

型理論に基づいた特定領域テキストからの動的な Taxonomy, Mereology 構成

A Type-Theoretic Dynamic Construction of a Taxonomy and Mereology from Texts of Specific Domains

緒方 典裕 Norihiro Ogata*

日本学術振興会特別研究員 Research Fellow of JSPS

Abstract. A dynamic construction method from texts of specific domains based on Discourse Representations expanded by a kind of constructive type theory which can treat expressions on types. For maintaining subtype relation, a non-monotonic system of subtype relation is introduced, and applied to constructions of taxonomy and mereology. In particular, we discuss what is an informative mereology based on type theory. This method is applied to a relatively small corpus of reports of Japan Stock Market and industrial goods, and a relatively efficient result is gained.

1 序

本稿では、シソーラスやオントロジーの中心的構造である、(i) *taxonomy* (is a kind of 関係)、(ii) *mereology* (*meronymy*) (is a part of 関係)¹の、特定領域テキストからの自動構成法を提案する。その体系は、(a) テキスト中の情報の拡散性、(b) テキスト中の情報の部分性、(b) 領域内概念の増加・変更可能性、という特性から、

- 構成される構造は選択されたテキストに対応して動的に変更可能でなければならない。
- 照応・省略解消などの談話処理により、テキスト内の情報を最大限に活用しなければならない。

という二つの点から「動的」である必要があり、本稿の提案する手法は、この二点において「動的」な構成法となっている。また、構成的型理論で拡張さ

れた談話表示に対して、型推論に基づいた情報抽出ルールによって、*taxonomy* や *mereology* に関する情報を抽出する。すなわち、型体系は、機能語の型スキーマ・型推論規則・型強制対応規則等からなり、*taxonomy*/*mereology* 構成は、型に関する問い合わせに対する、型推論による検索とみなされる。領域特定性から語の多義性によるノイズが少なく、機能語のみ型スキーマを与えるため未知語に対してロバストであり、型強制対応規則により、メトニミのノイズにも対応し、談話表示に基づくため照応・省略などの談話解析により資源が増大される。また、型割り当てには非単調論理を導入するために、領域内情報の増加や更新に対してもロバストである。型の情報抽出は型推論に基づくため、比較的少量のテキストからの構成が可能である²、型名自体が獲得されるという利点をもつ。*taxonomy* 構成の実装・実験

²同様な議論として、言語学的規則・語彙規則に基づいたテキストからのオントロジー獲得があるが (e.g., [4] の *external evidence/internal evidence, mention generator*, [5], [6])、規則自体の *validity* がわからないという不利な点がある。

*ogata @ is.s.u-tokyo.ac.jp

¹[2] 参照。

は株式ニュースに関して、mereology 構成については工業製品ニュースに関して行う。

$$\begin{aligned}
 1 \text{ という自然数} &\equiv (1(\vec{\lambda} y : type.(\overleftarrow{\lambda} x : y.x)(\text{自然数})) \\
 &\rightarrow_{\beta} (1(\overleftarrow{\lambda} x : \text{自然数}.x)) \\
 &\rightarrow_{\beta} 1
 \end{aligned}$$

2 理論的枠組み

構成的形理論 自然言語で、「パソコンなどハイテク機器」、「パソコンというハイテク機器」などの表現は、「パソコン」が「ハイテク機器」という型をもつ (パソコン : ハイテク機器と表す) という情報を直接表現しているが、型体系に基づいた文法であるカテゴリ文法 (Lambeck 計算) の枠内でも記述できない。そこで、次のように kind · type などの型のメタ概念を導入できる構成的形理論 ([1] の「 λ -cube」参照) 概念で拡張する。

