

『日本語日常会話コーパス』モニター公開版 —話し言葉研究の展開—

小磯 花絵

国立国語研究所音声言語研究領域

2019年3月4日

モニター公開版の概要

- 対象: 協力者20名により個人密着法で収録した50時間の会話

- 規模

時間	50時間	異なり話者数	237人
会話数	126会話	延べ話者数	392人
セッション数	116セッション	ディスクサイズ	286GB

- 提供するデータの種類:

データ種別	ハードディスク版	中納言版
映像データ	会話の様子を記録した映像データ(電話会話以外)	×
音声データ	話者ごとの音源と会話全体の音源	×
転記テキスト	2種類の単位・3種類のファイル形式	×
短単位情報	形態的側面に着目して規定した言語単位の情報	○
メタ情報	会話・話者に関するメタ情報	一部
全文検索システム	全文検索・短単位検索・検索結果からの動画再生	×

『日本語日常会話コーパス』

Corpus of Everyday Japanese Conversation, CEJC

- 対象: 日常場面の中で自然に生じた会話200時間
- 設計: 多様な会話をバランスよく納めたコーパス
- 収録法:

個人密着法(中心) 約180時間	<ul style="list-style-type: none"> ➢ 40名の協力者(5世代*男女*各4名) ➢ 協力者自身の日常場面の会話を約15時間収録 ➢ バランス等を考慮して4~5時間/1人を選定
特定場面法 約20時間	<ul style="list-style-type: none"> ➢ 個人密着法で収録の難しい場面の会話を補填(例)接客場面, 職場の会話

- 公開データ: 映像・音声・転記・各種アノテーション
- 公開時期: 2021年度末(予定)

本プロジェクトで目指す会話コーパス

- ❖ 日常場面で自然に生じるリアルな活動を記録
- ❖ 多様な場面・話者の会話を記録
- ❖ 音声データ・映像データを記録・公開

日常会話コーパスの収録方法

■ 個人密着法

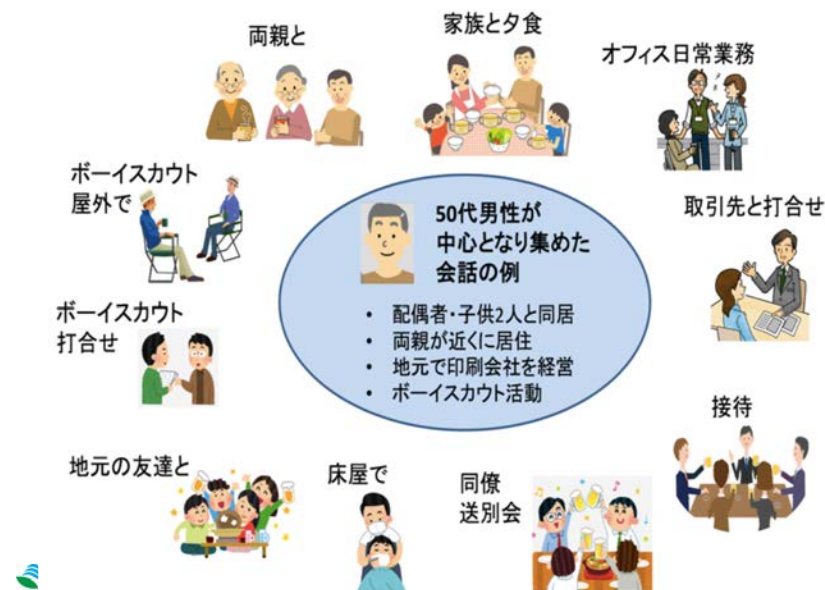
- ✓ 性別・年齢などの観点からバランスを考慮して選別された協力者に収録依頼（首都圏在住者、男女×年齢5世代×各4人=40人、職業偏らないよう配慮）
- ✓ 協力者の日常生活で自発的に生じるリアルな会話を協力者自身に記録
- ✓ コーパス構成比や倫理的問題等を考慮してコーパスに含める会話を選別
 - ・ 1協力者あたり約4-5時間を選別、計180時間（目安）

