Vol. 42 No. 10 情報処理学会論文誌 Oct. 2001

日本語名詞の意味マップの自己組織化

馬 青[†] 神崎 享子[†] 村田 真樹[†]
内元 清貴[†] 井佐原 均[†]

本論文は，日本語名詞の意味を連続的かつ可視的に表現する意味マップを神経回路網モデルSOMによる自己組織化によって自動構築する手法を提案する．共起する連体修飾要素の観点から，まず，意味マップの自己組織化に有効と思われる，連体修飾要素が名詞の具体的な内容を表すような名詞句を新聞から人手で収集し，その名詞句を用いた意味マップの構築を試みる．そして，大規模意味マップの構築にはデータ収集の自動化が不可欠という観点から，新聞から名詞及びそれと共起する形容詞と形容動詞を共起頻度の高いものから自動的に取り出して構成される名詞句を用いた意味マップの構築を試みる．計算機実験で得られた意味マップはまず実際に用いた学習データを用いて検討し，意味マップ上の名詞は全般的に学習データが示唆する意味で配置されていることを確かめる．そして，分類結果に可視性や連続性のない階層型クラスタリング手法との比較を行い，本手法の分類能力を評価する．更に，可視化能力を有す多変量解析手法が本タスクにうまく適用できないことを主成分の寄与率分析及び計算機実験を通じて明らかにし，提案手法の必要性を補強する．

Self-Organizing Semantic Map of Japanese Nouns

Qing Ma ,[†] Kyoko Kanzaki ,[†] Masaki Murata ,[†]
Kiyotaka Uchimoto [†] and Hitoshi Isahara[†]

A method is described for automatically constructing a semantic map, a visible and continuous representation in which Japanese nouns with similar meanings are placed at the same or neighboring points so that the distance between them represents semantic similarity. This is done by using the self-organizing neural network, SOM. From the point of view of common adnominal constituents, we first manually gather noun phrases whose adnominal constituents concretely describe the contents of head nouns from newspapers and construct a semantic map of the nouns using these noun phrases. Such types of noun phrases are thought to be effective for self-organizing a semantic map. Because it is indispensable to gather data automatically for constructing a large semantic map, we then construct a semantic map of the nouns using the noun phrases that consist of nouns and their co-occuring adjectives and nominal adjectivals. They are gathered automatically from newspapers in the order of the frequency of their co-occurrent words. Examination of semantic maps obtained in computer experiments showed that the nouns were mapped to the points corresponding to the training data. And, to objectively evaluate the SOM's ability in semantic classification, the semantic maps are compared to the results of classification by hierarchical clustering, which cannot give results with visible and continuous representation. Further, it is clarified that the multivariate statistical analysis such as principle component analysis and factor analysis cannot be used to construct semantic maps which reinforces the necessity of the proposed method for this task.

1. はじめに

単語間の意味的な類似性を自動的に判断する技術は，用例に基づく機械翻訳，文書分類，情報検索といった数多くの自然言語処理のタスクだけでなく，言語学の分野にも役立つ．これまでコーパスに基づく単語のクラスタリングの手法が多数提案されてきた（英語においては，例えば，文献1)と2)；日本語においては，例えば，文献3)）．しかしながら，単語のクラスタリングは基本的には単語をいくつかのグループに分類する技術である．このような手法では，グループ間の意味的な関係，または単語間の意味的距離を直観的に知ることが困難である．もっとも，単語を単純に特定のグループに分類すること自体でさえ困難な問題であるかもしれない．なぜならば，単語が意味的にいくつかのグループにまたがる可能性も十分ありうるからである．これらの問題を解決するためには，クラスタリングの

† 独立行政法人通信総合研究所
Communications Research Laboratory, Japan

代わりに単語を，意味的類似性を距離とする連続した意味空間☆にマッピングする技術が必要である．ここでは意味空間にマッピングされた結果を意味マップと呼ぶ．

このような単語の意味を連続的かつ可視的に表現する意味マップは自己組織化によって自動構築可能なものである．意味マップの自動構築には，単語間の意味関係を反映するようななんらかの情報を用いる必要がある．その情報は意味マップの使用目的に応じて，例えば，対象語を中心とした前後数個の共起単語列，動詞や名詞のような格関係にある単語，または係り先の単語といったような個別なものでもよいし，汎用性を考えてこれらを総合したものでもよいであろう．個別な情報を用いて構築される意味マップは特定のタスクに最適と考えてよいが，他のタスクには適用できない可能性が高い．一方，総合した情報で構築される意味マップは例えば分類語彙表[4]のような一般的な日本語名詞の意味表現に類するものと捉えることができる．このような意味マップは個々のタスクの角度から見れば最適でないかもしれないが，数多くのタスクに利用できよう．従って，どのような情報が大切かがはっきり分かっているタスクにおいては，意味マップをその情報を用いて構築したほうがよいであろうし，どのような情報が特に有用かがよく把握できないタスクにおいては不要なバイアスを避けるために汎用性の高い意味マップを構築したほうがよいであろう．このように，意味マップの構築はタスクを見極めた上で目的型と汎用型のどちらかを選定する必要があると考える．

これまで，英語においては，単語の意味マップを自己組織化によって自動構築する研究がいくつかなされてきた（文献 5) を参照）．しかし，日本語においては，このような研究が系統的に行われていない☆☆．また，英語などに関する研究においても，自己組織化に用いるデータは目的に応じて作ったものでもなければ汎用性を目指したものでもなく，単に目標単語と前後の 2 単語で構成される三つ組で構成されるものであった．上にも述べたように，実用的な意味マップの構築を目指すならば，まず目的型か汎用型かを決め，その上で，適切な言語構造をもつデータを用いる必要があろう．そして，言語構造を考慮するならば，言うまでもなく異なる言語の意味マップをそれぞれその言語の特性に合わせて構築していかなければならない．

上に述べたように意味マップには目的型と汎用型があるが，本稿では目的型意味マップを扱う．具体的には，連体修飾要素の観点からの日本語名詞の意味マップを構築することを目的とする．従って，本稿で扱う意味マップの構築には名詞を形容詞や形容動詞といった連体修飾要素を情報として用いる．このような意味マップは，著者らが進めている，連体修飾要素の意味的ふるまいをそれらと共起する被修飾名詞群との関係を調べることによって体系的に把握する研究[7]を手助けするために必要である．自動構築マシンとしては Kohonen が提案した自己組織化型神経回路網モデル（Self-Organizing Map，略して SOM）を用いる（詳細は文献 5) を参照）．研究の第一段階として，まず，意味マップの自己組織化に有効と思われる，連体修飾要素が名詞の具体的な内容を表すような名詞句を新聞から人手で収集し，その名詞句（66 個の被修飾名詞を含む 7,852 個の名詞句）を用いて名詞の意味マップ構築を試みた．そして，大規模意味マップの構築にはデータ収集の自動化が不可欠という観点から，8 年分の新聞から名詞及びそれと共起する形容詞と形容動詞を共起頻度の高いものから自動的に取り出して構成される名詞句（100 個の被修飾名詞を含む 33,870 個の名詞句）を用いて名詞の意味マップの構築を試みた．計算機実験の結果，まだ，実験の規模が小さいが，我々の直観に合う，つまり，意味マップ上の名詞は全般的に学習データが示唆する意味で配置されている，意味マップの構築が可能であることが分かった．

