

多世代会話コーパスに基づく話し言葉の総合的研究

小磯 花絵



コーパスの構築・公開とコーパスに基づく実証的研究

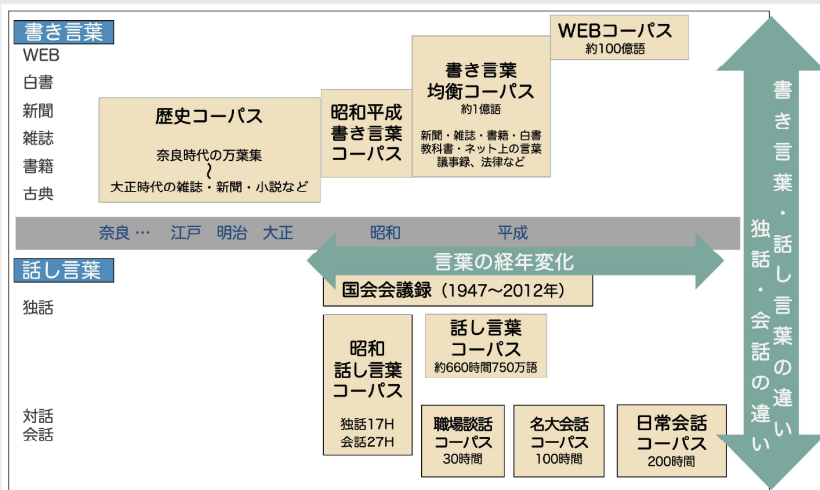
コーパスとは

- 実際に使用された言葉を記録したデータベース
- 目的に応じて対象をバランスよく収集
- コンピュータで扱えるように電子化
- 研究に必要な情報を付与（アノテーション）

