

〔研究ノート〕

## 映画英語字幕語彙の研究(1)

中井 紀明	福田 稔
萩原 弘子	池並 洋子
中井真理子	谷上れい子
友沢 昭江	

桃山学院大学総合研究所の共同研究プロジェクトの一つ（国際化時代の英語教育）として、洋画の台詞として現れた日常英語の収集研究が始まったのは二年前であった。<sup>①</sup> 1000本近い英語字幕入り洋画を集めて日常英語を研究し、英語教育の改善に寄与しようというのが本来の目標であった。もともと映像・音声しかなかった洋画も英語字幕が入ることによって教材の資料として注目を集めるようになってきた。映像・音声・文字などを使ったマルチメディアの英語教材を作るのには、洋画の英語字幕は最も便利で優れた資料である。英語字幕の文字情報は手で入力しなくてもどんどんコンピュータに取り込めるので、大量の日常英語が予算さえあれば短時間の内に極めて容易にデータベース化され、分析・研究しやすい姿でわれわれの前に現われてくる。しかも、これらの文字情報には現実の日常生活と同等同質の自然な音声と映像が付随している。われわれは印刷物のように言語行為を読むだけではなく、言語行為の実際の文脈・背景を現実そこにいるかのように、しかも、何度でも、体験することができる。日常英語のこのように重宝な資料は他には考えられない。

最近英語字幕入り映画が190本近く集まるまでになり、われわれの研究もようやく軌道に乗ってきた。<sup>②</sup> 予算が限られているので、購入できる洋画ビデオは目標には遥か届かないが、コンピューターなどの機器を利用した、台

詞の収集・頻度分析等の一連の作業に、われわれもかなり習熟して一通りの成果を出せるようになってきた。本稿では文章解析ソフト MICRO OXFORD CONCORDANCE PROGRAM を使用して語彙統計表などを出し、学習効果の上がる教材作成を考えてみたい。<sup>(3)</sup>

まず、映画で見られる日常生活の様々な状況の中で、どのような単語が最もよく使われているのかを見てみよう。次に挙げるのは174本の英語字幕全部を通して出現頻度1000以上の単語をアルファベット順に出したものである。174本で総単語数は1403572、語彙(異語)総数は39428、1本平均8066語である。<sup>(4)</sup>

a	(31623)	been	( 2781)	even	( 1024)	help	( 1391)
aah	( 1059)	before	( 1220)	ever	( 1212)	her	( 4618)
about	( 4872)	believe	( 1008)	find	( 1261)	here	( 6634)
after	( 1033)	better	( 1355)	first	( 1194)	hey	( 2251)
again	( 1043)	boy	( 1002)	for	(10592)	hi	( 1080)
all	( 7518)	but	( 5819)	from	( 3074)	him	( 5370)
always	( 1158)	by	( 2326)	get	( 6674)	his	( 3244)
am	( 1511)	call	( 1254)	give	( 1860)	home	( 1680)
an	( 2840)	can	( 4803)	go	( 5708)	how	( 4295)
and	(17948)	can't	( 2920)	god	( 1688)	I	(32069)
any	( 1491)	come	( 5743)	going	( 4003)	I'd	( 1472)
anything	( 1346)	could	( 2497)	good	( 5102)	I'll	( 4598)
are	( 8048)	day	( 1306)	got	( 5328)	I'm	( 6764)
around	( 1108)	did	( 3684)	great	( 1402)	I've	( 2538)
as	( 3275)	didn't	( 2176)	ha	( 3673)	if	( 5011)
at	( 4803)	do	( 9133)	had	( 2424)	in	(15039)
away	( 1321)	does	( 1085)	has	( 1856)	into	( 1215)
back	( 2938)	doing	( 1337)	have	( 9578)	is	(13549)
be	( 9619)	don't	( 9240)	he	( 8167)	it	(19366)
because	( 1195)	down	( 2313)	he's	( 3095)	it's	( 8122)

映画英語字幕語彙の研究(1)

just	( 5097)	nothing	( 1460)	sorry	( 1520)	up	( 5038)
know	( 6723)	now	( 4421)	sure	( 1432)	us	( 2519)
leave	( 1051)	of	(16861)	take	( 2297)	very	( 2218)
let	( 2009)	off	( 1768)	talk	( 1250)	wait	( 1210)
let's	( 1919)	oh	( 5116)	tell	( 2619)	want	( 4632)
life	( 1284)	ok	( 1433)	than	( 1163)	was	( 7508)
like	( 5727)	old	( 1275)	thank	( 2099)	way	( 1877)
little	( 2174)	on	(10183)	that	(13605)	we	( 6601)
long	( 1199)	one	( 3151)	that's	( 3141)	we'll	( 1486)
look	( 2886)	only	( 1742)	the	(32649)	we're	( 2172)
love	( 2397)	or	( 1707)	their	( 1002)	well	( 3796)
make	( 1579)	our	( 2095)	them	( 3150)	were	( 2418)
man	( 2569)	out	( 5324)	then	( 2049)	what	( 8542)
maybe	( 1171)	over	( 1814)	there	( 4749)	what's	( 1494)
me	(15028)	people	( 1547)	there's	( 1906)	when	( 2337)
mean	( 1273)	please	( 1895)	these	( 1408)	where	( 2021)
miss	( 1477)	put	( 1201)	they	( 3475)	who	( 2744)
more	( 1934)	really	( 1557)	they're	( 1270)	why	( 3331)
mr	( 2642)	right	( 5243)	thing	( 1224)	will	( 4729)
much	( 1629)	said	( 1661)	things	( 1114)	with	( 7361)
must	( 1826)	say	( 2626)	think	( 3580)	work	( 1261)
my	(11641)	see	( 2998)	this	(11482)	would	( 2786)
name	( 1025)	she	( 2984)	those	( 1173)	yeah	( 2113)
need	( 1486)	she's	( 1165)	thought	( 1230)	yes	( 3904)
never	( 2683)	should	( 1565)	time	( 2101)	you	(40202)
new	( 1201)	sir	( 1975)	to	(34387)	you'll	( 1432)
night	( 1609)	so	( 3685)	too	( 2193)	you're	( 4085)
no	( 7003)	some	( 2546)	two	( 1452)	your	( 8203)
not	( 7882)	something	( 1975)	uh	( 1523)		

