



The contextual factors of big, large and great in contemporary written and spoken texts: a corpus based analysis

Mercedita M. Reyes

University Science High School, Central Luzon State University, Science City of Muñoz, Nueva Ecija, Philippines
armier23@yahoo.com

Available online at: www.isca.in, www.isca.me

Received 3rd May 2017, revised 9th July 2018, accepted 17th July 2018

Abstract

Presumably, humans in general and not only language learners in classroom tend to choose words that are easily and readily retrievable in their minds whenever they do oral communication. This is relatively common in their second or foreign language. In the Philippine context, this has become the case in the use of the English language; hence some common synonyms have the tendency and are likely to be overused in communicating ideas. This paper evaluated the extent to which commonly used synonymous words of English were used in contemporary written and spoken texts, and it found out what context the difference in usage occurred. The Collins Word banks Online English corpus sampler and Web Concordancer were used. To get the linguistic data from the corpus, the researcher typed in the synonymous words big, large and great one by one and got a display of concordance lines. The output generated from the concordancer was made limited to 40 lines of concordance, with a maximum width of 250 characters. The sampled lines were selected at random. The demo facility provided and restricted the display to around 100 of collocates. These were the collocates that were found statistically most significant according to T score calculation. The study showed that most occurrences or appearances as per use of the word big in the native corpora were used to refer to physical size of objects, specifically following concrete nouns: big brother, big ears, big teeth, big tax, big boxers, big wall and the like. Large was preferred for abstract nouns: large quantity, large payments, large extent, large range, large discrepancy, large explosion, and large amounts. Great was preferred when ideas or concepts were intensified, hence the use of greater and greatest: greater power, greater press freedom, greater insights, greatest incentive, greatest business successes, greatest of all politicians. These outputs were taken from collocates that were statistically most significant according to T score calculation. Big, large and great were classified by their function in sentences. From among the three synonymous words, great, large, and big, large was used more often than big and great. Anent the observed phenomenon that some collocations with big happened when more specific and concrete meanings were conveyed and that they were likely to occur on cases where more concise expressions were preferred, hence the attribution of more concrete nouns, while abstract ones went with large and intensity was better shown by contemporary writers and speakers using great. But among the three, great as a modifier appeared to be more descriptive. It is inflected to show varying degrees, as in: greater than its minimum, greatest allocation, greatest of all, among others. Since limited analysis was done on the context of the three synonymous words, the specific function of synonymous words and the patterns of sentences into which they occurred were not examined. Further studies could be facilitated to augment the investigation on this part for more reliable findings and accurate descriptions on the use of language.

Keywords: Corpus-based analysis, synonyms, big, large and great, English language, contextual factors, collocates.

Introduction

Teacher's exploration in the language classroom is limitless, from simple to complex. The inability of students to express themselves using the target language always calls for introspection. More if some other greater concerns like syntax, grammatical construction, and lexicon are at play.

A teacher can investigate not only how the language is used, but also the level and degree that these patterns are being observed, and the many factors in context that directly and indirectly control students' usage and preference. Presumably, one could examine how often a particular grammar construction is used in

speaking versus writing or as it appeared in newspapers, magazines and the like versus fictional materials. Or one might want to examine the use of synonyms like destroy, terminate and annihilate among others or the usual and simple descriptors, pretty and beautiful or great/big/large and many others in order to establish their context and statistically define frequency of distribution.

Language or linguistic analysis using corpora can deal with almost all patterns of language often with very specific agenda. These agenda could cover areas such as discovering differences between male and female when it comes to usage of contractions in expressing queries, student's learning of past

form of the verb, or counterfactual indirect object switching patterns of students and many more.

Corpus linguistics had shown exemplar in providing a more scientific view of language than that of participatory or distant observation or mere introspection or even analysis based on personal or related anecdotes. Understandably, it was reasoned out that this may be explained by the fact that speakers really do not have primary access to the subliminal patterns through which the language is being used and being decoded¹.

