

# コーパスを活用したシノニム・語法研究における MI-score の役割 \*

井上 永幸

## 1. はじめに

世界で初めてコンピュータコーパスに依拠してつくられた *Collins COBUILD English Language Dictionary* (1987) の編集過程を詳述した Sinclair (1987) に言及するまでもなく、コーパス検索システムにおける統計値は辞書編集においてまさに救世主であり、辞書編集におけるコーパス言語学とともに発展してきた経緯がある。時間制約の多い辞書編集の現場では、執筆中の見出し語項目に関する単語を大規模コーパスで検索した際に、記述すべき単語や表現についての情報を短時間で効率的に提供することが課題とされてきたのである。McCarthy and O’Keeffe (2010) 及び Walter (2010) は以下のように述べている。

(1) McCarthy and O’Keeffe (2010), p. 7

As mentioned (see also Walter, this volume), the initial application of CL (i.e. Corpus Linguistics) in our profession was in the area of lexicography, and software and corpora were co-designed so that lexicographers could make better dictionaries.

(2) Walter (2010), p. 428

Any lexicographer working with the English language today must owe a huge debt to that evangelist of corpus lexicography, John Sinclair. While the *American Heritage Dictionary* of 1969 was the first to make use of corpus information, Sinclair and his team at Birmingham University, UK (in collaboration with Collins publishers), went far further in marrying the theory of large-scale corpus analysis with almost every aspect of the practice of lexicography in their groundbreaking *COBUILD* (Collins Birmingham University International Language Database) dictionary of 1987.

辞書編集における統計値の活用については、主として Church and Hanks (1990), Church et al. (1991), Church and Mercer (1993), Clear (1993), Church et al. (1994) などにおいて、効率的でより正確で適切な統計値を模索する議論が戦わされたが、特に、Church et al. (1994) は、MI-score (mutual information score) や *t*-score といった統計値を、シノニムの成立条件

と言える代替可能性 (substitutability) を判断する際の基準として使用することを提案していることで注目される<sup>1)</sup>。

さて、このような経緯で発達してきたコーパス言語学における統計値であるが、近年ではシノニム・語法研究に活用される場面も増えてきた<sup>2)</sup>。シノニム・語法研究では、特定の語と語の結合度の高さが問題にされることがあるが<sup>3)</sup>、その結合度の程度差を科学的で客観的な指標として明示するためにしばしば統計値が用いられるのである。辞書編集の際は、限られた時間の中でとにかく効率的に辞書項目中で言及すべき言語事実を選択するために統計値が用いられるわけであるが、シノニム・語法研究では、当該言語事実だけでなく関連語などにも目を配りながら分析・一般化を試みることになるため、できるだけ多様な類例に注目する必要がある。このように用途によって、コーパス検索システムに求められる機能や効用は当然変わってくるのである。本稿は、辞書編集における統計値としては低頻度語に敏感過ぎると批判を浴びることの多い MI-score について、シノニム・語法研究の立場から再考し、その有用性を主張するものである。コーパスとしては、公開されているコーパスの中では最も統計処理機能が充実しており大規模な WordBanks Online を利用する。

## 2. MI-score の概観

MI-score<sup>4)</sup> は特定の 2 語間について連想関係の強さを計る尺度で、予想以上に共起頻度が高い語の組み合わせで値が高くなる。共起頻度自体はそれほど高くなくても、当該単語との共起頻度が期待値以上であれば MI-score は高い値を示す。共起する 2 語が持つ意味的特性に焦点が当てられる傾向があり、慣用句、ことわざ、複合語、専門用語など比較的独自の言い回しを構成する語がリストの上位に現れることになる<sup>5)</sup>。たとえば、(3) は idea(s) の直前の位置に生起する単語の MI-score 順リストである。(3) の第 3 列目にある「共起頻度」とは特定の位置〔ここでは idea(s) の直前の位置〕における当該単語の頻度を、第 4 列目にある「頻度」とはコーパス全体に現れる当該単語の頻度を表す。

### (3) idea の直前の位置に生起する単語の MI-score 順リスト

	単語	共起頻度	頻度	MI-score		単語	共起頻度	頻度	MI-score
1	out-of-market	18	24	11.69436	26	Generate	3	39	8.40896
2	planet-formation	7	10	11.59483	27	dumbest	12	156	8.40896
3	pre-conceived	20	34	11.34387	28	half-baked	21	290	8.32181
4	direct-competition	3	6	11.1094	29	persecutory	3	42	8.30205
5	haziest	9	18	11.1094	30	dafter	3	42	8.30205
6	---Any	3	6	11.1094	31	earthshaking	3	43	8.2681
7	foggiest	37	93	10.7797	32	space-saving	5	73	8.2415
8	preconceived	131	338	10.74194	33	out-of-the-box	4	60	8.20251

9	self-oriented	4	11	10.64997	34	remotest	18	278	8.16038
10	faintest	194	661	10.34081	35	far-out	6	94	8.13977
11	FXs	12	48	10.1094	36	POTS	3	48	8.1094
12	hair-brained	3	12	10.1094	37	high-concept	3	48	8.1094
13	vaguest	35	145	10.05877	38	multiculturalist	3	49	8.07965
14	soberer	3	14	9.88701	39	19960922	4	70	7.98012
15	Godwinian	3	20	9.37244	40	daftest	3	53	7.96644
16	pie-in-the-sky	8	62	9.1552	41	slightest	198	3534	7.95167
17	over-valued	7	58	9.05877	42	keg	19	340	7.94794
18	new-fangled	15	129	9.00506	43	stupider	3	58	7.83638
19	hare-brained	8	70	8.98012	44	maddest	3	59	7.81172
20	cockamamie	4	38	8.86147	45	patentable	3	60	7.78747
21	half-formed	7	70	8.78747	46	money-saving	8	164	7.75185
22	Darwinist	6	60	8.78747	47	outworn	3	63	7.71708
23	Vygotskian	3	33	8.64997	48	harebrained	3	65	7.672
24	pseudoscientific	3	37	8.48491	49	'big	9	196	7.66462
25	crackpot	16	208	8.40896	50	testable	5	111	7.63691

※concordance size: 124818, Query: simple, attribute: word, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

このリストを一見しただけでは、idea(s) という単語とリストアップされた語との連想関係の強さを実感しにくいかもしれないが、それはまさに Baugh et al. (1996) が指摘している「低頻度語にまでも高いスコアを与えてしまう」という MI-score の特性が反映されている場合が少なくないからである。(4) を見られたい。

(4) Baugh et al. (1996), p. 45<sup>6)</sup>

A statistical index which is commonly used as an indication of the strength of a collocation is the *mutual information score*. At first the *CIDE* collocation tools used a slightly modified mutual information score (the “H-score”), but we found that this gave deceptively high values for collocates with low frequencies.

