

意味論とクラスター分析

Semantics and Analysis of Cluster

平澤 洋一*
Yoichi Hirasawa

1. カードデータの処理

語彙研究にカードはつきものであった。文献語彙を扱うにしろ、フィールドを定めて臨地調査するにしろ、素データの並べ換えのために、資料のカード化は当然の作業であった。語彙索引の作成となると、嫌気のさすような作業量を要求された。中でも圧巻だったのは、編集委員の一人として國學院大学日本文化研究所で作業を進めてきた『現代日本語方言大辞典』(8巻, 明治書院)の最終稿(調査カードをもとに400字詰め原稿用紙に清書)約25万枚をすべて手作業で五十音項目順かつ地点順に並べ直すという作業であった。

情報機器はここ2, 3年のうちに、驚くほどのスピードで言語学の世界にも入りこんできた。意味論の世界でも情報処理がらみの研究が増えてきた。意味論は、(a)語史、(b)語彙体系ないしは(c)方言語彙の研究が多く、(a)はデータベース、(c)では情報機器を用いた言語地図の作成やネットワーク法など、いくつかの新しい研究法が報告されてきたが、(b)の研究分野はまだまだ未開拓である。(b)では昭和50年代前半あたりから意味論の視点にたった類義語の研究が盛んになり、精力的な研究の成果が次々に公刊されていった感が強い。文献語彙や方言語彙の部分体系を明らかにしようとすることに研究者たちの関心が集まり、理論的な発展もあって、55年前後には体系化のためのきわめて有益な方法の提示もなされた。しかし、主に認知論や言語地理学をベースにした人体語彙・服飾語彙・食物語彙などについてのモデルの提示が行われたにすぎなかった。

ネックは、成分分析表から樹形図に移行する作業過程で要求される効率的な体系化理論と計算処理方法の欠如にあった。その結果、語彙体系の解明を目指した論文のほとんどが、語形群と意味特徴との関係を示すマトリックスの提示をもって筆を置かざるをえなかった。語彙の体系化は、精密な意味特徴の設定と意味の遠近の位置づけとを必要とするので、大量の計算処理を避けて通れない。したがって、それだけの潜在需要が研究者の側にあったところへ、個人用の情報機器が出回っ

* 城西大学女子短期大学部

たのであるから、意味の研究が新しい方向へ向かう素地が一気に整ったというわけである。

マトリックスから脱した研究者たちは、いろいろな工夫をこらして語彙体系表示のための樹形図を提案した。図1, 図2はその一例である。これは東京と多良間という2地点の遊戯語彙について心理的意味レベルでの部分体系を示す樹形図であり、枝分かれと弁別の特徴(=それぞれの語彙もしくは語を区別する意味特徴)の違いを明示できる利点があるものの、同一地点の語どうしまたは他の地点の語との意味的遠近関係などを表示できないという欠点をもつ。そこで本稿では意味の遠近関係を考慮した樹形図化の可能性について考察してみたい。部分体系として「移動動詞群」の一部を扱う。具体的にはアガル、ノボル、ウク、ウカブ、サガル、クダル、オリル、オチル、シズム、イク、クル、トオルの12語である。この種の調査は(a)研究者自身の言語を内省で記述するか、(b)何らかの調査表で面接調査するか、(c)文献調査するかである。ここでは(a)をとる。

あがる／のぼる

- S1 女は駅の階段をかけ(あがっていった／のぼっていった)。
- S2 歩いて富士山に(あがる／のぼる)。
- S3 2時間かけて船で川を(あがる／のぼる)。
- S4 さっと五、六人の手が(あがった／のぼった)。
- S5 50階建て高層ビルの屋上まで歩いて(あがる／のぼる)。
- S6 50階建てビルの屋上まで直通エレベーターで一気に(あがる／のぼる)。
- S7 木の葉が風に舞い(あがる／のぼる)。
- S8 子供の手を離れた風船が空に(あがっていく／のぼっていく)。
- S9 去年の同じ時期に比べ物価は8%も(あがった／のぼった)。
- S10 風呂から(あがる／のぼる)。
- S11 学校に(あがる／のぼる)。
- S12 雨が(あがる／のぼる)。
- S13 経営不振が噂に(あがった／のぼった)。
- S14 牛肉の刺身が食膳に(あがる／のぼる)。
- S15 能率が(あがる／のぼる)。
- S16 犯人が(あがる／のぼる)。

