

オントロジーを用いた型理論的な語彙意味論の試み

Combining Type-theoretical Lexical Semantics with Ontology

中村 絢子*¹ 峯島 宏次*^{1*3} 戸次 大介*^{1*2*3}
Ayako Nakamura Koji Mineshima Daisuke Bekki

*¹お茶の水女子大学
Ohanomizu University

*²国立情報学研究所
National Institute of Informatics

*³独立行政法人科学技術振興機構, CREST
CREST, Japan Science and Technology Agency

In this paper, we combine a framework of Dependent Type Semantics (DTS), a framework of natural language semantics based on dependent type theory, with a formal ontology to deal with lexical and world knowledge. We focus on lexical semantics for predicates that are sensitive to the part-whole structures of entities, such as “normal” and “malfunctioning”. We provide some illustrative examples of inferences involving lexical and world knowledge concerning these predicates, and prove them using the inference system of DTS combined with formal ontology.

1. はじめに

自然言語の形式意味論の中心的な課題のひとつは、文間の含意関係 (entailment) に体系的な説明を与えることである。自然言語処理の文脈においても、含意関係の認識は、意味処理において重要な位置を占めるタスクとして注目をあつめている。

これまで形式意味論は、量化や否定のような論理表現、また、複数性、テンス、アスペクト、モダリティといった機能表現に着目し、文の統語構造に即して計算される合成的意味 (compositional meaning) から体系的に生じる含意関係を明らかにしてきた。形式意味論の主たる説明対象は、自然言語の多様な含意関係のうち、こうした論理的・文法的特徴をもつ推論にあったといえる。

他方で、計算言語学では実世界テキストを対象とする大規模な意味解析も実現されつつあり (Bos [3])、パーザと合成的意味論を組み合わせた深い意味処理の研究が重要性を増している。こうした計算言語学のアプローチと形式意味論のアプローチを組み合わせることで、含意関係認識の研究のさらなる進展が期待される。

形式意味論の成果に基づいてより広範囲の推論を扱うさいに大きな問題となるのは、われわれが自然に行う推論の多くが、多様な世界知識・語彙知識に基づくものであるという点である。中村ら [14] では、ドメインを限定して構築された形式的オントロジーによって世界知識を記述し、それを自然言語意味論の枠組みの一つである依存型意味論 (DTS, Bekki[2]) と組み合わせることで、世界知識を要求する単調推論 (monotonicity inference) やいわゆる橋渡し推論 (bridging inference) を扱う手法について提案を行った。本稿では、この研究をふまえて、「故障している」「正常である」のような述語の語彙の意味に着目して、オントロジーとDTSを組み合わせた推論システムにおいて世界知識と語彙知識を伴う推論を扱う方法について述べる。特に、溝口ら [12] で提案されたオントロジー構築ツール「法造」を用いて構築された鉄道オントロジー (中村ら [14]) に基づいて、構造物の部分全体関係に依存した推論をDTSの枠組みで扱う方法を提案する。

従来の形式意味論は、文の意味論の枠組みとして発展し、さらに談話表示理論 (DRT, Kamp [5]) などの出現によって、80年代以降は談話の意味論 (discourse semantics) としても大き

な発展を遂げているが、これを語彙意味論 (lexical semantics) と接合する試みははまだ十分になされていない (cf. Asher and Pustejovsky [1])。本稿は、DTSの型理論的な枠組みに世界知識・語彙知識に基づく推論を扱うメカニズムを結合することで、形式意味論と語彙意味論とを組み合わせた研究への貢献を目指すものである。

2. 含意関係認識システムとオントロジー

この節では、DTSと形式的オントロジーを中心に、自然言語の推論関係を判定するシステムの概要を示す。

2.1 推論と世界知識

自然言語の文の間に成立する含意関係を認識するには、しばしば前提には明示されない世界知識を把握する必要がある。例えば、(1) は、「ラッセル車は除雪作業を行う列車である」という知識を用いることによって、前提から結論が導かれる単調推論の例である。

