ の男性 (多変量解析の結果に基づく)」では、割合がどう変わるのか、多種の文型を持つ *gets* についてはどうであろうか、との疑問が生じる。

理論的背景としては、言語スタイルの観点から、最も vernacular (日常口語的かつ通語的とも表現すべきか) であるとされる BNC のサブコーパス、「age 0-14 の少女」において、統語的に他のサブコーパスとは著しく異なった特徴を、上記の動詞群の文型の分析より明らかにすることにあると言える。つまり、vernacular style を特徴付ける動詞とその文型を探ることにある。

これまでの筆者の BNC の分析では、口語、文語、社会階層、地域、著者年齢、性別等の指標により、複数の仮想サブコーパスを構築し、それぞれのタグ頻度をカウント後、多変量統計解析の後に現れた尺度について言語スタイルに関する解釈を与え、そのスタイルを特徴づけるサブコーパスを特定した。同時に関連するタグや語彙を明らかにした。「age 0-14 の少女」のサブコーパスについての上記の紹介はその一例である。

現在、特定の動詞がどの文型であるかを自動認識するプログラムを構築しつつ、具体的な動詞の文型の判別を行っているが、詳細な分析結果についての発表を目指して準備中である。

【シンポジウム】

フリーオンラインコーパスの活用と英語学研究: BYU-Corpora を中心に

司 会 新井洋一 (中央大学)

最近、Web 上での直接検索やコンコーダンス表示が無料で可能な、フリーオンラインコーパスが一般的になりつつある。よく知られている無料のオンラインコーパスとしては、たとえば、Web データをそのまま検索対象にしている WebCorp (Kehoe & Renouf 2002, Renouf et al. 2007) が有名であるが、検索対象のデータの信頼性について議論の余地があり、広く利用されるまでには至っていない。一方、かつて CD-ROM で配布されていた BROWN, LOB コーパスは、現在、簡単な検索に限定すれば、Virtual Language Center の Web Concordancer で無料利用が可能になっている。また、最近の数年間でもっとも利用率が高いと思われる BNC については、有料のものとして小学館提供の BNC オンライン、無料のものとしても BNCweb や後に述べる BYU-BNC があり、CD-ROM 版の利用は減りつつある。

コーパスの CD-ROM 版からオンライン版へのこのような流れに決定的な影響を与えたのは、Brigham Young University の Mark Davies であろう。彼の来日講演 (於: 上智大学 2004 年) でその一部が公開された BYU Corpus は、今までの BNC (以下 BYU-BNC), TIME (以下 BNC-Time), Corpus of Contemporary American English (以下 BYU-COCA または COCA) に、今年新たに OED (以下 BYU-OED) を追加し、大規模な複数オンラインコーパスとして発展し続けている。無料利用が可能であるために、現在、世界中の研究者、教師、学生の注目を集めている。

本シンポジウムでは、まず最初に司会者から、Hunt et al. (2007), Anderson and Corbett (2009) を前提に、オンラインコーパスについての歴史と、現在の主要なフリーオンラインコーパスの紹介をおこなったあと、上記以外の、今まであまり紹介されていないものも含め、複数のオンラインコーパスについて整理・概観した上で、特に BYU-Corpora 全体に共通のコーパスデザインの特徴、検索方法を含む主要機能の概要、授業での活用を含む利用上の注意点、利用登録法などについて触れる。そのあと、以下の 3 人の講師がそれぞれ異なる専門分野から発表をおこなうことにしたい。具体的には、個々のオンラインコーパスとオリジナル CD-ROM 版や類似コーパスとの相違、オンラインコーパスの長所・短所、事例研究に必要な検索方法の具体例、ジャンル別分析および通時的分析の方法、具体的な英語学研究例や成果、などについて報告する予定である。

ANC との比較による BUY-COCA の活用と研究事例

講 師 林 裕 (関東学院大学)

