

---

# 科学技術のための英語のコーパスの分析

前 田 秀 夫

---

## は じ め に

筆者は、1994年から1995年にわたり、English for Science and Technology (EST) (科学技術のための英語)<sup>1)</sup>の語彙研究のために、*McGraw-Hill Encyclopedia of Science and Technology*<sup>2)</sup>をもとにデータベースを構築した。当時は、個人レベルでデータを解析するためのコンピュータのハードウェアやソフトウェアが整っていなかったために、作成したデータベースを十分に活用できなかった。その後、*McGraw-Hill Encyclopedia of Science and Technology*<sup>3)</sup>の第7版を入手する機会を得て、再度、データベースを構築した。

本稿は、このデータベース（以降、English for Science and Technologyの頭文字から EST コーパス<sup>4)</sup>と呼称する）をもとに、コーパス構築の検証と活用を論じて行く。

## 1. EST コーパスの概要

本コーパスのもとになっている *Encyclopedia of Science and Technology*<sup>3)</sup>は、科学・技術のほとんどすべての分野を網羅する20巻、総数13,450ページからなる辞典である。エントリー数は、7,500項目で、それぞれを、音響学

(Acoustics), 航行学(術) (Aeronautical engineering)<sup>5)</sup>, 農業(Agriculture)<sup>6)</sup> というように79分野に分類している。

本コーパスを構築するにあたっては、79分野から、特に、科学（物理学，化学など），工業，技術に関する分野に焦点を絞った。例えば，音響学 (Acoustics), 航行学(術) (Aeronautical engineering), 分析化学 (Analytical chemistry) などは入力対象分野であるが，農業 (Agriculture), 生化学 (Biochemistry), 食品工業 (Food engineering) などは除外した。結果として，40分野を選び，エントリーごとに，ほぼ完全な形でコンピュータに入力した。

入力は，光学式文字読み取り (Optical Character Recognition, OCR) ソフトウェアを用いて，手作業で進めた。入力作業後は，OCR の認識率による誤差の修正に合わせて，統計上の処理を考慮して，数式，化学式，単位，数字等は削除する作業を行った。結果として，総語数 (token)<sup>7)</sup> は，877,280語となった。次に，統計上の基本のデータとして，表1のようなアルファベット順に並べた異語数 (type)<sup>7)</sup> と頻度数 (frequency) の表を作成した。総異語数は，26,634語となった。

表1 異語数と頻度数の構築例

異語数	頻度数
a	22959
a. m.	7
abacus	15
abaft	1
abandon	1

## 2. EST コーパスの検証

EST コーパスは，科学・技術分野の英語を研究する目的のために構築し

た、言わば、特殊な目的のためのコーパス (special purpose corpus<sup>8)</sup>) である。一方、Brown Corpus<sup>9)</sup>や LOB Corpus は、汎用コーパス (general purpose corpus<sup>8)</sup>) と呼ばれている。この 2 つのコーパスの関係は、English for Specific Purposes (ESP, 特殊な目的のための英語<sup>10)</sup>) と English for General Purposes (EGP, 一般的な目的のための英語<sup>10)</sup>) の関係と類似している。

以上のことから本章では、語彙を頻度数を中心に EST コーパスと Brown Corpus を比較しながら、EST コーパスの妥当性と特徴を考察することにする。

## 2-1. 頻度数による上位50位までの語による検証

表 2 は、表 1 を頻度数に並べた上位50位までの語と、その語に対する Brown Corpus の頻度数の順位をまとめたものである。上位20位までは、冠詞や前置詞という機能語が占めている。この傾向は従来のほとんどのコーパスに見られる。例えば、上位 3 語 (the, of, and) の順位は、American Heritage Corpus, Birmingham Corpus, LOB Corpus, Wellington Corpus と同じである。<sup>11)</sup>

表の中で、EST コーパスを特徴づける語は、上記の 4 つのコーパスの上位語にはない used, system, energy, fig., surface, temperature である。逆に、汎用コーパスでは、上位に属する人称代名詞 (I, you, he, she 等) や be 動詞 (was, were) などは、EST コーパスの上位語の中にはない。

