

# Linggle

Calendar 2017



A Year of Language Exploration with  
Linggle, a linguistic search engine

Jason S. Chang

## January

S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

## February

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28				

## March

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

## April

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23/30	24	25	26	27	28	29

## May

S	M	T	W	T	F	S
1	2	3	4	5	6	
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

## June

S	M	T	W	T	F	S
						1
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

## July

S	M	T	W	T	F	S
						<b>1</b>
<b>2</b>	3	4	5	6	7	<b>8</b>
<b>9</b>	10	11	12	13	14	<b>15</b>
<b>16</b>	17	18	19	20	21	<b>22</b>
<b>23/30</b>	24/31	25	26	27	28	<b>29</b>

## August

S	M	T	W	T	F	S
			1	2	3	<b>4</b>
<b>6</b>	7	8	9	10	11	<b>12</b>
<b>13</b>	14	15	16	17	18	<b>19</b>
<b>20</b>	21	22	23	24	25	<b>26</b>
<b>27</b>	28	29	30	31		

## September

S	M	T	W	T	F	S
					1	<b>2</b>
<b>3</b>	4	5	6	7	8	<b>9</b>
<b>10</b>	11	12	13	14	15	<b>16</b>
<b>17</b>	18	19	20	21	22	<b>23</b>
<b>24</b>	25	26	27	28	29	<b>30</b>

## October

S	M	T	W	T	F	S
<b>1</b>	2	3	4	5	6	<b>7</b>
<b>8</b>	9	10	11	12	13	<b>14</b>
<b>15</b>	16	17	18	19	20	<b>21</b>
<b>22</b>	23	24	25	26	27	<b>28</b>
<b>29</b>	30	31				

## November

S	M	T	W	T	F	S
			1	2	3	<b>4</b>
<b>5</b>	6	7	8	9	10	<b>11</b>
<b>12</b>	13	14	15	16	17	<b>18</b>
<b>19</b>	20	21	22	23	24	<b>25</b>
<b>26</b>	27	28	29	30		

## December

S	M	T	W	T	F	S
					1	<b>2</b>
<b>3</b>	4	5	6	7	8	<b>9</b>
<b>10</b>	11	12	13	14	15	<b>16</b>
<b>17</b>	18	19	20	21	22	<b>23</b>
<b>24/31</b>	25	26	27	28	29	<b>30</b>

Winter, spring, summer or fall. Be sure to Linggle for recurring phrases in 1 trillion-word Web pages  
— Linggle searches as you type away, displaying results in a fraction of a second.

## Linggle Search Elements

Linggle is a Web-scale linguistics search engine, that retrieves lexical bundles in response to a given query.

### Search Elements

\* match zero or more words

[ play \* role ] --> [ play an important role ]

\_ match any word

[ listen \_ music ] --> [ listen to music ]

[ present a method \_ ] --> [ present a method for ]

~ synonyms

[ ~important role ] --> [ significant role ]

? optional word

[ listen ?to music ] --> [ listen to music ]

[ discuss ?about this ] --> [ discuss this ]

/ alternative word

[ go in/to school ] --> [ go to school ]

[ strong/heavy rain ] --> [ heavy rain ]

### part-of-speech

search for word with specific pos tag

[ v. a report ] --> [ submit a report ]

search for word with specific pos tag with  
search elements

[ shops ?prep. the street ] --> [ shops along the street ]

v. (verb) n. (noun) adj. (adjective) adv. (adverb)

prep. (preposition) det. (determiner)

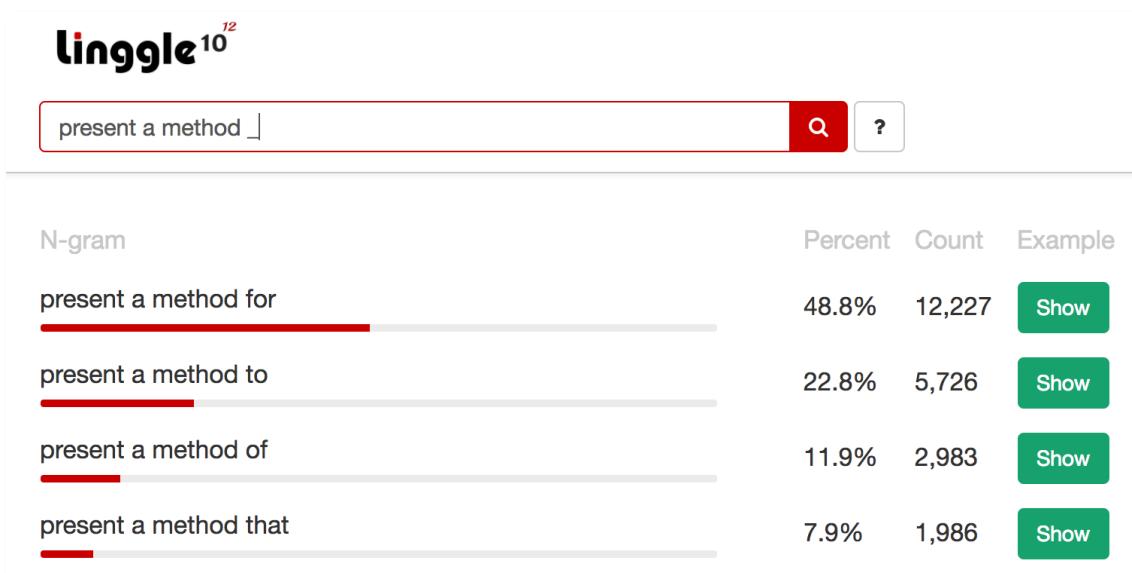
conj. (conjunction) pron. (pronoun)

