

## 文脈類似度を用いた項の異なる事態間関係の自動分類

隅田 飛鳥<sup>†1</sup> 服部 元<sup>†1</sup> 小野 智弘<sup>†1</sup>

事態間関係は、二つの事態間の含意関係、因果関係などの意味的な関係であり、Q&A システムや対話システムをはじめとした様々な自然言語処理アプリケーションへの応用が期待されている。事態間関係を項の共有関係から分類すると、〈富士山,ヲ,登頂する〉と〈富士山,ニ,登る〉のように、項の同一である関係と、〈大学院,ニ,進学する〉と〈修士,ニ,なる〉のように項の異なる関係とに大別できる。〈〉は事態を、〈〉内の要素はそれぞれ項、格、述語を指す。前者の関係は、語彙統語パターンをはじめ、文脈類似度を元に関係を推定する手法なども提案されており、これまで活発に獲得が試みられてきた。一方、後者の関係は、前者の関係よりも多様な関係が存在するため、獲得が容易ではない。そこで、本稿では、文脈類似度を元にクラスタリングを行うことで、項の異なる事態間関係を自動分類する手法を提案する。

## Clustering a Large Number of Event Relations that Have Different Arguments by Using Distributional Similarity

ASUKA SUMIDA,<sup>†1</sup> GEN HATTORI<sup>†1</sup> and CHIHIRO ONO<sup>†1</sup>

Event relations, which are semantic relations between events, play a crucial role for applications, such as the Q&A system and the dialogue system and so on. To follow whether event relations share their arguments, event relations are classified into two types - (1) the event relations that share arguments (ex. A relation between “(Fuji-san, wo, touchou-suru)” (To success in reaching the summit of Mt. Fuji.) and “(Fuji-san, wo, noboru)” (To climb Mt. Fuji.) ) and (2) the event relations that have different arguments (ex. A relation between “(Daigaku-in, ni, singaku-suru)” (To enroll to a graduate school) and “(Shu-shi, ni, naru)” (To become a student of master degree) ). Many researchers have attempted to acquire relations of the first type, by using lexico-syntactic patterns or distributional similarity. On the other hand, to extract relations of the second type is difficult because of the diversity in their relations. In this paper, we propose an automatic method that classifies relations of the second type to cluster event relation candidates into some groups that only include the same relations, by using distributional similarity.

表 1 主要な事態間関係とその定義

種類	表記	定義
同義関係	$e_1 = e_2$	事態 $e_1, e_2$ の意味が同一である関係 <sup>2)</sup>
含意関係	$e_1 \xrightarrow{E} e_2$	事態 $e_1$ が成立するならば、事態 $e_2$ が成立する関係 <sup>2)</sup>
因果関係 ※	$e_1 \xrightarrow{C} e_2$	事態 $e_1$ が発生すると、事態 $e_2$ が引き起こされる関係 <sup>2), 9)</sup>
準備-利用関係	$e_1 \xrightarrow{PU} e_2$	特定の名詞に対する準備表現 $e_1$ と利用表現 $e_2$ の関係 <sup>8)</sup>

※ 本稿では、乾ら<sup>9)</sup> が定義した 4 種類の因果関係のいずれかにあてはまる場合、因果関係が成立していると考え

(1) 〈富士山,ニ,登頂する〉 $\xrightarrow{E}$ 〈富士山,ニ,登る〉	(7) 〈吉野山,ニ,行く〉 $\xrightarrow{PU}$ 〈桜,ヲ,見る〉
(2) 〈山小屋,ヲ,予約する〉 $\xrightarrow{PU}$ 〈富士山,ニ,登る〉	(8) 〈桜,ヲ,植樹する〉 $\xrightarrow{E}$ 〈桜,ヲ,植える〉
(3) 〈山小屋,ヲ,予約する〉 $\xrightarrow{C}$ 〈山小屋,ニ,泊まる〉	(9) 〈桜,ヲ,植樹する〉 $\xrightarrow{E}$ 〈花,ヲ,植える〉
(4) 〈山小屋,ニ,泊まる〉 $\xrightarrow{\text{注意事項}}$ 〈トイレ,ガ,有料だ〉	(10) 〈風邪薬,ヲ,飲む〉 $\xrightarrow{C}$ 〈風邪,ガ,直る〉
(5) 〈ウコン,ヲ,飲む〉 $\xrightarrow{C}$ 〈二日酔い,ヲ,予防する〉	(11) 〈河川,ガ,氾濫する〉 $\xrightarrow{C}$ 〈堤防,ガ,決壊する〉
(6) 〈下剤,ヲ,飲む〉 $\xrightarrow{C}$ 〈便秘薬,ヲ,服用する〉	(12) 〈自転車,ガ,パンクする〉 $\xrightarrow{C}$ 〈タイヤ,ヲ,交換する〉

