The use of statistics and facts in contemporary political discourse

carr
centre for analysis of risk and regulation

carr report for the UK Statistics Authority

THE USE OF STATISTICS AND FACTS IN CONTEMPORARY POLITICAL DISCOURSE

Martin Lodge and Miran Norderland

July 2017

carr
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Email: risk@lse.ac.uk

Tel: + 44 (0) 20 7955 6577
Fax: + 44 (0) 20 7242 3912

Website:
http://www.lse.ac.uk/accounting/CARR/home.aspx

CONTENTS

INTRODUCTION

Questions about the use of statistics in the world of executive and electoral politics have received renewed attention over the past few years. In the UK, this was driven in part by concerns about the 'facts' that were marshalled by the respective sides of the Brexit debate, especially regarding the 'We send the EU £350 million a week, let's fund our NHS instead' bus advertising. Concern about the use of facts and statistics was combined with statements that questioned the role of 'experts' in informing policy conversations. Similarly, international election campaigns have given rise to concerns about 'post-factual democracies' and problems of moving political debates beyond social media 'echo chambers'.

This interest in the use of numerical language or statistics in political language is far from new and links to questions about the motivation and capacity of the messenger to appeal to their audiences, as illustrated in concern about deliberate misrepresentation in order to mislead target audiences. For example, it was said that the 2015 UK election was a 'statistical crossfire' in which different parties offered carefully crafted numerical statements to appeal to their voters. Concern about the quality of numerical claims has been further advanced due to the rise of Twitter as part of political contest and the limited means to communicate complex issues via 140 characters. Furthermore, there is concern about the ways in which data visualisation offers powerful tools to shape (and decontextualise) understanding of reliable and unreliable statistics.

Other trends might arguably have further added to a growing interest in 'numbers' in political statements. One such trend is said to be the rise of 'quantification' in the control of public services, whether this relates to questions of quality, efficiency or redundancy. Numerical statements might be said to give the impression of control.

This research cannot comment on whether there has been a rise in misleading statements or the association of numerical statements with particular themes. Instead, the research here focuses on the more straightforward question as to whether we can observe a rise in the use of numerical statements in political language over time by focusing on different data-chambers, namely major political speeches (1967–2016), speeches in parliament by parliamentarians (1967–2017) and, separately, by leading politicians (2010–17), tweets by MPs (2015–17) and government communications (2010–17) more broadly. We also briefly cover the 'demand' for numerical statements as indicated by 'Google trends' (2008–17). The term data-chamber highlights the study's interest in observing key words in certain self-contained areas. The chosen time periods reflect data availability. The research focuses on the UK in particular, focusing on 'macro' trends as well as select departments and senior politicians.

BACKGROUND

The interest in whether there has been a rise of numerical statements in political discourse links to a number of critical concerns in the contemporary study of public policy in general. As noted, they link to questions about (a) the value attached to statistical language as a rhetorical tool to lend authority to political statements, (b) broader control and management tools in public services, and (c) the wider issues on the role of expertise in policy argumentation.

One potential explanation for an increased use of numbers is to point to wider changes in public management, especially the move towards output and outcome performance management. At one level, the increased use of numbers in political language could therefore represent a growing shift towards managerialism in politics and therefore attempts by politicians to appeal to 'objective facts' in order to reduce the potential for political contests. In other words, the use of numerical statements points to an appeal for 'responsible government'. This shift as part of a wider interest in performance management has been accentuated by the promotion of 'big/open' data across public and private sectors.

A related explanation would point to the growing availability of numbers to assess individual organisational and system-wide performance as a source for growing political conflict. The use of numbers can also be said to have become increasingly politicised (rather than technocratised/managerialised). Interpretations of statistical estimates become political battlegrounds (such as the interpretation of the badger-culling trials) or different sources of statistics are put forward to make political points (such as the use of different crime statistics).

A third potential explanation points to the use of new forms of direct political communication. The recent US presidential elections and the Brexit referendum are said to have given rise to concerns about 'post-factual democracy'. The explanation for a growing use of numerical statements would be to highlight the ease with which they can be used in shortened political statements, especially in forms of a tweet. This, it might be argued, further advances the 'temptation' to carefully craft numerical statements that can easily be used for catchy headlines or 140 character tweets – and to data visualise 'accurate' statistics in ways that are misleading. For example, ONS data may be used to establish credibility, but in misleading ways. This, in turn, discredits 'good' statistics as the statistics rather than their communication become part of a political battle.

A fourth explanation would point to a more differentiated pattern. According to a 'blame avoidance' explanation, one would expect numerical statements to emerge in those areas where they can be used as a 'teflon' strategy. In other areas, where statistical trends over time might be used by opponents or where statements could become 'hostages of fortunes', we would expect either a low level of numerical statements or, at least, no significant change in frequency. By

drawing together different data-chambers, subsequent analysis is able to investigate whether the same actor (individual or organisational) exhibits similar frequency patterns.[1]

Regardless of explanation, the use of numerical statements in political language raises a number of issues for regulation. One is the question of *transparency*, namely to ensure that numerical statements display (in an accessible way) the appropriate level of information that its source and calculations are replicable. Another is the question of *accuracy*, namely the checking as to whether the claim is factually correct or part of a 'post-factual' arrangement. The third is *frequency*, namely the extent to which numerical statements are made, in which form and through which channel. Fourthly, there is also the question of *traceability* – is it possible to recover numerical statements once they have been made? This reflects questions about the removal of information (deleted tweets, content removal). Competencies are therefore required to deal with 'forensics' (the need to detect information even if it has been removed) and 'digital skills' more generally to identify digital footprints and their diffusion, even if it includes removed or re-categorised content. Finally, there is also the separate dimension, namely the extent and timeliness of 'response' – as debates about the responsibilities of social media providers suggest, concerns exist about the spreading of inaccurate numerical (and other) statements. Questions arise on how quickly state and non-state based organisations can respond to numerical statements, judge their accuracy and transparency, and not be accused of 'wrongful' interventions in political contests that violate rights of 'free speech'.

The challenges of transparency, accuracy, frequency and traceability represent the questions for regulatory oversight, in terms of information-gathering and behaviour modification as well as standard-setting. The wider question is how reactive or proactive regulatory oversight should be and whether a regulatory approach should take on pro-active interception strategies in this kind of communication environment. Such wider questions also include judgements about the appropriate level of damage control, in terms of addressing the potentially disproportionate effects of misleading information as it spreads across the wider communication environment.

[1] In addition, this introduces the question what the individual and cumulative effect of the use of numerical statements through one or multiple communication channels on any one individual recipient is.

OUR RESEARCH

This research focuses on the frequency of numerical statements in political discourse. It does so by focusing primarily on changes over time across different data-chambers, addressed to different audiences. Investigating changing levels of, or even 'punctuations' in, frequency addresses a fundamental underlying concern in the debate about the use of numerical statement; after all, should the number of numerical statements be said to be low and declining, then it might be said that the attention over the 'post factual democracy' (and appropriate ways of handling this phenomena) is over-blown. In contrast, growing (and differentiated) patterns would suggest the need to devote increasing resources on 'statistics checking' and debates about the appropriate organisations to conduct such 'statistics checking' in a timely and impartial manner. This might be particularly the case if the data-chambers (such as Twitter and Google) reveal different patterns than traditional data-chambers, such as parliamentary speeches.

To explore frequency of numerical statements in political statements over time requires a number of research design decisions. One key choice relates to data-chambers. Politicians and governments communicate in different formats and to different audiences. It is, for example, frequently said that politicians are increasingly resorting to social media in order to communicate directly to their supporters. Similarly, there are differences in constitutional norms as to when a senior politician makes statements as 'government' or as 'party politician' (and similarly, there are key issues regarding the involvement of civil servants in the preparation of statements). There is also limited recognition as to which capacity a particular individual is making a statement. This research therefore focuses on a diversity of data-chambers. Each one allows for different time-durations to be explored. These are:

a. **Government communications**. The website GOV.uk offers an archive of government communication in different forms – ranging from speeches of ministers to press releases – since 2010. This allows us to focus both on the trajectory over the course of the past seven years, and the use of numerical statements in certain key departments. We focus on the period 10 May 2010 to 18 April 2017.

b. **Major political speeches**. In particular, we concentrate here on major speeches as provided by the British Political Speech archive that offers major party conference speeches for the time scale of over a century. We focus on the period 1967–2016.

c. **Speeches in Parliament**, in particular the House of Commons. Hansard offers a range of (incompatible) forms of archives for different time periods. We distinguish between speeches by select senior politicians (10 May 2010 – 18 April 2017) and the use of numerical statements made in Parliament in general for period 1 January 1967 – 18 April 2017.

d. **Twitter** – this novel form of communication allows this study to focus on the frequency of numerical statements in individual tweets by UK MPs for the period of 10 March 2015 – 18 April 2017 (the period of the 56th UK Parliament).

e. In addition, in order to have some form of insight into the public 'demand' for numerical statements (measured in terms of attentiveness to the study's keywords), this study also assesses **Google trends** as one indicator as to how populations seek orientation by searching for 'facts'. We focus on 'Google trends' in Anglosphere countries – Australia, Canada, New Zealand, UK, USA – and 'worldwide' for the period 1 January 2008 – 18 April 2017.

The second key design dimension is the choice of numerical statements. It is impossible to gain a complete picture of all statements that are in some form or another associated with a numerical association. Similarly, there might be turns of phrase in the past that are no longer frequent in contemporary speech. The 'baseline' of words can also change over time. In the following, this study focuses on a number of keywords that have been found to be associated with numerical statements. In addition, while of relatively recent origin, this study explores whether the 'home' of statistics in the UK context has in itself become an increasingly frequent reference point in political statements. The keywords are:

a. Statistics
b. Facts
c. Figures
d. Numbers
e. Billion
f. Million
g. Thousand
h. Rates
i. Percentage
j. Numerical reference *(a figure that contains at least one zero)*
k. Office of National Statistics
l. UK Statistics Authority

The Appendix offers short research protocols for each of the different data-chambers. The following analysis looks across a different set of communication channels or data-chambers. This offers insights into how particular keywords – such as numerical statements – are treated in different settings where the same the same actor is communicating under particular conditions and with distinct audiences (and conventions) in mind. The following looks at these different chambers or channels separately as this offers a more fine-grained analysis of patterns. Future analysis might explore the interaction (if any) between the different channels. To assess whether keyword attributes change depending on peculiarities of a particular source, a standardisation protocol would be required that allows for data from different sources, scale and context to be uniformed. Future analysis might also conduct a 'tone' analysis or seek to explore the association of numerical statements with particular topics (such as 'inflation', 'unemployment' and such like).

Is there a growing demand for ‘facts’?

PUBLIC ATTENTIVENESS

Data-Chamber #01: Google Trends

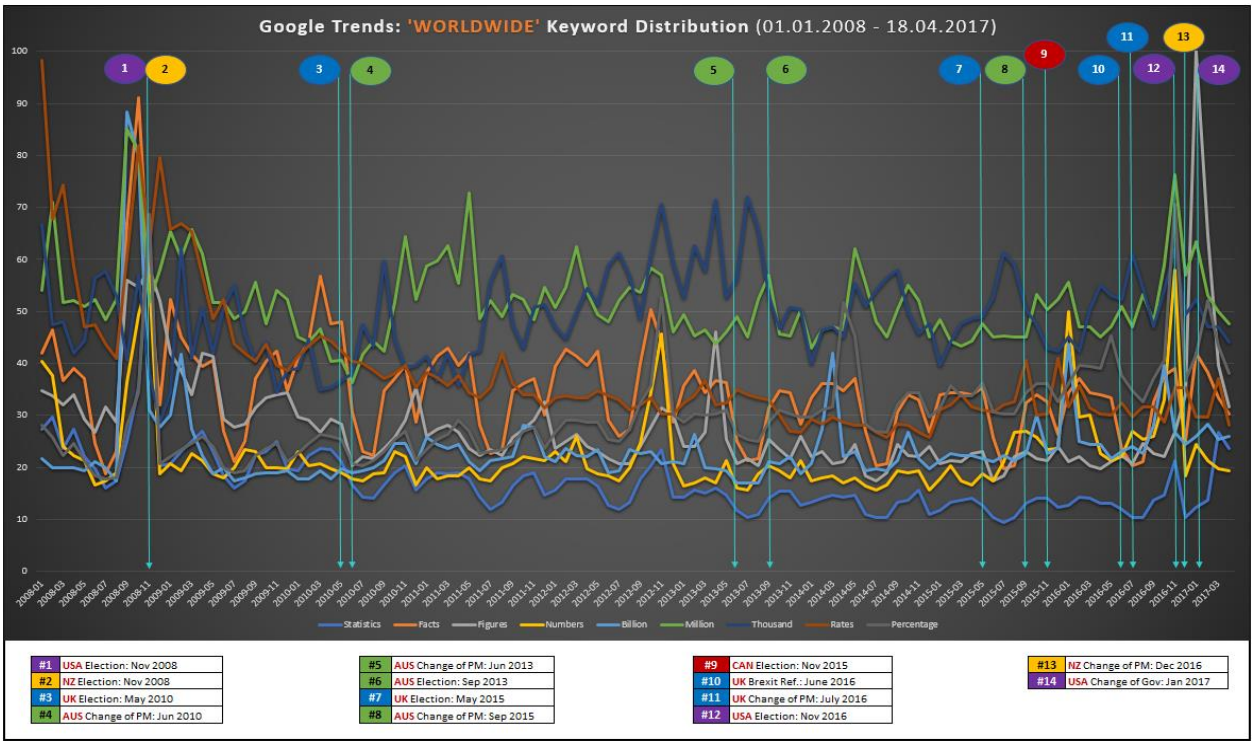
Observed Time-Line:	**1 January 2008 – 18 April 2017**
Source:	**Google search engine** (Google Trends database)
Data Unit:	**Keyword search interest** Averaged score for each Keyword using 3 filters: (1) All Categories + Web Search; (2) News Category + Web Search; and (3) News Category + News Search
Domain/s:	**Worldwide – Australia – Canada – New Zealand – UK - USA**
Keywords tested:	**Statistics, Facts, Figures, Numbers, Billion, Million, Thousand, Rates, Percentage**

To begin, the study looks at publics' 'demand' side. This angle provides for an insight as to whether there is a growing call for 'facts' or numerical statements, as expressed in google searches featuring our keywords globally and in select (English-speaking) countries. The search tool allows for an assessment for the time period January 2008 to April 2017. We investigate the UK, the US, Australia, New Zealand and Canada – we also look at 'worldwide' to examine how this select Anglosphere compares with global search trends.

The analysis was guided by 'Google trends'. While this tool offers a unique device to assess fluctuations in the 'searches' using particular words, the underlying methodology is notoriously non-transparent. According to Google's own definition, trends data are divided by the total searches of the geography and the time range. The resulting numbers are 'then scaled on arrange of 0–100 based on a topic's proportion to all searches on all topics'.[2] Total numbers of searches are therefore not indicative of total search volumes across different regions. Therefore, trend data in the context of this study offers insights as to whether there is, at particular times, a growing proportion of searches that use particular keywords that qualify as numerical statements.

[2] Google Trends methodology: https://support.google.com/trends/answer/4365533?hl=en (last accessed 7 June 2017).

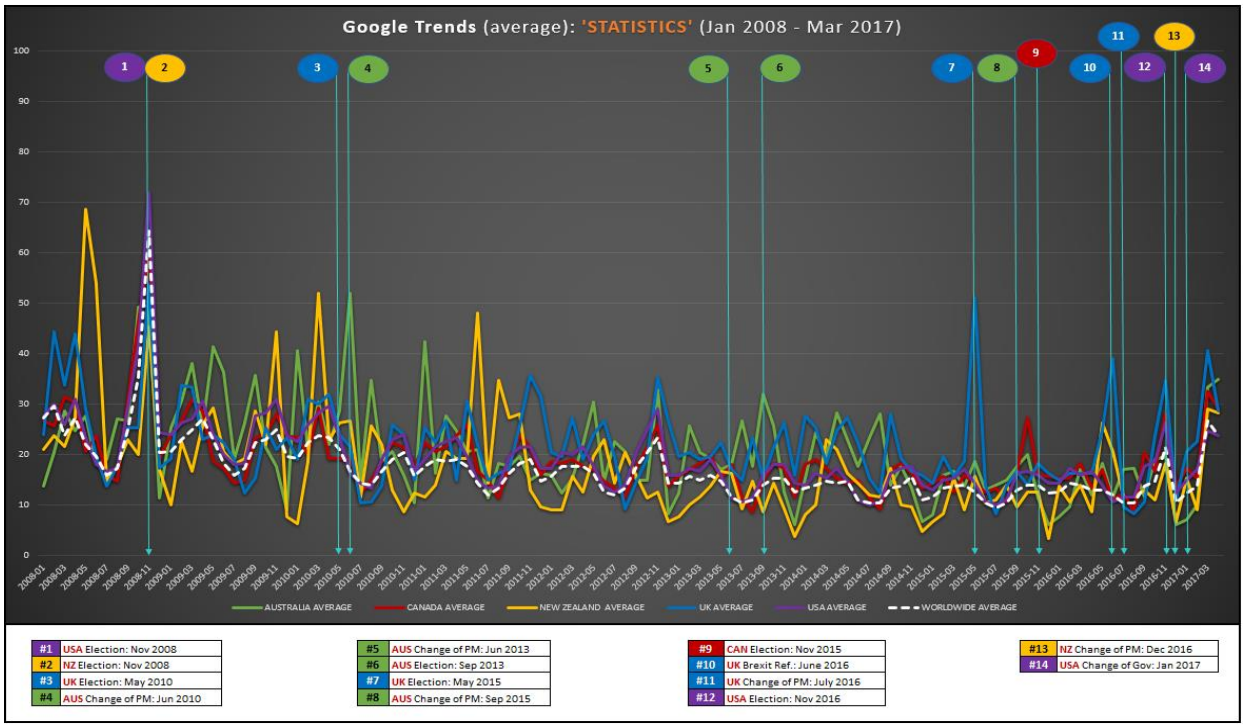
Graph 1. Google trends: 'worldwide' keyword distribution (1 January 2008 – 18 April 2017).



