



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

石川, 慎一郎

(Citation)

神戸大学国際コミュニケーションセンター論集, 1:13-27

(Issue Date)

2004

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/00517984>

(URL)

<https://hdl.handle.net/20.500.14094/00517984>



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—

石川 慎一郎¹

1. はじめに：EGP から ESP へ

近年の大学英語教育は、社会のニーズの変化をふまえ、専門教育・職業教育により積極的に関与しようとしている。一般教育・共通教育の一環としての EGP(English for General Purposes)は、専門分野での学びを前提とした EAP(English for Academic Purposes)や、職業英語教育としての EOP(English for Occupational Purposes)、あるいはそれらを包含する ESP (English for Specific Purposes)へと質的に変容しつつあるのである。

ESP 教育は EAP と EOP の直結性が高い自然科学の分野ではすでに広く実践されている。特に、工学専攻学生のための「技術英語」(西原, 2004; 松浦, 2004)、医学専攻学生のための「医療英語」(吉留, 2004)、看護学専攻学生のための「看護英語」(川北・白坂, 2004; 山本・大倉, 2004)などについては国内の先行実践も豊富である。

一方で、人文科学系の学問領域については、大学教育と将来の職業の関連性が明白ではないという事情もあり、いまだ専門分野に即応した英語教育の実践の取り組みは多くない。しかしながら、教養・共通教育としての外国語教育と専門教育のシームレスな融合を考えた場合、人文科学領域についても、専門や職業を意識した英語教育のニーズが今後さらに強まってくるものと思われる。

ESP 教育のメリットについて、山本・大倉(2004)は以下の3点をあげている。

- ・ 学習者の興味関心に合わせた英語教材を提供できる。
- ・ 学習者は専門分野の知識も同時に充実させられ、無駄がなく効率的である。
- ・ 一般英語(General English)に比べ学習者の動機付けが高まる。(p.60)

大学での ESP の教育システムを考える場合、重視すべきは「学習者の興味関心に合わせた英語教材」を使用することである。この場合の教材は、学習者が将来、学術・職業の分野で遭遇する可能性の高いものであることが望ましい。

EGP に対する EAP と ESP の相互関係性を概観した研究の中で、Dominguez & Rokowski (2002)は次のように指摘している。

¹ 神戸大学国際コミュニケーションセンター ishikawashin@anet.ne.jp

...the most important argument regarding selective learning is that adults need to obtain information they can apply outside the academic world. The use of authentic material can make the connection between both worlds possible, the academic and the professional, in a practical way. (First section, para. 6)

こうした「実物教材(“authentic material”)」を用いて ESP 教育を行う場合、鍵となるのは、学習者が産出・理解すべきテキストの特徴性の吟味である(Dudley-Evans, 2001, p. 134)。とりわけ言語材料の基礎となる語彙の特徴性を正しく認識しておくことは不可欠であり、この意味において、一般的な EGP 語彙と区別される ESP 語彙のリストが必要となる。

語彙学習にはそれ自身を目標として行う意図的学習(intentional learning)と、他の一般的な英語活動を行いながら、それに付随して自然に語彙を学んでいく偶発的学習(incidental learning)の両面があるとされる(望月・相澤・投野, 2003, pp. 97-98)。しかし、偶発的学習で修得できる語数はきわめて限られており(Aizawa, 1999), ESP のような特殊性の高い語彙を学習する場合は、信頼できる語彙表をふまえた意図的学習が重要である。

本稿では、先行実践の少ない人文科学系 ESP 教育の方向性を模索すべく、とくに「司法英語」というジャンルを取り上げ、コーパス・データに基づいた“authenticity”の高い ESP 語彙表開発に取り組みたい。

2. コーパスに基づく語彙表開発

実際の言語のありようを反映した“authenticity”の高い語彙表を開発しようとする場合、言語コーパスの利用は不可欠である。Leech, Rayson, & Wilson (2001)は、語彙表開発におけるコーパスの利点として次のように概括している。

Since the early days of Thorndike and Lorge, a big transformation has taken place through the development of computers and modern computer technology. Nowadays, a very large of collection of texts (normally called a corpus) can be stored and searched on a computer, and the frequencies of words in that corpus can be determined and listed by a fairly trivial computer program. (p. x)

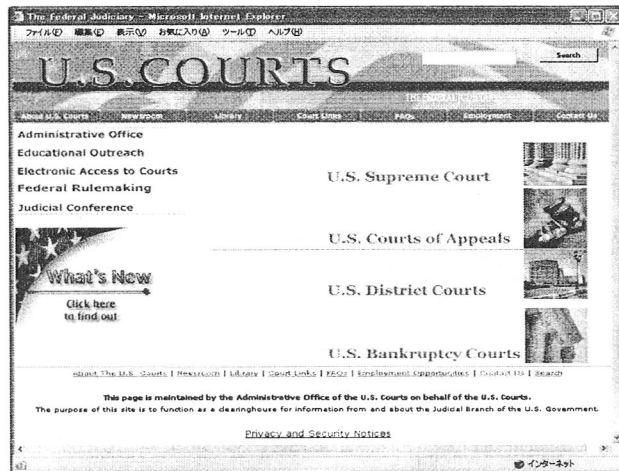
コーパスに基づく語彙表としては、古くは、米国語 100 万語を収集した Brown コーパスに基づく Kučera & Francis(1967)、英国語 100 万語を収集した LOB(Lancaster-Oslo/Bergen)コーパスに基づく Hofland & Johansson (1982)などがある。また近年では、1億語の英国語データベースである British National Corpus (BNC)に準拠した Leech, Rayson, & Wilson (2001)や、BNC と日本人学習者に特化したサブコーパスとの頻度調整によって選定された大学英語教育学会基本語改訂委員会(編)(2003)などがある。これらは主として EGP 向けの語彙表である。

