

歴史記述に対する概念分析の試み

白須裕之*

京都大学人文科学研究所

概要

本稿は歴史概念を理解するための概念モデルを提出するのではなく、歴史記述のための知識ベースを構築するということとはどのようなことであるか、ということを理解するための概念モデルを提出する一つの試みである。まず、歴史記述のプロセスでは何が本質的であるか、ということを取り上げ、従来の知識ベースを作成する枠組みに存在した問題点を議論する。特に従来行われてきた電子テキストにマークアップを行い、それに対するメタ情報の付加によって知識ベースを作成するという手法の問題点を述べ、歴史記述に対してどのような知識表現の枠組みが必要であるかを検討する。

A Representation Framework for Historical Knowledge

SHIRASU Hiroyuki*

Institute for Research in Humanities, Kyoto University

Abstract

We presented conceptual models of administrative geographies, government systems and offices in the Tang dynasty(618-907). There were a lot of disadvantages of our approach in the previous papers, in which we applied the method of "semantical markup and meta information" to build a knowledge base system for historical sources. Thus we try to suggest a conceptual model of constructing knowledge base for them. This paper discusses the historical discourse in the postmodern age at first, and presents a framework of knowledge representation for it. This framework is based on the OMG metamodel hierarchy and the reference model of Topic Maps.

1 はじめに

現在、我々は漢字文献の知識ベース化を試みるプロジェクトを推進中である [1]。そのプロジェクトの一環として、中国唐代 (618-907) の人物、官制、行政地理に関するオブジェクト指向分析/設計を行っている。その一つのステップが概念分析であり、その成果物が概念モデルである。概念モデルとは対象となる問題領域を理解するために、主要となる概念とその関連を抽出、表現するものである。文献 [3][5][6] では、各々行政地理、任官、官制に関する概念分析の成果を述べた。これらは歴史資料からその対象とする唐代の歴史的概念を記述、モデル化したものである。

概念分析では静的側面、動的側面、さらに機能的側面から考察する必要がある。知識ベースの構築を前提として、我々が作成した概念モデルは、特に機能的側面が明確でなく暫定的なモデルであった。本稿では我々の概念分析が何故暫定的なものであったのか

ということを出発点として、知識ベースを構築するとはどういうことか、そのためにはどういう枠組みが必要であるかを考察する。即ち、歴史記述のための知識ベースを構築すること、そのものの概念分析の試みである。これは以前提出した歴史概念を理解するための概念モデルではなく、歴史記述の知識ベース化を理解するための概念モデルである。

2 知識ベースの構築とは何か?

2.1 知識ベースの目標

本節では唐代知識ベースの構築のための前提から議論を始めよう。本プロジェクトの出発点となった文献 [7] では、プロジェクトの課題である「情報学によって漢字文化の発展を目指す上での課題」を、以下のように二点にまとめている。

1. 「多くの情報の中から必要とするものを素早く

*京都大学 21 世紀 COE 東アジア世界の人文情報学研究教育拠点

*Toward an Overall Inheritance and Development of Kanji Culture

的確に取り出すための仕組みを考えること」

2. 「我々が傳承してきた漢字文献を過不足なくデジタル化して後世に伝えることが出来るような方法を考えること」

まず、これらの課題を以下のようなユーザ目標レベルのユースケース¹として捉えることにしよう。

1. 多くの情報の中から必要とするものを素早く的確に取り出す。
2. 傳承してきた漢字文献を過不足なくデジタル化する。

ここでの主要アクタは、第一のユースケースでは情報を必要とする人であり、第二のユースケースでは情報を残す人である。知識ベースの場合、情報を残す人には二つの場合が考えられる。一方は歴史資料からテキストを作成する人であり、もう一方はテキストから歴史的事実を抽出する人である。

