



[Alexander, M.](#) (2023) Speech in the British Hansard. In: Korhonen, M., Kotze, H. and Tyrkkö, J. (eds.) *Exploring Language and Society with Big Data: Parliamentary Discourse across Time and Space*. Series: Studies in corpus linguistics (111). John Benjamins: Amsterdam, pp. 17-53. ISBN 9789027214065 (doi: [10.1075/scl.111.01ale](https://doi.org/10.1075/scl.111.01ale)).

This is the Author Accepted Manuscript.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/309442/>

Deposited on: 04 December 2023

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>

# Speech in the British *Hansard*

Marc Alexander  
University of Glasgow

## Abstract

This chapter gives a detailed textual and linguistic history of *Hansard*, the records of debates of the British Parliament from 1803 to the present, and accordingly the *Hansard Corpus*. It analyses how Parliamentary speech is recorded and presented across that period, examining the changes in direct and indirect speech types arising from commercial factors, pressure from Parliament, editorial practice, and the availability and quality of source material. The chapter concludes with a breakdown, for each period of *Hansard*'s history, of what the data for that period does and does not represent.

## 1. Introduction

[The student of politics] must be on [their] guard against the old words, for the words persist when the reality which lay behind them has changed.

Bevan 1952: 13

The richness of the records of proceedings and debates in the Parliament of the United Kingdom is profound historically, politically, and linguistically. However, the existence of these records under a single name – *Hansard* – can give a false sense of consistency. The span of *Hansard* is instead united only by its persistent discussion of topics of political importance, and only two features of the language of *Hansard* are consistent: its reports of debates are not verbatim (and always have been clearly signalled as such), and they are written in formal English. Other than its two-column layout, no other consistent features are evident.

This richness of content is profoundly exciting to linguists, and was why some time ago colleagues and I created the *Hansard Corpus* (originally spanning 1803-2003, its composition date, then expanded to 2005, and now to the present day). The corpus is an excellent example of humanities ‘big data’ – but again the single term ‘big data’ is not used consistently by researchers. At 1.6 billion words (consisting of 7.6 million ‘contributions’ such as speeches or interruptions) in its 2005 edition, the corpus is fairly big when considered as raw text; but big data in the humanities is, as Lorna Hughes (forthcoming, 2023) has pointed out, a function of *mass*, to which not only size but also density contributes. Complex, dense data has sufficient mass as to require ‘big data’ analytic techniques irrespective of its raw size, and here *Hansard* has not only a large volume of data, and a large ongoing velocity of new material for input, but also a huge complexity of style, content and internal textual variation contained within its virtual and physical covers. In addition to this complexity, the tagged *Hansard Corpus* contains several layers of semantic annotation, to allow for meaning-related searches of considerable delicacy, in addition to the usual grammatical, multi-word, and lemma annotations. Such density – highly complex data with multiple layers of annotation – requires big data techniques just as much as those resources which are ‘big’ solely in terms of their size.

Corpus linguists can be simultaneously enticed and frustrated by *Hansard*. The scope of records it represents – collecting words, thoughts, and topics of both great and middling social and political import across two centuries – is attractive for many types of study, and its size, which would take well over a decade of incessant silent reading to get through, makes it intractable for manual methods of analysis and ideal for corpus linguistics. However, its long history and internal complexity mean that it is challenging to understand precisely what it records, and this can be dispiriting in terms of making sure that an analyst is certain what *Hansard* is telling them.

To use *Hansard* well, we therefore must understand better how *Hansard* reports the speech of parliamentarians. This chapter therefore addresses for the first time the linguistic history of speech in *Hansard*, and by extension the *Hansard Corpus*. It describes the language of the corpus and the evolution of *Hansard* across time in order to build a comprehensive picture

of *Hansard's* representation of political speech in English over the last two hundred years. Such work allows us to understand what the resource can and cannot tell us about language in use over that period. While below I focus on linguistic points, the conclusions are equally significant for any investigations into the corpus in the social sciences and humanities.

The next two sections review the state of Parliamentary reporting before 1803 and introduce *Hansard* and the *Hansard Corpus*, while section 4 discusses the issue of what 'verbatim' means in the context of records of Parliamentary speech. Later sections then proceed chronologically through the evolution of *Hansard's* representation of speech, using the Leech and Short 2007[1981] model of speech representation (described in section 4). The chapter concludes with a breakdown, for each period of *Hansard's* history, of what the textual data for that period can and cannot tell researchers.

## 2. 'Tolerably well': reporting before Hansard

The history of parliamentary debates is similar to the description given of the history of a newspaper. The first day it is read with eagerness, the next day it is thrown away; after the lapse of some years it is worth its weight in gold.

Proposed wording by T.P. O'Connor for Parliament 1888a (xi)

Parliamentary reporting in Great Britain and the UK has a complex and tense history. For a significant period of time, Parliament considered the reporting of its debates to be a breach of its privilege, "partly because what was said in Parliament was felt to be unfit for public consumption, and partly to protect Members from the wrath of the monarch" (Vice and Farrell 2017: 3). This manifested itself in prohibitions such as a House of Commons resolution in 1722 ordering "That no Printer or Publisher of any printed News Papers do presume to insert in any such Papers any Debates or any other Proceedings of this House or any Committee thereof" (MacDonagh [n.d.]: 109). Newspapers and monthly magazines continued to test this resolve, with publishers being prosecuted and imprisoned for publishing debates while experimenting with various camouflage (most notably the

*London Magazine* and the *Gentleman's Magazine's* habit of reporting Parliamentary debates as if they were fiction from Ancient Rome or in the invented country of Magna Lilliputia, with personal names extremely thinly disguised). Regardless, such records were based either on "memory and hearsay" (as Thomas 1959: ix describes them) or were entirely invented. To underscore that last point, Arthur Murphy's 1792 *Essay* on Samuel Johnson, a sometime Parliamentary reporter better known for his lexicography, describes a dinner where a speaker praises a speech by Pitt the Elder, to which Johnson declares he wrote that speech himself in his garret:

I never had been in the gallery of the House of Commons but once. Cave [publisher of the *Gentleman's Magazine*] had interest with the door-keepers. He, and the persons employed under him, gained admittance: they brought away the subject of discussion, the names of the speakers, the side they took, and the order in which they rose, together with notes of the arguments advanced in the course of the debate. The whole was afterwards communicated to me, and I composed the speeches in the form which they now have in the Parliamentary debates.

Murphy 1792: 44-45

Instead of the assembled guests being frustrated by the linguistic or literal inaccuracy of the report of the speech, Murphy reports they "bestowed lavish encomiums on Johnson", including on his impartiality. Johnson replied that this was not quite true, rather that "I saved appearances tolerably well; but I took care that the WHIG DOGS should not have the best of it" (*ibid.*).

This is the environment which gave birth to *Hansard*. The legal prohibition on reporting any proceedings at all, let alone accurate transcripts, meant that accuracy of language was not a concern; generally following the argument in the report and being able to state what speaker spoke at what point was a sufficient advance. The history of the following changes in attitudes to reporting is vividly described in Vice and Farrell 2017, but in brief a rapid breakdown in enforcement of Parliamentary privilege in this area saw an increasing volume of reports being printed in the time leading up to 1803. This development was partly due to a blind eye being turned to the reports by MPs who found them advantageous (even when not written by Samuel Johnson), and also due to the continuing ascendancy of newspapers, who

reported on Parliamentary debates in increasing numbers. Although more reports were published, note-taking in the public gallery was still forbidden and so reports were still “pieced together from conversation, gossip, recollections of listeners, comments from Members, reports in rival newspapers, and the memory of reporters in the gallery” (McBath 1970: 28). Some reporters were said to have a prodigious memory, but even then few would argue for linguistic accuracy on the part of these reports other than in key phrases and topic words.

### 3. T.C. Hansard, *Hansard*, and the *Hansard Corpus*

The volumes of debates that became *Hansard* began as *Cobbett's Parliamentary Debates* in 1804, although for two years before that date debates were reported as a supplement to *Cobbett's Annual Political Register*. William Cobbett (1763–1835) had a zeal for providing the public with political information, and the new *Debates* were described by him in the *Register* as “the *only* compilation at all likely to be regarded as an authentic record of the Legislative Proceedings of the present time” (1804: 863). The word *compilation* here is essential: Cobbett employed no reporters but relied on a range of newspaper reports supplemented by various information, including corrections and speech texts sent by MPs who wanted to correct ephemeral newspaper accounts in the firmer record of the handsomely-bound volumes of *Debates*. Cobbett soon sold *Parliamentary Debates* to its publisher, Thomas Curson (T.C.) Hansard, whose name and that of his son and successor (also Thomas Curson Hansard) became the commonly used metonymic short title for the publication. In 1892, at the end of the family's involvement with the publication – father and son between them editing the debates for some 86 years – the name ‘Hansard’ was removed from the title of the *Debates* where it had been since 1829, but in 1943 the Speaker of the House of Commons instructed that the word *Hansard* would once again be printed on the title page (*HC Deb*, 4 August 1943, c2303).

*Hansard's* volumes are arranged into series, with each series containing numbered volumes. Inside each volume, reference is made not to pages but instead to columns (as *Cobbett's Annual Political Register* also reckoned its entries). While *Hansard* is the overarching term almost universally used, the

First to Fourth Series are formally titled some variation on the words *Parliamentary Debates*, and the Parliament-produced Fifth and Sixth Series are often referred to as the *Official Report* (split into the Houses of Commons and Lords, each report run by its own department), although those series also have *Parliamentary Debates* on their title page.

The First Series runs from first publication in 1804 to February 1820.<sup>1</sup> The Second Series, at the time known as the ‘new series’, begun following the accession of a new King, George IV, in January 1820. It spans April 1820 to July 1830, and the Third Series, October 1830 to August 1891, started with the accession to the throne of William IV in June 1830. The Third Series did not cease in 1837 with the beginning of Victoria’s reign but instead its end was due not to a change of monarch but to a change of publisher: with Thomas Curson Hansard senior having died in 1833 and Thomas junior retiring in 1888 (later dying in 1891), the business was at that point sold.

From this point until 1899, the work of producing the *Parliamentary Debates* would go through five different owners in eleven years – the business was subsidised by Parliament during this period to varying extents, but it was still near-impossible to run at a profit. The sixth owner during this period enjoyed a stretch of relative stability lasting nine further years until the end of the Fourth Series. Increasing dissatisfaction in Parliament with the subsidised commercial reports built over a long period, culminating with the establishment of the in-house *Official Report*, which inaugurated the Fifth Series in 1909 (with the House of Commons beginning a Sixth Series in 1981 after the thousandth volume in its Fifth Series; the separately-produced Lords *Hansard* instead continues with the Fifth Series).

No comprehensive survey of *Hansard* publication yet exists, but the key facts about each series are summarised in Table 1, and the size of each year’s contributions in words in the 1803-2005 corpus is shown in Figure 1.

---

<sup>1</sup> The first volume from 1804 covered 22 November 1803 to 29 March 1804, and so contained material from debates in 1803. 1803 is therefore used as the ‘start’ date of the contents of *Hansard*, while its publication dates begin in 1804.

