

日本語の移動の制約とその計算機上での実現について
—チョムスキーのバリアー(1986)の観点から—
Constraints of Movement in Japanese
and Parsing as their Implementation

奥村 学*, 柳田優子**, 田中穂積*

Manabu OKUMURA, Yuko YANAGIDA, Hozumi TANAKA

* 東京工業大学工学部

Tokyo Institute of Technology

** Cornell University

梗概

日本語の連体修飾節に修飾された名詞句や主題は、本来の位置から右あるいは左に移動し、移動前の位置にその痕跡を残すと通常考えられている。このような移動を解析する1つの方法は、移動した構成素の痕跡の位置を解析中に発見する機構をパーザに付加する方法である。言語学におけるGB理論は、この移動を説明する普遍的原理を与えている。本研究ではまず、GB理論の枠組で日本語の主題の移動とその制約の問題について論じる。次に、それらの制約を計算機上で実現する方法について述べる。また、制約を実現するための解析機構として、日本語の特徴を考慮した、効率的に痕跡を発見し、移動した構成素と対応付けるパーキング機構を提案する。

Abstract

Relativization and Topicalization in Japanese is considered to be rightward and leftward movement respectively. When movement occurs, trace remains in the original place in phrase structure. A way of analyzing these movement phenomena is to develop a mechanism which detects the places of traces during parsing. GB theory provides principles to explain movement phenomena. In this paper, we first discuss the constraints of topicalization in the framework of GB theory. Second, we describe how to implement these constraints on computers. we also propose the parsing mechanism which reflects the characteristics of Japanese and is able to find traces efficiently and to co-index with moved constituent elegantly.

1 はじめに

日本語の連体修飾節や英語の関係節は、被修飾名詞句や先行詞となる名詞句が右あるいは左に移動したことにより構成され、通常の文と比べ移動した名詞句が欠落した構文構造を取るが、この構文構造中には、移動前の位置にその痕跡が残ると考えられている。このような節を解析する1つの方法は、通常の文を解析するための文法規則以外に、痕跡となっている名詞句を1つ削除したような文法規則を1つ1つ記述することである。この方法の問題点は、「文法規則数が増大し、文法の見通しが悪くなる」ことである。もう1つの方法は、移動した構成素の痕跡の位置を解析中に発見する機構をパーザに付加する方法である。この方法では、記述する文法は通常の文に対するもののみでよいので、上の方法のような文法規則数の増大を回避できる。この方法を用いた文法記述形式としてXG(Extraposition Grammar)[9]などがある。

一方、言語学におけるGB理論[4]は、移動を説明する普遍的原理を与えている。GB理論に基づいた移動の制約をパーキングに用いる研究が最近見られる[3,11]。これらの研究は、

- 痕跡の位置と考えられる箇所の数を減らし探索空間を小さくする。
- XGなどの強力過ぎる文法記述形式に自然な形で制約を加える。

ものである。

本研究ではまず、GB理論の枠組で日本語の主題の移動とその制約の問題について論じる。次に、それらの制約を計算機上で実現する方法について述べる。日本語は、英語などと比べ

- 語順が自由である。

- 文脈から補うことができる情報は文中に陽に記述しないことが多い。

という特徴を持っている。そのため、英語などと同様に、文の構文構造をそのまま反映した文法規則を記述すると、

- 痕跡の位置と考えられる箇所が非常に多くなる。
- 得られる構文解析木の数が爆発的に増大する。

ため、効率を悪化させる。そのため、我々のパーズング機構は、

1. 痕跡の発見は、XG などのように統語的信息を用いるのではなく、意味解析用辞書中に記述された用言(動詞)の必須・任意格情報に基づいて行なう。
2. 構文解析木がごく小数(本研究で対象とする文の範囲では1つ)しか導出しない文法を用いる。
3. 文を読み進む過程で得られる情報をなるべく早期に用いるため、文頭から文末へ情報を伝播する機構 BUP-UTI[1]を解析に用いる。

ものである。1 により、我々の方法では、移動による痕跡と、文脈情報から補完される省略(0 代名詞)を、統一的に扱うことができる。

2 日本語の主題の移動とその制約

本章では、GB 理論の枠組を用いて日本語の主題の移動とその制約の問題について論じる。2.1 節で、ここでの議論で基本となる概念を紹介する。2.2 節で、主題化に関する過去の論文を検討し、その問題点を指摘する。2.3 節で、それらの問題をチョムスキーの下接の条件(Subjacency)と空範疇原理(ECP)を用いて解決する方法を述べる。

