確率モデルによる新分野の開拓

伊藤栄明／上田澄江

1. 確率モデルの発見

確率論は現実の現象と深くかかわりあっている。19世紀以前からの確率論は魅力にみちており、確率の計算は確率論の原点であると思う。確率論には様々なモデルがあり多様な個別性をたのしむことができる。新しい確率モデルの発見は確率論の研究の新分野をひらくものである。確率論の数学的基礎を与えるKolmogorov（1935）の公理系はよくしられており、現在は厳密な数学的な議論も展開されている。

水に浮かべた花粉のこきざみな不規則運動はBrown運動として知られている。Einstein（1905）は思考実験からBrown運動を考え、その確率モデルを考えた（アインシュタイン選集1、共立出版）。その研究は実験にもとづいたPerrin（1908）によるAvogadro数の計算にもちいられ当時問題になっていた原子論への有力な根拠をあたえた。多くのすぐれた研究者の貢献により、Brown運動は数学の研究対象としても厳密に議論されるようになった。硬貨を投げておもてが出れば正の方向に1歩、裏が出れば負の方向に1歩進むランダムウォークを連続化した確率モデルがBrown運動であると考えることができる。Brown運動の確率モデルと同等な確率モデルはEinstein以前にBachelier（1900）により株価の変動の理解のためにもちいられていたことを最近知った（楠岡成雄（2001）、数学通信、第6巻、2号）。Bachelierの著作のタイトルは投機の理論であったそうである。Brown運動という基本的な確率モデルは自然現象、経済現象の理解のための努力から発見されたようである。現象は数理に基本的な問題を提起し、数理的な研究からえられた成果は、さらに現象の理解にもちいられると言う例は多いであろう。Bachelier以前にも確率論において賭博の問題はふるくから議論され破産の問題は確率論の教科書にのべられている。金融工学の分野で確率論は基本的なものとなっている

ようである。Brown 運動を拡張した Lévy 過程が経済現象の理解にもちいられようとしている。Lévy 過程をもちいることにより Brown 運動では理解できない不連続なおおきな確率的変化もモデル化することができる。Brown 運動ほど基本的でなくても、確率論には様々な確率モデルがある。集団遺伝学において基本的である Fisher-Wright- モデルは数理的にも興味深いものである。物理学、生物学、工学などの各分野において固有の確率モデルが発見され、用いられている。まだだれも足を踏み入れてない分野に足を踏み入れることが新分野の開拓であり、既成分野のなかにある小さな問題を発端として新分野はでてくることも多いのではないかと思う。経済的にゆたかになりたいという欲求は人類社会でのふるくからのものであり、Bachelier は株価の変動の研究からとてつもない新分野に足をふみれた。Bachelier の研究も株式市場というものの発案がなければでてこなかったものである。経済活動以外にも選挙、言語、婚姻規則等人間の活動についての多くの魅力的な確率論の問題がある。株式市場は Bachelier の研究を生んだが、選挙制度も確率モデルと言う立場からも興味深いものである。人間社会の様々な制度、科学技術の変化は様々な確率モデルをうみだしてゆくと思われる。Bachelier や Einstein ほどでない研究者でも運よく面白い課題にであえば新しい数理の分野をひらけるかもしれないと思う。人間の活動を確率モデルとして理解しようといままで試みてきた例についてのべてみたい。

2. 選挙の確率論

選挙の結果は、誰の目にも触れる最も身近なデータである。各候補者は多くの資金、組織および時間をかけしのぎを削る。各候補者は大変な努力をするわけであるが、結果として思わぬ統計的法則に支配されてしまう。以前に日本で行われていた国会議員選挙における中選挙区制のデータについて票割りの確率モデルという視点から考えてみたい。

2.1. 票割りが行われない場合

票割りが行なわれない場合について、数理生態学においてよく用いられている考え方を適用する。ある地域に n 種の鳥がいるとすれば、おのおのの種の個体数はランダムに n 個に分割された棒の長さのいずれかに対応している。これは元村

