

漢字の分解と構成要素の計量的分析に基づいた学習漢字の最適な掲出順序の開発

ヴォロビヨワ・ガリーナ

キルギス日本語教師会、キルギス共和国

キーワード： 漢字の分解、書記素、構成要素、コード化、掲出順序

要旨

漢字学習の効果をより増大させるには、教科書等における漢字の掲出順序を最適化することがきわめて重要であろう。学習漢字の最適な掲出順序を明らかにするために、漢字の複雑さに関する要因に着目して、下記の5点を研究した。

- (1) 漢字の構造分解と書記素の分析。
- (2) 漢字の構造分解と構成要素の分析。
- (3) 漢字の複雑さの定義。
- (4) 漢字教材の計量的分析。
- (5) 最適な漢字の掲出順序の開発。

1. はじめに 日本語教師の経験から見ると非漢字系学習者にとって日本語の勉強では漢字の勉強こそ困難を引き起こすことである。困難な要素は次のようである

- (1) 漢字を非体系的に感じる。
- (2) 学習漢字の数が多。
- (3) 字形、字義、読み方の結び方が分からない。
- (4) 同時に字形、字義、読み方、筆順、単語を覚えなければならない。
- (5) 一時的に漢字を覚えても定着できない。
- (6) 漢字字典の調べ方は難しい。

漢字の勉強の道を踏んだ学習者の負担を軽くするのは教師の義務であるが、一人一人工夫しても漢字教育の問題が解決できない。非漢字系学習者を対象とした全体的な漢字教育法の案があれば教師も、学習者も楽になるであろう。本研究では主の一つの問題について検討する。それは学習漢字の掲出順序である。漢字教育法には漢字を覚えやすくするために、連想記憶法を使う教え方がある。連想記憶法は漢字の構造分解と一体不可分である。連想記憶法において、漢字の成り立ちのストーリーを効果的に使うには、先に構成要素、簡単な漢字、後で合体文字を教えるほうが望ましいと考えた。

2. 研究方法 漢字の諸性質のうち、計量可能な要素を検討した。計量的なアプローチとコンピューター技術に基づいて漢字の計量可能な要素を分析した。研究の段階は次のとおりである。

- (1) 漢字の構造分解をし、書記素の種類、それから部首以外の構成要素の種類を確定した。
- (2) 書記素と構成要素をコード化した。それから漢字のアルファベット・コード、シンボル・コードとセマンチック・コードの3つのコードのシステムを構築した。
- (3) 常用漢字 1945 字、学習漢字 1006 字、一般に使われている漢字教材の漢字コー

ドのデータベースを構築した。

- (4) そのデータベースをもとに書記素と構成要素の使用頻度、漢字教材の漢字の掲出順序を計量的に分析した。そのためには統計解析ソフト SPSS とせっかく作ってもらった特別なソフトを利用した。
- (5) 漢字の複雑さと使用頻度に基づき、漢字の最適な掲出順序の判定基準を検討した。

3. 漢字の分解と書記素の分析

3.1. 書記素の種類とコード化。漢字のアルファベット・コードとカタカナ・コード

漢字の一番細かい部分は書記素（つまり画、ストローク）である。書記素の種類は16から40までのいくつかのシステムが存在している。Fazzioli(1987:20)によると、漢字は、24種類の書記素で書くことができる。それぞれの書記素に、ローマ字の形と結びつけ、AからZと名前、つまりアルファベット・コードをつけた（表1）。

