



# TOEIC<sup>®</sup> Research Report NUMBER 3

“Utilizing the British National Corpus to Analyze TOEIC Tests: The Quantification of Vocabulary-Usage Levels and the Extraction of Characteristically Used Words”

「British National Corpusを活用したTOEICテストの分析  
— 語彙活用水準の定量化と特徴単語の抽出」

中條 清美 (日本大学生産工学部)

Michael Genung (日本大学生産工学部)



「British National Corpus を活用した TOEIC テストの分析  
－ 語彙活用水準の定量化と特徴単語の抽出」

中條 清美（日本大学生産工学部）  
Michael Genung（日本大学生産工学部）



## 目次

報告書 .....	1
1. 語彙活用水準の定量化 .....	2
1.1 BNC を活用した定量化の手順 .....	2
1.1.1 語彙活用水準の適切な判定基準 .....	2
1.1.2 BNC HFWL の作成 .....	2
1.1.3 テキストの語彙水準の測定方法 .....	3
1.2 3種類のビジネス英語テキストにおける BNC カバー率の測定 .....	4
1.2.1 3種類のビジネス英語テキストの語彙分析 .....	4
1.2.2 3種類のビジネス英語の語彙水準 .....	5
1.3 一般用途の英語テキストと英語習熟度テストにおける BNC カバー率の測定 .....	7
1.3.1 一般用途の英語テキストと英語習熟度テストの分析 .....	7
1.3.2 一般用途の英語テキストと英語習熟度テストの BNC 語彙水準の測定 .....	8
1.4 中学・高校教科書とビジネス英語テキストとの重なり .....	9
1.4.1 中学・高校教科書の語彙 .....	9
1.4.2 ビジネス英語のテキストに対する学校英語 教科書語彙の不足 .....	10
2. TOEIC テストに特有な語彙の抽出 .....	11
3. TOEIC 語彙リストの作成 .....	14
4. 結論 .....	14
参考文献 .....	15
Appendix TOEIC 語彙リスト(640 words) .....	17



## 報告書

本研究の目標は、(1) British National Corpus (以下、BNC) と中学・高校 (以下、JSH) 教科書の語彙を活用して、TOEIC テストの語彙水準を測定し、TOEIC に使用されている語彙の難易度を明らかにすること、(2) TOEIC テストに特有の単語を抽出すること、(3) JSH 教科書の語彙を習得した学習者の語彙を効果的に補足する単語リストを作成することである。前述の三つの目標はすべて達成された。第二の目標については、当初の計画よりも広範囲にわたる調査を実施することができた。筆者らは、本研究を進めるにあたって、ご協力くださった方々に心から感謝の意を表す。本研究の結果は以下のとおりである。

## 1. 語彙活用水準の定量化

本セクションでは、TOEIC 語彙の難易度を計測する際に BNC と JSH の語彙リストを活用し、TOEIC®テストの語彙水準を測定した。まず、**1.1** では、(BNC を活用した) 定量化の手順を説明する。次に、**1.2** では、BNC の高頻度語 (以下、BNC HFWL という) を利用して測定した、3 種類のビジネス英語テキストの語彙水準を提示する。**1.3** では、(BNC HFWL を活用して測定した) 一般用途の英語テキストと英語習熟度テストの語彙活用水準を示す。最後に、**1.4** では、JSH の語彙カバー率を活用して、異なる角度から 3 種類のビジネス英語テキストの語彙水準を測定し、結果について説明する。

### 1.1 BNC を活用した定量化の手順

#### 1.1.1 語彙活用水準の適切な判定基準

第一に、筆者らは、目標の英文を当該語彙でどの程度カバーできるかというカバー率を用いて内容理解の程度を計測することにした。言語を使いこなすために十分な言語知識があるか否かを示す言語知識の境界が存在するか否かについて、長い間関心が持たれてきたことは注目に値する (Nation, 2001)。従来、West (1926) をはじめとする経験豊かな教育者らは、文章を適正に理解するには、少なくとも 50 単語中 49 単語の知識がなければならない、との立場をとってきた。一方、Hatori (1979) や Johns (Bensoussan and Laufer, 1984) をはじめとする研究者は、後に Laufer (1989) が確認したように、95%の「カバー率」、つまり 20 単語中 19 単語の知識があるか否かが言語使用の成否を分ける境界であると考えた。その後、Hu and Nation (2000) は、学習者が主に楽しみのために自立的に読書を行なう場合は、テキストに使用されている単語の約 98%が既知語であることが必要であるとの結論を得た。しかし、最近の語彙学習および指導分野においては、内容理解に関するカバー率の境界は 95%とする考え方が主流のようである (Schmitt and McCarthy, 1997; Tono, et al., 1997; Read, 2000; Nation, 2001; Hayashi, 2002)。したがって、本研究ではこの 95%カバー率を目標水準に設定した。

#### 1.1.2 BNC HFWL の作成

1 億語を上回る単語で構成される BNC は、世界で最も信頼できるコーパス・リソースの一つとされている。BNC は、現代英語の書き言葉と話し言葉を代表するコーパスである (Leech, et al., 2001)。本研究の目的に合わせて、延べ 1 億語超のコーパスから基準リストを作成するにあたり、他の語彙リストとの比較に必要な大きさを確保するため、以下の手順によって語彙リストを作成した。

第一に、BNC コーパスを構成する延べ 100,106,029 語の中から、使用頻度が 5 回を超える語で構成される「all.num.05」と呼ばれる頻度リストを Adam Kilgarriff のウェブサ

イト、「BNC database and word frequency lists (BNC データベースと頻度順語彙リスト)」<sup>1</sup>からダウンロードした。

次に、このリストの比較可能性を効果的に最大にするために、頻度 100 以上の 38,683 語を以下の手順にしたがって整理した。(a)単語の屈折形を基本形のもとに集約した。たとえば、cat-cats、go-goes-went-gone-going といった単語の屈折形は、すべて基本形である cat、go に集約した。(b)固有名詞と数詞を除外した。(c)イギリス英語のスペリングをアメリカ英語のスペリングに変換した。(d)基本形が同形で異なる品詞の語は、同じ基本形に集約した。たとえば、answer という基本形には、名詞 4 つ、形容詞 2 つ、動詞 7 つの計 13 項目が集約された。

