



Title	情報理論と言語学〔Ⅰ〕：「情報量」と文の種類
Author(s)	森塚, 文雄
Citation	大阪外大英米研究. 1979, 11, p. 21-35
Version Type	VoR
URL	https://hdl.handle.net/11094/99033
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

情報理論と言語学〔I〕

一 「情報量」と文の種類

森 塚 文 雄

1 ま え が き 敗戦直後の虚脱と廃虚から僅か3分の1世紀足らずにしてフェニックスのように立ち直り、GNP自由世界第2位の経済大国として、今やその“集中豪雨の輸出攻勢”によって米国およびEC諸国の経済に深刻な脅威を与えるに至った、わが国経済力のあざやかな“変身”ぶりは、正に今世紀の一大奇蹟である。「東海道メガロポリス」の象徴する人口の急激な都市集中、矢つぎ早やに更新された超巨大タンカーの屯数記録、新幹線とATCさらにリニアモーターカーの開発、デジタル電子計算機の高速大容量化と急激な普及、衣料生活に革命をもたらした石油化学工業、“交通戦争”と“大気汚染”の元兇モータリゼーション、テレビと電話の電撃的普及、原子力平和利用の旗がしら原子力発電所等々。

わが国における叙上の経済界の超高度成長——もっとも数年前の“石油ショック”の痛手をうけて“減速低成長経済”へと軌道修正を余儀なくされはしたが——と並行して、戦(はん)近のわが国を象徴するものは“情報革命の時代”であると申して差支えないであろう。「情報産業」「情報科学」「情報社会」「情報システム」「情報センター」「情報処理」など、現代はまさに“情報氾濫の時代”と呼ぶにふさわしく、情報に関する記事が新聞、ラジオ、テレビなどマス・メディアを賑わさない日はただの一日も無いと言っても決して過言ではない。戦後の高度成長時代は、未開の新しい科学技術の分野の急速かつ総合的開発を要請し、またその成果が社会と経済の発展を促し、これがまた更に進んだ科学的技術の研究開発投資を誘導するといったフィード・バック(Feedback)——循環——が行われている。このフィード・バックまたは循環を形づくる連鎖が、通信と情報と制御であって、この意味からも現代を“情報社会時代”もしくは“情報革命

時代”と呼ぶことができよう。

およそ10万年以前と推測されている言語の形成は、事実上Homo sapiensの出現を意味するものであって、じらい言語は人類文化の主要な貯蔵手段としてまた人間相互の基本的な連絡手段として役立ってきた、即ち人類社会内における最も重要な情報の保存ならびに伝達的手段であったという点を考慮に入れば、われわれ言語学を志す者にとって近代情報理論についての一応の関心は不可欠であろうと思われる。けだし、「情報局」「情報屋」など芳しくない語感を帯びて用いられた戦前の“情報”とは全く異なって、現代の“情報”には新しい普遍的総合的な概念が包蔵されていて、この新しい学問の分野が「情報理論」であり、「サイバネティックス」であるからである。

本稿では、情報理論に関するいくつかの基本的概念について概略的説明を行ない、とくに「情報量」の見地から英文の分類について私見を加えたいと思う。

2 サイバネティックスと情報理論 最近わが国の数学者、工学者、心理学者、医学者、社会学者などの間で、サイバネティックス (Cybernetics) という言葉が話題に上っているが、この語は「舵手」^①を意味するギリシア語の κυβερνήτης (=Kubernētes) に由来するもので、創始者である米国の数学者 Nobert Wiener (1894-1964) がその名著 *Cybernetics* (1948, '61) の副題として「動物および機械における制御 (Control) と通信 (Communication)」^②と定義づけている通り、機械、生物体および人間社会における情報の生成、伝達、改造、蓄積、利用等に関する一般原理を取り扱う学問である。1919年以降MITのWiener教授の周囲には多数の神経生理学者、通信工学者、計算機研究者が群がっていて、自動機械、生物あるいは社会の如く従来は全く無縁のものとして別々の立場からのみ研究されていたものも、それらの制御機構の基本原理はフィード・バックとか通信などの共通の概念と方法とに基づいて統一的に記述説明が可能であるという確信のもとに、新しい境界領域の学問分野に探求の斧を入れた。情報理論 (Information theory) は、このサイバネティックスの立場に立って、電信、電話、ラジオのような狭義の通信のみならず、機械や動物の制御機構のなかで本質的に重要な役割を演じて

