1st Annual Conference of Teachers of English as a Foreign Language (ACTEFL 2017)

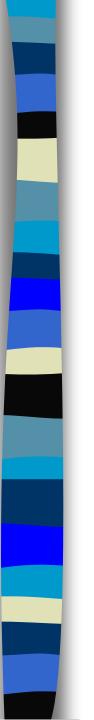
October 27, 2017 São Paulo, SP

Brasil

CORPUS LINGUISTICS AND TEACHING: hands-on work with corpora



Helmara Febeliana Real de Moraes, PhD U.S. Embassy – Brasília Regional English Language Specialist <u>moraeshf@state.gov</u> October 2017





- > Corpus Linguistics: what is it?
- > What is a *corpus*?
- Concordance lines
- ≻The corpora online
- What shall we explore? Discovering patterns
- Patterns in the classroom activities

CORPUS LINGUISTICS: WHAT IS IT?

"Corpus linguistics of the 80's and 90's can be defined as a branch of general linguistics that involves the analysis of large machinereadable *corpora* of running text, using a variety of software tools designed specifically for textual analysis." (Laviosa, 2002:6)



IT INVOLVES...

the analysis of language in use/empirical data - (new) findings and inferences

> observation vs. intuition/introspection

questioning: langue/parole (Saussure); competence/ performance (Chomsky)

Language is seen as a social phenomenon - uses which can be empirically studied

But is it a theory? A methodology? An approach?

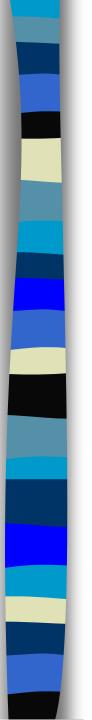


WHAT IS A CORPUS?

"A corpus is 'a collection of texts assumed to be representative of a given language, dialect or other subset of a language, to be used for linguistic analysis." (Francis, 1992, in Sinclair, 1992:2 apud Laviosa 2002:6)

- the researcher/linguist can work with samples of texts
- compilation criteria representativeness
- authentic material always obtained from sources of language in use
- computer-stored information it can be manipulated through specially developed search tools

A TEXT	A CORPUS
read whole	read fragmented
read horizontally	read vertically
read for content	read for formal patterning
read as a unique event	read for repeated events
read as an individual act of will	read as a sample of social practice
instance of parole	gives insights into langue
coherent communicative event	not a coherent communicative event



CONCORDANCE LINES

"A *concordance* is a collection of the occurrences of a word-form, each in its own textual environment." (Sinclair, 1991)

Search word: centralized and followed by its original co-text \rightarrow the words it occurred with in the *corpus*.

Example:

concerned about the way the western nations **lavishly** use the earth's diminishing resources. The theatre. [p] [p] Though many of the camps are **lavishly** equipped (even with electricity in the quide to Art Deco style. Paperback, 80 pages, lavishly illustrated in colour and black and white. the Roman invasion to the Battle of Britain. Lavishly illustrated with photographs, paintings and Elton John, Bernie Taupin [/h] [p] In this **lavishly** illustrated book the poetry of the lyrics of of classic pop, rock and soul covers. This lavishly produced and seductive selection includes by Lydia Greeves and Michael Trinick, and **lavishly** illustrated with colour photographs, it with a ferocity that is truly shocking in a **lavishly** equipped fourdoor sedan, particularly so for nearly three months," he says, pulling **lavishly** at the bottle. They had already lost one selected for detailed reproduction in this **lavishly** illustrated volume. Text is meanwhile kept of religious drag. [p] Piere et Gilles is a **lavishly** illustrated catalogue of some of their best



Aspects to observe:

- the horizontal <u>and</u> vertical axis;
- the most visible repetition of words that collocate with the node (to the right and to the left);
- the repetition of patterns in the light of grammatical categories;
- repetition at the semantic and functional levels (semantic field and semantic prosody);

Summarizing: we should read vertically, read fragmentarily and focus on repeated events.

(Tognini-Bonelli, 2001)

THE CORPORA ONLINE

> BNC (British National Corpus)

A closed *corpus* of 100 million words of British English, both spoken and written. It was concluded in 1995 and generates 50 examples with the search word.

