

言語資源活用ワークショップ 2018

Abstract 集

2018年9月4・5日(火・水) 『言語資源活用ワークショップ2018』
2018年9月6日(木) 『コーパスとしてのウェブテキスト活用シンポジウム』

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。



<https://goo.gl/XX6oWS>

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しております。発表IDの末尾に“-S”がついている14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4日中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5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. 2015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。
(大学や会社と同じでも研究室や部署が異なれば問題ありません。)
- c. 2015年10月以降、著者と共著で発表したことがある。
- d. 2015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 2018

2018年9月4日(火)

- 10:00-10:10 挨拶 (2F 講堂) 前川喜久雄
10:10-12:00 口頭発表 グループ 1 (2F 講堂)
- [O-1-01-S]
『現代日本語書き言葉均衡コーパス』のロシア語翻訳データの構築
..... 宮内拓也 (東京外国語大学/日本学術振興会: 学生)
..... Prokhorova Maria (東京外国語大学: 学生)
- [O-1-02-S]
中古語における形容詞テ形をめぐって
ー形容詞の意味分類との関わりからー
..... 菊池そのみ (筑波大学: 学生)
- [O-1-03-S]
日本語文における連用修飾語成分に見られるパラレルについて
のー考察
ー「赤く変わる」と「赤に変わる」は同じかー
..... 王棟 (東京外国語大学: 学生)
- [O-1-04-S]
連体助詞の「ノ」と文体の関係
..... 森秀明 (東北大学: 学生)
- 12:00-13:00 休憩
13:00-15:35 『包括的検索系』デモンストレーション (2F フロア)
13:00-14:15 ポスター発表 グループ 1 (2F フロア・多目的室)
- [P-1-01-E]
日本語日常会話コーパスに対する談話行為アノテーションの試
み: タグ選択が困難な事例に焦点を当てて
..... 居關友里子 (国立国語研究所)
..... 門田圭祐 (早稲田大学: 学生)
..... 伝康晴 (千葉大学/国立国語研究所)
- [P-1-02-E]
『了解』は使わないように」「了解です」
..... 高橋圭子 (フリーランス)・東泉裕子 (フリーランス)
..... 佐藤万里 (フリーランス)
- [P-1-03-E-S]
NWJC における敬語使用とレジスターとの関係
..... 金賢真 (大阪大学: 学生)
- [P-1-04-E-S]
学校お便り文書の高頻出語彙の縦断的研究
ー 4年生から6年生までの名詞・サ変名詞・動詞の分析
..... 今村桜子 (横浜国立大学: 学生)
- [P-1-05-S]
児童・生徒作文の日本語修辞ユニット分析と教員評価の検討
..... 田中弥生 (東京大学/国立国語研究所: 学生)
- [P-1-06-S]
日本語における慣用句の逸脱使用がもつ言語機能
ー形容詞の反義語への置き換えを手がかりにー
..... 鈴木あすみ (東北大学: 学生)
- [P-1-07]

日本語歴史コーパスの現代語辞書における未知語義判定システム

..... 田邊 紘 (茨城大学：学生)・古宮嘉那子 (茨城大学)
..... 浅原正幸 (国立国語研究所)・佐々木 稔 (茨城大学)
..... 新納浩幸 (茨城大学)

[P-1-08]

形態素解析器『Sudachi』のための大規模辞書開発

..... 坂本美保 (株式会社ワークスアプリケーションズ)
..... 川原典子 (株式会社ワークスアプリケーションズ)
..... 久本空海 (株式会社ワークスアプリケーションズ)
..... 高岡一馬 (株式会社ワークスアプリケーションズ)
..... 内田佳孝 (株式会社ワークスアプリケーションズ)

[P-1-09]

英語における前置詞句についての音響分析

..... 于曉陽 (九州大学：学生)・中島祥好 (九州大学)
..... 張一新 (九州大学：学生)・上田和夫 (九州大学)
..... 岸田拓也 (九州大学)

[P-1-10]

副詞の程度性の下位分類の試み

—「あまり・そんなに・それほど・たいして」を例に—

..... 劉時珍 (専門学校非常勤)

[P-1-11-E]

『日本語日常会話コーパス』構築における Praat の利用

..... 西川賢哉 (国立国語研究所)

[P-1-12]

多様な研究分野に利用可能な超高精細・高精度手話言語データベースの開発

..... 長嶋祐二 (工学院大学)・原大介 (豊田工業大学)
..... 堀内靖雄 (千葉大学)・酒向慎司 (名古屋工業大学)
..... 渡辺桂子 (工学院大学)・菊澤律子 (国立民族学博物館)
..... 加藤直人 (NHK 放送技術研究所)
..... 市川 薫 (千葉大学/工学院大学)

[P-1-13]

英語における頭子音連結の多変量解析

..... 張一新 (九州大学：学生)・中島祥好 (九州大学)
..... 于曉陽 (九州大学：学生)・上田和夫 (九州大学)
..... 岸田拓也 (九州大学)

14:15-14:20

休憩 (ポスター切替)

14:20-15:35

ポスター発表 グループ 2 (2F フロア・多目的室)

[P-2-01-E]

UD Japanese-BCCWJ の構築と分析

..... 大村舞 (国立国語研究所)・浅原正幸 (国立国語研究所)

[P-2-02-E]

LINE データベースの設計と属性付与の現状について

..... 宮寄由美 (国立国語研究所)

[P-2-03-E]

『日本語歴史コーパス (CHJ)』の教育利用の実践報告

..... 宮城信 (富山大学)・江口遼至 (金沢高等学校)

[P-2-04-E]

双方向 LSTM による分類語彙表番号を語義とした all-words WSD

..... 新納浩幸 (茨城大学)・鈴木類 (茨城大学)・古宮嘉那子 (茨城大学)

- [P-2-05-S]
『キングコーパス』の構築と活用
.....高橋雄太 (明治大学/日本学術振興会：学生)
- [P-2-06-S]
『明六雑誌』『東洋学芸雑誌』の特徴語から見る明治前期書き言葉の語彙特性
.....近藤明日子 (東京大学/国立国語研究所：学生)
- [P-2-07-S]
「飲み倒す」とはどういう意味なのか
—Google 検索を利用した日本語の低頻度複合動詞の分析—
.....SEO MINCHEOL(立命館大学：学生)
- [P-2-08]
先天性全盲ろう児の音声言語訓練長期記録の分析状況及び保存活動
.....菊池英明 (早稲田大学)・市川薫 (千葉大学)
.....岡本明 (筑波技術大学)・長嶋祐二 (工学院大学)
.....藤本浩志 (早稲田大学)・引田秋生 (元山梨県立山梨盲学校)
- [P-2-09]
『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性
.....馬場俊臣 (北海道教育大学)
- [P-2-10]
脚本テキストに基づくコーパス文体論の可能性
—テレビドラマ脚本に注目して—
.....松下晶子 (専修大学大学院文学研究科：学生)
.....丸山岳彦 (専修大学/国立国語研究所)
- [P-2-11-E]
『UniDic』を活用した語構造情報付与の試み
—『日本語歴史コーパス』に出現する語を対象に—
.....村山実和子 (国立国語研究所)
- [P-2-12]
Twitter で使われる「深い」の意味
—「強い」「すごい」と比較して—
加藤恵梨 (大手前大学)・山下紗苗 (明石工業高等専門学校：学生)
.....上泰 (明石工業高等専門学校)
- [P-2-13]
日本語の二重目的語構文の基本語順について
.....浅原正幸 (国立国語研究所)・南部智史 (モナシユ大学)
.....佐野真一郎 (慶應義塾大学)
- 休憩
- 15:35-15:45
15:45-17:00
口頭発表 グループ 2 (2F 講堂)
- [O-2-01-S]
比喩指標としての「感じる」
—文法形式と比喩の関係—
.....菊地礼 (中央大学：学生)
- [O-2-02-S]
日本語 wikipedia を用いた慣用句の構成性の数値化
.....岡田優也 (関西学院大学：学生)
- [O-2-03-S]
「XX (と)」、「XX な」、「XX しい」、の構造・文法機能
—畳語による生産性について—
.....陳祥 (筑波大学：学生)

2018年9月5日(水)

- 10:00-11:00 口頭発表 グループ3 (2F 講堂)
[O-3-01]
ニュースを対象にした手話マルチメディアコーパスの構築
加藤直人 (NHK 放送技術研究所)・内田翼 (NHK 放送技術研究所)
..... 東真希子 (NHK 放送技術研究所)
..... 梅田修一 (NHK 放送技術研究所)
[O-3-02]
ベイズモデルによる方言音声共通語化過程の分析
..... 前川喜久雄 (国立国語研究所)
- 11:00-12:00 招待講演 (2F 講堂)
[I-1-01]
言「考」不一致の言語学: コーパスはどこまで意識に迫れるか
..... 吉川正人 (慶應義塾大学)
- 12:00-13:00 休憩
- 13:00-15:35 『包括的検索系』デモンストレーション (2F フロア)
- 13:00-14:15 ポスター発表 グループ3 (2F フロア・多目的室)
[P-3-01-E]
「日常会話コーパス」活用環境の構築
..... 山口昌也 (国立国語研究所)
[P-3-02-E]
「よい子」って誰?
ー政策ニュース映画のナレーション表現に関する研究の一環としてー
.... 春木良且 (フェリス女学院大学)・田中弥生 (東京大学: 学生)
[P-3-03-E]
敬語接頭辞異形「お〜」「ご〜」両者の用例のある語について
..... 服部匡 (同志社女子大学)
[P-3-04-E]
撥音 (の解析) は機械 (UniDic) にとっても簡単ではなかったんだ!
ー BCCWJ を中心にー
..... 劉志偉 (埼玉大学)
[P-3-05]
『現代日本語書き言葉均衡コーパス』書籍サンプルに対する NDC 記号拡張アノテーションと NDC 形式区分を用いた「随筆」の文体分析
..... 加藤祥 (国立国語研究所)・櫻井芽衣子 (日本工業大学)
..... 森山奈々美 (津田塾大学: 学生)・浅原正幸 (国立国語研究所)
[P-3-06]
『現代日本語書き言葉均衡コーパス』に対する名詞述語文アノテーション
..... 今田水穂 (文部科学省)
[P-3-07]
クラウドソーシング発注文書におけるレジビリティの量的分析
..... 岩崎拓也 (国立国語研究所/一橋大学: 学生)
..... 井上雄太 (一橋大学: 学生)
[P-3-08]
話し言葉における代名詞「あれ」の用法の分布
..... 山崎誠 (国立国語研究所)
[P-3-09]

現職教員による児童・生徒作文の評価基準の分析

..... 宮城信 (富山大学) ・浅原正幸 (国立国語研究所)
..... 今田水穂 (文部科学省)

[P-3-10]

スペイン語における前置詞句の数・定性 7 前置詞のクラスタリ
ング

..... 喜多田敏嵩 (東京外国語大学：学生)

[P-3-11-E]

マルチアクティビティに伴う発話の分類：遂行発話と雑談

..... 天谷晴香 (国立国語研究所)

[P-3-12]

コーパスに基づく字順転倒漢語の網羅的把握の試み

..... 間淵洋子 (国立国語研究所)

[P-3-13]

実践医療用語の語構成要素抽出の試み

..... 内山清子 (湘南工科大学) ・岡照晃 (国立国語研究所)

..... 東条佳奈 (目白大学) ・小野正子 (西南女学院大学)

..... 山崎誠 (国立国語研究所) ・相良かおる (西南女学院大学)

休憩 (ポスター切替)

ポスター発表 グループ 4 (2F フロア・多目的室)

[P-4-01-E]

日本語の非流ちょう性

ーとぎれと延伸の数量調査からー

..... 佐々木藍子 (国立国語研究所／東京学芸大学：学生)

..... 砂川有里子 (筑波大学名誉教授) ・浅原正幸 (国立国語研究所)

[P-4-02-E]

『日本語日常会話コーパス』モニター公開版の概要

..... 小磯花絵 (国立国語研究所) ・天谷晴香 (国立国語研究所)

..... 居關友里子 (国立国語研究所) ・白田泰如 (国立国語研究所)

柏野和佳子 (国立国語研究所) ・川端良子 (国立国語研究所：学生)

.. 田中弥生 (国立国語研究所：学生) ・西川賢哉 (国立国語研究所)

..... 伝康晴 (千葉大学／国立国語研究所)

[P-4-03-E]

日本語学習者属性別の言語行為の対話自動生成への適用に関する
一考察

..... 太田博三 (放送大学：学生)

[P-4-04-E]

『現日研・職場談話コーパス』中納言版公開データの作成

..... 柏野和佳子 (国立国語研究所) ・大村舞 (国立国語研究所)

..... 西川賢哉 (国立国語研究所) ・小磯花絵 (国立国語研究所)

[P-4-05]

日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic

..... 首藤公昭 (福岡大学名誉教授) ・田辺利文 (福岡大学)

..... 高橋雅仁 (久留米工業大学)

[P-4-06]

語彙多様性指標の可視化と単回帰分析による TTR の補正

..... 今田水穂 (文部科学省)

[P-4-07]

二字漢語を構成する漢字の造語力の変化

ー『現代雑誌九十種の用語用字』データと『現代日本語書き言葉
均衡コーパス』の比較を通してー

..... 本多由美子 (一橋大学：学生)

