

# 英語コーパス学会 第24回大会資料

日時:2004年10月2日(土)午後1時より(正午受付開始)

会場:日本大学文理学部 百周年記念館2階 国際会議場

(<http://www.chs.nihon-u.ac.jp/>)

〒156-8550 東京都世田谷区桜上水 3-25-40

ワークショップ会場:6号館2階(621教室)

## 第 24 回大会プログラム

ワークショップ 10:00 ~ 11:45 6号館 2階(621 教室)

《コーパス言語学のための多変量解析入門》

講師 田畑 智司(大阪大学)

会場 百周年記念館 2階 国際会議場

受付開始 12:00 (百周年記念館 1階)

開 会 13:00

司会 赤野 一郎(京都外国語大学)

会長挨拶

中村 純作(立命館大学)

開催校代表挨拶

島方 洸一(日本大学文理学部長)

その他

研究発表 13:30 ~ 14:40

司会 杉浦 正利(名古屋大学)

N-gram から見る BNC のテキストジャンル

Unclassified に含まれるテキストのジャンル推定に向けて

後藤 一章(大阪大学大学院生)

品詞  $n$ -gram 分布に基づく NS/NNS 論文の分類モデルと日本人英語科学技術論文の特徴抽出

田中 省作(九州大学)

藤井 宏(九州大学大学院生)

富浦 洋一(九州大学)

徳見 道夫(九州大学)

実践報告 14:40 ~ 15:15

司会 野口ジュディー(武庫川女子大学)

英語学習のための大規模コーパス利用

梅咲 敦子(立命館大学)

休憩 15:15 ~ 15:40

シンポジウム 15:40 ~ 18:00

司会 深谷 輝彦(椋山女子学園大学)

《コーパスと言語理論：コーパスは言語理論にいかに関与しているか》

コーパスと生成文法

言語研究の目標・対象・方法論

講師 小野 隆啓(京都外国語大学)

コーパスと認知言語学

英語における姿勢維持動詞の意味と構文の発展について

講師 早瀬 尚子(大阪外国語大学)

コーパスと語用論

語用論はコーパスから何を引き出すことができるか

講師 田中 廣明(関西外国語大学)

閉会の辞

塚本 聡(日本大学)

懇親会 18:15 ~ 20:00

(場所：カフェテリア チェリー 会費：4,000 円)

# 発表要旨

## 【ワークショップ】

コーパス言語学のための多変量解析入門

田畑 智司（大阪大学）

コーパスから得られたデータをもとに言語使用の実際を記述しようとするコーパス言語学は、いわば、「言語の行動科学」であり、そこでは理論や仮説の検証のため、あるいは一見混沌としたデータから探索的に情報を抽出するための手段として統計学が援用されてきた。特に多変量解析は、Burrows (1987)や Biber (1988)以降、探索的データ分析(Exploratory data analysis)のための手法として使用域間の言語変異やテキスト類型論、著者推定や文体模写等の問題で効果的に活用されるようになってきている。

とはいえ、(主に紙面の制約によるものと推察されるが)コーパス言語学関連の雑誌やアンソロジーに所収されている論文では、使用するデータや解析手法に関する説明があまりにも簡略化され過ぎる傾向があり、人文系の研究者や学生にとって、そのような研究論文を読み解くのは容易なことではないと思われる。

そこで本ワークショップでは、コーパス言語学で使用される主な多変量解析法を取り上げ、ごく簡単な数学的背景、多変量解析でアプローチできるデータのタイプ、出力として得られる数値データやグラフの読み方などについて、初心者向けに解説するつもりである。まず前半部では、代表的手法のうち、クラスター分析 (cluster)、主成分分析 (PCA)、対応分析 (CA)を取り上げる。いずれの手法についても、雑誌論文やアンソロジー、モノグラフに現れる具体的な例をもとに、分析の手順や解析結果の読み方について、できるだけ平易な解説を試みたい。後半では、各統計手法を比較検討し、各手法の長所や短所、データ分析を行う際の条件などについて概説した後、多変量解析を行うための統計パッケージをいくつか紹介する。さらに、オープンソースで開発され無料公開されている統計処理言語(環境) R を用いて、時間の許す範囲でハンズオン・セッションも行いたい。