Access to the whole *corpus*: through purchase.

Online access: http://www.natcorp.ox.ac.uk/ http://corpus.byu.edu/bnc

COCA – The Corpus of Contemporary American English

- 520 million words (1990 2015)
- includes 20 million words each year
- equally divided among spoken, fiction, popular magazines, newspapers, and academic texts
- search different genres at different periods of time
- semantically-based research (little/small, republicans/democrats)
- free online access at

http://corpus.byu.edu/coca

Corpus of Contemporary American English (📄 🛃 🛃



	SEARC	H F	REQUENCY		CONTEXT		CONTEXT		CONTEXT		CONTEXT	
SEE CONTEXT: (CLICK ON W	ORD OR SELECT WORDS + [CONTEXT] [HELP]				TAL	COMPARE				
		CONTEXT	FREQ		ALL	96	MI					
1		MAN	766		345167	0.22	5.40					
2		GRASS	496		21403	2.32	8.79					
3		ORDER	352		111272	0.32	5.92					
4		BUILDINGS	319		26397	1.21	7.85					
5		TREES	279		46450	0.60	6.84					
6		ENOUGH	242		198460	0.12	4.54					
7		TALES	215		7575	2.84	9.08					
8		WOMAN	193		184603	0.10	4.32					
9		GLASS	191		52495	0.36	6.12					
10		WINDOWS	163		37372	0.44	6.38					
11		GUY	144		96632	0.15	4.83					
12		BLACK	128		201438	0.06	3.60					
13		TALE	126		12074	1.04	7.64					
14		PINES	117		3310	3.53	9.40					
15		BUILDING	112		96420	0.12	4.47					
16		BOY	94		83793	0.11	4.42					
17		GRASSES	86		2623	3.28	9.29					
18		GLASSES	80		15173	0.53	6.65					
	-											

Corpus of Contemporary American English (📄 🛃 🛃



SEARCH FREQUE		UENCY	Γ	CONTEXT				HELP			
SEE CONTEXT: (E CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP]							HIG	Η		COMPARE
		CONTEXT		FREQ		ALL	96	MI			
1		SCHOOL		61398		338174	18.16	8.64			
2		SCHOOLS		4332		94107	4.60	6.66			
3		LEVELS		3681		63292	5.82	6.99			
4		LEVEL		3161		113089	2.80	5.94			
5		BLOOD		2093		74673	2.80	5.94			
6		RISK		1822		77077	2.36	5.70	•		
7		HEAT		1769		60584	2.92	6.00	•		
8		QUALITY		1696		60350	2.81	5.95	•		
9		ENOUGH		1617		198460	0.81	4.16	1		
10		COURT		1610		123185	1.31	4.84	1		
11		ABOVE		1607		92240	1.74	5.26	1		
12		DEGREE		1532		42569	3.60	6.30	1		
13		HEELS		1268		9443	13.43	8.20	1		
14		POINT		1207		204593	0.59	3.69	1		
15		COST		1147		80102	1.43	4.97	I.		
16		RATES		1062		50856	2.09	5.52	1		
17		SPEED		1055		37675	2.80	5.94	1		
18		EXPECTATIONS		060		21555	4.50	6.62	1		



> WEBCORP

- accesses the World Wide Web as a corpus
- tool constantly revised
- provides more specific and detailed linguistic search
- results are presented in concordance lines (link to the web page)

