文字体系の統合による漢字情報の形式化
—『説文解字注』における音注を事例として —

白須裕之

概 要

本稿は人文学の研究に果たす形式的体系の意味を再考し、表象主義的な前提では捉えきれない情報技術の利用の形を模索するものである。文字という対象は概念的な存在であり、解釈や知識による揺れをもつ。このような歴史的な対象には体系の統合化の仕組みが欠かせない。文字情報の形式化としては「情報の流れ」理論的なアプローチが有効であったが、ここではその拡張として、文字体系の統合を具体的な実現に依存しない institution 理論による形式化を目指す。また、institution 理論によって定式化された論理プログラミングの枠組みを利用して、文字についての問合せ、推論等の実現を試みる。

Institutionalising Representational Systems for Chinese Characters
— A Case of Phonological Notes in *Shuowen Jiezi Zhu* —

Shirasu Hiroyuki

Abstract

Information technologies are the essential requisites for researching humanities. The concept of formalization system is an important subject underlying knowledge information technologies. The theory of *information flow*, in which J. Barwise and J. Perry tried to formalize F. Dretske's idea of representation, is valuable for formalizing the information on humanities. The theory of *institutions*, which is considered as an extension of it, could present general foundations for ontology-based semantic integration. The aim of this paper is to present to institutionalizing representational systems for Chinese characters.

1　はじめに

人文学研究は言うに及ばず、現代の学問研究において情報技術の必要性は自明のことと成りつつある。計算機の利用という面で、現代の利用形態はかつての人工知能研究とどのような違いがあると言えるのであろうか。そこには両者に共通する、世界に対する了解が隠されているように思う。

H. Dreyfus らの行なった人工知能批判の矛先は、人工知能研究者のもつ認知主義であったが、そのような批判を待つまでもなく、情報技術の利用には「表象主義」への陥穽がある。研究対象としての「表象されるもの」と、情報技術によって「表象するもの」とのあいだの指示関係が前提されている。このような志向性についての議論は「表象的志向批判」として、文献 [12] で提出されている。

筆者はこれまでに文献資料の電子的な再現/表象をテーマとして、テクストアーカイブズ構築の基礎理論を手掛けてきた [15][16]。また、N. Goodman の「記号体系」を手掛かりに、アーカイブズの機能を理解する試みを提出した [14]。これらは情報技術の利用を研究の為のツールとしてのみ捉える見方を批判するとともに、表象するものとしての「記号体系」の積極的な価値を考察するものであった。しかし、「表象主義」という枠組みに捉えられていなかったとは言えない[*1]。情報技術の利用によって「知識の再構築」を目指すのであれば、記号体系の働きについて再考することが欠かせない。

G.-G. Granger の比較認識論 [7] では、哲学的概念への記号体系の働きについて議論する。「対象のない概念」という考えを中心にすえ、哲学的な諸概念が集合論的な図式に還元不可能であり、「哲学的概念システムを具現化するのにいかなる形式的体系も有効に確立しえない」として、論理的還元の虚構性を指摘する。それでは我々には、概念システ

[*1] ここでの立場は、世界を有意味に表現するための単位として、文や命題の貢献を考える立場であるとして、文献 [12] で「表象主義 2」と名付けられている。

ムとしての形式化の道は閉ざされているのであろうか。Granger は個別的な学問領域の比較を通して、記号体系の役割を考察する。特に厳密性の根拠としての形式について、Granger の考えを文献 [13] では、「数学、論理学は対象そのものではなく、可能な対象の形式をもっぱら対象とする。(87 頁)」というように纏めている。

本稿では、「可能な対象の形式」に注目し、「情報の流れ」理論 [1] が情報を担うものの具体的な形ではなく、機能としての情報構造に注目していること、及び institution 理論 [6] が様々な論理体系の具体的な構文や意味の実現を抽象化して、圏論的な形式化を行なっていることに注目する。以下では漢字情報の形式化を取り上げるが、これは人文学研究における情報の形式化に対して、範型としての役割を模索するものである。

漢字情報の形式化

文字についての様々な概念を形式化するには、言語的な情報は勿論、引用や注釈などの文献学的な情報も重要である。また、文字という対象は概念的な存在であり、予め定義できるような確固とした存在とは異なり、どのように解釈し、どのような知識を持つかによって揺れがある。このような歴史的な対象であることを考慮すると、文字という対象にはオントロジー工学的な思考法に加えて、新たな解釈としての文字体系を従前の体系に統合化して、文字概念の更新を容易にする形式化が必要であろう。

オントロジー工学においては、従来よりオントロジーの統合について議論されている[*2]が、ここでは統合についても形式的であること目指す。文献 [16] では、文字体系の関係を表現系で定義したが、本稿ではこの方式を発展させて、統合を institution 理論の枠組みで表現する[*3]。実装としては、institution 理論で定式化した論理プログラミングの枠組み [4] を利用し、文字についての問合せ、推論等の実現を試みる。事例として段玉裁の『説文解字注』(以下、段注) の音注を取り上げる。

