

CLiC 1.6.1 User Guide

1. Introduction	1
2. Changes from CLiC 1.5 to CLiC 1.6	1
3. System requirements	1
4. The CLiC corpora	1
5. The CLiC analysis tabs	3
5.1. <i>Functions common to all tabs</i>	4
• Saving plain and annotated results	4
• Printing the results	5
5.2. <i>Concordance</i>	6
• Concordance – Search the corpora	6
• Concordance – Only in subsets	6
• Concordance – Search for terms	6
• Concordance – ‘Whole phrase’ or ‘Any word’	6
• Concordance – Co-text	6
• Concordance – Results	7
• Concordance – Basic sorting	8
• Concordance – KWICGrouper	9
• Concordance – Manage tag columns	13
5.3. <i>Subsets</i>	14
• Subsets – Show subsets	14
• Subsets – Results	15
• Subsets – KWICGrouper	16
• Subsets – Manage tag columns	17
5.4. <i>Clusters</i>	17
5.5. <i>Keywords</i>	18
Appendix 1: List of texts available in CLiC	20
Appendix 2: CLiC texts listed in A-Level and GCSE specifications	24

1. Introduction

The CLiC web app¹ has been developed as part of the CLiC Dickens project², which demonstrates through corpus stylistics how computer-assisted methods can be used to study literary texts and lead to new insights into how readers perceive fictional characters. CLiC Dickens started at the University of Nottingham in 2013; it is now a collaborative project with the University of Birmingham.

The CLiC Dickens project is funded by the Arts and Humanities Research Council, grant reference AH/P504634/1.

Project team: Prof. Michaela Mahlberg, Prof. Peter Stockwell, Viola Wiegand

2. Changes from CLiC 1.5 to CLiC 1.6

There have been several changes to the back-end; for further technical details, please refer to our GitHub repository.³ The focus of this user guide is the CLiC interface,⁴ which has changed substantially with version 1.6. 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

3. System requirements

The CLiC web app runs in your web browser (i.e. you will need an internet connection). We recommend recent versions of Chrome, Firefox or Safari. If the CLiC homepage displays a red error message instead of the table of contents, please check whether your browser needs updating. CLiC should also work on most (Apple and Android) mobile devices.

4. 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 1.6.1, released in November 2017, CLiC contains 138 texts, mainly from the 19th century; for a full list of titles please refer to Appendix 1. The procedure followed for retrieving, cleaning and importing the most recent texts is described

¹ <http://clic.bham.ac.uk>

² <http://birmingham.ac.uk/clic>

³ <https://github.com/birmingham-ccr/clic>

⁴ We are grateful to colleagues who have commented on previous drafts of this guide: Jamie Lentin, Anna Cermakova and Lorenzo Mastropiero. 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.

in detail on our GitHub repository.⁵ The texts can be selected individually and combined freely for analysis in any of the CLiC tools.⁶ You can also choose from one of our four pre-selected corpora: DNov – Dickens’s Novels (15 texts), 19C – 19th Century Reference Corpus (29 texts), ChiLit – 19th Century Children’s Literature Corpus (71 texts) and ArTs – Additional Requested Texts (23 texts). ArTs includes additional GCSE and A-Level titles (please see Appendix 2 for an overview of all CLiC texts listed in the AQA, OCR and Edexcel GCSE and A-Level English specifications).

In order to **select texts** in any of the CLiC analysis tabs, go to control bar on the right-hand side (circled in Figure 1 above). 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 5.2), as illustrated in Figure 2.

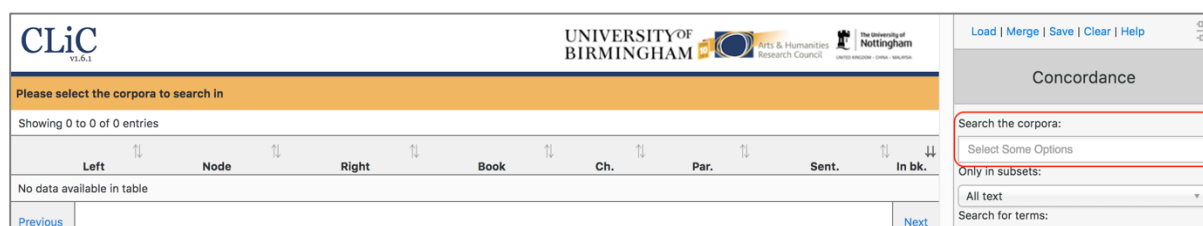


Figure 2: Selecting corpora in the Concordance tab (same procedure in Subsets and Clusters; for Keywords see Section 5.5)

You can select the pre-selected corpora in their entirety or you can pick individual books from them, effectively creating your own subcorpus. For example, you could select several books from Dickens, several books from the 19th Century Reference Corpus (19C) and several books from the 19th Century Children’s Literature Corpus (ChiLit). You could also create an author-based corpus, for example by selecting all of Jane Austen’s novels by typing *austen* into the textbox (which is not case-sensitive); this will pull out the relevant novels from any of the corpora for individual selection (see Figure 3).

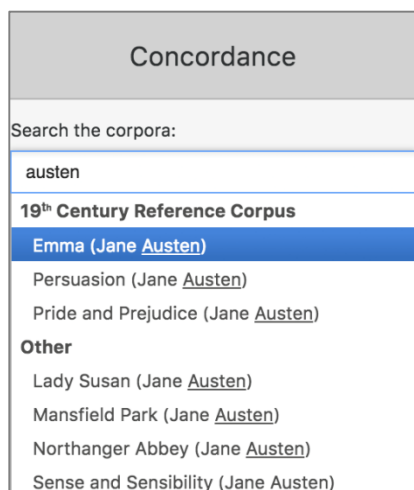


Figure 3: Example of creating a (temporary) author-based corpus: selecting all of Jane Austen’s novels

⁵ <https://github.com/birmingham-ccr/corpora/blob/master/README.pdf>

⁶ 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 for performance intensive queries. Moreover, any aspect of principled design would be lost when all texts were to be combined; we therefore advise choosing one of our pre-selected corpora or creating your own selection.

The CLiC corpora have been marked up to distinguish between several textual subsets of novels. Example 1 from *Great Expectations* below illustrates the subsets and Figure 4 shows how these are marked up in the chapter views, which can be retrieved from the 'in bk.' (in book) button in concordances (see Section 5.2) and the table of contents on the homepage.

