

# PLM と FCA による構文ネットワークの記述について

長谷部 陽一郎  
同志社大学\*

浅尾 仁彦  
ニューヨーク州立大学 [院] / 京都大学 [院]†

## 1 はじめに

認知言語学では構文は互いにあるいは内部的に関係し合い、ネットワークを構成すると考えられている (Goldberg 1995; 2006, Newman 1996, Mukherjee 2005)。しかし実際のところ、構文ネットワークとはどのような性質のものなのか必ずしも明確な共通理解は得られていない。構文ネットワークの構造を、単なる「イメージ」としてでなく、何らかの厳密な規定のもとに記述する方法を確立する必要がある。そこで本稿では構文という概念を、認知文法 (Cognitive Grammar) や語文法 (Word Grammar) などの認知言語学的理論との互換性を保ちながら再定義し、それに基づいた構文ネットワーク構造の明示化を試みる。またその中で、パターン束モデル (Pattern Lattice Model, PLM) と形式概念分析 (Formal Concept Analysis, FCA) という2つのモデルが有用であることを示す。

## 2 概念のネットワーク

構文に関する問題に取り組むに先だち、まずは概念一般のネットワーク構造について明確にしておきたい。認知文法では言語が音韻極 (Phonological Pole) と意味極 (Semantic Pole) という2つの極により構成されるとしているが (Langacker 1991, 2009)、ここで考えるのは意味極側の問題である。言語的概念にはモノあるいは関係という2つのカテゴリーが存在し<sup>1)</sup>、一般的に前者は名詞や名詞句によって表され、後者はそれ以外の語 (前置詞、動詞、形容詞、副詞など) や句で表される。

語よりも大きい単位の表現に対応する概念は、これらの要素が何らかの関係により結ばれることで成立する。ただしそのような関係もまた一種の概念であり、またモノが内部に関係概念を含むこともある。したがって、認知言語学における言語的概念とは一種の再帰的 / 自己参照的な構造として、内部に様々な度合いの複雑性を持つ。しかし基本的にはモノ的であるか関係的であるかという点について二分法的な判断が可能であり、それぞれのカテゴリーに属する概念は、話者の知識体系の一部として is-a 関係に基づくネットワーク構造の中に位置付けられる<sup>2)</sup>。例えば、英語におけるモノ的概念 dog は図1のようなネットワークを背景を持つ。

認知文法における schema network は is-a 関係に基づくネットワークであり、語文法においてもそれぞれの語が同様の構造を (taxonomy として) 背後に持つと規定

している。このような階層構造はモノ的であるか関係的であるかにかかわらず存在する概念ネットワークの基本的特徴であり、本稿の枠組みでも同様の想定を行う。

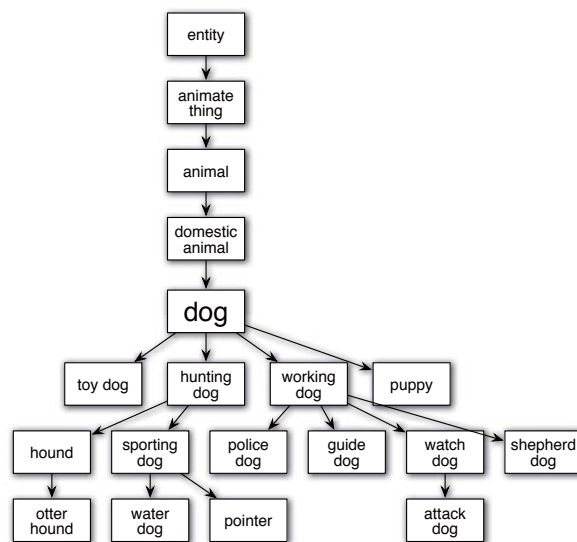


図1 モノ概念 (dog) の階層ネットワーク

## 3 概念としての構文

以上のような理論的基盤のもとに、構文および構文のネットワークという概念をどのように規定すべきか考えてみたい。まず、モノ的あるいは関係的という概念の二分法において、構文はそれ自体いずれか一方に属するという性質のものではない。従来の構文研究の中で論じられてきた構文の多く—例えば二重目的語構文や移動使役構文—は明らかに関係的概念を形成するタイプの構文である。一方で、(1) のような所有格名詞句にもある種の構文が関わっている。ここでの [X's Y] というパターンはモノ的概念を形成するタイプの構文である。

