

『窓ぎわのトットちゃん』語彙データ

田 島 毓 堂

1. はじめに

黒柳徹子著『窓際のトットちゃん』は1981年3月5日に初版第一刷りが発行されて以来、同氏が文庫本後書きで言うように、この種の出版としては異例なミリオンセラーになっている。発売後1年で450万部、3年後には600万部に達しようとしているという。「日本の出版界で、初めてのこと」であった。最初、当時はやりだったタレント本として目を向けようとしなかった人々にも受け入れられた結果、読者は、5歳から103歳の広い層にわたっているという。また、英語版を初め中国語には数種類の翻訳があり、韓国語・インドネシア語もあり、現在十数個の言語に訳されている。

このように、本書が世に迎えられたことは、何よりもその内容の魅力、就中、トモエ学園を創設し、ユニークな教育理念によっていわば現在で言えば「落ちこぼれ」とか「問題児」の教育に情熱を傾けた小林宗作氏の教育に対する社会的共感を呼んだことである。そのことを世に知らせたのは、軽妙な語り口で、現にマスコミ界で活躍している黒柳徹子氏の飾らぬ、わかりやすい文章にあったことは言うまでもない。何よりも、広い読者層がそれを示している。

当時の教育界などに与えた影響も大きい。なぜこの本がこんなにベストセラーになったかについても、マスコミに取り上げられ、朝日新聞ではこのことを「トットちゃん症候群」と言っている。トットちゃん症候群は『窓際のトットちゃん』を出版した講談社ではなく他の出版社から『トットちゃんベストセラー物語』などという本を出版させ、あらゆる角度から、なぜ売れたかを追究している。

本書を取り上げて語彙調査をし、それを比較語彙研究のデータとしようと考えたのは以上のようなことも当然関係する。中でも、その文章のわかりやすさは大切であり、ほとんど話し言葉に近い文章も多彩な語を提供している。私の採用した調査単位で延べほぼ8万語、異なりでほぼ4700語弱の中に、記号句読点（こういうものも全て一単位としている）は2割を越す頻度で用いられている。これが文章の性格に関係しないはずがない。同時に進めた他の作品では、1頁ほぼ250単位の中で、記号等に属するものが僅かに数個、2、3パーセントにしか満たないものとは対蹠的である。これは、差しあたり、最も目立った事を上げてみただけであり、そう簡単には本書の特徴を言うことはできない。本稿では、『窓際のトットちゃん』に関する各種の語彙データを提示する。これに関する分析は、別稿による。

2. 『窓際のトットちゃん』について

本書は、前記の通り、1981年3月5日初版第1刷発行、本稿では、同年7月28日発行の第25刷による。講談社発行。B6版、ハードカバー。本文は後書き19頁を含め、9頁から284頁まで276頁分。1行43字17行、表題は5行取り、大部分は、各章ごとの改頁は無い。あとがきは43字19行。途中13カ所にモノクロームの挿絵(15・27・45・77・116・131・135・159・194・203・221・243・257頁)、4カ所にカラー挿絵がある。カラー挿絵は、48-49、96-97、144-145、192-193頁の間に挟み込まれ、表側に絵、裏白になっている。そのほか、本の表紙、カバー、中扉・目次扉、4頁分の目次、本文の始まる前の8頁にも1頁のカラー挿絵がある。全て、いわさきちひろ氏の作品である。なお、285頁相当部分に黒柳徹子氏の小学校当時の写真と、81年当時の白黒写真が有り、裏面に本書掲載のいわさきちひろの絵の初出目録があり、「いわさきちひろ絵本美術館」の所在地とそこに原画のあることが記されている。

内容は、目次の終わった後、本文が始まる前の1枚の表^{おもて}(頁はないが7頁めに当たる)に「これは、第二次世界大戦が終わる、ちょっと前まで、実際に東京にあった小学校と、そこに、本当に通っていた女の子のことを書いたお話です。」とあるのに代表され、著者黒柳徹子氏(トットちゃん)が、トモエ学園で過ごした小学校での生活と校長先生と友達のことを書いたものである。「はじめての駅」から「さよなら、さよなら」に至る61章と、後書きから成る。本稿のデータとしては、上記7頁の文と目次をのぞき、1章から61章と「あとがき」までを採用する。「あとがき」は、単なる形式ではなく、本文の内容と密接な関係を持つので、あえて、本文と同等に扱った。文庫本(講談社文庫)も内容は同じであるが、さらに、「文庫本あとがき」9頁が加わる。本稿ではこの「文庫本あとがき」は対象から割愛した。

