



# #LancsBox X query cheat sheet

For version 6.0.0

Regular expressions are in blue

## Word queries

cup of tea  
adam's apple

## Wildcards

big\* (any suffix)  
\*tion (any prefix)  
very \* (any token)

## Smart queries

NOUN  
SPLIT\_INFINITIVE  
EMOTION

## Alternation

hit man | hitman  
red | green | blue  
(cup | mug) of tea

## Token queries

[word="[aeiou].*"]	Words starting with a vowel
[hw="like"]	Words with the headword <i>like</i>
[word="professional" pos="J.*"]	<i>professional</i> as an adjective
[word="jump.*" word!=".*ing"]	Words starting <i>jump</i> and not ending <i>ing</i>
[word=="LaTeX"]	<i>LaTeX</i> with case-sensitivity
[word="the" word!="THE"]	<i>the</i> but not in all caps
[pos="J.*" word=".*y"   pos="R.*"]	Adjectives ending with <i>y</i> or adverbs
[sem="F1" (pos="V.*"   pos="N.*")]	Food-related verbs or nouns

## Multi-token queries

[word="green blue"][]	Any word after <i>green</i> or <i>blue</i>
very{2} good	<i>very very good</i>
really []{1,3} day	<i>really</i> and <i>day</i> separated by 1-3 words
big (green blue)? NOUN	Any noun after <i>big</i> , <i>big green</i> , or <i>big blue</i>
[](plus []){2,10}	Lists like <i>W plus X plus Y plus Z</i>

- » Many corpora consider clitics (e.g. n't, 's, 've) separate tokens, but word queries handle them automatically
- » Not all corpora are tagged or use the same token tags (e.g. hw, pos, sem) and tagsets
- » These examples assume [CLAWS7 POS-tags](#) and [USAS semantic tags](#)

## Punctuation

/!/ /?/	! followed by ? (punctuation may be in between)
/' : ./	' consecutively followed by . (no punctuation in between)
okay : /punc="\.\ ?"/	<i>okay</i> consecutively followed by . or ?
/punc="' "/	' or "

- » Tagged corpora must wrap punctuation in <c> elements to support punctuation queries
- » Some punctuation marks can only be searched using punc=""

## XML elements

<u/>	<u></u> (utterance) XML elements (and contained text)
<text genre="fiction" subgenre=".*horror"/>	Horror fiction texts
<s> the	Sentence beginning followed by <i>the</i>
thanks </p>	<i>thanks</i> followed by end of paragraph
</s p" n="1">	End of a first sentence or first paragraph

- » These are examples for the BNC2014. Other corpora use different XML elements and attributes.

## Operators

red apple {within} <u gender="male"/>	<i>red apple</i> within male utterances
thanks {!within} no thanks	<i>thanks</i> not within <i>no thanks</i>
<s/> {containing} yes	Sentences containing <i>yes</i>
(<s/> {!containing} yes) yes	Sentences not containing <i>yes</i> but followed by <i>yes</i>
<s/> {containing} yes {within} <abstract/>	Sentences containing <i>yes</i> within an academic abstract

- » The leftmost operator has precedence unless parentheses are used
- » Y {containing} X or X {within} Y only requires Y to contain all the tokens of X