

多変量解析を用いた構文の記述

京都大学大学院 人間・環境学研究科

李在鎬(り・ちえほ)

lee@hi.h.kyoto-u.ac.jp

1. はじめに

本稿では多様な用法故に、一般的記述が難しいとされる格助詞「に」に関わる統語現象を統計学的手法に基づき、分析する。とりわけ[4]の方法論の用いてコーパスから得られたデータをクラスタ分析や判別分析といった多変量解析の手法で分析する。なお分析に関連した記述の問題は[1], [2]の構文文法の観点に立って考察する。

最終的には次の論点を明らかにする。語法の問題として a) 「に」の意味的分布は統語構造や名詞といった使用文脈の制約に支えられていること、b) 「に」の分布はそれが直接に共起する語(主として名詞)の制約を受けるクラスタ、統語的関係の制約を受けるクラスタ、両者(語と統語環境)の制約を相互に受けるクラスタが見られることを示す。次に本稿の分析手法の妥当性を示すべく、c) 複雑な言語現象に対し、システムティックな記述・分析を与えることが可能である、という見通しを示しておきたい。

2. 問題提起

深層格で代表される従来のアプローチでは動詞こそが文の中心であり、格現象はそこから生じる付随的現象と位置づけられ、(1)から(4)に示すように特定の動詞に特定の項構造を仮定する。

- (1) 行く (動作主, 着点)
- (2) あげる (動作主, 被動作主, 受益者)
- (3) 入れる (動作主, 主題, 着点)
- (4) 変える (動作主, 被動作主, 結果)

しかし、動詞の項構造を仮定する分析においてはしばしば一貫性の欠けた分析が見られ、とりわけ、次の3点の問題点が見られる。a) 無理な一般化を行う点、b) 記述の歯止めがない点、c) 記述の対象が不明確な点である。動詞を中心とした分析で

は[2]が指摘するように動詞の意味から予測困難な場合、無理な一般化をせざるを得ない。また動詞の意味の数だけ深層格を増やさなければならず、記述の歯止めがきかない[3]。これに関わる問題点は[5]で代表される従来の日本語分析においても明確に表れており、多くの研究は反証不可能な循環論を引き起こしている。

- (5) 態度: 子供が親に甘える。
- (6) 起因: 弟が借金に悩んでいる。
- (7) 部分: 父親が娘をひざにだく。
- (8) 資格: 彼をゲストに迎える。

上記の分析は[5]によるもので、ここで問題となるのは分析者自身は「に」格を記述しているつもりであっても(5)や(6)に見られるように、むしろ動詞の意味を記述していたり、あるいは(8)のように名詞句を記述していたり、さらには(7)のように名詞句と名詞句の関係を記述しているなど、観察対象に対する一貫性が保たれておらず、しばしば大きな問題を引き起こしている。

本稿では過去の研究が抱えたこうした問題点を克服すべく、動詞といった要素の制約をプリミティブには仮定せず、格現象の本質を(「に」の使用文脈としての)構文に還元する方向で分析を試みる。

3. 方法(データとコーディング)

本稿ではテキストが持つバイアスを排除すべく[6]の中から複数の作者による8冊の現代小説¹の冒頭の1000文字のみを取り出し、総文字数約8000字のミニコーパスを作成した。そこから「に」を含む203の文をサンプルとして採集した。次に各々

¹ 赤川次郎『女社長に乾杯!』/大江健三郎『死者の奢り・飼育』/村上春樹『世界の終わりとハードボイルド・ワンダーランド』/椎名誠『新橋鳥森口青春篇』/宮本輝『錦繡』/筒井康隆『エディプスの恋人』/五木寛之『風に吹かれて』/司馬遼太郎『国盗り物語』

のサンプルに対して二つの属性セットを用いてデータマトリックスを作成した。一つ目として「に」の共起語の性質 (FS1) として 6 ベクトル、二つ目として「に」が生じる統語関係 (FS2) として 11 ベクトル。そしてこれらのベクトルを用いて個々のサンプルを 1/0 で表現した。なお FS1 をめぐるすべての判断は [7] に準拠して行った。

- FS1:主体、場所、具体物、事、抽象物、抽象的關係
- FS2:文頭、後続_が、後続_を、後続_に、後続_から、先行_動詞、受動形、先行_形容詞、先行_が、先行_を、先行_に
- 例) 彼は訪問者をエレベーターに**乗**せた
→01100000100100000 (FS1: FS2:)

以上の作業によって、二値データで構成されたデータマトリックスが得られた。次に実際のクラスタ分析を行った。解析においては非階層的的手法 (K-means) と階層的的手法 (ワード法、平方ユークリッド距離で測定) をともに用いて検証した。

