

# **Rilevamento e analisi multimediale di meme misogini presenti in Rete: l'influenza del contenuto testuale**

**Relatore:** Francesca Gasparini  
**Co-relatore:** Aurora Saibene

**Tesi di Laurea Magistrale di:**  
Annalisa Bachir

Anno Accademico 2019-2020

# Il problema

Le manifestazioni di sessismo e misoginia online prendono nuove forme e sviluppano nuove strategie di **ostilità nei confronti delle donne.**

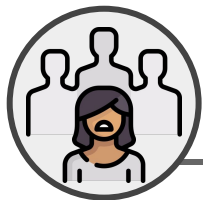
# Scopo della tesi

Studiare i fenomeni di *cyber sessismo* e *misoginia online* attraverso l'**analisi testuale dei meme.**

# Sessismo e misoginia online

---

Le principali **motivazioni** delle ostilità:



Volontà di **escludere** le donne da alcuni ambienti online



Ascesa di movimenti maschili **anti-femministi**



Influenza di una nuova **destra**



# I meme di Internet

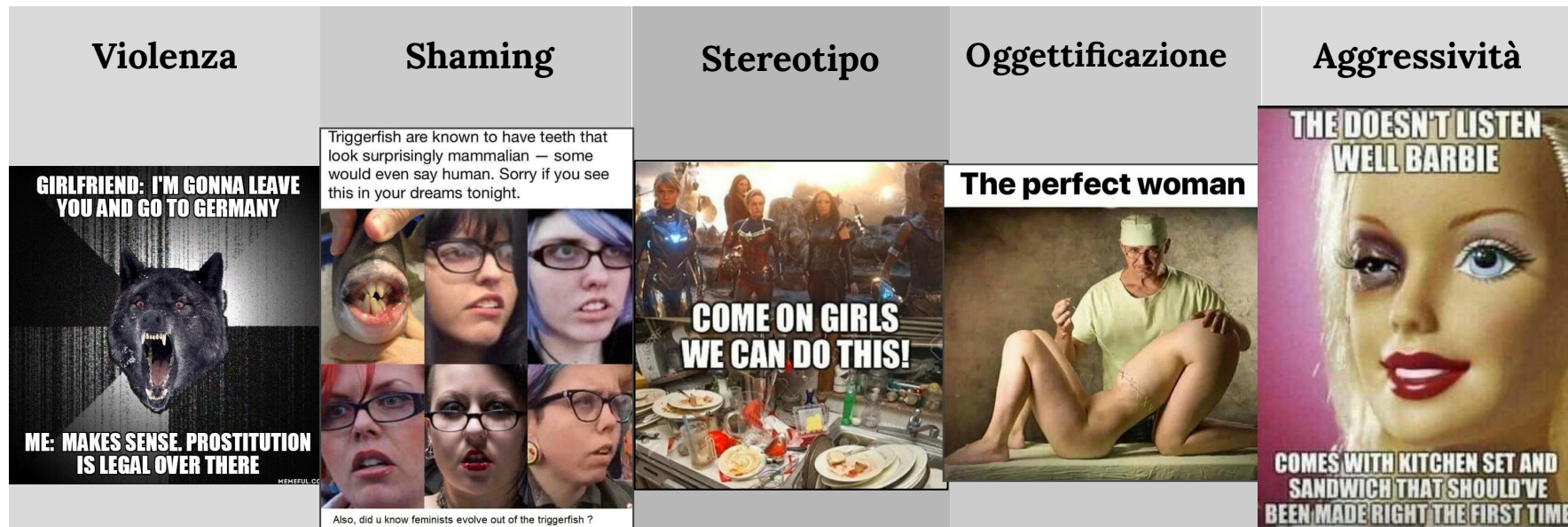
Imitazione, interpretazione e trasmissione.

Contenuto pittorico + testo + umorismo o ironia



# I meme misogini

Definizione di **meme misogino**: meme che mostra una scena offensiva, sessista o legata all'odio verso le donne.



**1**

**FASE PRELIMINARE**

# La creazione del campione

---

1



## Definizioni

Formulazione delle definizioni di meme, di meme misogino e di meme misogino aggressivo

2



## Ricerca delle parole chiave

Le parole chiave sono state utili per trovare risorse adatte a questo studio

3



## Download dei dataset

Download delle immagini da due siti:  
Reddit e 9GAG

4



## Ricerca dei doppi





È stato usato un software per la ricerca di immagini simili al fine di eliminarle

5



## Divisione dei dataset

Il dataset di immagini è stato diviso in base alle definizioni

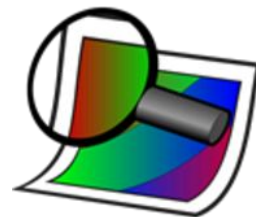
	Name	Size	Date	Dimensions
	zb7d27jb67g41.jpg	57,794	03/06/20 07:29:28	556 x 500
	9nqv7jc52n31.jpg	109,148	03/06/20 07:30:14	900 x 820
	zqn9ytlpux31.jpg	107,154	03/06/20 07:29:16	900 x 820
	dwvc0u4e2ce41.jpg	37,273	03/06/20 07:29:16	686 x 678

4



## Ricerca dei doppioni

È stato usato un software per la ricerca di immagini simili al fine di eliminarle



### Geeqie

Individua le immagini in base al livello di similarità pittorica.



