

小学校国語教科書に掲載されている単語の分析

——ラディカルを共有する漢字から構成される単語の
ファミリーサイズと出現頻度¹——

小河 妙子・藤田知加子・増田 尚史

(受付 2018年10月10日)

1. はじめに

言語に関する知識が人間の心の中にどのように表象され、これらの知識がどのように利用されているのかを解明することは、心理学的研究における従来からの重要なテーマである。その中でも、人間の意味処理過程を明らかにするために、視覚的に呈示された単語を読む認知処理が注目されてきた。単語の視覚的な認知過程は、形態・音韻・意味などの語彙知識の集合体である心的辞書 (mental lexicon) を検索し、入力刺激と一致する単語の意味を理解する過程と定義される。

意味処理に関する従来の見解は、主に意味や概念に関する記憶研究から得られてきた。その代表的な古典的モデルとして、意味ネットワーク・モデルがある (Collins & Loftus, 1975; Collins & Quillian, 1969)。これらのモデルは、ネットワークを構成する基本単位である個々の概念に対応するノードと、これらのノード間を連結するリンクから構成される。Collins & Loftus (1975) による活性化拡散モデル (spreading activation model) では、各々の概念はそれぞれ1つのノードで表され、意味的に関連するノード同士がリンクしていると仮定される。これらの概念ノードは意味的関連性が強いほど密接にリンクし、意味的関連性に基づき体制化されたネットワーク構造をなしている。このモデルでは、ある概念が認知される過程においては、その概念自身のノードが活性化されると、その概念と意味的に関連するノードにも活性化が拡散していくと仮定される。このような活性化拡散の枠組みは、以降の研究において、先行刺激の処理が意味的に関連する後続刺激の処理を促進するという意味的プライミング効果に関する多数の研究を牽引してきた (e.g., McNamara, 2012)。

近年では、従来からの意味記憶研究に加えて、インターネットの開発にともなう情報検索技術の著しい進展 (Manning, Raghavan, & Schütze, 2008 岩野・黒川・濱田・村上訳 2012) や、各国における大規模な言語コーパスや単語データベースの開発による影響を受け

1 本研究はJSPS 科研費 (基盤研究 (C) 課題番号16K04434) の助成を受けた。本論文の Summary 執筆にあたり助言を頂きました Terry Joyce 教授 (多摩大学) に心より感謝申し上げます。

(Baayen, Piepenbrock, & Gulikers, 1995; 国立国語研究所, 2011), 学際的な言語処理研究の視点から人間の意味処理が解明されつつある。例えば, 数千語から数万語規模の単語刺激を用いて, 視覚的あるいは聴覚的に提示される単語の語彙判断時間や音読時間などをデータベース化する研究 (megastudy) が英語 (Balota et al., 2007; Sánchez-Gutiérrez, Mailhot, Deacon, & Wilson, 2018), フランス語 (Ferrand et al., 2010), オランダ語 (Keuleers, Diependaele, & Brysbaert, 2010), 中国語 (Tse & Yap, 2018) などにおいて報告されている。

心理学の伝統的な研究手法を用いた研究に加え, より最近では, 言語コーパスに基づき構築される潜在意味解析 (Landauer & Dumais, 1997) などの計算論的アプローチによる意味空間モデルや, 意味情報検索技術としての Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) や WordNet (Fellbaum, 2005) の開発など, 計算機に理解可能な知識として概念を明示的に記述するためのオントロジー工学も急速に進展している。

1.1 漢字を材料とした先行研究

このような流れを受け, 日本語を材料とした単語認知研究においても, コーパス言語学的な手法を取り入れた言語処理研究が進展しつつある (Joyce, Masuda, & Ogawa, 2014; 小河, 2012, 2014; Ogawa & Fujita, 2017; 小河・藤田, 2018)。

日本語は漢字と仮名という二つの表記を併用しているという特徴を持つ (岩田, 1994; Kess & Miyamoto, 1999)。仮名表記語は, 文字の形態がシンプルであり, 形態情報と音韻情報との対応関係が規則的であるために, 読み書きを学習することは比較的容易である。一方, 漢字表記語は, 形態的特徴として, 非常にシンプルな視覚的要素から構成される漢字 (e.g., 「一」, 「人」), 画数の多い複雑な要素から構成される漢字 (e.g., 「鬱」, 「麟」) もある。また, 形態情報と音韻情報との対応関係も不規則的であり, 漢字表記語の読み書きの学習は, 日本語学習者 (日本語を母語とする子どもや他言語を母語とする学習者) にとって難しいと考えられる。

伝統的に, 本邦の小学校における国語教育では, 漢字は形態素文字 (Joyce, 2011) であることを利用し, その漢字の成り立ちに関する知識を手がかりとして読み書きを学習する手法が用いられている (市川, 1963)。例えば, 一つの形態素である「木」は, 植物の姿を表しており, 上部は木の枝葉であり下部は根を表す。形態素とは, それ以上小さな言語学的な単位に分けることができない単純形態であり, もっとも小さい個別な有意味単位である (Miller, 1991)。

形態素である「木」に着目すると, 一文字単語として「木」を表すが, ラディカル (radical: 部首あるいは部品) として漢字を構成する要素 (e.g., 「枝」) にも使用される。さらに, 「木材」, 「並木道」など二字熟語や三字熟語の構成要素としても使用される。このような特性

