

言語資源活用ワークショップ 2019

Abstract 集

2019年9月2日(月) - 4日(水)

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、アンケートにご協力ください。

- 2019年9月2日(月)～4日(水) 『言語資源活用ワークショップ2019』
アンケートフォーム <https://forms.gle/d2Pq7KmKigTKcMEv7>



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す。発表IDの末尾に“-S”がついている4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件に対し9月2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4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. 2016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. 2016年10月以降、著者と共著で発表したことがある。
- d. 2016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 2019

2019年9月2日(月)

- 10:30-10:40 ■挨拶 (2F 講堂) 前川喜久雄
- 10:40-11:55 ■口頭発表 グループ 1 (2F 講堂)
- [O-1-1-S]
文を包摂する名詞の形式的な特徴に関する考察
..... 泉大輔 (東京外国語大学: 学生)
- [O-1-2-S]
多義語のプロトタイプの意味認定における脱文脈化の活用—クラウドソーシングを用いた量的調査から—
..... 西内沙恵 (筑波大学: 学生)
- [O-1-3]
複数の脚本コーパスに現れた終助詞の比較分析
..... 松下晶子 (専修大学: 学生)
..... 丸山岳彦 (専修大学/国立国語研究所)
- 11:55-13:20 休憩
- 13:20-14:35 ■ポスター発表 グループ 1 (2F フロア・多目的室)
- [P-1-1-E]
自発発話におけるアクセント音調の動態分析
..... 佐藤大和 (東京外国語大学)
- [P-1-2-E]
共働き家庭の児童の LINE 使用—実例とフォローアップインタビューから—
..... 宮嶋由美 (明治大学)
- [P-1-3-S]
日本語学習者の助詞・動詞選択の傾向 —自動詞他動詞の比較を中心に—
..... 沖本与子 (一橋大学: 学生)
- [P-1-4-S]
「お/ご〜おき (下さい)」について
..... 井上直美 (埼玉大学: 学生)
- [P-1-5]
校歌の歌詞の言語的特徴に関する計量的研究 —滋賀県公立学校を対象として—
..... 陳曦 (立命館大学: 学生), 松本理美 (立命館大学: 学生)
..... 小椋秀樹 (立命館大学)

	[P-1-6]	発話の冒頭で使われる「まあ(ね)」について 加藤恵梨 (大手前大学)
	[P-1-7]	『現代日本語書き言葉均衡コーパス』における複合助詞の表記—動詞 の漢字表記に着目して— 本多由美子 (一橋大学: 学生)
	[P-1-8]	家庭での幼児の発話の修辞機能—脱文脈化の観点からの検討— 田中弥生 (神奈川大学/国立国語研究所) 小磯花絵 (国立国語研究所)
	[P-1-9]	明治・大正期の口語体会話文の位相差—語種率・品詞率を観点とし て— 近藤明日子 (国立国語研究所)
14:35-14:45		休憩 (ポスター切替)
14:45-16:00		■ポスター発表 グループ 2 (2F フロア・多目的室)
	[P-2-1-E]	「勿論」考 高橋圭子 (明治大学), 東泉裕子 (明治大学)
	[P-2-2-E]	地図課題対話における参照導入方法の特徴 川端良子 (国立国語研究所)
	[P-2-5]	漢字の形における統計則 太田守洋 (琉球大学: 学生), 山本健 (琉球大学)
	[P-2-6]	『現代日本語書き言葉均衡コーパス』書籍サンプルの NDC 情報増補 加藤祥 (国立国語研究所) 森山奈々美 (津田塾大学/国立国語研究所: 学生) 浅原正幸 (国立国語研究所)
	[P-2-7]	実践医療用語の語構成要素への分割と意味ラベル付与の試み 山崎誠 (国立国語研究所), 相良かおる (西南女学院大学) 小野正子 (西南女学院大学), 東条佳奈 (目白大学) 麻子軒 (大阪大学)
	[P-2-8]	文書領域情報を有する BERT の階層位置に関する考察

..... 欧陽恵子（茨城大学：学生），田中裕隆（茨城大学：学生）
..... 曹鋭（茨城大学：学生），白静（茨城大学：学生）
..... 馬ブン（茨城大学：学生），新納浩幸（茨城大学）

[P-2-9]

データ引用による言語資源活用文献の把握の可能性：BCCWJ の分析から

... 中渡瀬秀一（国立情報学研究所），加藤文彦（国立情報学研究所）
..... 大向一輝（国立情報学研究所）

16:00-16:10

休憩

16:10-17:00

■口頭発表 グループ 2 (2F 講堂)

[O-2-1]

All-words WSD と fine-tuning を利用した分類語彙表の語義の分散表現の構築

..... 柳沼大輝（茨城大学），古宮嘉那子（茨城大学）
..... 新納浩幸（茨城大学）

[O-2-2]

UniDic 非コアデータ—解析用 UniDic の ID 情報にひも付く追加情報の公開について—

..... 岡照晃（国立国語研究所）

2019年9月3日(火)

- 10:30-12:10 ■口頭発表 グループ3 (2F 講堂)
- [O-3-1]
配慮表現コーパス作成と分析手法の小考察 対人関係を考慮した対話システムの構築に向けて
..... 太田博三 (放送大学：学生)
- [O-3-2]
コーパスを用いて「は」と「が」に関する三上説を検証する試み
..... 庵功雄 (一橋大学)
- [O-3-3]
『日本語話し言葉コーパス』への声質情報付与と予備的分析
..... 前川喜久雄 (国立国語研究所), 西川賢哉 (国立国語研究所)
- [O-3-4]
類義副詞の文体を測る試み — 「まったく」・「ぜんぜん」・「すこしも」・「ちっとも」を例に—
..... 劉時珍 (東洋大学)
- 12:10-13:30 休憩
- 13:30-14:45 ■ポスター発表 グループ3 (2F フロア・多目的室)
- [P-3-1-E]
政策ニュース映画における否定的な表現の考察—戦後の社会課題と行政施策の可視化の試み—
..... 春木良且 (フェリス女学院)
..... 田中弥生 (神奈川大学/国立国語研究所)
- [P-3-2-E]
『日本語日常会話コーパス』の短単位解析：作業工程を中心に
..... 西川賢哉 (国立国語研究所), 渡邊友香 (国立国語研究所)
- [P-3-3]
『BTSJ 日本語自然会話コーパス』における反論ストラテジーの分析
..... 陳朝陽 (湖北第二師範学院/国立国語研究所)
..... 宇佐美まゆみ (国立国語研究所)
- [P-3-4]
手本として示される実習記録の語彙の特徴の分析
..... 山元一晃 (国際医療福祉大学), 浅川翔子 (慶應義塾大学)
- [P-3-5]
BERT を利用した教師あり学習による語義曖昧性解消
..... 曹鋭 (茨城大学：学生), 田中裕隆 (茨城大学：学生)
..... 白静 (茨城大学：学生), 馬ブン (茨城大学：学生)
..... 新納浩幸 (茨城大学)

	[P-3-6]	子どもによるやり取りへの参与の振る舞い—両親との会話の事例分析から— 居關友里子（国立国語研究所），小磯花絵（国立国語研究所）
	[P-3-7]	半教師あり語義曖昧性解消における各ジャンルの語義なし用例文の利用 谷田部梨恵（茨城大学：学生），佐々木稔（茨城大学）
	[P-3-8]	発話文表現文型辞書の設計と編纂 夏目和子（名古屋大学），佐藤理史（名古屋大学）
	[P-3-9]	BCCWJ 小説会話文への話者情報の付与とその活用 山崎誠（国立国語研究所），柏野和佳子（国立国語研究所） 宮嶋由美（明治大学）
14:45–14:55		休憩
14:55–16:10		■ポスター発表 グループ 4 (2F フロア・多目的室)
	[P-4-1-E]	国会会議録における言語表現の出現頻度に関する時間的変化モデルの検証 山口昌也（国立国語研究所）
	[P-4-2-E]	美容院におけるマルチアクティビティ：鏡越しの視線と発話 天谷晴香（国立国語研究所）
	[P-4-3]	『分類語彙表』と『岩波国語辞典第五版タグ付きコーパス 2004』の対応表 呉佩珣（筑波大学），近藤森音（東京大学） 森山奈々美（津田塾大学/国立国語研究所） 荻原亜彩美（津田塾大学/国立国語研究所） 加藤祥（国立国語研究所），浅原正幸（国立国語研究所）
	[P-4-4]	BERT を利用した単語用例のクラスタリング 馬ブン（茨城大学：学生），田中裕隆（茨城大学：学生） 曹銳（茨城大学：学生），白静（茨城大学：学生） 新納浩幸（茨城大学）
	[P-4-5]	

「日本語非母語話者の読解コーパス」から見える非漢字圏日本語学習者の辞書使用

..... クリスティナ・フメリヤク寒川（リュブリャナ大学）

[P-4-6]

節中のフィラー「エー」「アノー」「マー」の出現確率に関する要因

..... 渡辺美知子（国立国語研究所），白旗悠真（東京大学：学生）

[P-4-7]

『日本語日常会話コーパス』モニター公開版に見られる応答表現

..... 柏野和佳子（国立国語研究所）

[P-4-8]

日常会話音声に対する基本周波数推定の課題

..... 石本祐一（国立国語研究所）

2019年9月4日(水)

