

## 「通時音声コーパス」は可能か

丸山 岳彦 (国立国語研究所 言語資源研究系)<sup>†</sup>

### Possibility of a Diachronic Corpus of Spoken Japanese

Takehiko Maruyama (Dept. Corpus Studies, NINJAL)

#### 要旨

通時コーパスとは、通常、書き言葉を対象としたものが想定される。では、話し言葉を対象とした通時コーパス、すなわち「通時音声コーパス」はどのように実現可能だろうか。本稿では、「通時」「音声」「コーパス」という3つの条件について検討した後、「通時音声コーパス」の実現によってどのようなことが明らかになるのかについて、具体的な分析例を交えながら、その見通しを示す。

#### 1 はじめに

2004年に『日本語話し言葉コーパス』(CSJ)が完成・公開され、朗読音声ではない自然な発話(自発音声)の研究が飛躍的に進んだ。約651時間、752万語の音声データを収録したCSJは、音声学・音韻論・文法論などの音声言語研究に対して新しい研究データを提供しただけでなく、社会言語学におけるバリエーションの研究、音声認識・音声翻訳システムにおける音声処理・言語処理研究など、幅広い分野で利用されてきた。一方、CSJに収録された音声の中心が独話(モノログ)であったため、日常会話を収録した大規模コーパスの開発を求める声が以前から根強くある。これに対して現在、国立国語研究所で2016年度から始まる次期プロジェクトの一つとして、さまざまな場面における日常会話を大量に収録した会話コーパスの構築・公開が計画されている(小磯他, 2015)。CSJに加えて、日常会話の音声コーパスが整備されれば、現代日本語の話し言葉を対象とする言語資源がさらに充実したものになり、話し言葉研究のさらなる拡大・深化が期待される。

上記のような現状を踏まえた上で、ここでは少し視点を変えて、「通時音声コーパス」の構築は可能か、という問題について考えてみたい。通常「通時コーパス」と言えば、書き言葉を対象としたものが想定されるだろう。これに対して、「話し言葉の通時コーパス」、すなわち、さまざまな音声資料を時代ごとに集積し、話し言葉の史的研究に利用できるようなコーパスの構築は、実現可能だろうか。それは、話し言葉の研究に何をもたらすだろうか。本稿では、「通時音声コーパス」が満たすべき条件とその問題点を示した上で、具体的な分析例を交えながら、その可能性について論じる。

#### 2 「通時音声コーパス」の条件と制約

はじめに、「通時音声コーパス」を実現するために必要な条件について考えてみよう。ここでは、「通時」「音声」「コーパス」という3つに分けて、その条件を考える。

まず「通時」という点で言えば、「通時音声コーパス」は、複数の時代に録音された音声資料の時系列的な集積でなければならない。その縦断的な分析により、日本語の話し言葉の史の変遷を研究するために整えられた資料群である必要がある。次に、「音声」という観点からは、録音資料そのものが聴取できる状態になっていなければならない。音声コーパスの本質は音声データそのものであり、音声を文字化した転記テキストのみの集積は、真の意味での音声コーパスとは言えない。そしてその音声データは、可能な限り良好な音質で、特に会話の場合は話者ごとに別トラックの録音になっていることが望ましい。そして、「コーパス」であるからには、「用例を大量に偏りなく収集して電子

<sup>†</sup> maruyama @ ninjal.ac.jp

化し、検索用情報を付加したもの(前川, 2013)」というコーパスの定義を満たす必要がある。すなわち、さまざまなタイプの録音資料が大量にデジタル化され、その転記テキストや形態論情報、時間情報、話者情報、種々のメタデータなどがアノテーションされた状態であることが望ましい。

