話し言葉コーパスを用いた中学基本語彙の分析

山　本　五　郎
広島大学外国語教育研究センター

1．はじめに

本稿では，中学校英語教科書に使用すべき必修語として指定されている語彙リストと英語の話し言葉コーパスに基づく出現頻度順語彙リストを比較し，その差異について論じる。

英語学習において基礎となる中学校での学習内容について，新学習指導要領では，改訂前の「聞くこと」「話すこと」を重視した目標から，2012 年度から全面実施となった改訂後の要領では，「読むこと」「書くこと」を加えた 4 技能の総合的な英語力の育成を強調した目標へ移行している。しかしながら，目標の各項目内容に目を向けると，改定前の要領を踏襲し「英語を聞いて話し手の意向などを理解すること」に加えて「英語を用いて自分の考えなどを話すこと」に重きを置いており，基本的な学習指導方針に大きな変化はない。また義務教育以降の英語教育においても，国際的に競争力のある人材を育成するという目標のもと，文部科学省によるスーパーグローバルハイスクールやスーパーグローバル大学等事業が展開されており，より実践的な英語力の習得を後押しする傾向はますます強まっている。時世に鑑みると，日本の英語教育において所謂オーラルコミュニケーション力の養成が求められる傾向は，今後も変わらないであろう。このような背景の中，日本人英語学習者が文部科学省検定済の教科書を通して学ぶ語彙が，英語による日常会話で高頻度に使用される語彙をどこまで反映しているのかという点について知識を深めることは，より精度の高い教材開発を含め，今後の英語教育に資するものであるといえよう。

上記を踏まえ，本稿では，2000 年度の中学学習指導要領に指定されている必修語 507 語と，主に英語映画の台詞から構築した話し言葉コーパス[1] に基づく出現頻度順語彙リストを比較し分析する。

2．比較対象語彙リスト

中学学習指導要領の必修語彙リストについて，新学習指導要領では，その語数が従来の 507 語から 100 語へと大幅に抑えられており，中学校で学ぶべき総語数約 1200 語中に占める割合が低くなっている。塩見（2002）でも指摘されているように，検定済教科書であっても必修語以外の掲載語についてはテキストによってばらつきがあるため，本稿では検定済教科書での使用が必修とされる語彙をより広範囲に設定する目的で，2000 年度の学習指導要領において指定されている必修語彙 507 語を分析対象とした。

話し言葉コーパスに基づく頻度順語彙リストの構築にあたっては，コンコーダンサによる頻度検出を行った。文部科学省による学習指導要領改訂のポイントでは，語彙の選定について「綴りが同じ語は，品詞にかかわりなく 1 語と数え，動詞の語尾変化や形容詞や副詞の比較変化などのうち規則的に変化するものは原則として 1 語とみなす」と定めているため，この規定に沿って重複している語については同一の単語としてカウントした。同様に，オーラルコミュニケーションで頻出する短縮表現についても，短縮されていない例と同一とみなし，一語としてまとめた。ま

た，今回の比較分析の対象リストでは間投詞を含めなかったが，会話中高頻度に使用される間投詞表現については別途後述する。以上のように整理して話し言葉コーパスの頻出語を学習指導要領の必修語の語彙数にあわせ，上位から 507 語を抽出した。

３．比較分析結果

2000 年度の学習指導要領における必修語彙 507 語（C_1）と，話し言葉コーパスから抽出した同数の頻出順語彙リスト（C_2）を比較した結果，どちらの語彙リストにも出現した単語は 507 語中 309 語[2)] であった。語彙リスト間の類似度は Jaccard 係数によれば（1）のように数値化することができる。

（1）$Sim(C_1,C_2)=309/705$

以下，必修語彙と話し言葉頻出語彙のいずれか一方にしか出現しなかった単語を取り纏め，それぞれの語群の特性について記述するとともに，共通して現れた語についても英語教育における語彙学習の観点から考察する。

