

## 言語認知研究における NTT 日本語データベース 「日本語の語彙特性」の有効性

近藤公久(こんどう ただひさ)、天野成昭

NTT コミュニケーション科学基礎研究所

(要旨) 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]。本データベースには、新明解国語辞典[7]の約 8 万の見出し語に対する単語親密度、単語表記の妥当性などと JIS[8]に規定される約 7 千の文字に対する文字親密度、複雑度などが収録されている。それぞれのデータベースを用いて刺激の統制を行った認知実験結果から、データベースに収録されている特性値の有効性が示された。

Key words: 日本語データベース、単語親密度、単語表記、文字特性、漢字の読み

本稿では、言語認知研究における、NTT 日本語データベース「日本語の語彙特性」の有効性を示す実験結果を紹介する。各実験の詳細は、それぞれに対応する文献および日本語の語彙特性 1 巻から 6 巻の解説部を参照していただきたい。

### 親密度効果

単語の認知過程は単語の親密度の影響を受ける。単語親密度データベース[1]を用いて刺激を統制した、語彙判断反応時間(図 1) [9]、および雑音中の単語認知率(図 2) [10]において、統計的に有意かつ大きな親密度効果が示された。

### 言語能力推定テスト百羅漢の得点と音読潜時

言語認知過程は個々人の言語能力の違いの影響を受ける。そこで、言語能力を推定するための漢字単語の読みテスト[11]を単語親密度データベース[1]構築の予備的検討データ[12]から作成した。図 3 に、漢字単語の親密度ランクおよび被験者の百羅漢得点ランク毎の音読潜時を示す[13]。漢字単語の音読潜時においても大きな親密度効果がみられるとともに、百羅漢テストの得点ランク間で音読潜時に統計的に有意な差が見られた。

### 単語表記の妥当性効果

文字単語の認知過程は単語の親密度のみならず、その単語の表記形態にも影響を受ける。漢字表記優位語(漢字語)と漢字・ひらがな表記同等語(同等語)を単語表記データベース[2]から各 180 語検索し、これらを漢字および「ひらがな」

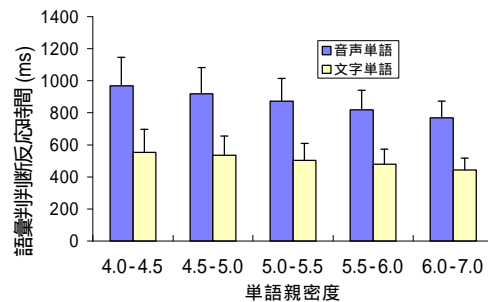


図 1 音声および文字単語親密度と語彙判断反応時間の関係  
被験者は 28 名、18-30 才。百羅漢得点 70-80。

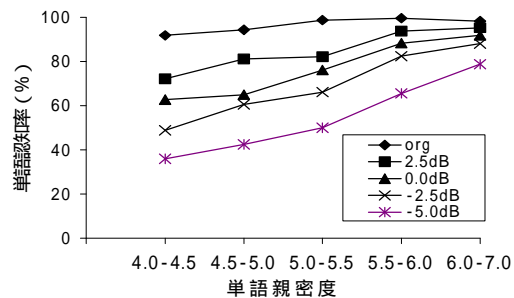


図 2 音声単語親密度と音声単語認知率の関係  
被験者は 40 名、18-30 才。百羅漢得点 70-80。

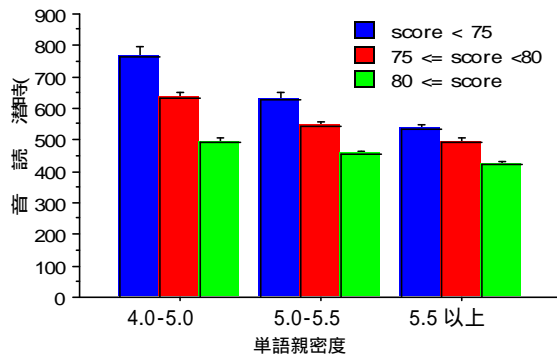


図 3 親密度ランクおよび百羅漢得点ランク毎の音読潜時(ms)  
各得点ランクの被験者数は、得点の低い方から、6,8,6 名。  
音読潜時は遅延音読潜時を引いた値。

で呈示した際の語彙判断反応時間を図 4 に示す[14]。漢字語は漢字呈示で語彙判断反応時間が有意に短く、同等語では差が見られなかった。この結果は、漢字表記とかな表記の違いによる語彙判断反応時間の差が表記の妥当性によって説明可能であることを意味する。

### 漢字の読みの妥当性・一貫性効果

Wydell and Kondo[15]は、文字 - 単語データベース[5,6]を用いて漢字の読みの一貫性(consistency)を計算し、漢字 2 文字の 1000 語に対する被験者 20 名の音読潜時を測定した。その結果、音読み 2 文字漢字単語における漢字の読みの一貫性効果が確認された(図 5)。この結果は、伏見ら[16]の結果と一致している。また、漢字の読みの妥当性および音読み訓読み[5]の影響が確認された(図 6)[13]。

