Nurturing vs Pragmatic Language in Discussions of Maternity and Paternity in British Parliamentary Discourse

Barbara Helene Schmehl
University of the Basque Country
bschmehl001@ikasle.ehu.eus

Abstract

This paper presents the findings of a preliminary investigation into the gendered language associated with the tokens "maternity" and "paternity" as seen in the British Parliament in the ParlaMint Corpus. Utilizing the NLP tools SpaCy, the study examines main verbs used and collocations by categorizing them as either "nurturing" or "pragmatic". A two-sided z-test is used to determine statistical significance of observed differences. The findings reveal that while "paternity" is associated with a higher frequency of pragmatic verbs, "maternity" does not show a statistically significant tendency to be associated with nurturing verbs, suggesting a complex interplay in the legislative discourse that challenges traditional notions of gender roles.

1 Introduction

This paper reports the initial findings of an exploratory delve into the uses of the tokens "paternity and "maternity" within the Great Britain portion of the ParlaMint Corpus (Kuzman et al., 2024) within the context of a graduate-level course "Language Technologies for the Digital Humanities". It is therefore, above all, an opportunity to practice skills and methods acquired in the course.

Notions of maternity have historically differed significantly from those of paternity. The discourse surrounding paternity and maternity leave policies has highlighted distinct perspectives and societal expectations tied to each. Moss and Deven (2015) in their review of paternity leave policies, note a prevailing focus on legal rights and responsibilities associated with fatherhood. This approach often overlooks the emotional and caregiving dimensions of fatherhood, presenting a less nurturing view of paternity when compared with maternity. Langan and Olsen (2001) further support this notion through a comparative analysis of paternity leave across various countries, revealing that such policies are generally less developed and shorter than maternity leave. This discrepancy is interpreted as underscoring a societal undervaluation of the caregiving role within the context of fatherhood.

Gatrell (2005) explores societal expectations by contrasting the emotional and physical caregiving aspects traditionally associated with maternity with the legal and economic perceptions of paternity. These views perpetuate gendered distinctions, positioning maternity as inherently nurturing while framing paternity through a more pragmatic lens. Similarly, Cameron (2005) examines the language and discourse surrounding gendered terms like "maternity" and "paternity," offering insights into how these terms reflect and reinforce societal attitudes towards gender roles. The use of specific collocations and linguistic patterns further illuminates how deeply ingrained these perceptions are within cultural narratives.

The present investigation takes an exploratory approach to the tokens "maternity" and "paternity" as they are used in British parliamentary discourse. Broadly, we would expect our results to support the findings of previous research - namely, that fatherhood and "paternity" is more closely associated with pragmatic considerations, while "maternity" is more closely associated with nurturing and caregiving. Specifically, we can pose the following as as research questions:

1. Are adjectival collocations with "maternity" more likely to be nurturing in nature than those associated with "paternity"? Likewise, are adjectival collocations with "paternity" more likely to be pragmatic?

2. Are the main verbs associated with the token "paternity" more likely to be pragmatic in nature than the main verbs associated with "maternity"? Likewise, are the main verbs associated with the token "maternity" more likely to

be nurturing?

In order to determine this, the Python NLP library SpaCy is used to preprocess and analyze the data on the linguistic level. Adjectival collocations as well as main verbs associated withour target tokens were assigned to lists for further analysis. The adjectives and verbs were categorized into nurturing or non-nurturing by the author without looking at which token the adjective or verb had been associated with originally. Overall sums for nurturing or non-nurturing words were calculated for both "maternity" and "paternity". Lastly, two-tailed z-tests were performed to determine whether any obseved differences were statistically significant.

2 Methods

Data for this investigation comes from the Great Britain portion of the ParlaMint Corpus, an open-source set of parliamentary corpora from 29 countries, with each individual corpus comprising between 9 and 126 million words (Kuzman et al., 2024). SpaCy (Honnibal and Montani, 2017), a commonly-used NLP library in Python, was used to analyze the corpus of Great British Parliament data. The analysis focused on two main areas: main verbs and adjectival collocations associated with the tokens "maternity" and "paternity".

2.1 Data Collection and Preprocessing

The dataset was comprised of XML documents from the ParlaMint-GB.TEI collection, which were processed to extract sentences containing the tokens "maternity" and "paternity." This involved reading each XML file, splitting the content into individual sentences, and filtering sentences that contained either of the target tokens.

