苗字資料による国内の移住パターン推定の試み

斎藤　成也
東京大学理学部人類学教室

人 類 学 雑 誌 別 刷
第91巻3号　昭和58年7月

人類誌．J. Anthrop.Soc. Nippon
91 (3) : 309-322 (1983)

苗字資料による国内の移住パターン推定の試み

斎藤　成也
東京大学理学部人類学教室

要約　国内地域間の移住パターンを推定する方法のひとつとして、苗字を標識とする分析を試みた。苗字の種類が膨大なため、今回は北陸地方のＴ市に着目し、Ｔ市のア行の全苗字409姓が、22府県から選んだ53市にどのように分布しているのかを検索した。資料源には電話帳を使用した。分布に地理的な偏りのある苗字がいくつか見出されたが、一般に総個数の多い苗字ほど分布の広くなる傾向がみられた。一方、KENDALLの順位相関係数・PENROSEのshape distance等の統計量を使ってT市と他の市との苗字分布における親和性を推定した結果、地理的にT市と近接する市ほどより高い親和性を示した。

日本人集団はその歴史からみて、少なくとも最近1000年間は、かなり隔離の程度が高いメンデル集団（繁殖単位）とみなされるが、日本列島の南北に長い地理的形状を考えると、その内部はいくつかの小集団に分割されていると予想される。これら小集団間の遺伝的分化について、ABO及びMN血液型資料を分析したNEI and IMAIZUMI（1966）は、どちらの血液型にも都道府県レベルで遺伝子頻度に有意な差のある事を示した。このような遺伝的分化の程度は、集団の有効な大きさとともに、小集団間の移住率にも左右される（木村、1960）。AOKI and OMOTO（1980）はABO血液型のA遺伝子頻度にみられる日本の南北にわたる地理的勾配を、過去に移住のなかった２集団がある時点から一定の移住率で混合を始めたとするモデルで説明している。

一方、地理的勾配のような単純なパターンの他に、様々の複雑なパターンを示す地域変異が、多くの形質-生体計測値（池田、1978）、骨計測値（山口、1978；HANIHARA *et al*.., 1981）、皮膚紋理（岩本、1978）、遺伝的多型（尾本、1978）等-で報告されている。これらの地理的変異を、自然淘汰の地域差で説明するのか、偶然による遺伝的浮動の結果として説明するのかについては議論の分かれる所であるが、自然淘汰や遺伝的浮動を原因として考える際にも、小集団間の移住率が高い場合には、移住パターンの問題を無視する事ができない。特に、過去においてなんらかの要因で小集団間に遺伝的分化が生じていても、近年になって活発な移住が起これば、それらの分化パターンは大きく変化してゆくと予想される。

従って、海外からの遺伝子流入がほとんどなかった最近1000年間の日本国内の移住パターンを推定する事は、生体計測値・皮膚紋理・遺伝的多型等、現在の分布状況のみの資料から過去の様子を推測する場合、きわめて重要な問題となる。また地域小集団間の移住パターンは、遺伝的側面だけでなく、動いてゆく人間に伴なう文化の伝播パターンにもつながり、その意味でも興味深い問題である。

国内の移住パターン推定に関してはこれまでにも、人口動態統計や国勢調査の資料を用いた臼井ら（1972, 1974）の研究、アンケート調査に基づく篠崎（1967）の研究が全国的な通婚パ

ターンを分析している。一方**YASUDA** and **KIMURA**（1973）と**YASUDA**（1975）は１都市に着目して、その住民における夫婦間距離・親子間距離等の分布を求めている。**KATAYAMA** and **TOYOMASU**（1979）と**KATAYAMA** *et al*..（1981）は近接する小集団間の移住パターンを分析している。本研究では苗字を標識として用い、その地理的分布から、全国的な移住パターンの推定を試みた。

苗字の発生・増殖・移住モデル

苗字が父方から伝えられる文化では、その伝達はＹ染色体上の遺伝子のふるまいと同様である（**CROW**, 1967）が、この事を利用して**CROW** and **MANGE**（1965）は同姓結婚（isonymy）の頻度から近交係数を推定する方法を提唱した。日本でも**YASUDA** and **FURUSHO**（1971a, 1971b）、片山他（1978）、**KATAYAMA** *et al*.（1981）が戸籍・住民票を使って同姓結婚率を求め、近交係数等を推定している。なお、神崎（1954）は**CROW** and **MANGE**（1965）が同姓結婚率と近交係数を理論的に結びつけた10年以上前に、親等数と婚姻同姓率との関係を論じている（**YASUDA**, 1983）。

このように、苗字は遺伝現象と深い関係にあるが、両者は次の視点からも結びつけられる。ある遺伝子座に突然変異が起きて新しい対立遺伝子が生じると、その中の大多数は短期間で集団から失われるが、時にはしばらくの間低い遺伝子頻度で集団中に存在する事がある。これは稀な対立遺伝子（rare allele）と呼び、その多くは未だ他の集団に拡まっていないと考えられるために、その集団の標識遺伝子とみなされる。例えば、血清タンパクのひとつであるTf（transferrin）には、電気泳動法により、モンゴロイドの標識とされるTf Dchinaが他の対立遺伝子と区別されるが、日本人では一般集団中に0.5～1.0％の遺伝子頻度で存在する（尾本、1978）。

一方苗字も、上述のrare alleleの場合と同じように考える事ができる。即ち、過去のある時点・ある場所で、一個人が新しい苗字を名乗る（突然変異に相当）と、以後その苗字は、いわゆる分家により徐々に個数を増加させる。実際に、**YASUDA** *et al*.（1974）は苗字を男系だけに伝えられる一座位上の対立遺伝子とみなし、その消長を、苗字の種類に関して増殖率に差がない、即ち淘汰上中立であると仮定して分析している。

苗字が増殖する間には、発生地から他地域（小集団）へ移住する事もおこる。移住先でも苗字が同じように増殖してゆくとすれば、その苗字の現在の地理的分布を調べる事によって、過去の移住を再構成できる。ここで遺伝子の場合と異なる点は、苗字が一般にきわめて種類が多いために、個々の苗字だけを調べるのではなく、多種類を調べる事によって、各小集団が相互にどのような移住パターンで結びつけられているかを推定できる事である。即ち、個々の苗字の地理的分布を見ているだけでは、その苗字に特有の原因等によって、移住パターンの全体像を反映する事はむづかしいが、多数の苗字分布からは、よりばらつきの少ない、統計的にも信頼のおける推定が可能になるだろう。