■ 特定場面法

個人密着法では収録の難しい場面

- ✓ 職場での会議・会合
- ✓ 店舗での接客場面、など

個人密着法による会話の収録例



『日本語日常会話コーパス』モニター公開

◆ 2018年12月4日

◆ 対象会話: 50時間・協力者20名 × 平均2.5時間

年代	職業・職種等	男性			女性			
		収録数	会話数	時間	収録数	会話数	時間	
20代	大学生	5	5	2.2h	大学生	7	7	2.6h
	大学院生	5	5	2.5h	大学生	5	10	2.6h
30代	自営業・自由業	4	4	2.8h	会社員・公務員等	5	6	2.7h
	会社員・公務員等	6	6	2.2h	専業主婦	7	7	2.8h
40代	会社員・公務員等	4	5	2.1h	会社員・公務員等	5	5	2.6h
	自営業・自由業	6	6	2.4h	パート・アルバイト	6	6	2.6h
50代	会社員・公務員等	7	7	2.4h	パート・アルバイト	6	6	2.6h
	会社員・公務員等	4	4	2.6h	会社員・公務員等	7	7	2.2h
60代	その他（非常勤講師）	9	9	2.1h	自営業・自由業	6	6	2.7h
以上	定年退職	6	8	3.0h	専業主婦	6	7	2.7h

日常会話コーパスを用いた研究の可能性

本プロジェクトで目指す会話コーパス

- ❖ 日常場面で自然に生じるリアルな活動を記録
- ❖ 多様な場面・話者の会話を記録
- ❖ 音声データ・映像データを記録・公開

日常会話コーパスを用いた研究の可能性

本プロジェクトで目指す会話コーパス

- ❖ 日常場面で自然に生じるリアルな活動を記録
- ❖ 多様な場面・話者の会話を記録
- ❖ 音声データ・映像データを記録・公開

日常場面で自然に生じるリアルな活動の記録

- 複数の活動の中での会話
家事(料理), 食事, 歯磨き, 宿題
- 話者の構成などが随時変化
父・母・長男・次男 ⇒ 母・父・長男・次男・祖母 ⇒
父・母・長男・次男 ⇒ 父・母・長男・次男 ⇒
父・母・長男・次男 ⇒ 父・母・長男・次男
- 会話の場が随時変化
食卓 ⇒ 勝手口 ⇒ 食卓 ⇒ リビング床

