

心像性データベースの構築と概要

○佐久間 尚子¹(さくまなおこ), 伊集院 睦雄¹, 伏見 貴夫², 辰巳 格³,

田中 正之⁴, 天野 成昭⁵, 近藤 公久⁵

¹東京都老人総合研究所, ²北里大学, ³LD・Dyslexiaセンター,

⁴京都大学霊長類研究所, ⁵NTTコミュニケーション科学基礎研究所

(要旨) 単語を文字呈示した場合と音声呈示した場合の2種の値を持つ約5万語の単語心像性データベースを刊行した。心像性は単語の感覚・運動覚的イメージの想起のしやすさを表し、単語の意味に関連する指標として有用である。複数の意味をもつ多義語などについては選択肢を用いて評定時の意味を特定した。これらの多義語などに対する選択肢別の頻度や心像性は多義語心像性データベースにまとめた。両データベースの概要を述べ、その特徴と利用法について概説する。

Key words: 語彙データベース, 単語属性, 心像性, 文字単語, 音声単語, 多義語

1. はじめに

人の語彙は数万語とも推定されている^{1,2}。この語彙の大きさに対し、通常の実験や検査で扱う単語は少数である。限られた単語数で言語処理の特性や障害の特徴を捉えるためには、目的の要因を操作し、目的外の要因が混入しない単語リストが必要となる。言語研究において多数の単語属性を統制する理由の一つはここにある。

我々は、単語の意味に関連した単語属性資料を得るため、文字呈示と音声呈示の2つの呈示モードにより約5万語の心像性評定実験を行い^{3,4}、心像性データベースを構築した⁵。心像性データベースは単語心像性データベースと多義語心像性データベースからなる。両者の構築方法の概略を述べ、2つのデータベースの特徴およびその利用法について解説する。

2. 心像性とは

心像性とは単語の感覚的・運動覚的イメージの想起のしやすさのことである。「りんご」や「テニス」のイメージに比べ、「はやり」や「プラス」のイメージは想起しにくい。一般に、具象語の方がイメージの想起は容易である。りんごの色、形、重さ、感触、味、香り、食べるときの音などはりんごのイメージとして想起されやすい。これらはりんごの意味の構成要素とも言える。心像性は単語の意味に深く関連した指標である。

3. 心像性データベースの構築

3.1 構築方針

心像性データベースの構築にあたり、以下の4つの課題を設定し、構築方針を定めた。

1) 大規模化 日本語の複数の表記や多数の品

詞を含む大規模な心像性資料を得るため、対象語として名詞、動詞、形容詞、副詞などの内容語で単語親密度4以上を基本とする約5万語を選定した。同じ基準の評定値を得るため、一人が対象語すべてを評定する方法を採用した。

2) 2つの呈示モード 従来心像性は文字呈示でのみ調べられてきたが、呈示モードの影響を調べるため、文字呈示と音声呈示の2種を採用した。音声呈示には単語親密度評定実験⁶に用いた音声ファイルを使用した(対象語は約4万語)。

3) 多義語の意味の区別 同音語(例、/カゼ/)や同字異音語(例、「一味」)、同字同音異義語(例、/サクラ/)など複数の意味をもつ多義語に関しては、あらかじめ意味を尋ねる選択肢を用意して評定時の意味を特定した。そして、最も多くの人が想起する第1意味を基に文字と音声のデータを組み合わせてデータベースを構築した。

4) 品詞情報の参照 本データベースシリーズの品詞データベース⁶を参照できるように、シリーズ共通のID番号を付与した。複数の品詞が対応している単語(例、{余り}は名詞と副詞)に対しては、第1意味に該当する共通ID番号を付与した。

3.2 構築過程の概要

心像性データベースの構築過程は図1の3つの過程からなる。第1過程では文字単語48,714語(同一語の異表記を含む)とそれらの4,073語に対する延べ10,632項目の選択肢を用いて心像性評定実験を行い、文字単語心像性データセットを構築した。文字単語の選択肢の選択頻度と選択肢別心像性の集計結果を基に多義語心像性データベース(文字単語版)を構築した。

第2過程では音声単語39,447語(同一語の異アクセント型を含む)とそれらの8,330語に対する

延べ 24,376 項目の選択肢を用いて心像性評定実験を行い、音声単語心像性データセットを構築した。音声単語の選択肢の集計結果を基に多義語心像性データベース(音声単語版)を構築した。