*2 オントロジーの統合については文献 [8] が参考になる。

*3 文献 [11] ではオントロジーの統合を institution 理論を使って議論しているが、表現系の議論はなされていない。

2　Institution 理論

本稿の理論的な基礎の多くが文献 [4] に基づいている。本節では後の議論に必要な範囲で、その考え方、及び結果について纏める[*4]。なお圏論についての基本事項は文献 [9] を参照願いたい。

文献 [5] では Institution 理論を応用して、論理プログラミングにおいて基本的である Herbrand 定理、モジュール性を Institution の具体的な構成に依存せずに証明する。本稿ではこのモジュール性を文字体系の表現に利用する。

定義 1 (institution) 以下から成る 4 つ組 $I = \langle \mathbb{S}ig, \mathrm{Sen}, \mathrm{Mod}, \models \rangle$ が institution であるとは、

1. 圏 $\mathbb{S}ig$ (対象 $\Sigma \in |\mathbb{S}ig|$ をシグニチャ、射 $\varphi : \Sigma \to \Sigma'$ をシグニチャ射と呼ぶ),
2. 関手 $\mathrm{Sen} : \mathbb{S}ig \to \mathbb{S}et$ (シグニチャ Σ に対して、集合 $\mathrm{Sen}(\Sigma)$ の要素を Σ-文と呼ぶ),
3. 関手 $\mathrm{Mod} : \mathbb{S}ig^{\mathrm{op}} \to \mathbb{C}at$ (シグニチャ Σ に対して、$\mathrm{Mod}(\Sigma)$ の対象を Σ-モデル、$\mathrm{Mod}(\Sigma)$ の射を Σ-モデル射と呼ぶ),
4. 関係 $\models_\Sigma \subseteq |\mathrm{Mod}(\Sigma)| \times \mathrm{Sen}(\Sigma)$.

$$
\begin{tikzcd}
 & \mathbb{S}ig \arrow[dl, "\mathrm{Mod}^{\mathrm{op}}"'] \arrow[dr, "\mathrm{Sen}"] & \\
\mathbb{C}at^{\mathrm{op}} & \models & \mathbb{S}et
\end{tikzcd}
$$

任意のシグニチャ射 $\varphi : \Sigma \to \Sigma'$ に対して、任意の $M' \in |\mathrm{Mod}(\Sigma')|$ と $e \in \mathrm{Sen}(\Sigma)$ が以下の充足条件 (satisfaction condition) を充すものを言う。

$$M' \models_{\Sigma'} \mathrm{Sen}(\varphi)(e) \iff \mathrm{Mod}(\varphi)(M') \models_\Sigma e.$$

$$
\begin{tikzcd}
\Sigma \arrow[d, "\varphi"'] & \mathrm{Mod}(\Sigma) \arrow[r, dash, "\models_\Sigma"] & \mathrm{Sen}(\Sigma) \arrow[d, "\mathrm{Sen}(\varphi)"] \\
\Sigma' & \mathrm{Mod}(\Sigma') \arrow[u, "\mathrm{Mod}(\varphi)"] \arrow[r, dash, "\models_{\Sigma'}"'] & \mathrm{Sen}(\Sigma')
\end{tikzcd}
$$

*4 管見の限り institution 理論に関する邦訳文献については存知あげない。以下で示す訳語は暫定的なものであり、一部は邦訳を示していない。

以下では、 Mod(φ) を $\_\restriction_\varphi$ と書き、縮約関手 (reduct functor) と呼ぶ。また、 $M = M'\restriction_\varphi$ であるとき、 M' を φ に沿った M の拡張と呼ぶ。文変換 Sen(φ) も単に $\varphi(-)$ と書くことにする。

定義 2 (表示 **presentation)** 組 $\langle\Sigma, E\rangle$ が表示 (presentation) であるとは、 Σ がシグニチャであり、 E が Σ-文の集合であるときを言う。ここで演算 $(-)^*$ を以下のように定義する。

- $E^* = \{M \in \mathrm{Mod}(\Sigma) \mid \forall e \in E. M \models_\Sigma e\}$,
- $\mathbb{M}^* = \{e \in \mathrm{Sen}(\Sigma) \mid \forall M \in \mathbb{M}. M \models_\Sigma e\}$.