## 1 Writing Ahead

Typing a few words and then a couple of underscores (e.g., <present a method \_> or <present a method \_\_>) allows you to find recurring phrases following the words. The results show that

- **present a method for (doing) something** or
- **present a method based on something**

are common grammar patterns. Note that a query for more than 5 words (e.g. <present a method \_\_\_>) returns nothing.



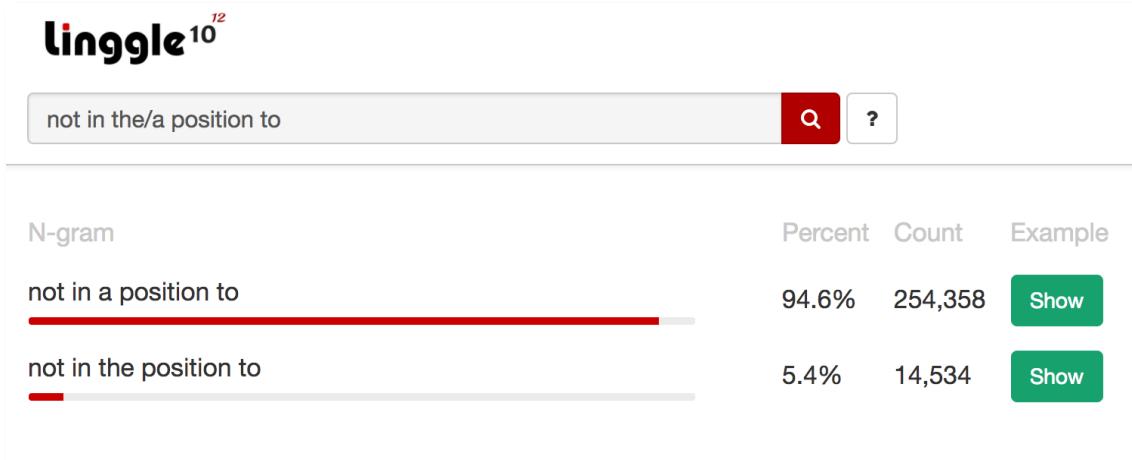
## 2 Checking whether a word is needed

A question mark followed by a word (e.g., **to**) allows you to check whether the word is needed in a phrase, as in <go ?to ask your mom> (see the figure).



### 3 Deciding on alternative phrases

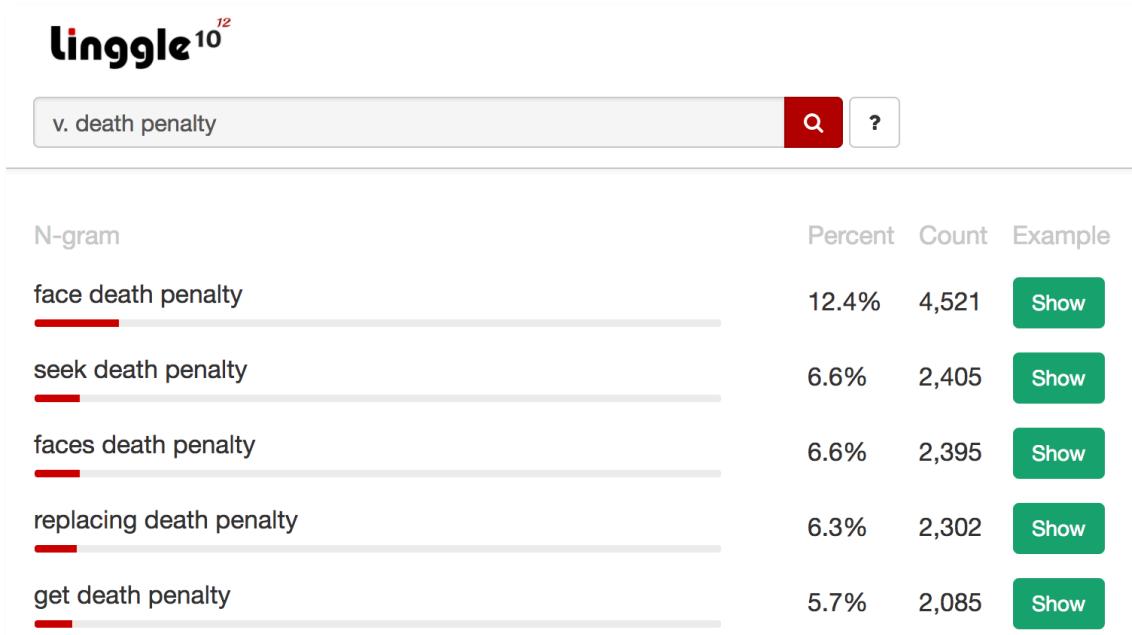
One of the test items of a *College Entrance English Exam* of Taiwan (2013), contains a problematic phrase, **not in the position to**, according to an expert. Linggle <**not in the/a position to**> and you will get the frequency ratios of **not in the position to** (4%) and **not in a position to** (94%). Phrases with a high frequency count are usually preferred.



