英 語 コ ー パ ス 学 会 第 16 回 大 会

日　時　　**2000 年 10 月 14 日(土)**

会　場　　**大東文化大学板橋校舎（〒175-8571 東京都板橋区高島平 1-9-1　℡ 03-5399-7319〔広報部〕）**
（東武東上線「東武練馬」駅北口より徒歩 20 分、同駅北口「大東文化会館」前よりスクールバス（5 分）．或いは
都営地下鉄三田線「西台」駅より徒歩 10 分．詳細は http://www.daito.ac.jp/ **参照）**

ワークショップ　10:00－12:00
《コーパス利用のための UNIX の最初歩》　　**講 師　名古屋大学　滝沢 直宏**
先着　30 名（予定）　参加費　会員無料・非会員 1,000 円　(申し込みは電子メール・郵便で事務局まで)

受付開始　12:30

開　会　13:00
1．会長挨拶　　**大東文化大学　齊藤　俊雄**
2．会場校挨拶　　**大東文化大学学長　須藤　敏昭**
3．その他

研究発表　13:20－14:50
司 会　日本大学　保坂　道雄　　明治大学　久保田俊彦

1．The Penn-Helsinki Parsed Corpus of Middle English に付加された統語情報　－分析への活用－
日本大学　塚本　聡

2．LOB コーパスにおける動詞の活用パターンとテキスト範疇
山形大学　岡田　毅

3．however の位置は何を示しているか
大阪女子大学　橋本喜代太

〈休　憩　14:50－15:10〉

特別講演　15:10－17:30
司 会　大東文化大学　山崎　俊次

Corpus Linguistics and Language Teaching
講 師　Northern Arizona University　Randi Reppen

Corpus Linguistics and the Study of English Grammar
講 師　Northern Arizona University　Douglas Biber

閉会の辞　　**大東文化大学　山崎　俊次**

《懇親会　17:45－19:30　　**会費 4,000 円》**
司 会　慶応大学　吉村　由佳

英 語 コ ー パ ス 学 会　(Japan Association for English Corpus Studies)
会長　齊藤俊雄　　事務局　770–8502 徳島市南常三島町 1-1　徳島大学総合科学部　中村純作研究室
TEL:　088-656-7129　　E-mail: jun@ias.tokushima-u.ac.jp　　**郵便振替口座**　00940-5-250586
URL　http://muse.doshisha.ac.jp/JAECS/index.html

◆大会当日、入会受付もいたしますので、お誘い合わせの上ご参加下さい（年会費　一般 5,000 円　学生 4,000 円）．
また「当日会員」としての参加も受け付けております（1,000 円）．

英語コーパス学会第16回大会レジュメ

◆ワークショップ《コーパス利用のためのUNIXの最初歩》

（講師　滝沢　直宏）

最近では，Linuxなどの UNIX 系オペレーションシステムの普及により研究者が UNIX 環境で電子テキストを扱うことが増えてきていますが, それでもまだ個人レベルではWindows や Macintosh を使うことが多いだろうと思われます．今回のワークショップでは，UNIX 環境で作業をしたことがない方を対象に，UNIX 環境でのテキスト処理の最初歩を, 実際に UNIX マシンに触れながら行います.

まず最初に，会場校のマシンから telnet（ネットワーク上の他のコンピュータを操作するための仕組み）によって遠隔地の UNIX マシンに接続し，「ログインの仕方」，「ファイルの内容の見方」，「コマンドの出力のファイルへの書き出し方」など，基本的な内容についてお話しした後，FTP (file transfer protocol，ファイル転送) によって, 一般に公開されている電子テキストを取得します．次に，そのファイルに対して，UNIX の様々なコマンドを実行することによって，例文検索その他の作業を行っていきます.