これらの手順を経て、最終的に、BNC の延べ 86,123,934 語を代表する 14,000 語の基本形リストが完成した。つまり、使用頻度に基づいて、BNC の頻度上位の語を 14,000 語の基本形に集約したものが、BNC 高頻度語リスト (BNC HFWL) ということになる。

### 1.1.3 テキストの語彙水準の測定方法

作成した BNC HFWL を用いて、計測対象の語彙リストとの重複部分の割合を計算することによって、比較が可能になる。また、語彙のカバー率を計算することもできる。カバー率は、テキストで用いられている語彙のうち、学習者が理解できると推測される語彙の割合をいう。

筆者らは、BNC HFWL との比較に基づいて、テキストやトランスクリプトの語彙水準を測定した。測定方法としては、テキストで用いられている語彙の 95% をカバーするのに必要な BNC HFWL の単語数を特定し、定量化する方法を用いた。言い換えれば、筆者らは、対象となるテキストの 95% の語彙を網羅するのに必要な数の単語を、BNC HFWL の上位から数えたことになる。このようにして BNC HFWL を利用して、さまざまなテキストの語彙水準を調査した。

---

<sup>1</sup> <http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/bnc-readme.html>

## 1.2 3種類のビジネス英語テキストにおけるBNCカバー率の測定<sup>2</sup>

### 1.2.1 3種類のビジネス英語テキストの語彙分析

筆者らは、まずビジネスに使用される58の話し言葉（主に対話）テキスト（以下、BNC ダイアログという）<sup>3</sup>を収集した。58のBNC ビジネス・ダイアログには、British National Corpusを構成する136の話し言葉のテキストから40件のビジネス会議、9件の相談、7件のインタビュー、2件のプレゼンテーションが選ばれた。次に、2001年4月から9月にかけて放送されたラジオ番組「NHK ビジネス英語」（以下、ビジネス英語という）の6か月分の放送テキストを収集した<sup>4</sup>。最後に、公開問題・模擬テストを含めた16種のTOEICテストを収集した。このうち6種は、一般公開されている公開問題より成る（以下、TOEIC 公開問題という）<sup>5</sup>。この中には、TOEIC テストで450点以下のスコアを取得した受験者の受験が推奨されているTOEIC Bridge®の模擬テストも1種含まれている（以下、TOEIC Bridge という）。残りの10種は、さまざまな筆者・出版社が発行した模擬テストである（以下、TOEIC 模擬テストという）<sup>6</sup>。表1は、3組のビジネス英語のテキストの語彙に関連する数値データ、選択した各テキストの延べ語数（単語の総数）と異語数（異なる単語の種類数）を示す。

収集したデータはすべて基本語形に集約した。すなわち、各項目について、同じ語幹を持つすべての屈折語を、基本形のもとに集約してアルファベット順に並べ、出現頻度を付加した。この集約にあたっては、他のリストと比較可能なリストを作成するために、BNC HFWL と同等の基準でリストを作成した。また、固有名詞と数詞については、「特定のテキストにおける出現頻度は高いものの、その他のテキストにはあまり出現し

<sup>2</sup> セクション1.2は、日本実用英語学会に提出した論文に詳述されている。“Comparing the Three Specialized Vocabularies Used in ‘Business English’, TOEIC, and British National Corpus Spoken Business Communications”と題するこの論文は、*Practical English Studies, Vol.11*の49ページから63ページに掲載されている。本論文には、TOEIC 運営委員会に対する謝辞が含まれている。

<sup>3</sup> これらは、BNCの話し言葉から得たデータである。

<sup>4</sup> NHKラジオ番組「ビジネス英語」の6か月分のテキストに使用されている単語の延べ語数は30,458語で、異語数は3,042語である。

杉田敏 (2001) やさしいビジネス英語、NHK、15、1-6

<sup>5</sup> 以下の3種の第一回～第三回テストは、一般公開されているTOEIC 公開問題より成る。

TOEIC 運営委員会(1981)第一回 TOEIC 問題集、東京、英文朝日

T. F. Communications (1997) ‘Practice TOEIC’ トーイックフレンズ、3(4)、24-53

TOEIC 運営委員会(1982)第三回 TOEIC 問題集、東京、英文朝日

The Chauncey Group International (2000) TOEIC 公式ガイド&問題集

The Chauncey Group International (2002) TOEIC 公式ガイド&問題集 Vol.2

The Chauncey Group International (2003) TOEIC Bridge 公式ガイド&問題集

<sup>6</sup> 10種のTOEIC 模擬テストは、T. F. Communications (1997)、Nagase (2000)、ALC (2002)などの出版社が発行した書籍から収集した。参考文献の詳細については、中條、牛田、山崎、Genung、内堀、西垣 (2004)を参照されたい。セクション2、セクション3では、公開問題・模擬テストの両方を使用した。セクション2、セクション3で用いた分析において信頼性のある結果を得るには10万語を超える規模のデータ(延べ語数)が必要であるため両方のデータを使用した。

ない...、また、テキストの文脈で意味が確定するため、事前に教授できない」(Nation, 2001: 19-20) との理由で、各言語材料から手作業で除外した。

表 1 三種類のビジネス英語テキストの語彙

出典	テキスト数	延べ語数	異語数
BNC のビジネス対話 (BNC ダイアログ)	58	474,613	6,878
NHK ビジネス英語 (ビジネス英語)	6	30,458	3,042
TOEIC 公開問題・模擬テスト	16	107,081	5,016

### 1.2.2 3種類のビジネス英語の語彙水準

筆者らは、表 2 に示すテキストデータの語彙水準を計測した。これらのテキストデータは、5 種類の TOEIC 公開問題、1 種類の TOEIC Bridge、6 ヶ月のラジオ番組「ビジネス英語」の放送テキスト、そしてビジネス関係の会議、協議、インタビューから収集した BNC の話し言葉のトランスクリプト 58 種類で構成される。これらの素材に使用されている語彙を BNC HFWL と比較して語彙水準を計測した。各テキストに使われている語彙のうち 95% をカバーするのに必要な BNC HFWL の単語数を特定・定量化し、対象テキストの語彙水準を決定した。結果は図 1 のとおりである。