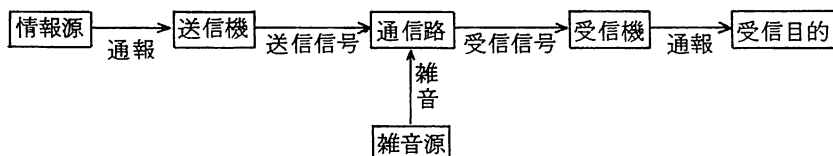
いる広義の通信という問題を研究しようとするもので、最近では数学、工学、医学、心理学から更に社会学、言語学などに至るまで広く各分野の専門家の均しく注目する“稔り豊かな発展の見込みのある”^③ 総合科学の有力な一翼を担おうとしている。そもそも情報理論は、19世紀中葉、英語のアルファベットの26文字の使用頻度を考慮してモールス符号の簡略化・能率化を計った英国の Lord Kelvin (1824-1907) をもって嚆矢(こし) と言われる^④ が、20世紀に入って先ずベル研究所の J.R. Carson と H. Nyquist の周波数帯域幅に関する別別の新しい研究(それぞれ '22年, '24年)があり、次いで H.O. Hartley は Nyquist らの研究の成果を一般化した。Hartley は、後述するように K 種類の符号をもつ符号系で N 個の符号を並べてつくる系列は K^N 通りが可能であり、このような系列の1つを受信したときの情報量 (Amount of information) は、この対数 $H = N \log K$ ととることにより、もっとも合理的に定義できることを示した。第2次大戦から戦後にかけて数学者や物理学者が電気通信技術者と協力して電子工学を中心とする新しい研究に当たった結果、通信の本質を情報伝送形式の広い立場から究明しようとする気運が生じ、又通信を妨害する雑音 (Noise) に関する統計数学的な解析の発達により、雑音が情報伝送において演ずる本質的な役割が理解され、情報理論は雑音をも考慮に入れて、統計数学的に処理されるべしとする認識が生まれた。この間 E.H. Armstrong の発明になる周波数変調 (FM) と並行するパルス技術、とりわけ雑音の妨害排除に絶大の効果をもつパルス符号変調 (PCM) の急速な発達が見られた。戦後 ('45~'48) に入ってから、E.W. Rice はランダム雑音 (Random noise) の確率統計論的な研究を行ない、次いで前述の Wiener や Harvard 大学の生物学者 Rosenbluth らが協力して現代の諸科学の中から通信と自動制御を根底とする分野を総合して新しい学問体系を樹立しようと意図してこれをサイバネティックスと名づけたが、'48年ベル研究所の C.E. Shannon (1916-) がその有名な論文「通信の数学的理論」^⑤ によって現在の情報理論の基礎を確立した。彼は Hartley の定義した情報量を確立過程論を導入して広い立場で定義し、また雑音の重要性を考慮に入れて情報とその通信路による伝送の本筋を捉えた理論を展開し、情報量

として後述のエントロピー (Entropy) とか、通信路の通信容量という新しい概念を導入したが、Shannon の画期的な業績によって情報理論は急速に新たな展開を見、今日では単に情報理論といえば普通 Shannon の理論を指していると見て差支えない。

3 情報理論の諸概念

情報理論では、それが常に電気通信技術と緊密な関係を保ちつつ発展したために、通信工学固有の用語を、本来の意味よりも遙かに拡大して用いている。およそ情報の伝達には、情報ないし通報 (Message) の発生源があってこれを情報源 (Information source) と呼び、これから発生する情報は送信機 (Transmitter) に入って電気信号に変換ないし符号化 (Coding) され、その符号化された情報つまり信号 (Signal) は種々の情報路 (通信路 (Channel)) を媒体として受信機 (Receiver) に入り、ここでもとの通報に復元化 (Decoding) されて受報者 (Receiver ; Destination) に伝わる。一般に通信路には伝送過程において妨害効果を生み出し送受信信号の間に誤差を起す雑音 (Noise) が作用するが、この雑音の有無によって通信路を雑音のある通信路 (Noisy channel) と雑音のない通信路 (Noiseless channel) とに分ける。受信機では、なるべく雑音に起因する誤差を除くようにして通報に復元し、これが受信目的 (Destination) にわたされる。最後に信号には、電信のごとき離散的 (Discrete) なものと、電話のごとき連続的 (Continuous) なものがある。一般に通信系は下図のようにモデル化することができる。

