

Talking About Yoga: A Corpus-driven Survey of Yoga-related Websites and Blogs

Barry Grossman

1. Introduction

The restraint of the modifications of the mind-stuff is yoga.

If you can control the rising of the mind into ripples, you experience Yoga
(Patanjali, 2012, v 1.2).

Out of compassion you have taught me the supreme mystery of the Self.

Through your words my delusion is gone ("The Bhagavad Gita," 2007, v 11.1-2).

The practice and study of Yoga has a very long and rich written history, the Vedic material being some of the oldest and extensive in the world (Gonda, 1977; "The Upanishads," 2007). Over the centuries, aphorisms, prose, poems and songs have exalted the union of individual consciousness (spirit/soul (?)) with the universal, advising and guiding the practitioner to sacrifice a limited, ego-based existence and eliminate ignorance and thereby achieve wisdom and liberation ("The Bhagavad Gita," 2007; Feuerstein, 2002; Guruji, 2018; Iyengar, 2001; Mallinson, 2017; Marshall, 2009; O'Neill, 2015; Simpson, 2021; "The Upanishads," 2007; "The Yoga Sutras of Patanjali," 1984). Various methods were created to achieve these goals; meditation (*dharana*, *dhyana* and *samadhi*), chanting, (*mantram*), breathing techniques (*pranayama*), physical postures (*asana*), spiritual song (*kirtan*), etc. Millions of books have been written about yoga. A Google Books search with the keyword *yoga* retrieved 'about 8,850,000' hits (in 1.6 seconds!¹) There is now a huge amount of online data such as yoga-related studios, retail goods, retreats, teacher trainings, life-coaching, etc. But are these websites all talking about the same thing? How is the term *yoga* delineated and used online and is it the same or different from the traditional definition given in the opening quotations ?

It would be a monumental task for even a large group of scholars to read through all the data, written and spoken, about yoga and analyze their contents without the help of computers. There is simply too much data, with new books, articles, web pages and blogs being added every day. This problem in particular is where research with the help of an electronic corpus is beneficial. A corpus is a collection of media; texts, movies, radio, web pages, blogs, etc. Some corpora contain only one medium (e.g., text) while others have many (e.g., text, video, audio, etc.) Some are time restrictive (e.g., 1900-2000 AD) and some are ongoing. Some are topic specific (e.g., news) and some include various sources. The data in corpora are marked (i.e., 'tagged') in some way to make it more easily analyzable; for example, by name, date, topic, media type, source, word, syntax, etc. And so, for example, a person interested in 'how mindfulness is being used during the COVID-19 pandemic' can access a specific corpus containing that data (e.g., <https://www.english-corpora.org/corona/>). At the time of this writing, the Coronavirus Corpus retrieved 1459 'concordance lines'

¹ Access date: Oct. 28, 2020.

using the search keyword *mindfulness*. That data can then be analyzed according to the particular research focus of the investigator.

The goal of the present research is to investigate how the word *yoga* is being used online in websites and blogs by discovering possible patterns in the data and analyzing those patterns.

2. Method

The research method used for the present discussion is ‘corpus-driven’, i.e., there is no preconceived hypothesis nor theory that is being tested. Data is mined from the corpus and then analyzed, and although no research agenda is completely unbiased, this method does have the advantage of relative objectivity and a broad theoretical and methodological spectrum from which the data can be analyzed. On the other hand, if data were collected and analyzed in order to test a hypothesis, for example, that “*war metaphors* are more frequent in the news during pandemics”, then this would constitute a ‘corpus-based’ approach. This method has the advantage of answering specific research questions, but narrows the scope of the researcher’s analysis (Barlow, 1996; Deignan, 2005; Gast, 2006; Gilquin & Gries, 2009; Hunston & Francis, 2000; Sinclair, 1991).

Two corpora were used for this investigation, The Intelligent Web-based Corpus (hereafter, ‘iWeb’) (Davies, 2018) and the Corpus of Global Web-Based English (hereafter, ‘GloWbE’, pronounced /globe/) (Davies, 2013b). The iWeb consists of 14 billion words taken from 95,000 web sites from six countries, current until 2017². The GloWbE includes about 1.9 billion words of text from 20 countries from 2012 until 2013. These corpora were chosen for the vast amount of data specific to Web pages and blogs as well as the ease at which data source comparisons could be made.

The first investigation examines the iWeb corpus’ main query page using the WORD search option and offers an overall ‘snapshot’ of information about the search term *yoga*. Information such as related topics, collocates (words that co-occur), clusters (related words grouped by syntax), Keyword in Context (KWIC), and often-used websites are discussed in Section 3.1. The second investigation used the LIST function of the iWeb corpus. A 500-item random sample was chosen for this examination, the results of which are discussed in Section 3.2. The third query compares word frequency calculations and Mutual Information (MI) Scores from the GloWbE corpus. These are presented in Section 3.3. The final investigation analyzes data by country in the GloWbE corpus. The results are presented in Section 3.4. Discussion and conclusions are presented in Sections 4 and 5, respectively.

3. Results

3.1. iWeb: WORD query for the search term *yoga*.

A screenshot of the WORD query employing the keyword *yoga* in the iWeb corpus is shown in Figure 1. At the top-left corner of the window is the keyword, its syntax, a definition and hyperlink to the etymology of

² See <https://www.english-corpora.org/iweb/> for more about the sources and selection process.

the word, shown as the symbol [E]. There is also an audio pronunciation (the speaker icon) and hyperlinks to videos that use the word *yoga* from the websites PlayPhrase, YouGlish, and Yarn. Below these links are hyperlinks to reference and translation sites Google, WordRef, Reverso and Linguee.

Figure 1. iWeb: Screenshot of the WORD query for the keyword *yoga*.

The screenshot shows the iWeb application interface with the following details:

- Top Bar:** iWeb: The 14 Billion Word Web Corpus. Includes icons for PDF, XML, CSV, ZIP, HELP, and a search bar.
- Search Bar:** WORD
- Contextual Buttons:** SEARCH, WORD, CONTEXT, HELP.
- Page Content:**
 - YOGA (NOUN) #3825:** A detailed card for the noun 'yoga' with a count of 3825. It includes:
 - Topics:** practice, class, exercise, meditation, fitness, workout, pose, muscle, body, teacher, mat, stress, studio, breath, stretch, posture, gym, instructor, exercise, training.
 - Collocates:** noun class, practice, teacher, meditation, mat, studio, instructor, pain.
 - Verb:** teach, practice, pose, stretch, exercise, pilates, dance, incorporate.
 - Adj:** hot, daily, gentle, certified, restorative, prenatal, spiritual, ancient.
 - Adv:** mindfully, avidly, kimberly, therapeutically, one-to-one, measurably, uncritically, radiantly.
 - Clusters:** A list of clusters related to yoga, such as 'yoga + yoga class + yoga classes + yoga practice + yoga teacher + yoga mat + yoga pants + yoga for + yoga studio'.
 - Websites / Virtual Corpora:** A list of websites including yogasix.com, sunshinyoga.com, yogaalliance.org, kidsyogastories.com, seattleyoganews.com, yogauonline.com, yogayoga.com, cosmickids.com, yogatreesf.com, and yogaaccessories.com.
 - Concordance Lines:** A table showing 10 examples of how 'yoga' is used in context across various websites. The columns show the source website, the sentence, and the highlighted words with their definitions and parts of speech.

