

# 3Q-7 文書中の単語認識における彩色の効果

内田友幸 田中英彦 東京大学 工学部

## 1 はじめに

近年、ネットワークや大規模文書データベースの普及などによって電子化された自然言語文書が大量に流通するようになってきている。しかし、これらの自然言語文書に対するユーザーインターフェースは従来の紙メディアと同様の白黒の表現が主流である。

これに対し、カラープリンタ、カラーディスプレイが普及し、色を利用したユーザーインターフェースが主流になってきている現在、色を利用する文書という選択肢も選び得ると考えている。現在我々は適切に文書を彩色する事でより表現力を豊かにしたり、より読みやすい文書にできる可能性について検討を重ねている。

本稿では、彩色文書が文書の判読時に与える影響を心理実験を用いて調査を行なったので、その実験の内容と考察について報告する。

## 2 心理実験概要

色を付けることによる心理的効果は多彩なものが期待できるが [1] 今回は「文書中要となる要素に対するアクセスを速く行なえる」点に着目した。

例えば、文書検索時にキーワードが彩色されていれば、キーワード周りの文書のポイントとなる部分へのアクセスを速めることが期待できる。

しかし、彩色による効果には悪影響も懸念される。例えば色に気をとられてしまって読解に時間がかかるなど、彩色が読解の妨害要因になる可能性もある。

そこで、これらに関する知見を得るため、次のような心理実験を行なった。

まず、新聞記事中の単語一つを赤字または太字で強調したものを用意し、ディスプレイ上に表示する。被験者には表示された新聞記事中の強調された単語の品詞による分類を行なってもらい、それにかかった時間を測定する。これにより、単語へのアクセス速度の強調の仕方による違いを測定できる。

また、同時に、同じ記事に対して単語を発見したら即座にキーを叩いてもらうことも行なった。これにより単語の発見と反応にかかる時間を測定し、分類にかかった時間との差をとることで単語の認識にかかる時間を測定できる。

## 3 心理実験内容

### 3.1 文書提示方法

被験者には一回の測定で新聞記事一つを提示する。記事は各記事毎に約 1200 × 500 ピクセルの一枚のビットマップファイルになっていて、この中で文字は 144DPI 相当で構成されている。そのうち一つの 2 字の漢字熟語が赤字、または太字で強調されている。

文字は 12 ポイント相当の明朝体で記述しており、赤字は赤色に着色し、太字は 14 ポイント相当のゴシック体で指定している。被験者にはこの指定の単語を基準に沿ってキーを押してもらい、表示からキーの押下までの時間を計測する。

画像の表示には三菱電気のダイヤモンド方式の 17 インチ CRT RD17G2 を利用し、被験者の目の位置から 1m 離れた位置に暗所にて表示させた。この際、表示される文字の 1 辺は 6mm であるので、視角にして 0.34 °に相当する。

なお、この CRT の輝度や色座標などの特性をミノルタの CL-100 によって計測した。この結果を表 1 に示す。

表 1. CRT 表示の輝度と色座標

| 利用色       | 輝度 (lx) | 色度 (CIE 1931 XYZ 表色系) |
|-----------|---------|-----------------------|
| 白 (文字の背景) | 38.5    | (0.292, 0.323)        |
| 灰 (記事の周囲) | 19.5    | (0.291, 0.332)        |
| 黒 (文字、太字) | 2.8     | -                     |
| 赤 (赤字)    | 6.9     | (0.623, 0.341)        |

### 3.2 被験者の操作

被験者はまず、強調された単語を品詞によって分類する。強調された単語は固有名詞かサ変動詞性名詞のどちらかであるので、なるべく早く判断をして 2 つのキーのどちらかを押すことで分類を行なう。なお、ここでのサ変動詞性名詞とは、するを後ろにつけて意味ができる熟語で実験、測定、調査、研究などのことである。

もう一つの操作は、強調された単語を発見するだけというもので、なるべく早く探し、発見した時点で指定のキーを押してもらった。

利用した記事は全部で 46 記事で 6 記事はトレーニング用とし、40 記事を 20 記事ずつに分け 1 群、2 群とする。被験者には赤字と太字のトレーニングデータを分類して慣れてもらった後に、1 群の太字と 2 群の赤字の組合せか、1 群の赤字と 2 群の太字の組合せのどちらかを順次提示する。

まずは 40 記事を品詞毎に分類してもらい、次に同じ 40 記事中の強調された単語を探索してもらった。

実験システムは AT 互換パソコンの MS-Windows3.1 上に実装した。Window システムの背景を灰色にし、中央に窓を開いて新聞記事を表示させた。新聞記事を提示する際にはまず窓内の白い背景の中央部に黒い文字で「X」を 1 秒間表示し、その後 1 秒間真白の窓を表示し、その後新聞記事を提示した。表示はキー入力が行なわれると同時にまた白い窓に戻るようにした。