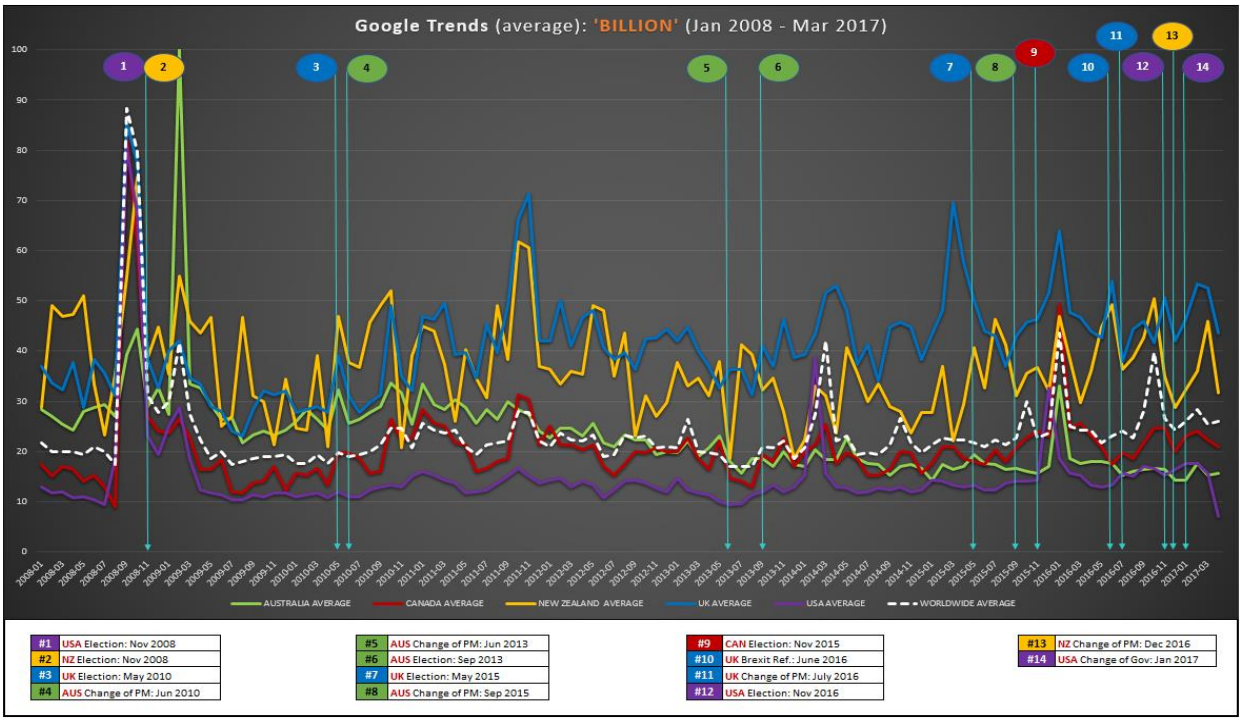
The graph above highlights the global searches for particular English-language keywords in global context. The vertical lines represent key political events, such as national elections and referenda or changes in government across the Anglosphere countries. Each country and search word has been provided with its own colour coding. The rather 'peaky' patterns suggest that searches are driven by particular demand surges, some of which can be associated with major political events. For example, searches for numerical keywords peak with the US elections in 2008 and 2016: there is an international alignment of scope and scale of interest.

The three graphs below highlight keyword searches for 'statistics', 'billion' and 'percentage'. New Zealand's punctuations and overall attention to the 'statistics' keyword is relatively strong in the first half of the time-period under investigation. They then dip below worldwide (i.e., not just Anglosphere) average (as indicated by the white dotted line). In the UK, it is the second half where punctuations and attention turn above worldwide average. The US and Canada are aligned with the average. The graphs also illustrate some national political key events in order to assess whether certain events are driving search activity to particular terms in particular countries. It is possible to suggest that there is a slight growth across countries in the search term 'percentage' (as illustrated in the white dotted line). Searches for 'statistics' are broadly stable, trends in the word 'billion' are somewhat bifurcated with more attention in some countries rather than others. It is also possible to suggest that some political events seem to be associated with searches of particular numerical statements, such as the UK Brexit referendum or a US presidential election. Most generally, US events seem to generate aligned peaks in searches in the sense of attention across Anglosphere countries correlating at the same time, but some peaks in searches cannot easily be identified as being driven by large political events (such as elections or changes in government).

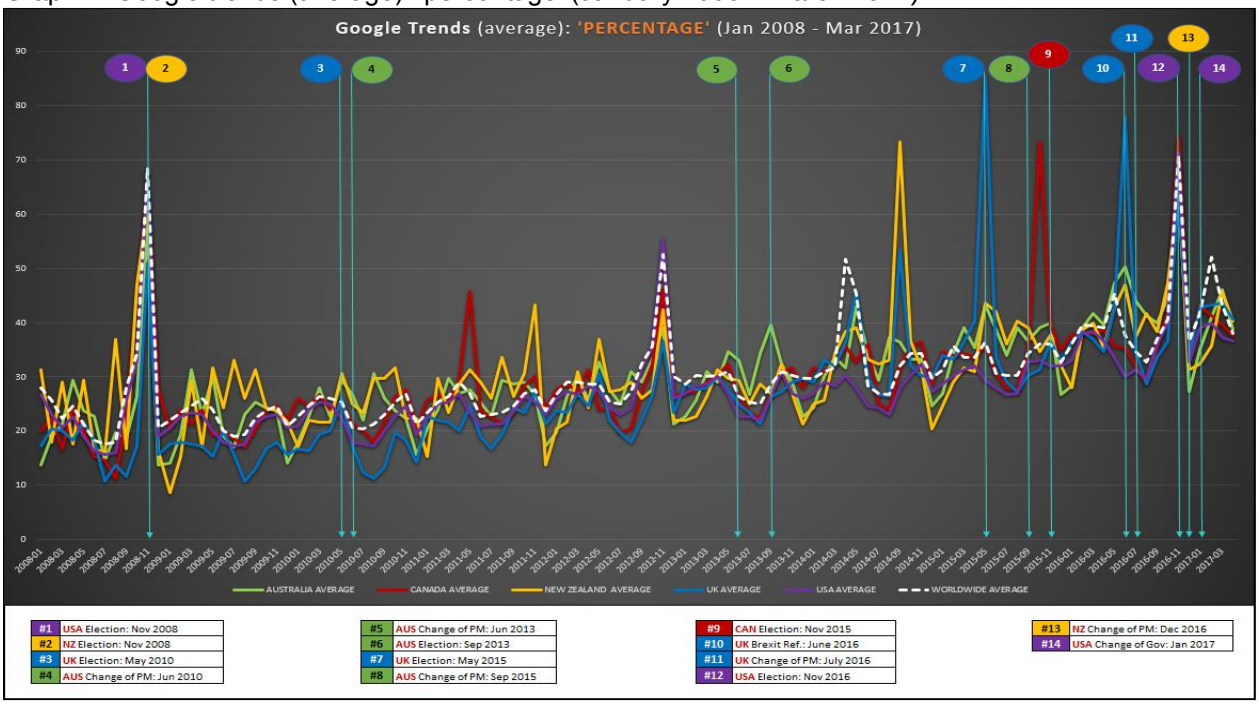
Graph 2. Google trends (average): ‘statistics’ (January 2008 – March 2017).



Graph 3. Google trends (average): ‘billion’ (January 2008 – March 2017).

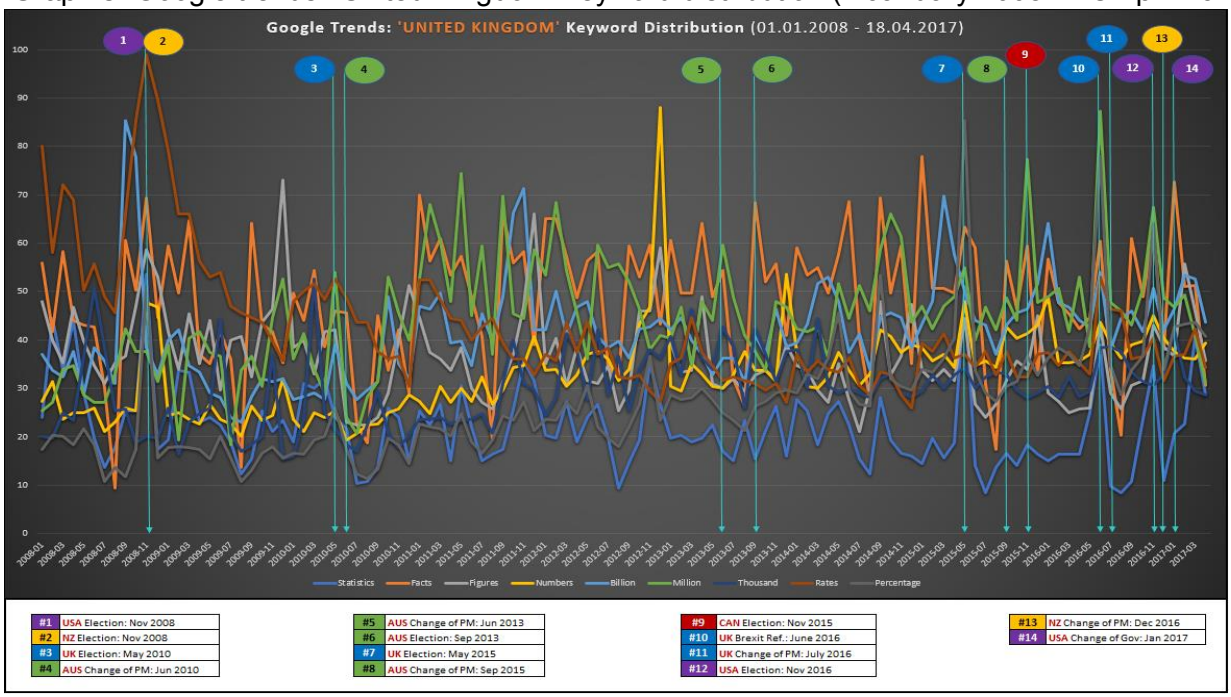


Graph 4. Google trends (average): 'percentage' (January 2008 – March 2017).



Turning to the analysis of national patterns, again no clear trajectories emerge. The graph for all keywords searched for in the UK highlights no particular pattern (see Graph 5 below). At best, there are some words, such as 'statistics' that are not as frequently searched for as 'million' or 'facts' or 'billion'.

Graph 5. Google trends: 'United Kingdom' keyword distribution (1 January 2008 – 18 April 2017).

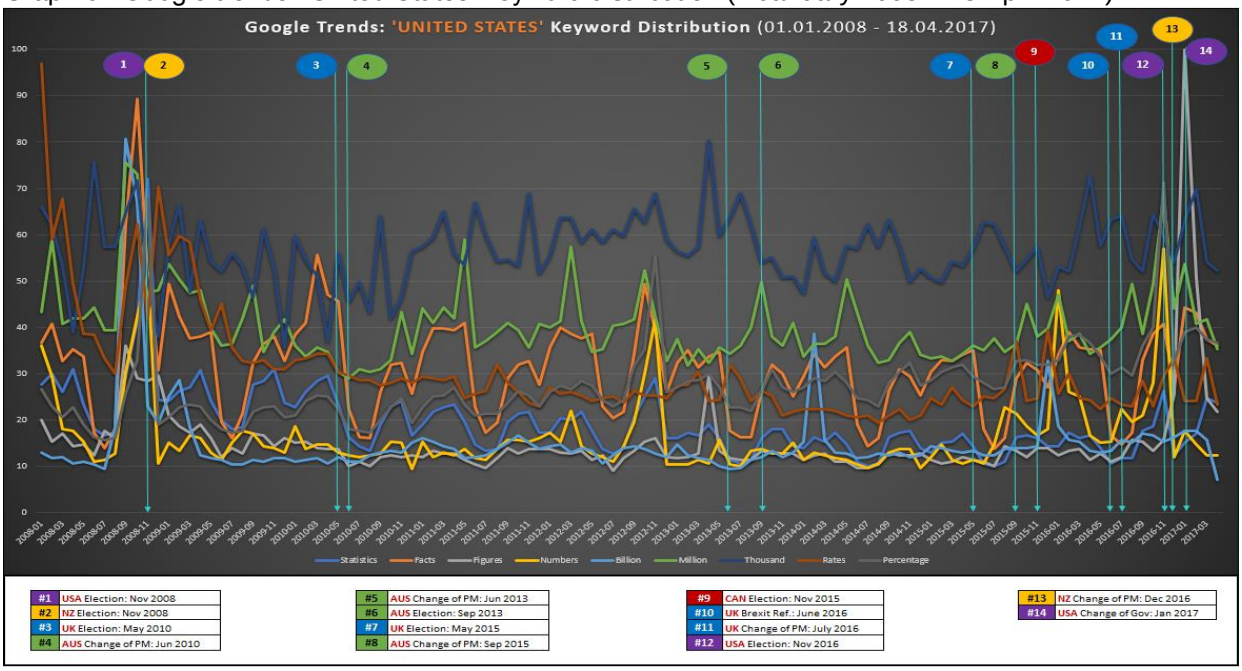


Similar peaky trends with somewhat different degrees of popularity of certain terms are observable in other English-speaking jurisdictions. In the case of the US, the influence of presidential elections in noticeable (and the dominance of terms are observable in other Englishany other jurisdiction); in Australia, like other jurisdictions, it is evident that some keywords are more frequently used than others. In neither jurisdiction is there clear evidence of a trend.

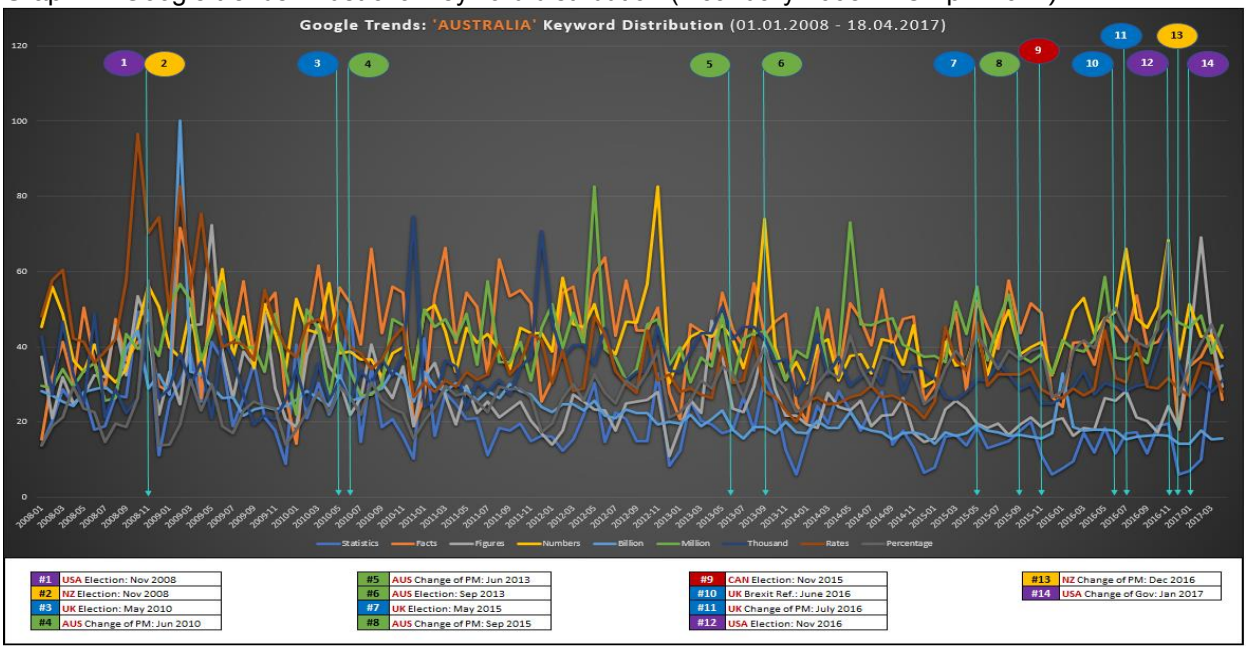
When comparing the different national examples, there is a much wider gap in the US context with regards to searches for different keywords (see Graphs 6 and 7). In contrast, in the Australian case, keyword searches are much more clustered. This suggests that the US audience searches for these words in much greater isolation than the Australian one. Put differently, the keywords are more loosely connected in the US context.

