

# 経済学英語の特徴語抽出に関する試み

—ミクロ経済学を例に—

久井田 直 之

## I はじめに

文部科学省によるグローバル化に対応した英語教育改革が小学校、中学校、そして高等学校で進む中で、大学における英語教育もグローバル化への対応が進み、大学で専門科目の授業を英語で行う大学も増えてきている。しかし、教員が教室英語に慣れていないことや教授法の知識や技能の不足などの問題が生じたり（中井 2011）、学生の専門科目を英語で学ぶにあたっての準備不足（語彙不足など）により、内容理解が不十分である場合があるとの声も現場の専門科目の教員から上がっている。田地野・水光（2005）が提案する、<sup>1)</sup>「一般学術目的の英語」（EGAP: English for General Academic Purposes）を学んでから「特定学術目的の英語」（ESAP: English for Specific Academic Purposes）に移行する流れで、専門科目を英語で学ぶことができれば、理想的であるが、専門科目を英語で学ぶまでに準備の時間をあまり費やすことができない場合も多い。

これらの問題を解決しようとする試みの1つに、学生の語彙不足を補い、効率的に専門語彙を習得するための専門科目の特徴語リストの作成がある。コーパスの作成が容易になったことで、<sup>2)</sup> 特徴語リストの作成も近年、積極的に行われるようになってきている。（cf. 石川 2005）しかし、多くの特徴語リストは大学である程度「一般目的の英語」（EGP: English for General Purposes）の語彙を習得した後の学生向けのもので、高校卒業後すぐに大学で専門科目を英語で学ぶ学生向けのものではなく、<sup>3)</sup> 社会科学系の専門科目である。経済学に関しては、特徴語リストの作成が進んでいない。

本研究では、高校教科書と経済学の英語のテキスト（本研究では、ミクロ経済学のテキスト）の語彙比較をコーパス分析により行い、経済学英語の特徴語抽出の方法を検討する。

---

<sup>1)</sup> 田地野・水光（2005）は、英語を「一般目的の英語」（EGP: English for General Purposes）と「特定目的の英語」（ESP: English for Specific Purposes）に分け、さらに ESP を「学術目的の英語」（EAP: English for Academic Purposes）と「職業目的の英語」（EOP: English for Occupational Purposes）に分けている。そして、EAP を「一般学術目的の英語」（EGAP: English for General Academic Purposes）と「特定学術目的の英語」（ESAP: English for Specific Academic Purposes）に分け、EGAP を学んだあとに、専門分野のための英語である ESAP を学ぶような流れが理想的であると提案している。

<sup>2)</sup> コーパスとは、大量の電子化された言語テキストデータのことである。

<sup>3)</sup> Sutarsoy, Kennedy, and Nation（1994）は、経済学のテキストと EAP の教材の語彙分析を行った。

## II 特徴語の定義と分析方法

### 1 特徴語の定義

内山・中條・山本・井佐原(2004:169)は、「特定目的の英語」ESPにおける重要単語について、以下のように定義している。

ESPにおける「重要な語彙」とは、ESPにおける「重要単語」からなる語彙であるが、ここの「重要単語」とは、学習者にとって、「学ぶ必要のある単語」である。

「学ぶ必要がある単語」という定義に注目し、本研究も独自に特徴語の定義を行う。従来の特徴語リスト作成の際には、中学校や高校での既習語は除外されていたが、本研究では、大学入学後すぐに専門科目を英語で学ぶ学生を想定した特徴語リストの作成を目指すため、「高校での未習語で学習者が学ぶ専門分野で高頻度使用される語」もしくは「高校での既習語であっても高校で学んだ語彙と異なる専門的な語彙で、専門分野で使用される高頻度語」を特徴語と定義する。

### 2 ミクロ経済学テキスト

世界で広く使用されているテキストである *Principles of Microeconomics* (Karl E. Case, Ray C. Fair and Sharon M. Oster) をテキストファイルに変換し、このデータをマイクロコーパスとした。

### 3 高等学校英語教科書

内外教育の採択部数のデータに基づき、6科目の採択部数一位の教科書 (All Aboard English Communication I, Crown English Communication II, Crown English Communication III, Vision Quest English Expression I Standard, Vision Quest English Expression II, Hello there! English Conversation) をテキストファイルに変換し、このデータを高校教科書コーパスとした。

