

連想実験に基づく概念辞書の構築と検索

3H-7

大熊智子[†] 石崎俊[‡]

慶応義塾大学 政策・メディア研究科[§]

はじめに

柔軟な辞書検索の対象となる知識ベースや、文脈理解に必要なと思われる、概念辞書の内容、即ち上位下位概念や属性、その概念がどのような環境で用いられるかなどの知識を得るために連想実験を行なった。

この連想実験は被験者に刺激として与える概念を4分割し [1]、自然物/人工物/人間/情報・コンピュータそれぞれの領域から取り出した概念（単語）を用いて、「上位概念は?」「動作環境は?」などといった課題を提示し、連想的に思いついた概念を任意の数だけ挙げてもらうというものである。

この実験結果から得られる概念の出現頻度、連想に要した時間などのパラメータを元に、概念辞書に記述する概念間のリンクに重み付けを行う。このようにして構築した概念辞書を用いて、概念構造から検索対象語の特定をはかる [2]。また、複数の単語の入力から、それらの単語間の概念構造を検出する手法を考察する。

1 連想実験の概要

人間が一つ概念からどのようにして、関連する概念を思い起こすかを調べるために連想実験を行った。このような実験を行うことで、人間の記憶の構造に対応した知識ベースを得ることが、最終的な目的である。

被験者に、自然物/人工物/人間/情報・コンピュータという4つの領域から取り出した10個の概念を、A、B二つのグループに分け刺激語としてランダムに提示する。その下に、課題を表示してその解答となる語を連想する順番に任意個入力してもらう。どうしても、概念が連想出来ない場合はそこを飛ばして次に進むことにする。事前に、実験の説明を行うが、課題の意図がよくつかめない場合には、随時 help 画面で注釈を参照するように指示しておく。Aグループには「海」「りんご」(自然物)「交通渋滞」(人工物)「奴隷」「手品師」「家族」「祖先」(人間)「辞書」「ネッ

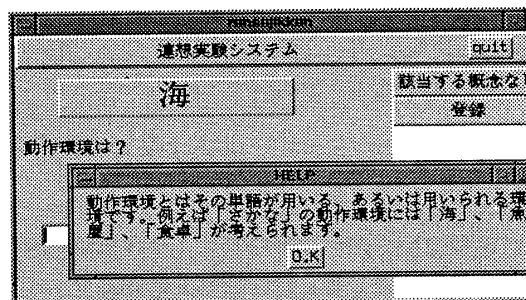


図 1: 実験中の画面

トワーク」「知識」(情報・コンピュータ)、Bグループには「泥沼」「卵」「金脈」(自然物)「鏡」(人工物)「ハッカー」「親友」「大人」(人間)「情報」「マルチメディア」「ディスプレイ」(情報・コンピュータ)を刺激語として与えた。

課題として上位/下位/同義語/属性/動作概念/動作環境を尋ねる。「動作環境は?」課題にはそれぞれ注釈が用意されており、動作概念には「動作概念とはその単語がする、あるいはされる動作です。例えば「さかな」の動作概念には「泳ぐ」、「釣る」、「焼く」が考えられます」、動作環境には「動作環境とはその単語が用いる、あるいは用いられる環境です。例えば「さかな」の動作環境には「海」、「魚屋」が考えられます。」という説明がある。

2 実験に基づく概念辞書の記述

2.1 実験結果

被験者一人につき、以下のようなデータを得る。概念の後にある数字は連想時にかかった時間である。

辞書

上位概念 { 本 6 }

下位概念 { 広辞苑 13 } { 英和辞典 14 } { 和英辞典 19 }
{ 国語辞典 17 } { 諺辞典 8 }

同義語 { 辞典 7 }

以下省略

Construction and Retrieval of Conceptual Dictionary based on Experiment of Concept Association

[†]Tomoko Ohkuma

[‡]Shun Ishizaki

[§]Keio University Media and Gavanance

表 1: 連想概念数と時間の平均

課題	概念数 (個)	時間 (時間)		
		1位	2位	3位
上位概念	1.07	68	27	11
下位概念	1.69	41	45	22
同義語	0.78	40	27	19
動作概念	1.61	20	12	16
動作環境	1.60	80	33	40

このようなデータから、以下のような傾向があることが分かった。

- ◎ 具体的概念の方が、抽象的概念より複数の被験者に共通の概念をより多く想起させている。
- ◎ 同じカテゴリに属する概念でも、日常生活に馴染みのある概念の方がより多くの概念を想起させている。
- ◎ 連想された概念の中には現実世界と異なる物がある。

2.2 実験に基づく概念辞書の記述

実験結果の集計データから、概念辞書の記述を行う。概念間にはコストを負荷し、その重さの順に単語を並べる。コストの負荷については次の段落で述べる。動作概念には、対象となる概念との格と場所格になると予想される動作環境を記述しておく。

表 2: 「りんご」の動作概念記述例

概念	コスト	格	場所格となる動作環境
食べる	1.71	を	食卓 病院
むく	5.00	を	食卓 台所
なる	5.01	が	果樹園 木畑 農園
落ちる	5.34	が	果樹園 木畑 農園

2.3 概念間リンクへのコスト負荷

重み付けを行うことによって、前述のように概念を辞書に記述する順番が明らかになり、概念フィルタリング [3] 実行時の候補の絞り込みを行うことが出来る。

実験結果から、「連想に要した時間」「連想順位」「出現確率」を用いて、それらの総和から概念間リンクに負荷するコストを計算する。ヒューリスティックに以下のような定式化を行った。a, b, c については、様々な値を用いて考察した結果、下のような場合が個々のデータの特徴を反映する一番妥当な値を生むと思われた。前述の個人のデータと下のコストの重み順の表を比較すると、並び方がほぼ対応しているのが分かる。刺激語と概念 W の間のコストを C として、

$$C = T \cdot a + j \cdot b + c/N$$

但し、

T: N 人の被験者が概念 W の連想に要した平均時間

j: 被験者が何番目に概念 W を連想したかの平均

実数: $a < b < c$ (下の例では $(a=0.001, b=1.0, c=5.0)$)

表 3: 「家族」の上位概念のコスト (上位 5 コ)

概念	コスト
社会	3.26
動物	3.68
人	4.01
人間	4.02
組織	7.01

3 概念フィルタリングによる辞書検索

以上のようにして構築したコスト付き概念辞書を用いて概念構造から検索対象語を特定する概念フィルタリングが可能になると思われる。また、候補となった語を絞り込むにあたって、コストの計算が有用になる。このような概念フィルタリングによる辞書検索を現在検討中である。

4 まとめと今後の課題

以上のように、本稿では連想実験を行なって知識を収集し、人間の記憶構造により近い概念辞書を構築する試みについて述べた。このように人の認知のメカニズムを考慮に入れて概念辞書を記述することによって、より柔軟な検索やメタファーなどの文脈理解 [4] が可能になると思われる。

しかし、連想実験による知識が豊富になればなるほど、概念フィルタリングなどの複雑な検索に実時間性を求めることが困難になって来ると思われる。検索作業の効率化が今後の課題である。

参考文献

- [1] 石崎・田中・今井: 柔軟な意味解析のための概念空間の定量化, 自然言語処理研究会 100-3(1994)
- [2] 大熊・石崎: 柔軟な辞書検索の機能について, 情報処理学会春季大会 (1995)
- [3] Kamran Parsaye, Mark Chignell Sretrag Khoshafian Harry wong/近谷英昭訳: 知的データベース: オーム社 (1992)
- [4] 今井・石崎: 比喩特徴発見における概念辞書の利用, 情報処理学会秋季大会 (1995)