Institution: University of Huddersfield
Unit of Assessment: 26 Modern Languages and Linguistics
Title of case study: Hansard at Huddersfield: Changing the Nature of Engagement with the Parliamentary Record Through an Online Interactive Database
Period when the underpinning research was undertaken: 2003 – 2019
Details of staff conducting the underpinning research from the submitting unit:

Names:	Roles:	Periods employed by submitting HEI:
1. Lesley Jeffries 2. Brian Walker 3. Matthew Evans 4. Alexander von Lunen	1. Professor 2. Research Assistant; Research Fellow; Senior Lecturer 3. Research Assistant; Senior Lecturer 4. Senior Lecturer	1. September 1990 – present 2. August 2012 – December 2015 3. January 2018 – July 2018; September 2018 – present 4. December 2014 - present

Period when the claimed impact occurred: March 2018 – 31 December 2020
Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

Hansard, the quasi-verbatim record of parliament, has been available online since 1997 but until the advent of the Hansard at Huddersfield project in 2018, its searchability was limited and the user-friendliness of its interface low. Jeffries and her team employed corpus linguistic techniques to create an entirely new web-based interface, known as Hansard at Huddersfield, for exploring the database. This is a user-friendly interface with a range of search functions and visualisations. It provides Hansard users (from students to parliamentary researchers) with an improved method of exploring political decision-making. Hansard at Huddersfield has changed the way in which the producers of Hansard perceive their data; influenced plans for Hansard's electronic future being made by the Parliamentary Digital Service; improved the way in which Hansard users access the data; and informed the work of a group planning new digital versions of parliamentary data outside the UK, in the Hellenic parliament of Greece.

2. Underpinning research

Hansard is a written record of what is said in the UK parliament. Parliament's electronic version comprises a standard database of texts, with metadata detailing information such as dates, debate titles and speakers. This kind of database and metadata searches are familiar to internet users. However, users are limited in what they can search for in the data itself, as they are restricted to searching for individual occurrences of words with no opportunity to identify patterns of usage. Jeffries and colleagues at Huddersfield set out to address this problem by drawing on their track-record of research in corpus linguistics, using a range of tools that enable the user to search across large bodies of textual data like Hansard, to see where words, phrases and topics occur in the database and identify patterns of usage. Their research combines this approach with qualitative interpretation of the results of searches. This kind of searchability, based on the patterning of words in texts, provides a new way of conceptualising the language of parliament, allowing patterns of language use and fluctuating levels of interest in topics to be discovered with relative ease and to be interpreted in context.

Since 2003, Jeffries, Director of the Institute of Applied Linguistics at the University of Huddersfield (UoH), has used corpus-based techniques to demonstrate the changing meanings of words in specific socio-political contexts. Her findings include tracking the developing meaning of a simple word (*water*) in the reporting of a time of drought **[3.1]** and this work fed into the concept of socio-political keywords developed by Jeffries and Walker (2012) **[3.2]**; (2017) **[3.4]** and (2019) **[3.5]** in their work on the language of the Blair years, where they proposed solutions to the methodological difficulties of bridging the gap between automated and manual analysis. Their approach combines automated search techniques such as finding the keywords in a dataset and producing concordance lines (lists of word occurrences within their context) for each keyword with Jeffries' critical linguistic framework for systematising the analysis of each occurrence. Their results demonstrated the changing meaning of a small set of keywords (*choice*; *reform*; *respect*; *terror*; *global* and *spin*) in the reporting of the Blair years. The approach was adopted by Jeffries and

Evans in 2015 **[3.3]**, who investigated a small but complete dataset of election manifestos. These projects produced key insights into the danger of misinterpreting automated results and the advantage of using a linguistic framework such as that developed by Jeffries to identify the textually-constructed meanings of words and their potential for ideological impact.

