ŁUKASZ STOLARSKI

Orcid: 0000-0002-2668-5509

Jan Kochanowski University in Kielce

EFFECTS OF SERVICE DURATION ON PARTICIPATION IN PARLIAMENTARY DEBATES: A HANSARD CORPUS ANALYSIS

K e y w o r d s: corpus analysis, Hansard Corpus, parliamentary debates, work participation, work productivity.

ABSTRACT

The aim of this study is to examine the relationship between MPs' length of service in the House of Commons and their level of participation in parliamentary debates throughout the 20th century. To accomplish this task, the version of the Hansard Corpus available at https://api.parliament.uk/historic-hansard/ (access: Dec 10, 2025) was used. With a Python script, full debate transcripts were extracted and annotated with details such as MPs' names, genders, and the dates of the speeches. After further processing of the data, the average number of word tokens spoken per year of parliamentary service was measured and analyzed as an indicator of participation. The overall results reveal an increase in verbal output during the initial years of service, followed by a stabilization of average word counts in later years. Nevertheless, an additional diachronic analysis indicates significant shifts in this pattern. During the first four decades of the 20th century, the relationship is largely linear. In later decades, however, it gradually evolves into an inverted U-shaped trajectory. Verbal output peaks earlier and declines among longer-serving MPs. Additionally, possible effects of gender differences were explored, but the results suggest that these diachronic shifts are not driven by changes in the proportion of male to female MPs throughout the century. The findings for the last three decades confirm assumptions found in the literature, but the patterns observed in earlier periods are not predicted by studies on contemporary tenure-productivity interactions. This underscores the need for more research on the diachronic evolution of these processes.

1. 1 AIMS AND STRUCTURE OF THE PAPER

The major aim of this study is to estimate the effects of MPs' service duration in the House of Commons throughout the 20th century on their participation in parliamentary debates, quantified by the average number of word tokens spoken per year of service in Parliament. This issue has never been investigated, and the general literature on the effects of job tenure on worker engagement does not provide a clear prediction.

The choice to focus exclusively on debates in the House of Commons and exclude the House of Lords from the analysis is based on the fact that many MPs transitioned from the former to the latter. Including data from the House of Lords would make the task of years-in-Parliament labelling, described in Section 3, more complex and prone to errors.

The paper begins with a general discussion of the British Hansard Corpus and other collections of parliamentary debates currently available (Section 2). Next, the methods applied in the study are outlined (Section 3). These include the procedures used to obtain and process the analysis data. Basic statistics of these data are also provided. The findings are presented in three parts. In Section 4.1, general results based on the entire 20th century are discussed. In Section 4.2, these data are examined separately for three different periods within the 20th century: 1900–1939, 1940–1969, and 1970–1999. This analysis reveals crucial diachronic changes in the way MPs participate in debates per years of service. Subsequently, Section 4.3 examines possible effects of MPs' gender. Finally, all these findings are summarized and discussed in Section 5.

2. BACKGROUND

2.1. Hansard Corpus

The name "Hansard" was originally used in reference to the collection of transcripts of parliamentary debates in Britain. It comes from the name of the first official printer to the Parliament at Westminster, Thomas Curson Hansard, although initial attempts at publishing parliamentary debates had been undertaken earlier by other printers, including William Cobbett (Alexander, 2023).

Comparable collections were later compiled and digitized for other Commonwealth countries, including Canada, Australia, and New Zealand. Moreover, in recent decades, similar documentation of parliamentary debates has been undertaken in numerous other parts of the world, e.g., in Africa (Parliament of South Africa, National Assembly of Kenya, National Assembly of Tanzania, Parliament of Ghana, Parliament of Uganda, National Assembly of Nigeria, etc.), Asia (Parliament of Malaysia, Parliament of Singapore, Legislative Council of Brunei, etc.), North America (Parliament of Jamaica, Parliament of Trinidad and Tobago, Parliament of Barbados), and Oceania (National Parliament of Papua New Guinea). These collections also happen to be referred to as "Hansard" and, consequently, the term has gained a more generic

meaning as a system for creating corpora of political debate rather than referring to any specific corpus of texts.

This paper focuses on the original British Hansard. However, there are numerous versions of this collection available online, differing in scope and annotation type. The most popular sources include "English-Corpora.org" (Alexander & Davies, 2022), which is also known for hosting many other popular corpora, such as the BNC (British National Corpus), COCA (Corpus of Contemporary American English), and COHA (Corpus of Historical American English). The version of Hansard it offers spans from 1803 to 2005 and encompasses around 1.6 billion word tokens. The graphical user interface available makes it possible to perform queries involving part-of-speech as well as semantic tagging. There is also functionality for filtering searches by specific decade, parliamentary house (House of Commons vs. House of Lords), and political party. The British Hansard is also available on the official UK Parliament website at https://hansard.parliament.uk. This version of the corpus encompasses both the historical data as well as texts of the newest debates, which are regularly updated. Additionally, one may find many other sources offering access to the corpus. For instance, the version available at https://api.parliament.uk/historic-hansard/, which is used in this study (see Section 3), offers an API for obtaining the entire contents of debates. The corpus can also be accessed through the Huddersfield web interface (Jeffries et al., 2023) and the platform described by M. Coole et al., (2020). Finally, transcripts of parliamentary debates in Britain have also been processed for specific purposes, although the resulting corpora may not be available through any concrete graphical user interface. This is the case in E. McGill & H. Saggion (2023), who attempt to align the text of debates with British Sign Language visible in corresponding videos, and also in F. Nanni et al., (2019), and O. Onyimadu et al., (2013), where different semantic annotations are introduced.

Various versions of Hansard have been used in numerous studies on syntax (Labat et al., 2023; Rickman & Rudanko, 2018), morphology (Micher, 2018), and semantics (Abercrombie & Batista-Navarro, 2019; Archer, 2017; Demmen et al., 2018). There is also a vast body of literature on sociolinguistic and interdisciplinary topics that analyze texts from political debates. Among these topics are differences in language use between men and women (Porro, 2022), the colloquialization and democratization of language (Hiltunen et al., 2020; Hiltunen & Vartiainen, 2024; Kruger & Smith, 2018), sentiment and emotion analysis (Abercrombie & Batista-Navarro, 2018; Jenkins, 2019; Leibel, 2024; Vardanega & Smith, 2016), politeness in language (Archer, 2017, 2018), and code-switching (Carpuat, 2014). Still, possibly the most common studies focus on various sociolinguistic and sociocultural aspects using lexical frequency analysis (Alexander & Struan, 2022; Appleton, 2023; Corbett, 2021; Demmen et al., 2018; Ingham, 2022; Kang, 2023; Kotze & Van Rooy, 2020; Smith & Korhonen, 2022; Soriano-Jiménez, 2024). Finally, it is worth noting that numerous papers examine linguistic aspects of parliamentary debate datasets explicitly referred to as "Hansard" in languages other than English, such as Malaysian (Awal et al., 2019, 2019;

Che Abdul Rahman et al., 2019; Rahman & Nadiah, 2024) and Nunavut (Joanis et al., 2020; Micher, 2018).

