

古辞書のテキ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 —

白須裕之

概要

解釈を含めた言語テキストの電子的な表現、特に文字の表現はデジタルアーカイブズの構築において重要な課題である。文字を様々な情報を伝える表現として理論化するには、「情報の流れ」理論的なアプローチが有効である。ここでは文献の翻刻を再考することを通して、この理論を漢字の転写（或いは翻字）という面に適用し、「文字転写の理論」を構成する。また、その応用として、文字の電子的表現を汎用的な符号化文字集合の上でどのように実現するかを述べ、現在実装中の古辞書のテキストアーカイブズについても触れる。

A Theory of Representational Systems for Transcription of Kanji Characters

SHIRASU Hiroyuki

Abstract

It is an important subject on constructing digital archives to formalize electronic representations of literary texts. We take notice of transcription of chinese scripts, on which Kanji characters transform information of interpretation of their texts through the writing medium. The aim of this paper is to present a theory of transcription system for Kanji characters using the theory of information flow, in which J. Barwise and J. Perry tried to formalize F. Dretske's idea of representation. In addition, we discuss its application to a foundation of digital archives for ancient Chinese dictionaries.

1 はじめに

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、テキストアーカイブズ構築の基礎理論を手掛けてきた [7][8][9]。それは文字とテキストの表現という二つの階層について、各層の対象オブジェクトが何如に構成できるか、また二つの層の相互作用が如何なるものであるかを理論化するものである。

本稿では文献の翻刻を手掛りに、特に文字の表現について考察した。文字を様々な情報を伝える表現として理論化するには、文献 [7] で述べた「情報の流れ」理論的なアプローチが有効である。ここではこの理論を漢字の転写（或いは翻字）という面に適用して、「文字転写の理論」を構成し、文字の電子的表現をどのように実現するかを述べる。また、その応用として、現在実装中の古辞書のテキストアーカイブズについても触れる。

2 文献の翻刻について

2.1 文字コードによる翻刻

計算機を使った文字表現の問題について議論するために、まず文字コードによる翻刻について考えてみよう。

文字コードと一言でいっても、その規格によって使われている概念が微妙に異なる。文字をある種のビット列で符号化し、それを計算機で使用したり、情報交換のために使用するのが文字コードであり、それを実際に運用する様々な規格が必要とされる。ここでは文字コードについての詳細を議論するのが目的ではないので、利用できる文字集合がなんらかの方式によって決まってい、その文字集合に属する文字に符号が割り当てられているといった、符号化文字集合を念頭に議論を進めよう。

文献 [11] は JIS 漢字についての論考であるが、文献学の立場から研究に使用することができる文

字の電子的表現について述べているので、文献の翻刻に対して議論する上で非常に参考になる。

文献は、解釈を経て釈文(解釈文字列)に翻刻され、その翻刻釈文が符号化されるのであって、当然、解釈者・翻刻の目的によって釈文も異なり、その結果は別の符号になる。(文献 [11] 頁 3)

文字の符号化は、この様に一旦解釈された文字列が符号化されるのであって、原文が常に同一の符号列に変換される事を意味しない。...「原文→釈文」の解釈は言語・学問の分野に属し、符号化・工業規格の分野ではない。字形を字体で解釈するのも言語の問題(「漢字の常識」)であって、符号化の問題ではない。同様に、書体間の対応付けも言語・デザイン(芸芸)の問題である。(文献 [11] 頁 4)

上で引用した文献の翻刻の様子を示したものが、図 1 である。文献が解釈・符号化の二つの段階を経て、符号列へと変換される様子を示している。

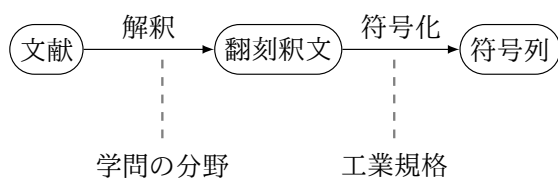


図1 文字コードによる翻刻

ここでの分析は「文献の符号化」において、解釈が果す役割を明確にしている。釈文(解釈文字列)で使われる文字は、文献の解釈と同時にその文字概念も構築されるはずであるが、翻刻に使用する文字コードの文字概念との関係が明確ではない。

