

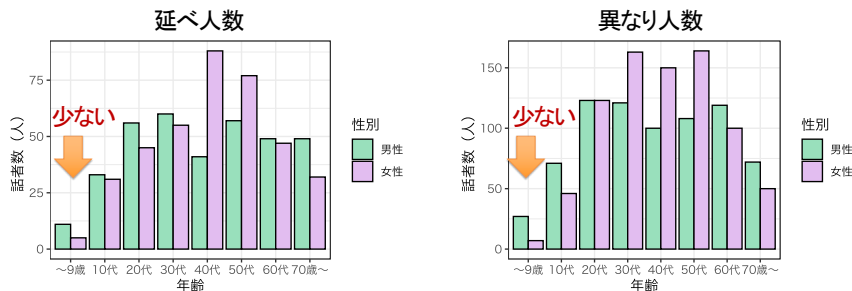
# 『子ども版日本語日常会話コーパス』の構築

小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子(国語研究所)  
 藤越(東京大学/国語研究所)・川端良子・田中弥生・西川賢哉(国語研究所)

## コーパス構築の背景

### ■ 『日本語日常会話コーパス』(2022年3月公開)

- 成人の調査協力者を中心に会話を収録
- 話者の年齢の分布を見ると男女ともに未成年者(特に10歳未満)が少ない



### ■ 国語研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」(2022~2027年度)

- 子どもを中心に多様な場面・話者との会話を収録した『子供版日本語日常会話コーパス』(仮)を新たに構築し、成人話者を中心とする『日本語日常会話コーパス』と接続
- 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

## コーパス設計

原則として『日本語日常会話コーパス』を踏襲

### ■ 設計方針

- 日常場面で自然に生じる会話を対象
- 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
- 音声だけでなく映像データも記録して公開

### ■ 収録法

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、毎月1時間程度、1~4年程度の中長期に渡り収録。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を収録。

### ■ 収録機材(基本収録)

- 映像 ZOOM Q2n-4k 2台(設定:1920×1080, 30fps)
- 音声 SONY ICD-SX1000 1台(設定:リニアPCM 44.1kHz, 16bit)

### ■ 規模

- 100時間の会話
- 延べ話者数900人、総語数120万語程度(これまでの収録実績やCEJCIに基づき試算)

## 収録状況

### ■ 対象児 協力世帯

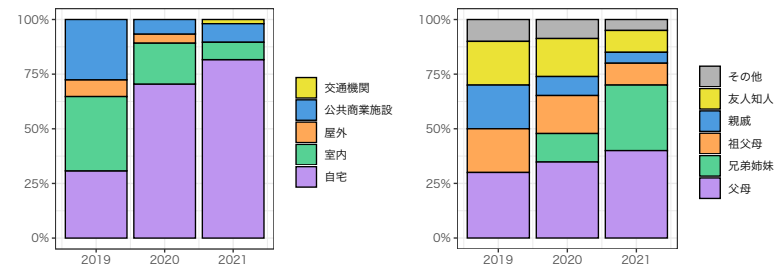
世帯ID	性別	収録開始時点の月齢	収録開始時期	収録期間(予定)	同居家族	備考
Y001	女	2歳6ヶ月	2019年6月	48ヶ月	父・母	
	女	0歳1ヶ月	2021年8月	22ヶ月		
Y002	女	5歳7ヶ月	2019年6月	36ヶ月	父・母	日韓バイリンガル
Y005	女	1歳7ヶ月	2020年1月	36ヶ月	父・母	日中バイリンガル
	男	0歳2ヶ月	2021年5月	20ヶ月		
Y006	男	6歳6ヶ月	2020年1月	36ヶ月	父・母・	
	女	1歳6ヶ月			姉(小学生)	
Y008*	女	0歳9ヶ月	2020年6月	18ヶ月	父・母	
Y009*	男	4歳2ヶ月	2020年6月	25ヶ月	父・母	
	女	1歳2ヶ月				
Y010	男	3歳8ヶ月	2020年9月	30ヶ月	父・母	
	男	0歳10ヶ月				
Y011*	男	4歳11ヶ月	2021年2月	14ヶ月	父・母	
	男	1歳6ヶ月				

\*世帯IDに“\*”を付けた3世帯は2022年8月現在で収録を終えている

### ■ 収録の様子



### ■ 収録状況



- 2019年: 自宅での家族との会話以外の会話も多い
- 2020,2021年: 自宅での家族との会話を中心 … 新型コロナウイルス感染拡大の影響