

# オントロジーを用いた自然言語の推論に向けて

中村 絢子<sup>1</sup>峯島 宏次<sup>2,4</sup>戸次 大介<sup>1,3,4</sup><sup>1</sup> お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース<sup>2</sup> お茶の水女子大学 シミュレーション科学教育研究センター<sup>3</sup> 国立情報学研究所<sup>4</sup> 独立行政法人科学技術振興機構, CREST

{nakamura.ayako,bekki}@is.ocha.ac.jp

mineshima.koji@ocha.ac.jp

## 1 はじめに

自然言語の含意関係認識では、推論の前提や結論の意味内容だけではなく、前提には明示されていない世界知識を必要とする場合があることはよく知られている。

(1) 前提 雪が積もると、ラッセル車が運行される。

結論 雪が積もると、除雪作業を行う列車が運行される。

(1) は、「ラッセル車は除雪作業を行う列車である」という知識を使って前提から結論が導かれる例である。

また、照応現象の一つであるいわゆる Bridging (Clark [2]) においても同様の世界知識が必要になる。

(2) 北斗星に乗って夕食を食べようと思った。しかし、食堂車はまだ開いていなかった。

(2) において第2文の「食堂車」を「北斗星の食堂車」と解釈するためには、「北斗星には食堂車がある」という知識を用いる必要がある。

このような知識の形式的記述には、オントロジーを用いることが有効である。溝口ら [8] によると、オントロジーとは、共通語彙 (概念) を提供する体系的な辞書であり、概念と概念間の関係 (意味リンク) を用いて記述される。概念間の関係を明示的に記述することによって、そのバックグラウンドにある暗黙的な情報を明示することが可能となる。人工言語ベースの形式的なオントロジーは、高い形式性をもち、従来の上位・下位概念に基づく概念関係の記述を超えたより精緻な知識の構造的記述が可能になるという利点がある。

乾ら [4] は、自然言語ベースのオントロジーを構築し含意関係認識へ適用している。これは、動詞の述語項構造を中心としたオントロジーである。本論文は、

溝口ら [7][8] の形式的なオントロジーに基づいて、動詞に加えて名詞によって表現されるより複雑な概念間の関係についても記述することを目的とし、自然言語推論に必要な世界知識を記述する枠組みとはどのようなものになるのかを考察する。

本研究では、鉄道というドメインに着目し、溝口ら [8] で提案されたオントロジー構築ツール「法造」を用いてオントロジーを構築する。また、そのオントロジーを Bekki [1] で提案された依存型意味論 (DTS) と接合することで、推論と照応のプロセスを形式化することを試みる。鉄道のオントロジーは、車両などの体系的な分類法がすでに存在し、日常的な推論への適用が豊富に見い出されるという点で、オントロジーを用いた推論研究の優れた題材を提供する。

## 2 オントロジーの構築

本節では、溝口ら [8] によるオントロジーの基本概念を導入した上で、推論・照応に関連する世界知識を表現した鉄道オントロジーの構築例を示す。

### 2.1 意味リンクとロール概念

ここでは、is-a リンク、part-of リンク、attribute-of リンクの3種類の意味リンクを取り上げる。列車は、気動車や電車等の異なる概念を含んでいる。これらを分けて記述する場合は、列車の集合の部分集合として気動車の集合や電車の集合があることを明記する必要がある。これを is-a リンクを用いて記述する。is-a リンクの根元を上位概念 (スーパークラス)、is-a リンクの先を下位概念 (サブクラス) と呼ぶ。is-a リンクを使用して種類を書き分ける場合は、部分集合同士が disjoint になるようにする必要がある。