2.1 チョムスキーの制約理論

2.1.1 下接の条件(Subjacency)

下接の条件は境界理論と呼ばれることもあり、次のように示されている。

下接の条件:

α -移動を適用する際に、境界接点(bounding node)を2つ以上越えてはならない。

この原理は、のちに Chomsky の Barrier[4]で、境界接点のかわりに障壁(barrier)という概念を使って改良が加えられている¹。

2.1.2 空範疇原理(ECP)

この原理を紹介する前に、まず空範疇という概念について説明する。GB 理論では、空範疇を異なったタイプに区分し、それらについて違った束縛条件を与える。空範疇のタイプは、以下の2つの素性の組合せに対応している。

1. 照応詞的(anaphoric) [+a]
2. 代名詞的(pronominal) [+p]

¹ 興味深い読者は「Barrier」pp.28-31を参考にして頂きたい。

GB 理論では4種類の空範疇が認められている。

-a,-p np-痕跡(あるいは「変項(variable)」)

+a,-p NP-痕跡

-a,+p pro(「小さい」pro)²

+a,+p PRO(「大きい」PRO)

ECPは移動によってできた空範疇1と2のタイプに対して適用される条件である。ECPは[8]³によって以下のように示されている。

ECP(Empty Category Principle):

痕跡は真統率されなければならない。

真統率の定義はここでは[4]のものを使う。

真統率:

α が β を真統率するのは、

(a) α が β を θ -統率するか、または

(b) α が β を連鎖(chain)⁴で統率するときに限る。

θ -統率:

α が β を θ -統率するとは、 α が語彙範疇であり、かつ α が β を直接 θ マークする。

θ -マーク:

α が β を直接に θ マークするとは、 α と β が姉妹であるときに限る。

2.2 日本語の主題

ここでは、主題化に関する HASEGAWA[6],SAITO[10]の説を検討し、その問題点を指摘する。

2.2.1 HASEGAWA[6]の説

HASEGAWA は、関係詞節からの主題に関して以下のことを指摘している。

関係詞節からの主題化が起こるのは、

1. 関係詞節が主文の主語にあり、かつ
2. その中の主語に限る。

例文:

- (1) その子_iは[[_{e_i}_jかわいがっていたね_jが]死んでしまった]
- (2)*? ね_jは[[_{e_i}_jかわいがっていた子供_iが]死んでしまった]

HASEGAWA は、上の例文のような主語/目的語の非対称性を[7]の GCR によって解決している。

GCR(Generalized Control theory):

代名詞的空範疇(2.1.2 節参照)は、もっとも近い名詞に束縛される。

² 3.4 のタイプの違いは、本論文の議論では重要ではないので、説明を省略する。

³ ECP は[8]によって理論化された。

⁴ α が β を連鎖的に統率するとは、 α が β に対してある構造的な関係を連鎖(chain)によって結ぶことであるが、詳しくは Chomsky の「Barrier」pp.16-28 を読んで頂きたい。

これによると、目的語の空範疇は必ず移動によって作られるタイプ1(2.1.2節参照)になるのに対し、主語には、タイプ1またはタイプ3(代名詞的)が可能になる⁵。すなわち、例文(1)の空範疇は代名詞的であり、2.1.1節で示した下接の条件が適用されないが、(2)は非代名詞的であり、下接の条件が適用され非文になる⁶。

-反例-

1. 関係詞節の目的語が主題化される例文

- (3) その金_jは[秘書が[e_ie_j受け取った人_iを]知っている]
(4) その金_jは[[e_ie_j受け取った人_iが]逮捕された]

2. 関係詞節の主語が主題化されない例文

- (5)* 太郎_iは[[e_ie_jきのう会った人_jが]公園にいる]
(6)* 太郎_iは[[e_ie_j外国映画でみた人_jが]来日している]
(7)* 太郎_iは[[e_ie_jきのう会った人_jが]事故に会った]

(3)-(7)の例文を HASEGAWA の例文(1), (2)と比較してみると、主語も目的語も主題化される場合とされない場合があり、これを HASEGAWA の枠組で解決することはできない。

