

# 名詞との共起関係に基づく構文の定義\*

李在鎬\* 黒田航\* 大谷直輝† 井佐原均\*

\*(独) 情報通信研究機構 知識創成コミュニケーション研究センター † 京都大学大学院

## 1 はじめに (本研究の目的と主張)

本研究の目的は二点ある。第一に、名詞との共起関係に基づいて日本語の構文 (construction) に実質のある定義を与える。第二に、(語文法を主要な領域とする従来のコーパス言語学に対して) 文文法 (sentence grammar) におけるコーパス利用の具体的な分析手法を提案する。ケーススタディとして移動や結果事象を表す文現象に注目して分析を試みる。

本研究の理論面、記述面での主な主張は次のようにまとめられる。a) 印欧語を中心とする従来の構文文法 (例えば Goldberg 1995, Michaelis and Ruppenhofer 2001, Langacker 2005 など) は (多かれ少なかれ生成言語学の対抗馬として「見かけの被覆率」を上げるために) 語の選択制限のような語彙特性に言及しない、抽象的なスキーマによって現象を説明し、結果的には過度の抽象化を行ってきた。こうした抽象的構文の記述手法は、日本語のようなタイプの言語には応用可能性が低く、汎用的モデルとしては不十分である。b) こうした問題点を解消するためには、(抽象度を抑えた) 下位レベルスキーマ (low-level schema) として構文 (constructions) を再定義する必要がある。具体的には、格パターンと名詞パターンの共起関係による記述が有効である。c) その具体的な方法論としてコーパスデータを用いて、格パターンと名詞パターンの共起の信頼度を統計的指標で検討する手法が有効である。なお、主張 a) の論証として作例に基づいて文の意味解釈に名詞の意味的制約が深く関与していることを示す。主張 b)、c) の論証として、Tスコアに基づくコーパスデータの分析結果を示す。

## 2 先行研究と問題提起

文や句レベルの形式と意味の問題に構文文法 (Construction Grammar) は様々な成果をあげた。とりわけ Fillmore (1988, 1989) の先見的研究に始まり、Goldberg (1995) による斬新な問題提起、さらには Croft (2002) や Tomasello (2003) や Goldberg (2006) による拡張と応用で、理論的完成度を高めることができた。彼らが提案する説明項としての構文 (construction) は構成要素間の有機的結合によって生成された記号の一種であり、表現全体として構成要素に還元できない特定の意味をエンコードするという特徴によって定義さ

\*本稿は日本語認知言語学会第6回全国大会での発表原稿をもとに作成されました。発表の機会を与えてくださった日本語認知言語学会の関係者の皆さんおよび司会の岡智之先生やフロアの皆さんに感謝します。なお、本稿の分析は KH Coder (樋口耕一氏作成: <http://khc.sourceforge.net>) を用いてなされました。この場を借り感謝します。

れる (Fillmore1988, Goldberg1995,2006, Ostman and Fried (eds.) 2005, Langacker 2005, )。構文文法ではこうした構文こそが言語や文法の基本単位であると位置づけており、標準的な見解の一つとして文法構文 (grammatical constructions; Fried and Boas (eds.) 2005) の存在が指摘されている。とりわけ Goldberg による項構造構文の研究では文法内の抽象度の高い形式 (e.g. SVOO) が特定の意味をエンコードすることを現象と理論の両面から示した。

- (1) a. Pat faxed Bill the letter.
- b. Pat sneezed the napkin off the table.
- c. She kissed him unconscious.
- d. The fly buzzed into the room.
- e. Sam kicked at Bill.