表1. 24種類の漢字の書記素とそのアルファベット・コード

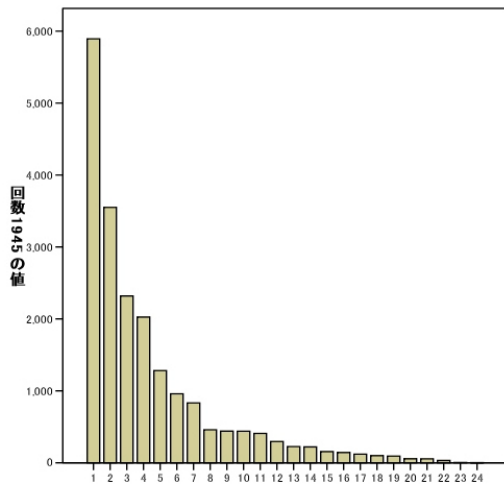
A	B	C	D	E	F
一	丨	丿	㇇	㇈	㇉
G	H	J	K	L	M
㇊	㇋	㇌	㇍	㇎	㇏
N	O	P	Q	R	S
㇑	㇒	㇓	㇔	㇕	㇖
T	U	V	W	Y	Z
㇗	㇘	㇙	㇚	㇛	㇜

漢字のそれぞれの書記素をローマ字のコードに変換し、筆順をローマ字の列で表した。その列を漢字のアルファベット・コードにした。同じようにそれぞれの書記素に、カタカナの形と結びつけ、アからンと名前（カタカナ・コード）をつけた。漢字の書記素の形からアルファベットとカタカナの形が思い浮かべられるようにした。

例 三 (AAA) (ナナナ) 川 (PBB) (ノトト) 玉 (ABAAQ) (ナトナナソ)
女 (KPA) (クノナ) 小 (JLQ) (オハソ) 也 (GBC) (ストシ)

3.2. 書記素の分析 1945字の常用漢字、1006字の学習漢字、一般に使用されている漢字教材のアルファベット・コードのデータベースを構築した。それらをもとに常用漢字などの書記素の使用頻度を計量的に分析した（図1、表2）。書記素の使用頻度のヒストグラムを描いた結果、常用漢字、学習漢字1006字などのそれぞれの群の中の書記素の割合はほとんど同じであることが分かった。

図1. 1945字の常用漢字。使用頻度の降順で並べた24個の書記素の使用頻度（回数）



X軸は1～24の番号、Y軸は回数である。

表2. 1945字の常用漢字。使用頻度の降順で並べた24個の書記素

書記素	書記素のコード		使用回数
	アルファベット	カタカナ	
一	A	ナ	5893
丨	B	ト	3550
ノ	P	ノ	2319
丶	Q	ソ	2026
冫	H	ロ	1282
丶	L	ハ	959
丶	O	キ	832
丶	V	ン	459
冫	G	ス	440
丶	S	ウ	439
丨	J	オ	408

フ	Y	フ	296
一	F	ア	225
丨	C	シ	221
ノ	U	ム	155
ノ	K	ク	143
丨	D	イ	120
丨	E	エ	99
ノ	T	レ	93
フ	M	リ	56
丨	R	モ	55
ノ	N	ヌ	32
フ	Z	ル	4
乙	W	ヒ	2

4. 漢字の分解と構成要素の分析 漢字の指導では構造分解と構成要素の分析にもとづく教え方がある。例えば、山田ボヒネック頼子(2007)では漢字の280の基本的な構成要素(原子、つまり最少意味単位)を習得してから、体系的増分式で1945字種の常用漢字を勉強する。学習者が自分で未知の漢字の「構成要素分析」をし、「全体総合」で意味取りができるという教え方が記述されている。

4.1. 構成要素の定義とコードのシステム 本研究は漢字の構造分解をするために、漢字の構成要素である部首に加え、部首ではないパターン(グラフィウム)も定義した。それから非漢字系学習者にとってなじみのアルファベットと数字を使って漢字の構成要素をコード化した。部首とグラフィウムのコードをもとに漢字のコード化ができた。漢字コードのシステムを二つ考えた。シンボル・コード(アルファベットと数字のコード)とセマンチック・コード(単語で表す意味的なコード)のシステムである。筆者が執筆した教科書『漢字物語Ⅰ、Ⅱ』に入っている518字の漢字ではグラフィウムが93種類(ヴォロビヨフ&ヴォロビヨワ 2007:176)と部首が140種類使われている。全部で基本的な構成要素が233種類使われている。

4.2. 構成要素のコード化 伝統的な部首のシステムにはそれぞれ1から214までの番号がついている。部首の番号をそのシンボル・コードにする。部首のセマンチック・コード(意味を表す単語)としてその主な意味を表す単語を利用することにした。