## 2-2. 頻度数による総語数と異語数の分布による検証

表 3 は、頻度を基準とした EST コーパスの総語数と異語数の分布を示す。また、図 1 は、表 3 を視覚的に表したものである。図で見ると、総語数の分布は、ほぼ等間隔に増加していることが分かる。筆者の計算では、Brown Corpus も、それぞれ、1000以上(47.2%)、500—999(52.8%)、300—499(58.0%)、100—299 (70.0%) とほぼ同じような結果になる。また、表 2 の上位50

表2 頻度数による上位50位までの順位比較

語	EST	Brown	語	EST	Brown
the	1	1	have	26	28
of	2	2	such	27	77
and	3	3	these	28	68
a	4	5	has	29	44
in	5	6	other	30	63
is	6	8	system	31	204
to	7	4	energy	32	1042
are	8	24	one	33	32
for	9	11	not	34	23
by	10	19	than	35	59
or	11	27	also	36	87
as	12	14	when	37	45
be	13	17	been	38	43
with	14	13	more	39	48
that	15	7	its	40	56
which	16	31	fig.	41	1512
from	17	26	two	42	69
an	18	29	surface	43	3622
at	19	18	but	44	25
on	20	16	between	45	124
this	21	21	into	46	58
it	22	12	most	47	84
can	23	61	temperature	48	739
used	24	148	they	49	30
may	25	70	through	50	92

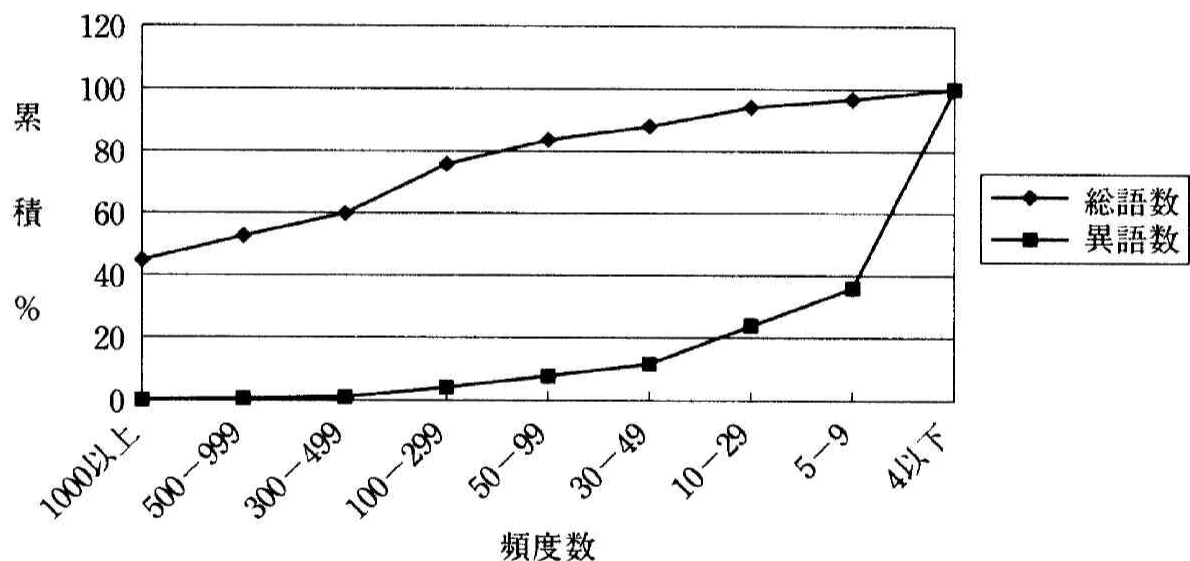
語の総語数の占める割合は、EST コーパスでは41.7%であり、Brown Corpus の上位50語が占める割合は41.2%でほぼ同じ割合である。これらの事は、EST コーパスが、総語数の分布では、バランスのとれたものであることを表す。

異語数の分布で見ると、EST コーパスの語彙の特徴が分かる。曲線の傾きが、49以下および、9以下で大きくなる。特に後者の場合は急激に増加して

表3 頻度数による総語数と異語数の分布

頻度数	総語数		異語数	
	累積語数	累積%	累積語数	累積%
1000以上	394152	44.9	75	0.3
500—999	462054	52.7	177	0.7
300—499	525508	59.9	343	1.3
100—299	664947	75.8	1169	4.4
50—99	734399	83.7	2157	8.1
30—49	773024	88.1	3165	11.9
10—29	827339	94.3	6395	24.0
5—9	848850	96.8	9602	36.1
4以下	877280	100.0	26634	100.0

