

## 2D-4

## 素性構造を入力とする英文生成

上田良寛、小暮 潔  
ATR自動翻訳電話研究所

## 1. はじめに

生成システムの実現部では、文法規則、辞書を解析システムと共有させる双方向システムが考えられている([1]など)。我々は、この方法を、ATRにおいて研究を行っている意図伝達方式[2]に結び付けることを検討している。ここでは、このような方針で設計している対話文翻訳システムの生成部について述べる。

## 2. 生成部における課題

生成部における課題と、それが双方向文法生成システムでどう解決されるかを示す。

1) 文法記述の容易さ。コントロールを明示的にしなければならない場合、文法記述には文法の全体の構造を把握する必要があり、文法記述は容易ではない。文法の一貫性を保つためには、解析用文法を変更した場合には、生成用文法においても同じ変更を行わねばならない。このため、文法・辞書を共有する双方向文法システムを採用する。

解析文法を用いて生成を行うと非文も同時に生成してしまう場合があるが、これは解析文法が十分でなかったことを意味している。このように、双方向文法システムは解析文法の検証にも用いることができ、解析文法の開発にとっても有効である。

2) 発話意図の反映。話し言葉においては、発話意図は重要な意味をもっており、翻訳結果にこれが正しく反映されなければならない。発話意図は言語に依存せず、原言語で適切に抽出すれば、そのまま目的言語へ渡すことができる。この方法を意図伝達方式と呼んでいる。コントロールを明示的に示したシステムでは、このようにして得られた発話意図の情報と表層の文との関係を明確にできないおそれがある。

双方向文法生成システムでは、宣言的な文法によりこの問題が解決される。さらに、過大生成した複数の生成結果から、考慮に入れていなかった発話意図が発見される可能性があり、意図伝達方式の研究に都合がよい。また、意図伝達方式では発話意図が適切に抽出されているので、Appeltが指摘するような意図を考慮して入力意味構造を作らねばならないという問題[1]はない。

ATRでは、解析の基礎となるメカニズムとして素性構造とその単一化を採用しており、解析結果の発話内容および発話意図は、素性構造で表現

される[2]。変換部[3]の入出力、生成部への入力も素性構造を採用する。

## 3. メカニズム

解析用文法は、CFG規則+素性構造の形式の制約で表されている。意味構造を表す素性構造を与えると、CFG規則を解析とは逆方向に適用しながら、句構造をつくる。この組立てはトップダウンに行われる。この句構造の各ノードには素性構造が与えられており、CFG規則を適用する際に親のノードおよび規則の制約から分配される。このため、句構造のノードは素性構造を共有することになる。

書き換え規則を逆に適用した後に、意味情報が適切に分配されないと、その後の書き換えは不可能になる。例えば、次のルールを考える。

$$V \rightarrow V X \quad (1)$$

$$(0 \text{ TRANS}) = (1 \text{ TRANS})$$

$$(1 \text{ SUBCAT REST FIRST}) = (2)$$

これは、「 $V \rightarrow V NP$ 」や「 $V \rightarrow V COMP$ 」に対応する規則であるが、 $X$ に対する素性構造の意味構造(TRANS素性)は与えられていない。これは、 $X$ 全体が $V$ のSUBCATに入り、この $V$ の中で詳細な記述がなされているからである。実際は辞書に記述されているので、 $V$ のほうの書き換えが進み、語彙まで書き換えた段階で初めて、 $X$ にTRANSが供給される。 $X$ のほうの書き換えは、それまで待たねばならない。このような待ち状態にあるノードは、生成途中の句構造ごとにリストで保持されている(ペンディングノードリスト)。

図1に与えられる素性構造に対する生成例を、図2を用いて説明する。

```
[[HEAD [[FORM FINITE]
[TENSE PAST]
[MODE DECLARATIVE]]]
[TRANS [[RELATION PERFECTIVE]
[ARG1 [[RELATION "SLEEP"]
[ARG1 [{"ARTHUR"の意味素性}]]]]]]]
```

図1 入力素性構造

1) 右側素性の素性構造をもつノードSは①の書き換えでNP, VPのノードを作る。NPのノードには意味構造が与えられない。VのSUBCATに入り、Vにより詳細な指定がされるまで待ち状態となる。

2) VPノードに「 $V \rightarrow VX$ 」が適用され(②)、 $X$ ノード(図では $VP_3$ )は、 $VP_2$ のSUBCATに追加され待ち状態となる。

3) VPノードはVに書き換えられ(③)、さらにターミナル“had”に書き換えられる(④)。“had”には、SUBCATの2番目( $VP_3$ )の意味記述と同じということ、 $VP_3$ が過去分詞であることが記載されて