($s \in \{type, prop, kind\}$, $* \in \{type, prop\}$), (A, B s.t., $A : s$, $B : s$)

$$\begin{array}{c}
 \frac{}{\vdash \langle \rangle : drs} \quad \frac{}{\Gamma \vdash x : A} \quad \frac{}{\vdash \Gamma : drs} \quad \frac{}{\Gamma \vdash * : kind} \\
 \frac{}{\Gamma \vdash A : s} \quad \frac{}{x \notin dom(\Gamma)} \\
 \frac{}{\vdash (\Gamma, x : A) : drs} \\
 \frac{}{\Gamma, x : A \vdash B : *} \\
 \frac{}{\Gamma \vdash (\overleftarrow{\lambda} x : A.B) : (\overleftarrow{\Pi} x : A.B)} \\
 \frac{}{x : A, \Gamma \vdash B : *} \\
 \frac{}{\Gamma \vdash (\overleftarrow{\lambda} x : A.B) : (\overleftarrow{\Pi} x : A.B)} \\
 \frac{}{\Gamma, x : A \vdash B : s} \quad \frac{}{x : A, \Gamma \vdash B : s} \\
 \frac{}{\Gamma \vdash (\overleftarrow{\Pi} x : A.B) : s} \quad \frac{}{\Gamma \vdash (\overleftarrow{\Pi} x : A.B) : s} \\
 \frac{}{\Gamma \vdash M : (\overleftarrow{\Pi} x : A.B)} \quad \frac{}{\Delta \vdash N : A} \\
 \frac{}{\Gamma \Delta \vdash MN : B[x := N]} \\
 \frac{}{\Delta \vdash N : A} \quad \frac{}{\Gamma \vdash M : (\overleftarrow{\Pi} x : A.B)} \\
 \frac{}{\Delta \Gamma \vdash NM : B[x := N]} \\
 \frac{}{\Gamma \vdash M : A} \quad \frac{}{\Gamma \vdash B : s} \quad \frac{}{A =_{\beta} B} \\
 \frac{}{\Gamma \Delta \vdash M : B}
 \end{array}$$

次の β 縮約をもつ。

$$\begin{aligned}
 (\overleftarrow{\lambda} x : A.Bx)(a) &\rightarrow_{\beta} Ba \quad (a)(\overleftarrow{\lambda} x : A.xB) \rightarrow_{\beta} aB \\
 (\overleftarrow{\Pi} x : A.Bx)(a) &\rightarrow_{\beta} Ba \quad (a)(\overleftarrow{\Pi} x : A.xB) \rightarrow_{\beta} aB
 \end{aligned}$$

この結果、次のように型を表す情報自体を対象言語中に表現できるようになる。

「一般的恒等関数」としての機能語

- (1) a. など, という, を中心とした $\equiv \vec{\lambda} y : type.(\overleftarrow{\lambda} x : y.x)$
 b. など, という, を中心とした $:(\overleftarrow{\Pi} y : type.(\overleftarrow{\Pi} x : y.y))$

「型一致」と述語

- (2) a. 平行である $\equiv (\overleftarrow{\lambda} x : \text{線}.\overleftarrow{\lambda} y : \text{線}.\text{平行}(y, x))$
 b. 平行である $:(\overleftarrow{\Pi} x : \text{線}.\overleftarrow{\Pi} y : \text{線}.prop)$
 c. l_1 が $(l_2$ に (平行である))

c が与えられると、 l_1, l_2 に関して型の情報を持たず、次のように T_1, T_2 という型変数を与えていても、「型一致」から型の情報が与えられる。

$$\begin{array}{c}
 l_1 : T_1 \\
 l_2 : T_2 \\
 \text{平行である} : (\overleftarrow{\Pi} x : \text{線}.\overleftarrow{\Pi} y : \text{線}.prop) \\
 (l_1(l_2 \text{ 平行である})) : prop \\
 T_1 := \text{線} \\
 T_2 := \text{線} \\
 (l_1(l_2 \text{ 平行である})) \rightarrow_{\beta} \text{平行}(l_1, l_2) \\
 \text{平行}(l_1, l_2) : prop \\
 l_1, l_2 : \text{線}
 \end{array}$$

3 非単調順序関係

1 節で述べたように、テキスト中の情報の拡散性、テキスト中の情報の部分性、領域内概念の増加・変更可能性、という特性から、テキストから抽出される情報は、部分的で不確実である。また、そのため、ユーザーが前提知識を持っていて、それを反映させたい場合もある。従って、構成される Taxonomy/Mereology は、非単調な体系となる。Taxonomy/Mereology は、ある種の順序と考えることができるので、非単調な順序体系が両者のメタ理論となる。提案する非単調部分順序体系は、*must*, *evidential*, *may* の三つの命題を区別する。*must* によって宣言された順序は、

