

事態含意名詞を用いた事態間関係知識の獲得

阿部 修也 乾 健太郎 松本 裕治
奈良先端科学技術大学院大学 情報科学研究科
{shuya-a,inui,matsu}@is.naist.jp

「～するために～する」のような2つの事態表現の共起パターンを利用してテキストコーパスから事態間の関係の知識を獲得する事態間関係知識獲得の問題について論じる。知識獲得の精度と規模を両立するには、意味的な制限の強い特殊な共起パターンを数多く用意することが望ましい。この問題に対し、本稿では、(a) 事態を直接的あるいは間接的に指す事態含意名詞を利用すること、(b) 実体間関係抽出のための共起パターン獲得技術を事態間関係獲得に応用することを検討し、これまでに得られた実験結果を報告する。

Extracting Inter-Eventual Relations from Text Data Using Event-Related Nouns

Shuya Abe Kentaro Inui Yuji Matsumoto
Graduate School of Information Science, Nara Institute of Science and Technology
{shuya-a,inui,matsu}@is.naist.jp

This paper addresses automatic extraction of inter-eventual relations such as causality relations and means relation from a large text corpus using linguistic patterns. The main concerns are (a) the impacts of using event-related nouns, as well as verbs, as indicators of events and (b) the effects of applying state-of-the-art techniques for pattern learning in entity relation extraction. The current results of our preliminary experiment are reported.

1. はじめに

人間に近い高度な言語情報処理能力を工学的に実現するには、辞書や文法などの言語知識の他に、大量の世界知識を計算機に与える必要がある。たとえば、「洗濯物を干したとたんに雨が降ってきた。」という文の意味は、「ある人が洗濯物を干した」「そこへ雨が降ってきた」といった表面的な内容に留まらない。人間ならば、「洗濯物を干したのはそれを乾かすことが目的であり、雨のためその目的が達成されなかった。つまり全体として悪いニュースを伝えている」といった「行間を読む」理解が可能である。こうした深い言語理解には、〈ものを干す〉と〈ものが乾く〉が行為-効果の関係にあることや、〈雨が降る〉と〈ものが乾く〉が負の因果関係にあるといった事態間の関係に関する知識が必要である。

こうした高度な言語理解の研究は1980年代に遡り、当時意図理解やプラン認識の名で盛んに研究されたが、それに必要な大量の言語知識や常識的世界知識を用意する方法論を持たなかったために、実用規模には発展しなかった。しかし、当時と現在では次の2点で状況が大きく異なる。

- 統計的言語処理の発展によって、形態素・係り受け解析や照応・省略解析などの基盤技術が実用的な水準に達してきた。

- 情報技術が日常生活に浸透しブログのような情報発信手段が普及したことにより、人間の日常的行動を記述した文書が大量に入手可能になった。この傾向は今後さらに顕著になると予測される。

以上のような背景から、大規模なテキストデータから事態関係知識を自動的に獲得するというアプローチが現実味を帯びてきており、すでにいくつかの試みが報告されている^{1),5),7),8)}。例えば、Inuiらは、接続助詞「ため」を含む複文(タメ複文)に現れる主節と従属節から事態間の因果関係を獲得する方法を提案しており、獲得した関係を Cause, Precondition, Effect, Means の4種類に高い精度で分類できると報告している⁷⁾。ただし、「書店に行く」と「書店で買う」のように、わざわざ言及するまでもない当たり前の関係をタメ複文のようなパターンから獲得のは限界があり、実際 Torisawa の報告によると、Torisawa の方法で獲得した知識の多くがタメ複文では獲得できないものであった⁸⁾。一方、Torisawa の方法は、動詞テ形接続や連用中止接続のように頻度が高く一般的な手がかりと、別途収集した格関係の統計を巧妙に組み合わせることによって、より常識的な事態間関係を獲得することを狙うもので、「モノの用途とその準備の関係」の獲得で成果を上げている。ただし、こうした方法を広く他の事態間関係に適用できるかどうかは今のところ明らかでない。