図 1 事態間関係の例

### 1. はじめに

事態間関係は、含意関係、因果関係などの二つの事態間の意味的な関係であり、Q&A システムや対話システムをはじめとした様々な自然言語処理アプリケーションへの応用が期待されている。主要な事態間関係の定義を表 1 に示す。本稿では、事態を項、格助詞、述語からなる組とし、以降、〈〉が事態を表し、〈〉内の要素がそれぞれ項、格助詞、述語を表す。事態間関係の具体例を図 1 に示す。これらの事態間関係を用いた、初心者向けガイドシステムを例としてあげる。このシステムに「富士山に登頂する」と入力された場合、例 1 の関係を用いることで、「富士山に登る」に含意されていることがわかるので、例 2 より、その準備として「山小屋を予約する」ようアドバイスすることができる。さらに、例 3, 4 を併用することで、「山小屋に泊まる」際には、「トイレが有料」かもしれないという注意をユーザに喚起させることも可能となる。例 4 は、表 1 に示す典型的な事態間関係ではないが、このような応用アプリケーションには有益な知識となる。本稿では、事態間関係を表 1 のような典型的な関係のみならず、様々な事態間関係ごとに分類することを目指す。

<sup>†1</sup> 株式会社 KDDI 研究所  
KDDI R&D Laboratories, Inc.

これまで、事態間関係の獲得には、語彙統語パターン、文脈類似度がよく用いられてきた。語彙統語パターンは、「〈事態 1〉ため〈事態 2〉」など、関係獲得を目的に設計されたパターンであり、高い精度で事態間関係を獲得できることが知られている<sup>1),8),9),11)</sup>。しかし、例 10 のように、一般に広く浸透している関係の場合、「風邪を治すため風邪薬を飲む」のようにパターンにあてはまる形で記述されることは少なく、語彙統語パターンによって獲得できる事態間関係の数には限界がある。一方、文脈類似度を用いた手法<sup>4),6),7)</sup>では、「同じ文脈で出現する単語は意味的に類似している」という原則<sup>3)</sup>に基づき、周辺文脈の同一性から、単語間の関係の有無を計測する。直接パターンによって表現されていなくても、周辺文脈が類似していれば、関係性を識別できるため、語彙統語パターンを用いた手法と比較し、より広範な関係の獲得が狙える。しかしながら、既存の文脈類似度を用いた手法では、例 1 のような、項が同一である事態間関係の獲得を目的に設計されており、例 2 のように、項の異なる事態間関係の獲得は難しい。項が同一である事態間関係だけでなく、項の異なる事態間関係を獲得できれば、自然言語処理アプリケーションでこれらの知識を応用する際に、より広範な入力への対応が可能となる。

本研究では、文脈類似度を用いて、項の異なる事態間関係を関係ごとに分類することを試みる。項の異なる事態間関係を観察したところ、「**同一の事態間関係が得られる場合には、二つの事態に含まれる項あるいは述語間の関係が同一の関係となりやすい**」という傾向が確認できた。例として、図 1 の項間の関係をみると、含意関係を表す例 1,8,9 では項間の関係がいずれも同義関係、あるいは上位下位関係であり、因果関係を表す例 5,10,11 では項間の関係が部分全体関係か、症状と薬の関係である。そこで、項および述語間の関係を手がかりとして、事態間関係を分類する。しかしながら、このような項間の関係には、上位下位関係などの典型的な実体間関係をはじめ、症状と薬の関係のように、これまで設計されていない関係が含まれており、学習データを用意して分類するのは容易ではない。そこで、項および述語の対を、関係の種類とクラスタが一対一に対応するように、クラスタリングを行う。