14:15-14:20
14:20-15:35

	[P-4-08]	方言音声に対するテキスト自動アライメントの試み 石本祐一 (国立国語研究所)
	[P-4-09]	単語の分散表現を用いた領域における出現単語の特徴分析 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-4-10]	形容詞感動文における曖昧性回避の条件 西内沙恵 (国立国語研究所)
	[P-4-11]	コーパスを用いた授業における成果と課題 ーノンネイティブ日本語教師研修における授業を例にー 清水まさ子 (国際交流基金日本語国際センター) 木田真理 (国際交流基金日本語国際センター)
	[P-4-12]	『日本語話し言葉コーパス (CSJ)』模擬講演における節頭フイ ラーの特徴 渡辺美知子 (国立国語研究所)・是松優作 (東京大学：学生)
	[P-4-13]	『国語研日本語ウェブコーパス』からの新規語彙素獲得の試み 岡照晃 (国立国語研究所)
15:45-16:35		口頭発表 グループ 4 (2F 講堂)
	[O-4-01]	アクセント音調の諸相とその動態形式 佐藤大和 (東京外国語大学)
	[O-4-02]	日本語複単語表現レキシコン JMWEL の概要 ー動詞性表現を中心にー 首藤公昭 (福岡大学名誉教授)・田辺利文 (福岡大学) 高橋雅仁 (久留米工業大学)
16:35-17:00		クロージング (2F 講堂)

Programme: コーパスとしてのウェブテキスト活用シンポジウム (SUWTC)

2018年9月6日(木)

- 13:00-13:15 趣旨説明
.....岡 照晃 (国立国語研究所)
- [セッション 1] 日本語研究に大規模ウェブテキストデータを扱うためには? (2F 講堂)
- 13:15-13:55 【初級編】ウェブの検索結果を利用する
.....荻野 綱男 (日本大学)
- 13:55-14:35 中の人が国語研日本語ウェブコーパス (NWJC) 使ってみた
- 【中級編】ウェブコーパスを”使って”みる-
.....岡 照晃 (国立国語研究所)
- 14:35-15:15 ウェブコーパスの表と裏
.....林部 祐太 (Megagon Labs)
- 15:15-15:30 休憩
- [セッション 2] 企業は大規模ウェブテキストデータをどのように活用しているか? (2F 講堂)
- 15:30-16:10 利便性のあるコーパス構築へのテキストマイニング取り組み
-ビジネス分析に役立つ解析手法開発-
.....三澤 賢祐 (Insight Tech)
- 16:10-16:50 **Wikipedia** を使った進んだ自然言語処理
.....山田 育矢 (Studio Ousia)

発表タイトルは当日変更になる場合がございます。

また発表時間は目安であり、必ずしもこのスケジュール通りに進行するわけではありません。

ご注意ください。

Contents

Abstract	14
『現代日本語書き言葉均衡コーパス』のロシア語翻訳データの構築 宮内拓也(東京外国語大学/日本学術振興会：学生)	[O-1-01-S]
Prokhorova Maria(東京外国語大学：学生)	15
中古語における形容詞テ形をめぐって 一形容詞の意味分類との関わりから一	[O-1-02-S]
菊池そのみ(筑波大学：学生)	15
日本語文における連用修飾語成分に見られるパラレルについての一考察 一「赤く変わる」と「赤に変わる」は同じか一	[O-1-03-S]
王棟(東京外国語大学：学生)	16
連体助詞の「ノ」と文体の関係	[O-1-04-S]
森秀明(東北大学：学生)	16
日本語日常会話コーパスに対する談話行為アノテーションの試み：タグ選択が困難な事例に焦点を当てて	[P-1-01-E]
居關友里子(国立国語研究所)・門田圭祐(早稲田大学：学生)	
伝康晴(千葉大学/国立国語研究所)	17
『「了解」は使わないように」「了解です!」	[P-1-02-E]
高橋圭子(フリーランス)・東泉裕子(フリーランス)・佐藤万里(フリーランス)	17
NWJCにおける敬語使用とレジスターとの関係	[P-1-03-E-S]
金賢眞(大阪大学：学生)	18
学校お便り文書の高頻出語彙の縦断的研究 一4年生から6年生までの名詞・サ変名詞・動詞の分析	[P-1-04-E-S]
今村桜子(横浜国立大学：学生)	18
児童・生徒作文の日本語修辞ユニット分析と教員評価の検討	[P-1-05-S]
田中弥生(東京大学/国立国語研究所：学生)	19
日本語における慣用句の逸脱使用がもつ言語機能 一形容詞の反義語への置き換えを手がかりに一	[P-1-06-S]
鈴木あすみ(東北大学：学生)	19
日本語歴史コーパスの現代語辞書における未知語義判定システム	[P-1-07]
田邊絢(茨城大学：学生)・古宮嘉那子(茨城大学)・浅原正幸(国立国語研究所)	
佐々木稔(茨城大学)・新納浩幸(茨城大学)	20
形態素解析器『Sudachi』のための大規模辞書開発	[P-1-08]
坂本美保(株式会社ワークスアプリケーションズ)	
川原典子(株式会社ワークスアプリケーションズ)	
久本空海(株式会社ワークスアプリケーションズ)	
高岡一馬(株式会社ワークスアプリケーションズ)	
内田佳孝(株式会社ワークスアプリケーションズ)	20
英語における前置詞句についての音響分析	[P-1-09]
于曉陽(九州大学：学生)・中島祥好(九州大学)・張一新(九州大学：学生)	
岸田拓也(九州大学)・上田和夫(九州大学)	21
副詞の程度性の下位分類の試み 一「あまり・そんなに・それほど・たいして」を例に一	[P-1-10]
劉時珍(専門学校非常勤)	21
『日本語日常会話コーパス』構築における Praat の利用	[P-1-11-E]
西川賢哉(国立国語研究所)	22
多様な研究分野に利用可能な超高精細・高精度手話言語データベースの開発	[P-1-12]
長嶋祐二(工学院大学)・原大介(豊田工業大学)・堀内靖雄(千葉大学)	
酒向慎司(名古屋工業大学)・渡辺桂子(工学院大学)・菊澤律子(国立民族学博物館)	
加藤直人(NHK放送技術研究所)・市川薫(千葉大学/工学院大学)	22

英語における頭子音連結の多変量解析	[P-1-13]	
張一新(九州大学：学生)・中島祥好(九州大学)・于曉陽(九州大学：学生)		
上田和夫(九州大学)・岸田拓也(九州大学)		23
UD Japanese-BCCWJ の構築と分析	[P-2-01-E]	
大村舞(国立国語研究所)・浅原正幸(国立国語研究所)		23
LINE データベースの設計と属性付与の現状について	[P-2-02-E]	
宮寄由美(国立国語研究所)		24
『日本語歴史コーパス(CHJ)』の教育利用の実践報告	[P-2-03-E]	
宮城信(富山大学)・江口遼至(金沢高等学校)		24
双方向 LSTM による分類語彙表番号を語義とした all-words WSD	[P-2-04-E]	
新納浩幸(茨城大学)・鈴木類(茨城大学)・古宮嘉那子(茨城大学)		25
『キングコーパス』の構築と活用	[P-2-05-S]	
高橋雄太(明治大学/日本学術振興会：学生)		25
『明六雑誌』『東洋学芸雑誌』の特徴語から見る明治前期書き言葉の語彙特性	[P-2-06-S]	
近藤明日子(東京大学/国立国語研究所：学生)		26
「飲み倒す」とはどういう意味なのか		
ーGoogle 検索を利用した日本語の低頻度複合動詞の分析ー	[P-2-07-S]	
SEO MINCHEOL(立命館大学：学生)		26
先天性全盲ろう児の音声言語訓練長期記録の分析状況及び保存活動	[P-2-08]	
菊池英明(早稲田大学)・市川熹(千葉大学)・岡本明(筑波技術大学)		
長嶋祐二(工学院大学)・藤本浩志(早稲田大学)・引田秋生(元山梨県立山梨盲学校)		27
『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性	[P-2-09]	
馬場俊臣(北海道教育大学)		27
脚本テキストに基づくコーパス文体論の可能性		
ーテレビドラマ脚本に注目してー	[P-2-10]	
松下晶子(専修大学大学院文学研究科：学生)・丸山岳彦(専修大学/国立国語研究所)		28
『UniDic』を活用した語構造情報付与の試み		
ー『日本語歴史コーパス』に出現する語を対象にー	[P-2-11-E]	
村山実和子(国立国語研究所)		28
Twitter で使われる「深い」の意味		
ー「強い」「すごい」と比較してー	[P-2-12]	
加藤恵梨(大手前大学)・山下紗苗(明石工業高等専門学校：学生)		
上泰(明石工業高等専門学校)		29
日本語の二重目的語構文の基本語順について	[P-2-13]	
浅原正幸(国立国語研究所)・南部智史(モナシユ大学)		
佐野真一郎(慶應義塾大学)		29
比喩指標としての「感じる」		
ー文法形式と比喩の関係ー	[O-2-01-S]	
菊地礼(中央大学：学生)		30
日本語 wikipedia を用いた慣用句の構成性の数値化	[O-2-02-S]	
岡田優也(関西学院大学：学生)		30
「XX(と)」、「XXな」、「XXしい」、の構造・文法機能		
ー畳語による生産性についてー	[O-2-03-S]	
陳祥(筑波大学：学生)		31
ニュースを対象にした手話マルチメディアコーパスの構築	[O-3-01]	
加藤直人(NHK 放送技術研究所)・内田翼(NHK 放送技術研究所)		
東真希子(NHK 放送技術研究所)・梅田修一(NHK 放送技術研究所)		31
ベイズモデルによる方言音声共通語化過程の分析	[O-3-02]	
前川喜久雄(国立国語研究所)		32
言「考」不一致の言語学：コーパスはどこまで意識に迫れるか	[I-1-01]	
吉川正人(慶應義塾大学)		32

「日常会話コーパス」活用環境の構築	[P-3-01-E]	
山口昌也(国立国語研究所)		33
「よい子」って誰？		
－政策ニュース映画のナレーション表現に関する研究の一環として－	[P-3-02-E]	
春木良且(フェリス女学院大学)・田中弥生(東京大学：学生)		33
敬語接頭辞異形「お～」「ご～」両者の用例のある語について	[P-3-03-E]	
服部匡(同志社女子大学)		34
撥音(の解析)は機械(UniDic)にとっても簡単ではなかったんだ！		
－BCCWJを中心に－	[P-3-04-E]	
劉志偉(埼玉大学)		34
『現代日本語書き言葉均衡コーパス』書籍サンプルに対するNDC記号拡張アノテーションとNDC形式区分を用いた「随筆」の文体分析	[P-3-05]	
加藤祥(国立国語研究所)・櫻井芽衣子(日本工業大学)		
森山奈々美(津田塾大学：学生)・浅原正幸(国立国語研究所)		35
『現代日本語書き言葉近郊コーパス』に対する名詞述語文アノテーション	[P-3-06]	
今田水穂(文部科学省)		35
クラウドソーシング発注文書におけるレジビリティの量的分析	[P-3-07]	
岩崎拓也(国立国語研究所/一橋大学：学生)		
井上雄太(一橋大学：学生)		36
話し言葉における代名詞「あれ」の用法の分布	[P-3-08]	
山崎誠(国立国語研究所)		36
現職教員による児童・生徒作文の評価基準の分析	[P-3-09]	
宮城信(富山大学)・浅原正幸(国立国語研究所)・今田水穂(文部科学省)		37
スペイン語における前置詞句の数・定性		
7前置詞のクラスタリング	[P-3-10]	
喜多田敏嵩(東京外国語大学：学生)		37
マルチアクティビティに伴う発話の分類: 遂行発話と雑談	[P-3-11-E]	
天谷晴香(国立国語研究所)		38
コーパスに基づく字順転倒漢語の網羅的把握の試み	[P-3-12]	
間淵洋子(国立国語研究所)		38
実践医療用語の語構成要素抽出の試み	[P-3-13]	
内山清子(湘南工科大学)・岡照晃(国立国語研究所)・東条佳奈(目白大学)		
小野正子(西南女学院大学)・山崎誠(国立国語研究所)・相良かおる(西南女学院大学)		39
日本語の非流ちょう性		
－とぎれと延伸の数量調査から－	[P-4-01-E]	
佐々木藍子(国立国語研究所/東京学芸大学：学生)		
砂川有里子(筑波大学名誉教授)・浅原正幸(国立国語研究所)		39
『日本語日常会話コーパス』モニター公開版の特徴	[P-4-02-E]	
小磯花絵(国立国語研究所)・天谷晴香(国立国語研究所)・居關友里子(国立国語研究所)		
白田泰如(国立国語研究所)・柏野和佳子(国立国語研究所)		
川端良子(国立国語研究所：学生)・田中弥生(国立国語研究所：学生)		
西川賢哉(国立国語研究所)・伝康晴(千葉大学/国立国語研究所)		40
日本語学習者属性別の言語行為の対話自動生成への適用に関する一考察	[P-4-03-E]	
太田博三(放送大学：学生)		40
『現日研・職場談話コーパス』中納言版公開データの作成	[P-4-04-E]	
柏野和佳子(国立国語研究所)・大村舞(国立国語研究所)・西川賢哉(国立国語研究所)		
小磯花絵(国立国語研究所)		41
日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic	[P-4-05]	
首藤公昭(福岡大学名誉教授)・田辺利文(福岡大学)・高橋雅仁(久留米工業大学)		41
語彙多様性指標の可視化と単回帰分析によるTTRの補正	[P-4-06]	
今田水穂(文部科学省)		42

二字漢語を構成する漢字の造語力の変化		
－『現代雑誌九十種の用語用字』データと『現代日本語書き言葉均衡コーパス』の比較を通して－	[P-4-07]	
本多由美子 (一橋大学：学生)		42
方言音声に対するテキスト自動アライメントの試み	[P-4-08]	
石本祐一 (国立国語研究所)		43
単語の分散表現を用いた領域における出現単語の特徴分析	[P-4-09]	
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)		43
形容詞感動文における曖昧性回避の条件	[P-4-10]	
西内沙恵 (国立国語研究所)		44
コーパスを用いた授業における成果と課題		
－ノンネイティブ日本語教師研修における授業を例に－	[P-4-11]	
清水まさ子 (国際交流基金日本語国際センター)		
木田真理 (国際交流基金日本語国際センター)		44
『日本語話し言葉コーパス (CSJ)』模擬講演における節頭フィラーの特徴	[P-4-12]	
渡辺美知子 (国立国語研究所) ・是松優作 (東京大学：学生)		45
『国語研日本語ウェブコーパス』からの新規語彙素獲得の試み	[P-4-13]	
岡照晃 (国立国語研究所)		45
アクセント音調の諸相とその動態形式	[O-4-01]	
佐藤大和 (東京外国語大学)		46
日本語複単語表現レキシコン JMWEL の概要		
－動詞性表現を中心に－	[O-4-02]	
首藤公昭 (福岡大学名誉教授) ・田辺利文 (福岡大学) ・高橋雅仁 (久留米工業大学)		46
Information		47
ポスター設営図		48
コーパス開発センター包括的検索系デモ		48
ランチスペース		49
ランチマップ		50

