国立国語研究所学術情報リポジトリ

談話コーパスの作成・公開・それを用いた研究

メタデータ	言語: jpn 出版者: 公開日: 2022-09-15 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/3673

談話コーパスの作成・公開・それを用いた研究

井上文子

1. はじめに

2020年，新型コロナウイルス感染症が世界中に拡大し，日本社会も激変した。日常生活の不安や不自由さだけではなく，医療，経済，教育，文化など，多方面に大きな影響が出た。移動や接触の制限によって，会場に集っての学会や研究会は中止となった。方言研究でも，フィールドワークのために現地を訪れたり，話者の方々に直接対面して調査をしたりすることができない状態であった。

そのような状況の中，日本方言研究会では，「方言研究支援プロジェクト」による取り組みがなされた。現地調査によらなくても可能な方言研究として，遠隔インタビュー調査・通信調査の提案や，今までに蓄積されたさまざまな言語資料・方言資料を活用する多様な研究方法が日本方言研究会のwebページに挙げられている。

そのうちのひとつ，利用可能な資料のひとつとして，方言談話資料がある。かねてから方言の会話を地域の生きたことばの記録として残す動きはあったが，それが研究のデータとして利用されることが多くなってきたのは比較的新しい。また，談話を対象とした研究はさかんにおこなわれてきたが，当初は，標準語談話に比べて，方言談話が扱われることは少なかった。最近は，方言談話を対象とした研究も活発になり，方言を分析するための基礎資料として，談話データが活用されている。それに伴って，大量の談話データを効率よく迅速に検索・処理できる，研究用の情報を付与したコーパスが必要とされるようになってきた。

近年，大量の言語データが整備され，大規模な言語コーパスの構築が進み，コーパス言語学と呼ばれる新たな言語研究が展開している。日本語方言の分野でも，全国規模の方言コーパスの構築が開始された。

本稿では，既存の言語資源である方言談話資料を利用して，談話コーパスを作成・公開する手順と，談話コーパスを活用した研究事例を紹介する。

2. 談話コーパスの作成・公開

2.1 方言談話資料

談話資料とは，一般に，日常的な会話などの音声を録音し，それを忠実に文字化し，共通語訳や注記をつけたものをいう。

全国的な方言談話資料は，ある程度まとまりのある期間に，同じ条件のもとに，一貫した方法で収録された資料であるので，多数のデータの比較・検討・分析が効率よく便利にできるという点で利用価値が高い。

記録としては，文字化のみでも貴重ではあるが，音声や映像があることによって客観性が保たれ，研究に用いた場合は科学的な証拠資料ともなり，論文に記載されている記述や分析の検証が可能となる。大規模な方言談話資料で，音声が付属しているものには，次の【表1】のようなものがある。

当初，紙媒体やソノシート・カセットテープなどのアナログ媒体で作成されたものも，その後デジタル化され，音声や文字化・共通語訳データが公開され，コンピュータでの利用が可能となっている。ただ，電子化されていても，データが横断的に検索できるものは今まで存在しなかった。そのような状況の中で，国立国語研究所において，『日本語諸方言コーパス』の構築が計画された。

『日本語諸方言コーパス』の原資料は，1977～1985年度に文化庁が実施した方言談話の収録事業「各地方言収集緊急調査」である。日本全国47都道府県のそれぞれで5地点程度，約200地点の方言談話を録音したカセットテープと，各地点約9時間分の方言談話の文字化・共通語訳などの手書き

【表 1】代表的な方言談話資料

資料名	収録時期	収録地点	収録時間	話者	談話形態	媒体	データ形式
CD-ROM 版 全国方言資料 （日本放送協会）	1952 ～ 1968 年	141 地点	約 30 時間	60 歳以上 男性・女性	自由会話, 場面設定会話	CD-ROM	冊子（pdf）, 音声（wav）
方言録音資料シリーズ （国立国語研究所）	1963 ～ 1971 年	14 地点	1 地点 約 20 ～ 130 分	老年層男性・ 女性中心	自由会話, 場面設定会話	冊子, https://mmsrv.ninjal. ac.jp/hogenrokuon_siryo/	冊子（pdf）, 方言文字化・共通語 訳（txt）, 音声（wav）
方言談話資料 （国立国語研究所）	1975 ～ 1981 年	20 地点	約 25 時間	60 歳以上 男性・女性, 20–30 歳台男性	自由会話, 場面設定会話	冊子, カセットテープ, https://mmsrv.ninjal. ac.jp/hogendanwa_siryo/	冊子（pdf）, 方言文字化・共通語 訳（txt）, 音声（wav）
全国方言談話データベース 日本のふるさとことば集成 （国立国語研究所）	1977 ～ 1985 年	48 地点	約 24 時間	60 歳以上 男性・女性	自由会話	冊子, CD-ROM, CD	冊子（pdf）, 音声（wav）, 方言文字化・共通語 訳（txt）
日本語諸方言コーパス （国立国語研究所）	1977 ～ 1985 年	56 地点＊	約 35 時間＊	60 歳以上 男性・女性	自由会話, 場面設定会話	https://www2.ninjal. ac.jp/cojads/	方言文字化・共通語 訳（CSV）, コーパス検索アプリ ケーション「中納言」
東日本大震災と方言ネット （東北大学方言研究センター）	2012 年～	東北 15 地点＊		70歳前後 男性・女性	自由会話, 場面設定会話	冊子, https://www. sinsaihougen.jp/	方言文字化・共通語 訳（pdf）, 音声（mp3）, 映像

＊更新がおこなわれるものについては，2020 年 12 月 19 日時点

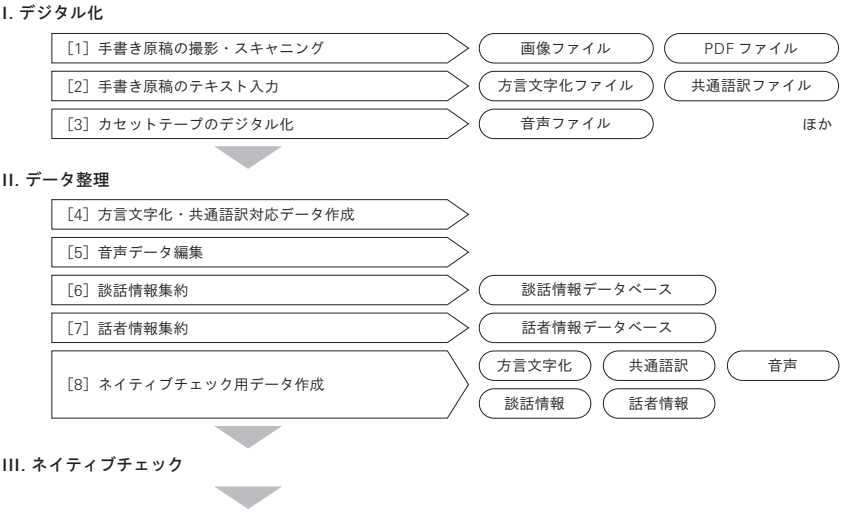
原稿が国立国語研究所に残されている。

長期間公開されなかった「各地方言収集緊急調査」報告資料の一部は，2001～2008年に『全国方言談話データベース 日本のふるさとことば集成』として刊行された。2019年には『全国方言談話データベース 日本のふるさとことば集成』のデータを利用して，全国の方言の談話を横断的に検索することのできる音声つき方言コーパスである『日本語諸方言コーパス』が構築された。