これらのユーザ目標には幾つか曖昧な点が残っている。第一のユースケースではどのような情報が存在するのか、その中からどういう情報が必要であるのかという点、第二のユースケースではどのような基準に従って文献を過不足なくデジタル化するのか、また、それに関連して、後世が必要とする情報を予めデジタル化することはできるのかという点である。従って、上のユーザ目標を理解するにはこれらを明らかにしなければならない。

2.2 歴史記述について

様々な学問分野で研究の道具そのものを再考することが行われているが、中でも言語論的転回は現代の多くの分野に影響を与えたと思われる。歴史学もその例外ではないであろう。このような言語論的転回以降の歴史学そのものを議論することは本稿の目的ではない。しかし前節で述べたような知識ベースにとって必要な情報とは何かということを明らかにするには、歴史学の営為を明らかにすることは避けて通れない。

文献 [8] では歴史家の仕事を次のように述べている。「歴史家の仕事は資料を読むとだけで終わるわけではなく、そこからしかるべきデータ（つまり、歴史的事実）を選択、抽出して、それを論文や本のかたちに

¹例えば、文献 [17]、63 頁参照。

²OWL については文献 [13] を参照。

まとめてゆかなくてはならない。」ここでは歴史的事実と歴史記述のプロセスが問題になっている。

- 歴史的事実とは何か — 歴史的な出来事を歴史上の事実（歴史学の対象）として成立させているのは、時間や場所などの概念枠、即ちフィクションの役割が重要であること。
- 歴史記述のプロセス — 歴史学の研究は資料から始まる訳ではなく、研究のための mind set を前提としていること。

以上のようなフィクションと mind set を知識表現として記述することはできないだろうか？ 前節での関連で言うと、情報の選択と抽出にはそのための概念枠が必要であり、利用者（後世を含めた読者）が欲しい情報を予め想定することはできないが、しかし、そのための概念枠を自由に設定できることが将来の知識ベースの姿と言えるのではないだろうか？ 以下にこのような知識ベースの姿を具体化していければと思う。

3 用語と準備

次節以降では、歴史記述を知識ベース化するにあたって必要となる要件について議論する。そのために本節ではその要件を議論するときに使用する用語について述べる。

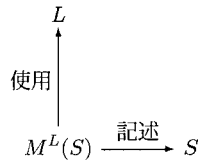
3.1 知識表現

以下で使用する知識表現の用語について、ここでまとめておく。知識表現は通常大きく二つの部分から構成される。第一はそこで使われる概念とその関係を記述する部分であり、第二はそれらの具体的な例を記述する部分である。

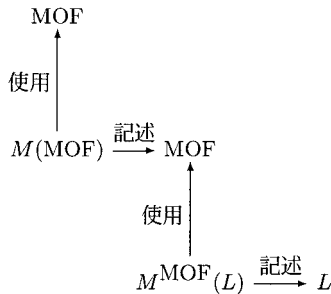
OWL²では前者をクラスとプロパティで記述し、後者を個体とそれらの同値関係などで記述する。また、UML では前者にはクラスと関連、後者にはインスタンスが含まれる。ここではこれら二つの部分を各々「語彙」と「表明」と呼ぶことにする。

3.2 メタモデルとその階層

ここではメタモデルとその階層について簡単にまとめておく。問題領域である「システム」に対して、興味のある側面をある「言語」で記述したものが「モデル」である。従って、一つのシステムに対して、注目する側面によって様々なモデルが提出できる可能性がある。モデルは言語の記述能力によっても異なるが、ここでは「言語」については計算機で扱えるということ以外制限を設けない。当面はUML, Topic Maps, OWL等のモデリング言語を想定している。



上の図はこの関係を示している³。言語 L によってシステム S の必要とする部分を記述したものが、モデル $M^L(S)$ である。例えば、文献 [3][5][6] で提出したモデルは、各々行政地理、任官、官制に関する唐代の歴史的なシステムを言語 UML で記述したモデルである。