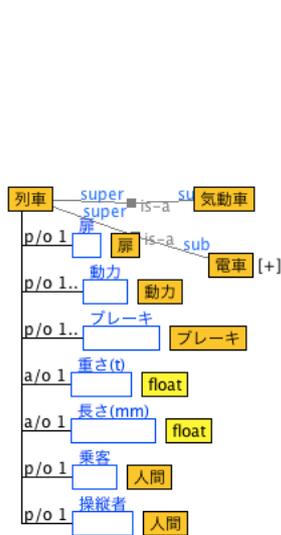


図 1: 列車に関するオントロジー

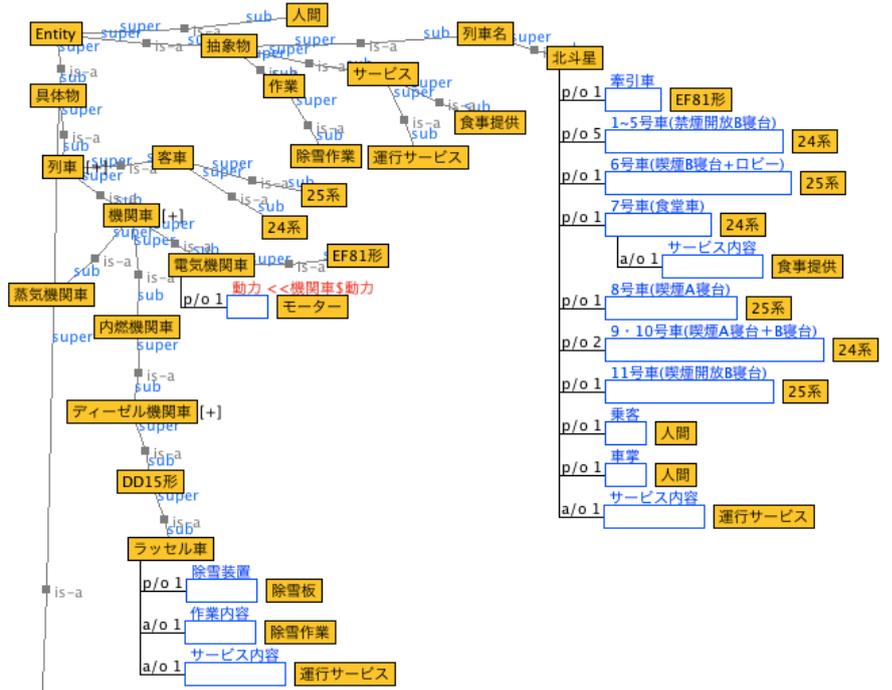


図 2: ラッセル車・北斗星に関するオントロジーの一部

各概念がもっている性質は、part-of リンクと attribute-of リンクを用いて記述される。図 1 は、「列車」は、扉、動力、ブレーキを部分としてもち、重さ・長さを属性としてもつ」という知識を階層的に表している。part-of リンク (p/o) は、「扉」や「ブレーキ」など、「列車」の空間的な部分を構成する概念に対して用いられる。一方、重さ・長さ等の物理量や色は、attribute-of リンク (a/o) を用いて記述される。

「列車」「気動車」「電車」「人間」など、コンテキストや他の概念に依存せず、それ自身の性質に基づいて規定される概念は基本概念と呼ばれる。一方、コンテキストに依存して決定される役割を表す概念は、ロール概念と呼ばれる。例えば、「乗客」「操縦者」は、ロール概念であり、「電車」という概念が定めるコンテキストの中での役割を表す。同様に、図 1 の「扉」「動力」「ブレーキ」「重さ」「長さ」もまた、「列車」を参照して規定されるロール概念である。ロール概念が規定する役割を担う基本概念は、クラス制約によって指定される。例えば、図 1 で、「重さ」「長さ」のクラス制約は float である。上位概念がもつクラス制約は、下位概念に自動的に引き継がれる。

## 2.2 オントロジーの構築例

ラッセル車と北斗星に関するオントロジーの一部を図 2 と図 3 に示す。

「ラッセル車」は、「列車」「ディーゼル機関車」の下位概念であり、同時に、除雪を行うことがその作業内容であるので、「抽象物」として「作業」「除雪作業」という概念を定義して「ラッセル車」の属性としている。

