

## 書き言葉と話し言葉の語彙分析

高橋 渉

### 1. 目 的

文字言語と音声言語の間には、その中に現れる語彙そのものや、また出現する語彙の使用頻度にも相違が見られることは従来から知られている。さらに使用語彙のセグメント数や、その語彙に含まれる音素の種類等、様々な点で相違が見られることも想像に難くない。

本稿では英語の書き言葉 (Written form) と話し言葉 (Spoken form) との間にはどのような語彙論上ならびに音韻論上の差異が見られるのかを5つのそれぞれ異なったタイプのデータから得られた使用語彙表ならびに使用語彙頻度順一覧表をもとに、パーソナルコンピュータを援用した計量分析を行ない、比較検討することを目的とする。

なお通常の方法でパソコンに入力したテキストデータから一つ一つの英単語を切り出し、アルファベット順にソートしたり、出現頻度数順にソートする方法は高橋 (1992) において詳細な記述があるので、本稿においては、パソコンを用いたテキストの分析の方法については最低限の記述にとどめることにする。

### 2. 使用するデータ

文字言語と音声言語の比較分析のために用いるデータは次の5種類である。

文字言語を代表するものとしては小説と科学論文の2種を選んだ。小説は現代アメリカの小説家 Raymond Carver による短編小説 “The Bath” を用いる。この短編小説は使用総語数2,354語で、そのうち会話を描写した部分 (すなわち引用句内の単語) は全体の20%にあたる471語である。

もう一種の文字言語データとして Kuhn (1962) の第1章 pp. 1-6までの2,009語を使用する。同書は経験科学における「パラダイム」という概念の重要性を説いたもので、同じ文字言語という範疇には入るものの、前記の小説とは使用語彙の性質が極めて異なることが予想される科学論文である。

音声言語のデータとしては以下に述べる3種のデータを使用する。いずれも実際に行われた会話を録音し、ついでその内容を文字に転記したものである。

Marckwardt & Quirk (1964) はアメリカとイギリスを代表する言語学者 Marckwardt と Quirk がアメリカ英語とイギリス英語の共通性を語り合った対談形式の会話である。BBC と VOA によってラジオで放送された内容が文字に転記されたものであるが、後述の2つのデータと非常に異なる点は、彼らの会話は極めて周到な準備のもとに行われた、日常の我々の会話とはやや異なる非常にアカデミックな内容の会話であることであろう。また

uh, eh, um など自然な会話の中では必ず生ずるであろう間投詞類も文字には転記されておらず、転記ののち若干の修正も加えられているという。本稿では第1章の Two languages or one のうち冒頭の彼ら以外の人物による一文(41語)のコメントを除いた全文を使用することとする。

最後に Carterette & Jones (1974) について説明しよう。同書はアメリカの小学校1年生(ほぼ6歳)、3年生(ほぼ8歳)、5年生(ほぼ10歳)、さらに成人(短期大学学生)のそれぞれの話者が、同じ Age-group の中で3人ずつが、インフォーマルなフリーディスカッションを行ったものを録音し、その全ての内容を逐語的に音声表記とアルファベットに転記したものである。前述の Marckwardt & Quirk (1964) と異なり、um, uh, などの言いよどみや、言い誤り、不要な繰り返しなども全て削除されずに Corpus の中に残されており、その意味ではまさにこのデータこそが、我々の日常の会話行動の記録といえる。

本稿では前述の4つの Age-group のうち、1年生の Corpus と成人の Corpus のそれぞれ一部をデータとして使用する。

以上述べた4つのデータは(1)のように略記して以後言及することとする。

(1)	データ名	略号
a.	Carver(1981)	W—NOVEL
b.	Kuhn(1962)	W—KUHN
c.	Marckwardt & Quirk(1964)	S—M & Q
d.	Carterette & Jones(1974)	
	成人の会話部分	S—ADULT
e.	Carterette & Jones ( <i>ibid.</i> )	
	小学校1年生の会話部分	S—CHILD
	(但し略号の語頭の W, S はそれぞれ Written form, Spoken form を示す)	

