

エラー分析 知識獲得グループ

日本語 Winograd Schema Challenge を題材として

柴田 知秀 *

京都大学/独立行政法人 科学技術振興機構 CREST

shibata@i.kyoto-u.ac.jp

1 はじめに

近年、大規模コーパスから常識的な知識を自動獲得する研究が進められている。知識のタイプは語・句の同義・上位下位関係、固有名クラス [1]、格フレーム [3]、スクリプト・事態間知識 [2, 6] など、様々である。以下に例を示す。

- 同義・上位下位
MacBook Air = MBA, MacBook Air → ノートパソコン
- 固有名クラス
ネイマール: サッカー選手, バンプレスト: 企業
- 格フレーム
{ 人, 男, 子供 } が { 犯罪, 犯行 } を 犯す
- スクリプト・事態間知識
X: { 男, 容疑者 } が 犯罪を 犯す → X が 逮捕される

このうち、同義・上位下位や固有名クラス、格フレームについては研究がかなり進んできており、言語資源として公開されているものもある。ここでは、現在研究されているスクリプト・事態間知識を題材として取り上げられている。

知識獲得の研究において難しいのは評価の問題である。例えば、システムが獲得した知識の中からランダムにサンプリングして精度を算出したとしても、獲得された知識が他の解析・タスクで有用であることやカバレッジが十分であることを示すのが難しい。獲得された知識を他のタスクに適用することにより有効性を示す外的な (extrinsic) 評価を行うことが考えられるが、通常、獲得された知識によって精度の変化が生じるのは少数であり、知識の有効性を示すのが難しい。

そこで、あるタスクに特化した評価セットを用いることが考えられる。例えば、テキスト含意認識では RTE 評価セットが整備されている。また、近年、常識的な知識の獲得を評価するものとして、英語では Winograd Schema Challenge(以降、WSC と呼ぶ) という評価セットが構築されている [4]。タスクとしては照応解析であり、例えば以下の問題では、照応詞 “she” に対して、先行詞候補 “Debbie” と “Tina” が与えられ、システムは正しく “Tina” と同定できるかどうか問われる。

(1) When Debbie splashed Tina, she got wet.

以降、下線をひいた語は先行詞候補、太字の語は正例の先行詞、波線をひいた語は照応詞を示す。

この評価セットでは、述語の選択選好などでは先行詞を同定することが困難で、常識的な知識が必要な問題が集められており、上記の問題では、「X さんが Y さんに水をかけると、Y さんが濡れる」という常識的な知識が必要となる。

この評価セットの興味深いところは、類似した問題がペアとなっていることで、先にあげた問題は以下の問題とペアになっており、以下では、先の問題では誤りであった先行詞 “Debbie” が正例の先行詞となっている。

(2) When Debbie splashed Tina, she got in trouble.

一般に照応解析では主語が先行詞になりやすいなどの統語的な情報が有効であるが、上記のように一部のみに変えた問題をペアにし、先行詞が異なるようにすることにより、統語的な情報が有効にならないようにし、システムが常識的な知識を保持しているかどうかだけを評価できるように工夫されている。

*メンバー: 小浜翔太郎 (京大), 井之上直也 (デンソー/東北大), 飯田龍 (NICT)

日本語においても常識的な知識獲得の評価を行えるように、英語 WSC を日本語に翻訳することにより日本語 WSC 評価セットを構築した [10]。本研究ではこの評価セットを用いる。

Levesque が作成したオリジナルの評価セットは 284 問と少数であるため¹、本研究では、Rahman らの研究で構築されたセットを用いる²。問題数は training が 1,322、test が 564 の計 1,886 問である。

2 日本語 WSC の構築

英語 WSC を日本語に翻訳することによって、日本語 WSC を構築した。構築された日本語 WSC の例を表 1 に示す。冒頭にあげた 2 例は表 1 の最初の 2 行のように翻訳されている。

翻訳に際しての注意点として、問題文にマッチするような常識的な知識が存在するかを評価できることが優先であり、日本語として自然かどうかの優先度は低いことがあげられる。例えば、以下の 2 問を日本語に翻訳する際、1 問目の照応詞 “they” は「彼ら」に、2 問目の照応詞 “they” は「それら」と訳するのが自然であるが、そうすると、システムはその違いを手がかりに先行詞を推定することができる。

- (3) Chevrolet had to recently recall their cars because they received complaints about the brakes in their product.
- (4) Chevrolet had to recently recall their cars so they could receive a repair in their brakes.

