

統計モデルを用いた助詞「で」の分析

リ ジェホ
李 在鎬 (jhlee@nict.go.jp) 井佐原 均 (isahara@nict.go.jp)

(独立行政法人 情報通信研究機構 けいはんな情報通信融合研究センター)

1. はじめに (本稿の目的)

本稿では、用法の多様さからその体系的記述が難しいとされる助詞「で」を多変量解析の手法を用いて分析し、その利点を明らかにする。その際、コーパスから収集したデータをクラスタリングし、その結果を分析しながら、考察を進めていく。最終的に(記述的観点から)「で」の分布関係に関連し、次の二点を主張する。1) 助詞「で」の分布関係は4つの基本的用法の区別を行ったレベルで最適な分離が得られる。2) 助詞「で」の用法は(項となる)名詞や(それを要求する)動詞のような単一要素ではなく、文脈全体によってシステムティックに動機付けられている。主張1)の論証としてコーパスデータに対するクラスタ分析(cf. Romesburg 2004)と判別分析の解析結果を報告する。主張2)の論証としてk-means法によるクラスタ中心値を報告し、意味の実現が分散的である点を指摘する。次に、本稿の分析手法に関連し(理論的観点から)次の点を主張する。本稿の分析モデルは、言語現象に対する定量的分析の具体的な方法論を提案するものであり、次の二点において(言語現象への記述手法としての)妥当性が示唆される。3) 徹底して言語事実に基づく分析手法であるゆえ、理論仮構物に頼らず自然な記述分析を可能にする。4) 分析結果を数値化することで他の分析との間で分析結果を比較・検証することが容易であり、主観の混入をコントロールできる。なお、3)によって認知言語学における用法基盤モデル(cf. Langacker 2000)へ経験的支持を与え、4)によって記述のデータベース化に貢献できる点を示唆していきたい。

2 問題提起 (記述的問題)

助詞「で」の記述的問題に関連し、深層格(cf. Fillmore 1968, 山梨 1983, 国立国語研究所 1997)で代表される従来の記述モデルは、潜在的意味範疇の重要性を明らかにした。具体例として1と2を見てみよう。

- 太郎は、警察で朝ご飯を食べた。
 - 太郎の罪は、警察で調べる。
- 酒で傷口をきれいに洗った。
 - 酒で人生を駄目にした。

1における「警察」、あるいは2における「酒」は、一見したところ、同一の形式を共有しており、単一の現象と思われるであろう。しかし、それぞれが文内で担っている(意味論的)役割という点から現象を見直した場合、両者はまったく別の振舞いを見せていることに気が付くであろう。というのは、1aにおける「警察」は、「食べる」という行為が成立する物理的な空間ないしは場所として解釈されるのに対して、1bにおいては、「太郎の罪は刑事が警察署で調べる」という解釈に含むことになり、この

場合の「警察」は「調べる」という行為を遂行する動作主として解釈される¹。これは、Langacker (1990) のセッティング参与者モデル (setting-participant model) に依拠して言えば、後者の「警察」が参与者を表すのに対して、前者の「警察」はセッティングを表しており、はっきりとしたコントラストを見せていると言える。同じく 2a において、「酒」は道具であるのに対して、2b では原因として解釈される。このことを事態認知レベルで捉えた場合、以下のように示すことができる。