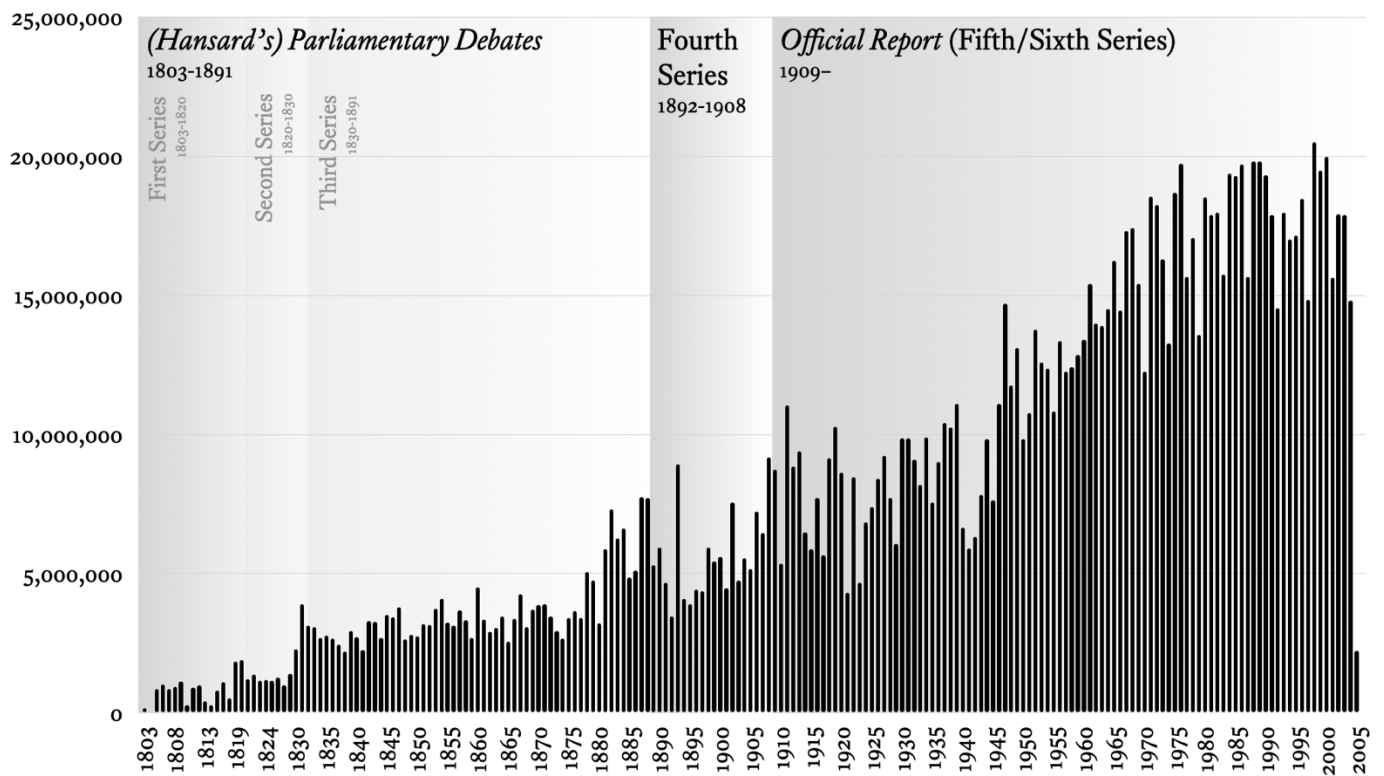


Figure 1: The size of *Hansard*, by year and showing Series.

Series	Title	Starting Volume	End Volume	Notes
First Series	<i>Cobbett's Parliamentary Debates</i>	Vol 1 (November 1803)	Vol 22 (March 1812)	Edited by William Cobbett, published by T.C. Hansard (senior). Note that the most commonly available copies from this period are the 1812 reprint series as <i>The Parliamentary Debates</i> under T.C. Hansard (senior).
	<i>The Parliamentary Debates</i>	Vol 23 (May 1812)	Vol 41 (February 1820)	Edited and published by T.C. Hansard (senior).
Second Series	<i>The Parliamentary Debates</i>	Vol 1 (April 1820)	Vol 20 (February 1829)	Edited and published by T.C. Hansard (senior).
	<i>Hansard's Parliamentary Debates</i>	Vol 21 (March 1829)	Vol 25 (July 1830)	Edited and published by T.C. Hansard (senior).



Third Series	<i>Hansard's Parliamentary Debates</i>	Vol 1 (October 1830)	Vol 356 (August 1891)	Edited by T. C. Hansard senior until his death in 1833 (vol 21), then T. C. Hansard junior (no distinction made on title page) until 1888, then editor unspecified. Published by the Hansards, then from 1845 'Printed and published for Mr Hansard by G Woodfall and Son' until 1850, then printing undertaken by Cornelius Buck (later Cornelius Buck and Son) 'at the office for Hansard's Parliamentary Debates' from 1850 to 1888, then 1889 onwards, 'The Hansard Publishing Union' ('printers, publishers, and proprietors of "Hansard's Parliamentary Debates" under contract with H.M. Government').
Fourth Series	<i>The Parliamentary Debates, Authorised Edition</i>	Vol 1 (February 1892)	Vol 199 (December 1908)	Produced by varying firms: Vols 1-7 (until August 1892) Reuter's; Vols 8-29 (until August 1894) Eyre and Spottiswoode; Vols 30-52 (until August 1897) Waterlow and Sons Ltd; Vols 53-70 (until May 1899) Economic Printing and Publishing Company/F. Moir Bussy; Vols 71-199 (until December 1908) George Walpole/Wyman and Sons Limited.
Fifth Series	<i>Parliamentary Debates (Official Report)</i>	House of Commons: Vol 1 (Jan 1909)  House of Lords: Vol 1 (Jan 1909)	House of Commons: Vol 1000 (March 1981)  House of Lords: (Ongoing)	Produced by the <i>Hansard</i> reporters of both Houses. Printers: H.M. Stationery Office
Sixth Series	<i>Parliamentary Debates (Official Report)</i>	House of Commons: Vol 1 (March 1981)	House of Commons: (Ongoing)	Produced by the <i>Hansard</i> reporters of the House of Commons. Note that the House of Lords continues to the present day with the 5th Series.

Table 1: The Series of *Hansard*, with publishers, personnel, and titles.

The transition between the Fourth and Fifth Series – a private business pre-1909 and two departments of the Houses of Parliament thereafter – is the key separation of the text into two major phrases. More fine-grained divisions can be distinguished from a careful examination of the publishing history of the debates and their linguistic characteristics, and the following sections undertakes this. Note that as the distinction between the first three series relates simply to the regnal years of the Parliaments concerned, in later figures I have combined them for simplicity.

A near-full set of all series was digitised in the mid-2000s (with a few minor omissions, some of which are now corrected online). The scanning and OCR of the over three million pages of *Hansard* then available was funded by Parliament itself, and this dataset was admirably made available through a website which was an open-access public beta using free and open-source software (including a public discussion group and issues log), in what was an unusual move both for the time and for Parliament. This availability led to a few experimental corpora being created at the University of Glasgow, following the presence of samples of *Hansard* material in other corpora such as LOB. Funds from the UK Jisc resulted in the 2011 *Parliamentary Discourse* project to create a tagged corpus of the full digitized ‘Historic Hansard’ (1803-2005; data beyond 2005 was available, but we set the end date at the UK General Election in 2005 to give a clear end date to the corpus). The many gigabytes of text collected were processed to produce one XML file per speech contribution. The metadata were cleaned and made suitable for corpus use, and then the text was tagged twice: once by the UCREL team at Lancaster (led by Paul Rayson), with CLAWS grammatical annotation (see Leech *et al* 1994) and USAS semantic annotation (Rayson *et al* 2004), and once by the GATE NLP team at Sheffield (Cunningham *et al* 2013). The Jisc funding was used to create an annotated offline corpus, although it was also loaded into Glasgow’s experimental ENROLLER high-performance computing platform (Anderson 2013). The troubles of distributing the corpus were substantial (primarily involving asking people to post external hard drives to Glasgow, where data would be copied over and posted back).

When planning the SAMUELS project (2014-15), funded under a specialised ‘big data’ call, the *Hansard Corpus* was an ideal large test case ready to use for that project’s semantic annotation system, and consequently SAMUELS put the semantically-tagged *Hansard Corpus* online for a wider user base (as Alexander and Davies 2015). Each word in the corpus files is therefore tagged across six dimensions (lemma, CLAWS part of speech tag, USAS semantic tag, a multi-word-expression marker, and two levels of SAMUELS semantic tags, each with up to three tags; see Piao *et al* 2017 for more on SAMUELS tagging). At Huddersfield, a daughter project to SAMUELS to update the time span of the corpus and make it much friendlier for non-academic users has also been recently completed (see Jeffries, this volume),

as well as other projects among members of the SAMUELS consortium and other contributors to the corpus.<sup>2</sup>

In addition to the *Hansard Corpus* as I describe it here (used either offline with the tagged files or online via a range of websites), there are also other corpora of *Hansard*. These have increased in recent years due to the wide availability of the original files alongside modern tools to obtain large amounts of online text and prepare them for corpus use. While there is therefore a large value of rich research which involve the creation of *Hansard* corpora designed for particular purposes, I am not aware of another *Hansard* corpus intended for general or broad re-use. The range of available versions of *Hansard* as corpora demonstrates its wide applicability for a range of research questions; this; this means that a deeper understanding of what *Hansard* contains, and how each stage of its development represents Parliamentary speech, is essential to use its data well. The following sections provide an overview of these key questions for the time span of 1803 to the present.

#### **4. ‘Fidelity is the first and indispensable requisite’**

For researchers using these reports, a key aspect of linguistic interest is the extent to which *Hansard* at various points provides a verbatim transcript of speech. Historical linguists are now used to asking the question which Culpeper and Kytö (2010: 3) frame as “why are written texts representing spoken face-to-face interaction as they are?”; sources from the history of English always approximate speech to various different degrees, and *Hansard* is no different. Many users of the resource, however, are not historical linguists, and so may not be accustomed to having to trace the relationship of the text to the speech it tries to represent. Beyond this, the issue of how ‘verbatim’ *Hansard* is – that is, how directly its text maps to the spoken acts in the Chamber – can be more complex than it first appears.

---

<sup>2</sup> See, for example, Alexander *et al* 2015, Alexander and Struan 2013, Alexander and Struan forthcoming, Archer 2017, Coole *et al* 2020, Piao *et al* 2017, and Wattam *et al* 2014. Colleagues who worked alongside me on the *Hansard Corpus* include Jean Anderson, Brian Aitken, Dawn Archer, Alistair Baron, Fraser Dallachy, Mark Davies, Jane Demmen, Lesley Jeffries, Christian Kay, Bethan Malory, Scott Piao, Paul Rayson, Andrew Struan, Brian Walker, and Stephen Wattam.

This section discusses this issue in more detail, and introduces the Leech and Short model of speech representation used in the following sections.

*Hansard*, at heart, is a written account, in the written mode, created as a written text, intended to be read as a written text, but ultimately based on speech. This means the text exists at a distance from the speech it records, as *Hansard* exists primarily as a legal record and a statement of the content of debates for posterity. In that context, it is clear it could never be ‘verbatim’ – any beginning student of linguistics knows how wholly unreadable linguistic transcripts are if one is interested just in their sense – and beyond this, the term ‘verbatim’ is a flawed one in this context. While the term can be literally glossed as ‘word for word’, it is more generally used to mean something like ‘representing what was actually said’. This can encompass a broad cline of meanings, such as:

- a segmented phonetic spectrograph, which is perhaps the most ‘accurate’ non-spoken representation of speech and the least intervened-with by a transcriber, although it will still involve intervention for ambiguous segments;
- a phonemic transcript where it is acceptable to edit fine phonetic detail, such as reducing allophones to a phoneme, or making choices of phonemic theory and application;
- a linguistic transcript for various purposes -- and where depending on those purposes it is acceptable to edit out features of low interest, such as hesitation phenomena or phonemic detail, but generally not to alter e.g. syntactic structure;
- a ‘readable’ transcript, where it is necessary to edit out features which hinder readability, given the different processing demands of speech and text);
- a ‘publisher’s-standard’ transcript, where it is acceptable not only to make a transcript highly readable but also to conform it to explicit and implicit rules of consistency and house style;
- and even more nebulous interpretations of ‘verbatim’ beyond this, depending on the use case underlying the work.

