CLiC 1.6 User Guide

1. Introduction

With the 1.6 release of CLiC in the autumn of 2017 the CLiC interface has changed completely.[1] The tabs for the analysis functions ('Concordance', 'Subsets', 'Keywords', 'Clusters') are now located on the side panel on the right (see Figure 1). The sidebar menu can be toggled by clicking on the menu icon ( ) in the top right corner in order to make more space for the analysis screen.
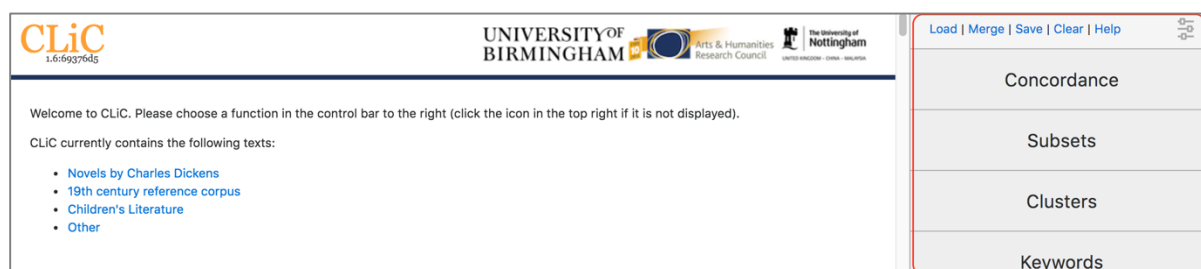

Figure 1: The CLiC home screen with the menu in the side panel on the right.

2. The CLiC corpora

The CLiC homepage, part of which is shown in Figure 1, acts as a table of contents for all the books and their chapters; when clicking onto any of the chapter links you are taken to the full text of that section. As of version 1.6 [released in November 2017], CLiC contains over 130 texts, mainly from the 19th century; for a full list of titles please refer to the appendix. The procedure followed for retrieving, cleaning and importing the most recent texts is described in detail on our Github page.[2] The texts can be selected individually and combined freely for analysis in any of the CLiC tools.[3] You can also choose from one of our three pre-selected corpora: *Novels by Charles Dickens*, *19th Century Reference Corpus* and the *Children's Literature Corpus*. In addition, there is a loosely defined collection of 'Other' texts which include additional GCSE and A-Level titles.

In any of the tools on the sidebar menu you can select any or all of the texts by picking the corpora from a drop-down list or typing their names into a textbox. For example, in the Concordance tool, once you have clicked on the Concordance tab, a textbox labeled 'Search the corpora' will appear (for more information on the Concordance tab, see Section 2.2), as illustrated in Figure 2.

---

[1] We are grateful to colleagues who have commented on previous drafts of this guide: Jamie Lentin, Anna Cermakova and Lorenzo Mastropierro. Yet, given the new interface, this user guide is still a work in progress. Please get in touch via clic@contacts.birmingham.ac.uk if you have further questions or suggestions for improvement.

[2] https://github.com/birmingham-ccr/corpora/blob/master/README.pdf

[3] Please note that we do not recommend selecting all CLiC texts at once, because the tool will not be able to handle this quantity of data and therefore will not show any results. Moreover, any aspect of principled design would be lost when all texts were to be combined; it is better to choose one of our pre-selected corpora or create your own selection.

Figure 2: Selecting corpora in the Concordance tab (same procedure in Subsets and Clusters; for Keywords see Section 2.5).

You can select the pre-selected corpora in their entirety or you can pick individual books from them (again from the drop-down or by typing the title), effectively creating your own subcorpus/subcorpora. For example, you could select several books from Dickens, several books from the *19th Century Novels Reference Corpus* and several books from the *Children's Literature Corpus*. You could also create an author-based corpus, for example by selecting all of Jane Austen's novels by typing "Austen" into the textbox; this will pull out the novels from any of the corpora (see Figure 3).

Figure 3: Example of creating a (temporary) author-based corpus: selecting all of Jane Austen's novels.