3. データ作成について

3.1 語彙調査単位

語彙調査の単位として、いわゆる「単語」を採用した。しかし、「単語」には定説となる定義が無いという。ただ、そんなことを言っていては始まらない。各種の語彙調査で用いられる調査単位は、国立国語研究所の実施した多くの語彙調査で、いろいろな調査単位を設定したことに示されるように一様ではない。索引作成においても、その立項方針は一定のものがないことは、かつて宮島達夫氏が苦言を呈されたとおりである²¹。調査単位について明言していないものについては、ともかくとして、何よりも一つの索引の中では統一された方針が必要であることは言を待たない。

もっとも、「総索引」は語を捜すのが目的であるから、各種の搜索法があることは長所である。目的の語に到達するのが第一の目的である。語数を勘定するのは総索引の第一目的ではないから、

索引に語数を知る資料としての役割を持たせるのはある意味ではコクかも知れない。勿論工夫次第で、一語を重複して掲出しなくても出来るはずで、私は各種の索引作成において、そのことを実践してきた。同一語は一カ所に、しかも、その構成要素からもその語に到達できるようにという方針で、最初の索引を編んだ(大久保同舟編『道元禅師全集 上・下』筑摩書房 1969・70)。これは総索引とは違いもともと語を数える目的はなかった。その後のいくつかの総索引^{註2}では、必ず語の計量にも役立つものを作ってきた。

それはともかくとして、語彙調査をするためには、なんと言っても、その調査単位が簡単で、誰がやっても間違いのない事が第一だと思う。国立国語研究所の調査単位はよく考えられていて、それはそれで大切であるが、なかなかそれで統一するのは難しい。規定はなるべく簡単でなければ実際の役に立たない。

それではどうすべきか。

言葉にとって発音というものは本質的なものである。記号等を除いて発音のない言葉はない。私はこのことに注目して国文法で言う「文節」を採用した。言語研究の世界ではあまり注目されない、と言うよりは、古びた、旧式なものという受け取り方もあるというが、発音を第一とした「文節」という単位は有用である。橋本進吉氏が提唱し、学校文法にも採用されたこの概念を語彙調査で活用する(ただ、この「文節」という名称はあまり適切ではないと思う。確かに、「文」の方から見るとその「節」になるものであるが、「単語」の方から見たときは、必ずしもびったりのイメージを持たない。韓国語で似たような単位として、「語節」というものがあるが、むしろこちらに一籌を輸する)。

語の単位として「文節」を基本とすることについては、すでに述べたことがある^{註3}。詳細はそれに譲るが、簡単に言えば、最初、対象資料を「文節」に分割し、その中から一つの自立語と、0個以上の付属語を析出するという方法である。「文節」は自然の発話において、それ以上短くすれば、意味を失うか、別の意味になってしまうという限界である。あくまで「自然」ということが大切だが、そもそも懐疑論者はその「自然」とは何かを問うであろう。それに対しては、あくまで「自然」な発話と言うより他はない。そうすると、すでに、文節否定の論拠にもされる付属語から始まるような「自然」な発話がある。これに関しては、拙著の中に例示した。現代語においても、引用を示すいわゆる格助詞「と」にはこの性格が著しい。自立語(この場合、引用する内容)と「と」は切り離されて、自然に発話されることはしばしばである。あまつさえ、その「と」が文頭にさえ立つ。「と、おっしゃいますと？」などと言う最初の「と」である。これは、この語の語源を示していると思われる。「とやかく」「ともかく」「とやこうや」「とにかく」「とまれかくなれ」などという句中に現れる「と」であり、本来自立要素であって、引用内容をもう一度「と」と繰り返しているのだと考えられる。従って、すでに枕草子などで非難しているいわば「と抜き」の発話は^{註4}本当はそれでいいのであり、「と」はむしろ重言なのである。勿論、こういう語源解釈もされているが、広く受け入れられている考えとも言えない。しかし、こう考え

