言語資源活用ワークショップ 2017

Abstract 集

2017 年 9 月 5・6 日 (火・水)　『言語資源活用ワークショップ 2017』
2017 年 9 月 7 日 (月)　『音声資源活用シンポジウム』

大学共同利用機関法人 人間文化研究機構
国立国語研究所 コーパス開発センター 編

アンケートのお願い

ワークショップ終了後以下のアンケートにご協力ください。

- 2017 年 9 月 5・6 日 (火・水)　『言語資源活用ワークショップ 2017』
  `https://goo.gl/45fheZ`

優秀発表賞

本ワークショップでは，全著者が学生である発表に対して，参加者の投票によって決める優秀発表賞を設定いたしました。発表 ID の末尾に “-S” がついている 10 件が投票の対象です。参加者は発表を聞いたうえで利益相反のない発表 1 件に 9 月 5 日中に受付にある投票箱に投票してください。投票用紙は当日受付時に配布いたします。

優秀発表賞に選ばれた発表は 9 月 6 日のクロージングで表彰いたします。

【利益相反 (Conflict Of Interest, COI) の判断基準】

下記 a-d のいずれかに該当すれば，COI ありと判断してください。

a. 審査対象発表の著者または共著者。

b. 2014 年 10 月以降，著者と同じ組織 (研究室や部署など) に属していた時期がある。
(大学や会社が同じでも研究室や部署が異なれば問題ありません。)

c. 2014 年 10 月以降，著者と共著で発表したことがある。

d. 2014 年 10 月以降，科研費等の公的研究費や企業からの委託・共同研究費に基づき著者と共同研究を行ったことがある。

Programme:言語資源活用ワークショップ 2017

2017 年 9 月 5 日 (火)

10:10-10:20　■挨　拶 (2F 講堂)　前川喜久雄

10:15-12:00　■口頭発表　A グループ (2F 講堂)

[O-A-1-S]
日本語特殊形容詞の装定用法の出現傾向について
.....王海涛（京都大学大学院：学生）

[O-A-2-S]
一般的な日本語テキストにおける助詞比率の規則性
.....森秀明（東北大学大学院：学生）

[O-A-3-S]
ウェブコーパス「梵天」による敬語研究–その活用可能性に関する事例的検討–
.....金賢眞（大阪大学大学院：学生）

[O-A-4-S]
非情の受身の「状態」の意味について
.....張莉（上海外国語大学：学生）

12:00-13:00　休憩

13:00-14:15　■ポスター発表　A グループ (2F フロア・多目的室)

[P-A-1-S]
従属節の意味分類基準策定について–鳥バンク基準互換再構築の検討–
.....松本理美（立命館大学大学院：学生）

[P-A-2-S]
現代日本語における公的場面で行われる談話に関する言語資源の現状と開発
.....馮荷菁（九州大学大学院：学生）

[P-A-3-S]
接続詞の直後に読点が打たれる要因について： 一般化線形モデルを用いた分析
.....岩崎拓也（一橋大学大学院：学生）

[**P-A-4**]

修辞ユニット分析における脱文脈化指数の妥当性の検証

. . . . . 浅原正幸（国立国語研究所）・田中弥生（東京大学：学生）

[**P-A-5**]

外の関係の連体修飾節を伴う名詞述語について

. . . . . 今田水穂（文部科学省）

14:15-14:20　**休憩 (ポスター切替)**

14:20-15:35　**■ポスター発表　B グループ (2F フロア・多目的室)**

[**P-B-1-S**]

話し言葉における接続詞の文体的特徴について

. . . . . 柴田好葉（立命館大学大学院：学生）

[**P-B-2-S**]

近代漢語の品詞性に見る多様性の画一化–形容詞用法を中心に–

. . . . . 間淵洋子（明治大学大学院：学生・日本学術振興会）

[**P-B-3-S**]

形容詞「ハヤイ」の意味と表記

. . . . . 髙橋雄太（明治大学大学院：学生）

[**P-B-4**]

nwjc2vec の fine-tuning

新納浩幸（茨城大学）・古宮嘉那子（茨城大学）・佐々木稔（茨城大学）

[**P-B-5**]

「お／ご～される」とその周辺

. . . . . 高橋圭子（フリーランス）・東泉裕子（フリーランス）

[**P-B-6**]

現代日本語書き言葉均衡コーパスの Universal Dependencies

. . . . . 大村舞（国立国語研究所）・浅原正幸（国立国語研究所）

15:35-15:45　休憩

15:45-16:35　**■招待講演 (2F 講堂)**

[I-1]

近代文献のデジタルアーカイブ化と研究・教育での活用–岩波書店『思想』を題材に

. . . . . 美馬秀樹 (東京大学)

16:35-17:25　**■口頭発表　B グループ (2F 講堂)**

[O-B-1]

CRF 素性テンプレートの見直しによるモデルサイズを軽量化した解析用 UniDic – unidic-cwj-2.2.0 と unidic-csj-2.2.0 –

. . . . . 岡照晃（国立国語研究所）

[O-B-2]

学習者コーパス研究における標本数の問題

. . . . . 石川慎一郎（神戸大学）

2017 年 9 月 6 日 (水)

10:20-12:00　**■口頭発表　C グループ (2F 講堂)**

[O-C-1]

鶴岡市共通語化調査データの確率論的再検討

. . . . . 前川喜久雄（国立国語研究所）

[O-C-2]

『日本語日常会話コーパス』のデータ公開方針–法的・倫理的な観点から–

. 小磯花絵（国立国語研究所）・伝康晴（千葉大学・国立国語研究所）

[O-C-3]

ツリーバンク検索への「UNIX 的」アプローチ

. . . . . 窪田悠介（筑波大学）

[O-C-4]

『源氏物語』コンコーダンスとその応用

. . . . . 近藤泰弘（青山学院大学）

12:00-13:00　**休憩**

13:00-14:15　**■ポスター発表　C グループ (2F フロア・多目的室)**

[**P-C-1**]

中国人日本語学習者の「念押し」表現に見る母語の影響–I-JAS のロールプレイにおける依頼表現に基づいて–

..... 迫田久美子（広島大学）・蘇鷹（湖南大学）・張佩霞（湖南大学）

[**P-C-2**]

単義語と共起する多義語に対する分散表現を利用した語義分析

..... 遊佐宣彦（茨城大学大学院：学生）・佐々木稔 (茨城大学)
古宮嘉那子 (茨城大学)・新納浩幸 (茨城大学)

[**P-C-3**]

書き言葉と話し言葉における外来語語末長音のゆれ

..... 小椋秀樹（立命館大学）

[**P-C-4**]

地域社会により順応するための方言教材作成のための方言データベースの開発について

大山浩美（アップルジャパン）・馬場良二（熊本県立大学）・和田礼子（鹿児島大学）
田川恭識（神奈川大学）・嵐洋子（杏林大学）・島本智美（熊本県立大学）
吉里さちこ（熊本大学）・大庭理恵子（熊本県立大学大学院・学生）

[**P-C-5**]

ナレーション分析を用いた川崎市政ニュース映画の映像理解の試み–市民アーカイブズ構築のための枠組みとして–

..... 春木良且（フェリス女学院大学）・田中弥生（東京大学：学生）
田村寛之（一般社団法人カワサキノサキ ）

[**P-C-6**]

『日本語歴史コーパス』における原文 KWIC 表示機能の実装

.....小木曽智信（国立国語研究所）・岡照晃（国立国語研究所）
中村壮範（マンパワーグループ株式会社）・八木豊（株式会社ピコラボ）

[**P-C-7**]

複合格助詞における丁寧形の機能 –『日本語話し言葉コーパス』を用いた分析–

.....本多由美子（一橋大学大学院：学生）

14:15-14:20　**休憩 (ポスター切替)**

14:20-15:35 **■ポスター発表　D グループ (2F フロア・多目的室)**

[P-D-1]

分類語彙表番号を用いた比喩表現収集の試み

. . . . . 加藤祥（国立国語研究所）・浅原正幸（国立国語研究所）

[P-D-2]

レジスター・位相の違いによる会話文の語彙的多様性

. . . . . 山崎誠（国立国語研究所）

[P-D-3]

言語形式に基づく児童作文の類型化

. . . . . 笹島眞実（東京学芸大学大学院：学生）

[P-D-4]