### 4 Finding Collocations

Linggle's unique and very best feature is that keywords can be combined with any part of speech to create a search query for finding collocations. For example, Linggle <**v. death penalty**> to find verb collocates of death penalty. The results are on par with what is available in a collocation dictionary.

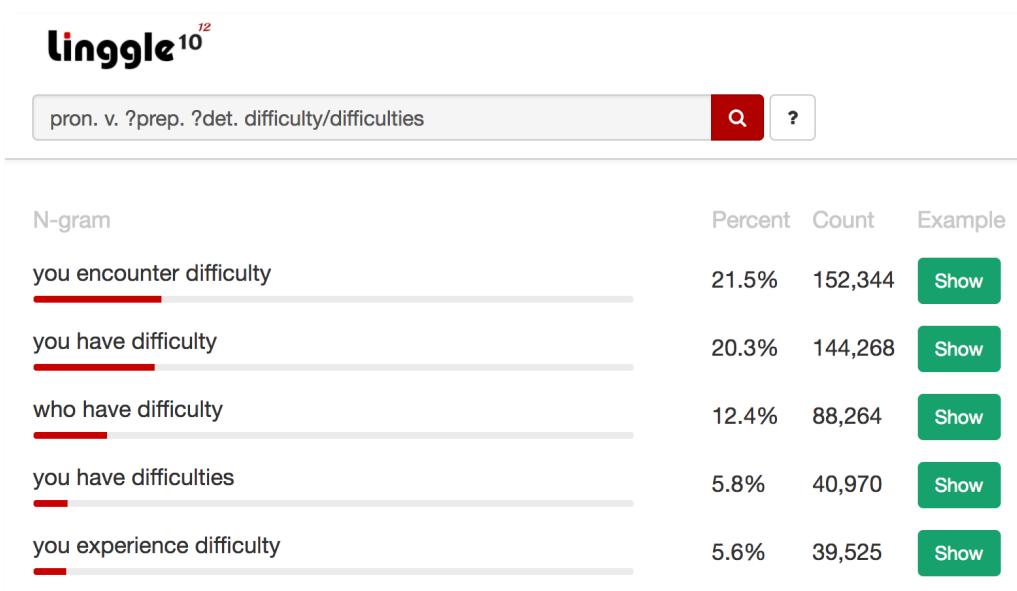
Recently the uncommon phrase graduation threshold has been used on websites of several universities in Taiwan. A simple Linggle query, <**graduation n.**> will bring to your attention the correct expression, **graduation requirements**, right on top of the search results.



## 5 Precise collocation queries

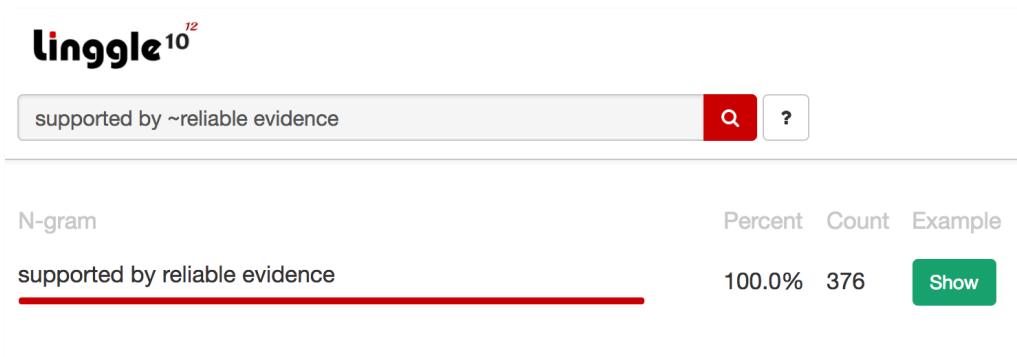
Linggle uses a simple strategy to handle queries that involve parts of speech. Yet, POS can be ambiguous at times, leading to less than perfect results when it comes to POS queries. For example, the query **<v. difficulty>** returns **learning difficulty** as the top result, while **<play n.>** brings back **play station!** In fact, both should be classified as NN.

