

国立国語研究所学術情報リポジトリ

A Method of the Analysis of Kanji Structure: A New Approach Based on Structural Decomposition and Coding

メタデータ	言語: jpn 出版者: 公開日: 2015-10-30 キーワード (Ja): キーワード (En): 作成者: ヴォロビヨワ, ガリーナ, ヴォロビヨフ, ヴィクトル, VOROBÉVA, Galina, VOROBÉV, Victor メールアドレス: 所属:
URL	https://doi.org/10.15084/00000469

漢字の構造分析に関わる問題

——漢字字体の構造分解とコード化に基づく計量的分析——

ヴォロビヨワ ガリーナ^a ヴォロビヨフ ヴィクトル^b

^a国立国語研究所 共同研究員

^bキルギス国立総合大学

要旨

本論文では、漢字の構造分析に関わる先行研究を概観し、いくつかの問題点を指摘した。それらの問題点を解決する手がかりを得るため、新たなアプローチに基づく漢字の構造分析の方法を提唱し、以下の5つの側面から検討をおこなった。

- (1) 常用漢字をカバーする「スタンダード化された構成要素のシステム」を開発する必要性を指摘した。その開発に成功すれば、漢字の認識や漢字習得の体系化が可能になる。
- (2) 構成要素のシステムについて先行研究における部首の扱い方を分析した結果、従来の部首を採用していないシステムが多いことが明らかになった。
- (3) 漢字字体の構造を明確に表現・表示できれば、漢字字体の計量的分析が可能になる。そこで、漢字の線型構造分解をおこなった上で漢字字体に関する独特のコード化を開発し、アルファベット・コードとシンボル・コードのシステムを構築した。漢字コード化の結果に基づいて漢字の画と構成要素の出現頻度の測定が可能になり、漢字の構成上の複雑さを定義する新たな指標を提案する可能性が開けた。
- (4) 英語の文や語構成における統語的階層構造と、漢字の階層構造が一部において類似していることを示した。漢字の階層構造分解をおこない、樹形図と数学的公式で漢字の階層構造を表示した。さらに藤村(1973)による漢字の階層構造の分析とコード化と、本研究による階層構造の分析とコード化を比較し、本研究の実用性を示唆した。
- (5) 漢字の構成上の複雑さを判定する新たな指標を定義し、複雑さによる常用漢字の分類をおこなった*。

キーワード：漢字の構造分析、常用漢字、漢字コード、線型構造分解、階層構造分解

1. はじめに

漢字は日本語学習において重要な学習項目である。漢字教育では日本語学習者を漢字系学習者と非漢字系学習者に区別し、それぞれに相応しい指導法の必要性が認められている。非漢字系日本語学習者にとって漢字学習が最大の困難であることは周知の事実である。これは表語文字（表意文字）である漢字が、非漢字系の学習者に馴染みのある表音文字であるアルファベットなどと大きく異なるからである。

以下、非漢字系日本語学習者、特に初期学習者の漢字学習の主な問題点を、先行研究の知見・

* 本稿の一部は国立国語研究所共同研究プロジェクト「文字環境のモデル化と社会言語科学への応用」（文字表記研究班）（プロジェクトリーダー：横山詔一）の研究成果である。国立国語研究所理論・構造研究系の横山詔一教授は2007年から我々の研究をご指導・ご支援してくださっている。津田塾大学非常勤講師の関麻由美先生は日本語の校閲をしてくださった。そして大阪大学の大学院生のDaniel Kobayashi-Better氏は要旨の英語訳をしてくださった。皆様に厚く御礼を申し上げる。

指摘に基づいて、次のような観点に立って整理する（ヴォロビヨワ 2014: 13）。

漢字そのものに内在する問題

- (1) 学習すべき漢字の数が多
- (2) 学習すべき漢字語彙が多
- (3) 漢字の字体が複雑である
- (4) 漢字を構成する要素が多
- (5) 個々の漢字に関わる情報が多い（字体, 字義, 読み方, 筆順, 部首など）
- (6) 字体, 字義, 読み方の間の関連性が明白でない
- (7) 音声情報が単一ではなく, 音訓読みも複数存在している

教授法の問題

- (8) 丸暗記に頼る非体系的な指導法が使用されることがある
- (9) 学習対象漢字の掲出順序が合理的ではないことがある
- (10) 漢字学習の時間的な制約がある
- (11) ICT（コンピュータやインターネット）の使用が不十分である

学習者の漢字学習に対する意識の問題

- (12) 漢字を非体系的に感じる
- (13) 自律学習方法がよく分からないことがある
- (14) 多様な漢字学習ストラテジーについて知らないことがある
- (15) 一時的に漢字を覚えても定着しない
- (16) 漢字辞典の調べ方を難しく感じる

日本語学習者は漢字学習における悩みが多岐に渡っており、学習を支援するためには様々な対策が必要となっていることが分かる。

2. 漢字の構造分解と構成要素

2.1 漢字の構成要素

漢字の構造分析をおこなうためにまず漢字の構成要素を確定する必要がある。漢字の構成要素というのは漢字の最小意味的単位である。例えば、「明」という漢字の構成要素は「日」と「月」である。構成要素数によって漢字は2つのグループに分けられる。1個の構成要素のみから成る漢字は「簡単な漢字, 単純な漢字, 単体漢字, 元素漢字」（例えば, 「人」, 「手」, 「鳥」）であり, 2個以上の構成要素から成る漢字は「複雑な漢字, 合体漢字, 合成漢字, 組み合わせ文字, 結合文字」（例えば, 「町」（田+丁）, 「字」（宀+于）, 「森」（木+木+木））である。

合体漢字は2種類の要素から成り立っている。それは単体漢字と漢字にならない構成要素である。例えば, 「右」=「ナ」+「口」である。ここで「口」は部首のリストに入っている単体漢字で, 「ナ」は現在漢字としては使われておらず, 「漢字の部品」となっている。武部（1984: 72）は, 漢字が, ある単位の組み合わせであるという考え方を導入すると「複雑な漢字が認識しやすくなる」, 「秩

序が生まれ見やすくなる」と記している。

2.2 先行研究による漢字の構成要素のシステムの比較分析

ここでは、漢字の構成要素（部首あるいはそれに相当する最小意味的単位）に関わる先行研究を概観する。漢字の構造分解と構成要素について検討した研究には、Wieger (1915/1965), Heisig (1977/2001), 白石 (1978), Hadamitzky & Spahn (1981), ハルペン (1987), Stalph (1989), Habein & Mathias (1991), Foerster & Tamura (1994), Demirci (1997), 鹿島 (2006), 齋藤 (2006), 山田ボヒネック (2008), ヴォロビヨワ (2011, 2014) などがある。これらは、いずれも漢字を構成要素に分解しているが、1945字の旧常用漢字¹などの漢字群をカバーするために扱う構造分解の方法と構成要素はそれぞれ異なっていた。

研究者によって漢字の構成要素の名称は異なり、「構成要素」, 「構成素」, 「primary element (根本要素)」, 「basic form (基本的な要素)」, 「字形要素」, 「形態要素」, 「字体素」, 「字体部分」, 「部品」, 「部品漢字」, 「元素」, 「原子」, 「要素文字」, 「パーツ」, 「パターン」, 「unit」, 「component」, 「radical」, 「primitive」, 「element」, 「sub-pattern」, 「formal element」, 「ideographic minimum」, 「kanjigrapheme」などという用語も使用されている。Kordek (2013: 74) によると中国と台湾には漢字のスタンダードの構成要素 (Basic components) システムを決める特別委員会がある。そして中国で扱うスタンダードの構成要素の数は560種類であり、台湾で扱うスタンダードの構成要素の数は517種類である。表1 (次頁) に示すように、日本の漢字の構成要素の各システムでは、構成要素が9種類から632種類まで扱われていることが明らかになった。

