

Seeing the glass half full: Using a learner corpus for development of writing skills

Alexandra García alexandra.garcia@sydney.edu.au



Outline

- Why the glass half-full?
- What is a corpus? (and what you can get from it)
- How can you use a corpus in a writing class?
- Online resources for independent learning



What is your first reaction to this text?

Many corporations benefit a lot from the telecommuting, from employees' perspectives, firstly, in this way, the time spend on transportation could be saved, that means the staffs would get more time to cope with the projects and documents which are related to work, the productivity could be boosted to a great extent. At the same time, more time staying at home help employees keep work-life balance and strengthen the band between families. For the employer, at first, they cut down enormous expense on the workplace renting, electric charge, water charge, etc., and more budget could be turn to the welfare of the staffs', it is a reasonable way to attract more talent staffs.





Q.1

STUDENT LUYJID29N JOIN ROOM



This used to be mine...

My writing sucks...

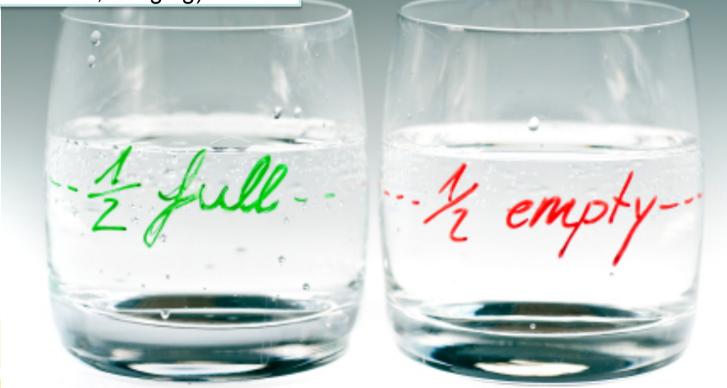
Many corporations benefit a lot from the telecommuting,. Ffrom the employees' perspectives, firstly, in this way, the time spentd on transportation could be saved., That means the staffs would get more time to cope with the projects and documents which are related to work, Thus, the productivity could be boosted to a great extent. At the same time, more time staying at home helps employees keep work-life balance and strengthen the boand between families. For the employer, at first, they cut down enormous expense on the workplace renting, electric charge, water charge, etc., and more budget could be turned to the welfare of the staffs'., Iit is a reasonable way to attract more talented staffs.

Sorry...



Paragraph Structure
Cohesive devices
Supporting details
Elaboration of ideas
Instances of academic lexis
Evidence of complex structures
(passive voice, hedging)

Punctuation
Use of articles
Verb forms
Subject + verb agreement
Instances of non-academic lexis
Plural forms





Compare it to this model

Telecommuting offers the possibility of significant economic benefits. By reducing the demand for mass transit, telecommuting minimises the cost of expanding expensive transportation infrastructure. Fewer cars on the road may also lead to a reduction in pollution and traffic accidents. This will help to alleviate the negative economic consequences associated with urbanisation. Furthermore, both employers and teleworkers themselves can benefit from valuable cost savings too. The reduced need for office space, higher staff retention rates as well as lower workrelated travel expenses all contribute to a strong financial argument in favour of telecommuting.



But can we measure the difference?

% of academic lexis

- in this student sample?
- in the previous model?



Q. 2-3



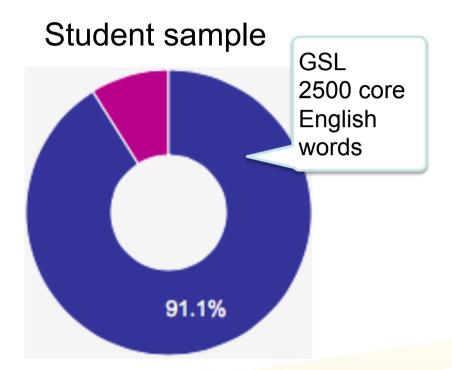


Three ways about it

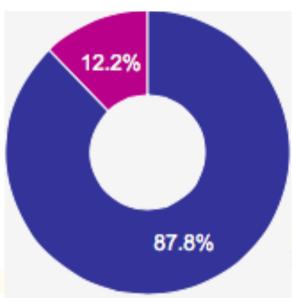
- New General Service List (Brezina & Gablasova 2013)
 2500 most frequent English words
- Academic Word List (Coxhead 2000)
 most frequent words in academic corpus
 not in the GSL 2000
- Readability Test Tool
 years of education required to understand a text
 word length (syllables, letters)
 sentence length