また上記データに生じる語彙の数は、なるべく同一のほうが、様々な分析の上で好ましいのであるが、実際上は同一に整えることは不可能であるので、S—CHILD に現われる単語数2,912語を最高に、下記の通りの総語数となった。従って以下の分析には、百分率による分析を行うこととする。

(2)	各データの総語数	
a.	W—NOVEL	2,354語
b.	W—KUHN	2,009語
c.	S—M & Q	1,886語
d.	S—ADULT	2,465語
e.	S—CHILD	2,912語

## 3. 分 析

## 3.1 Contraction

話し言葉と書き言葉との相違の一つとして従来から知られている Contraction(縮約)の出現率の違いについて考えてみよう。

Contraction とは概略(3)に見られるような音韻過程である<sup>(1)</sup>。

(3)

a.	is, has	→ [z]	he is	→ he's
b.	would, had	→ [d]	I would	→ I'd
c.	have	→ [v]	I have	→ I've
d.	am	→ [m]	I am	→ I'm
e.	are	→ [r]	We are	→ We're
f.	will	→ [l]	I will	→ I'll
g.	not	→ [nt]	is not	→ isn't

これらの Contraction はうちとけた話し言葉の音韻的特徴の一つにあげられるものである。我々のデータにおいてこれらの Contracted forms がどのような分布を示すのか、計量分析をしてみよう。

Contracted forms は文字としては「'」(=apostrophe)によって具現化されるので、データ中にどのくらいの「'」が生じるか検索を行ってみよう。データをそのまま入力して作成した MS-DOS 形式のテキストファイルを高橋(1992)に示した方法で、一語ずつの英単語に切り出したファイルから「'」が含まれる単語を取り出してみよう<sup>(2)</sup>。

(4) 縮約形の出現回数と割合

	'の出現回数	属格	縮約形	%
W—NOVEL	65	11	54	2.29
W—KUHN	5	5	0	0
S—M & Q	43	2	41	2.17
S—ADULT	112	0	112	4.54
S—CHILD	124	4	120	4.26

Contracted forms の出現率(百分率)では、S—ADULT と S—CHILD における縮約表記の出現率は W—NOVEL と S—M & Q のほぼ2倍にのぼることがわかる。また W—KUHN には一例の縮約形も出現せず、Spoken form であるはずの S—M & Q にも W—NOVEL 程度にしか現れないことがわかる。

さらに W—NOVEL は、その性質上登場人物の会話を描写した部分と、地の文とに二分

されるが、W—NOVEL 全体に生じる縮約表記54例中、50例が会話部分に生起している。これは全生起回数92.6%にあたる。前述の通りW—NOVEL の会話部分に生じる総語数は471語であり、地の文の総語数は1,883語であるのでW—NOVEL の会話文に占める縮約表記の割合は10.6%にものぼり、我々の全データ中の最大の出現率となることが特筆される。

対談形式で行われた会話文であるS—M & Qにおける縮約形の生起率がW—NOVEL 全体のそれと大差ないのは、S—M & Qが、我々が行ううちとけた友達同士の日常的会話と性質をかなり異にしているものだからであろう。

表(4)に見られる縮約表記の出現率の差異から、縮約形はくだけた会話文の特徴を構成する大きな要因であることがわかる。また本来作家が想像の上で作ったはずの小説の会話に縮約形が全データ中最大の割合（一割強）で生ずるのも、くだけた会話に Contracted formが多いという事実を知っている作家が、会話部分を作り上げる際に無意識のうちにその特色を強く表現した結果ではないかと考えられる。

### 3.2 語彙の長さ

高橋（1991）では、小学生の作文使用語彙に現れる様々な音素配列論上の制約を論じているが、その中で、小学生の作文使用語彙のうち、高頻度使用語彙群と低頻度使用語彙群との間にはセグメント数に明確な差異が存在することがわかった。すなわち、高頻度使用語彙群の一語当たりの平均セグメント数は低頻度使用語彙群のそれに比べて、全体の平均で42%も少ないのである。

