

NINJAL
National Institute for Japanese Language and Linguistics

プロジェクト紹介

『多世代会話コーパスに基づく話し言葉の総合的研究』
『コミュニケーション共生科学の創成』

小磯 花絵 (国立国語研究所)
2023年3月4日

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
National Institute for Japanese Language and Linguistics

1

① 国語研究所 機関拠点型基幹研究プロジェクト
『多世代会話コーパスに基づく話し言葉の総合的研究』

- プロジェクトリーダー：小磯花絵 (国語研究所・研究系)
- 研究期間：2022年4月～2028年3月 (6年間)

NINJAL
National Institute for Japanese Language and Linguistics

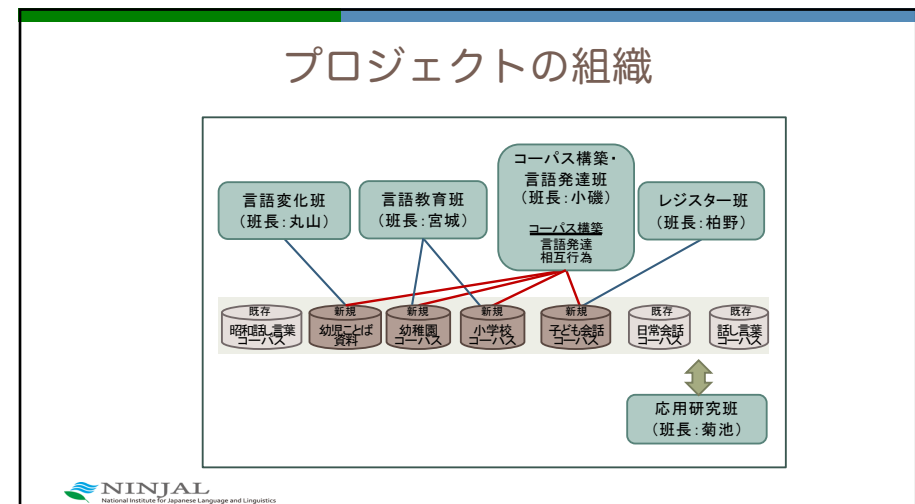
2

プロジェクトのねらい

- ① 『日本語日常会話コーパス』(CEJC)で不足する子どもを主対象とするコーパスを新たに構築し、CEJCと合わせることで、子どもから高齢層まで多世代に渡る話し言葉を分析できる研究基盤を構築する。
- ② 既存・新規のコーパスの分析を通してコーパスを育てる
 - 既存コーパス：研究で得られたアノテーション等を共有化
 - 新規コーパス：構築途中で研究に利用し評価してもらうことで設計に反映
- ③ コーパスを介して異なる分野の研究者と交流を持つ

NINJAL
National Institute for Japanese Language and Linguistics

3



4

子どもデータ拡充の背景

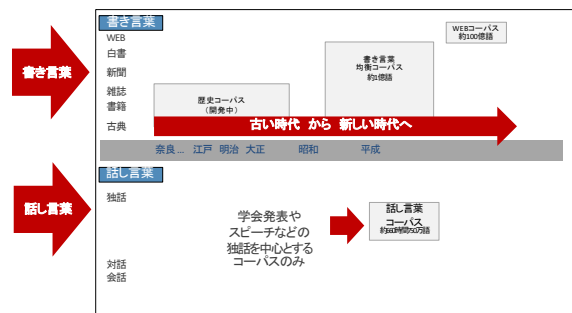
前期のプロジェクト

『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』

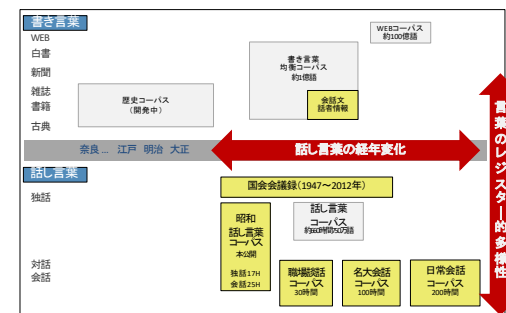
- ❖ プロジェクトリーダー：小磯花絵（国語研究所）
- ❖ 研究期間：2016年4月～2022年3月（6年間）
- ❖ 研究概要：