Online access: <u>http://www.webcorp.org.uk</u>



WebCorp LSE

A manager thing pair too be

1633

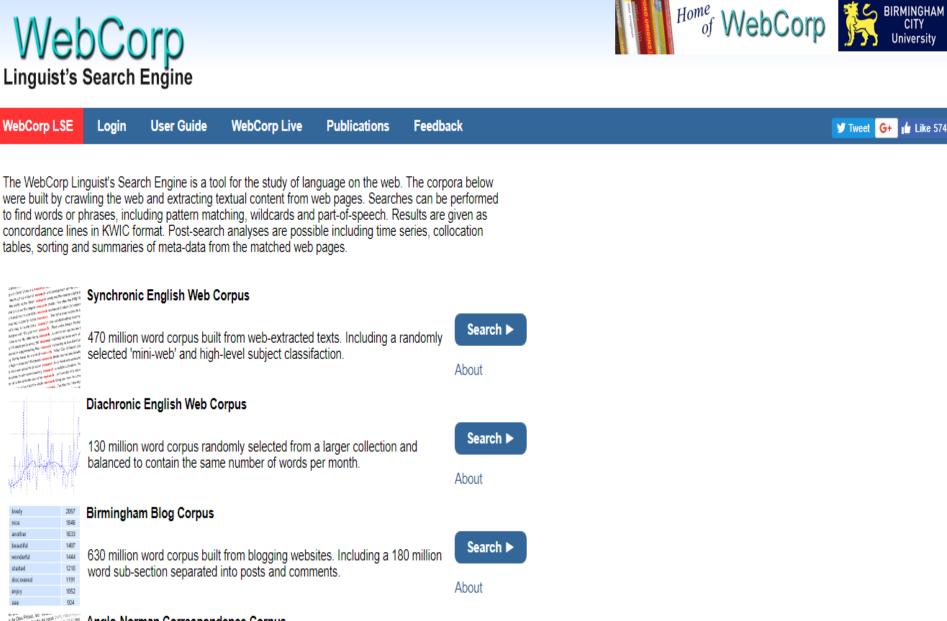
1210

1057

beautiful

andphi

discovered



У Tweet 🛛 🖌 👍 Like 574

(i) wse1.webcorp.org.uk/cgi-bin/SYN/search.cgi?q=tall_{N*},any,0,0,0-0-0 $\leftrightarrow \rightarrow C$

	4474 i	nstances of 'tall {N *}	r' (0 min 18 sec).		New Query Clone Help Profi
1: 2: 3: 4: 5: 6: 7:	McCain, or sidewalks would be unsafe as libs leapt from pitch their tents. Saddam had surrounded the area with a performing all their new tricks! Theres a competition jumping ring, jZRY 1 CaP Go Diego Go Video HwNn Gr8s [www.ahluwalia-law.com], in the centre. DIDYMOS and RASTOS, in. frock-coat costumes and right hand toward a chair on wheels, standing before a wheeled by Jeremiah Flintwinch of the cut-down aspect to her	tall buildings tall brick tall tunnel tall women tall hats tall writing tall cabinet	and psychologists lined up to explain that media professionals don't wall. There were only three points of entry. All the slide, a three-tiered winners stage with trophies, ribbons and treats porn a0Pc OTT Free Huge Collection Adult Movie Clips 2 , come down and. do a lively dance. KI-RAM. Colonel, what cabinet close shut up I am able to attend to . When she had unlocked and opened it, and had settled	ĺ	View Concordances 50 per page Show POS tags: Jump to 1 Go Span: 10 Unlimited results/doc
8: 9: 10:	with her hands laid separately upon the desk, and the No, not I. Madame, you are always right,' returned the the landlady Biraud, is it not Rigaud, madame,' returned the	tall cabinet tall Swiss tall Swiss	towering before her, she looked as if she were performing Doubtless you were enraged against that man, madame Ay, yes, Rigaud! To be sure.' The traveller's soup was succeeded by		Display Info: Number V
 11: 12: 13: 14: 15: 16: 17: 	and what doesn't. Nothing too complex, just having fun. A mine girl over action. across where he, behind his mousy 101. Sketches by Boz unlooked-for visitor. It was a singularly end of his journey. Mrs Wilfer was, of course, a mamma's cherished hopes that I should become united to a one historian of His day described Him as being a Tanya? Quite a big girl, I expect Yes, she's very	tall bow tall vinegar tall woman tall woman tall member tall man tall," Darya	stand will be set up for an O.P. then rovers , had venerated us. an intensification. We have no hostage that , dressed in deep mourning, and standing so close to the and an angular. 35 Her lord being cherubic, she was of society. It may have been a weakness, but if with chestnut-colored hair, parted in the middle, with short hair Alexandrovna answered shortly, surprised herself that she should respond so	1	Table Of Results Case insensitive: POS: Ignore ▼ Min frequency: 5 Top 1000 results. Summarise
18: 19: 20: 21: 22: 23: 24: 25: 26:	several voices began to talk all at once, and the a servant to take to somebody. In the evening a case containing the icon glittered, illuminated on all sides by past, and then all was still. The shadow of a behind the big stove was not long enough for a but got mixed, being used to senseless soldiers' talk. A only the payment is according to our strength," said the and there, all alone, she found the medical assistant, a what the general had said. The prison inspector was a t	tall nobleman tall man tall candles tall poplar tall person tall man tall man tall man	with the ring, getting more and more exasperated, shouted more , with long, grey hair and a white beard, entered the ornamented with golden spirals. The candelabra was filled with tapers, fell on the ground just opposite the window, and all to stretch out on. "And on this very bed," Nekhludoff with a small beard, a long nose, and a bass with the bass voice, evidently foreseeing how the matter would , with a blotchy face, who had for a long time g man, with moustaches and whiskers that twisted towards the corners		http://wse1.webcorp.org.uk/cgi-l
20: 27: 28: 29:	linen shirt and trousers on; it was that of a a probability function. Let de ne a fuzzy set of masked." "But, Olivia's last name isn't Laveau," Xander objected. The	tall man tall man tall men tall man	with a small beard and half his head shaved. The . Then say we are given a precise value of the led Olivia onto the dance floor as the music began		

