

論文

第二言語獲得における助詞「に」の習得過程の定量的分析

李 在鎬 ((独) 情報通信研究機構)

井佐原 均 ((独) 情報通信研究機構)

ディスクリプタ：助詞「に」、文法構文、第二言語習得過程

1. はじめに（本稿の目的）

本稿の目的は二点ある。一点目に、日本語学習者の言語使用を定量的に捉えることから、助詞「に」の習得に見られる複雑な構造を記述分析する。二点目に文法的現象に対するコーパスの具体的な利用法および方法論を提案する。主として OPI テストの発話データベースを利用し、日本語学習者の「に」の使用例を文脈的特徴からクラスタリングする。そこから「に」の運用の背後に存在する特性を明らかにする。

考察に際しては、まず「KY コーパス」(鎌田・山内 1999) における韓国語母語話者の発話データから助詞「に」の使用例を収集した。次に「に」の生起文脈を 18 変数でコード化し、クラスタ分析した。その結果、次の三点が明らかとなった。1) 初級、中級、上級学習者のいずれにおいても場所用法 (e.g., 大学に入る) がもっとも安定して観察されること。2) レベルが上昇するにつれ、対人的用法 (e.g., 先生にあげる) が増加し、対象用法 (e.g., 自転車に乗る) が減少すること。3) 個々のクラスタは (单一要素の制約によるものではなく) 要素間の複雑な相互作用の結果、形成されるものであることを示す。なお、主張 1) や主張 2) の論証としてレベル間のクラスタのクロス集計でそれぞれの分布を示す。主張 3) の論証として共起語の制約によって形成されたクラスタが存在する一方、統語的文脈や意味的文脈の制約によって形成されるクラスタが存在することを示し、クラスタリングにおける制約の非均質的側面を示す。最後にこれらの事実関係が第二言語習得の理論化に何を示唆するのかを簡単に考察する。

2. 記述的問題

ここでは、本稿のもっとも基本的な問題意識を明らかにすべく、一見当然のことにも思われ

Quantitative Analysis of Japanese particle Ni in Second Language Acquisition
LEE Jae Ho (National Institute of Information and Communications Technology)
ISAHARA Hitoshi (National Institute of Information and Communications Technology)

る次の問題から考察を始める。本稿では、助詞「に」の習得過程をめぐる作業仮説の一つとして次のような予測をする。それは（日本語学習者の言語使用において）助詞「に」の用い方は、「習得の過程で多様な変化を見せ、漸進的にその習得がなされていく」というものである。しかし、この予測の妥当性を十分な証拠でもって実証的に示すのは、決して容易なことではない。というのも、ここには二つの厄介な問題が関係しているからである。第一に、習得の過程で多様な変化を見せるという予測に関連し、「学習による変化の実態を記述する枠組み」が充分に確立されていない。第二の問題点として、助詞「に」の用い方の問題に関連し、「に」そのものの（記述レベルの）位置づけがそれほど明確ではないという問題がある。以下では、この二点の問題における背景を概略的に示すと同時に本稿の立場を明らかにする。

まず、第一の問題に関連し、次の点を確認しておきたい。学習者の言語とされ、学習の過程で生じる中間言語（interlanguage; Gass&Selinker 1994）にはターゲット言語、すなわち日本語の体系にもなければ母語にもない、独自の一般化や体系が観察されると言われている（cf. 森山 2000）。その一例として、場所表現と助詞の問題に関して言えば、地名や建物を示す名詞に対しては「で」を用いる傾向が強い反面、「前、後ろ」といった相対的な空間関係を表す名詞には「に」を用いる傾向が強いとされる。その証拠として、(1) の誤用例が報告されている（野田（編著）2001:30-31）。

- (1) a. 横浜で遊びに行きました。
b. 東京で住んでいます
- (2) a. タバコ屋の前に会うように行ってください。
b. 門の前に話しをしました。

(1)、(2) では「に」と「で」の場所用法における使い分けの誤用が観察される。(1)においては「に」で表現すべきものが「で」になっており、(2)においては反対に「で」で表現すべきものが「に」によって表現されている。これには、行為がなされる場所を表す「で」と移動先や存在地としての場所を表す「に」の用法をめぐる誤用として位置付けることができる。この背景には上記の野田（編著）（2001）が示した理由に加え、文法構造の決定が共起名詞の意味的特徴のみならず、文脈全体の相互作用によるものであることを十分に認識していないことから生じたものと理解することができる。こうした事実関係において重要なのは、この種の間違いが様々な背景の学習者に共通して見られる点であり、単純に母語干渉による問題としては片付けられない、ということである。このことから多くの研究者が共有する見方として誤用には独自の動機付けが存在するというものがある。この種の見方が定着できたのは寺村（1990）

による集大成と、その延長でなされた福間（1997）、市川（1997）、迫田（2001）、森田（2005）の忠実な観察と記述の成果であるといえよう。

一方、誤用分析を中心とする従来の中間言語アプローチは次の問題点を抱えている。それは、誤用例に基盤を置く現象分析では学習者の誤用例から、なぜ誤用を産出するのか、または、なぜ間違うのかを明らかにすることができるても、どのようにして正しく習得できるのか、という言語習得の本質的問題が必ずしも説明できるとは限らない（李 2004b）。こうした問題点を受け、本稿では正用例に見られる習得の過程を忠実に捉える記述モデルが必要であると判断し、その組織化における変化の様子を多変量解析で探索的に分析する。

第二の問題点として、助詞「に」の用い方の問題に関連し、「に」そのものの（記述レベルの）位置づけがそれほど明確ではないという問題がある。この問題点に関連し、以下では助詞「に」の実際の使用例をもとに、その多様な振舞いの一端を考えてみよう。

