

文字の指示概念に関する試論

白須裕之

東京大学大学院人文社会系研究科

次世代人文学開発センター

概要

本稿は文字の指示機能に注目して、文献学的な研究に使用できるようなテキスト表現を探索する。デジタルアーカイブズとしての史料等が持つべき機能には、原本をどのように解釈したのかを明示できるような枠組みが必要である。特に漢字文献を対象とした場合、文字（漢字）の解釈についての情報は大変重要である。本稿は符号化によらない文字のモデルである Chaon モデルを理論化の基礎とした。Chaon モデルでは文字をその性質（文字素性）の束として表現するが、文字概念が持つ状況依存性に注目すると、文字素性の見え方は文脈に依存する。本稿では Chaon モデルにおける解釈の再考により、文字概念を単独で扱わず、文字体系の中で表現するモデルを構築した。

On Reference of Kanji Characters

SHIRASU Hiroyuki

The Center for Evolving Humanities,

Graduate School of Humanities and Sociology, The University of Tokyo

Abstract

This paper presents a model of Kanji characters in order to represent digital texts in the philological researches. To understand the necessary and sufficient conditions for achieving philological texts we need to discuss formalization on referencing of Kanji characters. In this paper, we adopt the *Chaon model* as the representation of Kanji characters, which does not use coded characters, and on which any character is represented by the set of its character features. In addition we discuss representation of character systems and their interoperability. This paper proposes a formalization of semantic interoperability of character systems based on the Barwise-Seligman theory of information flow.

1 はじめに

文字処理の分野において、Unicode 等の大規模な符号化文字集合による技術が一定の成果をあげ、更に符号化によらない文字オントロジーによる手法 (Chaon モデル) と環境 (CHISE)[7][9] が大きな評価と期待を集めている。本稿は文字概念の状況依存性に注目して、文字の状況理論的な意味論の構築を目指すとともに、そのような知見から新たな文字オントロジーの設計を模索する。

文字、特に漢字がある特定の場所に存在するとき、どのような情報が得られるであろうか？例えば、書物や巻物等のテキストの中に出現する文字、或いはブラウザで表示されている Web 上の文字、検索サービスの検索窓に入力された文字、石に刻まれていたり、筆で書かれていたり、文字は様々な状態で我々に与えられる。

文字についての解釈はこのような文字の出現の様に依存しているとともに、解釈者の意図や状況にも

大きく依存している。本稿は文字がある文脈に存在していて、それを読む、或いは見る行為によって、どのような情報が得られるかを状況論的に議論し、文字に関する知識についての理論化を試みるものである。

2 言述の基本構造

本節では言述の基本機能を整理し、次節以降でここでの視点から発話と文字についての状況意味論を検討しようと思う。言述に共通の基本機能には、環境世界の事物を「言語」の使用によってシンボル化する機能と、「言語」が常に何かについて語るという働き、即ち、「言語」の指示の機能がある。即ち、

1. 言語の超越論的機能、
2. 言語の志向的機能ないしは志向性

とである。まず、第一の機能について文献 [6] より引用しよう。

最初の超越論的機能とは、言葉によって対象を意味づけ、指定する働きを指す。... われわれが環境世界の中で出会う諸々の事物を言語の網の目の中に掬い取り、分節化する機能のことである。... 言語を使う主体の側から言えば、それは「シンボル化能力」と言い換えることができる。(文献 [6] 208 頁)

また、第二の機能について次のように述べている。

第二の志向的機能とは、言語が言語外の対象に向かって超越する働きのことである。すなわち、言語が常に何かについて語るという働きをもつことであり、... その中核をなすのは言語の指示 (reference) 機能であろう。(文献 [6] 209-210 頁)

言語の志向性は、その分節化の様態によって以下のような「世界」の層を形成する。

1. 「身分け空間」 — 身体的な特性 (感覚、運動) によって分節化される世界であり、生命維持、生活の必要等の目的連関に根拠を持つ。
2. 「言分け空間」 — 言語使用による分節された世界であり、文化的な欲望に根拠を持つ。
3. 「テキスト空間」 — 理論的な枠組みによって構成された世界である。この理論的な枠組みには物理学のような理論に加えて、本稿で対象とする「文字」についての知識体系も含まれる。

分節化の様態は、各世界において指示行為の範囲にも影響を与える。「身分け空間」では、身体的行動の及ぶ範囲になり、〈今、ここ〉のような身体を中心とする軸を形成する。「言分け空間」では、身体的制約から離れた時空間が範囲となり、〈指標軸〉による相対化が行なわれる。また、「テキスト空間」では理論の対象、虚構の対象が指示行為の対象となる。