tall woman

30: trash, that one Yes!" A remarkable change came over the led Olivia onto the dance floor as the music began . "He perverts the old ways in his lust for power!"

gi-bin

rofile

.cgi?q=tall_{N*},any,0,0,0-0-0

4474 instances of 'tall {N*}' (0 min 18 sec).

Query Summary

Result	Frequency	Number of Documents
tall women	209	74
tall order	175	158
tall man	173	126
tall buildings	173	137
tall grass	132	108
tall terror	102	102
tall tales	101	85
tall trees	84	78
tall building	70	58
tall figure	64	41
tall tale	57	46
tall woman	54	41
tall glass	52	46
tall ship	43	26
tall models	39	39
tall tree	39	37
tall grasses	35	29
tall fescue	33	13
tall people	33	22
tall dark	29	27
tall Australian	28	22

☆
New Query Clone Help Pro
View Concordances
50 🔻 per page
Show POS tags: 🔲
Jump to 1 Go
Span: 10 🔹
Unlimited <pre>Image: Value of the second se</pre>
Display Info: Number 🔻
Table •
Table Of Results
Case insensitive: 🔲
POS: Ignore V
Min frequency: 5
Top 1000 results.
Summarise

http://wse1.webcorp.org.uk/cgi-t



We should have in mind:

- 1- What lexical patterns is the word part of?
- **2-** Does the word regularly associate with particular other meanings?
- **3-** What structure(s) does it appear in?
- **4-** Is there any correlation between the word's uses/ meanings and the structures in which it participates?

5- Is the word associated with (any position in any) textual organization?

(Hoey, 1997)

BUT WHAT EXACTLY SHALL WE EXPLORE? DISCOVERING PATTERNS

Collocation

A recurrent lexical combination which is non-idiomatic and cohesive; its elements are contextually restricted and co-occur arbitrarily. (Tagnin, 1998)

"The way words combine in a language to produce natural-sounding speech and writing." (Oxford Collocations Dictionary, 2002)

Classification	Base	Collocate	Example
Nominal	noun	noun	credit card
collocation			labor pains
Adjectival	noun	adjective	upright piano
collocation			outside chance
Verbal	noun	verb	pay a visit
collocation			muster courage
Adverbial	verb	adverb	pay dearly
collocation			lie outright
Adverbial	adjective	adverb	awfully sorry
collocation			deeply hurt

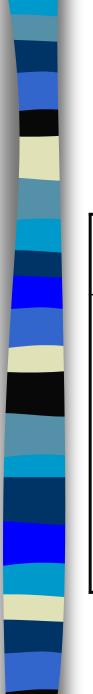


Colligation

The relation between a pair of grammatical categories \rightarrow a pairing of lexis and grammar. Example: *cases* + quantifiers - *in some cases, in many cases*: cases <some, many, most, more, both, several> (Stubbs, 2001)

"The interrelation of grammatical categories in syntactical structure."