われわれが中学校で習った、最も基本的な語ばかりである。予想以上に容易な語群である。文字を見てこれらの語の意味を理解しない大学生は少ないであろう。ところが、これらの語が半数近く占めるテキストを、画面を見ながら自然なスピードで聞くと、彼らの多くは全く分からないと言う。われわれの語学力も現代の中高生の語学力も基本的に読解力中心であり、聴解力を訓練されて大学に入ってくるものは皆無に等しい。われわれはこの点を是正したいと考えている。出現頻度の極めて高い基本語を中心に、文字を見れば分かる単語を組合せて聴解力を養っていきながら、英語は聞いたら全く分からないという学生の衝撃を少しでもやわらげてやりたいとわれわれは考えている。次に行う出現頻度調査なども聴解力養成のための、「学習者に優しい教科書」作成に利用するためである。

さて、ジャンル別に分けたものの中で最も量の多かったものは「ドラマ2」と分類したものであった。34本の作品から成り1本平均で8600語強、単語数では29万語強、語彙（異語）数で16000語強である。まず、今挙げた全体の超高出現頻度語と比較するために、「ドラマ2」の超高出現頻度の単語を挙げておく。先に挙げた全体の超高出現頻度語と順位などもほとんど変わらない。

you	(11247)	What	(2446)	It's	(1708)	like	(1186)
the	(9095)	this	(2274)	be	(1706)	go	(1178)
I	(8860)	your	(2203)	was	(1705)	Come	(1166)
to	(6824)	my	(2123)	Are	(1689)	out	(1089)
a	(6573)	for	(2080)	We	(1599)	him	(1056)
It	(3932)	on	(2049)	not	(1563)	but	(1047)
and	(3446)	I'm	(1995)	with	(1513)	right	(1043)
of	(3315)	Do	(1948)	all	(1420)	Oh	(1040)
in	(3172)	don't	(1923)	know	(1382)	got	(1036)
me	(3006)	have	(1858)	here	(1296)	they	(1020)
that	(2674)	No	(1821)	get	(1294)	want	(1020)
is	(2657)	He	(1762)	You're	(1200)	up	(1014)

映画英語字幕語彙の研究(1)

about ( 996)	just ( 945)	There ( 894)	Yes ( 849)
good ( 983)	so ( 922)	Did ( 892)	will ( 841)
That's ( 976)	her ( 913)	at ( 877)	going ( 824)
can ( 966)	how ( 913)	I'll ( 867)	
If ( 961)	She ( 913)	one ( 851)	

さて、MICRO OCP で作成した「ドラマ2」の語彙統計表を挙げてみよう。  
NUMBER SUCH の項に入っている単語は、後で挙げる100語テキストの語  
である。(各項目の説明は統計表の終わりにつけた。)

(a) FRE- QUENCY	(b) RELA- TIVE FRE- QUENCY	(c) NUM- BER SUCH	(d) WORDS IN FRE- QUENCY	(e) VOCAB TOTAL	(f) WORD TOTAL	(g) PERC. OF VOCAB	(h) PERC. OF WORDS	(i) PERC. OF WORDS IN FREQ.
1	0.00034	8340	8340	8340	8340	49.98	2.84	2.84
		[experiencing, turbulence, seatbelt, frequent, flier]						
2	0.00068	2359	4718	10699	13058	64.11	4.44	1.61
		[attendants]						
3	0.00102	1201	3603	11900	16661	71.31	5.67	1.23
4	0.00136	772	3088	12672	19749	75.93	6.72	1.05
		[bonus]						
5	0.00170	500	2500	13172	22249	78.93	7.57	0.85
		[landing]						
6	0.00204	410	2460	13582	24709	81.39	8.41	0.84
		[minor, observe]						
7	0.00238	293	2051	13875	26760	83.14	9.11	0.70
8	0.00272	257	2056	14132	28816	84.68	9.81	0.70
9	0.00306	204	1836	14336	30652	85.91	10.43	0.62
		[obviously, twist, cream]						
10	0.00340	192	1920	14528	32572	87.06	11.08	0.65
		[prepare]						
11	0.00374	143	1573	14671	34145	87.91	11.62	0.54
12	0.00408	116	1392	14787	35537	88.61	12.09	0.47
13	0.00442	118	1534	14905	37071	89.32	12.62	0.52

