

# 日本語文における2分木構造から 依存構造の導出

4H-2

河田 浩二

酒井 克英

牧野 武則

東邦大学理学部情報科学科

## 1 はじめに

依存文法 [1] は、導きだす依存構造が、格文法で表現した意味構造に類似していること、非終端記号のように人工の記号を用いなくて語と語の依存関係を表すなど利点が多い。実用のシステムでも採用され [2]、その有用性が検証されつつある。また、GPSG など単一化文法 [3] では語彙項目に語彙に係わる構文情報を集め、単一化操作で簡素な解析を実現している。

提案する語彙依存文法 [4] は、言語に共通の係り受けのモデルを設定し、語彙に依存する依存情報を語彙項目に集中し、さらに依存構造をネットワークで表現することで、簡素で柔軟な解析、生成を実現することを目的としている。しかし、実際に解析を行ったところ [5]、大局的な構造の把握に適さない、規則の自由度のため係り受けのスコープが限定できない、規則不足による解析の失敗する等の問題が起った。そこで、2分木構造 [6] を用いこれらの問題を解決し、より高速で正確な依存構造の導出法について紹介する。

## 2 係り受けモデル

依存文法は、語と語の依存関係から文の構造が決められる。この時、依存関係にある2つの語の一方を主要素、もう一方を従要素と呼び、例えば、2つの語が動詞と名詞であれば、主要素が動詞になり、従要素が名詞となる。この主要素と従要素の間関係は係り受け情報として与えられる (図1)。例えば、形容詞の場合、通常主要素として名詞をとり、その係り受け情報は「連体修飾」を持つ。この係り受け

情報は各語の情報、つまり語彙項目によって決定される。

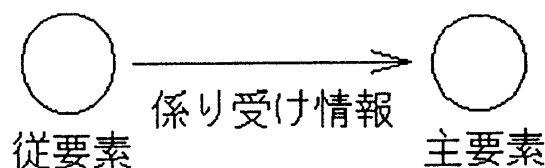


図1: 依存関係と係り受け情報

## 3 語彙項目

語から得られる情報を語彙項目といい、大きく2つに分けられる。一つはその語が主要素である場合、どのような従要素が係り受け可能であるかの情報 (主要素情報)、もう一方は従要素である場合、どのような主要素と係り受け可能であるかの情報 (従要素情報) である。主要素情報には、品詞と格情報があり、それらに対応する従要素情報として、修飾と格がある。

### 3.1 品詞

すべての語が持っている情報である。今回の日本語の実験では、名詞、動詞、形容詞、副詞、指示詞、連体詞、助詞、接続詞、特殊の9つに分類した。

### 3.2 格情報

格情報とは、その語に従要素として係り受け可能な格を示すものである。格情報は、動詞と形容詞が持っている。これらの格情報は接続可能な格の組み合わせにより動詞タイプとして定義し、それを基に動詞をグループ化する。例えば、タイプ1は「は」や「が」という格が係り受け可能で「優れる」、「終わる」やすべての形容詞がここに分類される。

### 3.3 修飾

修飾は、語の品詞、活用、格マーカーから得られ、名詞、代名詞に係り受ける連体修飾、及び、動詞、

No	文節	エントリ	品詞	機能	活用
1	青い	青い	形容詞	*	連体
2	鳥が	鳥	名詞	が	*
3	,	,	特殊	,	*
4	向こうに	向こう	名詞	に	*
5	飛んで	飛ぶ	動詞	*	連用
6	行った	行く	動詞	*	終止
7	。	。	特殊	。	*

(1(2(3(4(5青い鳥が)5,)4(6向こうに飛んで)6)3行った)2。)1

表 1: ワードリストの一部

形容詞、副詞に係り受ける連用修飾がある。

### 3.4 格

格は、修飾の情報を意味的により限定するものである。日本語の場合、主に格マーカーによって得られ、連体修飾では「の」、連用修飾では「が」、「を」、「から」、「提題」などがある。また、各格を持つ語はそれぞれの主要素に対し次のような意味を持つ。「が」を持つ語は、主要素のその行動の主体となる。また、「を」を持つ語は行動の対象であり、「から」や「へ」を持つ語はその行動の方向を示す。

## 4 構文解析

構文解析の入力として2分木構造でスコープ付けされたワードリストを用いる。表1は、例文「青い鳥が、向こうに飛んで行った。」のワードリストである。ここで、2分木構造で表されるスコープは、主要素となる語において従要素として取り得る語の範囲を示している。そして、このワードリストに次のような前処理を加える。まず、品詞とエントリから動詞タイプを付加し、次に、機能と活用から修飾と格を決定する。

構文解析は、この前処理がなされたワードリストのトップの語から係り受けを開始して行き最後の語の係り受けが決定された時点で解析終了となる。係り受けは次のような手順で行われる。

1. リストのトップから順番に係り受けを決定する語を選ぶ、これを対象語とする
2. 対象語の係り受け先の語の候補をスコープの最も近いところから選ぶ
3. 対象語の従要素情報と候補の主要素情報を照らし合わせる
4. 係り受けが可能なら、対象語の係り受け先はその候補となり 1. へ、不可能なら候補選抜のスコープを広げ 2. へ
5. 総てのスコープの候補で係り受けが失敗した場合、最も近いスコープの語に係り受けを決定する
6. 総ての語について係り受けが決定したら解析終了

実際に例文の2番目の文節「鳥が」を例にして係り受けの決定の様子を示す。対象語「鳥が」の最も近いスコープは、5番の括弧だがこれは自分自身なので一つスコープを広げる。次は4番の括弧だが候補は「、」であり、これには係り受けないので更にスコープを広げる。次は3番だが、ここには候補は無い(なぜなら6番の右括弧と3番の右括弧の間に語が無い)。そして、2番のスコープの候補「行った」は、対象語の格「が」を受ける事が可能なので、文節「鳥が」の主要素は文節「行った」となる。この様に例文の総ての語に対して解析を行った結果は図2のネットワークである。

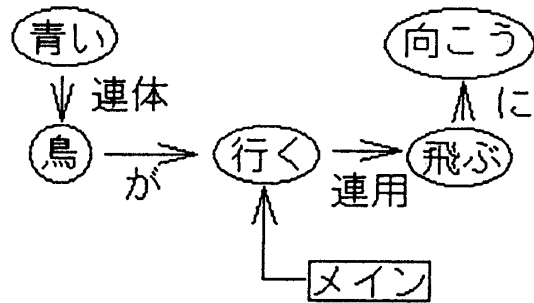


図 2: 例文の構文解析結果を示すネットワーク

## 5 おわりに

本稿では、2分木構造によって示されたスコープでガイドされた語彙依存文法による構文解析手法を提案した。その結果、係り受けのスコープが明示され、正確で曖昧性の少ない解析結果が選られた。また、解析アルゴリズムは one-path であり、高速かつロバスタな解析が実現できる。今後、より複雑な文に対して適用範囲を広げて行く予定である。

## 参考文献

- [1] 児玉「依存文法の研究」研究社出版(1986)
- [2] Muraki, K., "Augumented Dependency Grammar for Language comprehension," Conference of European Chapter of ACL 85(1985)
- [3] 黒川「自然言語処理入門」近代科学社(1988)
- [4] 牧野、納富「語彙依存文法について」情報処理学会自然言語処理研究会、82-14(1991)
- [5] 河田、牧野「語彙依存文法における語彙項目の記述について」情報処理学会第49回全国大会、3G-08(1994)
- [6] 酒井、河田、牧野「日本語文の経験則による2分木解析」情報処理学会第51回全国大会、4H-01(1995)