ると、当面、その「と」をどうするか（「と」自体を自立語とするか、自立語0の文節とするか）は問題として残るけれども筋は通ると考えられる。学校文法はその理論的な面について問題があり、助動詞等についての扱いも問題にはなるが、広くコンセンサスは得ていると思うので、大体それに依りつつ、若干の補正をして調査単位とする。

3. 2 コード付けの基準

さて、完成した『窓際のトットちゃん』語彙表の一々の語にコード付けをするに際して、調査単位となった一つ一つの一つの「単語コード」(WC)を与え、それぞれの語に構成要素ごとに「語素コード」(LC1・LC2…)を与える。

コード付けの基準の大本は『分類語彙表』1964（国立国語研究所）である（勿論、これは、『分類語彙表』でなければならないのではない。何らかの意味の一覧表に依ればよいのである。ただ、『分類語彙表』は広く使われていることのほか、実際によく出来ているという利点がある）。『分類語彙表』を基本として、さらに、他言語の語彙との比較を念頭に置き、文法質に対してもコードを与え、記号等、いやしくも何らかの「形」を持つものに対しては何らかのコードを与えることを根本の方針とする。ただし、敬意を含む語に対して、語形を特定する事は出来ないが、その敬意という意味に対して「.3590」、可能動詞等の可能の意味に対して「.1310」を、いわゆる「形」は特定できないが与えている。存在する意味にコードを与えているのであり、その語のどの部分に与えたかと聞かれても、そう答えるほか無い。用言語尾などについても、本来語源的には、「意味」のある成分が付着して成立したものであろうと考えられるが、現在、その大部分はその意味を特定できない。それ故、用言語尾としてのコードを与えておくにとどめる。その他、詳細については、数次の改訂を加え現在「コード付けの基準—単語コードと語素コード・比較語彙論のために（その5）—」^{註5}に至っている基準を参照されたい。この第5バージョンにもすでに改訂の必要がある項目もあるが、今後を期する。

4. 『窓際のトットちゃん』語彙データ

4. 1 全体の数値データ

異なり語数（記号等を含む）	4684 語
（記号等を除く）	4658 語
延べ語数（記号等を含む）	78252 語
（記号等を除く）	61116 語
一語平均使用回数（記号等を含む）	16.706
（記号等を除く）	13.121
度数階級（記号等を除く）	161

	(記号等を除く)	153
C50 ^{註6}	(記号等を含む)	17 語
	(記号等を除く)	32 語
C60	(記号等を含む)	38 語
	(記号等を除く)	68 語
C70	(記号等を含む)	88 語
	(記号等を除く)	147 語
C80	(記号等を含む)	225 語
	(記号等を除く)	351 語
C90	(記号等を含む)	740 語
	(記号等を除く)	1007 語
度数1の語数	(記号等を含む)	2272 語
	(記号等を除く)	2266 語
度数1の語の異なり語率 (記号等を含む)		48.505%
	(記号等を除く)	48.647%
度数1の語の使用率 (記号等を含む)		2.903%
	(記号等を除く)	3.708%

4.2 使用率上位の語(記号等を除いた C60 の語を示す)

表1 使用率上位の語

順位	語	度数	コード	備考
1	、	10948	16	
2	た(助動詞)	3612	9.0020	
3	の(格助)	3156	8.0010	
4	。	2774	16	
5	に(格助)	2720	8.0010	
6	て(接助)	2655	8.0040	
7	は(係助)	2376	8.0030	
8	を(格助)	1864	8.0010	
9	と(格助)	1549	8.0010	
10	が(格助)	1506	8.0010	
11	だ(助動詞)	1449	9.0050	
12	「	1016	16	合点符号を含む
13	」	996	16	
14	トットちゃん	860	15.2340	
15	も(係助)	742	8.0030	