(1) の例で注目すべきなのは、いずれの表現においても動詞の意味から全体の意味が予測できない点である。(1b) で言えば、*sneeze* は動詞そのものの意味として使役移動 (caused motion) を含意しないにも関わらず、使役移動構文 (caused motion construction) において生起している。このことは少なくとも動詞の下位範疇化として (1) は説明できないというのが Goldberg の構文文法の基本的な主張であり、概ね了承されているように思う。

以上の議論は英語の構文効果に関するものだが、日本語の構文の認定に英語と同様の議論をもち出すことはできない。最大の難点は、Goldberg 流に「構文は形式と意味の対 (form-meaning pairings) である」(Goldberg 1995:4) と定義した場合、日本語においては「形式」の定義が困難であるという問題がある。この点に関する例を挙げて、具体的に説明しよう。まず日本語における構文効果の存在を示唆する例として、(2) を考えてみよう。

- (2) a. サマータイムレジャーの人々が海外に消えた。
- b. 彼女もこれで主人にケーキを焼いたのだろう。

(2) で注目すべき点は“消える”や“焼く”のような移動や所有とは無関係な動詞が使用されているにも関わらず、全体の文意として移動や所有が喚起されている点である (李 2002, 2004)。このような含意は、明らかに動詞の下位範疇化による分析モデル (例えば Levin1993 や影山 1996 で提案されている動詞クラスモデル) では自然な分析が困難である。

次に問題となるのは、こうした構文の意味が何によって成立しているかである。この問題に対して本稿は名詞の制約が深く関与していると考える。(2b) を例にするなら、二格名詞が [+animate] という特徴をもつことが構文の意味の実現に強く関与している。この主張の論証として (3) を挙げる。

- (3) a. 彼女は旅行の前日にケーキを焼いた。
- b. 彼女は息抜きにケーキを焼いた。
- c. 彼女はドラ焼きサイズにケーキを焼いた。
- d. ??彼女は新品の皿にケーキを焼いた。(cf. 彼女は新品の皿でケーキを焼いた)

(3) では抽象物や事や抽象的關係や具体物を表す名詞を使って、(2b) の最小差対を作った。これらはそのいずれも移動や所有を文意として含意しないことを確認しておきたい。ただ、(3) における「に」の生起を単なる付加詞用法とし、コアから外す考え方もありえるが、これはこれで別の問題を派生し、本質的な解決にはならない。実際問題として、(3) が付加詞であるという立場において、(4) のような例はどう分析すれば良いのであろうか。

- (4) a. 太郎が神棚におみやげをあげた。
- b. 太郎が友達におみやげをあげた。

(4) で興味深いのは、(4a) 移動として解釈されるのに対して、(4b) では所有の意味が生じている点である。つまり、項構造構文の議論で言えば (4a) が使役移動構文のインスタンスとして解釈されるのに対して、(4b) は二重目的語構文のインスタンスとして解釈される。さて、ここで問題にすべきは (4) における解釈の相違を何に求めるべきかということにある。この問題に対して、経験科学のレベルでもっとも自然な一般化をするのであれば、それは動詞によるものでもなければ、統語形式によるものでもなく、名詞の意味クラスによる制約と考えるのが自然である。

以上のことから、日本語における構文の特徴づけに共起名詞の意味制約を取り入れる必要があることが示唆された。だが、問題は (名詞を構文の定義に含めた場合) 十分な記述のために必要となる意味的制約の具体性のレベルである。

### 3 代替案と方法論の紹介

#### 3.1 名詞の記述をめぐる技術的問題

前節の問題提起で名詞の貢献度を評価する必要性が示唆された。しかし、これが理論的に正しいとしても、どのような形で実際の分析に取り入れ、現象を記述していくべきかという点をめぐっては、いくつか技術的な問題が存在する。それは (??) のようにまとめることができる。