安定した情報であり、棄却されることはなく、背景知識や主観を表現する場合に使われる。*evidential* によって宣言された順序は、ある証拠から得られた情報であり、*must* によって宣言された順序と矛盾しないかぎり棄却されず、テキスト中の情報を表すのに使われる。*may* によって宣言された順序は、不安定な情報であり、棄却する証拠が無い限り仮定してもよいもので、仮説や推測を表すのに用いられる。これらは、 \prec を順序とした場合、次のような公理によって定式化される。

- $must(x \prec y) \rightarrow x \prec y.$
- $evidential(x \prec y) \wedge \sim must(y \prec x) \rightarrow x \prec y.$
- $may(x \prec y) \wedge \sim evidential(y \prec x) \wedge \sim must(y \prec x) \rightarrow x \prec y.$

ここでの \sim は論理的否定ではなく、「現理論で証明できない」ことを表す (prolog の Negation-as-Failure にほぼ等しい)。従って、理論が変更され、証明される命題が変化すれば、 \sim を含む規則の適用可能性がそれに依存して変化することで、非単調性を具現している。

4 Taxonomy 構成

Taxonomy (is_a_kind_of 関係) は型の部分関係 \leq とみなすことができる。つまり、

$$T \leq T' \text{ iff } \forall x.x : T \rightarrow x : T'$$

という部分型の定義により、トークンの型割り当て情報から部分型体系が構成でき、その結果それを taxonomy とみなすことができる。しかし、ユーザーが前提知識として部分型関係を知っていてそれを反映させたい場合や、ある語彙的知識から部分型関係の不確定な知識が得られ、それを反映させたい場合があるため、部分型関係は 3 節で述べた非単調規則によって構成される。また、トークンの型割り当て情報から構成する場合でも、テキスト内の情報の部分性から複雑な部分関係は構成しにくい。したがって、次のように、 \forall ではなく、一般化量子子である *most*

により構成する。

$$(3) \text{ a. } most(T, T') \rightarrow evidential(T \leq T'),$$

$$\text{ b. } most(T, T') \leftrightarrow |tok(T) \cap tok(T')| > 0.8|tok(T)|$$

ただし、 $tok(T)$ は型 T をもつトークンの集合、 $|X|$ は集合 X の濃度である。

4.1 ノイズ除去

型論理的メトニミ除去 「ハイテクや自動車など主力株」における「ハイテク」「自動車」は、業種や製品という型を表しているというよりは株を表している。しかし、このテキストから型情報を抽出すると、「ハイテク」「自動車」が「主力株」という型を与えられてしまう。このような型の変換が起こっていることを型論理的メトニミ (type-logical metonymy) と呼ぶことにする。型論理的メトニミは、このように型割り当ての際のノイズとなる。本稿では、「ハイテク株など主力株」のような同じ型を抽出できる表現があり、しかも、「ハイテク」と「ハイテク株」のように接尾辞が省略されているということがテキストから証明できる場合、省略された表現を棄却する規則を与えて、排除している。

トークン・型混同の除去 「銀行株など金融株」、「さくら銀行など銀行株」という表現があれば、「金融株」という型のトークンに「銀行株」という型が情報として抽出されてしまう。これは、「証券株」という表現がある不定のトークンを表す表現となっていることから生じる。このようなレベルの混同を避けるために先に型「銀行株:金融株」とトークン「さくら銀行:銀行株」と同時に宣言したものは、トークンとして宣言した情報「銀行株:金融株」を棄却する規則を与える。その情報の欠如を補うために、その型を持つトークンすべて (e.g., さくら銀行) にその型を与える (e.g., さくら銀行:金融株)。

4.2 さらなる形態素からの構造化

型名からの構造化 ある型名のグループが共通の最長の接尾辞 M を共有していた場合、 M をそれらの上位型とし、その証人としてのトークンとして「 $\text{tok}(M)$ 」を与える。(例えば、「株」「セクター」などの型はこのようにして構成された。)

