

漢字による初級・中級教科書分析

日本語教育項目マトリックス作成に向けて

虎尾 憲史 山元 啓史

要 旨

虎尾・山元(1999)での分析に続き、データベース化した日本語教科書の初級15種22冊、中級15種16冊の本文部分に出現する漢字の分析を行った。今回は各教科書の含有異漢字について、15種すべてに共出現するものから1種のみ出現するものまでの、それぞれの個数と各教科書内での割合の分析結果に現われた、各教科書の特性の異同による大まかなグルーピングを試みた。そして、各教科書の日本語能力試験対象漢字の含有率も分析し、3級以下を初級漢字、2級以上を中級以上漢字とした場合の、それぞれの含有比率の異同によるグルーピングも試みた。その結果、先の大まかなグルーピングとほぼ同様の結果となり、データベース分析による各教科書の含有異漢字の種類や重複、個数、出現の様子等の情報に基づく、教科書分類の可能性と各教科書の位置づけを明らかにすることができた。

【キーワード】 教科書データベース 共出現 異漢字 含有率 初級漢字
中級以上漢字 教科書特性

Basic and Intermediate Textbook Analysis by Kanji:

towards a Japanese teaching item matrix derived from textbook database analysis

Torao, Yoshinobu Yamamoto, Hilofumi

Following our analysis in Torao/Yamamoto(1999), we have made an analysis of the kanji in the database of the main text part of 15 kinds of 22 basic Japanese language textbooks and 16 intermediate ones. We analyzed the kanji of each textbook and calculated the number and percentage of kanji, ranging from those appearing in only one kind of textbook to those appearing in all 15 kinds of textbooks. These data also permit a rough grouping of textbooks. We have also analyzed the degree to which each textbook contains the kanji tested in the Japanese Language Proficiency Test (JLPT). We first divided the JLPT kanji into level 4 and 3 kanji (basic level kanji) and level 2 and 1 kanji (intermediate to advanced level kanji.) We then regrouped the textbooks according to the ratio of these two levels of JLPT kanji. The grouping of textbooks was more or less same using both methods, and shows the possibility of grouping and positioning textbooks based on their kanji as derived from the database analysis.

1. はじめに

虎尾・山元(1999)では、次節で述べる本研究の目的に沿った研究の一貫として、及川ほか(1990)の手法に従い日本語教科書データベースを作成し、そのうちの初級・中級がつながったセットとなっている、6種16冊分のデータを対象に、本文に出現する漢字の計量と比較を中心に分析を行った。そして、各教科書の出現異漢字の数と初・中級間の重複関係の実態を明らかにし、教育項目分析の一例を示した。しかしながら、紙幅規定の関係上、それ以外の分析結果については割愛せざるを得なかった。そこで本稿では、その後新たに追加されたものを含めた、現時点で作成済みのすべてのデータベースを対象として、虎尾・山元(1999)では扱わなかった観点からデータの再分析を行った結果をまとめることとする。

2. 研究目的

虎尾・山元(1999)でも述べたが、語学教育の最適化を考える上で、教材に含まれる内容や配列、分量に関する分析はきわめて重要である。これらを具体的かつ定量的に把握し、そのデータに基づいて教育の過程を考察することは、新人の教師のみならず、ベテランといわれる教師にも有益であると思われる。

学習過程においては、ある段階から次の段階に進んで学習を継続する際、前段階での既習項目が次段階にどの程度含まれているかが次段階の難易度を決定する大きな要因の一つとなる。前段階での既習項目が次段階に全く含まれていない場合、前段階の学習が次段階では全く役立たないこととなり、次段階が同じ分野の事柄であっても学習者にとっては全く新しい別のことを学習する段階となって、既習項目を活用した学習の効率化は望めなくなるばかりか、既習項目の定着を弱め忘却を早めることとなる。従って、各教育機関においては、段階間の学習項目の重複度をしっかり把握しておくことが必要となる。そして、どの程度の重複が学習に最適かについてはまだまだ十分に明らかにされていないのが現状ではあるが、各機関ごとに検討して設定し、教育/学習の効率化を図っていく必要がある。このような作業において大量の教材を対象とした分析を行う場合、省力化、迅速性、多用性等の観点から、対象教材のデータベース化とコンピュータによる分析が必須となってくる。

ある機関における教育対象項目から在籍学習者の既習項目を除いた部分が、その学習者へのその時点以降の教育必要項目となるわけであり、その機関の教育対象項目がデータベース化され、学習者の既習項目もそうならいれば、その学習者に対する教育必要項目はコンピュータ分析により瞬時にリストアップすることが可能となる。そして、その機関でゼロから学習を始めた学習者の、ある時点でのテストの結果から再教育/要復習項目を割り出したり、学習者が教育/学習過程のどこに位置しているか等も割り出せることとなる。テストもコンピュータ化されていれば、この作業は更に高速化できるであろう。同様に、その機関が既習学習者を新たに迎え入れる場合などには、教員、学習者双方にとって強力な味方となり、彼らがこれまでに使用した教科書の教育項目と受入コース

の教育項目を比較照合し、欠習項目を提示/教育することで、コースへの適応をより迅速化するなど、当該機関における教育/学習を効率化させることも可能となるであろう。

そこで本研究では、日本語教科書を入手可能なものから順次データベース化していき、その内容、配列、分量、各級相互の含まれ方の観点から定量的分析を行い、教育項目のマトリックスを作成して、その上で上述のような事柄がどの程度正確に行えるかを試みていこうとしている。これが本研究の大目的とするところであり、分析対象とする教科書の内容については、漢字、語彙、文法、表現を扱うべく作業を進めているところである。そして、虎尾・山元(1999)は本研究のその1、本稿はその2という位置づけで、まずは漢字に関する分析を継続しているところである。本稿においては、冒頭にも述べたように現時点で作成済みのすべての教科書データベースを対象として再分析を行い、各教科書の含有異漢字の種類、分量、含まれ方の異同がどのような特性として現われ、それが各教科書をどう位置づけるのかを明らかにすることを試みる。