2.2 Linguistic Analysis with SpaCy

The linguistic analysis was conducted using SpaCy's English language model, which provided part-of-speech tagging and dependency parsing. Only sentences containing "maternity" and "paternity" were processed to extract relevant linguistic features.

SpaCy's dependency parser was used to identify the grammatical relationships between words in each sentence. This allowed for the extraction of verbs associated with "maternity" and "paternity" by identifying verbs that had these nouns as their subjects or were closely linked to them in the sentence structure. Initially, the code misidentified "leave" as a verb due to its frequent occurrence within the compound noun phrases "maternity leave" and "paternity leave." To address this, the code was modified to better utilize SpaCy's syntactic dependency parsing capabilities. By analyzing the dependency tree, the code was adjusted to identify "maternity leave" and "paternity leave" as compound noun phrases and accurately extract verbs that govern these phrases. This approach enabled a more precise analysis of the verbs related to "maternity" and "paternity". Adjectival modifiers (amod) associated with the nouns "maternity" and "paternity" were extracted to identify common collocations.

2.3 Data Analysis and Visualization

The extracted verbs and collocations were analyzed using the Python libraries Pandas and Matplotlib. Frequency counts were calculated for each category, and bar graphs were generated to visually represent the distribution of collocations and associated verbs for "maternity" and "paternity."

Comprehensive lists of collocations and verbs were categorized into "nurturing" and "pragmatic" categories in a separate Excel sheet, where duplicates were removed and the data was manually inspected for non-relevant entries (such as XML tags). Both "paternity" and "maternity" entries were put in a singular list and randomized, in an attempt to mitigate bias. Categorization was done once by the author alone. Collocations and verbs were categorized separately (i.e., on different Excel sheets). Categorization was not obligatory; in other words, it was possible for a word to fall into neither category and remain unlabelled. For the purposes of this investigation, no single word could be categorized as both "nurturing" and "pragmatic". Words were categorized as either "pragmatic", "nurturing", or left un-categorized. In categorizing both verbs and collocations, the author followed Cameron (2005) as well as her own native-speaker intuition as to which verbs and adjectives were "caregiving" or "pragmatic". See 4 for a complete list of the categorizations.

3 Results

3.1 Associated Verbs

2,953 tokens were identified as the main verb of a sentence or clause containing the token "maternity" while 370 tokens were identified as the main

verb of a sentence or clause containing the token "paternity. The following figures identify the top 20 most frequent verbs for both tokens:
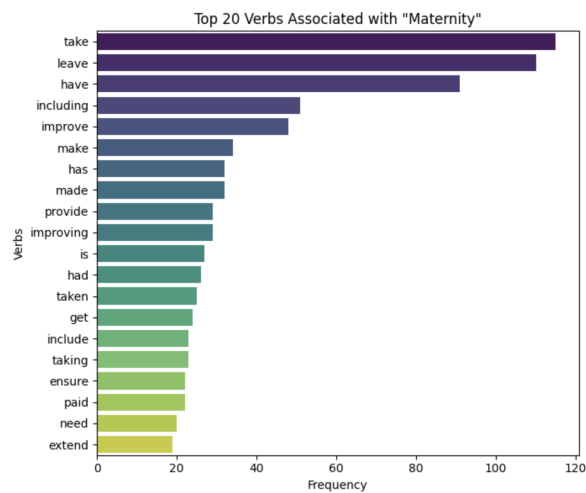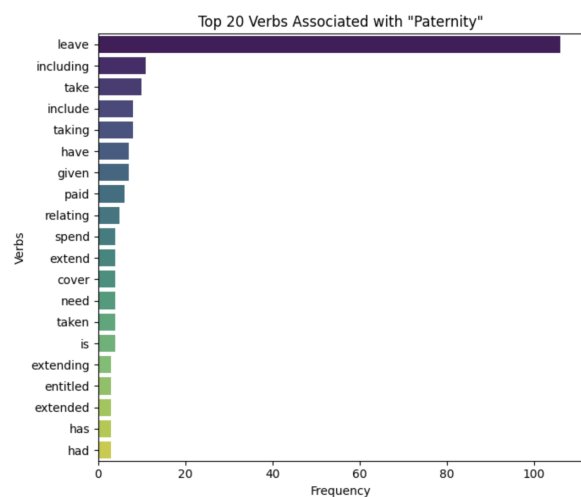