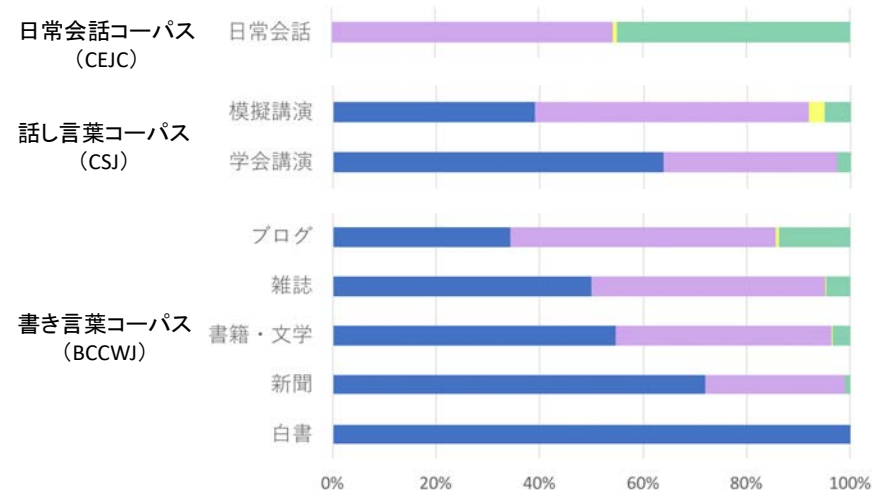
➡ 「収録」のために集められた状況では生じにくい

日常会話コーパスを用いた研究の可能性

本プロジェクトで目指す会話コーパス

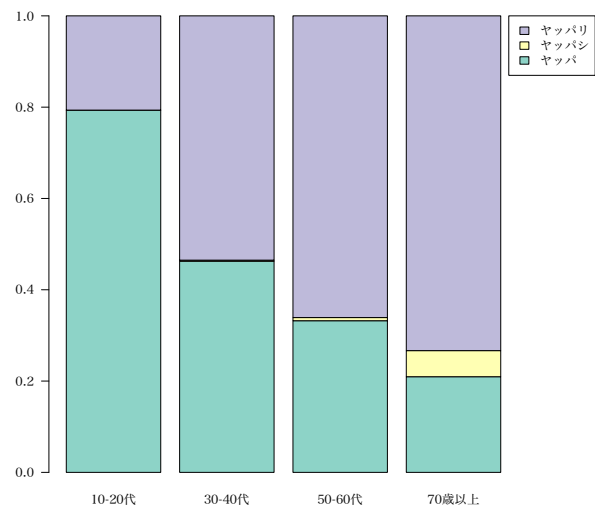
- ❖ 日常場面で自然に生じるリアルな活動を記録
- ❖ 多様な場面・話者の会話を記録
- ❖ 音声データ・映像データを記録・公開

