

生成語彙を用いた日本語辞書における形容詞の取り扱いについて

橋本喜代太* 松本裕治*

聖和大学人文学部* 奈良先端科学技術大学院大学情報科学研究科*

{kiyota-h, matsu}@is.aist-nara.ac.jp

1 はじめに

言語学・計算言語学における意味の扱いは、従来の真理条件意味論等による「文の意味」から「語の意味」に焦点が移ってきていると言える。R. Langacker らによる認知文法([1], [2]など)や Jackendoff らによる語彙概念構造([3]など)はその流れを汲むものと捉えることができ、特に前者は語の曖昧性についてプロトタイプ、イメージ・スキーマの操作など、単に語の意味を並べるのではなく、動的に意味の関連を付ける試みとして注目できるが、定式化という点で問題がある。これらに対して、Pustejovsky らによる生成語彙([4]など)は語の曖昧性を動的に捉える枠組の構築を目的とし、明示的な生成規則を持つ点で計算機による処理にも適している。

本稿では生成語彙の枠組を用いて日本語辞書を作成する際の形容詞の扱いについて考察する。生成語彙では形容詞を扱う上で選択束縛という機構が用意されているが、それだけでは十分でなく、問題点もある。本発表ではまず日本ごん形容詞の特徴を概観したあと、これまで既に提案されている生成語彙の形容詞の扱いについて見たのち、その問題点を 2 つ指摘し、その解決の方向を示すものである。

2 日本語の形容詞の特徴と意味記述

日本語の形容詞についてはこれまで[5], [6]をはじめとする研究があり、自然言語処理を目的としたものでも [7], [8], [9], [10], [11]などがある。具体的な語義記述としては[12], [13], [14]などがある。これらにもあるように、日本語の形容詞は次のような特徴を持つ。

- (1) a. 動詞と同様、活用変化を持ち、連体修飾・叙述の他に連用修飾が可能である。(ただし、語によっていずれかの用法が欠落しているものもある)
- b. イ形容詞の「～ばい」、ナ形容詞の「～めた」等の一部の生産的な形容詞化語尾を含む語を除けば、日本語の形容詞の数

は約一千に限られる。

- c. 語によって段階性の有無がある。
- d. 「小さい」のように相対的な概念を示すものがある。
- e. 意味記述において名詞・動詞に比べて曖昧性が大きいように見える。
- f. 大きく「感情形容詞」と「属性形容詞」に大別でき([5]など)、さらにいくつかの下位分類できる([10]など)。
- g. 動詞と同様に項構造を持つ。

このような特徴を持つなか、(1a)は英語における形容詞の記述をそのまま借用することができないことを意味している。また、自然言語処理で用いる辞書は常に未知語の獲得が大きな問題となるが、(1b)はこの点で形容詞は未知語としての出現頻度はむしろ少なく、この語義記述が十全になされれば、名詞や動詞の未知語の意味を学習する際に大きな手助けとなる可能性が高いことを意味している。(1c)～(1g)はおおむね他言語と共通する特徴であるが、特に(1e)は語義記述が主観的になりやすい危険性を増大させている。1980年代以降、主観性の排除を一つの目的として語義の機械学習も多くの研究で試みられているが、その多くは項構造など比較的学习しやすいものに限定されているのが現状であり、学習のための枠組という根本的な段階からの再検討が求められていると言える。

3 生成語彙

先に述べたように自然言語の意味を記述する際に従来の数え上げによるものでは各語義の関係が事実上適切に示せない([4])。これに対して生成語彙では明示的な記述枠組とともにいくつかの生成的な演算を用意することによって動的に語義の曖昧性を捉えようとする。

記述枠組としては項構造、事象構造、特質構造、語彙階層構造の 4 つの表現層を用意し、特に特質構造では材料、内容、部分などの構成要素を示す構成役割、上位概念、形、色などの静的な属性を表す形式役割、その概念を用いる目