- (1) 前提：ラッセル車が運行される。
結論：除雪作業を行う列車が運行される。

また、照応の解決にもしばしば世界知識が必要となる。以下は中村ら [14] で考察した橋渡し推論の例である。

- (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。

下線部の名詞句間の照応関係を確立するには、「北斗星には食堂車がある」という世界知識を用いる必要がある。

こうした世界知識を要する含意関係の導出や照応の解決を行うためには、1) 世界知識を形式的に記述し、2) それを明示的に与えられた前提・結論の意味表示と組み合わせることで推論を行う必要がある。本研究では、世界知識を記述するための枠組みとして人工言語ベースの形式的オントロジー (溝口ら [12]) を用いる。また、意味表示を扱う体系としては依存型意味論 (DTS) を採用する。

本研究が想定する自然言語の含意関係認識システムの全体像を図1に示す。まず入力となる推論の前提・結論を表す文は、CCGパーザの処理と意味合成を経て、DTSの意味表示への変換される*¹。一方、オントロジーによって記述された世界知

連絡先: 中村 絢子, お茶の水女子大学大学院人間文化創成科学研究科戸次研究室, nakamura.ayako@is.ocha.ac.jp

*¹ 組み合わせ範疇文法 (Combinatory Categorical Grammar, CCG) については, Steedman [9]、及び、戸次 [11] を参照。

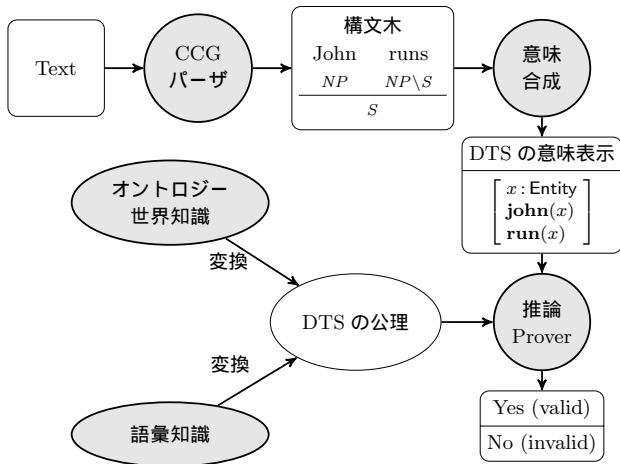


図 1: オントロジーを用いた含意関係認識システムの全体像

識と語彙知識は、DTS の公理へと変換され、文の意味表示と合わせて推論に使用される。

以下ではまず DTS について概観し、次に形式的オントロジーによる世界知識の記述について述べる。

2.2 依存型意味論 (DTS)

依存型意味論 (Dependent Type Semantics, DTS) は、Bekki[2] で提案された自然言語の形式意味論の一つであり、未指定項 (underspecified term) によって拡張された依存型理論に基づいて、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 では、自然言語文の意味表示を依存型理論 [7] の Π 型と Σ 型を用いて統一的に記述する。 Π 型は論理の全称量化に、 Σ 型は存在量化に対応する。 $(\Pi x : A) B(x)$ という Π 型をもつ項は、型 A をもつ任意の対象 a に対して、 fa が型 $B(a)$ をもつような関数 f である。 x が B に自由変数として出現しないとき、 $(\Pi x : A) B(x)$ は、含意 $A \rightarrow B$ に等しい。

$(\Sigma x : A) B(x)$ という Σ 型をもつ項は、型 A の対象 a と型 $B(a)$ の対象 b との組 (a, b) である。 x が B に自由変数として出現しないとき、 $(\Sigma x : A) B(x)$ は、連言 $A \wedge B$ に等しい。 Σ 型をもつ組には、射影関数 π_1 と π_2 が定義され、 $\pi_1(a, b) = a$, $\pi_2(a, b) = b$ という規則が与えられている。

DTS では Π 型と Σ 型を以下のように表記する。

	Π 型	Σ 型
標準的表記	$(\Pi x : A) B(x)$	$(\Sigma x : A) B(x)$
DTS の表記	$(x : A) \rightarrow B(x)$	$\left[\begin{array}{l} x : A \\ B(x) \end{array} \right]$
$x \notin fv(B)$ のとき	$A \rightarrow B$	$\left[\begin{array}{l} A \\ B \end{array} \right]$