3. 分析方法

今回の分析に用いた教科書データベースは、虎尾・山元(1999)でのそれと同形式でデータベース化したものであり、対象教科書の本文部分を1文1データとして検索標識を付加した構造のものである。作業も同様に、光学文字読み取り装置(OCR)でコンピュータにデータ対象部分を読み込み、修正・編集を加えていく形をとっている。初級が15種22冊、中級が15種16冊の日本語教科書本文データベースであり、表1には本稿内で用いる略称とともに各教科書名(著作/発行者)を示した。種別と冊数のずれは、分冊になっていても課番号が通し番号になっている教科書はまとめて1種とし、通し番号になっていないものは別種のデータベースとしているためである。下で述べる今回の分析においても、分析対象は各教科書の本文に出現する漢字を対象としており、併せて、毎日新聞の1993年1年分の記事データに出現する漢字を頻度順に並べた新聞漢字データベース、および、国際交流基金ほか(1994)をもとに作成した日本語能力試験対象漢字データベースも分析に用いている。⁽¹⁾

なお、教科書のデータベース化を本文部分に限定したのは、ほとんどの教科書において各課の本文中にその課で学習する事柄が含まれており、それに基づいて練習等の他の部分が構成されていること、本文以外の部分で新出事項が提出されることは少ないこと、教科書によってスタイルが異なり、一つの課に本文、練習、解説等が含まれている形式のものや分冊形式にされているもの、それらの長短や有無等、教科書毎に情報量が異なること、そして、本文部分のない教科書の存在は非常に稀であることなどの理由からである。さらに、今回の分析対象である出現漢字には、ルビを付けるなどして、必ずしも当該漢字を覚えさせる目的で本文内に提出されているわけではないものも含まれていると思われるが、学習者が繰り返し目にする本文内にあり、意欲的に覚える学習者もいるであろうし、覚えるまでには至らなくても当該漢字の認識力が多少はつくものと思われる。また、当該教科書を使用する機関によっては、出現漢字すべてを教育対象とするところもあるであろうし、

表1：分析対象教科書と略称一覧

略称	教科書名(著作/発行者)
初級	JEl Intensive Course in Japanese - Elementary - Part 1, 2 (対外日本語教育振興会)
	JFB 日本語初歩 (国際交流基金)
	WSD 外国学生用日本語教科書 - 初級 - (早稲田大学語学研究所)
	GY1 日本語Ⅰ (国際学友会日本語学校)
	KKS 日本語の基礎 , (海外技術者研修協会)
	BSJ Business Japanese , (日産自動車国際課)
	SK1 生活日本語 (文化庁)
	SK2 生活日本語 (文化庁)
	SFJ Situational Functional Japanese (筑波ランゲージグループ)
	TG1 日本語Ⅰ (東京外国語大学附属日本語学校教材開発研究協議会)
	TGB 初級日本語 (東京外国語大学附属日本語学校)
	BNB 文化初級日本語 , (文化外国語専門学校日本語科)
	SKS 新日本語の基礎 , (海外技術者研修協会)
	OGB Japanese for Today (Gakken)
IMJ An Introduction to Modern Japanese (O. Mizutani, N. Mizutani)	
中級	OGI INTERMEDIATE JAPANESE VOLUME 1,2 (大阪外国語大学留学生別科)
	Ijp 中級の日本語 (三浦昭, マグロイン花岡直美)
	TG2 日本語 (東京外国語大学附属日本語学校教材開発研究協議会)
	TGI 中級日本語 (東京外国語大学留学生日本語教育センター)
	TKD 日本語 中級 (東海大学留学生別科)
	GY2 日本語 (国際学友会日本語学校)
	JFI1 日本語中級 (国際交流基金)
	JFI2 日本語中級 (国際交流基金)
	ACC ちょっとひとこと (朝日カルチャーセンター [佐々木倫子])
	NBK 日本語でビジネス会話 中級編 本文冊 (日米会話学院日本語研修所)
	BNI1 文化中級日本語 (文化外国語専門学校)
	BNI2 文化中級日本語 (文化外国語専門学校)
	CKN 中級から学ぶ日本語 (荒井, 太田, 大藪, 亀田, 木川, 長田, 松田)
	SGJ 総合日本語中級 (水谷信子)
IJ3 日本語中級 J301 - 基礎から中級へ - 英語版 (土岐, 関, 平高, 新内, 鶴尾)	

教科書内では教育対象とされている漢字でも取捨選択して教育するところもあるものと思われる。最終的に漢字を覚えるか否かは個々の学習者によるところ大なわけであるから、本文内に提出され、覚える機会が提供されている漢字は、即ち当該教科書を使用した学習過程で覚えることが可能な、

当該教科書の漢字として分析対象とすることとした。

今回の分析には、山元作成のrtn2.pl、k_appear.pl、ksf(kanji statistic filter)、krf(kanji reference filter)の各プログラムを用いた。⁽²⁾手順としては、まずrtn2.plによって新聞漢字データベースと教科書データベース中の漢字を照合し、各教科書別に各漢字の出現頻度を計算した、表2のもととなるデータファイルを作成。次にこのファイルのデータからk_appear.plによって出現教科書数別に出現異漢字数を計算し、同条件の漢字が各教科書にいくつ出現しているかを表す、表3、表4のもととなるデータファイルを作成した。また、出現漢字の抽出、異なり漢字と同一漢字の抽出、それぞれの字数計算、出現頻度分析が可能なksfによって各教科書ごとの出現漢字データファイルを作成し、次にその各教科書漢字ファイルと日本語能力試験対象漢字データベースとの照合をkrfによって実行し、表6のもととなる両者間における異なり漢字と同一漢字の抽出と字数計算を行った。