- 10:30-11:20 ■口頭発表 グループ4 (2F 講堂)
[O-4-1]
『日本語日常会話コーパス』モニター公開版：研究の可能性
.....小磯花絵 (国立国語研究所)
[O-4-2]
『通時音声コーパス』の可能性と問題点 — 『昭和話し言葉コーパス』
の構築と分析—
.....丸山岳彦 (専修大学/国立国語研究所)
- 11:20-11:45 ■クロージング (2F 講堂)
11:45-13:00 休憩
- Universal Dependencies セッション (2F 講堂)
- 13:00-13:30 趣旨説明 Universal Dependencies の概要
.....浅原正幸 (国立国語研究所)
- 13:30-14:10 Universal Dependencies における日本語
.....金山博 (日本 IBM)
- 14:10-14:20 休憩
- 14:20-15:00 Universal Dependencies に基づく言語学研究的射程
.....伊藤薫 (九州大学)
- 15:00-15:40 統語的語の認定問題
.....村脇有吾 (京都大学)
- 15:40-15:50 休憩
- 15:50-16:20 現代日本語書き言葉均衡コーパスの Universal Dependencies —
UD Japanese コーパスの一例—
.....大村舞 (国立国語研究所)
- 16:20-17:00 GiNZA で始める日本語依存構造解析 — CaboCha, UDPipe,
StanfordNLP との比較
.....松田寛 (Megagon)

目次

Abstract	13
文を包摂する名詞の形式的な特徴に関する考察 [O-1-1-S]	
泉大輔（東京外国語大学：学生）	14
多義語のプロトタイプの意味認定における脱文脈化の活用—クラウドソーシングを用いた量的調査から— [O-1-2-S]	
西内沙恵（筑波大学：学生）	14
複数の脚本コーパスに現れた終助詞の比較分析 [O-1-3]	
松下晶子（専修大学：学生）・丸山岳彦（専修大学/国立国語研究所）	15
自発発話におけるアクセント音調の動態分析 [P-1-1-E]	
佐藤大和（東京外国語大学）	15
共働き家庭の児童のLINE使用—実例とフォローアップインタビューから— [P-1-2-E]	
宮嶋由美（明治大学）	16
日本語学習者の助詞・動詞選択の傾向 —自動詞他動詞の比較を中心に— [P-1-3-S]	
沖本与子（一橋大学：学生）	16
「お／ご～おき（下さい）」について [P-1-4-S]	
井上直美（埼玉大学：学生）	17
校歌の歌詞の言語的特徴に関する計量的研究 — 滋賀県公立学校を対象として — [P-1-5]	
陳曦（立命館大学：学生）・松本理美（立命館大学：学生）・小椋秀樹（立命館大学）	17
発話の冒頭で使われる「まあ（ね）」について [P-1-6]	
加藤恵梨（大手前大学）	18
『現代日本語書き言葉均衡コーパス』における複合助詞の表記 —動詞の漢字表記に着目して— [P-1-7]	
本多由美子（一橋大学：学生）	18
家庭での幼児の発話の修辞機能—脱文脈化の観点からの検討— [P-1-8]	
田中弥生（神奈川大学/国立国語研究所）・小磯花絵（国立国語研究所）	19
明治・大正期の口語体会話文の位相差 —語種率・品詞率を観点として— [P-1-9]	
近藤明日子（国立国語研究所）	19
「勿論」考 [P-2-1-E]	
高橋圭子（明治大学）・東泉裕子（明治大学）	20
地図課題対話における参照導入方法の特徴 [P-2-2-E]	
川端良子（国立国語研究所）	20
漢字の形における統計則 [P-2-5]	
太田守洋（琉球大学：学生）・山本健（琉球大学）	21

『現代日本語書き言葉均衡コーパス』書籍サンプルの NDC 情報増補	[P-2-6]	
加藤祥 (国立国語研究所)・森山奈々美 (津田塾大学/国立国語研究所：学生)		
浅原正幸 (国立国語研究所)		21
実践医療用語の語構成要素への分割と意味ラベル付与の試み	[P-2-7]	
山崎誠 (国立国語研究所)・相良かおる (西南女学院大学)・小野正子 (西南女学院大学)		
東条佳奈 (目白大学)・麻子軒 (大阪大学)		22
文書領域情報を有する BERT の階層位置に関する考察	[P-2-8]	
欧陽恵子 (茨城大学：学生)・田中裕隆 (茨城大学：学生)・曹鋭 (茨城大学：学生)		
白静 (茨城大学：学生)・馬ブン (茨城大学：学生)・新納浩幸 (茨城大学)		22
データ引用による言語資源活用文献の把握の可能性: BCCWJ の分析から	[P-2-9]	
中渡瀬秀一 (国立情報学研究所)・加藤文彦 (国立情報学研究所)		
大向一輝 (国立情報学研究所)		23
All-words WSD と fine-tuning を利用した分類語彙表の語義の分散表現の構築	[O-2-1]	
柳沼大輝 (茨城大学)・古宮嘉那子 (茨城大学)・新納浩幸 (茨城大学)		23
UniDic 非コアデータ解析用 UniDic の ID 情報にひも付く追加情報の公開について—	[O-2-2]	
岡照晃 (国立国語研究所)		24
配慮表現コーパス作成と分析手法の小考察 対人関係を考慮した対話システムの構築に向けて	[O-3-1]	
太田博三 (放送大学：学生)		24
コーパスを用いて「は」と「が」に関する三上説を検証する試み	[O-3-2]	
庵功雄 (一橋大学)		25
『日本語話し言葉コーパス』への声質情報付与と予備的分析	[O-3-3]	
前川喜久雄 (国立国語研究所)・西川賢哉 (国立国語研究所)		25
類義副詞の文体を測る試み — 「まったく」・「ぜんぜん」・「すこしも」・「ちっとも」を例に—	[O-3-4]	
劉時珍 (東洋大学)		26
政策ニュース映画における否定的な表現の考察—戦後の社会課題と行政施策の可視化の試み—	[P-3-1-E]	
春木良且 (フェリス女学院)・田中弥生 (神奈川大学/国立国語研究所)		26
『日本語日常会話コーパス』の短単位解析：作業工程を中心に	[P-3-2-E]	
西川賢哉 (国立国語研究所)・渡邊友香 (国立国語研究所)		27
『BTSJ 日本語自然会話コーパス』における反論ストラテジーの分析	[P-3-3]	
陳朝陽 (湖北第二師範学院/国立国語研究所)・宇佐美まゆみ (国立国語研究所)		27
手本として示される実習記録の語彙の特徴の分析	[P-3-4]	
山元一晃 (国際医療福祉大学)・浅川翔子 (慶應義塾大学)		28

BERT を利用した教師あり学習による語義曖昧性解消	[P-3-5]	
曹鋭 (茨城大学：学生)・田中裕隆 (茨城大学：学生)・白静 (茨城大学：学生)		
馬ブン (茨城大学：学生)・新納浩幸 (茨城大学)		28
子どもによるやり取りへの参与の振る舞い—両親との会話の事例分析から—	[P-3-6]	
居關友里子 (国立国語研究所)・小磯花絵 (国立国語研究所)		29
半教師あり語義曖昧性解消における各ジャンルの語義なし用例文の利用	[P-3-7]	
谷田部梨恵 (茨城大学：学生)・佐々木稔 (茨城大学)		29
発話文表現文型辞書の設計と編纂	[P-3-8]	
夏目和子 (名古屋大学)・佐藤理史 (名古屋大学)		30
BCCWJ 小説会話文への話者情報の付与とその活用	[P-3-9]	
山崎誠 (国立国語研究所)・柏野和佳子 (国立国語研究所)・宮崎由美 (明治大学)		30
国会会議録における言語表現の出現頻度に関する時間的変化モデルの検証	[P-4-1-E]	
山口昌也 (国立国語研究所)		31
美容院におけるマルチアクティビティ：鏡越しの視線と発話	[P-4-2-E]	
天谷晴香 (国立国語研究所)		31
『分類語彙表』と『岩波国語辞典第五版タグ付きコーパス 2004』の対応表	[P-4-3]	
呉佩珣 (筑波大学)・近藤森音 (東京大学)・森山奈々美 (津田塾大学/国立国語研究所)		
荻原亜彩美 (津田塾大学/国立国語研究所)・加藤祥 (国立国語研究所)		
浅原正幸 (国立国語研究所)		32
BERT を利用した単語用例のクラスタリング	[P-4-4]	
馬ブン (茨城大学：学生)・田中裕隆 (茨城大学：学生)・曹鋭 (茨城大学：学生)		
白静 (茨城大学：学生)・新納浩幸 (茨城大学)		32
「日本語非母語話者の読解コーパス」から見える非漢字圏日本語学習者の辞書使用	[P-4-5]	
クリスティナ・フメリヤク寒川 (リュブリャナ大学)		33
節中のフィラー「エー」「アノー」「マー」の出現確率に関する要因	[P-4-6]	
渡辺美知子 (国立国語研究所)・白旗悠真 (東京大学：学生)		33
『日本語日常会話コーパス』モニター公開版に見られる応答表現	[P-4-7]	
柏野和佳子 (国立国語研究所)		34
日常会話音声に対する基本周波数推定の課題	[P-4-8]	
石本祐一 (国立国語研究所)		34
『日本語日常会話コーパス』モニター公開版：研究の可能性	[O-4-1]	
小磯花絵 (国立国語研究所)		35
『通時音声コーパス』の可能性と問題点 —『昭和話し言葉コーパス』の構築と分析—	[O-4-2]	
丸山岳彦 (専修大学/国立国語研究所)		35
Information		36
ポスター設営図		37
ランチスペース		38
ランチマップ		39

Abstract

[O-1-1-S]