対象への指示機能の前提として、第一の超越論的機能による対象の類型的分節化を必要とする。指示の可能性は、言語共同体において生成維持されるコミュニケーションの連鎖を基礎とする¹。文字を考える上でも、このような指示機能のモデル化が必要である所以である。

¹ 「因果連鎖」とは言語共同体の歴史を貫く固有名使用の〈伝統〉なのであり、「指示」とは... 歴史的・社会的文脈の中での公共的な出来事なのである。(文献 [6] 頁 216-217)

² 例えば、文献 [11][13] を参照。

³ 厳密には、自然言語表現の意味は状況のタイプの間の関係で与えられる。

3 文字の指示機能

3.1 自然言語の状況意味論

状況意味論は 1980 年代に開発された意味論²で、自然言語表現の意味を状況間の関係として提出する³。我々は世界の一部としての状況に注目し、状況の間に類似性を見るとき、そこから情報を読みとる。このような情報を運ぶものの一つとして言語的表現を見るのである。ここでは状況が一級のオブジェクトとして扱われている。書記言語或いは文字の解釈において考慮すべき点を明確にするために、以下で発話についての状況意味論を整理しておく。

発話された文章の意味は、「談話状況」「埋め込み状況」「資源状況」と、文章が記述している「記述状況」との関係である。「談話状況」とは発話がなされている話し手と聞き手の現場状況で、その時空間位置等が関係する。「埋め込み状況」は発話を取り囲む状況で、発話された文章に含まれている指示物との結合を決める。「資源状況」は対話者の世界についての知識、「記述状況」は発話者が記述しようと意図している状況である。

状況は様々な出来事から構成される世界であるが、前節の議論で見たように、ここでの状況は「見分け空間」、「言分け空間」のような実在論的な空間である。

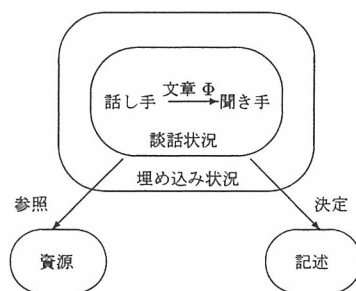


図 1: 発話行為に関わる状況

3.2 文字について状況意味論

書記言語に関わる世界は「言分け空間」及び「テキスト空間」である。特に文字の解釈に関わる世界は、「テキスト空間」の一部としての文字空間である。これは文字についての事実が作る空間であり、空間の

一部として、文字の解釈を決める状況が存在する。以下、状況の表現について考察する。

文献 [7][9] で提唱されている Chaon モデルは、文字の表現に符号化を使用せず、また、抽象文字、字形、字体などの文字の抽象度に依存するモデルも採用しない。文字の意味論的な側面に光を当て、文字についての情報、即ち「文字素性」の束によって文字を表現する。これは文字についての情報がどのような形で捉えられるかを、明示的に宣言するという意味で画期的な提唱と言える。

本稿は状況の表現を Chaon モデルに求める。即ち、ある文字 x がある文字素性 f として値 v を持つということを、ここでは事実として扱い、 $\langle f, x, v, p \rangle$ と書く。但し、 p は 1 か 0 であり、この関係が成り立つときは 1、成り立たないときは 0 とする。また、この事実が状況 s で成り立っているとき、以下のように書く。

$$s \models \langle f, x, v, p \rangle.$$

自然言語の発話と同様に、文字の意味も状況間の関係である。但し、関わる世界、状況が異なっている。

以下、文字の解釈に関わる状況について述べる。ある特定の文字 Φ が読み手 (受け手) に与えられたとする。

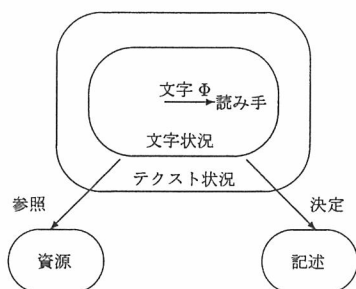


図 2: 文字解釈に関わる状況

- 文字状況 — 読み手が文字を読むときの状況、
- テキスト状況 — 文字を含むテキストが置かれている状況、
- 資源状況 — 文字の解釈で参照される知識資源、
- 記述状況 — 文字の解釈によって記述される状況。

テキスト状況には言語、スクリプト、書記系の違いや、デジタルテキストの場合はコーディングシステムが含まれる。発話の解釈と文字解釈に関わる状況で

は様々な点異なるが、これについては別稿で述べる予定である。但し、テキスト状況が文字 Φ の用例を形成すること、資源状況としては個々の文字についての知識、及びその形式化である文字体系が参照されることを強調しておこう。

