

貝塚データベースの数量的分析(1)

－データベース構築と数量化－

及川 昭文

茨城大学・人文学部

貝塚遺跡は、それが存在していた時代の社会、文化、自然環境等を復元するための有用な手がかりを与えてくれる。すなわち、貝塚遺跡の立地環境、分布状況、時代、出土遺物（とくに貝、魚骨、獣骨等）を調べることによって、いろいろな考古学情報を引き出すことができる。本研究は、まずこれまでに発見されている日本全国の貝塚遺跡（約4,000）のデータベースを構築し、ついでこれまでとは異なる視点から、新たな仮説を立て、数量的分析手法を利用してそれを検証しようとするものである。本報告では、現在構築の最終段階にある貝塚データベースについて、これまでの経過と数量化に関する諸問題について述べる。

Quantitative Analysis of Shell-mound Database

- Database Creation and Quantification -

Akifumi Oikawa

Department of Humanities, Ibaraki University

Shell-mound(Kaizuka) archaeological sites give us an important clue to understand the society, the culture, the natural environment in the period when those sites were occupied. We can get much archaeological information by investigating such as environments around sites, geographical distribution of sites, chronological period and remains(especialy shell, fish-bones and animal-bones).

Our research project aims (1) to creat the database of shell-mound sites whose number is about 4,000 all over Japan, (2) to build up new hypothesis from the new point of view, and (3) verify such hypothesis through quantitative analysis. In this report, the development of the shell-mound database which is now under the final stage and several aspects about quantification.

はじめに

考古学研究においては「どこで、どのような遺跡が発見され、どのような遺構・遺物が出土しているか」ということを知ることから研究は始まるといってもよい。これらの情報を網羅したデータベースの構築は、考古学研究にとって必要不可欠なものである。筆者は十数年前から考古学データベースの構築の研究¹⁾に取り組んでおり、これまでに佐賀県全遺跡約6000の遺跡データベースを完成しているが、これを全国規模に広げていくことは個人の研究者のレベルを超えている。現在、奈良国立文化財研究所で全国規模の遺跡のデータベース構築の計画があり、一日も早い本格的、かつ具体的な着手が待望されている。

一方、すべての遺跡を網羅的に収録したデータベースではなく、「青銅器を出土している遺跡」「前方後円墳」「高地性集落遺跡」等々、特定の遺跡のみを集めたデータベースの試みは、多くの研究者によって進められている。「貝塚データベース」もそのようなものの一つで10年ほど前から構築を始めた。当初は、酒詰仲男の「日本貝塚地名表」「日本石器時代食糧総説」に基づいて作成したが、前者は1959年、後者は1961年に著されたものであり、貝塚遺跡の数は約900程度であった。²⁾当然ながら、その後発見される貝塚遺跡の数は増え続けており、今日では4000以上に達している。5年ほど前から、これらの貝塚遺跡に関する資料の収集、データベース化を本格的に進めてきており、現在最終的な校正作業を行っている。³⁾

この貝塚データベース作成の目的は、まず「考古学研究のための基礎資料とする」ことであり、次に数量的分析を行い、新しい視点から貝塚遺跡の諸相を明らかにすることである。具体的な数量的分析研究は現在進行中でその詳細はまだ述べることができない。本報告ではこの貝塚データベース作成上の諸課題、作成の過程で遭遇した問題、及び質的データの数量化に関して具体的な方法や問題について述べることにする。

1. 貝塚データベースの構築

ここでは、貝塚データベースの具体的な内容、構築上の諸問題、及びデータベース構築支援ソフトウェアについて述べる。

1.1 データベースの種類と内容

考古学は遺跡を始め「モノ」をその研究対象としており、それらの「モノ」に関するあらゆる情報が研究を進めていく上で必要になってくる。発掘調査や分布調査等から生み出される分布図、写真、実測図、拓本、あるいは報告書等の文献といった多種多様な情報のデータベース化が検討されなければならない。⁴⁾貝塚データベースにおいても実測図や写真等についてのデータベース化が望ましいが、今回の研究においては次の3種類のデータベースを作成した。