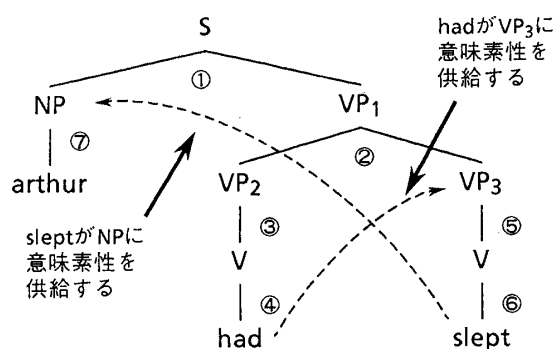


図2 生成例 ("arthur had slept")

いる。これでVP<sub>3</sub>に意味構造が与えられたことになり、VP<sub>3</sub>の書き換えが可能になる。

4) X (VP<sub>3</sub>)が、⑤、⑥と書き換えられ、“slept”まで書き換えられる。ここで最初のNPに意味構造が与えられ、“arthur”への書き換え(⑦)が可能になる。

#### 4. 書き換えに対する条件の付加

単に「書き換え+制約」で表されているだけでは、適用すべきルールを選ぶのには不十分な場合がある。先の例と似たような書き換え規則に、

V → V P (2)

(0 TRANS) = (2 TRANS)

(2 SUBCAT REST FIRST) = (1)

という規則がある。この2つの書き換え規則からえられる素性構造の意味構造は、どちらも述語1つとそれに対する2つの項からなるが、その違いは書き換え規則には記述されていない。例えば、図3のような構造では、“AT”のレベルの素性構造に対しては「V → VP」を、“PLAY”のレベルの素性構造に対しては「V → V X (NP)」を適用しなければならない。

[RELATION "AT"	
ARG1	[RELATION "PLAY"
	ARG1 [ダイエーホークス]
	ARG2 [野球]
ARG2	[平和台球場]

図3 素性構造の例

このように、書き換え規則だけでは、どの規則が適用されるかを決定できない。もしそのまま適用した場合、適用が進んでいってターミナルに達した段階で初めて正しかったかが分かる。また、同じルールを何度も適用してターミナルに達しない場合もある。

これを避けるために、書き換え規則に、親の素性構造の形を(おおまかに)指定する記述を設ける。これをASSERTIONSと呼ぶ。意味情報のうち辞書項目で決まるもの(動詞、前置詞ではRELATION素性)にはTYPEを指定できる。

書き換え規則「V → VP」には

(:EQ (:TYPE <0 TRANS RELATION>) P)

を、書き換え規則「V → V X (NP)」には

(:EQ (:TYPE <0 TRANS RELATION>) V)

をASSERTIONSとして記述する。

#### 5. 実験

現在、Symbolics 3600シリーズ上で生成プロトタイプを作成するとともに、D-PATRで用いられた英語文法をもとにした文法で実験を行っている。現在は、扱う言語現象を、ギャップがないもの(関係詞節、疑問文などを除いたもの)に絞っている。このため、発話意図を適切に抽出する段階には至っていない。図2は、この実験で生成した例である。

なお、日本語の解析文法[4]に多少の変更を加え、日本語生成実験も行っている。この文法では、発話意図を含めた解析/生成が可能である。

#### 6. 問題点と課題

実験を通じて問題点を検討した結果、次のような解決すべき課題があることが分かった。

##### 1) 効率の改善

現在の実装では効率は非常に悪い。これは、ある木構造のノードに対して適用した結果が2つ以上ある場合は、木全体をコピーしているため、全く同じ素性構造をもつノードに対して同じ書き換え規則を適用することや、細かい品詞分類を行っていないため書き換えが辞書項目まで進まないことと失敗したかどうか分からないことなどが原因としてあげられる。これを改善するために、構造共有化の導入や品詞の細分類(Longman 文法コードなどの利用)を考えている。

##### 2) 文法の強化

現在の文法は言語現象の一部しかカバーしていない。その一つにギャップの扱いがあり、生成メカニズムの変更も含めて検討する必要がある。

素性構造の形式にも問題がある。修飾の表現はその一つである。英語文法では、図1の素性構造に示すように、被修飾部全体を包みこむ形で表現されている。これは語順を指定することにもなってしまうし、スコープなど意味表現の意味にも関連してくるので、総合的な検討が必要である。

#### 7. おわりに

以上、現在検討を行っている生成システムについて述べ、その可能性を示した。今後、残された課題を解決する一方、将来的には音声との結合も研究していきたい。

本研究の機会を与えて下さいましたATR自動翻訳電話研究所 樽松 明 社長、および、有益な助言をいただいた相沢 輝昭 室長を初め言語処理研究室の諸氏に感謝します。

#### 参考文献

- [1] Appelt, D. E.: "Bidirectional Grammars and the Design of Natural Language Generation", TINLAP-3, 1987.
- [2] Kogure, K., et al.: "A Method of Analyzing Japanese Speech Act Types", Proc. 2nd MT Conference, 1988.
- [3] 長谷川、小暮: 「日英対話翻訳における意味構造変換手法」、情報処理学会大38回全国大会, 1989.
- [4] 吉本、小暮: 「日本語端末間対話解析のための句構造文法」、情報処理学会大37回全国大会, 1988.