

言語地図作成の電算化 『方言文法全国地図』第5集を例に 大西拓一郎

「日本語学」(明治書院)21-11(2002年9月号)

1. 言語地図の作成過程

言語データが広くパソコンで処理されている中、言語地図を描く作業は、多くが手作業にゆだねられてきた。まずは、これまでの問題点を整理してみる。

1.1. 手作業地図の限界

言語地図では、個々の語形に記号を与え、各語形が使われる地点に記号を置く表現法が広く採用されている。手作業では、ハンコや製図用テンプレートを利用して白地図上に記号を置き、地図を作成する。

でき上がってみると、うまく分布が示せなかった、記号が気に入らなかった、間違いが見つかった、汚れてしまった等々、さまざまな理由で書き直しが必要になることは少なくない。再度作成するには、同じ作業を繰り返さざるをえない。

作成した言語地図を論文や書籍などにする場合、二つの方法がとられる。一つは作成した地図をコピーして取り込む方法(オフセット法)で、もうひとつは手作業地図を原稿として、印刷所が書き直す方法(トレース法)である。第四集までの『方言文法全国地図』(GAJ)は、後者の方法をとっていた。規模の大小を問わず、専門家がトレースした方が印刷物としての品質は格段に高くなる。

トレース法では、原稿の作成と印刷用版下の作成という二段階での地図作成を要する。版下の作成は、印刷所とのやりとりなので校正が必要である。印刷所を介する以上、校正の繰り返しはコストアップにつながり、市販物だと価格に反映される。

1.2. 電算化に至るまで

データが整理されていれば、地図化の作業は、基本的に単純作業である。したがって、電算化は古くから試みられてきた(徳川1993)。なかなか実現できなかったのは、地図のような画像の扱いがパソコンでは難しかったことが大きい。また、地図が描けてもトレースほどの品質は望めなかった。しかし、ここ数年でパソコンやプリンタの性能が飛躍的に向上し、ソフトウェアの機能も充実した。

GAJは第5集(国立国語研究所2002)で全面的に地図作成を電算化させた。パソコン上で言語地図を描くことや研究室レベルでも解像度の高いプリントアウトが可能となり、データに基づく印刷所とのやりとりもできるようになった。なお、単に地図の電算化と言った場合、手作業で作成した地図をスキャナでデータ化することも含みえる。しかし、ここでは作成作業の大部分を自動化することを扱う。

2. GAJ作成の電算化 仕様と準備作業

2.1. 基本仕様・プラグインの開発

基盤ソフト イラストレータ

GAJの電算化は、アドビ社の画像ソフト、**イラストレータ**(Illustrator)上で行っている。イラストレータでは画像全体をレイヤーと各レイヤー上のオブジェクトの組み合わせとして構成する(図1)。**レイヤー**とは透明なOHPシートのようなもので、**オブジェクト**はそのシートに置いた切り紙、全体はそれらのOHPシートを重ねあわせた切り絵のようなものである。地図で言えば、白地図の海岸線・県境、個々の地点に置いた記号、地点番号を示す各文字列などすべてがオブジェクトである。以下では、地図の記号を**記号オブジェクト**、各文字列を**テキストオブジェクト**と呼び、区別する。

イラストレータを用いるのは、次の利点による。第一点は、ドロー系の代表的画像ソフトであることによる。画像ソフトは、データの扱いにおいて**ドロー系**と**ペイント系**に

分かれる。違いは画像を扱うデータの性質で、イラストレータに代表されるドロー系がベクトルデータであるのに対し、(デジカメの画像処理で広く利用される)フォトショップなどのペイント系はビットマップデータである。地理情報学の世界では、前者を**ベクタ形式**、後者を**ラスタ形式**と呼び、区別する。地図一般に関して、どちらが良いかは一概に言えない。GAJのように一定の地点に必要な記号を置くという様式では、ベクタ形式が適し、拡大・縮小をかけても曲線にギザギザが生じない。