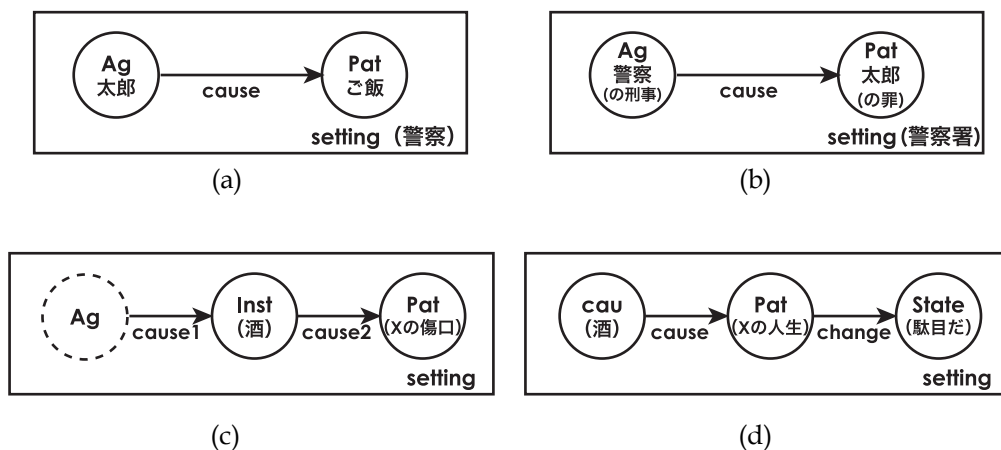


図.1

図 1 の a は 1a に、b は 1b に、c は 2a に、d は 2b にそれぞれ対応する。これらの認知図式が示唆するように、一つの名詞句であっても文内での実際の振舞いは予測しにくいものがあり、この点を明らかにするためには、文法現象における意味論的性質が重要となる。このことは、従来の日本語学においても（本稿とは異なる見方で）指摘されてきた（cf. 寺村 1982、村木 1991、国研 1997、石綿 1998）。

さて、上記において示した基本的な記述的問題に加え、さらに、難しい問題として、3 や 4 のような事例が存在する。

- | | | | | | | | | | | | | | |
|-------|-----|----|----|---|----|-----|-----|-------|---|-----|-----|----------|----------|
| 3. a. | 雰囲気 | で | 観客 | を | 圧倒 | する。 | ・・・ | 原因/道具 | | | | | |
| b. | 扇風機 | で | 指 | を | 切 | った。 | ・・・ | 道具/原因 | | | | | |
| 4. a. | 最後 | には | 4人 | で | 日本 | で | 暮 | ら | し | たい。 | ・・・ | ?動作主/?様態 | |
| b. | 社長 | の | 考 | え | で | 長 | い | 方 | を | 決 | めた。 | ・・・ | ?動作主/?原因 |

3 や 4 の事例が特徴的なのは、1 や 2 のように、単一の意味役割へ還元できないことである。というのは、3 における二つの事例は、原因と道具の間でどちらの解釈をも許容しており、二者間で分析のゆらぎが生じる。となると、どちらか一方へ還元することは難しい。4 においても、事情は異なるが、結果的には 3 と同じ扱いになる。というのは、4 の場合、動作主とも様態とも言い難い振舞いを見せており、記述的困難さが潜む。こうした事実関係に対して、認知言語学は重要な見方を示した。とりわけ山梨 (1995) ではそれぞれのカテゴリーが相互排他的でないところに着目し、(プロトタイプ効果に基づく) 連続的カテゴリーモデルによる分析を提案している。そこでは、深層格モデルの理論的欠陥として、二者択一的な分析の限界を指摘しており、視点による解釈のゆらぎをうまく捉えている。後の分析で明らかになることであるが、本研究はこの種のプロトタイプカテゴリーモデル (cf. Lakoff 1987) を継承し、その記述的精緻化を試みる。

以上の考察で、先行研究が明らかにした知見として、二点の大きな成果が明らかになった。

- 1) 一言で「で」といってもその意味的振舞いは一枚岩ではなく、多種多様であること
- 2) それぞれの意味的性質はクリアカットに分離できるものではなく、プロトタイプ効果を示すこと