## 2. 提案手法

本提案手法では、まず、Web 文書集合から事態間関係候補を生成し (Step1)、事態間関係候補に含まれる項および述語対をクラスタリングし (Step2)、事態間関係候補を意味的關係ごとにクラスタリングする (Step3)。以下、提案手法について詳しく説明する。

### 2.1 Step1. 事態間関係候補の生成

Step1 では、Web 文書集合から事態間関係候補を生成する。まず、Web 文書集合に格

解析処理を適用した結果から、述語とそれに対応する項と格助詞の組を一つの事態として抽出する。述語は動詞あるいは形容詞に、項は名詞に限定する。また、抽出した事態の出現頻度が  $\min f$  より小さい場合には、その事態を取り除く。

事態間関係候補を単純な組み合わせから生成すると関係候補数が爆発的に増大するため、項が類似している事態間関係候補に制限する。項の類似性は、類似度の高い名詞対の集合に含まれているか否かで判定する。名詞対は Web 文書集合から以下の語彙統語パターンを適用することで得られる。

$$n_1 \text{ の } n_2, n_1 \text{ pp } V n_2 \quad (1)$$

$n_1, n_2$  には名詞が、 $pp$  には助詞、 $V$  には任意の連体形の述語が代入される。また  $n_1, n_2$  の間に係り受け関係があるものに限定する。次に、抽出した名詞対から下式で表される相互情報量を計算し、その値が  $th$  以下の対を取り除き、類似度の高い名詞対の集合を獲得する。

$$PMI(w_1, w_2) = \log\left(\frac{f(w_1, w_2)N}{f(w_1)f(w_2)}\right) \quad (2)$$

但し、 $f(w)$  は語  $w$  の出現頻度、 $f(w_1, w_2)$  は語  $w_1, w_2$  が語彙統語パターン (式 1) を介して共起する頻度、 $N$  は名詞の出現頻度の総和である。これにより、周辺文脈が類似している名詞対を選別できる。

ここまでの処理で得られた類似度の高い名詞対を元に、事態間関係候補を生成する。具体的には、名詞対  $(n_1, n_2)$  中の  $n_1, n_2$  それぞれを項とする二つの事態を事態間関係候補とみなす。ただし、名詞  $n_2$  の曖昧性からくる事態間関係候補の生成誤りを減らすため、名詞  $n_2$  を含む事態  $\langle n_2, pp, p \rangle$  は、Web 文書集合中で少なくとも一回以上、“ $n_1 \text{ pat } n_2 \text{ pp}$ ” が述語  $p$  に係っている事態に制限する。 $pat$  は、語彙統語パターン (式 1) に相当し、助詞「の」あるいは助詞と連体形の述語を指す。例えば、名詞対 (吉野山, 桜) 中の“吉野山”から事態  $\langle \text{吉野山}, \text{ヲ}, \text{歩く} \rangle$  が、“桜”から、事態  $\langle \text{桜}, \text{ヲ}, \text{みる} \rangle$ ,  $\langle \text{桜}, \text{ト}, \text{名付ける} \rangle$  が得られたとする。“桜”を含む事態集合から、“吉野山の桜”と格助詞、述語が共起する事態に絞り込むと、“吉野山の桜と名付ける”という表現は出現しにくいので、事態間関係候補として ( $\langle \text{吉野山}, \text{ヲ}, \text{歩く} \rangle$ ,  $\langle \text{桜}, \text{ヲ}, \text{加工する} \rangle$ ) が生成できる。

### 2.2 Step2. 事態間関係候補中の項および述語対のクラスタリング

次に、生成された事態間関係候補中の項および述語対を、関係とクラスタが一対一に対応するようにクラスタリングする。ここでは、項、述語に関するどちらのクラスタリングもほぼ同じアルゴリズムで行うため、以下、項、述語のいずれかを指すものを語と表記する。