白石 (1978) は1975字種の漢字字体の基本形として307種類を提案している。ハルペン (1987) は旧常用漢字1945字種の漢字を分析した結果、元素 (原子) と認められる要素が308種類あるという結論に達した。Heisig (1977/2001) は2042字種の漢字をもとに229種類の primitive を考え、個々の漢字の基本的な意味を表すキーワードを決め、それをもとに漢字を覚えるシステムを構築した。Harbaugh (2009) は component と名づける構成要素を増分式で構成し、合成漢字を構築した。Harbaugh (2009) が扱っている182種の component の中で132種は部首のリストに入っている構成要素である。

また、山田ボヒネックの「KanjiKreativ」という旧常用漢字学習プログラムでは、漢字の280種類の基本的な最小意味的単位 (原子) の意味をアニメーションで教え、その後体系的増分式で旧常用漢字の覚え方を紹介している (山田ボヒネック 2008)。Demirci (1997) は個々の漢字を分析し、117字の「新部首」を扱って、「漢字家族木システム」を考案し、1字の漢字を家族の一員、または原子の核のように見立てる。同じ構成要素を含めた漢字を同時に覚えさせるシステムである。

¹ 本研究では「常用漢字」を以下のように表現することがある。
1981年内閣告示の常用漢字1945字→旧常用漢字
2010年内閣告示の常用漢字2136字→現常用漢字

表1 漢字の構成要素のシステムの例

出典	構成要素の 類型数	構成要素の名称	扱う 漢字数	カバーする 漢字数
Hadamitzky & Spahn (1981)	79	radical	1945	不詳
Demirci (1997)	117	新部首	1945	不詳
Harbaugh (2009)	182	component	約 4000	約 4000
坂野ら (2009)	215	部分 (part)	512	512
Wieger (1915/1965)	224	primitive	1500	1500
Heisig (1977/2001)	229	primitive	2042	2042
山田ボヒネック (2008)	280	原子	1945	1945
白石 (1978)	307	基本形	1975	1975
ハルベン (1987)	308	元素 (原子)	1945	1945
Foerster & Tamura (1994)	484	grapheme	1945	1945
Stalph (1989)	485	kanjigrapheme	1945	1945
鹿島 (2006)	631	文字単位	1945	1945
Smolensky (2004)	632	radical	1945	1945
ヴォロビヨフ (2011, 2014)	422 202 220	構成要素, そのうち: 部首 準部首 (グラフィウム)	2136	2136
本研究	307 202 105	構成要素, そのうち: 部首 準部首 (グラフィウム)	2136	2136

※「扱う漢字」から構成要素を抽出する。Hadamitzky & Spahn と Demirci を除く、「扱う漢字数」=「カバーする漢字数」のものについては、抽出した構成要素群だけを使って「扱う漢字」に含まれるどの漢字でも構成することができる。

以上のように、漢字字体に関わる研究には多様な構成要素のシステムが存在している。この事実から、漢字の構成要素を確定し整理しようとする傾向があることがうかがえる。一方で、部首以外で常用漢字をカバーする構成要素の種類はスタンダード化されていないことも明白である。また、上記の構成要素のシステムでは、文字符号化集合の国際規格ユニコード (The Unicode Standard) 7.0.0 (2014.06) (<http://www.unicode.org/versions/Unicode7.0.0/>) によってスタンダード化されている 214 種類の部首のシステムを構成要素として採用していないものもある。例えば Stalph (1989: 73) が扱っている 485 種類の kanjigrapheme の中で 184 種類は部首であるが、30 種類の部首を扱っていない。

上記の構成要素のシステムは、主に学習者が漢字字体を覚えるために連想記憶法を活用できるように工夫されている。Foerster & Tamura (1994), Demirci (1997), 坂野・池田・品川・田嶋・渡嘉敷 (2009) は構成要素のシステムに基づく漢字索引も作成している。表 1 には入っていないが、齋藤・川上・増田・山崎・柳瀬 (2003) は、JIS 第一水準に属する 2965 種類の漢字を分析し、構造分解をおこなって「形態要素 (部品)」を抽出し、その抽出作業によって得られた要素の出現頻度を測定した。

本研究でも漢字の構造分解と分析をおこない、部首のリストに入っている要素以外で使用さ

れている最小意味的単位を抽出し、それに準部首という名前を付け、部首と準部首を含む構成要素のシステムの案を作成した。2136 字の現常用漢字で使用される準部首を抽出する際に主に The Unicode Standard と CHaracter Information Service Environment project (CHISE) (<http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/>) を参考にした。体系化と統一化のため、従来の 214 種類の部首を分解せずにそのまま構成要素として使用する方法を提案している。

本研究における分析の結果、現常用漢字では 214 種類の部首の中から 202 種類の部首が使用されていることが明らかになった。抽出した準部首の種類数は 105 であり、扱っている構成要素の合計は 307 個である。

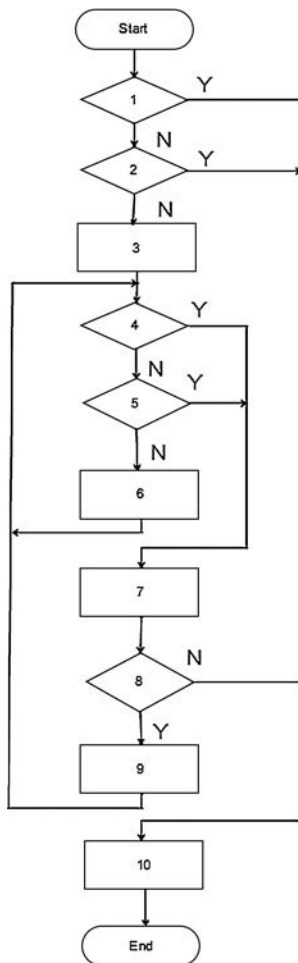
2.3 部首の使用

部首の中には、現在、漢字として使用される部首（例えば、「一」、「土」、「見」）と漢字として使用されていない部首（例えば、「一」、「彳」、「艹」）がある。前者は「単体漢字」、後者は「漢字の部品」となっている。いずれにしても、漢字の構成要素である 214 種類の部首が特別な役割を果たしている。伊藤（1988: 110）は 2 つの側面から部首についての知識の重要性を以下のように主張している。「部首を知ることが漢字の構成を理解する上で有効であり、漢字辞典で検索する上でも便利である。」

前田・阿辻（2009）によると、1615 年に作成された『字彙』は 214 部首で漢字を扱い、部首を画数順に並べた初めての字典だったという。その後、現在も漢字辞典では同じ部首のシステムが使用されている。文字符号化集合の国際規格 The Unicode Standard にはその 214 種類の部首の表が入っていて、漢字は 214 の部首順で並べられ、定められた部首番号が付いている。

