国語研究所共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」

研究代表者：小磯花絵（国立国語研究所）　研究期間：2022～2027 年度（6 年）

概要

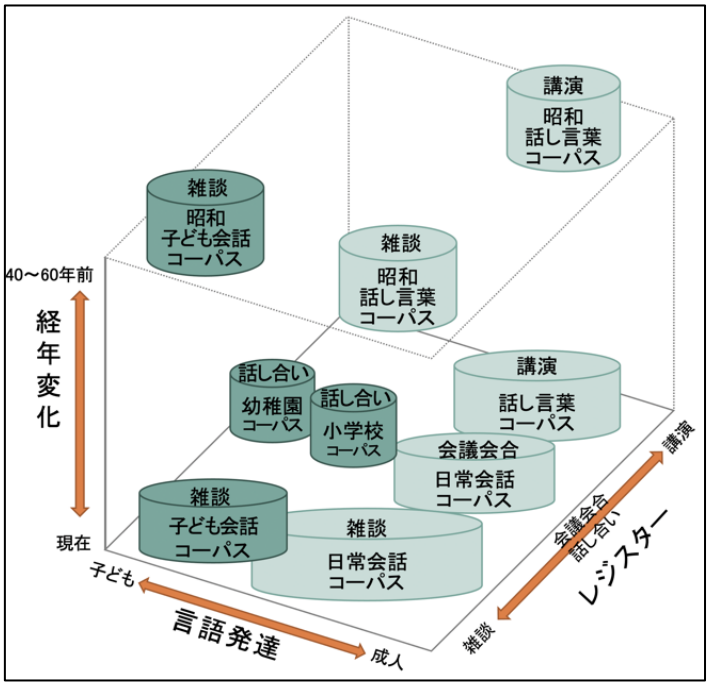
これまで乳幼児を含む子どもの言語発達に関する研究が数多く行われてきたが，発達研究は乳幼児に限られるものではなく，学童期，青年期，成人初期，壮年期，老年期など，多世代に渡り見ていく視点も不可欠である。特に，高齢者や子育て世代の孤立など，家族の問題が複数世代化し，多世代交流が国の重要施策の一つとして掲げられるなか，乳幼児から高齢者までの多世代を対象に，日常生活の中で交わされるリアルな会話を映像まで含めて記録したコーパスを構築・公開し，各世代の言語使用・コミュニケーション行動の実態やその発達・変化を実証的・多角的に研究することは，学術のみならず社会的にも重要な課題である。

現プロジェクト「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」（2016～2021 年度）では，多様な場面・話者によるの日常会話 200 時間をバランスよく集めた『日本語日常会話コーパス』（CEJC）を開発しているが，未成年者，特に 10 歳未満の子どもの数がかなり少ないという問題がある。乳幼児から高齢者までの多世代を対象とする話し言葉の実証的研究を広く推進するには，子どもを対象とする会話コーパスが不可欠である。

そこで 2022 年度か開始する新プロジェクトでは，子どもを中心とする多様な場面・相手との会話を含む映像付きコーパスを新たに開発し，成人中心の CEJC と接続させることにより，コミュニケーションを含む言語の発達・変化の過程を，子どもから高齢者まで多世代に渡り実証的に研究できる基盤を構築する。