図 1

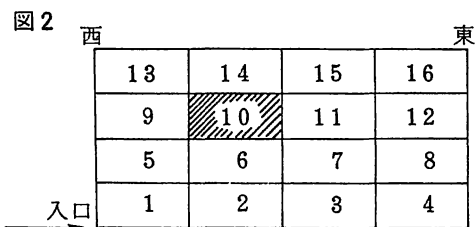


きわめて大まかな言い方をすれば、情報源は、たとえば自然語の 1 つの文字や記号の集合 X とそれに附随した確率分布 (Probability distribution) P_r の対によって表わすことができ、 X が有限のときは離散的、無限 (連続) の

ときは連続的，な情報源という。この情報源とともに情報理論の基本的な概念である情報路は，送信信号の集合 X と受信信号の集合 Y の間に $a \in X \rightarrow b \in Y$ なる写像 (Mapping) あるいは対応 (Correspondence) をつくりだす機構であって，それは X, Y とその間を結ぶ推移確率 (Transition probability) V_x の3要素によって特徴づけられ， $[X, V_x, Y]$ で表わされることがある。

情報理論には以上の外，2進信号の情報の最小単位としてビット (Bit) が，また情報の欠如の程度を表わす量としてのエントロピー (Entropy) や，所与の情報源がどれだけ無駄なものを含むかの度合いを示す冗長度 (Redundancy) などの重要な概念があるが，これらについては次節でやや詳細に述べることにした。

4 情報量，エントロピーと冗長度 情報量は情報理論のもたらした数学的成果のうち，最も重要な概念であって，1928年Hartleyによってはじめて提案されたもので，ごく平易な表現法を用いれば，情報とは同等な可能性をもつ幾つかの事柄の中から特定の1つを指定することであり，情報の量とは，その情報を得た人にとっての内容の豊富さのことである。同一の情報でも，過去の事柄を数多く知っている人よりも全く知らない人の方が，受ける情報の量が多いことは容易にうなずける。英文の通報でQという文字の次にUという文字が来るという情報の量はゼロに近く，inの次にtheがつづくという情報の量はatの次にtheが来るという情報の量より遙かに小さいであろう。下図のような4階建16



室のアパートに住む知人を訪問する人が，入口の管理人から知人の部屋は「10号室です」と教えられた方が「3階です」と教えられた場合より多くの情報量を受けることは自明である。

情報量の最も単純なタイプは，子供のじゃんけんに相当するHead or tail

(裏か表か?)という遊びでコインの表か裏が出る場合とか、サイコロをふって「丁」とでるか「半」とでるかの場合のように、それぞれ $\frac{1}{2}$ の確率をもつ2種の可能性の中から1つを選ぶ“二者択一”の場合であって、その選択によって生じる情報の量を基礎的な単位として設定し、それを二進数字(Binary digit)、略してビット(Bit)^⑥と名づける。上述の1回のコイン投げの結果「表(又は裏)が出た」という情報や、サイコロを1回振って「丁(又は半)が出た」という情報も、産院で父親が看護婦の口から「坊っちゃん(又は嬢ちゃん)です」と教えられたときの情報も、その量はすべて1ビットである。上掲のアパートの場合、訪問者が知人の部屋を全く知らなかった場合、その人に教えるのに、「16室のうちの10号室」と言ってもよく、「3階の西から2番目の部屋」といっても情報の量は同じはずである。後者の教え方では、1階から4階までの4個の可能性の中の3階であるという情報と、3階に4室あるという4個の可能性の中の西から2番目であるという情報との2個の情報が合成されて1個の情報を形づくっている。どちらの考え方をしても内容は同じことであって、その情報量は当然一致していなければならない。そのためには情報量は可能性の数の(2を底とする)対数^⑦で定義するのが便利である。なぜなら情報量として対数をとると $\log_2 16 = \log_2 4 + \log_2 4$ ($4=2+2$)であるから、2つの通報が合成されたときの情報量は、個々の通報の情報量の和で与えられるからである(「10号室です」の情報量は4ビット($\log_2 16=4$)、「3階です」と「西から2番目です」の情報量はそれぞれ2ビット($\log_2 4=2$)。かつてラジオの人気番組であった「二十の扉」でも、20回を限度にくり返されたYes、Noの答で正解に迷った人の得た情報の量は毎回1ビットで合計が20ビット(以内)である。