A number of researchers defined collocations from the perspective of partnership or group occurrence of words². Speakers or users of the language often fell flat in choosing the more appropriate, if not correct arrangement of words due to lack or absence of know how in the collocation properties and to their extent in the given vocabulary. Language learners are more likely to adopt simplified strategies in lexicon such as using a synonymous or L1- translated expression. Collocation may be defined as the customary appearances or succession of two or more words in a text, and is an essential element of language learning. However, due to the traditional grammar-based EFL pedagogy, the collocation property in relation to each item of vocabulary has been neglected in language class. When learners encounter a collocation problem, they tend to resort to one of the strategies of lexical simplification: synonym, avoidance, transfer and paraphrasing. She found out in her study that the most commonly used strategy; synonym taking can be viewed as a direct consequence of the unawareness of collocation borders between and among items of lexicon³.

Since learners tend to choose whatever word is easily and readily retrievable in their minds when putting the strategies in practice, certain commonly-used synonymous words are more likely to be overused in communicating ideas in the English language.

This paper examined how the commonly used synonymous words of English were used in contemporary written and spoken texts, and it found out what context the difference in usage occurred. This work investigated on this issue over the set of *great*, *large* and *big*.

Statement of the problem: This paper determined the contextual factors of synonymous words *big*, *large* and *great* as they were used in contemporary written and spoken texts. Specifically, it attempted to answer the following: i. How are related constructions *big*, *large* and *great* used differently in contemporary written and spoken texts? ii. How common are these related constructions in contemporary written and spoken texts? iii. Are these constructions used frequently in different registers of contemporary written and spoken texts? iv. Are there particular words that these constructions commonly occur with in contemporary written and spoken texts? v. What are the contextual factors that go with these constructions in contemporary written and spoken texts?

Methodology

The researcher used the descriptive method of research which is designed to gather information about present conditions. The purpose of which is to find new truth. In this work, the Collins Word banks Online English corpus sampler and Web Concordancer were used. The Collins Word banks Online English corpus sampler was reported to be composed of 56 million words of contemporary written and spoken texts, divided into sub-corpora: British books, ephemera, radio, newspapers, magazines with 36 million words, American books, ephemera and radio with 10 million words and British transcribed speech with 10 million words⁴.

To get a flavor of the type of linguistic data that a corpus like this can provide, the researcher typed in the synonymous words *big*, *large* and *great* one by one and got a display of concordance lines from the corpus, which were then downloaded in printable version for appending purposes. The lists are found in the Reference section.

The output generated from this demo facility was restricted to 40 lines of concordance, each with a maximum width of 250 characters. The lines displayed were selected at random. The demo facility provided and restricted the display to 100 collocates. These were the collocates that were found to be statistically most significant according to T score calculation. Meanwhile, the Web Concordancer had searched Articles1.txt for the synonymous words. To limit the search, from among the four options displayed: Equal to, Starts with, Ends with and Contains, the researcher opted the key word Contains to navigate the available data or the corpora. There were 1355 instances of *big*, 2104 instances of *great* and 1557 instances of *large*, all in one hit.

Results and discussion

In contemporary written and spoken texts, *Big* tends to describe a happening or a type of person: *big increase*, *big brother*, *big bike*, *big fan*, *big fight*. In the same case with other adjectives, *big* has displayed both the attributive and predicative functions; its attribution (*big* + noun) as in this example: *big eyes* and its predicative function (noun + be + *big*) like in this set: *focus group will be too big*. *Large* is used with nouns concerning numbers and measurements: *large amounts*, *large number*, *large quantities*. *Great* describes nouns which express feelings or qualities: *great accuracy*, *great fun*, *great injustice*, *great climax*.

The frequency of *great* indicates the highest occurrence with 2104 instances while the frequency of *big* and *large* was not too distant from each other, generating 1355 and 1557 instances respectively. To evaluate, analyze and assess in a bigger scale the collocation property of *big*, *large* and *great* in the learner corpus, the corresponding collocates and their frequency distribution are shown in tables.

