情報処理学会論文誌 Vol. 52 No. 1 293–307 (Jan. 2011)

WWWからの大規模動詞含意知識の獲得

橋本 力[†1] 鳥澤 健太郎[†1] 黒田 航[†2,†3]
デサーガ ステイン[†1] 村田 真樹[†4] 風間 淳一[†1]

テキスト間含意関係認識と呼ばれる技術は，深い自然言語理解を必要とするタスクにおいて重要な役割を果たす．この技術が実用レベルに至るには，大規模な含意知識ベースの構築が不可欠である．本稿では，動詞間含意関係知識の大規模な獲得を目的として，条件付き確率に基づく方向付き類似度尺度を提案する．提案手法の評価実験では，WWW 上の日本語 1 億文書から得られた 52,562 動詞（異なり）を対象とした．この動詞セットには，日常的に使用される動詞も特定の専門的な領域でのみ用いられるような動詞も区別せず含まれている．提案手法と先行研究の手法それぞれのスコア上位 20,000 位までの出力からランダムに選ばれた 200 サンプルを人手評価したところ，比較対象のすべての先行研究の手法の精度を提案手法の精度が上回ることを確認した．また，提案手法のスコア上位 100,000 の出力を人手評価したところ，大規模動詞含意知識ベースを構築する出発点としてリーズナブルな精度が得られていることを実験により確認した．

Large-scale Verb Entailment Acquisition from the Web

Chikara Hashimoto,[†1] Kentaro Torisawa,[†1]
Kow Kuroda,[†2,†3] Stijn De Sager,[†1]
Masaki Murata[†4] and Jun'ichi Kazama[†1]

Textual entailment recognition plays a fundamental role in tasks that require in-depth natural language understanding. For entailment recognition technologies to serve for real-world applications, a large-scale entailment knowledge base is indispensable. This paper proposes a conditional probability based directional similarity measure to acquire verb entailment pairs on a large scale. We targeted 52,562 verb types that derived from 10^8 Japanese Web documents, regardless whether they were used in daily life or only in specific domains. Evaluating 200 samples that were chosen randomly from the top 20,000 verb entailment pairs acquired by previous methods and ours, we found that our similarity measure outperformed the previous ones. For the top 100,000 results, our method worked well too.

1. はじめに

我々は，通常，次のような事柄を常識として知っている．たとえば，誰かがいびきをかいているならその人は寝ているだろうし，誰かが離婚したならその人はそれ以前に結婚していたはずである．また，誰かが勝訴したということはその人がそれ以前に他の誰かを告訴したということを意味する．このような，ある事態間に必然的に成立する関係を含意と呼ぶ．より正確には，動詞 1 が動詞 2 を含意するとは，動詞 1 の表す事態が成立するなら，動詞 2 の表す事態も成立しているということを意味する．WordNet3.0 [⋆1]でも同様の定義がなされている．含意に関する知識は，質問応答システムや自然言語インタフェース等の深い自然言語理解を必要とするシステムにおいて不可欠である．たとえば質問応答システムの場合，テキストデータベース中の「花子は太郎と離婚した」という文から「花子は（それ以前に）太郎と結婚した」ということを推論することができれば，「花子と結婚したのは誰か」という質問に答えることができる．この推論をするためには，質問応答システムが，「離婚する」という事態の成立が，それ以前にその離婚相手と結婚していたという事態の成立を含意するということを認識できる必要がある．

本稿では，「離婚する → 結婚する」[⋆2]や「勝訴する → 告訴する」といった含意関係が成立する動詞ペアを大規模に獲得するための新たな類似度尺度を提案する．動詞ペアの獲得元として，数多くのトピックが混在する世界最大のテキストデータベースである WWW を採用した．正確には，WWW 上にある日本語で書かれた 1 億文書から構築した，日本語ウェブコーパス[11] を使用した．

†1 情報通信研究機構
National Institute of Information and Communications Technology
†2 京都工芸繊維大学
Kyoto Institute of Technology
†3 早稲田大学総合研究機構
Comprehensive Research Organization, WASEDA UNIVERSITY
†4 鳥取大学
Tottori University

⋆1 WordNet3.0 には synset 間の含意関係知識が与えられている．たとえば，2 つの synset「divorce, split up」と「marry, get married, wed, conjoin, hook up with, get hitched with, espouse」の間に含意関係が付与されている．

⋆2 含意関係が成立する動詞ペアを「動詞 1 → 動詞 2」のように表記する．動詞 1 によって表される事態が動詞 2 によって表される事態を含意する．

本稿で提案する類似度尺度は，同じような文脈に出現する語は同じような意味を持つという分布類似度仮説[8] に基づいている．動詞の文脈として，Lin ら[14]，Szpektor ら[21] 等の類似度尺度と同様，注目する動詞の主語や目的語等の項を採用する．一方，先行研究と異なるのは，我々の類似度尺度が条件付き確率に基づくという点である．これにより，低頻度語を過大に評価しがちな相互情報量の使用を避けつつ高い精度を確保した．加えて提案手法では，コーパスの偏りによって生じる偶発的な高含意スコアを抑制する単純なトリックを取り入れている．これにより含意獲得の精度を大きく向上させることができた．評価実験では，提案手法と比較対象の先行研究の手法それぞれで動詞ペアを獲得し，それぞれのスコア上位 20,000 ペアからランダムに選ばれた 200 サンプルを人手でチェックした．その結果，提案手法の精度が比較対象のすべての先行研究の精度を上回った．また，提案手法のスコア上位 100,000 の出力を人手評価したところ，大規模動詞含意知識ベースを構築する出発点としてリーズナブルな精度が得られていることを確認した．

近年，自然言語処理技術の適用範囲は特定のドメインから不特定のドメインへと拡大しつつある．そのため，動詞含意関係データベース等の自然言語処理リソースも，日常用語か専門用語かを問わずあらゆる表現を幅広くカバーしなくてはならない．後述するように，提案手法は，日常的に使用される動詞だけでなく，「アポトーシスする → 死ぬ」や「ポワレする → 焼く」等の日常的には使用されない専門的な動詞からなる動詞含意を獲得することができる．

従来の動詞に関する含意知識獲得の研究の多くは，「X が Y を焼く」のような，変数と助詞，動詞から構成されるテンプレートの間の含意関係に焦点を当ててきた．確かに，テンプレート間の含意知識は動詞間のよりシンプルな含意知識に比べて，質問応答等の自然言語処理アプリーションにおいて有用であろう．しかし動詞ペアの中には，「離婚する → 結婚する」等，それ自体で（主語や目的語等を考慮しなくても）含意関係が成り立つものが存在する．そしてそのような動詞ペアは，今後自然言語処理コミュニティにおいて継続的に整備されるであろう含意知識ベースにおける核となる．本研究で動詞間の含意知識に焦点を当てるのは以上の理由による．しかし，後述するように，本研究の手法はテンプレート間の含意知識も十分な精度で獲得可能である．

以下本稿では，次章で関連研究を概観し，3 章で提案手法について詳述する．4 章では評価実験結果について報告する．最後に 5 章で結論を述べる．

2. 関 連 研 究

従来の動詞関連の含意知識獲得研究は，コンパラブルコーパスを用いるもの[3),10),18)] と用いないもの（分布類似度によるもの）[4),14),21),24)] の 2 種類に大きく分けることができる．以下，本章では，言い換えや推論知識の獲得も含める．言い換え表現対は，互いに相手を含意すると見なせる．

Shinyama ら[18] は，同じ出来事について記述された新聞記事群を言い換え事例が含まれるコンパラブルコーパスと見なし，固有表現を手がかりに言い換え知識を獲得した．この手法は，テキストが言い換えられても，その中の固有表現は言い換えられず元の表現のまま残り，かつ，複数の固有表現を共有しているテキストは言い換えの可能性が高い，という仮説に基づいている．Barzilay ら[3] も同様に，同一事件を扱っている新聞記事群をコンパラブルコーパスと見なし，そこから言い換え知識を獲得している．Ibrahim ら[10] は，非英語圏の小説の複数の英訳をコンパラブルコーパスと見なし，言い換え知識を獲得した．