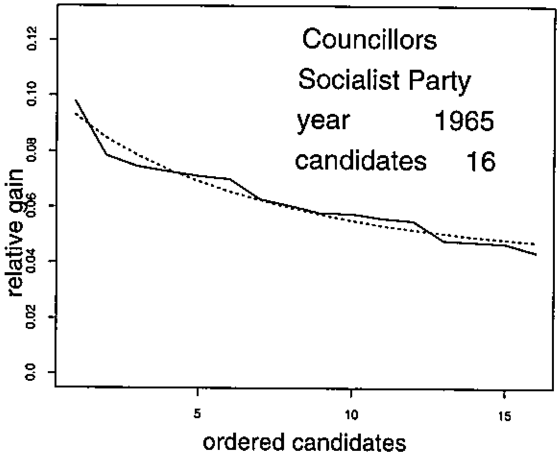

図 1.：1965 年社会党得票率分布

勲（1932)、篠崎吉郎、浦田直美（1953）らにより発展させられた考え方であり、R. H. MacArthur（1957）により明確に述べられ多くの人に知られるようになった。青山博次郎（1962）は同様な考え方をに日本の選挙に適用している。分割された棒を大きさの順にならべ i 番目の棒の期待値を計算すると

$$\frac{1}{m}\sum_{r=1}^{m-i+1}\frac{1}{m-r+1},\quad i=1,\cdots,m, \tag{1}$$

となる。

2.2. 選挙の票割りと路上駐車の類似性

選挙の票割りはどのように行なわれるのか。ここで問題にしているのは、どの候補者が一番多くの票を集めるかということではない。たとえば 1965 年の参議院全国区の社会党の得票率分布はどのような確率モデルで説明できるのだろうか。1 次元ランダムパッキングモデルによる理解を試みる（図 1)。

Bernal の考えたランダムパッキングの 1 次元のモデルとして路上駐車の問題が知られている。路上駐車の問題は、1 次元ランダムパッキングともいわれ、1958 年にハンガリーの A. Rényi により確率論的に議論された問題である。J. P. Flory は高分子化学の問題に関連して類似な問題を 1939 年に考えている。

いまある街路の片側に、A から B までのひと続きの駐車可能の区間があるとす

る。初めは全体があいている。そこへ1台の自動車が来て駐車するとする。Aから B までの間に納まる範囲にランダムに駐車するとし、その位置を PQ とする。2台目は A から P まで、または Q から B までの間に納まる範囲でランダムに駐車する。以下同様にして、次々と駐車していく。やがてどこにも駐車するすきまがないという状態になる。Rényi は「与えられた長さの駐車場に何台駐車できるか」という問題について解析的な結果を得た。

このような路上駐車の問題と選挙の票割りとどのような関係があるのであろうか。いま政党 A が支持票の票割りを行なっていくとする。支持票の総数は道の長さに対応し、車の駐車に必要なスペースというのは当選に最低必要な票の数に対応する。駐車された車は、政党 A により公認された候補者に対応し、政党 A は駐車可能なスペースがなくなるまで候補者をつめていくのである。隣り合う車の先端と先端の距離は各候補者の得票に対応すると考える。次の節において計算機実験を行ないデータと比較することにする。

2.3. 票割りの計算機実験

次のような実験を行なう。ここで円周上のランダムパッキングを考えるのは、道路の両端の影響を取り除くためである。

円周上に x 個の点 $0,1,2,\cdots,x-1$ が等間隔に並んでいるものとする。点 x は点 0 と一致し、一般に点 $x+y$ は点 y に一致するものとする。ある選挙区において政党 A の票が 1 から x まで番号づけられていたとする。政党 A は d 票の地盤を確保する能力のある候補者を次のようなやり方で公認していく。立候補希望者を優先順位により $C_1, C_2, \cdots\cdots$ と並べる。立候補希望者 $C_1, C_2 \cdots\cdots$ の地盤は円周上におのおの並んでいて、$[x_1, x_1+d), [x_2, x_2+d), \cdots\cdots$ のような区間になっているとする。ここで $[x_1, x_1+d)$ とは、x_1 から x_1+d-1 までの点を意味する。また $x_1, x_2, \cdots\cdots$ は乱数表から読み取られた数であるものとする。たとえば $x=8,000,000$ とすれば、7桁ずつ 8,000,000 未満の数字を読んでいけばよい。