2.2.2 SAITO[10]の説

SAITO は、主語/目的語は共に代名詞的であると主張している。この主張によると、主語/目的語の主題化は制約を受けない。一方、SAITO は、後置詞句(PP)は非代名詞的であり、移動により作られるタイプ1の空範疇を残すと主張している。そこで、PP の場合は移動制約である下接の条件が適用される。以下の文が非文なのはこの理由による。

- (8)* その自転車_jは[ジョンが[e_ie_j山を登った人_iを]知っている]
(9)* その会議_jは[[ジョンが e_ie_j会った人_iが] ジョンを招待した]

-主語/目的語を共に代名詞的とするこの問題点-

1. 主語/目的語を共に代名詞的とすることにより、HASEGAWA の示した主語/目的語の非対称性を扱うことができない(例文(1),(2)参照)。
2. 目的節からの主題化にも主語/目的語の非対称性が存在する事実を説明できない⁷。例文を以下に示す。

- (10) その金_jは[秘書が[e_ie_j受け取った人_iを]知っている]
(11)* その人_iは[秘書が[e_ie_j受け取った金_jを]知っている]

3. 後置詞句(PP)から主題化が可能な例

- (12) その会議_jは[[e_ie_j参加する国_iが]全部で40か国にもなる]

⁵ チョムスキーの束縛理論では、原理Bにより代名詞は統率範囲内で自由になければならない。目的語が代名詞的であれば、GCRにより主語に束縛されるので原理Bに反する。一方、主語は、統率範囲内で束縛されないで、代名詞的空範疇が可能である。HASEGAWA は、日本語の主語はタイプ4のPROであると主張する。

⁶ 下接の条件は移動による制約条件であることに注意して頂きたい。

⁷ SAITO も論文の中でこの問題を認めている。

- (13) その集会_jは[[e_ie_j参加する人たちが]よくパーティを開く]

- (14) そのホテル_jは[[e_ie_jとまる人_iが]金持ちばかりだ]

- (15) その店_jは[ジョンが[e_ie_j行った人_iを]知っている]

上で指摘した問題点は、主語/目的語を代名詞的、PPを非代名詞的とした SAITO の説では解決されない。

2.3 解決案

上に述べた問題点を解決するために本論文では次のことを提案する。

1. 日本語の複合動詞(主動詞+補助動詞)に対する主体者が、意味的に異なる場合、それを構造上に反映させる。
2. 下接の条件にある境界接点という概念を見直す。
3. 空範疇原理(ECP)を導入する(定義は2.1.2節を参照)。

1について述べる。HASEGAWA の例文(1),(2)で、「死んでしまった(好ましくない結果になる)」は意味的に2人の人物-死んだ人物とそれによって影響された別の人物-の存在を暗示的に含有する。これを構造的に示すと、(1)は以下ようになる。

- (16) その子_jは[t_i[[pro_ie_jかわいがっていたねこ_jが]死んでしまった]

すなわち、「死んだ」を主動詞とする節が、[+しまった]を動詞とする節にうめこまれた構造をしている。その結果、関係詞節内の空範疇は代名詞的であり、主題化は一番外側の節から起こる。(16)では、下接の条件はもとより、空範疇は連鎖的に主題と対応関係を結ぶのでECPに反しない。

(2)が不自然なのは、我々の世界で動物が好ましくない結果をしいられたと感じることができないという常識的知識により、(2)を(17)のように表すことができないからである⁸。

- (17)* そのねこ_jは[t_i[[e_jpro_iかわいがっていた子供_jが]死んでしまった]

唯一可能な派生方法は、関係詞節の内部から移動する場合である。

- (18)* そのねこ_jは[pro_i[[e_ie_jかわいがっていた子供_iが]死んでしまった]

ここで、2について述べる。SAITO, HASEGAWA では、日本語の場合、NPは下接の条件にある境界接点になると仮定しているが、ここでは、[5]らによって示された語彙範疇は最大投射されないという説を採用し、日本語の名詞句をN'と仮定する⁹。(18)では、主題がN'を通過することは問題ないが、その着地点(landing cite)に到着するまでにIP(=S)を2つ越えるために非文になる。同じ位置からの主題化が可能な例文(3),(4)の場合は、移動が境界接点IPを1つ越えるだけで、下接の条件を満たすために文法的である。さらに、名詞句をN'とすると、一般に関係詞節からの移動は、IPを1つ越えるだけなので、下接の条件に違反しない。しかし、このN'説を採用するためには、どうして(5)-(7)などは非文になるのかを説

