

SlopeQ for the BNC: Query Syntax

SlopeQ for the BNC uses the SlopeQ 2 query syntax. The examples below are customized to show how the syntax can be used for searching the data from the British National Corpus the access to which is provided by SlopeQ. 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 the query box in their plain orthographic form. The results are occurrences of the particular forms submitted in the query. Compare the query and the results:

[acknowledge](#)

| # | Left | Match | Right |
|---|---|-------|--|
| 1 | Courts and tribunals acknowledge | | that you are entitled to ' natural justice ' , ie that : |
| 2 | It may be that at a low level of a graduated test scheme teachers may wish to acknowledge | | a pupil 's recognition of only an equilateral or isosceles triangle as a triangle |
| 3 | You can see why he did n't want to acknowledge | | a family connection with someone who was convicted of murder . |
| 4 | The fund is voluntarily administered by members of the Trinity House of Leith , whose Master will acknowledge | | all contributions . |
| 5 | In this article , I want to acknowledge | | openly a number of tensions that I experience , and which I suspect others may recognise . |

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.:

[in vain](#)

| # | Left | Match | Right |
|---|---|---------|--|
| 1 | His sister Charlotte , hoping these efforts were not | in vain | , wrote to a friend : |
| 2 | they reflect a commitment to national identity that one will seek | in vain | in most parts of the Arab world . |
| 3 | the question whether the emancipators had laboured | in vain | is one which should be handled with care . |
| 4 | But he hoped | in vain | |
| 5 | The Roman Emperor Antonine , while he would have looked | in vain | for the wall he built across Scotland 2,000 years before , would have been quite at home in the vast amphitheatre of Celtic Park . |

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 without any space.

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=decide>

| # | Left | Match | Right |
|---|--|----------|--|
| 1 | the constable was required by section 7(4) to inform him that the specimen was to be of blood or urine and that it was for the constable to | decide | which ; |
| 2 | The “ yes ” campaign was also supported by the president of the African National Congress (ANC) , Nelson Mandela , who , whilst abhorring the notion of the white population | deciding | the political fate of the country , appealed to whites to vote “ yes ” for democratic change . |
| 3 | Despite the fact that the state has | decided | to put them to death , many prisoners remain positively patriotic |
| 4 | Since little is known specifically about the difficulty factors of reading tabular data it was | decided | to study these in a range of tables . |
| 5 | The person | decides | consciously or not to devote time and energy to a particular activity . |

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> advantage

| # | Left | Match | Right |
|---|--|------------------|--|
| 1 | Betty had | taken advantage | of her weakness , and here she was . |
| 2 | recent Budget changes give you genuine cause to wonder whether you are | taking advantage | of the concessions available to you , you should talk to an accountant . |
| 3 | Clearly some places , especially a number of fortified towns , were sited to | take advantage | of easily defended positions . |
| 4 | The Pension Loan | takes advantage | of the fact that your pension plan can provide a cash lump sum at retirement |
| 5 | the waves all round him eased and he | took advantage | of the lull to belt the last thirty yards to the beach . |

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.:

[issue|matter](#)

| # | Left | Match | Right |
|---|---|--------|--|
| 1 | It is an excellent way of introducing the | issue | of human rights to school children . |
| 2 | The democratic press expressed concern over the threat to freedom of expression by publishing on July 16 a second | issue | of Obshchaya gazeta , a joint paper published during the coup to overcome censorship . |
| 3 | There 's still the ' | matter | of other women . ' |
| 4 | justified the increased deficit " no | matter | how dangerous for the economy " because of growing social tensions . |
| 5 | but this is a much more serious | matter | if one abandons the assumptions of symmetry and reversibility . |

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=consist>|<lemma=comprise>](#)

| # | Left | Match | Right |
|---|---|------------|--|
| 1 | They usually | comprise | a circle of stake-holes set into shallow trenches with stouter posts located at the doorways |
| 2 | the first is the actual rise in new additional funds brought in , while the second | comprises | the increase in existing funds arising from investment performance |
| 3 | In brief , what this means is that every individual | comprising | the population of interest should have an equal chance of being selected for the sample . |
| 4 | My 1970-3 survey showed that 26% of the TV audience | consisted | of adult males earning less than K450 per annum . |
| 5 | It | consists | of spatial visualization , reasoning and experience . |
| 6 | to that extent change can be said to | consist | of change in community norms . |
| 7 | Under the 1924 Constitution the Republic has an executive President and a National Congress | consisting | of a 120-member Chamber of Deputies and a 22-member Senate ; |