### まとめ

NTT 日本語データベースの言語認知研究への適用例とその有効性を示した。本データベースが言語認知研究の発展に寄与することを期待する。

### 文献

1. 天野,近藤 “日本語の語彙特性1 単語親密度”。
2. 近藤,天野 “日本語の語彙特性2 単語表記”。
3. 天野,近藤 “日本語の語彙特性3 単語アクセント”。
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、NTT日本語データベースシリーズ、(三省堂、東京、近刊)。
7. 金田一,柴田,山田,山田 (編) “新明解国語辞典 第四版、(三省堂、東京、1989)
8. 日本規格協会 “情報交換用漢字符号 JIS X 0208-1990” (1990)。
9. 加藤,天野,近藤 “雑音を負荷した音声の単語良貨緯度に対する親密度の影響” 聴覚研究会資料、H-99-8 (1999)。
10. 天野,近藤 “音声単語の語彙判断に対する親密度の影響” 音響学会春講演論文集、I。
11. 近藤,天野 “漢字単語の読み能力テスト” 日心大会論文集、711 (1998)。
12. Amano, Kondo, & Kakehi “Modality dependency of familiarity ratings of Japanese words”, P&P 57(5), (1996)。
13. Kondo & Wydell (submitted)。
14. 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)。
15. Wydell & Kondo “Computation of phonology in Kanji words: What matters-consistency or ON/KUN difference” ICCPAL, Nagoya (1997)。
16. Fushimi, Ijuin, Paterson, & Tatsumi “Consistency, frequency, and lexicality effects in naming Japanese Kanji” JEP:HPP (1999)。

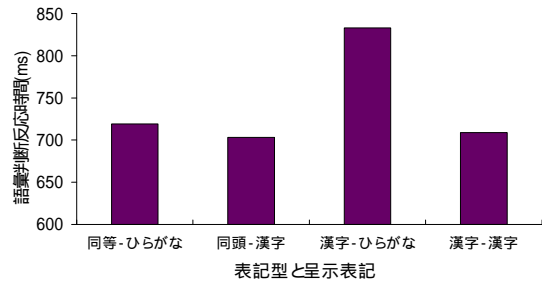


図 4 単語の表記型と呈示表記の関係と語彙判断反応時間  
被験者は 20 名、20-28 才。百羅漢得点 67-90。  
同等語および漢字語順に、ひらがな表記の妥当性は 3.28, 4.84 漢字表記の妥当性は、2.38, 4.91。

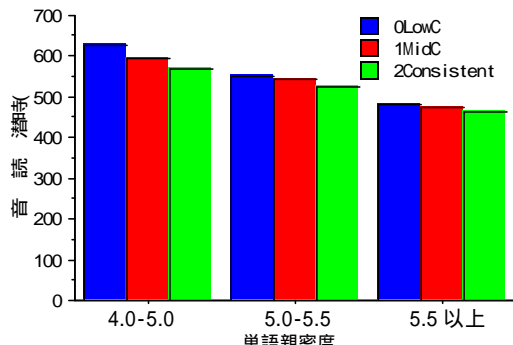


図 5 音読み 600 語に対する親密度ランク毎の一貫性効果  
一貫性のランクは 0LowC:60 未満、1MidC:60 以上 100 未満、2Consistent: 100 である。被験者は図 4 の実験と同じ。

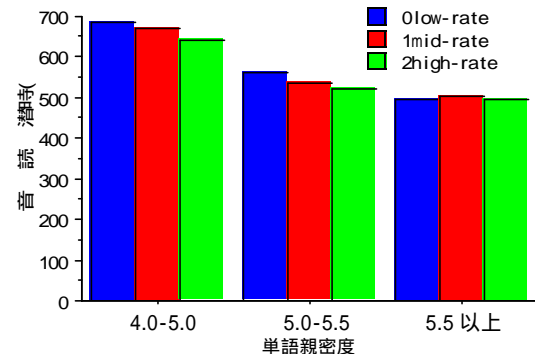
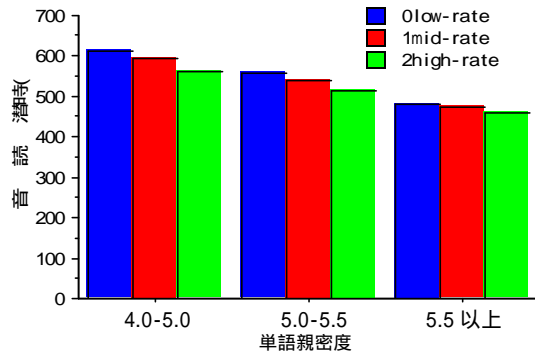


図 6 音読み 600 語(上)、訓読み 400 語に対する親密度ランク毎の漢字の読みの妥当性効果  
読みの妥当性のランクは 0low-rate:6 未満、1mid-rate:6 以上 6.5 未満、2high-rate: 6.5 以上である。  
被験者は図の実験と同じ。