これら二つの文字概念の関係を陽に示すように、文献の翻刻の様子を表現しなおしたものが図 2 である。但し、図 1 が文献が変換される様子を示しているのに対して、図 2 は、文献の翻刻に際して使用される「概念」の依存関係を示している。「釈文の文字概念」は文献の解釈から得られるが、それのみではなく、通常は翻刻するときの原則にも依存してい

る。こうして得られた「釈文の文字概念」を文字コードの文字概念に転写して、その後文献に対応する符号列が得られる。

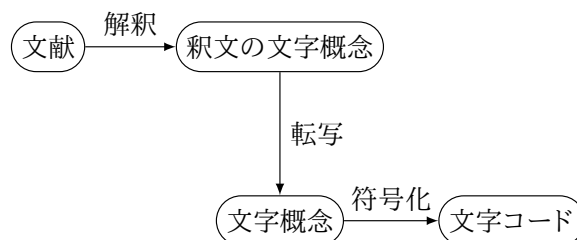


図2 釈文と転写

本稿はここで問題とした文字概念の間の「転写」の関係を一つの典型例として、「文字転写の理論」を構築する*1。この「転写」については文献 [11] の以下の指摘が重要な指針になる。

研究者が求める「原本に忠実な翻刻」とは、決して原本をそのまま引き写したものの事ではなく、原本の情報を伝えるべく適切に(明確化された方針によって)解釈された本文である。(文献 [11] 頁 5)

ここには研究に使用することができる文献や文字の電子的表現について、本質的なことが指摘されていると思う。奇しくも「情報」と「解釈」の言葉で表現されているが、我々の目指す翻刻とは、解釈を如何に情報化するかということを問うものであり、図 1 の左半分に対する情報学的なアプローチに他ならない*2。

2.2 文字概念について

以上見てきた文献の翻刻において、「転写」を理論化するための手掛かりを探るために、「文字概念」について、ここで簡単に議論しておこう。

*1 ここで想定している文字の転写とは、文字集合の間のマッピングを意味しない。このことは節 4 で明らかになる。また、一方の文字概念を文字コードの規格で決まるものに限定するという制約もない。

*2 電子化という情報の劣化を議論する場面が多かったが、電子化を「解釈」という積極的な行為として提出したものに、文献 [6] がある。

文字、特に漢字をどのようなものとして了解するかということに対しては、様々な立場が考えられる*3。まず文字と言語の関係をどのように規定するかということ言えば、大まかに分けて以下の二つの立場が想定できよう。

1. 文字は、言語のある要素を表現する視覚的な媒体であり、言語に属するものではない。
2. 文字は、言語の内的な要素であり、ある種の言語記号として取り扱う。

これらの立場は研究分野や研究対象とする文献の特徴にも依存するし、また文字言語と音声言語との関係をどのように捉えているかということにも関連する。第一の立場は、言語記号の表現面として音声を想定していて、文字には表現面としての地位が与えられていないことになる。また、第二の立場は文字を言語に属するものとして捉えているが、しかし、言語記号の一つと考えるのか、言語記号の表現面として想定するのかによっても、その立場を異にする。

通常の場合、文字コードにおける文字概念の規定も、第一の立場、すなわち言語要素を視覚的に表現するものとして、文字を捉える立場を採用している。さらにその文字集合は言語と独立であると考えるのが普通である*4。

文献 [10] では、統合化 Unification について述べた文脈で、“言語の「文字観念」とコード表の「文字観念」を安易に同一視した議論には従えない”とした上で、以下のように述べている。