Jeffries was invited to lead the Hansard strand of the SAMUELS project led by the University of Glasgow from 2014-2015, funded by the Arts and Humanities Research Council (AHRC) (£67,325.89). In 2018, Jeffries was successful in an application for AHRC follow-on funding (£80,510) to apply corpus linguistic techniques to the development of the Hansard at Huddersfield interactive web-based tool **[3.6]**, with the aim of presenting a user-friendly and efficient way to explore the data in Hansard from a textual point of view while providing guidance to users about the interpretation of findings. In addition to Jeffries, the core team comprised UoH Historian and Digital Humanities scholar Alex von Lunen, who applied his expertise in visualisation and computing to develop the project's overall approach; UoH Research Fellow Hugo Sanjurjo-Gonzalez (2018-2020), the programmer who combined database approaches with corpus methods and visualisations to produce the website and Research Assistant Fransina Stradling (2018-present) who worked with Jeffries on the public-facing aspects of the project. The resulting interface followed detailed work by Jeffries and the team on the structure and nature of the Hansard data and experimentation with corpus linguistic tools and visualisations to provide the best combination of these whilst ensuring that the raw data was always accessible to the user. The site provides more sophisticated searchability that end-users of Hansard can use with no training in corpus methods and, through provision of accessible online advice on the interpretation of results, without the need for extensive knowledge of linguistics. The website and supporting materials were developed in 2018, launched in 2019 **[3.6]** and further improved throughout 2020. The website provides unprecedented means for interested end-users to carry out detailed studies of the language used in the UK Parliament by observing patterns in word usage, initially through graphs and other visualisations but with the flexibility of always being able to access the data itself. Users can locate the discussion of topics through key terms and/or consider any emergent meaning of vocabulary based on frequency of use. They can also use the site's supporting materials to help them understand how to interpret the automated findings, avoiding the pitfalls of automation previously identified by Jeffries and colleagues **[3.2-3.5]**.

3. References to the research

<u>Evidence of the quality of the research:</u> Articles **[3.1]**, **[3.2]** and **[3.3]** were submitted to international high-ranking journals and went through a rigorous anonymous peer-review process before being accepted for publication. The book **[3.4]** was accepted for publication after anonymous peer-review and scrutiny by the publishing house. The book chapter **[3.5]** was submitted as a paper to an international conference where abstracts were peer-reviewed and the authors were invited to submit it to the edited collection where it was anonymously peer-reviewed a second time before acceptance. The website **[3.6]** is the output from the AHRC-funded project for which the proposal for funding (£80,510) was rigorously reviewed following the AHRC's internal procedures.

[3.1] Jeffries, L. (2003) 'Not a drop to drink: emerging meanings in local newspaper reporting of the 1995 water crisis in Yorkshire ', *Text* 23(4): 513-38. https://doi.org/10.1515/text.2003.020 [can be supplied on request]

[3.2] Jeffries, L. and Walker, B. (2012) 'Key words in the press: a critical corpus-driven analysis of ideology in the Blair years (1998-2007)', *English Text Construction* 5(2): 208- 29. https://doi.org/10.1075/etc.5.2.03jef [can be supplied on request]

[3.3] Jeffries, L. and Evans, M. (2015) 'The rise of choice as an absolute 'good': a study of British manifestos, 1900-2010', *Language and Politics* 14(6): 751-77. https://doi.org/10.1075/jlp.14.6.02eva [can be supplied on request]

[3.4] Jeffries, L. and Walker, B. (2017) *Keywords in the Press: The New Labour Years.* London: Bloomsbury. https://www.bloomsbury.com/uk/keywords-in-the-press-the-new-labour-years-9781350046269/ [can be supplied on request]

[3.5] Jeffries, L. and Walker, B. (2019) 'Austerity in the Commons: a corpus critical analysis of austerity and its surrounding grammatical context in Hansard (1803– 2015)', in Power, K. Ali, T. and Lebdušková, E. (eds) *Discourse Analysis and Austerity: Critical Studies from Economics and Linguistics*, pp. 53-79. Abingdon: Routledge.
https://www.taylorfrancis.com/books/e/9781315208190/chapters/10.4324/9781315208190-6
[can be supplied on request]