このように、高頻度に使用される語彙の方が、一般にセグメント数（ここではほぼ単語に含まれる文字数と同義）が少ないという事実は Statistical linguistics（統計言語学）でしばしば言及される「Zipfの法則」に従ったものである。高橋（*ibid.*）は前述のように小学生の作文、すなわち日本語の Written language というデータをもとに計量分析したものであるので、本稿では Written form と Spoken form のそれぞれを含む5つのデータの中で、文字数ごとの単語数はどのように分布するのかを、コンピュータを援用した計量分析で調査してみよう。

分析結果を示す前に、島岡他（1982：493）所収の研究社中英和辞典に含まれる7,945語の文字数別分布を表(5)に示す<sup>(3)</sup>。

(5)

文字数	語数	%	文字数	語数	%	文字数	語数	%
1	3	0.0	6	1,423	17.9	11	286	3.6
2	36	0.5	7	1,230	15.5	12	137	1.7
3	338	4.2	8	955	12.0	13	62	0.8
4	1,025	12.9	9	667	8.5	14	23	0.3
5	1,254	15.8	10	503	6.3	15	2	0.0

上記(5)表に示した通り、7,944語中最も多いのは6文字語で、全体の17.9%を占める。ついで5文字語(15.8%)、7文字語(15.5%)、4文字語(12.9%)等と続く。そして1文字語から6文字語までの合計は全体の半数を超える約51%となることがわかる。

次に我々のデータにおける分布を考えよう。

(6)

文字数	W_NOVEL		W_KUHN		S_M&Q		S_ADULT		S_CHILD	
	生起回数	%	生起回数	%	生起回数	%	生起回数	%	生起回数	%
1	71	3.02	34	1.69	59	3.13	187	7.59	238	8.17
2	350	14.87	354	17.62	387	20.52	460	18.66	623	21.39
3	735	31.22	336	16.72	389	20.63	490	19.88	615	21.12
4	490	20.82	297	14.78	333	17.66	599	24.30	726	24.93
5	256	10.88	212	10.55	219	11.61	283	11.48	319	10.95
6	204	8.67	115	5.72	148	7.85	169	6.86	173	5.94
7	117	4.97	171	8.51	152	8.06	114	4.62	106	3.64
8	74	3.14	117	5.82	87	4.61	56	2.27	61	2.09
9	38	1.61	96	4.78	47	2.49	42	1.70	31	1.06
10	12	0.51	134	6.67	28	1.48	42	1.70	9	0.31
11	4	0.17	64	3.19	21	1.11	13	0.53	6	0.21
12	3	0.13	41	2.04	8	0.42	7	0.28	3	0.10
13	—		18	0.90	7	0.37	1	0.04	2	0.07
14	—		10	0.50	1	0.05	2	0.08	—	
15	—		4	0.20	—		—		—	
総語数 (A)	2,354語		2,009語		1,886語		2,465語		2,912語	
総文字数 (B)	9,393文字		10,766文字		8,210文字		9,714文字		10,593文字	
一語あたりの平均文字数 (=B/A)	3.99文字		5.36文字		4.35文字		3.94文字		3.64文字	

表(6)が、我々の5つのデータ別に1文字語から15文字語までの生起回数と割合を提示したものである。但しW—KUHNには16文字語が4、20文字語が1、27文字語が1、生じるが作表上の都合から表(6)には含めなかった<sup>(4)</sup>。以下に表(6)からわかる事実を順に指摘し、それらに基づく考察をしよう。

まず各データごとの1文字語の生起回数とその割合を考えてみよう。

W—KUHNにおいてそれぞれ一回のみ生起するローマ数字“V”と“X”を別にする、

全データに生ずる一文字の英単語は一人称単数主格の代名詞 I と不定冠詞 A/a のみである。一文字語が生じる割合は、全データ平均で5.07%であるが、これに比してW— KUHN では極端に低い1.69%を占めるのみであるのに対し S— ADULT では7.59%、S— CHILD では最高の出現率である8.17%を占め、それぞれ平均をはるかに超えている。

一文字語のこのような出現率の差異は何に起因するのか考えてみると、これは前述の I と A/a のうち一人称単数主格代名詞 I の生起率の差が主たる要因であることがわかる。

(7)	I の生起回数	全一文字語に対する割合
W— NOVEL	12	16.9%
W— KUHN	0	0.0%
S— M & Q	24	40.0%
S— ADULT	142	75.9%
S— CHILD	169	71.0%