2.2. Effects of Work Tenure on Productivity

Research on the potential effects of work tenure on work productivity is difficult to summarize for two reasons. First, the very term productivity is rather ambiguous. In publications on job-related issues, it tends to be used alongside expressions such as "work performance" or "work engagement." Their exact meanings differ across publications and are also sometimes used interchangeably (Gagliardi et al., 2021). In the present study, no clear-cut distinction is made between these terms, and they are used in reference to work output quantified by the number of word tokens uttered by individual MPs. Second, these concepts are measured using different methods and metrics, which may produce dissimilar results and lead to seemingly contradictory conclusions.

Nevertheless, several broad tendencies can be identified. Some studies indicate no correlation between employment duration and work engagement (Rice, 2009), while others suggest that work productivity increases after three years (Batistuta & Safitri, 2024). There are also suggestions that productivity follows a "U-shaped" pattern: after initially high levels of productivity, there is a relative decrease lasting several years, which is eventually followed by an increase (Bakar, 2013; Coffman et al., 2002; Dongrey & Rokade, 2019; Swaminathan & Ananth, 2009). However, the largest number of publications point to a possible opposite trend, or an "inverted U-shaped" relationship. It is suggested that longer tenures result in "diminishing returns" in performance (Ng & Feldman, 2010; Shaw & Lazear, 2008). Likewise, "sunk costs," which refer to the investments that both employees and employers make in their relationship, can negatively impact productivity (Cohen, 1993; Meyer et al., 1989), as can "continuance commitment," where employees remain in their roles due to the perceived costs of leaving (N. J. Allen & Meyer, 1990; Uppal, 2017). Additionally, several studies discuss the impact of ageing on cognitive and physical abilities, which can also negatively affect performance (J. Allen & De Grip, 2012; Bryson et al., 2020; Bussolo, Koettl, and Sinnott, 2015; De Grip, 2006; Desjardins & Warnke, 2012; Giniger et al., 1983; Picchio, 2021; Van Loo et al., 2001). All such considerations have led some researchers to conclude that the relationship between tenure and performance follows an "inverted U-shaped" pattern, in which there is an initial increase, followed by plateauing and a consequent decrease in productivity (Gagliardi et al., 2021; Shaw & Lazear, 2008).

At this point, it must be recognized that the type of work carried out by politicians, as analyzed in this study, differs from many "standard" occupations in corporate environments. Therefore, the findings outlined above should be interpreted with caution and seen only as a broad framework for comparison. Still, a similar argument could be made for most professions. Work environments and demands differ widely across occupations, which makes direct comparisons with such research results inherently limited.

3. METHODS

To achieve the aims outlined in Section 1, the version of the Hansard Corpus available at https://api.parliament.uk/historic-hansard/ (access: Dec 10, 2025) was used. It provides access to full debate texts and allows searches for speeches by individual MPs. Moreover, in addition to its basic API functionality, its simple user interface makes it particularly suitable for the web scraping methods applied in this study.

An outline of the data collection and processing is presented in Fig. 1. First, a Python script was written to extract texts from all debates in the House of Commons throughout the entire twentieth century. Among other modules, the script utilized the "Beautiful Soup" library for pulling data out of HTML and XML files (Leonard, 2007). The collected data included each utterance (an uninterrupted sentence or group of sentences spoken by one person), the name of the MP who made the utterance, the debate title, and the debate year. In total, 5,118,968 utterances were recorded. However, for 205,976 of them no specific MP could be assigned, so these cases were removed, resulting in a final dataset of 4,912,992 utterances.
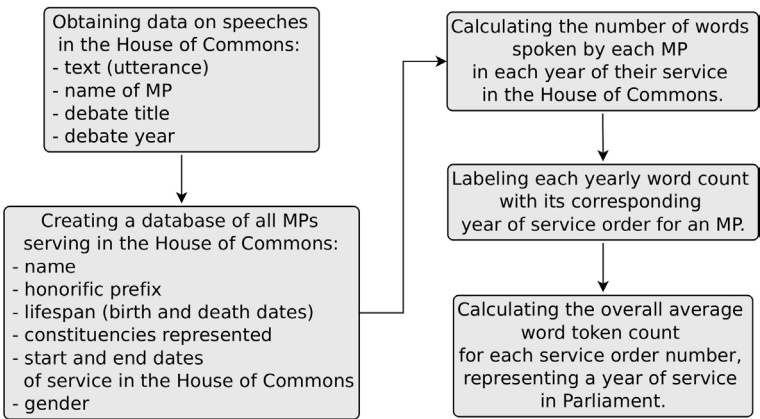
Obtaining data on speeches
in the House of Commons:
- text (utterance)
- name of MP
- debate title
- debate year

Creating a database of all MPs
serving in the House of Commons:
- name
- honorific prefix
- lifespan (birth and death dates)
- constituencies represented
- start and end dates
  of service in the House of Commons
- gender

Calculating the number of words
spoken by each MP
in each year of their service
in the House of Commons.

Labeling each yearly word count
with its corresponding
year of service order for an MP.

Calculating the overall average
word token count
for each service order number,
representing a year of service
in Parliament.

Fig. 1. Workflow for Data Collection and Processing

Next, a database of all MPs who served in the House of Commons was compiled. This was achieved with another Python script which extracted names from the section "people" in the Hansard Corpus. Additionally, data on honorific prefixes, life spans, constituencies represented, and service spans in the House of Commons were collected. Crucial to the analysis in Section 4.3, however, was to establish the gender of each MP. This process involved several stages. Initially, many honorific titles were used as an indicator of gender. For instance, MPs holding titles such as "Mr," "Sir" or "Lord" were classified as male, and MPs holding titles such as "Ms," "Mrs" or "Lady" were classified as female. This resulted in a majority of the MPs in the list being properly

categorized. Still, some titles such as "Dr" or "Professor" were not indicative of gender and applying additional methods was necessary. These methods included obtaining a list of names of all female MPs in the British Parliament from Wikipedia. This list was compared with the list of MPs found in the Hansard Corpus, resulting in improved gender classification. Finally, any cases which were still ambiguous because neither "title classification" nor "Wikipedia comparison" resulted in unequivocal categorization were manually verified.