ESP 語彙表としては、精神医学の診察で用いられる用語を集めた Dahl (1979)、コンピュータ技術用語を収集した James et al. (1994)などがあるが、データのサイズも限られており、データ処理の手法もかなり素朴なものである。本研究では、大学英語教育学会基本語改訂委員会(編)(2003)で試みられたデータ処理の手法を参照しつつ、司法英語についての信頼できる語彙表の開発を行う。

3. 米国司法コーパスの構築と参照コーパスの選定

一般に、英語語彙表を開発する場合、言語の主たる使用圏である英国と米国の言語データを一定のバランスで組み合わせる場合が多い。しかし、コンピュータ英語や医学英語などとは異なり、司法英語は当該国の司法制度と密接に関連している。司法とは社会的・文化的な制度であり、異なる司法制度を取る英国と米国の言語データを混入することは、得られた結果の信頼性を損なう恐れがある。そこで今回は、米国の司法英語データに限定して収集を行い、コーパスを構築することとした。

米国の司法英語に関する言語データの採取先としては、新聞の司法関係記事データベース、連邦最高裁判所のサイト、連邦政府のサイトなどが考えられるが、今回は、連邦裁判所(Federal Judiciary)が運営



する US Courts のサイト (<http://www.uscourts.gov/>)を選んだ。

これは、同サイトに米国司法制度改革の関連資料が数多く公開されており、最新の米国司法界の現状をふまえる上でとくに有益と判断したためである。

上記 US Courts のサイトより、"The Federal Court System in US", "US Probation Journal", "Code of Conduct for US Judges"など、pdf 形式で公開されている公式文書を幅広く収集した。pdfデータはすべてテキストファ

図 1. 米国連邦裁判所公式サイト

イルに変換し、変換エラーなどを手作業で修正して、本研究の基礎となる「米国司法文庫コーパス」(US Judiciary English Corpus, 以下 US-JEC)を構築した。US-JEC のサイズは、延べ語数(token)で約 238,000 語である。

以下、US-JEC の特徴的な高頻度語を統計的に抽出するわけであるが、作業の前提として、比較対象となる参照コーパス(reference corpus)を設定する必要がある。参照コーパスとして利用可能な一般英語のコーパスはBNCをはじめ各種存在しているが、ターゲットコーパスであるUS-JECが約24万語の米語データであることを考慮すれば、英国語データを含むものや、過度に大型のコーパスは参照対象として適切とは言えない。そこで本研究では、1990年代の米語100万語を各種のジャンルからバランスよく採集したFROWN (Freiburg-Brown)コーパスを参照コーパスとして利用することにしたい。

4. コーパスからの頻度データの抽出および統計処理

最初にターゲットコーパスであるUS-JECをコンコーダンサで語彙分析にかけ、粗の語彙頻度表を作成する。ついで参照コーパスである FROWN コーパスから同様の語彙頻度表を作成する。2つの語彙頻度を比較する際には比率的頻度(百分率や百万語当りの換算頻度など)を用いる場合が多いが、母データのサイズが異なっている場合は比率的頻度の信頼性は落ちることが知られている。たとえば同じ5%であっても24万語の5%と100万語の5%では統計的な重みが異なることになる。

そこで本研究では、最近のコーパス準拠語彙表のデータ処理の手法を踏襲し、対数尤度比

(log-likelihood score) を基準にして、ターゲットコーパスに有意に高頻度に出現している語を抽出することとした(大学英語教育学会基本語改訂委員会(編), 2003, pp.107-108; Leech, Rayson, & Wilson, 2001, pp.16-17; Stubbs, 2002, p.129). これにより, 異なる母データサイズを持つ 2 つのコーパスの頻度を正しく比較し, 特徴性を抽出することが可能になる.

なお, 対数尤度比に基づく「特徴度」の算出処理は, 最近の多くのコンコーダンスがサポートしているが, その呼び方は微妙に異なり, Scott (2004) では”keyness”, Anthony(2004a) では”keyword value”, Anthony(2004b) では”keyness value”とされている.

ターゲットコーパス(US-JEC)と参照コーパス(FROWN コーパス)を比較して特徴語リストを作成する場合, 特徴語の閾値をどこに設定するかは難しい問題であるが, 今回は特徴度(keyness)が 75 以上の上位 500 語に限ってリストとした. 下記はその一部である. TC Freq はターゲットコーパスにおける頻度, TC Ratio はその全体に占める構成比(%), RC Freq は参照コーパスにおける頻度, RC Ratio はその全体に占める構成比(%), P は有意水準である. なお RC Ratio は 0.01%に満たない場合は空欄となっている.