このような可視的な意味マップの構築には当然ながら SOM による自己組織化以外の方法も考えられる．その有力候補として例えば主成分分析や因子分析といった多変量解析手法を挙げることができよう．SOM を用いる必要性を見るために，このような多変量解析が本タスクに適応できるかどうかについて，同じデータを用いて，主成分や因子の寄与率分析及び計算機実験を行った．その結果，本タスクにおいて多変量解析の適用が困難であることが確認された．また，SOM の可視的な表現力とは別にその有効性を見るために意味分類能力も評価する必要があろう．そのために，分類結果に可視性や連続性のない，階層型クラスタリング手法をベースラインモデルとしてとらえ，同じデータを用いた計算機実験も行った．その結果，分類能力においては，SOM はクラスタリング手法よりやや優れていることが分かった．

以下，2 章では，自己組織化型神経回路網モデル SOM について簡単に紹介する．3 章では，SOM に

☆ つまり，意味的に類似性の高い単語どうしは近くに，意味的に類似性の低い単語どうしは遠くに配置されるような空間である．

☆☆ 日本語への自己組織化技術を利用したものとしては例えば清田らの電子ニュース記事の分類に関する研究[6]）があるが，意味マップの研究ではない．

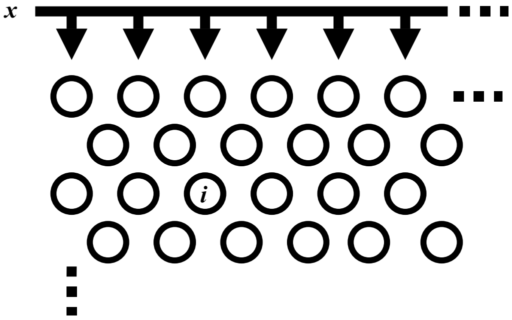

図 1　コホーネンの 2 次元 SOM
Fig. 1　Two-dimensional SOM.

よる意味マップの自己組織化にどのようなデータが有効か，またこのようなデータは SOM が取り扱えるようにどう符号化すればよいかについて述べる．4 章では計算機実験の結果について述べる．この中で，自己組織化によって得られた意味マップが妥当かどうかの学習データを用いた検討と，従来のクラスタリング手法による意味分類結果との比較による分類能力の検証を行う．そして，本タスクにおいては多変量解析がうまく機能しないという分析及び実験結果についても述べる．5 章では結論と今後の課題について述べる．

2. 自己組織化神経回路網モデル SOM

自己組織化神経回路網モデル SOM（図 1）は高次元入力を持つ 2 次元配列のノード（ニューロン）で構成され，以下に述べる自己組織化によって高次元データを 2 次元空間にその特徴を反映するように非線形的に射影（マッピング）することができる．

入力 $x = [\xi_1, \xi_2, \cdots, \xi_n]^T \in \Re^n$ ならば，個々のノード i はそれぞれ参照ベクトル $m_i = [\mu_{i1}, \mu_{i2}, \cdots, \mu_{in}]^T \in \Re^n$ を持つものとする．但し，参照ベクトルの要素 μ_{ij} はノード i と入力要素 ξ_j の間の重みであり，自己組織過程において少しずつ修正される．具体的には，入力ベクトル x が与えられたとき，まず，その入力をすべてのノードの参照ベクトルと比較し，ユークリッド距離の一番短いノードを活性化する．その活性化されたノードを勝者ノードと呼ぶ．即ち，勝者ノード c は以下のように選ばれる．

$$c = argmin_i\{||x - m_i||\} \tag{1}$$

そして，グローバルに自己組織化が行われるように，勝者ノードの近傍のノードも活性化させ，リラックス処理を行う．即ち，活性化されたすべてのノードに対し，それらの参照ベクトルを入力ベクトルに近付くように修正を行う．

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \tag{2}$$

ここで，t は学習回数で，$h_{ci}(t)$ は例えば以下のように定義された近傍関数である．

$$h_{ci}(t) = \alpha(t) \cdot exp(-\frac{||r_c - r_i||^2}{2\sigma^2(t)}) \tag{3}$$

但し，$r_c \in \Re^2$ と $r_i \in \Re^2$ はそれぞれ勝者ノード c と近傍ノード i の位置ベクトルである．従って，項 $||r_c - r_i||$ は近傍ノード i が勝者ノード c から離れて行くにつれ，h_{ci} が小さくなり $m_i(t)$ の修正量が小さくなることを意味する．また，$\alpha(t)$ は学習率で，$\sigma(t)$ は近傍の大きさ（半径）である．これらは時間と共に単調に減少していく関数であり，例えば以下のようなものである．

$$\alpha(t) = \alpha(0)\frac{T-t}{T}, \tag{4}$$

$$\sigma(t+1) = 1 + (\sigma(t) - 1)\frac{T-t}{T} \tag{5}$$

但し，$\alpha(0)$ は初期値で，T は総学習回数である．

自己組織化が終った後のマッピング処理は，単に入力 x に対する勝者ノードを選び出す [式 (1)] ことである．従って，上に述べた自己組織化（即ち，入力データに応じた参照ベクトルの修正）は入力にマッチするノードの整列過程と見なすことができる．即ち，マッチするノードを入力に近付ける過程である．但し，この処理はグローバルに行われるため，結果的には，似た入力への勝者ノードが近いところに配置され，そうでないノードは離れたところに配置されるようになる．

通常，学習は「整列」と「微調整」の二つのフェーズに分けて行われる．「整列」フェーズでは，$\alpha(t)$ と $\sigma(t)$ の初期値を共に大きく取り，時間と共に (4)(5) に従って減少して行く．ノードの配置の基本形はこのフェーズで形成される．一方，残りのフェーズでは $\alpha(t)$ と $\sigma(t)$ は小さい値のまま長時間をかけて初期フェーズで形成された基本形を微調整する．

3. 意味マップの自己組織化

本稿で取り扱う名詞の意味マップとは，意味的に類似性の高い名詞どうしが近くに配置され，意味的に類似性の低い名詞どうしが遠くに配置されるような，意味的類似性を距離とする 2 次元表現を指す．そして，意味マップの自己組織化とは，このような表現を自己組織化型神経回路網モデル SOM を用いて自動構築することである。

3.1 学習データ

著者らは連体修飾要素の意味的ふるまいを被修飾名詞を介して体系化する研究を進めている．このために，新聞などから集めた大量の被修飾名詞とそれと共起す

る連体修飾要素群をセットとした言語データに対する分析を行う必要がある．このような大量データを連体修飾要素の観点から分類し，その結果を可視的に表現できれば，分析作業の効率が飛躍的に向上することが期待できる．本稿はこのような視点で名詞意味マップの自己組織化を考える．

名詞の意味マップを自己組織化させるためには，名詞間の意味的な関係を反映できるような学習データが必要である．一般的に，名詞間の意味的な関係は名詞と共起する動詞や形容詞，さらには係り先の単語などを通じて結びつけることができる．本稿は上述したように，連体修飾要素の観点からの名詞の意味マップを構築したいので，名詞間の関係を共起する連体修飾要素を通じて結びつけることとした．実際，ある意味では，意味的に近い名詞どうしは多くの共通する連体修飾要素を持ち，意味的に遠い名詞どうしはあまり共通する連体修飾要素を持たないと考えることができる．ここで「思い」，「気持」，そして「観点」という三つの名詞を例として考えよう．「思い」と「気持」は意味的に近い名詞どうしのため，共通する連体修飾要素「うれしい」や「あこがれの」を持ち得る．しかし一方，「思い」や「気持」と意味的に遠い関係にある名詞「観点」は，同様な連体修飾要素を持つことができない．従って，本稿では，個々の名詞をそれと共起する連体修飾要素のセットを用いて定義し，そのセットを学習データとする．