It is evident that neither across countries nor within particular countries, is there a rise in searches for numerical statements. 'Peak' searches for particular terms occur throughout the period under investigation, some appear in the late 2000s, others in the early 2010s, and others towards the end of the period under investigation. In sum, demands for numerical statements are driven by particular attention-driving external events; some keywords (such as percentage, billion, million) seem to be more responsive to changing demand than others (such as 'facts' or 'statistics').

Graph 6. Google trends: 'United States' keyword distribution (1 January 2008 – 18 April 2017).



Graph 7. Google trends: 'Australia' keyword distribution (1 January 2008 – 18 April 2017).



Is there a growing supply of numerical statements?

GOVERNMENT COMMUNICATION

Data-Chamber #02: GOV.uk Info-Flows

Observed Time-Line:	**10 May 2010 – 18 April 2017**
Source:	**GOV.uk platform**
Data Unit:	**Number of references to our Keywords** in (1) Ministerial Speeches; and (2) Department's Press Releases relating to six Cabinet-level entities: PM – DPM – HO – DfE – DH - DWP
Domain/s:	**United Kingdom** (Government)
Keywords tested:	**Statistics** **Facts** **Figures** **Numbers** **Billion** **Million** **Thousand** **Rates** **Percentage**

Turning to the supply side of numerical statements, this study first turns to government communications, as represented on gov.uk. The timeframe of this resource is from 10 May 2010 (the beginning of the coalition government) to 18 April 2017 (the day the 2017 general election was called). The analysis focuses on ministerial speeches and press releases. We focused on both 'all departments' and on select departments that are associated with numerical statements. Table 1 below offers an indication of total numbers. It is noticeable that there was a decline in announcements in 2016 overall (maybe due to government changes). What is more noticeable is the changing patterns among ministerial speeches and press releases. Press releases decline after 2012, ministerial speeches increase somewhat.

Table 1. GOV.uk announcements published by all departments and across all policy areas (10 May 2010 – 18 April 2017). Note: orange fields indicate the largest info-flow distribution for each category.

	All Announcements	Ministerial speeches	Press releases
2010	5,651	530 (9.3%)	3,534 (62.5%)
2011	9,352	726 (7.8%)	5,885 (62.9%)
2012	9,363	690 (7.3%)	5,865 (62.6%)
2013	8,418	933 (11.0%)	3,327 (39.5%)
2014	9,509	947 (9.9%)	4,485 (47.1%)
2015	9,201	818 (8.9%)	4,326 (47.0%)
2016	7,755	751 (9.7%)	3,572 (46.0%)
2017	2,199	216 (9.8%)	987 (44.8%)

Table 2. GOV.uk announcements published by six departments and across all policy areas – combined total for all years (10 May 2010 – 18 April 2017)

Department	Speeches	Press Releases	Department	Speeches	Press Releases
Prime Minister (PM)	586	2,251	Education (DfE)	340	846
Deputy PM (DPM)	107	294	Health (DH)	77	884
Home Office (HO)	206	355	Work & Pensions (DWP)	107	1,007

Across the six departments of special interest (Table 2), the annual ranking in terms of producing speeches and press releases suggests that the prime minister publishes most statements with diverse patterns emerging across other departments over the different years and between speeches and press releases.

Tables 3 and 4 provide an overview of the speeches and press releases across the different departments of particular interest. The departments vary in terms of the frequency of making statements.

Table 3. Ranking order of speeches on GOV.uk: six departments by absolute frequency

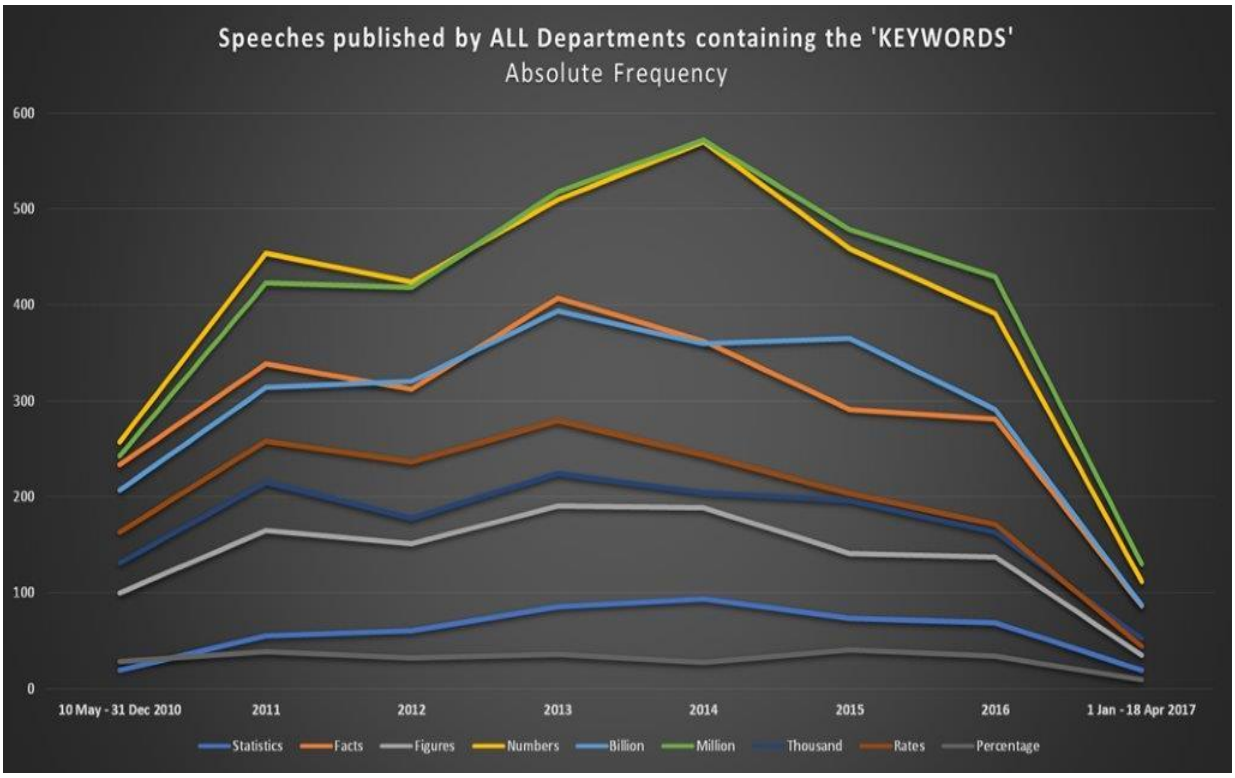
10 May – 31 Dec 2010	2011	2012	2013	2014	2015	2016	1 Jan – 18 Apr 2017
PM	PM	PM	PM	PM	PM	PM	PM
DPM	DfE	DfE	DfE	DfE	DfE	DfE	DfE
DfE	HO	DWP	HO	HO	HO	HO	HO
DWP	DWP	HO	DWP	DPM	DH	DWP	DWP
HO	DPM	DPM	DPM	DH	DWP	DH	DH
DH	DH	DH	DH	DWP	DPM		

Table 4. Ranking order of press releases on GOV.uk: six departments by absolute frequency

10 May – 31 Dec 2010	2011	2012	2013	2014	2015	2016	1 Jan – 18 Apr 2017
PM	PM	DH	PM	PM	PM	PM	PM
DWP	DH	PM	DH	DfE	DfE	DfE	DfE
HO	DWP	DWP	DWP	DWP	DWP	DWP	DWP
DfE	HO	DfE	DfE	DPM	DPM	HO	HO
DPM	DfE	HO	DPM	DH	HO	DH	DH
DH	DPM	DPM	HO	HO	DH		

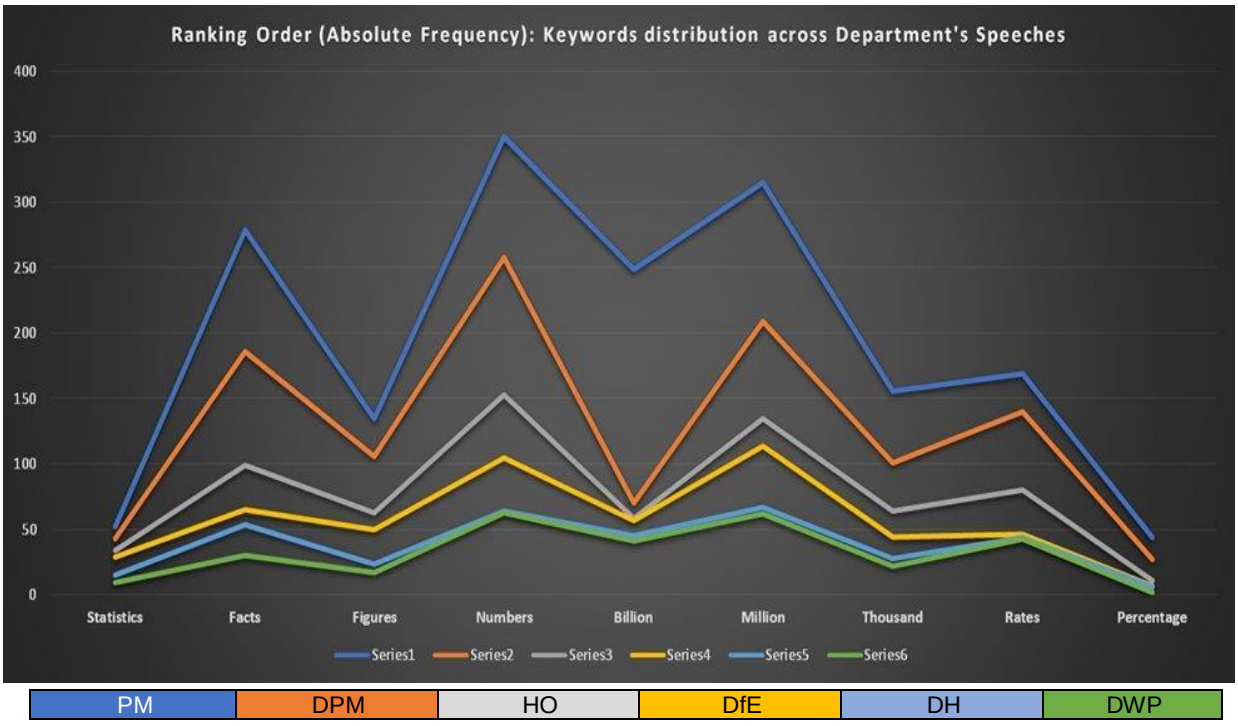
In terms of use of keywords, there are some differences across speeches and press releases. Turning to ministerial speeches first, Graph 8 below highlights patterns and frequencies across different keywords across all government departments. It reveals a growth in numerical statements until 2014 followed by a decline. It also reveals a broad ranking of keywords in terms of their relative frequency. 'Numbers' and 'million' dominate.

Graph 8. Speeches published by all departments containing the 'keywords'.



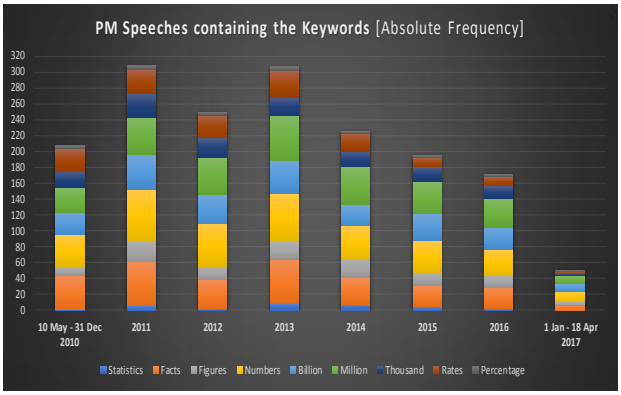
Graph 9 below focuses on the use of keywords by the six select departments. As noted already, the PM (blue line) dominates statements but on some terms more than others. The stability in terms of ranking of keywords is remarkable. The key variation is the extent of dispersion between the different departments in terms of the frequency of use of particular keywords.

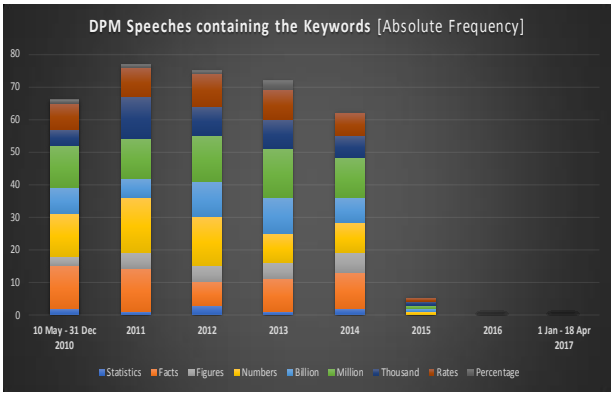
Graph 9. Ranking order (absolute frequency): keywords distribution across department speeches.

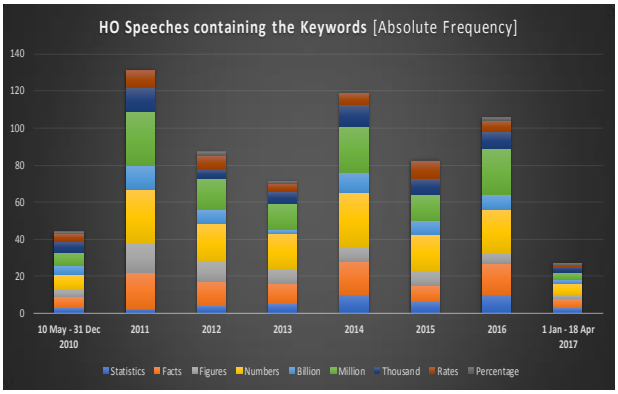


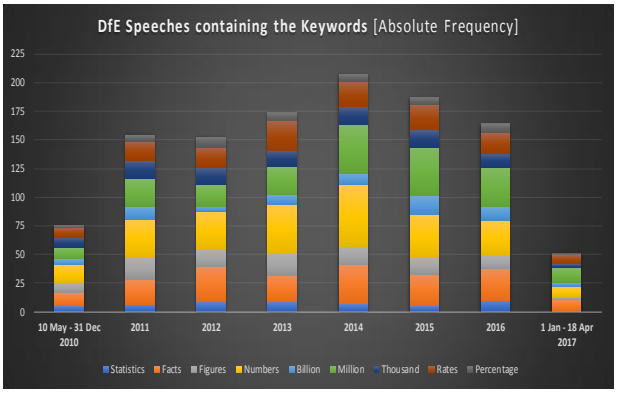
In terms of different departments, the compilation in Graph 10 highlights frequency as well as broad distribution of keywords. Even when taking into account different scales, different shapes are noticeable. It does not reveal a uniform trend in frequency in terms of numerical statements. The different absolute frequencies between departments is noticeable (as in speeches made by the Department of Health) as is the decline in speeches by DWP ministers.

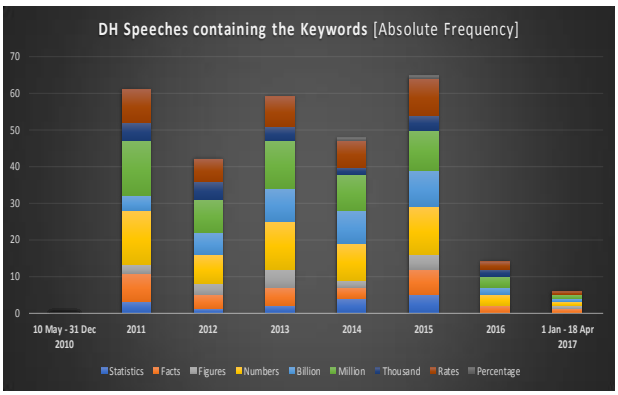
Graph 10 (compilation). Speeches form six departments containing the keywords (absolute frequency).

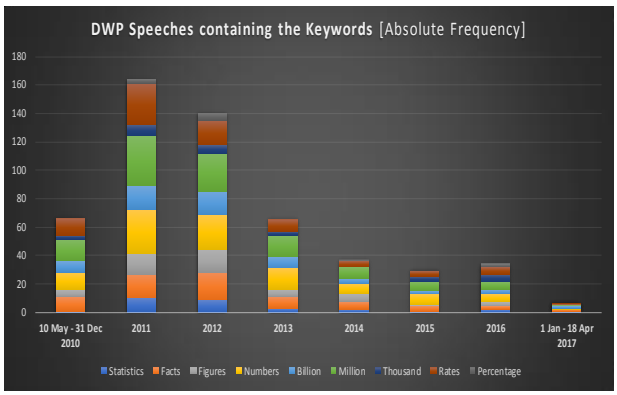












In terms of the relative use of particular keywords, there is broad similarity across departments.

