

# 明治前期の漢字活字と JIS 漢字包摂規準 — 『明六雑誌』 活字字形への、包摂規準適用実験 —

須永哲矢<sup>†1</sup> 堤 智昭<sup>†2</sup> 高田智和<sup>†1</sup>

国内規格 JIS X 0213 に定める「漢字の字体の包摂規準」の、明治前期漢字活字に対する有効性を、当時の雑誌『明六雑誌』の電子化を通じて検証した。『明六雑誌』第 1 号、第 26 号の 2 冊、計 7,442 漢字を対象に調査を行ったところ、JIS X 0213 文字集合のみで処理した場合のカバー率は 86%であった。これに対し包摂規準を適用して処理を行うと、199 設定されている包摂規準のうち 81 が実際に使用され、カバー率も 98%まで向上することが確認できた。

## **Kanji printing types of the early part of the Meiji period and the JIS unification standard ---Experimental application of the unification standard upon printing type forms in “Meiroke Zasshi”---**

TETSUYA SUNAGA<sup>†1</sup> TOMOAKI TSUTSUMI<sup>†2</sup>  
TOMOKAZU TAKADA<sup>†1</sup>

The domestic standard for *kanji* character codes, JIS X 0213, prescribes the “unification standard of *kanji* character forms”, a regulation to be applied to *kanji* variants. The paper examines the effectiveness of the unification standard over the printing types of the early part of the Meiji period, using “*Meiroke Zasshi*” as a sample. In the course of construction of an electronic corpus of “*Meiroke Zasshi*”, we conducted research upon 2 issues (1st and 26th). Among 7,442 *kanji* printing types included in the sample, 86% initially correspond to the character set of JIS X 0213. When the unification standard is applied to the processing, with 81 out of the total 199 detailed regulations being employed, another 12% printing types are newly processed, increasing the total processing rate to 98%.

### 1. はじめに

紙媒体の文書を電子化するには、規格として標準化された符号化文字集合に準拠し、それを運用することが、学術分野・実業分野を問わず、広く行われている。特に言語資料の電子化では、電子化に際してその都度、資料に出現した文字を文字集合のどの符号位置に対応させるべきかという問題（文字包摂の問題、粒度の問題）や、文字集合にない文字をどう扱うかという問題（規格外字の問題、文字セットの規模の問題）が指摘されている。

特に前者の文字包摂の問題に対しては、国内規格 JIS X 0208（第 1 次規格 1978 年）およびそれを拡張する形で開発された JIS X 0213（2000 年）で「漢字の字体の包摂規準」が定められている。



図 1 包摂規準の例（連番 8）

JIS X 0213 では連番で 199 の包摂規準が設定されており、

<sup>†1</sup> 国立国語研究所  
National Institute for Japanese Language and Linguistics  
<sup>†2</sup> 東京農工大  
Tokyo University of Agriculture and Technology

そこに示されている範囲内の差異であれば同一の符号位置の文字として処理することになる。

しかし、この包摂規準の有効性および妥当性に関しては、いまだ十分に検証されたとは言いがたい。図 1 をみてもわかるように、包摂規準は事実上、一般に旧字体と言われるような字体との差異を扱ったものが多い。しかし、そのような「旧字体」も、あくまで現代の活字字形の一つとして想定されたものであり、活字集合そのものが現代のものではない、「古い活字体」に対しても現行の包摂規準が有効であるかの検証は全くなされていない。現代の包摂規準の有効性を問う際に、古い活字に対する適用事例を問題にするというのは一見的外れなようであるが、包摂規準はその性質からして、やや古い活字資料を処理する際により適用の機会を与えられるものである。よって、現代活字を処理しているだけでは見えにくい有効性や限界が、現代以前の活字字形を対象とした検証を通してより明確に浮かび上がってくるのではないかと予想される。

そこで本研究では、あえて明治時代の活字に対して JIS X 0213 包摂規準の適用実験を行い、現行の包摂規準が、近代の活字に対してどの程度有効かを検証した。この適用実験の結果は、包摂規準そのものの有効性および妥当性を問う際の一つの材料となるはずである。

