

「日本語の語彙特性」を利用した認知実験研究 - 単語刺激の統制と非語刺激 -

近藤公久(こんどう ただひさ)、天野成昭

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

(要旨)我々は日本語の単語および文字の様々な特性を集めたデータベースを構築した[1,2,3,4,5,6,7]。本データベースを用いれば単語や文字の様々な特性を統制した実験を行うことが可能である。本発表では、本データベースを用いた実験を紹介する。また、音声単語親密度でランク分けしたとき、それぞれのランクに属す単語に出現するモーラおよびバイモーラの頻度分布を分析した結果から、超低親密度語を非語として用いることの妥当性を示す。

Key words: データベース、単語親密度、バイモーラ頻度、刺激の統制、非語

1. はじめに

本発表では、日本語の語彙特性データベース[1-7]を用いて単語および非単語刺激の選択および統制を行う方法について、これまで筆者らが実際に行った単語認知実験を例としてあげながら紹介する。また、被験者の言語レベルの統制方法についても紹介する。

2. 単語刺激の属性の統制

言語認知実験において、刺激として用いる単語の属性を統制することは非常に重要である。たとえば、各単語に対する「なじみ」の程度を表す主観的評定値である単語親密度[1]は単語の属性の1つであり、単語認知の容易さに強い影響を与えることが、語彙判断時間、音読潜時、単語認識率等において確認されている[1,8,9,10]。一方、単語認知に影響を及ぼす属性は、単語親密度だけではない。たとえば、表記妥当性[11]、頻度[7,12]、心像性[13,14]、獲得年齢[15,16]の影響も確認されている。また、このような単語の属性だけでなく、文字属性や音韻・音響属性も単語認知に影響を及ぼす。例としては、文字の親密度や頻度、漢字の読みの妥当性、複雑度、あるいは、アクセント、モーラ頻度、バイモーラ頻度などがあげられる。

以上であげた単語認知に影響を及ぼす属性を統制するには、「日本語の語彙特性」データベース[1-7]に収録された特性値、および、これらから計算されるさまざまな値を駆使すれば可能である(但し、心像性と獲得年齢については未公開データである)。しかし、可能であることと、実際の刺激セットを容易に作成できることとは必ずしも同意ではないこ

とに注意が必要である。

3. 非語刺激としての超低親密度語

言語認知実験においては、単語と共に非語を刺激に用いることが必要となる場合がある。これまでの実験では刺激に用いる非語を、音韻列のランダム発生、あるいは単語内の音韻の逆順化や置換で得ることが多かった。しかし、このような操作で得られる非語は、単語との偶発的一致や類似の影響、および音韻列としての日本語らしさが失われるという問題があった。ここではこれらの問題を回避する方法として、超低親密度語を非語として用いる方法の妥当性を検討する。

3.1 単語親密度と既知率

知らない単語はその人にとっては非語であるといえる。Amano, Kondo[17]において、音声単語親密度が2.0で大学生の既知率は約15%、2.5で約30%、3.0で約50%、4.0で約80%であった。よって、単語親密度が2.0から2.5以下の単語であれば非語として使用することが可能なレベルであると考えられる。

3.2 モーラの出現頻度と単語親密度

表1に単語親密度ランク間のモーラの出現頻度の相関を示す。頻度の計算方法などの詳細は天野, 近藤[18]を参照。結果から単語親密度ランク間のモーラの出現頻度の相関が高いことがわかる。すなわち、低親密度語であっても高親密度語とモーラの出現頻度分布が非常に類似していることが示唆される。

3.3 バイモーラの出現頻度と単語親密度

表2に単語親密度ランク間のバイモーラの出現頻度の相関を示す。頻度の計算方法などの詳細は同じく近藤, 天野[18]および天野,

近藤 [19]を参照。結果から、単語親密度ランク間のバイモーラの出現頻度の相関が高いことがわかる。すなわち、低親密度語であっても高親密度語とバイモーラの出現頻度分布がかなり類似していることが示唆される。

以上の分析から、超低親密度語（たとえば親密度 2.0 以下）は、語彙としてほとんどの人に知られていないが、日本語の音韻特徴をある程度保持しているといえる。これは、超低親密度語を非語として実験に用いることの妥当性を示している。しかし、刺激が漢字表記である場合には問題が生じる。なぜならば、超低親密度語の漢字表記は非常に難しく読めないことが多いからである。この場合には、必要に応じて易しい漢字への置き換えを要する。しかし、漢字を入れ替えると漢字の組み合わせによって造語として意味解釈が可能となることもあり、この操作は容易ではない。

4. 被験者の言語レベルの統制

言語認知実験においては、刺激の属性の統制とともに被験者の言語能力の統制も重要なポイントである。近藤、天野 [2,20]は、被験者の言語能力の統制のために漢字単語の読み能力テスト「百羅漢」を開発した。