- (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. [*Great Expectations*, Chapter 51]

- **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

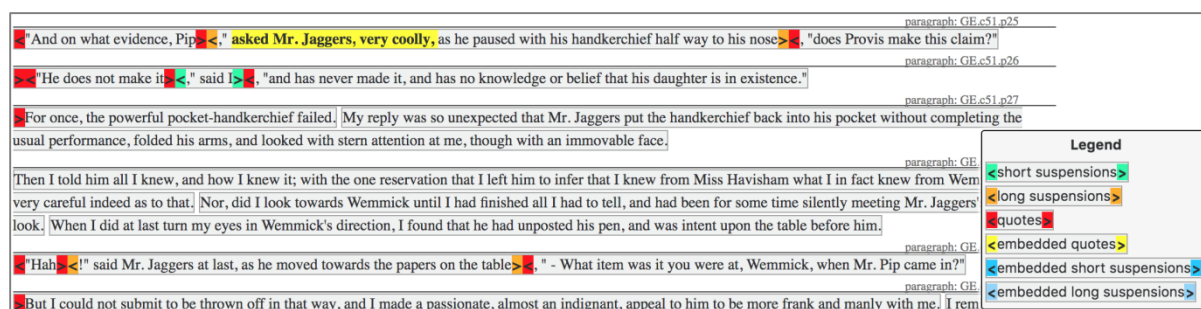


Figure 4: Chapter view of example (1) (retrieved via the 'in bk.' (in book) button in a concordance of *asked Mr Jaggers very coolly*), exemplifying the mark-up of subsets

A more detailed discussion of the subsets in CLiC – and the rationale behind this division – can be found in the open access article by Mahlberg et al. (2016).⁷

5. 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:

⁷ 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>]

- A. 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.
- B. The **‘Cluster’** and **‘Keywords’** tabs both show lists of frequent patterns (without context), but they differ in their applications. The Cluster tab lists frequent words and phrases in a single corpus. 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 Section 5.5 on keywords).

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.

5.1. Functions common to all tabs

At any point, you can close the menu on the right by clicking on the menu icon in the top right corner (; also see Figure 5).



Figure 5: Close the sidebar menu by clicking on the menu icon in the top right corner

- Saving plain and annotated results**

The buttons in the top row apply to all analysis tabs:

- **‘Load’**: You can upload a previously exported CLiC CSV file to restore your settings and your tag annotation (see the ‘tagging’ section below). The CSV file can only be reimported to CLiC if you haven’t made any changes to it. We would therefore recommend keeping the original copy for potentially loading it back into CLiC (as well as for your personal record) and saving any manual changes (e.g. comments, highlights, filtered lines) in a separate version. Also note that the ‘Load’ function will replace any existing tags in CLiC with those from the file: unlike ‘Merge’. see below.

- **‘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.
- **‘Save’**: Save your results, settings and annotated tags in a CSV file that can be opened in a spreadsheet viewer. The file contains a (shareable) link that will replicate your search settings.
- **‘Clear’**: Resets the search settings and any tag columns. (Identical to clicking the CLiC icon.)

For both **‘Load’** and **‘Merge’** the results/queries have to be compatible, i.e. they have to be based on the same node word.

- **Printing the results**

If you have already saved the results in a CSV file, you may want to print that file directly from your spreadsheet viewer. However, the CSV file will not preserve any of the colours or the highlighting that you may have created with the KWICGrouper function (see the subsection ‘Concordance – KWICGrouper’ in 5.2 below). In order to print the output in colour, go to the Chrome printing menu, click on ‘More settings’ and tick ‘Background graphics’ (see Figure 6; other browsers should have similar settings). The layout also tends to print best in Landscape. You can then “print” the output to a PDF file (as in Figure 6) or straight to your printer.

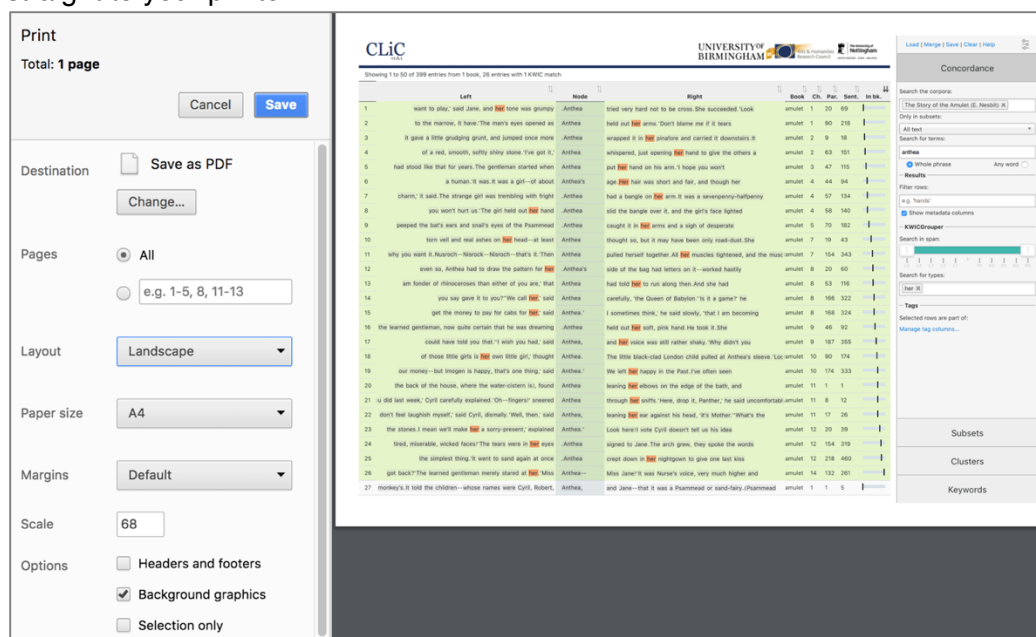


Figure 6: Settings for printing CLiC output in colour using the Chrome print menu

5.2. 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 4 on ‘The CLiC Corpora’ above).

- **Concordance – Search the corpora**

This is where you select a corpus to search in (cf. Section 4). 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 4).

- **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, select the relevant subset, for example non-quotes, and filter 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 (see Figure 7).

The screenshot shows the CLiC Concordance interface. At the top, there are logos for CLiC v1.6.1, the University of Birmingham, Arts & Humanities Research Council, and The University of Nottingham. Below the logos, it says "Showing 1 to 50 of 92 entries from 1 book".

	Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
1	HAVE endeavoured in this Ghostly little book, to raise the	Ghost	of an idea, which shall not put my readers out	carol	0	1	1	in bk.
2	leaped up, as though it cried, "I know him; Marley's	Ghost!"	and fell again. The same face: the very same. Marley	carol	1	96	252	in bk.
3	ge asked the question, because he didn't know whether a	ghost	so transparent might find himself in a condition to take	carol	1	109	273	in bk.
4	olve the necessity of an embarrassing explanation. But the	ghost	sat down on the opposite side of the fireplace, as	carol	1	109	274	in bk.
5	used to it. "You don't believe in me," observed the	Ghost."	I don't," said Scrooge. "What evidence would you have of	carol	1	110	275	in bk.

On the right side, there is a sidebar with the following controls:

- Buttons: Load | Merge | Save | Clear | Help
- Section: Concordance
- Search the corpora: A Christmas Carol: A Ghost Story of Christmas (Charles Dickens)
- Only in subsets: All text
- Search for terms: ghost

Figure 7: The ‘in bk.’ (in book) button leads to the chapter view of the occurrence

A Christmas Carol: A Ghost Story of Christmas, by Charles Dickens

PREFACE.