WWW の爆発的拡大とともにコンパラブルコーパスの収集は容易になってきてはいるが，それでもなお，収集可能な量は限られている．つまり上記の手法では動詞含意知識を大規模に獲得するのは困難であるため，本研究ではこのアプローチを採用しなかった．

コンパラブルコーパスを用いない手法の多くは分布類似度仮説[8] に基づいている．つまり，テキストを構文解析して述語とその項（主語，目的語等）のペアを獲得し，重みづけされた項を述語の特徴と見なして，述語間の方向性のある類似度，つまり含意関係らしさを計る．

Lin ら[14] は，2 つの変数からなるテンプレートを対象に，DIRT（Discovery of Inference Rules from Text）と呼ばれる言い換え知識獲得手法を提案した．以降では，2 つの変数からなるテンプレートを二項テンプレートと呼ぶ．たとえば「X が Y を焼く」は，2 つの変数 X と Y を持つ二項テンプレートである．一方，「X が眠る」は，1 つの変数 X を持つ一項テンプレートである．DIRT は，Lin[13] が提案した，一項テンプレート間の類似度を計算する次の式を用いている．

$$Lin(l,r) = \frac{\sum_{f \in F_l \cap F_r} [w_l(f) + w_r(f)]}{\sum_{f \in F_l} w_l(f) + \sum_{f \in F_r} w_r(f)} \quad (1)$$

l と r は一項テンプレート，F_x は一項テンプレート x の変数の値として出現する名詞の集合，$w_x(f)$ は一項テンプレート x の変数の値として出現する名詞 $f \in F_x$ の重みである．重

みとして一項テンプレート x と f の相互情報量（Pointwise Mutual Information，PMI）が用いられている．DIRT は，式 (1) で得られる一項テンプレート間の類似度 2 つの相乗平均を計算することで二項テンプレート間の類似度を計算する．たとえば，「X が ポワレする → X が 焼く」の一項テンプレート間類似度と「Y を ポワレする → Y を 焼く」の一項テンプレート間類似度の相乗平均が「X が Y を ポワレする → X が Y を 焼く」の類似度となる．この手法は，テンプレート間でより多くの名詞が共有されていれば，意味的により類似している，という仮説に基づく．我々の手法は二項テンプレートではなく一項テンプレートに基づいて動詞含意を獲得するので，後述の評価実験では，DIRT ではなく式 (1) をそのまま用いた．以下では式 (1) を Lin と表す．

DIRT は 2 つのテンプレートが意味的に類似しているかどうかは判定するが，どちらがどちらを含意するか，つまり方向性は示せない．Bhagat ら[4] は，LEDIR（LEarning Directionality of Inference Rules）と呼ばれる，DIRT のような方向性のない類似度尺度が出力する意味的に類似したテンプレートペアに対して方向性を付与する手法を提案している．この手法は，2 つのテンプレートが類似した文脈で現れ，かつ，一方が他方より広い文脈において出現するなら，後者が前者を含意する，という方向性仮説に基づいている．

Weeds ら[24] は，Precision と Recall と名付けられた概念から構成される，分布類似度の一般的な枠組みを提案している．以下では本提案手法と直接比較可能な Precision のみを取り上げる．Precision は次のように定義される．

$$Precision(l,r) = \frac{\sum_{f \in F_l \cap F_r} w_l(f)}{\sum_{f \in F_l} w_l(f)} \tag{2}$$

本研究においては，l と r は一項テンプレートに該当し，F_x は一項テンプレート x の変数の値として出現する名詞の集合，$w_x(f)$ は名詞 $f \in F_x$ の重みに該当する．重みとして相互情報量（PMI）が最も高精度だったことが Weeds ら[24] によって報告されている．Precision は，F_l が F_r によってどのくらいカバーされているかを調べる方向性のある類似度である．

Szpektor ら[21] は，BInc（Balanced Inclusion）という名前の方向性のある類似度計算法を提案している．BInc は Lin と Precision から構成される．

$$BInc(l,r) = \sqrt{Lin(l,r) \times Precision(l,r)} \tag{3}$$

l と r は一項テンプレートである．重みづけには相互情報量（PMI）を用いている．従来，テンプレート間の含意知識獲得研究では二項テンプレートを対象とするのが一般的だったが，Szpektor らは一項テンプレートの使用を提案した．一項テンプレートには，自動詞や受け身形の動詞，項が省略された動詞も扱えるという利点がある．日本語は項の省略が頻繁に起きるため，一項テンプレートの利点が特に際立つ．

後述するように，本研究の評価実験では，我々の手法は Lin，Precision，BInc のいずれよりも高い精度を示した．

上記以外の含意知識獲得手法として，Torisawa[23] は日本語の複文の構造を利用して動詞含意を獲得した．Pekar[17] は，テキスト中の局所的な文脈に存在する関連性のある節を同定することで動詞含意を獲得した．Zanzotto ら[26] は，動作主名詞を利用して動詞含意を獲得した．たとえば，“the player wins.” から「win → play」を獲得する．

動詞間ではなく，名詞間の含意知識獲得の研究として，Geffet ら[7] は，名詞 v が名詞 w を含意するなら，v の特徴を表すすべての素性が w でも現れる（そしてその逆もまた成立する），という分布包含仮説を提案している．

上位下位関係（IS-A 関係，たとえば「りんご IS-A 果物」）も名詞の含意関係と見なすことができる．上位下位関係の自動獲得の研究として，言語表現パターンを用いるもの[1),9)]，クラスタリングに基づくもの[5),16)]，HTML 文書の構造を利用するもの[19)]，Wikipedia の構造を利用するもの[15),20),25)] 等がある．

3. 提 案 手 法

本章では本研究の動詞含意知識獲得手法について詳述する．本手法では，まず一項テンプレート間の含意知識を獲得し，その獲得結果を動詞間の含意知識へと変換する．

3.1 節では我々が開発した一項テンプレート間の方向付き類似度尺度（以後 Score と呼ぶ）について，3.2 節では動詞含意知識の獲得源であるテンプレート共起名詞データベースについて，3.3 節では Score によるテンプレート共起名詞データベースからの動詞含意知識獲得のプロセスについて述べる．

以下では，一項テンプレートを $\langle p, v \rangle$（p は助詞で，v は動詞を表す）のように記述する．また，一項テンプレート間の含意関係を $l \rightarrow r$（ただし，$l = \langle p_l, v_l \rangle$，$r = \langle p_r, v_r \rangle$．$p_x$ は一項テンプレート x の助詞で，v_x は x の動詞）のように記述する．なお，以降では，単にテンプレートという場合，それは一項テンプレートを指すものとする．

3.1 条件付き確率に基づく方向付き類似度

我々が開発した方向付き類似度尺度 Score は次のように定義される．

$$Score(l,r) = Score_{base}(l,r) \times Score_{trick}(l,r) \tag{4}$$

l と r は一項テンプレートで，Score は「l→r」の含意関係らしさを表す．

Score$_{base}$ は Score の根幹にあたり，次のように定義される．

$$Score_{base}(l,r) = \sum_{f \in F_l \cap F_r} P(r|f)P(f|l) \quad (5)$$

F_x は一項テンプレート x の変数として出現する名詞（複合名詞を含む）の集合であり，$f \in F_x$ は F_x に属する名詞である．以降では，$f \in F_x$ を一項テンプレート x との共起名詞と呼ぶ．式 (6) に示すように，Score$_{base}$ は，与えられた 2 つの動詞 v_l，v_r の含意関係らしさに対応する条件付き確率 $P(v_r|v_l)$ を，対応する 2 つの一項テンプレート l，r の条件付き確率 $P(r|l)$ で近似し，さらに $P(r|l)$ を，l，r それぞれの共起名詞の重複の度合いで近似したものである．

$$P(v_r|v_l) \approx P(r|l) \approx \Sigma_{f \in F_l \cap F_r} P(r|f)P(f|l) \quad (6)$$

言い換えれば，一項テンプレート r，l 間で共有されている（特徴的な）共起名詞 f が多いほど r と l が「同時に観測」されている（$P(r|l)$ の値が大きい）と見なす．$P(r|l)$ の値が大きければ，$P(v_r|v_l)$，つまり動詞 v_r，v_l 間の含意関係らしさも高いと考える．