このような特性のある MI-score を使用する際に気を付けなければならないのは、MI-score のリストの上位にランクされていても、必ずしもキーワードとなる語〔ここでは idea(s)〕の典型的な連語ではなく、検索対象となったコーパスまたはそのコーパス内の特定のサンプルに特有の連語であることが多いということである<sup>7)</sup>。

コーパス検索システムのインターフェイスには、統計値リストを作成する際に各種統計値の特性を考慮したいくつかのオプションが用意されていることが多い。今回利用した WordBanks Online のプラットフォームである Sketch Engine でも、統計値リストを作成する際に最低頻度を設定することができるようになっている。(3) のリストを作成する際にも、デフォルトの

Minimum frequency in corpus: 5及び Minimum frequency in given range: 3という設定<sup>8)</sup>を生かしたリスト作成となっている。低頻度の項目ではMI-scoreの信頼度が下がるといったことを見越した最低限の配慮がされているといってもよいであろう。ただ、そのような配慮がされてあったとしても、上述のように、リストを一見しただけでは、idea(s)という単語とリストアップされた語との連想関係の強さを認識するのは難しいかもしれない。なぜなら、「共起頻度」が3以上のものに制限してあるとはいえ、この程度の制限では検索対象となったコーパスまたはそのコーパス内の特定のサンプルに特有の連語も混在してくる可能性が高いからである。試みに、(3)から、「共起頻度」が3のものだけを抜き出してみたのが(5)である。一番右側のコラムに、実際の用例を観察して、その3例が生起している状況を説明してある。

(5) idea(s)の直前の位置に生起する単語のMI-score順リストから共起頻度が3のもの

	単語	共起頻度	頻度	MI-score	説明
4	direct-competition	3	6	11.1094	3例中2例が同じソース
6	--Any	3	6	11.1094	3例とも同じソース；em dashを伴う定型の質問形式の一部
12	hair-brained	3	12	10.1094	hare(-)brainedの異形；第48位のharebrainedも参照
14	soberer	3	14	9.88701	3例ともソースは異なるが同箇所引用
15	Godwinian	3	20	9.37244	
23	Vygotskian	3	33	8.64997	3例とも同じソース
24	pseudoscientific	3	37	8.48491	3例中2例が同じソース
26	Generate	3	39	8.40896	3例中2例が同じソース
29	persecutory	3	42	8.30205	3例中2例が同じソース
30	dafter	3	42	8.30205	
31	earthshaking	3	43	8.2681	3例中2例が同じソース
36	POTS	3	48	8.1094	3例とも同じソース；ソーシャルメディアメッセージのタイトルの一部
37	high-concept	3	48	8.1094	
38	multiculturalist	3	49	8.07965	3例中2例が同じソース
40	daftest	3	53	7.96644	
43	stupider	3	58	7.83638	
44	maddest	3	59	7.81172	
45	patentable	3	60	7.78747	
47	outworn	3	63	7.71708	3例中2例が同じソース
48	harebrained	3	65	7.672	第12位のhair-brainedも参照

※concordance size: 51,777, Query: simple, attribute: word, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

(5)の中から、実際にどのような形でキーワードがヒットし抽出されているのか、いくつかサンプルのKWIC<sup>9)</sup>表示のハードコピーを示しておく。(6)を見られたい。

(6) idea(s)の直前に生起する語で高MI-scoreでありながら共起頻度が3の語の例

※行頭のdoc#XXXXXX [Xはアラビア数字]はソースとなるdoc IDを示す。

a. 第6位の ---Any<sup>10)</sup>

1	<input type="checkbox"/>	① doc#13951 > ---You'd think so, wouldn't you? </s><s> ---Any ideas ? </s><s> ---None. </s><s> She doesn't have any	
2	<input type="checkbox"/>	① doc#13951 real clotting. </s><s> It's bizarre. </s><s> ---Any ideas ? </s><s> ---I'd like to try him on the steroid-antibi	
3	<input type="checkbox"/>	① doc#13951 </s><s> She took Albert's hand. </s><s> ---Any ideas ? </s><s> ---There's no way to get out onto the low	

b. 第14位の soberer

1	<input type="checkbox"/>	① doc#13431 < into society with healthier sympathies and soberer ideas . </s><s> " James's peaceful army would perform u	
2	<input type="checkbox"/>	① doc#13818 < into society with healthier sympathies and soberer ideas ; they would tread the earth more proudly, the wom	
3	<input type="checkbox"/>	① doc#13893 < into society with healthier sympathies and soberer ideas . </s><s> " True to the Darwinian spirit, James cele	

c. 第23位の Vygotskian

1	<input type="checkbox"/>	① doc#13442 made the case earlier that many of the Vygotskian ideas overlap with behaviorist models of education as well	
2	<input type="checkbox"/>	① doc#13442 t al. </s><s> 's (1989) commitment to Vygotskian ideas about instruction within the zone of proximal develop	
3	<input type="checkbox"/>	① doc#13442 id, 1986). </s><s> Consistent with the Vygotskian ideas about scaffolding described in Chapter 8, parents see	

d. 第36位の POTS

1	<input type="checkbox"/>	① doc#17899 ath your current threshold. </s><s> Skype to POTS idea (Score: 5, Interesting) by dave1g (680091) on Satur	
2	<input type="checkbox"/>	① doc#17899 hat do you guys think? </s><s> Re: Skype to POTS idea (Score: 4, Insightful) by digitalvengeance (722523)	
3	<input type="checkbox"/>	① doc#17899 don't trust them at all. </s><s> Re: Skype to POTS idea (Score: 3, Informative) by Realistic_Dragon (655151)	

3 例中 2 例ないしは 3 例が同じソースからの場合は、MI-score が過剰評価している可能性が高くなるが、それでも、Godwinian や Vygotskian といった固有名詞派生の語や、hair-brained, soberer, pseudoscientific, Generate, persecutory, dafter, earthshaking, high-concept, multiculturalist, daftest, stupider, maddest, patentable, outworn, harebrained など、idea(s) の前に生起しても何ら不思議はない語もリストアップされている。このように、低頻度であっても idea(s) と全く無関係ではない語もリストアップされていることは無視できない事実である。辞書編集の現場では、こういった低頻度の紛らわしい例は排除してしまった方が効率は良いかもしれないが、シノニム・語法研究の観点からは、低頻度の排除は必ずしも良い結果ばかりをもたらすとは言えない。これに関しては、次節で詳しく見てゆくことにする。

