

2D-6

曖昧性のない文の生成

神岡 太郎 土屋 孝文  
(北海道大学 文学部)

安西 祐一郎  
(慶応大学 理工学部)

1. はじめに

文を解釈する側からすると、解釈の対象となる文が曖昧性を持つということは、それが非文であるということと同じぐらい迷惑なことである。最近では計算機による文生成の研究が盛んに行われるようになってきたにもかかわらず、このような曖昧性のない文をいかにして生成するかという問題が見落とされてきたように思われる。そこで我々は係り受け関係、特に述部とそれが支配する格との間の係り受け関係が一意に決まらないために、意味が曖昧になるような文を取り上げ、そのような文を生成しない文生成機構の構築を試みた。例えば「神岡は、昨日、土屋から、譲渡された、株を、売却した」という文では「昨日」が「譲渡された」に係るのか「売却した」に係るのかが一意的に同定できないので曖昧な文である。本論文では、次の2つの機構を文生成過程に組み込むことによって曖昧でない文を生成する方法について述べる。

- (1) 生成文の係り受け関係と意図された係り受け関係を対応させることにより曖昧性のチェックを行なう。
- (2) 曖昧性がある場合はそれを除去するように語句の出現順位と意味格マークを操作することによって曖昧な文を生成しないようにする。

これらの方法の特徴は、これまでの文生成が一方的に文を生成する側の立場だけから考えられてきたのに対して生成文を解釈する側の立場を考慮した点、すなわち、文生成過程で生成文を生成する側が一度解釈し直しているということである。以下、本論文で示される生成方法は、Sun-3上のQuintus-Prologによってインプリメントされている。

2. 曖昧性のない文を生成する枠組み

生成する1文で伝えるべき内容が決まったと仮定して、これを表したものを意味格構造とよぶ。この意味格構造から生成文の表層に現われない格を省いたものが表層格構造で、次にその表層格構造に意味格マークと語順情報を付加した生成文リスト表現を生成する。生成文リスト表現生成過程では述部とそれが支配する格に曖昧性のない生成文リスト表現だけが出力されるようにする。最後にこの生成文リスト表現をリダクションすることによって生成文を得る。

意味格構造と表層格構造は基本的には係り受け関係に基づいて表現されている。係り受け関係は、係り受けを起す側である係り句と係り受けを受ける側である受け句との間の関係であるが、一般には同一の受け句に複数の係り句に係るので、本論文ではそれらをリスト化した係り句リストと受け句を対応させて係り受け関係を示す。またここで扱う係り受け関係は受け句が述部か体言かによ

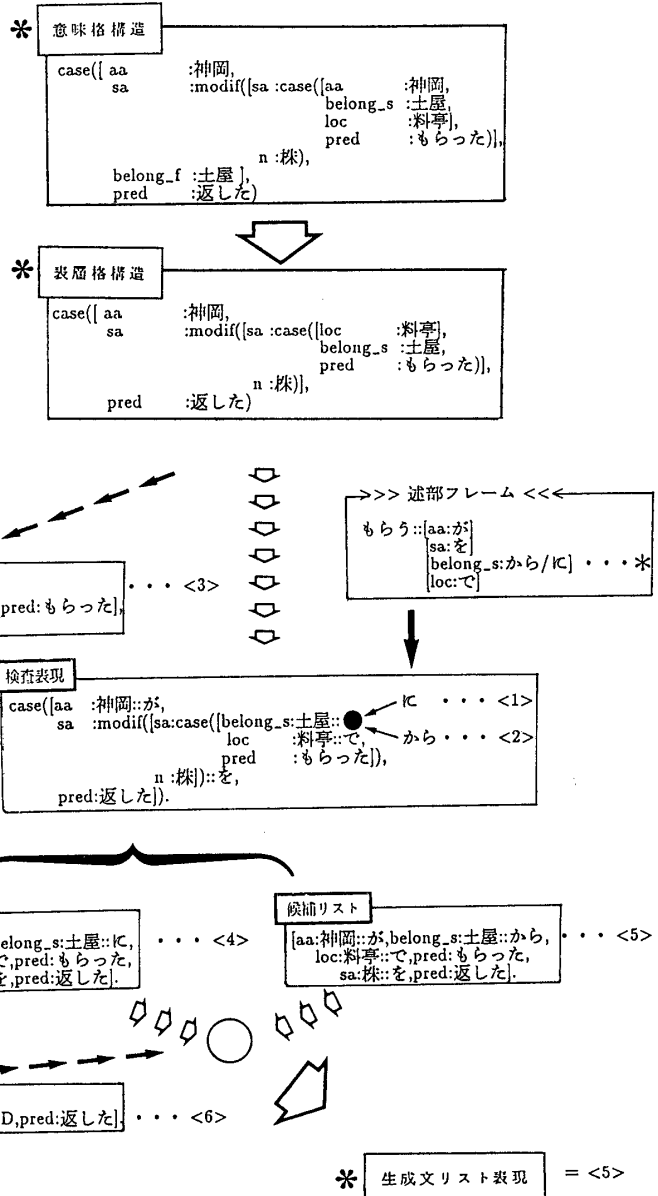


図1 曖昧でない文の生成過程

て分けられ、前者をcase係り受け、後者をmodif係り受けとよぶ。またcaseとmodifは係り受け関係名とよばれる。係り受け関係は次のような表現によって示される。

