

共参照解析における事象間関係知識の適用

杉浦 純 井之上 直也 乾 健太郎

東北大学大学院 情報科学研究科

{jun-s, naoya-i, inui}@ecei.tohoku.ac.jp

1 はじめに

語彙知識は、自然言語を理解する上で重要な要素である。近年では知識獲得技術の発展により、物事同義関係や上位下位関係、事象間の因果関係といった様々な語彙知識が大規模に獲得され、多くの研究者がその成果を利用できる環境ができてきた [5, 13, 1]。

実際に、名詞に関する知識（同義関係、上位下位関係など）については、いくつかの自然言語処理タスクにおいて利用され、ある一定の成果が上がっている。しかしながら、事象間関係知識については、その有用性を十分に示す成果は報告されておらず、その原因や解決策については深く議論されていないのが現状である [12, 7, 14]。

獲得した知識が言語解析の性能向上に寄与しない原因は、少なくとも二つあると考えられる。第一に、**知識の規模の問題**がある。言語解析に有用と考えられる知識を獲得できても、言語解析の一部の問題にしか適用できなければ性能向上は見込めない。第二に、**知識の使い方の問題**がある。有用な知識が大規模にあっても、それらを適用対象の文脈に合わせて使いこなせる枠組みがないと、性能向上どころか性能低下を招く可能性もある。

そこで本稿では、談話処理に事象間関係知識を適用する際の問題点を知識の使い方の観点から分析・整理する。本分析では、知識の規模の問題を排除するために、約7億文規模のコーパスからスクリプト知識（典型的に同時に起こる事象の集合に関する知識）を獲得し、分析に用いた。従来の、知識に基づく談話処理の研究において、これ程の大きな事象間関係知識を用いて課題の整理を行った研究はない。

課題を洗い出すための試金石としては、Winograd Schema Challenge (WSC) [8] を用いる。WSCは、*twin sentences* と呼ばれる、よく似た2文から構成される代名詞照応解析の問題集である。この問題集の特徴は、(1) 人間には簡単に解けるが、(2) 常識的知識を手がかりにしなければ解けない、といった、従来の統語的素性・意味的素性（述語の選択選好性など）の手がかりでは解析が困難な事例を集めた点である。例えば、下記のような問題が含まれている。

- (1) *John_i fired Dave_j because he_j had too many absences. (he=Dave)*
- (2) *John_i fired Dave_j because he_i had too many employees. (he=John)*

これら2文は、同一の文構造であり、違いは *absences* か、*employees* のみである。he には *John* も *Dave* も入りうるため、統語的な特徴および意味的な特徴は役に立たな

い。ここで重要な手がかりとなるのは、「fire する側は従業員を抱える立場」「欠勤が多い人は fire される」といった、事象間の関係知識である。知識の有効性のテストベッドとしては、古くから用いられてきた Recognizing Textual Entailment (RTE) [4] があるが、RTE は複合的な言語現象を解析する必要があるため、システムのどこに問題があるのか分かりにくい、という問題点があった。一方で WSC は、シンプルな1文から構成される問題集であり、知識の使い方の問題の課題を抽出するのに適したテストベッドと考えられる。

本稿では、まず大規模語彙知識の獲得と適用に関する既存研究と研究背景を述べる (2 節)。次に、大規模語彙知識を有効に使いこなすためのテストベッドとして、代名詞照応解析の追実験を行った結果を示す (3 節)。最後に、追実験における解析誤り事例を手で分析し、大規模語彙知識が有効に適用できていない様々な原因を事例とともに整理し、さらにそれらの解決方針を述べる (4 節)。

2 関連研究

2.1 談話処理への大規模語彙知識の適用

冒頭で述べたように、近年、大規模なウェブコーパスから推論規則を獲得し、言語解析に適用する研究が盛んに行われている。従来研究における推論規則の表現方法は、主に (1) 述語間関係 (e.g., *purchase ~ acquire*) [3, 6]、(2) 変項付き述語間関係 (e.g., *purchase_X ~ acquire_X*) [9, 2]、(3) タイプ付けされた変項付き述語間関係 (e.g., *purchase_X ~ acquire_X*, $X = COMPANY$, $COMPANY$ は X に入りうる名詞のクラス) [11, 1, 10]、の大きく3種類に分類できる (以後、これらを **推論規則** と呼ぶ。尚、本稿では、述語間の関連を示す記号として “~” を用いる)。