I HAVE endeavoured in this Ghostly little book, to raise the **Ghost** of an Idea, which shall not put my readers out of humour with themselves, with each other, with the season, or with me; May it haunt their houses pleasantly, and no one wish to lay it.

Their faithful Friend and Servant, C. D. December, 1843.

Figure 8: 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). This preface contains no quotes or suspensions; compare to the subset markup in the chapter view of Figure 4

- **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 (both in the node and co-text). 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 9.

Showing 1 to 8 of 8 entries (filtered from 124 total entries) from 1 book

	Left	Node	Right	In bk.
1	king in the dim light, like high-shouldered ghosts with their	hands	in their breeches pockets. Coffin-plates, elm-chips, bright-h-	
2	him; and another was so obliging as to put his	hands	in his pockets, in order that, as he was very	
3	at cocked, as usual; Master Bates sauntering along with his	hands	in his pockets; and Oliver between them, wondering where	
4	said the doctor; 'let me think.' Mr. Losberne thrust his	hands	into his pockets, and took several turns up and down	
5	vent to this result of experience, the doctor put his	hands	into his pockets, and walked up and down the room	
6	d asserted his superiority in a becoming manner, thrust his	hands	into his pockets, and sauntered towards the door, with muc	
7	cut short by the entry of Master Bates, with his	hands	in his breeches-pockets, and his face twisted into a look	
8	Bumble fixed his hat on very tight, and putting his	hands	in his pockets, followed his helpmate downstairs. 'Young lad,	

Search the corpora:
Oliver Twist (Charles Dickens) x

Only in subsets:
All text

Search for terms:
hands

Whole phrase Any word

Results

Filter rows:
pockets

Show metadata columns

Figure 9: 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 10). 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).

CLiC v1.6.1

UNIVERSITY OF BIRMINGHAM Arts & Humanities Research Council The University of Nottingham

Load | Merge | Save | Clear | Help

Concordance

Search the corpora:
Oliver Twist (Charles Dickens) x

Only in subsets:
All text

Search for terms:
head

Whole phrase Any word

Results

Filter rows:
eat

Show metadata columns

KWICGrouper

Search in span:

	Left	Node	Right	In bk.
1	eyes; and, running after him, bestowed a blow on his	head,	which would inevitably have beaten in any skull but a	
2	eyes. His hat was stuck on the top of his	head	so lightly, that it threatened to fall off every moment	
3	boy's face. There was its living copy. The eyes, the	head,	the mouth; every feature was the same. The expression wa	
4	as to say: looking at him with great complacency from	head	to foot, that she really didn't think it would have	
5	my death, or I'll be content to eat my own	head,	sir! This was the handsome offer with which Mr. Grimwig	
6	pass which will enable a gentleman to eat his own	head	in the event of his being so disposed, Mr. Grimwig's	
7	question, a very thick coating of powder. I'll eat my	head,	sir, repeated Mr. Grimwig, striking his stick upon the grou	
8	this bit of peel upon the staircase, I'll eat my	head,	and his too. "No, no, he has not had one	
9	that boy returns to this house, sir, I'll eat my	head.'	With these words he drew his chair closer to the	
10	exclamation: 'A beadle. A parish beadle, or I'll eat my	head."	Pray don't interrupt just now,' said Mr. Brownlow. 'Take a	

Figure 10: Filtering for the letter sequence *eat* returns forms of the verb *eat* and other words containing the sequence

- **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 tags (see ‘Manage tag columns’ 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 11, which shows a concordance of *fireplace* sorted first by book – so that results from *Barnaby Rudge* (*BR*) come first – and then ordered by the co-text on the right.

Showing 1 to 50 of 50 entries from 11 books

Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
1 Chigwell; where they widened and enlarged the	fireplace	for him, hung up the boiler, and furthermore p	BR	82	14	54	
2andlesticks in his hands, bowed them up to the	fireplace;	Hugh, striding in with a lighted brand and pile	BR	35	46	102	
3the window and secured it, and coming to the	fireplace,	prepared to sit down with his face to the close	BR	17	47	122	
4nately I found myself backing Traddles into the	fireplace,	and bowing in great confusion to two dry little	DC	41	31	85	
5a fall into the ashes of that remarkable kitchen	fireplace.	But we were not in condition to judge of this	DC	28	13	46	
6air, and stood on the hearth-rug in front of the	fireplace.'	Have the goodness to show Mr. Copperfield,' s	DC	38	9	36	
7ronmonger who had made Mrs. Crupp's kitchen	fireplace,	that it was capable of cooking nothing but chc	DC	24	19	55	
8vo little bird-like ladies.The clock ticks over the	fireplace,	the weather-glass hangs in the hall.Neither clo	DC	43	3	9	
9reet.It sat down, with its eyes upon the empty	fireplace,	and as it lost itself in thought there shone into	DS	59	91	330	
10Carker, taking up his usual position before the	fireplace,	and keeping on his hat, 'this is a bad business	DS	32	96	225	
11f himself, brooded and brooded over the empty	fireplace.	Now it lifted up its head, examining the lines a	DS	59	87	318	
12mself between his berth and a very little brass	fireplace,	serve out for self and friend.They soon reappe	DS	23	120	260	
13a kettle-holder on opposite sides of the parlour	fireplace.	The greater part of the furniture was of the pc	DS	7	6	18	
14point of learning, he said, hanging up over the	fireplace.	There were some books in it, too, and a fishing	DS	14	35	96	
15f similar volumes) gorged with burials.Over the	fireplace	was a ground-plan of the vaults underneath th	DS	5	111	245	

Figure 11: Concordance of *fireplace* in DNov (Dickens's Novels) – 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 2017.⁸

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:

- **'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.
- **'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.

⁸ 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/>

The total number of matching rows will be displayed at the top; the process is illustrated in Figures 12–17. Figure 12 shows the plain concordance lines as returned when searching for *fire* in Dickens's novels.

