

連想属性における感性語の特徴と定量的記述

安藤 まや 岡本 潤 石崎 俊
慶應義塾大学 政策・メディア研究科
〒252-8520 神奈川県藤沢市遠藤 5322
E-mail:{maya, juno, ishizaki}@sfc.keio.ac.jp

あらまし

本稿では、連想実験によって構築した連想概念辞書を用いて名詞とその属性を表わす感性語との定量的関連について述べる。名詞を刺激語とした連想実験の結果から、連想された感性語と刺激語との距離を計算する。その感性語から、名詞の上位・下位関係における属性継承やカテゴリー内における典型性などについて考察する。

1. はじめに

現在、辞書やコーパスには電子化されたさまざまな種類のものが存在する。EDR電子化辞書では大規模な概念体系が概念辞書として構築されている[8]。また、WordNet では、上位・下位関係以外に類義語、反意語といった情報も辞書化されている[9]。

しかしながら、実際に我々が自然言語処理に使用する際には、これらの情報だけでは不十分であり、多方面の情報が必要とされる。そのときの指針は人間がどのような多種類の情報を持ち、使用しているかという視点である。このような視点からの研究として、国語辞書の記述から属性を抽出し、ことばの類似性を判別するものなどがあるが[10]。本研究では、ことばに対する印象を表す感性語を扱う。感性語を連想という手段で獲得し、刺激語との距離を用いて定量化すると同時にさまざまな特徴を考察する。

筆者らはコンピュータネットワーク上で大規模な連想実験を行ない、連想語を収集して連想概念辞書を構築した。これは、我々が日常的に使用する概念空間をモデル化する目的で行なわれており、概念同士の距離を計算することによって概念空間を定量化したものである[2]。この概念辞書から、特に名詞の属性を表すことばとして形容詞を抽出し、名詞との共起関係について調べてきた[1]。

本稿では、連想属性に現れることばから形容詞に限定せずに、感性語についてカテゴリーごとに考察する。

2. 連想実験

連想実験は、被験者に刺激語を与えて、連想の制約を与えるための課題（上位概念や属性など）を提示して行なわれる。慶應義塾大学湘南藤沢キャンパス内のコンピュータネットワークを使用して行われており、被験者は、同キャンパスの学部生及び大学院生である。被験者は、事前に実験の目的や実験システムの操作方法などの解説を受ける。また、練習問題を解き、課題の意味などを学習してから、実際の実験を行なっている。

刺激語は、小学生用「学習基本語彙」[4]から意味的な重なりを除いた約 800 語の名詞とそれには含まれない基本的な名詞約 200 語から成る。これには、学習基本語彙を刺激語にして多くの被験者に連想された語の中から基本的な語を中心に選んでいる。

課題は、上位概念、下位概念、部分材料、属性、類義語、動作、環境の 7 課題で構成されている。被験者は刺激語とこの課題から連想を行なう。その際、これらの課題は刺激語ごとにランダムに表示される。

表1：課題説明

課題	課題説明	刺激語「ウサギ」の場合の連想例
上位概念	刺激語の特徴を含む広範なあるいは抽象的な概念。刺激語が属するカテゴリーを表す。	動物、哺乳類、生物、干支
下位概念	刺激語の特徴を持つより細かい概念。	ノウサギ、白うさぎ、アナウサギ
部分材料	刺激語の持つ部分、その全体像、また材料・原料。	目、耳、足
属性	刺激語の持つ特徴。	かわいい、元気、ぴょんぴょん
類義語	刺激語と同様の意味を持つ語。	ラビット
動作	刺激語が行なう、または行われること。	跳ねる、走る、食べる、捕獲する
環境	刺激語が用いる、または用いられる場。	月、野原、檻、昔話、山