図1 頻度数による総語数と異語数の分布



いる。頻度数が少なくなるに従い、異語数が急激に増加するということである。Brown Corpus や LOB Corpus のような汎用コーパスの異語数に比べて、EST コーパスの異語数が極めて多い結果である。しかも、本稿では詳しく論じないが、筆者の調査では、下位の頻度数の語はほとんどが専門用語 (technical word) である。

### 2-3. $\chi^2$ 検定による検証

EST コーパスの頻度数が、有意な数値であるかを調べるために、表4のように<sup>12)</sup> $\chi^2$ 値を求めてみる。

表4で得たすべての $\chi^2$ 値は、自由度1の有意度水準0.001の境界値10.83より大きいので、表4の語は統計的に0.1%レベルで有意な数値になる。

本節では、EST コーパスの頻度数50以上の語の $\chi^2$ 値を求め、コーパス全体の妥当性を検討する。

頻度数50以上の語は、表3によると、総語数で734,399語(83.7%)、異語数で2,157語(8.1%)あるので、これらの語の $\chi^2$ 値が有意な数値であれば、2-1, 2-2で論じたことから考慮して、全体として、EST コーパスの妥当性を証明できるであろう。

$\chi^2$ 検定の有意度水準は、語彙統計でよく使われる、0.001, 0.01, 0.05, 0.10の4つを使う。自由度は1なので、それぞれの境界値は、10.83, 6.63, 3.84, 2.71になる。

計算結果を有意度水準別に、語数の分布を示したものが表5である。表によれば、総調査語2,157語中、1,337語(62.0%)の語の頻度数が有意度水準0.001で有意な数値として認められたことになる。さらに、1,639語(71.3%)が統計上の有意度水準を上回っている。そして、わずか160語(7.4%)(表中では、その他)が有意度水準以下ということになる。

‘検定不可’とした458語(21.2%)は、Brown Corpusで頻度数が9以下の

表4 異語数の $\chi^2$ 値(例)

語	実測値		期待値		$\chi^2$ 値
	EST	Brown	EST	Brown	
the	72166	69970	66422.0	75714.0	932.49
of	40038	36410	35725.1	40722.9	977.45
and	24562	28852	24961.1	28452.9	11.98
a	22959	23129	21537.5	24550.5	176.13
in	21456	21337	19997.7	22795.3	199.64

表5  $\chi^2$ 検定による語数分布

有意度水準	語数	%表示
0.001	1337	62.0
0.01	83	3.9
0.05	71	3.3
0.10	48	2.2
その他	160	7.4
検定不可	458	21.2
計	2157	100.0

語である。これは、統計上の処置であるが、注意したいのは‘検定不可’の語が必ずしも意味のない語ではないということである。逆に、EST コーパスの場合 EST の特徴を持った語を示すこともある。汎用コーパスである Brown Corpus では、頻度数が少ないという逆の理由が成り立つからである。参考までに、‘検定不可’の語の中から、Brown Corpus で、頻度数 9 および 8 の語を下に挙げる。特に、イタリック体の語は EST コーパスで、100以上の頻度数がある。さらに、他の語、例えば, carrier, indicator, ions など EST の特徴を持った語である。

以上のことは、 $\chi^2$ 検定においても EST コーパスが高い信頼性を示していることになる。

analogous, analytical, barrier, boiling, cameras, carrier, compressed, compression, concentrations, constants, crystals, ensure, geometry, horizontal, impurities, indicator, ions, modes, parameters, rays, recovered

### 3. EST コーパスの活用例

本章では、EST コーパスの頻度数を中心にいくつかの活用例を考察する。

### 3-1. 差異係数 (difference coefficient) による EST の語分類の活用例

Hutchinson (1987)<sup>13)</sup> は、ESP の語彙を、structural (e.g. are, this, only, however), general (e.g. table, run, dog, road, weather, cause), sub-technical (e.g. engine, spring, valve, acid, budget), technical (auricle, schistosome, fissure, electrophoresis) の 4 つに分類している。また、Trimble (1985)<sup>14)</sup> は、EST の語彙を、①technical vocabulary, ②sub-technical vocabulary, ③noun compounds の 3 つに分けている。

