

BNCweb Intro

Jean Mark Gawron

Linguistics 503
San Diego State University

March 21, 2016

Outline

- 1 Intro
- 2 Changing context window
- 3 Sorting
- 4 Distributions
- 5 Downloads
- 6 Collocations
- 7 Dispersion

BNCweb

BNCweb is a pretty intuitive web-based interface to BNC you can use with your web browser.

BNCweb in CompLing Lab

<http://bulba.sdsu.edu/bncweb>

Requirements

Password and account name

Simple queries

Queries

introvertedness [no matches]

introverted

Results

Your query "introverted" returned 84 hits in 69 different texts.

- 1 A06 1335 But besides this more obvious point, there are subtler connections between voice and body: Cicely Berry observes that an **introverted** and thoughtful person often finds more difficulty in speaking and does not carry the thought through into the physical process of making speech.
- 2 A18 429 Razumikhin himself may or may not have come from the country, but he is certainly a member of the floating, unbelonging population of students and ex-students, and he records in simple puzzlement that Raskolnikov has been growing increasingly moody and suspicious and **introverted**; he has no time for anything, people are always in his way, and yet he lies about and does nothing a confirming echo of Raskolnikov on his bed telling Nastasya the maid that he is working, by which he means thinking.

Multiword queries and tokenization I

Multi word queries ok. Special characters must be escaped (preceded with “\”). Comma (“,”), question mark (“?”), plus (“+”), colon (“:”), at-sign (“@”), various kinds of braces, all special. Quotation mark NOT special

talk of the town	Good query
very very pretty	one hit
very, very pretty	syntax error
very , very pretty	still a syntax error
very \, very pretty	correct query, 5 results
“ pretty “	7 hits, all <i>pretty</i> in scare quotes

Lemmas, intervening stuff I

Lemmas

- {kick/V} curly braces signal lemma, {kick/N} and {kick/V} are different lemmas
- {kick} All instances of all lemmas with the form *kick*

Intervening stuff

- day >> 2 >> night *day* followed by *night* within a 2 word window (excludes *by day and by night*)
- night << 2 << day *day* followed by *night* within a 2 word window, but *night* will be the highlighted word
- day << 2 >> night *day* and *night* within 2 words in either order

Lemmas, intervening stuff II

Plus and Star

- `this (-{A})* {day}` *this* followed by any form of *day* with any number of adjectives intervening
- `this (-{A})+ {day}` *this* followed by any form of *day* with at least one adjective intervening
- `this + {day}` *this* followed by any form of *day* with exactly one word intervening
- `this ++ {day}` *this* followed by any form of *day* with exactly two words intervening
- `this + _NN1` *this* followed by exactly one word followed by a singular noun (note no curly braces to use BNC C5 tag).

Lemma/partof speech/sentences boundary

Query

```
{kick/V} <s> bucket_NN1
```

“All tokens of the LEMMA *kick* and the singular noun *bucket* occurring within a single sentence (in either order)”

Results

Your query "{kick/V} <s> bucket_NN1" returned 24 hits in 17 different texts

No	Filename	
1	A6W1120	At any speed, in any gear on the mile straight there is enough power to bury you hard into the thin bucket seats; every quick-fire gearchange though the massively solid Borg-Warner box kicks at the back end.
3	AC4 2431	Jinny was so startled that she nearly kicked the bucket over.
6	ATE 787	“Did you think I'd kicked the bucket, Ma?”

Combining ideas

{kick/V} the (-{A})* bucket_NN1

More BNCweb info

Hoffmann et al. (2008)

Reading

Getting started

ISBN

[Book flyer](#)

[Book extract](#)

[Quick tutorial](#)

978-3-631-56315-1

Looking at more context

Clicking on the highlighted target word takes you to a new screen in which more context is shown.