ところが実際には、当然のことながら、上記の条件を十全に満たすことは極めて困難である。例えば、「複数の時代に録音された音声資料」という点については、そもそも録音技術が開発されたのが19世紀後半、一般に普及し始めるのが20世紀に入ってからという事情を考えれば、通時音声コーパスは、20世紀以降に録音された音声のみに対象が限られるという制約がある<sup>1</sup>。書き言葉を対象とする通時コーパスが上代日本語(8世紀)以降の言語資料を扱えるのに対して、通時音声コーパスは扱える範囲が極めて狭く限定されることになる。次に、「音声そのものが参照できる状態」という点については、古い時代になるほど良質な音声データが期待できないという制約がある。後にも述べるように、1950年代に国立国語研究所が作成したさまざまな場面における会話の録音資料は、日常の会話場面をその場で実況録音したものであり、録音レベルの小ささやノイズの混入などによって聞き取りが困難な箇所も少なくない。また、デジタル化されていない音声資料は、原メディアの劣化などによって近い将来に聴取できなくなる可能性があるため、デジタル化の作業が急務となるが、そこには実務的・コスト的に大きな問題が生じることになる。さらに、「大量に偏りなく」という点については、これも当然のことながら、時代を遡るほど現存する音声資料の量は少なくなるため、そこに大規模コーパスとしての大量性や均衡性を期待することはできない。元来、話し言葉コーパスに均衡性を求めるのは非常に困難であるが(前川, 2013)、歴史的な音声資料を対象とする場合、その傾向はより顕著なものとなる。さらに、音源が残されていたとしても、著作権の問題または遺族の意向などによって、当該の音声資料を公開できない、という場合もある。

一口に「通時音声コーパス」と言っても、残されている音声資料が限られていることを考えると、その量の少なさや均衡性・多様性の偏りなどには目をつぶるしかない。言い換えれば、現存する音声資料をできるだけ幅広く収集し、それで賄うしかない。これは、例えば、上代日本語の研究者が限られた資料の中で研究せざるを得ない、という事情と同様である。古い言語資料を利用しようとする場合、このような資料の量的な制約は、必然的について回るものと言える。

そのような制約を踏まえた上で通時音声コーパスの実現を目指す場合、必要となるのは、できるだけ多様な音声資料の収集と、それを分類するメタデータの設計(丸山, 2012)という2点だろう。このうち前者については、現存する音声資料をできるだけ掘り起こし、研究資料として地道に整備していくしかない。国立国会図書館がウェブ上で公開している「歴史的音源<sup>2</sup>」のうち、演説や講演などの音源資料はその有力な候補の一つになるだろう。一方、後者については、音声資料をどのような分類基準によって整理し、話し言葉の位相の中に位置づけていくか、という点を検討することが求められる。収録時期はもちろんのこと、独話と会話、発話場面、発話者(性、年齢、出身地)、聞き手との関係、スタイルの高低、自発性の度合いなど、CSJで詳細に付与されたメタデータも参考にしながら、多様な音声資料を多角的に分類・分析するための指標の策定が必要となる。

### 3 「通時音声コーパス」が可能にする研究の事例

以下では、仮にある程度の規模を持った「通時音声コーパス」が実現されたと仮定し、それによってどのような言語研究が可能になるかという点について考えてみたい。ただし、20世紀以降の話し言葉を通時的に俯瞰できる音声コーパスは現在のところ存在しないので、ここでは限られた音声資料に基づいて分析例を示し、そこから将来的な展望を述べることにする。

<sup>1</sup> 金澤(2015)によれば、現存する最古の日本語録音資料は、1900年に川上音二郎一座が欧米興行を行なった際に録音した「オッペケペー節」であるという。

<sup>2</sup> <http://rekion.dl.ndl.go.jp/>

### 3.1 分析対象データ

ここでは、分析対象データとして、CSJに加えて、以下の2つの音声資料を用いる。

1. 「想隆社アカデミックリソースシリーズ 貴重音源コレクション 岡田コレクション I」
2. 『談話語の実態』録音資料

以下、前者を「岡田コレクション」、後者を「談話語データ」と略称する。

「岡田コレクション」とは、明治後期から昭和前期にかけて SP レコードに録音された音声資料のデジタル音源である。岡田則夫氏の収集した SP レコード 3.5 万枚のうち、165 音源、18.5 時間分の音声資料がデジタル化され、市販されている<sup>3</sup>。これらの音声データはすべて独話であり、演説、講演、講話、実況、法話、朗読などに分類される。音質の悪さにより、音声不明瞭なところも散見されるが、約 100 年前の音声資料の集積として、貴重な研究データとなることは間違いない。

この音声データに対して、国立国語研究所共同研究プロジェクト「多角的アプローチによる現代日本語の動態の解明」（2009～2015 年度、リーダー：相澤正夫）の中で、金澤裕之氏によって「岡田コレクション」の転記テキストが作成された。現在、その研究成果を収めた論文集の刊行が予定されている（相澤・金澤, 2015 予）。ここでは、このうち「演説」「講演」として分類された 109 講演、合計 14.5 時間分の音声データを用いる。異なり話者数は 76 人である。音声データの例を、表 1 に示す。

