



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in Linguistica
Classe LM-39

Tesi di Laurea

Deep learning e contenuti misogini: analisi linguistica dell'errore nei compiti di riconoscimento di misoginia nei meme

Relatrice
Dott.ssa Erica Biagetti

Correlatrice
Prof.ssa Elisabetta Fersini

Laureanda
Silia Claudia Pastori
n° matr. 2014723/ LMLIN

Anno Accademico 2022 / 2023

Indice

Introduzione	4
Capitolo 1: Il fenomeno della misoginia offline e online, e i meme come mezzo di comunicazione	6
1.1 Il fenomeno della misoginia offline da una prospettiva storica	6
1.2 La misoginia online	8
1.3 I meme come artefatti culturali e i meme di internet	11
1.4 Caratteristiche linguistiche dei meme	14
1.5 Lavori precedenti sulla misoginia nei <i>social media</i>	15
1.6 Presentazione del <i>task</i> MAMI di SemEval-2022	17
Capitolo 2: Prime fasi di analisi e di preparazione dei testi	23
2.1 Il software SketchEngine	23
2.2 Analisi nulle	24
2.2.1 Pulitura del testo	24
2.2.2 Confronti delle liste di frequenza.....	25
2.2.3 Analisi di tri- e tetragrammi, e categorie degli errori.....	26
2.3.4 Individuazione di <i>identity term</i>	32
2.2.5 Misure di impatto	33
2.2.6 Analisi della somiglianza fra i testi	34
2.3 Preparazione definitiva dei corpora	35
2.3.1 <i>Preprocessing</i> del testo	35
2.3.2 Processo di lemmatizzazione	36
2.3.3 Eliminazione di alcuni caratteri	38
2.3.4 Varietà lessicale dei dataset	39
Capitolo 3: <i>Identity term</i>	41
3.1 Scelta delle categorie	41
3.2 Descrizione delle etichette	42
3.3 Popolazione e processo di etichettatura	43
3.4 Normalizzazione delle forme grafiche	47
3.5 Analisi dei dati	48
3.5.1 <i>Identity term</i> femminili	50
3.5.2 <i>Identity term</i> uomini	53

3.5.3	<i>Identity term</i> professioni.....	57
3.5.4	<i>Identity term</i> sessualità ed estetica.....	60
3.5.5	<i>Identity term</i> offensivi.....	62
3.5.6	<i>Identity term</i> relazioni.....	63
3.5.7	<i>Identity term</i> di parentela	65
Capitolo 4: Analisi su bigrammi e trigrammi.....		69
4.1	Gli n-grammi.....	69
4.2	Misure di valutazione dell'impatto.....	70
4.3	Annotazione semantica di bigrammi e trigrammi	73
4.4	Analisi dei dati.....	75
4.4.1	Referenti del trigramma più comune	83
Capitolo 5: <i>Word embedding</i>.....		87
5.1	Gli <i>embedding</i> di parola	87
5.2	Analisi delle liste di similarità.....	88
5.3	Selezione delle parole per i <i>plot</i>	91
5.3.1	Criteri di raggruppamento delle parole in base alla semantica.....	93
5.3.2	Criteri di raggruppamento delle parole in base alla annotazione e alla predizione del modello	95
5.4	Analisi dei <i>plot</i>	99
Conclusione.....		105
Appendice		108
1.	Tabelle delle frequenze assolute degli <i>identity term</i>.....	108
2.	Tabelle con annotazione semantica di bigrammi e trigrammi FP e FN ..	112
3.	Liste di similarità	143
Bibliografia.....		159

Introduzione

Usando le parole di Simone de Beauvoir, "la rappresentazione del mondo come tale è opera dell'uomo; egli lo descrive dal suo punto di vista, che confonde con la verità assoluta" (de Beauvoir 1969). Il saggio della giornalista Caroline Criado-Perez (2020) dimostra, in tempi recentissimi, come questa concezione del mondo si tramuti in un mondo a misura di maschio, in cui la presenza delle donne non è pensata: dai giubbotti antiproiettile alla ricerca medica, dalle politiche di maternità alle toilette pubbliche, dal mito della meritocrazia fino alle cinture dell'automobile, dai trasporti pubblici alle offerte di lavoro, passando per gli iter per fare carriera nell'ambiente accademico fino alla gestione degli spazi nelle aree gioco per bambini, tutto è fatto apposta sulla base dell'uomo. L'uomo sembra costituire il centro, attorno al quale gli oggetti più vaghi girano come satelliti (Simpson 1993). Tutto ciò, oltre che nella percezione delle donne come variante non standard ma marcata, si tramuta anche nel linguaggio, per cui, secondo la legge androcentrica di Coates (1986), i comportamenti linguistici dei parlanti maschi sono visti come positivi, e determinano cosa è visto come normale, mentre i comportamenti linguistici delle parlanti sono visti come negativi, e devianti dalla norma. Questa visione del mondo si è trasferita inevitabilmente nel mondo online, che si è configurato fin dai suoi esordi come uno spazio inospitale per le donne, soprattutto nei social network (Poland 2016). In queste piattaforme, le donne adulte rappresentano il 78% degli utenti, contro il 66% degli utenti uomini,¹ ma nonostante ciò, per le donne i social non sono ambienti sicuri. Come molti report hanno dimostrato, negli ultimi anni il cybersessismo contro le donne è diventato la forma predominante di molestia online,² ed è quindi fondamentale combattere questo fenomeno per contribuire a colmare il divario che pervade tutti gli aspetti della vita privata e pubblica di donne e uomini.

Ispirata dai lavori di Busso, Tordini & Combei (2020), che analizzano la rappresentazione della violenza di genere nei media italiani, usando un approccio *corpus-based*; di Minnema et al. (2021), che usano la *frame semantic* per indagare come la percezione della responsabilità nei casi di femminicidio negli articoli di cronaca, così come fanno Meluzzi et al. (2021) ma concentrandosi sui casi di violenza domestica riportati sui giornali, ho cercato di combinare, per il mio lavoro di tesi, la linguistica, e in

¹ <https://www.statista.com/statistics/471345/us-adults-who-use-social-networks-gender/>

² Si vedano fra gli altri i report di Maeve et al. (2015), di UN Women (2015) e quello di Vox (2016).

particolare la linguistica computazionale, con lo studio dell'odio verso le donne, per rincorrere un ideale di parità. Ho quindi accolto con entusiasmo la possibilità di lavorare sul dataset di meme del *task* di identificazione automatica di contenuti misogini multimodali proposto da Fersini et al. (2022) nella campagna di valutazione semantica SemEval del 2022. Il mio compito è stato quello di analizzare dal punto di vista linguistico le ragioni degli errori di classificazione commessi dai modelli di *deep learning* proposti per la *challenge*, cercando di individuare eventuali pattern linguistici che caratterizzassero la lingua dei meme delle due classi di predizione, misogina e non misogina, sempre in un'ottica di comprensione dell'errore. Mi sono occupata quindi solo dei testi, lasciando le immagini sullo sfondo della mia analisi.

Nel primo capitolo introdurrò la misoginia come fenomeno offline e online, i meme come mezzi di comunicazione, e le caratteristiche linguistiche di questi ultimi. Parlerò poi di alcuni lavori che sono stati fatti nell'ambito del rilevamento automatico di *hate speech* e di misoginia, per finire con il *task* di cui mi sono occupata. Il secondo capitolo si divide in due parti: nella prima descrivo le analisi preliminari che ho svolto sui testi; nella seconda parlo invece di come ho preparato il testo per eseguire le successive analisi, che vengono descritte nel terzo e nel quarto capitolo, e che partono dall'annotazione semantica. Sono partita dalla parte più superficiale del linguaggio, il lessico, individuando una serie di nomi, gli *identity term*, che hanno referenti umani, con particolare distinzione fra genere femminile e genere maschile, per passare poi agli n-grammi di due o tre parole, sempre identificando quelle parole che potrebbero essere problematiche per i modelli. Nell'ultimo capitolo utilizzo invece gli *embedding* di parola per avere una rappresentazione interna delle logiche di comprensione semantica dei modelli.³

³ Colgo l'occasione per ringraziare la dottoressa Giulia Rizzi, che mi ha sostituita nei processi informatici di estrazione degli *embedding* e di lemmatizzazione.