16	いる (補)	729	10.1200	「…ている」
17	の (準体)	645	8.1000	
18	事	602	1.1010	
19	する (サ変)	555	8.0010	
20	で (格助)	528	2.3420	
21	言う	505	2.3120	
22	ない (助動詞)	497	9.1200	
23	成る	441	2.1220	
24	いう (形)	407	10.3120	「…という～」
25	!	376	16	
26	から (格助)	367	8.0010	
27	と (接助)	336	8.0040	
28	其の	321	3.1000	
29	から (接助)	306	8.0040	
30	みんな	292	1.1980	
31	時	263	1.1612	
32	けど (接助)	260	8.1130	
33	か (終助)	256	8.0060	
34	先生	245	1.2440	
35	此の	236	3.1000	
36	?	234	16	
37	其れ	231	1.1000	
38	思う	230	2.3060	
39	子	236	1.2310	
40	様う	224	1.1300	
41	学校	220	1.2630	
42	ん (準体)	220	8.1000	
43	有る	219	2.1200	
44	来る	216	2.1527	
45	校長先生	215	1.2630	
46	ママ	212	1.2120	
47	見る	202	2.3090	
48	ます (助動詞)	212	9.3590	
49	行く	194	2.1527	
50	そして	200	4.1110	
51	(190	16	
52)	186	16	
53	か (副助)	184	8.0070	
54	てる (連語)	181	20.1200	
55	無い	180	3.1200	
56	…	179	16	
57	いい	178	3.1330	

58	です(助動詞)	163	9.0050	
59	まで(副助)	160	8.1990	
60	とか(副助)	147	8.0070	
61	中	147	1.1700	
62	聞く	145	2.3092	
63	で(接助)	145	8.0040	「て(接助)」異形体
64	でも(接続)	145	4.1130	
65	分かる	143	2.3060	
66	ない(補)	137	10.1200	
67	たり(接助)	136	8.0020	「…でない」
68	トモエ	136	15.2630	
69	よ(終助)	133	8.0050	
70	居る	133	2.1200	
71	所	133	1.1700	
72	ながら(接助)	132	8.1620	
73	人	131	1.2020	
74	前	123	1.1760	
75	私	122	2.1200	
76	もう	121	3.1992	
77	や(並列)	114	8.0020	

以上の内、コード16のもの9項目が記号等である。記号等を含めて考えると、38位の「思う」までが、C60語彙である。

この中に、助詞・助動詞が29語含まれる。他の語彙調査とつきあわせた上でなければ確実なことは言えないが、「いる(補)」「事」「する(サ変)」「言う」「成る」「いう(形)」「其の」「時」「此の」「其れ」「思う」「よう」「有る」「来る」「見る」「行く」「そして」「無い」「いい」「中」「聞く」「分かる」「ない(補)」「居る」「所」「人」「前」「私」「もう」など、4割強の29語は、ほぼ現代日本語の「基幹語彙」²⁷⁾ と言って良いもので、残りの10語中「トットちゃん」「皆んな」「先生」「子」「学校」「校長先生」「ママ」「てる(連語)」「トモエ」の9語は『窓際のトットちゃん』のキーワード(基調語彙)的語彙だろう。ほかの「でも(接続)」はどうか、上に上げた中でも「みんな」「子」「そして」の3語は、もう少し他の語彙表と比較しなければ、簡単には判断できない。

4.3 『窓際のトットちゃん』語彙の意味分野別構造

4.3.1 単語コードによる構造

出現コード数 869コード、その内、『分類語彙表』枠内のコード779、新設コード90²⁸⁾次に、大分類²⁹⁾によるそれぞれの項目数(異コード数)、所属語数、その延べ数を示す。

表2 大分類による意味分野別構造

分類	コード数	所属語数	延べ語数
1.1	129	683	5064
1.2	51	346	3165
1.3	150	586	1906
1.4	72	398	1643
1.5	52	318	1595
1 計	454	2331	13373
2.1	70	500	3494
2.3	109	511	4375
2.4	1	1	2
2.5	16	54	159
2 計	196	1066	8030
3.1	54	384	4319
3.2 ^{注10}	2	2	2
3.3	36	207	735
3.5	17	87	261
3計	109	680	5317
4.1	7	37	756
4.3	13	101	432
4計	20	138	1188
1~4 計	779	4215	27908
5.	1	1	1
6.	1	1	1
7.	8	11	19
8.	13	104	22159
9.	11	30	6684
10.	15	43	1866
13.	1	2	2
14.	0	0	0
15.	29	233	2013
16.	2	27	17138
20.	9	17	461
5~20 計	90	469	50344

