

語彙概念構造のオブジェクト指向化について

山田隆弘

宇宙航空研究開発機構 宇宙科学研究所

tyamada@isas.jaxa.jp

1 はじめに

語彙の意味を記述するために語彙概念構造 (Lexical Conceptual Structure) [1][2]とされている方法が用いられている。この方法の特徴は、語彙の意味を基本的な意味要素に分解して記述することであるが、この方法を用いると、複数の関連する語彙間の意味的な関係を表すこともできるので、語彙を体系的に記述するための方式として注目を集めている [3][4]。

ところで、Saussure [5]が述べたように、言語とは「意味されるもの (signifié)」と「意味するもの (significant)」との対応関係として理解されるべきである (図1 参照)。言語表現の意味も、この両者の対応関係を基礎として記述されるべきである。

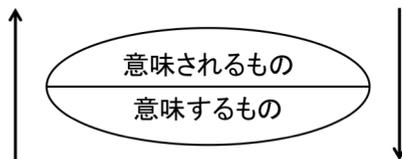


図1 「意味されるもの」と「意味するもの」

この対応関係を厳密に記述するには、「意味されるもの」の構造が「意味するもの (特定の言語表現)」の構造 (特定言語の語彙や文法) とは独立に規定されていなければならない。なぜならば、同一の「意味されるもの」を複数の言語によって表現することが可能であるからである。その場合、「意味されるもの」は、特定言語の辞書と文法を用いて特定言語の表現に変換される (図2 参照)。言語学においても言語処理においても「意味されるもの」を「意味するもの」と独立に規定するための試みは非常に少ないが、石崎[6]の提唱する概念世界と言語世界の対立は、これに相当すると思われる。

本論文では、「意味されるもの」の構造をオブジェクト指向モデリング[7]の手法に基づいて「意味するもの」とは独立に規定する方法を提案し、それによって語彙概念構造をさらに厳密に規定するための試みを示す。

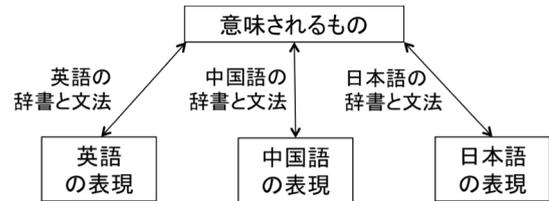


図2 「意味されるもの」と特定言語表現

2 オブジェクト世界

この節では、「意味されるもの」の構造をオブジェクト指向モデリング[7]の手法に基づいて規定する方法を説明する。オブジェクト指向に関する用語は、オブジェクト指向分野の標準言語である Unified Modeling Language (UML) [8]の用語を用いるものとする。

ここでは、意味される世界 (現実世界でも仮想世界でもよい) に存在する物をオブジェクトと呼び、意味される世界は一群のオブジェクトにより構成されるものとする。オブジェクトは、人間や機械のような物理的実体でもよいし、計画や規則のような概念的な存在でもよい。

各々のオブジェクトは、一つあるいは複数のアトリビュートを有する。各々のアトリビュートは値を有しており、時間とともにその値が変わり得るアトリビュートと値が不変であるアトリビュートとがある。前者のアトリビュートの値は、そのオブジェクトのその時点における状態を表し、後者のアトリビュートの値は、そのオブジェクト固有の性質を表す。(アトリビュートの概念は、竹内が[9]で使用している状態という概念をオブジェクト世界の要素として厳密に規定したものとなっている。)

各々のオブジェクトは、一つあるいは複数のオペレーションを有してもよい。各々のオペレーションは、そのオブジェクトが実行する行為を表し、オペレーションを起動するときには一つあるいは複数のパラメータを指定しても良い。

二つのオブジェクトは、一つあるいは複数の関係 (リンクと呼ばれる) で結ばれても良い。このオブ

ジェクト世界では、リンクの両端には名前を付けるものとする(ただし、UML ではこれは必須ではない)。三つ以上のオブジェクトが一つのリンクで結ばれることも可能である。リンクがアトリビュートやオペレーションを持って良い。

