Spokes for the BNC

This page contains the documentation of the search engine Spokes for the spoken data of the British National Corpus. Spokes provides access to the spoken component of the BNC.

Query syntax

Spokes uses the SlopeQ 2 query syntax. The examples below are customized to show how the syntax can be used for searching the spoken English data from the British National Corpus. For practical reasons the number of examples illustrating each query in this presentation is very limited. However, a link to a page with all the results is given for each query.

Surface queries

This is the simplest type of queries. Words are written in into the query box in their plain orthographic form. The results are occurrences of the particular forms submitted in the query. Compare the query and example results:

stuff

#	Left	Match	Right
1	Is it possible that erm , this , this wad of	stuff	that you 've had
2	The only change that I have got here are minor things that were references from a previous meeting all the difficult	stuff	comes from the Quiff and the suggestion forms that I have not looked at yet , and so if you want to do that around a table because I do n't know whether that s right .
3	This is what worries me , if you have got too much	stuff	going through one person that .
4	all sorts of	stuff	like that .
5	Erm well I 'm still on	stuff	from last term .

Queries of this kind can include a series of terms (positions in the query). The input is a series of words in regular spelling and the results show occurrences of the whole sequence, e.g.:

make up

#	Left	Match	Right
1	Yes , but you know for years , when they were growing up in their teens , I was the the tennis net between him and his daughters , where 's she going , who 's she going with , who 's bringing her home , is she wearing eye	make up	, why 's she wearing nylons and not ankle socks ,
2	Some say they 'll take over the children 's education completely as they try to help their youngsters	make up	for lost time .
3		Make up	you bloody mind !
4	And I 'd like you to both spend quite a bit of time just playing with the numbers that	make up	three sixty .

#	Left	Match	Right
5	Now it 's sticking out a mile , but the time home that	make up	another ruddy story to give

Queries of this kind can be fruitfully used with set phrases, some phrasal verbs and some collocations.

Base form queries

These queries make use of the lexical annotation of the data in the corpus. The results are the occurrences of different forms of the given word. The query is written in triangular brackets as an equation “lemma=”. After the equation mark you put the base form of the word.

The base form of the word is: the infinitive for verbs, the singular for nouns, the positive grade for adjectives. Since verbs are the most inflected words in English we will illustrate this functionality with a verb.

<lemma=go>

#	Left	Match	Right
1	I do n't particularly want to	go	
2	No doubt this	goes	through
3	I do hope we 're not	going	to extend our current lawyers who we use in this particular issue , erm , whatever this amount of money is
4	The the the treasurer has already	gone	through the report.
5	The children	went	to on tip toe but the frogs stopped singing.
6	if if those objections were were sustained , er you know	going	back to what was was said last week , then I think Selby might be in difficulties in in meeting its housing requirement .

This kind of query is a simple way to research the use of the word in all its forms.

In general it is possible combine various functionalities. For example, you can put a surface form and a base form as two terms in one query, e.g.

<lemma=take> time

#	Left	Match	Right
1	you know it takes time to get socialized into something and it	takes time	to get used to a way of working .
2	I can understand if they	take time	on a stray dog I often wondered if they realise
3	She lets me	take time	off , I mean she 's really understanding .
4	Yes , I mean it would be nice if it was possible , you know , after having	taken time	out , or some time out , over a period to be able to get back into a more permanent career structure , but maybe in the future this 'll come .
5	Well do you think it 's time that you started	taking time	off in lieu ?

#	Left	Match	Right
6	While they went into the firing blackboards to fire your rounds and you , all you was allowed to fire was five five and then you had to wait , take your turn course everybody had got to go in , they 'd got to check them as everybody got had fired their five there were n't one left up the spout like there was before , course it	took time	see , and it took us a day , I would say a day , to fire five rounds of ammunition .

Such queries are useful with collocations where one of the words always or usually takes just one form.

Operators

Alternative

This operator is represented by the pipe sign “|”. The words separated by the operator are variants of the query term. The results are occurrences of all of them, e.g.:

can|could

#	Left	Match	Right
1	Because I	could	n't find anything under procedures about signing the job descriptions and indeed this persons specification total , I do n't think but I mean is that something like you audit it on , its that I am not clear who should award job specifications .
2	Before you pass on that	could	I just ask one question , which is how far are B T with it ?
3	But	can	I just refer you back to the words of P B G three where it says quite clearly in paragraph thirty three .
4	I 'll just nip back and see if I	can	get the thing .

You can input base forms as variants in the query. In that case, occurrences of all forms of all the words will appear in the results, e.g.