政党 A はまず C_1 を公認する。C_k, $k \geq 2$ はすでに公認された候補者と地盤が重ならない場合に公認される。それ以外の場合は公認されない。このようにして、長さ d 以上のすきまがなくなるまで希望者を公認していく。その結果、M 人の候補者が公認されたとする。M 人の候補者の地盤を

$$[z_1, z_1+d), [z_2, z_2+d), \cdots\cdots, [z_M, z_M+d)$$

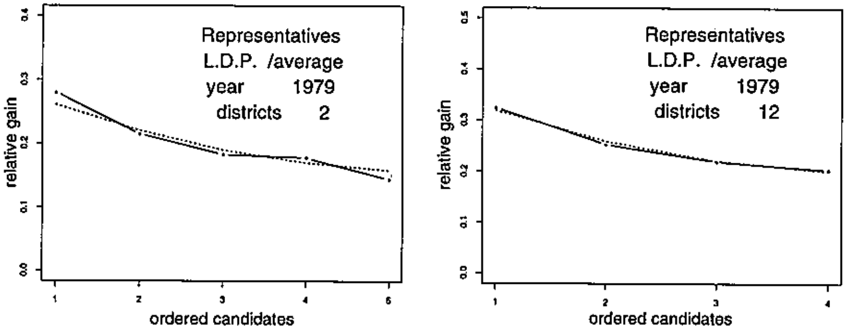

図 2.：1979 年衆議院選得票率分布

とする。$z_1, z_2, \cdots\cdots, z_M$ により円周上の点を M 個の組に分ける。各組に含まれる点の数を各候補者の得票に対応するものとし、これらを大きさの順に並べたものを次式とする。

$$Z_1 \geq Z_2 \geq \cdots\cdots \geq Z_M$$

$x \geq 4d$ の場合、上記の 1 次元ランダムパッキングについての Rényi の結果から、dM/x の期待値は約 0.748 とみなしてよい。データにあるのは政党 A により公認された候補者の数 m であり、d を $dm/x = 0.748$ により定めるものとする。$M = m$ であるような実験だけを選び出し、これらの実験により $Z_1, Z_2, \cdots\cdots, Z_m$ のおのおのについての平均値を求める。ここでは $M = m$ となる 10,000 回の実験により平均値 $E(Z_1), E(Z_2), \cdots\cdots, E(Z_M)$ を定める。

参議院選挙全国区 1965 年の社会党の結果を図 1 に示す。実線がデータを表わし、点線が $E(Z_1), E(Z_2), \cdots\cdots, E(Z_{16})$ の値を表わす。初めに述べたとおり、票割りは路上駐車のように行なわれたように見えるのである。上記 d, x は、ある程度大きければそれぞれの値は問題ではなく、d/x の値が重要であり、計算機実験では $md/x = 0.748$ を満たすように適当に d, x を定めてある。(Y. Itoh and S. Ueda "A random packing model for election" Ann. Inst. Statist. Math. Vol.31, 157–167 (1978)。

図 2 に 1979 年の衆議院選挙の得票率分布を示す。この年は 5 名の候補者が公認された選挙区が 2、4 名の候補者が公認された選挙区が 12 あった。実線がデータの平均値であり、点線が上に述べたシミュレーションの平均値である。

以上の場合、公認された候補者の数 m という 1 つだけのパラメータをもちいて

いる。適合度をよくするために調節したパラメータはない。したがって、これ以上単純な説明はありえない。各政党は票割りに努力し、公認された候補者はおのおの努力をするわけであるが、結果は思わぬ統計的法則に支配されてしまったということになる。このことは、逆に政党も候補者も与えられた条件で最大の努力をしたということを意味する。

政党の組織の状態により、票割りの能力はことなる。公明党、共産党は票割りの能力が高く、各候補者へほぼ均等に配分する。この場合、ランダムパッキングとは異なったやり方で票割りが行なわれている。

また時代によっても異なる。1947 年の社会党はランダムな棒の分割、1953 年から 1968 年の社会党はランダムパッキング、1971 年以後は公明党、共産党の票割りの方法に近くなってきている。自民党の場合、1965 年から 1974 年まではランダムパッキングである。1977 年は公明党、共産党の票割の仕方に似た図がでてきている。政党の支持率等の状況により、票割りの能力、方法は変わってくるわけである。上に述べた $E(Z_1), E(Z_2), \cdots\cdots, E(Z_m)$ の値は、票割りについて議論する上での客観的基準を与えているのである。