TOPICS, which here means words that often appear on websites that use the word *yoga*, are displayed at the top-right of the page. The most frequent 20 are: *practice, class, exercise, meditation, fitness, workout, pose, muscle, body, teacher, mat, stress, studio, breath, stretch, posture, gym, instructor, exercise* and *training*. These are shown in Figure 1, but clicking on the hyperlink TOPICS offers a more complete list. Even within the 20-item data set in Figure 1, an overall sense of the use of the term *yoga* can begin to be formed. Four words relate to specific yoga techniques; *meditation, breath, pose*, and *posture*. Two words refer to a yoga leader (traditionally called a ‘guru’, i.e., one who “closely supervised the spiritual progress of a disciple” (Feuerstein, 2003, p. 18)), instantiated in the data as *teacher* and *instructor*. The words *class, gym*, and *studio* refer to the locale at which yoga is practiced (traditionally called ‘ashram’, meaning “religious exertion” (Gabriel, 2017)). Four terms relate to general physical movement; *exercise* (2x), *fitness*, *workout* and *stretch*. Two terms relate to anatomy, *muscle* and *body*. The two terms *training* and *practice* are ambiguous as to what they refer, most likely (based on subsequent data) some combination of yoga-related physical, mental/emotional and/or spiritual techniques. The final term, *stress*, relates to the

mental/emotional condition of the yoga practitioner and/or teacher as well as environmental factors instigating this particular mental state.

Appearing below TOPICS in Figure 1 is a collection of COLLOCATES, i.e., words that appear physically close to the search term, typically at a range of four words to the left and right. The collocates are categorized by the syntax in which they appear, making it easy to distinguish between the noun *practice* and the verb *practice*, for example. The most frequent eight for each category are: NOUN: *class, practice, teacher, meditation, mat, studio, instructor* and *pant*; VERB: *teach, practice, pose, stretch, exercise, pilates, dance* and *incorporate*; ADJ: *hot, daily, gentle, certified, restorative, prenatal, spiritual* and *ancient*; and ADV: *mindfully, avidly, kimberly, therapeutically, one-to-one, measurably, uncritically, and radiantly*. Although axiomatic, looking with a critical eye at the data is paramount. Notice that the ADV *kimberly* has been mistakenly tagged by the automatic Part-of-Speech software (aka, ‘POS tagger’) as an adverb³, likely due to the *-ly* at the end of the proper name. In a detailed analysis, this is either deleted from the data set or tagged manually as a proper name. Even within this limited COLLOCATES section, we can again begin to glimpse what *yoga* ‘looks like’ in websites., i.e., *what it is, what it is used for, where it is practiced, from whom it is learned* and its *qualities and/or effects*.

Figure 2. iWeb: Two, three, and four-word CLUSTERS for the keyword *yoga*.

10501	yoga class	5348	to yoga	3315	yoga and meditation	1848	such as yoga	333	yoga is a great	281	activities such as yoga
8535	yoga classes	4712	in yoga	1537	yoga teacher training	1495	practice of yoga	322	yoga and tai chi	253	has been practicing yoga
7812	yoga practice	4000	doing yoga	1015	yoga and pilates	1241	benefits of yoga	247	yoga on the beach	193	as well as yoga
7452	yoga teacher	3899	do yoga	751	yoga can help	1058	as a yoga	231	yoga and meditation classes	183	has been teaching yoga
5984	yoga mat	3807	hot yoga	672	yoga and pilates	947	meditation and yoga	209	yoga or tai chi	166	different types of yoga
5208	yoga pants	3762	for yoga	616	yoga or meditation	927	in a yoga	190	yoga and pilates classes	153	different styles of yoga
5184	yoga for	3381	practicing yoga	588	yoga is not	854	in the yoga	170	yoga can help you	147	health benefits of yoga
4566	yoga studio	2915	with yoga	411	yoga or pilates	754	to a yoga	153	yoga in the park	139	i am a yoga
4174	yoga in	2823	practice yoga	395	yoga class in	696	to practice yoga	142	yoga teacher training in	134	go to a yoga
4095	yoga poses	2099	as yoga	378	yoga classes in	689	for a yoga	133	yoga teacher training program	134	have been practicing yoga
3918	yoga instructor	1973	like yoga	375	yoga and fitness	664	style of yoga	124	yoga is an ancient	132	i started doing yoga
2752	yoga can	1821	hatha yoga	365	yoga class at	595	on a yoga	121	yoga teacher training course	128	tai chi and yoga
2404	yoga teachers	1483	from yoga	362	yoga and other	584	new to yoga	119	yoga as well as	124	techniques such as yoga
2391	yoga with	1324	about yoga	330	yoga retreat in	573	pilates and yoga	117	yoga in the morning	119	in love with yoga
2196	yoga mats	1284	teaching yoga	327	yoga and tai	557	styles of yoga	113	yoga teacher training programs	104	have been doing yoga

At the center of the WORD query page from Figure 1 is the CLUSTERS section, which shows information related to strings of words from two to four in length, the asterisk before or after the keyword signaling whether the related word(s) come before or after the key term *yoga*. This can be useful to get a more detailed picture of how the word is being used. Figure 2 shows the most frequent 15 clusters for *yoga*

³ In this case, the retrieval of the term/name seems to be an error. Clicking on the hyperlink *kimberly* brings up an error message which reads, “*kimberly is not one of the 60,000 words of English, and so it's not possible to see the full range of information...*” This only helps to support the need for linguistic vigilance in these kinds of corpus studies; i.e., to not only look at the numbers/statistics, but also at the details of the language itself.

in each of four categories; the left-most column displays *yoga** (*yoga* plus one word to the right), the second column from the left is **yoga* (one word to the left of *yoga*), the next is *yoga*** and so on until the right-most column ****yoga*. These data, e.g., *yoga classes*, *types of yoga*, *yoga teacher training*, *yoga studio*, *benefits of yoga*, *yoga can help you*, *yoga is an ancient health benefits of yoga*, *yoga and meditation classes*, etc. help to corroborate the preliminary categorizations from the TOPICS and COLLOCATIONS sections above. At this point, the data are far from complete; however, with data-driven analyses, these kinds of preliminary data allow for low-resolution conceptions and hypotheses to be formed which can be tested when more data has been collected and analyzed.