文を包摂する名詞の形式的な特徴に関する考察

泉大輔（東京外国語大学：学生）

本稿はモダリティ形式を含む文が名詞に直接先行する形式（「母さん助けて詐欺」「早く帰れオーラ」など）について、ウェブコーパスを用いて事例を収集し、その出現状況および形式的な特徴を明らかにすることを目的とする。本稿で明らかになった形式的な特徴は、①当該の形式を形成する名詞は、抽象的な概念（活動、言語、様相、程度性、手法など）を表す名詞であり、和語よりも造語力の高い漢語または外来語が多いこと、②名詞に先行する文には主に心情や印象が述べられた発話形式が多く、その発話形式は後部要素である抽象概念の内実を言語化したものであること、③文末のモダリティ形式によって後続しやすい名詞には差があること、④当該の形式は一般的な複合名詞の持つ意味構造と同様に、後部要素が類概念を表し、前部要素の発話形式がその類概念を特徴づける種差として機能するという前後の意味関係を持つことである。

【口頭発表】9/2（月）10:40–11:05

〔利用する言語資源〕 国語研日本語ウェブコーパス (NWJC), 現代日本語書き言葉均衡コーパス (BCCWJ), 日本語日常会話コーパス (CEJC)

[O-1-2-S]

多義語のプロトタイプの意味認定における脱文脈化の活用—クラウドソーシングを用いた量的調査から—

西内沙恵（筑波大学：学生）

多義語のプロトタイプの意味の認定には、意味的出現の高頻度・想起の容易さ・用法上の制約の少なさ・歴史的出現の順序・習得段階など、様々な手法が提起されている。本研究では、意味の移り変わりを前提とした、再調査可能なデザインの量的調査による認定手法を提案する。調査では、多義的形容詞の実例と脱文脈化した語の類似度を調べ、その結果に基づいてプロトタイプの意味の認定を行う。発表では、この手法の妥当性を多角的に検討する。

【口頭発表】9/2（月）11:05–11:30

〔利用する言語資源〕 分類語彙表, 国語研日本語ウェブコーパス (NWJC), 現代日本語書き言葉均衡コーパス (BCCWJ), 分類語彙表番号付与済み『現代日本語書き言葉均衡コーパス』

[O-1-3]

複数の脚本コーパスに現れた終助詞の比較分析

松下晶子（専修大学：学生）・丸山岳彦（専修大学/国立国語研究所）

本稿では、脚本のテキストデータから構築したコーパスを分析した結果について述べる。今回、既に構築済みの『市川森一脚本コーパス』に加え、新たに『鎌田敏夫脚本コーパス』を構築した。これにより、違う書き手によって書かれた脚本の比較が可能となった。松下・丸山(2018)および松下(2019)では、『市川森一脚本コーパス』において主に終助詞に焦点を当て、分析を行い、性差が表れやすい終助詞の多用や、場面によって終助詞の表記が変わるといった点を明らかにした。今回、『鎌田敏夫脚本コーパス』に現れた終助詞の分布を分析し、『市川森一脚本コーパス』の終助詞の分布と比較した。その結果、2人の脚本家の間で終助詞の使用傾向が異なることがわかった。さらに実際の会話を収録した『日本語日常会話コーパス (CEJC)』で用いられる終助詞との比較を行ったところ、「話すことを前提として書かれた言葉」と実際の話し言葉では終助詞の分布が異なることがわかった。

【口頭発表】9/2 (月) 11:30–11:55

〔利用する言語資源〕 日本語日常会話コーパス (CEJC), 脚本コーパス

[P-1-1-E]

自発発話におけるアクセント音調の動態分析

佐藤大和（東京外国語大学）

本報告は、「日本語話し言葉コーパス (CSJ)」の東京方言話者2名（男性・女性）による独話音声の音調分析により、規範的なアクセントの型がどのような音調動態として実現されているかについて検討したものである。著者はこれまで主にアクセント拍近傍の音調特性に着目した内容を報告してきたが、今回は(1) 1~4アクセントを含む発話のまとまりの包括的音調特性の分析、(2) アクセント句の音調動態モデルに基づいた分析、等に留意した。特に後者では、アクセントの動態核は原則2拍とし、その“上昇-下降”、“平坦-下降”、“下降-下降”、“上昇-上昇”等の動態形式を分析する。また動態核に続いてアクセント単位の終結部 (Coda) を設定し、アクセントに直接関わる音調下降とアクセント単位の終結下降とを区別する必要性のあることを示す。さらに音調の急峻な下降や上昇で観測されるピッチの変曲点（音調制御開始点）は母音 onset 近傍で生起し、アクセント位置など発話の動的音調特性と関連することを述べる。

【ポスター発表】9/2 (月) 13:20–14:35

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[P-1-2-E]

共働き家庭の児童の LINE 使用—実例とフォローアップインタビューから—

宮崎由美 (明治大学)

本稿では、筆者が現在構築中の『M-ZAK LINE データベース』に収録されている小学 1 年生、小学 4 年生の兄弟とその母親との LINE データをもとに、LINE における小学生の言語活動と、その内容を分析する。分析にあたっては、児童の生活環境を詳しく聞き取り、なぜ LINE が利用されるのかという背景とともに、決して希薄化するばかりではない現代のコミュニケーションのあり方を、長男、次男、そして母親の「感情を表す表現」とともに考察を行った。

【ポスター発表】 9/2 (月) 13:20-14:35

〔利用する言語資源〕 M-ZAK LINE データベース

[P-1-3-S]

日本語学習者の助詞・動詞選択の傾向 —自動詞他動詞の比較を中心に—

沖本与子 (一橋大学：学生)

本研究は、助詞と動詞を組み合わせた問題文における学習者の助詞選択と動詞選択の傾向を自動詞他動詞の比較を中心に分析し、日本語学習者の回答特徴を把握することを目的とする。使用した言語資源は、松下 (2011)「日本語を読むための語彙データベース (VDRJ)」と日本語学習辞書支援グループ (2015)「日本語教育語彙表 Ver 1.0」である。この言語資源から動詞を抽出しテストを作成し、初級前半終了、中級後半終了の日本語学習者から延べ 137 名の回答データを収集した。回答データ分析の結果、各テストの平均正答率では大きな差は見られなかった。また、初級レベルはひらがなのみで出題したため、非漢字圏学習者の平均点が高い問題があり、中級レベルは漢字を使用し出題したため、全体的に漢字圏学習者の平均点が高かった。なお、対のある自他動詞の問題において、「を」を「が」と誤回答する率が高く、同様に、全ての動詞において自動詞より他動詞の方が回答において正答率が高かったことが分かった。日本語学習者の回答傾向を分析することで、特に自他動詞の教育実践への応用が期待できる。

【ポスター発表】 9/2 (月) 13:20-14:35

〔利用する言語資源〕 日本語を読むための語彙データベース (VDRJ) Ver. 1.1, 日本語教育語彙表 Ver 1.0

[P-1-4-S]

「お／ご～おき（下さい）」について

井上直美（埼玉大学：学生）

「ご承知おき下さい」などの「お／ご～おき下さい」という表現は、日本語教育の教材類に詳しい解説がなされていない項目である。「ご承知下さい」であれば問題ないが、「おき」が付加されたこの表現は、上級以上の日本語学習者が気になる学習ポイントだとして先行研究で指摘されている（劉 2015）。そこで、本研究では「お／ご～おき下さい」、および「お／ご～おきいただく」、「お／ご～おき願う」、「お／ご～おきのほど」等の関連する表現も考察対象とし、『現代日本語書き言葉均衡コーパス』を用いて用例調査を行い、これらのふるまいに注目した。その結果、「お／ご～おき」という表現は、ジャンル別では、国会議事録や自治体の広報で出現が顕著であること、共起する動詞の特徴としては、情報の認知（知覚・思考・記憶）を表す動詞（承知する・含む・見知る等）と出現すること、そして慣用的な表現として用いられていることが明らかになった。

【ポスター発表】9/2（月）13:20-14:35

〔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-1-5]

校歌の歌詞の言語的特徴に関する計量的研究 — 滋賀県公立学校を対象として —

陳曦（立命館大学：学生）・松本理美（立命館大学：学生）・小椋秀樹（立命館大学）

校歌は、その学校を象徴するものであり、校風、所在地の地理的特徴などが歌詞に歌われることが多い。式典で歌うなど、児童・生徒にとって身近なものでもある。しかし校歌の歌詞の言語的特徴について分析した研究は少なく、いまだ十分に明らかにされているとは言い難い。そこで筆者らは、滋賀県の公立小中高を対象に校歌の歌詞を各校 Web ページから収集し、コーパスを構築した。このコーパスを基に、MVR、受身形、動詞「V + あう」、連体修飾節といった観点から歌詞の言語的特徴について多角的な分析を行った。調査の結果、次のことが明らかとなった。(1) 名詞比率と MVR とによる分析から、小学校は「ありさま描写」的な歌詞が多く、中学校・高校になると「動き描写」的な歌詞になる傾向が見られる。(2) 受身形・動詞「V + あう」（「育まれる」「助け合う」等）については、小学校の歌詞に多く見られるが、中学校・高校になると減少する傾向が見られる。(3) 連体修飾節は小・中・高校の教科書（書き言葉）とは異なって補足語修飾節に集中する傾向が見られ、また主名詞は学年上昇とともに生徒や学校を表す名詞が減少し、自然や徳目を表す名詞が増加する傾向が見られる。

【ポスター発表】9/2（月）13:20-14:35

〔利用する言語資源〕 UniDic

[P-1-6]

