

みんなくりポジトリ

国立民族学博物館学術情報リポジトリ National Museum of Ethnology

Computer Use in Ethnological Studies

メタデータ	言語: jpn 出版者: 公開日: 2010-02-16 キーワード (Ja): キーワード (En): 作成者: 杉田, 繁治 メールアドレス: 所属:
URL	https://doi.org/10.15021/00004559

民族学研究における情報処理

—コンピュータ民族学について—

杉 田 繁 治*

Computer Use in Ethnological Studies

Shigeharu SUGITA

This short note discusses “Computer Ethnology”, one of the research sections in National Museum of Ethnology. The principal focus of “Computer Ethnology” is the construction of models and the simulation of complex phenomena. In addition, information processing, such as picture, speech and music analysis, and language processing, are useful and important aspects of the use of computers in ethnological studies.

- | | |
|----------------|------------------|
| 1. はじめに | 5. コンピュータ処理の特徴 |
| 2. 民族学研究の流れ図 | 6. コンピュータ民族学について |
| 3. 民族学研究における情報 | 7. あとがき |
| 4. 情報表現の変換 | |

1. はじめに

X線や放射性炭素の利用によって、考古学の議論がより客観的になり、さらに新しい方法論も取り入れて“新科学”への方向に進展しつつあるという。新しい文明の道具が発明されることにより、それまでは考えも及ばなかったことや、不可能に近いとしてあきらめていた事柄が、なんでもないことに変化してしまうという例はいくらかもある。現代ではコンピュータがそのような道具の一つであることに異論はなからう。

* 国立民族学博物館第5研究部

数多くの異なった分野で利用され、それを使わなかったもとの状態に戻ることが不可能になっている分野もある。また一見コンピュータとは無縁と思える分野で使用されると、意外性のために人々の興味をひくが、実情は極く単純な一般の事項に使用されているにすぎない場合も多い。

国立民族学博物館ではコンピュータを導入し、民族学研究の新しい展開を追求しようとしている。そして第5研究部に「コンピュータ民族学」という部門が設置されている。一体これはどのようなことを研究する部門なのか、その性格はまだはっきりしていないようである。研究を進める上での補助道具として、図書、文献、標本管理データなどの検索や統計処理、文書処理にコンピュータを利用することは確かに有効ではあるが、それは民族学固有の問題ではなく、すでに他の多くの分野で実用になっている。いわば現在では常識になっているコンピュータ利用の一つである。それでは研究者がかかえている民族学の香りが強い個々のデータを加工・変形させて、有意味な命題を引き出すための処理法を研究するのであろうか。コンピュータ民族学部門の研究者は他の研究者がコンピュータを利用しようとする時、その使い方、処理法について相談し、実際に手助けすること、つまり情報処理技術者として研究者とコンピュータとの仲介人（インターフェイス）としての役割を果たすのであろうか。

今までわが国の民族学研究にはコンピュータがほとんど活用されていなかったが、この分野に導入して道具として使いこなすためには、コンピュータ情報処理の専門家が必要である。しかしもし上記のような内容のみであるならば、わざわざコンピュータ民族学部門を研究部に置く必要はない。民博の組織でいえば情報管理施設のコンピュータ部門および共同利用研究所の利点を生かして他機関の情報処理研究者の協力で実現出来るはずである。研究部の中にあつて常日頃民族学研究者と接し、その研究内容やデータの性質をよく理解している情報処理研究者がいることは、コンピュータ利用が定着するまでの過渡期としては有効ではあろうが、もしそのような受身的な立場であるならば、やがてコンピュータ利用技術の発達によって不必要になってしまうに違いない。そのような臨時的な研究部門ともおもえない。

ではコンピュータ民族学とは何か。他の分野、たとえば経済人類学、宗教人類学や言語人類学などは経済、宗教、言語を対象にして、それらが各民族においてどのような特徴を有するか、その比較並びに相互の関係などを追求するものとして一応明確になっている。しかしコンピュータ民族学という場合コンピュータ自身が研究対象ではなからう。用具論の一部としてやがて扱われることはあってもことさらコンピュータを民族学での対象として取り上げる意味はない。

この研究ノートでは民族学研究における情報とその処理について考察し、コンピュータがどのような場面で活用出来るかを明確にすることによって、コンピュータ民族学とは何かを追求しようとするものである。

2. 民族学研究の流れ図

民族学研究の過程をデータ処理の観点からみれば図1のように表現出来るであろう。

まず実線で示した部分を考える。フィールド(F)を出発点として、その調査・収集から得られた資料(D)に基づき、人間生活に関する何らかの記述なり命題(R)を導出する。得られた命題が資料と矛盾しないかどうか、さらには実際のフィールドでの事実と合っているか否かを確かめるフィードバック過程も必要である。この研究の流れにおいて二種類の道具が使われている。図1でT1, T2と書かれた部分がそれである。T1はフィールドから資料を収集するためのものである。写真機、撮影機、計測器、録音器、望遠鏡、顕微鏡、などの他にX線分析、C₁₄分析、化学分析、カラー分析などの単能的機器によるもの、さらに標本写真や航空写真から計測するためのコンピュータによるパターン認識なども資料収集のための道具と考えることが出来る。T2は収集された資料を蓄積、加工、検索、比較することにより、法則を引き出すための