⁸ このことは、目的語が常に非代名詞的であるとするHUANGの主張とも一致する。

⁹ 下接の条件における境界接点は、最大投射XPでなければならない。Fukuiは、英語では、機能範疇は(例えばIPなど)XPに、語彙範疇はXに投射される。一方、日本語は、語彙範疇はXに投射されるのに対し、機能範疇は存在しないと主張する。現在では、日本語にも機能範疇(IPなど)があるという説が有力である。本論文でもIPの必要性を示した。

明でなければならない。ここで重要になるのが、空範疇原理(ECP)である。最後に、この ECP が本論文で扱った例文の文法性をどのように予測できるかを見ていく。

ECP は、空範疇は真統率されなければならないことを示している。空範疇が真統率されるためには、空範疇が θ -統率されるか、連鎖統率されなければならない(2.1.2 節を参照)。

ここでまず、ECP に違反しない例文から見ていく。(1)では、(16)に示してあるように、空範疇は、主題により連鎖的に統率されるので文法的である。(3)-(4)は、空範疇が目的語の場合であり、目的語は常に θ -統率されるので文法的である(同じ目的語でも(2)が非文法的なのは、ECP によるものでなく、下接の条件によるものだというのは前に述べた)。(12)-(15)は、空範疇が PP の場合であるが、PP が θ -統率されるかどうかは動詞によって決定される。(12)-(15)の例はいずれも、PP は θ -統率される。

次に、ECP に違反する例を見ていく。(5)-(7)の空範疇(主語)は、動詞により θ -統率されない。また、これらの空範疇は連鎖的にも統率されないで、非文になる。(8)-(9)の PP の空範疇も、上と同じ理由により非文になる。

以上で、日本語の関係詞からの主題語の下接の条件、ECP を適用することにより、その制約を記述できることを示した。

3 日本語の解析機構

本章では、我々の用いる日本語解析機構について述べる。3.1 節ではまず、文頭から文末へ情報を伝播する機構 BUP-UTI[1]について簡単に説明する。3.2 節では、BUP-UTI を用いた簡単な日本語解析システムによる解析例を示し、どのようにして痕跡を発見し、どのようにして痕跡と移動した構成素とを対応付けるかを説明する。3.3 節では、GB 理論に基づく移動の制約をパーキングに用いたいくつかの研究を概観し、その問題点を指摘する。3.4 節では、2.3 節で述べた制約を、3.2 節で述べる解析メカニズム上でどのように実現するかについて述べる。

3.1 左から右への情報伝播機構 BUP-UTI

本節では、BUP 系解析システム上でトップダウンな情報伝播を実現する BUP-UTI について簡単に述べる。以下では、BUP-UTI の基礎をなす BUP-XG の概要を述べるとともに、BUP-UTI の基本原理を説明する。

BUP-XG は、左外置処理を BUP 上で実現したものである。本節では、トップダウンな情報の流れを実現する基礎となる、BUP-XG におけるスタックの実現法について簡単に述べるとともに、BUP-UTI の基本原理を説明する。なお、BUP-XG の詳細は文献[2]を参照して頂きたい。

3.1.1 BUP-XG におけるスタックの実現法

Prolog 上で左外置処理をトップダウンに実行する方法として、XG (extraposition grammar)[9]がある。この方法では、左外置処理実現のため、スタックを用いている。ボトムアップ法をベースにする BUP-XG では、このスタックを実現するため、後述するように、goal 節を用いた BUP の動作特性を利用している。ここでは、このスタックの実現法について簡単に説明する。