表(7)から明らかなように、人称代名詞 I は S— ADULT や S— CHILD のようにくだけた日常会話の中に頻出することがわかる。また W— KUHN に 1 例も出現しないのはこれが科学論文という特殊な範疇に属するデータであるためであろう。但し、一人称単数目的格の代名詞 me に関しては、出現率にそれほどの差異は認められず、W— NOVEL 3 例、W— KUHN 無し、S— M & Q 3 例、S— ADULT 6 例、S— CHILD 9 例と極端なばらつきは見られなかった<sup>(5)</sup>。

以上の事実から、各データにおける一文字語の割合の差は、代名詞 I の生起率の差異によって生じていると言ってよいと思われる。

W— KUHN のデータ全体の一語あたりの平均文字数は、我々の扱う 5 つのデータ中最長の 5.36 文字である。数値は全データの平均文字数 4.91 文字に比べて、かなり高い数値と言える。後述のように、W— KUHN は他のデータに見られないほど長い文字の単語を含むので、平均文字数が長くなるのも当然ではあるが、興味深いのは W— KUHN の文字数別出現回数で最大を示すのは、表(6)から明らかな通り 2 文字語 (354 回, 17.62%) なのである。

では、W— KUHN の 2 文字語の内容に他データと異なる特徴があるのかその内部を見てみよう。

#### (8) 各データ別 2 文字語の頻度順一覧表

(但し生起回数20回以上)

単語	生起回数	単語	生起回数	単語	生起回数	単語	生起回数	単語	生起回数
W— NOVEL		W— KUHN		S— M & Q		S— ADULT		S— CHILD	
he	66	of	100	of	65	to	79	to	127
to	58	to	53	to	56	it	58	in	53
in	42	in	47	in	49	of	35	we	52

it	24	by	22	we	28	in	30	my	50
on	24	is	20	it	24	be	27	go	43
of	22			as	23	so	26	um	40
				is	20	oh	24	he	39
						uh	22	on	29
						do	21	it	27
						go	20		

(8)表からW—KUNH の 2 文字語の多さは、354語の 2 文字語のうち28.25%を占める of によっていることがわかる。実際次節で見ると、W—KUNH の頻度順単語一覧表でこの of は the について第二位を占めるのである。

頻度20回以上の語をリストした表(8)には現れない S—CHILD の of (18回)を含めての、各データの 2 文字語に占める of の割合は(9)のとおりである。

(9)	生起回数	%
W—NOVEL	22	6.29
W—KUNH	100	28.25
S—M & Q	65	16.80
S—ADULT	35	7.16
S—CHILD	18	2.89

この表から書き言葉、話し言葉を問わず、科学的内容の言葉に of は高頻度に生じるのに対し、子供の話し言葉には極めてまれにしか生じていないことが明らかである。

(8)の頻度順リストから二文字語のうち高頻度に用いられるものは、前置詞、代名詞、接続詞などの機能語がほとんどであることもわかる。そのうち特に前置詞 to はジャンルを問わず一様に高頻度に用いられている。またW—NOVEL で一位となる he は小説特有のもので、地の文の中で66例中62例 (94%) 用いられていることから明らかなように、作者が様々な人物描写を行おうとすれば必ず頻出するであろう単語である。

次いでどのジャンルでも一様に高頻度に用いられた 2 文字語は in であった。

以上のごとく、2 文字語の多くは各データにおよそ類似の単語が生起するが、of に関しては話し言葉、書き言葉という区別よりはデータの内容 (この場合、科学的か日常的か) にその生起率は依存することがわかった。

本節の最後に各データ別の一語当たり平均文字数について考えてみよう。下記の表(10)は参照の便宜上、前掲の表(6)の一部を重複して提示している。

表(10)

	W—NOVEL	W—KUHN	S—M & Q	S—ADULT	S—CHILD	全データ
総語数	2,354	2,009	1,886	2,465	2,912	11,626
総文字数	9,393	10,766	8,210	9,714	10,593	48,676
一語平均文字数	3.99	5.36	4.35	3.94	3.64	4.19