「やはり」の語形

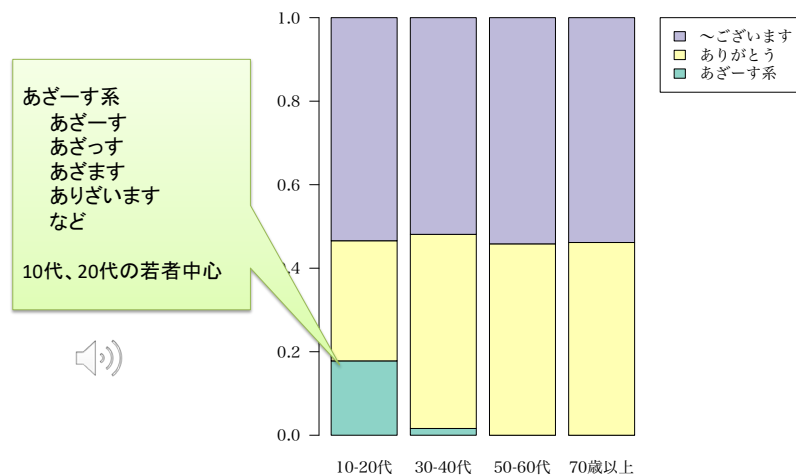


■ やはり ■ やっぱり ■ やっぱし ■ やっぱ

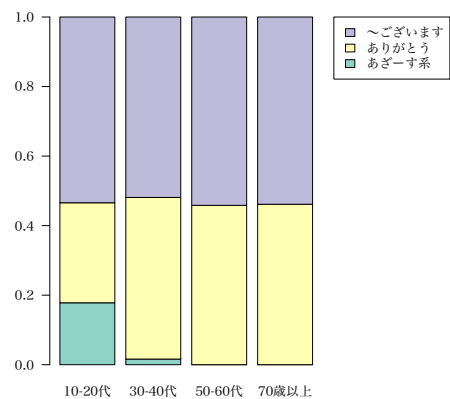
「やはり」の語形：話者の年齢別の傾向



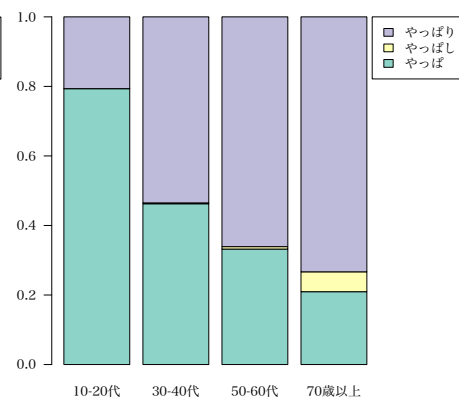
「ありがとう」



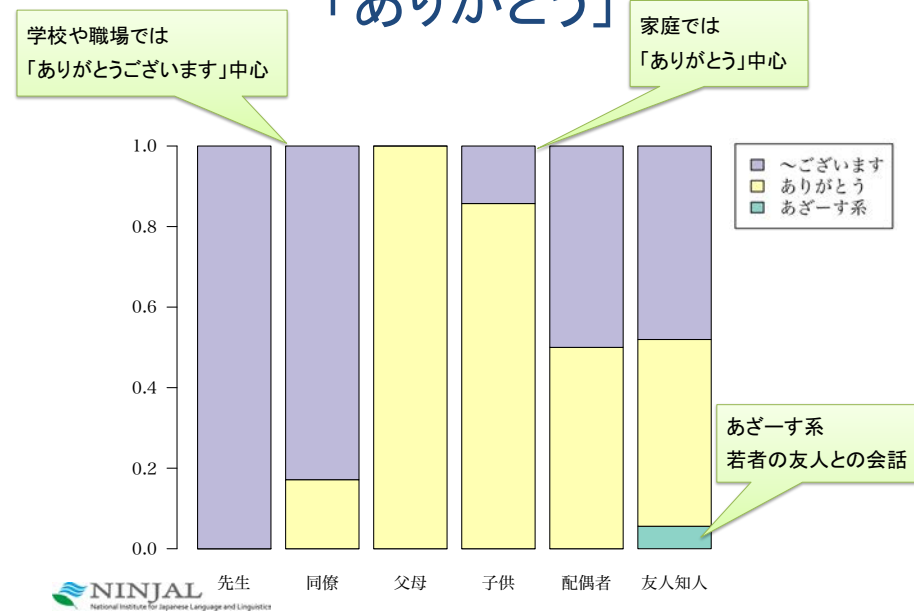
「ありがとう」



「やはり」

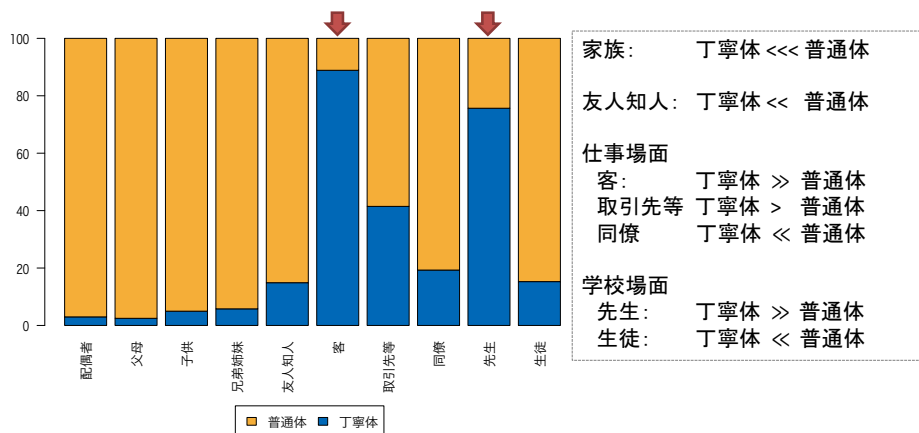


「ありがとう」



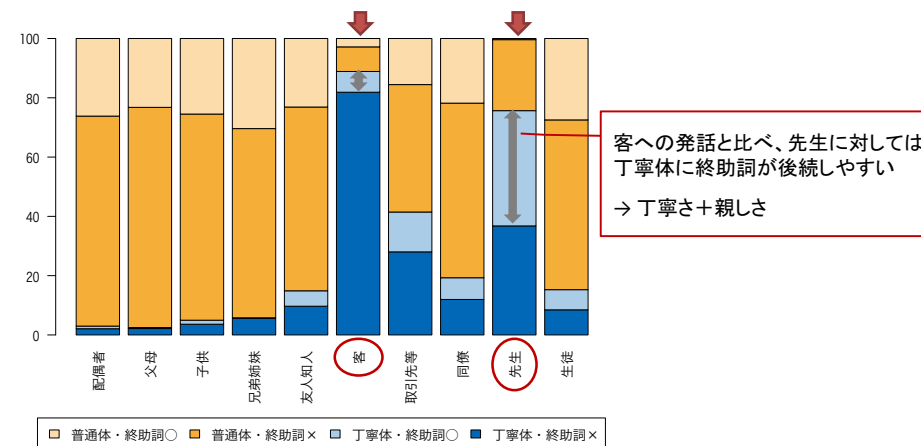
相手との関係性

丁寧体・普通体



相手との関係性

丁寧体・普通体 × 終助詞有無



店員と客の会話例

女将が料理を受ける場面 → 丁寧体が主

客3 お母さん すいません。
 女将 はい。
 客3 揚げパン 二つ。
 女将 はい。
 客3 また 切っ 切ってもらって:。