この大分類において、語数順、延べ語数順にしたのが、表3-1・表3-2である。自立要素だけで見れば(1.1~4.3以下の語と、15)所属語数順では、1.1(抽象的關係)、1.3(活動)、2.3(活動)、2.1(抽象的關係)、1.4(生産物・道具)、3.1(抽象的關係)、1.2(活動主体)、1.5(自然及び自然現象)の各項が300語を超える。延べ語数順では1.1(抽象的關係)、2.3(活動)、3.1(抽象的關係)、2.1(抽象的關係)、1.2(活動主体)、15(固有名詞)、1.3(活動)、10(補助用言)、1.4(生産物・道具)、1.5(自然及び自然現象)の各項が1000語を超える。

表 3-1 単語コード大分類集計 (語数順)

分類	Wcの例	語数	度数	語例
1.1	1.1113	683	5064	訳
1.3	1.3071	586	1906	イメージ
2.3	2.3030	511	4375	笑う
2.1	2.1521	500	3494	渡る
1.4	1.4151	398	1643	ホック
3.1	3.1550	398	4343	ぼっかり
1.2	1.2000	346	3165	我
1.5	1.5520	318	1595	林檎
15	15.2340	234	2015	ワット
3.3	3.3451	195	713	ジェントリー
8	8.1000	104	22159	の(準体)
4.3	4.3000	101	432	わははは
3.5	3.5030	87	261	わあーん
2.5	2.5810	54	159	薺る
10	10.1200	43	1866	らっしゃる(縮約・補)
4.1	4.1160	37	756	尤も
9	9.1200	30	6684	ん(否定)
16	16.0000	26	17136	♪
20	20.5050	17	461	苦がきゃ(連語)
7	7.1140	11	19	らしい(接尾)
13	13.0000	2	2	the
2.4	2.4280	1	2	包帯する
5	5.3590	1	1	御ん
6	6.1990	1	1	ー(挿入辞)

表 3-2 単語コード大分類集計 (度数順)

分類	Wcの例	語数	度数	語例
8	8.1000	104	22159	の(準体)
16	16.0000	26	17136	♪
9	9.1200	30	6684	ん(否定)
1.1	1.1113	683	5064	訳
2.3	2.3030	511	4375	笑う
3.1	3.1550	398	4343	ぼっかり
2.1	2.1521	500	3494	渡る
1.2	1.2000	346	3165	我
15	15.2340	234	2015	ワット
1.3	1.3071	586	1906	イメージ
10	10.1200	43	1866	らっしゃる(縮約・補)
1.4	1.4151	398	1643	ホック
1.5	1.5520	318	1595	林檎
4.1	4.1160	37	756	尤も
3.3	3.3451	195	713	ジェントリー
20	20.5050	17	461	苦がきゃ(連語)
4.3	4.3000	101	432	わははは
3.5	3.5030	87	261	わあーん
2.5	2.5810	54	159	薺る
7	7.1140	11	19	らしい(接尾)
2.4	2.4280	1	2	包帯する
13	13.0000	2	2	the
5	5.3590	1	1	御ん
6	6.1990	1	1	ー(挿入辞)

整数部分をはずして小数点以下一桁の「意味分野」によってみよう。この場合、コード5以上の項目は、15の固有名詞を除いていわゆる「意義質」ではなく「文法質」の性質が強いので、除外して考えなければならない。但し、コード5以上の項目でも、小数点以下一桁部分が0でないものは意味を認める事が出来るので集計する。

所属語数順で見ると、.1(抽象的關係)、.3(活動)、.2(活動主体)、.5(自然及び自然現象)、.4(生産物・道具)の順となる。延べ語数順でも同じである(表4)。ただ、コード数は、.3が一番多く、.1、.5、.4、.2の順になる。

どの観点で見ても、抽象的關係の意味分野に属するものが多い。

表4 意味分類別所属語数

意味分類	コード数	所属語数	出現度数
記号類	1	28	17138
.0	14	105	26375
.1	291	1675	17487
.2	66	556	5041
.3	329	1447	8436
.4	76	402	1648
.5	91	471	2127
合計	868	4684	78252