から、一つの形態素がラディカルや語の要素として、心的辞書に階層的に表現されていると推測される。このような形態素を共有する単漢字の単語間の心的辞書における意味的構造を明らかにするために、小河ら（小河，2012，2014；Ogawa, Fujita, Joyce, Kawakami, & Masuda, 2012）は、形態的に左右あるいは上下に分離可能な単漢字を用いて、タイプ頻度の高いラディカルを対象に、同一のラディカル（e.g., きへん, くさかんむり）を共有する漢字間（e.g., 枝－松, 花－草）の意味的類似性評定課題を大学生に対して実施し、多次元尺度法（multi-dimensional scaling: MDS）による検討を行っている。その結果、各ラディカルについて三次元の軸が抽出され、各ラディカルの辞書の定義に基づく次元に加え、複数のラディカルに共通する次元として、抽象－具象を表す軸やポジティブ－ネガティブを表す軸の存在が示唆されている。

図1は、小河（2014）で報告された「くさかんむり」を共有する単漢字のMDSの結果に、これらの漢字から構成される熟語群の一部を加えたものである。小河らの研究では、同じラディカルを持つ漢字を対象としているが、日本語の漢字表記における特徴として、形態素は単漢字の間だけではなく、同一の漢字を共有する単語（熟語）間にも階層を超えて構造化されていると考えられる。単漢字に加えて、これらの熟語間の結合関係を捉えることで、心的辞書における意味空間の構造を解明する手がかりとなると考えられる。また、このような形態素を中心とする階層的構造は、小学生から大人へと漢字の学習が進むにつれて、語彙発達過程に応じて変化していくと考えられる。しかしながら、このような意味空間の構造に関する発達的变化に着目した研究はなされていない。

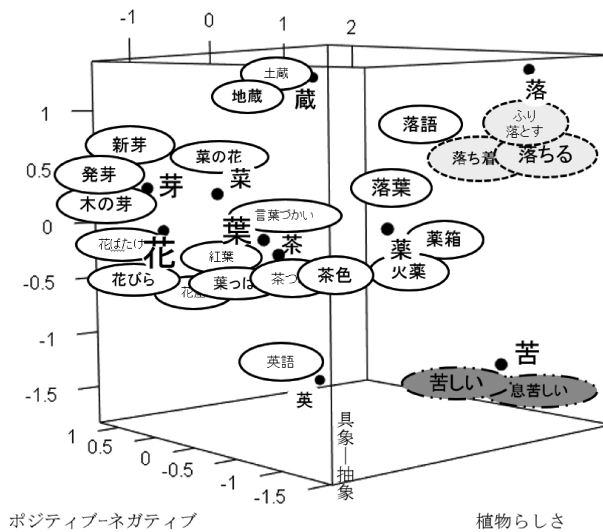


図1 「くさかんむり」ファミリーの意味空間の模式図。●は各漢字、楕円は単語を表す。実線は名詞、点線はそれぞれ動詞と形容詞を表す。

1.2 本研究の目的

本研究の目的は、心的辞書における意味空間の構造に関する発達的变化を検討する際に必要とされる、実験材料を整備することにある。具体的には、形態素を共有する漢字および単語について、小学校の国語教科書をコーパスとし、六学年分の国語教科書に掲載されている全単語（小河・藤田，2018）を対象として、同一ラディカルを含む漢字の数およびタイプ頻度、そしてその漢字から構成される単語ファミリーの成員と成員数および各成員の出現頻度、および単語ファミリー出現頻度を足し合わせた総出現頻度を算出し、一覧表を作成する。

以下に、本研究で用いる用語についての定義を述べる（図2を参照）。まず、「林」「根」「松」など「きへん」という同じラディカルをもつ漢字（radical neighbor）の一群を漢字ファミリー（kanji family）と呼ぶ。次に、同じ漢字を構成要素とする単語群を単語ファミリー（word family）と呼ぶ。例えば、漢字「根」を共有する単語ファミリーの成員には、名詞では「屋根」「根拠」など、動詞では「根ざす」が存在する。これらの各単語ファミリーに含ま

