概念の「乗物」についての考察

意味記述の単位と語彙記述の単位のズレを中心にした オントロジーと言語との対応づけの一般問題*

Some thoughts on the “vehicle” of concepts

How to deal with the mismatches between ontological units and linguistic units?

黒田 航† 李 在鎬† 渋谷 良方† 野澤 元† 井佐原 均†

2006 年 12 月 27 日

概要

WordNet や日本語語彙大系のようなシソーラスの有用性は (a) 語は概念を表わす，(b) 文の意味は構成語の意味として与えられた概念が合成された「複合的な概念」であるという前提の下で保証されている．ただ，この想定は自然言語には超語彙的意味 (superlexical meanings) [1] をもつ非線型表現 (nonlinear expressions [2, 3] が多いという事実を考えると，無条件に妥当とは言えない．複層意味フレーム分析 (MSFA) [4, 5, 6] を使った意味タグづけの作業で，この問題にどのように対処したかを紹介し，語の意味と文の意味の関係の一般問題をオントロジーと関係づけて理論的に議論する．

The usefulness of thesauri such as WordNet, A Japanese Lexicon, in NLP tasks relies on the assumption that (a) word meanings are concepts; (b) the meaning of a sentece $s = w_1 \cdot w_2 \cdots w_n$ is a “complex concept” given as a composition of the lexical(ized) concepts for $w_1, w_2, \ldots, w_n$. This assumption, however, cannot be seen as unconditionally valid as far as natural languages are full of “nonlinear expressions” [2, 3] with “supelexical meanings” [1]. Aware of recent advances in the research in (formal) ontology, we discuss the general problem of how to integrate lexical and sentential meanings in semantic analysis of sentences, based on our experience in semantic annotation using *Multilayered Semantic Frame Analysis* (MSFA) [4, 5, 6].

keyword

キーワード: シソーラス，オントロジー，複層意味フレーム分析，超語彙的意味，非線型表現，「概念はモノである」メタファー

Keywords: thesaurus, ontology, MSFA, superlexical meaning, nonlinear expressions, “Concepts are Things” metaphor

1 はじめに

1.1 シソーラスの限界

シソーラス (e.g., WordNet [7, 8], 語彙大系 [9]) はそれ自体はオントロジーではないが，語彙的要素を媒介に言語表現を形式オントロジー (e.g. SUMO [10, 11, 12], DOLCE [13]) に結びつけるために有益な意味資源である．その重要性は今さら強調する必要もないが，その利用価値には次のように，明らかな限界もある:

(1) **シソーラスは基本的に語の意味記述しか与えない**.

別の言い方をすれば,

(2) 単独の語の意味や変項を含まない特定の言い回し (慣用句) に還元できない**超語彙的意味** (superlexical meaning) や非語彙的意味をシソーラスを使って記述することには限界がある

ということである[1)]．非語彙的，超語彙的意味は §3 で示すように，実際の文章では決して例外的でも，稀でもないというのが現実である．とすれば,

(3) 文 $s = w_1 \cdot w_2 \cdots w_n$ の意味 (=文意 $M(s)$) を得るのに，$w_1, w_2, \ldots, w_n$ の個々の要素の意味をシソーラ

* この論文は「言語理解とコミュニケーション研究会」(01/31/2007 札幌コンベンションセンター) で発表される同名の研究発表の増補改訂版である．基本的には枚数制限で割愛された内容を補っている．本稿の以前の版への加藤鉱三 (信州大学) からのコメントに感謝したい.

† 情報通信研究機構 (kuroda ATMARK nict PERIOD go PERIOD jp)

1) 超語彙的意味の定義とその一例に関する心理実験は [1] で論じられている.

ここで私たちが超語彙的意味の単位と読んでいるものは，池原ら [2, 3] が**非線型表現**と呼ぶもの，**構文文法** (Construction Grammar) [14, 15] で **(文法上の) 構文** ((grammatical) constructions) と呼ばれる呼ばれるものとおおよそ等価である.

いわゆる慣用句の表現が固定されたものでなく動的なものであり，頻度が無視できるほど小さいわけではないという認識 [16] が得られて以来，**連語** (multi-word expressions) の網羅的記述の重要性が認識されている [17, 18, 19]．私たちの見地では，連語は構文/非線型表現/超語彙的意味単位の特殊な場合である.

スを参照して指定する (例えば SemCor のように WordNet の synset に指標づけする) だけでは文意が与えられる保証はない.

典型的には，次の場合が問題になる:

(4) 動詞，形容詞ばかりが意味上の述語を喚起するわけではなく，名詞もそのような効果をもつ (これは生成辞書 [20] が提唱するクォリア構造が有効だと考えられる理由の一つ).
(5) 文に n 個の意味上の述語 $P_1, P_2, \ldots, P_n$ が対応づけられた場合，これらの述語の項の統合をうまく表現するための手法が確立していない.
(6) 意味上の述語が，語彙的に特定しがたい，比較的長く複雑な単位 (=**超語彙的単位** (superlexical units)) によって喚起される場合が少なくない.

超語彙的単位には様々なクラスがあるが，わかりやすい例を挙げると，熟語や諺などの慣用表現がそれに該当する．類似の場合としては，重複要素 x, y による異なる意味 $M(x), M(y)$ のエンコード次の問題もあるが，これは上の問題に較べると対処が簡単である.