具体的にお話しするコマンドは，grep，sort，uniq などです．grep とは，大雑把に言えば，指定した文字列を含む行をテキストから抽出するためのコマンドです. この検索ではいわゆる「正規表現」を使います．正規表現を用いると，例えば，「小文字のアルファベット一文字の０回以上の繰り返し」は「[a-z]*」として表現され，「行頭にある This」は「^This」として表現されます．正規表現を使うことで, かなり自由な文字列検索が可能となります．sort というコマンドは，各行をアルファベット順あるいは数値順に並べ替える機能を持っています．テキスト中の単語の一覧表を作ろうとする場合，スペースやコンマなどの文字以外の記号を改行マークに置き換え，一行一語の構成にした上で sort をかけると，同じ語からなる行がアルファベット順に一箇所にまとめられます. その上で同じ語の行が複数回出ている場合に，その重複行を除去してくれる uniq というコマンドをかけると, ファイル中に用いられている単語の一覧表が作成されることになります．uniq には，重複行を除去する際にもともと何行重複していたかを表示する機能もあります．その機能を利用して実行した uniq の結果に，再び sort をかけると，出現回数順の単語表が作成されることになります．UNIX の各コマンドは，単純な機能しかもっていませんが, 複数のコマンドを組み合わせて使うことで, テキスト処理の上で結構便利なことができるわけです.

さらに，sed, awk, perl を用いた簡単な一行スクリプトも紹介します. これらのスクリプトも UNIX コマンドと同様に手軽に利用できます．時間の余裕があれば，こうしたスクリプトと UNIX コマンドの組み合わせ方も, 具体的にお話ししようと思います.

なお，参加される方々には，事前に，齊藤俊雄・中村純作・赤野一郎(編) 1998.『英語コーパス言語学：基礎と実践』(研究社)の第３章「コーパスを編纂する」（赤野一郎先生ご執筆）および第６章「コーパスに基づく語彙研究」（園田勝英先生ご執筆）をご一読下さいますようお願い致します.

◆研究発表

●The Penn-Helsinki Parsed Corpus of Middle English に付加された統語情報　－分析への活用－

（塚本　聡）

約 50 万語の中英語散文から成る The Penn-Helsinki Parsed Corpus o f Middle English（以下 Penn-Helsinki コーパス）には，統語情報に関して，名詞句（subject, direct object, indirect object etc.），動詞句（tensed main verb, tensed auxiliary, present/past participle etc.）などの品詞情報，さら main clause (quotation, question, imperative etc)，subordinate clause (adverbial, relative, that clause, infinitive clause, etc) などの文法標識が付加されている．この Penn-Helsinki コーパスが扱う中英語期の英語は, 変動が大きく，しかも，方言差をはじめ，フランス語・ラテン語など外国語との関係，時間の推移，文献の扱う内容の相違など, 非常に多くの条件によって英語が特徴づけられている．従って，ある一定の尺度でテキストを分類すれば, 他の視点での分類を視野の外において考慮することとなり, 包括的な分類をすることがきわめて困難である.

当コーパスを使用した今までの研究では, 同音異義語の判別のためなど，語彙レベルの利用を除けば，これらの情報がほとんど利用されていない．一方，文法カテゴリーや語彙などの情報を用いた著者推定に関しては, すでに多くの研究がある．今までの研究を考慮すると，この Penn-Helsinki コーパスにおいても，著者推定においてみられるように，これらの情報が分類・区分に大いに貢献できるであろうとの予測される．

当発表では, Penn-Helsinki コーパスにおけるこれらの文法標識を利用したテキストの分類・分析を行う．これらの標識と，テキストの分類との関係を検討し，もしあるならば, どのような文法カテゴリーがテキストの分類（ひいてはテキストの特徴）に大きく関わっているのか, これらの手法がどの程度の寄与ができるのかについて検討したい．

●LOB コーパスにおける動詞の活用パターンとテキスト範疇

（岡田　毅）

本発表の内容は次の３部から構成される．LOB コーパスに出現する全動詞の活用形を網羅した「動詞表」の自動作成手法とその意義・応用の可能性．合計 31 種類の活用パタンごとに代表的な動詞の抽出と，代表的な動詞のテキスト範疇に亘る分布傾向の調査・分析．LOB における 15 個のテキスト範疇が反映する活用パタンの傾向調査・分析．

