



# upL<sup>A</sup>T<sub>E</sub>X を用いた多言語文献目録の組版

## Typesetting of Multilingual Bibliography for Oriental Studies using upL<sup>A</sup>T<sub>E</sub>X

守岡 知彦 Tomohiko Morioka

京都大学人文科学研究所 tomo@zinbun.kyoto-u.ac.jp

KEYWORDS multilingual typesetting, private used character processing, upL<sup>A</sup>T<sub>E</sub>X

ABSTRACT This paper describes typesetting of “Annual Bibliography of Oriental Studies” (「東洋學文獻類目」; ABOS) as a case study of multilingual typesetting. ABOS is a multilingual bibliography of oriental studies, including various languages and scripts such as Japanese, Chinese, Korean, English, French, German, Russian, other European languages, Vietnamese, Thai, Latin transcriptions of Sanskrit, Tibetan, Arabic, etc., IPA phonetic symbols, Ancient Chinese scripts such as Oracle-Bone inscriptions (甲骨文字), Bronze inscriptions (金文), Chu bamboo scripts (楚簡文字) and their modern transcriptions (隸定文字), etc. Most of characters included in ABOS are included in UCS, however some characters/scripts are missing, for example Oracle-Bone script (e.g., 𠩺 and 𠩻), other ancient Chinese scripts (e.g., 𠩺 and 𠩻) and their modern transcriptions (e.g., 𠩺 and 𠩻). This paper describes the current typesetting system based on upL<sup>A</sup>T<sub>E</sub>X briefly.

## 1 はじめに

upL<sup>A</sup>T<sub>E</sub>X [12] を用いた「東洋學文獻類目」(以下、「類目」と略す)の組版の事例について述べる。

類目は東洋学研究に関する内外の文献について、その対象年の日付のある雑誌・論集所載の論文と単行本とを、それぞれ分類・排列し、著者索引を付けたものであり、現在は、京都大学人文科学研究所東アジア人文情報学研究センターがその編纂を行っており、2001 年度版以降、著者がその組版システムの開発と組版を行っている。

類目にはさまざまな言語で書かれた文献が収録されており、その組版システムは日本語、中国語、朝鮮語、英語、フランス語、ドイツ語、ロシア語、その他欧州諸語、ベトナム語、タイ語、サンスクリットやアラビア語等のラテン転写、IPA 発音記号、甲骨文字、金文、楚系文字、および、それらの隸定文字等のさまざまな言語・用字系 (script) をサポートする必要がある。類目に含まれる文字の大部分は Unicode に収録されたものであるが、甲骨文字 (例: 𠩺、𠩻) やその他先秦期の文字 (例: 𠩺、𠩻) およびその隸定字 (例: 𠩺、𠩻) 等で毎年必ず外字が発生する。そのため、Unicode にない字が扱える必要がある [9]。

ここでは、upT<sub>E</sub>Xを用いた現行の組版システムについて概説する。

## 2 「東洋學文獻類目」とは

「東洋學文獻類目」(類目)は、1934年度以来逐年編纂されている「東洋史研究文獻類目」を1963年度版から改称したもので、東洋学研究に関する内外の文献について、その対象年の日付のある雑誌・論集所載の論文と単行本とを、それぞれ分類・排列し、著者索引を付けたものである(図1)。