$E^* \subseteq E'^*$ のとき、 $E \models E'$ と書く。表示射 $\varphi : \langle\Sigma, E\rangle \to \langle\Sigma', E'\rangle$ とは、 $E' \models \varphi(E)$ を充すシグニチャ射 $\varphi : \Sigma \to \Sigma'$ を言う。

2.1 Herbrand 定理

ここでは論理プログラミングで基本的である Herbrand 定理の instituion における表現について述べる。

定義 3 (基本的 **basic)** シグニチャ Σ に対して、 Σ-文 e が基本的 (basic) であるとは、 Σ-モデル M_e が存在して、任意の Σ-モデル M に対して以下が成り立つことである。

$$M \models_\Sigma e \iff \text{モデル射} M_e \to M \text{が存在する}.$$

定義 4 (表現可能 **representable)** シグニチャ射 $\chi : \mathrm{Mod}(\Sigma) \to \mathrm{Mod}(\Sigma')$ が表現可能 (representable) であるとは、 Σ-モデル M_χ と同型射 i_χ が存在して、以下を可換にすることである。

$$\begin{tikzcd}
\mathrm{Mod}(\Sigma') \arrow[r, "i_\chi"] \arrow[dr, "\mathrm{Mod}(\chi)"'] & (M_\chi/\mathrm{Mod}(\Sigma)) \arrow[d, "forgetful"] \\
& \mathrm{Mod}(\Sigma)
\end{tikzcd}$$

但し、 $(M_\chi/\mathrm{Mod}(\Sigma))$ は圏 Mod(Σ) 上の恒等関手によるコンマ圏である。このとき Σ-モデル M_χ をシグニチャ射 $\chi : \mathrm{Mod}(\Sigma) \to \mathrm{Mod}(\Sigma')$ の表現 (representation) と言う。

定義 5 (query) シグニチャ Σ に対して、 Σ-問合せ (query) とは、 $(\exists\chi)\rho$ の形の存在冠頭 Σ-文である。ここで $\chi : \Sigma \to \Sigma'$ は表現可能であり、 $\rho \in \mathrm{Sen}(\Sigma')$ は基本的である。

定理 1 (Herbrand 定理**)** 始モデル $0_{\Sigma,E}$ を持つ表示 (Σ, E) を考える。各問合せ $(\exists\chi)\rho$ に対して、以下が成り立つ。

$$E \models (\exists\chi)\rho \iff 0_{\Sigma,E} \models (\exists\chi)\rho.$$

始モデル $0_{\Sigma,E}$ の $\chi : \Sigma \to \Sigma'$ に沿った拡張 M' (すなわち、 $M'\restriction_\chi = 0_{\Sigma,E}$) で、 $M' \models \rho$ のとき、 M' を $(\exists\chi)\rho$ の解と言う。 $0_{\Sigma,E}$ は (Σ, E) の Herbrand モデルと言われる。

2.2 モジュール性

以下では表示をモジュールとして扱うことができることを示す。問合せはモジュールのインポートに関して意味を持たせることができる。

定義 6 (liberal) シグニチャ射 $\varphi : \Sigma \to \Sigma'$ が liberal である[*5]とは、縮約関手 $\_\restriction_\varphi : \mathrm{Mod}(\Sigma') \to \mathrm{Mod}(\Sigma)$ が左随伴 (これを $(-)^\varphi$ と書く) を持つことである。

$$\begin{tikzcd}
M \arrow[r] \arrow[dr, "h"'] & (M^\varphi)\restriction_\varphi \arrow[d, "h'\restriction_\varphi"] & M^\varphi \arrow[d, dotted, "h'"] \\
& M'\restriction_\varphi & M'
\end{tikzcd}$$

$$\begin{tikzcd}
\mathrm{Mod}(\Sigma) \arrow[r, shift left, "(-)^\varphi"] \arrow[r, phantom, "\perp"] & \mathrm{Mod}(\Sigma') \arrow[l, shift left, "_\restriction_\varphi"]
\end{tikzcd}$$

定義 7 (semi-exact) institution I が semi-exact であるとは、モデル関手 $\mathrm{Mod} : \mathbb{S}ig^{\mathrm{op}} \to \mathbb{C}at$ が pullback を保つことである。

定義 8 (amalgamation square) 以下の可換なダイアグラムが amalgamation square であるとは、 $M_1\restriction_{\phi_1} = M_2\restriction_{\phi_2}$ を充す任意の Σ_1-モデル M_1 と Σ_2-モデル M_2 に対して、唯一の Σ'-モデル M' が存在して、 $M'\restriction_{\phi'_1} = M_1$ かつ $M'\restriction_{\phi'_2} = M_2$ を充すことである。

[*5] 性質 liberal は、モデルの自由性 freeness と密接に関係している。

$$\begin{array}{ccc} \Sigma & \xrightarrow{\phi_1} & \Sigma_1 \\ {\scriptstyle \phi_2}\downarrow & & \downarrow{\scriptstyle \phi_1'} \\ \Sigma_2 & \xrightarrow[\phi_2']{} & \Sigma' \end{array}$$

このとき M' を $M_1 \otimes_{\phi_1,\phi_2} M_2$ または単に $M_1 \otimes M_2$ と記す。semi-exact な institution において、シグニチャ射の pushout square は amalgamation square である。