In the case of *Hansard* across time, beyond a ‘publisher’s-standard’ transcript there is what we might call a transcript ‘for the ages’. Here intervention for readability and consistency is undertaken, but due to the

nature of the text as a record of fact there has to be some limited intervention beyond reporting what was said in order to render clear and uncontroversial some instances of what in the full context of speech was understood but not necessarily articulated. This last is the fluid space within which *Hansard* operates.

Terms such as ‘verbatim’ only highlight the scale of intervention which lie behind the purpose of a transcript, and in the practice of linguistics are often useless. Linguists interested in the act of speech naturally want transcripts to map as closely as possible to the spoken act, but unless we are the ones paying for them we cannot reasonably expect any given transcript to conform to whatever point of the cline above we would personally prefer. *Hansard* is entirely accurate as a record of *Hansard*, and is not entirely accurate as a record of the linguistic acts performed in the Houses of Parliament. A preferable term, from the perspective of the report, would be ‘faithful’: is *Hansard* faithful to the words spoken, given how the spoken act is tightly bound with the context of the speaker and the nature of the institution?

This issue was brought up by John Campbell, a Parliamentary reporter in the early nineteenth century who later became Lord High Chancellor (and so being one of the few people who both frequently reported on speeches and then had his own reported on), who when reflecting on his time as a reporter reflected that:

...there probably never was a parliamentary debater in whose language there was not some inaccuracy, and who did not fall into occasional repetitions. These are hardly perceived in the rapid stream of extemporaneous eloquence, and are corrected and remedied by the voice, the eye, the action of him to whom we listen; but blazoned on a printed page which we are deliberately to peruse, they would offend and perplex us. [...] Fidelity is the first and indispensable requisite, but this does not demand an exposure of inaccuracies and repetitions.

Campbell 1881: 106-7

This is a highly lucid expression of a key linguistic point: *Hansard*'s fidelity, across its long history, does not relate to the linguistic reproduction of syntactic constructions or hesitation phenomena, but to a ‘faithful’ rendering of what ideally would have taken place were humans able to clearly and

consistently speak in coherent prose. Therefore, to use *Hansard* effectively we must know from its history and text what *Hansard* consists of, rather than arbitrarily define a moving target of ‘accuracy’.

There are two simultaneous phenomena which must be taken into account to understand the language of *Hansard* and the *Hansard Corpus* across its two centuries: firstly, the changing relationship of the report’s fidelity to the language of its source speech, and secondly, the formal and informal language practices which surround the report across time, set by its various publishers and editors. In short, who was publishing what language, for whom, and with what constraints? Below, these two factors are discussed with regard to the three primary linguistic ‘phases’ of *Hansard* production they engender.

Throughout the following sections, terms from the Leech and Short 2007[1981]: 256ff (and the later Semino and Short 2004) stylistic model of speech representation are used. Here, a cline of ‘interference’ in a speech report is distinguished, based on quotation phenomena, syntactic dependence, pronoun type, tense, deictics, and the presence of reporting verbs. This relies on the presence of a narrator, which is different from a reporter in *Hansard* terms: a narrator has a distinct voice separate from the author, while a Parliamentary reporter is closer to an author narratologically-speaking. To illustrate these different types as they apply to *Hansard* (with some modifications to what is presented in Semino and Short 2004: 10), here is an example built on the Samuel Johnson anecdote reported above:

**NRSA *Narrative Report of Speech Acts*:** narrator in maximal control of report, no claim to quotation, speech described ‘at a distance’ rather than through any aspect of what was said, third-person pronouns.

Johnson looked at his friends and accepted their complements on his impartiality.

**IS *Indirect Speech*:** narrator in some control of report but content based on the speech act, no claim to quotation as the narrator uses their own words to represent that content, third-person pronouns.

Johnson looked at his friends and told them he had tried to be impartial and succeeded fairly well.

**FIS** *Free Indirect Speech*: narrator in some control of report, narrator's voice merged partly with the speaker's, no speech attributions, third-person pronouns.

Johnson looked around at his friends. He had certainly saved appearances tolerably well!

**DS** *Direct Speech*: narrator only has partial control of report outwith quoted sections, speech attributions (such as quotation marks) often present, tenses third person outwith quotes and first person inside quotes.

Johnson looked at his friends. 'I saved appearances tolerably well,' he said, modestly.

**FDS** *Free Direct Speech*: narrator has no apparent control over speech (or is not distinctly present), intermediary reporting phenomena like quotation marks and reporting verbs removed.

I saved appearances tolerably well.

In *Hansard's* case, a 'pure' and wholly faithful transcription without the presence of an intermediary would be FDS, with a neutral intervention to mark speaker name but without any narratorial 'presence'. Reports which summarise but without any claim to represent the actual words of a speaker all exist in a space occupied by NRSA, IS and FIS, where a reporter acts as narrator, summarising the words of a speaker in the third person either as NRSA, where the content was unimportant or heavily summarised, or as IS, where the speech is covered in more detail, and where relevant including flashes of FIS, for example to represent notable turns of phrase. Illustrations of each of these are given below.

We now turn to the major 'phases' of *Hansard* and what the language used in each can be said to represent.

## **5. 'Bound for the *bona fides*': The Hansards and the *Parliamentary Debates*, 1803-1888**

As already discussed, early *Hansard* is a 'compilation' of already-published reports supplemented by speech notes and other information supplied by

Members. At this time, its principal benefit – as well as being a well-printed volume dedicated to producing reports – was its lack of speed. *The Times* on its foundation as the *Daily Universal Register* advertised itself as providing faster reports of the previous night’s Parliamentary business than any other newspaper; *Hansard*, therefore, could not compete on speed but could on detail.

The First Series of 1803-1820 was particularly small by comparison to later volumes, as Figure 1 shows. The junior T.C. Hansard himself referred to this feature in a note on ‘Origin and Progress’ appended to an advertisement found at the end of some volumes in the 1840s. He says, for the First Series:

...owing to the rarity of long Debates and in the then imperfect state of Parliamentary reporting, the proceedings of each Session are reduced into a very small compass [...] being, for the most part, either the careful report of some Speech, (very frequently taken by a friendly Member in his place), or the draft itself of the Speech, or a report supplied by the Member or corrected by him; the materials derived from the gallery of the House are few, but very important, as being a digest of discussions on points of form and privilege by politically interested observers.

T.C. Hansard 1843: 2

T.C. Hansard here does not refer to the material taken from other publications, perhaps natural in an advertisement, and is clear that he does not employ his own reporters in the Gallery. As such, the two options he presents are a “careful report” (NRSA), a speech draft (FDS), or a report corrected by the member (which as a merger of NRSA and DS seems to occupy the FIS space, produced by the merger of narrator and speaker).

Examination of the early volumes themselves supports this analysis. One detailed example will illustrate the varieties present, taken arbitrarily from the House of Commons on 2 June 1819:<sup>3</sup>

---

<sup>3</sup> ‘Arbitrarily’ here means that many examples here were taken from debates which occurred on or around an anniversary of the date of the workshop from which some chapters in this volume derive, as a rough means of avoiding systematic selection bias. (Due to weekends and other days Parliament does not sit, often this means the nearest date to that anniversary.)



*Sir G. Warrender* rose to move the Navy Estimates. Upon this subject the hon. Baronet said he should think it unnecessary, if not presumptuous on his part, to make any speech, or to enter into any detail, as the question had already been so amply discussed, while all the details connected with it were so perspicuously stated, in the report of the finance committee; therefore, he felt that it rather became him to wait for any observations which might be made on the other side of the House, and to which he would endeavour to submit a satisfactory reply.

*Mr. Calcraft* expressed his surprise at the course pursued by the hon. Baronet, particularly in referring to, and wholly relying upon, the statements of the finance committee. Was it, from the observation of the hon. Baronet, to be taken for granted that he adopted the recommendation of the finance committee?

*HC Deb*, 2 June 1819, c823-4

The opening sentence is a clean NRSA/IS statement, although it is impossible to tell if it is one or the other without being certain if Warrender said “I stand to move the Navy Estimates” or something else. Erring on the side of caution, it is best considered NRSA. The remainder of Warrender’s speech and the beginning of Calcraft’s is likewise NRSA/IS, the third-person pronouns and reporting verbs keeping it at that end of Leech, Semino, and Short’s scale. The final sentence of the excerpt of Calcraft’s contribution here (the speech goes on to take up three full columns) poses some other issues. The use of the rhetorical question is a hallmark of FIS in fiction, and a favourite device of many authors; NRSA and IS would be more likely to specify via a reporting verb something like “He asked if it were to be taken for granted...”. It can therefore be labelled as FIS, although we should be careful to avoid the implied literariness of that term.

As the junior T.C. Hansard indicates, there are also stretches of *Parliamentary Debates* from the first series onwards in FDS. For example, in volume 5 of the First Series, a complex debate on a minor loan scandal involving William Pitt the Younger (*HL Deb* 14 June 1805, c385ff) begins with a long speech in FDS, introduced only by the name of the speaker and a long dash. The next speaker in this debate is introduced by “The Chancellor of the Exchequer rose and spoke to the following effect:— Sir, I do not think it necessary to endeavour to follow the hon. Gent...”. Here, the speech is introduced by a reporting phrase “spoke to the following effect”, which might

be thought to introduce NRSA but instead goes to a more direct first-person report. The third speech in the debate returns to clear NRSA: “Mr Henry Lascelles commenced his speech with observing, that it would ill become him, after what the house had heard...”. The remainder of the debate continues in NRSA, including contributions by the two speakers whose long introductory speeches were in direct speech. NRSA continues to be the default throughout these first three series, with only occasional interspersed moments in direct speech.

The example just cited is a good example of *Hansard* likely printing a speech from the speaker’s notes, as at this time such a long excerpt would be unlikely to be recorded in shorthand. (It would be useful to separate any extemporaneous FDS actually obtained from shorthand from the faux-FDS obtained from prepared notes, although in the historical *Hansard* context this is likely to be an impossible endeavour.) On the question of shorthand at this time, John Campbell, the one-time reporter quoted in the previous section, commented that “I knew nothing, and did not desire to know anything, of short-hand. Short-hand writers are very useful in taking-down evidence as given in a court of justice, but they are wholly incompetent to report a good speech” (Campbell 1881: 105).

With NRSA and occasional direct speech (of some form) the norm throughout the First to Third Series, the question of how extensive the direct speech is, and how it was obtained, becomes significant. Figure 2 reproduces Figure 1 but overlays the frequency of ‘I’ per million words. Setting aside the Fourth Series for now (discussed in the following section), the contrast between the *Hansard’s Parliamentary Debates* period and the first-person post-1909 *Official Report* is stark. Bearing in mind the small size of the First and Second Series, which accounts for the rather wild changes per million words in that period, the First to Third Series have on average slightly less than one third of the rate of occurrence of ‘I’ as the Official Report does. Approximately a quarter to a third of this period consists of first-person content (with significant year-by-year variation), although the significant length of individual speeches in the first person compared to the NRSA reports means that topics in the first person are substantially overrepresented compared to the third-person material. Similarly, Figure 3 shows the decline in the reporting verb ‘said’ – by no means the only

reporting verb used in *Hansard*, but indicative of how the first three series compare to the *Official Report* in terms of markers of NRSA. It should be noted that ‘I’ and ‘said’ are not unique markers of first-person and third-person wholly in their own right – ‘I’ can be used in a quote in a third-person piece of text, and ‘said’ can be used in first-person text (‘as I said earlier’ or ‘as the Government said’). ‘Said’, however, is heavily used in *Hansard*’s style pre-1909 to introduce third person speeches, along with close synonyms, and in the corpus the subjects of the verb form of ‘said’ are almost all noun forms representing people. The two words are good proxies, in aggregate, of the two major modes of speech representation with which we are concerned.