では様々な問題領域を記述する「言語」の意味を決めるにはどうしたら良いであろうか? それには「メタモデル」を記述すれば良い。例えば、UMLのメタモデルは言語 MOF で記述されている⁴。ではこの MOF の意味をどのように記述するかというと、リフレクションを使用して MOF 自身によって記述する。MOF は自分自身の意味を記述することができる程、記述力の強い言語である。こうして OMG が提出したメタモデルは 4 階層を構成することになる。本稿の目的は歴史記述という問題領域を扱う枠組みを提出することであるが、そのためにこのメタモデル 4 階層を使う。

³この表記は文献 [16] を参考にした。

⁴詳しくは文献 [18] を参照。

⁵文献 [15] では UML の使用は *informative* である。

⁶文献 [14] では Topic Maps に限らず、オントロジー、タクソノミー、語彙等を解釈するためのモデルをこの構成法を使って作成する。

- M3: メタモデルを記述する言語、例えば MOF。
- M2: UML, Topic Maps, OWL などの言語の意味を決める (メタ) モデルの層。本稿ではこのレベルで知識表現のための言語を記述したい。
- M1: 個々の問題領域の「語彙」を決めるモデルの層。M2 で定義した言語を使用してモデルを記述する。
- M0: 概念の具体例の層。「表明」を記述する。

メタモデル 4 階層では「語彙」は M1 で、「表明」は M0 で定義される。また「語彙」を定義する言語は M2 で定義されることになる。

文献 [3][5][6] では、概念モデルを記述するために UML を使用したが、他に Topic Maps や OWL 等の言語を用いることもできる。我々が考察の対象とする歴史記述のための枠組みを記述する言語を考える上で、これらの言語も非常に参考になる。文献 [15] では Topic Maps のデータモデルが UML によって記述されている⁵。Topic Maps, OWL の意味はこの OMG のメタモデル 4 階層においては M2 で定義できる。また、本稿で提出する枠組みも M2 で定義することになる。

3.3 参照モデル

節 5 では以下のモデル構成法を使用する。これは文献 [14] で提出された構成法⁶を単純化したものである。任意に与えられた二つの集合を L, S とする。ここで L はラベルの集合、 V はモデルを作るときの基本となる集合である。モデル $\mathcal{V}^L(S)$ は、以下の等式を充す \mathcal{V} の内で最小の集合とする。

$$\mathcal{V} = S \cup P_{fin}(L \times \mathcal{V})$$

但し、 $P_{fin}(X)$ は集合 X の有限部分集合全体の集合を表す。

4 知識ベース化のための要件

本節では従来の歴史記述を対象としたデータベース、知識ベースの問題点を述べ、歴史記述のための知識ベースに対する要件を明らかにする。

4.1 唐代知識ベースの概念モデル

まず唐代知識ベースの構築において行ってきた概念分析の問題点について述べる。これらの問題点を以下の要件記述の枠組みに反映する必要がある。

- 機能要件の不足 — 前節で歴史記述のプロセスについて概観した。史料を前にしての概念枠を予め設定できない以上、機能要件を明確化することは不可能と言わざるを得ない。概念枠の設定を知識表現にどのように取り込むかということを次節以降で述べる。
- 複数の概念モデルのマージ — 文献 [4] では行政地理の各文献が提出する概念モデルの違いについて述べた。これら複数の概念モデルを統一するような共通のモデルを提出することは、上の問題点とも関係して難しい。また、行政地理、人物、官制のような異なる種類の概念モデルを統合化する方法を見出せていない状況である。統一モデルの代りに、目的に合わせて概念モデルをマージできること、及びそのためのマージデータを定義する機構が必要である。
- 検索のための概念モデルの提出 — 上の問題点とも関わるが、selecting, filtering, rankingのような目的ごとに動的に概念モデルを提出する必要がある。いわゆる「読者の概念枠」の必要性である。