2.4. 参議院選挙における相転移

空間に球をランダムに配置していくランダムパッキングにより Bernal (1959) は液体の構造を説明することを試みた。鳥のなわばりの形は、2 次元ランダムパッキングにより説明されている (M. Tanemura and M. Hasegawa, Geometrical models of territory, J. Theor. Biol. 82 477–496 (1980))。参議院議員選挙の方法は 1983 年におおきくかわったが、第 2 次大戦後 1947 年以来おこなわれた選挙について考えてみたい。Bernal はランダムパッキングにより液体の構造を議論したが、選挙の票割りに物理学的イメージを与えることもできる。現在は存在しない政党であるが、たとえば社会党の参議院全国区について考える。1947 年から調べてみると、1947 年は棒のランダムな分割、1950 年は棒のランダムな分割とランダムパッキングの中間、1953 年から 1968 年はランダムパッキングとなる。1950 年ごろに相転移が起きたとみることができる。第 2 次世界大戦後の社会が安定化してゆく時期に対応しているものと考えられる。参議院選挙は 1983 年に比例代表制になったがこれは当時の 2 大政党であった自由民主党と日本社会党の双方に不利でない制度であった。すなわち総得票数にほぼ比例して各政党は議席をえられる制度である。

これは固体への相転移が起きたとたとえることができる。社会現象にも物理現象と同様に、相転移という現象があるのである (Y. Itoh and S. Ueda, Observed phase changes and random sequential packing model for elections in Japan, Physica D 142, 29–40 (2000))。

3. オーストラリアアボリジニの婚姻規則と確率モデル

オーストラリアアボリジニの婚姻規則について Morgan, Lévi-Strausse による研究が知られている。このような婚姻規則についてのポピュレーションダイナミックスはどうなるかということに数理という視点から興味を持った。オーストラリアアボリジニの婚姻規則については代数学的な視点からの研究がしられている。婚姻規則の安定性という問題すなわち半族間、系譜世代間の人数にかたよりがあった場合に、現実の婚姻はどのように行なわれるのであろうか。ランダムなであいということを仮定し確率モデルを考えると、複雑な婚姻規則は安定に存続しにくいのではないかという結論に導かれる。(伊藤栄明、統計数理研究所彙報、30、55–65 (1983))。人口がそれほど多くない集団において複雑な婚姻規則を維持するのは困難なのではないかと考えていたのであるが、この問題についてのフィールドワークが行なわれていることを知った (杉藤重信、民族学博物館研究報告別冊 15 号、251–275 (1991))。以下「」で囲まれた部分は「オーストラリアアボリジニー」(小山、松山、窪田、久保、杉藤、杉藤、松本編、産経新聞大阪本社 (1992)) 杉藤重信著部分から抜粋したものである。同論文にもとづいて、オーストラリアアボリジニの婚姻規則についてのべてみよう。「アボリジニの組織原理の分布を見てみると、むしろ厳しい環境の地域に複雑な組織をもつものが多く、逆に豊かな環境のところほど組織が単純であることに気づく。そこで、人口と婚姻規則の関係について、コンピュータシミュレーションしてみると複雑な婚姻規則ほど人口を減少させてしまうことがわかった。」次にのべるように、これら婚姻規則についてのポピュレーションダイナミックスを考えると一旦かたよりが生じた場合それをもとにもどすという構造になっていないことに気づく。これは「複雑な婚姻規則ほど人口を減少させる」という考え方を支持するものであると思う。

「ヨロンゴ (アーネムランドのガリウィンクでのアボリジニーの自称) を例にして彼らの人間の分類の枠組みをみていこう。ヨロンゴはまずイリチャとドワと

名づけられた半族に2分される。子供の半族は父親の半族とおなじである。さらに、それぞれの半族は4つの集団に分けられる。現地名ではマルク、英語名ではスキン、社会人類学ではサブセクションとよぶ。この4つのサブセクションは順に系譜がたどれるので、ある意味では世代に似ている。」「生まれた子供のサブセクションは、母がどのサブセクションかによって自動的に決まる。」

記号的にあらわすとこれらの規則は次のようになる。I はイリチャ、D はドア、m は男性、f は女性を表わすものとする。数字は世代に似た概念であるサブセクションをあらわす。

つぎのモデルはこれらのオーストラリアアボリジニの社会における規則を理想化したものである（伊藤栄明、統計数理研究所彙報 (1983)、厳佐庸、数理科学、1986年10月号、K. Tainaka and Y. Itoh, Physics Letters A 187, 49–53 (1994))。