Table 5. Ranking of keywords in ministerial speeches by a department (2010–2017) – absolute frequency.

PM	DPM	HO	DfE	DH	DWP
Numbers	Million	Numbers	Numbers	Numbers	Million
Million	Numbers	Million	Million	Million	Numbers
Facts	Facts	Facts	Facts	Rates	Rates
Billion	Billion	Thousand	Rates	Billion	Facts
Rates	Thousand	Figures	Figures	Facts	Billion
Thousand	Rates	Billion	Thousand	Thousand	Figures
Figures	Figures	Rates	Billion	Figures	Statistics
Statistics	Statistics	Statistics	Statistics	Statistics	Thousand
Percentage	Percentage	Percentage	Percentage	Percentage	Percentage

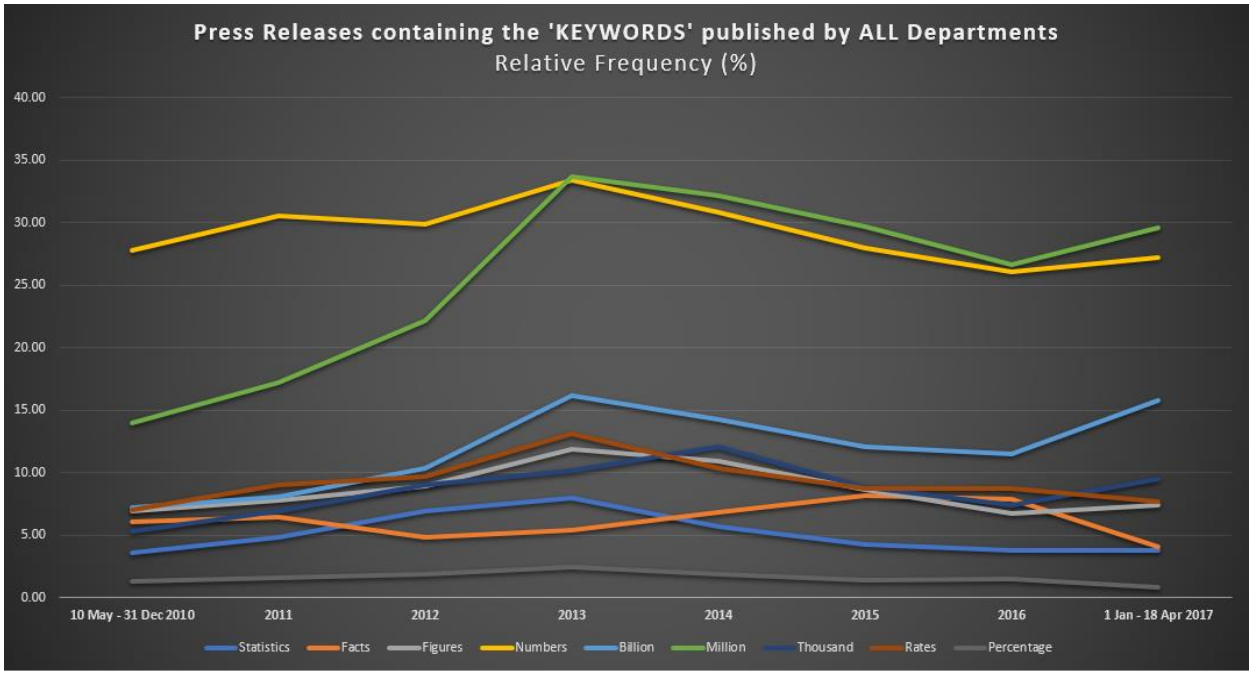
Ranking of keywords in ministerial speeches across the six departments highlights a broad stability of keywords from the different years. This is broadly matched by stability of keyword use by the six departments on their own.

Table 6. Ranking of keywords in ministerial speeches by year (6 departments) – absolute frequency

10 May – 31 Dec 2010	2011	2012	2013	2014	2015	2016	1 Jan – 18 Apr 2017
Numbers	Numbers	Numbers	Million	Million	Million	Million	Million
Million	Million	Million	Numbers	Numbers	Numbers	Numbers	Numbers
Facts	Facts	Billion	Facts	Facts	Billion	Billion	Billion
Billion	Billion	Facts	Billion	Billion	Facts	Facts	Facts
Rates	Rates	Rates	Rates	Rates	Rates	Rates	Thousand
Thousand	Thousand	Thousand	Thousand	Thousand	Thousand	Thousand	Rates
Figures	Figures	Figures	Figures	Figures	Figures	Figures	Figures
Percentage	Statistics	Statistics	Statistics	Statistics	Statistics	Statistics	Statistics
Statistics	Percentage	Percentage	Percentage	Percentage	Percentage	Percentage	Percentage

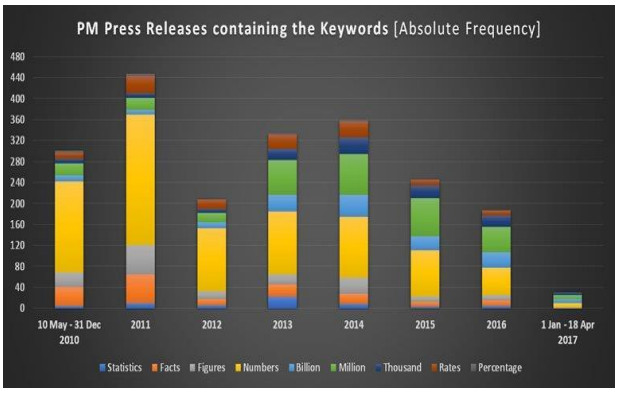
In terms of press releases and the use of keywords (Graph 11), there is no clear trend towards a growing use of keywords. Two words are frequently used – numbers and millions. Billion becomes more noticeable after 2012.

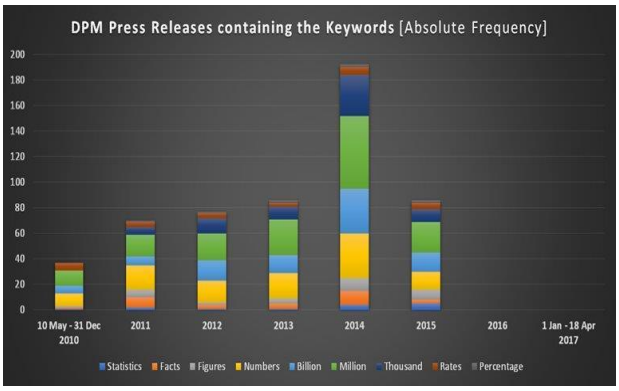
Graph 11. Press releases containing the ‘keywords’ published by all departments.



One can also consider the use by different departments in their press releases. Different patterns emerge here . Patterns do not reveal a unifrequency of press releases (note that the scales differ on the y-axis). Excluding the DPM from the discussion, it is puzzling why some departments are witnessing such a decline in frequency in press releases. An analysis of keyword use by different departments over time also does not reveal considerable differences in terms of word use.

Graph 12 (compilation). Press releases from the six departments containing the keywords.





There is broad stability (Table 7) in terms of keyword use across different departments. However, some more variety exists with other terms – such as 'facts'.

Table 7. Ranking of keywords in press releases by a department (2010–2017) – absolute frequency.

PM	DPM	HO	DfE	DH	DWP
Numbers	Million	Numbers	Numbers	Numbers	Million
Million	Numbers	Million	Million	Million	Numbers
Figures	Billion	Figures	Thousand	Statistics	Figures
Billion	Thousand	Billion	Figures	Rates	Rates
Facts	Figures	Thousand	Rates	Figures	Statistics
Rates	Facts	Statistics	Billion	Billion	Thousand
Thousand	Rates	Rates	Statistics	Thousand	Billion
Statistics	Statistics	Facts	Percentage	Percentage	Percentage
Percentage	Percentage	Percentage	Facts	Facts	Facts

Similar to ministerial speeches, there is a broad stability in the most popular keywords across years across departments. However, there is somewhat more variability in the less frequently used terms (Table 8).

Table 8. Ranking of keywords in press releases by year (6 departments) – absolute frequency.

10 May – 31 Dec 2010	2011	2012	2013	2014	2015	2016	1 Jan – 18 Apr 2017
Numbers	Numbers	Numbers	Million	Million	Million	Million	Million
Million	Million	Million	Numbers	Numbers	Numbers	Numbers	Numbers
Billion	Rates	Billion	Billion	Billion	Billion	Billion	Billion
Rates	Billion	Rates	Rates	Thousand	Thousand	Rates	Thousand
Figures	Figures	Thousand	Figures	Figures	Rates	Facts	Rates
Facts	Thousand	Figures	Thousand	Rates	Figures	Thousand	Figures
Thousand	Facts	Statistics	Statistics	Facts	Facts	Figures	Facts
Statistics	Statistics	Facts	Facts	Statistics	Statistics	Statistics	Statistics
Percentage	Percentage	Percentage	Percentage	Percentage	Percentage	Percentage	Percentage

SPEECHES BY POLITICAL LEADERS

Data-Chamber #03: **British Political Speech archive**

Observed Time-Line:	**1967 – 2016**
Source:	**British Political Speech archive**
Data Unit:	**Number of times our Keywords were mentioned in speeches by the political party leaders** (Conservatives – Labour – Liberal Democrats)
Domain/s:	**United Kingdom** (Political Parties)
Keywords tested:	**Statistics** **Facts** **Figures** **Numbers** **Billion** **Million** **Thousand** **Rates** **Percentage** **Numerical ref. (,0)** **Office of National Statistics / ONS** **UK Statistics Authority / UKSA**

A further area of interest concerns political speeches made outside the government and parliamentary arena. These speeches are arguably less regulated and they can also be seen as a barometer for the kind of tone that campaigns seek to establish. In the following, we therefore focus on major political speeches given by party leaders (Conservative, Labour and Liberal Democrat) as archived in the British Political Speech project. We focus on the period 1967–2016 and separate out leader speechers delivered at party conferences and speeches delivered at other occasions that have been included in the database.

The total universe of speeches is 264 (for the 1967–2016 period), including 77 Conservative, 111 Labour and 76 Liberal Democrat speeches, involving 1,144,592 words (average 4,336 words per speech).[3] A total of 3,069 keyword references are made in these speeches and the average keyword count per speech is 12.[4] There are no significant differences across parties in terms of their average speech length – Conservative 4,234 (1,108 unique) words, Labour 4,659 words (1,171 unique) words, and Liberal Democrat 3,966 (1,148 unique) words – or their use of average keywords (Conservative 12, Labour 10 and Liberal Democrat 13) (see Graph 13).

Focusing on leaders' speeches alone, no major differences emerge. The 143 speeches in the dataset amount to 779,894 words including 2,176 keywords. The average speech length varies by party (Conservative 5,181, Labour 6,687, Liberal Democrat 4,793), but the average use of keywords is broadly the same (Conservative 16, Labour 15, Liberal Democrat 15).

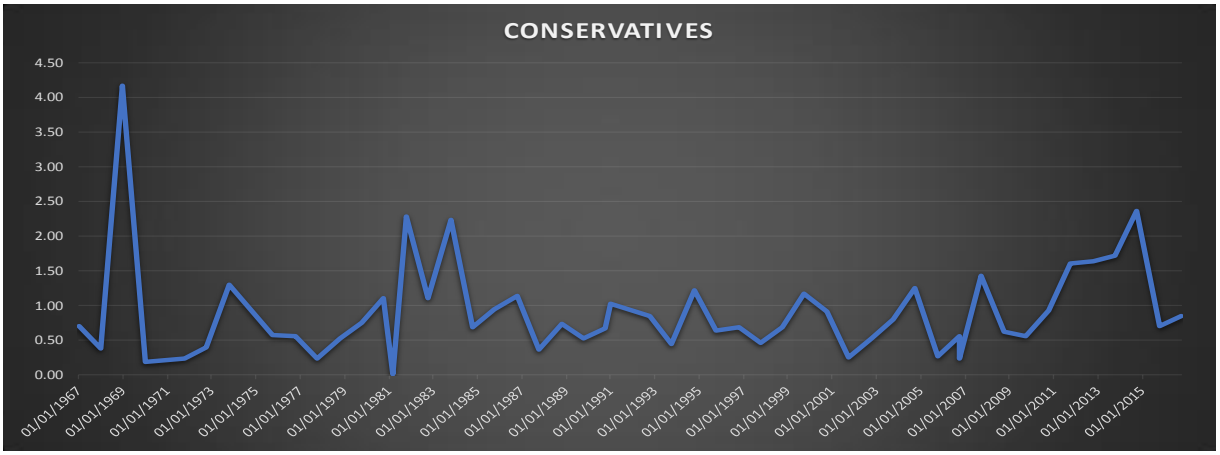
A number of issues arise from the analysis. One is whether the total number of words has changed over time (i.e. have speeches become longer/shorter over time). The other is that there might be

[3] The speeches contain 302,547 unique words with an average of 1,143 unique words.
[4] This involves the twelve keywords outlined earlier and the figure represents the combined total.

changing numerical statements over time, but the use of the different keywords should to some extent address the challenge of dealing with changing linguistic fashions, in particular we present aggregated keyword numbers so as to have an indicator of changes in the 'intensity' of numerical statements over time, and across parties.

Have the speeches gone longer and are they using more numerical references? As one can be seen in the Table 9 below, Labour speeches dominate across the ages, but there is no clear trend towards longer (or shorter) speeches over time. The same holds when focusing on 'total unique words'. In terms of combined keywords, Labour again features among the most frequent users with, again, no clear pattern emerging in terms of historical time-line.

Graph 13 (compilation). Leader's speech 1967–2016: total unique words relative frequency.



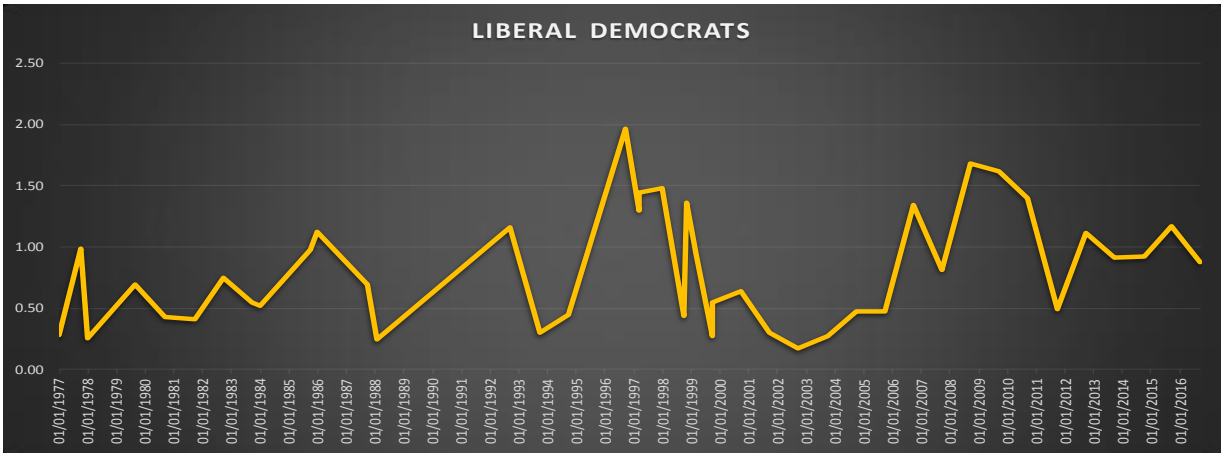


Table 9. (compilation). Top 15 speeches by total word count, unique word count and total number of combined keywords.