The New GSL



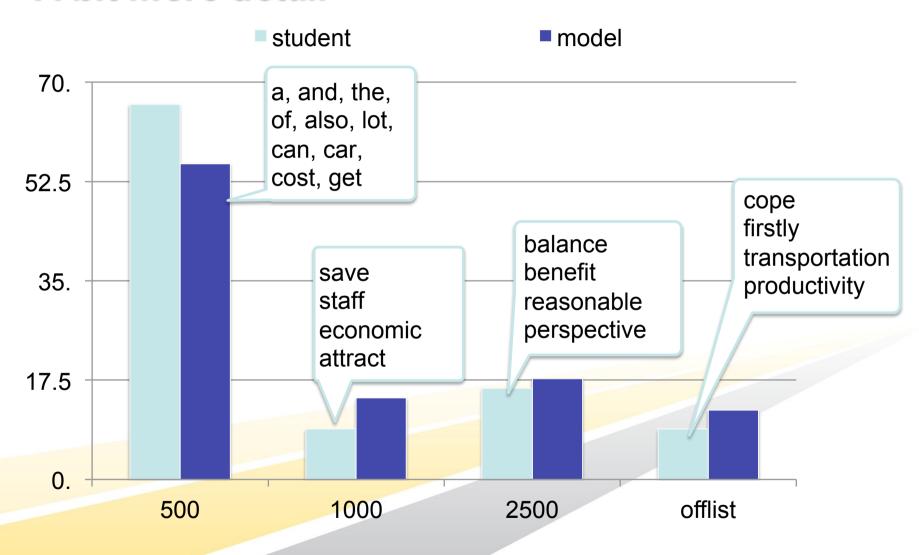




http://corpora.lancs.ac.uk/vocab



A bit more detail





Academic Word List

#Word Types: 18 #Word Tokens: 25

8.6%

			1	4	economic
#Word	Types: 9		2	3	benefits
#Word	Tokens:	10	3	2	benefit
			4	2	positive
1	2	trend	5	1	approach
2	1	access	6	1	contribute
3	1	affect	7	1	financial
4	1	approach benefit	8	1	flexible
6	1	enormous	9	1	furthermore
7	1	furthermore	10	1	infrastructure
8	1	undertake	11	1	major
9	1	welfare	12	1	potential
-			13	1	professional
	6.1%		14	1	promote
	3,2,0		15	1	pursue
			16	1	rigid
			17	1	transit
			18	1	trend
UNS	SW Global			-	

The Readability Test Tool The Model

This page has an average grade level of about 15.

It should be easily understood by 20 to 21 year olds.

The Student's

This page has an average grade level of about 14.

It should be easily understood by 19 to 20 year olds.



Text Statistics

	Model	Student's
Nº of contenece	17	11
N° of words	321	275
N° of complex words	92	51
Percent of complex words	28.66%	18.55%
Average words per sentence	10.00	25
Average syllables per word	1.89	1.57



But can I measure progress?

Yes, with a Learner Corpus

What's a corpus again?

'A helluva lot of text stored on a computer' -(Leech 1992)



The UEEC4 Corpus

- Writing Practice from UEECxx to UEECx (~300 words each)
- Entry 5.5 IELTS
- collected via email
- stored in separate week folders



What else?

'A concordancer makes the invisible visible'

(Tribble 1990)

AntConc

A freeware corpus analysis toolkit for concordancing and text analysis.

[AntConc Homepage] [Screenshots] [Help]

Downloads:

- Windows (3.4.4)
- Macintosh OS X 10.7-10.10 (3.4.3)
- Macintosh OS X 10.6 (3.4.1)
- Linux (3.4.3)
- Older versions

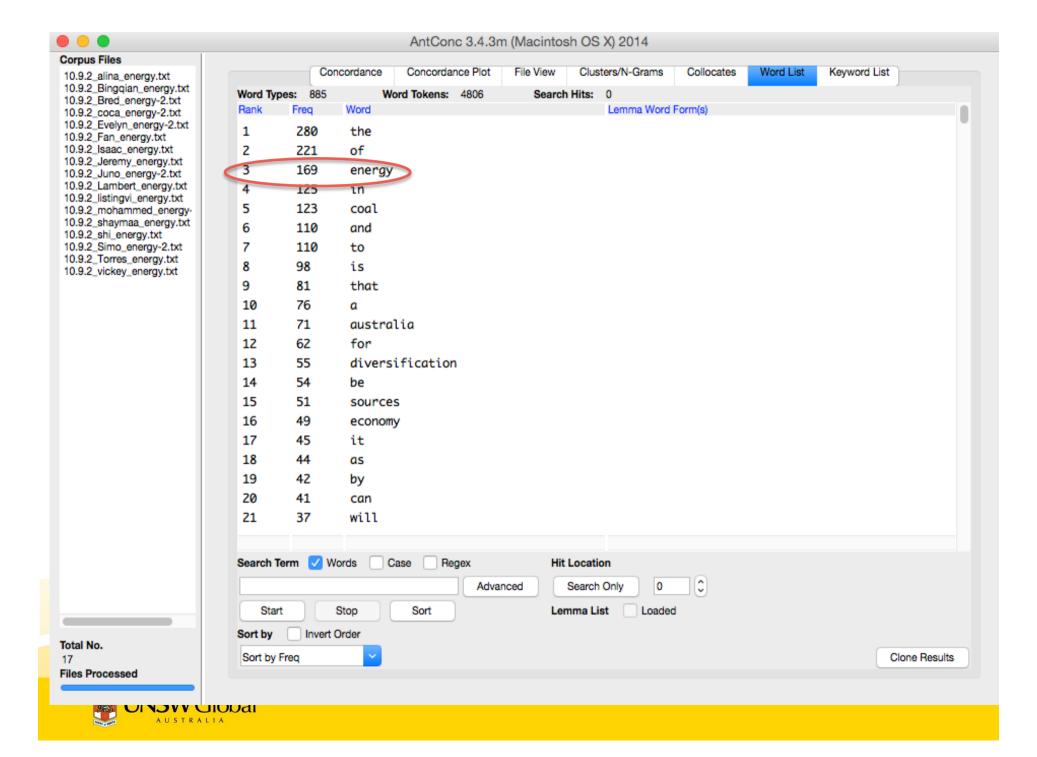
http://www.laurenceanthony.net/software.html