米語大規模コーパスとしてよく知られているものとしては、BNC と同じ 1 億語を目標としつつ、現在 Second Release で 2200 万語の American National Corpus (ANC) がある。これに加えて、フリーオンラインで利用可能な 3 億 8500 万語の Corpus of Contemporary American English (COCA) が、2008 年から公開されている。

本発表では、COCA に関する Davies (2009) を紹介したあと、(1) ANC と COCA のコーパスデザインの違いや特徴、特にジャンル区分の相違点、(2) 両者の話しことばテキストの違い、(3) 使いやすさの比較、などについて具体例を示しながら検討する。また (3) に関連して、ANC (DVD 版) を入手してから Xaira による検索にいたるまでの作業についても触れる予定である。

研究事例としては、BNCを元にした林(2005)を踏まえた上で、COCAを元にした英語形容詞の配列順序について考察する。前置修飾の形容詞が複数ある場合、順序に制約があることは古くから指摘されてきた。「サイズ」と「色」の順序(たとえば“the big red ball”のbigとred)のような、かなり固定した順序関係についてCOCAの特徴であるユーザーリストを活用して検索し、この順序に反する例を取り出して共通する要因を探りたい。

BYU-OEDの特徴とBYU-OEDを利用した研究事例

講師 山崎 聡 (千葉商科大学)

本発表では、まず、OED2 CD-ROM, Ver. 3 (2002), 同 Ver. 4 (2009) や OED Online にはない、BYU-OED の使い勝手の良さと、一方で、利用に際して不便を感じる点、注意を要する点を概観する。次に、BYU-OED を基本的なコーパスに据えた研究事例として、複合前置詞句 according to を取り上げ、「情報源を表す用法」の発達について考察する。考察に際して、OED の用例は文脈確認ができにくい点でしばしば批判を受けてきているが、やはり無料で可能なその克服法のひとつについても紹介したい。

OED2 では、“In a manner agreeing with, consistent with, or answering to; agreeably to” (s.v. **according** *ppl. a. and adv.* B4a) の語義(以下、adjunct 用法)は示されているが、*According to the police, his attackers beat him with a blunt instrument.* (LDOCE4) にみられる、情報源を表す disjunct (Quirk et al., 1985) あるいは evidential (Bolinger, 1990) の用法については明言がない。そこで BYU-OED をおもに使って 14 世紀以降の according to の引用例を検討する。まず adjunct 用法と disjunct 用法の違いを確認したうえで、以下の点を中心に報告する予定である。(1) 情報源を表す disjunct 用法は、どのような事例がきっかけとなって、いつ頃から発生したと考えられるか。(2) この用法の生起は、既存研究における Swan (1991) の指摘と類似の特徴を持つ。(3) この用法は、当初は通例文頭または動詞の前に生起し、文末は極めてまれであったが、時代とともにどのような変化が見られるようになっているか。またその傾向は英語学的にどのような説明が可能か。

複数のオンライン版 BNC と BYU-BNC を活用した英語学研究

講師 新井洋一 (中央大学)

世界的にもっとも利用頻度が高い BNC については、オンライン版のものとして、有料の小学館提供の BNC オンラインと、無料の BNCweb および BYU-BNC がある。本発表ではまず、これらの複数のオンライン版 BNC を、特に用例の検索機能の面から比較し、POS Tag の活用と検索式の組み合わせ法、ジャンル別分析法の面から考察する。また、特に意味概念データベースの WordNet との融合をはかった BYU-BNC の検索機能の長所とその問題点について触れるとともに、Davies (2007) で触れられながら現在利用できないものの、確実な将来性が見込まれる上位語 (hypernym) と下位語 (hyponym) に関わる検索機能について触れたい。