2.2 コーパス化の流れ

『日本語諸方言コーパス』を例に，未公開の既存資料であった「各地方言収集緊急調査」の生データを活用して，基礎データを作成する手順，コーパス化のプロセスなどを説明する。

既存の資料をコーパス化する場合，データ作成の観点からは，大きく分けて2段階の作業が必要となる。原資料から基礎データを作成する段階と，基礎データからコーパスデータを作成する段階である。まず，原資料から基礎データを作成する段階は，次の【図1】のような手順をとる。



【図1】基礎データ作成の流れ

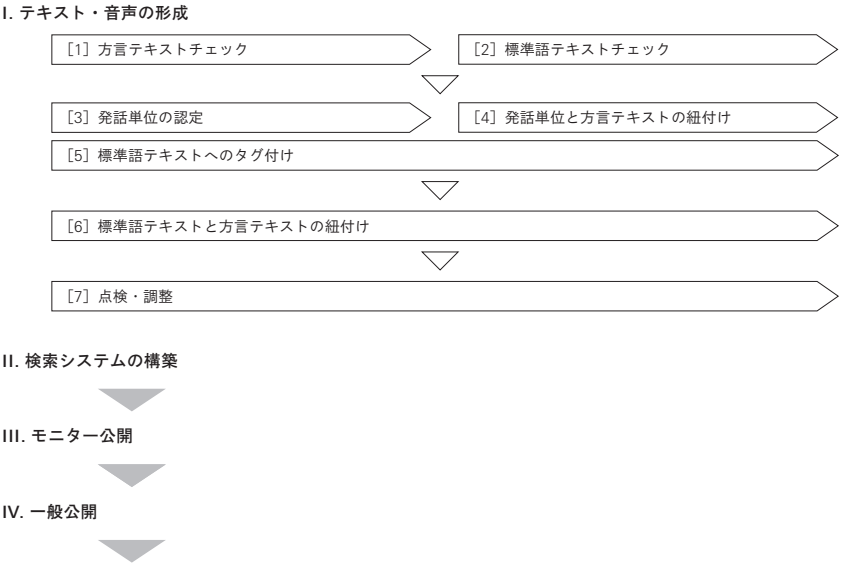
「I. デジタル化」では，原資料の保存とテキスト入力の利便性のため，手書き原稿を，スキャニング・撮影をして，画像ファイルとして保存する。手書き原稿の内容（方言文字化・共通語訳・注記・解説・談話情報・話者情報など）のテキスト入力をおこなう。カセットテープに録音された音声は，デジタル化し，電子ファイルとして保存する。「各地方言収集緊急調査」資料の場合は，約7,500本のカセットテープについて，サンプリング周波数22.050kHz，量子化ビット数16bitでデジタル化して，音声ファイルを作成した。なお，撮影・スキャニング・音声変換の方法・設定値や，テキスト入力の書式・表記，話者記号・発話番号の付与の仕方，記号の種類・使い方，注意事項などをマニュアルとしてまとめておく必要もある。

「II. データ整理」では，入力した方言文字化テキストと共通語訳テキストとの対応データを作成する。方言文字化・共通語訳の範囲に対応する音声データを，デジタル化した録音音声から切り出す。ひとまとまりの談話ごとに，談話番号・話者・発話番号・方言文字化・共通語訳・音声ファイル名などを対照できるファイルを作成する。原稿の記載から，収録した談話内容・話題・話者・収録地点・場所・日時といった情報を抽出し，談話情報としてまとめてデータベース化する。話者の性別・年齢・職業・間柄といった情報を集約し，話者情報としてまとめてデータベース化する。これらの情報は主要データとなる会話以外の部分であるが，コーパス作成のうえでも分析にあたっても，基本的で重要な情報である。以上の音声・方言文字化・共通語訳・談話情報などを総合して，ネイティブチェック用データセットを作成する。

「III. ネイティブチェック」は，データの信頼性を高めるため，当該方言を専門とする研究者やネイティブに依頼して，音声データをもとに，方言文字化データ・共通語訳データを照合・校閲してもらう。

次に，ネイティブチェック完了後のデータをコーパスデータにする段階は，次の【図2】のような手順をとる。談話データについて，テキストチェックとして，表記のゆれを修正し，方言文字化と標準語訳との対応を確認する。音声的な基準（0.2秒のポーズ）に従って区切りを設定し，発話単位として認定する。発話単位（音声）と方言テキストを紐付ける。発話単位認定，音

声と方言の紐付けと平行して，標準語へのタグ付けをおこなう。各方言における音声・方言テキスト・標準語テキストのセットの点検をおこなう。このような流れを経て，『日本語諸方言コーパス』データが作成される。木部暢子・佐藤久美子・中西太郎・中澤光平（2017）には，『日本語諸方言コーパス』構築のプロセスや，作業上の問題点，コーパスデータを用いた分析の注意点などが報告されている。



（木部暢子・佐藤久美子・中西太郎・中澤光平 2017: 59「図 1. CJD 作成の流れ」原図を編集）

【図 2】コーパス作成の流れ

コーパスデータの具体的な作成方針に関しては，「データ作成方針」（https://www2.ninjal.ac.jp/cojads/content/cojads_datamanual.pdf）に，『日本語諸方言コーパス』のデータの構成（収録データ，メタデータ，データの形式など），方言テキスト・標準語テキストの表記やタグ類，文節認定規則，タグ一覧について，詳細にまとめられている。このような指針に沿って整形およびタグ付けをおこない，メタ情報を付加することで，研究利用可能なコーパスデータが作成される。