このような MI-score の特性を踏まえ、筆者は MI-score のリストを利用する際は、共起頻度の高いものを優先してリストを読み解くことを推奨している<sup>11)</sup>。それに従えば、たとえば (3) からは、形容詞の最上級の形がいくつかリストアップされていることに気づく。そしてそれらの形容詞最上級には 2 つの用法が見られ、1 つは have を含む否定文で知識・理解の量や程度の少なさを強調する slightest (順位41, 頻度198), faintest (順位10, 頻度194), foggiest (順位 7, 頻度37), vaguest (順位13, 頻度35), remotest (順位34, 頻度18), haziest (順位 5, 頻度 9) [以上, 頻度順] である。(7) を見られたい。

- (7) Since the Cuban media never touches local crime, nobody had the **slightest idea** what the problem was. —doc#13471, 374068705

さらにもう1つの用法は、考えの愚かさにふれる **dumbest** (順位27, 頻度12), **daftest** (順位40, 頻度3), **maddest** (順位44, 頻度3) [以上, 頻度順] の3語である。(8)を見られたい。

- (8) That's the **dumbest idea** I ever heard. —doc#818, 63084176

特に、前者のような定型表現を弁別する際にMI-scoreは大きな威力を発揮する。*Longman Dictionary of Contemporary English* (Delacroix, L. (ed.), 2014, Sixth Edition, Harlow: Pearson Education Limited; LDOCE6) では、以下のように記述されている。

- (9) LDOCE6, s.v. IDEA, 2

**2 KNOWLEDGE** [C,U] a general understanding of something, based on some knowledge about it: . . . | not have the faintest/slightest/foggiest idea spoken: *I don't have the faintest idea what to get Rachel for her birthday.*

### 3. MI-score と他の統計値との比較

すでに述べたように、辞書編集における統計値活用とシノニム・語法研究における統計値活用では、その目的を異にする。シノニム・語法研究では、特定の言語事実の分析・一般化を試みることになるため、可能な限り多様な用例を対象に分析を試みる必要がでてくるからである。

MI-scoreは共起頻度の低い語であってもリストの上位に現れることがあるので、実際にどのように使われているのかKWIC表示を確認して、特定のコーパスソースにのみ依存する結果でないか、想定外の用法が紛れ込んでいないかなどを確認する必要がある。辞書編集の現場のような時間的な制約に縛られる状況では、低頻度語に過剰反応するMI-scoreのこういった特性が欠点としてみなされるため、MI-scoreに代わる他の統計値もいくつか提案されている。本節では、シノニム・語法研究の視点から、Sketch Engine上から利用できるMI-score及びMI3-score, logDice, log likelihoodといった統計値の比較検証を行い、指摘されている欠点があるにもかかわらずMI-scoreを利用する意味を考えてみたい。なお、本稿では、紙幅の関係からMI-score以外の3統計値については詳説は行っていないので、詳しくは注記に示した資料を参照されたい。

すでに述べたように、MI-score は 2 語間の連想関係の強さを測る尺度なので、例えば特定の動詞と強い連想関係のある副詞を調べようとするときには威力を発揮する<sup>12)</sup>。ここでは、一例として動詞 deny と連想度の高い -ly 副詞に焦点を当てて、MI-score とその他各種統計値のふるまいを観察してゆくことにする。(10) は動詞 deny の直前に生起する語の MI-score 順リストである。紙幅の都合で120位までにとどめてある<sup>13)</sup>。

## (10) deny の直前の位置に現れる単語の MI-score 順リスト

順位	単語	共起頻度	頻度	MI-score	順位	単語	共起頻度	頻度	MI-score
1	ungratefully	11	24	12.25331	61	Lipschitz	3	176	7.50437
2	strenuously	188	550	11.83014	62	Motassadeq	3	178	7.48807
3	categorically	238	957	11.37129	63	expressly	13	807	7.42286
4	vehemently	319	1294	11.35863	64	fiercely	52	3355	7.36719
5	Buthulezi	3	16	10.96381	65	ISU	6	394	7.34175
6	vigourously	3	17	10.87634	66	Rickards	3	208	7.24956
7	adamantly	64	440	10.59748	67	Navarro-Valls	4	280	7.24956
8	al-Dulaymi	3	25	10.31995	68	wrongfully	4	281	7.24442
9	flatly	143	1343	10.14747	69	blandly	4	283	7.23419
10	Ongaro	3	38	9.71588	70	Ben-Eliezer	6	448	7.15645
11	Zondi	4	51	9.70642	71	Azam	4	301	7.14522
12	steadfastly	47	639	9.61376	72	Jacko	14	1068	7.1255
13	ut-Tahrir	3	42	9.57149	73	unjustly	6	459	7.12146
14	emphatically	91	1290	9.55348	74	Ito	3	232	7.10583
15	Pechman	4	59	9.4962	75	Turley	4	311	7.09807
16	Mertel	3	47	9.40922	76	Hiss	3	235	7.08729
17	Graco	3	51	9.29138	77	Sevan	4	314	7.08422
18	stoutly	13	234	9.20892	78	Kenteris	4	332	7.0038
19	Lrc	3	57	9.13092	79	Gao	3	249	7.0038
20	hotly	65	1313	9.04256	80	Bashir	24	2029	6.97725
21	Tancred	4	83	9.0038	81	Gehring	3	260	6.94144
22	heatedly	7	150	8.95738	82	Pell	4	372	6.83968
23	MLIM	3	67	8.89772	83	Ujaama	3	279	6.83968
24	vigorously	132	2969	8.88748	84	Tsvangirai	9	838	6.83796
25	Makgoba	4	90	8.88699	85	Hann	3	292	6.77398
26	mainman	5	113	8.88059	86	Scott-Crossley	3	295	6.75924
27	Saklaoui	3	68	8.87634	87	fervently	7	693	6.74949
28	Weakland	4	95	8.80899	88	Bandar	4	412	6.69234
29	Hallman	3	74	8.75435	89	Khartoum	19	1969	6.68352
30	Cloete	3	76	8.71588	90	Goenka	3	313	6.67379
31	Chamblain	3	79	8.66003	91	Friedel	7	733	6.66853
32	Delucchi	10	268	8.63468	92	Airbus	21	2231	6.64769
33	angrily	125	3416	8.60654	93	persistently	7	749	6.63738
34	Cosby	26	759	8.51133	94	Islamabad	45	4845	6.62842
35	Boudria	4	123	8.43633	95	ETA	10	1115	6.57794
36	vociferously	6	192	8.37884	96	unlawfully	5	560	6.57149
37	Spellar	8	270	8.30203	97	initially	141	15796	6.57112
38	Sacher	3	106	8.23589	98	Libyans	4	459	6.53649
39	strongly	436	16731	8.11679	99	Pressman	3	350	6.5126

