

[共同研究：国際化時代の英語教育]

洋画「英語字幕」とコンピューター

中 井 紀 明*

『桃山学院大学人間科学』創刊号で私は大学英語教育に関する一つの提案をした¹⁾。それは現在二千本近くの洋画に付けられた英語字幕（キャプション）を大学英語教育の教材として積極的に使っているというものであった。その後、この提案を元に「国際化時代の英語教育」²⁾ という研究課題で八人の研究者とともに本学総合研究所の共同研究プロジェクトに申請したところ、幸いにも受け入れられ少なからぬ予算が付いた。この共同研究で「洋画英語字幕のコンピューターへの入力」という役割を与えられた私はこの夏百本近い洋画の英語字幕をコンピューターに取り込む作業をした。使用したハードは Future Technology Lab. 社の Futek Video Caption Adaptor FA710 と NEC の PC-9801RX, ソフトはメディアミックス社の Video Caption 編集システムである。一本一万語とすると百万語を多分越えるであろう膨大な英語字幕情報をこのハードとソフトの組み合わせで極めて容易にコンピューターに入力する事が出来た。大変な驚きであり望外の喜びであった。最終的には千本近い洋画英語字幕を我々は入力する予定であるが、現在でも既に約百万語の英語字幕データベースを構築したことになる。本稿では、共同研究でのコンピューター入力という役割を果たしながら観察した若干の事柄について考察し、合わせて英語字幕教材が映画の台詞情報として信頼できるものかどうか『素晴らしきかな人生』を例にとってコンピューターで検討してみたい³⁾

まず、最初に Video Caption Adaptor FA710 と NEC の PC-9801RX をテレビ、ビデオと接続し Video Caption 編集システムの指示に従って所定の手順を踏むと、コンピューターに英語字幕が全て大文字で次のように入力されてくる。『素晴らしきかな人生』の冒頭部分である。左側の数字は時間の経過を示している。

00:00:00 I OWE EVERYTHING TO GEORGE BAILEY.

*本学文学部

1) 拙稿「洋画使用の大学英語教育——一つの提案」『桃山学院大学人間科学』創刊号（1990年3月）

2) 総合研究所に提出した私の研究課題は次のようなものである。

我が国の英語教育は<読解力>養成を主たる目的にしてきた。それは西洋文化の受け入れという明治維新以来の国策に合致した語学教育で、文献解読を通して西洋を受け入れる語学教育であった。しかし到来した国際化時代に呼応した英語教育の必要性が叫ばれている。<読める>だけを目指す語学教育を脱しなければならない。読んで分かるだけでなく、聞いて分からなければならない。さらにこちらの意見も堂々と表現（言う・書く）出来なければならない。

読む・書く・聴く・話すの四点を総合的に養成する語学教育が、視聴覚機器・ソフトの目ざましい発達で夢の段階から実現可能の段階に入った。今後は各大学でも大学改革の流れの中で、総合的な語学力養成の体系的プログラム開発競争が行われるものと予想される。欧語 OCR, CLOSED CAPTION ADAPTOR などの電子機器を駆使して音声・画像・活字情報をコンピューターなどに収集し、OXFORD MICRO OCP（文章解析プログラム）で頻度分析を行い、顕著な学習効果をもたらす英語学習プログラムを、異分野の方々の参加も得て、開発したいと我々は考えている。

3) 本稿ではフランク・キャブラ（Frank Capra）監督の代表作『素晴らしきかな人生』（*It's a Wonderful Life*, 1946）を中心に取り上げた。内容・表現の豊かさの他に、この作品の英語教科書版が開文社から出ているからである。英語字幕を打ち出した台本が教養英語の教材としてふさわしいものか詳しく検討する際に、英語を母国語としない者にとって確認しにくい箇所をこの英語教科書のおかげで比較的容易に確認出来たことはありがたいことであった。編注者の曾根田憲三、渡辺幸俊、Marion G. Brasher 三氏に御礼を申し上げる。