具体的な英語学研究例としては、まず、話し言葉に特異な言語変種のひとつである I goes 表現について、最近の研究成果を参考にまとめてみることにする。次に、Dr Postlethwait **is accustomed to laugh heartily and reply** / The new businessmen-buyers **were accustomed to using debt as an investment tool** の例に見られる、be accustomed to の不定詞補文と動名詞補文選択の問題 (Rudanko, 2006) について、BYU-BNC を足がかりに、BYU-OED, BYU-TIME などとの連携によって、興味深い考察が可能であることを明らかにしたい。最後に、今までほとんど取り上げられていない「非人称主語受動態挿入文」(e.g. Julie, *it is believed*, tried to run for it) について、従来から研究例の多い人称主語能動態挿入文と比較しながら、特にその生起ジャンルの特性について考察し、その英語学的意味について明らかにする予定である。

《大会参加者へのご案内》

- ・自家用車でのご来場はできません。
- ・ワークショップの受付は「15号館 502 CALL 教室」の前で午前9時30分から行います。
- ・大会の受付は「11号館 1173 教室前」で正午から行います。
- ・昼食ですが、食事のできる場所は青学周辺にたくさんあり、正門、西門、東門（幼稚園・女子短大門）のいずれの出口から出ても探すのに困難はありません。ただし、せっかくの機会ですので、青学の学食をご利用いただくのが一番いいのではないかと思います。土曜日は午後8時まで開いています。学食は12号館または7号館から地下に降りていただくとあります（案内板がありますのですぐわかります）。
- ・校内は分煙措置がとられています。指定場所での喫煙にご協力ください。喫煙ルームは正門に入って左側の広場（7号館脇）にあります。
- ・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）。

キャンパスマップ



※正門に入って約50m先、右側にある15号館（ガウチャーメモリアルホール）の正面玄関の左側（正面はチャペルへの入り口になります）にある教室棟入り口を入るとすぐ前にエスカレーターがあります。これを使って3階に上がってください。

※大会本部＋発表者・講師控え室：11号館 1140 教室
予備教室（受付・お茶・談話室）：11号館 1172 教室



写真 1：大会会場は、正門を入れて20m くらいのところ、右手に「広報入試センター」の看板が見えますので、ここを右手に折れます。

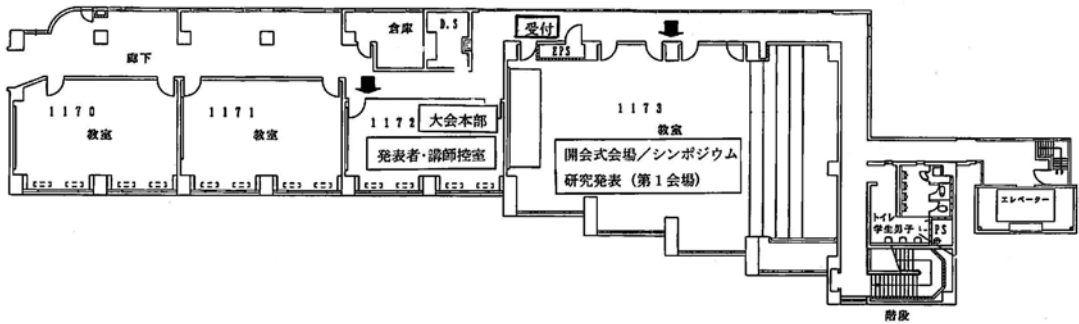


写真 2：写真 1 の場所で右に折れると、通路の奥に 11 号館が見えます。

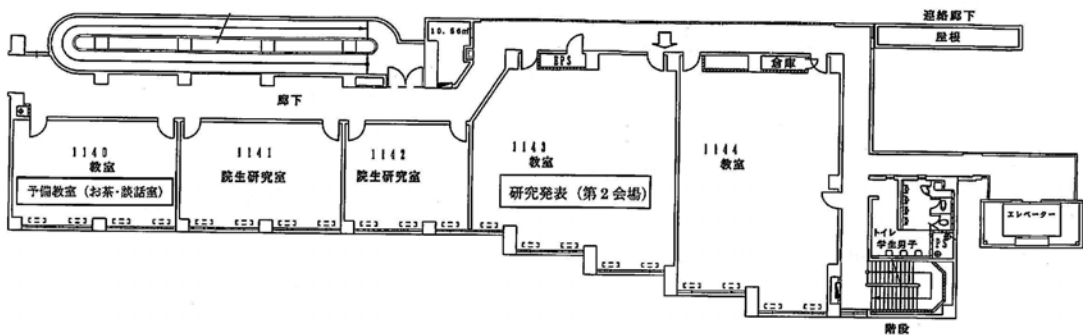
写真 3：ワークショップ会場は、ガウチャー・メモリアル・ホール（15 号館）の正面左手にある教室棟入口に入ってすぐのエスカレーターで 5 階へどうぞ。