移動動詞群は語彙論研究の早い時期から多くの研究家に取り上げられてきた。S1の「階段」やS2の「富士山」は移動の場所を表す。アガルもノボルも出発点から到着点にいたる主体の移動を表現する動詞であるから、出発点および到着点の格をとれるが、無条件でというわけにはいかない。S3でアガルが許容されないのは、到達までに時間のかかる主体移動だからである。「走って」「さっと」のような短時間かつ軽快な移動が顕著なS1やS4ではアガルが使いやすい。ノボルではそれが逆になる。したがって「屋上に～」のような表現要素の単純な文ではアガルもノボルも挿

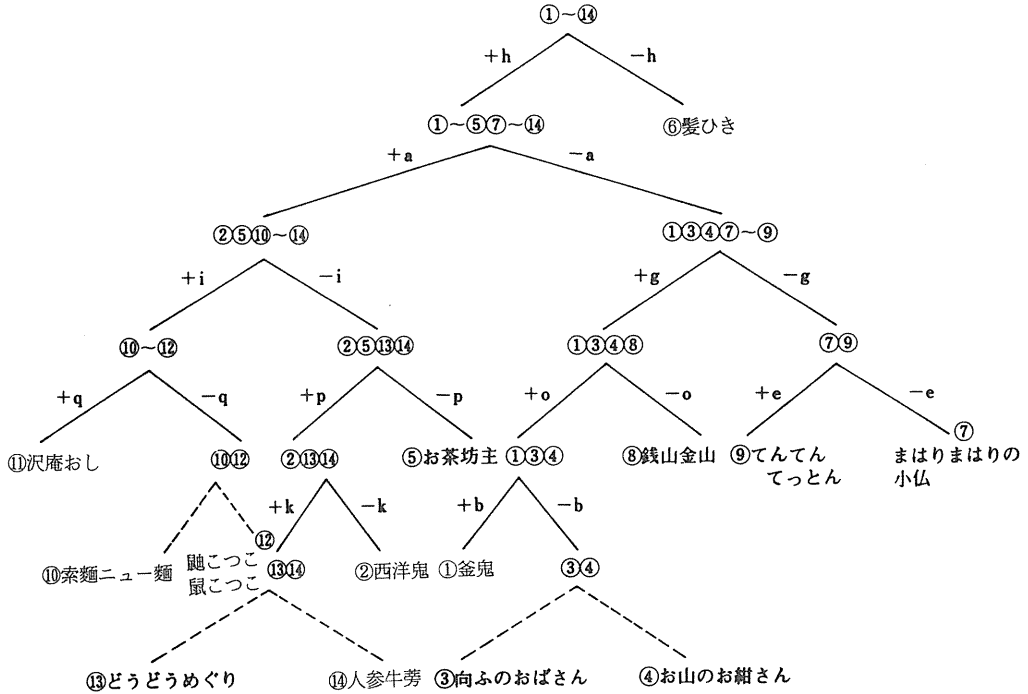


図1 心理的体系 (東京)

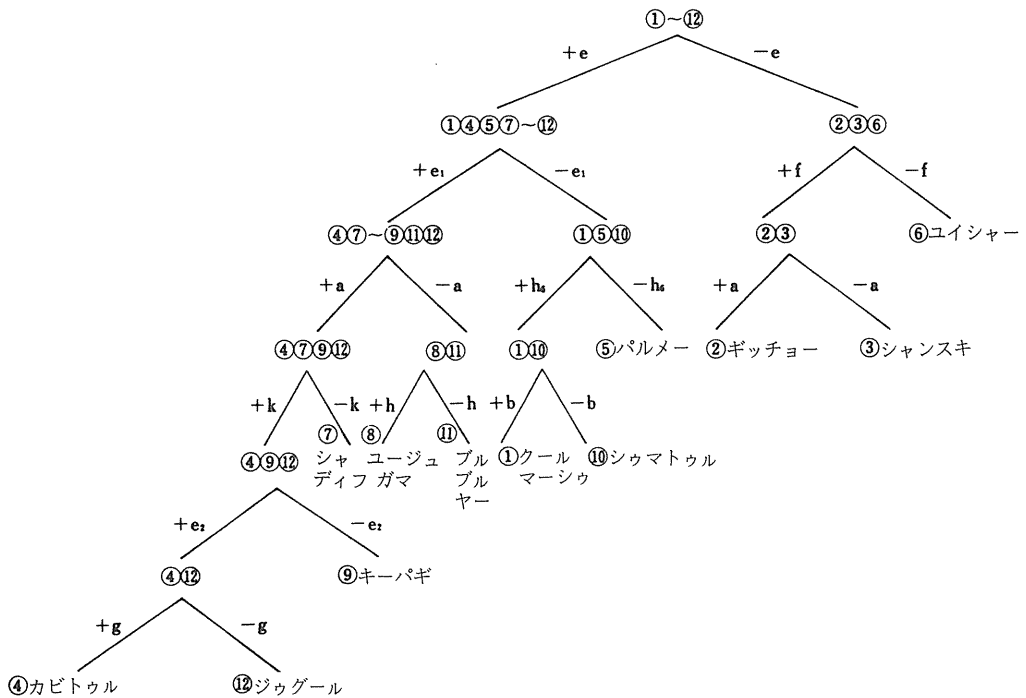


図2 心理的体系 (多良間)