遺跡・遺物 貝塚データベースといった場合、このデータベースを指し、表1のような項目が収録されている。現在までに蓄積されているレコード数は約4,000であるが、これは5年ほど前までに発見されている貝塚遺跡数で、現在はこれに数百の遺跡数の増加が見込まれる。

文献 上記の貝塚データベースを構築するために利用した1次資料である調査報告書を主体とした文献の書誌情報データベースである。約1,200文献が収録されている。

貝属性 約800種類の貝について、生息域（水深、潮間帯等）、生活圏（内海、河川、汽水域、地上

表1 貝塚データベース項目一覧

遺跡番号
遺跡名
遺跡名(よみ)
所在地
県・市区町村コード*1
時代コード
遺構コード
土器編年
絶対年代
文献番号*2
遺物-貝類
遺物-節足類(フジツボ、カニ等)
遺物-棘皮類(ウニ等)
遺物-魚類
遺物-両生類
遺物-は虫類
遺物-鳥類
遺物-哺乳類
遺物-人骨
遺物-植物
遺物-その他*3

*1 JIS X0401, X0402 に準拠

*2 文献データベース含まれている文献のID

*3 貝製装身具等の特記すべき遺物

等), 生活状態(付着, 着生等), 分布(本州以南, 瀬戸内海等)等の項目を収録した辞書ファイルである。

1.2 構築上の諸問題

一般的にデータベースの構築には多大な労力と時間を必要とするが, その過程で生じる問題には, 実際に携わった者でなければ理解できないものが少なくない。それらの多くは本質的な問題ではないが, データベースを利用する段階では大きな問題となることが多く, できるだけ早い時期に解決しておく必要がある。とくに, データの品質を一定に維持することは, データベース構築において, 最も重要なことの一つであり, そのためにいろいろな工夫が必要になってくる。

(1) データの均質性の維持

考古学データベースを作成するための基本的な一次資料は, 「発掘調査報告書」である。これらは一つの調査単位で作成されるが, その様式は統一されておらず, 報告書のそれぞれで遺跡や遺物の記述は異なっているといってもよい。一冊の報告書に複数の遺跡の報告がなされていることも少なくないが, 同じ報告書の中でも報告者が違えば, その表現や記述様式が異なっていることは珍しくない。このため, 一次資料からデータシートを作成する際に, データ記述の統一を図ることに多大な労力と時間を費やすことになる。今回のデータベース作成でとくに問題となったことをいくつか挙げると, 以下ようになる。

遺物名の表記の不統一 現時点で貝塚データベースに含まれている遺物の種類は, 約3000種類になる。この数字は, まだ未修正のものを含んでおり, 最終的には2800種類位になると予測している。このうち種類が多いのは貝類(約1700)魚類(約500)鳥類(約200)哺乳類(約300)で, 当然といえば当然であるが, 貝類が他を圧倒している。これだけの数になると, 人手で表記の統一を図るのは不可能であるため, チェックのためのプログラムを作成し, かなりの部分を機械的に処理できるようにした。とくに, 貝類については, 「原色日本貝類図鑑」「統原色日本貝類図鑑」(吉良哲明著, 保育社刊)に記載されている貝の名称(約2800)をコンピュータに入力し, この

辞書ファイルとのマッチングによるチェックを行った。これらのエラーが, 一次資料にあるのか, データシートへの転記, あるいは機械可読化の時点で起こったのか具体的に述べることはできないが, 一次資料のレベルですでに間違っている例も数多くある。不統一の具体的な例をいくつか挙げると, 次のようなものがある。(●印に統一)

- クジラ } 文字種(カタカナ, 仮名, 漢字)
- くじら }
- 鯨 }
- アマオフネ } ・濁点のあるなし
- アマオフネ } ・~カイ(ガイ)の有無
- アマオフネカイ }
- ウラウスガイ } 「ズ」と「ツ」
- ウラウツガイ }
- ニッポンジカ } 「ニッポン」と「ニホン」
- ニホンジカ }

