

# 熊本方言談話データベースの構築

田川 恭 識

## 0. 本稿の目的

筆者らの研究グループでは、現在、熊本在住日本語学習者向けの熊本方言の教科書作成に取り組んでいる(科学研究費補助金(基盤研究(B)(一般))平成18年度~平成20年度「地方中核都市在住外国人のための方言教材の開発-その理念の構築と実際」(代表:馬場良二)、以下、「教科書プロジェクト」と略称)。教科書プロジェクトは、複数の研究ブロックに分かれており、そのうちの主なものは、実際に教科書を執筆・編集する研究ブロック、そして教科書作成の基礎的資料として、当該地域における方言使用の実態を調査するための調査・研究を行うブロックである。本稿では、筆者が関わっている後者の研究ブロック-談話音声の収集・分析を行うための方法-について報告する。

## 1. 熊本方言について

熊本県は九州本島のほぼ中央に位置し(図1)、人口は平成19年の推計で1,828,289人<sup>※1</sup>であり、人口規模は九州地域で見ると福岡県の5,061,661人に次いで2番目に大きい。(ちなみに3位は鹿児島県の1,731,145人)。熊本県内の人口規模では、熊本市の670,361人が最も多く、次いで県の南部に位置する八代市(138,451人)が続く。熊本市は県の东北部に位置し(図2)、熊本県の経済的な中心地であるとともに、福岡市や鹿児島市と競合しつつ、九州中核都市の一翼を担っている。

熊本県は九州本島の中央に位置しているため、九州北部、南部双方の影響を受けながら地域文化を形成してきた。言語的にも同様であり、一口に「熊本方言」と言っても実際には地域差が大きい。秋山(1983)によれば、熊本県内の方言は大きく北部方言と南部方言に分けられ、さらに北部方言は東部地域とその他の地域の方言に分けられる(図3)。



図1 九州略図



図2 熊本県図(着色部分は熊本市)

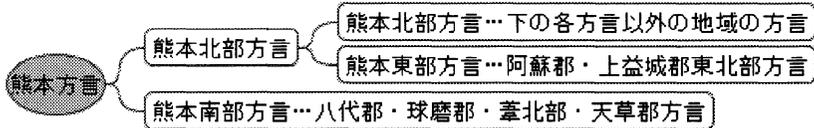


図3「熊本方言」の区画(秋山1983を改変)

音調の面でも、熊本県域の方言においては九州北部地域と南部地域の方言の影響が見られる。大きく九州方言のアクセントを見渡すと、九州方言のアクセント体系は、九州本島の東北部(大分県の全部と西南部以外の福岡県)に東京式アクセントが分布し、九州本島の西南部(佐賀県西南端と長崎県南部、熊本県西部をかすめて鹿児島県の全体)に型の区別が2つしかない所謂二型アクセントが、両地域の間接地帯である西南端以外の佐賀県、長崎県北部から大部分の熊本県、宮崎県にかけて音調の型によって語を区別しない一型アクセントが分布している(早田2001)。熊本方言のアクセントについては、平山(1951)、奥村他(1989)、秋山(1983)などにおいて、その特徴が記述されている。それらによれば、熊本県下における主流は一型アクセントであり、主に県の東部・東北部に分布、また県南西部において二型アクセントが分布している。また、現在の主流である一型アクセントは、二型アクセントから変化したものであり、二型アクセント地域における一型化あるいは変化の過程と見られる曖昧アクセントも見られるとしている。

以上、「熊本方言」と言っても地域ごとに性質の違いが認められる。しかし、音調の観点から見ると、大きく一型アクセントと二型アクセントに分けられるであろう。一型アクセントは、研究者によっては「無アクセント」や「無型アクセント」などと呼ばれることもあり、それぞれの定義は異なる。各用語についての説明は省略するが、少なくともそれぞれに共通しているのは、①当該方言では音調による語の弁別が行われないこと、②語を単位とする決まった音調の型は存在しない\*2、という点であろう。本稿で「熊本方言」と呼び、分析対象とするのは、上記のような特徴を持つ熊本北部方言である。その中でも、熊本市で使用されている方言に焦点を当てる。また、熊本市においても、年代差や性別などによって様々な言語変種の存在が考えられるが、本研究では特に10代後半から20代までの話者が日常的に使用する方言を主な分析対象とする。その理由は、教科書プロジェクトで想定している日本語学習者が、大学や高校の留学生といった若年層であることが挙げられる。一般的な方言調査では様々な年代のインフォーマントを揃えて調査を行うことが多いが(小西2007)、本研究においては以上の理由から若年層を対象と

