

日本語約5万語の心像性データベースの作成

佐久間尚子¹(さくまなおこ)、伊集院睦雄¹、伏見貴夫¹、辰巳格¹、

田中正之²、天野成昭³、近藤公久^{3,4}

(財)東京都老人総合研究所¹

京都大学霊長類研究所²、NTTコミュニケーション科学基礎研究所³、現NTTデータ⁴

(要旨) 我々は日本語約5万語の文字提示、音声提示による心像性評定実験を行い [9,10]、心像性データベースを作成した。その際、複数の意味をもつ単語については多くの人が最初に想起する意味を特定した。提示モダリティや属性間の関係を分析したところ、心像性の文字と音声の相関は非常に高く、心像性と単語親密度の相関も高かったが、親密度が高くなるにつれ心像性の異なる語が現れた。心像性と単語頻度の相関は低かった。表記別では片仮名語の心像性が高く、品詞別では名詞の心像性が高かった。

Key words: 単語心像性、単語親密度、単語頻度、データベース

1. はじめに

心像性 (imageability) とは、“単語の意味する内容を心的イメージとして思い浮かべる際の想起のしやすさ”を人に評定させた評定値であり、意味に関連する単語属性の代表として頻繁に用いられている [7,10]。我々は、今回、文字データと音声データの両方をもつ日本語の大規模な心像性データベースを作成した。単語の中には、同音異義語 (例、/kaze/は風、風邪など) や同字異音語 (例: 「一味」は /ichimi/、/hitoaji/)、多義語 (例: 「さくら」) があり、これらの意味の複数ある語については以下に述べる選択肢を用いて、各人の評定した意味を特定した。以下に、心像性データベースの概要を述べ、分析結果の一部を報告する。

2. 心像性評定実験

方法 単語親密度資料 [1] やその他の資料を参考に、文字単語 48,714 語 (主に親密度 4 以上の名詞、動詞、形容詞、副詞等で異表記を含む) を選定し、文字提示実験を行った。さらに 48,714 語中、対応する音声ファイル [2] のあった 39,447 音声 (異型アクセント音声を含む) を用いて音声提示実験を行った。参加者は、日本語を母語とする東京アクセントの大学生等計 64 名 (平均 21.8 歳) であり、「単語の意味内容を感覚的 (心的) にイメージする際の容易さ」を 7 段階 (1: 非常にイメージしにくい ~ 7: 非常にイメージしやすい) で評定 [8] させた。なお、提示された語が知らない文字表記ないし音声の場合は、“知らない” (以下、DK) と評定させ、心像性評定をさせなかった。

選択肢 実験に先立ち、複数意味の想起されうる単語を調べ、文字単語 4,147 語と音声単語 8,330 音声に対し、意味を尋ねる選択肢を作成した。各語の評定直後に選択肢を提示し、評定時に想起された意味を選択させた。なお、選択肢に該当する意味がない場合は、想起した意味を用紙に記録させた。

データの採用 本実験の評定値と後テストの評定値の相関が 0.5 以上だった文字実験 35 名と音声実験 31 名のデータを採用した。

3. 心像性データの集計

基本 DK 反応を除く評定値の平均を求めた。

選択肢別集計 各語について、選択肢別の選択肢数と平均評定値を求め、最も多くの人を選んだ選択肢をその語の第 1 意味とした。

文字と音声データの統合および共通 ID の付与

データベースの検索の利便性を考え、最初に心像性の文字データと音声データに対応する単語親密度資料 [1] の “共通 ID” を付与した。これらの中には表記ないし音声の同じ別の語 (例、「あれ」は代名詞と感動詞の 2 つの “共通 ID” がある) があり、選択肢のある項目については、第 1 意味と整合性のある共通 ID を選んだ。意味の整合性を判定するこれらの作業は、成人 3 名が独立に行い、判定の不一致が生じた場合は、協議して最終的な採否を決めた。

4. 心像性データベースの概要

本データベース 表記と音声 (アクセント別) と先の “共通 ID” の組み合わせが異なる項目を並べ、総数 54,224 項目のデータベースを作成した。これらの中には文字ないし音声データのみ項目もあり、両データのある項目は 42,660 項目であった。

付録データベース 文字単語 4,074 語 (注) における合計 9,350 の選択肢、および音声単語 8,330 語における合計 20,440 の選択肢とその選択順位、平均評定値からなる。

(注) 文字実験は第 1 期、第 2 期からなり、第 2 期に選択肢の一部を変えた。

5. 心像性データベースの分析

5.1 文字単語心像性と音声単語心像性の相関

心像性の提示モダリティ - 効果を検討するため、本データベース中、同音異義語、同字異音語、多義性のない 21,690 語を抽出した。文字単語心像性と音声単語心像性の平均値は、両者とも約 4.0 だった (表 1)。散布図 (図 1) を見ると、文字と音

