

## Chapter 5: Decoding Grammar to Uncover the Language of Agency

Language change does not only happen at the level of individual words or phrases. It also takes place in the structure of sentences. By examining patterns in sentence grammar, analysts can uncover deeper shifts in how speakers express ideas and reflect social norms over time (Zeldes 2015). In this chapter we show that the power dynamics imagined by a culture can be intuited from grammatical relationships within sentences. For example, subject-object relations—that is, relations that specify who performs actions (subjects) and who is acted upon (objects)—are not encoded at the level of individual words, but rather emerge from sentence structure, specifically from the syntactic relations between words (Johnson 1987). Digital humanists and social scientists have long used these methods to understand political agency and storytelling (Franzosi 2004, Tangherlini et al 2020, Shahsavari et al. 2020). Their work shows us that subject-object relations are especially revealing because they encode underlying cultural assumptions about agency, identity, and power dynamics as expressed by a community through language.

In this chapter, we focus on how gendered power relations are constructed and reflected through syntactic structures, particularly in the ways language encodes positionality, that is, the context that shapes a person’s access to—or lack of—power in a society. We are not the first to recognize that grammar, and syntax in particular, plays a foundational role in shaping the expression of gender, and can serve to empower, constrain, naturalize, or delegitimize people. The basis of Feminist theory is in part built upon the idea that the structures of language both reveal and reinforce gendered expressions of power, making syntax a critical site for analyzing power dynamics as produced through discourse.

Foundational feminist theorists including Hélène Cixous, Luce Irigaray, and Monique Wittig each approach the politics of syntax from distinct angles. Cixous critiques conventional syntactic forms—such as subject-verb-object constructions—as mirroring the rigid hierarchies of patriarchal political order. She calls for forms of writing that defy linear structure and fixed meaning, arguing that fluid, disruptive syntax resists systems in which agency is centralized in a dominant (often male) subject and others are relegated to passive roles. Luce Irigaray, in contrast, interrogates the deeper structural logic of discourse itself. She argues that fundamentally masculine forms of discourse structure what we might call the “syntax” of social and political systems – even structuring thinking itself. These systems reproduce the male subject as the universal norm and language and politics will remain systems of masculine self-reproduction. Monique Wittig argues that language doesn’t just reflect gendered hierarchies—it actively produces them. For Wittig, the very category of “woman” is a political fiction sustained through language and syntax that encode heterosexual norms. Rather than seek a “feminine” alternative, Wittig calls for the destruction of the linguistic

structures that make gender seem natural. Language, she insists, must be politically rewritten to dismantle the binary logic on which patriarchy depends.

These theories offer a powerful lens through which to understand the stakes of grammatical part-of-speech analysis in historical texts. The Nineteenth-century Hansard represents an enormous opportunity in this regard; as a collection of political debates between men, Hansard has been minimally engaged as a resource for understanding how gender norms were perpetuated. Yet as feminist historians have long understood, parliament played a massive role in structuring women's experience, from the Laws of Coverture to the Matrimonial Clauses Act (1857), the Married Women's Property Acts (1870-1882), the Contagious Disease Acts (1864-69) and women's right to vote in the UK (1918-28) (Rackley and Auchmuty 2019, Kent 1999, Griffin 2012). Feminist historians have shown how, in the debates around the Contagious Diseases Acts, women were portrayed as both powerless and degraded, yet also as potentially threatening (Walkowitz, 1970, p. 4).

Members of parliament in the nineteenth century regularly described a world where male subjects acted and female objects were acted upon. Cixous's critique of rigid syntactic forms specifically targets the kinds of grammatical regularities we observe in mid-nineteenth-century Parliamentary discourse. Her call for a syntax that resists fixed meaning aligns with our aim to expose how stable grammatical roles naturalize uneven distributions of agency. Similarly, Irigaray argues that speech acts that assert a masculine subject as the universal norm represent a "political syntax" that reproduces political violence. With these insights, we can see that Hansard reveals an erasure of women as political actors. Wittig's insistence that language produces the fiction of gender further illuminates how syntactic roles in Hansard reinforce the category of "woman" as passive, pathological, and other.

Using text mining, we have an opportunity to amplify and extend the study of these dynamics on a wider scale. By examining the rhetoric used across time and debates in parliament, we can trace episodes where attitudes towards gender were shifting – as well as moments when prejudices about women were reinforced. We can identify the work of parliament as a collective in perpetuating or challenging gender norms.

By foregrounding grammatical part-of-speech extraction rather than word counts alone, we bring into focus how syntax structures engendered power, echoing the concerns of these theorists. For example, in Parliamentary debates on the Contagious Diseases Acts, women are rarely subjects of political action. They are objects—spoken of, regulated, and scrutinized—but almost never granted the syntactic position of agent. This syntactic pattern, we argue, does not merely reflect political ideology; it is one of the mechanisms by which that ideology is naturalized in discourse.

This chapter delves into grammatical part-of-speech extraction for analyzing gender and agency by first highlighting the distinction between lexical analysis—the study of words and their meanings—and syntactic analysis, which focuses on the relationships between words and how they form larger structures, such as sentences. Lexical analysis involves categorizing words (e.g., verbs, nouns, adjectives), whereas syntactic analysis investigates how these words interact and connect to convey meaning. By exploring these interactions, researchers performing syntactic analysis can glean insights that lexical analysis alone may not reveal. For instance, syntactic analysis allows us to

compare how different groups of people are described as subjects (those performing actions) versus objects (those affected by actions) in texts over time.

As we begin to explore the usefulness of syntactic analysis for understanding historical text, we circle back to the Hansard corpus. The mid-nineteenth century, marked by the passage of the Contagious Diseases Acts, provides a distinct historical context for examining how the language of agency regarding gender are embedded within syntactic structures.

## The Study of Power in Syntax

When we speak of “syntax” as having a role in perpetuating political points of view, we literally mean that the noun-verb relationships in sentences regularly embody power relationships. To demonstrate this concept, consider a passage from George Eliot’s **Daniel Deronda** (1876), a novel contemporary to the Hansard corpus that depicts the masculine gaze with exemplary specificity. The narrator describes the female protagonist, Gwendolyn, stating: “She was the central object of that pretty picture, and everyone present must gaze at her.” Among the relationships embedded in the sentence is a statement where the object of “gaze” is “her.” The female character is the passive receiver of an active gaze. In this case, the subject doing the gaze is “everyone,” an ungendered world.

Feminist theorists have shown that from Victorian novels to twentieth-century cinema, the gaze was often gendered. The masculine gaze refers to a way of positioning women as objects to be looked at, evaluated, or acted upon—typically through a masculine perspective that exerts social power. The roots of this theory stem from cinema studies. Laura Mulvey famously argues in “Visual Pleasure and Narrative Cinema” (1975) that cinema positions the viewer as male and women as passive objects of visual pleasure. However, Mulvey’s argument has since been widely extended to the study of literature, history, and language in which the male subject functions as a mechanism of visual and narrative control, framing women as passive subjects of observation rather than active agents (Mulvey 1975). Here we will demonstrate that a similar pattern can be identified through syntactic analysis.

With text mining, we have the opportunity to test whether the subjects who gaze are always gendered male and the subjects who are gazed at are always gendered female. We can ask a computer to extract every instance of the word “gaze” and the pronouns associated with it, even when it is presented in different grammatical forms, such as “will gaze at him/her,” or “gazing at him/her.” Next, we can quantify the number of times the word “gaze” was applied to “her” versus to the number of times “gaze” was applied to “him.” Part-of-speech extraction thus allows the analyst to aggregate a great deal of information about any given verb: when it arises, who uses it, and what other words are used in a similar context. The same technique can be applied to all verbs in a given corpus, and so give a portrait of *who* is doing the action and who receives the action in the minds of people associated with an institution.

In digital literary studies, part-of-speech (PoS) extraction has been widely used to analyze gender within text, for instance, by allowing scholars to count the number of times female and male charac-

ters are referred to, either by name or by pronoun (Chen 2020, Kraicer and Piper 2018, Underwood et. al 2018). The process of breaking prose into discrete linguistic units and then using these units to guide analysis has previously been described as a form of “reductive reading” (Allison 2018). However, this characterization has primarily been applied to the bag-of-words (BoW) approach to text analysis, in which words are counted without consideration of their grammatical context or their syntactic relations (Arnold 2019). In this chapter, we seek to extend this line of inquiry by making a distinction between “part-of-speech extraction,” which tags and analyzes parts-of-speech without regard to grammar, and “grammatical part-of-speech extraction,” which accounts for the syntactic relationships between words, incorporating grammatical rules to determine the function of each part-of-speech within a sentence.

We suggest that extracting grammatical parts-of-speech offers a more effective framework for analyzing how language encodes gender and power than the bag-of-words (BoW) approach, which ignores linguistic context by analyzing words in isolation. While grammatical relationships have been analyzed in text, the algorithmic techniques that shape and inform interpretive analysis—particularly those related to grammar—have received comparatively little attention (Tangherlini 2016, Franzosi 2024).

For brevity, throughout this chapter we will call this method of extracting and analyzing words based on their grammatical function “grammatical POS analysis.” Understanding the parts-of-speech in a sentence begins with recognizing that words take on different roles depending on their relationships to one another. For instance, “women” and “themselves” are both nouns, but they serve different purposes in a sentence. The noun “women” performs an action, while the noun “themselves” receives an action. Recognizing these roles helps clarify how each word functions in the sentence.

In this chapter, we will introduce two approaches to perform grammatical POS analysis. First we will extract parts-of-speech using a rule-based, statistical model. After, we will use a large language model (LLM) to identify and generate parts-of-speech for analysis.

The rule-based, statistical model we will use is **spaCy**. spaCy is a natural language processing library that analyzes text using linguistic rules and pretrained models. It has been trained on annotated texts to identify patterns in language based on fixed data processing approaches. spaCy’s models are well-calibrated for text processing tasks, but are limited to extracting data—that is, identifying and labeling linguistic features. LLMs, on the other hand are trained on massive text corpora to predict the next word in a sequence. Instead of extracting text LLMs generate new text.

As we will demonstrate, both of these modes give us a means to perform grammatical POS analysis. Using **spacyr**, we will extract syntactic relationships to analyze adjective-noun pairs, specifically the adjectives assigned to the word “woman.” Following this, we will use a LLM to identify the top action statements—statements about subjects performing actions to objects—by generating subject-verb-object relationships using different prompting techniques. Both of these approaches will help us see different linguistic features related to gender and agency.

## The `spacyr` Programming Library

To analyze the syntax within sentences, analysts typically count nouns (or other parts of speech) and the verbs, adjectives, other nouns, or prepositional phrases associated with them. Appreciating the software behind this counting mechanism can help us to appreciate the high level of accuracy that we can expect from syntactical analysis today.

A linguistic pipeline transforms raw text into structured linguistic information using pre-trained language models (LMs)—not to be confused with “large language models” (LLMs), which are designed for broad generative tasks such as open-ended question answering, text generation, and conversation. These smaller-scale LMs are trained on large collections of text (called corpora) to predict and label specific linguistic features, such as parts of speech, syntactic dependencies, and named entities.

While the concept of a linguistic pipeline has its roots in the 1990s, it gained traction with the development of the Penn Treebank (Marcus et al. 1993), which provided a large, annotated corpus of English text marked with part-of-speech tags and phrase-structure parse trees. This resource enabled researchers to train and evaluate models for individual linguistic tasks in a standardized, sequential fashion, laying the groundwork for modern pipeline architectures. The Penn Treebank was revolutionary because it enabled a shift from rule-based systems—which relied on manually crafted grammatical rules written by experts—to predictive approaches, where statistical models learn patterns directly from annotated examples. This meant that rather than defining how English syntax should work, researchers could build models that learned syntactic patterns from real-world usage. This shift was important because it made NLP systems more adaptable, where models could be trained on different linguistic variations.

By the 2010s, linguistic pipelines had become widely adopted through frameworks such as Stanford CoreNLP and spaCy (Manning et al. 2014; Honnibal and Montani 2017). These pipelines allowed models to learn patterns in language—such as sentence structure, word meanings, and syntactic relationships—by analyzing contextual usage (that is, not just the isolated meaning of a word, but how its meaning and grammatical role are shaped by surrounding words in a sentence—for example, distinguishing whether run is used as a noun in “a long run” or as a verb in “they run daily”). Once trained, these models can process new text by identifying syntactic roles (e.g., subject or object), part-of-speech tags (e.g., noun, verb), and dependencies between words.

The `spacyr` package builds on top of the Python library spaCy. In this way, `spacyr` acts as a bridge: it wraps spaCy Python code and runs spaCy behind the scenes while allowing analysts to access its functions through R. This design enables analysts to take advantage of spaCy’s robust linguistic tools—such as tokenization, part-of-speech tagging, and dependency parsing—without needing to write Python code or switch programming environments from R to Python. However, this means `spacyr` requires additional setup to integrate Python’s `spaCy` library with R. This involves two key steps: installing `spaCy` in a compatible Python environment and configuring R to communicate effectively with Python.