### 4 ソフトウェア

本研究の目的である頻度表の作成や特徴語の抽出を行うために、構築したコーパスを分析するソフトウェアとして、AntConcを使用した。AntConcは、コーパス間の総語数の違いにあまり左右されずに特徴語を抽出するために用いられる統計指標、対数尤度比 (loglikelihood ratio) を自動的に算出する機能を備えており、本研究にとって有用であると判断した。

### 5 見出し語化

頻度表の作成と特徴語抽出のために、本研究は染谷氏のレマリストを使用し、見出し語化を行った。

例えば, price, prices, priced, pricing は price に集約される. cost のようなスペルが同じで, 名詞の cost と動詞の cost がある場合, この見出し語化では品詞の区別は行っていない.

### Ⅲ 分析結果

#### 1 コーパスデータの詳細

マイクロコーパスと高校教科書コーパスを, AntConc を用いて分析し, 明らかになったコーパスの語に関するデータが表 1 である. 延べ語数に大きな違いがあるものの, 異なり語数の違いに大きな差は見られない. 高校教科書ということで, 語彙習得の配慮から, 多くの語が意図的に使用されていることが異なり語数に影響していると思われる.

表 1 コーパスデータ

コーパス	異なり語数	延べ語数
マイクロ	6682	261406
高校教科書	5555	99667

#### 2 高校教科書コーパスのカバー率

高校教科書コーパスで使用された語がマイクロコーパスとどのように重なって使用されているかを分析し, 算出したのがカバー率である. 表 2 に見られるように, 異なり語数と延べ語数に大きな差が生じている. この結果から, ミクロコーパスには低頻度語が多くあるものの, 高頻度語の多くは高校教科書コーパスにも使用されているために延べ語数のカバー率が高くなっていると考えられる.

表 2 高校教科書カバー率

	異なり語数	延べ語数
高校教科書	45.56%	89.80%

#### 3 高頻度語のカバー率

Laufer (1989) は, 既知語数が総語彙数の 95% を上回ることが内容理解には必要であると述べている. その 95% を目安に分析すると, 表 3 が示すように, 高校教科書コーパスがマイクロコーパスの高頻度語上位 400 語までほぼカバーできている (95% 以上のカバー率). そして, ミクロコーパスの高頻度語上位 1798 語で, ミクロコーパス全体の延べ語数の 95% 以上になる. この 1798 語での高校教科書コーパスの延べ語数カバー率は 92.40% で, Laufer (1989) の示した 95% を若干下回ったものの高校までの既習語で内容の理解ができる可能性が示されたといえるだろう.

表 3 高校教科書コーパスのマイクロコーパス高頻度語カバー率

マイクロコーパス (延べ語数のカバー率)	高校教科書コーパス カバー率 (延べ語数)
高頻度語上位 50 語 (45.63%)	96.06%
高頻度語上位 100 語 (55.78%)	96.1%
高頻度語上位 200 語 (66.46%)	95.46%
高頻度語上位 300 語 (72.98%)	95.6%
高頻度語上位 400 語 (77.38%)	95.43%
高頻度語上位 1798 語 (95.37%)	92.40%

しかし、次の表 4 の異語数のカバー率に目を向けると、高頻度語上位 400 語までは 90% を超えるが、1798 語では 78% 弱まで落ちている。このことは、高頻度語上位は既習語が多く含まれるが、下位の語になるにつれて、未習語が増えることを意味しており、内容理解のためには、マイクロコーパスの高頻度語で、高校未習語に注意が必要であると言える。

表 4 高校教科書コーパスのマイクロコーパス高頻度語カバー率 (異語数)

マイクロコーパス高頻度語	高校未習語数	異語数カバー率
高頻度語上位 100 語	6	94%
高頻度語上位 200 語	14	93%
高頻度語上位 300 語	17	94.34%
高頻度語上位 400 語	24	94%
高頻度語上位 1798 語	395	78.03%

#### 4 頻度リスト

表 5 の頻度順の上位 30 語の多くは、両コーパス共通して機能語ではあるが、マイクロコーパスの上位には、price, firm, cost, market, demand, produce というテキストの内容に大きく関わる特徴的な内容語が含まれている。頻度リストの上位の内容語は、特徴語となる可能性が高く、学習者が専門科目を学ぶ前に理解しておくべき語である。