The next section of the WORD query page is called, WEBSITES/ VIRTUAL CORPORA and consists of the most frequent web pages from where the data was mined. Clicking on the ‘more’ hyperlink displays each webpage along with the most common words for each, allowing a quick glance at the types of ‘services’ each website represents (e.g., *yoga education*, *yoga retail*, *kids’ yoga*, *yoga related travel*, etc.) A screenshot of this is shown in Figure 3. Using the VIRTUAL CORPORA tool, it is possible to download the data and analyze the corpus with third party software in order to search and analyze the data according to individual parameters, but this is beyond the scope of the present discussion.

Figure 3. iWeb: WEBSITES section for the keyword *yoga*.

Freq	Per 1000	% Pages	Website (click for topics)	Words
415	19.1	86%	yogasix.com	yoga, mat, studio, workshop, fitness, class, teacher, training, challenge, practice, strength, teach, body, deep, benefit, student, experience, offer, learn, build,
290	18.7	84%	sunshineyoga.com	yoga, mat, foam, durable, workout, e-mail, comfortable, color, price, customer, please, quality, offer, call, add, provide, high, great, need, want,
2040	18.3	93%	yogaalliance.org	newsroom, yoga, alliance, teacher, registered, teaching, training, teach, register, practice, resource, community, standard, school, hour, program, class, education, member, language,
766	16.6	91%	kidsyogastories.com	pose, yoga, pose, kid, practice, sequence, daughter, book, story, animal, fun, movement, child, teacher, inspire, fun, celebrate, love, picture, teach,
2383	16.5	91%	seattleyoganews.com	yoga, meditation, workshop, teacher, studio, practice, teach, class, training, body, community, learn, life, lead, bring, experience, offer, student, start, help,
412	16.5	95%	yogaonline.com	yoga, posture, pose, anatomy, therapist, practice, practice, therapy, teacher, teaching, pain, body, principle, movement, discuss, physical, healthy, teach, course, common,
831	15.4	82%	yogayoga.com	yoga, meditation, teacher, therapy, teaching, practice, training, teach, class, course, student, body, program, learn, hour, location, begin, offer, bring, life,
373	15.0	86%	cosmickids.com	cosmic, yoga, posture, kid, adventure, brilliant, teacher, teach, class, story, adult, child, wonderful, video, book, learn, fun, sign, share, love,
508	13.4	93%	yogatreesf.com	yoga, alignment, meditation, teacher, teaching, training, practice, tree, teach, class, physical, body, learn, student, life, feel, experience, become, help, create,
1281	13.3	54%	yogaaccessories.com	yoga, practice, body, help, great, more, find, take, just,
1520	13.3	65%	yogamatters.com	yoga, practice, long, long, help,
275	13.1	76%	yogafinder.com	yoga, meditation, class, studio, practice, teacher, teach, body, mind, offer, community, learn, student, level, life, experience, help, day, find, more,
8202	12.6	98%	bookyogaretreats.com	yoga, yogi, retreat, meditation, retreat, relaxation, review, teacher, edit, relax, practice, instructor, spiritual, accommodation, wonderful, amazing, book, beach, beautiful, teaching,

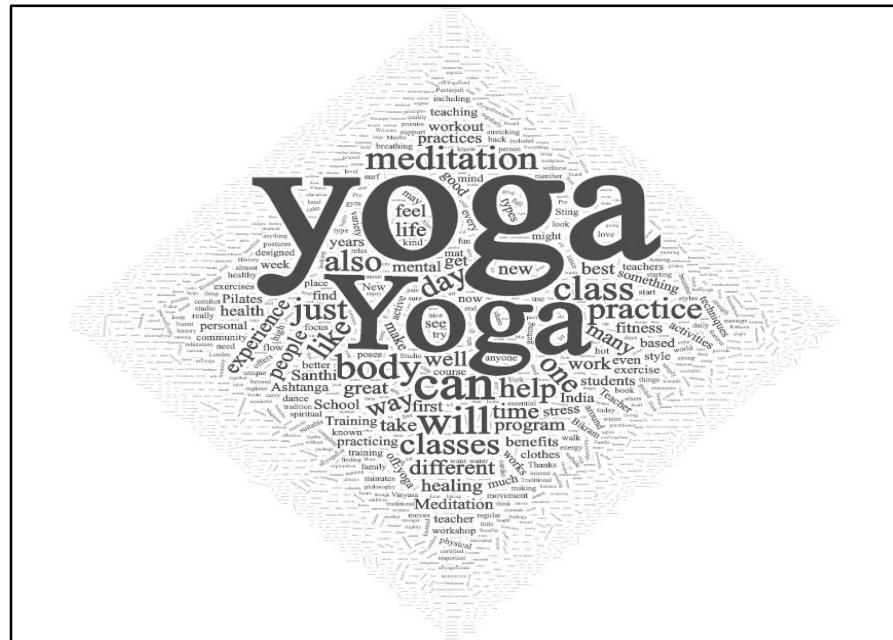
The WEBSITES function is convenient for discovering patterns that might otherwise be obscured. For example, notice the two websites for children, [kidsyogastories.com] and [cosmickids.com] and related words that do not appear in the other websites: *daughter, story, animal, fun, love, picture, wonderful, share*, etc. As another example, notice the words that appear on the retail-oriented website [sunshineyoga.com]: *mat, foam, durable, e-mail, price, customer, offer, call, need, want*, etc. Thus, the words that appear on websites display the type of service(s) offered and/or what their branding/marketing focus is. This is kept in mind for later as a working hypothesis for when more words are retrieved from the query and more encompassing data is analyzed.

The final section of the WORD query from Figure 1 is CONCORDANCE LINES, and consists of full phrases/sentences. The keyword *yoga* appears in the middle of the line with color-coded ‘tags’ for syntax and punctuation, called ‘Keyword in Context’ (KWIC). This allows for more linguistically-oriented analyses as well as being able to check for consistency in the meaning and sense of the keyword by analyzing the immediate context of the keyword. The short concordance list in Figure 1 will not be discussed here due to a more detailed discussion with a much larger data set presented below in Section 3.2.

3.2. iWeb: Data sample and Categorizations

A 500-item random sample of concordance lines for *yoga* (total $n= 367,006$) from the iWeb corpus was retrieved and visualized as a Word Cloud (Zygomatic) in Figure 4 and thematically categorized in Figure 5.