係り受け関係名(係り句リスト, 受け句)

係り句と受け句は意味情報と値を区切り記号:で結合したものによって示す(特に、case係り受け関係における係り句を格要素とよぶ)。以下、3.と4.では図1を例にして表層格構造生成過程と生成文リスト表現生成過程についてももう少し詳しく述べることにする。

3. 表層格構造生成過程

図1に意味格構造の例を示す。一番外側の構造は、係りリスト[aa:神岡,sa:modif(---,n:株),belong\_f:土屋]の各要素である係り句と受け句「pred:返した」に係り受け関係にあることを示している。意味情報に関しては、例えば「belong\_f:土屋」と受け句「pred:返した」に係り受け関係にあることから「返した」という動作が起こる前の「sa:modif(---,株)」の所有者が「土屋」であることを示している。

意味格構造に対して表層格構造は、意味格構造に含まれる格要素の中で、実際に文の表層に現われる格要素だけを残したもので、意

[確定規則1]: 同一の意味情報を持った複数の格要素が1つの述部に係ることはないで、同一の意味情報を持った複数の格要素が一つの述部の前に出現するとき、その中で出現順位が最も近い格要素をその述部と対応づける。

[確定規則2]: 対象となっている格要素より後に出現した述部の述部フレームを比較し、その格要素の意味格マークを持つ述部を抽出する。もし、それが1つである場合、その述部と格要素を対応付ける。

[確定規則3]: 意味格マーク「は」が付加されている格要素は主節の述部と対応づける。

Y中の要素のリスト内での出現順位を正の整数で表すこととする。

[連鎖規則1]: iを格要素の出現順位、nを述部の出現順位とする。i<nで、i<m<nなるmを出現順位とする述部が存在せず、iを出現順位とする格要素がnを出現順位とする述部に係る事が確定するならば、i<j<nであるjを出現順位とする全ての格要素はnを出現順位とする述部に係る。

[連鎖規則2]: jを格要素の出現順位、mを最後に出現する述部の出現順位とする。jを出現順位とする格要素がmを出現順位とする述部に係る事が確定するならば、i<jであるiを出現順位とする全ての格要素はmを出現順位とする述部に係る。ただし、i<l<jなるlを出現順位とする述部は存在しないものとする。

図2 確定規則群と連鎖規則群の例

意味格構造に対して冗長性を除去するためのルールを適用することによって得られる(ただし表層格構造の表現形式は意味格構造と同じである)。意味格構造から表層格構造へ変換するためには、例えば次のようなルール(R-1)(R-2)が使用できる。

(R-1)入れ子構造を持つ意味格構造において、内側と外側で同一の格要素を持っていれどどちらかの格要素が省略できる。

(R-2)内側と外側の述部間の因果関係から一方の格要素の値が推論できればどちらかの格要素が省略できる。

このような変換ルールの適用の程度に従って1つ以上の表層格構造が生成できる。図1の意味格構造の例については、「もらった」と「返した」のaaの値が一致しているので(R-1)によって内側の格要素「aa:神岡」を省略した表層格構造や、「もらう(もらった)」と「返す(返した)」の因果関係から(R-2)より「pred:返した」に係る「belong:f:土屋」を省略した表層格構造が得られる。図1の表層格構造の例は後者を示したものである。なお、意味格にルールを適用しないでもそのまま表層格とすることも可能である。

ただし、意味格構造から表層格構造へ変換するためのルールは論理的に公理化されるようなものではないので表層格構造が意味格構造の意味を保持しているという保証はできない。当然ながら、本来の文の構造である意味格構造から、他の格要素によって含意されるからという理由で、ある格要素を省略することは生成される文が曖昧になる可能性を増すことになる。しかし本論文では、まず曖昧性の原因を作る危険を侵してもできるだけ冗長性のない表層格構造を生成し、後の生成文リスト表現の生成過程で、語順や意味格マークの割り当てをどう行っても曖昧性が起こるようになってはじめて、表層格構造を生成しなおすという方法を採用している。

#### 4. 生成文リスト表現生成過程

表層格構造に対して、係り受け関係について曖昧性を持たないような語順の確定と意味格マークの付加を行ったものが生成文リスト表現で生成-検査型のアルゴリズムによって作り出される。以下では生成部と検査部のそれぞれについて説明する。