実験中に、連想時間、順序、頻度を記録しており、これらを使用して、刺激語と連想語の距離を計算している。刺激語と連想語との距離 D を線形性を仮定し、 α 、 β 、 γ を係数とした次の式を使用する。

$$D = \alpha \times T + \beta \times S + \gamma \times F,$$

$$T = \frac{1}{n} \sum_{i=1}^n t_i \times \frac{1}{60}$$

t_i = 被験者が連想に要した時間

s_i = 被験者が連想した語の順位

$$S = \frac{1}{n} \sum_{i=1}^n s_i$$

n = 連想人数

$$F = \frac{N}{n}$$

N = 被験者の合計数

これを線形計画法を用いて解く。最も距離の近い場合としては、「連想時間が短く」「一番始めに連想」されており「被験者全員が連想」した時を想定し、最も距離が遠い場合としては、「連想時間がある程度長く」「連想順位が大きく」「全被験者のうち一人だけが連想」した場合を想定している。このような境界条件の元に線形計画法を使用して解くと、係数 α 、 β 、 γ の値が次のように得られる。ここで、 α が0になったのは、連想時間の測定の中に仮名漢字変換などが含まれていて、測定値にバラツキが大きいためと思われる。

$$\alpha = 0, \quad \beta = 0.33, \quad \gamma = 0.77,$$

$$D = 0.33 \times S + 0.77 \times F.$$

距離を伴った刺激語「絵」の概念辞書データを例に示す。このように、刺激語と連想語の関連性が定量的に提示することが可能となる。

表2：絵に関する連想例

上位概念	芸術 (1.36) 趣味 (3.74) 娯楽 (3.74) 美術 (6.82)
下位概念	油絵 (1.35) 水彩画 (2.12) 風景画 (4.57) 版画 (7.48)
部分材料	絵の具 (1.39) 紙 (1.76) 線 (4.90) 色彩 (7.48)
属性	美しい (1.30) 上手い (4.07) 下手 (7.48) 写实的 (8.47)
類義語	絵画 (1.10) イラスト (6.49) スケッチ (6.82) 漫画 (7.15)
動作	描く (1.18) 売る (2.61) 鑑賞する (6.82) 載せる (8.47)
環境	美術館 (1.22) 学校 (2.090) 額 (6.82) 壁 (7.15)

被験者数は刺激語一語に対して多くは10人としているが、一部、被験者数を20名ないし50名に増やしたも

のもある。

今回データを使用している名詞を刺激語にした連想実験のほかに、感性語を刺激語にした連想実験も実施している。意味分類された 10 段階の感性語を刺激語として類語、反意語について同様の連想実験を行なっている。感性語を扱った研究に、感性語のもつ段階的な曖昧さを含んだ定量的な意味ネットワーク構築が存在する[11]が、このような段階性の解明などに連想データの応用が期待される。

3. 刺激語と連想属性

本稿では、連想実験の 7 課題の中から、属性に着目する。属性は、刺激語の持つ特徴を連想する課題で、連想語には主に形容詞、形容動詞、擬音語、擬態語などが含まれている。その中から、刺激語との距離が 5.0 以内のものに着目する。距離 5.0 以内という値は、被験者 10 名のうち、複数の被験者が比較的早い順位で連想を行なった場合に出るものである。今回、使用した概念辞書の総刺激語 1,035 語で、そのうち距離が 5.0 以内の属性が検出されたのは、967 語だった。つまり 9 割以上の刺激語について、何らかの有用な属性が抽出できたのである。ここで、属性が多く抽出された語、少なかった語、連想されなかった語について以下考察する。

まず、刺激語を分類語彙表にしたがって分類した。刺激語はすべて名詞であるため、分類語彙表中の「体の類」に従って分類している。ここでは、最も大きな 5 種の分類項目「抽象的關係」「人間活動の主体」「人間活動 - 精神および行為」「生産物および用具」「自然物および自然現象」について述べる。これらの項目には、さらに細かい項目として下記のようなものが含まれる。