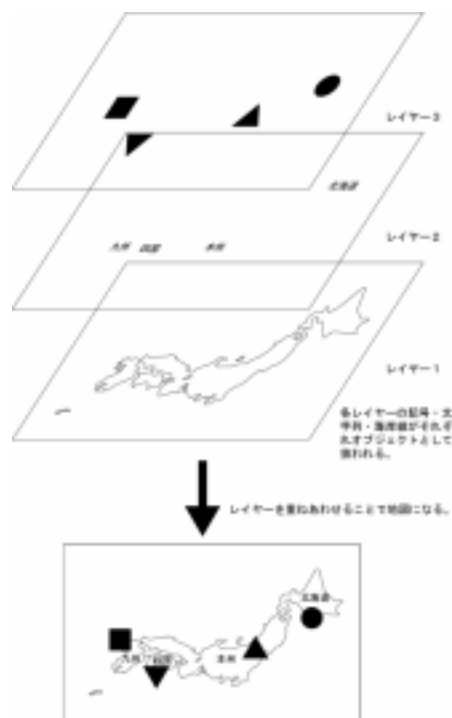


図1 イラストレータの概要

第二点は、イラストレータには、スウォッチと呼ばれる機能が備わっていることである。**スウォッチ**とは、各ファイルの中で記号や色など(正確には塗り)に名称を与えて管理する機能である。スウォッチを使えば、記号がコード化できる。この機能があったから電算化が実現したと言っても過言ではない。

第三点は、印刷業界で広く利用されていることである。ポストスクリプトと呼ばれるファイル形式に基づくことから、印刷所の業務用プリンタとの相性が良い。イラストレータで作ったデータは、版下に近い形で利用できるのも、トレース作業の省力化になる。

プラグインの開発

イラストレータは、地図作成を目的としたソフトではない。単体では、自動で地図を描くことはできない。必要な場所に手作業で記号を置いていくことは可能であるが、これでは作業場所を紙上からパソコン上に動かしたに過ぎない。

イラストレータでは、単体で持たない機能を追加する際に、**プラグイン**という方法が採用できる。必要な機能を持ったプログラムを作成し、イラストレータ本体に組み込むものである。言語地図は、特定の地点に必要な記号を置くことで完成する。ということは、地図上の調査地点に何等かの情報(GAJの場合、具体的には「地点コード」(後述)と呼ぶテキストオブジェクト)を与え、それを各地図が要求する記号に置き換えれば、言語地図になる理屈だ。このような機能を持ったプラグインを開発することでイラストレータは地図作成用ツールに変身する。

で利点の二番目に記したように、イラストレータにはスウォッチという機能があり、記号をコード化して管理することができる。ここに目をつけると、各調査地点に与えた地点コードとスウォッチのコードを介して記号データを繋げば、言語地図作成が自動化できるはずだ。このような機能を持つべく開発したのがウィンドウズ版イラストレータ

用プラグイン lms.aip である(以下、開発コードのまま、LMS(Language Map System に基づく)と呼ぶ)。

プラグインの仕様

言語地図作成用プラグイン LMS は、ウィンドウズ版イラストレータ、バージョン 8 で作動が確認されている(Mac には対応していない)。バージョン 9 以上での作動確認は十分ではない(バージョン 9 では、作動画面に多少不具合が生じるものの一応作動するようだ。現在市販されているバージョン 10 での作動は未確認)。

LMS は、イラストレータに三つの機能を追加する。a テキストオブジェクトをスイッチ登録された記号オブジェクトに置き換える。b 置き換えられた記号オブジェクトを別の記号オブジェクトに置き換える。c テキストオブジェクトを別のテキストオブジェクトに置き換える。簡単に言えば、オブジェクト間の置換機能に特化したプラグインである。

プラグインの組み込み方

通常の方法でインストールしたウィンドウズ版イラストレータの場合、エクスプローラを使えば、次のフォルダが見つかる(バージョン 8 の例であるが、それ以上のバージョンでも基本的に同様)。

c:\Program Files\Adobe\Illustrator 8.0\Plug-ins

このフォルダの中に lms という名前のフォルダを新規に作り、そのフォルダにプラグインファイルの lms.aip をコピーすれば、イラストレータ立ち上げ時に LMS は自動的に組み込まれる。