	平均値	標準偏差	相関係数
文字心像性	4.04	0.804	0.912
音声心像性	3.93	0.864	

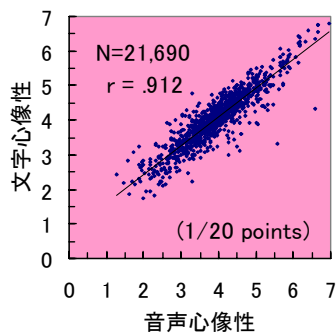


図1 心像性の文字と音声の相関

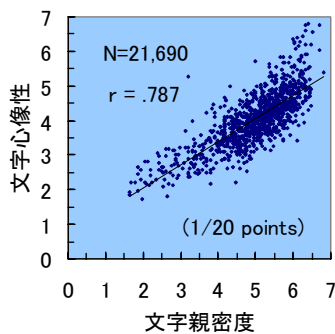


図2 心像性と親密度の相関

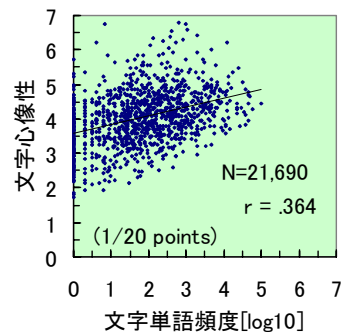


図3 心像性と頻度の相関

声で心像性の異なる語(例、「蜜柑、実父」)もあるが、相関係数は $r = .912$ と非常に高かった。

5.2 心像性と親密度と頻度の相関

先の21,690語における、文字単語心像性と、文字単語親密度[1]、単語頻度[4]との相関を示す(図2, 3)。単語頻度は読みを区別した表記頻度を対数変換した値を用いた。心像性と親密度の相関($r = .787$)は高いが、単語頻度との相関($r = .364$)は低い。この21,690語の文字単語親密度と単語頻度の相関は $r = .573$ だった。図2を見ると、親密度が高くなるにつれて、心像性の幅が広がる傾向が見える。すなわち、親密度の高い語には心像性の高い語(例、「牛乳」)から低い語(例、「経済」)まであり心像性の差が現れる。心像性評定値は親密度に影響されるが、心像性は親密度とは異なる単語属性であることが示唆される。

5.3 表記と品詞の影響

近年、名詞以外にも動詞や機能語の心像性に関心が高まっている[5,6]。先の21,690語について品詞[3]を検索し、表記と品詞の分布を検討した。図2に示されるように、親密度の低い単語では、心像性はほぼ親密度と共に低くなり、心像性の差が見えにくくなる。ここでは、文字単語親密度4.96(21,690語全体の文字単語親密度の平均値)以上の12,623語に限定して、表記と品詞別の単語数を数え(表2)、その平均文字単語心像性(表3)を比較した。なお、表2, 3の右端の総計欄には、21,690

表2 表記と品詞別の単語数 (文字単語親密度4.96*以上)

表記	品詞					合計	総計
	名詞	動詞	形容詞	副詞	他		
カタカナ	2153		61	2	38	2254	2876
ひらがな	251	162	95	364	126	998	1323
漢字	5649		178	21	399	6247	12446
記号		29				29	49
混合	1833	847	231	108	76	3095	4996
総計	9915	1009	565	495	639	12623	21690

表3 表記と品詞別の文字単語心像性の平均値 (文字単語親密度4.96*以上)

表記	品詞					合計	総計
	名詞	動詞	形容詞	副詞	他		
カタカナ	4.89		4.37	4.16	4.41	4.86	4.59
ひらがな	4.78	4.48	4.40	4.30	4.33	4.46	4.31
漢字	4.36		3.86	4.03	4.22	4.33	3.82
記号		4.73				4.73	4.21
混合	4.47	4.47	4.45	3.89	4.38	4.45	4.20
総計	4.51	4.47	4.25	4.20	4.27	4.47	4.04

*21,690語全体の文字単語親密度の平均値

語の全体に占める表記別単語数(表2)と平均文字単語心像性(表3)を付記した。

単語数(表2)は、表記や品詞によってばらつきがあり、文字単語心像性の平均値を求めるのに十分でないセルもあるが、表3の平均心像性を見ると、心像性が高いのは、表記別では、カタカナ、記号、ひらがな、混合(複数の表記からなる語: 例、「生ビール、思う、正しい」)、漢字の順であり、品詞別では、名詞、動詞、その他、形容詞、副詞の順だった。品詞の中で名詞の心像性が高いのは英語圏のデータ[5,6]に一致する。また、表記や品詞によって心像性評定値の範囲が異なり、たとえば、文字単語心像性の範囲は、名詞が2.66~6.94であるの対し、動詞は3.29~5.77と狭く、名詞以外の品詞では心像性の高い語が少ないことがわかった。これらのことから、品詞効果を検討する際にも心像性を考慮することが必要であると示唆される。

6. おわりに

心像性データベースの概要を述べ、分析の一端を紹介した。心像性は意味に基づく単語属性であるため、可能な限り、評定時の意味を特定して、意味の整合性のあるデータベースを作成した。今後さらに分析を進め、本データベースの有効性と問題点について検討する。

文献

1. 天野、近藤: 日本語の語彙特性1; 単語親密度
2. 天野、近藤: 日本語の語彙特性3; 単語アクセント
3. 天野、近藤: 日本語の語彙特性4; 品詞
4. 天野、近藤: 日本語の語彙特性7; 頻度 (以上、三省堂。1-3:1999. 4:2000.)
5. Bird, Franklin, Howard: Behavior Research Methods, Instruments, & Computers, 33, 73-79, 2001.
6. Chiarello, Shears, Lund: Behavior Research Methods, Instruments, & Computers, 31, 603-637, 1999.
7. 伏見: 第4回認知神経心理学研究会、2001.
8. Paivio, Yuille, & Madigan: *Journal of Experimental Psychology*, 76, 1-25, 1968.
9. 佐久間、伊集院、伏見、田中、辰巳、天野、近藤: 第2回認知神経心理学研究会、1999.
10. 佐久間、伊集院、伏見、田中、辰巳、天野、近藤: 信学技報 TL2000-15, 9-16, 2000.