3.1　中学必修語彙にのみ現れた語群

二つの語彙リストを比較した結果，2000 年度の学習指導要領における必修語のリストでのみ確認された単語は表 1 に示した 198 語であった。

表１　中学必修語でのみ確認された198語

across	dark	Friday	Monday	season	table
afternoon	December	fruit	moon	sell	tall
already	desk	garden	mountain	September	teach
among	dictionary	glad	mouth	seven	teacher
animal	draw	glass	near	seventeen	tenth
April	during	goodbye	nice	seventh	third
arrive	ear	green	niece	seventy	thirteen
August	early	ground	nineteen	shop	thirty
aunt	eight	grow	ninety	short	thousand
become	eighteen	hers	ninth	shout	Thursday
begin	eighth	hill	noon	sing	tree
bird	eighty	holiday	nose	sixteen	Tuesday
blue	eleven	hundred	notebook	sixth	twelfth
boat	eleventh	interesting	November	sixty	twelve
box	English	introduce	October	sky	twenty
bread	enjoy	invite	often	slowly	uncle
breakfast	evening	January	ours	small	useful
build	fall	Japan	paper	smile	usually
building	famous	Japanese	park	snow	vacation
bus	farm	July	pen	sorry	village
busy	February	June	pencil	spend	visit
card	fifteen	kitchen	picture	sport	wall
carry	fifth	lake	poor	spring	warm
chair	fifty	language	popular	star	wash
class	finish	large	quickly	station	Wednesday
clean	fish	learn	rain	store	whose
cloud	flower	lend	rice	strong	wind
club	fly	letter	rich	student	window
college	foot	library	rise	study	winter
color	forget	lunch	river	summer	write
cook	forty	March	sad	sun	yard
cry	fourteen	May	Saturday	Sunday	yellow
cup	fourth	milk	sea	swim	yesterday

3.1.1　中学必修語彙にのみ現れた語群についての考察

中学必修語彙リストでのみ確認された語群の特徴としては，数字や暦に関する語彙が多く取り入れられていることである。198 語中 56 の単語が, 数字や, 曜日, 月の名前についての語彙であった。話し言葉コーパスの頻出語彙リストでは，暦に関する語彙は観察されず，数字についても一桁の数字がごく少数観察されただけにとどまった。

その他，特定のトピックに関する語群としては，class, learn, library, notebook, pen, pencil, student, study, teach, teacher, のような学校や勉学に関連した語彙が目立った。加えて，日本における英語教育用のテキストであるため，Japan, Japanese という語が入っていることも特徴の一つとして挙げることができる。これらの単語が含まれていることから，必修語 507 語は，実践的な運用の度合いのみに頼ったものではなく，一般教養的な知識としての英単語を学習させることにも配慮して選定されていることが伺える。

また，話し言葉コーパス頻度順語彙リストでは観察されなかった動詞 26 語[3] が中学必修語彙では含まれていることが分かった。ごく基本的な動詞が，話し言葉における高頻度語彙として現れないのは，興味深い事実と言える。