定理 2 (問合せの変換 **Translation of queries**) semi-exact な institution において、全てのシグニチャ射が liberal であるとする。シグニチャ射 $\varphi : \Sigma \to \Sigma'$ は Σ-問合せ $(\exists\chi)\rho$ を Σ'-問合せ $(\exists\chi')\varphi_1(\rho)$ に写す。但し、以下のダイアグラムは pushout である。

$$\begin{array}{ccc} \Sigma & \xrightarrow{\chi} & \Sigma_1 \\ {\scriptstyle \varphi}\downarrow & & \downarrow{\scriptstyle \varphi_1} \\ \Sigma' & \xrightarrow[\chi']{} & \Sigma_1' \end{array}$$

定理 3 (解の変換 **Translation of solutions**) 上の定理と同じ条件において、 Σ-問合せ $(\exists\chi)\rho$ の解を M_1 とし、 N_1' を以下のように決める。

- $h_1 : M_1 \to N_1$ を $h : 0_{\Sigma,E} \to 0_{\Sigma',E'}\restriction_\varphi$ の χ-拡張とする,
- $N_1' = N_1 \otimes 0_{\Sigma',E'}$.

このとき N_1' は Σ'-問合せ $(\exists\chi')\varphi_1(\rho)$ の解である。

2.3 一階述語論理の Institution

ここでは文字体系の形式化に必要な範囲でのみ、一階述語論理の Institution $\mathcal{I}^{\mathbf{FOL}}$ について述べる。詳しくは文献 [5][6] 等を参照してほしい。

■シグニチャの圏 一階述語論理 **FOL** のシグニチャは以下の 3 つ組 $\langle S, F, P\rangle$ からなる。

- S はソート記号の集合,
- F は $S^* \times S$ で添字付けられた (関数記号の) 集合族 $F = \{F_{u,s} \,|\, (u, s) \in S^* \times S\}$、特に $F_{\lambda,s}$ はソート s の定数を表現する (但し、 λ は空列),
- P は S^* で添字付けられた (関係記号) の集合族 $P = \{P_u \,|\, u \in S^*\}$.

シグニチャ $\langle S, F, P\rangle$ から $\langle S', F', P'\rangle$ へのシグニチャ射は、各々を写す関数の 3 つ組 $\langle \phi_1, \phi_2, \phi_3\rangle$ からなる。

- $\phi_1 : S \to S'$ はソート記号の集合間の関数,
- $\phi_2 : F \to F'$ は $S^* \times S$ で添字付けられた関数族で、 $(\phi_2)_{u,s} : F_{u,s} \to F'_{\phi_1^*(u),\phi_1(s)}$,
- $\phi_3 : P \to P'$ は S^* で添字付けられた関数族で、 $(\phi_3)_u : P_u \to P'_{\phi_1^*(u))}$.

■モデルの圏 **FOL** のシグニチャ $\langle S, F, P\rangle$ のモデル M は以下の成分より構成され、これを使ってシグニチャの各記号が解釈される。

- S で添字付けられた空でない集合の族 $\{M_s \,|\, s \in S\}$ (集合 M_s をソート s の台集合と呼ぶ),
- $S^* \times S$ で添字付けられた関数族 $\{\alpha_{u,s} : F_{u,s} \to [M_u \to M_s] \,|\, u \in S^*, s \in S\}$ (関数記号 $\sigma \in F_{u,s}$ に関数 $\alpha_{u,s}(\sigma) : M_u \to M_s$) を対応させる),
- S^* で添字付けられた関数族 $\{\beta_u : P_u \to \mathcal{P}(M_u) \,|\, u \in S^*\}$ (関係記号 $\pi \in P_u$ に部分集合 $\beta_u(\pi) \subseteq M_u$ を対応させる).

以下、簡単のために $\alpha_{u,s}(\sigma)$ を M_σ、 $\beta_u(\pi)$ を M_π と書くことにする。

シグニチャ $\langle S, F, P\rangle$ のモデル射 $h : M \to M'$ は S で添字付けられた関数族 $\{h_s : M_s \to M_s'\}$ で、以下の条件を充たすものである。

$$\begin{array}{ccc} M_u & \xrightarrow{M_\sigma} & M_s \\ {\scriptstyle h_u}\downarrow & & \downarrow{\scriptstyle h_s} \\ M_u' & \xrightarrow[M_\sigma']{} & M_s' \end{array} \qquad \begin{array}{ccc} M_\pi & \hookrightarrow & M_u \\ {\scriptstyle h_u}\downarrow & & \downarrow{\scriptstyle h_u} \\ M_\pi' & \hookrightarrow & M_u' \end{array}$$

以上のように定義されたシグニチャ $\langle S, F, P\rangle$ のモデルとモデル射は圏を作る。これを $\mathrm{Mod}(S, F, P)$ と記す。