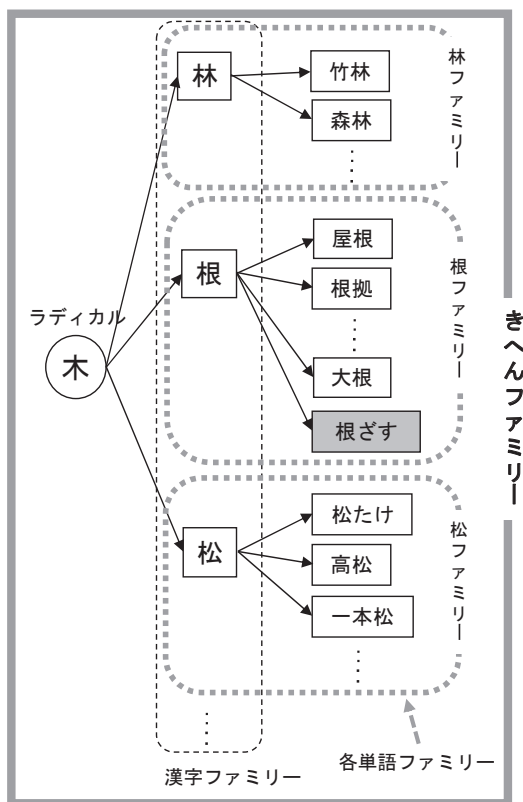


図2 ラディカル「木」を共有する「きへん」ファミリー。
白色の長方形は名詞を表し、グレーの長方形は動詞を表す。

れる単語数を単語ファミリーサイズ (word family size) とし、これらの成員の出現頻度を合計したものを単語ファミリー出現頻度 (frequency of word family) とする。さらに、同じラディカルを共有する複数の漢字の単語ファミリーの集合を、総単語ファミリー (total word family) と呼ぶ。例えば、「林」ファミリー、「根」ファミリー、「松」ファミリーなど、「きへん」を含む漢字を構成要素とする単語ファミリーの全てをまとめて「きへん」ファミリーと呼ぶ。この総単語ファミリーを構成する全単語の出現頻度を合計した値、つまり同一ラディカルを共有する各単語ファミリー出現頻度の合計を、単語ファミリーの総出現頻度 (total frequency of word family) と呼ぶ。ただし、一つの単語が同一ラディカルから構成される漢字を複数個含む場合 (例えば「松林」は、「松」ファミリーと「林」ファミリーに重複して出現する) には、単語ファミリーの各成員のファミリーサイズおよび総出現頻度を算出する際に、これらの重複する単語分の個数と出現頻度を引き算し、同一の単語は一回のみ含まれるように再計算した。

本研究では、小学校教科書で使用される教育漢字1,006字のうち、タイプ頻度の高いラディカルから構成される左右分離タイプおよび上下分離タイプの漢字について、上述の漢字および単語ファミリーのファミリーサイズや出現頻度を整理し、漢字学習の初期段階にある小学生の語彙発達の観点による心的辞書の体制化を明らかにするための言語材料を提供することを目的とする。

2. 方 法

材料の選定 小学生が六年間の国語科教育において学習する教育漢字1,006字の中から、小河 (2012, 2014) で報告された左右分離タイプのラディカル25種類と上下分離タイプのラディカル26種類を対象とした。これらのラディカルは、教育漢字の中で使用されるラディカルのタイプの頻度が高い順に左右分離タイプでは18位まで、上下分離タイプでは17位までを抽出したものである (後述の表1を参照)。

これらの51種類のラディカルから構成される漢字を含む単語を、以下の手順により光村図書小学校国語教科書から抽出した。まず、六学年を通してカウントした書字形の出現頻度表における13,340種類の単語 (小河・藤田, 2018) を対象に、51種類のラディカルを含む漢字から構成される単語の書字形を品詞別 (名詞・動詞・形容詞・形状詞・副詞) に抽出した。さらに、各ラディカルの由来となっている元の漢字を抽出対象として追加した。例えば「にんべん」では「人」を、「さんずい」では「水」を抽出した。

以上の手続きによって抽出された4,153語について、ラディカル毎にこれらの単語の書字形とともに、六学年の教科書における出現頻度、書字形、発音形の情報も抽出した。ただし、

動詞と形容詞は活用形によって出現頻度が異なるため、書字形の異なる同一の単語を一語として扱うために、書字形基本形に置き換えた。例えば、教科書では動詞「考える」は、「考え」・「考える」・「考えよう」・「考えれ」の四種類の書字形で使用されているが、これらをすべて書字形基本形「考える」に置き換えた。そのため、出現頻度はこれらの各書字形の出現頻度を合計した値が書字形基本形の出現頻度となる。この段階で、対象となる単語数は3,693語であった。

次に、登場人物や著者名は、単語ファミリーの一員として意味を担う構成要素として含めるのは妥当ではないと考えたため、固有名詞の人名(388語)を除外した。その一方で、固有名詞の地名(78単語)は出現頻度が高く(例えば「日本」の出現頻度は151回、「富士」は同25回)、単元の物語や説明文の中で重要な意味を担うため、分析対象に含めた。以上の手続きを通じて最終的に、合計3,305語を分析対象とした。

なお、本研究では、意味は同じでも視覚的な形状や布置が異なるラディカルは、別のラディカルとして単語ファミリーをカウントした。例えば、「にんべん」と「ひとがしら」は「人」という意味を表すが、別々のファミリーとしてカウントした。

3. 結果と考察

上記の手順で抽出した3,305語を対象として、同一ラディカルを含む漢字から構成される単語ファミリーの成員、各単語ファミリーのファミリーサイズと出現頻度、および単語ファミリーの総出現頻度を算出し、一覧表を作成した。

3.1 ラディカル毎の単語ファミリーサイズと単語ファミリー出現頻度および総出現頻度

表1は、ラディカル毎の単語ファミリーサイズと単語ファミリーの総出現頻度を示す。例の列には、各ラディカルを含む教育漢字一字を掲載した。漢字数は、教育漢字1,006字のうち、小河(2012, 2014)が報告した当該のラディカルを含む漢字の総数を表す。ラディカル毎に、単語ファミリーサイズおよび単語ファミリーの出現頻度を五つの品詞別に算出し、さらに総出現頻度として五品詞の合計を算出した。

3.2 単語ファミリー成員の一覧表

表2は、「くさかんむり」ファミリーのサイズと出現頻度を例として示した。対象漢字は、「花」から「著」まで16字あり、「くさかんむり」ファミリーの単語ファミリーサイズは147語、総出現頻度は1,545回であった。