40	Habs	13	511	8.0821	100	improperly	12	1425	6.48706
41	Maskhadov	10	394	8.07872	101	Avila	3	357	6.48403
42	Cedarbaum	5	200	8.05692	102	Hallett	3	359	6.47597
43	Gopalakrishnan	3	122	8.03307	103	Jaaskelainen	3	374	6.41691
44	USOC	13	538	8.00782	104	Hekmatyar	4	502	6.4073
45	repeatedly	287	12060	7.98581	105	explicitly	19	2401	6.39735
46	al-Sahhaf	3	130	7.94144	106	Shinawatra	4	508	6.39016
47	cruelly	24	1104	7.85528	107	Radwanski	4	513	6.37603
48	consistently	146	7100	7.77506	108	summarily	3	410	6.28433
49	Bagley	3	146	7.77398	109	routinely	29	4027	6.26133
50	indignantly	10	487	7.77299	110	Sneddon	5	709	6.23113
51	furiously	38	1888	7.74413	111	Conte	6	857	6.22065
52	Wragg	5	252	7.72349	112	Sixsmith	3	434	6.20226
53	unconstitutionally	3	152	7.71588	113	Kabila	3	438	6.18902
54	Clift	3	152	7.71588	114	woodwork	10	1466	6.1831
55	staunchly	10	520	7.6784	115	implicitly	7	1033	6.17357
56	Bail	4	214	7.63738	116	Greer	8	1230	6.1144
57	callously	3	163	7.61508	117	Smitherman	3	465	6.10272
58	robustly	5	281	7.56635	118	unfairly	13	2041	6.08422
59	-also	3	170	7.55442	119	brilliantly	22	3455	6.08381
60	Vilakazi	4	229	7.53964	120	Bondies	3	473	6.07811

※concordance size: 51,777, Query: simple, attribute: lemma, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

(11) は (10) の表から -ly 副詞のみを抜粋したもので、全部で41の副詞が示されている。一番左の列はこのリストのために新たに付けた順位、その右側には抜粋前の120位まで示されていた (10) の順位を残してある。

(11) deny の直前の位置に現れる -ly 副詞の MI-score 順リスト [(10) の抜粋]

順位	元 順位	単語	共起 頻度	頻度	MI-score	順位	元 順位	単語	共起 頻度	頻度	MI-score
1	1	<b>ungratefully</b>	11	24	12.25331	21	51	furiously	38	1888	7.74413
2	2	strenuously	188	550	11.83014	22	53	<b>unconstitutionally</b>	3	152	7.71588
3	3	categorically	238	957	11.37129	23	55	staunchly	10	520	7.6784
4	4	vehemently	319	1294	11.35863	24	57	callously	3	163	7.61508
5	6	vigourously	3	17	10.87634	25	58	robustly	5	281	7.56635
6	7	adamantly	64	440	10.59748	26	63	expressly	13	807	7.42286
7	9	flatly	143	1343	10.14747	27	64	fiercely	52	3355	7.36719
8	12	steadfastly	47	639	9.61376	28	68	<b>wrongfully</b>	4	281	7.24442
9	14	emphatically	91	1290	9.55348	29	69	blandly	4	283	7.23419
10	18	stoutly	13	234	9.20892	30	73	<b>unjustly</b>	6	459	7.12146
11	20	hotly	65	1313	9.04256	31	87	fervently	7	693	6.74949
12	22	heatedly	7	150	8.95738	32	93	persistently	7	749	6.63738
13	24	vigorously	132	2969	8.88748	33	96	<b>unlawfully</b>	5	560	6.57149
14	33	angrily	125	3416	8.60654	34	97	<b>initially</b>	141	15796	6.57112
15	36	vociferously	6	192	8.37884	35	100	<b>improperly</b>	12	1425	6.48706
16	39	<i>strongly</i>	436	16731	8.11679	36	105	explicitly	19	2401	6.39735
17	45	<b>repeatedly</b>	287	12060	7.98581	37	108	<b>summarily</b>	3	410	6.28433

18	47	<b>cruelly</b>	24	1104	7.85528	38	109	routinely	29	4027	6.26133
19	48	consistently	146	7100	7.77506	39	115	implicitly	7	1033	6.17357
20	50	indignantly	10	487	7.77299	40	118	<b>unfairly</b>	13	2041	6.08422
						41	119	<b>brilliantly</b>	22	3455	6.08381

※concordance size: 51,777, Query: simple, attribute: lemma, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

この表から動詞 deny は少なくとも4種類の -ly 副詞を左側に伴う可能性が高いことがわかる。すなわち、最も多くを占めるのが strenuously [第2位], categorically [第3位], vehemently [第4位] といった様態副詞である<sup>14)</sup>。また、イタリック体で示した strongly [第16位] は、much や greatly のような程度副詞<sup>15)</sup> のひとつであるし、太字イタリック体の repeatedly [第17位], initially [第34位], summarily [第37位] は時の副詞<sup>16)</sup> である。そして注目すべきは、太字で示してある ungratefully [第1位], unconstitutionally [第22位], wrongfully [第28位] といった話者の評価を表す副詞で<sup>17)</sup>、これらは (11) においては様態副詞に次いで多く現れる。第2節で指摘したように、MI-score は共起頻度が高いものを優先的に観察するのが効率的であるため、ここでも共起頻度に注目すると、話者の評価を表す副詞では cruelly [第18位, 共起頻度24], 次いで brilliantly [第41位, 共起頻度22] が共起頻度が高いが、前者は読者からの意見を求める定型文の類似ソースが8例含まれるため [(12a) を参照], 実質は17例となる。unfairly [第40位, 共起頻度13] も doc# は異なるが全くの同例が2件抽出されているため、実質は12例となる [(12b) を参照]。ungratefully [第1位, 共起頻度11] は同じソースからの定型疑問文に Sketch Engine が反応したもので、ここではカウントから除外する [(12c) を参照]。

