『ぐりとぐら』に見られる言葉
――絵本の語彙研究のための一考察――

The Words Found in “GURI and GURA”: A Study for the Lexical Study of Picture Books

平林　一利
Kazutoshi Hirabayashi

はじめに

子どもの言葉の発達における絵本との関わりは重要であり、また必要とされる[1)]。絵本が子どもの言葉の発達にどのように関わり合い、子どもの言葉の育みにどのように役立つのかを考える必要がある。そのためには幼児絵本にどのような言葉が用いられているのかを知る必要がある。

本稿では、その取り組みの一環として、幼児絵本に見られる語彙にはどのようなものがあるのかを知るために、幼児絵本に見られる語彙の数量的、意味的な調査を試みる。

なお、このような絵本を用いた語彙解析については、電子計算機を用いた大規模データの形態素解析調査がおこなわれておりコーパスも作成され評価されている[2)]。

調査資料

幼児用の絵本はブックスタート[3)]以後注目され、以前にも増し、出版されているため調査対象とすべき絵本はさまざまであるが、本稿では、長期にわたり子どもに読み継がれている『ぐりとぐら』シリーズの最初の作品である中川李枝子（作）・大村百合子（絵）『ぐりとぐら』を資料とする。『ぐりとぐら』を基礎資料とするのは、1963年の出版以来継続的に読み継がれており、いわゆる名作とされる作品であり、多くの人々に支持されているためである。これは、『てんじつき さわる えほん ぐりとぐら』[4)]として点字絵本として出版されていることからも、需要があり支持されており、絵本の語彙研究の基本的な資料作成の出発点として適切かと考えたためである。

調査方法

調査方法は、『ぐりとぐら』初版（236刷）をデータ入力し、それぞれの文に、地の文、会話文、歌の文の種別を付加した[5)]。そのうえで、国立国語研究所が公開している、形態素解析アプリケーション「web茶まめ」[6)] を用いて形態素解析をおこなった。

「web茶まめ」から出力されたデータから、単語（形態素）の切れ目と、「語彙素」「語彙素読み」「品詞」「語種」の情報を確認し、誤解析されている箇所については、手作業で修正を施した。また、形態素より長い単位として扱うべきと考える例についても修正を施した。なお、句読点、括り記号などは言葉ではないとみなし、今回の調査対象から除外した[7)]。

解析における処理

解析について

「web茶まめ」は通常の漢字仮名交じり文であれば解析にさほど問題は無いと思うが、すべて平仮名文である絵本では、いくつかの誤解析がおこなわれた。これらについては、手作業で修正を施した。また、品詞認定についても以下、単語の品詞の修正を施した。以下に修正例を示す。

単語の修正

・の　ねずみ　→　のねずみ（野ネズミ）　・かす　てら　→　カステラ

・繰り　→　ぐり　・り　ゅっくさ　っ　く　→　りゅっくさっく（リュックサック）

・うち（内）　へ　かえって（却って）　→　うち　→　家　へ　帰って

これらは、原文のまま平仮名文をそのまま「茶まめ」で解析をしたことによる結果であって分析ツールが原因ではなく、通常の漢字仮名交じり文にすることでこれらは減少した[8)]。

品詞の修正

ところ（名詞）　で（助詞）　→　ところで（接続詞）

ぽん（副詞）と（助詞）て（助詞）　→　て（手）名詞

形態素より長い単位での認定

「web茶まめ」の解析はかなり細かく、接頭辞、接尾辞（接辞的）についても解析されるため、以下のように接頭辞・接尾辞（接辞的）を含む語を一語として修正した。

２　ひき　→　２匹

お　さとう　→　おさとう（お砂糖）

お　なべ　→　おなべ（お鍋）

お　つき　さま　→　おつきさま（お月様）

お　りょうり　→　おりょうり（お料理）

ご　ちそう　→　ごちそう（ご馳走）

こむぎ　こ　→　こむぎこ（小麦粉）

ぼく　ら　→　ぼくら（僕ら）

もり　じゅう　→　もりじゅう（森じゅう）

などと解析されているため、修正をしている。これは、すべて平仮名で絵本を分析したことによる結果であると考えられるため、通常の漢字仮名交じり文で入力することによりこれらのものは解消できると思われる。

