

## 心的語彙の検索速度に影響する要因について (4)

久野 雅 樹

電 気 通 信 大 学

### 1 はじめに

本研究は、人間の認知システムにおける語彙処理のパフォーマンスと、単語自体にそなわっている特性とがどのようにからみあっているかを検討する一連の研究の第4報である。第1報(久野, 1995)では、3文字の和語動詞(e.g. ‘あおぐ’), 第2報(久野, 1996)では清音2字語(e.g. ‘アイ’), 第3報(久野, 1997a)ではカタカナ表記外来語(e.g. ‘アーケード’)を対象として、使用頻度、親近性、多義性などの単語がもつ様々な特性と語彙の検索速度の関連を検討した。本報告では対象をさらに拡張し、2文字の漢字熟語を素材として、心的辞書のしくみについて検討する。

語彙処理のパフォーマンスは、これまでと同様に、語彙判断課題(LDT=lexical decision task; ディスプレイに提示された綴りが単語であるか否かをできるだけ速く正確に判断する課題)の反応時間をその指標とした。単語特性のデータは、公刊されている調査類、新聞や辞書の電子化データに、自らの調査データを加えて用意した。一連の報告の基本的な発想として、精密な実験によって特定の仮説を検証するというよりも、様々な利用可能な情報を蓄積、統合し、データベース的な研究環境を整備することが念頭にある(久野, 1997b)。

### 2 方 法

#### (1) 素材

単語: 漢字2字熟語を1632用意した。素材選択にあたっては、従来の調査、研究で知見が蓄えられているものを使うことと、基本的な単語をなるべく網羅することを意図した。参照したのは、①梅本(1969)、②小川・稲村(1974)、③巖島ら(1991)、④情報処理振興事業協会技術センター(1996)、⑤国立国語研究所(1984)などである。②、③は、それぞれ400、600の漢字熟語についての特性調査であり、これらの素材は本研究でも全て素材として採用した。①、④は、広い視野で単語を集めていて利用可能な情報も多いので、見出しとなっている漢字熟語をほぼ全て用いることとした。以上に加え、日本語教育の資料である⑤で基本的な単語を補った。

語彙判断課題で使うのに、この1632語を4分して、408語ずつ、4系列の単語群を作成した。

非単語: 語彙判断課題においては、非単語が必要になる。この非単語は、漢字2字熟語のいずれか1文字について、別の漢字で置換して作成した(e.g. ‘愛情’ → ‘愛益’)。単語が4系列あるので、それぞれの系列に、別の系列の単語から作った非単語を割り当てた。作成時の実際の手続きとしては、おのおのの単語群で、1文字目と2文字目をランダムに組み合わせ直して、それが単語でないことを確かめた。

#### (2) 単語認識速度の測定: 語彙判断課題

被験者: 首都圏の大学生88名。4つの刺激系列に、ほぼ等人数を割り当てた。

#### 手 続 き

実験の制御はPC9801(NEC製)で行った。

個々の試行では、まず、ピープ音とともに注視点(+)がディスプレイの中央に500 msec 提示され、それが漢字2文字の綴りで置きかわる。被験者は、このターゲット刺激に対して、単語・非単語の判断をできるだけ

速く正確に行うことを求められる。反応はテンキーの‘1’と‘3’を用いて行う。ターゲットは最大 5,000 msec 提示され、被験者の反応とともに画面から消去される。そして、1,500 msec の試行間隔において、次の試行に移り、自動的に実験は進行する。

実験に当たっては、まず練習を 20 試行行った。練習では、被験者のセルフペースで進行できるようにし、被験者の反応後、正解（ターゲットが単語か非単語か）と反応の正誤を画面に示した。本試行は、24 セッション（1 セッションは 34 試行）で構成され、セッション間では休息をとれるようにした。

### (3) 単語の諸特性の指標

単語の特性について、以下の 23 の指標を算出あるいは参照した。

#### 1. 外的特徴による指標

- a. 新聞使用頻度(1 年分)：1994 年分の毎日新聞のテキストデータ（毎日新聞社、1995）に対して、RWC データベースワークショップ・テキストグループ（1996）がツールで形態素解析を施したものをを用いて使用頻度を集計。
- b. 新聞使用頻度(3000 文)：上記の毎日新聞のテキストデータは、その一部については、人手による修正が加えられた上で公開されている。頻度情報の精度について考えるために、これについても集計をおこなった。
- c. 仮名表記時の文字数：仮名書きした場合に、何文字になるかを調べた。
- d. 音節数：音節数についても確認した。

