国立国語研究所学術情報リポジトリ

子どもの会話コーパスの構築に向けて

メタデータ	言語: Japanese 出版者: 公開日: 2021-03-05 キーワード (Ja): キーワード (En): Corpus of Everyday Japanese Conversation (CEJC), Japanese Children's Corpus 作成者: 小磯, 花絵, 居關, 友里子, 柏野, 和佳子, 角田, ゆかり, 田中, 弥生, 宮城, 信, Sumida, Yukari メールアドレス: 所属:
URL	https://doi.org/10.15084/00003155

子どもの会話コーパスの構築に向けて

小磯花絵 *・居關友里子・柏野和佳子・角田ゆかり・田中弥生（国立国語研究所音声言語研究領域）
宮城信（富山大学人間発達科学部）

Towards the Construction of a Japanese Children's Corpus

Hanae Koiso, Yuriko Iseki, Wakako Kashino,
Yukari Sumida, Yayoi Tanaka
(National Institute for Japanese Language and Linguistics)
Shin Miyagi (Fuculty of Human Development, Toyama University)

要旨

現在構築中の『日本語日常会話コーパス』(CEJC) は、多様な場面・話者との会話を対象とし、映像を含めて公開するというものであるが、CEJC は成人の調査協力者を中心に会話を収集しているため、未成年者の発話が少ないという問題がある。そこで、幼児を含む子どもの会話コーパスの構築を目指し、8 世帯 10 名の子どもを対象とする会話収録と幼稚園での会話収録を進めている。成人話者中心の CEJC と接続させることで、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長期に渡り実証的に研究できる基盤を構築することを目指すものである。本発表では構築予定の子ども会話コーパスの設計方針や収録状況について報告する。

1. はじめに

これまで、乳幼児を含む子どもの言語行動に関する研究が数多く行われている。しかし国内においては、音声や語彙、文法面での発達過程が研究の対象とされることが多く、コミュニケーションに着目した研究の層は薄い。しかしコミュニケーションに問題を抱える子どもが社会問題となる中、子どものコミュニケーション研究を広く展開させることが求められるようになっている（高田ほか 2016,Buckley 2003）。また子ども、特に幼児を対象とする言語研究では、特に母親の子どもに対する影響が強いことから、家庭での母子間会話が研究の対象とされることが多かった。しかし成長するにつれ、公園で友達と話したり、親戚の集まりに参加したり、幼稚園・保育園といった集団生活を中心とする場で会話をするといったように、会話する相手や場面が多様化する。これに伴い、語彙等が増大するだけでなく、語用論的な発達も進む。現実のコミュニケーション行動を多角的に分析するには、子どもの社会を広くとらえたデータに根ざし、相手や場面なども考慮した上で、その発達過程を多角的な観点から解明することが求められる。

こうした研究を推進するために必要となるのが、子どもの言語行動を記録したコーパスであ

* koiso@ninjal.ac.jp

る。これまで CHILDES を中心に子どものデータの記録と共有化が積極的に進められてきた[1]（宮田 2004）。また幼児とその両親の自然発話 500 時間以上を集めた『NTT 乳幼児音声データベース』などもある[2]。しかし既公開データの多くが提供するのは、文字化テキストと音声、あるいは文字化テキストのみであり、映像まで含めて公開しているものはほとんど存在しない。コミュニケーション行動に関する研究を今後広く展開させるには、多様な場面・相手との映像データまで含めたコーパスの共有化が不可欠である。

そこで筆者らは、子どもが多様な場面・相手との会話を通してどのようにコミュニケーション行動を発達させるかを実証的に研究することを目指し、その準備研究として、子ども会話を対象とする映像を含めたコーパス構築に向けて子ども会話の収録を進めている。本発表ではコーパスの設計方針や収録状況について報告する。

2. コーパスの設計方針

2.1 構築の背景

現在、国立国語研究所「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」プロジェクトでは、『日本語日常会話コーパス』（*Corpus of Everyday Japanese Conversation*, CEJC）の構築を進めている（小磯ほか 2017; 小磯ほか 2020）。CEJC は、多様な話者による多様な場面の会話をバランスよく集めることを目指し、年齢と性別の観点からバランスをとった 40 名の成人の協力者（男・女 × 20 代・30 代・40 代・50 代・60 代以上 × 各 4 名）に収録を依頼してできるだけ多様な場面・話者との会話を収録してもらうという収録法を主として採用している。このように、成人の協力者を中心に会話を収集したことにより、話者の年齢に偏りが生じ