- (3) a. 太郎君が部屋に入った。
b. 新学期は4月に始まる。
c. 太郎が壁に釘を打つ。
d. 私は今月の家賃を大家さんに渡した。

助詞「に」は、(3) a のように典型的な移動先や方向や存在点を含む場所用法、b のような時間や時期を表す用法、c のように行為の対象を表す用法、さらには、d のように動作の相手を表す用法など、実に多様な使用文脈を持つ。多くの先行研究はこうした使用に対してどれだけの用法を認め、何の制約として記述すべきか、という点をめぐって様々な議論を展開してきた（cf. 寺村 1982、村木 1991、石綿 1998、森田 2002）。紙幅の都合上、先行研究を一つ一つ検討することはできないが、先行研究の多くに共通する二点の問題点を指摘しておく。一点目に、分析結果の多くが一分析者の内省に依拠しており、実証性に欠ける。実際問題として各々の分析結果を支えるべき、観察データの規模や分類基準などの「発見の手続き」が明らかにされておらず、反証不可能な分析となっている。二点目に、多くの先行研究が「に」に見られる様々な使用を個別現象として記述する方向へ議論が傾いていた故に、相互の関係性については必ずしも充分なビジョンを示せなかつたという問題点が挙げられる。こうしたことから場当たり的印象を払拭しきれず、充分な体系化はいまだなされていないのが現状と言えよう。

以上の考察が示す問題点は、助詞「に」の習得に限ったものではなく、文法現象の習得全般に渡る一般的問題である。本稿では、こうした問題に対してデータマイニングの手法として近年注目されているクラスタ分析をベースとした定量的分析モデルを提案する。これによって、

第一の問題点に対しては組織化の潜在的性質が発見でき、第二の問題点に対しては全体の相互関係が明らかにできる。

3. データと方法論

本稿では、前節で明らかにした先行研究の問題点を踏まえ、李（2004a）¹の枠組みに基づいて、発見の手順を明確に示した分析モデルを提案し、実証的に考察することを試みる。そのモデルの特徴として、次の3点をあげることができる。

- 1) コーパスを利用し、KWIC検索でデータを収集する。
- 2) 生起文脈に基づいてデータをコード化し、多変量解析の手法で定量的に分析する。
- 3) 分散意味論²の考えを導入し、制約を分散して記述する。

こうした方法論が本稿の議論にもたらす利点として、1) によってサンプル収集の一貫性が確保できる。2) によって言語的直感をうまくコントロールしながら、データ間の隠れた性質を捉えることができる。3) によって解析結果を自然に反映し、記述することが可能である。以下の3.1節では使用データの詳細を述べ、3.2節ではコード化の詳細を述べる。そして、3.3節では解析手法について述べる。

3.1. 使用データと分析モデル

本稿では一次調査資料として次のコーパスを利用した。それは、OPIテストによる日本語学習者の発話データベースとして広く知られている「KYコーパス」（鎌田・山内 1999）で、そこからKWIC検索でサンプルを収集した。なお、サンプル抽出に際しては、テスターの発話部をすべて除去することで、学習者の発話のみでコーパス全体を再構成した。また、分析対象の性質上、サンプリングの際には以下の二点の方針に基づいて行った。

方針1) 第二言語習得の発話データであることを考慮し、母語のバイアスを極力排除するため、単一の母語話者のみのデータを利用する。

方針2) 正用の組織化に対する一般化であるため、明らかな誤用例については除去する。こうした事情により韓国語母語話者のデータを利用し、サンプル抽出を行った。なお、韓国語母語話者全体のテキスト量を示す。

¹ 本稿のもとになっている李（2004a）では、新潮文庫内の8人の書き手による総計8000字のミニコーパスを作成し、そこから「に」を含む203のサンプルを収集した。そして、クラスタ分析を行った結果、5つのクラスタで最適な分離を得た。

² この考え方は従来の還元論的語彙主義に対するアンチテーゼとして開発されたものであり、次の考え方方に立つ。1) 言語の意味は特定の語彙に還元できない。2) それは複数の語に分散して表現される。3) 意味を複数の語の相互作用によって生起する現象として定式化する。この考え方をめぐる詳細は黒田・中本・野澤（2004）を参照されたい。

表1 テキスト量

区分	学習者数	一人当たりの平均文字数	標準偏差	最大	最小
初級	5	1199.8	626.58	1927	349
中級	10	3477.6	967.08	5631	2118
上級	10	4981.5	872.98	7657	3776

表1から直感的に伺える事実として、初級学習者においてはいずれの値においても発話量として極めて少ないことが分かる。一方、中級においては初級の二倍を超える発話量であることや標準偏差が高いことから、一見するとばらつきが激しくなっているようにも思われる。しかし、平均文字数の割合から考えると、(中級より) 初級のほうがばらついていると見るべきであろう。また、表1の量には反映されない特徴として初級学習者の発話の多くに関して、言い直しや一語文が多いことや極端な助詞の省略といった様々なバイアスがあり、(正用としての) 統語形式を基盤とする発話はさらに限定されてしまう。よって単純に表1の数値のみからは予測できない複雑さが内在する。こうしたデータそのものが含む様々なバイアスから有意義な結果を導き出すため、図1の分析モデルに従って3段階の手順で調査および分析を行った。

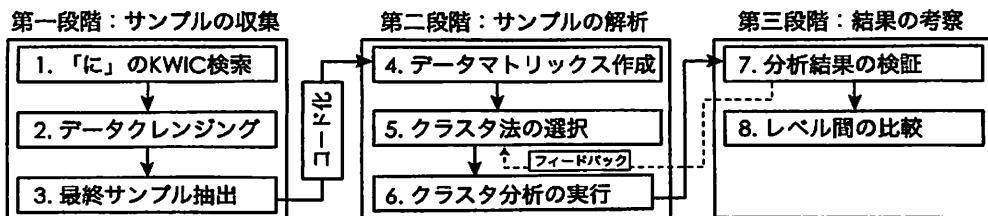


図2 分析モデル

まず、1. 「に」の KWIC 検索においては、Kwic Finder(Win 版 Ver3.01)を使用し、上述の一次調査対象に対して「に」を含むすべての文字列を収集した。そのトーカン頻度を以下に示す。