こうした研究の知見をまとめると、タメ複文やテ形接続のような特定の複文パターンを使って事態間関係のインスタンスを獲得するアプローチには次のようなジレンマがある。

- 特殊なパターンを使うと、より意味的に制約された関係のインスタンスを獲得することができるが、そうしたパターンは頻度が少なく、十分な量のインスタンスを獲得するのは難しい。また、常識的な関係のインスタンスはこうした特殊なパターンに現れにくい。
- より一般的なパターンは、常識的な関係のインスタンスとも共起し、頻度も高いが、所望の関係でないインスタンスも同様に共起するので、ノイズのフィルタリングに工夫が必要である。

すなわち、知識獲得の精度と規模を両立するには、意味的な制限の強い特殊な共起パターンを数多く用意することが望ましい。この問題に対し、本稿では、用言だけでなく体言の中にも事態を表す、あるいは含意するもの（以下、事態含意名詞）が多数あることに着目し、事態含意名詞を含むより広範な共起パターンを利用して、事態間関係のインスタンスを獲得する方法を検討する。

2. 事態含意名詞と事態間関係獲得

本稿では、行為者が意志的に行う**行為**とそれ以外の**出来事**（経験や状態、状態変化など）を併せて**事態**とよび、事態間関係の知識を獲得する問題を考える。我々の目標は、因果関係（「運動する」と「汗をかく」）や部分全体関係（「研究する」と「実験する」）、上位下位関係（「罪を犯す」と「盗む」）などを含む広い範囲の関係を獲得し、深い言語理解に資する知識ベースを構築することである。

事態を表す言語表現は「汗をかく」のような動詞句に限られるわけではない。次に挙げるように、名詞の中にも事態を直接表現するものや間接的に指すものが少なくない。本稿ではこうした事態を直接的あるいは間接的に参照する名詞を**事態含意名詞**と呼ぶ。

- **動詞形を持つ事態名詞**：窃盗、外食など
- **動詞形を持たない事態名詞**：雨、ガス欠、運動会など
- **事態の項を指す名詞**：犯人（「罪を犯す」という行為の行為者）、運転者など
- **特定の用途を持ったモノを指す名詞**：包丁（「食材を切る」ための道具）、研究室（「研究する」ための場所）など

事態を表す表現として上のような事態含意名詞まで含めて考えると、事態間関係獲得の手がかりとなる共起パターンは、2つの動詞句からなる複文のパターンに比べて飛躍的に広がる。例えば、事態含意名詞と動詞句の共起には、「雨でタイヤがすべる」、「食後に歯を

磨く」、「研究室で実験する」、「窃盗の容疑で逮捕される」のようなパターンが考えられるし、「入会時の登録」のような事態含意名詞どうしの共起も有用な手がかりになる可能性がある。1節で述べたように、我々は意味的な制限の強い特殊な共起パターンをたくさん集めたいので、事態含意名詞の利用によって共起パターンの候補のプールが拡大することは重要である。

次の問題は、膨大な数の共起パターンの候補からどうやって信頼性の高いものを探すかである。従来の事態間関係獲得の研究は、少数の適当な共起パターンを手で見つけて利用することを前提にしていたので、パターン発見の問題にはアプローチしてこなかった。一方、〈著作物名〉と〈著者名〉のような実体（entity）間の関係を抽出する、いわゆる関係抽出の研究では、実体の出現パターン（つまり共起パターン）の自動獲得が主要な課題の一つとして認識され、一定の成果を得るに至っている^{2),3)}。

以上のような背景を踏まえると、まずは、事態間関係知識の獲得を上の意味での関係抽出タスクの一種と見なし、関係抽出のための共起パターン獲得技術を事態間関係獲得に適用することの有効性を調べることが重要であろう。そこで我々は、共起パターン獲得技術の一例として、Pantelらが最近提案したEspressoと呼ばれるアルゴリズム³⁾を取り上げ、これを拡張して事態間関係獲得に応用することを試みている。

以下、本稿では、3節でPantelらの手法を概説し、4節でこれを事態間関係獲得に適用する際に必要となるいくつかの拡張を述べた後、5節で現在までに得られている実験結果を報告する。