発話の冒頭で使われる「まあ（ね）」について

加藤恵梨（大手前大学）

「まあ」は多くの用法や機能を有することが先行研究で指摘されている。本研究では、『名大会話コーパス』を用い、先行研究において「応答型用法」と呼ばれる、発話の冒頭で使われる「まあ（ね）」の用法および機能を調査・分析することを目的とする。分析の結果、「まあ（ね）」は、話し手の意見や質問に対して聞き手ははっきりとした意見を述べず、会話をつなぐために使うことが多いことが分かった。また、話し手の発言に対して聞き手が「まあ（ね）」と答えた場合、話し手は今話している話題を早めに切り上げ、別の話題へ転換していることが多かった。このことから、聞き手が「まあ（ね）」と応答した場合、話し手は話をさらに続けることが難しくなったり、聞き手が自身の発言に対してあまり興味を抱いていないと感じたりすることがあること等を指摘する。

【ポスター発表】 9/2 (月) 13:20-14:35

〔利用する言語資源〕 名大会話コーパス

[P-1-7]

『現代日本語書き言葉均衡コーパス』における複合助詞の表記 — 動詞の漢字表記に着目して —

本多由美子（一橋大学：学生）

「にあたって」や「にしたがって」などの動詞を含む複合助詞は、動詞の意味が形式化した表現である。本動詞としての意味が失われていれば、動詞は仮名で表記される傾向が強くなると予想されるが、実際には「に当たって」や「に従って」という漢字を用いた表記も見られ、その傾向は明らかではない。そこで本発表では『現代日本語書き言葉均衡コーパス』を用い、複合助詞 12 形式について複合助詞と同じ形式の用例を抽出し、複合助詞と本動詞の意味に分けて、表記を調査した。その結果、和語の動詞を含む複合助詞は仮名で表記される傾向が強く、「に関して」「に対して」のような漢語の動詞を含む形式は漢字で表記される割合が高いことが観察された。また、「をもって」は前に接する語の意味が表記の違いの要因になることが示唆されたが、「にあたって」は前に接する語および係り先の語の種類と、表記とに有意な差は見られなかった。

【ポスター発表】 9/2 (月) 13:20-14:35

〔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-1-8]

家庭での幼児の発話の修辞機能—脱文脈化の観点からの検討—

田中弥生（神奈川大学/国立国語研究所）・小磯花絵（国立国語研究所）

本研究は、幼児の言語コミュニケーションに関して、脱文脈化の観点から検討するものである。ここで脱文脈化とは、コミュニケーションが行われている時空と、その発話内容との、時間的・空間的距離の程度のことをさす。本発表では、国立国語研究所の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」において現在構築中の「日本語日常会話コーパス」モニター公開版データの中から、幼児と両親の自宅での食事場面での会話を分析の対象とし、幼児の発話について言語表現から修辞機能を特定し、親とのやりとりで幼児がどのように発話を展開させるか、脱文脈化の観点からの検討を行った。その結果、食事場面であるため、食事の話題や食卓に置いてあるものについての脱文脈化程度の低い修辞機能が多く使われるが、両親の発話をきっかけとして幼児も脱文脈化程度の高い修辞機能を使用していることがあきらかになった。

【ポスター発表】9/2（月）13:20-14:35

〔利用する言語資源〕日本語日常会話コーパス（CEJC）

[P-1-9]

明治・大正期の口語体会話文の位相差 —語種率・品詞率を観点として—

近藤明日子（国立国語研究所）

明治・大正期の話し言葉における語彙の位相差の実態を明らかにすることを目的として、国立国語研究所（2019）『日本語歴史コーパス 明治・大正編 I 雑誌』（短単位データ 1.2）から抽出した雑誌『太陽』の小説・戯曲中の口語体会話文を話者の性別・社会階層による計4種の話者属性により分類し、語種率・品詞率の観点から話者属性による位相差を分析・考察した。その結果、漢語率は男性>女性、知識層>非知識層の位相差が認められ、知識層男性が最も高く、非知識層女性が最も低く、その中間に知識層女性・非知識層男性が位置することが明らかになった。また、品詞率では知識層男性が他の話者属性から卓立して特徴的な傾向を有し、それは知識層男性の会話文の専門性の高さ、複雑さ、論理性の高さを示す可能性があることが明らかになった。また、語種率・品詞率の通時的変化を概観した結果、非知識層男性に多くの通時的変化が見られることが明らかになった。

【ポスター発表】9/2（月）13:20-14:35

〔利用する言語資源〕日本語歴史コーパス（CHJ）

[P-2-1-E]

「勿論」考

高橋圭子（明治大学）・東泉裕子（明治大学）

現代語の「もちろん」は「論ずる（こと）勿（なか）れ」という禁止表現から発生したと説明されることがある。近代以前のデータベースを検索すると、古代の六国史に代表される漢文体の文献では、「勿論」の用例は「論ずる（こと）勿れ／勿（な）し」という意味であり、否定辞「勿」と動詞「論」から構成される句であった。現代語とほぼ同様の意味の「勿論」の語の用例は、中世の古記録や『愚管抄』『沙石集』など和漢混交文体による仏教関連の文献から見られるようになる。用法は文末における名詞述語が主であった。近世には、ジャンルも文体も多様な文献に用いられ、文中や文頭における副詞用法や応答詞的用法も出現する。古代の禁止表現と中世以降の「勿論」の関連は不明だが、日本語のみならず中国語・韓国語においても漢字語「勿論」の研究が進められ、さまざまな知見が見出されている。通言語的な議論の深化が期待される。

【ポスター発表】9/2（月）14:45-16:00

〔利用する言語資源〕 六国史, 撰関期古記録データベース, 古記録フルテキストデータベース, 岩波日本古典文学大系本文データベース

[P-2-2-E]

地図課題対話における参照導入方法の特徴

川端良子（国立国語研究所）

対話で特定の対象が参照されるとき、話し手が発話によって指した対象と聞き手が想定した対象が同じであることはコミュニケーションの成功のために不可欠である。したがって、話し手は参照を行う際、聞き手との共有知識を想定し、聞き手が同じ対象を特定するように発話を計画する。しかし、参照対象に関する対話相手の知識が不確かな場合もあり、そのときは発話の計画が難しくなる。なぜなら、相手が知っている対象を知らないものとして対話に導入することは対話の状況によっては相手に失礼になることもあるし、逆に知らない対象を知ってるものとして対話を進行することでコミュニケーション上のトラブルが生じる可能性があるからである。本研究は、参照対象に関する聞き手の知識が不確実な状況で、話し手は実際にどのように対象を対話に導入するのか、その形式を『日本語地図課題対話コーパス』を題材に類型化し、それぞれのパターンが選択される要因について検討する。

【ポスター発表】9/2（月）14:45-16:00

〔利用する言語資源〕 日本語地図課題対話コーパス

[P-2-5]

漢字の形における統計則

太田守洋（琉球大学：学生）・山本健（琉球大学）

本研究では漢字の形を統計的に分析する。漢字のサイズを特徴づける基本的な指標として画数があり、画数が多い漢字ほど複雑な形である傾向がある。一方、漢字を構成する線の長さが形の複雑性を表すとみなすことができる。本研究では、漢字の線長をコンピュータのフォントを用いて計測し、線長と画数の関係を調べた。その結果、漢字の線長は画数に対しておおむねベキ乗則にしたがって増加することがわかった。さらに、フラクタル図形を基にした数学的なモデルを導入し、ベキ乗則の指数とフラクタル次元の関係を理論的に導出した。この関係を漢字のベキ指数に適用すると、フラクタル次元はおよそ2次元となった。すなわち、漢字の形は画数の増加とともに平面充満的に複雑化するといえる。

【ポスター発表】9/2（月）14:45-16:00

〔利用する言語資源〕 常用漢字オープンデータ (Joyo_Kanji), JIS 第1水準及び第2水準の画数データ (漢字辞典オンライン <https://kanji.jitenon.jp> より), MS ゴシック (フォント), 源ノ角ゴシック (フォント)

[P-2-6]

『現代日本語書き言葉均衡コーパス』書籍サンプルのNDC情報増補

加藤祥（国立国語研究所）・森山奈々美（津田塾大学/国立国語研究所：学生）

浅原正幸（国立国語研究所）

『現代日本語書き言葉均衡コーパス』の書籍サンプル（PB（出版）10,117 サンプル・LB（図書館）10,551 サンプル・OB（ベストセラー）1,390 サンプル）に付与された日本十進分類法（NDC）分類記号の補助分類を拡張した。また、開発当時 NDC 分類記号が付与されていなかったサンプル（「分類なし」）などの見直しもあわせて行った。作業は、国立国会図書館の NDC 情報を参照し、人手によって分類の確認と追加を進めた。本作業結果により、たとえば形式区分を利用し、ジャンルの分散する「随筆 (-049)」「理論 (-01)」「研究法 (-07)」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った。このほか、時代情報や小項目が追加されたサンプルもあり、今まで以上に詳細な分類が可能となった。本発表では、情報付与作業の方法と基礎情報を報告し、分類例を示す。本作業結果データは「中納言」の検索結果として利用可能となる。

【ポスター発表】9/2（月）14:45-16:00

〔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-2-7]

実践医療用語の語構成要素への分割と意味ラベル付与の試み

山崎誠 (国立国語研究所)・相良かおる (西南女学院大学)・小野正子 (西南女学院大学)
東条佳奈 (目白大学)・麻子軒 (大阪大学)

本発表では、電子医療記録に含まれる実践医療用語の語構成を明らかにするために、独自に設計した語構成要素への分割とそれに対する意味ラベルの付与を行い、意味ラベルによる語構成のパターンを調査した。調査対象は、ComeJisyoSjis-1 (111,664 語) から、『分類語彙表 増補改訂版』に収録されている語を含む約 7,000 語から抽出した 1,000 語である。これらを短単位よりやや長めの語構成要素に分割し、意味ラベルを付与した。意味ラベルは、石井 (2007) の複合名詞の語構造把握のための意味分類を参考にしたが、実践医療用語のために独自に設けたものも多い。分析結果から、以下のような点が明らかになった。(1) 語構成要素数が 2 個と 3 個のものが全体の 8 割以上を占める。(2) 意味ラベルは、「疾患」「身体部位」「状態」「症状」「医療行為」「時間」「生理」の 7 つで全体の約 8 割を占める。(3) 意味ラベルは、語頭により多く出現するもの（「身体部位」「時間」）や語末により多く出現するもの（「医療行為」「症状」「障害」）などがあり、分布に偏りが見られる。

