



USER GUIDE

Version December 2021

Table of Contents

1. INTRODUCTION.....	4
2. ACCESS TO THE SITE.....	4
3. OVERVIEW OF TOOLS.....	6
3.1 SEARCH: SEARCHING WITH PRE-DETERMINED SEARCH TERMS.....	7
3.1.1 BASIC SEARCH	8
BASIC SEARCH OPTIONS.....	8
1. Search term.....	8
2. Related terms.....	Error! Bookmark not defined.
3. Date range.....	9
4. Houses of Parliament.....	9
5. Reset/New Search.....	9
DISTRIBUTION GRAPHS.....	10
COMPARISONS.....	10
CONCORDANCE LINES.....	11
1. Concordance lines in document format or KWIC format	12
2. Showing debate title	13
3. Ordering by relevance.....	14
4. Excluding (meta)data from the display of concordance lines.....	14
5. Viewing more contributions	15
6. Download concordance lines.....	15
7. Clicking on MP's name	16
CONCORDANCE LINES FOR COMPARISON SEARCHES	16
COMPLETE CONTRIBUTIONS.....	17
3.1.2 ADVANCED SEARCH	18
ADVANCED SEARCH OPTIONS.....	18
1. Specific date range.....	18
2. Limiting by MP	19
3. Limiting by debate title	19
DISPLAY OF PATTERNS OVER MONTHS RATHER THAN YEARS	20
3.2 EXPLORE: SEARCHING WITHOUT PRE-DETERMINED SEARCH TERMS	20
3.2.1 EXPLORE BY WORD CLOUD	21
3.2.2 EXPLORE BY KEYWORDS	22
4. DISCLAIMER.....	26
5. NOTES.....	27

PERFORMANCE ISSUES	27
COMMENTS/SUGGESTIONS/BUG FIXES	27
CITING/REFERENCING HANSARD AT HUDDERSFIELD	27
6. ACKNOWLEDGEMENTS	28
APPENDIX I	29

1. INTRODUCTION

The Hansard at Huddersfield website aims to help users find new ways to search the content of parliamentary debates contained in Hansard, the official record of U.K. parliament, without the need for expert software.

This site contains nearly every contribution to parliamentary debate¹ from both Houses of the U.K. Parliament between 01st January 1803 and 16th December 2021, and allows for interactive engagement with the complex linguistic, historic and thematic patterns in the language of Hansard through simple searches.

Our application is inspired by the methods of corpus linguistics whose software is ideal for studying patterns of language use in large datasets but can be difficult to use for those without the relevant training in linguistics and/or statistics. There are also challenges in interpreting the search results of this software.

In order to provide the benefits of corpus searching to a wider group of Hansard users this site enables the user to carry out searches based on corpus linguistic search tools which are made more intuitive and accessible by combining them with interactive visualisations.

The Hansard at Huddersfield project started in March 2018 at the University of Huddersfield and was funded by the Arts and Humanities Research Council, grant reference AH/R007136/1. roject team: Prof. Lesley Jeffries (Principal Investigator), Prof. Marc Alexander (Co-Investigator), Dr Alexander von Lünen (Co-Investigator and Technical Lead), Paul Crossley (Programmer, Research Fellow), Fransina Stradling (Research Assistant). Former team member: Dr Hugo Sanjurjo González (Programmer, Research Fellow).

2. ACCESS TO THE SITE

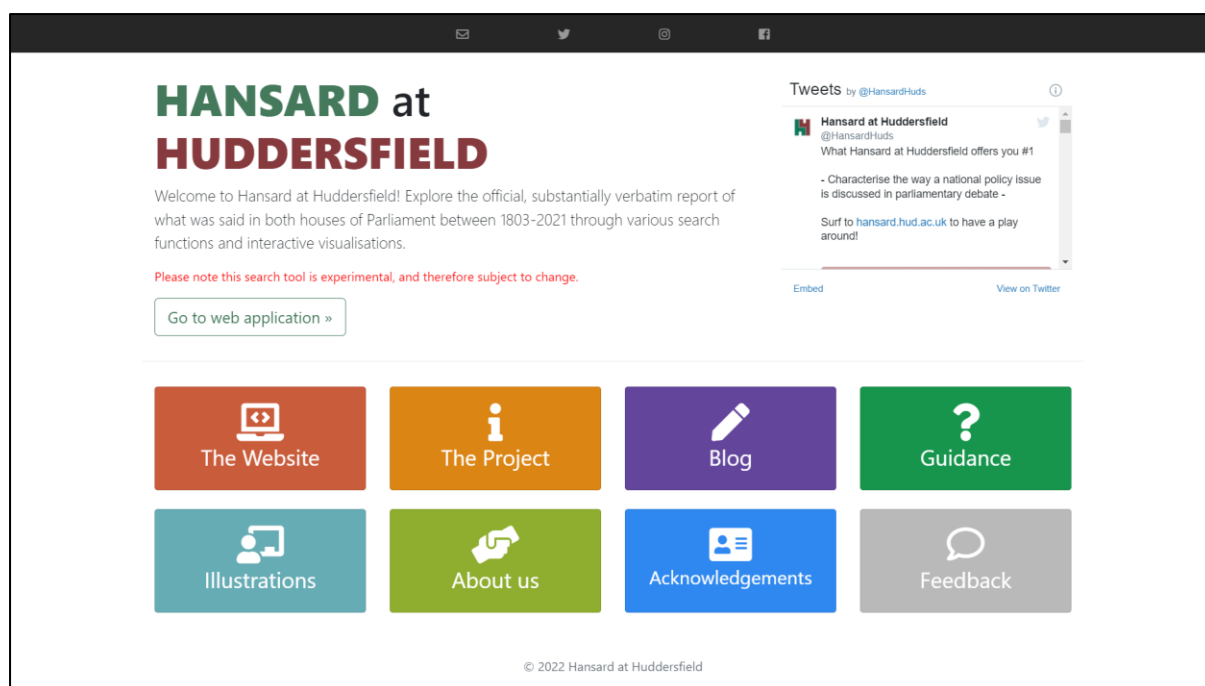
The Hansard at Huddersfield site works best in Google Chrome or Mozilla Firefox web browsers. It should also work well on most mobile devices that support Android or Apple.

Once you have landed on the front page of hansard.hud.ac.uk, you can access the search tools by clicking on the button 'Go to web application'.

Other information which can be accessed from the front page (see screenshot below) is:

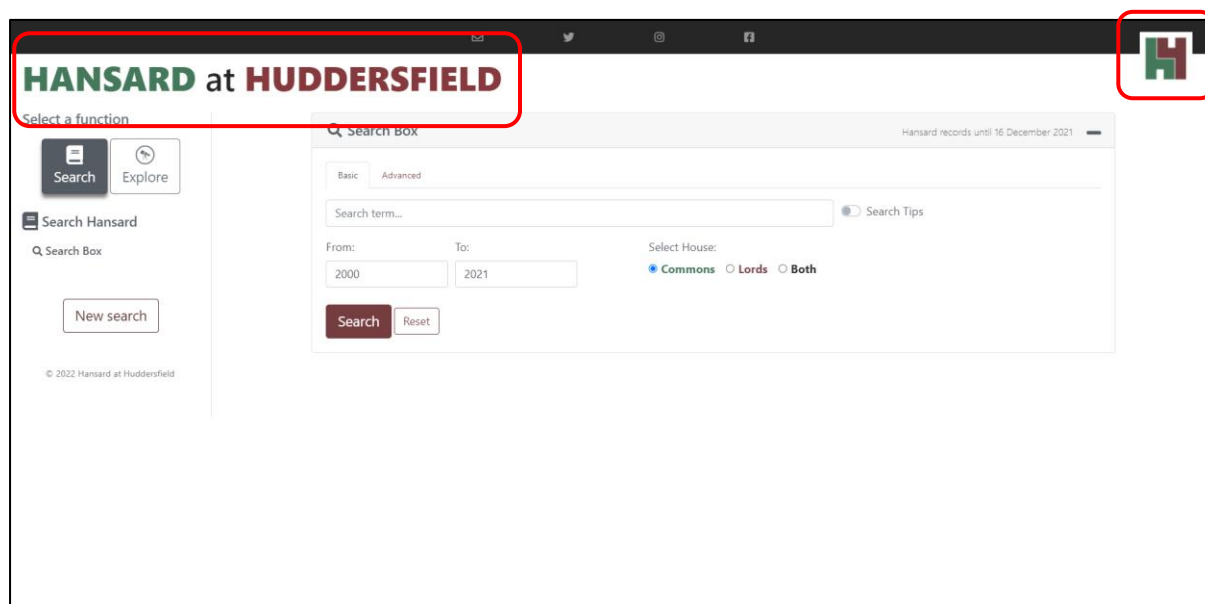
- A description of the website, including a brochure;
- A description of the project, including technologies used, outputs and events;
- Our blog, including further information and blogs by our users about how they are using the site (opens in a new window);
- Guidance, including this user guide, tutorials for how to use the site and how to cite it;
- Illustrations of usage to show in more detail how complex searches can be carried out on socio-political topics, including downloadable case studies and examples of researchers using the site;
- About us, i.e. where we are, who we are working with who our supporters are;
- Acknowledgements, i.e. whose contributions to the project were invaluable;
- A feedback form.

¹ Missing contributions are also missing from the official Hansard records. The following years are missing from the Commons: 1814, 1816 and 1829. The following years are missing from the Lords: 1816 and 1829. Hansard (Lords and Commons) is working on filling in these gaps in the data.



