

パーソナルコンピュータ上の対話型 言語地図作成支援ソフトウェア E G L

前川喜久雄
(国語学研究室)

0. はじめに

本稿では教育研究学内特別経費の配分を受けて筆者が開発中の、16ビットパーソナルコンピュータ上で稼働する言語地理学用のソフトウェア E G L (Editor for Geo-Linguistics) について報告する。この分野にコンピュータを導入すること自体、その目的・有効性の評価など、いろいろと議論することが可能であるが、本稿ではその種の議論は最小限にとどめ、もっぱら現在までに開発されたシステムの動作の紹介と今後解決すべき問題点の分析をおこなっている。

1. 開発の動機とシステムの特徴

鳥取大学国語学研究室では1984年度以来、鳥取・島根県境地域の言語地理学的調査を継続してきた。1988年夏までに364集落の調査を終了し、現在はその成果を『中海・大山周辺地域方言地図集』(Linguistic Atlas around Nakaumi and Daisen; LANDと略する)として公開するために分析を進めている⁽¹⁾。

言語地図の編集は一般に大変な時間と労力を要する作業である。国立国語研究所の『日本言語地図』の場合、一項目に平均200時間を必要としたというが⁽²⁾、LANDの編集も本学教育学部に大学院が設置されていないこともあって、かなりの困難が予想される。そこで、地図編集の能率を大幅に向上させるとともに人手による作業にともないやすい不正確を軽減するために、編集作業の全般を支援できるソフトウェアの開発をおこなうことにした。

言語地理学の分野でのコンピュータ利用にはいくつかの前例がある。その先駆けとなったのは徳川・山本(1967)であるが、一般に公開された実用レベルのプログラムとしては筑波大学の荻野綱男氏による方言調査分析用パッケージプログラム GLAPS が有名である(補注参照)。GLAPSは14000行程のFORTRANプログラムであり、統計パッケージのSPSSやBMDPのように大型機でのバッチ処理を想定している。筆者はこのパッケージプログラムを実際に使用したことはないが、そのマニュアルや、いくつかの解説記事(荻野1977; 1981a; 1981b; 沢木, 1984)から判断する限りではGLAPSの特徴は包括性と一般性の徹底した追求にあると言うことができそうである。言語地理学の分析で必要とされる編集・分析作業はまず漏れなくカバーされているし、大型コンピュータ間での移植性もきわめて高いようである。

このように高度に完成されたパッケージプログラムが既に存在する以上、新たに同じような動作をするプログラムを開発するのは無駄に思われるかもしれない。しかし、実際にはGLAPSにもいくつかの問題点を(その多くは上に述べた包括性と一般性の追求という長所の裏面と考えられるものであるが)指摘することができる。EGLシステムの開発にあたっては、その問題点を解決するために以下の新しい特徴を導入することを試みた。

① 対話型操作

例えばSPSS等の統計パッケージを利用するためには、やはり一種のプログラムを組む必要が生じる。このプログラムは主としてデータの入力フォーマットと統計処理プログラムのオプション選択に関する情報からなるもので、非常に簡素化されたものであるが、コンピュータにまったく経験のない人間にとっては(特に身近に親切な

指導者がいない場合)、これが意外と大きな困難となっている。GLAPSの場合にもGLAPS文法規則にしたがったプログラミングが必要とされ、同様の困難が生じる可能性が高い。

この問題点を解消するためにEGLでは言語地図の完成までに必要なすべての作業を画面上のメニューから選択する形で実行できることを目標とした。初心者でもつまづくことの少ない、いわばワープロ感覚での操作性を目標としたのである。この目標は現時点ではほぼ完全に達成されており、ファイル名の指定などごく一部の例外を除けば、ユーザーはテンキーとアルファベットキーの一部を押すだけでほとんどすべての操作を実行することができる。

② 多様な入出力

GLAPSで使用できる言語地図の出力装置はラインプリンターだけである。これは高い移植性を実現するための選択の結果と考えられるのだが、やはりラインプリンターの描く言語地図は表現力の点でものたりないものを感じさせる⁽³⁾。特に言語の地理的伝播に影響を与えることの多い河川・交通路・市町村境界などの地理情報を地図上に表現できないのでは、調査地域の地理に精通している調査者自身はともかく、それ以外の人間が眺める地図としては解釈に必要な情報を奪うことにもなりかねない。

EGLではパーソナルコンピュータの豊富なグラフィック機能を十分に活用し、各種の必要に応じてCRT上に説得力のある言語地図を作成できることを目標とした。また、デジタイザー、XYプロッタなどパソコン用に供給されている安価な入出力装置を積極的に利用することにしている。その結果、プログラムの移植性が著しく低下していることは事実であるが、使用機器の選択にあたってはできるだけ一般的な機種を選択するようにしている。この他に手書きで言語地図を清書する場合を想定して語形毎の地点番号一覧表を出力する機能を現時点でそなえており、また、プロッタ出力においては白紙に地形から作図する他に、あらかじめ地形などが印刷された白地図上に記号だけを印刷する機能もサポートする予定である。総じてさまざまな入出力形態に対応できることを目標としている。(ただし、XYプロッタによる出力に関しては現在まだプログラムが完成していないので、この報告書では構想を紹介するだけにとどまっている。)

③ 対話による語形編集

GLAPSでは言語調査で得られた語形を整数コードとしてコンピュータに入力する。したがって沢木(1984)も指摘しているように、調査資料の全体に目を通してすべての語形を洗いだしてからでなければデータ入力をおこなうことができない。また、マニュアルが親切にも指摘しているように意味と切り離された数字のコーディングにはエラーが発生しがちであり、特にキーボードに初めて触れる人間の場合はエラーが多発するものと予想される。

EGLはさきにも述べたように初心者に親切なシステムを目標としているので、通常の文字を用いて(2バイト系文字も可能)、調査票に近い姿で入力された語形を処理できるようにしている。文字による入力(意味と結びついた文字列の入力)を採用することによって入力エラーの発生率自体も低下するし、入力後のエラー検出も容易になる。しかし、入力された雑多な語形を ---- 換言すればS/N比の悪いデータを ---- そのまま言語地図に表現するわけにはいかないので、GLAPSが用意しているコード単位の強力な語形再編集の機能に代わる機能として文字入力された語形間の合併機能を備えることにした。後に詳しく説明するが、これはCRT上に表示された語形の一覧表を眺めながらカーソルの移動によって複数の語形をひとつの代表語形に合併していく対話型の編集機能である。この合併作業は従来の手作業による言語地図制作の過程で作図者が紙と鉛筆によっておこなっていたメモ作業を感覚的に大きなズレを生じない形でコンピュータ上に実現したものといえる。

2. 稼働する環境

現在EGLは次のような環境で稼働している。CRT上の言語地図をハードコピーするためにはこの他にプリンタが必要である。また()内はオプションの周辺機器である。

ハードウェア環境 NEC PC-9801シリーズパーソナルコンピュータ
メインメモリ384Kバイト以上
640×400ドットのカラーCRT
(ロジテック MYPAD-A3 K-510 mk2 デジタイザ)
(ローランド DXY-885プロッタ)

ソフトウェア環境 マイクロソフト/NEC MS-DOS (Ver.2.11 以上)
NEC N88-日本語BASIC (86) MS-DOS版
(インタプリタとして使用する場合は、コンパイルしての使用も可能だが別にランタイムライブラリが必要である。)

パーソナルコンピュータ本体としてNECのPC-9801シリーズを選んだのは性能によるわけではなく、たまたま筆者が以前から愛用していたこと、本学教育部には情報処理委員会の努力によって教育・研究用に自由に使用できるPC-9801UV21が11台設置されていることによるのだが、現在この機種が16ビットコンピュータ市場で占める位置を考えると、将来、学外への普及を考えた場合にもこの機種が適当であろうと思う。

ソフトウェアとしてMS-DOS上のBASICを選んだ理由は主としてデータの互換性を確保するためである。EGLで利用するデータファイルは市販のワープロやデータベースソフトを利用して作成するのが便利である。また、後に3.8で述べるようにEGLではMS-DOSの提供する各種のコマンドをチャイルドプロセスを介して利用できるようになっている。

3. EGLによる語形編集処理の流れ

フィールドワークから持ち帰った言語資料からEGLを使用して言語地図を作成するまでの処理のうち、語形編集処理の部分の流れ図をFIG.1に示した。以下ではこの流れ図にそってEGLの機能を紹介していく。

3.1 データファイルの作成

これはEGLを起動するまでに必要な作業である。EGLでは調査項目単位にデータファイルを作成する⁽⁴⁾。ワープロやエディタなどを利用して、調査票に記入されている語形をEGLが読める形のファイルにするわけだが、このファイルを以下では(EGL用)データファイルと呼ぶことにする。EGL用のデータファイルではひとつのレコードに以下のような順序でデータを並べる。各データの区切り(デリミタ)にはカンマ(,)を使用する。

地点番号, 語形1, 語形2, 語形3, 備考 ↓

「地点番号」は地図上の地点と一対一に対応し8桁以内でありさえすれば、どのようなものでも構わない。数字でも文字でも両者が混在していてもよい。筆者の研究室でLANDのデータを入力した際には国研式の8桁地点番号を使用した。「語形1」から「語形3」までは(地点番号に対応する)一人の話者が同一の質問項目に対して回答した語形が入る。つまりEGLは一地点で最高三つまでの語形が併用されている

データを処理できるように設計されている。「備考」には話者の内省などの情報を記入する。↓はレコードの終わりを示す標識（ターミネータ）であるが、ワープロで入力する場合には単にリターンキーを押して改行しておけばよい（BASICなどでデータを作成する場合にはCRコードを出力しておく。つまりただのシーケンシャルファイルである。）