表3は、同一ラディカル(例として「くるまへん」を記載)を共有する漢字の単語ファミ

表1 ラディカル毎の総単語ファミリーサイズと単語ファミリーの総出現頻度

順位	部品	例	漢字数	単語ファミリーサイズ						単語ファミリーの総出現頻度					
				合計	名詞	動詞	形容詞	形状詞	副詞	合計	名詞	動詞	形容詞	形状詞	副詞
左右分離タイプ															
1	にんべん	休	43	353	296	53	1	1	2	4,522	3,268	1,236	9	1	8
2	さんずい	池	41	277	220	44	8	3	2	2,056	1,577	403	64	4	8
3	ごんべん	計	32	254	207	46	0	1	0	6,049	3,597	2,449	0	3	0
4	きへん	村	31	194	184	5	0	4	1	1,542	1,501	22	0	18	1
5	いとへん	紙	28	168	134	27	4	2	1	1,564	1,255	263	41	3	2
6	てへん	打	22	173	124	44	1	3	1	1,153	936	188	1	22	6
7	ぎょうにんべん	行	15	113	98	13	1	0	1	1,040	748	290	1	0	1
8	こごとへん	院	13	56	48	6	2	0	0	242	222	18	2	0	0
9	のぎへん	科	12	75	64	10	0	1	0	483	454	28	0	1	0
10	つきへん	服	10	52	50	1	0	1	0	404	397	5	0	2	0
10	ひへん	明	10	141	123	6	5	6	1	1,710	1,567	39	52	47	5
12	かねへん	鉄	9	56	56	0	0	0	0	260	260	0	0	0	0
12	つちへん	地	9	92	89	3	0	0	0	446	419	27	0	0	0
14	しめすへん	社	7	47	41	5	0	1	0	229	209	19	0	1	0
15	いしへん	研	5	45	35	3	0	7	0	294	181	69	0	44	0
15	おんなへん	姉	5	47	38	4	1	4	0	300	159	99	1	41	0
15	くちへん	鳴	5	54	40	13	1	0	0	417	329	87	1	0	0
15	かたへん	貯	4	36	30	6	0	0	0	305	295	10	0	0	0
18	こめへん	放	4	27	27	0	0	0	0	199	199	0	0	0	0
18	しよくへん	軽	4	40	29	11	0	0	0	324	150	174	0	0	0
18	くるまへん	粉	4	41	34	5	1	1	0	176	127	38	10	1	0
18	かいへん	飲	4	13	12	1	0	0	0	47	45	2	0	0	0
18	ひへん	理	4	37	30	7	0	0	0	150	133	17	0	0	0
18	おうへん	畑	4	40	37	2	0	0	1	635	622	12	0	0	1
18	りっしんべん	快	4	50	43	3	3	1	0	585	557	8	8	12	0
上下分離タイプ															
1	うかんむり	字	26	164	140	18	1	4	1	1,663	1,574	82	1	5	1
2	くさかんむり	花	16	147	125	17	5	0	0	1,545	1,458	75	12	0	0
3	たけかんむり	節	13	70	63	3	1	3	0	590	473	90	7	20	0
4	なべぶた	六	12	147	107	39	0	1	0	1,946	1,707	237	0	2	0
5	ひとがしら	合	7	220	155	60	0	3	2	2,869	2,156	705	0	5	11
6	ひ	星	6	106	98	0	3	2	3	1,177	1,079	0	67	3	28
7	あおのうえ*	青	6	52	40	8	3	0	1	844	541	258	43	0	2
8	くち	号	5	69	65	3	0	1	0	542	531	10	0	1	0
8	た	異	5	49	31	18	0	0	0	818	229	589	0	0	0
8	ツ	学	5	42	37	5	0	0	0	652	585	67	0	0	0
8	つち	寺	5	51	36	11	1	1	2	318	175	98	34	9	2
12	夕	角	4	50	42	5	3	0	0	262	230	21	11	0	0
12	あみがしら	罪	4	18	13	5	0	0	0	115	38	77	0	0	0
12	おいかんむり	考	4	34	29	5	0	0	0	1,504	720	784	0	0	0
12	いまのうえ*	今	4	23	21	1	0	0	1	339	336	1	0	0	2
12	しょう	賞	4	16	13	0	0	2	1	39	33	0	0	4	2
17	ム	台	3	22	21	1	0	0	0	139	137	2	0	0	0
17	あめかんむり	雲	3	60	58	1	1	0	0	417	398	18	1	0	0
17	はちがしら	分	3	61	46	12	1	2	0	1,664	1,170	481	3	10	0
17	たつ	音	3	81	47	33	0	1	0	1,120	897	221	0	2	0
17	ふゆがしら	冬	3	17	17	0	0	0	0	114	114	0	0	0	0
17	め	見	3	95	53	40	0	2	0	1,731	692	1,021	0	18	0
17	やま	炭	3	51	50	0	1	0	0	422	421	0	1	0	0
17	ひつじ	義	3	17	16	0	1	0	0	92	51	0	41	0	0
17	じゅう	支	3	33	30	1	1	1	0	502	461	10	23	8	0
17	はかのうえ*	墓	3	15	12	3	0	0	0	55	40	15	0	0	0

* 部首としての名称がないため、便宜的に学年別配当の最も低い漢字一字の読みを用いて「うえ」をつけて部品の名称とした。

表 2 「くさかんむり」ファミリーの単語ファミリーサイズと単語ファミリー出現頻度

対象漢字	学年別 配当	単語ファミリーサイズ						単語ファミリー出現頻度					
		合計	名詞	動詞	形容詞	形状詞	副詞	合計	名詞	動詞	形容詞	形状詞	副詞
花	1	24	24	0	0	0	0	142	142	0	0	0	0
草	1	19	19	0	0	0	0	93	93	0	0	0	0
茶	2	12	11	0	1	0	0	28	26	0	2	0	0
荷	3	3	3	0	0	0	0	8	8	0	0	0	0
苦	3	12	7	2	3	0	0	23	14	3	6	0	0
菓	3	5	5	0	0	0	0	13	13	0	0	0	0
葉	3	20	20	0	0	0	0	911	911	0	0	0	0
落	3	22	9	13	0	0	0	224	155	69	0	0	0
英	4	3	3	0	0	0	0	10	10	0	0	0	0
芽	4	8	6	2	0	0	0	17	14	3	0	0	0
芸	4	8	8	0	0	0	0	15	15	0	0	0	0
菜	4	5	5	0	0	0	0	22	22	0	0	0	0
若	6	5	4	0	1	0	0	9	5	0	4	0	0
蒸	6	2	2	0	0	0	0	6	6	0	0	0	0
蔵	6	3	3	0	0	0	0	8	8	0	0	0	0
著	6	4	4	0	0	0	0	31	31	0	0	0	0
合計		147	125	17	5	0	0	1,545	1,458	75	12	0	0

