

# 説文小篆に対する漢字構造記述の試み

守岡 知彦

## 1 はじめに

現代の漢字字体は楷書を基礎にしているが、楷書は唐代初期までに長安宮廷写経に見られるような初唐標準字体というゆるやかな字体標準としてまとめられるに至ったが、その後、開成石経の字体が規範字体化し、[4][3]それが宋版によって定着しいわゆる『康熙字典』のような印刷用規範字体として通用するに至った。<sup>\*1</sup>この開成石経規範字体は説文解字の影響を強く受けていると考えられ、この問題を分析する上で楷書と説文解字の見出し字の小篆（説文小篆）の対応関係を記述することが重要といえる。

また、現代漢字と説文小篆の対応関係を記述することは、先秦時代～小篆等の古代の字体から隷書・楷書に基づく現代の漢字字体への変化（いわゆる隷変）において生じたバリエーションをまとめ、さまざまな形に変化した機能的部品を理解する上でも重要であるといえる。

そこで、鈴木俊哉氏に提供して頂いた説文小篆のグリフデータを元に CHISE 文字オントロジー [2] 上で文字定義を行い、それに漢字構造記述を付与することを試みた。

## 2 対象とした資料

鈴木俊哉氏に提供して頂いた説文小篆のグリフデータは京都大学人文科学研究所附属東アジア人文学情報学研究センターの東方学デジタル図書館で公開されている 2 種類の説文解字

- 説文解字 十五卷（漢許慎撰 宋徐鉉等奉敕校定 據汲古閣本重刊 有惠棟席世昌校語席湜識語圖記）<sup>\*2</sup>
- 説文解字十五卷 埤汲古閣説文解字校記一卷（光緒七年淮南書局據汲古閣第四次様本重刊）<sup>\*3</sup>

の見出し字の小篆を切り出し、元画像における矩形座標と幾つかのメタデータを付したものである。[6]

---

<sup>\*1</sup> 日本では初唐標準字体を大量に受容したために、近代以前の通用字体の基礎となっており、戦後の漢字改革で導入された当用漢字、常用漢字においてもこれらの通用字体が利用されたため、結果的に、中国に比べて初唐標準字体の影響が多く残っているといえる（その結果、いわゆる旧字体よりも新字体の方が古い字体である場合が少なくない）。

<sup>\*2</sup> <http://kanji.zinbun.kyoto-u.ac.jp/db-machine/toho/html/A020menu.html>

<sup>\*3</sup> <http://kanji.zinbun.kyoto-u.ac.jp/db-machine/toho/html/A024menu.html>

CHISE 文字オントロジーではこれらの説文小篆にそれぞれ `===shuowen-jiguge5` と `===shuowen-jiguge4` という字形粒度の CCS 素性 (ID 素性) [1] を割り当てた。

また、これらを統合した抽象字体用 CCS 素性として `=shuowen-jiguge` を定義した。<sup>\*4</sup>この CCS 素性を持つ小篆字体オブジェクトと `===shuowen-jiguge5` もしくは `===shuowen-jiguge4` という CCS 素性を持つ小篆字形オブジェクトの間に

小篆字体オブジェクト → `subsumptive` → 小篆字形オブジェクト


のように包摂関係素性 `->subsumptive` を張ることにする。但し、この2つの説文小篆字形オブジェクトの漢字構造が異なったり部品の形状が大きく異なるなど字体粒度の差異を越える大きな差異があると認められる場合、

小篆字体オブジェクト → `denotational` → 小篆字形オブジェクト

のように包摂関係素性 `->denotational` を張ることにする。

### 3 漢字構造の変換


もしある説文小篆字体と対応する現代漢字字体が同様の漢字構造を持っているとすれば、その現代漢字字体の漢字構造記述に現れる各部品を対応する説文小篆部品に置き換えれば説文小篆字体の漢字構造記述ができるはずである。

例えば、 は「説」に対応する説文小篆字体であるが、

説  $\xrightarrow{\text{漢字構造}}$   言 兌 (1)


言  $\xrightarrow{\text{説文小篆}}$   (2)

兌  $\xrightarrow{\text{説文小篆}}$   (3)

であるので、「 言 兌」を式 (2), (3) で書き換えることで、

 説  $\xrightarrow{\text{漢字構造}}$   言  兌 (4)

が得られる。

一般にある説文小篆字体に対応する現代漢字字体は複数ありうるが、このような漢字構造の対応関係が存在するものであればそのうちのどれを選んでも良い。例えば、「説」も  に対応する現