4 文字の状況依存性の表現

4.1 文字の電子化の事例

文献学的な研究に提供できるような電子テキストでは、原本を翻刻するに際して、どのような解釈を文字に与えたかが明示されている必要がある。このような解釈は前節で示唆した状況間の関係である。ここでは状況間の関係を与える文献学的な表現、即ち、文字の電子化の事例について見てみよう。

コード表をだましまし使う法 文献 [5] では、文献学的な研究に提供できるような電子化テキストを、既存の符号化文字集合を使って表現する便宜的な方法を示している。これには次のような二段階の解釈を伴う。

1. テキスト解釈者 (翻刻者) が原本に与えた解釈としての「文字観念」、
2. その「文字観念」のコード表との関係 (コードポイントの明示)。

この方法の利点は既存の文字コードを使うため、検索等の文字についての操作を実現する必要がないことである。ここから「文字観念」と複数の「文字観念」間の関係をモデル化する必要性を読みとることができる。

包摂規準による解釈 文献 [2] では、具体的な古辞書を対象として、辞書の各漢字を JIS の包摂規準及びその拡張によって「大漢和辞典」の漢字に関係付けている。ここからは「文字観念」を対応付ける方法として、包摂規準のような規則をモデル化する必要性を読み取ることができる。

Chaon モデルによる異体字表現 文献 [1] では、Chaon モデルを使った異体字表現のモデル化を議論している。ここで考慮すべきは字義階層とその階層に依存した文字間関係である。また、出典依存の情報の表現法も問題となる。文献 [1] では階層的素性名の

方式を採用する。即ち、文字間の関係を出典等を表現するドメイン識別子によって区別する。

ここからは「大漢和辞典」のような複数の文字(漢字)概念を総合した辞典をどのように扱うかを問題とすることができる。本稿での立場を述べておくと、異体字関係のような文字間の関係はある一つの文字体系を定義し、文字体系が異なれば文字概念は異なるとして扱う。但し、文字体系を統合するための方式をモデル化することによって、統合された文字概念には統合の対象となった文字概念の階層化が実現される。こうして総合的な辞典のモデル化によって、複数の文字体系をどのように総合したのかを知識化できる。

4.2 「文字素性」再考

文献[2]では、古辞書における漢字の字体注記とその処理手段について考察している。文字素性は漢字の価値規範、及び形・音・義についての注釈等を表現するものであるが、各々の特徴を纏めると以下のようになる。

音注は反切等の方法による体系的概念を持つ。形についての注である字体注記は、異形同字、類形別字の関係、漢字字形の規範や漢字の正しい用法を示すのが目的である。字体注記は文献相互の伝承がその構成に重要な役割を果たし、端的な様式による形式性を特徴とする。このように文字の性質は、文字素性が値を持つという枠組みだけでは捉え難い面、即ち、体系的、文献相互性を持つことが分かる。

文献[4]では、文字(漢字)を捉える枠組みとして、「形音義内在論」と「文脈依存論」の対立の構図を描いてみせ、それを超えるために漢字シソーラスの構築法を議論している。本稿はそれに対する一つの実現を試みる。

これまでのテキスト表現の例を検討してみると、文献学的に意味のあるテキスト表現のためには、以下のような文字の指示概念が必要であることが分かる。

1. テキスト解釈者(翻刻者)が原本に与えた解釈として、「文字観念」を表現できること、
2. 言語や時代、文献ごとに異なる「文字観念」を表現でき、またそれを統合できること、
3. 「文字観念」は文字間の関係を含む「文字体系」として扱えること、
4. 文字素性の性質を記述するための言語を備えること。

以降の節では、Chaon モデルを前提として、これらの要求を文字の指示機能としてモデル化する方法を検討する。

5 Chaon モデル再考

今まで議論した文字の解釈及びテキスト表現への要求を纏めてみると、以下のようになる。Chaon モデルに解釈は含めないが、文字が置かれている状況、読み手の置かれている状況、態度等が、文字を表現している文字素性の見え方に影響を与えている。従って、文字素性の見え方がどのように文脈によって変化するかを記述する方式を検討する必要がある。

文献[1]では素性名の階層化と素性値の構造化を検討していた。一つの Chaon モデルを前提とした場合には、このような方式を取らざるを得ないが、本稿では前半で検討したテキストを読む行為のモデル化との親和性を考慮して、文字の解釈が状況間の関係を決めるという発想からモデル化を行なう。

文字の表現は Chaon モデルで与えるが、ある文字の文字素性は他の文字との関係で決まるという発想から、文字を孤立して定義するのではなく、一つの文字体系として提出する。従って、文字素性についての性質を記述するための言語と、その解釈としての Chaon モデルという仕組みで文字体系を構成する。