00 : 00 : 02 HELP HIM, DEAR FATHER.
 00 : 00 : 08 JOSEPH, JESUS, AND MARY, HELP MY FRIEND MR. BAILEY.
 00 : 00 : 12 HELP MY SON GEORGE TONIGHT.
 00 : 00 : 15 HE NEVER THINKS ABOUT HIMSELF. THAT'S WHY HE'S IN TROUBLE.
 00 : 00 : 19 GEORGE IS A GOOD GUY. GIVE HIM A BREAK, GOD.
 00 : 00 : 23 I LOVE HIM, DEAR LORD. WATCH OVER HIM TONIGHT.
 00 : 00 : 27 PLEASE, GOD. SOMETHING'S THE MATTER WITH DADDY.
 00 : 00 : 32 PLEASE BRING DADDY BACK. PLEASE BRING DADDY HOME.
 00 : 00 : 34 WATCH OVER HIM TONIGHT.
 00 : 00 : 37 PLEASE BRING DADDY HOME.
 00 : 00 : 42 HELLO, JOSEPH.
 00 : 00 : 44 TROUBLE?
 00 : 00 : 47 LOOKS LIKE WE'LL HAVE TO SEND SOMEONE DOWN.
 00 : 00 : 50 A LOT OF PEOPLE ASKING FOR HELP
 00 : 00 : 52 FOR A MAN NAMED GEORGE BAILEY.
 00 : 00 : 54 YES. TONIGHT'S HIS CRUCIAL NIGHT.
 00 : 00 : 57 WE'LL SEND SOMEONE NOW. WHOSE TURN IS IT?
 00 : 00 : 59 THAT'S WHY I CAME, SIR.
 00 : 01 : 01 IT'S THAT CLOCKMAKER'S TURN AGAIN.
 00 : 01 : 04 OH. CLARENCE. HASN'T GOT HIS WINGS YET.
 00 : 01 : 06 WE'VE PASSED HIM UPRIGHT ALONG
 00 : 01 : 09 BECAUSE, YOU KNOW, SIR,
 00 : 01 : 12 HE'S GOT THE IQ OF A RABBIT.
 00 : 01 : 15 YES! BUT HE'S GOT THE FAITH OF A CHILD--SIMPLE.
 00 : 01 : 17 JOSEPH, SEND FOR CLARENCE.

次に Video Caption 編集システムの指示に従って、文頭と一人称代名詞“I”を除いた全ての大文字を小文字に変換する。

I owe everything to george bailey. Help him, dear father.
 Joseph, jesus, and mary, help my friend mr. Bailey.
 Help my son george tonight.
 He never thinks about himself. That's why he's in trouble.
 George is a good guy. Give him a break, god.
 I love him, dear Lord. Watch over him tonight.
 Please, god. Something's the matter with daddy.
 Please bring daddy back. Please bring daddy home.
 Watch over him tonight.
 Please bring daddy home.
 Hello, joseph.
 Trouble?
 Looks like we'll have to send someone down.
 A lot of people asking for help for a man named george bailey.

Yes. Tonight's his crucial night.
 We'll send someone now. Whose turn is it?
 That's why I came, sir.
 It's that clockmaker's turn again.
 Oh. Clarence. Hasn't got his wings yet.
 We've passed him upright along
 because, you know, sir,
 he's got the iq of a rabbit.
 Yes, but he's got the faith of a child--simple.
 Joseph, send for clarence.

この段階では固有名詞も全て小文字になっている。作中の固有名詞を全て探し出して、日本語ワードプロセッサ「一太郎」の「置換」機能で適切な処理をする。また下から5行目に見られるような up と right が一語になって出てきているのは、綴りの過ちであるから、市販のスペルチェッカーのようなソフトを使うと発見・訂正が簡単に出来るであろう。これらの過ちを工夫して効率よく全て直すと、次のような完全な英語字幕教材が出来上がる。

I owe everything to George Bailey. Help him, dear Father.
 Joseph, Jesus, and Mary, help my friend Mr. Bailey.
 Help my son George tonight.
 He never thinks about himself. That's why he's in trouble.
 George is a good guy. Give him a break, God.
 I love him, dear Lord. Watch over him tonight.
 Please, God. Something's the matter with Daddy.
 Please bring Daddy back. Please bring Daddy home.
 Watch over him tonight.
 Please bring Daddy home.
 Hello, Joseph.
 Trouble?
 Looks like we'll have to send someone down. A lot of people asking for help for a man named George Bailey.
 Yes. Tonight's his crucial night. We'll send someone now. Whose turn is it?
 That's why I came, sir. It's that clockmaker's turn again.
 Oh. Clarence. Hasn't got his wings yet.
 We've passed him up right along because, you know, sir, he's got the IQ of a rabbit.
 Yes, but he's got the faith of a child—simple.
 Joseph, send for Clarence.

筑波大学の久保田章氏らの研究⁴⁾によると、英語字幕は実際の台詞音声情報の80パーセント前後を復元（文字化）していると言う。今打ち出した英語字幕教材を開文社版の英語教科書の冒頭部と比較してみよう。

I owe everything to George Bailey... Help him, dear Father.

4) 久保田章, 大武博, 小張啓之「テキスト分析から見たビデオキャプションの教材としての可能性」LLA 第30回全国研究大会（1990年7月31日神戸国際会議場）

...Joseph, Jesus and Mary. Help my friend, Mr. Bailey.

Help my son, George, tonight.

He never thinks about himself, God, that's why he's in trouble.

George is a good guy. Give him a break, God...

I love him, dear Lord. Watch over him tonight...

Please, God, something's the matter with Daddy...

Please bring Daddy back. (Please bring Daddy home.)

(Watch over him tonight.)

(Please bring Daddy home.)

Hello, Joseph. Trouble?

Looks like we'll have to send someone down...a lot of people asking for help for a man named George Bailey.

George Bailey. Yes. Tonight's his crucial night. You're right. We'll have to send someone down immediately. Whose turn is it?

