

# 「全国方言文法辞典」におけるWebによる 調査データの報告システム開発について

林 良雄<sup>†1</sup>

方言研究の中でも方言文法には多くの調査データが必要である。そこで調査データを報告し、自動的にデータベース化を行うシステムを構築している。このシステムでは方言研究者の求める学術的なデータ収集を目的としている。これを使うことにより、多数の地方における統一した調査結果が素早く収集、共有でき、方言文法研究に役立てることが可能となる。まだ本運用には至っていないので本報告ではそのシステムの概要について述べる。

## On Development of Web Based Report System of Investigated Data for “Dictionary of National Dialect Grammar”

YOSHIO HAYASHI<sup>†1</sup>

The study of dialect grammar needs a lot of research data. We are constructing a investigation report system which automatically make databases of collecting data. The aim of his system is collecting academic data that dialect scholars need. By using this system, we can rapidly collect data from many region, can share and can serve the study of dialect grammar. Because it is not operated officially, we give an overview of the system.

### 1. はじめに

現代語の文法研究の成果を日本語諸方言の対照研究に応用する試みは幾つかの研究がおこなわれており、一連の成果を上げている<sup>1)</sup>。これらの研究は現代語（標準語）に限定せず地理的変異（方言差）および歴史的変異（時代差）を考慮に入れた日本語文法の全体像を把握することを目的としている。そのため、大量の調査データを必要とする。これまでの方言研究の中でも膨大な調査データの蓄積があるが、必ずしも文法項目を網羅するものではなく、文法の全体像を把握するためには不十分である。また、誰にでも容易に活用できる形になっているわけではなく、他研究での利用が十分になされているとは言えない。

そこで、関西大学文学部の日高が中心となって「方言文法研究会」を2001年に立ち上げ、日本語諸方言の文法を総合的に記述する『全国方言文法辞典』の編纂を目的に、既存の方言文献資料をデータベース化する作業および、現地調査による方言文法の記述的研究を進めている。例えば「日本方言大辞典<sup>2)</sup>」から助詞、助動詞、接辞類の記述を抜粋しデータベース化する作業や各種の方言談話資料類から用例をピックアップする作業を行った。また、現地調査のための統一的な調査項目の選定を行い、一部の項目については、すでにその調査報告を公開している<sup>3)</sup>。

### 2. 本研究の目的

前述の方言文法研究会では、現在、記述の出発点を助詞類に定め、共通調査項目の作成と要地方言の選定を進めている。この部分が確定すれば複数の調査員が担当の地域の調査を行い、データを蓄積することになる。最終的にはデータをまとめて「全国方言文法辞典《助詞編》」の冊子版の刊行及びウェブ版の公開を行うことを目標としている。この研究では異なる調査員が行った現地調査の結果を利用して、辞典項目の記述を行う。したがって、調査データは、誰でも利用できる形、すなわちデータベース化する必要がある。通常の調査研究において調査データはエクセルなどの個人的なデータファイルとして記録され、それを交換することによりまとめていく。最終的な結果は論文や報告書などの紙ベースで出され、電子化されたものは出されない、あるいは単純なWebのページとして公開されることとなる。

そうすると他の研究者の利用は紙ベースのものを再度コンピュータに入力し直すか、Webのページから必要な部分をコピーすることによって利用することになり、誰にでも容易にということにはならない。また、網羅的な研究を行うことも困難となる。

そこで、この研究ではWebを利用した調査報告システムを構築することとした。各地で調査を行った調査員がWebサーバーにアクセスし、調査結果報告のフォームに調査結果を入力する。入力されたデータは即時にデータベース化されていくシステムである。

これには次のような利点がある。

<sup>†1</sup> 秋田大学教育文化学部  
Akita University Faculty of Education and Human Studies

- (1)入力した時点で電子データとして扱うことができる。
- (2)どの調査地点からでも即時に報告が行え、従って報告の為にどこかで会議を開く必要がない。
- (3)調査報告がどのようになっているかを逐次全員見ることができできるようにすることにより、調査に参加している者全員がデータを共有できる。
- (4)入力時の形式を統一することにより、記号や記述の形式を統一することができる。