する。インフォーマントが若年層中心であるため、いわゆる伝統的な熊本方言の記述には結びつきにくい、当該方言地域における共通語化の様相やネオ方言といった社会言語学上の問題について考察する上でも有益な資料となると予想される。

## 2. 研究の概要

図4に熊本方言談話の収集・分析に当たっての大まかな見取り図を示す。色が塗られた右の三つの部分が、本稿での報告部分に相当する。以下、各節ごとにそれぞれの項目の説明を行なう。

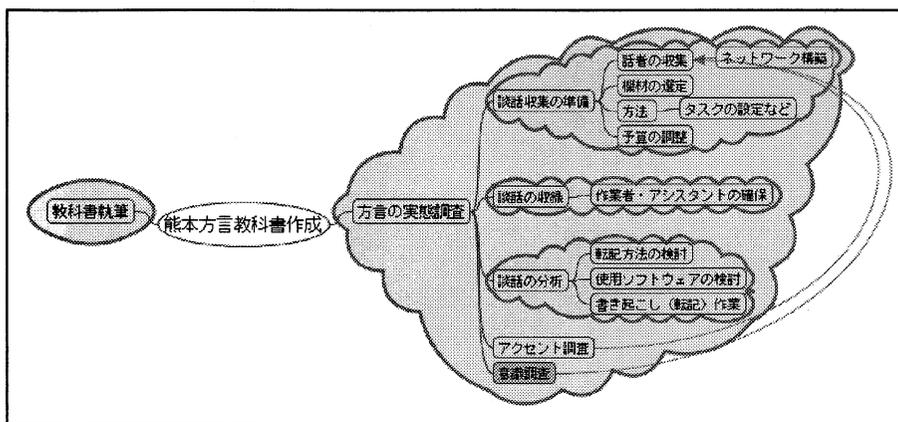


図4 研究の見取り図

### 2.1. 被調査者

現時点での調査協力者（インフォーマント）は、熊本方言話者男女20名（男性7名、女性13名）である。インフォーマントの主な社会的属性は学生で、熊本市内の大学に通学している大学生及び大学院生が中心である。調査時、拘束時間や調査内容に応じ、相当と見なされる謝礼が支払われた。

## 2.2. 談話の概要

### 2.2.1. 話者の関係

通常、我々は方言なら方言、共通語なら共通語というように、ある特定の言語のみを使用して生活しているわけではない。ある場面では方言で話し、また違う場面では共通語で話すというように、さまざまな言語（バラエティ）を使い分けている。逆に言えば、言語使用において、どのような場面なのかということが言語使用に大きな影響を与えるということになる（杉戸 1992、真田 2007）。場面と言うと、大学での講演や、会社での就職面接といった話の場や状況が想起されやすいが、話す相手との関係もまた、そこには含まれる。話者間の関係を親疎で見た場合、話者間の関係が「親」の場合、より砕けた普段の言葉が使用されると考えられる。また、反対に話者間の関係が「疎」の場合、より規範的な言葉が使用されるであろう（陣内 1996）。以上を考慮し、本研究では、親密な関係（友人同士等）にある話者間で行われた談話（親条件）と、初対面同士の話者による談話（疎条件）を収集する。これにより、話者ごとに親疎の談話が少なくとも一本ずつ収録され、2つの談話を比較することにより、関係（親疎）の違いによる言葉の使い分けについて考察が可能になる。現時点で収集した談話データは親疎条件合わせて18談話で、今後とも必要に応じて増える見込みである。

### 2.3. 談話のテーマ

談話の収録にあたっては、いくつかの方法が考えられる。そのうちの一つは、話者同士によって自由に談話を展開していくもので、「自由談話」と呼ばれることもある。また、調査者から話題を与え、それをテーマに会話することを依頼するという方法もある（久木田 2002）。これらの方法は、「自然な談話」が得られるという点で有益な方法であるが、(1) 口数が少ない談話参加者の場合、極端に発話数が少なくなる、(2) 談話が自由に展開されるが、反面、内容の統制がとれない、などの問題も挙げられる。これまでの談話研究では、以上のような「自由談話」もしくは「話題提供」という手法に則ったものが多かったようであるが、近年では何らかのタスクを会話参加者に与え、そのタスクの遂行過程の談話を収集するケースが増えてきた（代表的なものとして『日本語地図課題対話コーパス』や『日本語話し言葉コーパス』中の「課題指向対話」などが挙げられる）。本研究では、以上に挙げた「自由談話」及び「話題提供」の手法の問題点を考慮し、タスク指向型の談話を収集することとした。本研究の談話収集におけるタスクを参考として本稿末に挙げる。なお、1談話あたりの時間は20～30分程度であった。