表 1: 「岡田コレクション」に収録された音声資料の例

| 発表年  | 講演者 (生年)     | 講演タイトル        | 収録時間    |
|------|--------------|---------------|---------|
| 1915 | 尾崎 行雄 (1858) | 司法大臣尾崎行雄君演説   | 0:28:10 |
| 1916 | 大隈 重信 (1838) | 憲政ニ於ケル世論ノ勢力   | 0:17:14 |
| 1926 | 後藤 新平 (1857) | 政治の倫理化        | 0:12:54 |
| 1931 | 犬養 毅 (1855)  | 強力内閣の必要       | 0:04:09 |
| 1937 | 林 銑十郎 (1876) | 国民諸君ニ告グ       | 0:06:13 |
| 1941 | 近衛 文麿 (1891) | 日独伊三国条約締結に際して | 0:10:25 |

次に、「談話語データ」とは、国立国語研究所で 1950 年代から 1960 年代にかけて作成された録音資料である。国立国語研究所では、1948 年の設立当初から、話し言葉（東京方言および各地方言）の調査が進められていた。録音機を導入した調査の最も古いものは、記録を見る限り、1950 年 10 月に実施された福島県白河市での調査のようである。さらに、1952 年から「話しことば研究室」で開始された調査・研究では、さまざまな場面における日常談話が録音され、イントネーション、語・文節・文の長さ、文の構造、語の種類・使用度数・用法などが分析された。この研究成果は、1955 年の『談話語の実態』、1960・1963 年の『話しことばの文型 (1)(2)』という 3 冊の報告書にまとめられている（国立国語研究所, 1955, 1960, 1963）。このうち『談話語の実態』では、約 30 時間分の録音資料が作成され、うち約 10 時間分が分析に用いられた。

当時の録音資料は、現在、その大半がデジタル化されているものの、研究に用いるには未整理のままの状態になっている。筆者は以前、このうちの一部を抜き出し、音声を書き起こして新規に転記テキストを作成した。この「談話語データ」に含まれるのは、会話が 33 件で合計約 19.5 時間分、独話が 21 件で約 17 時間分である。ここに含まれる音声資料（略称）の例を、図 1 に挙げる。このうち会話は、街頭や室内でマイクで録音された一般人の雑談音声の主である。一方、独話はすべて当時国立国語研究所で行なわれた講義・講演などの音声である<sup>4</sup>。

<sup>3</sup> <http://www.nichigai.co.jp/database/sp/index.html>

<sup>4</sup> ただし、特に会話については、話し手の年齢、職業、出身地などを記録したフェイスシートや、正確な録音日、録音場所などの記録が残っていない（現時点では見つからない）ため、音声資料の詳細については不明なことが多い。

図 1: 「談話語データ」に収録された音声資料の例

会話: 「九段高校生」「三人の青年」「一研雑談」「鎌倉主婦」「ジイサン・バアサン」「魚屋小僧」  
 独話: 「助詞・助動詞」「国語講義」「日本語のアクセントほか」「新庁舎開き記念講演会」「国立国語研究所創立十周年記念講演会」

これら2つのデータの規模を、表2に示す。なお、「岡田コレクション」は収録時期によって「大正期」「昭和1ケタ」「昭和10年代」という3つに区分する。また「談話語データ」は独話と会話に分ける。総語数は、UniDic2.1.2+MeCab0.996による形態素解析の結果から補助記号を除いた数で示す。

表 2: 「岡田コレクション」と「談話語データ」の規模

|        | 岡田コレクション<br>(1915~1944年) |         |         | 談話語データ<br>(1950~1960年代) |          |
|--------|--------------------------|---------|---------|-------------------------|----------|
|        | 大正期                      | 昭和1ケタ   | 昭和10年代  | 会話                      | 独話       |
| 講演数    | 19                       | 52      | 38      | 33                      | 21       |
| 異なり話者数 | 16                       | 42      | 30      | (不明)                    | (不明)     |
| 収録時間   | 3時間                      | 6時間     | 6時間     | 19.5時間                  | 17時間     |
| 総語数    | 23,022語                  | 46,998語 | 49,070語 | 218,497語                | 182,619語 |

