

The Web phrases collection



The Learning Collocations collection

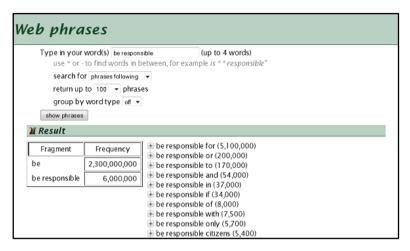
94

7. Language learning resources

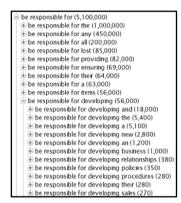
FLAX includes two large language learning resources: Web Phrases and Learning Collocations.

This section examines these resources and how they can be utilized in language learning. They contain authentic written text drawn from various sources and representing contemporary English use. The first, Web Phrases, is built from web text, which has been preprocessed and organized into a searchable database, along with a user interface and flexible searching and browsing facilities to cater for different user needs. The second, Learning Collocations, is built from the British National Corpus, the British Academic Written English and Wikipedia, again preprocessed and organized as a searchable database with a similarly flexible user interface, which in both cases the text.

The next two sections describe the Web Phrases and Learning Collocations resources. Following that, we examine how these resources can be used to support general and academic writing, using a series of scenarios that take place within a genuine teaching and learning context.



Searching for be responsible in Web Phrases



Further exploration of be responsible

```
be held responsible (1.700.000)
be solely responsible (95,000)
be a responsible (50,000)
be more responsible (26,000)
be directly responsible (24,000)
be fully responsible (23,000)
be financially responsible (22,000)
be primarily responsible (21,000)
be made responsible (16,000)
be personally responsible (12,000)
be the responsible (11,000)
be partly responsible (10.000)
be legally responsible (9,200)
be fiscally responsible (8,600)
be socially responsible (7,500)
he partially responsible (6.200)
be ultimately responsible (6.000)
be jointly responsible (5,700)
be hold responsible (5,200)
be largely responsible (4,800)
be entirely responsible (4.400)
be completely responsible (3 900)
```

Results of the search be * responsible

7.1 Web Phrases

The Web Phrases resource was developed from a database containing a vast set of word sequences in the English language, along with their frequencies. The data was collected from publicly accessible Web pages that together comprised approximately one trillion word tokens. The word sequences range from single words to units of five consecutive words. FLAX's Web Phrases resource contains 50,000 unique words, 14 million two-word sequences, 420 million three-word sequences, 500 million four-word sequences, and 380 million five-word sequences. It allows free exploration of word combinations, unconstrained by grammatical class. With it, users can study particular words and phrases to check whether and to what extent what they are writing represents common usage.

Users can find what words most commonly follow a particular word or phrase. The interface contains three parts, shown in the top illustration. Here, the user has typed the phrase *be responsible* into the upper part of the interface. A small statistical table shows the frequency with which this phrase, and any constituent prefix phrases (here, the word *be*), appears in the Web Phrases collection. Beside it is an expandable list (only the first ten lines are shown) that displays associated phrases in reverse frequency order, along with their frequency counts.

The most frequent words following *be responsible* are *for*, *or*, *to*, *and*, etc. (in that order). When the plus sign beside *be responsible for* is clicked, the tree expands as shown in the lower left-hand image, displaying the extensions of these phases, again in frequency order; here, the expanded phrase *be responsible for developing* has also been clicked. In addition the table and graph update accordingly (not shown in the illustration).

A phrase can be expanded up to five words, or until no further extensions are found. Once the phrase length reaches five words, the database contains no further extensions – but if you click on a five word phrase, the system will retrieve from the Web examples of text that use it

The illustration shows searching in the forward direction, but users can also search backwards by specifying the *phrases preceding* option – giving, in this case, phrases like *will be responsible*, *shall be responsible*, *may be responsible*.

Furthermore, an asterisk (*) can be used to stand for any word. This allows users to find words that occur between other words of a phrase. The lower right-hand illustration shows the adverbs that intervene between be and responsible – solely, directly, fully, etc.. Further asterisks can be added, for example, be ** responsible, be *** responsible, and be * responsible * the.

Finally, common words like *the*, *a*, *of*, and *to* are dominant constituents of phrases, and often make it hard for users to discern useful language patterns. The *group by word type* option addresses this problem by letting users look up the words following or preceding a phrase by their grammatical part of speech – preposition, verb, noun, adjective, etc..

