形容詞が内包する抽象的意味の抽出と自動分類の試み

神崎享子　馬青　山本英子　村田真樹　井佐原均
独立行政法人　通信総合研究所

1.　はじめに

本研究は、形容詞が内包する抽象的意味の様相を自己組織型意味マップ（SOM）を用いて求めるものである。また、本稿で述べる形容詞とは、形容詞と形容動詞を含める。

形容詞の意味は、複数の抽象的な意味カテゴリーに渡る。たとえば、「きつい坂道」の「きつい」は程度、感覚、場合によっては評価を表すし、また坂道のもつ特徴でもある。このように一語の形容詞でも、いろいろな意味を帯びるので、意味分類をしようと思っても、一つの意味カテゴリに納まらないことが多い。形容詞の意味は、抽象的でかつ多面的であるので意味分類することが難しい。従来の形容詞の意味記述や分類は、格パタンや連体、終止、連用などの用法、あるいは接尾語をつけたり共起する名詞などの情報を利用し、類義語、対義語などを調べている（西尾 1972, IPAL 1991）。

「属性」や「感情」などは、一般的によく知られた形容詞の意味的特徴であるのだが、そのような語を用いずに、抽象名詞を利用することで、形容詞の抽象的な意味、つまり意味カテゴリにあたるものをコーパスから抽出し自動分類する。

従来の我々の研究では、限られた語数で SOM の配置に従ってその類似性を検討してきた。また、SOM の分類能力に関しても他手法との比較により劣っていないことも明らかにした（馬 2001）。今回、構築した意味マップは、類似尺度として補完類似度を導入した。補完類似度に関しては山本・梅村（2002）で詳細に述べられているように、ある事象と事象が包含関係にある場合に強い類似尺度である。語どうしの関係も包含関係で関係付けてくれる可能性が高い。これと従来の類似尺度を用いた意味マップとを比較、検討する。そして、これらの意味マップにおいて、語がどのような分布になっているのかを、マップ上での語全体の位置関係と類似度の関係の高い二語間の関係とを合わせて検討する。

2.　形容詞の抽象的な意味を探るための意味的手がかり

抽象的な名詞の統語的役割に着目した先行研究には、根本(1969), 高橋(1975)などがあげられる。例えば高橋(1975)においては、

①やぎは性質がおとなしい
②ぞうは鼻が長い

の二例を比較し、①を側面語、②を部分語と仮に呼び、文中の役割が異なることを述べている。側面語になる単語は主語の示すものや人の側面を表すとともに、述語の示す属性の類概念（上位概念）を表す単語である。また、根本（1969）においても「色が白い」「速さがはやい」「年が若い」「背が高い」などは、「顔が赤い」などのような状態の持ち主を表す場合と違って同義反復的な性格が強いと述べている。このように、我々の言語活動の中にも、形容詞の上位概念を示すような用法がみられる。

神崎（1999）において形容詞の意味関係の分類を行ったところ、主語述語関係に変換できるものの中で、「ゆるやかな傾斜」にみられる「名詞＋「が」＋形容詞」というタイプと主語述語に変換できないものの中で「悲しい気持ち」のようなタイプの２種類に、上記のような関係に近いものがみられる[1]。つまり、形容詞が抽象名詞の具体的表現になっているものである。このタイプは、形容詞と抽象名詞が語彙的意味をお互いに共有していると考えられる。たとえば、「白い色」は「色」という意味を共有し、「白い」が「色」の属性値であり、「色」は「白い」の上位概念である。また、「悲しい気持ち」においても、「悲しい」は「気持ち」の属性値であり、「気持ち」は「悲しい」の上位概念である。このようなパタンは、形容詞の抽象的意味をコーパスから探るのに重要な手がかりになるのではないかと考える。

3.　データ

抽象名詞は、９４、９５年分の毎日新聞２年分から取り出した。抽象名詞と共起する形容詞、形容動詞は、毎日新聞１１年分、日本経済新聞１０年分、産業金融流通新聞７年分、読売新聞１４年分、新潮文庫１００選、新書版１００冊の中から用例を調べた[2]。抽出された抽象名詞は 365 語、形容詞の異なり語が 10525 語、のべ語数は 35173 語であった。最大共起語数 1594 語である。最初に作成されるのは、以下のような表である。