文字素性についての性質を表現する言語とは、例えば、「形」に注目した場合は包摂規準等の規則、「義」に注目した場合は漢字シソーラス、「音」に注目した場合は反切等の韻書の体系を意味する。これらを文字オントロジーと呼び、その解釈を文字具象オントロジーと呼ぶ。次節で文字体系のモデル化の方法を議論しよう。

6 文字体系とその統合

今までに文字の指示機能をどのようにモデル化すべきかについて、その要求事項を見てきた。ここでは実際に「情報の流れ」論的なオントロジーを使って、Chaon モデルの上にそのような指示機能のモデルを構成することを試みたい。

その概要は以下である。一つの文字体系は、Chaon モデルに対する指示機能をモデル化したオントロジーで表現する。この文字体系は一つの文脈において相互に関係しあう文字概念を定義する。また、文字体系間の文献相互性及び文字同定はオントロジー射で表現する。なお「情報の流れ」論的なオントロジーにつ

いては、節 A.2 に纏めたので参照願いたい。

6.1 Chaon モデルの形式化

最初に Chaon モデルを形式化する。文字素性の集合 F が与えられたとする。但し、全ての文字素性 $f \in F$ にはその値の領域 $\text{dom}(f)$ が決まっているとする。今、文字素性とその値の全ての組からなる集合を Fv とする⁴。即ち、

$$Fv = \{\langle f, v \rangle : f \in F, v \in \text{dom}(f)\}.$$

このとき、 F によって構成される Chaon モデルを $\text{Ch}(F) \subseteq \mathcal{P}(Fv)$ とする。

6.2 文字体系

まず、文字オントロジーを C とする。前節で議論したように、これには文字の性質を議論するための言語を定義する。例えば、この言語において、文字におけるある種のシソーラスは順序として、異体字や文字規範等は関係概念として、包摂規準等は規則として表現する。

この文字オントロジー C に従って構成した文字具象オントロジー $\tilde{C}$ が文字体系を与える。この構成には Chaon モデルを使用する。文字オントロジー C の Chaon モデル $\text{Ch}(F)$ における解釈を、二つの分類 $\mathbf{C} = \langle \text{Ch}(F), C, \models_C \rangle$ 及び $\mathbf{R} = \langle \text{Ch}(F)^+, R, \models_R \rangle$ で定義する。このとき、文字具象オントロジーは $\tilde{C} = \langle \mathbf{C}, \mathbf{R}, \leq_C, \leq_R, \sigma \rangle$ である。

さらに状況意味論で使用する事実の定義は以下のようを与える。文字概念 $c \in C$ 及び文字素性 f とその値 v に関して、 $\tilde{O} \models \langle \langle f, c, v; 1 \rangle \rangle$ は、 $\exists x \in \text{Ch}(F). x \models_C c, \langle f, v \rangle \in x$ で与える。また、関係素性 $r \in R$ に対して $\sigma(r) = \langle c, d \rangle$ であるとき、 $\tilde{O} \models \langle \langle r, c, d; 1 \rangle \rangle$ は、 $\exists x, y \in \text{Ch}(F). x \models_C c, y \models_C d, \langle x, y \rangle \models_R r$ で与える。

6.3 解釈による文字体系の構成

テキストを読むときには、ある文字体系を前提とする訳であるが、その読みによって新しい文字解釈が生れる場合がある。この新しい文字体系の構成を議論しよう。これはテキストの解釈で得られた文字オン

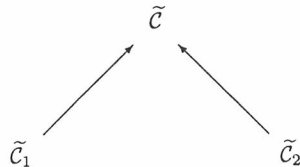
トロジーから、前提とした文字具象オントロジーへのマッピングを構成することである。

まず、前提とする基準の文字体系として、文字具象オントロジー $\tilde{C}$ が与えられたとする。テキストの読みによって得られた文字の解釈は、前節で議論したように文字（漢字）シソーラスで与えられ、それは文字オントロジー C によってモデル化される。ここでの問題は文字具象オントロジー $\tilde{C}$ を、文字オントロジー C の言葉で具体化することである。これには節 A.2 で述べた具象オントロジーの構成法を利用する。

6.4 文字体系の統合

二つの文字体系の統合を定義する方法を述べる。ここでの議論は、例えば複数の字書を解釈した各々の文字体系を、必要とする文字オントロジーに合せて統合する場合に相当する。

今、二つの文字体系を各々表現する二つの具象文字オントロジー $\tilde{C}_1, \tilde{C}_2$ が与えられたとする。このとき、二つの文字体系を統合するための言語である文字オントロジー C を設計し、共通の文字具象オントロジー $\tilde{C}$ を構成する。これらはチャンネルを構成している必要がある。下図はその様子を示している⁵。 $\tilde{C}$ は二つの文字体系をどのように統合したかを知識化している。



