

統語・意味情報を基にした複合形容詞・形容名詞形成モデル

高橋 幸† 佐藤 滋‡ 上原 聡‡

東北大学大学院国際文化研究科

†sachi@insc.tohoku.ac.jp ‡{satos, uehara}@intcul.tohoku.ac.jp

1 はじめに

我々は辞書情報を基にした複合語形成モデルの構築と、計算機上で自動的に生成するシステムの開発を目指している。それに伴い、本稿では辞書にどのような語彙情報を載せるべきかについて考察する。

形態素解析を行う際、複合語は辞書に登録されていなければ、記載単語に分割され、複合語としての情報を失うことになる。しかし、複合語は無限に造語されていくため、全てを辞書に登録することは現実的でない。また、従来の複合語解析・生成モデル[1][2]は複合語内の単語の共起頻度確率に基づくものが多いが、複合語の内部関係を、例えば主語－述語、目的語－述語というように表層的にしか分析できず、複合語全体の意味を捉えることは困難である。

そこで、我々は構成要素の持つ統語・意味情報を利用した、より言語学的な複合語生成システムを構築するために、辞書の記述において生成語彙 (Generative Lexicon) 理論[3]の適用を試みた。今回は名詞を前部要素に持つ複合形容詞・形容名詞 (「名詞＋形容詞」型、「名詞＋形容名詞」型)のみを対象とした。形容詞・形容名詞は、名詞と比べて語が自由に結合できず、数や組み合わせのパターンも限られている。構成要素を生成語彙の枠組みによって記述し、特に項構造と特質構造に着目して、複合語形成の文法的・意味的制約を導き出した。また、モデルの妥当性を評価するために、複合語の生成実験を行った。その結果も報告する。

2 辞書データベース作成

2.1 辞書登録語

まず、CD-ROM版『大辞泉』(1997年 小学館)から複合形容詞・形容名詞を抽出し、その構成要素の語を辞書データベースに登録した。ここで複合語とは自立語に分割できる語と考える。複合形容詞は393語(形

容詞全体の約31.8%)、複合形容名詞は412語(形容名詞全体の約26.3%)検索できた。構成する語の品詞種類や内部構造は様々であるが、前部要素に名詞がくるものが多く、形容詞は311語(約79.1%)、形容名詞は368語(約89.3%)であった。また、3構成要素以上から成る複合語は検出できなかったため、本稿では2構成要素の複合語のみを扱う。

2.2 生成語彙の枠組み

辞書の記述法としては生成語彙[3]を採用し拡張した。生成語彙は(1)項構造(argument structure):項構造を記述する、(2)事象構造(event structure):語が表わす事象とその時間的な関係を記述する、(3)特質構造(qualia structure):語の意味を表わし、4つの役割から成る、(4)語彙階層構造(lexical inheritance structure):語の概念的な階層構造を記述する、の4つの層から成る。我々のモデルでは、(1)項構造と(3)特質構造を中心的に用いて説明する。特に、特質構造は複合語全体の意味を捉えるのに有効であると考えられる。なぜなら、特質構造は単語の意味を静的に表示するだけでなく、語彙的な意味の解釈を作り出す「生成的な」働きがあるからである。名詞が挿入され複合語になると、転意により意味が変化する場合がある。例えば、複合語「根深い」の本来の意味は、「根が深く入っている」という物理的な物体「根」の状態を表わすものである。しかし、実際「根深い」は「しつこい、執念深い」という意味用法で用いられることが多い。この場合の「根」は「物事の起こるもと、根本、原因」という抽象的な概念を示している。このような転意は限りなく生じるため、複合語の意味生成を単語の一義的な意味情報だけで捉えることはできない。その点、生成語彙モデルは多数の語彙情報を持つ単語を統一的に記述できるような辞書モデルであるため、多数のフレーム処理によって複雑な複合語の意味を解析できる。

2.3 辞書記述

本節では、具体例を取り上げ語彙項目の記述法について説明する。図1は名詞「心」、図2は形容詞「苦しい」、図3は形容名詞「豊か」の辞書記述である。記述はHPSGと同様タイプ付き素性構造であり、四角の数字は同一指示を表わす変数である。ここで abstobj とは抽象的な概念を表わすのに用いている。