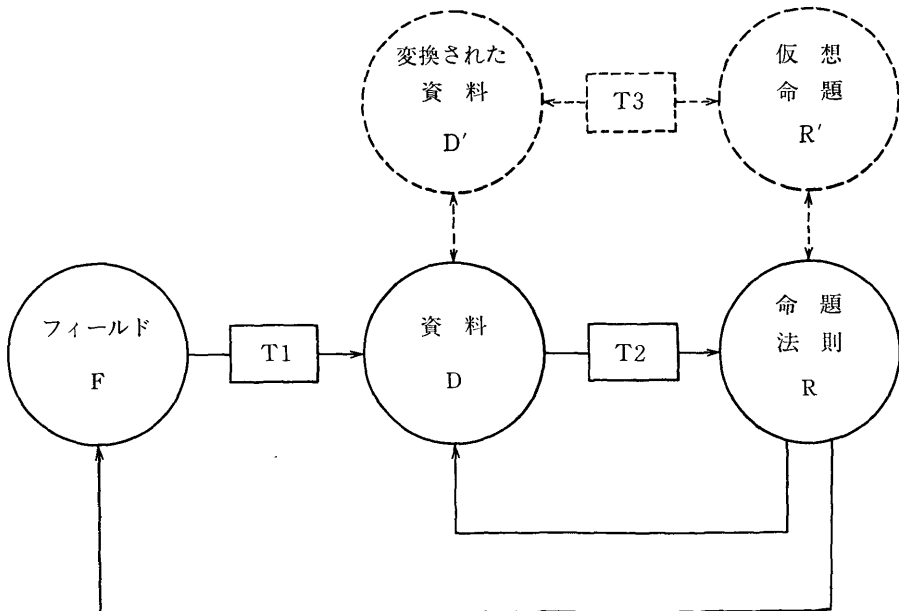


図1 民族学研究の流れ図

助けとなる道具である。コンピュータが大いに活用される部分である。

次に図1に於て点線で示された部分を考える。これは D から R へ直接的に行く道を考えるのではなく、一旦 D から D' (抽象化された資料, 仮想資料) へ移り, D' から R' (仮想命題) を導き出して, R' から R を得るという過程である。この場合 $D' \rightarrow R'$ はすでに手法が確立されていたり, その論理が明確になっている部分で, 例えば数学や論理学の世界のように自動的に推論が運ばれる。 $D \rightarrow R$ の過程は必ずしも単純でなくても, $D' \rightarrow R'$ は非常に直接的, 機械的である場合がある。しかし $D \rightarrow D', R' \rightarrow R$ への変換がなかなか困難であるが, 一旦うまい変換が見つければ $D' \rightarrow R'$ は簡単になるということである。発想の転換と呼ばれているものはこのように回り道を発見することに相当する。言語を例にとれば, 発話の集合 (D) からその文法 (R) を導く過程が実線で示した部分である。それに対し, 人工的な発話の集合 (D') を仮定し, その文法 (R') を求める, あるいは文法を仮定しそれによってどのような発話が生じうるかを調べるのが点線で示した部分である。数理言語学と呼ばれている分野に相当する。別の表現をすれば D', R' はモデルの世界であって, 現実 (F, D, R) のシミュレーションを行なっている部分と考えることが出来る。工学にしる理学にしる現象がよく分らない場合はモデルを想定し, 適当なパラメータを設定し, 現実の入力と出力の関係を満たすようにモデルを変えていくことが行なわれる。民族学では人間という複雑な集合を含む世界を扱っているために, モデルを構成するといってもそう簡単にはいかないが, 現実のデータをうまく説明する一つの理論を作るというだけではなく, あらゆる可能な事象を数え上げ, その一つ一つについて明確な論理的説明がつけられるようにするということが考えられる。このようにすべての可能性が与えられ, 現実にはその中の一つが生起したとすれば, それがどのような事情によるものかの説明が明確になる。

D から R への直接の道と, 一旦 D' へ移り R' を経て R に至る道程への大きな差は, $D \rightarrow R$ には人間の種々雑多な知識や偏見が含まれている可能性があるのに対し, $D' \rightarrow R'$ の過程は明確な論理 (アルゴリズム) の世界になっていることである。そこでは出来るだけ予見を排除する。仮説を立てそれを検証するというやり方に対し, むしろ仮説を生成するための手法といえる。 R' は一つの命題というよりは命題の集合, もしくは可能性の集合である。例えばあるデータ集団に対して項目 A と項目 B に注目し, その関係を統計を使って定量的に正確に求めることは $D \rightarrow R$ 的である。それに対しすべての項目間の関係をあらいざらい求めるというのが $D' \rightarrow R'$ 的である。

$D' \rightarrow R'$ の過程に入っている道具 $T3$ としてはコンピュータが中心的な役割をはた

すであろうと思われる。従来の民族学研究が主として図1の実線で示された過程のみであったのに対し、点線で付け加えられた D'→R' の過程が民族学研究に新しく導入されるべきもの、すなわちコンピュータ民族学に相当するものではないかと考えている。図1で3つの道具 T1, T2, T3 が使用されており、いずれにもコンピュータはかわりを持っているが、それぞれの役割、違いは何であるか。それを明らかにすることによってコンピュータ民族学の性格が浮かび上がってくる。

3. 民族学研究における情報

民族学研究はその対象が多様である。その情報形態から分類すると図2のように表現出来る。大きく分けると音や画像などのアナログ情報と文字や数値などのデジタル情報になる。アナログ情報は連続的な性質を持ち、そのままではいわゆるデジタルコンピュータで取り扱えないので、元の情報を出来るだけ失わないようにしてア