Abstract

[O-1-01-S]

『現代日本語書き言葉均衡コーパス』のロシア語翻訳データの構築

宮内拓也 (東京外国語大学/日本学術振興会：学生) ・ Prokhorova Maria(東京外国語大学：学生)

『現代日本語書き言葉均衡コーパス』(の一部のデータ)には、既に英語、イタリア語、インドネシア語、中国語の翻訳データが構築されているが、新たにロシア語の翻訳データを構築した。対象となるテキストは『現代日本語書き言葉均衡コーパス』新聞(PN)コアデータ16サンプル(総語数は短単位で全16,657語)とし、ロシア語翻訳データの総語数は13,070語となった。本データの構築にあたっては、日本語からロシア語へ人手による翻訳を行った。また、日本語とロシア語の言語構造の違いや各言語における新聞特有の書き方(タイトルや日付の書き方など)の違い等により、翻訳に困難を生じさせた箇所も多くあった。本発表では、翻訳データの構築方法、翻訳の際の留意点の詳細を述べる。さらに、この翻訳データの構築により、原データと並べることで疑似的な日露対訳コーパスとしての利用も可能であり、本データは日露対照研究への活用できると考えられる。本発表では、このような活用の可能性についても述べる。

【口頭発表】9/4(火)10:10-10:35

【利用する言語資源】現代日本語書き言葉均衡コーパス(BCCWJ)

[O-1-02-S]

中古語における形容詞テ形をめぐって—形容詞の意味分類との関わりから—

菊池そのみ(筑波大学：学生)

本発表は中古和文資料を対象とし、形容詞のテ形(連用形に接続助詞「て」が後接するもの)とゼロ連用形(連用形に接続助詞「て」が後接しないもの)との比較から中古語の形容詞テ形の出現傾向を明らかにするものである。主に『日本語歴史コーパス平安時代編』を使用し、形容詞テ形と形容詞ゼロ連用形の用例を抽出して両者の比較から以下の3点を明らかにした。まず、形容詞の連用形全体に占める形容詞テ形の割合は1割程度であり、動詞の場合はテ形が9割を超えることと対称的な結果が得られた。これに加えて現代語における同形式との比較によって通時的な変化についても問題を提起した。次に形容詞テ形の出現傾向は文章のスタイルに影響を受けないことを指摘した。最後に「あり」、「をり」などの存在動詞が後続する場合についてテ形の出現傾向を形容詞の意味分類を踏まえて分析した。その結果、テ形の場合には感情や評価を表す形容詞が多く、ゼロ連用形の場合には状態を表す形容詞が多いことが明らかとなった。

【口頭発表】9/4(火)10:35-11:00

【利用する言語資源】日本語歴史コーパス(CHJ)、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-1-03-S]

日本語文における連用修飾語成分に見られるパラレルについての一考察―「赤く変わる」と「赤に変わる」は同じか―

王棟 (東京外国語大学：学生)

本発表はこれまで注目されてこなかった日本語の連用修飾語に見られるパラレル「形容詞ク動詞」(以下 A ク V) と「名詞ニ動詞」(以下 N ニ V) の棲み分けに注目した論考である。本稿は大規模コーパス、現代日本語書き言葉均衡コーパス (BCCWJ) の用例に基づいて、「赤ク／赤ニ動詞」を「A ク／N ニ V」の例として考察した。その結果、「A ク V」と「N ニ V」の棲み分けについて以下のことを指摘する。(1) 通常、動詞の補語と修飾語の位置に現れるのは形容詞「赤ク」であり、「赤ニ」の使用は稀である。(2) 「赤ニ」は色によって表される意味概念の対立のある文脈に現れやすい。(3) 「色の多様性に言及する場合」と「色変化の始まりがある場合」と「色が詳細な説明を受ける場合」において、「赤ニ」の使用は文法的な面において義務的と言ってよいが、事象を説明する機能的な面において「赤ニ」の補足を行っている。

【口頭発表】 9/4 (火) 11:00-11:25

〔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-1-04-S]

連体助詞の「ノ」と文体の関係

森秀明 (東北大学：学生)

名詞の頻度と文体には強い関連性があり、硬い文体や客観的な文体ほど名詞の頻度が高い。一方、連体助詞のノは名詞の頻度に連動して増減する。それでは、ノと文体の関係はどうだろう。硬い文体ほど、名詞の増加以上にノが増加するのだろうか。本研究では BCCWJ 図書館書籍のサンプルに文体指標をつけた国立国語研究所 (2015) を利用し、文体の違いによる名詞とノの回帰直線の傾きの関係を調査した。回帰直線は外れ値に弱く、この除去が分析のカギとなる。図書館書籍では固有名詞や数詞が列举されるサンプルが存在し、これらが外れ値となっている。そこで文書構造タグの <figureBlock> と <list> が存在するサンプルを除き、普通名詞と普通名詞に接続するノに絞って回帰直線の傾きを調査した。その結果、文体別の傾きに変化は見られなかった。ノは、名詞に連動して増減するだけで、その頻度に文体の影響はない。ノの頻度は、人間の意志や個性とはほとんど無関係に増減している可能性がある。

【口頭発表】 9/4 (火) 11:25-11:50

〔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-1-01-E]

日本語日常会話コーパスに対する談話行為アノテーションの試み：タグ選択が困難な事例に焦点を当てて

居關友里子 (国立国語研究所) ・ 門田圭祐 (早稲田大学：学生) ・ 伝康晴 (千葉大学/国立国語研究所)

本研究では日常生活の中に生じた、具体的な文脈の中に埋め込まれた会話を扱った「日本語日常会話コーパス (CEJC)」に対する談話行為アノテーションの試みについて報告を行う。現在試行中の枠組みについて紹介した上で、実際のアノテーション作業の中で見出された談話行為の判断が困難な事例を示し、その要因について CEJC の特性を参照しながら議論する。

【ポスター発表】 9/4 (火) 13:00-14:15

〔利用する言語資源〕 日常会話コーパス (CEJC)

[P-1-02-E]

「『了解』は使わないように」「了解です！」

高橋圭子 (フリーランス) ・ 東泉裕子 (フリーランス) ・ 佐藤万里 (フリーランス)

近年、ビジネスマナーに関する書籍やウェブ上において、「了解」は上から目線の言葉で失礼なので使わないほうがよい、とする記述が少なからず見られるようになった。辞書などの記述からは、応答詞としての「了解」は、無線や軍隊・警察などで使用されていた専門語・職業語から、主に若者同士のメールに普及し、ビジネスシーンにも広がったことがわかる。その一方で、ビジネスシーンでの使用には違和感を覚える、目上や社外の人に対して使うのは失礼であるとする言説をも生み出すようになったようである。本発表では、各種コーパスなどを調査し、(1) 応答詞としての「了解」、(2) 丁寧な応答表現としての「了解しました」「了解いたしました」「了解です」、(3) 「了解」は失礼な言葉との言説、のそれぞれについて、その出現と広がりさまを検証する。

【ポスター発表】 9/4 (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), 日本語歴史コーパス (CHJ), 国語研日本語ウェブコーパス (NWJC), 名大会話コーパス

[P-1-03-E-S]

NWJC における敬語使用とレジスターとの関係

金賢眞 (大阪大学：学生)

発表者は、現在、NWJC を用いた敬語研究を志向し、その一環として、謙讓表現「ご～する」を尊敬用法で用いた「誤用例」の調査・分析を進めている。WEB コーパスは、この種の「誤用例」を収集するのに好適な言語資料だが、残念ながら、NWJC にはレジスターの情報が付与されていない。敬語の使用および誤用にはレジスターによる差があると考えられ、たとえば、ブログなどでは敬語の使用は多くないが、その分、誤用が現れやすい傾向があり、逆に、企業や政府機関のホームページなどでは敬語が多く使用されるものの、誤用は少ないと予想される。しかし、こうした予想を NWJC によって直ちに確認することはできない。そこで、本発表では、NWJC から得た「ご～する」の使用例を仮設したレジスターに分類し、動詞別の使用頻度や動詞ごとの誤用率がレジスターによってどのように異なるかを統計的に分析して、NWJC における敬語使用とレジスターとの関係を具体的に検討する。

【ポスター発表】 9/4 (火) 13:00-14:15

〔利用する言語資源〕 国語研日本語ウェブコーパス (NWJC)

[P-1-04-E-S]

学校お便り文書の高頻出語彙の縦断的研究

－ 4 年生から 6 年生までの名詞・サ変名詞・動詞の分析

今村桜子 (横浜国立大学：学生)

首都圏の公立小学校のお便り文書 (3 年分 712 部) を OCR ソフトでデータ化し、KH Coder で形態素解析を行った。本研究は、(1) 学年ごと (4～6 年生) の高頻出語や、(2) 学年による異同を明らかにし、外国人保護者の日本語支援へ提言をすることを目的とする。総語数は 4～6 年でそれぞれ 56,968 語、106,084 語、77167 語。異なり語数は 7,420 語、9,935 語、9,395 語であった。(1)4～6 年生の高頻出語を抽出し、品詞ごとの上位 100 語を明らかにした。(2)4 年生の上位 100 語が 5、6 年生の上位 100 語に含まれる割合は、名詞で 5 年の 74%と 6 年の 75%、サ変名詞で 83%と 81%、動詞で 82%と 79%に上ることが分かった。また、頻度グラフにより少数の高頻出語と多数の低頻出語が観察される。更に 6 年生に関連のあるサ変名詞「卒業」は 4 年生で 32 回、5 年生で 66 回、6 年生で 104 回 (20 位) 出現する。高頻出語の学習は次年度以降の読取りに一定の効果があり、学校文化理解や内容スキーマ活性化に有用であると提言する。

【ポスター発表】 9/4 (火) 13:00-14:15

〔利用する言語資源〕 KH Coder ・学校お便りコーパス (自作)

[P-1-05-S]

児童・生徒作文の日本語修辞ユニット分析と教員評価の検討

田中弥生 (東京大学/国立国語研究所：学生)

本研究は、作文評価への修辞機能と脱文脈化程度という新たな観点の提示を試みるものである。選択体系機能言語理論における英語談話分析手法の一つである修辞ユニット分析の手順を基に、テキスト内で用いられている修辞機能を特定し、脱文脈化程度が高い表現か低い表現か、すなわち事象が一般的なこととして表現されているか、特定の個人的なことやその場のことと表現されているかなどを示す。作文は従来、使用語彙、文字種、語種、文の種類や構造など様々な観点で評価されてきた。本発表では、児童・生徒によって同一のテーマで書かれた作文について、学年の違いによる修辞機能と脱文脈化程度の用い方の特徴を明らかにするとともに、小中学校の教員による評価の高低との関連から、作文評価における新たな観点の提示の可能性を検討する。

【ポスター発表】 9/4 (火) 13:00-14:15

〔利用する言語資源〕『児童・生徒作文コーパス』

[P-1-06-S]

日本語における慣用句の逸脱使用がもつ言語機能—形容詞の反義語への置き換えを手がかりに—

鈴木あすみ (東北大学：学生)

慣用句は一般に固定性が高く、語彙的柔軟性を欠くとされている。例えば、「食が細い」に対して「食が太い」は一般に慣用句としては認められず、慣用句辞書等への掲載も確認されない。しかし、慣用表現の一部を変形することで作られる逸脱的な言い回しは一種の言葉遊びとして、受け手の目を引くために文学作品や広告などで用いられることがある。本研究では『国語研日本語ウェブコーパス (NWJC)』を用いた調査を行い、「食が太い」のような耳慣れない句がブログや掲示板、Q&A サイトなど一般的な人々による日常的な書き言葉の中にも多用されていることを確認した。このようなウェブサイトでは、ブログ本文に対してコメントが寄せられるといったように、複数人が談話の参与者となっている場合がある。こうした性質をもつテキストにおける慣用表現の逸脱使用は、(1) 表現そのものを際立たせる詩的機能、(2) 受け手の注意を引いたり自分が談話に参加していることを確認したりする交話的機能を担っていると考えられる。

【ポスター発表】 9/4 (火) 13:00-14:15

〔利用する言語資源〕国語研日本語ウェブコーパス (NWJC)

[P-1-07]

日本語歴史コーパスの現代語辞書における未知語義判定システム

田邊絢 (茨城大学：学生) ・ 古宮嘉那子 (茨城大学) ・ 浅原正幸 (国立国語研究所) ・ 佐々木稔 (茨城大学)
新納浩幸 (茨城大学)

日本語歴史コーパス中の単語には、現代語と同様の意味で扱われている単語と、古語特有の意味を持つ単語がある。本研究では、この現代語にはない古語特有の単語の語義（言葉の意味）を未知語義と定義して、日本語歴史コーパス中から、未知語義を検出するシステムの提案を行う。具体的には、日本語歴史コーパス中の単語を、(1) 現代の分類語彙表でその単語の分類番号として登録されている語義をもつ語、(2) 現代の分類語彙表にある語義をもつが、現在その語義は、その言葉の語義として分類語彙表は登録されていない語、(3) その語義の定義が現代の分類語彙表にないため、分類番号が振られていない語、の3種類にクラス分けする。実験では、各単語について、出現書字形や見出しなどの8要素を基本素性として用いた。また、別の日本語歴史コーパスから word2vec を用いて、3種類の単語の分散表現のベクトル (50次元, 100次元, 200次元) を作成し、素性として加えた。それぞれ SVM を用いて正解率を比較したところ、日本語歴史コーパス中の未知語義の検出において、単語の分散表現のベクトルが正解率を向上させることが分かった。

