

# 日本語基本単語に対する連想語データベースの作成

テリー・ジョイス

(「大規模知識資源-21世紀COE」東京工業大学)

key words : 連想語 データベース 語彙連想マップ

言語刺激を使用する認知心理学実験は、その刺激の重要な語彙特徴を統制しなければならない。語彙親密度、使用頻度、表記親密度などを操作するためのデータベース(天野・近藤 1999, 2000; 横山・笹原・野崎・ロング 1998)はあるが、連想効果を統制するための包括的なデータベースは存在しない。従って、本研究の目的は、日本語の基本語彙における自由連想語の調査を行い、大規模連想語データベースを作成することである。

## 方法

回答者：大学生 1,004 名。

日本語基本語彙コーパスの作成：自由連想語の調査対象コーパスとして利用できる、適切な日本語の基本語彙コーパスが既存していないため、基本語彙コーパスを作成した。そのために、国立国語研究所(1984)、玉村(2003)、三省堂編修所(1991)を参考にした。その参考物の共通単語を同定して、5,000 項目の漢字と単語を選択した。大規模連想語データベースを効率的に作成するために、質問紙作成時にリスト内の連想をできる限り統制しながら、多数の回答者に対して個人の質問紙を自動的に作成できる方法を工夫する必要があった。そのために、調査対象コーパスをコード化(読み、表記、構成要素漢字、意味の分類など)した。

現在、Web を用いた質問フォーマットを作成中であるが、第一回目のデータ収集として、質問紙の調査を行うために、語彙コーパスから 2,000 項目をランダムに選択した。

質問用紙：各項目に対して 50 回答ずつを得るように、2,000 項目を 20 リスト(各リストを 100 項目)に分けた。リスト内の連想がないように各リストを調べた後、各リストの 50 名の回答者に対して個人の質問紙での提示順序の並べ替えをおこなった。

質問紙では、「印刷されている文字を見て、一番最初に思い浮かんだ日本語の単語を 1 つ、下線部に書いてください。意味的に関係がある単語なら何でもけっこうです。」という教示のほかに、日本語文字体系(自然な表記と思い浮かんだ単語の書き方など)に絡む指示も含んだ。

## 結果

大規模連想語データベースを作成するために、第一回目のデータ収集として、質問紙を 1,004 名の回答者に配って、約 100,000 弱の解答を得られた。そのデータを入力しながら、連想語データベース設計を行っている。

## 考察

本研究では作成している大規模連想語データベースを更に発展させるために、現在、語彙コーパスの残りの 3,000 項目を対象とした第二回目のデータ収集を行っている。また、近いうちに、Web を用いた質問フォーマットを公開する予定である。

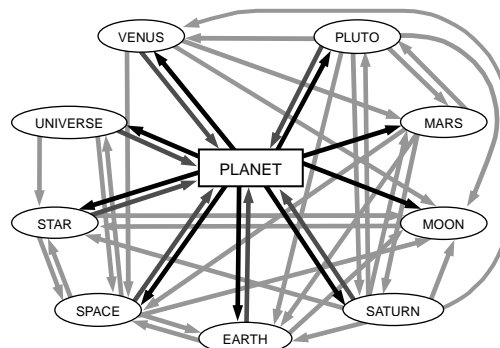


図 1. 語彙連想マップの概念 (Nelson & McEvoy, 2003)

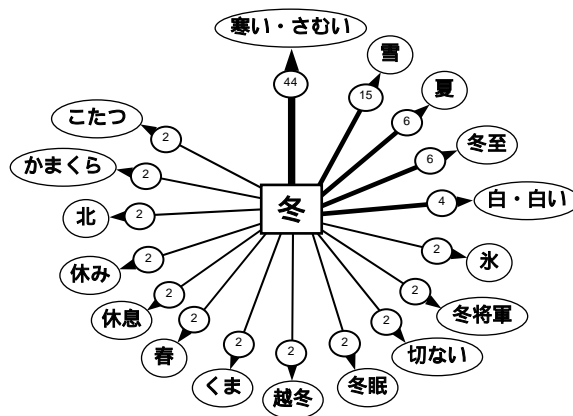


図 2. 現在のデータに基づく「冬」に対して連想語の集合

本研究で作成している日本語基本語彙に対しての連想語データベースは、認知心理学実験のための有用な資料となりうる。

連想語データベースの他の応用としては、図 1 に示しているように Nelson & McEvoy (2003) の概念に基づいた、語彙の重要な特徴と語彙どうしの接続性を捉えることのできる語彙連想マップとして利用することも考えられる。図 2 は、第一回目のデータ収集で得られた回答に基づいた「冬」に対しての連想語の集合である。このように連想語データベースを構築していくことにより、より完全な語彙連想マップができる。この語彙連想マップによって、日本語のコネクショニスト・モデル (e.g., レンマ・ユニット・モデル Joyce, 1999, 2002, 2004) における意味表象部分のモデル化が可能となる。

更なる応用は、ユーザ・フレンドリな検索方法および見出し語の追加・補足による辞書編集と包括的な漢字データベースに基づく日本語学習システムなどがある。

(Terry JOYCE)

本研究は、文部科学省 21 世紀 COE プログラム「大規模知識資源の体系化と活用基盤構築」の一部である。