$P(v_r|v_l)$ は v_l と v_r が含意関係にある場合，つまり，v_l の表す事態が成立するなら v_r の表す事態も必ず成立するという場合において 1 になる．つまり，もし直接的に $P(v_r|v_l)$ を推定でき，かつ，その値が十分大きいなら，含意関係「$v_l \rightarrow v_r$」が成立すると判断できる可能性がある．しかし，通常，含意関係にある 2 つの動詞を同時に観測することはほとんどないため[⋆1]，$P(v_r|v_l)$ を直接求めるのは難しい．そのため，我々は $P(v_r|v_l)$ を Score$_{base}$ で近似した．

Score$_{base}$ には，Torisawa [22] の研究に触発されたもう 1 つの論拠がある．それは，我々が予備実験の段階で得た，「動詞 v_l が表す動作のための道具の多くが，動詞 v_r が表す動作をその使用目的としているといえるなら，v_l と v_r の間には含意関係『$v_l \rightarrow v_r$』が成立しやすい」という観察結果に基づく．この議論では，助詞としてデ格（道具格）のみを対象とし，デ格でマークされた名詞（f）を道具と見なす．そして，$P(f|l)$ を，f が一項テンプレート l（の動詞 v_l）の動作を遂行する際に用いられる道具としての尤もらしさと見なす．一方，Torisawa [22] では，$P(r|f)$ は，一項テンプレート r（の動詞 v_r）の動作が道具 f の使用目的としてどの程度尤もらしいかを表すと見なされる．この議論に基づけば，$Score_{base}(l,r) = \Sigma_{f \in F_l \cap F_r} P(r|f)P(f|l)$ は，上記の観察結果を定式化したものと見なすことができる．つまり Score$_{base}$ は，$P(f|l)$ が大きく（f が l の動詞 v_l の動作のための道具として尤もらしく），$P(r|f)$ も大きい（r の動詞 v_r の動作が f の使用目的として尤もらしい）ような名詞（道具）f が多いほど，l と r（つまり v_l と v_r）の間には含意関係が成立しやすい，ということを表している．

例として，v_l として「ソテーする」を，v_r として「加熱する」を，対応する一項テンプレート l，r としてそれぞれ「X で ソテーする」「Y で 加熱する」をあげる．含意関係「ソテーする → 加熱する」は成立するが，実際，「ソテーする」の道具 f のほとんど（「フライパン」「中華鍋」「ガスコンロ」等）は，「加熱する」を使用目的として持つということができる．一方「ソテーする → 焼き払う」は成立しないが，このことは，「ソテーする」の道具 f のほとんどが「焼き払う」を使用目的として持たないことから予測することができる．

当初，本研究は，助詞をデ格に限定した上記の議論から出発したが，いくつかの予備実験を通して，他の助詞を対象に含めても Score$_{base}$ の精度を維持できることが分かったため，最終的に他の主な助詞[⋆2]を含むように Score$_{base}$ を一般化した．

Score$_{base}$ の $P(r|f)$ と $P(f|l)$ は次のように最尤推定により得た．

$$P(r|f) = \frac{freq(r,f)}{freq(f)} \quad (7)$$

$$P(f|l) = \frac{freq(l,f)}{freq(l)} \quad (8)$$

$freq(x)$ は x の出現頻度，$freq(x,y)$ は x と y の共起頻度である．

Score$_{base}$ は，Lin，Precision，BInc と，2 つの一項テンプレート間の共起名詞の重複を手がかりにするというアイデアを共有しているが，相互情報量による共起名詞の重みづけをしていない点がそれらの先行研究とは異なる．相互情報量は低頻度語を過大評価する傾向が知られており [27]，「アポトーシスする」や「ポワレする」等の日常的には使用されない専門的な（それゆえ低頻度な）動詞も対象とする本研究では相互情報量を用いない手法が望ましい．実際，4 章で述べるように，提案手法の低頻度動詞に対する頑健性が評価実験により確認された．

一方 Score$_{trick}$ は，Score$_{base}$ や Lin，Precision，BInc 等，分布類似度に基づく手法が陥りやすい問題を軽減する役割を果たす．その問題とは，本来含意関係にないテンプレートペ

⋆1 たとえば，含意関係にある 2 つの動詞「アポトーシスする」と「死ぬ」が同一文中に現れるのは意味的に冗長であり，そのため両者が同時に出現するのは稀である．

⋆2 具体的には，3.2 節で述べるように，助詞として「ハ」「ガ」「ヲ」「ニ」「デ」を対象に含めた．

アに対して，テンプレート間で共有される（少数の）共起名詞のうちの 1 つによって，高い類似度が誤って付与されるというものである．Score_{base} の評価実験（4.2 節）で実際にあった例として「X を 代行入手する → Y を 引用する」というペアをあげる（前者 l，後者 r とする）．このペアは含意関係にないが，実験で用いたコーパスにおいて，「メーカーカタログ等」と「作品」の 2 つの共起名詞を共有している．このうち，「メーカーカタログ等」を f とした場合の $P(r|f)P(f|l)$，より正確には $P(f|l)$ が非常に高い値になった．これは，実験で用いたコーパスにおいて，l（「X を 代行入手する」）の出現回数 61 回のうち 55 回が「メーカーカタログ等」と共起するというコーパスの偏りが原因だった．これにより，本来含意関係にないペア「X を 代行入手する → Y を 引用する」に高い類似度が付与された．コーパスに偏りがなければ，問題の l は，もう 1 つの共起名詞である「作品」や，コーパスでは l と共起していなかった（l の変数の値として出現しなかった）他の名詞，たとえば「チケット」「景品」「見積もり書」等とも「メーカーカタログ等」と大差ない回数で共起する可能性が高く，その結果，$P(\text{メーカーカタログ等}\,|l)$ が不自然に高い値にならなかった可能性がある．

Score_{trick} は，このようなコーパスの偏りに対応する手段として開発された．我々はまず，「X を 代行入手する → Y を 引用する」のような問題のあるペアを調べ，その結果，Score_{base} の値（類似度）に寄与する名詞がただ 1 つで，他の名詞の貢献が無視できるほど小さい場合，その類似度の信頼性は低い，という仮説を立てた．実際，「X を 代行入手する → Y を 引用する」の場合，Score_{base} の値の 99.99%以上が「メーカーカタログ等」によってもたらされていた．そして，この仮説を Score_{trick} として次のように定式化した．

$$Score_{trick}(l,r) = Score_{base}(l,r) - \max_{f \in F_l \cap F_r} P(r|f)P(f|l) \tag{9}$$

言い換えれば，類似度への貢献度が最大の名詞を無視することで，複数の名詞によって安定的に高い類似度が得られるペアだけを含意ペアと見なすようにした．これにより，偶発的に高い（誤った）類似度を抑制することができると考えられる．4 章で述べるように，実際にこの単純なトリックにより精度が大きく向上した．

Bannard ら[2] と Fujita ら[6] は Score_{base} と同様の方向付き類似度尺度を提案している．しかし Bannard らの手法は，類似度計算対象の表現の翻訳群を素性とし，対訳コーパスを必要とする点で本研究と異なる．一方，Fujita らの研究は，あらかじめ用意された言い換えパターンに合致する述語句ペア，つまり意味的な類似性がある程度保証されたペアのみを処理対象としている点が本研究と異なる．また Fujita らは，素性（述語句との共起名詞）の抽出元を，述語句をクエリとして得た WWW 検索結果のうちの上位 1,000 件のスニペットに限定している．一方本研究は，WWW 上の日本語で書かれた 1 億文書すべてを対象に素性（動詞との共起名詞）を抽出している．さらに，本研究は Score_{base} のみでなく，Score_{trick} も組み合わせることで頑健性を向上させている点がこれらの研究と異なる．

以上で述べた，我々の提案する方向付き類似度尺度 Score の特長をまとめると次のようになる．

- 低頻度語を過大に評価する相互情報量を用いない．
- 類似度への貢献度が最大の名詞を無視することで，コーパスの偏りによって生じうる偶発的に高い（誤った）類似度を抑制することが期待できる．