BUP-XG では、スタックの状態を構造体で表し、それを後続する文法カテゴリに次々受け渡していくことで、スタックを実現している。例えば、下の文法規則 1 に対して、2 のように 2 変数を付加したものを考える。2 の各文法カテゴリに付加された 2 変数(イタリック体)のうち、左の変数は、その文法カテゴリの解析を始める時点でのスタック(入力スタック)を表し、右の変数は、解析が終了した時点でのスタック(出力スタック)を表す。

$$s \rightarrow np, vp. \quad (1)$$

$$s(X0, X2) \rightarrow np(X0, X1), vp(X1, X2). \quad (2)$$

2 はトランスレータにより次の 3 に変換され、実際の解析は 3 の BUP 節を用いて行う。3 を用いた解析では、

$$np(G, [], I, X0, X1, XR) \rightarrow \{s(G)\},$$

$$goal_x(vp, [], X1, X2),$$

$$s(G, [], I, X0, X2, XR). \quad (3)$$

np の解析が成功すると、 $X0, X1$ にスタックの内容が返される。そして、np の出力スタック $X1$ が、次の文法カテゴリ vp の入力スタックに渡され、vp の解析が成功すると、 $X2$ に新しいスタックが返される。その後、np の入力スタック $X0$ 、vp の出力スタック $X2$ をそれぞれ s の入力スタック、出力スタックとして解析を続ける。

ここで注意したいのは、np の出力スタック $X1$ を vp の入力スタックに受け渡す点である。この vp に対する入力スタック $X1$ は、np を解析した結果得られた情報であり、goal 節を用いて vp を解析する時に、goal 節のボディで辞書引きされる文法カテゴリまで伝わる(図 1 参照)。これは、np から得られた情報(スタック)が、文法カテゴリ vp を経て、次に辞書引きされる文法カテゴリまで、トップダウンに流れることを意味している。このスタック(トップダウンな情報)の流れを実現するメカニズムが、BUP-UTI の基本原理となっている。次に、この基本原理について詳しく述べる。

3.1.2 BUP-UTI の基本原理

以上のような情報(スタック)の受け渡しに関する考察から、スタックの実現に用いた 2 変数をトップダウンに情報を流すための変数(この変数を以後、TI(Top-down Information)用変数とよぶ)として利用し、この変数で情報を左から右に(上から下)に流すことで、効率の良い解析システムの実現を可能にしたのが BUP-UTI である。

以下では、この BUP-UTI の基本原理について述べる。

BUP-UTI では、前節に示した 1 → 3 のように、ユーザが記述した文法は、BUP-XG トランスレータにより BUP 節等に変換される。そして、それらが図 1 の goal 節とともにボトムアップ構文解析システムとして直接動作する。ここで、述語 'goal' は、トップダウンな情報を用いないで解析を行う goal 節であり、一方、述語 'goal_x' は、トップダウンな情報を解析に用いる goal 節である。

解析は、次のような goal 節の呼び出しにより始まる。

$$? - goal(s, A, [she, smiled], []).$$

図 1: BUP-UTI における goal 節

従って、この BUP の動作特性を生かし、従来の goal 節を修正し、TI 用変数を導入して（図 1 参照）、情報をトップダウンに流すことを考える。前節で述べたように、ある文法カテゴリを解析して得られた情報が、次に辞引きする文法カテゴリに（上から下へ）伝わるのは、図 1 の述語 'goal_x' において、直前の文法カテゴリから伝わった入力 TI 用変数 XI が、次の単語を辞引きして呼び出される述語 C の入力 TI 用変数に受け渡されることによる（イタリック体の箇所参照）。

3.2 BUP-UTIを用いた簡単な日本語解析システム

以下に示す簡単な日本語文法を用いて、例文(3)

(3) その金_jは[秘書が[e_ie_j受け取った人_iを]知っている]

を解析してみよう。解析は文を左から右に読み進みながら行う。

文(s)→後置詞句(_),文(s). (g1)

$$\begin{aligned} \text{文}(S)[X_0, X_2] &\Rightarrow \text{述語}(V_p), \\ &\quad \{\text{interp}((v, n), V_p, X_0, _, X_1)\}, \\ \text{文}(S)[X_1, X_2] &. \end{aligned} \quad (g_2)$$
$$\text{文}(S)[X_0, X_1] \Rightarrow \text{述語}(V_p),$$

$$\{ \text{interp}((v, t), V_p, X_0, S, X_1) \}. \quad (g_3)$$

後置詞句()[X0,X1]==>名詞(N),助詞(P),
 {interp((n,v),P:N,X0, ,X1)}.(g4)