【ポスター発表】 9/2 (月) 14:45-16:00

〔利用する言語資源〕 分類語彙表, ComeJisyoV6

[P-2-8]

文書領域情報を有する BERT の階層位置に関する考察

欧陽恵子 (茨城大学：学生)・田中裕隆 (茨城大学：学生)・曹鋭 (茨城大学：学生)
白静 (茨城大学：学生)・馬ブン (茨城大学：学生)・新納浩幸 (茨城大学)

BERT は Transformer で利用される Multi-head attention を 12 層 (あるいは 24 層) 積み重ねたモデルである。各層の Multi-head attention は、基本的に、入力単語列に対応する単語埋め込み表現列を出力している。BERT の各層では低層から徐々に何からの情報を取り出しながら、その文脈に応じた単語の埋め込み表現を構築していると考えられる。本論文では領域適応で問題となる領域情報に注目し、BERT の出力の各層が持つ領域情報がどのように推移するのかを考察する。

【ポスター発表】 9/2 (月) 14:45-16:00

〔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-2-9]

データ引用による言語資源活用文献の把握の可能性: BCCWJ の分析から

中渡瀬秀一 (国立情報学研究所)・加藤文彦 (国立情報学研究所)

大向一輝 (国立情報学研究所)

言語資源データの引用情報調査に基づいて、そのデータを活用した研究文献の発見可能性について論じる。このために言語処理学会年次大会発表論文集を対象として「現代日本語書き言葉均衡コーパス」などの引用情報を調査した。本稿ではその結果と今後の課題について報告する。

【ポスター発表】 9/2 (月) 14:45-16:00

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[O-2-1]

All-words WSD と fine-tuning を利用した分類語彙表の語義の分散表現の構築

柳沼大輝 (茨城大学)・古宮嘉那子 (茨城大学)・新納浩幸 (茨城大学)

近年、単語を低次元の実数値ベクトルで表した分散表現が自然言語処理の様々な分野で利用されている。さらに、単語の分散表現や、その作成手法の応用により語義ごとの分散表現を作成する研究がされており多くのタスクで有効な結果を残している。しかし、一般に人手で語義が付与されたコーパスは量が少ないため、十分な量の語義が付与されたコーパスの確保は困難である。そこで、本稿では、語義を自動的に付与した大量の精度が低いコーパスから、作成した分散表現をもとに、人手で語義が付与された少量の精度が高いコーパスを用いて fine-tuning を行い、分類語彙表の語義の分散表現を作成し、その精度を検証した。その結果、分散表現の距離を用いた順位付けによる評価により、fine-tuning による精度の向上が認められた。

【口頭発表】 9/2 (月) 16:10-16:35

〔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-2-2]

UniDic 非コアデータ—解析用 UniDic の ID 情報にひも付く追加情報の公開について—

岡照晃 (国立国語研究所)

本発表では、形態素解析器『MeCab』用の電子化辞書である短単位自動解析用辞書『UniDic』(『解析用 UniDic』) のアペンドデータの公開について紹介を行う。『UniDic』は『MeCab』用の辞書の配布という外部公開形式をとっているが、v2.2 からその解析結果中に各短単位の ID 情報を出力するようになった。この情報を使えば、所外の研究者が自ら拡張した新たなカラムの情報を『UniDic』短単位にひも付く形で配布することができ、研究者間での共有も可能になる。本発表では、短単位の ID 情報について詳説し、それにひも付け、公開を行なっているアペンドデータ『UniDic 非コアデータ』を紹介する。

【口頭発表】 9/2 (月) 16:35–17:00

〔利用する言語資源〕 UniDic

[O-3-1]

配慮表現コーパス作成と分析手法の小考察 対人関係を考慮した対話システムの構築に向けて

太田博三 (放送大学：学生)

昨今、ディープラーニングの進展が急速に高まりを見せながらも、人間関係を配慮した対話ではなくなっている。自然な発話だけでなく、対人配慮した会話は、生活する上で必要不可欠である。この中で、本稿は、従来の制御文により対話応答文生成のアプローチを言語学の語用論の視点と配慮表現とを意識した推論を考察したものである。従来の意味論 (semantics) から今後の語用論 (pragmatics) への推論方法を比較考察し、どのような条件で成立しうるかも試みた、実用的なものである。

【口頭発表】 9/3 (火) 10:30–10:55

〔利用する言語資源〕 国語研日本語ウェブコーパス (NWJC), 現日研・職場談話コーパス, BTSJ 語用論的分析のための 1000 人自然会話コーパス

[O-3-2]

コーパスを用いて「は」と「が」に関する三上説を検証する試み

庵功雄（一橋大学）

「は」と「が」は日本語文法の根幹に関わる要素であるだけでなく、両者の違いを明らかにすることは日本語における予測の実相を考える上でも重要である。本発表では、「は」と「が」の違いに関する三上章の主張の妥当性を CSJ-RDB を用いて検証した。

【口頭発表】 9/3 (火) 10:55–11:20

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[O-3-3]

『日本語話し言葉コーパス』への声質情報付与と予備的分析

前川喜久雄（国立国語研究所）・西川賢哉（国立国語研究所）

『日本語話し言葉コーパス』コア中の母音に、声質研究用に各種音響特徴量を付与する試みについて報告する。母音の無声化等によって測定不可能な母音を除いたすべての母音を対象に、F0, インテンシティ, F1, F2 の平均値, jitter, shimmer, signal to noise ratio, H1*-H2*, H1*-A2, H1*-A3*等の声質関連情報、さらに発話中の位置に関するメタ情報などを付与し、RDB で検索可能とした。この情報の応用上の可能性を示すために、主要な音響特徴量が発話中の位置に応じてどのような変化を示すかを検討した。F0 やインテンシティだけでなく、H1 関連指標などにも発話末において一定の値に収束する傾向が認められた。

【口頭発表】 9/3 (火) 11:20–11:45

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[O-3-4]

類義副詞の文体を測る試み — 「まったく」・「ぜんぜん」・「すこしも」・「ちっとも」を例に—

劉時珍（東洋大学）

本発表はコーパスを用いて副詞の語彙レベルの文体を測る方法を試みるものである。4つの類義副詞、「まったく」「ぜんぜん」「すこしも」「ちっとも」を対象に、8つの形態的指標に基づき、「硬度」と「あらたまり」という2次元的な尺度を交差させることによって類義副詞の文体的な位置づけを明らかにした。2次元に分けた結果、4つの副詞がそれぞれの象限に収まった。「まったく」は硬度が高く、あらたまり度も高い。「ぜんぜん」は硬度はやや高いが、あらたまり度が低い。「すこしも」は硬度は低いが、あらたまり度が高い。そして、「ちっとも」は硬度も低く、あらたまり度も低いという結果である。一方、「硬度」という軸から見れば、「まったく」と「ぜんぜん」は近く、「すこしも」と「ちっとも」は似ているが、「あらたまり度」から見れば、「まったく」と「すこしも」は近く、「ぜんぜん」と「ちっとも」の文体がより似ていることが覗えた。「まったく」と「ちっとも」は最も距離が離れていることが分かった。

【口頭発表】9/3（火）11:45–12:10

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-3-1-E]

政策ニュース映画における否定的な表現の考察—戦後の社会課題と行政施策の可視化の試み—

春木良且（フェリス女学院）・田中弥生（神奈川大学/国立国語研究所）

本研究では、戦後各自治体が制作した行政映画（政策ニュース映画）である川崎市政ニュースを題材に、ナレーションに着目し、政策ニュース映画のコンテンツ分析を試みる。政策ニュース映画は、一般のニュース映画とは異なり、行政による施策が関係する内容が扱われる。本研究では、特にナレーションの否定表現に着目して、そのコンテンツ構造を分析する。それによって、ニュース映画で取り上げられている問題提起を把握することができるため、当時の行政側が認識している社会課題とそれに対する政策の構造が可視化できると仮説付けた。特に、昭和2, 30年代のニュース映画を対象にすることで、高度成長期直前から、現在社会の原型が出来上がるに至るプロセスにおける、社会課題を捉えることを試みた。

【ポスター発表】9/3（火）13:30–14:45

〔利用する言語資源〕 川崎市政ニュース映画ナレーション

[P-3-2-E]

『日本語日常会話コーパス』の短単位解析：作業工程を中心に

西川賢哉（国立国語研究所）・渡邊友香（国立国語研究所）

国語研で構築中の『日本語日常会話コーパス』（CEJC）の短単位解析作業について報告する。CEJCにおける短単位情報は、アノテーションの一つであるにとどまらず、(i) 発音に関する情報を唯一持つ、(ii) 他のアノテーション（長単位・韻律）の初期値作成の際の入力となる、(iii) 転記誤りを発見する際の有力な手掛かりとなる、などの点で重要なアノテーションであり、高い精度が求められる。作業は次のように進める。まず、MeCab+UniDicで自動解析したのち、短単位付加情報の一つである「発音形」を、音を聴取しながら人手で修正する。これにより、発音形の精度向上を図る。さらに、修正された発音形を尊重しつつ再び形態素解析を行なうことにより、発音形以外の短単位情報（境界・付加情報）の精度向上をも図る（例：初出店「ショシュツ/テン」→「ハツ/シュッテン」）。その後、短単位解析結果を、形態論情報管理ツール「大納言」で検索・修正できるようにし、引き続き解析誤りを修正していく。修正が進んだ段階で、境界・付加情報に揺れがないかを系統的にチェックする（例：「ミリ/メートル」「ミリ=メートル」）。