3.2　テンプレート共起名詞データベースの構築

本研究の動詞含意知識獲得は一項テンプレート間の含意知識の獲得から始まる．一方テンプレート間の含意知識は，本節で述べるテンプレート共起名詞データベースから獲得される．テンプレート共起名詞データベースは，テンプレート $\langle p, v\rangle$ とその共起名詞 n，$\langle p, v\rangle$ と n の共起頻度 f から構成される三つ組 $\langle n, \langle p, v\rangle, f\rangle$ の集合である．我々は，日本語係り受け解析器 KNP[12] で係り受け解析済みの日本語ウェブコーパス[11]（WWW 上の日本語 1 億文書）を用いて，次の手順でテンプレート共起名詞データベースを構築した．

(1)　日本語ウェブコーパス中の動詞を KNP で定義されている代表表記に変換する（ただし代表表記中の読みの部分は削除する）．

(2)　日本語ウェブコーパスから，係り受け関係にある名詞（n），助詞（p），動詞（v）の三つ組 $\langle n, p, v\rangle$（たとえば $\langle$原告, が, 勝訴する$\rangle$）とその頻度 f を抽出する．

(3)　三つ組 $\langle n, p, v\rangle$ をテンプレート $\langle p, v\rangle$ と共起名詞 n のペア $\langle n, \langle p, v\rangle\rangle$ に変換し，$\langle n, \langle p, v\rangle, f\rangle$ を得る．

(4)　次のいずれかに該当する $\langle n, \langle p, v\rangle, f\rangle$ を除外する．
- (a)　テンプレートの日本語ウェブコーパスにおける頻度が α 未満のもの
- (b)　動詞が受け身形，使役形，否定形のもの
- (c)　「する」「なる」「できる」等，動詞の意味が漠然としているもの
- (d)　助詞が「は」「が」「を」「に」「で」以外のもの

$\alpha = 20$ として得られたテンプレート共起名詞データベースには 127,808 のテンプレートが含まれており，動詞の異なり数は 52,562 語となった．動詞として「母子感染する」のような複合語も含まれている．

3.3　動詞含意知識の獲得

動詞含意知識は次の手順で獲得される．

(1)　テンプレート共起名詞データベース（3.2 節）から，テンプレートペアとその Score（3.1 節）の値からなるリストを生成する．

(2)　テンプレートから助詞と変数を取り去り，動詞ペアのリストに変換する．

(3)　この結果，複数の異なるテンプレートペア（たとえば「X が アポトーシスする → Y が 死ぬ」と「X は アポトーシスする → Y が 死ぬ」）が助詞と変数を取り去ることで同一の動詞ペア（「アポトーシスする → 死ぬ」）になる場合があるので，重複した動詞ペアを削除する．具体的には，Score 値が最も高いものだけを残し，他をすべて削除する．

(4)　Score 値上位 N 位内にある動詞ペアを取得する．

つまり，はじめにテンプレート単位の含意知識を獲得し，その後動詞単位に変換する．

本研究は動詞含意知識の獲得を目的としているが，4 章ではテンプレート単位の含意知識獲得の精度についても報告し，提案手法の有効性を主張する．テンプレート単位の含意知識として，スコア計算格テンプレートを用いたものと，ガ格テンプレートを用いたものの 2 種類を用意した．前者は上記手順 1 の結果である．すべてのテンプレートペアは Score 値の計算で使用した助詞（「は」「が」「を」「に」「で」のいずれか）のペアを持っている．後者は次の手順で得られる．

(1)　スコア計算格テンプレートペアのリストから，助詞がどちらも「が」以外であるものを取得する（たとえば「X を ポワレする → Y を 焼く」）．

(2)　それらテンプレートペアの助詞をどちらも「が」に変換する（たとえば「A が ポワレする → B が 焼く」）．

助詞として「が」を選んだのは，ガ格はどの動詞も備えている格だからである．ガ格テンプレートペアを評価する目的は次の仮説の検証にある．本提案手法により動詞間の意味的な含意関係を適切にとらえられているのであれば，スコア計算格テンプレートペアの含意関係の正しさはもちろんのこと，スコア計算に利用しなかった（かつ，どの動詞も必ず備えている）ガ格のテンプレートペアの含意関係の正しさも保証されているはずである．つまり，ガ格テンプレートペアの評価を通して，本提案手法の有効性を，動詞単体のペアやスコア計算格テンプレートペアの評価とは違った角度から評価する．さらに，ガ格テンプレートペアが含意関係として妥当なものであれば，スコア計算に用いられなかった助詞（上記の例の場合は「が」）を一項テンプレート含意ペア（たとえば「X を ポワレする → Y を 焼く」）に追加することで，多項テンプレート含意ペア（たとえば「A が X を ポワレする → B が Y を 焼く」）も獲得することができるということになる．

なお，テンプレート含意ペアが適格であるには，変数（上記の A と B，X と Y）の指示対象がテンプレート間で同じでなければならない．

最後に，本章で述べた提案手法は，2 章で取り上げた先行研究の手法と同様，1 つの動詞が含意する動詞を複数個獲得できることに注意する．つまり，「就職する」が「働く」，「入社する」，「勤める」等を含意するように，ある動詞が含意する動詞は 1 つとは限らないが，本提案手法はそれら複数の動詞を獲得する．実際，本提案手法により，「就職する」が含意する動詞として（スコア上位 5 万以内に限定すると）「働く」，「入社する」，「勤める」，「勤務する」を獲得できた．ただし，先行研究の手法と同じく，ある動詞が含意する動詞をすべて獲得できるわけではない．実際，ある動詞が含意する動詞をすべて列挙するというのは人間にとっても困難なタスクである．たとえば「就職する」が含意する動詞は，我々が思いつく限りは，上にあげたもの以外に，「労働する」，「就労する」，「就業する」，「就任する」，「就職活動する」，「内定する」があるが，これですべて列挙できたかは定かではない．動詞含意獲得の網羅性の確保という課題は，一足飛びに解決できるようなものではないが，現在我々が構築を進めている大規模動詞含意知識ベースを通して，議論を深めることができるものと考えている．

4. 評 価 実 験

本章では，まず 4.1 節で提案手法の動詞含意知識獲得の精度を先行研究（Lin [14]，Precision [24]，BInc [21]）と比較し，提案手法の優位性を示す．4.2 節では提案手法のコンポーネントである Score_{trick} に実装されたトリックの有効性を明らかにする．4.3 節では，提案手法が動詞単位の含意知識だけでなく，テンプレート単位の含意知識獲得にも有用であることを示す．最後に 4.4 節で，動詞含意知識獲得結果のスコア上位 10 万ペアの精度を示すことで，提案手法の出力が，大規模な動詞含意知識データベースを整備するための出発点として有用であることを主張する．

提案手法の評価として，作業者 3 名（いずれも著者ではない）が提案手法により獲得された含意知識の正解判定を行った．本研究では日常的には使われない動詞も対象とするため，作業者の知らない動詞も現れる．その場合，作業者は辞書（紙媒体か電子媒体かを問わない）あるいは WWW 上の情報によりその意味を確認しながら評価を行った．それでもなお意味が不明な場合は，その動詞を含むペアを不正解として扱った．複数の語義を持つ動詞に

関しては，そのいずれかの語義について当該のペアが含意関係として適切であると判断できれば正解とした．

作業者ペアごとの Kappa 値の平均は，動詞単位の含意知識獲得の評価では 0.579，テンプレート単位の評価では 0.568 となった．この Kappa 値は，作業者間の判断にまずまずの安定性がある（揺れが少ない）ことを示している．

評価指標として次の式で定義される Accuracy を用いた．

$$Accuracy = \frac{\text{正解ペア数}}{\text{全獲得ペア数}} \tag{10}$$

Accuracy はさらに，作業者 1 名以上が正解と判定した場合に正解とする Accuracy-1，2 名以上の正解判定で正解とする Accuracy-2，3 名とも正解とした場合の Accuracy-3 に分かれる．

前章の最後で述べたとおり，獲得した動詞含意知識の網羅性も重要な評価の観点である．網羅性を測定するためには，ある動詞についてそれが含意する動詞をすべて網羅したデータをあらかじめ用意する必要があるが，現在のところ，そのようなデータは，我々の知る限り，どの言語においても存在しない．そのため本研究では，網羅性の評価は見送り，Accuracy の観点でのみ評価することとした．