【ポスター発表】 9/4 (火) 13:00-14:15

【利用する言語資源】 分類語彙表, 日本語歴史コーパス (CHJ)

[P-1-08]

形態素解析器『Sudachi』のための大規模辞書開発

坂本美保 (株式会社ワークスアプリケーションズ) ・ 川原典子 (株式会社ワークスアプリケーションズ)
久本空海 (株式会社ワークスアプリケーションズ) ・ 高岡一馬 (株式会社ワークスアプリケーションズ)
内田佳孝 (株式会社ワークスアプリケーションズ)

我々は、汎用的な日本語形態素解析器『Sudachi』とその辞書を開発した。本稿では、Sudachi の辞書開発内容について述べる。我々は、まず、UniDic をベースとして、見出し表記、品詞、各種パラメータ等、形態素解析をするための辞書情報を整えた。次に、実用上 UniDic に不足している語句を見出しとして追加した。これには、NEologd から取り込んだ膨大な固有名称も含まれる。さらに、登録見出しについて、アプリケーションが利用しやすい形態素単位の整備、表記のゆれを同一視するための正規化表記の整備等を行い、辞書内容を充実させた。また、形態素解析精度の向上のため、UniDic 由来の見出しについても、弊害となる見出しの抑制や間違いの修正、形態素単位の調整を行った。我々のこれまでの成果は、最新版の辞書ソースに反映し OSS として公開している。

【ポスター発表】 9/4 (火) 13:00-14:15

【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)

[P-1-09]

英語における前置詞句についての音響分析

于曉陽(九州大学：学生)・中島祥好(九州大学)・張一新(九州大学：学生)・岸田拓也(九州大学)
上田和夫(九州大学)

英語における前置詞句のアクセントは、非英語母語話者にとって、学習が難点の一つとなる。本研究では、Kishida et al. (2016) が提案した起点移動因子分析という分析手段を用い、前置詞と直後の名詞又は名詞句からなる前置詞句に着目して、前置詞句の音響的特徴および知覚的な役割を調べることを目的とする。前置詞句の役割を明らかにするために、英語母語話者三名(男1名、女2名)が発話したイギリス英語音声データベースを構築し、研究を行った。収録された音声に対し、音素ごとにラベルを付けた。前置詞句を含む対象音声を20 臨界帯域に分割して、因子分析を行った。そこから、3 因子を抽出した。3300 Hz を超える周波数範囲と密接に関連する high factor の因子得点は、前置詞よりも名詞句の方が明らかに高いという結果が得られた。しかし、これ以外の因子得点については、明確な差が出なかった。以上の分析により、3300 Hz 以上の high factor は名詞句を知覚する際、重要な役割を果たしていることを示唆している。〔利用する言語資源〕 Fukuoka-Galway English Speech Database

[P-1-10]

副詞の程度性の下位分類の試み—「あまり・そんなに・それほど・たいして」を例に一

劉時珍(専門学校非常勤)

本発表では否定と呼応する4つの類義副詞「あまり」「そんなに」「それほど」「たいして」(以下を「4つの副詞」と呼ぶ)を例に、副詞の程度性について、『現代日本語書き言葉均衡コーパス』(以下を「BCCWJ」と呼ぶ)の調査結果に基づき、先行研究を踏まえた上で、副詞の程度性の下位分類について考察した。研究方法としてはBCCWJを用い、4つの副詞を、「たいして」(全数700例)を除き1000例ずつ無作為抽出し、それぞれの後ろの係り先に注目し、特に形容詞の評価性(『現代形容詞用法辞典』)に基づいて、4つの副詞の正の評価性と負の評価性を点数化した。その結果、「あまり」と「たいして」は負の評価性がより強く、「そんなに」は正の評価性がより強く、「それほど」は両方ほぼ同じという結果になった。以上のことから、「あまり」は負の評価性をより用いるものの、肯定形の「あんまり」という過度の否定により、程度性の部分性と評価性を両方持つことにより数多く使われ、一方、「そんなに」の正の評価性がより強いのは極限の程度を否定するだけの副詞であるためと結論づけた。程度性の下位分類としては「あまり」と「たいして」が同一の、「そんなに」と「それほど」が同一の下位分類に属することが考えられる。

【ポスター発表】9/4(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)、『現代形容詞用法辞典』

[P-1-11-E]

『日本語日常会話コーパス』構築における Praat の利用

西川賢哉 (国立国語研究所)

国立国語研究所で構築を進めている『日本語日常会話コーパス』(CEJC) のアノテーション作業 (書き起こし・短単位情報付与作業) を支援するために、無償の音声分析ソフトウェア Praat を利用したツールをいくつか開発した: (i) [Praat 起動] 必要な情報 (ファイル名・時刻情報等) が記された Emacs バッファ, あるいは形態論情報修正ツール「大納言」の検索結果画面から Praat を起動し、転記情報とともに当該箇所を表示するツール, (ii) [転記保存] Praat TextGridEditor 上で変更した転記を, CEJC 転記ファイル (タブ区切り形式) に上書き保存するツール, (iii) [メモ] TextGridEditor 上で選択された区間にある転記情報を, その他必要な情報 (ファイル名・時刻情報等) とともにクリップボードにコピーするツール, (iv) [別音声聴取] 当該会話に参加している別の話者の音声ファイルを追加で開くツール, など。これらのツールを用いることで、音声聴取をはじめとする、話し言葉コーパス構築に不可欠な作業が簡単な操作で行なえるようになり、作業の効率化および精度の向上が期待できる。

【ポスター発表】 9/4 (火) 13:00-14:15

【利用する言語資源】 日常会話コーパス (CEJC)

[P-1-12]

多様な研究分野に利用可能な超高精細・高精度手話言語データベースの開発

長嶋祐二 (工学院大学)・原大介 (豊田工業大学)・堀内靖雄 (千葉大学)・酒向慎司 (名古屋工業大学)

渡辺桂子 (工学院大学)・菊澤律子 (国立民族学博物館)・加藤直人 (NHK 放送技術研究所)

市川熹 (千葉大学/工学院大学)

手話は言語であるにもかかわらず、音声言語と比べて言語学、工学を含む関連諸分野での研究が進んでいない。本発表では、各個分野における手話研究および学際研究の推進を目的とした、様々な分野の研究者が共通に利用できる汎用的な日本手話の語彙データベース作成について報告する。言語学者の望むデータ形式と、工学や認知科学の分野で望むデータの形式は異なることが予想される。多分野での利用を可能にするためには、分析や解析内容に応じて手話の多視点の画像、3次元動作データ、深度画像など様々なデータ形式を含むことが望まれる。本データベースでは、120fps の超高精度な3次元動作、60fps の4K やフルHD の高精細の映像、30fps の深度センサーによる映像を異なるフレームレートで収録する。さらに、時間軸上で同期したこれらのデータを、各分析者が得意とするデータ形式で解析することを可能にする。データベース上の様々な形式データを同期解析できるアノテーション支援システムも開発する予定である。また将来的に、語彙だけでなく、文や会話文の収録ができるよう、方法を検討中である。これにより、様々な視点からの同一手話の解析が可能となり、手話言語に関する新たな知見が得られることが期待できる。

【ポスター発表】 9/4 (火) 13:00-14:15

【利用する言語資源】 多用途型日本手話言語データベース

[P-1-13]

英語における頭子音連結の多変量解析

張一新(九州大学：学生)・中島祥好(九州大学)・于曉陽(九州大学：学生)・上田和夫(九州大学)
岸田拓也(九州大学)

英語学習者にとって、子音連結を適切に発音・知覚することは大切である。本研究では、英語母語話者が発話した英語音声録音し、その音響的特徴の分析を行った。英語音節の頭子音連結や、頭子音とその次の母音をスペクトル変化の観点から調べ、録音した音声のスペクトル変化を因子分析した。得られた3因子から鳴音性と密接に関連する2つの因子が取り出された。一方の因子得点が高い場合、もう一方の因子得点は0に近く、因子空間におけるその散布図はL字型のように分布した。単語ごとの頭子音連結から母音に移る際の因子得点はこのL字型分布に沿って変化することが分かった。この際に、頭子音連結はL型の中央部に分布し、最も点が密集したのは、角の点に当たる原点の近くだった。第一子音から第二子音の間で、因子得点は統計的に有意に変化するが、第二子音から母音の間では因子得点に有意差はなかった。子音-子音-母音連結(CCV連結)の各音素に鳴音性と特に相関の高いmid-low factorの特徴がよく反映されていることが分かった。
【ポスター発表】9/4(火)13:00-14:15

〔利用する言語資源〕Fukuoka-Galway English Speech Database

[P-2-01-E]

UD Japanese-BCCWJの構築と分析

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。本論文では、日本語のコーパスであるUD Japanese-BCCWJについて紹介をする。UD Japanese-BCCWJは現代日本語書き言葉均衡コーパス(BCCWJ)に付随する係り受け情報などを組み合わせて、UDへと変換するプログラムの開発し構築したBCCWJのUniversal Dependencieである。このUD Japanese-BCCWJは日本語のUDの中でも1980文章、約126万単語を含む最大規模また複数のレジスターを内包したデータセットである。他の日本語UDと比較しつつ、UD Japanese-BCCWJの紹介をする。またUD Japanese-BCCWJの構築手順についても最新のものを説明する。

【ポスター発表】9/4(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)

[P-2-02-E]

LINE データベースの設計と属性付与の現状について

宮崎由美 (国立国語研究所)

本発表では、2016 年～ 2018 年に収集している「LINE」データベースの設計と現状について、具体例とともに報告する。

- (1) データ提供量 (2016 年～ 2018 年 6 月末)
- (2) データ提供者の属性
- (3) 言語研究用の加工について
- (4) データに付与される属性情報について

【ポスター発表】 9/4 (火) 14:20-15:35

〔利用する言語資源〕

[P-2-03-E]

『日本語歴史コーパス (CHJ)』の教育利用の実践報告

宮城信 (富山大学) ・ 江口遼至 (金沢高等学校)

本発表では、『日本語歴史コーパス (CHJ)』(以下, CHJ) の『徒然草』を高校の古典の授業で活用した実践を報告する。容易に古典作品の実例に当たることができる CHJ は, 日本語研究ばかりではなく, 豊富な資料を検索・提示することによって, 興味関心を高め, 作品に対する理解を深められる教育現場においても親和性が高いと考えられる。また, 新学習指導要領でも高校の現場においてもアクティブラーニング型の学習の必要性が説かれている。現在学校現場でも ICT 環境が整いつつあり, 今後大規模データを教材として ICT 機器を活用した学習の増加が見込まれる。国語科の学習内容の中でもおそらくもっともその実施から遠いと思われる分野の一つが古典の授業であり, 本研究は CHJ の教育利用の実践研究であるとともに, 国語科の ICT 活用とアクティブラーニング型授業開発を目的とした探索的研究実践でもある。

【ポスター発表】 9/4 (火) 14:20-15:35

〔利用する言語資源〕 日本語歴史コーパス (CHJ)

[P-2-04-E]

双方向 LSTM による分類語彙表番号を語義とした all-words WSD

新納浩幸 (茨城大学) ・ 鈴木類 (茨城大学) ・ 古宮嘉那子 (茨城大学)

国語研より公開されている BCCWJ に対する分類語彙表番号アノテーションデータを訓練データとして、分類語彙表番号を語義とした all-words WSD システムを作成した。all-words WSD は系列ラベリング問題であるためラベル付きデータがあれば CRF により容易に解決できそうだが、実際は語義の数 (ラベルの数) が膨大であるため CRF の学習は困難である。ここでは双方向 LSTM を用いた。LSTM は時系列データを扱うニューラルネットワークであり、系列ラベリング問題も解くことができる。自然言語処理では単語列を時系列データとして扱うことで、LSTM が利用できる。さらに単語列として文が入力される場合は逆方向からの解析も可能となり、双方向 LSTM も利用できる。作成できたシステムを大規模コーパスに適用し、分類語彙表番号の分散表現が構築できる。それをを用いて分類語彙表番号のクラスタリングも可能となる。
【ポスター発表】 9/4 (火) 14:20-15:35

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

[P-2-05-S]

『キングコーパス』の構築と活用

高橋雄太 (明治大学/日本学術振興会 : 学生)

本発表は、昭和期の雑誌『キング』を資料として構築した『キングコーパス』の設計と活用についてである。国立国語研究所の明治・大正期の『太陽』のコーパスに続く資料として、大衆雑誌の『キング』を選定し、1933年と1941年でコーパスを構築した。『キングコーパス』の設計については、資料の選定方法、字数、語数、記事数、著者数といった基礎的な規模を示し、また文体・記事ジャンル・品詞分布の観点から、『太陽』との連続性について検討した。『キングコーパス』の活用については、和語の表記を例に、「当用漢字音訓表」との比較を行った。「当用漢字音訓表」採用の和語の動詞を対象に、各語の使用表記の一致度を計測した。その結果、経年変化で一致度が高くなる傾向があることがわかり、「当用漢字音訓表」の設計の背景に、近代における用字法の変化があることが明らかになった。
【ポスター発表】 9/4 (火) 14:20-15:35

【利用する言語資源】 UniDic, 日本語歴史コーパス (CHJ)

[P-2-06-S]

『明六雑誌』『東洋学芸雑誌』の特徴語から見る明治前期書き言葉の語彙特性

近藤明日子 (東京大学/国立国語研究所：学生)