要素が共通の型談話指示対象の統合 共起情報から得られた型は基本的に変数であるので、要素が同じならば同一型とみなす。しかし、明示的に型名が選られる場合は、背景知識としてそのような型が存在するものとし、要素が同じでも統合しない。

要素の共通形態素からの構造化 共起情報から得られた型は基本的に変数であるので、それに明示的な型名を与えるために、要素が共通の最長の接尾辞を持つ場合、それを型名とした。(付録では「?」が附加されてある。)

5 Mereology 構成

本稿でいう「mereology」とは、直観的に、対象間の全体・部分関係、すなわち「 x が y の一部(分)である」という関係であるが、「一部(分)である」という概念自体が、「身体部位」や「部品」といった典型的な場合に限定せず、一般化した場合、曖昧である。そのため、mereology の定義を「情報のある」mereology とは何かという観点から議論し、次にその構成を述べる。

5.1 Mereology とは

純関係名辞との相違 「 x が y の一部(分)である」という情報は、テキスト中では「 y の x 」という名詞句で表現されることが多い。しかし、「山の上・車の横」などの表現は、次の例が示すとおり、曖昧である。

(4) a. 山の上で熊が出る。/車の横に傷がある。

b. 山の上に月が出る。/車の横に子供がいる。

(4a) でも (4b) でも「ある部分」をそれぞれ表しているには違いないが、(4a) は「山」や「車」に含まれる部分を表しているのに対して、(4b) は含まれない部分を表している。前者を **mereonym**、後者を **純関係名辞** と呼び、「 y の x 」の x を「関係対象」、 y を「関係先」と呼ぶことにすると、純関係名辞と mereonym の違いは関係対象が関係先に「含まれる」か否かという点にあることになる。「含まれる」という概念自体も文脈によって異なるため、「ある文脈 c において含まれる」という概念を \sqsubseteq_c と表記することにする。従って、ある語 μ が文脈 c において mereonym であることの条件は次のように定式化できる。

(5) $x : T_1, \mu : (\prod z : T_1.T_2), \text{の} : (\prod z : T_1.(\prod w : (\prod u : T_1.T_2).T_2)), \text{の} \equiv \lambda v : T_1. \lambda f : (\prod z : T_1.T_2). \lambda v : T_1.vf^3$, とすると、
 x の $\mu \equiv y \rightarrow x \sqsubseteq_c y$ ならば μ は文脈 c において mereonym である。

情報をもたない Mereology Mereology と概念構造の関係を考えて場合、概念構造上ではほとんど情報をもたないような mereology がある。例えば、「物質名詞」や *mass term* といわれる表現が表す対象は、部分が特にカテゴリー化されていることはない。「人体」の部分は「手・足・頭」などのカテゴリー化が行われ、「人体」自体ではないが、「水」の部分は「水」である。すなわち、情報を持たない mereology とは次のような型に関して単調性が成立するものといえる。

(6) $\forall x y.T : \text{type} \wedge x : T \wedge x \sqsubseteq_c y \rightarrow y : T$

従って、情報をもつ mereology とは (6) を満たさない関係 \sqsubseteq_c であるといえる。例えば、「水」でも化学のある文脈 d においては、次のような情報のある mereology が存在する。

³ μ の型割り当てはある関数としてのもので、「の」の型割り当ては、次の関数適用 *app* とほぼ同じである。

$\text{app} : (X \rightarrow Y) \rightarrow (X \rightarrow Y)$,

$\text{app} \equiv \lambda f : X \rightarrow Y. \lambda x : X. f(x)$.

