

依存型意味論を用いた日本語含意関係コーパスの分析に向けて

中村 絢子 (指導教員: 戸次 大介)

1 はじめに

2つの文が含意関係を持つとは一方の文が真であるときに他方の文も真である¹ということを目指す。文間の含意関係の有無を判定するシステムに機械学習させるための正解データを構築する方法論として金子ら [1] のものがある。現時点では、正解データの構築手法は人間の直観に基づいて策定されており、金子ら [1] の手法で提案された分類が言語学的に妥当であるかどうかの分析がなされていない。

本研究では、金子ら [1] の正解データの例文に Bekki[2] で提案された依存型意味論によって意味表示を与え、正解データの構築手法を分析する。また、推論の際に必要な知識を依存型理論を用いたオントロジーで記述する。

2 関連研究

本節では含意関係認識の機械学習のための正解データを構築する手法の提案として金子ら [1] の研究および Bekki[2] で提案された依存型意味論 (DTS) について述べる。

2.1 含意関係認識のための正解データ

金子ら [1] は、正解データの構築を通して、2つの文の間に成立するプリミティブな含意関係²を用いて1つの文を複数回書き換えて目的の文を生成する過程を提示している。

以下に文の書き換えの例を示す。〈t2〉は〈t1〉の「語順の入れ替え」という含意関係を用いた書き換えである。

〈t1〉 第1回バスケットボール欧州選手権は、1935年に開催された。〈\t1〉

〈t2〉 1935年に、第1回バスケットボール欧州選手権が開催された。〈\t2〉

しかし、文の書き換えの際に統語的なレベルで書き換える場合と意味的なレベルで書き換える場合が混在しているため、正解データの妥当性を分析することを困難にしていると考えられる。

2.2 DTS について

Bekki[2] で提案された DTS は、依存型理論を用いた自然言語の意味論である。

DTS が自然言語の意味論として優れている点は以下の4項目である。

1. 動的な束縛を表現することができる。
2. 証明論的に計算することができる。
3. 語彙化文法の意味部門として利用できる。
4. 到達可能性の条件を表現することができる。

実テキストを網羅した意味論は多くなく、DTS が実テキストにおいてどの程度の範囲を網羅しているかについては検証が待たれる。

3 分析

これらの点を踏まえ、本研究では、金子ら [1] で提案された正解データの例文に Bekki[2] で提案された DTS による意味表示を与え、金子ら [1] の正解データの構築手法を分析すると共に、DTS の実テキストに対する経験的な網羅性を検証する。

3.1 DTS を用いた意味表示

以下の文に DTS による意味表示を与えた例を示す。

- (1) 年功序列を基礎として、年長者が上司、若年者が部下となり、指揮・命令系統に不備が出ない。
- (2) 年功序列を基礎として、年長者が上司となる。

(1')

$$\begin{aligned} &(\lambda c)(\Sigma w:World)(\Sigma t:Time(w)) \\ &\Sigma u_1:((\Sigma e:Event(w)) \text{ とする } (w,e,pro, \text{年功序列, 基礎}) \wedge pres(e,t)) \\ &\Sigma u_2:((\Sigma e:Event(w)) \text{ とする } (w,e, \text{年長者, 上司}) \wedge pres(e,t)) \\ &\quad \wedge \text{となる } (w,e, \text{若年者, 部下}) \wedge pres(e,t)) \\ &\Sigma u_3:(\neg(\Sigma e:Event(w)) \text{ 不備がでる } (w,e, \text{指揮・命令系統}) \\ &\quad \wedge pres(e,t)) \end{aligned}$$

(2')

$$\begin{aligned} &(\lambda c)(\Sigma w:World)(\Sigma t:Time(w)) \\ &\Sigma u_1:((\Sigma e:Event(w)) \text{ とする } (w,e, \text{年功序列, 基礎}) \wedge pres(e,t)) \\ &\Sigma u_2:((\Sigma e:Event(w)) \text{ とする } (w,e, \text{年長者, 上司}) \\ &\quad \wedge pres(e,t)) \end{aligned}$$