## 2.4. 収録の様様

音声の収録のほとんどは、熊本県立大学のスタジオで行われた。収録の様様について模式図を示す。スタジオ内にはモニター室と録音室があり、2つの部屋は壁で仕切られているが、壁の一部はガラス張りになっており、モニター室から録音室の様子を見ることが出来る。

話者は録音室内にある椅子に腰をかけ、ヘッドウォーン型マイク（エレクトレットコンデンサー型鋭指向性マイク）を装着し、会話する。その間、調査者Aは話者から見えないようにスクリーンを挟んで座り、ヘッドフォンでモニタリングしながらミキサーで録音音声をコントロールする。ミキサーには半導体型録音機が繋がれており、ミキシングされた音声を録音する。また音声の録音に加え、話者から目立たない位置にビデオカメラを設置し、会話の様子を録画する。調査者Aはスクリーンを挟んでいるため、話者の様子は見えないが、ビデオカメラの液晶パネルによって観察が可能となる。調査者Bは収録の様子をモニター室から見守り、談話収録中の特記すべき事柄について記録を行う。モニター室の明かりは消されているため、収録スタジオからモニター室の様子は窺えない。

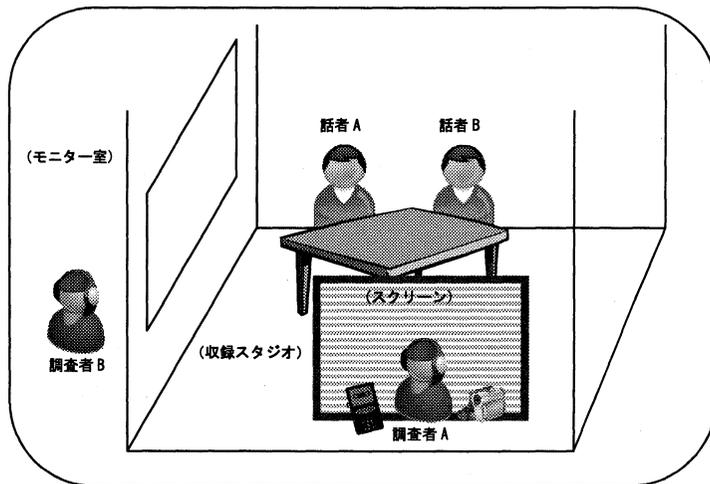


図5 収録の様様図

## 3. 音声の書き起こし

### 3.1. 書き起こしの方法・基準について

収録した音声を分析するに当たり、最初のステップとして音声の書き起こしを行なう。書き起こしの方法や基準については、研究者ごとに研究目的に沿ったものが考案されて

いるが、汎用性を考慮した書き起こし方法・基準としては、幼児の言語獲得研究のために開発された CHILDES (Child Language Data Exchange System) や日本人の談話行動研究のために開発された BTSJ (Basic Transcription System for Japanese) (宇佐美 1997) などが知られている。

本研究では、事前に行った予備調査の書き起こし資料を元に、書き起こし作業間で問題点を討議し、書き起こし基準を策定した。本調査で収集した談話の書き起こしにおいても、新たに問題点があればその都度報告し、作業者間で討議することで、書き起こし基準のブラッシュアップを行っている。

### 3.2. 書き起こし作業に使用するツール

音声の書き起こしは、音声自動認識システムを用いない場合、人手によって録音された音声を聞きながら、規則に則って転写していく。音声自動認識システムを使用した場合、ほぼ音声の実時間と同程度の時間で転写が可能であると考えられるが、人間の場合は発話スピードと同じ速さで転写することは不可能である。当然、人間の短期記憶の容量には限界があり、聞き漏らしや誤りが生じる。そこである程度の短い範囲を繰り返し聞きながら作業を行なう必要がある。カセットテープで録音された音声を書き起こす場合、トランスクリャーなどの機器を用いることで、同じ箇所を何度も聞き返したり、巻き戻し、早送りを行ったりと効率的に作業を行うことが出来る。また、近年ではコンピュータ上での書き起こし作業に特化したソフトウェアがフリーでダウンロード出来るようになってきた。これらのソフトを用いることで、本研究での資料のようなデジタルデータ化された音声についても書き起こし作業の効率化が期待される。しかし、談話音声は2人の対話によって構成されており、発話が重複する場合が少なくない。そのような重複部分の聞き取りは、時に困難を伴い、結果として書き起こしデータの信頼性が低下することになる。また、談話音声の書き起こしデータを効率的に利用するためには、ある発話が談話全体の中でいつ頃発話されたか、といった音声の時間情報も盛り込まれていることが望ましい。一般的に、書き起こし用のソフトを使用して書き起こしを行った場合、時間情報を付与したいと思えば、手作業によって時間情報を付与する必要がある。