- (5) a. 名詞の数え上げによる分析のディテールの問題
- b. 観察のバイアスと偶然率の問題

本稿の分析モデルの直感的印象として、(5a) の名詞のディテールの問題を解消する必要がある。というのは、名詞というカテゴリーが持つ特徴として、実際の使用に対する必要十分な定義が困難なことと、有限無数個のディテールが存在する。となると、名詞を概念レベルで定義を与え、その一つ一つを記述していくのは不可能とは言わないまでも現実的でないことは明らかと言えよう。さらには、統語パターンとの共起例を一つ一つ検討するのは、困難を極めることである。このことから名詞と統語パターンの共起を数え上げ、個別に記述していくことはあまり現実的ではない。同時に (5b) の問題として数ある名詞の中で、ある特定の名詞の制約を取り入れたとしてもそれが日本語の事実を必要十分に反映し、代表性の条件を満足するものだとは限らない。たまたま見つかった事例に対して、構文としての位置づけを与えたとしてもそれは意味ある記述にはならない。

こうした2点の記述的難問に対し、コーパスベースの記述モデルで問題の解決を図る。具体的には、3点の方法論を導入する。

- (6) a. コーパスからサンプルを収集する
- b. データベース型のシソーラスに準拠して個々の名詞の意味クラスを定義する
- c. 統計的指標でサンプルを評価する

(6a) と (6b) によって、(5a) の問題を解消し、(6a) と (6c) によって (5b) の問題を解消することを旨とする。

### 3.2 使用データ

まず (6a) の実践として、『新潮文庫 100 冊』の小説テキスト 6 冊分からそれぞれ冒頭 5000 文字を抽出し、3 万字のミニコーパスを作成した。これを茶せんで形態素解析した。総抽出語は異なりで 3066 語、延べで 19010 語あった。これに対して KH coder で「に」を含む単文を KWIC 検索をした。得られたサンプル数は、全部で 573 例、一冊平均で 95 例である。詳細を表 1 に示す。

区分	サンプル数
宮本輝『錦繡』	102 例
塩野七生『コンスタンティノーブルの陥落』	96 例
五木寛之『風に吹かれて』	93 例
村上春樹『世界の終りとハードボイルド・ワンダーランド』	94 例
椎名誠『新橋烏森口青春篇』	95 例
赤川次郎『女社長に乾杯』	93 例

表 1: 使用データ一覧

なお、紙幅の都合上詳細は述べないが、表 1 のデータについては、Lee and Otani (2005) を参照されたい。

### 3.3 分析方法

次に (6b) の実践すべく、(7) の作業を行った。

- (7) a. 名詞の意味クラスを特定した (e.g., [生物]、[具体物]、[抽象物] など)
- b. 意味クラスをパターン化した (e.g., [生物—具体物])

具体的には (8) に示す作業を行った

- (8) a. 母が父に会った。 [生物—生物]
- b. 母が公民館に行った。 [生物—場所]

なお、名詞の意味クラスの判断においては、主観の混入を可能な限り排除すべく、「日本語語彙大系」に準拠して行った (Lee and Otani 2005)。(7a)の対象となる名詞は、全部で2109個あった。(7b)で名詞のパターンを特定した結果、84タイプが見つかった。次に統語パターンと名詞パターンの共起数を計上していった。その一例を表2に示す。

統語パターン	名詞パターン	共起数
XがYに	生物—主体	19
XがYに	生物—場所	50
XがYに	生物—事	28
⋮	⋮	⋮
XがZをYに	生物—抽象的關係—場所	13
XがZをYに	生物—抽象的關係—抽象的關係	10
⋮	⋮	⋮

表2: データ解析見本

次に、(6c)の実践としてのT-score法で共起の強さを測定した。計算式は以下の通りです。

$$\frac{\text{共起数} - \frac{\text{統語パターン} \times \text{名詞パターン}}{\text{総数}}}{\sqrt{\text{共起数}}} \quad (1)$$

*t*-scoreは、*t*検定の手法を応用して、二つの単語の共起関係(相互の結びつき)の強さ(の確信度)を計る指標の一つである。他の統計的指標に比べ、単語の頻度も考慮に入れており、単語の頻度が少ない場合でも適切に判断できるといわれている(斎藤(編)1998, Gries 2003)。本稿のデータが小規模であることを考慮した場合、この手法は好都合と言えよう。本稿では個々の共起スコアは慣習化の度合の指標であると位置づける。