(12)

a. cruelly と deny の連語例で類似ソースの部分

17	<input type="checkbox"/>	① doc#17395	was you most looked forward to, and were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
18	<input type="checkbox"/>	① doc#17395	was you most looked forward to, and were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
19	<input type="checkbox"/>	① doc#17396	was you most looked forward to, and were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
20	<input type="checkbox"/>	① doc#17396	was you most looked forward to, and were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
21	<input type="checkbox"/>	① doc#17518	ures you most looked forward to, but were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
22	<input type="checkbox"/>	① doc#17521	ures you most looked forward to, but were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
23	<input type="checkbox"/>	① doc#17522	ures you most looked forward to, but were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	
24	<input type="checkbox"/>	① doc#17523	ures you most looked forward to, but were <b>cruelly</b>	<b>denied</b>	.	</s><s>	Send your suggestions for the "best" var	

b. unfairly と deny の連語例で類似ソースの部分

12	<input type="checkbox"/>	① doc#22557	Maasai believe Kikuyu farmers have been <b>unfairly</b>	<b>denying</b>	them water, the Kikuyus are equally resentful of the	
13	<input type="checkbox"/>	① doc#22560	Maasai believe Kikuyu farmers have been <b>unfairly</b>	<b>denying</b>	them water, the Kikuyus are equally resentful of the	

## c. ungratefully と deny の連語例

1	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Therein shall be damsels, refraining their
2	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Having complexions like rubies and pearl
3	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Shall the reward of good works be any o
4	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> And besides these there shall be two oth
5	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Of a dark green color. </s><s> Which, t
6	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> In each of them shall be two fountains p
7	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> In each of them shall be fruits, and palm
8	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Therein shall be agreeable and beauteou
9	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Having fine black eyes; and kept in pavil
10	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Whom no man shall have touched before
11	<input type="checkbox"/>	① doc#13438 erefore, of your Lord's benefits will ye ungratefully deny ? </s><s> Blessed be the name of thy Lord, posses

このように、共起頻度の少ないものに関しては検証の手間はかかるが<sup>18)</sup>、シノニム・語法研究ではこの確認が欠かせない。果たして、MI-score においては、話者の評価を表す副詞が 8 語確認できることになる。下に、brilliantly, unfairly, improperly の例を挙げておく<sup>19)</sup>。

(13) Baros could have added a third but was **brilliantly denied** by Monaco keeper Flavio Roma. —doc#401, 30460543

(14) Further, they said state law **unfairly denies** them specific rights guaranteed to opposite-sex couples by the state constitution. —doc#18214, 459154528

(15) After the 2000 election, it emerged that thousands of people, mostly blacks, were **improperly denied** the right to vote because they mistakenly appeared on a list of ex-felons. —doc#12443, 270689027

次に MI3-score を見てゆこう<sup>20)</sup>。deny の直前の位置に現れる -ly 副詞について、MI3-score 順に整列させたリストが (16) である。

## (16) deny の直前の位置に現れる -ly 副詞の MI3-score 順リスト

順位	元順位	単語	共起頻度	頻度	MI3-score	順位	元順位	単語	共起頻度	頻度	MI3-score
1	1	vehemently	319	1294	27.99346	19	45	previously	77	24613	17.59207
2	3	categorically	238	957	27.16092	20	47	publicly	59	12384	17.43058
3	4	strenuously	188	550	26.93932	21	50	<b>cruelly</b>	24	1104	17.02521
4	7	strongly	436	16731	25.65316	22	55	stoutly	13	234	16.6098
5	9	flatly	143	1343	24.46721	23	61	routinely	29	4027	15.9773
6	10	repeatedly	287	12060	24.31562	24	77	subsequently	29	7089	15.16142
7	13	vigorously	132	2969	22.97626	25	78	specifically	34	11587	15.14101
8	14	adamantly	64	440	22.59748	26	79	officially	33	10873	15.10356
9	15	emphatically	91	1290	22.56907	27	81	<b>brilliantly</b>	22	3455	15.00267

10	16	angrily	125	3416	22.5381	28	85	explicitly	19	2401	14.89321
11	17	consistently	146	7100	22.15471	29	86	expressly	13	807	14.82374
12	23	hotly	65	1313	21.0873	30	89	effectively	34	16008	14.67473
13	26	initially	141	15796	20.85023	31	92	heatedly	7	150	14.57209
14	27	steadfastly	47	639	20.72294	32	95	indignantly	10	487	14.41685
15	35	<del>ungratefully</del>	11	24	19.17218	33	96	staunchly	10	520	14.32226
16	38	fiercely	52	3355	18.76807	34	107	vigourously	3	17	14.04627
17	40	furiouly	38	1888	18.23998	35	117	improperly	12	1425	13.65699
18	43	firmlly	62	11201	17.79009	36	119	formally	20	6887	13.59497
						37	120	vociferously	6	192	13.54877

※concordance size: 51,777, Query: simple, attribute: lemma, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

(16) には、(11) の MI-score の表に現れていた cruelly [第21位, 実質共起頻度24], brilliantly [第27位], improperly [第35位] の他に effectively [第30位, 共起頻度34] が現れている<sup>21)</sup>。第15位の ungratefully はすでに述べたようにカウントから除外する。すなわち、MI3-score のリストでは、話者の評価を表す副詞が4語〔太字〕確認できたことになる。下に、effectively と cruelly の例を挙げておく。

(17) A newsreader later **effectively denied** that anything more sinister had taken place.

—doc#8718, 197536003

(18) She is a bright girl who loves learning and has been **cruelly denied** this right by the bullying.