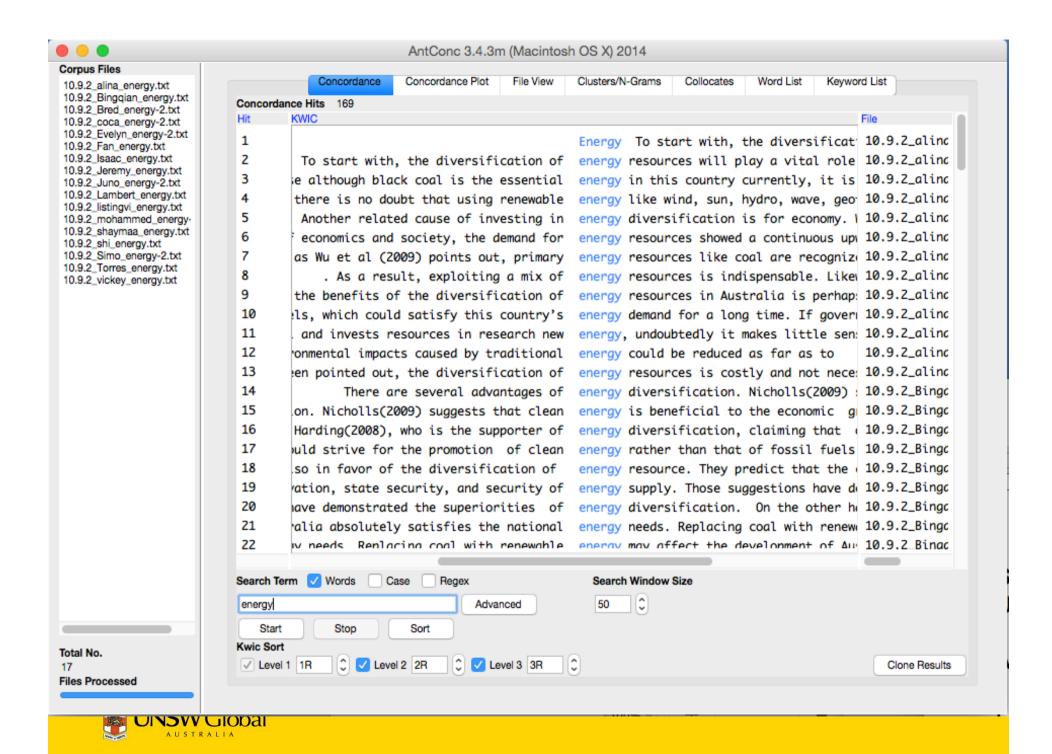


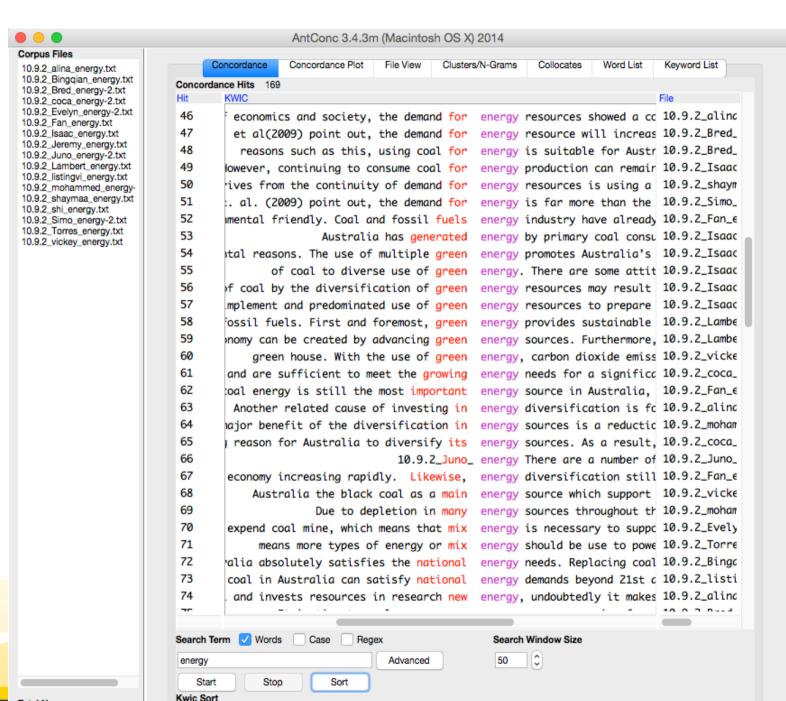
What kind of information do you think you can obtain from a concordancer?