思い：　うれしい　楽しい　悲しい ……
気持ち：楽しい　嬉しい　幸せな ……
観点：　医学的な　歴史的な　学術的な ……

4.　入力データの符号化

第３節の単語のリストを SOM の入力とするには、符号化する必要がある(Ma 2000)。

まず従来型の符号化について述べ、次に補完類似度の符号化について述べる。

---

[1] 高橋(1975)は構文上での考察であり、たとえば、「体が大きい」「顔がきれいだ」「足が速い」などは、部分語が側面語に変わったものだととしている。筆者は意味関係の分類は連体修飾関係を対象にしているが、もし、さきの例と同じ意味で「大きい体」「きれいな顔」「速い足」のような例があれば、それらは、「白い色」などと同じタイプとして扱う。

[2] 用例を検索するにあたっては通信総研で開発したツール Tea を利用した。

4.1. Ma(2000)の符号化

SOM への入力とするためには、単語のリストを符号化する必要がある（Ma 2000）。

ここで、一般にω 種類の名詞 $w_i(i=1,\ldots,\omega)$ が存在し、それらの意味マップを構築すると仮定する。このような場合、名詞 w_i は 以下のように連体修飾要素のセットで定義される。

思い＝ {悲しい、楽しい、幸せな、…}

$$w_i = \{a_1^{(i)}, a_2^{(i)}, \ldots, a_{\alpha i}^{(i)}\}$$

ただし、$a_j^{(i)}$ は、w_iと共起する j番目の連体修飾要素で、αi は w_iと共起する連体修飾要素の数である。これを符号化するためにここでは、名詞w_i と w_j間の距離d_{ij}を以下のような計算機式によって求めることとした。

$$d_{ij} = \begin{cases} \dfrac{(\alpha_i - c_{ij}) + (\alpha_j - c_{ij})}{\alpha_i + \alpha_j - c_{ij}} & i \neq j \text{ の場合} \\ 0 & i = j \text{ の場合} \end{cases}$$

ここで、α_i とα_jはそれぞれ w_i 、w_jと共起する連体修飾要素の総数で、c_{ij} は w_i 、w_j に共通する連体修飾要素の数である。上の式は、意味的関係 d_{ij}は、w_i 、w_jの間にどのくらい共通する連体修飾要素があるのかということを表す正規化された距離である。すなわち、d_{ij}が大きければ意味的な距離は遠く、d_{ij}が小さければ意味的な距離は近くなる。Ma(2000)では、この類似計算のあと「相関コーディング法」を用いる。ここで提案する相関コーディング法では、名詞 w_iをこの行列を用いて以下のような多次元ベクトルに符号化する。

$$V(w_i) = [d_{i1}, d_{i2}, \ldots, d_{i\omega}]^T$$

$V(w_i)$ はSOMへの入力であり、この多次元ベクトルを自己組織化によって、それらの間に存在する意味関係を顕在化し二次元空間に表現する。

4.2 補完類似度を用いた符号化

山本･梅村(2002)の補完類似度は、包含関係を取り出すことを得意とする類似尺度である。これをSOMの入力データの符号化に用いた。山本･梅村(2002)によれば、補完類似度は以下のような式になる。

今、 $\vec{F} = (f1, f2, \ldots, fi, \ldots, fn)$（$fi = 0$または 1)、
$\vec{T} = (t1, t2, \ldots, ti, \ldots tn)$（$ti = 0$または 1)
とする。

$$Sc(\vec{F},\vec{T}) = \frac{ad - bc}{\sqrt{(a+c)(b+d)}}$$

ここで、a は、二つのラベルが同時に現れるデータの数、b は $lab1$ が現れ、$lab2$ は現れないデータの数、c は、$lab2$が現れ、$lab1$ は現れないデータの数、d は、二つのラベルがどちらとも現れないデータの数である。本データに対してこの尺度を用いる際には、lab にあたるのが抽象名詞となり、a は、ある形容詞が二語の抽象名詞と共起しているパタン、b と c は、ある形容詞がそれぞれ一方の抽象名詞とだけ共起しているパタン、d は形容詞が両者ともに共起していないパタンということになる。補完類似度の数値を正規化し、そをもとに 4.1 節で述べた相関コーディング法を用いて抽象名詞を多次元ベクトルに変換して符号化する。