The CLiC corpora have been marked up to distinguish between several textual subsets of novels (see the example 1 from *Great Expectations* below).

- **quotes:** any text listed in quotes, i.e. mostly character speech but also thoughts or songs that might appear in quotes
- non-quotes: narration
    - and a special case of non-quotes, **suspensions**, which represent narratorial interruptions of character speech that do not end with sentence-final punctuation. Suspensions are further divided by length:
        - short suspensions have a length up to four words
        - long suspensions have a length of five or more words

CLiC
1.6

UNIVERSITY OF BIRMINGHAM · Arts & Humanities Research Council · The University of Nottingham UNITED KINGDOM · CHINA · MALAYSIA

(1) "And on what evidence, Pip," asked Mr. Jaggers, very coolly, as he paused with his handkerchief half way to his nose,"does Provis make this claim?"
"He does not make it," said I, "and has never made it, and has no knowledge or belief that his daughter is in existence."
For once, the powerful pocket-handkerchief failed. My reply was so unexpected that Mr. Jaggers put the handkerchief back into his pocket without completing the usual performance, folded his arms, and looked with stern attention at me, though with an immovable face.

A more detailed discussion of the subsets in CLiC – and the rationale behind this division – can be found in this open access article by Mahlberg et al. (2016).[4]

3. The CLiC analysis tabs

The homepage shows the table of contents of the books in CLiC. Click onto one of the tabs in the side bar to start your analysis. The CLiC logo will take you back to the homepage if you want to see the content page again at a later point.

The CLiC functions can be divided into two groups:

- The '**concordance**' and '**subsets**' tabs both display text (patterns) from the selected books in context. This is where you can analyse the use of particular words and phrases.
- The '**cluster**' and '**keywords**' tabs both show lists of frequent patterns (without context), but they have different applications. In the keywords tab, you can compare the frequency of words and phrases in one corpus with another; the tool will provide a list of those items that are significantly 'overused' in the first corpus (for more information, see the 'Keywords' section).
-

The CLiC analysis is based on 'white-space tokenisation'. This means that any sequence of letters that is not interrupted by a white space is considered a word. Some special cases arise, however:

- As mentioned in 2.1 below, the 'filter' function in the concordance and subset tabs works differently from the other search functions. The filter does not follow the tokenisation procedure but simply filters for character sequences, i.e. also punctuation. This means that you can filter for round brackets, colons etc. if this is useful for your research.
- Apostrophes: *Oliver* and *Oliver's* count as the same type in CLiC. Therefore, when you search for *Oliver* in *Oliver Twist* you will retrieve all instances of *Oliver* and *Oliver's* (and vice versa; 826 results). Note, however, that the filter searches by word form: So if you want to find only *Oliver's*, for example, you can filter the 826 rows and retrieve 100 entries of *Oliver's*.
- Two hyphens separate words: for example, *Char--lotte* in Oliver Twist (OT.c6.p20) "Oliver's gone mad! Char--lotte!" counts as two tokens.

---

[4] Mahlberg, M., Stockwell, P., de Joode, J., Smith, C., & O'Donnell, M. B. (2016). CLiC Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics. Corpora, 11(3), 433–463 [Open access, available from https://doi.org/10.3366/cor.2016.0102]

3.1. Functions common to all tabs

- At any point, you can close the menu on the right by clicking on the settings icon in the top right corner (see Figure 4).
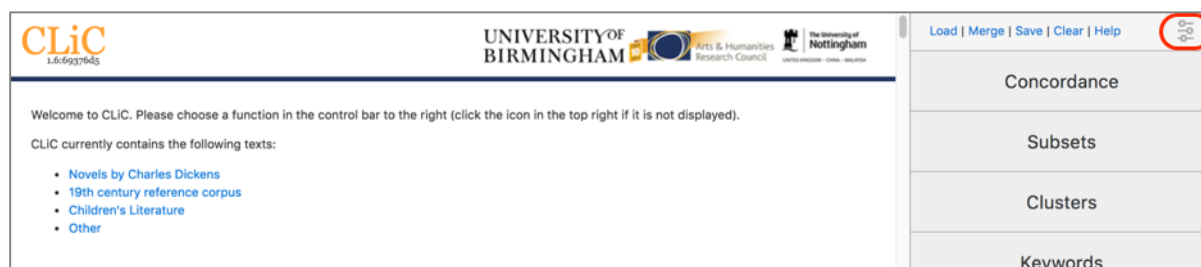

Figure 4: Close the sidebar menu by clicking on the menu icon in the top right corner.