 女将 揚げパン 二つ。 **中立形(注文繰り返し)**
 客2 じゃ (D 子)。
 女将 二個でいいですか。 **丁寧体**
 客3 二個でいいです。
 女将 はい。
 客2 チーズスティックもお願いします。
 女将 チーズスティック。 **中立形(注文繰り返し)**
 客2 はい。
 客4 すみません。あと 角煮:。
 女将 角煮。 **中立形(注文繰り返し)**
 客4 はい。 **丁寧体**
 女将 はい。ありがとうございます:。

女将が収録器具を見つけて客に話しかける (非接客場面) → 普通体に切りかわる

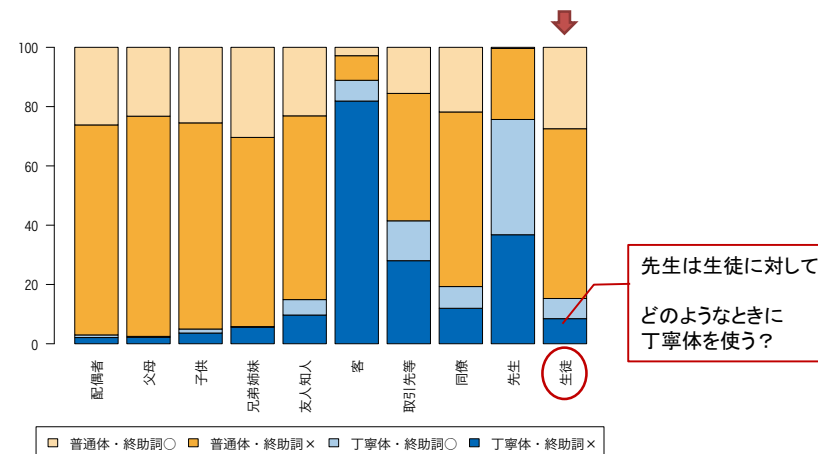
女将 えー。何:これ。
 客4 え。
 客1 ちょっとね。
 客3 もう 当然の 当然の疑問ですよ。
 客1 あー。まあまあまあまあま
 客2 確かに 確かに。
 客3 当然の疑問ですよね。
 女将 びっくり。
 女将 さっきから気になっちゃって。
 女将 かわいい。これ。

ていねい調の中に中立形が現れる場合

前文の一部を繰り返したり否定したりするだけの文が中立形になりやすい(野田 2003)