# La creazione del campione

---

1



## Definizioni

Formulazione delle definizioni di meme, di meme misogino e di meme misogino aggressivo

2



## Ricerca delle parole chiave

Le parole chiave sono state utili per trovare risorse adatte a questo studio

3



## Download dei dataset

Download delle immagini da due siti:  
Reddit e 9GAG

4



## Ricerca dei doppi

È stato usato un software per la ricerca di immagini simili al fine di eliminarle

5



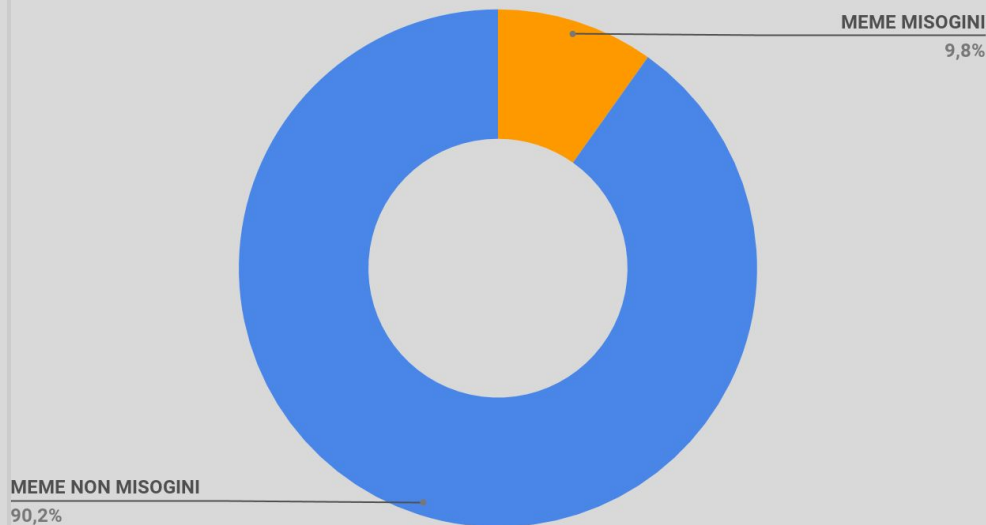
## Divisione dei dataset

Il dataset di immagini è stato diviso in base alle definizioni

# Il campione

Immagini analizzate	19.097
Meme non misogini	5.232
<b>Meme misogini</b>	<b>570</b>
Meme misogini aggressivi	85

Percentuale dei meme misogini e dei meme non misogini



2

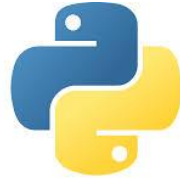
## **ESTRAZIONE E GESTIONE DEL TESTO**

# Gli strumenti utilizzati

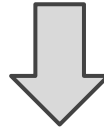
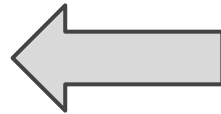
---



Preprocessing  
delle immagini



python™



Tesseract OCR  
Riconoscimento ed  
estrazione del testo

**NLTK**  
Natural Language Toolkit

Manipolazioni sul  
testo

# Estrazione del testo

---

Preprocessing delle immagini

Manipolazione delle immagini con nove **tecniche di preprocessing** delle immagini.

Estrazione testo con OCR

Uso di un sistema di Optical Character Recognition (OCR)  
Prove con differenti **metodi di segmentazione della pagina**.

Output salvati in un file csv

Nel csv vengono salvati l'identificativo delle immagini e i testi estratti con le varie tecniche di *preprocessing* e i vari metodi di segmentazione.

# Estrazione del testo

---

Preprocessing delle immagini

Manipolazione delle immagini con nove **tecniche di preprocessing** delle immagini.

Estrazione testo con OCR

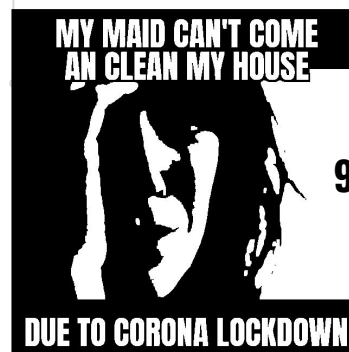
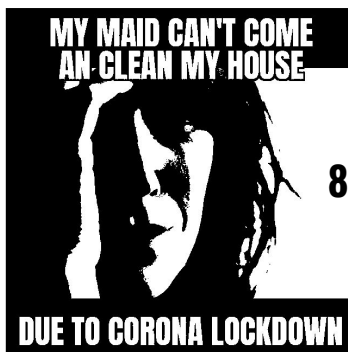
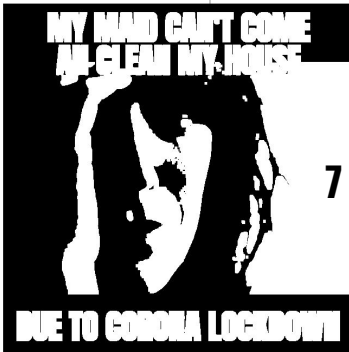
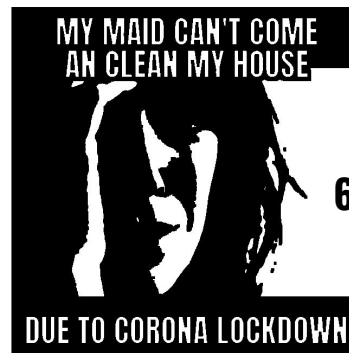
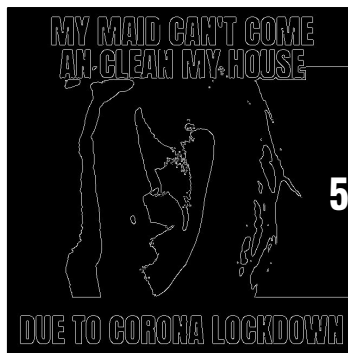
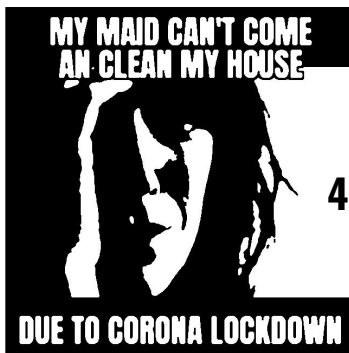
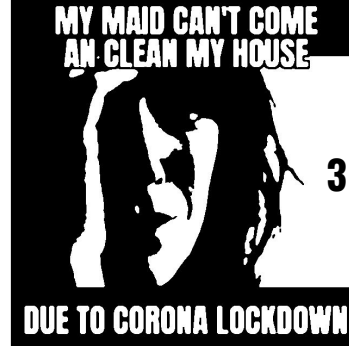
Uso di un sistema di Optical Character Recognition (OCR)  
Prove con differenti **metodi di segmentazione della pagina**.