There are other problems with using a simple search query like **<v. difficulty>**: missing verbal phrase (**run into**) and a lack of information related to articles in the retrieved phrases. For this, it would be better to use a slightly more complex query, e.g., **<pron. v. ?prep. ?det. difficulty>**. Here **pron.** is added so that **v.** can match a real verb (instead of seemingly correct verbs like **learning**). The optional POS wildcard **?prep.** after **v.** allows phrasal verbs (e.g., **run into**) to be found. The **?det.** before **difficulty** allows us to find the use of articles. See Page 5 for precise collocation queries, using **difficulty** and **acquire** as examples.



## 6 Searching for synonyms

A tilde before a keyword (e.g., **<~reliable>**) allows you to search for synonyms. For synonyms appearing in a certain context, simply add words before or after the keyword (e.g. **<supported by ~reliable evidence>**). You will find some relevant synonyms in context: e.g. **credible**, **verifiable**, and **robust**. You could also verify this: **a reliable friend**, yes; **a robust friend**, no. Press one of the plus signs, and you will see some example sentences (and the + sign turns into the — sign). Press the — sign and the examples disappear. *Synonyms in context with on-demand examples*, something you can not get with a paper thesaurus.



## 7 Well-spoken phrases

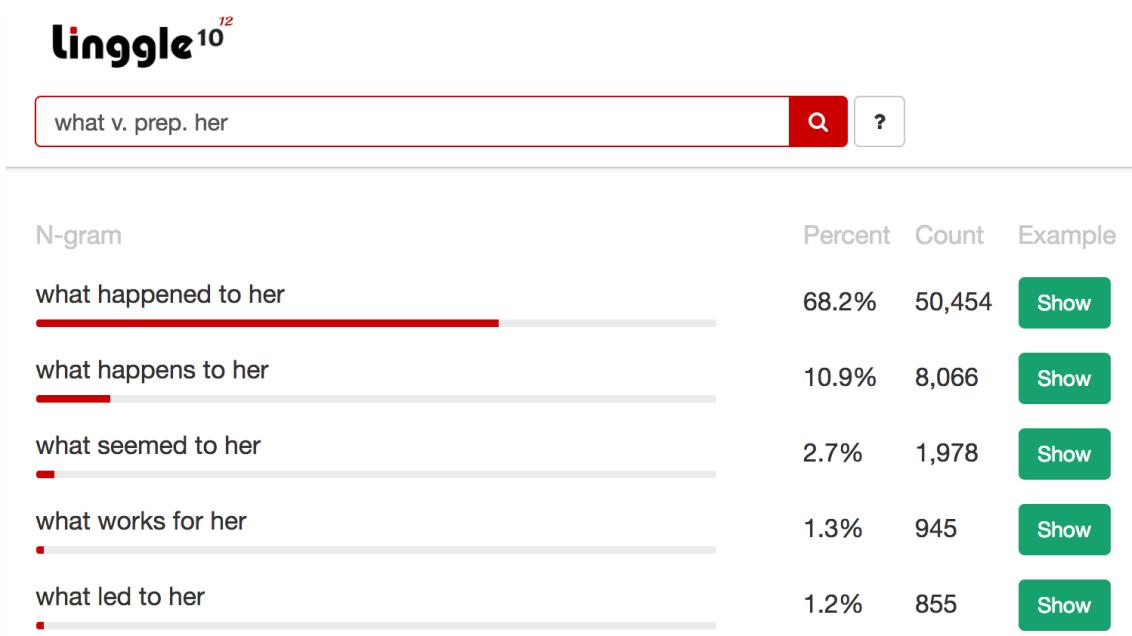
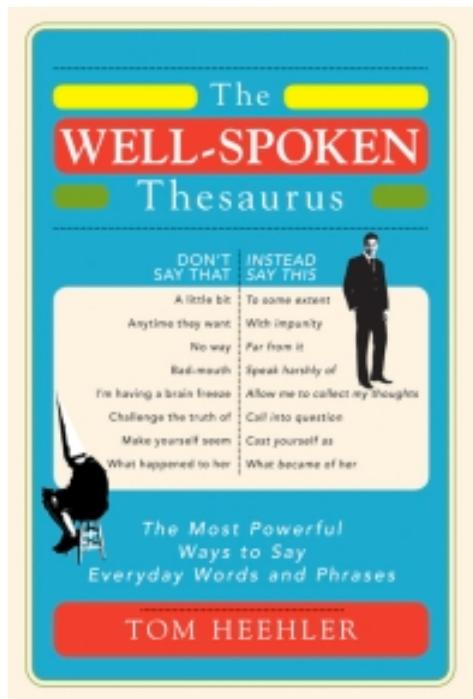
Sometimes, we are not just casually looking for any synonym, but rather phrases spoken by well-spoken people. While attending the Harvard Extension School, Tom Heehler, found that the eloquence of Harvard students put him to shame, spurring him to spend three years taking note of *Harvardian language*. The result is "The Well-Spoken Thesaurus" (TWST) full of fancier phrases like **what became of her** (to upgrade the commonly-used **what happened to her**).