二つ以上の文字体系についてもこの構成法を順次適用して、ある種の総合的な辞書を作ることができる。その辞書はどのような文字体系を含み、それらの統合の様子を知識化している。

6.5 具体例について

紙数の関係で、二三の例の概要のみを述べる。

符号化文字集合の場合を考えてみよう。この文字集合の作る文字体系の文字オントロジーを C_{CCS} とする。このとき文字素性の集合 F によって、 C_{CCS} を具体化した文字具象オントロジー $\tilde{C}_{\text{CCS}, F}$ を構成できる。

符号化は Chaon モデルから自然数の集合への関数 $f: \text{Ch}(F) \rightarrow \mathbf{N}$ で表現できる。この関数を使って自然に $\text{Ch}(F)$ の代りに $\mathbf{N}$ をベースにした、 C_{CCS} -具象オ

⁴文字素性の集合 F には文字間の関係素性は含まれていない。文字間の関係素性は文字オントロジーで、また、その実現は文字具象オントロジーで定義される。文字概念を決める際に関係素性も使用するには別の形式化が必要である。

⁵実際の構成については、紙数の関係で他の機会に譲りたいと思う。

ントロジー $\tilde{C}_{CCS,N}$ を構成できる。 $\tilde{C}_{CCS,F}$ と $\tilde{C}_{CCS,N}$ は共に文字オントロジー C_{CCS} の具体化である。

また、ある古辞書の解釈から得られる文字体系の文字具象オントロジーを $\tilde{C}_{\text{古}}$ とする。これを今構成した符号化文字集合で解釈するには、文字体系 $\tilde{C}_{CCS,F}$ と $\tilde{C}_{\text{古}}$ の統合が必要である。

7 おわりに

本稿は Chaon モデルを基礎として、文字の指示機能を表現するために、「情報の流れ」論的なオントロジーを使用した。本稿は理論的な側面に重点を置いたが、この枠組みで様々な文字現象を理論化し、文字に関する知識を陽に記述することは今後の研究課題としたい。また、文字に関係する解釈は非常に多岐に渡るため、ここではそのほんの一部をモデル化できたにすぎない。例えば、文字体系で表現できないような、出典に依存する情報については手を付けられていない。

前半では書記言語に対する状況意味論のスケッチを行なった。文字と文章では意味論の構築に大きな隔りがある。それは対象となる意味空間が異なるからである。書記言語に対する意味論は、この二つの意味空間の相互関係を記述できる必要がある。また、本稿では文字素性の見え方が、テキストの読みに従って動的に変化する様子をモデル化できていない。これらについても今後の研究課題としたい。

本稿では、文献学的な文字概念を表現するために、表象としての文字の解釈として Chaon モデルを使用した。文献 [8] では一般キャラクター論の立場から、人文情報学のあるべき姿を、視覚的表現と論理的表現の関係性に見ている。本稿がその解明の手掛かりとなることを期待する⁶。

本稿で取り上げることができなかったが、古辞書のデジタルテキスト化、漢字情報のデータベース化が多くの研究者の方々の手で行なわれている。このような試みを統一的に理解し、そのデータを活用する道として、本稿の枠組みのような再解釈が将来的に役立つであろう。本稿のアイデアを検証し、データベース化を実現するために、文字論・古辞書の研究者の方々との協力をお願いして、本稿を了えたいと思う。

謝辞 Chaon モデルに関して色々ご教示いただいた守岡知彦さんに感謝します。次世代人文学開発センターでお世話になっている下田正弘先生に感謝します。永田研宜さんには多くの助言を頂きました。概要論文に対して査読頂い

たお二人の方からは有益なコメントを頂戴しました。また、お世話になった多くの方の助力に感謝いたします。最後に妻留美と家族に感謝します。

参考文献

- [1] 秋山陽一郎, 守岡知彦, 浦田衣里: 階層的素性名を用いた異体字記述の試み, 情報処理学会研究報告, 2005-CH-67, 2005.
- [2] 工藤祐嗣: 古辞書における漢字の字体注記と処理手段, 博士学位論文, 北海道大学大学院文学研究科, 2003.
- [3] 白須裕之: 記号機能としてのアーカイブズ, 情報処理学会 人文科学とコンピュータシンポジウム「じんもんこん」2008, 2008.
- [4] 當山日出夫: 漢字コードをめぐる諸概念について— 漢字シソーラスは可能か?, 情報処理学会研究報告, 1994-CH-23, 1994.
- [5] 豊島正之: 「JIS に無い字」をめぐる, 月刊しにか 3(2), 大修館書店, 1992.
- [6] 野家啓一: 物語の哲学, 岩波現代文庫, 岩波書店, 2005.
- [7] 守岡知彦: 文字オントロジーに基づく文字処理について, 情報処理学会研究報告, 2006-CH-72, 2006.
- [8] 守岡知彦: キャラクターを考える, 人文情報学シンポジウム, 京都大学人文科学研究所, 2007.
- [9] 守岡知彦, 師茂樹: 文字素性に基づく文字処理, 情報処理学会研究報告, 2004-CH-62, 2004.
- [10] 師茂樹: 思想史としての文字情報処理 — 問題提起として, シンポジウム「文字情報処理のフロンティア: 過去・現在・未来」予稿, 2004.
- [11] J. Barwise, J. Perry: Situations and Attitudes, MIT Press, 1983. (邦訳 土屋俊他訳: 状況と態度, 産業図書, 1992.)
- [12] J. Barwise, J. Seligman: Information Flow: The Logic of Distributed Systems, Cambridge University Press, 1997.
- [13] K. Devlin: Logic and Information, Cambridge University Press, 1991.