Output salvati in un file csv

Nel csv vengono salvati l'identificativo delle immagini e i testi estratti con le varie tecniche di *preprocessing* e i vari metodi di segmentazione.

## Preprocessing delle immagini

1. Rescaling
2. Scala di grigi
3. Thresholding (T.)
4. T. con rimozione del rumore
5. T. con riconoscimento dei contorni
6. T. con erosione
7. T. con dilatazione
8. T. con skew correction
9. T. con opening



# Estrazione del testo

---

Preprocessing delle immagini

Manipolazione delle immagini con nove **tecniche di preprocessing** delle immagini.

Estrazione testo con OCR

Uso di un sistema di Optical Character Recognition (OCR)  
Prove con differenti **metodi di segmentazione della pagina**.

Output salvati in un file csv

Nel csv vengono salvati l'identificativo delle immagini e i testi estratti con le varie tecniche di *preprocessing* e i vari metodi di segmentazione.



# Estrazione del testo

---

Preprocessing delle immagini

Manipolazione delle immagini con nove **tecniche di preprocessing** delle immagini.

Estrazione testo con OCR

Uso di un sistema di Optical Character Recognition (OCR)  
Prove con differenti **metodi di segmentazione della pagina**.

Output salvati in un file csv

Nel csv vengono salvati l'identificativo delle immagini e i testi estratti con le varie tecniche di *preprocessing* e i vari metodi di segmentazione.

# L'output testuale



Esempio di output:

**Helping**

â€TM Lp -

= hyo? =o Ns

gn Sr A WS JE

he

[\$ Tm > AES

**Homeless women**

Il testo risulta  
sporco, ricco di  
caratteri  
speciali, lettere  
straniere e  
simboli

# Pulizia del testo

---

- Rimozione degli spazi ripetuti e dei segni di interpunzione
- Riduzione dei testi in minuscolo
- Rimozione di caratteri non alfanumerici
- *Tokenization*
- Rimozione delle *stopword*

Per ciascuna immagine sono stati estratti **27 testi differenti** usando le nove tecniche di preprocessing per ciascuno dei tre metodi di segmentazione.

È stato necessario selezionare i testi “migliori”.

I testi sono stati selezionati in base alla versione che conteneva il maggior numero di parole appartenenti al dizionario in lingua inglese.

Il programma si occupa anche di indagare quale tecnica di *preprocessing* e quale metodo di segmentazione della pagina hanno dato complessivamente gli *output* più soddisfacenti.

I risultati migliori si sono ottenuti con le seguenti combinazioni:

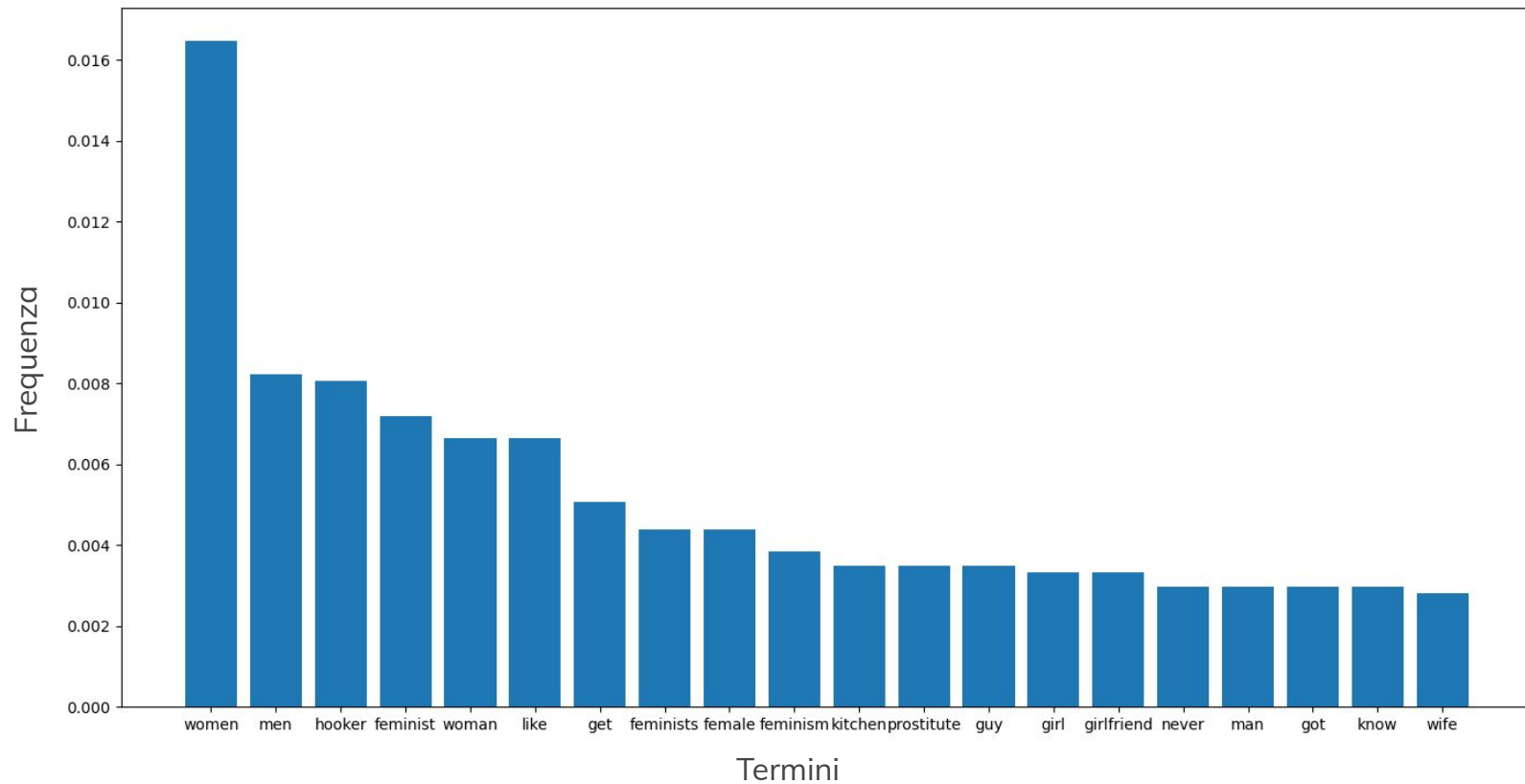
1. ***tresholding*** + metodo di segmentazione 6 → 2926 parole
2. solo *rescaling* + metodo di segmentazione 6 → 2916 parole
3. *skew correction* + metodo di segmentazione 6 → 2907 parole