種の同定 貝塚遺跡から発見される貝殻にしる動物の骨にしる, 必ずしもその種が同定できるとは限らない。非常に少量であったり, 腐敗や損傷がひどかったり, あるいはもともと同定しにくいものであったりして, 何であるかを特定できない場合には,

- ウシORウマ
- イノシシ?
- シジミ属
- タラ科種不明
- ニシン・ウグイ類
- ニシン科
- ニシン科の一種
- ニシン類
- ネズミ等中小動物の微細骨
- ブリ・タイ類似の魚
- 小型クジラ類

等と報告書に記載されている。正確を期すためには, 遺物そのものをもう一度調べ直すということが必要になるが, それは非現実的であり, かといってこれらのデータを削除することもせっかくの「情報」を捨てることになる。現在は例に挙げたようなまま入力されているが, できるだけもとの情報を生かし, かつ統一的に表記できるような方法を検討しているところである。

(2) 存在しない所在地

遺跡の所在地は基本的に一次資料である報告書等に記載されているとおりに入力されているが, 市町村合併や町名変更等の理由で, 現在は存在しない所在地となっている例が少なくない。また,

現在は正しくても、将来同様の理由で不正な所在情報となることもある。この問題を解決する方法としては、次の2つが考えられる。

①市町村合併や町名変更の情報を確実に収集し、それをデータベースに反映させる。

②物理的な位置情報を緯度・経度の数値データとして、あるいは遺跡の位置を示した地図をイメージデータとしてデータベースに取り込む。

①は費用、時間、人手という観点からみれば、あまりに非現実的である。②の物理的な位置情報は、むしろ積極的にデータベースに取り込むべき情報で、それが入力されることにより、遺跡分布図の自動作成や、あるいは空間分析といった研究に大いに資することは明らかである。現在はまだこの情報は入力されていないが、将来的には緯度・経度情報だけでもデータベースに取り込んでいくことを予定している。

(3) 読めない遺跡名

一般的に遺跡の名称は、その遺跡のあるところの地名、あるいはそれに類した名称（例えば、○〇塚とか△△洞と昔から呼ばれていたというような場合）等からつけられることが多い。したがって、難読の地名があるように遺跡名にも難読なものがある。また、一般的な読み方でなく、特殊な読み方をするような場合もあり、データベースとしては遺跡名の読みは必須項目といえることができる。

ところが、実際には一次資料である報告書のどこを探しても、遺跡名の読みが見つからないのが普通である。さすがに、最近では遺跡名にふりがなを振った報告書が多くなりつつあるが、データベース化を前提にした報告書の様式の標準化ということを検討する時機が到来しているのではないだろうか。

貝塚データベースにも「鉾切遺跡、海戸貝塚、磯原貝塚、土穴瀬貝塚、大鼠蔵貝塚、定留貝塚」のように遺跡名の読めないものが数十あり、これらは現地の人に問い合わせるしか方法がなく、これもデータベース構築に時間がかかる理由の一つである。

2. データベース構築支援ソフトウェアの開発

データベース構築の省力化、効率化、均質性の向上を図るには、これまで述べてきたような問題の解決を含め、強力な構築支援ソフトウェアの開発が不可欠である。貝塚データベースの構築にお

いても、かなりの数のプログラムを開発したが、これらはUPS(Utility Package for SDF*)と称している支援プログラム群に含まれるもので、このUPSは現在富士通の大型コンピュータ上で稼働している。以下、このUPSについてその概略を紹介する。 *SDF:Standard Data Formatの略

2.1 UPS 開発の目的

市販されている、あるいはメーカー等の提供するデータベース関連ソフトウェアには、一般的に構築を支援するプログラムは含まれていない。すなわち、機械可読化されているいないに関わらず、データはすべてエラーのないクリーンな状態にあることを前提としており、データの均質性の保持は利用者側で対応すべきものとなっている。そこで、いろいろな種類のデータベース構築を行う過程で、汎用性を持った支援プログラム群UPSの開発を継続して行ってきた。このUPS開発の主な目的は、①データの物理的、論理的なエラー・チェックの自動化と、②豊富な冊子体資料の作成の2つである。