- The buttons in the top row apply to all analysis tabs:
  - o **'Load'**: you can upload a previously exported CLiC csv file to restore your settings and your tag annotation (see the 'tagging' section below). Note that the 'Load' function will replace any tags on the website with those from the file.
  - o **'Merge'**: The 'Merge' function will add the tags from the CSV file to any pre-existing tags. You can also use this function when you have more than one CSV (for example with annotations from several researchers) so you can merge these in order to check to what extent the tag sets overlap or differ.
  - o **'Save'**: save your results, settings and annotated tags in a CSV file.
  - o **'Clear'** will reset the search settings and any tag columns.

For both 'Load' and 'Save' the results/queries have to be compatible, i.e. they have to be based on the same node word.

3.2. Concordance

Clicking onto the 'Concordance' tab will take you to the concordance view. In order to create a concordance, you will need to select a corpus to search in (see the section on 'The Corpora' above).

- **Concordance – 'Search the corpora'**

This is where you select a corpus to search in (cf. Section 2). The selection is very flexible and lets you pick a pre-defined corpus (*19th Century Novels Reference Corpus*, *Dickens's Novels* or *Children's Literature*) or choose your own subcorpus – either from books from only one of these corpora or combining books across the pre-defined corpora.

- **Concordance – 'Only in subsets'**

Here you can decide whether you want to search through 'all text' – the whole book(s) – or just one of the subsets: 'short suspensions', 'long suspensions', 'quotes' and 'non-quotes' (cf. Section 2).

- **Concordance – 'Search for terms'**

This is the fundamental parameter of the concordance search – it lets you determine the node word or phrase that forms the basis of the concordance. When you type your search word(s), keep in mind the notes from the tokenisation section above. The node has to be a

valid token according to the white-space tokenisation: for example, a search for a punctuation sign will not retrieve any results. If your research focuses on punctuation markers you can evade this issue by using the filter function in the subset tab: Go to the subset tab, selecting the relevant subset, for example non-quotes, and <u>filtering</u> the rows to the punctuation marker of interest.

- **Concordance – 'Whole phrase' or 'Any word'**

When you have entered several terms, you need to specify whether it is to be searched as one phrase (equivalent to using double quotes in a search engine, e.g. *dense fog*) or any of the words individually (*dense* and *fog*).

- **Concordance – Co-text**

The maximum number of words in the co-text is set at 10 on either side in a concordance (depending on the length of the words and the size of the screen you might see fewer). You can see the full chapter view by clicking on "in bk." (in book) button at the end of any row.

Figure 5: The "in bk." (in book) button leads to the chapter view of the occurrence.

Figure 6: The "in bk." view shows the whole chapter – in the case of this preface it is a very short chapter. (Note that all authorial text occurring before the official first chapter, is counted as 'chapter 0' in CLiC).

- **Concordance – 'Results'**

These options allow you to adjust the way the concordance output is displayed.

  - **Concordance – Results – 'Filter rows'**

This filter option lets you filter the concordance output by the rows that contain a particular sequence of letters (apart from the node). For example, searching for *hands* in *Oliver Twist* yields 124 results; when we use the option 'filter rows' and search for *pockets*, this is filtered down to 8 results as illustrated in Figure 7.

Figure 7: Concordance of *hands* in *Oliver Twist* filtered down to *pockets* in the co-text.

