

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Kollokationen und Korrigieren – Korpora als Hilfsmittel im Englischunterricht

Dr. Claudia Bubel

06.11.2008

- 1) Collocations
- 2) Teaching collocations
- 3) Corpora
- 4) Corpus data in the classroom
- 5) Corpora and assessment

In the morning I made some work in the garden, then I spent a rest for about an hour before going out to have some shopping in town. It was my sister's birthday and I wanted to do a special effort to cook a nice meal for her. I gave a look at a new Thai cookery book in the bookshop and decided to buy it. It has some totally easy recipes and I managed to do a good impression with my first Thai meal. I think my sister utterly enjoyed her birthday.

McCarthy/O'Dell. 2005. *English collocations in use*. Cambridge: Cambridge University Press.

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

To get the right word in the right place is a rare achievement.
(Mark Twain, 1868)

In the morning I (made) **did** some work in the garden, then I (spent) **had** a rest for about an hour before going out to (have) **do** some shopping in town. It was my sister's birthday and I wanted to (do) **make** a special effort to cook a nice meal for her. I (gave) **had/took** a look at a new Thai cookery book in the bookshop and decided to buy it.

It has some (totally) **very/quite/extremely** easy recipes and I managed to (do) **make** a good impression with my first Thai meal. I think my sister (utterly) **thoroughly/really** enjoyed her birthday.

The company words keep

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Collocations

combinations of words which occur naturally with greater than random frequency.

Adjective + noun	→ brilliant idea
Noun + prep + noun	→ a sense of pride
Noun + noun	→ service charge
Verb + noun	→ make a decision
Verb + preposition	→ filled with horror
Verb + adverb	→ whisper softly
Phrasal verbs	→ turn in
Adverb + adjective	→ blissfully unaware
Adverb + verb	→ half understand

The company words keep

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Colligations

type of collocation, where a lexical item is typically linked to a grammatical one.

e.g. *surprising, amazing, astonishing* but
*not surprising, *not amazing, *not astonishing*
→ only *surprising* colligates with the negative

richly, lavishly

lavishly colligates with verbs (*use lavishly*)

richly with adjectives (*richly fruity*)

both with –ed-forms (*lavishly decorated; richly decorated*)

Collocations provide learners with a powerful organisation principle of language.

(Michael Lewis. 2002. *Implementing the lexical approach: Putting theory into practice*. Boston: Heinle, p.26)

Teaching collocations

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Think bigger than the word: Make pupils aware of collocations and mis-collocations.

Don't explain – explore: Instead of explaining the meaning of a word, give co-text.

Make the most of what pupils already know: Use high frequency words and collect collocations that go with them.

(cf. Lewis, Michael (ed.) 2000. *Teaching collocation: Further developments in the Lexical Approach*. Boston: Thomson and Heinle.)

Teaching collocations

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

But:

hard to teach/mark/correct for non-natives because their teaching and assessment requires native speaker intuitions

Teaching collocations

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Collect collocations

- in written and spoken texts,
- in dictionaries (collocation dictionary entries),
- in corpora.

Correcting collocations

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Check collocations

-in dictionaries (collocation dictionary entries),

-in corpora.

Corpora

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

large collections of texts

stored electronically

consist of many million words of text

comprise various text types

from academic papers to casual conversation

words are tagged according to parts of speech (POS)

→ provide attested examples of words in context

→ *virtual native speaker* (Rühlemann 2008)

Corpora available on the web:

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

BROWN (1 million words - from written American English texts published in 1961)

<http://www ldc.upenn.edu/cgi-bin/ldc/textcorpus?doc=yes&corpus=BROWN>

also available through:

http://www.lex tutor.ca/concordancers/concord_e.html

BNC - British National Corpus (100 million words from a wide range of British English sources, both spoken and written, 1980s - 1993)

<http://www.natcorp.ox.ac.uk/>

also available through:

<http://corpus.byu.edu/bnc/>

Corpora available on the web:

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

COCA – Corpus of Contemporary American English (more than 385 million words from five text genres - spoken, fiction, popular magazines, newspapers, academic journals)

[http://www.americancorpus.org /](http://www.americancorpus.org/)

The Collins Cobuild Wordbank Online English Corpus

(56 million words of contemporary written and spoken text)

<http://www.collins.co.uk/corpus/CorpusSearch.aspx>

Corpora available on the web:

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Time Corpus (100 million words from Time magazine articles, 1923-present)

<http://corpus.byu.edu/time/>

English-German Translation Corpus (English texts with German translations by professional translators; different text types range from EU policy documents to speeches by political and public figures)

<http://www.tu-chemnitz.de/phil/InternetGrammar/>

MICASE - Michigan Corpus of Academic Spoken English (1.8 million words of transcribed academic speech events)

<http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>

The web as corpus

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

WebCorp

<http://www.webcorp.org.uk/>

Web Based Concordances

<http://www.niederlandistik.fu-berlin.de/cgi-bin/web-conc.cgi?art=google&sprache=en>

Google?