4.2 歴史文献のマークアップとメタ情報

さて従来、歴史文献をデータベース化するにあたって用いられる技術は、マークアップとメタ情報であった。[12]で始められた唐代研究ナレッジベースのプロジェクトは、「TEIのタグセットを用いて、XMLデータベースにより検索可能な、唐代文化に関する原点テキスト群」「事物・文章・発言、その他の多様な情報を管理するトピックマップ」をその基礎としている。

以下は歴史記述のための要求に関して、マークアップ言語が持つ問題点である。文献にマークアップを施すには、その問題領域に対するある種の概念モデルが必要である。更にある文献テキストから人物、地名、官職名などをマークアップして抽出する場合、テキストからの文字列の切り出しが行なわれる。これはその文献を読む人によって異なる場合がありえる。

⁷Concurrent Markup, Overlapped Markupの手法が提案されているが、一番目の問題と関連して、この問題の解決は難しいと思われる。

また、その切り出された文字列には意味付けという情報が付加される。マークアップとメタ情報という二つの意味付けの間の整合性をどのように取るかという問題も発生する。意味論的なマークアップの問題点は、マークアップのための概念枠を提示する機構がないことにある。以上をまとめると、

- マークアップは「語彙」に依存している。即ち、「語彙」の指定とマークアップを独立にすることはできない。
- 複数の「語彙」を一つのテキストにマークアップする必要がある。
- 「語彙」を固定しても、読みによって同一のテキストに複数のマークアップの可能性がある。テキストの「語彙」に対応する部分の指定に複数の可能性がある⁷。
- テキストにおける資源の指定は多様な要求がある。従来マークアップでは「語彙」における「個体」のみが指定されていたが、歴史記述では「表明」の典拠を指定する必要がある。指定の方法には多様な要求がある。例えば、「ある人物の生地はどこである」という表明の典拠が、一つの文であることもあれば、複数の文から機械的に推論したものである場合もあり、また、「語彙」でその表明を書けず、文章から人が推論したものもあるからである。
- マークアップとメタ情報の整合性 — マークアップによってテキストに埋め込まれた意味情報と、Topic Maps等の言語によって提供されるメタ情報の整合性をどのように取るかという問題。本来、このような意味論的な情報は電子テキストに持つべきではないと思われる。
- メタ情報の異説 — メタ情報に異説がある場合はそれを記述する必要がある。
- メタ情報が情報資源に含まれる場合、工具書のように始めからある種の概念モデルを想定して作成された文献の場合は、その知識表現はマークアップの方法では不十分。
- 信頼性評価の仕組みを取り入れることが難しい。即ち、「語彙」、テキストの校訂、意味付けなどに対して、本来分けるべき信頼性、典拠などを分けることができない。

これらはマークアップを意味記述のために用いる場合の問題点であって、校訂テキストの作成、デジタル化のためにはマークアップ言語を用いることは必須だと考えている。

4.3 歴史学的なオブジェクトの識別

歴史学的なオブジェクトを如何に識別するかという問題がある。ここで言う歴史学的なオブジェクトとは「表明」レベルの「個体」を意味する。Topic Maps では Topic の識別は topic identifier によって行なわれる。これがオブジェクトの識別のためのメタ情報だと考えることができる。しかし歴史学的なオブジェクトの場合は識別のためのメタ情報は、そのための工具書や論文、或いは研究者の解釈の中にしか存在しない。従って歴史学的なオブジェクトを識別するためには、そのためのメタ情報を明示的に記述する必要がある⁸。

また、概念モデルのマージと相俟って、歴史オブジェクトのマージも考慮する必要がある。従来の方法(特に Topic Maps)では概念と個体のマージを同じ方法で行っているが、これらは分ける必要があり、またマージのデータも自由に設定できる必要がある。

4.4 テキストと文献関連情報

デジタル化においてはテキストそれ自身の問題についても考慮する必要がある。文献 [9] では、校訂テキストを電子化するにあたっての問題点が指摘されている。また文献 [10] では、写本や翻訳文献等の複数文献にまたがる文献の関係情報の必要性が述べられている。歴史記述のテキストにおいても同様の問題が存在する。CVS や Subversion 等のバージョン管理システムが備えている機能も要件として捉える必要がある。