3.2　話し言葉コーパスにのみ現れた語群

話し言葉コーパスに基づいて作成した頻出順語彙リスト 507 語の内，中学での必修語彙に含まれていなかった単語は表 2 に示した 198 語であった。

表２　話し言葉頻度順語彙リストでのみ確認された198語

able	chance	follow	husband	ok	shot
actually	change	found	its	okay	shut
afraid	check	frank	Jack	order	side
against	children	free	Jesus	outside	sir
ahead	Christ	front	job	own	somebody
air	Christmas	fuck	Joe	part	sound
alive	clear	full	John	party	step
almost	company	fun	kid	pay	straight
alone	control	funny	kill	perfect	stuff
along	couple	gentleman	king	person	stupid
anybody	course	George	lady	phone	sweet
anymore	crazy	glad	later	pick	true
anyway	dad	god	law	piece	taken
around	daddy	goddamn	least	place	team
ass	damn	gonna	line	plan	thanks
ball	dance	gotta	lord	police	thing
bed	David	guess	lot	point	thought
behind	dead	gun	luck	power	told
believe	deal	hang	lucky	probably	tonight
bet	death	happen	ma(’am)	problem	top
better	die	Harry	matter	promise	touch
bill	doctor	hate	maybe	president	trouble
bit	dog	heart	men	pull	trust
bitch	Dr	hell	might	quite	truth
blood	else	hello	mind	real	wanna
body	end	hey	Miss	reason	war
brought	even	hi	moment	safe	whatever
business	everybody	himself	move	save	white
bye	exactly	hit	Mr	scared	whole
captain	fact	hold	Mrs	sent	wife
care	feet	honey	myself	set	wish
case	fight	hurry	nobody	shit	(New) York
cause	fire	hurt	number	shoot	yourself

3.2.1　話し言葉頻度順語彙リストにのみ現れた語群についての考察

話し言葉コーパスに基づいた頻度順語彙リストを作成するにあたり，今回は間投詞を語彙リストから除いたが，uh, huh, ah, などの表現が今回抽出した最頻度の語彙群と同等の頻度で会話中使用されていることが分かった。以下，（2）から（4）に例を挙げる。

（2）A: A little Coleman's mustard.
B: It's, *uh*, splendid.
A: Glad you like it.　(Meet Joe Black, 1998)

（3）A: See, the gorillas in the Pangani carry this recessive gene that pops up every four or five generations…
B: *Huh*.
A: With this rare form of gigantism. (Mighty Joe Young,1998)

（4）Yeah, he's, *ah*, he's gonna talk to the headmaster. (Scent of Woman, 1992)

また，be 動詞や法助動詞の短縮形（'m, 's, 'll, 've など）は，短縮されていない形とまとめて一語とカウントしたが, これらの短縮表現は口頭でのコミュニケーションにおいて多用されており，それぞれの非短縮形よりも使用頻度が高かったことは記しておかなければならない。とりわけ，is, has の短縮である 's，not の短縮である 't，am と are の短縮である 'm と 're については，今回抽出した 507 語の中でも，人称代名詞の I や you，冠詞の a や the，前置詞の to や of 等と並んで会話での使用頻出度が全体の上位 20 に入っていることが分かった。

中学での必修語として指定されていない単語で，話し言葉コーパスの頻出語彙リストに現れた語群で特徴的なものとして，固有名詞が挙げられる。David や George のような人名が，頻出語彙リストでは複数入っているのは興味深い事実である。York についてはその大半が New York として使用されたものであるが，ごく一部に Janet York や Alan York のような人名でカウントされたものもあった。いずれにしても, オーラルコミュニケーションでの使用頻度の高さに関わらず，これらのような固有名詞が検定済教科書で必ず取り上げられるべき必修語から外されているのは，英語教育上合理的だと言えるであろう。

上述したような個別のファーストネームに関連して，ファミリーネームと共起したり，呼びかけとして単独でも使用する Mr, Mrs, Miss などが頻出語彙に含まれていた点も，頻度順語彙リストの特徴の一つとして挙げることができる。

英語教育の観点から留意を必要とするのは，（5a）から（6b）の例に挙げたように，bitch, fuck などののののしり言葉が英語による会話では頻出するという事実である。これらの表現を日本人学習者が発表語彙（productive vocabulary）として学習し，使用することは差し控えるべきである。しかしながら，日常会話における頻出語彙約 500 語の中に見られるという事実がある以上，これらの語彙を受容語彙(receptive vocabulary)として学習する意味を否定することはできないであろう。

（5）a. Don't you know who I am?　I'm the Juggernaut, *bitch*. (XMEN the Last Stand, 2006)
b. I want to know who that *bitch* is. (Mr. and Mrs. Smith, 2005)

（6）a. A: I have a biology.
B: *Fuck* your biology. You always have excuses. (Murder by Numbers, 2002)
b. Where the *fuck* is the *fucking* tuna? (Bridget Jones's Diary, 2001)

また，今回作成した話し言葉コーパスの頻度順語彙リストでは，会話において，上述した Be 動詞や法助動詞の短縮とは別に，（7）から（9）に例示したような，gotta, gonna, wanna という発

音綴りによる短縮表現が多用されていることが観察された。

（7）You *gotta* tell me right now, is this man really your father? (August Rush, 2007)

（8）So you're not *gonna* let me stay in my own house? (Mona Lisa Smile, 2003)

（9）So do you *wanna* start grilling me now, or should we wait till another dinner?