そこで、日本語としては不自然ではあるが、いずれも「彼ら」と訳し、先行詞同定の手がかりにならないようにする。

- (5) シボレーは最近彼らの車をリコールした。彼らは自社製品のブレーキに関する苦情を受けたからだ。
- (6) シボレーは最近彼らの車をリコールした。だから彼らは車のブレーキの修理を受けることができた。

¹<https://www.cs.nyu.edu/davise/papers/WS.html> で公開されている。こちらのデータについては北海道大学のグループによって日本語に翻訳され、上記のページの末尾にて公開されている。

²<http://www.hlt.utdallas.edu/~vince/data/emnlp12/> で公開されている。

3 日本語 WSC の分析

日本語 WSC の training のうちの 100 問を検討したところ、問題を解くために必要な知識、また、評価セットから除外すべきという観点から以下のように分類できることがわかった。

1. 選択選好 (26)
2. 事態間知識 (22)
3. メタ知識 (12)
4. 除外 (不適切, 文化差) (18)
5. 上記以外の難問 (29)

括弧内の数は問題数を示す。また、カテゴリ 1,2,3 は複数所属可能とする。

以下では各カテゴリについて詳細に述べる。

3.1 選択選好

1 節では、選択選好で解くのが難しい問題が集められていると述べたが、項に「物」や「人」をとりやすいのような選択選好よりも広義の選択選好で解ける問題が存在する。例えば、以下の問題では述語「吠える」のガ格の選択選好は「猫」よりも「犬」の方が高いという知識から、先行詞を「犬」と同定することができる。

- (7) 猫は犬より賢い。彼らは理由無く吠えるからだ。

3.2 事態間知識

以下の問題では、「ある企業 X が破綻すると別の企業 Y が企業 X を買収する」という事態間知識から、「彼ら」の先行詞は「モトローラ」と同定することができる。

- (8) グーグルはモトローラを買収した。彼らが破綻していたからだ。

事態間知識だけでは解けず、問題と事態間知識のマッチングの際に同義知識が必要となる場合もある。以下の問題を解くには、事態間知識「X が Y に頼む ⇒ Y が断る」と同義知識「Y に頼む」=「Y に頼みごとをする」が必要となる。

- (9) ジェームズはロバートに頼みごとをした。しかし彼は断った。

デビーが <u>ティナ</u> に水をかけた。	<u>彼女</u> はびしょびしょになった。
デビーが <u>ティナ</u> に水をかけた。	<u>彼女</u> はめんどうをおこしたのだ。
バスの運転手は <u>子供</u> に怒鳴った。	<u>彼女</u> が彼女の車を運転した後のことだ。
バスの運転手は <u>子供</u> に怒鳴った。	<u>彼女</u> が窓ガラスを割ったからだ。
ジンボは <u>ボバート</u> のところから走って逃げた。	<u>彼</u> はひどい臭いがしたからだ。
ジンボは <u>ボバート</u> のところから走って逃げた。	<u>彼</u> が先に車に着きたかったからだ。
男は隣人の自転車を盗んだ。	<u>彼</u> は一台必要だったからだ。
男は <u>隣人</u> の自転車を盗んだ。	<u>彼</u> が一台余分に持っていたからだ。
メアリーは <u>スーザン</u> の部屋を掃除した。	そして <u>彼女</u> は感謝した。
メアリーは <u>スーザン</u> の部屋を掃除した。	そして <u>彼女</u> は頼みごとをした。

表 1: 日本語 WSC の例 (下線をひいた語は先行詞候補、太字の語は正例の先行詞、波線をひいた語は照応詞を示す)