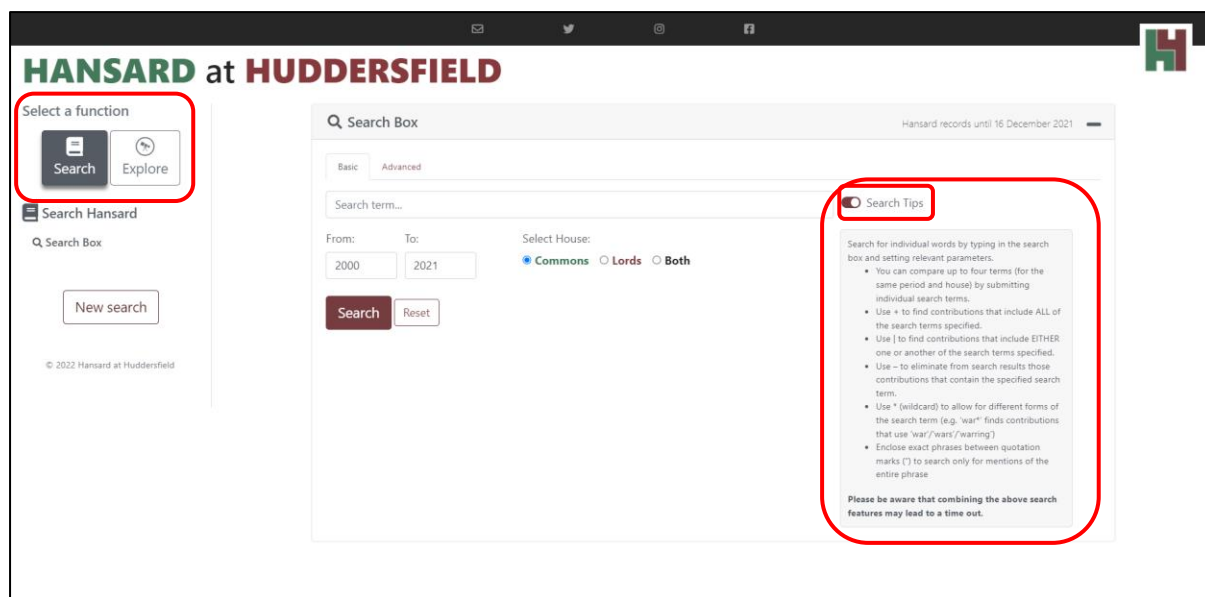
Front page [Hansard.hud.ac.uk](https://hansard.hud.ac.uk)

From the search screens you can return to the front page by clicking the words “Hansard at Huddersfield” at the top left or the logo at the top right of the screen:



Search Screen, access to front page via logo

The ‘Search Tips’ button next to the search box reminds you of the kinds of searches you can do with the search box. Click the ‘Search Tips’ button to display or hide these tips.



Search Tips

These tips have to do with the possibility to search using so-called Boolean search operators. These are commonly used to search databases for more complex searches that concern multiple search terms. On our site, you can use the following operators:

- Use + to find contributions that include ALL of the search terms specified.
- Use | to find contributions that include EITHER one or another of the search terms specified.
- Use – to eliminate from search results those contributions that contain the specified search term.
- Use * (wildcard) to allow for different forms of the search term (e.g. 'war*' finds contributions that use 'war'/'wars'/'warring')

Furthermore, note that you need to enclose exact phrases between quotation marks (") to search only for mentions of the entire phrase.

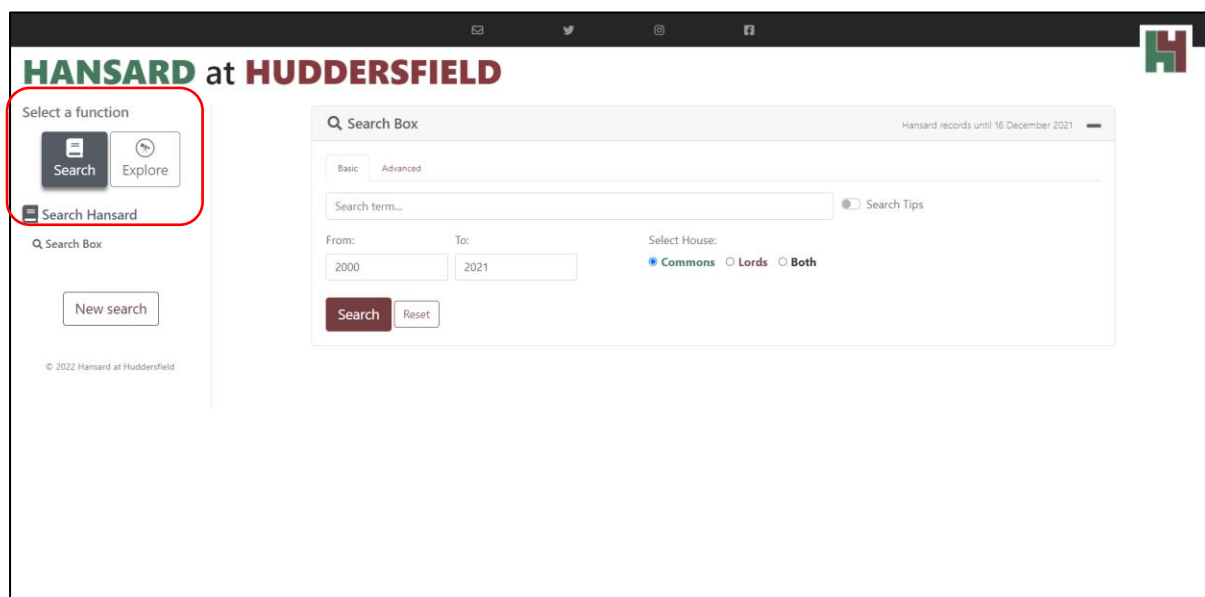
Please also be aware that combining the above search features may lead to a time out. The database cannot always handle the complexity of these searches.

3. OVERVIEW OF TOOLS

There are two main entry points to get through to the Hansard data:

1. Search requires you to choose your search terms (*Search*)
2. Explore allows you to search without pre-determined search terms (*Explore*).

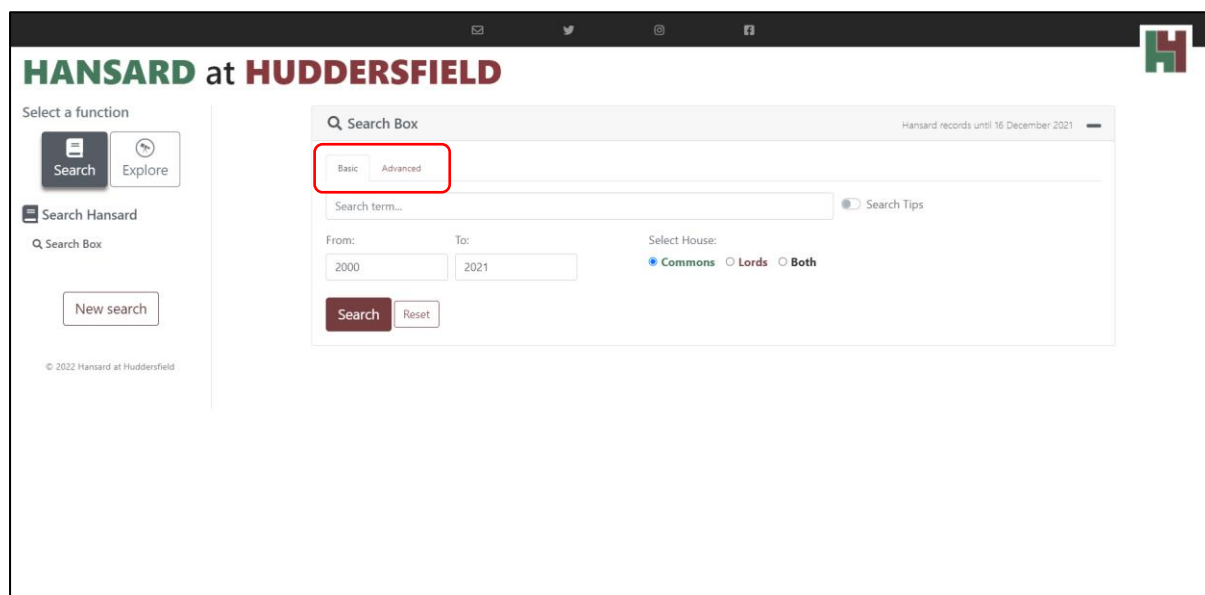
You can access either entry point by clicking on the large 'Search' or 'Explore' buttons under 'Search Function' in the top left corner of your screen:



Search Screen options

3.1 SEARCH: SEARCHING WITH PRE-DETERMINED SEARCH TERMS

The *Search* screen is the default screen to search Hansard. Accessing Hansard through this entry point is possible through two search boxes: the *Basic Search* and the *Advanced Search* (see screenshot 5). The Advanced Search box offers more options than the Basic Search and allows for limiting your search by debate title, by member and by whole dates. While the Basic Search screen is the default, you can access the Advanced Search screen by clicking on 'Advanced' under 'Search Box'.



Basic and Advanced Searches

3.1.1 BASIC SEARCH

The Basic Search screen allows you to search Hansard using any pre-determined search word or multi-word phrase within your chosen date range and for your selected house(s) of parliament. Any given search will give you a distribution graph, after which selection for a particular date range from the graph will give you a list of all contributions within that given date range.

When you search for a term on this page, after typing the term in the search box you can press 'Enter', or press 'Search'. You can include up to four terms in one search (see paragraphs 3.1.1.2.1 and 3.1.1.3.1 on comparison for more information) and you will see their distribution graphs overlaid on top of each other for your chosen date range and House(s) of Parliament.

The screenshot shows the 'HANSARD at HUDDERSFIELD' website. On the left, there's a sidebar with 'Select a function' (Search, Explore), 'Search Hansard', and a 'New Search' button. The main area is titled 'Search Box' and includes tabs for 'Basic' and 'Advanced'. It features a search term input field (1), date range selectors (2), house selection (3), and search buttons (4). The search buttons are labeled 'Search' and 'Reset'.

Basic Search start screen

BASIC SEARCH OPTIONS

1. Search term

This is where you type in the word or phrase (search term) you wish to search for. You can use a 'wildcard' (*) to allow for different forms of the search term. Thus, *war** would find *war*, but also *wars* and *warring* or *warlike*. Similarly, *asylum seeker** would find *asylum seekers* and *asylum seeking* as well as the singular *asylum seeker*. If you press <Enter> or select the 'Search' button after typing in your search term, you will see it in its context below.

The screenshot shows the 'HANSARD at HUDDERSFIELD' search page. On the left, there's a sidebar with 'Select a function' (Search, Explore), 'Search Hansard', and a 'New search' button. The main search area has a 'Search Box' with 'asylum' entered. Below it, 'From:' is 2000 and 'To:' is 2021. Under 'Select House:', 'Commons' is selected. A 'Search' button and a 'Reset' button are at the bottom. A red box highlights the search input field.