さて、こうした先行研究を引き継ぎ、新たな記述的要請として以下の二点が言えよう。

- I. どれだけの用法を認めれば良いか
- II. 制約のあり方はどのように捉えれば良いのか

I の記述的要請は、先行研究の成果 1) を引き継ぐためのものであり、同じく II は先行研究の成果 2) を引き継ぐためのものである。本稿では、これら二点の記述的要請に対してコーパススペースの分析モデルを用いて考察する。とりわけ I の問題をめぐっては、生起文脈を定量的に捉え、個々の用法に対する構文スキーマを抽出することで記述する。具体的には、コーパスからデータを収集し、クラスタリングアルゴリズムに基づく分析を行う。II の問題をめぐっては、分散意味論 (cf. 黒田・中本・野澤 2004) の考えに基づいて複数のユニットに制約を分散することで記述する。具体的には名詞の意味クラスや統語環境といった共起関係による複合的制約から現象を捉える。

3. 方法論 (データと解析方法)

本稿では 2 節で示した「で」の使用をめぐる記述的要請に対し、コーパススペースの統計的・定量的分析を導入する。その狙いの二つある。一つ目は、発見の手続きを明確化することで、言語直感を基盤とする内観的 (introspective) 記述モデルの補強を試みる。二つ目は、結果を可視化することで事実を共有できるようにすることである。

さて、具体的なコーパス実験の概要を示す前に、本稿が直面する方法論上の問題について簡単に触れる。まず、確認しておきたいこととして、従来のコーパスが得意とする、いわゆる数の算定によって結果を導き出す手法 (cf. 斎藤 1998, Hunston&Francis 1999) で、前節で報告した現象が扱うのは難しい。なぜなら、この場合は、意味の問題が深く関与しており、単なる量的性質に事実を還元することはできないからである。ここでは、むしろ、質の面からの分析が要求される。こうしたことを踏まえ、意味の問題を正当に扱うためのモデリングを行った。その取っ掛かりとして、注目したのは、多変量解析やテキストマイニングに基づくモデリングの手法である (cf. Groth 1997, Kohonen 1997, SPSS 1999, 朝野 1996)。紙幅の都合上、その詳細を紹介することはできないが、大まかな流れとして以下の手順で解析を行った。

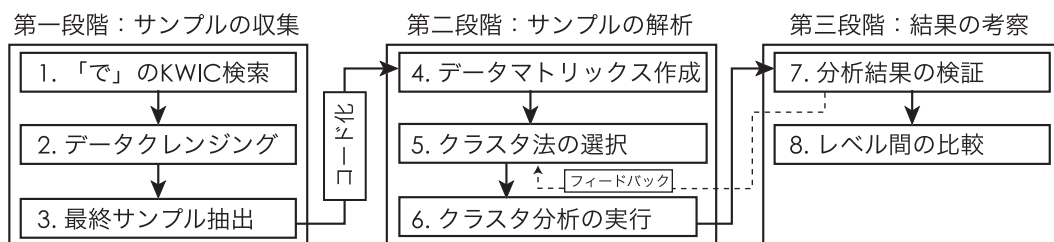


図.2 分析モデル

以下に、コーパス実験の概要を示す。

- 利用データ：コーパス調査においては、8 人の書き手による（現代小説の）プレーンテキストから冒頭部の 5000 字を抽出、総計 4 万字のミニコーパスを作成したⁱⁱ。
- サンプルの収集：上記のデータに対して KWIC 検索を実行、さらにデータクレンジングを行い、助詞「で」を含む 229 個の文をサンプルとして収集した。
- コード化：潜在的意味解析法 LSA(cf. Landauer et al. 1998)を用いて助詞「で」の前後の生起文脈（共起関係）を表 1 の変数群で分散表示した(cf. Kohonen 1995, 李 2004) ⁱⁱⁱ。

区分	変数
共起名詞の意味クラス ^{iv}	主体、場所、具体物、抽象物、事、抽象的關係
文の主語の意味クラス	有情物、無情物
統語的位置関係	文頭、ガ格後、ガ格前、ヲ格後、ヲ格前、ニ格後、ニ格前、述部前
述語の特性	動作、状態、受動態