Capitolo 1: Il fenomeno della misoginia offline e online, e i meme come mezzo di comunicazione

Nel primo capitolo inizio con una panoramica dei concetti principali della storia di genere, fornendo così un punto di vista storico sulle differenze percepite tra i sessi (paragrafo 1.1). Descrivo poi le caratteristiche più comuni dell'odio verso le donne come è espresso sul web, prestando attenzione alle tecniche di cybersessismo più utilizzate (paragrafo 1.2). Proseguo con la descrizione del meme di internet come mezzo di comunicazione, ma anche della sua origine come artefatto culturale (paragrafo 1.3), delineando di seguito le sue caratteristiche linguistiche (paragrafo 1.4) Infine, fornisco una visione d'insieme dei lavori precedenti sull'ambito della misoginia online (paragrafo 1.5), e presento la *challenge* MAMI lanciata nell'edizione del 2022 della campagna di valutazione SemEval, su cui si basa il mio lavoro (paragrafo 1.6).

1.1 Il fenomeno della misoginia offline da una prospettiva storica

Le guerre in Iraq e in Afghanistan combattute tra il 2001 e il 2012 sono costate 6.488 caduti agli Stati Uniti. Negli stessi anni le donne statunitensi vittime di femminicidio sono state circa 10.500 (Doyle 2021). Solo nel 2021, in tutto il mondo circa 45.000 donne sono state uccise per mano di un compagno o di un familiare.⁴ Si parla di femminicidio quando un uomo, tipicamente un membro della famiglia o un partner, uccide una donna per motivi di genere, dipendenti cioè dai ruoli sociali che sono imposti alle donne dall'uomo o dalla società (Merli 2015). Il genere indica i significati dati alle differenze percepite fra i sessi (Scott 1999), ossia tutti quegli attributi che vengono assegnati dalla società ad un sesso o all'altro, ma che non dipendono da altro se non dalla cultura della società stessa.⁵ La differenza fra genere e sesso risiede nella sua origine culturale, in opposizione a quella invece biologica del sesso; nel corso della storia proprio le differenze biologiche, che in quanto tali non possono essere messe in discussione, sono state usate per giustificare i ruoli complementari che le donne e gli uomini

⁴ <https://www.unwomen.org/sites/default/files/2022-11/Gender-related-killings-of-women-and-girls-improving-data-to-improve-responses-to-femicide-feminicide-en.pdf>

⁵ Gianini Belotti nel suo saggio *Dalla parte delle bambine* descrive a fondo come questi attributi ci vengono inculcati ancor prima di nascere (Gianini Belotti 1973 [2014]); lo studio venne poi ripreso da Lipperini (2007).

potevano ricoprire nella comunità. A partire dal XVIII secolo i due sessi iniziarono ad essere rappresentati dalla scienza come opposti,⁶ e in particolare con la Rivoluzione Francese la presunta inferiorità naturale delle donne fu alla base della loro esclusione dai diritti civili, riservati così agli uomini (Rose 2010).⁷ Nella seconda metà del '900 venne introdotta nelle scienze umane la categoria di *genere*, e in particolare nella storia delle donne: il genere quindi si riferisce a tutte le aspettative frutto di costrutti socio-culturali che vengono imposte a un individuo in base al suo sesso, e che lo caratterizzano da qualsiasi punto di vista. Il fattore fondamentale è che questi attributi si ordinano su una scala gerarchica, dove le donne occupano una posizione inferiore (Rose 2010). Il genere è, citando Scott (1999), uno dei mezzi principali attraverso cui si strutturano e si giocano le relazioni di potere.⁸ Kelly-Gadol (1976) e Zemon Davis (1976), le madri fondatrici della storia di genere, aggiungono al genere un aspetto strutturale: l'intersezionalità.⁹ Quando si parla di genere, non si può non parlare anche di razza e di classe, che entrano simultaneamente in campo, e che si intersecano fra di loro.¹⁰ Ogni individuo appartiene a più categorie sociali, che dialogano simultaneamente per creare relazioni di potere, a livello soggettivo, di gruppo e di istituzioni (Marchetti 2022). Nel presente lavoro si parlerà però soltanto di genere, sia perché l'obiettivo della classificazione dei modelli è il rilevamento dell'odio verso le donne tutte, sia perché il genere è la categoria più riconoscibile all'interno di un mezzo come i meme, e sarebbe perciò molto difficile approfondire tutti i numerosi aspetti che la discriminazione può colpire. Non si parlerà quindi di razza, di classe sociale, di disabilità, di religione, di orientamento sessuale, di appartenenza a minoranze, o di età.

Basandosi su una distinzione gerarchica, dove un attore domina sull'altra, il genere è motivo di discriminazione, e quindi di violenza. Ci sono tante forme che la

⁶ Precedentemente si parlava invece di un solo sesso, ovviamente quello maschile: l'unica differenza fra donne e uomini erano i genitali, che nelle donne erano rivolti verso l'interno, caratteristica che le rendeva inferiori all'uomo (Schettini 2012).

⁷ E nemmeno a tutti gli uomini, ma solo a quelli che rispettassero certe prerogative economiche e di classe sociale.

⁸ Per aggiungere un elemento all'importanza della gerarchia sociale che le differenze, presunte, di genere veicolano, si può notare l'odio di cui sono vittime le persone che non rispecchiano la dicotomia donna-uomo, fuoriuscendo da queste due categorie granitiche, ovvero soprattutto i transessuali e gli intersessuali; questi individui, mettendo in discussione la scientificità biologica degli attributi di genere, minano di conseguenza l'ordine gerarchico di potere collegato, violando i confini sociali fra donne e uomini (Schettini 2012).

⁹ Il termine *intersezione* è stato in realtà coniato nel 1989 da Crenshaw (1989).

¹⁰ In particolare per l'importanza che la razza assume nel discorso di genere si veda anche Ruiz e DuBois (2008).

violenza di genere può assumere, da quella psicologica, a quella economica,¹¹ verbale, passando da quella sessuale e a quella violenza fisica, fino ai risvolti che la violenza domestica ha sui minori che vi assistono, la violenza assistita.¹² Per citare alcune cifre, il 78% delle vittime di stupro sono donne e il 100% delle persone che commettono violenza sessuale contro le donne sopra i 18 anni sono uomini; il 92% delle persone che commettono molestie sessuali sono uomini; l'87% di chi perpetra stalking è un uomo e l'8,3% delle donne subisce stalking almeno una volta nel corso della propria vita, contro il 2,2% degli uomini; il 20% delle donne ha subito almeno una volta violenza da parte del partner, contro il 3% degli uomini; l'84% delle vittime di abusi domestici all'interno del matrimonio sono donne, e questi abusi sono perpetrati da uomini (Poland 2016). L'episodio culminante della violenza fisica di genere è il femminicidio. Nella violenza di genere, soprattutto in quella fisica, il concetto di onore maschile ha un ruolo importante. L'onore è uno degli attributi che culturalmente vengono imposti all'uomo, e che si proietta sulla morale della propria donna, e in particolare nella capacità, dell'uomo, di controllare i comportamenti sessuali femminili delle donne che appartengono al proprio nucleo familiare, o che intraprendono una relazione con loro (Rinaldi 2021):¹³ l'onore non è quindi una qualità insita, ma proviene dal riconoscimento sociale, dallo sguardo della comunità. Ecco che anche la violenza maschile dipende dagli attributi di genere, poiché è frutto di come storicamente si è affermato il concetto di mascolinità.¹⁴

1.2 La misoginia online

Con l'avvento di internet, le ideologie culturali dapprima limitate al mondo offline si sono trasferite anche sul Web, usando un mezzo diverso per veicolare lo stesso messaggio. Fra le tecniche usate dai cybersessisti per riproporre online la struttura sociale gerarchica propria del mondo offline ci sono il deragliamento, che ha come fine il controllo della conversazione spostandola su argomenti più comodi per l'utente;¹⁵ il

¹¹ Per dati sulle politiche di diminuzione del dislivello sociale fra donne e uomini a livello mondiale si veda il Gender Gap Report del 2022 https://www3.weforum.org/docs/WEF_GGGR_2022.pdf

¹² Per la violenza assistita si veda Gualco, Rensi e Fossa (2017).

¹³ Basti pensare al delitto d'onore in Italia, che riconosceva, solo per gli uomini, un'attenuante se l'omicidio era passionale.

¹⁴ Per altri studi sulla mascolinità in prospettiva storica si vedano anche Karras (2003), e Fiorino (2006).