しかし、異なる「部首」のシステムを構築する研究者もいる。例えば、長澤（1974）の部首索引では 295 種の部首が設けられて、「人、入」や「土、士」などは同じ部首として、「刀、リ」や「人、イ」などは別の部首として扱われている。また Matthews（2004）は、Nelson（1997）の辞書に入っている漢字の 214 部首による分類を分析した結果、伝統的な分類と違う分類のケースが少なくないと報告している。例えば「齒」という漢字は伝統的な分類では 211 番の部首であるが、Nelson（1997）では 77 番の「止」という部首に属している。同じように「服」という漢字は歴史的な分類では 74 番の「月」という部首に属しているが、Nelson（1997）では 130 番の「肉」という部首に属している。さらに、Demirci（1997）は 117 種の「新部首」のシステムを構築した。赤塚・阿部（1980）では 260 種の「部首」を扱っている。そして Hadamitzky & Spahn（1981: 373）は 214 種類の部首の中から 79 種類のみを抽出し、それぞれの部首に数字と文字から成るコードを付け、旧常用漢字の“Index by Radicals”を構築した。

これに対して、加藤（1988: 20）は統一した部首のシステムの利用について「なるべく新しい部首をつくらず、従来の部首の範囲で適当なものに変えて所属させるように工夫することが必要であろう」と記している。ヴォロビヨワ（2011, 2014）は体系化と統一化のために、従来の 214 種類の部首のシステムを採用することにした。



1. 確認：全体の漢字は部首のリストに入っているものであるか。
「はい」の場合「10」, 「いいえ」の場合「2」に進む。
2. 確認：全体の漢字は準部首であるか。
「はい」の場合「10」, 「いいえ」の場合「3」に進む。
3. 筆順にしたがい文字の最初の構成要素を抽出する。それから「4」に進む。
4. 確認：抽出された構成要素は部首のリストに入っているものであるか。
「はい」の場合「7」, 「いいえ」の場合「5」に進む。
5. 確認：抽出された構成要素は準部首であるか。
「はい」の場合「7」, 「いいえ」の場合「6」に進む。
6. 抽出された構成要素が部首のリストにも入っていない, 準部首でもない場合は, 構成要素の抽出を見直し, 「4」に進む。
7. 抽出した構成要素を確定する。
8. 確認：他にもまだ抽出されていない構成要素が残っているか。
「はい」の場合は「9」, 「いいえ」の場合は「10」に進む。
9. 次の構成要素を抽出し, 「4」に進む。
10. 構成要素を確定する。構造分解は終わりである。

図2 線型構造分解のアルゴリズム

漢字の線型構造分解の例

例1 安 = 宀 + 女

例2 国 = 口 + 玉

例3 雲 = 雨 + 二 + 厶

例4 謁 = 言 + 日 + 勹 + 匕

構造分解をおこなう際に注意すべきことがある。

- ・「穴」「糸」「老」「音」「麻」という構成要素は部首であり, 分解できそうだが分解しないことにする。
 - ・「行」「衣」という部首の中に別の構成要素が入る漢字がある。例えば, 「術」, 「街」, 「裏」, 「衰」。
- その場合は先に「行」, 「衣」を抽出し, 次に残った要素を抽出する。

例 5 術 = 行 + 木 + 丿

例 6 裏 = 衣 + 里

線型構造分解の意義は以下の通りである。

- ・漢字の構成要素の確定をおこない、字体と字義の関連の分析を可能にする。
- ・漢字群における構成要素の出現頻度の測定を可能にする。
- ・新しいタイプの漢字索引の構築を可能にする。

3.2 漢字のコード化

本研究では、漢字の構造を明確に表すことや漢字字体の計量的分析を目指し、漢字字体のコード化を開発した。これは、文字符号化集合の国際規格 The Unicode Standard とは異なるものである。漢字のコード化と、コード化に基づいた画と構成要素の使用頻度の測定や新しい検索法の案についてはヴォロビヨワ (2011, 2014) と Vorobeva & Vorobev (2012) に詳しく述べられている。

3.2.1 漢字字体のコード化に関する先行研究

先行研究には次の漢字字体のコード化とその利用の例がある。

・王雲五 (1925) による四角号碼 (しかくごうま) の漢字コード化

王雲五 (1925) (Wong Yunwu) により中国で考案された「四角号碼」, 日本語に訳せば「四隅コード」という漢字コードのシステムがある。漢字の四隅に書いてある画の形により 0 から 9 まで番号を付与する。5 桁の数字のコードの例は次の通りである。例えば, 漢字「法」のコードは「34131」である。四角号碼のコードは漢字の全体の構成を表さず, 四隅に書いてある画の形のみを示している。コード化の目的は漢字の検索のみである。

・Halpern (1988/1990) による System of Kanji Indexing by Patterns (SKIP) の漢字コード化

Halpern (1988/1990) は各漢字に 3 桁の数字のコードを付けた。そのために, まず, 漢字字体の 4 つの基本的なパターンを確定し, それぞれのパターンに 1 から 4 までの数字を当てた。この数字は漢字コードの最初の数字となる。漢字の部分の画数などがコードの 2 番目と 3 番目の数字となる。例えば, 漢字「相」のコードは「1-4-5」である。SKIP (System of Kanji Indexing by Patterns) のコードは漢字の全体の構成を表さず, 漢字構成のタイプと漢字の部分の画数などを表している。コード化の目的は漢字の検索のみである。

・若尾・服部 (1989) による筆順索引の漢字コード化

若尾・服部 (1989) によって作成されたコードは『くずし解読辞典』で使用されている。筆の動きが 8 つの方向の矢印で示されて, それぞれの方向には 0 から 7 までの決まった数字のコードが付けてある (若尾・服部 1989: 466)。漢字コードは, 最初に書くストロークのコードを含む 4 桁の数字のコードである。例えば, 漢字「仙」のコードは「5-1-4-4」である。コードは漢字の

全体の構成を表さず、筆順にしたがい先を書くストロークを示している。コード化の目的は漢字の検索のみである。

3.2.2 本研究における漢字字体のコードの例

ヴォロビヨワ (2011) は、漢字の構造分解に基づき、3つの漢字コードのシステムを構築し、それを土台にしたデータベースの構築について紹介した。本研究では現常用漢字すべてのコード化をおこない、漢字の「アルファベット・コード」と「シンボル・コード」(アルファベットと数字のコード)の2つのデータベースを構築した。表2に漢字のアルファベット・コードとシンボル・コードの例を示す。

表2 漢字のアルファベット・コード, シンボル・コードの例

漢字	アルファベット・コード	シンボル・コード
九	PR	4/5
逸	PYBHBAPCQMO	8PYB/162
新	SAQLAABPOPPAB	117/75/69
族	SAGPPAPAAPO	70/2PA/111

以下で上記の漢字コードの各々のタイプについて紹介する。

3.2.3 漢字のアルファベット・コード化

Zadoenko & Khuan (1993) は中国語の漢字指導では24種類の画を扱っている。ヴォロビヨワ (2011) が旧常用漢字と現常用漢字の個々の漢字を分解した結果、それらをカバーするのに Zadoenko & Khuan (1993) が指摘した24種類の画で必要十分であることが明らかになった。そこで、この24種類の個々の画にローマ字の形を結び付け、AからZのアルファベット・コードを付けた。その際、漢字の画の形からアルファベットの形が連想できるよう(思い浮かべられるよう)にした(表3)。