4.1　実験 1：動詞含意知識獲得の精度

評価対象のいずれの手法も 3.2 節で述べたテンプレート共起頻度データベースが出発点となる．本実験では，$\alpha = 200$ とした場合のテンプレート共起頻度データベース（V_{200}）と $\alpha = 20$ とした場合のテンプレート共起頻度データベース（V_{20}）の 2 種類を用いて動詞含意知識獲得を行った（α はテンプレート頻度．3.2 節を参照）．

評価対象は，獲得手法ごとに，スコア上位 20,000 の動詞ペアの中からサンプリングした 200 ペアである（つまり式 (10) の分母が 200）．公正を期すため，手法ごとの評価サンプルはすべて 1 つにまとめられ，シャッフルされたうえで作業者に提示された．つまり，どの評価サンプルがどの手法で得られたものか作業者に分からないようにした．

Lin は方向性のない類似度尺度のため（「$v_1 \rightarrow v_2$」か「$v_1 \leftarrow v_2$」か分からないため），次の手順により含意の方向を決めた．

(1)　Lin の手法で得られた動詞ペアのサンプル 200 ペアをコピーし，v_1 と v_2 を入れ替える．

(2)　「$v_1 \rightarrow v_2$」方向と「$v_1 \leftarrow v_2$」方向の Lin ペア計 400 をシャッフルし，そのすべての正解判定を行う．

表 1　V_{200} による動詞含意知識獲得結果
Table 1　Accuracy figures of verb entailment acquisition with V_{200}.

	Acc-1	Acc-2	Acc-3
Score	**0.690**	**0.520**	**0.335**
BInc	0.455	0.295	0.160
Precision	0.450	0.355	0.205
Lin	0.635	0.385	0.205

表 2　V_{20} による動詞含意知識獲得結果
Table 2　Accuracy figures of verb entailment acquisition with V_{20}.

	Acc-1	Acc-2	Acc-3
Score	**0.770** (+0.080)	**0.660** (+0.140)	**0.460** (+0.125)
BInc	0.450 (−0.005)	0.255 (−0.040)	0.125 (−0.035)
Precision	0.725 (+0.275)	0.545 (+0.190)	0.385 (+0.180)
Lin	0.590 (−0.045)	0.370 (−0.015)	0.160 (−0.045)

Lin で獲得された元の 200 ペアはいずれかの方向で正しい含意ペアと判定されれば正解と見なされる．これは LEDIR [4] の性能の上限を評価したことになる．

表 1，**表 2** にテンプレート共起頻度データベース V_{200} と V_{20} を用いた評価結果を示す．表 2 の括弧中の数値は，表 1 と表 2 の精度の差を表す．V_{200}，V_{20} いずれの場合も提案手法が最も良い精度を示している．また，V_{200} から V_{20} へと動詞セットを変更した結果，つまりより多くの低頻度動詞を対象とした結果，Lin と BInc が精度を下げたのに対し，提案手法と Precision は精度が向上した．情報爆発時代を迎えた現在，自然言語処理アプリケーションの高精度化の鍵は，ロングテールを構成する低頻度語への頑健性にあると我々は考えている．実験結果から，提案手法と Precision はこの性質を備えているといえる．さらに，提案手法は Precision に比べて，V_{20} の Accuracy-2 の場合で 0.115，V_{200} の Accuracy-2 の場合で 0.165 の差で高精度であることが分かった．

なお，V_{20} で動詞ペアをランダムに 200 ペア構成した場合の精度は，Accuracy-1 で 0.035，Accuracy-2 で 0.005，Accuracy-3 で 0.000 だった．

図 1 と**図 2** に，V_{200} と V_{20} を用いた場合の手法ごとのスコア上位 N 位の精度をあげる．この精度は，N を 1,000，2,000，3,000，と 1,000 ごとに区切ったうえで，200 サンプルの

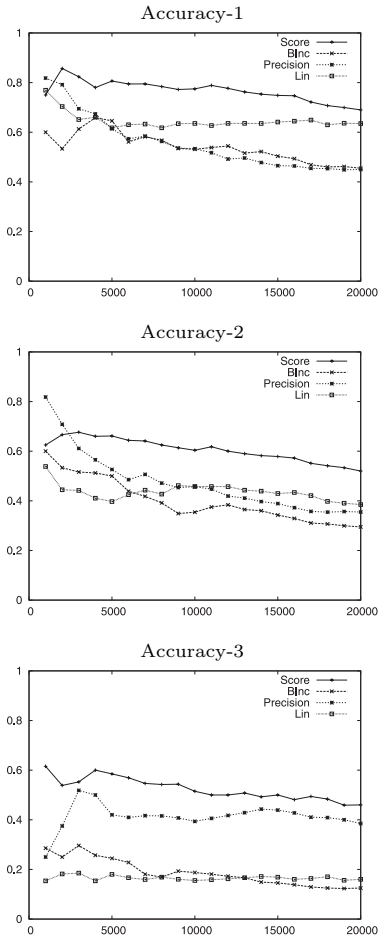


図 1　V_{200} による動詞含意知識獲得のスコア上位 N 位の精度
Fig. 1　N-best accuracy figures of verb entailment acquisition with V_{200}.

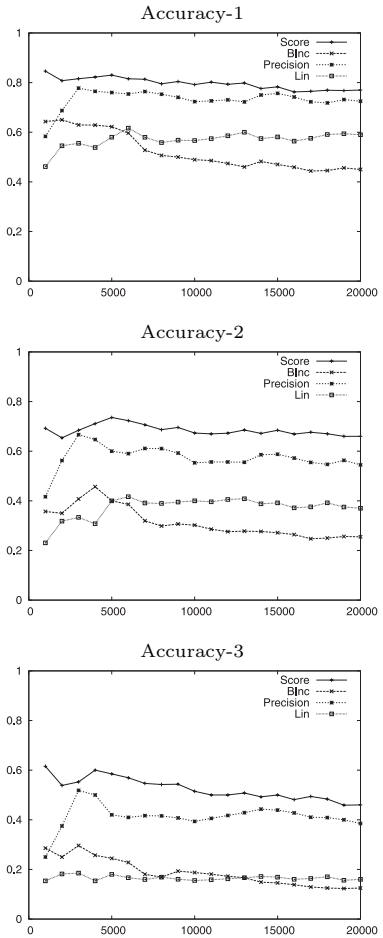


図 2　V_{20} による動詞含意知識獲得のスコア上位 N 位の精度
Fig. 2　N-best accuracy figures of verb entailment acquisition with V_{200}.

うち N 位以内にあるサンプルのみで測定したものである．本稿の精度のグラフはすべてこの方法によりプロットした．図 2 によると，Accuracy-1，2，3 のどの場合も，スコアトップから上位 20,000 位にかけて，提案手法が他の手法より高い精度を示している．

提案手法で得られた動詞含意知識を観察すると，例 [1] にあるような，専門用語的な動詞とそれを平易に説明するような動詞のペアが多く見られた．

[1]　a.「RSS 配信する → 届ける」

表 3　Score，Score_{trick}，Score_{base} の精度
Table 3　Accuracy figures of Score, Score_{trick}, and Score_{base}.