コード表は、それ自体構造的に「文字観念」を表現するものではあるが、言語の特定共時態に基づく「文字観念」の表現ではない。(文献 [10] 頁 21)

従って、各言語を越えて文字を統合し、文字集合

を規格化するための、統合規則や抽象的形状を決める規則は言語と独立に与えられる*5。

以上、様々な文字概念の捉え方があったが、情報交換の手段としての文字は、どのような情報を交換したいかという目的によって、その捉え方が異なるのは当然と言えよう。本稿で「文字概念」と言った場合には、言語の構成に積極的に参与し、言語記号として扱うことができるものを指すことにする。これは節 6 で述べるように、本研究の目的が解釈を含む、テキスト表現の可能性を最大限に探求したいからである。

2.3 注釈と引用

本稿で扱う文字概念は言語学的なものである。従って、言語単位としての張り合いや差異を議論できるような体系を構成する必要がある。文字は文字体系の中で取り扱われなければならない。では文字体系間の関係を記述する「転写」を、言語学的な原理によって理論化することができるであろうか？文字体系(すなわち、言語体系)間の関係を理論化することは、翻訳の理論を構成することと同様な困難が待ち受けているであろう。言語記号の表現と内容の結び付きは恣意的な関係であり、二つの言語体系間の関係を言語学的な概念だけで説明するのは難しい。なんらかの文献学的な理解が必要になる。ここでは言語学的に真正面から取り組むのではなく、注釈と引用という文献学的な面から議論しよう。

まず対象とする文献テキストの解釈があり、語のような言語単位が浮びあがる。言語単位の形成から文字相互の結び付きの強さが分かり、文字概念を決めることができる。もちろん解釈と言語単位の循環的な関係も見逃せない*6。

*3 詳しくは、例えば文献 [5] 等を参照。

*4 汎用的符号化文字集合 Unicode における文字概念については、文献 [3] を参照。また、汎用的符号化文字集合によらずに文字を表現する方法、Chaon モデルについては [13][14] を参照。

*5 文字コードによる文字概念をこのように扱う故に、文献 [11] の翻刻は図 1 の構成で十分なのである。現に文献 [10] には、具体的に「コード表をだましまし使う法」として翻刻の仕方が書かれているが、これは解釈の結果得られる文字概念の、文字コードを利用した表現になっている。

*6 言語単位というものを理解することは、言語学の重要な課題であり、テキストの電子的な表現においても、将来の主要なテーマになるであろう。

一つの文献テキストには、訓詁の伝統によって様々な解釈があり得るし、また、その文献を引用したり、注釈を加えたりすることによる異なる解釈が存在する。文献で使用された文字に対する文字概念は、こうした解釈に依存して、様々な文字概念が構成されることになる。

対象とする文献テキストは、引用や注釈に使われる特定の文字体系によって解釈されるのであるから、対象となる文献の文字体系から、引用や注釈に使われる文字体系へのある種の転写が起っていることになる。すなわち、引用や注釈によって、対象の文字にも訓詁としての歴史の重みが重ねられていく。ここでは訓詁としての歴史の重みを表現することも、「文字転写の理論」の役割と考える。

次節以降、文字体系間の関係を記述する「転写」を、「情報の流れ」理論を使用した「表現系」を使って記述する。文献 [1] には表現系を利用した、様々なタイプの体系間の関係の例があげられているが、「転写」もその一つとして扱うことができる。

なお文献の解釈によって得られる「釈文の文字概念」を構成することについては、文献 [8] で述べたテキスト分析が有効であるが、その応用については今後の課題とする。

3 「情報の流れ」理論

本節では本稿に必要な範囲で、「情報の流れ」理論について纏めておく。詳しくは文献 [1] を参照願いたい。

定義 1 (分類 Classification, 制約 Constraint) 分類 $\mathbf{A} = \langle \text{tok}(\mathbf{A}), \text{typ}(\mathbf{A}), \models_{\mathbf{A}} \rangle$ とは、トークンの集合 $\text{tok}(\mathbf{A})$ 、タイプの集合 $\text{typ}(\mathbf{A})$ 、その上の二項関係 $\models_{\mathbf{A}}$ からなる。