4. 結果

4.1 クラスタ分析

最終的に 5 つの (十分な非類似度を保証する) 有意義なクラスタが観察される。

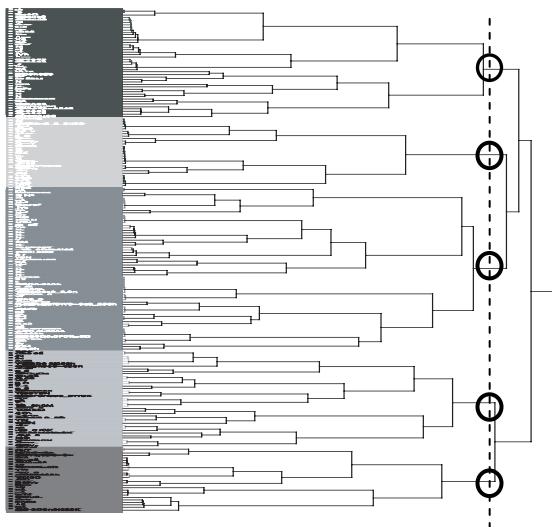


図.1 ワード法によるデンドログラム

以下、最終クラスタ中心をあげておく (K=5/反復回数は 12/初期の中心間の最小距離は 2.236)。

表.1 <最終クラスタ中心値>

区分	C1	C2	C3	C4	C5
主体	0.0213	0	0.1312	0.1177	0.1724
場所	<u>0.9362</u>	0.0204	0	0.1765	<u>0.7241</u>
具体物	<u>0.5957</u>	0	0.0164	0	<u>0.4138</u>
抽象物	0.0213	0.1225	0.4426	0.3529	0.0690
事	0	<u>1</u>	<u>0.5738</u>	<u>0.6471</u>	0
抽象関係	0.277	<u>0.4286</u>	0.4426	<u>0.4706</u>	0.2759
文頭	0	<u>0.5918</u>	0.1475	0	0.4828
後続_から	0.0213	0	0	0	0
後続_が	<u>0.6596</u>	0.2857	<u>0.6721</u>	0	0.3103
後続_を	0.1064	0	0	<u>1</u>	0
後続_に	0.0213	0.0408	0.0164	0	0
先行_動詞	<u>1</u>	0.1020	<u>1</u>	<u>0.8235</u>	0
受動態	0.1702	0	0.1148	0.1177	0
先行_が	0	0.3061	0	0	<u>0.4138</u>
先行_を	0.0213	0.2245	0	0	<u>0.5172</u>
先行_に	0	0.1429	0	0.1765	0
先行_形	0	0.2041	0	0	0.0690
ケース数	47	49	61	17	29

次にこれらのクラスタを実際の文内の動詞と付き合わせてみた結果、クラスタ 1 は行為を表す自動詞表現における場所にクラスタリングされていた。またクラスタ 2 は動詞に関する一貫性がなく副詞句に、クラスタ 3 は状態を表す自動詞表現の様態に、クラスタ 4 は状態変化を表す他動詞表現の結果に、クラスタ 5 は使役移動を表す他動詞表現の着点にクラスタリングされた。その詳細を以下において順に報告する。

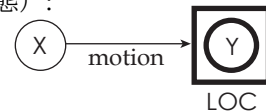
4.1.1 クラスタ 1

- 具体例

(9) 私がエレベーターに**乗**ったとたんに

- 形式: 「XがYにVする」

- 意味 (認知事態):



4.1.2 クラスタ 2

■ 具体例

(10) チップ代わりににマッチ棒が配られ、・・・

■ 形式: 「XにYがVする」

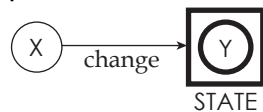
4.1.3 クラスタ 3

■ 具体例

(11) そのため私は無防備な気持ちになった。

■ 形式: 「XがYにVする」

■ 意味 (認知事態):



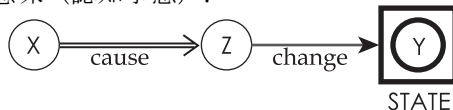
4.1.4 クラスタ 4

■ 具体例

(12) 水島は堅琴をかきたてました。曲は「春らんまん」をマーチ風に変えたものです。

■ 形式: 「XがZをYにVする」

■ 意味 (認知事態):



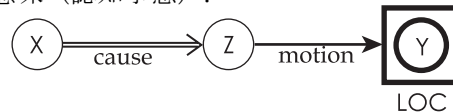
4.1.5 クラスタ 5

■ 具体例

(13) 私は右側のポケットに百円玉と五百円玉を入れ、左側に五十円玉と十円玉を入れる。

■ 形式: 「(Xが)YにZをVする」

■ 意味 (認知事態):