—doc#473, 35582964

続けて logDice を見てゆこう<sup>22)</sup>。(19) は deny の直前の位置に生起する -ly 副詞の logDice 順リストである。

(19) deny の直前の位置に生起する -ly 副詞の logDice 順リスト

順位	元順位	単語	共起頻度	頻度	logDice	順位	元順位	単語	共起頻度	頻度	logDice
1	1	strongly	436	16731	7.7042	19	32	furiouly	38	1888	4.53623
2	2	vehemently	319	1294	7.62178	20	40	specifically	34	11587	4.13609
3	3	categorically	238	957	7.20837	21	42	officially	33	10873	4.10937
4	4	repeatedly	287	12060	7.2028	22	43	routinely	29	4027	4.0899
5	5	strenuously	188	550	6.87932	23	45	effectively	34	16008	4.03878
6	6	flatly	143	1343	6.4629	24	50	subsequently	29	7089	4.01283
7	7	consistently	146	7100	6.34441	25	52	cruelly	24	1104	3.8945
8	8	vigorously	132	2969	6.30393	26	60	brilliantly	22	3455	3.70621
9	10	angrily	125	3416	6.21359	27	61	authority	52	80346	3.68892
10	11	initially	141	15796	6.09539	28	66	totally	28	23362	3.61008

11	12	emphatically	91	1290	5.81227	29	77	explicitly	19	2401	3.52251
12	17	adamantly	64	440	5.32777	30	82	formally	20	6887	3.48174
13	18	hotly	65	1313	5.32622	31	95	continually	17	4469	3.308
14	22	previously	77	24613	5.04569	32	96	originally	20	15864	3.27632
15	23	firmly	62	11201	5.01164	33	101	immediately	28	45937	3.23108
16	24	fiercely	52	3355	4.94984	34	108	promptly	15	4225	3.1337
17	25	publicly	59	12384	4.91323	35	109	simply	33	71537	3.13242
18	26	steadfastly	47	639	4.87687	36	117	stoutly	13	234	3.03391
						37	119	absolutely	20	28418	3.0307

※concordance size: 51,777, Query: simple, attribute: lemma, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

(19) のリストには, (11) の MI-score のリストにも見られた *cruelly* [第25位], *brilliantly* [第26位] のほか, MI3-score の (16) にも現れていた *effectively* [第23位] が見られ, 話者の評価を表す副詞3語 [太字] が確認できる。

さらに, log likelihood に目を向けてみよう<sup>23)</sup>。(20) は *deny* の直前の位置に生起する *-ly* 副詞の log likelihood 順リストである。

(20) *deny* の直前の位置に現れる *-ly* 副詞の log likelihood 順リスト

順位	元順位	単語	共起頻度	頻度	log likelihood	順位	元順位	単語	共起頻度	頻度	log likelihood
1	5	vehemently	319	1294	4473.297	23	84	specifically	34	11587	168.3617
2	7	strongly	436	16731	4052.164	24	88	officially	33	10873	165.5699
3	8	categorically	238	957	3341.865	25	91	subsequently	29	7089	162.3812
4	10	strenuously	188	550	2781.133	26	95	effectively	34	16008	147.1857
5	12	repeatedly	287	12060	2613.962	27	98	brilliantly	22	3455	142.3409
6	15	flatly	143	1343	1742.053	28	100	stoutly	13	234	140.7423
7	19	vigorously	132	2969	1369.153	29	106	explicitly	19	2401	131.1085
8	21	consistently	146	7100	1286.402	30	117	expressly	13	807	108.1363
9	22	angrily	125	3416	1246.948	31	126	formally	20	6887	98.62873
10	24	emphatically	91	1290	1030.159	32	133	continually	17	4469	92.76724
11	26	initially	141	15796	1007.019	33	137	totally	28	23362	91.05117
12	30	adamantly	64	440	822.185	34	139	indignantly	10	487	88.05481
13	34	hotly	65	1313	688.4052	35	140	staunchly	10	520	86.73651
14	41	steadfastly	47	639	536.0913	36	146	<b>improperly</b>	12	1425	84.28507
15	47	fiercely	52	3355	428.5626	37	148	<b>unfairly</b>	13	2041	84.11576
16	50	previously	77	24613	390.9236	38	152	promptly	15	4225	79.83145
17	51	firmly	62	11201	384.0407	39	173	originally	20	15864	66.90411
18	53	publicly	59	12384	348.0328	40	185	vociferously	6	192	57.91837
19	55	furiously	38	1888	333.0994	41	186	immediately	28	45937	57.40649
20	70	<b>cruelly</b>	24	1104	214.093	42	191	simply	33	71537	52.52959
21	74	routinely	29	4027	194.6974	43	195	fervently	7	693	51.69811
22	82	<b>ungratefully</b>	11	24	170.9177	44	200	persistently	7	749	50.61539

※concordance size: 51,777, Query: simple, attribute: lemma, Minimum frequency in corpus: 5, Minimum frequency in given range: 3

(20) には, (11) の MI-score リストにも見られた cruelly [第20位], brilliantly [第27位], improperly [第36位], unfairly [第37位] の他, (19) の logDice と同様に effectively [第26位] が見られ, 話者の評価を表す副詞は全部で6語 [太字] が確認できる。また, (11) と同様に, 第22位の ungratefully はカウントから除外する。

以上見てきたように, 動詞 deny と最も一般的に用いられる様態の副詞については, どの統計値でもその連想度の高さを一定程度示してくれるが, deny と特徴的に生起する話者の評価を表す副詞については MI-score が最も効果的に顕在化させてくれることが分かった。

#### 4. おわりに

低頻度語に関して過剰反応すると批判される MI-score をめぐって, シノニム・語法研究における利用価値の再考を試みた。コーパス検索システムは辞書編集を主な目的として開発・進歩してきた経緯があり, 厳しい時間制約のある状況で可能な限り統計値の信頼性を高いものにするということを最優先にして議論が進められる。一方, シノニム・語法研究においては, コーパス基盤的な (corpus-based) アプローチかコーパス駆動的な (corpus-driven) アプローチのいずれにせよ, 時間的制約よりも言語分析の際の新たな発見の可能性が優先されるべきであろう。低頻度語は, コーパスに含まれるかどうかはコーパス構築の際に特定のソースとして選ばれたテキストに特有の連語であるなど, 偶然による可能性が高くなるといったことがある程度想定されるが, 偶然であるかどうかは KWIC 表示を確認することで判断できるし, 低頻度語に過剰反応するという発想ではなく, 低頻度語にも反応してくれるので検証できる範囲が広がるととらえるべきであろう。シノニム・語法研究において統計値は新発見への単なる足がかりであり, 統計値だけで判断せず実際の KWIC 表示を確認しながら, 慎重に分析を進めてゆく必要がある。

(21) に示すように, Church and Mercer (1993) は, Church et al. (1991) を引用しながら, 各種統計値には長所短所があるのでそれらを把握し適材適所で活用することが重要であることを指摘して, 統計値を妄信することを戒めている。

(21) Church and Mercer (1993: 20) [下線は筆者]

Summary statistics such as mutual information and *t*-scores may have an important role to play in helping lexicographers to discover significant patterns of collocations, though the position remains somewhat controversial. Some lexicographers prefer mutual information, some prefer *t*-scores, and some are unconvinced that either of them is any good. Church et al. (1991) argued that different statistics have different

strengths and weaknesses, and that it requires human judgment and exploration to decide which statistic is best for a particular problem. . . .