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

定義 2 (情報射 Infomorphism) 分類 \mathbf{A} から分類 \mathbf{B} への情報射 $f : \mathbf{A} \rightarrow \mathbf{B}$ とは、二つの関数

$f^* : \text{typ}(\mathbf{A}) \rightarrow \text{typ}(\mathbf{B})$ と $f_* : \text{tok}(\mathbf{B}) \rightarrow \text{tok}(\mathbf{A})$ の対で、任意の $b \in \text{tok}(\mathbf{B})$ と $\alpha \in \text{typ}(\mathbf{A})$ に対して、 $f_*(b) \models_{\mathbf{A}} \alpha \Leftrightarrow b \models_{\mathbf{B}} f^*(\alpha)$ を充すものを言う (以下、適宜 $*$ を省略する)。

$$\begin{array}{ccc} \text{typ}(\mathbf{A}) & \xrightarrow{f^*} & \text{typ}(\mathbf{B}) \\ \vdots \Big\| \models_{\mathbf{A}} & & \Big\| \models_{\mathbf{B}} \vdots \\ \text{tok}(\mathbf{A}) & \xleftarrow{f_*} & \text{tok}(\mathbf{B}) \end{array}$$

定義 3 (シークエント Sequent, 分割 Partition) 集合 Σ に対して、 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 4 (帰結 Consequence, 理論 Theory) $\mathcal{P}(\Sigma)$ 上の二項関係 \vdash を Σ 上の帰結関係と言う。このとき $T = \langle \Sigma, \vdash \rangle$ を理論と言う。また、 $\text{typ}(T) = \Sigma$ と記す。

定義 5 (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 6 (局所論理 Local logic) 分類 \mathbf{C} 上の局所論理 $\mathcal{L} = \langle \mathbf{C}, \vdash_{\mathcal{L}}, N_{\mathcal{L}} \rangle$ とは、正則理論 $\text{th}(\mathcal{L}) = \langle \text{typ}(\mathbf{C}), \vdash_{\mathcal{L}} \rangle$ 、 $N_{\mathcal{L}} \subseteq \text{tok}(\mathbf{C})$ からなる。但し、 $N_{\mathcal{L}}$ は $\text{th}(\mathcal{L})$ の全ての制約を充たすトークンの集合である。 $N_{\mathcal{L}}$ の要素を正規トークン normal tokens と呼ぶ。

定義 7 (チャネル Channel) $\{\mathbf{A}_i\}_{i \in I}$ を分類の族とする。チャネル $\mathcal{C} = \{f_i : \mathbf{A}_i \rightarrow \mathbf{C}\}_{i \in I}$ とは、 \mathbf{C} への情報射の族であり、 \mathbf{C} をその核と言う。

定義 8 (表現系 Representation system)

表現系^{*7} $\mathcal{R} = \langle \mathcal{C}, \mathcal{L} \rangle$ とは、二項チャンネル $\mathcal{C} = \{f: \mathbf{A} \rightarrow \mathbf{C}, g: \mathbf{B} \rightarrow \mathbf{C}\}$ と、核 \mathbf{C} 上の局所論理 \mathcal{L} からなる。このとき \mathbf{A} のトークンを \mathcal{R} の表現と言う。

- $a \in \text{tok}(\mathbf{A})$ が $b \in \text{tok}(\mathbf{B})$ の表現 representation であるとは、 $\exists c \in \text{tok}(\mathbf{C}). f(c) = a \wedge g(c) = b$ のときを言い、 $a \rightsquigarrow_{\mathcal{R}} b$ と書く。また、 $c \in \text{tok}(\mathbf{C})$ が正規トークンのとき、的確な表現 accurate representation と言う。
- $\Gamma \subseteq \text{typ}(\mathbf{A}), \beta \in \text{typ}(\mathbf{B})$ に対して、 Γ が β を表示する indicate とは、 $f(\Gamma) \vdash_{\mathcal{L}} g(\beta)$ のときを言い、 $\Gamma \Rightarrow_{\mathcal{R}} \beta$ と書く。