また、 Σ 型に関しては以下の略記法を用いる。

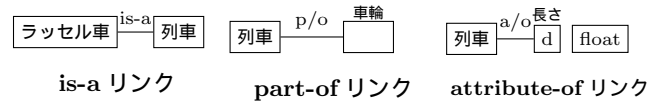
$$\left[\begin{array}{l} x : A \\ B \\ C \end{array} \right] \stackrel{\text{def}}{=} \left[\begin{array}{l} x : A \\ \left[\begin{array}{l} B \\ C \end{array} \right] \end{array} \right]$$

Π 型と Σ 型の導入規則と除去規則を図 2 に示す。

2.3 オントロジーの構築

溝口ら [12] で提案されたオントロジーの概要について述べる。形式的オントロジーは、概念と概念間の関係を記述する体系である。オントロジーが記述する概念として、基本概念とロール概念の 2 種類の概念がある。基本概念とは、「列車」「気動車」「電車」「人間」のようにコンテキストは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を指す。一方、ロール概念とはコンテキストに依存して決定される役割を表す概念を指す。例えば、「乗客」「操縦者」等は、「列車」というコンテキストの中の役割を示す概念である。また、「扉」「動力」「重さ」「長さ」等は、「列車」を参照して規定される概念である。

概念間の関係を記述するために、3 種類の意味リンクが用意されている。それぞれの例を以下に示す。



例えば、基本概念間の関係を述べる「ラッセル車は列車である」という知識は、is-a リンクを用いて記述される。ここで、「列車」は上位概念であり、「ラッセル車」は下位概念である。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。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ロール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。これらの意味リンクによって記述された知識を DTS の公理に変換することで 2.1 節で述べた推論の導出が可能となる [14]。それぞれの意味リンクから DTS の公理への変換方法は、4.1 節で与える。

3. 述語の分類：部分述語と全体述語

自然言語の推論の導出では、2.1 節で述べたような世界知識だけでなく、個々の語彙の意味に関する知識 (語彙知識) も重要な役割を果たす。この節では、形容詞的述語の語彙知識に注目する。dirty と clean のような反義語のペアは、部分と全体の関係に依存した推論を可能にすることが知られている (Yoon[10])。

- (3) a. Are the toys dirty?
- b. Are the toys clean?

例えば、(3a) では、the toys が指すおもちゃの集まりのうち、その一部が汚れていれば yes と答えることができるだろうが、(3b) の場合、おもちゃの全体がきれいでないといえれば、dirty を部分述語 (partial predicate)、clean を全体述語 (total predicate) と呼ぶ。

日本語の部分述語と全体述語の代表例を図 3 に示す。

全体述語	正常だ、安全だ、完成している
部分述語	閉まっている、乾いている、健康だ
部分述語	故障している / 異常がある、危険だ、未完成だ
部分述語	開いている、ぬれている、病気だ

図 3: 部分述語と全体述語の例

例えば、「開いている」と「閉まっている」というペアでは、ドアが少しでも開いていれば「開いている」と言えるだろうが、

$$\frac{\overline{x:A} \quad \dots \quad M:B}{\lambda x.M:(x:A) \rightarrow B} \text{ (III)} \quad \frac{M:(x:A) \rightarrow B \quad N:A}{MN:B[N/x]} \text{ (II E)} \quad \frac{M:\begin{bmatrix} x:A \\ B \end{bmatrix}}{\pi_1(M):A} \text{ (\Sigma E)} \quad \frac{M:\begin{bmatrix} x:A \\ B \end{bmatrix}}{\pi_2(M):B[\pi_1(M)/x]} \text{ (\Sigma E)}$$

図 2: DTS の Π 型と Σ 型の導入規則 (I 規則) と除去規則 (E 規則)

ドアが完全に閉まっていなければ「閉まっている」とは判断しにくい。この区別によれば、「正常だ」という述語は全体述語、「故障している / 異常がある」は部分述語に分類される。

部分述語と全体述語を区別するテストとして、「ほぼ」「少し」のような限定表現を付加するテストがある*2。「正常である」のような全体述語は「ほぼ」と自然に共起するが、「故障している」のような部分述語は、特別な文脈を用意しないかぎり「ほぼ」と共起しにくい。