英米評論 No. 5

14	0.00476	85	1190	14990	38261	89.83	13.02	0.40
15	0.00510	91	1365	15081	39626	90.37	13.49	0.46
16	0.00545	77	1232	15158	40858	90.83	13.90	0.42
17	0.00579	69	1173	15227	42031	91.25	14.30	0.40
18	0.00613	71	1278	15298	43309	91.67	14.74	0.43
19	0.00647	56	1064	15354	44373	92.01	15.10	0.36
20	0.00681	48	960	15402	45333	92.29	15.43	0.33
21	0.00715	37	777	15439	46110	92.52	15.69	0.26
22	0.00749	40	880	15479	46990	92.76	15.99	0.30
		[flight]						
23	0.00783	42	966	15521	47956	93.01	16.32	0.33
24	0.00817	32	768	15553	48724	93.20	16.58	0.26
25	0.00851	38	950	15591	49674	93.43	16.90	0.32
26	0.00885	31	806	15622	50480	93.61	17.18	0.27
27	0.00919	31	837	15653	51317	93.80	17.46	0.28
28	0.00953	28	784	15681	52101	93.97	17.73	0.27
29	0.00987	31	899	15712	53000	94.15	18.04	0.31
		[30]						
30	0.01021	30	900	15742	53900	94.33	18.34	0.31
31	0.01055	22	682	15764	54582	94.46	18.58	0.23
		[sign, cup]						
32	0.01089	31	992	15795	55574	94.65	18.91	0.34
33	0.01123	32	1056	15827	56630	94.84	19.27	0.36
34	0.01157	20	680	15847	57310	94.96	19.50	0.23
35	0.01191	18	630	15865	57940	95.07	19.72	0.21
36	0.01225	14	504	15879	58444	95.15	19.89	0.17
37	0.01259	14	518	15893	58962	95.24	20.07	0.18
38	0.01293	15	570	15908	59532	95.33	20.26	0.19
39	0.01327	17	663	15925	60195	95.43	20.49	0.23
		[suppose]						

映画英語字幕語彙の研究(1)

40	0.01361	9	360	15934	60555	95.48	20.61	0.12
41	0.01395	13	533	15947	61088	95.56	20.79	0.18
		[coffee]						
42	0.01429	15	630	15962	61718	95.65	21.00	0.21
43	0.01463	18	774	15980	62492	95.76	21.27	0.26
44	0.01497	13	572	15993	63064	95.84	21.46	0.19
45	0.01531	13	585	16006	63649	95.91	21.66	0.20
46	0.01565	15	690	16021	64339	96.00	21.90	0.23
47	0.01599	5	235	16026	64574	96.03	21.98	0.08
48	0.01634	6	288	16032	64862	96.07	22.07	0.10
49	0.01668	9	441	16041	65303	96.12	22.22	0.15
50	0.01702	9	450	16050	65753	96.18	22.38	0.15
		[20]						
51	0.01736	6	306	16056	66059	96.21	22.48	0.10
52	0.01770	13	676	16069	66735	96.29	22.71	0.23
53	0.01804	9	477	16078	67212	96.34	22.87	0.16
54	0.01838	9	486	16087	67698	96.40	23.04	0.17
55	0.01872	6	330	16093	68028	96.43	23.15	0.11
		[15]						
56	0.01906	9	504	16102	68532	96.49	23.32	0.17
57	0.01940	8	456	16110	68988	96.54	23.48	0.16
58	0.01974	13	754	16123	69742	96.61	23.73	0.26
59	0.02008	7	413	16130	70155	96.66	23.87	0.14
		[minutes]						
60	0.02042	6	360	16136	70515	96.69	24.00	0.12
61	0.02076	10	610	16146	71125	96.75	24.20	0.21
62	0.02110	5	310	16151	71435	96.78	24.31	0.11
63	0.02144	4	252	16155	71687	96.81	24.40	0.09
64	0.02178	7	448	16162	72135	96.85	24.55	0.15

英米評論 No. 5

65	0.02212	8	520	16170	72655	96.90	24.73	0.18
66	0.02246	7	462	16177	73117	96.94	24.88	0.16
67	0.02280	6	402	16183	73519	96.97	25.02	0.14
68	0.02314	8	544	16191	74063	97.02	25.20	0.19
69	0.02348	9	621	16200	74684	97.08	25.42	0.21
70	0.02382	8	560	16208	75244	97.12	25.61	0.19
71	0.02416	5	355	16213	75599	97.15	25.73	0.12
72	0.02450	4	288	16217	75887	97.18	25.83	0.10
73	0.02484	7	511	16224	76398	97.22	26.00	0.17
74	0.02518	6	444	16230	76842	97.26	26.15	0.15
75	0.02552	5	375	16235	77217	97.29	26.28	0.13
76	0.02586	2	152	16237	77369	97.30	26.33	0.05
77	0.02620	9	693	16246	78062	97.35	26.57	0.24
78	0.02654	8	624	16254	78686	97.40	26.78	0.21
79	0.02688	5	395	16259	79081	97.43	26.91	0.13
80	0.02723	4	320	16263	79401	97.45	27.02	0.11
		[gave]						
81	0.02757	5	405	16268	79806	97.48	27.16	0.14
82	0.02791	3	246	16271	80052	97.50	27.24	0.08
83	0.02825	3	249	16274	80301	97.52	27.33	0.08
84	0.02859	5	420	16279	80721	97.55	27.47	0.14
85	0.02893	7	595	16286	81316	97.59	27.67	0.20
86	0.02927	2	172	16288	81488	97.60	27.73	0.06
87	0.02961	4	348	16292	81836	97.63	27.85	0.12
88	0.02995	1	88	16293	81924	97.63	27.88	0.03
89	0.03029	4	356	16297	82280	97.66	28.00	0.12
90	0.03063	5	450	16302	82730	97.69	28.15	0.15
91	0.03097	4	364	16306	83094	97.71	28.28	0.12