2. Date range

The date range on the Basic search page is in full years only and is set to a default 2000-2019. You can change the date range by typing a new year into each box or by using the arrows.

This close-up shows the date range selection. The 'From:' field has '2000' and the 'To:' field has '2021'. A red box highlights both fields.

3. Houses of Parliament

You can choose between showing results from the House of Commons or the House of Lords only, or both. For some of the slower searches, choosing one or other House, rather than both together may be a way to achieve quicker results.

This close-up shows the 'Select House:' section. The 'Commons' radio button is selected, with 'Lords' and 'Both' also visible. A red box highlights the entire section.

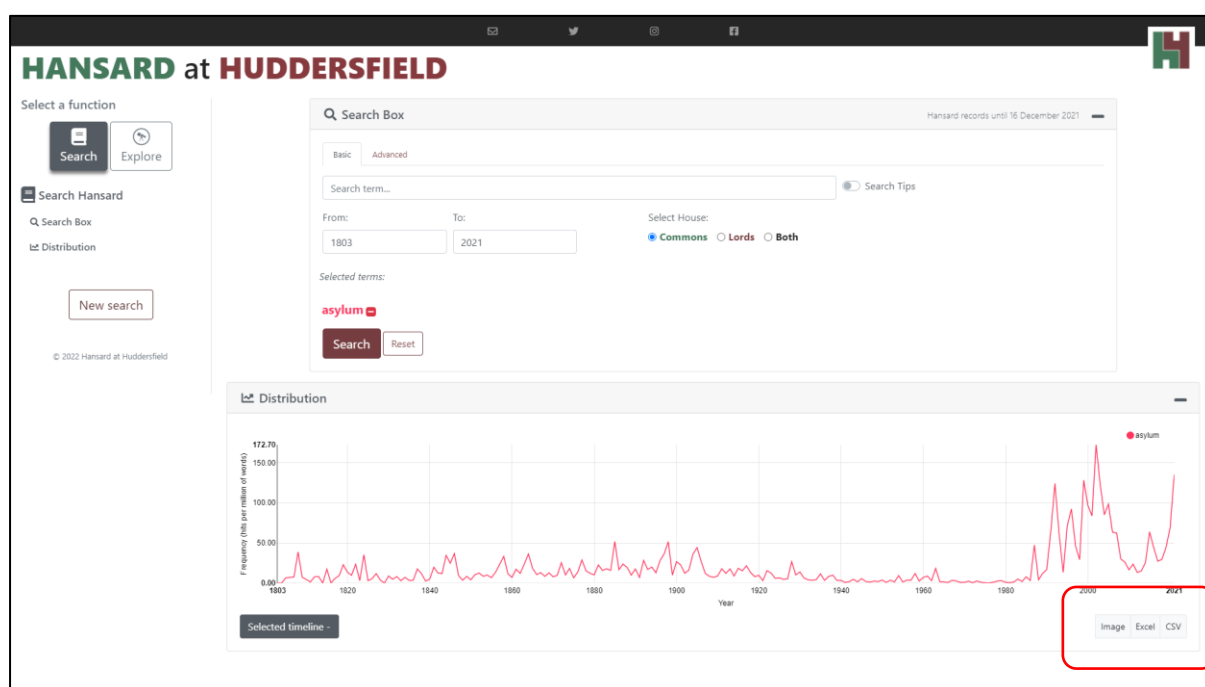
4. Reset/New Search

The 'Reset' button and the 'New Search' button will both delete any previously selected search terms and take all settings except for the dates back to their default.

This screenshot is similar to the first one, but with red boxes highlighting the 'New search' button in the sidebar and the 'Reset' button in the main search area.

DISTRIBUTION GRAPHS

After searching for your search term with the appropriate delimiters, you will see a frequency graph of the search term below the search box:



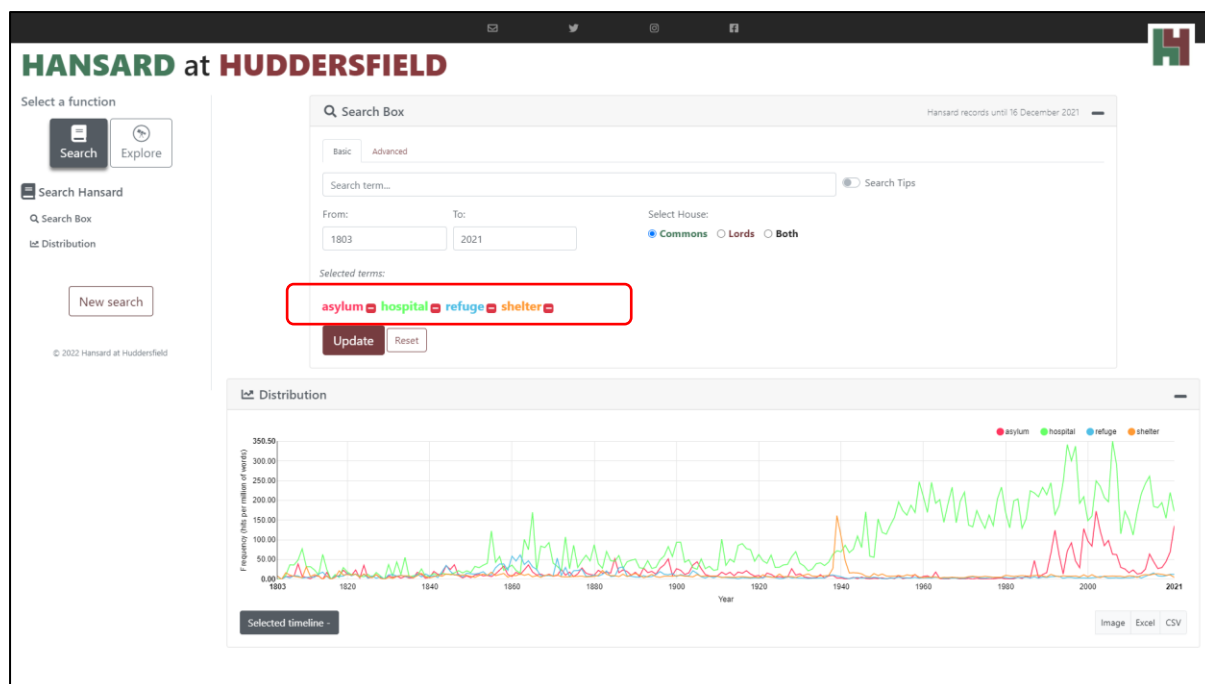
Distribution graph below Search Box

This graph is a representation of the number of occurrences of the search term *per million words*. This normalization of the data works to show comparative frequency of the term, but users should be wary of comparisons in the early part of the historical period when the overall number of words in Hansard was very much lower than it is today, meaning that an apparently high frequency might represent a rather low number of actual occurrences.

If you want to download the image for use in your own work, click on the 'Image' button in the right bottom corner of the graph (see screenshot above). This will initiate a download of the graph. Similarly, if you want to download the data in Excel or CSV format, click on the 'Excel' or 'CSV' buttons right next to the 'Image' button. **Please acknowledge Hansard at Huddersfield as your source and send us a copy of anything you publish (see paragraph 5.3 for more information)!**

COMPARISONS

You may want to compare search terms. This is as simple as typing another search term in the search box and clicking 'Update' (or hitting the enter key). You can compare up to four terms. These search terms will be listed just below the date boxes on the screen (see screenshot below). You will see that the distribution graphs of your search terms are overlaid on top of each other for your chosen date range and House(s) of Parliament. By clicking on the red minus symbols next to the search words you can delete search words from your distribution graph and overall search. If you just want to delete the distribution line of one of the words temporarily, click on the relevant word just above the graph itself.



Comparison of Search Terms

Remember that clicking on a related term to your search term using 'Related terms' will make your distribution graph look like a comparison graph too.

CONCORDANCE LINES

Clicking on the graph will bring up a list of concordance lines below the graph, which display the search term in its immediate context.

If you double-click on a single year, you will get the results for just that year. If you click on two different years, you will get the results for the whole period you have chosen. Here is the top of the list for the search term *poverty* in the year 1928:

<div> <div>Texts</div> <div>poverty 1928 - 1928</div> <div>195 contributions</div> <div>278 hits</div> </div>			
<div> <div>Download Selected</div> <div>Order by relevance</div> <div>Show debate title</div> <div>KWIC format</div> </div>			
	Date	Member	Contribution
<input type="checkbox"/>	1928-12-20	Mr John Tinker	... ries was trying to bring, about the reunion of particular Baptists with Anglo-Catholics... To-day after two and a-half years of dispute, had trade and inefficient organisation, there is the most horrible poverty and grave distress alike in South Wales and in Durham. I will make no comment on what the Government has done or left undone. The nation will pass judgment next summer. But the march of events h... [0 more]
<input type="checkbox"/>	1928-12-20	Mr. William Adamson	... tland, know the position to be quite as bad as it has been painted, and I ask the right hon. Gentleman not to place too much reliance on reports of the kind he has indicated because the destitution and poverty in our mining districts is undoubtedly very great. Those of us who are dealing with it day by day know the position better than it can possibly be known by any official who is only coming into touch... [0 more]

The number of contributions that your search word appears in is shown above the concordance lines with the word searched for and the period of concordance lines selected:

Texts	poverty 1928 - 1928	195 contributions
-------	---------------------	-------------------

Here, the date range is repeated (1928-1928) because only one year was selected. Note that the number of contributions is not the same as the number of occurrences or 'hits'. The term 'contribution' refers to the whole of any one speaker's turn and may range from a couple of sentences to a lengthy speech. As a result, a contribution may contain more than one hit.

