

文中の複数の語の(共)項構造の同時的、並列的表現法

Pattern Matching Analysis (Simplified) の観点からの「係り受け」概念の拡張

黒田 航[†] 飯田 龍^{††}

[†] 独立行政法人 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

^{††} 奈良先端科学技術大学院大学

E-mail: [†]kuroda@nict.go.jp, ^{††}ryu-i@is.naist.jp

あらまし 本研究は、PropBank [3] に代表される動詞の項構造、NomBank [6] に代表される名詞の(共)項構造のデータベース化を日本語についても行なうための効果的な実装方法として Pattern Matching Analysis (PMA) の簡略版 (PMA Simplified: PMAS) を用いる方法を記述する。PMA は黒田 [4] が提唱した、従来の係り受け解析と互換性を保ちつつ、表現力の不足を補う形で構文解析をするための手法である。PMA は助詞類(係り助詞、格助詞)にも動詞から独立した項構造を指定するが、この指定は今回の課題設定である「述語性をもつ要素の項構造の特定」には不用である。そのためこの論文の扱った範囲では PMA の解析の対象を動詞、形容(動)詞、事態性名詞のような述語性の明らかなものに限定し、PMAS とした上で、(i) 事態性名詞の項構造、(ii) 非事態性名詞の共項構造 (co-argument structure)、(iii) いわゆるゼロ照応の3つを分析した。

キーワード 多重パターンマッチング法による係り受け解析の拡張、項構造解析、共項構造解析、ゼロ照応解析、

Word-wise, parallel representation of (co-)argument structures in a sentence with Pattern Matching Analysis (Simplified)

Kow KURODA[†] and Ryu IIDA^{††}

[†] National Institute of Information and Communications Technology (NICT)

Hikaridai 3-5, Seika-cho, Souraku-gun, Kyoto, 619-0289, Japan

^{††} Graduate School of Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara, 630-0192, Japan

E-mail: [†]kuroda@nict.go.jp, ^{††}ryu-i@is.naist.jp

Abstract Based on the framework called Pattern Matching Analysis (PMA) [4], this paper proposes an “argument parsing” method that can be used to develop a lexical resource for Japanese that integrates features of PropBank [3], where argument structures of verbs are specified, and NomBank [6], where (co-)argument structures of event-evoking nouns are specified. PMA allows us to specify the (co-)argument structure of each word in a given sentence. We also suggest that PMA-based parsing allows us to integrate co-reference analysis [16] into the specification of (co-)argument structures.

Key words argument structure parsing, co-argument structure parsing, co-reference analysis, word-wise dependency parsing, Pattern Matching Analysis (PMA)

1. はじめに^(注1)

この論文で私たちは PMA [4], [10] の簡略版を用いて、(i) 事態性名詞の項構造、(ii) 非事態性名詞の共項構造、(iii) いわゆるゼロ照応の三つを分析し、PMA(S) が従来の係り受け解析が深い意味関係を表わせないという問題克服する記述法であることを主張する。

1.1 事態性のある名詞の振る舞い

述語/項構造 (predicate-argument structure; 以下では単に項構

造) の解析は自然言語の統語解析、意味解析に重要な役割を果たす。このことを早くから理解した人たちは PropBank [3] のようなデータベース化を進めて来ている。

述語である動詞、形容(動)詞が項構造をもつのは当然であるが、項構造をもつのは動詞に限られるわけではない。明らかに(1)の「届け」のような動詞派生の名詞は(少なくとも意味的には)項構造を本来的に内在させているし、(2)の「反対」「議論」のようなサ変名詞も(少なくとも意味的には)項構造を本来的に内在させていると言える:

(1) それは [彼]_{Arg0} の [[届け]_{V1} [物]_{Arg1}] でした。

(2) [彼]_{Arg0} の [反対]_{V1} のせいで [議論]_{V2} が紛糾した。Arg0, Arg1 は PropBank の記法に従い、おのおの外項(≈主語)、第一内項(≈直接目的語)を表わすとする。

(2) のような指定が「表層よりも深い」レベルでの意味処理

(注1): 私たちはこの論文の準備にあたって乾健太郎 (NAIST)、河原大輔 (NICT)、小町守 (NAIST) の各氏との議論から多くの示唆を得た。この場を借りてお礼を申し上げます。

を可能にするのは明らかであり、可能な限り多くの述語性をもつ語に対して、それらの項構造を生起文脈と共に指定したデータベースは、機械学習の教師用データとして有用性が見込まれている。この目的のために構築中の言語資源の代表例が PropBank [3] の補完として名詞の述語/項構造をアノテーションした NomBank [6] である。

1.1.1 項構造の認定基準の拡張の必要

NomBank に相当する名詞の項構造を明示した言語資源を日本語にも用意するための準備的研究として、小町ら [11] の研究がある。彼らは「動作性名詞=動詞由来の名詞, サ変名詞」と定義し、次の文で [電話]₂ を動作性名詞と認定している:

(3) 1階の入り口のところの [電話]₁ で [電話]₂ をかけた。

この小町ら [11] の定義にあるように、(i) 動詞派生の名詞、(ii) サ変名詞は項構造をもつ名詞の主なグループだと思われるが、項構造をもつと考えられる名詞はそれですべてというわけではない。例えば、次の例にある「反感」のように、「抱く」や「持つ」などの**支援動詞** (support verbs) [1], [2], [5] を用いないと述語として使えない名詞にも項構造が潜伏している場合がある。^(注2)