表2 初期データ量

区分	度数	一人当たりの平均値	標準偏差	最大(実数)	最小(実数)
初級	48	9.6	6.35	17	4
中級	479	47.9	22.66	99	20
上級	647	64.7	28.04	133	28

次に、2のデータクレンジングにおいてはノイズの除去、すなわち、本稿の調査対象ではない文字列を除去した。とりわけ、語中の「に」(e.g., こにちは) や誤用例 (e.g., 電話の中に言います) を除去した。これらの作業を行った結果、以下のサンプルが抽出できた。

表3 最終サンプル

区分	度数	一人当たりの平均値	標準偏差	最大(実数)	最小(実数)
初級	22	4.4	4.16	11	1
中級	295	29.5	10.01	51	15
上級	378	37.8	11.61	63	26

表3に示したデータ量が図1の3.最終サンプル抽出の結果となる。レベル間のデータ量に関してまず気がつくこととして、中・上級学習者のサンプル量に比べ、初級学習者のサンプル量が極めて少ない点である。このことは、本論の解析結果の信頼性にとって重大な問題になりかねないが、あえてこのようなデータを採用した背景には次の事情がある。本稿では、上記のサンプル抽出における方針1で示したように、第二言語獲得という特殊なコンテキストを考慮し、論者による人為的データの追加や削除は行っていない。なぜなら、そうした人為的操作は実際の解析においてはコントロールできない不確実要素を含むことになるので、さらに大きな問題を引き起こすと判断したからである。こうした事情により、現段階では表3のデータを用いることが妥当と考えられる。

3. 2. データのコード化について

第二段階への移行において、最初に行った作業は、各々のサンプルを計算機が認識できる表示に変換することである。この作業は、本稿の方法論の中核に当たる部分でもあるので、詳細に述べておく必要があろう。

まず、4のデータマトリックスを作成するに当たっては生起文脈の様々な情報を18の変数で(1/0の表現として)コード化した。まず、「に」と直接的な共起関係をなす語 (e.g., 太郎に、バスに、部屋に) の意味クラスに関連して6つの変数を用いた。次に個々の「に」が生じている意味的な文脈に関連して4つの変数を用いた。最後に統語関係、とりわけ項構造 (argument structure) に深く関与している形式的特徴に関連して8つの変数を用いた。

表4では、本稿が利用したすべての変数を示した。共起語と意味文脈に関しては具体例を示し、統語文脈に関しては構文スキーマを示すと同時に具体例を示した。なお、1から10の変数をめぐっての判断は主観の混入を防ぐと同時に作業の効率化をはかるため『日本語語彙大系

表4 変数一覧

区分	変数	具体例
共起語	1 主体	{ ガードマン / 彼女 / 社員 / 客 / 田中 / 妻 } に
	2 場所	{ 仙台 / 玄関 / 会社 / 部屋 / ポケット / 壁 } に
	3 具体物	{ 容器 / エレベーター / 電車 / 花瓶 / カメラ } に
	4 事	{ 具体的 / 頻繁 / 一緒 / 本当 / 妙 / 氷づけ } に
	5 抽象物	{ 時間 / 気持ち / 習慣 / 立場 / 問題 / 不安 } に
	6 抽象的関係	{ 最後 / こと / 逆 / 時 / 以外 / 第一 / 日 } に
意味文脈	7 生物主語	{ 山本 / 私 / 彼女 / 兄 / 猫 / 敵 } が太郎に話した。
	8 無生物主語	{ 台風大雨 } が本州に上陸する、{ 月 } が西に傾く
	9 動作述語	行く、こする、借りる、渡す、出かける、飲む
	10 状態述語	慣れる、似る、知る、見える、凍る、切れる
統語文脈	11 文頭	#に... (具体的に話します。)
	12 「が」に後続 (が#)	N1 が # に (みわこが 東京に 帰った)
	13 「を」に後続 (を#)	N1 を # に (みわこを 東京に 帰らせた)
	14 「に」に後続 (に#)	N1 に # に (9時に 東京に 着いた)
	15 「動詞」に先行 (#動詞)	# に 動詞 (東京に 送る)
	16 「が」に先行 (#が)	# に N2 が (東京に 春が 来た)
	17 「を」に先行 (#を)	# に N2 を (東京に みわこを 帰らせた)
	18 「に」に先行 (#に)	# に N2 に (東京に 9時に 着く)

CD-ROM 版』(NTT コミュニケーション科学基礎研究所) に準拠して行った。そして、7と8に関しては(主語の省略で)明示されていなくても文脈から明らかに分かる場合は(著者による)復元を行った。また、12、13、16、17についてもとりたて詞によって表層には明示されていなくても復元によるコード化を行った(e.g., 「太郎は言った」を「太郎が言った」と見なす)。

統語文脈の変数設定に関連し、次の点を補足しておく。「に」の生起文脈を形成する文脈は、「と」や「の」といった格関係、あるいは「は」や「も」といったとりたて詞など、表4のもの以外にもたくさんのディテールが存在する。さらには、埋め込み節や複文といった統語的制約によって実現される構造なども存在する。これらのディテールが存在するにも関わらず、本稿がわずか8変数のみで統語文脈の解析を行った背景として、次の点に注目してほしい。本稿の基本的目的は既述の(3)で見た文法格としての「に」の振舞いを明らかにすることである。こうした理由から文の命題構造に深く関与していると考えられる文法格をもとにコード化を行った。これらの点において上記の変数は記述的要請を満たすものと位置づけることができよ

う。なお、同様の理由から本稿では(4)に見られるすべての事例に対して单一のコード化を行った。

- (4) a. 太郎が部屋に入った。
- b. 太郎は妹の部屋に入った。
- c. 太郎は妹の部屋にはあまり入らない。
- d. 太郎は昨日も今日もあの部屋には入らなかった。