⁶なお芸術論からの回路を開く可能性として、文献 [3] を参照下さい。

- [14] Y. Kalfoglou, M. Schorlemmer: IF-Map: An Ontology-Mapping Method based on Information-Flow Theory, Journal on Data Semantics 1(1):98–127, 2003.

A 数学的な補足

A.1 「情報の流れ」理論の概要

本稿では情報を扱うための道具として、チャネル理論の基本的な部分を使用する。Shannon 等による情報理論が情報を量的に捉えるのに対して、チャネル理論は情報を定性的に扱い、また、「情報の流れ」に注目する。ここではその基本概念を纏めておく⁷。

定義 1 (分類 Classification, 制約 Constraint)

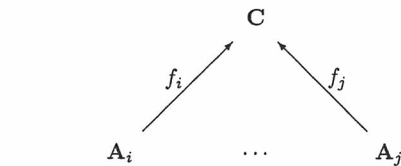
分類 $A = \langle \text{tok}(A), \text{typ}(A), \models_A \rangle$ とは、トークンの集合 $\text{tok}(A)$ 、タイプの集合 $\text{typ}(A)$ 、その上の二項関係 $\models_A$ からなる。

$\Gamma, \Delta \subseteq \text{typ}(A)$ とする。任意の $a \in \text{tok}(A)$ に対して、 $\forall \alpha \in \Gamma. a \models_A \alpha$ ならば $\exists \alpha \in \Delta. a \models_A \alpha$ が成り立つとき、 Γ と Δ は A に対する制約と言い、 $\Gamma \vdash_A \Delta$ と記す。

定義 2 (情報射 Infomorphism) 分類 A から分類 B への情報射 $f : A \rightarrow B$ とは、二つの関数 $f^* : \text{typ}(A) \rightarrow \text{typ}(B)$ と $f_* : \text{tok}(B) \rightarrow \text{tok}(A)$ の対で、任意の $b \in \text{tok}(B)$ と $\alpha \in \text{typ}(A)$ に対して、 $f_*(b) \models_A \alpha \Leftrightarrow b \models_B f^*(\alpha)$ を充すものを言う (以下、適宜 $*$ を省略する)。

$$\begin{array}{ccc} \text{typ}(A) & \xrightarrow{f^*} & \text{typ}(B) \\ \models_A \uparrow & & \uparrow \models_B \\ \text{tok}(A) & \xleftarrow{f_*} & \text{tok}(B) \end{array}$$

定義 3 (チャネル Channel) $\{A_i\}_{i \in I}$ を分類の族とする。チャネル $C = \{f_i : A_i \rightarrow C\}_{i \in I}$ とは、 C への情報射の族であり、 C をその核と言う。



⁷詳しくは文献 [12] を参照下さい。

⁸オントロジー工学では様々なオントロジー概念が導入されているが、本稿ではここで定義する意味で用語「オントロジー」を使用する。

⁹文献 [14] で導入されているようなシークエントで表現可能な、 C, R 上の演算子をオントロジーに含めてもよい。

¹⁰ σ には $\leq_R$ を保つような制約を設けても良いが今は詳細は省く。

チャネル C において、 $a \models_{A_i} \alpha$ が $b \models_{A_j} \beta$ という情報を運ぶとは、 $\exists c \in \text{tok}(C). f_i(c) = a \wedge f_j(c) = b$ かつ、 $f_i(\alpha) \vdash_C f_j(\beta)$ であることを言う。

定義 4 (シークエント sequent, 分割 partition)

集合 Σ が与えられたとする。 Σ の部分集合の組 $\langle \Gamma, \Delta \rangle$ を Σ のシークエントと言う。また、シークエント $\langle \Gamma, \Delta \rangle$ が $\Gamma \cup \Delta = \Sigma', \Gamma \cap \Delta = \emptyset$ を充すとき、 Σ' の分割と言う。