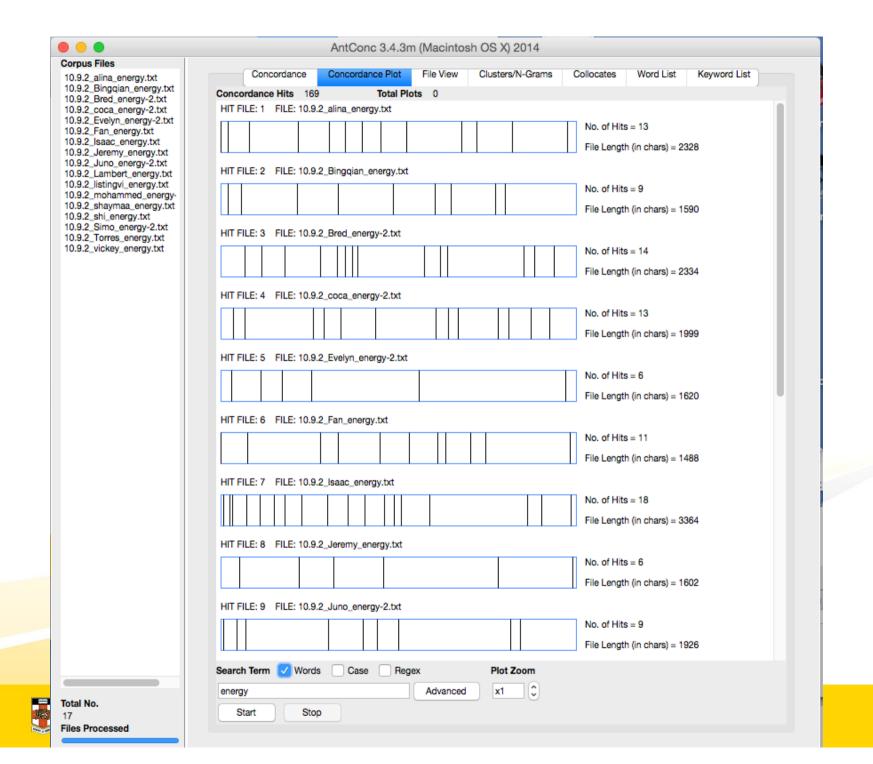


Clone Results

Total No.

Files Processed

✓ Level 1 1L





AntConc 3.4.3m (Macintosh OS X) 2014

10.9.2_alina_energy.txt 10.9.2_Bing qian_energy.tx 10.9.2_Bred_energy-2.txt 10.9.2_coca_energy-2.txt 10.9.2_Evelyn_energy-2.t: 10.9.2_Fan_energy.txt 10.9.2_Isaac_energy.txt 10.9.2_Jeremy_energy.txt 10.9.2_Juno_energy-2.txt 10.9.2_Juno_energy-2.txt 10.9.2_Lambert_energy.tx 10.9.2_Lambert_energy.tx 10.9.2_mohammed_energ 10.9.2_shaymaa_energy.tx