Your query "evidence" returned 21137 hits in 2323 different texts (98,313,429 words [4,048 texts]; frequency: 215 instances per million words) (1.135 seconds)

No	Filename	Hits 1 to 50	Page 1 / 423
1	A03.72	The authorities claimed they were conspiring to overthrow the government, which is dominated by a different community, the Moors, but offered no evidence to substantiate this claim.	
2	A03.221	On the highly publicized issue in AI's December report of the baby deaths, Amnesty said that although its team was shown alleged mass graves of babies, it was not established how they had died and the team found no reliable evidence that Iraqi forces had caused the deaths of babies by removing them or ordering their removal from incubators.	
3	A03.404	Although both were linked to the PKI, there is no evidence that either of them had knowledge of the events leading to the coup.	
4	A03.442	The previous May, 13 of his co-defendants were acquitted and released unconditionally on the grounds that there was insufficient evidence for prosecution.	
5	A03.579	The prosecution's evidence consisted of statements allegedly extracted under torture, and confiscated literature.	

67 The five were previously imprisoned from June until October 1990 for allegedly organizing a political party — all parties are prohibited.

68 The releases occurred shortly before two AI representatives arrived in Swaziland for talks with the government.

69 Mauritania

70 In April AI called on the Mauritanian Government to investigate reports that up to 200 political prisoners had died or been killed in military or police custody — some executed without trial and many others as a result of torture. 71 The victims were among some 3,000 black Mauritians arrested late last year. 72 The authorities claimed they were conspiring to overthrow the government, which is dominated by a different community, the Moors, but offered no evidence to substantiate this claim.

73 It took AI several months to compile details of the killings, but prisoners released in March 1991 provided information and themselves bore scars from torture. 74 In a few cases, prisoners are reported to have been deliberately executed without trial. 75 Most of the reported deaths, however, were due to torture in both military barracks and police stations.

76 Torture of political detainees in Mauritania has been routine since 1986, but it has never before been used on such a scale.

77 After the March 1991 releases, over 70 detainees were still held and were apparently due to be brought to trial in April.

Scrolling through context

Clicking on the left and right arrow will scroll back or forward through the context window. Here we click right.

A03: <s>-units 78 to 88 (of a total of 1051 <s>-units)

<< >> File info for A03 : Go! Show POS-tags Don't colour wordclass

78 However, the trial has been postponed indefinitely.

79 Mali

80 Over 150 men, women and children were killed by Mali's security forces in March, after a wave of pro-democracy demonstrations and riots. 81 In the wake of the killings the government led by President Moussa Traoré was overthrown.

82 At least 40 people were killed in Bamako on 22 March after a student demonstration was stopped by the security forces and rioting broke out.

83 A state of emergency was declared but at least another 40 people were killed the following day. 84 Some were shot dead when thousands of women, protesting against the killings, tried to march on the President's residence. 85 Up to 65 protesters were reported to have been burned to death when security forces set fire to a shopping centre in which they were seeking refuge.

86 Soldiers also shot at mourners burying the dead at two cemeteries in Bamako. 87 Amnesty International called on the Malian Government to halt the killings.

88 Following a general strike and calls for his resignation, the President was arrested on 26 March by fellow army officers.

Scrolling through context

Clicking on the left and right arrow will scroll back or forward through the context window. Here we click right.

A03: <s>-units 78 to 88 (of a total of 1051 <s>-units)

78 However, the trial has been postponed indefinitely.

79 Mali

80 Over 150 men, women and children were killed by Mali's security forces in March, after a wave of pro-democracy demonstrations and riots. 81 In the wake of the killings the government led by President Moussa Traoré was overthrown.

82 At least 40 people were killed in Bamako on 22 March after a student demonstration was stopped by the security forces and rioting broke out.

83 A state of emergency was declared but at least another 40 people were killed the following day. 84 Some were shot dead when thousands of women, protesting against the killings, tried to march on the President's residence. 85 Up to 65 protesters were reported to have been burned to death when security forces set fire to a shopping centre in which they were seeking refuge.

86 Soldiers also shot at mourners burying the dead at two cemeteries in Bamako. 87 Amnesty International called on the Malian Government to halt the killings.

88 Following a general strike and calls for his resignation, the President was arrested on 26 March by fellow army officers.

Sorting

Lexicographers like to sort their results to put similar contexts together.