The final counts obtained from the Hansard Corpus for all MPs who actively participated in parliamentary debates in the House of Commons in the 20th century, separated by gender, are presented in Table 1. These figures highlight a significant gender disparity (see also Table 3 for counts across different periods of the 20th century).

Table 1
Number of MPs taking part in debates in the House of Commons in the 20th century

Male MPs	Female MPs	All MPs
5,111	230	5,341

Punctuation was removed from all the texts collected, and they were split into words using the Python string method ".split()." The summary statistics of the word tokens obtained separated by MPs' gender and the periods within twentieth century analyzed in Section 4.2 is provided in Table 2. Overall, the portion of the Hansard Corpus used in this study encompasses over 800 million word tokens. About 773 million were used by male MPs, and only 30 million by female MPs. In relative terms, this is a ratio of 1 to 25. This disproportion decreased in the last three decades of the twentieth century (1 to 15), but was much more pronounced in the period before the Second World War. Between 1900 and 1939 male MPs used almost 230 million word tokens, while female MPs used less than 2 million, which indicates large disproportions in the participation in political discourse between the two genders. The ratio for this period is 1 to 141. These differences primarily reflect the disparity in the number of male and female MPs, as summarized in Table 1 (for the entire 20th century) and in Table 3 (for the periods analyzed in Section 4.3). However, this issue warrants further investigation in future studies.

Table 2
Word token counts in the House of Commons throughout 20th century

Period	Male MPs	Female MPs	All MPs
1900-1939	229,753,785	1,621,091	231,374,876
1940-1069	249,012,259	9,031,511	258,043,770
1970-1999	294,120,756	19,811,175	313,960,320
1900-1999	772,915,189	30,463,777	803,378,966

The next stage involved calculating the number of words spoken by each MP in each year of their service in the House of Commons (see Fig. 1). This was achieved using the aggregation functionality of the "pandas" data analysis package for Python (The pandas development team 2023). After that, another Python script was used to label each yearly word count with the corresponding year of service for each MP. For example, an MP who served from 1955 to 1964 would have their word count for 1955 labelled as "1," for 1956 as "2," and so on, with the word count for 1964 labelled as "10." The only exceptions requiring adjustment to this approach were MPs who began their service in the House of Commons before 1900. Such cases were identified using the MPs database mentioned above, which includes each MP's start year. The labels were then adjusted by adding the number of years served before 1900 to the established year-in-Parliament label. For instance, an MP who served from 1895 to 1904 would have their word count for 1900 labelled as "6" ((1900–1895)+1), for 1901 as "7" ((1900–1895)+2), and so on, with the word count for 1904 labelled as "10."

It should be noted that this labelling does not account for breaks in an MP's service. For instance, a given MP may have been elected for one term of five years, then not elected for the next term, but returned to the House of Commons for another term after that. This involves a span of 15 years, but the labels assigned to the word counts for this MP would include only 10 years of their actual activity in Parliament. Therefore, the labels do not necessarily represent consecutive years, although in many cases MPs served without any breaks.

Finally, the last stage of data processing involved calculating the average word token count for each year of service across multiple MPs. This was performed on various subsets of the data. The analysis presented in Section 4.1 includes all MPs, whereas the results in subsequent sections focus on cases filtered based on different periods within the 20th century and the gender of MPs.

4. RESULTS

4.1. MPs' Participation in the Entire 20th Century

Before discussing mean word token counts per year of service, which are the main focus in this study, it is useful to consider the overall retention of MPs throughout the entire 20th century. This is shown in detail in Fig. 2. The count for the first year is 4771, which is smaller than the total number of MPs mentioned in Section 3 (5,341). This is due to the fact that some MPs began their service in the House of Commons before 1900, and the year-in-Parliament labels for their yearly word counts were adjusted (see the discussion in Section 3). The plot shows that this number of MPs gradually declines each year. The longest tenure recorded is 55 years, achieved by only one MP.

It is worth noting that the bar plot does not exhibit patterns corresponding to regular parliamentary terms. The decline in MP numbers appears gradual rather than stepwise,

reflecting the fact that parliamentary elections were not held at fixed intervals during the 20th century. In the first decade, the maximum duration of a Parliament was seven years, as established by the "Septennial Act 1715." This was reduced to a maximum of five years by the "Parliament Act 1911." However, due to various political and historical factors, this theoretical maximum was often not followed, with elections being called early or, in some cases, parliamentary terms being extended, such as during both World Wars. These irregularities influenced how average values were calculated in this study, resulting in the absence of discernible "steps" in the area graph.

The shape in Fig. 2 suggests a strongly right-skewed distribution. To explore this pattern further, three different models are compared in Fig. 3, where the MP counts are presented on a logarithmic scale. The solid line, representing values predicted by a linear model, follows the observed values for the first 13 years but diverges afterward. From the 42nd year onward, the model even predicts negative values, which are not shown in the graph. Consequently, the overall fit is relatively poor ($R^2 = .798$). Attempts to fit a power-law model (the dotted line in Fig. 3) were even less successful ($R^2 = .660$). In contrast, an exponential model (the dashed line in Fig. 3) closely follows the observed values and captures MP retention most accurately ($R^2 = .951$, $F(1, 53) = 1031$, $\beta = -0.150$, $p < .001$, 95% CI [-0.159, -0.141]).

In the subsequent parts of the paper, MP retention is illustrated with area graphs; however, these are provided only for rough comparisons and the main focus is placed on line graphs representing word token counts.