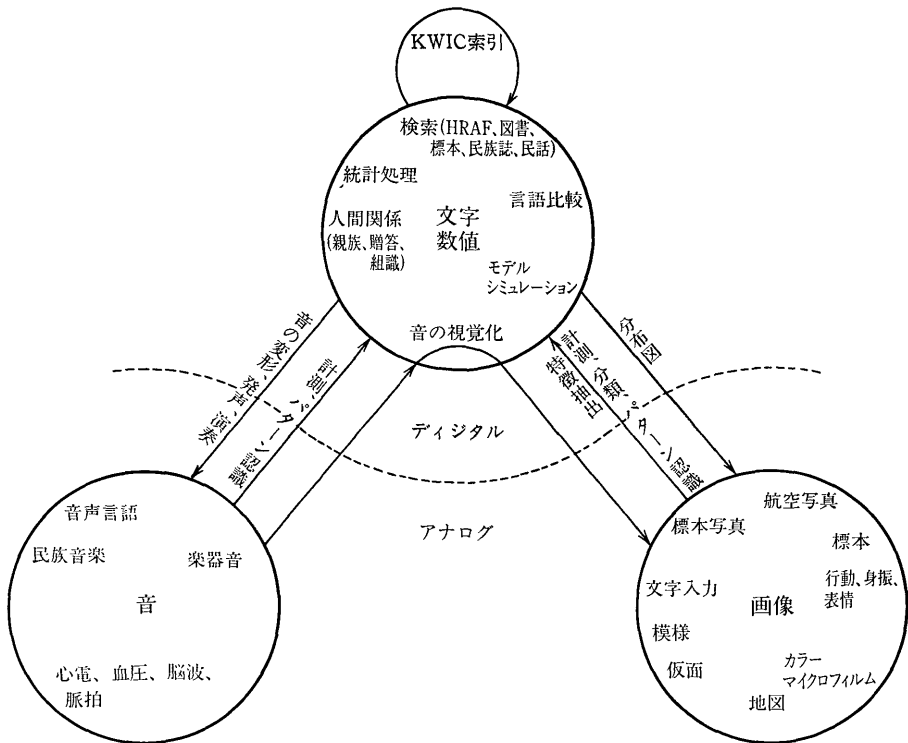


図2 民族学研究における情報

ナログ→デジタル変換を行なう必要がある。その逆にデジタル情報を人間の視覚や聴覚にうつたえるためにはデジタル→アナログ変換を行なう。

民族学研究で扱う一次元アナログ情報としては、音声言語、音楽、楽器音、計測波形データなどがある。音声言語ではトーンを客観的に示すためにピッチパターンを図で表現したり、音素の構造をさぐるためのコンピュータとの対話的手法、音から文字への変換などがある。また音楽では各種民族音楽の特徴を音響的に抽出し、五線譜ばかりでなく直観的な表示方法を考える必要がある。さらに楽器音の分析などを通じて民族による音の認識の比較研究が行なわれる。また動作分析あるいは反応分析の一部として、血圧、脈拍、脳波、心電図、などセンサーを被験者の身体につけて生活空間における行動の計測や、心理実験などを通じてその特徴抽出を行なうことも興味深い。その他化学分析装置からの出力も扱うことがある。

次に画像データと呼ぶものには二次元、三次元、それにカラーをともなうものがある。二次元情報としては、標本写真、航空写真、線図形、地図、映画フィルムなどがある。標本写真から土器ややじり、道具類の大きさや形状を計測したり、仮面、織物の模様、紋などの分析、航空写真から土地利用や作物の量などを測量すること、地図の重ね合せ、地図と他の情報との結合など多くの対象がある。三次元情報としては標本そのものの立体計測、それといろいろな角度からの三次元表示。踊り、身振り、行動、表情などの分析も画像処理の一部とみることが出来る。これについては定量的な研究がまだあまりなされていないようでコンピュータを活用して行なえば面白い結果が得られるであろう。

デジタル情報は文字、記号、数値がある。言語の系統を求めてその比較研究や祖語からの変化の規則を求めたり、民話の分析、民族誌、古文書などの分析から生活習慣の比較。人間関係（親族関係、贈答関係、勢力関係など）の記号的データの解析、生業・儀礼の特徴分析と比較など。

コンピュータでこれらの情報を扱うことの一つの大きな意味は媒体間の相互変換が可能になるということである。文章で示されたものを図として表示したり、図や写真として表現されているものを記号化するか、音声、音楽の特徴要素を図示したり、記号化するというような変換が容易に出来る。すべてのものが忠実にデジタル化出来るとはいえないが、デジタル化したものを媒介として他の形式に変換出来る装置が工夫されて来ており、直接2媒体間の変換を行なうよりも柔軟なことが出来る。

情報表現の変換を行なうことの意味は何かといえば、その情報をよりよく理解出来るようにするためである。その一例は「一覽性」という特徴を有効に利用することで

ある。目の前に全体像が一度に広げられると全体と部分の関係がよく見え、部分的にしか見えなかった場合に比べて理解が進む。この一覽性を更によくしたものが「図形化」である。統計処理の結果を数値で表現するのと、グラフで表現するのを考えてみれば明らかであろう。地図の上で分布を示すのも、単に地名と関連事項が列挙してあるのに比べれば格段の違いがある。

また計測という作業を考えてみよう。対象としてやじりや土器のサイズ、形状、とがりの角度などを現物又は写真から人間が計測する場合を想定してみる。操作は単純であって時間さえかければ誰でも出来ると思われるが、正確さ、再現性という点に疑問がある。人が異なれば結果が異なるし、また同じ人でも二回やれば両者が一致しないということがしばしば起る。人間は疲労したり、見誤りが必ず起きると考える必要がある。量が増えればなおさらである。このような場合コンピュータを使って写真から自動的に計測することが出来、その後の特徴抽出、分類、なども人手を介さずに正確に行なうことが出来る。物によっては微妙なところの計測が必要であって、それは単純な機械システムでは計測出来ないかも知れない。したがってすべて機械まかせ、すべて人間まかせではなく人間↔機械の結合システムをうまくする工夫を考えるべきである。

このような計測、情報媒体の変換又は情報表現の変換は図1でいえば T1, T2 を通る過程に相当している。データを処理したい目的がはっきりしており、それを正確、高速に行なう為の道具として利用している場合である。

4. 情報表現の変換

分 布 図

情報表現の変換例として文章で記述されたものから分布図を構成するものを示そう。これは民話とか産物、方言、言語、氏名、地名などいろいろな対象が考えられる。例えば韓国についての記録『東国輿地勝覧』に記載されているものの中から産物を取り上げ、その分布図や項目間の関係を知りたいとしよう。約340地域、産物の種類は約350種類ある。この記述から地図の上に産物の分布図を描くことは、手で書いてもそれ程手間のかかるものではない。一枚の地図の上にすべての産物を同時に表示したとすれば探すのがむずかしくなる。そこで産物毎に地図を作るということが考えられる。350枚の地図を描き、産物の指定があればそれを選び出せばよいのであるからかなり実際的である。しかしもし2つの産物についてそれが同時にとれる地域を捜したいと

