

日本語文間の意味関係解析システム InSeRA の開発研究

原田研究室 川端 崇央 (35599003)

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 の開発が行われてきた。しかし、SAGE は 1 文 1 文の意味解析を行うのであり、文間の関係を解析し、文章全体の文脈をとらえることはできない。一方、本研究室で進められているオブジェクト指向設計の動的分析(イベント時系列の解析)の自動化や判例文検索における事故状況の類似度解析などの応用を考えると、日本語文章の解析としては、意味解析だけでは不十分である。そこで本研究では、日本語文章の文間の意味的な関係を表す「文間深層格」を定義し、この文間深層格を接続表現などの表層的な情報や、EDR 電子化辞書による語の語意などを用いて自動的に決定するシステム InSeRA(INter SEntences Relation Analysis)を開発することを目的とする。

具体的には、文間深層格として本研究では、文間の理由、条件、時間的近接などの意味的な関係を表す 21 種類の文間深層格を定義した。

InSeRA の基本的な処理の流れを説明する。入力には SAGE の出力である意味フレーム群である。これはフレーム番号、日本語文中に現れた語の形、品詞情報、概念 ID、格関係、原文の文章番号、などの 10 の要素からなる。InSeRA は、該当文の文章番号 S_i の意味フレーム群とその前後の文 $S_{i \pm 1}$ の意味フレーム群の間に、我々が提案するヒューリスティックなルールを適用することによって、2 文間に文間深層格を割り当てる。文間深層格を決定するためのルールは大きく以下の 3 つに分類できる。 接続表現と EDR 電子化辞書の概念 ID による方法。

接続表現と助動詞の活用形による方法。 接続表現による直接的対応付け。 は例えば、「理由」「原因」「目的」のように同一の接続表現を持つものに対して、その中心用言の上位概念の違いによって、適切な文間深層格を割り当てる。 は「仮定」「論理的条件」のように同一の接続表現を持つものに対して、その中心用言にかかる助動詞の活用形の違いによって、適切な文間深層格を割り当てる。 は例えば、「時間的重複」をあらわす「その間」という接続表現を持つ文とその前文との間に、「時間的重複」という文間深層格を割り当てる。これらのルールにおいて判定した結果を、文と文との間にどのような文間深層格が成り立つかを記述したインターリレーション・フレームとして出力する。

なお、本研究室で同時期に開発を行った“オブジェクト指向分析における動的分析の自動化システム CAMEO/D”に、これらの意味フレームとインターリレーション・フレームを与えると人手による分析と同レベルの解析結果を得られることがわかった。