(10)から明らかなように、一語平均文字数の最長のはW—KUHNの5.36文字、最短はS—CHILDの3.64文字で、両者の間には1.72文字の差がある。また話し言葉は書き言葉より一般に短い語が多用されることもわかるが、S—M & QはW—NOVELより平均文字数が多くなっているのは、彼らの会話の内容が前述のようなアカデミックな内容だからであろう。

W—NOVELは地の文と会話文を区別して一語平均文字数を比較すると、会話中では8.87文字、地の文では4.02文字となり、会話文の中で、明らかに短い語が多用されていることがわかる。さらに子供の会話と大人の会話では、子供の会話が大人の会話より7%強、使用する単語の一語当たり平均文字数は短い<sup>(6)</sup>。

以上をまとめると、話し言葉は書き言葉より短い単語を使用すること、また子供の話し言葉は大人のそれよりさらに短い語を頻用することが明確になった。

### 3.3 生起回数順一覧表

本節では高橋(1992)の方法によって作成した各データ別の単語の生起回数順一覧表を提示し、そこに見られる差異を考察することにしよう。

(11)

W—NOVEL	W—KUHN	S—M & Q	S—ADULT	S—CHILD	W—E	S—E
256 the	137 the	87 the	142 I	169 I	the	the
79 and	100 of	65 of	109 you	149 and	of	and
68 said	56 and	56 to	79 to	127 to	to	I
66 he	53 to	49 in	58 it	69 a	in	to
59 a	47 in	41 and	55 the	67 the	and	of
58 to	47 that	36 you	45 a	53 in	a	a
55 she	36 scientific	35 a	45 and	52 we	for	you
42 in	32 a	33 that	40 that	50 my	was	that
35 was	32 science	28 this	39 know	49 you	is	in
30 his	22 by	28 we	39 they	43 go	that	it
29 her	20 is	26 English	35 of	40 um	on	is
24 it	19 are	24 I	30 in	39 he	at	yes
24 on	19 for	24 it	30 there	35 like	he	was
23 doctor	19 these	23 as	30 what	33 get	with	this
22 of	16 as	20 is	27 be	31 then	by	but
20 boy	15 what	17 but	27 like	29 on	be	on



20	man	14	have	16	are	27	yeah	29	was	it	well
19	this	14	they	15	well	26	don't	28	there	an	he
19	up	13	from	14	our	26	so	27	it	as	have
18	mother	13	which	13	for	24	oh	26	know	his	for
16	is	12	at	13	on	24	school	25	went		
16	with	12	be	13	so	22	uh	24	don't		
14	back	12	it	13	use	22	well	22	have		
14	for	12	or	12	about	21	do	22	that		
13	at	12	with	12	all	20	all	22	this		
13	from	11	been	12	be	20	go	22	when		
13	that	10	more	12	have	20	have	20	because		
13	woman	10	not	11	American	19	is	20	swimming		
12	be	10	research	11	British	17	think	20	well		
12	father	10	those	11	do	17	where	18	about		
12	I			11	not	16	going	18	it's		
12	then			11	think	16	it's	18	of		
11	bed			10	he	16	or	18	so		
11	husband			10	or	14	but	17	but		
10	but			10	shop	14	just	17	she		
10	he's			10	was	14	on	16	up		
10	went			10	would	14	then	15	is		
10	what					14	want	15	take		
						14	was	15	they		
						13	did	14	do		
						13	one	13	down		
						12	are	13	pool		
						12	had	13	see		
						12	high	12	all		
						12	how	12	feet		
						12	mean	12	his		
						12	that's	12	I'm		
						12	went	12	or		
						11	really	12	that's		
						10	been	12	uh		
						10	get	11	as		
						10	here	11	can		
						10	I'm	11	his		
						10	my	11	sometimes		
						10	out	10	school		
						10	when	10	swim		
								10	time		
								10	with		

表(11)は我々の5つのデータから10回以上の生起回数を持つ全単語をその生起回数と共に頻度順に列挙したものである。また参考のためにW—E、S—Eとそれぞれ仮称する頻度順語彙表をCrystal (1987: 86) から引用して、右側2列に付加した<sup>(7)</sup>。