<lemma=connect>|<lemma=join>

#	Left	Match	Right
1	We waited , tea-less all day for the gas man to	connect	the cooker .
2	and after say three o'clock in the afternoon they 'd do activities , work	connected	with
3	Ah , well you 'll have to take it up there then , you 'd be better off joining it at the back if you 're gon na	join	it
4	But no , but I ca n't could n't afford to be rude to him , I 've only just	joined	the choir .
5	Well want to know if Rick	joins	in the conversation will we get any vouchers for doggy food ?

As the examples show, the most natural application of such queries is with synonymous words.

It is also possible to use the alternative operator for one of the terms in a longer query. In the example below, the results are occurrences of all the sequences:

switch|turn off

#	Left	Match	Right
1	It do n't	switch off	I 'll have to go and do it .
2	well it 'll	switch off	when it 's done
3	You could	turn off	for Abergavenny again .
4	There 's something wrong with this you know , it wo n't	turn off	
5	Yeah but there was a	turn off	and I did n't know whether the turn off was at Cardiff and we went past it .

Note: the alternative operator can be applied to single terms in the query i.e. words, not whole phrases. That is why the above query has the form “switch|turn off” and not “switch off|turn off”. The latter query searches for phrases “switch off off” and “switch turn off”, none of which is present in the corpus.

Slop factor

This important functionality allows to search for a discontinuous string of words. The query specifies how many words may intervene between the terms of the query. This parameter is known as the slop factor and it is now set in the search menu, e.g.:

get over (slop factor = 1)

#	Left	Match	Right
1		get blisters over	me wherever it touches try and eat one , I 'd have them on my tongue and in my mouth and everywhere .
2		Get January over	and we shall be alright .
3	The singer had fought and won his battles against bulimia , drug and alcohol abuse and had publicly spoken about it to help others	get over	the nightmare of addiction.
4	we 'll	get it over	with somehow , said oh Sunday afternoon probably .
5	And I said oh he 's at home I said this week I said and that oh and she said well tell him to	get over	here and we 'll sort him out
6	you see you	get , over	the years it probably expands and it

Note: the provided number is the maximum number of intervening words. Strings with fewer or no intervening words will also appear in the results.

Also note that puctuation marks are counted as words, as is shown by the last result in the table above.

Slop factor queries obviously can have query terms other than surface forms. The following example illustrates again a phrasal verb , but all forms of the verb are allowed.

<lemma=take> up (Slop factor = 1)

#	Left	Match	Right
1	But if er if it all going to one side I think it 's going to be	take up	half the window wo n't it .
2	think those points perhaps ought to be	taken up	at the General Purposes Committee since er we have the problem of their decisions .
3	But the two little trays in the middle of that beautifully re-furbished kitchen with every mod con , the whole of the floor was	taken up	with sheets of news papers , two big litter trays and a sack full of .
4	Well , Christine	takes him up	there if they come here for the weekend .
5	I think what Mr has said , is quite right , is that this is a carry forward for this year , and er , we 've , we 're clobbering it	taking it up	really is what we 're saying .
6	Erm , the first time I met him was when he was doing a very tricky stitching job , I	took Mollie up	
7	It goes on to say and it came about that in thy journeying eastward they eventually discovered a valid plain and a land of Shinar and they	took up	growing there and they began to say each one to the other come on let us make bricks and bake and bake with a burning process , so bricks served as stone for them and

Slop factor with relaxed order

These queries allow intervening words up to the specified number and the query terms may appear in any order. In the run the query, one should uncheck the “order” box located next to the slop factor indicator.

good people (Slop factor = 1) ("order" unchecked)

#	Left	Match	Right
1	Furthermore , it does n't just take both into account in terms of some vague philosophical waffle , you know , that anybody could come to , sitting on a bar stool , after they 've had enough er , dry martinis , you know as people sometimes	good , people	sometimes bad .
2	You 've still got to get	good experienced people	in their fifties and er whilst a lot of people have experience or a lot of experience er from their their previous working life it 's not always easy to make the jump into
3	You still have to get	good people	in their fifties .
4	That 's nice , I 'm , it 's very	good that people	are going anyway , it 's got

#	Left	Match	Right
5	And one of the reasons why the black women in Liverpool put this exhibition together erm and they tried to pick out different kinds of jobs that women had to show that black women could do those jobs , and they were saying look , there 's lots of stereotypes erm about the kinds of jobs that black and Asian women can do that fit in with their personalities , and you know the world is wide open and you are able to do this and you do n't have to be erm a singer or a model or erm or erm erm a runner , you know , the whole stereotypes in terms of what black	people are good	at and they 're saying let's break away from this , let's show the kinds of things that we can do and we can do anything that we set our minds to and erm the exhibition is very positive , actually .
6	The jobs that	people need good	education to do .
7	This was , this is what I call an optimistic view , and this was Rousseau 's view of human nature , that basically	people were good	, and er , cooperative , and it was the bad things in human nature that had to be explained , not the good .