The line plot in Fig. 4 depicts the mean word token counts per year of service in the House of Commons throughout the 20th century. To ensure statistical representativeness, a minimum threshold of 30 MPs was set, so counts for years above 40 are not shown (the final three categories – 38, 39, and 40 years – are represented by 79, 56, and 41 MPs, respectively). It is readily apparent that active verbal participation increases sharply at the very beginning of service. The average word token count is 4,268 in the first year, rising to 9,011 in the second year and reaching 10,601 in the third year – more than double the amount observed in the first year. After that, the increase slows slightly for the next three years but accelerates again in the seventh and eighth years, surpassing 14,500 word tokens. This pattern of alternating slower and faster increases continues until approximately the thirteenth year of service, after which the mean values hover around 17,500 word tokens. However, likely due to diminishing sample sizes, the pattern becomes more irregular, which is visible in the "zig-zag" between the 31st and 35th years. Nevertheless, the highest mean word token count is recorded in the 35th year of service (20,628), and the values do not decrease even for longer tenures. In fact, the average word counts for 36th, 37th, 38th, 39th and 40th year of service oscillate around 18,500 word tokens. This increasing trend is only broken after the 43rd year of service, when the average values become highly irregular, but these measurements are based on samples smaller than 30 MPs and are not shown in Figure 4.
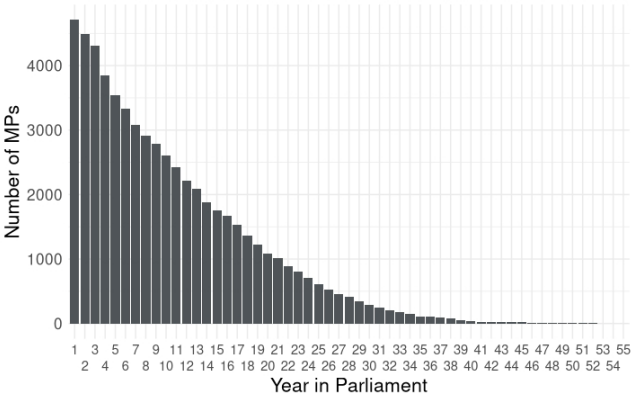
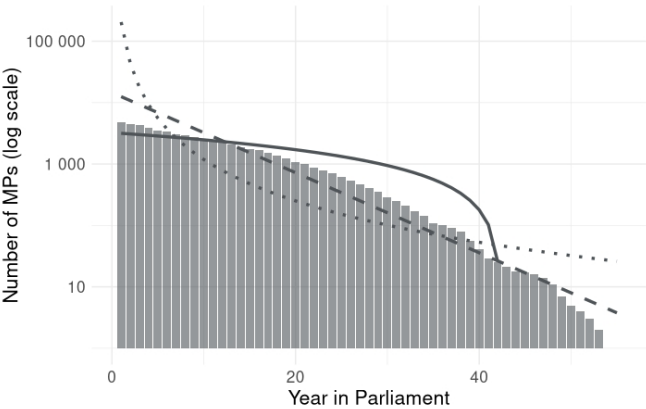*Fig. 2.* MP Retention by Years of Service

Fig. 3. Model Fits: Linear (solid), Exponential (dashed), and Power-Law (dotted)
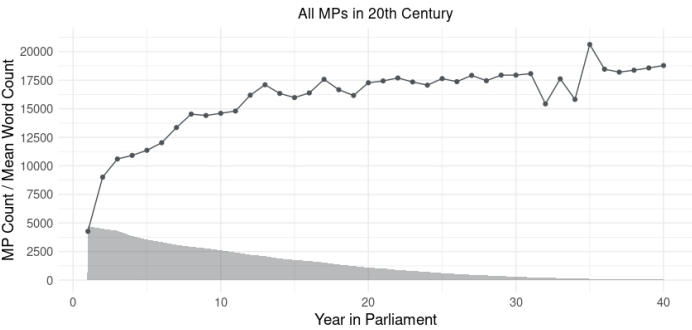

Fig. 4. MP Count (Area Graph) and Mean Word Count (Line Graph) per Year of Service in the House of Commons – All MPs in the 20th Century

4.2. Diachronic Trends in the 20th Century

The line plots in Fig. 5 illustrate the mean word token counts per year of service in the House of Commons across three periods of the 20th century. These periods are not based on historical events, except for the first, which ends with the outbreak of the Second World War. The division between 1940–1969 and 1970–1999 is largely arbitrary and was made to ensure relatively long periods in which the mean word token counts for several years of service would be represented by sufficiently large groups of MPs.

In the first four decades, shown in the left plot, word counts are almost linearly correlated with service duration. In fact, the results of a simple linear regression analysis performed on the measurements for the first 30 years in Parliament indicate a highly statistically significant trend (R^2 = .941, $F(1, 28)$ = 446.8, β = 0.00159, $p < .001$, 95% CI [0.00144, 0.00173]). This pattern changes only after the 30th year, where the values stabilize between 20,000 and 23,000, with the exception of the 35th year, for which the average word token count reaches 28,951. Finally, the counts start to decrease after the 40th year. However, these measurements are not shown in Fig. 5 because they are based on samples smaller than 30, which cannot be assumed to be statistically representative.

The trend in the middle line plot in Fig. 5 differs from the first period. The average word token counts for 1940–1969 rise much more quickly in the initial years in Parliament. In fact, the level of 11,000 word tokens is reached after just three years, whereas in the previous period, this took 10 years. Similarly, the threshold of 15,000 word tokens is reached within eight years, while in the 1900–1939 period, it took 14 years. Another notable difference is that the increasing trend slows down after the 10th year. While average word counts continue to rise up to about the 25th year in Parliament, the trend becomes less steep. Finally, after this point, a small but noticeable decline in average word counts is observed. The values fall below 20,000 but remain above 15,000.
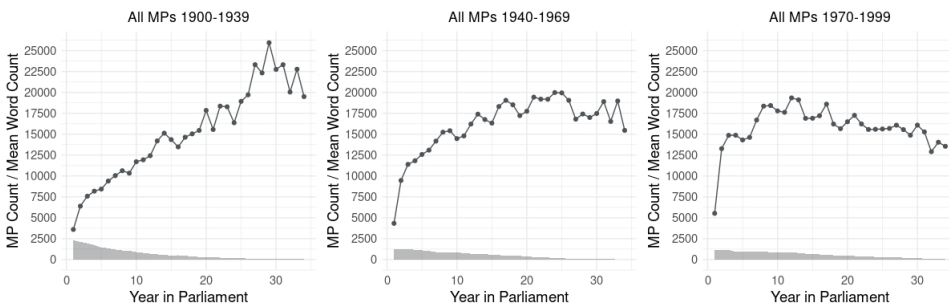

Fig. 5. MP Count (Area Graphs) and Mean Word Count (Line Graphs) per Year of Service in the House of Commons – All MPs in three different periods

The changes are even more dynamic in the right line plot in Fig. 5, which depicts the average word token counts per year in Parliament for the last three decades of the 20th century. The threshold of 15,000 word tokens is reached within just three years, compared to eight years in the 1940-1969 period and 18 years in the 1900-1939 period. This indicates a significant acceleration in MPs' participation in parliamentary debates over the twentieth century. The increasing trend continues, peaking in the 12th year with an average word token count of 19,342. After that, a distinct decrease in word counts is observed. Such a rapid reversal of the trend is not seen in either of the two previous line plots. This further underscores the accelerating changes in MPs' overall participation in parliamentary debates.

