

携帯電話を利用した 首都圏若年層の言語調査

鎌水兼貴[†]

首都圏の若年層の言語状況を調査するため、主に大学での授業時の利用を目的とした、携帯メールによるアンケート調査システムを開発した。携帯メールは回答側には手軽な手段だが、データの集積から研究への利用までを円滑に行うためには、専用のシステムの開発が必要となる。

本発表では、利用例とともにシステムの概要と利用例を紹介する。

Linguistic Survey of the Young Generation in the Tokyo Metropolitan Area Using Mobile Phone

Kanetaka Yarimizu[†]

To investigate the language use of the young people in the Tokyo metropolitan area, I developed a linguistic survey system using e-mail of mobile phone for mainly students of university. It would be easy to conduct the mobile phone survey. But it is difficult to accumulate data and to use them for linguistic study. Therefore, a dedicated system needs to be developed.

In this presentation, I show the overview of the system and the linguistic maps using this system.

[†] 国立国語研究所
National Institute for Japanese Language and Linguistics

1. はじめに

首都圏の言語研究は、その複雑性から、研究対象として敬遠される傾向があるように思われる。都市部は一般的に在住者の流動性が高いため、生まれた土地で生育し、仕事を得て、成人後もその土地に一生住み続ける、ということは稀である。特に大都市の場合には、他地域からの流入も多く、その土地の言語状況を捉えづらい環境にあるといえる。そうした大都市部においては、また地域社会における人々の交流も希薄であり、調査協力者を探すことも容易ではない。

さらに首都圏の言語は、いわゆる標準語（共通語）との言語的共通性が高いだけでなく、標準語自体をも変化させる力をもっている。そのため、首都圏の言語研究は、研究の枠組みそのものが捉えづらいという問題を抱えているあといえよう。

そうした中で、本研究では、首都圏における若年層の言語状況を調べるための効果的方法の追求するために、携帯メールを利用した調査方法の開発を行った。

2. 携帯メールによる調査

2.1 アンケート調査について

複雑な言語状況を解明するためには、まずは研究者にとって最も実施しやすい環境での調査を繰り返しながら、適切な調査内容を構築していくことが望ましいと考える。

研究者にとって実施しやすい調査は、おそらく大学生に対する授業時のアンケート調査ではないかと思われる。授業という半ば強制力をもった場を利用することで、学生から短時間に効率よく回答を集めることが期待できる。複数の大学で連携して授業でアンケートを行えば、現地で調査を行わなくとも広範囲のデータを集めることも可能である。

しかし、アンケート調査にも問題点がある。調査者が回答者一人一人と相対して質問をする面接式の調査と比較すると、アンケート調査は、質問の意図を回答者全員に徹底させることが困難である。もし質問文に疑問をもったまま回答したり、質問の意図を誤解した回答をしたり、回答意欲が低く適当に回答するなどのことがあっても、その場で再回答を促すことができないため、どうしても一定割合の不正確な回答が生じてしまう。

アンケート調査は回答者数を多くすることで、そうした回答の誤差を減らすことが一般的であるが、今回の目的のようにさまざまな角度から調査を繰り返したい場合には、データが大量に集まることもまた、問題となる。従来のアンケート用紙を用いた調査では、データ入力、集計に労力がかかってしまう。それだけでなく、授業時間

にアンケートをする場合は、授業への影響も最低限しなければならない。

そのため、データは最初から電子化されていることが望ましく、ウェブの入力フォームか携帯メールを利用したデータ送信システムの構築が必要となる。

もし短時間に集計結果を学生に示すことが出来るような調査システムを構築できれば、アンケート調査は、授業時間を邪魔するものではなく、授業に活用できるツールに変わるであろう。学生との関係も、調査を強制する/されるという関係から、協力的な関係を構築することが可能になると思われる。