Your query "easy" returned 14386 hits in 2717 different texts (98,313,429 words [4,048 texts]; frequency: 146.33 instances per million words) (0.372 seconds - retrieved from cache)

No	Filename	Hits 1 to 50	Page 1 / 288
1	A04_40	Pater's book is a classic, and easy to find in a library.	
2	A04_43	The library user who looks for books of art criticism is not necessarily going to have an easy ta	
3	A04_498	It is easy to sympathise with a traveller who writes: 'I have seen a quantity of things here — ch	fountains and pictures; and my brain is at
4	A04_831	Many readers must have sighed with regret that so few of Gainsborough's letters have survived, since the charm of his style is so fresh; it is easy to sympathise with him writing about his professional commitment to portraiture, on behalf of two fine ladies, his daughters:	

Your query "easy" returned 14386 hits in 2717 different texts (98,313,429 words [4,048 texts]; frequency: 146.33 instances per million words), sorted on position +1 (14386 hits) (data retrieved from cache)

Sort parameters: Position: Tag restriction: exclude Starting with letter:

No	Filename	Hits 1 to 50	Page 1 / 288
1	HH1_3505	ask if it would have succeeded? 'I am not so easy	a conquest!' she stormed. 'Does that answer your
2	C95_652	Such fantastic care could lead to a somewhat distorted view of how easy	a fish is. Astynax mexicana — A hard water Characin which
3	KCU_9955	fucking easy, what were the sex pistols? That's fucking easy	A group Yes, a punk group, that was one Who
4	HAC_5977	a hard disk does make a great deal of difference to how easy	a machine is to use. Personally I think that any manufacturer
5	B19_335	deny another, and so the stake must have been all too easy	a rendezvous for anyone who tried to maintain their integrity. Nevertheless
6	BNL_1535	persons do not use high street cash dispensers, they are too easy	a target for opportunistic thugs. Variety is important. Collect your
7	HW8_597	a couple of hours before hitting the road, I'm too easy	a target for the cops in that thing and I'm over

Sorting parameters

The sorted window has a new set of sort parameters that can be set. Position 1 right is the default, and often most useful, but we may want to start at a particular letter. Here we choose “t” and show part of p. 10 of the result.

Your query "easy" returned 14386 hits in 2717 different texts (98,313,429 words [4,048 texts]); frequency: 146.33 instances per million words), sorted on position +1 (14386 hits) (data retrieved from cache)

<input type="button" value="<"/> <input type="button" value="<<"/> <input type="button" value=">>"/> <input type="button" value=">"/>		Show Page: <input type="text" value="1"/>	Show Sentence View	Show in random order	New Query <input type="text"/> : <input type="button" value="Go!"/>
Sort parameters:		Position: <input type="text" value="1 Right"/>	Tag restriction: <input type="text" value="no restrictions"/>	<input type="checkbox"/> exclude	Starting with letter: <input type="text" value="all"/>
		Hits 1 to 50		Page 1 / 288	
No	Filename				
1	HH1.3505	ask if it would have succeeded? 'I am not so	easy	a conquest!	med. 'Does that answer your
2	C95.652	Such fantastic care could lead to a somewhat distorted view of how	easy	a fish is. Ast	exicana — A hard water Characin which
3	KCU.9955	fucking easy, what were the sex pistols? That's fucking	easy	A group Yes,	group, that was one Who
4	HAC.5077	a hard disk does make a great deal of difference to how	easy	a machine is	Personally I think that any manufacturer
5	B19.335	deny another, and so the stake must have been all too	easy	a rendezvous	one who tried to maintain their integrity. Nevertheless
6	BNL.1535	persons do not use high street cash dispensers, they are too	easy	a target for	nistic thugs. Variety is important. Collect your
7	HW8.597	a couple of hours before hitting the road, I'm too	easy	a target for	in that thing and I'm over
8	EEM.387	has been largely discredited. It is, in fact, so	easy	a target that	s reacting against it have constructed a revised view
9	J15.7	define precisely what we mean by 'money' — not as	easy	a task as it se	Money is more than just notes

Your query "easy" returned 14386 hits in 2717 different texts (98,313,429 words [4,048 texts]); frequency: 146.33 instances per million words), sorted on position +1 starting with letter T (5815 hits) (data retrieved from cache)