以下では、「岡田コレクション」「談話語データ」「CSJ」という3種類の音声資料を用いて、いくつかの分析例を示しつつ、「通時音声コーパス」の可能性について述べる。

### 3.2 イントネーションの分析

はじめに、イントネーションについて見てみることにする。ここでは、「談話語データ」に見られた、図2のようなイントネーションの型を取り上げよう。これは、1957年に録音された「三人の女性」という音声資料の中に現れた、「そしてーみりんとね、卵の黄身ね、それ使ってね、すり鉢でするのよ」という発話のピッチ曲線である。図を見ると、「黄身ね」の「ね」、「するのよ」の「よ」の部分、すなわち一部の句末・発話末において、ピッチが急激に上昇していることが分かる。当然、この上昇調は聞き手に対する質問や疑問を表すものではない。

図 2: 句末の上昇イントネーション (談話語データ「三人の女性」)

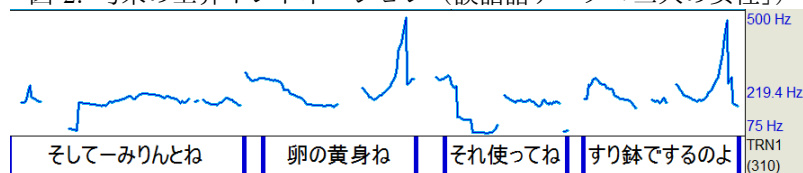
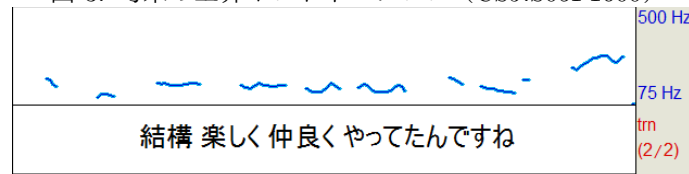


図2の上昇調を聞いた筆者がすぐに想起したのは、古い邦画の中で女優が話している場面であった。例えば、小津安二郎監督『東京物語』の中で、原節子が発話している台詞の中に、このような急激な上昇調が多数観察される。1950年代の若い女性は、このような発話が自然だったのだろうか。

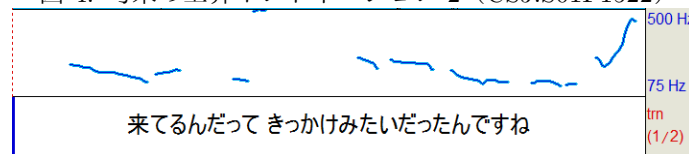
非疑問の文脈における句末の上昇調は、現代の話し言葉を取録したCSJでも観察することができる。例えば、図3の発話「結構楽しく仲良くやってたんですね」には、末尾に上昇調が認められる(句末境界音調はH%)。しかしながらこの上昇は、図2のような、急激な上昇調と同じとは認め難い。

図 3: 句末の上昇イントネーション (CSJ:S05F1600)



一方、同じく CSJ で観察された図 4 の発話「来てるんだって きっかけみたいだったんですね」の末尾に見られる上昇調は、図 2 の急激な上昇調に近いように思われる。

図 4: 句末の上昇イントネーション 2 (CSJ:S01F1522)



ここで注意したいのは、図 3 の話者の生年は 1970 年代前半（収録時は 20 代後半）、図 4 の話者の生年は 1940 年代後半（収録時は 50 歳前後）で、両者の間に 25 年ほどの年齢の開きがあるという点である。考えてみると、図 2 や図 4 のような句末の急激な上昇調は、現代でも高齢の女性の発話で観察されることがある。客観的な裏付けはないが、印象として、上品な高齢の女性が少し気取って話すような場面で、図 2、4 のような上昇調が現れるように思われる。さらに、周囲の子どもに図 2 の音声聞かせたところ、「おばあちゃんが話してるみたい」という印象が聞かれた。

図 2 の発話者である女性は、1950 年代の録音時に 20 歳代半ばだったとすれば、現在は 80 歳代になっている計算になる。ここから推測できるのは、彼女らは現在でも当時のイントネーションを（部分的に）保持しており、それを今でも使っている、ということである。それが現代の若い世代には「おばあちゃんみたい」あるいは「古い邦画の女優みたい」に聞こえることになる。若い世代の中で新しいイントネーションの型が出現し、古い型が使われなくなっていくという推移を考えれば、図 2 のような上昇調がどこかで衰退し、若年層が使わなくなった時期があるはずである。しかしその時期については、さらに多くの音声資料を準備し、縦断的・定量的に分析してみないと分からない。