映画英語字幕語彙の研究(1)

92	0.03131	6	552	16312	83646	97.75	28.47	0.19
93	0.03165	4	372	16316	84018	97.77	28.59	0.13
94	0.03199	5	470	16321	84488	97.80	28.75	0.16
95	0.03233	2	190	16323	84678	97.81	28.82	0.06
96	0.03267	3	288	16326	84966	97.83	28.92	0.10
97	0.03301	4	388	16330	85354	97.85	29.05	0.13
98	0.03335	3	294	16333	85648	97.87	29.15	0.10
99	0.03369	5	495	16338	86143	97.90	29.32	0.17
100	0.03403	1	100	16339	86243	97.91	29.35	0.03
101	0.03437	2	202	16341	86445	97.92	29.42	0.07
102	0.03471	2	204	16343	86649	97.93	29.49	0.07
103	0.03505	1	103	16344	86752	97.94	29.52	0.04
104	0.03539	7	728	16351	87480	97.98	29.77	0.25
		[minute, 10]						
105	0.03573	1	105	16352	87585	97.99	29.81	0.04
106	0.03607	2	212	16354	87797	98.00	29.88	0.07
107	0.03641	5	535	16359	88332	98.03	30.06	0.18
108	0.03675	1	108	16360	88440	98.03	30.10	0.04
109	0.03709	3	327	16363	88767	98.05	30.21	0.11
110	0.03743	2	220	16365	88987	98.06	30.28	0.07
112	0.03812	4	448	16369	89435	98.09	30.44	0.15
114	0.03880	2	228	16371	89663	98.10	30.51	0.08
115	0.03914	2	230	16373	89893	98.11	30.59	0.08
116	0.03948	3	348	16376	90241	98.13	30.71	0.12
117	0.03982	1	117	16377	90358	98.14	30.75	0.04
118	0.04016	1	118	16378	90476	98.14	30.79	0.04
119	0.04050	1	119	16379	90595	98.15	30.83	0.04
120	0.04084	4	480	16383	91075	98.17	30.99	0.16

英米評論 No. 5

121	0.04118	1	121	16384	91196	98.18	31.04	0.04
122	0.04152	2	244	16386	91440	98.19	31.12	0.08
123	0.04186	1	123	16387	91563	98.20	31.16	0.04
125	0.04254	1	125	16388	91688	98.20	31.20	0.04
126	0.04288	4	504	16392	92192	98.23	31.37	0.17
127	0.04322	3	381	16395	92573	98.24	31.50	0.13
128	0.04356	1	128	16396	92701	98.25	31.55	0.04
129	0.04390	1	129	16397	92830	98.26	31.59	0.04
130	0.04424	2	260	16399	93090	98.27	31.68	0.09
131	0.04458	3	393	16402	93483	98.29	31.81	0.13
132	0.04492	3	396	16405	93879	98.30	31.95	0.13
133	0.04526	2	266	16407	94145	98.32	32.04	0.09
134	0.04560	3	402	16410	94547	98.33	32.18	0.14
135	0.04594	3	405	16413	94952	98.35	32.31	0.14
136	0.04628	2	272	16415	95224	98.36	32.41	0.09
137	0.04662	1	137	16416	95361	98.37	32.45	0.05
138	0.04696	2	276	16418	95637	98.38	32.55	0.09
139	0.04730	1	139	16419	95776	98.39	32.59	0.05
140	0.04764	1	140	16420	95916	98.39	32.64	0.05
141	0.04798	4	564	16424	96480	98.42	32.83	0.19
143	0.04866	1	143	16425	96623	98.42	32.88	0.05
144	0.04901	1	144	16426	96767	98.43	32.93	0.05
145	0.04935	2	290	16428	97057	98.44	33.03	0.10
149	0.05071	2	298	16430	97355	98.45	33.13	0.10
150	0.05105	1	150	16431	97505	98.46	33.18	0.05
151	0.05139	1	151	16432	97656	98.47	33.23	0.05
152	0.05173	1	152	16433	97808	98.47	33.29	0.05
153	0.05207	5	765	16438	98573	98.50	33.55	0.26

映画英語字幕語彙の研究(1)

154	0.05241	2	308	16440	98881	98.51	33.65	0.10
156	0.05309	2	312	16442	99193	98.53	33.76	0.11
157	0.05343	2	314	16444	99507	98.54	33.86	0.11
158	0.05377	1	158	16445	99665	98.54	33.92	0.05
159	0.05411	1	159	16446	99824	98.55	33.97	0.05
160	0.05445	3	480	16449	100304	98.57	34.13	0.16
161	0.05479	1	161	16450	100465	98.57	34.19	0.05
162	0.05513	1	162	16451	100627	98.58	34.24	0.06
164	0.05581	3	492	16454	101119	98.60	34.41	0.17
165	0.05615	1	165	16455	101284	98.60	34.47	0.06
		[done]						
167	0.05683	1	167	16456	101451	98.61	34.53	0.06
170	0.05785	2	340	16458	101791	98.62	34.64	0.12
171	0.05819	2	342	16460	102133	98.63	34.76	0.12
172	0.05853	1	172	16461	102305	98.64	34.82	0.06
173	0.05887	1	173	16462	102478	98.65	34.87	0.06
174	0.05921	2	348	16464	102826	98.66	34.99	0.12
175	0.05956	2	350	16466	103176	98.67	35.11	0.12
176	0.05990	1	176	16467	103352	98.68	35.17	0.06
180	0.06126	2	360	16469	103712	98.69	35.29	0.12
181	0.06160	1	181	16470	103893	98.69	35.36	0.06
186	0.06330	2	372	16472	104265	98.71	35.48	0.13
188	0.06398	1	188	16473	104453	98.71	35.55	0.06
189	0.06432	1	189	16474	104642	98.72	35.61	0.06
192	0.06534	1	192	16475	104834	98.72	35.68	0.07
195	0.06636	2	390	16477	105224	98.74	35.81	0.13
198	0.06738	1	198	16478	105422	98.74	35.88	0.07
199	0.06772	1	199	16479	105621	98.75	35.94	0.07