男性 $I_{m,i}$ は女性 $D_{f,i}$ と結婚し、彼らの子供の所属するグループはつぎのような規則によるものとする。1人の $I_{m,i}$ と1人の $D_{f,i}$ の結婚により2人の $I_{m,i+1}$、あるいは1人の $I_{m,i+1}$ と1人の $I_{f,i+1}$、あるいは1人の $I_{f,i+1}$ と1人の $I_{m,i+1}$、あるいは2人の $I_{f,i+1}$ に変化する。すなわち

$$I_{m,i}+D_{f,i} \rightarrow \begin{cases} 2I_{m,i+1} \\ I_{m,i+1}+I_{f,i+1} \\ I_{f,i+1}+I_{m,i+1} \\ 2I_{f,i+1} \end{cases}$$

これらはそれぞれ1/4の確率をもつ。この規則を

$$I_{m,i}+D_{f,i} \rightarrow I_{m,i+1}+I_{f,i+1}$$

同様に

$$D_{m,i}+I_{f,i} \rightarrow D_{m,i+1}+D_{f,i+1}$$

とあらわすことにする。サブセクションの数が4であれば i は4ごとにもとにもどる。すなわち4を法として考える。この婚姻規則はイリチャの人数とドアの人数にちがいがある場合、あるいは対応する同世代の人数にちがいがある場合はうまく機能しないことはあきらかであろう。実際にはこれらの規則が厳密に行なわれているわけではなく遵守率（杉藤重信、民族学博物館研究報告別冊15号、251–275 (1991)）というものを考える。実際、第一選択、第二選択ということが行なわれ、例えば男性 $I_{m,i}$ は女性 $D_{f,i}$ と結婚することが多いが、女性 $D_{f,i+2}$ と結婚することが次に多い。イリチャとドアの互いに対応するサブセクションの間の人数の違

いが顕著になった場合は規則をゆるくする等のことが現実には行なわれているのではないかということが報告されている。地理的な構造を考えると通婚圏を考える必要があり、規則が機能するには通婚圏をひろくとる必要がある。

この規則の起源というのは興味があるところである。近親婚をさけるというのもひとつの理由だと思うが、交差いとこ婚を拡大解釈するシステムと考えることもできるのではないかと考えている。婚姻規則が成立する過程について確率モデルという視点からの理解を試みたいと思っている。

4. 古代社会の親族ネットワーク

民族学博物館における共同研究「数理民族学：その応用的研究」（代表者小山修三）に参加することから、古代メソポタミアのNuziの荘園の粘土版に記された名前のデータが一冊の書物、Nuzi人名資料、にまとめられていることを知った(Nuzi personal names, I. F. Gelb, P. M. Purves and A. A. Macrae, The Chicago University Press 1943)。「Nuziはイラク北東部キルクーク近郊ヨルガンテペにある青銅器時代の都市遺跡である。」「Nuziはアラプハという小国の一都市であり、その人口の大半がフルリ人という北方起源の民族であった。」「Nuziが成立した年代については不明であるが、当時の強国であるミタンニ王サウシャッタルの書簡が出土したことから、遅くともそれは前15世紀と推定される。」「Nuziの存続期間はアッシリアによって破壊されるまでの約一世紀であるが、この間にかかれ現在まで残された粘土版文書は約4000枚である。この多くが私宅に保存されていた個人的な契約書であり、その内容から当時の経済活動や親族関係等についてうかがい知ることができる。」（牧野久実、史学第60巻（1991）より抜粋）。Nuzi人名資料にある人名はほとんど男性の名前であるが、たとえばプヒセンニの子テヒプテイラという形に個人の名前が記されている。人名の表記が一通りではないこと、同名の人間が複数いること等、の問題があるが、このデータをもとにNuziの荘園の人口を推定することはできないかと考えた。このためにヌジ人名資料のデータベースを作成し、解析を行なった。（文部省科学研究費重点領域研究、系図データからの古代社会の人口推定、伊藤栄明、石黒真木夫、上田澄江、牧野久実（1998））単に人口の推定だけでなく親族関係の推定、社会に有力者が形成されてゆく過程等の問題についてもこれの活用が可能であると思われる。人口の推定はまだ行って

いないが、Nuzi 人名資料から系図を自動的に作成することを行った。人間社会において各個人は自分を中心とした親族関係を把握している。各個人の親族関係がつながってどのようなネットワークを構成しているかということについては日常あまり考えない。親族関係のないと思っていた知人との間に以外な親族関係を発見するということは世の中に時々あることである。二つの家系を統合するという系図の作成は親族関係についてのオーストラリアアボリジニの婚姻規則のフィールドワークにおいて問題になり、野外調査のためのソフトの作成も行われている(文部省科学研究費重点領域研究、文化人類学における家族、親族領域を中心として　フィールドデータの処理と分析、田中雅一、杉藤重信、窪田幸子 (1998))。