表3 漢字の24種類の画とそのアルファベット・コード

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

筆順にしたがい、漢字の画の代わりに当該のコードを当てて漢字のアルファベット・コードを作った (表2)。

3.2.4 本研究における構成要素と漢字のコード化

ここでは部首のコード化、準部首のコード化、漢字のコード化について紹介する。

・部首のコード化

本研究は、各部首に次の2種類のコードを付けた。

- (1) 部首に含まれているすべての画のアルファベット・コードの列を部首のアルファベット・コードとした。
- (2) 1から214までの従来の部首番号を部首のシンボル・コードとした。それは1716年に中国で完成された『康熙字典』の1から214までの部首の通し番号であり、その後多くの漢和辞典で使用され、文字符号化集合の国際規格ユニコード (The Unicode Standard) でも使用されている。

・準部首のコード化

現常用漢字から105種類の準部首を抽出した。次に部首と同様に、準部首にも2種類のコードを以下のように付けた。

- (1) 準部首に含まれるすべての画のアルファベット・コードの列を準部首のアルファベット・コードとした。
- (2) 準部首のシンボル・コードは、画数を表す数字と、筆順で3番目までの画のアルファベット・コードを含む。例えば、「万」—3AGP, 「五」—4ABH。同じコードの準部首がある場合は、コードの後ろに1からの数字を加えて区別するようにした。例えば、「乗」—9PAA, 「卸」—9PAA1。

・漢字のシンボル・コード

現常用漢字をカバーする漢字の構成要素のシステムを構築し、それに基づき、構成要素のコード化とそれをベースにした漢字のコード化をおこなった。漢字のシンボル・コードを作るために、筆順にしたがい、その漢字のすべての構成要素のシンボル・コードを書いた。漢字の構成要素のコードの間にはスラッシュ「/」を入れた (表2)。

3.2.5 本研究における漢字コード化の意義

漢字のアルファベット・コードおよびシンボル・コードの利用の意義は以下の通りである。コード化を利用して漢字字体の統計的分析などの問題が解決できる。

・漢字のアルファベット・コードを利用することで、次のことが可能になる。

- (1) アルファベット・コードによる漢字の筆順のコンパクトな表示
- (2) 漢字のアルファベット・コードのデータベースの構築
- (3) 上記のデータベースをもとにした画の出現頻度の測定
- (4) 上記のデータベースをもとにした漢字のアルファベット・コード索引の開発
- (5) 漢字の筆順の練習とテストのためのソフトの開発

・漢字のシンボル・コードを利用することで、次のことが可能になる。

- (1) 漢字のシンボル・コードのデータベースの構築
- (2) 上記のデータベースをもとにした漢字の構成要素の出現頻度の測定
- (3) 上記のデータベースをもとにした漢字のシンボル・コード索引の開発
- (4) 漢字の構成上の複雑さの表示
- (5) 複雑さによる漢字の分類
- (6) 漢字教材における漢字の掲出順序の分析

4. 漢字の階層構造分解

漢字のより深い構造分析をおこなうには、線型構造分解だけではなく、階層構造分解についても考える必要がある。ここでは言語学における階層的なアプローチの意義、漢字の階層構造分解について検討する。

4.1 階層的アプローチ

漢字の指導は学習対象漢字の選択と掲出順序の決定から始まる。漢字教材が本教材に準拠した副教材である場合、配列は本教材に漢字が登場する順序にしたがうため、構造的に複雑な漢字を簡単な漢字より早く教えることも少なくない。Gagne (1968) は学習階層論について述べている。そこでは、学習課題間の階層関係を考慮に入れることが必要であると主張している。そして学習順序を設計する際は、単純なものから複雑なものに進むよう考慮すべきだと記している。

階層的なアプローチの利用の方向は次のように考えられる。

- (1) 学問の内容の階層化
- (2) 「簡単な概念から複雑な概念へ」という教育過程の管理

概念間の関係を樹形図で示すと階層構造になる。合体漢字は階層構造を有する。階層数は漢字の複雑さによって異なる。漢字の構成上の複雑さの判定基準については後で述べる。上位階層と下位階層の関係には包含の関係がある。例えば、図3と図4の樹形図の矢印は、複雑な要素からそれが包含している簡単な要素への方向を示している。

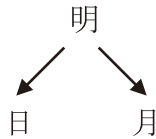


図3 漢字の樹形の例1

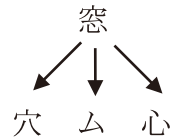


図4 漢字の樹形の例2

漢字の指導法では教育における階層的なアプローチが効率的だと考える。漢字の構成要素と画への階層構造分解を教育に応用した階層的なアプローチが重要である。「簡単な漢字から複雑な漢字へ」というアプローチというのは必ずしも先に単体漢字、その後組み合わせ漢字を教えることではない。必要な場合は組み合わせ漢字を先に教えるが、その際字体と字義の関連の理解のために漢字の階層構造と要素の意味も教える必要がある。構成要素の意味を組み合わせることで、学習者が未習の合体漢字の意味をある程度推測できるように教えたほうがよいということである。この点について、白石 (1978: 3) は「要素形的漢字学習指導法は、換言すれば、「基本形」と「基本形の積み重ね」との指導の方法である」と指摘している。

ちなみに、Mesarovic, Macko & Takahara (1970: 62) は複雑系 (complex system) を理解するためには階層的なアプローチが基本的な役割を果たしていると述べている。Mesarovic らは、マルチレベル分析をしないと解決できない重要な科学的な問題が増加していると主張しており、階層構造の分析方法を検討している。Mesarovic らは階層構造の理論を主に経済的な複雑系に対して応用したが、彼らの理論は漢字にも応用できる。松村 (2006) によると、複雑系は「多くの要素からなり、部分が全体に、全体が部分に影響しあって複雑に振る舞う系」である。漢字は構成要素と画の複雑な組み合わせであり、間違いなく複雑系であると言えよう。漢字の階層構造分析は漢字の体系的で効率的な教授法の開発の際に利用できる。

4.2 英語の文、英単語及び漢字の階層構造の類似点

Chomsky (1966: 73) が提唱した変形生成文法理論の基本概念に「深層構造」と「表層構造」がある。「深層構造」は文の意味の解釈を確定し、「表層構造」は文の音声解釈を確定する。Chomsky (1957/2002: 27) は“The man hit the ball.”という文の階層構造を図5のように表している。

その理論は漢字にも適用できる。漢字は思想を図式の方法で伝えるために作成された文字であり、目に見える形「字体」と、目に見えない (思想を表現する)「字義」がある。その2つの側面で漢字を分析することができる。Chomsky 理論の用語を援用して、「字体」は漢字の「表層構造」に、「字義」は漢字の「深層構造」に相当すると考えてみよう。漢字の「表層構造」は、視覚的な字体を表す要素の組み立て方を示している。それに対して、漢字の「深層構造」は、漢字の構成要素の組み合わせの意味の解釈を確定していると言える。

図5は、NP (Noun Phrase)一名詞句、VP (Verb Phrase)一動詞句、N (Noun)一名詞、V (Verb)一動詞、T (The)一定冠詞を示している。

階層構造を有する単語も存在している。Plag (2003: 39) は3つの形態素 (morphemes) から成

る英単語 “unregretful” の階層構造 (the relationship between the three morphemes) を図 6 のように示す。図 7 に示す 3 つの構成要素から成っている漢字「路」の階層構造も、図 6 の英単語の階層構造に似ている。