3. 関係獲得アルゴリズム Espresso

Pantelら³⁾は、信頼性の高いモノ間関係のインスタンスをシードとして与えて、そこからパターンを獲得し、そのパターンを使って新しいインスタンスを獲得するというブートストラップ的關係獲得手法を用いている。

Pantelらのブートストラップ的關係知識獲得手法は、4つの操作を繰り返し適用することでパターンの精度を高めて、結果として良いインスタンスを獲得する。

- パターンの獲得
- パターンのランキングと選択
- インスタンスの獲得
- インスタンスのランキングと選択

3.1 パターンの獲得

所望の関係のインスタンス $\{x, y\}$ を与えたとき、コーパスから x と y が含む文を抽出し、それらを一般化してパターンとする。このときインスタンス i の信頼性 r_i 式(2)を用いてスコアによって算出して上位 k 個のインスタンスのみを用いる。例えば、インスタンス “Itary, country” を与えたとき、テキスト “country such

as Itary” からパターン “Y such as X” を獲得する。

3.2 パターンのランキングと選択

式 (1) を用いてパターンの信頼性を評価する。

$$r_{\pi}(p) = \frac{\sum_{i \in I} \left(\frac{pmi(i, p)}{\max_{pmi}} \times r_i(i) \right)}{|I|} \quad (1)$$

I はインスタンスの集合、 p はパターン、 \max_{pmi} は全てのインスタンスと全てのパターンにおける PMI (式 (3)) の最大値である。インスタンス i の信頼性 $r_i(i)$ は式 (2) で与えられる。

3.3 インスタンスの獲得

パターン p を与えたとき、コーパスから p を含む文を抽出し、ここからインスタンス $\{x, y\}$ を獲得する。このとき $r_i(i)$ (式 (1)) の順に m 個のパターンのみを用いる。例えばパターン “Y such as X” を用いて、テキスト “country such as Itary” からインスタンス “Itary, country” を獲得する。

3.4 インスタンスのランキングと選択

式 (2) を用いてインスタンスの良さを評価する。

$$r_i(i) = \frac{\sum_{p \in P} \left(\frac{pmi(i, p)}{\max_{pmi}} \times r_{\pi}(p) \right)}{|P|} \quad (2)$$

P はパターンの集合。なお、シードインスタンスについては $r_i(i) = 1$ とする。

3.5 pmi の改善

Pantel ら⁴⁾ は、 pmi (式 (3)) が低頻度の事例に正のバイアスがかかることは良く知られた問題であり、それを解決するために式 (4) を用いると述べている。

$$\begin{aligned} pmi(x, y) &= \log \frac{P(x, y)}{P(x)P(y)} \\ &= \log \frac{C_{xy}}{N} \\ &= \log \frac{\sum_{i=1}^n C_{ix} \sum_{j=1}^m C_{yj}}{N} \times \frac{N}{N} \end{aligned} \quad (3)$$

$$\begin{aligned} pmi2(x, y) &= pmi(x, y) \\ &\times \frac{C_{xy}}{C_{xy} + 1} \times \frac{\min\left(\sum_{i=1}^n C_{xi}, \sum_{j=1}^m C_{jy}\right)}{\min\left(\sum_{i=1}^n C_{xi}, \sum_{j=1}^m C_{jy}\right) + 1} \end{aligned} \quad (4)$$

4. 事態間関係獲得への適用

Pantel らは実体間関係知識を獲得したが、我々は事態間関係知識を獲得する。そのため、Pantel らの手法に次の変更を加えた。

- Pantel らはインスタンスをシードとしたが、我々はパターンをシードとしてインスタンスを獲得

する。

- Pantel らは形態素列をパターンとしたが、我々はここを変更し、係り受け関係の部分木と事態のタイプによってパターンを指定する。
- パターンからインスタンスを獲得し、インスタンスからパターンを獲得するときのパラメータを変更した。