明治前期を代表する雑誌である『明六雑誌』『東洋学芸雑誌』の語彙と、明治中期以降の近代雑誌の語彙とを対照し、『明六雑誌』『東洋学芸雑誌』に有意に高頻度に出現する語群（特徴語）を抽出、そこから当時の書き言葉の語彙の特性を考察する。抽出される特徴語は、(1) 文語文に使用され口語文では使用されない、(2) 明治前期に使用され明治中期以降は使用されない、(3) 記事内容が専門的で他の雑誌には同種の内容の記事の掲載がない、等の複数の要因によりもたらされていると考えられる。対照する語彙を変えて特徴語抽出を行うことで特徴語を分類し、どの語がどの要因により特徴語となっているのか明らかにすることを試み、当時の語彙の重層的特性を探る。

【ポスター発表】 9/4 (火) 14:20-15:35

〔利用する言語資源〕 日本語歴史コーパス (CHJ), 『東洋学芸雑誌』コーパス

[P-2-07-S]

「飲み倒す」とはどういう意味なのか

—Google 検索を利用した日本語の低頻度複合動詞の分析—

SEO MINCHEOL(立命館大学：学生)

コーパスという用語の定義には、おおむね「大規模」という単語が登場する。しかし、そのような(大規模な)コーパスであっても、日常生活における使用頻度の低い言葉に関しては、そこから有用な情報を得ることが難しい。本研究では、意味記述が不十分だと考えられる日本語の低頻度語彙的複合動詞を取り上げ、Google の検索エンジンとクローラーを利用し、用例を網羅的に収集した。このような方法は、従来困難であった低頻度語彙の用例分析を可能とする。本稿では、低頻度複合動詞である「飲み倒す」を取り上げ、その特徴を記述し、類義語関係にある「飲み尽くす」「飲み潰す」「踏み倒す」との比較分析を行った。分析結果、「飲み倒す」は「酒を飲んでその代金を払わないままにする」という本来の意味よりも、「たくさん飲む」という派生的な意味での使用が顕著であることが明らかになった。また、類義語の中では、「飲み尽くす」との類似性が高いことがわかった。

【ポスター発表】 9/4 (火) 14:20-15:35

〔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), Google の検索エンジン、筑波ウェブコーパス

[P-2-08]

先天性全盲ろう児の音声言語訓練長期記録の分析状況及び保存活動

菊池英明 (早稲田大学) ・ 市川熹 (千葉大学) ・ 岡本明 (筑波技術大学) ・ 長嶋祐二 (工学院大学)
藤本浩志 (早稲田大学) ・ 引田秋生 (元山梨県立山梨盲学校)

昭和25年からの約10年間、山梨県立盲学校で2名の先天全盲ろう児に対する音声言語獲得訓練と生活指導が行われ、数万点に及ぶ時系列的・多角的記録と資料が残されている。故梅津八三東大教授が指導し、一貫して進めてきた盲人の認知行動・心理の研究の知見をベースに、先天盲ろう児への教育という未知の課題に対して取り組んだ科学研究の実践過程記録である。言語獲得が極めて困難な先天性盲ろう児に対する数万件の実践記録群は、おそらく世界で唯一の極めて貴重な資料であり、盲ろう児当事者から表出された点字や録音資料からは、学習の進行程度を直接見ることが期待される。言語獲得プロセスの解明や盲ろう児教育に重要な示唆が得られるであろう。しかし最も質の悪い時代の紙や録音テープ等に記録され劣化が著しいため、現在電子化保存とデータベース開発を進めている。データベース開発後は山梨県立盲学校に移管、公開する計画である。同活動及び発音訓練用の木製口模型などや訓練記録、訓練経緯及び同校での分析状況等を紹介する。

【ポスター発表】9/4 (火) 14:20-15:35

【利用する言語資源】山梨盲学校盲ろう児教育実践資料群データベース

[P-2-09]

『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性

馬場俊臣 (北海道教育大学)

『現代日本語書き言葉均衡コーパス (BCCWJ)』内の「図書館サブコーパス」のサンプルに対して文体情報を付与した『BCCWJ 図書館サブコーパスの文体情報』のデータを利用し、「図書館サブコーパス」内の語 (語彙素) を対象として各語の文体差を数値化する試みを行い、この試みが文体研究において有効性・可能性があることを示す。まず、『文体情報』の専門度、硬度などの文体に関する5種類の指標別に、各語の平均値を算出する方法を示す。次に、各語の平均値について専門度・客観度・硬度・くだけ度の4指標は相互に強い相関があること、品詞別では感動詞と接続詞の4指標の平均値とそのばらつきが他の品詞に比べて特異であり品詞の特徴を反映していること、語種別では和語・漢語・外来語の特徴の違いが平均値に反映していることなどの全体的傾向を示す。さらに、各語の平均値が、語の文体差に関する内省判断と強い相関があることを先行文献の調査結果と比較して示す。

【ポスター発表】9/4 (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), 『BCCWJ 図書館サブコーパスの文体情報』

[P-2-10]

脚本テキストに基づくコーパス文体論の可能性
—テレビドラマ脚本に注目して—

松下晶子 (専修大学大学院文学研究科：学生) ・丸山岳彦 (専修大学/国立国語研究所)

現在、「日本脚本アーカイブズ推進コンソーシアム」により、1950年代以降のテレビドラマの脚本を収集し、それらを体系的に保存・アーカイブする活動が進められている。脚本は、「話されることを前提とした書き言葉」という点で特徴的な書き言葉であるが、これまでの言語研究の中でも顧みられることは少なかった。収集した脚本をコーパス化して定量的に分析することにより、新たな言語学的利用の可能性が開かれると考えられる。そこで本発表では、脚本のテキスト化・コーパス化を試験的に実施した経緯を述べ、そのデータを使ってどのような言語研究が可能になるかについて論じる。故市川森一氏による、1970年代から2010年代までの脚本、32作品をテキスト化し、パイロットスタディを実施した。このような分析は、近現代における言語の通時的変化の研究、ある作家の作品に関するコーパス文体論的研究などにつながると考えられる。

【ポスター発表】 9/4 (火) 14:20-15:35

〔利用する言語資源〕 脚本テキストデータ

[P-2-11-E]

『UniDic』を活用した語構造情報付与の試み
—『日本語歴史コーパス』に出現する語を対象に—

村山実和子 (国立国語研究所)

国立国語研究所では、上代から近代までの日本語を通時的に研究するための基礎資料として、『日本語歴史コーパス (CHJ)』の構築を進めている。すべてのテキストを齊一な単位で分割し、形態論情報を付与している点が特長である。ただし、その単位認定は現代語のコーパスに準拠しており、時代によって「語」の認定に揺れが生じる場合もある。そこで本研究では、CHJに出現する合成語に対し、その語構造情報を付与することについて有効性を検討する。そのアノテーションの方法として、形態素解析辞書『UniDic』の登録語に対して「構成語情報」を付与することを試みる。UniDicは各時代のコーパスの実例と紐付いていることから、追加した情報に基づき、コーパス別・時代別といった様々な条件下でのデータ分析が可能になることが期待される。一方で、歴史資料を対象とする場合、その構成要素や結合度の認定には困難も伴う。UniDicが規定する最小単位以下の要素（接辞など）への対処についても考察する。

【ポスター発表】 9/4 (火) 14:20-15:35

〔利用する言語資源〕 UniDic, 日本語歴史コーパス (CHJ)

[P-2-12]

Twitter で使われる「深い」の意味 —「強い」「すごい」と比較して—

加藤恵梨 (大手前大学) ・ 山下紗苗 (明石工業高等専門学校 : 学生)
上泰 (明石工業高等専門学校)

近年, Twitter で「うれしみ」などの感情を表す「み名詞」が多用されており, それらは「深い」や「強い」などの形容詞と共起することが多い。本研究では, Twitter で使われる「感情を表す『み名詞』+が深い」がどのような状況で使われ, どのような意味を表すのかについて分析する。また, 類義表現である「感情を表す『み名詞』+が強い」「『み名詞』+がすごい」が表す意味についても分析し, それらの意味の違いを明らかにする。分析の結果, 「感情を表す『み名詞』+が深い」は「ある感情の積み重なりにより, その感情の程度が高くなっているさま」と「感情の程度が普通の程度を超えているさま」という二つの意味を表すということを述べる。

【ポスター発表】 9/4 (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), SNS

[P-2-13]

日本語の二重目的語構文の基本語順について

浅原正幸 (国立国語研究所) ・ 南部智史 (モナシユ大学) ・ 佐野真一郎 (慶應義塾大学)

本稿では日本語の二重目的語構文の基本語順について予測する統計モデルについて議論する。『現代日本語書き言葉均衡コーパス』コアデータに係り受け構造・述語項構造・共参照情報を悉皆付与したデータから, 二重目的語構文を抽出し, 格要素と述語要素に分類語彙表番号を付与したうえで, ベイジアン線形混合モデルにより分析を行った。結果, 名詞句の情報構造の効果として知られている旧情報が新情報よりも先行する現象と, モーラ数が多いものが少ないものに先行する現象が確認された。

【ポスター発表】 9/4 (火) 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)

[O-2-01-S]

比喩指標としての「感じる」 —文法形式と比喩の関係—

菊地礼 (中央大学：学生)

本発表は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 (BCCWJ) を用いて収集した比喩表現データを分析・考察する。中村 (1977) 『比喩表現の理論と分類』によれば直喩の指標となる語句は7類82種359号と多岐にわたる。しかし、直喩の典型である「よう」以外の分析はなされていない。本発表ではコーパスを用いた網羅的な用例収集を行い、分析に耐える量を確保する。その一例を本発表は動詞「感じる」によって示す。「感じる」は「AガBヲ」「AヲBト」「AヲBデ」等の10の構文を作るが、「AニBヲ」「AヲBニ」等の8つの構文で比喩を表わすことが可能である。しかし、直喩と認定できる例はその中から限定される。これは「感じる」が比喩指標として機能することが例外的事例であることを意味する。モダリティ形式として文法化することが比喩指標には求められるが、「感じる」は特定の構文環境においてのみ不完全ながら文法化を果たし、比喩指標と同様の機能を得る。

【口頭発表】9/4 (火) 15:45-16:10

〔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)

[O-2-02-S]

日本語 wikipedia を用いた慣用句の構成性の数値化

岡田優也 (関西学院大学：学生)

本研究は、日本語慣用句の構成性の度合いを数値化する手法を用いて、慣用句と一般連語句の差異について調査するものである。構成性の数値化は、日本語 wikipedia の全記事からなるコーパスをもとに、調査対象とするひとつの慣用句について構成的連語句分散表現ベクトルと非構成的連語句分散表現ベクトルの2つのベクトルをそれぞれ獲得し、この2つのベクトルについてコサイン類似度を計算することによって行う。構成的連語句分散表現ベクトルとは、連語句の構成要素である単語の分散表現ベクトルから加算的に計算されるベクトルである。一方、非構成的連語句分散表現ベクトルとは、連語句を内部構造のないひとまとまりの表現だと捉え、コーパスから直接的に獲得されるベクトルである。獲得した2つのベクトルのコサイン類似度が慣用句の構成性の度合いを反映するものであることを主張するため、一般連語句についても同じ手法により構成性の数値化を行い、慣用句において獲得された数値との間に統計的に有為な差が生じることを示す。

【口頭発表】9/4 (火) 16:10-16:35

〔利用する言語資源〕 日本語 wikipedia

[O-2-03-S]

「XX(と)」、「XXな」、「XXしい」、の構造・文法機能 — 畳語による生産性について —

陳祥 (筑波大学：学生)

日本語の重複表現の形式は様々であるが、本稿で取り扱うのは「畳語」と呼ばれ、2つの語基の重複によって形成される語である。日本語には畳語が数多く、名詞、動詞など様々な品詞からなる。畳語には生産性が限られているが、「色々な」のように後ろに「-な」、「細々しい」のように後ろに「-しい」などを付け加えることによって、新しい語として形成することが可能である。今回は畳語表現である「XXと」、「XXな」、「XXしい」という3種類を研究対象とする。データとして『現代日本語書き言葉均衡コーパス (BCCWJ)』を使い、3種類それぞれの共起関係や文法的振る舞いを明らかにする。また、構造上においては「色々な、*色々しい、色々(と)」、「*軽々な、軽々しい、軽々(と)」のように、3種類における「XX」の使い分けがあると見られ、3種類それぞれの構造が異なることを課題として分析した。例えば、派生元の品詞から見ると、「XXと」の派生元である「X」の品詞は、名詞あるいは形容動詞、動詞、形容詞、副詞であるものがあり、バリエーションが一番豊かであることが分かった。また、BCCWJを使用し、副詞としての畳語は形容動詞としても反復形容詞としても使われている用例が見られる。それは、副詞としての畳語は「XXと」と接続することが最も多いが、「XXな」や「XXしい」の用例も見られることから、使用用法のうち自由度が最も高いと考えられる。今回の考察から、畳語による生産性の理解の助けになることを期待している。

【口頭発表】9/4 (火) 16:35-17:00

【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)

[O-3-01]

ニュースを対象にした手話マルチメディアコーパスの構築

加藤直人 (NHK 放送技術研究所) ・ 内田翼 (NHK 放送技術研究所) ・ 東真希子 (NHK 放送技術研究所)
梅田修一 (NHK 放送技術研究所)

我々が現在構築している、ニュースの手話マルチメディアコーパスについて述べる。手話はコミュニケーションツールであるので、従来のコーパスも対話を対象にしたものが多く、ニュースを対象にしたものは少ない。我々のコーパスは、Eテレで日々放送している「手話ニュース」、「手話ニュース 845」、「週間手話ニュース」をデータベース化したものである。手話ニュースは毎日放送されているので、映像データの収集は容易である。また、ニュースを対象としているので、会話とは異なり構文構造がしっかりしていると考えられる。更に、放送するまでに複数人のチェックが入るので、そこで使われる言語表現の普遍性は高いと考えられる。手話コーパスは手話書き起こし、日本語書き起こし、手話映像で構成されており、現在までに約13万文とこの分野では比較的大きなコーパスである。また、日本語でコーパスを検索することも可能である。