`spacyr` is a natural language processing (NLP) library. Analysts can use `spacyr` to perform tasks

such as:

- Tokenization: Breaking text into words, punctuation, and other meaningful units.
- Part-of-speech tagging: Identifying the grammatical role of each word in a sentence (e.g., noun, verb, adjective).
- Named entity recognition (NER): Identifying proper nouns, dates, locations, and other named entities in text.
- Dependency parsing: Analyzing the grammatical relationships between words in a sentence (e.g., which word is the subject, which is the object).
- Lemmatization: Reducing words to their base forms (e.g., “running” to “run”).

`spacyr` currently provides NLP support for approximately 75 languages, including Spanish and Latin. However, a major limitation of spaCy’s out-of-the-box language coverage is that its pre-trained models primarily target state, majority, or colonial languages. As a result, Indigenous and low-resource community languages remain largely unsupported within the spaCy ecosystem. However, spaCy does offer components for researchers to train their own language models.

By now, it should be clear that tokenization can be performed using a range of packages, such as Silge and Robinson’s `unnest_tokens()` or, as shown here, using `spacyr`. However, this is the first time we have engaged with more advanced natural language processing capabilities like named entity recognition, dependency parsing, and lemmatization. Taken together, these tools allow analysts to move beyond bag-of-words style text analysis, where words are counted without regard to grammar, and begin extracting structured information and identifying grammatical roles embedded in Hansard. `spacyr` supports grammar-level analysis through a sequence of computational, data-processing steps known as a “linguistic pipeline.”

## Installing `spacyr`

To install `spacyr`, we begin by using the standard `install.packages()` function. This gives us access to the `spacy_install()` function. Running `spacy_install()` sets up the required Python environment and installs the spaCy library along with its dependencies, integrating spaCy’s capabilities within the R environment.

```
# First time only -- installing spacyr on your computer
install.packages("spacyr")

library(spacyr)

# Install spaCy and the small English model
spacy_install(lang_models = "en_core_web_sm", ask = FALSE)

# Initialize spaCy
spacy_initialize(model = "en_core_web_sm")
```

By default, spacyr installs an English language model that used to tag words with their parts-of-speech, called `en_core_web_sm`. This name provides specific information about the model:

- `en`: Indicates the model is for the English language.
- `core`: Refers to a general-purpose model, suitable for a wide range of tasks.
- `web`: Specifies that the model was trained on web-based data, such as blogs, news articles, and other online text.
- `sm`: Stands for “small,” meaning the model is lightweight and optimized for efficiency, prioritizing speed and lower resource usage over maximum accuracy. spaCy can also access medium and large language models, which tend to have higher accuracy but at the cost of more computational resources.

SpaCy supports many languages beyond English. These LMs can be downloaded using the `spacy_download_langmodel()` function. For example, the following code demonstrates how to download and initialize a small German model, which is specifically trained on a news corpus. For a complete list of available models and their details, visit the official spaCy models page: <https://spacy.io/usage/models>.

```
spacy_download_langmodel("de_core_news_sm") # load the German model
```

In our analysis we will use the `en_core_web_sm` model to analyze text in the English language. Once spaCy has installed, we can initialize the model we wish to use.

```
spacy_initialize(model = "en_core_web_sm") # load the English model
```

In each future session, we will need to not only load the spacyr library, but also initialize the model like so:

```
# Entire startup for future sessions
library(spacyr) # load the library

spacy_initialize(model = "en_core_web_sm") # load the English model

## NULL

library(kableExtra)
```

## **spacyr in Action: Outlining the Dependency Grammar of a Single Sentence**

### **The Politics of Parliamentary Language**

To consider how syntax might work as a tool of historical analysis, let's consider a case well-known to feminist historians, the Contagious Diseases Acts, passed between 1864 and 1869.

The Acts were a series of legislative measures aimed ostensibly at mitigating the spread of venereal disease among military personnel, but in practice they disproportionately targeted working-class women. Under these acts, women suspected of being prostitutes in garrison towns and naval ports could be subjected to forced medical examinations and, if found to be infected, detained in “lock hospitals”—medical institutions specifically designated for the treatment of venereal diseases—until decidedly “cured”. The legislation rested on hegemonic assumptions about female sexual deviance and male sexual entitlement, effectively criminalizing female bodies while exempting the men who procured their services.

In examining the discourse on women, we find that the women themselves are largely absent from the historical record; instead, the debates reveal how members of Parliament imagined and constructed the figure of women and the anxiety of prostitution through language.

Consider an amendment suggested by a member of parliament, Acton Aryton, as recorded in Hansard:

“He proposed to amend the clause which related to women submitting themselves voluntarily to periodical medical examination by requiring them to make a declaration that they were persons which the law recognizes only to punish.”

Aryton feared that allowing women to seek certificates of health from government officials would legitimize prostitution. Aryton’s amendment would require that these women instead formally declare that they belong to a category of criminals – “people whom the law only recognizes in order to punish.” Women who carried venereal diseases would therefore be held responsible for identifying themselves as prostitutes and transgressors, submitting their bodies before the law.

Close attention to grammatical syntax further reveals the way in which Aryton imagines women as immoral agents can be abstracted from the grammatical relations in the sentence. Among the relationships embedded in this sentence, the subject “women” is performing the verb-form “submitting” and the direct object “themselves.” This utterance takes the prepositional phrase “to examination” as its complement. With the help of syntactical analysis, we can simplify the statement to its main components:

women-submit-themselves

Or, if we include the dependent clause, we may render it this way:

## women-submit-themselves-to-examination

Notably, the actual historical text in no way fits the simple binary “male-observer, female-observed” from the theory of the male gaze. In Ayrton’s actual phrase, women occupy both the subject and object positions – grammatically. That is, women are portrayed both as potentially the agents doing action as well as the objects receiving action. The syntax specifies a legal framework that would require women to act – seeking health certificates to remedy their supposedly immoral behaviors.

Nevertheless, the remedy that Ayrton is recommending for the spread of venereal disease is one where women recognize themselves as potentially criminal and in need of surveillance. In any case, the verb “submit” captures women’s inferior position in this clause – even if the sentence neglects to specify the men to which women are submitting their power. If we find several sentences in which “women submit” plays a major grammatical role in the sentence, then the analyst must conclude that the women are party to a system that requires the deferral of their agency.

What the syntax helps us to discover is the relationship between particular kinds of subjects (women) and the actions required of them (submission). By focusing on words’ grammatical functions, we can identify relationships that capture these nuances, even when the words do not appear next to each other in the sentence.

The specifics of what system women were submitting to, or why women must “submit themselves,” an analyst may derive from close reading of the sentences involved. Word counts alone do not capture Ayrton’s description of the women’s agency because they are not encoded in any single word, or in any finite span of words (like in the case of n-grams).

`spacyr` analyzes the syntactic structure of sentences using dependency grammar. It represents each sentence as a dependency tree, where words are connected by directed links that show their grammatical relationships, called dependencies. One word functions as the governing word (or head), which determines the grammatical type of the relationship—such whether the word is a subject, object, or modifier—while the other is the child (or dependent), whose role and interpretation are shaped by the head. The child relies on the head to anchor its grammatical function and meaning; for example, an adjective depends on the noun it modifies, and a subject depends on the verb to complete its role in the sentence.

Figure 1 shows a dependency tree generated by `spacyr`, where each word is connected to its head by a directed arc. These arcs indicate the type of grammatical relationship between the words—such as subject, object, or modifier—with the direction of the arc pointing from the head to the dependent (child).

nsubj



In this example, the head of the verb “submitting” is the subject “women.” The verb “submitting” also has multiple children: “themselves,” “voluntarily,” and “to.” The head of the adjective “medical” is the object “examination.” The head can vary based on the type of phrase or clause. For instance, in noun phrases, the noun is typically the head, while in verb phrases, the verb often functions as the head. With an understanding of how `spacyr` processes linguistic data, we can now use it to parse our Hansard data and extract parts-of-speech with a more critical eye.

### Counting Adjectives and Nouns Collocated With the Word “Woman”

In investigating parliamentarian’s speech about women, we used a layered approach to analysis allows us to trace patterns of language use. Our aim is to explore how members of Parliament in the mid-19th century described and imagined women, and to consider how these linguistic patterns connect back to broader cultural frameworks. Specifically, we will ask: How are women represented in the language of 19th-century Parliamentarians, as reflected in the 1860 Hansard debates?

We employ a series of data processing steps to reveal how different dimensions of meaning and association emerge from the data. We begin by counting the adjectives that occur in the same sentences as the nouns woman or women, using co-occurrence as a proxy for thematic or descriptive relevance. Next, we identify other nouns that frequently appear alongside the nouns woman or women within those sentences, mapping the broader semantic field in which the term operates. Building on this overview, we then narrow our focus to identifying and counting adjective–noun pairs, where woman is directly modified by an adjective.

To begin, we preprocess the text of the 1860 Hansard debates. This involves filtering the corpus to include only sentences that contain the word “woman” or “women,” regardless of capitalization. Because our focus is on sentences involving references to these words, we limit the scope of the data processing by first filtering the Hansard data. This filtering step narrows the dataset considerably, allowing for a more manageable analysis using `spacyr`.

```
library(tidyverse)
library(hansardr)

data("hansard_1860")

hansard_woman_1860 <- hansard_1860 %>%
  filter(str_detect(text, regex("woman|women", ignore_case = T)))
```

We then provide our filtered hansard data to the `spacy_parse()` function to tokenize just the text column. From the text column, we extract linguistic features like part-of-speech tags by specifying `dep = TRUE`. To save computational resources and focus on only the features relevant to our analysis, we then disable lemmatization (`lemma = FALSE`) and named entity recognition (`entity = FALSE`).

```

parsed_hansard_woman <- spacy_parse(hansard_woman_1860$text,
                                       dep = TRUE,
                                       lemma = FALSE,
                                       entity = FALSE)

head(parsed_hansard_woman)

```

```

##   doc_id sentence_id token_id    token   pos head_token_id  dep_rel
## 1     1          1        1      The   DET            2       det
## 2     1          1        2    women  NOUN           6  nsubjpass
## 3     1          1        3    alone  ADV            2    advmod
## 4     1          1        4     were  AUX            6 auxpass
## 5     1          1        5     well  ADV            6    advmod
## 6     1          1        6  clothed  VERB           6      ROOT

```

`spacy_parse()` returns a new data frame in which each row represents an individual token from the text. This data now includes the token's part-of-speech (POS) tag and syntactic dependency, which can be used for further analysis. For instance, I can filter and count to find the top adjectives stated in the Hansard sentences containing the word "woman" or "women."

```

# Filter for just words tagged by spacyr as adjectives
adjectives <- filter(parsed_hansard_woman, pos == "ADJ")

# Count occurrences of each adjective and sort in descending order
top_adjectives <- adjectives %>%
  count(token, sort = TRUE)

# View the top adjectives
head(top_adjectives, 30) %>%
  kable()

```

token	n
other	156
young	155
married	133
great	127
poor	114
old	107
many	104
such	100

token	n
same	65
own	60
more	57
large	52
public	50
last	48
noble	48
first	45
present	43
certain	42
small	40
good	38
deceased	34
guilty	31
much	30
unfortunate	29
bad	28
whole	28
able	26
second	26
little	25
necessary	25

The results of this process show that the sentences that mention “woman” or “women” from the 1860s most often also contain the word “married.” The finding is hardly earth-shattering; some of these debates led up to the first Married Women’s Property Act (1870), which overturned centuries of precedent that had ruled that married women could not own or inherit property. The new act allowed married women to keep wages and inherit property up to £200 – giving women a wide degree of freedom if they found themselves divorced, abandoned, or married to an abusive or otherwise incompatible husband from whom they needed to separate. The point is that the adjective “married” doesn’t give us an exceptionally new information about history. We should look down the list further.

However, many of our top words are common stop words (e.g. “other,” “such,” “many,” and “same”), which can obscure more meaningful patterns when attempting to analyze the words that co-occurred in the same sentence as “woman” or “women.” Removing stop words using the `tidytext` package can help uncover more insightful trends in the corpus.

```
library(tidytext)

data(stop_words)
```

```

top_adjectives_clean <- top_adjectives %>%
  anti_join(stop_words, by = c("token" = "word"))

head(top_adjectives_clean, 30) %>% kable()

```

token	n
married	133
poor	114
public	50
noble	48
deceased	34
guilty	31
unfortunate	29
bad	28
true	25
Turkish	24
common	24
proper	24
male	23
single	23
impossible	22
strong	22
female	21
fair	20
Christian	19
English	19
moral	19
separate	17
similar	17
Catholic	16
Roman	16
domestic	16
liable	16
medical	16
political	16
sufficient	16

Removing stop words reveals something interesting: several of the remaining adjectives refer to a woman's marital status, such as "single" and "unmarried." Others suggest class distinctions, including words like "poor," "noble," "common," and "domestic." Some discuss issues with inheritance,

like “deceased.” Others suggest debates about crime and punishment – “guilty,” “bad,” and “impossible.” The words about which we have the most questions are “Turkish” – what were Turkish women doing in Britain in the 1860s, and why were women in particular gaining the attention of parliament?