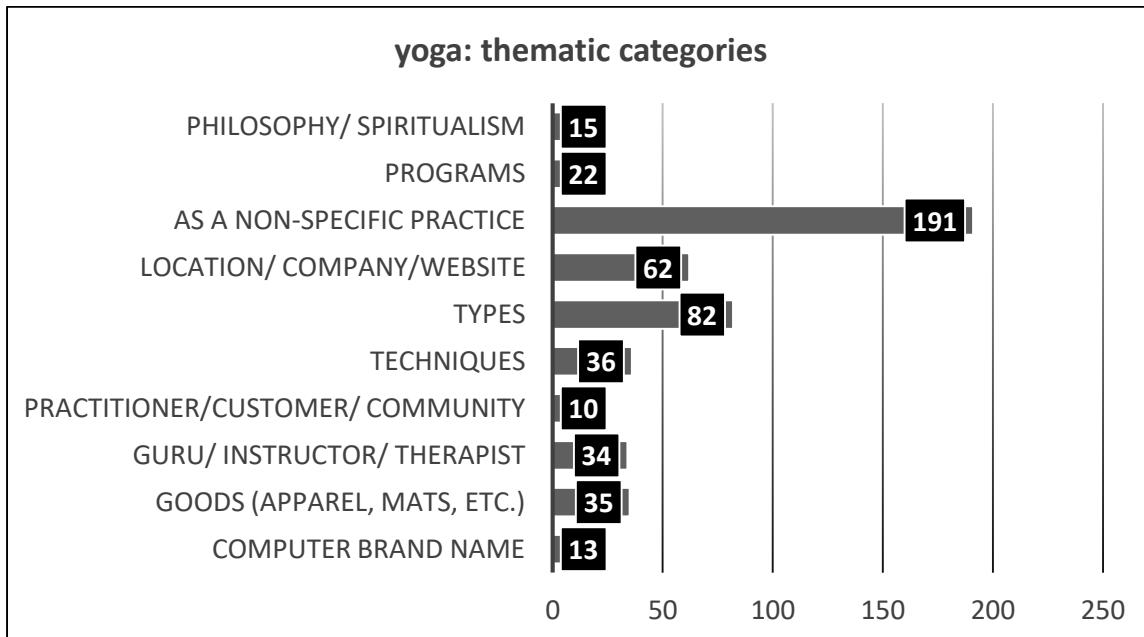
Figure 4. Word Cloud for *yoga*, iWeb, random sample ($n=500$).



As a visualization in Figure 4, the more frequent collocations that congregate around the keyword *yoga* are represented by the larger fonts; the smaller fonts represent less frequent terms. Although not very useful as a tool for analyzing data, this kind of visual representation allows for a rapid, gestalt understanding of high

and low frequency items in the data set. For example, besides the keyword *yoga*, the verbs *can* and *will* are very large in Figure 4, perhaps occurring in sentences such as, ‘Yoga *can* heal the body and mind’ and ‘Yoga *will* mobilize your spine’. These are not instantiated sentences in the data, but can be hypothesized as possible entries. Other high frequency terms are: *class(es)*, *practice*, *just*, *meditation*, *different*, *also*, *body*, *experience*, etc.

Figure 5. iWeb: Thematic categories for the term *yoga* (random sample, $n= 500$).



Ten general categories are proposed for the random sample ($n= 500$) data in Figure 5:

- 1) philosophy/spiritualism, 2) programs, 3) as a non-specific practice, 4) location/company name/website, 5) yoga type or subtype, 6) specific technique (posture, breathing, mantra, etc.), 7) practitioner, 8) instructor/therapist, 9) goods, and 10) brand name of personal computer. These ten thematic categories comprise 100 percent of the data from the 500-item random sample.

The word *yoga* used ‘as a non-specific practice’ is the most frequent with 191 ‘hits’ (38.2%). Here, some specific *yoga* technique, e.g., posture, breath, meditation, mantra, mudra, bandha, etc., is not elucidated. Examples one and two illustrate this use.

- 1) And if you are *practicing yoga* somewhere other than your house, avoid wearing any jewelry or belts or anything that... (iWeb:listaka.com:#219)
- 2) There are daily *yoga* and healthy-eating cookery classes for guests to enjoy, as well as massages and beauty... (iWeb:www.her.ie:#496)

What do *practicing yoga* and *yoga...classes* specifically mean? There are limitless possibilities and combinations of technique, but the usage here seems to be the default term for some combination of poses, breathwork, meditation, chanting, etc. However, even upon further investigation, detailed context is

sometimes simply not available, as in the last line of the blog post in Figure 6. The blogger writes about doing various kinds of exercise, but *also doing yoga* and *also doing lunges/squats*⁴, so it is not obvious here what the author is actually referring to when using the phrase *doing yoga*, except that for the writer, it does not include *squats* nor the *exercises* referred to in the previous sentence.

Figure 6. Screenshot of blog mentioning *yoga* as ‘a non-specific, generic practice’;
(<https://forums.runnersworld.co.uk/discussion/174374/newbie-runner-please-help-me-shift-a-stone>).

Newbie runner: please help me shift a stone!

 Cornwall 91
April 2011 in General running

Hi everyone (:

I'm new to this site and new to running in general and I'm in need of a little help to reach my goal. ***My main aim is to lose around 1 stone in the next 12-14 weeks***, which from a little bit of research seems like a fairly achievable goal. I currently weigh 9 stone and would like to get this down to 8 through running and maintaining a healthy diet.

Exercise

At the moment I work out 4 times a week, 3 x 20 mins HIIT on an exercise bike and 1 x 30 min spinning, all done at the gym. I will be living at home for the summer though and will have access to a treadmill, so I will need to change and increase the amount of exercise I am doing. I will also be doing yoga once a week as well as trying to incorporate lunges/squats etc.

The next most frequent use is ‘types of *yoga*’ with 82 (16.4%) instances, which specifically name some type or subtype of *yoga*, such as examples three and four.

- 3) *Bikram Yoga* is a process that can reduce the symptoms of many chronic diseases and is an... ([iWeb:islanddailydeals.com:#283](#))
- 4) I was very good about spending 30-60 minutes almost every single morning doing *pregnancy yoga*, or else going for a long walk in the park or along the river...
([iWeb:keeperofthehome.org:#185](#))

The following category includes examples where *yoga-related ‘companies, websites or locations’* are explicitly identified, instantiating 62 (12.4%) hits, shown in examples five and six.

- 5) Namit Kathoria is a *yoga teacher* based in *Yoga Sutra Shala*, Spain.
([iWeb:bookyogaretreats.com:#15](#))
- 6) ...bringing in ideas and teachings from several different schools. - *Yoga Agora* believes that every practitioner has his or her own valid reason to step on...
([iWeb:fitreserve.com:#451](#))

⁴ Lunges and squats are traditional *yoga* postures but they are given by the author as antithetical to the term *yoga*.