4 文字転写の理論

本稿で構築する「文字転写の理論」は、複数の文字体系の関係を記述するものである。文字体系の関係は文字体系に含まれる文字の間のマッピングを定義するものではない。また、文字だけを対象にするのではなく、文字の性質をどのように関係付けるかということも、その記述に含まれている。

文字の転写に対して、前節の「情報の流れ」理論を応用する。基本的な考えは、転写についての解釈を記述することが、そのまま文字体系の関係を記述することになるように理論を構成する。

文字体系は文字をトークン、文字の性質をタイプとする分類で与える。ここでは簡単のために、文字体系を表現する分類に対して、そのトークンが文字であり、文字に対して扱える性質が一つだけであるように制限を加える^{*8}。

表現に利用する文字体系を分類 \mathbf{A} 、翻刻の対象となる文字体系を分類 \mathbf{B} で与えたとする。このと

き \mathbf{B} から \mathbf{A} への転写は、表現系 $\mathcal{R} = \langle \mathcal{C}, \mathcal{L} \rangle$ でモデル化する。但し、二項チャンネル $\mathcal{C} = \{f: \mathbf{A} \rightarrow \mathbf{C}, g: \mathbf{B} \rightarrow \mathbf{C}\}$ 、及び局所論理 \mathcal{L} は、以下のように \mathbf{B} の言葉で書かれた事象や性質を、 \mathbf{A} の言葉で表現するように決められる^{*9}。

- \mathbf{B} の文字 b を \mathbf{A} の文字 a に転写するとき、 a が b の表現になる ($a \rightsquigarrow_{\mathcal{R}} b$)。
- \mathbf{B} の性質 β が、 \mathbf{A} の性質 α に反映されるとき、 $f(\alpha) \vdash_{\mathcal{L}} g(\beta)$ とし、それを理論に自然に拡張する。

体系 \mathbf{B} には元々なかった制約が、体系 \mathbf{A} の制約を通して、体系 \mathbf{B} に押しつけられるということも、表現系を使用した理論では記述できる。これは文字コードでの翻刻において、文字コードでは区別できない積文の文字があったり、文字コードでは区別があるが、積文では区別しない文字があるなど、常に付き纏う現象を記述できるということである。

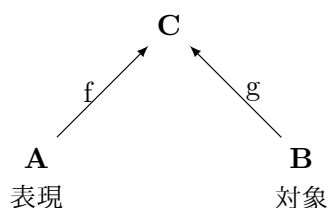


図3 文字転写のための表現系

4.1 文字転写の例

ここでは符号化文字集合 Unicode を使用した理論の構成の仕方を説明する^{*10}。

今、『説文解字』(これ以降、『説文』と略す)の親字を『廣韻』で解釈し、その実装を Unicode 上で実現すると仮定する。簡単のために、対象とする漢字の性質は中古音の韻と古音分部に限定する。表現系 $\mathcal{R} = \langle \mathcal{C}, \mathcal{L} \rangle$ の二項チャンネル $\mathcal{C} = \{f: \mathbf{A} \rightarrow$

^{*7} “Representation system” なる用語は元々は F. Dretske の構想に基づくもので、訳語としては「表象系」が適切かもしれない。「表象」という用語は表象主義を連想させるので、ここでは「表現系」を使用する。

^{*8} 分類によって表現できる一般の文字体系では、トークンが文字であるとは限らないし、複数の性質を取り扱うことも可能である。このような文字体系の一般的な構成法については、文献 [7] を参照してほしい。

