

情報科学分野の英語語彙親密度測定の試み

古樋 直己

情報科学部 情報メディア学科
(2020年11月24日受理)

An Attempt to Measure the Familiarity Levels of English Words and Phrases in
Information Science

by

Naoki FURUHI

Department of Media Science,
Faculty of Information Science

Abstract

This study aims to assess the level of comfort that students majoring in Information Science (IS) have with words and phrases used in the domain and, thereafter, sort the vocabulary according to an average level of familiarity. For the experiment, 78 undergraduate students were asked to rate, on a scale of 1 to 7, their familiarity with the English vocabulary from a set of 296 words and phrases commonly used in IS. The results reveal the familiarity of the students with domain-specific vocabulary and provide valuable insights for pedagogical matters.

キーワード ; 親密度, 単語, 語彙, 情報科学

Keywords ; familiarity, words, vocabulary, information science

1. はじめに

情報科学分野の英語の語彙、すなわち、単語や語句の親密度の測定を試みた。

まず、語彙の重要性、語彙の頻度、親密度について概説する。

英語運用能力の構成要素として語彙は重要な位置を占めている。語彙が重要であるとして、どのように習得するのかという学習方法を考えることと、学習すべき語彙を確定することが必要である。

習得の方法は別の機会に譲るとして、学習すべき語彙の選定に注目する。英語の語彙のすべてを習得することは非現実的である。したがって、どのような語彙を学習対象として選択し、集中的に習得を目指すべきか熟慮しなければならない。

主要な選定基準の一つとして、頻度が挙げられる。よく使われる語彙から覚えて使えるようにしようという発想であり、語彙選定を思い立った時に多くの人々がまず想起する基準であろう。

頻度を考えるにしても、分野や素材がどのようなものであるのかにより数値は様々な値を取るはずである。また、単語といっても様々な活用形を同一語と考えるのかどうか等、数え方によって結果が変わってくる。ただ、明確な規則を策定した上でなら、客観的な頻度データを得ることができ、頻度順の語彙リストを作成すれば、効率よく、重要な語彙を優先的に習得していけるはずである。実際に、学習用の語彙選定方法として、頻度を重要な指標としているものが多い。

しかし、資料を選び、語の定義をどう定めるのかの他にもまだ問題が残っている。たとえば、頻度が高いということならば、機能語がまずリストの上位に挙がってくる。前置詞、代名詞、助動詞等が機能語の例であるが、このような語彙はリストから除外すべきかもしれない。高頻度とはいえ、英語学習の初期段階ですでに深くはなくとも浅くは知っているはずである。

このように、頻度順のリストにも課題がある。そこで、他の指標として親密度が考案された。親密度とは、見聞きしたり、使ったりする度合いを表したものであり、頻度のみで学習語彙を選定する問題点を補完できる可能性がある。

この親密度は、いわば心的頻度であり、「当該語が学習者によって実際に処理される回数や処理の深さなどによって左右されるような頻度」であり、「同じ

回数遭遇しても学習されやすい単語と学習されにくい単語がある」こととも関係している。さらに、「英語としてではなくカタカナ語として見聞きされたもの」(横川, 2006, pp.69-70)もこの心的頻度である親密度に影響する可能性も考えられる。

横川(2006)では、英単語に対する親密度調査を約800名の大学生に対して実施した。一般的な英語のコーパスであるBNC (British National Corpus)を元にした Kilgarriff Lemmatised Frequency List から約3,000語を抽出し、調整を加えた2,999語を親密度調査の対象に挙げている。

なお、lemmatised とあるようにこのリストはレマ化された語彙が挙げられており、たとえば、have, has, having, had は have としてまとめて分類されている。つまり、辞書の見出し語のようなものがレマ化された表し方である。

親密度は7段階で評点を付けるように調査協力者に依頼している。より具体的には、とてもよく見聞きするに当てはまるなら7点、全く見聞きしないなら1点で、その中間は馴染みの度合いに応じて1点刻みで回答してもらう方式である。

また、西出・水本(2009)では、大学生22名と調査協力者数は小規模ながら、語数は8,000語と大規模な親密度測定を実施している。素材はアルク社の Standard Vocabulary List のレベル1から8までの8,000語である。このリストはレマ化されていない。