これまでは、説明の簡単化のために、情報量の定義として、幾つかの同等な可能性の中から1つを指定するという場合についてだけ考えたが、実際にはこの想定が当てはまらない場合が多い。受験生が合格発表つまり合否いずれかの情報によって受ける情報の量について考えても、合否の事前の予想——不確かさの度合い——はフィフティ・フィフティであるとは限らない。受験指導のベテランの高校や予備校の教師が永年の経験から合格ほぼ確実と見込んでいた受験生Aが予想

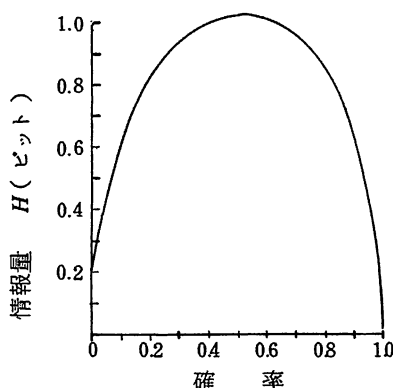
通り合格したという通報から得た情報の量は、及落ボーダーライン上の受験生Bが同様の通報から受けた情報量より遙かに小さく、Aが予想を裏切って落第したという情報の量と比べれば、更に小さいことがわかる。ジャーナリズムの格言に“犬が人に噛みつくのはニュースにならないが、人が犬に噛みつくのは立派なニュースだ”というのがあるが、情報量とは、いわばジャーナリストのニュース・バリューに外ならない。また、英語や日本語のような自然語の1つ1つの文字はその生起の可能性は同等ではなく、多数の十分長い文章や書物について各文字から現われる頻度を調べると、大体一定⁽⁸⁾となるから、その確率(Probability)を定めることができる。一般に或る情報源(たとえば、ある長さの英文)の元(元)(Element)1個あたりの担う平均の情報量を、元来熱力学の用語を用いてエントロピー(H)と呼び、十分長い1つの通報が生起する確率 p の逆数の対数をその通報の長さ n で割ったもので表わし、単位としてビットを用いる。すなわち

$$H = \frac{1}{n} \log \left(\frac{1}{p} \right)$$

エントロピーは、また特定の文字が選ばれる事象(Event)に着目すれば、事象の選択結果の“不確実さ”，つまり情報の欠如、の尺度、従って今後得られなければならない情報量と見ることができ、更に立場を変えて、ある情報源、たとえばアルファベットを構成する n 個(英語では26個)の文字を選択する際の選択の自由の尺度であると解釈することもできる。今、ある情報源のすべての文字が独立に等確率に生じる場合を考えてみよう。文字の種類を玉の色と考えて、色の同じ大きさの玉を同じ割合いで極めて多数、1つの壺の中に入れておき、これを十分かきまぜてから1つずつ取り出すとしよう。2色、たとえば白と黒、の場合と、4色、たとえば白、黒、赤、青の場合を比較すると、2色のとき玉の担っている情報量は1ビットであって4色のときは2ビットである。他方、玉の選択の不確実さは、2色の中から1色を選ぶよりも4色の中から1色を選ぶ方が2倍も大きい。また、玉の色の選択の自由の度合いも、色の多い方が大きいことは日常生活でのわれわれの直観と一致する。次に、色の種類は同じであるが、色に

よって確率が異なり、たとえば白・黒の玉が $p_1 : p_2$ の割合で入れてある壺を考えると、2種の玉の出現する確率の和 $p_1 + p_2 = 1$ であって、このとき白・黒の割合が同じで $p_1 = p_2 = \frac{1}{2}$ の場合エントロピーは最大で1ビットとなる。反対に $p_1 = 1$ または $p_2 = 1$ の場合、つまり全部白または黒の場合はエントロピーは最小でゼロになる。たしかに白・黒が同じ割合のとき色の選択は最も不確実であり、逆に選択の自由度は大きい。このようにエントロピーはわれわれの選択の不確実さおよび自由さの感覚と合致している。

図3 2つの可能性の場合の p と H の関係



一般に K 種の文字をもつ情報源では、すべての文字が等確率で、任意の文字 $P_i = \frac{1}{K}$ のとき情報量は最大で $\log K$ に等しい。また同じアルファベットをもつ2つの情報源を比較した場合、長さ n 字の可能な通報の数はエントロピー H が大きいほど大である。そして H が最大で通報の数が最

