

オントロジー理論に基づく非文字資料のデータ化可能性の検討

能登 正人・木下 宏揚

1. はじめに

人類の科学文明発達の歴史において、文字文化の果たしてきた役割は計りしれない。人類はその歴史において、大量の情報の中から知識を創出し、継承してきた。文字の発明は知識を記述し流通させる媒体を人類に与え、記述することの重要性を示した。一方、計算機の発明は知識を利用する新たな方法を提示し、高度に概念的な記述の恣意的解釈を排除した客観的検証を行う道具をもたらした。さらに、インターネット上で動作するウェブシステムの普及は、新しい知識を創出する新たな主体を出現させている。近年の計算機の処理能力と情報網の拡大化の進展によって明示化された、知識の計算機による流通と利用が一般化されるに従って、知識を扱う工学分野に対する期待が高まっている。また、人文科学的領域を扱う知識システムにおいては、人間が科学の発展に伴ってきたように、知識、あるいは、単に記号の羅列だけでは表現しきれない意味を積み上げる技術の必要性が強く求められている。このような背景のもと、対象世界の知識を体系化する「オントロジー (ontology)」に関する研究が近年盛んに行われている¹⁾。

しかしながら、民俗文化資料のような、文字として表現されない人間の文化継承を従来のオントロジーだけをベースとして蓄積・共有するのは困難である。そのため、既存メディアでは表現不可能な民俗文化研究の研究資料を「非文字資料」と定義し、非文字資料のデータベースに適した概念をオントロジーをベースに拡張することにより、非文字資料のデータベースを構築できる可能性がある。また、文化人類学の研究においては、各研究者が独自の形式で研究資料の多数を所有しているケースが多いことから、多種多様な非文字資料が存在する。このような状況下では、非文字資料を共有・流通する際に以下のような問題が生じる。

- ・非文字資料間同士の関係記述が不明確
- ・研究者間の相互利用・運用が困難

これらの問題の解決方法として、明示されていない要素を、複数の概念の対応付けという形で明示することにより、非文字資料の共有と流通が図られると考える。

本研究では、オントロジー基礎理論を踏まえながら、非文字資料に適したオントロジー記述に必要となる概念的性質を明らかにする。すなわち、従来のオントロジー基礎理論でのオントロジー構築では暗黙的であドホックに扱われてきた概念的性質を、明示的に体系化した枠組みで扱うことにより、非文字資料の共有と流通を図る。

2. 非文字資料

2.1 非文字資料の定義と体系化

本稿で扱う「非文字資料」は、文字媒体として記録されることなく受け継がれてきた民俗文化の変遷を対象とする。文字に記録されることなく継承されてきた人類文化の歴史は、人間の観念・知識・知恵・行為など幅広く、文字媒体として記録され続けてきた事象とは比較にならない伝承文化である。民俗文化の歴史に

において、非文字資料とは「伝承」によって行われる、文字媒体として記録されなかった人間の営みに関する資料である。民俗学における研究の手法としては、聞き取り調査、フィールドワーク、文書などの記録の観察、建築や日用品から民間伝承までさまざまな事物の観察などが用いられる。また歴史学、文化人類学、社会学や宗教学などと密接に関連し、ひとつの研究が民俗学とそれらの領域にまたがるものになっていることもある。また、研究資料の多数を当該研究者のみが個人的に管理・保管している例が多く、研究資料の流通を容易にするためには、構造や項目についての標準化を図るということも重要な要素となる。しかしながら、民俗学研究においては、思弁的作業が大きな比重を占めること、また研究手法が個別のかつ非定型であることから「標準化」ということに対しては問題がある。従って研究資料の標準化ということが民俗学研究の「均質化」に繋がる恐れがあり、民俗学分野における「研究の均質化」は研究の質の低下に繋がるもので、避けるべき要素である。特に、インターネットの普及により、これらの研究分野においても情報発信や研究資料の流通、さらには研究の支援システムとしてIT技術の導入が必要不可欠であり、このような二律背反をいかにして解決するかということは大きな課題である。