これまでに、談話解析に向けて語彙知識の活用を試みている既存研究がいくつか存在する [12, 7, 16, 14]。文献 [12, 7] では、変項付き述語間関係知識 (e.g. *purchase_X ~ acquire_X*) を用いて代名詞照応の解析を試みている。Rahman ら [12] は Winograd Schema Challenge に対し変項付き述語間関係知識を素性に用いた照応解析結果を報告しているが、十分な性能向上は見られなかった。

文献 [14] では、既存の語彙資源 [5, 13] から得た因果関係知識を用いた仮説推論により、文間に存在する述語間の関係を捉えることで談話関係解析を試みており、人手での小規模な検証から解析性能向上の可能性が報告されている。しかし、大規模な知識を用いた評価実験は行われておらず、知識を有効に扱うことができていないかどうかの検証も行われていない。

表 1: 推論規則を用いた照応解析追試結果

正解知識 / 対抗知識	Correct	Wrong
有り / 有り	35.0% (190/543)	24.0% (130/543)
有り / 無し	12.3% (67/543)	-
無し / 有り	-	12.0% (65/543)
無し / 無し	-	16.8% (91/543)
合計	47.3% (257/543)	52.7% (286/543)

そこで、本研究では、Rahman らの代名詞照応解析のアイデアにもとづき、規模を拡大した推論規則を用いて代名詞照応解析の実験を行い、解析誤りの分析を通し、推論規則が有効に扱えていない原因を整理しまとめる。また、適切な推論規則を適切に適用するための解決策を提示する。さらに、推論規則の適用以外の要素技術が解析に必要な問題に対する分析を行い、必要となる要素技術の洗い出しを行った結果を示す。

3 大規模な推論規則を用いた照応解析実験

3.1 実験方法

共有項を持つ述語対の獲得手法としては、文献 [2] の手法を拡張し、推論規則を事例レベルで獲得した（紙面の都合上、詳細は文献 [15] を参照されたい）。獲得の際は、共有項部分を変項化せず、述語間の項共有の事例をそのまま獲得した (e.g., *kill people ~ people die*)。知識獲得時のコーパスとしては、ClueWeb2012¹ の一部（約 2.2 億文書、6.9 億文）を用いた。また、文献 [2] の知識獲得手法では共参照解析器が必要となるため、解析器として Stanford CoreNLP² を用いた³。獲得した推論規則の総数は、約 2.3 億事例であり、変項付き述語間関係にして 15,065,290 であった。これは、Rahman らが用いた事象間関係知識 28,436 個に比べ、約 500 倍の規模である。

獲得した推論規則を用いる際は、代名詞と依存関係にある述語と、二つの先行詞候補のそれぞれと依存関係にある述語のペアから推論事例の集合を取得し、各先行詞候補への重み付き投票を行い、最尤先行詞を決定する。例えば、次の例を考えてみよう：

(3) *John_i fired Dave_j because he_j had too many absences.* (he=Dave)

例 (3) では、*X fire ~ X have* から取得した推論事例の集合と、*fire X ~ X have* から取得した推論事例の集合の和集合をスコア順にソートし、上位 k 個に含まれる推論事例の集合が、*John* と *Dave* のどちらをより多く支持するかを計算する。スコアの算出方法の詳細については、紙面の都合上文献 [15] を参照されたい。

3.2 実験結果

Winograd Schema Challenge のデータセット 564 問に対し、今回構築した推論規則を用いた照応解析結果を表 1 に示した⁴。1 列目は、入力文において、解析対象の代名詞を項とする述語と、正解の照応先である先行詞と項を共有する述語との関係知識（正解知識）の有無、および正解の照応先でない先行詞と項を共有する述語との関係知識（対抗知識）の有無の組み合わせを表し、それ