Figure 1: Top 20 Maternity Verbs

Figure 2: Top 20 Paternity Verbs

The analysis of the top 20 verbs associated with "maternity" and "paternity" highlights notable patterns in legislative language usage, reflecting the differing extents of discourse on each topic. For "maternity", the most frequently occurring verbs such as "take" (115), "leave", (110), and "have" (91) suggest active engagement with maternity issues, focusing on entitlements and actions. The presence of verbs like "improve" also points to discussions aimed at enhancing maternity-related policies.

Conversely, while the overall frequency of verbs associated with "paternity" is lower, this is proportionate to the fewer instances of "paternity" being mentioned in the corpus, not necessarily indicating less varied discussion. The verbs "leave", "include", and "take" lead the list, reflecting similar thematic concerns with entitlements and provisions as seen in maternity discussions. This consistency in verb usage across both categories suggests that while paternity is discussed less frequently, the nature of the discussions mirrors those of maternity.

3.2 Collocations

159 tokens were identified as adjectival collocations (amod) of the token "maternity" while 36 tokens were identified as as adjectival collocations (amod) of the token "paternity. The following figures identify the top most frequent collocations for both tokens:
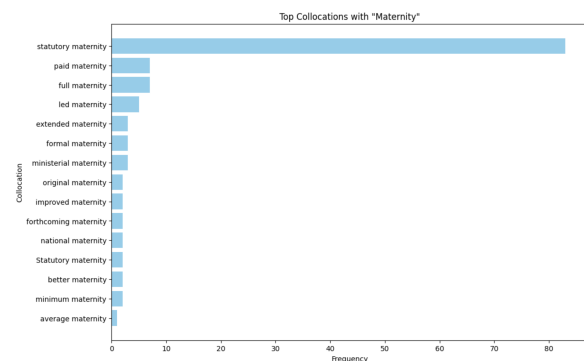

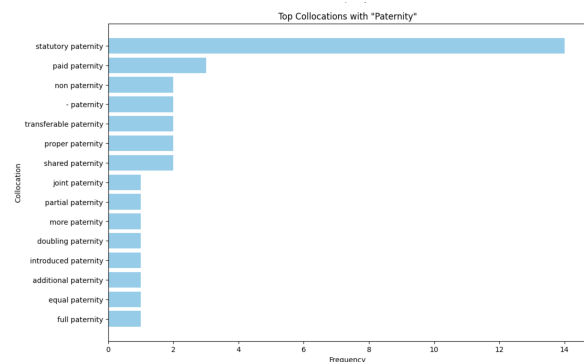Figure 3: Top Collocations with "Maternity"

Figure 4: Top Collocations with "Paternity"

For "maternity," the most frequently occurring collocation is "statutory maternity," appearing 83 times. This indicates a prominent focus on statutory aspects of maternity benefits or policies. Other notable collocations include "paid maternity" and "full maternity," both occurring 7 times.

As for "paternity", "statutory paternity" is the most frequent collocation, appearing 14 times, followed by "paid paternity" and "formal paternity," which occur 3 times each.

3.3 Categorization as Nurturing or Pragmatic

A two-sided z-test was conducted to assess the statistical significance of differences in language use between "maternity" and "paternity." The first hypothesis tested whether verbs and collocations associated with "maternity" are statistically more likely to be categorized as "nurturing" compared to those associated with "paternity." Similarly, the second hypothesis examined whether verbs and collocations associated with "paternity" are more likely to be categorized as "pragmatic" than those linked to "maternity." The following formula was used to determine significance for these hypotheses:

$$(H_0): \ p_1 = p_2$$
$$(H_A): \ p_1 \neq p_2$$
$$(\hat{p}): \ \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$
$$(\text{SE}): \ \text{SE} = \sqrt{\hat{p}\left(1 - \hat{p}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$Z = \frac{p_1 - p_2}{\text{SE}}$$

The following table shows the results of the results of the categorizations of verbs.