The next three categories are: ‘techniques’, ‘yoga goods’, and ‘instructor/leader’, having very similar frequencies at 36 (7.2%), 35 (7%) and 34 (6.8%), respectively. One of each is provided in examples 7, 9 and 10, respectively. For the ‘techniques’ category, not only the name of a specific technique but also references to techniques were included, as in example (8):

- 7) That's why the majority of *yoga poses* focus on opening up the hip joints...
(iWeb:hitchedmag.com:#352)
- 8) ...although if it were possible and reasonable for me to learn some *yoga techniques* from a book or something, I'd be open to doing that.
(iWeb:martialartsplanet.com:#412)
- 9) ...she he has no nice clothes, wears nothing but *yoga pants* and work out tops...
(iWeb:shaggytexas.com:#425)
- 10) Join *yoga instructor* Sarah B... for a story followed by a yoga lesson!
(iWeb:eventkeeper.com:#313)

The following category is ‘programs’, where the word *yoga* is used in the context of retreats, events, teacher trainings, conferences, etc. There are 22 (4.4%) instances in this category; two examples are provided in (11) and (12). Although retreats and teacher trainings are only somewhat similar in their particular content, they are treated here as a general type of ‘program’.

- 11) ...the island welcomes around 130,000 visitors per year; some for religious visits or *yoga retreats*, and others to simply be immersed in the outdoor tranquility.
(iWeb:calm.ac.uk:#403)
- 12) Our *Yoga Teacher Training* draws its elements from a variety of long established Yoga traditions and is... (iWeb:swiha.edu:#154)

The final three categories are ‘philosophy/spiritualism’, ‘computer brand name’, and ‘practitioner/customer/community’, with 15 (3%), 13 (2.6%) and 10 (2%) hits, shown in examples 13, 14 and 15, respectively.

- 13) In complexed (sic) *Yoga system of beliefs* the human soul Jiva Atma is liberated when we stop identifying with... (iWeb:pilatesandyogafitness.com:#199)
- 14) Are you sure its (sic) *a Yoga 2 Pro* and not a regular Yoga 2... They have slightly different specs... (iWeb:turbofuture.com:#2)
- 15) ...there are plenty of open-minded and caring *yoga practitioners* who are disabled in their own way by their ignorance of where to begin... (iWeb:yogitimes.com:#143)

The data from the 500-item random sample suggests that 10 thematic categories can account for the kind of ‘talk’ being conducted about *yoga*, ‘yoga as a non-specific practice’ being the most prolific.

3.3. Collocations and Categories

Looking at the data from a more statistical viewpoint, Figure 7 shows the 50 most frequent collocates of *yoga* in the GloWbE corpus, arranged from highest to lowest Mutual Information (MI) Score⁵. A high MI score shows statistically significant relevant words that appear close to the keyword and have a high level of statistical confidence of the outcome of the results. If the collocations at the bottom of Figure 7 are read from left to right, words further to the left (i.e., higher MI Scores) are more specifically related to *yoga* whereas collocations with lower MI scores can be applied more easily and broadly to non-yoga topics as well. In statistical terms, this means that the MI high-scoring words are less likely to occur near the keyword *yoga* by chance.

Figure 7. GloWbE: Most frequent 50 collocates of *yoga* by Mutual Information (MI) Score ($n = 24,783$).

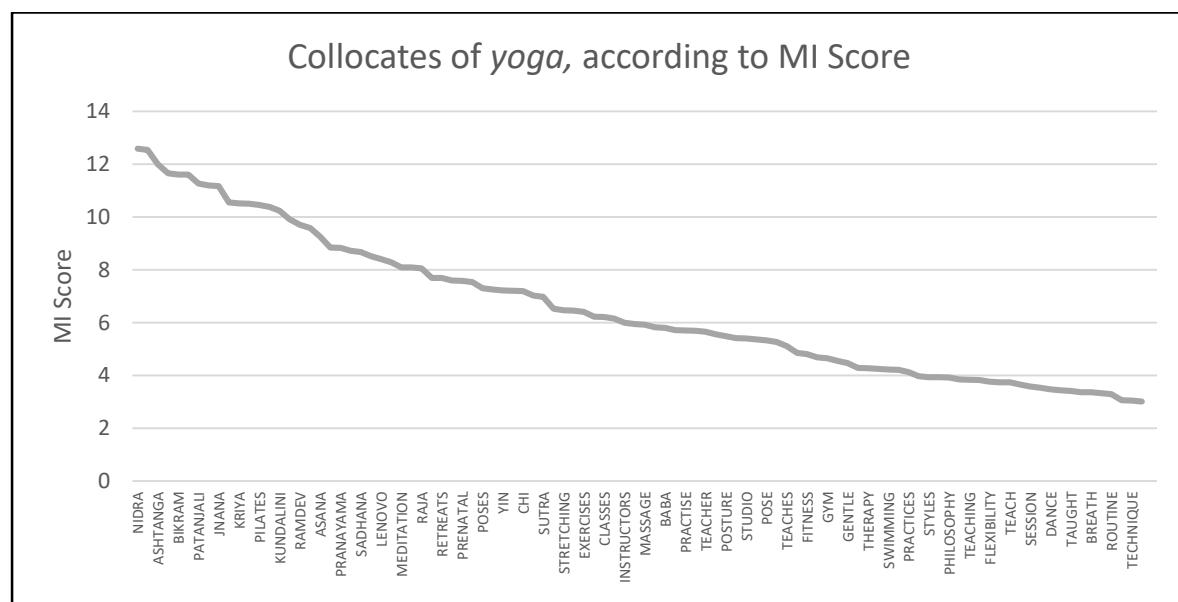


Figure 8. GloWbE: WORD search *yoga*, sorted by frequency ($n = 24,783$).

RANK	COLLOCATION	FREQUENCY	MI Score
1	YOGA	2057	8.72
2	PRACTICE	1794	5.56
3	MEDITATION	1366	8.09
4	CLASS	1132	4.55
5	CLASSES	1007	6.21
6	TEACHER	1001	5.66
7	HATHA	542	12.54
8	PILATES	542	10.45
9	EXERCISES	438	6.41

⁵ For details about MI Score calculations, see: http://www.scholarpedia.org/article/Mutual_information.

10	KARMA	434	8.09
11	PRACTICING	430	7.2
12	EXERCISE	401	4.21
13	STUDIO	385	5.4
14	INSTRUCTOR	357	7.25
15	MAT	344	7.69
16	PANTS	327	6.46
17	TEACHING	322	3.84
18	RAJA	319	8.05
19	BHAKTI	310	10.38
20	TEACHERS	305	3.74
21	POSES	299	7.3
22	KUNDALINI	286	10.23
23	BIKRAM	283	11.6
24	PRACTICES	259	4.13
25	HOT	258	3.44
26	BREATHING	251	5.71
27	POSTURES	249	9.58
28	SAHAJA	243	11.6
29	GURU	233	6.23
30	RETREAT	223	6.15
31	PATANJALI	218	11.27
32	JNANA	209	11.17
33	TECHNIQUES	208	4.25
34	ASHTANGA	204	11.99
35	STRETCHING	199	6.47
36	SPIRITUAL	197	3.83
37	CHI	196	7.19
38	TEACH	195	3.74
39	TAI	190	7.53
40	FITNESS	184	4.81
41	SUTRAS	181	9.91
42	MASSAGE	181	5.92
43	SESSION	177	3.58
44	RELAXATION	170	6.53
45	PHILOSOPHY	166	3.92
46	PATH	166	3.06
47	DANCE	156	3.47
48	TAUGHT	156	3.41
49	SESSIONS	154	4.28
50	WALKING	150	3.01