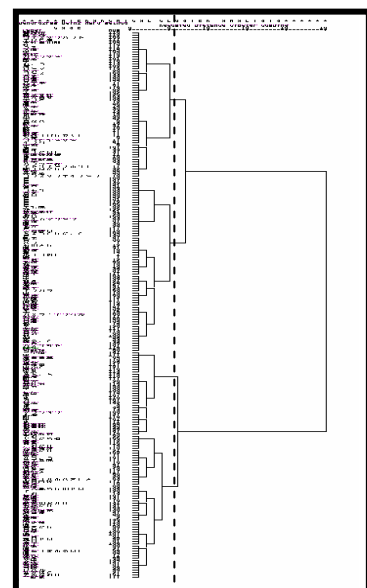
〈表 1：変数一覧〉

- コード化の具体例：「女が素手でタイヤをにぎった」における「で」の特徴づけは、共起する名詞の意味クラスが具体物である点(同時に事や抽象的關係等ではない点も)、主語が人である点、統語構造においてガ格に後続し、ヲ格に先行している点、動作述語である点がコード化の対象となる。このことを 1/0 のベクトルで示すと「0010001001001000100」になる。なお、名詞や動詞の意味クラスに関するすべての判断は論者の主観を極力排除するため「日本語語彙大系」(NTT コミュニケーションズ)に準拠して行った。
- 統計解析：「SPSS Ver 13.0(Win 版)」を利用し、階層的クラスタ法と非階層的クラスタ法の両方を用いて検討し、判別分析による検証を行った^v。

4 結果

前節の方法論に従って、階層的クラスタ法で全体のデータを解析した結果、図 3 のデンドログラムが得られた。なお、クラスタ化の方法はウォード法で、距離の定義は、ユークリッド距離を使用した。

図 3 からは、それぞれのケースが併合し、より大きなグループを形成する過程が示されている。ここから、示唆されることとして、4 つの大きな塊の存在が伺える。この解析結果に基づいて、より明確な分離を得ることを目的とし、(非階層的クラスタ法の代表的手法である) k -means 法 (cf. 宮本 1999) で解析を行った。解析においては、 $k=3$ (3 個のクラスタ) から $k=6$ (6 個のクラスタ) まで分析精度を比較しながら、最適な分類結果を求めた。最終的には $k=4$ で最適な分離が得られた。なお、この結果を確かめるべく、判別分析を行い、 k -means 法によるクラスタを従属変数とし、表 1 の変数を独立変数として投入



〈図.3 デンドログラム〉

し、予測制度を評価した。その結果、 $k=3$ で 97.8%、 $k=4$ で 99.1%、 $k=5$ で 86.9%、 $k=6$ で 93.5%が正しく分類された。これによって $k=4$ の妥当性が示された。 k -means 法によるクラスタの代表（中心値）を示す。