¹<http://lemurproject.org/clueweb12/>

²<http://nlp.stanford.edu/software/corenlp.shtml>

³推論規則抽出時には、同一文および隣接文の共参照解析結果のみを利用した。

⁴スコア計算が 30 分で終わらなかった事例 21 件は解析結果から除外した

表 2: 推論規則を用いた照応解析誤り分析結果

原因		対策	数
周辺文脈が考慮できていない (4.1 節)	(1.1) コピュラ	文脈スコア	58
	(1.2) 多義性(動詞句, 動詞の自他)	文脈スコア	18
	(1.3) 助動詞・否定表現	文脈スコア	50
	(1.4) 目的語や前置詞を伴うと因果関係にある述語の分布が変化する	文脈スコア	96
	(1.5) 談話関係	ルール	15
知識規模 (4.2 節)		同義・言い換え・単純推論	46
別の知識あるいは解析技術 (4.3 節)		選択嗜好	36
		極性	22
		世界知識	7
		高度推論	20
依存構造解析誤り (4.4 節)		-	49

ぞれにおける照応解析の正解数と誤り数およびその割合を示している。

正解知識と対抗知識のどちらも存在する問題文の中での正解率は 59.4% (190/320) であり、現段階でも、チャンスレート (50.0%) よりは高いことがわかった。なお、正解の先行詞候補との照応関係を支持する推論規則の数は平均 14,673 個、不正解の先行詞候補との照応関係を指示する推論規則の数は平均 7,691 個であった。

4 解析誤りの分析と課題の整理

前節では大規模な変項付き述語間関係知識を用いた実験結果を示した。しかし、これほど大規模な推論規則を適用した既存研究は無く、解析誤りの分析もこれまで詳しく行われてこなかった。本節では、実験結果に対する解析誤りを分析し、解析誤りの原因を洗い出し、推論規則適用時の課題を明らかにし、結果を報告する。分析は本実験において解析を誤った事例 286 個のうち 200 個に対し、なぜ現在の推論規則で解けていないのか、その原因を人手で洗い出し、さらにどのような要素を認識することができれば照応解析が可能であるのか、その解決策を人手で列挙し、整理を行った。

表 2 に照応解析誤りの原因、考えられる対策、誤り数を示した。このとき、一つの事例に対し誤りの原因は一つとは限らない。本実験で解けていない原因は大きく次の 4 つに分けられる⁵：(1) 周辺文脈が考慮できていない、(2) 推論規則が足りていない、(3) 述語間関係を表す推論規則とは別の種類の知識あるいは解析技術が必要、(4) 依存構造解析の誤り。

注目すべきは、知識の規模が足りていないことが原因で推論規則が適用できず照応関係の解析が行えていないという事例数は高々 46 件であり、周辺文脈が考慮できていないために推論規則が適切に適用できていないという事例が多数を占めているという点である。以降では依存構造解析誤りを除く 3 種類の照応解析誤りの原因について、事例とともに詳しく説明する。

4.1 周辺文脈が考慮できていない

推論規則を適用する際に、入力文および推論規則における述語の共有項以外の周辺文脈考慮できていないため、入力文に対し適切で無い推論規則を適用してしまう問題。この問題を更に細かく分類すると、以下の 5 つに分類することができる。それぞれを実例とともに説明する。

⁵この他に、そもそも問題文の文章が理解し難く、英語ネイティブであっても照応先が判断できないという事例が 3 件存在した。

(1.1) **コピュラ** 共有項とイコールの関係にある項が問題文に含まれており、共有項とイコールであることを認識しなければ照応先を特定できない問題。(例): *Watson_i beat Ken_j at Jeopardy because he_i is a superior machine.* この例では、現在の推論規則では $X \text{ beat} \sim X \text{ be}$ と $\text{beat } Y \sim Y \text{ be}$ との項共有のしやすさから照応先を特定することになるが、これは困難である。正しく解析するためには、*he* と *superior* がイコールの関係にあることを考慮し、 $X \text{ beat} \sim X \text{ superior}$ と $\text{beat } Y \sim Y \text{ superior}$ の項共有のしやすさを比較することで正しく X が項共有しやすくなり、照応先が解析できると考えられる。