実際，名詞句を構成する連体修飾要素と被修飾名詞の間には種々の関係が存在する．たとえば，「うれしい思い」の例では，連体修飾要素「うれしい」は名詞「思い」の具体的な内容を表す．一方，「太郎の思い」の例では，「太郎の」は「思い」の所有者を表す．連体修飾要素は名詞の属性として働く場合もある．例えば，「赤いリンゴ」の例では，「赤い」は「リンゴ」の属性である．これらの関係のうち，連体修飾要素が被修飾名詞の具体的な内容を表すものは特に自己組織化にとって有効であると考えられる．なぜなら，被修飾名詞の具体的な内容を表す連体修飾要素は被修飾名詞と共通する意味要素を持っているからである．すなわち，この種の名詞句の場合，被修飾名詞は，連体修飾要素の表す事柄を抽象化した意味を持つと考えることができる．例えば，「思い」は「うれしい」を抽象化した意味を持ち，「うれしい」は「思い」を具体化した表現で，互いに，「思い」という意味要素を共有している（詳細は文献 7) を参照）．

そのため，本稿では名詞の意味マップの構築に，まずこのタイプのデータを抽出して用いることにした．

表 1　収集されたデータの例
Table 1　Example of gathered data

名詞	連体修飾要素
思い	うれしい，あこがれの，いとしい，気の毒な $\cdots$．
気持	うれしい，あこがれの，いい加減な，うらやましい $\cdots$．
観点	アカデミックな，教育的な，歴史の，$\cdots$．
⋮	⋮

表 1は新聞から人手で収集した，連体修飾要素が名詞の具体的な内容を表すようなデータを示す．この表の 1 列目は被修飾名詞であり，2 列目は共起する連体修飾要素である．この表からも分かるように，連体修飾要素として，形容詞の他，形容動詞，そして「名詞＋の」も用いられている．但し，このタイプのデータは自動抽出が困難で言語学の専門家に頼らなければならないため，データの収集にコストがかかる．そのため，本研究では，さらに，連体修飾要素と被修飾名詞の関係を考慮せずに形容詞と形容動詞を共起語とする名詞句を新聞から自動抽出して用いた名詞マップの構築も試みることにした．

3.2　符　号　化

前の節で述べたように，個々の名詞は共起する連体修飾要素のセットで定義する．例えば，表 1から名詞「思い」を

思い＝｛うれしい，あこがれの，気の毒な, $\cdots$｝

のように定義する．ここで，一般に ω 種類の名詞 $w_i(i = 1,\cdots,\omega)$ 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．

$$w_i = \{a_1^{(i)}, a_2^{(i)}, \cdots, a_{\alpha_i}^{(i)}\} \qquad (6)$$

但し，$a_j^{(i)}$ は w_i と共起する j 番目の連体修飾要素で，α_i は w_i と共起する連体修飾要素の数である．このように定義された名詞を SOM が取り扱えるようにするためにはまずそれらを符号化しなければならない．符号化の方法としては，まず英語の意味マップの構築に用いられてきた「ランダムコーディング法」（詳細は文献 8) を参照）が考えられる．しかし，この方法は本研究においてはあまり有効ではないことを先行実験（詳細については文献 9) を参照）によって判明したため，本研究では以下に述べる相関コーディング法を新しく導入した．

相関コーディング法においては，まず，それぞれ二つの名詞間の意味的相関（あるいは意味的距離）を反映するものを求める．ここでは，名詞 w_i と名詞 w_j 間の意味的距離 d_{ij} を (6) の定義式を用いて以下のような計算式によって求めることとした．

表 2 名詞の相関行列
Table 2 Correlative matrix of nouns

	w_1	w_2	$\cdots$	w_ω
w_1	d_{11}	d_{12}	$\cdots$	$d_{1\omega}$
w_2	d_{21}	d_{22}	$\cdots$	$d_{2\omega}$
$\vdots$			$\ddots$	
w_ω	d_{w1}	d_{w2}	$\cdots$	$d_{w\omega}$

$$d_{ij} = \begin{cases} \frac{(\alpha_i - c_{ij}) + (\alpha_j - c_{ij})}{\alpha_i + \alpha_j - c_{ij}} & \text{もし } i \neq j \\ 0 & \text{その他} \end{cases} \quad (7)$$

但し，α_i と α_j はそれぞれ w_i と w_j が共起する連体修飾要素の数で，c_{ij} は w_i と w_j に共通する連体修飾要素の数である．従って，意味的関係 d_{ij} は名詞 w_i と w_j の間にどれだけ共通の連体修飾要素を持つかという意味での両名詞間の正規化された距離である．即ち，d_{ij} が小さければ，両名詞は意味的に近く，d_{ij} が大きければ，両名詞は意味的に遠い．

但し，個々の d_{ij} はあくまでも名詞 w_i と名詞 w_j を孤立に見る場合の関係であり，他の名詞も考慮にいれた場合，その他の名詞を参照系として考えるときのこの二つの名詞間の関係や，この二つの名詞と他の名詞との関係はこのような d_{ij} の集合を用いるだけでは反映できない．その意味ではこのような意味的関係は局所意味的関係と呼んだほうがより正確であろう．しかし，このような個々の局所意味的距離から表 2に示すような行列を作成すれば，各行はそれぞれ見出しとなる名詞とそれを除いた $\omega - 1$ 個の名詞との局所意味的距離から構成されていることが分かる．即ち，各行はある程度大局的にその見出しとなら名詞と他の名詞との意味的な関係を反映していると考えられる．

従って，本稿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．

$$V(w_i) = [d_{i1}, d_{i2}, \cdots, d_{i\omega}]^T \quad (8)$$

ここで，$V(w_i) \in \Re^\omega$ は SOM の入力である（即ち，$x = V(w_i)$，$n = \omega$）．即ち，SOM の役割は，このような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させて 2 次元的に表現することである．

4. 実験結果

4.1 データ

実験では，まず，毎日新聞から人手で収集した，連体修飾要素が名詞の具体的な内容を表すような名詞句を用いた（今後，このデータをデータ 1 と言う）．名詞の種類 ω は 66，連体修飾要素ののべ総数 $\sum_{i=1}^{i=\omega} \alpha_i$ は 7,852，連体修飾要素の異なり総数は 4,998 であった．このうち，名詞「状態」の連体修飾要素の数（$\alpha_{状態} = 391$）がもっとも多く，名詞「中 2」☆の連体修飾要素の数（$\alpha_{中2} = 4$）がもっとも少なかった．また，名詞「思い」と「気持」がもっとも多く共通の連体修飾要素を持ち（$c_{思い,気持} = 79$），多数の名詞が互いに共通する連体修飾要素を持たなかった．