■縮約関手　さらにシグニチャ $\langle S,F,P\rangle$ から $\langle S',F',P'\rangle$ へのシグニチャ射 φ に対して、縮約関手 $\_\restriction_\varphi$: $\mathrm{Mod}(S',F',P') \to \mathrm{Mod}(S,F,P)$ を決める。まず $\langle S',F',P'\rangle$-モデル M' に対して、$\langle S,F,P\rangle$-モデル $M'\restriction_\varphi$ は以下である。

- ソート記号 $s\in S$ のとき、$(M'\restriction_\varphi)_s = M'_{\phi_1(s)}$,
- 関数記号 $\sigma \in F_{u,s}$ のとき、$(M'\restriction_\varphi)_\sigma = M'_{(\phi_2)_{u,s}(\sigma)}$,
- 関係記号 $\pi \in P_u$ のとき、$(M'\restriction_\varphi)_\pi = M'_{(\phi_3)_u(\pi)}$.

モデル射 $h': M' \to N'$ の縮約　$h'\restriction_\varphi$ は、各ソート $s\in S$ に対して $(h'\restriction_\varphi)_s = h'_{\phi_1(s)}$ で定義できる。以上より関手 $\mathrm{Mod}: \mathbb{S}ig^{\mathrm{op}} \to \mathbb{C}at$ が定義される。

■文の集合　最初に有限の変数の集合 X を仮定する。但し、各変数はソートが決まっていて、ソート $s\in S$ をもつ X の部分集合を X_s と記す。また、変数 $x\in X_s$ のとき $x:s$ と書く。

まずソート s をもつ F-項の集合 $T_F(X)_s$ を以下を充す最小の集合とする。

- $X_s \cup F_{\lambda,s} \subseteq T_F(X)_s$,
- $\sigma \in F_{u,s},\ t_i \in T_F(X)_{u_i} (i\in u)$
  $\Rightarrow \sigma(t_1,\ldots,t_n) \in T_F(X)_s$.

FOL のシグニチャ $\langle S,F,P\rangle$ に対する文の集合 $\mathrm{Sen}(S,F,P)$ は以下のように定義される。

- 同じソートをもつ項 t と t' についての等式 $t=t'$、及び関係記号 $\pi\in P_u$、項 $t_i \in T_F(X)_{u_i}(i\in u)$ が作る関係式 $\pi(t_1,\ldots,t_n)$ は $\langle S,F,P\rangle$-文である。等号及び関係式を原子式という。
- ρ_1 と ρ_2 が $\langle S,F,P\rangle$-文であるとき、$\rho_1\wedge\rho_2$ は $\langle S,F,P\rangle$-文である。以下、他の命題論理記号についても同様。
- ρ が $\langle S,F\cup X,P\rangle$-文であるとき、
  $(\exists X)\rho$, $(\forall X)\rho$ は $\langle S,F,P\rangle$-文である。

シグニチャ射 $\varphi : \langle S,F,P\rangle \to \langle S',F',P'\rangle$ に対して $\mathrm{Sen}(\varphi)$ は、$\mathrm{Sen}(S,F,P)$ の文 ρ を $\mathrm{Sen}(S',F',P')$ の文 $\varphi(\rho)$ に写す。ここで $\varphi(\rho)$ は上の項、文の構成に関する帰納法で容易に定義できる。

以上で一階述語論理の Institution $\mathcal{I}^{\mathbf{FOL}}$ を構成できた。ここに $\mathcal{I}^{\mathbf{FOL}}$ の性質について纏めておく。

1. E を原子式の集合とする。このとき任意のモデル M に対して、$M\models E$ が成り立つこととモデル射 $0_E \to M$ が存在することは同値である。ここで $0_E = (T_E)/{=_E}$, $(0_E)_\pi = \{\langle t_1/{=_E},\ldots,t_n/{=_E}\rangle \mid \pi(t_1,\ldots,t_n) \in E\}$. 従って、原子式の論理積は基本的である.
2. $\mathcal{I}^{\mathbf{FOL}}$ は semi-exact (実は exact でもある),
3. X を定数の集合とするとき、シグニチャの拡大射 $\chi : \langle S,F,P\rangle \hookrightarrow \langle S,F\cup X,P\rangle$ は表現可能である。**FOL** では限量子を考えるとき、この形のシグニチャ射のみを扱う.

2.4　Horn 節論理の Institution

シグニチャ $\langle S,F,P\rangle$ に対する Horn 節 (Horn clause) とは、$(\forall X)(H\Rightarrow C)$ の形の $\langle S,F,P\rangle$-文である。ここで H は有限個の原子式の論理積、C は原子式である。

Horn 節論理の Institution $\mathcal{I}^{\mathbf{HCL}}$ とは、シグニチャとモデルに関しては一階述語論理の institution $\mathcal{I}^{\mathbf{FOL}}$ と同一であるが、文の集合に関しては Horn 節のみに制限したものである。以下、$\mathcal{I}^{\mathbf{HCL}}$ の性質について纏めておく。

1. 全ての表示射は liberal である,
2. Γ を Horn 節の集合とするとき、Γ-始モデルが存在する.