以上の問題点を回避するために、本研究では書き起こし作業において、MultiTrans というソフトウェアを用いることとした。MultiTrans の特徴の1つに、複数チャンネルの音声に対応していることが挙げられる<sup>\*)</sup>。本研究では、インフォーマント一人ずつにマイクを装着し、各話者の音声をLチャンネルとRチャンネルとに振り分け録音する。MultiTrans では、このような複数チャンネルの音声データを分離し、それぞれを独立

して書き起こすことが可能である。そのため、発話の重複部分も話者ごとに分離して書き起こすことができ、重複部分の書き起こしデータの正確性が増す。また、発話の時間情報も自動的に付与してくれるため、手作業による作業の誤りが減少する。MultiTransで書き起こしたデータは、LCF形式かXML方式で保存でき、テキストエディタで閲覧・編集が可能である。図6にMultiTransでの作業画面を示す。画面の上部が文字を入力する部分であり、その下の2層に別れた部分には波形が表示される。上段にはSpeaker1 (Rチャンネル) の音声波形、下段はSpeaker2 (Lチャンネル) の音声波形で、任意の長さの波形を選択し、繰り返し聞くことができる。

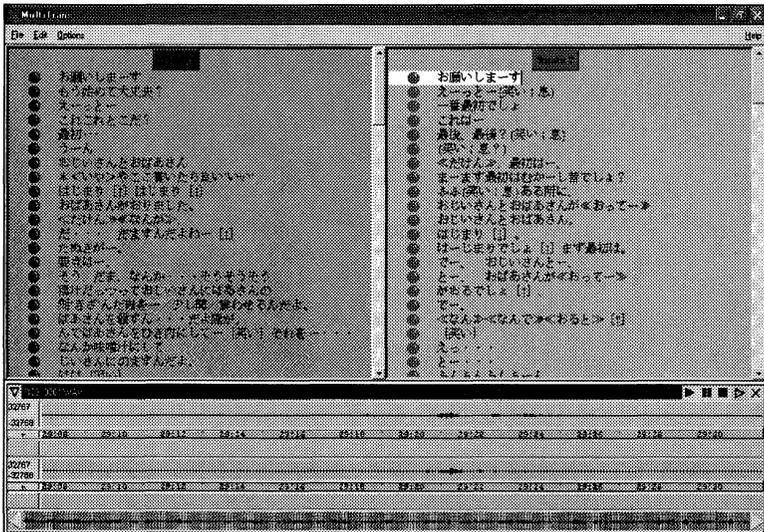


図6 MultiTrans の画面

#### 4. 談話音声の分析

書き起こしに続き、分析のためのステップとして、談話音声の音声の特徴についての記述を行なう。

音声の特徴、とりわけイントネーションの特徴記述の手法として良く知られたものに、ToBI (Tone of Break Indices) がある。ToBI そのものは英語の音声を対象としていたが、後に各言語向けに拡張され、やがて Venditti によって日本語向けに J\_ToBI (Venditti1997 など) が開発された。さらに現在では J\_ToBI を自発発話音声向けに拡張した X\_JToBI が開発されている (前川他 2004)。その他、言語学的関連付けは薄いが、F0 曲線の屈曲点を自動検出し、イントネーションのモデリングを行なう MOMEL



## 5. 今後の展望

以上、熊本方言談話データベースの構築の方法とその概略について述べた。平成 19 年度現在、談話音声の収録は大部分が終わっているが、親疎の談話が揃っていないインフォーマントがいるため、今後とも収録を続けていく予定である。書き起こし作業も進行中であり、今年度中に書き起こし作業を終了し、責任者によるチェックを終える見通しである。また、収録や書き起こし作業と平行し、収録した談話を資料とした研究も進行中である。来年度以降、熊本方言談話データベースを使用した研究成果の報告が期待される。