表 3：分類語彙表の分類項目（一部）

抽象的關係	力、位置・地点・場合、空間・場所、形・型・姿・構えなど
人間活動の主体	人間、相手・仲間、社会、団体など
人間活動 - 精神および行為	感覚、言動、創作・著述、文化・歴史・風俗、仕事など
生産物および用具	物品、食料、住居、道具など
自然物および自然現象	自然・物体・物質、宇宙・空、植物、動物など

これらの項目に所属する刺激語数、刺激語一語あたりの平均連想語数（距離 5.0 以内）、刺激語例は以下の通りである。

表 4：辞書データの分類（1）

	所属する 刺激語数	平均連想語数 (距離 5.0 以内)	刺激語例 (連想語例)
抽象的關係	161	2.36	力(強い) 時間(早い) 土地(広い) 玉(丸い)
人間活動の主体	117	3.09	人(優しい) 医者(偉い) 親友(楽しい) 学校(大きい) 政府(不透明)
人間活動 - 精神および行為	312	2.50	印象(良い) 方言(難しい) 作文(おもしろい) 七夕(ロマンチック) 漁業(大変)
生産物および 用具	280	3.57	品物(多い) 衣服(暖かい) 餅(白い) 建物(高い) 楽器(楽しい)
自然物および 自然現象	285	3.54	日光(明るい) 鉄(堅い) 台風(強い) 朝日(眩しい) 桜(美しい) 犬(かわいい)

「抽象的關係」および「人間活動 - 精神および行為」は、抽象的な意味を表わす名詞が多く含まれているため、一語あたりの平均連想語数が他に比べ、低くなっている。また、距離 5.0 以内の連想語が抽出されなかった語も、この 2 種の項目に多く含まれていた。「生産物および用具」や「自然物および自然現象」については、動植物、商

品など具体物を表わす身近な単語が多く含まれていたため、連想語数が多くなっているものと思われる。
次に、距離 5.0 以内の連想語が抽出されなかった刺激語の割合とその刺激語例を提示する。

表 5：辞書データの分類 (2)

	抽出されなかった 刺激語数の割合	連想されにくかった刺激語例
抽象的關係	13.0%	正体、機会、四季、順序、方角
人間活動の主体	3.4%	社会、団体、住民
人間活動 - 精神および行為	7.0%	期待、常識、感想、用心、参考
生産物および用具	2.8%	小包、材料、弓
自然物および自然現象	1.4%	波、岸、穂

特に「抽象的關係」で、方向を表わす刺激語（左、右、中央など）と思考に関わる刺激語（常識、関心、参考、結論など）の連想語数が少ない傾向が見られた。

4．属性継承

上位・下位関係にある刺激語を対象に、属性の類似度を見る。ここでは、三段階の階層になっているものを考察対象とする。

二つの刺激語の属性に関する類似度 S は距離を考慮に入れて刺激語 A、刺激語 B の間で次の様に定義する。表 6 にあるように、刺激語 A の各連想属性への距離を a_i 、刺激語 B の場合を b_i とする。距離の値は小さいほど関連性が強いものなので、それぞれの値の逆数をとって正規化する。そのように正規化した値を要素とするベクトルをそれぞれ X と Y とすれば、表 6 の右の式にあるように、ベクトル X の要素 X_i と Y の要素 Y_i を表すことが出来て、類似度 S はベクトル X と Y の内積として表現できる。ただし、 $||$ はベクトルのノルムを表し、また、表 6 で連想されていない場合の距離を 0 とする。

表 6：語彙行列

	刺激語 A	刺激語 B
連想語 1	1.35	2.18
連想語 2	3.73	0
連想語 3	4.07	4.31
⋮	⋮	⋮
連想語 i	a_i	b_i
⋮	⋮	⋮

$$X_i = \frac{1/a_i}{\sum \left(1/a_i \right)} \quad Y_i = \frac{1/b_i}{\sum \left(1/b_i \right)}$$

$$S = \frac{X \cdot Y}{|X| \cdot |Y|}$$

以下に属性の類似度および共通属性の例を提示する。

表 7：上位・下位関係における類似度 (1)