(The Devil Wears Prada, 2006)

2000 年度に出版された中学及び高校の検定済教科書 160 冊に出現した語彙の頻度分析を行った塩見（2002）による語彙一覧表に照らし合わせると，これらの発音綴りによる短縮表現が検定済教科書に現れる頻度は非常に低く，それぞれ，9 千語，4 千語，1 万 4 千語レベルの語彙として扱われていることが分かる。これらのようなくだけた口語表現を，学習対象として扱うべきかどうかという点については賛否が分かれるであろうが，検定済教科書ではオーラルコミュニケーションで最頻出の表現形式に触れる機会が限られているという事実は留意しておく必要があると言えよう。

3.3　共通語彙についての留意点

507 語から成る二つの語彙リスト中，共通項目である 309 語についても，単語の多義性等に焦点をあてて精査すると注目すべき側面が現れる。例えば，双方の語彙リストに含まれている cool について，話し言葉における使用例を見ると，その約 15％が談話での相槌表現として使われていることが分かる。相槌表現については，Maynard et al.（1986）や Cutrone（2010）によって談話上の機能が分析されているが，話し言葉コーパスに見られる cool の用法を先行研究による分類に沿って分析すると，“Display of understanding of content（10）” や “Agreement（11）” など，多様な用いられ方をしていることがわかる。

（10）A：My mother found the armadillo. She fainted.
B：For real?
A：Just like in the movies…screamed and dropped.
B：*Cool*.　(Simon Birch, 1999)

（11）A：Leave that to me. Do you understand?
B：*Cool*.
A：Do you understand?
B：Yeah, I get it.
A：Well, just say so. (The Rite, 2011)

これらの例が示すように，検定済教科書で見出し語としては取り上げられているものの，相槌表現のようなオーラルコミュニケーション特有の意味や用法については，学習対象である単語の第一義を学習しただけでは十分ではないものも少なくない。口頭でのコミュニケーション能力を育成するための会話表現の学習という観点から見れば，基本的な単語についても一歩踏み込んで学習に値する意味や用法を整理していく必要があると言えるであろう。

4．まとめ

本稿では，2000 年度の中学学習指導要領での必修語 507 語と，話し言葉コーパスから作成した同数の頻度順語彙リストを比較し，英語教育の観点からその差異やそれぞれの語彙リストについての特徴について述べた。必修語では暦や学校など特定のトピックに関連する語彙群が多く含

まれていることが分かった。また，話し言葉コーパスによる語彙リストでは，発音綴りによる短縮表現やののしり言葉が高頻度で出現することに加え，相槌表現などオーラルコミュニケーションにおける特徴的な用法にも注意が必要であることを確認することができた。

本稿の比較では，約 500 語という限定的な語彙を対象としたため，今回の比較によって検定済教科書で取り上げられている語彙の特性について余すところなく論じることは困難である。日本人英語学習者が検定済教科書で学ぶ語彙について，特定のコーパスとの比較という手法でその特性を把握するためには，より広範囲な語を分析の対象とする必要があるといえる。新学習指導要領で目標とされている中学校で学ぶべき約 1200 語に加え，高等学校での学習語彙を含めて研究対象とすることができれば，その比較・分析の結果は，実践的な英語力が求められる英語教育により直接的に還元できるものとなるだろう。

注

1）本稿で使用した話し言葉コーパスは，801 万 6 千語の「映画英語教育学会 Movie English Caption Database（Ver.3.0 2009）」を主として用い，科学研究費助成事業の支援を受けて構築中の英語スピーチコーパス（課題番号 25770205）をサブコーパスとして使用した。