表 1. 特徴語基礎リスト(上位 20 語)

Key word	TC Freq.	TC Ratio	RC Freq.	RC Ratio	Keyness	P
COURT	2,037	0.85	216	0.01	6,681.64	<0.01
COURTS	1,695	0.71	34		6,364.34	<0.01
FEDERAL	1,505	0.63	221	0.02	4,682.99	<0.01
JUDICIAL	1,206	0.51	26		4,512.59	<0.01
JUDICIARY	1,005	0.42	14		3,819.68	<0.01
JUDGES	1,029	0.43	32		3,780.55	<0.01
THE	17,171	7.19	62,351	4.23	3,599.07	<0.01
FS	910	0.38	0	0.00	3,588.39	<0.01
CASE	1,244	0.52	367	0.02	3,287.70	<0.01
ADMINISTRATIVE	860	0.36	27		3,157.42	<0.01
BANKRUPTCY	761	0.32	9		2,905.15	<0.01
DEBTOR	649	0.27	1		2,543.93	<0.01
CASES	864	0.36	177	0.01	2,510.60	<0.01
CONFERENCE	747	0.31	85		2,421.85	<0.01
C	852	0.36	259	0.02	2,230.45	<0.01
E	844	0.35	279	0.02	2,152.33	<0.01
DISTRICT	685	0.29	98		2,139.45	<0.01
OFFICE	830	0.35	272	0.02	2,122.50	<0.01
H	642	0.27	94		1,996.79	<0.01

この処理により, いわゆる司法英語において特徴的に用いられる語彙がかなりうまく抽出できた. しかし, ここにはいくつかの不適切なデータの混入が見られる. 1 つ目はエラー語(C, E, H など), 2 つ目は重複した派生形(court, courts など), 3 つ目は低頻度語である. エラー語は手作業で削除し, 派生形については上

位ランク語を基本形に戻して採用し、下位ランク語を削除することとした。ただし、名詞化・形容詞化している分詞形は、原則としてそのまま残した。また、ターゲットコーパスでの全体に占める構成比が 0.01%に満たない低頻度語はリストから削除した。さらなる問題としては、頭字語 (FS など) の混入と、司法英語語彙とは言いがたい一般語彙 (the, office など) の混入がある。前者については別リストを作成することとし、後者については一定の基準で削除することにした。

5. 頭字語リスト

頭字語には、全体を新たな発音として読む acronym と、それぞれのアルファベットとして発音する initialism があるが、司法英語に多く見られるのは後者である。これらの頭字語は、一般的な語とは機能を異にするものであり、別個の扱いが適当であろう。

そこで、500 語の特徴語基礎リストのうち、頭字語と判断される語を抽出し、「米国司法英語 ESP 語彙表」の本表に付随する付表とすることとした。表 2 に示すように、付表には全体で 12 語が含まれる。なお、ターゲットコーパスと参照コーパス間での出現頻度差の有意水準はいずれも $p < 0.01$ である。

表 2. 頭字語リスト(「米国司法英語 ESP 語彙リスト付表」)

Word	Original Meaning	TC Freq	TC Ratio	RC Freq	RC Ratio	Keyness
FS	Foreign Service	910	0.38	0	0.00	3588.39
CJRA	The Civil Justice Reform Act	241	0.1	0	0.00	949.75
ADR	Alternative Dispute Resolution	159	0.07	0	0.00	626.55
SIPC	Securities Investor Protection Corporation	56	0.02	0	0.00	220.65
ECF	Electronic Case File	44	0.02	0	0.00	173.37
PACER	Public Access to Court Electronic Records	41	0.02	0	0.00	161.55
SIPA	Securities Investor Protection Act	40	0.02	0	0.00	157.61
FJC	Federal Judicial Center	32	0.01	0	0.00	126.08
GSA	General Services Administration	27	0.01	0	0.00	106.38
CM	Case Management	43	0.02	21		94.73
CJA	Criminal Justice Act	24	0.01	0	0.00	94.56
FDO	Federal Defender Organization	24	0.01	0	0.00	94.56

各種の法律名称(CJRA, SIPA, CJA)や、長期化の進む司法裁判に代わる「代替的紛争解決手法」を指す ADR (大橋, 2002, p.29), あるいは判例情報電子化にかかる語(ECF, PACER)など、上記頭字語は近年の米国司法制度改革の鍵概念を指すものであり、関連文献を読む際に正確な理解は欠かせない。

しかしながら、これらの語は辞書ではいまだ十分な扱いがなされていない。大学生になじみの深い『ジーニアス英和辞典』第3版(大修館書店),『リーダーズ英和辞典』第2版(※補遺版の『リーダーズ・プラス』を含める)(研究社出版), *Collins COBUILD Advanced Learner's English Dictionary* [COBUILD4], *Longman Advanced American Dictionary* [LAAD]の4つの辞書を取り上げ, 上記の頭字語ないし元になった複合語が見出し語(追い込み含む)として記載されているかどうかを調査した結果を表3に示す。

表3. 頭字語の辞書エントリー状況

	ジーニアス		リーダーズ		COBUILD4		LAAD	
	複合語	頭字語	複合語	頭字語	複合語	頭字語	複合語	頭字語
FS	×	×	○	○	○	×	×	×
CJRA	×	×	×	×	×	×	×	×
ADR	×	×	×	○	×	×	×	×
SIPC	×	×	×	○	×	×	×	×
ECF	×	×	×	×	×	×	×	×
PACER	×	×	×	×	×	×	×	×
SIPA	×	×	×	×	×	×	×	×
FJC	×	×	×	×	×	×	×	×
GSA	×	○	×	○	×	×	×	×
CM	×	×	×	×	×	×	×	×
CJA	×	×	×	×	×	×	×	×
FDO	×	×	×	×	×	×	×	×

百科的事項に強い『リーダーズ』の場合でも4語しか記載がなく, その他のEFL学習辞書に至っては, 頭字語としても複合語としてもほとんど対応できていないことが明らかになった。この結果は, 司法英語のような専門性の高い文献を読む場合には, 専用のESP語彙表が必要となることを再確認させるものである。