これらの問題は，複層意味フレーム分析 (MSFA) [21, 22] を使った意味タグづけの作業で繰り返し現れた．私たちがこれらにどう対処したかを簡単に紹介する．本稿では紙面の都合上，一つの例に触れるに留めたが，発表の当日はより具体的な例を沢山挙げるつもりである.

1.2 言語学の意味記述/NLP の意味処理が目指すべきもの

ここで以下の議論のために次のことを確認しておきたい:

(7) (特定の文脈に置かれた) 文の意味がわかっている/理解できている状態とは正確にどんな状態のことなのかは —これまでの何十年の言語学，認知科学，心理学，哲学の研究にも関わらず— ハッキリとわかっているわけではない.

「文の内容が理解された状態とはこういう状態だ」という定義は幾つも存在する．だが，どれについても定義が正しいという証拠は十分とは言えないし，**どれかの定義が他の定義に対して排他的に正しいということはありそうにない**.

これは**言語学が自然言語の意味記述を始める前に，その最終地点を入念に定義しておく必要がある**ということであり，反面では，それは恣意的に設定できるということでもある．これは目標が低い場合に深刻になるだろう.

これはまた，**NLP で意味処理と呼ばれているものが何を達成するべきなのが事前にわかってるわけではいない**ということも含意している.

2 MSFA を使った文意記述の基本問題

2.1 意味記述の基本方針

MSFA は**文の意味を，それが語の意味の単純な合成として与えられないという可能性を考慮して，記述するための手法**として開発された．この背景にあるのは次の考慮である:

(8) 文の意味と語の意味の正確な関係はまだハッキリとわかっているわけではないことを積極的に認め，文意の構成性の確証バイアスから自由な文意の記述に努める必要がある.

これは健全な文の意味記述のために必要不可欠な条件であるが，一般に理解されているとは言い難いので，理由を説明する.

2.1.1 語の意味

意味記述の際のシソーラスの有効性は次の仮定の上に成立している:

(9) 語の意味は**概念** (concept) である (か，少なくとも概念として近似可能である).
(10) 文 $s = w_1 \cdot w_2 \cdots w_n$ の意味は，語 $w_1, w_2, \ldots, w_n$ のおのおの意味 $m_1, m_2, \ldots, m_n$ ($m_i = M(w_i)$) から何らかの形で構成可能である ($M(s) = H(m_1, m_2, \ldots, m_n)$.

ただし $M(x)$ は x の意味を決める関数とする.

$M(w_i)$ の型が概念であるということは (9) で要請されている．H が単なる合成関数であれば，$M(s)$ の型も概念となる．だが，(9), (10) は次の点で問題がないとは言えない:

仮に (9) が正しく，語の意味が概念で近似可能だとしても,

(11) 文の意味 $M(s)$ は —(10) がある解釈の下で規定しているように— 本当に (単なる) **複合的な概念**なのか?
(12) H の実態は何か?

がわからないと，正しいとも間違いとも言えない.

2.1.2 語よりも大きな単位の意味

語よりも大きな単位の意味を定義する**複合的な概念が何であるか**について，(10) から独立の定義が与えられていない点には注意が必要である，このため，(12) の答えとして「H は単純な合成関数だ」と言うことは論点先取かも知れない.

非常に多くの意味記述は —明示的, 非明示的に— (11) を受け入れ，$M(s)$ が多かれ少なかれ $m_1, m_2, \ldots, m_n$ の合成として与えられると仮定しているが，これで本当に有効な意味記述が達成できるのかは明らかではない.

次の (13) が示されない限り，(10) が正しいとは結論できないが，(13) の妥当性に関して十分な証拠があるとは言えない:

(13) 文意 $M(s_i)$ は一般に語の意味の合成と仮定しない限り，記述できない．

2.1.3 語の意味が文の意味に先立つを仮定する必要はない

もちろん，(10) を仮定しないと，次の問題に答えられないのは明らかである:

(14) 一般に文の意味が語の意味の単純な合成として与えられないのだとしたら，私たちは文の意味をどうやって知るのか?

(14) の説明は魅力的だが，実際のところ，無条件に (15) が正しいと考えるわけには行かない:

(15) $s = w_1 \cdot w_2 \cdots w_n$ を構成する w_i の意味が $M(s)$ に先立ってあらかじめ与えられている．

なぜなら，w_i の意味の決定と $M(s)$ の決定には相互依存性が認められる (この意味でも，単純な合成の関係にあるとは言えない)．だが，(10) は (15) が正しいことを仮定している．

実際，$M(s)$ がわかるには $W(s)$ の要素となっている語のおのおのの語義の曖昧性の解消が必要である．ところが，語義の曖昧性の解消は，s 内で相互依存的なプロセスである (一般にそれは n 個の語義の間に成立する多体問題になる)．これは，文意のレベルに単独の語の意味には帰着できない超語彙的意味=構文的意味があり，それが $M(s)$ の**解釈へのバイアス**として働いていない限り，この多体問題は解けない可能性がある[2)]．

2.2 MSFA の文意記述へのアプローチ

(15) が無条件に正しいとは言えないことは次の含意をもつ:

(16) シソーラスを使って，$s = w_1 \cdot w_2 \cdots w_n$ を形式的に構成する語の意味を，おのおの (概念として) 特定することは $M(s)$ を得るのに十分ではないかも知れない (し，もしかしたら必要でもないかも知れない)．