日本語話し言葉コーパスの F0 値再抽出に関する検討

. . . . . 石本祐一（国立国語研究所）・河原英紀（和歌山大学）

[P-D-5]

国会会議録における言語表現の時間的変化の予備的分析

. . . . . 山口昌也（国立国語研究所）

[P-D-6]

モンゴル語アクセント研究のためのデータベース (2)

西川賢哉（国立国語研究所）・玉栄（内モンゴル大・国立国語研究所）
前川喜久雄（国立国語研究所）

15:45-16:35 **■招待講演 (2F 講堂)**

[I-2]

コーパスを用いた談話情報の定量化

. . . . . 折田奈甫 (東北大学)

16:35-17:25 **■クロージング (2F 講堂)**

Programme:音声資源活用シンポジウム

2017 年 9 月 7 日 (木)

13:30–13:40　開会挨拶

13:40–14:15　**感情音声コーパスの構築と利用：いかにして生の感情を対象に研究を行うか**
……有本泰子（帝京大学）

14:15–14:50　**高齢者音声の収録・分析・認識**
……北岡教英（徳島大学）・入部百合絵（愛知県立大学）

14:50–15:25　**諸方言の文末イントネーション–日本語諸方言コーパスから–**
……木部暢子・大槻知世・佐藤久美子（国立国語研究所）

15:25–15:35　休憩

15:35–16:10　**実験室からの脱出：オープンスペースでのマルチモーダルインタラクション収録とその分析**
……坊農真弓（国立情報学研究所）

16:10–16:45　**複数人ユーザとロボットの音声対話**
……船越孝太郎（京都大学／ HRI-JP）

16:45–17:20　**対話システム構築のための音声言語資源**
……東中竜一郎（NTT）

17:20–17:30　閉会挨拶

目次

Abstract 13
日本語特殊形容詞の装定用法の出現傾向について [O-A-1-S]
王海涛（京都大学大学院：学生） 14
一般的な日本語テキストにおける助詞比率の規則性 [O-A-2-S]
森秀明（東北大学大学院：学生） 14
ウェブコーパス「梵天」による敬語研究–その活用可能性に関する事例的検討– [O-A-3-S]
金賢眞（大阪大学大学院：学生） 15
非情の受身の「状態」の意味について [O-A-4-S]
張莉（上海外国語大学：学生） 15
従属節の意味分類基準策定について–鳥バンク基準互換再構築の検討– [P-A-1-S]
松本理美（立命館大学大学院：学生） 16
現代日本語における公的場面で行われる談話に関する言語資源の現状と開発 [P-A-2-S]
馮荷菁（九州大学大学院：学生） 16
接続詞の直後に読点が打たれる要因について： 一般化線形モデルを用いた分析 [P-A-3-S]
岩崎拓也（一橋大学大学院：学生） 17
修辞ユニット分析における脱文脈化指数の妥当性の検証 [P-A-4]
浅原正幸（国立国語研究所）・田中弥生（東京大学：学生） 17
外の関係の連体修飾節を伴う名詞述語について [P-A-5]
今田水穂（文部科学省） 18
話し言葉における接続詞の文体的特徴について [P-B-1-S]
柴田好葉（立命館大学大学院：学生） 18
近代漢語の品詞性に見る多様性の画一化–形容詞用法を中心に– [P-B-2-S]
間淵洋子（明治大学大学院：学生・日本学術振興会） 19
形容詞「ハヤイ」の意味と表記 [P-B-3-S]
髙橋雄太（明治大学大学院：学生） 19
nwjc2vec の fine-tuning [P-B-4]
新納浩幸（茨城大学）・古宮嘉那子（茨城大学）・佐々木稔（茨城大学） 20
「お／ご〜される」とその周辺 [P-B-5]
髙橋圭子（フリーランス）・東泉裕子（フリーランス） 20
現代日本語書き言葉均衡コーパスの Universal Dependencies [P-B-6]
大村舞（国立国語研究所）・浅原正幸（国立国語研究所） 21
近代文献のデジタルアーカイブ化と研究・教育での活用–岩波書店『思想』を題材に [I-1]
美馬秀樹（東京大学） 21

CRF 素性テンプレートの見直しによるモデルサイズを軽量化した解析用 UniDic – unidic-cwj-2.2.0 と unidic-csj-2.2.0 – [O-B-1]
岡照晃（国立国語研究所） 22
学習者コーパス研究における標本数の問題 [O-B-2]
石川慎一郎（神戸大学） 22
鶴岡市共通語化調査データの確率論的再検討 [O-C-1]
前川喜久雄（国立国語研究所） 23
『日本語日常会話コーパス』のデータ公開方針–法的・倫理的な観点から– [O-C-2]
小磯花絵（国立国語研究所）・伝康晴（千葉大学・国立国語研究所） 23
ツリーバンク検索への「UNIX 的」アプローチ [O-C-3]
窪田悠介（筑波大学） 24
『源氏物語』コンコーダンスとその応用 [O-C-4]
近藤泰弘（青山学院大学） 24
中国人日本語学習者の「念押し」表現に見る母語の影響–I-JAS のロールプレイにおける依頼表現に基づいて– [P-C-1]
迫田久美子（広島大学）・蘇鷹（湖南大学）・張佩霞（湖南大学） 25
単義語と共起する多義語に対する分散表現を利用した語義分析 [P-C-2]
遊佐宣彦（茨城大学大学院 : 学生）・佐々木稔 (茨城大学)・古宮嘉那子 (茨城大学)・新納浩幸 (茨城大学) 25
書き言葉と話し言葉における外来語語末長音のゆれ [P-C-3]
小椋秀樹（立命館大学） 26
地域社会により順応するための方言教材作成のための方言データベースの開発について [P-C-4]
大山浩美（アップルジャパン）・馬場良二（熊本県立大学）・和田礼子（鹿児島大学）・田川恭識（神奈川大学）・嵐洋子（杏林大学）・吉里さちこ（熊本大学）・大庭理恵子（熊本県立大学大学院・学生） 26
ナレーション分析を用いた川崎市政ニュース映画の映像理解の試み–市民アーカイブズ構築のための枠組みとして– [P-C-5]
春木良且（フェリス女学院大学）・田中弥生（東京大学 : 学生）・田村寛之（一般社団法人カワサキノサキ ） 27
『日本語歴史コーパス』における原文 KWIC 表示機能の実装 [P-C-6]
小木曽智信（国立国語研究所）・岡照晃（国立国語研究所）・中村壮範（マンパワーグループ株式会社）・八木豊（株式会社ピコラボ） 27
複合格助詞における丁寧形の機能 –『日本語話し言葉コーパス』を用いた分析– [P-C-7]
本多由美子（一橋大学大学院 : 学生） 28
分類語彙表番号を用いた比喩表現収集の試み [P-D-1]
加藤祥（国立国語研究所）・浅原正幸（国立国語研究所） 28

レジスター・位相の違いによる会話文の語彙的多様性 [P-D-2]
山崎誠（国立国語研究所） 29
言語形式に基づく児童作文の類型化 [P-D-3]
笹島眞実（東京学芸大学大学院：学生） 29
日本語話し言葉コーパスの F0 値再抽出に関する検討 [P-D-4]
石本祐一（国立国語研究所）・河原英紀（和歌山大学） 30
国会会議録における言語表現の時間的変化の予備的分析 [P-D-5]
山口昌也（国立国語研究所） 30
モンゴル語アクセント研究のためのデータベース (2) [P-D-6]
西川賢哉（国立国語研究所）・玉栄（内モンゴル大・国立国語研究所）・前川喜久雄（国立国語研究所） 31
コーパスを用いた談話情報の定量化 [I-2]
折田奈甫（東北大学） 31