Nuzi 人名資料から Nuzi の古代社会における家系図を自動的に復元することが計算機の能力の飛躍的な向上により可能になった。計算機をもちいることにより古代社会における親族関係がうかびあがるのである。

粘土版文書はほぼ一世紀 5 世代にわたっているとみられ、Nuzi 人名資料はこの文書を人名によって分類した索引であり、同一人物とみなされる人の情報をひとまとめとして記述したものである。そこからは人名にかかわる親族関係、内容をしるした文献名その巻数行番号などが参照できる。Nuzi 人名資料は個人名の索引であるために家系図の観点からは重複して記載される。たとえば 3 人家族であれば 3 つの名前の箇所にそれぞれ記される。このようなデータであることを考慮し、家系図の復元を次のような手順で行った。

1. Nuzi 名の表記を統一し、同一人物が複数回別の表記で記載されないようにする。
2. 名前ごとの記述から、家系ごとの記述にする。家系に番号をつける。
3. 各家系について番号づけられたすべての家系を 1 番から番号順に比較し、統合できるか否かを判定する。2 つの家系間で 2 人の名前が一致し、その親族関係に矛盾がなければ同一家系とみなすことにし、2 つの家系を番号の若い方に統合する。これを最終の番号までつづける。結果にもとづき番号をつけなおす。
4. 統合される家系がなくなるまで 3. をくりかえす。

このアルゴリズムにより作成された家系図の多くはすでにほかの情報をもちいてしられていたものも多いが、おもいもよらない家系図を発見することもできた。図 3 は上記のアルゴリズムにより作成したものであるが考古学者が他の情報をも

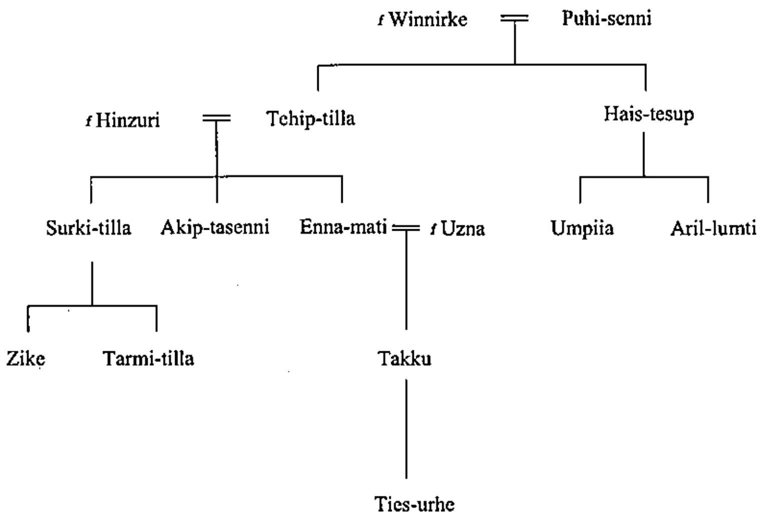

図 3.：テヒプティラの家系図

ちいて復元したものとよく一致している。

5. 語順規則の Ising モデル

語順規則は J. Greenberg により法則のかたちにまとめられ、言語類型論として知られている。角田太作（世界の言語と日本語、くろしお出版、(1991)）は世界の 130 の言語について次に示す 19 個の項目についての語順規則を表にまとめた。例えば、SOV の語順をもつ言語の多くは後置詞をもち、後置詞をもつならば所有格は名詞の前にくる。SVO の語順をもつ言語の多くは前置詞をもち、前置詞をもつ言語の所有格は名詞の後に位置することが多い。前者は日本語に近い語順をもつ言語の特徴であり、後者は日本語と逆の語順を示すものである。韓国語、ベンガル語が前者に属し、英語、タイ語は後者にあたる。階層クラスタ分析を用いると世界の言語は語順によって概ね 2 つのタイプに分類でき、この分類は、「名詞と側置詞」の順序によって説明される (Tsunoda, T., Ueda, S., and Itoh, Y. Linguistics, 33, 741–761 (1994))。