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:

[benefit|profit|gain from](#)

| # | Left | Match | Right |
|---|--|------------------|---|
| 1 | They are in a lot of trouble and we have got to | benefit from it. | |
| 2 | It was people who Labour claimed would | benefit from | its policies . |
| 3 | Both | gain from | being surrounded by the same protective shell |
| 4 | But this increase was minuscule when set against the potential | gain from | an improvement in industrial productivity which would make up only half the gap between Britain and its competitors . |
| 5 | But the Indians rarely | profit from | the mahogany trees cut from their land . |
| 6 | you 'll sell the ice cream you make good | profit from | those and every theatre and every promotional help is done . |

Note: you can have more than just two options of the query term, as shown in the example.

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. The allowed number of intervening words is set with the slider located below the search box. In this Wiki, the value of the Slop factor is indicated by the following expression: (Slop factor = 1,2,3... etc.).

[adopt policy \(Slop factor = 2\)](#)

| # | Left | Match | Right |
|---|--|-----------------------------|--|
| 1 | In practice most councils | adopt a policy | of partial delegation , so that routine matters may be dealt with expeditiously by the committees |
| 2 | despite the fact that most journals seek to | adopt a consistent policy | and insist on receiving manuscripts which follow a specified system . |
| 3 | the internal corporate debate about whether to | adopt a dolphin-safe policy | was ' epic , almost theological in tone . ' |
| 4 | so that they may | adopt a sensible policy | which we can then follow ? |
| 5 | But at the time that the county council had moved to | adopt the policy | , we may be in the working the context of regional policy to be issued by the Secretary of State . |

Note: the provided number is the maximum number of intervening words. Strings with fewer words also appear in the results. Strings with no intervening words will also be fetched, but there are none in the corpus in the example query.

Slop factor can be combined with other functionalities, for example alternative:

[wait in|on \(Slop factor = 1\)](#)

| # | Left | Match | Right |
|---|---|---------------------|--|
| 1 | Reassured that our political leaders are both aware of the problem 's growing dimensions and receptive to our rising anxieties , we | wait in | optimistic but realistic anticipation for crime to be at least effectively reduced . |
| 2 | But | wait , on | the political horizon there comes a general election . |
| 3 | Other nations , especially Japan | wait impatiently in | the wings with their offers of soft loans . |
| 4 | I 'll | wait in | ... the room , all right ? " |
| 5 | Even the Chancellor of Oxford University , Roy Jenkins , has to | wait outside on | the steps of the Clarendon buildings on a cold November morning |

Note: punctuation marks count as words for slop factor.

Slop factor with relaxed order

These queries allow intervening words up to the specified number and the query terms may appear in any order. This option is activated by unchecking the "order" box located next to the Slop factor slider ("order" unchecked).

wrong approach (Slop factor = 2) ("order" unchecked)

| # | Left | Match | Right |
|---|--|------------------------------|------------------------------------|
| 1 | On that | approach Halai looks wrong . | |
| 2 | Some believe the British | approach is wrong | . |
| 3 | Completely the | wrong approach | had been taken by the Government . |
| 4 | To ask about details before establishing the general context is to | approach from the wrong | direction . |
| 5 | What was | wrong with that approach | ? |

The relaxed order in pure form is available when the Slop factor is set to 0:

he would ("order" unchecked)

| # | Left | Match | Right |
|---|------------------------|------------|---|
| 1 | Surely | he would . | |
| 2 | He also said that | he would | press ahead with the controversial plans announced by his predecessor |
| 3 | Had she encouraged him | he would | have given more . |
| 4 | | Would he | be murdering people now in his struggle for existence ? |
| 5 | And | would he | be in bed ? |

The relaxed order may be combined with other functionalities. For example, in the next query we have slop with relaxed order, there is more than one term and the terms are base forms.