ただし、苗字の地理的分布から移住パターンを推定するには、苗字の発生が特定の小集団に限られる（単元発生）という仮定が必要である。この仮定の妥当性に関しては、「討論」で検討する。また日本の苗字は一般に男系で伝えられるため、女子が結婚に際し、苗字を変えて移住した場合、苗字分布に変化がないから、この様な移住を苗字分布から推定する事はできない。養子の場合も同様である。

調査方法と苗字資料

日本の苗字の総種類数は、現在12万前後と推定されてあり（日本ユニパック、1978）、きわめて膨大な数にのぼる。そのため今回はある一地域に着目し、そこに存在する苗字が他地域にどのように分布しているかを調べる方法をとり、調査する苗字の種類数を制限した。着目した地域は福井県武生市（1980年現在の人口6.7万人）であり、比較する地域としては、福井県を含む22府県から53市を選んだ（図１参照）。これら53市の人口は41万～３万（平均11万）にばらついている。比較する市の選択は、まず福井県内の他の６市、次に福井県に隣接する４府県（京都・滋賀・岐阜・石川）から12市、さらに、できるだけ地理的に偏らぬように選んだ17県から、１県あたり２～３市、計35市を選んだ。市の選択基準としては、県庁所在地等人口の大きい市はできるだけ避けるようにし、同一県から選ぶ複数の市は互い地理的に離れているようにした。しかし平野部に市が集中するという地形的な制約等があり、選ばれた市は原則通りには分布していない。

苗字資料源としては、電話帳（1979年～1981年発行のもの）を用い、今回はア行の苗字のみを調査した。着目した福井県武生市の電話帳ア行からは、409種類の苗字が得られた。これは日本の苗字総種類数の約1/300にあたる。表１にこれら409姓の苗字個数分布を示した。

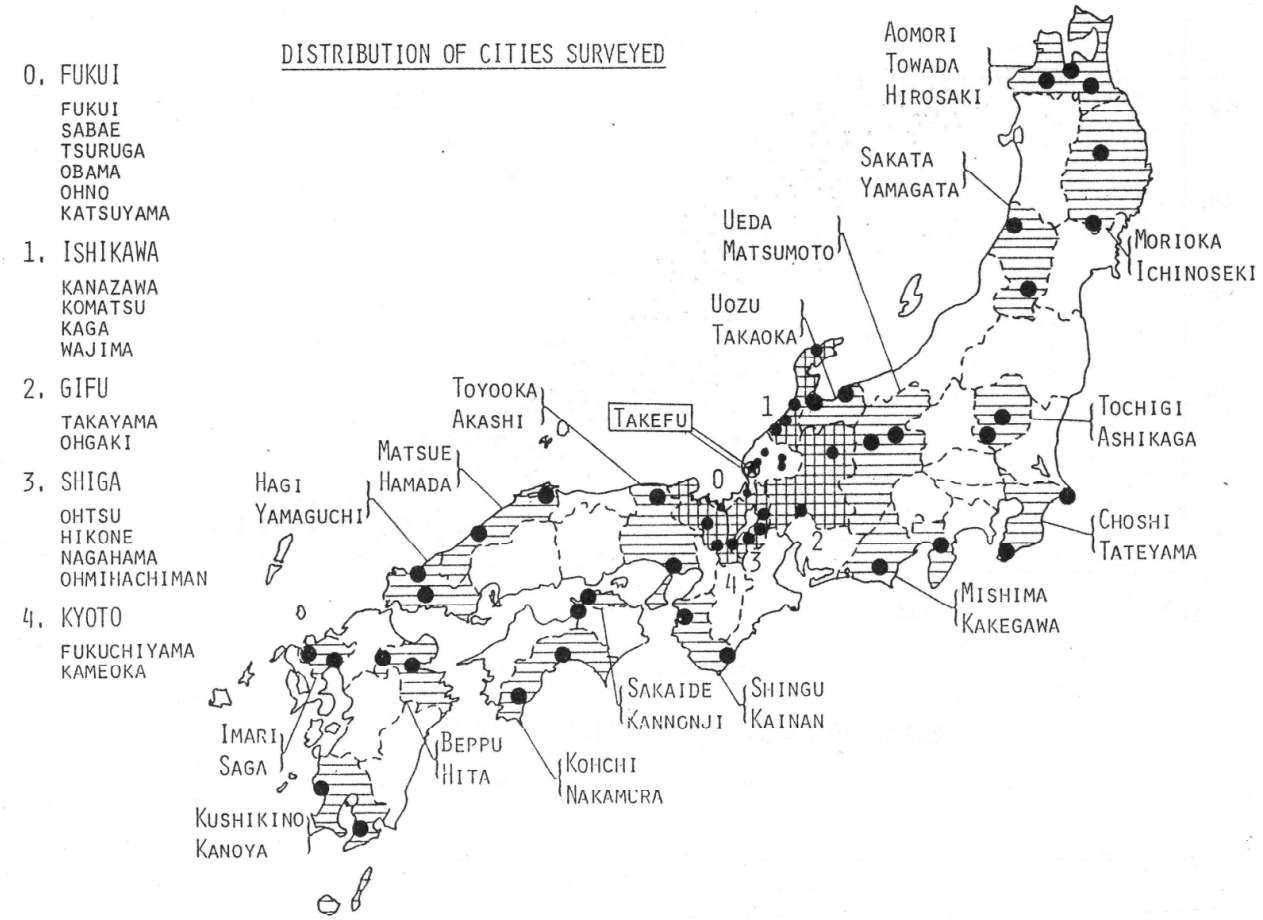

図　1　調査した54市の分布

表1　武生市から見出されたア行409姓の苗字個数分布

苗字個数	苗字種類数
1	160
2	61
3	32
4	25
5	19
6	13
7	10
8	8
9	6
10	11
11— 15	15
16— 20	13
21— 30	11
31— 40	8
41— 50	3
51—100	13
100＜	1
計	409

比較した他の53市の電話帳を使用した苗字個数の算定は、20個を越えるものについては第２桁以上の概数にとどめたが、統計処理の都合上、数値として中央値（例えば50個台ならば55）を与えた。個数の算定には、企業の地方支店と明記されているもの等、個人電話帳に掲載されてはいるが明らかな苗字とはみなされないものは含まれていない。

結　　果

武生市から見出されたア行の苗字409姓のうち、他の53市すべてに見出されたもの16姓、逆に武生市以外のどの市にも見出されなかった苗字が24姓あった。次に、苗字総個数（ある苗字が武生市を含む54市全体で見出された個数）と以下の2量との相関をみた。

