

2T-2 コンストレイントに基づく日本語の係り受け解析

杉村 領一、三吉 秀夫、向井 国昭

(財) 新世代コンピュータ技術開発機構

1. はじめに

一般に日本語の解析を行う場合、文節間の係り受けを基本にする手法(1)がよく用いられる。その問題になるのが、ある程度意味的なチェックを行っても係り受け関係に曖昧性が依然として存在する事である。従来このような場合、ある種のヒューリスティクスを適用して、解釈を1つに絞るという手法をとることが多い。それには、現状の技術レベルでは文脈理解のメカニズムはあまり解明されていず、機械翻訳のような応用システムでは解釈を一意に決める事が余儀なくされており、本来なら語用論的な知識を利用して更に精密な解析を行うべき所を、ヒューリスティクスに頼らざるを得ないという背景がある。従って、解釈の精度も必ずしも満足のゆくものではない。しかし、今後文脈理解の研究が進展し、そのメカニズムが解明されると、構文的な曖昧性を意味的な制約によって減少させるのと同様に、意味的な曖昧性を語用論的な制約によって更に減少させることが可能になる。本稿では、単文内での意味解析終了時の曖昧性を文脈情報との相互作用により解決する手法を提案する。本手法では、係り受けに関する各種のコンストレイント(制約)をCIL(2)の遅延実行機能によって記述している。これにより係り受けの制約を「静的」に記述することが可能になった。また、本手法により意味解析と文脈解析の接点に対する見通しも得られ、言語理解の精度を高めることが期待できる。

2. 係り受けのコンストレイント(制約)

本手法では、制約として、①係り受け関係間の制約、②係り受けの関係にある文節同士の文法的な制約、③係り受け関係にある文節同士の意味上の制約、④文の部分情報と文脈の情報間の制約、以上4種を静的に定義する。

単一文法の枠組みでは、②③の静的な記述が可能であるが、本手法では、日本語の依存構造上の制約である④をも静的に記述する事が可能である。

④については、現在のところ「文の意味と文脈の意味の間に、コピーレンシが成り立つ」という単純な制約だけを仮定しており、機械構造

についての研究は今後の文脈処理の課題である。

2. 1. 係り受け関係間の制約

文節間の係り受け関係間の制約として以下の3つを考える。

1) 非交差性

係り関係同士が交差してはならない。例えば下記のような関係は許さない。



2) 絶対修飾性

文末の文節を除く全ての文節は後方の文節を1つ以上必ず修飾する。係り受け関係は意味を強く反映しているので、係りの唯一性は成り立たない。

3) 文末文節非修飾性

文末の文節はそれが修飾する文節を持たない。

2. 2. 文節間の文法上の制約

修飾-被修飾関係にある文節には、連用修飾、連体修飾等の文法上の制約がある。この制約は、修飾-被修飾関係を否定する事しかできない。

2. 3. 文節間の意味的な制約

修飾-被修飾関係にある文節には、①格関係による制約、②副詞素性の一回性等がある。この制約も、修飾-被修飾関係を否定する事しかできない。

3. 制約による係り受け解析の処理手順

以下、例を用いて処理の概要を説明する。

STEP-1. 係り受け行列作成

まず、入力文が与えられると、これを文節単位にまとめ、各文節の係り受けの関係を示す係り受け行列を以下の様に作成する。

入力文 : 公園で、太郎が、走っているのを、見た。
文節No. : 1 2 3 4

—係り受け行列—

1	U1	U2	U3
	2	U4	U5
		3	U6
			4

上記の行列でU1からU6は論理変数であり、文節間の係り受けの状態を示す。例えば、U2は第1文節から第3文節への係りの状態を示し、1の場合係りがあり、0の場合係りが無いことを示す。変数に値が代入されていない時は、係りの有無が決定されていない。

STEP-2. 係り受け関係間の制約の適用

STEP-1で作成した行列の論理変数の間に、2.1で述べた制約を適用する。本例の場合、係り受け関係間の制約は次のように定義される。但しセミコロン(;)は論理和を、コンマ(,)は論理積を表す。