これによって、定理 2,3 より $\mathcal{I}^{\mathbf{HCL}}$ 上のモジュールでは、インポートにおける問合せに意味を持たせることができる。

3　文字体系の形式化

文献 [16] では、文字を機能として扱い、実現や過程に依存しない情報構造として形式化するため、「情報の流れ」理論に基づいた文字の表現を提出した。この理論に基づいた文字の表現は「分類」を用いる。文字の具体的な例示を「分類」におけるトークンで、文字の意味作用や性質を「分類」におけるタイプで表現する。

「分類」の圏は文献 [10] で定義された Twisted

Relation の圏 **Trel** と同じものである。Institution の圏は $\mathbb{C}at((-)^{\mathrm{op}}, \mathbf{Trel})^{\#}$ と同型であることが文献 [10] で示されている。ここで $\mathbb{C}at((-)^{\mathrm{op}}, T) : \mathbb{C}at^{\mathrm{op}} \to \mathbb{C}at$ は T への Hom 関手による indexed 圏、$\mathcal{K}^{\#}$ は indexed 圏 $\mathcal{K}$ に対する Grothendieck 構成を表わす。

この同型の対象部分に注目すると、institution は関手 $\mathcal{I} : \mathbb{S}ig \to \mathbf{Trel}$ と同じものであり、「分類」の圏はシグニチャの圏が singleton ときに対応している。これは「情報の流れ」理論で議論された概念を、institution 理論に拡張できる可能性があることを示す[*6]。以下では文献 [16] で議論した文字概念を、institution 理論で表現することを試みる。また、事例研究で問合せや推論に論理プログラミングを用いることを考慮して、Horn 節論理の institution を使用する。

3.1 文字体系の表現

Horn 節論理の institution $\mathcal{I}^{\mathrm{HCL}}$ に対して、一つの文字体系は $\mathcal{I}^{\mathrm{HCL}}$ 上の表示 $\langle \Sigma, E \rangle$ で表現する。ここで $\Sigma \in \mathbb{S}ig$ は $\mathcal{I}^{\mathrm{HCL}}$ のシグニチャ、$E \in \mathrm{Sen}(\Sigma)$ は Σ-文の集合である。

文字体系を表現するシグニチャ $\Sigma = \langle S, F, P \rangle$ は、具体的には以下のようなものとする[*7]。

- ソート記号の集合 S には文字を表現する char と、その他に韻 rime、古音分部 rgroup などを表現するソート記号が含まれる。
- 関数記号の集合族 F には少くとも、定数の集合 $F_{\lambda,\mathrm{char}}$ が含まれる。定数 $c \in F_{\lambda,\mathrm{char}}$ は特定の文字を表現する。
- 関係記号の集合族 P には文字の性質を表現する記号が含まれる。例えば、文字が古音分部に含まれるかどうかを示す関係記号として $\mathrm{in\_rgroup} \in P_{\mathrm{char,rgroup}}$ がある。

3.2 文字体系の統合

「情報の流れ」理論に基づいた文字体系の表現では、文字体系の統合はチャネル、或いは表現系で表わした。ここでは以下のようなシグニチャ射のダイアグラムで表現する。

$$\begin{tikzcd} & \Sigma & \\ \Sigma^1 \arrow[ur, "\varphi_1"] & & \Sigma^2 \arrow[ul, "\varphi_2"'] \end{tikzcd}$$

二つの文字体系を $\langle \Sigma^1, E^1 \rangle, \langle \Sigma^2, E^2 \rangle$ とする。この二つの文字体系を統合した体系 $\langle \Sigma, E \rangle$ は以下の条件を充たす。

- $\varphi^1(\Sigma^1) \cup \varphi^2(\Sigma^2) \subseteq \Sigma$,
- $\varphi^1(E^1) \cup \varphi^2(E^2) \subseteq E$.

ここで $\varphi^1, \varphi^2 : \mathrm{char} \mapsto \mathrm{char}$ とする。 $c \in F_{\lambda,\mathrm{char}}$ が存在して、 $\varphi^1(a) = c, \varphi^2(b) = c$ であるとき、 $a \in F^1_{\lambda,\mathrm{char}}$ は $b \in F^2_{\lambda,\mathrm{char}}$ の表現であると言う。核 c を介して二つの文字体系の文字が m 対 n で対応していることが表現できる。

4 段氏の音理

漢字には形音義の三つが関係している。形についての (計算機上での) 形式化には多くの研究がある。義については解決すべき多くの問題を抱えており、これからの課題であろう。本稿では音についての情報を中心に取り上げる。