このように英単語の親密度調査としては、一般的な英語語彙を対象としたものは、語数としては大規模なものが存在している。

そこで本稿では、調査対象の語彙を情報科学分野とした。分野を限定したのは、この分野の語彙について親密度の調査は見当たらないことと、学習者の専攻分野の語彙について親密度を把握しておくことは学習指導に有益と考えるためである。

さらに、日本語の中ですでに馴染み深いカタカナ語として定着しているかどうかも親密度に影響することを考え、本調査では、情報科学分野の英語語彙で、カタカナ語として通用するようなものを選定した。カタカナ語に着目したのは、語彙指導に有益なためである。

調査協力者は78名、調査対象とした語彙は約300項目である。

2. 方法

(1) 語彙の選定

調査に用いる情報科学分野の英語語彙は平井(2012)に挙げられている語句を採用した。512語句が掲載されているが、この中から296語句を採用した。

同書を選んだ理由、語彙の選定方法について述べる。本稿では、リスト作成及び調査実施上の人的・時間的制約により、情報科学分野の語彙が掲載されている既成の書籍を用いることにした。

今回の調査を計画する際に情報科学分野の辞典も検討したが、パイロットスタディのような位置づけと考えれば、最初からある程度語彙数が限定されている素材でもよいのではないかと考えた。そこで、平井(2012)を活用することにした。

この書籍には情報科学分野で用いられることが多い512語句の単語や語句が列挙されており、それぞれに和訳や関連の英語表現、解説等が掲載されている。語数も抑制されており、情報科学分野用の英語に関して1冊にまとまっていることもあって、今回の用途に適した素材と考えた。

512語句のうち296語句に限定した理由は、先述のように、パイロットスタディとしてならば、横川(2006)の約3,000語という語数を踏襲するまでもなく、これでまずは十分と判断したためである。まずは小規模でも実際に調査してみることを優先した。

512語句から200語句以上を削減する手順は次の通りである。まず、カタカナ語として通用しそうなものを優先した。より具体的には、平井(2012)においてカタカナ語の和訳が挙げられているものを採用した。和訳にカタカナ語が採用されているものの大部分は素材になっている英語の発音に準じたものである。

情報科学分野の英語であるから、容易に想像できるように、和訳としてカタカナ語が記載されているものが多い。たとえば、プロシージャ、スループット、オブジェクト等である。

調査に用いた語句は、付録に親密度の平均点(7点満点)の高い順に、その順位、標準偏差とともに全て掲載している。

(2) 調査協力者及び調査実施方法

情報科学を専門とする大学2年生を調査協力者とした。教養課程の英語科目の授業時間中に成績には無関係であることを説明した上で後期授業の最終回に実施した。

1枚50問のマーク用紙を6枚配布し300問解答

を依頼した。付録の一覧表の解答用紙の欄はその何枚目のマーク用紙であるのかを示している。また、問題番号の欄は解答用紙の1から50の番号である。

プレゼンテーションソフトウェアを用い、スライド1枚当たり25問を学生用机に設置されたディスプレイに提示した。スライド1枚当たり100秒つまり語句1つで平均4秒での解答を調査協力者に要求したことになる。

スライド2枚分で解答用紙1枚分の50問となり、これを6回繰り返して解答用紙6枚、スライド12枚で説明のスライドも含め20分強で、300項目の語彙について親密度を見聞きする度合いが最も低い1点から最も高い7点の7段階で回答するように指示した。

3. 結果

全296語句に対する結果は付録の一覧表の通りである。親密度の平均点順で挙げている。順位が同じものは平均点が等しいためである。たとえば、6位にはdataとe-mailが挙がっている。

標準偏差は調査協力者内での評点の付け方の分布を表している。上位2語句ではともに0.3台であり、平均点、つまり、順位が下がるにつれて、最初のうちは1から2程度、150位くらいからは2を超えるものも増えてくる。

この事実から、平均点が上位の語句ではほぼ全員について、その語句の親密度が高いが、平均点が下がるにつれて、調査協力者により親密度が異なる度合いが概ね大きくなるのがわかる。

また、上位の語句は情報科学を専門としない者でも見聞きすると思えるものが並んでいる。

さらに、最下位の語句から親密度の平均点が高くなっていく語句を概観すると、先述の上位の語句に比べて英語の語彙としてもカタカナ語としてもなじみが薄そうなものが並んでいる。

