

文章のわかりやすさ評価に関する研究

5C-5

～文法構造と記憶率について～

○宮下 尚子

上條 憲一

日本電気(株) C&C 情報研究所

1 はじめに

現代は情報が氾濫している時代であり、情報の受け手に対して、積極的に発信者側の意見・考えをアピールするためには、「わかりやすさ」を考慮した情報表現を行う必要がある。筆者らは、情報の1つの形態である文章に着目し、文章の「わかりやすさ」の定量化手法の確立と、それに基づいた文章の評価システムの構築を目指している。

文章のわかりやすさに影響を及ぼす要因として、字面要因・文法要因・構文要因・内容要因・効果要因の5つがあげられる[1]が、筆者らは構文要因に着目して研究を進めている。構文要因と文章のわかりやすさの関係を導き出すための評価実験を行う必要があるが、その手法を検討するために、記憶に関する心理実験を応用することを試みた。本稿では、この評価予備実験に関して、その目的、内容、結果について述べる。

2 評価実験

2.1 実験の目的

構文要因とは、単語の構成(文法構造)に関する要因である。最終的には、例えば係受けの個数等の文法構造に関する定量的要因で、1文におけるわかりやすさの定量化を行う事を目標としている。今回の予備実験では、「わかりやすさ」を「記憶のしやすさ」と見做し、記憶率で定量化した。本実験を行う際に、どのくらいの文長の文を使用したら良いか等の材料の検討と、またどのような要因を考慮したらよいのかの検討を目的としている。

2.2 実験計画

1. 材料

(a) 提示文 標準的な文として、新聞のコラムを材料とした。提示文は、文法構造以外の要因の影響を避けるため以下の基準により抜粋した。

- ・ 特殊な固有名称が含まれていない
- ・ 特定の分野の専門用語が含まれていない
- ・ 擬音語・擬態語が含まれていない
- ・ 「」や()が含まれていない

抜粋した文の漢字含有率は平均が30.75%、標準偏差が9.68%である。

(b) 材料数 コラムの文章の1文の長さの平均を計ったところ、文長が約30文字であったため、それをはさんで前後10文字ずつの間をとり、20(±2)文字、30(±2)文字、40(±2)文字の文長を持つ文長をそれぞれ12文ずつ用意した。

基準を満足した例文として、次のような文がある。

「赤い花が葉の中にもれるようにして咲いている。」
(文長20文字)

「飼育を担当する人たちの努力が実って、ペンギンの卵がかえった。」 (文長30文字)

「猛スピードで突っ込んで来た車は自転車に激突し、さらにコンクリートの塀を突き破った。」 (文長40文字)

2. 手続き

実験を開始する前に、被験者に対し記憶実験であることを教示し、最初にディスプレイ上に1文を提示する。その文を被験者に1回音読させ、音読終了後ディスプレイの文を消す。その後、音読した文を紙面上に書くことによって再生してもらう。1回の実験につき、各文長から3つずつの9文を与えている。同様の手続きを1人につき日を改めて4回に分けて行った。

3. 被験者 6名(平均年齢27才)

3 実験結果と考察

データは次の点に注目して解析を行った。

1. 文長と正解率の間の関係を調べる
2. 各文の文法構造の特徴と正解率の関係を調べる

各文の正解率を次式で定義する。

$$\text{正解率} = \frac{\text{再現性が高かった人数}}{\text{全体被験者数}}$$

ここで、「再現性が高い」文というのは、以下のうちのいずれかである。

1. 提示文と再現文が1字1句同じ
2. 置き変わっても意味が変わらない助詞で置き変わっている
3. 置き変わっても意味が変わらない単語で置き変わっている
4. 2と3の複合型

3.1 文長と正解率の関係

それぞれの被験者について、文長と再現性の高い文の個数の関係を表1に示す。そして、表2はこの表1を基に分散分析を行った結果である。この結果をみると文長のF値が93.83と非常に高くなっており、文長の効果に有意差が出ている($p < 0.01$)。しかし、この表において、細胞内誤差の平方和がかなり大きい事に注意しなければならない。この誤差が大きいということから、今回の実験では、評価した文長の幅が20,30,40と離れているため、文法構造が及ぼす影響が文長が及ぼす影響に隠れているという可能性がある。よって、この文長のとり幅が±10文字よりも小さい場合には、有意差が出ないかもしれないという可能性がある。

また、表3は各文における正解率を文長ごとに表したものである。これを見てわかるように、文長が同じであっても、1人も正解していない文もあれば、全員正解している文もあることから、文の記憶は一概に文長によるものではないという事がわか

る。

またここで、文の長さではなく、記憶をする上で意味的な単位（チャンク）と考えられる文節数と正解率の関係を調べてみた。結果を図1に示す。このグラフより、まず正解率の平均に関しては文節数が6までは一定レベルで高い値を保っており、文節数が7,8では約5割くらい、また文節数が9以上では、文節数10で高い値が出ているものの、これはデータ数が少なくあまり信用のできない数値であり、前後の数値から判断して除々に正解率が低くなっていると思われる。また、標準偏差に関して言えば、文節数6まではかなり低い値であるが、文節数が7,8のところではかなり値が大きくなっており、それ以上の文節になるとまた除々に低くなっている。この標準偏差が大きくてい部分の理由としては文法構造に大きく因っていると思われる。以上の事をまとめると、