コーパスを活用した実証的研究

- 言葉の使用実態や言語変化を調べる
- 背後の要因を調べる
- 言語変化を予測をする
- 言語処理・言語教育への応用、など

国語研究所が公開する代表的なコーパス



※ほかにも日本諸方言コーパスや、日本語学習者の書き言葉・話し言葉を含むコーパスなど

日本語日常会話コーパス

2022年
公開

映像を含む200時間の会話の記録

- 40名の協力者が収録した多様な場面・相手との会話
- 延べ1675人、異なり862人による会話

PTA役員引き継ぎ

ママ友ランチ会

夫の実家で

家族と食事しながら

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

家族で観光旅行

子供の宿題を見ながら

子供・夫のキャッチボールを見ながら

多様なアノテーション（研究用情報）

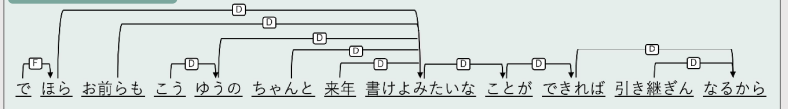
オンラインで検索可

単語・品詞情報

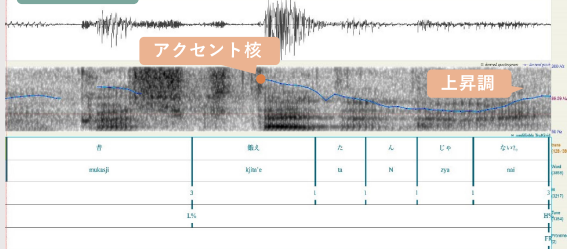
会話・話者に関するメタ情報

会話ID	前文脈	キー	後文脈	話者	性別	年齢	出身地	品詞	活用形	活用	話者間の関係	形式	活動	年齢	性別	出身地
T009_000	集合時間を結構いらい待ち指定しないとい	きつく	いない#それ#ゆつてるはね#あ#遅刻時間短	きつい	キツイ			形容詞一般	連用形一般	和	友人知人	会議・集会	課外活動	20-24	男性	愛知県
T002_015	あとグッズあの(そのたぶんあれば上)園さんに	引き継	けいては#あ#そう#うん#そう(そう)そう	引き継ぐ	ヒキツグ			動詞一般	未然形一般	和	友人知人	雑談	食事・付き合い	40-44	男性	長野県
K002_014	もあるよ#卵焼き(卵)になる#めちやまマの!	卵焼き	好き#メ#あ#卵焼き(卵)で#目玉焼き#なん	卵焼き	タマゴヤキ			名詞・普通名詞一般	名詞一般	和	家族	雑談	食事・家事・雑事	10-14	男性	神奈川県
K002_015	日玉焼き(卵)のあん	焼	じゃない#ヌスランブル(卵)エッグ#作れるもんね#持	焼	ヤキ			接尾辞・名詞一般	接尾辞一般	和	家族	雑談	食事・家事・雑事	50-54	女性	大阪府
C001_001	が(お母)は(大)丈夫です#長く(なり)ま	前書き	長く(なり)ま(い)たが#無事に	前書き	マエオキ			名詞・普通名詞一般	名詞一般	和	友人知人	雑談	付き合い	40-44	女性	東京都

係り受け情報



韻律情報



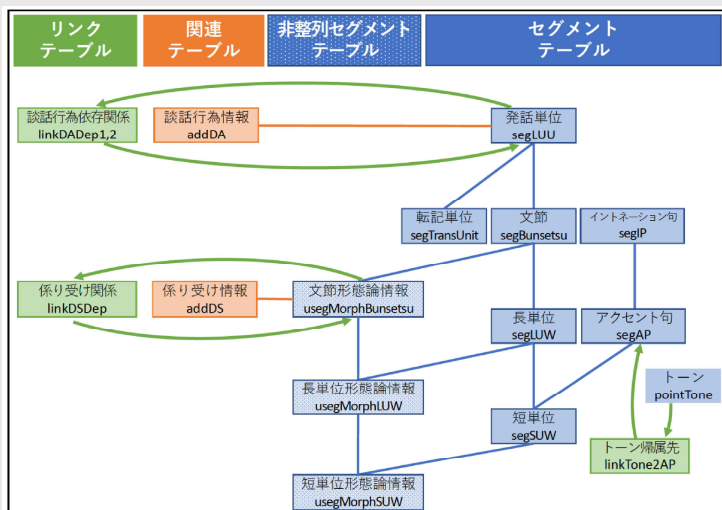
談話行為情報

話者A	情報提供	乗っけちゃったよ。
話者B	独り言	よいしょ。
話者A	謝罪	ごめんね。
話者A	情報提供	ちょっと見た目いいじゃないですか。
話者A	情報提供	中落ち食べよう。
話者A	挨拶	いただきます。

NINJAL
National Institute for Japanese Language and Linguistics

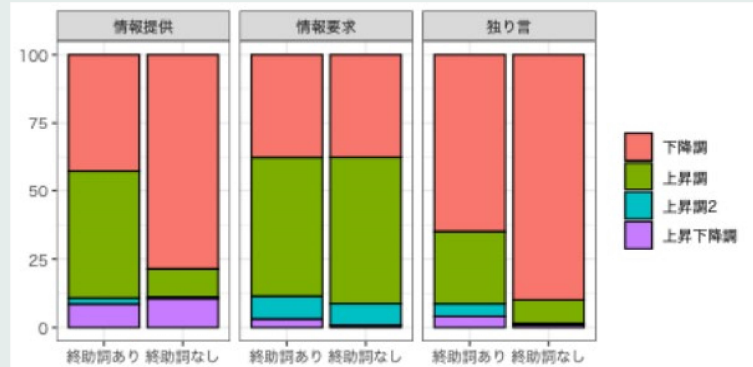
日本語日常会話コーパス

各種アノテーションを統合した リレーショナルデータベース (RDB)