That's why I came to see you, sir. It's that clockmaker's turn again.

Oh. Clarence. Hasn't got his wings yet, has he? We've passed him up right along.

Because, you know, sir, he's got the I. Q. of a rabbit.

Yes, but he's got the faith of a child. Simple. Joseph, send for Clarence.

下線部は変化していたり追加されている部分であり、()部は英語字幕版にありながら、開文社版には無い部分である。省略版の英語字幕版にはない台詞が開文社版にかなり見られることは当然予想されるが、省略版の英語字幕版にも開文社版で補足出来なかった台詞があることに注意したい。しかし、確かに英語字幕版のように登場人物は発言しているが、ここにある三箇所の()部分は極めて聴き取りにくく、教材としては教室では利用出来ないから省略されてあっても構わないであろう。

他に幾つか例を挙げてみよう。左が開文社版の台詞、右が英語字幕版で、開文社版からの引用の最後の()内は引用頁である。

Chain Gang! (7)

Make a chain, gang!

実際には開文社版のように主人公ジョージは言っている。“chain gang”という決まった表現をジョージはとっさに振って使っている。

Now you're talking. Gee whiz, I could use that as a raft in case the boat sunk.

I could use that as a raft in case the boat sunk. How much?

How much does this cost? (26)

Oh, this afternoon. I thought I'd give the kids a treat. (41)

省略されて無し。

英語字幕版には開文社版(26頁と41頁)の下線部のような頻出する慣用表現が省略されてしまっている。日本人学生には分かりにくいこの種のごく普通の表現がコンピューターでの頻度分析によって我々が捜し当てたい表現だから、英語字幕版にもデータとしてきちんと入力されていることが望ましい。

Putting a pool under this floor was a great idea. Saved us another building.

Putting a pool under this floor was a

Now, Harry, Sam, have a lot of fun.
There's lots of stuff here to eat and drink.
Lots of pretty girls around.
Thanks, Mr. Partridge.

Hey, George...

Hello, Violet.

Hello. What am I bid? (43)

英語字幕版の下線部はほとんど聴き取れないから教材として無視して良い。開文社版下線部の動詞“saved”の用法は重要だから、英語字幕版にないのは残念である。

Oh, oh!

Oh, oh!

Oh, no... I wanna...

Sir, my train, please!

A pox upon me for a clumsy lout.

Your...your caboose, my lady.

You may kiss my hand. (pp. 50-51)

英語字幕版下線部は、将来妻になるメアリーのガウンの紐を主人公が踏んでしまった時の台詞で、かろうじて聴き取れる所である。この言い方は教材に残っていた方が良いが、音声部分は教材としては難しすぎる。開文社版から削除されていることは理解できる。

Well, not just one wish, a whole hatful.

Mary, I know what I'm gonna do tomorrow and the next day and the next year and the year after that. I'm shaking the dust of this crummy town off my feet, and I'm gonna see the world. Italy... Greece...the Parthenon...the Colosseum.

Then I'm coming back here and go to college and see what they know... Then I'm going to build things. I'm gonna build air fields. I'm gonna build skyscrapers a hundred stories high. I'm gonna build bridges a mile long. (52)

ここでは開文社版が音声情報を完全に文字化している。開文社版の下線部のような具体的情報が英語字幕版から省略されているのは、多分画面上の制約のためである。経済性の原理が英語字幕版のここで働いているのである。

I want the Board to know that George gave up his trip to Europe to help straighten things out here these past few months. Good luck to you at school,

great idea. Now, Harry, Sam, have a lot of fun. Lot of pretty girls around.

Boys, I've got 1/3 of my dance card left.

Hello. What am I bid.

Uh-oh!

Oh, I'm on the thing here.

Sir, my train, please.

A pox upon me for a clumsy lout.

Your caboose, milady.

You may kiss my hand.

Not just one wish, a whole hatful.

I know what I'm going to do tomorrow and the next day. I'm shaking the dust of this crummy town off my feet, and I'm going [to see the world. Then I'll go to college and see what they know. Then I'll build things. I'll build airfields and skyscrapers, I'll build bridges a mile long.

George gave up his European trip to help straighten things out these past few months.

Good luck to you at school.

George. (58)

ここでは半角で18の句が17の句に変化している。半角でたった一つのスペースの差であるから音声情報のままで復元すれば良いのだが、英語字幕版は別の言い方を採用することによって経済性を追求したつもりなのである。

We made it! Close the door, Ernie. We made it! Look. Look, we're still in business. We've still got two bucks left. (105)

ここでは Ernie ではなく Eustace の方が明らかに正しい。英語字幕版から “make it” や “be in business” の慣用表現が省略されているのは誠に惜しい。

Congressman Black is here, to see you, sir. (118)

英語を母国語とする人たちにもこの下線部ははっきりしないようだ。何度も確認したが、私にもどちらが正しいか判断出来ない。これほど判断の難しいものについては学生向けの教材ではどちらでも良いのではないか。

Now, who do you think you are?
Wait a minute. Hello? Hello! Who is this? Oh, Mr. Welch? Okay, that's fine, Mr. Welch. Gives me a chance to tell you what I really think of your wife. George...