①は、人手と時間を要し、かつ単純作業の連続であるエラーチェック、修正の処理を可能な限り自動化することにより、その省力化・効率化のみならず、データの品質を一定の基準に維持することをめざすものである。

②は、構築されたデータベースをオンラインで利用するだけでなく、冊子体資料として活用できるようにすることをめざすものである。とくに考古学においては、オンラインで情報検索をするだけでなく、データベースに基づいて分布図やグラフ、あるいは地名表などの3次、4次資料を作成することが多く、そのためにはいろいろな種類の冊子体資料が不可欠なものとなってくる。

2.2 データ格納形式の標準化

多種多様なデータベースの構築をできるだけ標準的な手順で進められるように、SDFと呼んでいるデータの物理的格納形式を定めている。図1がそれで、非常に単純な構造となっているが、これまでの経験からいえば、イメージ等のバイナリー・データを除けば、ほとんどのデータベースのデータをこの形式で処理することが十分に可能である。UPSは、このSDFを前提として開発され

ているが、その他のアプリケーションでもデータの形式としては、この SDF のみを意識しておけばよく、汎用性のあるプログラムの開発が容易に行える。

この SDF には図 1 の形式とは別に、エディット形式 SDF がある。これは SDF レコードをメーカ提供のエディタ等で編集できる形式に変換したものである。物理的なファイル容量は増えるが、エディタの機能をフルに活用できるので、データの修正等を簡単に行うことができる。UPS のすべてのプログラムは、これらのいずれの SDF にも対応できるようになっており、また相互変換用プログラムも用意してある。

2.3 UPS の構成 (機能)

UPS は 1980 年代の中頃から開発が始まり、新種のデータベースや新しい問題に対応できるように、プログラムのバージョン・アップや追加が続

けられてきており、現在では約 100 本のメインプログラムと 200 以上のサブルーチンから構成されている。約 20 のサブルーチンを除いて、すべて FORTRAN 言語で書かれており、非常に高い可般性と汎用性を有している。

UPS 全体としての構成は次のようになる。

- ・データチェック
- ・データ変換 (自動修正を含む)
- ・レコード変換
- ・マージ処理 (項目及びレコード単位)
- ・ソート処理
- ・リスト出力
- ・KWIC(Keyword In Context) リスト作成処理
- ・統計処理
- ・分かち書き処理
- ・その他

UPS で処理できる文字は 1 バイト系の EBCDIC コードと 2 バイト系の JEF コードであるが、a e i o u の文字も 1 バイト系文字として取り扱う

