

第148回 STF 交流会報告

1. 日時： 2024年6月29日（土）14：20-6：50
2. 場所：きゅりあん 4階 第1G活動室
3. 参加者 10名（会場参加のみ）

●テーマ1. 「DNA が語る古代ヤポネシア」 後藤幸子

日経サイエンス 2024 月 2 月号記事が、出典。

1万6000年前から3000年前まで、日本列島で縄文文化を支えた縄文人の遺伝子は、今どこに？
縄文人集団と大陸から渡来した東アジア人集団が混血して現在の日本人集団になった。

今も日本人集団の中に存在している縄文人のかけらを現代日本人ゲノムから、抽出し、それらが示す「縄文人度合い」の地域差から、日本人集団が成立した過程がよりはっきり理解できるだろうと考えた。

大森貝塚が発見され、縄文土器や人骨が発見された際に、縄文人、縄文時代の名称がついた。ただし、この文化を担った人々は、現代日本人とは関係しない先住民族とみられていた。その後、縄文人と近隣集団との「混血説」、縄文人の小さな形態進化で現代人に移行した「進化説」があった。

1980年 形態人類学者の埴原和郎氏が千人分以上の現代日本人の頭骨を計測、多変量統計解析して「日本人の頭骨には地域的勾配がある」という特徴を確認した。日本人集団の形成は、縄文人と渡来人という2つの集団が共存、融合することで成立したという「二重構造モデル」が1991年に提唱され、以来これが1つの定説になっている。我々もそう習った。

形質だけで日本人集団の成立モデルが作られたが、今回は日本の現代人ゲノムから、縄文人に由来するとみられる変異を特定し、それに基づいた「縄文人度合い」を調査して日本の各地域の差異を調査した。日本人集団は、縄文人と渡来人の混血により生まれてくるが、その混血の程度には地域差がある。縄文人度合いが低い四国、近畿では、人口増加率、稲作の到来が早かった。縄文人度合いの濃い東北、関東では、人口増加率は低く、稲作の到来も遅かった。

DNA 解析で支持される人の移動ルートと、考古学で支持される稲作、畑作の伝来ルートは一致しない。今後、長江周辺の古代人ゲノム解析が進めば、伝来ルート判明の鍵となる。

お酒に弱いのは、渡来人由来。

混血はどこで起きたか？ 従来の想定は大陸からやってきた均一な渡来人集団が日本列島に渡来したであった。今回の朝鮮半島南端の古代人ゲノム解析からは、新石器時代から多様な混血状態の人がいた（縄文系と大陸系）ことが分かった。大陸系集団と在来系縄文集団との混血は、日本列島に来る前から、すでに始まっていたかもしれない。

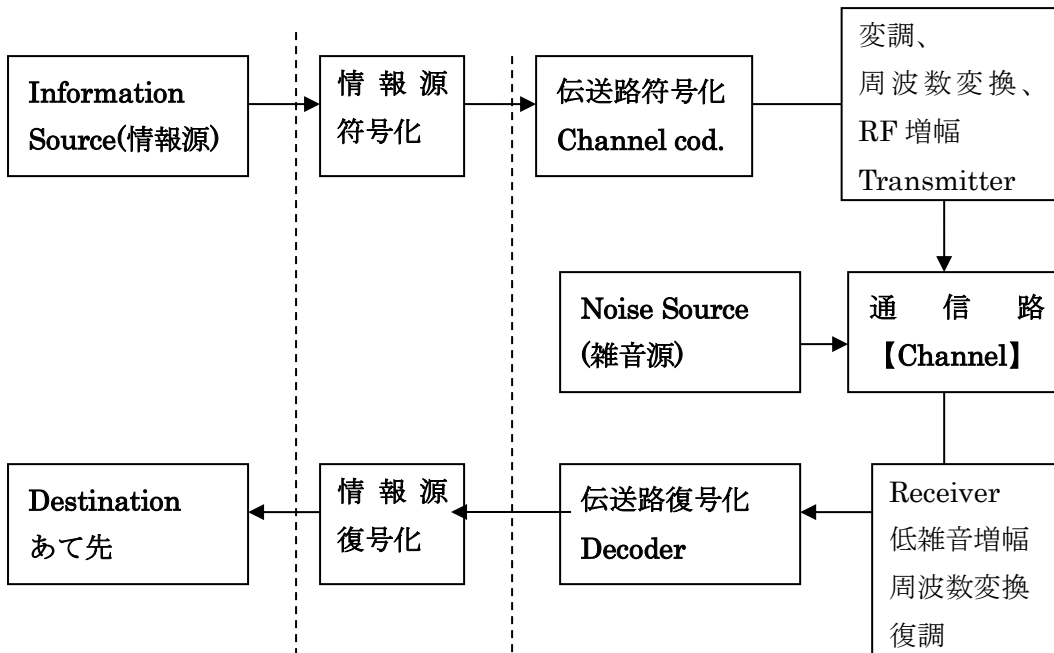
まず現在の情報通信技術の進歩を振り返った。

1948 ベル研究所紀要(Bell System Technical Journal)

- ◇ トランジスタの発明
- ◇ C.E.Shanon の通信の数学的理論
- ◇ 集積回路(IC)の発明と集積度の天文学的増大
- ◇ 電卓開発競争から生まれたマイコン(micro computer)
- ◇ マイコンから生まれたパソコン
- ◇ 通信網は電話網(回線交換)からデータ通信網(packet 交換)へ
- ◇ 異なるデータ通信網を相互に結ぶ通信網としての Internet の発展

上のように 1948 年は現在の情報化社会の技術的源となった年である。今回は特に C.E.Shannon の通信の数学的理論の概要を解説した。

1. 通信回線



文字、音、絵、動画 二進符号列
(電報の場合、dot / dash)