<input type="button" value="<"/> <input type="button" value="<<"/> <input type="button" value=">>"/> <input type="button" value=">"/>		Show Page: <input type="text" value="1"/>	Show Sentence View	Show in random order	New Query <input type="text"/> : <input type="button" value="Go!"/>
Sort parameters:		Position: <input type="text" value="1 Right"/>	Tag restriction: <input type="text" value="no restrictions"/>	<input type="checkbox"/> exclude	Starting with letter: <input type="text" value="T"/>
		Hits 451 to 500		Page 10 / 117	
No	Filename				
451	F9D.307	accumulated over 20, 30 or 50 years. It is not	easy	to 'prove' this. But I can offer three pieces	
459	HVS.944	full marks and von and er but it's all too	easy	to to focus on one or two aspects of say a	

Simple distribution

To look at various distributions, pull down the main search results action menu and click on distribution. The Adj *easy* is much more common in written than spoken language.

Your query "easy" returned 14386 hits in 2717 different texts (98,313,429 words [4,048 texts]; frequency: 146.33 instances per million words) ...272 seconds - retrieved from raskit

No	Filename	Hits 1 to 50	Page 1 / 288
1	A04_40	Pater's book is a classic, and easy to find in a library.	
2	A04_43	The library user who looks for books of art criticism is not necessarily going to have an easy ta	
3	A04_498	It is easy to sympathise with a traveller who writes: 'I have seen a quantity of things here -- ch	fountains and pictures; and my brain is at
4	A04_831	Many readers must have sighed with regret that so few of Gainsborough's letters have survived, since the charm of his style is so fresh; it is easy to sympathise with him writing about his professional commitment to portraiture, on behalf of two fine ladies, his daughters:	

Your query "easy" returned 14386 hits in 2717 different texts, sorted on *position +1* starting with letter *T* (5815 hits). The current solution set was found in 1959 texts. (1.124 seconds)

Categories:

Categories (for crosstabs only):

The following distribution was found:

Spoken or Written:

Category	No. of words	No. of hits	Dispersion (over files)	Frequency per million words
Written	87,903,571	5,416	1,726/3,140	61.61
Spoken	10,409,858	399	233/908	38.33
total	98,313,429	5,815	1,959/4,048	59.15

Desired text type:

Downloading results

Your query "whether <<s>> (or not)" returned 5321 hits in 1941 different texts (98,313,429 words [4,048 texts]); frequency: 54.12 instances per million words (0.962 seconds - retrieved from cache)

No	Filename	Hits 1 to 50	Page 1 / 107
1	A04_1325	From a critic's point of view, a label, whether ending in 'ism' or not, is convenient.	
2	A05_943	Shakespeare's play has an arranged duel which miscarries, and which takes off a divided, gambling man who has wondered whether or not it might be better to end his life.	
3	A05_1138	Whether it is or not, the poem can be called distinctive — distinctive both of Larkin and of Amis.	
4	A05_1318	If you were to tell me that there are people, like the man upstairs to whom you now threaten to turn yourself in, who actually do have a strong sense of themselves, I would have to tell you that they are only impersonating people with a strong sense of themselves — to which you could correctly reply that since there is no way of proving whether I'm right or not, this is a circular argument from which there is no escape.	

Your query "whether <<s>> (or not)" returned 5321 hits in 1941 different texts (98,313,429 words [4,048 texts]); frequency: 54.12 instances per million words

No	Filename	Hits 1 to 50	Page 1 / 107
1	A04_1325	From a critic's point of view, a label, whether ending in 'ism' or not, is convenient.	
2	A05_943	Shakespeare's play has an arranged duel which miscarries, and which takes off a divided, gambling man who has wondered whether or not it might be better to end his life.	
3	A05_1138	Whether it is or not, the poem can be called distinctive — distinctive both of Larkin and of Amis.	
4	A05_1318	If you were to tell me that there are people, like the man upstairs to whom you now threaten to turn yourself in, who actually do have a strong sense of themselves, I would have to tell you that they are only impersonating people with a strong sense of themselves — to which you could correctly reply that since there is no way of proving whether I'm right or not, this is a circular argument from which there is no escape.	

Now click go!