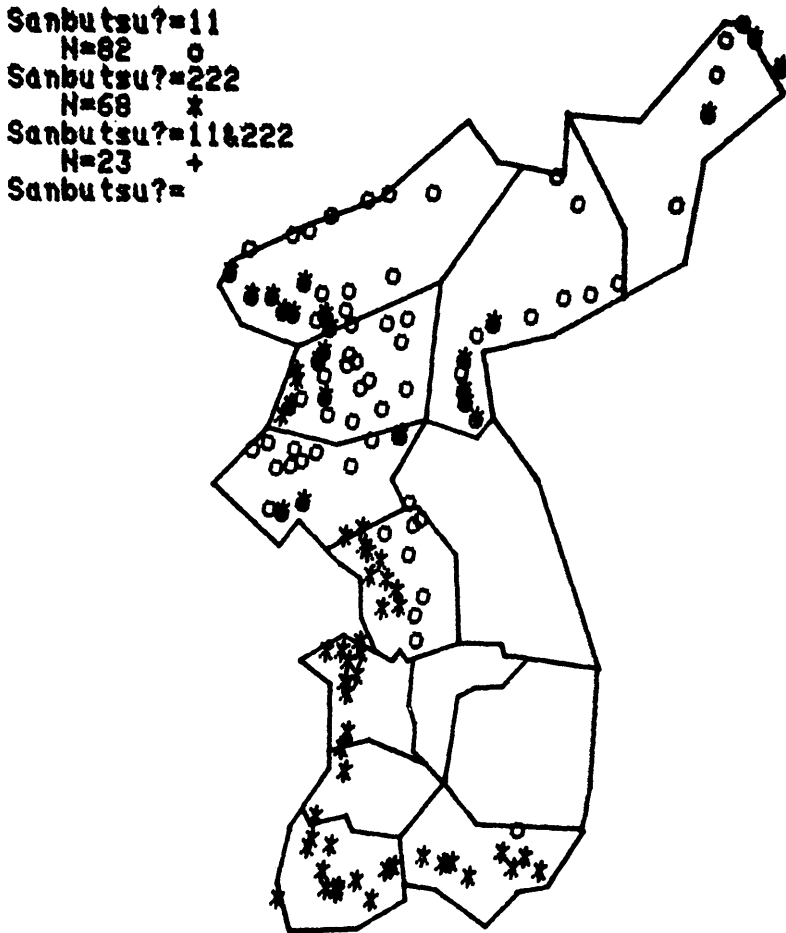


図3 産物の分布図の例

すればどうなるか。あるいは3つ、4つということになればもはや予めそのような地図を作成しておくことは不可能に近い。このような問題に対して、産物とそれがとれる地域名をすべてコンピュータに入れておき、別に地図情報もコンピュータに与えておけば、必要に応じて2つの組合せ、3つの組合せ、AND, OR の関係など任意の組合せに対して直ちに分布図を描くことが出来る。図3にその一例が示されている。質問としては産物名(のコード)を与えると、それらがとれる地域が1~2秒で地図の上に表示されコピーがとれる。『東国輿地勝覧』の場合、データ量は少ない方であるが、これが多くなってもコンピュータにとってはほとんど変わらない。データを入力する場合の手間がかかるだけである。さらに土地に関する他のデータ、標高、気温、

人口、川、山などを入れておき、それらの条件も組み合わせで表示させることも簡単である。この種の表示の活用は単に目的とする産物の分布を知るだけではなく、あらゆる産物の分布、組み合わせの分布を出力することによって今まで気がつかなかった発見をする可能性があるということである。地図上である地域にのみ集中しているものとか、他のものと分布が似ているとかが直観的に分かる。そしてある顕著な傾向があった時それが何に基づくものであるかを考えてゆくという方法がとれる。単に文字情報としてコンピュータに入っているデータを検索して取り出すだけではコンピュータの力を十二分に活用しているとはいえない。

KWIC

次に図形化ではないが、一覽性を持つデータ変形の例として KWIC (Key Word in Context) と呼ばれているものを示そう。これは民族誌なり民話なりのテキスト全文をコンピュータに入れておき、その中に含まれるあらゆる単語について、その前後関係、どの頁の何行目に出ているかの情報をつけて表示するものである(図4)。しかも Key となる語は辞書順に並べられているので捜そうとする語の部分は直ちに見つけることが出来る。これは図形化ではないが同じような語が同一箇所並んで出てくるので一種の図形と考えることが出来る。この index の特徴はそのテキストのすべての単語を含んでいる点である。普通図書の索引は著者や編者が重要と思うものだけが抽出され並べられている。しかしこれでは読む側に必要な情報が直に分かるとは限らない。テキスト全文を頭に入れるほど知っていたとしても見落すこともある。しかし KWIC を作れば冠詞一つまでも見落すことはない。もちろん目的によっては不要な語もあろう。もしリストの形で印字しておく場合には全く関係のない語の部分は省いておけばよい。しかしコンピュータの中ではすべてが対象になるようにしておき、必要な時に検索してその部分のみ抽出することが出来る。又逆にある単語がそのテキストに出現しているか否かがはっきりする。つまりある時代の民族誌には A という動物の記述があるか否か、B という習慣があったかどうかを知ろうとしても、普通の index では確実ではない。もしそれらが index にあれば明確であるが、ないとするとそれは本文中には出てくるかも知れないし又全く出ていないかも知れない。KWIC だとその在否がはっきりする。必ずしも KWIC の形をとらなくてもよい。とにかく対象となるテキスト全文を入力することによっていろいろな角度からの検索が可能になる。データ入力の内容が分からなくてもタイプさえ打てればよく、又適当な端末装置があればタイ文字、ハングル文字、漢字混り文などの文字も入力出来る。言語によっては単語単位への分割をしておく必要があるものもある。