2.2 携帯メールの利用

携帯メールやウェブを用いたアンケート調査自体は珍しいものではない。大学においてもコンピュータ端末を利用できる授業の場合には、容易に実施できるであろう。通常の教室で電子データによる提出をする方法は、学生の所有する端末、すなわち携帯電話を利用することになる。しかし、携帯電話の利用には通信料金が発生する。これは学生に対して有料サービスを行わせることを意味する。そのため携帯電話からのウェブアクセスは慎重であることが望ましく（学生に料金プラン上問題がないか確かめる必要がある）、通信料金を抑えられる携帯メールの利用が妥当だと思われる。

携帯メールの場合は、送信と同時にメールアドレスが相手に伝えられるため、個人特定が容易となる利点がある。つまり、一回登録すれば、その携帯電話から回答を送る場合には、名前を入力する必要がなくなるため、調査側の入力負担を軽減できる。複数回に調査を分割して実施することも容易である。また、誤入力時に回答者に修正を促すこともできる。

ただし、メールアドレスの蓄積でもあるため、システムの構築には細心の注意を払わなければならない。

2.3 方言研究における調査集計システム

方言研究においては、コンピュータを用いた言語地図の作成という目的があったため、それに付随してアンケート集計プログラムも作成されている。統計ソフトに近い体裁をもつ荻野綱男の GLAPS[1]や、大量の地図出力に適した高橋顕志の SUGDAS[2]が代表的であるが、どちらもデータ収集部分の自動化までは行われていない。データ収集が自動化されているものとしては、国立国語研究所で実施された「方言メール調査」[3]があるが、これは逆に言語地図の出力が自動化されていない。

選択式の調査のような回答の整理が必要ない調査の場合には、調査結果の自動集計が容易であるが、自由回答式の調査の場合には、回答のカテゴリー化に時間がかかるため、簡単には出力できない。高橋は GAJ-Sugdass[4]において、回答語形を画面上で選択する方式で、回答の整理を容易にしておき、本システムを作成する際にもこの点を重視する必要があるだろう。

3. システム概要

3.1 全体図

システムの概要を図 1 に示す。調査を行う段階順に大きく分類すると、調査管理、回答処理、データベース、結果表示の 4 つからなる。データベース部分以外はプログラムによって実現される。

基本的には、調査者がシステムを用いて調査項目を登録し、回答者がシステムの示す調査項目に対して回答のメールを送信することによって、自動的に回答データがデータベースに蓄積され、集計結果を即座に見ることが出来る、という流れになる。

なお、「リアルタイムに結果を見ることが出来る調査」ということで、この調査を "Real-time Mobile Survey (RMS)" (リアルタイム携帯調査) と呼ぶことにする。