的や機能を示す動作を表す目的役割、その概念を生み出す動作や原因を表す主体役割の 4 つの役割を用意している。

生成的な演算としては現在のところ、強制、共構成、選択束縛が用意されており、形容詞については選択束縛の適用が例示されてきている。例えば、「速い」という語は「速い車」、「速いタイピスト」、「速い演説」などの使い方が考えられるが、このそれぞれにおける「速い」は「何が速いのか」という点で共通性に乏しく、主辞となる名詞の意味がいかようであるかによって決定される。これについて、生成語彙では、特質構造における役割の中から修飾対象を選択的に指定することによって解決を図る。「速い」の場合は、それぞれ「動くのが速い車」、「タイプするのが速い人」、「話される速度が速い演説」ということである。このように「速い」はモノではなく事象の状態を示す。一方、事象という点では、それぞれの名詞には特質構造において、「車：動く」、「タイピスト：タイプをする」、「演説：話す」といった強く関連する事象が他の事象とともに与えられている。すると、「速い」は各名詞の意味全体を修飾するのではなく、特質構造内を見て、事象のもっとも適切なものを選択的に修飾する、と考えるのである。

このような見方は「小さい」のように従来から相対的な尺度を意味すると指摘されてきた形容詞の扱いも簡単にすると考えられる[15]。例えば、「小さい犬」と「小さい象」では、「小さい」が概略意味する大きさはまったく異なる。この大きさの違いは捉えられるのが望ましいが、選択束縛を使えば、特質構造の中の形式役割を適切に設定することで一応解決できる。

形式役割は上記で述べたように、上位概念、形、色などの静的な属性を表すものであった。「犬」、「象」はそれぞれその語彙構造の形式役割にその標準的な大きさが指定されていると考えるのは順当であろう。これを前提に、「小さい」はそれぞれの名詞を修飾する際に語の意味全体ではなく、この「大きさを指定している部分」を選択束縛すると考えればよい。

このような見方は従来よりはるかに細かな意味記述を比較的明示的な記述枠組と生成的な演算によって示すことができる。しかし、いくつかの問題点がある。これを次節で指摘し、検討する。

4. 生成語彙の問題点

まず、選択束縛は文脈によって同じ組み合わせでも異なった束縛になることがある。特に日本語の場合は文脈から了解可能なものが極力省略されるため、こうした語義記述から予想されないものが出てくる。

- (2) 1990 年入社組は出世の速い遅いにずいぶん隔たりがあるようだ。一番速いプログラマは誰だと思う？

この場合は当然、「プログラムを書くのが速いプログラマ」という意味ではなく、「出世するのが速いプログラマ」ということである。このようなものは事実上無限に考えられるため、語義に記載するわけにはいかない。別の言い方をすれば、生成語彙を用いた機械辞書・解析を実装したとしても、それだけでは、上記のような例では必ず選択束縛に失敗する。ローカルな解析で失敗した場合、前後の文脈から適切な事象を探してくる機構が必要である。ただし、注意すべきなのは、こうした機構は生成語彙による語彙構造とまったく切り離して考えなければいけないのではないということである。並行処理であろうと後処理であろうと、上記で言えば、「プログラマ：出世する」という事象を動的に語彙の意味記述の一部として与える機構さえ用意できれば、生成的な演算そのものはまったく同様に利用できる。

この点で、生成語彙が用意する語彙階層構造を利用できる可能性もある。語彙階層構造は語が表す概念間の階層構造を記述する層であるが、特質構造の役割ごとに独立に定義できるものとされている。「プログラマ」はその形式役割に「特定の職業に就く人」といった意味の役割が上位概念として付与されていると考えられる。こうした「職業」に出世に関わる属性が付与されていると考えるのは自然であろう。このとき、語彙階層構造を利用して、「プログラマ」は「職業」から出世に関わる記述を継承することが理屈上は可能である。具体的には超えねばならないハードルが多いが、生成語彙のように豊かな語彙記述を可能にする枠組の場合、このような解決法がともかくも考えられるのは利点の一つであると言えよう。