Table-1: Collocates of *Big* and frequency distribution.

Collocate	F
A	973489
Big	17260
Enough	15034
One	132786
Too	29066
S	484385
Deal	8791
Is	407114
There	149133
Business	13842
Difference	4193
Very	60986
Problem	13145
Like	91152
Companies	6188
Money	20707
Bang	581
Was	340423
Ones	5290
Really	36430
Great	21745
Names	3682
Thing	21689
Got	49055
Game	10082
It	494702
Man	29372
As	246656
Blue	5544
Star	6198
Quite	19835
Make	37874
With	290716
Fish	4273
Toe	525
words	9471
city	15855
house	22752
Small	15323
Screen	2531
Hit	6841
Brother	4005
Issue	7153
Breakfast	2050
The	2313407
Mistake	1645
Navajo	392
Band	4279

Fat	2785
Fan	1273
Guns	1395
Black	14690
Apple	946
Boys	4756
Question	9837
Such	31861
Cities	2021
How	50009
Race	5428
Eyes	8354
Issues	5055
Cuts	2310
Story	8155
Red	8357
Dog	3791
Step	4550
Break	6252
Making	12388
Surprise	2509
Bucks	301
Banks	2721
Match	6088
Spending	2998
Firms	1541
Win	8105
Budget	3608
Mac	407
Occasion	1855
Guy	3315
Isn	11173
For	376522
Name	28566
Boy	6055
Club	10579
Houses	3673
impact	2459
picture	9031
chub	670
corporations	329
breasts	590
word	7972
blow	1925
ears	1174
ben	1564
barbel	294
players	6036
hits	1048
boost	1258

Table-2: Collocates of *Large* and frequency distribution.

Collocate	F
A	973489
Of	1100578
Number	21954
Scale	2682
Very	60986
Numbers	4482
Small	15323
With	290716
Amounts	1122
Part	21256
Enough	15034
And	1129483
Companies	6188
Quantities	489
Large	10599
Room	13713
Amount	5483
In	765730
Extra	5094
Medium	1373
By	181034
Areas	6190
Sums	571
Are	198957
Money	20707
Quite	19835
Area	12877
Such	31861
Size	4814
Proportion	994
Flowers	2952
Population	2917
Extent	2284
Which	109080
White	15296
Group	16895
Bowl	1526
Sum	1160
Family	21020
House	22752
Red	8357
Quantity	492
Majority	3442
1	32618
Leaves	3256
Sizes	1956
Black	14690
Fairly	2777
Garden	6027

Cities	2021
Too	29066
Crowds	714
Eyes	8354
Measure	2349
Colour	4055
Windows	2147
Oil	5342
There	149133
Chunk	220
Selection	2417
Screen	2531
X	3645
Groups	6616
Or	156128
Chopped	625
Unusually	408
Onion	395
Pan	1300
Chunks	170
capacity	1737
corporations	329
houses	3673
trees	2792
Plastic	2516
Rooms	2578
Parts	3980
Shaped	1186
Crowd	2401
Families	4702
Writ	177
Rather	14112
Percentage	879
Pieces	2476
Pool	2194
Glass	3825
Portion	504
Store	2680
firms	1541
projects	1884
section	3665
yellow	2526
sae	186
doses	250
volume	1510
double	5180
loomed	135
heavy	4414
eggs	1230
hall	5588
company	16343

Table-3: Collocates of *Great* and frequency distribution.