上に示した文法を用いて解析を行なうと、最終的には図 2 のただ 1 つの構文解析木を得るが、その解析過程で情報は概

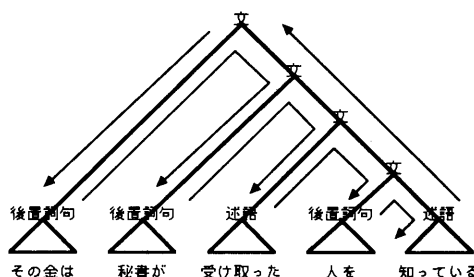


図 2: BUP-UTI を用いた解析の流れ

略矢印のように流れる。この情報の流れを利用し、この日本語解析システムでは、TI用変数を係り句(後置詞句や連体修飾節の用言)の意味情報のスタックとして用いて解析を行っている。すなわち、後置詞句を解析した意味解析結果を次々にスタックに付加して文末の方向(右方)に流し、述語(動詞)が現われた直後にそのスタックと動詞の意味を用いて全体の文の意味を決定している。このスタック中の要素は、未解析で受け句が決定されていない係り句を表しており、これらより後述に出現する任意の受け句に係りうる。

以下、実際の BUP-UTI の実行過程を見ることで、具体的にスタック情報の流れを追い、どのようにして痕跡を発見し、どのようにして痕跡と移動した構成素とを対応付けるかを説明する。解析は、

?-goal(文, A, [その金, は, 秘書, が, 受け取った, 人, を, 知っている], []).

の呼び出しで始まる¹⁰⁾.

1. goal節(g4)を用いて、最初の単語「その金」の辞書引きを行い、その文法カテゴリ「名詞」をヘッドにもつ文法規則を選択する。
2. 文法規則(g4)を適用する。そのボディにおいて、goal(助詞,P)を呼び出す。……これは「は」で成功する。
3. 意味解析プログラム interp を実行する。「花子は」の意味 (topic:その金) を入力 TI 用変数 X0 (内容は[]) の先頭に付加し X1 に返す。
4. 後置詞句の解析が終了する。X1 には[topic:その金]がバインドされている。
5. 文法規則(g1)を適用する。そのボディにおいて、goal(x(文,S,[topic:その金],X2))を呼び出す。

¹⁰ ここでは、説明を簡単にするため、「その金」、「受け取った」、「知っている」なども1つの単語として話しを進める。

6. 次の単語の辞書引きを行う。以下、1～5と同様の解析を繰り返し、その結果、goal_x(文,S',[が:秘書,topic:その金],X2')を呼び出す。
7. 次の単語「受け取った」の辞書引きを行う。そして、その文法カテゴリ「述語」をヘッドにもつ文法規則を選択する。
8. 文法規則(g2)を適用する。
9. 意味解析プログラム interp を実行する。「受け取った」の意味(Vp)と入力 TI 用変数 X0 ([が:秘書,topic:その金])を用いて意味解析を行う。属性 topic を持つ後置詞句(主題)は、連体修飾節から無条件には移動できない(2章で述べた制約参照)。そのため、この時点では、「その金は」が「受け取った」を修飾するという解釈は考えない。すると、結果として

[[([受け取った,(行為者:秘書,-),1),topic:その金],
 ([受け取った,-,1),が:秘書,topic:その金]]

が X1 にバインドされる。これは、ここまで読み進んだ時点で、「受け取った」の行為者が、「秘書」なのか、またはそれ以外の誰なのかの2通りの解釈があることを示している。

10. 再び、ボディにおいて「文」をゴールにした述語'goal_x'を呼び出し、次の「人」を辞書引きし、その結果、文法規則(g4)を適用する。
11. 「助詞」をゴールにした述語'goal'は「を」で成功し、意味解析プログラム interp を実行する。入力 TI 用変数 X0(

[[([受け取った,(行為者:秘書,-),1),topic:その金],

[[([受け取った,-,1),が:秘書,topic:その金]]

)中に連体修飾節となる用言(「受け取った」)があるので、被修飾名詞句(「人」)の意味と「受け取った」の意味を用いて意味解析を行なう。その結果、

[を:(人,行為者!([受け取った,-,1)),slash(を,受け取った),
 が:秘書,topic:その金]