4.3. Gender Differences in Parliamentary Participation

The results obtained in Section 4.2 need to be further explored for a possible interaction with MPs' gender. While the overall trends seem to consistently change across different periods, this may not be due only to diachronic shifts in parliamentary debate involvement. As indicated in Table 1, there is an overall large gender disparity throughout the entire 20th century, but these differences gradually decrease over time. This is shown in Table 3, which summarizes the number of male and female MPs in the different periods scrutinized in Section 4.2. These data are not directly comparable to the summary in Table 1 of general MP counts by gender. Because some MPs served across adjacent periods, these individuals are counted in both, and the overall number of MPs provided in Table 3 is larger than the actual number of MPs in the House of Commons provided in Table 1. Additionally, the counts cannot be compared directly across different periods, since the length of the periods is not uniform. The first one involves four decades, and the other two, three decades. However, what can be compared directly are the counts of male and female MPs in a single period, as well as the ratios of male to female MPs across periods.

The fact that the gender disparity decreases over time, as evidenced by the decline in the ratio of male to female values across the three periods in Table 3, may have a direct impact on the results reported in Section 4.2. At this stage, it is not clear whether the changes observed are due to an actual diachronic process, or if they appear as a result of increasing numbers of female MPs, who may exhibit different patterns of participation in political discourse; or both of these factors may play a role.

Table 3
Number of MPs taking part in debates in the House of Commons in different periods of the 20th century

Period	Male MPs	Female MPs	Ratio
1900-1939	2,925	26	113:1
1940-1969	1,739	62	28:1
1970-1999	1,608	177	9:1

Fig. 6 directly compares the mean word token counts for male and femaleMPs during their first 20 years of service. This shorter time span is presented due to the limited number of female MPs. Beyond the 20th year, the mean measurements for this group would be based on samples smaller than 30 (the 20th year in Parliament was served by 31 women). A visual inspection of the line plots suggests notable differences in the patterns of male and female MPs. Women appear to become more active earlier in their careers, surpassing an average of 14,000 word tokens as early as their third year of service, whereas men reach this level only after eight years. However, this conclusion is not supported by a more in-depth analysis, as discussed below.
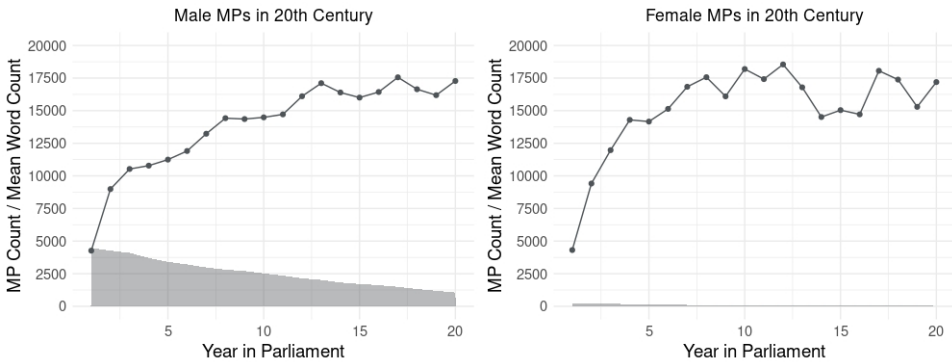

Fig. 6. MP Count (Area Graphs) and Mean Word Count (Line Graphs) per Year of Service in the House of Commons – Male vs. Female MPs

First, it is crucial to establish whether the diachronic changes in parliamentary debate involvement occur independently for male and female MPs. The line plots in Fig. 7 show the patterns across the three periods for the former group. They look very similar to the plots in Fig. 5 for all MPs. There is a more dynamic increase in mean word token counts in the middle plot than in the left one during the initial years, and an even more rapid increase in the right plot compared to the others. One may also observe a decrease in mean word counts after a peak has been reached in the period 1970–1999, even though the plot encompasses only the first 20 years in Parliament.

For women, such changes are shown in Figure 8, but only for two periods instead of three. The first four decades include too few female MPs to obtain any reliable measurements, and the plots in Fig. 8 are also based on relatively small samples and should be interpreted with caution. For the period 1940–1969, the samples drop below 30 by the 11th year (26 MPs), and the mean word count for the 20th year is based only on data from 13 MPs. The measurements for the period 1970–1999 are based on slightly larger samples, but they also fall below 30 MPs for the last four years shown in the second plot (27 MPs for the 17th year, 24 MPs for the 18th year, 22 MPs for the 19th year, and 18 MPs for the 20th year). Nevertheless, the patterns discernible in the plots

are consistent with the trends for male MPs. Increases in word counts during the initial years of parliamentary activity are greater for the period 1970–1999 than for 1940–1969.
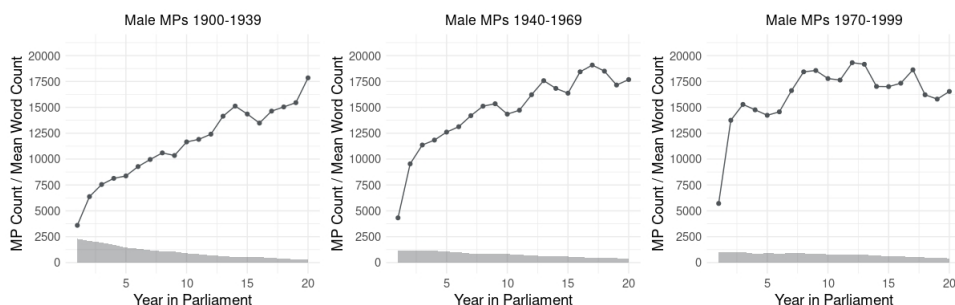

Fig. 7. MP Count (Area Graphs) and Mean Word Count (Line Graphs) per Year of Service in the House of Commons – Male MPs in three different periods
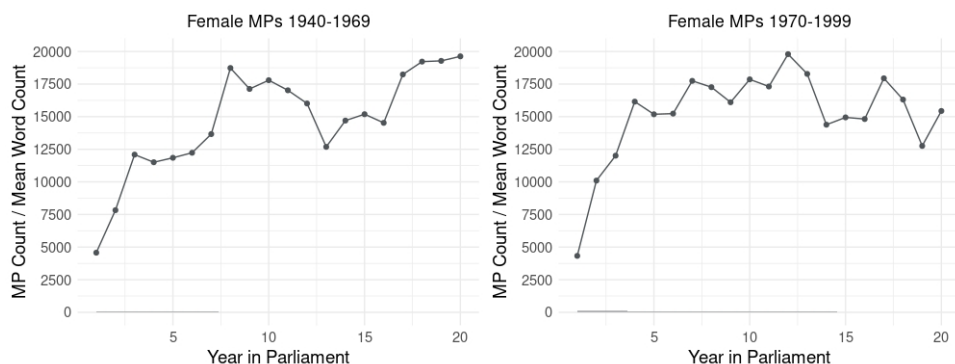