註) 「くさかんむり」ファミリーの成員には、一つの単語の中に「くさかんむり」を持つ漢字を複数含む単語が、8種類（合計の出現頻度15回）あるため（例えば、「草花」「落葉」など）、重複分を引き算した結果、合計欄が対象漢字毎の合計値と一致しない。

リーの成員を示す。各ラディカルにおいて、単語ファミリーの各出現頻度、品詞毎の書字形（動詞と形容詞では書字形基本形）、対象漢字、ラディカル、および発音形を示した。

その他のラディカルについては、誌面の制約上すべてを掲載することができないため、全51種類の単語ファミリーサイズと出現頻度、総出現頻度、および単語ファミリーの全成員は、Web上で公開している（第一著者のresearchmapを参照、<https://researchmap.jp/read00561116/>資料公開/）。

3.3 学年別の各漢字を共有する単語ファミリーサイズと単語ファミリー出現頻度

表4は、学年別配当毎に、各漢字を共有する単語ファミリーサイズと単語ファミリー出現頻度の平均値（*mean: M*）および標準偏差（*standard deviation: SD*）を示す。

より低学年で学習する漢字ほど、その漢字から構成される単語のファミリーサイズが大きく、かつ単語ファミリー出現頻度も高いのかを検証するために、単語ファミリーサイズと単語ファミリー出現頻度を従属変数として、品詞毎に学年の一要因（六学年）からなる分散分析を実施した。

単語ファミリーサイズに関する分散分析 単語ファミリーサイズに関する一要因の分散分

表3 同一ラディカルを共有する漢字の単語ファミリーの成員（「くるまへん」ファミリー）

品詞 (単語数)	単語ファミリー 出現頻度	書字形*	対象漢字	ラディカル	発音形	
名詞 (3)	1	軽	軽	くるまへん	ケー	
	1	軽そう	軽	くるまへん	ケーソー	
	1	軽重	軽	くるまへん	ケージュー	
	<u>3</u>					
形容詞 (1)	10	軽い	軽	くるまへん	カルイ	
形状詞 (1)	10					
名詞 (10)	1	軽々	軽	くるまへん	ケーケー	
	1					
名詞 (10)	7	運転	転	くるまへん	ウンテン	
	4	転校	転, 校	くるまへん, きへん	テンコー	
	2	回転	転	くるまへん	カイテン	
	2	自転	転	くるまへん	ジテン	
	2	転居	転	くるまへん	テンキョ	
	1	逆転	転	くるまへん	ギャクテン	
	1	転	転	くるまへん	テン	
	1	転換	転	くるまへん	テンカン	
	1	転入	転	くるまへん	テンニユー	
	1	変転	転	くるまへん	ヘンテン	
		<u>22</u>				
	動詞 (5)	28	転ぶ	転	くるまへん	コロビ
7		転がる	転	くるまへん	コロガル	
1		転がり出る	転	くるまへん	コロガリデル	
1		転げる	転	くるまへん	コロゲル	
1		転がり落ちる	転	くるまへん	コロガリオチ	
	<u>38</u>					
名詞 (5)	1	月の輪	輪, 月	つきへん, くるまへん	ツキノワ	
	1	指輪	輪, 指	くるまへん, てへん	ユビワ	
	1	首輪	輪	くるまへん	クビワ	
	1	日輪	輪, 日	ひへん, くるまへん	ニチリン	
	1	年輪	輪	くるまへん	ネンリン	
	<u>5</u>					
名詞 (4)	1	輸出	輸	くるまへん	ユシュツ	
	1	輸出入	輸	くるまへん	ユシュツニユー	
	3	輸入	輸	くるまへん	ユニユー	
	17	輪	輪	くるまへん	ワ	
	<u>22</u>					
名詞 (12)	30	車	車	くるまへん	クルマ	
	12	糸車	車, 糸	くるまへん, いとへん	イトグルマ	
	10	汽車	車, 汽	くるまへん, さんずい	キシヤ	
	7	列車	車	くるまへん	レッシュヤ	
	4	風車	車	くるまへん	カザグルマ	
	3	車庫	車	くるまへん	シャコ	
	3	水車	車, 水	くるまへん, さんずい	スイシャ	
	2	車道	車	くるまへん	シャドー	
	1	はしご車	車	くるまへん	ハシゴシャ	
	1	車たい	車	くるまへん	シャタイ	
	1	乗車	車	くるまへん	ジョーシャ	
	1	電車	車	くるまへん	デンシャ	
	<u>75</u>					

* 動詞と形容詞の書字形の列には書字形基本形が記載されている。

表 4 学年別の各漢字を共有する単語ファミリーサイズと単語ファミリー出現頻度の平均値 (M) および標準偏差 (SD)