## 2. 『明六雑誌』

本研究で包摂規準の適用実験に用いた『明六雑誌』は、明治7(1874)~8(1875)年の2年間にわたって発行された啓蒙雑誌で、近代日本における総合学術誌、学会誌の先駆けと位置づけられる。体裁は30字×13行の活字本、1号あたり12~24ページで全43号。広範な読者を獲得し、当時の社会への影響が大きかった点、また、記事の内容が幅広い分野にわたり、さまざまな語彙が取り出せる点などから、明治初期の日本語の様相を知るうえで欠かせない資料となっている。

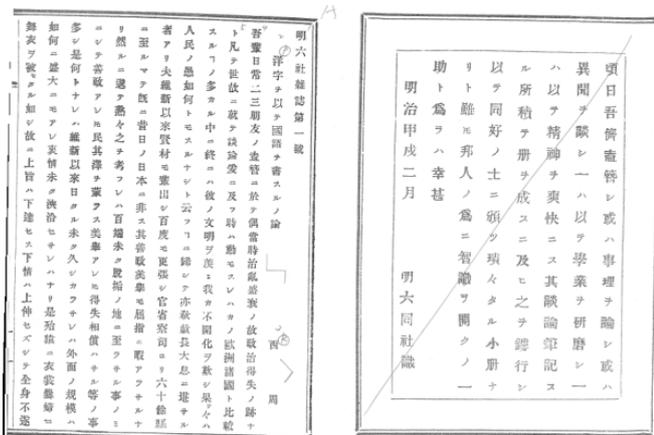


図2 『明六雑誌』

図3のとおり、『明六雑誌』に使用されている活字は現代のものとは異なり、異体漢字も多数出現する。このようにあえてやや厳しい条件下におくことで、JIS X 0213 包摂規準の有効性を見極めようというのである。

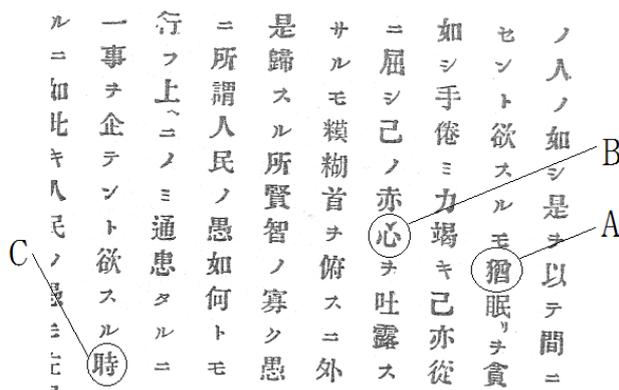


図3 『明六雑誌』の活字

## 3. 包摂規準適用実験

### 3.1 サンプルの選定とデータ作成

『明六雑誌』は全43号、発行期間が2年と短いこともあり、全巻を通じて使用されている活字に大きな変化はみられない。そこで今回は、サンプルとして各年の最初の号である第1号と第26号を選び、この2冊をJIS X 0213文

字集合および包摂規準を適用して電子化した。入力対象総文字数は、原資料汚れ・不鮮明のため判読不可12文字を除いて14,956文字、うち漢字は7,442文字である。

電子化に当たっては、JIS X 0213 文字集合および包摂規準に準拠して入力する[a]が、同時に調査に必要な情報を簡易タグの形で加えていく。図4に検証実験用のテキストデータ例を示す。

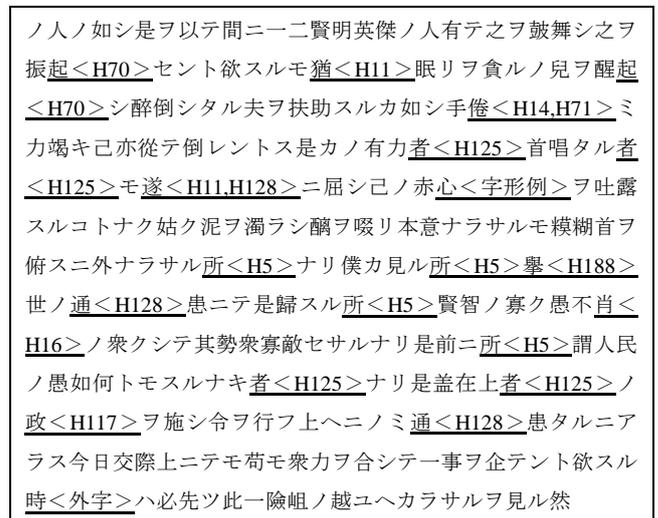


図4 検証実験用テキストデータ

まず、包摂規準の適用によって入力可能となった文字には簡易タグ<H>を付け、適用した包摂規準の連番を記しておく。例えば図3のA「猶」にあたる字は、包摂規準連番11の適用で包摂される。この場合は以下のように入力される。

