

# 『昭和話し言葉コーパス』の設計・構築と分析

丸山 岳彦

専修大学 / 国立国語研究所

## 1 はじめに

コーパスの類型を考える際、メディアの違いを軸にすれば書き言葉コーパスと話し言葉コーパスの対立があり、時代の違いを軸にすれば現代語コーパスと歴史コーパスの対立がある。このうち歴史コーパスは、実質的には、書き言葉コーパスである。日本語の文献資料は、奈良時代以降、1300年ほどの蓄積があり、これらをコーパス化することで日本語の歴史コーパスが得られる。国立国語研究所で構築が進んでいる『日本語歴史コーパス(CHJ)』は、その実践例である。

これに対して、「歴史コーパスとしての話し言葉コーパス」は実現可能だろうか。清水(2014)によれば、これまでに確認されている最も古い日本語の録音資料は、1900年のパリ万博で録音されたものである。つまり、日本語の録音資料の歴史は、せいぜい120年ほどしか遡れないことになる。さらに、当時の録音資料は断片的にしか残っておらず、話し言葉コーパスとして整備できるほどの質と量を備えているわけではない。

一定の質と量を備えた日本語の古い録音資料として注目されるのが、1950年代の初頭から国立国語研究所で作成された録音資料群である。当時、さまざまなタイプの日常会話や独話を録音し、それらを転記した上で、話し言葉の研究が行われていた。これらは、現代の話し言葉コーパスの源流として位置づけられる。

我々は、当時の音源を再利用して、現在『昭和話し言葉コーパス』の構築を進めている。これは、1950年代から1970年代にかけて、国立国語研究所が作成していた録音資料を利用し、独話25間、会話25時間のコーパスとして再編纂しようとするものである。これにより、今から60年ほど前の話し言葉の実態を明らかにすることができるのと同時に、現代の話し言葉コーパスと比較・対照することで、話し言葉がどのように変化してきたかを実証的に明らかにすることができる。

以下では、現在我々が構築を進めている『昭和話し言葉コーパス』の設計と構築について紹介する。さらに、分析の事例を紹介し、最後に今後の見通しについて述べる。

## 2 『昭和話し言葉コーパス』の設計と構築

はじめに、『昭和話し言葉コーパス』の設計と構築手順について述べる。

『昭和話し言葉コーパス』は、1952年以降、当時の国立国語研究所で作成された録音資料を新たに再編し、コーパス化しようとするものである(丸山2016)。国立国語研究所に初めて録音機が導入されたのは1951年頃で、可搬型のオープンリール型録音機(通称デンスケ、おそらくM1型)が使用された。1952年に話し言葉を専門的に分析する「第1研究室」が設置されて以降は、さまざまな場面における日常会話や、講演会での講演、挨拶、祝辞などの独話音声録音された。この録音資料を使った分析の結果は、『談話語の実態』(1955年)、『話しことばの文型(1)(2)』(1961、1963年)などの報告書にまとめられている。

特に日常会話の録音作業では、「地区・場所・性・年齢・教養・相手」という条件を設定し、各条件をなるべく広くカバーするような作業方針が取られた。図1に、『昭和27年度 国立国語研究所年報4』に載った表を転載する。

Reel No.	録音内容	録音可否	地区		場 所		性		年 齢		教 養		相 手				
			下 山	周 辺	家 庭	学 校	公 共	男 女	男 女	若 年	中 年	老 年	専 門	専 門	1 人	2 人	5 人 以上
3	I 夫 妻	可	x	x			x	x	x		x		x				x
7	T 家 雑 談	可	x	x			x	x	x	x	x	x	x	x			x
67	N 家 雑 談	可	x	x			x	x	x	x	x	x	x	x			x
86	ト タン 屋	可		x			x		x	x		x	x	x			x
76	じ い さ ん	可	x	x			x	x	x	x		x	x	x			x
93	魚 屋 雑 談	可	x	x			x	x	x		x	x	x	x			x
97	U 氏 談	可	x	x			x	x	x		x		x				x
61	学 生 談 1	可	x		x		x		x		x		x				x
68	井 戸 端	可	x	x			x		x	x	x	x	x	x			x
98	友 の 会 話	可	x	x			x		x	x	x	x	x	x			x
2	女 子 在	可	x	x			x		x		x		x				x

図 1: 当時の音声資料 (一部)