定義 5 (帰結 consequence, 理論 theory)

二項関係 $\vdash \subseteq \mathcal{P}(\Sigma)$ を Σ 上の帰結関係と言う。このとき対 $T = \langle \Sigma, \vdash \rangle$ を理論と言う。また、 $\text{typ}(T) = \Sigma$ と記す。シークエント $\langle \Gamma, \Delta \rangle$ が $\Gamma \vdash \Delta$ を充すとき、 T -無矛盾 T -consistent であると言う。

定義 6 (正則理論 regular theory) 理論 $\langle \Sigma, \vdash \rangle$ が以下を充すとき正則であると言う。

- Identity: $\alpha \vdash \alpha$,
- Weakening: $\Gamma \vdash \Delta \Rightarrow \Gamma, \Gamma' \vdash \Delta, \Delta'$,
- Global Cut: Σ' の任意の分割 $\langle \Sigma_0, \Sigma_1 \rangle$ に対して、 $\Gamma, \Sigma_0 \vdash \Delta, \Sigma_1$ ならば $\Gamma \vdash \Delta$ 。

定義 7 (理論から生成される分類) 正則理論 T

が与えられたとき、 T から生成される分類を $\text{Cla}(T) = \langle \text{tok}(\text{Cla}(T)), \text{typ}(T), \models_{\text{Cla}(T)} \rangle$ とする。但し、 $\text{tok}(\text{Cla}(T))$ は $\text{typ}(T)$ の無矛盾な分割全体の集合で、また、 $\langle \Gamma, \Delta \rangle \models_{\text{Cla}(T)} \alpha \iff \alpha \in \Gamma$ とする。

A.2 「情報の流れ」論的オントロジー

文献 [14] では、オントロジー工学で使用されているオントロジーとそのマッピングを、チャネル論を使って理論化している。ここではその概念を拡張して、文字オントロジーを表現するのに適したオントロジー概念を構築する⁸。

定義 8 (オントロジー Ontology)

⁹オントロジー $\mathcal{O} = \langle C, R, \leq_C, \leq_R, \sigma \rangle$ とは、 C, R が各々概念記号、関係記号の有限集合であり、 $\leq_C$ は C 上の半順序、 $\leq_R$ は R 上の半順序である。また、 $\sigma : R \rightarrow C^+$ は関係記号にアリティを指定する関数である。ここで $C^+ = \bigcup_n C^n$ とする¹⁰。

定義 9 (具象オントロジー Populated Ontology)

¹¹具象オントロジー $\tilde{\mathcal{O}} = \langle C, R, \leq_C, \leq_R, \sigma \rangle$ とは、 $C = \langle X, C, \models_C \rangle$ 及び $R = \langle X^+, R, \models_R \rangle$ が各々分類で、 $\mathcal{O} = \langle C, R, \leq_C, \leq_R, \sigma \rangle$ がオントロジーであるときを言う。このとき $\tilde{\mathcal{O}}$ を $\mathcal{O}$ -具象オントロジーと言うことにする。また、 $\tilde{\mathcal{O}}$ が正しい (correct) とは、任意の $x, x_1, \dots, x_n \in X, c, d \in C, r \in R, \sigma(r) = \langle c_1, \dots, c_n \rangle$ に対して以下が成り立つときを言う。

$$\begin{aligned} x \models_C c, c \leq_C d &\Rightarrow x \models_C d, \\ \langle x_1, \dots, x_n \rangle \models_R r, r \leq_R s &\Rightarrow \langle x_1, \dots, x_n \rangle \models_R s, \\ \langle x_1, \dots, x_n \rangle \models_R r &\Rightarrow x_i \models_C c_i \\ &\quad (i = 1, \dots, n). \end{aligned}$$

以下では具象オントロジーは正しいものだけを扱う。

定義 10 (オントロジー射 Ontology Morphism)

二つのオントロジー $\mathcal{O} = \langle C, R, \leq_C, \leq_R, \sigma \rangle$ と $\mathcal{O}' = \langle C', R', \leq_{C'}, \leq_{R'}, \sigma' \rangle$ に対して、オントロジー射 $\langle f^*, g^* \rangle : \mathcal{O} \rightarrow \mathcal{O}'$ とは、二つの関数 $f^* : C \rightarrow C'$ と $g^* : R \rightarrow R'$ の対であり、任意の $c, d \in C, r, s \in R, \sigma(r) = \langle c_1, \dots, c_n \rangle$ に対して以下の性質を持つ。

$$\begin{aligned} c \leq_C d &\Rightarrow f^*(c) \leq_{C'} f^*(d), \\ r \leq_R s &\Rightarrow g^*(r) \leq_{R'} g^*(s), \\ \sigma'(g^*(r)) = \langle c'_1, \dots, c'_n \rangle &\Rightarrow c'_i \leq_{C'} f^*(c_i) \\ &\quad (i = 1, \dots, n). \end{aligned}$$

$$\begin{array}{ccc} R & \xrightarrow{g^*} & R' \\ \sigma \downarrow & & \downarrow \sigma' \\ C^+ & \xrightarrow{(f^*)^+} & (C')^+ \end{array}$$