「相手との関係性」との関係

丁寧体・普通体 × 終助詞有無



先生と学生との会話例

大学生の運営メンバー数名と担当教師で運営する高校の講座の企画などについて、担当教師(高校時代の部活の顧問でもある)と相談。

この場面では日程がかわったことを誰が伝えるかについて語っている。

学生	でも若干変わったじゃないですか。	丁寧体 学生
学生	あの先生の振り返りとかでなくなったので。	
先生	ああああああ。それはどっちでもいいよ。	普通体 先生 +よ
学生	そうゆうのは 伝えたほうが。	
先生	うん。どっちでも。どっちがいいかな。	普通体 先生
学生	まずはその四人でやってもらってゆうのも: どっちから伝えたほうがいいですかね。	丁寧体 学生 +ね
学生	自分がやったほうがいいですか?。	丁寧体 学生
学生	うーん。	
学生	(R 大槻)先生 お願いしてもいいですか?。	丁寧体 学生
先生	うん。いいよ。	普通体 先生 +よ



先生と学生との会話例

教員と学生が講座の企画について打ち合わせの途中で話が脱線し、他のメンバー(女子)の印象について話している

先生	あの子たち いいよ。	普通体 先生 +よ
学生	はい。	
先生	あれ、いいよ。	普通体 先生 +よ
先生	あれはね。	
学生	でもみんなあれですね。	
学生	キャラってゆうとあれかもしれないけど。	
学生	違いますよね。	丁寧体 学生 +よね
学生	全然色が違うので。	
先生	ちよっと: 僕はもう(R 真帆)ちゃんお気に入り。	
学生	いや。でも。わかりますけど。	
先生	あの: なんなんかわからないか。	普通体 先生
学生	わかります わかります。	丁寧体 学生
先生	その 顔。	
先生	うーん。ごめんな。	丁寧体 先生 ←
先生	その美人かどうかって別としてかわいらしいです。	
学生	かわいいですよね。	丁寧体 学生 +よね