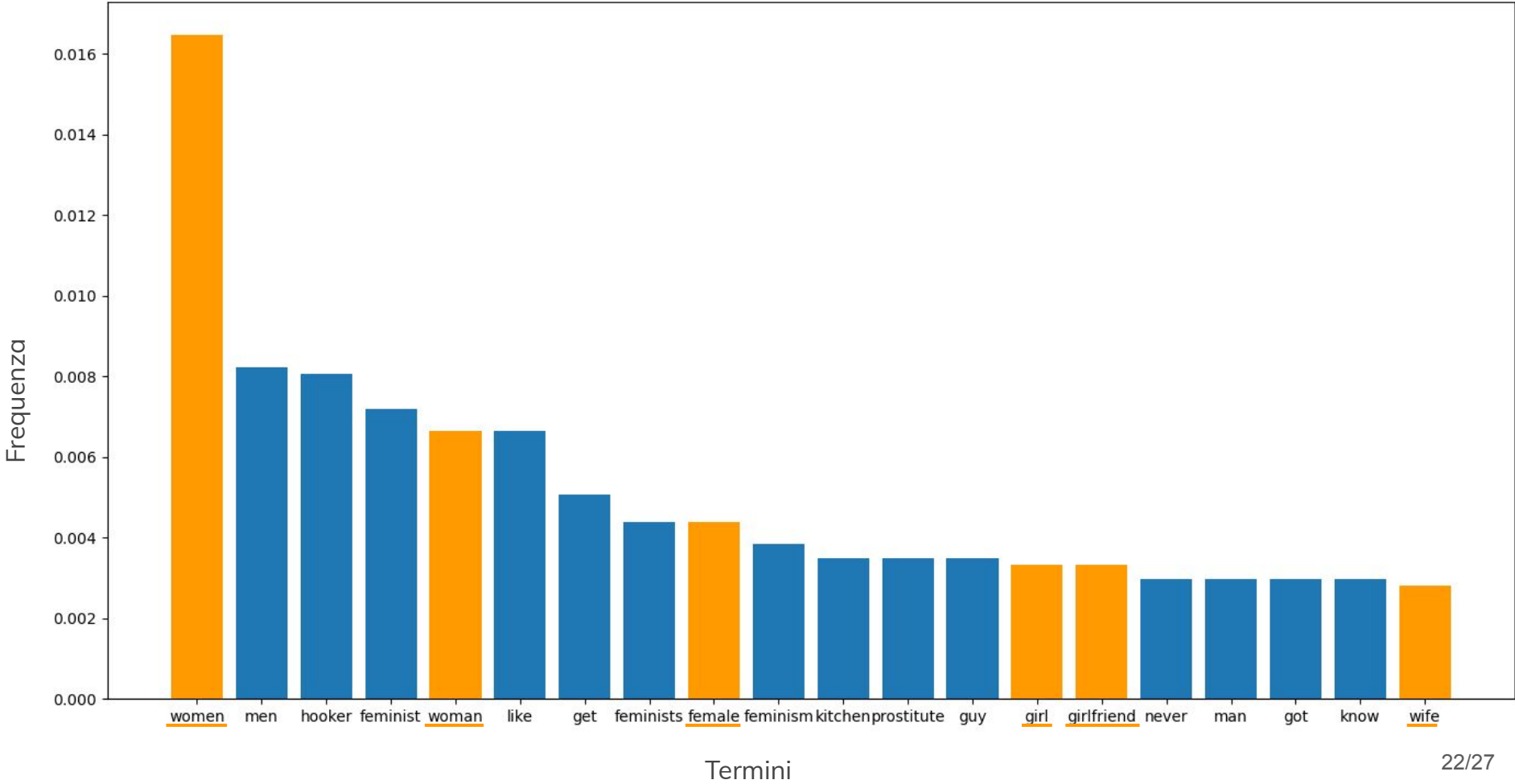
# **Analisi delle frequenze**

---

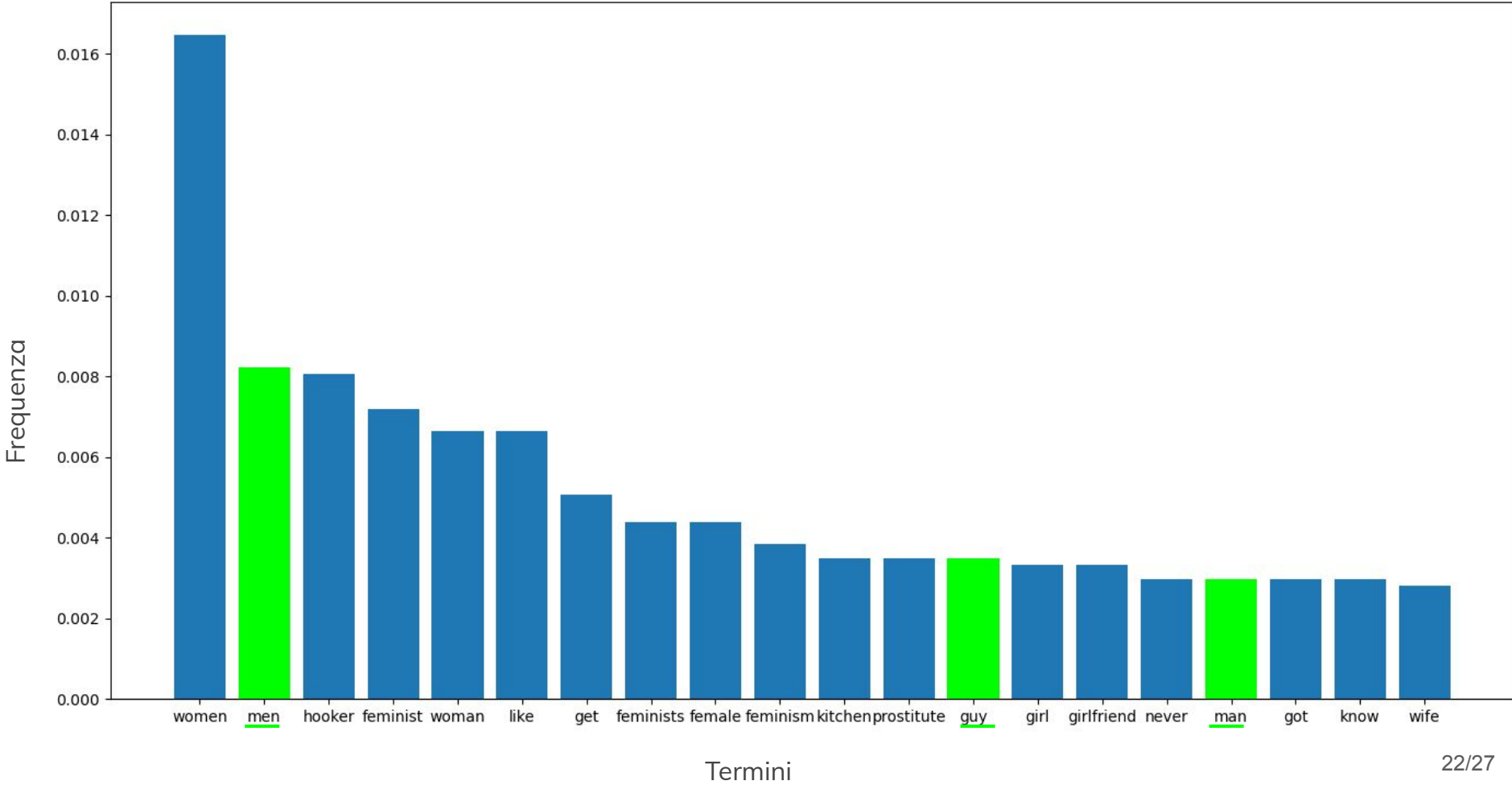
L'analisi delle frequenze rileva le parole presenti in maggior numero all'interno del *corpus*.