複数のアノテーションを関連付けた研究が可能に

単語品詞情報・談話行為情報・韻律情報を関連付けて分析した例

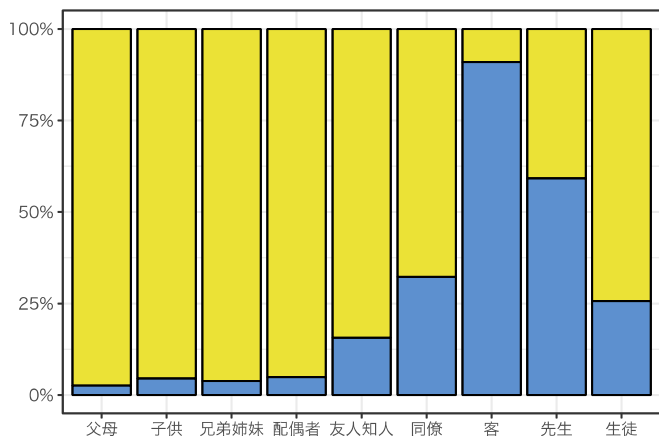


- ✓ 同じ情報提供でも「ね」などの終助詞がある場合は上昇調が多くなる
- ✓ 情報要求(質問)は終助詞の有無に関わらず上昇調が多い
- ✓ 独り言では情報提供ほどではないが終助詞があると上昇調が多くなる