学年別 担当	漢字数	単語ファミリーサイズ							単語ファミリー出現頻度					
		合計	名詞	動詞	形容詞	形状詞	副詞	合計	名詞	動詞	形容詞	形状詞	副詞	
1 年生	41	M	21.10	18.59	2.07	0.15	0.20	0.10	268.39	231.59	32.17	3.05	1.44	0.15
		SD	16.20	13.88	7.41	0.42	0.45	0.37	312.57	268.56	158.11	10.76	4.02	0.57
2 年生	70	M	15.37	12.43	2.53	0.21	0.16	0.04	248.29	165.27	80.24	1.61	1.04	0.11
		SD	12.43	8.62	6.65	0.63	0.58	0.20	337.51	188.44	207.89	6.34	5.61	0.64
3 年生	109	M	7.40	5.83	1.24	0.16	0.13	0.05	81.65	62.11	17.63	1.46	0.28	0.17
		SD	4.79	3.80	2.23	0.56	0.41	0.25	130.25	112.26	55.29	6.24	1.50	1.20
4 年生	98	M	5.64	4.70	0.77	0.04	0.09	0.04	34.07	24.98	8.09	0.16	0.48	0.36
		SD	3.91	2.76	1.93	0.20	0.45	0.20	48.51	31.47	35.09	1.04	4.12	2.57
5 年生	98	M	4.62	3.72	0.68	0.05	0.12	0.04	23.35	20.02	2.67	0.07	0.56	0.10
		SD	2.54	2.10	1.31	0.22	0.66	0.20	36.86	34.15	7.47	0.36	4.35	0.65
6 年生	95	M	2.59	2.18	0.33	0.03	0.04	0.01	10.25	8.80	1.06	0.15	0.22	0.02
		SD	1.66	1.50	0.69	0.17	0.20	0.10	11.96	11.77	2.52	0.83	1.75	0.20
全学年	511	M	7.83	6.46	1.12	0.10	0.11	0.04	85.88	64.74	19.60	0.85	0.56	0.16
		SD	9.05	7.33	3.63	0.40	0.48	0.22	190.87	136.72	97.41	4.93	3.70	1.33

析の結果、名詞材料では学年の主効果が有意であった ($F(5, 505) = 71.32, p < .001, \eta^2 = .41$)。多重比較 (Bonferroni 法、いずれも $p < .05$) の結果、ファミリーサイズは一年生において最も大きく、次に二年生が大きいことが明らかとなった。三年生および四年生のファミリーサイズは、六年生のそれよりも大きいことも明らかとなった。したがって、一年生および二年生に担当されている漢字から構成される単語は、より高学年で学習する漢字から構成される単語に比べて単語のバリエーションが多いことが示された。

次に、動詞材料でも学年の主効果が有意であり ($F(5, 505) = 4.19, p < .001, \eta^2 = .04$)、多重比較の結果、二年生のファミリーサイズが、四年生から六年生のファミリーサイズに比べて大きいことが明らかとなった。したがって、動詞では、二年生に担当されている漢字から構成される動詞のバリエーションが最も多いことが示された。

形容詞材料においても学年の主効果は有意であったが ($F(5, 505) = 2.96, p < .05, \eta^2 = .03$)、多重比較の結果、各水準間には有意差は認められなかった。また、形状詞 ($F(5, 505) = 0.84, n.s.$) と副詞 ($F(5, 505) = 0.97, n.s.$) では有意な学年の主効果は認められなかった。

単語ファミリーの出現頻度に関する分散分析 単語ファミリーの出現頻度に関する一要因の分散分析の結果においても、名詞では学年の主効果が有意であった ($F(5, 505) = 35.83, p < .001, \eta^2 = .26$)。多重比較の結果、出現頻度は一年生と二年生が、三年生以上と比べて高いことが明らかとなった。三年生の単語ファミリー出現頻度は、六年生のそれよりも高いことも明らかとなった。したがって、一年生および二年生の担当漢字から構成される単語は、高

学年で学習する漢字から構成される単語に比べて単語ファミリーの出現頻度が高いことが示された。

動詞においても学年の主効果が有意であり ($F(5, 505) = 9.13, p < .001, \eta^2 = .07$)、多重比較の結果、二年生の単語ファミリー出現頻度が、三年生から六年生のそれらに比べて高いことが明らかとなった。したがって、動詞では、二年生の配当漢字から構成される動詞の単語ファミリー出現頻度が最も高いことが示された。

形容詞においても学年の主効果が有意であった ($F(5, 505) = 3.64, p < .01, \eta^2 = .04$)。多重比較の結果、一年生の単語ファミリー出現頻度が四年生から六年生までのそれらに比べて高いことが明らかとなった。したがって、一年生の配当漢字から構成される単語の単語ファミリー出現頻度は高学年のそれらに比べて高いことが示された。

形状詞 ($F(5, 505) = 1.0, n.s.$) と副詞 ($F(5, 505) = 0.70, n.s.$) では有意な学年の主効果は認められなかった。

以上の結果から、名詞に関しては、低学年（一、二年生）で学習する漢字ほど、その漢字から構成される単語のファミリーサイズが大きく、また単語ファミリー出現頻度も高くなるという関係が示唆された。一方、動詞に関しては、二年生で学習する漢字から構成される単語のファミリーサイズが最も大きく、単語ファミリー出現頻度も高いことが示唆された。

表5は、単語ファミリーサイズが大きい漢字の名詞と動詞の上位各10位を示す。表5によると、名詞では、一年生に配当されている漢字が六個であり、二年生に配当されている漢字が四個であった。これらの漢字の多くは、象形文字であり自然の概念を表す基本語として位置づけられる (e.g., 「手」, 「日」)。小学校国語教科書では、低学年の間にこれらの基本語を学習し、その後、六学年を通してこれらの語から構成要される単語のバリエーションが増加し、語彙量を増やしていきけるように構成されていることが示唆される。