各種統計値の特徴を踏まえて活用すれば、一見欠点と思われる特徴も有効活用できることを示した。特に、MI-score のような癖の強い統計値については、入門書の記述だけに惑わされず、自分のなじみのある語でいろいろと試行錯誤の上、その特徴をつかむことが肝要である。

## 注

\* 本稿は、筆者が、「英語コーパス学会第44回大会ワークショップ2」（2018年10月6日、東京理科大学神楽坂キャンパス）において、「コーパス研究の作法－語法研究・辞書編集を通して－」と題して発表した内容の一部に加筆したものである。

- 1) 例えば、本稿で取り上げる MI-score は、言語研究者らを想定した原則無料のコーパスサイトとしては最大規模を誇る English-Corpora.org [https://www.english-corpora.org/mutualInformation.asp] や、有料の WordBanks Online [https://wordbanks.harpercollins.co.uk/]、Sketch Engine [https://www.sketchengine.eu/] はもちろん、WordSmith Tools [https://www.lexically.net/wordsmith/] といったパソコン上で動作する検索システムにおいてまでも利用可能である。
- 2) 井上 (2010)、井上 (2016)、井上 (2018) を参照。
- 3) 語と語の結合度 (collocability) に程度の差があることは、“collocation” を初めて専門用語として用いた Firth (1951: 194-95) ですでに指摘されている。
- 4) MI-score について、Church and Hanks (1990: 24) は「3 より大きい場合にその語の組み合わせが興味深いものとなる傾向がある」と述べ、Barnbrook (1996: 99) は「厳密な統計的有意な数値は明らかではないが、1.58を打ち切り点とすることができる」旨述べている。MI-score の考え方はそれほど新しいものではなく、情報理論の多くの教科書で扱われているという (Church et al. 1994: 159, Fano 1961: 28)。以下、Church and Hanks (1990: 23) より引用しておく：

What is “mutual information?” According to Fano (1961), if two points (words),  $x$  and  $y$ , have probabilities  $P(x)$  and  $P(y)$ , then their mutual information,  $I(x, y)$ , is defined to be

$$I(x, y) \equiv \log_2 \frac{P(x, y)}{P(x)P(y)}$$

Informally, mutual information compares the probability of observing  $x$  and  $y$  *together* (the joint probability) with the probabilities of observing  $x$  and  $y$  *independently* (chance). . . .

MI-score の詳細は、Church and Hanks (1990)、Church et al. (1991)、Church and Mercer (1993)、Clear (1993)、Church et al. (1994)、井上 (1994)、Biber et al. (1998: 266)、Oakes (1998: 63-65)、Ooi (1998: 77-79)、Hunston (2002: 70-71)、園田・高見 (2005)、井上 (2005) を参照。

- 5) Clear (1993: 282) を参照。
- 6) 引用文中にある *CIDE* /sáidi:/ とは、*Cambridge International Dictionary of English* (Procter, P. (ed.), 1995, Cambridge: Cambridge University Press) の略称。
- 7) Clear (1993: 282) は以下のように述べている：However, many of the collocates listed by the MI

score will be untypical and peculiar to the corpus or to particular text samples within the corpus.

- 8) “Minimum frequency in corpus: 5” はコーパスにおける頻度〔すなわち「頻度」の項〕が5以上のもの, “Minimum frequency in given range: 3” は指定範囲の共起頻度〔すなわち「共起頻度」の項〕が3以上のものという設定となる。
- 9) *Key Word In Context* の略。キーワードがどのような文脈で用いられているのかを, キーワードを中心に配置して見やすい形で表示したもの。
- 10) em dash なしの Any は109位で, 頻度293, MI-score は6.17294。
- 11) cf. 井上 (2016)。
- 12) cf. 井上 (2016)。
- 13) 当初, MI-score が6以上となる130までを表示予定だったが, 121位から130以内に -ly 副詞が現れないため, 120位までとした。
- 14) Quirk et al. (1985: 503) は adverbial を adjunct, subjunct, disjunct, conjunct の4種類に分類している。strenuously, categorically, vehemently といった副詞は, subjunct > narrow orientation > item > subject orientation > general といった下位分類をされることとなる。Quirk et al. (1985: 591) は関連する主語指向の副詞は様態副詞がほとんどを占めることを指摘している。
- 15) Quirk et al. (1985: 591) は strongly のような副詞を, subjunct > narrow orientation > intensifiers > amplifiers > boosters に分類している。
- 16) Quirk et al. (1985: 530-531, 543) では, time-position adjunct となる。
- 17) Quirk et al. (1985: 615) では, disjunct > content > value judgement of content に下位分類され, この型を Type (b) と呼び, 以下のような特徴について触れている [Quirk et al. (1985), p. 624]:  

[*Fortunately, / Wisely*] she consults her lawyer regularly.  
 = It is {*fortunate / wise*} that she {consult(s) / should consult} her lawyer regularly.

Most of the adverbs of Type (b) allow also a correspondence with an infinitive clause (commonly a replacement for a *that*-clause with putative *should*; cf 15.48f):  
*Foolishly*, Bill declined the invitation.  
 = It was *foolish* of Bill to decline the invitation.

なお, [ ] は contingent alternatives を, { } は free alternatives を表す [Quirk et al. (1985), p. x]。
- 18) Hoffmann et al. (2008: 154) では, 統計値のみから判断することを前提に, “Freq (node, collocater)” [最低共起頻度を指定] の設定項目は10以上に設定することを推奨している。母語話者のみによる辞書編集の現場などでは共起数10以上の統計値のみによる判断でよいかもしれないが, シノニム・語法研究の立場から幅広くデータを得た方がよいので, デフォルト設定の Minimum frequency in corpus: 5, Minimum frequency in given range: 3のままにしておくのが賢明であろう。また, 仮に10に設定しても, (11), (16), (20) に現れた ungratefully のように, 除外されるべき語が排除できるわけでもないことは心に留めておく必要がある。
- 19) 用例中の太字は筆者によるもの。以下, 同様。
- 20) MI3について詳しくは Oakes (1998: 171-172), *Lexical Computing* (2015: 4) を参照。