【口頭発表】9/5 (水) 10:00-10:25

【利用する言語資源】NHK 手話ニュースコーパス

[O-3-02]

ベイズモデルによる方言音声共通語化過程の分析

前川喜久雄 (国立国語研究所)

国語研による山形県鶴岡市における共通語化調査データのうち第1～3回調査の音声項目データをベイズモデルで分析し、方言音声共通語化過程の統計モデルを構築した。既に報告した第1回調査データと同様、第2回・第3回調査データも二項分布に基づくロジスティック回帰モデルを適用するには分散が大きすぎる(過分散状態)。そのため、ベルヌーイ分布の成功確率が種々の要因によって変動するベイズモデルを考案した。7種のモデルの性能をF値・平均予測誤差・WAICの三者で評価した結果、回帰直線の切片が話者と語彙の要因によって変動し、傾きが語彙の要因によって変動するモデルが最良モデルとなった。このモデルのF値は0.95に達しており、強い説明力を有している。さらにこのモデルにおける話者の個性性情報を「性別・言語形成地域・教育歴」の情報で置換したモデルを評価したところ、第2・第3回調査データについては、最良モデルとほぼ同等の性能を発揮するものの、第1回調査については性能がかなり低下することが判明した。

【口頭発表】9/5 (水) 10:25-10:50

【利用する言語資源】鶴岡調査データベース

[I-1-01]

言「考」不一致の言語学: コーパスはどこまで意識に迫れるか

吉川正人 (慶應義塾大学)

「ラ抜き言葉」のように、時に、ヒトの言語直感と実際の言語使用とが乖離することがある(前川 2006)。このような場合、よって立つ前提によって、直観か使用かいずれか一方が「正しく、もう一方はノイズにより歪められたものだ」とされることがほとんどである。しかしながら、直観も使用もランダムなものではなく、一定の傾向やパターンを示す以上は、どちらも言語に関する何らかのリアリティを持つものであり、特に文法について考える際には、いずれか一方を「歪み」として切り捨てることは適切ではないと考えられる。このように考えた際に、コーパスを利用して言語の規則性やパターンを記述・分析するようなアプローチでは、言語直観には反映されるが、言語使用には反映されない、あるいはされにくい側面は、少なくとも直接的には扱うことが難しいということが指摘できる。本発表では、言語直観、すなわち言語に対する「意識」的側面が言語使用の実態と乖離する事例をいくつか取り上げ、コーパスデータからそのような意識的側面にアプローチする可能性と限界について議論する。

【招待講演】9/5 (水) 11:00-11:50

[P-3-01-E]

「日常会話コーパス」活用環境の構築

山口昌也 (国立国語研究所)

本発表では、「日常会話コーパス」を活用するための環境構築について述べる。「日常会話コーパス」は動画・音声、転記テキストを含み、転記テキストには形態素解析結果などの言語学的な情報がアノテーションされている。本発表で提案する活用環境は、全文検索システム『ひまわり』と観察支援システム FishWatchr を統合することにより実現した。本環境を用いることにより、次のことが可能になる。(1)『ひまわり』で転記テキストを全文・単語検索し、当該位置の映像を FishWatchr で閲覧すること、(2)FishWatchr 上で動画再生位置に簡易なアノテーション（二つのユーザ定義ラベル、自由テキストを記述可能）を付与すること、(3)FishWatchr 上で転記テキストを表形式で表示し、選択した転記テキスト位置の動画を再生すること。また、動画の再生と同期させて転記テキストをスクロール表示すること。
【ポスター発表】 9/5 (水) 13:00-14:15

〔利用する言語資源〕 日常会話コーパス (CEJC)

[P-3-02-E]

「よい子」って誰？

－政策ニュース映画のナレーション表現に関する研究の一環として－

春木良且 (フェリス女学院大学) ・ 田中弥生 (東京大学：学生)

本研究では、昭和 20 年代後半から 30 年代に掛けて、各自治体で制作された、地域の復興を記録した行政映画（政策ニュース映画）のうち、神奈川県川崎市分を取り上げる。ニュース映画は、主に映像とナレーションから構成されている。ここではナレーション中で、特に地域の子供を表現するときに多用される「よい子」という表現に着目した。テーマによって、全くその表現が使われないものも多く存在する。どういふ子供が「よい子」なのか、よい子ではない子はなぜ、よい子とは呼ばれないのか、コンテンツとナレーションの相関から傾向を洗い出していく。

【ポスター発表】 9/5 (水) 13:00-14:15

〔利用する言語資源〕 川崎市政ニュース映画

[P-3-03-E]

敬語接頭辞異形「お～」「ご～」両者の用例のある語について

服部匡 (同志社女子大学)

「お～」と「ご～」のどちらも伴う語の存在が知られているが、コーパスでの網羅的調査が従来なかった。青空文庫・新聞記事データベース・自作ウェブコーパスの3種を用いた探索により、多数の語を発見した。青空文庫では、両形の用例のある形式は少なくとも173に昇り、同一著者が両形を併用している語も多い。時代小説では「疑似江戸語」と思われる例も見られた。また、高頻度語に絞った調査であるが、新聞記事では40、ウェブコーパスでは131の形式にオ/ゴ両形が見られた。尊敬語動詞句・一字漢語動詞・顧客向け使用などの観点からの観察を示す。

【ポスター発表】9/5 (水) 13:00-14:15

【利用する言語資源】UniDic, 新聞記事データベース, 自作ウェブコーパス, 青空文庫

[P-3-04-E]

撥音 (の解析) は機械 (UniDic) にとっても簡単ではなかったんだ!

— BCCWJ を中心に —

劉志偉 (埼玉大学)

現代語に限って考えても、日本語の撥音は実に種々雑多である。例えば、話し言葉には「君んち」「嫌んなる」「そんで」といった、くだけた言い方があるのに対し、書き言葉では「割れんばかりの拍手」「いざ行かん」「触れなば落ちん」等固い表現が挙げられる。また、用言の活用に関しては、いわゆる標準語においてだけでも「わかんない」「謝んなさい」「飛べんの」「かもしんない」「おもんない」のようにラ行音の全てが撥音化する場合がある。さらに近畿方言の「食べんで」「行きまんねん」等も考え合わせると、教科書では「飛ぶ」のテ形「飛んで」しか習わない日本語学習者にとって撥音が極めて難解であることは自明である。本発表では、BCCWJの非コアデータも視野に入れて、撥音の解析に関しては解析精度が98%に到底及ばないことを提示するとともに、具体的に「一般名詞」「オノマトペ」「漢語副詞」「漢字読み」「慣用語」「近畿方言」「呼称」「古典」「語尾」「固有名詞」「ぞんざい表現」「駄洒落」「同音異語」「動詞連用」「特定」「入力ミス」「話し言葉」「表記仮名」「表記仮名遣い」「表記漢字」「フィラー」「複合語」「(近畿以外)方言」「略語」「若者表記」「若者言葉」等の単純誤解析が多いことを明らかにする。また、コアデータの性格をもつ名大会話コーパスにおける撥音の誤解析が少ないことも、撥音の解析が解析機にとって簡単ではないことの証左となろう。

【ポスター発表】9/5 (水) 13:00-14:15

【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), 名大会話コーパス

[P-3-05]

『現代日本語書き言葉均衡コーパス』書籍サンプルに対する NDC 記号拡張アノテーションと NDC 形式区分を用いた「随筆」の文体分析

加藤祥 (国立国語研究所)・櫻井芽衣子 (日本工業大学)・森山奈々美 (津田塾大学：学生)
浅原正幸 (国立国語研究所)

現在、『現代日本語書き言葉均衡コーパス』に含まれる書籍サンプルに付与された日本十進分類法 (NDC) 分類記号 (主題による分類) に、補助分類を拡張する作業を進めている。国立国会図書館の NDC 情報 (8 版・9 版) を参照し、人手によって補助分類の確認と追加を行う。本発表は、現在までに作業の完了した図書館サブコーパス 10,551 サンプルについて、情報付与作業方法とその結果を報告する。本作業により、たとえば形式区分を利用し、ジャンルの分散する「随筆」「理論」「研究法」などのカテゴリで BCCWJ サンプルを分類することが可能となる。そこで、付与した形式区分「随筆」サンプル群を例とし、語彙特徴から文体的な傾向を調査した。さらに、柏野 (2015) の文体指標を用い、「随筆」の文体特徴として考えられてきた「主観的」で「軟らかく」「くだけている」傾向などを検証する。
【ポスター発表】 9/5 (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-3-06]

『現代日本語書き言葉均衡コーパス』に対する名詞述語文アノテーション

今田水穂 (文部科学省)

「現代日本語書き言葉均衡コーパス」コアデータに対して名詞述語文に関する文法情報付与を行った。付与したラベルの概数は名詞述語文の主語述語 13700 組、名詞述語の連体修飾節 3200、機能表現などの周辺のラベル 4600 である。主語は名詞主語とノ節主語に分類し、前者は is_a など述語との意味関係を、後者は分裂文焦点に対する文法関係を付与した。述語は通常の名詞述語の他、「X は Y になる」のような補語も若干数付与した。周辺表現ラベルは「する」が省略された漢語動詞、名詞述語由来の機能表現、述語が省略された節などを含む。既存の述語項構造データ (BCCWJ-PAS) と比べると名詞述語と他表現 (漢語動詞や非述語名詞) の判別に特化しており、PAS で述語ラベルが付与された名詞のうち 2800 例を名詞述語から除外、5200 例を追加している。本発表ではデータの設計とラベルの内訳を報告し、文法研究のための利活用の展望について説明する。
【ポスター発表】 9/5 (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-3-07]

クラウドソーシング発注文書におけるレジビリティの量的分析

岩崎拓也 (国立国語研究所/一橋大学：学生) ・ 井上雄太 (一橋大学：学生)

「顔の見えない相手」への文書による伝達を行う必要があるクラウドソーシングにおいて、まず重要になるのが読みやすさである。しかし、単にわかりやすい文法や表現を使用すればいいというわけではなく、読み手に自分(書き手)の意図を汲み取ってもらうためには、見やすい文書にしなければならない。可読性を上げる要素として考えられるのは、句読点や括弧といった符号(記号)、文と文の間の空白行などであるが、計量的な観点からの分析は多くない。本発表では、文書の見やすさ(レジビリティ)の一端を明らかにするため、クラウドソーシングで実際に使われた発注文書のデータベースをもとに、符号(記号)や空白行といった文章の周辺的事象に焦点を当て、その多寡と応募人数との相関の分析を行った。今回は、その結果を報告する。

【ポスター発表】9/5(水)13:00-14:15

【利用する言語資源】クラウドソーシング発注文書データベース(仮称)

[P-3-08]

話し言葉における代名詞「あれ」の用法の分布

山崎誠 (国立国語研究所)

代名詞「あれ」には、明確な指示用法(文脈指示、現場指示)のほかに指示はしているが、指示対象があいまいな、ぼかし用法がある。『新明解国語辞典』(第七版)では「はっきり口にしたくないこと、ちょっと忘れたこと、うまく言えないことなどの代りに用いる語。」としている。本発表では、話し言葉におけるこれら2用法の分布を調査し、その特徴を報告するものである。パイロット調査によると、『日本語話し言葉コーパス・学会講演』『同・模擬講演』『名大会話コーパス』からランダムに抜き出した例の観察では、およそ、どのコーパスでも指示用法と婉曲用法が半々程度という結果であった。婉曲用法では「あれなんだけど、～」という従属節になることが多かった。また、『名大会話コーパス』や『CSJ 模擬講演』で顕著なのは、「#でも、あの、あれですよ。#お金、あのー、ちゃんと替えてくれましたよ」(会話ID: data109, 開始位置: 143360)のように実際に言いたいことの前置きの使う場合が10%前後見られたことである。

【ポスター発表】9/5(水)13:00-14:15

【利用する言語資源】名大会話コーパス, 日本語話し言葉コーパス(CSJ), 多言語母語の日本語学習者横断コーパス(I-JAS)

[P-3-09]

現職教員による児童・生徒作文の評価基準の分析

宮城信 (富山大学)・浅原正幸 (国立国語研究所)・今田水穂 (文部科学省)

本発表は、児童生徒が作成した「手」という題の作文の評価のあり方に関する研究である。小中を問わず、作文の評価はほぼすべての教員が行っているが、その基準については経験則に基づくものであり、明確な基準は明示的ではない。本研究では教員間に共通する作文の評価基準があると仮定して、現職教員の協力を得て収集した作文を評価してもらい、その結果に基づいて評価基準の抽出、評価基準間の関連の分析を試みた。なお、作文の収集は同校・同題・同条件で1992年と2016年の2回実施された。作文を書く児童らの文章作成能力にも経年変化が見られる可能性があると考え、本研究ではその点にも着目して調査を実施した。具体的な分析の観点は以下ようになる。(1) 現職教員の児童の作文に対する総体評価と観点別評価の関係はどのようになっているか。(2) 教員の教育歴と作文の評価にはどのような関係があるか。(3) 1992年の作文と2016年の作文の評価はどのように異なっているか。

【ポスター発表】9/5 (水) 13:00-14:15

〔利用する言語資源〕児童・生徒「手」作文コーパス (1992年、2016年調査)

[P-3-10]

スペイン語における前置詞句の数・定性

7 前置詞のクラスタリング

喜多田敏嵩 (東京外国語大学：学生)

スペイン語学において、名詞の数・定性は活発な議論が展開されてきた分野であるが、前置詞句の項として生起する名詞に焦点を当てた研究は、部分的な記述が少数存在するのみである。本稿は、スペイン語において最頻出の前置詞である *de* および、*de* と可換な用法を有する6前置詞 *con*, *desde*, *en*, *para*, *por*, *sobre* の計7つを分析対象として、前置詞句の数・定性に関する Fernández Ramírez (1986) および Bosque (1996) の空間指示性による分類的記述の妥当性を検証するものである。データの収集にはコーパス検索ツール Sketch Engine において公開されている100億語規模の均衡ウェブコーパス esTenTen [2011, Eu + Am] を使用し、算出された7前置詞に後続する普通名詞の単数形生起頻度と、限定詞との共起頻度を2変数としたクラスタリングを実行した。その結果、これら7つの前置詞は *con*, *de*, *por*, *desde*, *en*, *para*, *sobre* の2群に分類され、Fernández Ramírez (1986) の挙げる分類基準と、*en* を *con*, *de*, *por* と同じクラスターに位置付ける Bosque (1996) の記述に検討の余地があることが分かった。