第二の問題としては、例えば「小さい犬」はそれ自体が曖昧である、ということが挙げられる。

- (3) 花子は小さい犬は怖くないが、大きい犬は

襲われそうで嫌いだ。

- (4) この鉄門を押して開けるなど、小さい犬にはとても無理で、象でも連れてこないとだめだろう。

この2例を見ると、(3)における「小さい」は「犬」という動物の標準的な大きさに比べて小さい」という意味で使われており、3節で示した考え方で処理できる例だが、(4)における「小さい」は「さまざまな動物の中での犬というものの大きさは小さい」という意味で使われている。この区別を適切に行う機構は今のところ用意されていない。

この解決に当たって、IPALで指摘される1語の複数の語義間における段階性の有無も関連性があると考えられる。

- (5) a. この本はとても古い本だ。
b. 古い住所は消してください。

IPAL辞書では(5)の両例の「古い」を別語義と認定し、aは段階性のある「古い」、bは段階性のない「古い」とする。この知見そのものは正しいと考えるが、それが数え上げによる別語義として記載すべきものであるかどうかは別問題であり、生成語彙の立場からすれば別語義と捉えるのは好ましくない。また、このような両例を持つ形容詞は少なくなく、その意味で、このような例を別語義にせず扱えるならば望ましいと言えるであろう。

(4)と(5)で共通しているのは、いずれもaでは特質構造の形式役割における特定の属性の値を選択束縛によって具体的に設定することができることによってこの修飾の妥当性が保証されているのに対して、bではそうした特定の属性の値を設定しているというよりも属性そのものを設定していると考えられるということである。もちろんこれは属性である。次例のように「犬」は常に「小さい」ものではない。

- (6) この穴を通っていけるほど犬は小さくない。

つまり、ここでも文脈によってその属性が一時的に指定されるという点では第一の問題と同じことである。

では、生成語彙の枠組では上のような違いはどのように扱えばよいであろうか。Pustejovskyらは[4]などにおいて形容詞については選択束縛

以外の生成的な演算を適用する事例を紹介していないが、これは共構成を適用することによって得られるものと考ええる。なぜなら、(4b)や(5b)では基本的には「犬」や「住所」という名詞が他の動物と、あるいは、現住所と比較して小さかったり古かったりするという意味記述を元から持っているとは考えがたく、一方、「小さい」、「古い」という形容詞側も、本来は各名詞に関わらず「その標準的な大きさ、経年に対して小さい、古い」というばくぜんとした尺度基準を提供しているだけであって、具体的に「すべての動物という種族の中で小さい方だ」、「現住所と比べて、それ以前の住所だ」という具体的な指定をあらかじめ持っているとは考えられないため、生成的な演算を適用するのが妥当であると考えられるのが第一の理由であり、しかもその生成的な演算は両者が同時に何らかの変化を被るような演算でなければならないのが第二の理由である。

つまり、(4a)、(5a)はともに形容詞として普通選択束縛によって解釈される例であり、一方、(5a)、(5b)は共構成によって解釈される例である。実際の処理においてそのいずれの演算を適用するかの決定は決して簡単ではないことが容易に予想されるが、この方向で更に検討するのが妥当であると考ええる。

ただし、形容詞における段階性の有無はこれと別に指摘しておく必要がある。IPALの記述や外池[7]では多くの形容詞で段階性が少なくとも語義によって截然と区別できるという前提があるように思われる。しかし、上の「古い」のようにもともと段階性を持つ形容詞の場合、一見段階性がないように見える(5b)のような使い方で段階性は観察可能である。

- (7) この記録ははこれまでに暮らされた場所の履歴を取るのが目的です。古い住所も覚えている限り書いてください。ただし、二十年以上前のものについては、そこまで古い住所は覚えてられることの方が少ないですから結構です。

上例の二つめの「古い」は明らかに段階性が観察できる。さらに[17]では段階性があるものとない形容詞でも「である」方へ視野が絞り込まれると、絞り込まれた「である」ことの近傍で程度を読み込むことができる」と指摘しているが、必ずしも「である」方に絞り込まれる必要はない。