a ngeralamo, 33. Toma waktu enase	manucia	dofu i firi, tataxi ana Jo	243/01
ma woru putera Achmad horu-isa lom	manucia	dofu firi-firi, so i kodih	243/20
ge fokolofino silom fohohe, sababu	manucia	dofu nga oti malo so rupe-	224/37
ano talunsa Gofornemen beto, mara	manucia	enase nga oho se tuan Gofu	251/32
kosolaha tuan residen ni fals, ge	manucia	enase nga oho toma tuan no	252/14
Jouu Allah Ta'ala tulung si hohira	manucia	fasik se munafi(k) nga ak	240/22
k se munafi(k) nga akal so sasala	manucia	firi-firi enase si sado ko	240/24
tomdii, dokasaa tersebut pasa ge,	manucia	sasa habari dahu-ie, waje,	227/09
awa i Jaha pasama ruwa. Toma enase	manucia	sanapu kolofino, mai habar	227/12
lano kalari oti ge, niscaya sasala	manucia	sanapu kolofino se bingu,	228/27
idin se Jou bobato sudo hoko tahan	manucia	se i ise idin enase riuwa.	226/12
toma soanyie-lamo, niscaya sasala	manucia	Islam se Nasarani se cina	233/09
u bobato sudo si dodemoroma sasala	manucia	kasaa pane-pane tomə oti;	236/27
esiden su-dole se Jouu Kolano lahi	manucia	ka una i oti-ma-awo, karan	239/03
u idin, sudo una ie Jai-Jai, tahan	manucia	ma nyiha-nyiha, si ma ha i	227/05
ma nyiha beto. Toma si futu enase	manucia	ma bati toma Santosa si ta	240/02
a se dero fala hutu si sioko se ma	manucia	moju, mara toma Santosa si	240/04
se ma nsofa se momote sakira-kira	manucia	nganyasiromdidi foloi, dok	227/16
labuan. Gilasila Jouu Kolano haka	manucia	nganyasiromodidi nsofa Tar	239/08
tuan Gufornemen Walanda ma lolahi;	manucia	nganyasiraense, silom, seb	251/24
ki ratumoi se ka ma ija-uwa; yaadi	manucia	nganyasiraense dokasaa ter	251/28
onga fala-ngasu-ngasu ma monara se	manucia	nganyasi rao-rao solaha, s	252/06
so se iskungir ma mantros moi-moi,	manucia	nganyasiromotoha foloi. Jo	262/09

図4 KWIC の例

HRAF

コンピュータでテキストの必要な箇所を検索するシステムと HRAF (Human Relations Area Files) のペーパーファイルシステムとを比べてみよう。HRAF システムの特徴は民族学に関する事柄を約800の項目に分類し、3桁のコードを定める。そしてテキストの各ページの文、節毎にどのような内容がそこに書かれているかを専門家が分析し、欄外にそのコードを記入しておく。そして先ず地域 (OWC コード) で大分類し、一つのキャビネット引出しをそれに当て、その地域内では項目毎に File カードを作りそれに該当するページのコピーを入れておくというやり方である。例えばマイクロネシア Yap 島の通過儀礼を調べようとすれば OWC コード OR22 のキャビネット引出しを開き、分析コード881 (思春期とイニシエーション)、561 (年令に

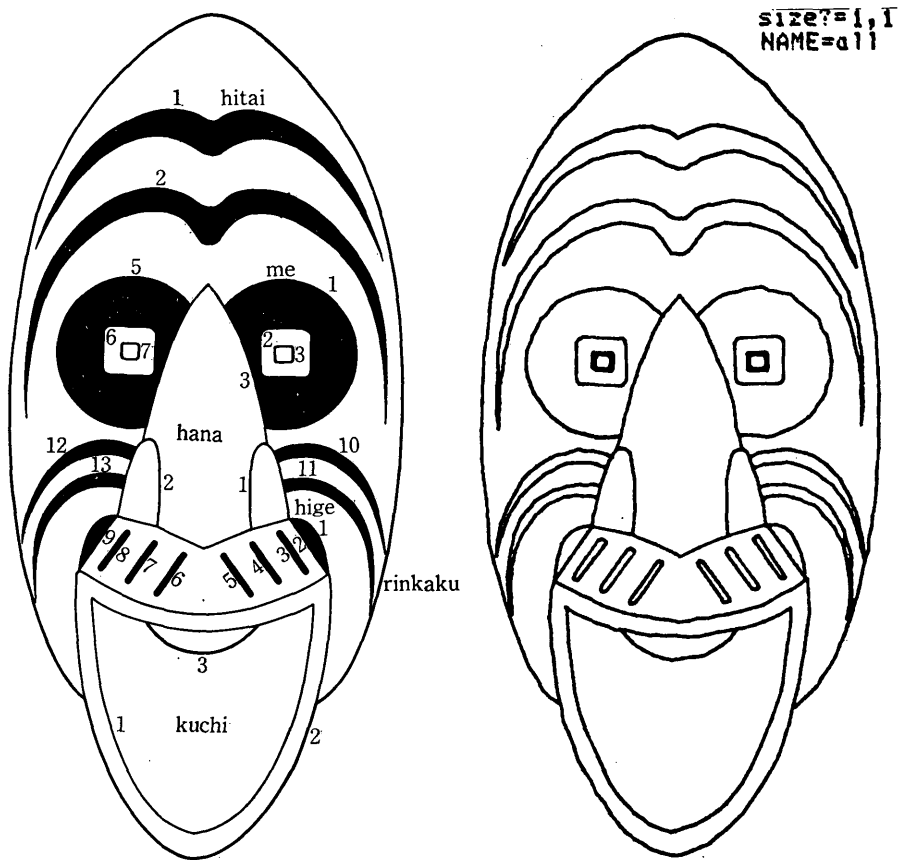


図5 仮面の変形

(a) 田主誠氏が描いたカナダ・イロコイ族の仮面

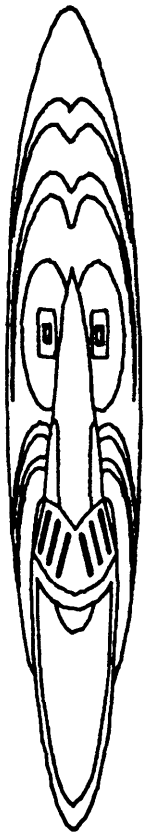
(b) コンピュータに入れたもの

よる成層), 585 (婚姻の儀礼) などの項目の File を見ればそこに 関連事項が記載されたページが入っている。

このシステムでは地域を指定してその中でいろいろな項目について調べるのは便利であるが, カテゴリーを指定して各地域での比較をする場合にはすべてのキャビネットを開けていかねばならない。又システム側では各ページ毎に高度な専門家によるめんみつな分析が必要である。利用者は自然単語では検索が出来ず, OCM コード表を参照して適当なカテゴリーを選択して検索しなければならない。現在の HRAF システムは原テキスト約50万ページ, 重複したページを含めると300万ページにもなり毎年10万ページ位増加しつつある。キャビネットの数にして110本必要で広い空間がいる。