- (1) a. John's book  
b. John's hometown  
c. John's comment  
d. John's death

ならば構文とは結局のところ単体としての語と同じで、違いは単に複数要素から構成されている点だけにあると考えて良いのだろうか? 近年の構文研究では確かにそのような見方も散見される。しかし、構文が語や形態素単体の持つ概念構造と決定的に異なるのは、それが一種の関数的性質を有している点にある。関数的である

\* yhasebe@mail.doshisha.ac.jp

† asaokitan@gmail.com

とはつまり項 (argument) を取り、何らかの値 (value) を返す働きをするということである。例えば [X's Y] という所有格名詞句構文であれば、X と Y というモノ型 (type) の2項を取り、同じくモノ型の値を返す関数的機能がみとめられる。

構文文法や語文法ではもともと関数という考え方が全面的に採用されている (Goldberg 1995, Hudson 2010)。一方、認知文法においてそれは理論的枠組みと合致しないように見えるかもしれない。しかし認知文法における e-site の具体化 (elaboration) による結合構造形成のプロセスなどは一種の関数適用に他ならない<sup>3)</sup>。また、ここで言う関数は数学的あるいは計算機的な文脈でのそれと異なり、一定の制限のもとに異なるプロトタイプ性を持つ要素を項として取ることが許される。[X's Y] 構文の事例を示した上の例でも、物理的な所有者-所有物の関係にある2要素を項とした (1a) は、(1b) - (1d) と比較してよりプロトタイプ的であると言える。

前節では、モノであれ関係であれ、概念は is-a に基づく階層的ネットワークの中に位置付けられることをみた。構文も同様であり、構文の形式に対応する概念はネットワークとして形成される。ここで注意すべきは、「概念の階層的ネットワークの中である形式に対応付けられる部分はある特定のノード以下に限られる」という (自明の) 事実である。例えばモノ概念 dog の概念ネットワークを示した図1において、dog という言語形式と対応付けられるのは中心に位置している dog ノード以下に限られる。それよりも上位に位置するノードの概念は animal などの表現によって表される必要がある<sup>4)</sup>。

同様のことが構文についても言える。二重目的語構文 [NP1 V NP2 NP3] は項となる名詞句の意味的性質により、またそれらの組み合わせにより、様々な下位構文概念を持つ。それらは単なる意味的バリエーションに過ぎないという見方も有り得るが、多くの場合そこにはある種の傾向が見られ、それぞれがパターンとしての際立ちを示す。例えば二重目的語構文で NP1 は生物要素であることが多い。しかし NP1 に非生物要素が来ることもある。そのようなときには高確率で NP3 は物質的な存在物ではなく、より抽象的なイベント (event) ないしは状態 (state) を表す概念が来る。各構文にはこのような意味パターンが複数存在し、is-a に基づく階層関係で結ばれていると考えられる。

ここで問題となるのが、いかにすれば構文の下位概念ネットワークを実際に記述できるかということである。単体としての語の概念ネットワークに関しては、これまでも WordNet プロジェクト (Fellbaum 1998) を始め複数の試みがなされてきた。構文の下位概念ネットワークについても、二重目的語構文についての Mukherjee (2005) や Newman (1996)、結果構文についての Boas (2003) など個別の研究はあるが、特定の事象を越えて体系化された方法論は確立されていない。そこで次節ではこの問題に対処するため、PLM と FCA の手法を導入することを提案する。

#### 4 構文ネットワークの記述

ここまで論じたように構文は一種の関数的概念であり、項を取って一定の出力を行う。項には基本的な型制約と意味制約のもとに多少の自由度があり、各項がどのように実現されるか、またどのように組み合わせられるかにより、下位構文の概念パターンが決定される。理論的にはシンプルであるが、実際には可能なパターンを1つ1つ手作業で導き出していくことは現実的に難しい。FCA を用いることで、これを半自動的に行うことが可能になる。FCA は Rudolf Wille らによって提唱された、階層的カテゴリ分析の手法であり (Ganter et al. 1999 2005)、一定のアルゴリズムに基づいて、2次元のマトリクス形式にまとめられたオブジェクトと属性の束から問題領域における概念 (concept) の集合を取り出し、それらの相互関係を明示化する<sup>5)</sup>。