(4) 彼は その相手に **反感** を {i. 抱い; ii. 持つ} た。

(5) *彼は その相手に **反感** した。

「反感」に項構造が潜伏しているのは、(6) の名詞句から (4) の意味内容が復元できることから明らかである。^(注3)

(6) 彼の その相手 {への; *に; に対する} **反感**; その相手 {への; *に; に対する} 彼 (??から) の **反感**

「反感」のような名詞は小町ら [11] の言う意味での動作性名詞 (act(ivity)-denoting nouns) とは言いがたい。このため、動作性名詞を含む、より一般的な名詞のクラスがあると想定し、それを暫定的に**事態性名詞** (event-denoting nouns) と呼ぶことにする。ここで重要なのは、名詞の動詞としての用法の有無とは別に述語性の有無の認定基準を設けるという点である。

1.1.2 事態性名詞の三つのタイプ

上の定義によれば、例えば対象を F : “ x が (y に) z を w ” というパターンに限定した場合、 z が述語性をもつ事態性名詞であるとき、表 1 の場合が考えられることになる (A, B, C では F が支援動詞構文 (Support Verb Construction: SVC) となる)。

表 1 F 内での z の事態性の観点からの分類

タイプ	z の例	項構造	w による支援	SVC?
A	恨み (< 恨む)	YES	随意的 (e.g., w =抱く)	YES
B	感動 ((する))	YES	随意的 (e.g., w =覚える)	YES
C	反感 (* (する))	YES?	義務的 (e.g., w =覚える)	YES
D	手 (* (する))	NO	義務的 (e.g., w =打つ)	NO

次の特徴をもつ名詞 z タイプ C の事態性名詞である: (a) 動詞由来、あるいはサ変名詞ではないが、項構造をもつ。 (b) 統語的に動詞になるためには支援動詞 w を必要とする。 (c) z の項構造は w の項構造と (整合的だが) 別モノである。

タイプ A 「恨み」のような名詞を便宜的に、**タイプ A の事態性名詞** と呼ぶ。これは動詞派生の事態性名詞と同一視可能。

(7) x が y に (対し (て)) **恨み** を {i. もつ; ii. 抱く} $\Leftarrow x$ が y {を; *に} **恨む**

タイプ B 「感動」のような名詞は**タイプ B の事態性名詞** と呼ぶ。このタイプはサ変名詞と同一視可能。

(8) x が y {に; ?に (対し (て))} **感動** を覚える $\Leftarrow x$ が y {*を; に} **感動** する

タイプ C 「反感」のような名詞を**タイプ C の事態性名詞** と呼ぶことにする。このタイプの名詞は動詞派生でもサ変名詞でもないが、述語性をもつ名詞である。

(9) x が y に (対し (て)) **反感** を {i. もつ; ii. 抱く} $\Leftarrow x$ が y {*を; *に} **反感** する

タイプ C の事態性名詞「反感」は**支援動詞による支援が項構造の統語的実現に不可欠な事態性名詞**であると考えられる。この理由から、このタイプ C の事態性名詞は**支援要求名詞**と呼んでもよいだろう。タイプ C の事態性名詞は動詞からの派生の関係の想定を許さないで、迂言表現とは言えない。

1.2 記述の方針

「 x が y に z を w 」という形の支援動詞構文 (w は z に対する支援動詞) は事態性名詞 z に関する迂言的表現だと見なされることが多い。この特徴づけは誤りではないが、支援動詞構文をイデオム=常套表現で片づけることになるなら、現象の表面的な扱いになる。実際、「常套句なのだから、有限個の実例を列挙すればそれで終り」という想定は、おそらく誤りである。(10) が支援動詞構文として使われる場合、 z を語彙的に完全に網羅するのは (仮に x と y を語彙的に固定しても) 不可能である:

(10) x が y に z を {抱く; 持つ} $z \Rightarrow$ {i. 反感; ii. 違和感; iii. 好感; iv. 親近感; v. 愛情; vi. 愛着}, \Leftarrow {i. *同感; ii. *実感}

タイプ A, B の事態性名詞は既成の辞書にある情報から特定可能で、{A, B} の認定はほぼ自動で行なえる。厄介なのは C, D の分離の問題である。NLP 的な課題はタイプ C の事態性名詞をなるべく効率的に—できるならば (半) 自動で—見つけることである。今のところ、既成の辞書 (日本語語彙大系, IPAL) にはタイプ C の事態性名詞が項構造をもつことを予測する情報は記載されていない。このため、タイプ C の名詞がタイプ D の名詞と同じ扱いを受ける。

1.2.1 事態に命名する名詞と事態を喚起する名詞

ここで重要な問題を明示化しておきたい。述語は (定義により) 項構造をもつが、その逆は言えるのか? つまり「**項構造をもつものはすべて述語である**」と言えるのか? この問題の答えは明らかに「項構造をもつとはどのような条件を満足することか」という基準に依存する。基準が甘すぎると、あまりに多くの語が項構造をもつことになり、好ましくない。

項構造の概念を拡大すぎないように注意しておくべきなのは「それ自体は述語ではないのに、まるで項構造をもっているかのように振舞う語がありうる」という可能性である。このような規定があてはまると思われる名詞が幾つか存在し、それはタイプ C, D の境界と関連している。例えば、(11) の「反感」が事態性名詞で項構造をもつとすれば、同様の理由から (12) の「手段」のような名詞にも項構造を認めるべきである。ただ、これを認めると「手段」の意味で使われる「手」にも同じことを考えなくては行けない。