Information 32
ポスター設営図・出店 33
ランチスペース 34
ランチマップ 35

Abstract

[O-A-1-S]
日本語特殊形容詞の装定用法の出現傾向について

王海涛（京都大学大学院：学生）

本研究は，『現代日本語書き言葉均衡コーパス』を対象にして，日本語の特殊形容詞（「多い，少ない，遠い，近い」）の連体修飾（装定用法）の使用実態と修飾構造ごとの特徴について考察したものである。形容詞は名詞を修飾する際，「連体形＋名詞」の形をとるが，特殊形容詞は，一般に「連用形＋の＋名詞」の形をとる。これまでに特殊形容詞が連用形によって装定用法を形成する理由について研究が行われてきたが，その使用実態を網羅的に調査・分析した研究は見られない。そこで，本研究では，BCCWJ から特殊形容詞の装定用法の用例を抽出し，その用例を修飾構造ごとに分類，それぞれの構造ごとに意味的な特徴の記述を行った。その結果，装定用法の「連用形＋の」と「連体形」の出現傾向には違いがあり，同じ特殊形容詞であっても，助詞のつく修飾構造の分布に偏りが見られた。また，「名詞＋に」が連体形の前にくる装定用法が容認されていることも明らかとなった。

【口頭発表】9/5 (火) 10:20-10:45

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[O-A-2-S]
一般的な日本語テキストにおける助詞比率の規則性

森秀明（東北大学大学院：学生）

日本語のテキストでは名詞比率に連動して動詞や形容詞などの比率が規則的に変化することが知られている。しかし名詞比率と付属語の関係は明らかにされていないため，『現代日本語書き言葉均衡コーパス』(以下 BCCWJ) 固定長・長単位データと『日本語歴史コーパス』(以下 CHJ) の長単位データを使用し，名詞比率と助詞比率の相関を中心に観察した。BCCWJ の中には，例えば商品名と値段が列挙されるなど，ほとんど助詞が使用されないサンプルが存在するため，「名詞比率 45% 未満・その他比率 30% 未満」のサンプルを仮に「一般的な日本語テキスト」と定義して調査した。この結果，連体助詞には名詞比率と正の相関が，接続助詞には負の相関があるなど，様々な相関が認められた。また注目すべきことに助詞の中分類ではこのように名詞比率との相関がありながら，それらを合計した大分類では，多くのテキストの助詞比率は 34% 前後とほぼ一定で，その比率は古典語でも同じであった。

【口頭発表】9/5 (火) 10:45-11:10

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ); 日本語歴史コーパス (CHJ)

[O-A-3-S]
ウェブコーパス「梵天」による敬語研究–その活用可能性に関する事例的検討–

金賢眞（大阪大学大学院：学生）

ウェブコーパスによる敬語研究の可能性について事例的に検討する。具体的には，謙譲表現「ご～する」が尊敬・丁寧表現として誤用される現象に注目し，ウェブコーパス「梵天」がこの問題の検討に有益なデータとなり得るか否かを，現代日本語書き言葉均衡コーパス及び日本語話し言葉コーパスとの比較により評価する。この三種のコーパスについて，「ご～する」の「～」の部分に使用される動名詞の異なり（種類）と，それぞれの誤用率を調査すると，「梵天」では多様な動名詞が収集でき，他のコーパスの結果をほぼ全てカバーすること，また，他のコーパスより誤用率の大きくなる動名詞の多いことが確認できた。このことは，ウェブコーパスが「ご～する」の正用・誤用を含む使用実態の調査に適切な資料であることを示すとともに，とりわけ誤用例の分析を充実させ得るという点で，コーパスを用いた敬語研究を更に発展させるものであることをも示唆している。

【口頭発表】9/5 (火) 11:10-11:35

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ); 日本語話し言葉コーパス (CSJ); 日本語ウェブコーパス「梵天」(NWJC)

[O-A-4-S]
非情の受身の「状態」の意味について

張莉（上海外国語大学：学生）

非情物が主語になる受身（非情の受身と略す）は古代語においては動作・作用が非情物に働きかけてくるという出来事の意味を表すのは稀であり，状態の意味を表すと言われる。「状態」の意味が現代語においてどの程度の割合を占めるのか。現代語の非情の受身の「状態」の意味をどう捉えるべきか。本稿は，BCCWJ を使い，述語動詞の構文形式を手掛かりに考察を行った。現代語の非情の受身の意味は，「状態」に限らず，「出来事」，「反復」など様々である。非情の受身は本質的には非情物主語が動作主の行為によって物理的な影響を受ける意味を表し，「状態」などは「動作主の行為によって物理的な影響を受ける」という意味に「る・ている」などの助動詞が加わって生じた文全体の意味である。通時的に見ると，状態の意味の占める割合の変化も，現代語の非情の受身が古代，近代に比べて多様であるのも，文体や文章の特徴の変化，すなわち言語生活の変化によるといえよう。

【口頭発表】9/5 (火) 11:35-12:00

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-A-1-S]
従属節の意味分類基準策定について–鳥バンク基準互換再構築の検討–

松本理美（立命館大学大学院：学生）

本発表は，「鳥バンク」基準互換 (池原, 2009)（以下「鳥バンク」）の日本語従属節の意味分類基準策定について再検討を行う。「鳥バンク」は日英対訳コーパス中の文型パターンの分析・分類を重ねグループ化された意味分類体系である。日本語–英語翻訳を意識した分類体系を，そのまま日本語分析のための従属節意味分類に転用することには若干の問題がある。『現代日本語書き言葉均衡コーパス』に対する「鳥バンク」基準の従属節アノテーション作業において，二人のアノテーターによる意味分類タグ付け作業の齟齬の分析を通し，日本語分析のための基準としての「鳥バンク」再構築を試みる。具体的には，意味・機能の階層の修正（分類階層に混在している意味と機能の整理），分類項目（日本語学的視点で節末形式や文例の再分類）についての議論を行い，安定した日本語の従属節意味分類基準の策定について検討する。

【ポスター発表】9/5 (火) 13:00-14:15

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-A-2-S]
現代日本語における公的場面で行われる談話に関する言語資源の現状と開発

馮荷菁（九州大学大学院：学生）

過去数年，語彙や日常会話などを中心とするコーパスが盛んになりつつある。だが，公的場面（例えば，ディベート，講演，法廷，国会会議など）で行われる談話を扱う言語資源が単一的であり，なかなかコーパスに入れないという現状となっている。また，深澤・ヒルマン小林 (2011) は「パブリックスピーキング」という用語を提出・定義した上で，「社会性および専門性の高い言語行動を行うためには，単なる『会話』だけでなく，公的な性質を持つ口頭コミュニケーションである『パブリックスピーキング』(Public Speaking) の能力が不可欠である」(p. 30) という課題も提示されている。このことから，公的場面におけるコミュニケーションのあり方を探る価値があると思われる。したがって，本研究では，公的談話をジャンルごとに分類したうえで，言語資源ないしコーパスを構築する試行を検討する。

【ポスター発表】9/5 (火) 13:00-14:15

〔利用する言語資源〕 JDA ディベート大会実録，NHK 文化講演会，国会会議録など

[P-A-3-S]
接続詞の直後に読点が打たれる要因について： 一般化線形モデルを用いた分析

岩崎拓也（一橋大学大学院：学生）

本稿では，接続詞の直後に読点を打つ場合の要因を探るためにコーパスを用いた計量的な調査を試みた。コーパスは『現代日本語書き言葉均衡コーパス』のコアデータを使用し，検索には『中納言』を用いた。分析は，一般化線形モデルを用いたロジスティック回帰分析から，接続詞の直後に読点が打たれる要因について検討した。その結果，接続詞の直後に打たれる読点は，接続詞の連接類型だけでなく，接続詞の前後の文字種や品詞，接続詞の語種，一文の長さや接続詞自体の文字数といった変数が影響を与えていることが確認された。その中でも，一文が長いほど接続詞の直後に読点が打たれやすいこと，接続詞の文字数が多いほど読点が打たれにくいこと，接続詞の連接類型が対比型と同列型の時に読点が打たれにくいことが明らかになった。

【ポスター発表】9/5 (火) 13:00-14:15

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-A-4]
修辞ユニット分析における脱文脈化指数の妥当性の検証

