

意味解析システム SAGE の精度向上と利便性の向上

山本 哲哉 (15801077)・小林 寛之 (15802033)・米澤 太一 (15802078)

原 田 研 究 室

1. はじめに

本研究では、EDR 電子化辞書を元に、日本語文章を意味解析し格フレーム群を生成するシステム SAGE の大幅な改訂を行い精度向上と利便性の向上を図った。

2. 意味解析システム SAGE

SAGE における意味解析とは、係り受け関係にある全ての文節の主辞および他の主要な形態素の語意と主辞間の深層格を決定し、文節頂点と深層格辺からなる意味グラフを生成することである。ここで主辞とは、文節の主要な語を指す。格の向きは、係り先 係り元とする。例えば、「人は発生するとフロアに並ぶ。」は図1の様に展開される。

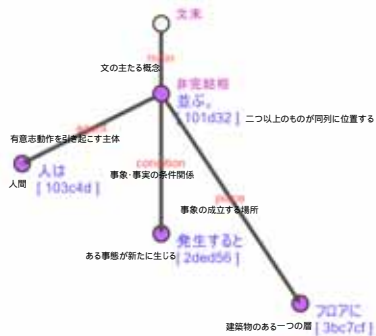


図1. SAGE2005 の解析結果

3. SAGE の精度向上

複合語、サ変名詞の検出、複文ルール、総主の構文に対しルール修正を行い、精度を大幅に向上させた。事例・結果の整理と把握のための補助ツールを作った。図1は意味グラフの視覚化ツール「vivit」による表示例である。また、概念マップの一覧表示「CidBrowser」も制作した。

EDR 辞書の情報を関係データベースとして構築した。EDR 辞書の情報を視覚的に確認できるように「Verifire」を作成し、条件を与えて辞書情報を検

索、EDR 辞書の誤りの検査ができる機能を加えた。誤り検査の結果を元に手で辞書情報の修正案を作成して、辞書内容を修正する機能も作成した。EDR 辞書修正用レコードを元にデータベースの情報を更新するための「AutoKoshin」も作成した。データベースの情報を SAGE 用の辞書である slz ファイルとして出力する機能も作成した。

文間格関係の解析

文内の接続表現や主述語の概念 ID などの情報を元に、文と文の間にある深層関係を調べ、SAGE の eos フレームに結果を返す。

(例) e: 8,null,逆接,[mn4]

4. SAGE の速度向上

概念の階層構造のコード化

従来は2概念の類似度を求めるために、EDR 概念木の最短経路を探索していたので時間がかかった。本研究では概念木の階層構造をコード化することで、類似度の算出速度を従来の約13倍の速度に向上させた。

EDR シリアライズ辞書

従来は EDR 電子化テキスト辞書から辞書を構築していたので、システムの起動に時間がかかった。そこで必要な情報のみが書かれた独自辞書を生成し、展開速度の向上を図った。

5. 結果

SAGE2005 の意味解析の精度を調べる評価実験を無作為に抽出した文章に対し行った。

表1. SAGE2005 の評価実験の結果

	コーパス101文		インターネット101文	
	全語意の正解率	全格の正解率	全語意の正解率	全格の正解率
SAGE2004	88.2%	86.6%	86.3%	85.6%
SAGE2005	94.4%	90.7%	93.0%	90.4%

表1. のように本年度の研究によって、SAGE の解析精度が実用域に達した事が検証された。