4.1 シードをパターンとする

Pantel らは “Itary, country” のようなインスタンスをシードとした。

この手法で良い結果を出すためには十分な量のシードインスタンスを用意する必要があるが、信頼度が高いシードを大量に用意するのは用意ではない。そこで、インスタンスの代わりにパターンをシードとし、パターンから獲得したインスタンスをシードと見なすことで、大量のインスタンスをシードとして与えた場合と同じ効果があると考え、パターンをシードとする。これに伴ない、 $r_i(i)$ (式 (2)) を計算するときパターン p がシードであれば $r_{\pi}(p) = 1$ とした。

4.2 動詞句

「切符を買う→電車に乗る」という事態間関係を考えた場合、「買う→電車に乗る」は「買う」を一般化しすぎたために不適切な関係である。「駅で切符を買う→電車に乗る」は事態間関係としては間違いではないが「駅で切符を買う」が特殊過ぎるために、この関係を利用するとき問題が生じる可能性がある。

事態が動詞句のとき適切に格を選ぶことで良い事態を獲得することができる。スコア関数は適切な格を持つ事態に高いスコアを与えるので、動詞句を格の有無によって展開してからスコア関数を適用することで適切な格を持つ動詞句を選ぶ。ここでは組み合わせの可能性が増え過ぎるのを防ぐために、動詞句の格を最大で1つとして展開した。例えば「駅で切符を買う」という動詞句を、「駅で買う」「切符を買う」「買う」に展開する。

4.3 事態含意名詞について

サ変名詞と接尾辞を組み合わせることで事態含意名詞を作成する。例えば、サ変名詞「マッサージ」に接尾辞を付与し、「マッサージ器」「マッサージ師」「マッサージ中」「マッサージ」(接尾辞なし)となる。これらはどれも、「マッサージする」という事態を含意していると考えられる。実験では、サ変名詞を含む名詞句を事態含意名詞と見なした。

4.4 パターン

我々が用いたパターンは次の要素の組み合わせからなる(例: 図1)。

- 係り受け関係*に基づく3つのパターン
 - 事態と事態が直接係り受け関係になっている

* 「河原ら⁹⁾ が収集した Web 上の 5 億文の日本語テキスト」の係り受け解析結果を利用した。

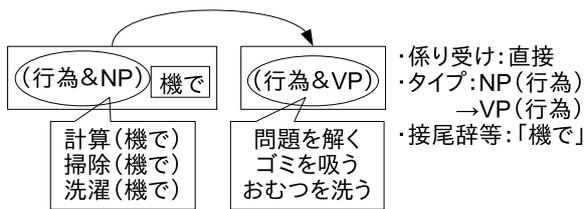


図1 共起パターンの例

表1 前後関係のための初期パターン

後続行為 _{NP}	の前に	先行行為 _{VP}
後続行為 _{VP}	前に	先行行為 _{VP}
後続行為 _{NP}	直前に	先行行為 _{VP}

- 事態と事態が任意の文節要素を介して係り受け関係になっている
- 事態と事態がそれぞれ共通の任意の文節に係っている
- 事態の接尾辞と助詞
 - 接尾辞: ~者, ~機, ~中, ...
 - 助詞: ~が, ~を, ~に, ~ために, ...
- 事態が行為か出来事かの区別
 - 行為: 走る, 食事をする, ...
 - 出来事: 風邪をひく, 事故にあう, ...

4.5 行為と出来事を区別するための辞書

事態が行為なのか出来事なのかを区別するために、LCS辞書^{12),13)}の動作主性の判断を用いた。しかし、ここで対処としている事態の種類と比べるとLCS辞書の規模が小さい。そこで、まずは人手で動作主性の有無を付与した。次に、分類語彙表¹⁰⁾を用いて、分類語彙表で同一のグループにある語は同じ動作主性を持っていると仮定し、動作主性が付与されていない語にも動作主性の有無を付与した。

4.6 実験条件

関係獲得のソースとなるコーパスには、河原ら⁹⁾が収集した「Web上の5億文の日本語テキスト」の約1/4を用いた。ただし、頻度5回以下の事態と頻度7回以下の共起パターン、文節の文字数が16文字を越えるもの、事態の動詞部分が分類語彙表に含まれていないものを除いた。