第3過程ではこれらの心像性データセットにシリーズ全体の共通ID番号^{6,7}を付与して、共通ID番号を基にデータベースを統合した。

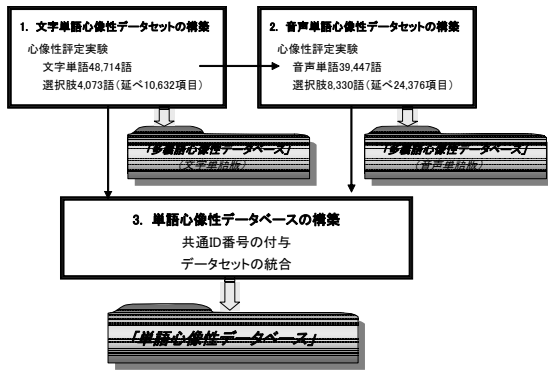


図1 単語心像性データベースと多義語心像性データベースの構築過程の概略図

4. 心像性データベースの内容

単語心像性データベースには 54,220 項目に対する 23 フィールドの情報がある。表 1 に各フィールドの名称を示す。54,220 項目のうち、文字単語心像性は 52,830 項目に、音声単語心像性は 44,052 項目に、両者は 42,662 項目に記載された。

多義語心像性データベース(文字単語版)には、表 2 に示す 9 フィールドの情報を、(音声単語版)には表 3 に示す 11 フィールドの情報を集録した。

表1 単語心像性データベースの 23 フィールドの名称

フィールド番号	印刷の有無	フィールド名
1		心像性エントリー番号
2	*	共通ID番号
3	*	共通ID注釈
4	*	読み
5	*	表記
6	*	アクセント型
7	*	文字単語評定人数
8	*	文字単語心像性
9	*	音声単語評定人数
10	*	音声単語心像性
11	*	選択肢の有無
12		文字単語と音声単語の第1意味異同の注釈
13		文字ID番号
14		文字刺激の表記
15		文字単語第1意味
16		文字単語第1意味の選択人数
17		文字単語第1意味の心像性
18		音声ID番号
19		音声刺激のカタカナ表記
20		音声単語第1意味
21		音声単語第1意味の選択人数
22		音声単語第1意味の心像性
23		音声ファイル名

表2 多義語心像性データベース(文字単語版)の 9 フィールドの名称

フィールド番号	フィールド名
1	文字単語選択肢エントリー番号
2	文字単語番号
3	* 文字ID番号
4	* 文字刺激の表記
5	* 文字選択肢の順位
6	* 文字選択肢の選択人数
7	* 文字選択肢の心像性
8	* 読み#
9	* 文字選択肢

表3 多義語心像性データベース(音声単語版)の 11 フィールドの名称

フィールド番号	フィールド名
1	音声単語選択肢エントリー番号
2	音声単語番号
3	* 音声ID番号
4	* 音声刺激のカタカナ表記
5	音声ファイル名
6	* アクセント型
7	音声選択肢の順位
8	音声選択肢の選択人数
9	* 音声選択肢の心像性
10	* 表記#
11	* 音声選択肢

5. 心像性データベースの利用上の注意点

心像性効果を検討する際には、心像性と他の単語属性間の相関関係を理解した上で心像性データベースを利用する必要がある。たとえば、心像性は単語頻度との相関が低く、単語親密度との相関が高い。一方、親密度が高くなるにつれて心像性が分化し、親密度の高い語では心像性の高い語(例、「牛乳」)から低い語(例、「経済」)まで広く分布する³。また、品詞によって心像性の分布は異なり、親密度の高い範囲では名詞以外の品詞に心像性の高い語が少ない⁴。本データベースの解説書⁵では、心像性と親密度⁶や単語頻度⁷、文字頻度⁷など複数の単語属性を統制することによって心像性の独自の効果を導き出す研究例を紹介している⁵。単語属性の特性を理解し、有効なデータベースの利用が望まれる。

<文献>

- 1) Amano & Kondo (1998), Proceedings of ICSP98.
- 2) 伊集院ほか (1999), 第 22 回日本失語症学会総会プログラム・講演抄録, p.103.
- 3) 佐久間ほか (1999) 第 2 回 CNP 抄録, 18-19.
- 4) 佐久間ほか (2002) 第 5 回 CNP 抄録, 48-49.
- 5) 佐久間ほか (2005) NTT データベースシリーズ日本語の語彙特性: 第 3 期第 8 巻単語心像性, 三省堂.
- 6) ・天野と近藤 (1999) NTT データベースシリーズ日本語の語彙特性: 第 1 期第 1 巻単語親密度, 第 3 巻単語アクセント, 第 4 巻品詞, 三省堂.
・近藤と天野 (1999) 第 1 期第 2 巻単語表記, 第 5 巻文字特性, 第 6 巻文字-単語, 三省堂.
- 7) 天野と近藤 (2000) NTT データベースシリーズ日本語の語彙特性: 第 2 期第 7 巻頻度, 三省堂.