情報通信回線の一般構成

ここで特筆すべきは 1948 年当時の通信網とは圧倒的に電話網であった事である。それに対して本論文の対象はデータ通信である。当時のデータ通信とは電信であり、その通信路符号はモールス符号であった。それに対して本論文は二進符号化を扱っている。それは情報量を可能な通報数の対数としていて定義した事に基づいている。対数の底に 2 を採ると情報の単位は binary unit 即ち bit となる。

2 情報量と情報エントロピー Information Quantity and Entropy

情報源, Information source

符号(文字)の集合 $\{ L_i ; i = 1, 2, \dots, n \}$

符号 L_i の生起確率 ; P_i

当然 $\sum P_i = 1$

情報量の定義

ある通報(文) S の運ぶ情報量 $I(S)$ の定義

$$I(S) = \text{Log}(1 / P(S)) = -\text{Log}(P(S))$$

$P(S)$; 通報 S の出現確率

これは通報の「意味」は全く捨象して事象の生起確率のみで定義される抽象的な定義である。

$$P(S) = 1 \quad \rightarrow \quad I(S) = 0$$

$$P(S) > P(T) \quad \rightarrow \quad I(S) < I(T)$$

即ち意外性の高い通報ほど情報量が大である事は直感的にも合理的である。

長さ N 文字から成る通報(文) S の運ぶ情報量

N が非常に大きい場合にはその文の中に含まれる各文字 L_i の数は典型的に $N \cdot P_i$ ($i = 1, 2, \dots, n$)

そうでない非典型的な文の出現確率は N が大きくなると急速に小さくなるので無視できる。

上の典型的な文の数 $M(N)$ は

$$M(N) = N! / \{ (N \cdot P_1)! \cdot (N \cdot P_2)! \dots (N \cdot P_n)! \}$$

これらの文は等確率で生起するので、通報 S の運ぶ情報量 $I(S)$ は

$$I(S) = \text{Log}(1 / M(N)) = N \cdot H$$

ここで

$$H = - \sum_{i=1, n} P_i \cdot \text{Log}(P_i)$$

これは次のように情報量の性質にかなっている。

- (1) ある文の伝える情報量はその文の長さに比例する。
- (2) H は上の定義式から分かるように一文字当たりの平均情報量と解釈できる。
 N 文字から成る文の運ぶ情報量が $N \cdot H$ で与えられる。

情報源の Entropy

上の H は情報源から生起する符号が運ぶ平均情報量である。

$$H = - \sum_{i=1, n} P_i \cdot \text{Log}(P_i) \quad \sum_{i=1, n} P_i = 1$$

Shannon は上の H を情報エントロピーと名づけた。

3 情報源符号化 Information Source Coding

情報源 S

Symbols; $\{S_i ; i=1,2,\dots,n\}$

Associated Probabilities; $\{P_i ; [i=1,n] \sum P_i = 1\}$

各符号 S_i を二進符号に変換する。

情報源符号化の目的; 生起確率の大きな情報源符号には短い伝送符号を割り当てることによって情報の伝送効率をできるだけ大きくすること。

これは情報 entropy の定義

$$H = [i=1,n] \sum P_i \cdot \text{Log}(1/P_i)$$

からも理解される。 $\text{Log}(1/P_i)$ は確率 P_i が小さい程大きな数になるのでそれに割り当てられるビット数は大きくなる。情報 entropy は符号化回路の出力における平均符号長と見る事もできる。

4 誤りのある通信路を通じた通信路符号化定理

Shannon は誤りの起きる通信路を通じてその通信容量 C (communication capacity, bits/sec)よりも低速 R の通信であれば送信側に於いて時間 T 即ち $RT(\text{bits})$ だけ蓄積して通信路符号化(受信側ではその逆動作、即ち複合化を行う)を行えば、 T を十分大きく設定する事により任意の値以下に通信誤り率を小さくできる事を証明した。

通信容量 Communication Capacity

Shannon は通信容量(伝送路符号化法を工夫して通信路を通じて伝送できる最大の情報量)が

$$C = W \cdot \text{Log} (1 + S / N)$$

で与えられる事を証明した。

ここで W は通信路の周波数帯域幅(Hz), S/N は受信部における信号対雑音電力比である。

結語

通信網と言えば圧倒的に電話網、電信と言えば Morse 符号であった時代に理論的便宜から二進符号を用いて情報理論は確立されたが、現在は実際すべて data 通信は二進符号で行われている。又電話は data 通信網に吸収されてしまっている。

Shannon が通信路符号化定理を証明するのに用いたのは Random 符号化であった。現在実用化されているのは勿論 random 符号化等ではない。抽象代数学に基づく Reed-Solomon 符号と信号間距離の積分を行う最適復号法(Viterbi)が主流である。何れも通信装置の信号処理能力の飛躍的向上に基づき実現可能となったものである。原理が証明されれば実用的な技術は雨後の筍のように開発される。Shannon の「通信の数学的理論」は力学における I.Newton の「自然哲学における数学原理」に匹敵する役割を果たしたと言えるであろう。

意見、所感等：

- ・今回は、仙台にお住まいの方が、所用の機会を利用してご参加くださり、久しぶりにお会いできて話が盛り上がりました。
- ・北方民族の日本への渡来のことも、合わせて論じたら、面白いとの意見がありました。
- ・情報技術の進歩は目覚ましいですが、市吉さんのお話は、大変有用な勉強の機会でした。Shannon の論文「通信の数学的理論」の果たした役割の大きさやその他の情報は、Wikipedia を読むと面白いです。

文責 後藤幸子