Showing 1 to 50 of 1,706 entries from 15 books

	Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
1	at a keeper's lodge and seeing the light of a	fire	upon the latticed panes, and smoke rising from t	BH	2	3	23	
2	to my Lady, who is on a sofa near the	fire,	shading her face with a hand-screen."It would be	BH	2	16	83	
3	among the national bulwarks.It happens that the	fire	is hot where my Lady sits and that the hand-scr	BH	2	26	101	
4	mother and I were sitting at the table before the	fire.	The clock ticked, the fire clicked; not another so	BH	3	9	41	
5	at the table before the fire.The clock ticked, the	fire	clicked; not another sound had been heard in the	BH	3	9	42	
6	my boxes, I asked him whether there was a great	fire	anywhere?For the streets were so full of dense t	BH	3	112	281	
7	it--and politely put an arm-chair for me by the	fire.	He then called my attention to a little looking-gle	BH	3	118	291	
8	tleman gave me as he spoke.He then stirred the	fire	and left me.Everything was so strange--the strai	BH	3	121	300	
9	an I went on, thinking, thinking, thinking; and the	fire	went on, burning, burning, burning; and the cand	BH	3	122	303	
10	ntleman were standing near a great, loud-roaring	fire.	A screen was interposed between them and it, a	BH	3	125	310	
11	in, and I saw in the young lady, with the	fire	shining upon her, such a beautiful girl!With such	BH	3	126	312	
12	se sitting in the window-seat, with the light of the	fire	upon us, talking together as free and happy as c	BH	3	128	316	
13	sat, he stood by us, in the light of the	fire,	talking gaily, like a light-hearted boy.He was ven	BH	3	130	321	
14	to talk about, and we talked about it; and the	fire,	which had left off roaring, winked its red eyes at	BH	3	130	324	

Figure 12: The first concordance lines of *fire* in DNov (Dickens's Novels) with the default sorting by 'in bk'

Showing 1 to 50 of 1,706 entries from 15 books

	Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
1	at a keeper's lodge and seeing the light of a	fire	upon the latticed panes, and smoke rising from t	BH	2	3	23	
2	to my Lady, who is on a sofa near the	fire,	shading her face with a hand-screen."It would be	BH	2	16	83	
3	among the national bulwarks.It happens that the	fire	is hot where my Lady sits and that the hand-scr	BH	2	26	101	
4	mother and I were sitting at the table before the	fire.	The clock ticked, the fire clicked; not another so	BH	3	9	41	
5	at the table before the fire.The clock ticked, the	fire	clicked; not another sound had been heard in the	BH	3	9	42	
6	my boxes, I asked him whether there was a great	fire	anywhere?For the streets were so full of dense t	BH	3	112	281	
7	it--and politely put an arm-chair for me by the	fire.	He then called my attention to a little looking-gle	BH	3	118	291	
8	tleman gave me as he spoke.He then stirred the	fire	and left me.Everything was so strange--the strai	BH	3	121	300	
9	an I went on, thinking, thinking, thinking; and the	fire	went on, burning, burning, burning; and the cand	BH	3	122	303	
10	ntleman were standing near a great, loud-roaring	fire.	A screen was interposed between them and it, a	BH	3	125	310	
11	in, and I saw in the young lady, with the	fire	shining upon her, such a beautiful girl!With such	BH	3	126	312	
12	se sitting in the window-seat, with the light of the	fire	upon us, talking together as free and happy as c	BH	3	128	316	
13	sat, he stood by us, in the light of the	fire,	talking gaily, like a light-hearted boy.He was ven	BH	3	130	321	
14	to talk about, and we talked about it; and the	fire,	which had left off roaring, winked its red eyes at	BH	3	130	324	
15	and sitting in an arm-chair at a table near the	fire,	was his lordship, whose robe, trimmed with beau	BH	3	132	328	
16	y," said Mr. Kenge, standing with his back to the	fire	and casting his eyes over the dusty hearth-rug a	BH	4	3	6	

Figure 13: Selecting types related to sitting from the KWICGrouper to group the concordance lines

Figure 13 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 *situation* 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 **situation**. I'm actually 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.]

The screenshot shows the CLiC concordance interface. The main table displays search results for the term 'fire'. The table has columns for Left, Node, Right, Book, Ch., Par., Sent., and In bk. The search results are filtered to show only the left side of the node (L1). The KWICGrouper settings are visible on the right, showing a search span of 1 and search types including 'a-sittin', 'sat', 'sit', 'sits', 'sittin', and 'sitting'.

Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
1 a question?" said I when we had sat before the	fire	a little while." Five hundred," said Ada. "Your cousin BH	BH	4	64	152	
2 ft clear, we approached the woman sitting by the	fire	to ask if the baby were ill. She only looked	BH	8	99	338	
3 put back an hour, and we were sitting round the	fire	with no light but the blaze when the hall-door su: BH	BH	9	27	77	
4 I had known the other! And we sat before the	fire,	and I had all the talking to myself for a	BH	13	94	233	
5 George is becoming thoughtful, sitting before the	fire	in the whitewashed room, which has a sanded flo: BH	BH	27	83	195	
6 to her boudoir, she sits down thoughtfully by the	fire,	and inattentive to the Ghost's Walk, looks at Ros: BH	BH	28	73	168	
7 nplacent. My Lady, as on that day, sits before the	fire	with her screen in her hand. Sir Leicester is partic: BH	BH	29	4	21	
8 invite me into her room, where she sat before the	fire	in a great chair; and, dear me, she would tell	BH	30	3	13	
9 my basket of keys, stop to sit down by her	fire	and accommodate myself for a little while to her, BH	BH	30	37	91	
10 npression of having seen Phil sitting by the frosty	fire	staring at him open-mouthed, and gallops away a: BH	BH	56	44	138	
11 to bed in her own room or sit by the	fire	in her own room, but must come forth with her	BH	58	86	233	
12 Miss Summerson," he said, "and sit down by the	fire.	Mr. Woodcourt, from information I have received	BH	59	47	116	
13 man to be. In the front kitchen, sitting by the	fire,	was Mrs. Snagsby, with very red eyes and a very	BH	59	50	127	
14 eak unkind. Come in and sit yourself down by the	fire;	there's a good dear--do.' Mrs Varden complied. Th: BR	BR	22	17	45	

Figure 14: The resulting 'KWICGrouper' 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 14 shows the resulting concordance lines according to the KWICGrouper settings after manually choosing types related to the action of sitting. 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.

The screenshot shows the CLiC concordance interface. The main table displays search results for the term 'back'. The table has columns for Left, Node, Right, Book, Ch., Par., Sent., and In bk. The search results are filtered to show only the left side of the node (L1). The KWICGrouper settings are visible on the right, showing a search span of 1 and search types including 'back'.

Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
1 and a cat sitting in the back part by a	fire.	The old man rises and comes forward, with anot BH	BH	10	52	168	
2 derby, pushing away his plate, and leaning back	.' Fire	away!" "I ha' coom," Stephen began, raising his ey HT	HT	11	18	58	
3 Lamble, standing with his back to the bedroom	fire,	with one hand under his coat-skirts, and all his OMF	OMF	34	53	122	
4 odsnap stood with his back to the drawing-room	fire,	pulling up his shirtcollar, like a veritable cock of OMF	OMF	11	147	318	
5 aby locked his outer door, and came back to his	fire.'	Well done you!" said Fascination to himself. 'Slow OMF	OMF	34	139	297	
6 behind, as he stood with his back to the kitchen	fire	to draw the damp out: which was not calculated GE	GE	6	4	20	
7 al students, in a small back parlour with a large	fire.'	But I am delighted to see you," said Mr. Ben PP	PP	54	147	230	
8 patiently put himself with his back to the sacred	fire	which had burned through so many years. There: LD	LD	46	48	135	
9 tle fellow, thoughtfully, as he looked back at the	fire	'I wonder why it didn't save me my Mama.' He DS	DS	8	27	69	
10 y," said Mr. Kenge, standing with his back to the	fire	and casting his eyes over the dusty hearth-rug : BH	BH	4	3	6	
11 said Mr. Skimpole, standing with his back to the	fire	after giving up the sketch when it was half finis: BH	BH	6	118	347	
12 id Mr. Turveydrop, standing with his back to the	fire	and waving his gloves condescendingly. "Go on, BH	BH	14	75	206	
13 as if they were pictures--and comes back to the	fire	repeating, "Sir Leicester Dedlock, Baronet, and : BH	BH	27	6	17	
14 n, stands on the hearth-rug with his back to the	fire,	watching the disappearance of Mr. Smallweed ar: BH	BH	27	60	138	
15 me again; and Mrs Pipchin, with her back to the	fire,	stood, reviewing the new-comers, like an old sol: DS	DS	8	91	222	
16 rom his chair, and standing with his back to the	fire.'	Have you anything there for me?" "I don't know t: DS	DS	13	21	60	