3.3 メタ知識

例えば以下の問題を考える。

- (10) メアリーはスーザンの部屋を掃除した。そして彼女は感謝した。

この問題を解くには、以下のような事態間知識が獲得できれば解くことができるが、このようなことがコーパスにある程度の頻度で書かれ、知識として獲得されるとは考えにくい。

X:{主体}がY:{主体}の部屋を掃除する ⇒ Y:{主体}がX:{主体}に感謝する

したがって、直接問題にマッチするような知識ではなく、以下の2つの知識があれば、問題を解くことができる。

- XがYに「良いこと」をすると、YがXに感謝する
- 部屋を掃除することは「良いこと」

ここでは、「XがYに良いことをすると、YがXに感謝する」のような汎化した知識をメタ知識と呼ぶことにする。

他の例としては以下がある。

- (11) ボブはジャックにオムレツを作った。彼は作り方を知っていたからだ。

この問題についても、これを解くための直接的な知識がコーパスから獲得されるとは考えにくく、以下のようなメタ知識が必要となる。

XがVした ⇒ XがVする方法を知っていた

上記のようなメタ知識を獲得した研究はなく、今後の課題である。

3.4 除外

もともとの英語の問題で不適切、もしくは、英語を日本語に翻訳して構築したために不適切となったもの、また、文化差により日本語の評価セットとして不適切なものなど、除外すべきものがある。

不適切

例えば、以下の問題では、ニコニコマークはもともと“smiley face”であるが、下記の文脈で何を指しているかがわからず、問題として不適切である。

- (12) 彼はニコニコマークに雪を乗せた。それは濡れていたからだ。

文化差

例えば、以下の問題では、「ゾンビ」に関する知識が必要となるが、日本語のテキストでは出現しにくいことから、以下の問題は評価セットから除外する。

- (13) ゾンビは生き残った人たちを追いかけた。彼らは空腹だったからだ。

3.5 上記以外の難問

上記には分類できない難問がある。例えば、下記のような問題は多段に推論を行えば解ける可能性があるが、現在の知識・解析では大変難しい問題である。

ガ	犬:2469, 愛犬:123, 飼い犬:85, ..., 猫:13, ...
デ	遠く:67, 外:29, 近く:20, ...

図 1: 動詞「吠える」の 3 番の格フレーム (名詞の後の数字はコーパス中での頻度を示す)

- (14) オーケストラは聴衆のブーイングを受けた。彼らはロックバンドを期待していたからだ。
- (15) 男は隣人の自転車を盗んだ。彼が一台余分に持っていたからだ。
- (16) 学究的な同好会はメンバーを失った。彼らに十分な資金がなかったからだ。

3 節の冒頭で述べたが、カテゴリ 1,2,3 については複数に所属することを許した。例えば、以下の問題は、選択選好に関する知識でも解け、また、「X が退屈する \Rightarrow X が居眠りする」という事態間知識でも解けると考えられるので、カテゴリ 1 と 2 に属している。

- (17) 学生たちは教授の講義中に居眠りをした。それらが退屈だったからだ。

4 利用した知識

現在の知識・解析システムでどのくらいの問題を解けるかを調べるために、以下にあげる知識を利用した。

4.1 格フレーム

選択選好に関する知識は例えば格フレームから得ることができる。図 1 に、河原らの手法 [3] で Web テキストから自動構築した格フレームを示す。動詞「吠える」の 3 番の格フレームのガ格において、「犬」、「猫」の頻度はそれぞれ 2,469、13 回であり、頻度に大きな差があることから、「彼ら」の照応先が「犬」であると推定できる。

4.2 事態間知識

事態間知識として、柴田らが Web コーパスから自動獲得したもの [6] を利用する。この知識では、事態を述語項構造で表し、一つの知識は「ある事態 E_1 が生じた後に、しばしば別の事態 E_2 が生じる」ということを表す。以下に例を示す。