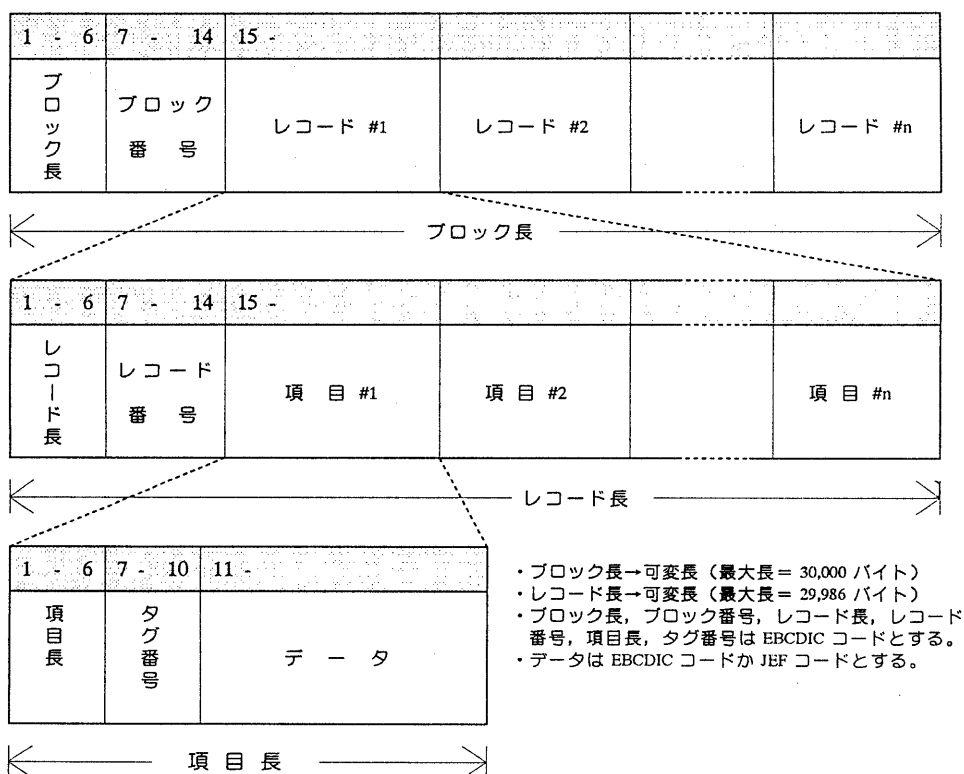


図 1 SDF (Standard Data Format)

ことができる。また、EBCDICコードではカタカナとアルファベットの小文字が同じコードに割り振られているが、これを異なるコードに割り振り、混在したテキストの処理も可能としている。

すでに述べたようにUPSの特徴は、データの

チェック・修正と冊子体資料の作成である。図2.3は、実際の処理結果の例であるが、いずれのプログラムも簡単なパラメータの指示で利用できるようになっている。

遺跡番号	遺跡名	コード類	出土遺物等
		PRD#23	
120001	南道合・遺跡 みなみちあい・いせき	ID1# 835 SAK# 369 CTY#12207 PRD#23,25,26,28	所在地：松戸市八ヶ崎264 編年：関山, 黒浜, 諸磯, 田戸下層, 堀ノ内 貝類：アサリ, オキシジミ, カガミガイ, カキ,
120002	新井・遺跡 あらい・いせき	CTY#12207 PRD#27,28	所在地：松戸市八ヶ崎11 編年：加曾利B, 加曾利E, 堀之内
120003	寺田・遺跡 てらだ・いせき	CTY#12207 PRD#40	所在地：松戸市馬橋 貝類：アサリ, ハマグリ
120004	千駄堀寒風・遺跡 せんだぼりさむかせ・いせき	ID1# 836 SAK# 370 CTY#12207 PRD#25,26,27,28	所在地：松戸市千駄堀1306 編年：阿玉台, 加曾利B, 加曾利E, 黒浜, 勝 貝類：アカニシ, アサリ, イタボガキ, イボキ, ツメタガイ, ハマグリ, マガキ, マテガイ 魚類：タイ 哺乳類：イヌ, イノシシ, ニホンジカ 人骨：有

図2 貝塚遺跡データベースのチェック用リスト

カイト ^ハ イタイク ^ン	関東州旅順管内山頭村	会大台山	遺蹟
カイト ^ン	備前熊山	戒壇	遺蹟考
カイチュウ		懐中	蠟燭
"		海中	出現鐘に就て
カイツ ^カ	栃木県藤岡町に於ける淡水	貝塚	
"	千葉県小見川町白井大宮台	貝塚	
"	周防国上郷	貝塚	
"	肥後国下益城郡当尾	貝塚	
"	武蔵東寺尾	貝塚	
"	富山県婦負郡小竹	貝塚	
"	詫田	貝塚	
"	伊予阿方	貝塚	
"	常陸国駒形	貝塚	
"	広島附近の	貝塚	
"	伊予新居郡の	貝塚	
"	筑前国糸島郡今津の	貝塚	
"	北海道爾志郡三ツ谷	貝塚	
"	伊勢大山田村	貝塚	
"	地形上より見たる	貝塚	
"	青森県上北郡早稲田	貝塚	
"	網走の堅穴及び	貝塚	
"	広島附近の	貝塚	(其二)
"	広島附近の	貝塚	(其三)
"	加賀国河北郡宇ノ気村上山田	貝塚	概報