(4)では(4a)の文法格に基づく単純な命題構造に対して、(4b)から(4d)のように様々な副次的要素をつけることができる。しかし、前述の事情から(4b)から(4d)の現象は(4a)と見なし、单一のコード化を行った。(4)を実際にコード化した場合、表5のようになる。

表5 コード化の実例

共起語						意味文脈				統語文脈							
主体	場所	具体物	事	抽象物	抽象的関係	生物主語	無生物主語	動作述語	状態述語	文頭	後続			先行			
											が	を	に	述語	が	を	に
0	1	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0

まず、コード化の対象となる特徴として、「部屋」という共起語が持った「具体物」であり、「場所」であるという意味的特徴、そして、意味文脈として「太郎」が生物であること、「入る」という動詞が動作を表すものであることを示す。最後に統語文脈として「が格」に後続し、述語（動詞）に先行していることを示す。そして、ここで例示したことと同様の作業を695例のすべてのサンプルに対して行った。

3.3. 解析手法について

本節では、図1の5から7の工程に関する解説を行う。まず、クラスタ分析においては階層的クラスタ法と非階層的クラスタ法を併用した。そして、クラスタリングの結果を検証する手法として判別分析を用いた。以下ではクラスタ分析の概要を示すと同時に適用方法の詳細を示す。

一般にクラスタ分析には、与えられた距離情報をもとに個々のケースを併合していくつかのグループにまとめていく「階層的なクラスタ法(cf. Romesburg 2004)」と、似たものが結果的に同じグループに入るよう集合を分割する「非階層的クラスタ法(cf. 宮本 1999)」とに大別して考えることができる。両者のもっとも大きな相違点は前者が(クラスタの数を指定せず)ボトムアップ的にケースとケースを結合していくものであるのに対して、後者はトップダウン

的に予めクラスタ数を決め、階層化せず单一の水準でクラスタを作る手法である。

それでは実際の分析における適用方法を示す。

- 1) 695 例すべてのデータを一つの母集団とし、階層的クラスタ法 (Ward 法) で分析を行う。デンドrogramを利用し、データ全体の類似関係や傾向を把握する。
- 2) 階層的クラスタ法から得た結果をもとに(非階層的クラスタ法の代表的手法である) K-means 法でクラスタリングする。解析においては 4 つから 6 つに(複数回に渡って) それぞれ解析する。
- 3) K-means 法のクラスタリングの精度を判別分析で評価する。正判別率に基づき、最適なクラスタ数を決定する。
- 4) レベル間でクロス集計をし、結果を解釈する。

本稿が階層的クラスタ法のみならず非階層的クラスタ法を使用した背景には、次の理由がある。階層的クラスタ法では、最初、各ケースをばらばらの一つのクラスターとみなして、近いクラスターを次々と統合することによって、最終的な分類結果を得る。この手法は、比較的性質の良い分類結果が得られることが知られている。ただし、分類対象の数が多い場合、分析や解釈上の負荷が大きいという短所があるとも指摘されている(朝野 1996)。それに比べ、非階層的クラスタ法は計算が速く、大きなデータセットに対して(きれいな分類結果を得ることができるので) 解釈がしやすいといった長所がある(SPSS 1999:206)。実際、本稿のように大量のデータを処理する場合には実質科学的な観点からクラスタ数の目星をつけてクラスタ数を変えて非階層的手法を数回実施したほうが効率が良い場合が多い。さらに、本稿の考察目的上、單なる分類だけでは充分とは言えず、個々のクラスタの内部構造を観察する必要があるが、これは階層的クラスタ法では難しい。K-means 法を用いた場合に得られる最終クラスタ中心値は、そのクラスタの特徴を代表的に示しており、本稿の目的に合致する。ただし、非階層的クラスタ法の欠点として一般にクラスタ数をあらかじめ特定する必要があることや、最適なクラスタ数の決定が難しいといった問題点があげられる。こうした問題点を踏まえ、本稿では階層的クラスタ分析との併用や判別分析といった複数の手法を導入する。なお、分析方法同士の関係として、階層的クラスタ法と非階層的クラスタ法が相互補間的関係にあるのに対して、判別分析はクラスタ分析に対する検証法として位置づけられる。

最後に本稿のモデリングの背景としてクラスタ分析の一般的特徴として、次の点を付け加えておく。クラスタ分析(の結果)の良し悪しは多変量解析の範囲だけでは決められないと指摘されることがある(朝野 1996)。というのは、クラスタ分析は因子分析などの他の多変量解析

の手法に比べ、非常に探索的な性質を持つ分析手法である。分析データが持つ様々な特徴を多角度的に探索していくことで分析対象、すなわち言語データに対する理解を深めることができる。こうしたこともあるってクラスタ分析では分析者がどのような条件を与えるかによって結果が大きく変わる。したがって一般に複数の手法を導入し、結果を検証することが必要であり、結果の解釈における分析者の直感が重要な要素として働くことも事実である。この種の柔軟性をマイナス要素と考えるか、プラス要素と考えるかは別として手法そのものが持つ限界が存在することを予め言っておきたい。

4. 結果と考察

4. 1. 階層的クラスタ法による分析結果

まず、第一段階の解析としてすべてのデータに対して階層的クラスタ法で分析を行った。クラスタ法の選択においてもっとも安定した結果を出すと言われている Ward 法を使用し、距離の定義はユークリッド距離を使用した。なおデータの標準化は行っていない。