英米評論 No. 5

203	0.06908	1	203	16480	105824	98.75	36.01	0.07
204	0.06942	1	204	16481	106028	98.76	36.08	0.07
205	0.06976	3	615	16484	106643	98.78	36.29	0.21
207	0.07045	2	414	16486	107057	98.79	36.43	0.14
		[wait]						
208	0.07079	2	416	16488	107473	98.80	36.57	0.14
209	0.07113	1	209	16489	107682	98.81	36.65	0.07
211	0.07181	2	422	16491	108104	98.82	36.79	0.14
214	0.07283	2	428	16493	108532	98.83	36.93	0.15
219	0.07453	2	438	16495	108970	98.84	37.08	0.15
220	0.07487	3	660	16498	109630	98.86	37.31	0.22
222	0.07555	1	222	16499	109852	98.87	37.38	0.08
223	0.07589	1	223	16500	110075	98.87	37.46	0.08
225	0.07657	1	225	16501	110300	98.88	37.54	0.08
226	0.07691	2	452	16503	110752	98.89	37.69	0.15
227	0.07725	1	227	16504	110979	98.90	37.77	0.08
229	0.07793	1	229	16505	111208	98.90	37.85	0.08
230	0.07827	1	230	16506	111438	98.91	37.92	0.08
231	0.07861	1	231	16507	111669	98.92	38.00	0.08
233	0.07929	1	233	16508	111902	98.92	38.08	0.08
236	0.08031	1	236	16509	112138	98.93	38.16	0.08
239	0.08134	2	478	16511	112616	98.94	38.32	0.16
240	0.08168	3	720	16514	113336	98.96	38.57	0.25
243	0.08270	2	486	16516	113822	98.97	38.74	0.17
246	0.08372	1	246	16517	114068	98.98	38.82	0.08
247	0.08406	2	494	16519	114562	98.99	38.99	0.17
249	0.08474	1	249	16520	114811	98.99	39.07	0.08
250	0.08508	2	500	16522	115311	99.01	39.24	0.17

映画英語字幕語彙の研究(1)

251	0.08542	3	753	16525	116064	99.02	39.50	0.26
253	0.08610	1	253	16526	116317	99.03	39.58	0.09
255	0.08678	1	255	16527	116572	99.04	39.67	0.09
257	0.08746	2	514	16529	117086	99.05	39.85	0.17
259	0.08814	1	259	16530	117345	99.05	39.93	0.09
262	0.08916	1	262	16531	117607	99.06	40.02	0.09
		[before]						
263	0.08950	1	263	16532	117870	99.07	40.11	0.09
264	0.08984	2	528	16534	118398	99.08	40.29	0.18
265	0.09018	1	265	16535	118663	99.08	40.38	0.09
271	0.09223	2	542	16537	119205	99.10	40.57	0.18
		[we'll]						
273	0.09291	1	273	16538	119478	99.10	40.66	0.09
274	0.09325	1	274	16539	119752	99.11	40.75	0.09
		[work]						
278	0.09461	1	278	16540	120030	99.11	40.85	0.09
282	0.09597	1	282	16541	120312	99.12	40.94	0.10
283	0.09631	2	566	16543	120878	99.13	41.14	0.19
285	0.09699	1	285	16544	121163	99.14	41.23	0.10
286	0.09733	1	286	16545	121449	99.14	41.33	0.10
		[am]						
288	0.09801	1	288	16546	121737	99.15	41.43	0.10
290	0.09869	1	290	16547	122027	99.16	41.53	0.10
		[thing]						
298	0.10141	1	298	16548	122325	99.16	41.63	0.10
302	0.10277	1	302	16549	122627	99.17	41.73	0.10
303	0.10312	1	303	16550	122930	99.17	41.83	0.10
304	0.10346	1	304	16551	123234	99.18	41.94	0.10
305	0.10380	1	305	16552	123539	99.19	42.04	0.10
306	0.10414	1	306	16553	123845	99.19	42.15	0.10

英米評論 No. 5

307	0.10448	1	307	16554	124152	99.20	42.25	0.10
310	0.10550	1	310	16555	124462	99.20	42.36	0.11
312	0.10618	1	312	16556	124774	99.21	42.46	0.11
317	0.10788	1	317	16557	125091	99.22	42.57	0.11
319	0.10856	1	319	16558	125410	99.22	42.68	0.11
325	0.11060	1	325	16559	125735	99.23	42.79	0.11
		[please]						
332	0.11298	1	332	16560	126067	99.23	42.90	0.11
336	0.11435	1	336	16561	126403	99.24	43.02	0.11
337	0.11469	1	337	16562	126740	99.24	43.13	0.11
340	0.11571	1	340	16563	127080	99.25	43.25	0.12
		[we're]						
342	0.11639	1	342	16564	127422	99.26	43.36	0.12
343	0.11673	1	343	16565	127765	99.26	43.48	0.12
349	0.11877	1	349	16566	128114	99.27	43.60	0.12
353	0.12013	1	353	16567	128467	99.27	43.72	0.12
355	0.12081	1	355	16568	128822	99.28	43.84	0.12
373	0.12694	1	373	16569	129195	99.29	43.97	0.13
375	0.12762	1	375	16570	129570	99.29	44.09	0.13
389	0.13238	2	778	16572	130348	99.30	44.36	0.26
390	0.13272	1	390	16573	130738	99.31	44.49	0.13
391	0.13306	1	391	16574	131129	99.32	44.63	0.13
		[give]						
394	0.13408	1	394	16575	131523	99.32	44.76	0.13
395	0.13442	1	395	16576	131918	99.33	44.89	0.13
		[something]						
397	0.13510	1	397	16577	132315	99.33	45.03	0.14
409	0.13919	1	409	16578	132724	99.34	45.17	0.14
410	0.13953	1	410	16579	133134	99.35	45.31	0.14
		[more]						