表 5 各コーパス頻度リスト上位 30 語

順位	マイクロコーパス		高校教科書コーパス	
	語	頻度	語	頻度
1	the	17137	be	4613
2	of	9999	the	4594
3	be	7168	a	3052
4	a	6810	to	3030
5	to	6507	i	2241
6	in	6392	of	1956

経済学英語の特徴語抽出に関する試み（久井田）

順位	マイクロコーパス		高校教科書コーパス	
	語	頻度	語	頻度
7	and	4726	in	1950
8	that	3803	and	1586
9	for	2640	you	1578
10	price	2149	have	1294
11	firm	1709	it	1177
12	have	1693	that	1093
13	cost	1625	for	892
14	on	1588	s	853
15	this	1546	do	742
16	as	1525	what	625
17	at	1389	this	616
18	it	1341	he	596
19	or	1308	with	566
20	market	1268	we	557
21	demand	1266	on	555
22	we	1236	my	546
23	you	1212	go	495
24	by	1198	t	481
25	will	1179	at	467
26	with	1159	as	464
27	from	1099	can	461
28	produce	1093	not	432
29	if	1081	about	423
30	s	1073	people	399

そして、高頻度語上位から内容語の高頻度語を抽出し、まとめたものが表6である。

表6 ミクロコーパス内容語上位20語

順位	語	マイクロ	高校教科書
1	price	2288	22
2	firm	2166	0
3	cost	1882	11
4	market	1465	42
5	demand	1400	4
6	produce	1095	11
7	good	963	206
8	income	919	3
9	more	915	205
10	curve	909	0

順位	語	マイクロ	高校教科書
11	output	899	0
12	product	862	14
13	tax	820	3
14	increase	792	26
15	supply	769	5
16	marginal	731	0
17	capital	688	10
17	unit	688	6
19	use	686	208
20	total	661	2

内容語 firm, curve, output, marginal 以外の語は、高校教科書コーパスにも出現し、表 6 からは既習語がほとんどであることがわかる。しかし、ここで注意すべきことは、これらの語の中には既習語扱いできないものが含まれている可能性があるということである。例えば、7 位の good は見出し語化されているため、詳細に調査を行う必要があり、good の内訳をみると、マイクロコーパスは名詞での使用がほとんどであるのに対し、高校教科書コーパスで、すべての good が形容詞として使用されていた。また他にも 17 位の capital の場合は、同じ名詞として使用されていたが、高校教科書コーパスでの使用は「首都」という意味での capital で、マイクロコーパスで使用される capital の意味とは異なる。カバー率の観点のみで、既習語と判断すべきではないことがこれらの例から示され、特徴語リストを作成するためには、特徴語である内容語の上位語から精査し、学習者に有益な情報を特徴語と共に示すことが重要であると考えられる。

## 5 特徴語リスト

最後に AntConc を用いて、統計的指数である対数尤度比に基づく特徴語の抽出を試みた。結果が表 7 である。

表 7 ミクロコーパス高指数上位 20 語

順位	頻度	指数	語
1	2166	1417.402	firm
2	2288	1347.158	price
3	1882	1140.316	cost
4	1400	916.142	demand
5	1465	768.533	market
6	1095	657.356	produce
7	909	594.838	curve
8	899	588.294	output
9	919	568.682	income

順位	頻度	指数	語
10	862	540.918	product
11	820	504.58	tax
12	769	480.516	supply
13	731	478.357	marginal
14	661	410.447	total
15	792	407.287	increase
16	688	396.578	unit
17	639	396.185	profit
18	634	384.405	industry
19	584	382.162	quantity
20	622	376.666	labor

表6の高頻度語上位と比較すると、表6の good, more, capital, use の代わりに、profit, industry, quantity, laborが入っている。このことから、指数順では、good や capital のような例を見落とす可能性があるため、頻度順に基づく高頻度語リストから特徴語を抽出することが妥当であると考えられる。表6と表7の語に大きな差が見られない要因として、高校教科書コーパスの総語数が少ないことが考えられ、指数による特徴語のあぶり出しを効果的に行うためには、高校教科書コーパスの語数を大幅に増やす必要がある。

