

理工学専攻修士論文要旨

提出年度 : 2008 年度
提出日 : 2009 年 1 月 26 日
専修コース : 知能情報コース
学生番号 : 35607054
学生氏名 : 梶川 由梨
研究指導教員 : 原田 実 教授

(論文題目)

意味グラフベースの言い換え知識の自動生成システム Blumen の開発研究

(内容の要旨)

原田研究室では 2005 年度に、意味グラフにおける言い換えエンジン Serpent を開発し、2006 年度に、Serpent 用言い換え知識自動生成システム BLUME を開発した。しかし BLUME では有用性のない言い換え表現対も抽出され、また Serpent 用言い換え知識は文脈を伴うものではなかった。そこで不要な言い換え知識生成を抑制し、また当研究室で新しく開発された文脈を伴った言い換えエンジン Priest 用の言い換え知識を自動生成するシステム Blumen を開発した。

BLUME と比較し、Blumen で追加された機能は主に 3 つある。

1 つ目は不要な言い換え知識生成の抑制である。BLUME では岩波国語辞典第四版電子データから言い換え表現対を抽出していたが、不要な言い換え表現対が抽出された場合はそれを手作業で取り除いていた。例えば「止まる,だ」「案ずるに,古風な言い方」などの言い換え表現対がそうである。Blumen ではこれらを自動で取り除く処理をする。

2 つ目は文脈を伴った言い換え知識の生成である。BLUME で生成された言い換え知識は文脈を考慮したものではなかったため、言い換えられる文章が言い換え知識の言い換え元の部分と一致すれば、文脈に関係なく言い換えが行われてしまい、不自然な言い換えが行われる事もあった。例えば「明るくなる 明ける」という言い換え知識で言い換えを行う場合、「空が明るくなる 空が明ける」という言い換えは自然だが「色が明るくなる 色が明ける」という言い換えは不自然となる。そのためこの場合は「空が」や「視界が」などの文脈を持つ時のみ言い換えるよう、文脈を考慮した言い換え知識を生成するようにした。

3 つ目は文脈を考慮して言い換えを行う言い換えエンジン Priest 用の言い換え知識を生成する事である。岩波国語辞典から抽出された言い換え表現対の両文を意味解析システム Sage で解析し、得られた意味グラフの情報を比べながらキーとなる部分や文脈などを抽出し、Priest 用の構文に従って言い換え知識を生成する。

こうして Blumen で生成された言い換え知識数は 6683 であり、BLUME で生成された言い換え知識数は 17297 なので、不要な言い換えを抑制する低減率は 38.64% である。また Blumen 生成の文脈を考慮した言い換え知識数は 1273 だったので、文脈考慮率は 19.05% である。よって不要な言い換え知識生成の抑制に成功し、文脈を伴った言い換え知識生成にも成功したと言える。