1.	S, O と V	私は本を買った（主語、目的語、動詞の順）
2.	名詞と側置詞	東京へ
3.	所有格と名詞	私の家
4.	指示詞と名詞	この家
5.	数詞と名詞	3 人の子供
6.	形容詞と名詞	大きい家
7.	関係節と名詞	昨日読んだ本
8.	固有名詞と普通名詞	名古屋大学
9.	比較の表現	この本はあの本より難しい
10.	本動詞と助動詞	食べている
11.	副詞と動詞	いつも速く歩く（副詞は動詞より前）
12.	副詞と形容詞	とてもきれい
13.	疑問の印	読みましたか
14.	一般疑問文での S, V 倒置	あれは富士山ですか？（倒置無し） Is that Mt. Fuji?　（英語は倒置あり）
15.	疑問詞	あなたは昨日どこへ行きましたか？（平序文式） Where did you go yesterday?（英語は文頭）
16.	特殊疑問分での S, V 倒置	あれは何ですか？（倒置無し） What is that?　（英語は倒置あり）
17.	否定の印	読まない（否定の印は動詞語尾）
18.	条件節と主節	もし明子が来れば、花子は行きます。
19.	目的節と主節	花子が出かけられるように、明子は家にいた。

アジア・ヨーロッパ地域の言語の角田による語順の表について最長距離法による階層クラスタ分析による結果が図 4 である。これらは、ヨーロッパ諸言語について自然な結果を与えていると思われる。一方、ヒンヅー語、ベンガル語、パンジャブ語等の北インドの言語が、ヨーロッパ諸言語から離れて日本語の近くに分類されることは興味深い。