我々の5つのデータはW—EやS—Eに比べれば十分な語彙数のサンプルとは言えないの

は当然であるが、それでもいくつか注目すべき点が見られるので、それらを列挙しよう。

まず我々のデータは書き言葉の2つとS—M&Qの、合計3データにおいてtheが頻度順位一位となっているのに対し、S—ADULTとS—CHILDではIが一位となっている。十分なサンプル数を持つW—E、S—Eとも一位はtheなのであるが、S—EにはW—Eには存在しないIが3位に入っているので、Iの頻出は話し言葉の重要な特徴の一つと考えて良いだろう。この点から考えると、S—M&Qは会話体（話し言葉）の形式をとっているとはいえ、極めて書き言葉に内容は近いと言える。

S—M&Qが書き言葉に性質が類似していることを示すもう一つの事実は前置詞ofの多用されることである。W—EでもW—KUHNでもofは第二位の順位を示すが、S—M&Qでも同じくofは第二位の頻度順位を示す。

次に、of以外の前置詞でW—E、S—Eを含む7つのデータを通して頻出するのはtoである。toがin, at, onなど他の前置詞より頻出するのは全データに共通する。このことからofは書き言葉に多用され、toは書き言葉、話し言葉を問わず多用される前置詞であることがわかる。

定冠詞theは不定冠詞aより頻出する傾向がS—ADULT、S—CHILDを除くデータで見られるが、S—ADULTとS—CHILDではそれほどの差は見られなかった。特にS—CHILDでは僅差で不定冠詞の方がランクは上位になっているが、この理由は我々のデータからは不明であった。

話し言葉にはwell, yeah, oh, um, uhなどの合いの手を意味する言葉、言いよどもの言葉が頻出するが、書き言葉には（当然のことながら）これらは見られないことを記しておく。

前述の通り、W—E、S—Eは十分なサンプル数を持っているので上位20語に一般動詞や普通名詞は存在しないが、我々の5つのデータ中にはそのデータ特有の語が頻出している。例えば一人息子の交通事故による入院の顛末を描いたW—NOVELにはdoctor, boyあるいは会話部分の発言者を明示するための動詞saidが多用されるのに対し、それぞれ既に述べた理由から、W—KUHNではscience, scientific, researchなどが、またS—M&QではEnglish, American, Britishなどの単語がそれである。

S—ADULTやS—CHILDではくだけた日常会話という性格からそれほど特有の語は多用されていないが、それでもschool, swimming, poolなどの語が見られる。しかしこれらの5つのデータそれぞれに特有に生じる単語群は、我々のデータが十分なサンプル数を持たないことを示しているだけのものであるから、これらの語は単に各データの持つトピックに関与する特別な単語と位置づけておくことにしよう。

#### 4 結 語

本稿は5つのそれぞれの性質の異なるデータに対してパーソナル・コンピュータを援用した計量分析を行い、書き言葉と話し言葉の相違点を明らかにしようと試みたものである。

3.1ではContractionに関する計量分析を行い、Contractionがくだけた会話体の重要な

特徴の一つであることが確認された。

3.2では語彙の長さに関して各データの文字数別生起回数一覧表をコンピュータによって作成し、その表を基にいくつかの考察をした。そしてここでも、各データの性質によって出現回数の分布には有為な差が認められた。

3.3では各データごとの単語の出現回数順の一覧表を作成し、各データ間の差異を検討した。

言語構造のどんなレベルにおいても、一見したところでは明らかではないいくつかの統計的規則性が存在することは周知の事実である。これらの規則性は十分な量のデータを近年我々の周辺に急速に普及したパーソナル・コンピュータを利用した分析をすれば比較的容易に解明することができるはずである。本稿はその趣旨にそって行った、語彙のレベルでのコンピュータを利用した計量分析の一つの試みである<sup>(6)</sup>。