[3.6] Jeffries, L., von Lunen, A., Sanjurjo González, H. and Stradling, F. (2019) *Hansard at Huddersfield*. https://hansard.hud.ac.uk/site/index.php

4. Details of the impact

Existing searchable interfaces for Hansard provide only a limited range of options for the user to search by speaker, debate title, date and word to find individual contributions to parliamentary debate. Some interfaces provide basic graph presentations of the results, but do not have direct clickable links to the data or allow general identification of patterns of language use. Hansard at Huddersfield addresses directly this problem of searchability.

Routes to Impact The development of the Hansard at Huddersfield interface involved extensive engagement with end-user groups. In 2018, two meetings in Huddersfield and two meetings in London involved consultation with 41 people from local governments (Leeds and Kirklees), local and national historical societies, political organisations, and parliamentary organisations including History of Parliament, Hansard reporters and editors, parliamentary IT staff, Parliamentary Office of Science and Technology, Education Policy Think Tank and non-linguist academics (historians and political scientists). Development of the site also involved bimonthly meetings and discussions between 2018-2020 with Hansard editors and later with the Parliamentary Digital Service (PDS) (2020). In March 2019, Jeffries launched the website at events in London and Huddersfield. The House of Lords event introduced the main features of the site and included a Q&A reception with over 40 attendees from History of Parliament, Commons Library, Institute for Government, Wikidata, Overseas Development Institute, political scientists, historians and members of both Houses of Parliament. The University of Huddersfield launch event included linguists, historians, librarians, archivists, and representatives of Leeds City Council and the Equality and Human Rights Commission. Between September 2019 and March 2020, five bespoke workshops demonstrating use of the site and training were provided for the British Library; POLIS thinktank in Leeds; Greenhead Sixth Form College in Huddersfield; The People's History Museum in Manchester, and the House of Commons Library, collectively reaching 67 participants. Throughout this period, Jeffries also engaged in several consultations with other parliaments' record-makers, including a round-table discussion with members of the Hellenic OCR team of the Greek parliamentary data service and an email exchange with the editor of the Australian parliament's Hansard. The Hansard team also maintains an active social media presence to engage a wider audience, with a Twitter following of over 660 people and organisations including MPs (e.g. Alison McGovern, Layla Moran and Jason McCartney); Peers (e.g. Lord Norton of Louth); parliamentary groups (e.g. Lords EU Committee Legal Advisor: Alexander Horne; Procedure Committee Parliament); research organisations (e.g. Study of Parliament Group; Political Studies Association sub-groups); think tanks (e.g. Rethink; Politaeia; Democratic Audit) and educational groups (e.g. Citizens Project; Westminster Skills Centre). These diverse pathways have resulted in changing the way parliamentary producers of Hansard perceive their data; influenced the PDS' plans for the electronic future of Hansard; improved the way in which Hansard users access data; and informed the work of a group in Greece creating digital versions of their own parliamentary data, as outlined below.

Changing the way in which the producers of Hansard perceive their data The principal impact of this project has been to change perceptions and behaviour in parliament among Hansard staff who produce the raw data, including the editors of Hansard (Lords) and Hansard (Commons). The Editor of Hansard (Lords) writes: "*The main impact that Lesley's work has had on me and my team is in relation to the way in which we perceive what we publish* […] *Her website* **[3.6]** *has demonstrated to my satisfaction that the types of search and visualisation available there can enhance the scrutiny of parliamentary debate by allowing users to find patterns of topic and linguistic usage that would not be accessible in the other online versions of official Hansard* [...] *The impact on our team's daily working has been to sharpen our awareness of the editing*

principles we apply in our work; any word can be important to researchers interested in parliamentary language and how we apply our terms of reference could skew their research if we edit inappropriately" **[5.1]**.

