

4 KWIC tool (key word in context)

The KWIC tool generates a list of all instances of a search term in a corpus in the form of a concordance. It can be used, for example, to:

- Find the frequency of a word or phrase in a corpus.
- Find frequencies of different word classes such as nouns, verbs, adjectives.
- Find complex linguistic structures such as the passives, split infinitives etc. using ‘smart searches’.
- Sort, filter and randomise concordance lines.

4.1 Visual summary of KWIC tab

The screenshot shows the #LancsBox v 3.0 interface with the KWIC tab active. The search term is 'new' with 1,181 occurrences. The concordance table has columns for Index, File, Left, Node, and Right. Callout boxes provide the following instructions:

- Save results (top right button)
- Left-double-click 'Index' to randomise concordance lines. (left side)
- Left-click concordance header to sort. (middle)
- Right-click concordance header to use advanced filter. (right side)
- Left-double-click concordance display to see text. (bottom left)
- Right-click inside to apply filter. (bottom right)
- Pull up the bottom panel. (bottom right)

Simple search

You can:

- Search for a word or phrase.
- Search for number ranges, e.g. >1930&<=1945
- Use * wildcards, e.g. new*
- Use case sensitive regular expressions, e.g. /[abc].*/
- Use case insensitive regular expressions, e.g. /dog|cat/i
- Search for punctuation, e.g. /.*\./p
- Use ‘smart searches’, e.g. PASSIVES, NOUNS

Advanced search

 Headword
 POS

You can:

- Search at different levels of annotation.
- Combine search terms at various levels.
- Use regular expressions, e.g. /N.*/

4.2 Searching and displaying results

#LancsBox supports powerful searching of corpora. The search box can be used for simple as well as advanced searches at different levels of annotation.

1. Simple searches: type in the word or phrase of interest in the search box in the top left-hand corner and left-click 'Search'.
2. Advanced searches: click on the triangle inside the search box (▼) to activate advanced searches at different levels of corpus annotation. You can type search terms as separate constraints into one or more advanced search boxes. For example, the following advanced search is a search for the lemma 'go'.

	Search
go	Headword
V*	POS

Text level empty → no constraint.

Headword is go.

AND

POS is any verbal use.

3. A concordance is generated. The search term, called the 'node', is positioned in the centre and highlighted (orange colour), with words displayed to the left and right of it.
4. KWIC displays basic information about the frequency of the search term and its distribution in texts; the second example shows an application of a filter (see Section 4.4):

Search Term research	Occurrences 158 (1.57)	Texts 13/15
-----------------------------	-------------------------------	--------------------

Read: The search term 'research' occurs 158 times in the corpus with the relative frequency 1.57 per 10k words in 13 out of 15 texts.

Search Term research	Occurrences 9/158 (0.09)	Texts 4/15
-----------------------------	---------------------------------	-------------------

Read: When a filter is applied (indicated by blue colour), the search term 'research' occurs 9 times out of 150 in the corpus with the relative frequency 0.09 per 10k words in 4 out of 15 texts.

4.3 Settings and full text pop-up

KWIC settings include Corpus, Context and Display options. KWIC also allows full-text pop-ups.

1. Corpus: this setting changes the corpus which is being searched. Note that different corpora can be searched in the top and bottom panel in split-screen view.
2. Context: this setting changes the number of words that are displayed in the concordance to the left and to the right of the node.
3. Display: this setting changes the display type. The 'Plain text' default can be changed to 'Text with POS', 'Lemmatized text' and 'All annotation'. The example below demonstrates these four display formats:

Plain text: The new life looks promising for Mr. Noyce.

Text with POS: The_DT new_JJ life_NN looks_VVZ promising_JJ for_IN Mr._NP Noyce._NP

Lemmatized text: the_DT new_JJ life_NN look_VVZ promising_JJ for_IN Mr_NP Noyce_NP

All annotation: [The{the}_DT] [new{new}_JJ] [life{life}_NN] [looks{look}_VVZ] [promising{promising}_JJ]
[for{for}_IN] [Mr.{Mr}_NP] [Noyce.{Noyce}_NP]

4. Full text pop-up: Double left-click on a concordance line to display the entire text with the appropriate line highlighted.

4.4 Sorting, randomising and filtering

KWIC concordance can be sorted alphabetically, randomised and filtered.

1. Alphabetical sorting: Left-click the concordance header (any column) to sort the column alphabetically in the A-Z (ascending) order; click again to re-sort alphabetically in the Z-A (descending) order. The sorting is indicated by arrows: A-Z (▲) and Z-A (▼).
2. Randomising: Left-double-click the header of the 'Index' column to randomise the concordance lines. Randomisation is indicated by the tilde sign (~).
3. Simple filtering: Right-click anywhere inside the concordance to activate the simple filter on that column. Input a word or phrase or a regular expression enclosed in forward slashes (/ /) and click 'Apply'. Filtering is indicated by light blue colour of the filtered text. The filter also updates the results (Occurrences and Texts) in the top display panel (see Section 4.2, point 4).
4. Advanced filtering: Right-click any part of the concordance header to activate the advanced filter. Select an exact column or position for filtering (see below), enter value and click 'Add' and 'Apply'. Filtering is indicated by light blue colour on text and in the results display panel (Occurrences and Texts).

An example of positions for advanced filtering:

L5	L4	L3	L2	L1	Node	R1	R2	R3	R4	R5
is	Mr.	Robert	Weaver	of	New	York.	One	of	his	tasks

► Did you know?

In 1992, when reviewing the state of the art in corpus linguistics, Leech (1992) considers a concordance program “[t]he simplest and the most widely-used tool for corpus-based research” (p. 114). 25 years later, a concordance program such as KWIC still belongs to the essential toolkit of a corpus linguist. The simple and direct access to data that a concordance program facilitates combined with more sophisticated functions such as sorting, filtering and randomising provides a powerful analytical technique.

Leech, G. (1992). Corpora and theories of linguistic performance. In: *Directions in corpus linguistics*, 105-122.