(11) K 首相 (の Y 神社参拝) への 中国政府 (から) の **反感**

(12) その問題 {の解決 (のため); を解くため} の (彼の) {**手段**; **手法**; **技術**; **策**}

(13) その問題 {の解決 (のため); を解くため} の (彼の) **手**
私たちは「手」が事態性名詞とは考えず、「反感」のような例とは区別すべきだと考える。「反感」は**事態を命名する名詞** (event-denoting nouns) であるが、「手段」や「手」はそうではなく (14) のフレーム [9]) で表現されるような**事態を喚起する名詞** (event-evoking nouns) であると考えよう:

(14) 〈利用者: x 〉が 〈目的: y 〉 (の; を 〈実現〉する) ために 〈手段: z 〉を {用いる; 使う}

(14) が記述しているのは〈手段の利用〉フレームであり、〈利用者〉、〈目的〉、〈手段〉はこのフレームの**フレーム要素** (frame elements) である。これに対し、「使う」「用いる」はフレームの**支配要素** (frame governor) が語彙化されたものである。一般にフレーム F のフレーム要素は F を喚起するだけだが、支配要素は F を喚起し、かつ F に名称を与える。

1.2.2 項構造と共項構造の区別

次にフレーム要素、フレームの支配要素の定義と、それに基づく項構造、**共項構造** (co-argument structure) の定義を示す:

(15) フレーム F を喚起する要素は F のフレーム要素 (の語彙的实现) である。語 w がフレームの支配要素である場合、 w は項構造をもつ。 w がフレーム要素であり、かつフレームの支

(注2): 解説のために英語の支援動詞の例を挙げる。例えば argue が名詞化され have an *argument* (with somebody) と make an *argument* (against; for) {somebody; some claim} に現われる時、have, make を (argument の) 支援動詞と呼ぶ。

(注3): 名詞化の際に、「に」→「へ」の形態変換が義務的に起こっている。

配要素でない場合、それは共項構造をもつ^(注4)。

以上の定義に基づいて、「手段」や「手」は〈利用 [手段の]〉フレームのフレーム要素であり、その支配要素ではないため、項構造はもたず共項構造のみをもつとすることができる。

これは「反感を抱く」で「抱く」が支援動詞と考えられるのに対し、「手段を{使う;用いる}」では「{使う;用いる}」を本動詞に相当すると考える根拠となっている。これらが本動詞であるということは、それが〈利用 [手段の]〉フレームの支配要素であるということである。

1.2.3 喚起のタイプ

喚起は語彙的な場合と連語的な場合とがある。「手段」は語彙的に〈利用〉フレームを喚起し、「手」は連語的に〈利用〉フレームを喚起する。

1.3 解析の方針

すでに述べた通り、「反感」のような語は、私たちが調べた限り、どの辞書にも述語性をもつ名詞としては記述されていない。従って、どんな語がタイプCの事態性名詞であるかは、逐一判別しなければならない。この作業は簡単ではないが、以上の項構造と共項構造の理論を仮定すると「タイプCの事態性名詞が述語として実現するためには支援動詞による支援が必要である」という予測が立つ。これが事実なら、この特徴を利用して作業を単純化できる見込みがある。

ただし、支援動詞と支援される名詞との組み合わせは単純ではないので、十分な実態調査と正確な記述が必要である^(注5)。この調査と記述の両方を兼ねる分析手法として、私たちは Pattern Matching Analysis (PMA) [4], [10] の簡略版 PMAS を用いる。PMA(S) の利用には PropBank が記述している動詞の項構造と、NomBank が記述している名詞の項構造を同一のデータベースに表現できるという効能もある。

1.3.1 PMA の役割

具体的な分析例を示す前に、PMA を導入する動機と利点について簡単に説明しておきたい。

まず、PMA による記述は (例えば HPGS や LFG による) 文法の実装に取って代わるものではない。PMA が担当するのは文法を書き下すことではなく、文法に反映されるべき統語論、意味論的事実の発見と、その高品質な記述である。従って、PMA は計算言語学者、工学者が実際の文法を書き下すための「前処理」を担当するものだとして理解するのが適当であろう。

PMA の対象は (少なくとも原理的には) 単文内の構造に限らない。木構造を使った解析とは違い、PMA では探索の対象は文字列ベースなので、文間の関係を解析する場合でも (個々の語から「見える」範囲が長くなるだけで) 解析の複雑度の増加は木構造の場合ほど大きくならない。

私たちが考えるシナリオは次の通りである。

(16) PMA は新しい事実の発見を支援し、発見された事実の簡潔で適切な記述を可能にするので、**言語学者は解析手法に PMA を利用すること自体から恩恵を得る。**

(17) PMA の与える記述は簡潔で適切ならばだけでなく、**言語学を専門としない研究者に対して最低限の知識で理解可能なもの、共有可能なもの**である。従って、言語学者が自分の研究のために具体的な例文に PMA を行ない、それを公開することは、**分散型の言語資源の (無償) 開発**となる。

恩恵とは次のようなものである: (i) PMA は文を読み/聞き流すだけでは気づきにくい統語、意味関係に気づかせる効果がある。 (ii) 木構造を使った解析にも同様の効果が伴うこともあるが、それにはしばしば全体の整合性のために部分の整合性が犠牲になるという代償がある。PMA は解析の結果を特定の構造 (例えば木構造) にムリやり当てはめることを強要しないので、このような危険は少ない。