MSFA は (16) に注意を払いながら文意のなるべく正確で具体的な記述を達成するための手法である．具体的には次の仮定の下で文の意味記述にアプローチする:

(17) 文 $s = w_1 \cdot w_2 \cdots w_n$ の意味 $M(s)$ は —それが「複合な概念」だと言っても内実がないので— ヒトが s を読んだり，聞いたりしたときに想起する**状況の集合**である．

[2)] [23] は「x が y を襲う」「y が x に襲われる」について，そのような解釈バイアスの存在を実験的に示した．

3 MSFA を使った超語彙的意味の認定と記述

本稿では紙面の都合上，一つの例に触れるに留めたが，発表の当日はより具体的な例を沢山挙げるつもりである．

3.1 非語彙的，超語彙的意味のクラス

文意が語の意味の単純な合成だと考えると扱いに困る非語彙的，超語彙的意味現象の代表例は，(6) に挙げた場合である: 単一の意味の超語彙的要素によるエンコード (これは熟語や諺などの慣用表現[3)]も含む)，単一の意味の不連続要素によるエンコード (異なる意味の重複要素によるエンコードも関係する，独立に扱うほど深刻ではいように思える)．

3.2 超語彙的意味の偏在性

慣用表現は表現が固定しているとは限らない．それから慣用表現と非慣用表現の境界は曖昧である．実際，変項をもつパターンとしても辞書化が困難な，自由度の高い定型表現は数多く存在する．従って，慣用表現と非慣用表現がうまく区別できることを前提にした意味記述は，あまり効力がない．

例えば (18) の一文[4)]を読んで，この一文から ⟨ ロバがキリギリスに憧れを感じていること ⟩ を読み取るのは難しくない:

(18) **ロバはキリギリスの歌声を聞いて魅了され，自分もあんな風に美しい声で歌ってみたいものだと考えた**．

ただ，それがどうしてなのかをハッキリ言うことは難しい．パターン (19) が ⟨X が感じている憧れ ⟩ に喚起に関与しているの確実だが，憧れがどの語彙的要素によってエンコードされているのか限定するのは至難である:

(19) X は自分が V し { たい | てみたい } (ものだ) と { i. 考えた; ii. 思った; iii. 感じた }

それは (18) と次の (20)，(21) との対比から明らかである:

(20) 彼は次の機会には別の手法を試し**てみたいと**{ i. **思った**; ii. **考えた**; iii. **感じた** }.

(21) 真一は恵の不意をつく挙動は困った**ものだと**{ i. 思った; ii. 考えた; iii. ?**感じた** }.

「X が自分が V てみたい (と { i. 思う; ii. 考える; iii. 感じる })」は一般に ⟨X の V することへの希望 ⟩ を表わすが，⟨ 憧れ ⟩ を表わすとは限らない．「X が S(な) ものだ (と { i. 思う; ii. 考える; iii. 感じる })」はせいぜい ⟨ 不可

[3)] 一言で慣用表現を呼ばれるものには実は下位クラスがある．佐藤 (名古屋大学) のグループによって辞書化が進められている [17, 18, 19].

[4)] 日英対訳データベース [24] を構成する「イソップ寓話」の一つ「ロバとキリギリス」の一文である．

避性〉を表わすもので，〈憧れ〉を表わすものではない．ところが，(19) は単なる〈希望〉ではなく，〈憧れ〉の意味に強く結びついている．

(19) に較べて (22) には，より強い〈憧れ〉との結びつきを認めることができる:

(22) X は自分も V し {たい | てみたい} (ものだ) と {i. 考えた; ii. 思った; iii. 感じた}

(19) と (22) との違いは「(自分) が」と「(自分) も」の違いであるが，これを係助詞の「も」の語彙的な意味として特徴づけるのは適切ではないだろう．憧れの喚起は分散的で，「も」が単独でエンコードしている意味ではないからである．これは (18) に MSFA を使って意味タグづけした (暫定的) 結果 (`http://www.kotonoba.net/~mutiyama/cgi-bin/hiki/hiki.cgi?c=view&p=msfa-aesop01-s01`) からも明らかであるが，詳細は紙面の都合で割愛する．

3.3 MSFA による記述の方針

このような現象は決して稀ではなく，繰り返し現れる．これに対処するため，私たちは次のように方針で記述を行った:

(23) もっとも大きな形式的単位 (複文) の超語彙的意味 (e.g.,〈因果性の指定〉) の認定を行い，それに続いて単文のレベルの超語語彙的意味の (e.g.〈憧れ〉) の認定を行い，それらを形式的特徴 (e.g.,「みたい (と考えた)」「ものだ (と考えた)」のような語彙的単位) の意味に関係づける．

(24) 不連続な要素による意味のエンコードを積極的に認める．

4 議論

4.1 文の意味と語の意味の正確な関係

文の意味 $M(s)$ がわかるには $W(s)$ の要素の意味がわかっている必要があるか? これは経験的に真ではない．実際，語の意味がわかることは文の意味がわかることの前提ではない．発話者 A が $s = w_1 \cdot w_2 \cdots w_n$ で伝えたいこと X ($M(s) \in X$) が理解できるためには $W(s)$ の全部の要素の意味がわかっている必要はない．これは，未知語の意味は推測できるし，それは多くの場合に当たっているという経験的を通じて，私たちがすでに知っていることである．