浅原正幸（国立国語研究所）・田中弥生（東京大学：学生）

本稿では修辞ユニット分析における脱文脈化指数の妥当性について検証する。修辞ユニット分析は，節に相当するメッセージが命題か提言か（発話機能）・空間上どこにあるか（中核要素）・時間軸上どこにあるか（現象定位）を認定して，その組み合わせに基づき「修辞機能」と「脱文脈化指数」を同定するものである。順序尺度である脱文脈化指数は経験的に決められたものではなく恣意的なものであった。この脱文脈化指数が一般にどの程度共有されるものなのかを確認するために，Yahoo! クラウドソーシングを用いた一対比較による被験者実験を行った。一対比較のデータを Bradley Terry 法に基づく一般化線形モデル，線形順序に変換し，脱文脈化指数の妥当性を検証した。結果，おおよそ脱文脈化指数の順序は共有される一方，中核要素や現象定位に対する日本語話者の見方を必ずしも反映していないということがわかった。

【ポスター発表】9/5 (火) 13:00-14:15

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-A-5]
外の関係の連体修飾節を伴う名詞述語について

今田水穂（文部科学省）

『現代日本語書き言葉均衡コーパス』述語項構造データを使って，人魚構文の抽出と，人魚構文になりやすい名詞の評価を行った。まず人魚構文を成立させる構造的要件を検討し，人魚構文を判別するための条件を設定した。次にコーパス中の名詞を語彙素ごとに集計し，対数尤度比統計量によって外の関係の連体修飾節の取りやすさと人魚構文へのなりやすさを評価した。また，nwjc2vec のベクトル表現を用いてそれらの名詞をクラスタリングし，類似語ごとに配列して可視化した。最後に個々の事例について調べ，抽出に失敗する事例の構造的特徴や，人魚構文になりやすい名詞の修飾節の特徴を検討した。

【ポスター発表】9/5 (火) 13:00-14:15

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ);nwjc2vec

[P-B-1-S]
話し言葉における接続詞の文体的特徴について

柴田好葉（立命館大学大学院：学生）

接続詞の文体的特徴に関する研究においては，書き言葉を対象とした調査はいくつかみられるが，話し言葉を対象とした大規模な調査は行われていない。そこで『日本語話し言葉コーパス』(CSJ) を用いて話し言葉における接続詞の文体的特徴を調査した。CSJ に出現した頻度 100 以上の接続詞の内「で」を除く 43 語を対象とし，石黒 (2008) を参考に意味用法の分類を行った。さらに講演種・自発性・発話スタイルの 3 項目それぞれを 5 段階に分け各接続詞に点数を付与し，その合計点数 (15 点満点) を元に層別した。その結果，合計点数が 13 点以上，5 点以下，その中間という 3 つの層に接続詞を分けることができた。層ごとに接続詞の意味用法をみると 13 点以上の層では偏りがなく多様な意味用法が確認できるのに対し，5 点以下の層では順接や逆接の接続詞に偏っていることなどが明らかになった。以上の結果を具体的な用例を含め考察し，話し言葉における接続詞の文体的特徴を明らかにする。

【ポスター発表】9/5 (火) 14:20-15:35

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[P-B-2-S]
近代漢語の品詞性に見る多様性の画一化–形容詞用法を中心に–

間淵洋子（明治大学大学院：学生・日本学術振興会）

本研究は, 現代日本語書き言葉の模索〜確立の時代にあたる近代に焦点を当て, 日本社会の近代化に伴い飛躍的に語彙を増やした漢語について, 語法（品詞性）の通時的変化の一側面を明らかにするものである。特に,「必要な」「偉大なる」「堂々たる」といった助動詞を伴い体言を修飾する形容詞用法を持つ二字漢語に着目し, 近代語コーパスと現代語コーパスの用例分析に基づいて, 使用実態と変化を確認した。その結果として, 近代と現代とで語法の変化が見られる語が少なくないこと, 変化の類型には, 形容詞用法を失うもの（「熟練」「親善」「優勝」など）, や減少するもの（「幼稚」「漠然」「優良」など）, 新たに形容詞用法を獲得するもの（「絶対」「任意」「均等」など）, 形容詞用法が増加するもの（「多用」「高度」「莫大」など）の 4 類型があること, 変化の背景には, 語義の拡張・変化や, 用法の固定化が関わるほか, 模索的で多様性を有していた状態から, 安定・定着に伴い画一化していく方向性が見出だせることを示す。

【ポスター発表】9/5 (火) 14:20-15:35

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ); 日本語歴史コーパス (CHJ)

[P-B-3-S]
形容詞「ハヤイ」の意味と表記

髙橋雄太（明治大学大学院：学生）

本研究では, 近代において複数表記語であるハヤイの意味を辞書の区分と近代雑誌コーパスの用例を基に分析した。その結果,[動作][作用][基準][期間] の 4 つの意味を立てられた。近代雑誌コーパスにおいては [基準] の頻度が高く,[期間][動作][作用] は低かったが, 大正期に [動作] の用例が増加し, 意味変化が起こることがわかった。次に, 近代雑誌コーパスを用い,「早」「速」「疾」の表記と,[動作] と [作用] を統合した [速度],[基準][期間] の 3 つの意味との対応関係を分析した。その結果,「早」は頻度が高く「疾」と「速」は低かったが,1909 年から [速度] の意味を「速」で表記する用例が増え,[基準] と [期間] の意味は「早」,[速度] の意味は「速」で書き分けられていくことがわかった。最後に, ハヤイの周辺語を調査し,[速度] の意味を「速」で表記するようになることには, 運動力学などにおける「速度」などの漢語が背景にあることを明らかにした。

【ポスター発表】9/5 (火) 14:20-15:35

〔利用する言語資源〕 日本語歴史コーパス (CHJ)

[P-B-4]
nwjc2vec の fine-tuning

新納浩幸（茨城大学）・古宮嘉那子（茨城大学）・佐々木稔（茨城大学）

nwjc2vec は国語研日本語ウェブコーパス (NWJC) から構築された分散表現データである。NWJC が超巨大コーパスであるため，nwjc2vec の品質はかなり高いと考えられる。ただし分散表現データを実際の自然言語処理システムに利用する際には，そのシステムが対象とする領域に依存した分散表現データの方が望ましい。これは領域適応の問題である。ここでは処理対象を新聞記事として，新聞記事 7 年分から構築した分散表現データと nwjc2vec を比較することでこの点を確認する。またこの問題の対処として nwjc2vec に対して少量の新聞記事を利用して fine-tuning を行い，その効果を確認する。

【ポスター発表】9/5 (火) 14:20-15:35

〔利用する言語資源〕 nwjc2vec

[P-B-5]
「お／ご～される」とその周辺

髙橋圭子（フリーランス）・東泉裕子（フリーランス）

「インターネットから お 申込み される と便利です」「ご 結婚 され て，何年になりますか」など，近年,「お／ご～される」という形式が, 尊敬表現として多用されている。しかし, この形式は,『敬語の指針』(2007) などでは不適切, 誤用とされている。ただし, これまで実際の用例に基づく分析・考察は管見ではほとんど行われてきていない。「お／ご～される」は 19 世紀末に成立したとされている。本発表では,『日本語歴史コーパス』『現代日本語書き言葉均衡コーパス』などから近現代語の用例を収集し量的・質的に分析する。また,「お／ご～なさる」「お／ご～する」「お／ご～できる」といった関連する形式についても分析し,「お／ご～される」およびその周辺の, 現代敬語における位置づけを考察する。

【ポスター発表】9/5 (火) 14:20-15:35

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ); 日本語歴史コーパス (CHJ)

[P-B-6]
現代日本語書き言葉均衡コーパスの Universal Dependencies

大村舞（国立国語研究所）・浅原正幸（国立国語研究所）

自然言語処理の分野では多言語かつ言語横断的な言語研究が盛んに取り組まれている。その言語横断的な言語研究の取り組みとして Universal Dependencies(UD) がある。UD では品詞や係り受け構造の標準・スキーマを定め，多言語のコーパスを提供している。本論文では，日本語コーパスである現代日本語書き言葉均衡コーパス (BCCWJ) を UD のスキーマへと変換したコーパスについて紹介をする。BCCWJ では日本語における文節単位の係り受け情報がすでに付与されている。この係り受け構造を基にして UD へと変換するプログラムの開発を行った。しかし，文節単位は UD の単語単位には沿っていない。そのため，BCCWJ で提供されている短単位と長単位というふたつの言語単位を単語の単位をして認定したコーパスを構築する。短単位と長単位について UD のスキーマに当てはめた場合，どのような係り受け構造ができるのかを示す。