(7) x : 水 \wedge y : 酸素 \wedge z : 水素 \wedge $y \sqsubseteq_d x \wedge z \sqsubseteq_d x$

これは、「水は酸素と水素から構成される」といった化学的話題に関する文から抽出できる。

抽象的対象の Mereology - イベント類 イベントのような抽象的対象も情報のある mereology ものと情報の無い mereology がある。aspectuality の観点からイベント類を、[7]に従って、状態・activity・accomplishment・achievement に分類した場合、状態・activity に関しては、(6) が成り立ち、mereology は情報が無い。一方、accomplishment・achievement に関しては、[3]のように、少なくとも「始まり・終わり・culmination(成就)・結果状態」という時間的 mereology のいくつかをもち、情報をもつ。例えば、「東京・NY間の飛行」というイベントには「離陸・飛行中・着陸」という情報のある mereology がある。さらに、儀式・行事・公演のような特殊にカテゴリー化された段階 (stage) をもつイベントもある。⁴「開幕・開会式・おいろなおいし」等である。イベント類の場合、個体や地理的対象が「空間的整合性 (spatial integrity)」をもつものに対して⁵、「時間的整合性 (temporal integrity)」をもった mereology であるといえる。また、個体を生成から消滅までのイベントとしてとらえると、「幼年期、青年期、大人、成年」など各段階をあらわす概念と個体の間には時間的整合性をもつ mereology があることになる。

抽象的対象の Mereology - 情報内容・属性類 情報内容や属性のような抽象的対象にも情報のある mereology と情報の無い mereology がある。これらも量的にとらえると、(6) を満たし、情報が無い mereology となる。例えば、次の例文で、

(8) a. この画像ファイル (50KB) の {20KB ぶん/大部分/一部} をダウンロードできた。

b. この鉄骨の強度は、以前の半分である。

⁴[2]も参照。

⁵個体に対しても「成分・材料」などは空間的整合性を持たない mereology である。

a における「20KB ぶん/大部分/一部」と「画像ファイル (50KB)」はともに同じ情報量というような型としてとらえられ、また、b における「この鉄骨の強度」と「半分」も同じ「程度 (degree)」というような型としてとらえられ、(6) を満たす。一方、両者も質的にとらえると、情報のある mereology をもつことがある。情報内容に関していうと、「話の {骨子・要点・要旨・補足・テーマ・余談}」のように整合性をもたないものや、「小説の {冒頭・クライマックス}」のように整合性をもつものがある。属性類に関していうと、(9a) のような属性記述文は、(9b-9c) のような部分関係を表す文に書き換えることができるため、

- (9) a. 塩酸には発煙性や刺激臭がある。
b. 塩酸のもつ性質には発煙性がある。
c. 塩酸のもっている {もの・特徴} には刺激臭がある。

「塩酸」のような種を属性の塊とみなすことで、種と性質・属性の間に情報のある mereology があるとみなすことが出来る。同様に「能力・機能」なども種との間の、整合性をもたない、情報のある mereology とみなすことができる。このようなとらえ方は、例えば、「人」という語が文脈によって「人体」や「人格」を表すことから支持される。また、「道具・プログラム・手法・アルゴリズム」などは、「どのような機能や能力 {があるか・もっているか・から構成されるか}」といったことが recognition の基準となるため、その一部の機能・能力を部分とみなすことは自然であるといえる。

Mereology のまとめ 以上述べた mereology は、(i) 関係先の型、(ii) 情報性、(iii) 整合性、という観点から次ぎのように整理できる。

● 整合的部分 (integrated part)

- 空間整合的部分: 部位名 (手、鼻、頭、胴など)、部品名や、相対的部分 (先、後ろ、本体など)、地理的・空間的位置 (北部、南部、北、上流、下流など) など

- 時間整合的部分: 個体段階名 (雛、親鳥、大人、子供など)、イベント段階名 (開会式、閉幕式、始まり、未到着など)、時間仮定的部分 (冒頭、クライマックス、エンディング、第2章など)

● 非整合的部分 (unintegrated part)

- 情報のある非整合的部分: 機能 (検索機能、オートコレクト機能など)、能力 (腕力、思考力、消化力など)、性質 (揮発性、発煙性など)、属性 (毒、刺激臭、酸味など)、材料、原料、成分、添加物など
- 情報のない非整合的部分: 等質的対象の mereology (物質名詞など)、数量の mereology

5.2 テキスト中の mereology 情報

テキスト中からは、以下のような点から内在されている mereology 情報が抽出できる。

述語に基づいた抽出 テキスト中の、項 x, y をもつ述語 R に関して、次のような形で mereology 情報抽出規則が定式化できる。

$$(10) R[x, y] \rightarrow evidential(x \sqsubseteq_c y)$$

R に関しては、一般に次のような述語は mereology 情報をもつ。

$$(11) y \text{ に } x \text{ がある, } x \text{ が } y \text{ を含む, } x \text{ が } y \text{ をもつ, } x \text{ が } y \text{ から (で) 構成される, } x \text{ が } y \text{ からなる}$$