(1.2) 多義性

(1) 述語が動詞句の一部を担っているか否か、あるいは (2) 述語が自動詞であるか他動詞であるかによって語義が異なる場合に適切で無い推論規則を適用してしまう問題。(例): *The boy_i laughed at his friend_j because he_j made the joke.* この例は *make* が *make a joke* という動詞句の一部を担っており、*laugh at* の主語と目的語のどちらが *he* と照応関係にあるかという事例である。しかし、現在の推論規則は *make a joke* が動詞句であるということが考慮できていないため、 $X \text{ laugh} \sim X \text{ make}$ と $\text{laugh at } Y \sim Y \text{ make}$ との項共有のしやすさの比較となってしまう、*make* と *make a joke* の語義は異なるため異なった語義の推論規則が適用されてしまい解析を誤る可能性がある。これに対し、*make a joke* を一つの述語として扱うことで、 $X \text{ laugh} \sim X \text{ make a joke}$ と $\text{laugh at } Y \sim Y \text{ make a joke}$ という推論規則による項共有のしやすさを扱えるようになり、正しく Y が照応関係であることが解析できるようになると考えられる⁶。一方、次の例は述語が自動詞であるか他動詞であるかによって語義が異なるために解析を誤った事例である。(例): *The lifeguards_i evacuated the swimmers_j from the public pool because they_j were in danger of getting sick from human wastes in the water.* この例では、避難した人が *they* と照応関係になるが、推論規則の獲得時にも適用時にも述語が自動詞であるか他動詞であるかは区別されていない。そのため、 $X \text{ evacuate}$ と $\text{evacuate } Y$ は X も Y も避難した人である可能性があり、判断することができない。これに対し推論規則の獲得時に述語が目的語を持っているかどうかを保持しておき、推論規則の適用時にも述語が目的語を持っているかどうかを、推論規則の項共有のしやすさのスコアに組み込むことで動詞が自動詞であるか他動詞であるかが一致している場合に項を共有しやすくなり、正しい推論規則が適用できるようになると考えられる。

(1.3) 助動詞・否定表現

否定表現や副詞、助動詞等による事実性や程度といったモダリティの違いを考慮できていないため、誤った推論規則を適用してしまう問題。(例): *Gorden_i got mad at his brother_j because he_j forgot to bring him chocolate.* この例では *forgot to* という動詞を制御する表現の存在により *bring* の事実性は *nonfact* であ

ることがわかる。しかし現在の推論規則では *forgot to* という表現は考慮されていないため、 $X \text{ mad} \sim X \text{ bring}$ と $\text{mad at } Y \sim Y \text{ bring}$ との項共有のしやすさを比べることになりどちらか判断するのは困難である。これに対し、事実性を変化させる *forgot to* を考慮することで $X \text{ mad} \sim X \text{ forget to bring}$ と $\text{mad at } Y \sim Y \text{ forget to bring}$ との項共有のしやすさの比較となり正解である Y が照応関係にあることが認識できると考えられる。

(1.4) 目的語・前置詞の付帯による述語間関係の変化

(a) 述語が目的語になにをとっているか、あるいは (b) 述語に前置詞が付帯しているかによって項を共有しやすい推論規則に変化が生じるため、適切で無い推論規則を適用してしまう問題。(例): *John_i fired Dave_j because he_j had too many absences.* この例では $X \text{ fire} \sim X \text{ have}$ と $\text{fire } X \sim X \text{ have}$ という推論規則の項共有のしやすさを比べることになるが、これでは X に入るタイプの類似度を用いてもどちらが正しいか判断することは困難である。推論規則に述語の目的語まで考慮することで $X \text{ fire} \sim X \text{ have absence}$ と $\text{fire } X \sim X \text{ have absence}$ の項共有のしやすさをを用いることで解析が可能になると考えられる。一方、次の例は前置詞が付帯しているかどうかを考慮しなければならない事例である。(例): *Bob_i cooked omelets for Jack_j, because he_j was hungry.* $X \text{ cook} \sim X \text{ be hungry}$ と $\text{cook for } Y \sim Y \text{ be hungry}$ との項共有のしやすさを比較すると、不正解である X の項共有のしやすさも高くなってしまい、正しく Y が項共有している推論規則が適用されない可能性がある。これに対し、付帯している前置詞を考慮することで $X \text{ cook for} \sim X \text{ be hungry}$ と $\text{cook for } Y \sim Y \text{ be hungry}$ との項共有のしやすさの比較となるため、 X が項共有しにくくなり、照応先は Y であることが解析可能になると考えられる。

