OUKA
The University of Osaka
Institutional Knowledge Archive

Title	品詞の情報量
Author(s)	樺島, 忠夫
Citation	語文. 1956, 17, p. 23-27
Version Type	VoR
URL	https://hdl.handle.net/11094/68496
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

品詞の情報量

樺島忠夫

話し言葉には内容の伝達に直接関与しない終助詞、間投助詞、感動詞のたぐいがしきりに挿入されたり、接続詞、副詞がほとんど意味もなく使われ、また論理的に重複した表現や言葉のくり返しが多く見られるという（はなしことば序・伊佐早敦子　国語国文二十二巻三号）そして、多くの人が注意しているように、このような一見無駄なことばは実は無駄ではくて、理解を容易にするための潤滑油のはたらきをしているのだ。これはなぜだろうか。

また私は類別した品詞の比率を、条件を異にする文章について測定し、これらが文章の成立条件によつて規則的な変化をすることを観察した（「国語学」十八輯、及び「国語国文」二十四巻六号の拙稿参照）即ち文章中の品詞の比率は名詞が最も高く、動詞がこれにつぎ、形容詞、副詞などを合併したものの比率は比較的低く、感動詞、接続詞を合併したものの比率が最も低い。そして、話し言葉、普通の書き言葉、新聞記事の三者を比較すると、名詞はこの順で比率が高くなり、これに反して形容詞の類や接続詞の類は比率が低くなつて行くのである。これもなぜだろうか。

この理由の一つとして、各品詞の持つ機能が異なり、種々の条件に対応して各品詞の出現率が変るからだということが考えられる。その品詞の機能として、品詞の情報量を考えこれによつて以上の問題を解決しようとするのが本稿の意図する所である。

たとえば次のような文を考える。

次の宝くじの当選番号は一〇三四八番だ。この文の中で最も重要な語は「一〇三四八番」である。この語を除いて

次の宝くじの当選番号はxだ。

とすれば、我々はこの文を受けとつても殆ど情報を受けとらないに等しい。しかし先の文から「の」と「だ」とをのぞいて

次の宝くじ当選番号は一〇三四八番。

としても我々が受けとる情報には殆ど変化がない。このように、ある文や文章（以下これを文章と略記する）中の語は各文章の意味を作りあげる働きをなすが、その場合に文章の意味を作りあげるのに強くあずかる語や弱くあずかる語がある。この度合を情報量とよぶことにする。これを別の表現で説明すれば次のようになる。

文章を作りあげる語の集合を考える。この集合の元である或る一つの語が欠けていて、それが何であるかわからないとする。これをxであらわそう。このxが、集合の他の元との関係においてとり得る値を考えるとき、xはaであり、a以外のものではあり得ないとすれば、xがaである確率は1となり、解は一つに定まる。このような場合、xがaであるという情報を得ても我々の知識は増さない

またxがとる値がaまたはbであるように二つの値をもつ場合や、一般にn個の値をもつ場合がある。このようにxがとり得る値の数が増せば増すほどわからなさは増すことになる。このような場合にxが特定のsであることがわかつたとすれば、わからなさが減っただけ多くの知識を得たことになり、大きな情報を得たと言える。この知識を得た量が情報量であるから、語の情報量を次のような方法で近似的に測ることができる。

即ち、ある文章から語を消去し、その文章を被験者に示して消去した語を再生させる。容易にしかも正しく再生される語は情報量の小さい語であり、その逆は情報量の大きい語である。

この方法は通信工学で用いられる通信理論の方法とは異なつている。通信理論では選択可能数の対数（2を底とする）をもつて情報量を示すことにしてある。したがつてこれにならい、語の情報量を数値でとらえようとするならば、語の使用頻度をとるか、または、以上でのべたような方法では、たとえば試行回数を非常に大きくした場合の、語が正しく再生される率pを考え、$\log(1/p)$をもつて情報量とすべきであろう。しかし、この場合の情報量は、通信理論の頻度による情報量と異なつて、文章の難易や語消去の度合となされ方によつて変ると考えられるし、また被験者の能力によつても変つてくる。以下では情報量を数値で測定するのではなく、品詞間の比較において、同一条件下におけるその大小を考えるにとゞまるから再生率をもつて近似的に情報量を考えるのである。

以上のような考えのもとに、品詞を機能によつて次の四つに類別し類別された品詞の各々が同一文脈中でもつ情報量の平均値の大小を考えよう。

N 名詞

V 動詞

A 形容詞、形容動詞、副詞、連体詞

J (1)感動詞、接続詞

(2)助詞、助動詞

品詞を右の四つに類別し、以下煩を避けるため、N、V、A、Jによつてそれ〳〵を表わす。たゞし本稿ではJの(1)と(2)とを区別しない。

調査の方法として、後に示すような、夫々三個の文章からなる第一群、第二群の資料を設け、四分の一の語を等間隔に抽出して消去し、被験者に消去した残りの文章を示してこれらを再生させ、類別した品詞毎に再生率を測定する方法をとつた。四分の一の語を消去した理由は、あまりに多くの語を消去すると再生が困難になるだろうし、また逆に消去が少なければ再生があまりに容易になるだろうと考えた所から両方を考慮して定めたのである。また等間隔に消去した理由は、もしランダムに抜いて、二語以上続けて消去された場合は再生が甚だ困難になるからである。