1. 文節数が6以内では、文法構造によらず記憶率が高い。
2. 文節数が7,8は、文法構造によって記憶率が支配される。
3. 文節数が9以上では、文法構造と文節数によって記憶率が変化する。

が考えられる。以上の考察を模式的に示すと図2のようになる。以上より、文法構造と記憶率の関係を調べる実験を行う際に適節な文節数は正解率が5割くらいであった7か8であり、今回1文節の平均が4.20文字であったことから、材料は約29～33文字の文長を持つ文が適切である事がわかった。

3.2 文法構造に関する考察

次に文法構造の係受けの距離と深さに関するデータ処理を行った。具体的には、係り受けの距離として「係り」と「受け」の間にある文節数、係り受けの深さとして入れ子の数をカウントしている。この2つの観点を取り入れて正解率との相関を文長ごとに算出した。具体的にはすべての係受け間の距離を足した距離総和を文節の個数で割った平均距離である「一個当りの距離」と、「受け」側の文節の個数を全体の文節の個数で割った「複雑度」との相関を算出した。この結果、「1個当りの距離」と正解率の相関係数は文長20文字では0.3884、文長30文字では0.2938、文長40文字では0.3344とあまり高い相関は得られていない。しかし、「複雑度」との相関を見てみると、文長20文字では0.1257、30文字では0.1001、文長40では0.4721となり、文長20文字、30文字では相関があまり見られないのに、文長40文字では相関係数が他に比べて大きく出ている部分が興味深い。この事は、文節の個数によって文法構造が効いてくる領域が出て来るという前述したモデルの妥当性を示していると思われる。しかし、全体の結果としては、考えていた要因と正解率の相関があまり明確になっていない。これには以下の原因等が考えられる。

1. 不適切な実験題材
(単語の影響をかなり受けていたものがあった)
2. 被験者の実験に対する動機付けの不統一
3. 文法構造で解析する際の、係受けの意味的な重み付けの未考慮
4. 係受けの曖昧度（読んでいる途中でどれくらい係受けの曖昧性が発生するか）

4 まとめ

今回の予備実験では、記憶率は文長に一概によらないという可能性を示し、また文法構造の定量化を導き出す記憶実験において、どのような事を考慮した材料が良いか、またどのような要因にポイントをおいて解析すれば良いかを導きだした。

今後は、今回導きだした結果を基に、文法構造とわかりやすさの定量化のための評価実験を行っていく。

参考文献

- [1] 浅野、小川: “学年レベルを基準とした文章の読みやすさの評価,” : 日人人間工学会関東支部大会講演集 C-19, 1989.

表1 各被験者ごとの正解数

被験者	20	30	40	計
1	12	10	4	26
2	11	8	2	19
3	7	7	1	15
4	11	7	3	21
5	11	9	5	25
6	12	9	4	25
合計	64	48	19	131

表2 表1に基づく分散分析

変動源	平方和	自由度	平均平方	F
文長	14.45	2	7.225	93.83**
被験者	2.58	5	0.516	3.04
交互作用	0.77	10	0.077	0.453
誤差内積	33.75	110	0.17	
全体	51.55	119		

表3 文ごとの正解率

文	20	30	40
1	5/8	0/8	1/8
2	8/8	4/8	0/8
3	8/8	8/8	1/8
4	5/8	6/8	1/8
5	5/8	3/8	0/8
6	8/8	2/8	0/8
7	5/8	5/8	1/8
8	8/8	5/8	2/8
9	8/8	2/8	2/8
10	4/8	5/8	3/8
11	5/8	5/8	3/8
12	1/8	2/8	4/8

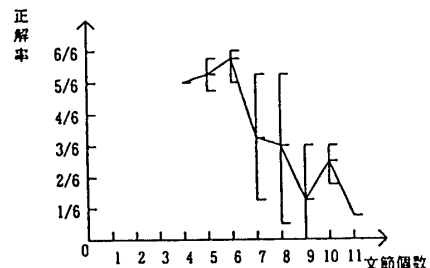
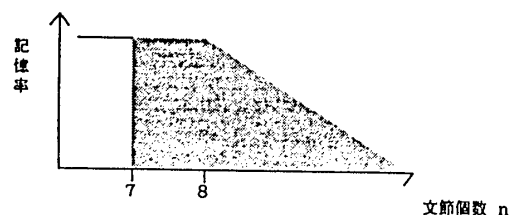


図1 文節個数と正解率の関係を表すグラフ



- $n < 7$: 文法構造に係わらず記憶率が高い
 $7 \leq n \leq 8$: 文法構造によって記憶率が変わる
 $8 < n$: 文法構造と文節個数によって記憶率が変わる

図2 文法構造と記憶率の関係を表す図