In other words, when searching for synonyms, we are often looking for what are called powernyms. Let us try <**what v. prep her**> to see if Linggle also can be used to find powernyms as with TWST. Sure enough, Linggle returns both **what happened to her** and **what became of her**. From the search results, it is apparent that **what happened to her** is ranked the highest, while **what became of her** falls somewhere in the middle. With its ranking mechanism, Linggle comes in handy, at times, for finding powernyms.

On that note, you would expect to find powernyms in a New York Times article, "**36 Hours in Taipei, Taiwan**", describing the must-see 101 tower and its surrounding:

In every direction lie city blocks and avenues winding among concrete-and-glass towers, with verdant hills rising in the distance. (Source: <http://goo.gl/EUekse>)

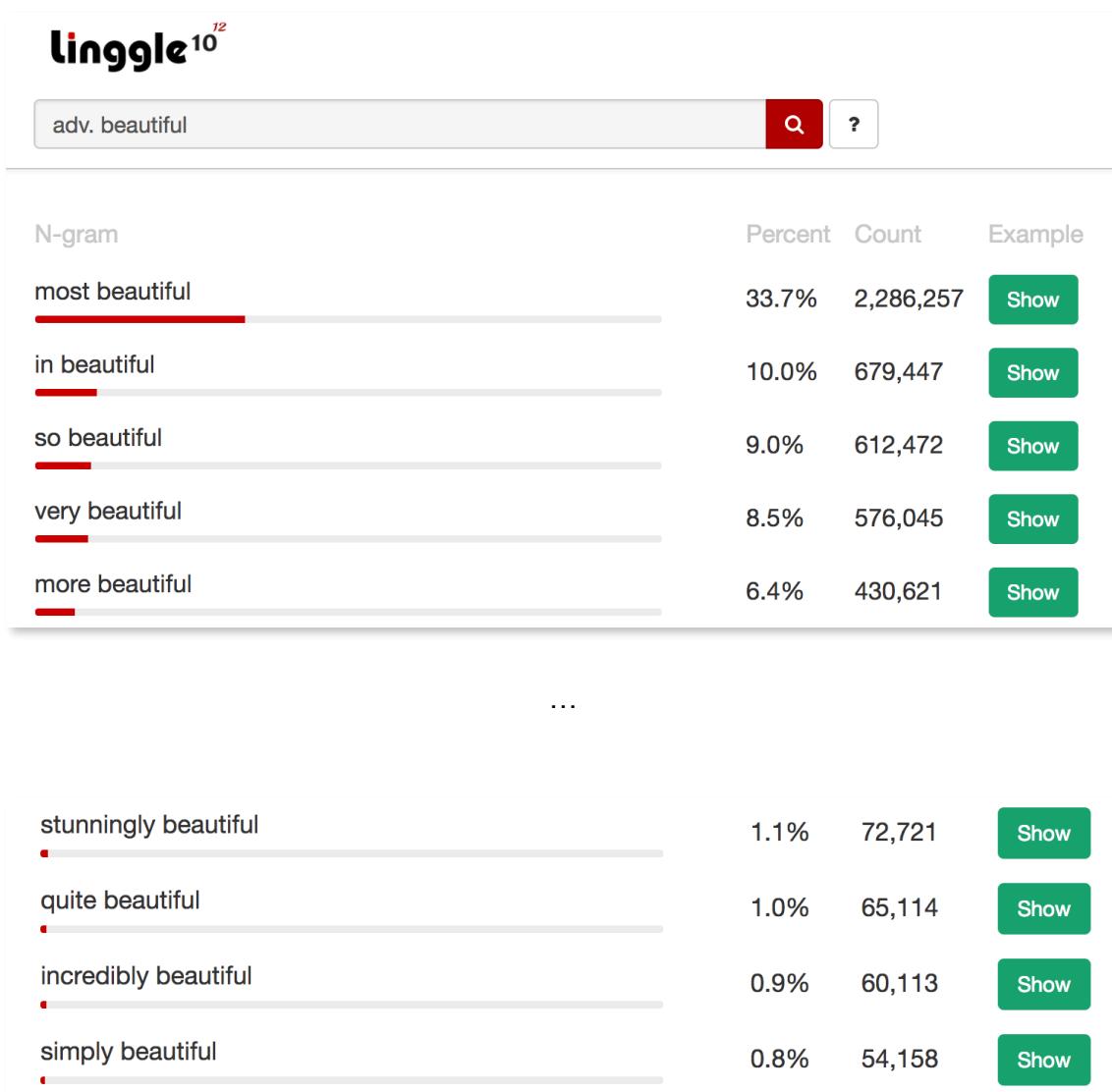
Normally, **green hills** would probably be used to describe the rural scenery, but NYT uses **verdant hills** to describe the view seen from the 89th floor of 101. Sure enough, checking Heehler's TWST, you will find, on page 181, <**green verdant**> followed by <**greenhouse conservatory**> .