【ポスター発表】9/5 (火) 14:20-15:35

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[I-1]
近代文献のデジタルアーカイブ化と研究・教育での活用–岩波書店『思想』を題材に

美馬秀樹（東京大学）

日本には，世界的に見ても稀なほど高度な人文学や芸術，文化創造の成果が集積されており，我が国の近代を通じた知と文化創造の蓄積は，21 世紀の人類の貴重な財産である。これら学術的，公共的知識資源をデジタル化し，蓄積，保存すると共に，アクセシビリティの向上等，高度に利活用可能な環境を迅速に構築する必要がある。特にテキスト資源に関しては，諸外国においても歴史的文献のデジタル化に関する取り組みが進んでおり，アルファベットのみならず，漢字やひらがなといった 2 バイト文字文化圏での技術開発に，日本だけではなく，海外での日本研究やアジア研究等，世界中の研究者より大きな期待が寄せられている。本講演では，1921 年創刊の岩波書店『思想』90 年分，約 9 千論文（16 万ページ）を題材とし，我々の研究室でこれまで取り組んできたデジタルアーカイブ化に関し，その手順と問題点を報告する。特に，OCR による文字認識精度の向上，デジタルアーカイブの（完全）自動構築に向けた研究開発について詳述すると共に，構築したアーカイブ構造化・可視化システムの，研究，及び学部・大学院教育での活用に関して，今後の方向性を含めて議論する。

【招待講演】9/5 (火) 15:45-16:35

〔利用する言語資源〕 岩波書店『思想』

[O-B-1]
CRF 素性テンプレートの見直しによるモデルサイズを軽量化した解析用 UniDic – unidic-cwj-2.2.0 と unidic-csj-2.2.0 –

岡照晃（国立国語研究所）

国立国語研究所で構築している短単位自動解析用辞書『UniDic』は，現在，形態素解析器 MeCab 専用の解析用辞書として使用・公開を行なっている．しかし解析用 UniDic で使用している CRF 素性テンプレートは，MeCab 用の他の辞書（『IPA 辞書』,『Juman 辞書』）に比べ，行数が多く，学習後のモデルファイルのサイズ（素性数）も他より大きなものとなっている．そこで今回，現代語用の解析用『UniDic』の CRF 素性テンプレートの再検討を行ない，より少メモリかつ，これまでとほぼ同等の性能の解析用辞書を実現した．

【口頭発表】9/5 (火) 16:35-17:00

〔利用する言語資源〕 UniDic; 現代日本語書き言葉均衡コーパス (BCCWJ); 日本語話し言葉コーパス (CSJ)

[O-B-2]
学習者コーパス研究における標本数の問題

石川慎一郎（神戸大学）

大量のデータを集めやすい母語話者コーパスと異なり，学習者コーパスでは，集められる標本数に物理的な制約がある。ここで問題となるのは，調査対象とする言語項目ごとに，どの程度の標本数を集めればある程度安定した結果が得られるかということである。内外の主要な学習者コーパスは，学習者の国（母語）ごとにモジュール構成を取っているが，1 モジュール当たりの標本数は，作文コーパスの場合，ICLE（英語・作文）で 243-982，ICNALE（英語・作文）で 200-800，日本語学習者作文コーパス（日本語・作文）で 144-160，また，発話コーパスの場合，LINDSEI（英語・インタビュー発話）で 50-53，ICNALE（英語・独話）で 200-600，I-JAS（日本語・インタビュー発話等）で 50 となっており，コーパスごとに大きな差がある。本論では，I-JAS の母語話者および学習者データを用い，分析するサンプル数を変化させた場合の基本的言語指標値の変化を概観し，その収束のポイントを検討する。

【口頭発表】9/5 (火) 17:00-17:25

〔利用する言語資源〕 多言語母語の日本語学習者横断コーパス (I-JAS)

[O-C-1]
鶴岡市共通語化調査データの確率論的再検討

前川喜久雄（国立国語研究所）

鶴岡市の共通語化調査（第 1〜3 回調査の音韻項目）では，時間の経過とともに共通語化が進行し，共通語化得点の平均値はいわゆる S 字カーブを描くことが報告されてきている。この発表ではデータの分散に注目して鶴岡データを再検討する。鶴岡調査音韻項目の共通語化得点は N=36 回のベルヌーイ試行からなる二項分布とみなすのが妥当であり，二項分布においては成功の確率が p のとき分散は Np(1-p) となる。データから計算される年代ごとの平均共通語化率を p とみなして理論的な分散値を計算し，データに観察される分散と比較すると，第 1 回調査では理論値の 7〜10 倍，第 2 回では 2〜6 倍，第 3 回では 1 〜7 倍の過分散が生じている。これは，共通語化には著しい個体差がともなうこと，そのため，共通語化過程の数理的理解には個体差を取り込んだモデリングが必須であることを示唆している。個体差を生み出す要因として，音韻項目のクラス，個々の音韻項目，被験者の個人差を用いた階層ベイズモデルを構築して比較した結果を報告する。

【口頭発表】9/6 (水) 10:20-10:45

〔利用する言語資源〕 鶴岡調査データベース

[O-C-2]
『日本語日常会話コーパス』のデータ公開方針–法的・倫理的な観点から–

小磯花絵（国立国語研究所）・伝康晴（千葉大学・国立国語研究所）

現在構築中の『日本語日常会話コーパス』には，自宅での家族との会話やレストランでの友人との会話，屋外の散策時の会話，職場・学校での同僚や学友との用談・会議など，実に多様な場面の会話が含まれる。こうした日常生活の中で記録された会話を，映像データを含めて公開するにあたり，肖像権や個人情報保護などの観点から，また著作物の写り込みについては著作権の観点から，検討する必要がある。本発表では，これまでに収録した大量の会話データをもとに具体的な問題を洗い出し，知財関連を専門とする弁護士との相談なども踏まえた上で整理したデータ公開方針について報告する。

【口頭発表】9/6 (水) 10:45-11:10

〔利用する言語資源〕 日常会話コーパス (CEJC)

[O-C-3]
ツリーバンク検索への「UNIX 的」アプローチ

窪田悠介（筑波大学）

国語研 NPCMJ コーパスは，(ゼロ代名詞や関係節空所などを含む) きめ細かな統語構造を付与したツリーバンクとして日本初のものであり，特に統語論や意味論など，今までコーパス利用があまりなされてこなかった分野でのコーパス活用を活性化させることが期待できる。一方で，木構造を検索し，そこから必要な情報を取り出す作業の (一見したところの) 複雑さのため，言語研究への活用は未だ模索段階を出ていない。本発表では，UNIX 系 OS での基本スキルである単純なコマンドを数珠つなぎにしてデータを加工する手法と，ツリー検索・加工に特化されたスクリプト言語の合わせ技によって，NPCMJ を用いて実際の言語研究に役立つ情報抽出が可能になることを示す。「(ガ/ノ交替の) ノ格でマークされた主語と共起する述語の頻度表を作る」というタスクを例に，コーパスからの情報抽出の具体的な手順を説明する。

【口頭発表】9/6 (水) 11:10-11:35

〔利用する言語資源〕 統語・意味解析情報付き現代日本語コーパス (NPCMJ)

[O-C-4]
『源氏物語』コンコーダンスとその応用

近藤泰弘（青山学院大学）

日本語歴史コーパスは，中納言のインターフェースで公開されているが，XML 形式の本体の全体を解析することで，より多くの情報を引き出すことができる。今回は，『源氏物語』の全文を対象に，短単位および長単位および N-gram による分割を施し，それぞれをコンコーダンス形式に出力した。また，それらを，既存のコンコーダンスである北山谿太編『源氏物語辞典』（平凡社・と比較対照することによって，コーパスを利用したコンコーダンスのメリット・デメリットを検証した。また，総語彙やセンテンスを各種の機械学習モデルによって分析し，『源氏物語』における語彙とトピックの分布の関係ついての知見を得ることができた。

【口頭発表】9/6 (水) 11:35-12:00

〔利用する言語資源〕 日本語歴史コーパス (CHJ)