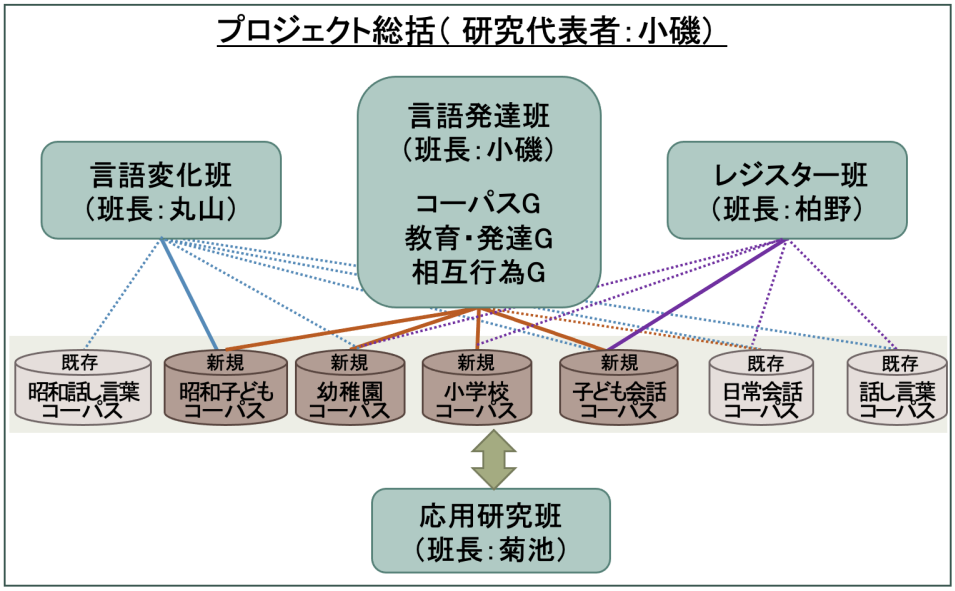
構築予定のコーパスは次の通り：

- 『子ども日常会話コーパス（仮）』：10名の子ども（収録開始時点で生後8ヶ月～6歳6ヶ月，うち2児はバイリンンガル環境）を中心とする日常会話を1～3年に渡り記録した映像付きコーパス（後述）。150時間程度。本プロジェクトの中心的コーパス。
- 『幼稚園話し合いコーパス（仮）』：年中組・年長組（予定）の園児による話し合い活動を中心に収録する映像付きコーパス。規模未定（コロナの影響で収録中断）。
- 『小学校話し合いコーパス（仮）』：1～3年生（予定）の児童による話し合い活動を中心に収める映像付きコーパス。規模未定（同）。
- 『昭和子ども会話コーパス（仮）』：昭和40～50年代に国語研で収録した子ども（1名、0～4歳の期間）の会話を対象とする音声付きコーパス。規模5時間程度。



新規・既存コーパスの位置づけ
（濃い緑が新規構築予定のコーパス）

このようにコーパスを整備した上で，話し言葉の特性を，①言語発達，②時代による言語の変化，③話し言葉のレジスター的多様性の観点から明らかにすると同時に，④各種話し言葉コーパスを活用した情報工学的応用可能性について探る。そのために次の通り4つの班を組織して研究を進める。



① 言語発達班：子どもを中心とする多様な場面（日常場面や幼稚園・小学校での話し合い場面など）の会話を収録した映像付きコーパス（『子ども日常会話コーパス』，『幼稚園話し合いコーパス』，『幼稚園話し合いコーパス』）を新たに構築した上で，成人中心の CEJC も活用することにより，コミュニケーションを含むことばの発達の過程を，子どもから成人まで多世代に渡り実証的に研究する。言語発達研究において層の薄い小学生などの学童期の児童やバイリンガル環境の子どものデータも拡充することで，国語教育や日本語教育も含めて発達研究を幅広く推進する。

② 言語変化班：成人を中心とする 50 年前の話し言葉（会話・独話を含む『昭和話し言葉コーパス』）と、現代の話し言葉を納めたコーパス（会話：CEJC，講演：『日本語話し言葉コーパス』）を活用し，成人の話し言葉の経年的な変化を実証的に研究する。また，50 年前の子ども（1 名）の会話を対象とする小規模コーパスを新たに構築し，子どもや子どもと接する母親の話し言葉の経年的な変化も研究する。

③ レジスター班：雑談や話し合いなど多様な場面の会話を含む成人中心の CEJC と，子どもの雑談中心の『子ども日常会話コーパス』や話し合いを中心とする『幼稚園話し合いコーパス』『幼稚園話し合いコーパス』を主対象に，国語研所有の多様なレジスターのコーパスも合わせて活用することにより，子どもから成人までの年齢層ごとに言葉のレジスター的多様性を実証的に研究する。

④ 応用研究班： 第3期から拡充してきた各種話し言葉コーパスを活用した情報工学的応用可能性について，基本周波数推定，音源分離，音声認識などの音響処理技術や，会話音声から盛り上がりや話者の感情・態度推定やターン認識などのコミュニケーション評価技術などの観点から検討する。高齢者や子育て世代の孤立という問題は技術的な支援も重要となる。こうした技術開発を見据えた基礎研究をコーパス開発と同時に進め，コーパスを評価する。

コーパスに関する補足説明

■　既存のコーパス

共同研究員は，プロジェクトの期間中，国語研が提供する有償のコーパスのうち以下を無償で研究に利用することができます。現プロジェクトの共同研究員が次期プロジェクトにも参画した場合には，引き続きこれらのコーパスを無償利用できます。なお『昭和話し言葉コーパス』は年度内に音声等生データを無償で一般公開するため共同研究員か否かに関わらず利用できます。

① 『日本語日常会話コーパス』 ※現在はモニター版の情報、今年度末公開(200 時間)。

https://www2.ninjal.ac.jp/conversation/cejc-monitor.html
http://doi.org/10.15084/00002540