また、文中のある語について考えると、

(1)彼は立派な男だ。

(2)非常に立派な男だ。

(1)の場合「立派な」を消去すると「彼は男だ」となり「立派な」が存在した痕跡はなくなつてしまう。しかし(2)の場合には「立派な」を消去しても「非常に男だ」となり、「非常に」と「男だ」との間に用言が存在した痕跡を残す。したがつて(1)のような場合よりは(2)のような場合に再生の努力が強く働くであろう。また調査の場合に

消去した語以外の部分にかつてに語を入れられると、それによつて文脈がまげられ、原文とはかなりちがつたものが作られる恐れがある。以上の理由によつて、語を消去した部分には△印をつけて消去を暗示することにした。しかし、文例に示すように、必要がないと思われる部分には語を入れなくてもよいことにして、再生を強制しないようにした。

また、一つの文章の語数はなるべく等しくなるようにえらんだ。これは文章の長さが再生率に作用するかもしれないという恐れによるためである。

調査に使用した文章を左に示す。

（注意）

次の三つの文章のところどころに△がつけてある。もしも△の部分に語をおぎなわなければ意味が不明になつたり、文法的におかしくなるとすれば、どういう言葉もおぎなうべきか。最も適切だと思われる語を入れよ。たゞしおぎなう必要がない場合には入れなくてもよい。その場合にはその所に×をつけよ。おぎなう言葉がわからない場合には?をつけよ。

（文例）

第一群

一、毎日同じ生活をくり返しているサラリーマンや、何かじらモタモタしたものを胸の中にもつている若いひとたちが、電車のツリカワにぶらさがつてそこらをながめまわすとき、いちばん心をひかれるのは、自分の生活に変化を与え得る力であろう。

二、わたしの家の飼犬が、どこが痛いのか気持が悪いのか、時々、黙つて軒下にうずくまつている。言語を持たない犬属には何も訴えられないのである。そんな時に言葉を持つ人類をつくづく幸福と思う。

三、日光見物の旅行のとき、東京で買物をして、いつものくせで、「おーきに」と言つて店を出ようとしたら、店の人が妙な顔をしたこと。その時、何かしら屈辱感のようなものを言葉の上ではじめて味わつたことを、今でも「おーきに」と言う言葉を使うたびに思い出す。

第二群

一、詩人の声や、美しい言葉のレコードがもし一般に売り出されたら、私たちの話し言葉に対する関心はもつと高まるだろうし、全国の方言や、種々の歴史的なできごとが音にふきこんであれば教育の上にもいろいろと役立つだろう。

二、話し言葉が書き言葉と一致するのは望ましいことではあろうがそれ〴〵の機能を考えてみるとむずかしいことである。しかし、今日のように、この二つが離れているのは、わたくしたちの言語生活にとつて好ましいこととはいえまい。

三、私が京都の家を出てからもはや三十年余りになる。父も母も他県人だし、偶然の結果たゞ京都に生れた、という来歴だけだが、初対面の人は私の顔を眺め、そして言葉をきいてすぐ関西、又は京都出身だと言いあてるのが普通である。

調査に使用した文章は雑誌「言語生活」からとつたものである。被験者は十九才前後、第一群三十二名、第二群三十三名、所要時間両群とも十五分以内である。文章中傍線を附けた語は消去されたも

のであることを示す。

問題に対する答は次のような規準によつて整理した。(1)原語と同じ語を再生したもの、(2)原語とは異なるが同義語、または幾分か原語の概念を縮小、拡大したもの。以上を正答とする。(3)文脈中で意味は通るが、原語の概念とは異なつたり、あきらかに誤答であるもの、(4)解答不能。以上を誤答とする。また、再生不要としたもの及び再生にあたつてあらわれた異なり語数をそれ〳〵項目をたてゝ分類集計した。その結果を表にまとめてかゝげた。

	品詞	調査語数	正答 計	正答 比率	誤答 計	誤答 比率	不要 計	不要 第率	異語数 計	異語数 一語平均
第一群(三十二名)	N	8	134	0.52	122	0.48	0	0	43	5.4
	V	5	96	0.60	64	0.40	0	0	11	2.2
	A	22	677	0.96	13	0.02	14	0.02	44	2.0
	J	6	53	0.28	45	0.23	94	0.49	26	4.3
第二群(三十三名)	N	14	171	0.37	288	0.62	3	0.01	124	8.9
	V	9	154	0.52	135	0.46	8	0.02	61	7.2
	A	15	369	0.75	77	0.16	49	0.09	57	3.8
	J	3	15	0.15	42	0.42	42	0.43	27	9.0