The relaxed order in pure form is available with the box unchecked and with no intervening words (Slop factor = 0):

did you ("order" unchecked)

#	Left	Match	Right
1		Did you	wish to ?
2		Did you	get that one Simon ?
3	What	did you	say ?
4	Well , if	you did	you do , you 'd use a damn sight more than twenty tapes a week so
5	Could I before you answer that question ,	you did	ask , did n't you , ask him to be hypothetical and say what would he say in a year 's time , and I think what
6	I think if	you did	n't go to Lucy 's party I do n't think .

The relaxed order may be combined with other functionalities. For example, in the next query we have slop with relaxed order, multiple terms in the query, which are base forms.

<lemma=question> <lemma=ask> (Slop factor = 2) ("order" unchecked)

#	Left	Match	Right
1	I think Richard 's	asking the question	shall we take this item now or later .
2	Can I go back and	ask you two questions	
3		Ask Gary that question	
4	It is now up to members , to make observations	ask questions	, and then an amendment has been put , which is er , accompanied to , direct negative over the recommendations .
5	as opposed to prompting	questions to ask	
6	That 's the	question we asked	

#	Left	Match	Right
7	All right that 's how we play the game a	question 's asked	about every record we play the question about this next record give us a ring on Nottingham three four three four three four if you know the answer is , Which local group had a hit with this one ?

Notice how each functionality contributes to shaping the set of the results. The use of base forms allows various forms of both words. Slop and relaxed order allows for a variety of constructions in which the queried words interact.

Negation (restricting results)

This operator excludes specified variants of query terms from the results. To be used effectively, it must be combined with query types that produce variation in the results. Negation is marked by a pipe sign with an exclamation mark “|!”, which is to be read as “but not”. You type it after the given query term **without space**. The example shows how it is used with a base form query. The specified form of the word is excluded from the results:

<lemma=see>|!see

#	Left	Match	Right
1	Try to think back who we	saw	at , I know Terence , , Terence Alexander was one .
2	Well I	saw	Murray , yeah , I did once .
3	Richard is responsible for	seeing	that it is done .
4	I mean she 's	seeing	people and
5	It 's not a case of being paranoid , it 's a case of seeking fairness for all the school children , and then some of them receiving disproportional , you know higher , er amounts of money , then that must be	seen	to be unfair , and I 'm certainly well supported in that .
6	He maybe	sees	a different process and and I do n't know about you Chairman but it would it would help me to see if to hear whether they they see a different process at work .

Regex queries

Queries of this type make use of special symbols and quantifiers. Each query is a formula describing a whole set of possible strings of signs (words, sequences of words). The results are occurrences of all these strings found in the data.

Wild card and quantifiers

A full stop “.” is a wild card, it stands for any sign.

A plus “+” is a quantifier: the preceding sign can appear one or more times.

An asterisk “*” is another quantifier: the preceding sign can appear zero or more times.

These symbols can be used on their own with standard signs, but it is most effective to combine the wild card with one of the quantifiers.

“.+” means that in this part of the query **any sign or sequence of signs** may appear.

“.*” means that in this part of the query **any sign or sequence of signs or nothing** may appear.

Note the difference between the quantifiers. If you use the plus, the preceding symbol (which may be any symbol if you use the wild card) needs to appear at least once in each item found. With the asterisk it may not appear at all.

Compare the examples:

trouble.*

#	Left	Match	Right
1	in the cell block and we and somebody was causing	trouble	many many years ago
2	Sorry to	trouble	you , do you own the caravan I 'm renting ?
3	Still giving you	trouble	Jim ?
4	who who do you think , I mean , , you know , the government 's making it pretty clear at the moment , Mrs , that that er , in these	troubled	times something 's gon na get squeezed because there 's so , many people out of work
5	You get some	troublemakers	here sometimes do n't you .
6	I 've got	troubles	now deep , deep troubles I might as well go in that way round .
7	She was a bit	troublesome	.