事態間関係として、行為の前後関係(ある行為の前にしばしば行なう行為)と行為の部分全体関係(ある行為を行なう間にしばしば行なう行為)を対象に実験する。

表1に前後関係の初期パターン、表2に部分全体関係の初期パターンを示す。

繰り返し回数を規定する。パターン(P_0 ; 初期値) → インスタンス(I_1) → パターン(P_2) → インスタンス(I_3 ; 結果)を1回の繰り返しとする。最初のパターン(P_0 ; 初期値) → インスタンス(I_1)を準備段階と考え、パターン(P_2) → インスタンス(I_3 ; 結果)を繰り返し

表2 部分全体関係のための初期パターン

全体行為 _{NP}	中に	部分行為 _{VP}
全体行為 _{NP}	時に	部分行為 _{VP}
全体行為 _{NP}	の間に	部分行為 _{VP}
全体行為 _{NP}	の途中で	部分行為 _{VP}
全体行為 _{NP}	の途中に	部分行為 _{VP}

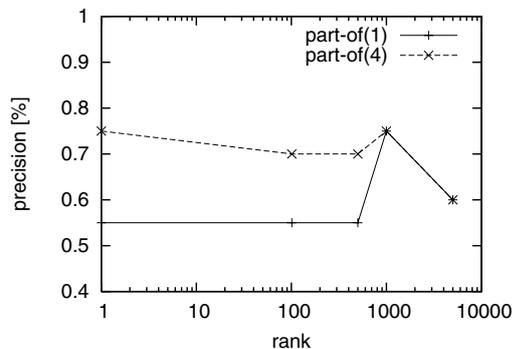


図2 部分全体関係の精度(繰り返し回数の違い)

と考えて、これを1回の繰り返しとした。同様に、パターン(P_0 ; 初期値) → インスタンス(I_1) → パターン(P_2) → インスタンス(I_3) → パターン(P_4) → インスタンス(I_5) → パターン(P_6) → インスタンス(I_7 ; 結果) → パターン(P_8) → インスタンス(I_9 ; 結果)を4回の繰り返しとする。

計算量の問題から、インスタンスを獲得するさいに頻度1000回以上のパターンを除いた。高頻度のパターンから生成されるインスタンスは低スコアになる可能性が高いため問題にはならないと考えたためである。

また、3.3で示した m を800、3.1で示した k を800とした。

5. 結果

5.1 精度

繰り返し1回、繰り返し4回、それぞれで獲得したインスタンスをスコア順に上位1~100件、101~500件、501~1000件、1001~5000件、5001~10000件の領域に分けた。それぞれの領域から各20組のインスタンスをランダムに抽出し、それを人手で評価した。

部分全体関係は図2のように、繰り返し1回の場合よりも繰り返し4回の方が精度は良い。また、5000位くらいから精度が下り始めた。

前後関係は図3のように、繰り返し1回の場合も繰り返し4回の場合も精度に差はなかった。これは、上位1~1000件くらいまでは繰り返し1回の時にも繰り返し4回の時にも、獲得したインスタンスの内容と順が共にほとんど等しかったため、結果として精度もほとんど変らなかった。

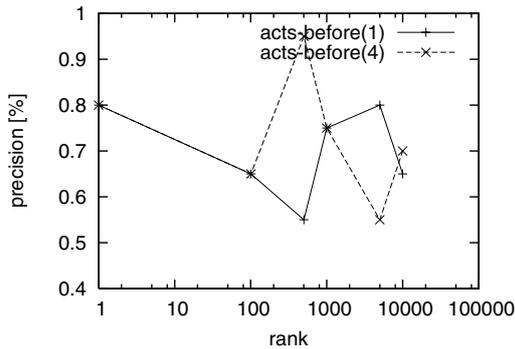


図3 前後関係の精度 (繰返し回数の違い)