¹¹ が X1 にバインドされる。ここでは、単語「受け取った」に対して、次のような辞書記述があることを仮定している。

[受け取った::
 [が \$ 人間 -> 行為者]
 (から \$ 人間 -> 起点)
 [を \$ もの -> 対象]]

辞書記述中、[]で囲まれた格は、用言(「受け取った」)に対して必須であり、()で囲まれた格は任意であることを表す。入力 TI 用変数にバインドされていた2つの解釈のうち、「秘書が受け取った」解釈が排除されるのは、被修飾名詞句「人」が「受け取った」のどの格も満たすことが

¹¹ (人,行為者!([受け取った,-,1))は、「受け取った」の行為者であるような「人」を表すものとする。

できないからである(満たすことができる「が」格はすでに「秘書」によって満たされている)。また、連体修飾節の用言と被修飾名詞句との関係を決めたので、用言の必須格のうち、欠落している格を抜き出す(slash(を,受け取った)。これは、抜き出した格に対する名詞句を痕跡として発見したことに相当する。

12. 文法規則(g1)を適用し、ボディで「文」をゴールにした述語'goal_x'を呼び出す。「知っている」が辞書引きされ、その文法カテゴリ「述語」をヘッドにもつ文法規則(g3)を適用する¹²。
13. 意味解析プログラム interp を実行する。下に示す「知っている」の辞書と、入力 TI 用変数

[を:(人,行為者!([受け取った,-,1)),slash(を,受け取った),
 が:秘書,topic:その金]

中の「が」、「を」格の名詞句(「秘書」、「人」)により、「秘書が人を知っている」という解釈を得る。

[知っている::
 [が \$ 人間 -> 行為者]
 [を \$ 人間 -> 対象]]

また、「知っている」が主文の動詞であるので、TI 用変数に残された情報

[slash(を,受け取った),topic:その金]

から、発見した痕跡'slash(を,受け取った)'と、移動した構成素(主題)'topic:その金'の対応付けを行なう。そして、最終的に解析結果として、

A = (知っている,(行為者:秘書,

対象:(人,行為者!([受け取った,(対象:その金,-),1)),
 -,1)

を得る。

3.3 関連研究

本節では、GB 理論に基づいた移動の制約をパーキングに用いている最近の研究[3,11]について概観する。これらの研究は、GB 理論で提案されている「c 統御原理」、「下接の原理」などを文法規則に対する制約として記述することで、

- 痕跡の位置と考えられる箇所数を減らし探索空間を小さくする。
- XG などの強力過ぎる文法記述形式に自然な形で制約を加える。

ものである。

しかし、[3]は、痕跡の位置を文法規則中に陽に記述しており、ただ単に痕跡と移動した構成素の対応をとっているだけであり、痕跡を発見する機構をパーザが持っているわけでは

¹² (g2)も適用可能であるが、解析を進めていくと失敗する。ここでは、説明を簡単にするため、その過程については述べてない。

ない。また、[3]は、我々と同様、BUP系解析システムを用いているが、従来のように、下から上へ伝わってきた情報を利用するのみなので、我々のように BUP-UTIを用いる場合と比べ、得られる情報を利用するタイミングが遅くなるので、不要な計算をすることが多くなる。

一方、[11]は、XGに制約を付加したものと考えられるので、痕跡を発見する機構を備えている。しかし、[11]は、構構的にXGの拡張であるため、トップダウンパーザである。そのため、従来主張されているように、左再帰的な文法規則を扱えないという問題を有している。

3.4 制約の実現

2.3節で述べた日本語の主題化に関する制約は、次のようにまとめることができる。

(2') 境界接点としてSのみを考慮して、下接の条件を満足する。

(3') ECPを満足するため、痕跡は動詞(用言)によって θ マークされていなければならない。

2.3節で提案した制約のうち、1は現在実現されていない。これは今後の課題である。以下では、この2つの制約の解析機構中での実現方法について述べる。

下接の条件を検査するには、痕跡が何個の境界接点を通過しているかを数えなければならない。そのため、痕跡を表すデータ構造である slash(C_格, V_動詞)に、もう1つ、通過した境界接点の数を表す引数を追加する。すなわち、

slash(C_格, V_動詞, N_境界接点の数)