Fig. 8. MP Count (Area Graphs) and Mean Word Count (Line Graphs) per Year of Service in the House of Commons – Female MPs in two different periods

These results confirm that diachronic changes occur regardless of MPs' gender. However, they do not entirely rule out the possibility that gender may still influence the dynamics of word count changes across different periods. To assess this possibility, male and female patterns must be directly compared within a single period in which the proportion of men and women remains relatively stable and does not change as much as in the entire 20th century. This is done in Fig. 9, which compares line plots for male and female MPs during the period 1970–1999. A visual inspection of these plots does not reveal any substantial differences in the dynamics of word count increases. In the first three years of service, the average word counts are higher for men, reaching 15,000. However, in the fourth year, women's word counts exceed those of men, surpassing 16,000. For subsequent years, the trends for both genders are similar. Overall,

no clear differences emerge between the two line plots, suggesting that gender does not play a crucial role in shaping the diachronic shifts in patterns observed in Section 4.2.
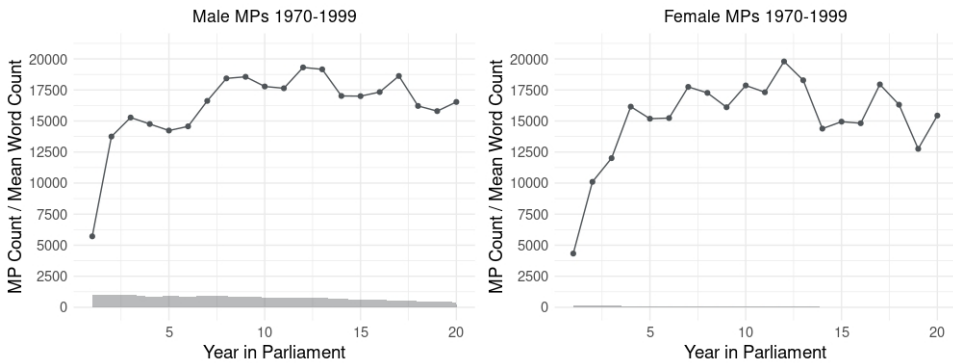

Fig. 9. MP Count (Area Graphs) and Mean Word Count (Line Graphs) per Year of Service in the House of Commons – Male vs. Female MPs

5. CONCLUSIONS

The results obtained in the present study reveal a complex picture of the relationship between duration of service in the House of Commons and the amount of verbal output. The analysis on the whole 20th century discussed in Section 4.1 shows a sharp increase in average word counts for the first three years in service, followed by a less steep rise up to the 13th year. After that point, the values mostly stabilize around 17,500 word tokens, and even for very long tenures they do not drop below 15,000 word tokens.

However, this general analysis misses important aspects which become evident when the data are scrutinized separately for different periods within the 20th century in Section 4.2. This additional diachronic investigation reveals consistent changes in the way MPs participate in parliamentary debates. For the first four decades, a steady, mostly linear increase throughout the 35 years in service is observed. The trend changes for the period 1940–1969, where the rise in mean word counts for initial years in Parliament is much faster, but later slows down, and after the 25th year a small but noticeable decrease in word counts is detectable. This shift in the dynamics of changes continues in the last three decades of the 20th century. The initial rise here is even steeper, and also the decline in values for longer tenures is more distinct and starts earlier.

An additional analysis of possible effects of MPs' gender does not reveal any consistent patterns. Even though there is a gradual increase in the number of female MPs throughout the 20th century, this trend seems not to be responsible for diachronic changes observed in Section 4.2. A direct comparison of men and women for the period 1970–1999 shows very similar patterns in the way male and female MPs engage in political debates in terms of the number of words uttered per year in service.

All these findings shed new light on the theoretical assumptions about the effects of work tenure on productivity outlined in Section 2.2. Assuming that the amount of verbal output is one of the possible metrics of work engagement for the political class, the results for the first four decades of the 20th century are surprising and show a pattern not predicted in any of the discussed proposals. The strong positive linear relationship between service duration and the average word count is a trajectory not encountered in studies on contemporary tenure-productivity interactions. However, in later periods, the trend shifts towards the inverted U-shaped relationship proposed in most publications, but it must be stressed that this pattern develops gradually.

In future studies, this diachronic shift in the dynamics of work productivity could be investigated using additional metrics beyond word counts, such as sentence counts, utterance counts, and speech counts. It would also be useful to consider factors such as the political orientation of individual MPs and whether they belonged to the ruling or opposition party. Moreover, parliamentary records from countries other than Great Britain, and in languages other than English, could also be investigated with respect to work productivity dynamics. Finally, such changes could also be examined diachronically in environments other than political discourse, as the academic discussion on the effects of work tenure on work productivity lacks such historical analyses.

REFERENCES

Abercrombie, G., & Batista-Navarro, R.T. 2018. "A Sentiment-Labelled Corpus of Hansard Parliamentary Debate Speeches". In: *International Language Resource and Evaluation Conference 2018: ParlaCLARIN Workshop*, Clarin, 43–47.

Abercrombie, G., & Batista-Navarro, R.T. 2019. "Semantic Change in the Language of UK Parliamentary Debates". In: *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, 210–15.

Alexander, M. 2023. "Speech in the British Hansard". In: *Exploring Language and Society with Big Data: Parliamentary Discourse across Time and Space*, eds. M. Korhonen, H. Kotze, & J. Tyrkkö. Amsterdam: John Benjamins Publishing Company, 17–53.

Alexander, M., & Davies, M. 2022. "Hansard Corpus (British Parliament) 1803–2005". *English Corpora*. https://www.english-corpora.org/hansard/ (access: Dec 10, 2025).

Alexander, M., & Struan, A. 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.

Allen, J., & De Grip, A. 2012. "Does Skill Obsolescence Increase the Risk of Employment Loss?". *Applied Economics*, 44(25): 3237–3245.

Allen, N.J., & Meyer, J.P. 1990. "The Measurement and Antecedents of Affective, Continuance and Normative Commitment to the Organization". *Journal of Occupational Psychology*, 63(1): 1–18.

Appleton, S.A. 2023. "Lexical Change in the Parliamentary Contributions of UK Foreign and Commonwealth Office Ministers, 1989–2015: A Corpus Linguistic Analysis". PhD Thesis. University of Birmingham.