<sup>\*4</sup> 現状、字体という概念は厳密には楷書系書体でしか定義できず、小篆における字体という概念は便宜上のものにならないが、漢字構造記述可能な抽象形状であって漢字構造が一定の範囲内にあると認められるものを仮に字体の代替物として用いることにする。

代漢字字体であるが、これも「説」と同様に




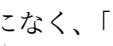





$$\text{説} \xrightarrow{\text{漢字構造}} \square\square\text{言兌} \quad (5)$$

$$\text{言} \xrightarrow{\text{説文小篆}} \begin{array}{c} \text{言} \\ \text{兌} \end{array} \quad (2)$$

$$\text{兌} \xrightarrow{\text{説文小篆}} \begin{array}{c} \text{兌} \\ \text{兌} \end{array} \quad (6)$$

から式 (4) が得られる。この「説」と「説」のように説文小篆と同様な漢字構造の対応関係を持つ現代漢字の集合を『同構造（字体）集合』と呼び、また、そのような関係を『同構造関係』と呼ぶことにする。また、同じ説文小篆字体に対応する現代漢字字体の集合を『同系（字体）集合』と呼び、また、そのような関係を『同系（関係）』と呼ぶことにする。


ここで、同系な字体同士が必ずしも同構造関係を持つ訳ではないことに注意しよう。つまり、隷変時のバリエーションの差によって異なる漢字構造になった現代漢字字体は多数存在する。逆に、同構造関係であっても異体字関係になく同系でない例も考えられるが本稿ではこの問題については議論の対象としないことにする。

ところで、「説」と「説」は同構造であり、その異なる部品「兌」と「兌」は同じ説文小篆字体  に対応する。ここで、「兌」の漢字構造を「兌儿」、 の漢字構造を「兌儿」とすると、「兌」と「兌」は同構造関係になく、「兌」が  の部品  に対応する一方、「兌」は  に対応しない。この時、「説」は  と2階の構造対応があると言い、「説」と  の対応関係は1階とすることにする。このように、「説」は「説」よりも構造対応の階数が高く、同構造集合の中で最も階数の高い現代漢字字体のことを『最近似（現代漢字）字体』と呼ぶことにする。

説文小篆字体の漢字構造記述の生成に際しては最近似字体の漢字構造を基に各部品の小篆部品への書き換えによって行うこととする。

## 4 漢字構造の調整



現代漢字と説文小篆で字体の輪郭パターンが異なるために、部品を組み合わせた時の食い込みの有無・差異が生じることがあり、このために IDC を変換する必要がある場合がある。

例えば、「進」に対応する説文小篆字体  の漢字構造を考える。

$$\text{進} \xrightarrow{\text{漢字構造}} \square\square\text{辵佳} \quad (7)$$

$$\text{辵} \xrightarrow{\text{説文小篆}} \begin{array}{c} \text{辵} \\ \text{佳} \end{array} \quad (8)$$

$$\text{佳} \xrightarrow{\text{説文小篆}} \begin{array}{c} \text{佳} \\ \text{佳} \end{array} \quad (9)$$

から  が得られるが、実際には  とすべきであろう。


これを一般化すると、

$$\boxed{\text{走}}x \xrightarrow{\text{説文小篆}} \boxed{\text{走}}x \quad (10)$$

となるが、同様なケースとしては

$$\boxed{\text{走}}x \xrightarrow{\text{説文小篆}} \boxed{\text{走}}x \quad (11)$$

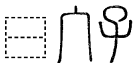
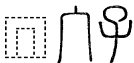
などが挙げられる。<sup>\*5</sup>

同様に、「字」に対応する説文小篆字体  の漢字構造は

$$\text{字} \xrightarrow{\text{漢字構造}} \boxed{\text{宀}}\text{子} \quad (12)$$

$$\text{宀} \xrightarrow{\text{説文小篆}} \text{宀} \quad (13)$$

$$\text{子} \xrightarrow{\text{説文小篆}} \text{子} \quad (14)$$

から  が得られるが、実際には  とすべきであろう。

これを一般化すると、

$$\boxed{\text{宀}}\text{子}x \xrightarrow{\text{説文小篆}} \boxed{\text{宀}}\text{子}x \quad (15)$$

となるが、同様なケースとしては



$$\boxed{\text{宀}}\text{子}x \xrightarrow{\text{説文小篆}} \boxed{\text{宀}}\text{子}x \quad (16)$$

などが挙げられる。

## 5 部品の問題


### 5.1 説文にない部品




説文解字は意符や声符として言及した部品の文字エントリーが欠けてるものがある。

例えば、説文の  (壯) の説解には「大也。从土，月聲。」とありこの声符は「月」であるが、説文には「月」の項目は存在しない。なお、甲骨文字には  があるので、当面はこれを流用することにした。