4.3.2 語素コードによる構造

次に、語素コードによって集計する。

同じく小数点一位までの分類で一覧表を作成する(→表5)。

表5 語素コードによる意味分類統計

意味分類	異コード数	語数	度数	該当語素数を含む語例
.1	479	2643	24281	一級
.2	123	629	5607	和田誠さん
.3	330	1811	11805	する(サ変)
.4	179	434	1806	廊下
.5	151	775	3961	陸軍
5.0	2	7	9	間違い
6.0	2	14	392	みんな
7.0	14	338	3286	っ(末尾音)
8.0	37	200	23032	を(格助)
9.0	19	233	6963	よう(助動詞)
12.0	2	1090	9922	似合う
13	1	2	2	the
16	1	30	17142	=
合計	1340	8206	108208	

ここでも、.1(抽象的關係)は、コード数でも、所属語数でも、延べ度数でも一番多い。次に、延べでは、助詞類・記号・.3(活動)、12.00902(用言語尾)、助動詞、.2(活動主体)、.5(自然及び自然現象)の順に多く、コード数順では1、.3、.4、.5、.2、所属語数では.1、.3、12.、.5、.2、.4、7.、9.、8.の順となる。^{註1)}

高頻度で出現するコードを示す。語数の順では、12.00902(動詞用言語尾)、7.342(複合サ変動詞語尾)、17.1950(漢字語素・数字)、5.3590(敬語接頭辞)、9.00503(形容動詞語尾)、7.0030

(形容詞語尾)、15.2340 (固有人名)、.3590 (敬意)、7.2020 (接尾辞「さん」「様」等)、7.0050 (造副詞接尾辞) が上位 10 項目である。その語素を含む語数は合計 8206 語¹²、その延べ度数は 108208 語になる (→表 6-1)。

表 6-1 高頻度出現コード (語数順10位まで)

Lc	語数	度数	漢字表記
12.00902	1088	9923	笑う
7.3420	227	509	わくわくする
17.1950	204	684	六ヶ月
5.3590	195	654	晩御飯
9.00503	188	643	禄だ
7.0030	155	1620	悪い
15.2340	112	1434	ワット
0.3590	104	565	リボン屋さん
7.2020	95	299	六人姉妹
7.0050	70	586	よくよく

一方、度数順では、16 (記号類)、8.0010 (格助詞)、12.00902、8.0040 (接続助詞)、9.0020 (助動詞三類¹³)、8.0030 (係助詞)、7.0030 (形容詞語尾)、9.0050 (助動詞別類)、15.2340 (固有人名)、3.1000 (こそあど)、7.0030 (形容詞語尾) である (→表 6-2)。

表 6-2 高頻度出現コード (度数数順10位まで)

Lc	語数	度数	漢字表記
16	30	17142	スコッチ・テリア
8.0010	18	11936	の (準体)
12.00902	1088	9923	笑う
8.0040	22	3894	初めて
9.0020	5	3774	大した
8.0030	10	3159	やら (副助)
7.0030	151	1620	悪い
9.0050	3	1615	です (助動詞)
15.2340	112	1434	ワット
3.1000	47	1351	他

同じものは、12.00902・15.2340 の 2 項目だけである。

4.3.3 度数 1 の語のデータ

4.3.3.1 単語コードによる

度数 1 の語は 2272 語ある。単語コードは 735 に及ぶ。

単語コードで、度数 1 の語が最も多く集まるコードは、15.2340 (固有名詞) 59 語、1.2410 (専門的・技術的職業) 36 語、1.1960 (単位・等・回・個・条) 34 語、15.2630 (固有社寺名・

学校名) 30 語、1.1962 (年・時・ワット・馬力等) 20 語、1.5520 (植物名) 17 語、2.3010 (気分・情緒) 16 語、1.3081 (方法) 15 語、3.1820 (形・平たい・丸い) 15 語、3.3010 (驚き・楽しい・快い) 15 語、1.5020 (色) 14 語、15.2590 (固有地名) 14 語等であり、多彩である (表 7-1)。

一方、1 語しか所属しないコードは 264 個、以下、2 語 174 個、3 語 95 個、4 語 66 個、5 語 41 個、6 語 36 個、7 語 18 個、8 語 10 個、9 語 5 個、10 語 5 個、11 語 5 個、12 語 3 個、13 語 1 個である。