Year	Total Word Count		Year
	All Speeches	Leader's Speech	
2007	Conservative	Conservative	2007
1971	Labour	Labour	1971
1967	Labour	Labour	1967
1975	Labour	Labour	1975
1985	Labour	Labour	1985
1989	Labour	Labour	1989
1978	Labour	Labour	1978
2013	Labour	Labour	2013
1970	Labour	Labour	1970
1972	Labour	Labour	1972
1973	Conservative	Conservative	1973
1979	Labour	Labour	1979
1976	Labour	Labour	1976
2007	Labour	Labour	2007
2012	Labour	Labour	2012

Year	Unique Word Count		Year
	All Speeches	Leader's Speech	
1971	Labour	Labour	1971
1975	Labour	Labour	1975
1984	Liberal Democrat	Liberal Democrat	1984
1970	Labour	Labour	1970
1972	Labour	Labour	1972
1983	Liberal Democrat	Liberal Democrat	1983
1967	Labour	Labour	1967
1976	Labour	Labour	1976
1986	Liberal Democrat	Liberal Democrat	1986
1985	Labour	Labour	1985
1978	Labour	Labour	1978
2008	Conservative	Liberal Democrat	1985
1985	Liberal Democrat	Labour	2007
2007	Labour	Liberal Democrat	1977
1977	Liberal Democrat	Liberal Democrat	1982

Top 15 speeches by the total number of combined keywords

Year	All Speeches	Leader's Speeches	Year
1969	Conservative	Conservative	1969
1967	Labour	Labour	1967
2000	Labour	Labour	2000
1990	Labour	Labour	1990
2003	Labour	Labour	2008
2008	Labour	Labour	2015
2015	Labour	Labour	2013
2013	Labour	Labour	1968
1968	Labour	Labour	1971
1971	Labour	Labour	2016
2016	Labour	Conservative	2014
2014	Conservative	Liberal Democrat	1996
1996	Liberal Democrat	Conservative	1981
2002	Labour	Labour	2007
1999	Labour	Labour	1998
1969	Conservative	Conservative	1969

Some more diversity emerges when focusing on particular keywords, but these changes do not suggest a rise in keywords over time. There are some interesting partisan variations with Conservative speeches dominating references to 'percentage', the Liberal Democrats feature more more prominently on 'thousand', whereas Labour dominates statements with numerical values, although other parties feature too.

Table 10 (compilation). Top 15 speeches by the number of individual keywords.

Year	STATISTICS		Year
	All Speeches	Leader's Speech	
2009	Labour	Conservative	1969
2005	Labour	Liberal Democrat	2016
1969	Conservative	Labour	2003
2016	Liberal Democrat	Conservative	2000
2009	Labour	Labour	2009
2007	Conservative	Liberal Democrat	2009
2003	Labour	Labour	2008
2000	Conservative	Conservative	2007
1990	Conservative	Liberal Democrat	1996
2009	Labour	Liberal Democrat	1993
2009	Labour	Labour	1990
2009	Liberal Democrat	Conservative	1986
2008	Labour	Liberal Democrat	1986
2008	Labour	Labour	1984
2007	Conservative	Conservative	1982

Year	FACTS		Year
	All Speeches	Leader's Speech	
1967	Labour	Labour	1967
1993	Labour	Labour	1993
2004	Labour	Labour	1984
1984	Labour	Labour	2013
2013	Labour	Liberal Democrat	2007
2010	Conservative	Conservative	2004
2007	Liberal Democrat	Conservative	1994
2006	Labour	Labour	1986
2004	Conservative	Labour	1983
1994	Conservative	Conservative	1978
1986	Labour	Labour	1969
1983	Labour	Liberal Democrat	2016
1978	Conservative	Conservative	2014
1969	Labour	Conservative	2013
2016	Liberal Democrat	Conservative	2012

Year	FIGURES		Year
	All Speeches	Leader's Speech	
1978	Labour	Labour	1978
1967	Labour	Labour	1967
1973	Conservative	Conservative	1973
1973	Labour	Labour	1973
1968	Labour	Labour	1968
1989	Labour	Labour	1989
1984	Liberal Democrat	Liberal Democrat	1984
1969	Labour	Labour	1969
1999	Labour	Labour	1983
1983	Labour	Labour	1971
1971	Labour	Conservative	2011
2011	Conservative	Conservative	2004
2009	Labour	Labour	1990
2007	Labour	Labour	1982
2004	Conservative	Labour	1981

Year	NUMBERS		Year
	All Speeches	Leader's Speech	
2004	Labour	Labour	2004
2002	Labour	Labour	2002
1981	Labour	Labour	1981
1976	Conservative	Conservative	1976
1971	Labour	Labour	1971
2014	Conservative	Conservative	2014
2013	Conservative	Conservative	2013
2005	Labour	Labour	2003
2003	Labour	Conservative	2011
2011	Conservative	Liberal Democrat	2010
2010	Liberal Democrat	Liberal Democrat	2006
2007	Labour	Labour	2005
2006	Liberal Democrat	Labour	1999
2005	Labour	Labour	1989
1999	Labour	Liberal Democrat	1986

Year	BILLION		Year
	All Speeches	Leader's Speech	
2000	Labour	Labour	2000
1998	Labour	Labour	1998
1998	Labour	Liberal Democrat	1998
1994	Labour	Labour	1994
1996	Labour	Liberal Democrat	1996
1996	Liberal Democrat	Labour	1991
1991	Labour	Labour	1990
1990	Labour	Labour	1989
1989	Labour	Labour	1997
2010	Conservative	Labour	2014
2007	Labour	Liberal Democrat	2012
2003	Labour	Conservative	2011
1997	Labour	Labour	2008
1996	Liberal Democrat	Liberal Democrat	2008
2014	Labour	Labour	2016

Year	MILLION		Year
	All Speeches	Leader's Speech	
1969	Conservative	Conservative	1969
1967	Labour	Labour	1967
2000	Labour	Labour	1968
1968	Labour	Labour	2016
2016	Labour	Labour	2013
2013	Labour	Labour	2008
2008	Labour	Labour	2000
2003	Labour	Conservative	1981
2005	Labour	Labour	1969
2004	Labour	Conservative	2014
2000	Labour	Labour	2012
1981	Conservative	Labour	2009
1969	Labour	Labour	2015
2008	Labour	Conservative	1994
2014	Conservative	Labour	1985

Year	THOUSAND		Year
	All Speeches	Leader's Speech	
1996	Labour	Labour	1996
1992	Liberal Democrat	Liberal Democrat	1992
2015	Liberal Democrat	Liberal Democrat	2015
2010	Labour	Labour	2010
2012	Conservative	Conservative	2012
2012	Conservative	Labour	2011
2011	Labour	Liberal Democrat	1996
1999	Labour	Conservative	1981
1997	Liberal Democrat	Labour	1971
1996	Liberal Democrat	Labour	2015
1981	Conservative	Liberal Democrat	2014
1971	Labour	Conservative	2014
2015	Labour	Labour	2013
2014	Liberal Democrat	Conservative	2010
2014	Conservative	Labour	2009

Year	RATES		Year
	All Speeches	Leader's Speech	
1990	Labour	Labour	1990
1998	Liberal Democrat	Conservative	1969
1969	Conservative	Conservative	1981
1981	Conservative	Labour	1996
2012	Conservative	Labour	1980
2010	Conservative	Conservative	2012
1998	Liberal Democrat	Labour	1998
1997	Liberal Democrat	Conservative	1990
1996	Labour	Labour	1986
1996	Labour	Labour	1984
1980	Labour	Conservative	1983
2012	Conservative	Conservative	1982
2008	Labour	Labour	1972
2003	Labour	Labour	1971
2002	Labour	Conservative	2013

Year	PERCENTAGE		Year
	All Speeches	Leader's Speech	
2011	Labour	Labour	2011
2012	Conservative	Conservative	2011
2011	Conservative	Conservative	2007
2007	Conservative	Labour	2002
2002	Labour	Labour	2001
2001	Labour	Liberal Democrat	1998
1998	Liberal Democrat	Labour	1990
1990	Labour	Conservative	1998
1988	Conservative	Conservative	1980
1980	Conservative	Liberal Democrat	1979
1979	Liberal Democrat	Conservative	1978
1978	Conservative	Conservative	1973
1973	Conservative	Labour	1971
1971	Labour	It only has 13 entries	n/a
n/a	It only has 14 entries	It only has 13 entries	n/a

Year	NUMERICAL REFERENCE		Year
	All Speeches	Leader's Speech	
2002	Labour	Conservative	1969
1969	Conservative	Labour	2000
2000	Labour	Labour	1967
1967	Labour	Labour	2007
2007	Labour	Labour	2015
2015	Labour	Conservative	2007
2007	Conservative	Labour	1972
2006	Labour	Labour	2004
2003	Labour	Labour	1985
1972	Labour	Liberal Democrat	2009
2004	Labour	Labour	1999
1985	Labour	Liberal Democrat	1998
2009	Liberal Democrat	Labour	1998
2009	Labour	Labour	1984
1999	Labour	Liberal Democrat	1997

Focusing solely on leaders' speeches offers a similar pattern. There is no evidence to suggest that speeches have gone longer and/or contain more numerical statements. In particular, the observed peaks mostly do not occur in the last decade.

SPEECHES IN PARLIAMENT

Data-Chamber #04: House of Commons Hansard records

Observed Time-Line:	All MPs: **1 January 1967 – 18 April 2017** Senior MPs: **10 May 2010 – 18 April 2017**
Source:	**House of Commons Hansard records**
Data Unit:	All MPs: **Number of times our Keywords were mentioned during debates** Senior MPs: **Number of references to our Keywords during debates**
Domain/s:	**United Kingdom** (Parliament)
Keywords tested:	**Statistics** **Facts** **Figures** **Numbers** **Billion** **Million** **Thousand** **Rates** **Percentage** **Numerical ref. (,0)** **Office of National Statistics / ONS** **UK Statistics Authority / UKSA**

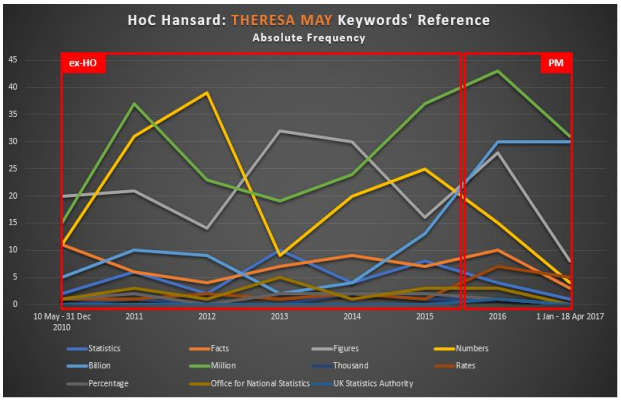
In this analysis, we focus on Hansard and the use of numerical statements in contributions in the House of Commons between 1967 and 2017 (verbal statements). We concentrate on two aspects – one is the total number of numerical statements (keywords) made in Parliament and, second, the reference to numerical statements by senior politicians in written and spoken contributions to the House of Commons between 10 May 2010 and 18 April 2017. In particular, we focused on those ministers associated with six departments discussed above as part of the analysis of gov.uk.

Focusing on senior politicians is complicated as careers bite the dust or take off – this is particularly evident. Again, apart from noting the position of the prime minister in making statements involving numerical statements, there is no evidence that references to numerical statements have increased over the period since 2010.

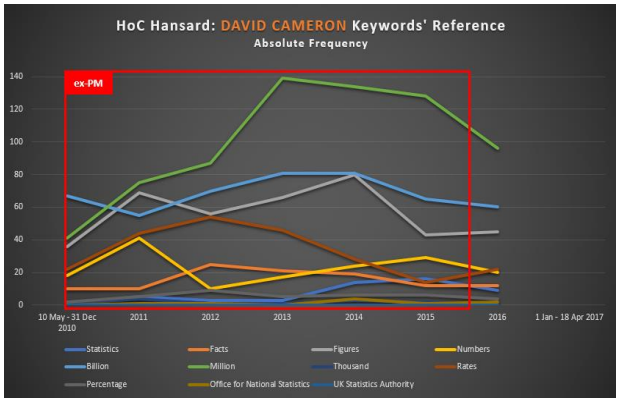
Graph 14 below highlight the use of keywords among senior politicians over the period under investigation (2010–17). We highlight periods where these politicians are in office in the six Departments of special interest. The data reveals that ministers are indeed associated with a greater tendency to make statements containing numerical references as can be seen by those MPs that are out of office in particular times. Ministers appear to make announcements with distinct numerical references – especially involving million and billion and 'figures'. Other words feature less frequently. There is no overall trend towards 'more volume' in terms of numerical statements emerging from these graphs.

Graph 14 (compilation). Senior MPs' references to keywords during parliamentary debates at the House of Commons (10 May 2010 – 18 April 2017).