表 2 語彙水準およびカバー率を計測したテキスト

出典	テキスト数	平均延べ語数	平均異語数
TOEIC 公開問題 (TOEIC テスト)	5	6,836	1,366
TOEIC Bridge	1	2,358	637
BNC ダイアログ	58	8,183	833
ビジネス英語	6	5,078	1,015

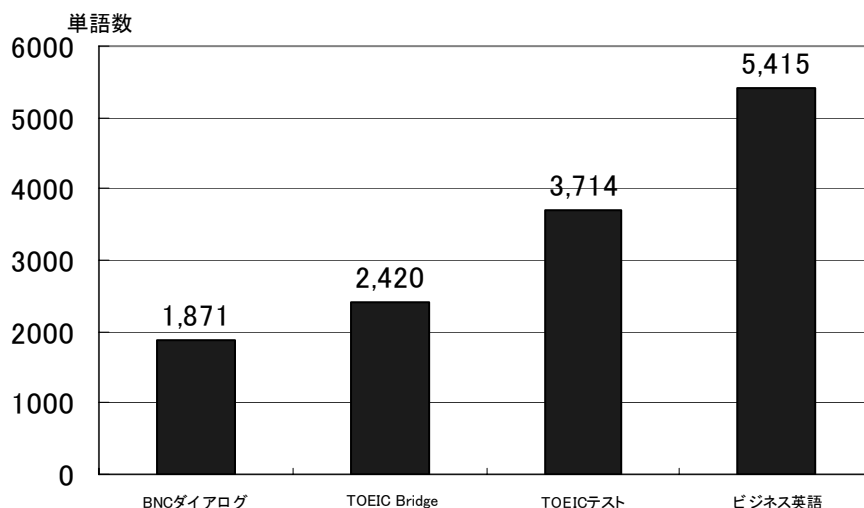


図1 BNC HFWLに基づく、TOEICテスト、TOEIC Bridge、および他の2種類のビジネス英語の語彙水準

図1の縦軸は、本研究で収集した各テキストで用いられている語彙の95%を網羅するのに必要なBNC HFWLの単語数を示す。5種のTOEICテスト、58種のBNCビジネスダイアログ、6冊のビジネス英語テキストの各語彙水準を平均したのがこのグラフである。5種のTOEIC公開問題に使用されている語彙の95%を理解するには、平均で、BNC HFWLに含まれている上位3,714語の知識が必要である。TOEIC Bridgeについては、上位2,420語の知識が必要である。図1のグラフを見ると、テキスト別の語彙水準は、予想どおりの結果になったといえる。会話の語彙水準が最も低く、TOEIC Bridge、通常のTOEICテストへと進むにつれて語彙水準は徐々に高まり、NHKラジオ「ビジネス英語」で急激に高まる。このグラフを見ると、3種類のビジネス英語テキストの中では、BNCダイアログの語彙水準が最も低いことがわかる。このことから、実際のビジネスコミュニケーションにおいて使用されている話し言葉は、共通語彙から最小限の単語を効率的に選択して構成されていることが推測できる。グラフを見ると、調査の対象となったNHKのラジオ番組「ビジネス英語」のテキストの95%を理解するには、BNCの頻度上位5,415語の知識が必要ながわかる。ラジオ「ビジネス英語」のテキストには、就職面接や人事関連など、典型的なビジネスの話題はもちろん、幅広い社会・経済問題に関連する話題も盛り込まれている。「TOEICテスト」の語彙水準が実際にビジネスで使用されている話し言葉（BNCダイアログ）の水準よりも高く、教材（ラジオ・ビジネス英語）の水準よりも低いことから、TOEICテストの語彙水準は、評価ツールとしては適切であると推測できる。

### 1.3 一般用途の英語テキストと英語習熟度テストにおける BNC カバー率の測定

#### 1.3.1 一般用途の英語テキストと英語習熟度テストの分析

表 3 と表 4 には、それぞれ筆者らが収集したテキストとテストをリストした。これらのテキストは二つのグループに分類できる。一つはタイム、ニューズウィークの記事、CNN ニュース、ABC ニュースのトランスクリプトといった一般用途の英語テキストや映画「タイタニック」、「クレイマー・クレイマー」のトランスクリプト、もう一つは TOEFL、英検 2 級、英検準 1 級、英検 1 級といった英語習熟度テストである。

表 3 一般用途の英語テキスト

出典	テキスト数	平均延べ語数	平均異語数
タイム(2002/6/17)	2	9,236	2,077
ニューズウィーク(2002/8/22)	2	9,051	1,989
CNN ニュース (2002/7/2,5,8,18,19,26, 8/8)	2	8,942	1,737
ABC ニュース (2002/8/28, 9/4)	2	9,357	1,715
映画 (タイタニック)	1	9,932	1,358
映画 (クレイマー・クレイマー)	1	9,886	988

表 4 英語習熟度テスト

出典	テキスト数	平均延べ語数	平均異語数
英検 2 級(2000 年・2001 年)	2	4,204	841
英検準 1 級(2000 年・2001 年)	2	6,039	1,445
英検 1 級(2000 年・2001 年)	2	7,278	1,760
TOEFL 模擬テスト(テスト A・B)	2	7,094	1,470

一般用途の英語テキストに選択されたものは、大学の英語の授業で ELT 教材としてしばしば使用されているものである。また TOEFL、英検、TOEIC テストは日本の多くの大学が英語の実用能力を評価する方法としてこれらのテストを採用しており、所定のスコアを取得した学生に英語の単位を与える大学も多い。

一般用途の英語テキストからそれぞれ 10,000 語の標本を 2 組ずつ収集した。映画については、各映画を 1 標本とした。英語習熟度テストについては、TOEFL テストから 2 組、3 レベルの英検テストから 2 組を収集した。次に、各標本から固有名詞と数詞を手作業で除外した。さらに、これらのテキストデータを基本形に集約して、出現頻度付きのアルファベット順リストを作成した。これらの語彙リストを作成する際には他の語彙リストとも比較可能なリストにするために、BNC HFWL 作成と同等の基準で語彙リストを作成した。