4.3.3.2 語素コードによる

度数 1 の語の語素コードは全部で 5402 個あり、異なりコード数は 1000 である。出現回数上位のものは、12.00902 (用言語尾) 508 個、17.1950 (漢字語素・数) 161 個、7.3420 (複合サ変語尾) 139 個、9.00503 (形容動詞語尾) 99 個、5.3590 (敬語接頭語) 90 個、15.2340 (固有人名) 60 個、7.0030 (形容詞語尾) 55 個、7.2020 (人を表す接尾辞) 52 個、.3590 (敬意) 48 個、17.3060 (漢字語素・思考) 43 個、7.0050 (造副詞接辞) 39 個、7.1960 (助数詞) 37 個、7.1300 (様態接尾辞) 35 個、7.1990 (強意接尾辞) 34 個、15.2630 (固有学校名) 33 個、2.1531 (込む) 31 個、6.1990 (強意接中辞^{註14}) 31 個、.1310 (可能の意味^{註15})・2.3393 (手の動作) 30 個である (表 7-2)。

この度数 1 の語を、小数点一位で意味の大分類を施すと、.1 (抽象的關係) が 363 (2079) とコード数、個数とも一番多い。以下、コード数順に配列すれば .3 (活動) 256 (1319)、.4 (生産物・道具) 131 (266)、.5 (自然及び自然物) 122 (479)、.2 (活動主体) 84 (416)、0 (文法質) がコード数 42 (延べ個数 833)、その他、13 と 16. となっている。

表 7-1 度数 1 の単語コード別語数

Wc	度数	語 例
15.2340	59	ワット
1.2410	36	六年生
1.1960	34	四日目
15.2630	30	ローチェスター大学
1.1962	20	六ヶ月
1.5520	17	林檎
2.3010	16	満足する
1.3081	15	よそい方
3.1820	15	緩い
3.3010	15	楽々
1.5020	14	四十八色
15.2590	14	緑が丘

表 7-2 度数 1 の語素コード別語数

Lc1	度数	語 例
12.009020	508	笑い合う
17.195	161	六ヶ月
7.342	139	連発する
9.005030	99	禄だ
5.359	90	御迷惑
15.234	60	ワット
7.003	55	忘れっぽい
7.202	52	六人兄妹
0.359	48	らっしゃる (縮約)
17.306	43	劣等意識
7.005	39	ぼっかり
7.196	37	六ヶ月
7.13	35	柔らかだ
7.199	34	弱さ
15.263	33	ローチェスター大学
2.1531	31	呼び込む
6.199	31	横っちょ飛び
0.131	30	揺ずれる
2.3393	30	遣う繰り

5. おわりに

本稿では、『窓際のトットちゃん』語彙の数量的データを示すと共に、意味分野別の構造を数量的に示した。これだけでは何も分からない。実際にコード付けしたリストはすでに発表してある²¹⁶。それに若干の補正を加えたものである。勿論まだいろいろミスがあり、なかなか完全な形で提示できないのは残念であるが、後日、このデータを基に、何らかの対照資料と比較することによって、『窓際のトットちゃん』語彙の性質を明らかにしようとして計画している。

不完全な内容であるが、以上によっても、語彙の数量的一般性は示されたと思う。今回は、品詞等による分析は一切行わなかったが、『窓際のトットちゃん』語彙の性格の一端、その表現法などはほの見えたであろう。

注

注1. 宮島達夫「総索引への注文」『国語学』76 1969

注2. 田島毓堂編『源氏物語絵巻詞書総索引』昭和45年11月(東海学園国語国文学会 B5、232頁、平成6年3月 汲古書院から再版 A5 380頁)

同「十六夜日記語彙表(上)」[近藤喜美子らと共編] 昭和48年12月 東海学園国語国文5号(pp55-65)

同「十六夜日記語彙表(下)」[近藤喜美子らと共編] 昭和49年3月 東海学園国語国文6号(pp57-67)

同「蒙古襲来絵詞語彙表」[川口孝子らと共編] 昭和50年3月 東海学園国語国文7号(pp59-74)

同「当麻曼陀羅縁起絵詞総索引」[加藤和子らと共編] 昭和50年3月 東海学園国語国文8号(pp59-70)

同「西行物語絵巻詞書本文並びに総索引」[加藤和子らと共編] 昭和51年3月 緑園11号(東海学園女子短期大学校友会誌)(pp1-21)

同「吉備大臣入唐絵詞・吉備大臣物語本文並びに総索引」[柿平五三子らと共編] 昭和51年3月 東海学園国語国文9号(pp145-168)