以上の背景から非文字資料をデータベース化し情報発信する研究環境の構築が望まれ、客観的解析手段を得ることによって研究の裾野が広がると考える。

2.2 メタデータ

「データに関するデータ」のことをメタデータといい、メタデータを記述するための代表的なメタデータ規則に国際標準である Dublin Core がある²⁾。Dublin Core は主に図書館に所蔵されている図書目録のように、情報資源の対象物が既存メディアで固定可能であることを前提としている。しかしながら、本稿で扱う非文字資料は、その特異性から Dublin Core による記述のみでは不十分であると考えられる。非文字資料が保持するデータは、メタデータも包括して、一つのデータとして重要な情報を含んでいる。従って、非文字資料は Dublin Core では記述不能なデータを多数保持し、また、メタデータと「真のデータ」の関係性から研究対象における「真の情報」を読み解くことにより、民族文化研究の貴重な研究資料となる。

非文字資料を体系化する際に問題となるのは、民俗学分野では、これまで研究資料として作成・蓄積されてきたデータベースを個人的に管理・保管しているものが多いため、研究資料としての情報が共有・流通されていない点が挙げられる。非文字資料が保有する情報を体系化するためには、民俗文化研究の対象世界、領域を情報化して情報発信する必要がある。そこで、対象世界の概念を明示的に記述、表現するオントロジーを構築することで非文字資料の共有と流通が図られる。非文字資料は、対象世界が同一対象物であっても異なる概念が複数存在する。これは、非文字資料を体系化し情報化する際、非文字資料特有のデータが、それぞれの概念の情報化をする立場の社会的・歴史的・文化的背景などにより世界観が異なるためである。そこで、概念体系の整理として、オントロジー構築を提供することが重要となる。オントロジー構築を行うことにより、研究者間において共通の研究対象の共有と流通が図られる。

3. オントロジー

3.1 基礎理論

近年、知識処理の分野において、オントロジーに関する研究が盛んである。オントロジーは概念化の明示的な記述であり、知識システムが対象とする世界のオントロジーを構築することで、対象となる概念の合意

による知識の共有・流通への貢献が期待できる。オントロジーとは本来哲学用語であり、「存在に関する体系的な存在論」という意味を持つが、情報工学では、「概念の明示的な記述」と定義される³。オントロジーの定義のうち、情報工学としてのオントロジーの中核をなす部分は「存在という対象物だけでなく、具体的な存在対象世界として考察したものであり、そこに現れる概念と関係を明示的に示し、明確な意味定義を与えたもの」とされる。計算機によって実世界をモデル化する時、我々は実世界に存在する「概念の存在」を認識し、概念として抽出する「概念化」を行っている。この時、他の概念との違いや関係性を同定することで、その概念を特徴付けて、その概念が示す意味を無意識的に把握していると考えられる。通常は無意識的・暗黙的であるこのような「概念化」を明示的に記述したものがオントロジーと呼ばれている。本研究では、非文字資料の共有と流通においてオントロジーの果たす役割に注目し、「従来暗黙的だった知識とオントロジー基盤概念を分節化し、組織化したもの」と捉える。

3.2 トップレベルオントロジー

オントロジーは一般に、概念、概念分類 (taxonomy)、概念間の関係記述から構成されている。概念は階層的に組織化されることが多い。トップレベルオントロジーは、対象世界において共通に見られるオントロジーを構成するオントロジー記述要素であり、一般性と抽象度の高いオントロジーである^{4,5}。また、概念階層の上位として、実体、事象、時間、空間、概念、関係、属性などといった基本的概念の分類と意味付けが行われている。トップレベルオントロジーは、ものの同一性とは何か、部分と全体とは何か、といった基本的な関係性に関する理論が構築されている。

文献⁶では、トップレベルオントロジーを以下のように分類し、これらの組み合わせにより、全部で12種類の基本カテゴリで構成されている。

1. 第一分類：「Physical」、「Abstract」
2. 第二分類：「Independent」、「Relative」、「Mediating」
3. 第三分類：「Continuant」、「Occurrent」

3.3 オントロジー構築概念

オントロジーは対象世界の概念化を明示的に記述したものと定義される。

オントロジー構築とは対象世界の概念と概念間との関係を示すものである。

以下にオントロジー構築概念を示す。

- ・ 任意の視点でのオントロジー生成

オントロジーの持つ性質として、共通性・合意性が挙げられる。構築されたオントロジーは多数の合意を得るべきである。しかし一方で、個別性や特殊性の許容領域が問題となる。従って、情報や目的が多様化する対象世界では、任意の視点で生成されたオントロジーが有用であると考えられる。

