

コーパス調査入門

2015 年度前期 水曜 2 限 (10:30-12:00)

担当

浅尾仁彦 (Yoshihiko Asao)

asao@lit.nagoya-u.ac.jp

文学部 1 階 121 号室

オフィスアワー: 月曜日 15~17 時。またはメールでのアポイントメント。

授業の目的

コーパスとは（通常電子的に利用できる形で）実際に使用された言語データを研究目的で集積したものであり、人文学や情報科学の広い分野で重要な研究資源になっています。この授業では、コーパスを扱うために必要な知識を基礎から学び、各自の興味にもとづいて、言語に関する問題を客観的な方法を学びます。

コンピュータの最低限の知識（文書作成ソフトによる文書作成や、ファイルをダウンロードして開く方法等が分かれば十分）と言語への関心を前提条件とします。また、自身でソフトウェアなどを導入できるコンピュータへのアクセスがあることが必要になります（必ずしも授業に持参する必要はありません）。

授業の概要

コーパスを用いた研究の意義やその限界について学ぶとともに、コーパスを実際に扱うための具体的な技術の習得に取り組みます。ただし、特定のソフトウェアの使い方を学ぶのではなく、ソフトウェアが世代交代しても変わることのない基本的な考え方を身につけることを重視します。

また、日本語の変化など身近なテーマについて扱ったコーパス研究の論文を読みます。

ウェブサイト

<http://asaokitan.net/courses/corpus2015/>

教科書

- 石川慎一郎 (2012) 『ベーシックコーパス言語学』

参考書

1. 藤村逸子・滝沢直宏 編 (2011) 『言語研究の技法 データの収集と分析』
2. 大名力 (2012) 『言語研究のための正規表現によるコーパス検索』
3. 浅尾仁彦・李在鎬 (2013) 『言語研究のためのプログラミング入門』

その他、授業中に適宜紹介します。

日程

日程は仮のものです。進捗状況、受講者の関心などをみて変更することがあります。変更はウェブサイトに反映させますので、チェックして下さい。

4/15	イントロダクション	
4/22	コーパスとしてのウェブ	教科書 1・2 章
5/13	テキストデータ	
5/20	コーパスの作成・入手	教科書 3・4 章
5/27	論文を読む (1)	
6/3	正規表現	
6/10	論文を読む (2)	
6/17	形態素解析	教科書 5 章
6/24	論文を読む (3)	
7/1	コロケーション	教科書 6 章
7/8	論文を読む (4)	
7/15	統計の初歩	
7/22	研究発表	
7/29	研究発表	
8/5	(予備)	期末レポート 〆切

成績評価の方法と基準

出席と授業への参加 30%

授業には毎回出席するようにしてください。授業の終わりに簡単な質問に答えたコメントシートなどを提出してもらうことで、出席点に代えます。

宿題 30%

授業内容の理解を問う実践的な宿題を 3 回予定しています。

期末レポート 40%

期末レポートではコーパスを用いた独自の研究をします。具体的なトピックについては自由ですが、5 月頃までにレポートのトピックについて提案してもらい、実現可能性などについて話し合います。学期末にレポートの内容について簡単なプレゼンをお願いします。

その他

- 携帯電話をマナーモードにしてください。
- 授業時間の 1/3 以上教室にいない場合は出席点を認めないことがあります。

論文の例

コーパスを利用した研究は無数にあり、ここに挙げたものは例にすぎません。ここにはない論文でも興味に応じて自由に選ぶことができます（事前に相談が必要です）。

- 音声・音韻
 - 佐々木・吉田 (2002) 複合名詞に対する複合語アクセント規則
 - 前川・五十嵐 (2006) 2 モーラ有核助詞の韻律上の独立性：『日本語話し言葉コーパス』の分析
 - 岡田 (2007) 「とりあえず『日本語話し言葉コーパス』検索」の可能性を検討する：「雰囲気」という単語の発音を例に
 - 荻野 (2013) 外来語の語形のゆれ：チックとティック
 - 荻野 (2013) 外来語の語形のゆれ（2）：バとヴァ
- 文法
 - 田野村 (1994) 丁寧体の述語否定形の選択に関する計量的調査：「～ません」と「～ないです」
 - 田野村 (2001) サ変動詞の活用のゆれについて：電子資料に基づく分析
 - 田野村 (2002) 形容動詞連体形における「な/の」選択の一要因：「有名な」と「無名の」
 - 荻野 (2006) 形容動詞連体形における「な/の」選択について：田野村氏の結果を WWW で調べる
 - 南部 (2007) 定量的分析に基づく「か/の」交替再考
 - 佐野 (2008) 『日本語話し言葉コーパス』に現れる「さ入れ言葉」に関する数量的分析
 - 田野村 (2009) サ変動詞の活用のゆれについて (続) 大規模な電子資料の利用による分析の精密化
 - 佐野 (2011) 現代日本語のヴォイスにおける進行中の言語変化に関する数量的研究：「ら抜き言葉」、「さ入れ言葉」、「れ不足言葉」を例として
- 語彙・語法
 - 杉本 (2009) コーパスからみた類義語動詞：「ねじる」と「ひねる」
 - 新屋 (2010) 類義語「状況」「状態」の統語的分析：コーパスによる数量的比較
- 談話・社会
 - 荻野 (2007) ブログにみる日本語の男女差
 - 藤村 他 (2011) 会話コーパスの構築によるコミュニケーション研究. 『言語研究の技法』
 - 平田 他 (2015) 国会会議録コーパスを用いたオノマトペ使用の地域比較
- 文字・表記
 - 柏野・奥村 (2012) 和語や漢語のカタカナ表記：『現代日本語書き言葉均衡コーパス』の書籍における使用実態
- 学習・教育
 - 坂口・鄭 (2007) ナガラ節に見られる日本語学習者の母語転移：韓国語母語話者が用いる「ながら」の始点的用法を中心に
 - 陳 (2009) 台湾人日本語学習者の「でも」の会話使用に関する縦断研究：使用位置及び共起表現を中心に
 - 陳 (2010) 第二言語としての二種類の複合動詞の習得：コーパスによる学習者の使用実態調査をもとに

- 朴 (2010) OPI データにみる日本語学習者と日本語母語話者による文末表現の使用：接続助詞で終わる言いさし表現を中心に
- 英語
 - Yoneoka (1998) “But they do it!”: Using corpora to research sentence-initial *ands* and *buts*.
 - Kjellmer (2009) Where do we backchannel?: On the use of *mm*, *mhm*, *uh huh* and such like.
 - Yoshimura (2007) A word-based approach to the study of adverbs: *completely*, *entirely*, and *really*.
 - Butler (2008) The very idea! A corpus-based comparison of IDEA, CONCEPT and NOTION and their formal equivalents in Spanish.
 - Tottie (2011) *Uh* and *um* as sociolinguistic markers in British English.