As a result of this change of perception and conviction of the value of the work, the Hansard (Lords) Editor initiated planning for new digital versions of the data: "*One of the outcomes of our collaboration with Lesley's team has been that I initiated a series of discussions with the Parliamentary Digital Service* [PDS]*, which is responsible for maintaining, among much else, the electronic version of Hansard for the public. We wanted to find a way to extend the project in order to maximise the potential for it to become sustainable* […] *The result of these discussions is that* […] [PDS] *has funded an extension to the Hansard at Huddersfield project* [November 2020] *to the tune of £44,312.26 (for one year in the first instance), the ultimate purpose of which is to work towards some kind of integration of the Hansard at Huddersfield approach to searchability* **[3.6]** *into the official site*" **[5.1]**.

Influencing plans for Hansard's electronic future As a result of this engagement with Hansard, Jeffries' work has also influenced the perceptions of the staff at PDS responsible for Hansard's digital presentation and for delivering parliament's Open Data policy. Jeffries and von Lunen have engaged in four meetings with staff of PDS, including the Director of Data & Search Transformation and the Head of Open Data. These meetings raised awareness of the problems inherent in the PDS configuration of the data for incorporating the Huddersfield approach: "...*the work of Lesley Jeffries and her team at Huddersfield University*... [is] *providing additional and beneficial search capability over the Hansard Data and* [their work is] *mitigating some of the issues that had resulted from the status of the* [existing] *Data platform*". Drawing on the Hansard research, "*a business case was produced and approved ... to extend it* [the collaboration between Parliament and Hansard at Huddersfield] *to derive learnings that will be fed into the strategic data platform.*" Through the collaboration with Huddersfield, *"We expect to better understand impacts of poor data quality or poor linkage, acquire knowledge of search requirements of external users and document requirements that will improve our taxonomies to feed into the target solution as we collaborate at the same time as existing Parliamentary users continue to get benefit from the service".* **[5.2]**.

Improving the way in which Hansard users access the data Mindful of the need to interpret the results of searches on their site with care, the Hansard at Huddersfield team produced materials specifically designed to develop the understanding and usage of the site by historical, political and policy researchers. Researchers in history and politics fields have been enthusiastic in taking up what the site has to offer. For example, a historian noted how the site's innovative use of corpus linguistics helps historians and the public exploit the Hansard database better: "*The* [Hansard at Huddersfield] *project is attempting to address the issue of inaccurate or inappropriate search results and thus reduce the likelihood that the end user's research will be skewed by an inability to understand or properly exploit the data available*" **[5.3]**. Other historians have utilised visualisations generated from the site to communicate their research with wider, non-academic audiences. For example, graphs generated by Hansard at Huddersfield have been used in one historian's public presentation "Remembering Peterloo: protest, satire and reform", hosted at the Palace of Westminster by the History of Parliament Trust, The Parliamentary Archives and the Citizens Project in July 2019 **[5.4]** and by another historian in a 2019 blog on the historic debtor sanctuaries of London **[5.5]**. Similarly, political researchers have used the site to make claims about the content of Hansard contributions, for example by using a Hansard at Huddersfield distribution graph that showed a nearly six-fold increase in relative frequency of the words *constituency* or *constituent(s)* in Commons Hansard between 1950 and 2019 as a "high-quality proxy" for determining MPs' constituency focus **[5.6]**. This author comments on the possibilities of advanced analysis enabled by the site, writing: "*Hansard At Huddersfield uses wordclouds to display the discussions of Parliament as a whole: this could be applied to MPs as individuals in order to convey information such as their 'focus' without the use of dubiously precise statistics"*; and they acknowledge the site's appeal in making parliamentary data accessible: *"Thanks are also due to* […] *Hansard At Huddersfield for providing easy-to-use access to Parliamentary data*" **[5.6]**. In addition, the take-up of the site by non-HE-based researchers has been encouraging, with the site being listed as a resource by the Parliamentary Archives and the History of Parliament Trust **[5.7]**.