【ポスター発表】9/3（火）13:30-14:45

〔利用する言語資源〕 UniDic, 日本語日常会話コーパス (CEJC)

[P-3-3]

『BTSJ 日本語自然会話コーパス』における反論ストラテジーの分析

陳朝陽（湖北第二師範学院/国立国語研究所）・宇佐美まゆみ（国立国語研究所）

本研究では、『BTSJ 日本語自然会話コーパス』の中から、異性友人間の討論会話5会話（計88分）を用い、反論の発話や談話を、その中に現れた言語形式（「疑問文」、「終助詞」、「接続詞」等）の観点からコーディングを行い、分析した。その結果、反論を行う際は、なんらかの形で「フェイス侵害度」を緩和するストラテジーが用いられており、言語形式としては、「よね、ね、さ、けど」などの「終助詞」、「疑問文」、「でも」などの「接続詞」がこの順によく用いられていること、「笑い」が共起する傾向、男性は女性より「疑問文」を多く使う傾向にあることなどが明らかになった。

【ポスター発表】9/3（火）13:30-14:45

〔利用する言語資源〕 BTSJ 日本語自然会話コーパス

[P-3-4]

手本として示される実習記録の語彙の特徴の分析

山元一晃 (国際医療福祉大学)・浅川翔子 (慶應義塾大学)

看護師を目指す学生のための実習記録を扱った教材に手本として示されている看護記録について、そこに用いられている語彙の分析を行った。今回、分析対象とした書籍に含まれる「アセスメント」、「看護計画」、(看護計画の)「実施・評価」「サマリー」を分析し、4 記録それぞれに含まれる語彙を頻度や対数尤度比といった指標を使って分析した。その結果、記録の種類や、記入する項目により、品詞の分布が異なり、名詞の割合が高いものと、名詞・助詞・動詞/助動詞がまんべんなく使われている項目があることがわかった。また、用いられている語についても、記録の種類や項目ごとの特徴があることが分かった。さらに、76.2% の語は、日本語教育で扱われる可能性があることが分かった。

【ポスター発表】 9/3 (火) 13:30-14:45

〔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-3-5]

BERT を利用した教師あり学習による語義曖昧性解消

曹鋭 (茨城大学：学生)・田中裕隆 (茨城大学：学生)・白静 (茨城大学：学生)

馬ブン (茨城大学：学生)・新納浩幸 (茨城大学)

BERT は Transformer で利用される Multi-head attention を 12 層 (あるいは 24 層) 積み重ねたモデルである。各層の Multi-head attention は、基本的に、入力単語列に対応する単語埋め込み表現列を出力している。つまり BERT は入力文中の単語に対する埋め込み表現を出力しているが、その埋め込み表現がその単語の文脈に依存した形になっていることが大きな特徴である。この点から BERT から得られる多義語の埋め込み表現を、その多義語の語義曖昧性解消ための特徴ベクトルとして扱えると考えられる。実験では京都大学が公開している日本語版 BERT 事前学習モデルを利用して、上記の手法を SemEval-2 の日本語辞書タスクに対する適用し、高い正解率を得た。

【ポスター発表】 9/3 (火) 13:30-14:45

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-3-6]

子どもによるやり取りへの参与の振る舞い—両親との会話の事例分析から—

居關友里子 (国立国語研究所)・小磯花絵 (国立国語研究所)

子どもの会話への関わり方は、大人ほど豊かではないことが予想される一方で、『日本語日常会話コーパス』に収められる家庭で収録された映像データには、子どもが大人とのやり取りに巧みに加わって行く様子も観察される。このような子どもの会話参与に関する詳細を明らかにする足がかりとして、本研究は『日本語日常会話コーパス』から抜き出した、子どもが両親とやり取りを交わしている二つの食事場面を扱った事例研究を行う。各事例の中で子どもたちが、やり取りを生み出すことや、それを進行、展開させることによつてどのように関わっているのかについて、言語的・身体的振る舞いを記述し、考察を行う。

【ポスター発表】 9/3 (火) 13:30-14:45

〔利用する言語資源〕 日本語日常会話コーパス (CEJC)

[P-3-7]

半教師あり語義曖昧性解消における各ジャンルの語義なし用例文の利用

谷田部梨恵 (茨城大学：学生)・佐々木稔 (茨城大学)

単語の語義曖昧性解消は、今日に至るまで様々な研究が行われており、教師あり学習を用いることで高い精度を出している。しかし、先行研究では学習用のデータが不足して誤る事例が多いことが指摘されている。新たに学習データを追加するには、用例文における単語の正解語義の割り当てに精通した専門家によるラベル付与が必要となるためコストがかかるという問題がある。この問題を解決するために、グラフベースの半教師あり学習を用いた語義曖昧性解消を提案し、語義なし用例文の利用による精度改善を行う。そこで、BCCWJの各ジャンルにおける語義なし用例文に対して語義曖昧性解消精度の比較を行い、どのような語義なしデータの利用が有効なのか分析を行う。実験の結果、BCCWJ全ての用例文を追加した場合よりも精度が低くなったが、今回扱ったジャンルの中では雑誌 (PM) に含まれる用例文を追加した場合が最も高い精度結果となった。そのため、ジャンルを限定して語義なし用例文を追加しても、語義曖昧性解消の精度にあまり効果がないことが明らかとなった。しかし、教師あり学習との語義曖昧性解消精度との比較を行った結果、グラフベースの半教師あり学習の語義曖昧性解消精度が高くなったため、グラフベースの半教師あり学習は学習データ不足の改善に有効であると考えられる。

【ポスター発表】 9/3 (火) 13:30-14:45

〔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-3-8]

発話文表現文型辞書の設計と編纂

夏目和子（名古屋大学）・佐藤理史（名古屋大学）

旧版の『日本語表現文型辞書』を改訂した、発話文表現文型辞書について報告する。旧版では、小説の発話文を生成する際に、「ある目的で発話する時、ある話し方をする人物は、この表現文型を使う」という情報を提供することを目標に、発話意図・話し方の特徴ベクトル・表現文型という3種類の情報から成る辞書のしくみを定めた。しかしながら、日本語の発話文では、話し相手との関係に応じて様々な表現文型が使い分けられる。そのため、新版では、話し方の特徴ベクトルを見直し、敬意の有無・距離・場面差・品格などを表す、待遇表現に関する要素を増強した。さらに、実際の発話文では、同じ人物が状況や感情に応じて、様々な表現文型を用いるため、辞書作成の目標を、「ある目的で発話する時、ある話し方で表すならば、この表現文型を使う」という情報の提供に変更した。辞書のサイズは、発話意図が68項目（旧版50）、話し方の特徴ベクトルは20次元（旧版8）、表現文型エントリが1,099（旧版675）である。

【ポスター発表】9/3（火）13:30-14:45

〔利用する言語資源〕 発話文表現文型辞書

[P-3-9]

BCCWJ 小説会話文への話者情報の付与とその活用

山崎誠（国立国語研究所）・柏野和佳子（国立国語研究所）・宮崎由美（明治大学）

本稿では「現代日本語書き言葉均衡コーパス」の図書館サブコーパスに含まれる小説（NDCで913, 923など）のサンプルにおける会話文に話者情報を付与した結果とそれを用いた分析について紹介する。付与したサンプル数は2663サンプルである。付与した話者情報は「話者名、性別、年齢層」（これらは必須）のほか、「話者の社会的属性（職業など）、会話相手の情報、会話モード（電話での会話、方言での会話、外国人の会話等）」なども全てのサンプルには付けている。「話者名、性別、年齢層」については、「中納言」の検索結果で表示することを計画している。また、その他の話者情報は、中納言のサイトからBCCWJ所有者に限りダウンロードできるようにする予定である。分析から分かったこととして以下の4点を挙げる。(1) 小説の全センテンスの約4割が会話文であること。(2) 性別では女性の会話文は全体の約3割であること。(3) 年齢層では約75%が成年層の会話であり、若年層は約20%、老年層は約5%であること。(4) 会話モードでは、電話による会話が全体の約4%程度あること。また、方言による会話文が約5000あり、その多くは大阪を中心とした関西の方言であること。

【ポスター発表】9/3（火）13:30-14:45

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス（BCCWJ）

[P-4-1-E]

国会会議録における言語表現の出現頻度に関する時間的変化モデルの検証

山口昌也 (国立国語研究所)

筆者は、言語表現の出現頻度に関する時間的変化モデルの構築を目的として、国会会議録 (衆議院・予算委員会、衆議院・本会議、1947-2012) を対象に、出現頻度の時間的変化が多い言語表現を抽出し、その特徴を分析してきた。その結果、「会議特有の表現の減少」や「改まった場を想定しない表現の増加」の傾向があることが明らかになった。この結果に基づき、聴衆としての国民が受け入れやすい表現に変化するという、メカニズムを持つモデルを作成した。本稿では、モデルの精緻化を行うため、発言者の肩書、発言の種類 (例：質問、回答、趣旨説明) などをパラメータとして、五つの仮説を立て、実際のデータで仮説を検証した。検証の結果は、いずれの仮説に対しても矛盾するものではなかった。

【ポスター発表】 9/3 (火) 14:55-16:10

〔利用する言語資源〕 国会会議録

[P-4-2-E]

美容院におけるマルチアクティビティ：鏡越しの視線と発話

天谷晴香 (国立国語研究所)