### 1.3.2 一般用途の英語テキストと英語習熟度テストの BNC 語彙水準の測定

一般用途の英語テキスト 6 種と習熟度テスト 4 種のそれぞれにつき 2 組ずつのサンプルの語彙水準を測定し、2 組の平均を求めた。TOEIC Bridge と TOEIC テストの語彙水準は図 1 の計算結果を使用した。一般用途の英語テキストとテストの語彙水準は、それぞれ 図 2 と図 3 に示すとおりである。

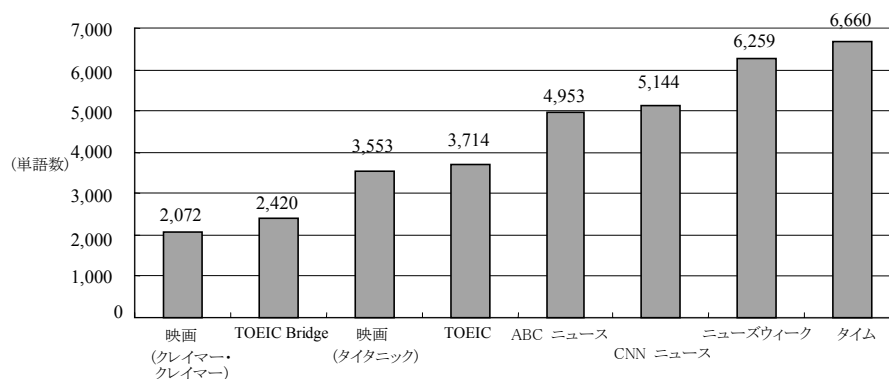


図 2 BNC HFWL を利用した TOEIC テスト、TOEIC Bridge、一般用途の英語テキストの語彙水準

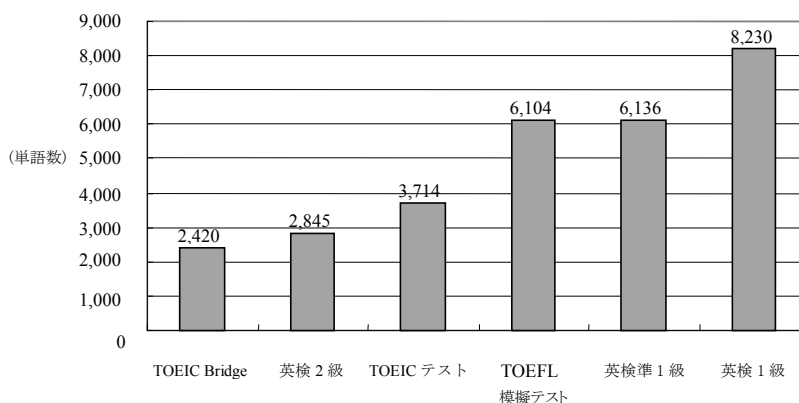


図 3 BNC HFWL を利用した TOEIC テスト、TOEIC Bridge、その他 4 種の英語習熟度テストの語彙水準

図 2 のグラフは、いくつかの興味深い結果を示している。第一に、6 種の一般用途の英語テキストは、語彙水準の点で、映画、テレビニュース、定期刊行物の三つのグループに分類できる。第二に、予想にたがわず、これら三つのグループには、話し言葉の語彙水準が書き言葉の語彙水準よりも低いという共通点がある。映画からテレビニュース、さらに定期刊行物であるニュースウィークとタイムへと移行するにしたがって、語彙水準は着実に高くなる。大学の英語の授業で ELT 教材として使用されることの多い TOEIC テストは、映画とテレビニュースという、種類の異なる話し言葉の中間に位置する。TOEIC Bridge は二つの映画の中間に位置する。

図 3 のグラフも、いくつかの興味深い結果を示している。6 つのテストを見ると、TOEIC Bridge から英検 2 級、TOEIC テストへと移行するにしたがって、語彙水準は徐々に高くなり、TOEFL テストと英検準 1 級に近づくにしたがって急激に高くなり、英検 1 級で最高に達している。英検 1 級の語彙水準は、十分な言葉の流暢さがあることを示すという点で、日本人の英語学習者の期待と一致する。中條・竹蓋 (1994) は、日本人の英語学習者がさまざまなコミュニケーションの目標を達成するには、約 7,000 語から 8,000 語に相当する語彙水準が必要だと推測した。彼らの推測は図 3 が示す英検 1 級テストの語彙水準に一致している。

図 3 を見ると、BNC HFWL を活用した測定では、TOEIC テストの語彙水準と、英検 2 級（高卒者の目標水準）の語彙水準の間には、約 900 語の差があることがわかる。こうした結果は、語彙水準という点で、TOEIC テストが大学生に到達可能な範囲内に含まれることを示すと考えられる。つまり、適切かつ規則的に厳しい語彙学習の訓練を受ければ、大学生は TOEIC テストの語彙水準に相当する語彙水準を獲得できる、と結論づけることは妥当といえよう。

#### 1.4 中学・高校教科書とビジネス英語テキストとの重なり

本研究の次の部分では、JSH 教科書<sup>7</sup>の語彙は、三種類のビジネス英語テキストで使用されている語彙をどの程度網羅しているかを計算した。

##### 1.4.1 中学・高校教科書の語彙

筆者らは、大学生又は大学卒業生が入学前に英語を学習した中学・高校（JSH）教科書のうち全国での採択率の高いものを収集した。それらは、中学 1 年から 3 年まで、および、高校 1 年から 3 年まで<sup>8</sup>、日本の学校で最も幅広く使用されている教科書の 1 つであるニューホライズン 1、2、3 とユニコーン I、II リーディングである。これらの教科書の延べ語数と異語数は表 5 のとおりである。各 JSH 教科書のデータから固有名詞と数詞を手作業で除外した後、基本形の JSH 語彙リストを作成した。