- (4) a. この列車は、ほぼ正常である 全体述語
 b. ?この列車は、ほぼ故障している 部分述語

一方、限定表現「少し」の場合、全体述語とは共起しにくい、部分述語とは自然に共起する。

- (5) a. ?この列車は、少し正常である 全体述語
 b. この列車は、少し故障している 部分述語

また、部分述語と全体述語はどちらも、肯定形 (P) がその反義語の否定形 ($\neg N$) を含意するだけでなく、否定形 ($\neg P$) がその反義語の肯定形 (N) を含意するという特性をもつ。

- (6) a. 正常だ \Rightarrow 故障していない $P \Rightarrow \neg N$
 b. 故障している \Rightarrow 正常でない $P \Rightarrow \neg N$
 (7) a. 正常でない \Rightarrow 故障している $\neg P \Rightarrow N$
 b. 故障していない \Rightarrow 正常である $\neg P \Rightarrow N$

通常の形容詞的述語の場合、肯定形 (P) はその反義語の否定形 ($\neg N$) を含意するが、否定形 ($\neg P$) はその反義語の肯定形 (N) を含意しない。

- (8) a. 重い \Rightarrow 軽い $P \Rightarrow \neg N$
 b. 軽い \Rightarrow 重い $P \Rightarrow \neg N$
 (9) a. 重くない \nRightarrow 軽い $\neg P \nRightarrow N$
 b. 軽い \nRightarrow 重い $\neg P \nRightarrow N$

適切な含意関係を導出するためには、こうした反義語との含意関係の相違を語彙知識として記述することも重要となる。

4. オントロジーから DTS への変換

オントロジーが記述する世界知識と語彙知識を DTS の公理に変換する方法を示す。

4.1 世界知識の変換

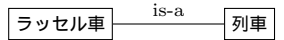
まず、オントロジーにおける「列車」「ラッセル車」「車輪」「食堂車」のような概念は、DTS の 1 項述語に対応付けられる。また、「 x は y の部分である」という部分全体関係を表す 2 項述語を $x \leq y$ と表す。 $x \leq y$ は、半順序関係とする。

中村ら [14] の提案に基づいて、3 種類の意味リンクは以下のような仕方で DTS の公理に変換される*3。

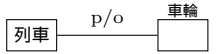
*2 英語については、Rotstein and Winter [8]、Kennedy and McNally [6] を参照。日本語の形容詞とその限定表現の共起関係については、西尾 [13] で詳しく論じられている。

*3 E は個体 (Entity) の型を表す。 \rightarrow は右結合であり、 $(x:E) \rightarrow$ ラッセル車 $x \rightarrow$ 列車 x は、 $(x:E) \rightarrow (\text{ラッセル車 } x \rightarrow \text{列車 } x)$ を意味する。

(10) is-a リンク

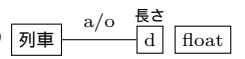
ラッセル車は列車である 
 DTS の表現 $(x:E) \rightarrow \text{ラッセル車 } x \rightarrow \text{列車 } x$

(11) part-of リンク

列車は部分として車輪をもつ 

DTS の表現 $(x:E) \rightarrow \text{列車 } x \rightarrow \begin{bmatrix} y:E \\ \text{車輪}(y) \\ y \leq x \end{bmatrix}$
 $(x:E) \rightarrow \text{車輪 } x \rightarrow \begin{bmatrix} y:E \\ \text{列車}(y) \\ x \leq y \end{bmatrix}$

(12) attribute-of リンク

列車は float 型の長さ d をもつ 
 DTS の表現 $(x:E) \rightarrow \text{列車 } x \rightarrow \begin{bmatrix} d:\text{float} \\ \text{length}(x) = d \end{bmatrix}$

本稿では、part-of リンクの翻訳としては、全体からその部分の存在を導く公理と部分から全体の存在を導く公理の 2 つを用いている。

4.2 語彙知識の変換

部分述語と全体述語にかかわる語彙知識はそれぞれ次のような DTS の公理として記述される。

(13) 全体述語 (正常だ)

$(x:E) \rightarrow (y:E) \rightarrow \text{正常 } x \rightarrow y \leq x \rightarrow \text{正常 } y$

(14) 部分述語 (故障する)

$(x:E) \rightarrow (y:E) \rightarrow \text{故障 } x \rightarrow x \leq y \rightarrow \text{故障 } y$