Figure 15: 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 (see chapter 6 of

Mahlberg 2013⁹) that the cluster *with his back to the fire* is prominent in Dickens's and 19th century novels by other writers. Figure 15 shows the first concordance lines of *fire* with *back* on the left (sorted to 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 14 and 15 there are only lines with one match, but no lines with more than one match. So, in Figure 15, 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 16, there are 27 lines in which *fire* co-occurs with both *back* and *his*. Most of these occurrences appear in the pattern *with his back to the fire*, as becomes obvious when we reverse the sorting on the left so that *the* occurs at the top in the first position to the left of *fire* – the L1 position. On the other hand, as we can see from Figure 17, Dickens's novels contain only instance of *fire* co-occurring with *back* and *her* (*with her back to the fire*).

The screenshot shows the CLiC v1.6.1 interface. At the top, it displays the logos for CLiC, the University of Birmingham, and the University of Nottingham. Below the logos, it states: "Showing 1 to 50 of 1,706 entries from 15 books, 27 entries with 2 KWIC matches, 141 entries with 1 KWIC match".

The main table displays concordance lines. The columns are: Left, Node, Right, Book, Ch., Par., Sent., and In bk. The first 10 rows show lines where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 11th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 12th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 13th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 14th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 15th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 16th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 17th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 18th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 19th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 20th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 21st row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 22nd row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 23rd row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 24th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 25th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 26th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms. The 27th row shows a line where 'fire' is the central term, and 'back' and 'his' are the matching terms.

On the right side, there is a search interface. It includes a search bar with the text "Search the corpora:" and a dropdown menu showing "Dickens's Novels". Below this, there is a section for "Only in subsets:" with a dropdown menu showing "All text". There is also a section for "Search for terms:" with a text input field containing "fire". Below this, there is a section for "Filter rows:" with a text input field containing "e.g. 'hands'". There is also a section for "KWICGrouper" with a search in span slider. At the bottom, there is a section for "Search for types:" with a text input field containing "back" and "his".

Figure 16: The 27 lines with two matches (here, *back* and *his*) are highlighted in a darker green

⁹ Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction*. London: Routledge.

Showing 1 to 50 of 1,706 entries from 15 books, 1 entries with 2 KWIC matches, 103 entries with 1 KWIC match

Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.
ome again; and Mrs Pipchin, with her back to the	fire,	stood, reviewing the new-comers, like an old solc	DS	8	91	222	
ment, smouldered within her like an unwholesome	fire.	All closely imprisoned forces rend and destroy.	TI	HT	29	34	89
otent smell of burnt pens when her feathers took	fire,	which acted as a restorative and woke her.	Mr W.	OMF	5	154	378
fire rose, and to dart back into it when the	fire	fell, were sufficient witnesses of his making him	LD	48	7	41	
as I have often known her sit and watch the	fire	- for an hour at a stretch."Ay, ay?Has resources	HT	19	34	73	
housebreaker, sat with her eyes fixed upon the	fire,	as if she had been deaf to all that passed	OT	19	46	89	
Still upon her knees, and with her eyes upon the	fire,	and the fire shining on her ruined beauty and he	DS	53	111	264	
on in a low voice, with her eyes upon the	fire.'	I, with poor father's grave not even cleared of ur	OMF	28	127	267	
fin, with her work abandoned, her eyes upon the	fire,	and her chin upon her hand, 'do you recollect, N	OMF	38	39	97	
the woman seemed to turn her face towards the	fire	as if to hide her bruised eye; nobody gave us	BH	8	84	276	
ochin,' said Mr Dombey, turning back towards the	fire,	'you know what to do with them, I presume.	You	DS	44	33	63
good girl.You had better take her in to the	fire,'	said the locksmith.'Master sounds unfeeling, min	BR	22	16	42	
sat down at the tea-board, with her face to the	fire.	When once again, in a tone more terrible than th	MC	25	130	351	
icture of a race-horse, with her head close to the	fire,	and her feet pushing the mustard off the dumb-	DC	17	87	223	
sit up with him, drew her chair close to the	fire	and went off into a series of short naps, cheque	OT	12	25	56	
y," said Mr. Kenge, standing with his back to the	fire	and casting his eyes over the dusty hearth-rug a	BH	4	3	6	

Search for types: back x her x

Figure 17: 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 18 illustrates the outcome of what a tagged concordance can look like. The tags are user-defined 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.

Showing 1 to 10 of 10 entries from 1 book

Left	Node	Right	In bk.	dick	you-one	oliver	rose	others
I know the doctor must be right, Oliver, because I	dream	so much of Heaven, and Angels, and kind faces th		✓				
asy state, between sleeping and waking, when you	dream	more in five minutes with your eyes half open, and			✓			
om what seemed to have been a long and troubled	dream.	Feebly raising himself in the bed, with his head re				✓		
pity and compassion had awakened some pleasant	dream	of a love and affection he had never known.Thus				✓		
all that is going on about us, and, if we	dream	at such a time, words which are really spoken, or			✓			
covert for the same reason.'It must have been a	dream,	Oliver,' said Harry Maylie.'Oh no, indeed, sir,' repli				✓		
nterview, which had more the semblance of a rapid	dream	than an actual occurrence, sank into a chair, and e				✓		
ar, and profligacy and riot have staggered home to	dream;	it was at this still and silent hour, that Fagin					✓	
of some old friend flashing on one in a vivid	dream.	I need not tell you he was snared away before		✓				
and all his recent life had been but a happy	dream.	But it was pure, earnest, joyful reality.They drove s		✓				

Manage tag columns...

Figure 18: 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 18) and create your own tag(s) through the 'Add new' option (see Figure 19). 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 one or more lines are selected you can click the tick next to the relevant tag in order to tag the line (see Figure 20). 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.

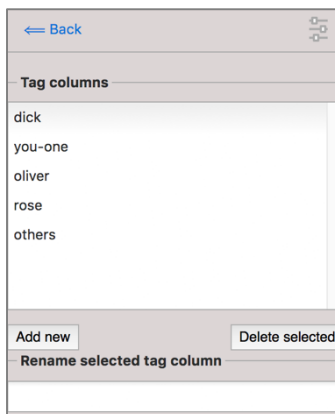


Figure 19: The menu for adding and renaming tags.

