

言い換え知識自動生成システム Blume の開発

荻原 由紀恵[†] 梶川 由梨[†] 中村 浩之[†] 原田 実[‡]

[†], [‡]青山学院大学 理工学部 情報テクノロジー学科

1. 背景・研究目的

原田研究室では昨年度、意味グラフにおける言い換えエンジン Serpent [1]を開発した。Serpent で使用する言い換えの知識情報はこれまで人手によって作成・追加を行っていたが、言い換える表現対には限られた文章にしか表れないものも多数存在し、人手で網羅的に収集するのは困難である。このため、本研究では辞書知識を用いて自動的に言い換え知識を生成するシステム Blume を作成する。

2. 言い換え知識自動生成システム Blume

今回作成した言い換え知識生成システムの構成を図1に示す。

例： 乱暴な振る舞いをする 荒れる

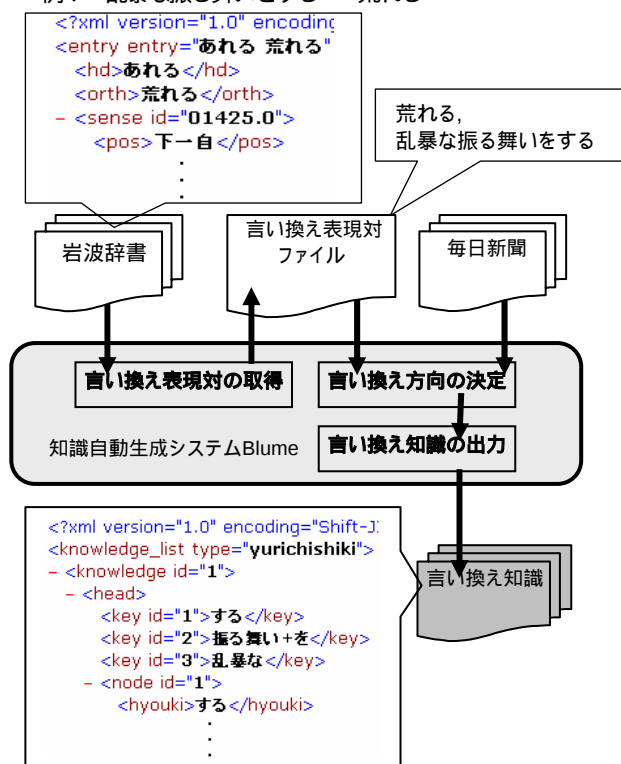


図1 Blume システム構成

An Automatic Paraphrase Knowledge Generating System Blume

Yukie Ogihara[†], Yuri Kajikawa[†], Hiroyuki Nakamura[†] and Minoru Harada[‡]

[†]Undergraduate school of Integrated Information Technology, Aoyama Gakuin University.

[‡]Department of Integrated Information Technology, Faculty of Science and Engineering, Aoyama Gakuin University.

2.1 言い換え表現対の取得

ある表現から別の表現へと言い換えを行うためには、それら表現の対が必要である。本研究では、言い換え表現対を『岩波国語辞典（電子データ）[2]』から自動的に取得する。抽出方法は見出し語の品詞ごとに分け、一般動詞、サ変名詞、形容詞、連語、副詞の5種類の抽出を行う。動詞の抽出法は鍛冶らの言い換え抽出手法[3]を基本とし、形容詞句はこれを拡張して抽出した。

i) 動詞句の抽出

辞書における動詞の定義文は、文末にある用言（主辞）と、それに副詞的にかかる文節が見出し語である動詞と同等の意味を成していることが多い。そのため、これらを取り出せば見出し語と定義文で言い換え表現対が取得できる。

例) 遮る:

行動や動きを邪魔してその先へまで

及ばないようにする ↓

遮る 及ばないようにする

ii) サ変名詞の抽出

サ変名詞の場合、定義文の文末が「こと」で終わっていることが大半であるため、その直前の語を主辞として動詞と同様の手順で抽出し、見出し語に「する」を加えてサ変動詞の言い換え表現対とする。