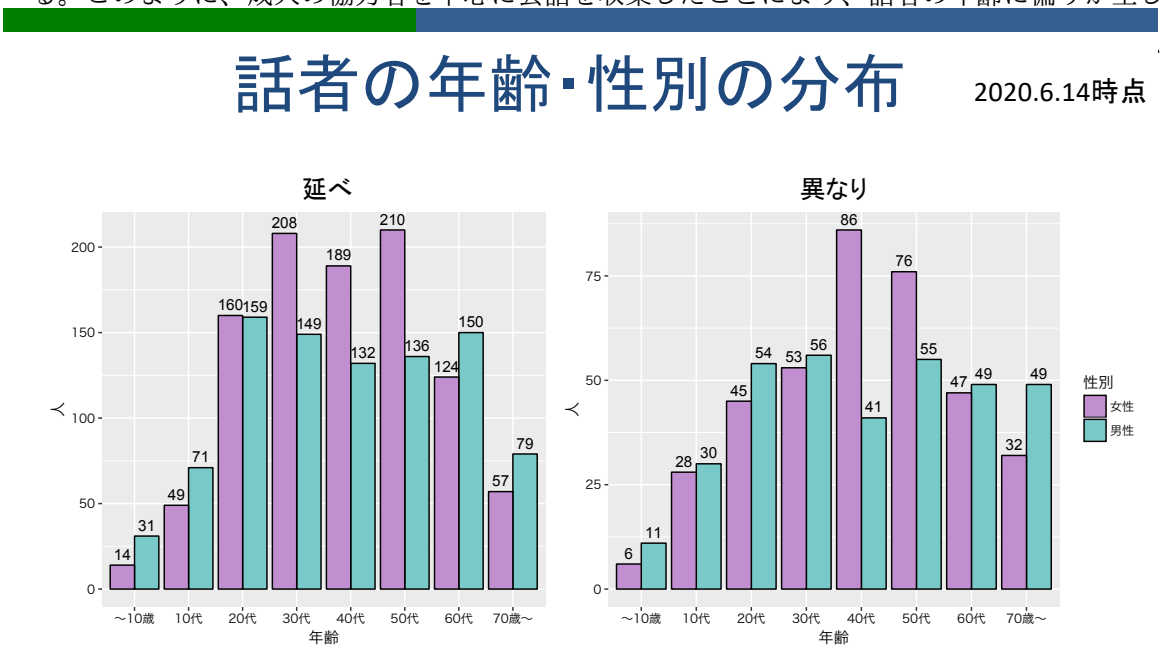

図 1　『日本語日常会話コーパス』に格納予定の話者の年齢・性別の分布（2020 年 8 月 1 日現在）

(1) https://childes.talkbank.org
(2) http://research.nii.ac.jp/src/INFANT.html

そのため、子どもを中心に多様な場面・話者との会話を収録したコーパスを新たに構築し、成人話者を中心とする CEJC と接続させることで、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長期に渡り実証的に研究できる基盤を構築することを目指す。

2.2　収録法

このように、新たに構築予定の子ども会話コーパスは、成人話者を中心とする CEJC と合わせて研究できるように設計する必要がある。そこで、CEJC を踏襲し、次の 3 つの方針のもとでコーパスを構築することとした。

1. 収録のために集められた状況での会話ではなく、日常場面の中で自然に生じる会話を対象とする
2. 子どもの成長とともに広がる多様な場面における多様な話者との会話を収録する
3. 音声データだけでなく映像データも記録して公開する

このようなコーパスを構築するために、現在、次の二つの方法で会話の収録を進めている。

個人密着法　調査協力世帯に収録機材等一式を貸し出し、対象とする子どもを中心とする多様な場面・話者との会話を、原則として毎月 1 時間程度、1〜4 年程度の中長期に渡り収録してもらう方法。調査者は収録に介在せずに、日常場面で自然に生じる会話の収録を目指す。

特定場面法　幼稚園での集団の会話など、個人密着法では収録が難しいと思われる場面の会話を、調査者が主体となり収録する方法。調査者が収録を調整し、実際の収録に関わることもあるが、作られた場面ではなく、日常場面で自然に生じる会話を対象とする。