By default, concordance lines show the first 10 contributions the search word appears in, but in reverse chronological order. Alongside the concordance lines, there are columns showing the date they were made and the MPs who made them. Optionally, you can also select a column showing the title of the debate in which the contribution occurred and/or order the contributions by relevance (which is a statistical measure of frequency of occurrence of the search terms within the contribution). You can also select to see more contributions listed, exclude metadata from the display, download selected contributions and click on MP's names for more information. Finally, you can choose to see your results in the Keyword in Context (KWIC) format (see next section). Here is what these options look like at the top of the concordance display:

The screenshot shows the search results for 'poverty' in 1928. The header indicates 195 contributions and 278 hits. Below the header, there are toggle switches for 'Order by relevance', 'Show debate title', and 'KWIC format'. A 'Download Selected' button is also present. The main table has columns for 'Date', 'Member', and 'Contribution'. A dropdown menu on the right allows selecting which columns to display: 'Date', 'Member', 'Contribution', and 'Debate title'.

	Date	Member	Contribution
<input type="checkbox"/>	1928-12-20	Mr John Tinker	... ries was trying to bring, about the reunion of particular Baptists wit after two and a-half years of dispute, had trade and inefficient organ horrible poverty and grave distress alike in South Wales and in Durha on what the Government has done or left undone. The nation will pa But the march of events h... [0 more]

Note that there is a drop-down menu on the right-hand side where you can turn the different columns on and off.

1. Concordance lines in document format or KWIC format

By default, the layout of the contributions containing the search term is arranged using the 'document format' (as above). This means that you will see these contributions listed chronologically with a large chunk of surrounding context on either side of the search term. Here contributions listed may contain more than one hit in each contribution. In order to see the whole contribution if it extends beyond the space in the table, you can click on it and the full contribution will display below the concordance list, with all the occurrences of the search term highlighted.

You can also change the display format to the KWIC format:

Texts **poverty** 1928 - 1928 195 contributions 278 hits

Download Selected ☐ Order by relevance ☐ Show debate title ☒ KWIC format 10 words ↻

	#...	Date ▼	Member ▲▼	Left Context	Hit	Right Context
<input type="checkbox"/>	1	1928-12-20	Mr John Tinker	had trade and inefficient organisation, there is the most horrible	poverty	and grave distress alike in South Wales and in Durham.
<input type="checkbox"/>	2	1928-12-20	Mr. William Adamson	of the kind he has indicated because the destitution and	poverty	in our mining districts is undoubtedly very great. Those of us
<input type="checkbox"/>	3	1928-12-20	Mr William Cove	the Avon Valley and see the misery and degradation and	poverty	that exists there. The fact of the matter is that
<input type="checkbox"/>	4	1928-12-20	Mr Thomas Shaw	they are in the condition of a continual struggle against	poverty,	which is breaking the hearts of some of the finest
<input type="checkbox"/>	5	1928-12-20	Mr William John	the Prime Minister, clearly indicates that there is distress and	poverty	in Scotland as well as in England and Wales, and

The term 'KWIC' is used to refer to concordance lines where the search term is placed centrally with its context listed either side and is an acronym for 'Key Word In Context'. If this option is selected, you will see concordance lines for each hit, rather than each contribution in your given search period. Note in this example (above), there are 195 contributions represented, but 278 hits.

You can change the number of words displayed on either side of the search word with the box next to the KWIC format button. Displaying results using a KWIC display may help you analyse the immediate context in which the search word appears in more detail and show how a word or phrase is being used at particular times or by specific (groups of) speakers.

Deselecting the KWIC format button will bring the display back to document format.

2. Showing debate title

By default, concordance lines will provide not only the contribution, but also the date this contribution was made and the MP who made it. To add the debate title to this display, tick the 'Show debate title' button:

Texts poverty 2008 - 2008 1191 contributions 278 hits

Download Selected ☐ Order by relevance ☒ Show debate title ☐ KWIC format

Unticking this button will hide the debate titles again. Debate titles can also be shown in both KWIC format and document format displays. Here are two lines in document format, with debate titles displayed in the right-hand column:

<input type="checkbox"/>	Date	Member	Contribution	Debate title
<input type="checkbox"/>	2008-12-18	Chris Bryant	... id that the hon. Member for East Dunbartonshire (Jo Swinson) talked a whole load of rubbish about tax credits. My experience is that one of the things that has rescued more people from poverty in some of the poorest families in the poorest communities in my constituency is the system of tax credits, which has also made it possible for many families to get into work. Yes, there have been difficult moments... [2 more]	Christmas Adjournment (Hansard, December 2008)
<input type="checkbox"/>	2008-12-18	David Amess	... there would never be another such disaster. Unfortunately, we all know what has happened. I was privileged to pilot the Warm Homes and Energy Conservation Act 2000, which should eliminate fuel poverty among vulnerable people by 2010 and for everyone by 2016. Recently, I was lobbied by Macmillan, which is concerned about the plight of cancer patients who need help with their heating bills because they ca... [0 more]	Christmas Adjournment (Hansard, December 2008)

3. Ordering by relevance

The default listing of concordance lines or contributions will be by date in reverse order. If you tick this box, the order will be by relevance, which is based on a calculation taking into account how often the query terms appear in the document, how close together the terms are in the document, and how important is the part of the document where they occur:

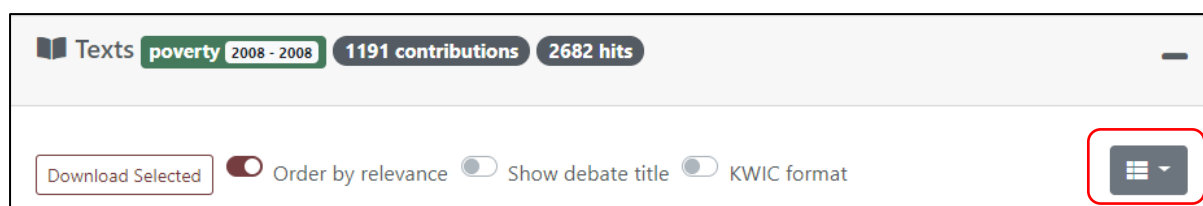
<input type="checkbox"/>	#...	Relevance	Date	Member	Left Context	Hit	Right Context
<input type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	congratulating the Scottish Affairs Committee on its valuable work on	poverty.	It is some years since I was a member of
<input type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	in opposition that led the assault on the scourge of	poverty	under the Tories; it was Labour that provided a detailed
<input type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	that provided a detailed analysis of the root causes of	poverty	under the Tories; and only Labour had the underlying values
<input type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	only Labour had the underlying values and ideology to address	poverty	in government and to make real progress in tackling poverty

The # column shows that these KWIC lines are all from the same contribution whose relevance is ranked number one. Relevance ranking for the document format doesn't need this column as it is the whole contribution that is ranked for relevance:

<input type="checkbox"/>	Relevance	Date	Member	Contribution
<input type="checkbox"/>	0.0981296	2008-05-01	Sandra Osborne	... I begin by congratulating the Scottish Affairs Committee on its valuable work on poverty . It is some years since I was a member of that Select Committee and a great deal of work has been done subsequently. We all know that the life chances of too many Scots have been strangled at birth, so it is timely to take a look at how far we have come, what progress has been made and where there is room for improve... [31 more]
<input type="checkbox"/>	0.098007	2008-03-13	Danny Alexander	... ampaing that he has done on the issue in his time in Parliament. The point I was trying to make is that this is the Budget in which the Government effectively abandoned their 2010 target of halving child poverty . It is worth reminding ourselves of a few of the facts about child poverty. In 2005-06, child poverty figures were still half a million higher than the target that had been set for 2004-05: th... [29 more]

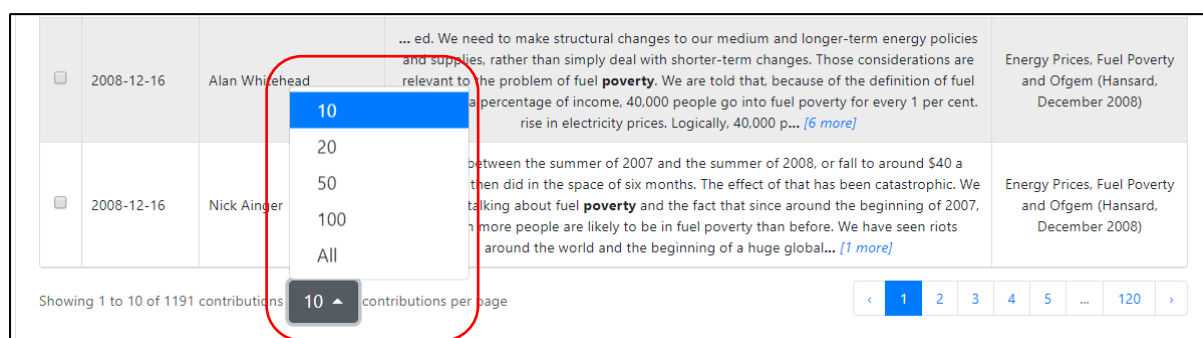
4. Excluding (meta)data from the display of concordance lines

To exclude (meta)data from the display, click on the drop-down menu with the 'list' icon in the top right-hand corner of the display and (de-)select any data you do not want displayed.



5. Viewing more contributions

By default, you will be shown the first 10 contributions that mention your search word. You can opt to increase the number of contributions shown by selecting a higher number of contributions displayed using the drop-down menu below the concordance lines:



6. Download concordance lines

While concordance lines give you a quick visual overview of the context in which your search word appears, downloading KWIC concordance lines into an Excel file may help you analyse the use of your search word better as it makes the concordance lines editable and even clearer. The latter may especially be helpful in case of a comparison between multiple search words.

To download KWIC concordance lines, select all the concordance lines you would like to appear in your Excel file. Clicking on the box at the top of the left-hand column selects all those contributions that are displayed on your screen at the time. If you want to download all concordance lines, you need to select 'All' in the drop-down menu at the bottom of the screen, otherwise you will only be downloading the default of all occurrences in the first 10 contributions, or whatever number you have previously chosen. Once you have selected the lines you want, click on 'Download Selected' button and choose between Excel and TSV format. Your file will be downloaded in the normal way.

Note that you can also download contributions from the document format. These will be delivered in a plain text format and need editing for further use.