X:{ 会社,{ 主体 }} が破綻する \Rightarrow Y:{ 会社 } が X:{ 会社,{ 主体 }} を買収する

この事態間知識では項の対応がとれており、この例では、最初の事態のガ格と次の事態のヲ格の対応が付いている。この情報を用いることにより、上記の問題の「彼ら」は「破綻」のガ格であるので、「買収」のヲ格である「モトローラ」が先行詞であることがわかる。

また、事態間知識のマッチングにおける同義の知識は分布類似度 [7] を利用した。

5 誤り分析

日本語 WSC の training のうちの 548 問に対して、4 節で述べた知識を適用して誤り分析を行った。以下の手法を使って精度を計算した。精度を表 2 に示す。

- CF: 正例・負例の頻度が 5 倍以上離れている³
- Event: 事態間知識における項のアライメントに基づいて決定
- Event+DS: 事態間知識のマッチングに分布類似度を利用
- CF+Event+DS: 全て利用

なお、CF+Event+DS においては、Event+DS を優先し、それで先行詞が決まった場合はそれを採用し、決まらなかった場合に CF で決定した。

5.1 格フレーム

以下に正解例を示す。格フレームを利用することにより、それぞれ「子供」「少年」が先行詞であると正しく同定することができている。

- (18) バスの運転手は子供に怒鳴った。彼女が窓ガラスを割ったからだ。
- (19) 少年は警官のところから走った。彼は犯罪を犯したからだ。

以下に誤り例を示す。

- (20) 男は乞食にいくらか金をやった。彼があまりにしつこかったからだ。

³一方の頻度が 0 の場合、もう片一方のものを採用した。

手法	correct	wrong	accuracy	match
CF	48	27	0.64	75
Event	12	7	0.63	20
Event + DS	18	11	0.62	29
CF + Event + DS	59	36	0.62	95

表 2: 実験結果 (CF:正例・負例の頻度が5倍以上離れている, Event:事態間知識における項のアライメントに基づいて決定, Event+DS:事態間知識のマッチングに分布類似度を利用, CF+Event+DS:全て利用)

「しつこい」のガ格において、「男」の頻度は38、「乞食」の頻度は0であったため、システムは「彼」の照応先を誤って「男」と同定した。しかし、この問題は選択選好で解けそうにないものであり、格フレームの利用により誤っても仕方ないものである。実際、誤りである27個について調べたところ、格フレームの誤りのものはなく、また、このうち3つは事態間知識の利用によって正しいものが選ばれていた。

5.2 事態間知識

以下に正解例を示す。

(21) ルイジはマリオを助けた。彼が困っていたからだ。

この例には以下の事態間知識がマッチすることにより、「困る」のガ格は「助ける」のヲ格であることがわかり、「彼」の照応先が「マリオ」であると正しく同定することができた。

$X:\{女, 友人\}$ が $Y:\{金\}$ に困る \Rightarrow $X:\{女, 友人\}$ を $Y:\{金\}$ で助ける

事態間知識によって生じた誤り7つを以下に分類した。

- 事態間知識の誤り [4]

(22) 看護師はその少女にペロペロキャンディーをあげた。彼女が泣き止むように。

この例には以下の事態間知識がマッチし、「泣きやむ」のガ格が誤って「あげる」のガ格に対応付いてしまった。

$X:\{私\}$ が $Y:\{ミルク, 飴\}$ をあげる \Rightarrow $X:\{私\}$ が泣きやむ

なお、格フレームの頻度をみると、「泣きやむ」のガ格における「看護師」の頻度は0回、「少女」の

頻度は10回であるので、照応先を「少女」と正しく同定することができる。

- 事態間知識は誤りではない [2]

(23) ジンボはロバートより良く食べる。彼はダイエット中だからだ。

以下の事態間知識がマッチし、「彼」の照応先を誤って「ジンボ」と同定してしまったが、下記の事態間知識自体は誤りではない。この問題を正しく解くためには、「良く」で比較していることを解釈することが必要となる。

$X:\{私\}$ が $Y:\{ゼリー, クッキー\}$ を食べる \Rightarrow $X:\{私\}$ が $Y:\{ゼリー, クッキー\}$ でダイエット

- 事態間知識マッチング時の誤り [1]