Note that the filter, when searching for character sequences does not necessarily search for complete words: for example, filtering a concordance of *head* in *Oliver Twist* for *eat* yields both occurrences of the verb *eat*, and the instance *threatened*, which contains the same sequence of letters (see Figure 8). The filter function is cruder than the KWICGrouper; it can be usefully applied to filter down a large set of results before you do a more fine-grained categorisation. You might want to filter down the results to rows containing similar word forms. For example, filtering for *girl* will also retrieve rows containing *girlish* and *girls*. Moreover, unlike the main concordance search and the KWICGrouper, the filter lets you search for particular types of punctuation (e.g. round brackets used in suspensions).

Figure 8: Filtering for the letter sequence *eat* returns forms of the verb eat and other words containing the sequence.

- o **Concordance – Results – 'Show metadata columns'**

Show the chapter, paragraph and sentence number for each row. (Used to be "Toggle Metadata" until CLiC 1.5). This illustrates where in the book you are and can be the basis for sorting (see section on sorting below).

- **Concordance – Basic sorting**

The concordance lines can be sorted by any of the columns in the concordance by clicking on the header, which will then be marked with dark arrows. For example, by clicking on "Left" the lines will be sorted by the first word to the left of the node and by clicking on "Right" by the first word on the right. If you have the metadata columns activated you can also sort by these, for example to sort all entries by chapter. Similarly, if you have created your own

CLiC 1.6

UNIVERSITY OF BIRMINGHAM · Arts & Humanities Research Council · The University of Nottingham · UNITED KINGDOM · CHINA · MALAYSIA

tags (see 'tagging' section below), you can sort for lines with a particular tag. Clicking on the same header a second time will reverse the order of sorting.

Note that you can create a 'sorting sequence' by clicking on various headers while pressing the shift key. For example, you could sort a concordance first by the words on the right and then by book, as illustrated in Figure 9, which shows a concordance of *bank* first by book – so that results from *Bleak House* (BH) come first – and then ordered

Figure 9: Concordance of *bank* in novels by Charles Dickens – fist ordered by book, then by the first word on the right.

- **Concordance – 'KWICGrouper'**

The KWICGrouper is a tool that allows you to quickly group the concordance lines according to patterns that you find as you go through the concordance. For a basic introduction to the KWICGrouper functionality (in the CLiC 1.5 interface) you can watch our KWICGrouper video tutorial from May 2017.[5]

The idea of the KWICGrouper is that you look for patterns as you search for particular words. Any matching lines will be highlighted and moved to the top of the screen. Among the matching lines we further distinguish between the lines based on how many matches they contain. A line with one match is highlighted in light green, lines with two matches are coloured in a darker green, those with three in purple and, finally, those with four in pink. (For lines with more matches than these, the colours with repeat.) The KWICGrouper gives you two options:

1. 'Search in span': Set the span for the KWICGrouper search. By dragging the slider you can adjust the number of words that will be searched to the left and right of the search term. The maximum (and default) span is 5 positions to either side.

---

[5]   The video is available from our CLiC Dickens blog at https://blog.bham.ac.uk/clic-dickens/2017/06/22/video-introducing-the-clic-kwicgrouper-function-to-group-concordance-lines/

2. 'Search for types': Choose one or more words to search for in the span. This is currently limited to single words, but there is no limit on how many words you add.

The total number of matching rows will be displayed at the top; the process is illustrated in Figures 10–15. Figure 10 shows the plain concordance lines as returned when searching for *fire* in Dickens's novels.

Figure 10: The first concordance lines of *fire* in Dickens's novels.

Figure 11 illustrates the process of choosing types (forms of words) from co-text surrounding *fire* in the concordance in order to group the concordance lines. The dropdown only contains those word forms that actually appear around the node term in the specified search span. Therefore, while *sitiwation* is listed here, it wouldn't be listed if we had searched for another node term or used other books; it only appears once in this set in the following context (example 2):

(2) I don't take no pride out on it, Sammy,' replied Mr. Weller, poking the fire vehemently, 'it's a horrid **sitiwation**. I'm activally drove out o' house and home by it.The breath was scarcely out o' your poor mother-in-law's body, ven vun old 'ooman sends me a pot o' jam, and another a pot o' jelly, and another brews a blessed large jug o' camomile-tea, vich she brings in vith her own hands.' [*Pickwick Papers*, Chapter LI.]