Texts **poverty** 2008 - 2008 **1191 contributions** **2682 hits**

Download Selected ☒ Order by relevance ☐ Show debate title ☒ KWIC format 10 words ↻

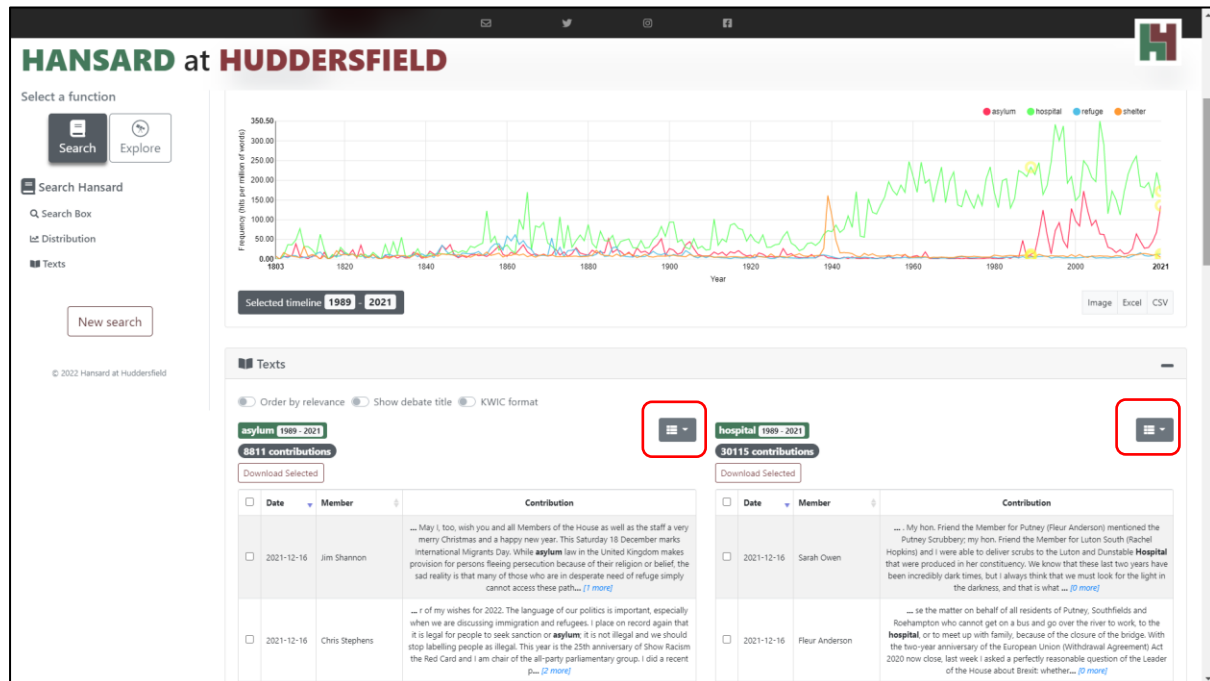
<input checked="" type="checkbox"/>	#...	Relevance ▼	Date ▲▼	Member ▲▼	Left Context	Hit	Right Context
<input checked="" type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	congratulating the Scottish Affairs Committee on its valuable work on	poverty.	It is some years since I was a member of
<input checked="" type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	in opposition that led the assault on the scourge of	poverty	under the Tories; it was Labour that provided a detailed
<input checked="" type="checkbox"/>	1	0.0981296	2008-05-01	Sandra Osborne	that provided a detailed analysis of the root causes of	poverty	under the Tories; and only Labour had the underlying values

7. Clicking on MP's name

If you click on the name of the MP whose contribution is being shown, you are linked through to a site that summarises their political career and links to all the contributions they have ever made to parliamentary debate recorded in Hansard.

CONCORDANCE LINES FOR COMPARISON SEARCHES

If you search for more than one search term and then click through to see the data, lists of concordance lines for each search term will be shown (see screenshot below) to allow easy comparison. You can change the displays for each concordance list individually as well as downloading them individually. If there are three or four terms in the search, the contributions will appear in two columns, with terms 3 and 4 below terms 1 and 2.



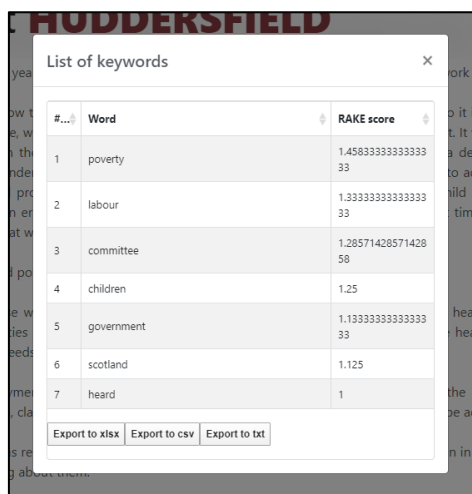
Concordance Lines for More than One Search Term

COMPLETE CONTRIBUTIONS

Clicking on a concordance line gives the full contribution behind that concordance line (see screen shot below). A close reading of its text will allow you to understand the contribution's key arguments, but you can also quantitatively find them using the "keywords" button right above the contribution (see the same screenshot).

This screenshot shows a detailed view of a contribution from the HANSARD at HUDDERSFIELD website. The top section displays the contribution's metadata: a checkbox, the number 0.0951392, the date 2002-04-24, and the member Mr. Shaun Woodward. Below this, a pagination bar shows 'Showing 1 to 10 of 627 contributions' and '10 contributions per page'. The main content area is titled 'Contribution' and includes a sub-section 'Analysis' with a 'Keywords' button highlighted by a red box. The text of the contribution follows, starting with '§ Mr. Shaun Woodward (St. Helens, South) The debate has been remarkable for the broad consensus evident on the matters that have been discussed. I suspect that this will not be the last asylum Bill to occupy my hon. Friend the Member for Walthamstow (Mr. Gerrard), as there will probably be another four such Bills in the next decade. I shall concentrate on children, and how they are to be protected. I ... [11 more]'. The text continues with several paragraphs discussing the asylum debate and the impact of the COVID-19 pandemic on refugees.

Clicking on this button will open a window superimposed on the screen, showing you the keywords of that document (contribution):



The screenshot shows a window titled 'List of keywords' with a close button (X) in the top right corner. The window contains a table with three columns: '#...', 'Word', and 'RAKE score'. The table lists seven keywords: poverty, labour, committee, children, government, scotland, and heard, each with a corresponding RAKE score. At the bottom of the window, there are three buttons: 'Export to xlsx', 'Export to csv', and 'Export to txt'.

#...	Word	RAKE score
1	poverty	1.4583333333333333
2	labour	1.3333333333333333
3	committee	1.2857142857142858
4	children	1.25
5	government	1.1333333333333333
6	scotland	1.125
7	heard	1

The hierarchy of the listing of keywords is determined by the RAKE score. RAKE is short for Rapid Automatic Keyword Extraction algorithm, an algorithm that determines keywords by analysing the frequency of occurrence and its co-occurrence with other words in the text. Unfortunately, the algorithm is only reliable when used on longer contributions, so these results should be interpreted with care. The length of lists of keywords may also vary depending on the length of the contribution.

You can download keyword lists including RAKE scores in either .xlsx, .csv or .txt formats by clicking on the relevant buttons at the bottom of the keyword list (see screenshot 16).

3.1.2 ADVANCED SEARCH

The Advanced Search screen allows you to do everything that the Basic Search Screen allows you to do, and more. Each contribution (from a short intervention to a long, prepared speech) has a range of identifiers built into the Hansard data. The Advanced Search screen will allow you to search along those identifiers: i.e. limit your search by debate title, by MP and by specific dates. It also allows you to display graphs that show the development of the usage of your search term per month rather than per year.

ADVANCED SEARCH OPTIONS

1. Specific date range

The advanced search page allows you to search for terms within a very specific date range. Please note that the dates are set out in YYYY-MM-DD order. Once you have a search term and a date range, you can click on the Search button (or press Enter) and you will see the distribution line of the search term. You can then explore the data itself in the same way as in the Basic search.

The screenshot shows the 'Search Box' interface. At the top right, it says 'Hansard records until 16 December 2021'. Below this are tabs for 'Basic' and 'Advanced'. The 'Basic' tab is active. There is a 'Search term...' input field. To its right is a 'Member...' input field. Below the 'Search term...' field are two date range fields: 'From:' with the value '2000-01-01' and 'To:' with the value '2021-12-16'. These two date fields are highlighted with a red rectangle. To the right of the date fields is a 'Debate title...' input field. Below the date fields are 'Search' and 'Reset' buttons. On the right side of the interface, there is a 'Search Tips' toggle switch, and a 'Select House:' section with three radio buttons: 'Commons' (selected), 'Lords', and 'Both'.

2. Limiting by MP

Typing the name of a particular MP will help you identify the contributions of a particular member of either House. It can be used on its own but will probably normally be used in conjunction with a search term and/or a description limiter. Please note that choosing non-matching limiters (such as dates when a member wasn't present) may lead to a long delay and a timed-out response.

Please also note that this function is only as good as the Hansard data behind it and that results for a specific MP might be skewed. Because Hansard has changed the notation of MP's names over time, sometimes you may not get all contributions from the MP you are interested in. Because some MPs have had the same or similar names, the results might also show contributions from multiple MPs. These issues might especially skew results for the period 1803-1907, but may impact results across the whole Hansard dataset.

This screenshot shows the same 'Search Box' interface as the first one. In this view, the 'Member...' input field is highlighted with a red rectangle. All other elements, including the 'Search term...' field, the date range fields ('From: 2000-01-01', 'To: 2021-12-16'), the 'Debate title...' field, the 'Search' and 'Reset' buttons, and the 'Select House:' options, are in the same state as in the first screenshot.