また、定義そのものである次のような文も mereology 情報をもつ。⁶

$$(12) x \text{ が } y \text{ の (一) 部位である, } x \text{ が } y \text{ の (一) 部品である, } x \text{ が } y \text{ の付属品である, } x \text{ が } y \text{ の付属物である, } x \text{ が } y \text{ の (一) シーンである, } x \text{ が } y \text{ の (一) 機能である}$$

また、領域特定の mereology 情報をもつ述語群がある。例えば、次のものは工業製品ニュース中に現れやすい述語である。

⁶ 「一部」や「一部分」という語は曖昧であり、ノイズを導入する結果となりやすい。

$$(13) x \text{ が } y \text{ を搭載している, } x \text{ に } y \text{ を組み合わせる, } x \text{ に } y \text{ を組み込む, } x \text{ の } z \text{ に } y \text{ を採用する, } x \text{ が } y \text{ を備える, } x \text{ が } y \text{ を内蔵している, } x \text{ に } y \text{ が付属している, } x \text{ に } y \text{ がバンドルされている,}$$

また、「 y に x が連載中」、「 y に x が収録されている」などは情報媒体とそこに含まれる各情報、「 x が y を新設する」、「 x が y を設ける」、「 x が y を統括する」などは組織とその組織に含まれる下位組織といった mereology 情報を内在する。また、能力や機能に関する mereology については

$$(14) x \text{ が } y \text{ ができる, } x \text{ が } y \text{ 可能である, } x \text{ を } y \text{ として使う}$$

といった構文から抽出できる。また、以上のように直接「含むこと」意味しなくても mereology を内在した、次のような文型もある。

$$(15) x \text{ が } y \text{ に怪我をする, } x \text{ から } y \text{ をもぎ取る, etc.}$$

型に基づいた推論 「 x の y 」という表現は、mereology の情報を含む場合もあるが、多義性が強い。しかし、次の例では、

$$(16) a. \text{ テレビが故障する}$$

$$b. \text{ チャンネルが故障する}$$

$$c. \text{ テレビのチャンネルが故障する}$$

a の「テレビ」は「テレビのチャンネル」の metonymy となりうる事がわかる。つまり、「故障する」は「部品」から命題への関数型の型をとり、 a のような表現では「全体」から「部品」への型変換が起こっていると考えることができる。したがって、 c の「テレビのチャンネル」は mereology の情報を内在していることがわかる。つまり、次のような型の推論によって「 x の y 」から mereology の情報を導出することが出来る。

$$(17) x \text{ の } y : T \wedge x : T \rightarrow may(y \sqsubseteq_c x)$$

また、二重主語の場合でも同様で、

(18) a. そのテレビは故障している

b. そのテレビはチャンネルが故障している

のような場合では「x は y が」の部分に mereology の情報が、(16) と同様の型変換という理由で、内在しているといえる。従って、次のような規則で形式化できる。

(19) x は : $T \wedge x$ は y : $T \rightarrow may(y \sqsubseteq_c x)$

Domain/Discourse-Driven Inference 一文だけではなく、談話全体から抽出できる mereology もある。例えば、「x の y」という、mereology の情報を含む場合もあるが、多義性が強い表現に対して、談話中の他の情報を利用して、mereology の情報を引き出すということが考えられる。次の工業製品の¹新製品ニュースの一部では、

(20) 6 月初頭に小売価格 300 ドルで販売される予定となっている Fireworks には、良くできたベクターツールおよびビットマップツールをはじめ、優れたイメージ最適化・圧縮機能や、HTML 自動生成機能などが含まれている。Fireworks のペインティングツールは、他のドローおよびベクターベースのイラストレーションプログラムと同じように、ポイントをクリックして、コントロールハンドルをドラッグすることによって線を描く。