## 8 Always favor higher frequency?

Looking back at the example, **what happened to her**, the answer is, "not necessarily." Consider for a second how novelists would describe their female characters. Simply describing them as **beautiful** is definitely not enough. You better add an adverb to clarify and amplify—just how beautiful, and in what way beautiful. You could of course get a copy of the *Macmillan Collocations Dictionary* to find<sup>9</sup> out, but just Linggling **<adv. beautiful>** will get you a lot of mileage, and faster.

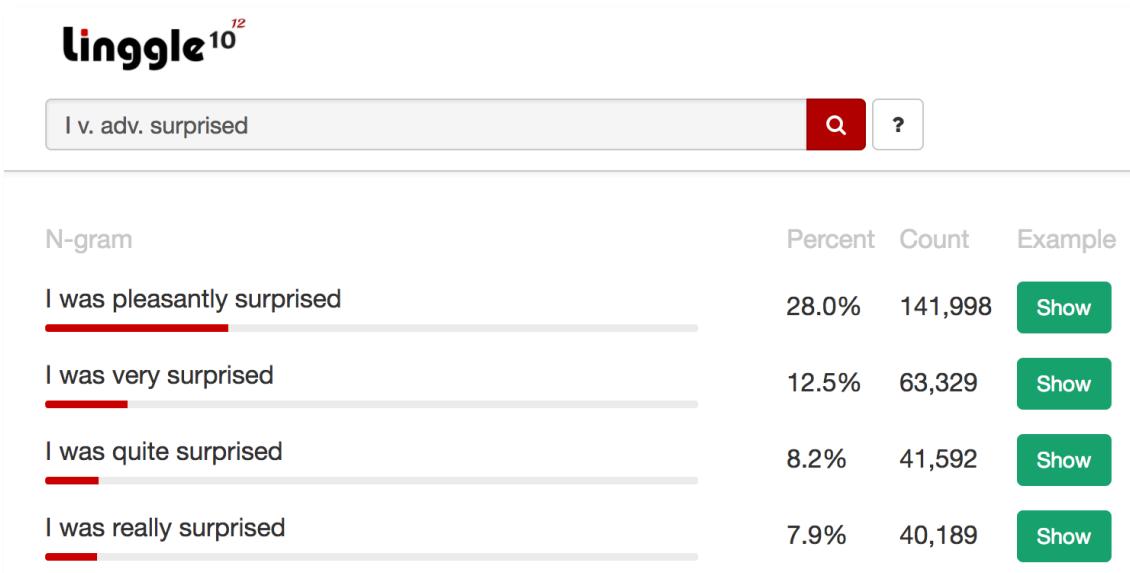
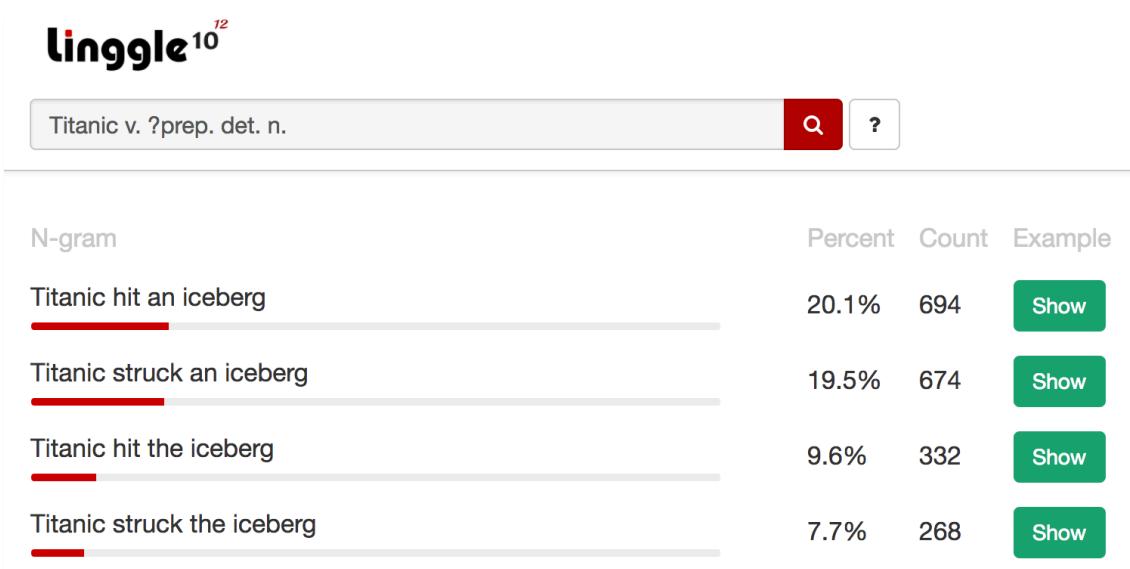
From search results of **<adv. beautiful>** (on the right) we can see that the phrases with the highest frequency, namely **most, so, very, how beautiful** are all too mundane. Pulitzer Prize winner, Michael Cunningham used **simply beautiful** in his novel, *The Hours*. Usually, words with middle frequency (e.g. **stunningly, incredibly, simply**) are more specific and creative.