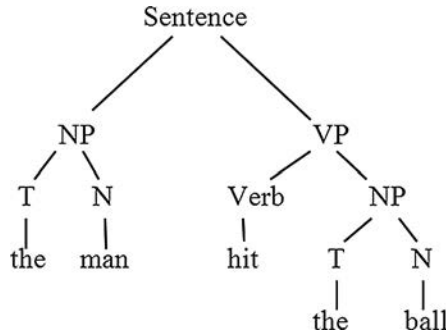


図 5 Chomsky (1957/2002: 27) による文の階層分解の例

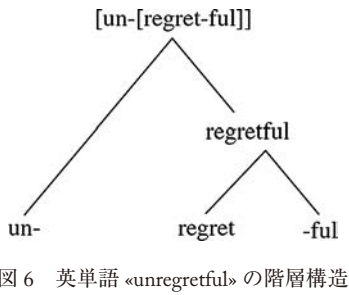


図 6 英単語 «unregretful» の階層構造

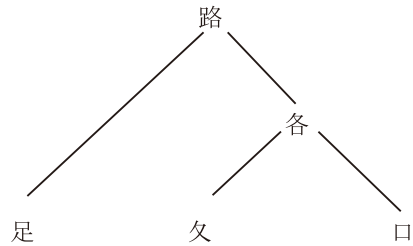


図 7 漢字「路」の階層構造

つまり、英語の文や語構成における統語的階層構造と、漢字の階層構造には類似点があり、言語学における文や単語の階層構造分析方法を漢字にも適用できそうである。本研究では漢字の「表層構造」と「深層構造」の関連を考慮に入れ、漢字の構造分解を検討する。漢字の構造分解と構成要素の意味の分析は、より深い字義の理解と記憶を促進することにつながる。漢字の構造分解では 2 段階が考えられる。それは構成要素と画への階層分解である (図 8)。

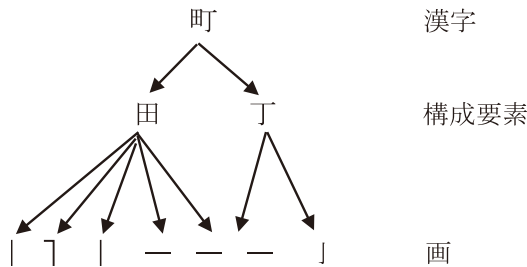


図 8 漢字の 2 段階の構造分解の例

4.3 藤村による漢字の階層構造の分析とコード化

漢字の構造分解とコード化の試みの例として、藤村（1973）の研究を紹介する。藤村（1973）は、漢字を階層的に分解し、それをもとに線型式で網羅的に漢字の構造を記述した。藤村（1973: 18）は、漢字の字体について「いろいろな要素，扁，旁，垂等から構成される複合的な構造を有するばかりでなく，またそれらの構成要素が画といわれる素子から成る」と述べている。

藤村（1973: 19）は、画の配置，結合のしかたに規則的な条件があるであろうということ を考慮に入れ，構成要素と画の階層的な記述に数字とアルファベットとギリシャ文字を用いて，漢字の要素を次のようにコード化した。例えば，横画を数字「1」で，縦画を数字「2」で，画と画の結合のしかたをアルファベットで表した。それから上記のコードによって画の結合配置のしかたを示した。例えば，「結合子 S はその左側の画（前画と呼ぼう）の頭（横画なら左端，縦画なら上端）と右側の画すなわち後画の頭とを重ねるように二画の相対位置を指定する。P は前画の尾と後画の頭とを一致させる。」（藤村 1973: 19-20）。

例 7 丁 1C2, 十 1X2(=2X1), 丄 2E1, 丿 1T2(=2T1), 丨 2P1, 二 1V1, 丩 2H2
(藤村 1973: 19 表 1 より)

また，画の特性点として頭と尾と中央をギリシャ文字で表した。

例 8 藤村（1973: 26）による「鱗」という漢字のコードは下記の通りである。

(4S14V2S12C2X1E1V3333)H((331X233)V(4S143H1φ1X2))

その中で各々の要素のコードは

ク : 4S14, 田 : 2S12C2X1E1, ヽ : 3, 米 : 331X233, タ : 4S143, 幺 : 1φ1X2

というようになる。

「鱗」という漢字の構造分解を図 9 に示す。(H は縦画 2 つを横に並べるもの，V は横画 2 つを縦に並べるものを表す。)

藤村（1973）によるコードは網羅的であるため複雑に見える。画の扱い方の特徴は，分解の際に折れ線である画を 2 つの線としてコード化する点にある。例えば，上記の例の「夕」という構

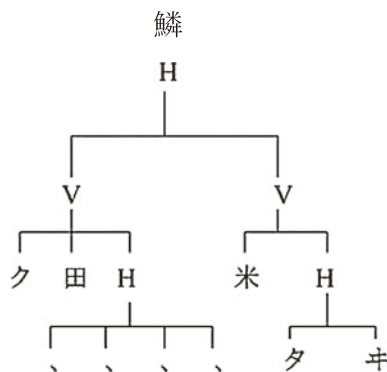


図 9 藤村（1973）による漢字「鱗」の構造分解

成要素は3つの画（一筆で書くもの）, 「ノ」, 「フ」, 「丶」から成っている。しかし藤村（1973）は4つの線, 「ノ」（コード4）「一」（コード1）「ノ」（コード4）「丶」（コード3）に分解している。

藤村（1973）が開発した構造の記述方法は独特で、網羅的に漢字の階層構造を文字・数字の列で表すことができる。しかし、残念なことに、実際にその記述方法が広範囲に活用されるようにはなっていない。また、漢字字体の記述のみを目的とし、そこから実際に教育現場などで使用する試みが見られない。

本研究では、2種類の漢字コードシステムを独自に開発した。それは、2節で記述されたアルファベット・コードとシンボル・コードのシステムである。これらは藤村（1973）と異なり、実用的である。藤村（1973）と本研究の漢字の構造分解とコード化は次のポイントにおいて異なっている。

(1) 目的

藤村（1973）の場合は、漢字字体の構造分解とコードによる記述を目的としている。それに対して、本研究では漢字字体の構造分解とコード化は手段である。目的は以下のような実践使用である。

- ①アルファベット・コードは漢字の筆順を表すため
- ②アルファベット・コードとシンボル・コードは画と構成要素の使用頻度の測定のため
- ③アルファベット・コードとシンボル・コードはコード索引の開発のため

(2) レベル

藤村（1973）の場合は構造モデルの中で構成要素と画を混ぜて扱っている。それに対して本研究では構成要素の構造モデルと画の構造モデルを分け、別の目的で使用している。

(3) 要素

藤村（1973）は部首も画も細かく分解するが、それに対して本研究では The Unicode Standard によってスタンダード化された部首と画を分解しないようにする。

(4) 要素の位置

藤村（1973）によるコードは要素のお互いの位置を表すコードである。それに対して本研究では漢字字体の網羅的な記述を目指さず、コードを複雑すぎるものにならないように、漢字の要素の筆順に基づき、順序のみを示して、要素のお互いの位置を表さないコードにする。上記の図9と同じ「鱗」という漢字の構造分解を図10に示す。

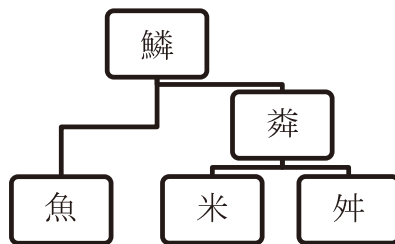


図10 本研究による漢字「鱗」の階層構造分解

本研究による「鱗」の階層構造の公式とコードは、次の通りとなる。

- ①漢字の階層構造の公式 「鱗=魚+彡 (米+舛)」
- ②アルファベット・コード 「PYBHBAALQQQQLABPOPYQABAB」
- ③シンボル・コード 「195/119/136」