も多いのは、すべての文字が等確率でかつ文字の並び方が独立の場合であって、例えば26字のアルファベットをもつ英語ならば長さ n 字の可能な通報の数は 26^n ($n = 4$ のときは $26^4 \approx 471,856$, $n = 5$ のときは $26^5 \approx 12,217,664$, $n = 6$ のときは $26^6 \approx 318,383,968$)と莫大な数に上る筈である。ところが実際に使用されている語数といえば、英語に限らずどの自然語でも、言語の内在法則としての複雑な文字系列生成の確率的法則という制約のために、右の数値とは比べものにならないほど小さい。今、エントロピー H なる情報源(A)と、これと全く同じアルファベットをもち、各文字が等確率で並び方が独立な第2の情報源(B)を想定しよう。(B)のエントロピーは同じアルファベットをもつ情報源の中で最大だ

から H_{max} と書き、(A)の長さ n 字の可能な通報の総数 $M_1(n)$ と (B)の長さ m 字の通報の総数 $M_2(m)$ とがもし等しくて $m/n = H/H_{max}$ であるとするれば、(A)では確率的構造が複雑なため情報量が小さく、そのため同じ情報を伝えるのに (B)では通報の長さを H/H_{max} まで縮めることになる。この量を相対エントロピー(Relative entropy)という。(A)から生じる長さ n の通報は (B)の通報に書き替えることにより、その長さを $n(H/H_{max})$ にできるから、残りの $\{1 - (H/H_{max})\}$ 字は不要となり、これが (A)の冗長度である。英語のアルファベット26文字について (B)の場合のエントロピーを F_0 で表わせば、英語の冗長度 H/F_0 は50%を優に超えと言われるが、これは相対エントロピーは50%以下であり、われわれは英文を書くとき、使用する文字のおよそ半分を自由に選べるだけで残りの半分以上は英語の内部構造によって強制的に定まることを意味する。つまり半分以上が、英語という形式を保つためのもので、いわば情報の伝達に直接役立ってはいないと解釈できる。

こう見てくると、言語における冗長度は、情報理論の立場からは無用の長物はおろか、情報伝送の能率を低下させる邪魔物のように思われるかも知れないが、実は冗長度には雑音や暗号解読などで非常に大きなプラスとなる面がある。ボナンザグラムやクロスワード・パズルのように、若干の文字を省いても原文が正しく復元できたり、

BEZT WISHES GOR VERY HAPPP BIRTFDAY

のような誤った英文電報も容易に判読して正しい英文になおせるのは、英語の冗長度が相当大きいお蔭である。仮りに英文の冗長度がゼロでエントロピーが F_0 に等しければ、すべての通報が意味のある文章となり、1文字欠落する毎に26通りずつの可能性が生じて、到底原文の推測ができないであろうし、また1字間違えると全く意味の異なる別の正しい文章となってしまうであろう。

5 情報量と英文の種類 前節で近代情報理論のもたらした数学的成果の中、最も重要な部分である情報量とエントロピー、冗長度を中心にその概略を述べたつもりであるが、ひるがえって考えればわれわれの日常の言語活動はすべて情報

の授受と見られるので、言語学を志す人々の情報理論に対する関心は当然のことであろう。以下情報量の多寡という観点から英文の分類を試みたいと思う。

先ず各種の文章ないし発話の分類基準として (a) Mono-bit information (単1ビット情報) か Multi-bit information (多重ビット情報) か, (b) Verbal information (言語による情報) か Non-verbal information (言語外情報) か, (c) Definite receiver (特定受報者) か Indefinite receiver (不特定受報者) か, などが考えられるが、実際の発話はこれらの異なった分類基準の組合せとして説明される。

(1) Information-eliciting utterance (情報誘引発話) 発話によって直接1名ないし数名の Definite receiver or receivers に広義の Information — 言語的もしくは言語外的 — を喚起することを目的とするもの。

(a) Verbal-information-eliciting utterance (言語による情報誘引発話) 普通に言う Interrogative sentence がこれに相当し、求めている情報量が1ビットか多重ビットかによって次の2種に大別する。

(i) Mono-bit-information-eliciting utterance (単1ビット情報誘引発話) いわゆる General question がこれに該当し、Yes or No の典型的“二者択一”の Verbal-information の供与を特定の受報者に求めるもの。

Is he an American? — Yes, he is. / No, he isn't. (何れの通報もその情報量は1ビット)

Are you ↗ ready, or are you ↘ not? — Yes, I am ready. / No, I am not. (Alternative general question の場合も同様である。)

その他、普通に Special question として取り扱われる Alternative (or Disjunctive) question (選択〔離接〕疑問) もこの範疇に加える。

Is your baby a ↗ boy or a ↘ girl?