表についてみると、第一群においても、第二群においても、Aは、N、V、Jとは大変ちがつた様子をしめしている。そこで、まずAをのぞいて、N、V、Jについて眺めてみよう。正答率はNが最も低く、V、Jの順で高くなつて行く。そして誤答率はこの順で逆に低くなつている(正答率は先にものべたように、文章の難易や、消去する語の抽出のされ方などによつて変る。したがつてあらわれた数値そのものはこゝでは問題にしない)また異なり語があらわれる様子を見ると、誤答率が高い品詞ほど異なり語数が大きいが、これは当然考えられることである。さきにのべたように文脈中でのその語が何であるかのわからなさが大きい語ほど大きな情報量をもつとすれば、N、V、Jの三つの間ではNが最も大きな情報量をもち、Vがこれに次ぎ、Jは小さな情報量をもつといえる。ではAはどうか。これは正答率も誤答率も小さい。これに反して文脈中で不要とされる率が第一群、第二群とも五割に近いのである。この事は、Aは取り去られても文章の意味には大して変化を及ぼさないような働きをもつ事が多い事を示しているのである。したがつてAも情報量は大きくないと考えてよい。

また同一品詞内で正答率の大小を調べると、たとえばNでは「とき」「こと」「人」「私」のような語の正答率大、即ち情報量が小であり、「日光見物」「京都」「三十年余り」のような語の正答率小、即ち情報量が大となつていることがわかるが、常識的に情報量の大小という事を考えてみて、この事は当然と言えよう。

以上のようにして、類別した品詞の情報量の大小を測定したので初めに出した問の答えを考えてみよう。

こゝで二つの文章A、Bを考える(したがつて文または文章をなさない、単なる語の集合は考えない)AまたはBの各語の情報量が

数値でとらえられたとして、各語の情報量を加え合わせ、それを語数でわったものを、その文章の情報量平均値とよぶことにする。もしAがBより大きな情報量平均値をもつならばAはBより少い語数で、（少い語数ということは短い時間でまたは狭い空間でということに置きかえられる）大きな情報を与え得ることになる。これを利用した言語現象は我々の周囲に多く見られる。火事になつて至急に避難しなければならないとき、我々は

「原因はわかりませんが、今、この家が燃えていましてね、たぶん我々の今居る所も焼けるでしょう。私はあなたも外に出て難を避けられるのが賢明だと思いますが、如何でしようか」

などとは言わない。

「火事だ！　逃げろ！」

と言う。情報量の大きなことばだけを使うのである。

このように情報量の大きな言葉を使うのは能率的であるが、その反面、理解度は低下する。情報量の大きい語ばかり並んでいる文章では、そのいくつかを見落したり聞落したりするとたちまち意味がわからなくなるからである。講義などにおいて、同じ言葉をくり返したり、または同じ意味を異なる表現で二度言いあらわすのも、文章の情報量平均値を小さくして、理解、把握を容易にするためだ。

波多野完治氏が『一体人が言葉からなかみを感得して行く過程は消化に似たところがある。胃袋というものは消化すべき必要にして充分なる量を送られれば、それでうまく消化機能が行われるというものではないらしい。むしろ多少余分なものが入つて胃がふくれると、胃の筋肉が充分な活動を始めるのである。馬に若草などをたべさせるのも、一つにはこのような必要なよけいものをあたえるという意味もあるのだ。文章とても同様である』というメイクルジョンの言葉を引いて「文章に変化をつけること、主題を言うだけでなくヴアリエーション（変奏）をつけることが、文章にあつても重要である。と言われた（文章心理学入門六十五頁）これも以上のよう意味で考えられる。

以上で話し言葉に一見不必要な遊び言葉が多くあらわれたり、重複した表現や、くり返しが多い理由の一つは自ら明らかであろう。話し言葉は話し手の記憶があいまいになつたり、推稿が時間の流れにしたがつてしか行われないという強制されたマイナスの原因だけではなく、時間的に流れて行くことばを耳でとらえるために生じる理解度の低下を防ぐために、文章の情報量の平均値を小さくするというプラスの理由がともに働いてそうなるのだ。こう考えなければ現象の説明がつかなくなることがある。

これが視覚的に固定される書き言葉では、読み手は自由に後もどりをして読んだり、二度読みなおすことができるから、文章の情報量の平均値をある程度高くしてかまわない。むしろその方が理解しやすい。したがつて、書き言葉は話し言葉に比して名詞（N）の比率が大きくなる。また新聞記事では、紙面及び他の記事との関係上なるべく小さい文章に大きい情報をもり込まねばならない。したがつて接続詞、感動詞の類（J(1)）や、形容詞の類（A）の比率が極めて小さく、名詞の比率が非常に大となつているのである。また、よく言われるように助詞の類（J(2)）さえ省略されるのだ。

以上によつて話し言葉にあらわれる遊び言葉の現象や、品詞の比率が条件によつて変化する理由が明らかになつたと思う。

——大阪府立北野高等学校教諭——