[P-C-1]
中国人日本語学習者の「念押し」表現に見る母語の影響–I-JAS のロールプレイにおける依頼表現に基づいて–

迫田久美子（広島大学）・蘇鷹（湖南大学）・張佩霞（湖南大学）

「念押し」表現とは，「今，週三日働いていますが，二日にしたいです。いいですか？」「昼食を食べるなら和食が好きです。よろしいですか？」の「いいですか？」「よろしいですか？」のように，話し手の意向や要望などのまとまりのある文の後に，聞き手に是非を確認する疑問文形式の表現を指す。本研究は，多言語母語の日本語学習者横断コーパス（I-JAS）のロールプレイに見られる「念押し」表現が中国語話者に多く見られたことから，母語の影響を考え，その可能性を探ることを目的とする。日本語学習経験のほとんどない中国語母語話者同士 12 組および日本語母語話者 15 組を対象として，同じロールプレイを実施し，「念押し」表現の出現を比較した。その結果，日本語母語話者には出現しなかったのに対し，中国語母語話者同士では，12 組中 8 組（67 %）の割合で「念押し」表現が使用されており，中国語の母語の影響の可能性が高いことが明らかになった。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 多言語母語の日本語学習者横断コーパス (I-JAS)

[P-C-2]
単義語と共起する多義語に対する分散表現を利用した語義分析

遊佐宣彦（茨城大学大学院：学生）・佐々木稔 (茨城大学)
古宮嘉那子 (茨城大学)・新納浩幸 (茨城大学)

多義語の語義曖昧性解消を行う際，周辺の共起単語を特徴として利用する。しかし，周辺の単語には多義語が多く含まれるため，正しく語義識別ができない問題がある。そこで，本研究では単義語と共起する多義語の語義分布について分析を行う。現代日本語書き言葉均衡コーパスを用いて，頻度の高い単義語について共起する多義語を分析し，単義語と共起しやすい多義語は特定の語義を持ちやすいのか調査する。さらに，分散表現を利用して多義語とのベクトルがどのような特徴を持つのか分析する。単義語は複数のベクトルに分かれず，ひとつのベクトルで表現されることと仮定し，単義語と頻繁に共起する多義語がどのような語義の特徴を持つのか調査する。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-C-3]
書き言葉と話し言葉における外来語語末長音のゆれ

小椋秀樹（立命館大学）

外来語表記のゆれには発音のゆれが関わっているといわれるが，表記のゆれと発音のゆれとの間にどの程度関連があるのかについては，十分に明らかにされていない。そこで，『現代日本語書き言葉均衡コーパス』（出版・書籍）と『日本語話し言葉コーパス』とを用いて外来語の語末長母音の表記と発音のゆれの実態調査を行った。調査の結果，長音符号を省略した表記の割合（符号無表記率）は 17.0%，短音で発音した割合（短音率）は 7.7% で，表記と発音との間にずれが見られた。この表記と発音とのずれの要因としては，(1)《エアー》《ソファー》《ボディー》等の特定の語において符号無表記や短音化が高い度数（比率）で生じていること，(2) 語末音「ティ」を持つ語において符号無表記が広範囲かつ高い度数（比率）で生じていることの 2 点が指摘できる。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ); 日本語話し言葉コーパス

[P-C-4]
地域社会により順応するための方言教材作成のための方言データベースの開発について

大山浩美（アップルジャパン）・馬場良二（熊本県立大学）
和田礼子（鹿児島大学）・田川恭識（神奈川大学）・嵐洋子（杏林大学）
吉里さちこ（熊本大学）・大庭理恵子（熊本県立大学大学院・学生）

言語には様々な異種が存在する。方言（ここでは特に地域方言）はその一つである。同じ日本語であっても様々な方言があり，別の言語であるかと思うほど理解し合えない時もある。本研究グループでは，「熊本方言を話せなくてもいいから理解できる」ということを目指し，留学生を対象とした熊本方言の特徴を学ぶ教科書『話してみらんね　さしより！熊本弁』（「さしより」とは共通語では「とりあえず」という意味の熊本方言である）を作成した。さらに中上級者用の教材としてデータベース型教材を同時に開発している。そのデータベースを作成する際に，熊本方言話者の会話データを収集し，デジタル化し，方言タグを付与し，方言データベースを作成した。そのデータベースでは，収集した会話データにおいて方言要素を抽出し，教材作成者（熊本方言話者）が内省を施し使用頻度が高いと思われる方言を選んだ。熊本方言といっても各地域によって大きく異なるため，熊本市で使われる方言に限定し，方言の世代差についても，大学生の使用を念頭において選定した。本稿では，熊本方言話者の会話データにおいてどのような方言要素を抽出し，方言タグを付与し，データベース化したのかについて述べる。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 熊本方言データベース

[P-C-5]
ナレーション分析を用いた川崎市政ニュース映画の映像理解の試み–市民アーカイブズ構築のための枠組みとして–

春木良且（フェリス女学院大学）・田中弥生（東京大学：学生）
田村寛之（一般社団法人カワサキノサキ）

報告者は，神奈川ニュース映画協会が作成し公開していた神奈川県ニュース映画のうち，川崎市に関するものについて，昭和 27 年から平成 19 年までの全 717 本を，川崎市市民文化局市民文化振興室，川崎市市民ミュージアムの協力のもとでアーカイブズ化するプロジェクトを実施している。現在，戦後の高度成長期にフォーカスを当て，川崎市を中心とした都市化，工業化の姿を明らかにするために，昭和 40 年までの高度成長期前期までの分析を行っている。映像そのものは，画質が低いのに加え，被写対象である地域の変化が激しいため，それだけでは分析，理解が困難である。そこで，そこに挿入されている，ナレーションやテロップなどの言語情報が重要な分析資料となっている。本発表では特に，そのナレーションに着目し，マルチメディアコーパスの構築とそれらを用いた映像解析に関して，現況を報告する。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 川崎市民アーカイブズ

[P-C-6]
『日本語歴史コーパス』における原文 KWIC 表示機能の実装

小木曽智信（国立国語研究所）・岡照晃（国立国語研究所）
中村壮範（マンパワーグループ株式会社）・八木豊（株式会社ピコラボ）

日本語史研究の基礎資料は，残された文献に見られる用例である。用例の原文は今日一般に用いられる表記とは大幅に異なる形である場合が少なくない。例えば，『万葉集』は万葉仮名で，キリシタン資料は当時のポルトガル語のローマ字で表記されている。こうした資料をコーパスとして形態論情報を付与し，現代人に読みやすいものとするためには，原文を校訂して漢字平仮名交じりにした読み下し本文を用意する必要がある。一方で，読み下し本文では失われてしまう情報も少なくないため，用例には原文を併せて表示することが求められる。『日本語歴史コーパス』では従来，原文情報を保持しつつ必要な修正を行った上で形態論情報を付与して公開してきたが，原文情報の提供は限定的だった。今回新たに，コーパス検索アプリケーション「中納言」上で，原文の前後文脈付きで検索結果を表示できる機能を実装した。本発表ではこの原文 KWIC 表示機能について述べる。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 日本語歴史コーパス (CHJ)

[P-C-7]
複合格助詞における丁寧形の機能 –『日本語話し言葉コーパス』を用いた分析–

本多由美子（一橋大学大学院：学生）

本研究では「～につきまして」や「～に関しまして」などの複合格助詞における丁寧形の機能について考察する。複合格助詞の丁寧形は講演など改まった場面での話し言葉に現れる表現であるが，一般的には丁寧さが加わった表現とされ，使用傾向についての分析はあまり行われていないようである。本研究では『日本語話し言葉コーパス』の講演のデータを用い，複合格助詞 21 形式について丁寧形と普通形の比較を通して分析を行った。その結果，複合格助詞の丁寧形は表現によって用いられる程度が異なり，また丁寧形が用いられる場合，「～につきましては」のように助詞の「は」や「も」がついた形式が普通形よりも多く用いられること，さらに普通形よりも丁寧形の方がより遠くに係る傾向があることが明らかになった。これらの結果から，複合格助詞の丁寧形は，丁寧さを加えるのみならず，普通形と比べてより広い機能を有することが示唆される。

【ポスター発表】9/6 (水) 13:00-14:15

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[P-D-1]
分類語彙表番号を用いた比喩表現収集の試み