6. 一般語彙の分離

本研究で開発を試みようとするESP語彙表はEGP語彙表とは峻別されるものであり, EGP語彙表の修得が終わった後の段階で取り組む独立した語彙表である(8章の図2参照)。つまり, 司法文献においていかに高い出現特徴度を持つ語であっても, それが同時に一般的文献にも頻出する場合はむしろEGP語彙として扱われるべきであり, ESP語彙表からは除外するのが適当と考えられる。

こうした考え方に立ち, 本研究では, 500語の特徴語リストのうち, 一般英語を代表する参照コーパスにおいて100回以上出現した115語を削除することにした。削除対象になった語のうち, 特徴度500以上の33語をリストしたのが表4である。

表4には, federal, office, management, plan, service(s), report, cost, program, information, staff, congress, systemといった語群が見られる。これらは司法文献の特徴語というよりも, むしろ一般英語のジャンルに属する語彙と言うべきであろう。なお, 表中, 司法英語の特徴をとくに強く伝えているものとしては, court, case, chapter, justice, rule, law, security, deathなどの語が注目される。

表 4. 司法文献に特徴的に出現する一般語彙

Keyword	TC Freq	TC Ratio	RC Freq	RC Ratio	Keyness
COURT	2037	0.85	216	0.01	6681.64
FEDERAL	1505	0.63	221	0.02	4682.99
THE	17171	7.19	62351	4.23	3599.07
CASE	1244	0.52	367	0.02	3287.7
OFFICE	830	0.35	272	0.02	2122.5
MANAGEMENT	658	0.28	133		1917.4
OF	8651	3.62	32273	2.19	1605.8
AND	7433	3.11	28002	1.9	1326.32
PLAN	506	0.21	176	0.01	1268.46
SERVICE	475	0.2	153	0.01	1220.82
REPORT	489	0.2	185	0.01	1190.82
COST	436	0.18	169	0.01	1052.49
PROGRAM	521	0.22	345	0.02	992.64
CHAPTER	394	0.17	135		992.49
INFORMATION	480	0.2	273	0.02	987.65
STAFF	419	0.18	183	0.01	966.88
CONGRESS	402	0.17	165	0.01	950.09
YEAR	655	0.27	771	0.05	845.9
FOR	2781	1.17	9229	0.63	736.94
MAY	660	0.28	939	0.06	715.64
SYSTEM	465	0.19	481	0.03	665.96
JUSTICE	277	0.12	118		645.33
TO	6143	2.57	26215	1.78	639.34
UNDER	479	0.2	549	0.04	632.38
ACT	344	0.14	270	0.02	594.54
IN	4888	2.05	20367	1.38	569.09
RULE	255	0.11	129		553.43
LAW	357	0.15	345	0.02	537.57
UNITED	381	0.16	406	0.03	533.23
DEATH	348	0.15	335	0.02	525.47
SECURITY	241	0.1	121		524.77
OR	1374	0.58	3993	0.27	508.96
STATE	417	0.17	525	0.04	507.6

これらの語彙は一般英語でも一定の頻度を持つものであるが、司法文献では限定的な意味合いのないシロケーションで用いられるのが通例である。たとえば、司法文献中の case はここでは”a question or

problem that will be dealt with by a court of law”(LAAD)の意味で用いられる。justice も一般的な「正義」というよりも, “the system by which people are judged in courts of law and criminals are punished”(LAAD)という限定的な意味での使用である。また, chapter はふつう広く書物・刊行物の中での区切りを指す語であるが, ここでは法律内の「章・条項」を指しており, death も death penalty (死刑) のコロンで多く使われている。

表 4 に記載された語は原則的には EGP に分類されるべきものであるが, すでに見たように, 場合によっては司法文庫の中で通常といくぶん異なる意味合いで使用されることもあるため, 留意が必要である。

ところで, 表 4 を検証すると, of と and が特徴度指数(keyness)順で 7 位と 8 位に入っていることに気づく(※さらに上位 30 語までに for, to, under, in, or も含まれる)。いわゆる内容語(content words)の頻度がテキストのジャンルによって影響を受けることは容易に想像できるが, 機能語(functional words)である of や and の頻度が司法文庫コーパスと参照コーパス間で大きなずれを見せているのは興味深い事実である。

ここで, 司法文庫における of と and の出現状況を概観しておきたい。両語の出現状況を観察するため, 中心語(NODE)に対して左右 5 語を測定スパンとして, 共起語頻度検証を行うこととする。

L5	L4	L3	L2	L1	NODE	R1	R2	R3	R4	R5
{ Total L }						{ Total R }				

表中の Total L は中心語の左方位置に出現した語(L5-L1)の総計頻度, Total R は右方位置に出現した語(R1-R5)の総計頻度, Total はそれらの総計(L5-R5)を示す,

表 5. of/and の共起語

of					and				
No	Word	Total	Total L	Total R	Word	Total	Total L	Total R	
1	THE	9563	4765	4798	THE	5199	2352	2847	
2	AND	2859	1290	1569	OF	2859	1569	1290	
3	A	2067	1145	922	TO	1840	913	927	
4	IN	1615	784	831	IN	1333	593	740	
5	TO	1598	844	754	FOR	1037	553	484	

上表に明らかなように, of の場合も and の場合も共起頻度上位 5 語の中に内容語は一切出現しない。このことは, 狭いスパンの中で, 機能語, とくに文の統語的複雑性を高める前置詞が反復して使用されていることを示す。下記は集中的に機能語が使われているデータの一例である(※強調は論者)。