5. ベースラインの意味マップと補完類似度を符号化に用いた意味マップ

本節では、従来型の意味マップ（ベースラインと呼ぶ）と補完類似度を用いた符号化による意味マップとを示す。

まず、分類結果の評価に関して、一つの座標上に配置された名詞の、共通する共起形容詞数で判断した。意味マップを見たときに、少なくとも同じ座標上にあるものは、かなり類似していなければ、直感的に類似している単語が近くに配置されているとは考えにくい。

次に両マップでグループの変わらない単語を調べる。両方のマップで、単語どうしがいつも同じ位置あるいは比較的近い位置にあれば、その単語の類似度は高い可能性がある。

最後に、単語がどのようにマップ上に配置されているのかについて調べる。意味マップ上に配置されている単語は、全体の中での類似性の位置付けであるが、そこに、SOM の符号化の際に計算した二語間の類似度計算で値の高いものから語どうしにリンクをはることにした。これによってある語の周辺の語との関係の強さがわかるので、マップ上の語どうしの関係性が視覚的にわかる。

意味マップは整列フェーズと微調整フェーズからなる。ベースラインの意味マップでは、整列フェーズが１万回、微調整フェーズは５万回の学習で得られたものである。座標は４５×４５で、半径７である。

一方、補完類似度を用いた意味マップの方は、ベースラインと同じ学習回数だと学習不足で、365 語の抽象名詞が意味マップ全体にランダムに散らばった状態になった。そこで学習回数を増やし、整列フェーズは３万回、微調整フェーズは１０万回にした。座標と半径の数値はベースラインと同じ４５×４５で、半径７である。（図１、２）

同じ座標上の語数：

同じ座標に集まっている名詞をトータルでみるとベースラインの意味マップの方が圧倒的に名詞がまとまっている。ベースラインの意味マップでは１４０語、補完類似度を使った意味マップでは、１０４語の名詞がところどころで座標を同じくしている。ベースラインの意味マップは、複数の抽象名詞が固まっているのに対し、補完類似度のマップは抽象名詞が散在している、というイメージになる（図１，図２）。

では、同じ座標上に複数の抽象名詞がある場合、どれくらい意味的に類似しているのか。それを調べるために、同じ座標上の複数の抽象名詞に共通する共起語数は、一つの抽象名詞の全共起語数の何割を占めているかを求めた（表１）。同じ座標上であれば、未分類であるか、とても意味が近いかのどちらかであると思われる。ベースラインも補完類似度も、共起語の重なりの割合がほとんどが 0%から 30%の間である。これは、お互いの関