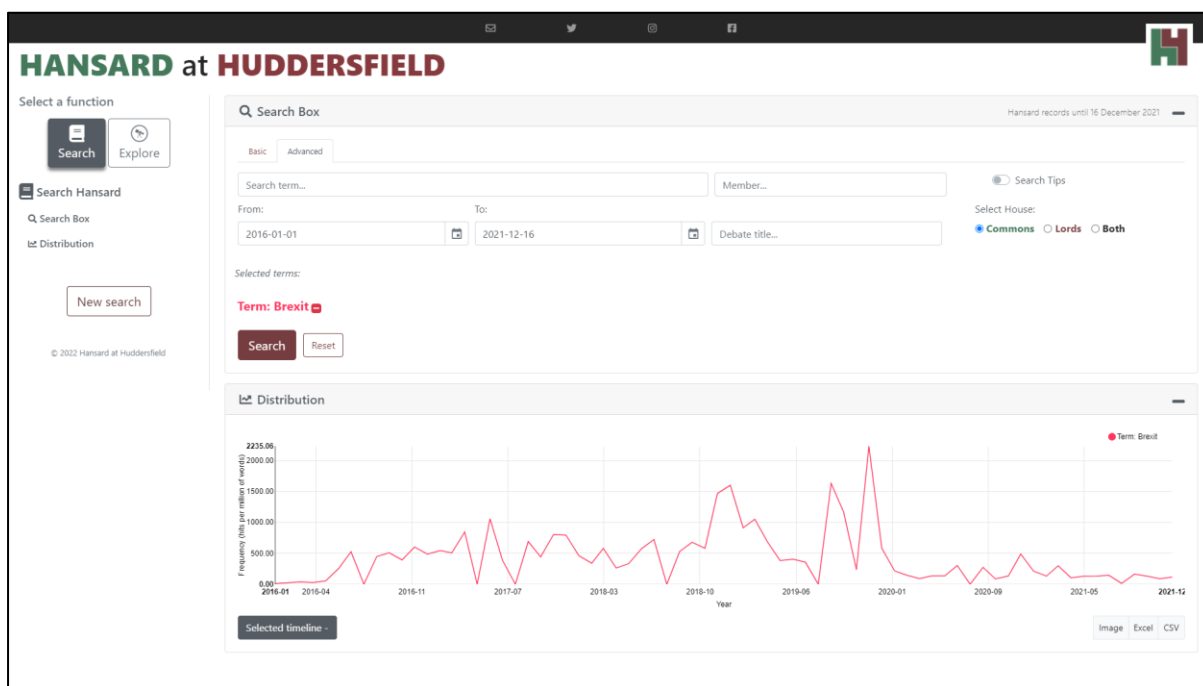
3. Limiting by debate title

If you know that you want to search only those items that relate to a particular topic or a particular Bill, this box can help you limit the search. It can be used on its own but will probably normally be used in conjunction with a search term and/or a member identifier. Please note that choosing non-matching limiters (such as dates when a member wasn't in office) may lead to a long delay and a timed-out response.

The screenshot shows the 'Search Box' interface. It has a 'Basic' tab selected and an 'Advanced' tab. The 'Search term...' field is empty. The 'Member...' field is empty. The 'From:' date is '2000-01-01' and the 'To:' date is '2021-12-16'. The 'Debate title...' field is empty and highlighted with a red rectangle. The 'Search' button is highlighted in red. The 'Reset' button is also visible. On the right, there is a 'Search Tips' toggle and a 'Select House:' section with radio buttons for 'Commons' (selected), 'Lords', and 'Both'. The top right corner indicates 'Hansard records until 16 December 2021'.

DISPLAY OF PATTERNS OVER MONTHS RATHER THAN YEARS

If you choose a search period of 5 years or less on the advanced search screen, the resulting graph will display the use of your search term by month rather than by whole year. The screenshot below shows an example of such a graph. It is also possible to compare up to four search terms for patterns of usage per month.



Monthly pattern

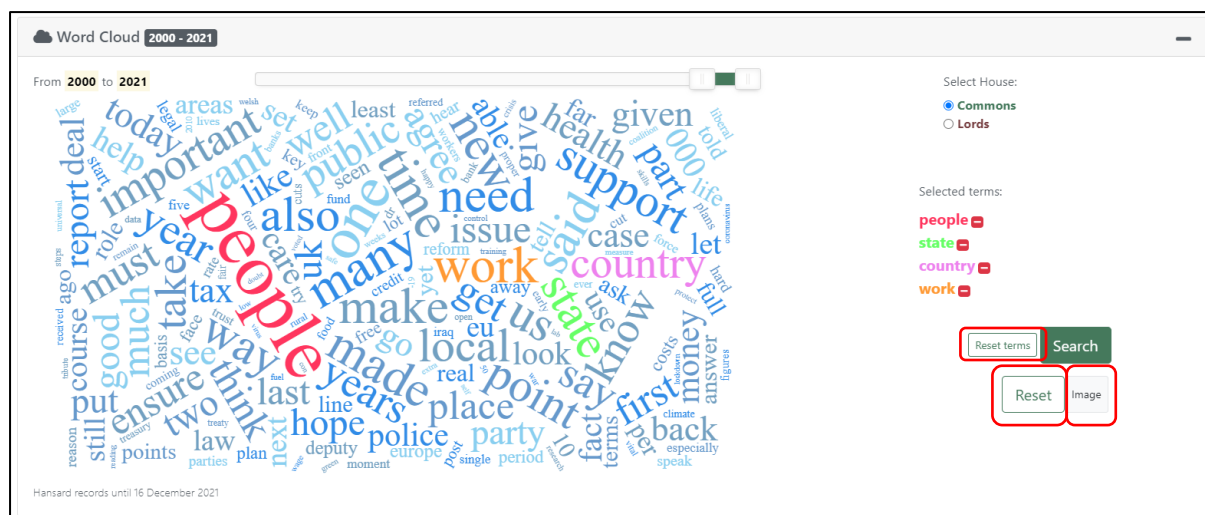
3.2 EXPLORE: SEARCHING WITHOUT PRE-DETERMINED SEARCH TERMS

Searching without having chosen pre-determined search terms is possible by word cloud and by keywords. You can access Hansard contributions through these word clouds and keyword displays.

You can select to display words for the House of Commons or the House of Lords. The option to search for the most frequent words in both houses is currently under construction.

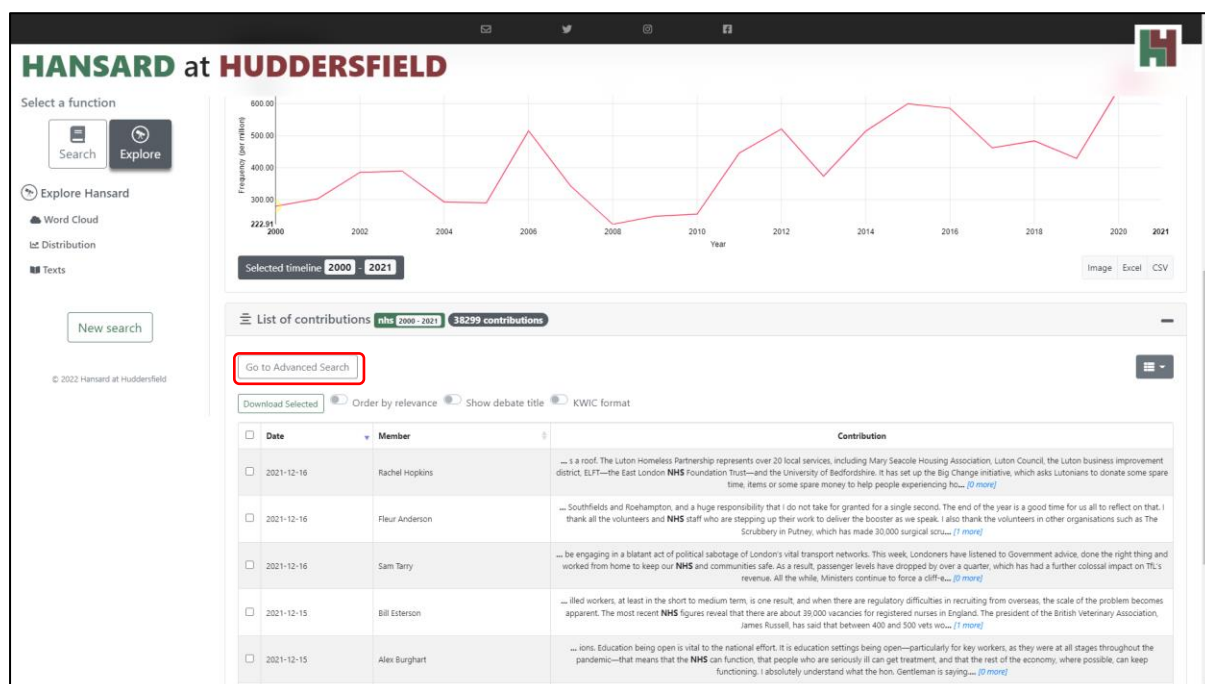


21



Note you can either reset the whole graph (with the 'reset' button next to 'image') or reset the individual terms you have selected (with the 'reset items') button.

From the concordance lines displayed after clicking through from a word cloud, you can also access the Advanced Search screen. To do so, click the button 'Go to Advanced Search' just below the 'List of contributions' bar (see screenshot):