言語研究を目的として、日常の多様な場面(地域、場面)、多様な話し手(性、年齢、学歴)から偏りなく日常会話をサンプリングして録音・収集するという方法論は、世界的に見ても極めて早い時期に行われた話し言葉研究と言える。

国立国語研究所では、1990年代以降、オープンリー

テープに記録されていた過去の音声資料をデジタル化する作業が進められてきた。このデータを利用して、2016年度より、国立国語研究所音声言語研究領域において、当時の音声データを『昭和話し言葉コーパス』として整備する作業を進めている。1952年から1974年に録音された音声資料、計50時間分(独話25時間、会話25時間)の音声データに対して、転記テキスト、形態論情報、メタデータなどを整備し、2020年度に一般公開することを予定している。

『昭和話し言葉コーパス』の構築作業は、(1)音声データの収集、(2)転記テキストの作成、(3)転記テキストと時間情報のアライメント、(4)形態論情報の付与、という順序で進めている。これと並行して、メタデータ的设计・付与を行っている。

2020年1月には、これまでに作業が完了した17時間分の独話データについて、「中納言」でのモニター公開を開始した。1955年から1974年にかけて録音された50講演(約17時間、異なりで33人の話者)の独話データ、180,688語が検索できるようになっている。

独話データの例として、1959年3月7日に行われた「国立国語研究所創立10周年記念」の「祝賀式」および「記念講演会」で録音された録音資料の一覧を表1に挙げる。また、中納言での検索結果の例を図2に示す。

表1: 収録された録音資料の例(独話)

祝賀式	
挨拶	西尾実「所長挨拶」31.4分
祝辞	橋本龍伍(文部大臣)3分
祝辞	兼重寛九郎(日本学術会議会長)2.5分
祝辞	関口隆克(国立教育研究所長)3.6分
祝辞	時枝誠記(国語学会代表理事)13.9分
祝辞	山本有三(日本芸術院会員)13分
祝辞	土岐善麿(国語研評議員会長)4.8分
祝辞	安倍能成(文部大臣)7.9分
祝辞	片山哲(衆議院議員)2.3分
記念講演会「現代語の発展のために」	
	西尾実「あいさつ」8.4分
	山田巖「明治初期の書きことば」61.2分
	林大「現代語の標準」48.3分
	大石初太郎「話しことばの文法」39.9分
	岩淵悦太郎「これからの日本語」25.5分

資料ID	話者	録音日時	録音場所	録音形式	録音時間	転記言語	形態論情報	メタデータ	収録時刻	収録時間
05-SC-001	西尾実	1959/03/07	国立国語研究所	独話	31.4分	中納言	あり	祝賀式	00:00	00:31.4
05-SC-002	橋本龍伍	1959/03/07	国立国語研究所	独話	3分	中納言	あり	祝賀式	00:31.4	00:34.4

図2: 「中納言」で『昭和話し言葉コーパス』を検索した結果

### 3 分析の事例

以下では、『昭和話し言葉コーパス』の分析を通して、当時の話し言葉の音調的・文法的な特徴をケーススタディ的に見ていくことにする。

#### 3.1 急激な上昇イントネーション

まず、イントネーションの型について見てみよう。1950年代の録音資料のうち、特に会話データを観察していると、図3のようなイントネーションが現れることがある。

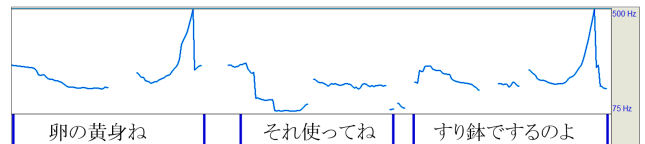


図3: 句末・発話末の急激な上昇イントネーション

これは、「三人の女性」という音声資料(1957年録音)に出現した、20代半ばの女性の発話である。図のピッチ曲線を見ると、「黄身ね」の「ね」、「するのよ」の「よ」、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。この上昇イントネーションは、無論、聞き手に対する質問や疑問を表すものではない。

このような句末・発話末における急激な上昇イントネーションは、小津安二郎作品をはじめとする1950年代の邦画において、女性の台詞の中でよく観察される。『昭和話し言葉コーパス』でも散見されることからすると、当時の女性の間で多用されていたイントネーションであることがうかがえる。自然な日常会話の音声記録したデータがあって初めて、当時のイントネーションの実態を捉えることが可能になっていると言えるだろう。