【ポスター発表】9/5 (水) 13:00-14:15

〔利用する言語資源〕esTenTen [2011, Eu + Am]

[P-3-11-E]

マルチアクティビティに伴う発話の分類: 遂行発話と雑談

天谷晴香 (国立国語研究所)

日常的に会話はマルチアクティビティ (複合活動) の中で行われることが多い。他の活動を主として行う場面での会話は、雑談のようなトピック会話に加えて、活動の遂行に関わる発話が見られる。トピック会話も主たる活動に関連して行われることが多く、全くその場の物事と関わらない発話は少ない。その場の物事に関する発話とその場がない物事に関する発話の区分は談話研究において重要である。Chafe (1994) は意識の宛先をもとにこれらの言語表現の分類を提示した。話し言葉においては話し手と聴き手が会話スペースを共有することがほとんどであるため、書き言葉に比べて場に即した言語活動が行われやすい。本発表では参加者が共同で料理活動を行いながら会話しているデータを扱う。料理行程を進捗させるための発話 (例: レシピを読み上げる、相手に指示を出す) と、料理の進捗に関わらない発話 (例: 言い間違いを正す、明日の予定を伝える) を分類し、それらの複合活動の中での構造的な位置と音声的特徴について報告する。

【ポスター発表】 9/5 (水) 13:00-14:15

〔利用する言語資源〕 日常会話コーパス (CEJC)

[P-3-12]

コーパスに基づく字順転倒漢語の網羅的把握の試み

間淵洋子 (国立国語研究所)

明治維新以降、西洋文明との接触により大量に流入した新しい物質・事象・概念を、日本語では多く「漢語」として受容したため、近代では漢語が飛躍的に語彙を増やしたが、この時期の漢語は、模索的・流動的であり、様々な点で多様で「ゆれ」の多い状態にあった。そのゆれの一つが、「華麗-麗華」、「遊戯-戯遊」のように、字順の相反する漢語対 (以下「字順転倒漢語」) である。これまでその事象の指摘や一部の語に対する研究は多く行われてきたが、全体像が明らかになっているとは言い難い。そこで、本研究では、発表者がこれまで行ってきた漢語における近代と現代との差異や、漢語の定着に伴う多様性収束の実態把握の一貫として、近代語コーパスと現代語コーパスを用いた字順転倒漢語の網羅的な実態把握と、その経年変化について報告する。字順転倒漢語の網羅的抽出方法、対となる漢語の意味関係に基づく分類、使用の有無や使用頻度を用いた経年変化の把握等を示し、語形の定着・淘汰の要因を分析する。

【ポスター発表】 9/5 (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), 日本語歴史コーパス (CHJ)

[P-3-13]

実践医療用語の語構成要素抽出の試み

内山清子 (湘南工科大学)・岡照晃 (国立国語研究所)・東条佳奈 (目白大学)・小野正子 (西南女学院大学)
山崎誠 (国立国語研究所)・相良かおる (西南女学院大学)

医療現場で用いられる電子カルテなどの記録文書（医療記録）に専門用語としての医療用語が大量に含まれている。医療記録に記載された言語情報を正確に理解・活用するためにはこれらの医療用語の理解が必要となる。医療記録に含まれる語には、複数の語からなる複合語や臨時一語も多く、これらは、病名、身体の一部名、処置名、薬剤名等、様々な用語から構成されている。しかし、現在はこの語構成要素の組み合わせのパターンや語構成要素間の関係などが曖昧である。そこで、本研究では複数の語からなる実践医療用語の語構成要素の抽出を試みる。ComeJisyoV5に含まれる用語を対象として MeCab 0.996 と UniDic-cwj-2.2.0 で形態素解析を行い、分割された単語の品詞情報を手がかりにして、単一単位となる得る語構成要素の条件を決定して抽出を行う。次に抽出した語構成要素に接続している複合語についても同様の分析を繰り返して抽出条件の整理を行なっていく。

【ポスター発表】 9/5 (水) 13:00-14:15

〔利用する言語資源〕 ComeJisyoV5

[P-4-01-E]

日本語の非流ちょう性

—とぎれと延伸の数量調査から—

佐々木藍子 (国立国語研究所／東京学芸大学：学生)・砂川有里子 (筑波大学名誉教授)
浅原正幸 (国立国語研究所)

これまでの日本語教育では、非流ちょうな発話の指導はほとんど行われていない。しかし、実際には母語話者であってもよどみのない流ちょうな発話を行うことはまれであるし、非流ちょうな発話が話し手のストラテジーとして用いられることや、聞き手の理解の促進につながることもある。そこで本研究では、非流ちょう性の要因となる「とぎれ」と「延伸」を取りあげ、「多言語母語の日本語学習者横断コーパス (I-JAS)」に格納された日本語母語話者データの数量調査を行った。その結果、ストーリーテリング (ST) とロールプレイ (RP) において、とぎれと延伸ではとぎれのほうが多いが、頻度に男女差がないこと、ST と RP のタスク間において、とぎれと延伸の生起の仕方に大きな違いがあることが分かった。

【ポスター発表】 9/5 (水) 14:20-15:35

〔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-4-02-E]

『日本語日常会話コーパス』モニター公開版の概要

小磯花絵 (国立国語研究所) ・ 天谷晴香 (国立国語研究所) ・ 居關友里子 (国立国語研究所)
白田泰如 (国立国語研究所) ・ 柏野和佳子 (国立国語研究所) ・ 川端良子 (国立国語研究所：学生)
田中弥生 (国立国語研究所：学生) ・ 西川賢哉 (国立国語研究所) ・ 伝康晴 (千葉大学／国立国語研究所)

国立国語研究所共同研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」では、200時間規模の日常会話を収めた『日本語日常会話コーパス』の構築を進めている。このコーパスは、多様な日常場面の会話を、映像まで含めて収録・公開するものであり、世界的に見ても新しい試みである。『日本語日常会話コーパス』の本公開は、プロジェクトの最終年度にあたる2021年度を予定しているが、コーパスの利用可能性や問題などを把握し今後の構築に活かすために、50時間のデータについて2018年度12月にモニター公開することを予定している。そこで本稿では、モニター公開データの仕様や特徴について報告する。
【ポスター発表】9/5 (水) 14:20-15:35

〔利用する言語資源〕 日常会話コーパス (CEJC)

[P-4-03-E]

日本語学習者属性別の言語行為の対話自動生成への適用に関する一考察

太田博三 (放送大学：学生)

最近、自然言語処理における対話システムや対話生成が注目されている。LINEなどのゆるい対話から、ビジネスでのコールセンターでの対話が、チャットボットの普及により身近なものになってコモディティ化している。しかし、実情としては、前者ではより自然な人間性な対話を、また後者では正確な対話応答が求められている。ここで、少し学術間のいいとこどりを行おうというのが本研究の目標である。具体的に、社会学のエスノメソドロジー・会話分析や談話分析における定性的な知見は有効である。また、自然言語処理や音声認識に基づく対話生成もビジネス上有益である。そこで、もう一度、国立国語研究所の日本語学習者会話データコーパスを用いて、相互行為をはじめとした効果検証を行った。この成果は対話破綻の定性的傾向や対話自動生成の技術に適用することで、実務の向上に資するものである。
【ポスター発表】9/5 (水) 14:20-15:35

〔利用する言語資源〕 UniDic, 国語研日本語ウェブコーパス (NWJC), NWJC2vec

[P-4-04-E]

『現日研・職場談話コーパス』中納言版公開データの作成

柏野和佳子 (国立国語研究所) ・ 大村舞 (国立国語研究所) ・ 西川賢哉 (国立国語研究所) ・ 小磯花絵 (国立国語研究所)

『現日研・職場談話コーパス』は、現代日本語研究会が作成した、首都圏の有職女性 19 名 (20 代～50 代) と、首都圏の有職男性 21 名 (20 代～50 代) の職場での自然談話を文字起こししたテキストを元に作成したコーパスである。その元となっている文字化テキストは、『合本 女性のことば・男性のことば (職場編)』(現代日本語研究会編, 2011 年, ひつじ書房) の付録 CD-ROM に収録されている。国立国語研究所に提供されたその文字化テキストを MeCab + UniDic で解析し、『現日研・職場談話コーパス』として公開する。本発表では、『現日研・職場談話コーパス』の概要と特徴を述べる

【ポスター発表】 9/5 (水) 14:20-15:35

〔利用する言語資源〕 UniDic, 現日研・職場談話コーパス

[P-4-05]

日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic

首藤公昭 (福岡大学名誉教授) ・ 田辺利文 (福岡大学) ・ 高橋雅仁 (久留米工業大学)

オノマトペの豊富さは日本語の特徴の一つとされるが、機械翻訳などの言語処理では十分な対応がなされていない。筆者らが開発した見出し数約 35,000 の、平仮名べた書き見出し、形態素分かち書き、構文機能、構文構造、内部修飾可否情報、文脈条件、連体・連用・動詞化情報などを与えた日本語オノマトペ共起表現 (コロケーション) レキシコン JMWEL_onomatopoeic の概要を報告する。

【ポスター発表】 9/5 (水) 14:20-15:35

〔利用する言語資源〕 日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic

[P-4-06]

語彙多様性指標の可視化と単回帰分析による TTR の補正

今田水穂 (文部科学省)

語彙多様性を評価する既存の指標には、延べ語数 N と異なり語数 $V(N)$ を入力とするもの、単語別の頻度を
入力とするもの、単語列を入力とするものなどがある。本発表では、これらの指標の特徴を整理し、「現代日本
語書き言葉均衡コーパス」(BCCWJ) を使用して指標値の分布を可視化する。 N と $V(N)$ を入力とする指標の
いくつかは、両者の間に冪乗則 $V(N) = aN^b$ を仮定している。TTR は $b=1$, R は $b=0.5$ として a を指標値と
して利用するが、1 では大きすぎ、0.5 では小さすぎる。そこで $V(N)$ と N の対数を単回帰分析して b の最適値
を推定し、TTR を補正することを考える。実際には冪乗則は成立しないため、この補正は近似的だが、比較的
簡単により補正を得ることができる。この補正值を他の指標と比較し、テキストサイズが指標値の平均やば
らつきに及ぼす影響を評価する。また、BCCWJ の 12 のサブコーパスについて b の値を推定し、一覧で示す。
[ポスター発表] 9/5 (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)

[P-4-07]

二字漢語を構成する漢字の造語力の変化

—『現代雑誌九十種の用語用字』データと『現代日本語書き言葉均衡コーパス』の比較を
通して—

本多由美子 (一橋大学 : 学生)

宮島 (1969) によると、明治時代に比べ現代では二字漢語を構成する漢字の意味は薄れており、漢語
の造語力が弱まった原因の 1 つであるという。このことを宮島 (1969) は『郵便報知』と『現代雑誌九十種』
のデータを用いて示した。本研究では『現代雑誌九十種』以降の傾向を捉えるために、宮島 (1969) と同様
の方法で『現代雑誌九十種』のデータと『現代日本語書き言葉均衡コーパス』の「新聞」「知恵袋」とを比
較した。比較の観点は二字漢語の構成漢字の「一字漢語の独立用例の有無」と「訓読みによる用例の有無」
である。その結果、(1) 『現代雑誌九十種』以降、「一字漢語」と「訓読み」の用例を有する漢字は減少傾向
にあり、(2) 『現代雑誌九十種』以降の「訓読み」の減少幅が大きいことが確認された。このことは、間接
的にはあるが二字漢語を構成する漢字の造語力が低下傾向にあることを示すものと考えられる。

[引用文献] 宮島達夫 (1969) 「近代日本語における漢語の位置」『教育国語』16, pp.17-44.