Is he an ↗ Oxford or a ↘ Cambridge man? (=‘I already know he is one or the other; tell me which he is.’)

Do you drink ↗ sherry or ↘ port? — Sherry, please.

更に又、普通“二者択一”を表わす Interrogative pronoun (or adjective)

“which”を伴う Special question も単 1 ビットの通報を求める発話となることが多い。

Which do you like better, spring or summer ? — I like summer better.

(ii) Multi-bit-information-eliciting utterance (多重ビット情報誘引発話)

ほとんどの Special question がこのグループに属すると考えられる。直接に求められる通報の情報量が何ビットであるかは一般に不定であって、個々の文脈、シチュエーションに依る変域 (Domain) の大小に応じて情報量は大いに異なる。例えば、*What day of the week is it today ? — It's Sunday. / What day of the month is it today ? — It's the fifth. / What day of the year is it today ? — It's April 1.* では Response [or Elicited] Information-amount は時空を超越して常に一定であって、それぞれ約 3 ビット、5 ビット、8.5 ビットであることが容易にうなずける。同じ Response である *He is seventy.* でも、それが *How old is your friend ?* に対するときの方が *How old is your grandfather ?* に対するときよりも情報量が大い。⑩

(b) Non-verbal-information-eliciting utterance (言語外情報誘引発話) これは言語によらない広義の Information、つまり各種の Action を誘発するもので、普通に Imperative sentence がこれに該当するが、例外的に

④ *Tell me where he lives [if he still lives].* の如く文頭の Verb の意味内容が Verbal-information を誘引する特別な発話では実質的に上掲の(1)(a)(i)および(ii)に相当するし、また

⑤ *Tell me the company you keep, and I'll tell you what you are.* (諺) では、実質的に後述の(2)(b)に入る。

(2) Non-information-eliciting utterance (非情報誘引発話) これは(1)と異なって、或る特定受報者に対して直接に言語的ないし言語外的情報の供与を求めるものでなく、一般に「独語などでない限り」不特定多数の偶然の受報者 (['Casual receiver(s)']と呼んでおこう) — 聴衆および読者をも含める — を対象とする。それ自身の Multi-bit-information-amountをもつが、この多重ビット情報には単純なものと複雑なものと考えられる。普通の Declarative sentence と Exclamatory sentence を区別なくこのグループに入れる。後者の特性から、

(a) Single non-information-eliciting utterance (単一非情報誘引発話) これは文章の構成の1つに関する多重ビット情報方式であって次のような強調構文がその例である。

It was *the boy* that [or who] broke the window yesterday. / It was *the window* that the boy broke yesterday. / It was *the act of breaking* that the boy did to the window yesterday. / It was *yesterday* that [or when] the boy broke the window.

(b) Complex non-information-eliciting utterance (複合非情報誘引発話) これは(a)の如くに文のどの要素をも強調しない普通の Statement で、その情報量は(a)よりも遙かに大きい。例、The boy broke the window yesterday. 尚、いわゆる Exclamatory sentence の内包する種々の emotive value は、情報の数量のみを問題とする情報理論では無視するのが妥当であろうと思う。

以上極めて粗雑ながら情報量という観点から、英文の新しい分類を取ってみたが、大方の御叱正を得て妥当な結論に一步でも近づきたいのが筆者の念願である。

〔未完〕

〔註〕④ Wiener によれば、「われわれの状況に関する2つの変量があるとし、その一方は、われわれには制御できないもの、他の一方はわれわれによって調節できるものとしよう。そのとき制御できない変量の過去から現在までの値に基づいて、調節できる変量の値を適当に定めて、われわれに最も都合のよい状況をもたらしたいと望んだとしよう。このとき、それを達成する方法がサイバネティックスに外ならない……」のである。

②③ N. Wiener: *Cybernetics (or Control and Communication in the Animal and the Machine)*. John Wiley.

④数学者によっては、H. Nyquist の'24年、'28年の論文、Hartley の'28年の論文をもって情報理論の萌芽とみなすものもある。

⑤ C.E. Shannon and W. Weaver: *The Mathematical Theory of Communication*, Univ. Ill. Press, '49.