Collocate	F
A	973489
Deal	8791
Of	1100578
Britain	13506
Great	21745
Is	407114
Was	340423
There	149133
Fun	4073
S	484385
Many	37516
Hall	5588
It	494702
With	290716
Success	5454
Value	5543
Western	5429
Grandfather	633
One	132786
Big	17260
Difficulty	1752
Interest	10305
Pleasure	2120
Street	11818
Majority	3442
importance	2136
Extent	2284
For	376522
Advantage	2564
Depression	1615
Pride	1548
War	16618
Has	124665
too	29066
white	15296
grandmother	784
man	29372
The	2313407
Ormond	72
Lengths	509
British	18587
grandchildren	354
Uncle	1159
Thanks	4549
Opportunity	3974
This	190140
Tradition	1854
Truly	1773
Okay	14530

Shame	1317
Care	10080
Expectations	1226
Adventure	1021
Relief	2850
grandparents	334
Had	151347
Made	36375
Source	3005
Sounds	3676
Aunt	598
Sense	9766
Joy	1722
Strength	2692
Yarmouth	62
Detail	1845
Grandson	269
Honour	980
Outdoors	274
Really	36430
achievement	1120
Our	52043
Debate	2785
Effort	3487
Skill	1349
Divide	662
Believer	159
Surprise	2509
Delight	842
excitement	1056
Escape	2000
Idea	9490
Emphasis	1310
Player	4078
Gift	2448
fan	1273
danger	2371
concern	3204
character	3097
friend	6807
mandala	60
asset	644
power	11720
but	224012
potential	3416
looks	5907
pity	729
love	14038
history	8769
century	6885
team	11707

The study showed that most appearances and occurrences of *big* in native corpora were used to refer to physical size of objects, specifically following concrete nouns: *big brother, big ears, big teeth, big tax, big boxers, big wall* and the like. Large was preferred for abstract nouns: *large quantity, large payments, large extent, large range, large discrepancy, large explosion, large amounts*. Great was preferred when ideas or concepts were intensified, hence the use of greater and greatest: *greater power, greater press freedom, greater insights, greatest incentive, greatest business successes, greatest of all politicians*.

These outputs were taken from collocates that were statistically most significant according to T score calculation. Big, large and great were classified by their function in sentences. But among the three, *great* seems to be more descriptive. It is inflected to show varying degrees, as in: *greater than its minimum, greatest allocation, greatest of all*. It was also found to be more attributive.

Conclusion

This paper investigated a set of corpora and examined how commonly used words of English were used in contemporary written and spoken texts, and found out the context wherein the difference in usage occurred.

Investigating this issue over a set of synonymous words, *big, large* and *great*, the results gave the following findings. From among the synonymous words, *big, large* and *great*, *large* was used more often than *big* and *great*. Additionally, the occurrence of collocations with *big* was observed when more specific meanings were expressed and that they were more likely to be present on cases where more concise expressions were preferred, hence the attribution of more concrete nouns, while

abstract ones went with *large* and intensity was better shown by contemporary writers and speakers using *great*.

In this work, limited analysis was done on the context of the three synonymous words. The specific function of synonymous words and the patterns of sentences into which they occurred were not examined. Further studies could be facilitated to augment the investigation on this part for more reliable findings and accurate descriptions of the words subjected into scrutiny.

References

1. Sinclair J., Mason O., Ball J. and Barnbrook G. (1997). Language independent statistical software for corpus exploration. *Computers and the Humanities*, 31(3), 229-255.
2. Hashemi M., Azizinezhad M. and Dravishi S. (2012). Collocation a neglected aspect in teaching and learning EFL. *Procedia-Social and Behavioral Sciences*, 31, 522-525.
3. Shih R.H.H. (2000). Collocation deficiency in a learner corpus of English: From an overuse perspective. In Proceedings of the 14th Pacific Asia Conference on Language, information and Computation, 281-288.
4. Krieger D. (2003). Corpus linguistics: What it is and how it can be applied to teaching. *The Internet TESL Journal*, 9(3). <http://iteslj.org/Articles/Krieger-Corpus.html>, March 2003.
5. Stern H.H. (1992). Issues and Options in Language Teaching. Oxford University Press. ISBN-13: 978-0194370660 Hans Jørgen Klarskov Mortensen, ICT4LT Website, Vordingborg Gymnasium, Denmark.
6. O'Donnell M.B. and Römer U. (2009). Michigan corpus of upper-level student papers.