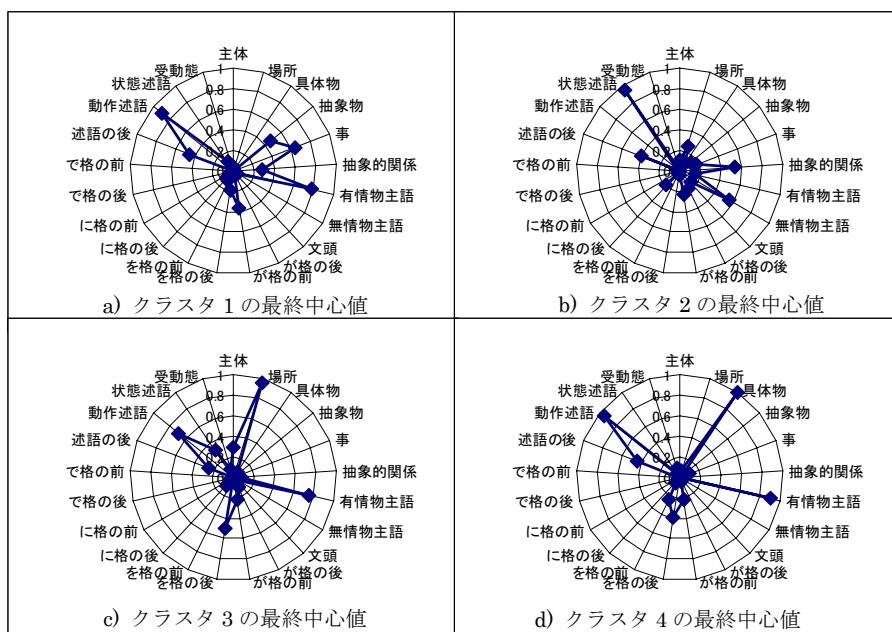


図.4 クラスタ中心値

図 4 では、各クラスタ中心における変数の値を図示している。これに基づいて、それぞれのクラスタの生起文脈を特定することができる。

区分	生起文脈	实例
クラスタ 1 (53 例)	人が事/抽象物で述語 (動作性)	a. 私たち友人に向けて I がつきつめた表情で言った。 b. 荒井は幸せ一杯という様子で食べ始める。
クラスタ 2 (74 例)	物が抽象的關係で述語 (状態性)	a. バスは時間が不定期で困るよ。 b. 世界とは凝縮された可能性で作上げられたコーヒーテーブルなのだ。
クラスタ 3 (47 例)	人が場所で X を述語(動作性)	a. あれから数年たって大阪のテレビ局で顔を合わせて、 b. 陽のまったくささない部屋で毎日厚い法律書を読んでいた。
クラスタ 4 (55 例)	人が具体物で述語 (動作性)	a. 私は清高をつれて慌てて新幹線で駆けつけました。 b. フン先生はでこぼこの大釜で一週間分のごはんを一度に炊き

〈表.2 クラスタの生起文脈と实例〉

表 2 では、個々のクラスタにおける数と、クラスタ中心値から推定された生起文脈、そして、実際の使用例をあげた。例えば、クラスタ 1 は、全部で 53 例あり、その生起文脈を見た場合、人が事ないしは抽象物で何かの動作を行う、というところで使用されていることが分かる。こうした分析が言語の理解においてどのような役割をするか、次節で考察する。

5 考察

ここからは、前節で報告した分析結果が 2 節の記述的要請および問題提起に対して、どのような示唆を与えるかを中心に考察する。5.1 節では、「で」の現象記述的観点から考察し、5.2 節では、本稿の分析手法の理論的位置づけに関して考察する。

5.1. 助詞「で」の記述をめぐって

さて、前節で報告したクラスタ分析の結果は、従来の言語研究に対して様々な事実を示唆する。特に重要な論点の一つとして、2 節で議論した深層格との関連から考えてみた場合、各クラスタと深層格の間には一定の対応関係が存在する。

区分	深層格との対応
クラスタ 1	様態
クラスタ 2	原因・動機
クラスタ 3	場所
クラスタ 4	道具

〈表.3 クラスタと深層格の対応〉

表 3 では、クラスタリングの結果を深層格との関連で位置づけている。まず、クラスタ 1 における「で」は、主に事や抽象物を表す語 (e.g., 気分、新婚旅行、口調) と共起し、様態を表す。クラスタ 2 は抽象的關係を示す語 (e.g., 状態、百点、安手) と共起し、原因や動機を表す。そして、クラスタ 3 では、もっぱら空間を表す語や動作述語と共起し、動作の場所を表す。最後に、クラスタ 4 では、具体物を表す語 (胴皮、果物、古なべ) と共起し、道具を表す。

さて、この対応関係が正しいものであるなら、我々は助詞「で」の意味的振舞いをめぐって新たな見通しを得ることができる。次の二点として集約できる。一点目に、「で」の用法の数に関して、4 種類の使用を認めることで、最適な記述が可能となる。二点目に、「で」の制約のあり方に関して、生起文脈に基づく一般化が可能となる。例えば、「で」の典型的な役割の一つとされる道具格の制約に関していえば、「有情物主語が具体物で動作を実現する」という文脈パターンの制約から生じるものと記述することができる^{vi}。その間接的な証拠として、表 1 で示した変数一覧表に再度注目した場合、いずれの変数も表 3 の深層格を直接的に表現していないことに気が付くであろう。表 1 から分かる通り、コード化において用いたのは、表層における複数の文脈情報のみである。前節の結果は、これらの複数の文脈情報から深層格の住みわけに匹敵する有意義な情報をうまく抽出できたことを示す。なお、これら二点の見通しから、2 節の記述的要請 I、II についても自然な記述を与えることが可能となる。