(5) 教育的側面

藤村 (1973) と異なり、本研究における目的は構造分解と構成要素の意味の分析を通して、漢字の字体と字義の連関を日本語教師と日本語学習者が漢字指導・漢字学習でより効果的に使用できるようにすることである。そこで漢字の構造分析をもとに作成された漢字教材 (ヴォロビヨワ 2007, ヴォロビヨフ & ヴォロビヨワ 2007) と漢字指導の手引き (Vorobev & Vorobeva 2013) で使用した。

4.4 漢字の階層構造分解, 樹形図と数学的公式

漢字の階層構造分解と漢字の構成は逆のプロセスをたどる。漢字の構成の例を図 11 に、階層構造分解の例を図 12 に示す。漢字の構成は、図 11 の例のように意味的単位から体系的増分式で漢字を構成する。矢印は簡単な要素から複雑な要素への方向を示している。2 個の構成要素を含む要素は第 2 層になる。「露」という漢字には構成要素の層が第 1 層から第 4 層までである。

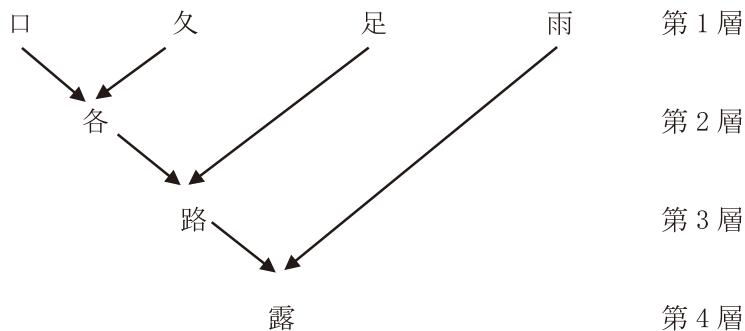


図 11 漢字の構成の例

漢字の構成要素への階層分解は、漢字を筆順にしたがい、順番に意味的単位に分解することである (図 12)。「露」は合体漢字であり、「雨, 足, 夕, 口」はその構成要素としての単体漢字であり、「路, 各」は中間漢字である。漢字「露」の階層的な数学的公式は次の通りである。

$$\text{露} = \text{雨} + \text{路} (\text{足} + \text{各} (\text{夕} + \text{口}))$$

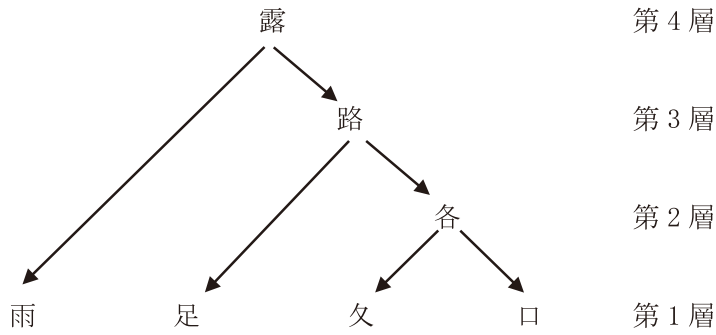


図 12 漢字の階層構造分解の例

これ以上分解できない要素を一番下の第 1 層に置く。階層構造分解の教育的な意義は、個々の漢字を最小意味の単位まで分解せずに、それらの合体である既習の形（中間漢字）までの分解とすることで十分であるため、連想記憶法を使用しやすいという点にある。「露」という漢字を覚える場合、「路」という漢字が既習であれば、「雨, 足, 夕, 口」をばらばらに覚える必要はなく、「雨」と「路」の 2 つを連想でつなげて覚えればよい。本研究ではこのように現常用漢字をすべて分解した。その際に 3 種類の間接漢字があることが明らかになった。

- 1) 現常用漢字群に入っている漢字 307 字
- 2) 現常用漢字群に入っていないが、The Unicode Standard に入っている漢字 276 字
- 3) 現常用漢字群にも、The Unicode Standard にも入っていないが、「文字鏡単漢字」（文字鏡研究会 2002）に入っている漢字 135 字

現常用漢字に含まれる準部首を抽出し、部首のリストと準部首のリストに入っている 307 種類の構成要素でそのすべての常用漢字をカバーすることができた。現常用漢字のシンボル・コードと階層構造の公式の例を表 4 に示す。

表 4 漢字のシンボル・コードと階層構造の例

漢字	シンボル・コード	階層構造の公式
花	140/9/21	艹 + 化 (イ + 匕)
学	42/14/39	灬 (灬 + 冫) + 子
气	84/2PO	气 + 丩
休	9/75	亻 + 木
空	116/48	穴 + 工
校	75/8/88	木 + 交 (亠 + 父)

4.5 漢字の階層構造における 3 種類の構成要素の役割

上記で述べた構造分析では、漢字の構成要素間の関係については分析をおこなわなかったが、漢字の構成要素の関係は実際には簡単ではなく、以下の漢字の構成要素の役割を考慮に入れながら慎重に分析する必要がある。Kordek (2013: 79) によると、漢字の個々の構成要素の役割は以

下の通りである。

- 1) (Ph) – Phonetic, 音符 (読み方を表す)
- 2) (Sig) – Signific, 意符 (漢字の字義の確定のヒントになる)
- 3) (Sym) – Symbolic, 記号 (他の漢字との違いを示す)

上記の概念を考慮に入れ、漢字の階層構造を表した例が図 13 である。音読みと構成要素の役割が示してある (<http://www.chise.org/>)。

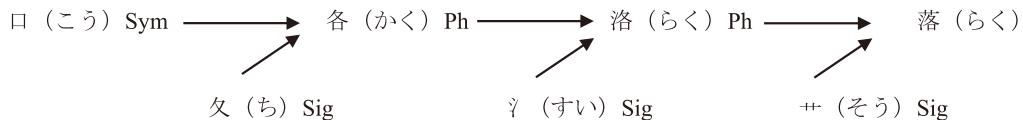


図 13 4 層の漢字「落」の階層構造の例

上記の構成要素の役割を考慮した研究は今後の課題としたい。

5. 漢字の構成上の複雑さの判定基準と複雑さ指標の定義

5.1 漢字の構成上の複雑さの判定基準

先に、教育には「簡単な概念から複雑な概念へ教える」という原則があると述べた。この原則にしたがって学習対象漢字の合理的な掲出順序を検討する際は、「簡単な漢字から複雑な漢字へ」教えるようなアプローチを考慮に入れる必要があると考える。そのためには、まず、漢字の複雑さの判定基準を定義しなければならない。

漢字の複雑さの要因について先行研究には次のような言及がある。海保・野村 (1983: 122) によると、「漢字は構成要素の多さと多次元的配列という点で複雑である」。例えば、「口」と「貝」という構成要素から「貝」あるいは「唄」の形、「田」と「丁」から「町」あるいは「早」の形ができる。

また、賀集・石原・井上・齋藤・前田 (1979) は漢字の構成上の複雑さの要因について検討し、漢字を種々の線分や点の組み合わせとして考え、画数が漢字の複雑さを表す要因であろうと指摘する。あわせて、「画」よりも下位レベルの「漢字の線」という概念を導入し、それにより漢字の複雑さを確定する案を示している。賀集・石原・井上・齋藤・前田 (1979: 104) は「漢字の線」について次のように記している。「“山”や“口”のように、いずれも3画であるが、4つの線で構成されている」。さらに真田・横山 (2007: 24) は、漢字の複雑さを計るには画数、ドット数、主観的複雑度だけでなく、画の長さ、交点の数などを数える方法があると分析している。