表5 名詞と動詞の単語ファミリーサイズ上位10語とその単語ファミリー出現頻度

順位	名詞				動詞			
	対象漢字	学年別 配当	単語ファミリー サイズ	単語ファミリー 出現頻度	対象漢字	学年別 配当	単語ファミリー サイズ	単語ファミリー 出現頻度
1	手	1	59	575	合	2	49	500
2	日	1	58	824	見	1	36	1,008
3	人	1	50	1,237	立	1	33	221
4	地	2	46	315	思	2	17	585
5	水	1	39	216	読	2	16	1,189
6	木	1	37	296	付	4	15	69
7	山	1	34	342	落	3	13	69
8	体	2	31	295	分	2	12	481
9	合	2	29	227	言	2	11	503
10	色	2	29	167	打	3	11	56

一方、動詞では、単語ファミリーサイズが大きい漢字の上位10位には、二年生に配当されている漢字が多い (e.g., 「合」, 「思」)。上位の漢字は、例えば、「合う」は、「話し合う」「読み合う」「合わせる」「伝え合う」など多くの動詞の構成要素として使用されている。このような漢字は、二年生以上で学習し、これらの漢字を構成要素に持つ動詞のバリエーションが二年生以降で増えていくことが示唆される。

4. お わ り に

本研究の結果、小学校国語教科書に掲載されている五つの品詞（名詞・動詞・形容詞・形状詞・副詞）に限定した全単語を対象として、教育漢字において使用されるタイプ頻度の高いラディカルを共有する漢字群および単語群の一覧表が利用可能となった。これらの材料をもとに、ラディカルを中心とする単語ファミリーの意味的活性化が、単語の認知過程において果たす役割を検討する実験が可能となるだろう。

意味を担う最小単位である形態素が意味処理において果たす役割に関する研究は、英語 (Taft, 1979) やオランダ語 (Bertram, Baayen, & Schreuder, 2000; Schreuder & Baayen, 1997) を材料とした研究において報告されている。例えば、Schreuder & Baayen (1997) および Bertram et al. (2000) は、オランダ語における形態素ファミリーサイズ (morphological family size) の効果について検討している。形態素ファミリーとは、屈折 (e.g., *tablet*, *tabular*) や複合 (e.g., *tablespoon*, *timetable*) によって、ある語幹 (stem) から派生した単語のセットであり、形態素ファミリーサイズとは、派生語や複合語を合わせた単語のタイプ数である。Schreuder & Baayen (1997) は、名詞を用いた語彙判断課題を実施し、名詞の形態素ファミリーサイズが小さい条件よりも、大きい条件において、反応時間が短くまた反応が正確であるという結果を報告している。さらに、Bertram et al. (2000) では、複合語の語幹単語の形態素ファミリーサイズが語彙処理に影響することを報告している。彼らは、一連の実験結果に基づき、形態素ファミリーサイズによるタイプ頻度効果は、語幹単語のトークン頻度とは独立に生起し、意味的に透明な形態素ファミリーメンバーによって駆動されると論じている。そして、形態素ファミリーサイズ効果は、心的辞書に貯蔵される意味的に透明な、形態素を共有する単語間における活性化拡散に依存するという仮説を支持すると主張している。日本語の漢字を含む単語についても、このような形態素ファミリーサイズやファミリー出現頻度による効果の検証が期待される。

本研究では、教育漢字においてタイプ頻度の高いラディカルを対象として単語ファミリーの一覧表を整備したが、ここで注意すべき点が二つある。第一に、本研究では左右分離と上下分離のタイプを分けて別々に単語ファミリーをカウントしているが、前述の「にんべん」

と「ひとがしら」のように、同一ラディカルが形を変えて様々な位置に利用されることがある。したがってこの例では、「人」を形態素として含む単語ファミリーは、「にんべん」ファミリーと「ひとがしら」ファミリーの両方を含むと考えられる。ただし、両者は形態的類似性の観点からは異なる単語群として捉えることが妥当である。また、「思」などの下部品として使用される「心」は、同じく「心」に由来する「りっしんべん」とは形態的に大きく異なるという特徴を持つが、「心」を下部品に含む漢字は多く、「志」や「忘」など、教育漢字に14種類が含まれる。さらには、「慕」の下部品も元々の由来は同じである。本研究では、ラディカルが下部品として共有される場合を対象としていないために、心を含むこれらの漢字を「心」ファミリーに含めていない。

第二に、本研究では、左右分離タイプと上下分離タイプの漢字のみを対象とし、ラディカルとして左部品と上部品のみに注目した。しかし、左右分離タイプの左部品が右部品として含まれる漢字や、上下分離タイプの上部品や下部品として含まれる漢字も存在する。例えば、「木」は、左部品としては「きへん」であるが、「休」の右部品、「杳」の上部品、「楽」の下部品などにも使用される。使用される位置が異なる場合も、形態素を共有するという観点では、これらも「木」ファミリーの成員であるともみなしうる。

本研究では、このような位置をも含めた形態的類似性を区別した上で、ラディカルを共有する単語ファミリーをカウントした。研究の目的に応じて、これらの位置の異なる、あるいは共通の由来を持つが形態が異なるラディカル単語ファミリーを合算するなどの操作も必要となるだろう。位置情報をも含めた形態的類似性を前提とした上で形態素が担う意味から生じる意味的類似性と、位置情報に関わらない同一形態素を共有することによる意味的類似性とが単語認知過程に異なる影響を及ぼすのか否かについても、今後の検討課題として残される。