以下の事例では、段注及び『六書音均表』を中心に扱う。ここで段氏の音理について必要な範囲でその概略を示す[*8]。

段氏の音理では、中古音の二百六韻をどのように上古の十七分部に配するかということが問題となる。一つの韻を分割して、一方はある部に、もう一方は他の部に配するということはしない。『廣韻』の韻と古音分部の関係を記述したものが『六書音均表』「今韵古分十七部表」である。例えば、上平一東韻は九部に所属する。

各字の中古音の韻への所属は、段注に記されている反切下字の『廣韻』での所属とする。また上古音は段注に記されている分部とする。中古音の韻の古音分部への所属は「今韵古分十七部表」に照して、決めることになる。段注での上古音の分部が「今韵古分十七部表」と異なるとき、その漢字の音韻は「今韵古分十七部表」に対して不規則な音

*6 Institution 理論はプログラミング言語の意味論や仕様記述等を、理論的に扱うために構築され、「情報の流れ」理論とは独立に発展したものである。

*7 音韻に関係する用語については後の節 4を参照。

*8 詳しくは、例えば [18] 等を参照してほしい。

韻変化をしたことになる。段注では「古音在幾部」と記されている。

例えば、五篇上可部の「奇」は『廣韻』の「上平聲五支」に属し、「今韵古分十七部表」では十六部であるが、段注では「古音在十七部」と記されている。十七部から十六部へと音韻変化をしたことが分かる。

5 事例

5.1 実現の方式について

実現方式及び事例の説明では、論理プログラミング言語 Prolog*[9]の用語を使用する。

事例研究で利用したプロトタイプでは、基本となるデータはデータベース管理システム `SQLite`*[10]によって実装し、述語表現及びデータベースへの問合せには Prolog の一実装である `SWI-Prolog`*[11]を使用している。 `SQLite`と `SWI-Prolog`の連携には、 *proSQLite*[2] を使用した。

まず、ある文献を解釈して得られる文字体系を一つのモジュールとして表現する。モジュールは文字についての性質を表現したプログラム節の集合 (institution $\mathcal{I}^{\mathbf{HCL}}$ 上の表示に相当) である。文字体系上での問合せは、そのモジュールを構成しているプログラム節の集合に対して行なわれるが、文字体系の統合は節 3.2で述べた方式を使用する。

5.2 基本的な述語

『説文解字注』『六書音均表』『廣韻』等の各々の文献を解釈して得られる文字体系を、別々のモジュールで表現している。各々のモジュール拡張子は `dz`、`ly`、`gy` である。以下の基本的な述語の説明には、述語の型とモード宣言の部分を示す。なお記述の関連でここでの述語名には省略がある。

■『説文解字注』の基本的な述語

- head_char(?Char:atom),
- pelement(?Char:atom, ?Pelem:atom),
- fanqie(+Char:atom, -Fanqie:atom),
- rgroup(+Char:atom, -Rgroup:atom).

それぞれの述語の意味は、Char が見出し字であること、その聲符が Pelem であること、その反切が Fanqie であること、その古音分部が Rgroup であることを示す。

■『六書音均表』の基本的な述語

- pelement_rgroup(+Char:atom, -Rgroup:atom),
- rime_rgroup(+Rime:atom, -Rgroup:atom).

それぞれの述語の意味は、聲符 Char の古音分部が Rgroup であること、今韻 Rime が古音分部 Rgroup に所属することを示す。

■『廣韻』の基本的な述語

- ph_pos(?Char:atom, ?Iname:atom, ?Iclass:atom, ?Rname:atom, ?Hu:atom, ?Div:atom, ?Fclass:atom).

『廣韻』の基本的な述語としては様々なものが考えられるが、ここでは文字の音韻地位に関連する述語についてのみ述べる。 ph_pos は Char の音韻地位を示す述語*[12]で、 Iname は声母、 Iclass は声母類 (唇音、牙喉音などの種類を示す)、 Rname は韻母、 Hu は呼 (開合)、 Div は等、 Tone は四声、 Fclass は韻類 (重紐 A,B 等を示す) である

5.3 段注における音韻の不規則変化

Fu が反切 F の反切上字であることを ufq(F, Fu)、Fl が反切 F の反切下字であることを lfq(F, Fl)であるとする。字 Char の今韻が古音分部 Rm に所属するという述語を rgroup_in_mc(Char, Rm)、上古音において分部 Ro に所属するという述語を rgroup_in_oc(Char, Ro) とする。

これらの述語及び「今韵古分十七部表」に対して不規則な音韻変化したという述語を、以下のように表現できる。

```
rgroup_in_mc(C, Rg) :- gy:fanqie(C, F),
  lfq(F, Lfq), gy:rime(Lfq, R),
  ly:rime_rgroup(R, Rg).
rgroup_in_oc(C, Rg) :- dz:rgroup(C, Rg).
ph_changed(C) :- rgroup_in_mc(C, Rm),
  rgroup_in_oc(C, Ro), Rm \== Ro.
```

