

第3回例会プログラム

日 時 1994年4月2日(土)

会 場 大阪大学大学院言語文化研究科棟

開会の辞 13:00 大阪大学 齊藤俊雄

総 会 13:10-13:30

帰朝報告 13:30-14:50 司会 徳島大学 中村純作

1. 「オーストラリア・ニューカースル

における Literary Computing の状況」 有明高専 田畑智司

2. 「Bank of English について」 島根大学 井上永幸

研究発表 15:45-16:00 司会 西南学院大学 久屋孝夫

「タグ付きコーパスの分析について—多変量解析法を用いて」

豊田高専 高橋 薫・中京大学 古橋 聡

講 演 16:00-16:50 司会 大阪大学 玉井俊紀

「AWKによるコーパス処理について」 東京大学 山縣宏光

閉会の辞 16:50 追手門学院大学 丸谷満男

懇 親 会 17:00- 会場 大阪大学言語文化部大会議室

第3回例会発表のレジюме

○田畑 智司「オーストラリア・ニューカースルにおける Literary Computing の状況」

ニューカースル大学(The University of Newcastle, NSW, Australia)で行われているコン

ピュータ援用英語英文学研究についてレポートする。ニューカースル大学では1980年代前半から John Burrows によって、Jane Austen を中心とする英文学作品のテキスト・データベースの構築及びそれに基づいた文体分析が始められ、その後 The Centre for Literary and Linguistic Computing (CLLC) が設立された1989年以来、チームプロジェクトの形でコーパスの規模が拡大され、分析の対象とされている作家は、17世紀から現代にまで及んでいる。Burrows の代表的な研究は、Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method (Clarendon Press, Oxford, 1987)等の著作によっても知られているが、そのほかにも現在 CLLC で進行中の研究の動向について報告したい。

○井上 永幸「Bank of English について」

English Department at the University of Birmingham と Collins Publisher (現在の HarperCollins)との合弁事業である COBUILD Project について、コーパス構築の進行状況やコーパス管理の方法について報告し、Bank of English をネットワーク端末から効果的に検索・出力して利用するための lookup program 及び lookup program による検索結果について統計的情報を与える picture program については、膨大な数の用例におぼれることなく、簡単に連語分析の指針を示してくれる MI-score と t-score を具体例を交えて示し、語法研究への応用の可能性について考察する。

○高橋 薫・古橋 聡

「タグ付きコーパスの分析について—多変量解析法を用いて」

総単語数が百万語から成る LOB Corpus の注目すべき特徴は、15種類のカテゴリー分類があることと、すべての単語に品詞情報を表すタグが付してある点である。

特にタグには139種類の分類があり、このタグについての頻度に関する分析は、カテゴリーごとの特徴を規定するのに有益であると考えられる。たとえば、口語的な文体の特徴と文語的なものとをタグの使用頻度によって予測し、それぞれの文体をモデル化することも可能である。

今回の発表では、LOB Corpus のパソコンでの解析方法をソフト面で簡単に紹介しながら、タグについての数値的な解析方法について触れ、上述のカテゴリーについての特徴を示す。

○山縣 宏光「AWKによるコーパス処理について」

第1回の研究会の席上 **AWK** という名前が出てきたことをご記憶の方もおられると思います。これは1種の簡易言語で、起源的には **UNIX** 上で **ED** を発展させた **SED** と同じく、単なるフィルターとして使われてきましたが、これに様々な機能を加えて、全く別のものと言っていい **AWK** ができました。**AWK** はワークステーションだけでなく、パソコン上でも使えますので、親しみやすく便利です。

コーパスの処理の際、問題が起こるのは、コーパスが計算機の論理によって設計されているためで、たとえば、機械的に行が切られていたり、ハイフンが入れているといったことで、単純に検索したのでは望ましい結果が得られないことがあります。こうした問題にどう対処するかということについてお話したいと思います。