プラグインの適用方法

プラグインを機能させるには、テキストオブジェクトと記号オブジェクトのスイッチ名称とをマッチさせるデータが必要である。LMS では、コンマ区切りのテキストファイル(いわゆる CSV 形式、以下では CSV ファイルと呼ぶ)を用いる(LMS ではエクセルのファイルも利用可能であり、併用(特定地点の使用語形が複数に及ぶこと)は別の書式でも作動するが、ここではデータベースの基本的な形式にのっとった方法を紹介しておく)。

簡単な適用例を示す。例えば、**図 2**のようにテキストオブジェクトを置いておく。

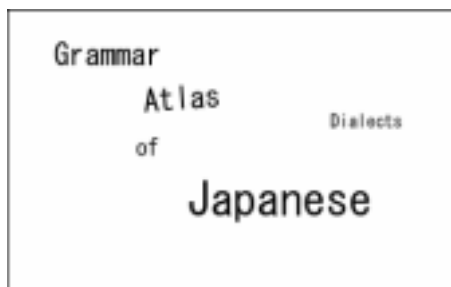


図 2

次に、**図 2**の 'Grammar'、'Atlas'、'Dialects' と記号オブジェクトを作り、それぞれに 1, 2a, 2b, 3 と名称を与えてパターンスイッチに登録する(**図 3**)。

図 2の Grammar を **1**、Atlas を **2a** と **2b** の併用、of を **3**、Japanese を **1** と **2a** の併用、Dialects を **2b** にしてみよう。必要な CSV ファイルは**表 1**のようなデータになる。



図 3 登録されたスウォッチ

Grammar, 1
Atlas, 2a
Atlas, 3
of, 2a
Japanese, 1
Japanese, 2b
Japanese, 3
Dialects, 3
表 1

このデータをプラグイン LMS にかけると、図 2 は図 4 のように書き換えられる。併用の場合は自動的にアークで各記号が繋がれる。この適用例からわかるようにテキストオブジェクトのサイズ(フォントの種類も)は無関係である。ただし、テキストオブジェクトの傾きは記号オブジェクトに反映される(図 2 の「Atlas」とそれに対応する図 4 の一連の記号が同様に傾いていることに注目)。なお、テキストオブジェクトやスウォッチの名称は、全角(2バイト、ひらがな・カタカナ・漢字)文字も一応使用可能であるが、誤作動の可能性があるので、半角(1バイト、アルファベット・数字)文字を使うことを強く推奨する。また、デフォルトで登録されているパターンスウォッチは削除しておこう。

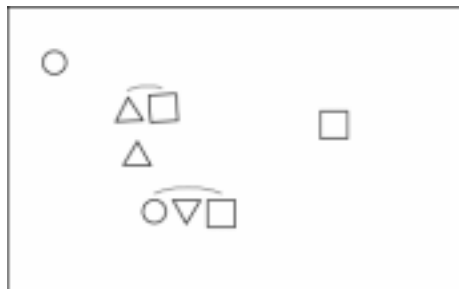


図 4

この例で用いた図 2 の Grammar・Atlas・of・Japanese・Dialects のそれぞれの一連の文字列(テキストオブジェクト)が言語地図上の調査地点に与える情報(例えば地点番号)が各語形に与える記号と類推すれば、どのように言語地図作成に活用できるか理解されるだろう。

2.2. 準備作業 白地図と記号

白地図

海岸線や県境など地理情報を白地図として用意することが必要であるが、できあいのものであれば、デザインエクステンション社の MAPIO シリーズが便利である(<http://www.dex.ne.jp/>)。GAJ の場合は、第四集までと同じ様式のものを作成した。

白地図には、調査地点にテキストオブジェクトを組み込むことが必要である。GAJ ではもともと各調査地点に調査地点番号という 6 桁のコード情報を用意しているが、6 桁の数字は、面積を要するとともに入力ミスを誘いやすい。そこで数字とアルファベットを組み合わせた三六進法 2 桁(0,1,2,3,...,a,b,c,...x,y,z の順で 2 桁に組み合わせる)に従い調査地点番号順にコードを与えた(表 2)。これを**地点コード**と呼んでいる。