「Fireworks のペインティングツール」から mereology の情報を引き出したいが、不確実である。しかし、「Fireworks」がトピックである製品名であることがわかれば、「x の y」という表現で x が製品なら、mereology 情報である可能性が高いため、文脈からそれが導出できればよい。それは同一談話内の「x を発表・開発・販売」という部分を参照すれば得られる情報である。したがって、工業製品の¹新製品ニュースに関しては次ぎのような談話駆動型の mereology 導出規則が考えられる。

(21) a. x の y : $T \wedge object(x) \rightarrow may(y \sqsubseteq_c x)$

b. $(\exists ez. 販売(e, z, x) \vee 開発(e, z, x) \vee 発表(e, z, x)) \rightarrow object(x)$

Anti-Mereology 領域が特定のであると、その領域中のテキストに、ある型をもつトークンがある部

分を持つことに関して否定的な情報がある場合、その情報は単に否定情報をもつのではなく、その型をもつトークン全般もしくは一部に関しては肯定的で、むしろそのトークンの特殊性を述べている。これを **anti-mereology** と呼ぶことにする。例えば、次のテキストでは、「Fireworks」というグラフィック作成プログラムに関する anti-mereology の情報が述べられているが、このことは、「グラフィック作成プログラム」という型に関していえば、逆に mereology の情報を述べていることになる。

(22) もちろん、ブラシやゴムスタンプ、バケツといったツールもある。ぼかし、シャープ、にじみといったツールはないが、Fireworks では、フィルタによってぼかしおよびシャープの効果を加えるようになっている。

従って、トークンレベルの anti-mereology は型レベルではある意味で mereology となる、という関係があるため、anti-mereology (23a) からは (23b) のような非単調推論規則が抽出できる。

(23) a. $a \sqsubseteq_c b \wedge a : T \wedge T' : type \wedge b : T' \wedge T' : type$

b. $x : T \wedge T' : type \wedge y : T' \wedge T' : type \wedge \sim x \sqsubseteq_c y \rightarrow x \sqsubseteq_c y$

anti-mereology の情報をもつ一般的な文型は (24) のようなものである。

(24) x が y を欠いている, x に y がない, x が y を持たない

6 実験

実装 実装は、現在、形態素解析器 ChaSen(奈良先端大松本研) の出力に Perl によるパターンマッチによりテキストから型情報を抽出し、それを Prolog で実装した型推論のフラグメントのプログラムに与える、という形で行っている。

実験対象 Taxonomy に関しては、株式ニュース 3 3 6 記事 (4 8 2 8 文) に対して行い、共起述語による型情報は生起数 2 以上とした。結果は付録参照。株式領域の重要な概念である「株・銘柄・セクター・

材料・市場・値」などが明示的に適切に構造化できた。また、「売り・買い・市場のムード」などイベント・状態を表す抽象名詞は明示的な型は、共起する述語による型でグループ化できた。さらに、述語のクラスターリングも同時に行えた。Mereology に関しては、対象とその仕様記述が多いと見られる、工業ニュースのヘッダーに「販売・開発・発表」を含むものを収集して現在実験中である。一記事からの mereology 構成の結果は、付録を参照のこと。

実験の評価に向けて Taxonomy/Mereology 構成の評価は、同型語/meronym の検索の精度とみなすことができるので、検索結果に対する precision/recall によって与えられる。

参考文献

- [1] H. P. Barendregt. Lambda calculi with types. In S. Abramsky et al., editor, *Handbook of Logic in computer Science*, pp. 118-309. Clarendon Press, Oxford, 1992.
- [2] D. A. Cruse. *Lexical Semantics*. Cambridge University Press, Cambridge, 1986.
- [3] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht, 1993.
- [4] Inderjeet Mani and T. Richard MacMillan. Identifying unknown proper names in newswire text. In Branimir Boguraev and James Pustejovsky, editors, *Corpus Processing for Lexical Acquisition*, pp. 41-59. The MIT Press, Cambridge, 1996.
- [5] Woojin Paik, Elizabeth D. Liddy, Edmund Yu, and Mary McKenna. Categorizing and standardizing proper nouns for efficient information retrieval. In Branimir Boguraev and James Pustejovsky, editors, *Corpus Processing for Lexical Acquisition*, pp. 61-73. The MIT Press, Cambridge, 1996.
- [6] Klemens Schnattinger and Udo Hahn. Constraining the acquisition of concepts by the quality of heterogenous evidence. In Gerhard Brewka, Christopher Habel, and Bernhard Nebel, editors, *KI-97: Advances in Artificial Intelligence*, pp. 255-266. Springer, Berlin, 1997.
- [7] Zeno Vendler. *Linguistics in Philosophy*. Cornell University Press, New York, 1967.

