

心的語彙の検索速度に影響する要因について (3)

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の処理に単語自体の特性がどのように影響するか、心的なパフォーマンスが単語の特性という説明変数によってどの程度予測できるか、について検討する一連の研究の第3報である。第1報(久野, 1995)では、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’)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、カタカナ表記の外来語を素材として、心的語彙の処理について検討する。

日本語の語彙は、和語、漢語、(非漢語の)外来語、という3群に大きく分けられるが、カタカナ外来語は以下に記すような点で認知実験の素材として興味深く、また扱いやすい語彙群である。

- ・正書法が熟している：ひとつの語に対する表記法として漢字とかなの両方が存在することが多い和語とは異なる。このため、単語処理の自然なプロセスを見るのに都合がよく、考慮に入れなければならない特性が少なくてすむ(e.g. 久野, 1996の表記頻度というような特性を取り上げる必要がない)。
- ・語彙性が高い：名詞が中心で一つ一つの単語の独立性が高い。語彙性が高く、正書法が安定しているため、各種の特性値の概念が明快に定義しやすく、テキストコーパスなどを用いた計量的な処理も行いやすい。
- ・表記が音声と対応している：カタカナは表音文字である。表意文字である漢字を用いた語の場合、一般に、表記文字は意味をになっているが、それ自体が語の読みと直接に対応しているわけではない。カタカナ外来語においては、表記がそのまま音声をあらわしており、直接に意味には対応していない。このことから音声と意味との関連について検討するのに具合がよい。この点について、本研究では単語の多義性を取り上げて検討している。
- ・語彙の規模：一般に広く使われているカタカナ外来語は、せいぜい数千である。この一連の研究の背景として、言語素材による認知実験においては素材の特性を事前にコントロールすることは原理的に難しい、という認識があり、なるべく広くまんべんなく網をかけてデータを収集・蓄積することを目指している。この網をかけるユニットとして、カタカナ外来語の語彙は手頃である。
- ・使用の増大：カタカナ外来語の多用は、近年、いよいよ著しい。カタカナ外来語は、語彙の時代的な変化を考える上でも、重要なグループである。

なお、心的辞書の処理のパフォーマンスは、前報同様、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、単語特性のデータは、①公開されている調査、辞書資料、②主観的な評定による調査資料、に加え、③新聞、辞書のテキストコーパス、を利用して用意した。

2 方法

(1) 素材

単語

一般的に用いられているカタカナ外来語についてほぼ網羅することを念頭に素材を選択した。素材の抽出にあたっては、『岩波国語辞典 第5版』(西尾他 編, 1994; 実際の作業は イースト, 1995 の電子辞書検索ソフトを

用いて行った),『分類語彙表』(国立国語研究所,1964;実際の作業は 国立国語研究所, 1994 の電子化テキストによる),新聞の語彙調査である国立国語研究所(1971)の「度数順外来語表」などを参照した。明らかな複合語,単語としての独立性の低い語,カタカナ以外の表記も多用される語,難度の高い語,性的な語,などを除き,表記を標準的なものに調整・統合し,最終的に調査・実験の対象として 1600 語のカタカナ外来語を得た。語彙判断課題においては,この 1600 語を 4 分して 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,カタカナ外来語の 1 文字について,子音を置換することで作成した擬似単語である(e.g. ‘アーケード’ から ‘アーケーボ’)。単語が 4 群あるので,それぞれの群に,別の群の単語から作った非単語を割り当てた。