項構造 (ARGSTR) には、 $ARG_1, ARG_2, ARG_n, \dots$ で示される必須項 (true argument), D_ARG で示され、任意格にあたる暗黙項 (default argument), 及び S_ARG で示され、語義に含まれている格要素を表わす影項 (shadow argument) がある。例えば「迷惑」という形容名詞¹は「迷惑をかける対象 (GOAL)」、例えば「他人」や「近所」を補語としてとることができる。目標格 (GOAL) は D_ARG として捉えられる。一方、 S_ARG にあたる「kick」の場合の「leg」に相当するものである。尚、語がとる項については一部 [4] を参考とした。

kokoro	
ARGSTR=	$\left[\begin{array}{l} ARG_1=\boxed{1} \left[\begin{array}{l} \text{human} \\ \text{FORMAL}=\text{animate} \end{array} \right] \end{array} \right]$
QUALIA=	$\left[\begin{array}{l} \text{CONSTITUTIVE}=\text{part_of_x} \end{array} \right]$

図 1: 「心」の辞書記述

kurusii	
ARGSTR=	$\left[\begin{array}{l} ARG_1=\boxed{1} \left[\begin{array}{l} \text{human} \\ \text{FORMAL}=\text{animate} \end{array} \right] \\ D_ARG_1=\boxed{2} \left[\begin{array}{l} \tau \end{array} \right] \end{array} \right]$
EVENTSTR=	$\left[\begin{array}{l} E_1=e_1.\text{state} \\ E_2=e_2.\text{process} \\ \text{RESTR}=e_2 < \alpha e_1 \\ \text{HEAD}=e_1 \end{array} \right]$
QUALIA=	$\left[\begin{array}{l} \text{create} \\ \text{FORMAL}=\text{kanasii}(e_1, \boxed{1}) \\ \text{AGENTITIVE}=\text{exp_act}(e_2, \boxed{1}, \boxed{2}) \end{array} \right]$

図 2: 「苦しい」の辞書記述

事象構造 (EVENTSTR) は、今回は直接的に用いていないが、辞書には記述した。形容詞、形容名詞は基本的に主辞事象 (HEAD) として状態 (state) の値を取る。

特質構造 (QUALIA) は、(a) 構成役割 (constitutive role): 材料、内容、部分 (part of x) 等の構成要素を表わす、(b) 形式役割 (formal role): 形、色等の外的

¹形容名詞が名詞を修飾する際に付く繫辞「～な」は copula「～だ」の連体修飾形であると考え、本分析では語幹のみの形で記述している。

な属性を表わす、(c) 目的役割 (telic role): その概念の目的や機能を示す、(d) 主体役割 (agentitive role): その概念を産み出す動作や原因を表わす、の4つの役割から成っている。

yutaka	
ARGSTR=	$\left[\begin{array}{l} ARG_1=\boxed{1} \left[\begin{array}{l} \text{physobj} \end{array} \right] \\ ARG_2=\boxed{2} \left[\begin{array}{l} \text{abstobj} \end{array} \right] \\ D_ARG_1=\boxed{3} \left[\begin{array}{l} \tau \end{array} \right] \end{array} \right]$
EVENTSTR=	$\left[\begin{array}{l} E_1=e_1.\text{state} \\ E_2=e_2.\text{process} \\ \text{RESTR}=e_2 < \alpha e_1 \\ \text{HEAD}=e_1 \end{array} \right]$
QUALIA=	$\left[\begin{array}{l} \text{create} \\ \text{FORMAL}=\text{yutaka}(e_1, \boxed{1}) \\ \text{FORMAL}=\text{yutaka}(e_1, \boxed{2}) \\ \text{AGENTITIVE}=\text{R}(e_2, \boxed{3}, \boxed{2}) \end{array} \right]$

図 3: 「豊か」の辞書記述

また、辞書登録のユーザインターフェイスを開発し、データを管理・修正する際の負荷を軽減した。

3 複合語形成モデル

3.1 項構造における制約

辞書データ中の「名詞+形容詞 (形容名詞)」型複合語を、述語に対して名詞がどのような文法機能 (「主語」「直接目的語」「間接目的語」「補語」「付加部 (副詞的要素)」) を持っているかによって分類すると、表1のような結果が得られた [5]。

表 1: 名詞の文法機能による複合語の生産性分布

	形容詞	形容名詞
主語 (ガ格)	306 語 (98.4%)	321 語 (87.2%)
直接目的語 (ヲ格)	0	0
間接目的語 (ニ格)	0	0
補語	0	36 語 (9.8%)
付加部	5 語 (1.6%)	11 語 (3.0%)
合計	311 語	368 語

表1より、形容詞・形容名詞と結び付いて複合語となれるのは、付加部を除けば、主語や補語となる名詞であることが分かる²。この制約は、2.3 節の辞書では ARGSTR に記述するものである。主語は ARG の値で、補語は D_ARG の値に相当する。この場合の補語とは GOAL の意味役割を持つものであった。