4．分析結果と考察

以下では、表2から表7までにまとめた各分析結果をまず概観し、次に結果に現れた各教科書の特性に基づく大まかな分類を試みることによって、それぞれの位置づけがどのようになるかを明らかにしていく。

4．1．出現漢字の頻度と個数分析

表2は、初級・中級教科書の漢字別出現頻度分析結果がどのようなものかを示すため、初級のその途中を省略して一覧表にしたものである。途中省略無し、表2のもととなる分析結果は1236行、中級のそれは1915行に及ぶ長大なデータとなっており、即ち、初級教科書15種に出現した異漢字総数が1236字、中級15種では1915字であることがわかってはいるが、本稿では全部は紹介しきれない長さのものであるため、全体の紹介は機会を改めることとする。しかし、表2の見方については説明しておく。総出現数の行の各列の数値は、3行目に略称を記した各教科書のデータ中に出現した漢字の総数を示し、異漢字数の行には出現漢字のうちの異なり漢字の数を示した。3行目のApより下の15から1までの数字は、その右欄のKjより下の各漢字が出現した教科書の種数を示し、Kjの右以降の各列の教科書名の下の数値は、各教科書データに出現したKj欄の漢字の個数を示している。また、表2中の は、データ表示の途中省略を示したものである。

表 2：初級教科書における出現書種数と漢字別の漢字出現頻度一覧

総出現数		2180	2453	4893	4427	3600	2869	5360	2846	2561	3208	2657	6895	4265	3160	1764
異漢字数		291	283	714	570	429	536	568	460	355	278	445	731	531	537	315
Ap	Kj	JEI	JFB	WSD	GY1	KKS	BSJ	SK1	SK2	SFJ	TG1	TGB	BNB	SKS	OGB	IMJ
15	日	78	165	107	105	79	47	57	34	21	164	95	174	105	88	30
15	一	51	47	54	39	34	24	22	22	5	77	29	39	9	31	14
15	大	27	23	34	40	40	101	29	24	24	22	46	58	34	28	18
15	人	52	79	61	66	69	12	239	156	30	88	83	85	46	69	30
14	長	6	3	8	4	2	11	14	3	0	4	3	7	14	2	5
14	自	6	0	6	12	16	6	15	7	5	10	8	27	6	8	6
14	合	6	0	8	11	2	16	12	6	7	3	4	19	8	7	1
14	部	3	1	17	61	12	12	8	8	4	6	3	71	17	4	0
13	二	17	35	62	24	0	9	3	12	2	45	23	7	0	19	21
13	三	18	35	40	15	1	4	10	1	0	37	14	22	0	17	12
13	業	4	0	9	9	0	15	6	3	4	4	9	27	3	5	3
13	立	2	3	5	7	2	7	4	0	2	4	3	6	8	6	0
12	同	1	2	1	3	3	0	2	0	2	6	1	5	4	2	0
12	五	11	28	21	13	1	6	1	0	0	44	10	4	0	14	9
12	九	8	20	16	15	0	3	0	2	0	21	2	6	1	4	5
12	発	5	6	7	3	1	1	0	0	0	3	3	9	5	3	1
11	十	33	65	62	23	0	9	0	1	0	112	29	12	0	15	30
11	回	3	0	3	0	4	1	8	7	0	5	1	6	3	4	0
11	員	0	1	11	10	0	22	102	50	61	0	0	61	27	3	7
11	八	5	18	12	7	0	2	2	0	0	26	5	2	0	7	8
10	四	8	26	24	12	0	0	2	0	0	35	5	7	0	7	4
10	開	0	0	1	3	2	3	5	5	1	0	0	14	2	3	0
10	七	5	15	13	5	0	1	0	0	0	16	5	5	0	8	5
10	関	0	0	6	1	0	1	6	0	1	0	1	15	2	1	1
9	連	0	0	7	0	0	10	11	1	1	0	1	7	8	1	0
9	相	0	0	5	1	0	15	12	1	2	0	2	3	1	0	0
9	百	3	7	6	2	0	3	2	0	0	15	11	0	0	5	0
9	決	0	0	9	1	0	2	2	2	2	1	0	11	1	0	0
8	的	0	0	7	1	0	9	0	0	2	0	1	8	0	1	1

8	力	1	0	3	1	0	2	0	1	0	0	1	3	0	1	0
8	法	0	2	1	0	0	1	2	5	0	0	3	1	1	0	0
8	化	0	0	4	0	1	2	0	1	0	5	2	5	0	5	0
7	对	0	0	4	0	0	4	3	1	0	0	1	7	1	0	0
7	米	0	0	2	0	1	2	0	0	0	1	1	0	1	0	1
7	期	1	0	0	0	0	7	1	0	0	0	1	2	0	1	1
7	要	0	0	3	0	0	7	2	1	0	0	0	6	2	2	0
6	市	0	0	5	0	0	1	4	0	0	0	2	3	0	2	0
6	議	0	0	2	0	2	3	0	0	0	0	3	0	7	4	0
6	表	0	0	3	2	0	2	0	0	0	0	2	2	7	0	0
6	治	0	0	2	2	1	0	0	1	0	0	0	3	0	2	0
5	選	0	5	7	0	0	0	0	1	0	0	0	2	1	0	0
5	制	0	0	0	0	0	2	0	0	0	0	3	1	0	1	1
5	島	0	0	2	0	3	0	2	0	0	0	0	11	0	1	0
5	支	3	0	0	0	0	5	4	2	0	0	0	0	0	1	0
4	政	0	0	2	0	0	0	0	0	0	0	1	1	0	2	0
4	氏	0	0	2	0	0	1	1	0	0	0	1	0	0	0	0
4	性	0	0	0	0	0	2	0	0	0	0	1	2	0	1	0
4	界	0	0	1	0	0	1	0	0	0	0	0	1	0	2	0
3	民	0	0	1	0	0	0	0	0	0	0	8	0	0	2	0
3	戰	0	0	0	0	0	0	0	0	0	0	1	2	0	4	0
3	県	1	0	3	0	0	0	2	0	0	0	0	0	0	0	0
3	区	0	0	0	0	0	1	2	0	0	0	0	8	0	0	0
2	勝	0	0	3	0	0	0	0	0	0	0	0	7	0	0	0
2	総	0	0	0	0	0	0	0	0	2	0	0	4	0	0	0
2	改	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
2	派	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0
1	委	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	判	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
1	認	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	策	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

