Language Concreteness

What is language concreteness?

Brysbaert et al. (2014) argue that concreteness measures the degree of a concept denoted by a word that refers to a perceptible entity. In contrast to abstract words, concrete words are psycholinguistically easier for people to remember. In psychological research, concreteness has been an important and popular variable due to the availability of norms of words.

Previous research related to concreteness measurements includes the MRC database (Coltheart, 1981), English Lexicon Project (Balota et al., 2007), and the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012). But they are limited in word samples, word choices, and measurements in the concreteness ratings. For example, in measuring the adjective, five senses including touch, hearing, seeing, smelling, and tasting are not equivalently presented in the concreteness rating. So to improve the existing limitations from previous concreteness measuring approaches and scales, Brysbaert et al. collected new ratings for larger numbers of stimuli and developed a more standard word list for future studies. In the experiment, participants used a 5-point rating scale from abstract (language-based) to concreteness (experience-based) to measure their feelings of words. In total, they measured about 40 thousand (37,058 words and 2,896 two-word expressions) generally known English word lemmas.

What do concreteness results tell us?

This file is the result of Brysbaert et al. research, which includes about 40 thousand words with their concreteness.

	Word	Bigram	Conc.M	Conc.SD	Unknown	Total	Percent_i	SUBTLEX	Dom_Pos
1	Word	Bigram	Conc.M	Conc.SD	Unknown	Total	Percent_i	SUBTLEX	Dom_Pos
2	roadsweeper	0	4.85	0.37	1	27	0.96	0	0
3	traindriver	0	4.54	0.71	3	29	0.9	0	0
4	tush	0	4.45	1.01	3	25	0.88	66	0
5	hairdress	0	3.93	1.28	0	29	1	1	0
6	pharmaceutics	0	3.77	1.41	4	26	0.85	0	0
7	hoover	0	3.76	1.23	4	29	0.86	162	0
8	shopkeeping	0	3.18	1.19	1	29	0.97	0	0
9	pushiness	0	2.48	1.24	1	30	0.97	0	0
10	underdevelop	0	2.37	1.4	0	30	1	0	0
11	tirelessness	0	2.28	1.28	1	30	0.97	0	0
12	oldfashioned	0	2.26	1.02	0	27	1	0	0
13	wellmannered	0	2.25	1.14	2	30	0.93	0	0
14	dismissiveness	0	1.83	1	1	30	0.97	0	0
15	spitefulness	0	1.8	0.76	0	25	1	0	0
16	untruthfulness	0	1.73	0.92	4	30	0.87	0	0
17	dispiritedness	0	1.56	0.71	3	28	0.89	0	0
18	sled	0	5	0	0	28	1	149	Adjective
19	plunger	0	4.96	0.2	0	26	1	48	Adjective
20	human	0	4.93	0.26	0	28	1	6363	Adjective
21	waterbed	0	4.93	0.27	0	27	1	27	Adjective
22	cymbal	0	4.92	0.28	1	25	0.96	13	Adjective
23	ginger	0	4.92	0.27	0	26	1	327	Adjective
24	bobsled	0	4.9	0.41	0	29	1	48	Adjective
25	cardboard	0	4.9	0.41	0	29	1	138	Adjective
26	olive	0	4.9	0.31	0	30	1	375	Adjective
27	dogsled	0	4.89	0.32	0	27	1	2	Adjective
28	rubber	0	4.86	0.74	0	29	1	714	Adjective
29	soybean	0	4.82	0.77	0	28	1	25	Adjective
30	tangerine	0	4.81	0.62	0	27	1	38	Adjective
31	headrest	0	4.8	0.76	0	30	1	6	Adjective
32	eucalyptus	0	4.77	0.57	0	30	1	25	Adjective
33	saltwater	0	4.77	0.82	0	30	1	31	Adjective

NLP Suite → lib → concretenessLib → Concreteness_ratings_Brysbaert_et_al_BRM.csv

Word: the word

Bigram: whether it is a single word or a two-word expression

Conc.M: the mean concreteness rating

Conc.SD: the standard deviation of the concreteness ratings

Unknown: the number of participants indicating they did not know the word

Total: the total number of participants who rated the word

Percent: percentage participants who knew the word

SUBTLEX: the SUBTLEX-US frequency count (on a total of 51 million; Brysbaert & New, 2009)

Dom_Pos: the POS tagging (adjective, adverb, article, conjunction, determiner...)

How to understand the output file

To calculate the concreteness for each sentence in the text file, the algorithm would count the number of words found in Concreteness_ratings_Brysbaert_et_al_BRM.csv (# Words Found, Percentage). Then, the mean and median concreteness scores are computed.