10.9.2_shi_energy.txt

10.9.2_Torres_energy.txt

10.9.2_vickey_clean%20er

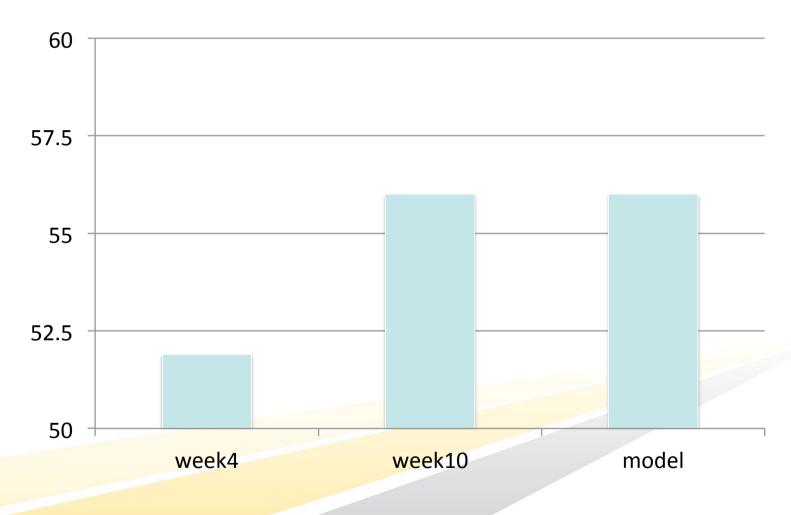
File View Clusters/N-Gram Word Lis Keyword Lis Concordanc Concordance Plo Collocate Total No. of Cluster Types 29 Total No. of Cluster Tokens 155 Cluster Rank Frea Range 12 1 37 energy sources 2 19 6 energy resources 3 12 energy sources in 12 eneray sources in australia 4 8 5 energy is 6 7 energy diversification 7 7 6 energy source 8 7 energy sources is 9 4 energy resources is 3 3 10 energy can 3 11 2 energy needs 2 12 energy demand energy diversification is 13 2 2 14 energy economy 15 2 2 energy for Search Term V Words Regex N-Grams Cluster Size Case Advanced Min. 2 energy Start Stop Sort Min. Frea. Min. Range 2 Sort by Invert Order Search Term Position 2 On Left On Right Sort by Freq Clone Results

Total No. 15

Files Processed



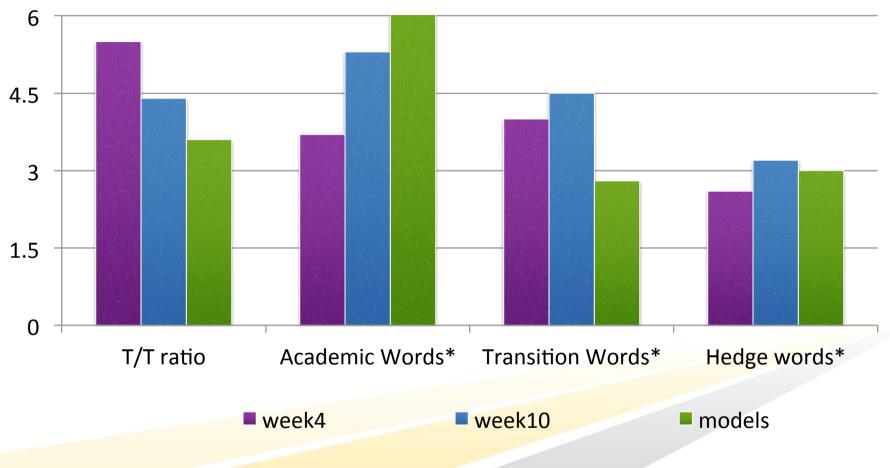
Words that appear only once



% Hapax Legomena



Features of Academic Writing



*per 100 words



What other lists would you create to measure students' academic language skills?

- cause & effect language
- reporting verbs
- non-academic language
- Abstract nouns



Q. 4





Not a bag of words...

Passive voice?

- Wild cards: be|is|are|was|were|been
- 88 concordance lines
- 88/17 (#texts) = 5.17 instances per text

So what?



How can I use these data to help my students?

1	environmentalists considerate that if satellites	are abolished, fraction of satellites is the biggest
2	. At the same time, some poverty countries	are argued that it is not necessary to
3	astronauts may be died. These space programs	are thereby seem to be a waste of
4	9, about 19,000 pieces of debris larger than 5 cm	are tracked, while an estimated 300,000 pieces lar
5	the innovations which created from space industry	are widely use in people's daily life.
6	of technology and medicine, those problems may	be addressed in the future so that the
7	will discuss whether the space exploration should	be advocated or not. It will examine the
8	is one of the first organs can	be affected because of the reduction in exercise.
9	some people claim that more money should	be allocated in this field of science in
10	time, opponents allege that harmful effects may	be caused by space technologies. Firstly, economis
11	d stability. Personally, space exploration should	be continue for long term consideration. Torres.
	·	





How can I use these data to help my students?

	√ or x
vironmentalists considerate that if satellites are abolished, fraction of satellites is the bigge	
the same time, some poverty countries are argued that it is not necessary to	
tronauts may be died. These space programs are thereby seem to be a waste of	
ut 19,000 pieces of debris larger than 5 cm are tracked, while an estimated 300,000 pieces	
iovations which created from space industry are widely use in people's daily life.	
hnology and medicine, those problems may be addressed in the future so that the	
iscuss whether the space exploration should be advocated or not. It will examine the	
is one of the first organs can be affected because of the reduction in exercise.	
some people claim that more money should be allocated in this field of science in	
opponents allege that harmful effects may be caused by space technologies. Firstly, econom	
tability. Personally, space exploration should be continue for long term consideration. Torres.	



Looking at text development - Theme

There can be some positive influences bought by the space exploration.

