

句をノードとするネットワークを用いた構文および意味構造の同時解析

4 H-6

山田憲司 荒木健治 栃内香次
北海道大学工学研究科

1 はじめに

これまでもネットワークを用いた意味解析手法 [1] は存在するが、これは格構造の解析を目的としたものであり、意味解析において格構造を解析するだけでは不十分であると考えられる。日本語には修飾語が頻繁に現れるが、格構造解析だけではそのような構造を解析することができないという問題点が存在する。

本稿では句をノードとするネットワークを用いて構文構造および意味構造の同時解析手法を提案する。句とは、形態素を一定の規則でまとめたものである。また、意味構造の解析に独自に拡張した意味関係を用いることにより、従来の深層格では解析できなかった修飾語の意味構造解析も行う。具体的には、入力文の形態素解析結果から句を生成し、この句を利用して意味構造解析を行う。次に、それぞれの句について表層表現上の意味の解析を行う。システムは解析した結果をネットワークに登録し、以降の解析時に参照する。以下では本システムのネットワーク構造、解析手法を示すとともに実験結果から本手法の有効性について考察する。

2 ネットワークの構造

本手法のネットワークでは、ノードは単語またはは句と、リンクは単語間の関係と対応する (図 1)。

2.1 単語ノード

本システムのノードを単語ノードと呼ぶ。単語ノードは単語の表記とその品詞情報、または句の表記とその構成要素の情報を保持する。また、3. 1 で述べている活用型の句を構成要素である場合は、この単語の表層表現中での意味の情報も保持する。

2.2 表層リンク

本システムのリンクのうち、単語間の表層関係を表すリンクを表層リンクと呼ぶ。表層リンクは基本形を表すノードと活用形を表すノードとの間や、句

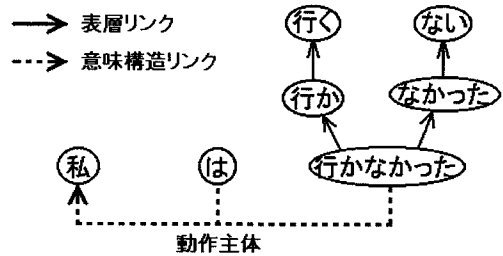


図 1. ネットワークの例

表 1. 使用する意味関係

深層格	状態格	動作・状態の存在しない格
動作主体	状態主体	所有格
動作対象	状態対象	特定格
動作経験者	対抗状態主体	
動作道具	状態時間	
動作結果	状態理由	
動作源泉	状態場所	
動作目標		
対抗動作主体		
動作時間		
動作理由		
動作場所		

とその構成要素のノードとの間を接続する。

2.3 意味構造リンク

本システムのリンクのうち、単語間の意味構造を表すリンクを意味構造リンクと呼ぶ。意味構造リンクは意味構造の係る元の語を表すノードと係り先の語を表すノード、そしてそれらを接続する語を表すノード間を接続し、それらが表す意味関係情報を保持することによって意味構造を表す。このリンクが保持する意味関係情報は Fillmore の深層格 [2] に独自の意味関係を加えたものを使用する (表 1)。

3 解析手法

最初に入力文をネットワーク中の単語ノードを用いて形態素解析する。次に形態素解析結果を用いて句を生成する。そして、句を用いてネットワーク中の意味構造リンクを探索することにより意味構造解

*Sequential analysis of the syntax and the semantic structure using the network which makes a phrase a node

*Yamada Kenji, Araki Kenji, Tochinnai Koji

*Graduate School of Engineering, Hokkaido University

析を行う。最後に単語ノードが保持する表層表現中での意味情報を探索することにより句の表層表現上での意味を解析する。

3.1 句の生成

システムは形態素解析によって得られた形態素とその品詞情報から、以下の規則に基づいて句を生成する。句には、活用のある単語をまとめて生成する活用型の句と、活用のない語から生成する非活用型の句の2種類がある。活用型の句は連続する動詞、形容詞、判定詞、助動詞、終助詞、形容詞性名詞接尾辞、形容詞性述語接尾辞、動詞性接尾辞をまとめて生成する。非活用型の句は上記以外の品詞1つで生成される。

3.2 表層表現上の意味の解析

活用型の句において、句の構成要素のノードを探索していき、そのノードの表層表現中での意味情報を取得していくことにより、句の表層表現上の意味を解析する。

4 実験

以上のシステムを実際に作成し、実験を行った。実験のための入力文として英語教材[3]の「花咲か爺さん」と和訳より単文34文(意味構造135組)を用意した。この意味構造は独自に拡張した意味関係を用いて第一筆者が作成したものを用いた。また、今回はネットワーク中に形態素解析を行うのに十分な量の単語ノードを登録できるほどの意味構造データを用意できなかったため、形態素解析には形態素解析ツールのJUMANを用いた。実験結果について意味構造の解析率の推移を図2に、表層表現の解析率の推移を図3に示す。図2中の意味誤りのグラフは、意味構造の解析は正しいものの、意味関係の解析が誤っているものを表す。

4.1 考察

図2を見ると、意味誤りの割合が約45%とかなり高いことがわかる。これは最初の段階では、ネットワーク中に意味構造リンクが蓄積されていないために起こっていると考えられる。一方、実験後半になると、今度は入力文中に新しい意味関係が出現するため、やはり意味関係の解析が失敗してしまう。この意味誤りの割合はさらに実験データを増やすことにより減少させることが可能だと考えられる。

5 おわりに

本手法では意味構造の正解析率は約16%とあま

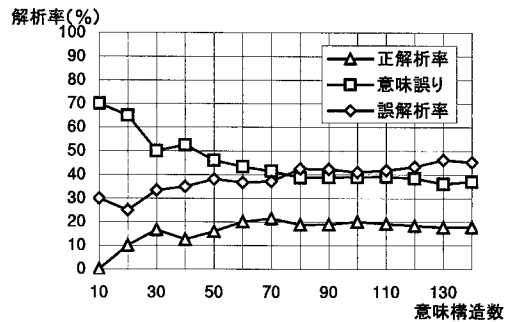


図2. 意味構造の解析率の推移

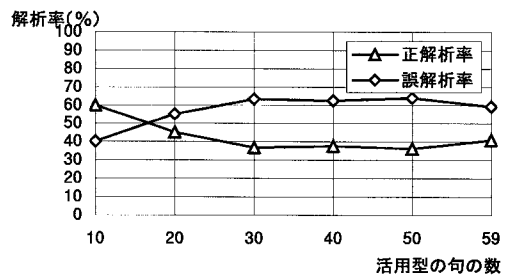


図3. 表層表現の解析率の推移

り良い値ではなかったが、表層表現の正解析率は42%であり、ある程度の成果が得られたと考えられる。

今後の課題としては、意味構造の解析アルゴリズムの改善が考えられる。また今回使用した意味関係は文の意味を表すには不十分であるため、この検討も今後の課題である。

参考文献

- [1] 竹内俊行, 荒木健治, 柁内香次: Marker Passing形式のパーザにおける力学的制約を用いた意味吊り上げモデルによる格構造解析手法, 電子情報通信学会論文誌 D-II Vol.183-D-II No.12 pp.2755-2766, 2000.
- [2] C.Fillmore (著), 田中春美, 船城道雄 (訳): 格文法の原理, 三省堂 (東京), 1980.
- [3] 遠山顕, モナ遠山: NHK ラジオ英会話入門 2001年2月号, 2001.