4.5 知識ベース化のための要件

以上、知識ベース化のための要件について述べた。ここでその要件についてまとめておく。

- テキストの発展、異読などの Version 管理情報

⁸OWL では「個体」の同定、識別を記述するプロパティが予め用意されている。

⁹文献 [3] を参照。これらを Topic Maps や OWL で自然に記述することは難しい。例えば、Topic Maps の Scope を使って時間軸に沿った情報を記述することがすぐに思い浮かぶ。しかし、このような情報とイベント情報との統合化を素直に表現することは難しいと思われる。これについては別考を用意している。

¹⁰Topic Maps では書誌情報や意味論的な制限(時間や場所)を scope で扱うことが良く行なわれるが、本稿では書誌情報の「語彙」、意味論的制限の「語彙」を設計して、「語彙」のマージで扱う。「語彙」がある意味で、節 2.2 における mind set とフィクションを記述する。

を記述できる。

- マークアップと「語彙」を連動させる仕組みを提供する。
- 「語彙」の統合化、マージの仕組みを提供する。
- 「表明」の典拠と評価を定義する仕組みを提供する。
- 作者、使用者による自由なメタ情報の追加、変更ができる。
- 歴史オブジェクトの同定、識別判断を定義できる仕組みを提供する。
- 歴史オブジェクトの動的振舞いが記述できる。

最後の歴史オブジェクトの動的振舞いが記述できるというのは、例えば、行政地理における「行政区」のオブジェクトが名前の変更、改廃、所属の変更などの歴史的なイベントを記述する枠組みを要求しているからである⁹。

5 歴史記述のための知識表現の枠組み

本稿で扱う歴史資料は主に文献として表現されたもの、史料を対象とする。非文献資料である考古資料、民族資料等については更に考察を必要とする。以下では前節で議論した知識ベースの要件を踏まえて、歴史記述のための知識表現の枠組みを提出する。前節で歴史記述にとって「マークアップとメタ情報」という枠組みを持つ問題点を見た。本節では別のアプローチ¹⁰として、「抽象テキスト」「証拠」「典拠」「表明の構成」という概念を提案する。

5.1 抽象テキスト

「抽象テキスト」は後程述べる「典拠」が行いやすいように作られている。以下の図は抽象テキストを構成するオブジェクトである。

「単位」はテキストを構成するための単位であって、文献 [2] で提案されている Chaon モデルの文字の

ようなものを考えている。この問題については別稿で考察することにする。

「テキスト」は「単位」から Composite パターンを用いて構成される。また各「テキスト」は前節で見たテキストの間の引用、翻訳、校訂などの文献間の関係情報を持つ。また「文献」は「テキスト」から構成されていて、様々な書誌情報を持つ。こうした仕組みによって、「テキスト」オブジェクトは自身の親に対する位置、書誌情報などを知ることができる。

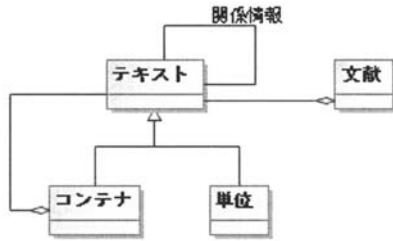


図 1: 抽象テキスト

抽象テキストはある意味で文字列の集合を適当に指定できれば良い。例えば、ある文字列を C とし、 C の部分列全体の集合を $Seq(C)$ とする。この集合の各要素に文字列間の位置情報、文献関係情報を追加したものを抽象テキストと考えれば良い。以下では抽象テキストとそれに含まれる「テキスト」オブジェクト全体の集合を同一視する。