Figure 11: Selecting types related to sitting from the KWICGrouper to group the concordance lines.

Figure 12: The resulting 'KWICGrouped' concordance lines: the selected types are listed in the search box on the right; and in the case of this example it is suitable to restrict the search span to only the left side of the node.

The KWICGrouper only searches through a number of words to the left and right of the node term, as specified by the search span. Figure 12 shows the resulting concordance lines according to the KWICGrouper settings on the right. Apart from the selected search types the search span has also been restricted to the left side so that clearer patterns of sitting by the fire become visible.

Figure 13: The first lines of *fire* co-occurring with *back* (i.e. one KWICGrouper match) are highlighted and moved to the top.

Apart from looking for characters sitting by the fire, it might also be of interest to look for characters standing by the fire. We have shown in our previous work (chapter 6 of Mahlberg 2013[6]) that the cluster *with his back to the fire* is prominent in Dickens's and 19th century novels by other writers. Figure 13 shows the first concordance lines of *fire* with *back* on the left.

The output from the KWICGrouper lists at the top of the screen the number of lines that contain any number of matches. In the case of Figures 12 and 13 there are only lines with one match, but no lines with more than one match. So, in Figure 13, the message says "36 entries with 1 KWIC match", this means that 36 lines contain both *fire* and *back*. This function becomes useful when we now look for gendered pronouns. As shown in Figure 14, there are 27 lines in which *fire* co-occurs with both *back* and *his* (mostly of the pattern *with his back to the fire*). On the other hand, as we can see from Figure 15 that Dickens's novels contain only instance of fire co-occurring with *back* and *her* (*with her back to the fire*).

---

[6] Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction.* London: Routledge.

Figure 14: The 27 lines with two matches (here, *back* and *his*) are highlighted in a darker green.

Figure 15: Only one line contains both *back* and *her*; it is highlighted and shown above single match lines.

- **Concordance – 'Manage tag columns'**

Once you have identified lines with patterns of interest, you might want to place these into one or more categories. CLiC provides a flexible tagging system for this. Figure 16 illustrates the outcome of what a tagged concordance can look like. The tags are user-customised so you can create tags that are relevant to your project. In this case, occurrences of *dream* in *Oliver Twist* have been tagged according to who is dreaming.

Figure 16: Tagged concordance lines of *dream* in *Oliver Twist*.

In order to tag the lines, click on "manage tag columns" (shown in the bottom right corner of Figure 16) and create your own tag(s) through the "Add new" option (see Figure 17). You can rename a tag by selecting it from the "Tag columns" list and renaming it in the text box. Once you have created your tag(s), you can click "back" to return to the menu. Now you can select the relevant concordance lines by clicking on them and you will see that the sidebar contains the list of your tags. Once your lines are selected you can click the tick next to the relevant tag in order to tag the line (see Figure 18). An extra column will appear for each tag and you can sort on these columns as mentioned in the sorting section above. Selected and tagged rows will be automatically deselected when you click on (i.e. select) a new row.

Figure 17: The menu for adding and renaming tags.

Figure 18: Select a line (by clicking on it) in order to apply an existing tag; once tagged, the tick will appear green for the selected line.

3.3. 'Subsets'

The subsets option can display the full subset of your choice for the selected corpus; click onto the dropdown to select a relevant subset (short suspensions, long suspensions, quotes, non-quotes). For example, you could display all quotes in Dickens's novels for further analysis. You can then use the filter option to narrow down the lines and further search and group the lines using the KWICGrouper. As in the concordance tab, you can create tags to annotate rows with noteworthy patterns.

- **Subsets – Results**

Like in the concordance tab, this allows you to adjust the way the concordance output ('table') is displayed.

- o **Subsets – Results – Filter rows**

This filter option lets you filter the concordance output by the rows that contain a particular sequence of letters, as described in Section 3.2 on the filter function in the concordance tab.