	果物・ぶどう	果物・マスカット	ぶどう・マスカット	乗り物・車	乗り物・スポーツカー	車・スポーツカー
--	--------	----------	-----------	-------	------------	----------

類似度	0.712	0.736	0.756	0.734	0.556	0.623
共通属性	おいしい、甘い、すっぱい、丸い、高い	おいしい、甘い、すっぱい、丸い	おいしい、甘い、すっぱい、ツブツブ、緑色	高い、便利、速い	高い、速い	高い、速い、危険

これは、どの段階においても比較的高めの類似度が見られる例である。

「果物」の下位概念で刺激語となっている語が他に5語あったが、ほとんどの刺激語に対して「おいしい」「甘い」が非常に近い距離で、また「すっぱい」「丸い」が比較的距離の近い連想語として存在していたため、類似度は最低でも0.535と、いずれも高い値になっている。同様に、「乗り物」は、「速い」が非常に距離が近い状態で連想されており、たいていの刺激語でも「速い」が、距離が近い状態で連想されていた。そのため、全体的に類似度の値が高くなっている。「果物」と異なる点としては、一部、類似度の値が低い刺激語が存在したことである。「モノレール」「船」などは「速い」という連想語がなかったため、類似度の値が0.062、0.161と極端に下がっている。

よって、カテゴリ「果物」では、「おいしい」「甘い」が、「乗り物」では「速い」が継承されている場合が多いと言える。

次にこれらと異なる傾向の類似度の値をもつカテゴリをあげる。

表8：上位・下位関係における類似度（2）

	楽器・ピアノ	楽器・グランドピアノ	ピアノ・グランドピアノ	家具・いす	家具・ロッキングチェア	いす・ロッキングチェア
類似度	0.399	0.280	0.709	0.386	0	0.195
共通属性	重い、高い、硬い、面白い	重い、高い、大きい、美しい	重い、高い、大きい、ピカピカ	重い、硬い	-	気持ちいい、ツヤツヤ、フカフカ

カテゴリ「楽器」の場合は、共通した連想語は存在するが、それぞれ各々異なった距離の値をもっていたため、類似度の値が小さくなっている。また、他の「楽器」に属する刺激語も類似度の値が0.20以下と小さい。また、カテゴリ「家具」についても、同様の傾向が見られる。これらのカテゴリでは、多少の属性継承はみられるものの、「果物」「乗り物」のように継承された「属性」を明示するまでには至らなかった。

このようなカテゴリは、感性語での属性継承が強く見られない。それは、本来「属性」というものが感性語以外の要素によっても表すことができるものであるからだと考える。事実、連想実験データで属性以外の課題で類似度を測定したところ、カテゴリごとに異なった傾向が見られた。全体的には、課題「部分材料」の類似度が高くなっていたが、一部では、さほど高くないものも見られた。

以上のように、属性継承が強く出るカテゴリとそうでないものがあるため、属性を表す表現を感性語のみにとどめず、幅を広げていく必要があると考えている。その際、何を基準に属性記述とみなすのか、が課題となるだろう。また、今後一語あたりの被験者数を増やし、カテゴリ別の属性継承の傾向についてさらなる分析を重ねていきたいと考えている。

5．典型性と属性

カテゴリに所属する語には、典型的な例となり得るものと、そうでないものが存在する。例えば、「スズメ」と「ペンギン」は、共にカテゴリに含まれるが、「スズメ」の方が鳥らしい鳥として認識されている。このような典型的な例は、「鳥」を刺激語にした場合、課題「下位概念」で多くの被験者に早い順位で連想される場合が多い。

ここでは、下位概念で連想された語と属性で連想された語に着目し、カテゴリーに属する名詞の典型性と属性の関連性について述べる。手法としては、カテゴリーの「下位概念」に属する距離の近い連想語を抽出し、その連想語を刺激語とした時の属性とカテゴリーの属性の類似度を比較する。