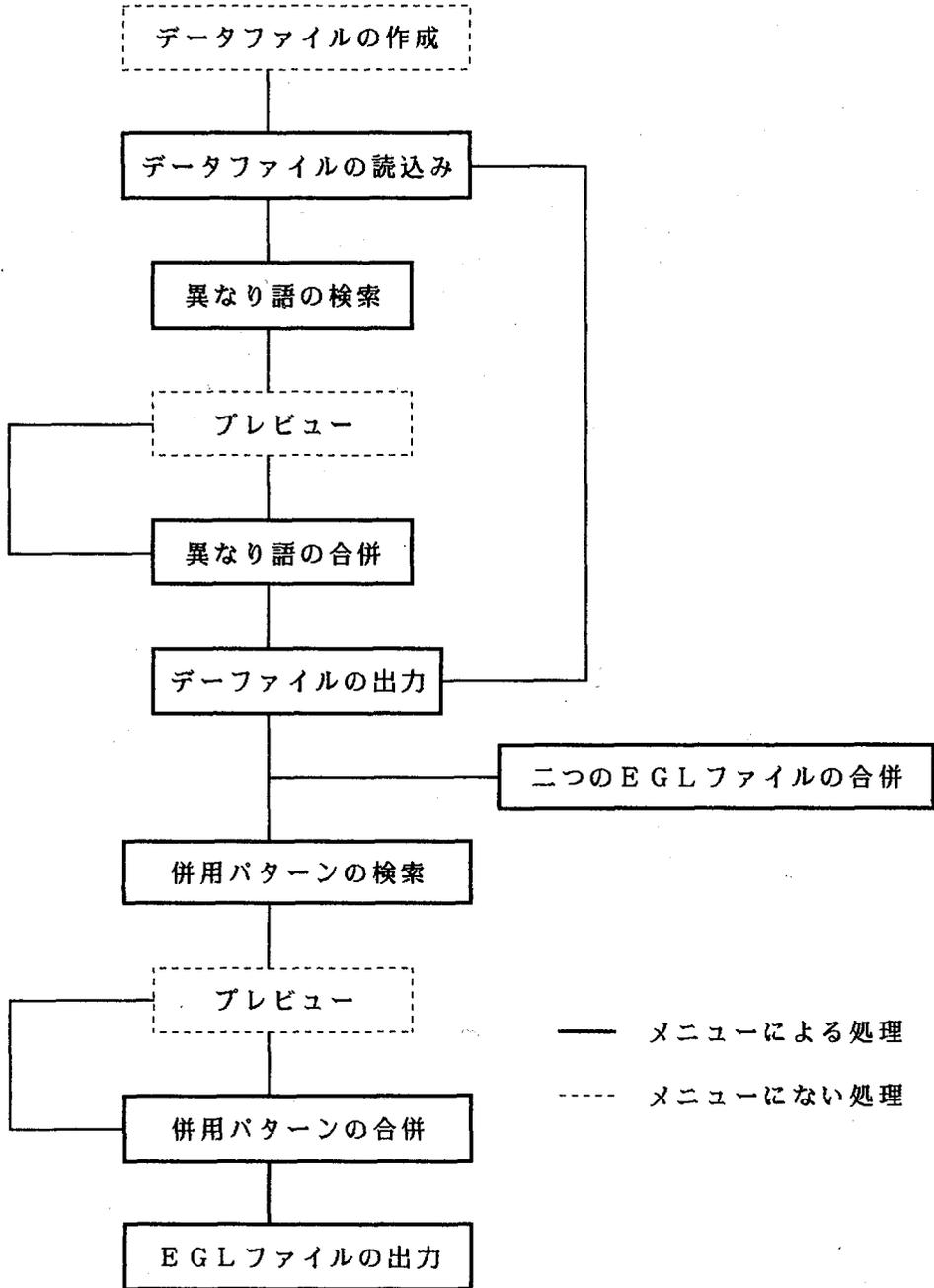


FIG.1 EGLによる語形編集の流れ（本文参照）

「語形1～3」の長さは一応それぞれ20文字(20バイト)までとする⁽⁵⁾。実は255文字までなら読み込むことが可能なのだが、後にCRT上で語形を編集する際に語形の一部しか表示されないことになる。この場合にも最終的な編集結果は保証されるので必要ならば20文字以上の「語形」を入力しても構わない。「備考」の場合は同様に一応65文字までとする。「語形2」以下が存在しない場合は単にカンマだけを打っておけばよい。つまりEGL用データファイルではひとつのレコードに必ず4個の、そして4個だけのカンマが含まれていることになる。先にも述べたようにEGLではカンマをデリミタとして使用しているので「語形」などのデータ中にはカンマを含めてはならない。また「語形1」には必ず何か記入されていなければならない。LANDのデータ入力では無回答の地点にはNRを、調査票の改訂による未調査地点にはNQを記入している。

ひとつのファイルに含まれるレコード数(地点数)は現在500を上限として設定している。これは暫定的な設定であり、コンピュータのメインメモリに640Kバイトを実装した場合、1000レコードぐらいは楽に処理することができる。ただし現実の問題としては調査地点数2400の『日本言語地図』を別格とすれば多くの言語地図は500地点以下であろうし、またこれ以上の地点数になるとCRTに記号を表示する際にもよほどうまく地点が分散していないと見分けることが困難になるだろう。

ファイルの名前(ファイルディスクリプタ)はMS-DOSの制限にかなっていればなんでもよいが、拡張子にEGL, HST, TTNは用いないほうがよい。これらの拡張子はEGLシステムで独自の意味をもっているからである⁽⁶⁾。LANDのデータファイルではQ100.DATのようにDATを用いた。

3. 2 EGLの起動とファイルの読み込み

EGLシステムを起動するためにはEGLのシステムディスクをドライブAにセットし、MS-DOSのプロンプトA>が表示されている状態でEGLとキーインしてリターンキーをおせばよい⁽⁷⁾。ディスクからEGLシステムが読み込まれEGLのタイトルが表示される。この状態で何かキーを押すと画面が変化し、スタートメニューが表示される(FIG.2参照)。(ちなみにEGLを終了させるためにはスタートメニューか、後述するメニュー1が表示されている状態でF10を押せばよい。)

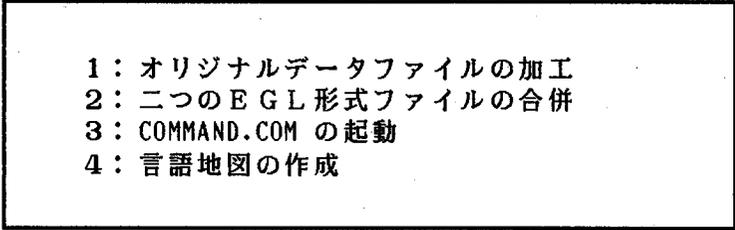
- 
- 1: オリジナルデータファイルの加工
 - 2: 二つのEGL形式ファイルの合併
 - 3: COMMAND.COMの起動
 - 4: 言語地図の作成

FIG.2 スタートメニュー(本文参照)

ここで既に作成されてディスクに格納されているデータファイルをEGLに読み込むためには「1」を選択する。画面が変化し、ドライブ名とファイル名を指定するように要求があるので、ドライブ名には読み込みたいデータファイルが格納されているフロッピーディスクがセットされているドライブの番号(A:, B:, C: など。詳しくはMS-DOSの解説書参照)を指定してリターンすると、ここで指定されたドライブ名が以後のファイル処理のためのデフォルト値として採用される⁽⁸⁾。続いてファイル名を指定するとEGLは指定が正しいかを確認してくるので、正しいければYを、誤っていればNを押す。Yが押されてファイルが無事に読み込まれると⁽⁹⁾、新しくメニュー1が表示される(FIG.3参照)。

- 1: 現在の編集状態を見る
- 2: 異なり語形の検索
- 3: 異なり語形の合併 (2の実行後に可能)
- 4: 併用パターンを検索
- 5: 併用パターンの合併 (4の実行後に可能)
- 6: スタートメニューに戻る
- 7: COMMAND.COM の起動

FIG.3 メニュー1 (本文参照)

3.3 異なり語の検索と合併

メニュー1はEGLによる語形編集処理の中核的な部分を担っている。通常考えられる処理の流れとしては、まず「2」を選んでデータファイルの中に含まれているすべての異なり語を洗い出す処理をおこなう。「2」が選択されるとメニュー2が表示されシステムは第何答までを処理の対象とするかを尋ねてくるので、必要に応じて「第一答だけ」「第二答まで」「第三答まで」の中から選択する。ファイルによっては「語形1」しか記入されていなかったり、高々「語形2」までということがあるであろう。そのようなときにここで適切な選択をおこなえば処理時間がかかなり短くなる。EGLは検索処理を、「異なり語の検索」→「異なり語のソート」→「各異なり語の頻度計算」の3段階に分けて実行するが、処理中の画面には各段階の処理が何パーセント終了したかが表示される。すべての処理が終了すると画面には異なり語の一覧表が表示され、異なり語の合併処理に進む (FIG.4 参照)。