まず、以後の処理の性能および速度向上のため、クラスタリングする前に語対の相互情報

量を計算し、その値が閾値  $th$  以下のものを取り除く。相互情報量は式 2 を利用する。ただし、述語については、 $f(w_1, w_2)$  を述語  $w_1, w_2$  が係り受け関係にある頻度、 $N$  を述語の出現頻度の総和とする。

次に、語対を意味的關係とクラスタが一对一に対応するようにクラスタリングする。クラスタリングの入力となる特徴ベクトルは、係り受け関係にある語を用いる。例えば、名詞間関係候補 (吉野山, 桜) が与えられたとき、Web 文書集合から係り受け関係を調査することで、“吉野山”には述語集合 { 行く, 立ち寄る, 咲く, 植樹する } が、“桜”には述語集合 { 守る, みる } が得られる。また、同様に、(花, 桜) が与えられた場合、“花”には述語集合 { 咲く, 植樹する, みる } が、“桜”には述語集合 { 咲く, 植樹する, みる } が得られる。特徴ベクトルは、語対が与えられたとき、各語ごとにベクトルを生成し、次元が重ならないように結合することで得られる。各語のベクトルは、その係り受け語集合の各語をベクトルの次元に、語とその係り受け語が係り受け関係にある頻度を値とすることで得られる。

このようにして生成した特徴ベクトルからクラスタリングを行うと、種類の異なる関係候補を誤って類似していると計算してしまう危険性がある。例えば、先ほどの例であげた名詞間関係候補 (吉野山, 桜), (花, 桜) では、共に一つ目の名詞に係る述語として“咲く”, “植樹する”が共通して出現し、二つ目の名詞は同一であるため、これらの二つの関係候補のベクトル間には共通要素が多く、類似しているといえる。そのため、これらの関係が同一のクラスタに所属するように分類できる。しかしながら、実際には、前者の関係候補が部分全体関係であるのに対し、後者の関係候補が上位下位関係であるため、異なるクラスタに所属することが望ましい。

そこで、本稿では、このような判定誤りを避けるため、係り受け語集合の差が類似している関係候補は共通の関係にあるだろうと仮定し、分類を行う。ここでいう係り受け語集合の差とは、語対の各構成要素から得られる係り受け語集合から、共通する要素を取り除いた集合である。例えば、名詞間関係候補 (吉野山, 桜) では、各名詞に係る述語集合に共通する述語“咲く”, “植樹する”を除くことで、“吉野山”に係る述語集合の差は { 行く, 立ち寄る } に、“桜”に係る述語集合の差 { 守る, みる } になる。同様に (花, 桜) では、各名詞に係る述語集合に共通する述語“植樹する”, “咲く”, “みる”を除くことで、“花”の述語集合の差は空集合になり、“桜”の述語集合の差は { 守る } となる。上記の例で二つの関係候補の述語集合の差を比較すると、一つ目の名詞の述語集合の差の間は共通する述語がなく、二つ目の名詞の述語集合の差の間には、共通する述語として、“守る”しかないため、ベクトル間には一つしか共通要素しかないため、述語集合間の共通要素を除く前と比較して、類似度が

低くなる。そのため、異なる関係を同一クラスタに分類する誤りを減らすことができる。

### 2.3 Step3. 事態間関係候補のクラスタリング

Step3 では、事態間関係候補を関係ごとに分類するため、事態間関係候補中の項、述語間の対が所属するクラスタ、および格助詞を手がかりとして、クラスタリングする。具体的には、まず事態間関係候補  $(\langle n_1, pp_1, p_1 \rangle, \langle n_2, pp_2, p_2 \rangle)$  から項  $(n_1, n_2)$ 、格助詞  $(pp_1, pp_2)$ 、述語  $(p_1, p_2)$  の対を抽出する。つづいて、項  $(n_1, n_2)$ 、述語  $(p_1, p_2)$  の対それぞれが所属するクラスタ ID とそのクラスタの重心からの距離を算出する。各事態毎に、得られたクラスタ ID をベクトルの各次元に対応させ、そのクラスタの重心からの距離を値とするベクトルを生成する。さらに、格助詞  $(pp_1, pp_2)$  の種類をベクトルの各次元に対応させ、存在する場合には対応するベクトルの次元の値を 1 とするようにして、ベクトルを生成する。ここまでできた 3 つのベクトルを次元が重ならないように結合する。このようにして生成したベクトルが類似しているということは、事態間関係候補に含まれる項および述語間の関係がそれぞれ同一の関係をもっており、事態対間が類似していることを示唆する。そこでベクトルの類似度を元にクラスタリングすることで、意味的に関連しやすい事態間関係候補ごとに分類できる。