プラグイン LMS の利用にあたっては、白地図での地点コードの埋め込みが重要である。注意すべきことの第一点は、テキストオブジェクトの左下隅(原点)を調査地点に置くことである。ここがプラグインを使って記号を置く際の基準点になる(記号オブジェクトの中心がここに置かれる)。第二点は、テキストオブジェクトの傾きは、プラグイン適用後の記号オブジェクトに反映される(図 2・4 の「Atlas」参照)ことである。地図の様式によるが、任意に記号を傾けることが基本になる場合がある。GAJ の場合、地図上の記号の水平垂直は、経度緯度線を基準とし、地図用紙の縦横には従っていない。

地点番号	地点コード
0228.96	00
0246.88	01
0247.31	02
:	:
9249.94	mc
9311.67	md
9313.46	me

表 2

白地図は、どのような地図であっても地図作成の基盤である。面倒に感じられるかもしれないが、ここで手を抜くと必ず後悔する。時間を惜しまず、慎重に作業したい。

記号セット

記号セットは、手作業地図のハンコセットに相当するものと考えればよい。GAJ の場合は、基本的な記号は決まっているので、これをイラストレータのデータとして作成した。後述するように、このデータは公開を予定している。各自で地図を作る際は、これを利用すると便利だろう。

3. GAJ 電算化作業の実際 「ありがとう(総合図)」を例に

白地図と記号セットを利用した地図化の手順について、第5集 270 図「ありがとう(総合図)」を例に示す。その際、イラストレータ上での各メニューからの操作手順を [ファイル] [終了] のように表現する。なお、イラストレータの基本的な操作方法については、詳しく記す余裕がない。マニュアル本の類が多く出版されているので、それらを参照してほしい。

3.1. 必要なデータの作成

凡例と記号

凡例用のファイルを作成し、そこに各見出しの文字列とそれぞれの見出しに対応する記号を置く。GAJ では、音声記号を多く使うが、研究室内の作業ではフリーで公開されている SIL IPA フォントを利用している (<http://www.sil.org/computing/fonts/encore-ipa.html>)。GAJ で利用する音声記号のかなりの部分がカバーされ、Mac と互換性があるとともにポストスクリプトにも対応している。

次にそれぞれの見出し語形に番号を与えるが、この番号が見出しに対応する各記号に与えるスウォッチの名称(記号コードと呼ぶ)となる。見出しと番号のレイヤーを別にしておけば、最終的に記号コードを表示しない出力形式が選べる。

記号の作成では、基本記号セットから、必要な記号をコピーし、色や傾きを与え、凡例上に配置する。補助記号(基本記号の上下左右に追加する微小な線など)を付加する場合は、該当する記号をグループ化しておく。記号の準備が整ったら、個々の記号をスウォッチに登録する。登録に用いる名称は、各見出しに与えた記号コードである。以上で図5の凡例ができる。

CSV ファイルの作成

電算化地図を作成するにあたり、記号コードと地点データ(GAJ では地点コード)の対応表を作成することが必要である。データの作成は、CSV 形式に書き出すことが可能であれば、何を用いてもよい。

3.2. 地図化

記号(スウォッチ)データの読み込み

白地図の元ファイルをコピーして、作業用の白地図にする。ここに 3.1. で作成したスウォッチデータを凡例ファイルからコピーして読み込む。スウォッチのコピーは [ウ



ウィンドウ] [スウォッチライブラリ] [その他のライブラリ]でファイル一覧のウィンドウに入り、凡例ファイルを選択する。この段階でコピー元の凡例ファイルに登録されたスウォッチがウィンドウで一覧表示されるので、コピー先(作業用白地図)のスウォッチパレットにドラッグアンドドロップする。なお、コピーする際は、コピー元のファイルは閉じておかないと警告表示が出て作動しない。読み込みが終わったら、コピー元のスウォッチ一覧ウィンドウは閉じておこう。

プラグインの実行

プラグイン LMS を実行することで地図化を行う(立ち上げ方は、**図6**)。プラグインの機能のうち、ここでは[2)テキスト オブジェクト置換]を選択する。なお、プラグインでは、テキストオブジェクトを「テキスト」、記号オブジェクトを単に「オブジェクト」と表現しているので注意。