具体的には、まず図2のような2次元テーブル上でオブジェクトと属性をそれぞれ行と列に対応付けて記述する。このようなテーブルをコンテキスト表というが、FCA ではコンテキスト表から得られるオブジェクト集合すなわち外延 (extension) と属性集合すなわち内包 (intension) の組を「概念」とみなす。そしてこのような概念を網羅的に抽出して階層関係に落とし込むことが計算プロセスとしての形式概念分析である。図2はいくつかの鳥類のカテゴリーを問題領域としたコンテキスト表であるが、これを形式概念分析にかけると、図3のような完備束 (complete lattice) と呼ばれるグラフ構造が得られる。

	飛行可	水泳可	捕食行動	家禽・ペット	産卵
ペンギン		x	x		x
ワシ	x		x		x
カナリア	x			x	x
ニワトリ				x	x

図2 コンテキスト表

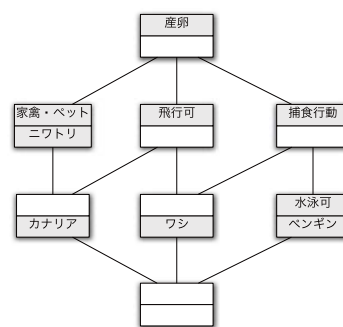


図3 完備束

FCA における完備束において最上位ノードに位置付けられるのは、問題領域のすべてのオブジェクトを含む概念である。一方、底のノードに位置付けられるのは問題領域に存在するすべての属性を含む概念である。これら2つを両極として、その中間階層にすべての可能な概

念が配置される。図3ではFCAの慣例に従って表示の縮約を行っているためやや分かりづらいが、最上位ノードの下側ボックスにはペンギン、ワシ、カナリア、ニワトリというすべてのオブジェクトが含まれている。一方底ノードの上部ボックスには飛行可、水泳可、捕食行動、家禽・ペット、産卵という属性がすべて含まれている。このようにして、FCAでは問題領域内に存在する概念をすべて列挙し、その分布と相互関係を図式的に表示する。

この手法を用いることによって、構文の概念ネットワーク(=下位構文ネットワーク)を実際の事例の中から抽出し、グラフとして可視化することができる。先に述べた通り、異なる性質を持った複数の項を取る構文のあらゆる可能性を手作業で抽出したり関連づけたりすることは難しいが、FCAは数学的なアルゴリズムとして規定されているため、オブジェクトと属性に関するデータさえ用意すれば、それに応じた結果を自動的に得ることができる<sup>6)</sup>。

	NP1 生物	NP1 非生物	NP2-NP3 生物-情報	NP2-NP3 生物-物質	NP2-NP3 生物-抽象物	NP2-NP3 生物-事態	NP2-NP3 非生物-事態
give	x	x	x	x	x	x	x
tell	x	x	x			x	
send	x		x	x			
show	x	x	x		x		
get	x			x			
ask	x		x		x		
cost		x		x			
buy	x			x			
wish	x					x	

図4 二重目的語構文のコンテキスト表

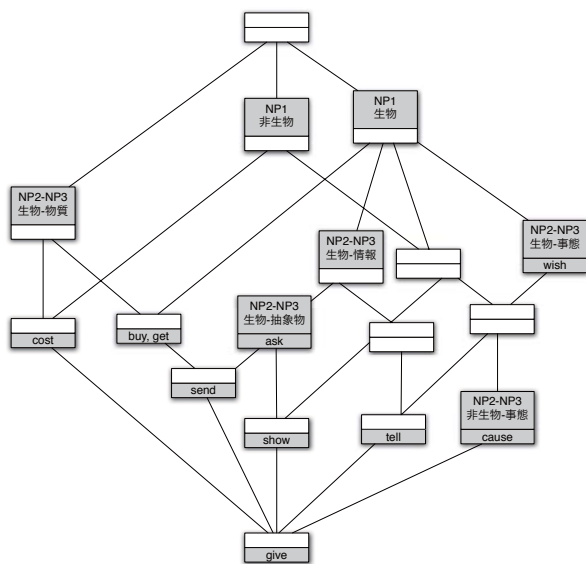


図5 二重目的語構文の概念ネットワーク

このような想定に基づいて長谷部(2010)はFCAの手法を用いてコーパスデータから二重目的語構文の下位構文抽出を行い、ネットワーク構造として可視化した。そこではICE-GBコーパスから得られた二重目的語構文の事例588件を対象に、出現した動詞の基本形をオブジェクトとして、実例において見出された意味パターン