Will you get out and let me handle this? Hello? Hello? What? Oh, you will, huh? Ok, Mr. Welch. Any time you think you're man enough. Hello...any...Oh... (171)

この所では最初の下線部は英語字幕版の方が正確に音声情報を文字化しているように思える。しかし二行目の下線部はどちらも正しくないように思える。私には “What does he say?” と主人公は言っているように思えるのだが、いずれにしてもこの音声情報は聴き取り難く教科書版では省略してあっても良いだろう。

That does it! Out you two pixies go, through the door or out the window. Look, Nick, what's wrong?
And that's another thing. Where do you come off calling me Nick?
Well, Nick, that's your name.
What's that got to do with it? I don't know you from Adam's off ox. (206)

開文社版は音声情報を完全に文字化している。下線部最初の二箇所は良く聴く言い方で、英語字幕版でも残して置いて欲しいものである。最後の下線部も、慣用表現の頻度調査のためにデータとし

Close the door, Eustace. We've still got two bucks left.

Congressman Blatz is here, sir.

Hey, you! I'll knock your block off!
What is that? Hello! Who's this?

Oh, Mr. Welch. That's fine.

I'll tell you what I really think of your wife!

George—

Will you let me handle this? Hello! Hello! What? Oh, you will, huh? Ok, Mr. Welch. Any time you think you're man enough, you —Hello. Any—

Out you two pixies go, through the door or out the window.

Nick, what's wrong?

Where do you come off calling me Nick?

Well, Nick, that's your name.

What's that got to do with it?

て英語字幕版に残して欲しかった。

Hey... hey. Where did the Building and Loan move to?
The building and what?
 The Bailey Building and Loan. It was up there.
 They went out of business years ago.
 Hey, Violet! Hey, listen, that's Violet Bick?
I know. I know. Beat it.
 I know that girl!
 Ernie, take me home. I've gone off my nut! Ernie! (213)

Hey, hey, where did the Building and Loan move to?
 What?
 The Bailey Building and Loan.
 That folded years ago.
Liar!
I know every big shot in this town! And I'll have you kicked off the police—
 Hey, Violet! Hey, that's Violet Bick! I know that girl!
Take a walk. Beat it.
 Hey, Ernie! Ernie!
 Ernie, take me home. I'm going off my nut.

英語字幕版の“folded”は経済性のために短く言い替えたのであろうが、実際は警察官が開文社版のように言っている。頻度分析のデータとして考えた時には、“go out of business”と“fold”では明らかに違っている。経済性だけをねらったの言い替え表現が多いデータは頻度分析の対象としては望ましくない。この点が英語字幕版の欠点であることは明らかである。

英語字幕版の“Liar”“Take a walk”までは開文社版から省略されている。聞き取り難いところだったからやむを得なかっただろう。バイオレット・ビックとおぼしい若い女が叫んでいる所で、日本人には聞き取り難いが確かにこう言っている。

開文社版の下から三行目に当たる所は“I know. I know. Take a walk. Beat it.”と警察官が言っている。つまり二つの版を合わせれば完全な文字化になる。最後の行は開文社版の方が正しいと思う。

次に文章解析ソフト OXFORD MICRO OCP を使って完全台詞情報（開文社版を基に英語字幕と私の判断を生かしたもの）と英語字幕情報を比較分析することにする。まず二つの版の語彙統計表を比較してみよう。ここで言う語彙とは、重複して現れない、あるいは他と識別できる単語の総体である。〔 〕内の数値は英語字幕版のものである。表の後に続いているのは(a)から(i)までの各項目の簡単な説明である。

It's a Wonderful life (完全台本) 語彙統計表

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
FREQUENCY	RELATIVE	NUMBER	WORDS IN	VOCAB	WORD	PERC.	PERC.	PERC. OF
[出現頻度]	FREQUENCY	SUCH	FREQUENCY	TOTAL	TOTAL	OF	OF	WORDS IN
	[相対頻度]	[単語の個数]	[出現頻度と該当単語個数の積]	[累積語彙数]	[累積単語数]	[語彙の累積百分率]	[単語の累積百分率]	[出現頻度ごとの単語の百分率]
1	0.00591	1071	1071	1071	1071	50.02	6.33	6.33
[1	0.00759	995	995	995	995	51.10	7.56	7.56]
2	0.01181	317	634	1388	1705	64.83	10.07	3.74
[2	0.01519	303	606	1298	1601	66.67	12.16	4.60]
3	0.01772	179	537	1567	2242	73.19	13.24	3.17