Table 1: Verb Categorization

	Counts		Percentages	
	Nur	Prag	Nur	Prag
Maternity	309	611	10.46%	20.69%
Paternity	39	153	10.54%	41.35%

When investigating whether "nurturing" verbs were more associated with "maternity" than "paternity", results yielded a z-score of -0.0454 and a p-value of 0.5181, indicating that verbs categorized as "nurturing" are not statistically more likely to be associated with the token "maternity" than with "paternity". This indeed suggests that the "nurturing" verbs may be more closely associated with "paternity".

When investigating whether "pragmatic" verbs were more associated with "paternity" than "maternity", results yielded a z-score of 8.9034 and a p-value of 2.71×10^{-19}, which strongly suggests that verbs associated with "paternity" are more likely to have been categorized as "pragmatic" when compared with verbs associated with "maternity".

Table 2: Collocation Categorization

	Counts		Percentages	
	Nur	Prag	Nur	Prag
Maternity	2	115	1.26%	72.33%
Paternity	0	23	0.00%	63.89%

When investigating whether "nurturing" collocations were more likely to have been associated with "maternity", results yielded a Z-statistic of 0.6713 and a p-value of 0.2515. We therefore cannot reject the null hypothesis and cannot conclude any statistically significant differences in the likelihood of a "maternity" collocation being categorized as "nurturing" vs. a "paternity" collocation.

When investigating whether "pragmatic" collocations were more associated with "paternity" than "maternity", results yielded a z-score of -1.0052 and a p-score of 0.8426. There is no statistical significance. In fact, the negative z-score indicates that there may in fact be a greater association between "maternity" and collocations that have been categorized as "pragmatic".

4 Conclusions

The present investigation into the linguistic categorizations of the tokens "maternity" and "paternity" within the British Parliamentary Corpus has yielded insightful and surprising results. The analyses align with some aspects of previous research but also provide new directions for understanding the nuances of gendered language in legislative discourse. We did not find support for the hypothesis that "maternity" is more likely to be associated with nurturing verbs than "paternity." In fact, the two-sided z-test revealed no significant difference in the use of nurturing verbs between "maternity" and "paternity," with a z-score of -0.0454 and a p-value of 0.5181. This suggests a more complex interplay of nurturing connotations with both terms than traditionally perceived, potentially challenging the notion that maternity is inherently more nurturing in legislative language.

The hypothesis that "paternity" is more likely to be associated with pragmatic verbs when compared with "maternity" was strongly supported by our results. The z-test yielded a significant z-score of 8.9034 and a p-value of 2.71×10^{-19}, indicating a distinct emphasis on the pragmatic aspects of paternity over maternity. This finding corrobo-

rates with Moss and Deven's (2015) observations of paternity being framed within a more legalistic and economic context, as well as Langan and Olsen's (2001) findings on the underdevelopment of paternity policies.

For collocations, however, neither nurturing nor pragmatic categories showed significant differences between "maternity" and "paternity." This suggests that while verbs may carry distinct gendered connotations, the adjectival modifiers surrounding these terms do not exhibit a strong bias towards nurturing or pragmatic aspects in parliamentary discourse. The lack of significant findings in collocations may also point to a more standardized or formalized use of language in legislative texts that dilutes personal biases or stereotypes.

Given these outcomes, further research using a more comprehensive dataset from the ParlaMint Corpus is recommended. Expanding the analysis to include a broader array of linguistic markers and a larger corpus could provide a more definitive picture of how maternity and paternity are framed in political discourse.

The present investigation highlights the capabilities of NLP tools in uncovering underlying patterns in political language and underscores the need for ongoing scrutiny of how perceptions of gender are perpetuated through seemingly neutral institutions.

Limitations

In this section, I would like to highlight the numerous potential limitations of this preliminary investigation which may affect any interpretation of the findings.

The categorization process used in this study was highly subjective. The differentiation between "nurturing" and "pragmatic" verbs and collocations was determined by the author based on existing literature and personal intuition. Despite the randomization utilized, this approach introduces a significant degree of subjectivity, which could lead to biases in how tokens are categorized. Future research could benefit from a more standardized or automated categorization method, potentially incorporating machine learning techniques that could learn from a broader set of training data to reduce personal bias.

There was an issue in the initial coding process where "leave", commonly occurring in the noun phrase "maternity leave" and "paternity leave", was misidentified as a verb. This misclassification likely skewed the verb frequency data, suggesting that improvements in the code's ability to discern between usage as a noun and a verb are necessary. Enhancing the NLP parsing algorithms or refining the syntactic filters used in the analysis could help mitigate such errors.