¥ used as a noun				
research + noun	 research project research programme market research research findings research team 	237 👶 209 👶 141 🗳	research contract research work research students research interests research methods	103 102 95 82 80 ≫more
adjective + research	social research social research social science research social research asocial research method quantitative social rese.	s	nedical research 159 ch 20 earch 11 ch 4 search 3	90 70 60 59 56 >>> more
noun + of + research	social work research social survey research social survey research social science research comprisal social research contemporary social res social scientific research social scientific research	n search n	esearch rch arch esearch arch	34 26 25 23 22 >>> m ore
research + preposition + noun	social science research social research social policy research in-depth social science	computing project	ts 1 earch 1 n researc 1 between	research 9

	research	hypothesis	scientific	empirical	method	prediction	academic	researcher	
	houraliss	quantitative	iournal	knowledge	criticism	evidence	varv	outcome	
	👶 scientifi	ic method		3		scientific infor	mation	1	
	👶 embodi	ied in the scienti	ic method	1	*	scientific meth	ods	1	
	variety	of scientific matt	ers	1	*	scientific testi	ng	1	
	applicat	ion of the scient	ific method	1	*	classical scien	tific methods	1	
	rounds	of scientific test	ing	1	*	scientific rese	arch	1	
M (🧀 alternat	ive to purely sci	entific methods	1	*	use the scient	tific method	1	
Res	🚑 purely s	cientific		1	4	provides scier	ntific information	1	tablish novel facts, solv
orol	🚑 in the s	cientific method		1	*	funding for sc	ientific research	1	tablish noverracts, son
pioi	ቆ scientifi	ic matters		1					
exte.									
uni Su									
	lefinition	s	-						
M c		_							
M d	earch can b	e defined as t							
M d	earch can b	e defined as t							mind, to establish nove method. [Wikipedia]
M d	earch can b s, solve new	e defined as to or existing pr	oblems, prov						
M c Rese acts exte	earch can b s, solve new nded definit	oe defined as the or existing protions from wikt	oblems, provi ionary						
M c Rese facts exte	earch can b s, solve new nded definit	e defined as to or existing pr	oblems, provi ionary						mind, to establish nove method. [Wikipedia]
M c Rese acts exte	earch can b i, solve new nded definit elated to	oe defined as the or existing protions from wikt	oblems, provi ionary			new theories	, usually using	a scientific	
M c Rese acts exte	earch can b s, solve new nded definit	oe defined as the or existing protions from wikt	oblems, provi ionary			new theories		a scientific	
M d Rese acts exte	earch can b i, solve new nded definit elated to	oe defined as the control of the con	oblems, provi ionary			new theories	, usually using	a scientific	
M c Rese facts exte	earch can b s, solve new inded definit elated to Social so Science	oe defined as the control of the con	oblems, provi ionary			new theories Resi	, usually using	a scientific	
M c Rese facts exte	earch can b s, solve new nded definit elated to Social so Science Academ	oe defined as the control of the con	oblems, provi ionary			Resi Peer Disc	, usually using earch and deve review	a scientific	

Searching for research in Learning Collocations

98

7.2 Learning Collocations

The Learning Collocations resource contains a host of collocations covering most English words. It is split into three databases, each sourcing collocations from a different body of text: standard and general English (the British National Corpus), specialized academic written text (the British Academic Written English Corpus), and a contemporary crowd-sourced resource (Wikipedia). Word combinations that follow 17 common syntactic patterns (verb + noun, noun + noun, adjective + noun, noun + of + noun, etc.) are extracted, sorted by frequency, and presented in an easy-to-use interface, with links to their original context.

This resource works as follows. The user types a term of interest and selects one of the three databases: standard, academic, and contemporary English. In the illustration, the term is *research*, with the contemporary English (Wikipedia) database. The system retrieves and displays collocations and other information about the term.

At the top are family words, that is, inflected and derived forms of the query term — in this case researches, researching, researched, researcher and researchers. Clicking any of these will re-invoke a search using it as the query term. Links to the term's synonyms and antonyms appear on the right: standard resources (WordNet and Roget's thesaurus) are used to identify these.

The illustration shows collocations of the word *research* "used as a noun", and four syntactic patterns are visible: <u>research + noun</u>, <u>adjective + research</u>, <u>noun + of + research</u>, <u>research + preposition + noun</u>. There are five further patterns, that can also be shown. And <u>research</u> can also be "used as a verb", with a further seven patterns.