¹⁵ Sul controllo che gli uomini operano costantemente nelle conversazioni offline con le donne si vedano Maltz &orker (1982).

mansplaining, ossia quando un uomo spiega ad una donna un argomento che la donna conosce meglio dell'uomo stesso; la molestia verbale basata sul genere, che è una delle forme più comuni di molestia online; abusi e minacce (Poland 2016), come per esempio lo "stupro virtuale", in cui viene simulato e/o descritto uno stupro (Spallaccia 2017). Se il deragliamento e il *mansplaining* rimangono quasi invariati fra spazi offline e online, le molestie basate sul genere e gli altri abusi e minacce conoscono online delle forme peculiari. Fra queste, l'uso di immagini accompagnate da un testo, come nel caso dei meme, è molto comune, e si serve spesso di un'oggettificazione sessualizzata e violenta della donna (Spallaccia 2017). Anche l'impersonificazione, il *doxxing* e lo *swatting* sono forme tipiche dell'online. Si parla di impersonificazione quando un utente crea un sito internet o un account rubando l'identità di un'altra persona, con l'intenzione di danneggiarla; il *doxxing* avviene quando vengono diffuse online informazioni personali come il nome, l'indirizzo di casa o del posto di lavoro, il numero di telefono, al fine di minare la sicurezza personale della vittima; lo *swatting* consiste invece nel chiamare le forze dell'ordine per minacce di suicidio, rapimenti di persona, incendi o altre situazioni di pericolo del tutto inventate che coinvolgono la vittima, ma che richiedono un intervento immediato della polizia (Poland 2016; Spallaccia 2017).

L'organizzazione no profit *Working to Halt Online Abuse* (WHOA) ha calcolato che dal 2000 al 2013 il 70% delle vittime di abusi online è composto da donne, e il 47% dei molestatori è uomo (ma ben il 22% del sesso dei molestatori è ignoto).¹⁶ Gli abusi, le molestie, le minacce, l'odio perpetrati su internet hanno conseguenze sulla vita reale delle vittime. Secondo Poland (2016) e Spallaccia (2017) le ripercussioni degli abusi online possono presentarsi su più fronti: gli autori distinguono ripercussioni emotive e psicologiche (come depressione, disturbi alimentari, autolesionismo o suicidio),¹⁷ sociali, economiche (che coinvolgono le spese legali per combattere le molestie ricevute online, oppure i costi di un eventuale trasferimento nel caso di *doxxing*), lavorative (per i potenziali danni all'immagine professionale della persona). Un caso esemplare di abusi che sono iniziati online ma che sono sfociati nel mondo offline è il caso di *Gamergate*, cominciato nel 2014. Si tratta di una campagna di odio nei confronti di alcune donne coinvolte nella sottocultura dei videogame; fra queste, la sviluppatrice di videogame Zoe Quinn fu vittima, a partire dal 2013, di molestie, minacce di morte e

¹⁶ Dati presi dal sito <http://www.haltabuse.org/resources/stats/index.shtml>

¹⁷ Per la cronaca italiana basti pensare al caso di Tiziana Catone del 2016, suicidatasi a causa della divulgazione di materiale pornografico e del conseguente odio di cui è stata vittima sui social.

di stupro, *revenge porn*, diffusione del proprio indirizzo e altre informazioni personali in seguito a un articolo diffamatorio diffuso dal suo ex compagno (Poland 2016).

La potenza dei meme come mezzo di comunicazione per la diffusione di ideologie è studiata da DeCook (2018), che mette in luce come un movimento politico di estrema destra, i Proud Boys, suprematisti bianchi e anti-femministi, si servano proprio dei meme come strumento di propaganda politica sui social network. L'autrice evidenzia anche come la comunità stessa del gruppo politico sia a sua volta attiva e coinvolta nella creazione di meme che sono portatori delle ideologie politiche del gruppo, e come l'anonimato garantito da alcune piattaforme renda ancora più arditi i contenuti dei meme. Creare meme è un'operazione anonima, e l'anonimato è una forma di protezione non trascurabile quando si parla di ideare contenuti che vanno contro l'individuo o contro la morale e la sensibilità comuni, offrendo un grado di libertà che altrimenti difficilmente viene conosciuto, incentivando in qualche modo gli utenti a spingersi oltre ciò che sarebbe loro consentito in un contesto reale (Davison 2012).

Per tutte le ragioni descritte sopra, a partire dalle radici storiche della misoginia, alle conseguenze fisiche che l'odio verso le donne può causare, fino all'anonimato, i meme in Figura 1 (n. 2010, n. 8190, n. 6474 e n. 7255) non sono né innocui né divertenti, ma veicolano oggettificazione, soggiogamento, stereotipi di genere e subordinazione, violenza di genere, e, data l'influenza che possono avere nella propagazione di un certo tipo di cultura che vede le donne come esseri inferiori e quindi non degni di accedere agli stessi spazi della società occupati dagli uomini, è importante che la circolazione di meme misogini venga rilevata e censurata.

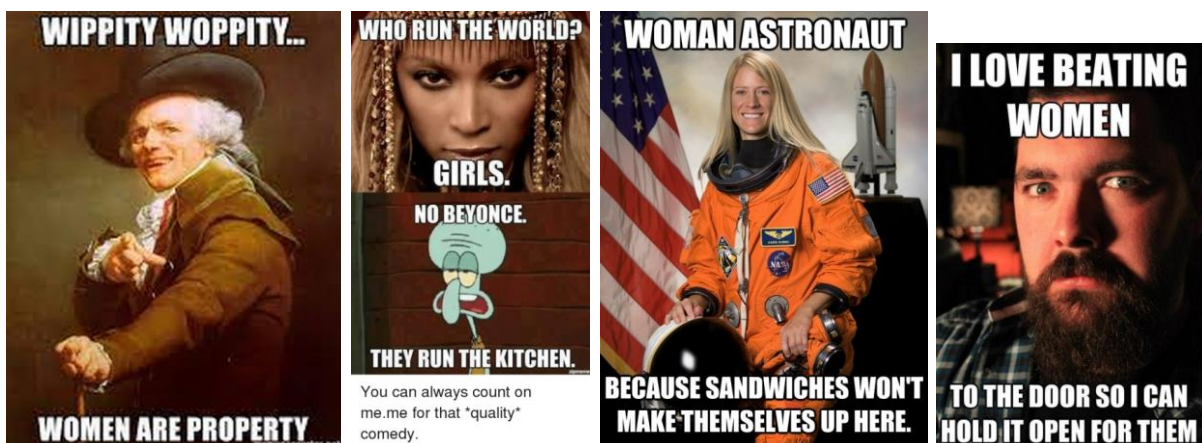


Figura 1. Meme n. 2010, n. 8190, n. 6474, n. 7255.

1.3 I meme come artefatti culturali e i meme di internet

Quando si parla di meme, la forma più nota che questi assumono è quella del meme di internet, un mezzo di comunicazione multimediale che fa uso di un'immagine cui è sovrapposto un testo. Ma il meme come concetto ha radici molto più profonde. Susan Blackmore nell'articolo *The power of memes* (Blackmore et al. 2000) afferma che i meme sono tutto ciò che non è comportamento genetico, ma che contribuisce all'evoluzione della specie, definendo il meme come "an idea, behavior, style or usage that spreads from person to person within a culture", includendo quindi nel concetto di meme le istituzioni politiche e finanziarie, la scienza e anche la lingua. Anche Davison (2012) concettualizza i meme come comportamenti culturali che hanno alla base processi mentali di osservazione e di apprendimento, e che possono diffondersi e cambiare istantaneamente. Secondo Blackmore i meme sono una forma di informazione, e diventano meme solo quelle informazioni che hanno fortuna fra la specie umana perché vengono selezionate e copiate. È possibile che alcuni meme vengano proprio creati per questo scopo "riproduttivo", quasi virale. Fra questi, le catene di Sant'Antonio hanno il fine di essere diffuse, e per spingere la popolazione a condividere le informazioni della catena queste contengono minacce e promesse, a seconda che la catena venga o meno condivisa dal soggetto, esattamente come hanno fatto le grandi religioni: per garantire la propagazione del culto, si promette il paradiso e si minaccia l'inferno (Dawkins 1976). Sempre secondo Blackmore, i mezzi di comunicazione sono quindi fondamentali per la trasmissione dei meme culturali, e con l'avvento di internet la loro diffusione può essere istantanea, e la loro prolificità è aumentata.¹⁸

Davison (2012) definisce i meme di internet come "a piece of culture, typically a joke, which gains influence through online transmission", e cita i primi due archetipi di meme di internet: la prima emoticon realizzata con i segni di interpunzione (:)) nei primi anni '80, e un sito pieno di GIF di criceti danzanti creato nel '99, che è stato visitato diverse migliaia di volte al giorno durante il primo mese. Questi esemplificano le caratteristiche essenziali di un meme di internet, ossia l'uso e la visualizzazione. Infatti, i meme di internet sono considerabili come i prodotti di una cultura digitale partecipativa,

¹⁸ Henry Plotkin, sempre all'interno della stessa pubblicazione (Blackmore et al. 2000), muove delle critiche alla concezione di Blackmore: pensare che la cultura sia solo un insieme di meme è troppo semplicistico, e la realtà che il termine *meme* descrive è così ampia da far perdere significato alla parola.

perché la creazione di un meme non è mai una produzione isolata ma una riproduzione di meme già esistenti; modelli, credenze e stereotipi condivisi da una comunità, che possono venire da attualità, film, pubblicità e altro, vengono riutilizzati e rielaborati per creare qualcosa di nuovo, configurando quindi il meme come artefatto collettivo (Ostanina-Olszewska & Majdzińska-Koczorowicz 2019; Dancygier & Vandelanotte 2017).