To begin with, the innovations and inventions generated during the projects have changed human's life.

For instance, the material invented and used on the 'moon boot' has been widely applied at present to improve shock absorption and provide superior stability and motion control.



Compare the following sets of concordances

From your writing

cial violent. However, this essay will argue that the violent video games do not contribute t are insufficient evidences to demonstrate that the violent video games could cause the v For instance, the number of people who played the violent video games in United States ha introduced such as radio, movies and television, the violent video games have been made. or societies try to control the student players, the violent video games still the most popul

From the Web

end to sales of violent video games to minors", Freedom Forum,
Retail employees who sell violent video games to minors would face
to its armpits in violent video games, violent books and magazines,
More ». Study says that violent video games can improve vision. Mar.
that goes on about violent video games, it's nice to hear,



	What is the pattern for the use of 'not only' at the beginning of a sentence?			
1:	Not only	is it important to hit all the right nutritional notes,		
2:	Not only	that, but even when I try to open AcroRd32.exe in		
3:	Not only	did it include a prohibition on employing firkas [the fourth		
4:	Not only	do the Americans tolerate this bloodshed, but at least one		
5 :	Not only	would he win the presidency by a comfortable margin, but		
6:	Not only	should we keep passengers from sitting next to a terrorist,		
7:	Not only	were the objects of his strenuous art pleasing to look		
8:	Not only	do we get to hear the spin-doctor coaching candidates received		
9:	Not only	did the Bush Administration out a covert intelligence officer working		
10:	Not only	does it provide plenty of different, age-appropriate fun for babies		



A bottom-up approach

- Identify examples and non-examples
- Compare and contrast with output from authentic corpora
- Elicit grammatical patterns from examples (e.g. Passive voice= to be + verb-ed)
- Identify colligation (word + grammatical form e.g. due to + NG)
- Guess the missing word (search term)
- Identify shades of meaning (e.g. claim)
- Disambiguate synonyms (citizens vs humans)

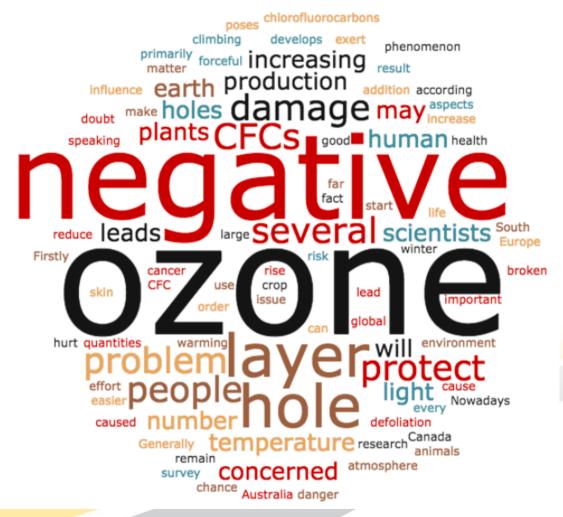


What can my students do independently to improve their writing?





An image worth a thousand words...



http://www.wordclouds.com



Find a synonym, but which?

Synonyms for negative

adj bad, contradictory

adverse

gloomy

pessimistic

unfavorable

weak

abrogating

annulling

anti

con

contrary

contravening

denying

disallowing

disavowing

dissenting

gainsaying

impugning

invalidating

iaundiced

naysaying

neutralizing

nullifying

opposing

recusant

refusing

rejecting

removed

resisting

against

antagonistic

balky

colorless

counteractive

cynical

detrimental

Common

dissentient

nugatory

privative

repugnant

resistive

unaffirmative

unenthusiastic

uninterested

unwilling





"adverse effects"

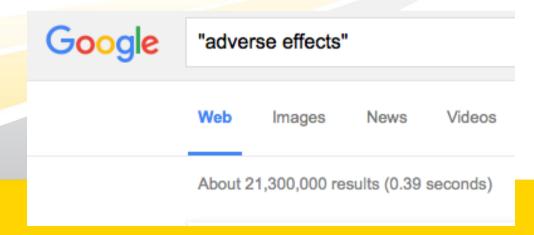
Google Search

I'm Feeling Lucky



Vocabulary: Other words for 'negative' +

	impact	effect	consequences
detrimental	> 530K	<3.3 M	225K
deleterious	180K	<3 M	215K
adverse	4.4 M	22 M	2 M





But is it academic?