本研究では、コーパスとして小学校の国語教科書を用いた。はじめに述べたとおり、漢字の学習を始めたばかりの子どもの心的辞書は、基本概念を表す形態素を核として、言語の階層的構造の中で、形態的、音韻的、および意味的に関連する語彙を徐々に増やすように発達していくと想定される (Schröter & Schroeder, 2017)。成長過程において環境から受け取る言語刺激の量や質によって、個々人が各々の意味空間を構築してゆく。人間の言語発達過程に沿った語彙量や様々なジャンルを含むコーパスに基づく心理実験やコーパス分析を通じて、子どもから大人への心的辞書の発達過程や言語知識の意味空間構造を解明することが期待される。

引用文献

- Baayen, R. H., Piepenbrock, R., & Gullikers, L. (1995). *The CELEX lexical database [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, **39**, 445–459. doi: 10.3758/BF03193014
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, **284**, 34–43.
- Bertram, R., Baayen, R.H., and Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, **42**, 390–405.
- Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407–428.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 240–248.
- Fellbaum, C. (2005). WordNet and wordnets. In B. Keith. (Ed in chief), *Encyclopedia of Language and Linguistics, Second Edition* (pp. 665–670). Oxford: Elsevier.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, **42**, 488–496. doi: 10.3758/BRM.42.2.488
- 市川 本太郎 (1963). 教育漢字指導法——学年配当と字源解説 学芸図書
- 岩田 誠 (1994). 読み書きと脳—特集 認知科学の現状と展望 認知科学, **1**, 19–30.
- Joyce, T. (2011). The significance of the morphographic principle for the classification of writing-systems. *Written Language & Literacy*, **14**, 58–81. doi:10.1075/wll.14.1.01joy
- Joyce, T., Masuda, H., & Ogawa, T. (2014). Jōyō kanji as core building blocks of the Japanese writing system: Some observations from database construction. *Written Language & Literacy*, **17**, 173–194. doi: 10.1075/wll.17.2.01joy
- Kess, J. F., & Miyamoto, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing*. Philadelphia: John Benjamins.
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, **1**, 174. doi: 10.3389/fpsyg.2010.00174
- 国立国語研究所 (2011). 現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese: BCCWJ). http://pj.ninjal.ac.jp/corpus_center/bccwj/
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211–240.
- Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*, Cambridge, England: Cambridge University Press.
- (マニング, C., ラガバン, P., シュッツェ, H. 岩野 和生・黒川 利明・濱田 誠司・村上 明子 (訳) (2012). 情報検索の基礎 共立出版)
- McNamara, T. P. (2012). *Semantic priming: Perspectives from memory and word recognition*. New York, NY: Psychology Press.
- Miller, G. A. (1991). *The Science of words*. New York: Scientific American Library.
- 小河 妙子 (2012). 教育漢字を対象とした部品 (部首) を共有する漢字群の意味的類似性に関する検討 東海学院大学紀要, **5**, 217–223.
- 小河 妙子 (2014). 教育漢字を対象とした部品 (部首) を共有する漢字群の意味的類似性に関する検討 II—上下分離漢字の分析 東海学院大学紀要, **8**, 161–168.
- Ogawa, T., & Fujita, C. (2017). Morphological analysis of sentences in the Japanese language textbooks for sixth graders in Japanese elementary schools. *Bulletin of Tokai Gakuin University*, **11**, 149–155.
- 小河 妙子・藤田 知加子 (2018). 小学校国語教科書に掲載されている単語の種類数と出現頻度 読書科学, **60**,

241-253.

- Ogawa, T., Fujita, C., Joyce, T., Kawakami, M., & Masuda, H. (2012). Semantic similarities among radical-neighbors of kanji characters based on multi-dimensional scaling. *The 14th International Conference on the Processing of East Asian Languages*. Nagoya, Japan.
- Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods*, **50**, 1568–1580. doi: 10.3758/s13428-017-0981-8
- Schreuder, R., & Baayen, R. (1997). How complex simplex words can be. *Journal of Memory and Language*, **36**, 118–139.
- Schröter, P., & Schroeder, S. (2017). The Developmental lexicon project: A behavioral database to investigate visual word recognition across the lifespan, *Behavior Research Methods*, **49**, 2183–2203. doi: 10.3758/s13428-016-0851-9
- Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic Syllabic Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, **18**, 21–39.
- Tse, C. S., & Yap, M. J. (2018). The role of lexical variables in the visual recognition of two-character Chinese compound words: A megastudy analysis. *Quarterly Journal of Experimental Psychology*, **71**, 2022–2038. doi: 10.1177/1747021817738965

Summary

Analysis of words appearing within Japanese language
textbooks for elementary school:

Family sizes and token frequencies for words
consisting of kanji with shared radical

Taeko Ogawa, Chikako Fujita, and Hisashi Masuda

Japanese kanji are a morphographic script, such that, for example, 木 represents the morpheme/ki/ *tree* as a single kanji character, which can also appear as either the left radical of another kanji 松/matsu/ pine or as the first consistent of the two-kanji compound word 木材/mokuzai/ *wood*. Within Japanese language education at elementary schools, morphological knowledge concerning the formation of kanji is explicitly utilized as clues for learning to read and write kanji. In this study, which analyzes a corpus of elementary school books, we count and present as summary tables the following five characteristics for all kanji and words sharing radicals; namely, word family members and type frequencies, the token frequencies of each member of a word family, the token frequency of each word family, and the total token frequencies of word families. These tables of information will enable researchers to suitably control for the lexical materials of experiments that seek to elucidate the structures of the mental lexicon from the developmental perspective elementary school students during the early stages of learning kanji characters.