Come già detto, i meme di internet sono contenuti multimediali, in cui materiale linguistico e materiale iconografico si fondono, per creare un significato compositivo, ma che a volte può necessitare di conoscenze esterne al meme stesso. I meme hanno diverse potenzialità comunicative, dal momento che, secondo Grundlingh (2017), possono soddisfare tutti i tipi di atti illocutivi comunicativi della teoria di Bach e Harnish (1980), ossia gli atti costitutivi, direttivi, commissivi e di riconoscimento. I meme possono avere caratteristiche strutturali molto diverse fra loro, ed è difficile riuscire a individuarle tutte. Per questo qualsiasi tassonomia risulterà almeno in parte parziale, ma si possono elencare i tre tipi più diffusi:¹⁹

1. Meme a macro immagine (Ostanina-Olszewska & Majdzińska-Koczorowicz 2019; Grundlingh 2017; Davison 2012): il materiale iconografico consiste in una sola immagine, a cui è apposto un testo, che spesso è suddiviso in due parti, il *bottom text* e il *top text*. L'immagine in questi casi è quasi più importante del testo per comprendere il meme, e spesso il contenuto è umoristico. Ne è un esempio il primo meme in Figura 2;
2. Meme *reaction shot*: è una variazione dei meme a macro immagine, dove però l'immagine rappresenta la reazione al testo scritto (come nel secondo meme in Figura 2), ma può anche non avere materiale linguistico;
3. Meme standard/meme-modello: sono i meme diventati *mainstream*, riconosciuti da tutta, o quasi, la comunità di internet. È il caso del terzo meme in Figura 2.

¹⁹ Una tassonomia più completa è fornita da Milner (2012).



Figura 2. Meme n. 11574, n. 7173, n. 10868.

Nonostante i meme possano assumere molte forme, la maggior parte di essi appartiene alle tre categorie elencate: i creatori di meme sfruttano modelli già esistenti, operando delle innovazioni che però restano in parte limitate al modello stesso, anche per quanto riguarda l'espressività (Nissenbaum & Shifman 2017), nonostante le possibilità comunicative che i meme assolvono. Secondo lo studio di Dancygier e Vandelanotte (2017) il fine principale dietro la creazione di un meme è condividere un'esperienza emotiva del creatore che sia facilmente condivisibile anche dai fruitori del meme stesso. Per questo la compartecipazione verso l'esperienza descritta diventa un fatto fondamentale della diffusione dei meme, che quindi si propagano più facilmente all'interno di comunità virtuali che, quanto più condividono degli ideali e una visione del mondo, tanto più saranno coese e forti. Da un punto di vista sociale, l'esperienza emotiva che chi crea un meme rappresenta è conservativa, cioè in linea con la concezione del mondo egemone (Nissenbaum & Shifman 2017). Questo vuol dire che gli autori dei meme più diffusi sono i gruppi socialmente dominanti (Nissenbaum & Shifman 2018), portatori cioè dei valori culturali che predominano, e quindi spesso i meme riflettono i modi in cui vengono vissuti i rapporti con i gruppi discriminati e con le minoranze (Miltner 2014; Philips & Milner 2017), così come riflettono e rafforzano gli stereotipi e i pregiudizi con cui questi gruppi vengono percepiti (Frazer & Carlson 2017). L'importanza che i meme possono avere nella creazione e nel rafforzamento di un'ideologia e la loro potenzialità come mezzo di comunicazione politica sono studiate nel lavoro di DeCook (2018), che analizza meme del *social network* Instagram diffusi da un gruppo suprematista bianco e occidentale antifemminista di estrema destra, i Proud Boys, movimento affiliato del più ampio alt-right. Il gruppo usa i meme come mezzo di propaganda politica, sfruttando la funzione dei *social network* come strumento

di socializzazione soprattutto per i giovani, e di creazione per questi ultimi di significati dei loro mondi. Il meme risulta inoltre essere particolarmente funzionale nella creazione e nel rafforzamento di ideologie politiche grazie anche all'immediatezza del messaggio e alla sua facilità di propagazione.

1.4 Caratteristiche linguistiche dei meme

La lingua dei meme si caratterizza per la peculiare forma grafica che prevede: un'immagine che viene accompagnata da un testo scritto, e per questo si può collocare nella categoria del linguaggio multimediale, dove il significato del meme è spesso comprensibile soltanto combinando il significato del testo al significato dell'immagine. La predominanza dell'uno o dell'altro non è sistematica; Pranesh & Shekhar (2020) hanno però notato che il testo è la parte che più spesso fornisce informazioni sufficienti e più chiare per quanto riguarda meme con contenuti d'odio. A livello linguistico, la lingua dei meme può essere considerata come una sottocategoria delle lingue del web, e ne condivide alcune delle principali caratteristiche.

L'uso massivo della scrittura sul web ha portato a un abbassamento della serietà e della formalità che venivano attribuite al testo scritto prima dell'avvento di internet. La frequenza sempre maggiore che si fa della scrittura è una delle motivazioni principali dell'abbassamento di registro che si nota nella lingua usata sul web, soprattutto nei contesti dei social media, dei forum e delle chat simultanee (Prada 2015): la differenza fra scritto e parlato diventa in questi contesti più labile, in un livello diamesico che non può più essere nettamente distinto fra il parlato e lo scritto, che si fondono insieme per creare uno scritto-parlato, secondo la definizione del linguista Giovanni Nencioni (1976). Questa compresenza di tratti tipici sia dello scritto che del parlato, ma più tendenti al parlato, dipende dall'immediatezza della progettazione del discorso, grazie alla quale si tende ad usare lessico ad alta frequenza e strutture sintattiche molto semplici, a volte anche appartenenti a varietà substandard (Prada 2015). Data la particolare diffusione che hanno fra i giovani, i meme contengono molte parole dello *slang* e una sintassi non solo molto semplice ma anche non conforme con la grammatica dell'inglese standard, caratteristiche dovute in parte anche alla brevità dei testi (Setpal & Sarti 2020), che portano alla ricerca di un'espressività immediata dei contenuti. Disgrafie, abbreviazioni ed uso ludico-espressivo dei caratteri caratterizzano

soprattutto le comunicazioni fra giovani, secondo Prada (2015; 2016), che riassume le disgrafie nelle seguenti categorie:

- Errori di battitura o refusi (per esempio i numerosi casi in cui non viene scritto l'apostrofo fra soggetto e ausiliare, come "lm" al posto di *l'm*, o "youd" al posto di *you'd*; oppure forme come "vetran" anziché *veteran*, "gentleman" per *gentleman*);
- Competenza linguistica carente;
- Tachigrafie, abbreviazioni, single (come i comuni "btw" o "idk" per *by the way* e *I don't know*);
- Fonosimbolismi (per citarne alcuni, i gergalismi grafici *gonna* o *gotta* per *going to* e *got to*, oppure "gurlz" come forma ludica per *girl*, o ancora le ripetizioni di fonemi o interpunzione con uso espressivo, come "NOOOOOOOO" o "ccckkkkk!!!! FFFFTTECHN C CHIFFFF PFFFFTCH FFFFFTT CHHT PFFCHH CHH CHHCH");
- Emoticon.

La punteggiatura è piuttosto scarsa, data anche la brevità dei testi, tranne quella funzionale alla resa di intenzioni prosodiche e pragmatiche, come punti esclamativi, interrogativi e di sospensione, e le virgolette. Un'altra caratteristica tipica delle lingue del web in generale sono le neoformazioni, come "fugly" o "shemale", che sono rispettivamente un aggettivo poco rispettoso per indicare una persona molto brutta, e un sostantivo, anche questo offensivo, con cui ci si riferisce alle donne trans che non hanno cambiato i genitali tramite operazione chirurgica.

Sia per il livello diatopico che per quello diafasico dei creatori dei meme non ci sono dati precisi, ma secondo Nissebaum e Shifman (2018) i maggiori autori sono giovani uomini bianchi e di classe media, e, dato che il repertorio dei meme è in inglese, si può pensare che buona parte di questi siano provenienti da paesi anglofoni.