In cases under chapter 12 (adjustment of debts of a family farmer) and 13 (adjustment of debts of an individual with regular income)... (“BASICS Bankruptcy”)

Provides support to the courts for federal defenders, probation and pretrial services officers, clerks of court, circuit executives, staff attorneys, conference attorneys, court reporters,

interpreters, *and* bankruptcy administrators including the development of budgets, allocation of resources, *and* management of national programs. ("2000 Report of the Director")

前者の例では、1つの名詞句内での of の重複使用が見られる。また、後者の例では、*and* を介在した名詞句の羅列が起きている。司法英語が、前置詞によって名詞句を次々に結合させてゆく統語的特徴を持っている可能性をここで指摘しておきたい。

7. 司法英語 ESP 語彙表の確定

特徴語基礎リスト 500 語から、エラー語・重複派生形・低頻度語などを除去し、さらに頭字語と参照コーパスで 100 回以上の出現度を持つ EGP 語を削除した結果、256 語が残り、これを「司法英語 ESP 語彙表(本表)」の候補語とすることにした(Appendix)。なお同表はそれぞれのコーパスでの頻度と特徴度のみを記載することとし、構成比については省略する。

8. まとめ:ESP 語彙指導への示唆

本研究で試みたように、特定 ESP ジャンルのコーパス・データと、より一般性の高い参照コーパスを比較し、それぞれの頻度データの統計的比較をふまえて特徴語を抽出するアプローチは、ESP 語彙表の開発に高い有用性と汎用性を持っていることが示された。言語データに基づくこうした専門分野別語彙表の開発は、英語教育が EGP から ESP への傾斜を強めていくなかで、今後、大規模に取り組まれるべき課題である。

ESP 語彙指導に関して留意すべきは、EGP 語彙との有機的な連続性を確保することである。EGP 語彙表のサイズについては、古くは West (1953) の「General Service List」の 2000 語(※ただし、これは独自の基本形見出しとなっており、実際には約 2300 語になるとされる)、大学英語教育学会が制定した「JACET4000」(旧版)の 4000 語、「JACET8000」(新版)の 8000 語、さらにはアルク社が選定した「標準語彙水準 SVL12000」の 12000 語など、さまざまな見解が存在しているが、EGP としてあまりに多くの語を取り込むよりも、精選された EGP の上に、分野別の ESP 語彙表を組み合わせるほうが学習の効率性が高まるものと推測される。

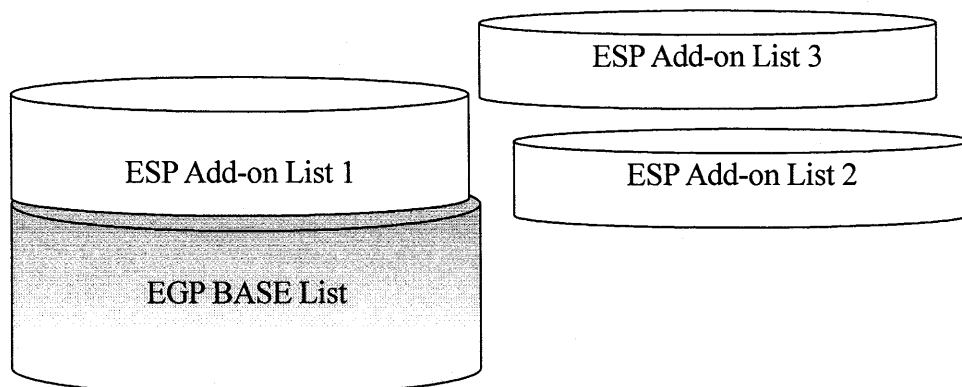


図 2. 語彙指導の概念図

上記の概念図に示すように、EGP 語彙表がさまざまな学習者の共通基盤であるとするれば、ESP 語彙表は必要に応じてそれに上積みする「追加リスト(add-on list)」として位置づけられるべきものである。

工学・医学・看護学・法学・経済学・社会学など、自然科学系・人文系を含めた幅広い学問分野や職種に対応した追加リスト群(ESP Add-on List 1, 2, 3...)をあらかじめ用意しておき、学習者の個々のニーズに応じて、EGP 語彙表と適切な ESP 語彙表を組み合わせた語彙カリキュラムを提示していくことが、今後の英語教育、とくに語彙指導において求められることとなる。

今回は、米国連邦裁判所発行の司法文献に基づき、「司法英語」というジャンルに限って試行的な語彙表の開発を行ったわけであるが、今後は本研究の成果をふまえ、大学英語教育における EGP 語彙表の適切なサイズの検証、他の専門分野の ESP 語彙表の開発、当該分野の専門研究者を交えた ESP 語彙教材の開発などに取り組んでいきたい。

注記

本研究は、文部科学省の科学研究費基盤研究(B)(2)「大規模コーパスを用いた日本人英語学習者用語彙リスト構築と教材分析システムの開発」(研究代表者:相澤一美, 課題番号:16320076)の助成を受けて実施されたものである。また、本研究の内容の一部は、2004 年 11 月 27 日、神戸大学国際文化学部において開催された言語文化学会第 18 回大会において、「法学系学生のための ESP 語彙表開発の試み:米国司法コーパスと FROWN コーパスの比較による特徴語の抽出および ESP 語彙オンライン学修システムの構築に向けて」の題で発表を行った。