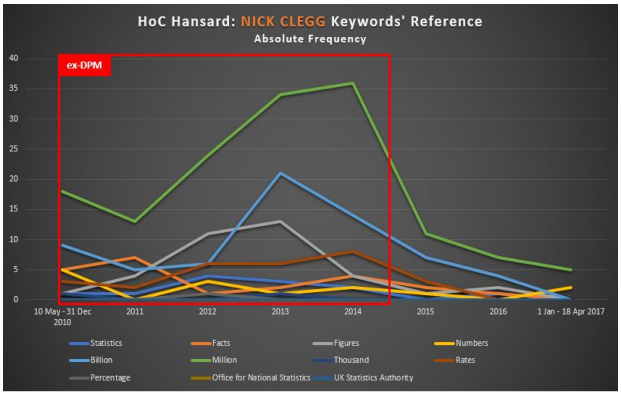
[PM / HO] Theresa May



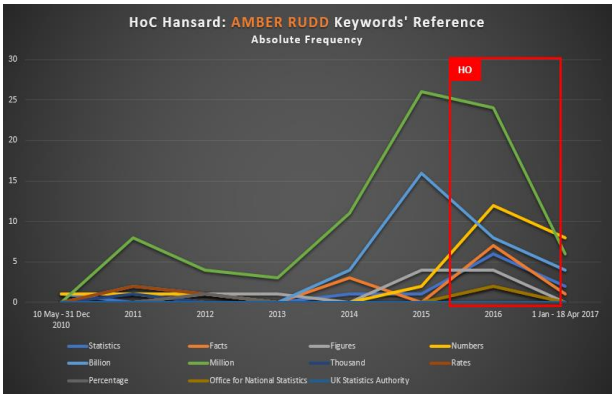
[PM] David Cameron



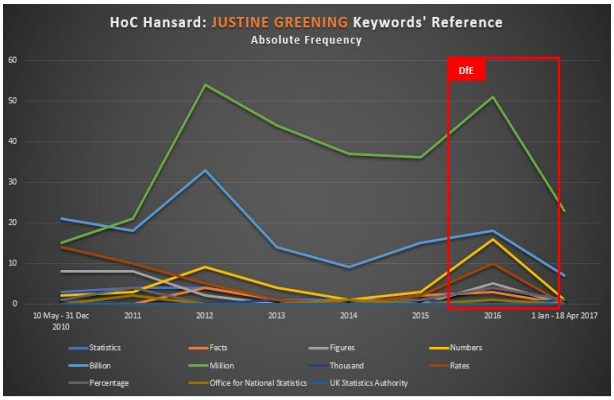
[DPM] Nick Clegg



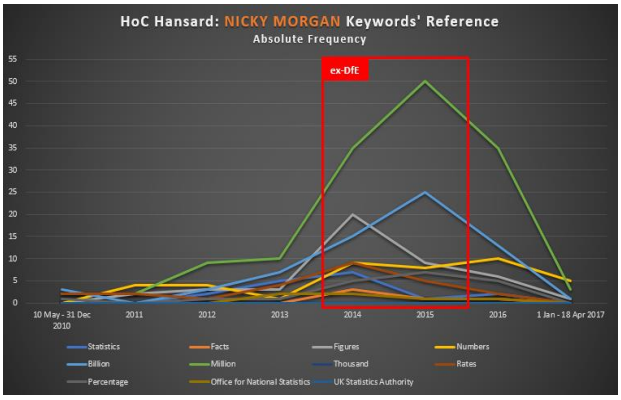
[HO] Amber Rudd



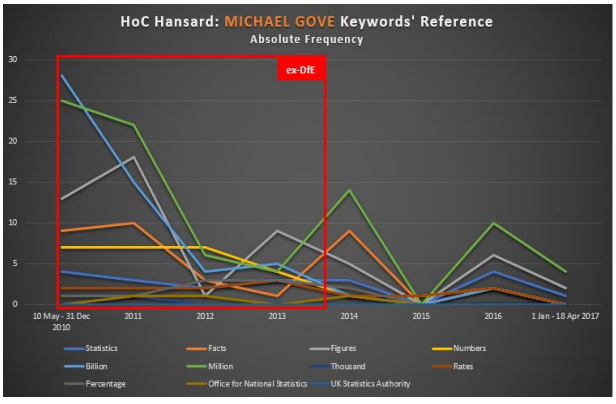
[DfE] Justine Greening



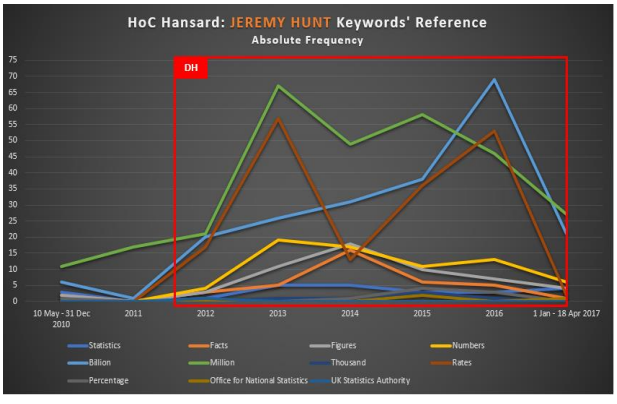
[DfE] Nicky Morgan



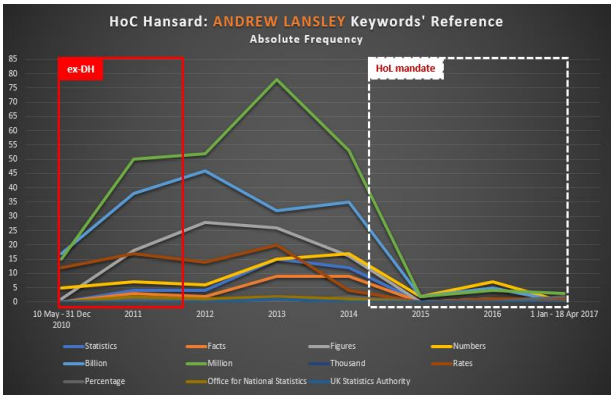
[DfE] Michael Gove



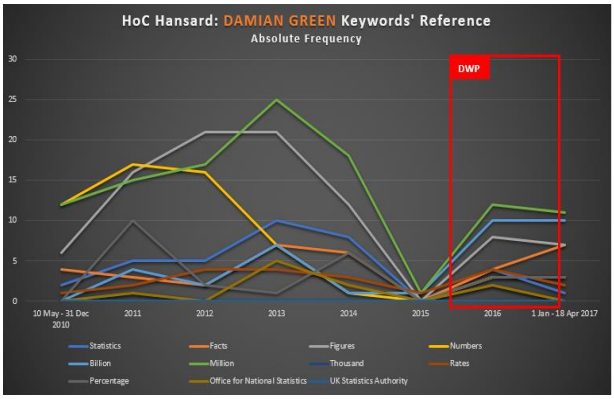
[DH] Jeremy Hunt



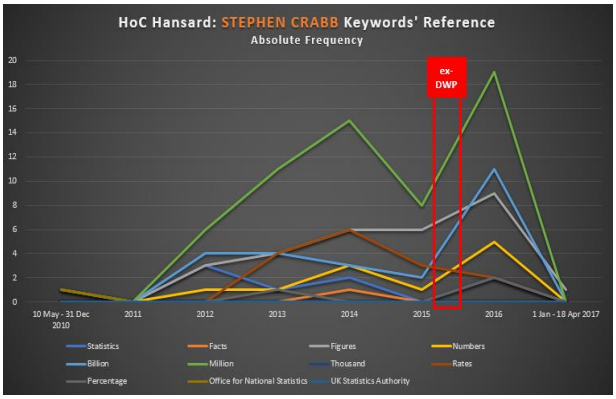
[DH] Andrew Lansley



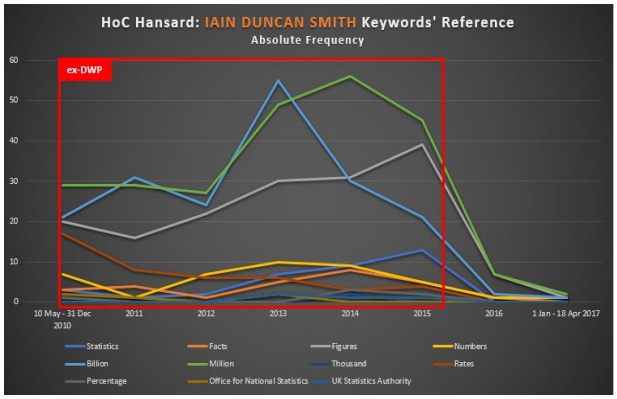
[DWP] Damian Green



[DWP] Stephen Crabb

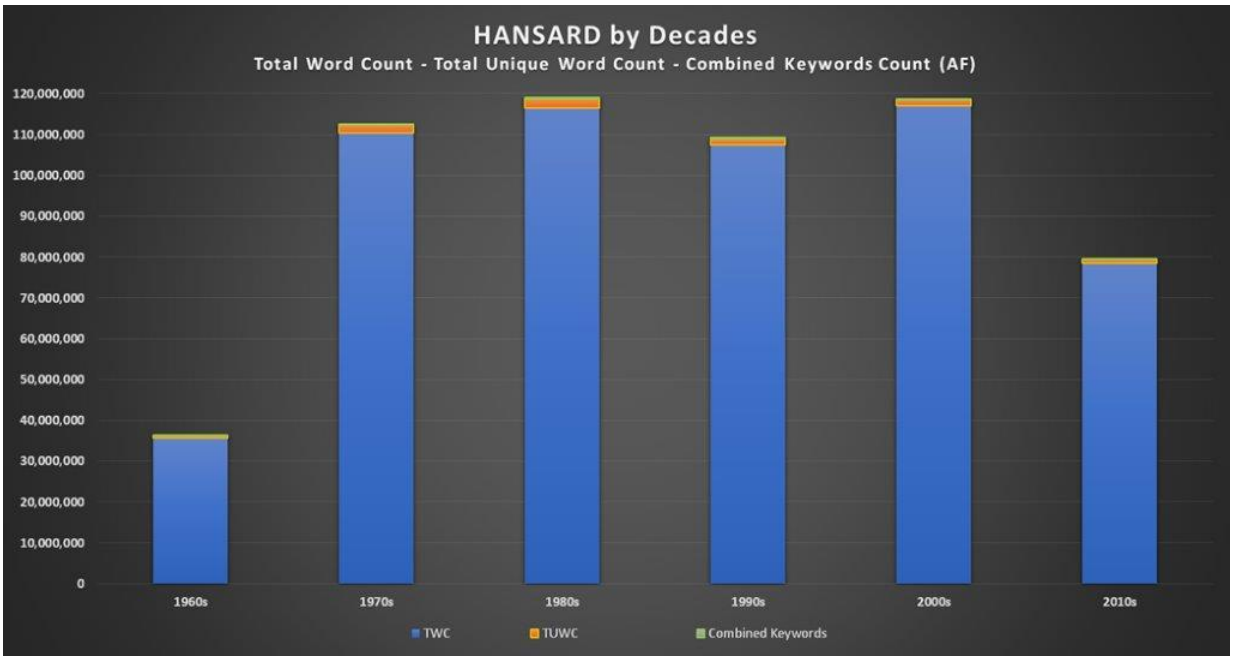


[DWP] Iain Duncan Smith



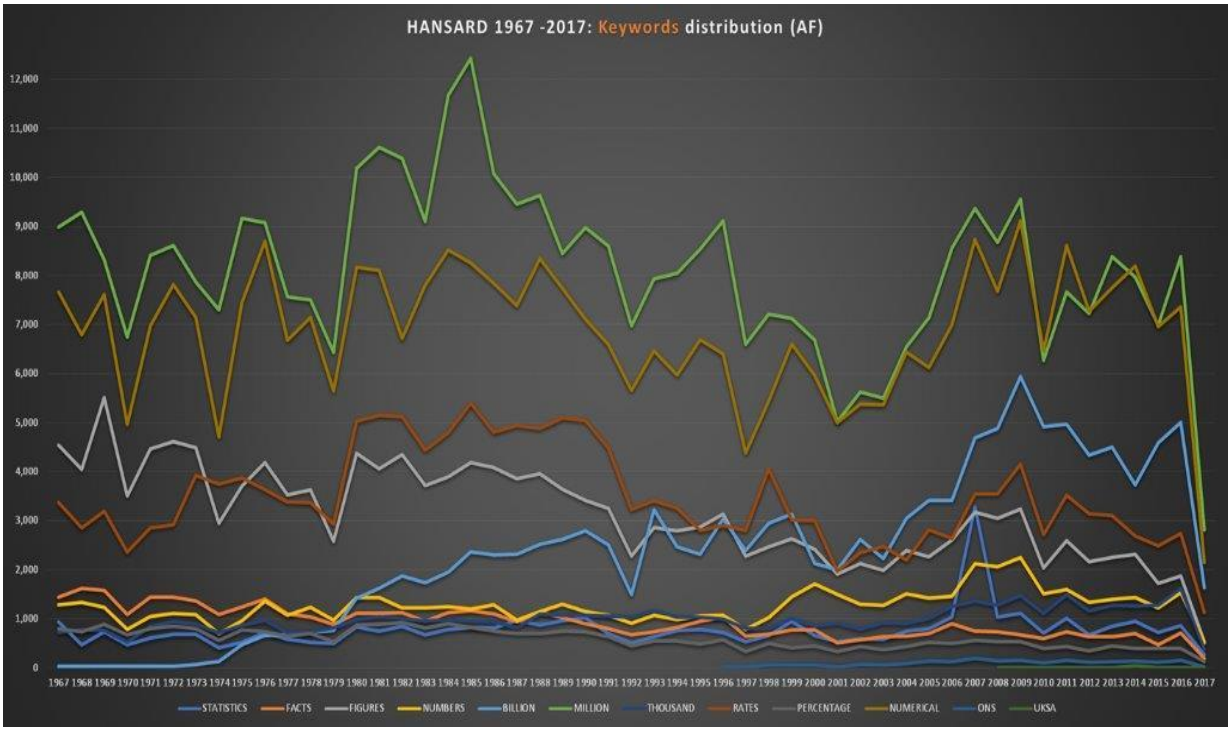
Turning to Hansard records since 1967, Graph 15 below highlights the amount of 'words' recorded in Hansard by decade. The total sample includes 565,674,118 words (an annual average of nearly 11.1 million words per annum). Note that the data for the 1960s and the 2010s do not cover complete decades. However, the four complete decades suggest broad stability and it is not likely that the 2010s will perform differently overall. The total ration of keywords used during these parliamentary exchanges has also not changed much. In terms of relative frequency of the combined keyword count across the total word count, the 1960s score 0.2454, the 1970s 0.2381, the 1980s 0.2922, the 1990s 0.2517, the 2000s, 0.2369 and the 2010s 0.2748. So in terms of 'peak decade' it was the 1980s where most references to numerical statements was made: a total of 0.29 per cent of all words contained a keyword. Generally, one cannot support the claim that there has been an rise in the use of numerical statements in the House of Commons since 1967.

Graph 15. Hansard by decades: total word count, total unique word count, combined keywords count.



Turning to the keywords, it is not evident that there has been a considerable rise of references to particular numerical statements over time. Instead, numbers peak in various years – for example, it is likely that the peak in 'millions' is associated with the rise in unemployment in the early 1980s and the rise of 'billions' in the late 2000s with the financial crisis and bank bail-outs. It is noticeable that the word 'numerical reference' more or less tracks the keyword 'millions'. 'Rates' continue to feature highly, but here attention might reflect interest in either unemployment or information (or both).

Graph 16. Hansard 1967 – 2017: keywords distribution (absolute frequency).



SOCIAL MEDIA

Data-Chamber #05: Twitter

Observed Time-Line:	**10 March 2015 – 18 April 2017**
Source:	**MPsonTwitter.co.uk**
Data Unit:	**Number of times our Keywords were mentioned in MPs' Tweets during the 56th UK Parliament**
Domain/s:	**United Kingdom** (Parliament / Politics)
Keywords tested:	**Statistics, Facts, Figures, Numbers, Billion, Million, Thousand, Rates, Percentage, Numerical ref. (,0), @ONS, @UKStatsAuth, Office of National Statistics / ONS, UK Statistics Authority / UKSA**

Finally, we focus on 'Twitter' as a relatively recent social media phenomenon (launched in March 2006) that has been at the heart of controversy over 'fake news' in the US in particular. It has become a dominant tool for political statements, in contrast to, for example, Facebook, that contains more personal accounts and statements. The message on Twitter is also much more focused than on Facebook. Analysing Twitter usage by MPs is limited by the technology – we focused on the last parliament as Twitter only stores the last 3,000 tweets of a particular account and we could rely on a specialised site that stored tweets by MPs over the course of the previous parliament. Furthermore, not all MPs are on Twitter, Twitter frequency varies and, even more importantly, the number of followers varies. This means that frequency needs to be considered in combination with potential 'reach' of an individual tweet. This allows us to focus not just on the frequency of tweets, but also their potential reach in terms of followers and re-tweets. Twitter is therefore unique among the data-chambers in this study in terms of allowing us to get a sense of the reach of a message.

Our analysis is based on a specific portal devoted to storing MPs' tweets (www.mpsontwitter.co.uk). This facility has preserved tweets by MPs over the past 12 months. At the time of this study, 85.23 per cent of MPs had twitter accounts (a very high proportion in comparison with other European jurisdictions). Among these, 264 Conservative MPs (out of 330), 205 (out of 229) Labour, and 8 (out of 9) Liberal Democrat (as well as all 54 Scottish National and 23 'others' (out of 28)) were on Twitter. The sample of tweets containing a numerical statement is 12,976 with 3,640 making a numerical reference, 3,324 referring to 'million', 158 refer to ONS (Office for National Statistics) and 12 to UKSA (UK Statistics Authority). The 12,976 tweets that contain at least one of the keywords consist of 4,568 Conservative, 6,578 Labour, 255 Liberal

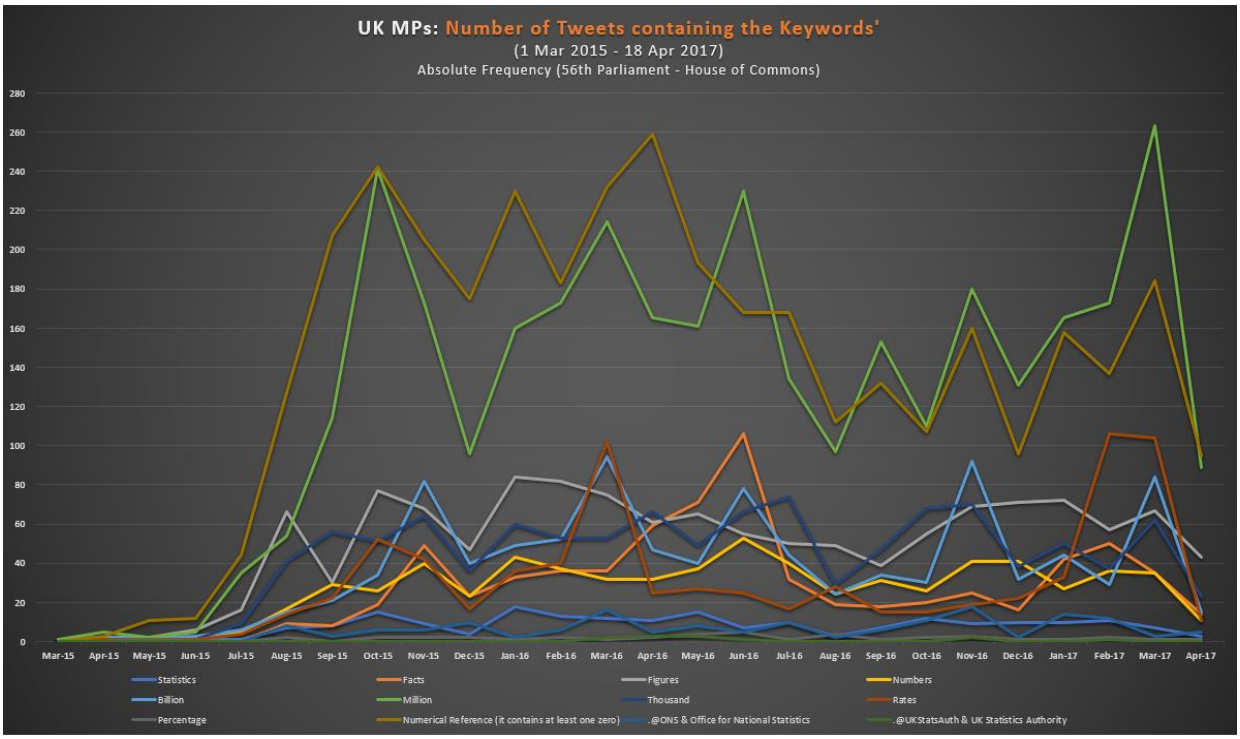
Democrat and 1,107 SNP tweets. No major difference exists in the ranking of keywords used across the different parties as illustrated in Table 11 below.