もしこのシステムをコンピュータ化したとすればどうなるか。

size?=1,2
NAME=all



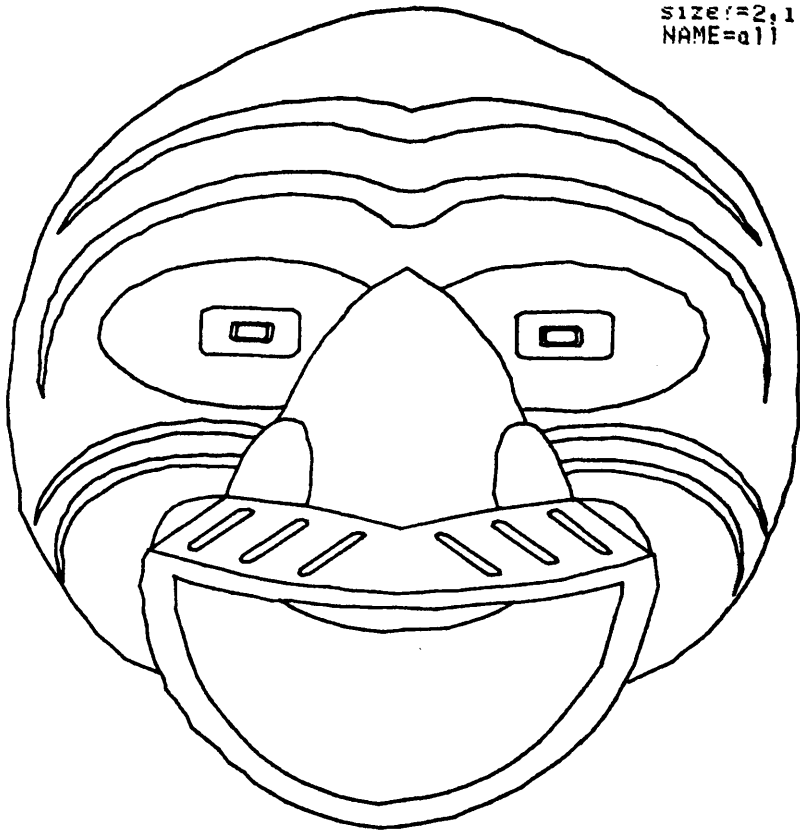
(c) 縦横の比を
2:1にし
た場合

先に述べたテキストの KWIC を作るのと同じ要領で, 対象テキスト全文をコンピュータに入力する。分析も何も必要でなく, 自然語で質問をすればその単語が含まれるページがすべて出てくるようにすることが出来る。もう少し柔軟にするならばシソーラス (類語集) を作っておき, 一つの単語を入力すればそれとよく似た概念の他の単語をも含むページが出てくるようにすることが出来る。このシステムではいくつかの単語の組合せでも検索が出来, 地域コードからでも検索出来る。

コンピュータを使った HRAF システムは非常に便利ではあるがペーパーシステムに比べ現在ではコストが高つく。例えば1ページに2000文字含まれるとすれば50万ページでは1000 MB (1MBは 10^6 キャラクターに相当する) 必要である。高速検索のための工夫をすれば少なくともその2~3倍の容量を必要とする。しかしテキスト分析の費用と空間の問題を考えればやがてコンピュータシステム化の方向にいかざるを得ないと思われる。その中間の方式としてマイクロフィッシュと選択された自然語での index を用いる方法もあるが, コンピュータを活用することの意義としてすべてをつくすという特徴を生かさねばならない。

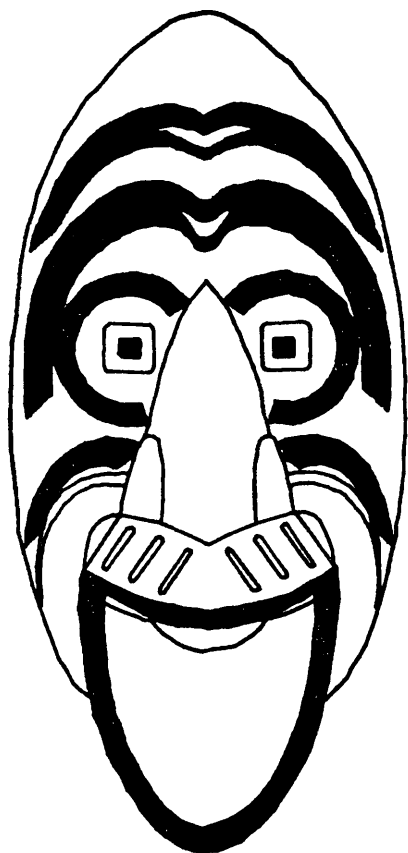
仮面の変形

次に図形から図形への変換例を示そう。民族仮面がある。これらの特徴を抽出し, 民族毎, 或は儀礼の種類による仮面の分類を試みるとしよう。仮面の写真からコンピュータによって顔の輪郭,