引用文献

- Aizawa, K. (1999). *A study of incidental vocabulary learning through reading by Japanese EFL learners*. Unpublished PhD Dissertation. Tokyo Gakugei University.
- Anthony, L. (2004a). Antconc (Version 2.6.0) [Computer software]. Retrieved January 10, 2005, from <http://www.antlab.sci.waseda.ac.jp/>
- Anthony, L. (2004b). Antconc (Version 3.0) [Computer software]. Retrieved January 10, 2005, from <http://www.antlab.sci.waseda.ac.jp/>
- Brumfit, C. J., & Johnson, K. (1989). *The communicative approach to language teaching*. Oxford, UK: Oxford University Press.
- Dahl, H. (1979). *Word frequencies of spoken American English*. Detroit, MI: Verbatim.
- 大学英語教育学会基本語改訂委員会 (編). (2003). 『大学英語教育学会基本語リスト JACET List of 8000 Basic Words』. 東京: 大学英語教育学会.
- Dominguez, G. A., & Rokowski, P. E. (2002). Bridging the gap between English for academic and occupational purposes. *English for Specific Purposes World*, 2 (1). Retrieved January 10, 2005, from http://www.esp-world.info/Articles_2/issue_2.html
- Dudley-Evans, T. (2001). English for specific purposes. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 131-136). Cambridge, UK: Cambridge University Press.
- Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A*

- multi-disciplinary approach*. Cambridge, UK: Cambridge University Press.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.
- Hofland, K., & Johansson, S. (1982). *Word frequencies in British and American English*. Bergen, The Norwegian Computing Centre for the Humanities.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge, UK: Cambridge University Press.
- James, G., Davison, R., Cheung, A. H. Y., & Deerwester, S. (Eds). (1994). *English in computer science: A corpus-based lexical analysis*. Hong Kong: Hong Kong University of Science and Technology.
- 川北直子・白坂佳代. (2004). 「意味のあるオンライン/オフライン学習環境とは？—看護大学英語プログラム実践についての考察より—」. 『外国語教育メディア学会(LET)第44回(2004年度)全国研究大会発表論文集』, 52-55.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Harlow, UK: Pearson Education Limited.
- 松浦千佳子. (2004). 「日本人エンジニアの英語論文への文化的影響」. 『外国語教育メディア学会(LET)第44回(2004年度)全国研究大会発表論文集』, 106-109.
- 深山晶子(編). (2000). 『ESPの理論と実践:これで日本の英語教育が変わる』. 東京:三修社.
- 望月正道・相澤一美・投野由紀夫. (2003). 『英語語彙の指導マニュアル』. 東京:大修館書店.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle.
- 西原貴之. (2004). 「工業英検3・4級レベルの学習項目に対する学習者のニーズ分析」. 『第30回全国英語教育学会長野研究大会発表要綱』, 530-533.
- 大橋真由美. (2002). 「行政手続に関する調査研究」. 『平成14年度提携プロジェクトⅡ—行政手続に関する調査研究—報告書』, 一橋大学総合法政政策実務提携センター, 29-39.
- Scott, M. (2004). Oxford wordsmith tools (Version 4.0) [Computer software]. Oxford, UK: Oxford University Press.
- Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford, UK: Blackwell.
- Thorndike, E., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College, Columbia University.
- West, M. (1953). *A general service list of English words*. London, UK: Longman, Green and Company.
- 山本淳子・大倉孝昭. (2004). 「看護学生のコミュニケーション能力向上を目指したCALLシステム」. 『外国語教育メディア学会(LET)第44回(2004年度)全国研究大会発表論文集』, 60-62.
- 吉留文男. (2004). 「医療関係者の視点に基づくESP教材分析について」. 『第30回全国英語教育学会長野研究大会発表要綱』, 534-535.

Appendix. 司法英語 ESP 語彙表(本表)