Table 11. Keywords ranking by political party (frequency)

Conservative	Labour	Liberal Democrats	Scottish National Party
Numerical	Numerical	Numerical	Numerical
Million	Million	Million	Million
Figures	Thousand	Thousand	Billion
Billion	Figures	Billion	Thousand
Rates	Billion	Figures	Figures
Thousand	Numbers	Statistics	Numbers
Facts	Facts	Numbers	Rates
Numbers	Rates	Facts	Facts
ONS	Statistics	Rates	Statistics
Statistics	ONS	UKSA	Percentage
Percentage	Percentage	ONS	ONS
UKSA	UKSA		

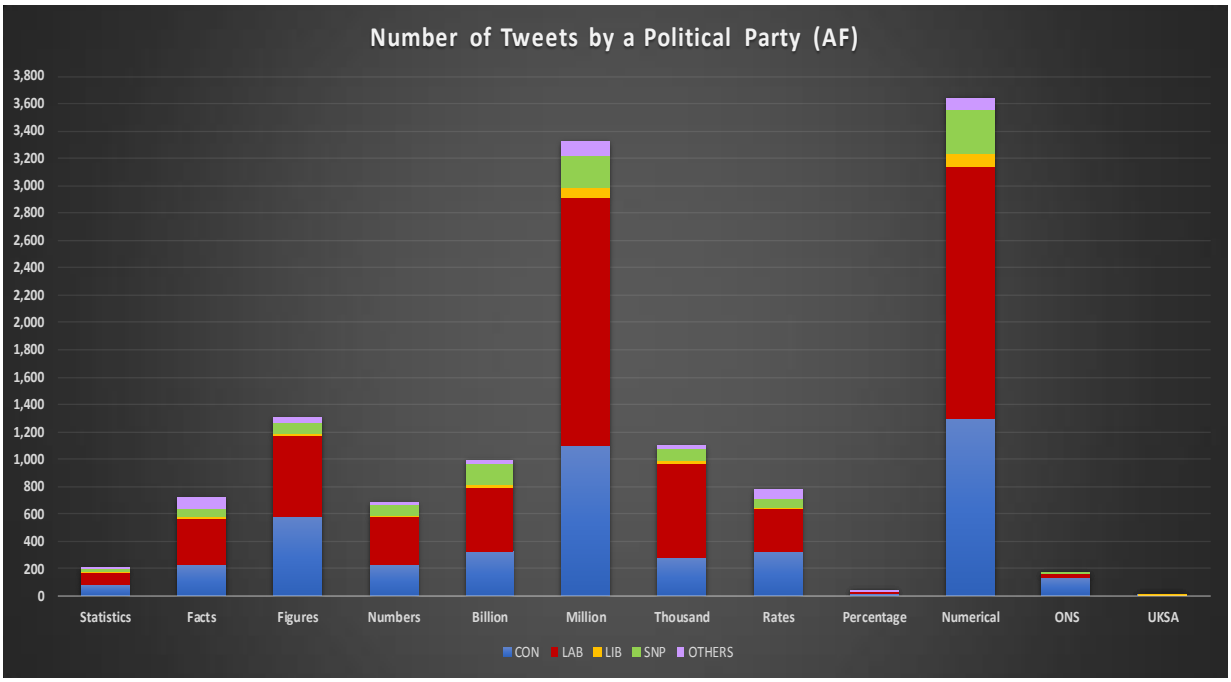
Graph 17 below highlights the number of tweets that contained keywords in terms of their absolute frequency. Numerical references (3,640 in total) and 'millions' (3,324 in total) stand out as the most frequently used keywords. Other keywords appear regularly with some peaks. The ONS attracted a total of 158 tweets, UKSA 12.

Graph 17. UK MPs: number of tweets containing the keywords (1 March 2015 – 18 April 2017).



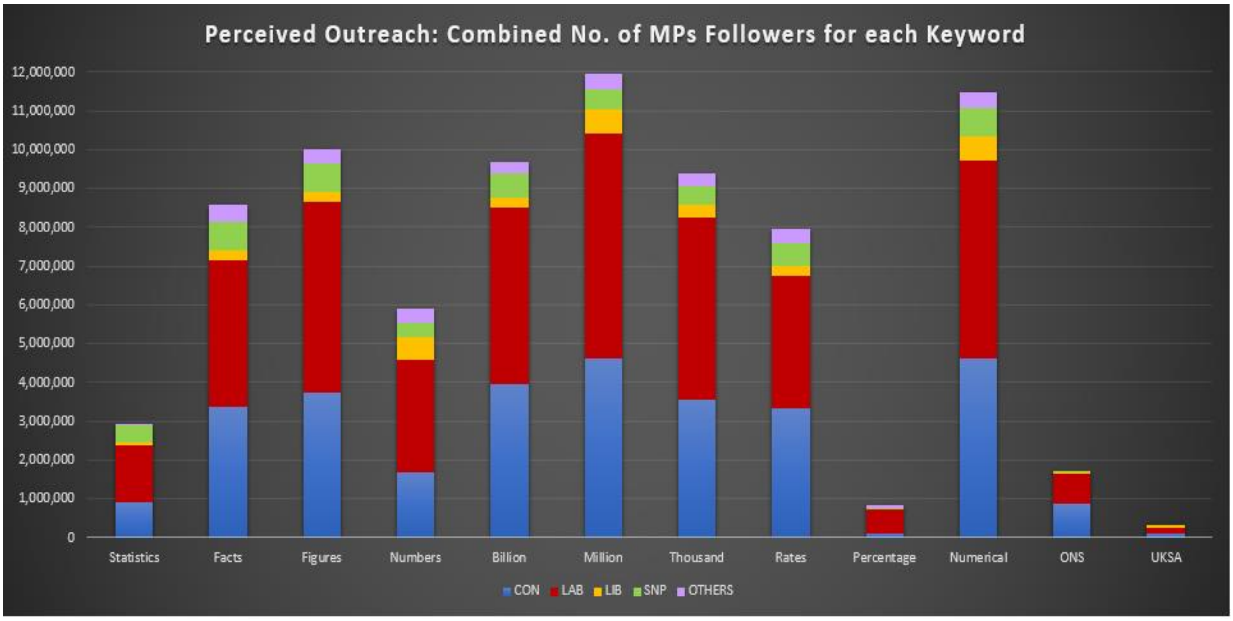
As noted, Labour is dominant in terms of tweets. Graph 18 below indicates the partisan attention to different keywords. In relative proportion, all parties (as already discussed above) do not vary significantly in terms of their level of attention to different keywords, despite differences in tweet frequency.

Graph 18. Number of tweets by a political party (absolute frequency).

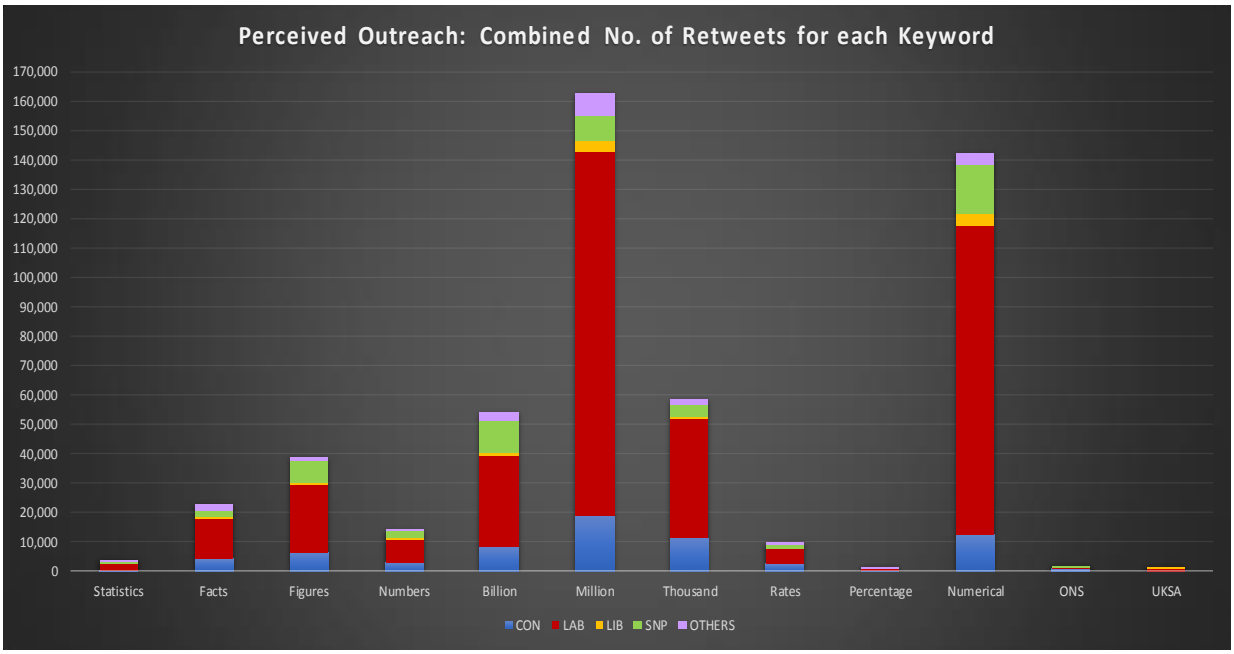


As noted, differences exist in terms of frequency, followers and re-tweets. The following graphs build the above characterisation of tweet frequency and highlight how attention to different keywords varies when viewed in terms of these different criteria. Of particular relevance is the attention to select keywords that are being re-tweeted. It might suggest that these keywords are associated with messages that gain specific traction or attention.

Graph 19. Perceived outreach: combined number of MPs' followers for each keyword.



Graph 20. Perceived outreach: combined number of re-tweets for each keyword.



The time period for twitter-based analysis is arguably too short to make statements in response to the wider question regarding trends over time. However, the varied patterns and the cyclical nature of the tweeting behaviour suggests that tweets and the use of numerical statements in tweets follows broadly the parliamentary timetable, namely peaks during spring and autumn statements, the release of key statistics (such as unemployment, immigration or inflation), or the opening of parliament.

IMPLICATIONS

The analysis has been based on non-reactive measures: gaming effects can therefore be excluded. The study is based on publicly available open information and can be regarded as an example of investigating maximum transparency (although this maximum transparency is impeded by the considerable challenges of cleaning up data sources for analysis). The communication crosses the different channels (and chambers), is intended to be public and for the consumption of different publics.

Overall, we cannot suggest that there has been a rise in the use of specific numerical statements over time. The patterns are highly 'peaky' and suggest that wider political and other events drive interest and supply rather than underlying 'cultural' restrictions. However, even though frequency patterns are broadly stable, the potential amplification effects of the digital world represent critical challenges when it comes to the political use of statistical information. Accessibility via digital forms (Hansard and gov.uk) have improved considerably. However, it is more likely that it is traces found on Google and Twitter that will create attention in a digital communication environment.

In terms of the wider implications of this study, the observed pattern does not suggest that political discourse has become increasingly 'quantified'. What might have changed is the way in which 'numbers' are used and with what consequences, but there has not been a major shift in attention. Further analysis might reveal whether numerical statements are taking place in different contexts (such as moving from discussions of unemployment to hospitals).

The examination of different communication channels (and within channels) reveals diverse patterns. Some keywords enjoy sustained popularity over time, others are consistently infrequent. The differences across the different data-chambers suggest that any further investigation needs to consider the separate and cumulative impact of the use of numerical statements across these different areas of communication. Given that there is arguably no overall trend towards greater use of numerical statements, regulatory resources are required to support differentiated detection tools to explore individual modes of communication in 'real time'. At the same time, while outside of this study, it is arguably not just the numerical statement as such which requires attention, but also its visualisation. For regulators, this requires a dashboard approach that assesses the use of numerical statements on a continuous basis across media rather than a fire alarm system.

ANNEX

Abbreviations: Observed Executive Offices and the Cabinet-level departments:

PM	Office of the Prime Minister	**DfE**	Department for Education
DPM	*Office of the Deputy Prime Minister	**DH**	Department of Health
HO	Home Office	**DWP**	Department for Work & Pensions

*DPM office was active during the Coalition Government (May 2010 – May 2015)

A#1 Keywords

This study explores whether the 'home' of statistics in the UK context has in itself become an increasingly frequent reference point in political statements. We have used the following keywords to determine the attention frequency in the context of different data-chambers and time-frames:

1. **Statistics**
2. **Facts**
3. **Figures**
4. **Numbers**
5. **Billion**
6. **Million**
7. **Thousand**
8. **Rates**
9. **Percentage**
10. **Numerical reference** *(a figure that contains at least one zero)*
11. **Office of National Statistics** (ONS / @ONS)
12. **UK Statistics Authority** (UKSA / @UKStatsAuth)

Due to search (tool) restrictions – algorithmic inability to isolate a keyword in their original context – certain words cannot be accurately identified and by extension counted as a representative sample. Therefore, we have decided to apply a mix-and-match approach when using different search platforms. For example, a 'numerical reference' (a figure that contains at least one 0) was only applied when we could manually check most of the result. Also, @ONS and @UKStatsAuth terms were only used in the context of our Twitter analysis as they were only applicable for that particular domain. We are confident that this approach has significantly reduced the margin of error across our research.

The following table lists keywords in relation to the observed data-chambers:

Data-Chamber #01	Data-Chamber #02	Data-Chamber #03	Data-Chamber #04A	Data-Chamber #05
Google Trends	**GOV.uk Info-Flows**	**British Political Speech archive**	**HoC Hansard records**	**Social Media - Twitter**
Statistics	Statistics	Statistics	Statistics	Statistics
Facts	Facts	Facts	Facts	Facts
Figures	Figures	Figures	Figures	Figures
Numbers	Numbers	Numbers	Numbers	Numbers
Billion	Billion	Billion	Billion	Billion
Million	Million	Million	Million	Million
Thousand	Thousand	Thousand	Thousand	Thousand
Rates	Rates	Rates	Rates	Rates
Percentage	Percentage	Percentage	Percentage	Percentage
		Numerical ref. (,0)	*Numerical ref. (,0)	Numerical ref. (,0)
		Office of National Statistics (ONS)	Office of National Statistics (ONS)	Office of National Statistics (ONS)
		UK Statistics Authority (UKSA)	UK Statistics Authority (UKSA)	UK Statistics Authority (UKSA)
				@ONS
				@UKStatsAuth

*Note that the 'numerical reference' keyword was only observed in the context of parliamentary debates at the House of Commons – speeches by all MPs; it was not observed in the context of selected group of senior MPs.

	Public Attentiveness
Data-Chamber (Source)	#01: Google Trends
Data URL	https://trends.google.co.uk/trends/
Original Data Format	• web-based Google Trends database (accessible to public)
Description	Google Trends is a public web facility of Google Inc., based on Google Search, that shows how often a particular search-term is entered relative to the total search-volume across various regions of the world, and in various languages. Non-real time data is a random sample of Google search data that can be pulled from as far back as 2004 and up to 36 hours prior to your search. Search results are proportionate to the time and location of a query. Each data point is divided by the total searches of the geography and time range it represents, to compare relative popularity. Otherwise places with the most search volume would always be ranked highest. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. Different regions that show the same number of searches for a term will not always have the same total search volumes.

Data-Scope	1 January 2004 – present
Observed Time-Line	1 January 2008 – 18 April 2017 (News category + News search filter is only available from 1 January 2008, and for the sake of data standardisation across all search filters and country domains, we have decided to adopt this date as our starting point)

Observed Data-Unit	Keyword search interest relative to the highest point on the chart for the given region and time (value range: 0–100)
Observed Domain/s	Worldwide – Australia – Canada – New Zealand – UK – USA [The Frequency of Public Attention]
Observed Keywords	9 keywords
large-N	Worldwide + 5 Anglosphere countries
small-N	United Kingdom

Data Protocol	• To correct for bias, reduce irrelevant associations with our keywords, and to standardise data-sets, we have used Google's triple-filter to calculate an average score for each keyword in the context of our selected geographic domains and fixed time-line: (1) All categories + Web search; (2) All categories + News search; and (3) News category + News search. • Results were downloaded in CSV format. • All the formatting, calculations, analysis and visualisations were conducted in the Excel programme.

	Government Communication
Data-Chamber (Source)	#02: GOV.uk Info-Flows
Data URL	https://www.gov.uk/government/announcements
Original Data Format	• web-based GOV.uk search platform (accessible to public) • info-flow results were displayed in .HTML format
Description	GOV.uk is a United Kingdom public sector information website, created by the Government Digital Service to provide a single point of access to HM Government services. It officially replaced Directgov and the online services of Business Link on 17 October 2012, and was launched as a beta on 31 January 2012. It contains information and services for citizens and businesses, detailed guidance for professionals, and information on government and policy.

Data-Scope	10 May 2010 – present
Observed Time-Line	10 May 2010 – 18 April 2017

Observed Data-Unit	**Number of Keyword references in GOV.uk Announcements:** • Ministerial Speeches • Departments' Press Releases (counting how many announcements have made a reference to our keywords – not the number of times the keyword itself was mentioned in the actual text) In the context of our research we have decided to focus only on two types of the announcements: Speeches and Press Releases. These two info-flows reflect strong agenda-setting communication and high public outreach.
Observed Domain/s	United Kingdom [The Frequency of Government Attention]
Observed Keywords	9 keywords
large-N	GOV.uk Announcements: (1) all info-flow categories for all departments across all policies (2) all speeches and press releases for all departments across all policies
small-N	GOV.uk Announcements: • Speeches and Press Releases relating to six Cabinet-level entities and in relation to all policies: PM –DPM – HO – DfE – DH – DWP

Data Protocol	• We have used GOV.uk search tool to set up a filter formula for each Keyword in the context of eight annual samples: (1) Keyword + (2) Speech / Press Release designation + (3) All policies + (4) Select a Department + (5) All locations and (6) Specify annual time-frame (e.g. 01.01.2012 – 31.12.2012). • Please note that the 2010 sample started from 10 May 2010; and 2017 sample ended on 18 April 2017. • The results were manually inputted into an Excel table. Due to scope and scale of the results, we were unable to open each speech / press release and manually check the information. We have relied on GOV.uk algorithmic accuracy. • Excel programme was used to format data, perform calculations, analyse patterns and create visualisations (graphs / tables).