表3：初級教科書間共出現書種別漢字数（上段）、および各教科書内比率（下段）一覧

書種 異数	JEI 291	JFB 283	WSD 714	GY1 570	KKS 429	BSJ 536	SK1 568	SK2 460	SFJ 355	TG1 278	TGB 445	BNB 731	SKS 531	OGB 537	IMJ 315
15 73	73 25.09	73 25.80	73 10.22	73 12.81	73 17.02	73 13.62	73 12.85	73 15.87	73 20.56	73 26.26	73 16.40	73 9.99	73 13.75	73 13.59	73 23.17
14 37	34 11.68	32 11.31	37 5.18	36 6.32	35 8.16	32 5.97	35 6.16	30 6.52	34 9.58	34 12.23	36 8.09	37 5.06	37 6.97	37 6.89	32 10.16
13 45	35 12.03	36 12.72	45 6.30	43 7.54	40 9.32	41 7.65	41 7.22	35 7.61	33 9.30	38 13.67	41 9.21	45 6.16	38 7.16	43 8.01	31 9.84
12 43	29 9.97	29 10.25	42 5.88	39 6.84	38 8.86	36 6.72	34 5.99	26 5.65	26 7.32	32 11.51	39 8.76	43 5.88	37 6.97	35 6.52	31 9.84
11 46	26 8.93	29 10.25	46 6.44	41 7.19	30 6.99	35 6.53	37 6.51	29 6.30	27 7.61	21 7.55	36 8.09	45 6.16	36 6.78	42 7.82	26 8.25
10 45	16 5.50	22 7.77	43 6.02	42 7.37	30 6.99	34 6.34	35 6.16	21 4.57	28 7.89	13 4.68	32 7.19	41 5.61	39 7.34	36 6.70	18 5.71
9 35	9 3.09	14 4.95	34 4.76	25 4.39	26 6.06	25 4.66	24 4.23	16 3.48	18 5.07	11 3.96	22 4.94	29 3.97	27 5.08	26 4.84	9 2.86
8 71	13 4.47	15 5.30	62 8.68	51 8.95	37 8.62	43 8.02	50 8.80	38 8.26	29 8.17	21 7.55	39 8.76	57 7.80	54 10.17	43 8.01	16 5.08
7 64	13 4.47	10 3.53	54 7.56	37 6.49	34 7.93	31 5.78	30 5.28	25 5.43	24 6.76	10 3.60	26 5.84	51 6.98	48 9.04	34 6.33	21 6.67
6 61	11 3.78	5 1.77	48 6.72	36 6.31	21 4.90	35 6.53	28 4.93	24 5.22	9 2.54	5 1.80	25 5.61	43 5.88	32 6.03	32 5.96	12 3.81
5 76	12 4.12	7 2.47	45 6.30	33 5.79	20 4.66	32 5.97	37 6.51	30 6.52	10 2.82	6 2.16	24 5.39	54 7.38	36 6.78	21 3.91	13 4.13
4 90	7 2.41	6 2.12	45 6.30	30 5.26	18 4.20	35 6.53	38 6.69	23 5.00	10 2.82	7 2.52	23 5.17	54 7.39	27 5.08	27 5.03	10 3.17
3 115	7 2.41	2 0.71	48 6.72	38 6.67	9 2.10	32 5.97	36 6.34	30 6.52	14 3.94	2 0.72	17 3.82	53 7.25	21 3.95	30 5.59	6 1.90
2 156	4 1.37	1 0.35	49 6.86	32 5.61	12 2.80	32 5.97	34 5.99	24 5.22	10 2.82	3 1.08	6 1.35	54 7.39	16 3.01	24 4.47	11 3.49
1 279	2 0.69	2 0.71	43 6.02	14 2.46	6 1.40	20 3.73	36 6.34	36 7.83	10 2.82	2 0.72	6 1.35	52 7.11	10 1.88	34 6.33	6 1.90