5.2 語彙と証拠

「証拠」とは「表明」のための証拠となるテキスト断片のことである。すなわち「証拠」はテキストに対する一つのビューを提供する。「典拠」は「語彙」付きの「証拠」の集合である。「語彙」を $T = (V, R)$ とする¹¹。但し、この二項組は各々、クラス変数の集合 $V = \{v_1, \dots, v_n\}$ 、型付きの関係変数の集合 $R = \{r_0, \dots, r_m\}$ である。各関係変数の型は $r: v_1 \times v_2$ などを示す。関係は二項以上の多項関係であってもかまわない。「証拠」、「典拠」を以下のように定義する。

- 語彙 $T = (V, R)$ 、抽象テキスト S に対して、
- 証拠 $W^T(S): T \longrightarrow S$
- 典拠 $A(W) = \sum_{l \in V \cup R} W^T(l)$

¹¹本稿の目的が知識表現そのものでなく枠組みを提出することであるため、「語彙」として非常に単純なものを仮定する。

ここで $W^T(S)$ は抽象テキスト S から作られた「証拠」である。クラス変数 v に対しては、 $W^T(v)$ はこのクラスに関するテキストオブジェクトの集合であり、関係変数 r に対しては、 $W^T(r)$ はこの関連に関するテキストオブジェクトの集合である。関係変数が $r: v_1 \times v_2$ である場合、 $W^T(r)$ 中のテキストオブジェクトには、 $W^T(v_1)$ 、 $W^T(v_2)$ のオブジェクトの一つが含まれている。また、テキストによっては $W^T(v)$ 及び $W^T(r)$ は空集合であることもある。 $\sum_{l \in X} W^T(l)$ は dependent product で、 $(l, x) \in \sum_{l \in X} W^T(l)$ は $l \in X \wedge x \in W^T(l)$ と同値。

語彙の例を「人物とその生地」とする。対応するモデルにはクラス「人物」、「場所」とその人物が生れた場所とを結びつける関連「生地」が含まれている。この語彙で、例えば文献『旧唐書』列伝から、人名、地名の証拠を作成する訳であるが、読みに従って証拠の作成は複数可能であることが分かる。

5.3 表明

「表明」とはメタモデル階層 M0 のモデルであり、以下のようにテキストと「語彙」によって決まる「証拠」から構成される。

- 語彙 $T = (V, R)$ 、抽象テキスト S 、証拠 $W = W^T(S)$ が与えられたとき以下のように定義する。
- 表明 $M^T(W) \subseteq \mathcal{V}^T(A(W)) / \sim$

ここで右辺の T は $V \cup R$ であり、 \mathcal{V} は節 3.3 で述べたモデル構成法である。また、同値関係 \sim は個体の同定関係である。この同値関係を定義する方法は別に考慮しなければならない。また、「表明」はモデルで真なるもののみを含むが、真なるものを選択する方法についても述べない。ここでの興味は表明を構成する方法だからである。

語彙の例を先程と同じ「人物とその生地」とする。人名、地名に対応する証拠を「人物」、「場所」の意味に対応付ける必要がある。これは複数の箇所に出現する人名を一人の人物に関係付けることに相当する。

以上で抽象テキストから典拠、表明を構成する方法が定義できた。以下ではこの枠組みを使用して、複数のテキスト、複数の語彙、更に検索をどのように扱うかについて述べる。

5.4 複数テキストの扱い

議論を単純化するためにテキストが二つである場合を考えよう。実はこの問題は二つの抽象テキストから一つの抽象テキストを構成すれば良い。全ては表明の構成によって吸収できる。

- 語彙 T と二つのテキスト S_1, S_2 、二つの証拠 $W_1 = W^T(S_1), W_2 = W^T(S_2)$ から
- 表明を $M^T(W_1 \cup W_2)$ で定義する。

語彙の例として先程と同じ「人物とその生地」を考える。複数の文献、例えば『新唐書』『旧唐書』の列伝に対して、その文献の「人物」、「場所」を対応付けることが必要である。これは表明の構成、即ち同値関係に記述されている。