次に，8 年分の毎日新聞から共起する形容詞と形容動詞の頻度順から自動的に集めた形容詞／形容動詞と名詞で構成されるような名詞句を用いた（今後，このデータをデータ 2 と言う）．名詞の種類 ω は 100，形容詞／形容動詞ののべ総数 $\sum_{i=1}^{i=\omega} \alpha_i$ は 33,870，形容詞／形容動詞の異なり総数は 4,023 であった．このうち，名詞「もの」と共起する形容詞／形容動詞の数（$\alpha_{もの} = 1,870$）がもっとも多く，名詞「政治家」と共起する形容詞／形容動詞の数（$\alpha_{政治家} = 192$）がもっとも少なかった．また，名詞「もの」と「こと」に共通して共起する形容詞／形容動詞の数がもっとも多く（$c_{もの,こと} = 764$），名詞「活動」と「気分」に共起する形容詞／形容動詞がもっとも少なかった（$c_{活動,気分} = 6$）．

4.2 SOM による実験結果

両データに対し，SOM への入力の次元 ω が異なってそれぞれ 66 と 100 に設定された以外は，すべてのパラメターを同じように設定した．SOM は 13×13 の 2 次元配列のノードで構成し，近傍の形状は六角形にした．整列フェーズにおいては，学習総回数 T を 10,000 に，学習率の初期値 $\alpha(0)$ を 0.1 に，そして，近傍の初期半径 $\sigma(0)$ を 13（即ち，SOM の直径に等しい）に設定した．微調整フェーズにおいては，学習総回数 T を 100,000 に，学習率の初期値 $\alpha(0)$ を 0.02 に，そして，近傍の初期半径 $\sigma(0)$ を 13 に設定した．

図 2(a) にデータ 1 を用いて得られた名詞の意味マップを示し，図 2(b) にこれを人手で意味的に七つの領域（グループ）に分けた図を示す．以下，得られた結果が妥当かどうかを各グループごとに学習データを使って（即ち，連体修飾要素の観点から）検討を行う．

＜感情・感覚：思い，気，気持，気分，くらい，ほど，感じ 1，感情，うち，空気，念，声，情，影，そぶり，ため息，どん底，色 2＞

左上から左下までの領域にマッピングされた名詞は

☆ 添字は同じ単語の（連体修飾要素の観点からの）異なる意味を区別するために用いている．例えば，「くらやみの中」と「短い中」の場合では，連体修飾語群の性格が異なる（前者が状態，後者が範囲を表す）ので，「中 1」（状態），「中 2」（範囲）などとして区別しておいた．

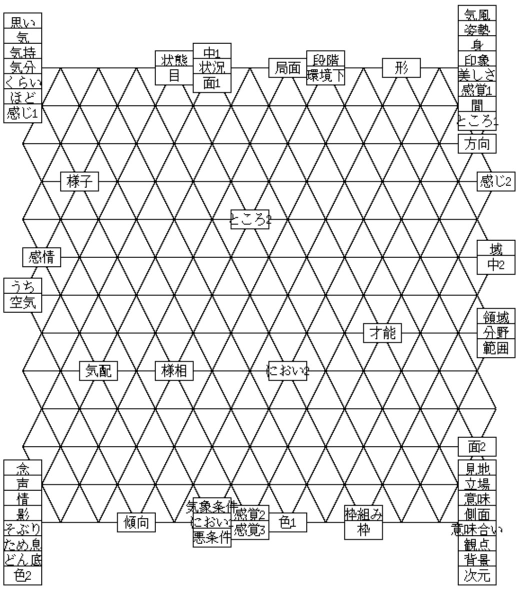

図 2(a) データ 1 を用いた意味マップ
Fig. 2(a) Semantic map using data #1.

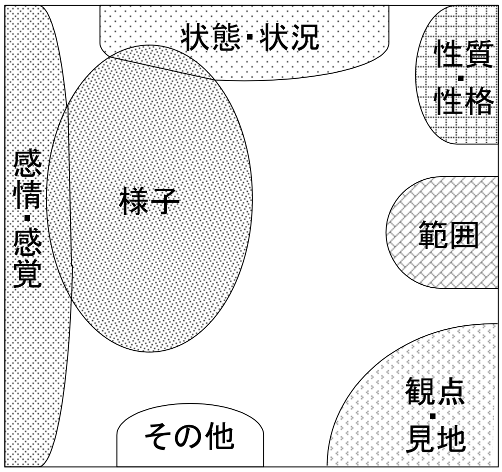

図 2(b) 意味マップに対応する意味グループ
Fig. 2(b) Semantic groups corresponding to the map.

基本的に＜感情・感覚＞を表すものと考えてよい．この領域の連体修飾関係は「悲しい思い」／「あやふやな気」／「拍子抜けの気分」／「疲労の色」などの感情や感覚を表す表現である．このグループは直接感情を表すもの（思い，気，…）と感情の程度を表すもの（くらい，ほど，どん底）を含んでいるが，分布の傾向としては，「思い」や「気持」など内的な感情の表現と，「声」や「ため息」など感情を表に出す場合の表現が分離されている．この領域には，左端中央に「空気」が出現しているが，これは「気まずい空気」などのように「雰囲気」を表すものである．連体修飾要素からみると，「空気」は＜感情 感覚＞の名詞と共通の連体修飾要素を多く持つが，それだけではなく＜様子＞の領域の名詞（様子，気配，様相）とも共通する連体修飾要素が多い（例えば「不穏な空気」）．従って，図 2(b) の意味グループの図では＜様子＞の領域の一部を＜感情・感覚＞の領域に重ねて図示している．また，同様の例として，名詞「様子」は＜様子＞の領域だけでなく，＜感情・感覚＞の領域（例えば「感激の様子」）や，＜状態・状況＞の領域（例えば「悪戦苦闘の様子」）とも連体修飾要素を共有している．実際，図 2(a) をみると，名詞「様子」の位置は，＜感情・感覚＞や＜状態・状況＞の領域に寄っていることが分かる．従って，＜様子＞の領域を＜感情・感覚＞の領域とだけでなく＜状態・状況＞の領域とも一部分重ねて図示した．

＜状態・状況：状態，目，中１，状況，面１，局面，段階，環境下，形＞

上中央の領域にマッピングされた名詞は＜状態・状況＞を表すものである．その表現としては「無意識の状態」／「ひどい目にあう」／「忙しい中」／「有利な局面」／「なしくずしの形」などがある．その中で，被修飾名詞「形」が，「孤立無縁の形で解任され …」のように状態や状況を表すだけでなく，「正式な形」や「最善の形」といった特徴ともとれるような表現も含むため，「形」が隣接する＜性質・性格＞領域寄りに位置していると考えられる．また，被修飾名詞「ところ２」は「土壇場のところ」や「あぶないところ」といったように状態・状況を表していると考え，このグループにマッピングされることを期待していたが，中央部の離れているところにマッピングされてしまっている．しかし，実際の学習データを調べた結果，「つらい」，「苦しい」，「冷たい」といった「ところ２」の共起語は，＜状態・状況＞の名詞よりも，＜感情・感覚＞の名詞（例えば「思い」）とより多く共通していることが分かった．従って，「ところ２」がこの両グループのほぼ中間にマッピングされているのはこのためであると考えられる．一方，名詞「面１」は「意地っぱりの面」や「独りよがりの面」など性質や性格を表すため，隣接の＜性質・性格＞領域にマッピングされる