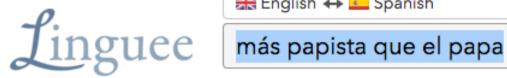




Articles (include patents)
 Case law

Stand on the shoulders of giants





áéíóúüñ

Q

Dictionary Spanish-English

```
más que adv ■) - rather adv (i)
papista adj 

papist adj 

papist adj 

Papa prop.n m ■ O Pope n (i)
papa f = 0 — potato n = 0 . spud n = 0
[...]
```

© Linguee Dictionary, 2016

External sources (not reviewed)

Este Parlamento no debe ser más papista que el Papa. This House really should not try to be more Catholic than

⇒ europarl.europa.eu

the Pope. □→ europarl.europa.eu

[...] europeas de acuerdo en esta liberalización, yo no voy a ser más papista que el papa y, por tanto, mañana votaré a favor. ⇒ europarl.europa.eu

[...] are in agreement on uncoupling, I will not be more catholic than the Pope, and I shall therefore be voting in favour tomorrow. ⇔ europarl.europa.eu

Esto significa que el Parlamento sería más papista que el Papa, porque la propia Comisión Europea, basándose en las estadísticas [...] ⇒ europarl.europa.eu

This means that Parliament would be more royal than the king, because the European Commission itself, on the basis of the [...] ⇒ europarl.europa.eu

Lo que ocurre es que, en este caso, la Comisión ha sido más papista que el Papa y -como ha puesto de relieve el Sr. Fernández Martín- más librecambista que las propias multinacionales. ⇒ europarl.europa.eu

The problem is that in this case the Commission has been more papist than the Pope and, as Mr Fernández Martín has pointed out, more in favour of free trade than the multinationals themselves. ⇒ europarl.europa.eu

Are we being more Catholic than the Pope?

- How much passive voice?
- Can we use 'I'?
- How about 'a lot' or 'bad'?





APAStyle.org

Resources

Exit

✓ 12 of 25 >

Reducing Bias in Language

Acknowledge participation.

When you write about the roles of individuals in an experiment, use language that portrays them as active participants, rather than passive recipients of the experiment.

For example, choose phrases such as the following:

- The students completed the survey instead of The survey was administered to the students.
- The children viewed objects on the screen instead of The children were shown several on-screen objects.
- We collected data from the participants instead of The participants were run.

Write about the people in your study in a way that acknowledges their participation but is also consistent with the traditions of the field in which you are working. Thus, although descriptive terms such as *college students*, *children*, or *respondents* provide precise information about the individuals taking part in a research project, the more general terms *participants* and *subjects* are also in common usage.





"will argue"

Scholar

About 508,000 results (0.07 sec)



Q.5

STUDENT

LUYJID29N

JOIN ROOM





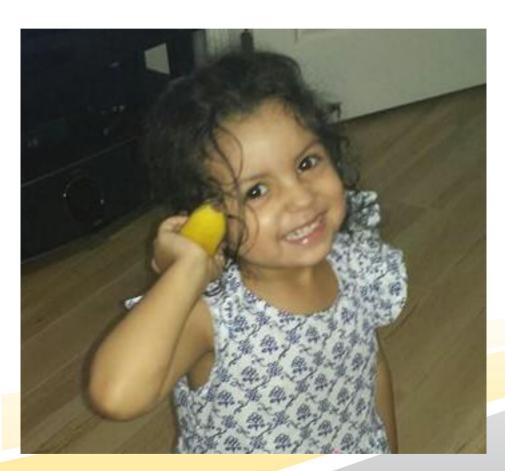
Stylish Academic Writing

HELEN SWORD

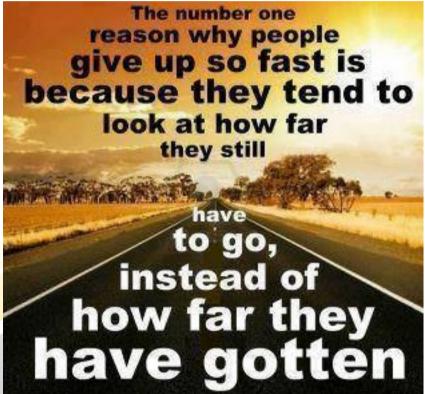




Coda



Language Development





Thank you!



Seeing the glass half full: Using a learner corpus for development of writing skills

Alexandra García

I. What is your first reaction to this text?

Many corporations benefit a lot from the telecommuting, from employees' perspectives, firstly, in this way, the time spend on transportation could be saved, that means the staffs would get more time to cope with the projects and documents which are related to work, the productivity could be boosted to a great extent. At the same time, more time staying at home help employees keep work-life balance and strengthen the band between families. For the employer, at first, they cut down enormous expense on the workplace renting, electric charge, water charge, etc., and more budget could be turn to the welfare of the staffs', it is a reasonable way to attract more talent staffs.



- How academic is the lexis in this text? Can you offer a percentage?
- How about the following model?

Telecommuting offers the possibility of significant economic benefits. By reducing the demand for mass transit, telecommuting minimises the cost of expanding expensive transportation infrastructure. Fewer cars on the road may also lead to a reduction in pollution and traffic accidents. This will help to alleviate the negative economic consequences associated with urbanisation. Furthermore, both employers and teleworkers themselves can benefit from valuable cost savings too. The reduced need for office space, higher staff retention rates as well as lower work-related travel expenses all contribute to a strong financial argument in favour of telecommuting. (from UEECx, Assessment Details)



What kind of information do you think you can obtain from a concordancer?