A:Q50.DAT 異語 第三答まで

↑ ↓ ← → : スクロール F8: プレビュー ESC: 終了

NO	異語形	FREQ	NO	異語形	FREQ
1	NR	(6)	2	イ-	(1)
3	イイ	(1)	4	イリ	(2)
5	カ	(2)	6	キョ-	(14)
7	キョ-ト	(1)	8	キョ-ト-	(1)
9	クサワノカ	(1)	10	ジ(ツ)ョ-ト	(1)
11	ジョ-キジョ	(1)	12	ジョ-クイ(キ)-	(1)
13	ジョ-ゴ	(1)	14	ジョ-ト	(30)
15	ジョ-ト[y]	(10)	16	ジョ-ト-	(16)
17	ジョ-ト-[y]	(4)	18	ジョ-ト-シ	(10)
19	ジョ-ト-シ[y]	(1)	20	ジョ-ト-スイ	(3)
21	ジョ-トシ	(1)	22	ジョ-トシ[y]	(1)
23	ジョ-トス(シ)	(1)	24	ジョ-トリ[y]	(1)
25	ジョ-トルイ	(1)	26	ジョト[y]	(1)
27	ジョ	(2)	28	ゾ-クイ	(4)
29	ゾ-	(2)	30	ゾ-ト	(9)
31	ゾ-ト[y]	(1)	32	ゾ-ト-	(5)
33	ゾ-トリ	(1)	34	ゾ-トリ	(1)
35	ゾウト-シ(y)	(1)	36	ゾウトウシ	(1)
37	チ(ツ)ョトチ(ツ)ョト	(1)	38	チョ	(10)
39	ツインツイン	(1)	40	ツツ	(1)

合併の対象を選択してリターンキーを押してください (BSキーで指定終了)

FIG.4 異なり語合併処理中の画面 (LANDの調査項目50番「トカゲ」に含まれる異なり語形の一覧。語形NO6, 8は既に合併され白色で表示されており、代表語形NO7は水色である。NO8を囲んでいるのが疑似カーソル。)

最初に「合併のヒストリーを記録しますか？」という質問があるが、ここで Y を選択すると以後の合併処理の全過程が現在のデータファイル名の拡張子を HST に変更したファイル（ヒストリーファイル）に記憶される。ヒストリーファイルは編集作業のメモの役割を果たすものだが、将来はこれを利用して過去におこなった編集作業をシステムが自動的に再現する機能を実現する予定である⁽¹⁰⁾。

ヒストリーに関する選択が終わると続いて「代表形を選択してリターンキーを押してください」というメッセージが表示され、ここから実際の合併処理が開始される。FIG.4 では、LAND の調査項目「トカゲ」のファイル (Q50.DAT) に含まれる 67 の異なり語のうち最初の 40 個が表示されている。このうち /キョー/ /キョト/ の 2 語形をまとめて /キョト/ に合併するでしょう。

カーソルキー (→←↑↓) で語形を囲む疑似カーソルを移動させて /キョト/ にあわせ、リターンすると /キョト/ に関する表示 (通し番号と語形と頻度) が黄色から水色に変化する。これで /キョト/ が代表形に選択された。ここで「合併の対象を選択してリターンキーを押してください」というメッセージが出るので、/キョー/ と /キョト/ を順次選択してリターンすると、今度は選択された語形が黄色から白に変化し、頻度の表示が消えてなくなる。合併対象は一度に 30 個まで選択することが可能だが、選択が終了したら BS キーを押すか最後に選択した語形にカーソルをあわせてそのまま再度リターンすればよい。すると画面が変化し、選択された代表形および合併対象語形の確認がおこなわれる。

ここで N が押されると画面は代表語形の指定前の状態に復帰するが、Y が押されるとシステムはメモリに読み込まれた語形を検索し、合併対象語形があればそれを代表形に書きかえて代表形の頻度を調整する。この処理が終了すると再び異なり語の一覧表が表示され、代表語形の入力待ちの状態となる。一旦合併がおこなわれると、それ以前の状態を復元するためにはデータファイルの読み込みからやりなおしとなるので確認は慎重におこなった方がよい。また、合併処理の結果を確認したければエスケープキー (ESC) を押してメニュー 1 に戻り、「1」を選択することによってファイルの内容を確認できる。もちろんこの後さらに合併処理を継続することも可能である。

40 以上の異なり語を含むファイルの場合、最下行の語形に疑似カーソルをあわせた状態で ↓ キーを押すと画面が半分だけスクロールして次の 20 語形が示される。異なり語がある限り 500 語までなら何画面でもスクロールしてゆくが、最後の異なり語を含む画面ではもうスクロールをおこなわない。もちろん逆にスクロールして元の画面に戻ることも可能である。このスクロール操作は ROLL UP, ROLL DOWN キーでも可能である。

3. 4 代表語形の手換えと指定の取り消し

このような合併処理を好きなだけ繰り返して語形の編集を進めるのだが、適当な代表語形が検索された異なり語の中になく場合は、代表語形を書きかえることもできる。そのためには書きかえたい語形に疑似カーソルを合わせた状態でホームクリアキー (HOME CLR) を押してからリターンすればよい。画面下部に「新しい語形 (20 文字以内) >> 」という表示が出るので新語形を入力する。

また既に代表語形に指定してしまった語形を他の語形に合併することも可能である。この場合、画面表示が水色の語形を合併対象に指定しようとする時、システムは警告音を発し「エラー 代表形は合併の対象になりません」と言ってくるので、まず代表形の指定を解除する必要がある。そのためには水色の語形に疑似カーソルをあわせ、デリートキー (DEL) を押してからリターンすればよい⁽¹¹⁾。

このようにしてすべての合併処理が終了したらエスケープキーを押してメニュー 1 に戻り、「4: 併用パターンを検索」へと進むのだが、その前に合併処理作業においてしばしば威力を発揮する「プレビュー」の機能について説明を加えておく。

3. 5 プレビュー

ここまで語形の合併処理の方法を述べたが、実際の編集作業では二つの語形AとBを合併してよいかが語形を見ただけではわからないことが多い。その場合、決め手となるのは語形の地理的分布であることが多く、それを知るためにはともかく一回地図を描いてみる必要がある。もちろん必ずしも実際の地図を描く必要はなく、頭の中で分布を視覚化できればそれでよいのだが、国研方式の地点番号では地点間の関係を復元するのが容易ではない。また、A、Bの頻度が高い場合には想像力にも限界がある。このような場合にそなえて語形の編集作業中に特定の語形の地理的分布をCRT上で確認できるようにしたのがプレビュー機能である。ただし、プレビューを利用するためには、調査地点の座標を指定した「地点ファイル」(4. 2参照)を作成しておく必要がある。

プレビューを利用するためには異なり語の一覧表がCRTに表示されている状態でF8を押す。すると画面が変化するので、まず(最初にこの機能を利用する場合に限って)、地点ファイル名を指定する。続いて分布を確認したい語形を入力し、表示に用いるグラフィック記号を指定すると(5種類用意されている)、すぐにCRTに分布が示される(FIG.5参照)。指定された語形が存在する地点には指定したグラフィック記号が描かれ、それ以外の地点には1ドットの白点が打たれる。地点ごとに記号の色が異なることがあるが、これは第一答・第二答・第三答の別を水色・赤・黄色の三色で区別しているからである。ここでリターンキーを押せばもとの編集画面に復帰するが、続いて他の語形の分布を確認することもできる。その場合以前の分布を重ねて新しい分布が表示されるので最大5種類の語形の分布を同時に表示することができる。実際にEGLを利用して言語地図の作成をおこなうと、プレビュー機能によって思わぬ分布を発見することがある。

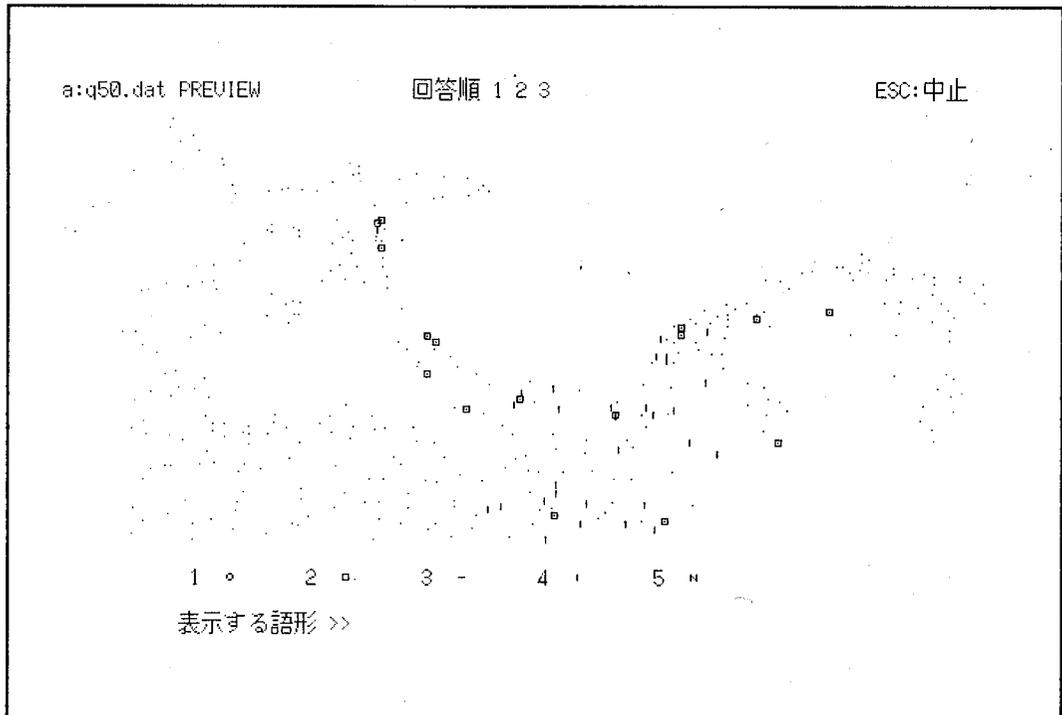


FIG.5 プレビュー画面 (FIG.4 中の/ジョート/[四角]と/ジョート/[縦線]の分布を重ねて表示している。四角は縦線を取り囲むように、いわゆる周圈的分布を示しているのので相対的に古い語形であると考えられる。この2語形は合併しない方がよいであろう。 FIG.8 と比較すれば調査地域の地理がわかる。)