## 【研究発表】

N-gram から見る BNC のテキストジャンル

— unclassified に含まれるテキストのジャンル推定に向けて—

後藤 一章（大阪大学大学院生）

British National Corpus (BNC) は現在公開されている英語コーパスの中で最も大規模なものの1つであり、研究や教育など様々な場面で利用されている。特に、David Lee (2001) は各テキストを詳細なジャンルに分類し、BNC を利用したレジスター研究の重要な基盤を形成したと考えられる。

しかし、Lee の分類は主に言語外的な要因(著者やタイトル名など)を基準として行われたものであり、言語内的な基準からの検討はほとんどなされていない。Nakamura (1995)では言語内的な要因を基にしたテキスト分類の重要性が示唆されており、BNC に関しても言語内的な観点からの分類を行う必要があると思われる。さらに、Lee は話し言葉 (Spoken) におけるいくつかのテキストを unclassified (未分類) のままにしているが、言語内的な基準を利用することで、これらを分類できる可能性がある。

本発表の目的は、Lee によって分類されているテキストを N-gram モデルを利用して言語内的な観点から見直し、さらに unclassified に含まれているテキストのジャンル分類を試みることであり、N-gram モ

デルは著者推定に有効であることが既に示されており (Hoover (2001)), テキストジャンルの判別にも有効である可能性が高い。プログラムを利用して話し言葉に含まれている各テキストの単語と品詞タグの bi-gram 統計を取り, その結果を多変量解析にかけてジャンル間の相対的な距離を測った。発表は, データや処理の手続きを具体的に示しながら進めていく予定である。

### 品詞 $n$ -gram 分布に基づく NS/NNS 論文の分類モデルと 日本人英語科学技術論文の特徴抽出

田中 省作 (九州大学)

藤井 宏 (九州大学大学院生)

富浦 洋一 (九州大学)

徳見 道夫 (九州大学)

英語科学技術論文は, 論文という性格上綴りや文法的な誤りといった稚拙な誤りはほとんど無く, 事前にネイティブ・チェックなどが施されることも少なくないにもかかわらず, 母語話者 (Native Speaker; NS) と非母語話者 (Non-Native Speaker; NNS) の書いたもの間には質的に大きな隔りがある。本研究では, そのような英語科学技術論文における NS/NNS の微妙な差異を明らかにすることを目的に, 論文から得られる品詞  $n$ -gram 分布に着目し, 文書分類モデルに基づいた NS/NNS に特徴的な品詞  $n$ -gram 分布の抽出手法とその結果について報告する。

NS/NNS 論文の文書分類モデルは, 論文から算出される品詞  $n$ -gram 分布を特徴量としたパターン認識問題として定式化する。NS が書いた英語科学技術論文 229 本, NNS として日本人が書いた英語科学技術論文 63 本に対して  $n=3$  とした分類実験では 97.9% という高い精度を得た。さらに, NNS に正しく分類された NNS 論文から, “論文中で頻繁に出現し, なおかつその論文を介して NS/NNS において分布傾向の差が大きい品詞  $n$ -gram 分布” を優先的に抽出した。その結果, “冠詞 形容詞” につづき日本人は直ちに名詞句を構成する傾向が高いこと (NS はさらに修飾を加える), “名詞 be” につづき日本人は NS に比べ前置詞を利用する割合が極めて低いこと, 日本人はほとんど利用しない現在進行形を NS は利用することや, 句動詞や動詞の自他などに関して興味深い相違が認められた。

### 【実践報告】

#### 英語学習のための大規模コーパス利用

梅咲 敦子 (立命館大学)