4. 語句選定上の制約と今後の課題、結果の教育への活用

調査対象に選定した英語の語句であるが、横川(2006)の選定作業にならって、情報科学分野の語彙コーパスを作成し、頻度順リストを得て、それを元に何らかの客観的な手法で語彙を選択することも考えられるが、今回は平井(2012)を活用した。

他にも、特に情報科学分野は時代によって用いら

られる語彙の変化が激しく、今回の調査語彙に含めていた dial-up はもはや使われることはまれであろう。このような今はもう使われるのが稀な語句は除外すべきであったかもしれない。

ただ、上述のような制約や課題はあるが、今回の結果を指導者が知ることで、これくらいはカタカナ語として学生が知っているであろうという推定の精度を上げることが可能となり、語彙指導に有益な情報が得られたと考えている。

今後はこれらの語句をカタカナで示してみた時にどのような結果となるかも興味深いと考えている。カタカナ語と親密度という観点から情報科学分野の語彙を指導することは親和性が高いと思われる。

一般的な英語の高頻度語 3,000 のうち約半数はカタカナ語として用いられている (Daulton, 2008)。情報科学分野の場合、半数をはるかに上回る割合の語がカタカナ語として通用しているのではないだろうか。平井(2012)の掲載語彙でも約 500 のうち 300 をはるかに超える語句に和訳としてカタカナ語が付けられていた。

今回の調査で、カタカナ語として通用することが多い語句を英語で示した場合、調査協力者がどの程度、見聞きした覚えがあるのかを確認することができた。得られた結果によって、学習者の予備知識をある程度推定することも可能である。今後の語彙指導において手がかりの一つとして活用したい。

参考文献

- Daulton, F.E. (2008). *Japan's built-in lexicon of English-based loanwords*. Clevedon, UK: Multilingual Matters.
- 西出公之・水本篤 (2009). 「英単語 8000 語についての親密度測定の試み」『都留文科大学大学院紀要』13. 57-92.
- 平井通宏 (2012). 『キクタンサイエンス 情報科学編』, 東京: アルク.
- 横川博一(編) (2006). 『日本人英語学習者の英単語親密度 文字編』, 東京: くろしお出版.

付録

付録 親密度平均点上位の語彙一覧

Appendix Words and phrases in order of average scores of familiarity

順位	語彙	平均点	標準 偏差	順位	語彙	平均点	標準 偏差
1	internet	6.91	0.33	26	click	6.56	1.27
2	program	6.88	0.39	26	double-click	6.56	1.13
3	computer	6.83	0.67	28	sleep mode	6.54	1.18
4	online	6.81	0.77	28	wiki	6.54	1.23
5	error	6.79	0.82	30	global	6.53	1.09
6	data	6.76	0.69	30	log in	6.53	1.14
6	e-mail	6.76	0.71	32	desktop	6.52	1.28
8	personal computer	6.73	0.75	33	log out	6.51	1.25
8	download	6.73	0.91	34	icon	6.49	1.19
10	account	6.72	0.90	34	bit	6.49	1.02
11	smartphone	6.69	0.98	36	bookmark	6.47	1.20
12	website	6.68	0.90	37	database	6.46	1.03
13	password	6.65	1.18	38	version	6.45	1.35
14	install	6.64	0.98	39	chat	6.44	1.28
14	webpage	6.64	1.04	39	command	6.44	1.13
16	page	6.62	0.98	41	backup	6.42	0.96
16	software	6.62	1.02	41	home page	6.42	1.45
18	save	6.59	1.10	41	menu	6.42	1.25
18	file	6.59	1.04	41	host	6.42	1.11
20	log	6.58	0.91	45	access	6.41	1.34
21	object	6.58	0.82	46	load	6.40	1.24
22	upgrade	6.58	0.85	46	freeze	6.40	1.28
22	server	6.58	1.05	48	code	6.38	1.03
24	screen	6.57	1.14	49	check	6.36	1.28
25	upload	6.56	1.09	49	IP address	6.36	1.27