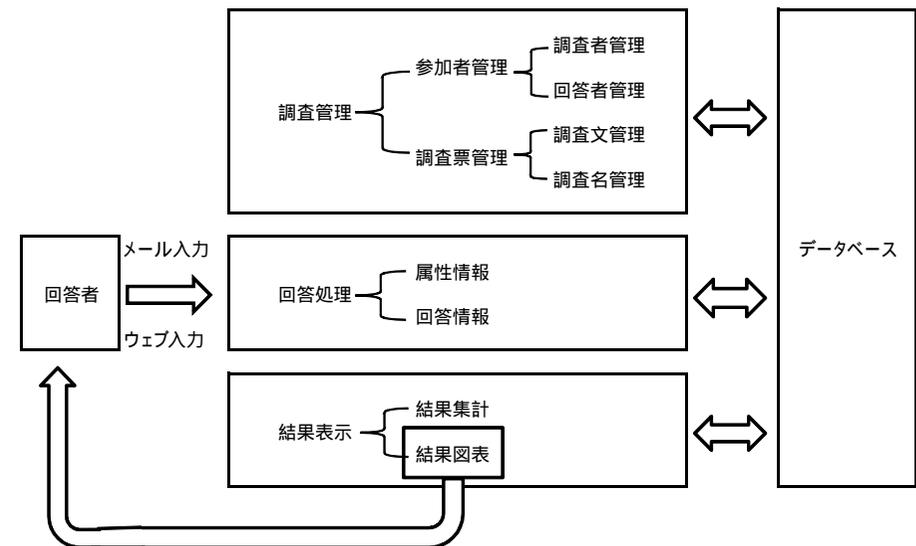


図 1 システム構成図

3.2 各段階の説明

以下、RMS システムの各段階を説明する。なお、現時点（2011 年 8 月）では実装されていないものもある。

3.2.1 調査管理

調査管理の部分は、調査そのものを管理する部分である。a 参加者管理、b 調査票管理という二つの部分からなっている。

a.参加者管理

調査者と回答者が異なるさまざまな調査を一括して管理するため、まず、参加者の管理が重要な部分となる。参加者管理は、調査者管理と回答者管理の二つに分かれる。

調査者管理：個人的な調査をする場合と、複数の調査者でグループを作る場合とがあるため、調査者には、個人の管理のほかに、グループ単位の管理が必要となる。システムへのアクセス件の管理もするため、調査者はIDとパスワードを登録する。

回答者管理：回答処理中の属性情報の登録によって行われる。複数の授業に出席する学生によって回答者が重複することがある場合には、回答の共有レベルの指定に従って、回答者を統合することが出来る。

b.調査票管理

基本的には調査は調査票単位で管理される。しかし、同じ調査票で異なる対象で行うこともあれば、異なる調査票を同じ集団で行う場合もある。このほかにも調査票の改訂などもあり、単純に調査票 = 調査とはならないことがある。分析時の調査結果同士の比較のことも考慮すると、システム上では、個々の質問文単位で管理し、調査者ごとに、その質問文を「調査名」を与えてグループ化することにする。

そのため、調査票管理は、調査文管理と調査名管理に分かれる。

調査文管理：質問文、質問の種類、絵や音声の有無、選択肢、関連質問などの調査文の情報が管理される。作成者は共有許可を与えることができる。

調査名管理：複数の調査文をグループ化し、一つの調査として扱うことができるようにする。調査文はさまざまな整理方法があるため、タグ付けや階層化などが必要になると思われるが、「言語地図データベース」[5]のような独立したデータベースを別に作成して、本システムから参照するようにしたほうがよいかもしれない。

3.2.2 回答処理

調査実施時に回答を受け取る部分であり、システム全体の重要な部分の一つである。多くの場合、回答はメールで受け取るのだが、ウェブ上での入力にも対応することに

する。回答メールは、サーバーのメールアドレスに送信され、メールサーバーの転送機能を用いて、メール内容をプログラムに入力し、あらかじめ指定した方法によってデータベースに蓄積する。データは回答者の属性情報、質問への回答情報にわかれる。

a. 属性情報

回答者は、最初から調査に参加できるわけではない。最初に登録を済ませておくことで、複数回調査する場合に個人特定をする必要がないようにする。登録は回答者の属性情報の登録時に行われる。属性情報送信用のメールアドレスに送信することで完了する。町丁目単位の生育地、転居歴、性別、配偶者・両親の出身地といった情報がデータベースに登録される。

調査時に携帯電話を忘れたり、電池の残量不足で携帯メールが使用できない場合がある。こうした場合に対処するために、登録時に特定のIDを1つ以上登録しておくことが推奨される（学生対象の場合には学籍番号や氏名が一般的）。これにより、ウェブで入力した場合でも回答の蓄積が可能となる。

メールアドレスや学籍番号、氏名といった情報が自動的にデータベース化されてしまうので、学生の許諾が無い限りは一定の調査期間が終了したら、個人を特定可能な情報はデータベースから自動的に切り離すようにする。

b. 回答情報

回答データは、回答データ送信用のメールアドレスに送信することで完了する。件名には前述の調査管理で指定した「調査名」を指定することで、自動的にその調査に対する回答とみなされる。本文には、行頭に設問記号であるアルファベット1字が入り、続いて選択肢や自由回答が入る（図2）。