『ぐりとぐら』に見られる語の量的な性質

修正し整理したデータをもとに、『ぐりとぐら』から得た語彙を集計し異なり語数191語、述べ語数605例を得た。以下このデータをもとに考察する。

語種および品詞の構成比率

『ぐりとぐら』に見られる語種は表１のとおりである。なお、頻出する「ぐり」「ぐら」は、和語扱いとする。

表１　語種の異なり語数と述べ語数（総数）

総数	和語	漢語	外来語	混種語	計
異なり語数	163	16	8	4	191
述べ語数	556	23	18	8	605

語種を異なり語数・述べ語数ともに、和語が大多数を占めていることがわかる。全数調査の場合日本語の特性として助詞・助動詞が多数を占めるため、ここから付属語である助詞・助動詞を抜き自立語のみで集計したものが表２である。

表２　語種の異なり語数と述べ語数（自立語のみ）

自立語のみ	和語	漢語	外来語	混種語	計
異なり語数	134	16	8	4	162
述べ語数	277	23	18	8	326

助詞・助動詞は基本的に和語であるので、和語の数値が減少するのみであるが、和語が多数使用されることには変わりは無い。大規模データによる語種の割合は多く『現代雑誌90種の用語用字第三分冊分析』[9)] のものが参考にされるが、その語種別分布は、表３の通りである。

表３　現代雑誌90種の用語用字における語種分布

	和語	漢語	外来語	混種語	計
異なり語数	36.7%	47.5%	9.8%	6.0%	100%
述べ語数	53.9%	41.3%	2.9%	1.9%	100%

『ぐりとぐら』に見られる語彙の種別を割合で示すと、自立語のみにおいて表４のとおりである。

表４　『ぐりとぐら』における語種分布

	和語	漢語	外来語	混種語
異なり語数	82.8%	9.8%	4.9%	2.5%
述べ語数	84.9%	7.1%	5.5%	2.5%

一般的な文章に見られる語種の割合では、異なり語数では和語と漢語の使用率が逆になることが多い[10]が、『ぐりとぐら』に見られる語種は異なり語数、述べ語数ともに和語が多用されていることがわかる。また、助詞、助動詞を含むと、和語の割合が異なり語数で85.4%、述べ語数では、91.9%になりその使用率はさらに増加する。

なお、蓮井（2014）は、文体がやわらかいとされる資料には、漢語の使用率が低く、文体が堅いとされる資料では、漢語の使用率が高いとしている[11]。

『ぐりとぐら』に見られる語彙の量的構造

田中（2016）の語彙の量的構造の研究方法を参考に『ぐりとぐら』に見られる言葉を数量的に考察する。先に示した異なり語数、述べ語数をもとに図１と図２を作成した。

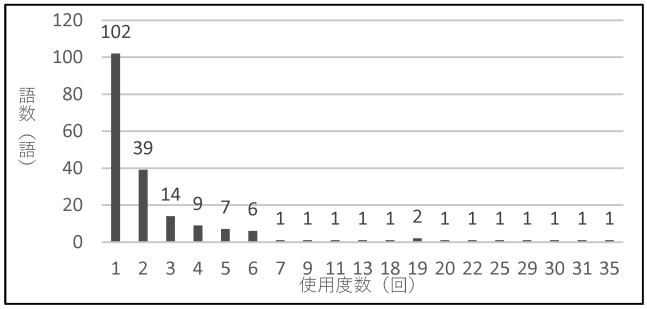

図１　語の度数分布

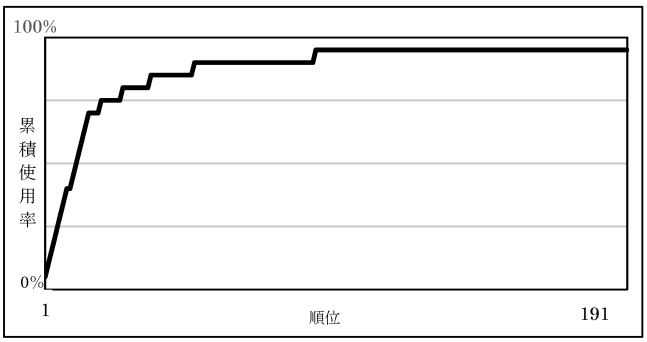

図２　語の累積使用率

図１は、異なり語数、述べ語数をもとに度数分布をグラフにまとめたものであり、図２は、頻度の高いものを横軸に示し、その使用数の高いものから順位をとり、縦軸に累積使用率を示したものである。