For each pattern, ten collocations are displayed, along with their frequencies. Clicking one – *social research*, in the illustration – brings up a superimposed panel displaying similar collocations, with links that retrieve samples in context from the original text. Learners can use the cherry icon to place collocations of interest into their cherry basket (Section 5.5).

The panel beneath shows words that are related to the query term: in this case, *hypothesis*, *scientific*, *empirical*, *method*, *prediction*, *academic*, *researcher*, Here, *scientific* has been clicked, bringing up a panel showing collocations associated with that word: *scientific method*, *scientific information*,

At the bottom, FLAX gives a definition of the query term (from Wikipedia), and related topics (also derived from Wikipedia). Mousing over a topic brings up its definition.

Typing more than one word retrieves collocations containing all the query terms, regardless of word order. This is a good way to expand a learner's knowledge of correct usage. For example, the query benefit public yields public benefit, benefit to the public, benefit the public, reserved for the public benefit, and so on.



Searching for alcohol in Learning Collocations



Searching for advertising in Learning Collocations



Synonyms of ban in Learning Collocations

7.3 Preparing for essay-writing

Knowing collocations of a term that is central to what students are writing about can help them express ideas fluently and accurately. Teachers often recommend collecting a few topic-specific collocations and expressions before starting writing. The Learning Collocations described in the previous section can help.

As an example, we use an essay entitled "Alcohol advertising: Should it be banned?" The first step is to identify keywords for the topic: *alcohol*, *advertising*, and *ban* are obvious candidates. Next the learner seeks collocations that are germane to the topic. This can stimulate a brainstorming process, because students might encounter new and inspiring ideas. Thus it is a good idea to collect several collocations, even though many might not end up in the text. Here are four sample collocations for each keyword:

alcohol	advertising	ban
alcohol consumption	effects/power/impact of	ban on tobacco advertising
amount of alcohol	advertising	advertising ban
excess alcohol	heavy advertising	legislation to ban
addicted to alcohol	funded by advertising	supported the ban

It is not always possible to find exactly matched collocations. However, words can sometimes be substituted or added to relate a collocation to the topic. In the Table, ban on tobacco advertising can be changed to ban on alcohol advertising, heavy advertising to heavy alcohol advertising, and legislation to ban to legislation to ban alcohol advertising.

Sample sentences can be a fruitful source of useful expressions. In the upper image, correlations between alcohol consumption and incidences is a good example. From this expression, plausible sentences like "There is a positive correlation between advertising familiarity and alcohol consumption" can be constructed. Likewise, some collocations in the above Table can be further extended, such as a fair amount of alcohol, driving with excess alcohol, and become addicted to alcohol.

Collecting and using semantically similar collocations is an effective way to avoid repetition and increase lexical variety, like this:

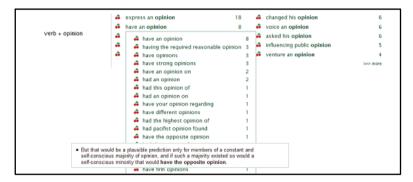
This following section focuses on the effects of alcohol advertising on increasing drinking behaviors. The power of alcohol portrayal on The impact of liquor advertisements on

Here, the repeated use of synonyms *effects*, *power*, and *impact* also helps to increase the cohesion of the writing.

Another way of reducing repetition is to explore synonyms and their collocations. For example, the synonyms *forbid and prohibit* for the word *ban* are shown in the bottom image.



Search for *related* in Learning Collocations



Search for opinion in Learning Collocations



Search for *profit* in Learning Collocations

7.4 Choosing the appropriate word

Do the phrases "learn knowledge" and "big rain" sound odd to you? Many students have difficulty in finding the right words to express their ideas, because they lack collocation knowledge or are unduly influenced by their mother tongue. (Both these phrases are natural in Chinese). As a result, they tend to formulate inappropriate word combinations, or overuse general modifiers such as *more*, *very*, *bad*, *good*, etc. This is particularly noticeable in <u>verb + noun</u>, <u>adverb + verb</u>, and <u>adjective + noun</u> combinations, as in the following sentences, where infelicitous phrases appear in bold:

Alcohol advertising is actively related to alcohol consumption, and the consumption can lead to fatalities.