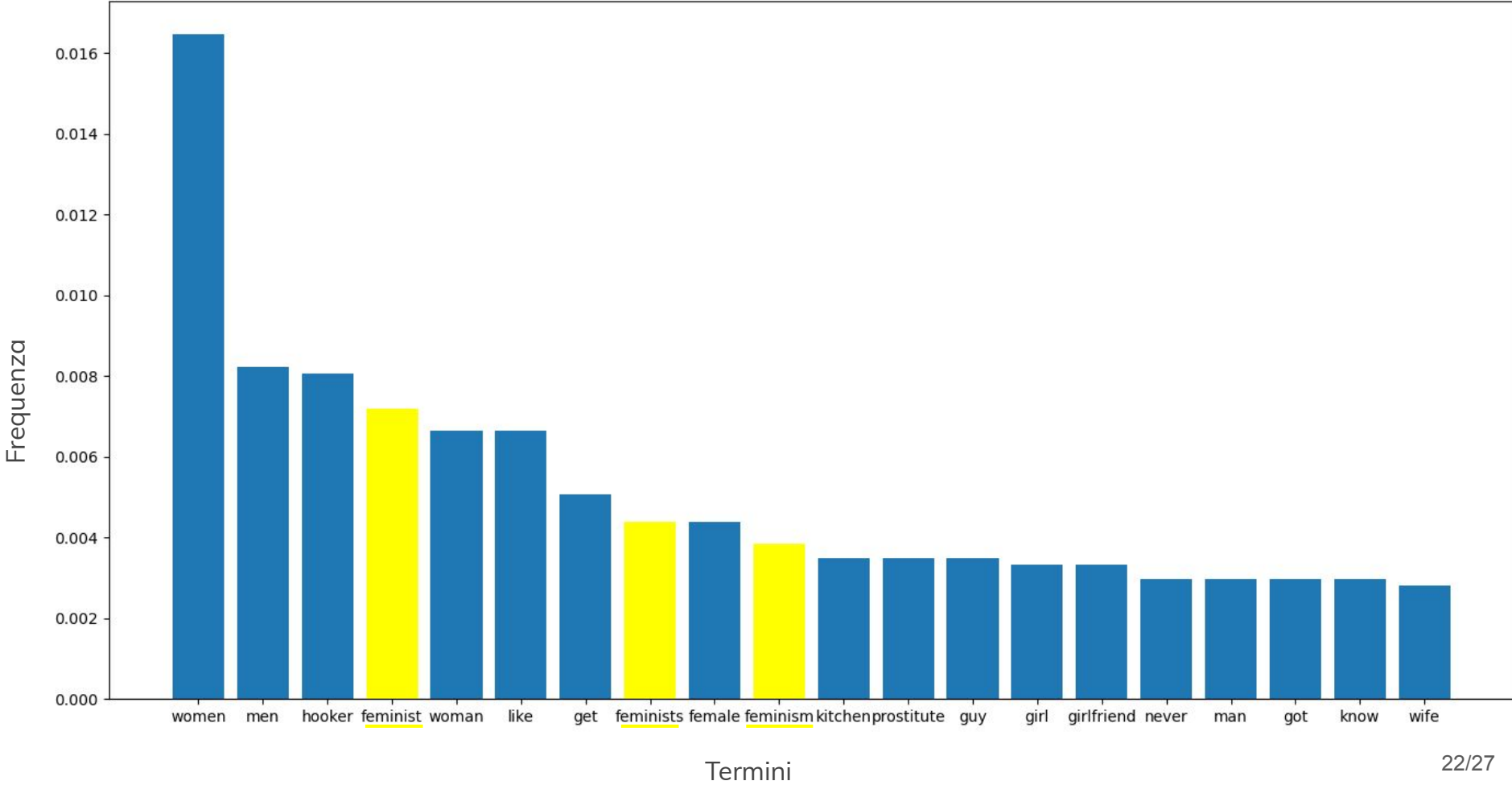
# Il grafico

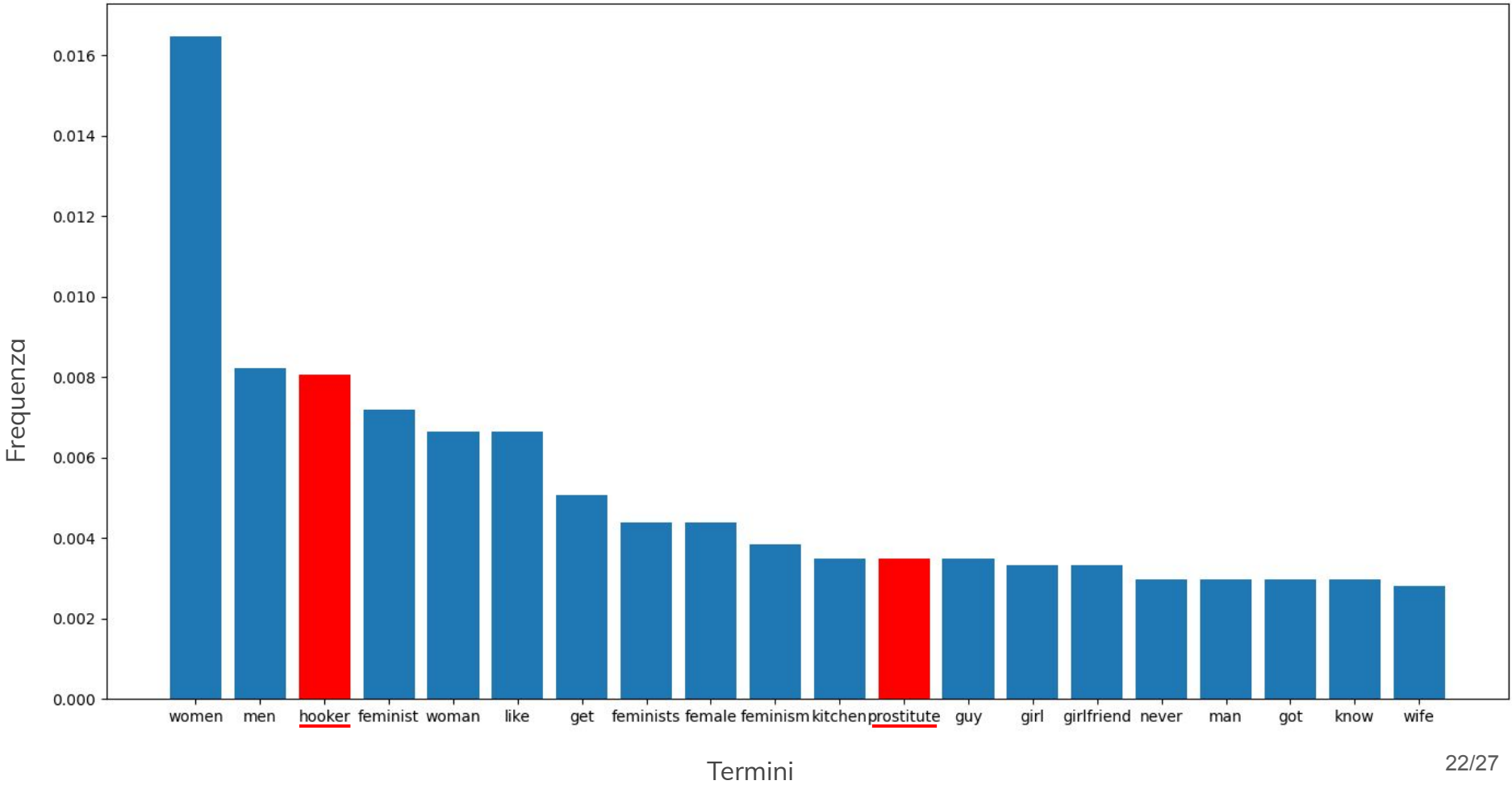


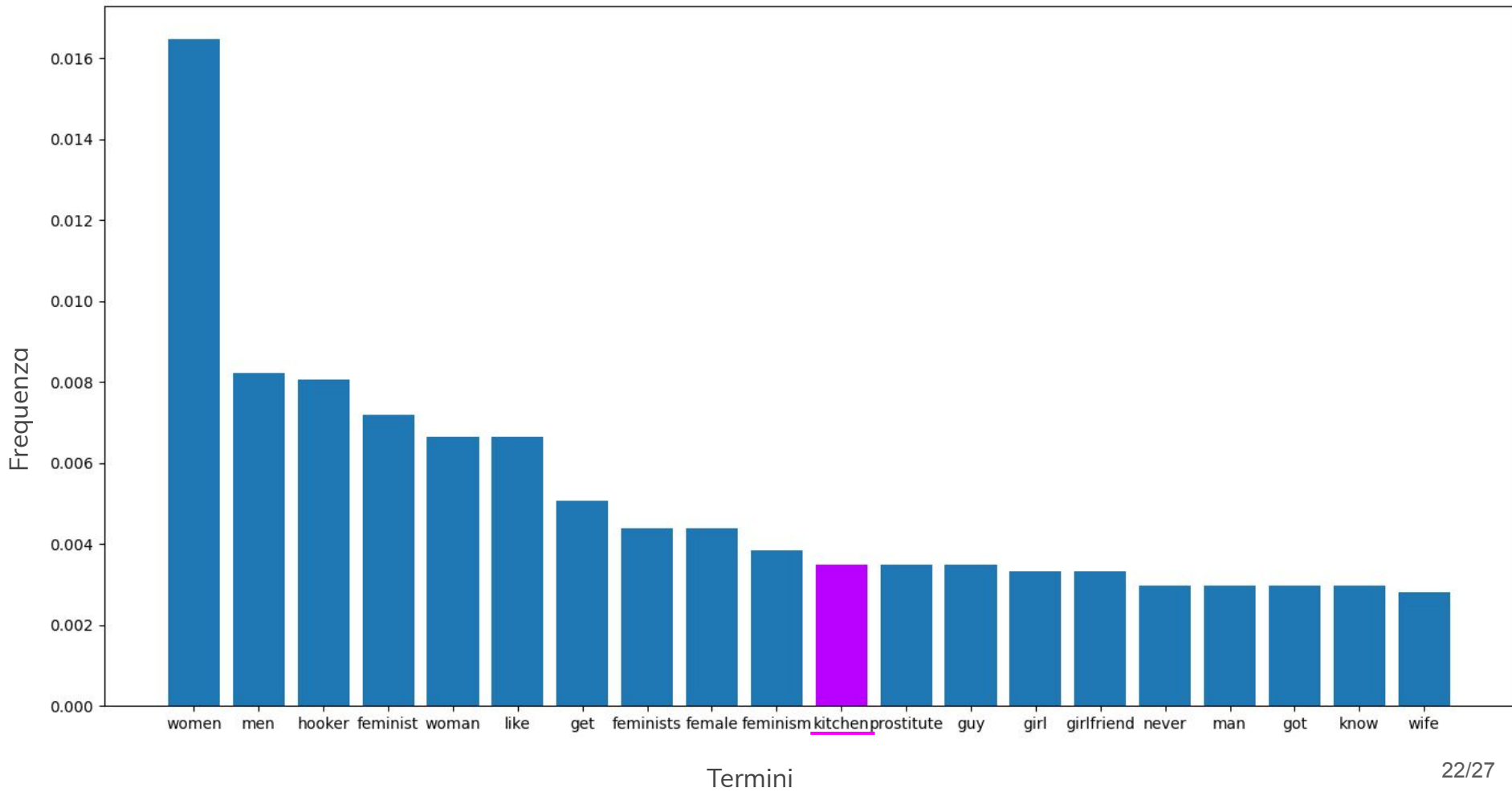




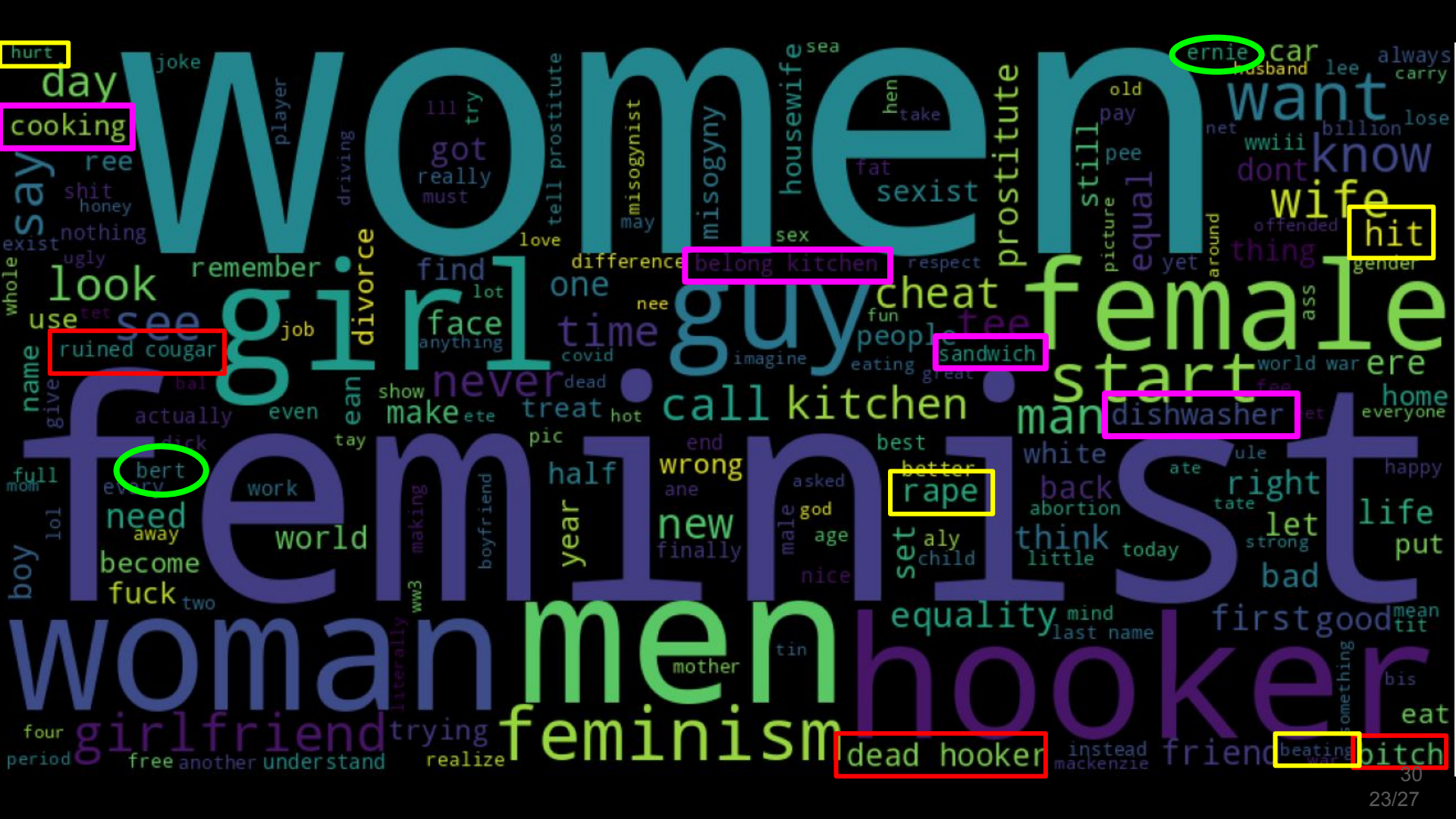












hurt

cooking

ruined cougar

bert

belong kitchen

sandwich

dishwasher

rape

dead hooker

beating

bitch

# **Analisi sulla vicinanza semantica**

---

Il presupposto teorico della vicinanza semantica sostiene che parole fisicamente vicine tra loro abbiano una **relazione semantica** più stretta tra loro rispetto a parole lontane.