Figure 20: Select a line (by clicking on it) in order to apply an existing tag; once tagged, the tick in the sidebar will appear green for the selected line. A tick will also be added to the tag column in the concordance itself

5.3. Subsets

The Subsets tab can display the full subset of your choice for the selected corpus. Therefore, you can retrieve all quotes or all long suspensions, etc. in any of the books or pre-selected corpora for further analysis. Note that we find this option most useful for the smaller subsets, i.e. quotes and suspensions; if you select the whole 'non-quotes' subset the output may become unwieldy.

- **Subsets – Show subsets**

Click onto the dropdown '**Show subsets**' (see Figure 21) to select a relevant subset (short suspensions, long suspensions, quotes or non-quotes). You will also need to choose a corpus.

Subsets

Search the corpora:

Select Some Options

Show subsets:

Select an Option

Short Suspensions

Long suspensions

Quotes

Non-quotes

Figure 21: The basic subset options

Figure 22 shows sample lines from the subset of long suspensions in *Oliver Twist*. You can then use the filter option to narrow down the lines and group them using the KWICGrouper. As in the concordance tab, you can create tags to annotate rows with noteworthy patterns.

CLiC v1.6.1				
UNIVERSITY OF BIRMINGHAM Arts & Humanities Research Council The University of Nottingham UNITED KINGDOM · CHINA · SINGAPORE				
Showing 1 to 50 of 315 entries from 1 book				
Left	Node	Right	In bk.	
1 a beadle's.'Lor, only think,'	said Mrs. Mann, running out,--for the three boys had been removed by this time,--'	only think of that!That		
2 or proper conduct, Mrs. Mann,'	inquired Mr. Bumble, grasping his cane,	'to keep the parish officers		
3 relaxed.'Well, well, Mrs. Mann,'	he replied in a calmer tone;	'it may be as you		
4 said Mrs. Mann.'Well, well,'	said the beadle, evidently gratified with the compliment;	'perhaps I may be.Perhaps		
5 quite at his ease.'Boy,'	said the gentleman in the high chair,	'listen to me.You know		

Figure 22: The first few lines from the subset of 'long suspensions' in *Oliver Twist*

- Subsets – Results

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

- Subsets – Results – Filter rows

The filter option lets you filter the output by the rows that contain a particular sequence of letters, as described in Section 5.2 on the filter function in the Concordance tab. For example, you could filter suspensions for particular speech verbs like *cried* (Figure 23).

CLiC v1.6.1				
UNIVERSITY OF BIRMINGHAM Arts & Humanities Research Council The University of Nottingham UNITED KINGDOM · CHINA · SINGAPORE				
Showing 1 to 21 of 21 entries (filtered from 315 total entries) from 1 book				
Left	Node	Right	In bk.	
1 the office!"D--n me!"	cried the old gentleman, bursting out with the rage he had kept down so long,	'd--n me!I'll--"Clear		
2 no reply.'God forgive me!"	she cried after a while,	'I never thought of this		
3 kissed the matron.'Mr. Bumble!"	cried that discreet lady in a whisper; for the fright was so great, that she had quite lost her voice,	'Mr. Bumble, I shall scream		
4 her?--what about--I know!"	she cried, jumping fiercely up: her face flushed, and her eyes starting from her head--"	I robbed her, so I		
5 on in intense dismay.'Hark!"	cried the Dodger at this moment,	'I heard the tinkler.'Catching		

Figure 23: Filtering long suspensions in *Oliver Twist* for *cried*

Left	Node	Right	In bk.
1 said Mrs. Mann. 'Well, well,'	said the beadle, evidently gratified with the compliment;	'perhaps I may be. Perhaps	
2 Perhaps she does see me,'	whispered Oliver, folding his hands together;	'perhaps she has sat by	

Figure 24: Filtering the co-text of long suspensions for *perhaps* in *Oliver Twist*

Note, however, that the filter will search 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 24).

- **Subsets – KWICGrouper**

If you want to restrict your search to the subset itself, the KWICGrouper is the better option; it will also highlight your search terms, as described in Section 5.2 on concordances. The Subset KWICGrouper works like the Concordance KWICGrouper, with the exception of its search span which operates only on the subset itself. See Figure 25 for an illustration of the Subset KWICGrouper searching for lines with *cried*, *screamed* and *sobbed*.

Left	Node	Right	In bk.
1 are the--''No, no, sir,'	sobbed Oliver, clinging to the hand which held the well-known cane;	'no, no, sir; I will	
2 the office!' 'D--n me!'	cried the old gentleman, bursting out with the rage he had kept down so long,	'd--n me!!' 'Clear	
3 I don't care for that,'	screamed the girl, struggling violently with the man,	'the child shan't be torn	
4 no reply. 'God forgive me!'	she cried after a while,	'I never thought of this	
5 kissed the matron. 'Mr. Bumble!'	cried that discreet lady in a whisper; for the fright was so great, that she had quite lost her voice,	'Mr. Bumble, I shall scream	
6 her?--what about--I know!'	she cried, jumping fiercely up: her face flushed, and her eyes starting from her head--'	I robbed her, so I	
7 on in intense dismay. 'Hark!'	cried the Dodger at this moment,	'I heard the tinkler. 'Catching	
8 pace with me, to-day. 'Halloa!'	cried the doctor, letting down the front glass in a great hurry, and shouting to the postillion;	'something very short of flying	
9 infant's sake. 'She sold it,'	cried Monks, with desperate eagerness;	'did she sell it?Where	

Figure 25: The search span of the Subset KWICGrouper applies to the subset; not to the co-text

- Subsets – Manage tag columns

Showing 1 to 50 of 3,891 entries from 69 books, 216 entries with 1 KWIC match

	Left	Node	Right	Book	Ch.	Par.	Sent.	In bk.	male	female
1	relation of yours?" "No, no!"	cried Tony, his eyes growing bright,	"I haven't got any relation	alone	15	10	26		✓	
2	up to her face." "Eh!"	cried old Oliver in a wailing voice,	"I don't know whatever the	alone	18	25	57		✓	
3	at 300, Fitzroy Street.'Jane,'	cried Cyril with great presence of mind,	'go and get the plate	amulet	7	181	399		✓	
4	of the Babylonian Queen.' Jimminy!	cried Robert, and ran down the steps,	'here she is!' 'Here!' he	amulet	8	72	151		✓	
5	him in a second." "Sawyer,"	he cried in a stern voice,	"is that pony made of	beauty	11	4	13		✓	
6	he said." "Oh, oh, oh!"	cried the boy as he struggled about among the thorns;	"I say, come and help	beauty	13	2	7		✓	
7	he bid seventeen shillings." "Eighteen,"	cried his rival, a short, cheery, cherub-faced little dealer, whose neighbours adjured him to	"sit quiet like a good	brass	2	37	84		✓	
8	Curiouser and curiouser!' "	cried Alice (she was so much surprised, that for the moment she quite forgot how to speak good English);	'now I'm opening out like	alice	2	1	1		✓	
9	somebody else" --but, oh dear!	cried Alice, with a sudden burst of tears,	'I do wish they WOULD	alice	2	13	50		✓	
10	the rats and--oh dear!	cried Alice in a sorrowful tone,	'I'm afraid I've offended it	alice	2	24	99		✓	
11	like Tony was." "Yes, yes!"	cried Dolly, eagerly, raising herself up in his arms;	"It's a nice place, and	alone	17	12	31		✓	
12	unspeakably horrible.' "Oh, go away,"	cried Anthea; but Cyril cried,	'No; stay!' The creature hesitated	amulet	7	158	352		✓	