- ・ オントロジー統合・変換

対象世界を、ある目的によって概念化する際に、同一対象世界において異なるオントロジーが構築される。このようなオントロジー構築がされた場合、目的に応じて、オントロジー間の統合や変換を行い、研究者が情報を付与する必要がある。従って、オントロジーの統合・変換は対象の同定や明示化、記述において重要な問題となり検討される必要があると考える。

4. 非文字資料におけるオントロジー構築

4.1 民具収集分類法における民俗学分野の概念体系

民俗学分野において、民具はそれぞれの地域や文化によって、その対象世界となる概念は多く存在する。特に分類体系法においては、対象とする世界が同一であっても、個別の目的や意図で分類した体系が多数存在する。表1は民具収集分類法の一部を示したものである。用途分類は、消費者分類法といわれ、消費者にとっての効用や用途による分類を基本としている。一方、機能分類は、主に民俗学統計基準として利用される分類体系法である。これらの同一対象世界における分類体系の違いは、目的や対象世界が、どの属性に含まれるかに注目するかによって、異なる概念化が行われた結果による。

対象物	用途	機能
茶碗	和食器	固型の食べ物を入れる容器
皿	洋食器	液体の食べ物を入れる容器
箸	和食器と共に使用	食べ物を口に運ぶ物
スプーン	洋食器と共に使用	食べ物を口に運ぶ物

表1 「用途分類」と「機能分類」

4.2 民具収集分類法におけるオントロジー構築

民族文化において研究対象とする世界は概念の集合体である。対象世界を記述する際には概念化を暗黙的に行う必要があり、概念化は対象とする世界の概念とその関係を示し、それらを体系化したものがオントロジーとなる。すなわち、オントロジーにおける全体概念とは、対象世界全体の概念を指し、関係概念とは、対象世界同士での関係性の概念を指す。

民族文化研究の一事例である民具収集法を体系化することは、概念を体系化し理解する上で重要であると考えられる。民俗学では、「茶碗」という物体に対して、単に「食べ物を入れる容器」という概念だけではなく、「食べ物を口に運ぶ物」という概念でもあることから、「スプーン」という概念も同時に成立する。すなわち、茶碗という食器の分類カテゴリにおいて、2つの意味概念を併せ持つ。このような認識概念の相違が、民俗学分野における民具というカテゴリの中で重要な意味を持つものと考えられる。

対象世界を概念化する際、一つの対象物でも複数の概念を持つ。オントロジーの体系化は、概念の明示的記述により成立し、オントロジーを構築することで、対象世界となる知識の共有と流通が行われる。図1に民具分類の体系図を示す。