## 3. 実験

提案手法の有効性を評価するため、河原らの 5 億文コーパス<sup>5)</sup> から、ランダムにサンプリングした 126,348,981 文に対して、提案手法を適用した。本コーパスには、格解析機能付きの構文解析器 KNP3.0<sup>\*1</sup> が既に適用されている。項および述語間の関係のクラスタリングには CLUTO<sup>\*2</sup> を、事態間関係のクラスタリングには MAHOUT<sup>\*3</sup> を用い、ベクトル間類似度をコサイン類似度とする K-means アルゴリズムを採用した。また、相互情報量の閾値  $th$  は 0 に、事態の最低出現頻度を  $minf$  は 5 に設定した。

まず、Step1 により 461,142 対の名詞間関係候補が抽出でき、これを元に 16,918,687,634 対の事態間関係候補が生成できた。つづいて、Step2 にて、述語間関係候補が 40,809 対得られ、クラスタ数を名詞間関係候補で 5,000、述語間関係候補で 1,000 と設定して分類を行った。クラスタリングされた結果からそれぞれランダムに 5 個のクラスタを選択し、それぞれ 100 個について評価を行った。評価基準は、名詞間関係の上位下位関係、部分全体関係は<sup>2)</sup>

\*1 <http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/knp.html>

\*2 <http://glaros.dtc.umn.edu/gkhome/views/cluto/>

\*3 <http://mahout.apache.org/>

に、概念-属性関係は<sup>10)</sup>に、述語間関係は表1に従う。なお、クラスタのサイズが100個に満たない場合は、全て評価した。その結果得られた、名詞間関係の例を表2に、述語間関係の例を表3に示す。表の行は抽出したクラスタを、列は、左から順に、クラスタID、クラスタ内での関係の割合および実際に得られた例を示す。ただし、どのクラスタにも出現しなかった関係は記述しなかった。表2より、名詞間関係は、全体的な傾向として部分全体関係と概念-属性関係が得られやすく、特にクラスタ1からは部分全体関係が得られやすく、クラスタ4からは概念属性関係が得られやすい傾向が得られた。また、表3より、述語間関係は、全体的な傾向として含意関係が得られやすく、特にクラスタ2から含意関係が、クラスタ3から因果関係が、クラスタ4から対義関係が得られやすい傾向が得られた。以上の結果から、Step2の段階では、クラスタ毎に関係を分離しやすい傾向が得られた。

Step2の結果を用い、Step3により、事態間関係候補を関係ごとにクラスタリングした。その際、Step2で類似度が低いため削除した述語対候補を含むものを除き、13,871,814対の事態間関係候補を入力とし、クラスタ数を1,000と設定してクラスタリングを行った。クラスタリングされた結果からそれぞれランダムに5個のクラスタを選択し、それぞれ100個について評価を行った。評価基準は表1に従う。実際に分類した事態間関係の例を表4に示す。表の見方は表2、表3と同様である。また表中の“↔”は誤った関係を表す。表4より、全体的な傾向としては、因果関係が出現しやすく、特にクラスタ3から得られやすいことがわかった。一方、クラスタ3以外のクラスタは、そのほとんどに関係性がなかった。この主な原因としては、1) 事態が単独で成立せず他の文脈を必要とする点、2) 事態間の関係性が十分に捉えられていない点である。原因1は、事態〈近く, 二, 引く〉のように、任意格と述語の関係が多く、周辺文脈なしに事態が成立しない点である。この原因を除くために、予め任意格であるか否かを格フレーム辞書等を利用して取り除くことが考えられる。原因2は、〈太平洋沿岸, 二, 移動〉↔〈カプセル, ガ, ない〉のように、個々のイベントが単独では成立しても、共起して成立しないという点である。この問題は、各事態の周辺文脈を利用することで解消する可能性がある。