<sup>7</sup> 使用した JSH 教科書は次のとおりである。

*New Horizon English Course 1, 2, 3* 浅野博他(1999 年)東京書籍

*Unicorn English Course I, II, Reading* 末永國明他(2002 年)文英堂

<sup>8</sup> 日本の学習者が中学・高校で使用するものは、1 種類の教科書シリーズのみである。日本人学習者が語彙を獲得する最も典型的な例をシミュレーションするために、筆者らは、先に述べた最も採択率の高い 2 種類の教科書シリーズ、言い換えれば、最も幅広く使用されている教科書シリーズを選択した。日本の中学・高校教科書は、文部科学省が定める学習指導要領にしたがって執筆されている。学習指導要領は約 10 年に 1 度改訂される。本研究では、1990 年代の終わりに執筆された教科書の語彙を検討した。現在、(改訂された学習指導要領にしたがって 2000 年代初めに改訂された)教科書が販売されたため、その語彙水準を再検証する必要がある。

表 5 中学・高校教科書の語彙

教科書	延べ語数	異語数
ニューホライズン 1、2、3	9,440	1,124
ユニコーン I、II リーディング	36,678	3,478
合計 (JSH 教科書の語彙)	46,118	3,747

#### 1.4.2 ビジネス英語のテキストに対する学校英語教科書語彙の不足

次に、筆者らは、JSH 教科書に使用されている語彙が、ビジネス英語のテキストに使用されている語彙とどの程度重なるかを計算した。ビジネス英語のテキストと JSH のテキストのそれぞれについて単語の重なりを調べ、JSH 教科書に含まれるビジネス英語の語彙の割合を計算した。結果は表 6 のとおりである。この計算の目的は、ビジネス英語のテキストの語彙水準を推定するとともに、各テキストのコミュニケーション目標を達成するために増強する必要のある語彙量を明らかにすることにある。

表 6 JSH 教科書語彙のカバー率

	BNC ダイアログ	TOEIC Bridge	TOEIC テスト	ビジネス英語
JSH 教科書	93.1 %	94.7%	88.7%	89.0%

表 6 を見ると、JSH 教科書の語彙の知識では、本研究の調査対象である TOEIC テスト、TOEIC Bridge、他のビジネス英語の二種のテキストのいずれについても、十分に対応できないことがわかる。Laufer (1982, 1992) や Nation (2001) といった研究者は、学習者が文章の意味を理解するには、その 95%を網羅する語彙水準が必要だと指摘する。つまり、未知の単語は 20 単語あたり 1 単語でなければならない。表 6 は、JSH 教科書の語彙は TOEIC Bridge の 94.7%、TOEIC テストの 88.7%を網羅している。これは、TOEIC Bridge では 18.9 単語あたり 1 単語、TOEIC テストでは 8.8 単語あたり 1 単語、未知の単語があることを意味する。このことは、TOEIC Bridge、TOEIC テストのいずれについても、学習者がこれらを理解できる知識水準に達していないことを意味する。JSH 教科書は General English (標準英語) のテキストに分類されること、それゆえ JSH 教科書はビジネス英語に代表される ESP (English for Specific Purposes: 特殊目的のための英語) 語彙とは異なる標準英語の語彙使用を代表することを考えると、ビジネス英語に求められるコミュニケーション能力を獲得する基礎となる JSH 語彙を増強したい学習者にとっては、ビジネス分野の専門用語をある程度補完する必要があると結論づけられる。

## 2. TOEIC テストに特有な語彙の抽出<sup>9</sup>

次に、筆者らは TOEIC テストに特有な語、すなわち、BNC HFWL といった「標準的に使用される」語彙に分類される語とは明らかに異なる語の抽出を試みた。こうした単語があるために、教育関係者や学生から、TOEIC の語彙は「標準的に使用される」単語とは異なるとの声があがっている。Nation (2001:18) は、「専門用語を抽出する一つの方法は、一般的なコーパスにおける出現頻度と、特殊なテキストにおける出現頻度を比較することである」と示唆している。こうした示唆を実行に移すべく、中條・内山 (2004) は、複数の統計的指標を利用して、前述の出現頻度を比較し、TOEIC 特徴語の単語リストを習熟度レベルごとに抽出する手法を提案した。

中條・内山 (2004) は、出現頻度、ダイス係数 (Manning and Schütze, 1999)、補完類似度 (CSM) (Wakaki and Hagita, 1996)、対数尤度比 (Dunning, 1993)、カイ二乗検定 (Hisamitsu and Niwa, 2001)、イエーツの補正公式 (Hisamitsu and Niwa, 2001)、コサイン (Manning and Schütze, 1999)、自己相互情報量 (Church and Hanks, 1989) という 8 種類の統計的指標を利用して TOEIC 特徴語を抽出した。BNC HFWL を基準リストとして利用して、16 種の TOEIC 公開問題・模擬テスト<sup>10</sup>から作成した語彙リストにこれらの指標を適用した。図 7 に統計的指標別に抽出された特徴語の上位 20 単語をまとめた。

表 7 TOEIC 公開問題・模擬テストにおける特徴語上位 20 単語の抜粋

ランク	出現頻度	ダイス係数	対数尤度比	カイ二乗検定/ イエーツの補正/ コサイン/CSM	相互情報量
1	the	company	office	check-out	cross-cultural
2	be	what	refer	downtown	discontinue
3	a	will	employee	e-mail	cookbook
4	to	office	will	upcoming	reorder
5	of	question	company	hamburger	short-sleeved
6	in	refer	question	copier	comfortingly
7	you	follow	what	ferryboat	lost-and-found
8	will	new	sale	teal	ferryboat
9	have	man	please	beverage	taxicab
10	for	you	hotel	interoffice	carefulness
11	and	employee	customer	reimburse	preempt
12	I	woman	follow	accordance	no-smoking
13	do	service	vacation	vacation	below-mentioned
14	on	sale	store	payload	paper-recycling
15	it	a	computer	alumni	security-cleared
16	at	at	a	sightseeing	checkpoint
17	we	please	service	salespeople	prepackaged
18	what	business	business	newsstand	fabricate
19	this	do	mail	forfeit	conditionally
20	they	how	woman	requisition	first-come-first-served