調査者が事前に登録をしていない調査については、調査名を空欄にすることで行うことができる。

メール作成 新規	
宛先	chosa@xxx.xx
件名	shuto1
本文	a 1 b 2 c ウザイ d 思わない

図 2 回答メール作成画面例

回答の制限時間を設ける場合には、設問記号も不要となり、本文に回答するだけでデータの蓄積が可能である。

データの登録状況を後から確認可能にするため、送られてきたメールは、オリジナルのメール、メール内容を所定の項目に振り分けたもの、振り分けた内容を指定方法によって変換したもの、という3つの状態で蓄積される。の変換は、後述する結果集計の段階でも行うが、ここでは文字コード・改行コードの変換や、表記(半角・全角、ひらがな・カタカナなど)の統合のような、簡単な処理だけを行う。

3.2.3 データベース

調査管理や回答処理で蓄積されたデータは、データベースに格納される。データベースといっても、テキストファイル(XMLファイルの場合もある)として調査者単位で、ディレクトリにて管理される単純なものである。

テキストファイルでしかないものの、個人情報が入ったファイルが多いため、今後はセキュリティも考慮にいれて、データの暗号化なども考慮に入れる必要があるだろう。重要な課題である。

3.2.4 結果表示

結果表示部分は、集まったデータの処理に関する部分である。a 結果集計、b 結果図表の2つに分かれる。

a. 結果集計

結果の集計部分は、データベースに格納されたデータを指定した方法で整理・分類する部分である。回答の修正もここで行うことになる。

出力方法は二種類で、全データの表形式による出力、回答の基礎集計出力がある。

全データの表形式による出力: 回答情報を、回答者×質問の1枚の表として扱い、その表の中のセルを選択することで、回答の閲覧を行い、修正・整理を行うことができる。表は、属性情報と回答情報とに分けて表示される。

修正画面では、元のメール内容も表示され、変換が正しいかをチェックすることができる。回答ミスによる項目のずれの修正などもここでを行い、データベースに結果を反映させることができる。属性情報の部分では、結果図表において言語地図を表示するための、住所情報から緯度経度情報への変換も行う(図3)。

生育地	緯度	経度	信頼度
東京都江戸川区平井	35.703	139.85	5
静岡県静岡市清水区興津中町	35.056	138.52	5
兵庫県相生市赤坂	34.814	134.48	5
東京都新宿区東五軒町	35.706	139.73	5
埼玉県上尾市小泉	35.969	139.57	5
神奈川県厚木市松枝	35.448	139.36	5

図3 住所から緯度・経度への変換例

回答の基礎集計出力: 項目ごとの異なり語形を出力し、回答を整理するための処理である。回答処理の段階では、項目ごとに回答を項目ごとに振り分けることが中心であった。選択式の質問の場合には、結果集計段階での処理はあまり必要としない。しかし、自由記述式の質問の場合には、分類が難しい回答もあらわれる可能性が高く、調査者が内容を読んで分類することも多くなる。異なり語形の出力だけでは、回答が多岐にわたった場合に分類が難しくなるため、回答語形間の類似度[6]などを用いるなど、分類の効率性を高めていきたいと考えている。

b. 結果図表

結果の図表による出力部分も、このシステムの重要な部分の一つである。前述の集計結果を、地図やグラフによって視覚的に表示するもので、これについては、以下に結果の表示例を示すことにする。

4. 利用例

現時点(2011年8月)では、以上のシステム全てを実装しているわけではないが、回答者を登録し、調査を実施し、地図を出力する一連の流れについては完成している。ここでは、2011年6~7月に実施した調査テストの結果の一部を紹介する。

以下は、筆者が担当する、埼玉県・東京都・神奈川県の上3大学における4つの授業において行った調査結果を地図化したものである。回答者は授業の受講者であり、全体で300人程度の回答が集まった。

調査は、まず質問文と選択肢を板書にて示し、回答方法の指示は筆者が口頭で行った。

図4は、「燃やす」の意味で「モス」を用いるかどうかについて尋ねた結果の出力例である。この質問は選択式で、「1:使用する, 2:聞いたことがある, 3:知らない」から選択するものである。それぞれ地図出力時には、「1: , 2: , 3: -」の記号を対応させた。記号の位置は、回答者が「5~15歳までに最も長く居住した場所」としたが、転居歴は考慮せず表示したため、方言学でいう「生え抜き話者」にはなっていない点に注意する必要がある。