（Ⅰ）苗字の見出された市の数（図2a参照）

（Ⅱ）苗字分布の変動係数（図2b参照）

この場合の変動係数は、苗字個数の市間の標準偏差を、その平均で割った値である（付録参照）。図2aから、苗字総個数の多い苗字ほど、一般に多数の市に分布する事がわかる。この一般的傾向からのずれには2通りあり、総数の多い割に分布する市数の少ない場合は図の下方に、逆の場合は上方に飛び出す。図2aには前者の場合がより多く示されている。図2bの場合も苗字総個数が増えるに従って、分布のばらつきの尺度である変動係数も小さくなる傾向がみられるが、図2aほど明瞭ではない。これらの図をもとに、特に地理的分布に偏りのある苗字2種類（小野寺と表）を選び、図3a・3bに示した。

次に、調査した409姓の資料から、武生市と他の53市との苗字分布に関する親和性を、以下の4つの示標に基づいて算出し、その地域差を図示した。

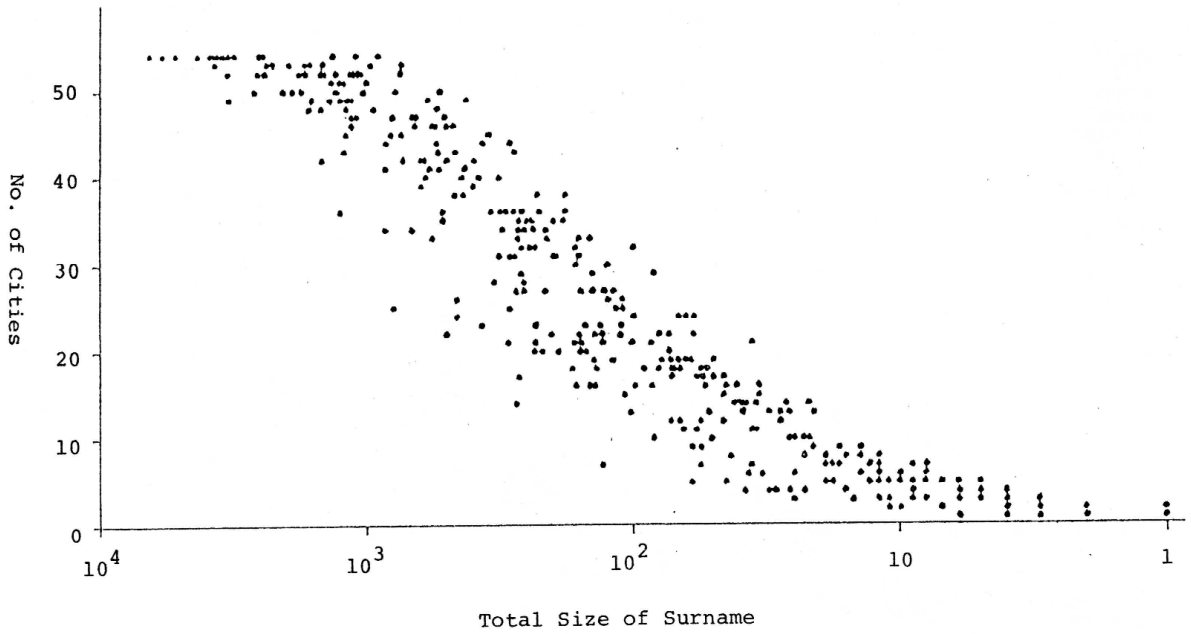

図2a　苗字総個数と苗字の見出された市数との相関

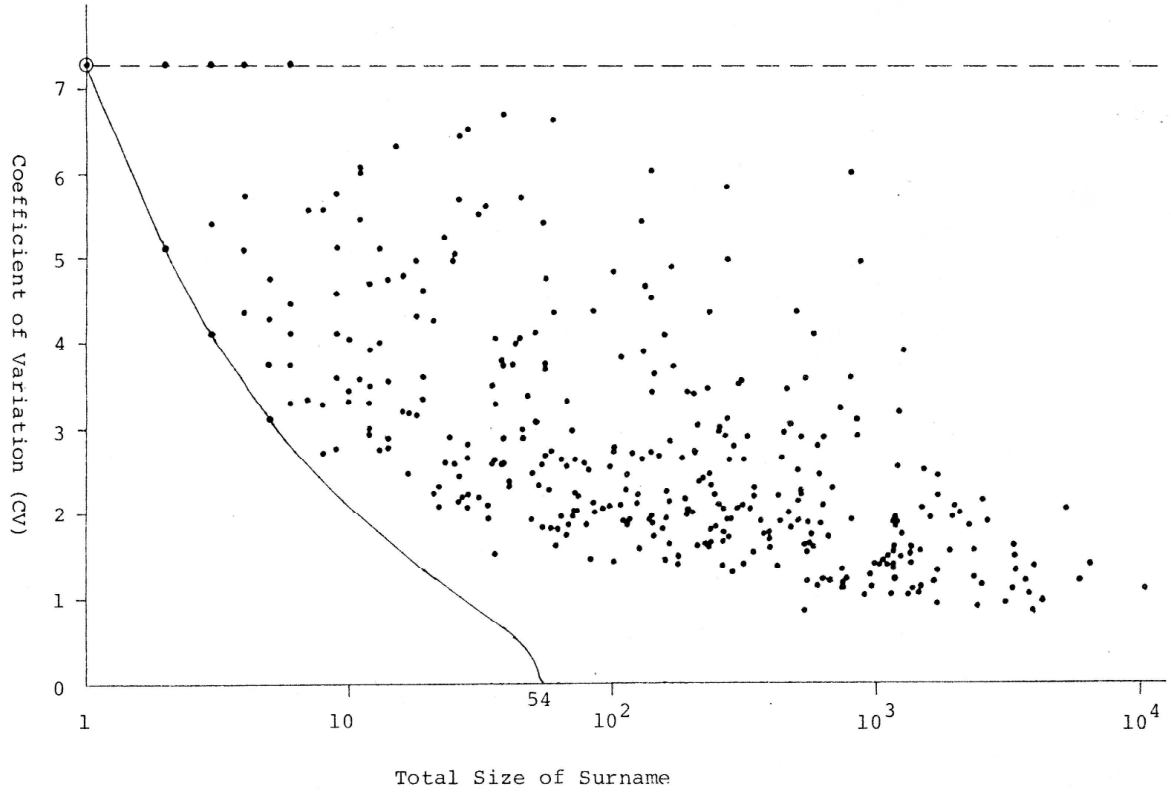

図2b　苗字総個数と変動係数との相関　　破線：変動係数の最大値，実線：変動係数の最小値（付録参照）

図3a，b 及び図４**a～d** では，作図の都合上，日本全体図中には，石川・福井・滋賀の三県のみは，各々以下の市のみを図示し，その他の市は図右下の小地図（全市を含む）に示した。石川県：輪島・金沢，福井県：武生・大野・敦賀・小浜，滋賀県：大津