1.5 Lavori precedenti sulla misoginia nei *social media*

I modelli di *deep learning* hanno iniziato ad essere impiegati per i compiti di rilevamento automatico di *hate speech* solo nell'ultimo decennio (Razavi et al. 2010; Nobata et al. 2016; Zhang & Luo, 2018; Zampieri et al. 2019; Capozzi et al. 2019), mentre gli studi rivolti specificamente all'identificazione della misoginia risalgono a pochissimi anni fa (Anzovino et al. 2018). Nel campo della misoginia, la categoria di *hate speech*

probabilmente più diffusa, ci si è concentrati spesso sui post o sui commenti di Twitter (García-Díaz et al. 2021; Pamungkas et al. 2020), anche grazie ad alcuni *task* come *l'Automatic Misogyny Identification* (AMI) presentato nella campagna di IberEval2018 (Fersini et al. 2018) e in quella di Evalita-2020 (Fersini, Nozza, Rosso 2020), che hanno stimolato la ricerca in questo campo (Ahluwalia et al. 2018; Pamungkas et al. 2018; Frenda & Bilal 2018; Muti & Barrón-Cedeño 2018). Altri lavori correlati al rilevamento di contenuti misogini sono Guest et al. (2021), che ha creato e annotato un dataset di post e serie di commenti relativi a conversazioni del *social network* Reddit, sviluppando una tassonomia gerarchica dei diversi tipi di misoginia, oppure l'approccio stilometrico di Anzovino et al. (2018). Alcuni studi sono stati fatti per indagare l'*unintended bias* misogino delle macchine, cioè la tendenza non intenzionale dei modelli ad attribuire ad alcune parole un peso maggiore nella classificazione misogina, anche se il contesto non è di odio (Nozza et al. 2019; Fersini et al. 2021). Ci sono poi una serie di studi che indagano diversi aspetti di misoginia online, usando un approccio di analisi critica del discorso, come i lavori che indagano il mondo degli *incel* (Maxwell et al. 2020; Furl 2022; Gemelli 2021), dei suprematisti bianchi di estrema destra (DeCook 2018), o di *hate speech* misogino su Twitter e Facebook diretto a personaggi pubblici femminili (Spallaccia 2017).

Nel campo della classificazione automatica di meme, nel 2020 sono state proposte ben tre diverse *challenge* per il rilevamento di *hate speech*, una in SemEval, una in Evalita, e una da Facebook, lanciata dal *social network* per la creazione di sistemi multimodali contro i discorsi d'odio (Kiela et al. 2020). Nell'edizione di SemEval è stato proposto il *task* Memotion Analysis – the Visuo-Lingual Metaphor (Sharma et al. 2020). Nell'edizione Evalita è stato presentato il *task* DANKMEMES, The Memeing of Life: Memes, Multimodality and Politics (Milani et al. 2020), che ha promosso per la prima volta la ricerca nella classificazione automatica di contenuti multimodali in italiano.²⁰ Tutti i *task* si concentrano sulla *sentiment analysis* e sull'*hate speech*: Memotion Analysis richiede di individuare il tipo di sentimento veicolata dal meme, distinguendo tra positivo, negativo e neutro; di specificare di che tipo di emozione si tratta fra umorismo, sarcasmo, offesa, e motivazionale; e di esplicitare l'intensità dell'emozione. DANKMEMES invece usa meme sulla politica principalmente italiana, e richiede

²⁰ A questo task hanno partecipato soltanto quattro squadre: Fiorucci (2020), Brezzano et al. (2020), Vlad et al. (2020), Setpal & Sarti (2020).

innanzitutto di rilevare se si tratti o meno di un meme, poi di individuare l'eventuale presenza di discorsi d'odio, e infine di riconoscere l'evento politico di cui si parla.

Altri studi che invece combinano contenuti multimodali e misoginia, sono quelli di Fersini, Gasparini e Corchs (2019), specifico sui meme, e quello di Gasparini et al. (2018), che si concentra invece sulle pubblicità sessiste sia combinando testo e immagine, sia guardando solo all'immagine. C'è poi un lavoro che combina la classificazione automatica con il parlato, con l'obiettivo di riconoscere gli stati emotivi del parlato durante i processi in tribunale (Fersini, Messina, Archetti 2012).

Pochissimi sono invece i lavori in cui si parla dell'analisi dell'errore da un punto di vista linguistico dei modelli, in particolare per la misoginia. I paper che hanno partecipato alle tre *challenge* di classificazione automatica di misoginia che danno indicazioni strettamente linguistiche sugli errori commessi dai modelli sono molto pochi. Fra questi, ci sono Fenda et al. (2018), Van Aken et al. (2018), Lippe et al. (2020), Guest et al. (2021), e Tao e Kim (2022).

1.6 Presentazione del *task* MAMI di SemEval-2022

Le campagne di valutazione semantica SemEval (Semantic Evaluation) sono workshop internazionali a cadenza annuale, con lo scopo di promuovere la ricerca nel campo dei sistemi di analisi semantica del linguaggio naturale. I *task* dei workshop vengono selezionati tra quelli proposti da gruppi di ricercatori, e nella sedicesima edizione del 2022 (Emerson et al. 2022) ne sono stati presentati dodici, fra cui MAMI – *Multimedia Automatic Misogyny Identification* (Fersini et al. 2022). La *challenge* MAMI ha visto partecipare più di cento gruppi di ricercatori,²¹ e si è distinta, secondo gli organizzatori di SemEval, per l'originalità del compito proposto, e per la qualità scientifica dei dataset forniti e dei metodi di valutazione. La *challenge* presenta un dataset di 10.000 meme sui cui allenare i modelli a svolgere due *sub-task*, che dovranno poi essere testati su un dataset di 1.000 meme. Il primo *sub-task* prevede una classificazione binaria *misogino* o *non misogino*, mentre per partecipare al *sub-task* B è richiesto di riconoscere, nel caso in cui il meme sia misogino, a che categoria di misoginia questo appartiene: stereotipo, *shaming*, oggettificazione, violenza. L'originalità della sfida consiste in due

²¹ Per citarne alcuni, Sherzod, Cheema ed Ewert (2022), che si sono classificati primi nel *task* B; Zhi et al. (2022), arrivati primi anche loro nel *task* B, e terzi nel *task* A; Zhou et al. (2022), i secondi del *task* A; e Sharma et al. (2022), classificatisi terzi nel *task* A.

fattori. Il primo è che l'oggetto da classificare è un contenuto multimediale, per cui è necessario creare dei sistemi che siano capaci di gestire e combinare informazioni provenienti da due diverse fonti, il testo e l'immagine. Il secondo motivo è che nei meme selezionati l'odio è spesso espresso in modo implicito, mentre l'esplicitezza ha caratterizzato i precedenti compiti di riconoscimento di contenuti d'odio. Tutti i dati e maggiori informazioni sul *task* sono disponibili al link Github <https://github.com/MIND-Lab/SemEval2022-Task-5-Multimedia-Automatic-Misogyny-Identification-MAMI->.

I meme sono stati presi da due social media, Twitter e Reddit, e da alcuni dei siti di creazione di meme più popolari, quali 9gag.com, il sito più famoso e il primo a diventare virale, knowyourmeme.com, imgur.com, memecrunch.com, imgflip.com, memeshappen.com e me.me.com.²² La selezione è stata fatta tramite *site scraping* e download manuale usando parole chiave o *hashtag*, tramite ricerca di discussioni con termini femminili come soggetto, ricerca di nomi di donne famose e degli eventi in cui sono state coinvolte, ricerca di discussioni o account esplicitamente anti-femministi e misogini, come il movimento Men Going Their Own Way (MGTOW). Sono stati così raccolti circa 15.000 meme, che sono stati annotati manualmente tramite piattaforme di *crowd-sourcing*, facendo annotare ogni meme tre volte da tre annotatori diversi. I 10.000 meme del dataset di *training* sono stati selezionati fra i 15.000 a seconda del grado di accordo di annotazione. Gli annotatori si distribuiscono eterogeneamente per età e genere. Invece, il dataset di 1.000 meme di *testing* è stato annotato da tre esperti di misoginia. Per svolgere l'annotazione di entrambi i *task* sono state fornite delle linee guida. Per il primo *sub-task* l'indicazione per il riconoscimento di contenuti misogini è stata: "a meme is misogynous if it conceptually describes an offensive, sexist or hateful scene (weak or strong, implicitly or explicitly) having as target a woman or a group of women. Misogyny can be expressed in the form of shaming, stereotype, objectification and/or violence", mentre "a meme that does not express any form of hate against women" è stata l'indicazione per classificare un meme come non misogino. A queste generiche definizioni sono stati affiancati degli esempi di meme sia misogini che non. Sono state poi date altre indicazioni per il secondo *sub-task*, differenziando il tipo di misoginia, quindi *shaming*, stereotipizzazione, oggettificazione, violenza, etichette che possono essere compresenti in uno stesso meme. Agli annotatori è stato inoltre chiesto

²² Quest'ultimo non è più raggiungibile.

di assegnare un punteggio da 1 a 3 al grado di misoginia del meme, dove 1 è il grado minore e 3 quello maggiore.