4. ま と め

本稿では、項が異なる事態間関係を項と述語を関係毎にクラスタリングした結果を手がかりとして関係毎に分類する手法を提案した。Step2の項、述語単独でのクラスタリングでは、比較的關係毎に分類できるようになったものの、Step3の事態間関係のクラスタリングでの分類は不十分であった。今後はStep2の精度をさらに向上させることに加え、単独で

表2 Step2で分類した名詞間関係の例

ID	クラスタサイズ	上位下位関係の割合	部分全体関係の割合	概念属性関係の割合	例
1	52	4% (2)	35% (18)	10% (5)	(フィレンツェ, メディチ), (習熟, 度合い), (車体, 傾き), (素食, 店), (猫好き, 人間), (当期, わが国経済), (丘, 住宅地), (宮沢賢治, 童話), (関西電力, 石川博志社長), (警視庁, 刑事)
2	43	0% (0)	30% (13)	9% (4)	(塩山, 街), (塩分, 働き), (塩基, 置換), (塩谷, 特徴), (塩ラーメン, 店), (塩分, 鉛直分布), (塾, 教室), (塩野, 本), (塩分, 摂取量), (塩漬, 桜)
3	87	7% (6)	16% (14)	14% (12)	(美, 塊), (リーダー格, 警備員), (債権回収, 手紙), (美容, 商品), (私, 議員活動), (シャツ, 襟), (駒, 小屋), (真心, 種), (飼い主, 足元), (電源, 入り)
4	26	4% (1)	27% (7)	19% (5)	(お買い物, お客), (お買い物, フレーズ特集), (お近く, ホームセンター), (お近く, 公共職業安定所), (お近く, アニメイト), (お車, 状態), (お買い物, 金額), (お近く, 民生委員), (お車, 詳細), (お身体, 具合)
5	64	3% (2)	27% (17)	6% (4)	(財源, 均衡), (調達コスト, 上昇), (貴方, 味覚), (自分, ベース配分), (振込手数料, 額), (内臓, 成分), (ミュージシャン, 曲), (欧米, 企業), (貴社, お役), (法令, 合憲)

表3 Step2で分類した述語間関係の例

ID	クラスタサイズ	同義関係の割合	因果関係の割合	含意関係の割合	対義関係の割合	例
1	35	0% (0)	7% (1)	28% (4)	0% (0)	(いう, 呼び出す), (いう, 定義), (よい, 呼び出す), (よい, 定義), (ディスパッチ, 呼び出す), (リセット, 呼び出す), (ロック, 呼び出す), (保証, 呼び出す), (停止, 呼び出す), (再, 呼び出す)
2	13	0% (0)	53% (8)	0% (0)	0% (0)	(ない, 面白い), (分析, 取り上げる), (報道, 取り上げる), (工夫, 取り上げる), (思える, 面白い), (批判, 取り上げる), (検証, 取り上げる), (考察, 取り上げる), (要望, 取り上げる), (解説, 取り上げる)
3	15	7% (1)	0% (0)	0% (0)	0% (0)	(いう, 引き続く), (いく, 引き続く), (保有, 引き続く), (強める, 引き続く), (徹底, 引き続く), (持続, 引き続く), (推移, 引き続く), (点検, 引き続く), (発展, 引き続く), (監視, 引き続く)
4	29	0% (0)	0% (0)	31% (9)	17% (5)	(いく, とる), (いく, 大), (よい, 独立), (上昇, 低下), (下ろす, 持ち上げる), (代理, 補佐), (公表, 指示), (制御, 独立), (営む, 独立), (投稿, 間違える)
5	53	0% (0)	0% (0)	15% (8)	0% (0)	(いう, コミット), (いう, ダイエット), (いう, 仕組む), (いう, 伝承), (いう, 侵す), (いう, 凌ぐ), (いう, 分断), (いう, 分配), (いう, 到来), (いう, 創造)