リリース以来，約四半世紀を経ても，英語コーパス研究の対象として，高い評価を受け続けている Tagged LOB コーパスに出現した全ての動詞の５つの活用形: _VB(原形), _VBZ(３単現), _VBD(過去形), _VBN(過去分詞形),_VBG(ing 形)を反映させた「動詞表」を作成し，これに準拠して LOB コーパスにおけるテキスト範疇ごとに動詞の用法に関してどのような性格と傾向が見られるかを分析する．これと同時に，一般的な概念として同種の範疇に分類されがちなテキストであっても, 動詞の用法という観点からは, 別個の範疇に所属させるべきではないかというような疑問も提示できればと考える．

本発表で言及する「動詞表」は 1993 年に完成しており, MS-DOS 上で稼動する一連の処理過程を経て生成される．「動詞表」の大きな意義は，特定動詞の特定の活用形を集計するだけではなく, ５つの活用形のうちのいずれかがコーパス中では出現しないような場合に, 実際に出現した活用形に準拠して欠落した活用形を自動生成し, 結果的に全ての動詞の活用パタンに関する特徴を的確に把握できるところにある．

５つの活用形全てを伴って出現している動詞類を，FLAG01 と呼び，_VBG だけが欠落しているパタンを示すものを FLAG02，_VBZ のみで用いられている動詞を FLAG24 などと分類すれば，その可能な組み合わせ FLAG 総数は 31 個になる．５つ全ての活用形で現れる動詞の出現回数は, 全動詞の token 数 104136 の 81.557% に相当する 84930 回に登るが，一方で，_VBZ 形のみが欠落して用いられている動詞は 6.27%に当たる 6529 回出現している．更に_VBN でのみしか用いられないという，活用形の観点からは特異と思われる動詞でも，936 回の出現数を伴ったりする．このような性格を持つ 31 種類の FLAG ごとに典型的な動詞を抽出し, それらが, 15 個の LOB のテキスト範疇にどのように分布しているかを探ることにより, written コーパスにおける動詞の具現形の特徴を探求する．

実際の調査・分析に当たっては，特に_VB と_VBN について慎重な下位分類化が必要とされる．これらの活用形が実際のコーパス中で用いられている環境を, 例えば不定詞内，否定辞先行，法助動詞先行，完了形内，受動態内等の別に整理した．

最後に，Biber, et al. (1999, pp.367-371)での the most common lexical verbs に対して，「動詞表」における FLAG に着目した再整理を試みてみたい．

●however の位置は何を示しているか

（橋本　喜代太）

いわゆる接続副詞の however は周知の通り, 次のように文中位置が可変である．

(1) a. However, John is kind.
　　b. John is, however, kind.
　　c. John is kind, however.

こうした however はいわゆる文間の接続関係を示すものとされてきた．このように考える限り，(1)の however 位置の違いが何を意味しているかは導けない．また，談話意味の研究として however と but との違いを前提の捉え方の違いから導くといった研究は関連性理論の枠組みなどで少なからずあるが, ここでも比較対象である but が等位接続詞であり語順が固定されているためか，however の位置の違いが何を意味しているかは触

れていない．規範的にはFowlerやGarnerの語法書で触れられており，前文との対比焦点となる語句がhoweverの左に来るとする．記述的な研究はきわめて少ないが，Bell (1994)は基本的に同じ立場に立ち，同時に，Ifantidou (1994)と同じく，however が文末に近づくほどいわゆるafterthought のニュアンスが強くなるとしている．

しかし，規範的，記述的問わず，これまでのhoweverの位置についての主張はいずれもきわめて単純な文しか扱っておらず，例えば，

(2) Be aware that comment, however, is . . .

のような例は説明がつかない．

本発表では，Brown コーパス等の既製コーパスだけでなく，発表者が作成した現代アメリカ英語コーパス(CAMEC)を利用して，2万例以上のhowever用例を収集し，記述的な観察を行なう．CAMECは新聞，雑誌を中心とした書き言葉5000万語，インタビューや記者会見，CNN等の討論番組を中心とした話し言葉1000万語からなる建設中のコーパスである．これらにより，まず，文末のhoweverが書き言葉ではafter-thoughtであり得ない例が多数発見された．