のようである。また、境界接点がSであるので、Sを構成する最後の構成要素である用言(動詞)を解析する時点で、slash中の変数Nを1つ加算することにする。したがって、3.2節のステップ11で抜き出された痕跡は、

slash(を, 受け取った, 1)

となる。この痕跡は、文末の「知っている」が解析される時点までに、これ以上境界接点を通過しない。その結果、主題と対応付けを試みる時点でも、下に示すように下接の条件を満足している。

[slash(を, 受け取った, 1), topic: その金]

主題との対応付けを行なう時点で、下接の条件を満足しているかどうかの検査を行なう。すなわち、その時点での slash 中の変数Nを調べ、Nが1であることを確認する。Nが2以上の痕跡は、下接の条件を満足しないので、移動による痕跡とはみなすことができず、省略された0代名詞と判断され、文脈情報から補完を行なう。

次にECPの実現方法について述べる。ECPを満足するためには痕跡は動詞によって θ マークされていなければならない。一般に、 θ マークする名詞句は動詞によって様々であるので、その情報は、それぞれの動詞について辞書中に記述する方法が適切であると考えられる。すなわち、「受け取った」については、

[受け取った::

[が \$ 人間 → 行為者]

[から \$ 人間 → 起点]

[を \$ もの → 対象]] θ

のようである。そして、3.2節のステップ11のように、痕跡を抜き出す時点で、辞書の情報から θ マークされていない格要素は、痕跡として抜き出せないようにすることにより、ECPは満足できる。

4 おわりに

GB理論の枠組で日本語の主題の移動とその制約の問題について論じた。また、それらの制約を計算機上で実現する方法について述べた。我々は、日本語の特徴を考慮した、効率的なパーズング機構を用いた。すなわち、

1. 痕跡の発見は、XGなどのように統語的情報を用いるのではなく、意味解析用辞書中に記述された用言(動詞)の必須・任意格情報に基づいて行なう。
2. 構文解析木がごく小數(本研究で対象とする文の範囲では1つ)しか導出しない文法を用いる。
3. 文を読み進む過程で得られる情報をなるべく早期に用いるため、文頭から文末へ情報を伝播する機構 BUP-UTI[1]を解析に用いる。

ものである。1により、我々の方法では、移動による痕跡と、文脈情報から補完される省略(0代名詞)を、統一的に扱うことができる。

今後の課題としては、以下のものが挙げられる。

- 2.3節で提案した制約のうち、1は現在実現されていない。これは、その実現のためには、2,3よりもより深い意味的知識を必要とするためである。
- 本研究で述べた実現方法は、制約をプログラムとして記述するもので、手続き的であり、ごく普通のユーザが制約を記述できるかという問題がある。制約の記述は宣言的に行ない、それを実行プログラムに変換するようなインターフェースを今後考慮する必要がある。

参考文献

- [1] 奥村学, 田中穂積. BUP系解析システム上でのトップダウンな情報の制御について. 情報処理学会論文誌, 29(11):1043-1050, 1988.
- [2] 今野聡, 田中穂積. 左外置を考慮したボトムアップ構文解析システム. コンピュータソフトウェア, 3(2):19-29, 1986.
- [3] H. Chen, I. Lin, and C. Wu. A new design of prolog-based bottom-up parsing system with government-binding theory. In *Proc. of the 12th International Conference on Computational Linguistics*, pages 112-116, 1988.
- [4] N. Chomsky. *Barrier*. The MIT Press, 1986.
- [5] N. Fukui. *A theory of category and projection and its applications*. PhD thesis, M.I.T., 1986.
- [6] N. Hasegawa. On the so-called 'zero pronoun' in Japanese. *The Linguistic Review*, 4, 1984.

- [7] J. Huang. On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15, 1984.
- [8] H. Lasnik and M. Saito. On the nature of proper government. *Linguistic Inquiry*, 15:235–289, 1984.
- [9] F. Pereira and D. Warren. Definite clause grammar for language analysis - a survey of the formalism and a comparison with augmented transition networks. *Artificial Intelligence*, 13(3):231–278, 1980.
- [10] M. Saito. *Some asymmetries in Japanese and their theoretical implications*. PhD thesis, M.I.T., 1985.
- [11] E. Stabler. Restricting logic grammars with government-binding theory. *Computational Linguistics*, 13(1-2):1–10, 1987.