(d) 縦横の比を1:2にした場合

目、鼻、口、ヒゲなどの要素を正確に取り出すことは面白いパターン認識の問題ではあるが、まだ現在のところむつかしいのでその部分は人間が抽出することによって必要なデータが得られたとしよう。例えば図5の(a)は田主氏によってデザイン風にかかれたカナダインディアン、イロコイ族の仮面である。これを図形入力装置のタブレットを用いて、目や口、鼻、ヒゲなどを人間がその部分の線を追跡することによって入力することが出来る。そして各々の部分に名前をつけておく。つまり仮面の要素をばらばらにしてコンピュータに入れておく。この要素の位置関係、大きさなどから仮面の特徴分類をすることも出来るが、人間の視覚にうったえる方法も有効である。例えば図5の(b)はコンピュータに再現させたものの図、図5の(c)はもとの図を縦横の比を2:1にした場合、(d)は1:2にした場合、(e)は一部分を黒くぬった場合である。要素の位置関係はいずれももとのままである。これを見ても、単純な変形



```
size?=1,1  
NAME=me1,  
NAME=hige10,  
NAME=hige12,  
NAME=kuchi  
NAME=hana  
NAME=hige  
NAME=me5,  
NAME=me  
NAME=hitai,  
NAME=rinkaku  
NAME=kuchi2,  
NAME=me4,  
NAME=me8,  
NAME=
```

(c) 一部分を黒くぬった場合

でもずいぶん印象が異なることが分かる。更に目や鼻、口の位置を相対的に移動させることや部分的な変形も行なえば、ある民族の仮面を他の民族の仮面に近似させていくことが出来るであろう。このような変形を通じて仮面の特徴を抽出していくことも出来る。その他模様なども変形という操作を通じて特徴があらわになることがある。

5. コンピュータ処理の特徴

コンピュータが誤って動作したために混乱に陥ったというような記事を見かけることがある。しかしこれは正しい表現ではない。たいていの場合は入力データの誤りか、人間が作ったプログラムの誤りである。ハードウェアの故障は別問題と考えれば、いずれも人為的なもので、コンピュータは与えられた情報に従い、与えられた指示通り

正しく動作したのである。つまりコンピュータは正確に間違う、ごみを入れたらごみが出てくるといわれるゆえんである。しかしこのことがコンピュータを使う上で一つの特徴になっている。人間の場合は融通性があり、データにうまく適応して判断を変えていくことが出来る。これは一見長所のようにあるが、別の見方をすればデータの意味にひっぱられて一定のアルゴリズムを守ることが出来にくいという欠点になる。コンピュータは予め与えられた指示以外のことはやらない。つまり途中で規則が不明確に変化してしまうということがなく終始一貫して同一の規準を保っているという特徴がある。

次にコンピュータの規模にもよるが大量のデータを蓄積・検索することが出来るという特徴がある。百件のデータを並べ換えるのも数十万件のデータを並べ換えるのも大した違いはない。検索にしても同じである。これはあたかも人間の持つ一覽性を拡大したような機能を持っていると考えることが出来る。

またいくつかのデータがあり、各々は一応独立しているが、共通項目が含まれているような場合コンピュータでそれらを結び合わせて、あたかも一つの大きなデータ集団があるかの如くに取扱うことが出来る。いわゆるデータベースといわれているものの本質はそこにある。

また処理速度に関しては、パターン認識やケースバイケースの処理を必要としたり、量が少ない場合は人間が行なった方が早い場合もあるが、計算や単純ではあるがくり返しくり返し行なう処理の場合にはコンピュータの高速性が発揮される。人間がやれば1時間かかるものを機械が1分で行なったとしてもそれ程利点はないが、1年かかるものが1時間で出来るというようなことはざらにあることである。しかしコンピュータ内部での処理は早くても、データの準備や入出力に時間がかかるという場合が多い。

コンピュータはプログラムを変えることによっていろいろな仕事をさせることが出来る。元の情報形態が何であれ一旦デジタル量に変換されれば取扱いに区別はない。コンピュータが広い分野に利用されるのもこの性質のためである。しかも共通言語を用いて開発されたプログラムは他のコンピュータでも実行出来る場合が多いのでプログラムの蓄積が行なわれる。しかしデータ構造に依存するようなプログラムはデータに合わせて修正しなければならず全くの汎用というものは少ない。

研究者が自分でプログラムが書けることが望ましいが、処理したい問題が充分細かく分析され、他人に正確に伝達することが出来るようになっていけば、プログラムそのものは自分で必ずしも書けなくてもよい。むしろ自分でプログラムを書くことの危

険性は、いろいろな条件を設定することが面倒になってプログラムを出来るだけ簡略化してしまう傾向が出ることである。プログラムを書くことに時間をかけるよりは問題を分析するのに時間を費す方が効果は大きい。コンピュータ関係者の協力を得てプログラムを開発すればよい。

コンピュータを使う時に注意せねばならぬことは、プログラムの文法的誤りがなければもっともらしい結果を出してくることである。よく結果を吟味しないととんでもない誤りをおかしている場合がある。データが多くなるといろいろなケースのチェックが出来にくいので誤りに気がつきにくい。自分でプログラムを書いていない場合には特に注意を要する。

6. コンピュータ民族学について

第4節で示した、分布図、KWIC、HRAF 検索、仮面の変形などの例は、単純ではあるが、手作業では膨大な時間がかかり又見落しなどがあるという点でコンピュータがあって始めて実現出来たことである。しかもそれは図書カードや標本管理データの検索や統計計算にコンピュータを使うというのに比べて民族学とのかかわりが強いような印象を与える。確かにそれらを使って民族学の研究が進むということは事実である。しかしそれらはたまたま民族学研究に関係するデータを扱っただけで、手法そのものは一般的なものである。情報工学の分野で開発されているデータ処理法：画像処理、音声・音響の処理、言語処理、蓄積・検索技法などを実際に民族学で発生するデータに適用すること即ちコンピュータによる民族学情報の処理がコンピュータ民族学なのであろうか。もしそうだとすればコンピュータ利用技術が進み、豊富なアプリケーションプログラムや特殊なハードウェア機器が出来、誰でもが簡単に情報処理装置を扱えるようになればコンピュータ民族学部門は発展的に解消されていくものとなる。そのような下請け的な機能ではなく、もっと積極的な意味で他の研究部門と共存していかなければならない分野としてのコンピュータ民族学があるはずである。しかも我が国の民族学研究においてはいままで欠けていたと思われる分野である。それは『コンピュータが研究を行なうようなやり方で民族学研究を行なう』ということである。これはある目的があって、それに関するデータをこうすればこうなるということが分かっているが、量が多く手間と時間がかかるのをコンピュータを道具として使って早く正確に実現するというコンピュータ情報処理とは少し異なる。つまりコンピュータ民族学という場合、コンピュータは実際的な処理道具というよりは、象徴的な意味