図3 KWICリストの例(考古学雑誌の論文名)

3. 数量化

貝塚データベースを数量的に分析するためには、まずデータベースに含まれる個々のデータを数量化しなければならない。貝塚データベースに収録されている各項目のデータは、尺度としてはすべて名義尺度で表された、いわゆる nominal data である。これらの nominal data を数量化するためには、まず「考古学データ」の特質と問題点を、十分に理解しておくことが必要である。

3.1 考古学データの特質と問題点

考古学は「モノ」を対象とすることから、それらの「モノ」から得られる計測値（測定値）等、多種多様な量的データがあり、また土器様式や文様、あるいは古墳の形式といった質的データの種類も豊富である。しかしながら、これらのデータが、そのまま数量的分析に利用できるかという点、必ずしもそうではない。

一次資料としての報告書 考古学研究に必要な情報源は、多くの場合「報告書」に求められる。報告書を作成するのは実際に発掘調査を担当した専門家であるが、彼らがすべての分野に精通しているわけではない。縄文時代を専門とする者が前方後円墳の調査を担当させられるというようなことは珍しくない。その結果は当然報告書に反映されてくることになる。また、報告書が完成するまでには、発掘調査そのものから、遺物の整理、土器の復元、測量図や実測図の作成等、実に多くの人々が作業を分担することになるが、それらの人々の間で一定の基準や客観性を維持することは非常に困難であるという問題もある。

そして、一つの帰結として、信頼性の高い情報を得るためには0次情報である「モノ」そのものにあたらなければならないということになり、考古学研究においては「一次情報」の収集に、多大な時間と労力をかけることになる。このような状況は、「報告書」が研究のための一次資料として十分に機能していないということを意味する。

遺物の量的把握 一つの遺跡から、どのような遺物がどのくらい出土したかという量的把握は、数量的分析の第一歩であろう。ところが、この情報を得ることは非常に困難である（不可能に近いといった方がいいかもしれない）というのが現実で

ある。

まず、必ずしも遺跡の全ての範囲（遺跡の範囲をどうやって認定するかという根元的な問題もあるが）を発掘するわけではないから、すなわち出土した遺物の量は全体ではなく部分でしかないということがある。また、重要でない（と調査者が思っているだけという場合も少なくないが）遺物、例えば土器片については、最初から量的に把握しようと考えていないということもある。完形の土器も土器片も、得ようとする情報の種類によっては、どちらも同じ「情報量」を有していると考えられるが、出土した土器片についての情報を報告書から得ることは、ほとんどできない。

量的に把握したくてもそれが著しく困難な場合もある。今日の発掘調査は、ほとんどが行政発掘と呼ばれるものであるが、それには限られた予算で、限られた期間内に調査を終了しなければならないという制約がある。すべての遺物についてその量的データを得ることは、予算や時間をふくればあきらめることになり、結果的にそれは許されないということになる。

基準と精度 勝坂式、加曾利式、遠賀川式といった土器様式は、その編年とともに考古学研究に不可欠な要素としてある。しかしながら、それらの土器様式がはっきりと定義され、そして研究者の共通理解の上で使用されているかという点、現実にはそれらを使う研究者によってその定義は、違いの程度は別にして、異なっているということがある。また、土器や石器の計測は日常的に行われているが、そこにおいても計測項目や精度には微妙な違いがみられ、多くの場合「測定誤差」に関する考慮は希薄である。

このように考古学データにおいては基準（定義）や精度に大きなバラツキがあり、それが数量化を行う際の大きな問題としてある。

非再現性 遺跡の発掘調査は、遺跡の破壊でもある。すなわち、調査が終わればほとんどの場合、遺跡は消滅してしまうことになり、遺跡から情報を得るチャンスは一度しかないといえる。新しい技術や手法が開発されても、それらを適用して、再度調査を行い新しい情報を得ることはできない。10年前の報告書から得られる情報がいくら少なくても、その量を増やすことは不可能である。