## 註

- (1) Contraction に関する制限など詳細は新言語学辞典ならびに同辞典に引用されている文献を参照のこと。
- (2) 「'」の出現回数をパソコンで検索する際には、属格の'sも区別なく検索結果に出力されてしまうことに注意する必要がある。ここでは、付随する作業として属格を示す'sを検索結果から除外した。例 grandmother's house.
- (3) 表(5)は島岡他 (*ibid.*) の表 6-3 をそのまま示したものである。明示されてはいないが同辞典中の7,945語とは見出し語の左肩に\*, \*\*, \*\*\*, ◇のマークが付された基本語の総計であろう。また同表には誤植(文字数1の項の6.0は0.0の誤植:ただし本表では訂正済み)やパーセンテージを計算する際の誤差が存在する。また語数の合計も先に述べた7,945語に一語不足する7,944語であるがこれを説明する事情は不明であるので、そのまま提示しておくことにする。
- (4) 各データ中の最長語である27文字語は development-by-accumulation であるが、これは我々の英単語切り出し法では-(=ハイフン)を単語の区切り文字(デリミタ)としなかったため、いわば方法論上の制約から最長語となったものであることを付記しておく。
- (5) 二人称代名詞 you においては単数、複数、主格、目的格は語彙リストからは区別できないので、これら全てをまとめた数になるが、you の生起回数はW—NOVEL…8, W—KUHN…0, S—M & Q…36, S—ADULT…109, S—CHILD…49という結果が観察される。S—ADULTでは全語数のうち、You/you は4.42%を占めるのに対し、S—CHILDでは1.68%に過ぎず、一人称単数代名詞 I の S—ADULT, S—CHILD 総語数に占める割合がそれぞれ5.76%, 5.80%と大差のないのに比べれば出現率にアンバランスが生じている。これは子供の会話の内容が大人のそれに比べて自分のことを述べるばかりで相手に質問をすることや相手の事情に配慮をするといった例が少ないためではないかと思われるが、会話の内容分析、あるいは疑問文の生起率など意味・統語に関する調査は別の機会にゆずることにする。
- (6) W—KUHN に使用される単語に長いものが多い、S—CHILD では短いものが多いのは5つのデータそれぞれの中に生起する4文字語以下の語の占める割合、10文字以上の語の占める割合(それぞれ4文字語、10文字語を含む)を示した下表からもはっきりとかがえる。

	W—N	W—K	S—M	S—A	S—C
4	69.9	50.8	61.9	70.4	75.6
10	0.8	13.8	3.4	2.6	0.7

(単位%)

- (7) W—Eは新聞の英語に現われる語彙の頻度順一覧表で書き言葉を代表する。S—Eは世界の三大コンピュータコーパスの一つと目される The London-Lund Corpus of Spoken English における頻度順リスト上位20語である。ただし生起回数はW—E, S—Eとも原表にも記されていない。このリストの詳細については Crystal (*ibid.*) を参照のこと。
- (8) Crystal (*ibid.* 86-87) では統計的規則性のいくつかを紹介されている。例えば我々が発する言葉の60%は子音, 40%が母音であること, 日常会話で用いる音節構造の約1/3が cat のような CVC の構造を持つこと, 我々の使用する頻度順上位50語で我々が日常書く言葉の全体の45%を占めること, などである。

### 使用データ

- Carterette, Edward & Margaret Jones. 1974. Informal speech. Alphabetic & phonemic texts with statistical analyses and tables. Berkeley, California : University of California Press.
- Carver, Raymond. 1981. The bath. In Iwamoto & Morita 1988. American scenes today. 東京：朝日出版社
- Kuhn, Thomas. 1962. The structure of scientific revolutions. Chicago : The University of Chicago Press.
- Marckwardt, Albert & Randolph Quirk. 1964. A common language. In Kenkyusha pocket English series. 東京：研究社

### 参考文献

- 荒木一雄他 編 1982. 新英語学辞典 東京：研究社
- 島岡 丘, 伊藤健三 他 1982. 「英語学と英語教育」英語学大系第12巻 東京：大修館
- 高橋 涉 1991. 「日本語における音素配列論上の制約」信州大学教育学部紀要第72号 pp. 221—231.
- . 1992. 「パーソナル・コンピュータを利用した英語語彙分析法」信州大学教育学部教育工学センター紀要第8号 pp. 61—72.
- Crystal, David. 1987. The Cambridge encyclopedia of language. Cambridge : Cambridge U. P.  
(1992年8月27日 受理)