同じアトリビュート、オペレーション、リンクを共有するオブジェクトの集合をクラスと呼ぶ。あるクラスのオブジェクトと別の(あるいは同じ)クラスのオブジェクトを結びつける同じ名前を持つリンクの集合をアソシエーションと呼ぶ。

オブジェクト世界は、クラス定義とオブジェクト定義の2段階に分けて定義される。クラス定義は、クラス、アトリビュート、オペレーション、アソシエーションを定義したものであり、複数のオブジェクトに共通する特徴を規定する。オブジェクト定義は、オブジェクト、アトリビュート値、リンクを定義する。記述論理[10]の用語を用いると、クラス定義は TBox に相当し、オブジェクト定義は ABox に相当する。

「意味されるもの」の構造は、オブジェクト世界として規定され、これは「意味するもの」を表現する言語の文法には依存せずに規定できる。ただし、アトリビュートやオペレーション等の名前付けに自然言語を用いると、これらの意味がその言語に依存する可能性もあるので、注意が必要である。

簡単なオブジェクト世界の例を UML の記法に従って図3に示す。図3の上半分はクラス定義であり、下半分はオブジェクト定義である。

この世界には、人間と組織という二つのクラスが存在する。生年月日と住所は、人間クラスのオブジェクトが有するアトリビュートである。見ると食べ

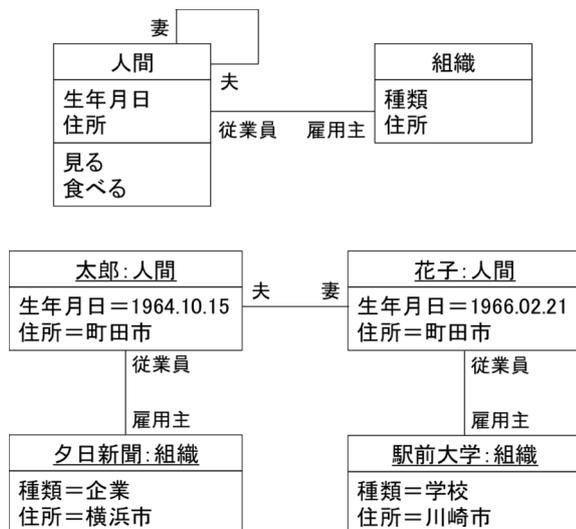


図3 簡単なオブジェクト世界の例

るは、人間クラスのオブジェクトが有するオペレーションである。また、アソシエーションが二つ定義されていて、これらは、二つの人間クラスのオブジェクトの間には夫-妻というリンクを張ることができること、および、人間クラスのオブジェクトと組織クラスのオブジェクトとの間には従業員-雇用主というリンクを張ることができることを示している。

この世界には四つのオブジェクトが存在する。太郎と花子というオブジェクトは人間クラスに属し、夕日新聞と駅前大学というオブジェクトは組織クラスに属す。これらのオブジェクトのアトリビュート値も示されている。これらのオブジェクトのオペレーションは、クラスで規定されているものと同じであるので、オブジェクト定義においては省略されている。また、太郎と花子の関係がリンクとして示されている。太郎と花子の雇用主もそれぞれリンクを用いて示されている。

3 従来の語彙概念構造

ここでは、従来の語彙概念構造の代表例として Van Valin [11]のものを紹介する。これを取り上げる理由は、これが現時点では最も網羅的に語彙概念構造を規定していると思われるからである。

Van Valin は、動詞を6種類に分類し、それぞれの種類毎に語彙概念構造の表現方法を規定している。これらの動詞種類は、表1に示すように4つ特徴によって分類されている。表1の4つの特徴の意味は、表2に示す。

表1 Van Valin の動詞分類

	static	dynamic	telic	punctual
State	+	-	-	-
Activity	-	+	-	-
Achievement	-	-	+	+
Semelfactive	-	+, -	-	+
Accomplishment	-	-	+	-
Active accomplishment	-	+	+	-

表2 動詞分類で使用了特徴の意味

特徴	意味
static	起こらない(起こっていない) こと
dynamic	行動を伴うこと
telic	終了点が存在すること
punctual	持続性があること