映画英語字幕語彙の研究(1)

423	0.14395	1	423	16580	133557	99.35	45.45	0.14
426	0.14497	1	426	16581	133983	99.36	45.60	0.14
432	0.14702	1	432	16582	134415	99.36	45.74	0.15
441	0.15008	1	441	16583	134856	99.37	45.89	0.15
450	0.15314	1	450	16584	135306	99.38	46.05	0.15
	[or]							
457	0.15552	2	914	16586	136220	99.39	46.36	0.31
460	0.15654	1	460	16587	136680	99.39	46.51	0.16
462	0.15723	1	462	16588	137142	99.40	46.67	0.16
484	0.16471	1	484	16589	137626	99.41	46.84	0.16
488	0.16607	1	488	16590	138114	99.41	47.00	0.17
489	0.16641	1	489	16591	138603	99.42	47.17	0.17
	[some]							
492	0.16743	1	492	16592	139095	99.42	47.34	0.17
495	0.16846	1	495	16593	139590	99.43	47.50	0.17
500	0.17016	1	500	16594	140090	99.44	47.67	0.17
504	0.17152	1	504	16595	140594	99.44	47.85	0.17
	[would]							
515	0.17526	1	515	16596	141109	99.45	48.02	0.18
521	0.17730	1	521	16597	141630	99.45	48.20	0.18
544	0.18513	2	1088	16599	142718	99.47	48.57	0.37
545	0.18547	1	545	16600	143263	99.47	48.75	0.19
550	0.18717	1	550	16601	143813	99.48	48.94	0.19
551	0.18751	1	551	16602	144364	99.48	49.13	0.19
552	0.18785	1	552	16603	144916	99.49	49.32	0.19
565	0.19228	1	565	16604	145481	99.50	49.51	0.19
568	0.19330	1	568	16605	146049	99.50	49.70	0.19
573	0.19500	1	573	16606	146622	99.51	49.90	0.20
	[back]							

英米評論 No. 5

574	0.19534	1	574	16607	147196	99.51	50.09	0.20
586	0.19942	1	586	16608	147782	99.52	50.29	0.20
608	0.20691	1	608	16609	148390	99.53	50.50	0.21
612	0.20827	1	612	16610	149002	99.53	50.71	0.21
619	0.21065	1	619	16611	149621	99.54	50.92	0.21
632	0.21508	1	632	16612	150253	99.54	51.13	0.22
634	0.21576	1	634	16613	150887	99.55	51.35	0.22
645	0.21950	1	645	16614	151532	99.56	51.57	0.22
646	0.21984	1	646	16615	152178	99.56	51.79	0.22
656	0.22325	1	656	16616	152834	99.57	52.01	0.22
671	0.22835	1	671	16617	153505	99.57	52.24	0.23
688	0.23414	1	688	16618	154193	99.58	52.47	0.23
737	0.25081	1	737	16619	154930	99.59	52.72	0.25
748	0.25456	1	748	16620	155678	99.59	52.98	0.25
754	0.25660	1	754	16621	156432	99.60	53.24	0.26
778	0.26476	1	778	16622	157210	99.60	53.50	0.26
824	0.28042	1	824	16623	158034	99.61	53.78	0.28
841	0.28620	1	841	16624	158875	99.62	54.07	0.29
849	0.28893	1	849	16625	159724	99.62	54.36	0.29
		[yes]						
851	0.28961	1	851	16626	160575	99.63	54.65	0.29
867	0.29505	1	867	16627	161442	99.63	54.94	0.30
877	0.29846	1	877	16628	162319	99.64	55.24	0.30
892	0.30356	1	892	16629	163211	99.65	55.54	0.30
894	0.30424	1	894	16630	164105	99.65	55.85	0.30
913	0.31071	3	2739	16633	166844	99.67	56.78	0.93
922	0.31377	1	922	16634	167766	99.68	57.09	0.31
945	0.32160	1	945	16635	168711	99.68	57.41	0.32
		[Just]						



映画英語字幕語彙の研究(1)

961	0.32704	1	961	16636	169672	99.69	57.74	0.33
966	0.32874	1	966	16637	170638	99.69	58.07	0.33
976	0.33215	1	976	16638	171614	99.70	58.40	0.33
983	0.33453	1	983	16639	172597	99.71	58.74	0.33
996	0.33895	1	996	16640	173593	99.71	59.08	0.34
GT 1000	0.34065	48	120253	16688	293846	100.00	100.00	40.92

[have, you, this, no, I, to, go, don't, a, me, what, is, this, here, was, that, the, get, with, for, be, in, I'm, and, they]