先行研究をまとめてみると、漢字の構成上の複雑さの要因として、構成要素数、画あるいは画より細かい要素の数、漢字の中の要素の位置、の3点が考えられることが明らかになった。しかし、漢字の構成上の複雑さの判定基準は決まっていない状況である。

5.2 漢字の複雑さに関する指標の定義

漢字の構成上の複雑さを判定する規準（以下、指標とよぶこともある）を定義するため、2つの客観的な性質、構成要素数と画数からできる $V = (NE, NS)$ という2次元のベクトル（行列）で漢字の複雑さを定義することにした（ヴォロビヨフ 2011: 26）。

NE - 漢字の構成要素数

NS - 漢字の画数

例えば、「森」の構成要素数（NE）は3、画数（NS）は12であり、複雑さのベクトルの指標は $V(\text{森}) = (3, 12)$ である。

5.3 複雑さによる現常用漢字 2136 字の分類

漢字の構成上の複雑さを判定する指標の定義に基づいて、現常用漢字のシンボル・コードのデータをもとに個々の漢字の複雑さの指標を算出した。本研究で扱う 307 種類の構成要素をベースにすると、現常用漢字は、1 個から 7 個の構成要素から成り立っている。その構成要素数による分布を表 5 に示す。

表 5 構成要素数による現常用漢字の分布

構成要素数	1	2	3	4	5	6	7
漢字の数	191	693	736	367	123	21	5

現常用漢字の中でもっとも多いのは 2 個と 3 個の構成要素から成る漢字である。1 ~ 3 個の構成要素からできている漢字は $191 + 693 + 736 = 1620$ 字であり、全体の常用漢字の 75.8% を占めている。つまり、常用漢字の中に、構成要素が多くてその組み合わせにおいて複雑な漢字はそれほど多くないという事実が表 5 から分かる。表 6 中の数字は構成要素数（1 ~ 7）と画数（1 ~ 23）による各グループの漢字数を示す。例えば構成要素数 = 2 と画数 = 5 の漢字数は 60 である。

表 6 構成要素数と画数による 2136 字の現常用漢字の分布の一部

↑ 構 成 要 素	7	0	0	0	0	0	0	0	0	0	0	0	1	2	2	0	0	0	0	0	0	0	0	0	
	6	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	4	4	1	1	4	2	0	0	0
	5	0	0	0	0	0	0	1	0	2	4	13	20	20	7	21	8	8	12	3	1	1	0	2	
	4	0	0	0	1	0	1	7	10	27	52	48	59	34	39	33	23	9	12	7	4	1	0	0	
	3	0	0	1	3	10	26	51	96	84	82	80	82	72	46	42	25	12	11	8	2	3	0	0	
	2	0	4	10	35	60	62	73	83	67	71	64	47	40	21	17	17	12	4	1	2	1	2	0	
	1	2	8	20	30	29	25	21	15	14	9	8	4	3	1	2	0	0	0	0	0	0	0	0	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		
	画 数→																								

常用漢字の中には画数が 29 の漢字も 1 字あるが、表 6 では画数が 23 までを示した。

6. おわりに

本研究は、漢字の構造分解の方法を紹介し、漢字の構造分析に関わる問題を検討した。

- (1) 現常用漢字をカバーする「スタンダード化された構成要素のシステム」を開発する必要性について述べた。その開発に成功すれば、漢字の認識や漢字習得の体系化もできる。
- (2) 構成要素のシステムについて先行研究における部首の扱い方を分析した結果、従来の部首を採用していないシステムが多いことが明らかになった。
- (3) 漢字字体の構造を明確に表現・表示できれば、漢字字体の計量的分析が可能になる。そこで、漢字の線型構造分解をおこなった上で漢字字体に関する独特のコード化を開発し、アルファベット・コードとシンボル・コードのシステムを構築した。漢字コード化の結果に基づいて漢字の画と構成要素の出現頻度の測定が可能になり、漢字の構成上の複雑さを定義する新たな指標を提案することができるようになった。
- (4) 英語の文や語構成における統語的階層構造と、漢字の階層構造が一部において類似していることを示した。漢字の階層構造分解をおこない、樹形図と数学的公式で漢字の階層構造を表示した。さらに藤村（1973）による漢字の階層構造の分析とコード化と、本研究による階層構造の分析とコード化を比較し、本研究の実用性を示した。
- (5) 漢字の構成上の複雑さを判定する新たな指標を定義し、複雑さによる現常用漢字の分類をおこなった。

今後の課題は大きく2つある。1つは、漢字の構造分解とコード化で得られた知見とデータに立脚しながら漢字字体に関する計量言語学的な検討をおこなうことである。もう1つは、漢字の階層構造における3種類の構成要素の役割を考慮した研究を進めることである。

参考文献

- 赤塚忠・阿部吉雄（1980）『旺文社漢和辞典』東京：旺文社。
 坂野永理・池田庸子・品川恭子・田嶋香織・渡嘉敷恭子（2009）『イメージで覚える「げんき」な漢字512』東京：The Japan Times。
 Chomsky, Noam (1957/2002) *Syntactic structures*. Second Edition. Berlin and New York: Mouton de Gruyter.
 Chomsky, Noam (1966) *Cartesian linguistics: A chapter in the history of rationalist thought*. New York and London: Harper & Row Publishers.
 Demirci, Muharrem (1997) 『日本語・トルコ語新常用漢字辞典』 Istanbul: J.K.E.M.
 Foerster, Andreas and Naoko Tamura (1994) *Kanji ABC: A systematic approach to Japanese characters*. Boston: Tuttle Publishing.
 藤村靖（1973）「漢字の構造」『月刊言語』2(7): 19-27. 東京：大修館書店。
 Gagne, Robert (1968) Learning hierarchies. *Educational Psychologist* 6: 1-9.
 Habein, Yaeko and Gerald B. Mathias (1991) *The complete guide to everyday kanji*. Tokyo; New York; London: Kodansha International.
 Hadamitzky, Wolfgang and Mark Spahn (1981) *Kanji & kana revised edition: A handbook of the Japanese writing system*. Rutland, VT: Tuttle Language Library.
 ハルペン・ジャック（1987）『漢字の再発見』東京：祥伝社。
 Halpern, Jack (1988/1990) *New Japanese-English character dictionary*. Tokyo: Kenkyusha.
 Harbaugh, Rick (2009) 『中文字譜 漢英字源字典』USA :Yale University.
 Heisig, James W. (1977/2001) *Remembering the kanji*. Vol. 1. Tokyo: Japan Publications Trading Co. Ltd.
 伊藤芳照（1988）「第3章 外国人学習者に対する表記の指導」『日本語教育指導参考書 14 文字・表記の教育』