Figure 8 shows the same collocation set arranged according to frequency and includes the MI Score as well. Notice that more/less frequent does not necessarily equate with more/less relevance to the keyword (i.e., MI Score), although this is often the case with single-corpus and/or small-scale investigations. That being said, a word's frequency rate analyzed together with its MI Score provides a more confident picture of each collocates' statistical profile. For example, the highest MI Scores in Figure 8 are: *Hatha* (MI=12.54), *Ashtanga* (MI=11.99), *Bikram* (MI=11.6), *Sahaja* (MI=11.6), and *Patanjali* (MI=11.27). The first four are branches or (sub)types of yoga and the final term is the name of a yoga Saint who wrote the renowned

Patanjali's Yoga Sutras (an ancient text that elucidates the purpose and methods of yoga (approx. 425-325 CE (Mallinson, 2017)). These terms differ from the five most frequent words: *yoga*, *practice*, *meditation*, *class(es)*, *teacher* which refer to the pragmatic, utilitarian activity of 'doing yoga'. Combining frequency and MI Score results, the seven most relevant words that appear with the keyword *yoga* in web pages and blogs in the GloWbE corpus are: *Hatha* (Freq= #7, MI Score= 12.54), *Pilates* (Freq= #8, MI Score= 10.45, *Bhakti* (Freq= #19, MI Score= 10.38), *Kundalini* (Freq= #22, MI Score= 10.23), *Patanjali* (Freq= #31, MI Score=11.27), *Jnana* (Freq= #32, MI Score=11.17) and *Ashtanga* (Freq= #34, MI Score= 11.99). *Hatha*, *Bhakti*, *Kundalini*, *Jnana*, and *Ashtanga* are branches or types of yoga⁶, *Patanjali* is the yoga Saint referred to previously and *Pilates* is a yoga-associated exercise program.

3.4. GloWbE: Analysis by Country

Another query using the GloWbE contains data which is tagged by country. In a CHART query ($n= 33,937$), shown in Figure 9, country of origin is ordered according to frequency per million. Statistically, because of the different amounts of online data for each country, a per million frequency calculation 'levels the playing field', so to speak.

Figure 9. GloWbE: Frequency per million for *yoga*, ranked by country.

<u>Rank</u>	<u>Country</u>	<u>Frequency per million</u>
1	India	84.32
2	New Zealand	37.65
3	Canada	26.33
4	Hong Kong	21.93
5	Singapore	20.64
6	Sri Lanka	19.66
7	Jamaica	18.32
8	Malaysia	18.03
9	South Africa	17.48
10	Australia	15.15
11	Philippines	14.2
12	United States	13.03
13	Tanzania	11.86
14	Ireland	11.85
15	Ghana	11.71
16	Bangladesh	9.9
17	Great Britain	8.05
18	Pakistan	7.51
19	Kenya	6.62
20	Nigeria	2.27

⁶ *Kundalini* historically refers to the dormant, potential energy that lies at the base of the spine whereas *Ashtanga* historically refers to the eightfold yogic method laid out by Patanjali. However, these terms were found to instantiate as 'types or subtypes of yoga' in the data.

India, being the historical birthplace of yoga along with the recent formal acceptance of yoga by government agencies (Basavaraddi, 2015), tops the chart with a frequency per million (hereafter, fpm) of 84.32, more than double that of New Zealand (fpm= 37.65), in the number two spot, and more than threefold of the number three position, Canada, (fpm= 26.33). Three Asian countries, Hong Kong, Singapore, and Sri Lanka make up the third to sixth most frequent, with 21.93, 20.64 and 19.66 fpms, respectively. Asian countries make up half of the top ten.

It is notable that the United States, having had over 16.8 billion dollars-worth of yoga-related business in 2016 (Alliance, 2016), the most of any country, ranks 12th. African nations comprise one-fourth of the total, represented by five countries: South Africa (fpm= 17.48), Tanzania (fpm= 11.86), Ghana (fpm= 11.71), Kenya (fpm= 6.62) and Nigeria (fpm= 2.27). It is important to be mindful of the fact that a country's low order ranking here does not necessarily mean that its population is doing less yoga, but low online traffic in websites and blogs⁷ is suggestive of less communication about yoga, and it might be extrapolated that communication and activity are related. But this kind of research is better suited to sociologists or at least, socio-linguists, who are better equipped for this type of investigation.

Concerning the ubiquity of yoga around the globe, online presence of 'yoga talk' of the most frequent 20 countries is represented by five of seven continents: Asia, North America, Africa, Australia and Europe. Thus, it can be suggested from the data here that *yoga* is an activity that certainly has world-wide attention.

4. Discussion

Throughout the centuries, yogic texts have explored concepts and methods of: unification with God, self-awareness, morality/ethics, desire, self-fulfillment, ecstasy states, ritual and psycho-physical purification, ritual and psycho-physical sacrifice, mastery of body and mind, individual-societal relationships, universal-individual causality, happiness, suffering, friendship and many more. It has been the progenitor of Hinduism, Buddhism, Jainism and perhaps even Christianity (A Google search of *Jesus studied yoga* retrieved 5,200,000 hits!) It is promoted as a possible spiritual, mental, physical, and sociological panacea for any ailment one may be suffering from. It is currently a newly-flourishing topic for modern medical and psychological research from a Western methodological paradigm with growing evidence of positive effects (IAYT, 2020). There are more yoga types and styles than ever before, with some of the more interesting being: goat yoga, SUP yoga, beer yoga, sling yoga, nude yoga, yoga raves, ganja yoga, tantrum yoga, and the list goes on.

Yoga has been on the rise around the world and especially in the USA. Investigating this, data retrieved from a query using the Corpus of Historical American English⁸ (Davies, 2013a), shown in Figure 10, notes the upward trend from the 1920's and then a frequency spike in the 1970's. This can most likely

⁷ Websites and blogs do not represent all Internet traffic. Other online media platforms contain large amounts of 'yoga talk' containing valuable data. However, due to space restrictions, these cannot be investigated here.

⁸ The 'COHA' corpus is a balanced corpus consisting of 400 millions words of American English.

be accounted for by the acclaimed visits to the USA of Swami Vivekananda in 1893, Paramahansa Yogananda in the 1920's and then Indra Devi in the 1950's. The book and television appearance of American Richard Hittleman in 1961, along with publications and access to teaching by B.K.S. Iyengar, Amrit Desai and Swami Rama began the surge in the 1970's that has lasted until the present (Deslippe, 2018, 2019; Hammond, 2007).

Figure 10: COHA: frequency of *yoga* from the years 1810-2009.