図4をみると記号が関東地方周辺に著しい偏りを示している。これは当然のことで、首都圏1都2県の大学に通学する学生のデータであるために関東地方に集中している。大学別に表示すればその大学の通学圏がわかるだろう。

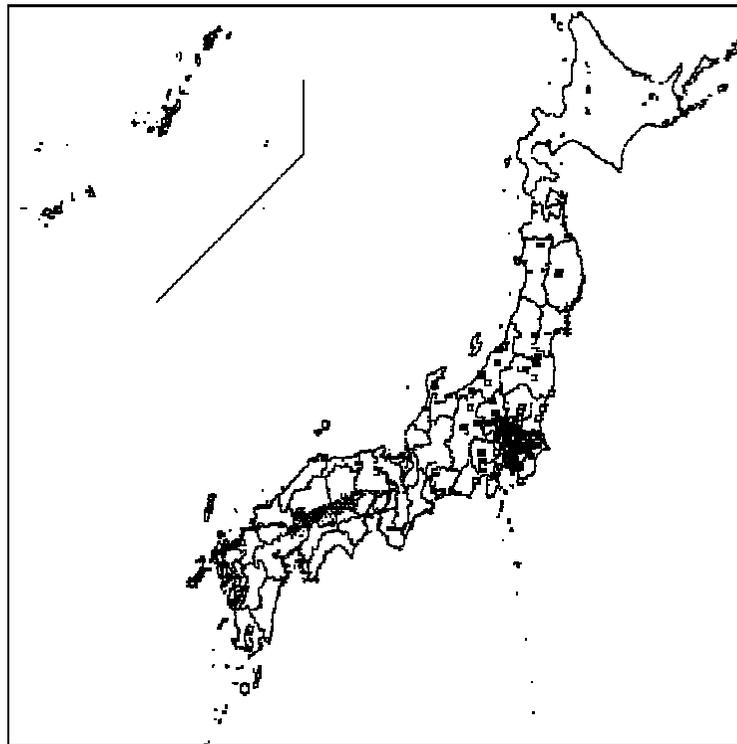


図4 「モス」の使用(全国)

この状態では結果の分析が難しいため、関東地方を拡大した結果が図5である。みると「モス」の使用者は、北関東寄りに偏っており、東京に近いほど、不使用者が増加していることがわかる。全体としては聞いたことがある人という人が多いのだが、それでも東京中心部では知らないという人が目立っている。