Download form

Download concordance

Output format options

Choose operating system on which you will be working with the file:	<input type="text" value="UNIX (incl. OS X)"/>
Print codes (numbers) or full values for metatextual categories:*	<input type="text" value="full values"/>
Mark query result in sentence (format: <<< result >>>):	<input type="text" value="yes"/>
Size of context:	<input type="text" value="1 <s>-unit"/>
Download both tagged and untagged version of your results:*	<input type="text" value="yes"/>
Write information about order of categories at the beginning of file:*	<input type="text" value="no"/>
Format of output: KWIC or list:*	<input type="text" value="List"/>
Include corpus positions (required for re-import)*	<input type="text" value="Yes"/>
Include URL to context display*	<input type="text" value="Yes"/>
Enter name for the downloaded file:	<input type="text" value="gawron"/>

Default results (1 example)

1 A04 1325 From a critic 's point of view , a label , <<< whether >>> ending in
‘ ism ’ or not , is convenient .

From_PRP a_ATO critic_NN1 's_POS point_NN1 of_PRF view_NN1 ,_PUN a_ATO
label_NN1 ,_PUN <<< whether_CJS >>> ending_VVG in_PRP ‘_PUQ
ism_UNC ’_PUQ or_CJC not_XXO ,_PUN is_VBZ convenient_AJO ._PUN

Written Written books and periodicals W:ac:humanities_arts

1985-1993 Academic prose Beginning sample Book

Informative: Arts

High unknown UK and Ireland Male Sole

Adult Mixed Medium n/a n/a n/a n/a n/a n/a

n/a n/a n/a n/a n/a n/a

[http://bulba.sdsu.edu/bncweb-cgi/context.pl?text=A04&qname=nosol&](http://bulba.sdsu.edu/bncweb-cgi/context.pl?text=A04&qname=nosol&refnum=0&theData=0&len=0&showTheTag=0&color=0&begin=1325&spids=1&interval=11&first=yes&urlTest=yes)

[refnum=0&theData=0&len=0&showTheTag=0&color=0&begin=1325&spids=1&](http://bulba.sdsu.edu/bncweb-cgi/context.pl?text=A04&qname=nosol&refnum=0&theData=0&len=0&showTheTag=0&color=0&begin=1325&spids=1&interval=11&first=yes&urlTest=yes)

[interval=11&first=yes&urlTest=yes](http://bulba.sdsu.edu/bncweb-cgi/context.pl?text=A04&qname=nosol&refnum=0&theData=0&len=0&showTheTag=0&color=0&begin=1325&spids=1&interval=11&first=yes&urlTest=yes)

77845 77845

Abbreviated results

- 1 A04 1325 From a critic 's point of view , a label , whether ending in 'ism ' or not , is convenient .
- 2 A05 943 Shakespeare 's play has an arranged duel which miscarries , and which takes off a divided , gambling man who has wondered whether or not it might be better to end his life .
- 3 A05 1138 Whether it is or not , the poem can be called distinctive — distinctive both of Larkin and of Amis .
- 4 A05 1318 If you were to tell me that there are people , like the man upstairs to whom you now threaten to turn yourself in , who actually do have a strong sense of themselves , I would have to tell you that they are only impersonating people with a strong sense of themselves — to which you could correctly reply that since there is no way of proving whether I 'm right or not , this is a circular argument from which there is no escape .

Collocates

Definition

The *collocates* of a word w are the words that co-occur with w most often. [First pass]

Definition

The *collocates* of a word w are the words that, statistically, are most strongly associated with w .

Requesting collocates

Your query "powerful" returned 7062 hits in 1919 different texts (98,313,429 words [4,048 texts]; frequency: 71.83 instances per million words)

(Click for information - refresh (slowly))

No	Filename	Hits 1 to 50	Page 1 / 142
1	A03.341	Professional fundraisers utilize the powerful medium of advertising in their national autumn 1990 about Iraq's treatment of Kurds and its human rights record over a decade.	
2	A04.627	A powerful sitter may also impose a requirement that the portrait looks impressive, consequent power struggle in a picture; Queen Elizabeth I of England was as firm as the appearance matters less than the imprint of authority.	
3	A04.1170	The persuasive tax advantages for the rich to contribute works of art to public collections also play a part, contributing a measure of prestige shared between artists and the powerful in society.	
4	A04.1540	As a sound is a mantra, so an image is a yantra, the two used together making a powerful combination for spiritual exercises.	
5	A05.365	Another powerful smell, of decaying rubbish, came mainly from the ash-pits at the far end of the corridor, but also from a deposit scattered over the floor.	