2019年5月には，『日本語諸方言コーパス』（モニター版）が公開された。オンラインでは，コーパス検索アプリケーション「中納言」を通して利用することになるため，中納言に登録する必要がある。2020年3月には，webサイト『日本語諸方言コーパス（COJADS）』（https://www2.ninjal.ac.jp/cojads/index.html）が開設され，メタ情報付きテキストデータ（CSV形式）が「データDL」（https://www2.ninjal.ac.jp/cojads/index.html?targ=data）からダウンロードできる。

2.3 データの公開

大量のデータを整備して大規模なコーパスを作成するには，相当の資金や労働力が必要であり，少なからず時間もかかる。大規模コーパスの構築は，組織としての事業でなければ難しいのが現実である。ただ，個人でも，過去に収録した談話などをデータとして公にすることは可能であるし，意味のあることだと考える。一例をあげると，2011年11月から収録してきたロールプレイ会話の音声・文字化を，『方言ロールプレイ会話データベース』[1]（http://hougen-db.sakuraweb.com/）として公開し，随時データを追加している。

収録したロールプレイ会話は，「同性の親しい友人同士2名が電話で会話をおこなうもの」「同性の同輩2名と同性の先輩1名を1グループとして，2名がペアとなり，総当たりで，電話で会話をおこなうもの」である。「文句を言う」「頼む」「慰める」「誘う」「出欠を確認する」「手伝いを申し出る」「本を持っているか尋ねる」「遅刻の連絡」のいずれかの場面を設定している。話者は，若年層（大学生～20歳代）・高年層（60～70歳代）の女性・男性である。現時点での掲載地点は，首都圏・関西・秋田・愛知・相生・広島・熊本・人吉・大分・鹿児島・沖縄となっている。それぞれの会話の文字化を載せ，音声を再生して確認できる。調査方法，場面設定，話者の性別・出身地・生年といった話者情報をまとめ，文字化や音声処理についての凡例を示している。

収録にあたっては，会話や音声の公開などについての許諾をとり，公表にあたっては，話者の個人情報やプライバシーに十分配慮し，話者情報や会話

の内容が個人を特定できる形で公表されることのないよう，発話中の当事者の個人名は記号に置き換え，個人名にあたる音声は伏せ音に加工するなどの対応をおこなった。

この調査のモデルとなったのは，2006～2008年度に各地の大学で収録されたロールプレイ会話である。勧誘・依頼などの場面設定で，大学生2名が会話を実演したロールプレイの音声と文字化が，日高水穂（2009）『ロールプレイ会話データベース』（http://hougen.sakura.ne.jp/hidaka/kaiwa/）で公開されている。日高水穂（2012，2014）では，このデータを用いて配慮表現についての分析をおこなっており，収録したデータを研究データとして活用しつつ，公開するという姿勢が感じられる。

3. 談話コーパスの利用

3.1 談話データを利用した研究

談話資料は，内省や質問紙調査では把握するのが困難である言語形式や言語運用について，実際の使用場面における状況を確認できる点が有効である。話者本人も無意識のうちに用いていることばや，話者自身には意味の説明の難しい語や文法形式についても，会話から文脈や場面が読み取れることによって，観察者にとっては分析が可能になるという利点もある。また，談話には，発音，アクセント，イントネーション，語彙，文法形式，文表現など，いろいろな要素が含まれるため，1地点の記述・分析をおこなうのにも，多地点を比較するのにも，多面的な利用が可能なデータである。

ただし，談話資料については，次のような留意点もある。談話資料は，ありのままの方言を記録したものであっても，方言の一部分を切り取ったものに過ぎず，記録されている範囲のことしかわからない。質問紙調査では求める項目を効率的にたずねることができるのに対して，談話資料では求める項目が得られるとは限らない。また，談話の録音にあたっては，さまざまな場面にわたっての収録が難しいため，分析対象とする場面が限定される。

【表 2】談話データを利用した研究事例

対象単位・分野	地域	概要	利用資料	文献
イントネーション	東京	文中の文節末でのイントネーションの使用実態，使い分け	日本のふるさとことば集成，ほか	郡史郎（2016）
オノマトペ	全国	方言特有のオノマトペの出現状況，談話の特徴	日本のふるさとことば集成	三井はるみ・井上文子（2007）
対称詞	全国	形式・用法・使用量の地域差	日本のふるさとことば集成	山本空（2015）
接続詞	全国 13 地点	バリエーション，意味類型，語構成からみた方言特徴	日本のふるさとことば集成	甲田直美（2018）
助詞	広島，東京，大阪	対格標示のバリエーションの計量的把握	日本のふるさとことば集成	小西いずみ（2015）
授受表現	全国	使用頻度の分布状況，中央語の史的変遷過程を反映した地理的分布	全国方言資料，方言談話資料，ほか	日高水穂（2007）
推量表現	全国	形式のバリエーション，特徴と地域差の概観	日本語諸方言コーパス	木部暢子（2020）
副詞・応答詞・フィラー	富山県富山市，富山県砺波市	多機能形式「ナ（ー）ン」「ナモ」の談話における運用を記述	日本のふるさとことば集成，各地方言収集緊急調査資料	小西いずみ（2018）
あいづち	全国 8 地点	あいづち要素の出現傾向，頻度，使われる形式の特徴	日本のふるさとことば集成	舩木礼子（2016）
待遇表現	大阪	「デス」「マス」の転訛形出現率，語形，場面差	全国方言資料	村中淑子（2020）
談話構造	全国	結婚のあいさつを構成している要素を分析して，談話型を記述	方言研究ゼミナール幹事団編（1991）	沖裕子（2006）
談話展開	東京，関西	文末・接続詞・感動詞などをキーワードとして，談話展開パターンを分析	（個人収録データ）	久木田恵（1990）
談話展開	東京，大阪，仙台	計量的に談話標識の出現傾向を比較し，談話展開の方法の地域差を解明	（個人収録データ）	琴鍾愛（2005）
言語行動 感謝表現	全国	感謝表現・謝罪的感謝表現，場面と言語選択，親疎の関係と場面適切性	方言談話資料	沖裕子（2006）
言語行動 配慮表現	九州 4 地点	依頼における都市性と配慮性・積極性，働きかけに対する姿勢の異なり	方言ロールプレイ会話データベース，全国方言資料	松田美香（2021）
言語行動 配慮表現	秋田	依頼・受諾・断りにおける配慮の談話の構成要素，談話展開の型	ロールプレイ会話データベース	日高水穂（2014）
言語行動 弔問の会話	全国	弔問の会話の部分構成，発話要素の現れ方，儀礼性と心情性の地域差	全国方言資料	小林隆（2018）
言語行動 挨拶	全国	挨拶の定型性，挨拶会話の展開における地域的な違い，発想法の違い	全国方言資料	小林隆（2017）