### 3.3 文法形式の分析 (1)

次に、話し言葉の中に現れる文法形式に着目してみよう。ここでは、助動詞「まする」という例を取り上げる。「まする」は、近世初期に「ます」への移行が始まった形式と言われるが（服部, 2011）、「岡田コレクション」の中には、次のような「まする」の例が散見される。(1) は文末、(2) はト節、ガ節の述語句に「まする」が現れている。

(1) 明治 17 年、先帝陛下の御齡お五つの頃と記憶を致しております。

(間部詮信「大行天皇御幼兒を偲び奉りて」1927 年)

(2) 今日、新聞などを見ますると、誠に嘆かわしいことがたくさんありまするが、一に良心を顧みないで悪魔の聲にだまされて... (牧野元次郎「良心運動の第一声」昭和 10 年代)

一方、「談話語データ」の中にも、次のような「まする」の例が観察された。それぞれ、カラ節、ト節、ケレドモ節の述語句に「まする」が現れている。

- (3) 非常に予算の窮屈な、あー、時代でありますから、えー、それでもって...  
(山本有三「国立国語研究所十周年記念式典」1959年)
- (4) ラジオニュースの書き方というような本を見ますと、えー、ニュースには...  
(波多野完治「新庁舎開き記念講演会」1962年)
- (5) 新しい字引きが二〇万語を収載すると書いてありますけれども、その中の二万語しか...  
(林大「新庁舎開き記念講演会」1962年)

ただし、(1)~(5)に挙げた話者の同じ講演中には、「感激致しておる次第であります。」(間部)、「論語のうちであったかと思ひますが」(牧野)、「難しいんでありますから」(山本)、「聞いておりますと」(波多野)、「差もありますけれども」(林)という用例があることから、同一話者内において「ます」と「まする」の使用に揺れが生じていると言える<sup>5</sup>。

そこで、「岡田コレクション」と「談話語データ」の独話、CSJの「コア」に含まれる独話(177講演、約41時間分)を対象として、「まする」と「ます」の出現数を集計した。結果を表3に示す。

表3: 各資料における「ます」「まする」の出現数

|     | 岡田コレクション    |             |             | 談話語データ<br>独話  | CSJ<br>コア 独話 |
|-----|-------------|-------------|-------------|---------------|--------------|
|     | 大正期         | 昭和1ケタ       | 昭和10年代      |               |              |
| ます  | 271 (86.6%) | 752 (89.8%) | 903 (92.9%) | 3,918 (98.8%) | 5,604 (100%) |
| まする | 42 (13.4%)  | 85 (10.2%)  | 69 (7.1%)   | 48 (1.2%)     | 0 (0%)       |

「まする」は大正期には13.4%を占めていたが、時代が下るに従って「ます」に置き換わっていく様子が見て取れる。現代のCSJでは「まする」の用例は一つも見つからなかった。

次に、「岡田コレクション」「談話語データ」の2つを対象として、「まする」にどのような要素が後接しているかを集計した。上位10位までの結果を表4に示す。「助。」は助詞を表す。

表4: 「まする」に後接する要素

| 岡田コレクション   |            |            | 談話語データ<br>独話 |
|------------|------------|------------|--------------|
| 大正期        | 昭和1ケタ      | 昭和10年代     |              |
| 11 と(接続助.) | 28 ば(接続助.) | 16 句点(文末)  | 13 けれども      |
| 6 句点(文末)   | 11 が(接続助.) | 13 が(接続助.) | 8 から(接続助.)   |
| 5 が(接続助.)  | 10 名詞句     | 8 と(接続助.)  | 7 と(接続助.)    |
| 4 ゆえに      | 7 と(接続助.)  | 7 ば(接続助.)  | 6 し(接続助.)    |
| 3 ならば      | 5 に(格助.)   | 5 名詞句      | 6 名詞句        |
| 3 けれども     | 5 から(接続助.) | 4 の(準体助詞)  | 6 ば(接続助.)    |
| 3 から(接続助.) | 5 句点(文末)   | 4 から(接続助.) | 4 が(接続助.)    |
| 2 の(準体助.)  | 4 けれども     | 4 か(終助詞)   | 2 に(助動詞)     |
| 2 に(格助詞)   | 3 や(終助.)   | 3 や(終助詞)   | 1 ために        |
| 1 ば(接続助.)  | 3 という(引用節) | 3 に(格助詞)   | 1 ゆえに        |