表4：中級教科書間共出現書種別漢字数（上段）、および各教科書内比率（下段）一覧

書種 異数	OGI 733	Ijp 577	TG2 1203	TGI 1115	TKD 872	GY2 1404	JEI1 418	JEI2 1166	ACC 782	NBK 627	BNI1 695	BNI2 1047	CKN 872	SGJ 717	IJ3 456
15 136	136 18.55	136 23.57	136 11.31	136 12.20	136 15.60	136 9.69	136 32.54	136 11.66	136 17.39	136 21.69	136 19.57	136 12.99	136 15.60	136 18.97	136 29.82
14 102	99 13.51	90 15.60	102 8.48	102 9.15	99 11.35	102 7.26	77 18.42	102 8.75	97 12.40	96 15.31	98 14.10	101 9.65	101 11.58	100 13.95	62 13.60
13 70	63 8.59	58 10.05	70 5.82	65 5.83	62 7.11	70 4.99	33 7.89	66 5.66	64 8.18	62 9.89	63 9.06	67 6.40	68 7.80	61 8.51	38 8.33
12 90	70 9.55	56 9.71	90 7.48	85 7.62	81 9.29	87 6.20	42 10.05	83 7.12	73 9.34	60 9.57	72 10.36	82 7.83	85 9.75	69 9.62	45 9.87
11 81	64 8.73	42 7.28	80 6.65	77 6.91	71 8.14	79 5.63	23 5.50	73 6.26	59 7.54	47 7.50	61 8.78	67 6.40	68 7.80	55 7.67	25 5.48
10 90	55 7.50	34 5.89	87 7.23	83 7.44	63 7.22	87 6.20	26 6.22	82 7.03	57 7.29	42 6.70	52 7.48	77 7.35	73 8.37	57 7.95	25 5.48
9 88	44 6.00	32 5.55	84 6.98	76 6.82	52 5.96	86 6.13	24 5.74	78 6.69	53 6.78	36 5.74	47 6.76	60 5.73	59 6.77	43 6.00	18 3.95
8 106	50 6.82	29 5.03	95 7.90	83 7.44	68 7.80	98 6.98	16 3.83	88 7.55	44 5.63	41 6.54	38 5.47	75 7.16	62 7.11	37 5.16	24 5.26
7 114	43 5.87	25 4.33	83 6.90	83 7.44	53 6.08	101 7.19	11 2.63	90 7.72	46 5.88	27 4.31	34 4.89	78 7.45	59 6.77	45 6.28	20 4.39
6 125	29 3.96	21 3.64	98 8.15	86 7.71	52 5.96	106 7.55	10 2.39	84 7.20	46 5.88	21 3.35	31 4.46	67 6.40	53 6.08	34 4.74	12 2.63
5 109	26 3.55	18 3.12	75 6.23	66 5.92	39 4.47	71 5.06	12 2.87	62 5.32	29 3.71	14 2.23	22 3.17	42 4.01	36 4.13	19 2.65	14 3.07
4 132	15 2.05	13 2.25	65 5.40	55 4.93	29 3.33	90 6.41	3 0.72	73 6.26	33 4.22	16 2.55	19 2.73	52 4.97	26 2.98	26 3.63	13 2.85
3 164	18 2.46	7 1.21	60 4.99	63 5.65	30 3.44	109 7.76	4 0.96	52 4.46	19 2.43	13 2.07	8 1.15	59 5.64	23 2.64	17 2.37	10 2.19
2 174	9 1.23	5 0.87	42 3.49	30 2.69	21 2.41	90 6.41	1 0.24	49 4.20	13 1.66	9 1.44	5 0.72	50 4.78	8 0.92	8 1.12	8 1.75
1 334	12 1.64	11 1.91	36 2.99	25 2.24	16 1.83	92 6.55	0 0.00	48 4.12	13 1.66	7 1.12	9 1.29	34 3.25	15 1.72	10 1.39	6 1.32

表3は、初級教科書15種について、表2のもととなっているデータから各出現異漢字そのものと出現頻度の表示を省き、各教科書別、出現教科書数別に、共通して出現した異漢字の個数を求め、それの各教科書内の割合を付したものである。表4は、中級教科書15種について、表3同様に作成したものである。表3と表4の見方を説明しておく、1行目の上段は教科書名、下段は含有異漢字数を示しており、2行目以降については、最左列の上段は出現教科書種数、下段は上段の書種数に共出現する各教科書ごとの異漢字数について、15種全体で見て重複を排した出現異漢字数を示している。第2列目以降の上段は、各教科書含有異漢字のうちで最左列上段の出現教科書種数に共出現した異漢字数であり、下段は上段の異漢字数が各教科書の含有異漢字数に占める割合を示している。

表3、表4を見ると、15種の初級教科書すべてに共出現した異漢字数は73字、中級教科書15種では136字という結果である。虎尾・山元(1999)での6種の教科書分析の結果に比して、各教科書に共出現する異漢字数に大幅な減少が見られる。初級は6種分析時の158字から半数以下に、中級は6種時の510字から4分の1近くにまで減少している。分析対象が増えれば共通部分が減っていくのは当然であり、ある程度の減少は予想していたが、分析対象を6種から15種に増やしただけでこれほど減少してしまうとは予想外であった。ともあれ、各教科書の作成者ら全員が、結果的に初級漢字、中級漢字として、共通に認識していたと判断できる漢字の数が73字と136字という結果は、彼らの漢字の級別意識の共通性が極めて希薄であったことを示していると言えよう。

そして、すべてに共通という観点からだけでなく、複数に共通という観点に目を転じた場合はどうなるかについての結果も、ここで表5に示しておく。2種以上欄にある957、1581という数値は、15種の教科書作成者のうち、少なくとも2種以上の教科書作成者間に共通して、初級漢字、中級漢字と判断されたと言える漢字の数である。これは、表3、表4の1種にのみ出現する異漢字数279字と334字を異漢字総数の1236字と1915字から引いた数であり、以降同様の計算で、表5には2種以上から15種までの各場合の初級、中級の異漢字数をまとめた。複数に共通という観点の場合、すべてに共通という観点の場合と反対に、分析対象の増加に比例して異漢字数も増加するが、必然的に共通性の内容も異なってくる。いくら複数とは言え、2種や3種程度では偶然性が高まり、信頼性も低くなる。しかし、何種以上をもって信頼できる数とするかの判定はここでの目的ではないので、表5によって別の観点からの分析結果をも示しておくにとどめる。

表5：初級、中級教科書の種数別出現異漢字総数一覧

	2種以上	3種	4種	5種	6種	7種	8種	9種	10種	11種	12種	13種	14種	15種
初級	957	801	686	596	520	459	395	324	289	244	198	155	110	73
中級	1581	1407	1243	1111	1002	877	763	657	569	479	398	308	238	136

