

大局スタックを用いた依存構造の解析

1 R-8

上原 悟

山室 由紀子

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、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、その有用性が検証されつつある。また、GPSG など単一化文法 [2] 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。

語彙依存文法 [7] 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、生成の実現を目的としている。しかし、実際には、大局的な構造の把握に適さない [8]、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等の問題が起こったため、日本語の持つ基本構造を経験的に得た小数の規則を用いて依存構造を導出する方法を示した [9]。

大局的な構造の把握のために、文構造を接続助詞等の機能語が持つ「文中での切れ目の強さ」や「係り先の制約」を行ない、文内構造を推定する試みがなされてきた。また、節の内部に含み得る要素による接続助詞等のレベル分けする試みも様々な形で行なわれている [3] [5]。そこで、機能語レベル情報を取り入れ、さらに、機能語に依存する用言の性質を捕えた依存構造の導出法を提案する。とくに、提題化された語の共有を考慮するため大局スタックを用いて、大局構造の導出に焦点を当てた。

2 語彙依存文法

2.1 係り受けモデル

依存文法では、語と語の依存関係から文の構造が決められる。依存関係にある 2 つの語の一

方を主要素、他方を従要素と呼ぶ。主要素と従要素の間の関係は係り受け情報として与えられる (図 1)。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。

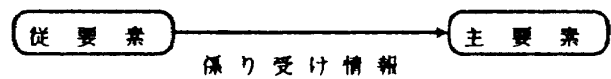


図 1: 依存関係と係り受け情報

2.2 語彙項目

要素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。一つはその要素が主要素となる場合、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。主要素情報は、品詞と格情報であり、それらに対応する従要素情報として、修飾と格がある。

3 機能語の大局構造指示

文構造を解析するにあたり、文法的な構文情報 (格、修飾) だけから、構文解析を行なうとあいまいさが残る。このあいまいさを減少させるために適切な文内構造を推定する試みがなされてきた。接続助詞や用言の活用語尾、それ自身に文構造を推定する力があることに焦点を当て、機能語のレベル分けにより文構造を決定している [5] [6]。本論も機能語の働きを考慮し局所構造を導入する。局所構造内においての主述語が持つ依存関係から局所構造同志を結ぶ。さらに、主題共有情報をもつ局所構造は、大局スタックを用いて、つながりを持つすべての局所構造の依存関係を調べ、大局的な依存構造を導出する部分を担う。

4 構文解析

形態素解析された情報をいわゆる文節単位の要素に分け、語彙依存文法に基づき情報を付加させる (前処理)。この情報を要素リストと呼ぶ。要素リストを元に 2 段階処理の構文解析を行なう。

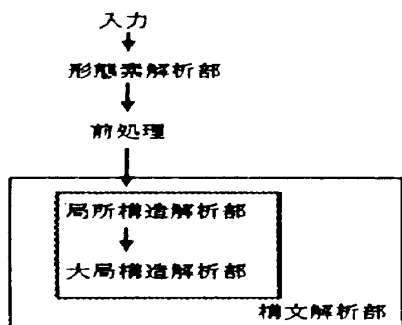


図 2: 解析の流れ

● 局所解析部

- 入力文に対して、機能語のチェックを行い、局所構造の範囲を決定。
- 局所構造内で依存ネットワークを導出。
- 局所構造内において、主述語情報を保持。
- 提題化された要素を含むなら、大局スタックを用いて、大局構造解析部へ。

● 大局解析部

- 一文一格の原則における意味上の制約を考慮。
- 局所構造同志を結ぶ。
- 大局スタックに保持された構造は局所構造内の主述語情報により、係り受けを決定。

例文「1階述語論理式は充足可能性を保持したまま節だけを対象することにより節に変換できるので一般性が失われるわけではない。」の構文解析の流れを示す。

まず、機能語を含む局所構造の範囲を決める。例文では、「たまま」、「ことにより」、「ので」の機能語があるので、局所構造は大きく5つできる(図3)。図4において(2)の局所構造を示す。構造内での主述語を中心に、語彙依存文法に従い、依存関係を結び、主述語情報がその構造の持つ全ての情報である。また、提題語を含む構造は大局スタックを用いる。局所構造内のメイン述語の情報によって、局所構造動詞の依存関係を取り、主題情報を要求する。この時大局スタックとの依存関係をとる。例文の場合、すべての局所構造において、マッチングを取ると、各々の主題が大局スタック内の情報を取りうるので、依存関係が成立し、主題を考慮した大局構造を導出できる。

5 性能評価

提案した方法は現在評価中である。形態素解析されたデータを、語彙依存文法にそったデータにし、システムに導入。結果は人手により判断し、評価を行なう。

- (1階述語論理式は) (1)
- (充足可能性を保持したまま) (2)
- (節だけを対象にすることにより) (3)
- (節に変換できるので) (4)
- (一般性が失われるわけではない。) (5)

図 3: 局所構造の決定

各局所構造ごとに依存ネットワークを作る。
(2)の局所構造の場合

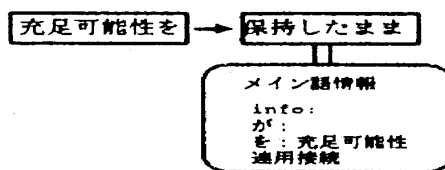


図 4: 局所内依存関係-(2)の局所構造

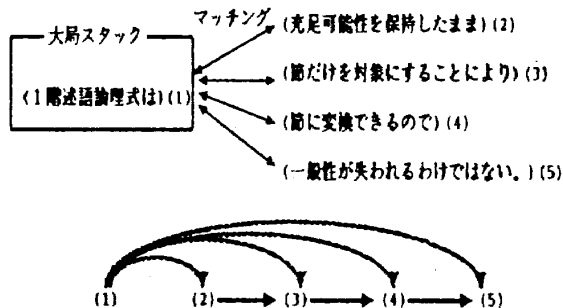


図 5: 大局構造ネットワーク

6 終わりに

機能語が文構造を支配する機能を用いて、大局構造の導出法を提案した。また、各局所構造の情報より、主題共有の機能も付加した。今後、提案した方法に沿ったシステムを構築し、依存構造を導出、大局構造の把握・主題共有を比較検討する。

参考文献

- [1] 児玉「依存文法の研究」研究社出版(1986)
- [2] 黒川「自然言語処理入門」近代科学社(1988)
- [3] 南「現代日本語の構造」大修館書店(1974)
- [4] 益岡、田窪「基礎日本語文法」くろしお出版(1992)
- [5] 亀井、村木「Lexical Discourse Grammarの提案」言語理解とコミュニケーション研究会(1986)
- [6] 土井、村木、亀井「語彙文脈文法とその大局的係り受け決定への応用II」言語理解とコミュニケーション研究会(1991)
- [7] 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)
- [8] 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-8(1994)
- [9] 酒井、上原、牧野「日本語の基本係り受け構造をもとにした依存構造による解析」情報処理学会第53回全国大会 2H-2(1996)