日常場面で自発的に生じる会話約200時間を収録した大規模な日常会話コーパスを構築し、それに基づく分析を通して、日常会話を含む話し言葉の特性を、レジスター・相互行為・経年変化の観点から多角的に解明することを目指す。

「日常会話コーパス」プロジェクト開始時 (2016.4)



「日常会話コーパス」プロジェクト終了時 (2022.3)



『日本語日常会話コーパス』

Corpus of Everyday Japanese Conversation, CEJC

2022年3月末公開

- 対象：日常生活の中で生じるリアルな会話
- 規模：200時間、577会話、述べ話者数1675名、語数240万語
- 特徴
 - 多様な場面・多様な話者の会話
 - 日常場面で自然に生じるリアルな活動
 - 音声・文字化テキストだけでなく映像データも公開
 - 多様なアノテーション

9

『日本語日常会話コーパス』

調査協力者を中心に多様な場面・話者との会話を収録

50代男性が中心となり集めた会話の例

- 配偶者・子供2人と同居
- 両親が遠くに居住
- 地元で印刷会社を経営
- ボーイスカウト活動

成人の調査協力者40名（年齢・性別でバランス）

年齢	男性		女性	
20代	学生	先生	学生	会社員等
	学生	先生	学生	会社員等
30代	会社員等	自由業	会社員等	自由業
	会社員等	会社員等	自由業	専業主婦
40代	会社員等	先生	会社員等	パートタイム
	会社員等	自由業	パートタイム	自営業
50代	会社員等	先生	会社員等	自営業
	会社員等	先生	会社員等	自由業
60歳～	定年退職	自由業	専業主婦	自営業
	定年退職	先生	会社員等	自由業

10

『日本語日常会話コーパス』 話者の分布

話者数

語数

性別

- 男性
- 女性

子どものデータが少ない
→ 協力者を成人に限定したため、未成年者は協力者の家族が中心

11

子どもの発話データ CHILDES

Collection	Description	Collection	Description
Bilingual	Children learning two or more languages.	Celtic	Irish and Welsh
Chinese	Cantonese, Mandarin, Taiwanese	Clinical	Children with various language disorders
Clinical-MOR	English-speaking children with language disorders tagged morphologically	Dutch/Afrikaans	Dutch and Afrikaans
EastAsian	Korean, Indonesian, Thai	Eng-AAE	North America
Eng-NA	North America	Eng-UK	United Kingdom
French	French	Frogs	Frog story narratives
German	German	Japanese	Japanese
Narrative	Other narratives	Romance	Catalan, Italian, Portuguese, Romanian
Scandinavian	Danish, Swedish, Icelandic, Norwegian	Slavic	Croatian, Czech, Polish, Russian, Serbian, Slovenian
Spanish	Spanish	Other - 1	Arabic, Basque, Berber, Cree
Other - 2	Estonian, Farsi, Greek, Hebrew	Other - 3	Hungarian, Jamaican, Ngun, Sesotho, Tamil, Turkish
XLang	Crosslinguistic studies		

<https://chilDES.talkbank.org/access/>

12

子どもの発話データ NTT 乳幼児音声データベース

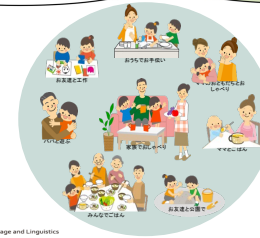
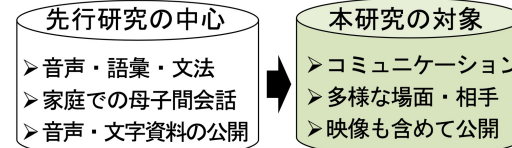
3家庭5名の**幼児とその両親の自然発話**を幼児の誕生直後から最大5年間にわたって断続的に録音した縦断的データベース。
音声・転記・各種アノテーション等を公開。