2. PMA と PMAS の簡単な解説

2.1 PMA の例

s		太郎**	は**	花**	で**	玄関**	を**	きれい**	に**	飾っ**	た**
p1	太郎	太郎*	P:??							V	IV
p2	は	T	は*							{N, V}	
p3	花	S	P(S):(が)	花*	P(O):(を)	O	P(O):(を)			V{2}:(使っ)	IV
p4	で	S	P(S):(が)	X	で*					V{2}:??	IV
p5	玄関	S	P(S):(が)			玄関*	P(O):(を)			V{2}:??	IV
p6	を	S	P(S):(が)			O	を*			V{2}:??	IV
p7	きれい					S	P(S):(が)	きれい*	P(C):(に)	V{2}:(なっ)	IV
p8	に					S	P(S):(が)	C	に*	V{2}:(なっ)	IV
p9	飾っ	S	P(S):(が)	X	P(X):(で)	O	P(O):(を)			飾っ*	IV
p10	た	S	P:??							V	た*

図1 (18) の PMA

PMA は係り受け解析 [14] に対して上位互換性をもつ意味関係の分析法である。これは真に受けることの難しい主張だと思われるので、例文 (18) の PMA を示し、それが何を意味しているかを解説しながら実証することにした。

(18) 彼は花で玄関をきれいに飾った。

(18) の PMA を図1に示す。図1の p_1, p_2, \dots, p_{10} は、文 $s = (18)$ を構成する部分パターン (subpatterns) である。文全体の意味構造と統語構造は、部分パターンの情報を列ごとに統合して与えられる (一行目にある $[w_i^{**}]$ (e.g., [太郎**], [は**], ..., [た**]) は w_i^* が現われている列の統合の結果を記号化したもの)。

PMA は文の部分の構造を詳細に記述し、それらを「移動」を使わず、「変数」と「値」の多次的な対応づけとして表現することで文の全体の構造との関係を表示する。

PMA の設計思想の一つは、「見えない要素の数は可能な限り減らす」ことである。この際、「見えない」要素は、他の要素と「重なっている」だけだと考える。実際、語ごとに独立に依存関係を認定すると「見えない」(照応的)要素 (e.g., S, O, O2) が見えるようになる。

2.1.1 部分パターンの構成

部分パターン p_i は、文 $s = w_1 \dots w_i \dots w_n$ の i 番目の語 w_i の依存関係を記述するが、PMA の考える依存関係は係り受けの関係よりも明確な解釈をもっている。 p_i は (i) w_i が述語 (e.g., 動詞、形容動詞、事態性名詞) の場合はその項構造を、(ii) w_i が非述語 (e.g., 「本」のような非事態性名詞) の場合にはその共項構造を記述したものである。

2.1.2 部分パターンの定義

部分パターン $\{p_1, \dots, p_n\}$ は文 $s = w_1 \dots w_n$ を構成する語の一つ一つについて定義される。例えば、(18) を s とし、 $W(s) = \{w_1, w_2, \dots, w_n\}$ とすると、

(19) $W(s) = \{w_1 = \text{彼}, w_2 = \text{は}, w_3 = \text{花}, w_4 = \text{で}, w_5 = \text{玄関}, w_6 = \text{を}, w_7 = \text{きれい}, w_8 = \text{に}, w_9 = \text{飾っ}, w_{10} = \text{た}\}$

から、 $P(s) = \{p_1, p_2, \dots, p_{10}\}$ が定義される。

S, O, O2, ... のような指定は単に統語素性の指定 (e.g., [if x is S, then x is NP]) を行なうばかりではなく、選択制限の表現のために意味素性の指定も行なうとする。これにより、部分パターンによって文中での語 w_i の選択制限が、すべての i について同時に表現されることになる。現時点で、選択制限とは何である

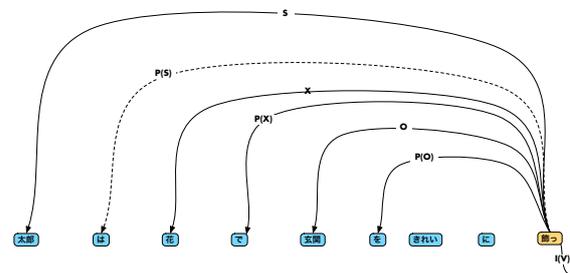


図2 (18) の「飾っ」の依存構造

(注4): この分析はクォリア構造 [7] を使った分析に類似するが同一ではない。クォリア構造は共項構造の特殊な場合でしかない。

(注5): 簡単な予備調査の結果を <http://clsl1.hi.h.kyoto-u.ac.jp/~kkuroda/data/supported-nouns-and-supporters-v1.xls> に公開している。