(1.5) 接続表現

推論規則の適用時に、述語の極性や事実性を反転させる効果を持つ接続詞が問題文に存在していることを考慮できていないため、誤った推論規則を適用してしまう問題。(例): *Justin Bieber_i sold many more albums than Jack Johnson_j, but he_i is also very unpopular among older adults.* もしこの例における *but* が *and* であつたら *he* の照応先は変わると考えられる。つまり、推論規則を適切に適用するためには *but* を考慮する必要がある。具体的には、 $X \text{ be unpopular}$ との項共有のしやすさをはかるのでは *but* を考慮できていないために正しくない推論規則における項共有のしやすさを比較することになってしまう。極性を反転させる効果のある *but* を考慮し、*unpopular* の極性を反転させ、 $X \text{ sell} \sim X \text{ be popular}$ と $\text{sell than } Y \sim Y \text{ be popular}$ という推論規則の項共有のしやすさを比較することで正解である X の照応関係が解析できると考えられる。

これら5つの周辺文脈が考慮できていない問題のうち、接続表現以外の4つの問題に対しては次のような対策が考えられる。知識獲得時に変項付き述語とそのタイプのみでなく、周辺文脈に関する情報も保持して獲得しておき、知識適用時にも周辺文脈に関する情報の類似度をス

⁶ある程度の動詞句の認識は *Staford CoreNLP* により行っている

コア計算に組み込む手法が挙げられる。このアイデアを実際に実装しWSCに適用した結果については [15] で述べられている。

接続表現問題には、あらかじめ極性の反転が起こる接続表現のリストを用意し、ルールベースで推論規則の極性反転の処理を行った上で推論規則を適用する手法が考えられる。

4.2 推論規則の規模問題

推論規則を用いることによって識別可能と考えられるが、直接的に適用可能な推論規則が現在の大規模推論規則からは得られない問題。ここで確認しておく、本実験では3節で示したとおり、約2.2億文書から獲得した非常に大規模な推論規則の知識ベースを適用している。そのため、推論規則の規模問題は知識獲得の規模を大きくすれば解決するといった単純な問題ではないと考えられる。(例): *The Detroit Lions_i beat the Dallas Cowboys_j because they_j were unprepared.* この事例では、*beat X ~ X unprepared* という項共有の推論規則は今回の大規模語彙知識から得られなかった。この原因としては *unprepared* という表現自体が使われることが少ないために、*beat X* と直接項を共有する文が無かったのではないかと考えられる。一方で、*beat X ~ X be weak* と *X weak ~ X be unprepared* という推論規則は存在ことが確認できたため、これらの推論規則をつなぎ合わせることでこの事例における項共有の手がかりにすることが可能である。このように推論規則を多段に組み合わせることや、推論規則の汎化や同義表現の言い換え、あるいは知識獲得規模の拡大などによって知識規模問題が改善できると考えられる。

4.3 別の種類の知識あるいは解析技術問題

述語の持つ極性情報や固有名詞に関する世界知識、選択選好などといった、推論規則以外の常識的知識や解析技術を必要とするため照応解析を誤る問題。照応解析に必要な知識および解析技術としては以下の4つに分類される。

選択選好

述語の選好性によって照応先が定まる事例。(例): *The scientists_i gave the chimps some bananas_j because they_j were ripe.* *be ripe* の選好性により *they* と *banana* が照応関係であると定まる。選好性を正しく推定するモデルを用意し適用する必要がある。

極性

述語のもつ極性の認識により照応先が定まる事例。(例): *Apple_i defeated Microsoft_j in the war because they_j lack creativity.* 述語 *X defeated Y* のもつ極性により、Apple および Microsoft の極性はそれぞれ Apple: Positive, Microsoft: Negative となる。また、*Z lack creativity* の持つ極性から *they* の極性は *they*: Negative であることがわかる。これを手がかりに、すなわち *they* の照応先は Microsoft であることが定まる。述語のもつ極性に関する知識が必要となる。

世界知識

問題文における固有名詞に関する事前知識によって照応先が定まる事例。(例): *Gladiators_i are viewed*

better than Slaves_j because they_i fight to the death in the arena. 「Gladiators はアリーナの中で死闘する映画である」という事前知識によって *they* の照応先が Gladiators であることが定まる。固有名詞に関する世界知識が必要となる。

