

关键词分析与对数似然比 (LL) 计算结果

下面列出了对数似然比 (LL) 计算公式，以及性别、种族、地域各领域中 LL 值显著高于临界值（95%置信水平下为 3.84， $p<.05$ ）的关键词。

Table 1 对数似然比 (LL) 计算方法 (Rayson et al. 2000; 2008)

	目标语料库 Target Corpus (i*)	参照语料库 Reference Corpus (i*)	总计 Total
目标词频率 Frequency of Target Word (O*)	a	b	a+b
其他词频率 Frequency of other words	c-a	d-b	c+d-a-b
Total (*N)	c	d	c+d

$$LL_i = -2 \ln \lambda = 2 \sum_i O_i \ln \left(\frac{O_i}{E_i} \right), \text{ where } E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

Table 2 性别仇恨言论的关键词统计

Type (汉语)	性别领域 词频 (Target)	全领域 词频 (Reference)	Keyness (LL值)	Type (韩语)	性别领域 词频 (Target)	全领域 词频 (Reference)	Keyness (LL 值)
强奸	385	535	357.10	한남	462	613	371.32
恶心	838	2303	245.64	페미	341	499	241.49
出轨	195	206	236.44	맘충	194	254	158.06
男女	170	238	156.22	보지	224	378	128.56
直男癌	151	221	131.86	한남충	133	184	101.18
女权	278	657	115.39	냄저	100	132	80.61
家暴	109	140	109.87	자지	142	260	71.01
重男轻女	117	199	84.42	김치녀	82	113	62.66
渣男	59	73	61.75	이기(야)	225	541	62.57
同性恋	66	125	40.69	메갈	66	96	47.00
性骚扰	35	42	37.75	김치년	64	93	45.63
渣	41	65	32.46	자대이	61	88	43.92
女拳	63	133	32.43	한녀	45	51	43.01
gay	37	54	32.40	부랄	74	128	40.74
垃圾	119	353	28.02	(이)노	122	278	38.82
嫖娼	28	37	27.39	피싸개	38	46	33.84
男同	28	37	27.39	남협	39	50	32.53
活该	66	156	27.36	재기	95	220	29.05
基佬	14	20	12.57	여협	53	93	28.49
绿茶	6	7	11.71	소추	45	72	27.99

Table 3 种族与国籍仇恨言论的关键词统计

Type (汉语)	种族领域 词频 (Target)	全领域 词频 (Reference)	Keyness (LL值)	Type (韩语)	种族领域 词频 (Target)	全领域 词频 (Reference)	Keyness (LL 值)
黑人	6610	6697	1690.32	조선족	222	271	371.17
中国	2661	3305	377.26	짱깨	224	284	364.48
白人	1332	1360	330.15	중국	160	259	213.68
歧视	3638	5069	327.04	다문화	112	157	168.64
美国	1401	1526	295.01	난민	97	149	135.47
亚裔	970	975	249.24	무슬림	98	197	105.79
黄种人	377	379	96.70	한국	154	487	89.51
华人	336	350	79.14	똥남아	55	77	82.84
他们	1149	1703	75.11	왜구	53	77	77.49
印度人	288	299	68.38	탈북자	57	100	70.59
种族	275	286	65.02	외노자	43	54	70.30
非洲	252	256	63.05	북한	57	112	63.00
国家	436	544	60.47	중국인	38	50	60.02
教育	416	515	59.29	필리핀	37	49	58.16
犯罪率	245	260	55.06	흑인	33	49	47.38
奴隶	221	236	48.87	개슬람	74	242	40.46
日本	370	470	48.12	중동	27	39	39.65
政治	240	267	47.57	개새끼	47	122	37.08
反对	253	307	38.56	조선인	15	24	20.20
黑命贵	128	128	33.22	민족	21	47	20.05

Table 4 地域仇恨言论的关键词统计

Type (汉语)	地域领域 词频 (Target)	全领域 词频 (Reference)	Keyness (LL值)	Type (韩语)	地域领域 词频 (Tar get)	全领域 词频 (Referenc e)	Keyness (LL值)
河南人	1158	1190	654.18	전라도	471	602	1047.84
河南	743	762	420.43	홍어	383	490	850.45
北京	674	700	374.42	쌍도	216	280	474.79
东北人	654	670	370.55	경상도	119	143	272.64
井盖	550	558	315.88	라도	113	208	196.74
黑	927	1306	298.83	대구	78	104	168.29
偷	474	498	258.74	서울	62	106	113.81
外地人	441	452	249.54	새끼	358	2373	95.49
台湾	404	425	220.07	절라도	29	34	67.26
台湾人	318	325	180.61	명청도	27	30	64.43
四川人	294	294	172.18	촌놈	26	30	60.82
山东人	275	276	160.23	머구	29	41	60.34
新疆	275	284	153.85	사투리	25	31	56.21
素质	496	764	129.27	지역	34	67	56.10
骗子	300	381	119.24	부심	25	37	50.54
排外	213	225	115.25	통수	21	40	35.57
香港	221	241	113.95	설라도	15	19	33.31
mmp	54	56	30.01	통구이	20	40	32.60