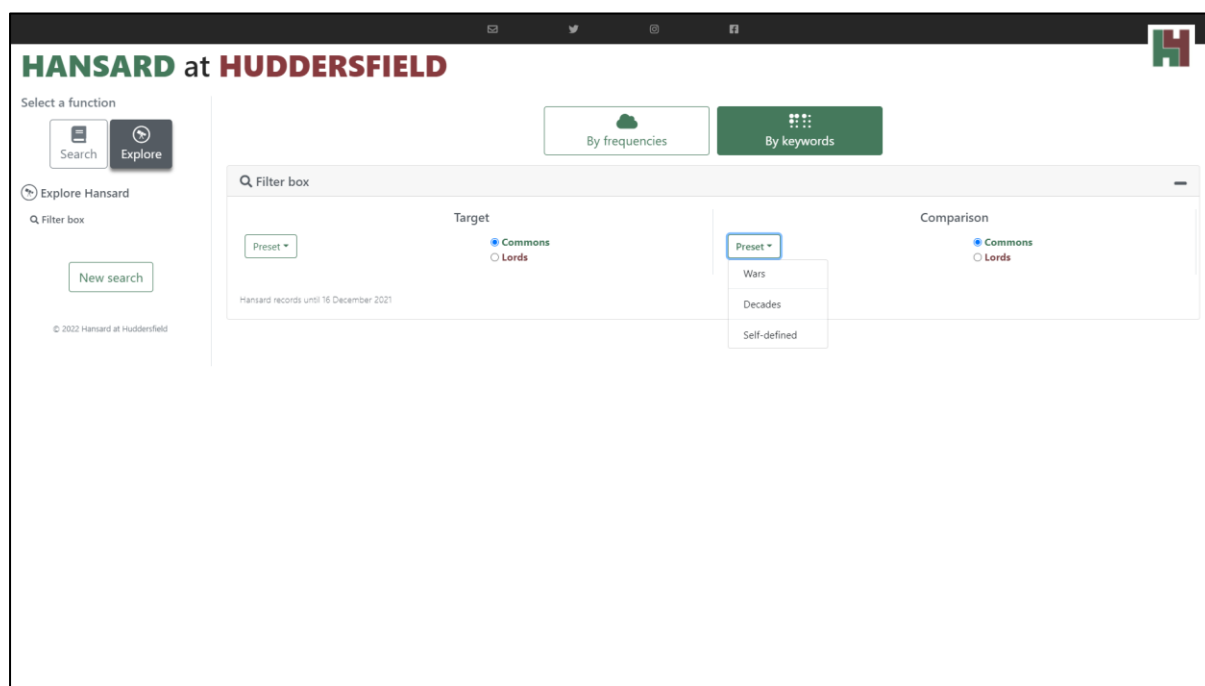
3.2.2 EXPLORE BY KEYWORDS

The keyword function allows you to see which words in one corpus (i.e. “body of text”) are key to that text compared with another corpus. Keywords are often used in corpus linguistics to answer questions like: “What is the body of text mainly about?”, “What are the major themes or concerns of the text?” and “What makes a text distinct from another text?”.

Keywords computed with this function are those words which occur relatively more frequently in one corpus compared to another corpus. To do this, we compare the corpus we are interested in knowing keywords for (the ‘target corpus’) with a ‘comparison’ corpus. The

algorithm behind the comparison then calculates which words are key by carrying out a statistical test called ‘Log-likelihood’. The higher the log-likelihood score, the more confident we can be that the frequency difference between the two corpora is not just by chance. The keyword function on our site only shows the keywords of which we are 99.99% certain this difference is not by chance.

The screenshot shows the first screen upon clicking on the “by keywords” button on the Explore screen. On the left hand side you select the target corpus you would like keywords computed for, while on the right hand side you select the corpus you want to compare the target corpus with.



Keyword Start Screen

Our site currently has three options for selecting your target and comparison corpora, which include self-defining both corpora and selecting pre-set corpora. It is also possible to self-define the target corpus and choose a pre-set comparison corpus or vice versa.

Bear in mind that the computation of keywords from self-defined corpora will take longer than the pre-set corpora² of wars and decades. The pre-set corpora have been calculated in advance, which reduces the search time, but the self-defined corpora require the algorithm to calculate the keywords afresh and this can take some time, depending on the size of the dataset chosen.

1. **Self-defined**

You can define your own target and comparison corpora using time period, member, house of parliament and search term parameters. Examples of corpora you may select include: a time period, a time period with a specific search term (i.e. a corpus that includes all contributions that mention “climate”) and a time period with a specific member (i.e. a corpus that includes all contributions by a specific member).

² We are working on increasing the variety of pre-set corpora for this function. We hope to include, for example, parliamentary sessions and/or governments in due course.

For limitations with selecting a corpus of all contributions by a specific member, please refer to section 3.1.2 Advanced Search > Advanced Search Options > Limiting by MP.

The screenshot shows a 'Filter box' with two main sections: 'Target' and 'Comparison'. Each section contains a 'Self-defined' dropdown menu, a 'Search term...' input field, 'From:' and 'To:' date pickers (both set to 2018-01-01 and 2021-12-16 respectively), and a 'Member...' input field. Below these are radio buttons for 'Commons' (selected) and 'Lords'. At the bottom of each section are 'Search' and 'Reset' buttons. A footer note states 'Hansard records until 16 December 2021'.

2. Wars

If you want to compare keywords for wars Britain fought between 1803-2019, we have preselected a number of important wars in this period. Your war corpus will contain all contributions by all MPs within the time period of the war³ for either the House of Commons, the House of Lord or both. Click on the buttons in front of the war names to select the wars and the buttons in front of 'Commons', 'Lords' or 'Both' to select the house of parliament. The keywords for the war periods have been precalculated and will therefore take slightly less time to compute.

If the war you want to create a corpus for has not been preselected by us, you can self-define the start- and end dates of the war with our self-define option.

This screenshot shows the 'Filter box' with the 'Wars' dropdown selected in both 'Target' and 'Comparison' sections. Below the dropdowns, a list of wars is displayed with radio buttons for selection. In the 'Target' section, the 'Independence' war is selected. In the 'Comparison' section, the 'War of 1812' is selected. The 'Commons' radio button remains selected in both sections. The 'Search' and 'Reset' buttons are still present at the bottom, along with the 'Hansard records until 16 December 2021' note.

3. Decades

All decades between 1803-2021 (excluding 2021) have been pre-set as corpora too.

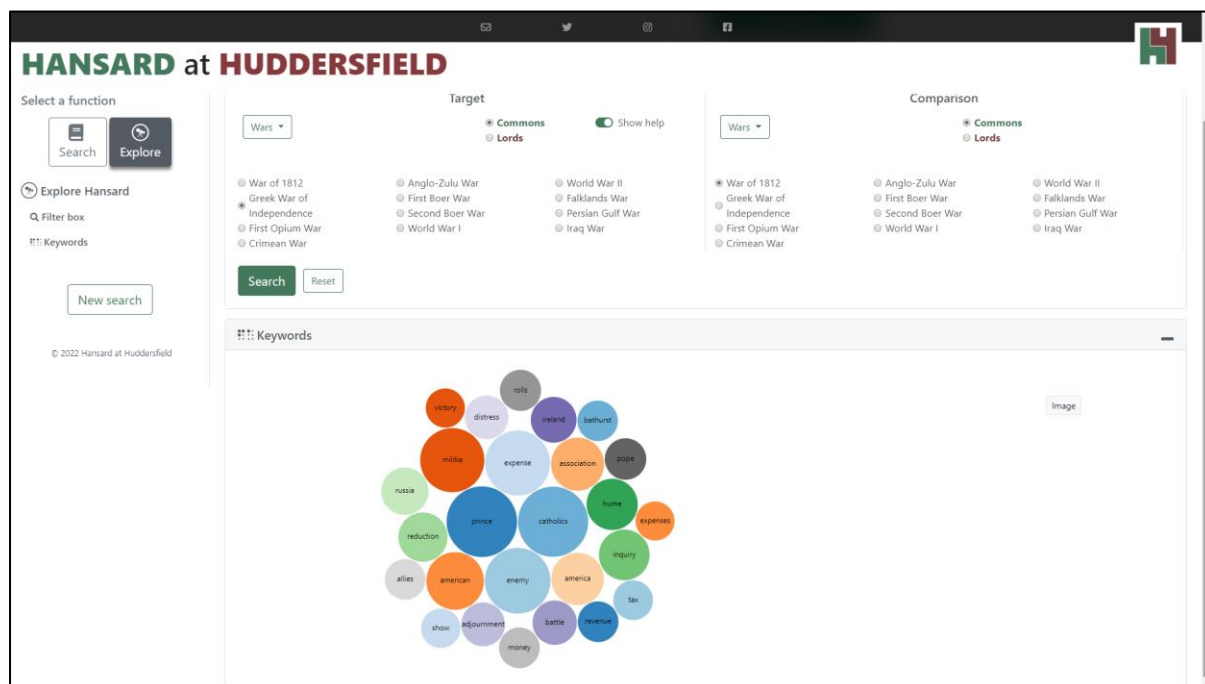
³ See Appendix I for a list of the start- and end dates used for each war.

Decades run from 1st January of the starting year to 31st December of the end year. Once you start typing the first digit of a decade, you should see a list with all decades that include that digit appear from which you can then select one. You can also choose for your “decade corpus” to include contributions from either House of Parliament or both houses. Keywords for the decades have been pre-calculated and should therefore not take long to compute.

The screenshot shows the 'Filter box' interface. It is divided into two main sections: 'Target' and 'Comparison'. Both sections have a 'Decades' dropdown menu. Below the dropdown in the 'Target' section is a text input field labeled 'List of decades...'. To the right of the input field are 'Search' and 'Reset' buttons. Below the buttons, it says 'Hansard records until 16 December 2021'. Both sections also have radio buttons for 'Commons' (selected) and 'Lords'. The 'Comparison' section has a dropdown menu showing a list of decades from 1901-1910 to 1971-1980.

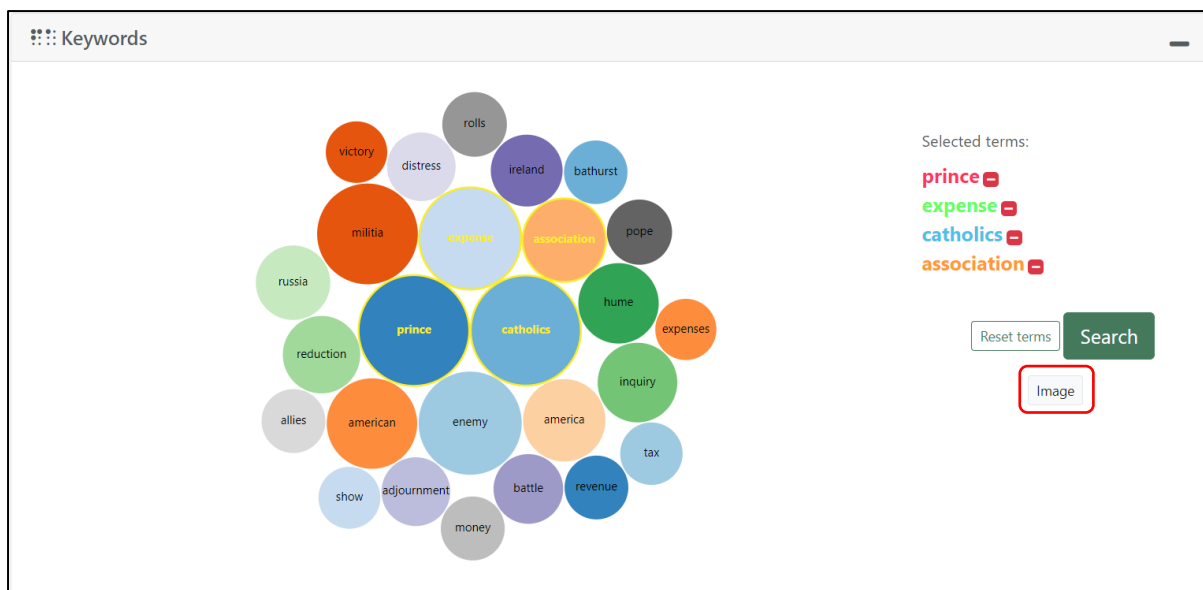
After you have selected a target corpus and a comparison corpus, you can click ‘Search’. Next to the ‘Search’ button you also find the ‘reset’ button that you may use if you want to reset the parameters selected for both of your corpora.