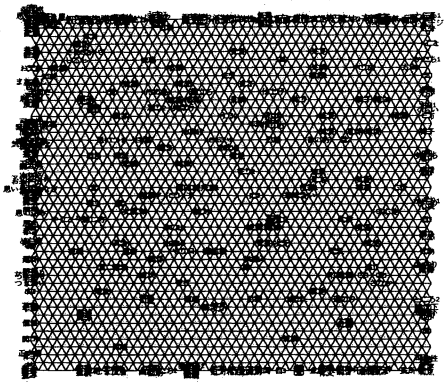

図１　ベースラインの意味マップ

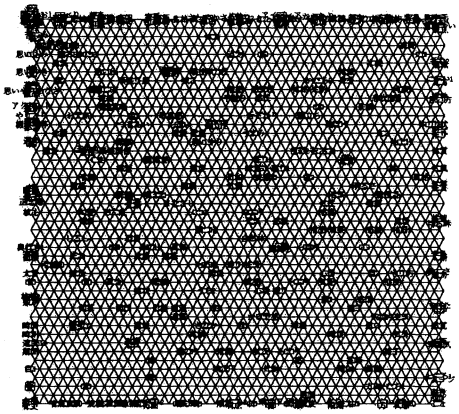

図２　補完類似度を用いた意味マップ

係が薄いか、あるいは、同じ座標上の複数の名詞の中で、共通する形容詞はあるものの比較的共起語が多いので重なりの度合いが低くなる名詞であるかのどちらかの可能性がある。たとえば、ベースラインの意味マップで、同じ座標に複数の名詞がありながら、すべてに共通する形容詞がない場合（０％）をみると、「間隔·思い込み」や「圧力·線·値·都合」などがある。これらの名詞の意味を考えると直感的にも座標上同じ位置を許すほどの類義語であるとは思いにくい。同じ座標上の複数の名詞の中で、比較的共起語が多いために重なりの度合いが低くなっている例としては、たとえば「色·色彩·彩り」は一つのグループであるが、「色」の共起語が比較的他の２つの抽象名詞と比べて多いため、形容詞の重なりの割合が10%程度となっている。全体的には、共通する共起語が１０％未満の場合には、名詞は直感的にもうまく分類されていない場合が多い。ベースラインの意味マップと補完類似度を類似尺度として用いた意味マップとを比べると、両者とも、同じ座標上に複数の単語が分類されていても共通する共起語の割合がそれほど高くはないが、両者の比較をするならば、ベースラインの方が、同じ座標上に共起語に重なりのある単語が位置しているので、うまく分類されているイメージになる。

表１　同じ座標に位置する抽象名詞の、重なりの割合ごとの個数

重なりの割合(%)	ベースライン	補完類似度
0-1	22	16
1-10	27	21
10-20	40	22
20-30	17	21
30-40	10	10
40-50	4	2
50-60	10	5
60-70	0	0
70-80	0	2
80-90	1	1
90-100	0	0
100	9	4

ここで、もう一度、前の２つのマップを比較すると、補完類似度の意味マップはそもそも単語が散在している。もし、単語を一つ一つ厳密にマップ上に配置するとしたら、同義語以外には同じ座標に置かれないわけなので、そういう意味では、補完類似度の意味マップをもう少し検討してみる必要がある。

両方のマップに共通する語のセット：

ベースラインの意味マップ上で同じ座標もしくは近辺にある単語群は、補完類似度を用いた意味マップ上では、どのようになっているかを調査する。どの尺度を使っても安定している単語のグループは、強い類似性を示す可能性が高い。前述のように、ベースラインの意味マップで同じ座標に位置する抽象名詞の数は１４０語であった。この１４０語の名詞のうち補完類似度の意味マップにおいて近い位置にプロットされたもの（半径２以内）は８７語あった。半径３以内を含めると９６語になる。両方のマップに共通して近い位置にある８７語をあげると、次のようになる。

「幸福感　心遣い　情愛　配慮　思いやり　温かさ　気立て」「願い　情熱」「言葉　意見　評価」「触感　手触り　感受性　感性　舌ざわり」「におい２　香り１　味覚　味」「色２　色彩２　彩り１」「活力　若さ」「緊張　緊張関係」「かかわり　つながり」「傾斜　勾配」「核　核心　眼目　急所　骨子　重み　正念場」「家系　血筋　血統　家柄」「昔　大昔　老舗」「円　曲線　図形　面３」「奥行き　空間　面積」「歳月　格好２」「勢い　速度　時間　時刻　期間」「強度　力２　語気」「品格　階級」「数　量」「音程　角度２」「利点　順番」「順　程度　可能性」「うち１　中１　一方」「時　状態　方」「ところ１　イメージ　印象　面」「美しさ　魅力　人柄」

以上の結果をみると、補完類似度を用いた意味マップでは、同じ座標上の単語は少ないが、類似

している単語は基本的には近くに位置していることがわかる。

意味マップの単語間の関係：

名詞の意味マップは相関行列値に基づいた類似度によるので、ある名詞の全体の中での類似関係が求められたのだが、この単語のマップを二語間の類似度に従って更にリンクさせると、ある語の周辺の語との直接的な関係の強さがわかる。ここにリンクつきのベースラインのマップと補完類似度を用いたマップをあげる。類似度の低いものもすべてリンクしてしまうとマップは線で埋め尽くされてしまうので、見やすさの便宜のため類似度の値が高いものだけリンクを張っている。