を属性として形式概念分析が行われた。

オブジェクトの数(=動詞の異なり語数)は47、属性の数は18で、得られたグラフはかなり複雑なものであった。そのためここで詳細を示すことはできないが、FCAを用いた構文ネットワーク記述の方法を示すという本稿の目的に合わせ、ここでは代表的な10の二重目的語動詞と大幅に単純化された7つの意味パターン属性を用いたコンテキスト表(図4)を示す。これを形式分析にかけると、図5のような概念ネットワークが得られる。

以上のように、FCAを用いることで従来よりも厳密かつ一般化された方法で構文ネットワークの記述を行うことが可能になる。加えて、FCAを通じて理論的研究の成果をコーパスなどから得られた事例に適用することで説明力・記述力を測り、さらなる理論の修正や発展につなげられるという利点も期待できる。

## 5 構文内ネットワークと構文間ネットワーク

ここまでの議論は基本的に、ある構文が内部にどのような概念ネットワークを含み持つかに関するものであった。しかし、広い意味での構文ネットワークを考えるにあたっては、異なる構文間の関係やある構文が包摂する部分構文についても検討する必要がある。前者については例えば二重目的語構文とto与格構文との関係があり、後者の例としては二重目的語構文と単純他動詞構文との関係がある。これらの問題を射程に入れるには意味的・概念的なネットワークに加え、形式面でのネットワーク構造を明らかにしなければならない。

認知言語学において構文は「必ずしも個々の構成要素から予測されない全体としての(創発的)性質を持った言語単位」として広く理解されている。このこと自体に問題はない。しかし多くの場合、構文は要素の複合的構造(compositional structure)としての性質を併せ持っている。また、各部分構造が一種の構文としての役割を果たしていることが有り得る。そこで、異なる構文間あるいは包摂関係にある構文間においてどのような部分パターンが共有されているかを見極め、必要に応じてそれらのある種の構文として認定していく作業が重要となる。

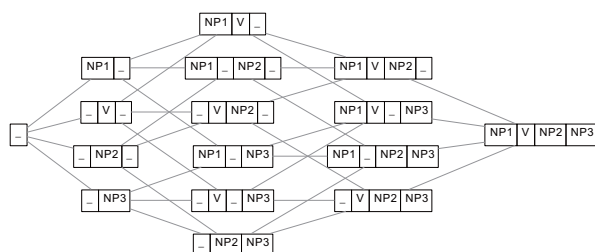


図6 二重目的語構文の部分パターン

このような目的において、Kuroda(2009)や吉川(2010)などで提案・考察されているPLMの手法が有用である。理論的詳細については上記の研究に譲るが、PLMを用いることで、パターンの構成要素とその組み合わせを階層的に表示し、互いの関係や頻度・定着度の度合いを可視化することができる。例えば、二重目的語構文を構成

するパターン [NP1 V NP2 NP3] を PLM を用いて表示すると図 6 のようになる。

図 6 において各ノード中のボックスは形式上の分節(語, 形態素, etc.) を表しており, アンダースコア ( ) はパターン中の変項を意味する。アンダースコア以外の部分は実際には何らかの具体的形式を与えられた定項であるが, ここでは NP や V といった表示により一般化して表示している。すなわち, 図 6 の最右部ノードのパターン [NP1 V NP2 NP3] は実際には

- (2) a. John gave mary a book.  
b. John told her the truth.

といった構文事例に対応する。変項を含んだパターンはこのような事例から経験的に得られる部分パターンであり, 各部分パターンに対応する事例との遭遇あるいは生産を繰り返すことにより, 次第に話者の中で各ノードに異なる重みが与えられ, 構文としての形式パターンが定着していくと考えられる。

## 6 まとめ

本稿では認知言語学的な枠組みの中で構文の構造をより体系的に記述するために 2 つの提案を行った。1 つは概念ネットワークの考え方を敷衍し, 構文を語彙的要素と同様の言語的概念として捉えることである。ただし構文には様々な性質の項を取り, その組み合わせに応じて異なる値を返すという特徴がある。この点については語彙的な概念と明確に差別化される。もう 1 つの提案は, 構文に関する事象をその時々でのやり方で記述するのではなく, より厳密に規定された方法で行うために PLM および FCA の手法を用いることである。