(j) TYPE/TOKEN RATIO : 0.05679

(k) TOTAL WORDS READ = 293841

TOTAL WORDS SELECTED = 293841

TOTAL WORDS PICKED = 293841

TOTAL WORDS SAMPLED = 293841

TOTAL WORDS KEPT = 293846

(l) TOTAL VOCABULARY = 16688

(a) FREQUENCY [出現頻度] 総単語数293841のテキスト中で1回しか出てこない単語が(c) 8340個もある。1000回以上も出てくる超高「出現頻度」語が48個ある。(語彙統計表の前に掲げた超高出現頻度語のリストを参照。)

(b) RELATIVE FREQUENCY [相対頻度] (a) の単語が総語数の中で占める割合。特定の単語にわれわれがどの程度出会うかを示す数値。 $(\text{出現頻度数} \div \text{総単語数}) \times 100$ 。293841語に1回しか出てこない単語にわれわれが出会う遭遇率は、100語のテキストでは0.00034になる。11247回出てくる you は100語テキストで3.8個出ていることになる。(a) が9095の the は3.1個、(a) が8860の I は3個出てくる。

(c) NUMBER SUCH [単語の個数] (a) の単語の数。

(d) WORDS IN FREQUENCY [出現頻度と該当単語個数の積] (a) × (c)。

1回しか出てこない単語に我々は8340回出あう。5回しか出てこない単語

語には2500回である。

- (e) VOCAB TOTAL [累積語彙数] 語彙 (異語) 総数。 (c) の累計。  
 (a) が5までの異語の合計は  $8340 + 2359 + 1201 + 772 + 500 = 13172$  である。  
 (f) WORD TOTAL [累積単語数] 単語総数。 (d) の累計。 (a) が5  
 までの単語の合計は  $8340 + 4718 + 3603 + 3088 + 2500 = 22249$  である。  
 (g) PERC. OF VOCAB [語彙の累積百分率] ( (e) の百分率表示) (a)  
 が1の単語の (g) は49.98である。つまり語彙 (異語) の半分は29万語  
 に1回しか出てこない。語彙の81.4%が1～6回しか出てこない。また、  
 語彙の9割強が15回以下の低出現頻度である。しかし、この数値はテキ  
 ストが29万語にも上る膨大なものだからである。現実のテキストではこ  
 の数値ははるかに低くなる。(後の100語テキストの分析参照。)  
 (h) PERC. OF WORDS [単語の累積百分率] ( (f) の百分率表示) 語彙  
 の点では81.4%を占める出現頻度1～6の単語が、語数の点では8.4%に  
 過ぎない。  
 (i) PERC. OF WORDS IN FREQ. [出現頻度ごとの単語の百分率] ((d)  
 の百分率表示) ( (d) ÷ 総語数) × 100。われわれが個々の単語に出くわ  
 す遭遇率を (b) が表すのに対して、出現頻度ごとの遭遇率をこの数値  
 は示している。(1) の数値2.84は  $(8340 ÷ 293841) × 100$  から求められ  
 た。29万語に1回しか出てこない単語が、8340もあるということは、100  
 語テキスト中では、大体2.8個は1回しか出てこない単語だということだ  
 である。(a) が1000以上の単語のここの数値40.92は、総単語数の4割以  
 上は1000回以上の超高出現頻度語であることを意味する。  
 (j) TYPE/TOKEN RATIO [語彙総数/総単語数] テキストの語彙の  
 豊富さを示す数値だが、総単語数を同じにして比較しなければ意味がな  
 い。  
 (k) TOTAL WORDS READ [読み込まれた総単語数]  
 (l) TOTAL VOCABULARY [語彙総数]

さてわれわれがこのプロジェクトに着手する際に考えていたことは、出現頻度を調べて、数値の高いものから順に学習者に提示して行けば、学習効果が上がるのではないかということである。しかし、30万語近い量から選び出した出現頻度1000以上の48語は、ことさら工夫して大学生に提示しなければならないような難しい単語ではない。このような超高出現頻度の単語が、大体どのテキストでも総単語の4割以上を占めている。

例として「ドラマ2」の中から、無作為に選んだ約100語のテキストを読んでみよう。*Nothing in Common*の冒頭の部分である。( )内は直前の語の「ドラマ2」全体での出現頻度数である。(ト書き部は省略している。)

Have (1858) you (11247) done (165) this (2274) before (262) ?

No (1821) .

Obviously (9) you have.

I (8860) have to (6824) go (1178) back (573) to work (274) .

No, you don't (1923) .

Yes (849), I have to go.

Wait (207) a (6573) minute (104) .

Just (945) give (391) me (3006) 10 (104) minutes (59) more (410),  
or (450) 15 (55) or 20 (50) minutes more.

What (2446) is (2657) this (2274) thing (290) here (1296) ?

Am (286) I supposed (39) to twist (9) this or something (395) ?

(noise) .

Was (1705) that (2674) you?

(pilot): We're (340) experiencing (1) some (489) minor (6) turbulence (1) .

Please (325) observe (6) the (9095) seatbelt (1) sign (27) .

Would (504) you get (1294) me a cup (27) of (3315) coffee (41)  
with (1513) cream (9) .

(pilot): flight (22) attendants (2), prepare (10) for (2080) landing

(5)

I have to get back to work.

We'll (271) be (1706) landing in (3172) 30 (29) minutes.