同様に、「由」も幾つかの字で声符として言及されているにもかかわらず、説文に文字としての項目がない。この字の場合、甲骨文字は形状が大きく異なるため、当面は現代字の「由」を流用しているが、こうすると現代字の検索に悪影響を及ぼすため、小篆専用の部品を作ることが望ましい。



<sup>\*5</sup> 「迨」や「速」の説文小篆は「𠂔」を用い(10)の例外である。





あるいは、説文の  (𪗇) の説解には「齒差跌兒。从齒，佐聲。」とありこの声符は「佐」であるが、説文には「佐」の項目は存在しない。<sup>\*6</sup>このケースの場合、分解して IDS で表現することは可能である。




 (萬) の場合、説解に「蟲也。从内<sup>\*7</sup>，象形。」とあるが、この「萬」の本体である  の項目はない。しかしながら、 のようにこの省略形が使われる場合がある。この部品は象形であるので他の部品の組合せで表現することは望ましくなく、こうした場合、小篆部品として追加することが望ましい。

## 5.2 省略された部品







説文解字は意符や声符が元の字を省略した部品であると述べる場合がある。この場合、その省略した部品字体の文字エントリーは存在しない。また、どのように省略したか（省略された部品字体がどのような部品をどのように組み合わせたものか）といった部品字体の形状に関する説明は存在しない。

例えば、 (余) の説解には「語之舒也。从八，舍省聲。」とあり、この声符は「舍」の省略形であることが判るがその具体的な形状に関する説明はない。「舍」ではこれを「从宀，中，象屋也。」としているが、甲骨文字や金文などからこれが「余」の初文であることが判る。そこで、当面は甲骨文字  を流用することにした。

 (𪗇) の場合、説解に「臥也。从宀，侵聲。」とあり、この声符は「侵」の省略形であることが判るがその具体的な形状に関する説明はない。「侵」の説文小篆は  であるが、 と見比べれば、その右側の  (𪗇) であることが判る。なお、この部品は生産性が高く、現代漢字との対応関係の記述の観点からも、小篆部品として追加することが望ましい。


 (𪗇) の場合も同様に、説解に「玉色。从玉，熒省聲。」とあり、この上部の部品  火火八 は  (𪗇) の省略形であることが判る。前述のように IDS で表現可能であるが、これは現代漢字の部品「𪗇」（簡体字では「𪗇」、日本の簡略字形では「𪗇」）に対応し、この部品の生産性の高さを鑑みれば、対応する小篆部品を定義した方が良いといえる。







## 5.3 複合部品・繁体部品等

 (𪗇) の右側の部品   は  (𪗇) の異体であると考えられる（漢字構造的には の方が  の簡略字といえる）が、説文にはこの文字の項目はない。前述のように IDS 表

<sup>\*6</sup> 徐鉉の注によれば、「臣鉉等曰：説文無佐字。此字當从尪，傳寫之誤。」とある。ちなみに、「佐」が出て来るのはこの1箇所だけなのである。











<sup>\*7</sup> 段注では、「𪗇」に作る。

記は可能ではあるが、現代漢字との関係や小篆の異体部品関係やこれを部品として持つ小篆字体およびその簡略字体  を部品として持つ小篆字体との関係を記述する上でこの複合部品も小篆部品として追加することが望ましい。






 (春) は、説解に「推也。从艸从日，艸春時生也，屯聲。」とあるように、 (艸) と  (屯) と  (日) という3つの部品からなる字であるが、上2つの   は現代漢字の部品「𦰇」に対応し、この部品の生産性の高さを鑑みれば、対応する小篆部品を定義した方が良いといえる。



## 5.4 異体部品








「回」に対応する説文小篆字体  は

-   :  (烟),  (網)
-   :  (迴),  (坳),  (駟),  (洞),  (肩)
-   :  (詠)

のように文字(や版)によってその部品字体が変化している。

「其」の場合、説文小篆字体は  であるが、 や  のように部品としては   という字体もとる。

「言」も「詩」の古文  (詛) で通常の小篆部品  とは異なる字体をとる。

また、 (遣) の右側の部品のように説文の「盧」の項目の見出し字  と部品も食い込みの有無も異なっているようなケースもある。なお、「遣」の説解に「籀文从盧」とあり、この部品は通常の小篆ではなく籀文の部品なのかも知れない。例えば、 (置) では  に似た字体の部品が使われている。また、 や  では  と構成部品は似ているが部品も食い込みの有無が異なっている。<sup>\*8</sup>