3. 6 併用パターンの検索

異なり語の合併処理が終了すると、続いてメニュー1の「4」を選び、「併用パターンの検索」を実行する。異なり語の場合と同じ手順で操作すればよいのだが、ここでEGLが何をおこなっているのかを簡単に説明しておく。

EGLでは最初から各地点で（最高三つまでの）語形の併用があることを想定している。第一答=A・第二答=B・第三答=Cを（A, B, C）と書くことにし、下のようなデータがあったとする。

	地点1	(A, B, C)
仮想データ1	地点2	(A, D, B)
	地点3	(B, A, C)
	地点4	(D, A, B)

このデータに「異なり語の検索」を施すとA, B, C, Dの4語形が得られるわけだが、ここで語形Bを代表語形Aに合併処理すると、データは以下のように変化する。処理前には異なった順序対であった地点1と地点3が同一のパターンになっていることに注目してほしい。

	地点1	(A, A, C)
仮想データ2	地点2	(A, D, A)
	地点3	(A, A, C)
	地点4	(D, A, A)

このデータに対し「併用パターンの検索」の検索を実行すると、まずデータは以下のように書き改められる。

	地点1	(A, C,)
仮想データ3	地点2	(A, D,)
	地点3	(A, C,)
	地点4	(D, A,)

この後に検索が実行され、その結果3つのパターンが検出される。つまり「併用パターンの検索」ルーチンは①合併処理によって生じた語形の重複を調整し、②調整済みの併用パターンを単位とした異なり語（併用パターン）を検出するのである。そしてEGLは最終的に言語地図を描くにあたっては、この併用パターンを単位にして記号の割当などの処理をおこなうように設計されている。（この処理方法は通常の言語地理学における「併用形」の処理法といささか異なったものである。この点に関しては6で再度とりあげる。）

3. 7 併用パターンの合併

併用パターンの検索が終了すると画面には併用パターンの一覧が表示され、これに対して合併処理をおこなうことができる⁽¹²⁾。処理の方法は異なり語の場合と全く同一であり、プレビューの機能も利用できる⁽¹³⁾。ここでもまた必要ならば語形（パターン）の分布を確認しながら合併を進めていくのだが、調査項目によっては調査地域の全体にわたって全国標準語ないし有力な地域共通語が広く分布していることがあり、その場合、これらの語形が他の小地域語とまんべんなく併用されるので併用パターンの数が非常に多くなってしまふ⁽¹⁴⁾。このような時に利用するための機能として「無視語形の指定」と「回答順の無視」がある。

「無視語形の指定」は検索に先だって数個の語形を無視するように指定するものであり、先に3. 6で用いた仮データを再度利用して説明すれば、仮想データ2の検索

にあたってAを無視語形に指定すると検索結果は

```

          地点 1 ( C, , )
仮想データ 4  地点 2 ( D, , )
                地点 3 ( C, , )
                地点 4 ( D, , )
    
```

となる。もう少し複雑な例では、仮想データ5に対して無視語形1=B、無視語形2=Cと指定した場合、仮想データ6の結果を得る。

```

          地点 1 ( A, B, C )
仮想データ 5  地点 2 ( B, C, A )
                地点 3 ( B, C, )
                地点 4 ( C, , )

          地点 1 ( A, , )
仮想データ 6  地点 2 ( A, , )
                地点 3 ( C, , )
                地点 4 ( C, , )
    
```

つまり、複数の無視語形が指定されている場合、先に指定したものが優先して無視される。また無視語形のみからなるレコードでは優先度の高い語形から無視してゆき、最後に残った語形ひとつだけが残される。無視語形の指定をおこなうにはメニュー1で「併用パターンの検索」を選んだ際に表示されるメニュー3で「4：無視語形の指定」を選択すればよい。

次に「回答順の無視」であるが、これは検索において(A, B)と(B, A)、あるいは(A, B, C)と(A, C, B)(B, A, C)(B, C, A)などを同一のパターンと見なす指定である。この指定もメニュー3でおこなえばよいが、現在の「回答順の無視」プログラムにはバグが含まれており正確な動作が保証されない(近日中に対策を講じる予定である)。

3. 8 EGLファイルの出力とCOMMAND.COMの起動

FIG.1に示した語形編集作業が終了したファイルは、その状態を保存するために一度外部に出力する。EGLが出力する編集済みのファイルを「EGL形式のファイル」もしくは「EGLファイル」と呼び、以後言語地図の作成にはEGLファイルを利用する。EGLファイルはデータファイルと同じくシーケンシャルファイルである。

EGLファイルとデータファイルを比較すると、以下の点で違いがある。①EGLファイルの先頭第一行にはそのファイルに含まれる併用パターン単位の異なり語の総数Nが記入されている。②2行目からN+1行までにはN個の異なり語の語形と頻度が記入される。ただし、語形と頻度の間には後にグラフィック記号を示す記号を記入することになるので、デリミターが一個加えられている。③N+2行以降には地点番号とその地点の併用パターンが記入されている。同一の地点についてデータファイルとEGLファイルの相違を概念的に示すと以下のようなものである(DはAに合併されたとする)。EGLファイルでは併用パターンが一個のデータとして記録されている点が重要である。

データファイル		EGLファイル			備考
地点	語形 1	語形 2	語形 3		
01,	A,	B,	C,		AよりBが古 ↓
02,	A,	D,	B,		↓
03,	B,	C,	D,		Bは昔使った ↓

EGLファイル (N+2 行目以降)

地点	併用パターン
01,	A & B & C ↓
02,	A & B ↓
03,	B & C & A ↓

EGLファイルの出力をおこなうためには、まず併用パターンの一覧表が表示されている状態かメニュー3の状態ではFを押す。すると画面が変化し、各種出力用のメニュー4が示されるので、「3: パタン毎の処理結果 ==> ファイル」を選択する。システムは出力先ファイルのデフォルト名称(データファイルの拡張子をEGLに代えたもの)を示し、確認を求めてくる。もちろん名称は自由に変更可能である。ファイルの出力が終われば画面はメニュー4に復帰する。以上で一応の語形編集作業は終了する。続いて他のファイルを編集したければESCキーを2回押してメニュー1に戻り、「6」を選んでスタートメニューに戻る。

ところでメニュー1にはこれまで説明していなかった機能「7: COMMAND.COMの起動」があるので、ここで説明しておく。これはチャイルドプロセスと呼ばれるMS-DOSの機能を利用して、EGLのシステムの中からプログラムを中止せずにDOSが提供する各種のコマンドを利用するためのもので、MS-DOSのコマンドインタプリタであるCOMMAND.COMのコマンド(内部コマンド)すべてが無条件で使用できる。例えば先ほど出力したEGLファイルが本当に出力されているかどうかを確認したければ、DIRコマンドで確認できるし、TYPEを用いればファイルの内容をCRTやプリンターに出力することも可能である。COMMAND.COMに含まれない外部コマンド --- FORMAT や CHKDSK --- の場合はMS-DOSのファイルであるFORMAT.COMやCHKDSK.COMをシステムディスクに入れておけばよいのだが、メインメモリーが384Kバイト以下ではメモリー不足のために利用できない⁽¹⁵⁾。MS-DOSからEGLへ復帰するためには、EXITと入力すればよい。

3.9 ふたつのEGLファイルの合併処理と現在の編集状態の保存

EGLファイルは地図作成に用いる他に、複数項目を合併した言語地図を作成するためにも使用する。この機能を利用するためにはスタートメニューで「2: 二つのEGL形式ファイルの合併」を選択し、システムの指示に従って二つのファイル名を指定する。EGLは第一のファイルの併用パターンを「語形1」、第二のファイルの併用パターンを「語形2」とみなして「併用パターンの検索」を実行する。その際、「無視語形の指定」「回答順の無視」も指定可能である。

この機能は意味的に関連のある2項目の複合地図を作成したり、あるいは多くの地図に共通する音韻項目を一枚の地図にまとめる場合に有効である。FIG.10ではLANDの調査項目で合拗音に関する10項目を一枚にまとめている。このように3枚以上の地図を合併するためには「二つのEGL形式ファイルの合併」を繰り返し使用すればよい。ついでながら、ちょっとしたテクニックを用いればEGLファイルを単独で読み込み、併用パターンの合併から処理を始めることも可能である⁽¹⁶⁾。

また、先に3.8でEGLファイルの出力をおこなうために利用したメニュー4で「4: 現在編集中的数据ファイル ==> ファイル」を選択することにより、合併処理などを施したデータファイルをディスクに書き出すことができる。これは編集の途中で一度EGLを終了しなければならない場合に利用するために考えた機能であるが、オリジナルのデータファイル中で語形があまりに細分化されすぎていて、合併処理にむやみに時間がかかるような場合には、ある程度まで合併を進めたファイルを保存しておくことによって次回からの分析の手間をはぶくことも可能である。