べきであるが，学習データがそれほど少なくなかったにもかかわらず，うまく分類されなかった．

<性質・性格: 気風，姿勢，身，印象，美しさ，感覚１，間，ところ１，方向，感じ２>

右上の領域にマッピングされた名詞は基本的に<性質・性格>を表すものと考えてよい．その表現としては，例えば，「質実剛健の気風」/「強気の姿勢」/「一文なしの身」/「やさしい印象」/「清楚な美しさ」/「親切なところ」/「楽な方向」などがある．この領域にマッピングされる被修飾名詞「感覚１」は，例えば，「ユーモラスな感覚」や「優美さの感覚」といった性質や特徴を表す連体修飾関係を表すものである．一方，被修飾名詞「感覚２」は「恐怖の感覚」や「痛みの感覚」といった<感情・感覚>を表現するため，<感情・感覚>の領域に出現すべきであるが，この表現の学習データが少なく<感情・感覚>に正しくマッピングされた名詞との共通する連体修飾要素が少なすぎるため，<その他>に入ってしまった．また，この領域に位置する名詞「感じ２」は，「好人物な感じ」や「独り善がりな感じ」などのような性質や性格を表す場合と「NTT 有利の感じ」というような状態や状況を表す場合があるために，本来なら隣接の<状態・状況>との間にマッピングされるべきであった．それはもともとこれに関連する学習データが少ない上，状態や状況よりも性質や性格の表現が多いためであると考えられる．しかし，名詞「間」は「忙しい間」や「留守の間」のように状態や状況などを表すため，隣接の<状態・状況>領域にマッピングされるべきであるが，それの学習データがそれほど少なくなかったにもかかわらず，うまく分類されなかった．

<範囲： 域，中２，領域，分野，範囲>

右中央の領域にマッピングされた名詞は基本的に<範囲>を表すものと考えてよい．それらの被修飾名詞は，「常識の域」/「ユーモラスな中（に情味をただよわせている）」/「政治的な領域」/「科学的な分野」/「合理的な範囲」などのように範囲を表している．この領域は，（域，中２）と（領域，分野，範囲）とがやや分離している．「域」と「中２」は例文からも分かるように<性質・性格>の表現と近く，「領域」と「分野」と「範囲」は次に述べる<観点・見地>と近いと考えられる．また，「才能」がこの語群の近くに位置している．それは「音楽の才能」や「文学的な才能」といった「分野」と共通する表現が多いためであると考えられる．

<観点・見地： 面２，見地，立場，意味，側面，意味合い，観点，背景，次元，枠組み，枠>

右下の領域にマッピングされた名詞は<観点・見地>を表すものである．この表現領域には，「技術の面」/「財政的な見地」/「否定の立場」/「文化的な背景」/「政治的な枠組み」などの例がある．「意味」，「意味合い」，「側面」などもこの表現領域に含まれている．これは例えば，「精神的な意味（での体質改善）」/「肯定的な意味合い」/「教育的な側面」などのように，これらの被修飾名詞はある観点や立場を表明する意味機能を持っているためと考えられる．

<様子： 様相，気配，うち，空気，感情，様子>

これらの名詞は，その配置が少々互いに離れるが，基本的に<様子>領域をなすと考える．この表現領域には例えば，「深刻な様相」/「重苦しい気配」/「うやむやのうち」/「不穏な空気」などの様相や雰囲気を表す表現である．また，被修飾名詞「感情」は，<感情・感覚>の名詞とだけでなく，この領域の名詞とも共通の連体修飾要素を多く持っているために，この両領域の重なった領域に入っている．

<その他： 傾向，気象条件，におい１，悪条件，感覚２，感覚３，色１>

この領域にマッピングされた名詞は，全体としてのこれらの位置付けが分からないので，<その他>にしておいた．

以上，すべての領域について述べた．ここで一つ注意されたいのは，例えば，前述した<観点・見地>グループにおいては，名詞（面2，見地，$\cdots$，次元）と名詞（枠組み，枠）の間の意味上の微妙な違いは異なる領域にマッピングされることによって表現されていることである．もちろん，分け方によってはこの二つの領域にマッピングされる名詞を二つのグループに分けることも可能である．しかし，この二つの領域が近接しているため，これらの中の名詞が意味的に近いという結果には変わりがない．似たようなことは<観点・見地>グループと<範囲>グループの間にも言える．この二つのグループの名詞は意味的に相似しているため，それぞれ隣どうしの領域に位置するグループにマッピングされている．実際，<範囲>グループの名詞「分野」や「領域」は<観点・見地>グループに分類されてもおかしくない．今回は，これらを二つのグループに分けたが，いずれにせよこの二つのグループの中の名詞が意味的に近いという結果に変わりがない．このように，意味マップは意味の相似性を連続的に表現しているため，意味マップを用いれば，どの名詞がどのグループに分類できるかを知ることだけでなく，名詞どうしが互いにどれぐらい意味的に近いかを知ることもできる．そして，名詞間の意味的相似を見

る際にはそれらの名詞が同じグループに属するかどうかを考える必要はなく，また，その名詞がどのグループに分類すべきかが決まらない場合でも意味的な近さの比較が可能である．

以上に述べたように，自己組織化されたマップは全般的に妥当である，すなわち，学習データあるいは連体修飾要素の観点から妥当な名詞の分類結果を得たと言ってよい．うまく行かなかった例もあるが，本実験に限って言えばそれは基本的に学習データのスパースネス問題に起因するものであった．その例として <その他> にマッピングされた名詞（傾向，気象条件，$\cdots$，色 1）がある．学習データを調べた結果，これらの名詞は他の名詞との間に共通する連体修飾要素が極端に少ないことが分かった．例えば，名詞「気象条件」は他の名詞との間に最も多い場合でも三つの共通の連体修飾要素しか持っていない．

次に，データ 2 を用いた実験結果について述べる．図 3(a) にデータ 2 を用いて得られた名詞の意味マップを示し，図 3(b) にこれを人手で意味的に分類した図を示す．この意味マップを大まかに見て，まず，感性的活動 <感情> を表す名詞（表情，態度，気持，感じなど）と理性的活動 <思考活動> を表す名詞（意見，判断，考え，批判など）がそれぞれ左下と右上の対角領域に，そして，社会性 <社会的トピック> を表す名詞（社会，環境，情報，政治，数字など）と文化性 <文化活動> を表す名詞（作品，音楽，舞台，プレーなど）がそれぞれ上中央と下中央の領域にマッピングされていることが分かる．そして，右の大半の領域を占める <行動・展開> グループは <文化活動> のグループとともに，感性と理性の表現領域の中間に配置されている．この配置は，その結果の必然性を別にして，文化活動を含む人間の種々の行動は理性と感性の総合作用によって発生するという人間の外的行為と内的精神活動との関係を反映している，というように解釈を与えることが可能であろう．そして，外的行為の中も，文化活動のほうがより感性に頼っているという意味で，<文化活動> が <行動・展開> よりも <感情> の近くに配置されていることも解釈可能となろう．残りの領域は，他の領域との相対的な関連性が薄く，それぞれ独立していると考える．具体的に，左上の領域にマッピングされた名詞（時代，時，$\cdots$, 国）は基本的に <時・場所> を表すものとしてとらえることができ，左中央の領域にマッピングされた名詞（方，子供，$\cdots$，商品，人物，物，本，女性，$\cdots$，男）は基本的に <人・物> を表すものと考えることができる．そして中央にマッピングされた名詞（関係，意味，$\cdots$，力）はどのグループにも分類されていないものとしてとらえることができる．このように，ただ単に共起語の頻度順で選んだ，我々人間にとってもその意味付けや分類が非常に困難と思われる名詞群も，SOM を用いることによってある程度我々人間の直観に合う意味マップを得ることができる．