- 13167 叶伟奇 著: 晚清福建船政铸币史, 中國錢幣 2006 年 1 期 p.12-14  
 13168 弓月 著: 华俄道胜银行及其在新疆发行的金币券:新疆钱币图说(2), 中國錢幣 2006 年 1 期 p.61-63  
 13169 孟海华 / 金普军 / 毛振伟 著: 西藏铜币发展史简述, 東南文化 2006 年 1 期 p.83-85  
 13170 徐渊 著: 百元债券百年纪:兼与《中华革命军第一票》作者商榷, 中國錢幣 2006 年 4 期 p.34-39  
 13171 李涛 他 著: 西藏铜币的科技分析, 中國錢幣 2006 年 1 期 p.42-44  
 13172 段洪刚 著: 论开国纪念币一文铜元的造币厂归属问题, 中國錢幣 2006 年 1 期 p.20-23  
 13173 徐渊 著: “中华民国中央银行钞票”揭秘, 中國錢幣 2006 年 1 期 p.29-38  
 13174 孙浩 著: 中国近代机制币大师耿爱德生平简介:在华外籍人士小传(5), 中國錢幣 2006 年 3 期 p.89-90  
 13175 弓月 著: 中央银行新疆省流通券:新疆钱币图说(5), 中國錢幣 2006 年 4 期 p.71-72  
 13176 张小杰 著: 广东中国银行特别券, 中國錢幣 2006 年 1 期 p.39-41  
 13177 弓月 著: 殖边银行新疆专用红钱票:新疆钱币图说(4), 中國錢幣 2006 年 3 期 p.91-92  
 13178 弓月 著: 阿尔泰通用银券:新疆钱币图说(3), 中國錢幣 2006 年 2 期 p.69-70  
 13179 孙浩 著: 老上海中央造币厂的洋技师赫维特:在华外籍人士小传(3), 中國錢幣 2006 年 1 期 p.58-59  
 13180 郭洲 著: 陕北地方实业银行及其发行的纸币, 中國錢幣 2006 年 4 期 p.40-41  
 13181 罗词安 著: 新发现湘赣省工农银行壹元“永新”券, 中國錢幣 2006 年 1 期 p.49-50  
 13182 李怀德 著: 江西省万载县革命烈士纪念馆藏工农兵银行石印版介绍, 中國錢幣 2006 年 2 期 p.52-53  
 13183 张建新 著: 川陕苏区“改制代用币”初论及藏品辨析, 中國錢幣 2006 年 1 期 p.47-48  
 13184 姚杰 著: 国家博物馆藏陕甘边区南梁油布币探究, 中國錢幣 2006 年 3 期 p.73-78  
 13185 解武军 / 毛赛蓉 著: 湘赣省革命战争公债券考略, 中國錢幣 2006 年 3 期 p.69-72  
 13186 周银霞 著: 陕甘宁边区票证及教育文献, 文博 2006 年 2 期 p.65-71  
 13187 马怀光 著: 介绍两张中国人民银行解放初期金融票据改用的支票, 中國錢幣 2006 年 2 期 p.57-59  
 13188 陈桥 著: 中国人民志愿军随军银行及其“志愿军赔家汇款特种票汇”考述, 中國錢幣 2006 年 2 期 p.54-56  
 13189 伊斯拉菲尔·玉苏甫 / 安尼瓦尔·哈斯木 著: 新疆博物馆馆藏波斯萨珊朝银币, 中國錢幣 2006 年 3 期 p.80-85  
 13190 津村 貞輝子 著: サーサニ式銀貨につけられた「擦痕」は何か?:新疆ウチャ出土一括コインのカウンターマークとの関係から見た新知見, オリエンツ 49 巻 2 号 p.40-69  
 13191 李涛 / 毛振伟 / 金普军 著: 西藏铜币的 XRF 分析及其来源初探, 西藏研究 2006 年 1 期 p.83-89

### 1.13.8 古文書

- 13192 池田 知久 / 西山 尚志 谈: 出土资料研究同样需要:“古史辨”派的科学精神:池田知久教授访谈录, 文史哲 2006 年 4 期 p.21-30  
 13193 李天虹 著: 释曾侯乙墓竹简中的“𠂔”, 古文字研究 26 輯 p.303-307  
 13194 萧圣中 著: 曾侯乙墓竹简编联与缀合关系的局部调整与考察, 武漢大學學報(人文科學版) 59 卷 5 期 p.606-613  
 13195 何有祖 著: 楚简释读七则, 江漢考古 2006 年 1 期 p.91-93  
 13196 侯乃峰 著: 竹书《周易》“采”字补释, 古文字研究 26 輯 p.357-359  
 13197 冯胜君 著: 有关战国竹简国别问题的一些前提性讨论, 古文字研究 26 輯 p.314-319  
 13198 刘乐贤 著: 楚秦选择术的异同及影响:以出土文献为中心, 歷史研究 2006 年 6 期 p.19-31  
 13199 刘信芳 著: 上博藏竹书所载殷高宗政令及相关问题, 中國歷史文物 2006 年 5 期 p.60-62  
 13200 刘信芳 著: 信阳楚简 2-04 号所记车马器研究, 古文字研究 26 輯 p.233-236  
 13201 刘信芳 著: 关于竹书“错别字”的探讨, 考古 2006 年 10 期 p.66-76  
 13202 卢盛江 / 魏静 著: 上博楚简《诗经》献疑, 南開學報(哲學社會科學版) 2006 年 3 期 p.126-133  
 13203 史杰鹏 著: 《孔子诗论》简中的“諄言”和传世文献中相关字词疏证, 古文字研究 26 輯 p.339-342  
 13204 吴辛丑 著: 楚竹书《周易》训诂札记, 古文字研究 26 輯 p.320-324  
 13205 周鳳五 著: 上博五<姑成家父>重編新釋, 臺大中文學報 25 期 p.1-23  
 13206 周鳳五 著: 上博楚竹書《彭祖》重探, 傳統中國研究集刊 1 輯 p.273-281  
 13207 姚小鸥 著: 《孔子诗论》第二十九简与周代社会的礼制与婚俗, 北方論叢 2006 年 1 期 p.1-3