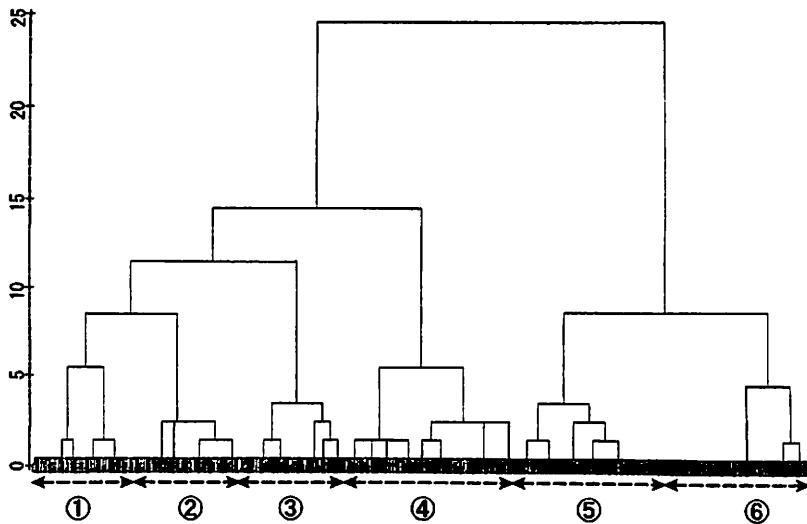


図2 デンドログラム（樹状図）

図2のデンドログラムからデータ全体の傾向を把握することができる。図示の便宜上、横に圧縮しており、末端部の黒ずみに 695 例が位置付けられている。最下段の番号にそってクラスタの代表例を示す。

- ① ちょっと国のその母国語に興味をもち始めて、 . . .

- ② 授業が終わって〈ええ〉自転車に乗って〈ええ〉留学生会館へ戻って
- ③ 先生は外国人に〈ええ〉教える時〈ええ〉一番困ったことは何ですか。
- ④ 気持ちが落ちつかないで不安になったと思うんですけど
- ⑤ 今はあのコンピューター関係の専門学校に通っています
- ⑥ 皆は結婚して他の所に住んでいました。

図2の左から順に見ていった場合、①においては抽象物を表す名詞と共にし、動作や状態が及ぶ対象をマークする用法のクラスタができている。また②では具体物を表す名詞と共にし、主として動作の対象を表す用法のクラスタ、③では主体を表す名詞と共にし、行為の相手や受益者などを表す用法のクラスタ、④では抽象的関係や事を表す名詞と共にし、時間を含め、様態や属性などを表す用法のクラスタ、⑤では場所を表す名詞と共にし、移動先を含む場所を表す用法のクラスタ、⑥では位置関係や指示詞など抽象的関係概念を前提とする名詞と共にし、存在や位置を表す用法のクラスタができている³。

次に（2節の問題提起にもあった）データ間の相互関係の問題をめぐってクラスタの分岐パターンを見た。まず、⑤や⑥の場所用法と①から④の非場所用法とで大きな分岐が見られる。その次に非場所用法の中で④の様態や属性を表す用法とそうでないものとに分岐している。そして、①と②の対象を表す用法と③の対人関係を表す用法とに分岐している。これらの分岐事実から場所用法としての「に」は他の用法に比べ独立性が強いものと考えることができる。

さて、こうしたクラスタのすみわけは共起語のオントロジ的特徴をうまく反映しており、一見納得する。しかし、助詞「に」の用法のクラスタという点からは必ずしも最適なものとは言えない。というのは、記述言語学的視点に基づき、「に」の機能的特徴として捉えなおすと（タイプレベルではなく）トークンレベルの変種として扱うべきものが存在する。具体的には、①と②は共起語が具体物か抽象物かの違いであり、行為の対象という点においては単一のものと扱って良いと考えられる。また、⑤と⑥においても存在や移動といった動詞レベルの特徴は異なるものの、場所用法として一括して捉えて良いと考えられる。この観察をもとに、4つのクラスタで改めてクロス集計を行った。なお、説明の便宜上、クラスタ番号は上記の記述および図2にあわせて振り直した。

³ ⑤と⑥は共起名詞のレベルで、⑤が具体的場所を表す語（関西空港、部屋、岡山、研究室など）と共にしているのに対して⑥は抽象的関係を表す語（後ろ、前、表現、左側など）と共にしていること、一応の相違は観察される。とは言うものの⑥においても「後ろに行く」といった事例が含まれていることや、⑤においても「部屋にある」といった事例が含まれていることを考慮した場合、⑤と⑥はクリアカットに分かれるものではなく、何らかの連続的関係にあるものと捉えるべきであろう。

表6 レベル間クラスタの対応

学習レベル		1 (①、②)	2 (③)	3 (④)	4 (⑤、⑥)	合計
初級	度数 (%)	8(36.4)	2(9.1)	3(13.6)	9(40.9)	22(100)
中級	度数 (%)	91(30.8)	48(16.3)	32(10.8)	124(42.0)	295(100)
上級	度数 (%)	109(28.8)	46(12.2)	76(20.1)	147(38.9)	378(100)
合計	度数 (%)	208(29.8)	96(14.1)	111(16.0)	280(40.1)	695(100)

表6では、学習レベル内の各クラスタの分布を示している。例えば、初級1ではクラスタ1が8例、クラスタ2が2例、クラスタ3が3例、クラスタ4が9例あったことを表している。また、クラスタ番号の下の番号は図2の①から⑥との対応関係を示したものである。つまり、①と②（行為の対象）はクラスタ1、③（相手、受益者）はクラスタ2、④（様態、属性）はクラスタ3、⑤と⑥（場所）はクラスタ4と対応していることを示している。この表6によってまず明らかになったことは、いずれのレベルにおいても、場所用法としての「に」がもっとも生産的であったこと、次に行行為の対象を表す用法が多く見られたことである。様態と相手や受益者を表す用法については明確な順位づけは難しいが、データ全体においては前者の方が若干多いさて、以上の分析によって得られた見通しをもとに非階層的クラスタ法による分析を行った次節ではその分析結果を報告し、言語習得における段階性の問題を考察する。

4. 2. 非階層的クラスタ法による分析結果

前節の階層的クラスタ法では、個々のサンプルのボトムアップ的な結合の様子が示され、全体の分布が把握できた。だが、この分析においては各クラスタの中身については必ずしも明らかではない。こうした問題点を踏まえ、非階層的クラスタ法の代表的手法であるK-means法による解析を行った。