The above query yields words that starting with the letter sequence “trouble”. The very word *trouble* is also included. To exclude it from the results, you need to use the other quantifier as below.

trouble.+

#	Left	Match	Right
1	I was initially	troubled	by a philosophical problem .
2	Like many people , some years ago was	troubled	by nuisance telephone calls .
3	British Rail tell us that the 6.30 Aylesbury to Marylebone train service has been cancelled this evening , while the buses continue to operate a	trouble-free	service .
4	You get some	troublemakers	here sometimes do n't you .
5	I know they 're probably the same people that lived here then but er There 's all this talk about problems and	troubles	er we we 've never noticed it .
6	Erm used for	troubleshooting	when it does n't start .

#	Left	Match	Right
7	I 'd got a	troublesome	cough developed , and now looking back through the years , er it would have been a sort of hay-feverish condition that I have been a bit bothered with .

The use of wild card without quantifiers is also possible, but probably less useful for linguistic research. It allows for crossword-like queries which yield word forms of a set number of letters, containing set letters at certain positions, e.g.:

h...l.

#	Left	Match	Right
1	they 're trying to	haggle	with him now
2	This is I think there is an administrative question how we	handle	this and it seems to be running very well .
3	Sometimes you	hardly	get any time to cross the road before the actual lights turn , and particularly the handicapped and the blind
4	And what you 're suggesting	Harold	is that er he 's going back now to er er the erm er the idea of the , I 'd suggest that what would have been the alternative was erm international exploitation .
5	and the lad who was dealing with it all , we had some	hassle	from another quoter .
6	Pardon , oh yes , commended , not	highly	.

The elements of regex queries can be combined with other functionalities of the syntax. For example, it is possible to query for a whole group of similar-looking words in all their forms.

<lemma=time.+>

#	Left	Match	Right
1	Well actually a girl friend of mine came round here yesterday , oh , lunch	timeish	was it , and erm , Henry forget his ball
2	And , obviously to get them to do the right job in the right , within the right	timescales	.
3	Right so I 'll be the	timekeeper	
4	and I think all true clothes , I mean true sort of	timeless	costume , you can wear at anytime .
5	but it can be so organized on the	timetable	that they are withdrawn at a different period each week
6	you might 've had some	timetables	like they have on the train

Another possibility is the alternative combined with the wild card and quantifier, e.g.

differ.+|vari.+

#	Left	Match	Right
1	If we 're talking about outside auditors , whether it 's a letter , a fax or whatever wo n't make any	difference	as they wo n't quibble over things like order forms and stuff .

#	Left	Match	Right
2	He did n't bring a sample home of that cos he , he brought the same design but in a	different	colour , but it was more greeny
3	Yar , I think I got a copy but I just sort of filed it with all the Quality Manual stuff , as there were various	different	things which needed .
4	There are	various	ways about that as there are with many road schemes er where there are structure plan policies for a particular scheme
5	In that context , how does an outer northern relief road impact	differently	in terms of traffic relief on Knaresborough as opposed to an inner Northern relief road ?
6	So that there is a measure of the bias and the flat distribution is a measure of er erm the fact that we do n't have an estimator that 's got this minimum	variants	property right .
7	Erm , the only variations which are actually going outside the Committee control are as a result of the internal market	variations	which are going on down .
8	you know erm not that they will but er you know it 's , it 's one of these things that , that sort of looks like there 's a bit of	variety	on it when you 've got mm

Grammatical queries

Spokes for the BNC makes use of the BNC tagset. We will not present the tagset in full here, but only discuss its application for our search engine. A list of all the tags is available online.

The tags are three-symbol codes, which classify the exact grammatical form of the given form in the corpus. For example, “NN1” marks a singular common noun, “AJC” marks a comparative adjective and “VVG” marks an –ing form of a lexical verb. You can use exact tags like the ones above, but it is even better to make underspecified queries by means of the wild card (and quantifiers). This is possible because the tags form an ordered system. Thus, all tags starting with “N” mark nouns, all tags starting with “V” mark verbs, in particular lexical verbs are marked by tags with “VV” at the beginning, and all tags starting with “AJ” mark adjectives.

All grammatical queries have the same formula. They are written in triangular brackets as an equation “pos=” (pos stands for part-of-speech). The tag or regex tag is put immediately after the equation sign. Our first example will use an exact tag, the one for past tense of lexical verbs.

<pos=VVD>

#	Left	Match	Right
1	The conservatives	asked	to be represented on that committee and they were refused by the ruling groups .
2	we we we tried , but we never	got	anywhere , and on various issues I 've said , well perhaps if we got us an all party delegation , at least we could have done no worse .
3	You 've already curtailed when , you	said	you 've already spoken , you can not again .