4.2 クラスタ間の関係

クラスタ間の親疎関係を表すクラスタ中心間の距離を以下に示す。

表.2 〈クラスタ中心間の距離〉

区分	1	2	3	4	5
1	-	1.940	1.640	1.676	1.386
2	1.940	-	1.294	1.440	1.572
3	1.640	1.294	-	1.258	1.697
4	1.676	1.440	1.258	-	1.873
5	1.386	1.572	1.697	1.873	-

表 2 ではクラスタ間の (非類似度を示す) 距離が示されている。これによって、クラスタ 2 は 1 や 5 と遠い関係であるのに対して、(相対的に) 3 や 4 と近い関係にあることが分かる。

5. 考察

5.1. 「に」の多義性とその動機付け

本節では、4 節で示したクラスタ分析の結果が従来、助詞「に」をめぐって行われてきた分析をどのように評価し、かつ統合しうるかを検討していく。主として「に」をめぐる構造的な多義性の動機づけに関する問題を考察し、結論的には三つの非対称的分布関係が見られる点を指摘する。まずは、

(14) から (16) の具体例から考えてみよう。

(14) a. ビスケット生地を円形タルト型に焼いたものに、果物やソースを...

b. もえは食の細い先生に色々と工夫をこらしたビスケットを焼いた。

(15) a. 一文銭で銀を見つけたのは金花を嫁に決めたあとと分かり.....

b. 子どもにこの学校を決めた理由

(16) a. 進歩党は幣原を空席の総裁に受け入れると発表した。

b. ボランティア活動の機会の提供として施設にボランティアを受け入れることである。

(14) から (16) の例は、a と b の対は、仮に語順による意味的制約を捨象した場合、構造的には同じ構成要素で成り立っていると言える。(14) で言えば、動詞「焼く」に対して、「が格」、「を格」、「に格」といった要素によって構成されている。これらの例については、次の点に注目してほしい。(14) a で、「に」格は、ビスケット生地を焼いた結果として生じる変化後の状態・様態を表すのに対して、(14) b では、単に焼いたビスケットの移動先・受け手を表している。前節のクラスタで言えば、(14) a はクラスタ 4 の具体例であるのに対して、(14) b

はクラスタ5の具体例である。同様の交替が(15)や(16)のペアからも観察できる。

さて、次の問題として、(14)から(16)でみた様々な交替現象を動機付けるものは何であろうか。果たして、動詞といった単一の要素のみの制約として、これらのデータを説明することは可能であろうか。本論の問題提起が正しいのであれば、以上の交替現象はむしろ動詞の制約とは別の次元で起きているものと捉えるべきであろう。

こうした分析は決して、動詞の制約を否定するものではない。むしろ、以上の考察は構文現象が決して一枚岩ではないことを示すものであり、その動機付けに関しても単一の制約を仮定することは正しくないことを示唆するからである。とりわけ名詞や構文の制約に注目し、表1から問題を捉えてみた場合、次のような非対称的(動機づけの)システムが見えてくる。クラスタ1の場合、場所や具体物といった共起語に制約されつつ、動詞に先行し、主語に後続するという統語環境に制約されていることが分かる。一方、クラスタ2の場合、事や抽象的關係といった共起語に制約されるものの、文頭であること以外には統語環境による制約はそれほど顕著ではない。さらに、クラスタ3の場合、事といった共起語の制約があるものの、主語に後続し、動詞に先行するという統語関係に強く制約されていることが分かる。クラスタ4やクラスタ5に関しても同様の視点から観察した場合、以下のようにまとめることができる。

表3 〈クラスタの非対称性〉

区分	クラスタ1	クラスタ2	クラスタ3	クラスタ4	クラスタ5
FS1	○	○	×	×	○
FS2	○	×	○	○	○

表3は、クラスタ1からクラスタ5が、どの変数セットにより強い制約を受けるかを示している。それによると、三つの動機づけのシステムが見えてくる。まず、一つ目として共起語に強い制約を

受けるクラスとしてクラスタ2、二つ目として統語関係に強い制約を受けるクラスとしてクラスタ3とクラスタ4、最後に、3つ目として共起語と統語関係の両方に制約を受けるクラスタ1とクラスタ5がある。

6. まとめ

以上の考察から「に」に関わる構文現象を動詞の意味的制約を仮定せず、その分布関係を適切に捉えることができることを示した。また、分析に際して多変量解析の手法を用いることで「文脈」に対する定量的分析が可能であること、親疎関係をはじめとする様々な性質を明らかにできることが分かった。

〈参考文献〉

- [1] Croft, William. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.
- [2] Goldberg, Adele. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- [3] Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar, Vol. 1, Theoretical Prerequisites*, Stanford, Calif.: Stanford University Press.
- [4] 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
- [5] 村木新次郎 (1991). 『日本語動詞の諸相』、ひつじ書房.

〈データ集〉

- [6] CD-ROM版 新潮文庫の100冊、新潮社
- [7] 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.