(24) ネズミは猫から逃げてずっと台所まで走った。それが追いかけてきたからだ。

この問題に対しては以下の事態間知識がマッチした。現在は主辞の動詞に対してマッチさせているため、「走る」にマッチしている。「猫」が係っており、また、省略解析により「ネズミ」がガ格であることがわかることにより、「逃げる」にマッチさせれば、マッチング時の誤りを防ぐことができると考えられる。

$X:\{俺, 私\}$ が $Y:\{中\}$ を走る \Rightarrow $X:\{俺, 私\}$ が $Y:\{中\}$ に追う

6 関連研究

Winograd Schema Challenge(WSC)[4]が提唱されて以降、いくつかの研究がこの問題を解き、現在の解析器・知識の問題点などを議論している [5, 9, 8]。

Rahmanらは、機械学習を用いて正例先行詞と負例先行詞のランキング問題として解いている [5]。素性として、Chambersらが獲得した事態間知識、Google検索のヒット件数、FrameNet、極性、接続詞などから得られた様々なものを利用している。

杉浦らは事態間知識を大規模コーパスから獲得し、それを英語WSCに適用し、その解析誤りの分析を行っている [9]。解析誤りの主な要因は、周辺文脈が考慮できていない、推論知識が不足している、別の種類の知識が必要、依存構造解析誤りなどであったと報告されている。

井之上らの手法 [8] では、まず、大規模コーパスからChambersらの方法を用いて、周辺文脈付きの事態間知識を獲得する。そして、問題文と類似している近傍 k 個の事態間知識に基づき、先行詞を同定している。実験の結果、類似度関数に文脈の類似度を考慮することにより精度が向上したと報告されている。

ここにあげたような英語WSCの分析と日本語WSCの分析を通して英語・日本語における知識獲得の比較を行うことが今後の課題としてあげられる。

7 おわりに

本論文では英語の Winograd Schema Challenge(WSC) を日本語に翻訳することにより構築した日本語WSCを用いて、必要な知識を分析し、次に、知識を適用した時の誤り分析を行った。

構築した日本語WSCの評価セットは公開し、様々なシステムによる誤り分析を行う予定である。また、英語WSCと日本語WSCの分析を通して英語・日本語における知識獲得の比較を行う予定である。

謝辞

本研究は科学技術振興機構CREST「知識に基づく構造的言語処理の確立と知識インフラの構築」の支援のもとで行われた。

参考文献

[1] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka, and Tom M. Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010)*, 2010.

- [2] Nathanael Chambers and Dan Jurafsky. Un-supervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*, pp. 789–797, 2008.
- [3] Daisuke Kawahara and Sadao Kurohashi. Case frame compilation from the web using high-performance computing. In *Proceedings of LREC-06*, 2006.
- [4] Hector J. Levesque. The Winograd Schema Challenge. In *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, 2011.
- [5] Altaf Rahman and Vincent Ng. Resolving complex cases of definite pronouns: The winograd schema challenge. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 777–789, 2012.
- [6] Tomohide Shibata and Sadao Kurohashi. Acquiring strongly-related events using predicate-argument co-occurring statistics and case frames. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP2011, poster)*, pp. 1028–1036, 2011.
- [7] Tomohide Shibata and Sadao Kurohashi. Predicate-argument structure-based textual entailment recognition system exploiting wide-coverage lexical knowledge. *Special Issue of ACM TALIP on RITE (Recognizing Inference in Text)*, Vol. 11, No. 4, pp. 16:1–16:23, 2012.
- [8] 井之上直也, 杉浦純, 乾健太郎. 共参照解析のための事象間関係知識の文脈化. 言語処理学会第20回年次大会論文集, pp. 717–720, 2014.
- [9] 杉浦純, 井之上直也, 乾健太郎. 共参照解析における事象間関係知識の適用. 言語処理学会第20回年次大会論文集, pp. 713–716, 2014.
- [10] 柴田知秀, 小浜翔太郎, 黒橋禎夫. 日本語 winograd schema challenge の構築と分析. 言語処理学会第21回年次大会 論文集, 2015.