K-means法におけるクラスタ数の指定においては階層的クラスタ法でも示唆されたように4つのクラスタ(K=4)と6つのクラスタ(K=6)の間で複数の可能性が認められた。このことを受け、K=4、K=5、K=6で探索的に分析してみた。そして分割の最適度を評価すべく、判別分析を行った。なお、判別分析の際、所属クラスタを従属変数にし、クラスタ化で用いた表2の変数群を独立変数として投入した。

表7 判別分析結果

区分	判別率	交差妥当化の判別率
K=4	97.7%	97%
K=5	97%	95.8%
K=6	92.5%	89.5%

表7においてはK=4からK=6のクラスタ化に対する判別率を示した。僅差ではあるもののK=4においてもっとも高い判別率を示していることが確認できる。このことから4つのクラスタで、最適な分離が得られたと判断し、K=4でのクラスタ中心とクロス集計をもとに考察を進める。なお、説明の便宜上、階層的クラスタ法の結果と類似関係が認められるものには同じ番号を振って、クラスタ番号を統一した。

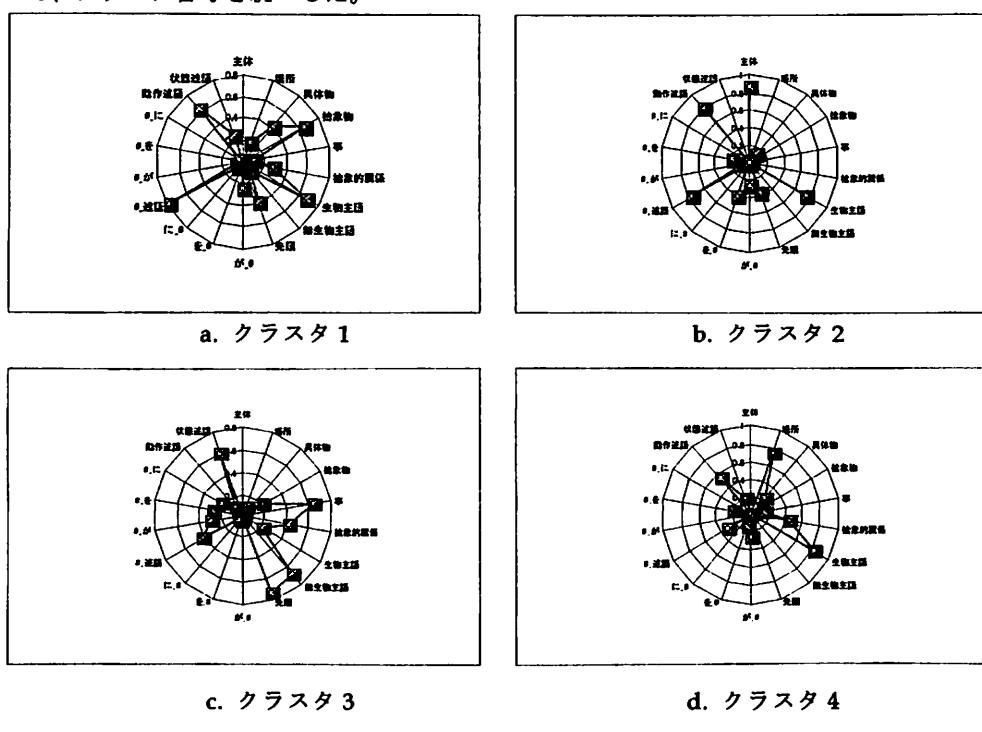


図3 クラスタ中心値

図3ではクラスタ中心値をグラフ化した。そして、これをもとに、個々のクラスタを特徴づけた場合、以下のようにまとめることができる。

表8 クラスタの特徴⁴

区分	主語の特徴	共起語の特徴	述語の特徴	統語的特徴
クラスタ1	生物主語	抽象物/?具体物	動作	動詞が後続する
クラスタ2	生物主語	主体	動作	動詞が後続する
クラスタ3	無生物主語	事/?抽象的関係	状態	文頭
クラスタ4	生物主語	場所	?動作	—

表8ではK-means法によって得られたクラスタの大まかな特徴を記述している。これについていくつか興味深い事実が明らかになった。それはクラスタ化における貢献度の点から各変数を見直した場合、それらは必ずしも均等なものではない、ということが示唆される。例えば、クラスタ4のように述語の特徴や統語的特徴は必ずしも明確ではないが、共起語の特徴として場所を表す名詞であり、意味文脈として主語が生物である、という特徴が大きく関与しているクラスタがある一方、クラスタ1や3の場合は、共起語の意味クラスよりは文頭であることや述語の直後に生じるといった統語的特性が大きく関与しているクラスタもある。さらには、クラスタ2のように主語や述語、そして共起語の意味クラスや統語文脈のいずれにおいてもバランスよく動機付けているクラスタもある。このことは、助詞「に」の動機付けの問題に関して次の見方を示唆する。それは、「に」のすみわけが動詞や名詞といった単一の要素の制約には還元できない、ということである。この点において本稿が意図する分散意味論、すなわち複数の要素に制約を分散し、記述するという手法は、言語の事実に適した手法であると言えよう⁵。

さて、次の課題として学習レベル内のクラスタの数を確認しよう。

図4では各レベルにおけるクラスタの分布を示した。まず確認したいこととして階層的クラスタ法の解析結果と比較した場合、少々の誤差、特に中級におけるクラスタ数のずれは見られるものの大まかな傾向においては同じ分布を示しており、安定した解析結果が得られたと言えよう⁶。というのは、階層的クラスタ法での分析同様、場所に関連づけられるクラスタ4の用法が

⁴ 変数の頭に付いている「?」は貢献度に対して無いとも有るともはつきりとは評価できないことを意味する。

⁵ 反対に本稿の分析結果が示唆する分散的見方は、影山(1996)のような(動詞の意味のみからすべての文現象を捉えようとする)還元論的動詞クラスモデルに対し、経験面・理論面での問題提起を可能にし、再考の必要性を示唆すると言えよう。