Gli organizzatori della *challenge* hanno fornito ai partecipanti i meme, le annotazioni manuali, la trascrizione dei testi dei meme fatta con il riconoscimento ottico dei caratteri Google Cloud Vision,²³ e tre modelli di base per l'elaborazione del testo, dell'immagine, e dell'insieme di testo e immagine. Per il primo *sub-task*, il modello di rappresentazione profonda del testo è il modello pre-addestrato USE (Cer et al. 2018); per la rappresentazione profonda del contenuto dell'immagine è stato fornito un modello di classificazione ottimizzato basato su VGG-16 (Simonyan & Zisserman 2014); per testo e immagine insieme è stata invece prevista una concatenazione dei due modelli precedenti, attraverso una rete neurale a un solo strato.

Il modello di rappresentazione profonda del testo Universal Sentence Encoder (USE) codifica le frasi in vettori di *embedding* perfezionati, ed usa due modelli di codifica. Il primo codificatore costruisce degli *embedding* di frase usando un sotto-grafo che calcola gli *embedding* di parola all'interno della frase tenendo conto anche del contesto, cioè di tutte le altre parole, e le rappresentazioni di queste parole vengono convertite in un vettore di codifica della frase. Il secondo codificatore è una rete di mediazione profonda che crea *embedding* di frase tramite gli *embedding* di parole e di bi-grammi. Entrambi i codificatori producono come *embedding* di frase con un vettore di 512 dimensioni. Il modello è stato allenato su Wikipedia, articoli internet, siti di domanda e risposta, e forum di discussione.

I dati che ho usato per il presente lavoro fanno riferimento al dataset da 10.000 meme, su cui i creatori del *task* hanno, al di fuori della campagna di valutazione, testato le capacità di classificazione dei modelli. Le due classi di predizione 1 (misogino) – 0 (non misogino) che uso, fanno riferimento proprio a questa classificazione. Durante l'esposizione, parlerò di GT+, GT-, TP, TN, FP e FN, che sono single in uso nei compiti di classificazione automatica per indicare gli elementi dati in input in base alle predizioni e alle annotazioni fatte su di questi: per annotazione si intende il livello di classificazione gold, e quindi tutti gli elementi che appartengono all'una o all'altra classe, a prescindere dal fatto che siano stati predetti correttamente o meno. Per predizione si intende invece

²³ Per maggiori dettagli si può visitare il sito internet dell'OCR al link <https://cloud.google.com/vision/docs/ocr?hl=it>

il livello di classificazione del modello, a prescindere in questo caso dalla correttezza o meno di quest'ultima. Le sigle sono:

- GT+ (*ground truth* positiva): tutti i meme che appartengono alla classe positiva, cioè misogina, a livello di annotazione
- GT- (*ground truth* negativa): tutti i meme che appartengono alla classe negativa, cioè non misogina, a livello di annotazione
- TP (*true positive*): i meme che sono stati predetti dai modelli come misogini, e che lo sono anche a livello di annotazione
- TN (*true negative*): i meme che sono stati predetti dai modelli come non misogini, e che lo sono anche a livello di annotazione
- FP (*false positive*): i meme che sono stati predetti come misogini, ma che a livello di annotazione sono non misogini
- FN (*false negative*): i meme che sono stati predetti come non misogini, ma che a livello di annotazione sono misogini.

Tutte le liste di frequenza estratte con SketchEngine su cui ho lavorato dal secondo capitolo in poi, quindi le liste di frequenze di parole, bigrammi e trigrammi, sono accessibili al link Github <https://github.com/MIND-Lab/Misogynous-N-Gram-Lists>.

Come detto, l'annotazione manuale dei 10.000 meme è stata fatta dal *crowd*, quindi da individui che non hanno nessuna competenza in materia di misoginia, e le linee guida fornite agli annotatori su che tipi di contenuti dovessero essere considerati misogini e quali no sono molto generali. Questo ha comportato che l'interpretazione misogina dei meme fosse soggetta alla sensibilità personale, e non sostenuta da solide basi teoriche: lungo tutto il lavoro mi sono imbattuta in una suddivisione per classe di annotazione misogina – non misogina molto poco coerente. I meme in Figura 3 sono tutti annotati come non misogini, mentre quelli in Figura 4 come misogini: nel meme n. 991, essere una prostituta appare la caratteristica principale di un personaggio cinematografico femminile, attribuzione molto sminuente nei confronti di una donna; nel meme n. 1876 il femminismo viene accusato di voler dominare gli uomini, stravolgendo il concetto di femminismo e ignorando anche la dominazione millenaria che invece proprio gli uomini hanno avuto e hanno tutt'oggi sulle donne; il meme n. 2339 è invece una sorta di rappresentazione di una molestia sessuale, poiché se una donna incinta invita qualcuno a "*feel her baby*", l'invito è inequivocabilmente a toccarle il grembo, non a penetrarla. È difficile, a partire da questi dati, ricavare delle idee sui criteri sottostanti alla percezione di cosa sia misogino e di cosa non lo sia, soprattutto confrontando il

memes n. 9640, annotato come non misogino, con i memes n. 8092, annotato invece come misogino, che trasmettono un concetto molto simile: un marito che tradisce la moglie in un caso con la cognata e nell'altro con una prostituta.²⁴ L'annotazione appare tanto più strana guardando il meme n. 7067, un'altra situazione di tradimento dove ad essere tradito è un ragazzo, e non c'è nessun giudizio morale nemmeno implicito nei confronti della ragazza. Difatti tutti i memes in Figura 3 e Figura 4 sono stati predetti in modo errato dalla macchina, e si potrebbero fare molti altri esempi. La mancanza di coerenza a livello di annotazione gold è sicuramente un grande problema soprattutto se si parla di comprendere i motivi di una predizione sbagliata.



Figura 3. Meme n. 991, n. 1876, n. 2339, n. 9640, tutti annotati come non misogini.

²⁴ Nel meme n. 8092 viene usata la parola offensiva *hooker*, ma, come si vedrà più avanti (cfr. paragrafo 3.5.1), nonostante la non neutralità del termine, questo viene usato molto di frequente anche nei memes annotati come non misogini.



Figura 4. Meme n. 6132, n. 7067, n. 8092, tutti annotati come misogini.

Capitolo 2: Prime fasi di analisi e di preparazione dei testi

In questo capitolo presento i dati e le strategie usate per l'analisi, iniziando con la descrizione del software impiegato per ricavare la maggior parte dei dati (paragrafo 2.1). Dal paragrafo 2.2 al paragrafo 2.2.6 presento le strategie usate inizialmente per analizzare i dati, ma che si sono rivelate inefficaci, evidenziandone le criticità. Nella seconda parte del capitolo parlo invece dei metodi di preparazione dei dati in vista delle analisi definitive, usando alcuni procedimenti di pulizia e normalizzazione del testo (dal paragrafo 2.3 al 2.3.4).

2.1 Il software SketchEngine

Lo strumento che ho usato di più per le analisi sulla lingua dei meme è SketchEngine (Kilgarriff et al. 2004). Si tratta di un software creato nel 2004 per condurre analisi automatiche su *corpora* che possono raggiungere decine di miliardi di parole. Il software consente l'accesso a 600 *corpora* già annotati, in più di 90 lingue, ma offre anche la possibilità di creare i propri *corpora*. Fra le sue funzioni di investigazione del *corpus* ci sono la *Wordlist* e gli *N-grams*, di cui mi sono servita. La *Wordlist* è uno strumento che genera liste di frequenza di parole a partire dai token del *corpus*; le parole possono essere distinte in base a varie caratteristiche, come le parti del discorso, o parole che iniziano, finiscono o contengono i caratteri desiderati dall'utente. Si può decidere di escludere una serie di parole, e di farsi restituire solo le parole che hanno una certa frequenza minima o massima. Ci sono poi le opzioni di esclusione delle non parole, e di non fare distinzione fra caratteri minuscoli e maiuscoli. La lista viene restituita su due colonne: nella prima ci sono le parole, e nella seconda il valore della frequenza assoluta corrispondente. Anche lo strumento *N-grams* produce delle liste di frequenza, ma di sequenze di token, che possono andare da uno a sei. Sono disponibili anche per gli *n-grammi* tutte le opzioni delle liste di parole, ma la frequenza minima perché un *n-gramma* venga restituito è di due occorrenze.