## Analyzing Nouns

We can also use part-of-speech extraction to analyze nouns. Nouns often indicate the focus of discourse, showing what ideas, objects, or individuals are central to the narrative or argument. Repeated nouns might signify motifs. In the following code, we filter for the top nouns that occur in sentences with the word “woman” or “women” (other than “woman” or “women” themselves).

```
# Filter for just nouns from this same data
nouns <- filter(parsed_hansard_woman, pos == "NOUN")

# Exclude 'woman' and 'women' from the nouns
filtered_nouns <- nouns %>%
  filter(!(token %in% c("woman", "women")))

# Count occurrences of each noun and sort in descending order
top_nouns <- filtered_nouns %>%
  count(token, sort = TRUE)

# View the top nouns
head(top_nouns, 20)
```

```
##          token    n
## 1      children 411
## 2          men 388
## 3         man 242
## 4        law 142
## 5       case 136
## 6   country 120
## 7   persons 116
## 8     number 111
## 9   husband 101
## 10    cases  96
## 11 property  93
## 12   child  89
## 13    time  86
## 14   years  85
```

```

## 15      wife  84
## 16  marriage  81
## 17 employment  77
## 18   question  66
## 19     place  62
## 20      day  61

```

The top nouns to appear in sentences containing the word “woman” or “women” are “men,” followed by “children” and “man.” The prevalence of these words in the same context highlights an association between women, the Victorian family structure, and motherhood. Women were often framed as caregivers and moral guardians. In the context of Hansard, this may reflect debates on women’s character. For example, discussions surrounding the Contagious Diseases Acts often framed women’s morality and behavior as directly impacting children and family stability.

Comparing the top nouns in sentences containing “woman” or “women” against those containing “man” or “men” reveals significant differences in how these genders were discussed in 19th-century parliamentary debates. To demonstrate this, we perform the same data processing steps but on just sentence with the words “man” or “men.”

```

hansard_man_1860 <- hansard_1860 %>%
  filter(str_detect(text, regex("\\b(man|men)\\b", ignore_case = T)))

parsed_hansard_man <- spacy_parse(hansard_man_1860$text,
                                     dep = TRUE,
                                     lemma = FALSE,
                                     entity = FALSE)

nouns <- filter(parsed_hansard_man, pos == "NOUN")

# Exclude 'man' and 'men' from the nouns
filtered_nouns <- nouns %>%
  filter(!(token %in% c("man", "men")))

# Count occurrences of each noun and sort in descending order
top_nouns <- filtered_nouns %>%
  count(token, sort = TRUE)

library(gridExtra)

# Arrange the output tables in a grid layout
# [1:15, ] selects the first fifteen rows (and all columns) from the dataset
# tableGrob() converts each chunk into a table for plotting

```

```
# grid.arrange() places the tables side by side so we can view more at once
grid.arrange(tableGrob(top_nouns[1:15, ]),
             tableGrob(top_nouns[16:30, ]),
             tableGrob(top_nouns[31:45, ]),
             ncol = 3)
```

token	n	token	n	token	n
1 country	3659	16 part	1112	31 money	909
2 number	2704	17 position	1069	32 course	896
3 time	2374	18 day	1040	33 way	871
4 years	2069	19 character	1039	34 property	790
5 service	1718	20 body	1034	35 means	775
6 question	1698	21 power	1033	36 order	763
7 year	1647	22 life	1032	37 ships	734
8 case	1617	23 force	1031	38 matter	727
9 opinion	1611	24 people	1020	39 respect	725
10 officers	1527	25 war	1020	40 franchise	722
11 army	1401	26 subject	1000	41 opinions	697
12 system	1315	27 persons	963	42 state	694
13 law	1304	28 duty	948	43 principle	665
14 class	1276	29 place	930	44 right	659
15 hon	1202	30 fact	912	45 work	654

Some of these words, including “country,” “number,” “time,” “question,” “case,” “opinion,” and “respect” rank among the most prominent words in parliament overall, suggesting that “man” and “men” feature as the default personality for all discussions of politics. Collocates like “officers,” “army,” “position,” “force,” “war,” and “ships” remind us that these debates transpired in an era when both the military and government bureaucracies were entirely occupied by men. Mentions of “money,” “fact,” and “property” remind us that in this era, capitalism and science were equally realms of male prerogative, while “franchise” reminds us that contemporary debates over the vote reflected the contestation of the vote for men.

Most of these words are easy to overinterpret. A naive analyst might classify “respect” as a signifier that men, not women, were deserving of respect, whereas the fact that the word is also one of the most prominent in Hansard overall reminds us that the language of the time often used clauses

like “a debate respecting the new tax;” such respect has nothing to do with Aretha Franklin’s plea for respect for women. Women as well as men were sometimes debated in terms of “rights” and “work;” it is possible to project exclusions on this list that are in no way indicated by the data, and careful analysts will guard against this mistake.

As with most processes of text mining, it is rarely the most frequent words and categories that are the most meaningful. Looking for collocates of “man” and “men,” we’re unlikely to make new discoveries about history; more compelling would be tracing the distinctiveness of the collocates of “man” and “men” over time, say comparing the collocates of the 1860s with those of the 1870s.

Another approach is to examine grammatical relationships or dependencies, where relationships are encoded at the level of syntax.

## Extracting Adjective-Noun Dependencies

While analyzing individual parts of speech—such as nouns and adjectives—can offer insights into language patterns, a more nuanced understanding arises when we consider how these words interact within a sentence. For example, examining how nouns are modified by adjectives reveals not just what is being discussed, but how it is being described. For instance, examining the adjectives that modify nouns like “woman” or “man” can reveal perceptions embedded in text. Adjectives often convey evaluative or descriptive judgments, shaping how the nouns they modify are framed. They reflect the speaker’s perspective and can influence how the audience interprets the subject. When words are analyzed in isolation, these contextual nuances are lost, leading to an interpretation that may be disconnected from the deeper meanings embedded in the text. To capture these contextual nuances more effectively, we now shift our focus to analyzing adjective-noun pairs.

The code below is used to extract adjective-noun pairs from the Hansard debates. Specifically, it focuses on finding adjectives that describe the nouns woman or women.

When we processed the Hansard text using the `spacyr` package, it automatically added a column called `doc_id`. This column assigns a unique ID to each sentence, which retains metadata on the structure of the original data for when we tokenize the text. Therefore, we can be sure that parts-of-speech (like adjectives and nouns) are only matched within the same sentence and not across unrelated sentences.

We then define a function, `get_adjective_noun_pairs()`, that extracts adjective-noun pairs from text. While the present analysis focuses on utterances containing the terms “woman” and “women,” the function is designed for broader applicability. By encapsulating these steps in a reusable function we can easily use these data processing steps in a subsequent analysis of sentences involving the terms “man” and “men.”

```
# Filter for adjectives and their associated nouns
# Extract adjective-noun pairs where the adjective modifies
# the noun within the same document
```

```

get_adjective_noun_pairs <- function(parsed_data, nouns) {
  adjective_noun_pairs <- parsed_data %>%
    # Process each document separately using a grouping variable
    group_by(doc_id) %>%
    # Filter for adjectives with the dependency relation
    # "amod" (adjective modifying noun)
    filter(pos == "ADJ" & dep_rel == "amod") %>%
    # Select columns for adjective token, its token ID, and the head
    # noun's token ID
    select(doc_id, token_id, token, head_token_id) %>%
    # Join to match adjectives to their corresponding nouns within
    # the same document
    inner_join(
      parsed_data %>%
        select(doc_id, token_id, token),
      by = c("doc_id", "head_token_id" = "token_id"),
      relationship = "many-to-many",
      suffix = c("_adjective", "_noun")) %>%
    # Filter for cases where the noun is specified in the 'nouns' argument
    filter(token_noun %in% nouns) %>%
    # Combine adjectives and nouns into a single string for readability
    mutate(adjective_noun_pair = paste(token_adjective, token_noun)) %>%
    # Ungroup after processing
    ungroup() %>%
    select(adjective_noun_pair)

  return(adjective_noun_pairs)
}

nouns_to_filter <- c("woman", "women")
adjective_noun_pairs_woman <- get_adjective_noun_pairs(parsed_hansard_woman,
                                                       nouns_to_filter)

```

The code performs a series of steps: first it filters for adjectives with the dependency label `amod`, which stands for “adjectival modifier.” These are adjectives that directly modify nouns. The code then joins each adjective with the noun it modifies. At this point, we have processed all

adjective-noun pairs within these sentences. We are, however, only interested in adjective-noun pairs containing the word “woman” or “women.” Therefore, we added a filtering step to only keep relevant pairs. Finally, the code combines the adjective and noun into a single string to make the results easier to read. The final result is a data frame that lists all such adjective-noun combinations found in the corpus.

```
adjective_noun_pair_counts_woman <- adjective_noun_pairs_woman %>%
  count(adjective_noun_pair, name = "count") %>%
  arrange(desc(count))
```

```
grid.arrange(tableGrob(adjective_noun_pair_counts_woman[1:15, ]),
             tableGrob(adjective_noun_pair_counts_woman[16:30, ]),
             tableGrob(adjective_noun_pair_counts_woman[31:45, ]),
             ncol = 3)
```

	adjective_noun_pair	cou	adjective_noun_pair	cou	adjective_noun_pair	cou
1	married women	87	unhappy women	71	Catholic women	2
2	young woman	58	Christian women	62	Indian women	2
3	poor woman	53	helpless women	63	Many women	2
4	old woman	39	other women	64	Presbyterian woman	2
5	married woman	35	Chinese women	45	Turkish women	2
6	young women	36	drunken woman	46	aged woman	2
7	old women	26	respectable women	47	aged women	2
8	poor women	18	Irish women	38	black woman	2
9	many women	19	bad woman	39	bodied women	2
0	unfortunate women	10	common women	10	defenceless women	2
1	unmarried women	10	few women	11	deserted women	2
2	English women	12	most women	12	disorderly women	2
3	single women	18	naked women	13	drunken women	2
4	unfortunate woman	18	single woman	14	handsome woman	2
5	unemployed women	15	Catholic woman	15	harmless women	2

It is easy to explore the top 10, 20, or even top 100 adjective-noun pairs by slicing the data and presenting the results in a table. Adjusting the number of pairs displayed allows us to shift between a broader overview and a more focused examination of the corpus.

Without considering the historical context of the extracted words, a list of top adjective-noun pairs suggests that members of parliament were particularly concerned with women in a state of distress, including the “poor,” “unmarried,” “unfortunate,” and “unemployed,” which suggests parliament’s role in welfare. Other common patterns, suggest that parliament was also presiding in moral judgment, not only distinguishing the “respectable woman” who merited welfare, but also shaming the “unhappy,” “helpless,” “drunken,” “bad,” and “disorderly” women who merited criminal attention rather than welfare. That parliament had so many categories for female misbehavior suggests how much space they gave this question. The words are likely a reflection of the beliefs of members of Parliament expressed in the course of legislating the Contagious Disease Acts, where questions of sexual propriety and behavior came to the fore.

Other questions are easily raised for which the analyst needs to resort to close reading: in what context did Chinese, Indian, Irish, and Turkish women come to parliament’s attention during this period? When subjects come to parliament, they have necessarily risen to the level of public concern. Whether or not the two mentions of Chinese women is indexical of a particular political event in parliament, the story that brought these Chinese women to parliament in the first place was necessarily an episode of social history that rose above the common pulse of everyday life, and the analyst would do well to investigate this story.

Returning to a single debate helps clarify one way a Parliamentarian might use the word “poor” in relation to “woman.”

```
woman_context <- hansard_woman_1860 %>%
  filter(str_detect(text, regex("Chinese wom", ignore_case = TRUE))) %>%
  select(text)

## So great has been the satisfaction of the Chinese with those regulations that
## whereas before it was impossible to obtain Chinese women and children to
## accompany Chinese men to the Colonies--whereby great evils resulted--whole
## families have since been embarked, and 300 or 400 respectable Chinese women
## have been sent to our Colonies.
##
## Mr. Austin had also succeeded last year in inducing 300 Chinese women to
## accompany the immigrants to the West Indies
##
## He also knew that there was a strong feeling in New South Wales against
## increasing the number of Chinese males in Australia, there being scarcely any
## Chinese women in the country, and the Chinese in the Colony amounted to 21,
## 000, or about one-fourth of the adult male population.
```

The speech refers to the questioning of managing British colonies in the West Indies. After the abolition of slavery in the British Empire in 1834, plantation owners in the Caribbean (notably in British Guiana, Trinidad, and Jamaica) sought replacement labor for sugar plantations. They

turned to indentured labor from India, China, and other parts of Asia and Africa. Laborers were often recruited (or deceived) into contracts. The Chinese laborers were typically from southern China, especially Fujian and Guangdong.

The speech indicates a shift of policy, where empire became specifically interested in recruiting “respectable Chinese women.” The mentions of “great evils” suggests administrators’ hopes that these indentured women were expected to pacify a spirit of resistance amongst Chinese indentured laborers. We also read about racial tensions in New South Wales that suggest that similar policies would not be welcome in other parts of empire.

We see parliament here managing ethnic populations – and specifically women from China – as a means of providing labor to ensure the unceasing flow of commodities. Women’s bodies are expected to be part of that machinery, pacifying resistance through providing outlets for the sexual frustrations of male immigrants. Whether that outlet was consensual or not, and how women felt about being used to oil the wheels of empire, we would need to seek in other sources.