このような制約があるにしても，現実の言語活動が得られるという利点は大きく，また，対象とする要素も多岐にわたるため，方言の実態を解明するうえで，談話資料は欠かすことのできない基本的なデータであるといえよう。

談話データは，文法形式の地域差，会話の中で用いられる敬語の形式や運用，あいづちやフィラーの出現状況，談話の構造や展開，発話機能の特徴，言語行動，など，さまざまな研究に用いられている。【表2】は，具体的な談話データを利用した研究の一部を，研究の対象とする言語の単位・分野別に挙げたものである。

3.2 談話データの分析の観点

すでに談話データを利用したさまざまな研究事例があるが，ここでは，『日本語諸方言コーパス』（モニター版）を検索し，談話データがどのような観点で利用できるかを整理してみる。そのことによって課題も見えてくる。対象としたのはいわゆる間投助詞である。具体的には，次のようなものである[2]。短単位検索の「助詞 - 終助詞」でヒットしたものから，前後の文脈を確認して，間投助詞と判断したものを採用した。検索結果を通して，音声を確認できるのは便利であった。

ホイチ Xサント Xサンガナ アッコン ウチガタニ キチョッタンジャ。／そして Xさんと Xさんがね あそこの 私の家に 来ていたんだ。
（大分県大分郡挾間町・女・1904年生・74歳）

ホイテ ウエノ アノ ソラヤマノノー イケノ ハタニ オッタ ユー コトオ シットライ。／そして 上の あの 空山の ね 池の そばに いた という ことを 知っているよ。　（愛媛県松山市・女・1914年生・67歳）

ン ソレオネー ミッカカンノ アイダニ アノ ナク ナリマス。／ン それを ねえ 三日間の 間に あの なく なります。
（京都府京都市・女・1912年生・69歳）

マー イマー トーキョーデ サーギ ヤッテルケンドサー アー ジシンガ キタラ ドコエ ニ ヒナン スル ニゲルナンツエ ユッタッテサー アー ユー ニゲル マモ アーニモ アリャー シネーガネ。／まあ 今は 東京で 騒ぎを やっているけれどさあ ああ 地震が 来たら どこへ ニ 避難する 逃げるなんて 言ったってさあ ああ いう 逃げる 間も なにも ありは しないじゃない。（埼玉県児玉郡上里町・男・1904 年生・77 歳）

ンー チョエト ナ n デルドヨ。 ビリーッテ ユー オンダド。／うん ちょいと なでるとね。 びりっと いう ものだって。
（秋田県湯沢市・女・1904 年生・73 歳）

もちろん，得られる間投助詞は，限られた話者・話題・場面に出現したものである。使用には個人差があり，臨時的に用いられたものもあるかもしれない。しかし，それはどのような資料にもあることなので，そのうえで利用のしかたを考えていきたい。簡単ではあるが何らかのヒントになればと思い，分布，使い分け，変化の方向などの観点でみていく。

3.2.1 分布の観点

『日本語諸方言コーパス』の地点別の間投助詞の出現状況を示したものが【表 3】である。代表的な形式として，ナ系・ノ系・ネ系・サ系・ヨ系・ヤ系・ユ系をたてた。これ以外のものはその他とした。用例数の多寡は問わず，該当形式が出現するかどうか，出現した形式の欄に印をつけて，分布の概観を見る。

ある程度の地点を並べて比べてみると，ひとつの地域にどのような形式のバリエーションがあるのか，特定の形式がどのような範囲で使われているのか，などに気がつく。密度は粗いが，多くの地点を比較すると，間投助詞の各形式がそれぞれ独自の分布領域を持っており，その地理的広がりや境界を全国的な視点で概観することができる。

【表 3】『日本語諸方言コーパス』における間投助詞の出現状況

地点	収録時間	話者	ナ系	ノ系	ネ系	サ系	ヨ系	ヤ系	ユ系	その他
北海道中川郡豊頃町	00:37:16	3人(男1・女2)	●		●	●				
青森県弘前市	00:36:34	3人(男2・女1)	●			●				●
岩手県遠野市	00:46:59	2人(男1・女1)	●			●	●			
宮城県仙台市	01:46:54	10人(男6・女4)	●		●	●		●		●
秋田県湯沢市	00:25:52	3人(男1・女2)	●				●	●		●
山形県東田川郡櫛引町	00:26:52	3人(男1・女2)	●	●			●			
福島県大沼郡会津高田町	00:51:07	3人(男3・女0)	●		●		●			●
福島県大沼郡昭和村	00:41:40	5人(男3・女2)	●			●	●			
茨城県高萩市	00:32:38	2人(男2・女0)			●					
茨城県水戸市	00:38:57	2人(男1・女1)			●					
栃木県日光市	00:34:47	3人(男1・女2)	●		●	●	●			●
群馬県前橋市	00:39:35	4人(男2・女2)			●	●				
群馬県甘楽郡下仁田町	01:02:33	3人(男3・女0)	●	●	●	●				●
埼玉県児玉郡上里町	00:38:03	3人(男2・女1)			●	●				
千葉県長生郡長生村	00:37:56	2人(男1・女1)	●		●	●	●			
東京都台東区	00:34:58	2人(男1・女1)	●		●	●				
神奈川県小田原市	00:44:12	4人(男3・女1)	●		●	●	●			
新潟県糸魚川市	00:36:45	4人(男3・女1)	●		●	●				
富山県砺波市	00:22:02	3人(男1・女2)			●					
石川県羽咋郡押水町	00:21:32	3人(男2・女1)			●					●
福井県勝山市	00:23:13	3人(男1・女2)		●	●					
山梨県塩山市	00:26:40	3人(男2・女1)			●					
長野県木曽郡開田村	00:26:40	3人(男2・女1)	●			●				
岐阜県中津川市	00:14:10	3人(男2・女1)		●						
岐阜県岐阜市	01:01:58	3人(男2・女1)		●	●					●
静岡県静岡市	00:23:49	3人(男2・女1)				●				
愛知県常滑市	01:32:17	3人(男2・女1)	●	●	●					
三重県志摩郡阿児町	00:28:21	4人(男3・女1)	●	●	●	●				