- (8) a. 彼の工作物は丸い。
 b. 彼の工作物はほぼ丸い。
 c. 彼の工作物は丸いと言ってよいのだろうが、私としては、もうちょっと丸い方がよかったと思う。

a と b は外池の指摘通りであるが、c のような例を考えると必ずしも「である」方の近傍である必要はない。

もちろんあらゆる非段階的な形容詞がこのように段階性を持つと主張するつもりはない。例えば「ない」は基本的に非段階的であり、「ほとんどない」のような例も従前から指摘される通り、対象が物質的にたくさんあるのが前提の場合にしか使えない。

- (9) a. !ぼくの部屋にはテレビはほとんどない。
 b. この倉庫にはテレビはほとんどない。

しかし、これは「少ない」のようなある程度段階的な形容詞でも観察されることであり、その観点からすれば、段階性の有無は語義の重要な一要素なのではなく、単にどのような語と結びつか、文脈がどうであるか、といったことによって導き出されるものであって、一種の結果である可能性がある。この点については稿を改めて論じたい。

5 おわりに

本発表では生成語彙における形容詞の取り扱いについて 2 つの点で問題を提起し、併せて形容詞の段階性の有無という基準を再検討する必要も指摘した。その基本的な立場は生成語彙を利用することを前提としながらも、生成語彙が目標とするレベルの細かな語彙記述のためには適宜文脈を利用する必要があるというものであった。実際に辞書を作成し、処理を行うという前段階として以上のような点は最低限一定の解決を必要とすると考えた。

また、生成語彙は本発表だけでも分かる通り、これまでの「言語的意味」と「世界知識」とを峻別する発想にない。とりわけ特質構造ではむしろ積極的に世界知識を取り込んでいこうとする発想が強い。狭義の言語学では特に言語を他の知識から独立させて論じようとする雰囲気強いが、むしろ世界知識を必要に応じて柔軟に取りこめるような枠組を模索する必要があるの

ではないか。そうした観点から生成語彙のような枠組を発展させていくことが語彙意味論の一つの方向性であると考えられる。

参考文献

- [1] R. W. Langacker. *Foundations of Cognitive Grammar, vol 1: Theoretical Perspectives*. Stanford: Stanford University Press. 1987.
- [2] R. W. Langacker. *Foundations of Cognitive Grammar, vol 2: Descriptive Applications*. Stanford: Stanford University Press. 1991.
- [3] R. Jackendoff. *Semantic Structures*. Cambridge, MA: The MIT Press. 1990
- [4] J. Pustejovsky. *The Generative Lexicon*. Cambridge, MA: The MIT Press. 1995.
- [5] 西尾寅弥. 『形容詞の意味・用法の記述的研究』. 国立国語研究所報告 44. 東京:秀英出版. 1972.
- [6] 佐伯哲夫. 「動詞性述語と形容詞性述語」. 北原保雄(編)『日本語の文法・文体(上)』講座日本語と日本語教育第 4 巻, 東京:明治書院. pp. 117-142. 1989.
- [7] 外池俊幸. 「形容詞の段階性/非段階性」. 『ソフトウェア文書のための日本語処理の研究-10—IPAL(Basic Adjective)をめぐる—』, 情報処理振興事業協会技術センター, pp. 17-76. 1990
- [8] 岡田直之. 『語の概念の表現と蓄積』. 電子情報通信学会. 1991.
- [9] 橋本三奈子・青山文啓. 「形容詞の三つの用法: 終止、連体、連用」. 『計量国語学』, 18-5, pp. 201-214. 1992
- [10] 内海彰・掘浩一・大須賀節雄. 「自然言語処理のための形容詞の意味表現」. 『人工知能学会誌』, 8-2, pp. 192-200. 1992
- [11] 宮島達夫. 「形容詞の語形と用法」. 『計量国語学』, 19-2, pp. 94-104. 1993
- [12] 情報処理振興事業協会. 『計算機用日本語基本形容詞辞書 IPAL(Basic Adjectives)』解説編・辞書編. 情報処理振興事業協会. 1990.
- [13] 飛田良文・浅田秀子. 『現代形容詞用法辞典』. 東京:東京堂出版. 1991.
- [14] EDR 電子化辞書
- [15] 松本裕治. 「意味と計算」. 『意味』岩波講座言語の科学第 4 巻, pp. 125-167, 東京:岩波書店. 1998.