図１から、度数が上がるに従いその語数が減少していく様子が見てとれる。また、図２からは、使用度数の高い一部の語だけで語彙全体のかなりの部分を占めていることがわかる。図２では数値が明確にできないが、異なり語数の上位10％で約52％を占めている[12)]。

なおこのような語彙の量的な構造は、多くの日本語の文章において共通するものである[13)]。

語彙のランク

高頻度の語彙と低頻度の語彙にはどのような語が所属しているのかを確認する。このことを考えるために、図１の元となった語別頻度表をもとに高頻度語彙から低頻度語彙へとして田中（2016）[14)]を参考にランク別に５段階に分けた。

このランクに従い、各ランクにおける品詞構成を示す。なお、ランク分けは、以下の通りである。

表５　使用度数による語彙のランク

ランク	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ
度数（回）	21～	11～20	4～9	2～3	1
語数（語）	6	6	22	52	106

「web茶まめ」で形態素解析したものを、11品詞に分類し、データを見やすくするために、類似性の高いものを次のａ～ｄの４種にまとめ、それぞれの数値を示した。

ａ名詞・代名詞　　ｂ動詞・形容詞　　ｃ形状詞・副詞・連体詞・接続詞・感動詞
ｄ助詞・助動詞

表６　語彙のランクと品詞

ランク	a	b	c	d
Ⅰ	0%	0%	0%	100%
Ⅱ	33.3%	16.7%	0%	50%
Ⅲ	27.3%	36.4%	9.0%	27.3%
Ⅳ	36.5%	30.8%	17.3%	15.4%
Ⅴ	43.4%	30.2%	20.7%	5.7%

表６から読み取れることは、まず最も高頻度のランクⅠの語彙は、ｄの助詞と助動詞である。２番目に高頻度のランクⅡの語彙も、ｄが半数と高い比率を占めている。ランクが上がるにともない、ｄに所属する語彙の減少が見られ、それに対してａ～ｃの語彙の増加が見られる。

ランクⅠ、Ⅱに見られる語彙の内訳を品詞別に示すと以下のようになる。

ランクⅠ　助詞「て、と、の、を」 助動詞「た、ます」

ランクⅡ　助詞「が、は」 助動詞「だ」 動詞「為る」

これらの語は、どのようなことを表現する際にも用いられる語であると言うことができるものである。ランクⅡに名詞が含まれるがこれは、登場キャラクターの「ぐり」と「ぐら」であるので、絵本の性格からして多用されるものである。高頻度と低頻度の中間にあたるランクⅢでは、名詞「お鍋、たまご、カステラ」など、絵本のストーリーに関わる語が多く見られる。動詞は「いう、いく、いる」など、形容詞は「おおきい、よい」などが見られるが、これらはさまざまな文章に見られる一般的なことを表現する語であると思われる。他にも動詞には「つくる、できる」があるが、絵本のストーリーに関連する語であり、また一般的なものと考えることのできる語である。低頻度になるランクⅣになると、高頻度とはことなり、ａ名詞、ｂ動詞・形容詞の比率が高くなり、ｄ助詞・助動詞の比率が大きく低下している。そして最も低頻度のランクⅤでは助詞・助動詞を含まず、ａ～ｃで94.3％を占めている。名詞を原文のまま示す。

あいだ、あさ、あと、いえ、いっしょ、うでぐみ、おく、おつきさま、かお、からっぽ、かんがえ、かんしん、くり、くりーむ、けち、げんこつ、ごちそう、ごらん、ころ、ざいりょう、さんせい、たきぎ、たまごやき、つくり、どうぶつ、どんぐり、なか、なみだ、におい、にまい、のねずみ、はな、ばん、べっど、まっち、まんなか、みち、みんな、め、めだまやき、もり、もりじゅう、よういい

