

日本語 Winograd Schema Challenge の構築と分析

†‡ 柴田 知秀 † 小浜 翔太郎 †‡ 黒橋 禎夫

† 京都大学 ‡ 独立行政法人 科学技術振興機構 CREST

{shibata, kohama, kuro}@nlp.ist.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。

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を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。

本研究では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。次に、この問題を解くために必要な知識を分類・分析した。

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 を構築する。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

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。

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)

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。

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)

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)

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。

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。

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。

このような知識は例えば、柴田らが Web コーパスから自動獲得した事態間知識 [6] から得ることができる。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。

$X:\{ \text{会社}, \langle \text{主体} \rangle \}$ が破綻する $\Rightarrow Y:\{ \text{会社} \}$
が $X:\{ \text{会社}, \langle \text{主体} \rangle \}$ 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。

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 \Rightarrow 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。

このような同義の知識は WordNet や分布類似度 [7] から得ることができる。

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。

$X:\{ \langle \text{主体} \rangle \}$ が $Y:\{ \langle \text{主体} \rangle \}$ の部屋を掃除する $\Rightarrow Y:\{ \langle \text{主体} \rangle \}$ が $X:\{ \langle \text{主体} \rangle \}$ 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 $\Rightarrow X$ 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。

- (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 ⇒ 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 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。

5 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、日本語 WSC を構築し、次に、日本語 WSC を解くために必要な知識を分析した。分析によれば、必要な知識は、選択選好、事態間知識、メタ知識などに分類できることがわかった。

構築した日本語 WSC の評価セットは公開する予定である。今後の課題としては、日本語 WSC を解くシステムを構築し、現状の知識・解析器の問題点を検討することや、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などがあげられる。

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。また、翻訳は国立情報学研究所 宮尾研究室にて行われた。ここに記して感謝の意を表す。

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. Unsupervised 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.