Figure 26: Tagging subsets – here, long suspensions in ChiLit containing *cried* are tagged for character gender

Just like in the Concordance tab (Section 5.2), subset rows can be annotated with user-defined tags. Figure 26 shows a potential application of tagging subsets: long suspensions in the 19th Century Children's Literature (ChiLit) corpus containing *cried* are tagged for whether the crying character is male or female. Note that this screenshot just illustrates the technique; it does not represent the actual gender distribution of *cried* in the ChiLit long suspensions.

5.4. Clusters

The output of the cluster tool generates frequency lists of single words and 'clusters' (repeated sequences of words). Clusters are also called '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.) CLiC currently supports 1-grams (single words), 3-grams (like *i don't know*) 4-grams and 5-grams (*what do you mean by*), as illustrated in Figure 27.

CLiC 1.6.4aa999a

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

Load | Merge | Save | Clear | About

Concordance

Subsets

Clusters

Please select a subset

Showing 0 to 0 of 0 entries

Cluster	Frequency
No data available in table	

Previous Next

Search the corpora:
Oliver Twist (Charles Dickens) ✕

Only in subsets:
Select an Option ▼

n-gram:
1-gram
3-gram
4-gram
5-gram

Figure 27: Cluster options

As in the other tabs, you can restrict the search to a particular subset (see Figure 27 – ‘**Only in subsets: Select an Option**’) so that, for example, you can create frequency lists for clusters in quotes (or any of 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.

5.5. 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)¹⁰. 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 28):

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

¹⁰ 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

Keywords

Target corpora:

Select Some Options

...within subset:

Select an Option

Reference corpora:

Select Some Options

...within subset:

Select an Option

n-gram:

1-gram

p-value cut-off:

0.0001

– Results

Filter rows:

e.g. 'hands'

Figure 28: The settings for the keywords tab require you to select two sets of corpora for the keyword comparison – target and reference – and their corresponding subsets

CLiC
v1.6.1

UNIVERSITY OF
BIRMINGHAM

Arts & Humanities
Research Council

The University of
Nottingham
UNITED KINGDOM · CHINA · SINGAPORE

[Load](#) | [Merge](#) | [Save](#) | [Clear](#) | [Help](#)

The results are limited to 3000 rows. Generally there will be fewer results. Only overused (positive) keywords are displayed.

Showing 1 to 4 of 4 entries

	N-gram	Target frequency	Ref frequency	Expected target	Expected ref	LL	P
1	what do you mean by	8	14	0.59	21.41	29.82	p < 0.0001
2	not a bit of it	6	6	0.32	11.68	27.11	p < 0.0001
3	have you got to say	5	5	0.27	9.73	22.59	p < 0.0001
4	what have you got to	5	7	0.32	11.68	20.26	p < 0.0001

Previous
1
Next

Target corpora:

Oliver Twist (Charles Dickens) x

...within subset:

Quotes

Reference corpora:

19th Century Reference Corpus x

...within subset:

Quotes

n-gram:

5-gram

p-value cut-off:

0.0001

Figure 29: Key 5-word clusters in *Oliver Twist* 'quotes' compared to 'quotes' in 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 Reference Corpus (quotes), we retrieve the results displayed in Figure 29 (for a p-value of 0.0001). The keyword output is by default ordered by the log-likelihood (LL) value, the 'keyness' statistic used here (for more details on the calculation, please refer to Rayson and Garside, 2000).

The frequency threshold of 5 used for the cluster tab is not 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.

Appendix 1: List of texts available in CLiC**Pre-selected corpora in CLiC**

DNov (15 texts)	Dickens's Novels
ChiLit (71 texts)	19 th Century Children's Literature Corpus
19C (29 texts)	19 th Century Reference Corpus
ArTs (23 texts)	Additional Requested Texts (containing more set titles from A-Level and GCSE exam specifications).

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

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

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

Table 2: Titles in ChiLit; a corpus compiled by Anna Cermakova for the GLARE project¹¹

Title	Author
<i>Vice Versa; or, A Lesson to Fathers</i>	F. Anstey
<i>The Brass Bottle</i>	F. Anstey
<i>The Coral Island: A Tale of the Pacific Ocean</i>	R. M. Ballantyne
<i>Peter Pan</i>	J. M. Barrie
<i>The Secret Garden</i>	Frances Hodgson Burnett
<i>Alice's Adventures in Wonderland</i>	Lewis Carroll
<i>Through the Looking-Glass</i>	Lewis Carroll
<i>The Surprising Adventures of Sir Toady Lion with Those of General Napoleon Smith</i>	S.R. Crockett
<i>Mrs. Overthway's Remembrances</i>	Juliana Horatia Ewing
<i>Jackanapes</i>	Juliana Horatia Ewing
<i>Moonfleet</i>	J. Meade Falkner
<i>Eric; Or, Little by Little</i>	F. W. Farrar
<i>Adventures in Wallypug-Land</i>	G. E. Farrow
<i>The Golden Age</i>	Kenneth Grahame
<i>Dream Days</i>	Kenneth Grahame
<i>The Wind in the Willows</i>	Kenneth Grahame
<i>King Solomon's Mines</i>	H. Rider Haggard
<i>Allan Quatermain</i>	H. Rider Haggard
<i>Winning His Spurs. A Tale of the Crusades</i>	G. A. Henty
<i>With Clive in India; Or, The Beginnings of an Empire</i>	G. A. Henty
<i>Tom Brown's Schooldays</i>	Thomas Hughes
<i>Mopsa the Fairy</i>	Jean Ingelow
<i>Wood Magic, a Fable</i>	Richard Jefferies
<i>Madam How and Lady Why; Or, First Lessons in Earth Lore for Children</i>	Charles Kingsley
<i>The Water-Babies</i>	Charles Kingsley
<i>Stalky and Co</i>	Rudyard Kipling
<i>The Jungle Book</i>	Rudyard Kipling
<i>Prince Prigio</i>	Andrew Lang
<i>At the Back of the North Wind</i>	George McDonald
<i>The Princess and the Goblin</i>	George McDonald
<i>The Three Mulla-mulgars</i>	Walter de la Mare
<i>Masterman Ready; Or, The Wreck of the Pacific</i>	Frederick Marryat
<i>The Settlers in Canada</i>	Frederick Marryat
<i>The Children of the New Forest</i>	Frederick Marryat
<i>Feats on the Fiord</i>	Harriet Martineau
<i>The Crofton Boys</i>	Harriet Martineau
<i>The Peasant and the Prince</i>	Harriet Martineau
<i>The Settlers at Home</i>	Harriet Martineau
<i>A World of Girls: The Story of a School</i>	L. T. Meade
<i>The Cuckoo Clock</i>	Mary Louisa Molesworth
<i>The Tapestry Room: A Child's Romance</i>	Mary Louisa Molesworth
<i>The Carved Lions</i>	Mary Louisa Molesworth
<i>The Book of Dragons</i>	E. Nesbit
<i>The Story of the Treasure Seekers</i>	E. Nesbit
<i>Nine Unlikely Tales</i>	E. Nesbit
<i>The Railway Children</i>	E. Nesbit
<i>Five Children and It</i>	E. Nesbit