	Acc-1	Acc-2	Acc-3
Score	**0.770**	**0.660**	**0.460**
Score_{trick}	0.725	0.610	0.395
Score_{base}	0.590	0.465	0.315

b.「ミッドシップマウントする → 積む」

一方，精度の低かった Lin と BInc の出力（それぞれ例 [2a]，[2b]）を観察すると，誤りの多くのは専門用語等の低頻度動詞からなるペアだった．

[2]　a.「クラッキングする → 構築保守する」

b.「水槽飼育する → 試験放流する」

これらの動詞ペアはいずれも，何らかの関連性はあるが含意関係にあるとはいえないものである．

4.2　実験 2：Score_{trick} の有効性

表 3 に Score 全体，Score_{trick} のみ，Score_{base} のみの精度を，**図 3** にそれぞれのスコア上位 N 位の精度の変化をあげる．テンプレート共起頻度データベースとして V_{20} を用いた．評価対象は実験 1 と同様，スコア上位 20,000 位内の動詞ペアからサンプリングした 200 ペアである．この結果から，Score_{base} よりも Score_{trick} が，Score_{trick} 単体よりも Score_{base} と Score_{trick} 両方を掛け合わせた Score が高い精度を示しているのが分かる．

Score_{base} により誤って獲得された動詞ペアの 1 つに 3.1 節で例としてあげた「代行入手する → 引用する」がある．この動詞ペアの間で共有されている共起名詞は「作品」と「メーカーカタログ等」の 2 つしかなく，Score_{base} の値（類似度）の 99.99%以上は「メーカーカタログ等」によってもたらされたものだった．我々が提案するトリックは，ただ 1 つの共有名詞により（偶発的に）高い類似度がもたらされることを防ぐ仕組みであり，これを導入することで上記のような誤った動詞ペアを下位に押し下げることができる．

4.3　実験 3：テンプレート含意獲得の精度

本節では，実験 1 で良好な結果を示した提案手法と Precision の 2 つの手法を取り上げ，テンプレート単位の含意獲得の精度を示す．テンプレート共起頻度データベースは V_{20} である．

評価実験では，3.3 節で述べた手順により獲得したスコア計算格テンプレートペアとガ格テンプレートペアを作業者が正解判定した．作業者には，テンプレートペア間の対応する格

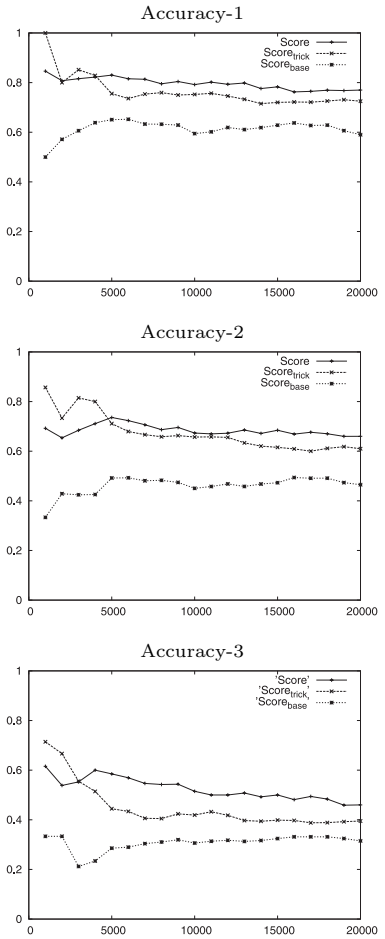


図 3　Score，Score_{trick}，Score_{base} による動詞含意知識獲得のスコア上位 N 位の精度
Fig. 3　N-best accuracy figures of verb entailment acquisition by $Score$, $Score_{trick}$ and $Score_{base}$.

には同じ名詞が入り，かつ，どの名詞が当該格に出現しても含意関係が成り立つもののみを正解とするよう指示した．つまり，動詞含意の場合と異なり，変数に入る名詞に関する制限を正解条件に追加した．評価対象は先の実験と同様，スコア上位 20,000 位内のテンプレー

表 4　テンプレート含意獲得結果の精度
Table 4　Accuracy figures of case frame entailment acquisition.

	Method	Acc-1	Acc-2	Acc-3
スコア計算格テンプレート	Score	**0.655** (−0.115)	**0.510** (−0.150)	**0.300** (−0.160)
	Precision	0.565 (−0.160)	0.430 (−0.115)	0.265 (−0.120)
ガ格テンプレート	Score	**0.665** (−0.105)	**0.515** (−0.145)	**0.315** (−0.145)
	Precision	0.490 (−0.235)	0.325 (−0.220)	0.215 (−0.170)

トペアからサンプリングした 200 ペアである．

表 4 に評価結果をあげる．**図 4** と**図 5** にスコア計算格テンプレートとガ格テンプレートを用いた場合のスコア上位 N 位の精度の変化をあげる．表中の括弧内の数字は動詞含意知識獲得の精度（実験 1 の表 2）との差を表す．この結果から，テンプレート単位の含意獲得においても，提案手法が Precision の精度を上回ることが分かる．提案手法と Precision のいずれも，動詞含意知識獲得の場合と比べて 10%ほど精度が低下した．これは，変数に入る名詞に関する上述の制限が追加されたためと考えられる．

獲得されたスコア計算格テンプレートペアとガ格テンプレートペアの例（正解）をそれぞれ例 [3] と [4] にあげる．

[3]　a.「X を 立ち食いする → X を 食べる」
　　b.「X で マリネードする → X を 入れる」
[4]　a.「X が NBA 入りする → X が 入団する」
　　b.「X が 製粉する → X が 挽く」

4.4　実験 4：動詞含意知識獲得上位 100,000 の精度

提案手法と Precision による動詞含意知識獲得結果のうち，スコア上位 100,000 位内の動詞ペアからサンプリングした 200 ペアを評価した．テンプレート共起頻度データベースとして V_{20} を用いた．**表 5** にあるとおり，スコア上位 100,000 ペアにおいても，提案手法の精度が Precision を上回った．本研究における動詞含意知識獲得結果は，人手によるアノテーションを経たうえで研究コミュニティで広く共有されることを想定している．この実験結果において注目すべきは，提案手法がスコア上位 100,000 位に至るまで 48%（Accuracy-2）という，人手によるアノテーションの出発点としてはリーズナブルな精度を保っている点である．48%という精度は，スコア上位 100,000 位とかなり下位まで評価対象としているこ

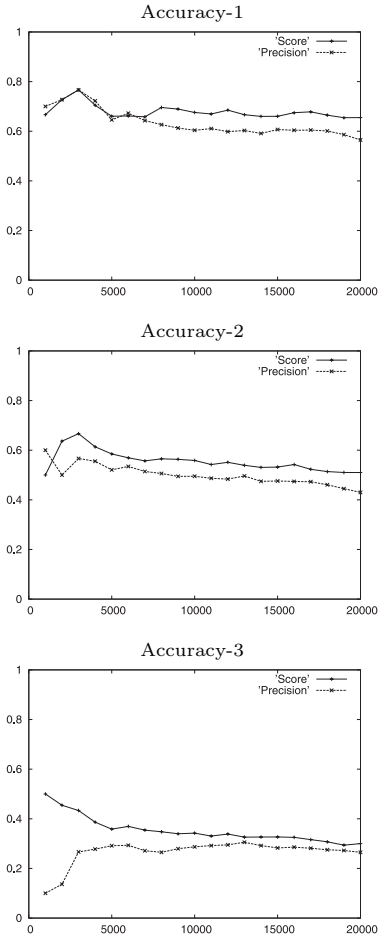


図 4　スコア計算格テンプレートを用いた場合の Score と Precision によるテンプレート含意獲得のスコア上位 N 位の精度
Fig. 4　N-best accuracy figures of template entailment acquisition by Score and Precision using scoring case slots.

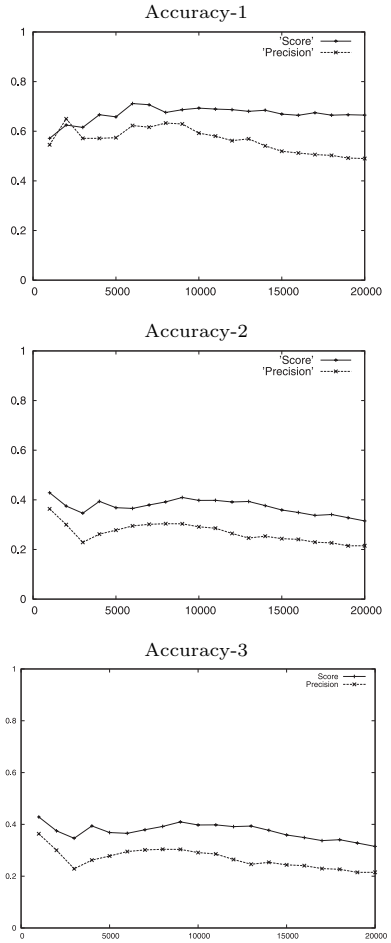


図 5　ガ格テンプレートを用いた場合の Score と Precision によるテンプレート含意獲得のスコア上位 N 位の精度
Fig. 5　N-best accuracy figures of template entailment acquisition by Score and Precision using nominative case slots.