4. E G Lによる言語地図の作成

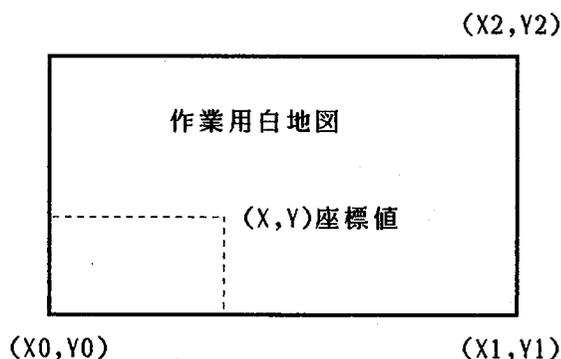
4. 1 地形情報の扱い方の特徴

以下では編集済みのE G LファイルからC R Tに言語地図を描くまでの方法を説明する。E G Lでは通常の言語地図に盛り込まれる地理的な情報を2種類に分けて取り扱う。ひとつは海岸線・河川・等高線・交通路など普通の地形図が表現している情報であり、これらは「地形ファイル」に書き込まれる。もうひとつはデータファイルおよびE G Lファイルの地点番号に対応する調査地点の平面座標の情報であり、これは「地点ファイル」に書き込まれる。そして地形ファイルと地点ファイルとは完全に独立しているので、相互に自由な組み合わせが可能である。従来の言語地理学でも、海岸線と交通路だけの白地図と等高線を描いた白地図の2種類を用意して必要に応じて使い分けるといった工夫がおこなわれていたが、E G Lでは複数の地形ファイルを用意することによって、これと同じ効果を得ることができる。しかも ---言うまでもなく--- 語形の情報もまたE G Lファイルという形で独立しているので、異なった表現効果の言語地図を簡単に何種類も作成して見比べることができる。これは手作業による地図作成では絶対に体験することのできなかつた世界である⁽¹⁷⁾。

4. 2 地形ファイル・地点ファイルの作成

地形ファイル・地点ファイルはともにB A S I Cのシーケンシャルファイルであり、基本的には共通した性格のものである。まず地形ファイルを例にとって説明しよう。地形ファイルを作成するためには、調査地域の白地図に該当する作業用白地図を前もって作成しておく必要がある。ただし、この地図は完成した（すぐにスタンプを押せるような）ものである必要はない。適当な大きさの地形図の上に調査地点が示されている四角形のものであれば何でもよい（極端な場合、フリーハンドで描いた地図であってもよい）。そしてこの地図上のすべての点に対し何らかの方法で一意的に二次元の直交座標上の値を決定できるようにしておく。例えば、すべての調査地点を含むような長方形を描いてその左下を座標原点(0,0)とし、あとすべての地点を座標原点からのX, Y座標で表せばよい。

ファイルの先頭第一行には作業用地図の作図範囲の左下隅の座標(X0,Y0)を入力する。第二行には作図範囲の右下隅の座標(X1,Y1)を、第三行には作図範囲の右上隅の座標(X2,Y2)を入力する。これらの値をもとにコンピュータは以後入力される座標値を正規化してファイルに記憶するので同じ地図を何回入力してもその度に座標がずれる心配はほとんどない（例えば一回目はセンチで二回目はインチで座標を入力してもよい）。



四行目以降には地形を表現するための具体的な座標値の情報が書き込まれるが、E G Lではすべての地形情報を「ひと筆書き」の要領で表現している⁽¹⁸⁾。例えば海岸線を表す複雑な曲線はC R T上では細かく指定された座標の間を短い直線で結んで表

現する。この場合、地形ファイルではそこからひと筆書きを開始する「書き始めの点」とそこまでひと筆書きを中止する「書き終わりの点」を指定しておく必要がある。このふたつの点以外はすべて前後の点と直線で結ばれるわけである。地形ファイルの四行目以降ではひとつのレコードが「番号」「X座標値」「Y座標値」「フラグ」の4つの数値から構成され、デリミタのカンマで区切られている。「番号」は一回の「ひと筆書き」に含まれるすべての点に対して1から順に与えられる自然数の通し番号であり必ず1ずつ増加する。「X座標」「Y座標」は既に述べた方法によって決定される地図上の点の座標である。「フラグ」は現在のEGLシステムでは利用していない値なので、デリミタだけ打っておけばよい。

EGLシステムは「番号」が1のレコードを読むと、これを「書き始めの点」とする。そしてこれ以後では「番号」が増加していく限り、新しく読んだレコードの座標と直前のレコードの座標との間を直線で結んでゆく。もしあるレコードで「番号」が1に復帰していた場合は直前の座標が「書き終わりの点」になり、現在のレコードを新たに「書き始めの点」とする。したがって「番号」が1のレコードが続く場合にはそれぞれの座標に点だけを表示してゆく。LAND用に作成した地形ファイル(LAND.MAP)の冒頭部分を参考までに示すと以下のようなものである。ここではフラグの値に1を入力している。

レポート	情報	ファイルの内容
1	(X0,Y0)	---> 141,741 ↓
2	(X1,Y1)	---> 3738,742 ↓
3	(X2,Y2)	---> 3758,2364 ↓
4	地点 1	---> 1, 11, 1165, 1 ↓
5	地点 2	---> 2, 20, 1165, 1 ↓
6	地点 3	---> 3, 31, 1173, 1 ↓
7	地点 4	---> 4, 40, 1182, 1 ↓
8	地点 5	---> 5, 43, 1193, 1 ↓

S ラインスタイル

次に地点ファイルの作成であるが、最初の3行に座標正規化のための情報を入力する点は地点ファイルを同一である。4行目からは個々の調査地点のX, Y座標を入力すればよいのだが、各レコードの先頭にはEGLファイルと対応する地点番号が必要である。「番号」や「フラグ」は必要ない。LAND用に作成した地点ファイル(LAND.TTN)の冒頭を参考までに示す。

レポート	情報	ファイルの内容
1	(X0,Y0)	---> 142, 740 ↓
2	(X1,Y1)	---> 3738,742 ↓
3	(X2,Y2)	---> 3757,2362 ↓
4	地点 1	---> 64024309, 424, 1576 ↓
5	地点 2	---> 64024338, 416, 1542 ↓
6	地点 3	---> 64024475, 504, 1514 ↓
7	地点 4	---> 64024491, 456, 1489 ↓
8	地点 5	---> 64025359, 419, 1414 ↓

地形ファイル、地点ファイルともに座標の数字が大きな値になっているが、これはA3版の白地図を読み取り精度0.1mmのデジタイザーで入力しているためである。EGLにはロジテック社のデジタイザーMYPAD-A3(MODEL K-510 mk2)を使用して簡単に地形・地点ファイルを作成できるユーティリティープログラム(DZIN.BAS)が付属しているが、本稿ではその使用法を説明するスペースがない。

4.3 EGLファイルの読み込みとグラフィック記号の割当

EGLで言語地図を描くためにはまずEGLファイルに記録されている何種類かの語形（併用パターン）にどのような記号（ハンコ）を割り当てるかを決めなければならない。そのためにはスタートメニューで「4：言語地図の作成」を選び、画面にメニュー5が表示されたら（FIG.6 参照）「1：ファイル読み込みとグラフィック記号の割当」を選ぶ。ここで地図を描きたい項目のEGLファイル名を指定すると画面には併用パターンの合併処理時に似た併用パターンの一覧表があらわれる（FIG.7 参照）。以後、この画面上でシステムと対話しながら語形とハンコの関係を決めていくのである。

- 1：ファイル読み込みとグラフィック記号の割当
 - 2：CRTに言語地図を作成
 - 3：グラフィック記号の調整
 - 4：COMMAND.COMの起動
 - 5：語形編集に戻る
 - 6：凡例のプリンタ出力

FIG.6 メニュー5（本文参照）

記号割当 A:Q50.EGL F6:画面↓ F7:画面↑ B:系列選択

NO	語形	頻度	NO	語形	頻度
1	NR	N (28)	2	SONOTA	■ (4)
3	ギョート	● (16)	4	ギョート & ジョート	⊙ (1)
5	ジョート	○ (27)	6	ジョートー	□ (16)
7	ジョートーシ	◇ (14)	8	ジョートーシ & ゴーリ	⊕ (1)
9	ゾーキ	∠ (9)	10	ゾート	∩ (14)
11	ゾーリ	● (3)	12	チョロ	△ (9)
13	チョロ & ジョート	▲ (3)	14	ツインツインサ	▽ (2)
15	トカ	■ (7)	16	トカ	(194)
17	ヤモリ	↑ (16)	18		(0)
19		(0)	20		(0)
21		(0)	22		(0)
23		(0)	24		(0)
25		(0)	26		(0)
27		(0)	28		(0)
29		(0)	30		(0)

使用する記号は? >>

1 □	2 □	3 ■	4 ■	5 ■	6 ⊙	7 □	8 □	9 ■	10 ■
11 ■	12 ■	13	14	15	16	17	18	19	20

FIG.7 ハンコ割り当て処理中の画面（FIG.4と同じ「トカゲ」の項目である。既にすべての語形にハンコが割り当てられており、画面下部には矩形系列のハンコの一覧が表示されている。13～20にはまだハンコが登録されていない。）

EGLで使用可能なハンコは、①円、②矩形、③線、④三角、⑤小記号、⑥他、の6系列に分類されており、現在約70種類のグラフィック記号がハンコとして登録されている。最初に「どの系列のハンコを使いますか?」とのメッセージが画面下部に出るので、使いたい系列を指定すると、その系列のハンコが画面下部の枠中に一覧表となって示される。使用したいハンコの番号を入力し、続いてそのハンコを割り当てたい語形の番号(語形の横に示されている)を指定すると語形一覧表中にハンコが表示され、割り当てが終了したことが分かるので、次の割り当てに進む。既に割り当ての済んだ語形に別のハンコを指定すると、画面中のハンコは自動的に書き換えられる。ハンコや語形の番号の代わりにBを入力するとひとつ前の画面に戻るので、「どの系列のハンコ...」のメッセージまで戻ればハンコの系列を変更することができる。ひとつの画面には30までしか語形が表示されない。これ以上の語形を含むファイルの場合は、F6、F7で語形一覧表をスクロールさせながら割り当てを続ける⁽¹⁹⁾。