*9 例えば、文献 [3] 等を参照してほしい。
*10 `http://www.sqlite.org`
*11 `http://www.swi-prolog.org`
*12 ここでは簡単のため、データベース理論における正規化に相当する手順は考慮していない。

ゴール ?- head_char(C), ph_changed(C). により、不規則な音韻変化をした字を列挙できる。ここでモジュールのインポートには、文字体系の統合化のための変換が必要である (節 3.2を参照)。

5.4 中古の重紐と上古分部

段注の例と同様にして、上で述べた基本的な述語、特に中古の音韻地位を使用して、重紐のある中古韻と上古分部との関係を議論することができる*13。重紐と上古分部の相関関係の強さがかつての上古の部分けに影響している。相関が強いほど、早い時期に部分けが行なわれている。

6 おわりに

本稿では、文字体系の表現と統合化を Horn 節論理の institution を利用して形式化し、また実装例として、論理プログラミングを用いた*14。この表現と統合化の方式を institution の具体化に依存しない形で形式化すること、及び「情報の流れ」理論の様々な概念の institution への拡張は今後の課題である。また、文字体系の表現を institution を用いて形式化することは、オントロジー工学で使用されるオントロジーマッピングなどの統合化技術との親和性を獲得し、多くの実装を利用する可能性を拡げるものとなろう。

文字体系における概念 (文字概念や文字についての情報を表現する概念) は、新しい設定においては新たな解釈がなされるため、複数の文字体系の統合化を容易にする仕組みが必要とされる。新しい媒体としてのテクストアーカイブズはそのような操作を可能にすることにより、これまでにない経験を生むとともに、その表象を通して新しい概念を創出する。これが更なる統合化という循環的な歩みを踏むことになろう。将来のテクストアーカイブズはそのような経験の場を提供するものとなる。

■謝辞 『説文解字注』の読書会「点注会」の皆さま、特に古勝隆一先生にはいつも有益なご教示を戴き、とても感謝しております。概要論文に対する査読者の方々、助言や助力を賜わった多くの方に感謝致します。最後にいつも支えてくれる妻留美と娘に感謝します。

*13 詳しくは、例えば文献 [17] を参照してほしい。

*14 文字体系の統合化を論理プログラミングで実用化するには、モジュールのインポート時にシグニチャ射の機能を自動的に実現する仕組みも必要であろう。

参考文献

[1] J. Barwise, J. Seligman: *Information Flow: The Logic of Distributed Systems*, Cambridge University Press, 1997.

[2] S. Canisius, N. Angelopoulos, L. Wessels: *ProSQLite*: Prolog file based databases via an *SQLite* interface, *Proc. of Practical Aspects of Declarative Languages*, LNCS 7752, 2013.

[3] W.F. Clocksin, C.S. Mellish: *Programming in Prolog*, Springer, 1981.

[4] R. Diaconescu: Herbrand Theorems in arbitrary Institutions, *Information Processing Letters* 90:29–37, 2004.

[5] R. Diaconescu: *Institution-independent Model Theory*, Birkhauser, 2008.

[6] J.A. Goguen, R. Burstall: Institutions: Abstract model theory for specification and programming, *J. ACM* 39(1):95–146, 1992.

[7] G.-G. Granger: *Pour la Connaissance Philosophique*, Éditions Odile Jacob, 1988. (邦訳 植木哲也訳:「哲学的認識のために」, 法政大学出版局, 1996.)

[8] Y. Kalfoglow, M. Schorlemmer: Ontology Mapping: the State of the Art, *The Knowledge Engineering Review* 18(1):1–31, 2003.

[9] S. MacLane: *Categories for the Working Mathematician*, Springer, 1971, 2nd ed., 1998.

[10] G. Roşu: Kan Extensions of Instituions, *J. Universal Computer Science* 51(8):482–492, 1999.

[11] M Schorlemmer, Y. Kalfoglow: Institutionalising Ontology-Based Semantic Integration, *Applied Ontology* 3:131–150, 2008.

[12] 門脇俊介: 理由の空間の現象学 — 表象的志向性批判, 創文社, 2002.

[13] 近藤和敬: グランジェの科学認識論 —「操作 -対象の双対性」、「形式的内容」、「記号的宇宙」, 所収 金森修編: 「エピテモロジー— 20 世紀のフランス科学思想史」, 慶應義塾大学出版会, 2013.

[14] 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」論文集, 2008.

[15] 白須裕之: 文字の指示概念に関する試論, 情報処理学会「人文科学とコンピュータシンポジウム」論文集, 2008.

[16] 白須裕之: 古辞書のテクストアーカイブズ構築について — 文字転写の理論とその応用, 情報処理学会「人文科学とコンピュータシンポジウム」論文集, 2011.

[17] 頼惟勤: 中國における上古の部と中古の重紐, 國語學 28, 1957.

[18] 頼惟勤監修, 説文会編: 説文入門, 大修館書店, 1983.