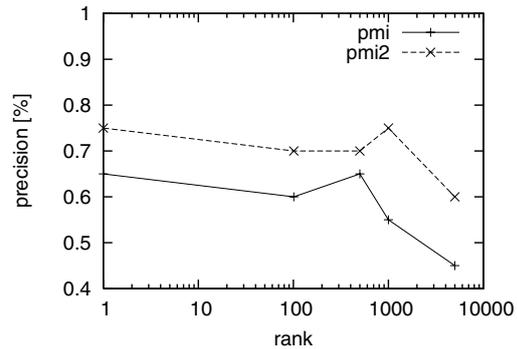


図4 pmi と pmi2 の違い

表3 初期パターンと意味的に類似したパターンの例

全体行為 _{NP}	時に	部分行為 _{NP}
全体行為 _{NP}	時	部分行為 _{NP}
全体行為 _{NP}	時の	部分行為 _{NP}
全体行為 _{NP}	中の	部分行為 _{NP}
全体行為 _{NP}	の際に	部分行為 _{VP}
全体行為 _{NP}	時間中に	部分行為 _{VP}

表4 初期パターンの特殊化になっている特殊なパターンの例

全体行為 _{NP}	時に改めて	部分行為 _{VP}
全体行為 _{NP}	時の	部分行為 _{NP} メッセージ
全体行為 _{NP}	締結時に	部分行為 _{VP}
全体行為 _{NP}	時に	部分行為 _{NP} が発生する
全体行為 _{NP}	に当たっての	部分行為 _{VP}
全体行為 _{NP}	内容に応じた	部分行為 _{NP}

5.2 獲得したパターン

最終的に獲得されたパターンは、初期パターンと意味的に類似するパターン、あるいは特殊化されたパターンが多かった。

表3と表4に部分全体関係で獲得されたパターンの例を示す。表4は初期パターンの特殊化されたパターンの例であるが、こうしたパターンは想起するのが困難であり、今回試みた手法が共起パターンの発見に役立つことを示唆している。

5.3 pmi を修正した効果

3.5で述べたように、Pantelら⁴⁾は pmi (式(3))の代りに pmi2 (式(4))を用いて、低頻度のインスタンスに負のバイアスをかけている。これがどの程度、精度向上に寄与しているかを調べるため、部分全体関係(繰返し4回)に pmi を用いた場合と pmi2 を用いた場合の精度を比較した結果を図4に示す。全体的に pmi2 を用いた方が精度が高い。よって、低頻度のインスタンスへ負のバイアスが精度向上に有効であると言える。

5.4 動詞句のみで獲得できる事態対の数

動詞句のみの共起よりも、事態含意名詞を用いることでより多くの事態間関係を獲得できると考えて、事態含意名詞を用いた。これが正しいことを確認する。

表5 両方の事態がともに動詞句として一度でもコーパス中で共起したインスタンスの例

全体行為	部分行為
攻撃する	光弾を発射する
コピーする	クリップボードへ追加する
締結する	約束する
決済する	クレジットカード番号を送信する
受信する	メールをチェックする

表6 事態間関係を用いて初めて獲得できたインスタンスの例

全体行為	部分行為
授業する	説明する
面接する	気持ちをアピールする
共用する	パスワードで保護する
契約する	調印する

部分全体関係(繰返し4回)のインスタンスの上位1000件のうち、両方の事態がともに動詞句として一度でもコーパス中に共起した(パターンは問わない)インスタンスはわずか279件であった。つまり、1000件中721件のインスタンスは動詞句のみの共起で獲得できる見込みがまったくないということである。

このことから、事態含意名詞を用いることで、動詞句のみの共起を用いた場合よりも、より多くの事態間関係を獲得することができる見込みがあることがわかる。

動詞句のみの共起で獲得できる関係(表5)と獲得できない関係(表6)の例を挙げる。

6. まとめ

実験の規模の拡大

本稿で報告した評価実験は規模が小さいため、十分に信頼できる結果を得るには至っていない。まずは実験の規模を拡大し、より正確な知見を得ることが急務の課題である。また、獲得サイクルの繰返しの中で徐々にパターンを増やす実験や、動詞句のテンスやアスペクト情報などによって共起パターンの粒度を