## 9 Tip-of-the-tongue word

Perhaps you have experienced this: when writing an email, you wanted to use a phrase that you kind-of know: **I was presently surprised**. But then you felt like something was off. Perhaps the word **presently** was wrong. At times like this, you could turn **presently** into an **adv.**, and just Linggle <**I was adv. surprised**>. Voila, mystery solved. Looks like you spelled **pleasantly** as **presently**. It is surprising that many people made the same mistake. Google search showed 148,000 Websites contain this error (as of Dec. 25, 2015).

I came across a flyer with the line, The airlines Titanic that running into an ice **mountain**. I asked myself: do **ice mountain** and **that running** sound right? Linggle <**Titanic v. ?prep. det. n.**> to find out. Using **ice mountain** for **iceberg** would be strange. With Linggle's help, you could write better: **The Titanic of the airline industry that struck an iceberg**.



## 10 Building up word power with Linggle

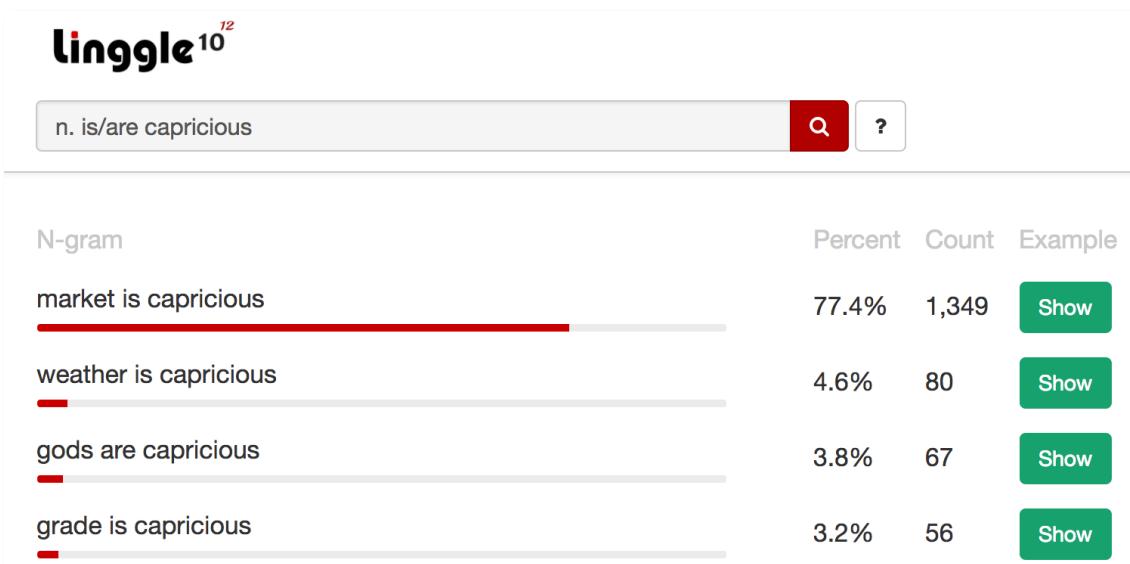
Suppose you want to learn the adjective **capricious**. You can Linggle <n. is/are capricious> to find out things that are **capricious** (see the figure on the right):

- capricious **market**, capricious **God**, capricious **weather**
- capricious **grade**, capricious **rule**, capricious **definition**

We can also Linggle <adj. and capricious> to find adjectives related to **capricious**:

- **arbitrary** and capricious, **mooody** and capricious, **whimsical** and capricious
- **unpredictable** and capricious, **wild** and capricious, **sudden** and capricious
- **uncertain** and capricious, **erratic** and capricious, **wanton** and capricious