<sup>9</sup> これらの手順とその結果は、2004年3月に関東甲信越英語教育学会に提出、発表された論文、中條清美・内山将夫、「統計的指標を利用した特徴語抽出に関する研究」に記述されている（『関東甲信越英語教育学会研究紀要』第18号、99-108ページ）。この論文には、TOEIC 運営委員会に対する謝辞が含まれている。

<sup>10</sup> このデータは、すでに表 1 で使用している。このデータは異語数 5,016 語、延べ語数 107,081 語に相当する。

これらの結果に基づいて、筆者らは、出現頻度とダイス係数を使用する方法は入門レベルに、対数尤度比は初級レベルに、カイ二乗検定、イエーツの補正公式、コサイン、補完類似度は中級レベルに、自己相互情報量は上級レベルに適した特徴語の特定に有用であることを提案した。

各学習者に適切な習熟度レベル別の単語リストを作成することは、ベテラン教師にとっても容易ではないため、統計的指標を利用して専門領域の特徴語を特定する方法は、有益かつ有望な方法である。

さらに、Chujo and Genung (脚注<sup>2</sup>で述べた論文の中で)は、これらの統計的指標を利用することで、三種類のビジネス専門用語を特定できたことを報告した。Chujo and Genung (2004)は、対数尤度比と自己相互情報量を活用して、前述の三種類のビジネス語彙に特有の単語を特定し、TOEIC 語彙と、BNC の話し言葉データおよびラジオ番組「ビジネス英語」に含まれる語彙を比較した(30,458 語)。表 8 は比較結果の抜粋である。

表 8 自己相互情報量を活用して抽出した三種類のビジネス語彙に特有な上位 20 単語の抜粋

順位	BNC ダイアログ	ビジネス英語	TOEIC テスト
1	photocopy	healthcare	refund
2	pallet	online	merchandise
3	erase	acupuncture	consulate
4	byte	high-tech	photocopy
5	spreadsheet	gourmet	supervisor
6	bearing	clout	bookcase
7	advertiser	fickle	pharmacy
8	worksheet	yoga	reimbursement
9	conductivity	well-informed	memo
10	raffle	severance	instructional
11	costing	teamwork	receptionist
12	divisional	high-speed	dishwasher
13	folder	gadget	payroll
14	accrue	burner	incorporated
15	assignment	surf	fax
16	turnaround	carnation	relocate
17	chute	enroll	typist
18	overtime	perk	banquet
19	seconder	overboard	renovation
20	empowerment	chemotherapy	enroll

三つのリストは興味深い対照をなしている。自己相互情報量は、各ビジネスコミュニケーションに使用される特徴的な「ビジネス専門用語」を特定するのに有効であった。たとえば、BNC ダイアログには、pallet, costing, accrue, turnaround など、貿易、流通、金融のビジネス活動に使用される単語が含まれている。また、BNC ダイアログには、photocopy, erase, byte, spreadsheet, worksheet, folder など、事務、特にコンピューター関連

の事務に使用される単語も含まれている。一方、ラジオ・ビジネス英語の語彙は、聞き手の注意を引くために番組が選んださまざまなトピックの影響を受けている。Healthcare, gourmet はその一例である。また、代替医療のトピックに関連して、acupuncture, yoga, chemotherapy といった語彙が使用されている。また、この番組は IT を取り上げることが多いためか、online, high-tech, high-speed, surf, gadget といった単語が上位 20 単語に含まれている。上位 20 単語には、当然、perk, severance, teamwork といった人事関連の単語も見られる。

TOEIC テストの上位 20 単語には、supervisor, receptionist, payroll, relocate, typist といった人事関連の単語、photocopy, bookcase, memo, fax といった事務関連の単語、refund, reimbursement といった会計関連の単語、pharmacy, renovation, dishwasher といった日常生活関連の単語が含まれ、同テストがビジネスや日常のコミュニケーションに関連する幅広いトピックを網羅していることがわかる。

Chujo and Genung (2004) では、中條・内山 (2004) が実施した最初の実験に基づいて、ノイズ単語を減らすことによって統計的指標の適用方法を向上させた。結果、表 7 に見られる自己相互情報量による特徴語上位 20 語 (1 番右の列) と、表 8 の 4 列目に示す「TOEIC テスト」の上位 20 語の間には相違が見られる。

### 3. TOEIC 語彙リストの作成<sup>11</sup>

筆者らは、学校英語教科書（JSH）語彙を習得した学習者の語彙を効果的に補完する単語リストの作成を試みた。適切な出現頻度と幅広い領域を網羅するという二つの基準にしたがって、（Appendix に添付の）640 語で構成する TOEIC 語彙リストを作成した。このリストは、TOEIC と JSH 教科書の語彙のギャップを埋め、指導者と学習者の教授・学習活動の効果を高めるのに役立つと考えられる。このリストに含まれる TOEIC 語彙の有効性は、TOEIC 公開問題のカバー率を計算して確認した。計算にあたっては、日本の現状に即した状況を設定した。すなわち、最も採択率の高い中学教科書（ニューホライズン、1、2、3）に見られる非常に基本的な語彙と高校教科書に見られる入門レベルの語彙（すなわち中学教科書のレベルを超えない）を習得した大学生が、初級英語を学習するという設定である。こうした学生の有する推定語彙力では、（T. F. Communications が 1997 年に公開した）第二回 TOEIC 公開問題に使用されている語彙の 77.1%しか網羅していない。同じ学生が 8 単位に相当する大学英語教科書の語彙を学習すると（1 週間に 2 回英語の授業を受け、8 単位取得に必要な大学英語教科書に出現するすべての語彙を習得すると仮定した場合）、第二回 TOEIC 公開問題に使用されている単語の 94.2%を網羅する語彙を習得することになる。これに TOEIC 語彙リスト（640 語）を加えると、網羅できる語彙は 96.9%へと増加する。前述したとおり、語彙の獲得と知識に関連した第二言語の読解力の分野における最新の知見では、有意な知識の境界は 95%とされている（Nation, 2001:146）。これらの理由によって、筆者らは、TOEIC に特有の語彙リストは、テキストの理解に必要な語彙の境界とされる 95%を達成するのに有効であると結論した。