#### 2. 連想調査による指標

以下の 8 つの指標は、梅本(1969)が、210 の刺激語について、大学生 1000 人を対象として自由連想（刺激語 1 つにつき制限時間 10 秒）を求め、その結果を整理したものである。

- e. 無連想価：その刺激語に対して、連想反応を全然もたなかった者の数。
- f. 最多反応頻度：その刺激語に対してなされた連想反応の中で、最も多くの人が反応した言葉の頻度。
- g. 反応種類数：その刺激語に対してなされた反応の種類（異なり語数）。
- h. F 価 (From 連想)：発散連想。その刺激語から、調査で用いられた他の刺激語を想起した反応の総計。
- i. T 価 (To 連想)：収束連想。その刺激語が、調査リスト内のほかの刺激語に対する連想反応としてあらわれた頻度の総計。使用頻度、熟知度などに近い内容の特性であると考えられている。
- j. 名詞連想率：その刺激語に対して名詞を想起した者の比率。
- k. 形容詞連想率：同じく形容詞を連想した者の比率。
- l. 動詞連想率：同じく動詞を連想したものの比率。

#### 3. 主観的評定による指標

以下の 3 指標は、大学生を対象とした小川・稲村(1969)、巖島ら（1991）の報告による。評定はいずれも 7 段階で数値が大きいほどその特性を強くもっていることになる。各単語、各特性について、前者 30 人、後者は約 100 人の評定の平均を特性値としている。

- m. 心像性：心像性は、イメージの浮かびやすさで、小川・稲村は、調査のときの説明で、イメージの浮かびやすい語として‘馬車’を、浮かびづらい語として‘流動’を挙げている。
- n. 具象性：同じく、例を挙げれば、‘教科書’は具体的で‘発揮’はそうではない。
- o. 学習容易性：覚えなさいと言われたら、覚えやすいと思うかどうかによって評定したもの。

Table 語彙判断課題の反応時間と単語の諸特性の相関

	LDTの反応時間		a. 新聞使用頻度		o. 学習容易性		LDTの反応時間 aの影響を除く		LDTの反応時間 oの影響を除く		LDTの反応時間 a, oの影響を除く	
(1)外的特徴による指標												
a. 新聞使用頻度(1年分)	-0.508	*** 1630	——	1632	0.471	*** 988	——	1630	-0.400	*** 987	——	987
b. 新聞使用頻度(3000文)	-0.494	*** 1630	0.929	*** 1632	0.468	*** 988	-0.070	** 1630	-0.400	*** 987	-0.088	** 987
c. 仮名表記時の文字数	-0.016	1630	0.010	1632	-0.042	988	-0.011	1630	-0.009	987	-0.000	987
d. 音節数	-0.006	1630	-0.007	1632	-0.018	988	-0.009	1630	-0.004	987	0.003	987
(2)連想調査による指標												
e. 無連想値	-0.070	115	0.058	115	-0.175	+ 113	-0.044	115	-0.180	+ 113	-0.129	113
f. 最多反応頻度	0.172	+ 115	-0.223	* 115	-0.029	113	0.046	115	0.197	* 113	0.104	113
g. 反応種類数	-0.155	+ 115	0.206	* 115	-0.032	113	-0.037	115	-0.150	113	-0.070	113
h. F値 (From 連想)	-0.195	* 115	0.075	115	0.096	113	-0.189	* 115	-0.173	+ 113	-0.179	+ 113
i. T値 (To 連想)	-0.434	*** 115	0.442	*** 115	0.572	*** 113	-0.228	* 115	-0.185	+ 113	-0.099	113
j. 名詞連想率	-0.054	115	0.241	** 115	0.150	113	0.124	115	0.023	113	0.126	113
k. 形容詞連想率	0.107	115	-0.299	** 115	-0.005	113	-0.103	115	0.093	113	-0.053	113
l. 動詞連想率	-0.038	115	-0.088	115	-0.251	** 113	-0.117	115	-0.194	* 113	-0.202	* 113
(3)主観的評定による指標												
m. 心像性	-0.055	+ 987	0.004	988	0.386	*** 988	-0.070	* 987	0.241	*** 987	0.170	*** 987
n. 具象性	0.032	987	-0.039	988	0.299	*** 988	0.016	987	0.289	*** 987	0.229	*** 987
o. 学習容易性	-0.609	*** 987	0.471	*** 988	——	988	-0.474	*** 987	——	987	——	987
p. 親近性	-0.448	*** 231	0.196	** 231	0.359	*** 207	-0.417	*** 231	-0.366	*** 207	-0.378	*** 308
q. 主観的漢字表記頻度	-0.437	*** 101	0.445	*** 101	0.477	*** 55	-0.189	+ 101	-0.464	*** 55	-0.252	+ 55
r. 主観的ひらがな表記頻度	0.196	* 101	-0.076	101	-0.044	55	0.204	* 101	0.173	55	0.170	55
s. 主観的カタカナ表記頻度	0.179	+ 101	-0.096	101	-0.220	55	0.159	101	0.064	55	0.023	55
(4)語彙表参照による指標												
t. 見出し登録数	-0.118	*** 1630	0.094	*** 1632	0.099	** 988	-0.083	*** 1630	-0.078	* 987	-0.062	+ 987
u. 文字共有語彙数	-0.239	*** 1630	0.132	*** 1632	0.326	*** 988	-0.203	*** 1630	-0.084	** 987	-0.079	* 987
v. 1字目共有語彙数	-0.200	*** 1630	0.087	*** 1632	0.283	*** 988	-0.184	*** 1630	-0.061	+ 987	-0.075	* 987
w. 2字目共有語彙数	-0.222	*** 1630	0.100	*** 1632	0.258	*** 988	-0.200	*** 1630	-0.100	** 987	-0.094	** 987