EST コーパスの語彙も同じように分類できるわけだが、語義的な問題があり、統計上では処理が難しい。本節では、1つの試みとして、差異係数を使って考察する。

差異係数は、語の頻度数を (EST コーパス - Brown Corpus) ÷ (EST コーパス + Brown Corpus) の式にあてはめて求める。その値 (DC 値) は、+1.00 と -1.00 の間にある。正の値の場合は、EST コーパスの方にその語が多く出現し、負の値の場合は、その逆になる。例として、Hutchinson (1987)<sup>13)</sup> が分類した語で検討する。結果を表 6 に示す。語は EST コーパスの頻度数の多い順から並べてある。実測値は、両コーパスの実際の頻度数を示す。

DC 値を求めるために、EST コーパスでは、100万語相当語を計算した (4 列目)。最後列の  $\chi^2$  値は本表では参考とする。

structural な語は、当然のように両コーパスで頻度数は多い。are を除けば DC 値にあまり差はない。しかし、are は structural な語として異論はないので問題とはならないだろう。一方、general あるいは sub-technical な語には、DC 値にばらつきが多く特定しにくい。run, road, dog は負の値が大きく、engine, acid では、逆の結果が出ている。technical な語は fissures (頻度数: 4) 以外、両コーパスになかった。

以上のことから分かることは、DC 値は語彙の領域を限定することはできないが、特徴のある語彙をある程度提起できる。参考までに、表 5 の調査語 2,157 語で、DC 値の分布を見ると、正の値の語は 1,794 語 (83.2%)、負の値



表 6 差異係数による EST の語分類

語	EST			Brown	DC 値	$\chi^2$ 値
	順位	実測値	100万語当り	実測値		
are	8	10217	11646	4394	0.45	3157.98
this	21	4216	4806	5146	-0.03	10.85
only	64	1123	1280	1747	-0.15	66.64
however	92	833	950	552	0.26	100.13
engine	239	399	455	50	0.80	320.27
acid	252	380	433	13	0.94	393.87
table	420	256	292	198	0.19	16.97
cause	495	226	258	130	0.33	40.08
weather	1151	101	115	69	0.25	11.03
run	1814	61	70	212	-0.50	65.27
spring	2108	51	58	127	-0.37	23.40
valve	3304	28	32	0	1.00	
road	3909	21	24	197	-0.78	
budget	7417	7	8	59	-0.76	
dog	12867	2	2	75	-0.95	

の語は363語 (17.8%) であった。この差は、前章で述べたように、両コーパスの異語数によるものである。表7では、調査語の中から、アルファベットのaで始まる単語をDC値順に並べたみた。

表から類推できることは、DC値が1.00—0.80間で、Hutchinson (1987)<sup>13)</sup>の分類の technical な語が多く、順に sub-technical, structural あるいは general (0.20以下) というように語が分散しているようである。いずれにせよ、DC値では、ESTの語彙をある程度示唆できるわけだが、限定することは不可能なことが分かる。

例えば、1.00—0.80間にある armature, acid, acceleration 等は、technical な語であるし、0.20以下にある along, and, able 等は、structural あるいは general な語である。しかし、0.80—0.20間の語の分類には問題が起きてくる。atmosphere, apparatus, access 等は専門用語でもあるし、一般的に使わ