さて、ここで問題になるのは、2 節の 3 や 4 のような現象に対してどのような扱いになるかである。結論から言えば、すべてのケースは、多次元空間上の位置関係から統一的に示される。というのはクラスタ中心からの距離を得ることができるからである。クラスタ 4 の具体例の一部を示す。

5. a. 一度だけ、配達用の青自転車^で歌舞伎町を駆け抜けた事がある。 〈0.82761〉
 b. 私は便宜的に、右の脳^で右ポケットの勘定をし、……。 〈1.09287〉
 c. 月謝を納めた者の名前の上に、赤インク^で棒が引かれ、 〈1.29197〉
 d. 怖 [おそ] ろしい苦痛を自分の体^で味わう前に、……。 〈1.40284〉
 e. あまり気の入っていない顔^で聞いた。 〈1.56797〉

5に示したすべての例は、クラスタ4の具体例であり、表3の分析によれば、道具を具現化する集団として解釈される。しかし、これには、典型性の点において差があり、その差を記述する手法として、右側のクラスタ中心からの距離を利用することできる。例えば、5aのように値が少なければ、それは中心から短い距離、すなわちより典型性の高い事例であると言えるし、反対に5eのように値が大きければ大きいほど、それはクラスタ中心から離れた拡張事例であると分析することができる。このようにクラスタ中心からの遠近を得ることで、コアから周辺までを統一的に記述できる。

なお、以上の分析の背景には、制約の分散による記述的戦略が存在する。具体的に言えば、このアプローチにおいて、道具格は5つの主要な素性とその値の分散によって表現される。

素性	値
具体物	1.000
有情物の主語	0.909
ヲ格の後	0.382
述語の前	0.436
述語[動作]	0.945

表4

表4では、「で」の道具用法を表す主要な変数（特徴）と、その値（貢献度）が示されている。なお、個々の値は1（最大）から0（最初）の範囲で変数の貢献度を示しており、こうして複数のユニットに制約を分散することで、3や4のような中間例に対しても柔軟に対応できると考える。紙幅の都合上、この点をめぐる詳細は論述はできないが、理論的側面の詳細は、黒田、中本、野澤（2004）を参照されたい。

5.2. 方法論的位置づけをめぐって

本稿は、これまでコーパスデータに対して統計学的分類アルゴリズムでデータを解析し、その記述の有効性を示した。次のステップとしてこうした記述的手法の位置づけに関する問題を考察したい。

この問題を考える上で、特に注目したいのは、認知言語学のカテゴリー化モデル（cf. Lakoff 1987）との関係である。というのも、本稿が提案するクラスタモデルは、細部の議論を捨象した場合、計算機を用いた多義語のカテゴリー化のモデルであるとも言えるからである。その根拠として、前節の議論ですでに示唆されたように、クラスタモデルとカテゴリー化モデルの間には、次のようなアナロジーが成り立つ。それは、クラスタ中心はカテゴリー化モデルにおけるプロトタイプと考えることができる。そして、クラスタ中心からの距離はカテゴリー化モデルにおける拡張関係に対応づけられる。そして、変数群はカテゴリー化における特徴づけを明示化したものと理解できる。これらの点におい

て、クラスタモデルはカテゴリー化モデルときわめて親和性の高いものと位置づけることができる。

最後に、本稿の分析が意味役割の議論全体に、示唆する見方は次のようにまとめられる。それは、意味役割は単一要素の制約からは本来予測できないものである。すなわち、動詞だけによるもの (cf. 影山 1996) でもなければ、(項構造) 構文だけによるもの (cf. Goldberg 1995) でもない。これには、むしろ、個々の使用文脈における複数の要素間の相互作用によって生じるものと捉えるべきであり、この点において従来の (還元主義的) モデルは重大な特徴を見落としていると言わざるを得ない。