ここで以下のような問題を考えてみよう。二つのオントロジー $\mathcal{O}, \mathcal{O}'$ と、その間のオントロジー射 $\langle f^*, g^* \rangle : \mathcal{O} \rightarrow \mathcal{O}'$ 、及び $\mathcal{O}'$ -具象オントロジー $\tilde{\mathcal{O}'}$ が与えられたとする。このとき適切な $\mathcal{O}$ -具象オントロジー $\tilde{\mathcal{O}}$ を構成するにはどうすれば良いだろうか？ そのために以下の定理を準備する¹²。

定理 1 $\mathcal{O} = \langle C, R, \leq_C, \leq_R, \sigma \rangle, \mathcal{O}' = \langle C', R', \leq_{C'}, \leq_{R'}, \sigma' \rangle$ をオントロジー、 $\tilde{\mathcal{O}'} = \langle C', R', \leq_{C'}, \leq_{R'}, \sigma' \rangle$ を $\mathcal{O}'$ -具象オントロジーとする。但し、 $C' = \langle X', C', \models_{C'} \rangle, R' = \langle (X')^+, R', \models_{R'} \rangle$ とする。また、二つの情報射 $f : B \rightarrow C', g : Q \rightarrow R'$ が与えられたとする。もし $\langle f^*, g^* \rangle : \mathcal{O} \rightarrow \mathcal{O}'$ がオントロジー射であるならば、任意の $x, x_1, \dots, x_n \in X', c, d \in C, r, s \in R, \sigma(r) = \langle c_1, \dots, c_n \rangle$ に対して以下が成り立つ¹³。

$$\begin{aligned} f_*(x) \models_B c, c \leq_C d &\Rightarrow f_*(x) \models_B d, \\ g_*(x_1, \dots, x_n) \models_Q r, r \leq_R s &\Rightarrow g_*(x_1, \dots, x_n) \models_Q s \\ g_*(x_1, \dots, x_n) \models_Q r &\Rightarrow f_*(x_i) \models_B c_i \\ &\quad (i = 1, \dots, n). \end{aligned}$$

$$\begin{array}{ccc} C & \xrightarrow{f^*} & C' \\ \models_B \downarrow & & \downarrow \models_{C'} \\ Y & \xleftarrow{f_*} & X' \end{array} \quad \begin{array}{ccc} R & \xrightarrow{g^*} & R' \\ \models_Q \downarrow & & \downarrow \models_{R'} \\ Z & \xleftarrow{g_*} & (X')^+ \end{array}$$

□

具象オントロジー $\tilde{\mathcal{O}}$ の構成法 さて、今問題にしているオントロジーの構成は以下の手順で行なう。

1. オントロジー $\mathcal{O}$ の順序関係 $\leq_C, \leq_R$ を含む二つの最小正則理論を $T = \langle C, \vdash_T \rangle, S = \langle R, \vdash_S \rangle$ とする。
2. 定理 1 のおける分類を $B = \text{Cla}(T), Q = \text{Cla}(S)$ とする¹⁴。
3. オントロジー $\mathcal{O}$ を持つ具象オントロジーは $\tilde{\mathcal{O}} = \langle C, R, \leq_C, \leq_R, \sigma \rangle$ とできる¹⁵。

ここで $C = \langle X', C, \models_C \rangle$ 及び $R = \langle (X')^+, R, \models_R \rangle$ は以下を充す分類である。

$$\begin{aligned} x \models_C c &\iff f_*(x) \models_B c, \\ \langle x_1, \dots, x_n \rangle \models_R r &\iff g_*(x_1, \dots, x_n) \models_Q r. \end{aligned}$$

¹¹筆者は寡聞にも populated ontology の訳を見ていない。ここでは仮に具象オントロジーとしておく。

¹² $\tilde{\mathcal{O}}$ の適切さを次の定理が定義している。

¹³定理の各定義にもとづいて容易に証明できる。

¹⁴ここで、二つの情報射 $f : \text{Cla}(S) \rightarrow C', g : \text{Cla}(S) \rightarrow R'$ を適切に構成しなければならない。

¹⁵ $\tilde{\mathcal{O}}$ が $\mathcal{O}$ -具象オントロジーであることは定理 1 から帰結する。