加藤祥（国立国語研究所）・浅原正幸（国立国語研究所）

比喩表現類型に意味情報を付与し，意味的な要素の結合を比喩表現の指標とした用例収集を試みた。まず，日本語比喩表現の類型を『分類語彙表増補改訂版』に基づく意味分類で整理した。具体的には，国立国語研究所報告 57『比喩表現の理論と分類』に挙げられた結合類型（5,537 種類）に含まれる要素の自立語すべてに人手で分類語彙表番号を付与し，比喩表現類型を意味分類において整理したデータを作成した。次に，意味分類の結合を比喩表現の指標として，分類語彙表番号を付与した「現代日本語書き言葉均衡コーパス」のコアデータの一部（約 19 万短単位）を調査した。本発表では，本手法によって取得できた BCCWJ の比喩表現例を示すとともに，比喩表現類型の意味分類結合を用いた比喩表現収集可能性について考察する。

【ポスター発表】9/6 (水) 14:20-15:35

〔利用する言語資源〕 分類語彙表; 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-D-2]
レジスター・位相の違いによる会話文の語彙的多様性

山崎誠（国立国語研究所）

本発表は，以下の 5 つのコーパスを用いて，日本語の会話文の多様性をレジスターや位相（話者の性別，年代）の観点から語彙的に分析するものである。使用したコーパスは，『日本語話し言葉コーパス』(CSJ) の学会講演及び摸擬講演，『日常会話コーパス』(CEJC・構築途中のもの)，『名大会話コーパス』『女性の言葉・男性の言葉（職場編）』，『現代日本語書き言葉均衡コーパス』中の小説会話文である。分析の単位はいずれも短単位である。分析の方法は，品詞構成比，上位語，対数尤度比による特徴語の比較である。特徴語はコーパス間の比較に加えて，性別と年代による比較も行った。品詞構成比では，名詞，副詞は，CSJ 学会講演と日常会話・名大とが対照的な分布を示し，また，終助詞，感動詞-フィラーは，CSJ 学会講演・CSJ 摸擬講演と日常会話・名大とが対照的な分布を示すことが分かった。特徴語では，コーパス間で感動詞（一般，フィラー），終助詞，人称代名詞の分布に違いが見られた。また，これらの語の使用において，性差の違いのほうが年齢層の違いよりも特徴的な語数が多いことが観察された。

【ポスター発表】9/6 (水) 14:20-15:35

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ); 日常会話コーパス (CEJC); 名大会話コーパス; 日本語話し言葉コーパス (CSJ); 女性の言葉・男性の言葉（職場編）

[P-D-3]
言語形式に基づく児童作文の類型化

笹島眞実（東京学芸大学大学院：学生）

本発表は，児童作文の文章を資料とし，言語形式に基づく類型化を目的に，分類指標と手続を提案する。先行研究 (小林,1953・土部・宝示,1963) により，分類指標を接続詞の有無時間の経過を示す語の有無文数の 3 観点とし，類型を「散叙型」(「バラバラ文」)・「連叙型」(「ダラダラ文」)・「話題型」の 3 類型とした。指標に基づき小学 3〜5 年生の計 140 名の作文を対象に分類調査を実施したところ，3 年生では「散叙型」が約 3 割，「連叙型」が約 4.5 割と分布が二分したのに対し，4・5 年生では「連叙型」が全体の約 6〜7 割を占めるとの結果が得られた。これらより，中学年期に「散叙型」から「連叙型」へと，類型の変化が生じると考えられた。ただし，より熟達した文章を考慮すると，「散叙型」及び「連叙型」の下位分類も含めた 5 ないし 6 類型とすることが妥当であり，類型化の指標に文章の一貫性や ttr を加えた改良が考えられ，今後の課題としたい。

【ポスター発表】9/6 (水) 14:20-15:35

〔利用する言語資源〕 児童作文

[P-D-4]
日本語話し言葉コーパスの F0 値再抽出に関する検討

石本祐一（国立国語研究所）・河原英紀（和歌山大学）

本稿では，日本語話し言葉コーパス (CSJ) の音声から高精度な基本周波数 (F0) 抽出法による F0 推定を行なった結果について報告する。CSJ に現在付与されている F0 値は，波形振幅の正規化相互相関に基づいた手法により推定されたものである。波形振幅の相関に基づく手法は波形の周期性を利用するため，周期性が乱れた箇所では本来の F0 とはかけ離れた値を誤検出したり，非周期区間とみなされて F0 の抽出が全くできないことがある。実際に CSJ に付与された F0 情報を確認すると，明らかな抽出誤りの箇所が見受けられる。F0 は CSJ に付与されている韻律ラベリングの基となっているため，F0 抽出精度は韻律ラベリングの精度にも関わる。そこで最新の高精度な F0 推定法によって F0 値を求め，現在付与されている値との違いを調べるとともに，韻律ラベルへの影響について述べる。

【ポスター発表】9/6 (水) 14:20-15:35

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[P-D-5]
国会会議録における言語表現の時間的変化の予備的分析

山口昌也（国立国語研究所）

本発表では，国会会議録 (衆議院・予算委員会，1947-2012) を対象に言語表現の時間的変化の予備的な分析を行う。本研究の最終的な目標は，時間的変化のモデルを構築することであるが，今回はモデルの構築に先立ち，モデルの雛形を構築する。まず，第 1 回〜48 回 (1947/1965)，および，第 145 回〜182 回 (1999/2012) の 2 区間における文字 5gram の比較に基づいて，頻度変化の大きな表現を抽出したのちに，全期間 (1947〜2012) の経年変化を計測した。さらに，会議録作成時に行われる整文の影響を考慮した上で，抽出した表現の特徴を分析し，(a) 会議特有の表現が減少傾向にあること，(b) 改まった場を想定しない表現が増加傾向にあることを示した。さらに，この分析結果に基づいて，国会の議論における発話状況を考慮した時間的変化モデルの雛形を提案した。

【ポスター発表】9/6 (水) 14:20-15:35

〔利用する言語資源〕 国会会議録

[P-D-6]
モンゴル語アクセント研究のためのデータベース (2)

西川賢哉（国立国語研究所）・玉栄（内モンゴル大・国立国語研究所）
前川喜久雄（国立国語研究所）

モンゴル語アクセントの音声学的特徴を把握するために筆者らが設計と実装を進めている単語読み上げ音声データベースについて報告する。CSJ-RDB を参考に，「語」「音節」「音素」「分節音」という，階層関係が認められる 4 つの単位を設定したうえで，以下のテーブルから構成される RDB（リレーショナルデータベース）を構築した：(i) 上記の単位ごとに発話中の要素を記述したテーブル［単位テーブル］，(ii) 単位間の対応関係を記述したテーブル［関係テーブル］，(iii) 音響情報（F0，インテンシティー）を記述したテーブル［音響情報テーブル］，(iv) メタ的情報（話者情報など）を記述したテーブル［メタ情報テーブル］。これらのテーブルを相互に関連付けることにより，「母音の持続長を，それが所属する音節の構造／音節の位置ごとに比較する」といった，複数の単位にまたがる分析を容易に行なうことができる。