(Firth apud Tognini-Bonelli, 2001:5)



The verb 'undergo'

passive or modal	+undergo	+adjective (typical adj.)	+abstract noun (typical lexical fields)
forced to		further	medical procedure
required to		extensive	testing
must		major	training
etc.		severe	change
		etc.	a trauma
			etc.





 a discontinuous sequence of two words, positioned at one word remove from each other

- highly selective of their collocates. Examples:
 - a + ? + of
 - an + ? + of
 - be + ? + to
 - too + ? + to
 - for + ? + of
 - had + ? + of
 - many + ? + of

AN + ? + OF:

- Measurement and quantifiers (army, average, inch, ounce)
- Indicating part of something (edge, end, evening,hour)
- Specification of an attribute (array, index)
- Support for the noun following of (act, example, expression, object)
- Indicating an activity (extension, explanation, invasion)
- Indicating a quality or circumstance (absence, awareness)
- Indicating a relationship (enemy, officer)

(Hunston & Francis, 2000)

> SEMANTIC PROSODY

"(...) many uses of words and phrases show a tendency to occur in a certain semantic environment. For example, the verb *happen* is associated with unpleasant things - accidents and the like."

(Sinclair apud Hoey, 1997)

CAUSE AND PROVIDE: EXAMPLES

> CAUSE \rightarrow frequently with words for unpleasant events: problems, death, damage, concern, trouble, cancer, disease, etc.

> **PROVIDE** \rightarrow with words denoting things which are desirable or necessary: information, service(s), support, help, money, protection, food, care, etc.

(Stubbs, 2001)



BUT WHY ARE PATTERNS IMPORTANT?

> sharing patterns and meaning

Verb+at+noun:

- speaking loudly and unpleasantly:
 "I just laughed at him." "I shouted at her."
- communicating with a facial expression or gesture: "We just smiled at each other."
 "You used to wave at me."

 \rightarrow Arranging words in groups which share a meaning helps making sense of and remembering the pattern.

> patterns are a source of mistakes

• important to accuracy and cause a lot of mistakes when mis-used. Example:

The noun *objective*:

* Teachers have the objective to help learners acquire natural English.

Pattern to be used:

'noun + of + -ing' - the objective of helping (aim, function, purpose and role)

knowing a word with its patterns encourages fluency

• learners can produce not just one word, but a series of words, a phrase, together.

Patterns in the classroom

• Teachers should make sts aware that words have patterns which are meaningful \rightarrow we should not try to teach all of them, but show them and emphasize their importance.

(Hunston, 2001)

Going back to the corpora online...

Some uses for concs (teaching):

band

- different meanings of the same word
- ≻ well
 - different functions of one same word

> shelter

- different grammatical roles of the same word

depend

- the co-occurrence of grammatical patterns

Now considering the keywords listed in the next slides:

- > can you guess what the material is about?
- Do they belong to the same area/text typology?

Ν	Key word	Freq.	%	RC. Freq.	RC. %	Keyness	Ρ
1	#	219513	12,78843689	1604421	1,61304605	525899,875	0.0*
2	TBSP	11259	0,655929267	569		87326,48438	
3	MINS	9261	0,539529383	575		71195,13281	
4	OZ	9633	0,561201453	1283		70734,53125	
5	MINUTES	13556	0,78974843	17662	0,017756946	68495,21094	
6	ADD	11541	0,672358096	8162		67718,97656	
7	CHOPPED	8690	0,506263912	1037		64328,43359	
8	PAN	8586	0,50020504	1862		60316,16406	
9	BUTTER	8547	0,497933	2086		59270,89844	
10	OIL	10511	0,612352133	10158	0,010212608	57460,625	
11	PEPPER	7202	0,419575691	951		52908,8125	
12	INGREDIENTS	7235	0,421498209	1242		51997,95313	
13	HEAT	8621	0,502244115	5794		51103,83984	
14	TSP	6491	0,378154099	394		49941,17188	
15	SALT	7606	0,443112016	2932		49685,11719	
16	UNTIL	13004	0,757589877	40128	0,04034372	48350,22266	
17	COOK	7230	0,421206921	3543		45452,01563	
18	COOKING	6844	0,398719251	2592		44821,87891	