なお、現代の話し言葉でこのような型のイントネーションが存在するかどうかは、現在構築中の『日本語日常会話コーパス(CEJC)』などを詳しく調べてみないと分からない。が、少なくとも若い世代の女性の中には見受けられないように思われる。過去60年の間に、このようなイントネーションが徐々に廃れていったことが推測される。

#### 3.2 「ます」と「まする」のゆれ

次に、助動詞「ます」と「まする」のゆれについて見る。以下、(1)は1959年「国立国語研究所10周年祝賀式」における山本有三氏(当時72歳)の祝辞、(2)は1955年における波多野完治氏(当時50歳)の講演、(3)は1955年における林大氏(当時42歳)の講演からの例である。

- (1) a. 窮屈な、あー、時代でありますから、  
b. 問題というのは難しいんでありますから、
- (2) a. 書き方というような本を見ますと、  
b. 放送を、おー、聞いておりますと、
- (3) a. 掲載すると書いてありますけれども、  
b. 非常に、あの、時代的な差もありますけれども、

これらは、同一話者による同じ講演の中で観察された例の組である。いずれも同じ接続助詞の直前で、「ます」と「まする」の両方が用いられており、二つの文法形式の間でゆれが生じていることになる。

現代の話し言葉を収録した『日本語話し言葉コーパス (CSJ)』『日本語日常会話コーパス (CEJC) モニター公開版』で「まする」の例を検索すると、CSJに3例「まする」の出現が認められるが、いずれも古典の例文を読み上げているものであり、自然発話としての「まする」の例は一切ない。一方、国会会議録を検索すると、1990年代までは「まする」の例が一定数見つかるほか、現代でもなお少数の例が見つかる。

- (4) a. 財源を十分積み立てることが望ましい姿であります、 (1994年12月)  
b. その対象となった方々の人数をお答え申し上げますと、 (2019年11月)

日常的な話し言葉の中では「まする」が消失した一方、政治的な発言の場面においては、現在でもなお、「まする」が残存しているものと考えられる。

### 3.3 「ますという」と

独話データを観察していると、「ますという」という耳慣れない接続が見つかることがある。

- (5) a. どういうことが考えられてきたかと申しますという、おー、言語を、そういう流動の姿においてでなくて、 (時枝誠記 1959年)  
b. そうして、建物のほうと言いますという、10年たっても、今もって、木造の貸家住まい。 (山本有三 1959年)  
c. 総合雑誌の調査で得られた結果と照らし合わせてみますという、なかなか、あ、よく選ばれていると思われ。 (林大 1959年)

(5)は、1959年の国立国語研究所創立10周年祝賀式および記念講演会からの例である。この場合の「という」は引用節ではなく、文脈としては「申します」と「言います」と「照らし合わせてみます」とのよう

な条件節トと等価である。このような「ますという」は、筆者の語感では非常に違和感のある言い方であるが、『昭和話し言葉コーパス』の独話データの中では、12例が見つかった。

一方、『日本語話し言葉コーパス (CSJ)』でも、ごく少数ながら類例が見つかった。「ますという」は、実は現代でも確認することができることになる。

- (6) それで見ていきますと言うと (Fあ) (Fお) かなりの (D (? ん)) ことが (A07M0652)

### 3.4 「ますのです」

さらに、会話データの中には、「ますのです」という、これもまた不自然な接続が観察される (7例)。

- (7) a. はい。そこに おります ですよ。 (1952年)  
b. 並んで できてます なのです。 (1952年)  
c. 調査も ございます なんですけど。 (1958年)

ところがこれらの例も、『日本語話し言葉コーパス (CSJ)』で18例観察することができる。

- (8) a. 色々 あります なんですけども (S11M1380)  
b. これはもう既にものを書いて 発表してあります なんです (A06M0046)  
c. 朝日が真正面に入る部屋に一人で おります なんです (S08F0486)

(8)の発話者は、いずれも高齢層 (収録時に60代後半以上) に該当する。発話者の年齢が、「ますのです」の出現に影響している可能性はある。

なお、国会会議録の中にも「ますのです」の例が複数見つかる。

- (9) a. 私はこのように 思います なんですけれども、 (1981年3月)  
b. 各省庁間とのまた打ち合わせ等も ございます のんです、 (1987年9月)  
c. 長官の領域に 移っております なんですから、 (1999年8月)

「ますのです」は「不自然な接続」ではなく、少し前の時代の話し言葉では (あるいは現在もなお) 許容されていた (されている) 形なのかもしれない。一見不自然に見える接続が、どのような場面でどの程度使用されているのかについては、さらに広範囲のデータから事例を収集して検討する必要がある。