(1)　各市で見出された苗字種類数（図4a）

(2)　武生市以外に最大４市までしか分布していない苗字種類に関する、各市で見出された苗字種類数（図4b）

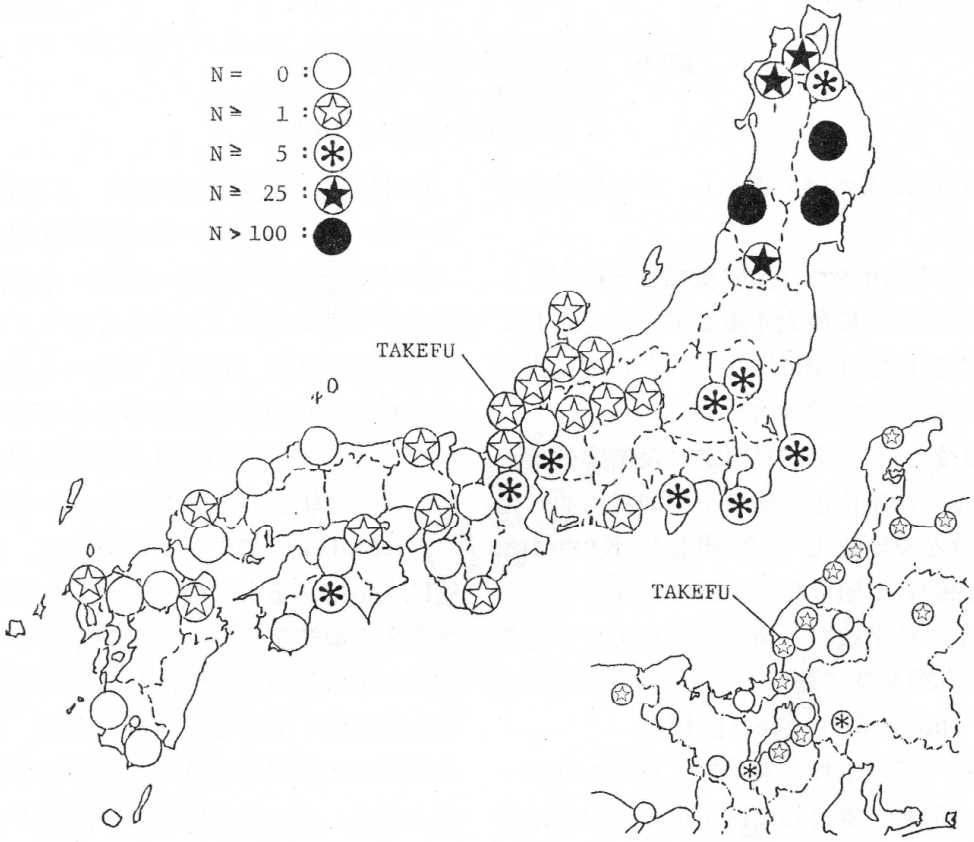

図3a　東北地方に分布の偏る例(小野寺)
苗字総個数：1256　　分布する市数：36

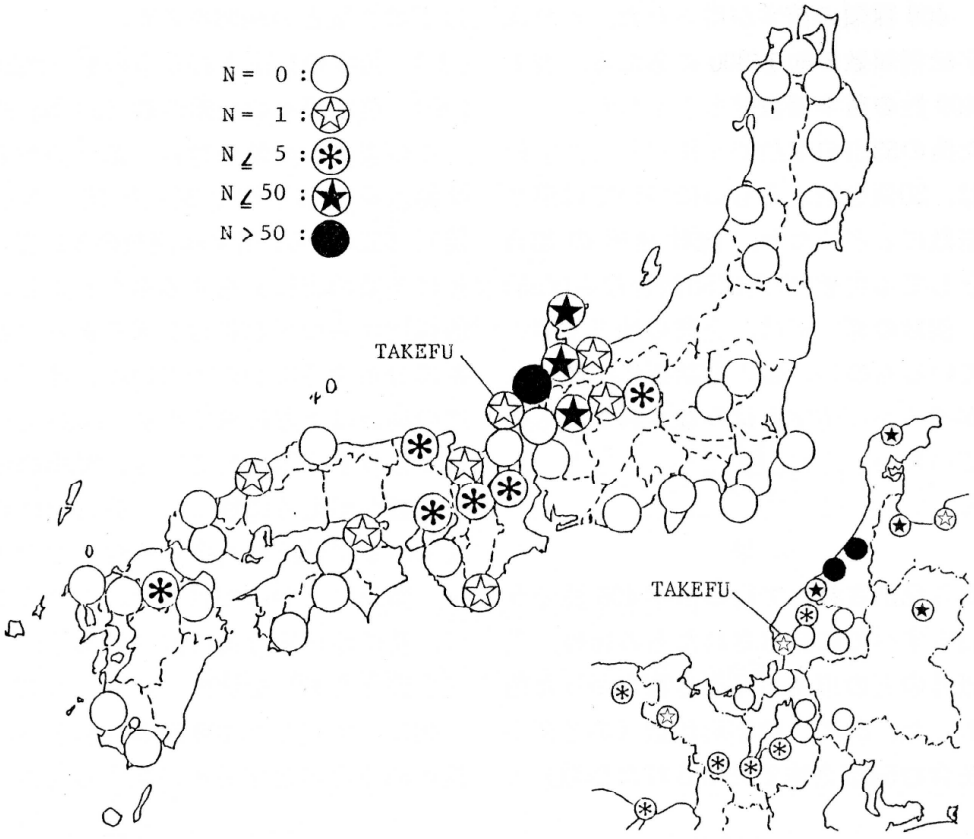

図3b　北陸地方に分布の偏る例（表）
苗字総個数：297　分布する市数：21
N：苗字個数（図3**a**，3**b**とも）