美容院では鏡越しに会話が行われる。鏡越しの視線に注目し、美容院の特殊な物的環境の中で参加者が美容施術中の日常会話が成立するプロセスを記述し明らかにする。通常の会話に伴う視線の移動パターンとの違いはどのように見られるか。本発表では日常会話コーパスから4人のデータ提供者による美容施術場面を扱う。マルチモーダルな事例分析を通して美容院における顧客の視線パターンを例示し、日常生活にある複合活動場面における視線ストラテジーの多様性と普遍性の発見を目指す。

【ポスター発表】 9/3 (火) 14:55-16:10

〔利用する言語資源〕 日本語日常会話コーパス (CEJC)

[P-4-3]

『分類語彙表』と『岩波国語辞典第五版タグ付きコーパス 2004』の対応表

呉佩珣（筑波大学）・近藤森音（東京大学）・森山奈々美（津田塾大学/国立国語研究所）
荻原亜彩美（津田塾大学/国立国語研究所）・加藤祥（国立国語研究所）
浅原正幸（国立国語研究所）

『分類語彙表』の見出し語と『岩波国語辞典第五版タグ付きコーパス 2004』に含まれる国語辞典見出し語との対応表を作成した。分類語彙表は統語・意味に基づいて見出し語を分類したシソーラスであるが、その語義を規定する語釈文を含んでいない。そこで、岩波国語辞典の見出し語と対照させることで対応表を構築し、統語・意味分類と語釈文を結びつける作業を行った。作業は、見出し語表記による2部グラフを構成し、対応する見出し語対を抽出することによる。本作業は5人の作業員により平行して進めた。本作業結果により、『現代日本語書き言葉均衡コーパス』に付与された2種類の語義情報（分類語彙表番号・岩波語義タグ）との対照比較ができるようになった。本発表では、情報付与作業の方法と基礎情報を報告する。

【ポスター発表】9/3（火）14:55-16:10

〔利用する言語資源〕 分類語彙表, 岩波国語辞典第五版タグ付きコーパス 2004

[P-4-4]

BERT を利用した単語用例のクラスタリング

馬ブン（茨城大学：学生）・田中裕隆（茨城大学：学生）・曹鋭（茨城大学：学生）
白静（茨城大学：学生）・新納浩幸（茨城大学）

事前学習モデルである BERT は入力文中の単語に対する埋め込み表現を出力するが、その埋め込み表現はその単語の文脈に依存した形となっている。つまり BERT から得られる単語の埋め込み表現はその単語の意味を表現していると考えられる。本論文では、この点を確認するために BERT から得られる単語の埋め込み表現を利用して、その単語の用例のクラスタリングを行う。実験では日本語版 BERT 事前学習モデルを利用して、単語「意味」の用例クラスタリングを行った。語義曖昧性解消のための標準的な特徴ベクトルや分散表現から構築した特徴ベクトルからクラスタリングを行う場合と比較することで、BERT から得られる単語の埋め込み表現が、より適切に意味を表現できていることを示す。

【ポスター発表】9/3（火）14:55-16:10

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ)

[P-4-5]

「日本語非母語話者の読解コーパス」から見える非漢字圏日本語学習者の辞書使用

クリスティナ・フメリヤク寒川 (リュブリャナ大学)

非漢字圏日本語学習者が日本語を読む際、読めない漢字やわからない言葉を辞書で調べるのに相当な時間を費やすが、学習者がどんな辞書をどのように使い、そこでどんな工夫と困難点がみられるかを探るために、国立国語研究所で開発中の「日本語非母語話者の読解コーパス」に収録されたデータの一部を、辞書使用という観点から分析した。このコーパスのデータは、日本語学習者が普段使っている辞書などのツールを使いながら各自が選んだ文章を読み、理解したことを母語で話す場面を録音（一部録画）、文字化したデータである。調査対象はヨーロッパ各地の大学で日本語を習っている初級から上級までの27名のデータである。分析の結果、辞書選びの段階から、検索を実行し、得られた情報を文脈へ適用する段階まで様々な工夫と困難点がみられたが、より効果的な辞書使用の指導に向けた辞書使用分析の枠組みとデータアノテーションについて述べる。

【ポスター発表】9/3 (火) 14:55-16:10

〔利用する言語資源〕 日本語非母語話者の読解コーパス

[P-4-6]

節中のフィラー「エー」「アノー」「マー」の出現確率に関係する要因

渡辺美知子 (国立国語研究所)・白旗悠真 (東京大学：学生)

「エー」「アノー」「マー」はインフォーマルなスピーチに最もよく出現するフィラーである。これらの使用に関係する要因に違いはあるだろうか。本研究では、これらの出現に関係すると思われる以下の5つの要因の影響力を、ロジスティック回帰分析を用いて検討した。①当該節直前の境界の種類、②節中語数、③話者の性別、④年齢、⑤講演経験。『日本語話し言葉コーパス (CSJ)』コア中の107模擬講演を対象に調べた。分析の結果、境界の種類と節中語数はどのタイプのフィラーの出現確率とも関係していることがわかった。境界の種類は節頭だけでなく節中のフィラーの出現率にも関係していた。語数との関連性は節頭よりも節中で大きかった。年齢と講演経験の影響は「アノー」にのみ観察された。一方で、性別の影響は「エー」と「マー」にのみ観察され、「アノー」には観察されなかった。フィラーの種類によって、出現確率に深く関わっている要因に違いのあることが明らかになった。

【ポスター発表】9/3 (火) 14:55-16:10

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ)

[P-4-7]

『日本語日常会話コーパス』モニター公開版に見られる応答表現

柏野和佳子 (国立国語研究所)

森山卓郎 (2015) (「感動詞と応答」友定賢治 (編) 『感動詞の言語学』ひつじ書房, pp.53-81) では、新情報遭遇に対する反応の応答形式について論じられている。たとえば、①強化や懐疑はない「導入類」である「ふうん」「なるほど」「わかりました」、②先行情報導入に対して確認や想定などの強化を伴う「強化類」である「ああ」「本当だ」、③情報導入に対して意外感を示す、懐疑的態度の場合もある、「意外表示類」である「うそ」「本当?」「はあ?」。本発表では、『日本語日常会話コーパス』モニター公開版を用いて、新情報遭遇に対する応答表現として用いられる語や表現を明らかにする。たとえば、「無理」「全然」「だめ」、「いいね」「さすが」「すごい」、「知っている・知らない」「わかる・わからない」、「やばい」「まじ・まじか」など。

【ポスター発表】 9/3 (火) 14:55-16:10

〔利用する言語資源〕 現代日本語書き言葉均衡コーパス (BCCWJ), 日本語日常会話コーパス (CEJC)

[P-4-8]

日常会話音声に対する基本周波数推定の課題

石本祐一 (国立国語研究所)

音声コーパス構築において韻律ラベリングを行うためには、音声波形の基本周波数を抽出することで声の高さを数値化し、上昇・下降の程度を観察することが必要となる。一般に、『日本語話し言葉コーパス』に収録されているような雑音がほとんど存在しないクリーンな音声に対する基本周波数推定は容易であるが、日常場面のような周囲に様々な音が存在する環境で収録された音声に対しては各種の雑音の影響や発話の重複により誤った推定がなされる場合がある。本稿では『日本語日常会話コーパス』モニター版を基に、推定エラーが生じやすい日常会話音声に対して雑音抑圧や音源分離といった音声信号処理を利用することで、音声コーパス構築に向けてどの程度の基本周波数推定を行うことができるかを示す。

【ポスター発表】 9/3 (火) 14:55-16:10

〔利用する言語資源〕 日本語日常会話コーパス (CEJC)

[O-4-1]

『日本語日常会話コーパス』 モニター公開版：研究の可能性

小磯花絵（国立国語研究所）

本発表では『日本語日常会話コーパス』（CEJC）モニター公開版を用いることでどのような研究の可能性が開けるかを、コーパスを用いた予備的分析を通して具体的に示す。プロジェクトの開始時点では、国語研究所コーパス開発センターが提供するコーパスは書き言葉に集中しており、話し言葉については独話を主対象とする『日本語話し言葉コーパス』のみであった。日常会話を対象とする CEJC のモニター公開により、書き言葉・話し言葉を含む多様なレジスターを対象に、言葉の使用傾向を多角的に捉えることができるようになった。このことを、並列節を導く接続助詞や縮約の分析を通して具体的に示す。また CEJC が多様な話者・多様な会話を収録していることによって、話者の年齢、場面、聞き手との関係性などが言葉の選択に与える影響の分析が可能となった。この点を丁寧体・普通体の分析を通して示す。

【口頭発表】 9/4 (水) 10:30–10:55

〔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), 日本語日常会話コーパス (CEJC)

[O-4-2]