グラフィウムのシンボル・コードはアルファベットと数字を含めたコードにする。詳しい説明はヴォロビヨワ(2007:22)にある。グラフィウムのセマンチック・コード(意味的なコード)を主にHeisig(2001)とHensha11(1998)によってつけた。

表3. グラフィウムのコード化の例

グラフィウム	アルファベット・コード	シンボル・コード	セマンチック・コード	画数	漢字の例
乍	PABAA	5PAB	make	5	作
世	ABBAE	5ABB2	generation	5	葉

4.3. 漢字のコード 漢字のコード化の目的は次のようである。

- (1) コードは学習者に漢字の構成を説明するとき手助けとなる。
- (2) 漢字のコードのデータベースを構築し、様々な計量的分析ができる。
- (3) データベースの計量的分析に基づき、学習漢字の最適な掲出順序の開発ができる。
- (4) 漢字が簡単に検索できるためにコードに基づき、新しいタイプのアルファベット・コード索引とシンボル・コード索引とセマンチック・コード索引を作る。

漢字のシンボル・コードを作るためには、筆順に従ってその漢字の構成要素のコードを書いた。セマンチック・コードを作るためには、筆順に従って最初の二つの構成要素の主な意味を表す単語を書いた。

表4. 漢字のアルファベット・コードとシンボル・コードとセマンチック・コードの例

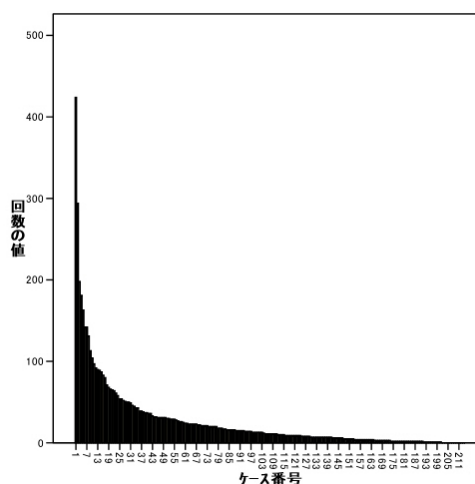
漢字	アルファベット・コード	シンボル・コード	セマンチック・コード
七	AC	1/5	one/hook
九	PR	PR	nine
気	PAARPO	84/P0	steam/sheaf

4.4. 構成要素の分析 常用漢字1945字、学習漢字1006字、一般に使われている漢字教材のシンボル・コードのデータベースをもとに漢字の構成要素の使用頻度の計量的分析をした(表5、図2)。図2のX軸は1~214の番号、Y軸は回数である。

表5. 1945常用漢字の214の部首の使用頻度
(使用頻度の降順で並べた回数)
(表の上の部分)

部首	部首番号	使用頻度 (回数)
一	1	424
口	30	294
丿	4	198
人; 亻; 亼	9	181
日	72	163
八; ㄨ	12	142
木	75	142
水; 氵; 氷	85	131

図2. 1945常用漢字の214の部首の使用頻度
(使用頻度の降順で並べた回数)



5. 漢字の複雑さの定義 NTTデータベース(近藤公久・天野成昭 1999)の「文字の特性データベース」(6,847字の漢字)から1945字の常用漢字を取り出し、「構成要素数」と「使用順位」を加え、計量的に分析した。統計解析ソフトSPSSを利用し、漢字の書記素数(画数)、構成要素数、NTTデータベースによる親密度、複雑度、ドット数24、ドット数32という要素の相関係数を求め、漢字の複雑さの定義を考察した。相関係数を分析した結果、二つの要素、構成要素数と書記素数で漢字の複雑さを定義した。

表6. 計量可能な漢字の要素のデータベース (表の上の部分)

漢字	親密度	複雑度	画数	ドット数24	ドット数32	構成要素数	使用順位
一	6.96	1.04	1	28	37	1	2
右	6.75	2.58	5	126	181	2	601
雨	6.67	3.29	8	189	246	1	796
円	6.67	2.75	4	153	181	3	48
王	6.58	2.62	4	108	137	1	739