Year	Frequency (n=746)	Frequency per million
1810	0	0
1820	0	0
1830	0	0
1840	0	0
1850	0	0
1860	0	0
1870	1	0.05
1880	0	0
1890	0	0
1900	2	0.09
1910	2	0.09
1920	8	0.31
1930	9	0.37
1940	25	1.03
1950	8	0.33
1960	38	1.58
1970	88	3.7
1980	38	1.5
1990	193	6.91
2000	334	11.3

With more than 100 years of growing popularity worldwide and the huge number of current online data, the research question can again be proposed, "According to the Web, what is yoga?" To get a different sense of how people are currently 'talking about' yoga, an iWeb corpus query with the phrase *yoga is* was conducted (random sample $n=100$; total $n=13,055$). The first 20 of the random sample are listed below:

Yoga is...

Amazing

(One plus of yoga is) I felt much more loose and relaxed

a spiritual practice to many people

effective for overall general conditioning

a typical Indian activity

a state of balance to attain control

is not a workout, but it is still a physical practice

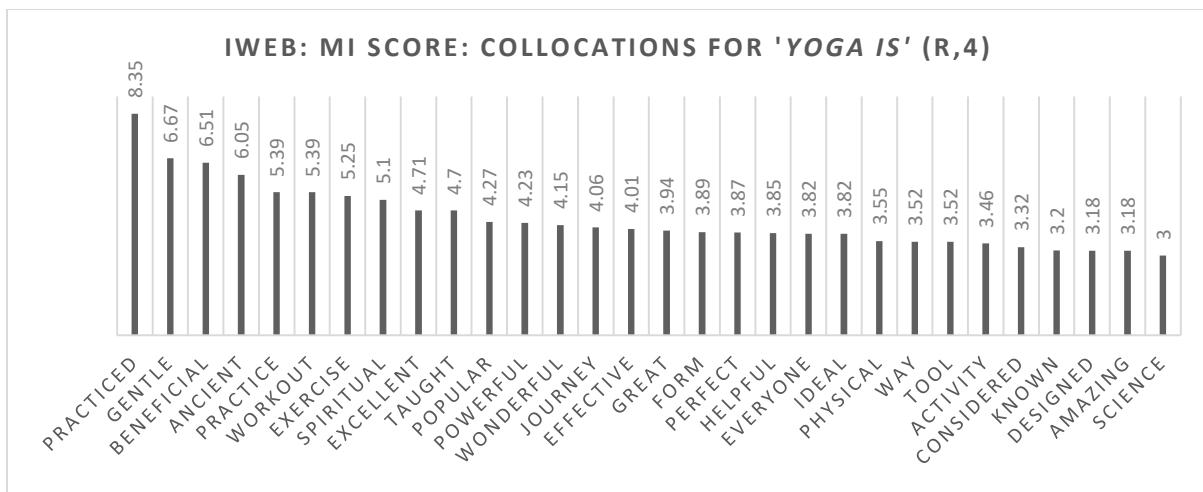
(Baby yoga is) another way to get rid of the stresses of everyday life

traditionally practiced at sunrise and sunset

*very Pisces to begin with, but any area of life can be helped
 a panacea of sorts
 effective at transforming bodies from the inside out
 a method of systematic discipline of all external and internal organs
 (Power Yoga is) Michelle's drug of choice as it helps alleviate the daily chronic pain
 (Pregnancy Yoga is) a gentle and relaxing style of yoga
 (If you think yoga is) just for little women in stretch pants, think again.
 as much about lengthening the muscles as it is about strengthening them
 (The very nature of yoga is) to help you stretch and get longer
 a viable alternative for these children and adolescents
 composed of a series of asanas, or postures, connecting with breathing
 (Ashtanga Yoga is) often referred to as power yoga
 Knowing Yourself
 not for babies
 a full body (including the mind) experience
 the only one that I feel actually helps bring much needed calming*

Even within this short list, various aspects of *yoga* are referred to: body, mind, spirit, stress reduction, pain reduction, breath work, self-awareness and self-control. In Figure 11, collocations up to four places to the right of the keywords *yoga is* are arranged according to the highest 30 MI Scores ($MI \geq 3.0$). It is possible to describe attitudes about *yoga* using this method; especially notable are the 15 adjectives instantiated: *gentle, beneficial, ancient, spiritual, excellent, popular, powerful, wonderful, effective, great, perfect, helpful, ideal, physical* and *amazing*. Analyzing the full random sample ($n=100$) also revealed no connotatively ‘negative’ adjectives. Thus, it can be suggested that most of the online ‘talk about’ *yoga* is positive and is considered a physically and spiritually beneficial practice.

Figure 11: iWeb: Highest MI Scores for collocations (R, 4) of *yoga is* ($n=30$).



Another question may be asked here; Whom is yoga for? Is it for the spiritual aspirant, the stressed-out mom or the fitness buff? Investigating this, an iWeb query using the search criteria *yoga is for* (random sample; $n=100$; total $n=284$) was analyzed. 59% of the post-term collocates (R, 4) were some form of *everyone* (*everyone*, *everybody*, *all people*, *every body*, *anyone*, etc.), as shown in examples 16 through 18.

- 16) But the truth is, *yoga is for everyone*, and we want to help you get started
(iWeb:yogainternational.com:#2).
- 17) *Yoga is for everybody*, not just the young and flexible (iWeb:retreat.guru:#81).
- 18) Now I'm of the mind that *yoga is for all sorts of people* and it can be altered to fit your needs... (iWeb:www.oregonlive.com:#274).

Thus, the majority of Web ‘talk’ seems to adhere to the belief that *any person* is able to do *yoga*. As a result of or perhaps because of this, yoga has been accepting of numerous theories, methods, types and practices, as long as the practice led to the goal of liberation of the mind/spirit. The Bhagavad Gita itself advocates two types of yoga, Bhakti and Karma⁹ and is accepting of diversity, “All paths, Arjuna, lead to Me” (“The Bhagavad Gita,” 2007, v 4.11). Within the Yoga Sutras of Patanjali, teachings advocate the yoga practices of: Jnana, Japa, Bhakti, Karma, Tantra and Kriya (Patanjali, 2012). Spirituality is both a personal and social experience, leading some individuals to seek isolation while others feel the need to share that experience. Thus, throughout the history of spiritual practices (including religions), there are hermits who live in the forests and caves but also small communities and great cathedrals. Every major spiritual practice has numerous branches where people with different ideas created new paths towards some notion of salvation, liberation, union, God... It is through words that we share these experiences. The different ways that prophets, saints, seers and monks choose to convey their experiences to their audience makes a particular type of impact on them and that message is filtered and analyzed through each individual mind. This creates both a shared vision of reality within the group that supports cohesion, but there is also individual interpretation that could cause rifts to appear in the theoretical/theological underpinnings of that community. Given these ever-changing circumstances, it is indeed noteworthy that yoga has survived and-thrived across the globe.