Figure 19: Filtering the co-text of long suspensions for *perhaps* in *Oliver Twist*.

CLiC
1.6

UNIVERSITY OF BIRMINGHAM   Arts & Humanities Research Council   The University of Nottingham   UNITED KINGDOM · CHINA · MALAYSIA

Note that the filter searches through the whole row and therefore also accounts for words in the context, not only in the subset itself. For example, when searching through the subset of long suspensions in *Oliver Twist* and filtering rows for *perhaps* the results originate only from the co-text, as *perhaps* does not occur in long suspensions (see Figure 19).

3.4. Clusters

The output of the cluster tool generates frequency lists of single words and clusters (or 'n-grams', where 'n' stands for the length of the phrase). If we choose a '1-gram' (single word), we retrieve a simple word list. (In *Oliver Twist*, for example, the top 10 words retrieved from this tool are *the*, *and*, *to*, *of*, *a*, *he*, *in*, *his*, *that* – all function words, as we would generally expect.) A cluster is a repeated sequence of words. CLiC currently supports 1-grams (single words), 3-grams (like *i don't know*) and so on up to 5-grams (*what do you mean by*).

Figure 20: Cluster options.

As in the other tabs, you can restrict the search to a particular subset so that, for example, you can create frequency lists for clusters in quotes (or the other subsets). You can save the resulting list as a CSV file (for example for use in a spreadsheet viewer) by clicking the 'save' button at the top. Note that the CLiC 'Cluster' tab will display words and clusters with a minimum frequency of 5.

3.5. Keywords

The keywords tool finds words (and phrases) that are used significantly more often in one corpus compared to another. CLiC incorporates the keyword extraction formula reported by Rayson and Garside (2000)[7]. Apart from comparing single words, CLiC also allows you to compare clusters. Whereas the cluster tab focuses only on one corpus, the Keywords function can compare cluster lists. You have to make selections for the following options (also see Figure 21):

---

[7] Rayson, P. and Garside, R. (2000). Comparing corpora using frequency profiling. In proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000). 1-8 October 2000, Hong Kong, pp. 1-6, retrieved from http://ucrel.lancs.ac.uk/people/paul/publications/rg_acl2000.pdf

- **'Target corpora':** Choose the corpus/corpora that you are interested in.
- **'within subset':** Specify which subset you are comparing (or simply choose 'all text')/
- **'Reference corpora':** Choose the reference corpus to compare your target corpus to.
- **'within subset':** Specify the subset.
- **'n-gram':** Do you want to compare single words (1-grams) or phrases (2-grams up to 5-grams)

Figure 21: The settings for the keywords tab. Here you need to select two sets of corpora for the keyword comparison – target and reference – and their corresponding subsets.

Figure 22: Key 5-word clusters in Oliver Twist quotes compared to the 19th Century Reference Corpus.

Note that you have to select a subset for each of the two corpora or you'll see the error message: "Please select a subset". So, for example, when comparing 5-grams in *Oliver Twist* (quotes) against the 19th century novels (quotes), we retrieve the results displayed in Figure 22 (for a p-value of 0.0001).

The frequency threshold of 5 used for the cluster tab is <u>not</u> applied to the keyword tab, so that all frequencies are compared. The keyword output shows the top 3000 results (for most comparisons, you will yield fewer results, though). Moreover, CLiC only generates so-called 'positive keywords': those that are 'overused' in the target corpus than in the reference corpus, but CLiC does not generate 'negative' or 'underused' keywords.

4. Appendix: List of texts available in CLiC

> **Pre-selected corpora in CLiC**
> DNov   Novels by Charles Dickens
> ChiLit   Children's Literature
> 19C   19th century reference corpus
> Other   A collection of 'other' texts with more set titles from A-Level and GCSE exam specifications.

Table 1: Titles in the DNov, 19C and Other pre-selected corpora.