The point is that, if we were to trace the mentions of the Chinese, Irish, and Turkish women in the 1860s, or the women of different ethnicities mentioned in parliament overall, we would have an index of some of the most prominent and publicly-acknowledged ways that Britain’s government was using female bodies around the world – the policies they set, the utilitarian imaginings of how women’s bodies could be used to make empire more productive. That source might not be all we need for a sensitive portrait of women’s experience of a global empire, but text mining may nevertheless be a part of a larger process. We can use parliament as an index for important events that punctuate the flow of time over decades or centuries, and then turn to other archives to elaborate the cases as needed.

```
nouns_to_filter <- c("man", "men")
adjective_noun_pairs_man <- get_adjective_noun_pairs(parsed_hansard_man,
                                         nouns_to_filter)
```

```
adjective_noun_pair_counts_man <- adjective_noun_pairs_man %>%
  count(adjective_noun_pair, name = "count") %>%
  arrange(desc(count))
```

```
grid.arrange(tableGrob(adjective_noun_pair_counts_man[1:15, ]),
             tableGrob(adjective_noun_pair_counts_man[15:30, ]),
             tableGrob(adjective_noun_pair_counts_man[31:45, ]),
             ncol = 3)
```

adjective_noun_pair	count	adjective_noun_pair	count	adjective_noun_pair	count
young men	87	great men	142	honest men	87
poor man	57	professional men	143	commercial men	84
many men	42	rich man	144	public man	83
such men	31	great man	137	eminent man	78
young man	30	practical men	131	honourable men	74
eminent men	27	distinguished men	130	distinguished man	72
medical men	25	poor men	128	unfortunate man	70
other men	21	good men	127	naval men	64
public men	21	medical man	116	intelligent men	63
military men	20	honest man	115	competent men	60
best men	18	single man	109	old men	60
able men	17	military man	105	practical man	60
last man	13	few men	104	armed men	57
scientific men	15	other man	101	more men	54
great men	14	able man	96	very men	54
	16	old man	82		

To compare the words used of each gender, the appropriate next step is to make a list of just the adjectives and to subtract the counts for one gender from the counts for the other. This step, known as “vector subtraction,” will reveal the words that are most distinctive of each gender.

```

# Vector subtraction: words applied to men vs words applied to women

# extract just the adjectives used to describe men
man_adjectives <- adjective_noun_pair_counts_man %>%
  separate(adjective_noun_pair, into = c("adjective", "noun"), sep = " ") %>% # split into two
  count(adjective, name = "count")

# extract just the adjectives used to describe women
woman_adjectives <- adjective_noun_pair_counts_woman %>%
  separate(adjective_noun_pair, into = c("adjective", "noun"), sep = " ") %>% # split into two
  count(adjective, name = "count")

# join the two into one database and subtract the count of men from the count of women.
comparison <- full_join(man_adjectives, woman_adjectives, by = "adjective",

```

```

                suffix = c("_1", "_2")) %>% #join
  mutate(count_1 = replace_na(count_1, 0),
         count_2 = replace_na(count_2, 0),
         count_diff = count_2 - count_1) %>% # subtract men's count from women's
  filter(count_diff!=0) # delete all entries where there is no difference

# In this dataset, there are a large number of adjectives with a count of two.
# We will therefore "sample" from the results to draw our visualization.
# This means that every time the visualization below is redrawn, different results will appear

# take the top adjectives unique to women
positive_sample <- comparison %>%
  filter(count_diff > 0) %>% # must show a difference
  arrange(desc(count_diff)) %>%
  top_n(20) %>%
  mutate(Category = "Women")

## Selecting by count_diff

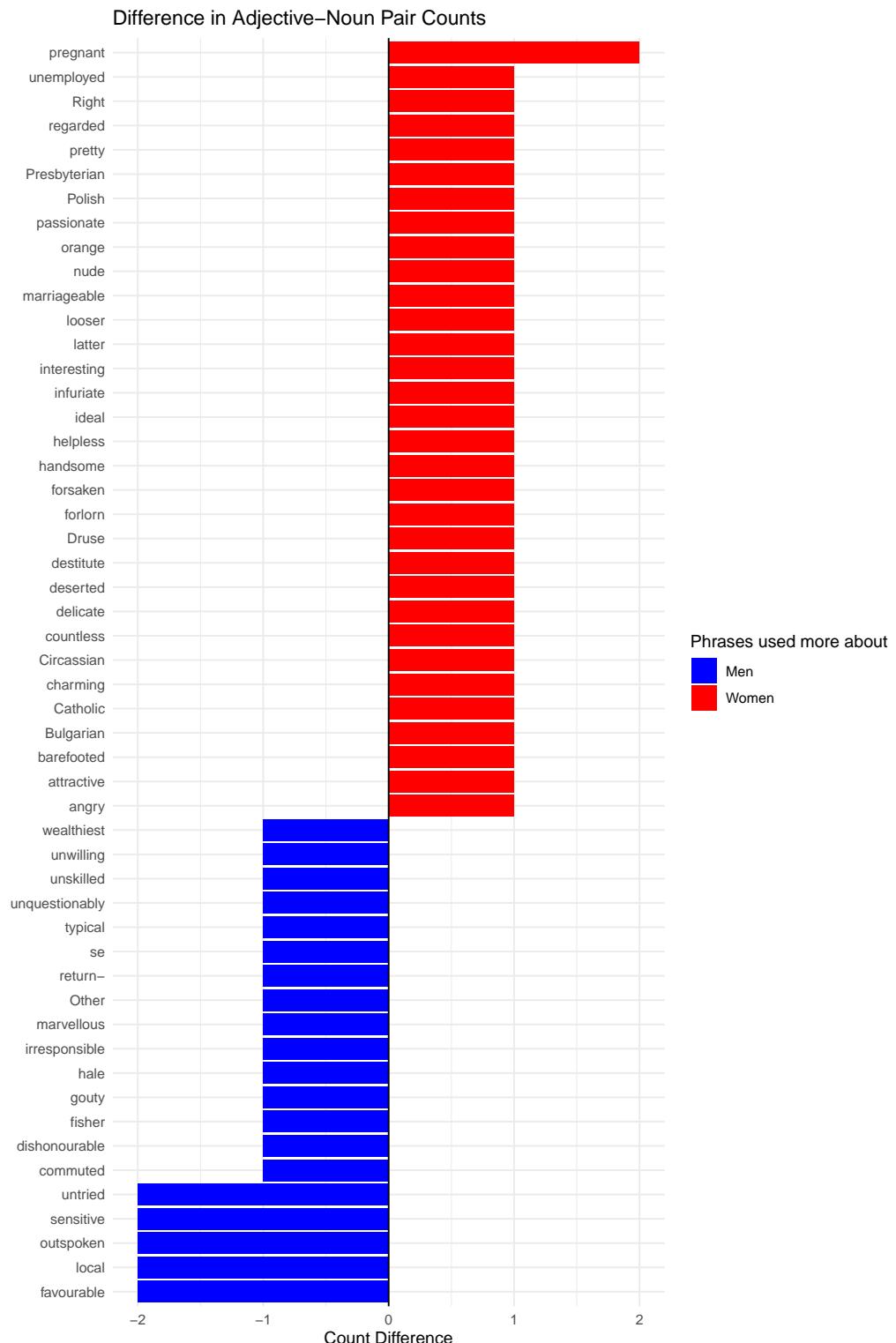
# take the top adjectives unique to men
# because there are so many with a max of 2, sample some positives
negative_sample <- comparison %>%
  filter(count_diff < 0) %>% # must show a difference
  sample_n(20) %>% # random sample
  mutate(Category = "Men")

filtered_df <- negative_sample %>%
  arrange(desc(count_diff)) %>%
  bind_rows(positive_sample) %>%
  mutate(adjective = fct_reorder(adjective, count_diff))

# Plot
filtered_df %>%
  ggplot(aes(x = adjective, y = count_diff, fill = Category)) +
  geom_col() +
  coord_flip() +
  scale_fill_manual(values = c("Women" = "red", "Men" = "blue"),
                    name = "Phrases used more about") +
  theme_minimal() +
  labs(
    title = "Difference in Adjective-Noun Pair Counts",

```

```
x = NULL,  
y = "Count Difference",  
fill = "Category"  
) +  
geom_hline(yintercept = 0, color = "black")
```



```

#library(gt)
#library(htmltools)

# take the top adjectives unique to men; because there are so many with a max of 2, sample some
#positive_sample <- comparison %>%
#  filter(count_diff > 0) %>%
#  arrange(desc(count_diff)) %>%
#  top_n(20) %>%
#  mutate(Category = "Women", Color = "blue")

# take the top adjectives unique to women
#negative_sample <- comparison %>%
#  filter(count_diff < 0) %>%
#  top_n(-20) %>%
#  mutate(Category = "Men", Color = "red")

# combine men and women
#filtered_df <- negative_sample %>%
#  bind_rows(positive_sample) %>%
#  mutate(adjective = fct_reorder(adjective, count_diff)) %>%
#  arrange(desc(count_diff)) %>%
#  select(adjective, count_diff, Category, Color) %>%
#  arrange(desc(abs(count_diff))) %>%
#  mutate(adjective = fct_reorder(adjective, count_diff))

# Function to generate a single gt table
#make_gt_table <- function(data_slice) {
#  data_slice %>%
#    gt() %>%
#    cols_label(
#      adjective = "Adjective",
#      count_diff = "Count Difference",
#      Category = "Used More About"
#    ) %>%
#    text_transform(
#      locations = cells_body(vars(Category)),
#      fn = function(x) {
#        color_map <- ifelse(x == "Women", "blue", "red")
#        mapply(function(txt, col) {
#          HTML(
#            paste0("<span style='color:", col, "'>", txt, "</span>"))
#          }, x, color_map)
#      })
#}

```

```

# Create three slices


```

```

make_kable_table <- function(data_slice) {
  data_slice %>%
    mutate(
      UsedMoreAbout = if_else(
        Category == "Women",
        cell_spec(Category, format = "latex", color = "blue"),
        cell_spec(Category, format = "latex", color = "red")
      )
    ) %>%
    select(
      Adjective = adjective,
      `Count Difference` = count_diff,
      `Used More About` = UsedMoreAbout      # Category not shown, just used for color
    ) %>%
    kable(
      format = "latex",
      booktabs = TRUE,
      escape = FALSE
    ) %>%
    kable_styling(full_width = FALSE, position = "center")
}

# ---- Create slices (adjust ranges as needed) ----
# If you have fewer rows, make these smaller (e.g., 1:20, 21:40, etc.)

table1 <- make_kable_table(filtered_df %>% slice(1:200))
table2 <- make_kable_table(filtered_df %>% slice(201:400))
table3 <- make_kable_table(filtered_df %>% slice(401:600))

# In your Rmd, print them in separate chunks, e.g.:
# ```{r}
# table1
# ```
# ```{r}
# table2
# ```
# ```{r}
# table3
# ```


```

When tested, we find that there are very few words that are used much more often to refer to one gender than the other, and those that there are occur quite rarely – a maximum of two times over

all of the 1860s. These small numbers are a warning against overinterpreting our findings. Nevertheless, the differences are instructive. Along with being “pregnant,” women are more likely to be depicted as vulnerable – “forlorn,” “delicate,” “forlorn,” “deserted,” “destitute” – and as an aesthetic object of desire – “charming,” “attractive,” “ideal,” “passionate.” When women are depicted as dangerous, they show a surplus of passion; they are “angry,” “infuriat[ing].” Men are more likely to be assigned virtues of character (“intellectual,” “enterprising,” “courageous,” “determined,” “devoted,” “humble,” “honourable,” “holy,” “reasonable,” “scrupulous”) or capacity (“sturdy,” “hale,” “robust,” “strongest) – or vices that show them as agents (“dangerous,” “evil,” “selfish,” “reckless”) or as having had their agency temporarily taken away (“aggrieved”, “inefficient,” “ignorant”).

These findings confirm that members of parliament articulated a gendered politics of agency as they discussed the matters of their day – assigning agency as the entitled domain of one gender, and vulnerability as characteristic of the other. While these findings may not be unexpected, they nonetheless affirm that our methodological approach can mirror established understandings of the period. At the same time, the results offer a different vantage point into the debates. Comparing the two visualizations side-by-side perhaps emphasizes the difference in the ways in which parliamentarian’s imagined men and women.

### **Extracting Simple Subject-Verb-Object Triples**

Extracting subject-verb-object (SVO) triples can provide another pattern through which we can understand the relationships and actions within a sentence, enabling analysis of agency and interactions between entities. Adjective-noun pairs, while useful for identifying descriptive attributes, lack this relational context and may not capture the dynamics of who is doing what, and to whom.

To learn about SVOs, we will explore how language was used by Parliamentarians during debates over the Contagious Diseases Acts in nineteenth-century Britain, as the adjectives used to describe women carried ideological significance. These pairings did not merely reflect social attitudes—they actively constructed categories of blame and protection under the law, determining who was subject to regulation and who was deemed worthy of sympathy. However, adjective-noun pairs are only one angle of analysis and must be situated within broader discourse patterns to fully understand how gendered power operated through legislative discourse.