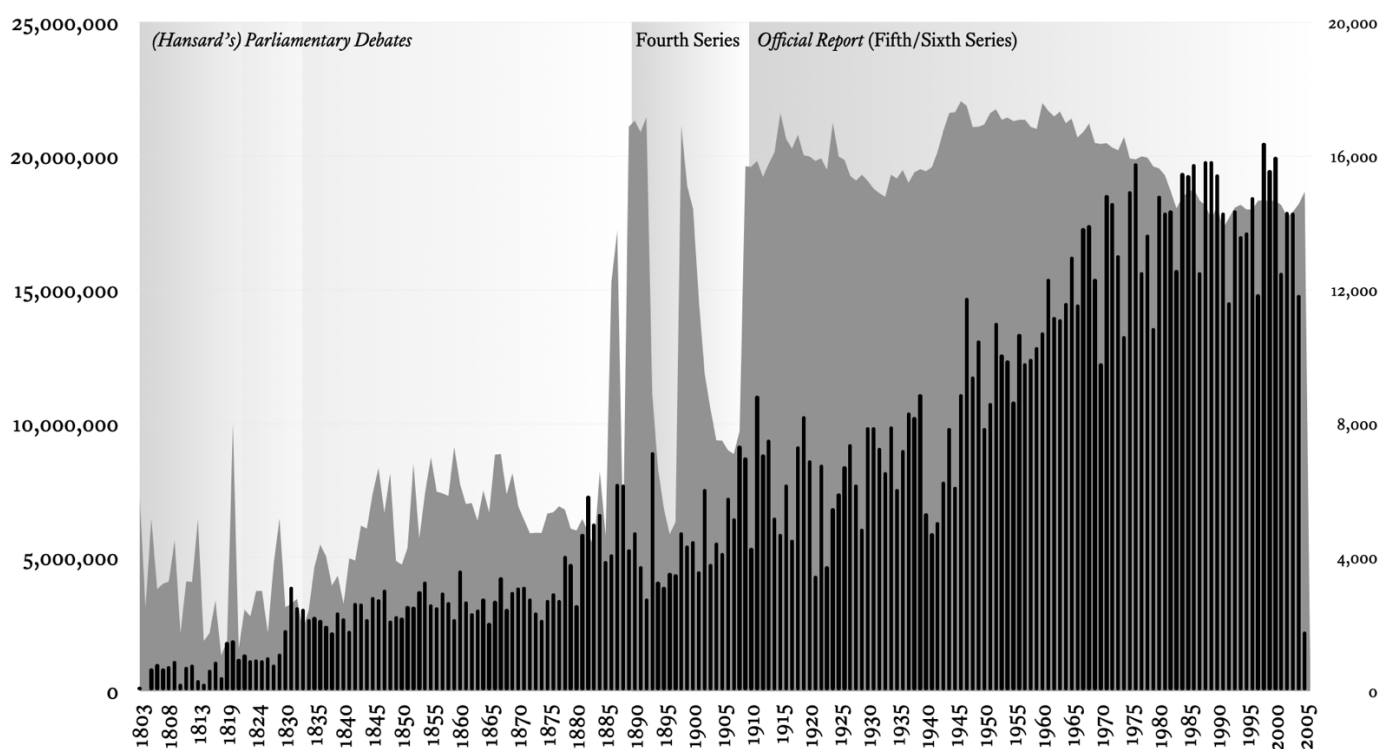


Figure 2: ‘I’ per million words in the *Hansard Corpus* alongside the number of words contained per year, and major Series markers.

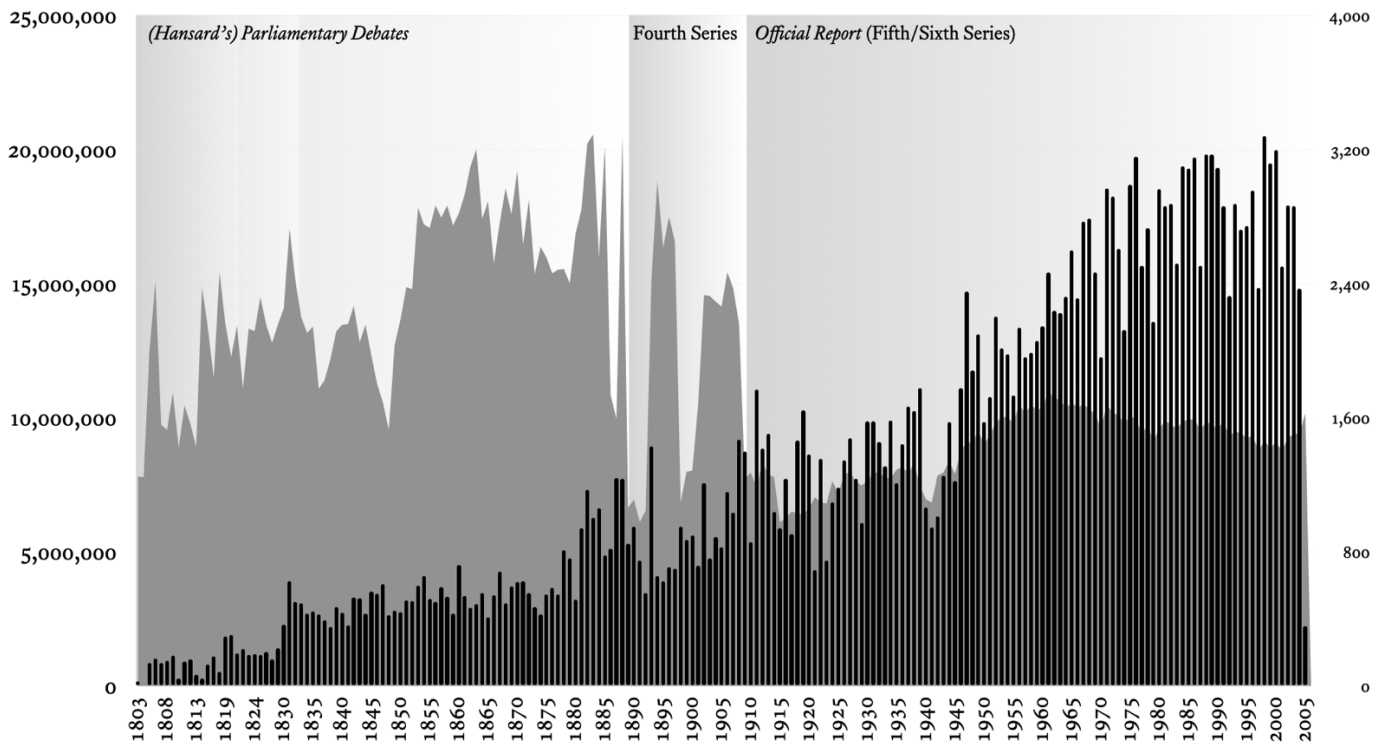


Figure 3: 'said' per million words in the *Hansard Corpus* alongside the number of words contained per year, and with major Series markers.

What gets to be represented as direct speech? T.C. Hansard's 'Origin and Progress' note in his advertisement continues on to discuss the Third Series, and says that:

Peers and Hon. Members, of all parties, of the highest influence and importance, in and out of office, deeply interest themselves in the completeness of the Work – and that IN THE COURSE OF SESSION 1842, UPWARDS OF **Seven Hundred**, and in 1843 **Six Hundred-and-fifty**, SPEECHES WERE REVISED BY THE DIFFERENT SPEAKERS – now, when it is considered that very few speeches but those of length and importance are thus cared for, that much of the volumes is occupied by Minutes of Proceedings, Lists of Division, and minor Debates, it will be evident that HANSARD presents a mass of Reports of the most valuable and authentic character.

Hansard 1843: 2, formatting in the original

In 1843, the *Hansard Corpus* records 2,855,357 words across 4,257 contributions in Commons, and 728,551 words across 1,108 contributions in Lords, so while 650 revised speeches may seem quantitatively small in that context, these 650 will qualitatively represent longer and more major proceedings. (A significant number of the other contributions are very short, too, representing brief questions, interjections, or clarifications rather than the stereotypical and lengthy content-driven speech.) Such correction continued for some time, prompting one member to comment in a debate in 1877:

I am bound to say that no one recognizes more than I do the able way in which Mr. Hansard has conducted his difficult work during this long number of years [...] Unfortunately, however, this system gives rise to a great deal of inconvenience. Members have to spend a considerable amount of time in correcting their speeches. I am told that several hon. Members are occupied for a very long time in correcting their contribution to the debates. Indeed, I am told that one right hon. Gentleman, who has occupied a very prominent position in this House, has actually been put to the trouble of spending two solid hours in correcting an hour's speech.

Hon. Charles Hanbury-Tracy, *HC Deb* 20 April 1877, c1552

For a period from 1889 to the early twentieth century, speeches corrected by a speaker bore a useful asterisk noting the report had been so revised (from 1850-1889 the asterisk meant the speech was taken from a pamphlet or other authorised speech report). With regards to the bulk of speeches not revised in this way or otherwise reported in the first person, some minor elements are either omitted or poorly recorded.

The state of clarity of the debates overall during this period is shown by the frankness of the *Times*, perhaps the foremost newspaper of record of the time, about the limitations of its 'parliamentary intelligence'. For example, on 4 June 1847 Lord Campbell rose to discuss bills of Scottish property law and the *Times* report on Parliament on that day stated only that "The noble and learned lord shortly explained the objects of these bills, but the details were inaudible in the gallery". *Hansard* decides to simply list the bills as being presented, mentioning Campbell's name in a list, presumably due to a lack of available evidence or considering the speech too minor to investigate any further. As a further example, on 29 April 1836 *Hansard* neglects to

mention either the Earl of Haddington's observations on the Irish Constabulary, which the *Times* reports in some detail, or the response by Viscount Duncannon, on which the *Times* reporter – although able to hear Haddington with some thoroughness – has to report simply that he “said a few words in reply, which were wholly inaudible in the gallery”. Where the sources are silent and the speech unimportant, the limitations of *Hansard's* NRSA sources and the need to ensure what went into the reports was what had been substantiated, meant that silence would often fall over some reports.

This leaves direct speech from a source other than the speaker. *Hansard* does not wholly recast existing reports but instead copies extremely freely – to modern eyes, rather too freely – from other reports such as the four major London newspapers the *Times*, the *Morning Post*, the *Morning Chronicle*, and the *Morning Herald*. The time lag of *Hansard* compared to newspapers perhaps excuses this copying on the grounds that *Times* subscribers, for example, were paying for a first look at proceedings, and *Hansard* reprinting these some weeks later for a different purpose would bring little financial harm to the newspaper. This is why, during these first three Series, the idea of a particular editorial direction in *Hansard* is very hard to discern; the editorial matter in a direct speech is often unchanged from the source. This can be easily seen through what Vice and Farrell memorably call “asides and stage directions” (2017: 21) – where these are present in the source, then *Hansard* will allow them through unchanged. An example from the corpus is when Sir Robert Peel, on 9 February 1843, rouses the Commons into a frenzy:

...I will exhibit to you, in the course of ten months more, that same Governor-general at the head of 40,000 men, having effected the evacuation of the kingdom in which we suffered such reverses—having, on the scene of every former disaster, retrieved our honours—[the remainder of the sentence was rendered wholly inaudible by an enthusiastic burst of applause from all parts of the House]. I will show you these dispirited sepoy converted into an army, excited by enthusiasm...

*HC Deb*, 9 February 1843, c372

This speech, and the rest of the debate, is taken directly from the *Times* of the next morning (p.5), right down to the phrasing of the interruption.

*Hansard* cannot be easily described as having a clear linguistic style here, given that other than switching round brackets for square ones it absorbs interjections and other editorial matters from its sources.