全体述語は、複合物全体について成り立つならその部分についても成り立つ。一方、部分述語は部分について成り立つならそれを含む全体についても成り立つ。反義語間の関係についても同様の公理として記述可能であるが、ここでは省略する。

5. オントロジーを用いた DTS による推論

世界知識と語彙知識を用いた推論がどのように行われるかを考察する。まず、次の推論を取り上げよう。

- (15) 前提: 列車は正常である
 結論: 車輪は正常である

前提と結論はそれぞれ (16) と (17) に示される DTS の意味表示に対応する*4。ここでは、前提文の「列車」が指示する対象を個体の型 E をもつ変数 a によって表している。

*4 結論文の「車輪」は、中村ら [14] で論じた橋渡し推論により、「列車の車輪」と解釈する必要がある。単純化のため、ここでは照応・前提がすでに解消された後の意味表示を想定している。

$$\begin{array}{c}
\frac{q : \left[\begin{array}{l} \text{車輪 } x \\ x \preceq a \end{array} \right]}{\pi_2 q : x \preceq a} \quad (\Sigma E) \quad (2) \quad \frac{p : \left[\begin{array}{l} \text{列車 } a \\ \text{正常 } a \end{array} \right]}{\pi_2 p : \text{正常 } a} \quad (\Sigma E) \quad (1) \quad \frac{a : E \quad f : (x : E) \rightarrow (y : E) \rightarrow \text{正常 } a \rightarrow x \preceq a \rightarrow \text{正常 } x}{fa : (y : E) \rightarrow \text{正常 } a \rightarrow y \preceq a \rightarrow \text{正常 } y} \quad (\Pi E) \\
\frac{\frac{fa x : \text{正常 } a \rightarrow x \preceq a \rightarrow \text{正常 } x}{fa x(\pi_2 p) : x \preceq a \rightarrow \text{正常 } x} \quad (\Pi E)}{fa x(\pi_2 p)(\pi_2 q) : \text{正常 } x} \quad (\Pi E) \\
\frac{\frac{\lambda x. fa x(\pi_2 p)(\pi_2 q) : \left[\begin{array}{l} \text{車輪 } x \\ x \preceq a \end{array} \right] \rightarrow \text{正常 } x}{\lambda x \lambda y. fa x(\pi_2 p)(\pi_2 q) : (x : E) \rightarrow \left[\begin{array}{l} \text{車輪 } x \\ x \preceq a \end{array} \right] \rightarrow \text{正常 } x} \quad (\Pi I), (1)}{\lambda x \lambda y. fa x(\pi_2 p)(\pi_2 q) : (x : E) \rightarrow \left[\begin{array}{l} \text{車輪 } x \\ x \preceq a \end{array} \right] \rightarrow \text{正常 } x} \quad (\Pi I), (2)
\end{array}$$

図 4: (15) の推論 (「列車は正常である」 \Rightarrow 「車輪は正常である」) の導出例。(II) に伴う *type* 条件の導出は省略している。

$$(16) \quad \left[\begin{array}{l} \text{列車 } a \\ \text{正常 } a \end{array} \right] \quad (17) \quad (x : E) \rightarrow \left[\begin{array}{l} \text{車輪 } x \\ x \preceq a \end{array} \right] \rightarrow \text{正常 } x$$

「正常である」は全体述語であるから、(13) の公理が使用可能である (その証明項を f とする)。この他に、 $a : E$ と前提に対応する (16) の型をもつ証明項 (これを p とする) の存在を仮定し、結論 (17) の証明項を構成した証明木を図 4 に示す。

次に部分述語を含む推論の例を検討する。

$$(18) \quad \frac{\text{前提: 車輪が故障している}}{\text{結論: 列車が故障している}}$$

前提と結論はそれぞれ、(19) と (20) の意味表示に翻訳される。

$$(19) \quad \left[\begin{array}{l} u : \left[\begin{array}{l} x : E \\ \text{車輪 } x \end{array} \right] \\ \text{故障 } \pi_1 u \end{array} \right] \quad (20) \quad \left[\begin{array}{l} u : \left[\begin{array}{l} x : E \\ \text{列車 } x \end{array} \right] \\ \text{故障 } \pi_1 u \end{array} \right]$$