表7 DC 値の分布例

語	DC 値	語	DC 値	語	DC 値
aberration	1.00	analytical	0.76	area	0.29
aberrations	1.00	atom	0.75	appears	0.28
ac	1.00	approximate	0.75	adjusted	0.27
accelerator	1.00	allows	0.73	available	0.24
acids	1.00	associated	0.73	additional	0.23
acoustic	1.00	ash	0.72	absence	0.22
acoustical	1.00	aircraft	0.71	although	0.22
alkali	1.00	artificial	0.70	appearance	0.21
alloys	1.00	applied	0.69	an	0.17
alpha	1.00	absolute	0.68	avoid	0.16
alternating	1.00	application	0.68	approach	0.14
altitude	1.00	attached	0.66	areas	0.14
ambient	1.00	advantages	0.66	automobile	0.13
ammonia	1.00	accelerated	0.66	added	0.13
ammonium	1.00	atmosphere	0.65	along	0.13
amplifier	1.00	amounts	0.64	acting	0.13
amplifiers	1.00	air	0.63	appear	0.12
amplitude	1.00	approximately	0.61	according	0.11
analog	1.00	accurate	0.59	actual	0.10
anode	1.00	accuracy	0.56	as	0.09
antenna	1.00	approaches	0.54	assumed	0.09
aperture	1.00	agent	0.54	activity	0.07
appreciable	1.00	addition	0.53	a	0.06
aqueous	1.00	accomplished	0.52	adequate	0.03
armature	1.00	automatic	0.49	apart	0.02
astronomical	1.00	analysis	0.49	apply	0.01
atmospheric	1.00	agents	0.49	and	-0.02
attenuation	1.00	arrangement	0.49	at	-0.02
audio	1.00	astronomy	0.49	allowed	-0.05
auxiliary	1.00	apparatus	0.47	action	-0.05
aviation	1.00	affect	0.46	across	-0.09
axial	1.00	appropriate	0.45	another	-0.12
axle	1.00	are	0.45	about	-0.14
alloy	0.97	acts	0.45	account	-0.17
acid	0.94	access	0.44	annual	-0.20
absorption	0.94	achieved	0.43	add	-0.21
applications	0.92	amount	0.38	any	-0.25
angular	0.88	above	0.36	actually	-0.26
angles	0.87	active	0.36	almost	-0.29
atoms	0.85	applying	0.35	among	-0.31
aluminum	0.85	assembly	0.35	around	-0.32
acceleration	0.84	ability	0.35	aid	-0.33
arch	0.83	advantage	0.34	all	-0.36
absorbed	0.81	advanced	0.34	after	-0.40
atomic	0.80	automatically	0.33	always	-0.43
adjacent	0.79	achieve	0.33	alone	-0.44
angle	0.78	allow	0.33	act	-0.48
arc	0.77	average	0.32	against	-0.48
array	0.77	apparent	0.31	able	-0.50
analogous	0.76	also	0.30	away	-0.52
axis	0.76	arranged	0.30	again	-0.69

れる語でもある。

また, are (0.45), also (0.30) 等は structural あるいは general な語であるが, air (0.63), advantage (0.34) 等は分類分けが難しい。

さらに, 本稿では, 異語数で処理をしているため, alloys (1.00), alloy (0.97) では問題にはならないが, allows (0.73), allow (0.33) は, 別の処理が必要になってくる。

### 3-2. 関連語と異語による語の活用例

表8は, 動詞 change とそれに関連する動詞の屈折形<sup>15)</sup>を含めてまとめたものである。表のようにすると, 各動詞の特徴がある程度分かりやすくなる。

表8 change とその関連語の基礎データ例

語	EST 実測値	Brown 実測値	DC 値	語	EST 実測値	Brown 実測値	DC 値
adapted	25	13	0.37	modified	59	13	0.68
adapting	1	13	-0.86	modifies	4	0	1.00
adapts	2	13	-0.73	modify	22	0	1.00
adjust	23	13	0.33	modifying	10	0	1.00
adjusted	50	33	0.27	transform	40	0	1.00
adjusting	25	11	0.44	transformed	14	25	-0.22
adjusts	12	0	1.00	transforming	6	0	1.00
alter	19	15	0.19	transforms	9	0	1.00
altered	28	22	0.19	transmuted	1	0	1.00
altering	6	0	1.00	varied	52	42	0.17
change	464	240	0.38	varies	140	11	0.87
changed	57	95	-0.19	vary	139	34	0.65
changes	406	131	0.56	varying	88	42	0.41
changing	111	44	0.49				
convert	56	82	-0.12				
converted	107	20	0.72				
converting	40	0	1.00				
converts	39	0	1.00				

EST の実測値で見ると、change, vary は change, changes, vary, varies の形の使用例が多いが、convert, modify は、converted, modified の形の方が多い。DC 値で見ると、adjust や alter よりも、convert, modify, vary といった語の方が正の値が大きく、EST の語彙の特徴を示しているようである。