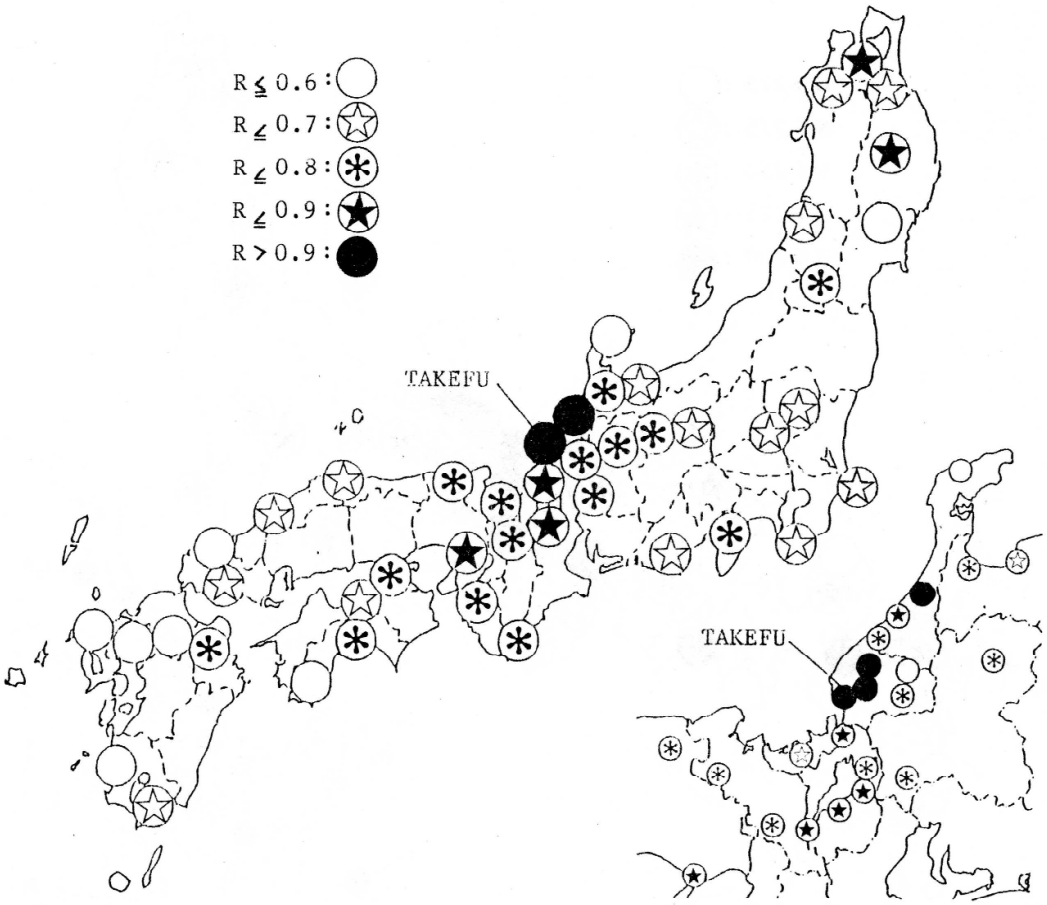

図4c　武生市との親和性（3）：KENDALLの順位相関係数（R）を示標としたもの

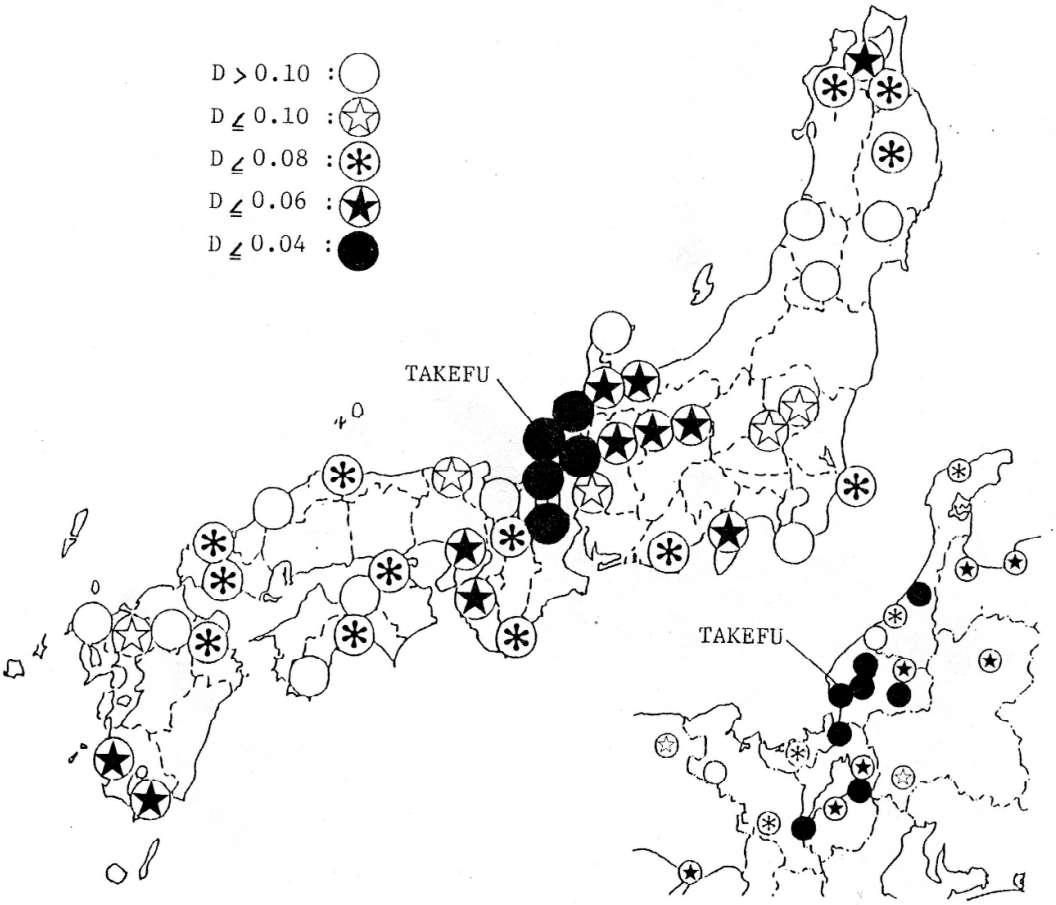

図4d　武生市との親和性（4）：PENROSE の shape distance（D）を示標としたもの

(3) 409姓を武生市で見出された苗字個数順に並べた順位と、他の市で同様に並べた順位とを比較して算出した**KENDALL**（1962）の順位相関係数（図4c）