月齢	収録月数	収録時間	収録回数
0-30	25	161	316
0-54	50	140	720
0-60	61	68	398
0-60	61	66	290
0-59	50	106	691
計	247	541	2415



13

本プロジェクトで目指すコーパスの方向性

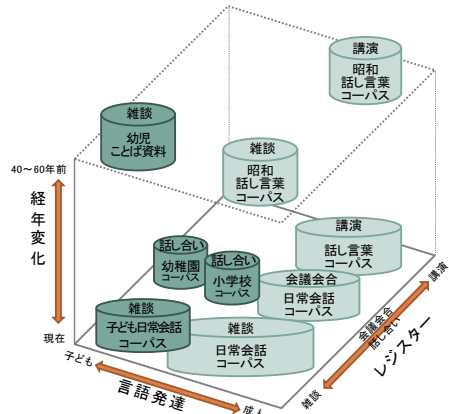


「日常会話コーパス」と接続することで多世代に渡ることばの変化を研究できる基盤整備を目指す



14

研究の基盤となるコーパスの構築



15

子ども版日常会話コーパス 収録対象児

協力世帯ID	性別	収録開始時点の月齢	同居家族	調査協力開始年月	収録期間	備考
Y001	女	2歳6ヶ月	父・母	2019年6月 (妹:2021年8月)	4年間(予定)	CEJC参加者
	女	0歳1ヶ月			2年間弱(予定)	
Y002	女	5歳7ヶ月	父・母	2019年6月	4年間(予定)	日韓バイリンガル
	男	1歳7ヶ月			3年間	
Y005	男	0歳2ヶ月	父・母	2020年1月 (弟:2021年5月)	1.5年間	日中バイリンガル
	男	6歳6ヶ月			3年間	
Y006	女	1歳6ヶ月	父・母・ 姉(小学生)	2020年1月	3年間	
	女	1歳6ヶ月			1.5年間	
Y008	女	9ヶ月	父・母	2020年6月	1.5年間	
	男	4歳2ヶ月			2年間	
Y009	男	1歳2ヶ月	父・母	2020年6月	2年間	
	女	1歳2ヶ月			3年間(予定)	
Y010	男	3歳8ヶ月	父・母	2020年9月	3年間(予定)	
	男	0歳10ヶ月			1年間	
Y011	男	4歳11ヶ月	父・母	2021年2月	1年間	
	男	1歳6ヶ月				



8世帯12(14)児 計100時間

16

Y001 収録の例

- 自宅：
 - お絵描きなど遊びながら
 - お昼ごはん食べながら
 - 父親と遊びながら
 - 洗濯物を畳むお手伝いをしながら
- 祖母宅：
 - 祖母と一緒に
 - はとことシャボン玉で遊びながら
- その他：
 - お散歩しながら
 - お買い物しながら
 - 飲食店でランチをしながら
 - 友人宅で友人2人とその子どもと遊びながら



コロナの影響で2020年からは収録場面・話者が限られる



17

幼稚園での会話の収録

■ 収録開始時期：2019年～（コロナの影響で収録停滞中）

■ 対象： 年中児・年長児

■ 場面：	話し合い	遠足 グループ名相談
	話し合い	遠足 隊長決め
	話し合い	クラス内に貼る写真 撮影場所相談
	話し合い	演劇 役決め
	お話タイム	台風のはなし
	お話タイム	誕生日の人に質問
	お話タイム	活動報告



18

小学校での授業の収録

- 対象：1～3年生
- 場面：国語や道徳、総合などの授業での児童（原則4名のグループ）の活動

これまで収録した活動の例

教科	学年	活動内容
国語	1年生	単語の中から別の単語をみつける学習。グループで例を決めて発表準備。
国語	2年生	好きな本のブックレビューを書き、紹介するために資料を作りスピーチ練習。
国語	2年生	似た意味・反対の意味の言葉を見つけてノートに整理。
総合	3年生	オリジナル味噌汁のレシピの作成、紹介用の資料をPCで使い話し合いながら作成。
道徳	3年生	身近にあるバリアフリーについて意見交換。