[3	0.02278	147	441	1445	2042	74.22	15.50	3.35]
4	0.02362	88	352	1655	2594	77.30	15.32	2.08
[4	0.03037	87	348	1532	2390	78.69	18.15	2.64]
5	0.02953	70	350	1725	2944	80.57	17.39	2.07
[5	0.03797	60	300	1592	2690	81.77	20.43	2.28
6	0.03544	38	228	1763	3172	82.34	18.73	1.35
7	0.04134	37	259	1800	3431	84.07	20.26	1.53
8	0.04725	35	280	1835	3711	85.71	21.92	1.65
9	0.05315	22	198	1857	3909	86.74	23.09	1.17
10	0.05906	26	260	1883	4169	87.95	24.62	1.54
中略								
100	0.59060	1	100	2111	10802	98.60	63.80	0.59
102	0.60241	1	102	2112	10904	98.65	64.40	0.60
103	0.60832	1	103	2113	11007	98.69	65.01	0.61
105	0.62013	1	105	2114	11112	98.74	65.63	0.62
108	0.63785	3	324	2117	11436	98.88	67.54	1.91
110	0.64966	1	110	2118	11546	98.93	68.19	0.65
112	0.66147	1	112	2119	11658	98.97	68.85	0.66
116	0.68509	1	116	2120	11774	99.02	69.54	0.69
120	0.70872	1	120	2121	11894	99.07	70.25	0.71
133	0.78549	1	133	2122	12027	99.11	71.03	0.79
134	0.79140	1	134	2123	12161	99.16	71.82	0.79
138	0.81502	1	138	2124	12299	99.21	72.64	0.82
139	0.82093	1	139	2125	12438	99.25	73.46	0.82
141	0.83274	1	141	2126	12579	99.30	74.29	0.83
145	0.85637	1	145	2127	12724	99.35	75.15	0.86
147	0.86818	1	147	2128	12871	99.39	76.02	0.87
167	0.98630	1	167	2129	13038	99.44	77.00	0.99
168	0.99220	1	168	2130	13206	99.49	77.99	0.99
189	1.11623	1	189	2131	13395	99.53	79.11	1.12
197	1.16348	1	197	2132	13592	69.58	80.27	1.16
199	1.17529	1	199	2133	13791	99.63	81.45	1.18
201	1.18710	1	201	2134	13992	99.67	82.64	1.19
209	1.23435	1	209	2135	14201	99.72	83.87	1.23
286	1.68911	1	286	2136	14487	99.77	85.56	1.69
373	2.20293	1	373	2137	14860	99.81	87.76	2.20
389	2.29743	1	389	2138	15249	99.86	90.06	2.30
411	2.42736	1	411	2139	15660	99.91	92.49	2.43
479	2.82896	1	479	2140	16139	99.95	95.32	2.83
793	4.68344	1	793	2141	16932	100.00	100.00	4.68

(j) TYPE / TOKEN RATIO	開文社版	英語字幕版
	: 0.12645 (総語彙数/総単語数)	0.14784
TOTAL WORDS READ	=16932 (読み込まれた総単語数)	13170
TOTAL WORDS SELECTED	=16932 (処理対象となった単語数)	13170
TOTAL WORDS PICKED	=16932 (解析対象となった単語数)	13170
TOTAL WORDS SAMPLED	=16932	13170
TOTAL WORDS KEPT	=16932	13170

TOTAL VOCABULARY = 2141 (処理対象となった語彙数) 1947

- (a) FREQUENCY [出現頻度] (テキスト中の単語の「出現頻度」。このテキストでは「出現頻度」1から793までの単語が出てくる。ちなみに793回出てくる単語というのは you である。)
- (b) RELATIVE FREQUENCY [相対頻度] (出現頻度(a)の単語がテキスト全体の中で占める割合。最後の「出現頻度ごとの単語の百分率」に対して、「相対頻度」の数値は頻度ごとではなく個々の単語ごとのいわば「遭遇の度合い」を表している。特定の単語に我々がどの程度出会うかをこの数値は示している。16932語に1回しか出てこない単語は、100語からなるテキストでは $0.00591 (1 \div 16932)$ 語に1回出てくることになる。)
- (c) NUMBER SUCH [単語の個数] (テキスト中で、出現頻度が(a)である単語の語彙数。この「単語の個数」欄を見ると出現頻度1の単語が1071個あり、793回も出てくる単語が1個しかないことを知る。)
- (d) WORDS IN FREQUENCY [出現頻度と該当単語個数の積] (テキスト中で、出現頻度が(a)である単語の総数 = (a)×(c)この欄では、1回しか出てこない単語にこのテキストで我々は1071回出会うということを知る。出現頻度5の単語は 1個あるからこれらの単語にテキスト中で会う回数は延べ 5回である。)
- (e) VOCAB TOTAL [累積語彙数] (語彙(異語)総数=(c)の累計。「出現頻度」1の異語は1071個ある。出現頻度を2にまで上げると $1071 (出現頻度1) + 317 (出現頻度2) = 1388$ になる。これを出現頻度5回までにすると $1071 + 317 + 179 + 88 + 70 = 1725$ になる。つまり総語彙数2141の8割以上が、1回から5回までという比較的低い頻度になっていることがここで分かる。)
- (f) WORD TOTAL [累積単語数] (単語総数=(d)の累計。この数値は次のように計算する。出現頻度5までの単語の総数は $(1071 \times 1) + (317 \times 2) + (179 \times 3) + (88 \times 4) + (70 \times 5) = 2944$ である。語彙(異語)の点では80.57% (この数値が「語彙の累積百分率」)を占める出現頻度5までの単語も総単語数16932の内17.39% (この数値が「単語の累積百分率」)を占めるにすぎない。))
- (g) PERC. OF VOCAB [語彙の累積百分率] ((e)の百分率表示)
- (h) PERC. OF WORDS [単語の累積百分率] ((f)の百分率表示)
- (i) PERC. OF WORDS IN FREQUENCY [出現頻度ごとの単語の百分率] ((d)の百分率表示。「出現頻度と該当単語個数の積」の数値を総語数で割ったものである。出現頻度1の6.33は $1071 \div 16932$ の式から出たものである。例えば100語からなるテキストの中で、1回しか出てこ