Alongside the *Times*, the *Morning Chronicle*, and other newspapers, a new competitor to *Hansard* throughout the first quarter of its Third Series (from January 1828 until October 1841) was *The Mirror of Parliament*. This was founded and edited by John Henry Barrow, now better known as Charles Dickens' uncle. Vice (2018) effectively makes a case for how important Dickens' time working for the *Mirror* was for Dickens' later life, and otherwise the *Mirror* itself has escaped much scholarly attention. Aside from some brief mentions in the histories of *Hansard* (including Vice and Farrell 2015 and Jordan 1931), the *Mirror* is cited only occasionally by modern scholars working within the narrow period of its existence, a fact not helped by the lack of an easily-accessible digital version: these main secondary sources are Everett S. Brown's 1955 article and brief discussions in McBath (1970: 29-30) and in Jupp (1998: 203-205). This is regrettable as the *Mirror* is for much of its existence a superior source to *Hansard* for a range of purposes: speeches are often – but not always – reported in more detail, it is consistently FDS, it is much more willing to give stage directions, including indicating laughter and exclamation marks, and Jupp (1998:204) suggests the length of speeches in the *Mirror* match contemporaneous speech timings at a normal speaking pace. In the same debate as was quoted above, where Hanbury-Tracy mentioned the length of time members spent correcting *Hansard's* record of speeches, the former and future Prime Minister William Ewart Gladstone commented:

At the time when the curiosity and eagerness of the public about Parliamentary debates was at high-water mark [...] an attempt was boldly made by a gentleman named Barrow, to produce a verbal report of the proceedings of Parliament. He succeeded and carried it on for several years; and, for those years, do not hesitate to say, that Barrow's *Mirror of Parliament* is the primary record, and not *Hansard's Debates*, because of the greater fulness which Barrow aimed at and obtained. It is within my own recollection—in the year 1833 or 1834—just after the Reform Bill, that Gentlemen, who wanted to correct their speeches, did it for Barrow's *Mirror of Parliament*. I grant that after, I should think, five or six years, it was found impossible to carry it on. Private enterprize would not

sustain reporting so carried on; and, after the lapse of that time, there was a decline in the appetite of the public for it.

*HC Deb* 20 April 1877, c1576-77

As Gladstone states, the cost of producing a report of the quality and length of the *Mirror* was a significant financial challenge, and after some drawn-out troubles the *Mirror* ceased publication in 1841.

The establishment of the *Mirror* in the late 1820s led to some mudslinging between Barrow and T.C. Hansard, giving us some rare public discussions of policy and practice. The first volume of the *Mirror* began with a Prospectus setting out the aims of the work, which dismissed *Hansard* as “a mere abstract of the reports contained in the daily papers” (which had already been called “decidedly ineffective”) and was “excessively abridged” (1828: 1-2). This led T.C. Hansard to issue an advertisement in the *Gentleman’s Magazine* on 25 April 1828, announcing more frequent publication schedules of his own work and addressing some of the claims of the *Mirror*’s Prospectus.<sup>4</sup> These include criticising the *Mirror*’s fuller reporting, as it spends much space on “trite, temporary matters” and seems to think “*length* is, upon all occasions, to be preferred before *strength*” (1828: 5). Relevant here is that T.C. Hansard insists “Hitherto, the practice of giving parliamentary speeches in the first person has been reserved for great occasions” (1828: 7), rather than the standard practice for which Barrow aimed. There is even an accusation that Barrow simply borrowed from newspapers (as *Hansard* did) and then “changes them, by turning them from the third into the first person, and so disfigures them [...] by repetition, circumlocution, and amplification”. The existence of Barrow’s ruinously expensive shorthand reporters (including his nephew Dickens) contradicts this claim, although it is an interesting accusation, and it is certainly likely that Barrow did not have enough reporters for the completely full report the *Mirror* produces (Jordan 1931: 438).

During this first phase we also have more insight into the editorial practice of *Hansard* from discussions of subsidy. The issue of subsidising the Reports was first raised in 1877, after a period of dissatisfaction with what was recorded in Parliament. The fundamental issue was that *Hansard*’s primary

---

<sup>4</sup> I am indebted to John Vice for this reference.



sources at that point were newspapers, and newspapers had two main problems when it came to the debates: they only wished to pay to report and print material likely to be of interest to their readers, and they had daily printing deadlines to meet (so when Parliament met late into the night, later material went unrecorded while the reporters left to file what they had by the deadline). Consequently the junior T.C. Hansard was granted money in that year by the Exchequer to employ his own shorthand writers in order to record what was neglected by newspapers (discussions after midnight and unpopular but important work such as Committees or private bills); the remainder of his material was still collated from newspapers. In evidence to a Select Committee in 1862, he stated “The reports of the Debates are collected from a great number of sources; it is a compilation very carefully and very laboriously edited,” but was later drawn to admit “I use the *Times* frequently” (Parliament 1862: 41). When asked about accuracy and detail at that Committee, he replied simply, “I hold myself bound for the *bona fides* of the reports, not for their literal accuracy” (1862: 39). In another Select Committee in 1888, he was asked from where he took his sources for collation:

I obtain them from the London newspapers, the country newspapers, and from the reports supplied by the Press Association. They are then passed into the hands of the collators, who collate all those sources together; they take the public documents that relate to the subject, and, in fact, compile and edit the whole.

Parliament 1888a: 9

The “country newspapers” were those who would pay extra for reporters to take shorthand notes of speeches by their local MPs, who were often underserved by the metropolitan newspapers.

To illustrate this range of sources for *Hansard*, we can look at a speech made on 27 March 1840, where Viscount Palmerston, at the time Foreign Secretary, was asked to discuss a minor crisis between Constantinople and Egypt. One section of his speech’s report from the *Times* reads:

His hon. friend seemed to be under the impression that the British Government and Lord Ponsonby stimulated the Sultan to renew hostilities against the Pacha of Egypt. In this his hon. Friend was much mistaken. The

Pacha was the aggressor. In the first instance he declared his determination to throw off his allegiance to the Sultan, and make himself the independent sovereign of the province which he was appointed to govern.

*The Times*, 28 March 1840, p.4

The *Morning Chronicle*, another major newspaper source of debates, has for the same part of the debate:

His honourable friend thought that the British government and Lord Ponsonby, the British Ambassador at Constantinople, has stimulated the Sultan to renew hostilities against the Pacha of Egypt. He (Lord Palmerston) could assure him that he was entirely mistaken. In the first place, it was the Pacha who was the aggressor, and not the Sultan, inasmuch as it was the Pacha who in the first instance, publicly declared his determination to throw off his allegiance, and make himself the independent sovereign of the provinces over which he was appointed to govern.

*The Morning Chronicle*, 28 March 1840, p.4

Published a few weeks later, *Hansard* copies the *Morning Chronicle* word for word, although with some changes for house style, as follows:

His hon. Friend thought that the British Government and Lord Ponsonby, the British Ambassador at Constantinople, had stimulated the Sultan to renew hostilities against the Pacha of Egypt. He (Lord Palmerston) could assure him that he was entirely mistaken. In the first place, it was the Pacha who was the aggressor, and not the Sultan, inasmuch as it was the Pacha who, in the first instance, publicly declared his determination to throw off his allegiance, and make himself the independent sovereign of the provinces over which he was appointed to govern.

*HC Deb* 27 March 1840 c193

Finally, the same section in the *Mirror of Parliament* is:

My honourable Friend thinks that the British Government and Lord Ponsonby, the British Ambassador at Constantinople, stimulated the Sultan to renew hostilities against the Pacha of Egypt. I can assure him that he is entirely mistaken. In the first place, my opinion is, that the Pacha is the aggressor, and not the Sultan; inasmuch as the Pacha, in the first instance, publicly declared his determination to throw off his allegiance, and make

himself the independent sovereign of the provinces of which he was appointed the governor.

*The Mirror of Parliament*, 27 March 1840, p2042

With regards to other newspapers, the *Morning Post* gives only a brief third-person summary of this exchange and the *Standard's* report is identical to the *Times*. It is notable that the *Morning Chronicle* and the *Mirror of Parliament* match closely but are not identical: as well as the first-person reporting of the *Mirror*, some small changes of phrasing (“over which he was appointed to govern” versus “of which he was appointed the governor”, and “In the first place” versus the quasi-hedge in “In the first place, my opinion is”) indicate that they were likely independently produced by shorthand writers. By contrast, *Hansard* simply reprints the *Morning Chronicle* with “hon. Friend” in place of “honourable friend”, including the parenthetical “Lord Palmerston” to clarify the reference of the preceding pronoun. As this was part of the *Morning Chronicle's* editorial decision-making for clarity, it then becomes part of *Hansard's*. Presumably Palmerston actually said “I”, as the *Mirror* records, and so the *Morning Chronicle* decides to render this directly as “he” – but due to the five possible male referents in the passage (admittedly not all referents being equally likely), they decide to insert a clarificatory parenthetical rather than restate Lord Palmerston’s name or give his ministerial title again during the actual report of the speech, which does happen elsewhere in *Hansard*.

Overall, then, in this complex period of *Hansard* from 1803-1888 (the unsubsidised period covered by T.C. Hansard senior and junior), we see the following linguistic characteristics:

- The dominant mode, speech-by-speech, is NRSA, generated by reports in the *Times*, the *Morning Chronicle*, and some others.
  - o For much of this period, there is a chance newspaper reports may give more detail than *Hansard* for NRSA reports of particular days or debates.
  - o From 1877 to 1888, subsidy by Parliament means that *Hansard* is more detailed than the newspapers on what T.C. Hansard junior calls the “four points”; “debates upon Private Bills set down by Order; the discussions in Committee upon

- Public Bills; the discussions in Committee of Supply; and the discussions which take place after midnight or half-past 12” (Parliament 1888a: 6).
- Sections of NRSA may appear FDS-like. See below for a discussion of NRSA which is adapted from FDS.
  - There are occasional editorial interventions, varying according to the source.
- The secondary mode, which is more common proportionally in the text than speech-by-speech, is FDS. This is reproduced where a “great occasion” of a speech occurs (in the judgement of the editor), *Hansard* may reproduce a period of direct speech (generally FDS) obtained most often by either taking the speech notes of the speaker (which T.C. Hansard would check against the notes of the speech to check for veracity), and sometimes by sending a report to the speaker for correction. It is highly unlikely all of a debate will be FDS, and it is most likely only the lengthier opening speeches of a major debate are represented as such, with later contributions (including those by the opening speakers) reverting to NRSA.
    - For 1828-1841, the *Mirror of Parliament* also provides comprehensive first-person speeches derived from shorthand of high accuracy.
  - Editorial points of style are not easily discernible as being that of *Hansard* rather than a source.

Therefore, if an individual speech in the corpus from this period is relatively long and high in first-person markers such as pronouns and the present tense, it can be processed with a reasonable degree of confidence as a FDS speech. By ‘relatively long’, the average length of a speech in both Houses across the Second and Third Series (excluding the very short First) is approximately 360 words, so anything above 500 could be considered comparatively ‘long’. Given the formal speaking context of the time it should be noted that such speech will be a reading of a highly-formal written-to-be-spoken text rather than spoken language *per se*. *Hansard*’s ability to cross-check a speaker’s notes with published versions taken from shorthand notes (after the mid-1820s), adds to the likely accuracy of such speeches.

The NRSA material, the majority of the corpus at this period, can be used as a description of major topics discussed – though not necessarily representative of the language in which it was discussed.

Some of what is registered as NRSA will actually be pronoun- and tense-shifted FDS (for an example, see Palmerston's 1840 speech above). With apologies to Leech and Short, I call this 'Narrator's Report of Direct Speech', or NRDS. As this does not present itself as direct speech its editorial standards of accuracy may not be the same as true FDS, and other than length of speech there are few simple linguistic indices to mark this type, so care should be taken in trying to distinguish NRSA from NRDS.