このように部品に複数の部品字体が存在する場合、そのどれを選ぶかはケースバイケースとなる。また、このような異体部品の書き分けを行う場合、異体部品が説文小篆の別体として存在する場合はそれを使えば良いが、ない場合は部品字体を追加する必要がある。



## 5.5 部品対応の曖昧性




現代漢字の部品字体が複数の小篆部品に対応する場合、そのどれを選ぶかはケースバイケースとなる。説文の説解が参考になる場合も多いが、これを自動化するためには説解の機械可読化を行わなければならない。ただ、数の少なさを鑑みれば、人手で確認する方が現実的かも知れない。また、

\*8 食い込みの有無は説文小篆でも現代漢字でもしばしば揺れるので、字体としては捨象した方が良いのかも知れない。

将来的には OCR 技術の利用も考えられる。

## 5.6 版による差異

例えば、「半」に対応する説文小篆  は段註では  となっており、漢字構造が微妙に異なっている。

あるいは、汲古閣系の版本の間でも小篆字体が異なる場合がある。[5] 例えば、「正」の説文小篆は `===shuowen-jiguge5` では  であるが、`===shuowen-jiguge4` では  である。なお、部品字体としては `===shuowen-jiguge5` でも  が用いられる。

こういう場合、記述対象となる小篆の版を明確化し（記述対象となる版に ID を付与するなど）、漢字構造を記述すべきであろう。この際、必要に応じて異体部品を整備することも重要であろう。

## 6 おわりに

CHISE 文字オントロジーに収録された現代漢字字体に対する漢字構造記述と現代漢字と説文小篆の対応関係のデータを用いて、説文小篆に対する漢字構造記述の自動生成を試みた。本稿で述べた手法を用いて、京都大学人文科学研究所所蔵汲古閣通行本説文解字 (`===shuowen-jiguge5`) の見出し字にある説文小篆 11114 字体に対しその 80% 程度、9000 字体程度が変換できた。なお、この結果生成した説文小篆の漢字構造記述は CHISE 漢字構造データベースの一部として

<https://gitlab.chise.org/CHISE/ids/-/blob/master/IDS-SW-JIGUGE.txt>

で公開している他、CHISE-wiki (EgT) 等の CHISE 関連 Web サービス上でも公開している。

本稿で用いた手法を適用して良い結果を得るためには、現代漢字字体と説文小篆字体の適切な対応関係の記述、特に、『最近似字体』を適切に設定するとともにその漢字構造を適切に記述することが重要であるといえる。このことは、多粒度漢字構造記述における異なる粒度間での文字間の包摂関係と部品間の包摂関係に関する両対応（自然変換）を実現するのと同様といえ、いわば超抽象文字粒度として説文小篆をすえた上で、現代漢字と説文小篆の間に文字（字体）間の関係と漢字構造（部品）間の両対応（自然変換）を実現するような記述を求める問題と捉えることができる。そして、これを実現することは漢字構造記述に対する有効な制約となり、その恣意性を低め、安定性を向上させることに寄与するといえる。

## 謝辞

説文小篆のグリフデータを提供して頂いた鈴木俊哉氏に感謝する。

## 参考文献

- [1] 守岡 知彦. *CHISE のデータ形式 (Ver.0.1)*. <http://git.chise.org/~tomo/character/chise-format.pdf>. Aug. 2017.
- [2] Tomohiko Morioka. “Multiple-policy Character Annotation based on CHISE”. In: *Journal of the Japanese Association for Digital Humanities* 1.1 (Nov. 2015), pp. 86–106.
- [3] 池田 証壽. “漢字字体史の資料と方法：初唐の宮廷写経と日本の古辞書”. In: **北海道大学文学研究科紀要** 150 (Dec. 2016), pp. 201–236.
- [4] 石塚 晴通 and 高田 智和. “漢字字体と文献の性格との関係 — 「漢字字体規範史データベース (石塚漢字字体資料)」の文献選定”. In: **漢字字体史研究 二 — 字体と漢字情報**. Ed. by 石塚晴通 監修 et al. 勉誠出版, Nov. 2016, pp. 349–359.
- [5] 鈴木 俊哉. “清刊大徐本説文解字の版本評価の再検討に向けて”. In: **広島大学大学院総合科学研究科紀要. II, 環境科学研究** 11 (Dec. 2016), pp. 77–100.
- [6] 鈴木 俊哉. “テンプレートマッチを用いた岩崎本説文解字影印本の加筆部分推定”. In: **東洋学へのコンピューター利用 第 30 回研究セミナー**. Mar. 2019, pp. 231–285.