I'm (1995) a frequent (1) flier (1), and (3446) they (1020) gave (80) me a bonus (4)

出現頻度別語彙の分布状況を NUMBER SUCH の該当箇所を参照しながら纏めてみると、29万語テキストと100語テキストの違いがはっきり出てくる。

出現頻度	単語数	総語数に占める割合	語彙	語彙全体に占める割合
1000以上	52	50.5% (40.9%)	25	36.2% ( 0.29%)
999~20	36	35% (43.9%)	30	43.5% ( 7.7 %)
19~ 1	15	14.6% (15.1%)	14	20.3% (92.01%)

( ) 内は「ドラマ2」全体の数値。

このテキストは「ドラマ2」全体の平均よりは超高出現頻度語が多い、つまり単語は概して易しいと言えるであろう。しかし、出現頻度1、つまり29万語に1度しか出てこない単語が5回も出てくる。「ドラマ2」全体では平均して100につき2.84だから、比較的多いと言えるだろう。語彙に関して言えば、100語テキストでは低・超低出現頻度語が20.3%なのに、29万語テキストでは、92.01%にもなっている。これは不思議なことではない。テキスト語数が大きくなるに連れて、ここの数値も大きくなっていく。つまりテキストの量が大きくなっていくということは、高・超高出現頻度語のここでの数値が低くなり、低・超低出現頻度語の数値が上がっていくということである。簡単に言えば、テキストの語数が大きくなるに連れ、大きな語彙が必要になるという当然のことを意味している。全体的に見ると低・超低出現頻度語が9割を越える数値になるが、現実にはわれわれが対峙しているテキストの部分部分では、低・超低出現頻度語は約2割にすぎない。約8割は高・超高出現頻度語で文字としては十分大学生にも分かる単語である。

われわれがテキスト編集で特に注意しなければならないのは、極めて容易なテキスト、つまり読めば十分に理解できるテキストでも、耳と目からだけでは大学生は殆ど理解できないということである。彼らの語学力は読解力に偏重している。今回の調査で分かったように、洋画では語彙の点では想像以上に易しい単語が多い。それでも聴いて分からないのは、我々の語彙力が意味中心で、決して音中心ではないということである。学習者の中にある、意味中心の文字の知識を利用しないでは、いくら聴かせても聴解力はつきにくいのではないだろうか。日本の大学生には文字をてこに音を教えたほうがよいと考えられる。

われわれには英語字幕データベースから教科書を編集する際の原則がいくつか見えてきた。テキストに採用すべき箇所を選ぶ時には(1)高・超高出現頻度語に関しては、最も使われている用法・意味のものから順に提示していった方が学習効率が良い(2)低・超低出現頻度語については、文字を見ても学習者が理解できないような単語・熟語は少なくとも初期の段階では避ける。

この100語テキストでも have の違った使い方が見られる。1858の実例をコンコーダンスで調べて、断然多い用法を先に提示した方がよい。日本の大学生は、聴解力よりは読解力の方が格段優れているので、文字では分かるような超低出現頻度の語も混ざっていてもよいだろう。この100語テキストでは出現頻度1の語も、turbulenceを除けば、読めば分かるであろう。しかし1作品全体のシナリオを教科書にした市販の教科書は聴解力を訓練する初歩の教科書としては難しすぎる。最初から自然なスピードに慣れさせることができるという洋画教材のメリットを生かすために、語彙の点では注意深くテキストを選ぶ必要がある。ただでさえ聴解力のない学習者を、読んでも分からないような難解な単語・熟語で最初から鞭打たなくてもよいだろう。コースの最初の方に提示するのは、高出現頻度語が圧倒的に多い部分を集めたものがよいだろう。

[注]

- (1) 中井紀明「洋画「英語字幕」コンピュータ」桃山学院大学『総合研究所報』16巻3号(1991年3月)参照。
- (2) 英語字幕入りビデオから英語字幕を収集するのに次の個人・大学から協力を得た。ここで深い感謝の気持ちをわれわれは表明しておきたい。福田稔、萩原弘子、加藤映子、桃山学院大学、中井紀明、ノートルダム清心女子大学、大阪女学院短期大学。
- (3) 最初の予定では収集したデータ全てを一括して分析する予定であった。しかしわれわれの集めたデータはパソコンのレベルを越えているらしくて、MICRO OCPが七万行を越えた辺りで(40本から50本の間であろうか)動かなくなることが分かった。総語数百五十万語を越えるデータのコンコードダンスなどはとてもパソコンのレベルでは処理できないことも分かってきた。とりあえず全体を「喜劇」とか「ドラマ2」とかに分けて分析した。困っていたところ、大阪大学言語文化部の三木邦弘氏が助け船を出して下さり、単語数・出現頻度数などを出すプログラムを書いて下さった。われわれは氏に深く感謝している。最初に掲げた出現頻度1000以上の単語は同氏のプログラムで作成したものである。

京都大学大型計算機センターと桃山学院大学計算機センターに大型計算機用のOXFORD CONCORDANCE PROGRAMが近く入る予定なので、本格的なコンピュータ分析は近い内に大型計算機で瞬時の内に行えるであろう。1000本を目標にしているビデオの収集も急がなければならない。

- (4) 語彙総数39428という数字にしても、厳密に計算すると激減するであろう。コンピュータは文法解析までしないので、単純に単語の形だけから語数を数えている。sのついた複数形、短縮形、動詞の過去形など全て異語に数えている。

中井紀明	桃山学院大学	福田稔	帝塚山学院大学
萩原弘子	大阪府科学教育センター	池並洋子	桃山学院大学
中井真理子	山陽学園短期大学	谷上れい子	桃山学院大学
友沢昭江	桃山学院大学		