For both modes there should be some wariness attached to frequency counts in the corpus in this period. Content-heavy words in the open grammatical classes will be over-represented in sections of FDS and NRDS and so comparatively under-represented in sections in NRSA. However, the rationale that FDS and NRDS is used for important speeches and speakers means that the effect of this is likely to be mitigated somewhat. Thus a topic of significance, aimed to be measured by corpus frequency, will come up as high-frequency, and a topic of low frequency will still come up as low-frequency, but relative measures, especially in the middle of the frequency curve, will be distorted compared to a uniform corpus. As very few studies focus on such concepts or lexical items this may not pose a very significant issue – but for large diachronic studies discussing concepts which were important at one time but only of middling importance in another period, it is a consideration which must be taken into account.

## **6. *Hansard* without Hansard: The Chaotic *Authorised Edition*, 1889-1908**

[...] on certain days if the matters under discussion happen to be dry and do not appear to be interesting, the records of Parliament are practically nil. If the matter is sensational and the subject is one of great interest, then you have a full report, but not otherwise. Now, Sir, I maintain that in this House our records ought to contain not only the sensational but also the dry discussions.

The public, via Parliamentary subsidy (as well as subscriptions to the volume from individuals and libraries), would increasingly fund the production of *Hansard* after the retirement of the junior T.C. Hansard in 1888. The new operators naturally sought a profit from this enterprise; anyone involved in a large and long-term linguistic project could tell them that the great twin complexities of language and discourse are always more expensive and time-consuming than even pessimistic assessments estimate. When the subsidy contracts require a particular turnaround of revisions and publishing, a lack of time can only be compensated by yet more expense.

So it was with the publishers attracted by the idea of putting their name to one of the most prestigious publications of the age: fixing the nation's words of power in time and print by collecting them from reserved seats in the heart of Parliament, with the whole soon to be augmented with the title of the "Authorised Edition", in words suggestive of the King James Bible. Even after Parliament stepped in with limited subsidy, the cost of producing what Parliament both wanted and was willing to pay to subsidise could never match the income from that subsidy plus commercial sales. Members of both Houses continued to be dissatisfied with the report (Port 1990: 180), and each successive publisher took different approaches to attempt to satisfy their contract.

The remainder of the Third Series, from 1889 to 1892, was fulfilled by the Hansard Publishing Union, who purchased the rights from T.C. Hansard of the *Hansard* name in 1888 and then put in the lowest of 12 bids to a tender from Parliament for further subsidy. This subsidy came in return for certain changes to the content of the reports – as the contract of 1888 specified (signed by the immediate predecessor of the Union), "the Contractor shall exercise his own discretion as to the fullness of the Reports given, provided always that in no case shall any speech be reported at less than one-third of its length as delivered" (Parliament 1888b: 2). All speeches were to be recorded equally fully, regardless of the type of debate or Bill or its timing. This transformed *Hansard* from its pre-1888 form, and is why the 1889 date (rather than the beginning of the Fourth Series in 1892) marks the start of the second linguistic phase of the *Debates*.

The Hansard Publishing Union was short-lived, with its principal handicap being that it was run by a fraudster, who funded generous dividends to its shareholders from a series of large loans. Regardless, by employing dedicated reporters to fulfil its contract it was unlikely to make a profit: McBath (1970 :35) suggests that the Union’s investment may have been an error due to confusing the profitable government printing branch of the wider Hansard family (the Luke Hansard side) with the minimally profitable debates branch (the T.C. Hansards). With the Union’s collapse at the end of the 1891 session, a Fourth Series was begun in 1892 with the subtitle “Authorised Edition” and no mention of *Hansard* (all to avoid confusion with the now-disgraced Union; see Vice and Farrell 2017:25). As outlined in Table 1, a series of publishers then followed. Figures 2 and 3 clearly show the erratic linguistics of this period. Figure 4 shows the data from 2 and 3 within this 1889-1908 period in more detail as line graphs, with one year of context from the Third and Fifth Series on either side.

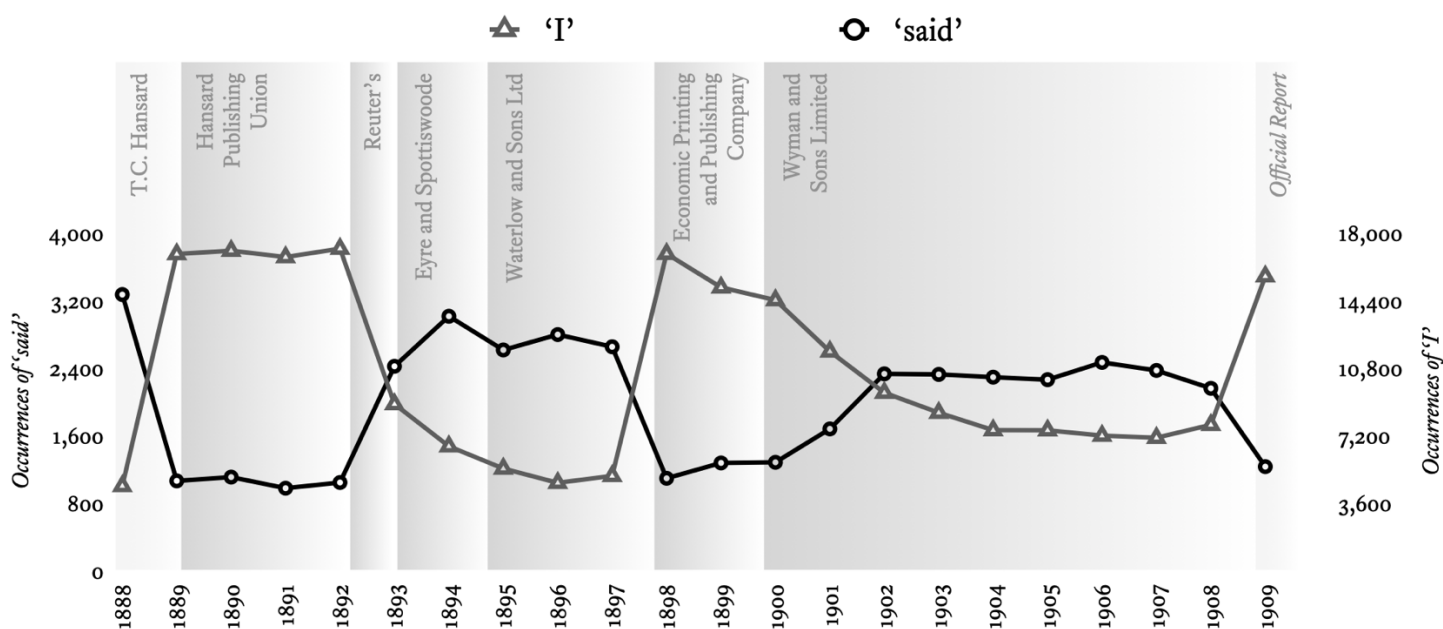


Figure 4, 'I' per million words and 'said' per million words in the *Hansard Corpus* between 1888 and 1909, with publishers indicated.

Taking the ratio of 'I' to 'said' as an index of first-person speech, there are clear shifts shown in in Figure 4: first from T.C. Hansard's primarily-NRSA accounts to the Hansard Publishing Union's first-person accounts, then Reuter's middling ratio during their brief involvement, and then the next two publishers, including Waterlow and Sons, returning generally to T.C. Hansard's earlier ratio of mainly NRSA. In 1898 the Economic Printing and Publishing Company, then a long period of Wyman and Sons Limited, begun a return to more first-person accounts, though these declined over the decade (most likely because of the high costs of accurate shorthand), with a rapid reversal with the start of the *Official Report*. The return to more third person NRSA reporting during the publishing period of Waterlow and Sons is easily explained:

One interesting experiment was made in the three years, from 1895 to 1898, when Messrs. Waterlow, the Government printers, were the contractors. The report was furnished by the staff of *The Times*, whose daily record of the proceedings in Parliament formed the chief source from which, under the old arrangements, *Hansard's Debates* were compiled. The reporting staff of *The Times* turned out two separate and distinct reports – one for *The Times* and the other for *The Parliamentary Debates*...

McDonagh [n.d.]: 433

*The Times* during this period stayed in the third-person NRSA style it generally adopted, and reporting in the third person except for major speeches would save the shorthand writers of the *Times* time in producing the two 'distinct' reports.

The challenges of this period preclude easy generalisations, but its general linguistic characteristics are:

- Each publisher has a broadly different dominant ratio of speech types, but with the exception of the two periods January 1889-February 1892 and January 1898-April 1899 the dominant mode is NRSA or NRDS, generated by newspaper reports and other reporters as in the *Hansard's Parliamentary Debates* phase. Occasional speeches are given in FDS following the pattern of the previous phase.



- From 1895 to 1898 speech notes would be prepared by the same reporters as reported *The Times*.
- From January 1889 to February 1892, the dominant mode is FDS from shorthand reporters.
- From January 1898 to April 1899, the dominant mode is FDS from shorthand reporters.
- As required by contract, speech reports could be anywhere between no less than one-third of its length as delivered to full-length as delivered.
- Given the chaotic publication history, editorial points of style are not easily discernible. Stability was only found by Wyman and Sons Limited for their nine-year period as publisher, but Figure 4 demonstrates how their style shifted across time as they sought a profitable balance (likely with a higher reliance on newspaper reports).

Any individual speech in this period is hard to categorise easily, but the characteristics above will apply. The concerns about relative frequencies from the previous section still apply, and could in places be even more erratic. Nonetheless, the same fundamental principle – that the key material, topics, and content words will shine through – also applies.

## **7. ‘Something like literary shape’: The *Official Report*, 1909 to present**

During the period of the Fourth Series, the House of Commons convened Select Committees on Parliamentary Debates (the Select Committee in 1888 that generated the new system of subsidy was jointly with the Lords). The 1893 committee supplied a consensus that there should be a full report “in all cases in the first person” which was “substantially verbatim” and with relatively minimal correction by Members (Parliament 1893: iv). Parliament being unwilling to dedicate the funds for this, the contractual arrangement continued with minor changes until the next Select Committee report in 1907. This Committee criticised the system of putting the reports out for tender, whereby financial incentives existed to lengthen or shorten volumes according to subsidy, whereby the lowest tender was taken up irrespective of

a detriment in quality, and which usually still relied on “newspaper cuttings” (Parliament 1907: v). At this point the financial arguments for an in-house report were more substantial than before – while the benefits were greater – and so on 14 May 1908, the reporters employed by Wyman and Sons recorded in the *Parliamentary Debates* the decision to end the contractual outsourcing of reports, beginning in 1909.

The 1893 committee recommended a full Report, “which, though not strictly verbatim, is substantially the verbatim Report, with repetitions and redundancies omitted, and with obvious mistakes corrected; but which, on the other hand, leaves out nothing that adds to the meaning of the speech or illustrates the argument” (Parliament 1893: 2). The 1907 report repeated this word-for-word (Parliament 1907: iii), and this phrasing remains *Hansard’s* terms of reference to the present day (Vice and Farrell 2017: 25). In the 1908 debate surrounding establishing the *Official Report*, the Financial Secretary who introduced the relevant Select Committee report gave a further gloss:

Perhaps I might just express to the House what is the meaning of the word ‘full,’ used in its technical sense. It means a verbatim report, trimmed of all those excrescences and redundancies with which Members are perhaps in the habit of filling up the matter of their speeches. In fact, a ‘full’ report puts into something like literary shape the efforts with which we endeavour to express our thoughts.

*HC Deb* 14 May 1908 c1357-8

Matters of reports being somewhere between a third-to-full-length of the original were therefore dismissed, and the Official Report from the outset was one of first-person full transcripts, transformed and tidied only as the terms of reference permit.