【ポスター発表】9/6 (水) 14:20-15:35

〔利用する言語資源〕 モンゴル語

[I-2]
コーパスを用いた談話情報の定量化

折田奈甫（東北大学）

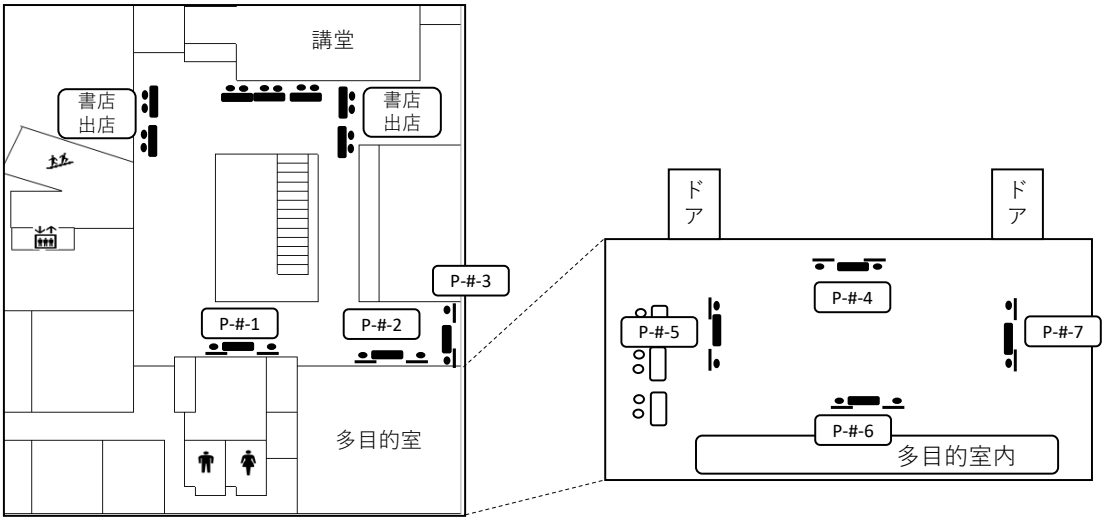
新情報/旧情報，顕著性，話題性，特定性などの談話情報は，言語の理解・産出・獲得において様々な役割を果たす。本講演では，このような談話情報を自然言語処理の技術や資源を活用して定量化し，言語獲得や言語産出のモデリングに用いた研究を紹介する。代名詞の統語的カテゴリ獲得をモデル化した研究では，大人の被験者がコンテキストから代名詞の先行詞をどれくらい推測できるかを実験し，その結果をシミュレーションに用いた。談話における指示対象の話題性を定量化した研究では，トピックモデルにより話者の指示表現選択を捉えられることを示した。また，話者の指示表現選択を確率モデルにより形式化した研究では，談話における指示対象の顕著性，新情報/旧情報，特定性を確率として表し，ニュースコーパスにおける話者の指示表現選択を予測した。これら一連の研究は，コーパスを用いて談話情報を定量化することが，言語獲得や産出の問題に対し明確な説明を与えるのに役立つことを示している。最後に，現在作成中の絵本述語項構造アノテーションコーパスについて紹介したい。

【招待講演】9/6 (水) 15:45-16:35

〔利用する言語資源〕 CHILDES;Ontonotes; 現代日本語書き言葉均衡コーパス (BCCWJ)

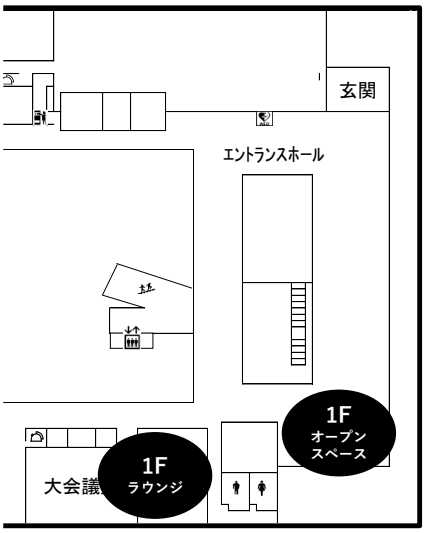
Information

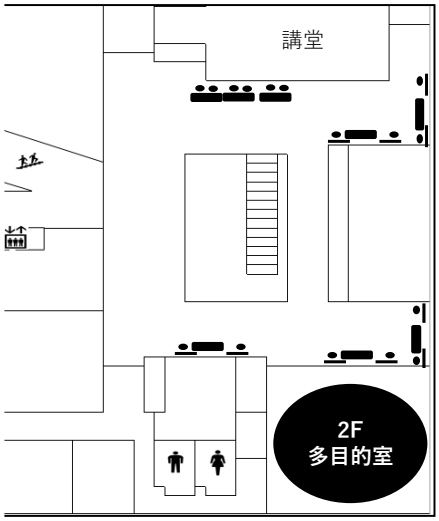
ポスター設営図・出店

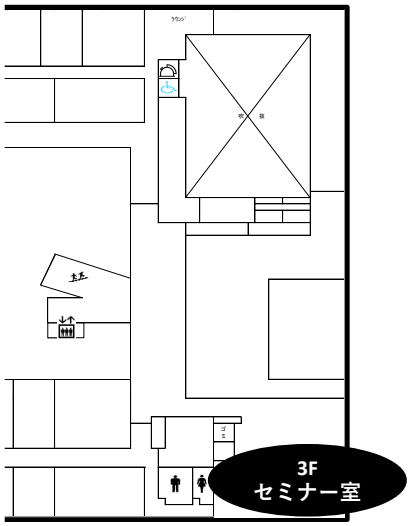


- P-#-5, P-#-6 は電源が利用できます。
- 無線 LAN は eduroam をご利用ください。
- ポスター掲示用ついたての貼り付け可能面積は横 120cm× 縦 150cm です。
- ポスター掲示用資材は国語研で準備いたします。
- 書店各社の出店がございます。

ランチスペース



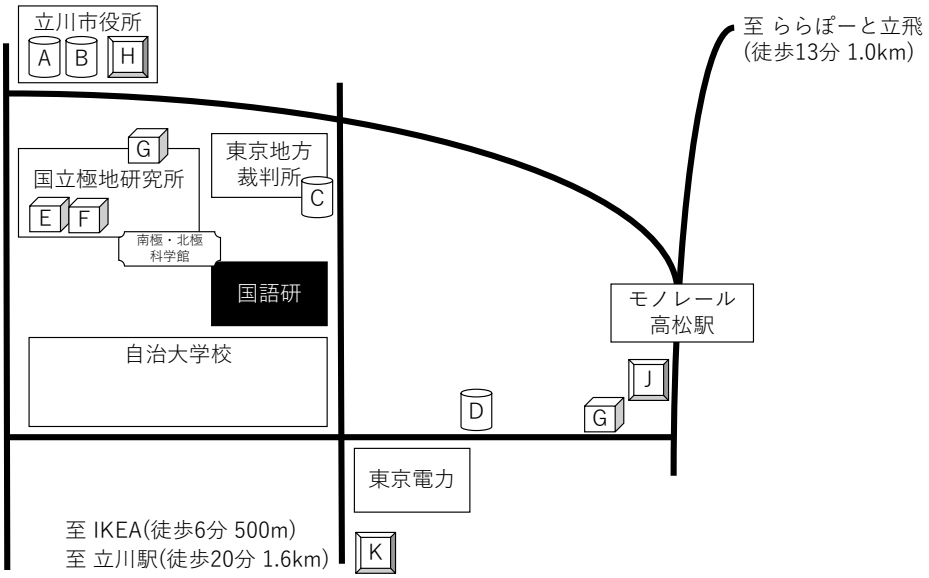




- **2F** 講堂内では飲食しないでください。
- **ゴミの分別にご協力をお願いいたします。**
  ゴミ箱に弁当ガラを捨てる際にはワリバシを別のごみ箱に分別して捨ててください。
- ランチスペースは 1F オープンスペース・1F ラウンジ・2F 多目的室・3F セミナー室の 3 カ所です。
- 2F 多目的室はポスター発表会場を兼ねており，椅子が片付けられておりますが，適宜椅子・机を移動して食事していただいて結構です。移動した椅子は元通りに戻しておいてください。

ランチマップ

国語研近隣は，食事できる場所が限られております。



- お食事処
  - A 立川市役所食堂 (市役所内 3F) 11:00-15:00
    ラーメン・スパゲッティ 300 円程度，日替定食 500 円程度
  - B Café ハーモニー (市役所内 1F) 10:00-16:00
    トースト 300 円程度，ピラフ 600 円程度
  - C 東京地方裁判所食堂 (裁判所内 B1) 8:00-20:00
  - D 中華料理 瑞京 11:00-14:30, 17:00-23:00
- お弁当
  - E チコマート お弁当屋 (極地研内 1F) 11:00-13:00
    各種お弁当 300 円-500 円
  - F ハイジ カレー弁当 (極地研内 1F) 11:00-13:00
    各種 500 円
  - G 瑞京 中華弁当 (ワゴン 2 か所) 11:00-13:00
    日替わり 2 種 500 円
- コンビニエンスストア
  - H ポプラ (市役所内 1F)
  - J ミニストップ (モノレール高松駅南)
  - K セブンイレブン (東京電力南)

言語資源活用ワークショップ 2017 Abstract 集

発行日：平成 29 年 8 月 29 日

発行者：国立国語研究所コーパス開発センター

連絡先：〒190-8561 東京都立川市緑町 10-2　　lrw@ninjal.ac.jp