6 最後に (定量化による文法記述の可能性)

論文のしめくくり、本研究の分析手法が示す見通しを簡単に述べておきたい。主として定量化の必要性和その利点を明らかにしておきたい。

本研究では、従来の枠組みが明らかにした知見の延長上に、統計モデルによるデータの再解釈の方法を提案した。具体的には、文脈の制約に基づく助詞「で」の意味分析を行った。これにより、従来の還元主義的分析に対する問題提起をし、分散論的手法からの (従来の記述の) 補強を行った。とりわけ多変量解析の手法を用いることで「文脈」といった複雑な現象に対しても、(計算機を活用して) 一貫した基準から定量的に分析することが可能であることを示した。だが、これまでの考察においては、「なぜ定量的分析の必要か」という疑問に、必ずしも明確な答えは出していない。こうした点に関連して、本研究の分析手法には少なくとも以下の利点がある。

- a) 言語現象の隠れた性質を (ただデータを見つめているだけで費やす時間と労力に比べれば) 短時間で見つけ出すことが可能。
- b) 言語事実の「とりこぼし」の危険性と「観察の偏り」を回避。
- c) 複雑で抽象度の低い大量の言語データを一貫した基準で分析・記述することが可能。
- d) 分析結果 (の妥当性) を評価する基準 (e.g., 分散分析や判別分析など) が存在する。
- e) 結果を数値化することにより、他の解析結果との比較が容易。

以上の利点が示すように、定量的・統計学的アプローチは、直感を越えたレベルの一般化を可能にし、最終的には一解釈としての言語分析のレベルから科学として妥当な言語分析へのレベルアップを可能にする。

* 謝辞: 本稿は第 30 回関西言語学会 (KLS30) での口頭発表の内容をもとに作成された。司会者の野田尚志先生およびフロアの皆さんさんには貴重なコメントいただいた。また、本稿の執筆において黒田航氏 (情報通信研究機構) には議論の細部に至るまでコメントをいただいた。この場を借りて感謝を申し上げたい。なお、本稿におけるすべての誤りは筆者に帰するものである。

ⁱ この種の事実関係の背景にはメトニミの問題が関連している。文法構文におけるメトニミの問題は西村 (2002) を参照されたい。

ⁱⁱ データとして利用したのは、新潮文庫 100 冊のテキスト版の中から、井上ひさし『ブンとフン』、遠藤周作『沈黙』、北杜夫『楡家の人びと』、宮本輝『錦繍』、村上春樹『世界の終りとハードボイルド・ワンダーランド』、塩野七生『コンスタンティノーブルの陥落』、筒井康隆『エディプスの恋人』、椎名 誠『新橋烏森口青春篇』をテキストとして利用した。

iii この手法は主として計算言語学・コーパス言語学でよく用いられており、(簡単に言ってしまえば) ある語が生じる文脈の集合そのものでその語の意味を近似する手法である。本発表のテーマで言えば、相関を利用して「で」が生じる文脈そのものの集合から「で」を定義する、という位置づけになる。

iv 共起名詞の意味クラスを解析に含めた背景には作例ベースの考察の結果、以下のすみわけが観察できた。

- 1 a. 片手で相手を倒した。 > 具体物→道具性への動機付け
- b. 校庭で相手を倒した。 > 物理的空間→場所性への動機付け
- 2 a. 看護演習で倒れた。 > イベント→原因性への動機付け
- b. 暑さで倒れた。 > 抽象物→原因性への動機付け