^{*9} この表現系では \mathbf{A} が表現、 \mathbf{B} が対象である。チャンネルの左右が転写の向きと逆になっていることに注意。

^{*10} Unihan(Unicode Han Database) は、Unicode の文字集合の典拠になった文字集合へのマッピング、標準的な辞書への参照などの情報を保持している [4]。ここでは Unicode で翻刻するのではなく、標準的な辞書への参照情報を利用する。

$C, g : B \rightarrow C$ は以下である。

- ・『廣韻』による分類 A は、 $\text{tok}(A)$ を『廣韻』の親字の集合、 $\text{typ}(A)$ を『廣韻』における韻とし、その間の二項関係は親字の韻への所属とする。
- ・『説文』による分類 B は、 $\text{tok}(B)$ を『説文』の親字の集合、 $\text{typ}(B)$ を古音分部^{*11}とし、その間の二項関係は親字の分部への所属とする。
- ・Unicode による分類 C は、 $\text{tok}(C)$ を Unicode の文字集合とし、 $f : \text{tok}(C) \rightarrow \text{tok}(A), g : \text{tok}(C) \rightarrow \text{tok}(B)$ は Unihan で決められた辞書参照とする^{*12}。このとき $\text{typ}(C)$ は f, g より文献 [7] の定理 1 と同様の方法で決定する。

局所論理 \mathcal{L} は、全ての $\alpha \in \text{typ}(A), \beta \in \text{typ}(B)$ に対して、『六書音均表』『今韵古分十七部表』の韻と古音分部の関係 $f(\alpha) \vdash_{\mathcal{L}} g(\beta)$ から決まる制約から自然に定義する。また、正規トークンの集合 $N_{\mathcal{L}}$ は『今韵古分十七部表』に従う『廣韻』の親字である。

$a \rightsquigarrow_{\mathcal{R}} b$ は『廣韻』の親字 a が篆文 b を表現していることを示し、 $\alpha \Rightarrow_{\mathcal{R}} \beta$ は韻 α が古音分部 β に属することを示す。 $a \models_{\mathbf{A}} \alpha, b \models_{\mathbf{B}} \beta, a \rightsquigarrow_{\mathcal{R}} b$ であっても、 $\alpha \Rightarrow_{\mathcal{R}} \beta$ であるとは限らない。このとき『説文解字注』では「在古音分部 β 」と表現している。

4.2 文字検索の例

「文字転写の理論」の応用として、文字の検索をどのように解釈すれば良いかについて、その例を示す。『説文解字注』（これ以降、『段注』と略す）の親字に対応する『説文』の親字を検索すると仮定する。このときその検索を、例えば以下のような意味に解釈することができるだろう。

1. 似た字形を持つ漢字を検索する。

^{*11} 上古音の分部として、段玉裁に依るものを使用する。詳しくは、段玉裁『説文解字注』上海古籍出版社,1981. 等を参照。

^{*12} Unihan では『廣韻』への参照は定義されているが、『説文』への参照は定められていない。ここでは『漢語大字典』への参照を通して、間接的に『説文』への参照を決めることにする。

2. 同じ説解を持つ漢字を検索する。

以下ではこれらの二つの検索の意味を、「文字転写の理論」を使って説明してみよう。

4.2.1 字形による検索

電子テキストの場合は通常、字形による一番目の意味を採用していることが多いであろう。通常の文字コードによる翻刻では、この解釈に文字コードの同一性を持って代えている。

表現系では、例えばチャンネルにおいて共通の核を持つ漢字で解釈できる。『説文』による分類を A 、『段注』による分類を B とする。また、表現系は節 4.2 と同様に構成する。『段注』の親字 a に対応する『説文』の親字 b は、 $\exists c \in \text{tok}(C). a = f(c), b = g(c)$ を充す。

この解釈は f, g がどのような意味で定義されているかに依存する。Unicode を使用して『説文』を扱う場合には、字体による概念である Unicode best match が使用されている場合が多いように思う。文献 [2] では以下のように説明されている。