N	Key Word	Freq.	%	RC. Freq.	RC. %	Keyness	Р
1	SHALL	8787	1,354663849	19817	0,019923532	53643,37109	0.0*
2	OR	19341	2,981740475	370166	0,372155935	46328,03906	
3	AGREEMENT	6116	0,942884266	13242	0,013313186	37717,21875	
4	ANY	11025	1,69968915	122180	0,122836813	36809,73438	
5	SUCH	6016	0,927467585	107326	0,107902959	15037,39648	
6	FRANCHISEE	1448	0,223233551	46		14186,47754	
7	COMPANY	3024	0,466200471	35947	0,036140244	9689,033203	
8	HEREUNDER	775	0,119479284	45		7463,502441	
9	LENDER	824	0,127033457	380		6809,098145	
10	FRANCHISOR	696	0,10730011	24		6805,135742	
11	HEREOF	700	0,107916772	35		6774,625977	
12	PURCHASER	1006	0,155091822	1676		6613,243652	
13	SELLER	999	0,15401265	1640		6589,92041	
14	APPLICABLE	946	0,145841807	1413		6376,626465	
15	BORROWER	742	0,114391781	280		6282,437988	
16	DATE	1758	0,27102527	17192	0,017284421	6238,036133	
17	NOTICE	1422	0,219225213	9028		6138,090332	
18	THEREOF	757	0,116704285	466		6010,709961	

Other uses for keyword lists:

to explore some vocabulary before reading a text

> to ask sts what they think the text is about

Phraseologies in ESP - results of a *corpus*-based research

Phraseology (I)	Equivalent form (P)
finely chopped	bem PICADO, (bem) PICADINHO
peeled and finely chopped	DESCASCADO e (bem) PICADO
seeded and finely chopped	sem sementes, PICADO; PICADO (e) sem sementes
onions, finely sliced	cebolas (cortadas) em rodelas (bem) finas; cebolas (cortadas) em fatias (bem) finas; cebolas cortadas fino/fininho
finely grated	ralado
freshly ground black pepper	pimenta-do-reino moída na hora
freshly grated nutmeg	noz-moscada ralada na hora
freshly chopped parsley	salsa/salsinha (bem) picada na hora; salsa/salsinha picadinha na hora
cover and simmer gently	tampe (a panela) e cozinhe em fogo baixo/em fogo brando/lentamente/em fogo lento

simmer gently for a few minutes	cozinhe em fogo baixo/em fogo brando/ lentamente/em fogo lento por alguns minutos
over a pan of gently simmering water	em banho-maria
heat gently until melted	aqueça em fogo baixo até dissolver/derreter; aqueça ligeiramente/levemente até dissolver/ derreter
gently fry	fritar/frite ligeiramente/levemente/rapidamente; dê uma ligeira fritada; refogar/refogue rapidamente/ levemente/ligeiramente
gently bring to the boil	cozinhe em fogo baixo até ferver/levantar fervura; leve ao fogo baixo até ferver/levantar fervura
lightly floured surface	superfície levemente/ligeiramente polvilhada com farinha; superfície levemente/ligeiramente enfarinhada
until lightly browned	até começar a dourar; até que comece a dourar; até dourar levemente/ligeiramente
lightly oiled bowl	tigela untada com um pouco de óleo
thinly sliced	CORTADO em fatias (bem) finas
peeled and thinly sliced	DESCASCADO e CORTADO em fatias (bem) finas
thinly spread	espalhe uma camada fina de + ingrediente

Adverbial collocations in Business - some examples

basic and intermediate levels: work closely (with)

advanced level: adversely affect, fall sharply, highly competitive, mainly due to (?), rapidly grow, rise sharply, work closely (with)

(Santos, 2011)



Work closely with – basic level

Step 1: present six concs (one at a time) – headlines (news)

Each line of business works closely with Risk Management through its own risk committee

Contractors from Spain and Ireland, and Ulster Bank worked closely with GBM in Madrid