図6

図7の「テキスト オブジェクト」ウィンドウが表示されたら、ファイル一覧のアイコンをクリックし(), 3.1. で作った CSV ファイルを選択する。また、置換作業対象とするレイヤー(GAJ では地点コードを置いたレイヤー)を選択する()。実行時のオプションが用意されているが(), 通常は「すべてのテキストを残す」「新しいレイヤを作成する」「併用地点のレイヤを作成する」を選んでおこう(特にバージョン9ではテキストを残すようにしておくのが無難)。



図7

準備が整ったところで、「OK」ボタンをクリックする()と、一気に言語地図が描き出される(**図8**)。ただし、最終版にするためには多少の調整が必要になる。

整形

プラグイン LMS の概要を**図9**に示した。LMS にできるのは、テキストオブジェクトを記号オブジェクトに一定の手順で置き換えることまでである。それゆえ、併用の一連の記号は、配列されることで面積をとり、周囲の記号と重なる場合がある。このような重なりを回避するように手作業で記号をずらして、見栄えを向上させる作業を**整形**と呼ぶ。



図8

手作業のGAJでは、併用の多い地点から順次押印し、最後に単用を押印するという手順を踏んでいた。手順で単用と併用のレイヤーを分けたのは、この品質を引き継ぐために、併用にはなるべく手をつけず、おもに単用で整形作業を行うことを念頭に置いたからである。この整形作業で、地図は完成する。

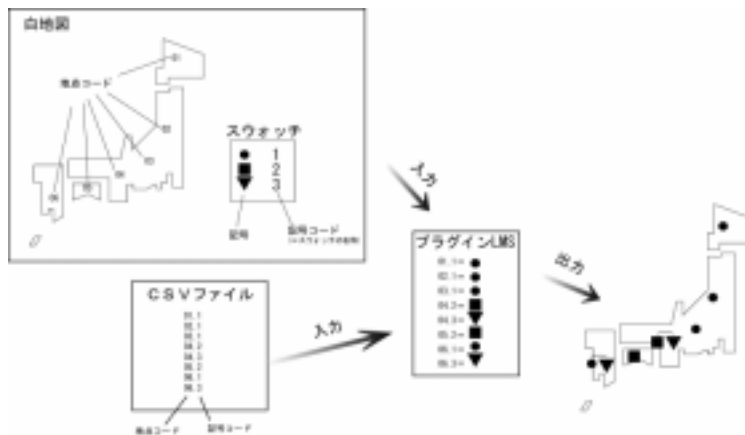


図9 LMSの概要

3.3. データとプログラムの公開

プラグインプログラム LMS, 白地図, 基本記号セット, 各地図で用いた CSV ファイルなどは, 国立国語研究所の方言研究のホームページ (<http://www.kokken.go.jp/hogen>) で公開を予定している。利用者自身の希望に応じた GAJ の書き換えやデータ検索のほか, GAJ 以外の言語地図の作成にも広く利用されることを期待する。

4. 電算化方言文法全国地図(iGAJ)の応用

電算化作業で作成した GAJ は, 作成グループ内では iGAJ と呼びならわされている (illustrator で作った GAJ に依る作業上のコードネーム)。iGAJ データは, さまざまな角度から利用が可能である。この点について説明する。

4.1. 書き換え

LMS の「テキスト オブジェクト」機能で作成した iGAJ 地図は, LMS により特定の記号オブジェクトを別の記号オブジェクトに置き換えることが可能である。例えば, 22 として登録していた をすべて に置き換えたい場合, に 999 のようなコードを与え

てスウォッチに追加登録し，プラグインの立ち上げ時に「オブジェクト オブジェクト置換」を選択すればよい。この機能は，多少の手直しや，別の記号で作成し直した場合の見栄えを確認する場合に有効である。ただし，大幅な書き換えを行う場合は，基本手順で作直す方が地図の品質は高くなる。

4.2. 略図の作成

GAJ は，カラーの大判であり，そのままでは論文などに引用しづらい。そこで，見出しを簡略化した地図に描き直すというケースは少なくない。これを実行するためには，CSV ファイルのデータを利用して，見出しを統合・捨象する必要がある。この作業には，研究者それぞれの意図が強く反映される。