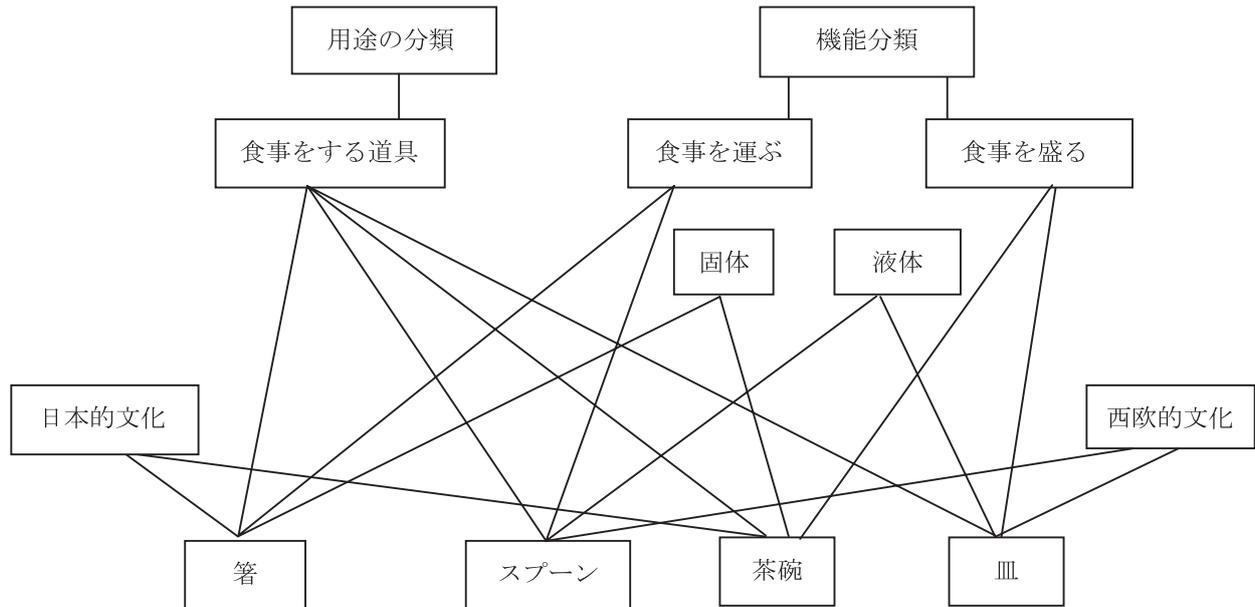


図1 民具分類体系

オントロジーにより体系化された民具収集分類法を、従来のトップレベルオントロジーを拡張することによりオントロジー構築を行った。図2に民具収集分類法におけるオントロジー構築を示す。なお、図中の「用途分類」および「機能分類」は、図1で示した「用途分類」と「機能分類」にそれぞれ相当する。

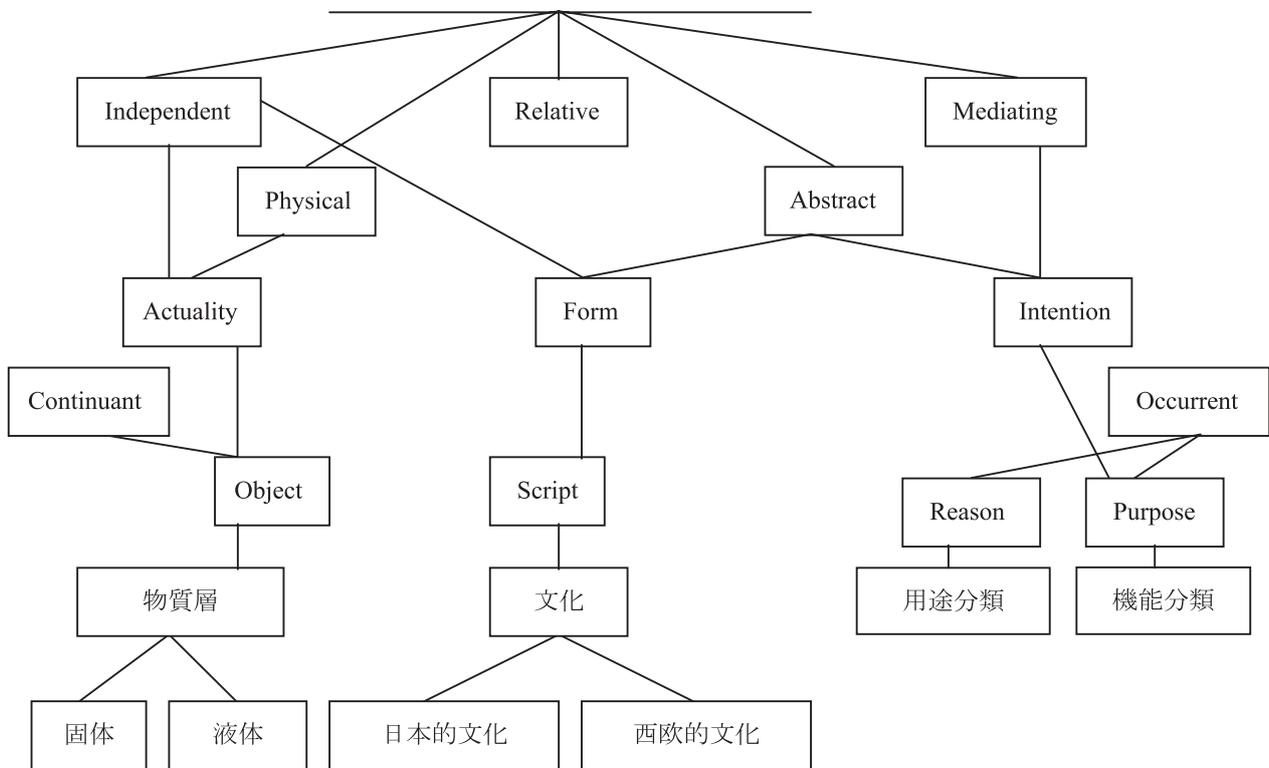


図2 民具収集におけるオントロジー構築

オントロジー基礎理論に基づいたトップレベルオントロジーによるオントロジー構築を示したことにより、民具収集分類法における以下の点が明示化された。

- ・一つの概念（茶碗）に対して、2つの上位概念（食べ物盛る・運ぶ）から異なる意味が多重継承される。
- ・もとの概念の意味（茶碗）の多重継承は全く異なる他のカテゴリ（スプーン）を含む。