では逆に，語の意味が全部わかれば $M(s)$ がわかるか? これは (10) に保証されているが，経験的にそうなのかはわからない．

4.2 意味記述のパラドックス

以上のことから明らかなるのは，意味記述が次のパラドックスに直面しているということである:

(25) **語の意味の決定と文の意味の決定の循環性**: 漠然とでも文の意味がわかっていなければ語の意味は決めようがないが，語の意味がまったく不明の状態では文の意味は決めようがない．

このパラドックスを現時点では未解決のパラドックスとして認識し，(26) のような中間目標を掲げる場合，現時点で (14) に答えられないことは深刻な問題ではない:

(26) 適当な文 s_i の意味 $M(s_i)$ の**可能な限り正確な記述**を与えることを中間目標にし，$M(s_i)$ の説明を与えるという目標は —(27) を見込んで— 先送りにする[5].

(27) $m_1, \ldots, m_n$ の合成であると仮定しないで特定された $M(s)$ の (冗長な) 記述が十分に蓄積されてから，$m_1, \ldots, m_n$ と $M(s)$ の差分を求め，$m_1, \ldots, m_n$ と H の内実を決定する．

これは現時点でもっと経験科学的に健全な言語の意味に対する記述方略であると考える．

4.3 「概念はモノではない」はずなのに...

シソーラスがオントロジーといかに関係しているかを不問にしても，それが有意味な意味資源と見なされるためには，次のような前提が必要である:

(28) 知識の基本的要素=単位は概念である．

(29) 概念は語彙項目 (多くの場合には語) によって表わされる

「概念はモノではない」としばしば警告的に言われるが，その効果は薄い．概念の研究者，オントロジーの研究者の多くが事実上，概念をモノとして扱っている．その背後には次の二つのメタファー [25, 26] があるように思われる:

(30)「概念はモノである」(Concepts Are Things)

(31)「概念を分類することは (自然) 物を分類することである」(Classification of Concepts Is Classification of (Natural) Things)

言うまでもなく，これらはメタファー以上のものではない可能性が高い．

「概念はモノである」メタファーは間接的にシソーラスの構築者の仕事を支えている．それは「概念を分類することは (自然) 物を分類することである」という作業方針を与える．「ヒトは概念 (concepts) を操る」[6]という規定を，多くの人が無反省に受け入れているが，この規定は「概念はモノである」というメタファーなしでは意味をもたない．だが，概念はモノではないのだから，それらを自然物として扱う方法には限界がある．実際，概念

[5] この結果，言語学者の仕事は減る．

[6] とはいえ，ヒトが知的活動を脳内で行っているとき，操っているものすべてが概念かどうかはわからない．

の分類は自然物の分類と同じようは行かない[7].

特に大きな問題は，すでに問題にした「意味の成立する単位が語に限らない」という問題である．

4.4 概念化の単位は本当に語か?

シソーラスは基本的に語を単位に概念記述を行う．これは事実上**語が「概念の乗物」と考えられている**ことに等しい．

これは広く受け入れられている見方だが，すでに §3 で問題にしたように，概念 (化) がどんな言語単位によって表わされているか=エンコードされているのかは自明ではなく，文意の重要な部分が超語彙的単位によってエンコードされている可能性が十分にある．従って，これは次の可能性を排除する「危険」な意味観でもある:

(32) 概念化と言語形式との対応関係は構成的なものというより分散的なものかも知れない．

これが正しいとすると，それは「概念の乗物が語である」というシソーラスの利用の前提と部分的に矛盾する．

4.4.1 語の意味，概念，概念化，オントロジーの関係

(17) で触れたことだが，$s = w_1 \cdot w_2 \cdots w_n$ の意味 $M(s)$ が「複合な概念」だと言っても内実がない．

特に (形式) オントロジーの研究という形を取らなくても，概念 (化) の研究は古くから存在する [27]．それにもかかわらず，相変わらず概念が何であるかに関して，研究者の間の同意を見ていない．この意味を正確に理解するために次のことは確認しておいた方がいいだろう:

(33)「語の意味は概念 (concepts) で近似で (きるもので) ある」という定義の有効性は，「概念が何であるか」を決める定義の有効性に依存し，

(34)「概念が何であるか」は「概念化 (conceptualization) が何であるか」を決める定義の有効性に依存し，

(35)「概念化 (と概念) が何であるか」を決める定義は，今のところ異なる研究分野の間 (e.g., 認知科学，人口知能，知識工学，オントロジー研究，哲学) で十分に共通理解が得られている事柄ではない．

4.4.2 動詞は本当に「概念」を表わすのか?

一般に「語は何らかの概念を表わす」と考えられている．だが，(33)–(35) のような事情があるとすれば，この定義にはどれほどの実質があるのだろうか?

名詞に関しては基本的にそれでいいとしても，それ以外の品詞になるとだんだん怪しくなってくる．例えば形容 (動) 詞や動詞が表わしているのは本当に概念なのか?

動詞が概念を表わすとしても，それが「語が表わすものが概念だ」と定義した結果なら，空虚である[8]．MSFA が (17) のように考え，状況という説明概念を使って文の意味を与えようとするのは，語が表わすものが概念 (化) と言っても，概念化の実質がない限り空虚だという反省からである．

4.4.3 概念 (化) は何のために存在するか?