2.2 Analisi nulle

Una prima fase del lavoro di estrazione e analisi dei dati non ha portato risultati utili, perché la metodologia si è rivelata essere inadeguata. Alcune di queste analisi sono state quindi intraprese, e poi abbandonate prima di essere concluse, ma sono state utili sia per familiarizzare con il *corpus* di meme, sia per capire dove concentrare le analisi successive.

2.2.1 Pulitura del testo

Per l'estrazione delle liste di frequenza di parole e di n-grammi ho fatto inizialmente alcune prove usando le opzioni di pulitura del testo offerte da SketchEngine, ossia l'eliminazione delle non parole²⁵ e la *case sensitivity*. Le liste di frequenza, di cui parlo poco sotto, sono state estratte inizialmente senza selezionare queste due opzioni. Le liste degli n-grammi senza l'eliminazione delle non parole contenevano però troppi segni di punteggiatura, e ho quindi selezionato solo l'opzione di eliminare le non parole. Non sono intervenuta in modo più invasivo perché l'obiettivo era proprio intervenire il meno possibile all'interno di un linguaggio non standard come quello dei meme, che è molto informale e quindi molto prone a contenere disgrafie. Intervenire troppo avrebbe potuto comportare il rischio di non catturare alcune peculiarità che sono invece importanti per il livello diafasico, ma anche solo lo scegliere l'opzione *case sensitive* ha inficiato molto in particolare i valori delle frequenze assolute di queste prime analisi: la lingua dei meme fa largo uso di frasi scritte interamente con caratteri maiuscoli, e SketchEngine, con la *case sensitive*, restituisce come due elementi diversi parole o n-grammi che si distinguono solo per il carattere maiuscolo o minuscolo, quando in realtà il modello USE non fa questa distinzione. Le analisi definitive sono state invece fatte con un testo ripulito e normalizzato in modo un po' più dettagliato.²⁶ Le trascrizioni dei quattro dataset sono state gli input dati a SketchEngine per l'estrazione di una serie di dati:

1. Liste di frequenza di:

²⁵ Da SketchEngine, "Non-words (also spelt nonwords) are tokens which do not start with a letter of the alphabet. Examples of non-words are numbers, punctuation but also tokens such as 25-hour, 16-year-old, Important, 3D. Tokens such as post-1945, mp3 or CO2 are words because they start with a letter." The regular expression Sketch Engine users to identify non-words is `[^[:alpha:]].*`.

²⁶ Si veda il paragrafo 2.3.

- a. *Type*
 - b. Lemmi
 - c. Nomi
 - d. Aggettivi
2. N-grammi:
- a. Bigrammi
 - b. Trigrammi
 - c. Tetragrammi.

2.2.2 Confronti delle liste di frequenza

Nella linguistica dei *corpora* uno dei metodi di analisi preliminare vede protagoniste le liste di frequenza, che servono per notare fenomeni tipici o di interesse su cui poi fare ulteriori analisi (Spina 2001). Quindi, per ogni lista di frequenza delle parole e degli n-grammi ho usato Excel confrontando i numeri di frequenza assoluti degli elementi target delle liste. Tramite i confronti delle frequenze assolute volevo verificare quali fossero le parole o gli n-grammi che apparissero più di frequente o al contrario meno spesso, soprattutto fra i dataset GT e i corrispettivi con classificazione sbagliata. A questo scopo ho tentato di isolare gli elementi che avessero frequenze assolute particolarmente alte o basse, oppure frequenze uguali o quasi, calcolando alcuni valori numerici, usando per esempio le frequenze normalizzate calcolate dal software, oppure ricavando le percentuali dei valori delle frequenze assolute. Avendo a disposizione *corpora* di dimensioni molto diverse, per confrontare le frequenze assolute queste devono essere normalizzate (Spina 2001). Ho quindi provato a sottrarre questi valori percentuali fra coppie di dataset per ordinare le liste in ordine crescente o decrescente così da avere valori molto vicini allo 0 per elementi che avessero frequenze relative molto simili, e valori molto vicini a 1 per elementi le cui frequenze relative al contrario fossero molto diverse.

Questi tentativi sono problematici per due ragioni principali collegate fra loro. Poiché le liste di frequenza sono composte da migliaia di parole, è difficile evidenziare gli elementi più interessanti senza avere un coefficiente o un valore che tenga conto dell'impatto che questi elementi hanno per i modelli. Avere un indice che restituisse l'importanza di un certo elemento all'interno dei modelli e creare su questo una gerarchizzazione delle liste di frequenza delle parole e degli n-grammi è invece un

fattore fondamentale per riuscire a ricavare informazioni potenzialmente significative da migliaia o centinaia di dati. Un coefficiente non definitivo è quello descritto al paragrafo 2.2.5, mentre il coefficiente definitivo è descritto al paragrafo 4.2.

2.2.3 Analisi di tri- e tetragrammi, e categorie degli errori

I trigrammi e i tetragrammi sono delle unità linguistiche che possono catturare caratteristiche interessanti all'interno di un testo, o di una serie di testi, come nel caso del dataset preso in considerazione. Per questo ho iniziato un'analisi semantica e pragmatica dei trigrammi del dataset GT+ (non di GT- perché ho interrotto questo tipo di lavoro una volta finiti i trigrammi del dataset misogino). Mi sono concentrata in particolare su quei trigrammi che fossero presenti anche nei due dataset classificati in modo scorretto, per cercare di trovare delle caratteristiche sia del contesto linguistico che della componente visiva che potessero effettivamente rendere difficile la classificazione dei modelli. A questo livello di analisi sono emerse le categorie principali di ambiguità per la classificazione dei modelli, che spesso si combinano fra di loro. Per questo motivo, questo tipo di analisi è stata molto importante per capire alcune criticità dei modelli. Un *leitmotiv* di buona parte delle categorie è collegato a una ragione molto generica, l'uso cioè di linguaggio aggressivo, o sessualmente esplicito, o genderizzato, anche quando si trova in contesti che non sono misogini, che è un problema molto comune in compiti simili (Guest et al. 2021). Le categorie che ho individuato sono:

1. Contenuti ironici e contenuti sarcastici.²⁷ Per esempio, il meme FP n. 2082 (fig. 5) ironizza sull'inutilità di pagare una prostituta per venire masturbati, ma non contiene nessun contenuto d'odio nei confronti delle donne ("Man: Hey... How much for a hand job? Hooker: \$10. You want one? Man: Nope... I was just curious how much I save when I do it myself."). Tuttavia, il meme contiene le parole *hooker* e *hand job* che possono essere ambigue, e inoltre è accompagnato da un'immagine comune in meme GT+, cioè una donna dall'atteggiamento sessualizzato che parla ad un uomo all'interno di una macchina, scena che facilmente può essere identificata come una prostituta che parla a un potenziale cliente. Altri due esempi sono i meme, sempre FP, n. 558 e 10038 (fig. 5). In entrambi i casi l'immagine è neutrale, ma nel primo caso il testo ha due parole

²⁷ Numerosi studi di linguistica computazionale si sono concentrati sull'implicitezza del contenuto di sarcasmo e ironia, per esempio nell'edizione di Ewallta del 2018 è stato proposto un *task* sul rilevamento dell'ironia nei tweet (Cignarella 2018).

che possono essere soggette ad *unintended bias*, *woman* e *dick*, e nel secondo caso si tratta di un meme con due immagini, una delle quali è un'immagine *reaction shot* che all'interno del dataset viene spesso usata per veicolare contenuti misogini, combinata con un testo contenente *women* e *men*.

Man: Hey... How much for a hand job?
 Hooker: \$10. You want one?
 Man: Nope... I was just curious how much I save when I do it myself.



Figura 5. Meme n. 2082, n. 558, n. 10038.