「北斗星」とは、ある特定の列車に付けられた名前である。そこで「抽象物」の下位概念として「列車名」を定義した。一方で、「北斗星」を構成する車両は、具体物である。(2) に現れる「食堂車」は、ロール概念であり、さらに食事提供を行う車両であるため「食堂車」の属性として「食事提供」を付加している。

このように一つの概念を定義し、その概念にかかわる世界知識を表現するには、is-a リンクによる上位・下位概念の記述だけでなく、part-of リンクと attribute-of リンクを用いて機能のような内包的な特性まで記述することが重要となる。抽象物の導入、及び、基本概念とロール概念の区別により、こうしたより正確な知識表現が可能となる。

## 3 オントロジーを用いた推論

以上のオントロジーが、どのように推論・照応の計算に使用されるかを考察する。

### 3.1 依存型意味論

推論を行うために、Bekki [1] で提案された依存型意味論 (Dependent Type Semantics, DTS) を用いる。DTS は、依存型に基づく自然言語の意味論であり、含意関係の導出と照応の解消を「証明の構成」として統一的に扱える点に特徴がある。

DTS の意味表示において中心的な役割を果たすのは、いわゆる  $\Pi$  型と  $\Sigma$  型である (Martin-Löf [6])。 $\Pi$  型は論理の全称量化に、 $\Sigma$  は存在量化に対応する。照応関係を適切に扱うためには、(3a) に例示されるような動的な束縛関係を表現できる体系が必要となる。1 階述語論理のような従来の論理体系では、代名詞を変数とみなし、2 文を連言で結びつけると、(3b) のように変数  $x$  が自由変数にとどまってしまうという問題があった (いわゆる「ロバ照応」の問題)。

- (3) a. Someone entered. He smiled.  
 b.  $(\exists x \text{ entered } x) \wedge \text{smiled } x$   
 c.  $(\Sigma u : ((\Sigma x : \text{Entity}) \text{ entered } x) \text{ smiled } \pi_1 u)$

一方、依存型理論では、 $\Sigma$  型を用いて、(3c) のような意味表示を与えることができる。ここで、第 1 文は、 $(\Sigma x : \text{Entity}) \text{ entered } x$  という部分に対応し、この型 (= 命題) の証明項を  $\Sigma u$  によって取り出すことで、先行詞 (量化表現) と代名詞の関係を適切に表示することが可能となる。また、 $\Sigma$  型と  $\Pi$  型の推論規則により、先行詞の到達可能性を適切に表現することも可能である (詳しくは Bekki [1] を参照)。

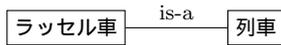
DTS では  $\Pi$  型と  $\Sigma$  型に対して以下の表記を用いる。

$$(x : A) \rightarrow B(x) \stackrel{\text{def}}{=} (\Pi x : A) B(x)$$

$$\left[ \begin{array}{l} x : \text{Entity} \\ B(x) \end{array} \right] \stackrel{\text{def}}{=} (\Sigma x : A) B(x)$$

この表記を用いて、前節で導入したオントロジーの 3 種類の意味リンクは、それぞれ以下のように DTS に翻訳される。

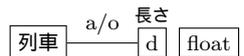
#### 1. is-a リンク

ラッセル車は列車である   
 $(x : \text{Entity}) \rightarrow (\text{ラッセル車 } (x) \rightarrow \text{列車 } (x))$

#### 2. part-of リンク

ラッセル車は除雪装置をもつ   
 $(x : \text{Entity}) \rightarrow (\text{ラッセル車 } (x) \rightarrow \left[ \begin{array}{l} y : \text{Entity} \\ \text{除雪装置 } (y, x) \end{array} \right])$

#### 3. attribute-of リンク

列車は float 型の長さ  $d$  をもつ   
 $(x : \text{Entity}) \rightarrow (\text{列車 } (x) \rightarrow \left[ \begin{array}{l} d : \text{float} \\ \text{length}(x) = d \end{array} \right])$