ハンコの割当てが終了したら、ファイルにハンコの指定を記録するために、一度EGLファイルを出力しておかねばならない。「どの系列のハンコ...」の画面に戻り、Fを入力すればよく、ファイル名は通常変更する必要がない⁽²⁰⁾。これで言語地図を描くためのすべての準備が終了した。

4.4 CRTに言語地図を描く

言語地図を描くための指定は簡単である。メニュー5に戻って「2: CRTに言語地図を作成」を選び、地形ファイル・地点ファイル・EGLファイルをそれぞれ指定するとEGLはCRTに言語地図を作成する。

ここでシステムがおこなっている処理を概観しておく、最初に地形ファイルを読み込まれ、記録されている各種地形情報がひと筆書きされる。システムは地形ファイルの先頭3行に記録された座標値をもとに作業用地図の座標値をCRT上の表示エリア内におさまるように線形変換するので、指定したエリアが作業用地図と相似でない場合には、作成された地図にもゆがみが生じる。現在、表示エリアの指定はプログラム中で指定しているが、これはメニューでおこなう方式に変更する予定である。

続いて地点ファイルとEGLファイルを読み込まれ、両者の地点番号が照合される。矛盾がなければ、やはり線形変換によって調査地点の座標値が定まる。何か矛盾が見つかったら(EGLファイルにしかない地点番号が見つかった場合など)システムは警告を発生し、矛盾が生じたデータの位置を知らせてくる。この警告は無視してもよく、その場合システムは処理を進めるが、出力される地図の正確さは保証の限りでない。

最後にシステムはEGLファイルの先頭に書き込まれている異なり語の情報をもとに、各地点にどのハンコを表示するかを決定する。これが終了するとCRT上の調査地点の座標に各種のグラフィック記号(ハンコ)を表示する。

これで一枚の言語地図がひとまず完成したわけであるが、ハンコの割り当てが気に入らなければF9を押してメニュー5に戻り、割り当てを調整すればよい。気に入った地図ができあがったら、地図が表示されている状態でCを押すことによって画面のハードコピーをとることができる(もちろんプリンターが接続されていなければならない)。これだけでは地図上のハンコと語形の関係が示されないの、語形の凡例がほしい場合にそなえてメニュー5では「6: 凡例のプリンタ出力」を用意しているが、このルーチンは現在まだ稼働していない。その代わりにハンコ割り当ての画面でもハードコピーをとれるようにしてあるのでこれを利用することができる(ただし、30以上の語形がある場合はいささか見にくい凡例になってしまう)。

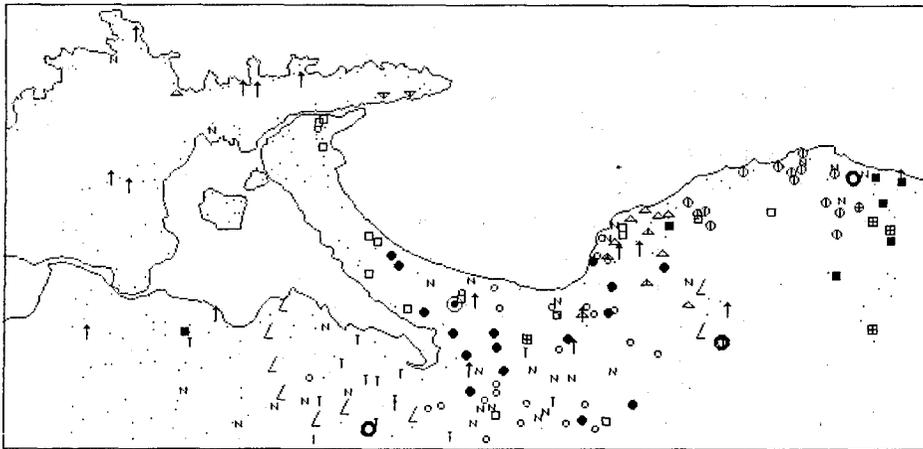
5. EGLによる地図作成の実例(LANDの編集作業から)

ここでは現在編集作業を進めているLANDから、EGLを使用して描いた言語地図を紹介する。スペースの関係で3枚にとどめるが、できるだけEGLの機能がよく発揮されている地図を選び、処理に要した時間も報告する。

5. 1 「トカゲ」

FIG.8 は先に異なり語やプレビュー機能の紹介に用いた「トカゲ」の言語地図の完成図である。松江を中心とする調査地域の西部には標準語形「トカゲ」だけが分布しているのに対し、調査地域の中央部では米子が中心となって多くの方言形（「ゾーキ」・「ゾート」・「ジョート」・「キョート」など）が生み出されている。一方、大山町海岸部から東の地域もまた独自の変化を示しており、「チョロ」・「ジョートシ」・「トカ」などが固有の分布領域を確保している。

FIG.8 上部の言語地図、下部の凡例ともにCRT画面のハードコピーであり、凡例にはメニュー5「グラフィック記号の調整」の画面を利用している。この項目の場合事前に授業でとりあげたことがあり分布の概要が既にわかっていたので、語形の編集を始めてから地図の出力を得るまで、2回の描き直しを含めて30分程度であった。



記号割当 A:050.EGL

F6:画面↓ F7:画面↑ F9:メニュー F:出力 C:ジョー

NO	語形	頻度	NO	語形	頻度
1	NR	(28)	2	SONOTA	(4)
3	キョート	(16)	4	キョート & ジョート	(1)
5	ジョート	(27)	6	ジョート	(16)
7	ジョートシ	(14)	8	ジョートシ & ゾーリ	(1)
9	ゾーキ	(9)	10	ゾート	(14)
11	ゾーリ	(3)	12	チョロ	(9)
13	チョロ & ジョート	(3)	14	ツインツイン	(2)
15	トカ	(7)	16	トカ	(194)
17	マロ	(16)	18		(0)
19		(0)	20		(0)
21		(0)	22		(0)
23		(0)	24		(0)
25		(0)	26		(0)
27		(0)	28		(0)
29		(0)	30		(0)

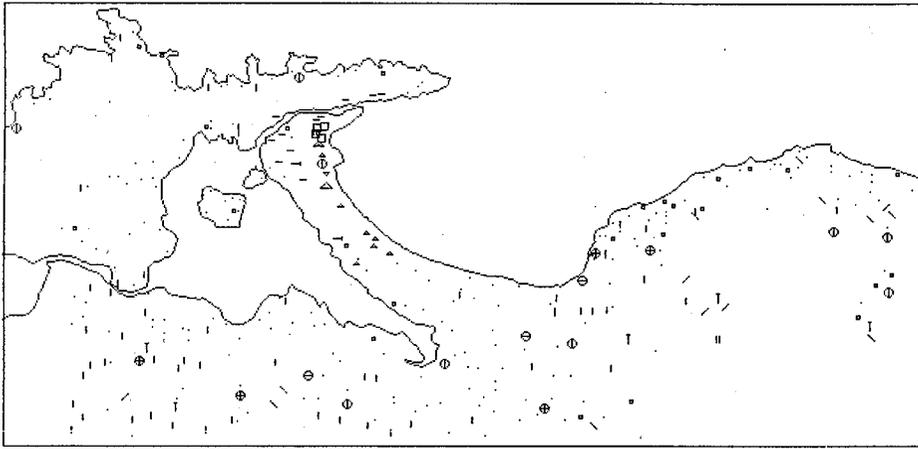
FIG.8 「トカゲ」の言語地図

(FIG.7 に示したEGLファイルを地図化したもの。上半分はCRT上に描かれた地図を、また下半分は記号割当画面をそれぞれ同じ紙にハードコピーしたものである。FIG.9, FIG.10 も同様に作成している。)

5. 2 「オタマジャクシとメダカ」

LANDでは弓浜半島特に境港を中心に「オタマジャクシ」に多くの方言形が認められることが調査の初期から明らかになっていたが（前川,1986）、その原因の一部が「メダカ」をあらわす語形との混同にあることを示すために、両項目を合併して示したのが FIG.9 である。「オタマジャクシ」に関しては /オタマジャクシ/・/キヤコ/、「メダカ」に関しては /メダカ/・/ネハ/・/ハイ/ が調査地域の全域に分布する主要な語形なのだが、弓浜半島の美保湾側（外浜）では「オタマジャクシ」を表す語形が /ネハ/ となっているので「メダカ」を区別するために /カワコ/ という新語形が生じている。また、両者を区別せずに /メダカ/ もしくは /ネハ/ と呼ぶ地点が中海をとりまくように分布していることも注目に値するであろう。

FIG.9 を作成するにあたっては「オタマジャクシ」と「メダカ」のファイルを事前に少数の語形に整理しておき、それを「二つの E G L 形式ファイルの合併」機能で読み込んで分析した。ファイルの整理から合併処理を経て地図を描くまで（2回の描きなおしを含めて）1時間を少し下回った。