「岡田コレクション」では文末に「まする」が現れる場合が見られるのに対して、「談話語データ」では文末位置の「まする」は皆無であった。これは、国会会議録に現れる「まする」を調査した上で

<sup>5</sup> なお、各話者の生年は、間部詮信が1878年(明治11年)、牧野元次郎が1874年(明治7年)、山本有三が1887年(明治20年)、波多野完治が1905年(明治38年)、林大が1913年(大正2年)である。

「主文末で用いられることがほとんどないという顕著な特徴がある」と述べた服部 (2011) の見方と符合する。一方、「岡田コレクション」では低い順位にある接続助詞ケレドモに後続する場合は、「談話語データ」ではトップとなっている。時代が下るにつれて文末で言い切る形が避けられ、並列節（ケレドモ節）でつなぐ形が好まれるように変化した、ということだろうか。

これらの観察から考えると、「まする」は、講演や講義、演説など、改まった発話スタイルの独話の中で、比較的少数の話し手により用いられていたことが予想される。ここで言う「少数の話し手」の条件には、おそらく、生年が強く影響しているだろう。すなわち、時代が進むにつれて若い世代の中で「まする」が消失していったということである。しかしながら、「まする」という形がいつごろ独話の中から消失していったのか、何年生まれの話し手までが「まする」を保持していたのか、という点については、データ量が不足しているため、現時点では分からない<sup>6</sup>。

### 3.4 文法形式の分析 (2)

さらに、終助詞の出現状況について見ておこう。ここで取り上げるのは、「談話語データ」の会話における、以下のような終助詞（の接続）の例である。

- (6) a. 私だったら九州に行きたいわ。（「相模女子大生」）
- b. 払うとしたら大変ですわね。（「友の会」）
- c. あんたんとこのお魚、美味しいわよ。（「魚屋小僧」）
- d. 先生とお話してきましたのよ。（「鎌倉主婦」）

これらはいずれも、女性の発話の末尾に現れた終助詞である。なお、この場合の「わ」「のよ」は、下降調ではなく上昇調（非疑問）である。現代の若い世代の女性の会話で、「わ」「わね」「わよ」「のよ」などの形が現れる場面は、非常に想定しにくいように思われる。もしくは、ふざけて「お嬢様」を演じているような場面（役割語としての使用）が想起される。

一方、先のイントネーションの場合と同様、高齢の女性が話者であると想定すると、これらはかなり自然に聞こえるように思われる。印象として、上品な高齢の女性が少し気取って話すような場面において、「～するわ。」「～したのよ。」と上昇調で話すのは、非常に自然に感じられる。

ここで、「談話語データ」と、CSJに含まれる対話（58対話、約12時間分）とを比較してみよう。である。発話末に現れる終助詞を抽出し、その一部を両者で比較した結果を表5に示す。

表 5: 対話の発話末に現れる終助詞

|            | ～わ  | ～わね | ～わよ | ～のよ | ～よ    | ～ね    |
|------------|-----|-----|-----|-----|-------|-------|
| 談話語データ（会話） | 153 | 296 | 116 | 296 | 1,675 | 5,752 |
| CSJ（対話）    | 4   | 2   | 0   | 0   | 391   | 4,165 |

表5からは、「わ」「わね」「わよ」「のよ」の出現数が、CSJよりも「談話語データ」の側に圧倒的に多いという事実を見て取ることができる。なお、ここでは両者ともイントネーションの型を考慮していないため、上昇調の用例は、特にCSJでさらに少なくなるはずである。

このような発話末尾の終助詞がいつごろから若年層で用いられなくなったのか、という点についても、やはり、さらに多くの音声資料を検討してみなければ分からない。

<sup>6</sup> 服部 (2011) は、現代でも国会会議録の中で少数ながら「まする」の用例が観察されることを報告している。

#### 4 まとめ

以上、本稿では、「岡田コレクション」「談話語データ」「CSJ」という3つの音声資料を用いて、イントネーションや文法形式に関する分析例を示してきた。実際の音声資料をもとに、話し言葉の通時的な変化を分析するという点では、いずれも興味深い事実を指摘することができたと思われる。