この結果は3つの大学の結果を合わせたものであるが、同じ関東地方に在住していても言語の使用に差があることを効果的に示すことが出来た例ということができるだろう。

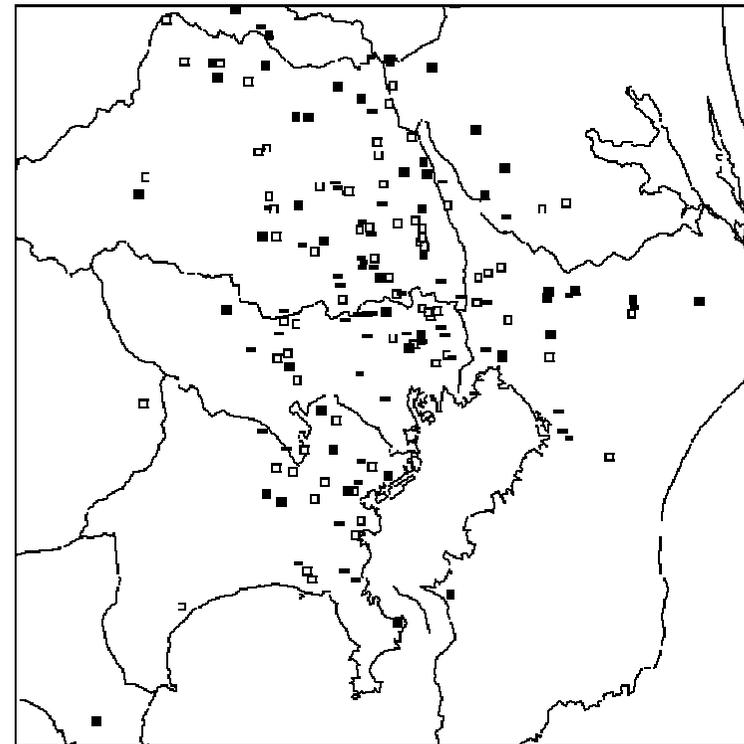


図5 「モス」の使用(関東地方)

つづいて図6は、「ツマグロオオヨコバイ」という虫について、「バナナムシ」という呼称を用いるかどうかについて尋ねたものである。これは東京西部でこの呼称があるということを聞いたため急に行った調査で、即時にデータを収集し結果を示すという、RMSのシステムには都合のよい事例であると思われる。

図6の凡例は、図4・図5と同様である。東京都南部の回答者が欠けているのが難点であるが、埼玉県南部から神奈川県北部にかけて使用者が分布していることがわかる。千葉県での使用者はほとんどなく、東京都の東部においても、使用()や聞いたことがある()という回答がほとんどないため、東京都でも東西に使用の差があるように見える。しかし、この差がどこから来るものなのか、白地図では解釈しにくい。

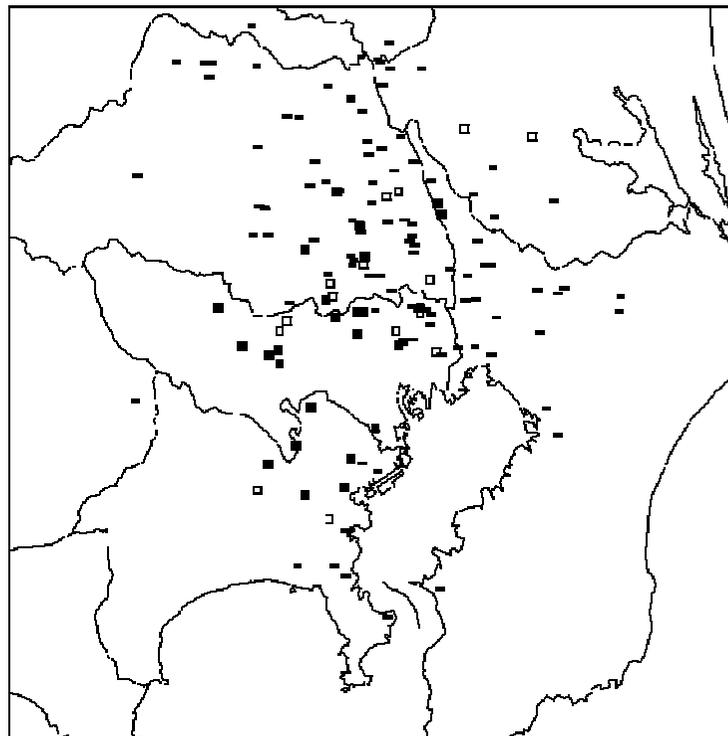


図6 「バナナムシ」の使用(関東地方)

そのため、河川のデータを重ね合わせてみたものが図7である。東京都心部での使用について、河川からみてみると、荒川によって使用が分断されていることがわかる。埼玉県と東京都の境でも荒川の南西側に多く使用者がいることから、現時点では「バナナムシ」の勢力は、東京都西部が中心であり、北東側の分布は荒川を越えていないのではないと予想できる。

調査を行うまで、「バナナムシ」の呼称の使用者が地理的に偏っているということはわからなかったため、授業で地図が示されたときには学生の反響が大きかった。首都圏では学生に地域差の例を示すことが難しいため、リアルタイムで結果を示すことができたことは、教育効果的にもよかったと思われる。