表 5　スコア上位 100,000 の動詞含意の精度
Table 5　Accuracy figures of the top 100,000 verb entailment pairs.

	Acc-1	Acc-2	Acc-3
Score	**0.610**	**0.480**	**0.300**
Precision	0.470	0.295	0.190

とを考慮しても，けっして高いものではないが，作業者に予備的なアノテーションをしてもらったうえで作業の負担について質問したところ，十分に許容範囲内であるとの回答が得られたため，この精度をリーズナブルなものと判断した．

図 6 にスコア上位 100,000 以内における精度変化をあげる．スコア上位 20,000 位前後までは，提案手法の精度が Precision の精度を下回っているが，これはスコア上位 20,000 位までのサンプル数が約 40 サンプルと少ないためであると考えられる．実際，実験 1 では，スコア上位 20,000 位を 200 サンプルで評価した結果，提案手法の精度が Precision の精度を上回るという結果が得られている．

例 [5] にスコア上位 100,000 位以内にあった正解動詞含意ペアの例を，その順位，正解と判定した作業者の人数ともにあげる．

[5] a.「裁定する → 定める$^{\star 1}$」 ..... 6,081 位（3 人）
b.「混ぜる → 入れる」 ..... 8,272 位（2 人）
c.「テレビ中継する → 観る$^{\star 2}$」 ..... 19,515 位（3 人）
d.「増員配置する → 配置する」 ..... 24,926 位（3 人）
e.「ブロックする → 防ぐ」 ..... 38,140 位（3 人）
f.「準加盟する → 所属する」 ..... 39,478 位（3 人）
g.「全面改正する → 施行する」 ..... 44,455 位（3 人）
h.「ブログチェックする → 観る」 ..... 53,113 位（3 人）
i.「参加出品する → 出展する」 ..... 54,366 位（3 人）
j.「離任する → 就任する」 ..... 57,653 位（2 人）
k.「ジンギスカンする → 焼く」 ..... 59,771 位（2 人）
l.「集中連載する → 載る」 ..... 62,745 位（3 人）
m.「属性変更する → 設定する」 ..... 66,126 位（2 人）

⋆1 この「定める」は「決定する」の意味で解釈された．
⋆2 視聴者，あるいはテレビ局関係者が観るという意味で正解と判定された．

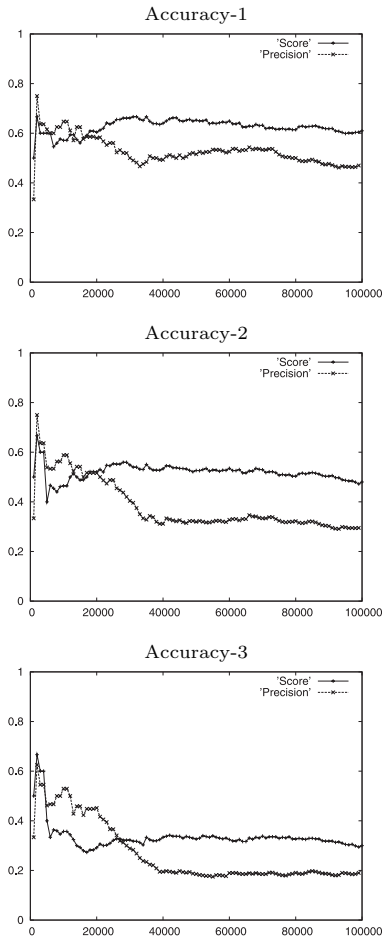


図 6　動詞含意知識獲得のスコア上位 N 位の精度（スコア上位 100,000 位内が対象）
Fig. 6　N-best accuracy figures of verb entailment acquisition (until Top 100,000).

n.「主演する → 出演する」 ..... 75,533 位（3 人）
o.「買置きする → 購入する」 ..... 94,526 位（2 人）
p.「用心する → 注意する」 ..... 99,399 位（3 人）

例 [6] は，同じくスコア上位 100,000 位以内にあった不正解の動詞ペアと，その順位，不正解と判定した作業者の人数である．

[6] a.「身柄送検する → 現行犯逮捕する」 ..... 4,389 位（2 人）
b.「デイトレする → 儲ける」 ..... 6,752 位（3 人）
c.「返金保証する → 返金する」 ..... 7,595 位（2 人）
d.「買取査定する → 稼ぐ」 ..... 29,614 位（2 人）
e.「呟く → 書く」 ..... 38,790 位（3 人）
f.「モニタープレゼントする → 発売する」 ..... 46,111 位（2 人）
g.「分解脱臭する → 吸着する」 ..... 51,696 位（3 人）
h.「事前公表する → 掲載する」 ..... 60,741 位（3 人）

これらはすべて，2 つの動詞の表す事態の間に何らかの関連はあるが，含意関係にあるとはいえないものである．このほかにも，上位 100,000 位内には例 [7] のような，矛盾関係にある動詞ペアがあった．

(7)「欠場する → 出場する」 ..... 40,504 位

提案手法も含め本研究で使用した手法はすべて分布類似度に基づくものであるため，共起名詞を共有する動詞ペアであれば，矛盾関係にあるものにも，何らかの関連性が認められるだけのものにも含意スコアが割り当てられる．分布類似度の高い動詞ペアの中から含意ペアとそれ以外を区別するモデルの開発は今後の課題である．

5. 結　　論

本研究は WWW からの大規模な動詞含意知識の獲得に取り組み，WWW 文書の偏りや低頻度動詞に対して頑健な方向付き分布類似度尺度を提案した．また，実験により次の点を明らかにした．

(1)　我々が提案する方向付き分布類似度尺度 Score は，WWW からの大規模な動詞含意知識獲得において，これまでに提案された分布類似度尺度である Lin，Precision，BInc より高精度である．
(2)　我々が開発し Score_{trick} として実装したトリックは WWW からの大規模な動詞含意知識獲得の精度を大きく向上させる．
(3)　Score はテンプレート単位の含意知識獲得にも有効である．
(4)　Score で得られた動詞含意知識はスコア上位 100,000 位に至るまで（人手アノテーションの出発点として）リーズナブルな精度を保つ．

提案手法は，これまであげた獲得例にあるとおり，WWW という広大な情報の海から，日常的に使用される動詞の含意知識から特定の専門分野の動詞の含意知識まで幅広く獲得することができる．

我々は現在，本研究で比較した 4 手法（Score，Precision，Lin，BInc）により獲得した動詞含意知識を人手でチェックすることで，動詞含意知識データベースを構築している．これまでに約 30,000 ペアの正例（含意が成立している動詞ペア）と約 38,000 ペアの負例（含意が成立していない動詞ペア）を整備しており，そのデータは「ALAGIN 動詞含意関係データベース」という名称で高度言語情報融合フォーラム ALAGIN [⋆1]から配布されている．現在のバージョンでは，「離婚する → 結婚する」や「離任する → 就任する」のような時間的にずれのある前提条件関係といえるペアや，「売る → 買う」，「勝つ → 負ける」のように視点あるいは動作主体が異なる同一事態のペアは含んでいない．今後このデータベースは，上記のような未収録の種類のペアを他と区別したうえで収録する等，さらに継続的に規模を拡大していく予定である．