このシステムで調査結果を蓄積したデータベースからデータを抽出することにより辞典の編纂およびWebの公開を行うことが最終目標である。

### 3. システムの要件について

本研究は科研費の配分を受けて一昨年度から5カ年で実施することになっている。今年度は三年目としてWebのシステムの利用開始を予定している。既にシステムは一応の完成を見ている。この章ではこのシステムの考え方を述べる。

本研究で行う方言調査の場合、直接話者から聞き取りを行う。調査では話者がどのような環境で育ったかが重要である。そこで話者情報として表1の情報を扱うこととした。通常は一人の話者に複数の質問を行うことになる。このことを考えると話者の基本データの入力と調査データの入力は独立させたほうが良いと判断される。

共通調査項目は方言文法研究会で検討のうえ、条件表現68項目、逆接表現のケレドモ・ノニ類25項目、テモ類46項目の調査例文が設定されている(表2)。

目的において述べているように、このプロジェクトでは調査結果を共有することが必要である。そのために他の調査員の結果を検索する機能を備える必要がある。以上のことを踏まえ、このシステムは図1のような流れとなっている。なお、システムはCentOS 5.7上のapache2.2.3およびphp5.3で開発している。また、データベースとして

表1. 話者データ項目

	データ項目	説明
地点情報	調査地点	都道府県, 市区町村は必須
	地点概要	自由記述
話者情報	話者氏名	必須
	話者性別	必須
	話者生年月日	西暦年は必須
	調査時満年齢	必須
	現住所	都道府県, 市区町村は必須
	電話番号	省略可
	出身地※	現住所と異なる場合
調査概要	外往歴	出身地と異なる地域に居住した経験がある場合
	その他	自由記述
	調査者	ログイン情報から
調査概要	同席者	調査者以外に調査に立ち会った人
	調査場所	調査を行った場所
	調査日時	
	その他	自由記述

※言語形成期(6~12歳)に主に過ごした土地。

表2. 共通調査項目の一部

逆接表現(テモ類)	共通調査項目例
1. 従属節用法	
1-1. 仮説的用法	
(01)動詞述語, 推量 走っても, 間に合わないだろう。	
(02)動詞述語否定形, 推量 走らなくても, 間に合うだろう。	
(03)形容詞述語, 推量 今は寒くても, 午後になれば暖かくなるだろう。	
(04)名詞述語, 推量 雨でも, 試合はあるだろう。	
(05)形容動詞述語, 推量 いやでも, ちょっとは我慢できるだろう。	
逆接表現(ケレドモ・ノニ類)	共通調査項目例
1. 従属節用法	
1-1. 因果関係の不成立・予想外	
(01) 動詞述語・非過去形接続 あんなに食べる {けれども/のに}, やせている。	
(02)動詞述語・過去形接続 せっかく木を植えた {けれども/のに}, 枯れてしまった。	
(03)形容詞述語・非過去形接続 あんなに家が近い {けれども/のに}, いつも遅刻する。	
(04)形容詞述語・過去形接続 あんなに家が近かった {けれども/のに}, いつも遅刻する。	
(05)名詞述語 もう出発の時間 {けれども/なのに}, まだ来ない。	
条件表現	調査項目例
1. 従属節用法	
1-1. 仮説的用法(予測的条件)	
(01)動詞述語, 推量 あした雨が降れば, 船は出ないだろう。	
(02)動詞否定述語, 推量 あした雨が降らなければ, 船は出るだろう。	
(03)形容詞述語, 推量 あした波が高ければ, 船は出ないだろう。	
(04)形容動詞述語, 推量 あした波が静か {であれば/なら}, 船は出るだろう。	
(05)名詞述語, 推量 あした雨 {であれば/なら}, 船は出ないだろう。	