記号割当 a:q49_65.egl

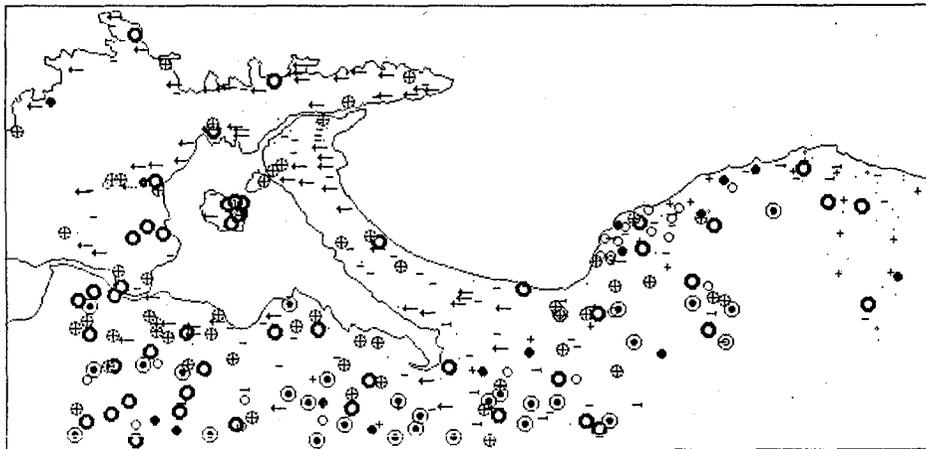
F6:画面↓ F7:画面↑ F9:メニュー F:出力 C:終了

NO	語形	頻度	NO	語形	頻度
1	sonota	h (29)	2	オタマジャクシ & カワコ	^ (2)
3	オタマジャクシ & ネハ	· (185)	4	オタマジャクシ & ハイ	、 (9)
5	オタマジャクシ & メダカ	· (26)	6	キヤコ & ネハ	! (58)
7	キヤコ & ハイ	、 (5)	8	キヤコ & メダカ	! (5)
9	タツホッホ & カワコ	△ (1)	10	タツホッホ & ネハ	- (11)
11	タツホッホ & メダカ	- (2)	12	トテケ & ネハ	□ (4)
13	ネハ & カワコ	^ (9)	14	ネハ & ネハ	⊕ (9)
15	メダカ & カワコ	- (1)	16	メダカ & ネハ	⊕ (5)
17	メダカ & メダカ	⊕ (3)	18		(0)
19		(0)	20		(0)
21		(0)	22		(0)
23		(0)	24		(0)
25		(0)	26		(0)
27		(0)	28		(0)
29		(0)	30		(0)

FIG.9 「オタマジャクシ」と「メダカ」の言語地図
 (& の左側が「オタマジャクシ」、右側が「メダカ」を表す語形である。)

5.3 「合拗音の残存率」

島根県出雲地方を中心とするいわゆる出雲方言には中世日本語の姿を反映する合拗音（「火事」「西瓜」「外国」などを / kwazi / / suikwa / / gwaikoku / と発音する現象）が認められる。LANDの調査地域は出雲方言の東端にあたるので、調査票には合拗音に関する10項目を含めておいた（ただし、1984,85の両年度は3項目のみ）。現在、日本全国で合拗音は急速に標準的な / ka / / ga / に変化しているが、その標準語化の進展は単語によって大幅に異なっていることが知られている（前川,1988）。FIG.10では合拗音に関する10項目を合併して、そのうち何項目に合拗音が残存しているかを分析した。残存率が20%以下の地点は米子・境港の市街地を除けば地図の東端、鳥取県東伯郡に集中しており、従来報告されてきた出雲方言の東端にほぼ一致している。また弓浜半島・島根半島の全域で残存率が50%前後であり、交通の中心である山陰本線・国道9号線沿線よりもかえって標準語化が進んでいるように見えることもFIG.10の特徴である。ただし、この地域は1984,85年度の調査地域にあたるので調査項目が少なく、その影響もあるであろう。FIG.10を作成するには合拗音項目のファイルに含まれる語形を1=合拗音が記録されている、0=記録されていない、N=調査票の変更・調査ミスなどにより記録がない、の3種に分類して整理し、合併している。EGLのシステムディスクに含まれるADSUBというプログラムは、このように語形が数量で表現されたファイルの合併処理を補助するプログラムである。ADSUBとEGLを利用した編集作業では、10個のファイルの編集作業から4回の地図書き直しを含めて、FIG.10の完成まで2時間弱であった。



記号割当 B:KW%.EGL

F6:画面↓ F7:画面↑ F9:メニュー F:出力 C:コピー

NO	語形	頻度	NO	語形	頻度
1	0%	(35)	2	10%	(4)
3	100%	(50)	4	20%	(21)
5	30%	(53)	6	40%	(16)
7	50%	(63)	8	60%	(22)
9	70%	(16)	10	80%	(55)
11	90%	(29)	12		(0)
13		(0)	14		(0)
15		(0)	16		(0)
17		(0)	18		(0)
19		(0)	20		(0)
21		(0)	22		(0)
23		(0)	24		(0)
25		(0)	26		(0)
27		(0)	28		(0)
29		(0)	30		(0)

FIG.10 「合拗音の残存率」の言語地図

6 おわりに

6.1 今後の改良点

EGLは1987年の6月からプログラミングを開始し、現在までに約8カ月を経過している。語形の編集から地図を描くまで一応の機能が動くようになったのが1987年の11月末であり、その後バグをとったり機能を拡張したりしながら現在にいたっている。現在のプログラムでもLANDの分析・編集に一定の効果をあげているが、まだ未完成の部分、改良を要する機能が多い。現時点では次に示すような機能の追加・改良を考えている。

①XYプロッタによる作図機能 EGLの本来の狙いは満足のいく地図ができるまでCRT上で試行錯誤を重ね、最終的にはXYプロッタで高品質の地図を清書できるようなシステムである。プロッタへの出力プログラムは現在一部が稼働しているが、そこでは既に地形などが印刷された白地図にあわせて各種の記号(ハンコ)と凡例だけを記入していく方式を採用している。また、プロッタのコマンド体系としてはできるだけHP-GLとの互換性を確保するよう努力している。

②凡例のプリンタ出力機能 現在BASICのコピー命令によっている凡例の出力をプリンタによる印字に変える。そのために各種のグラフィック記号を「外字」としてプリンタに登録する方式を考えているが、その登録方法がプリンタ毎に違い、苦慮している。また例えばNECのPC-PR201の場合最高85文字までしか外字登録ができないので、必要に応じてユーザーが登録する記号を選択するオプションが必要だろう。①とも関係するが、現在普及し始めているレーザービームプリンタをプロッタ・プリンタとして利用できればこの問題の解決策となるだろう。

③地図の拡大・縮小機能 現在の主流である640×400ドットのCRTでは解像度に限界があり、地点が接近している場合にはハンコが重なってしまうことがある(LANDでは境港市と名和町の海岸部で地点が接近している)。このような場合に地図の一部を自由に拡大表示できる機能が必要である。BASICではこの種の機能がWINDOWコマンドとして用意されているので比較的容易に実現できるはずである。

④データファイルのエディット機能 データファイルの作成・編集のための簡易エディタを作成する。

⑤語形編集機能の拡張 EGLの最大の特徴である対話型語形編集機能を次の点でさらに強化する。ア)地点番号を指定して特定のデータを直接編集できるようにする、イ)文字列を指定してそれを含む語形を検索・表示できるようにする、ウ)現在全く利用していない「備考」を編集に利用できるようにする、等。

⑥異なり語単位の地図作成機能 現在のEGLシステムでは地点1(A, B)、地点2(A, C)の場合に両地点が語形Aを共有している事実は編集に利用することができない。両者は別個の併用パターンだからである。従来の言語地理学ではこのような場合に、同一地点にふたつの記号(そのうちひとつが共通のもの)を与える処理がなされており、GLAPSもそれにならっている。ところでEGLではそもそもシステムの開発がLANDを念頭においておこなわれたので、注(14)に述べた調査上の特色や③に述べた地点の接近などを考えて現在の方法を採用した。

調査地点が接近した地図の上で、次に使うハンコの大きさを考慮しつつ適当な隙間を見つけてハンコを押す作業は人間にはさほどむずかしくないが、これを一般的におこなうプログラムを書こうとすると非常な困難が予想される。しかし、調査地点が充

分に散在している地図を処理することもあるし、また従来の慣習を尊重することも必要である。そこで③の拡大表示機能と組み合わせた形で、異なり語を単位として一地点に複数のハンコを押す機能を追加することを考えている（GLAPSはこの方式を採用している）。

この他にも注に述べたものを含め多くの改良が必要であろう。またEGLは現在約5000行のBASICプログラムであるが、これ位の規模になってくるとプログラム中にバグが含まれているのはまず確実である。今後、演習などで多くのユーザーに利用してもらいながら問題点を洗いだしていこうと思う。

6. 2 調査現場での活用

最後に近年普及が著しいラップトップ型のパーソナルコンピュータでEGLを使用するメリットについて述べておきたい。ラップトップ型パソコンとは本体、キーボード、CRTに相当する表示装置を一体化させた軽量の持ち運び可能なパソコンである。最新型のラップトップ機は記憶容量や処理速度の面で通常のデスクトップ型パソコンと同等かそれを上回る性能を実現しているし、グラフィックRAMを備えるようになってきているので、EGLを改訂することなく稼働させることができる。EGLは既に収集された資料をまとめて入力して分析する目的にも無論使えるが、ラップトップ型パソコンの進歩により、より効果的な利用法として、調査中に調査現場で利用することが考えられるようになった。EGLでは語形をコーディングする必要がないので、ラップトップ機を調査地に持ち込めば、一日の調査が終了したところで資料をすべてそのまま入力することができる。調査地点の座標も簡単に入力できるから、レビュー機能を利用すれば収集したばかりの資料の分布を確認することができるはずである。筆者の研究室では予算不足からこのような方法を実際に試す機会をまだ得ていないが、実現すれば言語調査のあり方が質的に変化する可能性があるのではないかと思う。従来から各種の臨地調査用マニュアルでは現地でその日のうちにその日の資料を整理することの重要性が例外なく強調されているのだから。