例えば，3 で作った GAJ270 図では，アリガトの類が「ございます」にあたる部分も含めて細分化されている。これらを単にアリガトとしてまとめてしまう，といった手順を踏むことで，描き直すことにしよう。まずは，新たな統合を考え，原図と略図の見出しの対応をデータ化する。次に CSV ファイルのデータに対し，整理・並べ替えなどの処理を行い，結果を新たな CSV ファイルとして書き出す。また，必要な記号のスウォッチへの登録，凡例の作成など地図作りに必要な一連の作業を行う。最後にプラグインを実行させることで図 10 のような略図が作成される（なお，この図では，事前に略図用の白地図を新規に作成し，その上でデータ処理を行っている）。



図10 ありがとう

4.3. 閲覧用ファイルの作成

イラストレータは種々のファイル形式への出力をサポートしている。これにより閲覧を目的とした地図データが作れる。例えば、jpeg や gif 形式なら、ブラウザでの閲覧が可能であり、PDF 形式であれば、フリーでダウンロードできるアcroバトリーダーでの閲覧が可能になる(ただし、単純に書き出した場合、凡例に含まれる音声記号フォントの表示は保証されないので注意したい)。

5. 電算化言語地図の将来

ここではイラストレータを利用した GAJ の方法を紹介したが、他にも地図作成の電算化が実行されている。『大阪府言語地図』(岸江他 2001)や『富山県言語動態地図』(富山大学 2001)はエクセルとファイルメーカープロを利用する(岸江他 2000, 中井他 2000 参照)。福嶋秩子らによる SEAL(福嶋 2001)や高橋顕志による『中国・四国言語地図-1999-』(<http://home.hiroshima-u.ac.jp/hoogen/chus/>参照)の手法は、独自に開発したプログラムによる成果である。

以上のように言語地図の作成をパソコン上で行うことが可能になったことにより、研究者が言語地図を利用して論文を作成する際にぶつかる大きな壁 言語地図の作成と論文への掲載手段 も解消されようとしている。このことで言語地図を利用した研究の活性化が期待される。

最後に電算化言語地図の大きな課題にふれておこう。現段階の電算化は「地図作り」に目的が特化されている。本来、地理情報は、すべてが地続きの連続面(正確には三次元の曲面)上にある。最低限経度緯度のような情報は、共有化されてしかるべきものであるが、GAJ を含め、現在の電算化地図では扱いが統一されていない。

地理学においては、GIS(Geographical Information Systems)がしばらく前から脚光を浴びている(矢野 1999, 中村他 1998)。GIS の手法を応用すれば、分布に対し、数量化処理(例えば、具体的な距離について、少数の例外を除けば「中央から遠い・近い」程度の大まかで主観的な処理しか我々は行ってこなかった)を実行したり、各地の言語地図の情報をリンクさせたり、言語以外の情報と重ねあわせたり(例えば、山脈・交通網・海路・藩境などについて、行ったにしてもこれまた大ざっぱであった)など、さまざまな角度からの分析が可能になる。将来的には、方言の分布形成シミュレーションも期待される。

地図の作成も GIS の一環ではあるが、初歩である。本格的な GIS を目指した言語地図データのインフラ整備が求められる。パソコンで地図が描けただけで喜んでいるようでは将来は危ういと肝に銘じたい。

文献

- 岸江信介・中井精一・鳥谷善史・石田祐子(2000)「エクセルとファイルメーカープロを利用した言語地図の作製」『徳島県言語地図』作製を例として」『大阪樟蔭女子大学日本語研究センター報告』8
- 岸江信介・中井精一・鳥谷善史(2001)『大阪府言語地図』(近畿方言研究会)
- 国立国語研究所(2002)『方言文法全国地図』第5集(財務省印刷局)
- 徳川宗賢(1993)「電子計算機の「言語地図」作成への適用」『方言地理学の展開』(ひつじ書房)
- 富山大学人文学部日本語学研究室(2001)『富山県言語動態地図』
- 中井精一・坂口直樹(2000)「データベースソフトによる富山県言語動態地図の作製について」『富山大学人文学部紀要』33
- 中村和郎・寄藤昂・村山祐司(1998)『地理情報システムを学ぶ』(古今書院)
- 福嶋秩子・福嶋祐介(2001)『パソコンによる言語地理学 その方法と実践 SEAL ユーザーズマニュアル第5版』
- 矢野桂司(1999)『地理情報システムの世界』(ニュートンプレス)