個人密着法による会話の収録例

PTA役員引き継ぎ

家族と食事しながら

子供の宿題を見ながら

ママ友ランチ会

夫の実家で

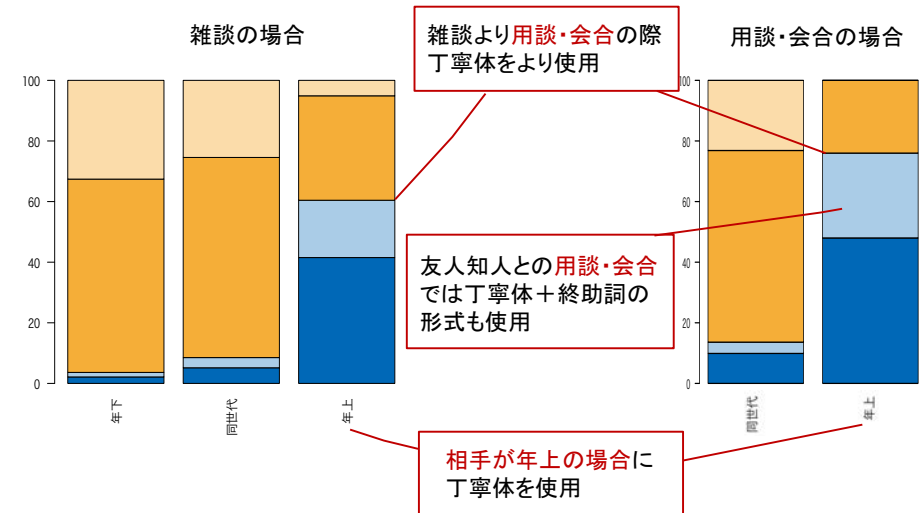
家族で観光旅行

子供・夫のキャッチボールを見ながら

30代女性 専業主婦

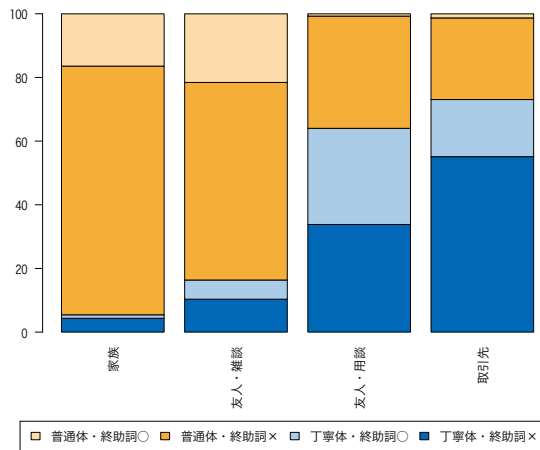
- 配偶者・子供2人と同居
- 両親が近くに居住
- 夫の実家によく行く
- PTA活動に積極的に参加
- ママ友と趣味の会にも参加

友人知人との会話: 世代との関係



協力者K002

50代女性
娘・息子と同居
自営店舗をレンタルスペースとして活用



普通体 >>> 丁寧体
家族

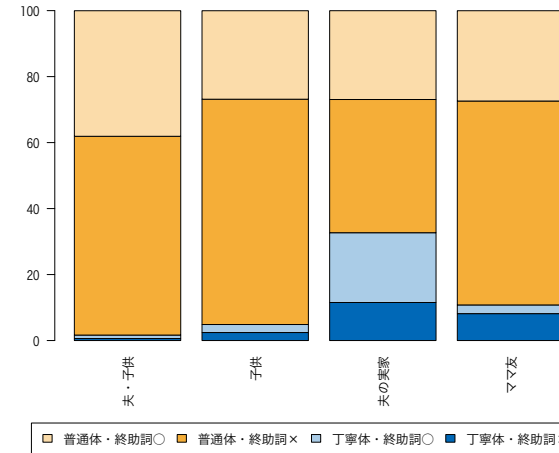
普通体 >> 丁寧体
友人と雑談

普通体 < 丁寧体 + 終助詞
友人と仕事について用談

普通体 << 丁寧体
取引先と用談

協力者T003

30代女性
夫・二人の小学生の息子と同居
夫の実家に頻繁に帰省
ママ友との交流が多い



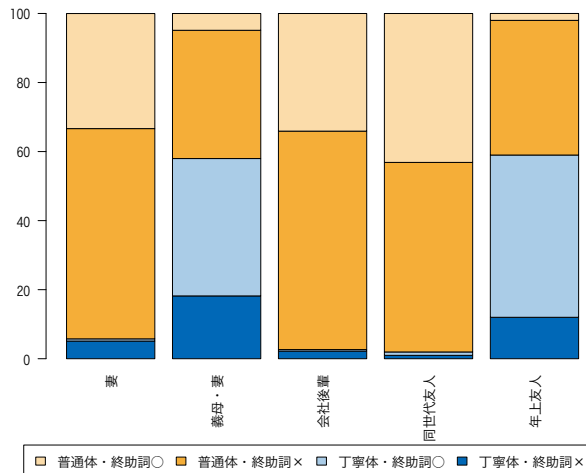
普通体 >>> 丁寧体
子供+夫
子供

普通体 >> 丁寧体
ママ友

普通体 > 丁寧体 + 終助詞
夫+夫の家族(義理の父母弟)