There has been a similarly positive response to the site from sixth-form students, particularly in relation to their individual projects for the Extended Project Qualification (EPQ). The Head of EPQ at Greenhead College, Huddersfield, notes the value of being able to search across a dataset and not just find single moments/speakers/debates: "*Students working on topics where the political dimension is relevant have found the website a very useful way into the data and the visualisations provided by the site have been used by at least one student in their final reports as well*" **[5.8]**. The site has also impacted the behaviour of teachers who introduced to their students and listed it as a resource: "*We have made certain that the EPQ students are aware of this resource by showing students how to use it in lessons and putting a link on the EPQ moodle* [Virtual Learning Environment] *page direct to Hansard at Huddersfield*" **[5.8]**.

Feedback from users on social media also demonstrates users' appreciation of the new possibilities opened up by the site:

- @edwardwood99: "Great presentation by @HansardHuds showing off their brilliant search tool in @UKParliament. All sorts of possibilities for linguistic analysis"
- @TheVictCommons: "Congratulations @HansardHuds – it was great to see how your website has developed since we first saw it & we look forward to exploring how we can use it in our @HistParl research"
- @generalising: "Back from the launch of @HansardHuds. Really promising site and well worth a play – some very fun ways to get into the data compared to the old Historic Hansard." **[5.9]**

Informing the work of a group in Greece working on parliamentary data Whilst not all parliaments have data to match the extent of the UK Hansard, there are the beginnings of digitisation around the world. In Greece, for example, Jeffries' approach has informed the early stages of an open data project using a crowdsourcing approach to the digitisation and processing of parliamentary data. The Hellenic Optical Character Recognition (OCR) Team is acknowledged as an interoperability solution by EU's Joinup.eu, as well as an innovative practice by the Inter-Parliamentary Union, and currently counts 40 members in 10 countries. The Leader of the Team at the Hellenic Government of Greece writes: "*In Hansard at Huddersfield, we were able to identify a competent and reliable partner.* [...] *Through talks, we agreed to the exchange of knowhow, methods and code to jointly tackle common challenges that are related not only to data openness but also to increased visibility and understanding of parliamentary data sources. Our frequent exchange with Hansard at Huddersfield has certainly sparked further innovation within our team and it is not exaggerated to say that the Huddersfield approach has influenced the design and development of Xtralingua, our latest open tool for textual analysis*" **[5.10]**.

5. Sources to corroborate the impact

[5.1] Testimonial: Editor of Hansard (Lords).

[5.2] Testimonial: Director of Data and Search Transformation for Parliamentary Digital Service, UK.

[5.3] Brinkley, L. (2018) 'Innovation versus tradition in historical research methods:

the "Digital Turn"', *Emergence* X: 34-48, p. 45.

[5.4] Navickas, K. (2019). '"Right of public meeting" in Hansard' [PowerPoint Slide]. Paper presented at *Remembering Peterloo: Protest, Satire and Reform.* London, 11 July 2019.

[5.5] Levin, J. (2019, October 8) 'Resource: Cobbett's Parliamentary History', *Alsatia*.

[5.6] McKay, L. (2020) 'Does constituency focus improve attitudes to MPs? A test for the UK', *The Journal of Legislative Studies* 26(1): 1-26, pp. 1 and 21-22.

[5.7] Compilation of webpages that cite Hansard at Huddersfield as a resource for researchers including The Victorian Commons (part of History of Parliament Trust) and Parliamentary Archives.

[5.8] Testimonial: Head of Extended Project Qualification Greenhead College, Huddersfield, UK.

[5.9] Compilation of social media feedback.

[5.10] Testimonial: Hellenic OCR Team Leader, Hellenic Government of Greece