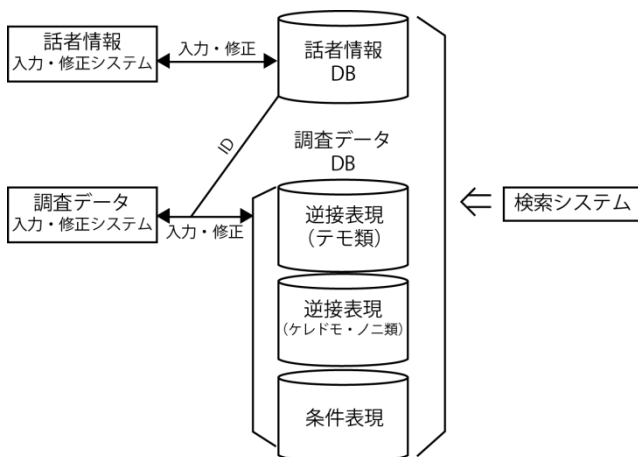


図1 システムの概要

MySQL5.1 を使用している。以下ではこの流れの各部分について述べる。

#### 4. データ入力

このサイトのトップメニューは図2の通り報告を行うシステムと、方言研究会のサイトへのリンクが張られている。方言研究会のサイトでは moodle を使い、意見やドキュメントの交換をおこなっている。

報告を行うシステムでは話者の氏名や住所などの個人情報情報を扱うので、パスワードで保護している。これはデータベース (MySQL) を使ったベーシック認証により実現しており、このシステムに入った時点でユーザーID とパスワードが求められる。データ報告システムのトップメニューは図3のとおりである。

報告者はまず、調査の対象となる話者の情報を入力しておく(図4)。表1に示した必須項目は赤字で示し、目立つようにしている。さらに、入力の際には必須項目のチェックを行い、入力忘れの防止を行っている。

現在準備している共通調査項目は、話者情報を入力したのち、これらの調査したデータの入力を行うことになる。



図2 トップページ

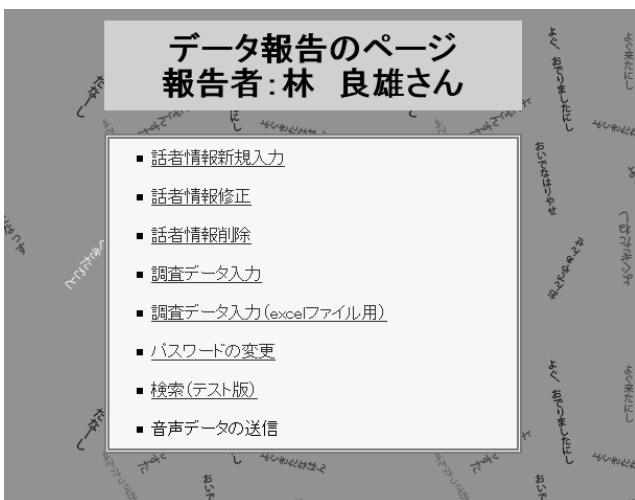


図3 トップメニュー



図4 話者情報入力画面

調査員が担当する話者の情報のみを入力、確認するために図5のようにその調査員が入力した話者のみのリストから選択するようにしている。

一つの調査項目については調査例文に対する方言訳と文法性判断 (○:自然, ×:不自然, ? :やや不自然), それに対する注記を1セットのデータとしている。ただし、一つの例文に対して複数の言い回しが存在する場合があるので5セットを一つの例文に用意する(図6)。

調査項目はかなりの数があり、入力に手間がかかる。また、Web入力であると入力したデータそのものは手元に残らない。そこで、Excelファイルのアップロードができるようにした。ExcelファイルにWebの調査と同じような項目を同じような書式で記述し、そのファイルをアップロードすればデータが入力されるようにした。これについては後述する。