How do you think you can use concordance lines in a writing class?

1	environmentalists considerate that if satellites	are abolished, fraction of satellites is the biggest
2	. At the same time, some poverty countries	are argued that it is not necessary to
3	astronauts may be died. These space programs	are thereby seem to be a waste of
4	9, about 19,000 pieces of debris larger than 5 cm	are tracked, while an estimated 300,000 pieces lar
5	the innovations which created from space industry	are widely use in people's daily life.
6	of technology and medicine, those problems may	be addressed in the future so that the
7	will discuss whether the space exploration should	be advocated or not. It will examine the
8	is one of the first organs can	be affected because of the reduction in exercise.
9	some people claim that more money should	be allocated in this field of science in
10	time, opponents allege that harmful effects may	be caused by space technologies. Firstly, economis
11	d stability. Personally, space exploration should	be continue for long term consideration. Torres.



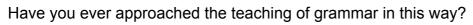
What is the purpose of the following activity?

Compare the following sets of concordances From your writing

violent. However, this essay will argue that the **violent video games** do not contribu insufficient evidences to demonstrate that the **violent video games** could cause the instance, the number of people who played the **violent video games** in United States introduced such as radio, movies and television, the **violent video games** have been made. or societies try to control the student players, the **violent video games** still the most popul

From the Web

end to sales of violent video games to minors", Freedom Forum,
Retail employees who sell violent video games to minors would face
to its armpits in violent video games, violent books and magazines,
More ». Study says that violent video games can improve vision. Mar.
that goes on about violent video games, it's nice to hear,



	What is the pattern for the use of 'not only' at the beginning of a sentence?		
1:	Not only	is it important to hit all the right nutritional notes,	
2:	Not only	that, but even when I try to open AcroRd32.exe in	
3:	Not only	did it include a prohibition on employing firkas [the fourth	
4:	Not only	do the Americans tolerate this bloodshed, but at least one	
5 :	Not only	would he win the presidency by a comfortable margin, but	
6 :	Not only	should we keep passengers from sitting next to a terrorist,	
7:	Not only	were the objects of his strenuous art pleasing to look	
8:	Not only	do we get to hear the spin-doctor coaching candidates received	
9:	Not only	did the Bush Administration out a covert intelligence officer working	
10:	Not only	does it provide plenty of different, age-appropriate fun for babies	

What else you can do:

- Identify examples and non-examples
- Compare and contrast with output from authentic corpora
- Elicit grammatical patterns from examples
- (e.g. Passive voice= to be + verb-ed)
- Identify colligation (word + grammatical form e.g. due to + NG)
- Guess the missing word (search term)
- Identify shades of meaning (e.g. claim)
- Disambiguate synonyms (adverse vs. injurious)

Free Online resources

	What you can do with it	Where you can get it
AntConc	Obtain frequency lists, concordance lines, word clusters, key words and collocates	http://www.laurenceanthony.net/software.html
General	Analyse the lexis of a text to determine	http://corpora.lancs.ac.uk/vocab/
Service List	what percentage of it belongs to the core 2500 English vocabulary list	

Readability	Identify how many years of education are	http://www.webpagefx.com/tools/read-able/
Test Tool	required to understand a text. Calculate	
	text statistics such as # of sentences.	
WebCorp	Obtain concordance lines for a word or	http://www.webcorp.org.uk/live/
	expression from the whole Web	
Synchronic	Obtain concordance lines for a word or	http://wse1.webcorp.org.uk/cgi-bin/SYN/index.cgi
English Web	expression from the whole Web – Sort	
Corpus	concordances by adjacent words	
Google scholar	Find out whether an expression is	https://scholar.google.com.au/
	regularly used in academic texts	
Linguee	Find in-context translations for short	http://www.linguee.com/
	expressions (Chinese available but not	
	Arabic)	
BNCweb	Concordance lines from the British	http://bncweb.lancs.ac.uk/bncwebSignup/user/login
	National Corpus (can be sorted)	.php
MICUSP	Michigan Corpus of Upper Level student	http://micusp.elicorpora.info/
	Papers. Retrieve paragraphs containing	
	search term. Search can be filtered by	
	text type, student level and/or discipline	
Academic	research based list of most common	http://www.victoria.ac.nz/lals/resources/academicw
Word List	academic words (not in the GSL)	ordlist/awl-headwords
Wordcloud	obtain a visual representation of word	http://www.wordclouds.com
	frequency from a text	Intp.//www.wordclouds.com
Stylish	By Helen Sword. Corpus Describes	Ebook available from most university libraries
Academic	effective and not so effective academic	
Writing	writing based on corpus research	