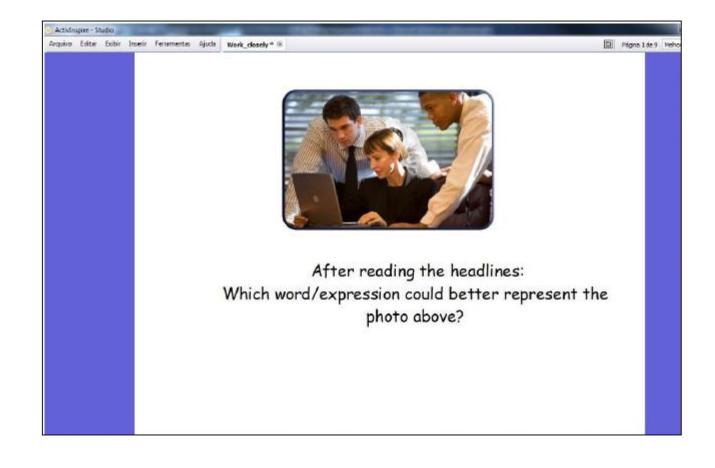
The Centre works closely with HSBC's Global Research sector heads

The Group Chairman works closely with the Company Secretary to ensure that accurate, timely and clear information flows to the Board.

LeMasters said he would work closely with Replay's entertainmentindustry investors to devise schedule formats for Replay

Mr Matsuo worked closely with the Cabinet Secretariat deciding politicians' travel budgets

Step 2: using visual aids, present the picture/question below, so that sts shall come up with the unit:



Step 3: present the concordances together:

EXERCISES

N Concordance

3

- its own risks. Each line of business works closely with Risk Management through
- 2 from Spain and Ireland, and Ulster Bank worked closely with GBM in Madrid, the
- for the bank and its clients. The Centre works closely with HSBC's Global Research
 - to the Board. The Group Chairman works closely with the Company Secretary to
- 5 do in the future." LeMasters said he would work closely with Replay's
- 6 Japan's most senior politicians. Mr Matsuo worked closely with the Cabinet Secretariat
- Look at the lines above. They were taken from the texts you've read. What can you observe about the verb "work"?

Step 4: provide some gap-filling exercise for practice

Complete the sentences below with:

a) a verb:

We have also _____ closely with the U.S. government and with a number of other institutions.

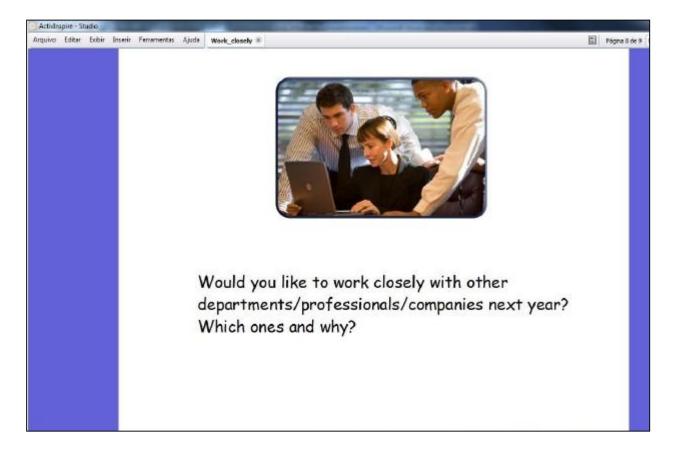
b) an adverb:

We will work ______ with UNITE and other employee representatives to reduce staff number.

c) a preposition:

A Group Corporate Sustainability unit was formed in 2007 to work closely ______ individuals and businesses in all customer groups.

Step 5: sts use the phraseology to talk about their reality:



(Santos, 2011)

Checklist for developing *corpus***-based activities:**

- Have a clear idea of the point that you want to teach;
- Select the *corpus* that is the best resource for your lesson;
- Explore the *corpus* completely for the point you want to teach;
- Make sure that your directions are complete and easy to follow;
- Make sure that your examples focus on the point that you are teaching;
- Provide a variety of ways for interacting with the materials;
- Use a variety of exercise types;
- If you are using computers, *always* have na alternative plan or activity in the event of computer glitches.

(Reppen, 2010)

And now let's analyze some concordance lines...

Thanks!!

Helmara