ところで，概念 (化) について何がわかっていないかというと，**ヒトが概念 (化) を何に使っているかがわかっていない**という問題に帰着されるのではないだろうか? これが正しいかどうかはわからないが，もしそうだとすると，(33)–(35) の問題を次のような形でもう少し掘り下げてみることができるように思える:

(36) 概念が自然物と同じように存在するものであるならば，それが存在する理由を問うのは意味がない (自然物が存在する理由は多くの場合，人知を超える) けれど，概念が自然物と同じように存在するものではない，**概念が本質的に人工物だとしたら，それらがヒトにとって存在する理由は，ちゃんと存在している可能性がある**．

これはちゃんと定式化されれば，(形式) オントロジーの利用法にうまい制約をかけるように思われる．

4.5 いつ，どんな形でオントロジーが必要か?

以上の議論のまとめとして言えることは，当たり前のことだが (37) だと思われる:

(37) 何がシソーラスに指定されていて，何が指定されていないかをちゃんと理解しておくことが必要である

ここで次のような基本的な問題に立ち返ってみるのも意味のないことではないだろう:

(38) 自然言語処理 (NLP) や言語学でオントロジーを利用する目的は何だろうか?

この問いの答えは案外，自明ではなかったりする．これは結局，いわゆる「意味処理」で何がしたいのか?という問題に帰着することになるが，意味処理と言っても，浅い処理なのか，深い処理なのかで，大きくやることが違う．浅い処理なら，表層パターンの一致だけで，それなりのことができる．深い処理となると，どこまで深くするかが問題になる．作りの良いオントロジーがあれば，意味処理の深さ，記述の粒度をコントロールできるという利点がある．

4.5.1 シソーラスに表わされていない意味関係

基本的にシソーラスには幾つかの語の (意味の) 間の関係が指定されているが，多くの場合，それは is-a (= subsumption) 関係，part-of 関係のような分類的関係 (taxonomic relations) に限られる．(39) や (40) に現れる ⟨ 行為の目的 ⟩，⟨ 行為の目的の実現手段 ⟩ のような主題

---

[7] 多くのシソーラスの開発者が既成のシソーラスに満足できず，「正しい分類」に到達しようとして自分のシソーラスの開発を始める．

[8] 動詞が表わしているのがコトであるというのは有意義な説明ではない．せいぜいそれはモノではないことがわかるだけである．

関係 (thematic relations) はシソーラスには指定されていないことの方が多い.

(39) 卵 (の白身) (instantiates 食材; instantiates 出発点) で (おいしい (is-a 質)) メレンゲ (instantiates 料理; instantiates 製品; instantiates 目標) を作る

(40) 卵 (の白身) (instantiates 食材; instantiates 出発点) を泡立て (instantiates 手段; instantiates 経路) て (おいしい (is-a 質)) メレンゲ (instantiates 料理 is-a 製品; instantiates 目標) を作る

関係抽出 [28, 29, 30, 31] の狙いの一つは，このような主題的関係を同定することである．格フレーム辞書 [32, 33] の高精度化が要望されている背景には，このような事情がある．

[34] が指摘しているように，格フレーム辞書が与える述語の共起関係は FrameNet [35, 36] がデータベース化を進めている意味フレームの具現化だと考えられるので，主題的関係を網羅する言語資源としては FrameNet が提供するフレームのデータベースも有望だろう．

5 なぜ (NLP に) オントロジーなのか?

5.1 より深い意味処理のために

オントロジーの重要性が意識されるようになったのは，表面的にはシソーラスに記載されている基本的意味関係を補って，より深い意味処理が必要だと自覚されてきたからである．より深い意味は，オントロジーという形で定式化された一般的知識構造に言及しないと対処できない．当然，単なるシソーラスを越えた意味資源=知識ベースが必要であるという意識が広がっている理由になっている．その種の知識ベースは巷ではオントロジーと呼ばれている．ただ，この背景には別の側面もある．

5.2 オントロジーの再興の理由

近頃，オントロジー (の構築) は大人気である．ネコも杓子もオントロジーという感じだ．オントロジー，オントロジーと騒ぐ人の中にはオントロジーが何であるかわかっていない人もいる．この一因には Semantic Web [37] への (過度の) 期待もあるのだろう．

様々な分野でのオントロジーの再興は，それ自体は悪いことではない．NLP の分野でのオントロジー再興は少なくとも部分的には，自然言語の文を相手にした意味処理の高度化の要求から来ている．これは自然言語処理の始まりの頃から潜伏していた問題だが，どちらかというと後回しになっていた．大きな規模で十分に深い処理をするのは至難だったからである．だが，この問題はインターネット時代になって尖端化した．今日では全地球規模で夥しい量の情報が自然言語の形でやり取りされるようになっり，大量の自然言語データを効率良く「処理」する必要は，今や至るところに存在する．

テキストマイニングや文書分類の基盤技術として，フリーの形態素解析器 (e.g., juman, chasen) が大きく貢献したのは周知のことである．だが，要求は常に高度化し続ける．NLP 内部では係り受け解析の品質はまだ十分ではない (新聞コーパスへの overfitting も含める) という自覚があり，まず被覆率の向上のため，大量データから自動獲得した格フレーム辞書 [32] で対応を試みているが，精度の向上は，大きな課題である