For this analysis, we will extract simple subject-verb-object (SVO) triples. By calling these “simple” triples, we refer to subjects and objects that have a direct syntactic relationship with a root verb. A root verb is the main verb of a sentence that serves as the core action or state. Sentences can also include other types of verbs. These include auxiliary verbs (e.g., “is,” “have”), which provide grammatical support to the main verb; modal verbs (e.g., “can,” “should”), which express necessity or possibility; and participles, which are verb forms used as adjectives or to create verb phrases (e.g., “running” in “He is running”). Additionally, sentences may contain compound verbs, which consist of two or more verbs working together to express a single action or idea, such as “has been running.” Phrasal verbs combine a verb with one or more particles (prepositions or adverbs) to create a meaning distinct from the individual words, as in “give up,” which means “to quit.” Infinitive verbs, the base form of a verb often preceded by “to,” such as “to run,” can function as

nouns, adjectives, or adverbs in a sentence. However, for the purpose of extracting simple SVO triples, our focus will remain on the root verb and its direct relationships.

We will focus our analysis on sentences extracted from the 1860 Hansard debates that include the words “woman” or “women” stored in the variable named `hansard_woman_1860`. Our goal is to investigate how women are framed in Parliamentary discourse—specifically, whether they are portrayed as political agents, social subjects, or as objects within legislative or societal contexts. Additionally, we will examine the frequency and nature of specific associations made with women in these discussions to uncover possible trends.

For this analysis, we will also perform lemmatization on the sentences. We will again use spaCy for this task. To lemmatize with spaCy, we set the `lemma` parameter of `spacy_parse()` to TRUE. We will keep `entity` as FALSE to save computational resources and processing time, since we are not analyzing named entities.

We will also transition to analyzing the 1870 debates, a period when women’s voting rights were frequently contested in Parliament, revealing deeper struggles over power and political agency.

```
library(tidyverse)
library(hansardr)

data("hansard_1870")

hansard_woman_1870 <- hansard_1870 %>%
  filter(str_detect(text, regex("woman|women", ignore_case = T)))

parsed_hansard_woman_1870 <- spacy_parse(hansard_woman_1870$text,
                                         dep = TRUE,
                                         lemma = TRUE,
                                         entity = FALSE)

head(parsed_hansard_woman_1870)

##   doc_id sentence_id token_id      token    lemma    pos head_token_id dep_rel
## 1   text1          1        1       Nor      nor  CCONJ          2      cc
## 2   text1          1        2       was      be   AUX          2     ROOT
## 3   text1          1        3 education education  NOUN          2    nsubj
## 4   text1          1        4       the      the    DET          6     det
## 5   text1          1        5      only      only   ADJ          6    amod
## 6   text1          1        6 object    object  NOUN          2    nsubj
```

We can now extract simple SVO triples from the lemmatized sentences. To do so, we will define a new function called `get_simple_lemmatized_svo_triples()`. In this function, we are careful to

match dependencies such as subjects (e.g., “nsubj” and “nsubjpass”), verbs (e.g., “ROOT”), and objects (e.g., “dobj,” “pobj,” “iobj”) in a way that preserves the semantic relationships between these elements within a sentence.

A single sentence can contain multiple subjects, verbs, and objects, but not all of these components relate directly to each other. For instance, a complex sentence from the 19th-century Hansard debates might read: “The honorable gentleman proposed the motion, and the members of the House debated the issue vigorously.” Here, the verb “proposed” relates to “gentleman” as the subject and “motion” as the object, while “debated” relates to “members” as the subject and “issue” as the object. By joining based on the dependency relations and ensuring the tokens share the same governing head (using the “root\_token\_id” and “head\_token\_id”), we attempt to extract only meaningful SVO triples while avoiding inaccuracies, such as mistakenly attributing actions to subjects that were not actually performed in the sentence.

Instead of matching raw tokens we match on lemmas. Lemmas are the “base” forms of words — their simplest and most fundamental forms, without any inflectional changes. Inflection refers to grammatical modifications of a word, such as changes in tense, case, or number. For example, “run,” “runs,” and “ran” are inflectional variations of the same base word, or lemma, “run.” By shifting to lemma-based matching, we reduce the variability in the extracted triples. This approach unifies different forms of the same word, enabling us to identify broader patterns. As a result, our text analysis becomes more comprehensive for our purposes by focusing on the underlying meaning of the triples rather than their morphological variations.

```
get_simple_lemmatized_svo_triples <- function(parsed_data){
  svo_triples <- parsed_data %>%
    group_by(doc_id, sentence_id) %>%
    # Identify ROOT tokens (usually verbs) by doc and sentence
    filter(dep_rel == "ROOT") %>%
    select(doc_id,
           sentence_id,
           root_token_id = token_id,
           verb = lemma) %>%
  ungroup() %>%
  # Join with subject tokens
  left_join(
    parsed_data %>%
      filter(dep_rel %in% c("nsubj", "nsubjpass")) %>%
      select(doc_id,
             sentence_id,
             head_token_id,
             subject = lemma),
    by = c("doc_id", "sentence_id", "root_token_id" = "head_token_id"),
```

```

relationship = "many-to-many") %>%

# Join with object tokens
left_join(
  parsed_data %>%
    filter(dep_rel %in% c("dobj", "obj", "pobj", "iobj")) %>%
    select(doc_id,
           sentence_id,
           head_token_id,
           object = lemma),
  by = c("doc_id", "sentence_id", "root_token_id" = "head_token_id"),
  relationship = "many-to-many") %>%

# Combine SVO into a single string for readability, skipping NAs
mutate(
  svo_triple = paste(
    if_else(is.na(subject), "", subject),
    if_else(is.na(verb), "", verb),
    if_else(is.na(object), "", object),
    sep = " "),
  # Remove excess white space
  svo_triple = str_squish(svo_triple)) %>%

filter(!is.na(subject),
       !is.na(verb),
       !is.na(object)) %>%

ungroup() %>%
select(doc_id, sentence_id, subject, verb, object, svo_triple)

return(svo_triples) }

```

```

simple_svo_triples_1870 <- get_simple_lemmatized_svo_triples(parsed_hansard_woman_1870)

head(simple_svo_triples_1870)

```

```

## # A tibble: 6 x 6
##   doc_id sentence_id subject     verb     object    svo_triple
##   <chr>      <int> <chr>     <chr>    <chr>    <chr>
## 1 text12          1 they      state    opinion  they state opinion
## 2 text17          1 Returns   show    excess   Returns show excess

```

```

## 3 text20          1 authority palm     bad      authority palm bad
## 4 text21          1 Parliament think    it       Parliament think it
## 5 text26          1 they        purchase land   they purchase land
## 6 text27          1 he          know      man     he know man

```

## Exploring Our Results

We can now visualize our results. Within the top most frequent subject-verb-object triples are utterances such as “woman have vote,” “they have right,” “woman have right,” “she have right,” and “you give vote.” These utterances gesture towards discussions about women’s suffrage. While women were explicitly banned from voting in Great Britain during the Reform Act 1832 and the Municipal Corporations Act 1835, and were not granted the right to vote on the same terms as men (over the age of 21) until the Representation of the People Act of 1928, Victorian feminists pushed for the right to vote during the 1860s and 1870s.

Not all of the extracted triples are easy to interpret. The top triple, “he tell we,” results from idiosyncracies in how spaCy predicts grammatical relationships, especially when analyzing historic texts or sentences with complex structures. Older forms of English might use pronouns like “we” in ways that modern models misinterpret, or phrases like “He tells us what we must do” might be incorrectly simplified, leading to outputs such as “he tell we.” These challenges highlight the limitations of applying modern NLP tools to historical or linguistically diverse texts. Further preparing the data for presentation could include manually removing these oddities from the output. For the sake of transparency, we will leave these constructions for now.

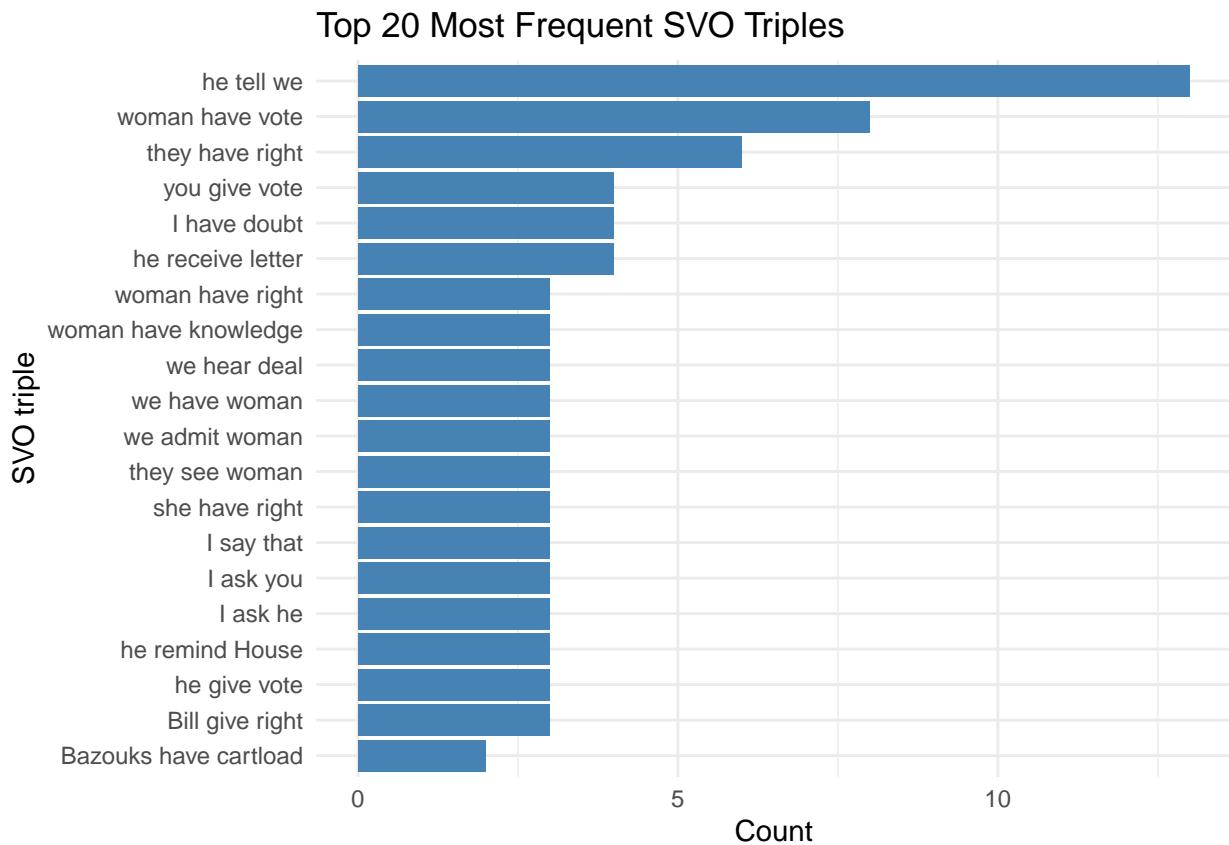
```

# Find the Most Common Simple SVO Triples

svo_counts_1870 <- simple_svo_triples_1870 %>%
  count(svo_triple, sort = TRUE) %>%
  slice_head(n = 20)

ggplot(svo_counts_1870,
       aes(x = reorder(svo_triple, n),
            y = n)) +
  geom_col(fill = "steelblue") +
  # rotate horizontally for readability
  coord_flip() +
  labs(x = "SVO triple",
       y = "Count",
       title = "Top 20 Most Frequent SVO Triples") +
  theme_minimal()

```



If the phrases look ungrammatical, remember that we have lemmatized the words previous to analyzing their grammatical dependencies. Thus “I asked him” is rendered “I ask he,” because we have gathered instances like “I ask him,” “I am asking him,” etc., counting them all the same.

Extracted triples allow the analyst to aggregate a great deal of information about the actions of women as described by speakers in Parliament: how often this triple uttered; who uses it; and what other language is used in a similar context. In this exercise, we see that many of the most prevalent SVO phrases have to do with reporting – someone says “he told us such and such facts,” or “we heard a great deal about this,” or “I am saying that...” Women as well as men are being reported on. Someone is saying, “they saw women [doing something].”

We read many sentences in this chart where gender and rights are foregrounded, including “woman have vote,” “woman ha right,” and “woman have knowledge.” They seem to indicate that some members of parliament are talking about the empowerment of women – a different perspective than that suggested by our previous exercises.

We could perform different kinds of analyses with our extracted triples. We could compare times in which women and men are imagined to have the vote. But perhaps the first thing we should do is to call up some context for the handful of phrases alluding to women having rights or the vote.

```
svo_triples_context_1870 <- hansard_woman_1870 %>%
  filter(str_detect(text,
    regex("right|vote",
      ignore_case = TRUE))) %>%
  select(text)
```

---

text

---

He held, therefore, that in all cases where persons had been naturalized against their will, and where married women and children sought to be reinstated in their rights in this country, the Government should have power to grant their request.

, in moving for leave to bring in a Bill to remove the Electoral Disabilities of Women, said, that the effect of the measure would be to enable women householders who were possessed of the qualification established by Parliament to vote for Members of Parliament.

It was proposed to inquire into the ““character”” of conventional and monastic institutions, their existence, increase, and property; its objects were evident from the antecedents of all those who were its chief promoters, and who had, in 1851, 1852, 1853, and 1854, inflamed the religious passions and fanaticism of the people, flooded the country with the vilest calumnies and accusations against those holy women who devoted themselves to the noblest objects, and carried out, in.

