

2 日本語コーパスの使用法と解析

麗澤大学言語研究センターおよび言語科学会(2008年度会員講習会)共催

講師: **李 在鎬 (LEE, Jae-Ho)** (情報通信研究機構)

玉岡 賀津雄 (麗澤大学)

日時: 2008年7月5日(土) 午後1:00~4:30

場所: 麗澤大学生涯教育プラザ1階プラザホール
(千葉県柏市光が丘2-1-1)

李 在鎬

日本語研究のためのコーパス使用法紹介

近年、ウェブの普及と進化により、コーパス基盤の研究環境は大きく変わりつつある。本発表では、ウェブブラウザ上で利用可能な言語資源を中心に、日本語研究のためのコーパスを紹介する。また、形態素解析技術など、自然言語処理の研究成果を組み込んだコーパス管理ツールも紹介する。紹介の際には、具体的な入手方法や利用手順など、利用者の観点から情報を提供し、コーパス日本語学の活性化につなげたい。

玉岡 賀津雄

エントロピーと冗長度の指標を使ったコーパス共起頻度の分析

クロード・シャノン(Claude Shannon)は、『通信の数学理論(A Mathematical Theory of Communication)』(1948)で、「エントロピー(entropy)」と「冗長度(redundancy)」という概念を発表した。エントロピーは情報量の尺度の一つであり、表現の種類と使用頻度から一つの値を算出して、あいまいさや乱雑度の増減を示すことができる。冗長度は表現の多様性と使用頻度から一つの値を算出して、無駄の程度を表すことができる。エントロピーと冗長度の尺度を組み合わせ、コーパスの共起頻度を分析する手法を紹介する。

案内のHP <http://r-linc.org/>

アクセスのHP <http://www.reitaku-u.ac.jp/introduction/2007030914094082.html>

参加費無料・事前予約不要 (どなたでもご参加いただけます)

問合せ先: 玉岡賀津雄(麗澤大学)

Email: ktamaoka@gc4.so-net.ne.jp