

意味解析システム SAGE の Web 化と 連体・使役・受身における意味解析の精度向上

井村 裕 (15599011) 沓掛 俊樹 (15599032) 佐藤 直美 (15599045)
原 田 研 究 室

原田研究室ではこれまで、EDR 電子化辞書に記載された情報を元に、日本語文を意味解析し格フレーム群に変換するシステム SAGE を開発して、語意精度 80.1%、深層格精度 80.6%を達成している。しかし、実利用するにはさらなる精度向上が必要なので、SAGE の評価実験を繰り返し行ったところ、不正解の原因として、連体節における補足関係を示す深層格の正解率が低い、使役・受身を含む文章の正解率が低い、などがあった。そこで、これらの原因に焦点を絞りより精度の高い意味解析を行うシステム SAGE2002 の開発を行った。

用言が体言を修飾する連体節（例：「借りた本」）において、従来の SAGE では語意と補足関係を表す深層格（本 object 借りた）の精度が低かった。そこでまず、係り受け解析から得た 2 文節から連体節の関係（名詞を修飾する関係）があるかどうかを判定する。そのような関係が得られなければ通常通りの語意と格の優先率を求める計算を行う。連体節の関係があると判定されたらこの 2 文節に対して、被修飾名詞「本」が連体節中の述語「借りた」に対する補足語の係りに仮定し（「借りた本」「本を借りた」として語意と格の優先率を求めてから最適解釈木を生成するように改良した。

従来の Sage では、接辞'(r)areru' '(s)aseru'が付くと表層格が変わるが深層格はかわらない使役表現や受動表現に対応していなかった。そこで、使役表現・受動表現を含む文は能動態に直したり、深層格を独自に決定するルールを作成することで解析精度を向上した。

また、SAGE の実用化に向けて現在 Windows 上のみで動作している SAGE を Unix 上でも動作するようにし、Web 上で利用・公開できるよう Web Server 版を開発した。さらに、以前からの様々な要求に沿って、形態素解析・係り受け解析との一貫した連動、文節番号の追加、KNP による解析結果の品詞の追加、生成される格フレーム要素の選択を可能にするといった改良を行った。

EDR 電子化辞書のコーパス辞書に記述されている例文から 100 文ランダムに選び、SAGE が生成した格フレームと、コーパス辞書からの意味解析済みデータを格フレーム形式に変換したコーパスフレームより、語意・深層格・係り受け関係をそれぞれ照合するシステム EvalSAGE による評価実験を行った結果、語意正解率は 89.2%、深層格正解率は 89.0%であった。よって、SAGE2002 は実利用を開始できる精度に至ったといえる。