事例 ①

丁寧体・普通体の選択

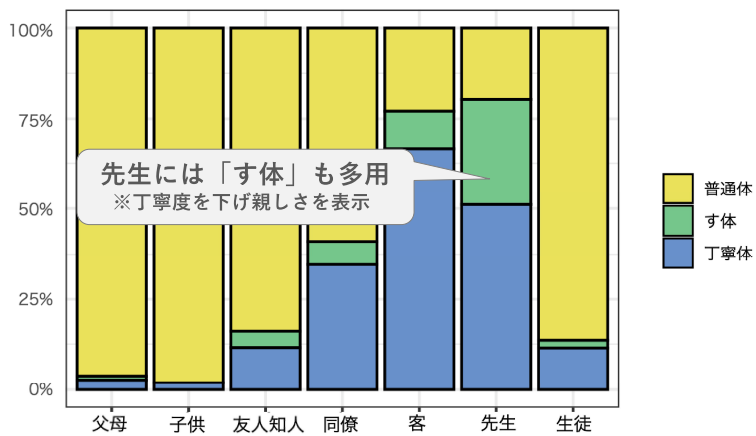
話し手から見た聞き手との関係性別に見ると



家族に丁寧体はほとんど使用せず

先生に対する丁寧体使用率は高くない？

20・30代男性に限定し「す体」も入れると



先生には「す体」も多用
※丁寧度を下げ親しさを表示

す体

面白いです → 面白いっす

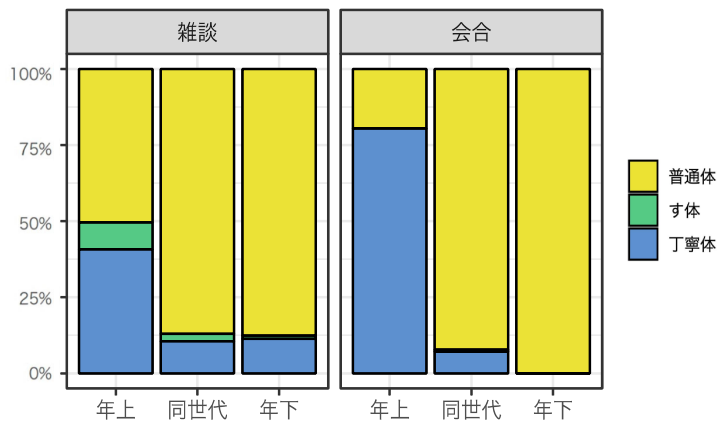
ご飯です → ご飯す

行きます → 行くっす

事例 ①

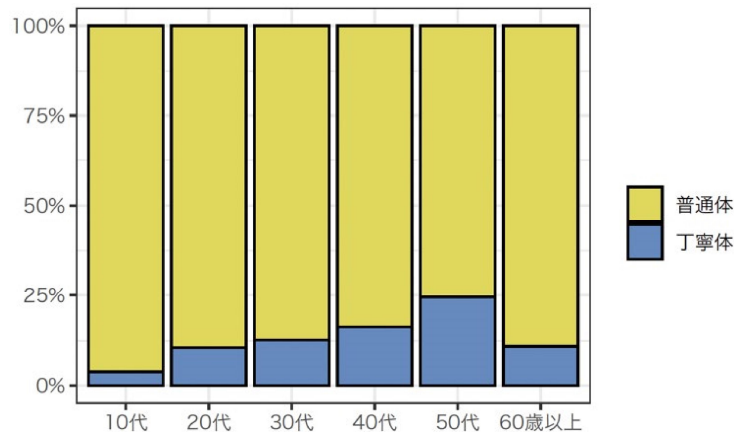
丁寧体・普通体の選択

友人知人に限定して場面別・年齢の上下別に見ると



年齢の上下 (年上か否か) や場面 (雑談・会合) により丁寧体や「す体」の使用率が変わる

年齢別に見ると



年配の人の場合、会話の相手は同世代か年下

昭和話し言葉コーパス

主に1950年代に東京（下町・山手・多摩）で収録した会話などをコーパスとして整備し2022年に公開

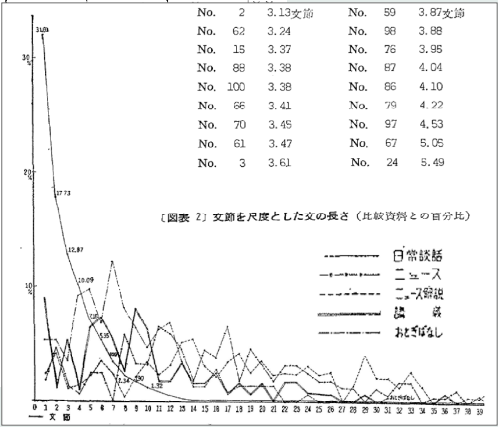
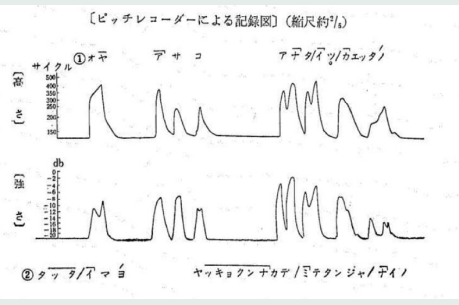


国立国語研究所(1955)『談話語の実態』収録データ p.3

Reel No.	略称	録音状態	地区	場所	性別	年齢	職業	相手	手
			下町	山手	多摩	男女	男女	男女	男女
			町	手	辺	男	女	男	女
3	I 夫妻	可	×	×		×	×	×	×
7	T 家庭談	可	×	×		×	×	×	×
67	N 家庭談	可	×	×		×	×	×	×
86	トクン屋	可	×	×		×	×	×	×
76	じいさん	可	×	×		×	×	×	×
93	魚屋雑談	可	×	×		×	×	×	×
97	U 氏談	可	×	×		×	×	×	×
61	学生談	可	×	×		×	×	×	×
66	井戸端	可	×	×		×	×	×	×
98	友の会	可	×	×		×	×	×	×
2	女子座	可	×	×		×	×	×	×
59	松根屋	可	×	×		×	×	×	×
57	富島屋	可	×	×		×	×	×	×
100	柳屋談	可	×	×		×	×	×	×
5	松岡船	可	×	×		×	×	×	×
25	床屋	可	×	×		×	×	×	×
51	一研雑談	可	×	×		×	×	×	×