4.2. 出現異漢字の教科書内比率分析

次に、表3、表4の下段に示した、各教科書内での割合に目を向けると、共出現する教科書数が減るにつれてほぼ段階的に割合も減っていくものと、そうではないものがあることがわかる。大まかに分けると、表3ではJEL、JFB、KKS、TG1、TGB、SKS、IMJが段階的、WSD、GY1、BSJ、SK1、SK2、SFJ、BNB、OGBが非段階的となり、表4ではTG1、TGI、GY2、JFI2、BNI2以外は段階的タイプに分けられよう。非段階的タイプの教科書には含有異漢字数の多いものが多いが、すべてがそうではなく、必ずしも含有異漢字数の多少と比例関係にあるとは言いきれないようである。表3においては、上に出現し共出現教科書数が多い漢字ほど初級漢字であり、下に現れて共出現教科書数が少ない漢字ほど初級から離れて中級漢字に近づいた漢字であると考えられる。そうすると、含有異漢字数がそれほど多くない教科書で、共出現教科書数が減るにつれて共出現異漢字数も段階的に減っていく様子がはっきりとしているものほど、初級漢字を多く含む、漢字面では典型的な初級教科書と見ることができると思われる。このような基準で表3の教科書を順に並べてみると、概ねTG1、JFB、JEL、IMJ、KKS、TGB、SKS、SFJ、SK2、OGB、GY1、BSJ、SK1、WSD、BNBの順となる。そして、TG1からSKSまでは典型的初級教科書の範囲内にあり、SFJからSK1までは含有異漢字数の割には中級漢字を多く含み、WSDとBNBは含有異漢字数の多さに比例して中級漢字も多く含む教科書であるという分類が可能となる。ここで、中級教科書の分類に進む前に、中級教科書中にも初級教科書に出現した漢字が重複して出現することを考慮しておかなければならない。krfによる分析では、初級15種1236字と中級15種1915字の重複字数は1193字という結果が出ており、初級側から見ると96.5%、中級側から見ると62.3%が重複している。これを考慮して中級教科書の分類を考えると、表4においては、上に出現し共出現教科書数が多い漢字ほど初級漢字と重複することが多いものであり、下に現れて共出現教科書数が少ない漢字ほど初級とは重複しない中級以上の漢字であると考えられる。そうすると、表4の場合は、概ねJFI1、IJ3、IJp、NBK、BNI1、SGJ、OGI、ACC、CKN、TKD、TGI、BNI2、TG2、JFI2、GY2の順となり、JFI1が漢字面では初級教科書に近く、IJ3からTKDまでは典型的な中級教科書の範囲内にあり、TGIからGY2は含有異漢字数の多さに比例して中級以上の漢字を多く含む教科書であると分類することができよう。

4.3. 出現漢字と能力試験漢字の分析

さて、表6は、各教科書別に、日本語能力試験対象漢字の級別含有数と級別漢字数に占める割合（カバー率）、および教科書含有異漢字数に占める割合、そして、試験対象外漢字数と教科書含有異漢字数に占める割合を示したものである。

表6：教科書別日本語能力試験対象漢字の含有数とカバー率
および教科書含有異漢字数に占める割合一覧

	書名(異数)	4級(/80)/(異数)	3級(/165)/(異数)	2級(/755)/(異数)	1級(/1036)/(異数)	外(/異数)	
初級	JEI (291)	78(97.50)(26.80)	123(74.55)(42.27)	79(10.46)(27.15)	9(0.87)(3.09)	2(0.69)	
	JFB (283)	78(97.50)(27.56)	115(69.70)(40.64)	86(11.39)(30.39)	3(0.29)(1.06)	1(0.35)	
	WSD (714)	80(100.0)(11.20)	160(96.97)(22.41)	362(47.95)(50.70)	106(10.23)(14.85)	6(0.84)	
	GY1 (570)	76(95.00)(13.33)	158(95.76)(27.72)	270(35.76)(47.37)	62(5.98)(10.88)	4(0.70)	
	KKS (429)	66(82.50)(15.38)	134(81.21)(31.24)	200(26.49)(46.62)	25(2.41)(5.83)	4(0.93)	
	BSJ (536)	72(90.00)(13.43)	127(76.97)(23.69)	270(35.76)(50.37)	63(6.08)(11.75)	4(0.75)	
	SK1 (568)	71(88.75)(12.50)	126(76.36)(22.18)	284(37.62)(50.00)	81(7.82)(14.26)	6(1.06)	
	SK2 (460)	62(77.50)(13.48)	100(60.61)(21.74)	235(31.13)(51.09)	58(5.60)(12.61)	5(1.09)	
	SFJ (355)	60(75.00)(16.90)	102(61.82)(28.73)	161(21.32)(45.35)	32(3.09)(9.01)	0(0.00)	
	TG1 (278)	78(97.50)(28.06)	126(76.36)(45.32)	71(9.40)(25.54)	3(0.29)(1.08)	0(0.00)	
級	TGB (445)	80(100.0)(17.98)	132(80.00)(29.66)	217(28.74)(48.76)	15(1.45)(3.37)	1(0.22)	
	BNB (731)	79(98.75)(10.81)	148(89.70)(20.25)	385(50.99)(52.67)	105(10.14)(14.36)	14(1.92)	
	SKS (531)	63(78.75)(11.86)	133(80.61)(25.05)	273(36.16)(51.41)	60(5.79)(11.30)	2(0.38)	
	OGB (537)	74(92.50)(13.78)	150(90.91)(27.93)	246(32.58)(45.81)	58(5.60)(10.80)	9(1.68)	
	IMJ (315)	73(91.25)(23.17)	99(60.00)(31.43)	116(15.36)(36.83)	22(2.12)(6.98)	5(1.59)	
	中級	OGI (733)	78(97.50)(10.64)	152(92.12)(20.74)	368(48.74)(50.20)	117(11.29)(15.96)	18(2.46)
		IJp (577)	76(95.00)(13.17)	139(84.24)(24.09)	278(36.82)(48.18)	82(7.92)(14.21)	2(0.35)
		TG2 (1203)	80(100.0)(6.65)	165(100.0)(13.72)	643(85.17)(53.45)	289(27.90)(24.02)	26(2.16)
		TGI (1115)	78(97.50)(7.00)	158(95.76)(14.17)	597(79.07)(53.54)	264(25.48)(23.68)	18(1.61)
		TKD (872)	80(100.0)(9.17)	154(93.33)(17.66)	450(59.60)(51.61)	169(16.31)(19.38)	19(2.18)
GY2 (1404)		80(100.0)(5.70)	160(96.97)(11.40)	667(88.34)(47.51)	435(41.99)(30.98)	62(4.42)	
JFI1 (418)		63(78.75)(15.07)	120(72.73)(28.71)	199(26.36)(47.61)	35(3.38)(8.37)	1(0.24)	
JFI2 (1166)		74(92.50)(6.35)	151(91.52)(12.95)	584(77.35)(50.09)	322(31.08)(27.62)	35(3.00)	
ACC (782)		75(93.75)(9.59)	147(89.09)(18.80)	412(54.57)(52.69)	139(13.42)(17.77)	9(1.15)	
NBK (627)		76(95.00)(12.12)	137(83.03)(21.85)	310(41.06)(49.44)	93(8.98)(14.83)	11(1.75)	
級	BNI1 (695)	70(87.50)(10.07)	145(87.88)(20.86)	363(48.08)(52.23)	110(10.62)(15.83)	7(1.01)	
	BNI2 (1047)	78(97.50)(7.45)	152(92.12)(14.52)	494(65.43)(47.18)	288(27.80)(27.51)	35(3.34)	
	CKN (872)	72(90.00)(8.26)	155(93.94)(17.78)	489(64.77)(56.08)	153(14.77)(17.55)	3(0.34)	
	SGJ (717)	69(86.25)(9.62)	146(88.48)(20.36)	374(49.54)(52.16)	123(11.87)(17.15)	5(0.70)	
	IJ3 (456)	52(65.00)(11.40)	104(63.03)(22.81)	220(29.14)(48.25)	69(6.66)(15.13)	11(2.41)	