順位	語彙	平均点	標準 偏差	順位	語彙	平均点	標準 偏差
51	mouse	6.35	1.51	76	digital	6.13	1.69
52	flag	6.33	1.22	77	adapter	6.13	1.43
52	zip	6.33	1.47	78	hacker	6.12	1.54
54	source code	6.31	1.05	78	package	6.12	1.38
55	window	6.29	1.53	80	security	6.09	1.73
56	hard disk	6.29	1.30	81	channel	6.08	1.34
57	tablet computer	6.28	1.43	82	swipe	6.07	1.48
58	scanner	6.28	1.34	82	client	6.07	1.52
59	byte	6.27	1.44	84	crash	6.06	1.28
59	blog	6.27	1.41	85	supercomputer	6.04	1.70
61	memory	6.27	1.34	86	double-tap	6.03	1.47
62	help	6.26	1.56	87	portal	6.01	1.62
63	interface	6.26	1.37	87	prototype	6.01	1.58
63	tab	6.26	1.40	89	hardware	6.00	1.57
65	task	6.24	1.23	90	format	5.96	1.49
66	disk	6.24	1.37	91	template	5.95	1.72
67	screenshot	6.22	1.56	92	resource	5.93	1.60
67	font	6.22	1.40	93	touchscreen	5.91	1.71
69	drive	6.21	1.16	93	simulation	5.91	1.81
70	pointer	6.21	1.29	95	taskbar	5.91	1.73
70	operating system	6.21	1.47	96	platform	5.90	1.51
72	graphics	6.18	1.34	97	firewall	5.88	1.55
73	bug	6.17	1.59	97	tweet	5.88	1.98
74	job	6.14	1.33	99	visit	5.87	1.82
75	volume	6.13	1.44	100	router	5.86	1.73

順位	語彙	平均点	標準 偏差	順位	語彙	平均点	標準 偏差
101	head	5.86	1.59	126	decode	5.47	1.87
102	function	5.85	1.53	127	open source	5.47	1.92
103	default	5.85	1.85	128	flash memory	5.46	1.58
103	hit	5.85	1.67	129	record	5.42	1.76
105	point	5.84	1.68	130	cookie	5.38	1.98
106	spam	5.83	1.68	131	shell	5.37	1.87
106	clock	5.83	1.59	132	port	5.36	1.74
108	field	5.81	1.52	132	streaming	5.36	1.97
109	domain	5.79	1.55	132	blue screen	5.36	2.08
109	multimedia	5.79	1.81	135	packet	5.31	1.77
109	local area network	5.79	1.51	136	archive	5.30	1.99
112	export	5.76	1.53	137	multitasking	5.26	2.02
112	alphabet	5.76	2.16	138	assembler	5.24	2.04
114	ground	5.73	1.67	139	content	5.24	1.92
115	compiler	5.71	1.82	140	fault	5.21	1.96
116	flow chart	5.71	1.70	141	Internet service provider	5.19	1.78
116	register	5.71	1.63	142	filtering	5.17	2.26
118	directory	5.69	1.95	143	buffer	5.16	2.00
119	release	5.68	1.97	144	flick	5.14	1.98
120	navigate	5.64	1.78	145	macro	5.14	2.07
121	protocol	5.63	1.80	146	asterisk	5.13	2.24
122	folder	5.62	2.01	147	hub	5.11	2.13
123	debug	5.55	1.96	148	layer	5.10	2.11
124	array	5.50	1.88	149	attachment	5.10	1.92
125	clipboard	5.47	1.95	149	native	5.10	2.09

順位	語彙	平均点	標準 偏差	順位	語彙	平均点	標準 偏差
149	cyber-	5.10	2.11	176	cloud computing	4.78	2.14
152	plug-in	5.09	1.98	177	multiprocessing	4.76	2.19
152	property	5.09	2.06	178	freeware	4.75	2.14
154	avatar	5.06	2.37	179	token	4.71	2.26
155	bus	5.01	1.91	179	bits per second	4.71	2.17
156	video clip	5.01	2.11	181	kernel	4.68	2.02
157	screen saver	5.00	2.01	181	supervisor	4.68	2.07
157	patch	5.00	2.08	183	breakdown	4.66	2.14
159	toolbar	4.99	2.18	184	slash	4.64	2.40
160	domain name system	4.96	1.85	184	browser	4.64	2.38
161	function key	4.96	2.01	186	algorithm	4.64	2.47
162	activate	4.95	2.00	187	beta	4.62	2.03
163	mainframe	4.94	2.00	188	motherboard	4.56	2.24
164	cracker	4.92	2.11	188	proxy	4.56	2.04
165	troubleshooting	4.91	2.22	190	moderator	4.53	2.17
166	social gaming	4.88	2.37	191	virus	4.51	2.31
167	pop-up	4.88	2.05	192	radio button	4.49	2.21
168	search engine	4.86	2.12	192	intranet	4.49	2.35
169	application service provider	4.86	1.97	194	gateway	4.49	2.06
169	traffic	4.86	2.15	195	Ethernet	4.47	2.30
171	solution	4.83	2.20	196	prompt	4.44	2.10
172	slot	4.82	2.08	197	fragmentation	4.43	2.22
173	spyware	4.82	2.02	198	transaction	4.40	2.16
174	digital video interface	4.81	1.95	199	modem	4.38	2.27
174	locale	4.81	2.20	200	scalar	4.37	2.19