に使われていると解すべきである。第5節でも述べたようにコンピュータという言葉から我々が持つイメージは、杓子定規、融通がきかない、感情がない、正確、高速、大量データを扱う、複雑な計算もできる……などがある。もちろんコンピュータは人間が作ったプログラムによって動作しているわけであるからコンピュータが主体性を持っているわけではない。しかし一旦プログラムが作られると与えた情報のみに基づいて正確高速に処理を進めていく。そのアルゴリズムは明確である。この意味では先に述べた情報表現の変換例などもコンピュータ民族学の考えにそったものとみなすことも出来る。これらは特定の問題に対する処理というよりはコンピュータが作った民族学資料と考えられる。元の資料を加工しているだけで全体的な情報量は同じであるかもしれないが、人間が見るという状況では加工されたデータ（分布図、KWIC、など）ははるかに多くの情報を与える。

人間は柔軟で豊富な知識を持っているから問題解決に必要な情報、不必要な情報を意味的にすばやく選択する。しかしこれは裏を返せば個々のケースによって勝手な解釈をしがちであり、また予め自分の考えられる範囲でしか対象のデータを見ていないといえる。またどのような知識を導入したのか明確になっていない場合が多い。ところがコンピュータは与えられた情報以外は用いないし、プログラムの指示通り忠実に行なう。これは一面では欠点であるが、このような性質を積極的に民族学研究に取り入れようとするのがコンピュータ民族学であると考えられる。

この性質を満たす内容として最もふさわしいものは「モデル」によるシミュレーションである。図1に於て人間の思考による $D \rightarrow R$ の過程で扱う問題に対して、別のアプローチとして過程 $D' \rightarrow R'$ でも平行して同じ問題を扱う。そして R と R' を比較検討することによってその問題の内容をより詳細に理解することが出来る。また R' としてはあらゆる可能なケースが得られるからそこから新たなヒントが得られるということが期待出来る。つまり現実に意味があるかないかということを考慮することなくあらゆるパラメータの組合せによるシミュレーションを行なうことがコンピュータでは容易であるが、これによって思いもよらぬ意味がよみとれることがある。また逆に $D \rightarrow R$ の過程から新しい情報が得られて、モデルを変更していくなど相互に影響しながら進んでいく。 R と同じ R' を得るパラメータとアルゴリズムは $D' \rightarrow R'$ をたどれば明確になっている。

ここで一番問題となることは $D \leftrightarrow D'$ 、 $R' \leftrightarrow R$ の対応をどのようにつけるかということである。現実の D は必ずしも明確なデータでないかも知れない。また個体としては明確ではあっても $D' \rightarrow R'$ の世界で扱うために抽象化しなければならないがどの

ような情報を取り出すか。文化分析の理論というのはこの部分に当るものである。モデルというのは要素と関係で対象を記述するということである。そして同一のアルゴリズムによって大量のデータに適用し、分類なり関係なりを求めるのがコンピュータ民族学のやり方であろう。単にモデルを作るというだけではコンピュータ民族学にふさわしくない。象徴的な意味でコンピュータが使われているとしてもやはりコンピュータを使って定量的モデル、ダイナミックモデルを扱わないと存在意義がうすれる。

7. あとがき

民族学研究を進める上での補助道具としてのコンピュータ情報処理の問題と、「コンピュータ民族学」の性格について考察した。コンピュータが両者に関係しているが、前者では実際の処理装置として、後者では象徴的な意味に使われていると解すべきだと考えている。実際的な情報処理の問題としては、民族学研究は豊富な対象データをかかえている。画像処理、音声・音楽、言語処理など情報工学の専門分野の人々が協力して民族学研究で発生する本当のニーズがある具体的データを取り上げて研究する方向に進めば、我が国の情報処理研究も更に前進するものと思われる。

一方、象徴的な意味でのコンピュータ民族学は「理論民族学」、或は現実の世界ではなくモデルを取扱っているという意味で「仮想民族学」とでも呼ぶ方がふさわしいかも知れない。しかしモデルを用いてシミュレーションを行なう時にコンピュータを活用することはいうまでもない。

情報処理即ち図1でいう T1, T2 のレベルの研究はすでにいくつかの試みを行なっているが、コンピュータ民族学の問題といえるものについて具体的に成果があるわけではない。我々は先ず民族学へのコンピュータ導入という過渡期の役割としてデータ処理に対する新しい道具の開発から手がけている。この過程において実際のデータとその分析に習熟することにより、そのデータをいろいろに抽象化することの意味が明確になってくるものと期待している。文化の伝播、文化変容、言語の系統や変遷などの問題を文化とは全く関係ない物理現象のモデルやカタストロフィ理論、グラフ理論など数学モデルなどと関連させてみるなどの試みもやってみる必要があるだろう。

欧米に比べて少なくとも10年は遅れをとっている我が国の民族学研究におけるコンピュータ利用ではあるが、昭和54年1月より民博にも本格的な情報処理装置が導入された。外部の方々の御協力も得てこれから精力的に利用技術を開発していかなければならぬと思っている。

常日頃民族学におけるコンピュータの利用について御討論いただいている本館助教授栗田靖之氏，小山修三氏に感謝致します。

文 献

CLARKE, David L.

1968 Analytical Archaeology, Methuen & Co. LTD.

HYMES, Dell (ed.)

1965 The use of computers in anthropology, Mouton & Co.

坂井利之

1970 『情報学』 筑摩書房。

杉田繁治

1977 「コンピュータと民族学」『民博通信』no. 1, pp. 30-33.

1978 「コンピュータ民族学の役割」『コンピュータピア』No. 5 コンピュータエイジ社, pp. 62-67.

梅棹忠夫

1966 「文化分析の構想」『人文学報』22 京都大学人文科学研究所, pp. 149-164。

ウィリー, G. R./サブロフ, J. A.

1979 『アメリカ考古学史』小谷凱宣訳, 学生社。