Title	Author	Corpus
Emma	Jane Austen	19C
Persuasion	Jane Austen	19C
Pride and Prejudice	Jane Austen	19C
Sense and Sensibility	Jane Austen	Other
Mansfield Park	Jane Austen	Other
Northanger Abbey	Jane Austen	Other
Lady Susan	Jane Austen	Other
Lady Audley's Secret	Mary Elizabeth Braddon	
Agnes Grey	Anne Brontë	
The Tenant of Wildfell Hall	Anne Brontë	Other
Jane Eyre	Charlotte Brontë	
The Professor	Charlotte Brontë	
Wuthering Heights	Emily Brontë	
The Last Days of Pompeii	Edward George Bulwer-Lytton	
The Awakening[a]	Kate Chopin	Other
Antonina or, the Fall of Rome	Wilkie Collins	
Armadale	Wilkie Collins	
The Woman in White	Wilkie Collins	
The Moonstone	Wilkie Collins	Other
Heart of Darkness[b]	Joseph Conrad	Other
A Christmas Carol in Prose; Being a Ghost Story of Christmas	Charles Dickens	Other
A Tale of Two Cities	Charles Dickens	DNov
Barnaby Rudge	Charles Dickens	DNov
Bleak House	Charles Dickens	DNov
David Copperfield	Charles Dickens	DNov
Dombey and Son	Charles Dickens	DNov
Great Expectations	Charles Dickens	DNov
Hard Times	Charles Dickens	DNov
Little Dorrit	Charles Dickens	DNov
Martin Chuzzlewit	Charles Dickens	DNov
Nicholas Nickleby	Charles Dickens	DNov
Oliver Twist	Charles Dickens	DNov
Our Mutual Friend	Charles Dickens	DNov
Pickwick Papers	Charles Dickens	DNov
The Mystery of Edwin Drood	Charles Dickens	DNov

CLiC
1.6

UNIVERSITY OF BIRMINGHAM 10 · Arts & Humanities Research Council · The University of Nottingham · UNITED KINGDOM · CHINA · MALAYSIA

The Old Curiosity Shop	Charles Dickens	DNov
Sybil, or the two nations	Benjamin Disraeli	19C
Vivian Grey	Benjamin Disraeli	19C
The Hound of the Baskervilles	Arthur Conan Doyle	19C
The Sign of the Four	Arthur Conan Doyle	Other
Daniel Deronda	George Eliot	19C
The Mill on the Floss	George Eliot	19C
Silas Marner	George Eliot	Other
Middlemarch	George Eliot	Other
A Room with a View	E. M.Forster	Other
Cranford	Elizabeth Gaskell	19C
Mary Barton	Elizabeth Gaskell	19C
North and South	Elizabeth Gaskell	19C
The Yellow Wallpaper	Charlotte Perkins Gilman	Other
Jude the Obscure	Thomas Hardy	19C
The Return of the Native	Thomas Hardy	19C
Tess of the D'Urbervilles	Thomas Hardy	19C
The Portrait of a Lady – Volume 1[b]	Henry James	Other
The Portrait of a Lady – Volume 2[b]	Henry James	Other
What Maisie Knew	Henry James	Other
Women in Love	D. H. Lawrence	Other
Twelve Years a Slave	Solomon Northup	Other
Frankenstein	Mary Shelley	19C
The Strange Case of Dr Jekyll and Mr Hide	Robert Louis Stevenson	19C
Dracula	Bram Stoker	19C
Vanity Fair	William Makepeace Thackeray	19C
The Small House at Allington	Anthony Trollope	19C
Adventures of Huckleberry Finn	Mark Twain	Other
The War of the Worlds	H. G. Wells	Other
The Return of the Soldier	Rebecca West	Other
The Picture of Dorian Gray	Oscar Wilde	19C

Notes:
[a] The Gutenberg source that we used for The Awakening (see https://github.com/birmingham-ccr/corpora/blob/master/README.pdf) also contained short stories by Chopin. These have been removed before the main text was imported into CLiC.
[b] These titles will be added in mid-November.

Table 2: Titles in ChiLit; a corpus compiled by Anna Cermakova for the GLARE project[8].