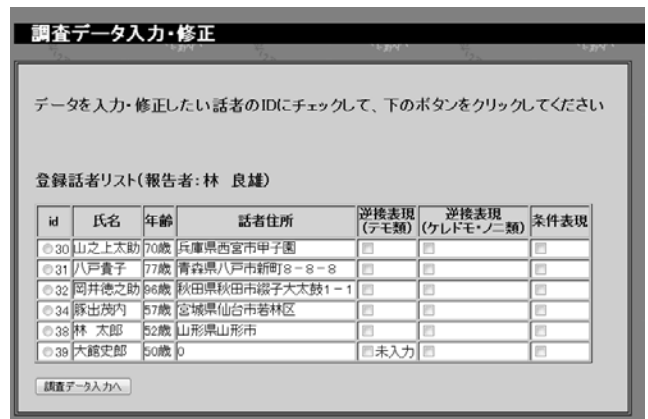


図5 話者の選択画面



図 6 データ入力画面

## 5. 検索機能

データの入力は今後進められることとなるが、次に必要となるのが検索を行うシステムである。ここでの検索機能は、ある言葉が入っているものという絞込みよりも、調査項目や調査地点が同じものを全てリストアップするという考え方である。方言研究者にとっては Web 上で絞り込んでみるという作業より、一覧を見ながらじっくり検討することが好まれる。あるいは研究用の資料としても一覧表示が必要となる。そこで次の三種類のメニューを用意した(図7)。まず、調査員自身が入力したデータの検索である。検索は調査をした調査員自身がデータ確認を行う場合と他の調査員が研究等でデータを調査する場合があると考えている。前者の場合では話者の基本情報まで出せるようにするが、後者では話者情報の中の個人情報部分、氏名や住所などを表示してはならない。

そこで、この調査員自身が自分の入力したデータを確認したり、まとめてダウンロードするための機能を用意した(図8)。どのユーザーが検索しているかは php のスーパーグローバル変数である PHP\_AUTH\_USER を使ってログイン時のユーザー情報をもとに判断し、そのユーザーだけが報告した話者とその調査結果を表示するようにしている。従って、ユーザーはあまり意識することがないようになっている。

次は全調査員のデータを検索できるものである。これには二種類のメニューを用意した。どちらも、調査項目のカテゴリと調査地点で絞り込むが、一つは話者単位でデータを

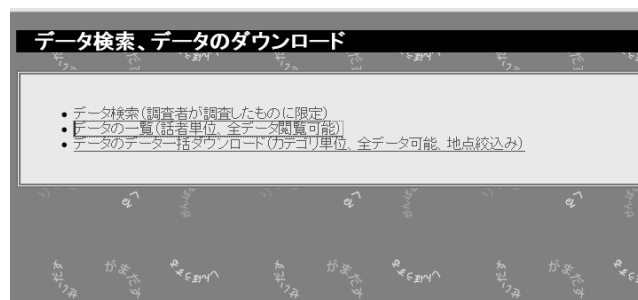


図 7 データ検索メニュー

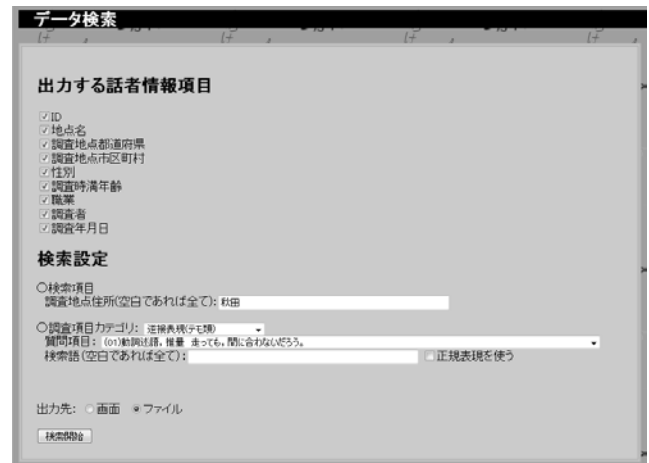


図 8 調査員自身のデータの検索画面

表示するもの、もう一つは一人の話者に限らず条件にあうデータを全て出力するものである。

後者があれば前者のデータも含まれるが、後者の方は出力する情報量が多いため、単なる一覧を出力するのに対し、前者は一人の話者だけに絞っているため(図9)、調査データ入力と同じ形式(図6)で出力することができる。