ない単語 6.33 個に我々は遭遇する。出現頻度 1 の単語に 100 語テキストの中で我々が「遭遇する」率は 6.33%だと解釈してもよいだろう。一番下の出現頻度 793 の単語 1 個だけを分かっているならば、テキストの総単語数の 4.68% の単語は分かったことになるとも言える。

(j) TYPE / TOKEN RATIO

(総語彙数 / 総単語数)

(テキスト中の使用語彙の豊富さを示す値 = (e) ÷ (f) この値が 1.0 の時、1 度も同じ単語が使用されなかったことを意味する。)

総単語数で 16932 → 13170, 総語彙数で 2141 → 1947 の変化である。総単語数では 22.2% なのに、総語彙数では 9.06% しか減少していないことに注目したい。この点を詳しく見てみよう。英語字幕版に縮小された時に、開文社版語彙での出現頻度の低さ高さがどう影響を受けているだろうか。矢印の次に来るのは、英語字幕版で 1 回から 5 回の出現頻度の単語リストに出てこない開文社版の単語である。() 内は減少率を示している。

開文社版出現頻度 1 (1071 語) → 176 (16.71%)

開文社版出現頻度 2 (317 語) → 13 (4.1%)

開文社版出現頻度 3 (179 語) → 8 (4.47%)

開文社版で 1 回しか出てこない単語の 16.71% が省略されてしまっているのに、出現頻度が 2, 3 になると語彙の点では 4% 前後しか省略されていない。

一方もっと出現頻度の高い語で出現頻度を比較してみると次のようになる。

you	793→590 (26%)	What	168→125 (26%)	no	112→72 (35%)
I	479→360 (25%)	Oh	167→72 (56%)	all	110→65 (40%)
The	411→339 (17%)	here	147→107 (27%)	don't	108→83 (23%)
to	389→303 (22%)	me	145→109 (25%)	is	108→79 (27%)
a	373→309 (17%)	this	141→107 (24%)	your	108→96 (11%)
and	286→188 (34%)	on	139→98 (29%)	Have	105→82 (21%)
George	209→154 (26%)	Well	138→56 (59%)	right	103→64 (37%)
it	201→149 (25%)	My	134→112 (16%)	do	102→74 (27%)
of	199→141 (29%)	know	133→75 (44%)	I'm	100→74 (26%)
in	197→162 (17%)	for	120→98 (18%)	Now	97→53 (45%)
That	189→145 (23%)				

開文社版の高出現頻度の単語は英語字幕版での縮小・省略率が概して高い。これはつまり総単語数で 16932 → 13170 への変化 (22.2% の縮小) に高出現頻度の単語が関わり、総語彙数で 2141 → 1947 への変化 (9.06% の減少) に低出現頻度の単語が関わるということではないか。大ざっぱに言えば、高出現頻度の単語群が総単語数の変化を説明し、出現頻度 1 の単語が語彙の変化を説明するのではないか。793 回使われている you は例え 590 回に縮小されようと、高出現頻度の単語として生き残ってくるはずである。高出現単語群は多少減ってもデータとしての価値は余り変わらない⁵⁾。問題は出現頻度一回の語彙群がどのくらい英語字幕で失われていくかということである。

5) 既に指摘したように確かに惜しいデータが英語字幕版で省略されている場合が出現頻度 1 以外にもある。Now you are talking. の talking は開文社版で 16 回 (talk は 13 回, talked は 3 回) 使われているが、英語字幕版で 8 回 (talk は 12 回, talked は 2 回) に減っている。この減った部分に日常生活でよく使われるこの慣用表現が入っているのである。