Unicode 4.0 *best match* ... contains an encoded *kaishu* form that is the current best available match for a seal form (in terms of the ongoing variant mapping/encoding process). The typeface for this concordance field was produced using undocumented features of *Wenlin* software. (文献 [2] 頁 365)

篆文に基づいて楷書へのマッピングを決め、部分字形に『説文』による解字に矛盾しないように、対応関係を決めていて、一つの見識を示していると思う。しかし、結局は引用でも述べられているように、その対応関係は実装に任されていて、それを厳密に文書化することはできていない。

4.2.2 説解による意味検索

第二の解釈は、同じ説解を持つ漢字を対応させる。『段注』には段玉裁の正字意識の反映として、屢々篆文の改変が認められる。このとき『段注』の

親字が『説文』の親字にどのように対応しているかは、第一の解釈では意図したものが得られない。そこで対応関係として、例えばここで提出する第二の解釈も有用になるであろう。

同じ説解を持つ漢字を対応させるということ、
「転写の理論」を使って述べよう。『説文』による分類を **A**、『段注』による分類を **B** とする。各々の親字のタイプには、その説解を取る。『説文』と『段注』の説解は必ずしも同じ字句ではないが、同じと見做すことができるものについては、理論の二項関係で関係付ける。このとき『段注』の親字 a に対応する『説文』の親字 b とは、 $\text{typ}(a) \vdash g(\beta)$ であるような説解 β を持つ親字 b のことである ($b \models \beta$)。

5 古辞書のテキストアーカイブズ

前節の理論を応用して、古辞書のテキストアーカイブズを WEB システム *Han Morphism System* として実装中である。なおこのシステムは近々公開の予定である。



図4 Han Morphism System の画面表示の例

本システムは UniHan の辞書へのマッピングを主に利用して、『廣韻』『説文』『段注』などの基本的な古辞書の情報を検索・表示する。図4は構築中のこのシステムの表示画面の例である。漢字「久」を検索して、各辞書が提供する情報を表示している。『廣韻』からは韻（及び、それが属する古音分部）、小韻、反切が、『説文』『段注』から説解（と注）が表示されている。また、韻や部首からはそれに属する漢

字の情報を表示する頁へのリンクがはられている。辞書間の連携、特に文字や韻、古音分部の対応、検索は転写の理論に基づく。

6 テキストの新たな表現に向けて

本研究の課題は、我々が近年手に入れた電子化という表現媒体の可能性を十分に活かすべく、言語テキストについてのデジタルアーカイブズ構築のための基礎理論を提出することにある。

人類は有史以前より、自分達が遭遇した事件、事実、それらに対面したときの気持ち、想念、考えなどを表現し、書き残すための様々な道具や技術を手に入れてきた。そのような技術の中から文字も生れた。我々が文字を記すために使用してきた媒体（例えば、甲骨、金石、竹簡、紙、版本、活字などの媒体）や、その媒体に記録するための技術が移り変ってきたことによって、その文字概念も少しずつ変様を受けてきたであろう。

文献 [12] では、近年のメディア技術によって、文字の性格が以下のように変様することに注目する。

ある出来事の発生とその発信、受信に関してはもはや私たちは本や新聞がメディアの中心だった時代と全く同じ能力を使っているわけではないのである。そして見方を変えれば文字を読む者はその人自身一つの「装置」とであることもできる。文字を読む者は「装置を備えて」おり、その装置を「外化」することで新たな技術が登場する。（文献 [12] 頁 127）

文献 [12] では、文字を読む行為について触れているが、文字概念そのものには触れていない。過去の歴史において、文字を表現する方法が文字概念をも変えてきたと同様に、文字コードによる技術が結果的に文字概念を少しずつ変えてきた。電子書籍や電子出版においては、まだその萌芽しか見られないが、近年のメディア技術は確実に文字概念にも変様を及ぼさずにはおかない。