これらは、あくまでも統計上の見方であることはしばしば指摘したところであるが、語彙を選定する際の強力な基礎データにはなる。そして、コーパスの活用・研究には大変重要な役割を果たす。

### 3-3. 関連語等からの活用例

EST の語彙を絞り込む過程で、その関連語の頻度数などをまとめたもの（表9）や連語を調べるデータベース（表10）の作成もコーパス活用の重要な要素である。

例えば、EST コーパスと Brown Corpus の中から、adjust に関連するすべての語とその頻度数を表したのが表9である。この表からまず注目したい語は、EST コーパスにだけある adjustability, adjustable である。表8とは違った視点から、調査対象の語を特定できる。

表9 adjust とその関連語例

語	実測値	
	EST	Brown
adjust	23	16
adjustability	2	0
adjustable	26	0
adjustable-blade	1	0
adjusted	50	33
adjusting	25	11
adjustment	33	35
adjustments	15	20
adjusts	12	2

表10 adjustable と adjustment の連語例

onal to that pressure. An adjustable scale, with a vernier scale, allows a readi  
 he transformer to provide adjustable leakage reactance between windings (Fig. 5)  
 s field tapped to give an adjustable welding current with drooping volt-ampere c  
 its the output voltage is adjustable, and the welders are designed for either fl  
 signed for either flat or adjustable droop volt-ampere characteristics. In addit  
 his may be done by manual adjustment ; however, an automatic feed mechanism is mo  
 hich is provided to allow adjustment of the intereye piece distance to match the  
 other schemes of current adjustment are used, most of which employ the principl  
 ed in two forms. The null adjustment method requires adjustment of a calibrated

\* 表中の下線は筆者が加えた。

表10は、主に連語関係を調べるために、表9の中から adjustable, adjustment を選び、その一部を紹介するものである。調査対象の語が句、節、文の中でどのような働きをしているかを、連語中心に検討するには、このようなデータが必要になってくる。

## お わ り に

EST コーパスの構築、検証、活用という3つの柱で論考を進めてきたが、活用面での論議が十分ではなかった。今後の課題としたい。

コーパスの構築のためのデータ収集、コンピュータへの入力作業等には、時間と労力が非常にかかる。個人レベルでは、大容量のコーパスの構築は、

いかに最新のハードウェア、ソフトウェアを駆使しても限界がある。これからの手法としては、CD-ROM やインターネット上からのデータ収集が主流となろう。問題となるのは、いかに適切なデータを収集できるかということになる。

コーパスの検証については、EST コーパスのような特殊な目的のためのコーパスでは、比較できるコーパスが少なく、現在のところ既存の汎用コーパスを利用する方法しかない。このことも今後の課題となろう。

コーパスの活用については、当然のことながら、コーパス研究の中核となるものである。英語の語彙、文法、文体等の研究のためのコーパスの利用、英語教育の中でのコーパスの応用等、筆者にとっては、まだまだ未知の分野であるが、研究を続けて行きたいと考えている。関係諸氏のご指導、ご叱正を賜れば幸いである。

---

#### 注

- 1) *English for Specific Purposes : A learning-centred approach* (Hutchinson, 1987, p. 17) の 'The tree of ELT' によれば、EST は English for Specific Purposes (ESP) から枝分れした 1 つの型 (type) になる。
- 2) Parker, S. P. (ed.) (1986) *McGraw-Hill Encyclopedia of Science and Technology*, 5th ed., New York : McGraw-Hill Company.
- 3) Parker, S. P. (ed.) (1992) *McGraw-Hill Encyclopedia of Science and Technology*, 7th ed., 20 Vols., New York : McGraw-Hill, Inc.

本辞典は、第 7 版を発刊後、1996 年、1997 年にそれぞれ年鑑 (yearbook) を発刊している。さらに、関連する辞典も下記のように多数、発刊している。

- Parker, S. P. (ed.) (1993) *McGraw-Hill Dictionary of Scientific and Technical Terms*, 5th ed., New York : McGraw-Hill, Inc.
- Parker, S. P. (ed.) (1994) *McGraw-Hill Concise Encyclopedia of Science and Technology*, 3rd ed., New York : McGraw-Hill, Inc.
- マグロウヒル科学技術用語大辞典編集委員会 (1997) 『マグロウヒル 科学

技術用語大辞典 第3版』, 日刊工業新聞社。

・ Parker, S. P. (ed.) (1997) *McGraw-Hill Dictionary of Engineering*, New York : McGraw-Hill, Inc.