# Analisi sulla vicinanza semantica

<b>Feminist</b>	Testi meme misogini	woman, yes, sea, like, want, file, <b>wwiii</b> , man, <b>rape</b> , abortion, today, people, understand, wife, <b>mad</b>
	Testi meme non misogini	<b>world</b> , 1st, feminists, standards, rights, <b>rape</b> , toxic, teaches, touch, real, slap, thinks, <b>killing</b> , actually, <b>3rd</b> , understand, assume
<b>Rape</b>	Testi meme misogini	change, understand, say, yes, shit, sexual, nice, nothing, start, crisis, really, <b>feminist</b>
	Testi meme non misogini	toxic, touch, standards, teaches, rights, killing, slap, thinks, sae, <b>feminists</b> , <b>3rd</b> , sentence, assume, ario, hate, understand, real, <b>1st</b> , word

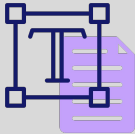


# Spunti per le future ricerche

---



**Preprocessing delle immagini:** considerare altre tecniche e combinazioni di tecniche di *preprocessing*



**Manipolazione del testo:** correggere le parole, integrare operazioni come lo *stemming* e la *lemmatization*.



**Analisi del *sentiment*** per individuare i meme misogini in modo da segnalare in tempi rapidi il contenuto come violento o offensivo.