The overall number of collocations associated with both "maternity" and "paternity" was relatively small, totaling 159 for "maternity" and 36 for "paternity". This discrepancy raises questions about the comprehensiveness of the data extraction process. It is possible that not all relevant instances were captured, thus missing potential collocations.

These limitations highlight the need for a more robust analytical framework and suggest that the findings should be interpreted with caution. Further investigation with improved data handling and categorization methods could provide a more accurate and comprehensive understanding of the linguistic patterns associated with "maternity" and "paternity" in the corpus studied.

References

Deborah Cameron. 2005. Language, gender, and sexuality: Current issues and new directions. *Applied Linguistics*, 26(4):482–502.

Caroline J. Gatrell. 2005. *Hard Labour: The Sociology of Parenthood.* PoliPoint.

Matthew Honnibal and Ines Montani. 2017. Spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. Software available from https://spacy.io. Version 2.x.

Taja Kuzman et al. 2024. Linguistically annotated multilingual comparable corpora of parliamentary debates in english. http://hdl.handle.net/11356/1910. ParlaMint-en.ana.

Mary Langan and Wendy Olsen. 2001. Fathers' leave: A cross-national comparison. *Journal of European Social Policy*, 11(2):159–176.

Peter Moss and Fred Deven. 2015. Leave policies and research on fathers' leave: What have we learned? *Comparative Social Research*, 31:33–68.

Appendix

COLLOCATION	nurturing (y/n)	pragmatic (y/n)
ministerial	NaN	y
formal	NaN	y
statutory	NaN	y
consistent	NaN	y
minimum	NaN	y
increased	NaN	y
extended	NaN	y
disciplinary	NaN	y
additional	NaN	y
formalised	NaN	y
paid	NaN	y
poorest	NaN	y
poor	NaN	y
paediatric	NaN	y
national	NaN	y
average	NaN	y
unlawful	NaN	y
proper	NaN	y
transferable	NaN	y
partial	NaN	y
happy	y	NaN
great	y	NaN
busy	NaN	NaN
enhanced	NaN	NaN
full	NaN	NaN
specific	NaN	NaN
modern	NaN	NaN
better	NaN	NaN
extraordinary	NaN	NaN
more	NaN	NaN
fairer	NaN	NaN
new	NaN	NaN
led	NaN	NaN
individual	NaN	NaN
longer	NaN	NaN
appropriate	NaN	NaN
real	NaN	NaN
multi	NaN	NaN
central	NaN	NaN
ordinary	NaN	NaN
own	NaN	NaN
black	NaN	NaN
forthcoming	NaN	NaN
low	NaN	NaN
existing	NaN	NaN
stronger	NaN	NaN
improved	NaN	NaN
possible	NaN	NaN
original	NaN	NaN
shared	NaN	NaN
equal	NaN	NaN
non	NaN	NaN
stand	NaN	NaN
doubling	NaN	NaN

VERB	nurturing (y/n)	pragmatic(y/n)
matters	y	NaN
attached	y	NaN
include	y	NaN
including	y	NaN
accept	y	NaN
offers	y	NaN
pledged	y	NaN
supported	y	NaN
visit	y	NaN
improving	y	NaN
ensure	y	NaN
drawn	y	NaN
touch	y	NaN
enjoy	y	NaN
enhance	y	NaN
played	y	NaN
invited	y	NaN
welcome	y	NaN
protecting	y	NaN
strengthening	y	NaN
given	y	NaN
hosting	y	NaN
supporting	y	NaN
agree	y	NaN
congratulate	y	NaN
inspired	y	NaN
give	y	NaN
creating	y	NaN
focused	y	NaN
believe	y	NaN
believes	y	NaN
agrees	y	NaN
agreed	y	NaN
interact	y	NaN
benefited	y	NaN
cared	y	NaN
learn	y	NaN
carry	y	NaN
stress	y	NaN
invested	y	NaN
undertaking	y	NaN
fought	y	NaN
share	y	NaN
improves	y	NaN
helped	y	NaN
embracing	y	NaN
achieve	y	NaN
thank	y	NaN
appreciate	y	NaN
promote	y	NaN
value	y	NaN
attach	y	NaN
involved	y	NaN
sharing	y	NaN
protected	y	NaN