【入力例】

セント欲スルモ猶<H11>眠リヲ貪

また、JIS規格に則って文字処理をする際の手引となる『JIS漢字字典』[1]には、全体に対する包摂規準とは別個に、各漢字に個別字形例が示されている。図3のB「心」の字は、いちばん上の点のはねており、このような差異に対する包摂規準は設定されていないため、包摂規準のみの適用では外字となる。ただし、『JIS漢字字典』の「心」の項には康熙字形として同様の字形が収録されており、これを参照すれば包摂することが可能となる。このような場合には、別個に「字形例」という情報を記録する。

【入力例】

ニ屈シ己ノ赤心<字形例>ヲ吐露ス

JIS X 0213規格外となるものは「=」を入力するが、通用字のどの字に当たるか明らかな場合は、「=」にせず、通

a JIS X0213 準拠としたが、厳密には包摂規準の適用対象外となる康熙別掲字は使用せず、康熙別掲字に対しても包摂規準通りの字体包摂を行う。

用字を入力したうえで、外字であることを記入する。例えば図3のC、「時」は、左側の「日」の形に差異がみられるが、このような差異については包摂規準もなく、『JIS 漢字字典』の字形例にも見られない。このような場合は、「時」を入力したうえで「外字」という情報を記録する。

【入力例】

一事ヲ企テント欲スル時<外字>

3.2 計測結果 - 『明六雑誌』漢字カバー率 -

以上の形式で作成したテキストデータをもとに、『明六雑誌』1号、26号の漢字7,442字に対し、包摂規準を適用せずJIS X 0213 文字集合のみで表現できた漢字、包摂規準の適用で表現可能となった漢字を計測、カバー率を算出した。

表1のとおり、JIS X 0213 文字集合のみの漢字カバー率は85.96%、1,045字が外字となる。これに対し包摂規準を適用して漢字処理を行うと、外字10,45字中909字までが処理可能となり、カバー率も98.17%にまで向上する。

なお、これに加えて『JIS 漢字字典』の字形例を参照した場合、さらに29字が処理可能となり、カバー率は98.56%となる。

表1 JIS X0213 文字集合と『明六雑誌』の漢字 (のべ)

	X0213のみ	包摂規準適用	『JIS 漢字字典』 字形例参照
処理可能 文字数	6397	7306	7335
	第1水準: 5632	第1水準: 6513	第1水準: 6542
	第2水準: 764	第2水準: 792	第2水準: 792
	第3水準: 0	第3水準: 0	第3水準: 0
	第4水準: 1	第4水準: 1	第4水準: 1
新たに処理 できる文字 総数	-	909	29
		第1水準: 881	第1水準: 29
		第2水準: 28	第2水準: 0
		第3水準: 0	第3水準: 0
		第4水準: 0	第4水準: 0
外字総数	1045	136	107
カバー率	85.96%	98.17%	98.56%

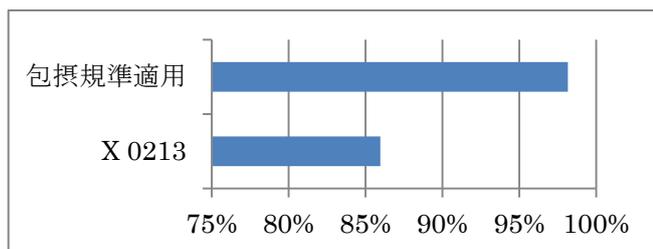


図5 カバー率の比較

なお、『明六雑誌』1号、26号の異なり字数は1,607字、うち包摂規準の適用を受けたのはその1割以上に及ぶ170字であった。

表2 JIS X0213 文字集合と『明六雑誌』の漢字 (異なり)