Archer, D. 2017. "Mapping Hansard Impression Management Strategies through Time and Space". *Studia Neophilologica*, 89(sup1): 5–20.

Archer, D. 2017. 2018. "Negotiating Difference in Political Contexts: An Exploration of Hansard". *Language Sciences*, *68*: 22–41.

Awal, N.M., Jaludin, A., Abdul Rahman, A.N.C., & Abdullah, I.H. 2019. " 'Is Selangor in Deep Water?': A Corpus-Driven Account of Air/Water in the Malaysian Hansard Corpus (MHC)". *GEMA Online Journal of Language Studies*, *19*(2).

Bakar, R.A. 2013. "Understanding Factors Influencing Employee Engagement: A Study of the Financial Sector in Malaysia". PhD Thesis. RMIT University.

Batistuta, A.I.F.I., & Safitri, R. 2024. "Comparison of Employee Tenure on Employee Productivity". *Jurnal Ekonomi, Bisnis & Entrepreneurship*, *18*(2): 524–32.

Bryson, A., Forth, J., Gray, H., & Stokes, L. 2020. "Does Employing Older Workers Affect Workplace Performance?". *Industrial Relations: A Journal of Economy and Society*, *59*(4): 532–62.

Bussolo, M., J. Koettl, & E. Sinnott. 2015. *Golden Aging: Prospects for Healthy, Active, and Prosperous Aging in Europe and Central Asia*. World Bank Publications.

Carpuat, M. 2014. "Mixed Language and Code-Switching in the Canadian Hansard". In: *Proceedings of the First Workshop on Computational Approaches to Code Switching,* 107–15.

Che Abdul Rahman, A.N., I. Ho Abdullah, I.S. Zainuddin, & A. Jaludin. 2019. "The Comparisons of Ocr Tools: A Conversion Case in the Malaysian Hansard Corpus Development". *Malaysian Journal of Computing (MJoC)*, *4*(2): 335–48.

Coffman, C., A. Gopal, & G. Gonzalez-Molina. 2002. *Follow This Path: How the World's Greatest Organizations Drive Growth by Unleashing Human Potential*. New York: Warner Books.

Cohen, A. 1993. "Age and Tenure in Relation to Organizational Commitment: A Meta-Analysis". *Basic and Applied Social Psychology*, *14*(2): 143–59.

Coole, M., P. Rayson, & J. 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 the Second ParlaCLARIN Workshop*, 23–27.

Corbett, J.B. 2021. "Terminology and the Evolution of Linguistic Prejudice: The Conceptual Domain of "Irishness" in the Historical Thesaurus of English and the Hansard Corpus of British Parliamentary Speeches". *Tradterm*, *37*(2): 515–37.

De Grip, A. 2006. "Evaluating Human Capital Obsolescence". *Researchcentrum voor Onderwijs en Arbeidsmarkt, Faculteit der Economische Wetenschappen. ROA Working Papers* No. 2E

Demmen, J., L. Jeffries, & B. Walker. 2018. "Charting the Semantics of Labour Relations in House of Commons Debates Spanning Two Hundred Years: A Study of Parliamentary Language Using Corpus Linguistic Methods and Automated Semantic Tagging". In: *Discourse Approaches to Politics, Society and Culture*, eds. Michael Kranert and Geralinde Horan. Amsterdam: John Benjamins Publishing Company, 81–104.

Desjardins, R., & A.J. Warnke. 2012. *Ageing and Skills: A Review and Analysis of Skill Gain and Skill Loss Over the Lifespan and Over Time (Working Paper No. 72)*. OECD Education Working Papers.

Dongrey, R., & V. Rokade. 2019. "Implication of Tenure on Employee Engagement". *Open Access Journal of Humanities*, *2*(3): 34–37.

Gagliardi, N., E. Grinza, & F. Rycx. 2021. "Can You Teach an Old Dog New Tricks? New Evidence on the Impact of Tenure on Productivity". *IZA Discussion Papers* (No. 14432)

Giniger, S., A. Dispenzieri, & J. Eisenberg. 1983. "Age, Experience, and Performance on Speed and Skill Jobs in an Applied Setting". *Journal of applied psychology*, *68*(3): 469.

Hiltunen, T., J. Räikkönen, & J. Tyrkkö. 2020. "Investigating Colloquialization in the British Parliamentary Record in the Late 19th and Early 20th Century". *Language Sciences*, *79*: 101270.

Hiltunen, T., & T. Vartiainen. 2024. "A Corpus-Pragmatic Analysis of Linguistic Democratisation in the British Hansard: Comparing the Two Houses". *Journal of Historical Pragmatics*, *25*(2): 245–73.

Ingham, R. 2022. "The Cradle of Parliamentary Democracy?: Collocates of "Democracy" in Hansard 1803–1900". In: *Auto- und Heterostereotypie im Europa des 19. Jahrhunderts*, Linguistik in Empirie und Theorie/Empirical and Theoretical Linguistics, eds. A. Fábián, A. Owzar, & I. Trost. Berlin, Heidelberg: Springer Berlin Heidelberg, 191–206.

Jeffries, L., F. Stradling, A.V. Lünen, & H. Sanjurjo-González. 2023. "Hansard at Huddersfield: Streamlined Corpus Methods and Interactive Visualisations to Pursue Research Aims beyond Corpus Linguistics". In: *Studies in Corpus Linguistics*, eds. M. Korhonen, H. Kotze, & J. Tyrkkö. Amsterdam: John Benjamins Publishing Company, 89–117.

Jenkins, P.S. 2019. "Text Mining the British Hansard". In: *War and Happiness*, Cham: Springer International Publishing, 41–53.

Joanis, E. et al. 2020. "The Nunavut Hansard Inuktitut–English Parallel Corpus 3.0 with Preliminary Machine Translation Results". In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2562–72.

Kang, N. 2023. "On Consist of and Made up of: A Corpora-Based Analysis". *Theory and Practice in Language Studies*, *13*(4): 840–46.

Kotze, H., & B. Van Rooy. 2020. "Democratisation in the South African Parliamentary Hansard? A Study of Change in Modal Auxiliaries". *Language Sciences*, *79*: 101264.

Kruger, H., & A. Smith. 2018. "Colloquialization versus Densification in Australian English: A Multidimensional Analysis of the Australian Diachronic Hansard Corpus (ADHC)". *Australian Journal of Linguistics*, *38*(3): 293–328.