(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 141 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,それがカタカナ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 試行を行った。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,20 セッション(1 セッションは 40 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 9 つの指標を算出あるいは参照した。内容的には,語長,頻度,親近性,多義性,近似語量,の 5 種類に分けることができる。それぞれ,先行する諸研究において,単語の心的処理に対して影響することが指摘されているものである。語長,頻度,親近性については,それらが単語の処理に影響するのは,現象の存在自体は推測しやすく,問題はむしろ影響の程度の推定であると言える。多義性については,多義語の方が処理が速いとする研究(e.g. Millis & Button,1989),近似語量(neighborhood size)については,音韻・綴りを共有する近似語が多い単語の方が認識が速いとする研究(e.g. Andrews,1989)がある。

a.語長:単語の表面的な特性として,文字数を算出した(2~9 文字)。

b.雑誌頻度'56:使用頻度については,調査対象,調査時期,調査方法の異なる 3 通りの指標を用意した。まず雑誌における頻度調査データとして,国立国語研究所(1962)の「五十音順語彙表」を参照した。これは,1956 年の時点における雑誌 90 種を対象とするものである。頻度はパーミルで記載されているが,これを 1000 倍して,100 万語あたりの度数として処理した。

c.新聞頻度'66:新聞における頻度調査データとして,国立国語研究所(1971)の「度数順外来語表」の度数を参照した。これは,1966 年の朝日,毎日,読売の 3 新聞の頻度調査である。

d.新聞頻度'94:1994 年分の毎日新聞のテキストデータ(毎日新聞社,1995)について,RWC データベースワー

Table 語彙判断課題の反応時間と単語の諸特性の相関

	LDTのRT	a. 語長	b. 頻度'56	c. 頻度'66	d. 頻度'94	e. 親近性	f. 評定多義性	g. 辞書多義性	h. 近似語	i. 重みづけ近似語
LDTの反応時間			-0.229	-0.351	-0.409	-0.566	-0.299	-0.149	-0.037	-0.079 +
a. 語長	0.561									
b. 雑誌頻度'56	-0.286	-0.178		0.452	0.343	0.175	0.060 *	0.053 *	0.046 -	0.044 -
c. 新聞頻度'66	-0.401	-0.207	0.472		0.478	0.314	0.152	0.172	0.105	0.120
d. 新聞頻度'94	-0.474	-0.263	0.373	0.505		0.408	0.196	0.144	0.045 -	0.065 +
e. 親近性	-0.506	-0.069 +	0.184	0.321	0.411		0.329	0.119	0.065 +	0.089 *
f. 評定多義性	-0.400	-0.290	0.108	0.202	0.257	0.334		0.381	0.289	0.334
g. 辞書多義性	-0.240	-0.213	0.089 *	0.209	0.192	0.130	0.418		0.205	0.378
h. 近似語量	-0.380	-0.635	0.147	0.211	0.200	0.094 *	0.398	0.290		0.957
i. 語義重みづけ近似語量	-0.406	-0.633	0.146	0.222	0.214	0.112	0.431	0.421	0.974	

Note: 各数値の右肩に印がないものは $p < .0001$; 他は, * $p < .001$, + $p < .01$, * $p < .05$, - $p \geq .05$

対角左下が2変数間の相関係数; 右上のイタリック体の数値は, 語長の影響を除いた偏相関係数。

クショップ・テキストグループ(1996)が形態素解析を加え, 人手修正を施した3000記事分を対象に頻度の集計をおこなった。

e. 親近性: 大学生23人の評定値の平均。評定は7段階(数値が大きいほど親近性が高い)。

f. 評定多義性: それぞれの表記に対して, いくつの意味があると思うかを, 13人の大学生に評定してもらい, その平均をそれぞれの単語の多義性とした。

g. 辞書多義性: 『岩波国語辞典 第五版』(西尾他 編, 1994)でいくつの数の語義が挙げられているかをカウントした。『岩波国語辞典 第五版』に見出しのないもの(主に固有名詞)については, 語義数を1とした。

h. 近似語量: 今回対象としたカタカナ語では, たとえば, ‘アクセル’に対して, ‘アクセス’ ‘アクリル’がそれに当たる。『岩波国語辞典 第五版』において, 当該語と1字のみ異なる表記の語(カタカナ語; 見出しが複数でも表記が同一なら一語とした)の数をカウントした。

i. 語義重みづけ近似語量: 近似語量を多義性の程度によって重みづけしたものである。ここでは, 『岩波国語辞典 第五版』に記載されている語義数をカウントすることで求めた。