付録：実験結果の一部

最新の実験結果は <http://godiva.is.s.u-tokyo> において公開中である。

株式領域の taxonomy の一部

```

yes:
40: ?- graph_fact2.
top: vfp
|--株.....[sok(株)]
|--全株.....[さくら銀行, 銀行, 住友銀行, 証券, 野村証券]
|--消費者金融.....[アイフル, アコム, シンキ, プロミス, 武富士]
|--[が] [間 a 位].....[銀行]
|--銀行株.....[さくら銀行, 安田信託, 横浜銀行, 紀陽銀, 三井信託, 住友銀行, 足利銀, 長銀]
|--東海銀行, 東京三菱銀行, 富士銀行, 北陸銀
|--[が] [急上昇].....[安田信託]
|--建設株.....[熊谷組, 国土開発, 住友建設, 三井建設, 鹿島, 住友建設, 西松, 大塚建設, 飛島建設]
|--セネコン株.....[佐藤工業]
|--材料株.....[ダイニック]
|--住居材料株.....[ガンゼ産業, コアサ, 川島織物, 東急車両, 東京特殊電線, 権助]
|--住居株.....[東洋化学, 日本ドライケミカル]
|--自動車株.....[トヨタ]
|--低位株.....[skk, a 円台, セッツ, フジタ, コアサコーポ, 株価 x ケタ銘柄, 建設, 殖産住宅, 川崎製鉄, 大成プレハブ, 日本レー, 不動産, 油研工業]
|--内閣低位株.....[石油, 鉄鋼, 不動産]
|--[が] [売買停止].....[セッツ]
|--[に] [乗せる].....[a 円台]
|--鉄鋼株.....[skk, 生友金属, 新日鉄, 川崎製鉄]
|--大鉄鋼株.....[新日鉄, 川崎製鉄]
|--鉄鋼セクター.....[skk, 新日鉄]
|--電線株.....[フジタ]
|--土地再開発関連株.....[三菱地所]
|--内閣株.....[紙パルプ, 石油, 鉄鋼]
|--低位内閣株.....[建設, 鉄鋼, 不動産]
|--品薄株.....[白旗五]
|--不動産株.....[三井不動産]
|--優良株.....[cvt データ, sok, ソニー, トヨタ自動車, 電気機器, 富士写真フィルム, 本田技研工業]
|--区際優良株.....[Doys, asc, sok, アドバンテスト, キヤノン, セブンイレブン, ソニー, トヨタ自動車, 花王, 三共, 村田, 東京エレクトロ, 日立, 富士写真フィルム, 富士通, 武田薬品, 本田技研工業]
|--主力ハイテク株.....[sok, ソニー]
|--ハイテク優良株.....[ソニー]
|--[を] [上げる].....[支配, 儲]
|--[を] [下げる], [を] [消す], [を] [飛ばす], [を] [保つ], [を] [崩す], [を] [戻す].....[儲]

```



```

yes:
27: ?- graph.
top:
|--fireworks
|--skml 自動生成機能
|--イメージ
|--ポップアップメニュー
|--イメージ最適化/圧縮機能
|--エクスポート
|--グループ化
|--ビットマップツール
|--ビットマップ編集ツール
|--ベクターツール
|--ベインティングツール
|--レイヤー設定
|--切り取り機能
|--macroedia
|--製品
|--オブジェクト
|--ゴムスタンプ
|--ツール
|--パスワ
|--ブラシ
|--ベクター
|--ブラシ
|--ペイントスタイル
|--ブラシ
|--ペン先
|--水彩筆
|--文房
|--油絵筆

```



```

no
28: ?- halt.

```