表 4 Step3 で分類した事態間関係の例

ID	クラスサイズ	因果関係の割合	含意関係の割合	例
1	18977	7%	0%	〈 フォント, ガ, 表示 〉 → 〈 指定, ガ, 分割 〉, 〈 ボタン, ガ, 表示 〉 → 〈 保存, ガ, 受信 〉, 〈 何文字, ガ, 表示 〉 → 〈 行, ガ, 実行 〉, 〈 データ, ガ, 受信 〉 $\xrightarrow{C}$ 〈 集計, ガ, 表示 〉
2	815	7%	0%	〈 認可, ヲ, 取り消す 〉 → 〈 設立, ガ, 取得 〉, 〈 中止, ヲ, 求める 〉 → 〈 使用, ガ, 対処 〉, 〈 中止, ヲ, 考える 〉 → 〈 使用, ガ, 自然 〉, 〈 向き, ヲ, 変更 〉 → 〈 印刷, ガ, 文字 〉
3	2629	24%	5%	〈 企業, ガ, いる 〉 $\xrightarrow{C}$ 〈 協力, ヲ, 進める 〉, 〈 企業, ガ, 組織 〉 $\xrightarrow{C}$ 〈 再生, ヲ, 推進 〉, 〈 企業, ガ, 依頼 〉 $\xrightarrow{C}$ 〈 広告, ヲ, 制作 〉
4	490	9%	0%	〈 太平洋沿岸, ニ, 移動 〉 → 〈 日本, ガ, ある 〉, 〈 表面, ニ, 開発 〉 → 〈 封筒, ガ, ある 〉, 〈 制度, ニ, 生まれ変わる 〉 → 〈 国民, ガ, ある 〉, 〈 職員, ニ, 及ぶ 〉 → 〈 一般, ガ, ある 〉
5	1322	3%	0%	〈 仕方, ヲ, 言う 〉 → 〈 把握, ヲ, よい 〉, 〈 仕方, ヲ, 検討 〉 → 〈 把握, デ, 比較 〉, 〈 仕方, カラ, 記述 〉 → 〈 把握, ガ, ある 〉, 〈 仕方, ガ, 自由だ 〉 → 〈 把握, ガ, ある 〉, 〈 仕方, ヲ, 習得 〉 → 〈 把握, ガ, ある 〉

成立する事態に制限し、事態の周辺文脈を併用することで、性能が向上する可能性がある。

## 参 考 文 献

- 1) Chklovski, T. and Pantel, P.: VerbOcean: Mining the web for fine-grained semantic verb relations, *Proceedings of EMNLP*, Vol.4, pp.33–40 (2004).
- 2) Fellbaum, C.: *WordNet: An Electronic Lexical Database*, The MIT Press (1998).
- 3) Harris, Z.: Distributional structure, *Word*, Vol.10, No.23, pp.146–162 (1954).
- 4) Hashimoto, C., Torisawa, K., Kuroda, K., Saeger, S.D., Murata, M. and Kazama, J.: Large-scale verb entailment acquisition from the web, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pp.1172–1181 (2009).
- 5) Kawahara, D. and Kurohashi, S.: A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis, *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, Association for Computational Linguistics, pp.176–183 (2006).
- 6) Lin, D. and Pantel, P.: DIRT-discovery of inference rules from text, *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp.323–328 (2001).
- 7) Szpektor, I. and Dagan, I.: Learning entailment rules for unary templates, *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pp.849–856 (2008).
- 8) Torisawa, K.: Automatic acquisition of expressions representing preparation and

utilization of an object, *Proceedings of the Recent Advances in Natural Language Processing*, pp.556–560 (2005).

- 9) 乾 孝司, 乾健太郎, 松本裕治: 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得, *情報処理学会論文誌*, Vol.45, No.3, pp.919–933 (2004).
- 10) 徳永耕亮, 風間淳一, 鳥澤健太郎: 属性語の Web 文書からの自動獲得と人手評価のための基準, *自然言語処理*, Vol.13, No.4, pp.49–67 (2006).
- 11) 阿部修也, 乾健太郎, 松本裕治: 項の共有関係と統語パターンを用いた事態間関係獲得, *自然言語処理*, Vol.17, No.1, pp.121–139 (2010).