<lemma=abuse> <lemma=right> (Slop factor = 1) ("order" unchecked)

| # | Left | Match | Right |
|---|--|-------------------|---|
| 1 | This form of trespass may also be committed by | abuse of right | of entry . |
| 2 | and the defendant subsequently | abuses that right | , then he becomes a trespasser ab initio (from the moment of entry) |
| 3 | (it has always seemed unjust that what was deplored then as a human | rights abuse | Was later marketed under the brand name of acid house) . |
| 4 | its leadership replaced and those officers guilty of gross human | rights abuses | brought to trial , had led to the deadlock in the latest round of UN-mediated peace talks |
| 5 | says that ' all too often , the lack of choices open to the homeless , mean that their | rights are abused | , |

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 allow for a variety of constructions in which the queried words interact.

Negation (restricting the results)

This operator excludes specified variants of query terms from the results. Consequently, 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”. The example shows how it is used with a base form query. The specified form of the word is excluded from the results:

<lemma=remain>|!remaining

| # | Left | Match | Right |
|---|--|----------|--|
| 1 | We remain | remain | cautious for 1993 , but Bowater remains ready to move wherever the upturn in the business is sighted . |
| 2 | Why he did not pay the full amount must remain | remain | a mystery . |
| 3 | and , for a while , the Times of Zambia remained | remained | the sole daily paper . |
| 4 | But Grisedale remains | remains | a sad place . |
| 5 | However , the next owner dismantled the walls in 1685 , since when it has remained | remained | in a state of decay . |

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 predefined 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 directly 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:

[develop.*](#)

| # | Left | Match | Right |
|---|---|-------------|--|
| 1 | for the thesis is that , once | developed | , it is extremely difficult to resist its application . |
| 2 | We grow and | develop | as a result of those interactions . |
| 3 | So , if there are limits to what we can do , and if the | development | of technology brings costs as well as benefits |
| 4 | If we really want to help children | develop | an awareness of which aspects of mathematics are universal |
| 5 | | Developed | by the Dutch physicist J.R.J. van Asperen de Boer , it employs a camera equipped with an infra-red vidicon that allows an image of a small section of a panel to be transmitted to a television screen . |
| 6 | Being a late | developer | , I had been small and slim in comparison to most of my peers , and my appearance had been a large part of my character . |

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

[develop.+](#)

| # | Left | Match | Right |
|---|---|----------------|---|
| 1 | There is no shortfall in the East Lothian HP4 supply , nor any difficulty in the availability or | developability | of land . |
| 2 | upon the main road frontages , through those which | developed | a system of irregular internal streets and lanes , to the few with some form of organized street plan . |
| 3 | CNES | developed | a series of subsidiary companies , entrusted with commercial applications |
| 4 | IBM sells the Stratus machines as System/88 , but the new contract includes a cross-licensing agreement giving IBM and Stratus some rights to each other 's patents , suggesting that IBM may have designs on | developing | its own fault-tolerant Unix for the RS/6000 . |

| # | Left | Match | Right |
|---|---------------------|-------|---|
| 5 | It was developments | | in sampling theory from statistics which weakened the force of both these assumptions , and quite early on in Britain . |

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.:

d....e

| # | Left | Match | Right |
|---|--|--------|--|
| 1 | The answer to these questions (particularly the last) has important ramifications for the raison | d'etre | of the Chinese Wall mechanism . |
| 2 | The report says that the methods used by scientists at the Department of the Environment to calculate future acid rain | damage | are " flawed " . |
| 3 | The author , Susan Hill , was given a large advance to write a sequel to | Daphne | du Maurier 's Rebecca and Emma Tennant is penning a follow-up to Jane Austen 's Pride and Prejudice . |
| 4 | Each year in February and March a different | decade | is remembered , with the festival reaching its climax in the year 2000 . |
| 5 | She was shrewd enough to see that , if all the saintly characters were killed off , readers might be inclined to | deduce | that ' the sure reward of virtue is a fatal accident . ' |
| 6 | He kept her waiting , pushing her | desire | to the limits , enjoying the torment , the agony that was surely written all over her face . |
| 7 | In the event of my dying before remarriage , I | DEVISE | and BEQUEATH all of my real and personal estate whatsoever and wheresoever not already disposed of as to my freeholds in fee simple and as to my personal estate absolutely to the issue of my union with JACQUELINE MYRTLE MITCHELL |
| 8 | McGinlay scored his eleventh goal of the season on Saturday , though he was on the losing side against | Dundee | , and would relish a move to Celtic , no matter their current state . |