Of the 43 Select Committee on Publications and Debates reports (a Commons committee established from 1909 to 1967) which mention *Hansard*, almost all are concerned with publication dates and distribution processes (and, during wartime, about reformatting type and binding standards to save paper).<sup>5</sup> Of the 1,924 pages of reports in this series, there

---

<sup>5</sup> In addition to this, the most endearing part of these reports is their recurrent dedication to choosing the image for the House of Commons Christmas Card.

are only two mentions of matters of language and of the ‘trimmed’ nature of the reports. The second is the briefest, in a single paragraph in 1944:

It was brought to the notice of Your Committee that in the official report of Members’ speeches the words “shall not” were consistently printed for “shan’t”, and that similar slight verbal alterations were frequently made, so that emphasis and vividness were sometimes lost. After consideration Your Committee formed the opinion that the Editor of the Official Report should continue to exercise his discretion with regard to minor matters of grammar and spelling in Members’ speeches.

Parliament 1944: 3

No definition is given of “minor”, but in the general spirit of the *Official Report* – with which MPs and Lords alike were generally very pleased – the common sense of the editors is left to be the best judge.

An equally brief statement is given in the committee report in 1930, but with a far more detailed backstory:

Your Committee considered the question how far alterations and corrections made by Members in the text of the Official Report are admissible, but in view of the statement on this subject made by Mr. Speaker on the 14<sup>th</sup> November, they consider it unnecessary to make any recommendation with regard to such corrections.

Parliament 1930: iv.

The statement referred to here was made on 14 November 1929, in a brief debate of significance to the study of the language of *Hansard*. Hugh O’Neill MP, who was later to serve as Father of the House, disputed a change made by a government minister to *Hansard*, claiming the minister had openly stated that he “altered what I actually did say into what I ought to have said”. O’Neill continued, “Surely, if we are going to have official reports in this House, they should be reports of what is actually spoken in the House [...] The OFFICIAL REPORT ought to be a kind of arbiter whose relentless accuracy is beyond any question” (*HC Deb*, 14 November 1929, c2229). The minister responsible replied that “Owing to the confusion and noise, an announcement [by this minister] was not heard in the Gallery, and a slip was sent to be filled in”. (Such slips are common even now.) A second complaint was then raised by another MP, who objected that they had asked

supplementary questions based on the erroneous figures given by the minister, and that as the original statement was corrected, “my supplementary questions looked positively ridiculous in the OFFICIAL REPORT”. This, however, was not a complaint about *Hansard*’s practice; instead, the MP continued to say that if only he had been informed about the minister’s corrections, “I would willingly have made any arrangement that would have suited the hon. Gentleman, so that both our statements would at any rate have been consistent”. The intricacies of the debate reveal on the one hand a desire from some MPs that *Hansard* reflects the debate as said in the House, but on the other merely an interest that the key records of state are consistent. The Speaker of the House then proceeded to give an official statement about *Hansard* and the ability for members to make corrections:

As regards the first question that he raised, it is, unfortunately, the case that it is not given to all Members to speak with the grammatical precision which is desirable when the spoken word has to be transcribed for the OFFICIAL REPORT. An hon. Member obviously is entitled to correct an error which arises in his speech and to put the speech in grammatical form, but at the same time it is often the case that an hon. Member is imperfectly heard in the Gallery and that a word, perhaps a very important word, is omitted from his speech. Obviously, again, he is entitled to insert that word if it has been spoken in the House and has not been heard in the Gallery. I am ashamed to say that that has happened to myself. [...] if any great difference occurs between what is reported in the OFFICIAL REPORT and what an hon. Member thinks ought to be there, inquiries by the reporters in the Gallery are made, and great care is taken that no untrue report appears in the OFFICIAL REPORT.

*Ibid*, c2231

This is one of the most explicit statements *Hansard* records about itself in the era of the *Official Report*. While some MPs were concerned about the literal accuracy of the facts of *Hansard* (although the known alterations of grammar were not under dispute), some MPs were far less concerned about reflecting what was said and were more concerned with the consistency of the record, and the official position of the House was that errors in fact and grammar were permissible to be corrected, and that when a reporter’s evidence failed them or was imperfect, the Member speaking was the arbiter of what was and was not to be recorded. Practically speaking, nothing else could be reasonably counter-proposed, and so this statement by the Speaker

of the House, alongside the terms of reference of *Hansard*, is the firmest public statement available of the contents of the record of debates.

Accounts of the mechanism of creating *Hansard* in the twentieth and twenty-first century are widely available, and the editors and reporters from both Houses are lucid and enthusiastic communicators of the history, process, and structure of their roles and the place of *Hansard* – in line with the opening-up of Parliament in general to the public in recent years, the public have also been invited to learn more about the *Official Report*. Excellent sources from within include Vice and Farrell 2017, Sutherland and Vice 2015, Hansard Writing Team 2018, Sutherland and Farrell 2013, and Browne 2017.

While the mechanism of creating the *Official Report* has changed with the evolution of technology in the twentieth century, the fundamentals have remained the same. Reporters take ‘turns’ in the House of five to ten minutes each, then are responsible for producing the report for that timed period (nowadays within 45 minutes of the end of a five minute turn). From 1909 until the 1970s, this consisted of shorthand notes, then reporters could work instead from audio tapes, and now reporters use digital audio as their source, creating the report directly from them. A brief period of correction or clarification is allowed, with reporters having the ability to pass notes directly to speakers asking for their speech notes or for clarity on particular points.

The view of modern Parliamentarians is that *Hansard* remains:

...faithful to the Members’ words, accurately conveying the nuance of their argument and preserving their speaking style, while also, with the slightest of editorial touches, producing a fluent and readable report that will serve as a working document, a legal record and a historical resource.

Bercow 2017: i

This objective is also aided by the rapid accessibility of the *Official Report* online, on the same day as the speeches it records.

Other authors, notably Mollin (2007) and Slembrouck (1992) have analysed modern *Hansard* in comparison to the audiovisual sources of the Chambers in action, and have discussed what features of spoken speech do not survive

the process of producing the ‘full’ report (notably repetition, hesitation phenomena, the generic *you*, contractions, the *going-to* future, and often intensifying adverbs and proximal determiners). In the present volume (Kotze, Korhonen *et al.*), an analysis of the Australian *Hansard* between 1946-2015 also reveals the effects of editorial practices in the transcription in that Parliament. Without retreading the work of these writers, in the course of working on the linguistics of the British *Hansard* I have been permitted to informally shadow reporters working ‘live’ on occasion over the past seven years in Parliament, and while a fuller analysis is forthcoming there are some initial observations which are relevant for this article.

Firstly, it is interesting from a linguistic perspective that while the audio of the reporter’s turn is available digitally, a reporter must still attend the relevant House for the duration of their turn. This is essential for the interpersonal situatedness of the debate (cameras will not necessarily be at an appropriate angle to see to where a speaker gestures or points, for example) as well as to take notes of key words and to hear the opening words of each speaker’s turn (as all of the hundreds of microphones in the Houses are not permanently ‘live’, the process of recognising a speaker is speaking means that the first few words of their speech, particularly in the House of Lords, is wholly audible in the chamber and not necessarily audible on a recording). Interventions – only when they merit a response from the current speaker – are included and these are not often picked up by microphones or cameras. Reporters also note key terms in the debate, particularly those which would be worth looking up or requesting clarification on before undertaking the production of the report, such as when a speaker mentions a place, or a report, or (on one occasion) their dog. Members seem to be generous with their speech notes, but reporters refer to them while working with audio rather than unquestioningly copying them.

Producing the report itself involves a significant amount of procedural text, which is the main part of a speaker’s words that are reported ‘as they should be’ rather than as they actually were; if a speaker makes a minor error such as a wrong form of address for a colleague or of Parliamentary process (such as saying ‘this chamber’ instead of ‘this Committee’ or ‘this House’, or ‘you’ instead of ‘your Lordships’, or referring to the Government in the singular, or producing direct second-person speech), then these are not unnecessarily

recorded and exposed but silently altered to fit agreed protocol. Hesitations, false starts, verbal tics, speech errors, stutters and similar phenomena are removed.

Other changes which do not always happen but have been observed include changing 'this' to 'the' and 'among' rather than 'amongst', moving adverbs like 'only' elsewhere in a sentence for clarity, removing much otiose repetition (such as 'very very'), removing deictics not necessary to the meaning, removing fillers (such as 'I mean' or 'so') and removing many sentence adverbs when they are likewise not key to the meaning of the sentence (such as parenthetical 'actually' or 'really'). Particularly interesting for linguists is the common removal of otiose hedges (such as 'I believe' and 'I think') when these do not indicate actual commitment to the sentence – reporters are careful to retain hedges where they are explicit indicators of perception, which often requires careful judgement. For this sort of decision, reporters can spend a fair amount of time working through difficult matters, often consulting colleagues and with the ability to consult more senior editors if necessary.

While there is always the desire from many researchers, discussed above, to have a 'full' transcript including all of these omitted features, it is difficult to criticise any of the changes made. As an item of linguistic anthropology, the more I observed the reporting staff, the more I also came to understand the ways in which effort is taken not to level each speaker's style; stylistic features are still allowed to be present in the record, although with some of the more stylistically-explicit markers removed. Reporters discussed instances where a speaker was making a clear point through repetition or redundant phrases – such as when aping overly bureaucratic language – and so such repetition would 'naturally' be kept in. At least twice I asked why a larger change was not made to a speech and the editor responded that such a change would be 'gratuitous'. The phrasing used by John Campbell above (1881: 105-7), that a reader should not be perplexed seems to be the principal motivation behind every change, while not permitting errors of fact or the sense of what was said to be altered. (A 'corrections page' is produced in *Hansard* for any errors of fact to be later corrected; see Parliament 2007.)

The period of the *Official Report* is much simpler to summarise than the previous periods. It is FDS throughout, except where in the first half of the period a formal question is asked by a speaker of a Minister where there is a small section of IS (this is a formulaic section, used when a question is asked with an initial ‘to’, such as ‘to ask the Secretary of State for Health whether...’, and rendered as ‘[speaker name] asked the Secretary...’). On the matter of editorial points, house style has clearly evolved – the modern British *Hansard* does not edit out split infinitives anymore, for example – and a style guide does exist; future research planned on this area will investigate that particular question in more depth, but editing due to house style generally limits itself to the types of changes listed above. The content nouns, adjectives, and verbs spoken in a debate are very well recorded in the *Official Report*, with the exception of compound verbs of process, such as ‘give an assurance to’ or ‘be opposed to’, which would generally be changed to ‘assure’ or ‘oppose’. The post-1909 report is therefore rigorously and semantically full for main content.

## 8. Conclusion: *Hansard* for eternity, *Hansard* for linguists

Finality is not the language of politics.

Benjamin Disraeli, *HC Deb* 28 February 1859, c998

The preceding discussion has investigated in detail what the language of *Hansard* over its two-century run represents. The purpose of *Hansard* is to be a record for the Parliament of the day and for the ages, not to be a linguistic corpus, and a modification of that purpose requires a detailed knowledge of the content, which this chapter aims to supply.

The story of *Hansard*, as we have seen, is one of improvements in quality and reductions in internal heterogeneity. The analyses above give the following outline:

**1803-1876** Primarily NRSA/NRDS (NRSA often taken from summaries in the *Times*, *Morning Chronicle*, and other newspapers) with some longer speeches in FDS where the speech is particularly significant (often taken from a speech text, when available). Compared to the speech events in Parliament, corpus frequency results from *Hansard* will be



relatively overrepresented in the FDS/NRDS sections which editors consider a 'great occasion'.

**1877-1888** As above, but with subsidy generating a) more detail on committee discussions and other areas newspapers do not report for lack of interest and b) more details post-midnight. FDS/NRDS sections still fairly overrepresented, but a larger volume of NRSA is proportionally present from the subsidised topics.

**1889-February 1892** FDS from shorthand. Compression of unimportant speeches. Corpus results likely to be overrepresented for speeches of importance which will be reported in more detail.

**March 1892-1897** As 1877-1888. FDS/NRDS sections still relatively overrepresented.

**1898-April 1899** FDS from shorthand. Speeches of importance likely overrepresented per 1889-1892.

**May 1899-1908** Generally as 1877-1888 but variable in ratio of NRSA/NRDS to FDS. FDS/NRDS sections proportionally overrepresented.