### 3.5 「ます」「た」と「です」の接続

会話データの中には、「ます+です」や「た+です」という接続が多く観察される。「おいしかったです」のような、形容詞のタ形に「です」が後接する場合は現在では広く許容されているが、問題は動詞・助動詞のタ形に「です」が後接する場合である。「ますです」が30例、形容詞以外の「たです」が66例、それぞれ見つかった。

- (10) a. ここへ来ますです。 (1952年)  
b. あれに654と書いてあったです。 (1957年)  
c. 商売は何をなさったですか。 (1952年)  
d. 入りませんでしたですけどね。 (1952年)  
e. 怖い面もありましたですよ (1952年)

一見、非文のようにも見えるこのような例は、実は、『日本語話し言葉コーパス (CSJ)』でも散見される。

- (11) a. 右端に示してありましたですが (A02M0094)  
b. 絵を 書いたです (S01M0764)  
c. 静かで いいところでしたですね (S03M0174)  
d. 静かに 横たわっていたですね (S02M0076)

特に(11d)のような例の存在を考えると、特に若年層における日常生活の会話の中ではかなり広範に使われている可能性がある(ただし、『日本語日常会話コーパス (CEJC) モニター公開版』の中で確認されたのは、「持ってきたですか (T013.006)」という1例だけであった)。

前川(2007)は、このような「タ形+です」の用例が存在することを指摘した上で、「これらの用例が用いられたであろう文脈を想像してみる。すると私などは(中略)非文と断定しにくく感じられてくる。合理化の契機が与えられれば、むしろ適格文に思えてくる」と述べている。上記の「ますという」と「ますのです」「ますです」などの例に対しても、前川(2007)の主張は通用するように思われる。文法的な規範意識から逸脱するように見えるこれらの例が、話し言葉の中にたまたま出現した誤りなのか、実は体系的に存在している文法的な形式なのか、この点は内省で即断することなく、過去から現在に至る広範なデータを収集して、注意深く吟味する必要があるだろう。

## 4 まとめと展望

本稿では、「歴史コーパスとしての話し言葉コーパス」という視点から、現在構築中の『昭和話し言葉コーパス』を取り上げ、その設計・構築について紹介し、分析の事例を示した。1950年代に録音された日常会話・独話の音声データを再編して構築した『昭和話し言葉コーパス』を、現代の話し言葉コーパス(『日本語話し言葉コーパス (CSJ)』『日本語日常会話コーパス (CEJC)』など)と比較・対照することにより、過去60年間で生じた話し言葉の変化を、アクセント・イントネーション・語彙・文法など、さまざまな観点から明らかにすることができるだろう。本稿では上昇イントネーションと「一見不自然に見える文法形式」に焦点を当てたが、話速やポーズ長、BPMの実態など、音声データならではの特徴について比較・分析することを、今後の課題の一つとしておきたい。

『昭和話し言葉コーパス』のうち、独話データについてはモニター公開中であるが、今後は会話データの整備を含め、2020年度末の本公開に向けて、構築作業を進めていく予定である。

さらに将来的な課題として、二つの点を挙げておきたい。1点目は、録音資料の追加・拡充である。現存する過去の録音資料をさらに発見(発掘)し、音声コーパスとして整備することにより、より幅広いレジスターを備えた「通時音声コーパス」を実現することができるだろう(丸山2015)。2点目は、異なる性格を持つ録音資料群をカタログ化するためのメタデータの設計である。各録音資料を体系的に整理し、それらがどのような性格を持つ資料なのかを正確に知るためには、適切なメタデータを設計・付与しておく必要がある。

### 参考文献

- [1] 国立国語研究所(1955)『談話語の実態』秀英出版。
- [2] 国立国語研究所(1960,63)『話しことばの文型1,2』秀英出版。
- [3] 清水康行(2014)『百年前の日本語を聴く』日本女子大学。
- [4] 前川喜久雄(2007)「コーパス日本語学の可能性—大規模均衡コーパスがもたらすもの—」『日本語科学』22, 13-28。
- [5] 丸山岳彦(2015)「「通時音声コーパス」は可能か」『第8回 コーパス日本語学ワークショップ 予稿集』, 29-36。
- [6] 丸山岳彦(2016)「『昭和話し言葉コーパス』の計画と展望—1950年代の話し言葉研究小史—」『専修大学人文科学研究月報』第282号, 39-55。