⁶ このことが間接的に示唆するものとして次のことが言える。まず、階層的クラスタ法がボトムアップ的分析手法であるのに対して、非階層的クラスタ法はトップダウン的手法とされ、両者は別のアルゴリズムに基づいてデータを分類する(3.3節)。この両者の分析結果、すなわち表6と図5の数値間で大よその一致を見たということは分析手法によって結果がゆれないことを意味しており、データおよびコーディングの安定性を間接的に示唆する。ただ、中級では相対的に大きな差を見せており、若干の不安定さが認められるが、初級や上級における安定さから考慮すると、おそらく中級という過渡期的なレベルに由来する分布と考えられ、分析結果そのものの信頼性を否定する証拠ではないと考えられる。

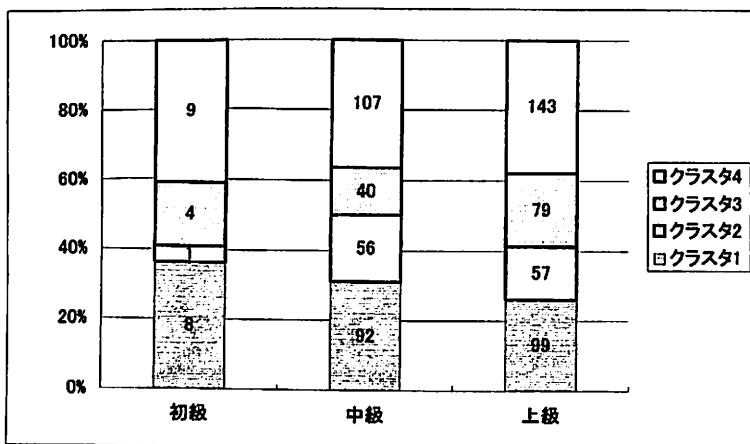


図4 レベル別のクラスタ分布図

もっとも安定して、使用されていることが分かる。一方、クラスタ1と2においては、少し興味深い変化が見られる。まずクラスタ2においては、中級と上級の差がそれほど明確でないに對して、初級と中級において大幅な変動が見られる。すなわち中級以降において対人関係の文脈における相手、そして授受構文における受益者のマーカーとして助詞「に」が頻繁に用いられるようになる。次にクラスタ1においては、レベルが増加するにつれ、使用比率が逆に減少傾向を示しているのである。この点に関しては他の要素へ分化、すなわち助詞「に」が持つクラスタ1の機能が他の形態、例えば「へ」や「まで」や「から」によって代用されている可能性が考えられる。ただし、これはあくまで推論の域を脱しておらず、今後何らかの形で検証してみる必要がある。

さて、非階層的クラスタ法の利用に関連し、この手法を用いた場合、中心的事例と周辺的な事例を実際に得られる数値として（クラスタ中心とその中心から各事例への距離という形で）表現される。クラスタ4を例に考えてみよう。

- | | |
|---|--------|
| (5) a. 日本語を勉強して〈え〉日本の大学院 <u>に</u> （うん）入って | 0.5306 |
| b. みんながテク(*韓国地名:大邱) <u>に</u> 住んでいます。 | 1.1681 |
| (6) a. プサンからソウル <u>に</u> 行って教員研修生みんなが集まって | 0.5306 |
| b. 救急車と早くここ <u>に</u> 来て下さい | 1.0918 |
| c. プサンは南の方 <u>に</u> ある都市で暖かいところです | 1.4824 |
| (7) a. だからびっくりしました、あの温泉 <u>に</u> 入ったとき、 | 0.5306 |
| b. 車の中 <u>に</u> かばんをわすれてしまいました | 1.2532 |
| c. 自分が目指すところ <u>に</u> 向かって一生懸命やってください | 1.5824 |

(5) では初級学習者の発話例、(6) では中級学習者の発話例、(7) では上級学習者の発話例を示した。同時にその右側にはクラスタ中心（図 4）からの距離を示した。これによって個々の事例が持つ位置づけが明らかになる。というのは、クラスタ中心に対する遠近として相互関係が示されている。具体的には (5a) や (6a) や (7a) の場合、クラスタ中心から非常に近い値を示しており、このことから中心的な事例として位置付けることができる。その一方、(5b) や (6c) や (7c) のようにクラスタ中心から離れており、周辺例として位置付けられるものもあれば、中間例として (6b) や (7b) のような事例も存在する。こうした事実関係に対し、数値化による一貫した分析を提供できるのは非階層的クラスタ法を用いる利点の一つと言える。

以上で明らかにした分布関係が偶然でない限り、中間言語が持つダイナミズムに対して二点の興味深い示唆を与える。一点目に学習の時系列的変動は単なる増加のみを意味するものではない、という見方が可能になる。というのも学習の進展はそれぞれの用法の使用が均等に増加していく形で進むのではなく、学習のレベルによって使用の増加が著しいものと相対的に頻度が減少するものが入り混じっていることが示されたからである。二点目に言語習得における事実関係を単純にある用法を使用するか、しないかの二項対立的な見方で捉えるべきではないという見方が可能になる。助詞「に」の用法に関して言えば、初級ではなかった用法が中級や上級において存在する、といった単純化は正しくない可能性を示唆する。

5. 最後に（結論とまとめ）

本稿では、第二言語習得の途中過程で表れる様々な中間言語を記述する手法として、助詞「に」の使用状況を調査し、多変量解析に基づく分析モデルを提案した。本節ではこれまでの分析によって明らかになった事実関係を簡単に整理し、今後の課題となるべき問題を示す。

