

複文の格表現とその意味解析システム SAGE の開発研究

原田研究室 大野博之(35500003)

原田研究室では、昨年度までに日本語文章の意味解析を行う SAGE を開発している。これは、文内の語の意味と語間の格関係を EDR 電子化辞書からの統計処理を行って決定し、格フレームとして出力する。しかし、単文における解析を中心にしており、補足節・副詞節・連体節・並列節などからなる複文の接続節がからむ深層格の解析は充分とは言えない。また、解析精度も実利用するには十分なレベルではない。

そこで本研究ではまず、複文における語と節の関係を表す格表現を新たに定め、複文で表されている内容の構造も格フレームで精密に表現できるようにした。また、この新表現に従い、従来の語間の解析に主として用いた EDR 電子化辞書情報からの統計的な格決定に加えて、補足節・副詞節・連体節・並列節毎に語と節（その中の中心となる述語）あるいは節と節間の格を、接続助詞、文節の主辞の品詞と語意などから決定するルールベースを用意した。具体的に補足節の解析では、「調子が悪い v2 こと n に気づいた v1」といった文があった場合、表層的な文字「こと」とそれに係る文節が用言かどうかで「調子が悪いこと」が補足節であると判断する。そして補足節内の中心となる述語 v2 と形式名詞 n の間を that 格とする。これは、「v1 that v2」という方法も考えられるが、「v1 object n that v2」とした方が、v1 と n の間の関係をより精密に表現できるので、形式名詞 n にも一つの格フレームを与えることにした。連体節の解析では、「私は太郎が n1 撮った v1 写真 n2 を見た」といった文があった場合、係り受け関係にある 2 文節において用言が体言に係っているかで「太郎が撮った写真」が連体節であると判断する。そして修飾関係を表す which 格 (n2 which v1) と修飾節中の関係を表す格 (v1 object n2) の両者を生成することにした。並列節の解析では、「太郎は音楽が好きで、花子は映画が好きだ」といった文があった場合、係り受け関係にある 2 文節が用言同士であることとその間に並列関係があることから「太郎は音楽が好きで」が並列節であると判断する。並列節の表す意味には、連結関係、選択関係、逆接関係、順接関係があり、and, or, reverse, sequence などを並列性のある用言間の深層格として解析することにした。副詞節の解析では、「テレビを見ながらご飯を食べる」といった文があった場合、係り受け関係にある 2 文節が用言同士で、その間に並列関係がない場合に副詞節（「テレビを見ながら」）であると判断する。副詞節が表す意味には、時、原因・理由、条件・譲歩、付帯状況・様態などがあり、論理的な記述や時系列的記述の多い文章などを分析した結果、reason, cause, logical, timing, purpose, when, during, などを並列性のない用言間の深層格と定義し、解析を行うこととした。

また、SAGE の高精度化のために、精度の高い係り受け解析システム「南瓜」と「KNP」に対応させた。また、複合語における構成語の語意を数詞も含めてすべて決定できるように改良し、さらに係り受け関係がなくても、ゼロ代名詞との格の補完ができるように改良した。

複文およそ 350 文における解析精度（正解率）を手作業で調べたところ、従来の SAGE と比べて、補足節は 0.0% から 90.5%、連体節は 42.5% から 69.3%、副詞節は 8.9% から 77.0%、並列節は 58.8% から 64.7% へ精度が向上した。よって、複文における深層格の解析精度は充分に向上したと言える。