**1909-** FDS from shorthand and audiovisual recording, occasional formulaic IS sections decline by 1970s. No over-/under-reporting.

In general, those speeches thought to be contemporaneously more significant from 1803 to 1909 will be overrepresented in a corpus, either as an editorial choice (marked in the NRSA-form phases by a shift to FDS and in the FDS-form phases by longer text) or as a source reporter/editor's choice (for NRSA-form through NRDS, for FDS-form by longer text). Debates contemporaneously considered to be less important will be proportionally underrepresented. The presence of topics (through either lexical items or semantic tags) will be well-recorded, but their relative frequency in this period is relatively uncertain. Research which focuses on topics of major contemporaneous importance (such as war, foreign affairs with major powers, major expenditure, etc) has more solidity in terms of frequency due to the likelihood of results being within the relatively overrepresented category. Frequency results for matters which were not of contemporaneous interest (including much regional matters not affecting London) should be treated with suspicion. Useful qualitative support can be sought from the *Mirror of Parliament* for 1828-1841. Useful comparisons can also be gained by contrasting material from 1880-1908 with the specific periods of 1889-February 1892 and 1898-April 1899 as the nature of reporting and selecting

material shifted; a lack of significant change in this period is likely to indicate frequency counts are more robust than otherwise.

In the period 1909 to present, a ‘full report’ omits sociolinguistic and conversation-analytic phenomena, elides grammatical variation and information structure, and generates a written-to-be-spoken text from written-to-be-spoken and spontaneous speech alike. It is substantially a highly accurate reflection of spoken semantics, and therefore of content.

Those factors of publishing history, editorial policy, and shifting priorities affect how linguists, humanities scholars, and social scientists should view and use the text of *Hansard*. It leaves a vast range of research questions still open, and much easily addressable using big data corpus techniques. As well as being the first textual history and linguistic analysis of speech representation in *Hansard*, it is hoped the present chapter can act as a handbook to the text for future research. The accumulated work of two centuries and many hundreds of hands – T.C. Hansard senior and junior, the army of *Times* and other newspaper shorthand reporters, the early *Hansard* collators, those reporters employed by the many commercial publishers entrusted with the work, and as the careful and judicious staff of the *Official Report* – deserves not only its established pre-eminence in history but also careful curation throughout its growing range of further digital afterlives as an immense quantitative source of information on British culture, language, and society.

### **Funding**

This work was supported by the Leverhulme Trust (grant PLP-2019-341). The Arts and Humanities Research Council in conjunction with the Economic and Social Research Council (grant AH/L010062/1) funded the creation of the annotated *Hansard Corpus 1805-2003*.

### **Acknowledgements**

This chapter was enriched by the comments of two anonymous reviewers and by generous advice from Jeremy Smith, Andrew Struan, and John Vice. For the purpose of open access, the author(s) has applied a Creative

Commons Attribution (CC-BY) licence to any Author Accepted Manuscript version arising from this submission.

## Bibliography

- Alexander, Marc & Andrew Struan. 2013. 'In countries so unciviliz'd as those?': the language of incivility and the British experience of the world. In Martin Farr & Xavier Guégan (eds.) *The British Abroad Since the Eighteenth Century, Volume 2: Experiencing Imperialism*. London: Palgrave Macmillan. 232-249.  
doi:10.1057/9781137304186\_13
- Alexander, Marc & Andrew Struan. 2022. "In barbarous times and in uncivilized countries": Two centuries of the evolving uncivil in the *Hansard Corpus*. *International Journal of Corpus Linguistics* 27(4): 480-505.
- Alexander, Marc & Mark Davies. 2015. *The Hansard Corpus, 1803-2005*. Available online at <http://www.english-corpora.org/hansard>.
- Alexander, Marc, Fraser Dallachy, Scott Piao, Alistair Baron & Paul Rayson. 2015. Metaphor, popular science, and semantic tagging: Distant reading with the *Historical Thesaurus of English*. *Digital Scholarship in the Humanities* 30(s1). 16-127, doi: 10.1093/llc/fqv045
- Anderson, Jean. 2013. ENROLLER: An experiment in aggregating resources. In Wendy Anderson (ed.) *Language in Scotland: Corpus-Based Studies*. Amsterdam: Rodopi.
- Archer, Dawn. 2017. Mapping *Hansard* Impression Management Strategies through Time and Space. *Studia Neophilologica* 89(s1). 5-20.  
doi:10.1080/00393274.2017.1370981
- Barrow, John Henry. 1828. Prospectus. In John Henry Barrow (ed.) *The Mirror of Parliament* vol. 1. London: Winchester and Varnham.
- Bercow, John. 2017. Foreword, in Vice and Farrell 2017.
- Bevan, Aneurin. 1952. *In Place of Fear*. Kingswood: Heinemann.
- Brown, Everett S. 1955. John Henry Barrow and the *Mirror of Parliament*. *Parliamentary Affairs*, 9(3). 311-323.
- Browne, Charlie, 2017. Being a *Hansard* reporter: 10 things you thought you knew. *Commons Hansard Blog*, 24 July 2017.  
<<https://commonshansard.blog.parliament.uk/2017/07/24/being-a-hansard-reporter-10-things-you-thought-you-knew/>> (1 February 2021).
- Campbell, John, 1st Baron Campbell. 1881. *Life of John, Lord Campbell, Lord High Chancellor of Great Britain; consisting of a selection from his autobiography, diary, and letters. In two volumes*. Ed. by the Hon. Mrs Hardcastle. 2nd edn. London: John Murray.

- Cobbett, William. 1804. *Cobbett's Weekly Political Register*, vol V. London: Cox and Baylis.
- Coole, Matthew, Paul Rayson & John Mariani. 2020. Unfinished Business: Construction and Maintenance of a Semantically Tagged Historical Parliamentary Corpus, UK Hansard from 1803 to the present day. In *Proceedings of ParlaCLARIN II Workshop*, Marseille, France. 23–27.
- Culpeper, Jonathan, and Merja Kytö. 2010. *Early Modern English Dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.
- Cunningham, Hamish, Valentin Tablan, Angus Roberts & Kalina Bontcheva. 2013. Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. *PLoS Comput Biol* 9(2): e1002854. doi:10.1371/journal.pcbi.1002854.
- Hansard, Thomas Curson (junior). 1828. *Hansard's Parliamentary Debates* [advertisement]. In Sylvanus Urban (= Edward Cave, ed.) *The Gentleman's Magazine and Historical Chronicle*, vol. xcvii (January-June 1828). London: Nichols and Son. Sections separately numbered; Hansard's advertisement begins approx. 800 pages in.
- Hansard, Thomas Curson (junior). 1843. Hansard's Parliamentary Debates [advertisement appended after the formal end of the debates volume]. In *Hansard's Parliamentary Debates*, Third Series, vol. LXXI (31 July-24 August 1843). London: Thomas Curson Hansard. 1-4.
- Hansard Writing Team. 2018. *The Hansard Story*. London: House of Commons Library.
- Hughes, Lorna M. Forthcoming, 2023. 'The ties that bind: curation and creation in the digital humanities', in Prescott, Andrew, Lorna M. Hughes, Claudine Moulin & Edward Vanhoutte, *The Cambridge Companion to the Digital Humanities*. Cambridge: Cambridge University Press.
- Jordan, H. Donaldson. 1931. The Reports of Parliamentary Debates, 1803-1908. *Economica* 34. 437-49.
- Jupp, Peter. 1998. *British Politics on the Eve of Reform: The Duke of Wellington's Administration, 1828-30*. London: Palgrave Macmillan.
- Leech, Geoffrey, Roger Garside & Michael Bryant. 1994. CLAWS4: The tagging of the British National Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)* Kyoto, Japan. 622-628.
- Leech, Geoffrey & Mick Short. 2007. *Style in Fiction*, 2nd edn. London: Routledge. [Reprints certain unaltered chapters from the 1981 first edition.]
- MacDonagh, Michael. [n.d., likely 1913]. *The Reporters' Gallery*. London: Hodder and Stoughton.
- McBath, James H. 1970. Parliamentary reporting in the nineteenth century. *Communications Monographs*, 37(1).

- Mollin, Sandra. 2007. The *Hansard* hazard: Gauging the accuracy of British parliamentary transcripts. *Corpora* 2 (2). 187–210.
- Murphy, Arthur. 1792. *An essay on the life and genius of Samuel Johnson LL.D.* London: Longman *et al.*
- Parliament. Select Committee on Parliamentary Proceedings. 1862. *Report*. HC 373. London: Her Majesty's Stationery Office.
- Parliament. Joint Select Committee of the House of Lords and House of Commons. 1888a. *The Cost and Method of the Publication of Debates and Proceedings in Parliament*. HC 284. London: Henry Hansard and Son.
- Parliament. House of Commons Papers. 1888b. *Copy of contract between MacRae, Curtice & Co., Limited, and the Controller of Her Majesty's Stationery Office, for preparing, printing, and publishing reports of debates and other proceedings in both Houses of Parliament, to commence from the 1st day of January 1889*. HC 445. London: Henry Hansard and Son.
- Parliament. Select Committee on Parliamentary Debates. 1893. *Report*. HC 213. London: His Majesty's Stationery Office.
- Parliament. Select Committee on Parliamentary Debates. 1907. *Report*. HC 239. London: His Majesty's Stationery Office.
- Parliament. Select Committee on Publications and Debates Reports. 1930. *Report*. HC 164. London: His Majesty's Stationery Office.
- Parliament. Select Committee on Publications and Debates Reports. 1944. *Report*. HC 110. London: His Majesty's Stationery Office.
- Parliament. House of Commons Procedure Committee. 2007. *Report*. HC 541. London: Her Majesty's Stationery Office.
- Piao, Scott, Fraser Dallachy, Alistair Baron, Jane Demmen, Stephen Wattam, Philip Durkin, James McCracken, Paul Rayson & Marc Alexander. 2017. A time-sensitive *Historical Thesaurus*-based semantic tagger for deep semantic annotation. *Computer Speech and Language* 46. 113-135. doi: 10.1016/j.csl.2017.04.010
- Port, M. H. 1990. The Official Record. *Parliamentary History* 9. 175-83.
- Rayson, Paul, Dawn Archer, Scott Piao & Tony McEnery. 2004. The UCREL semantic analysis system. In *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic Labelling for NLP Tasks in Association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal. 7–12.
- Semino, Elena & Mick Short. 2004. *Corpus Stylistics: Speech, Writing and Thought. Presentation in a Corpus of English Writing*. London: Routledge.
- Slembrouck, Stef. 1992. The parliamentary *Hansard* 'verbatim' report: The written construction of spoken discourse. *Language and Literature* 1 (2). 101–19.

- Sutherland, Lorraine & Stephen Farrell. 2013. *The History, Workings and Future Challenges of Hansard*. UK Parliament Open Lecture, 11 December 2013. <<https://www.youtube.com/watch?v=yaLWV9gfiSo>> (1 February 2021).
- Sutherland, Lorraine & John Vice. 2015. *Hansard: The Ear of History*. London: House of Lords Library and House of Lords Hansard.
- Thomas, P.D.G. 1959. Sources for Debates of the House of Commons 1768-1774. *Bulletin of the Institute of Historical Research* (special supplement 4).
- Vice, John. 2018. Charles Dickens and Gurney's Shorthand: 'That savage stenographic mystery'. *Language and History* 61. 77-93.
- Vice, John & Stephen Farrell. 2017. *The History of Hansard*. London: House of Lords Library and House of Lords Hansard.
- Wattam, Stephen, Paul Rayson, Marc Alexander & Jean Anderson. 2014. Experiences with Parallelisation of an Existing NLP Pipeline: Tagging *Hansard*. In *Proceedings of the 9th International Conference on Language Resources and Evaluation* (LREC '14), Reykjavik, Iceland. 4093-4096.