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=pain.+>**](#)

| # | Left | Match | Right |
|---|---|---------|--|
| 1 | It 's also potentially the most disastrous — even Rainey has been caught out by the | painful | highside crash . |
| 2 | Not one drop of | paint | hit the wall he was supposed to be painting , but it had done a marvellous job of covering the path and half a garage door . |

| # | Left | Match | Right |
|---|--|--------------|--|
| 3 | Alcohol tends to be cross-addictive with other mood-altering chemicals such as minor tranquillisers , antidepressants , sleeping tablets and some other prescribed drugs such as Pethidine , Morphine and many other | pain-killers | , anti-histamines and the various illegal drugs such as cannabis , amphetamines , LSD , Mescaline , Ecstasy , magic mushrooms and other hallucinogens , cocaine , heroin and opium . |
| 4 | For instance , instead of sending work to the paint shop to be painted , the | painters | were in many cases permanently located at points where work was sufficiently finished . |
| 5 | He | painted | scenes in an amateurish fashion from the medieval romance of Sir Degrevant on the lower walls of the drawing room , some of which have survived . |
| 6 | This was founded in 1696 ; separate music and | painting | academies were established in 1833 . |

The above query yields forms of various words starting with *pain-*. The word *pain* and its inflected forms like *pains* are not included because of the quantifier used in the query.

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

[vari.+!variety](#)

| # | Left | Match | Right |
|---|--|--------------|---|
| 1 | Alternatively , other methods assume some | variability | about a boundary between mastery and non-mastery . |
| 2 | This is because | variable | scribal usage is likely to be functional in some way , just as spoken variation is functional (as suggested in chapter 2) |
| 3 | Thus , ME language states , being so variable , should in principle be suited to the same kind of analysis that we use in present-day social dialectology , and by using | variationist | methods we should be able to explore at least some of the constraints on variation that might have existed in ME . |
| 4 | Distinguishing between a multiplicity of variables in accounting for policy | variations | is , of course , a hazardous exercise , but despite the methodological difficulties it seems clear that a simple agency model , with local authorities implementing national policies with little or no discretion , is far from accurate . |
| 5 | The knowledge of the likelihood of such modifications and of the probable direction of the changes as suggested by the Hungarian data may be helpful in the explanation of | variance | found in the relationship among developed and , recently , also among developing countries . |
| 6 | | Varied | programme is envisaged included participation in the Bridge Crosses , a number of services on 3rd & 10th May , schools , Presbytery , church representatives , and Theology & Development students . |
| 7 | ME sources , however , also contain variation that may be relevant to non-standard | varieties | and casual styles of speech ; hence , there may be considerable time-depth to these variables also . |

The query results do not include examples of the exact form *variety*, but as can be seen the plural *varieties* is included.

Grammatical queries

SlopeQ 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 word 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 these, but it is even better to make underspecified queries by means of the wild card (and quantifiers). This is possible because the tags form a neatly 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 etc.

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 the third person singular present tense of lexical verbs.

`<pos=VVZ>`

| # | Left | Match | Right |
|---|---|----------|--|
| 1 | I 'm sure if ever the occasion | arises | when I want advice on insurance , you 're the first person I 'll come to . ' |
| 2 | they can move on to full doctor status and for many students the chance to experience life in another country more than | makes | up for the extra years of study . |
| 3 | It also | requires | only a fraction of the fees of other European Universities . |
| 4 | The trust has now drawn up detailed plans and | claims | a living museum at Bletchley Park could attract at least 100,000 visitors every year . |
| 5 | I do n't suppose anyone really | wants | to see him , do they ? |

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.+>` or `<pos=N.*>` or `<pos=N..>`. Of course, these are not identical as regex queries, but because all tags have three symbols, there will be no difference in the results. This use will be illustrated in further examples.