- 4) コーパス (corpus) の定義は, 齋藤俊雄・中村純作・赤野一郎編 (1998) 『英語コーパス言語学—基礎と応用』 研究社, p. 18 から, 以下を引用する。

ある特定の言語, 方言もしくはその他のヴァリエティを代表し, かつ言語研究に使用されることが想定され, コンピュータに蓄積, 処理された話し言葉, 書き言葉のテキストの集合体

- 5) 参考までに, 航行学 (術) (Aeronautical engineering) の分野でエントリーされている項目を下記に示す。

Aeroelasticity, Aeronautical engineering, Aeronautics, Agricultural aircraft, Aileron, Air-cushion vehicle, Air transportation, Aircraft, Aircraft design, Aircraft instrumentation, Aircraft rudder, Aircraft testing, Airframe, Airplane, Airship, Airspeed indicator, Altimeter, Aspect ratio, Autogiro, Automatic horizon, Autopilot, Aviation, Ballistic missile, Ballistic range, Balloon, Blimp, Convertiplane, Dirigible, Drone, Elevator (aircraft), Elevon, Flight, Flight characteristics, Flight controls, Flight dynamics, Flight science, Flutter (aeronautics), Fuselage, General aviation, Glide-path indicator, Glider, Guided missile, Helicopter, Hypersonic flight, Inertial guidance system, Kite, Landing gear, Launch complex, Military aircraft, Missile, Nose cone, Parachute, Rate-of-climb indicator, Rocket-sled testing, Seaplane, Short takeoff and landing (STOL), Sonic boom, Stability augmentation, Stabilizer (aircraft), Stall-warning indicator, Strake, Subsonic flight, Supercritical wing, Supersonic flight, Tail assembly, Transonic flight, Turn and bank indicator, Ultralight aircraft, Vertical takeoff and landing (VTOL), Wing, Wing loading, Wing structure, Yaw indicator

- 6) 参考までに, 79分野を下記に示す。\*印が付いた40の分野が本コーパスを構築している分野である。

\*Acoustics, \*Aeronautical engineering, Agriculture, \*Analytical chemistry, Animal anatomy, Animal systematics, Anthropology and archeology, \*Astronomy, \*Atomic, molecular, and nuclear physics, Biochemis-



try, Biophysics, Cell biology, \*Chemical engineering, \*Civil engineering, \*Classical mechanics and heat, Comparative physiology, \*Computers, \*Control and information systems, \*Design engineering, Developmental biology, Ecology and conservation, \*Electrical power engineering, \*Electricity and electromagnetism, \*Electromagnetic radiation and optics, \*Electronic circuits, \*Fluid mechanics, Food engineering, Forestry, General physiology, Genetics and evolution, Geochemistry, Geology (mineralogy and petrology), Geology (physical, historical, and sedimentary), Geophysics, \*Graphic arts, Immunology, \*Industrial and production engineering, \*Inorganic chemistry, Invertebrate zoology, \*Low-temperature physics, Mathematics, \*Mechanical and power engineering, Medical microbiology, Medicine and pathology, \*Metallurgical engineering, \*Meteorology and climatology, Microbiology, \*Microscopy, \*Mining engineering, Mycology, \*Naval architecture and marine engineering, \*Navigation, \*Nuclear engineering, Oceanography, Organic chemistry, Paleobotany, Paleontology, \*Petroleum chemistry, \*Petroleum engineering, \*Physical chemistry, \*Physical electronics, Physical geography, Physiological and experimental psychology, Plant anatomy, Plant pathology, Plant physiology, Plant taxonomy, \*Propulsion, Psychiatry, \*Radio communications, \*Soils, \*Solid-state physics, \*Space technology, \*Telecommunications, \*Theoretical physics, \*Thermodynamics, Vertebrate zoology, Veterinary medicine, Virology

- 7) 総語数 (token), 異語数 (type) の定義は, Carroll, J. B., Davies, P. and Richenman, B. (1971) *The American Heritage Word Frequency Book*, USA: American Heritage Publishing. による。以下, 本書から引用する。

The distinction between types and tokens, conventionally used in discussions of word frequency, is frequently employed throughout this work. A type is a particular word, counted just once, regardless of how many times it occurs; a token is any of the individual occurrences of the type. For example, the type 'the' has 373, 123 tokens (occurs 373, 123 times) in the AHI Corpus.