→ e.g. extended search: site: uk

also: <http://cli.la.asu.edu/togoogleornot.htm>

Corpus data

– frequencies

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Give information on how frequently a word occurs or how frequently words co-occur.

e.g. *beautiful, handsome, ugly, attractive, good-looking, pretty, unattractive*
(cf. Thornbury 2002:68ff.)

most frequent (among the 3000 most frequent words of English):

1) *beautiful*

2) *attractive*

3) *pretty*

least frequent:

negative adjectives *ugly, unattractive*

Corpus data – frequencies

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

BYU-BNC: BRITISH NATIONAL CORPUS (100 MILLION WORDS, 1980s-1993)
Mark Davies / Brigham Young University

DISPLAY: CHART LIST COMPARE WORDS

SECTIONS: NO SORT: RELEVANCE

SEARCH STRING: WORD(S): men women CONTEXT: [aj*] 3 0

SECTION 1: IGNORE SPOKEN FICTION NEWSPAPER ACADEMIC MISC MIN FREQ: 4

SECTION 2: IGNORE SPOKEN FICTION NEWSPAPER ACADEMIC MISC MIN FREQ: 0

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

WORD 1 (W1): MEN (0.97) WORD 2 (W2): WOMEN (1.03)

	WORD	W1	W2	W1/W2	SCORE		WORD
1	ARMED	92	0	184.0	189.6	1	PREGNANT
2	HOLLOW	45	0	90.0	92.8	2	BATTERED
3	MASKED	34	0	68.0	70.1	3	SALVADOREAN
4	EVIL	33	0	66.0	68.0	4	INTERNATIONAL
5	LESSER	23	0	46.0	47.4	5	ETHNIC
6	DRY	16	0	32.0	33.0	6	LESBIAN
7	MERRY	15	0	30.0	30.9	7	ACTUAL
8	MOUNTED	13	0	26.0	26.8	8	PRIMIPAROUS
9	GUILTY	12	0	24.0	24.7	9	RAPED
10	LEARNED	12	0	24.0	24.7	10	SEPARATED

FIVE MINUTE TOUR

You can also search for **collocates (words nearby a given word)**, which often provides insight into the example, you can search for the most common nouns *near thick*, adjectives *near smile* (sorted by relevance), starting with *clos** near eyes.

You can also include **register information** directly as part of the query. This allows you to see how words and many different types of written texts. We can easily find which words and phrases occur much more frequent such as *-ness* words in poetry, adjectives in tabloid newspapers, nouns in advertisements, good + [noun] in fiction.

frequency lists from the British National Corpus <http://corpus.byu.edu/bnc/x.asp>

Corpus data

– KWIC concordances

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

like you would be as ignorant as you are ugly. A gwerbret's a kind of lord, see, the mc
chambers of Westminster, it looked like ugly revolution. The Communists saw themselves
e, that he would do nothing about it--an ugly, shrewd look--so in the end I gave him sc
wful, stupid, incompetent depressing and ugly the nation is [p] Just about every aspect
WEDDING [p] Buena Vista, 15, 1994 [p] AN UGLY duckling becomes a swan in this queasy Au
ond opinion All right," he said, "you're ugly, too .." Terribly embarrassing, I bumped
In this saga of the good the bad and the ugly, there is good news for the ugly. Women s
pianist's descent into madness caused an ugly public row. Harvey Weinstein, co-founder
' weren't rich and famous? Power makes an ugly man attractive [p] Bienvenida Sokolow,
ves out;Diary [/h] [b] P.H.S [/b] [p] AN UGLY rumour is doing the rounds in the kitchen
an explosive derby match boiled over in ugly scenes. [p] The Manchester United manage
ine, would go as follows Belinda is fat, ugly and spotty. She goes into the cake shop a
ul on Pembridge, which again sparked off ugly scenes between the sides. [p] Arsenal we
d not become Labour leader because he is ugly and that Margaret Beckett's vote will suf
: "I used to think of myself as a really ugly duckling, with a big nose and ears [p] Ar
d as basic insurance against becoming an ugly misfit. Actually, beauty has been paired
' stood, on its hind legs, and a massive, ugly head. Leaphorn inspected the declivity ir
s waiting to be put away, of the nicked, ugly furniture. They were trapped up here, tra
ll, no mussed hair or sloppy clothes, no ugly faces not even any average faces. Everyor
abbed boric water on his cuts and on the ugly gash on his forehead. "Shit, you hardly e
also happened to be lazy, stubborn, and ugly. All things considered, he was the only f

concordance for ugly from <http://www.collins.co.uk/corpus/CorpusSearch.aspx>

Corpus data in the classroom

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Example: *what's the point* (Rühlemann 2008)

What's the point of having a phone when you never answer it?