Note: 各組み合わせ、左端の数字が相関係数 (か 偏相関係数), \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , +  $p < .10$   
 イタリックの小さ目の数字は相関係数算出に用いたデータ数である。

p.親近性: 各種の言語実験で最も基礎的な主観的特性。小川・稲村, 梅本のリストの一部を中心に, 大学生を対象として著者が行った調査による (7段階評定で, 数値が大きいほどなじみがある; 各語を 50 人が評定)。

q.主観的漢字表記頻度; r.主観的ひらがな表記頻度; s.主観的カタカナ表記頻度: 浮田ら (1996) は, 1 つの単語を漢字, ひらがな, カタカナの 3 つの表記で示し, それぞれどの程度見かけるか尋ね, 「よく見る」と答えた者の比率をその表記に対する主観的頻度としている (大学生を対象; 各単語, 201~216 人が評定)。

#### 4. 語彙表参照による指標

t.見出し登録数: 多義性に関する指標の 1 つということを主に意図して算出した。中野 (1996) による語彙表の電子データの索引を使い, それぞれの熟語が, 何度, 本文中に出てきているかを調べた。

u.文字共有語彙数; v.1 字目共有語彙数; w.2 字目共有語彙数: 似た綴りの語の数が, 単語認知に影響することが多くの研究者によって指摘されている。漢字熟語の場合, 構成要素は 2 字なので, 1 字目, 2 字目, いずれかを共有すれば, 近似語ということになる。v と w を足し合せたものが u である。

### 3 結果と考察

語彙判断課題 (LDT) の反応時間 (RT) は, 単語に対する正反応のみを対象とし, 外れ値 (200 msec 以下, 2,000 msec 以上, 各個人内で全反応の平均から  $\pm 2.5$  標準偏差 以上隔たった反応) は予め除いた。88 人の被験者のうち, 遂行成績に問題のある (有効試行が 7 割未満, もしくは, 単語・非単語の少なくとも一方で誤反応が 3 割以上) 10 人も, 今回の分析から除いた。また, 頻数データは, 分布の歪みを補正し, 極端な値の影響を