(4) 各市の苗字頻度（苗字個数を市人口で割った値）を基に算出した**PENROSE**(1954) の shape distance（図4d）

(1)〜(4)すべての示標において、どれもほぼ一定の傾向、即ち、着目した武生市に地理的に近い市ほど、武生市と苗字分布に関する親和性の高い事が見出された。

(1)は最も単純な示標だが、示標(3)、(4)と似たパターンを示しているのは興味深い。ただし人口の大きな市（高知・明石・金沢等）の親和性がやや他のものより大きくなっているようである。示標(2)は「考察」で述べる理由により、分布する市数の少ない苗字のみで親和性を表示した。武生市近傍の市ほど親和性が高いという傾向は示標(1)よりも顕著である。示標(3)で用いた、**KENDALL**の順位相関係数は、市によって値に大小はあるものの、調べられた53市すべてが武生市の苗字順位とは0.1％水準で統計的に有意な差はない。これは苗字総個数の多い苗字は全国に広く分布しているために、どの市でも上位の苗字が似た順位になっているためだと考えられる。逆にいえば、相関係数の大小は、下位の、総個数の少ない苗字の寄与が大きい事になり、示標(2)と同様のパターンを示している事がうなずける。示標(4)のみは、「考察」に述べるように、苗字個数を各市の人口で割った頻度から**PENROSE**のshape distanceを計算した。示標(1)、(3)と共通した点であるが、武生市と地理的に離れているにもかかわらず、青森市にやや高い親和性が観察され、またこれは示標(4)のみの現象だが、鹿児島県の２市（鹿屋・串木野）の親和性がやや高まりをみせている。これらが、検索した苗字種類数が少ないために生じた誤差なのかどうかについては、今後調査する苗字の種類数を増やす等の再検討が必要であろう。

考　　察

1．苗字の単元発生について

苗字の地理的分布から小集団間の移住パターンを推定するためには、苗字が単元発生する（ある小集団のみで発生した）という仮定が必要である事は既に述べた。以下ではこの仮定の妥当性を検討する。

日本の苗字は、丹羽（1980）によればその９割までが地名からとられたものであると言う。遠く離れた２地点に同一地名の付けられている事はよくあるが、それぞれの地名から独立に同一苗字の発生する場合が考えられる。この時、実際の移住がなかったにもかかわらず２地点に同一苗字が独立に発生するから、現在の苗字分布による移住の推定に謝りを導入する事になる。このような多元発生は、地名から派生した苗字に限らず、同じ職業の者が独立にその職業に由来する同一苗字を名乗る事からも生じるであろう。地名・職業の他にも苗字のもととなった原因が多数あるが、それらも多元発生を導く事があるだろう。

どの苗字が単元発生であり、どれが多元発生であるかを識別するには、各苗字の発生過程を正確に把握する必要があるが、多数の苗字を扱う場合、これはきわめて困難である。しかし大雑把な基準として、総個数のきわめて大きな苗字は多元発生したと仮定し、地域的な比較から除外する事が考えられる。

苗字が毎世代同じ増殖率で増えてゆくと仮定すれば、t世代前に発生した苗字の世代毎の増殖率をmとした時、現在の苗字個数は$(1+m)^t$ 個となる。苗字によって増殖率mにはばらつきがあるだろうが、すべての苗字の平均増殖率は、日本人口の増殖率と一致するはずである。平安時代初期の日本人口は約5×10^6人と推定されている（沢田、1926）が、これは現在の人口約1×10^8人のおよそ1/22にあたる。即ち、約1000年間で苗字個数は22倍になるはずである。ただし

この増殖率には消失した苗字の寄与があるので、現存する苗字に限った平均増殖率はより大きな値となる事が予想される。

今、２世代毎にひとつの分家を生ずる苗字を考えると、世代毎の増殖率$m=\sqrt{2}-1$であるから20世代後（１世代25年とすれば500年後）にようやく発生時の１個から1024個に増える。しかしこの値ですら、平安初期以降現在までにせいぜい人口が数十倍にしか増加していない事と比較すれば大きすぎる。20世代にわたってこのような高い増殖率（$m \fallingdotseq 0.41$）で増え続ける事は、生物学的にみて矛盾している。

ところが日本でよくみかける姓のなかには苗字個数が10^5を越えるものが少なくない。表2にいわゆる「佐久間ランキング」（佐久間、1968）に基づき、苗字個数のクラス別（８段階）に苗字を分類したもの（歴史資料調査会、1976）のどのクラスに、今回調査した409姓が分布するかを示した。このうち82姓が苗字個数$\boxtimes \times 10^4$個以上のクラスに属している。これら莫大な個数を有する苗字は、到底生物学的な増加だけでは説明できず、なんらかの社会的要因が加わっていると考えられる。

これらの要因の結果として、現実の人間の移住を伴なわず、ある苗字の情報のみが他の小集団へ伝わり、拡がっていった事も考えられる。個数の大きな苗字の場合には、このような飛び火現象が各地・各時代で多数生じたと思われる。従って、図4bで使用した示標のごとく、苗字総個数の少ない，即ち分布する市数の少ない苗字だけを取り出してみる事が必要である。今後は地域的な比較をする際に、調べる苗字の種類を、例えば佐久間ランキング（上位4000位）に出ていない苗字に限るなどの基準を設け、なるべく単元発生の仮定を満たすと思われる苗字に限定するべきであろう。

表2　今回検索した409姓の「佐久間ランキング」における位置

Rank	Range of P*	N**	A***
1— 13	$50.0 \leqq P$	1	1
14— 40	$20.0 \leqq P < 50.0$	10	11
41— 101	$10.0 \leqq P < 20.0$	13	24
102— 370	$5.0 \leqq P < 10.0$	58	82
371— 694	$3.0 \leqq P < 5.0$	38	120
695—1269	$1.0 \leqq P < 3.0$	52	172
1270—2147	$0.7 \leqq P < 1.0$	30	202
2148—4000	$P < 0.7$	64	266
4001≦	—	145	409

*P：苗字個数（$\times 10^4$）
**N：苗字種類数
***A：苗字種類数の累積数

2.　電話帳を苗字資料源とした事について

今回の調査では、資料収集が簡便な事もあり、電話帳を資料源とした。個人別電話帳に載せられている苗字１個が１世帯にあたると考えられる。即ち今回検索した苗字個数の分布は、世

帯数の分布に置きかえられる。しかしこの場合に、以下の様な偏りが生ずる。

(1)　すべての世帯が電話を持っているわけではない。また持っていても電話帳に載せない場合がある。

(2)　親から子へ伝える苗字（戸籍上の姓）以外の苗字を載せる場合がある。

(3)　同一漢字で発音の異なる苗字は区別がつかない。

(1) については、1979年現在で70％以上の世帯が電話を持っており（日本電信電話公社、1981）そのうちで電話帳に載せていない少数の世帯を除いても、電話帳の苗字個数分布は十分に実際の世帯数分布を反映すると考えられる。
(2)については、大規模で組織的な変名が行われている場合はともかく、散発的に芸名・雅号等が載せられている場合には、あまり全体の結果には影響を与えないだろう。(3)については、苗字は電話帳に発音・漢字の種類を考慮した順序で並べられているので、同一漢字で発音の大きく異なる苗字（例：東、アズマとヒガシ）に関しては問題がない。よく似た発音の場合（例：大庭、オオニワとオオバ）両者を区別できない可能性があるが、この様な例は少ないと思われる。