一方、一つのカテゴリの中の調査項目は多数ある。一つの調査項目についても5例記述できるようになっている。これをディスプレイに直接表示するととても見づらい。そこで、結果をファイルに出力することができるようにもした。

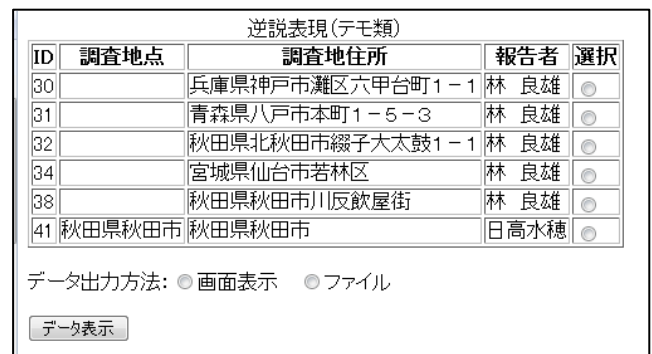


図 9 検索結果の話者リスト

## 6. その他の機能

このシステムには前章までで述べた機能以外に類似するシステムには通常ある機能を備えている。例えばデータの削除やパスワードの変更などである。ただ、この調査は言語学的な素養がないと正確なデータが集められないこと、他者の悪戯などがあると困ることから自由に登録できるような機能は供えていない。調査員はプロジェクトのリーダーである日高に申し出て、それを筆者がシステムに登録することになっている。

4章でも触れたようにデータ入力を Excel ファイルをアップロードすることにより行う機能も備えている(図10)。調査データの入力画面と同じようなシートを入力した Excel ファイルを用意しておき、利用者にはそのファイル

に調査データを入力してもらい、そのファイルをアップロードするとファイル中の話者 id を読み、話者データベースで検索する。話者 id は話者データベースのキーであり、一意に決まる。もし、その話者が他の調査員のものであれば処理が停止される (図 11)。Excel ファイルの読み込みは PHPExcel ライブラリを使うことにより実現されている。

ただし、幾つかの難点が存在すると思われる。まず、データを入力するセルが決まっているので勝手に行や列、セルを削除したり追加すると正しく読み込めない。また、基本的にはアクティブなセルを読むようになっているが、調査データがアクティブになっていない状態あるいは違う調査データのシートがアクティブな状態で保存されていれば、データが読めない。Web 入力よりはトラブルが多いことが予想される。

メニューにあるがまだ実装していない機能が音声ファイルの送信 (アップロード) である。計画当初は調査時の音声ファイルを同時に集めることも考えていたが、限られたグループ内での公開でも話者の了解が必要と思われるし、現在のプロジェクト自身では文字情報で正確な情報を得ることが先決であるので、この部分については今後の検討ということにしている。

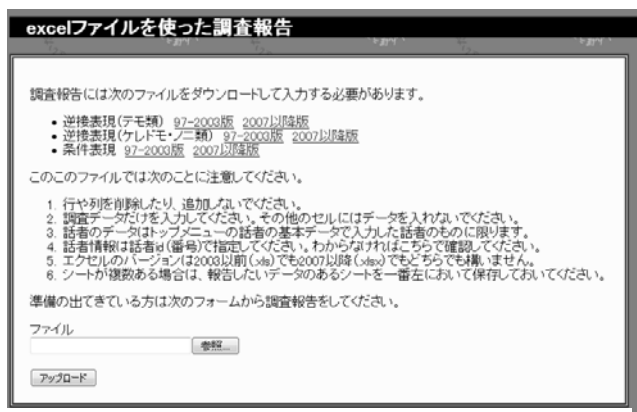


図 10 エクセルによる調査データ報告の画面

林 良雄さんの報告  
 katei.xlsx をアップロードしました。

林 良雄さんの扱っている話者に ID= の人はいません。  
 現在登録されている話者は次の通りです。

id	氏名	調査時年齢	話者住所
30	山之上太助	70歳	兵庫県西宮市甲子園
31	八戸貴子	77歳	青森県八戸市新町8-8-8
32	岡井徳之助	96歳	秋田県秋田市綴子太太鼓1-1
34	豚出茂内	57歳	宮城県仙台市若林区
38	林 太郎	52歳	山形県山形市
39	大館史郎	50歳	