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 ± 2.5 標準偏差 以上隔たった反応)は予め除いた。また, 3つの使用頻度, 辞書多義性, 2つの近似語量の6指標については, 分布のゆがみが大きいので, それぞれ対数変換(辞書多義性以外は0をとりうるので, 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9つの特性値の相関行列をTableに示した。RTは語長と最も高い相関を示している。特定の語長の単語群を対象とした前二報では, 得られなかった知見である。使用頻度とRTの相関は, 新聞頻度'94, 新聞頻度'66, 雑誌頻度'56と古くなるにしたがって低くなる。これは集計手続きやメディアの差も無論関与しているだろうが, 時代変化を多分に反映していると思われる。なお, 新聞から使用頻度を求めるのに, さらに大きなコーパスを用い, 細かい修正を加えて(表記のゆれや, 複合語への対処など)計算することが望まれるが, それをしなくとも, かなりの精度の情報が得られることがわかったのも大きな収穫である。さらに, 新聞という特

定の活字メディアに限った出現頻度が、RT との間に、親近性の評定特性値に近い相関を示していることも重要である。これは、前二報では見られなかったこと (LDT の RT と新聞頻度'66 の間の相関は何通りか出したが、.25 に満たない) で、カタカナ外来語の語彙性の強さと表記の定着が結果をクリアなものにしているのだと思われる。多義性、近似語量の効果が本研究では明確に出たことも、同様の理由によるところが大きいだろう。

特性相互の構造をつかむために、因子分析 (主成分分解、バリマックス回転) をおこなったところ、4 因子で、以下のような解釈しやすい結果が得られた (第 4 因子まででの累積寄与率 78.1%)。

第 1 因子 = 語彙構造の密度の因子 (語長、近似語数、語義重みづけ近似語数)

第 2 因子 = 頻度 (雑誌頻度'56, 新聞頻度'66, 新聞頻度'94)

第 3 因子 = 親近性 (親近性)

第 4 因子 = 多義性 (評定多義性、辞書多義性)

こうした結果もふまえ、LDT の RT を、単語の諸特性から予測する重回帰分析を行い、各種のモデルを検討した結果、現段階で妥当と思われる回帰式として、次のようなものが得られた。

$$RT = 0.461 \cdot \text{語長} - 0.369 \cdot \text{親近性} - 0.177 \cdot \text{新聞頻度'94} - 0.097 \cdot \text{評定多義性}$$

(標準偏回帰係数はいずれも $p < 0.0001$ で有意 ; $R^2 = 0.568$)

4 おわりに

今回、カタカナ外来語を素材としたことで、表記、語形変化、語の複合など、日本語を対象とした認知研究がしばしば直面する微妙な問題とあまり取り組まずにすみ、かなりすっきりとした結果を得ることができた。現在、次のステップとして、2 文字の漢字熟語 (基礎的なものが中心) を対象にデータの分析、検討を行っている。漢語は、語彙性が強く、表記が定着しているという点では、カタカナ外来語と同様であるが、表記-音韻-意味の対応システムがカタカナ外来語と異なる語彙群であり、日本語の独自性を生かした知見の蓄積が期待できる。

引用文献

- Andrews, S. 1989 Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 802-814.
- イースト 1995 ViewDic 岩波国語辞典 イースト.
- 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.
- 国立国語研究所 1962 現代雑誌九十種の用語用字 第一分冊 総記および語彙表 秀英出版.
- 国立国語研究所 1964 分類語彙表 秀英出版.
- 国立国語研究所 1971 電子計算機による新聞の語彙調査(II) 秀英出版.
- 国立国語研究所 1994 [フロッピー版] 分類語彙表 秀英出版.
- 毎日新聞社 1995 CD-毎日新聞'94 日外アソシエーツ.
- Millis, M. L., & Button, S. B. 1989 The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition*, 17, 141-147.
- 西尾 実・岩淵悦太郎・水谷静夫 編 1994 岩波国語辞典 第五版 岩波書店.
- 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-95-2 (RWC テキストデータベース に収録 ; 実際の公開は 1997) メディアドライブ.