部分述語である「故障している」には、(14) の公理が適用可能である。また、「列車は車輪の一部としてもつ」というオントロジーの part-of リンクによって記述された知識から次の公理が得られる。

$$(21) \quad (x : E) \rightarrow \text{車輪 } x \rightarrow \left[\begin{array}{l} y : E \\ \text{列車 } y \\ x \preceq y \end{array} \right]$$

前提 (19) の証明項の存在を仮定すると、結論の型 (20) の証明項を構成することができ、(18) の推論が導出可能である。

DTS による上記の証明は、Coq (Castéran and Bertot [4]) を用いて検証・自動化が可能である。Coq とは、依存型理論に基づく定理証明支援系であり、タクティクスと呼ばれる証明手続きを用いることによって、証明の検証が行われる。また、Ltac と呼ばれるタクティクスの記述言語を用いることで、証明探索の手続きを定義することが可能であり、定理証明器としての役割も果たしうる。上記の形の証明はいずれも、比較的単純なタクティクスの組み合わせにより自動化可能である。

6. おわりに

本論文では、中村ら [14] の鉄道オントロジーに基づき、形式的オントロジーで記述された世界知識と語彙知識を用いた推論を依存型意味論の枠組みで扱う方法について述べた。特に、鉄道オントロジーに関与する述語を部分述語・全体述語に分類し、その語彙知識を DTS の公理に変換することで、構造物の部分全体関係を参照する含意関係が導出可能であることを示した。

今後の課題として、オントロジーの部分全体関係のよりきめ細かな分類を行うことが挙げられる。例えば、(14) の部分述語の公理を用いると、以下の推論が導出される。

$$(22) \quad \frac{\text{前提: 列車の座席の一部に異常がある。}}{\text{結論: 列車に異常がある。}}$$

しかし、列車の座席のごく一部に異常があるからといって列車全体に異常があると判断することは難しい。「列車に異常がある」状況とは、列車の走行に支障をきたすような異常があることを指すことが多い。よって、列車の走行に直接関与しない座席の一部は、列車にとって必要不可欠な部品であるとは言えない。列車が走行するという目的に対して必要な部品とそうではない部品をオントロジー上で分け、推論できる範囲を制限する必要がある。また、定理証明支援系による自然言語推論の証明の検証と自動化についても、さらに研究を進めていく予定である。

参考文献

- [1] N. Asher and J. Pustejovsky. The metaphysics of words in context. *Journal of Logic, Language and Information*, 2000.
- [2] Daisuke Bekki. Representing anaphora with dependent types. In *Logical Aspects of Computational Linguistics*, pp. 14–29. Springer, 2014.
- [3] Johan Bos. Wide-coverage semantic analysis with boxer. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, pp. 277–286. Association for Computational Linguistics, 2008.
- [4] Pierre Castéran and Yves Bertot. *Interactive theorem proving and program development. Coq'Art: The Calculus of inductive constructions*. Springer Verlag, 2004.
- [5] Hans Kamp and Uwe Reyle. *From Discourse to Logic; Introduction to the Modeltheoretic Semantics of natural language*. Reidel, 1993.
- [6] Christopher Kennedy and Louise McNally. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, pp. 345–381, 2005.
- [7] Per Martin-Löf. Intuitionistic type theory. *Naples: Bibliopolis*, 1984.
- [8] Carmen Rotstein and Yoad Winter. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, Vol. 12, No. 3, pp. 259–288, 2004.
- [9] M. Steedman. *The syntactic process*. MIT Press, 2000.
- [10] Youngeun Yoon. Total and partial predicates and the weak and strong interpretations. *Natural language semantics*, Vol. 4, No. 3, pp. 217–236, 1996.
- [11] 戸次大介. 日本語文法の形式理論 -活用体系・統語構造・意味合成. くろしお出版, 2010.
- [12] 溝口理一郎, 古崎晃司, 笹島宗彦, 來村徳信. オントロジー構築入門. オーム社, 2006.
- [13] 西尾寅弥. 形容詞の意味. 用法の記述的研究, 第 44 巻. 秀英出版, 1972.
- [14] 中村絢子, 峯島宏次, 戸次大介. オントロジーを用いた自然言語の推論に向けて. 言語処理学会第 21 回年次大会発表論文集, pp. 309–312, 2015.