	Document ID	Docum	Sentence ID	Sentence	Concreteness (Mean score)	Concreteness (Median score)	Standard Deviation	# Words Found	Percentage	Found Words	All Words
2	1	/Users/	1	As an introduction to	3.075714296	3.14	1.055722241	14 out of 17	82.35%	(introduction, 3.0), (conundr introduction. conundrum, complexity, aii, smith, bicameral, novel, consider, conversation, h
3	1	/Users/	2	The mother is dead, a	3.478333333	3.33	0.778316559	6 out of 7	85.71%	(mother, 4.6), (dead, 4.07), (c mother, dead, conversation, take, place, george, memory
4	1	/Users/	3	But let's not get too c	1.92	1.955	0.47644517	4 out of 4	100.00%	(let, 2.26), (get, 2.38), (compl let, get, complicate, yet
5	1	/Users/	4	In George's memory,	N/A	N/A	N/A	0	0.00%	N/A	['george']
6	1	/Users/	5	In George's memory,	2.806666667	3	0.370045042	3 out of 5	60.00%	(memory, 2.38), (last, 3.04), (george, memory, last, may, italy
7	1	/Users/	6	Mother, daughter and	3.263333333	3.3	1.117169041	15 out of 20	75.00%	(mother, 4.6), (daughter, 4.7 mother, daughter, george, brother, henry, travel, south, home, cambridge, england, see, ea
8	1	/Users/	7	Mother and daughter	4.0275	4.335	0.967552795	4 out of 5	80.00%	(mother, 4.6), (daughter, 4.7 mother, daughter, talk, restoration, fresco
9	1	/Users/	8	As Ms. Smith writes it "But which came first?	3.3775	3.285	0.755088295	4 out of 4	100.00%	(smith, 3.81), (write, 4.22), (cr smith, write, come, first
10	1	/Users/	9	her mother says.	3.59	3.59	1.428355698	2 out of 2	100.00%	(mother, 4.6), (say, 2.58) mother, say
11	1	/Users/	10	The chicken or the eg	4.885	4.885	0.120208153	2 out of 2	100.00%	(chicken, 4.8), (egg, 4.97) chicken, egg
12	1	/Users/	11	The picture underneat	4.2525	4.39	0.381870397	4 out of 4	100.00%	(picture, 4.52), (underneath, picture, underneath, picture, surface
13	1	/Users/	12	The picture below can	3.145	2.74	0.919909416	4 out of 5	80.00%	(picture, 4.52), (come, 2.72), (picture, come, first, george, say
14	1	/Users/	13	Because it was done f	2.61	2.61	0.212132034	2 out of 2	100.00%	(do, 2.46), (first, 2.76) do, first
15	1	/Users/	14	But the first thing we	3.4	3.19	0.656031842	10 out of 10	100.00%	(first, 2.76), (thing, 3.17), (se first, thing, see, mother, say, time, thing, see, one, surface
16	1	/Users/	15	So does that mean it c	2.66	2.72	0.14	3 out of 3	100.00%	(mean, 2.5), (come, 2.72), (fir mean, come, first
17	1	/Users/	16	And does that mean t	2.901666667	2.75	0.973456042	6 out of 6	100.00%	(mean, 2.5), (picture, 4.52), (i mean, picture, know, may, well, exist
18	1	/Users/	17	This same question, a	2.626666667	2.37	1.044234169	9 out of 10	90.00%	(question, 3.36), (many, 2.37 question, many, others, arise, halve, novel, though, drastically, different, circumstance
19	1	/Users/	18	This is not a book to b	3.01	3.175	1.115445294	12 out of 12	100.00%	(book, 4.9), (read, 3.56), (pas book, read, passively, even, read, especially, since, take, figure, two, part, connect
20	1	/Users/	19	To amp up her games	3.377857143	3.445	0.685881552	14 out of 16	87.50%	(amp, 3.59), (smith, 3.81), (o; amp, gamesmanship, smith, publish, two, edition, sequence, halve, vary, copy, copy, georg
21	1	/Users/	20	"How to Be Both," a fi	3.075454545	3.25	1.108010502	11 out of 12	91.67%	(finalist, 3.79), (year, 3.25), (n finalist, year, man, booker, prize, accessible, yet, ordinary, george, story, come, first
22	1	/Users/	21	It's knottier and more	N/A	N/A	N/A	0	0.00%	N/A	['knottier']
23	1	/Users/	22	It's knottier and more	2.73	2.55	0.930161276	13 out of 18	72.22%	(challenge, 2.11), (begin, 2.5 knottier, challenge, begin, read, francescho, del, cossa, renaissance, painter, whose, first, na
24	1	/Users/	23	And it's knottier and i	N/A	N/A	N/A	0	0.00%	N/A	['knottier']
25	1	/Users/	24	And it's knottier and i	3.324666667	2.9	1.239768796	15 out of 17	88.24%	(challenge, 2.11), (wonder, 1 knottier, challenge, wonder, artist, keep, conjure, girl, kind, magic, tablet, portrait, blonde, t