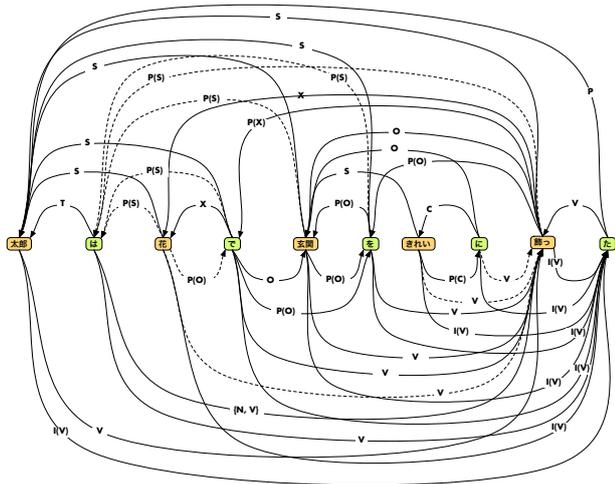


図3 (18) 全体の依存構造

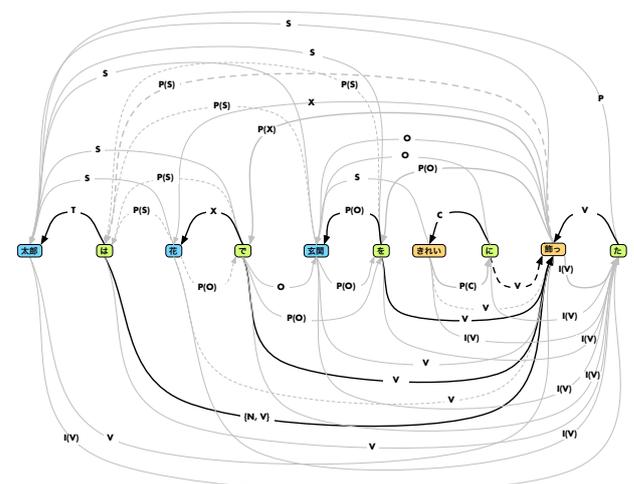


図4 (18) の従来の依存構造分析: 上下が逆になるのは、係り受け解析では、係り受けの関係は厳密には述語の項の実現の関係だと定式化されていないからである。係り受け関係は下半分にある「左から右への」「認可者探し」関係にすぎない。だが、この規定は厳密には「きれいに」のような副詞には当てはまらない。例えば「きれいに」は係り先の「飾った」に認可されているのではなく、逆に係り先を項に取っているからである。

か、それはいかに記述すべきか、という問題は、まだ本格的な解決を見ていない問題だということは強く意識している。第一著者は部分パターンがその記述単位になると期待している。

2.1.3 従来の依存構造解析との比較

係り受けが正確にどんな関係であるかは(それが依存関係の一種であるという緩やかな定義しかないため)明らかではないが、PMA で解釈される依存関係は**述語と項の実現関係に限られる**。別の言い方をすれば、述語と項の実現関係として解釈できない関係は依存関係とは見做さない。

任意の統語パターンと全体パターンとの関係は依存構造として解釈可能である。例えば、 p_9 「飾つ」と s の他の語彙要素との関係は図2に示す依存構造を表わすと解釈可能である。ただし、図2にあるのは「飾つ」のみについての部分的な記述なので、 s 全体の依存構造の記述は、 p_1, \dots, p_{10} の個々の依存構造を「重ね合わせ」たもので、図3にあるものになる。図3で指定されている構造は依存構造解析の一種である。

図3では、右から左への(つまり後に現われた要素から先に現われている要素への)認可の指定は上半分に示し、左から右への(つまり前に現われた要素から後に現われる要素への)要求(あるいは期待)の指定は下半分に示し、(値の指定の食い違いなどで)指定が部分的に破られている関係は破線で区別した。

(18)の係り受け解析と呼ばれているのは、図4にある制限付きの依存構造解析である。図4にある依存構造はKNP[15]や南瓜(CaboCha)[13]が出力する構造的に等価である(ただしKNPとCaboChaで出力の詳細は異なる)。

係り受けに表現される関係は常にPMAに表現され、その逆は真ではないので、PMAは係り受け解析に対して上位互換の関係にあると言える。

PMAは係り受け解析を拡張するために導入した指定[12],[17]の多くを含んでいる。例えば「名詞の格フレーム」はPMAでは名詞の共構造として表現される。ただ、PMAは語の構造と文の構造との関係の表示を第一に考えるため、[12],[17]の行なった拡張とPMAの間には完全な対応関係があるわけではない。

2.2 PMAを使った事態性名詞の項構造の記述

PMAを使って、支援される名詞の項構造と支援する動詞の項構造との融合をどう記述するかを説明する。

- (20) x が y に恨みをもつ $\Leftarrow x$ が y {を; *}に} 恨む
- (21) x が y に感動を覚える $\Leftarrow x$ が y {*を; に} 感動する
- (22) x が y に反感をもつ $\Leftarrow *x$ が y {を; に} 反感する
- (23) y が x に反感を買う $\Leftarrow x$ が y に {*反感; ケンカ} を売る

(20)のPMAを図2.2に、(22)のPMAを図6に示す。S, O, O2はおのおの主語、直接目的語、間接目的語 NP の存在する位

p0		x**	が**	y**	に**	恨み**	を**	もつ**
p1	x	x*				V		U
p2	が	S	が*			V		U
p3	y	S		y*		V		
p4	に			0	に*	S	(が)	V=ある
p5	恨み	S		0		恨み*		
p6	を	S				0	を*	V
p7	もつ	S				0		もつ*

図5 (20) のPMA

p0		x**	が**	y**	に**	反感**	を**	もつ**
p1	x	x*				V		U
p2	が	S	が*			V		U
p3	y	S		y*		V		
p4	に			0	に*	S	(が)	V=ある
p5	反感	S		0		反感*		
p6	を	S				0	を*	V
p7	もつ	S				0		もつ*