基本的に意味を無視した処理の効率は，頭打ちの傾向にあり，現実問題として，意味処理はまだまだである．更に言えば，形態素解析，係り受け解析のいずれについても，間接的に意味処理が入っている．これらの技術の精度の向上には意味処理の高度化が不可欠である．

この目的に使えるほぼ唯一の意味資源がシソーラス (e.g., 日本語語彙大系 [9], 分類語彙表 [38], WordNet [8]) であった．だが，その利用価値には限界に感じられている．**「単なるシソーラス以上の何かが必要だ」と多くの研究者が痛感している**．これがオントロジーに対して大きな期待が生じる理由であるように私たちには思われる．

5.3 形式オントロジーと形式ばらないオントロジーの狭間

「オントロジー」と呼ばれているものの実態は何か? これがどうも，あまりハッキリしているとは言えない．オントロジーに関する色々な理論 [39, 40, 41, 42, 43, 44, 45] によって定義が提案され，そのうちの幾つかは広く流通しているが，共通理解と言えるのは「オントロジーは (十分に) 形式的なものである (べきだ)」という点ぐらいである．「オントロジーは概念化の明示化である (An ontology is a(n explicit) specification of a conceptualization)」という定義 [39, 40] がもっともよく流通しているが，これがどれぐらい意味をもつのかは怪しい面がある．というのは，正直なところ**概念化が何であるかはよくわからない**というのが現状であるように思うからである[9]．そして，概念化が何であるかがわからなければ，**概念が何であるか**がわからない．

以上のような理由で，NLP や言語学の研究者の多くは，いわゆる**形式オントロジー** (formal ontologies) と**形式ばらないオントロジー** (informal ontologies) との狭間で，両者の板挟み状態にあるように思える．形式ばら

[9] この点は [46] でも触れられている．[39, 40] は次のように述べている: “A body of formally represeneted knowledge is based on a *conceptualization*: the objects, concepts, and other entities that are assumed to exist in some areas of interest and the relationships that hold among them [47]. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.”

だが，これはあまりに漠然とした定義であり，それが**世界観** (Weltanshauung = worldview) [48] の簡略版であるという以上の特徴は読み取れない．

ないオントロジーの代表格は，WordNet，日本語語彙大系，分類語彙表に代表されるシソーラス類である．実際，NLP でオントロジーと言うと，これらだと理解される[10]．

シソーラスは SUMO [10, 11, 12]，DOLCE [13]，Cyc [51, 52] のような (形式化された) オントロジーより手軽に使える，形式ばらないオントロジーのような位置を占めている．形式オントロジーは多くの形式オントロジーの利用には事前知識が必要で (まだまだ) 気軽に使えるものではない以上，本格的なオントロジーの代用品としてのシソーラスに多くの研究者の関心が寄せられるのは避けようがないように思う．

脚光を浴びている NLP の課題の一つである，自動獲得された意味関係をオントロジー化する (ontologizing/ontologization) 課題 [28, 53] で「正解」を表現するのに使われているのは WordNet である．だが，多くの研究者が指摘するように，WordNet は本格的なオントロジーではない[11]．

とはいえ，自然言語の意味をオントロジーを使って記述するとなると，次のような問題が生じる:

(41) 自然言語の意味記述は，形式オントロジーが理解できる専門家によって行われることを前提にするわけには行かないし，形式オントロジーの完成を待っているわけには行かない．

実際，Semantic Web [37] の構想が実現される日が本当に来るのかはわからないし，当面は形式化への肩入れは極端でない方がいい．これが現状である以上，当面はシソーラスを拡張し，形式ばらないオントロジーの充実に心がけるのがもっとも現実的な路線なのだろう．

参考文献

[1] 中本, 李, 黒田：“日本語の語順選好は動詞に還元できない文レベルの意味と相関する: 心理実験に基づく日本語の構文研究への提案”, 認知科学, **13**, pp. 334–352 (2006). 「文理解」特集号.

[2] 池原, 徳久, 村上, 佐良木, 池田, 宮崎：“非線形な重文複文の表現に対する文型パターン辞書の開発”, 情報処理学会研究報告, **NL-170**, 25, pp. 157–164 (2005).

[3] 池原, 阿部, 竹内, 徳久, 村上：“意味的等価変換方式のための重文複文の統語的意味的分類体系について”, 情報処理学会研究報告, **2006-NL-176**, pp. 1–8 (2006).

[4] K. Kuroda and H. Isahara: “Proposing the MULTILAYERED SEMANTIC FRAME ANALYSIS OF TEXT”, The 3rd International Conference on Generative Approaches to the Lexicon, pp. 124–133 (2005). [Revised version is available as: `http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/msfa-gal05-rev1.pdf`].

[5] 黒田, 井佐原：“意味フレーム分析は言語を知識構造に結びつける: 文 “*x* が *y* を襲う” の理解を可能にする意味フレーム群の特定”, KLS 25: Proceedings of the 29th Annual Meeting of Kansai Linguistic Society, 関西言語学会 (KLS), pp. 326–336 (2005). [増補改訂版: `http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/sfal-osou-kls29-rev2.pdf`].

[6] 黒田, 井佐原：“複層意味フレーム分析 (MSFA) による文脈に置かれた語の意味の多次元的表現: 実例に基づく msfa の設計思想の解説”, 日本認知言語学会論文集, 第 6 巻, pp. 171–181 (2006). Available as: `http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/kuroda-isahara-06-jcla-paper-submitted.pdf`.