例) 絶食: 食を断つこと

↓
絶食する 食を断つ

iii) 形容詞句の抽出

定義文末の形容詞句と副詞的にかかる文節を抽出する。

例) 怪しい: 不思議で気味が悪い

↓
怪しい 気味が悪い

iv) 副詞・連語の抽出

定義文をそのまま取り出した後、人手によって有用なものを選別する。

2.2 言い換え方向の決定

抽出した言い換え表現対は、どちらからどちらへ言い換えを行うのかを決定する。今回は以下の2点に着目して言い換える方向を決定した。ルールの優先順位は > である。

意味グラフにおける文節ノードがより少なくなる方へ言い換える(図2参照)
より使用頻度の高い語へ言い換える
なお、毎日新聞より検索した件数を頻度とした。

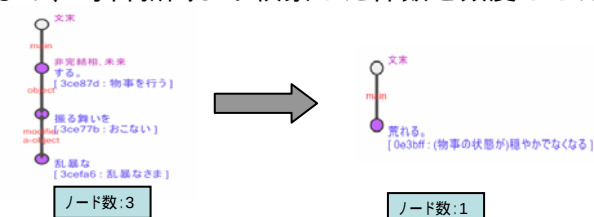


図2 文節ノード数が減少する方へ

2.3 言い換え知識の生成

図3に示すように言い換え表現対の両辺を意味解析システム Sage [4]で解析し、得られた意味グラフの情報を元に言い換え知識を生成する。

・Sage解析データ(意味グラフ情

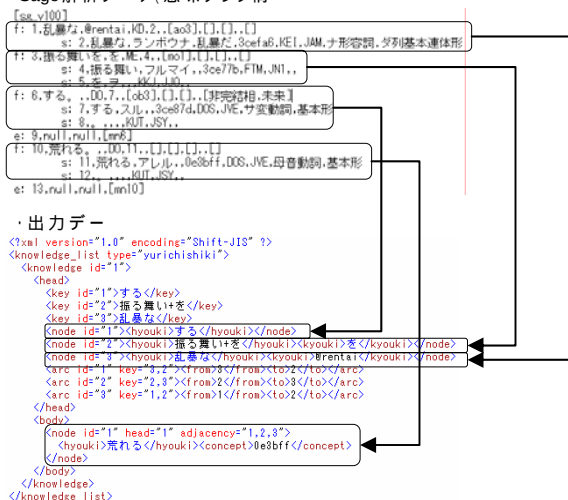


図3 Sage 解析結果と出力 XML の対応

3. 実験結果と評価

本研究では『岩波国語辞典タグつきコーパス2004(56,000語収録)』から言い換え表現対を取得した。表現対の取得件数・知識生成件数を表1に示す。Blumeによって動詞句を中心とした前年度の知識数419個を大幅に上回る17297個の知識を獲得できた。またこれらを使ったSerpentによる言い換え例を図4に示す。

表1 Blumeでの言い換え知識生成結果

見出し語の品詞	自動取得 表現対	取得*1 表現対数	知識*2 生成数
サ変名詞(サ変動詞)	8155	8032	7787
動詞	11015	10534	8045
形容詞	685	649	400
副詞	1195	1061	534
連語	993	776	531
合計	22043	21052	17297

*1 自動取得後に人手で誤った表現対を修正・除外した。

*2 EDR 辞書に存在する一語の表現対は意味グラフ上の言い換えは不要とみなし、本研究では知識生成を省いた。

例1) 別の機会に 改めて

前: ...「また別の機会に議論しましょう」と、軽く話を受け流した...

後: ...「また改めて議論しましょう」と、軽く話を受け流した...

例2) 体を休める 療養する

前: ...メドレーリレーが終わったら体を休めたい。

後: ...メドレーリレーが終わったら療養する。

* Serpentの仕様上、活用は基本形だが意味グラフ上は正しい

図4 Blumeが生成した知識による言い換え例

4. まとめと今後の課題

辞書知識から自動的に言い換え知識を生成することで、人手で作成困難な言い換え知識を大量に獲得することに成功した。また、鍛冶らによる辞書からの抽出手法を形容詞句にも拡張し、Serpentと連動させて意味グラフ上での言い換えが実現できた。課題は主に次の3つである。

4.1 言い換え知識の適用条件等の精密化

本システムは取得された語の意味グラフ部分のみ言い換えを行う。文脈に応じて適用すべき場合の条件の取得や言い換え元とその文脈との深層格の変換方法の指定等が今後必要である。

4.2 変数を使用した言い換え知識生成

言い換え表現対に変数を用いれば、部分一致による汎用的な言い換え知識の生成が可能になる。

4.3 岩波辞書データの抽出アルゴリズムの改善

普通名詞・形容動詞等の言い換えも行う。

5. 謝辞

岩波国語辞典の元データを研究利用のために公開された水谷静夫先生を初めとする編者の方々、メタデータの研究利用を許諾された三菱電機株式会社、および調整の労を取って下さった株式会社岩波書店および特定非営利活動法人言語資源協会(GSK)に感謝いたします。

6. 参考文献

- [1] 前澤敏之：意味グラフに対する言い換えエンジンの開発研究，修士論文，青山学院大学理工学部情報テクノロジー学科，2005。
- [2] 西尾実・岩淵悦太郎・水谷静夫(編)：岩波国語辞典第6版，岩波書店，2000。
- [3] 鍛冶伸裕・河原大輔・黒橋禎夫・佐藤理史：国語辞典とコーパスを用いた用言の言い換え規則の学習，言語処理学会第8回年次大会，2002。
- [4] 山本哲哉・小林寛之・米澤太一：意味解析システム SAGE の精度向上と利便性の向上，青山学院大学理工学部情報テクノロジー学科卒業論文，2005。