経験的に (1) は (2) を含意している。また、図 2 に示すように、DTS では、(1') から (2') は依存型理論の推論規則 (ΣE) によって導出可能である³。よって、金子ら [1] の正解データは理論的な含意関係と一致する。しかし、次の例では説明が難しい。

- (3) ルキウス・ドミティウス・アウレリアヌス (以下ルキウス) の突然の死で皇帝不在の中、マルクス・クラウディウス・タキトゥス (以下マルクス) が帝位につく。

¹自然言語処理においては「含意関係」という用語はもう少し広い意味で用いられている。

²これを「基本文関係」と呼ぶ

³(ΣE) はペアの第一要素、または第二要素を取り出す操作に相当する。

$$\begin{array}{c}
P : (\lambda c)(\Sigma w : World)(\Sigma t : Time(w)) \\
\Sigma u_1 : ((\Sigma e : Event(w)) \text{ とする } (w, e, pro, \text{年功序列, 基礎}) \wedge pres(e, t)) \\
\Sigma u_2 : ((\Sigma e : Event(w)) \text{ となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t)) \\
\wedge \text{ となる } (w, e, \text{若年者, 部下}) \wedge pres(e, t) \\
\Sigma u_3 : (\neg(\Sigma e : Event(w)) \text{ 不備がでる } (w, e, \text{指揮・命令系統}) \wedge pres(e, t)) \\
\hline
(\Sigma E) \frac{\pi_1 P : (\lambda c)(\Sigma w : World)(\Sigma t : Time(w))}{\pi_1 P : (\lambda c)(\Sigma w : World)(\Sigma t : Time(w))} \\
\Sigma u_1 : ((\Sigma e : Event(w)) \text{ とする } (w, e, pro, \text{年功序列, 基礎}) \wedge pres(e, t)) \\
\Sigma u_2 : ((\Sigma e : Event(w)) \text{ となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t)) \\
\hline
(\Sigma E) \frac{\pi_2 \pi_1 P : \Sigma u_2 : ((\Sigma e : Event(w)) \text{ となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t)) \wedge \text{ となる } (w, e, \text{若年者, 部下})}{\pi_2 \pi_1 P : (\text{となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t)) \wedge \text{ となる } (w, e, \text{若年者, 部下}) \wedge pres(e, t)} \\
\equiv (\Sigma x : (\text{となる } (w, e, \text{年長者, 上司}))(\text{となる } (w, e, \text{若年者, 部下}))) \\
\hline
(\Sigma E) \frac{\pi_1 \pi_2 \pi_1 P : \text{となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t)}{\pi_1 \pi_2 \pi_1 P : \text{となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t)} \\
\hline
(\Sigma I) \frac{(\pi_1 \pi_1 P, (\pi_1 \pi_2 \pi_1 P, \pi_1 \pi_2 \pi_2 \pi_1 P)) : (\lambda c)(\Sigma w : World)(\Sigma t : Time(w)) \Sigma u_1 : ((\Sigma e : Event(w)) \text{ とする } (w, e, pro, \text{年功序列, 基礎}) \wedge pres(e, t)) \Sigma u_2 : ((\Sigma e : Event(w)) \text{ となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t))}{(\lambda c)(\Sigma w : World)(\Sigma t : Time(w)) \Sigma u_1 : ((\Sigma e : Event(w)) \text{ とする } (w, e, pro, \text{年功序列, 基礎}) \wedge pres(e, t)) \Sigma u_2 : ((\Sigma e : Event(w)) \text{ となる } (w, e, \text{年長者, 上司}) \wedge pres(e, t))}
\end{array}$$

図 2: (3') から (4') の導出

(4) アウレリアヌスの突然の死で皇帝不在の中、マルクス・クラウディウス・タキトゥスが即位する。

(3')

$$\begin{array}{l}
(\lambda c)(\Sigma w : World)(\Sigma t : Time(w)) \\
\Sigma u_1 : ((\Sigma e : Event(w)) \text{ 突然の死 } (w, e, \text{ルキウス}) \wedge pres(e, t)) \\
\Sigma u_2 : (\neg(\Sigma e : Event(w)) \text{ 存在する } (w, e, \text{皇帝}) \wedge pres(e, t)) \\
\Sigma u_3 : ((\Sigma e : Event(w)) \text{ つく } (w, e, \text{マルクス, 帝位}) \wedge pres(e, t))
\end{array}$$