4.3 他手法との比較

4.3.1 階層型クラスタリング手法

SOM が意味マップの構築に有効であることを確かめるためには，その分類能力を評価する必要がある．そのため，同じデータ 1 とデータ 2 を用いて階層型クラスタリング手法☆との比較実験を行った．クラスタリングにかけられたデータは SOM への入力ベクトルと同様，3.2 節に述べた相関コーディング法を用いて構成した．表 3と表 4はそれぞれデータ 1 とデータ 2 に対する分類結果（1 列目と 2 列目）を示す．但し，得られたクラスタリング結果は階層的になっているため，ここに示した分類結果は以下に述べる方法で得た．一般的に，階層型クラスタリング結果から表に示したような分類結果を得るためには二通りの方法が考えられる．一つの方法は，グルーピングしようとする階層のレベルを決め，その決められたレベルでデータを自動的にグルーピングすることである．しかし，この方法で得られるグループは，その中のデータが極端に多かったり少なかったりするため，偏った分類結果しか得ることができない．そこで，本比較実験では，SOM で得た意味マップをグループ分けする場合と同様，階層的にクラスタリングされた結果をできるだけ最適であるように人手でグルーピングするという方法を取った．

まず，図 2 の意味マップと表 3のクラスタリング結果を比較してみよう．基本的にこの両者はよく対応が取れていると言える．分かりやすくするために，表の右側には，それぞれのクラスに対応すると思われる意味マップ上のグループ名を付け加えた．具体的には，表 3のクラス 1 と 4 は図 2(a) の <感情・感覚> グループ，クラス 2 は <性質・性格> のグループ，クラス 3 は <状況・状態> と <様子> グループ，クラス 5 と 6 は <範囲> と <観点・見地> のグループ，そしてクラス 7 が <その他> グループに対応付けすることができる．もちろん両者間に細かいところに色々な違いもある．ここではまず，両者ともうまく分類ができていな

☆ パッケージソフト SPSS for Windows 6.0J の階層型クラスタリングプログラムを用いた．但し，クラスターの結合方法としてはグループ間の平均連結法を，ベクトルやクラスタ間の距離計算には平方ユークリッド距離を採用した．

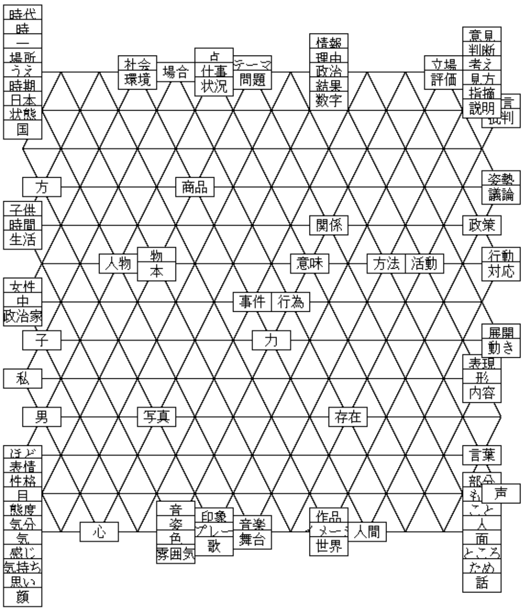

図 3(a) データ 2 を用いた意味マップ
Fig. 3(a) Semantic map using data #2.

表 3　クラスタリング手法によるデータ 1 の分類結果
Table 3　Classified results for data #1 by clustering method

クラス	名詞	意味マップとの対応
1	ほど くらい 気 感じ 1 気分 思い 気持 感情 様子	＜感情・感覚＞
2	ところ 1 面 1 空気 姿勢 印象 感覚 1 気風	＜性質・性格＞
3	ところ 2 目 身 様相 気配 状態 状況 局面 段階 環境下 形 うち 中 1	＜状況・状態＞＜様子＞
4	そぶり ため息 どん底 影 念 情 色 2 声	＜感情・感覚＞
5	枠 枠組み におい 2 色 1 観点 見地 面 2 背景 側面 意味 立場 意味合い 次元 領域 才能 分野 方向	＜範囲＞＜観点・見地＞
6	域 範囲 間	＜範囲＞
7	感じ 2 傾向 悪条件 気象条件 におい 1 中 2 感覚 3 感覚 2 美しさ	＜その他＞

い例をみてみよう．例えば，図 2(a) の意味マップにおいては，学習データの不足のため我々人間も分類できない名詞「色 1」と「におい 2」はそれぞれ＜その他＞と他の名詞群から離れて孤立に，そして「忙しい間」や「留守の間」など状況の意味を持つ名詞「間」は隣接の＜性質・性格＞のグループに配置されている．

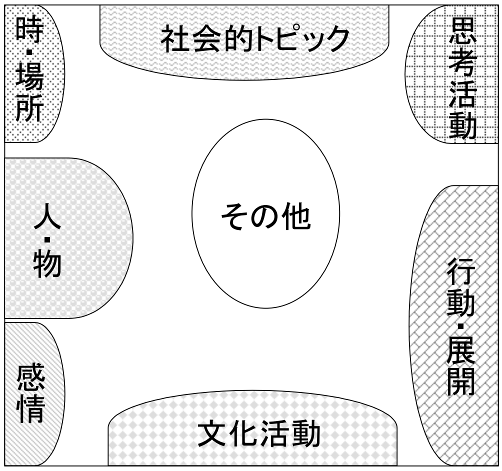

図 3(b) 意味マップに対応する意味グループ Fig. 3(b) Semantic groups corresponding to the map.

一方，クラスタリングの結果では，その三名詞はクラス 5 と 6（＜範囲＞と＜観点・見地＞）に分類されている．次に，意味マップの結果が正しく，クラスタリングの結果が間違っている例としては次のようなものがある．「清楚な美しさ」など＜性質・性格＞に属する名詞「美しさ」は意味マップでは正しく配置されているが，クラスタリングにおいては＜その他＞に相当するクラス 7 に分類されている．同様に「一文なしの身」など＜性質・性格＞に属する名詞「身」も意味マップでは正しく配置されているが，クラスタリングにおいては＜状態・状況＞と＜様子＞を一まとまりにしたものに相当するクラス 3 に分類されている．＜感情・感覚＞と＜様子＞に属する名詞「空気」は意味マップでは正しく配置されているが，クラスタリングにおいては＜性質・性格＞に相当するクラス 2 に分類されている．そして「好人物の感じ」のように＜性質・性格＞に属す場合もある「感じ 2」は意味マップではこの領域に配置されているが，クラスタリングにおいては＜その他＞に相当するクラス 7 に分類されている．さらに，意味マップでは意味の微妙な違いを表現できているのに対し，クラスタリングではうまく反映できていない例として次のようなものがある．クラスタリングの結果においては，クラス 3 は状況と様子の名詞を一まとまりにし，クラス 5 も＜観点・見地＞だけでなく「領域」や「分野」といった＜範囲＞の名詞も含んで