Collocations... Go!

- New Query
- Thin...
- Frequency breakdown
- Distribution
- Sort
- Collocations...**
- Download...
- Categorize hits...
- Save current set of hits...

Collocation settings window

BNC Collocation Settings

Calculate over sentence boundaries:

no :

Include lemma information:

no :

Maximum window span:

+/- 5 :

Submit

Example: *powerful* collocates

Collocation parameters:

Information:	<input type="text" value="collocations"/>	Statistics:	<input type="text" value="Log-likelihood"/>
Collocation window span:	<input type="text" value="1 Left"/> - <input type="text" value="3 Right"/>	Basis:	<input type="text" value="whole BNC"/>
Freq(node, collocate) at least:	<input type="text" value="5"/>	Freq(collocate) at least:	<input type="text" value="5"/>
Filter results by:	Specific collocate: <input type="text"/>	and/or tag: <input type="text" value="any adjective"/>	<input type="button" value="Submit changed parameters"/> <input type="button" value="Go"/>

There are **9846** different types in your collocation database for "[word="powerful"%]". (Your query "powerful" returned 7062 hits in 1919 different texts) (18.369 seconds)

No.	Word	Total No. in whole BNC	Expected collocate frequency	Observed collocate frequency	In No. of texts	Log-likelihood value
1	most	97,999	35.401	647	448	2549.7677
2	a	2,164,238	781.812	2301	1160	1989.3687
3	more	209,333	75.620	671	472	1749.4068
4	tool	2,180	0.788	75	64	537.7234
5	very	119,400	43.132	260	200	501.9577
6	influence	10,324	3.729	92	77	414.2255
7	and	2,616,708	945.263	1564	890	347.4752
8	force	15,497	5.598	85	74	304.1962
9	lobby	1,118	0.404	39	36	280.6709
10	rich	6,691	2.417	61	51	277.2885
11	weapon	1,938	0.700	42	39	262.2437
12	than	144,567	52.224	207	172	261.3607
13	enough	31,123	11.243	96	89	242.6597
14	is	990,281	357.730	675	454	225.2373
15	immensely	642	0.232	27	24	204.5034

Example: *powerful* noun collocates

Collocation parameters:

Information:	<input type="text" value="collocations"/>	Statistics:	<input type="text" value="Log-likelihood"/>
Collocation window span:	<input type="text" value="1 Left"/> - <input type="text" value="1 Right"/>	Basis:	<input type="text" value="whole BNC"/>
Freq(node, collocate) at least:	<input type="text" value="5"/>	Freq(collocate) at least:	<input type="text" value="5"/>
Filter results by:	Specific collocate: <input type="text"/>	and/or tag: <input type="text" value="any noun"/>	<input type="button" value="Submit changed parameters"/> <input type="button" value="Go"/>

There are **9846** different types in your collocation database for "[word="powerful"%]". (Your query "powerful" returned 7062 hits in 1919 different texts)

No.	Word	Total No. in whole BNC	Expected collocate frequency	Observed collocate frequency	In No. of texts	Log-likelihood value
1	influence	7,831	0.981	72	61	477.6129
2	tool	2,177	0.273	49	44	412.5753
3	weapon	1,938	0.243	35	34	279.1873
4	force	13,308	1.667	48	45	230.2471
5	man	58,595	7.338	64	55	164.1834
6	voice	16,127	2.020	38	35	151.2464
7	body	24,588	3.079	42	27	141.8188
8	lobby	944	0.118	16	16	125.5725
9	header	713	0.089	15	10	124.2254
10	effect	22,545	2.823	37	33	122.1823

Min frequency may matter

$k = 3$, medicine

References I

Hoffmann, S., S. Evert, N. Smith, D. Lee, and B. P. Prytz. 2008.
Corpus Linguistics with BNCweb: a Practical Guide.
New York: Peter Lang.