Otherwise, what's the point of being alive?

What's the point of advertising a Porsche if you don't have one?

Corpus data in the classroom

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Example: *if-clauses expression requests* (Rühlemann 2008)

→ pattern of grateful if

*In order to be able to make seating arrangements I would be **grateful if** you **would** please, please and return the slip below no later than ...*

*I'll be very **grateful if** you **could** bring me in a bowl.*

*I would be **grateful if** they **would** contact me.*

*We should be **grateful if** you **would** search the register of the above company.*

*I should be **grateful if** he **would** write to me.*

*I'd be ever so **grateful if** I **could** send you about 10 of these sets to distribute amongst fellow teachers.*

*I would be most **grateful if** you **could** send the print as soon as possible.*

Corpus data in the classroom

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Use familiar texts

e.g. play concordances

<http://victorian.lang.nagoya-u.ac.jp/concordance/>

what is the use of them? They seem, as a class, to have **absolutely** no sense of moral responsibility.

It does not thrill. It produces **absolutely** no vibrations

you led me to believe, Miss Fairfax, that you were not **absolutely** indifferent to me.

About everything I should fancy. You have such an **absolutely** trivial nature.

moments of physical courage of which we women know **absolutely** nothing.

without my consent until she comes of age. That consent I **absolutely** decline to give.

Simplify and pre-select concordance lines

Corpus activities

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Example: Finding synonyms

BYU-BNC: BRITISH NATIONAL CORPUS (100 MILLION WORDS)
Mark Davies / Brigham Young University

DISPLAY ?

CHART LIST COMPARE WORDS

SECTIONS SORT ?

SEARCH STRING ?

WORD(S) ?

CONTEXT ?

POS LIST ?

USER LISTS ?

SECTION ?

1 2

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT
1	<input type="checkbox"/>	BEAUTIFUL
2	<input type="checkbox"/>	LOVELY
3	<input type="checkbox"/>	ATTRACTIVE
4	<input type="checkbox"/>	WONDERFUL
5	<input type="checkbox"/>	STRIKING
6	<input type="checkbox"/>	SUPERB
7	<input type="checkbox"/>	MAGNIFICENT
8	<input type="checkbox"/>	HANDSOME

Corpora and assessment

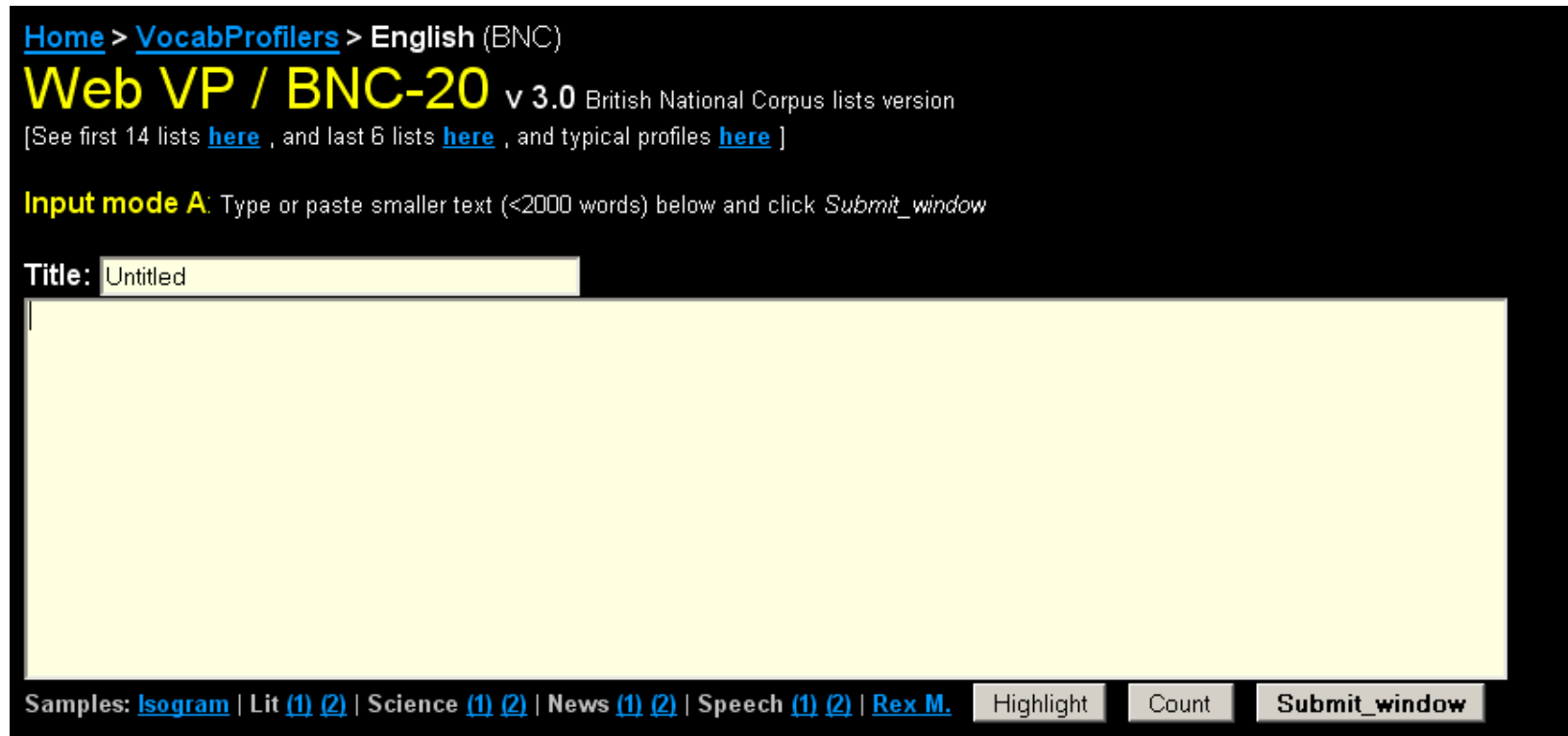
Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Evaluating the range of vocabulary in student texts