¹¹ See <http://birmingham.ac.uk/glare> for more information on the GLARE project.

<i>The Story of the Amulet</i>	E. Nesbit
<i>The Tale of Peter Rabbit</i>	Beatrix Potter
<i>The Tale of Squirrel Nutkin</i>	Beatrix Potter
<i>The Tale of Benjamin Bunny</i>	Beatrix Potter
<i>The Tale of Two Bad Mice</i>	Beatrix Potter
<i>The Tale of Jemima Puddle-Duck</i>	Beatrix Potter
<i>The Tale of the Flopsy Bunnies</i>	Beatrix Potter
<i>The Fifth Form at Saint Dominic's: A School Story</i>	Talbot Baines Reed
<i>The King of the Golden River; or, the Black Brothers: A Legend of Stiria</i>	John Ruskin
<i>Black Beauty</i>	Anna Sewell
<i>Holiday House: A Series of Tales</i>	Catherine Sinclair
<i>Kidnapped</i>	Robert Louis Stevenson
<i>Treasure Island</i>	Robert Louis Stevenson
<i>Jessica's First Prayer</i>	Hesba Stretton
<i>Little Meg's children</i>	Hesba Stretton
<i>Alone in London</i>	Hesba Stretton
<i>The Rival Crusoes; or The Shipwreck</i>	Agnes Strickland
<i>The Rose and the Ring</i>	William Makepeace Thackeray
<i>Leila at home</i>	Ann Fraser Tytler
<i>The Happy Prince, and Other Tales</i>	Oscar Wilde
<i>The Heir of Redclyffe</i>	Charlotte M. Yonge
<i>The Daisy Chain, or Aspirations</i>	Charlotte M. Yonge
<i>The Dove in the Eagle's Nest</i>	Charlotte M. Yonge
<i>The Little Duke: Richard the Fearless</i>	Charlotte M. Yonge

Appendix 2: CLiC texts listed in A-Level and GCSE specifications

Table 3: Overview of CLiC texts listed in the AQA, Edexcel and OCR A-Level / GCSE specifications.

Author	Title	AQA GCSE Eng Lit (8702)	AQA AS Lang Lit (7706)	AQA A-Level Lang Lit (7707)	AQA AS Eng Lit A (7711)	AQA A-Level Eng Lit A (7712)	AQA AS Eng Lit B (7716)	AQA A-Level Eng Lit B2 (7717)	OCR GCSE Eng Lit (J352)	OCR AS Eng Lit (H072)	OCR A Level Eng Lit (H472)	OCR AS Lang Lit (H074)	OCR A-Level Lang Lit (H474)	Edexcel GCSE Eng Lit	Edexcel AS Lang Lit	Edexcel A-Level Lang Lit	Edexcel AS Eng Lit	Edexcel A-Level Eng Lit	Total
Jane Austen	<i>Emma</i>					✓	✓												2
Jane Austen	<i>Mansfield Park</i>					✓													1
Jane Austen	<i>Northanger Abbey</i>					✓													1
Jane Austen	<i>Persuasion</i>				✓	✓													2
Jane Austen	<i>Pride and Prejudice</i>	✓							✓					✓					3
Jane Austen	<i>Sense and Sensibility</i>										✓								1
Mary Elizabeth Braddon	<i>Lady Audley's Secret</i>																✓	✓	2
Anne Brontë	<i>The Tenant of Wildfell Hall</i>					✓													1
Charlotte Brontë	<i>Jane Eyre</i>	✓		✓	✓				✓		✓	✓	✓	✓					8
Emily Brontë	<i>Wuthering Heights</i>			✓	✓										✓	✓	✓	✓	6
Kate Chopin	<i>The Awakening</i>			✓	✓														2
Wilkie Collins	<i>Moonstone</i>				✓												✓	✓	3
Wilkie Collins	<i>The Woman in White</i>				✓														1
Joseph Conrad	<i>Heart of Darkness</i>																✓	✓	2
Charles Dickens	<i>A Christmas Carol</i>	✓												✓					2
Charles Dickens	<i>Great Expectations</i>	✓							✓					✓	✓	✓			5
Charles Dickens	<i>Hard Times</i>				✓		✓										✓	✓	4
Charles Dickens	<i>Oliver Twist</i>						✓												1
Sir Arthur Conan Doyle	<i>The Sign of the Four</i>	✓																	1
George Eliot	<i>Silas Marner</i>													✓					1
George Eliot	<i>The Mill on the Floss</i>			✓	✓														2
George Eliot	<i>Middlemarch</i>				✓														1
E. M. Forster	<i>A Room with a View</i>			✓	✓										✓	✓			4
Elizabeth Gaskell	<i>North and South</i>				✓														1
Charlotte Perkins Gilman	<i>The Yellow Wallpaper</i>				✓														1
Thomas Hardy	<i>Tess of D'Urbervilles</i>			✓	✓	✓	✓				✓				✓	✓	✓	✓	9
Henry James	<i>The Portrait of a Lady</i>									✓									1
Henry James	<i>What Maisie Knew</i>																✓	✓	2
D. H. Lawrence	<i>Women in Love</i>									✓									1
Solomon Northup	<i>Twelve Years a Slave</i>											✓	✓						2
Mary Shelley	<i>Frankenstein</i>	✓	✓	✓		✓					✓			✓			✓	✓	8
Robert Louis Stevenson	<i>The Strange Case of Dr Jekyll and Mr Hyde</i>	✓							✓					✓					3
Bram Stoker	<i>Dracula</i>		✓	✓							✓				✓	✓	✓	✓	7
William Makepeace Thackeray	<i>Vanity Fair</i>					✓													1
Mark Twain	<i>The Adventures of Huckleberry Finn</i>										✓						✓	✓	3
H. G. Wells	<i>The War of the Worlds</i>								✓								✓	✓	3
Rebecca West	<i>The Return of the Soldier</i>					✓													1
Oscar Wilde	<i>The Picture of Dorian Gray</i>					✓					✓						✓	✓	4