The auto-generated output of the line chart shows the mean concreteness and the median concreteness score of all sentences given index 1, 2, and 3 in all text files. But we encourage you to look at the "Excel data" spreadsheet to manually create a line graph for more sentences.

Concreteness Scores by Sentence Index
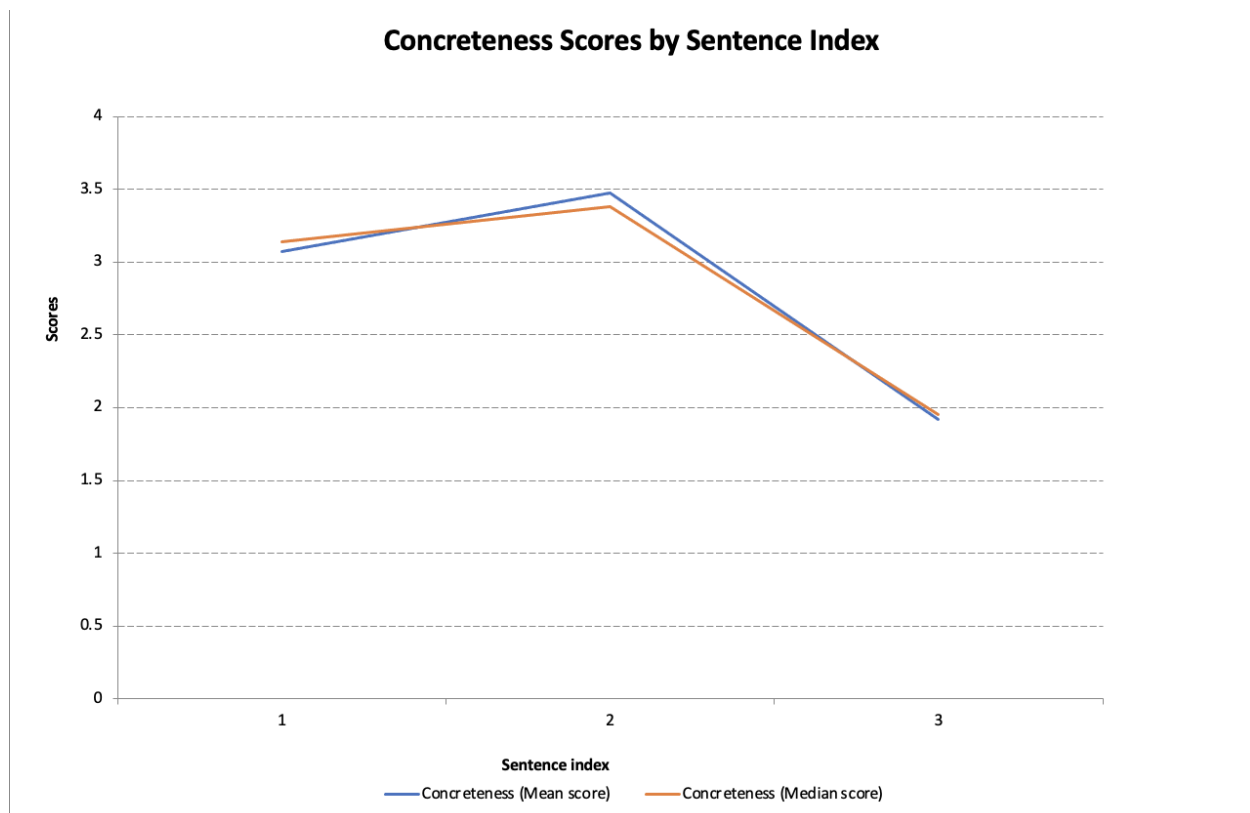
Example from the New York Times Book Review Corpus (Spring 2021 Project)

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R.

(2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33, 497–505.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–90.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding Part-of-Speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, 44, 991–997.

Brysbaert, M., Warriner, A.B., Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904–911.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44, 287–304.