Title	Author
Vice Versa; or, A Lesson to Fathers	F. Anstey
The Brass Bottle	F. Anstey
The Coral Island: A Tale of the Pacic Ocean	R. M. Ballantyne
Peter Pan	J. M. Barrie
The Secret Garden	Frances Hodgson Burnett
Alice's Adventures in Wonderland	Lewis Carroll
Through the Looking-Glass	Lewis Carroll
The Surprising Adventures of Sir Toady Lion with Those of General Napoleon Smith	S.R. Crockett

---

[8] See http://birmingham.ac.uk/glare for more information on the GLARE project.

Mrs. Overtheway's Remembrances	Juliana Horatia Ewing
Jackanapes	Juliana Horatia Ewing
Moonfleet	J. Meade Falkner
Eric; Or, Little by Little	F. W. Farrar
Adventures in Wallypug-Land	G. E. Farrow
The Golden Age	Kenneth Grahame
Dream Days	Kenneth Grahame
The Wind in the Willows	Kenneth Grahame
King Solomon's Mines	H. Rider Haggard
Allan Quatermain	H. Rider Haggard
Winning His Spurs. A Tale of the Crusades	G. A. Henty
With Clive in India; Or, The Beginnings of an Empire	G. A. Henty
Tom Brown's Schooldays	Thomas Hughes
Mopsa the Fairy	Jean Ingelow
Wood Magic, a Fable	Richard Jefferies
Madam How and Lady Why; Or, First Lessons in Earth Lore for Children	Charles Kingsley
The Water-Babies	Charles Kingsley
Stalky and Co	Rudyard Kipling
The Jungle Book	Rudyard Kipling
Prince Prigio	Andrew Lang
At the Back of the North Wind	George McDonald
The Princess and the Goblin	George McDonald
The Three Mulla-mulgars	Walter de la Mare
Masterman Ready; Or, The Wreck of the Pacific	Frederick Marryat
The Settlers in Canada	Frederick Marryat
The Children of the New Forest	Frederick Marryat
Feats on the Fiord	Harriet Martineau
The Crofton Boys	Harriet Martineau
The Peasant and the Prince	Harriet Martineau
The Settlers at Home	Harriet Martineau
A World of Girls: The Story of a School	L. T. Meade
The Cuckoo Clock	Mary Louisa Molesworth
The Tapestry Room: A Child's Romance	Mary Louisa Molesworth
The Carved Lions	Mary Louisa Molesworth
The Book of Dragons	E. Nesbit
The Story of the Treasure Seekers	E. Nesbit
Nine Unlikely Tales	E. Nesbit
The Railway Children	E. Nesbit
Five Children and It	E. Nesbit
The Story of the Amulet	E. Nesbit
The Tale of Peter Rabbit	Beatrix Potter
The Tale of Squirrel Nutkin	Beatrix Potter
The Tale of Benjamin Bunny	Beatrix Potter
The Tale of Two Bad Mice	Beatrix Potter
The Tale of Jemima Puddle-Duck	Beatrix Potter
The Tale of the Flopsy Bunnies	Beatrix Potter
The Fifth Form at Saint Dominic's: A School Story	Talbot Baines Reed
The King of the Golden River; or, the Black Brothers: A Legend of Stiria	John Ruskin
Black Beauty	Anna Sewell
Holiday House: A Series of Tales	Catherine Sinclair
Kidnapped	Robert Louis Stevenson
Treasure Island	Robert Louis Stevenson

Jessica's First Prayer	Hesba Stretton
Little Meg's children	Hesba Stretton
Alone in London	Hesba Stretton
The Rival Crusoes; or The Shipwreck	Agnes Strickland
The Rose and the Ring	William Makepeace Thackeray
Leila at home	Ann Fraser Tytler
The Happy Prince, and Other Tales	Oscar Wilde
The Heir of Redclye	Charlotte M. Yonge
The Daisy Chain, or Aspirations	Charlotte M. Yonge
The Dove in the Eagle's Nest	Charlotte M. Yonge
The Little Duke: Richard the Fearless	Charlotte M. Yonge