	Speeches by Political Leaders
Data-Chamber (Source)	#03: British Political Speech archive
Data URL	http://www.britishpoliticalspeech.org/speech-archive.htm
Original Data Format	• web-based searchable database (accessible to public) • each speech is displayed as a text in .HTML format
Description	British Political Speech (BPS) is an online archive of British political speech and a place for the discussion, analysis, and critical appreciation of political rhetoric. The archive currently holds texts of speeches given by Conservative, Labour and Liberal/Liberal Democrat Party leaders going back to 1895.

Data-Scope	1895 – 2016
Observed Time-Line	1967 – 2016 The objective was to have a 50-year sample and to observe the Political Speech and Hansard within the same time frame (1967–2016/17). Note that the Archive did not hold any speeches that were delivered in 2017.

Observed Data-Unit	Number of Keywords in the speech by party political leaders (Conservatives, Labour and Liberal Democrats) Counting how many times each Keyword was mentioned in selected speeches.
Observed Domain/s	United Kingdom [The Frequency of Political Attention]
Observed Keywords	12 keywords
large-N	All speeches (1967–2016)
small-N	Leader's speeches / Top 15 Leader's speeches (1967–2016)

Data Protocol	• Use BPS's internal search tool to filter all speeches that were delivered between 1967 and 2016. • Copy the list into an Excel sheet and format/contextualise data. • Open each speech in the web browser, copy the speech text, and use a NotePad to create a .TXT format file which is free from .html formatting and irrelevant text. • **For each Speech (.txt file) we had to establish:** ***Total Word Count*** (TWC) – we have used UltraEdit and Notepad++ programmes to count the number of words in each speech. As we had two counts, the final total was the average of both results. ***Total Unique Word Forms*** (TUWF) – to count the number of unique and non-repetitive words we have used Voyant-Tools.org software. ***Keyword Frequency*** – we have used UltraEdit and Notepad++ programmes to count the number of times each keyword was mentioned in the speech. • Results were formatted, calculated, analysed and visualised in an Excel programme. We have used the following formula to calculate the Relative Frequency: Keyword Count x 100 / TWC or TUWF.

	Speeches in Parliament [A]
Data-Chamber (Source)	#04: House of Commons Hansard records (using TheyWorkforYou.com platform)
Data URL	https://www.theyworkforyou.com/pwdata/scrapedxml/debates/
Original Data Format	• web-based file directory in .XML format (accessible to public) • records are segmented in series of smaller files (e.g. 1967 records were spread across 162 individual files)
Description	***Hansard*** is the traditional name of the transcripts of Parliamentary Debates in Britain and many Commonwealth countries. Current UK records stretch from 1804 to present. ***TheyWorkForYou.com*** launched in 2004, taking data and information from official parliamentary sources and adding features that make them easier to understand.

Data-Scope	1 Nov 1935 – present
Observed Time-Line	1 January 1967 – 18 April 2017 The objective was to have a 50-year sample and to observe the Political Speech and Hansard within the same time frame (1967–2016/17).

Observed Data-Unit	Number of times the Keywords were mentioned during the parliamentary debates (on annual basis) at the House of Commons – taking into the account speeches by all MPs across all parties and in relation to all debates/policies/issues.
Observed Domain/s	United Kingdom [The Frequency of the Legislative Attention]
Observed Keywords	12 keywords
large-N	Parliamentary debates 1967 – 2017
small-N	n/a

Data Protocol	• It is important to note that there are several web-based databases that host UK Hansard records (Hansard-Corpus.org; TheyWorkforYou.com and the official UK Parliament). However, none of them have a capacity to run a keyword query that encompasses the entire digital archive (1804 – present). In addition, there are serious discrepancies between these platforms as to how they count and/or define the keyword frequency when scanning through the records. As a result, one gets different results for the same record/year. • In order to address these obstacles, we have decided to download the raw-data files from TheyWorkforYou.com server and re-format all the Hansard files from 1967 to 18 April 2017. This was a time-consuming exercise, but it was the only way that we could assure greater accuracy of the results. • Once the .xml files were downloaded, we had to (1) convert them into a .TXT format using a Notepad++; and (2) amalgamate all the

Data Protocol – cont.	individual files to form one massive annual record for each year in question (e.g. records for 1967 were spread across 162 individual files). Once that was accomplished, we could proceed with the keyword count. • The objective was to count how many times each Keyword was mentioned during the parliamentary debates and record those results in the context of annual sample. We have focused only on the keyword frequency and not its relationship with the MP, debate, particular policy/issue and/or the political party affiliation. • **For each Annual record we had to establish:** ***Total Word Count*** (TWC) – we have used UltraEdit and Notepad++ programmes to count the number of words in each speech. As we had two counts, the final total was the average of both results. ***Total Unique Word Forms*** (TUWF) – to count the number of unique and non-repetitive words we have used Voyant-Tools.org software. ***Keyword Frequency*** – we have used UltraEdit and Notepad++ programmes to count the number of times each keyword was mentioned in the speech. • Results were formatted, calculated, analysed and visualised in an Excel programme. We have used the following formula to calculate the Relative Frequency: Keyword Count x 100 / TWC or TUWF.

	Speeches in Parliament [B]
Data-Chamber (Source)	#04: House of Commons Hansard records (using official Hansard Online platform)
Data URL	https://hansard.parliament.uk/
Original Data Format	• web-based searchable database (accessible to public) • results listed on the website in .HTML format
Description	***Hansard*** is the traditional name of the transcripts of Parliamentary Debates in Britain and many Commonwealth countries. Current UK records stretch from 1804 to present. ***TheyWorkForYou.com*** launched in 2004, taking data and information from official parliamentary sources and adding features that make them easier to understand.

Data-Scope	10 May 2010 – present
Observed Time-Line	10 May 2010 – 18 April 2017

Observed Data-Unit	Number of Keyword references made during the parliamentary debates at the House of Commons by a selected group of senior MPs. (counting how many references were made to our keywords by the selected group of senior MPs as part of their speeches at the House of Commons – not the number of times the keyword itself was mentioned during those debates)
Observed Domain/s	United Kingdom [The Frequency of the Legislative Attention]
Observed Keywords	11 keywords
large-N	12 senior MPs who occupied the executive function within the six Cabinet-level entities: PM – DPM – HO – DfE – DH – DWP
small-N	n/a

Data Protocol	• We have used the official Hansard Online platform search tool to setup a filter formula for each Keyword in the context of our time-line and focus on 12 senior MPs: (1) Keyword + (2) MP's name + (3) All references + (4) Specify annual time-frame (e.g.: 01.01.2012 – 31.12.2012). The results were displayed within the webpage in .html format. • Please note that the 2010 sample started from 10 May 2010; and 2017 sample ended on 18 April 2017. • The results were manually inputted into an Excel table. Due to scope and scale of the results, we were unable to open each record and manually check the information. Hence, why we refer to number of 'reference' that have been made in relation to our keywords; and not the exact number of times each keyword was mentioned by those select 12 MPs. We have relied on Hansard's algorithmic accuracy. • Excel programme was used to format data, perform calculations, analyse patterns and create visualisations (graphs / tables).

	Social Media
Data-Chamber (Source)	#05: Twitter (via MPsonTwitter.co.uk platform)
Data URL	http://www.mpsontwitter.co.uk/
Original Data Format	• web-based searchable database (accessible to public) • results listed on the website in .html format and downloaded in .CSV format
Description	MPsonTwitter.co.uk provides an interactive way to explore the way in which Members of the UK parliament use the social media website Twitter

Data-Scope	10 March 2015 – present (reflecting the mandate of the 56th UK Parliament)
Observed Time-Line	10 March 2015 – 18 April 2017

Observed Data-Unit	Number of times our Keywords were mentioned in tweets posted by the MPs during the 56th UK Parliament
Observed Domain/s	United Kingdom [The Frequency of the Legislative/Political Attention]
Observed Keywords	14 keywords
large-N	All tweets with keyword references
small-N	Top 15 tweets by outreach factor (number of retweets / favourites)

Data Protocol	• We have decided to use the existing MPsonTwitter.co.uk platform for two reasons: (1) it is the only tool currently available that aggregates and contextualises tweets in reference to UK MPs – it is incredibly resourceful, accurate and very user friendly; (2) it allows us to observe the Twitter traffic in the context of a single parliamentary mandate. • We have used their search tool to filter out all the tweets that make reference to our Keywords. In addition, we have also tested how many times the MPs have made a reference to ONS and/or UKSA by using their official Twitter Handles (@ONS / @UKStatsAuth). • Search formula was as follows: (1) Keyword + (2) Literal Search option + (3) Date specification (10.03.2015 – 18.04.2017) + All MPs option. • The results were displayed both in .html format and were available for download in .csv format (our preferred option). • Downloaded .csv format files were formatted, analysed and visualised using the Excel programme. In the process, we had to manually check all the results – especially in when it came to 'rates' and 'numerical reference' terms – in order to assure that the final count was as accurate as possible.

A#3 Senior MPs/Politicians

Throughout our research we have focused on six Cabinet-level entities:

1. Office of the Prime Minister [PM]
2. Office of the Deputy Prime Minister [DPM] (active during the 'Coalition' government 2010–15)
3. Home Office [HO]
4. Department for Education [DfE]
5. Department of Health [DH]
6. Department for Work and Pensions [DWP]

As such, it was deemed important to focus on senior ministers (incumbent postholders and their predecessors) in charge of these six entities in the period between 10 May 2010 and 18 April 2017. The objective was to assess their level of association with our Keywords (number of references they have made in the House of Commons); and whether any patterns emerge in the context of their legislative vs. executive mandate. The following textual and visual chronology maps-out their executive roles since May 2010. Roles in **red** indicate the positions that are of interest to this study.

THERESA MAY

- **Prime Minister** (13 July 2016 – present)
- **Home Secretary** (10 May 2010 – 13 July 2016)
- Elected MP (since 1997)

DAVID CAMERON

- **Prime Minister** (10 May 2010 – 13 July 2016)
- Elected MP (June 2001 – 12 September 2016)

NICK CLEGG

- **Deputy Prime Minister** (10 May 2010 – 8 May 2015)
- Elected MP (May 2005 – 8 June 2017)

AMBER RUDD

- **Home Secretary** (July 2016 – present)
- Secretary of State for Energy and Climate Change (May 2015 – July 2016)
- Parliamentary Under-Secretary of State for Climate Change (July 2014 – May 2015)
- Parliamentary Private Secretary to the Chancellor of the Exchequer (September 2012 – October 2013)
- Elected MP (May 2010)

JUSTINE GREENING

- **Secretary of State for Education & Minister for Women and Equalities** (July 2016 – present)
- Secretary of State for International Development (September 2012 – July 2016)
- Secretary of State for Transport (October 2011 – September 2012)
- Economic Secretary to the Treasury (May 2010 – October 2011)
- Elected MP (May 2005)

NICKY MORGAN

- **Secretary of State for Education** (July 2014 – July 2016)
- Minister for Women and Equalities (April 2014 – July 2016)
- Financial Secretary to the Treasury (April 2014 – July 2014)
- Economic Secretary to the Treasury (October 2013 – April 2014)
- Elected MP (May 2010)

MICHAEL GOVE

- **Secretary of State for Justice & Lord Chancellor** (May 2015 – July 2016)
- Chief Whip of the House of Commons & Parliamentary Secretary to the Treasury (July 2014 – May 2015)
- Secretary of State for Education (May 2010 – July 2014)
- Elected MP (May 2005)

JEREMY HUNT

- **Secretary of State for Health** (September 2012 – present)
- Secretary of State for Culture, Olympics, Media and Sport (May 2010 – September 2012)
- Elected MP (May 2005)

ANDREW LANSLEY

- **Leader of the House of Commons** (September 2012 – July 2014)
- Lord Privy Seal (September 2012 – July 2014)
- Secretary of State for Health (May 2010 – September 2012)
- Elected MP (May 1997 – March 2015) – Appointed Peer at the House of Lords (August 2015)

DAMIAN GREEN

- **Secretary of State for Work and Pensions** (July 2016 – 11 June 2017)
- Minister of State for Policing and Criminal Justice (September 2012 – July 2014)
- Minister of State for Immigration (May 2010 – September 2012)
- Elected MP (May 1997)

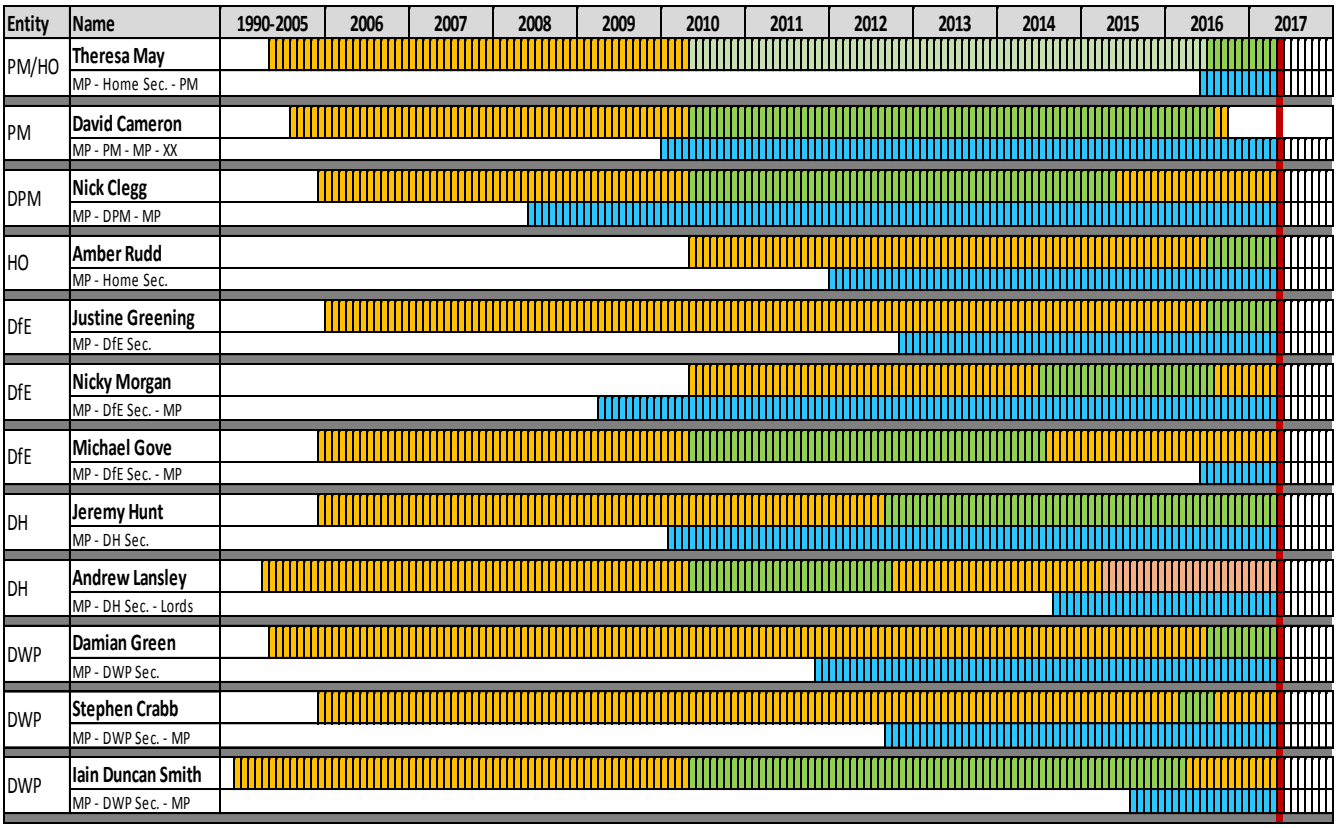
STEPHEN CRABB

- **Secretary of State for Work and Pensions** (March 2016 – July 2016)
- Secretary of State for Wales (July 2014 – March 2016)
- Parliamentary Under-Secretary of State for Wales (September 2012 – July 2014)
- Elected MP (May 2005)

IAIN DUNCAN SMITH

- **Secretary of State for Work and Pensions** (May 2010 – March 2016)
- Elected MP (April 1992)

Graph 21. Senior MPs: Legislative - Executive mandate + Twitter adoption time-line



LEGEND

- Politician's MP mandate
- Politician's Executive mandate (our focus)
- Theresa May's HO Executive mandate
- Politician's post-Executive mandate (switch from MP to Peer status)
- Politician's Twitter adoption time-line
- 18.04.2017 (end of our time-sample)

carr
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Email: risk@lse.ac.uk

Tel: + 44 (0) 20 7955 6577
Fax: + 44 (0) 20 7242 3912

Website:
http://www.lse.ac.uk/accounting/CARR/home.aspx