このように具体的な事柄を表す語が多く、ｂ、ｃの動詞、形容詞、副詞などにも、具体的な様子や関係を表す語が多い。低頻度語には、そのような性質の語が多く含まれるのである。

語彙ランクと語種

ここでは語彙のランクと語種の面から考える。

表７　語彙のランクと語種

ランク	和語	漢語	外来語	混種語
Ⅰ	100％	0％	0％	0％
Ⅱ	100％	0％	0％	0％
Ⅲ	100％	0％	0％	0％
Ⅳ	76.9％	9.6％	7.7％	5.7％
Ⅴ	85.7％	10.6％	2.8％	0.9％

以上のようにランクⅠ～Ⅲはすべて和語で占められていることがわかる。ランクⅣでは、漢語、外来語、混種語が見られるようになるが合計20％程度であり、和語が大多数を占める。ランクⅤで

も和語が85％を超している。これは幼児用絵本という特性によるものであり、通常はランクⅢ～Ⅴにおいて漢語の比率が上がることになる[15)]。

意味分野による分析

最後に、意味分野から考える『分類語彙表』[16)] によって自立語を分類すると、以下のような結果となった。その他には、助詞、助動詞、判別出来ないものを含めた。

表８　分類語彙表による分類

部門	語数	%
１．抽象的関係	68	35.40％
２．人間活動の主体	6	3.10％
３．人間の活動―精神および行為	38	19.80％
４．生産物および用具	21	11.00％
５．自然物および自然現象	27	14.00％
６．その他	32	16.70％

『分類語彙表』による分類では、『ぐりとぐら』には抽象的関係の語彙が多く見られた。具体的には、１．抽象的関係には、「間　朝　晩　真ん中　奥　二匹　二枚　落ちる　転がす　流す　居る」など、時間、空間、量、作用、存在に分類される語が多く含まれる。次に多い、３．人間の活動―精神および行為には「好き　考え　名前　言う　話す　歌う　賛成　お料理　関心　作る　煮る」など、心、言語、生活、交わり、行為に分類される語が見られる。さらに、４．生産物および用具には、「薪・マッチ、エプロン、クリーム、卵焼き、お砂糖、バター、カステラ、かまど、お鍋」など資材、食料、道具など絵本のストーリーに直接関わりのある語の存在が確認できる。さらに、２．人間活動の主体には、「皆、僕ら、家、世」など、人間、社会に関わる語が確認でき、５．自然物および自然現象では「石、お月様、栗、団栗、野ネズミ、鼻、目、痛い」など、自然、天地、動物、身体、生命に関わる語彙が見られた。

これらは子どもたちの生活の場において身近なものとして使われる語であると考えられる。子どもたちに絵本を読み聞かせることにより、または、自ら絵本を読むことにより、これらの語彙に触れることとなり、言葉が育まれていくものと考えられる。

まとめと課題

絵本と子どもの言葉の育みを考えるために、今回『ぐりとぐら』に見られる語彙について計量的、意味的に考察を試みた。特に語種の分類においては、一般的な文章における語種の比率と異なることがわかった。これは、語彙ランクと語種の関係でも同じであった。また、意味的な分類からは、

抽象的な意味の語や、人間の活動に関わる語が多く見られた。

今回の調査は『ぐりとぐら』1冊のみをデータ化したものであるので、用例数が限られる。そのため本稿で示した数値的な考察のデータは参考程度になるが、調査対象を増やすことにより、絵本における語彙の特性などが明確になると考えられる。

また、基本的なデータ作成においては、入力方法の工夫も必要であり、また一語の認定についての考察も以降の研究の課題となる。

本稿は、絵本から子どもたちが、実際にどのような言葉を育んでいくのかということを目的とする研究の実験的な考察を試みた。

参考文献

青葉ことばの会編（2016）. *日本語研究法【近代語編】*：おうふう.

浅木尚実（編）（2015）. *絵本から学ぶ子どもの文化*：同文書院.

大久保　愛（1993）. *乳幼児のことばの世界*：大月書店.

計量国語学会【編集】（2009）. *計量国語学事典*：朝倉書店.

国立国語研究所（2004）. *分類語彙表　増補改訂版*：大日本図書.

田中牧郎（2016）. 日本語研究法【近代語編】：あおば言葉の会編, *『浮雲』の量的構造.*（pp53～64）：おうふう.