- 21) effectively は MI-score のリストでは234位 [MI-score は4.4998]。
- 22) logDice について詳しくは, Rychlý (2008), 恒川 (2020) を参照。また, Sketch Engine の以下のサイトにも解説が見られる。 [https://www.sketchengine.eu/my\\_keywords/logdice/](https://www.sketchengine.eu/my_keywords/logdice/)
- 23) log likelihood について詳しくは, Dunning (1993), Oakes (1998: 42, 172, 189), Hoffmann (2008: 84-85, 152-154), 堀田 (2011) を参照。

## 参考文献

- Barnbrook, G. (1996) *Language and Computers: A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.
- Baugh, S., A. Harley & S. Jellis (1996) "The Role of Corpora in Compiling the Cambridge International Dictionary of English," *International Journal of Corpus Linguistics*, 1:1, pp. 39-59.
- Biber, D., S. Conrad and R. Reppen (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Brezina, V. (2018) *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.
- Church, K., W. Gale, P. Hanks and D. Hindle (1991) "Using statistics in lexical analysis," U. Zernik (ed.) *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc. pp. 115-164.
- Church, K. W., W. Gale, P. Hanks, D. Hindle and R. Moon (1994) "Lexical Substitutability," in Atkins, B. T. S. and A. Zampolli (eds.) *Computational Approaches to the Lexicon*. Oxford: Oxford University Press. pp. 153-177.
- Church, K. W. and P. Hanks (1990) "Word Association Norms, Mutual Information and Lexicography," *Computational Linguistics*, 16: 1, pp. 22-29.
- Church, K. W. and R. L. Mercer (1993) "Introduction to the Special Issue on Computational Linguistics Using Large Corpora," in *Computational Linguistics*, 19:1, pp. 1-24.
- Clear, J. (1993) "From Firth Principles: Computational Tools for the Study of Collocation," in Baker, M., G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins. pp. 271-292.
- Dunning, T. (1993) "Accurate Methods for the Statistics of Surprise and Coincidence," *Computational Linguistics* 19:1, pp. 61-74.
- Fano, R. (1961) *Transmission of Information*. Cambridge, Mass.: MIT Press.
- Firth, J. R. (1951) "Modes of Meaning," in Firth, J. R. *Papers in Linguistics: 1934-1951*. London: Oxford University Press. pp. 190-215.
- Greenbaum, S. (1969) *Studies in English Adverbial Usage*. London: Longman.
- Hoffmann, S., S. Evert, N. Smith, D. Lee and Y. B. Prytz (2008) *Corpus Linguistics with BNCweb—a Practical Guide*. Frankfurt am Main: Peter Lang.
- 堀田隆一 (2011) 「#696. Log-Likelihood Test[corpus][bnc][statistics][lltest]」 hellog～英語史ブログ。

- <http://user.keio.ac.jp/~rhotta/hellog/2011-03-24-1.html>
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- 井上永幸 (1994) 「Cobuild Corpus: The Bank of English とは何か」『英語教育と英語研究』第11号 (島根大学教育学部英語科教育研究室) pp.31-51.
- 井上永幸 (2005) 「コーパスに基づく辞書編集」, 齊藤等, pp.207-228.
- 井上永幸 (2010) 「辞書編集におけるコーパス活用」『英語語法文法研究』第17号 (英語語法文法学会), pp.5-22.
- 井上永幸 (2016) 「語法研究の要一 副詞一コーパスを活用した辞書編集の立場から一」『英語語法文法研究』第23号 (英語語法文法学会), pp.20-35.
- 井上永幸 (2018) 「コーパス研究の作法一語法研究・辞書編集を通して一」英語コーパス学会第44回大会ワークショップ2, 2018年10月6日 (土), 東京理科大学神楽坂キャンパス.
- Kilgarriff, A., P. Rychlý, P. Smrž and D. Tugwell (2004) "The Sketch Engine," *Proceedings of EURALEX 2004*, pp. 105-116.
- Lexical Computing Ltd. (2015) "Statistics used in the Sketch Engine."  
<https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf>
- McCarthy, M. and A. O'Keeffe (2010) "Historical perspective: What are corpora and how have they evolved?," in O'Keeffe, Anne and Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, pp. 3-13.
- Oakes, M. P. (1998) *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh Univ. Press.
- Ooi, V. B. Y. (1998) *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Rychlý, P. (2008) "A Lexicographer-Friendly Association Score," in *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*, pp. 6-9.
- 齊藤俊雄・中村純作・赤野一郎 (編) 『英語コーパス言語学ー基礎と実践ー【改訂新版】』研究社出版.
- Sinclair, J. M. (ed.) (1987) *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London and Glasgow: Collins ELT.
- 園田勝英・高見敏子 (2005) 「コーパスに基づく語彙研究」, 齊藤等, pp.121-143.
- Thomas, J. (2016) *Discovering English with Sketch Engine*. Second Edition, Updated. Versatile.  
<http://versatile.pub>
- 恒川元行 (2020) 「logDice 係数はどのような共起指標か」『言語文化論究』45 (九州大学大学院言語文化研究院), pp.35-44.
- Walter, E. (2010) "Using Corpora to Write Dictionaries," in O'Keeffe, Anne and Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, pp. 428-443.
- Zernik, U. (ed.) (1991) *Lexical Acquisition: Using On-line Resources to Build a Lexicon*. Englewood Cliff, NJ: Lawrence Erlbaum.

井上永幸

## 辞書

Delacroix, L. (ed.) (2014) *Longman Dictionary of Contemporary English*. Sixth Edition. Harlow: Pearson Education Limited. [LDOCE6]

Procter, P. (ed.) (1995) *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press. [CIDE]

Sinclair, J. (ed.) (1987) *Collins COBUILD English Language Dictionary*. London & Glasgow: Collins ELT.

## コーパス

WordBanks Online. (tokens: 551,531,292; words: 462,302,063; 2023/8/24). Glasgow: HarperCollins Publishers. <https://wordbanks.harpercollins.co.uk/>

## The Role of MI-score in Studies of English Synonyms and Usage Using Corpora

INOUE Nagayuki

The strategy on the statistics in corpus linguistics has been developing with the evolution of lexicography. This article is intended to show that the MI-score, which is often criticized for having a tendency to overestimate low-frequency pairs that are too specific to a particular corpus and may not generalize very well, is one of the most appropriate solutions when you try to identify interesting associations among words in corpora, especially when you would like to obtain interesting adverbials associated with verbs and adjectives.