2）必修語と頻度順語彙リストのどちらでも確認された語 309 語

a	car	fine	in	my	same
about	catch	first	into	name	say
after	child	five	is	need	school
again	city	food	it	never	second
ago	close	for	just	new	see
all	cold	four	keep	news	send
also	come	friend	kind	next	shall
always	cool	from	know	night	she
am	could	game	last	no	should
an	country	get	late	not	show
and	cut	girl	leave	nothing	sick
another	daughter	give	left	now	since
answer	day	go	let	of	sister
any	dear	good	life	off	sit
anyone	different	great	light	old	six
anything	dinner	hair	like	on	sleep
are	do	half	listen	once	so
as	does	hand	little	one	some
ask	door	happy	live	only	someone
at	down	hard	long	open	something
away	drink	has	look	or	sometimes
back	drive	have	lose	other	son
bad	each	he	love	our	soon
be	easy	head	make	out	speak
beautiful	eat	hear	man	over	stand
because	either	help	many	people	start
before	enough	her	may	plane	stay
between	ever	here	me	play	still
big	every	hers	mean	please	stop
black	everyone	high	meet	pretty	story
book	everything	him	mine	put	street
both	excuse	his	minute	question	such
boy	eye(s)	home	money	read	sure
break	face	hope	month	ready	take
bring	family	hot	more	really	talk
brother	far	hour	morning	red	tell
but	fast	house	most	remember	ten
buy	father	I	mother	ride	than
by	feel	idea	much	right	thank
call	few	if	music	room	that
can	find	important	must	run	the

their	till/until	under	way	why	wrong
them	time	understand	we	will	year
then	to	up	week	with	yes
there	today	us	welcome	without	yet
these	together	use	well	woman	you
they	tomorrow	very	what	wonderful	young
think	too	wait	when	word	your
this	town	walk	where	work	yours
those	try	want	which	world	
three	turn	watch	white	worry	
through	two	water	who	would	

3）表 1 より，抽出した動詞

中学必修語でのみ確認された動詞26語

arrive	clean	fall	invite	sing	wash
become	cook	finish	learn	spend	
begin	cry	forget	lend	study	
build	draw	grow	sell	teach	
carry	enjoy	introduce	shout	visit	

参考文献

塩見知之（2002）．文部省検定済　中学校・高等学校教科書に現れた英語の語彙．北星堂書店．

Cutrone, P. (2010). The Backchannel Norms of Native English Speakers: A Target for Japanese L2 English Learners. *University of Reading Language Studies Working Papers* Vol.2. 28–37.

Maynard, S. (1986). On back-channel behaviour in Japanese and English conversation. *Linguistics 24*, 1079–1108.

ABSTRACT

A Corpus-based Analysis of English Vocabulary in Junior High School Textbooks

Goro YAMAMOTO
Institute for Foreign Language Research and Education
Hiroshima University

The Japanese government's policy on English education has been focusing on developing students' communication abilities so that students can accurately understand and appropriately convey ideas and information. Administrators also strive to foster students' positive attitudes toward communication in English. To fulfill the objectives, the Ministry of Education, Science and Culture has set up some pedagogical guidelines not only for how the target language should be taught at schools but also for textbook development. In particular, guidelines have been provided for what range of English vocabulary students should learn in Japanese compulsory education. However, almost no research has shed light on the validity of the vocabulary list for officially-approved English textbooks.

The objective of this paper is to analyze the 507 vocabulary items that must appear in each officially-approved Japanese junior high school English textbook. For the purpose of this study, a vocabulary list was derived from an English speech corpus to capture how well the textbook vocabulary reflects a group of English words that most frequently appear in oral communication. A comparative analysis of the two lists reveals a number of interesting features of the English words which students learn in the early stages of their studies. This paper explicates characteristic groups of nouns and verbs in each vocabulary list, as well as some colloquial expressions observed in the speech-corpus-based vocabulary list.