Some people argue that the alcohol product advertising should be banned and others **keep the opinion** against it.

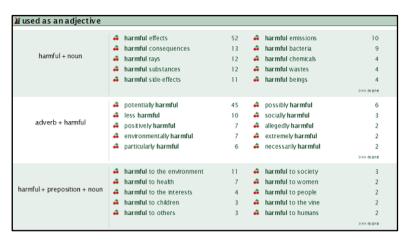
While many alcohol companies are enjoying **lucrative profits**, their alcohol advertising activities are being challenged by the general public and researchers.

In the first example, the student has used the adverb *actively* in an attempt to emphasize the strong correlation between *alcohol advertising* and *consumption*. In the second, *keep* is not an appropriate verb to associate with the noun *opinion*. The last example, *lucrative profits*, is a bizarre combination: *lucrative* is commonly used with *business*, *market*, *career*, etc., but not with *profit*.

The Learning Collocations resource is an excellent source of plausible word combinations. It is fairly easy to locate appropriate verbs or adjectives for a particular noun, or appropriate adverbs for a particular verb. Here are some collocates of *related*, *opinion*, and *profit*, retrieved from Learning Collocations:

related	opinion	profit
closely related highly related clearly related	express an opinion have an opinion voice an opinion	substantial profit increased profit considerable profit

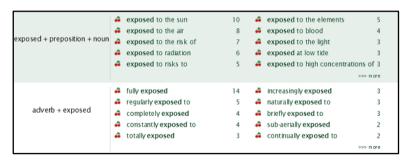
In the first example sentence above, *closely*, *highly*, and *clearly* are all far more appropriate than *actively*. In the second, *express*, *have* and *voice* all seem to fit the context. In fact, this sentence can be further improved by including *have an opposite opinion*; a student can find this by examining the extensions of *have an opinion* shown in the middle image. In the third example sentence, *lucrative can be* replaced with *substantial*, *increased*, or *considerable* to express the intended idea.



Search for *harmful* in Learning Collocations

	eventually lead to	39	4	probably lead to	14
	inevitably lead to	31	4	simply lead to	13
adverb + lead	♣ well lead to	30	4	necessarily lead to	9
	ultimately lead to	24	4	usually lead to	7
	asily lead to	21	4	undoubtedly lead to	6
					>>> more

Search for lead in Learning Collocations



Search for exposed in Learning Collocations

7.5 Hedging and boosting

Adding adverbs to qualify statements is a common rhetorical device, particularly in academic writing. But students often have trouble hedging or boosting statements appropriately and precisely. As a result, they overuse general adverbs such as *very*, *more*, and *much* to weaken or strengthen their claim, and sometimes invalidate statements by choosing highly specific qualifiers. Consider these:

Alcohol is very harmful to their physical and psychological health.

It is a common sense that the more ads we are exposed to, the more likely we are to be seduced to drink and may drink excessively, which **inevitably leads to** disasters while driving.

Smart (1988) however had reviewed many other research and admitted that the link between the advertising and consumption was weak and awaiting more comprehensive research, while at the same time confirmed that alcohol drinkers were definitely exposed to alcohol advertising and their consuming behaviors were in fact continuing to increase.

The *very* in the first example is probably the most common adverb used by novice writers to add extra strength to a statement. Students commonly rely on such adverbs to help voice opinions because of their restricted vocabulary knowledge. These adverbs are weak and ambiguous, and should be avoided in academic writing. In the second and third examples, the adverbs *inevitably* and *definitely* are used to express a high degree of certainty. However, these qualifiers are too extreme. Excessive drinking does not *necessarily* lead to driving disasters, and not all alcohol drinkers are influenced by liquor advertisements.

The Learning Collocations resource can help writers find appropriate hedges and boosters. Here are some examples that are commonly associated with *harmful*, *lead to* and *exposed to*, expressing various degrees of certainty:

harmful	lead to	exposed to
potentially harmful	probably lead to	regularly exposed to
possibly harmful	easily lead to	constantly exposed to
apparently harmful	usually lead to	increasingly exposed to
particularly harmful	ultimately lead to	naturally exposed to
extremely harmful	inevitably lead to	continually exposed to

	a significantly reduce	34	4	thereby reduce	16
	considerably reduce	21	4	gradually reduce	11
adverb + reduce	greatly reduce	20	4	actually reduce	10
	further reduce	18	4	well reduce	9
	substantially reduce	17	4	drastically reduce	9
					>>> m ore