『通時音声コーパス』の可能性と問題点 — 『昭和話し言葉コーパス』の構築と分析—

丸山岳彦（専修大学/国立国語研究所）

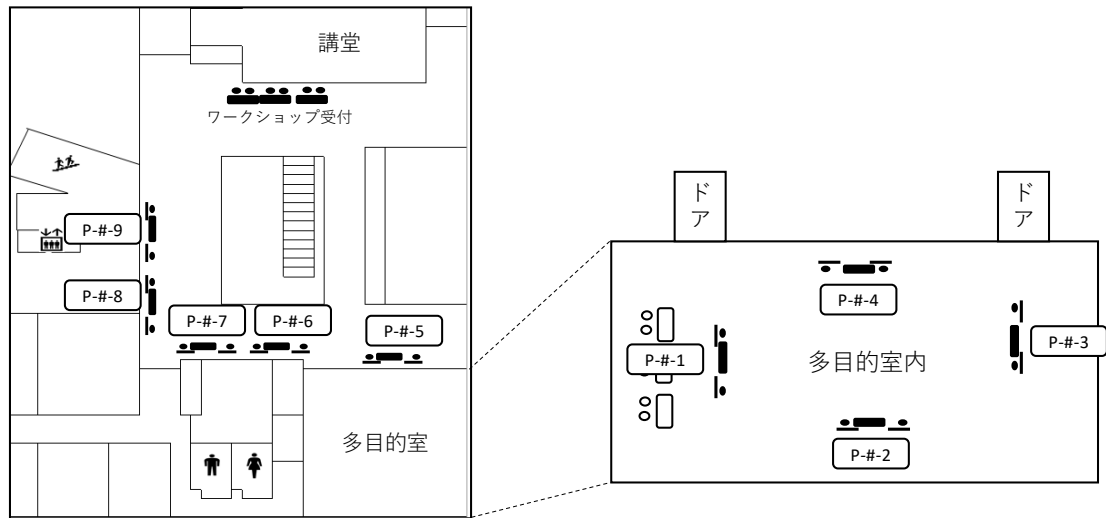
本稿では、話し言葉の経年変化を知るための「通時音声コーパス」が持つ可能性と問題点について述べる。過去の録音資料を大量に収集し、別の時代の録音資料と比較することにより、話し言葉がどのように変化してきたかを実証的に明らかにすることができる。本稿では、「通時音声コーパス」が言語研究に果たす可能性について論じた上で、現在構築している『昭和話し言葉コーパス』を例として、過去の録音資料をコーパス化する際に生じる作業上の問題点について、具体例を交えながら論じる。さらに、『昭和話し言葉コーパス』に見られるいくつかの言語現象を取り上げ、過去の録音資料をコーパス化して分析することの意義について述べる。

【口頭発表】 9/4 (水) 10:55–11:20

〔利用する言語資源〕 日本語話し言葉コーパス (CSJ), 日本語日常会話コーパス (CEJC), 昭和話し言葉コーパス

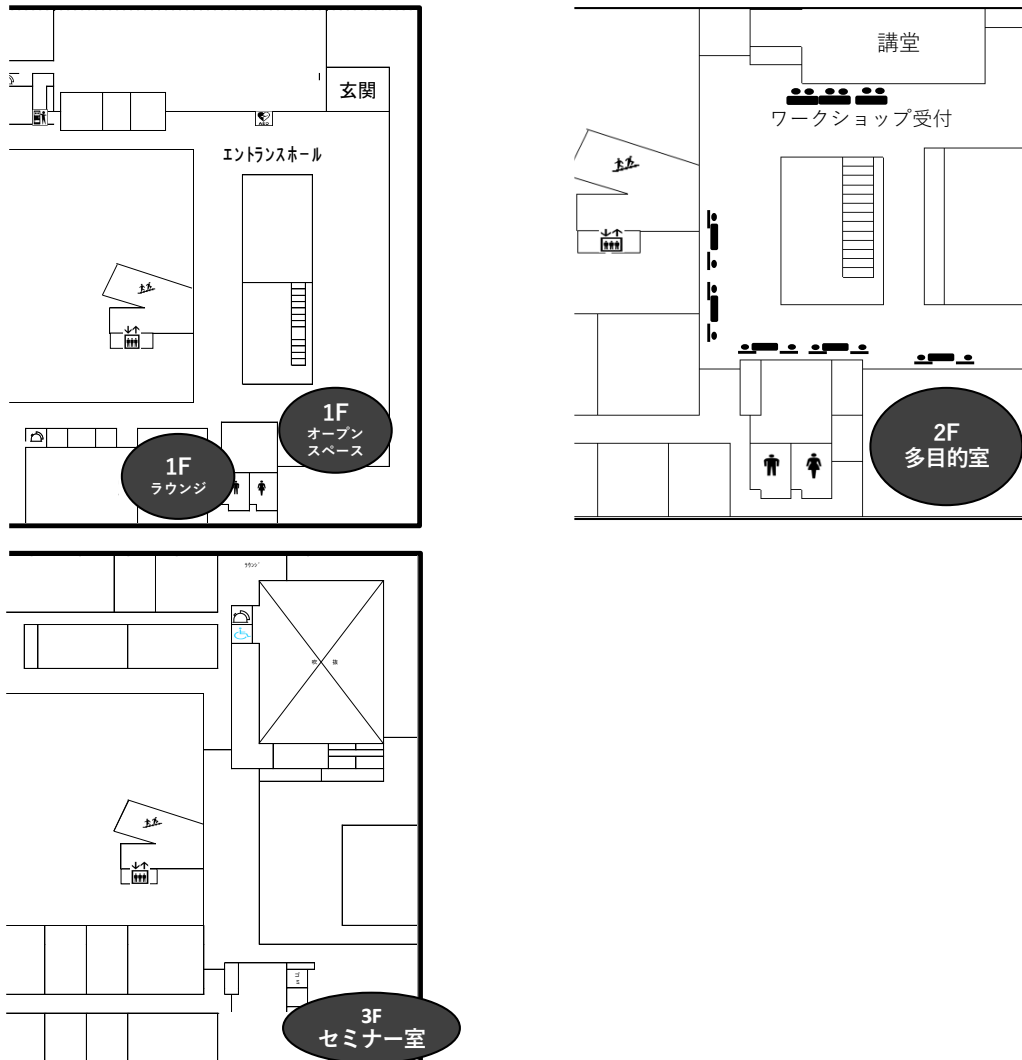
Information

ポスター設営図



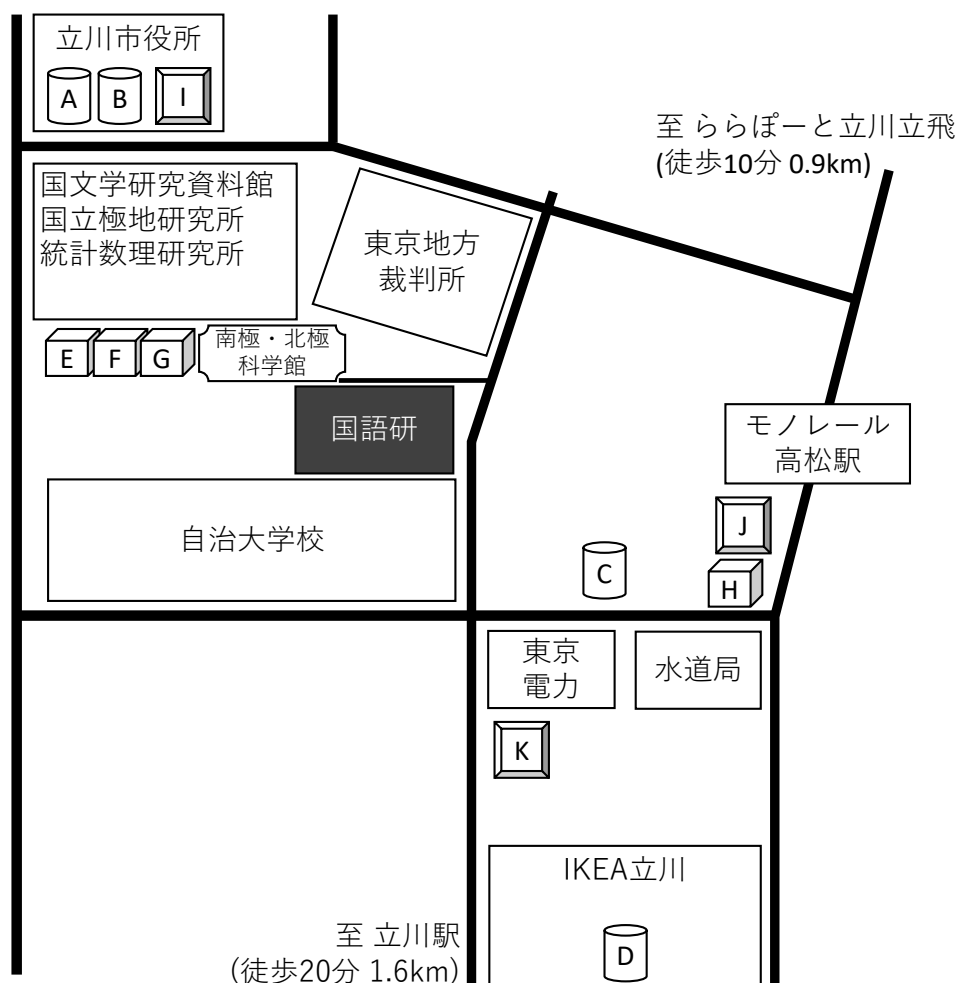
- P-#-1, P-#-2 は電源が利用できます。
- 無線 LAN は eduroam か、受付に掲示している SSID とパスワードでご利用ください。
- ポスター掲示用ついたての貼り付け可能面積は横 120cm× 縦 150cm です。
- ポスター掲示用資材はワークショップ事務局で準備いたします。

ランチスペース



- 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ください。
- 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。

ランチマップ



- お食事処

- A 立川市役所食堂 (市役所内 3F) 11:00-15:00
- B Café はあもにい♪ (市役所内 1F) 10:00-16:00
- C 中国料理 瑞京 11:00-14:30, 17:00-23:00
- D IKEA 立川内 1F ビストロ 2F レストラン 10:00-21:00

- お弁当

- E ハローランチ (南極・北極科学館前) 11:30-13:00
- F カレー工場 HAIJI (南極・北極科学館前) 11:30-13:00
- G 中国料理 瑞京 (南極・北極科学館前) 11:30-13:00
- H お弁当ワゴン販売 (モノレール高松駅南) 11:30-13:00

- 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 2019 Abstract 集

発行日：令和元年 8 月 26 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