滋賀県甲賀郡甲賀町	00:39:34	3人（男2・女1）	●	●			●			
京都府京都市	00:24:00	5人（男3・女2）			●					
大阪府大阪市	00:28:20	6人（男4・女2）	●		●					
大阪府河内長野市	00:59:55	6人（男4・女2）	●		●		●			
兵庫県相生市	00:30:35	2人（男1・女1）	●		●	●				
兵庫県洲本市	00:28:33	4人（男3・女1）	●	●						
奈良県五條市	00:34:01	2人（男1・女1）	●		●					
和歌山県田辺市	00:31:46	3人（男1・女2）	●	●	●					
鳥取県米子市	00:35:20	2人（男1・女1）	●							
島根県仁多郡仁多町	00:35:20	3人（男1・女2）			●					
岡山県小田郡矢掛町	00:27:43	2人（男1・女1）	●	●						
広島県広島市	00:38:17	3人（男1・女2）	●	●	●					
山口県豊浦郡豊北町	00:37:19	3人（男2・女1）	●	●						
徳島県阿南市	00:36:51	3人（男2・女1）	●		●		●			
香川県観音寺市	00:36:06	3人（男1・女2）	●							
愛媛県松山市	00:32:56	2人（男1・女1）	●	●	●	●				
高知県高知市	00:33:47	2人（男1・女1）			●					
福岡県北九州市	00:29:29	4人（男2・女2）	●		●					
佐賀県佐賀市	00:20:59	3人（男1・女2）	●		●					●
長崎県平戸市	00:24:37	2人（男1・女1）	●		●					
熊本県熊本市	01:00:17	7人（男4・女3）	●							
熊本県球磨郡錦町	00:21:11	3人（男2・女1）	●							
大分県宇佐市	00:59:53	3人（男1・女2）	●							●
大分県大分郡挾間町	01:19:47	8人（男5・女3）	●							
宮崎県宮崎市	00:22:40	2人（男1・女1）	●				●			●
鹿児島県揖宿郡頴娃町	00:34:37	3人（男2・女1）	●		●					
沖縄県国頭郡今帰仁村	00:18:29	2人（男1・女1）				●	●	●		●
沖縄県平良市	00:18:29	2人（男1・女1）					●	●	●	●

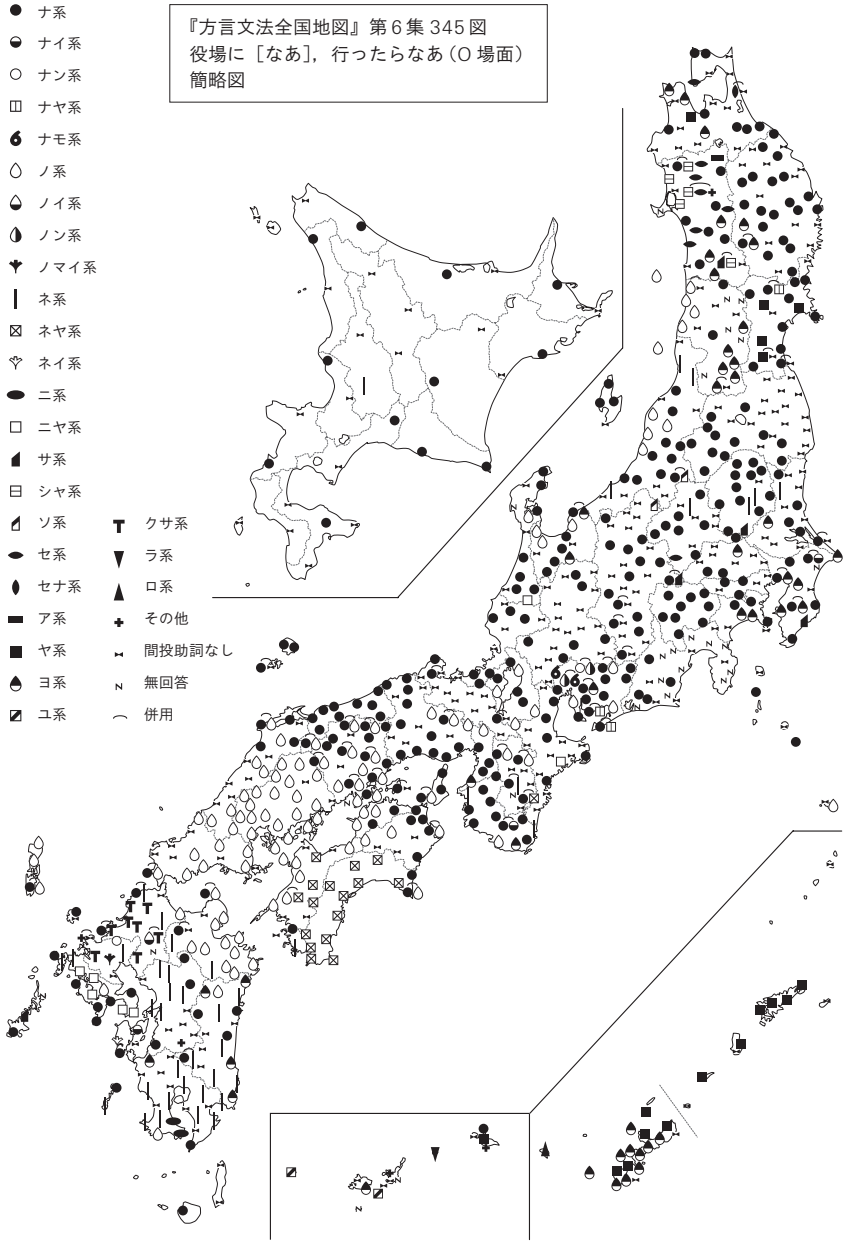
中納言検索結果データダウンロード 2020 年 12 月 19 日現在，
収録時間は web サイト参照，地点別の話者数・性別はデータを確認したうえでの異なり数

ナ系　北海道から九州までのかなり広い範囲，中部・北陸には少ない
ノ系　中部・近畿・中国・四国中心
ネ系　北海道から九州北部までの広い範囲，関東・北陸に顕著
サ系　北海道から中部に集中，西日本にはあまり見られない
ヨ系　東北・関東・沖縄に多い
ヤ系　東北・沖縄
ユ系　沖縄

間投助詞の全国分布を調べたものとしては，国立国語研究所編『方言文法全国地図』第6集343～348図の「間投表現」の言語地図がある。【図3】は，「345　役場に［なあ］，行ったらなあ（O場面）」の簡略図である。凡例・記号などは簡略図では変更している。質問文は次のようなものであった。「役場に①なあ」は，名詞を含む文節につく間投助詞，「行ったら②なあ」は動詞を含む文節につく間投助詞を問う設問である。