[7] G. Miller: “Wordnet: An online lexical database”, International Journal of Lexicography, **3 (4)**, (1990).

[8] C. Fellbaum Ed.: “WordNet: An Electronic Lexical Database”, MIT Press (1998).

[9] NTT コミュニケーション科学研究所：“日本語語彙大系”, 東京: 岩波書店 (1997).

[10] I. Niles and A. Pease: “Towards a standard upper ontology”, Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, October 17-19, 2001. (Eds. by C. Welty and B. Smith) (2001).

[11] A. Pease and I. Niles: “IEEE Standard Upper Ontology: A progress report”, Knowledge Engineering Review: Special Issue on Ontologies and Agents, **17**, pp. 65–70 (2002).

[12] I. Niles and A. Pease: “Linking lexicons and ontologies: Mapping WordNet to Suggested Upper Merged Ontology”, Proceedings of the International Conference on Information and Knowledge Engineering (IKE-03), Las Vegas, Nevada, June 23–26, 2003 (2003).

[13] A. Gangemi, N. Guarino, C. Masolo and A. Oltramari: “Sweetening WordNet with DOLCE”, AI Magazine, **24 (3)**, pp. 13–24 (2003).

[14] C. J. Fillmore, P. Kay and K. O’Connor: “Regularity and idiomaticity in grammatical constructions: The case of let alone”, Language, **64**, 3, pp. 501–538 (1988).

[15] A. D. Goldberg: “Constructions: A Construction Grammar Approach to Argument Structure”, University of Chicago Press, Chicago, IL (1995).

[16] I. Sag, T. Baldwin, F. Bond, A. Copestake and

[10] これは [46] が指摘するように，実際には誤りに近く，オントロジーの研究家 [49, 50, 13] は WordNet の問題点を指摘し，拡張を提唱している．

[11] [54, 55] は WordNet には (意味の) 型 (types) と役割 (roles) の区別がないと指摘した．例えば `animal` synset には `chordate`, `larva`, `fictional animal` や `work animal`, `domestic animal`, `mate`, `captive`, `prey` が一緒に含まれているが，前者は types で後者は roles であると [55] は言う．同様の指摘は，日本語語彙大系 [9] の名詞概念の分類体系について [56] によって同様の指摘が独立に行われている．

[41] は型と役割の区別の欠如の他に「is-a 関係の使いすぎ」(is-a overloading) を WordNet の問題点として挙げている．is-a 概念同士の包含/包摂関係を表わす is-a 関係と instance-of 関係は区別する必要があると論じている: “The problem with ISA when considering linguistic ontologies like WordNet is that it is intended as a *lexical* relation between words, which not always reflects an *ontological* relation between classes of entities of the world.”

D. Flinckinger: “Multiword expressions: A pain in the neck for NLP”, Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (Mexico City), pp. 1–15 (2002).

[17] 尾嶋, 佐藤, 宇津呂：“日本語慣用句用例データベースの構築法”, 言語処理学会第 12 回年次大会発表論文集, pp. 456–459 (2006).

[18] 橋本, 佐藤, 宇津呂：“自動検出のための慣用句の分類と語彙的情報”, 言語処理学会第 12 回年次大会発表論文集, pp. 825–828 (2006).

[19] 橋本, 佐藤, 宇津呂：“依存構造照合に基づく慣用句自動検出”, 言語処理学会第 12 回年次大会発表論文集, pp. 829–832 (2006).

[20] J. Pustejovsky: “The Generative Lexicon”, MIT Press (1995).

[21] 黒田, 井佐原：“意味フレームを用いた知識構造の言語への効果的な結びつけ”, 信学技報, **104 (416)**, pp. 65–70 (2004). [増補改訂版: http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/linking-l-to-k-v3.pdf].

[22] Y. Shibuya, K. Kuroda, J. H. Lee and H. Isahara: “Specifying deeper semantics of a text using MSFA”, IEIECE Techical Report, **106**, 299, pp. 27–32 (2006). NLC2006-27 (2006-10).

[23] 黒田, 中本, 野澤, 井佐原：“意味解釈の際の意味フレームへの引きこみ効果の検証: “*x* が *y* を襲う ” の解釈を例にして”, 日本認知科学会 第 22 回大会 発表論文集, pp. 253–55 (Q–38) (2005). [増補改訂版: http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/frames-attract-readings-jcss22.pdf].

[24] 内山, 高橋：“日英対訳文対応付けデータ”, http://www2.nict.go.jp/x/x161/members/mutiyama/align/index.html (2003).

[25] G. Lakoff and M. Johnson: “Metaphors We Live By”, University of Chicago Press (1980). [邦訳: 『レトリックと人生』 (渡部昇一ほか 訳). 大修館.].

[26] G. Lakoff and M. Johnson: “The Philosophy in the Flesh”, Basic Books (1999).

[27] G. L. Murphy: “The Big Book of Concepts”, MIT Press (2002).

[28] P. Pantel: “Inducing ontological co-occurrences vectors”, Proceedings of ACL-05, pp. 125–132 (2005).

[29] M. Pennacchiotti and P. Pantel: “A bootstrapping algorithm for automatically harvesting semantic relations”, Proceedings of Inference in Computational Semantics (ICoS-06), pp. 87–96 (2006).