大規模コーパスの教育現場での利用は, これまでは英語を専門とする学部・大学院における英語研究を目的としたものが中心であった。しかし, さまざまな検索が比較的簡単に行えるようになってきた。本発表では, 大規模コーパスを大学の英語教育一般に利用する効果的な方法の構築をめざして, 小学館コーパスネットワークシステム (SCN) による British National Corpus の検索を, 発表者が大学学部のリーディングとリスニング授業に利用している実践例を報告し, コーパス利用法と検索例について, 学習目的と外国語習得理論の点から考察する。あわせて, 検索例から得られた英語学上の成果についても触れたい。実践は, 情報教室を常時使用できる 3 クラスと, 常時利用できない 1 クラスにおけるものである。発表では (1) リーディングとリスニングの理解を助けるための検索, (2) リーディング, リスニングの授

業における語彙力を高めるためのコーパスの利用, (3) perception から production へ英語力を高めるための検索について例示する。また, その検索例としては, succeed at, raise an eyebrow, effect/influence/impact on, have a reputation for, This paper / This study に後続する動詞他を取り上げる。利用法としては, 学部指定教材に出てくる表現の検索以外に, オンラインで読める論文のアブストラクトを利用したリーディングと, それを応用したライティングについて紹介する。

## 【シンポジウム】

### コーパスと言語理論

コーパスは言語理論にいかに関与するか

司会 深谷 輝彦 (椋山女学園大学)

このシンポジウムでは, 次の3つの問題を共通テーマとして, 各講師が得意とする言語理論の枠組みでコーパスの役割を論じる。

1. コーパス分析は言語理論にいかに関与できるか。
2. 理論言語学者が将来使ってみたいコーパスとはどのようなものか。
3. 言語理論がコーパスから引き出したい情報とは何か。

ここでいう言語理論としては, 生成文法, 認知言語学, 語用論を想定する。まず司会者から, シンポジウムへの導入を行う。次に, 各講師は英語を例に取りながら, それぞれの分野の典型的分析事例を提示し, 次いでその分析事例とコーパスがどのように相互作用を持ちうるのかに関して, 上の3つの論点から議論を深める。コーパスに対するスタンスを明らかにしつつ, さらにこのようなコーパスが欲しい, このような検索をしたい, というコーパスに対する注文も積極的に出す予定である。そうすることで, 聴衆からの多様な反応を期待したい。

### コーパスと生成文法

言語研究の目標・対象・方法論

小野 隆啓 (京都外国語大学)

人間の自然言語には多くの側面があり, 単一の言語理論ですべてが解明できることはあり得ない。言語の研究が理論に裏付けされたものでなければならないことは Chomsky(1957), *Syntactic Structures* の preface において明確に述べられており, いかなる言語研究も何らかの言語理論をその基礎においている。言語理論は, それぞれ言語研究の目標と対象を明確に規定し行われる。生成文法の目標は普遍文法(Universal Grammar: UG)の解明にあり, その研究対象は I-language である。そのため研究発達の当初から an ideal speaker-listener, in a completely homogeneous speech-community (Chomsky 1965: 3)を研究対象とし, 内省(introspection)による研究を続けている。このような生成文法の目標と研究対象に対して, 昨今のコンピュータ技術とソフトウェアの進展を基盤におくコーパスを用いた研究方法が研究者に大きな関心を持たせている。本発表では, 言語研究の目標と研究対象, 生成文法研究におけるコーパスの果たす役割を, 1980年代に非常に精力的に行われた補文内からの wh 移動(wh-movement)などを例にとり, コーパス分析からの貢献, 生成文法研究に望まれるコーパスなどについて考察してみたい。

## コーパスと認知言語学 英語における姿勢維持動詞の意味と構文の発展について

早瀬 尚子 (大阪外国語大学)

認知言語学では、「実際に用いられている言語データが私たちの言語知識を反映し、またさらに形成する源となる」とする、Usage-Based model(用法基盤モデル)が提案されている。しかし実際には、これまで「認知言語学」を標榜する内外の研究の多くが、旧来的なデータ分析、つまり、専ら母語話者の言語直観を重視した作例のみに基づいて成されていたのが現状であった。ここ数年の談話分析や言語習得の研究に導かれて、認知言語学の中でもコーパスを積極的に利用する意義および必要性が認識されるようになってきたが、まだ途上である感は否めない。