節2でも述べたように、本研究の文字翻刻の構想においては、テキストの新たな表現へと、その一步を踏み出したように思う。文字の符号化から始まっ

た電子化は、すでに文字概念の再考の流れを止めることはできない。我々が現在立ち合っているのは、人文学における解釈や情報を追求する試みであり、またそれらを表現する枠組みの革新である。デジタル技術時代の人文学である Digital Humanities の枠を越えて、それらは「外化のプロセス」[12]として捉えた方が適切なのかもしれない。

7 おわりに

本稿では文献の翻刻の意味を再考することによって、「情報の流れ」理論を適用した文字転写の理論を提出した。また、その応用として、現在構築中である古辞書のテキストアーカイブズについて述べた。漢字の表現については、文献 [13] の先行研究があり、そのアプローチは個々の文字に注目した集合論的なものであるが、本稿は寧ろ文字体系の關係に注目する。

本稿では漢字の性質として中古音の韻、古音分部、説解しか扱えなかったが、漢字についての様々な性質を記述するためには、文献 [7] で提唱した定性的情報理論に基づく文字オントロジーを使用する必要がある。本稿で得られた知見を基に、漢字の様々な面を更に理論化し、将来的には総合的な漢字オントロジーの構築を目指す。

■謝辞 *Han Morphism System* の Ver.0.3.1 までの実装は、一般財団法人 人文情報学研究所の援助によって作成されました。所長の永崎研宣さんには様々な便宜をはかって頂きました。ここに感謝致します。このシステムで使用している『廣韻』、『段注』のデータは「漢字データベース」に依っています。データの作成に携わっていらっしゃるプロジェクトの皆様には感謝します。

日頃から『説文解字注』についてご教示下さる「点注会」の皆様、特に古勝隆一先生には多くの教えを頂いております。野村英登さんには文献の解釈と翻刻について色々ご教示頂きました。概要論文に対して、査読頂いたお二人の方、プログラム委員長の鈴木卓治様には有益なコメントを頂戴しました。その他、多くの助言、助力を賜わった多くの方々に感謝致します。最後にいつも支えてくれる妻留美と娘に感謝します。

参考文献

[1] J. Barwise, J. Seligman: Information Flow: The Logic of Distributed Systems, Cam-

bridge University Press, 1997.

[2] R.S. Cook: 《説文解字・電子版》 Shuo Wen Jie Zi - Dianzi Ban: Digital Recension of the Eastern Han Chinese Grammaricon, PhD Dissertation, Department of Linguistics, University of California, Berkeley, 2003.

[3] J.D. Allen, et al. eds.: The Unicode Standard, Version 5.0, The Unicode Consortium, 2006.

[4] J.H. Jenkins, R. Cook, eds.: Unicode Han Database (Unihan), Unicode Standard Annex38, 2006 – 2010.

<http://www.unicode.org/reports/tr38/>

[5] 犬飼隆: 言語の要素としての文字の位置, 日本語学論集「小松英雄博士退官記念」, 三省堂, 1993.

[6] 白須裕之: 記号機能としてのアーカイブズ, 人文科学とコンピュータシンポジウム論文集, 2008.

[7] 白須裕之: 文字の指示概念に関する試論, 人文科学とコンピュータシンポジウム論文集, 2008.

[8] 白須裕之: テキストアーカイブズにおける文字概念について, 人文科学とコンピュータシンポジウム論文集, 2009.

[9] 白須裕之: 書記行為の再現系としてのテキストアーカイブズ, 人文科学とコンピュータシンポジウム論文集, 2010.

[10] 豊島正之: 「JIS に無い字」をめぐって, 月刊しにか 3(2), 大修館書店, 1992.

[11] 豊島正之: JIS 漢字批判の基礎知識, 第 8 回「日本語の文字と組版を考える会」, 講演予稿, 1998.

[12] 中村大介: フランス技術哲学の現在 — ベルナル・スティグレールの技術哲学の射程 —, 人文論究 53(3), 関西学院大学, 2003.

[13] 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.

[14] 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.