## 4 結果と考察

本節では、前節の方法論を用いて、コーパスデータを解析した結果を報告する。4.1節では、二項関係(e.g.,「XがYに」「YにXが」)における共起度の上位10項目を報告し、4.2節では、三項関係(e.g.,「XがYにZを」「XがZをYに」)における上位10項目の共起度を報告する。4.3節では、本稿のデータ解析結果を先行研究の分析に付き合わせながらその位置づけをめぐる考察を行う。

### 4.1 二項関係に見られる共起度

二つの項による構文の共起関係を計り、以下の結果を得た。

共起パターン	共起頻度	t-score
a. 生物が 場所二	50	4.0437
b. 物が 抽象的關係二	30	2.4563
c. 生物が 事二	28	2.2628
d. 生物が 抽象的關係二	28	2.2506
e. 場所二 物が	19	1.354
f. 生物が 主体二	19	1.338
g. 物が 事二	17	1.1104
h. 物が 場所二	15	0.852
i. 生物が 具体物二	14	0.715
j. 物が 抽象物二	13	0.5963

表 3: 二項関係における共起度

表 3 の共起パターンに対応する具体例を (9) に示す。

- (9)
- a. ぼくが百貨店ニュース社に入ったのはまったくの偶然だった。
  - b. 六世紀半ば、東ローマ帝国の勢力圏が最大に達したのである。
  - c. 最近の会社の営業活動について彼は (= が) かなりの早口に喋りだした。
  - d. 当時二十五歳だった私も (= が) 三十五になりましたが、
  - e. 暗い空に巨大なお化け煙突の影 (= が) 見えた。
  - f. 山本は (= が)、今日も純子に会えるという、それだけを楽しみに、会社へ向かう
  - g. 一都市の陥落が一国家の滅亡につながる例は…。
  - h. ビザンチン領であるマケドニアも (= が)、属国になって年貢金と…。
  - i. 男は (= が) 電話に出た。
  - j. 機械の故障や係員の不注意が現実に取り得ないと言っているわけではない

「X が Y に」でもっとも安定しているのは、生物が場所に何らかの動作（特に移動）を行う、という経験のスキーマを具体化する場合である。次に続くものとして、無生物が抽象的な状態に陥るといった経験のスキーマを具体化する場合である。

一般に t-スコアは基準値 2 を満たない場合は、偶然である可能性が高く、統計的には意味をなさないものと解釈される（斎藤（編）1998: 140）。これに照らし合わせて考えてみた場合、二項関係において慣習化された構文は a から d までと結論づけることができる。

これらの事実関係は従来の枠組みで分析した場合、a は自動移動構文であり、b は自動詞結果構文、c は様態構文、d は自動詞結果構文に該当する。この対応関係の理論的位置づけの詳細は、4.3 節で考察するが、この結果が偶然でない限り、英語研究で言われてきた項構造構文が日本語においても独立に定義可能であることが確認できたであろう。

## 4.2 三項関係に見られる共起度

三つの項による構文現象の共起関係を計り、表4の結果を得た。具体例は(10)に示す。

共起パターン	頻度	t-score
a. [生物ガ][場所二][具体物ヲ]	12	2.46
b. 生物ガ 具体物ヲ 場所二	10	2.159
c. 生物ガ 具体物二 具体物ヲ	10	2.158
d. 生物ガ 具体物ヲ 具体物二	8	1.825
e. 生物ガ 具体物ヲ 抽象的關係二	6	1.447
f. 生物ガ 抽象的關係二 具体物ヲ	6	1.446
g. 生物ガ 主体二 具体物ヲ	5	1.234
h. 生物ガ 主体二 抽象的關係ヲ	4	0.998
i. 生物ガ 抽象的關係二 事ヲ	4	0.998
j. 生物ガ 場所二 場所ヲ	3	0.73