6. 漢字教材の計量的分析 一般に使用されている漢字教材の漢字の掲出順序を分析した結果、アプローチが色々あると明らかになった。例えば、徳弘康代(2008) に含めた2100字の漢字は「よく使う順で」並べてある。しかし、複雑な漢字が簡単な漢字より早く出ること、また合体文字がその構成要素である漢字より早く出るとは漢字教科書の問題点になっている。21冊の漢字教材の17472字(延べ数)の漢字のシンボル・コードのデータベースを構築し、それぞれの教科書の漢字の掲出順序を計量的に解析した。分析するためには、特別なコンピューターのソフトを作ってもらった。分析した結果、21冊の漢字教材の中には合体文字がその構成要素である漢字より早く出ることが42%に達することが明らかになった(表7)。そのような順序には「逆の順序」という名前をつけた。

表7. 掲出順序の分析の例(「逆の順序」のケースの%の降順)

名前	漢字数	逆の順序のケース	
		数	%
Kanji pictographix (Rowley 1992)	1225	522	42.6
日本語学習のためのよく使う順 漢字2100(徳弘康代2008)	2100	751	35.8
BASIC KANJI BOOK Vol. 1, 2 (Kano 2003) Intermediate KANJI BOOK Vol. 1, 2 (Kano 2003)	1098	231	21.0
Kanji in Context (西口光一 2004)	1947	374	19.2
1006学習漢字 (赤尾文夫 2002)	1006	189	18.8
みんなの日本語初級 I、II 漢字練習長 (西口光一 2000)	530	87	16.4
Kanji Book (Naganuma 1951)	1689	262	13.0
Remembering the Kanji (Heisig 2001)	2042	72	3.5
ストーリーで覚える漢字300 (ボイクマン総子2008)	300	6	2.0

「逆の順序」の具体的な例として日本の小学校で教えている「木」という漢字の位置の説明をする(赤尾文夫 2002)(表8)。

表8. 学習漢字1006字の提出順序の分析の例

構成要素である漢字	掲出順序	合体文字	掲出順序	順序の差
木	74	休	13	61
		校	23	51
		森	40	34
		村	54	20
		本	72	2

「木」という漢字を教える前にその漢字を構成要素として含めた合体文字「休」、「校」、

「森」、「村」、「本」を教えるのが、もし先に「木」を教えたら合体文字の意味がもっと説明しやすくなると思う。

7. 学習漢字の最適な掲出順序 漢字の複雑さと使用頻度に基づき、漢字の最適な掲出順序の判定基準を検討した。漢字の書記素数、構成要素数、新聞での使用順位の属性情報を用いて漢字をソートし、その複雑度表を作った(表9)。それをもとに勉強の目標に相応しい漢字の最適な掲出順序が検討できる。例えば、表10は1006学習漢字の「木」「休」「校」「森」「村」「本」(表8)の漢字の最適な掲出順序の案を示している。

表9. 常用漢字の複雑度表(上の部分)

漢字	画数	構成要素の数	使用順位
一	1	1	2
乙	1	1	1955
十	2	1	3
二	2	1	4
人	2	1	5
八	2	1	39

表10. 表8の漢字の最適な掲出順序の案

漢字	画数	構成要素の数	使用順位
木	4	1	189
本	5 (+1)	2 (+1)	11
村	6 (+2)	2 (+1)	170
休	6 (+2)	2 (+1)	334
校	10 (+6)	3 (+2)	151
森	12 (+8)	3 (+2)	569

8. 漢字教材作成 国際交流基金日本語国際センターのプログラムに参加させていただいたおかげで、連想記憶法に基づき、漢字の成り立ちのストーリーを中心に漢字を518字含めた初級の漢字教科書『漢字物語Ⅰ、Ⅱ』(ヴォロビヨワ 2007、ヴォロビヨフ&ヴォロビヨワ 2007)を日本語とロシア語で執筆した。これは『みんなの日本語初級Ⅰ、Ⅱ漢字』の漢字掲出順序と同じ順番で学習を進めるように工夫されている。合体文字がその構成要素である漢字より早く出るケースが16.4に達する。本書を独立した教材として使う場合は学習する順番を自分で決めることができるように、この教科書に漢字の複雑度表を付けた。それは三つの要素、構成要素数、書記素数、新聞での使用順位(横山詔一・笹原宏之・野崎浩成・エリク=ロング 1998)を用いて漢字をソートして作った表である。