#### IV まとめ

本研究は、高校卒業後すぐに専門科目を英語で学ぶ学生を想定し、高校教科書コーパスとマイクロコーパスの語彙比較を行い、特徴語の抽出の方法を検討した。高校教科書コーパスとマイクロコーパスの重なりに焦点をあてながら、マイクロコーパス中で高頻度に用いられている内容語は特徴語リストに収録されるべき語で、コーパスの比較上既習語扱いされる語の中にも特徴語が含まれている場合があるため、個々の内容語の使用例を精査する必要があることを示した。

Nation (2013: 306) は、専門分野の背景知識がない状態で、文中の語彙の推測や辞書を活用しての語彙理解は難しいと指摘している。高校卒業後すぐに専門科目を英語で学ぶ学生にとって、語彙の問題は内容理解に大きな影響を与える。本研究の示した特徴語抽出方法の例は、内容理解を補助する特徴語リストの作成に貢献できると考えられる。

今後の課題としては、さらにコーパスを大きくすること、マクロ経済学との比較分析や特徴語リストの作成があげられる。今後もこれらの課題に取り組み、学習者に有益な特徴語リストの作成を目指したいと考えている。

## 謝辞

本研究の分析に際し、日本大学経済学部 齋藤 哲哉准教授には、多くのご教示を賜った。心より感謝申し上げる。

## 参考文献

- Anthony, L. (2014). AntConc (Version 3.4.4w) [Computer Software]. Retrieved July 7, 2015, from <http://www.antlab.sci.waseda.ac.jp/>
- Karl E. Case, Ray C. Fair and Sharon M. Oster. (2014). Principles of Macroeconomics, Pearson.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), Special language : From humans thinking to thinking machines (pp.316-323). Clevedon : Multilingual Matters.
- Nation, I.S.P. (2013<sup>2</sup>). Learning Vocabulary in Another Language, Cambridge : Cambridge University Press.
- Someya, Y. (n.d.) "e\_lemma\_no\_hypen." Retrieved July 7, 2015, from [http://www.antlab.sci.waseda.ac.jp/software/e\\_lemma\\_no\\_hypen.zip](http://www.antlab.sci.waseda.ac.jp/software/e_lemma_no_hypen.zip)
- Sutarsyah, C., Kennedy, G., and Nation, P. (1994). "How useful is EAP vocabulary for ESP? A Corpus-Based Study." RELC Journal, 25, 34-50.
- 石川慎一郎. (2005). 「司法英語 ESP 語彙表構築の試み - FROWN コーパスと米国司法文献コーパスの比較に基づく特徴語の抽出 -」『神戸大学国際コミュニケーションセンター論集』1, 13-28.
- 内山将夫, 中条清美, 山本英子, 井佐原均. (2004). 「英語教育のための分野特徴単語の選定尺度の比較」『自然言語処理』Vol.11, No.3, 165-197.
- 時事通信社. (2015). 「15 年度高校教科書採択状況 - 文科省まとめ (下)」『内外教育』, 6392, 4-11.
- 田地野彰, 水光雅則. (2005). 「大学英語教育への提言 - カリキュラム開発へのシステムアプローチ -」竹蓋幸生, 水光雅則 (編)『これからの大学英語教育』岩波書店, 1-46.
- 中條清美, 内山将夫. (2004). 「統計的指標を利用した特徴語抽出に関する研究」『関東甲信越英語教育学会研究紀要』第 18 号, 99-108.
- 中條清美, 吉森智大, 長谷川修治, 西垣知佳子, 山崎淳史. (2007). 「高等学校英語教科書の語彙」『日本大学生産工学部研究報告 B』第 40 巻, 71-92.
- 中井俊樹. (2011). 「英語による授業のための F D の課題」『留学 交流』2011 年 9 月号, 1-7.

## 高校教科書

Hello there! English Conversation	(酒井志延他, 東京書籍)
Vision Quest English Expression I Standard	(野村恵造他, 新興出版社啓林館)
Vision Quest English Expression II	(野村恵造他, 新興出版社啓林館)
CROWN English Communication II	(霜崎實他, 三省堂)
CROWN English Conversation III	(霜崎實他, 三省堂)
ALL Abroad! Communication English I	(清田 洋一他, 東京書籍)