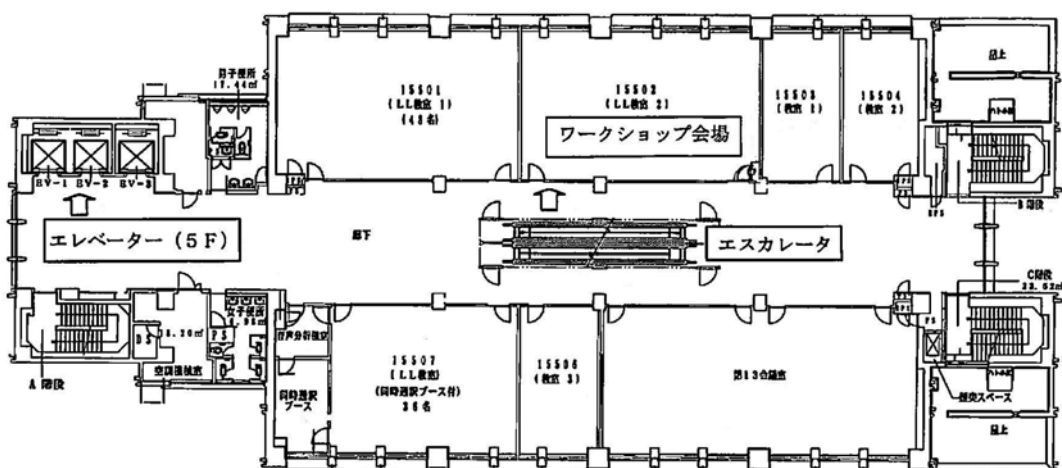
写真 4：懇親会場の IVY HALL 青学会館です。



平面図 1：大会会場 11 号館 7 階

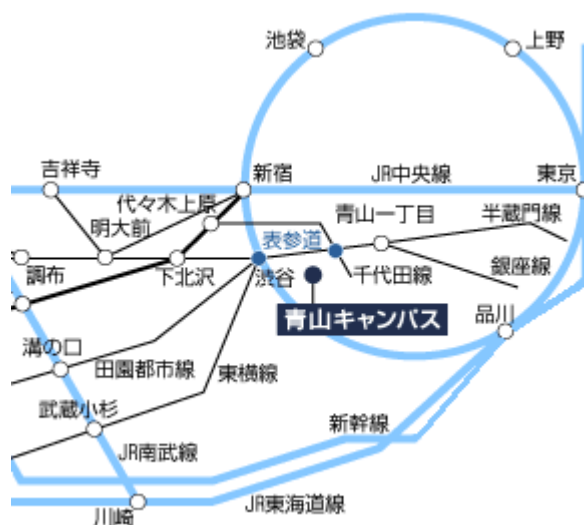


平面図 2：大会会場 11 号館 4 階



平面図3：ワークショップ会場 15号館 5階

アクセスマップ



- ・ JR 山手線，東急線，京王井の頭線「渋谷駅」宮益坂方面の出口より徒歩約 10 分。
- ・ 地下鉄「表参道駅」B1 出口より徒歩約 5 分のところにあります。

2009 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

〒770-8502 徳島県徳島市南常三島町 1 丁目 1 番地

徳島大学総合科学部 井上永幸研究室内

電話・ファックス：088-656-7125

email: inoue@ias.tokushima-u.ac.jp

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