	X0213のみ で処理可能	包摂規準適用 で新たに 処理可能	『JIS 漢字字典』 字形例 参照で新た に処理可能	最終 的な 外 字	計
	1395	170	8	34	1607
異 な り 字 数	第1水準: 1129	第1水準: 152	第1水準: 5		
	第2水準: 265	第2水準: 18	第2水準: 3		
	第3水準: 0	第3水準: 0	第3水準: 0		
	第4水準: 1	第4水準: 0	第4水準: 0		

4. 『明六雑誌』に適用された包摂規準と、対象文字

適用された包摂規準を連番ごとに見ると、全199の包摂規準のうち、81が実際に『明六雑誌』の文字処理に使用されたことがわかる。包摂規準が適用されたのべ909字は、異なりにして178字、表3にその一覧を示す[b]。

また、のべ30字以上または異なり5字以上に適用された、使用頻度の高い包摂規準を図6に示す。

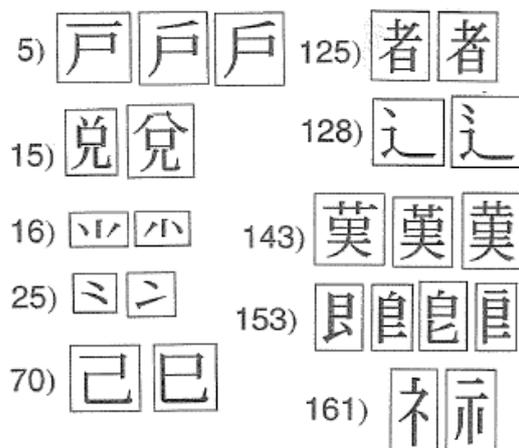


図6 使用頻度の高い包摂規準  
 (のべ30字以上または異なり5字以上に適用)

b なお、一つの文字に対し複数の包摂規準を適用する場合がありますため、表3ののべ字数、異なり字数総計は表1、表2の字数を超える。

表 3 『明六雑誌』に適用された包摂規準と、対象文字

包摂規準連番	適用文字数(のべ)	適用文字数 (異なり)	適用文字							
			微	程	聖					
1	8	3	微	程	聖					
2	2	2	耘	籍						
3	26	1	害							
4	0									
5	64	2	所	絜						
6, 7	0									
8	4	1	教							
9	0									
10	1	1	歳							
11	24	4	益	猶	兼	遵				
12	3	1	判							
13	9	1	平							
14	10	4	騰	藤	勝	倦				
15	38	5	税	送	説	脱	悦			
16	33	6	幣	尚	弊	消	肖	蔽		
17	3	1	率							
18	13	4	習	摺	翰	翼				
19	0									
20	1	1	弱							
21	7	3	暖	曖	採					
22	3	1	判							
23	0									
24	7	3	故	固	姑					
25	22	3	於	終	寒					
26	1	1	空							
27, 28	0									
29	28	1	術	述						
30~32	0									
33	18	3	次	恣	資					
34, 35	0									
36	3	1	望							
37	0									
38	5	1	恐							
39~41	0									
42	12	4	造	慥	告	鵠				
43	1	1	唐							
44~47	0									
48	1	1	吳							
49	3	1	捨							
50	0									

51	7	2	亡	望						
52	5	1	那							
53	0									
54	11	2	急	婦						
55~68	0									
69	3	2	抗	冗						
70	21	6	杞	起	改	港	忌	撰		
71	2	1	倦							
72	5	1	産							
73~76	0									
77	7	1	化							
78	1	1	叱							
79	18	1	全							
80	18	2	内	納						
81	3	2	免	危						
82	3	3	若	勸	模					
83	2	1	茲							
84~87	0									
88	9	1	要							
89	0									
90	1	1	堅							
91~96	0									
97	2	1	并							
98	0									
99	1	1	墨							
100	0									
101	5	2	增	層						
102	0									
103	1	1	裨							
104,105	0									
106	3	3	格	格	降					
107~111	0									
112	2	1	專							
113	0									
114	9	1	及							
115	0									
116	1	1	延							
117	16	2	踪	政						
118~124	0									
125	103	5	著	者	諸	緒	都			
126,127	0									
128	116	28	送	述	造	慥	適	過(ほか)		
129	1	1	琢							
130	4	1	德							