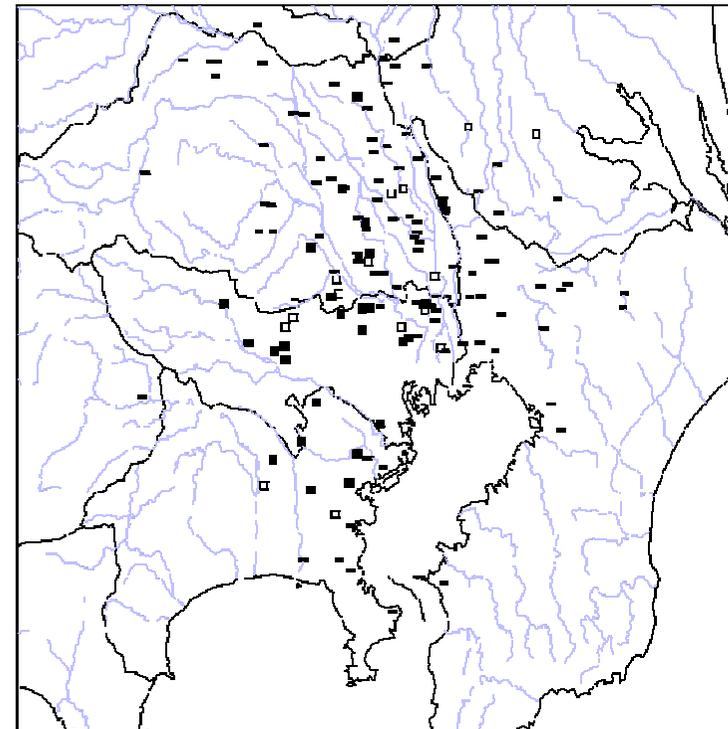


図7 「バナナムシ」の使用(関東地方・河川表示)

5. まとめ

以上、携帯メールを用いた調査"RMS"のためのシステムの概要と、その利用例を示した。今後システムを利用して、首都圏における多様な言語状況の解明に役立てていきたい。

現状ではすべての機能が実装されていないため、当面の課題としては、各機能の実装のほか、利用を通じて生じた問題点の改良に努めていきたいと考えている。機能的な改良だけでなく、操作性の改良なども通じて、授業という貴重な調査場面でより効果的かつ資料性の高い調査を可能にすることを目指したい。

また、方言資料のデータベース化という点では、方言調査データのXML化の研究[7]や、方言資料を統一的に扱うための研究[8]も進行中であり、それらと組み合わせていくことで、日本の方言研究の発展にも寄与できるものと考えられる。

参考文献

- 1) 荻野綱男, パソコン版GLAPSの開発, 文部省科学研究費研究成果報告書, (1996).
- 2) 高橋顕志, 「四国言語地図」を作る, 日本語学(明治書院) 11-8, (1992).
- 3) 大西拓一郎・鎌水兼貴・三井はるみ・吉田雅子, 方言の形成過程解明のための全国方言調査方言メール調査報告書, 国立国語研究所共同研究報告 10-2, (2011).
- 4) 高橋顕志, 日本語方言の層位 GAJ-Sugdas 2006, (2007).
- 5) 竹田晃子, 言語地図データベースの特徴と利用方法, 国立国語研究所共同研究プロジェクト「方言の形成過程解明のための全国方言調査」公開研究発表会, (2011).
- 6) 鎌水兼貴, 「方言文法地図」における共通語化の状況 多変量解析を用いた分析, 日本語学(明治書院) 26-11, pp.112-119, (2007).
- 7) 鎌水兼貴・小西いずみ・松丸真大, 方言調査データのXMLによるデータベース化, 第88回人文科学とコンピュータ研究発表会, (2010).
- 8) 鎌水兼貴, 多様な方言資料を統一的に扱うための検索システムの開発, 日本語学会 2011年度秋季大会, (2011 予定).

本研究は、国立国語研究所共同研究プロジェクト(萌芽・発掘型)「首都圏の言語の実態と動向に関する研究」(プロジェクトリーダー:三井はるみ)の研究成果の一部として発表するものである。