いる．それに対し，意味マップにおいてはそれらのグループ間に重なりはあるものの，離れた場所に配置することによりそれらの違いもはっきり表現している．一方，意味マップの結果が間違っていてクラスタリングの結果が正しい例は以下の一個しか見つけることができなかった．すなわち，「意地っぱりの面」のように＜性質・性格＞に属する名詞「面１」は意味マップでは隣接の＜状態・状況＞に配置されてしまっているが，クラスタリングにおいては正しく＜性質・性格＞に相当するクラス２に分類されている．

次に，図 3 の意味マップと表 4のクラスタリング結果を比較してみよう．この二つの結果の対応付けは表 3と図 2 の対応付けに比べてやや困難である．しかし，例えば，クラス１は＜行動・展開＞，クラス２は＜人・物＞，クラス３は＜時・場所＞，クラス４は＜思考活動＞，クラス５は＜社会的トピック＞，クラス６は＜文化活動＞，そして，クラス７は＜感情＞に大雑把に対応づけることも可能であろう．ここでも分かりやすくするために，表の右側には，それぞれのクラスに対応すると思われる意味マップ上のグループ名を付け加えた．詳細に見ると，例えば，意味マップの＜感情＞グループには１１個（少し離れている名詞「心」を含むと１２個）の感情に関する名詞が入っているのに対し，クラス７には（「態度」と「感じ」を除く）９個の名詞しか入っていない．＜社会的トピック＞に入った「結果」や「数字」が＜思考活動＞に相当するクラス４に入っている．また，例えば＜文化活動＞に入っていいはずの名詞「プレー」，＜行動・展開＞に入るべき名詞「動き」や「展開」，そして，＜人・物＞に入るべき名詞「男」や「私」などは＜その他＞に入っている．一方，逆にクラスタリング結果のほうが妥当なところも見られる．例えば名詞「人間」はクラスタリングの結果においてはクラス２（＜人・物＞）に入っているが，意味マップでは＜文化活動＞に配置されている．また，文化に属すと思われる名詞「写真」はクラスタリングの結果においてはクラス６（＜文化活動＞）に入っているが，意味マップにおいては＜文化活動＞の名詞群から少し離れているところに配置されている．そして，共通してうまく分類できていない例も見られる．例えば＜人・物＞に入るべき名詞「人」は意味マップにおいてもクラスタリングの結果においても＜行動・展開＞のグループに分類されている．これは恐らく「人」は＜行動・展開＞の動作主であるため，「人」と＜行動・展開＞の中の名詞の共起頻度が高いため生じた現象として解釈できよう．

表 4　クラスタリング手法によるデータ 2 の分類結果
Table 4　Classified results for data #2 by clustering method

クラス	名詞	意味マップとの対応
1	もの こと 人 ところ 言葉 表現 内容 形 話 声 面	＜行動・展開＞
2	物 本 商品 人物 政治家 子供 子 方 女性 人間	＜人・物＞
3	日本 国 時 時期 時代 場所 一 状況 ため うえ 部分 点 生活 時間 中	＜時・場所＞
4	結果 数字 指摘 説明 理由 場合 発言 批判 意見 考え 判断 立場 見方 評価	＜思考活動＞
5	政策 活動 議論 行動 方法 問題 テーマ 仕事 情報 政治 社会 関係 環境	＜社会的トピック＞
6	作品 舞台 音楽 歌 音 写真 色 感じ 印象 イメージ 雰囲気 姿 世界 存在	＜文化活動＞
7	ほど 気 気持 思い 心 顔 表情 目 気分	＜感情＞
その他☆	動き 展開 プレー 力 意味 行為 事件 対応 姿勢 態度 男 性格 私	＜その他＞

このように，SOM とクラスタリング手法の両者の比較について，厳密な評価を下すことはできないが，SOM を用いた手法は意味的分類を可視的連続的に表現できるだけでなく，分類能力においてもすくなくとも従来のクラスタリング手法よりやや優れていることが言えよう．

4.3.2　多変量解析手法

可視的な 2 次元意味マップの構築には SOM による自己組織化以外の方法も考えられる．その有力候補として主成分分析や因子分析といった多変量解析手法を挙げることができる．しかし，以下に述べるように，本タスクにはこのような多変量解析手法がうまく適用できない．

主成分分析も因子分析も，多くの変量の持っている情報を少数個の主成分あるいは因子に縮約するための解析手法である．両者の基本的な考え方には相違があるが，少数個の変量で多変量データを説明するという意味で共通性があり，同じ課題に利用されることが多い．そして，もとの変量の持っている情報の大部分を説明できるようにするために主成分あるいは因子を幾つまで必要とするかは，それらの寄与率（即ち，それぞれの分散が全体の中で占める割合）の累積値によって決まる．表 5と表 6は意味マップの構築に用いた 2

☆ 本手法は階層型クラスタリング手法であるため，下位の層で分類されていないものは「その他」にした．

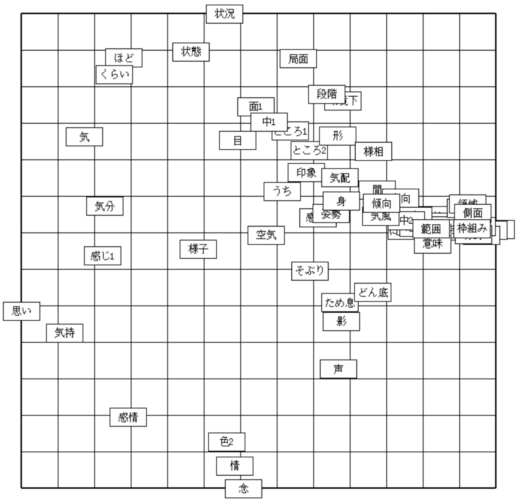

図4 主成分分析結果のプロット図
Fig. 4 Plotting graph of the result of principle component analysis

種類のデータに対し主成分分析☆により得られた，最大の１０個の主成分の寄与率及び累積寄与率を示す．

表 **5** データ１の主成分分析結果
Table 5 Principle components analysis for data #1

主成分	寄与率 (%)	累積寄与率 (%)
1	4.511	4.511
2	3.221	7.732
3	2.498	10.230
4	2.278	12.508
5	2.116	14.623
6	1.875	16.499
7	1.863	18.362
8	1.814	20.176
9	1.780	21.956
10	1.762	23.718

表 **6** データ２の主成分分析分析結果
Table 6 Principle components analysis for data #1

主成分	寄与率 (%)	累積寄与率 (%)
1	9.093	9.093
2	5.882	14.975
3	4.302	19.277
4	4.024	23.301
5	2.825	26.126
6	2.421	28.547
7	2.196	30.743
8	2.031	32.774
9	1.754	34.528
10	1.543	36.072

この二つの表から分かるように第１と第２主成分が合わせた累積寄与率はデータ１とデータ２に対してそれぞれ 15%と 20% にすぎない．１０個の最大の主成分の累積寄与率もそれぞれわずか 23%と 36%であった．一般的には主成分の累積寄与率が 70 ないし 80%以上大きくならないと，多変量データを正しく縮約できないとされている[10]．したがって，明らかにデータ１とデータ２を用いた意味マップの構築においては，多変量解析を用いることは困難である☆☆．実際，データ１を第１主成分と第２主成分を軸とする２次元平面にプロットする計算機実験も行ってみた．その結果，図４に示しているように，意味マップに相当する＜感情・感覚＞グループ（全エリアのほぼ半分を占める左下の領域）と＜状態・状況＞グループがある程度形成されたとしても，それ以外の約三分の二のデータが真中の小さなエリアに集中してうまく分類できなかった．