##### 4.1. 生成部

表層格構造中の各係り受け関係について語順を決定した検査表現を生成する。検査表現は係り受け関係名を述語名とし、表層格構造の係り句リストに受け句を挿入したリストを引数とする。リスト内の要素の位置は語順を表し、::は付加された意味格マークを格要素に接続する記号である。例えばmodif係り受けでは全ての受け句が係り句の前に出現するので、受け句を係り句リストの最後尾に付加する。このような語順決定を表層格構造のリストの各段について行なえばよい。

入力 候補リストX、コピーリストY

出力 変数が具体化したY

[step0]: Zの初期値をXとする。

[step1]: Zの要素が述部のみであるとき終了する。

[step2]: Zの頭部の要素について確定規則が適用できるとき、その要素をY中の変数と対応づける。ただし、対応する変数が存在しないときは本アルゴリズム全体が失敗する。続いてstep4へ。

[step3]: Zの尾部についてstep1を適用する。

[step4]: Xについて連鎖規則が適用できるとき、その要素をY中の変数と対応づける。ただし、対応する変数が存在しないときは本アルゴリズム全体が失敗する。step4を繰り返す。

[step5]: Xについて連鎖規則が適用できないとき、step3へ。

図3 候補リストとコピーリストの対応付けアルゴリズム

語順決定手続きのうち、case係り受けについては同時に意味格マークの決定が行われる。意味格マークの決定は、受け句のpred格の値について、述部フレーム内に記述された意味格マーク決定の知識を参照することによって行われる。図1右は述部フレーム「もらう」の記述例である。/は意味格マークの非決定性を示しており、例えば図1の水は「もらう」の格要素belong\_sには、意味格マーク「から」あるいは「に」が付加されることを表している。case係り受けでは意味格マークを付加された格要素(係り句)の語順に非決定性があるので、例外を除いて順列によって1つの語順を選択する。このように語順決定と意味格マーク決定に非決定性が存在するために、後続する検査部に失敗した場合、可能な語順と意味格マークの組合せが全て生成される。図1<1>、<2>は、それぞれ意味格マークとして「に」を選択した場合、「から」を選択した場合に対応する。

##### 4.2. 検査部

検査部では、検査表現がcase係り受け関係において曖昧性がなく、システムが意図したcase係り受け関係として理解されることを検査する。検査のためにまず、表層格構造を係り受け関係caseを中心にリストに変形する。このリストはシステムが意図している構造を示しているの意図リストとよばれる。図1<3>は図1の表層格構造を変形した意図リストである。一方、検査表現に同様な変形を行い、さらに平坦化したリストを候補リストとよぶことにすると、候補リストを解釈し係り受け関係に従って構造化していったのが意図リストと一致すれば検査は成功することになる。図1<4>は意味格マーク<1>を選択したときの検査表現を変形した候補リストである。

この検査の実現のため、意図リストにおいて格要素だけを変数に置き換えたリスト(コピーリストとよぶ)を用意する。候補リスト中のある格要素とある述部の係り受け関係が確定できると判断できた時、コピーリスト中で対応のついた述部を含むリスト内の変数を具体化することにする。ここで、曖昧性がないということは、候補リストの格要素を全てコピーリストの変数と対応づけられることで示せる。なぜなら、係り先が曖昧な格要素はコピーリストの変数と対応づけることができないからである。また、意図リストとコピーリストの要素が一致することを調べることで、システムが意図した係り受け関係として理解されうことを検査できる。図1<6>は意図リスト<3>のコピーリストである。

このような係り先を決定する規則は、ある要素について係り受けの制約から係り先を決定する確定規則と、ある要素が確定することに連鎖して係り受けの制約から他の格要素の係り先を決定する連鎖規則とからなる。図2に確定規則群と連鎖規則群の例を示す。図3に候補リストとコピーリストを入力として、確定規則と連鎖規則の適用によって、コピーリストの変数へ割り当てを行うアルゴリズムを示す。このアルゴリズムが出力するコピーリスト(このとき変数は全て具体化している)と意図リストの要素の対応がつくとき、検査部は成功し候補リストを解として出力する。

図1の例について説明すると(図中の白丸に対応する)、<4>の2番目の要素「土屋::に」と3番目の要素「料亭::で」には、確定規則も連鎖規則も適用できず、本アルゴリズムが失敗する。生成部へ戻りして生成される候補リスト<5>とコピーリスト<6>について本アルゴリズムを適用する。候補リスト<5>の2番目の要素「土屋::から」の係り先は図2の確定規則2によって「pred:もらった」と確定される。この格要素が確定したことによって、連鎖規則1より<5>の3番目の要素「料亭::で」も「pred:もらった」と対応づけられる。従って本アルゴリズムは成功し変数が具体化されたコピーリスト<6>が出力される。<6>と意図リスト<3>の要素について、リストが集合を表わしていることとすると全ての要素に対応がつくので候補リスト<5>が生成文リスト表現となる。