### 3 類目電子化の歴史

#### 3.1 汎用機時代

類目の電子製版は 1981 年度版<sup>1</sup> から行われている。この版から類目の編集は従来のカードベースのものから旧京都大学大型計算機センターの汎用機を用いて行われるようになった。この時、組版システムとして L<sup>A</sup>T<sub>E</sub>X ベースのものが用いられた。この汎用機の文字符号は EBCDIC と JEF であったが、それだけでは到底文字が足りないので、アクセント付きのラテン文字やキリル文字、ハングル、漢字等を独自の外字を用いて表現していた。これをここでは“JEF-CHINA3”と呼ぶことにする。ちなみに、この“CHINA3”というのは当時の類目の検索用データベースの名前である<sup>2</sup>。電子化されたことにより計算機を用いた検索が可能になった訳だが、このことは類目システムを単に組版だけのものではなく、検索サービスも意識して設計・実現・運用しなければならないことを意味している。この汎用機に基づく類目システムの運用は 2000 年度版で終了し、京都大学人文科学研究所附属漢字情報研究センター<sup>3</sup> において後継システムを開発・運用することとなった。その際、この大量に作られた類目外字は徐々にメンテナンス上の問題を引き起こしていき、後にレガシー問題の要因のひとつともなった。

#### 3.2 UTF-8 化以降

2001 年度版から PC や UNIX 系サーバー環境を前提とした新システムに移行することとなったが、ここでは文字符号として UTF-8 を用いることとなった。この結果、以前に比べて、外字の量は著しく少なくなった。しかしながら、それでも、毎年、外字は発生している（表 1）。

表 1. 類目外字の出現数。

対象年	種類	のべ
2001	13	15
2002	12	15
2003	10	14
2004	11	11
2005	15	16
2005 補遺	1	1
2006	41	42

もし、外字が繰り返し使われるならば、その外字を UCS [3] に追加すれば良い訳だが、表 2 に示すように、類目外字の平均利用回数は 1.2 回であり、大部分の外字は 1 回限り

1. 収録対象年が 1981 年のもの。

2. CHINA1, CHINA2 というデータベースも存在した。

3. 現、京都大学人文科学研究所東アジア人文情報学研究センター。

の使用であることが判る。

表 2. 新規外字の再利用度 (「複数回」は複数回利用された文字数、「多年度」は他の年に利用された文字数を表す)。

対象年	新規	複数回	多年度
2001	13	2	1
2002	12	2	2
2003	8	2	0
2004	8	2	2
2005	13	1	0
2005 補遺	1	0	0
2006	38	0	0

こうしたことから、常に新しい外字に対処するための工夫が必要であるといえる。このため、著者が提案する CHISE の文字オントロジー [7, 8] を用いて外字を管理するようにしている。

2001 年度版以降の組版システムとしてはこれまで幾つかの変遷を辿っている。

### 3.2.1 Ω/CHISE

最初、2001 年度版には Ω (Λ) ベースのものをを用いた。ここでは外字処理等のために、苫米地等流氏と宮崎泉氏が開発した “Ω/CHISE” [14] をを用いた。この Ω/CHISE は Ω が提供する ΩTP という仕組みを用いて、『GT 明朝』 [13] や『漢字庫』 [2] 等の大規模外字集合や “Kage” [5, 6, 4] をを用いた IDS 形式 [3] による漢字部品の組合せ表現 (漢字構造情報) からのオンデマンドでのグリフ合成といった機能を提供していた。しかしながら、柔軟である半面、類目のような頁数の多い文書を組版するには重すぎるという問題があった。また、日本語・中国語、ハングル、欧文等が混在する場合のスペースの処理に難があった。