3.2 貝塚データベースの数量化

どのような数量化を行うかは、どのような仮説を立てているか、あるいは何を明らかにしようとしているのかということに対応してくるが、仮説を立てるためには、まずそのための材料が必要である。

クロス集計 nominal data の数量化は、まずクロス集計から始まるというよいだろう。貝塚データベースに現在収録されている項目のうち、クロス集計が可能なのは、

- ・ 都道府県
- ・ 市区町村
- ・ 時代
- ・ 遺構
- ・ 遺物 (その他を除いた10種類)

の各項目の組み合わせである。

現在貝塚データベースに収録されているレコード数は4020であるが、遺物のうち、節足類(137)棘皮類(80)両生類(51)は虫類(204)人骨(160)植物(60)等が出土している遺跡数は()内の数字で、全遺跡数に比べて非常に少ない。これらについては、単純にクロス集計しても意味のある値が得られる可能性は低いと思われる。細かい単位ではなく、大まかな単位でのクロス集計の方が有効ではないかと考えている。ただし、どのような単位にするかは、十分に検討する必要がある。

「遺物の量的把握は難しい」ということはすでに述べたが、この貝塚データベースにおいてもその問題を抱えている。すなわち、アサリが出土しているということは分かって、その量は分からないということで、集計の結果得られる数字はあくまでアサリを出土している「遺跡数」でしかない。そのような数値から、意味のある情報を取り出すにはどのようなクロス集計、あるいは数量化

を行えばよいか、今後十分に検討していかなければならない。

貝属性ファイルを利用した数量化—マイクロからマクロへ 貝塚データベースにみられる貝は約1700種であるが、その内訳をみると

100以上の遺跡から見つかるもの	……	約 50
10 //	……	約 260
10以下の遺跡からのみ出土	……	約1400
1遺跡からしか出土しないもの	……	約 800

となる。これからいえることは、1遺跡でしか見つからないような貝にこだわっていると、全体像がなかなか浮かび上がってこない恐れがあるということである。

そこで、貝属性ファイルを参照し、個々の貝を属性で分類しそれを数量化する、すなわちマイクロな視点からマクロな視点への転換を図る。例えば、暖かい海で採れる貝、寒い海で採れる貝、浜辺で採れる貝、深いところでしか採れない貝といった大分類に基づいてクロス集計等を行う。

地理的位置情報 貝塚遺跡の立地しているのが、海岸近くなのか山間部なのかは、遺跡の性格を考察する上で重要な情報である。都道府県・市町村コードを利用することによって、ある程度の地理的位置を特定できるが、空間分析等を行うためにはより正確な位置情報が必要である。この位置情報の追加は今後の課題である。

注)

- 1) 及川「考古学データベースの構築」『シンポジウム「考古学とコンピュータ」1991 論文集』pp.66-85, 1991
- 2) 及川他「貝塚データベース—その作成と応用」『国立民族学博物館研究報告』5(2), pp.439-470, 1980
- 3) 科学研究費総合研究(A)「データベースを利用した貝塚遺跡の総合的研究」1990-92
- 4) 及川「情報化時代の考古学」『考古学ジャーナル』294, pp.4-7, 1988

表2 貝類、魚類、哺乳類の出土遺跡数(上位10種類)

貝 類		魚 類		ほ 乳 類	
ハマグリ	1457	スズキ	346	シカ	1023
アサリ	1037	クロダイ	299	イノシシ	812
アカニシ	904	マダイ	278	イヌ	401
サルボウ	821	ボラ	150	タヌキ	264
シオフキ	804	マグロ	118	クジラ	202
オキシジミ	803	タイ	97	サル	137
カキ	740	エイ	94	ウマ	134
ハイガイ	730	コチ	89	アナグマ	132
ヤマトシジミ	655	サメ	84	ノウサギ	110
ツメタガイ	623	ヒラメ	76	イルカ	81