After you have clicked search, the keywords of the target corpus will be displayed in a bubble chart:



From your bubble chart you can select up to four items to search for them within your chosen target corpus. The four terms will be shown in concordance lines just as everywhere

else on the site. You can also download it in all the same ways. If you want to save a copy of the bubble chart on your computer (for example to compare bubble charts from different target corpora, or to include in a report or publication), click on the 'Image' button to the right-hand side of the screen (see screenshot).



As with the word cloud function, you can access the Advanced Search screen from the concordance lines (see screenshot):

The screenshot shows the 'HANSARD at HUDDERSFIELD' website. The top navigation bar includes the site name and a logo. Below the navigation bar, there are two main sections: 'Keywords' and 'Texts'. The 'Keywords' section features a bubble chart and a list of selected terms ('expense', 'catholics', 'association'). The 'Texts' section displays a table of concordance lines. The table has columns for 'Date', 'Member', and 'Contribution'. The first table is for 'prince' (1821-02-21 to 1829-09-12) with 174 contributions. The second table is for 'expense' (1821-02-21 to 1829-09-12) with 1417 contributions. Both tables have a 'Download Selects' button and a 'Go to Advanced Search' button (highlighted with a red box). The 'Contribution' column contains text excerpts, such as '... ght bring their force, would have a claim to rights of blockade; and we should be bound to recognize the legitimacy of the claim. He had also a strong objection to a ring-leader of rebels being styled the **prince** regent of Portugal. What was the meaning of those words? Did they not imply, that the person of whom they were used was exercising legitimate authority? He did not say, that jurists would construe... [2 more]'. The second table shows a similar excerpt for 'expense'.

4. DISCLAIMER

All the Hansard data behind this website has been generously made available to us by the Parliamentary Digital Service. Any errors or omissions that you may encounter in the text of Hansard as you explore it through our website are also there in the original Hansard proceedings parliament holds.

The historical content (1803-2005) of Hansard may exhibit some character and structure errors. These errors are a by-product of the digitisation process undertaken to make this content available online; a process that consisted of scanning physical copies of bound Hansard volumes and afterwards converting these scans into an online format suited to web processing.

Furthermore, a few bound volumes from the early 1800s have not made it onto the online version of Hansard yet. While the Parliamentary Digital Service looks for ways of solving that, please note that this may have a skewing effect on the frequencies of search words as well as distribution graphs.

It is also helpful to keep in mind that Hansard is “substantially verbatim” record of parliamentary proceedings. That means that it literally records everything a member says, but with repetitions, hesitations and obvious mistakes edited out. This also means that there are only minimal references to non-verbal interactions between members, such as “the member rose” or “laughter”. For processing ease these references have been taken out of Hansard on our site completely. The effect is that it makes Hansard much more readable, but it loses some explanatory information about the way some contributions may be worded in response to previous contributions or other occasions prior to the contribution. We recommend cross-checking with [Hansard.parliament.uk](https://hansard.parliament.uk) should you wish to do so.

Furthermore, the Historic Hansard data up until 1909 was not even always substantially verbatim. Before that, reports were often less comprehensive and often drawn together from press reports. This means that non-appearance of a particular search term in 19th century Hansard cannot be taken to mean it was not used in debate. This information and more about the History of Hansard can be found in a publication by Hansard reporter Stephen Farrell and Hansard Editor John Vice: <https://www.parliament.uk/documents/lords-library/History-of-Hansard.pdf>.

5. NOTES

PERFORMANCE ISSUES

The performance of searches can be very slow when doing a multi-word search over a long period of time. One way to improve performance is to shorten the period of time.

Loading of the query behind “Related terms” on the Basic Search screen and “Disambiguation” in the Advanced Search screen can take a long time due to the processing required behind the search in the database.

COMMENTS/SUGGESTIONS/BUG FIXES

Many revisions we have made on our site have been the result of feedback we received from users. Have you found any bugs, suggestions for improving search functions or any other feedback on the website? You can contact us by either emailing us at hansard@hud.ac.uk or by using the form on the site under “Feedback”. We will do our best to respond to your feedback and get back to you as soon as we can.

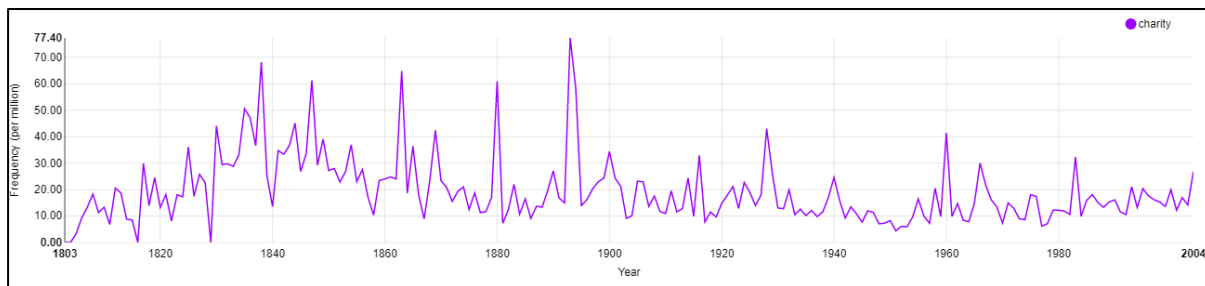
CITING/REFERENCING HANSARD AT HUDDERSFIELD

All works which use or refer to this site should acknowledge the source by means of bibliographic citation. If you have used it in your work, please let us have details and (if possible) copies of any internal or published work drawing on the website. We are also

particularly interested in hearing from you how you have used the site and for what purpose. You can contact us on hansard@hud.ac.uk or use the feedback form on the site under “Feedback”.

If you would like to use a **graph** that you have downloaded from our site, please use the following citation format: Author (Date). “Search word, years” [Graph]. Publisher. Available online at: Website Title.

Example:



Hansard at Huddersfield (2020). “Charity, 1803-2004” [Graph]. University of Huddersfield. Available from: <https://hansard.hud.ac.uk>.

When citing the **entire** website in-text, please give the whole address of our site: <https://hansard.hud.ac.uk>.

Example:

Hansard at Huddersfield is a wonderful interactive site to study the language of parliament (<https://hansard.hud.ac.uk>).

The citation format for footnotes or the reference section of publications is: Author, (Date). Title. Publisher. Available online at: Website title.

Example:

Hansard at Huddersfield Project (2018). ‘Hansard at Huddersfield’. University of Huddersfield. Available online at: <https://hansard.hud.ac.uk>

6. ACKNOWLEDGEMENTS

We are very grateful to the Arts and Humanities Research Council for funding the project (grant reference AH/R007136/1). We would also like to extend thanks to the following end-users who commented on earlier versions of our site:

[Campaign for Nuclear Disarmament](#), [Centre for Disability Studies](#), [Democratic Audit](#), [Equality and Human Rights Commission](#), [Hansard](#) (both Lords and Commons), Dr Gavin Hart, [Heritage Quay](#), [History of Parliament](#), [House of Commons Library](#), [Institute for Government](#), [Kirklees Democracy Commission](#), Dr. Thomas Loughran, [LKMco](#), [Migration Yorkshire](#), Dr Andy Mycock, [People’s History Museum](#), [Parliamentary Digital Service](#), [Parliamentary Office of Science and Technology](#), [The Green Party](#), [Theos](#), [TheyWorkForYou](#), [The Young Foundation](#), Andrew Gray and [Wikidata British Politicians](#).

APPENDIX I

The start- and end dates were chosen based on British involvements in the wars starting or ending.

What	Start	Finish	Reason start	Reason Finish
War of 1812	18.06.1812	30.12.1814	American declaration of war by president James Madison	Treaty of Ghent signed into law by the British Prince Regent
Greek War of Independence	21.02.1821	12.09.1829	Start of revolts of Filiki Eteria against the Ottoman Empire	Battle of Petra, last major engagement of war with Ottomans surrendering
First Opium War	04.07.1839	29.08.1842	Battle of Kowloon, first armed conflict of the First Opium War	Signing of the Treaty of Nanking
Crimean War	28.03.1854	30.03.1856	British Declaration of War	Treaty of Paris signed
Anglo-Zulu War	11.01.1879	04.07.1879	High Commissioner of Southern Africa Bartle Frere deems state of war to exist. British invade Zululand.	Zulu armies defeated by the British
First Boer War	20.12.1880	23.03.1881	First skirmish between Transvaal Boers and British troops in Potchefstroom	Peace Treaty signed
Second Boer War	11.10.1899	31.05.1902	British Declaration of War	Treaty of Vereeniging
World War I	04.08.1914	11.11.1918 OR 28.06.1919	British declaration of War	Armistice on 11/11 OR Treaty of Versailles

World War II	01.09.1939	08.05.1945	German invasion of Poland	Nazi Germany formally surrenders
Falklands War	02.04.1982	14.06.1982	Argentinian invasion of the Falklands	Argentinian surrender to the British
Persian Gulf War	02.08.1990	28.02.1991	Operation Desert Shield after Iraqi invasion of British-occupied Kuwait	Operation Desert Storm
Iraq War	19.03.2003	22.05.2011	Invasion of Iraq by US, UK, Australian and Polish troops	Withdrawal of last British remaining forces