#	Left	Match	Right
4	then I think you 're like the man who	wanted	a cake and you 'll find that he wanted to eat it , so I think you 've got to come to terms and be realistic .
5	I did n't think , I thought they	hated	the sight of each other .

A formula covering a series of tags can be obtained by using the wild card(s). For example, in order to search for nouns in general you can input <pos=N.+>, <pos=N.*> or <pos=N..>. These are obviously not identical as regex queries, but because all tags have three symbols, there will be no difference in the actual results. This use will be illustrated by further examples.

It is possible to combine grammatical search with other functionalities. The following query yields sequences of the form *fit* followed by any noun.

fit <pos=N.+>

#	Left	Match	Right
1	she was complaining about how her keep	fit class	they bloody put on rave music .
2	you look a pretty	fit guy	erm
3	they had to play it again the following Friday and Charlton ran out two-one winners and Walsall finished with nine	fit men	and still that 's the football I suppose .
4	she wanted to go before she was too old and crotchety to get there because David 's he 's in no	fit state	to go anywhere any more .
5	That 'll fit , that 'l	fit Duvane	.
6	it 's going out he said the curtain does n't	fit flush	against the wall , because the wall leans , you know

Note: the query searches for the exact form *fit* and so we get results with both the adjective and the verb, but not inflected forms of either.

This kind of query is very good for researching collocates of a given word.

The next query is quite similar in its design but it also involves the slop factor. The results are sequences starting with a preposition and ending with the word *relief* with one intervening word possible.

<pos=PRP> relief (Slop factor = 1)

#	Left	Match	Right
1	After , after the years of hardship and loss and then everything came	as a relief	, course we were still at war with the Japanese and people were still in Burma
2	I remember very well erm going to the , my father applying	for relief	, and er we had to go and face the erm Court of Referees .
3	Are you suggesting that apart	from this relief	you 've been set you might see Wogan tonight or something you could think
4	They 've started to restore my beauty at last she signed	with relief	
5	Judge David told them , their lives had been devoted	to the relief	of pain and suffering but

The next query combines the grammatical search with the use of a base form.

<lemma=fight> <pos=PRP>

#	Left	Match	Right
1	Sending delegates to Congress would be a small step in the	fight against	centralization .
2	But after losing a protracted	fight for	asylum both parents were deported .
3	You know Annie got into a	fight with	Elsa 's boyfriend .
4	Also on the programme : Percy the penguin , the sole survivor of a mystery illness at the Cotswold wildlife park at Burford , is	fighting for	its life .
5	he writes a poem in which there 's a battle and there 's a character called Cruelty , who comes and	fights against	a character called Mercy .
6	The demonstrators want Britain to apologize for the executions of nine men who	fought for	the island 's independence in the fifties .

Note that each of the functionalities above is used for a separate term in the query. Yet, it is possible to use both base form query and grammatical query with one term. The labels “lemma=” and “pos=” need to be taken in the same pair of brackets then.

<lemma=drink pos=v.*>

#	Left	Match	Right
1	And she	drank	it and drank it !
2	She never , she never	drank	alcohol at all
3	Well there are then	drink	it .
4	Do n't	drink	it too quick !
5	You start off	drinking	then you smoke and you , yeah
6	Have n't	drunk	my coffee yet .

In general, it is possible to add grammatical specification to any other kind of query term by writing it immediately after the term (without a space so that it is the same term). Below we have a regex query for words starting with *four-*, but they must be tagged as adjectives.

four.+<pos=AJ.>

#	Left	Match	Right
1	Er it would be a van that would take a a	four-bedroomed	house with ease .
2		Four-letter	word .
3	The	four-man	team from the democratic Labour movement claim more than £1 million given by Russian miners , was intended for their striking British colleagues .
4	It 's	four-poster	beds .
5	That 's the kind of gimmicky , artificial sort of exercise , when in fact erm in a sense what , what is wanted is , as in	fourteen-year-old	Judith 's case , to , to get into that situation right from the start .

#	Left	Match	Right
6	They have	four-wheel	drive , they have anti-lock brakes , they have power steering .
7	Thirty	four-year	old Michael Shorey entered no plea to charges that he murdered Patricia Morrison and Elaine Forsyth in Holloway in July .

From:
https://docs.pelcra.pl/ - **PELCRA**

Permanent link:
https://docs.pelcra.pl/doku.php?id=spokes_for_bnc

Last update: **2025/04/10 08:52**