同 p.54

国立国語研究所(1960)『話しことばの文型』p.265



事例 2

「終助詞」経年変化

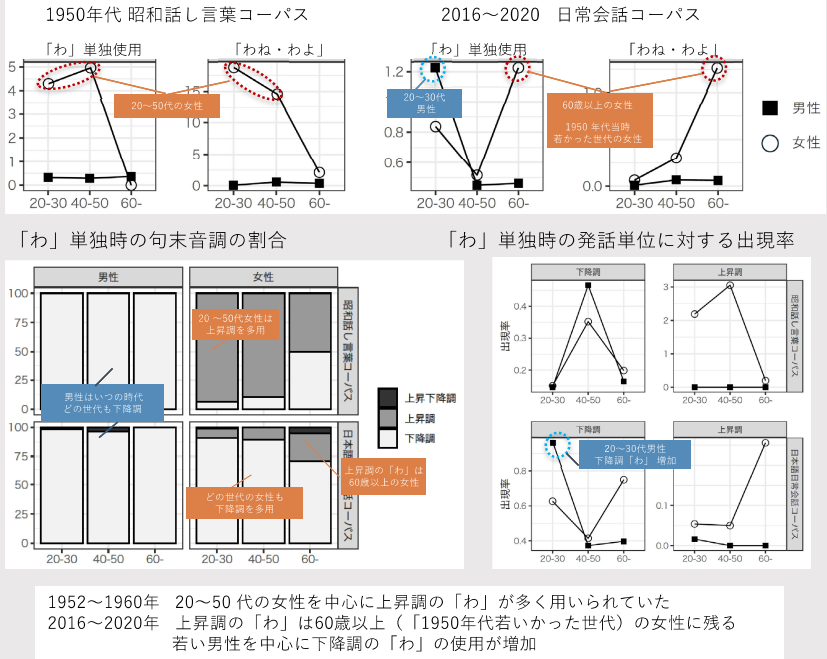
終助詞使用の男女差

- 女性が多く用いるとされる終助詞の例
面白いわ、面白いもの、面白いかしら
 - 男性が多く用いるとされる終助詞の例
面白いぜ、面白いぞ、面白さ
- ⇒ 女性語としての終助詞の女性使用が減っている
男性語としての終助詞を女性も使うようになっている

1950年代の雑談と2016～2020年の雑談

終助詞	1950年代 昭和話し言葉コーパス SSC				日常会話コーパス CEJC			
	男性		女性		男性		女性	
わ	22	0.59%	343***	12.80%	183	0.78%	331**	1.02%
かしら	11	0.30%	55***	2.05%	3	0.01%	38***	0.12%
もの	9	0.24%	28***	1.05%	4	0.02%	25**	0.08%
ぞ	4	0.11%	0	0%	42***	0.18%	9	0.03%
ぜ	3	0.08%	0	0%	40***	0.17%	6	0.02%
さ	8	0.21%	1	0.04%	13	0.06%	17	0.05%

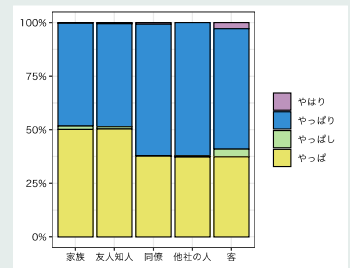
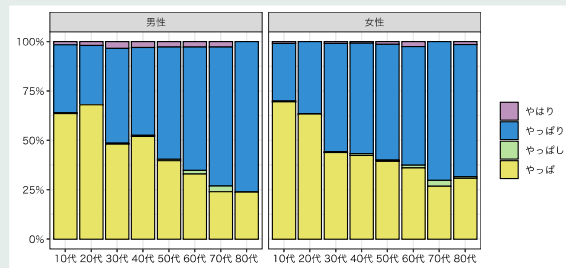
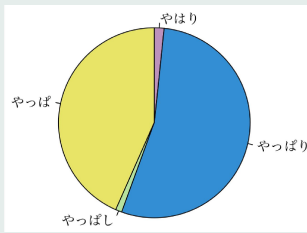
終助詞「わ」の結果（一部）