日本語学習者の助詞「に」の使用に対して本稿が明らかにした計量的事実は次の二点である。一点目は場所用法としての「に」がもっとも安定して見られることが示された。これは第二言語獲得のデータのみならず、助詞「に」が持つ典型的な機能をうまく反映した事実と言える。二点目に、対象用法は初級から上級に行くにつれ、減少する傾向であることが示された。その原因については、本稿のデータのみでは必ずしも明らかではないが、他の助詞や機能表現へ「に」の機能を分散している可能性が考えられる。ただ、この点に関しては推測の域を脱しておらず、今後、稿を改めて検証する必要があると言える。

次に、現象記述的観点からは次の点が明らかになった。それは個々のクラスタの形成を動機付ける要因が均等ではないことが示された。というのも、統語的要因に制約されているクラス

タもあれば、共起語の意味クラスに制約されているクラスタもあるなど、動機付けの問題として非均質性が観察される。このことから動詞クラスモデルの問題が示唆された。

さて、今回の調査・分析においてはデータの安定性に関して二点の問題点があった。一点目に初級学習者のサンプル数が少なく、数値としての安定性が必ずしも保証されているとは言えないこと。二点目に会話データを利用したことによる格助詞の省略現象への対策が充分でなかったことが挙げられる。このことから、今後の課題として分析対象データの網羅性と精緻性の向上が必要と言えよう。

謝辞

本稿の執筆に当たって多くの方々にお世話になった。特に査読者の先生には本稿の諸論点に関する貴重なアドバイスやご指導をいただいた。そして中本敬子さん（京都大学）には統計解析全般に関して数々のアドバイスをいただいた。山梨正明先生（京都大学）や黒田航さん（情報通信研究機構）には暖かいご指導と励ましをいただいた。この場を借りて感謝申し上げる。

〈参考文献〉

- 朝野照彦（1996）『入門 多変量解析の実際』、講談社。
- 石綿敏雄（1998）『現代言語理論と格』、ひつじ書房。
- 市川保子（1997）『日本語誤用例文小辞典』、イセブ。
- 影山太郎（1993）『動詞意味論』、くろしお出版。
- 黒田航、中本敬子、野澤元（2004）「意味フレームに基づく概念分析の理論と実践」、山梨正明ほか（編）「認知言語学論考第4巻」、ひつじ書房、pp.133-269。
- 迫田久美子（2001）「学習者の誤用を生み出す言語処理のストラテジー(1)・場所を表す「に」と「で」の場合」、『広島大学日本語教育研究』11、広島大学日本語教育学講座。
- 寺村秀夫（1982）『日本語のシンタクスと意味I』、くろしお出版。
- 寺村秀夫（1990）「日本語学習者の日本語誤用例集」（科学研究費 特別推進研究 「日本語の普遍性と個別性に関する理論的及び実証的研究」 代表者井上和子、 分担研究「外国人学習者の日本語誤用例集、整理及び分析」 資料）
- 長友和彦（1993）「日本語の中間言語研究—概観」、『日本語教育』81号 1・17、日本語教育学会。
- 野田尚史（編著）（2001）『日本語学習者の文法習得』、大修館書店。
- 福間康子（1997）「作文から見た初級学習者の格助詞「に」の誤用」『九州大学留学生センター紀要』8、九州大学留学生センター、pp.61-74。
- 宮本定明（1999）『クラスタ分析入門：ファジィクラスタリングの理論と応用』、森北出版株式会社。
- 村木新次郎（1991）『日本語動詞の諸相』、ひつじ書房。
- 森田良行（2002）『日本語文法の発想』、ひつじ書房。
- 森田良行（2005）『外国人の誤用から分かる日本語の問題』、明治書院。
- 森山新（2000）『認知と第二言語習得』、図書出版啓明。

- 山梨正明（2000）『認知言語学原理』、くろしお出版。
- 李在鎬（2003）「認知事象の複合的制約に基づく結果構文再考—構文現象の体系的記述を目指してー」、『認知言語学論考』No.3 183-262、ひつじ書房。
- 李在鎬（2004a）「助詞「に」の定量的分析への試み:語法研究の新たな手法を求めて」、『日本認知言語学会論文集』No.4、pp.55-65。
- 李在鎬（2004b）「自動移動構文の習得からみた文法習得:中間言語的アプローチの精緻化に向けて」、『2004年日本語教育学会国際研究大会予稿集1』、pp.71-76。
- Braidi, Susan. M. (1999). *The Acquisition of Second-Language Syntax*. Oxford: Oxford University Press.
- Fillmore, Charles J. et al. (1988). "Regularity and idiomacticity in grammatical constructions: the case of *let alone*", *Language*, 64, pp. 501-538.
- Gass, Susan. M. & Larry Selinker. (1994). *Second Language Acquisition: An Introductory Course*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Goldberg, Adele, E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Kohonen, T (1995) *Self-Organizing Maps*, Berlin: Springer Series in Information Sciences, Vol. 30.
- Langacker, Ronald. (1987). *Foundations of Cognitive Grammar, Vol.1 Theoretical Prerequisites*, Stanford, Calif.: Stanford University Press.
- Langacker, Ronald. (1999). "A Dynamic Usage Based Model". in Barlow, Michael and Suzanne Kemmer (eds.) *Usage Based Models of Language*. Stanford: CSLI Publications, pp.1-64.
- Romesburg, Charles. (2004) *Cluster Analysis for Researchers*. Florida: Robert E. Krieger Publishing Company.
- SPSS. (1999) *Data Mining With Confidence*. SPSS.
- Tomasello, Michael. (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.
- White, Lydia. (1989). *Universal Grammar and Second Language Acquisition*. Amsterdam: John Benjamins.
- Yamanashi, Masaaki. (2002) "Cognitive Perspectives on Language Acquisition". *Studies in Language Sciences*. Vol.2. pp.106-116.

〈言語資源〉

- 鎌田修・山内博之（1999）「KY コーパス Version 1.1」(<http://www.opi.jp/shiryo/kycorp.html>)
- NTTコミュニケーション科学基礎研究所、『日本語語彙大系 CD-ROM 版』、岩波書店

(2005年2月28日受付、2005年11月8日最終稿受付)