図6 (22) のPMA

置を、V, Uはおのおの動詞、助動詞の位置を表わす^(注6)。

図2.2の p_5 が「恨む」から派生した「恨み」の項構造を指定し、図6の p_5 が派生元のない「反感」の項構造を指定する。項と文法役割の関係は単純なので、 $S=Arg0, O=Arg1, O2=Arg2$ のような対応を考えているが、 $Arg0, Arg1, Arg2$ を明示してもよい。ただ、動詞が非能格用法で現われた場合、 $S=Arg1$ と書くことになる。これは一長一短である。PMASでは判断の容易な $Argx$ の表記を採用している。

次にPMAとPMASの違いを実例を用いて示すことにする。

(24)のPMASは図7に示すものである。

(24) K首相は迂闊な靖国神社参拝で中国政府{の; に; から}反感を買った。

「K首相」には「反感」と「買った」が異なる項役割を与える。事態性名詞「反感」が「K首相」に $Arg0$ を与え、「買っ

(注6): 支援動詞は、支援される名詞を本動詞と見れば、助動詞Uに相当する。

s	K名詞**	は**	述語**	な**	補語**	参照**	で**	中国**	政府**	に**から**の**	反感**	を**	買った**
p1	K名詞*	は*											
p2	Arg0	P: (は)	述語*	(だ)									
p3	Arg1			な*	Arg0[1,2]	Arg0[2,2]							
p4	Arg0	P: (の)			補語**	参照**							
p5	Arg0	P: (の)					で*	中国*	政府*				
p6													
p7													
p8	Arg0	P: (の)						Arg2[1,2]	Arg2[2,2]	に*	Arg1	P: (を)	V: (もたれた)
p9	Arg1[1,3]	P: (の)			Arg1[2,3]	Arg1[2,3]	P: (に)	Arg0[1,2]	Arg0[2,2]	P: (の)	反感*	P: (を)	V: (もたれした)
p10													
p11	Arg0	P: (の)	Reason[1,4]	Reason[2,4]	Reason[3,4]	Reason[4,4]	P: (で)	Arg2[1,2]	Arg2[2,2]	P: (が)	Arg1		買った*

図7 (24)のPMAS

た」がArg2を与える(「反感」は「買った」のArg1)。このような項役割/意味役割の多重性/複合性を表わすことができるのがPMA(S)の最大の特徴である。

PMAとPMASとの大きな違いは格助詞の項構造と非事態性名詞の共項構造を記述するか否か、語の意味関係を表わす要素として何を選ぶか(Argxか{Subj, Obj, Obj2}か)である。PMAは格助詞の項構造も非事態性名詞の共項構造も記述するが、PMASは原則としてそれを省く。

3. PMAで係り受け解析の先へ行く

PMAが係り受け解析に対して上位互換性をもつと上で述べた。これは根拠を挙げずにはなしえない強い主張なので、十分とは言えないが補足的な証拠を挙げることにする。

3.1 「ゼロ照応」の問題をPMAで再解釈する

PMAの効果が顕著なものの一つに照応解析がある。飯田ら[16]のゼロ照応解析の研究では、正解データを作成する際、係り受け解析に明示されない項を表わすゼロ代名詞を便宜的に述語の直前に書き、その先行詞の同定を作業者に依頼した。

具体例を一つ挙げて論点を明確にしよう。次の文(25)で述語部「口にせず」のガ格要素の実現値は[(国史跡・秋田城跡に派遣された)役人]であるが、例えば図8にあるKNPの係り受け解析結果はこの情報を表わしていない。

(25) 奈良、平安時代に中央政府の最北の出先機関だったとされる国史跡・秋田城跡に派遣された[役人]_iは、サケやマスなどを食材にした郷土料理は[こと]_jが[こと]_jが分かった。

この情報を保証するのに[0_iガ]を補完して、0_iの先行詞を同定するのが従来の手法である。別の言い方をすれば、照応解析が必要なのは、一面では係り受けの構造自体によってうまく表示されておらず、並列構造の解析によってうまく補完できていない情報を補うためである。

以下で簡単な例を挙げて要点を説明する。

- (26) [その男]は風邪をひき、一週間寝込んだ。
- (27) [その男]_iは風邪をひき、[彼]_jは一週間寝込んだ。
- (28) [その男]_iは風邪をひき、(0_iが)一週間寝込んだ。

(26)で述語部「風邪をひき」のガ格を実現するのは「その男」である。もう一つの述語部「(一週間)寝込んだ」のガ格を実現するのは、やはり「その男」であるが、この関係は従来の係り受け解析[14]では(「係り先は一つに限る」という原則によって抑制され)表現されない。

この問題は代名詞「彼」が明示的に現われている(27)では生じない。それは、代名詞「彼」が「(一週間)寝込んだ」の主語(ガ格名詞)の存在を保証し、「前方照応詞(anaphor)「彼」が「その男」を先行詞(antecedent)としている」という前方照応(anaphora)の関係(=参照(reference)の一種)を認定すれば、解決可能だからである。ゼロ照応解析は、(26)では(従来の統語解析では)「表層に現われていない」と判断される「(一週間)寝込んだ」の主語を、(28)のように適当な位置にゼロ代名詞0の存在を仮定して、非ゼロ代名詞が現われている(27)と同じ形で問題を解く方法である。

確かに、ゼロ照応解析は(正しく行われれば)求められている解を与える。だが、ゼロ照応解析という名称は解析の実態を表

わしていない。

「ゼロ代名詞の実現の位置を問題にしない」というのは正しい答えを与えるための照応概念の拡張だが、それが照応の定義に沿ったものであるかは疑問である。というのは、**ゼロ代名詞が存在しなければ、そもそもゼロ照応というものは成立せず、ゼロ代名詞が存在するならば、それはどこか特定の位置に存在しなければならないからである。**