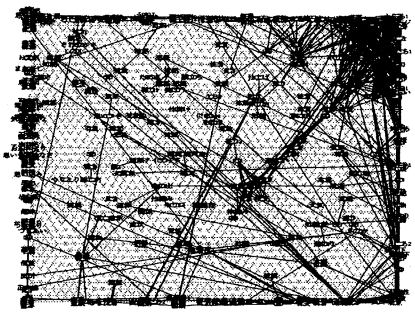

図3　ベースラインの二語間の類似関係

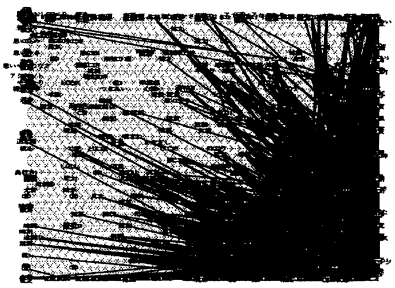

図4　補完類似度の二語間の類似関係

図3と図4を比較すると、ベースラインの意味マップは右上と右下のコーナーのあたりに線が密集しているが、基本的には、全体の単語の配置と二語間の類似度との対応は説明しがたい。補完類似度を用いた意味マップの方は右下を基点に左上へ放射状に単語が配置されている。補完類似度の特徴である包含関係が、マップ全体の単語の配置に反映されていることがわかる。さきに同じ座標軸上の複数の名詞の共起語を比較した結果、両マップに劇的な分類結果の差が出たわけではなかった。同じ座標上になくても、近くの座標に類似した語はプロットされていた。従って、意味マップ上の名詞の分類がどのような結果になったかについては、マップ全体の単語の配置に包含関係を反映した補完類似度を用いた意味マップがわかりやすい。

6. まとめ

補完類似度による意味マップは、右下が共起語が最多の「こと」という抽象名詞が配置されており、「こと」を基点として単語の広がり方に方向性がみられる。「状態」「方向」「意味」「様子」「感覚」「印象」などの抽象名詞が「こと」の近くに位置しており、更にそれを基点に放射線状に広がっている。右下から離れるほど名詞が具体化していく。マップの右上には、「観点」や「立場」などの抽象名詞があるが、これらは、「意味」と強い包含関係を示している。抽象名詞の広がりを解説すると次のようになる。「状態」から「段階」「程度」「数」「量」「時間」「距離」「空間」「奥行き」と分布していく。「方向」から「傾向」「兆候」「影響」「評価」などへと広がり、「状態」と「方向」の広がりの中間には「局面」「状況」「情勢」などが位置している。「様子」の方向には、「顔つき」「しぐさ」「そぶり」などがあり、「感覚」は、「気持ち」と「感触」、「印象」などと関係を強くし、「印象」は「感覚」や「人柄」などの人やものの特徴などと関係していく。一般的に、程度を表す「傾斜」などは「評価」の近くに、「温度」は「思いやり」「情熱」「気候」などの近くに、「弾力性」は「強さ」や「根性」などの近くに、「触感」や「味覚」「色」などは「感受性」の近くに「感受性」は人の特徴を表す抽象名詞の近くに位置している。得られたマップから形容詞は、「状態」「方向」「意味」「様子」「感覚」「印象」などの上位概念ではほとんどの形容詞が共起しうることから、形容詞は、一般的にこれらの抽象概念を複合的に持っていると考えられる。マップ上では、語の散らばり方に方向性があり、上記の抽象概念を基点に右下隅から左上へと放射線上に上位から下位概念へと広がっている。

参考文献

西尾寅弥　形容詞の意味用法の記述的研究
国立国語研究所　秀英出版　1972

情報処理振興事業協会技術センター　計算機用
日本語基本形容詞辞書 IPAL—解説編—　1991

馬青 神崎享子 村田真樹 内元清貴 井佐原均「日本語名詞の意味マップの自己組織化」情報処理学会論文誌 vol.42, No.10 , 2001

山本英子 梅村恭司「コーパス中の一対多関係を推定する問題における類似尺度」自然言語処理　Vol.9 No.2 2002

根元今朝男　「「が格」の名詞と形容詞とのくみあわせ」「電子計算機のための国語研究Ⅱ」国立国語研究所 1969

高橋太郎「文中にあらわれる所属関係の種々相」国語学 103 国語学会 1975

神崎享子　井佐原均「形容詞類の連体用法にみられる連用的な意味」計量国語学 Vol.22 No.2 計量国語学会 1999

Qing Ma, Kyoko Kanzaki, Masaki Murata, Kiyotaka Uchimoto, and Hitoshi Isahara (2000) *Self-Organization Semantic Maps of Japanese Noun in Terms of Adnominal Constituents*, In Proceedings of IJCNN'2000, Como, Italy, vol.VI.