⑥ヨーロッパでは、情報量の最初の指導者である Hartley を記念して、ピ

ットの代りにハートレーと呼ぶならわしである。

⑦情報量を、可能性の数の対数として定義することの今1つの便利さは、情報論では非常に大きな数の可能性を取り扱う必要があるからである。例えば、情報がA、B 2種類だけの文字の系列として表記されている通報である場合、1文字、2文字、3文字の長さの通報の数はそれぞれ $2^1=2$ 、 $2^2=4$ 、 $2^3=8$ 、というふうに、長さが1文字増す毎に通報の数は2倍になる。かくして、4文字では $2^4=16$ 、5文字では $2^5=32$ 、一般に n 文字では 2^n 個の通報が存在する。更にA、B、Cの3種類の文字があれば、1文字ます毎に通報の数が3倍になるから、一般に n 文字では 3^n となり、一般に文字の種類が K であれば、長さ n 字の通報は K^n だけ存在する。英語の26種類の文字を用いて、たとえば5字の短い通報と送信するときでも $K^n = 26^5 = 11,881,376$

という数になり、日本語の50種類の仮名文字を用いてする10文字の長さの通報に至っては $K^n = 50^{10} = 97,656,250,000,000,000$

という龐大な数値となる。ところが K^n の対数をとると、 $n \log K$ となり、ずっと扱い易い数となる。

⑧今、日本語、英語などの自然語を構成する各文字の使用頻度をとってみると、英語、ドイツ語、フランス語、スペイン語では何れもEが最も大きく、和文ではIが最も大きい。第2位の使用頻度の文字は、T(英語)、N(ドイツ語)、N(又はS)(フランス語)、I(イタリア語)、A(スペイン語)、O(日本語)である。尚、英文の文字の中、EおよびZの使用頻度(全体を1として)はそれぞれ0.1073, 0.00063であると言われる。

(9) Cf. Is he an \nearrow Oxford or a \nearrow Cambridge man ? ('Has he studied at Oxford or Cambridge—I don't care which—as opposed to the University of London, the German universities, etc. ?') / Do you drink \nearrow sherry or \nearrow port ? ('Do you drink [such strong wines as] sherry or port ?')

参 考 文 献

数理科学 1965, 5, 特集：情報理論（鎮目恭夫「情報理論とは何か？」／国沢・梅垣「情報理論の数学的構成」／榎本肇「工学における情報理論」／竹内啓「情報理論と統計」／渡辺修「言語学における情報理論」

数理科学 1966, 8, 特集：言語（中島健蔵「文学の研究と数学」／長谷川欣佑「文法理論」／野崎昭弘「計算言語学入門」／細井勉「暗号と字引」）

本多波雄「情報理論入門」(昭44,3)日刊工業新聞社

関英男「情報理論」(昭44,3)オーム社

北川敏男「情報科学への道」(情報科学講座A・1・1)共立出版KK

奥野治雄「情報処理基礎概論(1)」|—システム工学編—(昭47,1)オーム社

同氏「情報処理基礎概論(2)」|—プログラミング技法編—(昭46,11)オーム社

森政弘「制御と情報」(NHK情報科学講座④)(昭43,8)

喜安善市「通信と情報」(NHK情報科学講座⑤)(昭43,9)

渡辺茂「認識と情報」(NHK情報科学講座⑥)(昭43,7)

北川敏男「社会と情報」(NHK情報科学講座⑧)(昭43,10)

森口繁一「初等数理統計学」(昭45,4)培風館

尾崎繁雄「大学教養数学」(昭42,3)森北出版KK

矢野健太郎「教養の数学」(昭42,3)裳華房

遠山啓「現代数学対話」(1967,5)岩波書店

矢野健太郎「新しい数学」(1968,10)岩波書店

矢野・茂木「行動科学者のための基礎数学」(昭47,3)裳華房

スモリャンスキ編

宮本・松田・松野訳「情報数学ハンドブック」(1972)森北出版KK

D. Fabun: *Communications—The Transfer of Meaning*, 1960 (The Glencoe Press)

M. Gross and A. Lentin: *Introduction to Formal Grammars*, 1970 (Springer-Verlag)

G. Hemphill: *A Mathematical Grammar of English*, 1973 (Mouton)

J.C. Hancock: *An Introduction to the Principles of Communication Theory*, 1961 (McGraw-Hill)

C. Cherry: *On Human Communication*, 2nd ed., 1957, 1966 (The M.I.T. Press)