また学習者が漢字の検索のときなじみあるアルファベットと数字のコードで漢字を調べるためには『漢字物語Ⅰ、Ⅱ』のシンボル・コードとセマンチック・コードのデータベースを構築し、それからコードの順番に並べ替え、新しいタイプの漢字の「シンボル・コード索引」と「セマンチック・コード索引」をつけた。

9. まとめと今後の課題 本研究のテーマになった漢字の構造分解、書記素と構成要素の使用頻度、学習漢字の最適な掲出順序の開発は総合的な漢字教育改善の側面である。非漢字系日本語学習者を対象とした漢字教育法の案に下記の功を含める必要があると思っている。

- (1) 漢字の階層性の分類、書記素と構成要素の規格化と応用。
- (2) 学習用漢字の最適な量や提出順序。

- (3) 漢字の字義、読み方、書き方、熟語の教え方。
- (4) 字義を記憶する方法。
 - (4.1) 漢字の歴史や成り立ち。
 - (4.2) 連想記憶法。
 - (4.3) 長期にわたる学習効果の持続法。
- (5) 漢字字典の調べ方の合理化。新しいタイプの索引の開発。

今後本研究をもとに最適な漢字の掲出順序に改訂して約 1000 字の漢字を含めた漢字教科書を執筆する予定である。

参考文献：

1. 赤尾文夫 (2002) 『1006 字の正しい書き方』 旺文社.
2. ヴォロビヨワ・ガリーナ (2007) 『漢字物語 I』 ビシケク.
3. ヴォロビヨフ・ヴィクトル、ヴォロビヨワ・ガリーナ (2007) 『漢字物語 II』 ビシケク.
4. 近藤公久、天野成昭 (1999) 『NTTデータベースシリーズ『日本語の語彙特性』 第 5 巻 文字特性』 三省堂.
5. 徳弘康代 (2008) 『日本語学習のためのよく使う順 漢字2100』 三省堂.
6. 西口光一、新矢麻紀子 (2000) 『みんなの日本語初級 I 漢字』 スリーエーネットワーク.
7. 西口光一、新矢麻紀子 (2001) 『みんなの日本語初級 II 漢字』 スリーエーネットワーク.
8. 西口光一、河野玉姫 (2004) 『Kanji in Context』 The Japan Times.
9. ボイクマン総子、渡辺陽子、倉持和菜 (2008) 『ストーリーで覚える漢字300』 ころしお.
10. 山田ボヒネック頼子 (2007) 「KK2.0 (KanjiKreativ) Eラーニング:1945常用漢文字学習プログラム」—体系的・増分式「識字力育成」が日本語教育に齎すインパクト」、『ヨーロッパ日本語教育12報告・発表論文集』 169-175頁.
11. 横山詔一、笹原宏之、野崎浩成、エリク・ロング (1998) 『新聞電子メディアの漢字』 三省堂.
12. Fazzioli Edoardo (1987) *Chinese calligraphy*. New York –London: Abbervillpress-Publishers.
13. Heisig James W. (2001) *Remembering the Kanji*. Japan Publications Trading Co. Ltd.
14. Henshall Kenneth G. (1988) *A Guide to remembering Japanese characters*. Tuttle Publishing.
15. Kano Chieko et al. (2003) *BASIC KANJI BOOK* Vol. 1, 2. Bonjinsha.
16. Kano Chieko et al. (2003) *Intermediate KANJI BOOK* Vol. 1, 2. Bonjinsha.
17. Naganuma Naoe (1951) *Kanji Book*. 長風社.
18. Rowley Michael (1992) *Kanji pictographix*. Berkeley, California: Stone Bridge Press.