²先行研究 [6] では、述語と項で形成される複合語形成において述語と複合するのは内項 (あるいは内項が複数ある時には最も内側の直接項) の名詞であるとする、「直接項の法則」を取り上げている。

3.2 複合語形成過程

次に、辞書の統語・意味情報を利用した複合語形容詞・形容名詞形成の流れについて説明する。

まず、手順1において項構造における制約を適用する。前部要素である名詞(w1とする)が述語である形容詞、形容名詞(w2とする)の主語もしくは補語となれるかどうかを判定する。w1とw2のARGSTRのARG1の値が満たせば、手順2として特質構造における意味関係の整合を分析する。そして最終的に複合語全体の意味情報を出力する。

もし、w1とw2のARGSTRのARG1の値が整合しなければ、w2のAGR2、D_ARG、S_ARGの値との整合を試みる。このようにして、次々と項制約を適用していく。w1とw2の値が最後まで整合できなければ、次に意味素性に頼った複合語形成が行われていると考え、手順2の分析に移る。

手順2では、特質構造における値の整合の判断を人手によって行う。意味的な制約として先ず挙げられるのが、形容詞の種類と名詞の属性である。形容詞と形容名詞は意味用法によって、対象の属性を表わす「属性」と、感覚や感情を表わす「感覚・感情」に大別できる[7]。その境界は曖昧であるが、「豊か」は前者に、「苦しい」は後者にあたると考えられる。「感覚・感情」形容詞・形容名詞は、主語に無生物のものがとれないという制約がある。

例えば、複合形容名詞「心豊か」について分析すると、abstobjの値を持つ「心」と abstobjの値を持つ名詞(他には「感情」等)を項(主語)に持つことのできる「豊か」の間では矛盾が起きない。よって、複合語形成の条件は満たしているといえる。「彼は豊かな人だ」と表現できるが、図3においてRGSTRの値にhumanの値が与えられていない。これは彼の“経済力”や“感性”等、ある部分、しかも inanimate なものの属性を指していると考えられるため、概念的には animate の値を持つ名詞は主語となれないと考えた。

一方、「苦しい」は animate の値を持つ名詞を項(主語)としてとることはできるが、inanimate な physobj の値を持つ名詞は取ることができない。しかし、“心”等は animate の素性値を持ち、形容詞「苦しい」の第1項となれる人間と、部分-全体の関係にある。そのため、主語になり得ると考えられる。よって、項制約が適用され、複合語となる条件を満たす。

図4に複合語「心豊か」の意味出力を記す。「心」が挿入されたことにより、「心」のARGSTRである human

の値を持つ名詞しか項(主語)に持てないという情報が記される。

また、同じように「心強い」等も図5で表わすように複合語化することにより、属性的な要素が感情・感覚的な意味を持つようになり、例えば「彼は私にとって心強かった」というように心理的な感情を表わせるようになる。

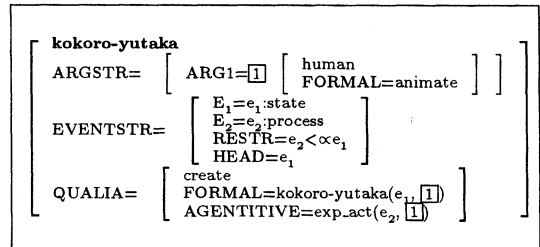


図4: 「心豊か」の辞書記述

3.3 本モデルの問題点

本モデルの問題点として、出力された複合語の意味を捉えきれない現象が存在する。「根」と「深い」で複合語「根深い」となる場合、「深い」と「根深い」の意味構造の違いは明らかにできる。しかし、「心苦しい」のような複合語と単純語「苦しい」の意味構造の違いは生成語彙の枠組みでは捉えきれず、両者とも同じ意味構造を持つことになってしまう。これは、“苦しく”感じるところが、そもそも“心”そのものであるためである。複合語の場合、右側の構成要素の意味は複合語へ継承される[8]。よって基本的に「心苦しい」が「苦しい」の意味を継承するものであるが、全く同じというわけにはいかない。複合語の複雑な意味概念構造をどのようにして記述していくかは今後の課題である。

構成要素間には、主語-述語という文法的に説明ができる関係から、文法では説明できない関係まである。挿入される名詞は後部要素である述語から格を付与されるが、文中ではガ格やヲ格をとらない場合がある。それらの単語は修飾的役割、つまり付加部である。例えば形容詞では「目新しい」、「涙もろい」、形容名詞では「一目瞭然」等があった。このタイプの複合語に関しては今回は分析を行わなかった。