### 3.2 Bridging

Bridging は、談話表示理論 (DRT, Kamp [3]) を代表とする従来の形式意味論の枠組みでは扱うことが難しいことが知られている (Krahmer and Piwek [5])。DRT では、照応を解消するプロセスは、照応表現の指示対象を先行談話の談話指示物とマッチングする束縛 (binding) という操作と、適切な先行詞がない場合に働く accommodation という操作の二つに分類される。しかし、Bridging はこのどちらのケースでもなく、談話で明示的に与えられた情報と世界知識を組み合わせることで先行詞を推論するという点に特徴がある。

DTS では、Bridging を含め照応の先行詞を探索するプロセスが、含意関係の導出と同じく、証明の探索プロセスとして定式化されており、このため、世界知識との相互作用を体系内で統一的に扱うことが可能である。DTS では、照応を扱う演算子として @ 演算子が用意されている。(2) の談話を DTS に翻訳したものを図 3 に示す。(2) の第 1 文と第 2 文の意味表示は、動的連言  $M ; N$  によって結合される。 $M ; N$  は  $\Sigma$  型を用いて次のように定義される。

$$M ; N \stackrel{\text{def}}{=} \lambda c. \left[ \begin{array}{l} u : Mc \\ N(c, u) \end{array} \right]$$

ここで  $c$  は先行談話の情報をエンコードした証明項であり、第 2 文の意味表示  $N$  は、 $c$  と第 1 文の意味表示の証明項  $u$  のペア  $(c, u)$  を引数として渡される。この文脈情報の受け渡しにより、動的束縛を扱うことが可能となる。Bridging の引き金となる照応表現「食堂車」は、@ 演算子により表示される。第 2 文の意味表示において @ 演算子は、文脈  $c$  から、Entity 型の対象  $x$  とそれが食堂車であることの証明項を要求する項であり、その第一要素を射影  $\pi_1$  により取り出すことで、意図された照応関係が手に入る。@ の型は、型推論により特定することが可能であり、この場合、

$$@ : \left[ \begin{array}{l} \gamma \\ \left[ \begin{array}{l} u : \left[ \begin{array}{l} x : \text{Entity} \\ \text{Hokutosei}(x) \end{array} \right] \\ \text{EatIn}(\pi_1 u) \end{array} \right] \end{array} \right] \rightarrow \left[ \begin{array}{l} x : \text{Entity} \\ \text{BuffetCar}(x) \end{array} \right]$$

となる (先行文脈  $c$  の型を  $\gamma$  としている)。この型をもつ証明項を実際に構成し、@ と置き換えることが、照

$$\underbrace{\lambda c. \left[ \begin{array}{l} u: \left[ \begin{array}{l} x: \text{Entity} \\ \text{Hokutosei}(x) \end{array} \right] \\ \text{EatIn}(\pi_1 u) \end{array} \right]}_{\text{第 1 文の意味表示}} ; \underbrace{\lambda c. \text{NotOpen}(\pi_1(@c : \left[ \begin{array}{l} x: \text{Entity} \\ \text{BuffetCar}(x) \end{array} \right]))}_{\text{第 2 文の意味表示}} = \lambda c. \left[ \begin{array}{l} v: \left[ \begin{array}{l} u: \left[ \begin{array}{l} x: \text{Entity} \\ \text{Hokutosei}(x) \end{array} \right] \\ \text{EatIn}(\pi_1 u) \end{array} \right] \\ \text{NotOpen}(\pi_1(@c, v) : \left[ \begin{array}{l} x: \text{Entity} \\ \text{BuffetCar}(x) \end{array} \right]) \end{array} \right]$$

図 3: (2) の DTS による意味表示

応の解消プロセスに対応する。「北斗星には食堂車がある」という part-of リンクによって記述されている知識は、

$$w : (x : \text{Entity}) \rightarrow (\text{Hokutosei}(x) \rightarrow \left[ \begin{array}{l} y: \text{Entity} \\ \text{BuffetCar}(y, x) \end{array} \right])$$

