

The Academic Vocabulary List

Mark Davies and Dee Gardner, Brigham Young University

Sample of core academic words. Complete lists and detailed information at: www.academicwords.info.

rank	word	PoS	frequency	ratio	dispersion	range
39	value	n	49,900	1.97	0.96	9
41	role	n	49,426	1.81	0.95	8
42	difference	n	49,081	1.93	0.93	8
43	analysis	n	48,500	2.85	0.93	8
44	practice	n	47,915	2.06	0.95	8
46	thus	r	46,287	2.69	0.95	8
47	control	n	45,690	1.64	0.93	9
48	form	n	45,275	1.90	0.95	8
49	report	v	44,666	1.64	0.92	8
51	significant	j	44,415	2.44	0.93	8
52	figure	n	44,084	1.92	0.92	8
53	factor	n	43,871	2.43	0.92	8
54	interest	n	43,498	1.51	0.95	9
56	need	n	42,193	1.82	0.95	9
57	base	v	42,082	1.80	0.95	9

In order to be included in the list, the word must meet the four criteria shown below, and the words are then listed by frequency in COCA-Academic.

Columns	Limit	Explanation
word		The word; actually the lemma (e.g. [report] = {report, reported}, etc)
PoS		Part of speech (<u>n</u> oun, <u>v</u> erb, <u>a</u> djective, <u>a</u> dverb, etc)
frequency		Frequency in COCA Academic (120 million words)
ratio	1.50	A ratio showing how much more frequent the word is in Academic than the rest of COCA (per million words); e.g. 1.50 = 50% more frequent
dispersion	0.80	How "evenly" the word is spread across Academic (0-1). This is the Juilland "d" measure.
range	7	In how many of the nine domains does the word occur with at least 20% of the expected frequency (based on the size of that domain)
	Not greater than 3.0	The word (lemma) can also not occur at more than three times the expected frequency in any domain. For example, the word <i>federal</i> occurs at 3.69 the expected frequency in Law/PolSci, so it is not included as a "core academic" word.
domains		The spreadsheet also includes the raw frequency of the lemma in each of the nine domains, as well as a ratio of frequency to expected frequency (based on the size of the domain, in words). These are not shown here, for reasons of space.