20

1970年代に国語研で収録した子ども会話データ

■ 対象：1974年生まれの1男児（と妹）の記録

- ✓ 1歳児のことばの記録（幼児のことば資料3）
- ✓ 2歳児のことばの記録（幼児のことば資料4）
- ✓ 2歳・3歳誕生日のことばの記録（幼児のことば資料1）
- ✓ 3歳前半のことばの記録（幼児のことば資料5）
- ✓ 3歳後半のことばの記録（幼児のことば資料6）
- ✓ 4歳誕生日のことばの記録（幼児のことば資料2）

■ 研究資料室所蔵音源（約100時間）と合わせて再整備



22

② 人間文化研究機構 共創先導プロジェクト

『コミュニケーション共生科学の創成』

- プロジェクトリーダー：菊澤律子（国立民族学博物館）
- ユニットリーダー：小磯花絵（国立国語研究所）
- 研究期間：2022年4月～2028年3月（6年間）



28

プロジェクトのねらい

- 「コミュニケーション弱者」と呼ばれてきた人たちが、コミュニケーションのバリアを意識せず、他の人々と同等に社会活動に参加できるようにするために、現状のメカニズムを解明するなどの基礎研究を実施。
- 研究者・技術者が「障害」の対象を決め一方的な支援を行うという姿勢ではなく、当事者を含む参加者同士の相互認識と自然な相互支援に結び付けることを目指す。



29

研究組織

・ 研究班

- 視覚型コミュニケーション（民博・国語研中心）
- 非視覚型コミュニケーション（民博中心）
- 社会の成員の特性とコミュニケーション（国語研中心）



30

視覚型コミュニケーション班

1. 手話談話のデータベースを構築・公開する。
2. 1のデータベースの分析に基づき、聴覚障害者自身が抱くコミュニケーション問題を浮き彫りにする。
3. 2の検討をふまえ、聴覚障害者が対等の立場で参加するための、会議やイベント実施のマニュアル化を行い、国内における聴覚障害者対応の社会還元を行う。



31

視覚型コミュニケーション班

手話談話収集調査

1. ろう者3名、聴者（手話学習者）3名の合計6名
2. ろう者同士およびろう者者と手話学習者（初級、中級、手話通訳士）による手話談話収集調査
3. オンラインで約30分間、事前に用意したトピックに関する計13談話（内訳：ろう者同士4談話、ろう者と手話学習者9談話）を記録、文字化

32

社会の成員の特性とコミュニケーション班

自閉症スペクトラム症(ASD)のコミュニケーション問題

幕内充氏（国リハ）ほかとの共同研究

▶ 日常会話コーパス

- うち60名の話者を対象に自閉症スペクトラム指数（AQ）や共感化指数（EQ）、システム化指数（SQ）などを調査
- 文末における終助詞「よ」「ね」の使用傾向とAQ指数等との関係
- F0等音声特徴量から見た同調傾向とAQ指数等との関係



34

社会の成員の特性とコミュニケーション班

新たなコーパスの構築・公開

■ 沖村宰氏（昭和大発達障害医療研）・幕内充氏（国リハ）ほかとの共同研究

ASD者と調査者との雑談会話

- ASD者10名（比較のために定型発達者10名）
- 音声・転記・形態論情報・談話行為情報などを公開予定
- 各種心理テストなども含む（IE-SQ, IRI-J, TIPI-J, MAIA）



36

社会の成員の特性とコミュニケーション班

新たなコーパスの構築・公開

■ 鈴木あすみ氏・和田真氏（国リハ）ほかとの共同研究

ASD者と調査者とのインタビュー会話・雑談など

- ASD者10名
- 映像・音声・転記・形態論情報などを公開予定
- 各種心理テストなども含む（AQ, EQ, SQ, CAARS, IRIなど）



37