本稿で指摘したようないくつかの慣用表現が英語字幕版に生き残っていない。これは誠に惜しい省略であり、可能な限り実際に我々がテープを聴きまた母国語としての英米人の話し手の助けを借りながら修正・補足すべきである。しかし先に挙げた開文社版出現頻度1の1071語の内、英語字幕版で省略された179語（1071語の16.71%にもなる）は次に見るように大部分基本的な単語であり、他作品データが入力されるにつれて基本的な単語にふさわしい出現頻度になることが予想される。across, affairs, Adam's, airport, Alright, also, apologize, Atta. awful, balmy, barn, bath, between, Blaine, bleed, blue, bub, bud, calf, cap, capsule, card, caress, celebrate, cemetery, claim, Colosseum, company, configuration, contemptible, convincer, crowd, da, decided, deep, desk, difference, dirty, doggonedest, dogs, downtown, dozens, dreams, dressing, drove, ears, earth-time. eaten, eating, Eats, Elmira, Employer, enjoy, Eve, everybody, Everytime, explorers, faces, Fade, fatted, Fault, feed, feeling, Female, fifty-five, fifty-six, figure, fix, flier, flying, forty, four-F, fourth, Français, freshman, George's, glasses, goat, golden, gotta, Greece, half-dozen, having, hears, Hill, hm, Hmmm, Horace, husky, interested, Jane, joint, lands, lay, lady, length, loafer, Lookit, Ma, manage, Mark, Martha, mind's, Momma, Mommy, money-grubbing, mostly, naturally, Near, need't, nephew's, nine, ninety, ninety-nine, ninety-three, Nobody, nor, nowadays, number, one-horse, ours, ox, page, palms, Papa, Parlez, Parthenon, Partridge, Peanut, personally, picked, piled, pines, played, plugged, Porgie, pull, pulling, pulls, queen, reverse, ring, rise, rubbing, salary, Samarkand, saving, screwy, Sea, sealegs, seed, serious, seventeen, seventeen-fifty, Shadows, share, Shhh, sled, slight, sis, smell, spoke, sport, song, sounded, spending, stood, stuck, studs, stood, sung, sunrise, talent, Thar, taxicab, taxi-driver, team, Tears, telling, terrific, thrill, Tilly, Tilly's, touch, tourist, treat, trust, tux, twelve, twenty-fourth, twenty-seven, Two-cents', unhappy, upstairs, usually, vous, ward, watering, Whatta, wives, wives, Wow, wring, Yep (OXFORD MICRO OCP は文章解析する時にまだ文法判断まではしないので、次のような語まで1語として数えている。everybody'd, everybody'll, What're, What've, When'd. F-r-a-n-k-i-n (frankincense のスペルを母親が子供に教えているところ。) 次は()内の語を言おうとして、途中で遮られるか言い淀んだものである。thous (thousand), tempe (temperature), th (this), tha (that), Unc (Uncle)。さらに電話番号を棒読みしたものまで (two-four-seven) 1語になっている。上記の単語リストからはこのような語は省略したが、統計上の数値から引くようなことはしなかった。)

以上例を挙げながら、英語字幕版の信頼性をコンピューターで検討してきた。本稿で指摘してきた聴き取り難い所は大部分が電話での相手の声などであるから、「物理的に」英語を母国語とする人々にも難解な所である。英語を母国語とする人々の間でも音声情報の文字化の際に違うことがあるのは、大部分この「物理的」原因のせいであろう。このような箇所の正確さを狙って時間を費やすよりは、出現頻度の多い動詞とか慣用表現を早く見つけようとした方が良いのではないだろうか。また分析するデータ量が多ければ、言い替え表現があると言う英語字幕データの限界はある程度克服される。当座の我々の目的「英単語・慣用句出現頻度調査」の為には英語字幕をスペルチェッカーで校正した程度で良い。語彙調査の点では、英語字幕版でも単語の出現頻度分析データとして十分満足のいく結果が得られるというのが本稿の結論である。

最後に OXFORD MICRO OCP が簡単に打ち出してくる統計表の利用法を考えてみよう。私が特に注目しているのは相対頻度の数値である。すでに述べたようにこれは我々が特定の単語に出会

う「遭遇率」である。これを利用して学生が頻繁に遭遇するであろう単語・慣用句を探し出すのである。限られた本数・語数だとデータの片寄りが考えられるので、まず洋画100本分約100万語の台詞データをきちんと編集することを当面の目標としよう。上映時間が平均して2時間とすると、2時間×100本=200時間となる。1日で聞いたりはなしたりする言語体験が仮に10時間とすると、200時間の濃密な洋画言語体験は日常生活での言語体験の20日分に相当すると考えて良いであろう。(別の比喻で考えてみよう。1冊のペーパーバック小説が仮に6万語前後の単語で書かれているとすると、100万語データベースはペーパーバック16冊前後の量に相当する。)これが達成された段階で相対頻度の数値が高い単語を探し出したら良い。相対頻度が高いということは、テキストの中でその単語に我々が会える確率が高いということである。その段階で出現頻度1とか2などという単語は、先ほどの比喻で言えば、密度の濃い言語体験を毎日10時間約3週間に渡って行ってやっと1度か2度しか現れないということである。基礎を勉強している学生にこのような単語を最初から提示する必要は全く無いだろう。基礎段階の学生には、相対頻度の極めて高い単語を慣用句・用法・意味別に整理分類して慣用句・用法・意味などで高い頻度のものから提示していったら良い。