### 4. 結論

本研究によって、TOEIC テストと TOEIC Bridge に特有の語彙の特徴が明らかになった。本研究の結論は以下のとおりである。

1. TOEIC Bridge に必要な語彙は、（高卒者に望ましい目標水準とされる）英検 2 級よりも少ない。一方、JSH 教科書の語彙は、TOEIC Bridge の語彙の 94.7%を網

---

<sup>11</sup> 本論文では、TOEIC 語彙リストの作成の説明にあたる本セクション 3 はセクション 1 とセクション 2 の後に記述されているが、資料を収集し、出現頻度とレンジを組み合わせた伝統的な方法によって TOEIC 語彙リストを作成することは、本プロジェクトの「基礎」を成すため、実際のプロジェクト運営にあたっては、セクション 1 とセクション 2 より先に行なわれた。本プロポーザルの修正と承認の遅れにより、セクション 3 は、プロジェクトの正式な開始を待たずに開始された。これは、本プロジェクトの対象であるすべての作業を完了するには 1 年以上の期間を要することが予想されたためである。従い、セクション 3 は財政的支援なしで実施された。TOEIC 語彙リストの作成方法は、中條・牛田・山崎・Genung・内堀・西垣（2004）に詳述されている。

羅している。これらのデータによって、TOEIC Bridge に使用されている語彙はほぼ JSH 教科書の範囲内にあり、高校英語を習得した学習者にとっては望ましい英語習熟度の目標レベルにあることが確認された。

2. 内容理解に関して、TOEIC テストに必要な語彙は、TOEIC Bridge、英検 2 級、一般的な洋画よりは多いが、テレビのニュース番組や定期刊行物よりは少ない。BNC HFWL を利用した計算によると、TOEIC テストと TOEIC Bridge が必要とする語彙の差は 1,294 語である。この結果は、語彙水準の点では、TOEIC Bridge を極めた学習者が適切かつ着実な語彙学習を行えば、TOEIC テストのレベルに到達できることを示唆している。また、JSH 教科書の語彙は、TOEIC テストの 88.7%しか網羅しておらず、テストに出題される文章を十分に理解するには不十分である。
3. JSH 教科書の語彙の不十分さを補完するために、640 語で構成する単語リストを作成し、その有効性を確認した。
4. 統計的指標の適用は、学習者の語彙水準に対応した習熟度別の TOEIC 語彙リストを抽出するうえで有効と考えられる。

今後の研究の課題は、JSH 教科書データの改訂、統計的指標を活用した TOEIC 特徴語の習熟度レベル別の定義、専門語彙を特定するさらに有効な統計的指標の発見などである。

## 参考文献

- Bensoussan, M. and Laufer B. (1984) Lexical Guessing in Context in EFL Reading Comprehension. *Journal of Research in Reading*, 7, 1: 15-32.
- 中條清美, 竹蓋幸生 (1994) 「現代英語のキーワード『プラスα2000』一定義と効果の検証一」, 『千葉大学教育実践研究』, 1, 253-267.
- 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』18: 99-108.
- 中條清美, 牛田貴啓, 山崎淳史, マイケル・ジナング, 内堀朝子, 西垣知佳子, (2004) 「ビジュアルベーシックによる TOEIC 用語彙力養成ソフトウェアの試作Ⅲ」, 『日本大学生産工学部研究報告』, 37, 29-43.
- Chujo, K. (2004) “Measuring Vocabulary Levels of English Textbooks and Tests Using a BNC Lemmatised High Frequency Word List.” In J. Nakamura, N. Inoue, & T. Tabata (eds.) *English Corpora under Japanese Eyes*. Amsterdam: Rodopi, pp.231-249.
- Chujo, K. and Genung, M. (2004) Comparing the Three Specialized Vocabularies Used in ‘Business English,’ TOEIC, and British National Corpus Spoken Business Communications. Japan Association for Practical English, *Practical English Studies*, 11: 49-63.
- Church, K. W. and Hanks, P. (1989) Word Association Norms, Mutual Information, and Lexicography. *Proceedings of ACL-89*: 76-83.

- Dunning, T. E. (1993) Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19, 1: 61-74.
- 羽鳥博愛他. (1979) 『英語指導法ハンドブック 4<評価編>』 東京: 大修館書店.
- 林洋和. (2002) 『英語の語彙指導』 広島: 溪水社.
- Hisamitsu, T. and Niwa, Y. (2001) Topic-Word Selection Based on Combinatorial Probability. In NLPRS-2001: 289-296.
- Hu, M. and Nation P. (2000) Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, 13, 1.  
([http://www.vuw.ac.nz/lals/staff/paul\\_nation/marcella.rtf](http://www.vuw.ac.nz/lals/staff/paul_nation/marcella.rtf)).
- Laufer, B. (1989) What Percentage of Text Lexis Is Essential for Comprehension? in C. Lauren and M. Nordman (eds.), *Special Language: from Humans Thinking to Thinking Machines*. Clevedon: Multilingual Matters. 316–323.
- Laufer, B. (1992) How Much Lexis Is Necessary for Reading Comprehension? in L. Arnaud and H. Bejoint (eds.), *Vocabulary and Applied Linguistics*. London: Macmillan. 126–132.
- Leech, G., Rayson P., and Wilson A. (2001) *Word Frequencies in Written and Spoken English*. Harlow: Pearson Education Limited.
- Nation, P. (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Manning, C. D. and Schütze, H. (1999) *Foundations of Statistical Natural Language Processing*. The MIT Press.
- Rayner J.C.W. and Best. D.J. (2001). *A Contingency Table Approach to Nonparametric Testing*. Boca Raton: Chapman & Hall/CRC.
- Read, J. (2000) *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Schmitt N. and McCarthy M. (1997) *Vocabulary, Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- T. F. Communications. (1997) Practice TOEIC, *TOEIC Friends*, 3, 4: 24-53.
- 投野由紀夫. (1997) 『英語語彙習得論』 東京: 河源社.
- Wakaki, M. and Hagita, N. (1996) Recognition of Degraded Machine-Printed Characters Using a Complementary Similarity Measure and Error-Correction Learning. *IEICE TRANS. INF. & SYST.* E79-D, 5.
- West, M. (1926) *Learning to Read a Foreign Language*. London: Longman, Green & Co.