このシンポジウムでは、認知言語学における構文研究の一例として、英語における姿勢維持動詞を用いたコピュラの用法の成立に焦点を当て、その構文の成立・発展の動機づけについて、コーパス利用によるサポート、論証の可能性を探ってみたい。

まず、英語における姿勢維持動詞(sit/stand/lie)がコピュラになっていくその変遷の手がかりをコーパスで追う。これらの動詞は共時的に見て付随動作と共起する用例(e.g. sit and V または sit-ing/-ed)が少なくないこと、また「座る」「立つ」「横たわる」といった本来の意味では必ずしも用いられていない例も散見されることなどをコーパス資料から導く。このことから、sit/stand/lie が assertive な意味ではなく、むしろ主語の存在を導入するだけの提示機能(presentational function)を果たしていることを見る。この機能は姿勢維持動詞のコピュラの用法への道を開くものと考えられる。

さらに、この提示機能という条件が満たされる限りにおいて、他の動詞であっても類似の表現形式が可能であること、つまり「構文」として成立する可能性があることを、コーパスにより示す試みをする。

## コーパスと語用論 語用論はコーパスから何を引き出すことができるか

田中 廣明 (関西外国語大学)

通常なら言えない文でもある文脈では言えたり、ある文脈に限ってこちらの意味になるとされる場合がある。前者は「有標性」、後者は「一義化、意味確定」と言われ、語用論的な問題とされる。問題は、われわれはどのようにその文脈(語用論的)情報をその文に読み込み、理解しているのかである。この発表ではコーパスからその情報をいかに引き出せるかに焦点を当てる。試みに、begin a book(本を読み/書き始める)、begin the cake(ケーキを食べ/焼き始める)、begin the dictionary(辞書を作り/?調べ始める)、??begin the rock(岩を?始める)を取り上げる。bookは「(読者が)読むもの」でもあるし、「(著者が)書くもの」でもある。主語が「読者」なら「読み始める」、「著者」なら「書き始める」となる。cakeは「食べるもの」でも「焼くもの」でもあるが、料理のレシピでは「焼き始める」という意味にしかならない。begin the dictionaryは「辞書を調べ始める」にはならないようである。??begin the rockはどのように文脈を与えても不自然である。これらの意味確定に文脈の果たす役割は何であろうか。コーパスからそのkey wordとなるものは見いだせないであろうか。読み込み方に一定の推論過程がないであろうか。

一般的には、意味論は文脈に左右されない意味を、語用論は文脈によって左右される可変的な意味を扱うとされる。その境界線は現在の研究でも不明確なままであるが、文脈情報は可変的な変数としてとらえられる。そのすべてを列挙するのは不可能に近い。では、その変数の取り入れ方はどうなっているであろうか。それをコーパスの典型例から説明することができないであろうか、というのがこの発表の主眼となる。

## 大会参加者へのご案内

- 車でのご来場はできません。
- ワークショップの受付は6号館2階621教室前で午前9時30分から行います。
- 大会の受付は百周年記念館（正門の真向かい）1階で正午から行います。
- 第2体育館にカフェテリア(学生食堂)、売店があります。
- 下高井戸駅周辺には多数飲食店があります。
- 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。
- 会員でない方も、「当日会員」として参加していただけます(1,000円)。