3.　親和性を表わす示標の意味と有効性

今回の調査で得られた資料から、着目した武生市と他の市との苗字に関する親和性を、４つの示標を使って表わした（図4a～4d）。これら相互の比較を行う前に、種々の示標の表わす親和性と移住パターンとの関係を考察する。

再び遺伝子と対応づけするために、苗字に関する親和性を、集団間の遺伝的親和性を表わす尺度である遺伝距離と比較する。現在までに様々な遺伝距離が提唱されているが、どの場合でも、調べられたすべての遺伝子座において集団間で一致する対立遺伝子がなければ最小の親和性を与え、逆にすべての遺伝子座において各対立遺伝子の遺伝子頻度が一致すれば最大の親和性（遺伝距離は０となる）を与えている（根井、1977）。苗字資料に対しても遺伝距離と同様の尺度が考えられるが（LASKER,1977；KATAYAMA and TOYOMASU、1979）、これは人口の少なく、かつ近接した隔離集団間を全数比較する場合には有効だが、それらよりもはるかに人口の多い都市間の比較には不都合である。なぜなら、苗字の種類は平均的な一遺伝子座上の対立遺伝子数に比べてきわめて多いため、二集団間で意味のある比較をするには、標本数をかなり大きくとらなければならないからである。日本で最も頻度の高い苗字（佐藤・鈴木）ですら、苗字全体では各々２％以下の頻度にとどまっている。（田中、1972）。

従って今回の検索では、一つの市に着目してその市に存在する苗字の一部分（ア行の苗字）をあらかじめ選択し、それらの苗字が他の市にどのように分布するかを調べる方法をとった。ある集団の遺伝子頻度を推定するために一般に行われる任意抽出とはまったく異なる標本抽出方法であるが、苗字種類の選択が苗字の全種類からの任意抽出であれば、これらの分布から集団間の親和性を推定する事には意味がある。

遺伝距離で２集団を比較する場合、前提として過去のある時点で共通祖先集団が分裂し、以後各分集団が遺伝的に少しずつ離れていくとしているため、十分高い移住率で混合しない限り、両分集団はその遺伝距離を時間の経過とともに増大させてゆく。一方苗字は、ごく最近になって生じた世代間伝達系であるため、すでに多数の小集団が存在している事を前提としている。生ずる苗字は各小集団間ですべて相互に異なると仮定しているため、小集団間の苗字に関する親和性は、移住のおこらない限り０である。従ってふたつの小集団間に苗字分布に関して親和性が認められれば、その程度に応じた移住が過去にあったと推定される。

今回は、親和性を示す尺度として４つを選択した（結果参照）が、(1)苗字種類数の比較は最

も単純な示標であり、武生市に存在する409姓のなかで、ある市に見出される数が多いほど、その市と武生市は活発に移住を行っていると考えられる。これは苗字の発生・増殖・移住モデルからただちに理解されよう。(2)分布市数の少ない苗字だけの比較は(1)の変形であり、上記モデルに、よりあてはまると考えられる苗字を選んだものである。(3)**KENDALL**の順位相関係数は、苗字個数の大小関係を比較するものだが、ふたつの小集団間で十分長い期間にわたり相互に移住がおこれば、両者の苗字の種類が似かよってくるばかりでなく、各苗字の個数も似てくる事が期待される。従って、順位相関係数の高いほど、移住が活発であると考えられる。(4)**PENROSE**のshapedistanceも(3)と同様に、苗字個数分布に基づくものだが、偏差の二乗和を扱うため、苗字の様にその種類によって個数の著しく異なる場合には、shapedistanceもsizedistanceと同様、大きさにひきずられて変動する。そのため、苗字の発生・増殖・移住モデルからすれば個数を比較する事が好ましいが、この示標の場合だけは、市の人口で個数を割った苗字頻度からshapedistanceを算出した。

苗字分布に関する親和性から移住パターンを推定する場合，次のような問題点がある。今，２集団Ａ・Ｂの苗字分布を比較して，互いに共通苗字の種類が多く，高い親和性を与えたとしよう。この結果から，集団ＡとＢの間には直接に活発な移住が相互に生じていると推定したくなるが，以下の様な可能性もある。即ち，別の集団Ｃが存在し，そこから集団Ａ・Ｂの両者へ一方通行の活発な移入がおこっている場合である。この時，Ａ・Ｂ間で認められた親和性は，両集団間の実際の移住の反映ではなくなる。

結論と展望

日本国内の移住パターンを推定するため苗字資料を用いた解析を試みた。一都市に着目し，その市と他の市との苗字分布に関する親和性を，いくつかの示標を使って表した。その結果，着目した市と地理的に近い市ほど，より親和性が高いという傾向が，どの示標の場合にも見出された。

親和性の大小が，どのように移住パターンと結びつくかという問題は，まだ厳密な対応付けがなされておらず，今後の検討を要する。また，検索する苗字種類は，移住標識として適切なものを選択する事が重要である。単元発生の仮定を満たさないと考えられる苗字個数の莫大なもの，逆にきわめてまれな苗字で，全国的にもほとんど分布していないものの両極端を効率よく除く事が必要であろう。

今後は，今回のように一都市に着目するのではなく，各都市間の親和性を推定できる様な苗字種類をあらかじめ選択して検索を進める事が望ましい。

謝　辞

本研究を進める間，一貫して指導・激励していただいた，東京大学理学部人類学教室の尾本惠市教授，並びに放射線医学総合研究所の安田徳一博士に，深く感謝の意を表する。

文　献

Aoki, K. and K. Omoto, 1980: An analysis of the ABO gene frequency line in Japan: a migration model. J. Anthrop. Soc. Nippon, 88: 109-122.

Crow, J. F., 1967: Genetics and medicine. Heritage from Mendel (ed. A. Brink), Univ. Wisconsin Press, Madison, 351-374.

Crow, J. F. and A. R. Mange, 1965: Measurement of inbreeding from the frequency of marriages between persons of the same surnames. Hum. Biol. Quart. 37: 199-203.