「解析の目的は項構造解析であり、その実装法として(たまたま)ゼロ照応解析の形に落としている」というのはNLPの共通理解かも知れないが、それは背景知識のない言語学者が「ゼロ照応解析」という名称を見てすぐに理解できる内容ではない。

これは名称の問題だが、**単なる名称の問題ではない**。誤った名称は誤った特徴づけに通じ、誤った特徴づけが行われているときには、例えば根本的な問題が生じたとき、それを「基本」に戻って解決することを難しくする。

何が問題なのかは、表2に示す(26)のPMAを見れば一目瞭然だろう。従来の構文解析では「(一週間)寝込んだ」の主語が表層に現われていないと判断されるが、PMAでは(統語「移動」はナシで)ゼロ照応詞として不可視化された要素を表層に実現されている要素として記述するので、その必要は生じない。「その男」とSubjとの関係を照応の関係と呼ぶのは不自然である。

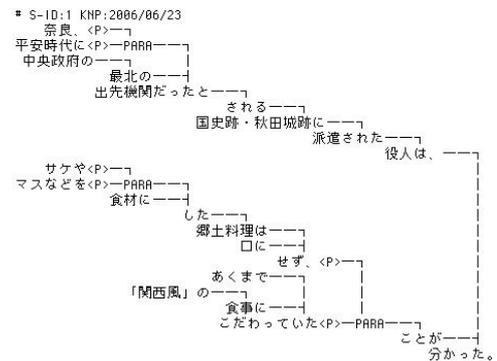


図8 (25)の係り受け解析結果(KNP 2.0の出力)

表2に示した(26)のPMAでは、「その男」の列と述語(e.g., 「風邪をひき」「寝込む」)のある行の交点にSubj=Arg0が現われていれば、「その男」が「風邪をひき」「寝込む」のSubj=Arg0の実現値であることが表わされる。これはSubj=Arg0, Obj=Arg1がゼロ代名詞であることは必ずしも意味しない。

以上のことから確実に言えることは、**飯田らがやっているゼロ照応解析は実質的に範囲を文内に限定しないPMAであり、Subj=Arg0, Obj=Arg1(特に「先行詞から見た」場合に)ゼロ代名詞と見立てているだけだということである**(注7)。

3.2 係り受け関係の拡張の「受け皿」としてのPMA

これはどういうことなのか? 私たちは次のことを主張する:
(29) **ゼロ照応関係解析と項構造解析は、同一の解析の別の面であるか、手段と目的の関係でしかなく、**

s	その男	は	風邪	を	ひき	,	一週間	寝込んだ
p1-5	その男	は	風邪	を	ひき			
p6	α		β ₁ [1,2]	β ₁ [2,3]	β ₁ [3,3]	,	β ₂ [1,2]	β ₂ [2,2]
p7-9	Subj	(が)					一週間	寝込んだ

表2 (26)のPMA(パターンの部分的統合を行なったもの): β₁, β₂は並列構造をなす要素, αはβ₁, β₂に共通の要素を表わす. α=NULLのとき、並列構造は単純な対比を表わす

(注7): PMAの解析の範囲は本来、文内には限られないので、飯田らは境界設定なしのPMAをやっていたと言える。

(30) 飯田らの解析が項構造解析 (argument structure parsing) と呼ばれなかったのは、従来の言語学にその「受け皿」になるような統語構造の記述理論がなかったからである。

これが正しいなら、照応関係タグづけの作業 (http://c1.naist.jp/~ryu-i/coreference_tag.html) と PMA を使った項構造構造解析との統合が可能という見通しが得られる。それと同時に次の言い方もできる。ゼロ照応解析の実質が PMA ならば、照応関係解析/タグづけは一部の言語学者が感じているような「言語学外の仕事」ではなく言語学の問題であり、言語学者が工学者と協力して進めるべき仕事のひとつである。

4. 係り受け解析の先へ行くために: 終りに代えて

係り受け解析の技術も成熟を迎えつつあり、それに伴って限界も見えるようになってきた。実際、限界を克服するために幾つかの拡張 [12], [17] も行われている。それでも係り受け解析の先へ行くために何をすべきなのかはわかっているわけではなく、暗中模索と言ってよい状態である。PMA は係り受けの先に進むために検討すべき可能性の一つであると私たちは考える。この点を図 8 の解析結果を基に考察する。

4.1 PMA の効用

図 8 で KNP は個々の部分の解析に成功しているが、部分の情報の最後の「統合」の段階で誤っていると見える。この解析での誤りは「役人は、」の係り先が「分かった」になっている点、その一点のみである。係り先が「こだわっていた」であれば並列構造の情報を利用して「口にせず」の主語 (i.e. ガ格名詞) が「役人」であることは比較的容易に表現できる。だが、これを KNP に「教える」のは簡単なことではない。ある程度以上に複雑になった解析システムの改良は「アチラを立てればコチラが立たず」的性質をもつものだからだ。

図 8 の解析結果が満足に行くものでないのは、おそらくアルゴリズムが云々とかいう問題ではなく、言語資源がまだ十分でないため、もっと正確に言えば、例えば並列構造解析の精度向上に繋がるような言語事実の記述が、量、質共に十分でないためである。どうして今だにそういう状態なのか? それはおそらく言語学者が言語資源として「使える」ような記述を残して来なかったからである。