入可能であるが「高層ビルの屋上まで～」となると、アガルが使いにくくなる。が、この文に「高速エレベーターで一気に」「直通エレベーターで」といった短時間移動の表現が加わると、アガルが使いやすくなる。移動の仕方の有する意味情報量がアガルのもつ〈主体が出発点から到着点まで移動する〉〈移動時間が短い〉という意味特徴の束が表現する意味情報量に近づくからである。

アガルが重点的に表現しようとするのはあくまで移動の時間的短さ・軽快さであって、移動の距離の多少ではないから、「ジェット機で一気に1万メートルの上空までアガル。」は許容されることになる。S3では多くの時間的経過をとまなう移動経路としての経路格「川」が選択されているので、アガルは使えない。

このほかの面でもアガルとノボルでは差がありそうだ。「遮断機がアガル。」「風船が空に舞いアガル。」とはいっても「遮断機がノボル。」「風船が空に舞いノボル。」とはいえない。これらを説明するのに、移動するのは遮断機の一部としてのバーの部分であり〔-自動〕素性の風船であるから、ノボルでは意味特徴〈主体が〔+自動〕の全体的な移動に限定される〉が関与的、アガルではそれが非関与的になるといえばうまくいきそうにみえるが、「鳥が舞いノボル。」が許容されないから、有意な意味特徴にはならない。意味特徴〈移動時間が短い〉の範囲内で説明がつく。

S9 昨年と同じ時期に比べ物価は8%もアガった。

S10 風呂からアガル。

S11 学校にアガル。

S12 雨がアガル

これらは一般に派生用法と呼ばれ用例であり、S1～S7に比べると意味の抽象度が高くなる。その結果、「(物価が) 上昇する」「(風呂から) 出る」「(学校に) 入る/入学する」「(雨が) 晴れる」のように近似表現が出てくる。S10では出発点が焦点化されている。それ以外の文では到着点が焦点化されている。このような意味構造ではノボルは使えない。時間的経過に関心が寄せられた伝達表現になっていないからである。「火の手がアガル。」「経営不振が噂にノボった。」「牛肉の刺身が食膳にノボった。」「能率がアガル。」のような用例にも、この基準は適用できよう。

派生用法の意味特徴について最低限の整理はここでやっておかなければならない。たとえば「効果アガル。」「人気アガル。」「犯人アガル。」のアガルは「出る」「高まる」「捕まる」などの意味にとれるので、意味特徴〈到着点〉が焦点化され強調され、その結果、多義化が生まれるとみておきたい。

アガル・ノボルの意味特徴と語形に関するこれまでの議論をまとめてみよう(意味特徴の中に〈経路〉を入れるのは、それが焦点化されるトオルという動詞が移動動詞群の中に存在して体系を形成するからであり、トオルを含めた移動動詞群の意義素は後述する)。意味特徴と語形との関係は次のように示すことができる。

〈主体が出発点から経路点・経路を通過して到着点に移動する〉→アガル, ノボル

- 〈上方への移動〉→アガル, ノボル
 〈出発点〉〈焦点化〉→アガル
 〈到着点〉〈焦点化〉→アガル
 〈経由点・経路〉〈焦点化〉→ノボル

2. 意味と体系

アガル・ノボルとは逆の方向に主体が移動することはオリル・サガル・クダルの担う意味領域となる。意味特徴としての〈下方への移動〉が加わる。

- S17 二階からオリル。
 S18 長い石段をオリル。
 S19 父は一階にオリていった。
 S20 物価が先月よりサガった。
 S21 畑に霜がオリている。
 S22 軒から大きなツララがサガっていた。
 S23 川をクダル。

アガル・ノボルもオリル・サガル・クダルも、〈主体が出発点から経由点・経路を通して到着点まで移動する〉という点では共通の意味特徴をもつ移動動詞ではあるが、移動状態についての意味特徴の焦点化が異なる。

シズム・イク・クル・トオルの4語についてもアガル・ノボルと同様な意味分析を加え、移動動詞12語についてまとめると、次のようになる（二重ヤマガッコは焦点化された意味特徴を表す便宜的な記号とし、必須成分に対する選択成分とする）。

