

メンタル・レキシコン研究(XVI)

漢字二字熟語の意味的透明性のデータベース構築に向けて (1)

○増田尚史¹・Terry Joyce²・小河妙子³・藤田知加子⁴・川上正浩⁵

(¹広島修道大学人文学部・²多摩大学・³東海学院大学・⁴南山大学・⁵大阪樟蔭女子大学)

キーワード：意味的透明性，漢字二字熟語，データベース

Toward construction of semantic transparency database of two-kanji compound words (1)

Hisashi MASUDA¹, Terry JOYCE², Taeko OGAWA³, Chikako FUJITA⁴ and Masahiro KAWAKAMI⁵

(¹Hiroshima Shudo Univ., ²Tama Univ., ³Tokai Gakuin Univ., ⁴Nanzan Univ., ⁵Osaka Shoin Women's Univ.)

Key Words: semantic transparency, two-kanji compound word, database

問題と目的

語彙処理研究を進める上で、語彙データベースの整備は必要不可欠である。特に、出現頻度などの物理量に加えて、親密度のような心理量に関するデータベースの構築は、新たな研究課題の発見につながる可能性も秘めている。

日本語の語彙については、近年、親密度(天野・近藤, 1999; 天野他, 2008)や心像性(佐久間他, 2005)に関する大規模なデータベースが構築されている。また、日本語の語彙において中心的位置を占める漢字二字熟語を対象に、統語構造のデータベース化の試みもなされている(Masuda & Joyce, 2005)。しかしながら、単語の意味情報に関しては、シソーラスのようなデータベースは存在するが、心理量のデータベース化は行われていない。特に、日本語以外の諸言語を実験材料とする研究では(e.g., Monsell, 1985), 単語の意味的透明性が語の認知過程に影響を及ぼすことが報告されていることから、漢字二字熟語のような複合語を多数有する日本語について、意味的透明性のデータベースを構築することは喫緊の課題である。

本研究の目的は、漢字二字熟語の意味的透明性に関するデータベースに向けた第1段階として、調査対象とする熟語材料の妥当性と、意味的透明性の評定手続きの妥当性について検討することにある。この目的に向け、本研究では多数の漢字二字熟語を調査対象とした上で、意味的透明性として、試験的に熟語の意味とその第一構成要素すなわち第一文字目の漢字の意味との関連性についての評定課題を実施した。

方法

調査協力者：日本語を母語とする大学生 42 名が本調査に参加した。

調査対象項目：調査対象とする漢字二字熟語として、日本工業規格(JIS)第一水準に登録されている常用漢字 2 字からなる語(固有名詞を除く)のうち、現代日本語書き言葉均衡コーパス(BCCWJ)を基に算出した長単位出現頻度(Joyce, Hodošček, & Nishina, 2010; in press)が 49 以上の 10,015 語を選定した。

手続き：まず、全調査協力者に共通して評定作業を求める語として、10,015 語の中から無作為に 115 語を抽出した。次に、残りの 9,900 語を 990 語ずつの 10 リストに無作為に分類した上で、無作為に選定したいずれか 1 リストを各調査協力者に割り当てた。各調査協力者は、第 1 日目に共通の 115 語について評定し、第 2 日目から第 10 日目の 9 日間で、割り当てられたリストの語に対して、1 日あたり 110 語ずつの評定作業を行なった。各リストに対して評定を行なった協力者数は 3 ないし 5 名であった。1 日あたりの評定作業は、まず各熟語の読み方と意味のそれぞれの知識の有無について回答し、次に熟語の意味と、その第一文字目の漢字の意味との関連性について 6 件法(非常に／かなり／少し関連性がある、あまり／ほとんど／全く関連性がない)で回答することであった。1 日あたりの評定作業時間は、いずれの協力者も 30 分以内であった。

結果と考察

まず、熟語の読み方と意味の既知性に着目して、全熟語を次の 3 種類に分類した。(1)未知語：評定を行なったすべての協力者が、その読み方あるいは意味を知らないと回答した熟語(112 語, 1.12%)。(2)不完全既知語：評定者のうちのいずれかが読み方あるいは意味を知らないと回答した熟語のうち、未知語を除いた語(2,704 語, 27.00%)。(3)既知語：未知語と不完全既知語とを除いた残りの語(7,199 語, 71.88%)。これら 3 タイプの熟語の長単位出現頻度(対数変換値)の平均値(と標準偏差)は順に、1,931 (0.225), 2,039 (0.300), 2,441 (0.535)であり、いずれのタイプ間にも有意差が認められた。このことは、熟語の出現頻度とその読み方や意味の既知性に影響を及ぼしていることを示す。

次に、既知語のみを対象に、出現頻度と意味的透明性の平均評定値の散布図(図 1)を描き、両者間の相関係数を求めたところ、その絶対値はきわめて小さな値であった($r = -.061$)。出現頻度が高いものの、意味的透明性の平均評定値が低かった熟語の例としては、“当時、今度、絶対、大事、本人”などが認められた。逆に出現頻度は低いものの、意味的透明性の平均評定値が高かった熟語の例には、“仮設、整体、死別、胃酸、筆先”などが認められた。

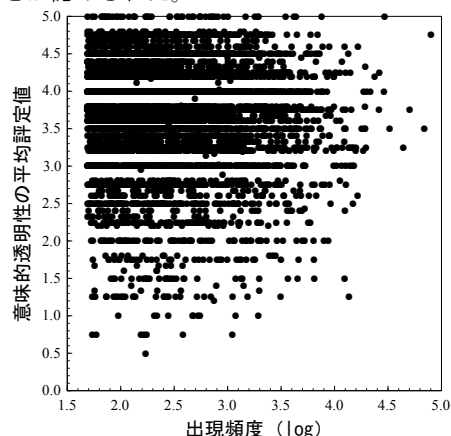


図 1 出現頻度と意味的透明性の平均評定値の分布

次に、意味的透明性の平均評定値は高いあるいは低いものの、評定者間によるばらつきが大きい熟語を探索的に調べたところ、たとえば“英語、空手”のように、第一構成要素の漢字の意味が多義的であるものや、“温床、銀行”のように、熟語の意味表示の際に第一構成要素の意味を比喩的に利用しているものが認められた。このことは、実験条件として意味的透明性を操作する際に、単にその高低のみに着目した材料選定が実験結果を不明瞭にする危険性を示唆している。

(本研究は科研費(基盤研究(C)23530966)の助成を受けたものである。)