収録法も CEJC で採用したものを参考に設定したが、個人密着法については子ども収録ならではの変更もある。CEJC では成人の調査協力者に約 3 ヶ月間、集中して会話を収録してもらったが、子どもの場合、経年的にデータを収集することで可能となる研究も多くなることから、1 ヶ月あたりの収録時間を少なめにし、1〜4 年程度の中長期に渡り収録してもらうこととした。

3.　会話の収録状況

3.1　個人密着法による子ども会話の収録

■対象児・協力世帯　現在、表 1 に示す 8 世帯 10 名の子どもを対象に収録を進めている。個人密着法による子ども会話の収録では、特に子どもが小さいうちは家庭での収録が中心となる。そのため、兄弟のいる場合や祖父母と同居する場合、また家庭がバイリンガル環境の場合など、できるだけ家族構成などに多様性を持たせるようにしている。また、1〜4 年程度にわたり経年的に調査する予定であるが、その中で対象とする子どもの年齢の多様性を確保するため、収録開始時期の月齢についても幅を持たせている。

表 1　対象児・協力世帯の情報

収録開始時期	性別	収録開始時点の月齢	同居家族	備考
2019 年 6 月	女	2 歳 6 ヶ月	父・母	
2019 年 6 月	女	5 歳 7 ヶ月	父・母	日韓バイリンガル
2019 年 8 月	男	3 歳 8 ヶ月	母	
2020 年 1 月	女	1 歳 7 ヶ月	父・母	日中バイリンガル
2020 年 1 月	女	1 歳 6 ヶ月	父・母・姉（小学生）	
	男	6 歳 6 ヶ月		
2020 年 6 月	女	8 ヶ月	父・母	
2020 年 6 月	女	1 歳 2 ヶ月	父・母	
	男	4 歳 2 ヶ月		
2020 年 7 月	男	2 歳 0 ヶ月	父・母・祖父・祖母	

■収録調査の流れ　収録調査の方法は原則として CEJC に準ずる（田中ほか 2018）。協力世帯に依頼することの概要は次の通りである。先述の通り調査者は収録に介在しないため、親戚や友人・知人など、会話に新たに参加する人（参加者）に対して収録調査の趣旨を説明しデータ収録・公開に関する同意をとってもらうなど、その内容は多岐に渡る。

1. 参加者に対して収録調査の趣旨やデータ公開の方法などについて説明
2. 参加者に対してデータ収録・公開に関する同意書への署名を依頼
3. 参加者に対してフェイスシート（話者の性別や出身地など）への記入を依頼
4. 収録の日時や使用機材、参加者等の情報を記録
5. カメラ・IC レーダーを用いた会話の収録（毎月 1 時間程度）
6. 収録データの調査者への提出
7. 調査方法や収録した会話に関する調査者との打合せ

■会話の収録　CEJC では、映像収録用にカメラを 3 台、また音声収録用に会話全体を録るための IC レコーダー 1 台と各話者の音声を録るための人数分の IC レコーダーを使用するといったように、多くの機材を用いる収録方法を採用した。これにより、高い精度の映像・音声を記録することができた。しかし子ども収録の場合、育児中の家庭に収録を依頼するため、収録を準備する手間を軽減する必要がある。そこで表 2 に示すように、カメラ 2 台、IC レコーダー 1 台を用いて収録することとした[(3)]。このように IC レコーダーを 1 台に減らしたため、カメラについては、音声を高精度に収録できる小型の機器を選択することとした。採用した ZOOM Q2n-4K は、非圧縮で高いサンプリング周波数の音声を収録することができる（設定：リニア PCM 48kHz, 24bit）。1 台の IC レコーダーと 2 台のカメラを分散して配置することによって、

(3) 収録を進めながら、協力世帯から意見をもらい、機材を確定させた。そのため、GoPro など異なる機材で収録したものが一部含まれている。

例えば部屋の中で遊びながら移動するような場合などであっても、各話者の音声をある程度の精度で録音することが可能となる。収録の様子を図 2 に示す。

表 2　個人密着法で使用する収録機材

品名	設定	使用台数

図 2　個人密着法による収録の様子

このように、CEJC と比べて収録機材をかなり制限してはいるものの、育児中の家庭では、機材の準備が間に合わないことや、屋外での収録のためにこれらの機材を持ち出すことが難しいこともある。そこで表 2 の機材を基本としつつも、必要に応じて、スマートフォンなど容易に利用できる機材を用いてもよいこととした。子ども収録の場合、協力世帯の負担を軽くして中長期に渡り継続して調査に協力してもらうために、このような柔軟な収録方法を採用することが重要となる。