#### 4 実験結果

被験者の数は 6 名で、分類にかかった時間、単語の探索にかかった時間の測定でそれぞれ 240 個の素データを得た。分類所要時間のヒストグラムを図 1 に示す。

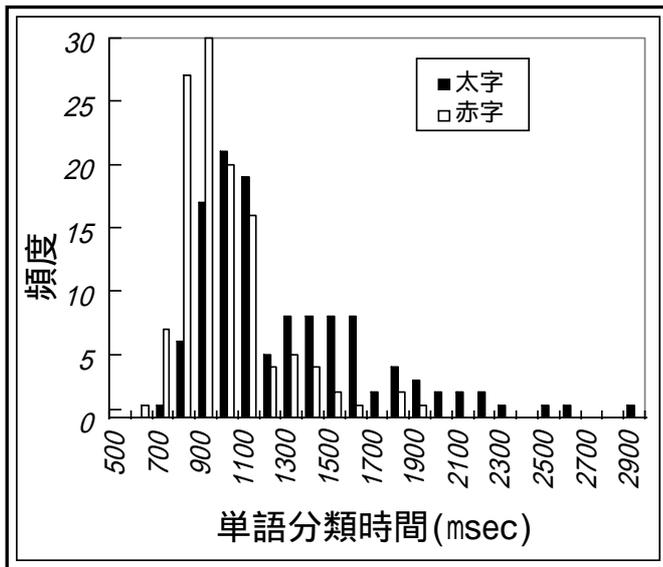


図 1: 単語分類所要時間のヒストグラム

赤字の方が左側に鋭いピークをもっているのに対し、太字は右側になだらかに減衰していつている。これは太字よりも赤字の方がポップアウトが確実に行なわれていることを示唆していると考えられる。

また、このデータの統計的解析の結果を表 2 に示す。この t-検定の結果は 8.08 であり、95% の区間での境界値  $\pm 1.98$  の範囲外なためこの 2 群は分離しているといえる。このことから赤字の方が単語の分類にかかる時間は少なくなると言える。

表 2. 単語の分類時間

|            | 赤字   | 太字                 |
|------------|------|--------------------|
| 平均         | 941  | $1.24 \times 10^3$ |
| 標準誤差 (95%) | 21.2 | 38.1               |
| 標準偏差       | 232  | 417                |

分類に要した全時間は「単語の位置の探索 + 文字の知覚 + 単語の同定 + 単語の概念の同定 + 分類判断 + 手の筋肉への指示 + キーの押下」となると考えられ、単に単語の探索して押すだけに要した時間が「単語の位置の探索 + 手の筋肉への指示 + キーの押下」であると考えられるので、その差であるところの時間は「文字の知覚 + 単語の同定 + 単語の概念の同定 + 分類判断」という単語の認識所要時間と考えられる。

この認識所要時間のヒストグラムを図 2 に示す。また、統計解析の結果を表 3 に示す。t-検定の結果が 2.87

なのでこの 2 群も分離しているといえる。この結果から赤字の方が太字よりも認識の所要時間が大きいことがいえる。

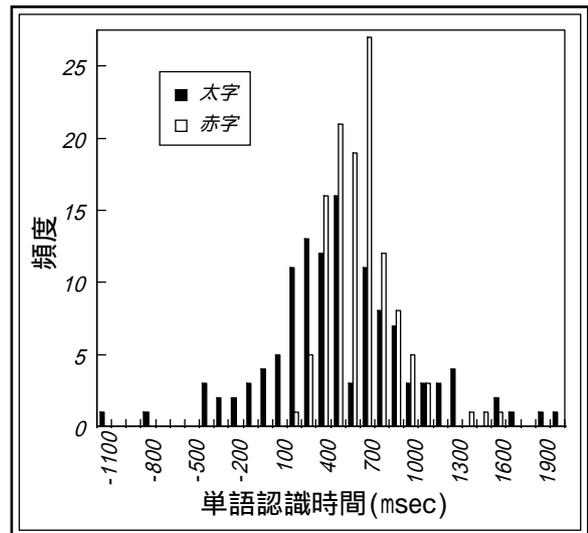


図 2: 単語認識所要時間のヒストグラム

表 3. 単語の認識所要時間

|            | 赤字   | 太字   |
|------------|------|------|
| 平均         | 604  | 462  |
| 標準誤差 (95%) | 21.5 | 45.4 |
| 標準偏差       | 235  | 497  |

#### 5 考察

実験の結果から文書中の読ませたい単語を赤色に彩色することは太字で指定するより所要時間の短縮につながるということが言える。

しかし、この短縮は主に単語探索時間の短縮によってもたらされているものであり、見つかった単語を認識するのにかかる時間は逆に赤字の方が大きくなってしまっている。

これは太字よりも赤字の方が読みにくいということを示していると言えるだろう。

この読みにくい原因には赤が黒よりも輝度が高いことによるコントラストの低下、太字のフォントの違いによる読みやすさの効果、なども考えられるが色情報が読解過程を妨害していることも十分に考えられる。被験者の感想の中にも文字が赤いとチラチラするというものがあり、これがこの読解過程の妨害に該当しているとも考えられる。

色彩は感情に影響を与えるなど多彩な心理的な効果をもたらすので、そのような効果が何らかの形で単語の概念の同定などの過程で妨害をもたらしている可能性もある。

本研究の目的は彩色などの表現を用いて文書を読みやすく加工することにある。そのため、このような彩色による読解過程の妨害の効果については丁寧に検討していきたい。

#### 参考文献

- [1] 内田 友幸 田中 英彦: “色情報の付加によるテキストの可読性向上の試み” 情処第 50 回全国大会, 4N-6(March 1995).