5. Conclusions

Collocational and statistical analyses of the term *yoga* in web pages and blogs suggests that ten thematic categories can account for the data. They are: 1) philosophy/spiritualism, 2) programs, 3) as a non-specific, general practice, 4) location/company name/website, 5) type or subtype of *yoga*, 6) techniques, 7) practitioner, 8) instructor/therapist, 9) retail goods, and 10) brand name of personal computer. The most frequent is ‘yoga as a general, non-specific practice’. This suggests that what is being practice and called

⁹ The meaning of *Karma* here is that of Karma Yoga, i.e., working in the world for the sake of others. This lessens the power of the Ego, allowing the aspirant access Universal consciousness.

yoga may be assumed to be understood by the audience. Whether this is actually the case (i.e., the assumption and/or the understanding) cannot be confirmed without further investigation. Language is an imperfect tool and so the chances are that a statement such as, *I do yoga* will have multiple interpretations as to what that means in practical terms.

However ambiguous the language about the actual practice of *yoga* is, based on data from adjective collocations with *yoga is*, there is little doubt that *yoga* is viewed in a positive light and is seen as *beneficial*, *wonderful*, *helpful* and *effective*. Concerning *whom yoga is for*, the data suggests that *anyone* can do *yoga*. It was also seen that the ‘talk of *yoga*’ covers the body, mind and spirit, with many types of programs, teachers and retail goods to support a huge variety of needs and objectives.

India is the birthplace of *yoga* and has a prolific amount of web content related to *yoga*. Asian countries account for half of the ten most-frequent. The United States, which conducts the largest amount of *yoga*-related business in the world, ranks 12th in web traffic. Although the two factors are probably not directly related, an interesting survey would be the amount of online retail language compared to the actual amount of sales in different countries. This would perhaps be better analyzed by a marketing specialist or economist, but the data could be interesting and perhaps reveal the attitudes of money and social status to *yoga* in different communities.

In my personal experience as a *yoga* practitioner and instructor (Yoga Alliance® E-RYT500), I see a need for language specificity and clarity in the *yoga* community as to *what yoga is* without losing the astonishing diversity that has grown over its long history. This can most quickly and realistically be achieved through the use of precise language by *yoga* teachers and schools in educational, marketing and public relations materials on the web and other media outlets.

Finally, linguistic evidence from websites and blogs suggests that the use and meaning of the term *yoga* is broad and multifarious, just as there are types of people with different ways of understanding, all with different needs and proclivities. However, if the meaning and use of the term *yoga* is to be understood in any meaningful way in the rapidly expanding online, worldwide community, linguistic precision concerning what ‘*yoga* is’ and ‘is not’ becomes a necessary conversation. It is my hope that this research has added something useful to that discussion.

References

- Alliance, Y. J. a. Y. (2016). 2016 Yoga in America Study Conducted by Yoga Journal and Yoga Alliance Reveals Growth and Benefits of the Practice. Retrieved from https://www.yogaalliance.org/Get_Involved/Media_Inquiries/2016_Yoga_in_America_Study_Conducted_by_Yoga_Journal_and_Yoga_Alliance_Reveals_Growth_and_Benefits_of_the_Practice
- Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics*, 1(1), 1-37.
- Basavaraddi, I. V. (2015). National Institute of Yoga. Retrieved from https://www.nhp.gov.in/yoga_mty
- The Bhagavad Gita (E. Easwaran, Trans.). (2007). In. California, USA: Nilgiri Press.
- Davies, M. (2013a). Corpus of Historical American English. Retrieved from <https://www.english-corpora.org/coha/>

- Davies, M. (2013b). Corpus of Web-Based English. Retrieved from <https://www.english-corpora.org/glowbe/>. <https://www.english-corpora.org/glowbe/>
- Davies, M. (2018). The iWeb Corpus. Retrieved from <https://www.english-corpora.org/iweb/>.
- Deignan, A. (2005). *Metaphor and Corpus Linguistics* (Vol. 6); John Benjamins Publishing.
- Deslippe, P. (2018). The Swami Circuit: Mapping the Terrain of Early American Yoga. *Journal of Yoga Studies*, 1, 5-44.
- Deslippe, P. (2019). Yoga Landed in the U.S. Way Earlier than You'd Think - And Fitness Was Not the Point. Retrieved from <https://www.history.com/news/yoga-vivekananda-america>
- Feuerstein, G. (2002). *The Yoga Tradition: Its History, Literature, Philosophy and Practice*: Laurier Books, Limited.
- Feuerstein, G. (2003). *The Deeper Dimension of Yoga; Theory and Practice*. Boston, USA: Shambala.
- Gabriel, R. (2017). 6 Important Lessons from the Ashram [Online Website]. 2021. Retrieved from <https://chopra.com/articles/6-important-lessons-from-the-ashram>
- Gast, V. (2006). *The scope and limits of corpus linguistics : empiricism in the description and analysis of English*. Würzburg: Königshausen & Neumann.
- Gilquin, G., & Gries, S. (2009). *Corpora and experimental methods : a state-of-the-art review*.
- Gonda, J. (1977). *The Ritual Sutras* (J. Gonda Ed.). Wiesbaden: Otto Harrassowitz.
- Guruji, O. (2018). *Yoga Through Third Eye*. Indore, India: Institute of Yoga Sciences & Research.
- Hammond, H. (2007). The Timeline and History of Yoga in America. Retrieved from <https://www.yogajournal.com/yoga-101/yogas-trip-america/>
- Hunston, S., & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*: John Benjamins Publishing.
- Iyengar, B. K. S. (2001). *Light on Yoga*. London, UK: Thorsons.
- Mallinson, J. S., M. (2017). *Roots of Yoga*. Great Britain: Penguin Random House UK.
- Marshall, B. (2009). *The Perennial Way* (Expanded Edition ed.). North Carolina, USA: Realface Press.
- O'Neill, M. (2015). *On Yoga; The Architecture of Peace*. Germany: Taschen.
- Patanjali. (2012). The Yoga Sutras of Patanjali (S. S. Satchidananda, Trans.). In *The Yoga Sutras of Patanjali*. Virginia, USA: Integral Yoga Publications.
- Simpson, D. (2021). *The Truth of Yoga*. New York, USA: North Point Press.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*: Oxford University Press.
- Therapists, I. A. o. Y. (2020). Retrieved from <https://www.iayt.org/>
- The Upanishads (E. Easwaran, Trans.). (2007). In *The Upanishads: Introduced and Translated by Eknath Easwaran*. Canada: The Blue Mountain Center of Meditation.
- The Yoga Sutras of Patanjali (S. Satchidananda, Trans.). (1984). In *The Yoga Sutras of Patanjali: Translation and Commentary by Sri Swami Satchidananda*. Yogaville, Virginia: Integral Yoga Publications.
- Zygomatic (Producer). (11/30/2020). WordClouds.com.