語順規則の変化を時間的、地理的な確率過程としてとらえることは自然なことであると思われる。確率モデルとしてとらえるという視点にたてば、言語が分裂、

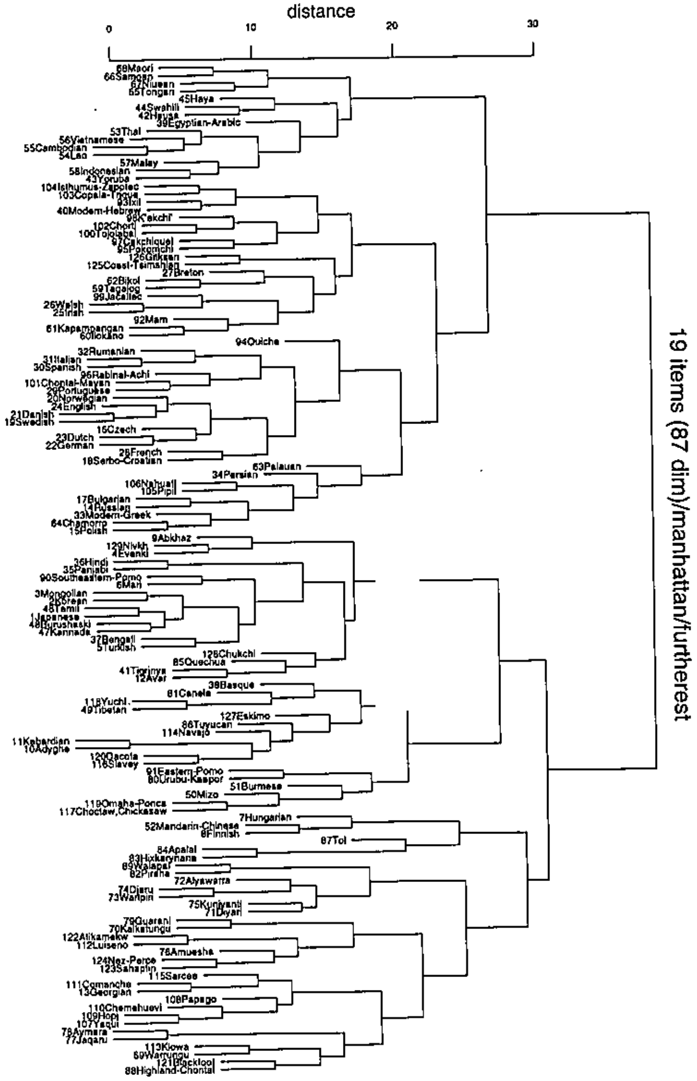

図 4.：語順表の階層クラスタ分析

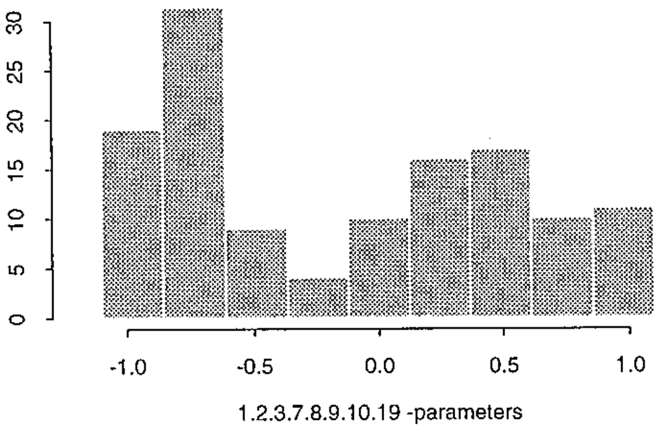

図 5.：8 項目の度数分布

統合、相互作用を繰り返しながら変化してゆき、たまたま語順が非常に近い対が現れたという解釈もありうる。例えば、角田による語順表においてタミル語と日本語の間の違いは、イタリア語とスペイン語ほどに近い。このことは偶然の結果であるのか、それとも別の観点からの議論から必然性が示せるのか、興味のあるところである。日本語、タミール語、朝鮮語、蒙古語の相互の近さについて比較言語学からの研究も知られている（藤原明、日本語はどこから来たか、講談社現代新書（1981））。語順規則について集団遺伝学の場合のように中立説を仮定すれば、語順規則から言語の系統を考えることができると思われるが、各言語は地球の表面に分布しており、相互に影響し合っている。生物種とちがい系統というとらえ方だけではとらえきれないことはあきらかであろう。DNA 列を解析し生物種の系統を考える際に、変化の速い部分とおそい部分を区別して考えないと混乱した結果がえられる。同様な理由から、言語の場合語順規則と基礎語彙の両方を同時に考えるということでなく、語順規則のみにより、階層クラスタリング（図 4）を考えるということにより立場がより明確になるように思われる。名詞と側置詞の項目およびそれと関連の強い項目、全部で 8 個の上記の項目、1, 2, 3, 7, 8, 9, 10, 19 を考え、日本語とおなじ場合を+1 逆の場合を −1 とする。現実には+1、−1 といいきれない中間的なものもあるがそれらは、+1 と −1 の間の値に数値化し、各言語について 8 個の項目についてのそれらの数値の算術平均を考える。図 5 はそれについての度数分布をあたえたものである。

これを説明する確率モデルとして多数決モデルというのはどうかと考えている。プラスとマイナスふたつの箱があり、それぞれに n_+、n_- 個の粒子がはいっていたとする。総粒子数を n 個とし、各粒子には1から n まで番号がふってある。各時点でランダムに3個の粒子を指定し、3粒子のうちの2個がプラスと1個がマイナスにはいっていたとすれば指定されたマイナスの1粒子が箱プラスに移るものとする。同様に3粒子のうちの2個がマイナスと1個がプラスにはいっていたとすれば指定されたプラスの1粒子が箱マイナスに移るものとする。3個とも同じものの組みあわせである場合移動はおこらないとする。この操作を繰り返していくとき、それぞれの箱の粒子数はどのように変動するのであろうか。あきらかに最終的にはどちらかの箱に全部の粒子が集まるはずである。このモデルはランダムにであった3粒子が、多数決のルールにより変化するというモデルと考えることができる。さらに各粒子はプラスからマイナスへある確率で突然変異するとし、マイナスからプラスへもある確率で突然変異すると仮定することにより図5において観察されたような2個の山のある度数分布がえられるが、このような確率モデルにより語順規則の変化を説明できるのではないかと思う。日本語と同じ語順を+1、逆の語順を −1 とあらわしたが、各項目の変化は独立に起きるのではなく、関連する項目のプラス、マイナスとの相互作用で変化するはずであり多数決モデルというのは自然であるように思う。さきにのべた、SOV の語順をもつ言語の多くは後置詞をもち、後置詞をもつならば所有格は名詞の前にくる。SVO の語順をもつ言語の多くは前置詞をもち、前置詞をもつ言語の所有格は名詞の後に位置することが多い、等からこのことは納得できるであろう。非常に単純化したモデルであるが、例えばこのような多数決による確率過程で語順規則の変化を理解できないかと考えている。上記多数決モデルは磁性体のイジングモデルといわれているものに類似なモデルであり、語順規則のイジングモデルとでもいってよいものである。このモデルによりデータを理解できるということは、上記8項目の全部がプラスあるいは全部がマイナスという状態が安定であるということをある意味で示すものかもしれない。