表 4: 三項関係における共起度

- (10)
- a. (私が) 壁に粘土をなげつけたときのような扁平な音が聞こえた。
  - b. 先を回って、薬を箱に補給したり、集金したりする帳面である。
  - c. 山本はせっせと髪にクシを入れ、出社の支度を急いでいた。
  - d. 父は経済紙を手に取りながら、…。
  - e. (私が) 出っ張りを削り、真ん丸な顔を長目にこね直し、
  - f. 荒井は嬉しげに茶碗を妻の方へ差し出しながら、…。
  - g. 私は体の不自由な自分の息子に、清涼な山のたたずまいやたくさんの星々を見せてやりたくなりました。
  - h. あなたはやがて何気なく私に視線を向け、
  - i. 動けないとなると途端に痲癩を起こして、
  - j. 高根は上野の近くの自宅敷地に蛇舎を設け、四百匹近くの蛇を飼っており、

この結果から、日本語の三項関係において、もっとも安定しているのは、ある生物 X がある場所や具体物の Y に別の具体物 Z を移動させる経験のスキーマを具体化する場合である。言い換えれば使役移動構文としての使用がもっとも顕著と言える。ただ、本稿が用いたデータにおいては e や f に見られる結果構文、g や h に見られる二重目的語構文としての使用には統計的有意差は見られなかった。また興味深い点として a と b、c と d はいずれも語順交替によるものである。この点をめぐる詳細な考察として中本・李・黒田(近刊)を参照されたい。

三項関係についてはケースの絶対量として非常に少ない故、あまり信頼できる結果とは言いがたいが、数値としては a から c までのパターンの定着度が確認できる。

### 4.3 項構造構文に対する示唆

以上の結果について次の点を強調しておきたい。それは、二項関係で顕著だったように統計的な有意が認められるレベルで現象を切り取った場合、上位にランクするものはいずれも項構造構文として指摘された現象の典型的なパターンである。これによって(日本語においても)項構造構文の実在が確認された。具体例として、「XガYニVする」パターンと名詞の共起を考えてみよう。

- (11) a. あなたはまだ私に気づかず、… → 生物ガ主体二  
b. じゃ、もう私は会社に行かなくていいんだ。 → 生物ガ場所二  
c. 私は山本と違って、バスに乗る前に… → 生物ガ具体物二  
d. 私がその問題にかかわらねばならないことが、… → 生物ガ抽象物二  
e. 彼は我々との意図不明な共同生活に加わった。 → 生物ガ事二  
f. 純子が学生時代に会得したもので、… → 生物ガ抽象的關係二

(11)は主語名詞が生物のデータであるが、本稿の分析はこれらの非対称性をうまく数値化している。例えば、(11a)と(11b)を取り上げてみた場合、前者の共起指数が1.338で基準値2を満たないのに対して、後者の共起指数は4.0437で、慣習化された構文として位置づけられる。このことを項構造構文との関連で見直した場合、(11b)は李(2002)が指摘した自動移動構文の具体例であるが、(11a)についてはそのような特徴づけが難しい。

このことは、日本語の文法構文を統語パターンと名パターンの共起に基づいて定義することの妥当性を示唆する。つまり文法構文を単純に統語形式のみで定義するのではなく、名詞の意味クラスに関する情報を含めて捉えるべきであるという本稿の主張をサポートする。このことを2節の問題提起に関連づけて考えてみた場合、重要な理論的含意を持つ。従来の構文研究(例えばKay(2002)などが提案するUnificationベースの構文文法など)は統語カテゴリーのような極めて抽象的な形式のクラスで構文を定義し、記述のカバー率をあげることに主眼が置かれてきた。しかし、本稿のアプローチはそれとは反対の方向を目指す。その理由として2節の事実が示す通り、日本語の分析では抽象的な構文の定義法は有効ではない。本稿は名詞の意味クラスまでも含めた、下位レベルスキーマ(low-level schema)として文法構文を定義したとき初めて見えてくる事実があることを指摘する。それと同時に、本稿の提案するアプローチは従来は十分に考慮されて来なかった選択制限と構文効果との関係を明示化する。