図 11 話者の id が合わなかった際の画面

メニューにはないが、データの表示における機能として、更新日時に関するものがある。調査は新規に入力される以外にも、後で修正が加えられることが考えられる。調査データを見る場合に、以前見たものと内容が変わっていることがありうるということである。現在はデータが更新されても、文字の色等に変化はなく、一目では分かり難い。そこで以前から変更があった更新日の色を変えてほしいという要望がある。ただ、そのデータを閲覧した人が前回いつ閲覧したのかが分からなければその目的に十分にこたえられないであろう。ただ、初めての入力ではなく単にデータに修正が加えられたという情報だけを表すことはできるので、その方向で機能を追加していくつもりである。

## 7. 今後について

今年度から実際の調査に入ることになっている。これまでではテスト的に何人かのユーザーに利用してもらって、現在のところ大きなトラブルはない。ただ、本格的な運用が始まると、問題点が見つかると思われるので、今後はその対応を行っていく。また、新しい調査項目も追加されることになっているので、報告のシステムも更新していくことになる。

今後は収集したデータをどのように活用、あるいは全国方言文法辞典の Web 版とするかの検討に入る。第 1 章でも述べたように本研究会ではすでに全国方言文法データベースを一部作成している (図 12)。公開しているのは接続句関係の「原因・理由」だけであり、まだまだ多数の項目が残っている。

その中に会話例あるいは例文集が公開されており、簡単な地図上にその調査地点が表示され (図 13)、それぞれの地点での会話例や例文集が見られる (図 14, 15, 16)。

この Web のページはデータベースという名称になっているが、html で作られている単純なページであり、本格的なデータベースシステムを利用したものではない。これはこの研究会に情報の専門家がなかったためではあろうが、もう一つ理由がある。それは集めたデータをそのまま生で使うことはなく、必ず方言の専門家が元データを検討を加え、確認したうえで纏めたものを公開するという高い専門性のためでもある。

今回のプロジェクトでも、このシステムでデータを集めるのは基礎資料である。その基礎資料をもとに言語学的な検討を加え、そこで得られた結果を改めてデータベース化することになると思われる。ただし、多くの調査データが集まる可能性があるため、そのデータの更なる活用を考えることも今後の課題であろう。