高度推論

事前知識や前提などの組み合わせといった複雑な書き換えによって照応先が定まる事例。(例): *Michael_i decided to freeze himself in cryo-stasis even though his father_j was against it, because he_i hopes to be unfrozen in the future when there is a cure available.* *in the future when there is a cure available* という部分から今病気であることや今は治せないことという前提や、今凍結し未来に解凍することで未来へ行けるという事前知識などを推論し解釈することで *he* の照応先が Michael であることが定まる。

5 おわりに

従来の研究において、推論規則を用いることで共参照解析性能の向上が試みられてきたが、十分な有用性が示されずその原因の検証も深くなされてこなかった現状に対し、これまでにない大規模な推論規則を用意し、照応解析タスクへの適用を通し、推論規則適用時の誤り分析を行った。誤りの原因と課題を洗い出した結果、推論規則の規模が足りていない問題以上に、周辺文脈を考慮できていないことが顕著な問題であることが明らかになった。今後、周辺文脈も考慮した大規模な推論規則を構築するとともに、談話関係認識といった談話処理への応用を行っていく予定である。

謝辞 本研究は、JST 戦略的創造研究推進事業 CREST の一環として行われた。

参考文献

- [1] N. Chambers and D. Jurafsky. Unsupervised Learning of Narrative Schemas and their Participants. In *ACL*, pp. 602–610, 2009.
- [2] Nathanael Chambers and Daniel Jurafsky. Unsupervised learning of narrative event chains. In *ACL*, pp. 789–797, 2008.
- [3] T. Chklovski and P. Pantel. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In D. Lin and D. Wu, editors, *Proceedings of EMNLP*, pp. 33–40. ACL, 2004.
- [4] I. Dagan, B. Dolan, B. Magnini, and D. Roth. Recognizing textual entailment: Rational, evaluation and approaches - Erratum. *Natural Language Engineering*, Vol. 16, No. 1, p. 105, 2010.
- [5] Christian Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [6] Andrew Gordon and Reid Swanson. Identifying personal stories in millions of weblog entries. In *Third International Conference on Weblogs and Social Media, Data Challenge Workshop*, San Jose, CA, 2009.
- [7] N. Inoue, E. Ovchinnikova, K. Inui, and J. Hobbs. Coreference Resolution with ILP-based Weighted Abduction. In *COLING*, pp. 1291–1308, 2012.
- [8] Hector J. Levesque. The Winograd Schema Challenge. In *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, 2011.
- [9] D. Lin and P. Pantel. Dirt: discovery of inference rules from text. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference*, pp. 323–328, 2001.
- [10] Oren Melamud, Jonathan Berant, Ido Dagan, Jacob Goldberger, and Idan Szpektor. A two level model for context sensitive inference rules. In *ACL*, pp. 1331–1340, 2013.
- [11] Patrick Pantel, Rahul Bhagat, Bonaventura Coppola, Timothy Chklovski, and Eduard H Hovy. Isp: Learning inferential selectional preferences. In *HLT-NAACL*, pp. 564–571, 2007.
- [12] A. Rahman and V. Ng. Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge. In *Proceedings of EMNLP-CoNLL*, pp. 777–789, 2012.
- [13] J. Ruppenhofer, M. Ellsworth, M.R. Petrucci, C.R. Johnson, and J. Scheffczyk. FrameNet II: Extended Theory and Practice. Technical report, 2010.
- [14] Jun Sugiura, Naoya Inoue, and Kentaro Inui. Recognizing implicit discourse relations through abductive reasoning with large-scale lexical knowledge. In *Workshop on Natural Language Processing and Automated Reasoning*, pp. 76–87, 2013.
- [15] 井之上直也, 杉浦純, 乾健太郎. 談話処理のための推論規則文脈化の一手法. 言語処理学会第20回年次大会予稿集, p. To appear, 2014.
- [16] 井之上直也, 乾健太郎, E. Ovchinnikova, J. R. Hobbs. 大規模世界知識を用いた仮説推論による談話解析の課題と対策. 言語処理学会第18回年次大会予稿集, pp. 119–122, 2012.