平均的な言語学者は、平均的な言語処理研究者より、実際の言語事例を見ていない。驚くべきことだが、これは本当である。大半の言語学者は誰かの論文に出てきた作例をいじくりまわす以外のことはしていない。この結果、言語学者が提供する分析の多くは NLP が要求する一般性、網羅性を欠く。

だが、それ以上に問題なのは、これが事実であることを自覚している言語学者は極く極く一部であるという現状である。このような現状は何とか打破されるべきである。

統語構造の記述モデルが「使える」モデルであるためには**記述の対象となる文を選ばない**ことが必要である。構文解析モデルは (言語学の論文にしか出てくるような単純な作例ではなくて) **実際の文**の構造を記述するためのモデルだからである。PMA がこの条件を満足しているかは未知数であるが、見こみがないわけではない。実際、KNP が取りこぼしている述語/項の関係は図 2 に示した PMA には正しく表現されているし、それは PMA を使えば発見に苦労する類いのものではない。

私たちの期待通りであるならば、PMA は言語資源として使えるような**統語構造、意味構造の高品質な記述のための記述モデル**を提供する可能性がある。それは (少なくとも今のところ) どんな言語現象の説明を与えるものではなく、**説明すべき現象を明確にする (のを手伝う)** だけであるが、それは「弱点」でも「欠点」でもないはずだ。

4.2 PMA(S) の現状と将来への展望

もちろん、これは今の段階ではせいぜい「逸話」的なものである。PMA(S) が本当に「使える」解析法かどうかは、それが十分な量の、一貫した記述を生成できるかどうか試してみないとわからない。現在、そのための小規模な評価実験を行なっている。京大コーパスから選んだ 42 文に対する第一著者の PMAS を見

本にして、第三者 (言語学を専攻する院生) がその記述をどの程度再現できるものかを調査している最中である。この調査の結果を見て仕様を定め、将来的にはレポジトリを儲け、そこに PMA(S) による構文解析の結果を蓄積してゆくことを考えている。最初のデータは <http://c1sl1.hi.h.kyoto-u.ac.jp/~kkuroda/data/pma-samples-of-japanese-v1.xls> として公開している。

文 献

- [1] T. Fontenelle. *Turning a Bilingual Dictionary into a Lexical-Semantic Database*, Vol. 79 of *Lexicographica Series Maior*. Max Niemeyer Verlag, Tübingen, 1997.
- [2] M. Gross. Constructing lexicon-grammar. In B. T. S. Atkins and A. Zampolli, editors, *Computational Approaches to the Lexicon*, pp. 213–263. Oxford University Press, Oxford, 1994.
- [3] P. Kingsbury, M. Palmer, and M. Marcus. Adding semantic annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference*, 2002.
- [4] K. Kuroda. *Foundations of PATTERN MATCHING ANALYSIS: A New Method Proposed for the Cognitively Realistic Description of Natural Language Syntax*. PhD thesis, Kyoto University, Japan, 2000.
- [5] I. A. Mel'čuk. *Dependency Syntax: Theory and Practice*. State University Press of New York, Albany, N.Y., 1988.
- [6] A. Meyers, R. Reeves, and Catherine Macleod. NP-external arguments: A study of argument sharing in english. In *The ACL 2004 Workshop on Multiword Expressions: Integrating Processing*, Barcelona, Spain, 2004.
- [7] J. Pustejovsky. *The Generative Lexicon*. MIT Press, 1995.
- [8] I. Sag, T. Baldwin, F. Bond, A. Copestake, and D. Flinckinger. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, Mexico City, Mexico, pp. 1–15. 2002.
- [9] 黒田航, 井佐原均. 意味フレームを用いた知識構造の言語への効果的な結びつけ. 信学技報, Vol. 104 (416), pp. 65–70, 2004. [増補改訂版: <http://c1sl1.hi.h.kyoto-u.ac.jp/~kkuroda/papers/linking-1-to-k-v3.pdf>].
- [10] 黒田航, 李在鎬. Pattern Matching Analysis (PMA) を用いた日本語の結果構文の共述語分析. 小野尚之 (編), 結果述語の意味論 (仮題). ひつじ書房, 印刷中.
- [11] 小町守, 飯田龍, 乾健太郎, 松本裕治. 共起用例と名詞の出現パターンを用いた動作性名詞の項構造解析. 言語処理学会 12 回大会発表論文集, 2006.
- [12] 河原大輔, 黒橋禎夫, 橋田浩一. 「関係」タグ付きコーパスの作成. 言語処理学会第 8 回年次大会発表論文集, pp. 495–498, 2002.
- [13] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842, 2002.
- [14] 黒橋禎夫, 長尾真. 並列構造の検出に基づく長い日本語文の構文解析. 言語処理, Vol. 1, No. 1, pp. 35–58, 1994.
- [15] 黒橋禎夫, 長尾真. 日本語構文解析システム KNP. <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html>.
- [16] 飯田龍, 乾健太郎, 松本裕治. 文の構造を利用した文内ゼロ照応解析. 言語処理学会第 12 回大会発表論文集, 2006.
- [17] 笹野遼平, 河原大輔, 黒橋禎夫. 名詞句格フレーム辞書の自動構築とそれを用いた名詞句の関係解析. 自然言語処理, Vol. 12, No. 3, pp. 129–144, 2005.