その一方で、多様な発話者・発話場面の違いを考慮しつつ話し言葉の変化の過程を通時的に分析しようとする観点に立つと、やはり、データの偏りや量の不足といった点が目立つ。つまり、分からないことが多い。この点を補完するためには、2節で論じたような問題意識を持った上で、より多くの音声資料を掘り起こし、「通時音声コーパス」の整備を進めていく必要があるだろう。

2006年、UCL (University College London) から“DCPSE” (Diachronic Corpus of Present-day Spoken English) が公開された。これは、1960年代後半から1990年代前半までのイギリス英語の話し言葉を収録し、形態論情報・統語構造情報などがアノテーションされた通時音声コーパスである<sup>7</sup>。Aarts et al. (2015) は、助動詞 *must*, *may*, *shall* の使用が時代とともに大幅に減少したこと、*would*, *could*, *should* も減少したこと、一方で *will* が増加したことなどを、数量的に明らかにしている。2節で論じたような「通時音声コーパス」の十全な仕様という点では疑問も残るが、実際に「通時音声コーパス」を作成し、話し言葉の動態を数量的に明らかにした、優れた実践例と言える。

現存する古い日本語音声の録音資料（蠟管レコードやSPレコード（落語、演説）など）については、かねてから清水康行や金澤裕之による詳細な調査・分析がある（清水, 1988, 1994, 2011; 金澤, 1991, 2000, 2015）。今後は、現存する録音資料をより幅広く、時代縦断的に収集し、話し言葉の動態を捉えるための通時的研究に利用できるよう、「通時音声コーパス」として整備を進めていくことが重要になるとと思われる。

謝辞：本研究は JSPS 科研費 24520523 の助成を受けたものです。

#### 参考文献

- Aarts, B., Bowie, J., & Wallis, S. (2015). Profiling the English verb phrase over time: modal patterns. In Taavitsainen, I., Kytö, M., Claridge, C., & Smith, J. (Eds.), *Developments in English: expanding electronic evidence*, pp. 48–76. Cambridge University Press.
- 相澤正夫・金澤裕之（編）（2015 予）. 『(仮) 戦前期 SP 盤レコードが拓く日本語研究』. 笠間書院.
- 服部匡 (2011). 「話者の出生年代と発話時期に基づく言語変化の研究—国会会議録を利用して—」. 『計量国語学』, **28** (2), 47–62.
- 金澤裕之 (1991). 「明治期大阪語資料としての落語速記本と SP レコード」. 『国語学』, **167**, 15–28.
- 金澤裕之 (2000). 「録音資料の歴史とその可能性」. 『日本語学』, **19** (11), 197–208.
- 金澤裕之 (2015). 「録音資料による近代語研究の今とこれから」. 『日本語の研究』, **11** (2), 133–140.
- 小磯花絵, 土屋智行, 渡部涼子, 横森大輔, 相澤正夫, 伝康晴 (2015). 「均衡会話コーパス設計のための一日の会話行動に関する調査—中間報告—」. 『第7回コーパス日本語学ワークショップ予稿集』, 27–34.
- 国立国語研究所 (1955). 『談話語の実態』. 国立国語研究所報告 8. 国立国語研究所.
- 国立国語研究所 (1960). 『話しことばの文型 (1) —対話資料による研究—』. 国立国語研究所報告 18. 秀英出版.
- 国立国語研究所 (1963). 『話しことばの文型 (2) —独話資料による研究—』. 国立国語研究所報告 23. 秀英出版.
- 前川喜久雄 (2013). 「コーパスの存在意義」. 前川喜久雄（編）, 『講座 日本語コーパス 1 コーパス入門』, pp. 1–31. 朝倉書店.
- 丸山岳彦 (2012). 「大規模コーパスの利用とメタデータの役割」. 『第1回コーパス日本語学ワークショップ予稿集』, pp. 203–210. 国立国語研究所.
- 清水康行 (1988). 「東京語の録音資料」. 『国語と国文学』, **65** (11), 129–143.
- 清水康行 (1994). 「録音資料に聴く 20 世紀初めの東京語」. 『国学院大学日本文化研究所紀要』, **73**, 191–230.
- 清水康行 (2011). 「欧米の録音アーカイブズ: 初期日本語録音資料所蔵機関を中心に」. 『国文目白』, **50**, 29–19.

<sup>7</sup> 1960年代から1970年代までの音声は London-Lund Corpus から、1990年代の音声は ICE-GB から、それぞれ約40万語ずつが採録されている。http://www.ucl.ac.uk/english-usage/projects/dcpse/