- 1) 絶対修飾

$$(U1;U2;U3),(U4;U5),U6 = 1 \text{ (true)}$$
- 2) 非交差

$$\neg(U2,U5) = 1 \text{ (true)}$$

CILには論理変数間の命題論理の制約をそのまま記述する機能があるので、これにより、上記制約を定義できる。

上記の絶対修飾と非交差の制約は互いに関係している。たとえば、U2が1になると、2)の制約より、U5は0になる。この結果、1)の制約よりU4は1になる。そして、これらの計算は全てCILによる論理変数間の制約定義により、自動的に行われる。また絶対修飾性によりU6は必然的に1になる。

STEP-3. 文法/意味的制約の適用

次に文法/意味的な制約を解く。本例の場合では、「公園で」は「太郎が」に係りえない。このため、U1は0となる。その結果この文に関する部分情報PI-1が得られる。係りえる部分は変数のまま残しており、これと意味構成の間の制約を述語'u_sem'で定義する。

1	0	U2	U3
	2	U4	U5
		3	1
			4

PI-1:

```

J#(rel/見た,
  agent/UV1:u_sem(UV1,太郎,U5),
  obj/({rn}/走っている,
    loc/UV2:u_sem(UV2,公園,U2),
    agent/UV3:u_sem(UV2,太郎,U4),
    pol/1),
  loc/UV4:u_sem(UV4,公園,U3),
  pol/1
)

```

述語'u_sem'は、係り受けと、係り受けから得られる部分意味との間の以下の制約を示す。

- 1) 係り受けから得られる部分意味が、文の意味として成立するなら、係り受けはある。
- 2) 係り受けがあるなら、係り受けから得ら

れる部分意味がある。

```

u_sem(A,B,Sign) :-
  1) freeze(A,pv(F,(A == B,Sign = 1:true))),
  2) freeze(Sign,pv(F,(Sign == 1,A = B:true))).

```

前述したように、文法/意味的な制約は、係り受けが有りえない事は示唆するが、必ず成立することは示唆しない。また、係り受けの制約は、係りの有無双方を示唆できるが、一意に係りを決定できない。

STEP-4. 文と文脈情報間の制約の適用

STEP-3の結果に、文の意味と文脈上の意味の制約を適用し、更に曖昧さを減少させる。ここでは、以下の制約を考える。

- 1) 文の意味と、文脈の意味の間には、コピーレンシが成り立つ。
- 2) コピーレンシが成り立たない時には、文脈の切り換えがおきている。

本例の場合、「走っている場所は公園である。」という文脈情報CI-1が得られたとしよう。

CI-1:

```

J#(rel/走っている,loc/公園,pol/1)

```

部分情報PI-1と文脈情報CI-1とのマッチングを取ると、公園がloc/に代入され、係りの制約から「太郎が」が「走っている」に係り、その結果新たな部分情報PI-2が得られる。

PI-2:

```

J#(rel/見た,
  agent/UV1:u_sem(UV1,太郎,U5),
  obj/({rel/走っている,
    loc/公園,
    agent/太郎,
    pol/1),
  loc/UV4:u_sem(UV4,公園,U3),
  pol/1
)

```

マッチングが取れないときには、文脈とコピーレンシが満足されていないので、新しい文脈を定義して、文の部分情報を入れる。

4. おわりに

本稿ではCILのコンストレイントによる日本語の係り受け解析、及び文脈解析との接点を述べた。係り受け情報から部分情報を求める方法等、未解決の問題も多いが文脈解析への糸口はつかめた。今後、本方式を用いて談話理解システムを開発してゆく予定である。

【参考文献】

- [1] 吉田、二文節間の係り受けを基礎とした日本語文の構文分析、信学論(D), 55-D,4, 447.
- [2] 向井ほか、CIL言語マニュアル、ICOTテクニカルメモ, TM-242, 1986.