こうすることの理論的利点として、Goldbergを中心とする従来の構文文法が、抽象的・理論的存在物として特徴づけてきた文法構文の存在を、よりグラウンド化され、身体化されたレベルで特徴づけることができ、研究プログラムとしての一貫性(例えば、身体基盤主義に基づく言語分析)を保持できる。

## 5 まとめ

本研究では、従来の構文文法を批判的観点から検討し、名詞の意味クラスと統語パターンの共起関係による分析モデルを提示した。その際、発生する記述的ディテールの問題に対して、コーパスデータに対する統計的指標を用いることで解決できることを示した。

また、解析結果が示唆する事実として日本語の項構造構文は名詞と統語パターンがシステムティックに連動していることが明らかになった。これによって、従来の(名詞の貢献度を捨象して議論していた)項構造構文の定義法は言語事実に対する不十分な記述であることが示され、名詞の意味クラスを構文の定義に含めなければならないことが示された。

## 参考文献

- [1] Croft, William. (2002). *Radical Construction Grammar*. Oxford: Oxford University Press.
- [2] Fillmore, Charles J. (1985). "Syntactic Intrusions and the Notion of Grammatical Construction". *BLS* 11. pp. 73-86.
- [3] Fillmore, Charles J. (1988). "The mechanisms of Construction Grammar". *BLS* 14, pp. 35-55.
- [4] Fillmore, Charles J. (1989). 生成構造文法による日本語の分析一試案、柴谷方良・久野すすむ編『日本語学の新展開』、東京: くろしお出版, pp. 11-28.
- [5] Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- [6] Goldberg, A. E. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13 4, pp. 327-356.
- [7] Gries, Stefan Th. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1, pp. 1-27.
- [8] Hunston, S. and Francis, G. (1999). *Pattern Grammar: A Corpus Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- [9] 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.
- [10] Kay, P. (1997). *Words and the Grammar of Context*. Stanford, CA: CSLI Publications.
- [11] Kay, P. (2002). "An Informal sketch of a formal Architecture for Construction Grammar". *Grammars* 5, pp. 1-19.
- [12] 黒田 航 (未刊). 日本語の Pattern Matching Analysis の簡単な実例. [<http://cls1.hi.h.kyoto-u.ac.jp/~kuroda/papers/sample-pma-of-japanese.pdf> として入手可能]
- [13] Langacker, R. W. (1987). *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- [14] 李在鎬 (2004). 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. 『認知言語学論考 No. 3』. ひつじ書房, pp. 183-262
- [15] LEE, J.-H. and Otani, N. (2005). "Toward a description of its grammatical phenomena based on a cluster analysis" *International Cognitive Linguistics Conference 9*, pp. xx-yy. (Seoul, Korea).
- [16] Michaelis, L. A. and Ruppenhofer, J. (2001). *Beyond Alternations: A Constructional Model of the German Applicative Pattern*. Stanford: CSLI Publications.
- [17] 尾上 圭介 (2001). 『文法と意味 I』. くろしお出版.

- [18] 中本 敬子・李 在鎬・黒田 航 (近刊). 語順は超語彙的意味と相関する: 使役所有/移動を例とした心理実験と日本語の構文効果研究への提案. 『認知科学』
- [19] 斎藤 俊雄・中村 純作・赤野 一郎 (編) (1998). 『英語コーパス言語学—基礎と実践』. 研究社.
- [20] Stefanowitsch, A. (2003). A construction-based approach to indirect speech acts. In K.-U. Panther and L. Thornburg (eds), *Metonymy and Pragmatic Inferencing*. Amsterdam and Philadelphia: Benjamins, pp. 105–126
- [21] 「CD-ROM 版 新潮文庫の 100 冊」. 新潮社.
- [22] NTT コミュニケーション科学基礎研究所 (199x). 『日本語語彙大系 CD-ROM 版』. 岩波書店.