注

(1) 調査の詳細に関しては前川(1986)、同(1988)を参照していただきたい。現在までにこの調査の資料を用いて5編の卒業論文が鳥取大学教育学部に提出されており(美坂,1986;足立,1987;伊奈垣,1988;徳橋,1988;深田,1988)、このうち1988年に提出された3論文では分析の道具としてEGLが広範囲に使用されている。なお、伊奈垣真吾、徳橋毅、深田高史の三氏はEGLの最初のユーザーとして多くのバグを発見し、プログラムの改善について非常に有益なコメントを与えてくれた。注(16)で述べたEGLの機能は開発時に筆者が想定していなかった用法を彼らが試行錯誤の過程で発見したものであることを付言して深謝したい。

(2) 徳川・山本(1967,p.27)による。LANDの場合には『日本言語地図』に比較すれば調査地点数が少ないためにこの数字がそのままあてはまるわけではない。それでも従来の手作業の経験では一枚の地図の作成に数日を要することがめずらしくなかった。このような能率で作業をおこなうと、一旦一枚の地図が完成すると何か問題点が残っている可能性があっても、つい「まあ、いいだろう」となってしまう。コンピュータによる作図能率改善がもたらす大きなメリットのひとつは再分析の意欲の確保である。

(3) 萩野(1975)を読むと、GLAPSも当初はプロッタ出力をおこなっていたことがわかるが、その後のバージョンアップの過程で出力をラインプリンタに限定したら

しい。また今のバージョンに移行した後に執筆された荻野(1977,1981b)では今後プロット出力をサポートする予定があると述べているが、現在はまだ実現していないようである。

(4) EGLでは調査項目単位にファイルを作成するので、GLAPSのように複数の調査項目を自由に組み合わせた編集をおこなうことができない。この欠点を補うために後述する「二つのEGL形式ファイルの合併」機能を設定しているが、それでもGLAPSに比べて複数項目のデータ編集の機能が劣っていることは事実である。

ここで何故調査項目単位にファイルを作成するかを説明しておく。EGLは対話型のシステムなので、ユーザーをイライラさせないために処理の高速性が要求される。そのために一旦ファイルを読み込んだ後はすべての処理をメインメモリの中でおこなっている。現在の16ビットパーソナルコンピュータの主流は8086系のCPUを採用しており、そのメインメモリは最大640Kバイトであるので、この値からDOSとBASICとEGL本体に必要な容量を差し引くと、極端に大きなファイルを同時に読み込むことができないのである。

(5) ここで「文字」というのは1バイト系の文字を考えている。ワープロの用語ではいわゆる「半角」の文字であり、漢字、平仮名などの「全角」文字は2バイトを使用するので、「2文字」になる。

(6) 拡張子でファイルを選別することはしていないので、データファイルの拡張子にEGLやTTNを指定しても正常に読み込まれるし、その後の処理も正常におこなわれる。ただし、ファイル管理を混乱させないためにこれらの拡張子は使用するべきでないし、HSTを使用すると後に述べるヒストリーファイルを作成した際にデータファイルを消去してしまう危険があるので絶対に使用してはならない。

(7) EGLのシステムディスクにはMS-DOSのシステム以外に通常以下のファイルを入れておく。

- 1) EGL.BAS 2) CGL.BAS 3) N88BASIC.EXE 4) EGL.BAT
- 5) EGL.EXE 6) CGL.EXE 7) N88BASIC.LIB
- 8) DZIN.BAS 9) ADSUB.BAS 10) ADSUB.EXE 11) ADSUB.BAT

ただしインタープリター版だけで使用するならば5)～7)は不要であり、またコンパイラ版だけならば1)～4)が不要である。8)～9)は地図の入力やEGLファイルの加工に使用するユーティリティープログラムであり、EGLの稼働には必要でない。この他に、「地形ファイル」や「地点ファイル」もシステムディスクに入れておいた方が便利である。

(8) デフォルト値とは「ユーザーが何も指定しない場合にシステムが勝手に決定する値」である。例えば処理結果を新しくファイルに書き出すとき、新ファイル名だけを指定するとそのファイルはデフォルトのディスクドライブに書き出される。

(9) ファイル名の誤りなどなんらかのエラーが発生するとEGLは画面にエラーメッセージを表示して中止状態に入り、ユーザーに現在の処理を続行するか中断するかを選択を要求してくる。ディスクドライブの蓋を閉め忘れた場合などはエラーの原因を取り除いて続行すればよいが、通常は中断を選択した方がよい。その場合システムはメニュー1に戻るようになっている。

(10) GLAPSの場合プログラムという形で語形編集の記録が残されるが、EGLでは対話型処理のため、どのような処理をおこなったかはユーザーの記憶に頼るしかない。この欠点を補うためにヒストリーファイルを導入した。

(11) 既に合併対象となった(白の)語形を代表形に指定しようとしたり、再度合併しようとした場合にもシステムは警告をおこなう。

(12) 併用パターンの一覧表では併用された二つ以上の語形はその間を & のマークで区切られて表示される。

(13) ただし、併用パターンに関しては回答順の区別が存在しないので色による区別はおこなわれない。

(14) LANDの調査では標準語の分布を知ることも調査の目的のひとつであったので(深田,1988 参照)、明らかに標準語形が方言形と併用されている場合にもそのまま両方ともデータファイルに記入している。最初から標準語形をはぶいておけば事態は多少改善されようが、われわれの経験では一見明らかかな標準語と思えたものでも実際には限定された地理的分布を示したり、同じ方言形の分布領域の中に標準語を併用する地域とそうでない地域とが分かれていたりすることがあった。

(15) 正確に説明するとメインメモリーが不足するために十分なチャイルド領域を確保できず、そのために外部コマンドを利用できないのである。チャイルド領域の確保はEGLのプログラムがBASICのCHILD命令を用いておこなう処理なので、現在のEGLではMS-DOSの外部コマンドを使用するためにはメインメモリーを確保すると同時にEGLプログラムの一部を変更しなければならない。EGLの起動時にメインメモリーの容量を質問し、自動的に対応するようなシステムの方が望ましいと思っている。

(16) そのためには、データファイルのすべての併用パターンをひとつに合併してしまい、語形を他のファイルに絶対に含まれていないものを書き換える。これをEGLファイルとして出力すると、たとえば全地点が\$のファイルができる。このファイルと編集したいEGLファイルとを合併し、「無視語形」に\$を指定すれば、編集したいファイルの併用パターンだけが編集画面に表示される。

(17) EGLが動きだしてから、卒業研究や演習で学生が提出する言語地図が美しくない時に遠慮なく「やりなおし」と言うことができるようになった。手作業の時代に、LANDの白地図上の364地点を資料リストをながめながら数十種類のゴム印で正確に埋めてゆく作業を純粋に審美的な観点から命ずるのはなかなか勇気のいることであった。

(18) 現在のEGLでは線の種類を選ぶことができず、実線だけを使用する。従って破線を作りたいときには、細かく「ひと筆書き」を繰り返さなければならない。将来は各種の線(実線、破線、鎖線など)をシステムの方で用意したい。

(19) 記号割当画面は異なり語合併画面とはちがって1画面全体がスクロールする。現在、3画面目まで用意しているので最高90までの併用パターンにハンコを割り当てるのが可能である。この設定も本文で述べた最大地点数と同じく経験的に決めた値であるが、必要ならばプログラムを少し書き換えるだけで画面数を増やすことができる。異なり語・併用パターンなどと共に、将来はこうした設定もユーザがメニューで変更できるようにした方がよいだろうと思う。

(20) ここでEGLファイルを再度出力するのはファイルの先頭にどの併用パターンとどのハンコが対応するかの情報を書き加えるためである(3.8の②参照)。

参考文献

I 引用したもの

- (1) 荻野 綱男 (1975) コンピュータによる方言地図の作成と解析 『日本方言研究会第20回研究発表会原稿集』.
- (2) 荻野 綱男 (1977) 一般的言語地図作成システムGLAPS 『計量国語学』 11巻8号.
- (3) 荻野 綱男 (1981a) 『方言調査用パッケージプログラムGLAPSの使い方』 自家版 (科研費報告書).
- (4) 荻野 綱男 (1981b) コンピュータが描く言語地図 『月刊 言語』 10巻10号.
- (5) 沢木 幹栄 (1984) コンピュータを使った方言研究 『日本語学』 3巻1号.
- (6) 徳川 宗賢・山本 武 (1967) 電子計算機で言語地図をつくる試み 『計量国語学』 40.
- (7) 前川 喜久雄 (1986) 中海周辺における言語革新と境港市 『鳥取大学教育学部研究報告 (人文・社会科学)』 37巻1号.
- (8) 前川 喜久雄 (1988) 合拗音消滅の言語的要因 --- 山陰での言語調査から -- 『Sophia Linguistica』 No.22 (印刷中).

II LANDに関する卒業論文 (すべて鳥取大学教育学部国語学研究室蔵)

- (1) 足立 達夫 (1987) 『鳥取県中・西部境界地域の方言地理学的研究』
- (2) 伊奈垣 真吾 (1988) 『方言調査の調査規模と語彙数の関係 --- 中海・大山周辺地域方言地図集と日本言語地図および中国地方五県言語地図との比較 ---』
- (3) 徳橋 毅 (1988) 『方言地理学に於ける誘導形の研究』
- (4) 深田 高史 (1988) 『中海・大山周辺地域方言地図集における標準語形の方言地理学的研究』
- (5) 美坂 朗生 (1986) 『美保湾・中海周辺地域の方言地理学的研究 -- 弓浜半島を中心に ---』

補注

パーソナルコンピュータで稼働する言語地図作成システムとしては、EGLより早く1983年にPC-8801で稼働するSEALSが福嶋秩子氏によって発表されている。本稿でも当然これを紹介すべきであったのだが、マニュアルが入手できず、またGLAPSのように豊富な解説論文があるわけでもないので残念ながら一切言及することができなかった。EGLを正式に公開するまでにはぜひ参考にしたいと思う。