### 3.2.2 pL<sup>A</sup>T<sub>E</sub>X + OTF パッケージの利用

こうしたことから、2002 年度版からは pL<sup>A</sup>T<sub>E</sub>X と OTF パッケージ [11] をベースとしたものをを用いることにした。ここでは “Ω/CHISE” を流用した “CHISE-OTF” パッケージを用い、大規模外字集合や類目外字等を利用できるようにした。pL<sup>A</sup>T<sub>E</sub>X ベースにすることでスペースやスピードの問題は解決し、OTF パッケージをベースとすることにより漢字系の外字の問題は解決したが、ロシア語やベトナム語、タイ語といった各種言語・スクリプトの組版はあまり容易ではない。オリジナルの L<sup>A</sup>T<sub>E</sub>X 用にはさまざまな言語・スクリプト用のパッケージが作られているが、その多くは pL<sup>A</sup>T<sub>E</sub>X ではそのまま利用することはできないからである。

### 3.2.3 up $\text{\LaTeX}$ への移行

この問題を解決するために、2005 年度版からは up $\text{\LaTeX}$  を用いることにした。up $\text{\LaTeX}$  は p $\text{\LaTeX}$  の機能を備えており、OTF パッケージも利用可能である。また、オリジナルの  $\text{\LaTeX}$  用のパッケージもそのまま使うことができる。そこで、CHISE-OTF を up $\text{\LaTeX}$  に移植し、従来からの漢字系の外字をサポートするとともに、オリジナルの  $\text{\LaTeX}$  用のパッケージを用いて非漢字系の各種言語・スクリプトの組版を行うようにした。

以下では、この up $\text{\LaTeX}$  ベースの組版システムについて概説する。

## 4 chise-tex.el

類目の組版は、プログラムによって自動生成された  $\text{\LaTeX}$  ソースに対して、各種言語・用字系用パッケージのためのタグを付加し、up $\text{\LaTeX}$  および dvipdfmx を用いて行っている。このタグを付加する処理は XEmacs CHISE [10] 上において chise-tex.el という Emacs Lisp プログラムが提供する utf-8-jp-tex という coding-system<sup>4</sup> として実現されている。タグとして表現された文字は読み込み時にデコードされるので、利用者はこれらのタグを意識することなく編集することが可能になっている。外字もまたこの中で処理されている。

chise-tex.el が付加するタグには、以下のものがある：

各文字個別のもの 文字毎に  $\text{\TeX}$  のタグに変換する。

アクセント記号関連 合成済み文字 (precomposed character) を基底文字と結合文字 (combining character) のシーケンスに分解して  $\text{\TeX}$  のタグ化するものである。(  $\text{\LaTeX}$  標準のもの<sup>5</sup>、他、苦米地等流氏らによる extaccent.sty<sup>5</sup>, および、福井玲氏作の TIPA [1] を用いている。)

文字符号関連 符号化文字集合<sup>6</sup> に  $\text{\TeX}$  のタグ名を関連付け  $\backslash\text{tagname}\{\text{codepoint}\}$  のように表現するものである<sup>7</sup>。GT や「漢字庫」などのグリフセット、類目外字、タイ文字等で用いている。

用字系指定タグ ギリシア文字やキリル文字のためにフォントエンコーディングを切替えるために必要となったもので、ギリシア文字やキリル文字の文字列をそれぞれ

$\backslash\text{GreekScript}\{\dots\}$  と  $\backslash\text{CyrillicScript}\{\dots\}$

というタグで囲むものである。これは、類目の元データにおいて言語タグが基本的に付与されていないためである。

こうした仕組みにより、エディタ上では多言語化や多字化のために特別な記述を行うことなく編集することができるようになっている。

4. 文字符号を変換するための機構。

5.  $\Omega$ /CHISE の一部として配布されている。

6. CHISE における CCS 素性。

7. SGML の実体参照を  $\text{\TeX}$  風にしたものである。utf-8-jp-tex は utf-8-jp-er に対応する。

## 5 非漢字用の各種パッケージの利用

**TIPA** 発音記号や複雑なアクセント記号が付いたラテン文字等を扱うために、福井玲氏作の“TIPA” [1] を用いている。

**タイ文字** タイ文字を扱うために、NECTEC で配布されていた MikT<sub>E</sub>X 用のパッケージを用いている。このパッケージではタイ文字を TIS 620 で記述する必要があるので、chise-tex.el を用いて、UTF-8 で書かれたタイ文字を \ThaiTIS{XX} (但し、XX は TIS 620 での符号位置) というマクロに一旦変換し、chiseotf.sty を介して利用するようにしている。