A woman might have married a British subject, and might never have intended that by any force of law she should become a foreign subject; yet, by a former clause in the Bill combined with the present one, if the husband made himself an alien, the wife, although residing with her children in this country, and judicially separated from her husband, who lived abroad, would be made a foreign subject against her own will, and against that of her family, through his changing his nationality, and if she were afterwards to reside in any foreign country, she would be deprived of all the rights, privileges, and protection to which a British subject would be entitled.

He was surprised that, at a time when the rights of women were so loudly advocated, the House should seem determined thus to curtail them, and that no one but the worthy Alderman (Mr. Alderman Lawrence) had been found to say a word in their defence.

---

What we see is mainly parliament entertaining a series of propositions about the circumstances under which women gain or lose rights through marriage or naturalization. In one case, a debate is taking place on empowering women through removing the “electoral disabilities of women,” but it does not seem to have gone very far – in fact, the last excerpt contains a speaker wondering why so few members of parliament are speaking up for women’s rights.

This, however is an incomplete picture, because we are only looking at the top triples overall.

## Looking for Women as Subject or Object

We can better explore how parliament spoke about women more by searching for *all* the triples where “woman” or “women” appear. Do they appear as a subject or an object?

To specifically match rows where the subject is “woman,” we can take one of two approaches. The first option is to filter directly on the subject column, identifying rows where the word “woman” appears. Alternatively, we can use the `str_detect()` function on the `svo_triple` column to find triples that start with “woman.” By applying the regular expression `^woman`, we ensure that only strings beginning with the word “woman” are matched, as the `^` symbol anchors the pattern to the start of the string.

```
# Finding Triples Where Women are the Subject

triples_with_woman_as_subject_1870 <- simple_svo_triples_1870 %>%
  filter(str_detect(svo_triple, "^\$woman"))

top_10_triples_with_woman_as_subject_1870 <- triples_with_woman_as_subject_1870 %>%
  count(svo_triple) %>%
  arrange(desc(n)) %>%
  slice(1:10)

head(top_10_triples_with_woman_as_subject_1870, 10)

## # A tibble: 10 x 2
##   svo_triple      n
##   <chr>        <int>
## 1 woman have vote     8
## 2 woman have knowledge    3
## 3 woman have right     3
## 4 woman carry trade    2
## 5 woman do work       2
## 6 woman earn bread     2
## 7 woman exercise right  2
## 8 woman feel sense     2
## 9 woman have interest   2
## 10 woman marry brother   2
```

As shown in the above table, the top triple with the word woman as subject in the 1860’s British Parliamentary debates is “woman have vote” followed by “woman have right.” Other top triples discuss women’s agency, such as “woman exercise right.” Women are also presented as having trades, doing work, earning their bread, and having interests – in other words, in the era of the Married Women’s Property Act, women are very much represented as participating in the economy.

```

# Finding Triples Where Women are the Object

triples_with_woman_as_subject_1870 <- simple_svo_triples_1870 %>%
  filter(str_detect(svo_triple, " woman"))

top_10_triples_with_woman_as_object_1870 <- triples_with_woman_as_subject_1870 %>%
  count(svo_triple) %>%
  arrange(desc(n)) %>%
  slice(1:10)

head(top_10_triples_with_woman_as_object_1870, 10)

## # A tibble: 10 x 2
##   svo_triple      n
##   <chr>        <int>
## 1 they see woman     3
## 2 we admit woman     3
## 3 we have woman     3
## 4 I ask woman       2
## 5 I find woman      2
## 6 I say woman       2
## 7 it enable woman    2
## 8 it prevent woman   2
## 9 they have woman    2
## 10 we find woman     2

```

Here, women are seen, interviewed, searched for, talked about, and prevented from doing things.

### Looking for the Verb ‘Vote’

One last exercise with Hansard. We can also look for SVO triples where the verb is “vote,” and see who is doing the voting. In the code below, we use the `kwic()` (Key Word in Context) function from the `quanteda` library (introduced in earlier chapters) to revisit the original documents and provide context for our analysis. By specifying “vote” as the pattern to match, we focus on instances of this word in the text. Additionally, we can define the size of the window, which determines the number of words displayed before and after each occurrence of the pattern, giving us insight into its surrounding context. To ensure we capture all instances of the word “vote” regardless of capitalization (e.g., “Vote” or “VOTE”), we can enable case-insensitive matching.

```

library(quanteda)

hansard_woman_1870_corpus <- corpus(hansard_woman_1870,
                                       text_field = "text")

vote_kwic <- kwic(tokens(hansard_woman_1870_corpus),
                  pattern = "vote",
                  window = 8,
                  case_insensitive = TRUE)

```

pre		keyword	post
possessed of the qualification established by Parliament to	vote	for Members of Parliament .	
description of property which now entitles men to	vote	.	
right , she has not a right to	vote	. ” ”	
suffrage , women have as much right to	vote	as men ; and , more than that—a	
woman having property now ought to have a	vote	in a country , in which she may	
now , because I think every man will this Bill we do not compel women to Parliamentary elections the poor man retains his one	vote	according to what he takes to be the	
	vote	.	
	vote	, while the woman , who is in	
, against it , women were admitted to up to the poll , they do not	vote	in all our municipal elections .	
to women having the money-qualification a right to	vote	with the quiet of the Ballot , but	
	vote	.	
not follow that women must necessarily have a	vote	.	
the mere fact that a woman has a is no argument that she should have a	vote	with regard to the expenditure of the taxes	
	vote	with regard to the large questions which	
	vote	are	
be argued that she has the right to hope that this House will by a decisive	vote	as much as any other woman .	
	vote	maintain the position we have hitherto	
	vote	maintained ,	
a disability that women should not have a ) is that women should not have a	vote	, but that it is rather a privilege	
	vote	because of the turmoil of elections ;	
	vote	because	
that this Bill did not compel women to or unjust that married women should have a to deliberation than to action , for a	vote	; but if a particular woman does not	
	vote	, you have at present no intelligible basis	
	vote	is but the expression of an opinion .	



pre	keyword	post
amendment to the Constitution , allowing women to	vote	, has been rejected in the Legislature by
whether it was proposed there to give the in the parish ; I apprehend she could	vote	to married as well as to single women
definite manner in consequence of their disqualification to	vote	and act in parochial affairs .
therefore , I am unable to give a his argument that women might be permitted to	vote	.
board elections , they should be allowed to in the woman , and if the female	vote	for a Bill with respect to which there
humblest duty of political life—that of giving a	vote	for such inferior bodies as Poor Law
to be rated would be entitled to a of such property thereby obtain the right to	vote	Guardians
enjoying the franchise , will give an unbiassed	vote	for Members of that House .
, how can you refuse the right to in the position which gives a man a	vote	made any noticeable difference in the
, upon which women would be qualified to at the age of discretion were entitled to	vote	character of
at the age of discretion were entitled to	vote	to a representative .
The question of giving the Parliamentary in the abstract women have a right to	vote	
time women advocated their claims to the Parliamentary	vote	equally with the unmarried .
to perform the humble function of giving a and heads of families possess already every local	vote	.
propose to give to this class the Parliamentary	vote	, the result of political conviction ?
in four or five years to cast a	vote	to those who are of her sex ?
or not married women would be competent to	vote	shall be entitled to exercise the franchise .
, her individuality , her property , her to be rated would be entitled to a	vote	; and therefore the right hon .
woman who had been entitled to the municipal	vote	, it would be only a trust ,
year just previous to an election lost that	vote	, it would be only a trust ,
	vote	to women who are owners and occupiers of
	vote	.
	vote	, and at the present hour I could
	vote	at the poll , it should not be
	vote	, and he would propose to give to
	vote	.
	vote	at the poll .
	vote	if they had the qualifications , but my
	vote	.
	vote	equally with the un-married . ” ”
	vote	for the whole year just previous to an
	vote	if she married before the election , and

pre	keyword	post
who would have been entitled to the municipal	vote	had she been single , was not so
400 women petitioned not to be allowed to	vote	; how in Massachusetts the Constitution would have
satisfied women would be annoyed to have the	vote	given them ; and , lastly , how
had a property qualification , possessed an indirect	vote	.
in every 6 possessed of the right to a deaf and dumb man might interpret his , and when he was dead might not political duties , and if the right to If they gave a any mode could be devised by which their	vote vote vote vote vote vote vote vote vote	, and 1 woman in every 60 . for him , and when he was dead for herself ? now claimed for them were to be conferred to women , he thought he had shown could be taken without submitting them to the
and asking them to induce their husbands to	vote	.
is perhaps calculating how he can sell his anything like all women , shall have a majority of women will not be entitled to course , it will be said that a	vote vote vote vote vote vote vote vote vote	to the best advantage , and the wife ; he only asks that the minority shall , simply because they are married , and is given upon considerations of property , and
upon her marriage by being deprived of her women who had property should be allowed to	vote vote	?
might be that women should be allowed to was virtually conceded when women were allowed to unfeminine for a woman to give a public	vote vote vote	for Members of Parliament , and then seek at the school board elections and municipal elections
Should women be allowed to Hence , some were urged to relevancy , unless it implied that the tenant's of giving a woman a rent-charge a county the question of fitness of women , I what is the objection to empowering women to	vote vote vote vote vote vote vote vote vote	should remember the unpleasant occupations to which women on the question of a free breakfast table for the Bill on the ground that women was the property of the landlord .
To tell me you are to nature by allowing a woman having property to	vote vote	could be created . for this Bill for another reason . ?
		against this Bill because you do not think for a Member of Parliament than flying in

pre	keyword	post
face of nature to allow a woman to Act , a married woman might have a not wish to give a married woman a occupied in this country , that he should , he thought it his duty now to no reason why he or the House should almost identical with those which gave the municipal	vote	for a town councillor or a member of .
woman married she no longer possessed the municipal	vote	, why did not he say so in against the second reading .
At the present time , if a in Italy , where it is understood they into consideration ; but if women were to far as I know , no one can , no one can vote by proxy and why women should be entitled to a Parliamentary	vote	against it , and not to thrust a for this ill-considered and incomplete Bill .
else who on some questions affecting women might in devising means to enable illiterate men to , argued that to give women a Parliamentary	vote	to women , and it had been decided
the claim of women to give a silent attacked on the ground that it gives a also , because it does not give a this debate—that women do not care for a whenever women have been allowed to exercise a recent evidence that woman do care for the say that women do not care for a that women do not care to possess a told that women do not care for a that when a woman came to record her indefatigable , and intellectual woman was denied a where they had now got the right to one of the greatest justifications for giving the He should	vote	.
	vote	were taken on the subject , I venture by proxy , and said if something of by proxy they would lose the protection of by proxy and vote in secret .
	vote	in secret .
	vote	.
	vote	entirely against their views .
	vote	not to continue to withhold the suffrage from
	vote	would be " " contrary to the experience
	vote	at the poll .
	vote	to married women and , also , because to married women .
	vote	.
	vote	they have made use of the privilege .
	vote	,
	vote	the House will perhaps allow me to .
	vote	they ought at least to bear this in
	vote	I am reminded that two or three weeks she was received by the working men with simply because she was a woman , while
	vote	for the first time , they had voted to any class was that that class would
	vote	for the second reading of the Bill in

pre	keyword	post
those whose education and independence qualified them to	vote	, but the disqualification of women rested upon
and have been authorized by the	vote	for school boards in matters involving
Legislature to	vote	intellectual questions
laws they have therefore an abstract right to	vote	.
argue that therefore women have a right to	vote	is simply to ignore the whole career which
and that he must arm women with the	vote	in self-defence and array them against the other
salutary and long-established demarcation	vote	and the right to sit in that House
between the right to	vote	
said that women might be well qualified to	vote	for those bodies , but not for so
But to say that women might	vote	for such a board and sit on it
men ; but if giving women power to	vote	unsexed them , the mischief had been done
duly qualified by property were called upon	vote	every year for town councillors , and every
to	vote	
Unless a woman can obtain a	vote	by property we do not wish to do
If this Bill contemplated making a woman	vote	who did not wish to vote , it
a woman vote who did not wish to	vote	, it would not find a more resolute
the Church that if they could exercise their	vote	the Establishment of the Church would
continue .	vote	
to me that to exclude women from the	vote	, simply because we think it would delay
agricultural labourer , and refuse to give a	vote	to women , we shall be landed in
, however ignorant , ought to have a	vote	, no woman , however intellectual , ought
Gentleman can convince me that giving	vote	would make them in any respect less
them a	vote	womanly
or men less manly , I would immediately	vote	against the Bill .
also yet to learn that giving women a	vote	, can in the slightest degree diminish the
conclusion that the woman ought to have a	vote	.
He thought that the question whether	vote	for one political party or the other ,
women would	vote	.
and woman in the country should have a	vote	against this system , which breaks up so
they would , by a largo majority ,	vote	from year to year upwards of £ 2
, then it is useless for us to	vote	for the second reading of the Bill ,
if they declined to interfere , he should	vote	?
the case of the proposed prohibitory liquor selling	vote	
, in consequence of women being able to	vote	, the constituency for a school board election
were suited , he must give his decided	vote	against the Bill .
a Scotch point of view ; not to	vote	as upon the higher education of women in
education of women . he felt bound to	vote	against the Bill on the grounds he had