#	Key word	Keyness	TC Freq.	RC Freq.	#	Key word	Keyness	TC Freq.	RC Freq.
1	JUDICIAL	4512.59	1206	26	46	PILOT	419.65	151	37
2	JUDICIARY	3819.68	1005	14	47	REVIEW	417.57	191	95
3	JUDGE	3780.55	1029	32	48	PROCEDURE	413.9	167	60
4	ADMINISTRATIVE	3157.42	860	27	49	CHIEF	408.18	188	95
5	BANKRUPTCY	2905.15	761	9	50	RESOURCE	404.47	185	92
6	DEBTOR	2543.93	649	1	51	EXECUTIVE	400.32	180	86
7	CONFERENCE	2421.85	747	85	52	PETITION	394.77	122	14
8	DISTRICT	2139.45	685	98	53	LAWYER	384.08	158	60
9	CIVIL	1438.87	479	84	54	BRANCH	374.31	145	46
10	PENALTY	1260.31	355	18	55	REQUIREMENT	363.45	144	49
11	FISCAL	1230.78	405	67	56	EVALUATION	362.65	133	35
12	PRETRIAL	1161.34	298	1	57	EMPLOYEE	354.85	155	69
13	TRIAL	1102.16	377	75	58	REDUCTION	354.03	139	46
14	CREDITOR	1000.26	268	6	59	LITIGANT	346.75	88	0
15	COUNSEL	964.56	269	12	60	INTERNET	342.81	87	0
16	TRUSTEE	886.65	235	4	61	AUTHORIZE	335.04	112	20
17	PROBATION	883.99	230	2	62	APPELLATE	328.25	86	1
18	LITIGATION	882.75	259	20	63	ADDITIONAL	327.77	152	78
19	DISCOVERY	855.99	275	40	64	SALARY	323.99	107	18
20	DELAY	843.44	246	18	65	IMPLEMENT	321.37	100	12
21	CRIMINAL	830.1	271	43	66	SENTENCE	300.91	81	2
22	CIRCUIT	656.47	207	27	67	CONFIRMATION	298.87	88	7
23	DISCHARGE	643.96	178	7	68	REQUEST	296.31	109	29
24	ELECTRONIC	640.74	209	33	69	COURTHOUSE	291.3	86	7
25	ATTORNEY	619.84	196	26	70	DEMONSTRATION	286.42	102	24
26	BENEFIT	617.56	253	95	71	COURTROOM	278.29	85	9
27	JUDGESHIP	606.85	154	0	72	LEGISLATION	274.23	134	77
28	DEFENDANT	605.34	183	18	73	APPROPRIATION	269.88	85	11
29	SUPERVISION	588.01	168	10	74	OFFICE	265.38	96	24
30	RECOMMENDATION	582.49	192	32	75	CLERK	260.79	93	22
31	DEBT	562.75	168	15	76	FUNDING	260.46	114	51
32	FILING	542.16	156	10	77	REORGANIZATION	260.06	66	0
33	JURY	540.22	198	52	78	CONDUCT	257.01	107	42
34	APPEAL	531.6	172	26	79	OPERATION	256.47	133	86
35	IMPLEMENTATION	509.64	167	27	80	LIQUIDATION	256.12	65	0
36	MAGISTRATE	496.5	126	0	81	PRINCIPLE	255.83	129	79
37	GUIDELINE	475.12	147	17	82	CLAIM	246.13	130	87
38	OFFICER	468.5	209	98	83	DISPOSITION	240.39	77	11
39	APPOINT	455.29	146	21	84	STATUTORY	238.73	65	2
40	BUDGET	452.24	202	95	85	INTERPRETER	236.41	60	0
41	ADVISORY	450.14	138	15	86	ENSURE	235.4	97	37
42	FILE	441.46	151	30	87	FACILITY	233.09	106	52
43	PROCEEDING	437.97	137	17	88	REPRESENTATION	233.09	106	52
44	CODE	429.68	162	47	89	AUTOMATE	231.86	69	6
45	DEFENDER	425.46	124	9	90	APPOINTMENT	226.52	84	23

#	Key word	Keyness	TC Freq.	RC Freq.	#	Key word	Keyness	TC Freq.	RC Freq.
91	REGARDING	223.57	98	44	136	SITE	153.96	92	76
92	MAIL	223.06	83	23	137	BENCH	153.35	61	21
93	QUALITATIVE	222.95	64	4	138	SENATE	152.12	88	69
94	PARTY	220.15	121	87	139	STATUTE	150.8	45	4
95	OFFENDER	216.57	77	18	140	HOURLY	149.82	42	2
96	PAYMENT	210.69	77	20	141	ELECTRONICALLY	148.5	40	1
97	FINDING	209.37	96	48	142	ANNUAL	146.38	84	65
98	DOCKET	208.83	53	0	143	ASSIGNMENT	146.29	51	11
99	LEGISLATIVE	207.64	83	29	144	COMMUNICATION	145.28	62	26
100	ACCOUNTING	205.51	73	17	145	STAFFING	144.61	39	1
101	PERSONNEL	204.12	99	56	146	DEADLINE	144.34	42	3
102	AMENDMENT	199.63	73	19	147	TYPICALLY	143.62	70	40
103	AGENCY	196.09	106	74	148	CASELOAD	141.84	36	0
104	STATISTICAL	194.84	74	22	149	DOCUMENT	140.71	81	63
105	IMPACT	191.75	105	75	150	FORMULA	140.54	41	3
106	SCHEDULING	191.37	51	1	151	TELEPHONE	139.95	71	44
107	APPROVE	188.79	87	44	152	PLAINTIFF	135.75	41	4
108	ALTERNATIVE	188.68	95	58	153	MOTION	135.67	48	11
109	UNSECURED	181.25	46	0	154	DISCLOSURE	135.33	86	78
110	MATTER	180.5	97	67	155	DETERMINE	135.09	89	85
111	REDUCE	180.1	74	28	156	FEE	134.71	51	15
112	JURISDICTION	179.1	73	27	157	COMMISSION	134.57	83	72
113	RECOMMEND	178.5	57	8	158	CONGRESSIONAL	134.06	64	35
114	SAVING	178.11	72	26	159	INTERCIRCUIT	133.96	34	0
115	DISPUTE	177.84	74	29	160	COMPLETE	133.6	84	75
116	PROSECUTION	177.52	65	17	161	TECHNIQUE	131.52	85	79
117	SUBCOMMITTEE	175.14	60	12	162	ENACT	129.19	47	12
118	WORKLOAD	174.85	50	3	163	EXPENSE	128.89	64	38
119	REFORM	171.03	102	84	164	RECEIVE	127.82	72	54
120	COMPENSATION	170	86	53	165	INTERPRET	124.21	43	9
121	SETTLEMENT	168.6	67	23	166	ASSET	124.05	56	27
122	CONSTITUTION	168.43	81	45	167	ADJUSTMENT	123.64	71	55
123	VACANCY	167.97	45	1	168	RESOLUTION	120.96	71	57
124	NATIONWIDE	167.7	57	11	169	OBTAIN	119.34	52	23
125	QUANTITATIVE	167.7	57	11	170	ENHANCE	119.26	50	20
126	AUTOMATION	165.86	60	15	171	PROCUREMENT	119.12	34	2
127	ASSESSMENT	165.12	77	40	172	IMPROVE	118.76	73	63
128	WEB	164.43	57	12	173	SUBMIT	118.49	53	25
129	REQUIRE	163.87	67	25	174	EFFICIENT	116.75	61	40
130	ASSIST	158.31	66	26	175	MANAGE	116.53	39	7
131	RESPONSIBILITY	158.31	66	26	176	AUTHORIZATION	115.32	52	25
132	PANEL	158.26	73	37	177	PHASE	115.04	63	45
133	SUPREME	158.15	82	53	178	MANDATORY	114.66	42	11
134	HEARING	157.63	63	22	179	QUALIFIED	114.53	57	34
135	MUNICIPALITY	157.61	40	0	180	DATABASE	114.33	45	15

