



第56回名古屋大学大学院人文学研究科・日本語教育学分野公開講演会
主催：平成30年度名古屋大学大学院人文学研究科研究プロジェクト経費「研究者・大学院生のための言語テキスト検索およびテキストマイニングのワークショップの開催」



2018年12月15日（土）

14：30～17：30



名古屋大学東山キャンパス
全学教育棟北棟406



参加費無料・事前申し込み不要

講師：石田基広 教授

徳島大学大学院・社会産業理工学研究部

テーマ：Rによる言語・テキスト分析

内容：フリーソフトのRを利用して言語データ（テキスト）を分析する方法を紹介します。最初にRで言語データを解析するための準備等について説明します。次にRで各種コーパスデータを利用する方法や、テキストを単語（形態素）などの単位に分割し、出現頻度表にまとめる方法を解説します。取り上げるのは英語と日本語ですが、他言語の解析も可能です。続けて、頻度表に対する各種操作（品詞別の集計やtf-idfの算出など）や、統計的分析手法について紹介します。時間的余裕があれば、センチメント分析（単語のネガポジ分類）や、単語出現の時系列表示、単語分散表現についても取り上げます。また、テキストデータを可視化する技法（ワードクラウドやネットワークグラフなど）についても紹介します。

【備考】会場でノートパソコンを操作する場合は、下記の機能をインストールしておいてください。座学として聴講する場合は必要ありません。

(1) R (<https://cran.ism.ac.jp/>) および RStudio Desktop free 版

(<https://www.rstudio.com/products/rstudio/download/>) をインストールする。

(2) RStudio を起動して左下の Console に以下のように入力し Enter を押す（機能が追加インストールされる。新規にフォルダを作成するか等の確認ダイアログが現れた場合は Yes を選ぶ。）
`install.packages(c("dplyr", "tidytext", "udpipe", "wordcloud2", "igraph"), dependencies = TRUE);`
`install.packages("RMeCab", repos = "http://rmeCab.jp/R")`

(3) フリーの日本語形態素解析器 MeCab (<http://taku910.github.io/mecab/#download>) をインストールしておく。

- Windows では mecab-0.996.exe をダウンロードしてダブルクリックするだけ。

- Mac では 3 つの手順 (Apple 社が提供するコマンドラインツールの導入、MeCab のインストール、辞書 IPA DIC のインストール) が必要。詳細は

<https://sites.google.com/site/rmeCab/home/install> を参照のこと。

(4) Mac ユーザーのみ XQuartz (<https://www.xquartz.org/>) をインストールしておく。

連絡先：名古屋大学大学院人文学研究科・日本語教育学分野(教授) 玉岡賀津雄 tamaoka@nagoya-u.jp

名古屋大学大学院人文学研究科・英語教育学分野(准教授) 三輪晃司 kojimiwa@nagoya-u.jp

名古屋大学大学院人文学研究科(助教) 張婧禕 jingyizhang@nagoya-u.jp