これらの動詞種類に対する語彙概念構造の記述例

を表3に示す。ここで、アポストロフィ付きの太字で示されている語は述語であり、()内に示されているのは述語の項である。また、表3で使用されている用語の意味を表4に示す。

表3 Van Valin の語彙概念構造の例

動詞分類	語彙概念構造
State	live-in' (太郎, 町田市)
Activity	do' (太郎, [eat' (太郎)])
Achievement	INGR work-for' (太郎, 隔日新聞)
Semelfactive	SEML see' (太郎, 絵)
Accomplishment	BECOME live-in' (太郎, 鎌倉市)
Active accomplishment	do' (太郎, [eat' (太郎, うどん)]) & INGR consumed' (うどん)

表4 語彙概念構造の記述で使用する用語の意味

用語	意味
INGR	(ingressive) 動作が開始されること
SEML	(semelfactive) 状態変化を伴わずに瞬間的に起こること
BECOME	その状態になること

4 オブジェクト指向語彙概念構造

本節では、オブジェクト世界を用いて語彙概念構造を記述する方式を示す。本方式の特徴は、語彙概念構造をオブジェクト世界の状態や変化として記述することである。表1に示した Van Valin の動詞の6分類のうち、State をここでは状態と呼び、「オブジェクト世界がどうなっているか」として記述する。それ以外の5分類は、ここでは事象(Event)と総称し、「オブジェクト世界で何が起きたか、または、起きているか」として記述する。

状態の概念構造は、以下の3つのいずれかとして記述される。

- 1) オブジェクトの存在
- 2) オブジェクトのアトリビュート値
- 3) オブジェクト間のリンクの存在

事象の概念構造は、以下の4つのいずれかとして記述される。

- 4) オブジェクトの生成あるいは消滅
- 5) オブジェクトのアトリビュート値の変更
- 6) オブジェクト間のリンクの生成、消滅、変更
- 7) オブジェクトのオペレーションの実行

オブジェクト指向を用いて語彙概念構造を記述するための具体的な方法には、いろいろなものがあり

得るが、以下ではこれを日本語文で表すものとする。より形式的な記述方式は別途報告する。

上の各分類の語彙概念構造の実例を表5に示す。ここでは、「」内の日本語表現(意味するもの)に対応する語彙概念構造(意味されるもの)を<>内に示す。

表5 オブジェクト指向による語彙概念構造の記述例

分類	記述例
オブジェクトの存在	「太郎は人間である」→ <太郎は人間クラスのオブジェクトである>
アトリビュート値	「太郎は町田市に住んでいる」→ <太郎オブジェクトの住所アトリビュートの値は町田市である>
リンク	「太郎は夕日新聞に勤務している」→ <太郎オブジェクトの雇用主は夕日新聞オブジェクトである>
オブジェクトの生成・消滅	「花子が1966年2月21日に生まれた」→ <人間クラスの花子オブジェクトが生成され、生年月日アトリビュートの値は1966.02.21である>
アトリビュート値の変更	「太郎が鎌倉市に引っ越した」→ <太郎オブジェクトの住所アトリビュートの値が鎌倉市に変更された>
リンクの生成・消滅・変更	「太郎が隔日新聞に転職した」→ <太郎オブジェクトの雇用主は隔日新聞オブジェクトに変更された>
オペレーションの実行	「太郎がうどんを食べた」→ <太郎オブジェクトはうどんをパラメータとして食べるオペレーションを実行した>

事象の記述において、事象が進行中であるか(例:太郎は食べている)、あるいは、既に完了しているか(例:太郎は食べた)の区別も記述する必要がある。また、状態についても事象についても、いつ、どこで、どのようにして、という付加情報を追加すべき場合もある。このような詳細な記述方法については別途報告する。

各々の分野毎に、その分野で使用すべき標準的なクラス定義と標準的な述語の語彙概念構造を定め、それらをその分野の標準辞書として発行すべきである。その分野における意味記述は、標準辞書に準拠して記述すべきである。