中川李枝子・山脇百合子（1963）. *ぐりとぐら*：福音館書店.

中川李枝子・大村百合子（2013）. *てんじつき　さわるえほん　ぐりとぐら*：福音館書店.

中川素子・吉田新一・石井光恵・佐藤博一（編集）（2011）. *絵本の事典*：朝倉書店.

田中牧郎（2016）. 日本語研究法【近代語編】：あおば言葉の会編, *『浮雲』の量的構造.*（pp53～64）：おうふう.

蓮井理恵（2014）. *動詞・形容詞・副詞における語種比率（RJF）を用いた文体分析－公人スピーチ・「天声人語」・女性ファッション誌記事を事例に－*. 学習院大学大学院日本語日本文学10号.

平　博順・藤田早苗・小林哲生（2012）. *絵本テキストにおける高頻度語彙の分析*：情報処理学会関西支部　支部大会　講演論文集.

藤田早苗・小林哲生・奥村優子・服部正嗣（2017）. *幼児の語彙獲得と絵本コーパスの関係を探る*：言語処理学会第23回年次大会　発表論文集.

注　釈

1）浅木尚実（編）（2015）. *絵本から学ぶ子どもの文化*：同文書院, 大久保　愛（1993）. *乳幼児のことばの世界*：大月書店. 中川素子・吉田新一・石井光恵・佐藤博一（編集）（2011）. *絵本の事典*：朝倉書店.

2）日本では2000年の「子ども読書年」を契機に, 翌2001年にスタートした. 中川素子・吉田新一・石井光恵・佐藤博一（編集）（2011）. *絵本の事典*：朝倉書店.

3）浅木尚実（編）（2015）. *絵本から学ぶ子どもの文化*：同文書院, 中川素子・吉田新一・石井光恵・佐藤博一（編集）（2011）. *絵本の事典*：朝倉書店.

4）中川李枝子・大村百合子 (2013). *てんじつき　さわるえほん　ぐりとぐら*：福音館書店.

5）地の文, 会話文, 歌に分類をしているが, データ量が少ないため今回の報告では, それぞれについての分析はしていない.

6）「Web茶まめ」は,「大学共同利用機関法人　人間文化研究機構　国立国語研究所」が開発した形態素分析ツールである. 日本語の語彙分析ツールとしてその信頼性から研究・教育に多く利用されているため, 本稿でも基礎解析に利用した.

7）「！」「？」は物語の流れの中において重要な記号であるが, 今回は絵本の中の語彙についての調査であるため除外している.

8）これらの「野ネズミ」「カステラ」「リュックサック」とすることで正確に解析される. ただし「ぐり」に関しては品詞解析では「副詞（オノマトペ)」または「繰り（動詞)」に解析されることがあるので, いずれにせよ, 手作業での修正は必要になる.

9）計量国語学会【編集】(2009). *計量国語学事典*：朝倉書店.

10）和語では抽象的な意味を表す語彙が少ないとされ, それに変わって漢語がその役割を担っている. そのため, 具体的な語には和語が多く, 抽象的な語には漢語が多いとされ, 日本語では, 多くの表現をするために異なり語として漢語の比率が高くなる. しかし, 基本的な語が多い和語は, その使用率において漢語を上回る.

11）蓮井理恵 (2014). *動詞・形容詞・副詞における語種比率（RJF）を用いた文体分析－公人スピーチ・「天声人語」・女性ファッション誌記事を事例に－*. 学習院大学大学院日本語日本文学10号.

12）なお本来は度数の高い一部の語だけで, 語彙全体のかなりの部分を占めるが, 今回の調査は資料が一冊であり語彙の数量が限られていることによる.

13）計量国語学会【編集】(2009). *計量国語学事典*：朝倉書店. では, フランス語・スペイン語・ロシア語・ドイツ語など, 他の言語の文章にも共通するものであることが述べられている.

14）田中牧郎 (2016). 日本語研究法【近代語編】：あおば言葉の会編, *『浮雲』の量的構造*. (pp53～64)：おうふう.

15）同上

16）国立国語研究所 (2004). *分類語彙表　増補改訂版*. 大日本図書における中項目による分類をした.