Search for reduce in Learning Collocations

	inguistic behaviour	37	disruptive behaviour	31
	anti-social behaviour	37	strange behaviour	30
adjective + behaviour	a individual behaviour	34	acceptable behaviour	29
	a undesirable behaviour	34	chaotic behaviour	27
	a unacceptable behaviour	34	devlant behaviour	26
	<<< less			>>> m ore

Search for behaviour in Learning Collocations

solve the problem	123	4	have the problem of	21
a tackle the problem of	37	4	face the problem of	20
overcome this problem	35	4	causing the problem	16
a resolve the problem	33	4	see the problem	14
address the problem of	27	4	cure the problem	14
	 tackle the problem of overcome this problem resolve the problem 	 tackle the problem of overcome this problem resolve the problem 33 	a tackle the problem of 37 a overcome this problem 35 a resolve the problem 33 a	cackle the problem of 37 cackle the problem of 37 causing the problem 35 causing the problem 33 causing the problem 33 causing the problem

Search for problem in Learning Collocations

adjective + effect	immediate effect adverse effect overall effect good effect profound effect	136	nnficant effect 11: mulative effect 11- ect effect 110 sired effect 100 matic effect 100 >>> nor	4 0 5
--------------------	--	-----	--	-------------

Search for effect in Learning Collocations

	A heavy drinking	32	4	underage drinking	7
	excessive drinking	20	4	under-age drinking	6
adjective + drinking	a moderate drinking	9	4	late-night drinking	5
	clean drinking water	8	4	responsible drinking attitudes	4
	sensible drinking	7	4	serious drinking	4
					>>> more

Search for drinking in Learning Collocations



Search for drinker in Learning Collocations

7.6 Improving formality

Formality and precision are both important features of academic writing. However, students often overuse colloquial language, and their writing comes out as too informal and lacking in precision. Here are three different ways this can occur.

- 1. Using generic quantifiers
 - a. Due to this, the consumption of alcohol product has reduced a lot.
 - b. If alcohol advertising was banned then this **sort of** behavior would decrease.
- 2. Overusing general words
 - a. Drinking alcohol will hurt health and make public health problems.
 - b. The majority of binge drinkers do not think they are problem drinkers so they could have **bad** effect on their classmates.
- 3. Failing to employ topic-specific collocations
 - a. Drinking too much alcohol can change our behaviors.
 - b. Banning alcohol advertising makes **people who love alcohol very much** decrease.

Students can consult the Learning Collocations resource to find precise expressions that help them avoid colloquial usage. The suggestions in the Table below relate to the example sentences above.

For sentence 1a, the Table suggests replacing a lot by a more expressive word: significantly, considerably, or greatly. Likewise, sort of in 1b could be replaced by undesirable, unacceptable, or deviant. For 2a, the verbs cause, raise and pose are commonly associated with the noun problem. In 2b, substituting serious, damaging or disastrous for bad adds strength. The cumbersome expressions in 3a and 3b can be replaced by topic-related collocations, heavy (or excessive, or serious) drinking instead of drinking too much alcohol, and heavy (or regular, or habitual) drinker for people who love alcohol very much.

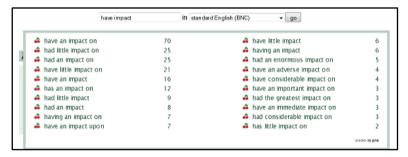
1a. reduce a lot	2a. make public health problems	3a. drinking too much alcohol
significantly reduce considerably reduce greatly reduce	cause the problem raise the problem pose the problem	heavy drinking excessive drinking serious drinking
1b. sort of behavior	2b. have bad effect on	3b. people who love alcohol very much
undesirable behavior unacceptable behavior deviant behavior	have serious effect on have damaging effect on have disastrous effect on	heavy drinker regular drinker habitual drinker



Search for benefit in Learning Collocations



Synonyms of link in Learning Collocations



Search for have impact in Learning Collocations

108

7.7 Increasing text variation

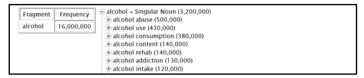
A common problem in student writing is repetition, repetition, repetition. Unless deliberately used for dramatic effect, repetitive writing is boring writing. Here we illustrate how Learning Collocations can be used to enliven the examples below, taken from a student essay.