相対頻度の極めて高い単語は最も基本的な単語だろうが、音として聞いた時にも基本的な単語であるかどうかは疑わしい。活字としての単語の意味・文法は教えているが、単語の音の面での意味・文法は教えてこなかったのではないか。活字にしたら即座に分かる単語が自然なスピードで話されているのを聴くと全く分からないということが頻繁に起こっている。例えばこの『すばらしきかな人生』の出現頻度10以上の単語は殆ど基本的な単語であるが、聴き取りの点でも基本的なものとして理解できているかどうかは疑わしい。高出現頻度の単語は個々の慣用句・用法・意味別のより精度の高い出現頻度を調べ、音としても基本的な単語になるよう徹底的に指導する必要がある。

各作品で5回以下の低出現頻度の単語については、出現頻度5も出現頻度4も等しく出現頻度1にして各作品を横断するよりに出現頻度を調べたらどうか。特定の状況に固有の出現頻度ではなくさまざまな状況・分野に渡って普遍的に使われる単語が見つかるのではないか。

また特定の状況・分野に特有の表現を探す必要も出てくるであろう。このためには洋画台本台詞情報データベースを状況・分野別に再編成して、その状況・分野別により精度の高い出現頻度調査を行う必要がある。この為に各作品の高出現頻度の単語を調べれば良い。例えば恋愛の場面に使われる表現を恋愛もの洋画の台詞情報や各作品の恋愛の場面の台詞を集めて探せば良い。

本稿では、英語字幕で日常会話表現データを収集し、頻度調査することを考察したのだが、実用段階に入っている欧語 OCR (文字読み取り装置)を使えば我々のデータは飛躍的に増大する。最近購入したマッキントッシュ IIsi と DATACOPY 社の ACCUTEXT というソフトを使えば、活字情報も英語字幕を上回る正確さ(本をコピーしたもので90%をはるかに越える)でスキャナーが自動的に読み取り、次々とコンピューターに入力・処理してくれる。我々は本をコピーし20枚毎セットし、入力処理終了後に数パーセントのミスプリをワードパーフェクトなどのソフトで直していけば良い。これを使えば、ウィスコンシン大学出版部から出ている映画台本シリーズ⁶⁾、南雲堂から出た英和对訳シナリオシリーズ⁷⁾も英語字幕情報データベースに付け加えることが出来る。さらに「英米の大学の講義で使われる単語」、「フレッシュマン・イングリッシュの教材に頻出する単語」、「経済学の大学院で読む教科書に頻出する形容詞、動詞」などの調査は、これらの機器・ソフトを使えば、簡単に出来るようになる。専門書講読に必要な語彙の修得を助けて我国あるいはアメリカの大学院での専門教育の一助になるよう配慮してやることも我々の大切な務めであるから、

6) Wisconsin / Warner Bros Screenplay Series としてすでに私の手元にあるだけで *The Sea Hawk* など20冊以上に及ぶ。

7) 『昼下がりの情事』(1965年出版)など20作品ほど出ているはずであるが全て絶版になっている。

学習効果をよく考えた教材作成・カリキュラム編成にこれらの機器・ソフトが果たす役割・効用は計り知れない。

現在教養英語担当者は数多くの大学英語教科書の中から「講読」用の評論とか小説などの文学作品、さらに時事的な新聞雑誌記事などを教材を選んで訳読中心の授業を行っている。このような状況下での洋画ビデオ教材利用促進を考えなければならない。教科書を選ぶような感覚で、教養英語担当者が市販洋画ビデオを教材として選び易いようにしなければならない。そのために是が非でも『洋画台本・台詞・映像・音声辞典』を編集する必要があるだろう。教師が利用し易いようにこの辞典はあらゆる便宜を計るべきである。どういう単語・慣用句がどの作品のどこに使われているかを担当者それぞれがこの辞典で調べ、英語字幕台本あるいは完全台本英語教科書版と共にその場面の映像と音声を教室で学生に示すことが出来るようになると良い。現在、活字教科書のみで小説、評論などを「講読」している教養英語担当者が、映像・音声を伴った教材を僅かな労力で自由に編集・作成できるようにこのような辞典、洋画台本教科書を整備しようというのが我々の意図である。英米の映画を英語教育に利用したいという英語教師の熱意に英米の映画会社がなかなか理解を示さないから、完全台本の教科書版が年に数冊しか出版されないのは大変遺憾なことである。ソニーとか松下がアメリカの映画会社を買収することになると、状況が大きく変わって日本の英語教育のために洋画ビデオが大いに利用し易くなる可能性がある。それに備えて、聴き取りを実際にやりながら英語字幕版を修正して完全台本を一本一本着実に作っていく必要があることも付記しておく。