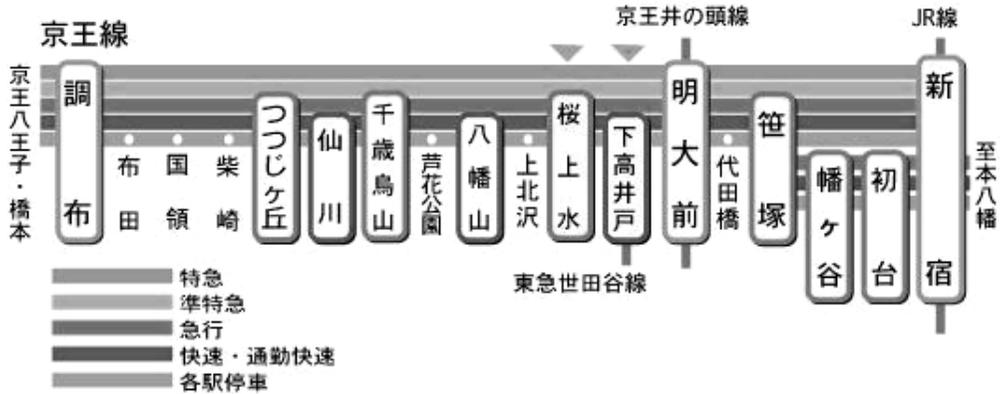
## CAMPUS MAP



- |            |                   |
|------------|-------------------|
| Ⓐ 本館       | Ⓜ 総合体育館           |
| Ⓑ 1号館      | Ⓝ 新図書館(平成17年完成予定) |
| Ⓒ 2号館・図書館  | Ⓝ 研究棟             |
| Ⓓ 3号館      | Ⓔ 地球システム科学作業棟     |
| Ⓔ 4号館      | Ⓟ 地理・物理実験棟        |
| Ⓛ 5号館      | Ⓢ 化学実験棟           |
| Ⓜ 6号館      | Ⓝ 中庭グラウンド         |
| Ⓝ 7号館      | Ⓝ サークル棟           |
| Ⓝ 8号館      | Ⓝ サークルセンター        |
| Ⓝ 百周年記念館   | Ⓝ 受付案内・正門         |
| Ⓝ 第2体育館・食堂 |                   |

## 《会場へのアクセス》

### 交通機関

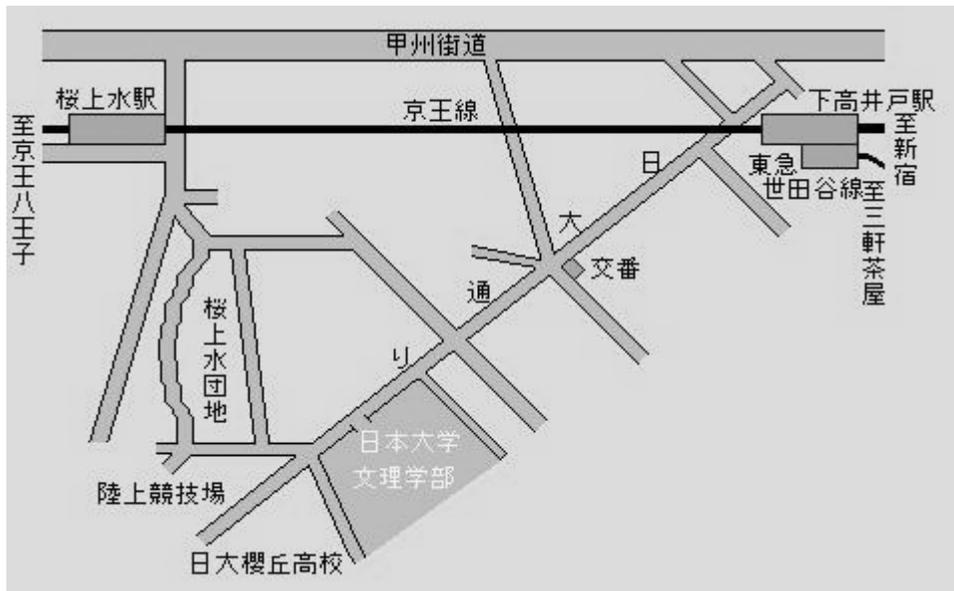


京王線新宿 - 下高井戸間 乗車時間：10分 料金：150円

京王線新宿 - 桜上水間 乗車時間：12分 料金：150円

小田急線豪徳寺 - 世田谷線山下 - 下高井戸間 乗車時間：5分 料金：130円

### 最寄り駅から会場まで



2004年9月 発行  
 編集・発行 英語コーパス学会  
 代表者 中村 純作  
 事務局 〒615-8558 京都市右京区西院笠目町 6  
 京都外国語大学 赤野研究室内  
 TEL: 075-322-6103 FAX: 075-322-6246  
 E-mail: i\_akano@kufs.ac.jp  
 URL: <http://muse.doshisha.ac.jp/jaecs>