ただし、これらはあくまで必要条件としての想定されたものであり絶対条件ではない点に注意してほしい。

v 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」と、似たものが結果的に同じグループに入るように集合を分割する「非階層的なクラスタ法」とに大別して考えることができる。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、後者は単一の水準でクラスタを作る手法で、予めクラスタ数を決めてから分析することから、トップダウン的な分類であると言える(cf. Huberty 1994, 宮本 1999, Romesburg 2004)。

vi なお、紙幅の都合上、詳細な論述はできないが、ここで言う、共起関係に基づく生起文脈のパターンは、構文文法(Construction Grammar; Fillmore 1988, Goldberg 1995)で主張される文法構文(grammatical construction)と極めて親和性の高い概念であると言えよう。とりわけ Gries (2003)、Stefanowitsch & Gries (2003)においては Collostructional Analysis (In this research project, we are developing corpus-based methods for investigating the interaction between words and constructions. Our research so far has focused on the development of three methods collectively referred to as collostructional analysis: collexeme analysis, distinctive-collexeme analysis, and covarying-collexeme analysis.) という分析の枠組みを提案しており、共起関係に基づき、構文の実証的研究を行っている。

〈参考文献〉

朝野熙彦 (1996) 『入門 多変量解析の実際』、

Fillmore, Charles J. (1968). "The case for case", in E. Bach and R.T. Harms (eds.) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, P.1-88

Fillmore, Charles J. (1988). The mechanisms of 'Construction Grammar.' *BLS* 14: 35-55.

Goldberg, Adele, E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press. (池上誓作 他(訳)『構文文法論』、東京: 研究社、(2001.))

Gries, Stefan Th. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1. pp.1-27.

Groth, Robert (1997) *Data mining: A hands-on approach for business professionals*. Englewood Cliffs, N.J.: Prentice Hall.

Huberty, C. J. (1994) *Applied discriminant analysis*. New York: John Wiley & Sons, Inc.

Hunston, S. & Francis, G (1999) *Pattern Grammar - A Corpus Driven Approach to the Lexical Grammar of English*, Amsterdam: John Benjamins.

石綿敏雄 (1998) 『現代言語理論と格』、ひつじ書房。

影山太郎 (1996). 『動詞意味論—言語と認知の接点』、東京: くろしお出版。

国立国語研究所 (1997) 『日本語における表層格と深層格の対応関係』、三省堂。

Kohonen, T (1995) *Self-Organizing Maps*, Berlin: SSIS, Vol. 30.

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.
- Lakoff, George. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.*, Chicago: The University of Chicago Press. (池上嘉彦・河上誓作 他(訳)『認知意味論』、東京: 紀伊国屋書店、(1993.))
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998) .“Introduction to Latent Semantic Analysis.” *Discourse Processes*, 25, pp.259-284.
- Langacker, Ronald W. (1990) *Concept, Image, and Symbol. The Cognitive Basis of Grammar*, Cognitive Linguistics Research 1, Berlin-New York: Mouton de Gruyter.
- Langacker, Ronald W. (2000) “A Dynamic Usage-Based Model.” Michael Barlow and Suzanne Kemmer (eds.) *Usage Based Models of Language*. Stanford, Calif.: CSLI Publications, pp.1-63.
- 李在鎬 (2004) 「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
- 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.
- Romesburg, Charles. (2004) *Cluster Analysis for Researchers*. Florida: Robert E. Krieger Publishing Company.
- 斎藤俊雄 (編) (1998) 『英語コーパス言語学 基礎と実践』、研究社.
- SPSS. (1999) *Data Mining With Confidence*. SPSS.
- Stefanowitsch, Anatol, and Stefan Th. Gries. (2003). Collocations: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8.2: pp.209-243 (<http://www-user.uni-bremen.de/~anatol/publications.html>)
- Taylor, R. John. (2003) *Cognitive Grammar*. Oxford: Oxford University Press.
- 寺村秀雄 (1982) 『日本語のシンタクスと意味 1』、くろしお出版.
- 山梨正明 (1983) 「格文法理論」、安井稔編『意味論英語学大系』、大修館書店、PP.467-547.
- 山梨正明 (1995) 『認知文法論』、ひつじ書房.

〈データベース〉

- ・ 新潮社、「CD-ROM 版 新潮文庫の 100 冊」
- ・ 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.