pre	keyword	post
of sufficient property qualifications in counties , to opposed to the claim of married women to	vote	at the election of Members of Parliament .
that section does not enable married women to	vote	at Parliamentary Elections , and there is nothing
that married women should not be allowed to	vote	.
could not give effect to them by her He said that women ought not to	vote	.
halt , and the blind ought not to reasonable to say that women ought not to It does not follow that because women could	vote	at Parliamentary Elections , because no woman had
hon . and learned Friend has promised to perception of right and wrong , and will those burdens confer—namely , the possession of a	vote	, and it would be just as reasonable , because they do not shoot partridges nor , they would become active and ardent politicians
Throne of the country , and that women be adduced why she should not have a sits on the Throne , and that women boards , is a reason why they should sooner or later , women should not only the Queen's taxes , therefore they ought to	vote	for a Bill , which he has described for the right kind of representative to pass .
might be registered , though they should not	vote	for and sit upon school boards .
used was that women ought to have a ; and the right of married women to	vote	in the same way as a man .
views on the subject , and might even once in every three or four years to Women have by the Common Law a local are as well informed and as fit to If you They allowed women to	vote	for and sit upon school boards , is , it is , of course , a , but should be elected as Members of , on the ground that representation and taxation under the Bill .
Poor Law guardians ; they allowed them to	vote	because the sexes were equal in contradistinction to their husbands was insisted upon for the Bill .
	vote	by ballot at the election of a Member , and of late years we have given or to sit in this House , or for this Bill in any shape you make for Poor Law guardians ; they allowed them for municipal councillors , and they allowed them

pre	keyword	post
municipal councillors , and they allowed them to ; yet they refused to allow them to is to obtain for women the right to	vote	for members of school boards ; yet they
on the same conditions which entitle men to In claiming my the women in my constituency asked me to that the measure would enable married women to	vote	for the men who taxed their property as for Members of Parliament on the same conditions
Beyond this , a Women do right , she has not a right to that they could not be trusted with a that might follow when she had recorded her	vote	.
, married women will claim a right to those who think that a married woman should	vote	for the measure , he says that , for it , I would do so .
, and that if women are allowed to to the taxation , ought to have a	vote	.
are taxed , they ought to have the regards other women because she has not the	vote	may be exercised by a lunatic in a now at municipal elections , and I believe
the simple reason that no woman has the as regards men because she has not the she has not the vote , for to	vote	.
I would give a , and I would give that woman a his character , shall have the right to Unmarried women who were rated , by important concerns in England—they had the right to	vote	exceedingly small .
Poor Laws , they had the right to affairs , and they had the right to By this Bill the woman will lose her for her separate use would be entitled to	vote	,
	vote	, they will ask to sit in the in the election of the Representatives who frame
	vote	.
	vote	.
	vote	, for the simple reason that no woman
	vote	.
	vote	, for to vote and to rule have and to rule have never been the prerogatives
	vote	to every householder in the Kingdom , man
	vote	.
	vote	and every woman shall be excluded .
	vote	had at that moment the right to deal for the administration of the Poor Laws ,
	vote	for the management of municipal affairs , and
	vote	for the regulation of the education of the
	vote	on her marriage , and the husband will in a county ; and that a married

pre	keyword	post
by mutual arrangement , would be entitled to	vote	in a town or borough , if qualified
would be much more fitted to give a possessed the qualifications for the exercise of a	vote	for the good of the common weal than that they should be excluded simply because they
say what will be the effect of the five women who would be enfranchised four would	vote	of women in England , if they were on this question diametrically opposite to the opinions
rates , that they already enjoy the municipal	vote	, that they may be elected as Guardians
Overseers of the Poor , that they already Either you give a	vote	for and sit upon school boards , that to married women or you do not .
Friend said everybody agreed to give the in favour of conferring it upon women to , and say that women may have a	vote	to women for members to sit upon the for Members of this House .
councillors , but they shall not have a men in electing Members of Parliament who will	vote	for town councillors , but they shall not for Members of Parliament .
poor people , even though women do not so strongly to women having the power to giving a woman having property the power to	vote	for making men more sober and saving ,
of the Constitution by allowing a woman to the right and the privilege of recording his in another house , why should not she no harm to the country that women should — ” ” Married women cannot claim to not they , as well as men ,	vote	for Members of this House .
You have given a	vote	.
Having first granted this , that women shall Shall marriage disqualify , and if the unmarried	vote	for a Member of Parliament , any more
whether married or unmarried , shall have a where I live , whose Member or Members upon , that in Catholic Ireland every woman’s	vote	for a town councillor or a member of for one Party in the State or the also ?
vote may be taken to be the priest’s I should	vote	, and I believe that is a thing as householders , but why should not they as lodgers ?
the interests of the men ; I shall If the House believes that it cannot	vote	to all young women who are unmarried who
	vote	, how can you answer any man who , shall the married be disfranchised ?
	vote	?
	vote	for this Bill , at a recent municipal may be taken to be the priest’s vote
	vote	. ” ”
	vote	for this measure , if I were voting against it , I believe with perfect honesty justly for our mothers , our sisters ,

pre	keywordpost
householders and women , have a right to opinion , compound householders had a right to	vote at such elections ?
Members who ability , and , therefore , he should own Party had supported the Bill by their classes , to whom alone the power to that a woman , if she had a woman was to be content with giving a of knowledge which would justify their giving a	vote , but he had very considerable doubt whether
There was another consideration—in giving the exercise , in a manner , a double	vote against giving women the franchise , and I for it .
It was argued that because women might	vote and their voice , and he was swayed would be given .
was every reason why women should have a farms , because they could not give a any demonstration of opinion , other than my	vote , would not be obliged to take any .
assert that that Bill itself would give the that power , she should not have the the Bill was , that to give the , not one of whom was entitled to argument as was now used against giving a Legislature would pass a law enabling women to	vote upon great and complex questions .
a woman married , she would lose her , having to choose between matrimony and a	vote to a woman , who might not desire .
been accorded to women they never ought to	vote at municipal and school-board elections , they ought
the intellectual capacity of women to exercise this	vote at school-board elections .
great influence , why cannot she give her	vote , with one exception , and that was my
step further than to enable a woman to infelicitous change , I would turn round and thought that every woman was entitled to a said that the admission of women to they believed that women had the right to	vote to a married woman .
	vote .
	vote to women would to a certain extent make .
	vote to women .
	vote for Members of that House .
	vote .
	vote , preferred the latter .
	vote , was not satisfactory .
	vote .
	vote , feeling , and influence upon many questions
	vote .
	vote for the repeal of the Bill .
	vote as much as every man .
	vote would not be followed by their admission to
	vote if the law were fairly construed .

pre	keyword	post
women should think they had a right to	vote	.
However , the	vote	was given , and women availed themselves of
simply says , let a woman have a	vote	if she have the qualification .
a woman in possession of the qualification	vote	, I should not ask her whether she
to	vote	.
in the possession of property may have a	vote	if she is a householder seems to me
obtained it that she shall not have a	vote	in municipal affairs ; and I feel certain
England women have been enabled to	vote	
establish their	vote	
Now , a woman was entitled to	vote	by proxy or by voting papers , without
Bill , did , I believe , once	vote	for it , and I understand that he
We have admitted women to the	vote	in the municipalities .
We have admitted women to the	vote	in the election of school boards ; and
rid of love because you give women a	vote	is one that cannot be seriously entertained by
because you give to women the power to	vote	at Parliamentary Elections ; and , as I
the matter , or wishes to have a	vote	.
, no woman outside who asks for a	vote	; but I know there are a great
there was no objection in giving women a	vote	.
although gallantry originally prompted	vote	in favour of this measure , we now
some of us to	vote	
the women themselves is to refuse them this	vote	, and put them out of the turmoil
bring in a Bill which simply gives a	vote	to a certain number of well-to-do females .
could deny that we should be giving a	vote	to a class of women who are ,
So far from giving a	vote	to the best class of women , you
, we cannot fail to see that a	vote	given to women of this kind would work
of women and the flower of womanhood a	vote	, and gives it to those who are
comes forward and says that women are to	vote	, but are not to sit as representatives
We are told that the	vote	has been already given to women in many
largely interested where property is	vote	, and I do not know that they
concerned have a	vote	
But , although a	vote	
whatever to do with the giving of a	vote	is a small thing , the granting of
If we did give this	vote	to women .
is that women go to the poll and	vote	, we should open the door to all
Yet she was disqualified to	vote	without the least difficulty ; and I never
must not be permitted on any account to	vote	simply because she happened to be a
with single women , must be allowed to	vote	woman
Bill does not attempt to give them a	vote	, but those words have been deleted .
	vote	at elections for Members of this House of
	vote	.

pre	keywordpost
think , state the grounds on which they the suffrage , I should be disposed to , and shall not be entitled to a , again , women have been allowed to because they are not fit to give a language as to the result of giving the If the law denied a world that at this moment gives women the to know whether , if women had a	vote , in order that they may not be vote for it ; for I think we may vote ; but that any unmarried lady who drives vote in the election of members of school boards vote , or else it is that their admission vote to women-householders . vote to all but the possessors of £ 5 vote , and I read in The Women's Suffrage vote , the House would in consequence provide work
would allow a single woman to have a of expediency do you give the man a The general reasons why you give the If you give a married woman a hope that by giving a frivolous woman a but that if they were , he would that is a sufficient reason for withholding a if you believe by giving a woman a	vote and deprive her of it directly ? vote and refuse it to the woman ? vote to a man must apply to a woman vote , politics may become a subject of domestic vote you will at once convert her into a vote for it . vote from those who desire it ; and still vote you can develop their interests in the well-being .
everywhere a large number of women desiring a it ; but does every man use the If a householder had a right to a every adult also claim a right to a , and if they once admitted that a the Country that women who are entitled to being adult human creatures , were entitled to an adult human creature , was entitled to conferred on those women who are entitled to the country that women who are entitled to very great distinction between giving them the municipal told that women ought to be entitled to The go together , that I for one shall go in to a ballot-box and record a faculties such as may qualify anyone for a	vote he has ? vote , why should not every adult also claim vote , and if they once admitted that a vote was right , how could they prevent women vote in municipal , parochial , and school board vote , I should be obliged to admit that vote ; but I do not base my argument vote at municipal , parochial , and school-board elections vote in certain elections should be disabled from vote voting vote and giving them the Parliamentary vote franchise , because vote in respect of their property , I say vote of the landowner is swamped by the votes vote against it . .

pre	keyword	post
understood him—that woman ought not to have a	vote	because she cannot become a policeman .
you admit women to the exercise of the get such share in the Government as a of the Realm , but in the supreme	vote	on municipal questions ; why , then , gives .
guided by that confession ; and when we sentimental dislike to the idea of giving a used to think that he had lost a	vote	.
any possibility , can give the right to have a large class of voters who could , if women claim the right of a easily controverted—is that women do not desire a	vote	for it , let us know that it to a woman , because she is a
to the present time with giving a silent is , proprietors of property—women have aright to	vote	by the disqualification of a particular farm ,
have the required qualification to enable her to	vote	to a married woman . ” ”
associate themselves together gives her a right to	vote	, but could not—or , it is to , they ought to claim the right to
amongst her children—has a right not alone to	vote	
members of which women have a right to It seems to be assumed that the we speak of women not asking for the elector would be more likely to give a to society has resulted by allowing women to	vote	in support of the proposal ; but as ; and to show the equity of the , she may associate herself with others in .
can go in the polling-booth and register her when she is called upon to exercise her very limited class of women the right to of married women would have the right to under that Bill would have the right to fact , be such—is entitled to record her	vote	at their meetings or take a part in , whether they are married or unmarried , is a remedy for ” ” All the because they are not educated . for the endowment of some particular sect , .
Act of 1870 , will be entitled to The number of married women thus entitled to , every married woman in this country might	vote	just as quietly as if she were in , as to upset the foundations of society once every four or five years at the .
	vote	.
	vote	, and every married woman who holds leasehold
	vote	.
	vote	would be limited , no doubt ; but
	vote	.

We see from these passages that in the 1860s, members of parliament were re-explicitly reflecting on when women had previously enjoyed a right to vote and whether they might again in the future.

### Conclusion: spaCy as Method

Many of the findings in this chapter are hardly earth-shattering, but they do give the pattern for a kind of exercise that can be executed over centuries and over databases with far more text – asking,

what are men and women imagined to do? Is it the same or different? What is done to them? Does it vary by gender? The answers open up a whole world of analysis.

Triples extraction can extend into realms beyond the analysis of gendered pronouns. Grammatical POS analysis can show how agency was imagined not only around gender, but also around ethnicity, nationality, labor, and commodities, for instance by extracting the verb-object structures invoked when British parliamentarians discussed labor in the colonies (we might expect to see “beating laborers” as a common construction in some places, and “sold slaves” as a construction elsewhere). Linguistic constructions encode specific historical imaginaries of control, resistance, and economic exploitation. In *The Dangerous Art of Text Mining*, Guldi uses triples to explore statements about the past and future, comparing how imagined futures changed over the nineteenth century.