5.5 複数の語彙

議論を単純化するために二つの語彙 T_1, T_2 と二つのテキスト S_1, S_2 がある場合を考えよう。語彙 T_1 に基づく証拠 $W^{T_1}(S_1)$ と、語彙 T_2 に基づく証拠 $W^{T_2}(S_2)$ が与えられていると仮定する。二つの語彙をマージした語彙 $T_1 \times T_2$ に対応した証拠が作れば問題はないが、証拠を新たに作成することは困難である。従って、以下のような方法を使う。

- マージした語彙 $T_1 \times T_2$ と、二つの証拠 $W_1 = W^{T_1}(S_1), W_2 = W^{T_2}(S_2)$ から以下により表明を定義する。
- 表明 $M^{T_1 \times T_2}(W_1 \cup W_2)$

語彙 $T_1 \times T_2$ に対応した表明を証拠 $W^{T_1}(S_1)$ と証拠 $W^{T_2}(S_2)$ から構成する訳である。

語彙の例として、「人物とその生地」、「行政区」の二つを取り上げるとする。複数の文献としては『旧唐書』列伝、『新唐書』地理志を扱う。このとき『新唐書』地理志にある「行政区」を『旧唐書』列伝の「場所」として扱う必要がある。

5.6 検索、ビュー

検索やビューとは、新しい「語彙」を与えて、既存の「証拠」から「表明」を構成することである。

- 既存の語彙 T 、テキスト S 、証拠 $W^T(S)$

¹²文学資料については文献 [11] を参照する。

- 新しい語彙 T'
- 語彙の対応関係 $T : T \longrightarrow T'$

が与えられたとき、新しい表明は既存の表明 $M^T(W^T(S))$ からモデル変換で得られる。また、このモデル変換を自動的に定義できるように、語彙の対応関係 $T : T \longrightarrow T'$ を定義しなければならない。実際の検索やビューを実現するには、複数の語彙、テキストを前提として行なわれるため、上で述べた変換の前に語彙とテキストのマージが行なわれる必要がある。

6 課題

歴史記述を知識ベース化するということはどういうことかということを出発点に、本稿では「歴史記述のための知識表現の枠組み」を提出した。UML, Topic Maps, OWL 等の個々のモデリング言語で歴史記述を扱えないと主張したい訳ではなく、却ってこの枠組みは既存の言語で記述できるものばかりである。

本稿の目的は、これらの言語で記述するにしても、新しい言語を設計するにしても、歴史記述のために必要な概念がどのようなものであるかを明らかにすることであった。それには「マークアップとメタ情報」という枠組みを一旦解体して、歴史記述に必要な概念を再構成する必要がある。そのためのヒントが歴史記述のプロセスを再考することである。本稿はあくまでそのための試みである。今後、ここで述べた枠組みを M2 モデルとして詳細化する必要がある。以下は今後の課題である。

- 歴史学において、歴史記述の知識ベース化で扱える事象と扱えない事象。本稿で議論したように歴史記述で知識ベース化として扱える情報は、概念モデルの網にかかるものに過ぎない。
- 信頼性記述の枠組み。文献 [1] では信頼性評価の仕組みを提出した。この仕組みは「抽象テキスト」、「表明」、「語彙」に適用する必要がある。信頼性は本来的にテキストと読みの二つの層に必要であり、本稿でのテキストと読みの分離は、信頼性を記述するために適していると言える。
- 文学と歴史学¹² — 言語論的転回以降、歴史学に対して文学論は欠かせない。歴史的な事実は文学のフィクションとどう異なるのか、文学資

料と歴史資料の電子化という問題に対して考察の必要がある。

7 おわりに

歴史記述の知識ベースで何ができるかと言えば、歴史記述のプロセスそのものを概念分析して、その中からソフトウェアシステムとして計算機で扱える事柄を抽出する必要がある。本稿はそのための拙い第一歩だと考えている。