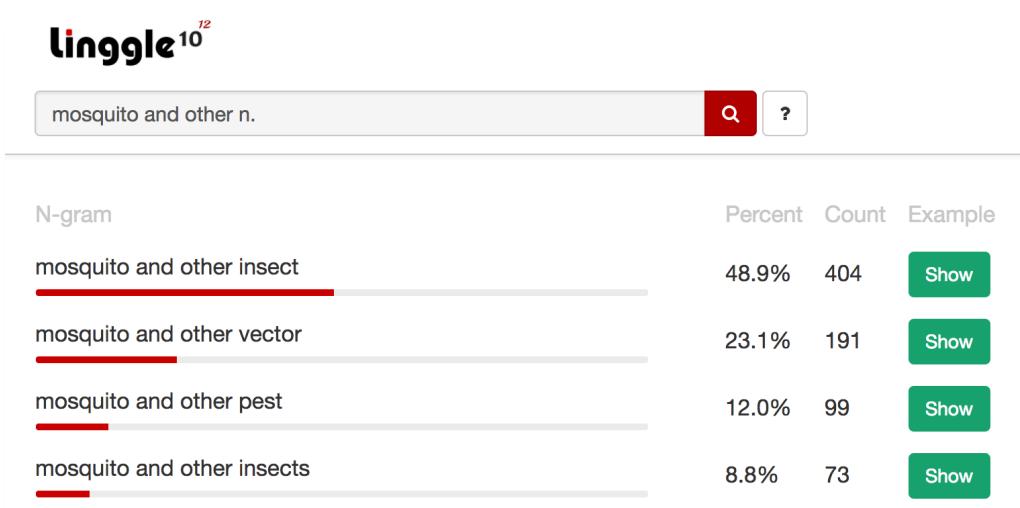
In this manner, it becomes easy to remember that the meaning of **capricious** is **fickle**, what its synonyms are, and that a typical way of using **capricious** is **capricious weather**.



## 11 Using Linggle to paraphrase

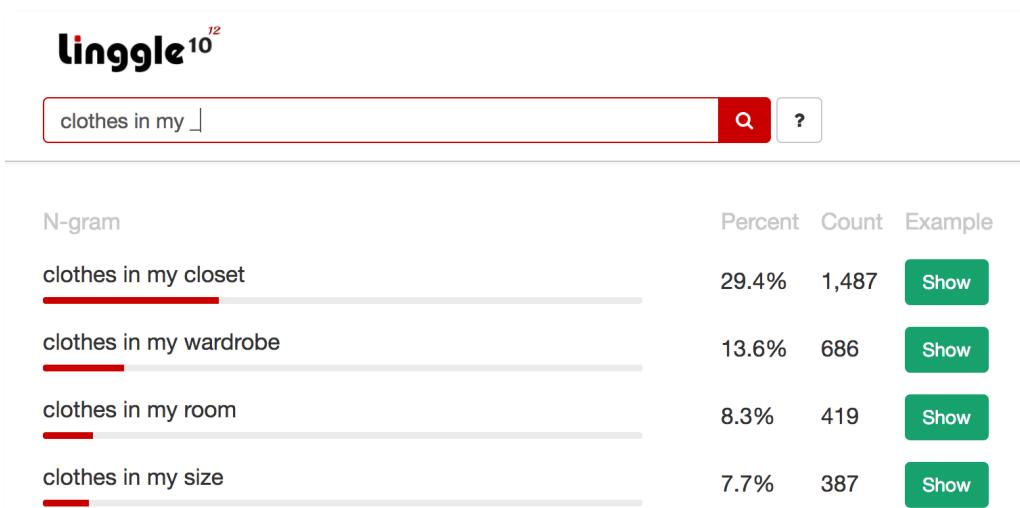
When writing, repeated use of the same word is frowned upon. We use the word **mosquito** to demonstrate how Linggle can help you put it *another way*. You can create a query (or surface pattern), **<mosquito and other n.>**, to find **insects, pests, vectors, bugs**, and **nuisance**, all words that can be used to replace **mosquito**. The word **nuisance** stands out in the list. This paraphrase is interesting here because most thesauri would never place **nuisance** and **mosquito** in the same category.

For compound nouns (e.g. **malaria mosquito**), previous research in computational linguistics has indicated that an additional verb can reveal the relationship between the two words as a way of paraphrasing. For this, we can use **<mosquito that v. malaria>**. From the search results on the right, we can see that we can paraphrase **malaria mosquito** as mosquito **that transmits** malaria, mosquito **that carries** malaria, and mosquito **that spreads** malaria.



## 12 Brainstorming

Linggle can also be used for brain storming. For example, in order to find out what to write about **clothes**, you can start by Linggling **<clothes in my \_>**. From the search results, you would get **closet** and **wardrobe**. Repeated Linggle and/or Google searches would yield sentences containing such relevant words as **out-of-season, built-in cabinets, accessories, tops, denim, drawers, organizers, wash, hue**, and **corral**, words that you would likely to find in a thesaurus, but probably not all of them.





Find out more at [linggle.com](https://linggle.com)



Find out more at [nlplab.cc](https://nlplab.cc)