## 注

- \*1 ウィキペディア調べ (<http://www.wikipedia.org/>)
- \*2 郡 (2006) では熊本市方言において、文節を基本単位として音調の型が見られるとし、当該方言のアクセントを「非定型アクセント」であるとしている。
- \*3 <http://chilides.psy.cmu.edu/>
- \*4 小磯 (2006) などを参照されたい。
- \*5 前川 (1990)、前川 (1992)、前川・吉岡 (1992)、前川 (1997)、前川他 (1999)、馬場 (2004)、郡 (2006)、田川 (2005) など。
- \*6 <http://www.speech.kth.se/wavesurfer/>

**付記:** 本研究は、科学研究費補助金 (基盤研究 (B) (一般)) 平成 18 年度～平成 20 年度「地方中核都市在住外国人のための方言教材の開発 - その理念の構築と実際」(代表: 馬場良二) に基づく。

## 参考文献

- 秋山正次 (1983) 「熊本県の方言」『講座方言学 9-九州地方の方言 -』国書刊行会。
- 宇佐美まゆみ (1997) 「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ) の開発について」『日本人の談話行動のスクリプト・ストラテジーの研究とマルチメディア教材の試作』文部省科学研究費一般 (C) 研究成果報告書。
- 奥村三雄編 (1989) 『九州方言の史的研究』, 東京: 桜楓社。
- 小磯花絵 (2006) 「会話データの構築法 - 収録と書き起こし」『講座社会言語学第 6 巻 方法』, 東京: ひつじ書房。
- 久木田恵 (2002) 「方言の表現: 会話(談話)」『朝倉日本語講座 10 方言』, 東京: 朝倉書店。

- 郡史郎 (2006) 「熊本市および周辺の非定型アクセント方言における語音調と音調句の形成」『音声研究』10:2,43-60.
- 小西いずみ (2007) 「新時代の方言調査法」『シリーズ方言学 4 方言学の技法』, 東京: 岩波書店.
- 真田信治(2007)「発話スタイルと方言」『シリーズ方言学 3 方言の機能』, 東京: 岩波書店.
- 陣内正敬 (1996) 『地域語の生態シリーズ九州篇 地方中核都市方言の行方』, 東京: おうふう.
- 杉戸清樹 (1992) 「言語行動」『社会言語学』, 東京: おうふう.
- 田川恭識 (2005) 「熊本方言における「問いかけ」と「非難の問いかけ」のイントネーションパタンについて」『第 16 回社会言語科学会研究大会発表論文集』, 社会言語科学会.
- 馬場良二 (2004) 「熊本方言における引用の助詞「て」を含む節の韻律について」『日本音響学会聴覚研究会資料』, 34:4,H-2004-42.
- 早田輝洋 (2001) 「単語声調と九州方言」『音声研究』5:3,4-9.
- 平山輝男 (1951) 『九州方言音調の研究』学界之指針社.
- 前川喜久雄 (1990) 「無アクセント方言のイントネーション (試論)」『音声言語 IV』近畿音声言語研究会.
- 前川喜久雄 (1992) 「熊本無アクセント方言のイントネーション」『月刊言語』98,66-74.
- 前川喜久雄(1997)「アクセントとイントネーション-アクセントのない地域-」『日本語音声 [1] 諸方言のアクセントとイントネーション』, 東京: 三省堂.
- 前川喜久雄・吉岡泰夫 (1992) 「熊本無アクセント方言のイントネーション: イントネーションモデルの知覚実験」『日本語音声の研究と日本語教育』文部省科学研究費重点領域研究「日本語音声」国際シンポジウム資料,35-42.
- 前川喜久雄・楨洋一・吉岡泰夫 (1999) 「発話の丁寧さの知覚におよぼす語彙的要因と韻律的要因の関係とその社会差」『電子情報通信学会技術報告』TL99-19,9:16.
- 前川喜久雄・五十嵐陽介・菊池英明・米山聖子 (2004) 「日本語話し言葉コーパス」のイントネーションラベリング Version1.0」日本語話し言葉コーパス附属文書.
- Venditti, Jennifer J. (1997) "Japanese ToBI Labelling Guidelines." , Ohio State University Working Papers in Linguistics 50: 127-162.

おはなし

下のお話の中から1つ選び、2人で物語を完成させてください。出来るだけ、詳しく、正確に思い出してください。