5. 考察

民俗学分野で分類されている民具収集分類法から、用途分類と機能分類の2つを詳細化し、オントロジー基礎理論に基づくトップレベルオントロジーを構築した。また、定義内容を詳細化し、概念的性質の相違点の明確な分類を行った。これらの概念を導入することで、従来は専門的な対処がされていた概念定義の伝播を、汎用な枠組みで明示的に扱うことが可能となった。概念の明示化により対象世界が組織化・体系化されることで、オントロジーを介した知識の共有と流通の向上に貢献すると考える。

概念体系の構造や複数の体系間の関連を明示化することは、情報工学的オントロジーの観点において、システムと人間の共存による対象世界の共通理解を想定した場合に重要な役割を果たす。このような対象世界の根底となる概念は、一般的に暗黙的に共有していると仮定されているものである。オントロジーは、このような対象世界の背後の暗黙的な概念を明示化した記述であり、知識の共有と流通を向上させるものである。これらの暗黙的概念は、対象世界に関する知識や、意図を表していると考えられる。非文字資料における対象世界は、メタデータと「真のデータ」との関係記述から「真の情報」を推論で導き出す民俗学特有の資料である。対象とする世界の領域概念は、多数存在するが、対象とする範囲などオントロジーとして記述すべき内容は異なる。これは、対象世界の種類や領域が異なれば、必要とされる概念は自ずと異なるからである。また、同じ対象物であっても、ある概念についてどの程度まで、定義内容を記述するかにより内容も異なる。以下に、非文字資料に適したオントロジー構築を行う際、対象世界となる内容記述の検討項目を示す。

- ・ 専門知識の収集
- ・ 専門知識の分析・組織化（体系化）

「専門知識の収集」では、構築しようとしているオントロジーに必要な専門知識を収集し、オントロジーで記述する概念を明らかにすると考える。収集された専門知識に含まれる概念が多数存在するが、オントロジー記述される概念は、対処とする範囲などにより異なる。そこで、収集された多くの専門知識から目的に応じて必要な概念を抽出する必要がある。「専門知識の分析・組織化」では、収集された専門知識の内容を分析し、抽出された概念を体系的に組織化する。上記に示した内容記述を経て記述されたオントロジーを知識として利用するには、オントロジー構築が不可欠であり、対象世界の共有と流通に大きく貢献すると考える。

6. おわりに

本稿では非文字資料とは何かを定義し、オントロジーの概要について述べ、非文字資料の事例を用いて、オントロジーが知識共有・流通に有効であることをオントロジー構築により示した。同じ対象世界を領域として対象としていても、目的により複数のオントロジーが存在する。よって、既存のオントロジーを蓄積・共有し、別のオントロジーを構築することが求められる。構築したオントロジーを利用することで、オントロジーに基づいた知識を記述することができる。オントロジーに基づいて知識の記述を行うには、オントロ

ジーで記述された背景情報に基づきモデルを構築していくことが必要不可欠であると考え。そのためには、形式化や標準化による情報の共有化ではなく、情報の内容とその利用（知識・知恵の伝承）を共有し、活用していくことが重要である。非文字資料の共有と流通は、知識の共有・流通であり、民族文化を継承していく上でも非常に重要な要素を含んでおり、取り組んでいく必要性は十分にあると考える。本稿が、民族文化の情報発信と問題提起になれば幸いである。

謝辞

本研究に協力していただいた、神奈川大学大学院工学研究科電気電子情報工学専攻の木下慶子さんに感謝する。

参考文献

- 1 T. R. Gruber: A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, Vol.5, No.2, pp.199-220 (1993).
- 2 <http://www.dublincore.org/>.
- 3 溝口理一郎：オントロジー工学、オーム社 (2005)。
- 4 J. F. Sowa: Distinction, Combination, and Constraints, *Proc.of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing* (1995) .
- 5 J. Top and H. Akkermans: Computational and Physical Causality, *Proc.of 12th International Conference on Artificial Intelligence*, pp.1158-1163 (1991) .
- 6 J. F. Sowa: Knowledge Representation: Logic, Philosophical, and Computational Foundations, *Brooks/Cole* (2000) .