参 考 文 献

1) Ando, M., Sekine, S. and Ishizaki, S.: Automatic Extraction of Hyponyms from Japanese Newspaper Using Lexico-syntactic Patterns, *Proc. LREC '04* (2004).
2) Bannard, C. and Callison-Burch, C.: Paraphrasing with Bilingual Parallel Corpora, *Proc. 43rd Annual Meeting of the Association for Computational Linguistics* (*ACL2005*), pp.597–604 (2005).
3) Barzilay, R. and Lee, L.: Learning to Paraphrase: An Unsupervised Approach using Multiple-Sequence Alignment, *Proc. HLT-NAACL 2003*, pp.16–23 (2003).
4) Bhagat, R., Pantel, P. and Hovy, E.: LEDIR: An Unsupervised Algorithm for Learning Directionality of Inference Rules, *Proc.Conference on Empirical Methods in Natural Language Processing* (*EMNLP2007*), pp.161–170 (2007).
5) Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D.S. and Yates, A.: Unsupervised named-entity extraction from the web: an experimental study, *Artificial Intelligence*, Vol.165, No.1, pp.91–134 (2005).
6) Fujita, A. and Sato, S.: A Probabilistic Model for Measuring Grammaticality and Similarity of Automatically Generated Paraphrases of Predicate Phrases, *Proc. 22nd International Conference on Computational Linguistics* (*COLING2008*), pp.225–232 (2008).
7) Geffet, M. and Dagan, I.: The Distributional Inclusion Hypotheses and Lexical Entailment, *Proc. 43rd Annual Meeting of the Association for Computational Linguistics* (*ACL2005*), pp.107–114 (2005).
8) Harris, Z.: Distributional Structure, *Word*, Vol.10, No.2–3, pp.146–162 (1954).
9) Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora, *Proc. 14th conference on Computational linguistics*, pp.539–545 (1992).
10) Ibrahim, A., Katz, B. and Lin, J.: Extracting Structural Paraphrases from Aligned Monolingual Corpora, *Proc. 2nd International Workshop on Paraphrasing* (*IWP2003*), pp.57–64 (2003).
11) Kawahara, D. and Kurohashi, S.: Case Frame Compilation from the Web using High-Performance Computing, *Proc. 5th International Conference on Language Resources and Evaluation* (*LREC-06*), pp.1344–1347 (2006).
12) Kawahara, D. and Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis, *Proc. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (*HLT-NAACL2006*), pp.176–183 (2006).
13) Lin, D.: Automatic Retrieval and Clustering of Similar Words, *Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* (*COLING-ACL1998*), pp.768–774 (1998).
14) Lin, D. and Pantel, P.: Discovery of Inference Rules for Question Answering, *Natural Language Engineering*, Vol.7, No.4, pp.343–360 (2001).
15) Oh, J.-H., Uchimoto, K. and Torisawa, K.: Bilingual Co-Training for Monolingual Hyponymy-Relation Acquisition, *Proc. 47th Annual Meeting of the Association for Computational Linguistics* (*ACL-IJCNLP-09*), pp.432–440 (2009).
16) Pantel, P. and Ravichandran, D.: Automatically Labeling Semantic Classes, *Proc. Human Language Technology and North American Chapter of the Association for Computational Linguistics Conference* (*HLT/NAACL-2004*), pp.321–328 (2004).
17) Pekar, V.: Acquisition of verb entailment from text, *Proc. main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics* (*HLT-NAACL2006*), pp.49–56 (2006).
18) Shinyama, Y., Sekine, S. and Sudo, K.: Automatic Paraphrase Acquisition from News Articles, *Proc. 2nd international Conference on Human Language Technology Research* (*HLT2002*), pp.313–318 (2002).
19) Shinzato, K. and Torisawa, K.: Extracting hyponyms of prespecified hypernyms from itemizations and headings in web documents, *20th International Conference on Computational Linguistics* (*COLING-2004*), pp.938–944 (2004).
20) Sumida, A., Yoshinaga, N. and Torisawa, K.: Boosting Precision and Recall of Hyponymy Relation Acquisition from Hierarchical Layouts in Wikipedia, *Proc.*

⋆1 http://www.alagin.jp/

6th International Conference on Language Resources and Evaluation (*LREC-2008*) (2008).
21) Szpektor, I. and Dagan, I.: Learning Entailment Rules for Unary Template, *Proc. 22nd International Conference on Computational Linguistics* (*COLING2008*), pp.849–856 (2008).
22) Torisawa, K.: Automatic Acquisition of Expressions Representing Preparation and Utilization of an Object, *Proc. Recent Advances in Natural Language Processing* (*RANLP'05*), pp.556–560 (2005).
23) Torisawa, K.: Acquiring Inference Rules with Temporal Constraints by Using Japanese Coodinated Sentences and Noun-Verb Co-occurences, *Proc. Human Language Technology Conference of the Norh American Chapter of the ACL* (*HLT-NAACL2006*), pp.57–64 (2006).
24) Weeds, J. and Weir, D.: A General Framework for Distributional Similarity, *Proc. Conference on Empirical Methods in Natural Language Processing* (*EMNLP2003*), pp.81–88 (2003).
25) Yamada, I., Torisawa, K., Kazama, J., Kuroda, K., Murata, M., De Saeger, S., Bond, F. and Sumida, A.: Hypernym discovery based on distributional similarity and hierarchical structures, *EMNLP '09: Proc. 2009 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2009*), Association for Computational Linguistics, pp.929–937 (2009).
26) Zanzotto, F.M., Pennacchiotti, M. and Pazienza, M.T.: Discovering asymmetric entailment relations between verbs using selectional preferences, *Proc. 44th Annual Meeting of the Association for Computational Linguistics and 21th International Conference on Computational Linguistics* (*COLING-ACL2006*), pp.849–856 (2006).
27) 久光　徹，丹羽芳樹：統計量とルールを組み合わせて有用な括弧表現を抽出する手法，情報処理学会自然言語処理研究会資料 NL-122，pp.113–118 (1997).

(平成 22 年 7 月 29 日受付)
(平成 22 年 10 月 4 日採録)

橋本　　力（正会員）

1999 年福島大学教育学部卒業．2001 年北陸先端科学技術大学院大学博士前期課程修了．2005 年神戸松蔭女子学院大学大学院博士後期課程修了．京都大学情報学研究科産学官連携研究員を経て，2007 年山形大学大学院理工学研究科助教，2009 年より独立行政法人情報通信研究機構専攻研究員．現在に至る．自然言語処理の研究に従事．博士（言語科学）．言語処理学会会員．

鳥澤健太郎（正会員）

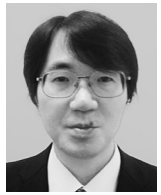
東京大学大学院理学系研究科博士課程専攻中退後，同研究科助手．その後，科学技術振興事業団さきがけ研究 21 研究員（兼任），北陸先端科学技術大学院大学助教授，准教授を経て，2008 年 NICT に，MASTAR プロジェクト言語基盤グループ，グループリーダーとして着任．自然言語処理，特に Web 上の言語処理，Web からの知識獲得，獲得された知識の活用方法の研究に従事．けいはんな連携大学院教授を兼務．博士（理学）．

黒田　　航

元独立行政法人情報通信研究機構知識創成コミュニケーション研究センター MASTAR プロジェクト言語基盤グループ短時間研究員．現在，京都工芸繊維大学，早稲田大学総合研究機構．京都大学から人間・環境学博士を取得．言語学と自然言語処理を融合する研究に従事．

デサーガ　ステイン

2006 年北陸先端科学技術大学院大学知識科学研究科博士課程修了．博士（知識科学）．北陸先端科学技術大学院大学研究員を経て，2007 年に情報通信研究機構に入所．2008 年に NICT MASTAR プロジェクト言語基盤グループに専攻研究員として着任．自然言語処理を用いた知識獲得の研究に従事．



村田　真樹（正会員）

1970 年生．1993 年京都大学工学部電気工学第二学科卒業．1997 年同大学大学院工学研究科電子通信工学専攻博士課程修了．博士（工学）．同年京都大学にて日本学術振興会リサーチ・アソシエイト．1998 年郵政省通信総合研究所入所．独立行政法人情報通信研究機構主任研究員を経て，現在，鳥取大学大学院工学研究科情報エレクトロニクス専攻教授．自然言語処理，情報抽出の研究に従事．2005 年 FIT2005 論文賞受賞．共著書に『事例で学ぶテキストマイニング』（共立出版）等がある．言語処理学会，人工知能学会，電子情報通信学会，計量国語学会，ACL 等の各会員．

風間　淳一（正会員）

独立行政法人情報通信研究機構知識創成コミュニケーション研究センター MASTAR プロジェクト言語基盤グループ主任研究員．2004 年東京大学大学院情報理工学系研究科コンピュータ科学専攻博士課程修了．博士（情報理工学）．同年北陸先端科学技術大学院大学情報科学研究科助教．2008 年より現職．自然言語処理の研究に従事．