Labat, S., H. Kotze, & B. Szmrecsanyi. 2023. "Processing and Prescriptivism as Constraints on Language Variation and Change: Relative Clauses in British and Australian English Parliamentary Debates". In: *Exploring Language and Society with Big Data: Parliamentary Discourse across Time and Space*, eds. Minna Korhonen, Haidee Kotze, & Jukka Tyrkkö. Amsterdam: John Benjamins Publishing Company, 250–76.

Leibel, M.S. 2024. "Such Cold, Dispassionate Terms Fail Us': Reading Hansard as an "Archive of Feelings". *Parliamentary Affairs*, *77*(2): 328–49.

Leonard, R. 2007. "Beautiful Soup Documentation". https://www.crummy.com/software/BeautifulSoup/bs4/doc/

McGill, E., & H. Saggion. 2023. "BSL-Hansard: A Parallel, Multimodal Corpus of English and Interpreted British Sign Language Data from Parliamentary Proceedings". In: *Proceedings of the Second International Workshop on Automatic Translation for Signed and Spoken Languages*, 38–43.

Meyer, J.P. et al. 1989. "Organizational Commitment and Job Performance: It's the Nature of the Commitment That Counts". *Journal of applied Psychology*, *74*(1): 152.

Micher, J. 2018. "Using the Nunavut Hansard Data for Experiments in Morphological Analysis and Machine Translation". In: *Proceedings of the Workshop on Computational Modeling of Polysynthetic Languages*, 65–72.

Nanni, F., S. Menini, S. Tonelli, & S.P. Ponzetto. 2019. "Semantifying the UK Hansard (1918–2018)". In: *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, IEEE, 412–13.

Ng, T.W.H., & D.C. Feldman. 2010. "Organizational Tenure and Job Performance". *Journal of Management*, *36*(5): 1220–50.

Onyimadu, O. et al. 2013. "Entity-Based Semantic Search on Conversational Transcripts Semantic". In: *Semantic Technology*, eds. H. Takeda, Y. Qu, R. Mizoguchi, & Y. Kitamura. Berlin, Heidelberg: Springer Berlin Heidelberg, 344–49.

Picchio, M. 2021. "Is Training Effective for Older Workers?". *IZA World of Labor,* 121.

Porro, G. 2022. "Language and Gender: A Corpus-Assisted Discourse Analysis of British Politicians' Language Use". Thesis. University of Padova.

Rahman, C.A., & A. Nadiah. 2024. "The Portrayal of Women: A Corpus Analysis of the Representation around the Word'Wanita'in the Malaysian Hansard Corpus". *GEMA Online Journal of Language Studies*, *24*(4).

Rice, J.K. 2009. "The Role of Tenure as a Moderator to Work Engagement and Job Satisfaction". Thesis. San José State University.

Rickman, P., & J. Rudanko. 2018. "Null Objects and Sentential Complements, with Evidence from the Corpus of Historical American English and Hansard". In: *Corpus-Based Studies on Non-Finite Complements in Recent English*, Cham: Springer International Publishing, 75–96.

Shaw, K., & E.P. Lazear. 2008. "Tenure and Output". *Labour Economics*, *15*(4): 704–23.

Smith, A., & M. Korhonen. 2022. "Parliamentary Hansard Records and Epicentral Influence in Australia, New Zealand and Papua New Guinea". *World Englishes*, *41*(3): 475–90.

Soriano-Jiménez, C. 2024. "Language Change in British Parliamentary Discourse: A Corpus-Based Study of Power and Authority Markers, 1930–2005". *Atlantis. Journal of the Spanish Association for Anglo-American Studies*: 49–68.

Swaminathan, J., & A. Ananth. 2009. "Impact of Demographic Factors on Employee Engagement: A Study with Reference to Vasan Publications Private Limited, Chennai". *MPRA Paper* No. 39768.

The pandas development team. 2023. "Pandas". https://pandas.pydata.org/

Uppal, N. 2017. "Uncovering Curvilinearity in the Organizational Tenure-Job Performance Relationship: A Moderated Mediation Model of Continuance Commitment and Motivational Job Characteristics". *Personnel Review*, *46*(8): 1552–70.

Van Loo, J., A. De Grip, & M. De Steur. 2001. "Skills Obsolescence: Causes and Cures". *International Journal of Manpower*, *22*(1/2): 121–38.

Vardanega, M., & L. Smith. 2016. "Analysis of Sentiment Direction Based on Two Centuries of the Hansard Debate Archive". *Stirling: University of Stirling unpublished manuscript.*

Wpływ długości pełnienia mandatu na udział w debatach parlamentarnych: analiza korpusu Hansard

Słowa kluczowe: analiza korpusowa, korpus Hansard, debaty parlamentarne, zaangażowanie zawodowe, produktywność zawodowa.

STRESZCZENIE

Celem tego badania jest analiza związku między długością sprawowania mandatu przez posłów w Izbie Gmin a ich udziałem w debatach parlamentarnych w XX wieku. W tym celu wykorzystano wersję korpusu Hansard dostępną pod adresem „https://api.parliament.uk/historic-hansard/" (access: Dec 10, 2025). Za pomocą skryptu w Pythonie pobrano pełne transkrypcje debat i opatrzono je adnotacjami dotyczącymi m.in. nazwisk posłów, ich płci oraz dat wystąpień. Po dalszym przetworzeniu danych obliczono i przeanalizowano średnią liczbę wypowiedzianych tokenów leksykalnych na rok pełnienia mandatu parlamentarnego jako wskaźnik aktywności w debatach. Ogólne wyniki wskazują na wzrost liczby wypowiadanych słów w pierwszych latach pracy, po którym następuje stabilizacja wartości średnich w kolejnych latach. Dodatkowa analiza diachroniczna ujawnia jednak istotne zmiany w tym trendzie. W pierwszych czterech dekadach XX wieku zależność ta ma w dużej mierze charakter liniowy. W późniejszych okresach stopniowo ewoluuje jednak w trajektorię o odwróconym kształcie litery U – maksymalna liczba wypowiadanych słów przypada na wcześniejszy etap kariery parlamentarnej, po czym następuje spadek wśród posłów z dłuższym stażem. Dodatkowo zbadano możliwy wpływ różnic pod względem płci, jednak wyniki sugerują, że obserwowane zmiany diachroniczne nie wynikają ze zmieniającego się stosunku liczby posłów do posłanek w ciągu stulecia. Ostatnie trzy dekady XX wieku potwierdzają założenia obecne w literaturze, jednak trendy obserwowane we wcześniejszych okresach nie są sugerowane w badaniach nad współczesnymi zależnościami między długością stażu a produktywnością. Sugeruje to potrzebę dalszych badań nad ewolucją tych procesów w perspektywie diachronicznej.