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

`beautiful <pos=N.+>`

| # | Left | Match | Right |
|---|-----------------------------|--------------------|--|
| 1 | A paint-effect wall makes a | beautiful backdrop | , whether you try your hand at sponging or go in for a more adventurous colour-wash finish . |

| # | Left | Match | Right |
|---|---|------------------------|--|
| 2 | HOMES on a north Belfast street have rooms with a | beautiful view | today . |
| 3 | Miss Harker removed her bonnet , a | beautiful item | with long blue ribbons , and looked round for somewhere to hang it . |
| 4 | the woman could only hope that moving her right away from the influence of the people she went around with into these | beautiful surroundings | might bring her back to herself . |
| 5 | And you gave me everything , my | beautiful Maggie | . |

This kind of query is very good for researching collocates of a given word that are from a specific grammatical class.

The next query involves the slop factor and a base form. The results are sequences of any form of the word *derive* followed by a preposition with one intervening word possible.

<lemma=derive> <pos=PRP> (Slop factor = 1)

| # | Left | Match | Right |
|---|---|--------------------------|---|
| 1 | More frequently these | derive from | species still found in tropical and semi-tropical habitats |
| 2 | The payments pursuant to the discretionary power | derive from | an overseas source and come within Drummond v Collins 6 TC 525 . |
| 3 | As a result , it is impossible to | derive egalitarianism in | the Marxist sense from a Biblical foundation . |
| 4 | the development of local government reflected economic organization and the political processes which | derived from | it . |
| 5 | A cosmopolitan group | derived mainly from | crosses between bush roses and wild species , many noted for their vigour , scent and exuberant flowering . |
| 6 | The model is | derived by | the processes of data analysis |

It is also possible to apply base form query and grammatical query to a single term. The labels "lemma=" and "pos=" need to be taken in the same pair of brackets then. The following query yields occurrences of all forms of the verb *approach*, but not of the noun.

<lemma=approach pos=V..>

| # | Left | Match | Right |
|---|---|------------|---|
| 1 | As they | approach | a grazing herd of , say , wildebeest , they spread out in line abreast and begin the stalk . |
| 2 | she added : ' A customer is a customer , | approach | them in all the same way . |
| 3 | When The Art Newspaper | approached | the British Museum they pointed out that they do not possess the Altar of Cybele , |
| 4 | When CDC 's intention became clear , it was | approached | by a variety of potential partners , Ousley says , and ' came very close to changing ' its decision |

| # | Left | Match | Right |
|---|-----------------------------|------------|---|
| 5 | We think now , as Christmas | approaches | , and that elusive fishkeeping present becomes top priority , is the perfect time to get the unbiased recommendations of our experts as to what they would buy , given the chance . |

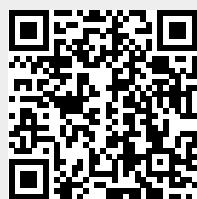
In general, it is possible to add the grammatical specification to any other kind of query term by writing it immediately after the term (without a space). Below we have a regex query for words ending in *-fish*, but they must be tagged as nouns. In this way we exclude adjectives like *selfish*.

[.+fish<pos=N.*>](#)

| # | Left | Match | Right |
|---|--|---------------|--|
| 1 | Experienced , mature | Angelfish | can fill two or three Amazon Sword leaves with eggs |
| 2 | Over-exploitation has led to a collapse in numbers of | bluefish | tuna , swordfish and cod in the Atlantic . |
| 3 | A popular aquarium fish , the range of the Redfin | Butterflyfish | stretches from the tropical Indo-Pacific to South Africa . |
| 4 | We have a 10 gallon tank with an undergravel filter containing two common goldfish , one fancy | goldfish | and a Moor . |
| 5 | Ian Lucas spots some new opportunities with mid-price | catfish | . |
| 6 | Echinoderms , like | starfish | , urchins and sea-cucumbers can be found here . |

From:

<https://pelcra.pl/> - PELCRA



Permanent link:

https://pelcra.pl/doku.php?id=slopeq_for_bnc

Last update: **2017/02/03 00:15**