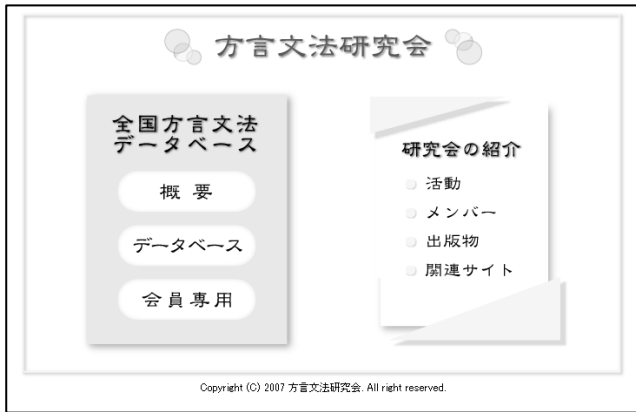


図 12 Web版全国方言文法データベース

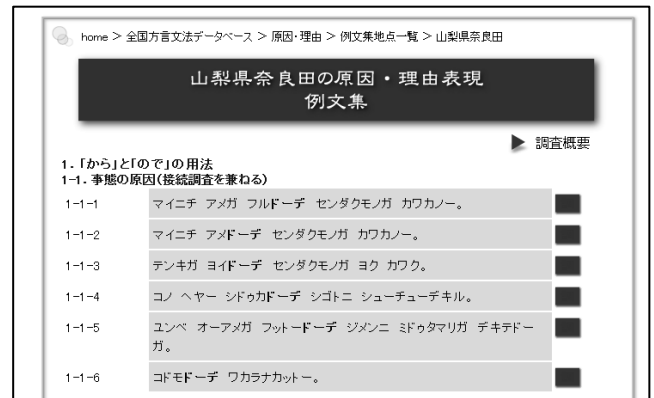


図 15 例文集

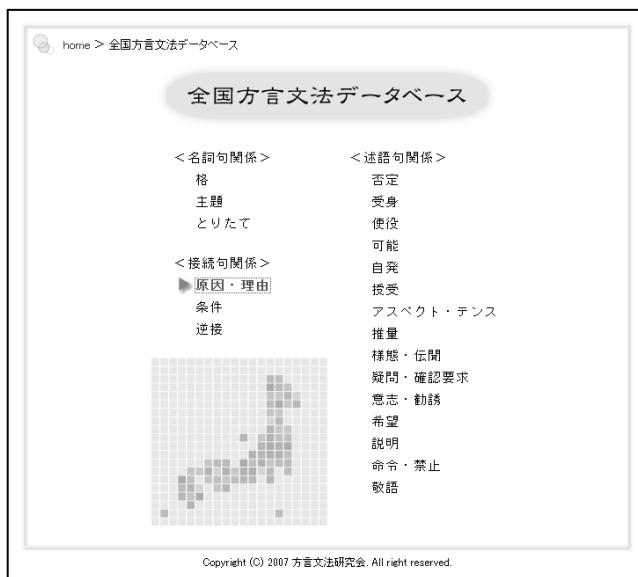


図 13 調査予定項目

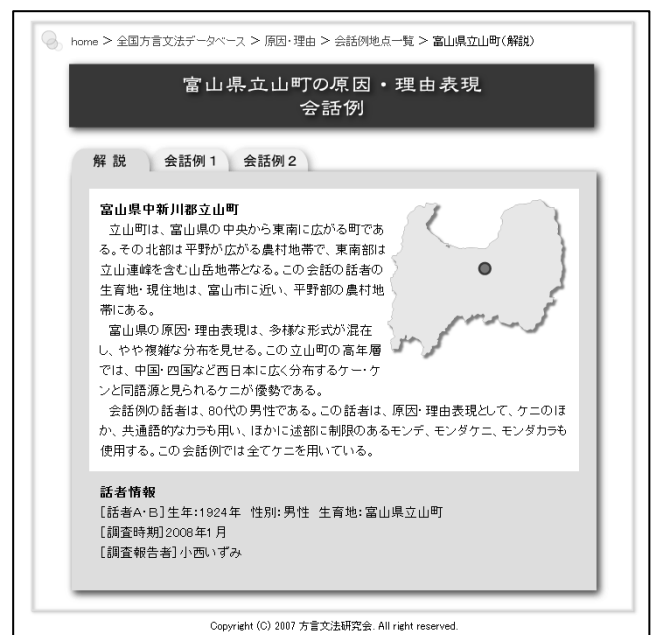


図 16 会話例

会話例 1, 会話例 2 のタブでは実際の音声聞ける。



図 14 調査地点一覧

### 謝辞

本研究は科学研究費補助金「日本語諸方言の文法を総合的に記述する『全国方言文法辞典』の作成とウェブ版の構築」(基盤研究(B)課題番号 21320086)の成果の一部である。

### 参考文献

- 1) 沼田 善子, 野田尚史編集「日本語のとりたて—現代語と歴史的变化・地理的変異」, くろしお出版 (2003).
- 2) 徳川宗賢監修「日本方言大辞典」, 小学館 (1989).
- 3) 方言文法研究会編「全国方言文法辞典データベース(Web版)」, <http://hougen.sakura.ne.jp/>.