これらの手法を用いた構文研究については未だ蓄積も少なく, 今回の限られた紙面では論じられなかった課題や問題も存在する。しかし, 認知言語学における構文研究はこれまで以上の関心と期待を集める一方で, 方法論上の限界について指摘されることも少なくない (e.g., Bod 2009)。本稿はこのような観点から, PLM や FCA という新たな道具立てを導入することで従来の研究を活かしつつ新たな展開が可能なることを示すものである。

## 謝辞

本稿の執筆にあたっては, その元となった日本認知言語学会大 11 回全国大会予稿集原稿の段階から, 黒田航氏と吉川正人氏に多くの有益なコメントを頂いた。この場を借りて感謝の意を表したい。なお, 本稿で示された研究の一部は日本学術振興会科学研究費補助金若手研究 B (課題番号: 21720179) の助成を受けて行われた。

## 注

<sup>1)</sup> 認知文法ではモノ的概念を thing, 関係的概念を relation と称している (Langacker 1991, 2009)。語文法ではこれらを entity concept および relational concept と呼んでいる (Hudson 2007, 2010)。

<sup>2)</sup> Hudson (2010) はこのような is-a 関係を primitive relation の 1 つとして, より具体的な conceptual relation と区別している。

<sup>3)</sup> Langacker 自身, semantic function という表現を (機能ではなく関

数という意味で) 用いることがある。例えば認知文法におけるグラウンディング (grounding) は命題的概念を発話の「いま・ここ」に接地させるための semantic function であると定義されている (Langacker 1991: 549)。

<sup>4)</sup> 図 1 は WordNet 3.0 から得られた dog に関する synset データに基づいている。

<sup>5)</sup> FCA はもともと言語研究のために開発されたものではなく, それ自体は汎用的なデータ分析手法である。自然科学や計算機科学といった分野の他, 近年では社会科学などの分野でも利用されている。

<sup>6)</sup> FCA を用いても, オブジェクトと属性のセットを決定する作業を自動化するのは難しい。この作業には事例についての十分なりサーチと理解が不可欠である。逆に言えば, FCA のような計算的手法を用いる場合でも, これまでに積み重ねられてきた理論的研究は変わらぬ価値を持つということである。

## 参考文献

- Boas, H. C. 2003. *A Constructional Approach to Resultatives*. Stanford: CSLI.
- Bod, R. 2009. "Constructions at work or at rest?" *Cognitive Linguistics*, Vol. 20, No. 1, 129-134.
- Fellbaum, C. ed. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Ganter, B, G. Stumme, and R. Wille. eds. 2005. *Formal Concept Analysis: Foundations and Applications*. Berlin: Springer.
- Ganter, B and R. Wille. 1999. *Formal Concept Analysis: Mathematical Foundations*. Berlin: Springer.
- Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*.
- Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- 長谷部陽一郎. 2010. 「構文のネットワークモデルについて—二重目的語構文を中心に」, 山梨 正明他 (編) 『認知言語学論考 No.9』, 東京: ひつじ書房 81-137.
- Hudson, R. 2007. *Language Networks: The New Word Grammar*. Oxford: Oxford University Press.
- Hudson, R. 2010. *An Introduction to Word Grammar*. Cambridge: Cambridge University Press.
- Kuroda, K. 2009. "Pattern lattice as a model for human linguistic knowledge and performance," in *Proceedings of PACLIC 23: The 23rd Pacific Asia Conference on Language, Information and Computation*, 278-287.
- Langacker, R. W. 1991. *Foundations of Cognitive Grammar, vol 2, Descriptive Application*. Stanford: Stanford University Press.
- Langacker, R. W. 2009. *Investigations in Cognitive Grammar*. Berlin: Walter de Gruyter.
- Mukherjee, J. 2005. *English Ditransitive Verbs: Aspects of Theory, Description and a Usage-based Model*. Amsterdam: Rodopi.
- Newman, J. 1996. *Give: A Cognitive Linguistic Study*. Berlin: Mouton de Gruyter.
- 吉川正人. 2010. 「パターンの生産性に見る統語発達: パターン束モデルに基づく習得プロセスの検証」, 『日本認知科学会第 27 回大会発表論文集』 235-241.