- Ackoff and Emshoff (1975) confirmed that the increase of advertising activity on the alcohol brand was positively linked with the sales, hence the increasing consumption of the product. Smart (1988) however admitted that the link between the advertising and consumption was weak and awaiting more comprehensive research. Saffer (1997) focused on alcohol consumption and motor vehicle fatalities and revealed positive link between the two.
- Some people will argue that some alcohol products also have some benefits such
 as the use for medicine. However, everything has both sides, it is up to how people
 use. Even though some alcohol products have some benefits, the drawbacks of
 alcohol products overweight the benefits. Therefore, the alcohol product advertising
 should be banned.
- 3. In the long run, it has more advantages to ban alcoholic product advertising on the whole in terms of the healthier and sustainable development of the country, although it may have big impact on the sales of alcohol companies as frequently argued as their evidence by the opponents. For example, ... It is unwise to invest even one dollar on alcohol advertisements, which have bad impact on people's health.

First, deploy synonyms to avoid overusing the same word. For example 1, the Synonyms button in Learning Collocations shows that associate and relate are synonyms of the word link. Further checking the collocations of these two words, and their noun forms (association and relation), yields useful phrases: associate with or association between and relate to or relation between. These are plausible alternatives for link with and link between.

Second, consider using other members of the same word family (e.g., verb, noun, adjective and adverb). The word *benefit* is frequently overused in student writing, particularly its noun form — as in the phrase *have benefits*. Searching for *benefit* in Learning Collocations generates the family word *beneficial*, and also verb usages such as *benefit consumers*, *benefit greatly from*, *able to benefit from*, and *benefit from the use of*.

Third, <u>have + adjective + impact on</u> occurs several times in the example essay in conjunction with weak adjectives like <u>big</u>, <u>bad</u>, <u>small</u>, and <u>great</u>. Searching for phrases by putting multiple words in the query box — in this case <u>have impact</u> — provides an effective way of finding alternatives, such as <u>enormous</u>, <u>considerable</u>, <u>significant</u>, <u>little</u>, <u>adverse</u>, and <u>minimal</u>. Other verbs associated with <u>impact on include assess</u>, <u>examine</u>, <u>consider</u>, <u>minimize</u>, <u>reduce</u>, and <u>measure</u>.



Search for alcohol in Web Phrases



Search for advantage in Web Phrases



Search for begin to in Web Phrases



Search for has * influence in Web Phrases



Search for exposure in Web Phrases

7.8 Checking grammar

Should I use a determiner? And which one: *a*, *an*, or *the*? Is *famous for* or *famous with* correct? Does the word *research* have a plural form? The Web Phrases collection can be used to solve grammatical problems such as these. We will examine the five examples in the Table below: singular *vs.* plural form, appropriate determiners, verbs followed by the word *to*, adjectives *vs.* adverbs, and appropriate prepositions.

1. Singular vs. plural	Eemund (2012) has found the association between alcohol intakes and asthma.
2. Determiners	Alcohol companies take an advantage of the opportunity to promote their products on TV.
3. Verb + to	This situation may make young people begin to drinking alcohol.
4. Adjective vs. adverb	Alcohol products has negatively influence on the social harmony.
5. Prepositions	A study has found that exposure on alcohol advertising encourages people to drink.

- Consulting Web Phrases for phrases that follow the word alcohol shows that the overwhelming majority of nouns are singular: alcohol abuse, alcohol use, alcohol consumption, alcohol rehab and alcohol intake.
- 2. Language learners find it hard to use determiners correctly, because there are many exceptions to the general rules. For example, we say have an advantage, but not take an advantage. In a search for phrases preceding advantage, take advantage (without a determiner) tops the list.
- 3. The word to can be used either as a preposition or as an infinitive marker (used with a verb's base form to indicate that it is in the infinitive). Determining its true function can be difficult, particularly when it follows a verb for example, in contribute to and expose to it is a preposition. Checking Web Phrases for phrases following verb + to helps distinguish them. Here, begin to should be followed by a base form verb: begin to drink.
- 4. Students often confuse a word's adjectival and adverbial forms. Using "wild-card" search in Web Phrases, has * influence generates many <u>adjective + influence</u> phrases, including negative influence.
- Web Phrases makes it easy to find the most commonly used preposition for a particular word. Here, the most common preposition following exposure is to: exposure to.