[30] P. Pantel and M. Pennacchiotti: “Espresso: Leveraging generic patterns for automatically harvesting semantic relations”, Proceedings of the COLING/ACL-06, pp. 113–120 (2006).

[31] K. Shinzato and K. Torisawa: “Acquiring hyponymy relations from web documents”, Proceedings of HLT-NAACL-2004, Boston, MA, pp. 73–80 (2004).

[32] 河原, 黒橋：“格フレーム辞書の漸次的自動構築”, 自然言語処理, **12**, 2, pp. 109–131 (2005).

[33] 笹野, 河原, 黒橋：“名詞句格フレーム辞書の自動構築とそれを用いた名詞句の関係解析”, 自然言語処理, **12**, 3, pp. 129–144 (2005).

[34] 中本, 黒田：“「逃れる」の階層的意味フレーム分析とその意義: 「言語学・心理学からの理論的，実証的裏づけ」のある言語資源開発の可能性”, 言語処理学会第 12 回大会発表論文集, pp. 592–595 (2006). 発表 P4-1.

[35] T. Fontenelle Ed.: “FrameNet and Frame Semantics”, Oxford University Press (2003). A Special Issue of *International Journal of Lexicography*, 16 (3).

[36] K. H. Ohara, S. Fujii, H. Sato, S. Ishizaki, T. Ohori and R. Suzuki: “The Japanese FrameNet project: A preliminary report”, Proceedings of PACLING ’03, pp. 249–254 (2003).

[37] T. Berners-Lee, J. Hendler and O. Lassila: “The semantic web”, Scientific American, **May**, (2001).

[38] 国立国語研究所：“分類語彙表 (増補改訂版)”, 大日本図書 (2004).

[39] T. R. Gruber: “A translation approach to portable ontology specifications”, Knowledge Acquisition, **5**, pp. 199–220 (1993).

[40] T. R. Gruber: “Toward principles for the design of ontologies used for knowledge sharing”, International Journal of Human-Computer Studies: Special issue on Formal Ontology in Conceptual Analysis and Knowledge Representation, pp. 907–928 (1995).

[41] N. Guarino: “Some ontological principles for designing upper level lexical resources”, Proceedings of the First International Conference on Language Resources and Evaluation (Granada, 28–30 May 1998) (Eds. by A. Rubio and Others), ELRA, Paris, pp. 527–534 (1998).

[42] 溝口：“オントロジー研究の基礎と応用”, 人工知能学会誌, **14**, 6, pp. 45–56 [977–988] (1999).

[43] 溝口：“オントロジー工学”, オーム社 (2005).

[44] 溝口：“特集「開発されたオントロジー」”, 人工知能学会誌: 特集「開発されたオントロジー」, **19**, 2, pp. 135–193 (2004).

[45] J. F. Sowa: “Knowledge Representation: Logical, Philosophical, and Computational Foundations”, Brooks/Cole, Pacific Grove, CA (2000).

[46] N. Guarino and P. Giaretta: “Ontologies and knowledge bases: Towards a terminological clarification”, Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (Ed. by N. Mars), Amsterdam, IOS Press, pp. 25–32 (1995).

[47] M. R. Genesereth and N. Nilsson: “Logical Foundations of Artificial Intelligence”, Morgan Kaufmann, San Mateo (1987).

[48] W. Dilthey: “Dilthey’s Philosophy of Existence: Introduction to *Weltanschauungslehre*”, Vision, London (1960).

[49] A. Gangemi, R. Navigli and P. Velardi: “The OntoWordNet Project: Extension and axiomatization of conceptual relations in WordNet”, Proceedings of the International Conference on Ontologies, Databases and Applications of Semantics (ODBASE2003) (2003).

[50] A. Gangemi, R. Navigli and P. Velardi: “Axiomatizing WordNet glosses in the OntoWordNet project”, Proceedings of the Workshop on Human Language Technology for the Semantic Web and Web Services, 2nd International Se-

mantic Web Conference (ISWC2003) (2003).

[51] D. Lenat, R. V. Guha, D. Pittman and M. Shepard: “Cyc: Towards programs with common sense”, Communications of the ACM, **33**, 8, pp. 30–49 (1990).

[52] D. Lenat: “Cyc: A large-scale investment in knowledge infrastructure”, Communications of the ACM, **38**, 11, pp. 33–38 (1995).

[53] M. Pennacchiotti and P. Pantel: “Ontologizing semantic relations”, Proceedings of Conference on COLING/ACL-06, pp. 793–800 (2006).

[54] N. Guarino: “The ontological level”, Philosophy and the Cognitive Science (Eds. by R. Casati, B. Smith and G. White), Holder-Pivhler-Tempsky, Vienna, pp. 443–456 (2004).

[55] A. Oltramari, A. Gangemi, N. Guarino and C. Masolo: “Restructuring WordNet’s top-level: The *OntoCean* approach”, Workshop Proceedings of OntoLex ’02, Ontologies and Lexical Knowledge Bases, LREC2002, Las Palmas, Spain, May 27, 2002 (Ed. by K. Simov), pp. 17–26 (2002).

[56] 黒田, 井佐原：“意味役割名と意味型名の区別による新しい概念分類の可能性: 意味役割の一般理論はシソーラスを救う?”, 信学技報, **105**, 204, pp. 47–54 (2005). [増補改訂版: `http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/roles-save-thesauri-rev1.pdf`].