例えば, 'To produce goods or services, a business needs to buy goods or services.' という文では, 総語数は13語で, 異語数は9語となる。EST コーパスでは, コーパス構築の際, 統計処理上の理由から, 化学式, 数式等を削除したが, コーパス全体の量から考えると, 削除語の影響は, きわめて少ないと思われるので, 総語数, 異語数という用語はそのまま区別しないで使用する。

8) 齋藤俊雄・中村純作・赤野一郎編『前掲書』p. 19。

9) 本コーパスは, アメリカ英語の資料約100万語からなるサンプルコーパスである。コンピュータを使って編纂した初めての電子コーパスでもある。完成年度が1964年で, 現在では古さを感じられるが, アメリカ英語の語の使用頻度の比較等では現在でもよく使われている。

本稿で使用した Brown Corpus の数値は, Hofland, K. and Johansson, S. (1982) *Word Frequencies in British and American English*, Bergen: Norwegian Computing Centre for the Humanities, pp. 472-544 から得た。

10) 山崎真稔・高橋貞雄・佐藤久美子訳 (1988) 『ロングマン応用言語学用語辞典』南雲堂, p. 122。

11) Kennedy, G. (1998) *An Introduction to Corpus Linguistics*, London: Addison Wesley Longman Limited, p. 98.

12) 齋藤俊雄・中村純作・赤野一郎編『前掲書』pp. 79-83。

13) Hutchinson, T. and Waters, A. (1987) *English for Specific Purposes: A learning-centred approach*, rpt. Cambridge: Cambridge University Press, 1990, pp. 165-166.

Robinson, P. C., (1991) *ESP Today: A Practitioner's Guide*, London: Prentice-Hall, p. 28. では, 語彙を以下のように分類している。

- Specialist vocabulary is identified as the first of three levels of vocabulary differentiated by researchers in ESP (specifically EAP) ....
- The second level of vocabulary that has been identified is variously called semi-technical, sub-technical or general scientific/technical and comprise words which occur in a number of scientific or technical areas.
- The third level of vocabulary is general and non-academic, such as would be central to an EGP course.

14) Trimble, L. (1985) *English for Science and Technology: A discourse*

- approach*, 3rd ed., Cambridge : Cambridge University Press, 1990, p. 128.
- 15) Godman, A. and Payne, EMF (1979) *Longman Dictionary of Scientific Usage*, rpt. Hong Kong : Longman, 1981, p. 32.

#### 参考文献

- Barnbrook, G. (1996) *Language and Computers*, Edinburgh : Edinburgh University Press.
- Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus linguistics : Investigating language structure and use*, Cambridge : Cambridge University Press.
- Garside, R., Leech, G. and McEnery, T. (Eds.) (1997) *Corpus Annotation*, London : Longman.
- 橋本光憲 (1997) 「ビジネス用語の一考察——ビジネスとファイナンスの英語——」『国際経営論集』No. 11, 神奈川大学経営学部。
- 橋本光憲 (1997) 「General English の一考察——Business English, Special English との比較において——」『国際経営フォーラム』第 8 号, 神奈川大学経営学部。
- Jordan, R. R. (1997) *English for Academic Purposes : A guide and resource book for teachers*, UK : Cambridge University Press.
- Kennedy, G. (1998) *An Introduction to Corpus Linguistics*, London : Addison Wesley Longman Limited.
- Nation, I. S. P. (1990) *Teaching and Learning Vocabulary*, New York : Newbury House Publishers.
- Oakes, M. P. (1998) *Statistics for Corpus Linguistics*, Edinburgh : Edinburgh University Press.
- Ooi, V. B. Y. (1998) *Computer Corpus Lexicography*, Edinburgh : Edinburgh University Press.
- Schmitt, N. and McCarthy, M. (Eds.) (1997) *Vocabulary : Description, Acquisition and Pedagogy*, Cambridge : Cambridge University Press.
- Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (Eds.) (1997) *Teaching and Language Corpora*, England : Longman.