本稿の歴史記述のための言語では表現力が主な要因になった。しかし、知識表現のための言語を設計する上で推論について議論することも欠かせない。推論の計算量はその語彙に大きく依存するため、その観点からも語彙を検討していく必要があるであろう。

謝辞 永崎研宣さん、明星聖子さんには論文の別刷を頂き、本稿を書く上で多くの示唆を賜りました。深く感謝いたします。また、筆者をマークアップ研究班に参加させていただき、オントロジーについて学ぶ機会を与えてくださった C. Wittern さんに感謝いたします。本稿を書くにあたり以下の方々にお世話になりました。深く感謝いたします。安岡孝一さん、永田知之さんには唐代知識ベースのプロジェクトにおいてお世話になっております。山田崇仁さん、牛根靖裕さん、秋山陽一郎さんには日頃より人文学、計算機科学、歴史学についての議論を通じて多くの影響を受けております。特に山田崇仁さんには原稿を読んでもいただき有益なコメントを頂きました。守岡知彦さんには人文情報学について有益なお話を沢山うかがっております。最後に妻留美に感謝します。

参考文献

- [1] 秋山陽一郎, 白須裕之, 永田知之: 中国古典学知識ベースにおける信頼性評価モデルの一試案, 第 17 回 東洋学へのコンピューター利用, 2006.
- [2] 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.
- [3] 牛根靖裕, 白須裕之, 山田崇仁: 唐代行政地理の概念モデル, 情報処理学会研究報告, 2007-CH-73, 2007.
- [4] 牛根靖裕, 白須裕之, 山田崇仁: 「唐代の行政地理」に関する「史料」問題, 第 18 回 東洋学へのコンピューター利用, 2007.
- [5] 白須裕之: 唐代任官情報の概念モデル — 時間、文献に依存する情報のための分析パターン —, 情報処理学会研究報告, 2006-CH-72, 2006.
- [6] 白須裕之: 唐代官制の概念モデリングの試み, 第 18 回 東洋学へのコンピューター利用, 2007.
- [7] 高田時雄: 漢字文化の全き継承と発展のために, 漢字と文化 創刊号, 3-4, 2003.
- [8] 富山太佳夫: フィクション抜きの実実は存在するか, 森明子編著『歴史叙述の現在』人文書院所収, 129-144, 2002.
- [9] 永崎研宣: デジタルアーカイブと校訂テキスト, 情報処理学会「人文科学とコンピュータシンポジウム」, 1-8, 2003.
- [10] 永崎研宣: シラブルを最小単位とする仏教哲学文献データベースについて, 情報処理学会研究報告, 2006-CH-71, 2006.
- [11] 明星聖子: 文学デジタルアーカイビングをめぐる理論的考察, 情報処理学会「人文科学とコンピュータシンポジウム」, 153-160, 2006.
- [12] C. Wittern(訳: 秋山陽一郎): 唐代ナレッジベースに向けて, 漢字文献情報処理研究 5, 20-24, 2004.
- [13] M. Dean, G. Schreiber eds.: OWL Web Ontology Language Reference, W3C Recommendation, <http://www.w3.org/TR/owl-ref/>, 2004.
- [14] ISO 13250-5: Topic Maps Reference Model, <http://www.isotopicmaps.org/tmrm/>, 2006.
- [15] ISO 13250-2: Topic Maps — Data Model, <http://www.isotopicmaps.org/sam/>, 2006.
- [16] A. Kleppe, J. Warmer, W. Bast: MDA Explained, Addison Wesley, 2003. (日本語訳「MDA 導入ガイド」インプレス)
- [17] C. Larman, Applying UML and Patterns, 2nd Edition, Pearson Education, Inc., Pub., 2002. (日本語訳「実践 UML 第 2 版」ピアソン・エデュケーション)
- [18] Object Management Group: Unified Modeling Language, Infrastructure ver.2.1, <http://www.omg.org/ptc/06-04-03>, 2006.