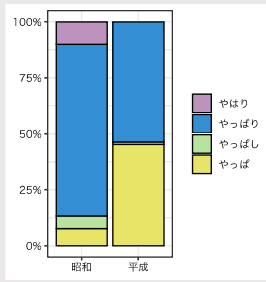
事例 ③

「矢張り」の語形

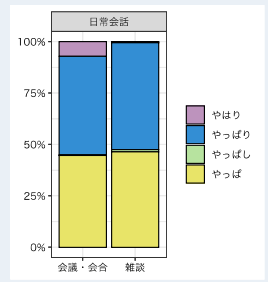
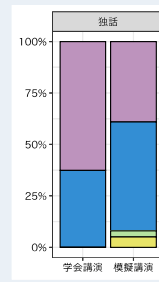
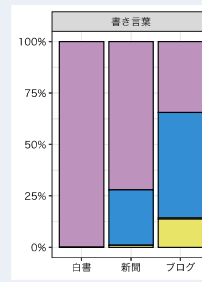
「日本語日常会話コーパス」だけ利用した分析



「昭和話し言葉コーパス」も活用



書き言葉コーパスや独話中心のコーパスも活用



事例 ④

自閉スペクトラム傾向と言語使用の関係

自閉スペクトラム症 (ASD)

- ASDは社会的コミュニケーション・社会的相互作用に困難が見られることが中核症状の1つ
- ASDの障害程度は幅がある
(定型発達者にもASD傾向の強い人、弱い人がいる)

ASD者の言語特徴の1つ

- ASDの幼児は終助詞をほとんど使わない
(佐竹・小林1987; 綿巻1997) ※少数の事例研究

日本語日常会話コーパスの活用

60人を対象に次の追加調査を実施

ASDの特性全般に着目

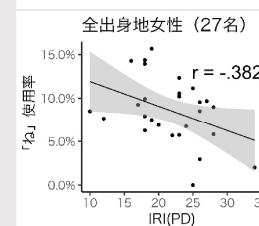
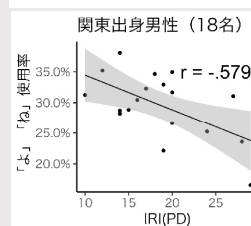
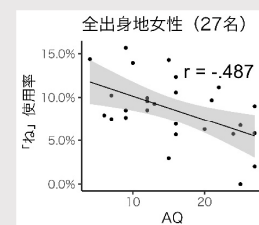
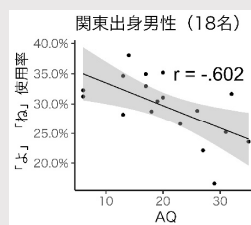
- 自閉スペクトラム指数 (Autism-Spectrum Quotient AQ)

社会的・情動的特性に着目

- システム化指数 (Systemizing Quotient, SQ)
- 対人反応性指標 (Interpersonal Reactivity Index, IRI)
- 共感化指数 (Empathizing Quotient, EQ)
- トロント・アレキシサイミア尺度 (TAS-20)

国立障害者リハビリテーションセンター研究所との共同研究

終助詞「ね」「よ」との関係 (一部の結果に限定)



自閉傾向が高いほど
終助詞使用率が低い

対人認知能力の傾向
がASD者に近いほど
終助詞使用率が低い

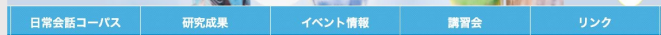
直江・南部・鈴木・小磯・幕内 (2023) 「日本語母語話者の日常会話における終助詞「よ」「ね」の使用と自閉傾向の関係—日本語日常会話コーパスを用いた検討—」『第46回社会言語科学会研究大会予稿集』pp.102-105.

多世代会話コーパスプロジェクト



プロジェクトのホームページ

<https://www2.ninjal.ac.jp/conversation/>



ホーム >> 日常会話コーパス

メニュー

データの仕様・関連文献	関連データ・ツール	利用ガイドラインほか
データ概要	調査協力者一覧	利用目的の範囲・利用方法
会話・話者のメタ情報	会話一覧	成果の公表に際して
転記テキスト	語彙表・語数表	授業や演習等での利用について
アノテーション	その他のデータ	CEJCを用いた研究業績一覧
音声・映像データ	ツール	F A Q
関連文献	リレーショナルデータベース	バグ報告・更新情報



子ども版日常会話コーパス

- 子どもの活動を中心に記録した
- 2025.3～試験公開開始
- 2027年本公開予定



<https://www2.ninjal.ac.jp/conversation/cejc-childM.html>