よって、辞書データを基にして辞書に登録できた複合形容詞は306語、複合形容名詞は357語であり、総構成要素数は287語であった。

4 実験と評価

辞書データに存在しない複合語の例を我々のモデルで解析できるか検討し、モデルの妥当性を評価した。

4.1 実験方法

対象データは『天声人語』1989年の1年分と、CD-ROM版『新潮文庫の100冊』(1995年新潮社)に収録された5作品から検索した複合形容詞・形容名詞である。また、解析する複合語は、辞書に存在しないものを対象とした。複合形容詞・形容名詞の抽出にあたっては、先ず形態素解析システム『茶筌』[9]で形態素解析を行い、品詞の並びに注目して“…名詞・形容詞(もしくは形容名詞)”のパターンで現れた文字列を複合形容詞・形容名詞であると判断した。データからは、述べ52語、21種類の複合形容詞、述べ128語、42種類の複合形容名詞が検索できた。実験は種類別の複合語に対し行った。

モデルについては計算機上に実装した。検索語の前面要素である名詞と、後面要素である形容詞もしくは形容名詞を入力し、我々のモデルの制約により複合語を形成することができるかを検討した。構成要素が辞書データに存在しないものについては、辞書の記述・登録後に実験を行った。

4.2 結果と考察

表2に実験結果を示す。本モデルはテキストデータ中の90%以上の複合語に対して適用できた。80%以上の複合形容詞・形容名詞形成が項構造による制約で説明がついた。複合形容詞は全てが項制約で形成される複合語形成であり、全てに「主語ー述語」の関係が成り立った。先行研究[10]の指摘にあるように、英語では意味レベルで複合語が生成されるが、日本語の複合形容詞・形容名詞は項構造レベルの生成が多いことが分かる。

表2：モデルによる複合語生成実験結果

	手順1	手順2	生成
	項構造での整合	意味構造での整合	不可能
複合形容詞	21語(100%)	0語	0語
複合形容名詞	34語(81.0%)	5語(11.9%)	3語(7.1%)
合計	55語(87.3%)	5語(7.9%)	3語(4.8%)

また、解析できなかった複合形容名詞の例としては、前述した付加部となるような名詞が挿入された例の他に、3つ以上の構成要素によって成り立つもの“カード

利用可能”のようなものがあつたためである。

5 おわりに

本稿では複合語の構成要素の統語・意味情報を基にした複合語形成モデルを提案した。辞書の記述にあたっては生成語彙の枠組みを採用し、文解析の場合と同様に複合語形成にも適用できることが分かった。また、実際のテキストデータ中の複合語を基にした語形成実験結果により、モデルの有効性を示した。

今後は、他の複合語の構造に対しても適用し、辞書データベースを拡張する予定である。辞書に登録されていない未知語の複合語解析にも対応できるように、既に登録された語の情報を活かしデータベース作成を行う。辞書が整備され多くの情報を網羅できれば、自由な入力に対し複合語を生成するようなシステムの開発が可能となる。

また、名詞と形容名詞には「大柄」「真面目」のように重複して存在するものが多い。こうした関連する単語同士を関連付けて辞書登録を行う予定である。

謝辞

本研究は一部、日本学術振興会科学研究費(No.12610548)の支援を受けて行われている。

参考文献

- [1] 太田悟, 宮崎正弘: 複合語用例データベースを用いた複合名詞の構造的曖昧さの絞り込み法. 情報処理学会第53回全国大会, pp.2-9, 1996.
- [2] 西野哲朗, 藤崎哲之助: 漢字結合語の確率的構造解析. 情報処理学会論文誌, Vol.29, No.11, pp.1034-1044, 1988.
- [3] Pustejovsky, J.: *The Generative Lexicon*. MIT Press, 1995.
- [4] NTTコミュニケーション科学研究所: 日本語語彙体系 5. 岩波書店, 1997.
- [5] 高橋幸, 佐藤滋: 辞書データを基にした日英語形容詞・形容名詞の複合語形成に関する一考察. 言語処理学会第6回年次大会発表論文集, pp.91-94, 2000.
- [6] Levin, B. & Rappaport, M.: The Formation of Adjective Passives. *Linguistic Inquiry*, No.17, pp.623-661, 1986.
- [7] 西尾寅弥: 国立国語研究所日本語教育指導参考書 13 形容詞の意味・用法の記述的研究. 秀英出版, 1972.
- [8] Williams, E.: On the Notions 'Lexically Related' and 'Head of a Word'. *Linguistic Inquiry* 12, pp.245-274, 1981.
- [9] 松本裕治, 北内啓, 山下達雄, 平野善隆: 日本語形態素解析システム『茶筌』version 2.0, 1999.
- [10] 影山太郎: 形態論と意味. くろしお出版, 1999.