3.2　特定場面法による幼稚園での会話の収録

前述の個人密着法により、家庭での家族との会話を中心に、祖父母やいとこなどの親戚との会話や友達との会話など、多様な会話が収録されている。しかし、協力世帯が収録・公開の承諾を得た相手・場面が対象となるため、幼稚園や保育園での会話を収録することは難しいという問題がある。幼稚園や保育園は多くの子どもとの集団生活が始まる場であり、語彙や文法的な発達だけでなく、語用論的な側面も含めて言語発達が大きく進む場である（杉山 2009）。そこで、調査者が主体となり幼稚園での会話を収録することとした。

■対象児・対象場面　対象とするのは、収録調査への協力が得られた幼稚園における、年中児・年長児（各 2 クラス、計約 70 名）である。

この幼稚園では、演劇会での役決めや遠足のグループ名決めなど、特定のテーマで園児同士が話し合いをすることがある（以下、「話し合い」）。話し合いはイベントなどに応じて不定期に開催される。話し合いでは、教員は調整役にまわり、子どもの自主的な行動が尊重される。そのため、各自の意見を出し合った上でジャンケンや多数決などで決定するなど、子どもを中心とする合意形成の過程を観察することができる。またイベントの内容によっては、クラスを越えて話し合いが持たれることもある。話し合いについては、主として調査者が立ち会える場合に収録する。

こうした不定期に開催される話し合いとは別に、ほぼ毎日、クラスごとに、一日の活動内容や特定のテーマ（例えば台風）などに関して、教員が司会役となり子どもに発話してもらう時間が設けられている（以下「お話しタイム」）。お話しタイムの会話については、月に 1～数回、各クラスの教員に依頼して収録してもらうこととした。

■会話の収録　幼稚園での会話を記録する機材を表 3 に示す。話し合いやお話しタイムでは、原則として園児は輪になって座る。そこで、輪の中心に 360 度撮影可能なカメラと IC レコーダーを配置し、園児たちの行動を中心に記録する。また調査者が収録に立ち会える場合には、カメラを 2 台、会話の輪の外に配置し、会話の様子を俯瞰的に記録する。

表 3　特定場面法による幼稚園の調査で使用する収録機材

		品名	設定	使用台数
映像	中央配置	Kodak PIXPRO SP360 4K	1440 × 1440, 60 fps	1
	脇配置*	ZOOM Q4n	1920 × 1080, 30fps	2
音声		SONY ICD-SX1000	リニア PCM 44.1kHz, 16bit	1

＊会話を俯瞰的に収録するカメラは、調査者が収録に立ち会える場合に限定して使用

4.　おわりに

本発表では、子どもの会話コーパス構築の準備研究として現在進めている子ども会話の収録調査について報告した。成人話者を中心とする CEJC と接続させ、子どもから高齢者まで長期に渡り実証的に研究できる基盤を構築することを目指し、設計や収録法はできるだけ CEJC を

踏襲するようにした。しかし、子どもを養育中の世帯や幼稚園の協力を得て中長期に渡り調査を行うために、CEJC から変更したところもある。今後、調査が進む過程で、協力世帯・幼稚園からの意見を取り入れ、収録法の見直しが生じることもある。

謝　辞

本研究は、科研費 20H01264、国立国語研究所フィージビリティ型プロジェクト「子供の言語コミュニケーション研究に向けた基盤整備」および同基幹拠点型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」による成果である。

参 考 文 献

B. Buckley (2003)　*Children's Communication Skills: From Birth to Five years*, Routledge.

小磯花絵・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉 (2017)「『日本語日常会話コーパス』の構築」『言語処理学会第 23 回年次大会発表論文集』775-778.

小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉 (2020)「『日本語日常会話コーパス』モニター版の設計・評価・予備的分析」『国立国語研究所論集』18, 17–33.

宮田 Susanne 編 (2004)　『今日から使える発話データベース CHILDES 入門』ひつじ書房.

杉山弘子 (2009)　「幼児の話し合い活動とコミュニケーションの発達との関連」『尚絅学院大学紀要』57, 91–102.

高田明・嶋田容子・川島理恵編 (2016)　『子育ての会話分析』昭和堂.

田中弥生・柏野和佳子・角田ゆかり・伝康晴・小磯花絵 (2018)　「『日本語日常会話コーパス』の構築—会話収録法に着目して—」『国立国語研究所論集』14, 275–292.