親しい友達にむかって，「今日，役場に①なあ，行ったら②なあ」のように言うとき，「役場になあ，行ったらなあ」のところをどのように言いますか。

「親しい友達」に対する「くだけた形式」の間投助詞として回答された形式は，ナ系・ノ系・ネ系・サ系・ヨ系・ヤ系・ユ系など，談話データに出現したものと共通しているが，分布状況には多少の相違がある。言語地図では，ナ系は北海道から九州までの広い範囲に見られるが，広島・山口ではノ系が大勢を占め，四国南西部には別語形が分布する。ノ系は中国・四国・九州・近畿の一部や東北・北陸の日本海側の各地にも広がっている。ネ系はおもに九州に分布する。サ系はほとんど現れない。このほか，特徴のある形式として，ネヤ系が高知・愛媛に，クサ系が福岡・佐賀に，ニヤ系が長崎などに見られる。これらの地域の特徴的な形式は，談話データには出現していない。



【図3】『方言文法全国地図』における間投助詞の分布

なお，『方言文法全国地図』の調査期間は1979～1982年，話者は最低限1925年以前に生まれた男性であり，『日本語諸方言コーパス』の原資料「各地方言収集緊急調査」の収録期間は1977～1985年度，話者は原則として60歳以上であるため，時期や話者に大きな差があるわけではない。

言語地図よりも談話のほうが，併用する形式が多く複雑な状況である。言語地図に現れた形式は地域の代表形と意識された形式であり，談話に現れた形式は言語使用の幅を反映した形式であり，調査方法の違いと言えるかもしれない。

3.2.2 使い分けの観点

茨城県高萩市（ネ），香川県観音寺市（ナ）などのように，使用される間投助詞が1種類のみの地点がある一方で，多くの地点では，同一地点で複数の間投助詞が使われている。中には，栃木県日光市（ナ・ネ・サ・ヨ・その他）や群馬県甘楽郡下仁田町（ナ・ノ・ネ・サ・その他）のように，多数の形式が併用されている地点もある。

東京都台東区では，【表4】のように，ナ・ネ・サが使われている。話者の性別によって集計してみると，ネは男女ともに多用されているが，サには多少の性別の偏りがあるように見てとれる。また，【表5】は，2021年1月に中納言に追加登録されたデータによる場面設定会話であるが，自由会話には現れなかったヨが男性に頻出し，ナも若干使用が増えている。

【表4】『日本語諸方言コーパス』
自由会話に現れる間投助詞（東京都台東区）

	男性	女性	計
ナ	1	0	1
ノ	0	0	0
ネ	68	85	153
サ	25	4	29
ヨ	0	0	0
計	94	89	183

【表5】『日本語諸方言コーパス』
場面設定会話に現れる間投助詞（東京都台東区）

	男性	女性	計
ナ	4	0	4
ノ	0	0	0
ネ	6	5	11
サ	7	1	8
ヨ	17	0	17
計	34	6	40

複数の形式が出現することについては，使用の傾向的性差や，自由会話と場面設定といった談話形態による違い，話題による差などが考えられそうである。そのほか，概観的に話すか，臨場感たっぷりに具体的に話すかといった話しぶり，会話場面の状況依存度の程度，などが影響する可能性もある。

間投助詞の場面差ということについては，国立国語研究所編『方言文法全国地図』第6集343～348図の「間投表現」の言語地図で，「親しい友達に対するくだけた形式」「近所の知り合いに対するやや敬意のある形式」「この土地の目上の人に対するもっとも敬意のある形式」を求める3場面を設定し，話し相手への待遇度の異なりによる間投助詞の形式をたずねている。その出現形式や分布領域を概観すると，全体的な傾向として，待遇度が高くなるにつれナ・ノが減少し，ネが増加する。しかし，場面にかかわらず同じ語形が用いられる地域もあれば，ナとネについて逆の使い分けが見られる地域もある。ナとノの併用地域についても，上位場面と下位場面でナ・ノの切り換えが逆になる地域が観察され，同じ形式の間投助詞でも地域により待遇価や場面の現れ方に違いがあることがわかる。

収録時期はかなり異なるが，参考までに，『方言ロールプレイ会話データベース』の先輩・後輩の会話を見ることにする。2015年収録，愛知県津島市・海部郡佐織町出身の高年層男性[3]の会話である。【表6】を見ると，同輩同士A・Bの会話では，ヨが使われているが，先輩であるCとの会話には，ヨは使われず，ネが多用される。話し相手によって，ヨからネへの切り換えが見てとれる。ヨよりもネのほうが待遇価が高いとみられる。

また，東京の大学に通う埼玉県出身の女性[4]が，友人と先輩に対して，同様の内容を話す談話では，【表7】のように，同輩同士A・Bの会話では，サ・ネが頻出しているが，話し相手が先輩Cになると，後輩から先輩に対しては，間投助詞はまったく現れない。その代わりか，文節末を延ばすような発音が頻繁に観察された。先輩から後輩にはサ・ネが少し使われているが，友人同士の会話に比べると非常に少ない。話し相手が友人から先輩に変わると，敬語が現れ，スピーチスタイルも変化する。間投助詞の現れ方もそれに伴う。首都圏の若年層女性では，待遇度が上がると，間投助詞は使われなくなる傾向があるとみられる。

【表 6】『方言ロールプレイ会話データベース』
間投助詞の使い分け（愛知・高年層・男性）

かけ手→受け手		かけ手	受け手
同輩→同輩	A → B	A ヨ 4 A ネ 1	B ヨ 2 B サ 1
同輩→同輩	B → A	B ヨ 4 B ナ 1	A ヨ 4
先輩→後輩	C → A	C　0	A ネ 2
先輩→後輩	C → B	C ネ 1	B ネ 5
後輩→先輩	A → C	A ネ 4	C ネ 1
後輩→先輩	B → C	B ネ 5	C　0

【表 7】『方言ロールプレイ会話データベース』
間投助詞の使い分け（首都圏・若年層・女性）

かけ手→受け手		かけ手	受け手
同輩→同輩	A → B	A サ 20	B サ 1
同輩→同輩	B → A	B サ 21 B ネ 3	A　0
先輩→後輩	C → A	C サ 3 C ネ 1	A　0
先輩→後輩	C → B	C サ 3	B　0
後輩→先輩	A → C	A　0	C　0
後輩→先輩	B → C	B　0	C　0

それぞれの地点で使われている形式のバリエーションを洗い出して，基本的な性質や特徴，用法を記述し，実際の使用状況や文脈がわかれば，これらのことも分析が可能となる。1 地点に複数の形式がある場合，それぞれの形式はどのように使い分けるか，社会的属性によって使用する形式が違っているか，場面によって切り換えられているか，その要因にはどういうものがあるか，が気になるところである。また，同じ形をとっていても地点ごとに機能が異なるという事象も観察されるため，形式の現れ方や使い分けにどのようなパターンがあるのか，類型化にも興味がわく。記述・分析のためには多量のデータが必要であり，比較のためには同じ条件で収録された複数の地点の多様な話者の十分な量のデータが必要となる。