(4')

$$\begin{array}{l}
(\lambda c)(\Sigma w : World)(\Sigma t : Time(w)) \\
\Sigma u_1 : ((\Sigma e : Event(w)) \text{ 突然の死 } (w, e, \text{アウレリアヌス}) \wedge pres(e, t)) \\
\Sigma u_2 : (\neg(\Sigma e : Event(w)) \text{ 存在する } (w, e, \text{皇帝}) \wedge pres(e, t)) \\
\Sigma u_3 : ((\Sigma e : Event(w)) \text{ 即位する } (w, e, \text{マルクス}) \wedge pres(e, t))
\end{array}$$

(3) と (4) は意味的にはおおそ同じことを述べた文であるが、人名が省略されていたり、(3) の「帝位につく」という言葉と同じ意味で (4) の「即位する」という言葉が使われていたりするなど意味表示に違いが見られる。この 2 文の含意関係を認識するためには、「帝位につく」という言葉と「即位する」という言葉が意味が近いという知識が必要になり、またその知識を記述することが重要になる。

3.2 オントロジー

オントロジーとは概念間の関係を記述したものであり、本研究では Patrick[3] に基づき部分型をもつ依存型理論を用いて記述する。たとえば Patrick[3] では「リンゴは色を持つ」という概念は以下のように記述される。

$$\Pi u : (\Sigma x : Apple.Color[x]) HasColor[\pi_1 u]$$

Π は全称量子化 (\forall) に対応する働きをする。よってこの式は全てのリンゴは何かしらの色を持つことを表している。

これを踏まえ、(3) の「帝位につく」という概念と (4) の「即位する」という概念の関係を記述することを考える。「帝位につく」という概念は「即位する」という概念の下位概念と捉えることができる。部分型を持つ依存型理論で下位概念を記述する場合は、以下のように記述することができる。

$$\text{帝位につく} \preceq \text{即位する}$$

また、部分型を推論する場合には以下の規則が必要になる。

$$\frac{\Gamma \vdash M : A \quad \Gamma \vdash A' : Type \quad A \preceq A' (Sub)}{\Gamma \vdash M : A'} (Sub)$$

「帝位につく」という概念と「即位する」という概念の関係を、部分型ではなく DTS を用いて記述すると以下ようになる。

$$\begin{array}{l}
(\Pi w : World) Actual(w) \rightarrow \\
(\Pi e : Event(w)) (\Pi x : Entity(w)) \\
\text{帝位につく } (w, e, x) \rightarrow \text{即位する } (w, e, x)
\end{array}$$

上の記述と推論規則を利用すると (5') から (6') を導出することが可能である。

4 考察・まとめ

金子ら [1] が作成した含意関係認識のための正解データの一部を Bekki[2] で提案された依存型意味論で記述するによって金子ら [1] の正解データの構築手法の妥当性と依存型意味論の実テキストに対する網羅性を検証した。一方の文が他方の文を含意するような例では依存型理論の推論規則を適用することによって比較的容易に 2 つの文が含意関係を持つことを証明することができる。また、推論規則のみでは証明が困難な類義語への書き換えなどは依存型理論を用いたオントロジーを用いることによって、意味論と同じ体系で知識の記述が可能になった。金子ら [1] で提案された手法で書き換えた場合の書き換えの結果が一致するかどうかの計算は今後の課題である。また、記述した知識を実際に意味論で推論する際に利用できるようにしていく予定である。

参考文献

- [1] 金子 貴美, 戸次 大介, 宮尾 祐介. “基本文関係に分解した日本語含意関係認識アノテーション”, 人工知能学会第 27 回全国大会. (2013)
- [2] Bekki, Daisuke. (to appear). “Dependent Type Semantics: An Introduction”, the 2012 edition of the LIRa yearbook: a selection of papers, University of Amsterdam. (2012)
- [3] Patrick, B., Richard, D. “A type-theoretical approach for ontologies: The case of roles”, Applied Ontology 7 pp.311-356. (2012)
- [4] 兼岩 憲. “記述論理と Web オントロジー言語”, オーム社. (2009)