Hanihara, K. *et al.* Commission of Anthropological Investigation of Modern Japanese Crania), 1981: Statistical data for modern Japanese crania-distance coefficients based on metric data. J. Anthrop. Soc. Nippon, 89: 519-522.
池田次郎， 1978：日本人の生体計測， 人類学講座6「日本人II」（池田次郎編）,雄山閣， 東京,101-141.
岩本光雄， 1978：日本人の皮膚隆線系， 人類学講座6「日本人II」（池田次郎編）,雄山閣， 東京,63-100.
神崎　貢， 1954：婚姻同姓率， 生物統計学雑誌， 2：292-298.
Katayama, K., T. Kudo, M. Suzuki and H. Matsumoto, 1981: Genetic studies in Tobishima: II. Population structure. J. Anthrop. Soc. Nippon, 89:427-437.
Katayama, K. and T. Toyomasu, 1979: A genetic study on the local populations in Mie Prefecture IV. The genetic relations among the Kamishima, Toshi, Momotori and Toba populations. J. Anthrop. Soc. Nippon, 87:377-392.
片山一道・豊増翼・松本秀雄， 1978：三重県離島における遺伝学的研究II， 神島の集団構造について， 人類誌， 86：83-94.
Kendall, M. G., 1962: Rank correlation statistics (3rd ed.), Griffin, London.
木村資生， 1960：集団遺伝学論， 培風館， 東京,233-266.
Lasker, G. W., 1977: A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations. Hum. Biol. 49:489-493.
根井正利， 1977：遺伝距離， 人類学講座10「遺伝」（松永英・尾本惠市編）,雄山閣， 東京,29-62.
Nei, M. and Y. Imaizumi, 1966: Genetic structure of human populations. III. Local differentiation of blood group gene frequencies in Japan. Heredity, 21:9-36.
日本電信電話公社， 1981：図説電信電話サービス， 日本電信電話公社総裁室広報部， 東京.
日本ユニパック（編）， 1978：日本の苗字， 日本経済新聞社， 東京.
丹羽基二，1980：姓氏，鷹書房，東京，5.
尾本惠市，1978：日本人の遺伝的多型，人類学講座6「日本人II」（池田次郎編）,雄山閣， 東京,217-263.
Penrose, L. S., 1954: Distance, size and shape. Ann. Eugen. 18:337-343.
歴史資料調査会（編）， 1976：日本人の代表姓氏2600， 歴史読本臨時増刊「日本姓氏総覧」：198-299.
佐久間英，1968：日本の名字， 谷川商事， 山梨.
沢田吾一， 1926：奈良朝時代の人口の研究（1）,史学雑誌， 37：108-142.
篠崎信男， 1967：通婚圏に関する一考察， 人口問題研究所年報， 12：48-52.
田中康仁， 1972：日本人の姓と名の統計， 言語生活， 254：72-79.
臼井竹次郎・方波見重兵衛・金子功， 1972：東は東・西は西一本籍地三ブロック別ならびに東西日本の初婚夫妻の組合せ， 日本医事新報， 2506：70-71.
臼井竹次郎・方波見重兵衛・金子功， 1974：明治・大正・昭和の人口移動． 公衆衛生院研究報告， 23：89-121.
山口敏， 1978：日本人の骨， 人類学講座6「日本人II」（池田次郎編）,雄山閣， 東京,143-174.
Yasuda, N., 1975: The distribution of distance between birthplace of mates. Hum. Biol. 47:81-100.

Yasuda, N., 1983: Studies of isonymy and inbreeding in Japan. Hum. Biol. 55:263-276.
Yasuda, N., L. L. Cavalli-Sforza, M. Skolnik and A. Moroni, 1974: The evolution of surnames: an analysis of their distribution and extinction. Theoret. Pop. Biol. 5:123-142.
Yasuda, N. and T. Furusho, 1971a: Random and nonrandom inbreeding revealed from isonymy study. III. Small cities of Japan. Am. J. Hum. Genet. 23:303-316.
Yasuda, N. and T. Furusho, 1971b: Random and nonrandom inbreeding revealed from isonymy study. IV. A group of farm-villages in Japan. Jpn. J. Hum. Genet. 15:231-240.
Yasuda, N. and M. Kimura, 1973: A study of human migration in the Mishima district. Ann. Hum. Genet. 36:313-322.

（1982年6月8日受付）

付　録

（1）　変動係数について

今，ある苗字の地理的分布をT個の市で調べ，各市において見出された苗字個数がn_i個（i=1,.....T）になったとする。この時，苗字総個数(N),及び一市あたりの平均苗字個数(M)は,

$$N=\sum_{i}^{T} n_i, \quad M=\frac{N}{T}$$

で与えられる。また，苗字個数の市に関する分散（σ^2）,及び標準偏差(SD)は,

$$\sigma^2=\frac{1}{T}\sum_{i}^{T}(n_i-M)^2, \quad SD=\sqrt{\sigma^2}$$

となる。ところが苗字個数のばらつきといっても，例えばある市だけに見出され，残りの市には見出されなかった苗字に対しては，その苗字総個数(N)が何個であっても，他の市へ拡がっていないという意味で同一のばらつきを示す値を与えたい。そのために，標準偏差(SD)を平均(M)で割って規格化したものが変動係数(CV)である。実際，n個の苗字が1市だけにかたまって見出された場合,

$N=n, M=N/T, SD=M\sqrt{T-1}$,

$CV=SD/M=\sqrt{T-1}$

となり，変動係数(CV)は苗字総個数(N)によらず，調べた市数(T)だけで定まる。今回の調査では$T=54$であるから，１市（即ち着目した武生市）のみに見出された苗字の変動係数(CV)$=\sqrt{53}$(≃7.28）である。（図2b参照）。

（2）変動係数の最小値について

上述した様に，変動係数の最大値は調べた市数(T)だけで定まるが，最小値は苗字総個数(N)にも依存する。$1\leq N\leq T$の場合を考えると，この時最小値を与える分布は，N個の市に各１個苗字が分布し，$(T-N)$個の市には各０個分布しているもので，この時変動係数(CV)は,

$$CV = \sqrt{(T-N)/N}.$$

この関数は，$1 \leq N \leq T$ で単調減少であり，$N=T$，即ちすべての市に１個ずつ苗字が分布した状態で０になる。図2b に このグラフを示した。一般に，$(i-1)T \leq N \leq iT$ $(i=1,2,\ldots)$ における変動係数の最小値は

$$CV = \sqrt{(i-1)(1-[N/T])/M}$$

（$i=[N/T]$，$M=N/T$，$[\ ]$ はガウスの記号）となり，$N>T$ の範囲ではこの最小値は十分小さい（<1.0）ので，図2b には図示していない。