A 意味特徴

- 〈1〉=〈主体が出発点から経由点・経路を通して到着点に移動する〉
 《2》=〈出発点〉〈焦点化〉
 《3》=〈経由点・経路〉〈焦点化〉
 《4》=〈到着点〉〈焦点化〉
 〈5〉=〈方向性をもつ〉
 〈6〉=〈上下への方向性をもつ〉
 〈7〉=〈上方へ〉
 〈8〉=〈下方へ〉
 〈9〉=〈表現者・話し手への方向性をもつ〉
 〈10〉=〈重力を上回る浮力を受ける〉

<11>=<重力による直線的で急速な移動>

B 意味構造

- ① アガル=<1>《2》《4》<5><6><7>
- ② ノボル=<1>《3》<5><6><7>
- ③ ウク=<1>《2》《3》《4》<5><6><7><10>
- ④ ウカブ=<1>《4》<5><6><7><10>
- ⑤ サガル=<1>《2》《4》<5><6><8>
- ⑥ クダル=<1>《3》<5><6><8>
- ⑦ オリル=<1>《2》《3》《4》<5><6><8>
- ⑧ オチル=<1>《2》《4》<5><6><8><11>
- ⑨ シズム=<1>《4》<5><6><8>
- ⑩ イク=<1>《2》《3》《4》<5><-9>
- ⑪ クル=<1>《2》《3》《4》<5><9>
- ⑫ トオル=<1>《3》

C 構文構造

基本構造=主体格(が)——出発点格(から)——経由格(を)——到着点(に/まで)——移動する

表1 移動動詞群成分分析表

	1	2	3	4	5	6	7	8	9	10	11
①アガル	+	±	-	±	+	+	+	-	-	-	-
②ノボル	+	-	±	-	+	+	+	-	-	-	-
③ウク	+	±	±	±	+	+	+	-	-	+	-
④ウカブ	+	-	-	±	+	+	+	-	-	+	-
⑤サガル	+	±	-	±	+	+	-	+	-	-	-
⑥クダル	+	-	±	-	+	+	-	+	-	-	-
⑦オリル	+	±	±	±	+	+	-	+	-	-	-
⑧オチル	+	±	-	±	+	+	-	+	-	-	+
⑨シズム	+	-	-	±	+	+	-	+	-	-	-
⑩イク	+	±	±	±	+	-	-	-	-	-	-
⑪クル	+	±	±	±	+	-	-	-	+	-	-
⑫トオル	+	-	±	-	-	-	-	-	-	-	-

注1：必須成分(=< >)で示した意味特徴)を+で、選択成分(=《 》で示した焦点化の意味特徴)を±で、その他の成分を-記号で表す。

注2：変数1~11は前頁の意味特徴の番号を表す。

ウカブ・シズムは「水にウカブ。」「海にシズム。」のように、いかにも〈出発点〉〈水面・海面〉〈焦点化〉ないしは〈到着点〉〈水面・海面〉〈焦点化〉を有する動詞に見えるが、「空に雲がウカブ。」「捜査線上に犯人像がウカブ。」などの用例があるから〈水面・海面〉に限定されないことは明かである。これらも〈出発点〉〈焦点化〉ないしは〈到着点〉〈焦点化〉で説明がつく。ウクは「水底から泡がウキだした。」「水中にゴミがウイている。」「水面に魚がウイている。」などといえるから、《2》をとることも《3》をとることも《4》をとることもできる動詞と思われる。「東京から電車でイク。」「裏山の道をイク。」「アメリカにいったことがある。」「大阪からきた。」「道なき道を5時間もかけてクル。」などといえるイク・クルも同様である。

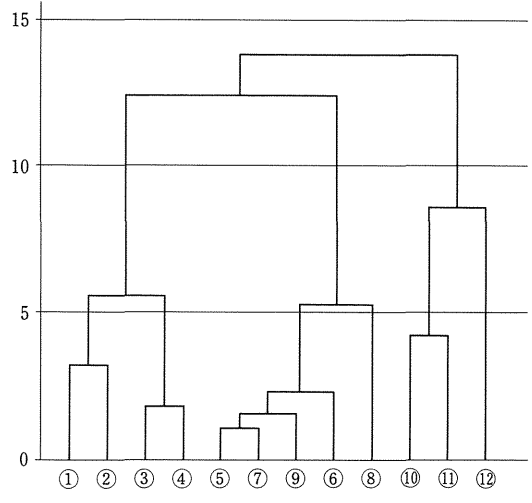


図3 移動動詞群デンドログラム

12語の移動動詞の意味的類似と対立のおおよその様子は表1から読みとることができるものの、具体的にどのような体系になるのかは分からない。また、語彙は常に同じ意味間隔・類似性で互いに接し合っているわけではないから、樹形図はマトリックスに込められた意味の類似と対立の強弱を正確に反映しているのが理想と思われる。