さらに細かくする実験など、調査すべき項目は数多く残っている。

関係抽出との違い

従来の関係抽出は実体を表わす固有表現間の関係の抽出が中心であり、そこでは固有表現そのものの抽象度の選択は問題にならなかった。一方、4.2 で述べたように、事態間関係抽出では事態表現の抽象度を定める問題が顕現する。今回は事態間関係を抽出するときのスコアが最適な抽象度の事態を自動的に選択すると期待して実験を行ったが必ずしも期待通りの結果が得られたとは限らない。これについてはさらに工夫が必要である。

機能動詞、複合辞

「病院に通う」と「通院する」、「仕事を始める」と「仕事する」、「スーパーで買い物をする」と「スーパーで買う」等、同じ事態だが異なる表現になっている場合がある。今回の実験では1億文強のデータを用いたが、それでもこうした表現の多様性を無視できる程に十分な量の共起データを集められたとは言えない。実験結果は、データがまだまだスパースであることを示している。したがって、今度は同じタイプの事態を表す事態表現の多様性を吸収する技術を効果的に組み込む必要がある。

述語項構造解析技術の適用

テキストからより高い精度で事態の共起を獲得するためには、省略された格を補ったり、2つの事態の項の間の共参照関係を同定したりする述語項構造解析が欠かせない。これについては、飯田らのゼロ照応解析技術⁹⁾や小町らの事態性名詞の述語項構造解析技術¹¹⁾を利用することを検討している。

7. 謝 辞

「Web上の5億文の日本語テキスト」の使用許可を下された情報通信研究機構の河原大輔氏と京都大学大学院の黒橋禎夫氏に感謝いたします。

参 考 文 献

- 1) Chklovski, T. and Pantel, P.: Global Path-based Refinement of Noisy Graphs Applied to Verb Semantics, *In Proceedings of Joint Conference on Natural Language Processing (IJCNLP-05)* (2005).
- 2) Lin, D. and Pantel, P.: DIRT - Discovery of Inference Rules from Text, *In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2001* (2001).
- 3) Pantel, P. and Pennacchiotti, M.: Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations, *the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pp.113-120 (2006).
- 4) Pantel, P. and Ravichandran, D.: Automatically La-

beling Semantic Classes, *In Proceedings of Human Language Technology / North American chapter of the Association for Computational Linguistics*, pp. 321-328 (2004).

- 5) R. Girju, A.B. and Moldovan, D.: Automatic Discovery of Part-Whole Relations, *Computational Linguistics*, Vol.32, No.1, pp.83-135 (2006).
- 6) Ryu Iida, Kentaro Inui, Y.M.: Anaphora resolution by antecedent identification followed by anaphoricity determination, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol.4, No.4, pp.417-434 (2005).
- 7) Takashi Inui, Kentaro Inui, Y.M.: Acquiring causal knowledge from text using the connective marker tame, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 4, No. 4, pp.435-474 (2005).
- 8) Torisawa, K.: Acquiring Inference Rules with Temporal Constraints by using Japanese Coordinated Sentences and Noun-Verb Co-occurrences, *In Proceedings of Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL06)*, pp.57-64 (2006).
- 9) 河原大輔, 黒橋禎夫: 高性能計算環境を用いたWebからの大規模格フレーム構築, 情報処理学会自然言語処理研究会 NL-171-12, pp.67-73 (2006).
- 10) 国立国語研究所: 国立国語研究所資料集 14 『分類語彙表一増補改訂版』, 大日本図書 (2004).
- 11) 小町守, 飯田龍, 乾健太郎, 松本裕治: 共起用例と名詞の出現パターンを用いた動作性名詞の項構造解析, 言語処理学会第12回年次大会論文集 (2006).
- 12) 竹内孔一, 乾健太郎, 藤田篤: 語彙概念構造に基づく日本語動詞の統語・意味特性の記述, レキシコンフォーラム, Vol.2, pp.85-120 (2006). ひつじ書房.
- 13) 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央, 阿部修也: 分類の根拠を明示した動詞語彙概念構造辞書の構築, 情報処理学会自然言語処理研究会 NL-169-18, pp.123-130 (2005).