抑えるために、対数変換（0 をとりうる場合、特性値に 1 を加えて変換）した値を分析に用いた。

LDT の RT および諸特性の相関の一部を Table に示した。RT は、主観的特性である学習容易性と客観的特性である使用頻度と高い相関を示している。前者の学習容易性は、いわば単語の難易度であり、よく使い、よく知っている単語は早く認識されるという結果である。後者の使用頻度と RT の相関係数が .5 程度という数値は、前報のカタカナ外来語の際に、今回の b と同様の頻度集計をして、RT との間に .47 の相関係数を得ているのよりも、若干大きい。高相関が得られた理由として、漢字熟語の単語としての安定度と、実験に用いた素材が頻度に関して広く分布していることが挙げられる。なお、b の 3000 文のサンプルは、888,000 形態素で、1 年分の a の約 3000 万形態素から見るとずいぶん少ないのだが、両者に RT との相関係数の上での違いは見られなかった。100 万語ぐらいで頻度情報の精度は、本研究のようなねらいならば、ほぼ天井に達するということかもしれないが、b の形態素解析は、人手修正が加えられているということも、幾分か、今回の結果に影響しているだろう。

ほかに RT との相関が高いのは、親近性、T 価、主観的漢字表記頻度である。ただし、因子分析などで、諸特性の構造を検討する分析では、頻度、学習容易性、親近性、T 価は 1 つの因子を形成するが、主観的表記頻度はそれとは別の因子を形成する特性で独自の要素を多分にもっている。心像性と具象性は、RT そのものの相関はないのだが、学習容易性の影響を除いた偏相関は符号が正で有意となるのが興味深い。見出し登録数、文字共有語彙数は、相関係数自体は高くないが、RT 他と一貫して相関が認められる。なお、あまり予想していたわけではないが、仮名表記時の文字数、音節数は、RT を含め他の変数とほとんど関連が見られなかった。

#### 4 おわりに

今回は、23 の特性を取り上げたが、これは実際に分析の対象としたものの一部である。また、本報告で取り上げていても多義性のように満足できるレベルに至っていない特性があるし、分析に取り込んでいない変数ももちろんある。たとえば、漢字 1 字ずつの特性を考慮に入れることもできるだろう。かくして、この先、多くの側面から分析を深められる余地があることは、本研究のようなデータベース的相関研究の大きなメリットである。

なお、本報告の素材に関しては、ひらがな、カタカナで画面に提示して語彙判断を行った場合の分析も進めている。日本語における表記と認識との関連は、最近、ますます大きな関心が寄せられているトピックである。普通、漢字で書くものを仮名で見て、ひっかかったりしつつも、まあなんとか対処できるというのは、柔軟に働いている日本語認識システムについて考えるための格好の手がかりとなるであろう。

#### 引用文献

- 久野雅樹 1995 語彙の検索速度に影響する要因について 言語処理学会第 1 回年次大会発表論文集, 361-364.  
久野雅樹 1996 心的語彙の検索速度に影響する要因について (2) 言語処理学会第 2 回年次大会発表論文集, 409-412.  
久野雅樹 1997a 心的語彙の検索速度に影響する要因について (3) 言語処理学会第 3 回年次大会発表論文集, 181-184.  
久野雅樹 1997b 心理実験と日本語 日本語学, 16, 44-52, 明治書院.  
巖島行雄・石原治・永田優子・小池雇生 1991 漢字二字名詞 600 語の諸属性調査 - 心像性, 具象性, 学習容易性 - 日本大学心理学研究, 12, 1-19.  
情報処理振興事業協会技術センター 1996 計算機用日本語基本名詞辞書 I P A L (Basic Nouns) (第 3 版; FTP で公開)  
国立国語研究所 1984 日本語教育のための基本語彙調査 秀英出版.  
毎日新聞社 1995 CD-毎日新聞'94 日外アソシエーツ.  
中野 洋 1996 「分類語彙表」形式による語彙分類表 (増補版) 文部省科学研究費報告書.  
小川嗣夫・稲村義貞 1974 言語材料の諸属性の検討 - 名詞の心像性, 具象性, 有意味度および学習容易性 - 心理学研究, 44, 317-327.  
RWC データベースワークショップ・テキストグループ 1996 RWC-DB-TEXT-95-1, -95-2 メディアドライブ.  
浮田潤・杉島一郎・皆川直凡・井上道雄・賀集寛 1996 日本語の表記形態に関する心理学的研究 心理学モノグラフ, 25 日本心理学会.  
梅本堯夫 1969 連想基準表 - 大学生 1000 人の自由連想による - 東京大学出版会.