## 6 CHISE-OTF の入手先

CHISE-OTF は現在、公開に向けてソースを整理中である。将来的には、“uptex-chise”パッケージにまとめる予定である。現在の状態を anonymous CVS で入手することができる。

**CVS システム経由の入手** 標準的な CVS システムを利用してリポジトリの内容を取得できる。例えば以下のように操作する。

```
### サービスを初めて利用する時に行う ( 2 度目以降は不要 )
# login
% cvs -d :pserver:anonymous@cvs.m17n.org:/cvs/chise login
CVS password:   何も入力しないで Enter
# checkout
% cvs -d :pserver:anonymous@cvs.m17n.org:/cvs/chise co uptex-chise
### 作業ディレクトリの中身を最新のソースにする ( 2 度目以降 )
# update
% cvs -z9 update -Pd
```

**WWW 経由の入手** 上述の CVS リポジトリの内容は <http://cvs.m17n.org/viewcvs/chise/uptex-chise/> から参照することができる。

**フォント関連** CHISE-OTF では Ω/CHISE, および、そこから派生した “chise2otf” モジュール ( Ω/CHISE パッケージに含まれている ) のフォント設定を流用している。Ω/CHISE パッケージの最新版もまた anonymous CVS で入手することができる [14]。

## 7 おわりに

upL<sub>A</sub>T<sub>E</sub>X [12] は pL<sub>A</sub>T<sub>E</sub>X の利点を継承しつつオリジナル T<sub>E</sub>X 用のパッケージも利用できる点で優れており、upL<sub>A</sub>T<sub>E</sub>X と OTF パッケージ、および、chise-tex.el (CHISE-OTF) を用いることで、XEmacs CHISE [10] 上で従来 pL<sub>A</sub>T<sub>E</sub>X では扱えなかった文字も含めて

特に意識することなく見えるままに編集することができるようになり、類目のような多言語・多字種文書の組版が容易になった。特に、これまで扱ってなかった文字体系が必要となった場合、オリジナル  $\text{T}_{\text{E}}\text{X}$  用のパッケージが容易に利用できることは非常に重要なポイントである。

## 参考文献

1. 福井玲, *TIPA*. <http://www.l.u-tokyo.ac.jp/~fkr/computer.htm>
2. 漢字庫. <http://www.sinica.edu.tw/~cdp/zip/hanzi/hanzicd.zip>
3. International Organization for Standardization (ISO), *Information technology — Universal Multiple-Octet Coded Character Set (UCS)*, March 2003, ISO/IEC 10646:2003.
4. Koichi Kamichi, *Kage*. <http://fonts.jp/kage/>
5. 上地宏一, 漢字フォント自動生成サーバ “影 KAGE” の構築 文字コードの枠組みを越える次世代漢字処理の提案, 漢字文献情報処理研究 3 (2002), 143–147.
6. Koichi Kamichi, *KAGE — an automatic glyph generating engine for large character code set*, Proceedings of the Glyph and Typesetting Workshop, 21st Century COE program “East Asian Center for Informatics in Humanities—Toward an Overall Inheritance and Development of Kanji Culture—” Kyoto University, February 2004, pp. 85–92.
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, 人文科学とコンピュータ研究会報告 2006 (2006), no. 112, 25–32, 2006-CH-72.
8. Tomohiko Morioka, *CHISE: Character processing based on character ontology*, Large-scale Knowledge Resources (LKR2008), LNAI, no. 4938, March 2008.
9. ———, 類目外字における “Old Hanzi”, 東洋学へのコンピューター利用第 20 回研究セミナー, 全国文献・情報センター人文社会科学学術セミナーシリーズ, March 2009.
10. Tomohiko Morioka et al., *XEmacs CHISE*. <http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/xemacs/>
11. 齋藤修三郎, Open Type Font 用 VF. <http://psitau.at.infoseek.co.jp/otf.html>
12. 田中琢爾, *upT<sub>E</sub>X*. <http://homepage3.nifty.com/ttk/comp/tex/uptex.html>
13. 東京大学多言語処理研究会, GT 書体フォント 2000. <http://www.l.u-tokyo.ac.jp/GT/>
14. 苔米地等流, 宮崎泉,  $\Omega/\text{CHISE}$ . <http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/omega/>