多変量解析がこのタスクにうまく適用できない主な理由は今回扱ったデータが強い非線形性（つまり線形分離不可性）を持っているためであると推測される．つまり，既存の主成分分析や因子分析手法は，非線形処理を行う SOM と違って，その主成分や因子はもとの変量の線形結合だけで求められているため，線形処理しか行っておらず，非線形性の強いデータ処理には対処できないと思われる．

5. おわりに

本稿は，日本語名詞の意味を連続的かつ可視的に表現する意味マップを神経回路網モデル SOM による自己組織化によって自動構築する手法を提案した．共起する連体修飾要素の観点から，まず，意味マップの自己組織化に有効と思われる，連体修飾要素が名詞の具体的な内容を表すような名詞句を新聞から人手で収集し，その名詞句を用いた意味マップ構築を試みた．そして，大規模意味マップの構築にはデータ収集の自動化が不可欠という観点から，新聞から名詞及びそれと共起する形容詞と形容動詞を共起頻度の高いものから自動的に取り出して構成される名詞句を用いた意味マップの構築を試みた．学習データの符号化は，英語などの意味マップの構築に用いられて来たランダムコーディング法の代わりに相関コーディング法を新たに導入した．計算機実験で得られた意味マップはまず実際に用いた学習データを用いて検討した．その結果，意味マップ上の名詞は全般的に学習データが示唆する意

☆ パッケージソフト SPSS for windows の因子分析プログラムを用いた．

☆☆ 因子分析において座標軸への種々の回転操作を行っても累積寄与率をあげることができなかった．

味で配置されていることが分かった．そして，本手法の有効性を見るために，同じデータを用いた階層型クラスタリング手法による意味分類の計算機実験も行った．その結果，意味マップは可視性や連続性を有すだけでなく，分類能力においてもクラスタリング手法よりやや優れていることが分かった．更に，可視化能力を有す多変量解析手法が本タスクにうまく適用できないことを主成分の寄与率分析及び計算機実験を通じて明らかにし，提案手法の必要性を補強した．このように，意味マップの自動構築において，その規模がまだ小さいが，提案手法の有効性・必要性が本研究によって明らかにされた．

今後の予定としては，まず，提案手法の改善及び拡張を行う．これはSOMのアーキテクチャ，コーディング，及び学習データの構成という三つの側面から考えられる．アーキテクチャにおいては，これまでサイズが固定のSOMを用いていたため，問題に応じた最適なサイズの設定が困難という問題がある．そのため，問題駆動のサイズ可変型SOMを導入する必要がある．学習データのコーディングにおいては，提案した方法に共起語の頻度を使用していないという問題がある．しかし，共起語の頻度情報はその共起語が修飾する名詞にとってどれだけ重要かを測る尺度と考えることができる．従って，共起語の頻度情報を重みとしてどう相関コーディング法に組み込むかを考える必要がある．学習データの構成においては，これまで，目的型意味マップを構築することとして，名詞を共起する連体修飾要素のみで定義している．汎用型意味マップの構築を目指すためには，動詞などを含むより広範囲な情報（例えば名詞の係り先のすべての単語）の利用も考慮する必要がある．以上のように提案手法の改善や拡張を行った後，目的型と汎用型の両方からより大規模な日本語単語の意味マップの構築を行っていく予定である．

謝辞 多変量解析に関し，助言を頂いた通信総合研究所の白土保主任研究員及び内山将夫研究員に感謝する．

参 考 文 献

1) D. Hindle: Noun classification from predicate-argument structures, *ACL'90*, pp. 268-275, (1990).
2) P. F. Brown, et al.: Class-based n-gram models of natural language, *Computational Linguistics*, 18(4), pp. 467-479 (1992).
3) 森 信介，西村 雅史，伊東 伸泰：クラスに基づく言語モデルのための単語クラスタリング，情報処理学会論文誌，38(11)，pp. 2200-2208 (1997)．
4) 国立研究所：分類語彙表，大日本図書 (1964)．
5) T. Kohonen: Self-organizing maps, Springer, 2nd Edition, 1997.
6) 清田 陽司，黒橋 禎夫，中村 順一，長尾 真：構文情報を用いた電子ニュース記事のクラスタリングシステムの作成と評価，信学技報，NLC98-17, pp. 15-22 (1998-07)．
7) K. Kanzaki, Q. Ma, and H. Isahara: Similarities and differences among semantic behaviors of Japanese adnominal constituents, *ANLP/NAACL'2000 Workshop: Syntactic and Semantic Compexity in Natural Language Processing Systems*, Seattle, Washington, pp. 59-68 (2000).
8) H. Ritter and T. Kohonen: Self-organizing semantic maps, *Biological Cybernetics*, 61, pp. 241-254 (1989).
9) Q. Ma, K. Kanzaiki, M. Murata, K. Uchimoto, H. Isahara: Self-Organization Semantic Maps of Japanese Nouns in Terms of Adnominal Constituents, *IJCNN'2000*, Como, Italy, Vol. VI, pp. 91-96 (2000).
10) 田中 豊，脇本 和昌：多変量統計解析法，現代数学社 (1983)．

(平成13年1月24日受付)
(平成13年8月 3 日採録)

馬　青

1983年北京航空航天大学自動制御学部卒業．1987年筑波大学大学院理工学研究科修士課程修了．1990年同大学院工学研究科博士課程修了．工学博士．1990～93年株式会社小野測器勤務．1993年郵政省通信総合研究所入所．現在独立行政法人通信総合研究所主任研究員．人工神経回路網，知識表現，自然言語処理の研究に従事．日本神経回路学会，言語処理学会，電子情報通信学会，各会員．

神崎　享子

1998年早稲田大学院文学研究科博士課程修了．同年郵政省通信総合研究所特別研究員．現在独立行政法人通信総合研究所自然言語グループ専攻研究員．言語学，自然言語処理の研究に従事．言語処理学会，計量国語学会，日本言語学会，国語学会，各会員．

村田　真樹（正会員）

1993年京都大学工学部卒業．1995年同大学院工学研究科修士課程修了．1997年同大学院工学研究科博士課程修了．博士（工学）．同年京都大学にて日本学術振興会リサーチ・アソシエイト．1998年郵政省通信総合研究所入所．現在独立行政法人通信総合研究所研究員．自然言語処理，情報検索，機械翻訳の研究に従事．情報処理学会，人工知能学会，言語処理学会，電子情報通信学会，ACL，各会員．

内元　清貴（正会員）

1994年京都大学工学部卒業．1996年同大学院修士課程修了．同年郵政省通信総合研究所入所．現在独立行政法人通信総合研究所研究員．自然言語処理の研究に従事．言語処理学会，情報処理学会，ACL，各会員．

井佐原　均（正会員）

1978年京都大学工学部卒業．1980年同大学院工学研究科修士課程修了．博士（工学）．同年通商産業省電子技術総合研究所入所．1995年郵政省通信総合研究所入所．現在独立行政法人通信総合研究所自然言語グループリーダ．自然言語処理，機械翻訳の研究に従事．情報処理学会，言語処理学会，人工知能学会，日本認知科学会，各会員．