② 『日本語話し言葉コーパス』　学会講演や一般人のスピーチなど独話中心のコーパス

https://ccd.ninjal.ac.jp/csj/manu-f/overview.pdf
https://ccd.ninjal.ac.jp/csj/

③ 『現代日本語書き言葉均衡コーパス』 多様なレジスターの書き言葉を収めたコーパス

https://ccd.ninjal.ac.jp/bccwj/

■　次期プロジェクトで開発・公開するコーパス

一般公開に先立ち，共同研究員の方には準備の整ったデータから段階的に限定公開し，研究に活用していただきます。2 年目から公開予定です。それまでの期間は，上記既存のコーパスやお手持ちのデータ・コーパスを活用して研究を進めて頂くことになります。

① 『子ども日常会話コーパス（仮）』

以下に示す 14 名の子どもを中心とする日常会話を 1〜3(4)年に渡り記録した映像付きコーパス。データ規模 150 時間（予定）。CEJC と同様，次の方針のもとで収集しています：(1)日常場面の中で自然に生じる会話を対象，(2)子どもの成長とともに広がる多様な場面・話者との会話を収録（祖父母や従兄弟，友達との会話など。ただし現在はコロナの影響で家庭での収録が中心），(3)映像データも記録・公開。

性別	収録開始 時点の月齢	同居家族	備考	収録期間 (いずれも予定)
女 女	2 歳 6 ヶ月 0 歳 1 ケ月	父・母		3(〜4)年間 1.5〜2 年間
女	5 歳 7 ヶ月	父・母	日韓バイリンガル	3(〜4)年間
男	3 歳 8 ヶ月	母		2.5 年間
女 男	1 歳 7 ヶ月 0 歳 2 ヶ月	父・母	日中バイリンガル	3 年間 1.5 年間
女 男	1 歳 6 ヶ月 6 歳 6 ヶ月	父・母 姉（小学生）		3 年間 3 年間
女	8 ヶ月	父・母		2 年間
女 男	1 歳 2 ヶ月 4 歳 2 ヶ月	父・母		2 年間
男	2 歳 0 ヶ月	父・母・祖父母		3 年間
男 男	1 歳 6 ヶ月 4 歳 11 ヶ月	父・母		1 年間

② 『幼稚園話し合いコーパス（仮）』

収録協力の得られた幼稚園（某大学付属幼稚園）では，演劇会での役決めや遠足のグループ名決めなど，イベント等にあわせ，特定のテーマで園児同士が「話し合い」をすることがあります（不定期）。こうした話し合いとは別に，ほぼ毎日，クラスごとに一日の活動内容や特定のテーマなどについて，教員が司会役となり子どもに発話してもらう時間も設けられています。前者の話し合い活動を中心に後者も一部含める形で，2020 年度よりテスト収録を実施し，その後，本収録を開始しましたが，コロナにより収録が中断しています。現時点で収録再開の見通しは立っていません。

③ 『小学校話し合いコーパス（仮）』

収録協力の得られた小学校（同大学付属小学校）では，授業の一環として児童のグループワークが行われています。このうち 1〜3 年生(予定)の児童による活動を中心に収録する予定です。2022 年 5 月頃からの本収録に向けて，1 月よりテスト収録を開始しました。主に授業中の 4 人 1 組の話し合い活動の場面を収録した物ですが，ワークシートを活用した場面，PC を活用した場面，教師が介入する場面など，できるだけ多様な状況を収録することを目指しています。ただし，今後のコロナの状況等により計画が変更になることもありえます。

④ 『昭和子ども会話コーパス（仮）』

1974 年生まれの男児とその家族（主として母親）との会話を収録した音源が国語研究所に残されています。現在は，次の報告書によりその文字化資料が公開され，電子化テキストが CHILDES で公開されています。本プロジェクトでは，音源が発掘できたものについて，その一部を音声付きコーパス（5 時間程度）として整備する予定です。

「幼児のことば資料１」2 歳・3 歳の誕生日の 1 日の記録
「幼児のことば資料２」4 歳の誕生日の 1 日前の日の記録
「幼児のことば資料３」1 歳期（1 歳〜２歳 11 ヶ月まで）の記録
「幼児のことば資料４」2 歳期（２歳〜２歳 11 ヶ月まで）の記録
「幼児のことば資料５」3 歳前半期（3 歳〜3 歳 5 ヶ月末まで）の記録
「幼児のことば資料６」3 歳後半期（3 歳 6 ヶ月〜3 歳 11 ヶ月まで）の記録