協力者T001

30代男性
会社を退職し起業準備
妻と同居
近所に住む妻の母(義母)との交流が多い

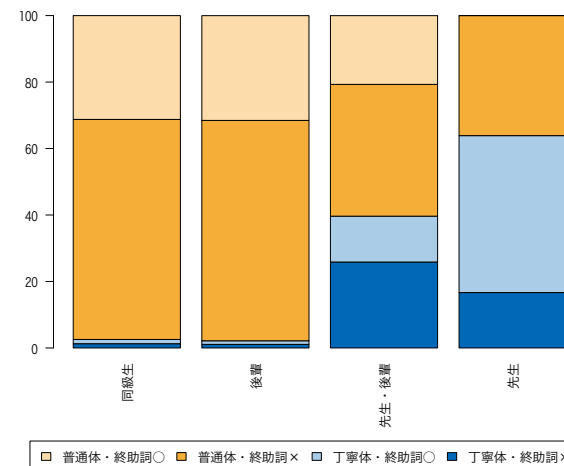


普通体 >>> 丁寧体
妻
同世代の友人+妻
会社後輩

普通体 ≒ 丁寧体 + 終助詞
義母+妻
年上の友人+同世代の友人

協力者T006

20代男性の学生で就活中
TAで先生や後輩とも交流あり
サークル(趣味のプロレス)顧問の先生と仲がよい



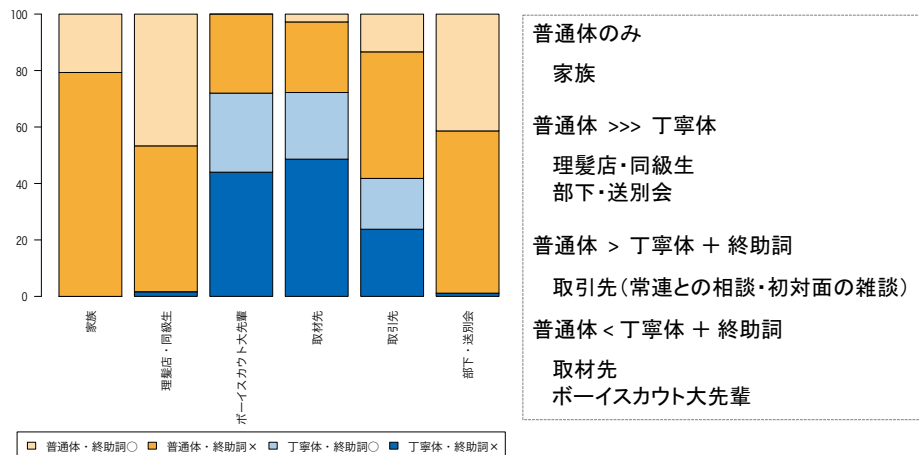
普通体 >>> 丁寧体
同級生
後輩

普通体 >> 丁寧体
TA関係の先生+後輩

普通体 < 丁寧体 + 終助詞
サークル顧問の先生

協力者T015

50代男性
妻・子供2人と同居
地元で印刷会社を経営
ボーイスカウト活動
地元の友人との交流も

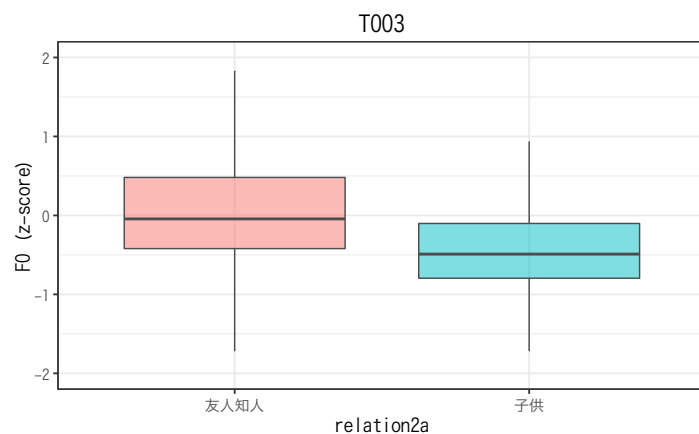


日常会話コーパスを用いた研究の可能性

本プロジェクトで目指す会話コーパス

- ❖ 日常場面で自然に生じるリアルな活動を記録
- ❖ 多様な場面・話者の会話を記録
- ❖ 音声データ・映像データを記録・公開

声の高さの分布



まとめ

『日本語日常会話コーパス』(CEJC)

- モニター公開版のデータの特徴
 - 話者・会話の形式などの分布
- CEJCを用いた研究の可能性
 - 日常場面で自然に生じるリアルな活動
 - 多様な場面・話者の会話
 - 音声データ・映像データの記録・公開
 - 会話コーパスの追加で広がるレジスター的多様性