#	Key word	Keyness	TC Freq.	RC Freq.	#	Key word	Keyness	TC Freq.	RC Freq.
181	ADOPT	113.62	64	48	226	VOLUNTARY	86.81	38	17
182	PRIOR	112.93	69	59	227	NET	86.7	50	39
183	TIMELY	110.97	44	15	228	LAWBOOK	86.68	22	0
184	BAR	110.6	81	89	229	PRESENTENCE	86.68	22	0
185	PENDING	110.6	46	18	230	EMPLOYMENT	86.4	55	50
186	AMOUNT	110.58	83	94	231	CONSULTATION	86.39	33	10
187	DATE	109.72	74	73	232	CUSTOMER	86.17	42	24
188	HABEAS	109.59	34	4	233	RETIREMENT	85.37	40	21
189	INVOLVING	109.34	61	45	234	ALLOW	84.47	59	61
190	NOTICING	109.2	36	6	235	ENTITLED	84.47	42	25
191	COMPREHENSIVE	108.85	54	32	236	SUPPLEMENTAL	84.41	30	7
192	ADMINISTRATOR	108.58	44	16	237	APPROVAL	83.79	46	33
193	NORTHERN	108.57	66	56	238	TRACK	83.79	62	69
194	ESTATE	108.38	65	54	239	SUMMARY	82.83	42	26
195	TEMPORARY	105.41	47	22	240	LIEN	82.74	21	0
196	STATISTICS	105.37	55	36	241	OPERATIONAL	82.44	34	13
197	ENDORSE	104.28	39	11	242	SETTING	82.36	63	73
198	TYPE	103.81	75	81	243	POSSESSION	82.28	39	21
199	ASSISTANT	103.67	57	41	244	PROCEDURAL	82.19	31	9
200	OBLIGATION	103	43	17	245	TITLE	81.85	62	71
201	JUROR	102.46	40	13	246	MANUAL	81.48	41	25
202	DECIDE	102.32	55	38	247	SUPERVISE	80.34	25	3
203	MARSHAL	101.92	28	1	248	UPDATE	80.15	27	5
204	RELIEF	100.9	58	45	249	AUDIT	80.08	26	4
205	RELEASE	100.76	64	58	250	REPRESENTATIVE	80.01	53	51
206	EMERGENCY	100.21	69	70	251	ONGOING	79.75	37	19
207	CONDUCTING	99.09	39	13	252	OVERSIGHT	79.59	31	10
208	EFFECTIVENESS	98.72	43	19	253	SEEK	79.46	57	61
209	CONTRACT	98.24	58	47	254	CERTIFYING	78.8	20	0
210	EXISTING	97.5	70	75	255	CONVICTION	76.16	37	21
211	OUTREACH	96.23	28	2	256	OFFENSE	75.37	40	27
212	CHAMBER	96.02	35	9					
213	POSITION	94.91	60	54					
214	GUIDANCE	94.73	43	21					
215	ACTION	93.53	70	79					
216	LAWSUIT	92.14	39	16					
217	VIDEOCONFERENCING	91.51	28	3					
218	POSTAGE	90.33	25	1					
219	DUTY	89.87	46	29					
220	CONTINUING	89.48	49	35					
221	SECURE	89.08	33	9					
222	FUND	88.9	38	16					
223	RESOLVE	88.47	41	21					
224	CORPORATION	88.29	46	30					
225	COMPLAINT	88.11	35	12					

Development of an ESP Vocabulary List for College Law Students: Based on the Statistical Comparison of Word Frequencies in US Judiciary English Corpus and Freiburg-Brown Corpus

Shin'ichiro Ishikawa

Recent college education has begun to put increasing emphasis on job preparation training. This overall trend has caused college English education to shift gradually from EGP (English for General Purposes) to ESP (English for Specific Purposes).

When conducting an ESP educational program, we need to use "authentic material" that reflects the verbal features of the text in the particular ESP genre. The authenticity of the text is largely dependent on the vocabulary, which demonstrates the importance of developing an ESP vocabulary list for students.

Several ESP vocabulary lists have been compiled, but, in the genre of law or jurisprudence, few lists that are available are based on reliable frequency data. Therefore, in this paper, we aim to develop a corpus-based ESP vocabulary list for college law students.

For this purpose, we compiled a US Judiciary English Corpus (US-JEC), which includes approximately 240,000 words used in official judiciary documents. We then compared a raw frequency word list created from US-JEC as a target corpus with a list created from Freiburg-Brown (FROWN) Corpus as a reference. Frequencies taken from the two different base data were adjusted on the basis of log-likelihood statistical scores.

Thus, statistical key words in the US JEC were selected. After deleting a group of error words, acronyms and EGP words that occur at a considerable rate in the reference corpus, we finally came up with a proposed brand-new ESP vocabulary list for college law students.