まずカバー率を見ると、初級教科書については、4級漢字は75～100%の範囲内であって十分な率であると思われるが、含有漢字数がさほど少なくはないSFJ、SK2、SKSのカバー率が70%代で、含

有漢字数の少ないJEL、JFB、TG1のほうが100%に近いカバー率を示している点が目につく。3級漢字は60～96.97%とやや範囲が広がってはいるものの、まずまずの値を示していると言えそうだが、4級同様SK2、SFJがIMJとともに低い値に留まっており、他は高率を維持している。2級漢字は9.4～50.99%と率がぐんと下がり幅も広がっているが、ここからは含有漢字数の少ないJEL、JFB、TG1、IMJが低率で、含有漢字数の多いWSDとBNBが高率となっている。1級漢字では0.29～10.23%と非常に低いカバー率で、特にJEL、JFB、TG1は極端に低率であるが、SK1、SK2、SFJは含有漢字数の割には高率を示している。これが初級15種の実態であり、各教科書の特性がそれなりに現れていると思われる。特性を示すカバー率の差異は作成者のコントロールによるところが大なのではないかと思われるが、それが無いから現れた差異ではないかと思われる教科書もあり、この点の判断には更に分析を重ねる必要がある。

中級教科書については、4級漢字は65～100%、3級漢字は63.03～100%と同様のカバー率となっており、ともにIJ3とJFI1の低率が目立っている。中級でも2級漢字になると26.36～88.34%と広範囲にばらついてはいるが、カバー率の差は含有漢字数の差にほぼ比例しており、健全なばらつきであると言える。しかし、1級漢字の3.38～41.99%のばらつきはそうでもない。JFI1、IJ3、Ijp、NBKの低率とGY2の高率が目立つが、やはりJFI1の低率は含有漢字数の差を考慮しても特に際立っている。中級15種ではJFI1とIJ3が全般的に低いカバー率を示し、それにIjp、NBKが続く形となっているが、全体的に初級15種よりは他と異なる特性を示す教科書は多くないように思われる。

表7：能力試験漢字含有率 3級以下：2級以上対照表

分 類	比 率	初級	3以下	2以上	中級	3以下	2以上	比 率	分 類
初級レベル	7 : 3	TG1	73.38	26.62	JFI1	43.78	56.22	4.5 : 5.5	弱く中級寄り
		JEL	69.07	30.93	Ijp	37.26	62.74		
		JFB	68.20	31.80	IJ3	34.21	65.79	3.5 : 6.5	強く中級寄り
弱く初級寄り	5.5 : 4.5	IMJ	54.60	45.40	NBK	33.97	66.02		
弱く中級寄り	4.5 : 5.5	TGB	47.64	52.35	OGI	31.38	68.62	3 : 7	中級レベル
		KKS	46.62	53.38	BNI1	30.93	69.07		
		SFJ	45.63	54.36	SGJ	29.98	70.01		
中位中級寄り	4 : 6	OGB	41.71	58.29	ACC	28.39	71.61	2.5 : 7.5	弱く上級寄り
		GY1	41.05	58.95	TKD	26.83	73.17		
強く中級寄り	3.5 : 6.5	BSJ	37.12	62.87	CKN	26.04	73.97	2 : 8	中位上級寄り
		SKS	36.91	63.09	BNI2	21.97	78.03		
		SK2	35.22	64.79	TGI	21.17	78.83		
		SK1	34.68	65.32	TG2	20.37	79.63		
		WSD	33.61	66.39	JFI2	19.30	80.71		
中級レベル	3 : 7	BNB	31.06	68.95	GY2	17.10	82.91	1.5 : 8.5	強く上級寄り