表の刺激語は左から距離の近い下位概念の順で提示している。括弧内の数字はカテゴリーとの距離である。「×」は刺激語として実験が行なわれていないことを表す。

表 9 : 「家具」の類似度

	いす (1.27)	たんず (1.91)	机 (1.95)	テーブル (3.07)	棚 (3.64)
類似度	0.386	0.273	0.151	0.228	×
共通属性	重い、硬い	大きい、重い	大きい、高い	大きい、高い	

表 10 : 「乗り物」の類似度

	車 (1.1)	飛行機 (2.11)	電車 (2.13)	バイク (2.53)	自転車 (2.68)
類似度	0.734	×	0.570	0.488	0.570
共通属性	高い、便利、速い		大きい、安い、遅い、便利、速い	楽しい、楽、速い	高い、安い、楽しい、便利、速い

「家具」と「乗り物」については、下位概念として最も短い距離で連想されている語との類似度の値が高くなっている。しかし、二番目以降の距離で連想された語については、多少の揺れはあるものの、比較的類似度の値は高くなっている。

現段階では、それぞれのカテゴリーの下位概念として連想された語が必ず実験を行なわれているわけではないため、はっきりしたことは言えないが、典型性と類似度との間に何らかの関連があるのではないかと考えている。

しかし、一方で、下表のような結果もある。

表 11 : 「動物」の類似度

	イヌ (2.2)	サル (2.45)	ネコ (2.82)	鳥 (3.11)	人間 (3.37)
類似度	0.161	×	0.243	0.138	0.219
共通属性	かわいい、恐い、大きい		温かい、かわいい、冷たい、素早い	素早い、速い、かわいい	弱い、賢い、強い

「動物」の場合は、下位概念の連想結果と、属性の類似度は必ずしも一致していない。また、典型性に関わらずおおよその語において高い類似度の値が出た「果物」なども存在している。そのため、属性継承と同様に、カテゴリーごとの特徴を見極める必要があると考えている。

6 . 今後の課題

今回の考察で、感性語のみの属性傾向が多少見えてきた。今後は、属性とは何か、判断基準をどう定めるのか、について考え、属性要素を増やしてさらなる考察を加えていきたい。また、刺激語の充実と一語あたりの被験者の増加が必要である。現在、一部の刺激語に付いて被験者数を一語あたり 50 名まで増加させている。以上のような観点から、名詞と属性の関連性についてさらなる考察を加えていきたいと考えている。

謝辞

連想実験の被験者となってくださいました皆様、また、辞書構築にあたりお力添えと助言をくださいました石崎研究室の皆様に心より感謝いたします。

参考文献

- [1] 安藤まや・石崎俊, 名詞と形容詞における共起関係の研究---連想実験に基づく定量的アプローチ---, 信学技報 TL98-26, pp51-58, 1999.
- [2] 岡本潤・石崎俊, 概念辞書の構築と概念空間の定量化---連想実験による概念空間の抽出---, 情報研報 Vol.99.
- [3] 岡本潤・内山清子・石崎俊, オンライン連想実験システムと学習基本語彙の概念辞書化, 情報処理学会報告 自然言語処理 118-18, 1997.
- [4] 甲斐睦朗・松川利広編, 語彙指導の方法 [語彙表編], 光村図書, 1996.
- [5] 国立国語研究所編, 分類語彙表, 秀英出版, 1993.
- [6] 株式会社日本電子化辞書研究所, EDR 電子化辞書仕様説明書, 1993.
- [7] 株式会社日本電子化辞書研究所, 日本語単語辞書 Version1.5, 1995.
- [8] 株式会社日本電子化辞書研究所, 概念辞書 Version1.5, 1995.
- [9] Christiane Fellbaum, WordNet: An Electronic Lexical Database, The MIT Press Cambridge, 1998.
- [10] 笠原要・松澤和光・石川勉, 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌 Vol.38, No.7, 1997.
- [11] 小林秀行・岩岡俊行・奥村成吾・太田俊二, 概念構造を記述する感性語意味ネットワークの構築, 第一回日本感性工学会大会論文集, 1999.