[ポスター発表] 9/5 (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), 『現代雑誌九十種の用語用字』

[P-4-08]

方言音声に対するテキスト自動アライメントの試み

石本祐一 (国立国語研究所)

音声コーパスの構築にあたり、音声に対する発話・音素・韻律などの各種ラベル付与が作業者の大きな負担となっている。この負担軽減を目的としてラベリングを自動化する試みが行われており、音声認識技術を利用した転記テキストの自動アライメントシステムがすでにコーパス構築の補助として稼働し始めている。しかし、システムの音声認識部を構成する音響モデル・言語モデルが標準語を基に設計されていることから、現在のところは標準語を主とした音声へのシステム利用にとどまっております。標準語とは異なる特性を持ちうる方言音声に対してはシステムの有効性が不明である。そこで本稿では、方言音声に対する転記テキストの自動アライメント性能について調べた結果について報告し、方言音声コーパスの構築におけるテキスト自動アライメントシステムの実用可能性について述べる。

【ポスター発表】 9/5 (水) 14:20-15:35

〔利用する言語資源〕 日本語諸方言コーパス (CJD)

[P-4-09]

単語の分散表現を用いた領域における出現単語の特徴分析

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)

自然言語処理では、コーパス中の単語に対して意味的な特徴をベクトルで表現し、様々な自然言語処理タスクにおいて利用することが多い。単語をベクトルで表現することにより、単語間の類似度を計算し、単語や文の違いなどの比較が可能である。これまでの研究では、単語ベクトルを生成するためには、ひとつの大規模な文書データから生成する必要があった。そのため、書籍・新聞・雑誌など、文書の分野による分散表現の比較や特徴分析は行われていなかった。すなわち、分野による単語の類似性や違いは明らかになっていない。そこで、本研究では領域ごとの単語ベクトル生成手法を提案し、各領域における単語の特徴分析を行う。書籍・雑誌・新聞の3つの領域の日本語コーパスに対し、指定した対象単語に対して単語ベクトルを作成する。対象単語のベクトルを用いて、対象単語が類似する単語を各領域において抽出し、単語の使用傾向などの分析を行う。実験の結果、同一単語であっても他領域で使われる意味と異なった語義で使用されている単語があることが分かった。加えて、動詞は領域からの依存度が低い、副詞は領域への依存度が高いなど品詞によって領域の依存度が異なるといった傾向があった。また、書籍領域では様々な種類の語句、新聞領域では政治関連語句、雑誌領域においてはカタカナ語が多く登場するなど、類似した単語には領域によって特徴がみられた。

【ポスター発表】 9/5 (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)

[P-4-10]

形容詞感動文における曖昧性回避の条件

西内沙恵 (国立国語研究所)

感動文では、分脈情報のほか、眼前の情景も意味の推測に大きく寄与するとされる。しかし、「(子供の表情を回想して)この顔のおかしかったこと。」のように、過去の事態も表現しえることから、対象が目の前に実在することは必須要素ではないことが明らかである。本研究は、「青っ! <*色が青い / **未熟だ>」や「くさい! <*嫌なにおいがする / **怪しい>」のような多義語による感動文の分析から、先行研究で規定される形式的な文法枠組みに語用論的記述を加えようとするものである。多義語は、共起語や文脈情報などの言語要素が意味選択の大きな手がかりとなる。言語的手がかりがない状態では、多義のいずれの意味も選択される可能性があるのにもかかわらず、眼前性も担保されない状態で出現し、意味の解釈が可能になるのはどのような条件によるのか。実例の観察から出現状況を分析し、意味の選択に語彙の特性と身体性がかかわっていることを論じる。

【ポスター発表】 9/5 (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), 日本語話し言葉コーパス (CSJ), 名大会話コーパス

[P-4-11]

コーパスを用いた授業における成果と課題

ーノンネイティブ日本語教師研修における授業を例にー

清水まさ子 (国際交流基金日本語国際センター) ・ 木田真理 (国際交流基金日本語国際センター)

筆者らは将来的に、ノンネイティブ日本語教師 (以下,NNT とする) がコーパスを用いて日本語を学んでいく授業を開発しようと考えている。本発表は、このような授業の開発にあたり、実際に NNT がコーパスを用いて興味のある語を調べた際に、彼らがどんな発見をし、どんな所でつまづくのか調べる、いわば授業開発のためのパイロットスタディである。本調査の結果,NNT はコーパスを使用することによって、対象語の新たな用法やレジスター、またはどんな時に使用するのかについての発見をしていた。しかし、そもそも検索対象語を正しく認識していない場合や、検索結果の解釈の誤りなど,NNT ならではの問題も確認された。本発表ではこのような調査の結果と考察をもとに,NNT がコーパスを使用する際に踏まえるべき点を示した。

【ポスター発表】 9/5 (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), 筑波ウェブコーパス

[P-4-12]

『日本語話し言葉コーパス (CSJ)』模擬講演における節頭フィラーの特徴

渡辺美知子 (国立国語研究所) ・ 是松優作 (東京大学：学生)

文境界、節境界にはフィラーがよく観察される。これは、文や節などの談話の切れ目では、その先の発話内容や表現を考えるのに時間が必要なためと考えられる。本研究では、文頭、節頭でのフィラーの使用に影響する要因を、節頭フィラー全体ならびに頻度の高い「エー」「アノ」「マー」について CSJ コア中の模擬講演を対象に調べた。考察対象とした要因は、1) 話者の性別、2) 年齢、3) 学歴、4) 講演経験、5) 直前の境界の種類、6) 節中語数、7) 節頭の接続詞の有無、の7つである。節頭フィラーの出現確率との間に有意な関連が見られたのは、要因1), 5), 6) のみであった。フィラーの出現確率は男性話者の方が女性話者よりも高く、強い節境界の方が文境界よりも高かった。頻度の高い3種類のフィラーを個別に見ると、エーのみ、文境界の方が強い節境界よりも出現確率が高かった。エーは、アノ、マーよりも、深い談話境界で用いられる傾向のあることが明らかになった。【ポスター発表】 9/5 (水) 14:20-15:35

〔利用する言語資源〕 日本語話し言葉コーパス

[P-4-13]

『国語研日本語ウェブコーパス』からの新規語彙素獲得の試み

岡照晃 (国立国語研究所)

『国語研日本語ウェブコーパス (NWJC)』は、国立国語研究所がこれまでに公開してきた『現代日本語書き言葉均衡コーパス』、『国語研日本語ウェブコーパス (NWJC)』は、国立国語研究所がこれまでに公開してきた『現代日本語書き言葉均衡コーパス (BCCWJ)』や『日本語話し言葉コーパス (CSJ)』と異なり、形態論情報をすべて形態素解析器『MeCab』と『解析用 UniDic』を使って自動付与している。『BCCWJ』や『CSJ』といった既存のコーパスの整備の際には、コーパスアノテーションと同時に、形態論情報のデータベースである『UniDic DB』に新規短単位語彙素を追加していた。そのためコーパス整備と同時に『UniDic DB』も拡張されてきたが、『NWJC』は全自動で構築されたため、新規短単位語彙素の検出とDBへの登録が行われておらず、その箇所での自動解析誤りを残したままとなっている。そこで本研究では、形態素解析を介さず、文字 N-gram の出現頻度と連接頻度の情報から文字 N-gram の分散表現を作成し、『NWJC』から『UniDic DB』に未登録の新規短単位語彙素の候補を列挙する方法について述べる。これによりDBのさらなる拡張がのぞめるだけでなく、『UniDic DB』のエクスポートデータで作成される『解析用 UniDic』も拡張されるため、それをを用いた再解析によって『NWJC』中の誤解析を減らすことにもつながる。

【ポスター発表】 9/5 (水) 14:20-15:35

〔利用する言語資源〕 国語研日本語ウェブコーパス (NWJC), UniDic

[O-4-01]

アクセント音調の諸相とその動態形式

佐藤大和 (東京外国語大学)

日本語の規範的なアクセントの型が、自発発話の中でどのような音調動態として実現されているかについて研究を進めている。本報告では、「日本語話し言葉コーパス (CSJ)」における東京方言話者 (女性) 1 名が発話した 11 分程度の独話音声を用い、約 520 のアクセント単位の分析から得られた内容について報告する。アクセント単位の音調は、大きく分類すると卓立型と非卓立型音調があり、同一アクセントであっても両者によって音調形式が異なっている。また、アクセントが知覚される拍 (アクセント拍と呼ぶ) におけるピッチの“上昇”, “下降”, “平坦”, “上昇下降”等の音調形式が、音響上のアクセント位置と関連すること、特にアクセント位置が後続拍にくる遅下がり現象と“上昇”ピッチが密接な関係にあることを示す。アクセントの遅下がりに関しては、強調的発話やフットリズムとの関連にも触れる。更に、アクセント拍以降の音調の降下特性とアクセント単位の音調動態モデルについても議論する。

【口頭発表】 9/5 (水) 15:45-16:10

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[O-4-02]

日本語複単語表現レキシコン JMWEL の概要

－動詞性表現を中心に－

首藤公昭 (福岡大学名誉教授) ・ 田辺利文 (福岡大学) ・ 高橋雅仁 (久留米工業大学)

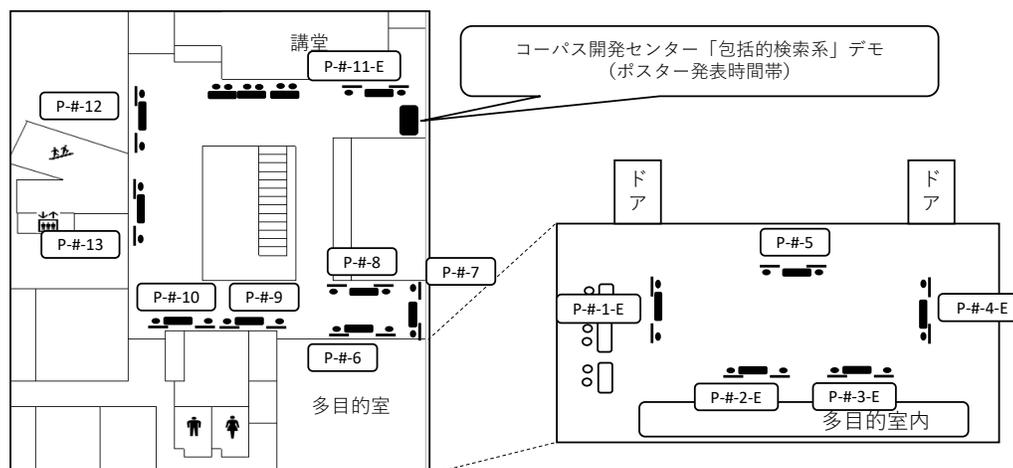
コロケーション, 決まり文句, 慣用句, 準慣用句などの定型表現とその派生表現約 140,000 見出しからなり、平仮名ベタ書き見出しのほかに、形態素分かち書き, 構文機能, 構文構造, 内部修飾可否情報, 文脈条件, 呼応情報, 語釈などを与えた日本語複単語表現レキシコン JMWEL の概要を動詞性表現を中心に報告する。(本論文は言語処理学会 2018 言語資源賞受賞論文に修正・加筆を施したものです。)

【口頭発表】 9/5 (水) 16:10-16:35

〔利用する言語資源〕 日本語複単語表現レキシコン

Information

ポスター設営図



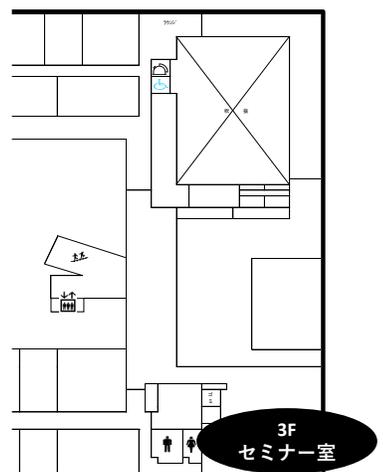
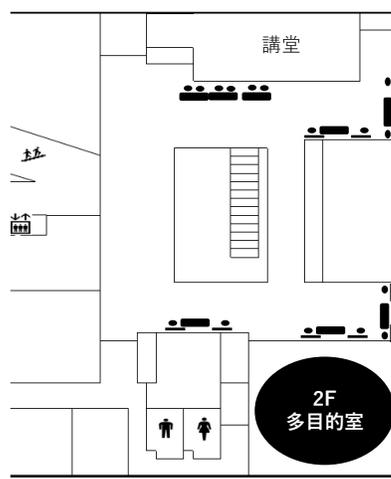
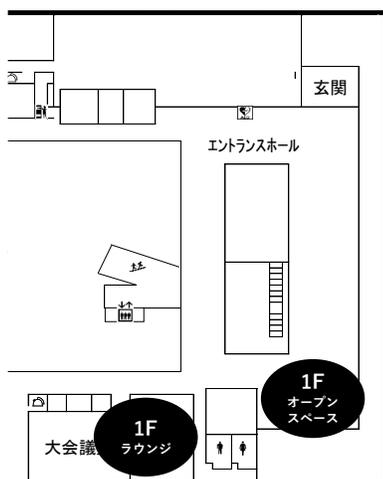
- P-#-1, P-#-2, P-#-3, P-#-4, P-#-11 は電源が利用できます。
- ポスター掲示用ついたての貼り付け可能面積は横 120cm× 縦 150cm です。
- ポスター掲示用資材は国語研で準備いたします。

コーパス開発センター「包括的検索系」デモ

国立国語研究所コーパス開発センターは「日本語言語資源の包括的高度共同利用環境の整備」プロジェクト (2016-2021 年) において、2021 年の公開を目指して「包括的検索系」の開発を進めております。本ワークショップではポスター発表時間帯に開発中の「包括的検索系」のデモを行います。

- 9月4日(火) 13:00-15:35 (ポスター発表1・2の時間帯)
- 9月5日(水) 13:00-15:35 (ポスター発表3・4の時間帯)

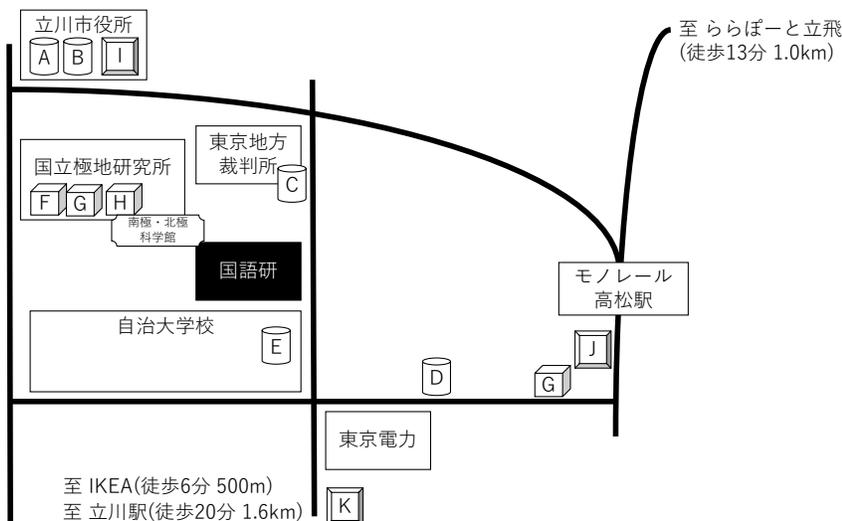
ランチスペース



- **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 4 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
- E 自治大大学校食堂 11:45-13:30
正門受付で名前・電話番号等を記入することにより利用できます

● お弁当

- F 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
- G ハイジカレー弁当 (極地研内 1F) 11:00-13:00
各種 500 円
- H 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

- I 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 2018 Abstract 集

発行日：平成 30 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