百羅漢の得点は、大学生 1410 名の平均が 55.1 点(100 点満点)、標準偏差 14.8、最低点 16、最高点 88 であり、十分分散している(近藤、天野 [21])。また、天野、近藤 [22]では百羅漢の得点と語彙数の推定値間に中程度の相関があることを確認している。さらに、Kondo, Wydell, Amano [23]では、百羅漢の得点が音読時や語彙判断時間に影響を及ぼすことを示している。以上の結果から、百羅漢は言語認知実験の被験者の言語能力の統制に有効なテストであるといえる。経験からいえば、安定した結果を得るには最低でも 60 点以上の被験者を使用すれば良いだろう。

5. おわりに

本稿ではスペースの都合で掲載できなかったが、日本語語彙特性データベースの検索の実例や単語属性と単語認知モデルについても時間の許す限り紹介する予定である。

文献

1. 天野, 近藤 “日本語の語彙特性1 親密度”.
2. 近藤, 天野 “日本語の語彙特性2 表記”.
3. 天野, 近藤 “日本語の語彙特性3 アクセント”.
4. 天野, 近藤 “日本語の語彙特性4 品詞”.
5. 近藤, 天野 “日本語の語彙特性5 文字特性”.
6. 近藤, 天野 “日本語の語彙特性6 文字-単語”.
- *1-6はすべて、(三省堂, 東京, 1999) .
7. 天野, 近藤 “日本語の語彙特性7 頻度” (三省堂, 東京, 2000) .

8. 加藤, 天野, 近藤 (1999) “雑音を負荷した音声の単語了解度に対する親密度の影響” 聴覚研究会資料, H-99-8.
9. 天野, 近藤 (1998) “音声単語の語彙判断に対する親密度の影響” 音響学会春, 1.
10. 近藤, Wydell (2001) “漢字 2 文字単語の読みの過程と単語および漢字の特性との関係” 認知科学会.
11. Ainsworth-Darnell & Kondo “Beyond orthographic depth: Similarities in the processing of words in Kanji and Hiragana” The Annual Meeting of the Linguistic Society of America, New York (1998).
12. 加藤, 天野, 近藤 (2001) “音声単語認知における親密度と頻度の影響” 音響学会春, 1.
13. 佐久間, 田中, 伏見, 伊集院, 辰巳, 天野, 近藤 (2000) “文字呈示, 音声呈示による日本語約 5 万語の心像性評価” 思考と言語研究会資料.
14. 佐久間, 田中, 伏見, 伊集院, 辰巳, 天野, 近藤 (2000) “48 カテゴリーの語想起” 日心大会.
15. 近藤, 天野 (2000) “日本語単語の獲得年齢と親密度の関係” 音響学会春, 1.
16. 近藤, 天野, 柴原, 佐久間, 伊集院, 伏見, 辰巳, 田中 (2000) “日本語単語の親密度、心像性、獲得年齢と読みの過程の関係” 認知科学会.
17. Amano & Kondo (1998, December) “Estimation of mental lexicon size with word familiarity database” Proceedings of International Conference on Spoken Language Processing, 5, 2119-2122.
18. 天野成昭, 近藤公久 (2001). “単語親密度ランク間におけるバイモーラ頻度の相関” 音響学会秋, 1.
19. 近藤公久, 天野成昭 (2001). “単語親密度ランク間におけるバイモーラ頻度の相関” 音響学会秋, 1.
20. 近藤, 天野 (1998) “漢字単語の読み能力テスト” 日心大会.
21. 近藤, 天野 (2001) “漢字単語の読み能力テスト「百羅漢」の得点傾向” 日本心理学会.
22. 天野, 近藤 (2000) “漢字単語の読み能力と語彙数” 日心大会.
23. Kondo, Wydell, & Amano (1998, March) “Language processing and reading skills in Japanese.” Poster session presented at the CUNY Conference on Human Sentence Processing, New Brunswick, NJ.

表 1 モーラ頻度の単語親密度ランク間での Pearson の相関係数 (N=130)

親密度	1~2	2~3	3~4	4~5	5~6
2~3	.976				
3~4	.963	.981			
4~5	.953	.972	.995		
5~6	.912	.931	.973	.984	
6~7	.891	.867	.909	.916	.942

表 2 バイモーラ頻度の単語親密度ランク間での Pearson の相関係数 (N=6224)

親密度	1~2	2~3	3~4	4~5	5~6
2~3	.904				
3~4	.867	.945			
4~5	.854	.918	.953		
5~6	.799	.869	.916	.951	
6~7	.761	.777	.801	.810	.848