(3) Colquitt Policeman Tom Williams said. "Being at the polls was just like being at church. I didn't smell a drop of liquor, and we didn't have a bit of trouble". The campaign leading to the election was not so quiet, however. It was marked by controversy, anonymous midnight phone calls and veiled threats of violence. (Brown: A01, 1760-1810)

このように文末のhoweverは上記第1文に対する対比として，第2，3文からなるまとまりを結び付けていることがある．また，この例からも見て取れるように，however の位置は単にhoweverの左に何が来るか，だけではなく，howeverの右に何が来るかも考える必要があり，ほとんどの場合，文中にhoweverが置かれるときは左右両方に前文の対応部分との対比が見られる．この右側に来るものが十分新情報としてまとまっていれば，左側が明確に前文と対比になっていなくてもかまわない．

(4) John is kind. He is, however, greedy.

こうした however の特徴は(3)に見るように文という単位を超えて働くこともあり，また，(2)のように Be aware that のようなBiber et al.(1999)でstanceと分類されるような表現を超えて働く．これはいずれも however が純然とした文文法の枠外にあることも示している．

こうした事実観察を経て，本発表ではさらに規範的な語法書で however の左側に対比焦点が来るとしているのは，それが典型例であるだけであり，音調などさまざまな原因が複合的に関係してくることを示す．同時に，有標な文頭配置は対比焦点を生み出すとよく言われるが，それについても再考すべきであることを指摘する．また，howeverの位置については規範的な語法書で繰り返し触れられるように，確かに最適とは言い難い例も多く，コーパスによる生の用例が万能とは限らないが，多数の用例を検討することから，記述的に見て適切と思われるhoweverの用法を確定したい．

Bell, D.M. (1994) Cancelative Discourse Markers, Ph.D. Dissertation, Boston University.

Biber, D. et al. (1999) Longman Grammar of Spoken and Written English, London: Longman.

Ifantidou, E. (1994) Evidentials and Relevance, Ph.D. Dissertation, University College London.

◆特別講演

●《Corpus Linguistics and Language Teaching 》

（講師　Randi Reppen ）

This presentation will explore recent developments in the field of corpus linguistics and the implications of corpus linguistics for English language instruction. The presentation will focus on two major areas: English language instruction and ESL/EFL teacher training.

As language teachers and teacher trainers we are responsible for providing the best possible instruction for our students. This includes textbook and material selection and the actual instructional implementation of the textbooks and materials that we have selected. How do we make those pedagogical decisions? How do we determine what our learners need to learn? In the past we often relied on our intuitions, past teaching successes, or the advice of experts. However now, with the emergence of corpus linguistics, we can have a solid basis for the pedagogical decisions that we make. Corpus linguistics can provide vital information for language instruction and teacher training by allowing us to look at how language is used in various situations. By knowing what language our students will encounter, we can better prepare our students for the actual language demands that they will face.

In addition to informing pedagogical decisions (what we teach or when we teach particular structures), corpus

linguistics can also play a major role in the area of materials development. Using information from corpora and corpus-based studies we can design and develop materials that are based on real language use rather than artificial or made up examples. Language learners can be involved in working with actual language use and getting the practice needed to be successful in real life situations, thus transforming our classrooms from insulated language experiences into springboards for language learning that equip learners to interact with native speakers and a variety of situations.

●《 Corpus Linguistics and the Study of English Grammar 》

（講師 Douglas Biber）

This talk describes how corpus-based analyses can be employed for the study of English grammar, with a focus on the presenter's experiences working on the Longman Grammar of Spoken and Written English (LGSWE). Two major themes are developed: 1) the new kinds of descriptive information about English grammar made available through corpus-based research, and 2) the range of research methodologies required for such analyses.

The talk is organized around the research methods required for corpus-based analyses of grammar. I will begin with discussion of lexical analyses which might be done using a concordancing package, and then move on to a discussion of the range of automatic and interactive computer programs used for the basic research underlying the LGSWE.

For each research methodology, the talk will provide one or two example analyses, illustrating the new kinds of linguistic patterns that can be discovered using corpus-based techniques. These example analyses will include studies of individual words, grammatical classes, lexico-grammatical associations, and syntactic patterns. A secondary theme developed through these studies will be the unreliability of intuitions as predictors of language use, even for the most common patterns.