131	3	1	徴					
132	2	1	徴					
133~ 135	0							
136	5	1	歩					
137~ 138	0							
139	4	1	強					
140	1	1	歴					
141	5	2	廣	横				
142	2	1	僅					
143	50	3	難	漢	艱			
144,145	0							
146	15	3	情	請	精			
147~ 150	0							
151	1	1	頼					
152	0							
153	31	5	概	嚮	節	既	卿	
154	0							
155	1	1	飾					
156	1	1	像					
157~ 159	0							
160	1	1	獵					
161	45	6	視	社	神	福	祉	祈
162	0							
163	10	1	旅					
164	2	2	蔗	庶				
165	0							
166	1	1	虚					
167	2	1	録					
168,169	0							
170	1	1	忍					
171~ 177	0							
178	16	3	輸	愈	愉			
179~ 187	0							
188	7	1	學					
189	4	1	船					
190~ 196	0							
197	1	1	庶					
198~ 199	0							

## 5. 『明六雑誌』にみられる JIS 外字字形

『明六雑誌』1号および26号中、JIS X0213 規格外字はのべ107字、異なりにして34字であった。今回の作業で外字認定された異なり34字の全例を以下に示す(図7~9)。

外字とされたものの大部分は、JIS X0213 内字のいずれかに対し僅かな字形差があるものであった。なお、34字中10字はUnicode[c]で表現可能な差異である。本調査では包摂規準をなるべく厳密に適用し、包摂してよいという明確な根拠が見いだせない場合は外字とすることとした。そのため、本調査で外字と認定された文字の中には『常用漢字表』の「デザイン差」等の解釈によっては外字処理に回さずに済む可能性があるものも含まれる。また、既存の包摂規準の中には適用可能なものはないが、類例が見いだせるというものも多い(図8)。全体を見渡しても34字中33字はJIS X0213 内字に対する字形差とみなすことができる文字であり、字形差レベルにとどまらない規格外字は1文字のみである(図9)。

図7 『明六雑誌』JIS 外字・その1 (10字)  
 (Unicode では表現可能な差異)



※左から、JIS X0213, Unicode 別字, 『明六雑誌』実字形

c ここでは Unicode4.0 を参照している。

図 8 『明六雑誌』 JIS 外字・その 3 (23 字)

藏藏藏	訴訴訴
善善善	時時時
万万万	博博博
良良良	斗斗斗
解解解	候候候
會會會	胸胸胸
華華華	嘩嘩嘩
序序序	号号号
覽覽覽	寧寧寧

續續續	参考包摂規準 27) 壘壘 28) 楞楞
-----	-------------------------

記記記	配配配
改改改	参考包摂規準 67) 𠄎𠄎 70) 𠄎𠄎 71) 𠄎𠄎𠄎
犯犯犯	

※左から、JIS X0213、『明六雑誌』字形、『明六雑誌』実字形

図 9 『明六雑誌』 JIS 外字・その 3 (1 字)

璿

※Unicodeに「璿」あり。亦 ←→ 余

## 6. おわりに

以上、『明六雑誌』の漢字活字に JIS X0213 包摂規準の適用を試みたところ、包摂規準なしで 86%であった漢字カバー率が包摂規準の適用で 98%に向上することが明らかになった。また、包摂規準の適用を受けた漢字は全体の 1 割以上にのぼり、199 の包摂規準のうち実際に 81 が使用されたことで、JIS X0213 包摂規準の有効性はある程度確認できたと言えよう。しかし一方で、包摂できなかった字形をみると、包摂規準を設定して処理すべきレベルの差異と考えられるものも多数見受けられ、近代活字用に包摂規準を拡張するという方向性も考えられる[2]。近代活字での検証は始まったばかりであり、今後『明六雑誌』以外の活字資料も視野に入れ、さらなる検証をしていきたい。

### 参考文献

- 『JIS 漢字字典』増補改訂版，日本規格協会（2002）
- 須永哲矢，堤智昭，高田智和：明治前期雑誌の異体漢字と文字コードー『明六雑誌』を事例としてー，人文科学とコンピュータシンポジウム論文集 2011，pp.381-388(2011)