3.2.3 変化の観点

過去に収録された談話資料は，当時の言語事象・言語生活を解明するだけではなく，現在の状況に至る方言の歴史を探究できる方言データでもある。収録方法が異なるなどの理由で，過去に収録されたデータと現在収録したデータを厳密には同列に論じることはできない場合もあるが，その方向性を捉えるという点では利用価値がある。

『日本語諸方言コーパス』の 1977 年収録の大阪府河内長野市では，次の

例のように間投助詞にナ・ネ・ヨが使用されるが，サは使用されていない。

アノネー　ショーガツノネー　アノー　ソノー　ハジメカラネ　ツマリ　アノー　シ　ジュンビカラー　ズーット　ショーガツノネ。　ナニ　ソレカラ　イッペン　アノー　ショーガツノ　イッ　ジューズメトカ　アノ　オジュートカナー　セチリョーリヤケド　オセチリョーリナ。　アー　ユー　モノオネ　イッペン　チョット　ハナシ　シテクレヘンカ。／あのねえ 正月のねえ あの その 初めからね つまり あの シ 準備から ずーっと 正月のね。なに それから いっぺんに あの 正月の イッ 重詰とか あの お重とかなあ お節料理だけれど お節料理な。ああいうものをね 一度 少し 話を してくれないか。

（大阪府河内長野市・男・1901年生・76歳）

『方言ロールプレイ会話データベース』では，2011年～2014年に，大阪府河内長野市出身の大学生を含む会話[5]を収録している。その会話には，ナ系の間投助詞も用いられているが，サが頻繁に出現している。

あのさー、ちょっと聞きたいことあんねんけどさー、あーのさ、今、ちょっとなー、探し物しててー、本やねんけどー、んで、おまえ、持ってへんかなーと思ってー。（大阪府河内長野市・男・1990年生・22歳）

や、今日、その、家出たときにさー、あの、チャリンコの、か、かごにさー、や、すっごいかわいい野良猫入っとってさー、そんなん、見つめるほかないやん。　（大阪府河内長野市・男・1992年生・22歳）

さまざまな時期に収録された談話データは，比較することによって，その地域の方言の歴史的な変化の過程を知る手がかりとなる。新たな変種の発生や，他方言の移入，標準語の影響などを解明することのできる情報となる。

杉村孝夫（2018）は，大分県方言の豊富な談話資料を用いて，依頼表現の世代差と経年変化に関する分析をおこなっている。大分県では，約60年間

に3度（1954～1958，1983～1985，2009～2011）にわたって同じ場面設定で収録された談話資料があり，それぞれの時期で，高年層（50～80代）・青年層（20～30代）・少年層（10代の中学生）の3世代の談話も収録している。そのため，世代差という見かけ上の時間の時間による変化と，実時間による経年変化の両方の変化を把握することができた。

また，杉村孝夫（2019）では，青森，東京，大阪，福岡について，『全国方言資料』と2010年代に追跡調査した談話の60年間の実時間の経年比較も試みている。このような実時間の言語変化を探るための追跡調査の地点や話題の選定にあたっては，椎名渉子・小林隆（2017）が役に立つ。

4. おわりに

『日本語諸方言コーパス』のような談話コーパスが構築されることによって，大量のデータに基づいて効率的に研究を進めるための基盤ができつつあり，方言研究は新たな局面を迎えようとしている。ただ，まだ，現状では，収録時間は十分ではなく，地点の密度も高いとは言えない。まずは，今後，「各地方言収集緊急調査」の未公開データを整備してデータ量を増やすことが急がれる。加えて，『方言録音資料シリーズ』『方言談話資料』など，国立国語研究所内で権利が処理できるものから『日本語諸方言コーパス』に追加していく方法が考えられる。

もちろん，今後，新たな収録をおこなえるような状況になれば，新規に談話を積極的に収集することができるであろう。既存資料は，話者の属性・人数・人間関係が地域によって違ったり，談話の内容・話題が地域によって異なっていることも多い。新規に収録する場合は，既存資料では限界があった点についてコントロールが可能となり，統一的な条件での談話収集が実施できる。さらに，新規収録の場合は，方言音声の録音とあわせて，談話場面を録画し，非言語情報を記録することも必要となる。『東日本大震災と方言ネット』など，身振りや視線などの非言語行動も含めて方言を総合的に記録し，

談話行動を総合的に分析する手がかりとする試みもあり，映像による資料は今後ますます増加すると思われる。なお，新規計画の場合でも，既存資料もあわせて活用できるような計画を立てるのが理想的である。

また，以前から各地の研究者や有志の人々は，個人で調査・収録をおこない，地域のことばを記録してきた。限定された地域のものが多いが，貴重なデータである。このような資料を発掘し，著作権者の許諾や話者のプライバシー保護に丁寧に対応しながら，権利関係について処理したうえでデジタル化していく努力も必要である。状況が許せば，話者本人や収録者・協力者と協働して，談話の収録や，録音資料の文字化，ネイティブチェックやフォローアップインタビューをおこなったり，方言音声・方言テキストを直接検索できるようにするための形態素辞書を作成したり，という市民参加型の活動も有意義である。

そして，これまで個別に収録・記録がおこなわれていた，さまざまな資料を集約・蓄積し，言語資源として整備してコーパスを構築し，あらゆる人々が広く共同利用できるように公開・維持・管理をおこなう基盤となる組織やシステムの必要性が高まってきている。

データに基づく実証的研究が重視される現在では，データを分析して研究成果として発表するとともに，そのもとになったデータを研究の科学的な証拠として公開することが一般的になってきている。それに加えて，後世への言語資源の継承という意味でも，収録データの保存・公開の側面にも関心を持っていただけたらと思う。

注

1 『方言ロールプレイ会話データベース』（http://hougen-db.sakuraweb.com/）のデータは，下記の課題の共同研究者，研究分担者，研究協力者によって得られたものである。また，調査地域の研究者・協力者・話者の方々にもご尽力いただいた。国立国語研究所萌芽・発掘型共同研究プロジェクト「方言談話の地域差と世代差に関する研究」（平成 22-25 年度）（プロジェクトリーダー：井上文子）共同研究者：熊谷智子・小西いずみ・高橋顕志・日高水