順位	語彙	平均点	標準 偏差	順位	語彙	平均点	標準 偏差
201	wide area network	4.36	2.24	226	utility	3.94	2.18
202	search box	4.35	2.00	227	spread	3.90	2.19
203	surround sound	4.35	2.29	228	wild card	3.87	2.26
204	drop-down menu	4.35	2.09	229	stand-alone	3.83	2.20
205	architecture	4.28	2.26	230	mailing list	3.81	2.41
206	mirroring	4.27	2.13	231	universal serial bus	3.76	2.23
207	error correcting code	4.26	2.10	232	failsafe	3.75	2.12
207	hard copy	4.26	2.17	233	plug and play	3.73	2.15
209	thread	4.22	2.22	234	throughput	3.71	2.10
210	surfing	4.22	2.16	235	pinch	3.70	2.10
210	idle	4.22	2.25	236	post office protocol	3.69	2.13
212	chip	4.18	2.25	237	right-click	3.69	2.47
213	malware	4.16	2.30	238	cache	3.65	2.32
214	hyperlink	4.16	2.37	239	voice clip	3.64	2.13
215	media access control	4.13	2.15	239	at sign	3.64	2.19
216	data mining	4.12	2.04	241	worm	3.62	2.18
217	dialog box	4.10	2.06	241	footprint	3.62	1.93
218	failure	4.08	2.31	241	adware	3.62	2.23
219	wizard	4.03	2.21	244	smart card	3.59	2.21
220	antivirus software	4.01	2.25	245	relational database	3.58	2.19
221	applet	3.99	2.02	246	cursor	3.56	2.46
222	broadband	3.96	2.34	247	subroutine	3.56	2.23
222	uncheck	3.96	2.11	248	webcam	3.55	2.22
224	dock	3.95	2.15	249	clip art	3.54	2.07
225	system administrator	3.95	2.15	250	spreadsheet	3.52	2.20

順位	語彙	平均点	標準 偏差	順位	語彙	平均点	標準 偏差
251	gadget	3.51	2.11	275	secure sockets layer	2.74	1.88
252	front-end	3.49	2.15	277	aspect ration	2.71	2.11
253	configuration	3.47	2.23	278	walk-through	2.67	1.86
254	data warehouse	3.46	2.05	279	handheld	2.64	1.82
255	prefix	3.45	2.07	280	resilient	2.63	2.00
256	roaming	3.41	2.14	281	ubiquitous computing	2.63	2.06
257	digit	3.38	2.02	282	bandwidth	2.60	1.87
258	boot	3.31	2.17	283	ergonomics	2.56	1.82
258	rich text format	3.31	2.05	284	thin-film transistor	2.56	1.79
260	peer-to-peer	3.28	2.36	285	disaster recovery	2.55	1.78
261	video graphics array	3.25	1.99	286	pane	2.51	1.47
262	scam	3.23	2.18	287	spoofing	2.49	1.76
263	parity	3.17	2.13	287	Trojan horse	2.49	1.88
264	thumbnail	3.16	2.52	289	eXtended graphics array	2.45	1.78
265	phishing	3.15	2.09	290	twisted pair	2.44	1.93
266	nest	3.13	2.16	291	authoring	2.32	1.77
267	solid-state	3.13	2.02	292	foolproof	2.29	1.71
267	toggle button	3.13	2.19	293	netiquette	2.26	1.72
269	shareware	3.12	2.08	294	wrap round	2.18	1.68
270	voice over IP	3.10	2.01	295	dail-up	2.12	1.47
271	procedure	3.06	1.94	296	hoax	2.01	1.63
272	legacy system	3.00	2.10				
273	podcast	2.86	2.04				
274	ripping	2.84	2.09				
275	Carrier Sense Multiple Access	2.74	1.91				