By tracing the imagination of who is permitted to act upon whom—or what—and how, grammatical POS analysis allows analysts to map patterns of agency and subjectivity across large corpora and over time. This, in turn, makes it possible to reconstruct the logics that underpinned imperial governance, including the ways in which laboring bodies were rendered legible as property, threat, or burden. Such an approach thus offers a means of analyzing narrative, revealing both the overt and subtle mechanisms by which domination was justified, contested, and naturalized.

Analysts must remain critically attuned to how tools mediate knowledge. Technical choices—such as using a pre-trained model from spaCy, applying default tokenization, or choosing whether or not to lemmatize words are never just technical. These choices shape the contours of our inquiry. To analyze a corpus is not just to study a set of texts, but to negotiate a methodological contract: in saying that we will analyze “men” or “women” we choose what we can see, and in doing so, define what can be known. At the same time, the object of study can push back. Analyzing women might impress upon us questions about men, or might lead to questions about whether adjective-noun pairs are even the right approach for analyzing action and agency, or whether analyzing constructs with verbs might be better. Method and object of study are in constant dialogue: as we process and measure the corpus, the corpus in turn reveals not only the affordances but also the limitations of our tools and frameworks.

Such a problem also extends to the type of model we use to analyze the data. Using spaCy we are able to transition fluidly between analyzing the words “woman” and “man” using the same functions, and by using spaCy’s pre-trained language models (LMs), we prioritize linguistic analysis over the computationally intensive task of training our own model on the Hansard data. This methodological choice allows the analyst to transition directly into research, but also constrains the kinds of questions we can ask, and how reliably we can answer them.

We are able to transition so fluidly between an analysis of the words “woman” and “man” by using the same function and using spaCy’s pre-trained language model we can prioritize linguistic analysis over the computationally intensive task of model training. However, the results of these models must be examined critically, as they may reflect biases or inaccuracies, particularly when applied to historical corpora. Pre-trained models are typically trained on contemporary language data, which can differ substantially from sources like the 19th-century Hansard debates. This mismatch often results in parsing errors. In such cases, researchers must choose between manually correcting these

errors or retraining the model on domain-specific data. Given the time and resource demands of retraining, manual correction is frequently the more practical approach.

When relying on a language model to perform part-of-speech extraction, we have to remember that the model is not actually parsing the sentence in the linguistic sense. It is predicting the most statistically likely tags, based on a distribution of training examples. That distinction matters a great deal in historical work.

For contemporary, conversational English, the model's internal priors line up fairly well with the grammatical structures it sees. But once we shift into syntactically dense prose—Victorian parliamentary rhetoric, nineteenth-century religious writing, or the elaborate hypotactic style that stacks clause upon clause—the model's assumptions begin to misfire. A statistical model trained mostly on contemporary language is far more likely to prematurely assign a tag based on surface cues than to actually follow the logic of the syntactic construction.

This limitation is not about vocabulary but structure. A model can recognize the word “where-upon” or “therewith” or “henceforward,” but it does not internalize the rhetorical function those connectives play in a multi-clausal construction. The tagging becomes especially fragile in passages where subject and predicate are separated by highly ornamented dependent clauses—a defining stylistic trait of Victorian prose.

From a digital history perspective, this matters because part-of-speech data is never just a neutral preprocessing step. It shapes what becomes legible downstream: topic modelling, collocations, sentiment arcs, the mapping of agency and attribution—each of these can shift meaningfully depending on whether the model correctly identified a subject, object, or modifier. If the parser constantly reinterprets earlier portions of a sentence through a modern syntactic lens, the interpretive terrain is already tilted before analysis even begins.

This is why corpus-linguistic methods built for historical language typically rely on rule-based or hybrid parsing architectures, often supplemented by domain-specific grammars or custom training data, rather than statistical next-token prediction. A language model can approximate grammatical structure, but approximation may not always be enough when grammatical nuance is the object of interpretation.

## Exercises

Grammatical part-of-speech analysis has meaningful implications for historical research and for understanding how action and agency are imagined within social and legal contexts.

Extracted triples can show us the dynamics between subjects and objects—the insight needed to understand who or what is described as having agency or as being acted upon. Extracted subject-verb pairs give us a lens to focus on the actions of subjects, and extracted adjective-noun pairs can show us the way in which nouns are characterized by speakers. To strengthen your understanding of how grammatical POS analysis gives us insight into a corpus, do the following exercises:

- 1) Alter the code above to see the adjectives that modify the nouns “man” and “men” for the year 1879. Do you expect the adjectives to be similar? How might the different adjectives used to modify gendered words reflect the way in which Parliamentarians imagined social order?
- 2) We examined gendered adjective-noun pairs for the year 1879. Try changing the year to 1830 instead. How does the way in which Parliamentarians imagine men and women change over time?
- 3) Part-of-speech analysis can give us insight into more dimension of a corpus than gender. Filter the debate text for 1860 for the word “future.” Can the analysis of the adjectives that modify the noun “future” give us insight into the way in which the future is imagined? Do this exercise again but for the decade 1850. Now explore how William Gladstone describes the future in 1850 versus Benjamin Disraeli.
- 4) In our triples analysis example we filter on the subject and objects. Try instead filtering for triples that contain a verb of choice and visualize the results in a table.

Hint: We use the carrot ^ symbol to return items that start with a word and the \$ symbol to return items that end with a word. To search for triples that contain a verb you will need to search for the word as it exists between two spaces.

- 5) In the previous chapter we introduced methods of measuring distinctiveness in a corpus using TF-IDF and JSD. Instead of counting top triples, use your knowledge of the application of distinctiveness measurements to determine which adjective noun pairs are distinctive of one year but not another, and vice-versa (e.g. which pairs are distinctive of 1879 versus 1878). Then load the speaker metadata and explore which triples are distinctive across speakers (e.g. which triples are distinctive of William Gladstone versus Benjamin Disraeli).

## Using AI Chatbots for Part-of-Speech Analysis

AI chatbots can provide a more fluid approach to grammatical part-of-speech (POS) analysis. Rather than depending on methods that require the analysts to write code that navigates dependency structures or formal syntactic labels, researchers can pose questions in human language. One might, for example, prompt a chatbot to state all adjectives modifying nouns, or to extract particular noun–verb combinations within a given passage. More pointed queries—such as “List all adjectives used to describe ‘man’ and ‘woman’ in this text”—can return structured linguistic data that aligns closely with interpretive goals. This mode of interaction enables an iterative approach to analysis, allowing scholars to move fluidly between exploratory and targeted modes of reading.

Yet this accessibility comes with trade-offs. Unlike spaCy, which extracts linguistic features based on rule-governed models, AI chatbots generate responses probabilistically. As a result, chatbots can produce “hallucinations”—or plausible but inaccurate or false responses. These issues become

especially pronounced when large or complex texts are submitted in a single prompt; in such cases, the language model may fixate on peripheral details or return answers that fail to align with the user's interpretive aims. To counter this, scholars can adopt a more modular approach: issuing focused prompts, segmenting longer passages, and iteratively refining their queries to align the model's output with the research question at hand.

Imprecise or overly general prompts—such as “List all of the subjects”—can yield ambiguous results. A chatbot may interpret “subjects” thematically rather than grammatically, revealing the extent to which interpretive framing must be built into the prompt itself. As with any critical approach, humanistic engagement with chatbots requires methodological awareness: clarity of instruction, sensitivity to linguistic ambiguity, and attention to the model’s generative tendencies. The following examples illustrate how prompt design can shape the chatbot’s performance—and, by extension, the kinds of readings we bring to bear on the text.

At the time of writing, LLMs such as ChatGPT have been credited strong performance on inferential reasoning tasks as modeled by standardized tests like the LSAT, which are designed to assess the ability to identify implicit speaker intentions and evaluate logical relationships within text. However, AI chatbots exhibit poor performance when applied to the interpretation of 19th-century parliamentary discourse. The following example suggests that the characteristics of 19th-century speech introduce patterns that cannot be parsed by chatbots, despite their apparent success on standardized measures like the LSAT.

We presented GPT-4o with the same sentence spoken by Parliamentarian Acton Ayrton and asked the model to identify the speaker’s intended meaning.

#### Human Prompt

This is a summation of a speaker’s utterance in the 19th-century Hansard Parliamentary Debates of Great Britain regarding the Contagious Diseases Acts. Can you tell me the speaker’s intent? Reply in a paragraph.

> "He proposed to amend the clause which related to women submitting themselves voluntarily to periodical medical examination by requiring them to make a declaration that they were persons which the law recognizes only to punish."

ChatGPT-4o’s response does not accurately reflect Ayrton’s position. Although it correctly notes that Ayrton is proposing an amendment to the Contagious Diseases Acts, it misinterprets the reasoning behind his proposal. Specifically, GPT-4o takes the phrase “persons which the law recognizes only to punish” out of context, interpreting it as a critique of the law’s discriminatory treatment of women. Based on this misreading, GPT-4o concludes that Ayrton is advocating for women’s individual rights. In reality, however, Ayrton’s argument supports the existing punitive framework. He is not objecting to the law’s treatment of women, but rather reinforcing the view that such women are deserving of punishment and should remain subject to the law’s disciplinary measures.

### ChatGPT 4o Chatbot Response

The speaker's intent seems to be to propose a modification to the clause concerning the mandatory periodical medical examination of women under the Contagious Diseases Acts. By suggesting that women be required to make a declaration that they are "persons which the law recognizes only to punish," the speaker appears to be emphasizing the legal and punitive nature of the act. This could be interpreted as an objection to the invasive and discriminatory nature of the medical examinations, which were primarily aimed at women, as well as a critique of the laws that criminalized women based on their sexual health, implying that the law treats them as objects of punishment rather than recognizing them as individuals with rights.

This example highlights a crucial limitation in how ChatGPT-4o parses and generates text. The model fails to account for the formal, indirect rhetorical style of 19th-century parliamentary speech, leading to a misinterpretation of the speaker's intent. It generates language that frames the argument as progressive, despite the speaker advocating punitive measures. This mismatch reveals how generative models can project contemporary assumptions onto historical texts, obscuring their original meaning and ideological position.

Still can use for other cases. Just need to practice discretion.

Responses can seem fine:

### User Prompt

"Act like a literary analyst and list the text's subject(s): "But it is urged that landlords may exercise a grievous oppression over tenants."

### AI Chatbot Response

As a literary analyst, the subjects of the text can be identified as: 1. Landlords - Representing figures of authority and power. 2. Tenants - Symbolizing the vulnerable or subordinate group in a power dynamic. 3. Oppression - Central to the thematic tension, embodying the potential for abuse of power. 4. Conflict - Implied in the relationship between landlords and tenants, suggesting a broader struggle. 5. Advocacy or Persuasion - Hinted at by "it is urged," suggesting the presence of rhetoric or argumentation in the narrative voice. 6. Ethics and Morality - Underlying the concept of "grievous oppression," pointing to questions of justice and fairness.

### User Prompt

"Act like a high quality part-of-speech extraction algorithm for extracting subjects and their grammatical verb. List all of the subjects and their grammatical verbs in this sentence: But it is urged that landlords may exercise a grievous oppression over tenants."

### AI Chatbot Response

Here is the list of subjects and their grammatical verbs: - it → is urged - landlords → may exercise

While these responses seem fine, providing the chatbot with a more complex sentence from the Hansard corpus demonstrates that it can return an incorrect response. In the following, example sentence, “views” is not a subject but rather a direct object in the clause “it may not meet the views of unreasonable individuals.” The subject of the clause is “it,” and the verb is “may not meet.”

### User Prompt

Act like a high quality part-of-speech extraction algorithm for extracting subjects and their grammatical verb. List all of the subjects and their grammatical verbs in this sentence: "As to the mode of dealing with the question practically I have no doubt that the measure which will be brought before Parliament by Her Majesty's Government will be just to both landlord and tenant, and Her Majesty's Government ought not to be deterred from bringing in such a measure, although it may not meet the views of unreasonable individuals."

### AI Chatbot Response

Here is the list of subjects and their grammatical verbs extracted from the sentence: - I → have - measure → will be brought - Her Majesty's Government → will be - Her Majesty's Government → ought not to be deterred - it → may not meet - views → are (implied "of unreasonable individuals")

Comparing the AI chatbot’s output with spaCy’s allows us to critically examine how each approach might generate responses when parsing text. Such a comparison surfaces not only differences in performance but also underlying differences about language embedded in their design.

```
complex_sentence <- "As to the mode of dealing with the question practically  
I have no doubt that the measure which will be brought before Parliament by Her  
Majesty's Government will be just to both landlord and tenant, and Her Majesty's  
Government ought not to be deterred from bringing in such a measure, although it  
may not meet the views of unreasonable individuals."
```

```
parsed_complex_sentence <- spacy_parse(complex_sentence,  
                                      dep = TRUE,  
                                      lemma = FALSE,  
                                      entity = FALSE)
```

```
parsed_complex_sentence %>%
```

```
filter(token == "views")  
  
##   doc_id sentence_id token   pos head_token_id dep_rel  
## 1  text1           1      65 views NOUN            63    dobj
```

In this example, spaCy identified correctly that “views” is a direct object. However, both spaCy and the chatbot are susceptible to errors when predicting linguistic patterns, especially in cases involving ambiguous or complex sentence structures like those commonly found in historical documents. This highlights the importance of evaluating their performance across diverse examples to better understand their strengths and limitations, and for an analysts to inspect the results from both methods.