## Appendix TOEIC 語彙リスト(640 words)

便宜上、640 語を 200 語、200 語、240 語に分けて、3 つの表を作成した。

### TOEIC 語彙リスト(最初の 200 Words)

accept	case	department	fare	law	prefer	review
according	cash	deposit	fax	license	prepare	route
activity	cause	detail	feature	list	president	sale
additional	charge	development	fee	local	price	seat
advertisement	check	director	field	maintenance	profit	secure
affect	claim	division	figure	major	project	service
agreement	clerk	earn	file	manage	promotion	shipment
airport	clothing	economy	finance	manufacturer	property	staff
alarm	code	editor	financial	marketing	protect	stock
amount	committee	electrical	firm	meal	protection	submit
analysis	company	emergency	force	medical	publisher	suggest
analyze	complaint	employee	forecast	meeting	purchase	suit
announcement	complete	employment	form	membership	quarterly	supply
apartment	complex	enclose	forward	million	range	theater
application	concern	encourage	function	nearby	rate	trade
appointment	consider	engineer	furniture	obtain	receipt	traffic
appropriate	construction	environment	guarantee	occur	receive	transfer
approximately	consumer	estate	handle	officer	reception	transportation
area	contain	estimate	household	official	recommend	update
attend	convenience	examine	improvement	operate	regional	various
automobile	corporate	excellent	incorporate	operator	regular	
availability	corporation	expand	indicate	organize	relation	
available	cost	expansion	individual	original	remain	
bank	crowd	expect	inflation	outstanding	remove	
benefit	daily	expense	information	passenger	repair	
bid	deal	expensive	installation	payment	representative	
bill	decision	experience	intend	permit	require	
brand	degree	express	investment	personal	reserve	
budget	delivery	extension	item	power	responsibility	
candidate	demand	factor	labor	predict	responsible	

TOEIC 語彙(次の 200 Words)

access	cancel	discuss	industry	pack	reimbursement	supervise
accommodate	capital	display	inspection	package	rent	supervisor
account	career	distribute	install	payroll	rental	supplier
accountant	cater	document	insurance	personnel	replace	suspend
adjustment	chain	downtown	inventory	PIN	replacement	tax
advance	check-out	due	invest	policy	request	technical
agency	chief	earnings	investor	postpone	requirement	technician
agent	client	economic	invoice	present	research	temporary
ahead	colleague	efficiency	lately	previous	reservation	terminal
annual	commercial	electricity	lawyer	prior	resident	transit
appliance	competitive	employer	loan	procedure	resume	unit
applicant	conference	envelope	location	product	retailer	upcoming
apply	confirm	equipment	luggage	professor	retire	urgent
approve	consultant	executive	machinery	promote	revise	valid
architect	contract	exhibit	mail-order	promptly	safety	vehicle
arrange	convention	export	management	proposal	salary	vice
arrangement	copy	extend	manager	propose	savings	visa
arrival	coverage	facility	manufacture	provide	secretarial	warranty
assembly	credit	former	measure	purchaser	secretary	weekly
assistant	currency	franchise	merchandise	quality	security	workshop
associate	current	frequently	minimum	quarter	seminar	
assure	customer	full-time	monthly	rapidly	session	
audit	decline	fund	negotiation	recently	single	
auditor	decrease	furnish	newsstand	receptionist	site	
award	defective	gas	offer	refund	skilled	
baggage	delay	guide	operation	regarding	sponsor	
beverage	demonstrate	headquarters	opportunity	region	statement	
boss	departure	identification	order	register	strategy	
brochure	detergent	immediately	organization	registration	strike	
buyer	discount	import	overtime	regulation	subscription	

TOEIC 語彙(最後の 240 Words)

acceptance	chairperson	efficient	landfill	permanent	signature
accommodation	closure	eligible	lease	pharmacy	sincerely
accomplishment	comfort	eliminate	leather	photocopy	specialist
accordance	commission	ensure	lecture	physician	specifications
accumulation	commitment	enterprise	legal	preliminary	specified
accurate	commuter	entry	load	previously	spokesperson
achievement	compensation	establishment	locate	priority	spouse
acquisition	competitor	evaluate	long-term	privilege	stack
administration	compile	evaluation	lounge	prompt	steadily
admit	complicated	exceed	majority	qualified	storage
advise	complimentary	exclusive	malfunction	rearrange	subscribe
agenda	comply	fabric	maturity	reasonable	subscriber
allowance	compound	facilitate	mechanical	recession	summary
alter	comprehensive	fasten	medication	recommendation	surcharge
alternative	conduct	fiscal	memorandum	recruitment	surgical
analyst	congratulations	fitness	merger	reduction	survey
anticipate	considerable	folder	motor	reference	tablet
appendix	consideration	formal	multinational	reimburse	taxable
appoint	consult	format	municipal	reject	telecommunications
architectural	contractor	frequent	nationwide	relatively	tenant
assemble	convert	garage	negotiate	relevant	total
assign	coordinator	garment	notify	relocate	township
assignment	copier	grocery	occupant	removal	transaction
assistance	coupon	hike	occupational	renewal	traveler
attendance	deadline	identify	option	reorganization	trend
attendant	declare	impact	orientation	requisition	union
audiovisual	defect	implement	originate	reschedule	upgrade
authorize	demonstration	inadequate	outlet	residence	up-to-date
auto	depart	incentive	output	residential	urban
bankruptcy	description	inconvenience	overall	resign	user
bookkeeping	designated	incur	overdue	restriction	vacate
breakdown	dispute	inexpensive	overnight	retail	venture
cabinet	distribution	injury	oversee	retain	vessel
cancellation	diversify	innovative	pamphlet	retirement	via
capacity	dock	inquiry	participate	revenue	violation
cargo	documentation	inspector	patent	round-trip	voucher
carrier	domestic	institution	payable	rush	wage
catalog	double	insure	paycheck	section	warehouse
certificate	draft	interoffice	penalty	self-addressed	withdraw
chairman	duty	itinerary	periodic	shift	workstation



**TOEIC<sup>®</sup>**