同「男衾三郎絵巻詞書本文並びに総索引」[平野由美子と共編] 昭和51年9月 東海学園国語国文10号(pp107-126)

同「国宝粉河寺縁起絵巻詞書本文並びに総索引」[西岡真佐子と共編] 昭和52年3月 東海学園国語国文11号(pp112-121)

同「華厳縁起絵巻詞書本文並びに総索引」[田口陽子らと共編] 昭和52年8月 東海学園国語国文12号(pp53-97)

同「長谷雄草紙詞書本文並びに総索引」[丹羽一恵と共編] 昭和52年8月 東海学園国語国文12号(pp98-107)

同「絵師草紙詞書本文並びに総索引」[丹羽一恵と共編] 昭和52年10月 『熊谷武至教授古稀記念国語国文論集』(笠間書院)(pp447-462)

同「豊明絵草子詞書本文並びに総索引」[中村悦子と共編] 昭和53年3月 東海学園国語国文13号(pp184-207)

同「葉月物語絵巻詞書本文並びに総索引」[土方弘美と共編] 昭和52年3月 東海学園国語国文14号(pp87-100)

同『正法眼蔵随聞記語彙総索引』[近藤洋子と共編] 昭和56年8月(法蔵館 B5、573頁)

同『栄花物語語句索引』[松村博司・進藤義治と共編] 昭和61年11月(名古屋大学出版会 A5、553頁)

注3. 拙著『比較語彙研究序説』1999 付章第2「語の単位と定義」

注4. 枕草子「ふと心おとりとかするものは」の段。ただ、ここで言う例は、引用の「と」とはちょっと違

うかも知れないが、本質的に通じる所があるように思う。関西方言で「行こうと思う」を「行こう思う」などと言うのと通じると思う。こういう「と」を省いても言葉は通じる。

注5. 田島毓堂「コード付けの基準—単語コードと語彙コード・比較語彙論のために(その5)—」『名古屋大学文学部研究論集 文学47』2002

注6. C50、C60等については、田島毓堂稿「語彙指標—語彙の数量的側面と語彙研究への視点—」『日本語研究と日本語教育』名古屋大学出版会 1992

同「古典作品の特徴語彙の分析のための一試論—C50とD10と—(その一)」平成5年10月『日本語学』12-10(明治書院)(pp108-118)

同「古典作品の特徴語彙の分析のための一試論—C50とD10と—(その二)」平成5年11月『日本語学』12-11(明治書院)(pp115-123) 参照

なお、この『窓際のトットちゃん』語彙の度数別統計表を参考表として掲げた。

注7. 林四郎「語彙調査と基本語彙」『電子計算機による国語研究Ⅲ』1971

注8. 注5の基準参照。

注9. 『分類語彙表』では1から4の整数が、その語の品詞的性質を表す。1は体言、2は用言、3は相言(形容詞、形容動詞、連体詞、副詞の一部)、4はその他で、感動詞、接続詞、陳述副詞等である。小数点1位で纏める仕方を「大分類」と称する。

注10. 『分類語彙表』に3.2というコードはあり得ないが、我々の「コード付けの基準」による単語コードには、若干『分類語彙表』には出現しないコードが出現する。当面そのままにしておく。

注11. 7(接尾辞)、9(助動詞)、8(助詞)の語数には、それを構成要素として含む語が入っている。

注12. この語数は総異なり語数より大きいが、それぞれの語彙を構成要素として含む語に重複のあるからである。

注13. 『岩波古語辞典 基本助動詞解説』1974による。ここに示されるのは古典語の助動詞であるが、それに準じて「た」を第3類助動詞に分類した。9.005はそこで「別類」とされる助動詞で、断定の助動詞「だ」が中心である。

注14. 拙稿「日本語接中辞論—語彙論の基盤「コード付けの中から」—」『名古屋大学文学部研究論集 文学48』2002

注15. 「可能の意味」とは、コード付けの基準に示してあるが、「可能動詞」に付けたコードである。例えば「書ける」ならば「書く」のコード2.3150と用言語尾12.00902の他、可能の意味の存在に対して「.1310」のコードを与える。

注16. 拙編「『窓際のトットちゃん』コード付き語彙表<評価版>」『比較語彙研究の試み10』2003