と翻訳される。この知識を用いて問題の証明項を構成し、@に代入すると、最終的に (2) の談話の意味表示として次が得られる。

$$\left[ \begin{array}{l} v: \left[ \begin{array}{l} u: \left[ \begin{array}{l} x: \text{Entity} \\ \text{Hokutosei}(x) \end{array} \right] \\ \text{EatIn}(\pi_1 u) \end{array} \right] \\ \text{NotOpen}(\pi_1(w(\pi_1(\pi_1 v))(\pi_2(\pi_1 v)))) \end{array} \right]$$

## 4 考察

3種の意味リンクをもつオントロジーを構築し、それを基に DTS の意味表示を与えることによって、形式的に表現された世界知識を用いて自然言語の推論と照応の解決を行うことができることを示した。他に、鉄道オントロジーが関与する推論の興味深い例として、「方向」のような空間的概念や時間にかかわる推論が挙げられる。例えば、以下は列車の進行方向が関与する例である。

- (4) (全 17 両編成の列車で、1~11 号車が新青森行き、12~17 号車が秋田行きであるという掲示を見ながら) この列車の前から 6 両目は、秋田行きですか?

この質問に答えるためには、「下りの東北新幹線では、17 号車が先頭であり、かつ、号車番号は昇順である」という知識が必要である。また、「上り」「下り」は、時間経過によって変化する(折り返す)こともため、時間を扱う必要がある。また、「前から 6 両目」という言葉には、「進行方向先頭から数えて 6 両目」という隠れた知識が存在する。「前」「後」のような空間的概念は、視点(進行方向)を隠れたパラメータとして要求し、これらの知識を表現するためには、空間と時間にかかわる概念に関して、現時点でのオントロジーを拡張する必要がある。

## 5 まとめ

本論文では、溝口ら [8] が提案したオントロジーに基づいて、従来の上位・下位概念にかかわる知識だけでなく、ルール概念を含むより精緻な概念間の関係の形式的表現を与え、それを自然言語推論に適用することを試みた。オントロジーにより形式的に表現された知識は、Bekki [1] で提案された DTS を用いて意味表示に翻訳され、推論と Bridging を含む照応解決を証明構成プロセスとして統一的に扱うことができることを示した。時間・空間が関係する複雑な推論に必要なオントロジーを構築することは今後の課題である。また、構築したオントロジーから自動的に DTS の意味表示に翻訳し、計算機による自動推論を実現する方法についても今後考察していく予定である。

## 参考文献

- [1] Bekki, Daisuke. "Representing Anaphora with Dependent Types", LACL2014, LNCS 8535, pp.14-29, Springer. (2014)
- [2] Herbert H.Clark. "Bridging". In *Theoretical issues in natural language processing*, pp. 169-174. (1975)
- [3] H.Kamp and U.Reyle. *From Discourse to Logic*. Kluwer. (1993)
- [4] 乾健太郎「事態オントロジー:言語に基づく推論のためのコトに関する基本知識」言語処理学会第 13 回年次大会ワークショップ「言語的オントロジーの構築・連携・利用」論文集, pp. 27-30. (2007)
- [5] E. Krahmer and P. Piwek. Presupposition projection as proof construction. In: Bunt, H., Muskens, R. (eds.) *Computing Meanings*, Kluwer. (1999)
- [6] Martin-Löf, P. *Intuitionistic Type Thoery*, Bibliopolis. (1984)
- [7] 溝口理一郎『オントロジー工学』オーム社. (2005)
- [8] 溝口理一郎他『オントロジー構築入門』オーム社. (2006)
- [9] 周俊, 国府裕子, 古崎晃司, 溝口理一郎. 「OWL を用いたルール概念表現モデルに関する考察」人工知能学会第 22 回全国大会. (2008)