表2 一致係数による類似度

ファイル名 = Idou-dousi41 (一致係数)
 データの数 = 12
 変数の数 = 11

	(1)	(2)	(3)	(4)	(5)	(6)
(1)	1.000					
(2)	0.727	1.000				
(3)	0.818	0.727	1.000			
(4)	0.818	0.727	0.818	1.000		
(5)	0.818	0.545	0.636	0.636	1.000	
(6)	0.545	0.818	0.545	0.545	0.727	1.000
(7)	0.727	0.636	0.727	0.545	0.909	0.818
(8)	0.727	0.455	0.545	0.545	0.909	0.636
(9)	0.727	0.636	0.545	0.727	0.909	0.818
(10)	0.727	0.636	0.727	0.545	0.727	0.636
(11)	0.636	0.545	0.636	0.455	0.636	0.545
(12)	0.455	0.727	0.455	0.455	0.455	0.727
	(7)	(8)	(9)	(10)	(11)	(12)
(7)	1.000					
(8)	0.818	1.000				
(9)	0.818	0.818	1.000			
(10)	0.818	0.636	0.636	1.000		
(11)	0.727	0.545	0.545	0.909	1.000	
(12)	0.545	0.364	0.545	0.727	0.636	1.000

3. 記号 ± の扱い

このことをもう少しはっきりさせるため、語形どうしの類似度を考慮してクラスター化する方法をとってみよう。表1の+記号と±記号に別々の意味をもたせるため、(1)+記号に2を、±に1を、-記号を0に与えて量的データのマトリックスを作り、(2)群平均法によるクラスター分析を行って、(3)デンドログラムで表示すると、図3のようになる。この図では、最初に①②③④⑤⑥⑦⑧⑨/⑩⑪⑫という分節がなされ、つづいて①②③④/⑤⑥⑦⑧⑨という分節になっていて、弁別の特徴も明確である。もちろん樹形図は類似度を反映した形状になっている。

表1の+記号に1を±および-記号に0を与えて質的データにしたマトリックスをもとに、一致係数による①～⑫相互の類似度を求めたのが表2である。この表2から、①～⑫はそれぞれ相互の類似度が高いので一つのクラスターにりそうなのが読み取れるが、群平均法でクラスター分析した最初の分節型が①/②～⑫になってしまう。この型の分節を説明できる弁別の特徴がマトリックスの中に見出せない。

これとは別に表1の+記号および±記号に1を-記号に0を与え、数量化理論Ⅲ類による質的データの類分け→類似性の処理→クラスター分析という手順での分析方法をとると、分節される型(たとえば移動動詞群①～⑫が最初に分節される型は群平均法では①③④⑤⑦⑧⑨⑩⑪/②⑥⑫、ワード法では①③④/⑤⑦⑧⑨/②⑥⑩⑪⑫となる)の有意味性を説明できる弁別の特徴が発見できない。最短距離法や重平均法による分析についても、結果は同様であった。この事実、マトリックスから体系化の作業に映る場合に、±記号の存在価値を無視してはならないことを意味する。

最後に、言語が「恣意的」であるために起きる問題点について。既出のS1ではノボルを選択する被験者(日本人も外国人も)が混在するのに、S2ではほぼ全員が迷うことなく「のぼる」を選択するのはなぜか。S2の文を見たとき、歩いて富士山に登る行為が必然的に行為者の時間と汗と忍耐とを伴う行為であることを被験者は容易に表象化し認知することができるのに対し、S1では行動が抽象的にしか表現されていないので、つまり行為者としての女がどのくらいのスピードで階段を移動していくのか、手に重い荷物をいくつも持っているのか手ぶらなのか、女は若いのか年輩者なのか、といった一切の具体情報の提示がなされていないので、駅の階段を移動していく女の表象を描くのにためらいを伴う若干の時間がかかるのであろう。

S3の場合でも、S1ほどには顕著ではないものの、語彙選択のパラツキが観察された。「2時間かけて」「船で」という制限があるので簡単にノボルが選択されるところであるが、これらを焦点化させずにS3という情報を受け止め表象させた被験者は、無意識のうちに「船」を「舟」に置換させ、船頭が竿を操る舟が静かに滑るように川を上がっていく、などと捉えたようである。