5 本方式の特徴

本方式の特徴は、「意味されるもの」をオブジェクト世界の状態や変化として記述することである。これによって従来の意味論では曖昧であった概念のある面では精密に規定できるようになる。

表5では紙面の都合により実例のみを示したが、本来は一般的な語彙概念構造として規定すべきである。例えば、「引越す」という動詞の語彙概念構造は、次のように規定される。

「XがYに引越す」→

<X オブジェクトが自分の住所アトリビュートの値をYに変更する>

上記の規定により、「Xが引越す」という場合、「X オブジェクトは可変の住所アトリビュートを持たねばならない」ことが導かれる。これにより、組織も可変の住所アトリビュートを有するので、組織も引越し可能であることが導かれる。これは、従来の意味論における意味役割の概念よりも精密に述語（動詞等）と項（名詞等）の間の意味関係を規定できることを示している。

また、従来の語彙概念構造で用いられている特徴（表2参照）のうち、staticは「オブジェクト世界の状態」に、telicは「アトリビュート値の変更」にほぼ対応するが、dynamicとpunctualの定義は必ずしも厳密ではない。その結果、動詞種類のAchievementとAccomplishmentの区別も厳密ではない。（詳細は別途報告するが）オブジェクト世界との関連で規定を行う方が、より厳密に語彙概念構造を規定できる。

ただし、オブジェクトを用いれば全ての曖昧さを排除できるわけではない。アトリビュートやオペレーション等の名前に自然言語を用いる場合は、その言語に内在する曖昧さが持ち込まれる可能性がある。また、オブジェクト世界の構成も一意に決まるとは限らず、一つのことをアソシエーションでもアトリビュートでもオペレーションでも表わせる場合があり、何らかの基準が必要となる。

6 将来の課題

本方式に関する将来の課題は、以下の通りである。

- 1) オブジェクト世界およびオブジェクト指向語彙概念構造のさらに精密な規定の作成。
- 2) 従来の語彙概念構造との詳細な比較。
- 3) 複数の述語を持つ意味構造、および、談話の意味構造の記述への拡張。
- 4) 代表的な分野における標準的なクラスと標準的な語彙概念構造を定義した標準辞書の作成。

- 5) 本方式を利用した応用技術の開発。例えば、
 - UMLモデルの意味を厳密に規定する方式
 - 文書を厳密に記述する方式
 - 情報ベースや知識ベースの構築
 - セマンティックウェブの構築

7 おわりに

本論文では、「意味されるもの」の構造をオブジェクト指向モデリングの手法に基づいて「意味するもの」とは独立に規定する方法を提案し、それによって語彙概念構造がどのように記述されるかを示した。この研究は、まだ始まったばかりであり、上述のように将来の課題も多い。しかし、本方式は様々な分野で様々な用途に応用できるはずであり、今後は視野を広げつつ研究を進めたい。

参考文献

- [1] Kearns, K.: *Semantics*, Second Edition, Palgrave Macmillan (2011).
- [2] Jackendoff, R.: *Semantic Structures*, MIT Press (1990).
- [3] 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央: 語彙概念構造に基づく事象上位オントロジーの構築, 言語処理学会第13回年次大会, pp. 859-862 (2007).
- [4] 乾健太郎: 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」, pp27-30 (2007).
- [5] Saussure, F.: *Cours de Linguistique Générale*, Payot (1915).
- [6] 石崎俊, 自然言語処理, 昭晃堂 (1995).
- [7] Blaha, M., Rumbaugh, J.: *Object-Oriented Modeling and Design with UML*, Second Edition, Prentice Hall (2005).
- [8] Booch, G., Rumbaugh, J., Jacobson, I.: *The Unified Modeling Language User Guide*, Second Edition, Addison-Wesley (2005).
- [9] 竹内孔一, 石原靖弘, 竹内奈央: 述語項構造のシソーラス分類と意味役割の設計について, 人工知能学会全国大会 (2013).
- [10] Baader, F., et al.: *The Description Logic Handbook: Theory, Implementation and Applications*, Second Edition, Cambridge University Press (2010).
- [11] Van Valin, R. D.: *Exploring the Syntax-Semantic Interface*, Cambridge University Press (2005).