穂・松田美香・三井はるみ，研究協力者：酒井雅史・野間純平・森勇太／基盤研究（C）「方言ロールプレイ会話における談話展開の地域差に関する研究」（平成 25-27 年度）（研究代表者：井上文子）研究分担者：熊谷智子・小西いずみ・日高水穂・松田美香・三井はるみ，研究協力者：大久保北斗・酒井雅史・白坂千里・田中渉・利岡真帆・野間純平・森勇太・山本空／基盤研究（C）「地域的多様性の教材としての参加型方言データベースの構築」（平成 29 年度 – 平成 31 年度）（研究代表者：井上文子）研究分担者：小西いずみ・竹田晃子・日高水穂・松田美香・三井はるみ，研究協力者：山田豊樹

2 カタカナは方言文字化，漢字仮名交じりは共通語訳。（　）内は，収録地点・性・生年・収録時年齢。話者記号，あいづち，タグは省いて示す。スペースも適宜変更した。 n は前鼻音を表している。以下，『日本語諸方言コーパス』の用例については同様。

3 愛知・高年層・男性・グループ 2（A：愛知県海部郡佐織町・男・1945 年生・70 歳，B：愛知県津島市・男・1944 年生・71 歳，C：愛知県海部郡佐織町・男・1929 年生・86 歳）

4 首都圏・若年層・女性・グループ 4（A：埼玉県川口市・女・1991 年生・22 歳，B：埼玉県比企郡鳩山町・女・1991 年生・22 歳，C：埼玉県秩父市・女・1991 年生・22 歳），首都圏・若年層・女性・グループ 5（A：埼玉県狭山市・女・1995 年生・18 歳，B：埼玉県川越市・女・1994 年生・19 歳，C：埼玉県所沢市・女・1991 年生・22 歳）

5 方言文字化は漢字仮名交じりによって表記。（　）内は，出身地・性・生年・収録時年齢。話者記号，あいづち，凡例の記号は省いて示す。スペースも適宜変更した。以下，『方言ロールプレイ会話データベース』の用例については同様。

引用文献

井上文子（2013–）『方言ロールプレイ会話データベース』http://hougen-db.sakuraweb.com/

沖裕子（2006）『日本語談話論』，大阪：和泉書院.

木部暢子・佐藤久美子・中西太郎・中澤光平（2017）「『日本語諸方言コーパス』の構築について」『言語資源活用ワークショップ発表論文集』1，57–68.

木部暢子（2020）「諸方言コーパスに見るモダリティ形式のバリエーション―推量表現の地域差―」田窪行則・野田尚史編『データに基づく日本語のモダリティ研究』41–62，東京：くろしお出版.

久木田恵（1990）「東京方言の談話展開の方法」『国語学』162，1–11.
琴鍾愛（2005）「日本語方言における談話標識の出現傾向―東京方言、大阪方言、仙台方言の比較―」『日本語の研究』1（2），1–18.
郡史郎（2016）「間投助詞のイントネーションと間投助詞的イントネーション―型の使い分けについて―」『言語文化研究』42，61–84.
国立国語研究所はなしことば研究室編（第1巻のみ地方言語研究室編）（1965–1973）『方言録音資料シリーズ』全15巻，非売品.
国立国語研究所編（1978–1987）『方言談話資料』全10巻，東京：秀英出版.
国立国語研究所編（2006）『方言文法全国地図6』，東京：財務省印刷局.
国立国語研究所（2019–）『日本語諸方言コーパス』（Corpus of Japanese Dialects：COJADS）https://www2.ninjal.ac.jp/cojads/
国立国語研究所編（2001–2008）『全国方言談話データベース 日本のふるさとことば集成』全20巻，東京：国書刊行会.
甲田直美（2018）「接続詞の語形変化と音変化―方言談話資料からみた接続詞のバリエーション―」小林隆編『コミュニケーションの方言学』271–291，東京：ひつじ書房.
小西いずみ（2015）「広島市方言の対格標示―談話資料による計量的把握―」『国語教育研究』56，13–24.
小西いずみ（2018）「富山県方言の「ナ（ー）ン」「ナモ」―否定を表す多機能形式の談話での運用―」小林隆編『感性の方言学』253–271，東京：ひつじ書房.
小林隆（2017）「談話からみた挨拶の定型性―「おはよう」の地域差をめぐって―」『方言の研究』3，77–101.
小林隆（2018）「儀礼性と心情性の地域差―弔問の会話に見る―」小林隆編『コミュニケーションの方言学』65–92，東京：ひつじ書房.
椎名渉子・小林隆（2017）「談話の方言学」小林隆・川﨑めぐみ・澤村美幸・椎名渉子・中西太郎『方言学の未来をひらく―オノマトペ・感動詞・談話・言語行動―』207–337，東京：ひつじ書房.
杉村孝夫（2018）「大分県方言の依頼談話」小林隆編『コミュニケーションの方言学』115–152，東京：ひつじ書房.
杉村孝夫（2019）「不祝儀の場面における談話の変容―大分県3時期の談話を中心として―」『方言の研究』5，65–89.
東北大学方言研究センター（2012–）『東日本大震災と方言ネット』https://www.sinsaihougen.jp/
日本放送協会編（1999）『CD-ROM版 全国方言資料』，東京：日本放送出版協会.
日高水穂（2007）『授与動詞の対照方言学的研究』，東京：ひつじ書房.

日高水穂（2009）『ロールプレイ会話データベース』http://hougen.sakura.ne.jp/hidaka/kaiwa/

日高水穂（2012）「「察し合い」の談話展開に見られる日本語の配慮言語行動」三宅和子・野田尚史・生越直樹編『「配慮」はどのように示されるか』91–112，東京：ひつじ書房.

日高水穂（2014）「談話の構成から見た現代語の配慮表現」野田尚史・高山善行・小林隆編『日本語の配慮表現の多様性―歴史的変化と地理的・社会的変異―』261–278，東京：くろしお出版.

方言研究ゼミナール幹事団編（1991）『方言資料叢刊 第1巻 祝言のあいさつ』，広島：広島大学教育学部国語教育学研究室方言研究ゼミナール.

舩木礼子（2016）「方言談話におけるあいづちの出現傾向―老年層方言談話資料から―」『方言の研究』2，165–191.

松田美香（2021）「九州4地点の依頼談話における配慮表現と積極的言語行動―九州における方言談話の特徴と分布―」『国立国語研究所論集』20，1–20.

三井はるみ・井上文子（2007）「方言データベースの作成と利用」『シリーズ方言学4 方言学の技法』39–89，東京：岩波書店.

村中淑子（2020）『研究叢書530 関西方言における待遇表現の諸相』，大阪：和泉書院.

山本空（2015）「方言談話における対称詞の使用量の地域差」『国文学』100，482–466.

（いのうえ・ふみこ　国立国語研究所准教授）