- 72-145. 東京：国立国語研究所。
- 海保博之・野村幸正 (1983) 『漢字情報処理の心理学』 東京：教育出版。
- 鹿島英一 (2006) 『漢字の情報理論』 福岡：白沙ヶ濱。
- 賀集寛・石原岩太郎・井上道雄・齋藤洋典・前田泰宏 (1979) 「漢字の視覚的複雑性」 『人文論究』 29(1): 103-121. 関西学院大学人文学会。
- 加藤彰彦 (1988) 「第 1 章 日本語の表記の基準」 『日本語教育指導参考書 14 文字・表記の教育』 1-49. 東京：国立国語研究所。
- 『康熙字典』 (1715) 中国：中華書局。
- Kordek, Norbert (2013) *On some quantitative aspects of the componential structure of Chinese characters*. Poznan: Wydawnictwo Rys.
- 前田富祺・阿辻哲次 (2009) 『漢字キーワード事典』 東京：朝倉書店。
- 松村明 (編) (2006) 『大辞林 第三版』 東京：三省堂。
- Matthews, Laurence (2004) *Kanji fast finder*. Tokyo: Tuttle Publishing.
- Mesarovic, Mihajlo D., D. Macko and Yasuhiko Takahara (1970) *Theory of hierarchical, multilevel systems*. New York: Academic Press.
- 文字鏡研究会 (編) (2002) 『パソコン悠悠漢字術 2002 今昔文字鏡徹底活用』 東京：紀伊國屋書店。
- 長澤規矩也 (編) (1974) 『新明解漢和辞典』 東京：三省堂。
- Nelson, Andrew N. (1997) *The new modern reader's Japanese English character dictionary*. Boston: Tuttle.
- Plag, Ingo (2003) *Word-formation in English*. Cambridge: Cambridge University Press.
- 齋藤洋典 (2006) 「漢字の認識と発達」 前田富祺・野村雅昭 (編) 『漢字のはたらき 朝倉漢字講座 2』, 132-168. 東京：朝倉書店。
- 齋藤洋典・川上正浩・増田尚史・山崎治・柳瀬吉伸 (2003) 「意味処理における情報統合過程の解明」 『平成 15 年科学研究費補助金研究成果報告書 基盤研究 (B) (2)』 301-363.
- 真田治子・横山詔一 (2007) 「漢字の諸性質の計量言語学的研究 (1)」 『社団法人情報処理学会研究報告 IPSJ SIG Technical Report』 2007(9): 17-24.
- 白石光邦 (1978) 『要素形的漢字学習指導法』 東京：桜楓社。
- Smolensky, Vadim [Смоленский Вадим] (2004) *Японско-русская словарь иероглифов 6.2.YARXI* [Японско-русский словарь иероглифов 6.2 YARXI]. <http://www.susi.ru/yarxi/> (2014 年 11 月 6 日参照) .
- Stalph, Jürgen (1989) *Grundlagen einer Grammatik der sinojapanischen Schrift*. Wiesbaden: Otto Harrassowitz.
- 武部良明 (1984) 「漢字の単位について」 『講座日本語教育』 20: 58-72. 早稲田大学語学教育研究所。
- ヴォロビヨフ・ヴィクトル, ヴォロビヨフ・ガリーナ (2007) 『漢字物語 II』 Bishkek: Lakprint.
- Vorobev, Victor [Воробьев В.М.] and Galina Vorobeva [Воробьева Г.Н.] (2013) *Metodicheskoe posobie dlya prepodavaniya yaponskikh ieroglifov* [Методическое пособие для преподавания японских иероглифов] (Handbook for teaching Japanese kanji). Kyrgyz Republic [Кыргызская Республика], Bishkek: KNU [Бишкек, КНУ]. <http://japanese.iiiep.kg/ТОС_Metod.htm>.
- ヴォロビヨフ・ガリーナ (2007) 『漢字物語 I』 Bishkek: Lakprint.
- ヴォロビヨフ・ガリーナ (2011) 「構造分析とコード化に基づく漢字字体情報処理システムの開発」 『日本語教育』 149: 16-30.
- ヴォロビヨフ・ガリーナ (2014) 『構造分解とコード化を利用した計量的分析に基づく漢字学習の体系化と効率化』 東京：ノースアイランド。
- Vorobeva, Galina and Victor Vorobev (2012) An analysis of efficiency of existing kanji indexes and development of a coding-based index. *OPEN JOURNAL SYSTEMS: Acta Linguistica Asiatica* 2(3), Slovenia: University of Ljubljana, <<http://revije.ff.uni-lj.si/ala/article/view/180/318> pp. 27-59>.
- 若尾俊平・服部大超 (編著) (1989) 『くずし解読字典』 東京：柏書房。
- Wieger, Leon (1915/1965) *Chinese characters: Their origin, etymology, history, classification and signification: A thorough study from Chinese documents*. New York: Dover Publications.
- 王雲五 (Wong Yunwu) (1925) 『號碼檢字法』 上海：商務印書館。
- 山田ボヒネック頼子 (2008) 「KK2.0 (KanjiKreativ) E ラーニング：1945 常用漢字学習プログラム一体系的・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報告・発表論文集』 12: 169-175.
- Zadoenko, Tamara [Задоеико, Тамара] and Shuin Khuan [Шуин Хуан] (1993) *Osnovy kitajskogo jazyka: Vvodnyj kurs*. [Основы китайского языка: вводный курс - 基礎漢語]. Moscow: Nauka [М:Наука].

関連 Web サイト

CHaracter Information Service Environment project (CHISE)

<http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/> (2014 年 12 月 26 日参照)

文字符号化集合の国際規格ユニコード (The Unicode Standard) 7.0.0

<http://www.unicode.org/versions/Unicode7.0.0/> (2014 年 11 月 6 日参照)

文字符号化集合の国際規格ユニコード (The Unicode Standard) 7.0.0 に属している Unihan Database

<http://www.unicode.org/cgi-bin/GetUnihanData.pl?codepoint=6BCD> (2014 年 12 月 26 日参照)

Smolensky, Vadim [Смоленский Вадим] (2004) *Японско-русский словарь иероглифов 6.2.YARXI* [Японско-русский словарь иероглифов 6.2 YARXI]

<http://www.susi.ru/yarxi/> (2014 年 11 月 6 日参照)

A Method of the Analysis of Kanji Structure: A New Approach Based on Structural Decomposition and Coding

VOROBÉVA Galina^a VOROBEV Victor^b

^aProject Collaborator, NINJAL

^bKyrgyz National University

Abstract

In this study, we specify some important problems concerning the analysis of kanji structure, after reviewing previous research on the subject. After discussing a method of analysis of kanji structure using a new approach, we suggest the following solutions:

- (1) We emphasize the need to develop a standardized system of elements that covers all of the 2136 Jōyō kanji. If such a system were successfully developed, it would be possible to systemize the recognition and acquisition of kanji.
- (2) By analyzing how previous investigations deal with radicals as part of an element system, we find that many systems that did not use the traditionally defined set of radicals have been developed.
- (3) The quantitative analysis of kanji form can be made possible by clear representation and indication. In order to achieve this, after carrying out the linear decomposition of kanji, we develop a unique code system of kanji form and constructed alphabetic code and symbolic code systems. The results of this kanji coding make it possible to measure the frequency of kanji strokes and elements, and open up the possibility of creating a new indicator of the structural complexity of kanji.
- (4) We show that the syntactical hierarchical structure of sentence and word structure in English shares some similarities with the hierarchical structure of kanji. Conducting a hierarchical decomposition of kanji, we present the hierarchical structure using tree diagrams and mathematical formulae. We also compare the hierarchical analysis and coding in our study with those of Fujimura (1973), showing the practicality of our study.
- (5) We define an indicator of the structural complexity of kanji, and classify the 2136 Jōyō kanji by their complexity.

Key words: analysis of kanji structure, Jōyō kanji, kanji code, linear decomposition, hierarchical decomposition