次に、日本語能力試験対象漢字の3級以下を初級漢字、2級以上を中級以上の漢字として分けて考えた場合、それらの含有率の差異によって各教科書の位置づけがどのようになるかを見てみる。表7は、初級、中級の教科書別に3級以下と2級以上の含有率を表6から求めて降順にソートした結果に、比率と分類を付加したものである。比率は、大まかに0.5ポイント区切りで表すこととした。分類は、7対3の比率で初級漢字の含有率が高いTG1、JEL、JFBの3種が第1グループとなっていることから、これを起点に初級レベルとして、以降0.5ポイントごとに6.5対3.5を強く初級寄り、6対4を中位初級寄り、5.5対4.5を弱く初級寄り、5対5を中間レベル、4.5対5.5を弱く中級寄り、4対6を中位中級寄り、3.5対6.5を強く中級寄り、そして、初級レベルの反対の3対7を中級レベルというふうにグルーピングしていくこととし、表中にはグループごとの区分線も示した。表7からもわかるように、初級教科書15種は、TG1、JEL、JFBが初級レベル、IMJが弱く初級寄り、TGB、KKS、SFJが弱く中級寄り、OGB、GY1が中位中級寄り、BSJ、SKS、SK2、SK1、WSDが強く中級寄り、BNBが中級レベルという6つのグループに分類できる。中級教科書15種は、JFI1が弱く中級寄り、Ijp、IJ3、NBKが強く中級寄り、OGI、BNI1、SGJ、ACCが中級レベル、TKD、CKNが弱く上級寄り、BNI2、TGI、TG2、JFI2は中位上級寄り、GY2が強く上級寄りという6グループに分けることができる。

さて、先に表3、表4を用いて行った大まかな分類では、初級15種はTG1、JFB、JEL、IMJ、KKS、TGB、SKSを典型的初級教科書の範囲内にあるものとし、SFJ、SK2、OGB、GY1、BSJ、SK1は含有異漢字数の割には中級漢字を多く含むもの、WSD、BNBは含有異漢字数の多さに比例して中級漢字も多く含む教科書とした。これを表7での分類と比較すると、表7のほうが細分化されてはいるものの、SKSとSK2を除いてほぼ同様の順番で上から並んでいる。SK2は順序が異なってはいても中級寄りの分類であるが、SKSは典型的初級教科書の範囲内グループと強く中級寄りのグループに分かれている。分類が異なった原因は、表3の8種と9種共出現字数の教科書内比率が高いことが影響したものである。

中級15種については、JFI1を初級教科書に近いものとし、IJ3、Ijp、NBK、BNI1、SGJ、OGI、ACC、CKN、TKDは典型的中級教科書の範囲内にあるもの、TGI、BNI2、TG2、JFI2、GY2は含有異漢字数の多さに比例して中級以上の漢字を多く含む教科書という分類を表4から行った。表7の分類と比較すると、いくつか順序が異なるものはあるが、表7でのグルーピングは同じであり、同様の分類ができていたものと思われる。

5. おわりに

以上、初級15種、中級15種とも、各教科書の含有異漢字の種類や重複、個数、出現の様子等の情報に基づいて、表3、表4の分析からは3グループずつに、表6、表7の分析からは6グループずつに含有漢字面からの教科書分類が可能となり、各教科書の位置づけを明らかにすることができた。しかし、今回の分析は、教科書の分類自体を目的としたものではなく、教科書データベースの作成で可能となる様々な分析による日本語教育項目マトリックス作成の一貫として、各教科書の漢字情

報分析から得られた特性に従ってそのグループ分けを行ったものである。そして、使用教科書が他書に比してどのような性格を持ち、どのような位置づけにあるかを知った上で用いること、次レベルの教科書との関係等、教師が事前に把握しておくべき重要な事柄が教科書データベース分析によって導き出せることを示したものである。従って、ここではグループ数の問題や分類の精度、どこで級別の線引きをするか等々の問題にまでは立ち入らず、それらの問題は稿を改めて論じることとしたい。

注

- (1) 新聞漢字データベースは、注(2)で説明するksfプログラムによって、1993年1年分の毎日新聞記事データに出現する漢字の抽出と頻度計算を行い、最も出現頻度の高かった漢字から順に並べたものである。データは1行に漢字1字とその出現頻度順位という構成で、全4211行のデータベースとなっている。日本語能力試験対象漢字データベースは、国際交流基金ほか(1994)の各級別漢字表をもとに作成した。データは1行に漢字1字とその級別という構成で、全2036行のデータベースとなっている。
- (2) rtn2.plは、新聞漢字データベースの各漢字と教科書データベース中の漢字を照合し、各漢字の出現回数の計算結果を各教科書別にリスト表示させるプログラムであり、これによって表2のデータを作成した。k_appear.plは、rtn2.plの出力データから出現教科書数別に出現異漢字数を計算し、各教科書別に同じ条件の漢字出現数をリスト表示するプログラムである。これによって表3、表4の上段のデータを得ている。ksf(kanji statistic filter)は、テキストデータ解析プログラムであり、テキストに出現した異漢字を頻度の高いものから順に出現回数と共にリストし、漢字の出現総数と異漢字数、その他の文字数等の計算結果を出力するものである。本研究では、主に各教科書別の出現異漢字データファイル作成や出現漢字総数、出現異漢字数計算などにksfを用いている。krf(kanji reference filter)は、漢字データファイル照合プログラムであり、2つの漢字データファイルを照合して、一致は1、不一致は0を漢字と共に出力する。本研究では、主に各漢字データファイル間の重複関係の分析に用いている。

参考文献

1. 及川昭文・加納千恵子・藤田正春・吉岡亮衛・他(1990)『パソコンによる外国人のための日本語教育支援システムの開発』日本語教育支援システム研究会
2. 国際交流基金・日本国際教育協会編(1994)『日本語能力試験出題基準』凡人社
3. 虎尾憲史・山元啓史(1999)「初級・中級教材による日本語教育項目マトリックス:(1)漢字」『日本語教育方法研究会誌』Vol.6, No.1:14-15