<http://www.lextutor.ca/vp/bnc/>

→ type-token ratio



The screenshot shows the interface of the 'Web VP / BNC-20 v 3.0' tool. At the top, there is a breadcrumb trail: 'Home > VocabProfilers > English (BNC)'. Below this, the title 'Web VP / BNC-20 v 3.0' is displayed in large yellow font, followed by 'British National Corpus lists version'. A note in brackets says: '[See first 14 lists [here](#), and last 6 lists [here](#), and typical profiles [here](#)]'. Underneath, the 'Input mode A' instruction reads: 'Type or paste smaller text (<2000 words) below and click *Submit_window*'. There is a 'Title:' label followed by a text input field containing 'Untitled'. Below the title field is a large, empty yellow text area for input. At the bottom of the interface, there is a row of links: 'Samples: [Isogram](#) | Lit [\(1\)](#) [\(2\)](#) | Science [\(1\)](#) [\(2\)](#) | News [\(1\)](#) [\(2\)](#) | Speech [\(1\)](#) [\(2\)](#) | [Rex M.](#)'. To the right of these links are three buttons: 'Highlight', 'Count', and 'Submit_window'.

Corpora and assessment

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

Correctness? Appropriateness? Idiomaticity?

Enter a word combination / form in a BYU corpus or Google / Webcorp.

Look at frequencies!

Examples:

Then he goes to bed and sleep.

I had a problem to send the letter.

There's her car keys.

(s. Rühlemann 2008)

Bibliography

Saarland

Landesinstitut für Pädagogik
und Medien (LPM)

- Kreyer, Rolf. 2007. Das Korpus im Klassenzimmer. *Praxis Fremdsprachenunterricht* 6. 12-16.
- Lee, David. no date. Bookmarks for corpus-based linguistics. [<http://devoted.to/corpora>]
- Lewis, Michael (ed.) 2000. *Teaching collocation: Further developments in the Lexical Approach*. Boston: Thomson and Heinle.
- Morgan, John; and Mario Rinvoluceri. 2004. *Vocabulary*. Resource Books for Teachers. Oxford: Oxford University Press.
- Mukherjee, Joybrato. 2002. *Korpuslinguistik und Englischunterricht: Eine Einführung*. Frankfurt: Peter Lang.
- Perrin, Geoff. 2007. Korpusbasiertes Lernen. *Praxis Fremdsprachenunterricht* 6. 17-21.
- Rühlemann, Christoph. 2008. The British National Corpus – Wie können Lehrerinnen und Lehrer das BNC nutzen. *Praxis Fremdsprachenunterricht* 2. 43-47.
- Rühlemann, Christoph. 2008. The British National Corpus – Wie können Schülerinnen und Schüler das BNC nutzen. *Praxis Fremdsprachenunterricht* 3. 32-36.
- Thornbury, Scott. 1999. *How to teach grammar*, Sample lesson: Teaching verbs that take both infinitive and –ing form, using concordance data, 65-68. Harlow: Pearson – Longman.
- Thornbury, Scott. 2002. *How to teach vocabulary*. Harlow: Pearson – Longman.