2. Casi di sensibilizzazione sulla violenza maschile, che spesso associano un testo volto a responsabilizzare sull'ambito a un'immagine che presenta una scena di violenza più o meno esplicita, come un uomo che picchia una donna, o una donna con tumefazioni sul viso. Ne sono esempi i meme n. 8016 e 8370 (fig. 6), anche in questi casi riconosciuti come misogini, quando in realtà non lo sono. Oltre alle immagini, che sono inequivocabilmente violente, i testi contengono parole che possono essere ambigue, quali *sex* e *sexual assault* nel meme n. 8016, o la combinazione di *woman* e *sexual assault* nel n. 8370. Ci sono però anche alcuni casi in cui l'immagine non è violenta, come nel meme n. 1054 (fig. 6), che raffigura una foto a mezzobusto di un uomo, accompagnata da una lunga citazione dello stesso.²⁸ Alcune delle parole di questo testo sono *sperm*, *groceries* e *house*, che forse hanno potuto confondere i modelli. Un caso opposto, sempre FP, dove l'immagine non è violenta, ma il testo contiene un pattern ricorrente nei meme misogini, è il numero 3607 (fig. 6), dal testo "X WHAT DO YOU CALL A WOMAN IN REVEALING CLOTHES? imgflip.com HER NAME", con una doppia immagine del primo piano di un ragazzo prima dallo sguardo

²⁸ Riporto il testo della citazione: "I think women are foolish to pretend they are equal to men. They are far superior and always have been. Whatever you give a woman, she will make greater. If you give her sperm, she will give you a baby. If you give her a house, she will give you a home. If you give her groceries, she will give you a meal. If you give her a smile, she will give you her heart. She multiplies and enlarges what is given to her. So, if you give her any crap, be ready to receive a ton of shit!"

sorridente e poi un po' confuso. Il pattern *What do you call a* è presente sia nei meme GT+ che in quelli GT-, ma con una differenza sostanziale: nei meme GT+ il complemento oggetto è sempre una donna,²⁹ mentre in quelli GT- i complementi oggetto sono o animali, o uomini (in un solo caso), e in un caso soltanto una donna (meme n. 3550, in fig. 6). In quest'ultimo caso però la parola che segue *what do you call a* non è direttamente un termine femminile, ma due aggettivi seguiti poi da *prostitute*: "WHAT DO YOU CALL A CHUBBY CHINESE PROSTITUTE".³⁰ Quindi nonostante il testo veicoli un messaggio contro l'oggettificazione della donna, e l'immagine sia neutrale, il pattern *what do you call a woman* ha probabilmente comportato il mancato riconoscimento del meme come non misogino.

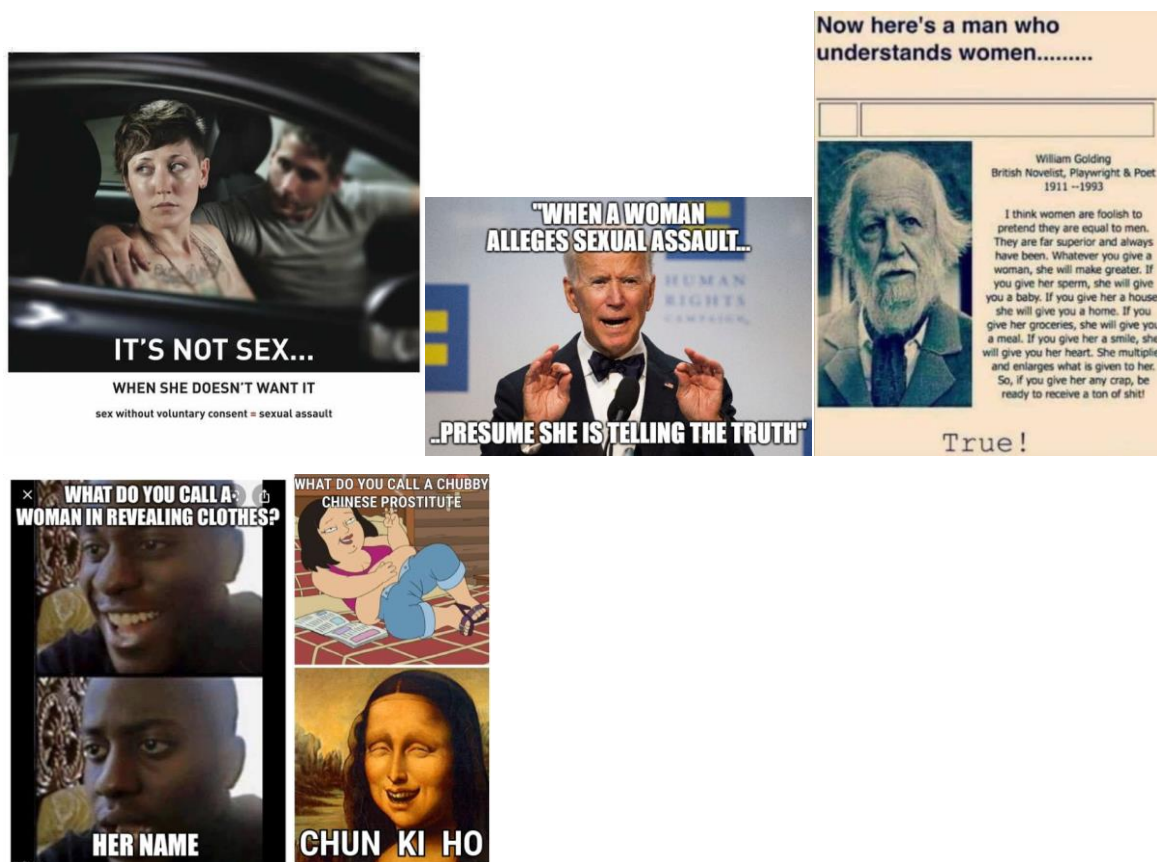


Figura 5. Meme n. 8016, n. 8370, n. 1054, n. 3607, n. 3550.

3. Non riconoscimento del genere del referente, come i meme FP n. 1828 e n. 5308, che recitano rispettivamente "MAKE ME A SANDWICH I MADE HIM A SUB", e "She needs to tell those guys to get back in the kitchen and make her a damn sandwich",

²⁹ In un meme la combinazione di parole è presente due volte, la prima volta con complemento oggetto un uomo, ma la seconda volta il complemento oggetto è una donna.

³⁰ Anche in GT+ ci sono casi in cui la parola successiva a *call a* non è direttamente un sostantivo femminile ma un aggettivo (per esempio nei meme n. 10949 e 247), ma l'immagine è quella di una donna, e sono presenti anche parole come nel primo caso *fat woman* e *rape* e nel secondo caso *dumb blonde*.

dove una delle due immagini rappresenta una donna in posizione di potere rispetto all'uomo, mentre l'altra è neutrale. Dato che la struttura *make a sandwich*³¹ è molto usata nei meme misogini, un bias non intenzionale dei modelli potrebbe avere assegnato a questa combinazione di parole un peso molto alto verso la classificazione misogina di contenuti simili, portando le macchine a ignorare il resto della struttura sintattica e dei referenti collegati ai pronomi. Un altro caso esemplificativo in cui nella errata classificazione è coinvolto solo in testo è il meme FP n. 8126. L'immagine del meme raffigura il cosiddetto *Socially Awkward Penguin* (Grundlingh 2017), e che quindi non fornisce indicazioni sulle persone coinvolte nell'atto comunicativo. Per comprendere il meme è fondamentale il testo, "HAD SEX FOR THE FIRST TIME EVER. I ASKED HER IF SHE LIKED IT SHE: "IF YOU WERE A MALE PROSTITUTE, I WOULD ASK FOR MY MONEY BACK" MEMEFUL.COM": l'unico elemento che fornisce indicazioni sul genere di *prostitute* è l'aggettivo preposto *male*, che elimina ogni possibilità di contenuto misogino, ma evidentemente questa combinazione non è stata riconosciuta dai modelli.



Figura 6. Meme n. 1828, n. 5308, n. 8126.

4. Immagini di anime: anche in questo caso si tratta di un bias verso la classificazione misogina. Alcuni meme misogini sono formati da disegni di anime che rappresentano donne con un corpo sessualizzato, ed è quindi probabile che il modello che si occupa del riconoscimento delle immagini riconduca lo stile di disegno tipico dell'anime a un contenuto misogino, come nel meme n. 8172 (dove sono anche scritte

³¹ Proprio lo stereotipo di ordinare alle donne di fare un panino, chiaramente a un uomo, viene individuato da Citron (2009), insieme all'invito a ritornare in cucina, come uno fra i più comuni cliché misogini atti a rinforzare gli stereotipi di genere.

le parole *man* e *girl*), o nel meme n. 5467 (che contiene anche le parole *girl* e *wife*), in Figura 8.



Figura 7. meme n. 8172 e n. 5467.

5. Non corrispondenza fra contenuto misogino o meno del testo e contenuto misogino o meno dell'immagine. Questa categoria di errori spesso si combina con altri fattori. Per esempio, i tre meme FN n. 2945, 8557 e 2636, che hanno tutti e tre immagini simili e sono tutti varianti di una stessa barzelletta, e quindi la non corrispondenza fra misoginia del testo e dell'immagine si combina in questi casi con il mancato riconoscimento dell'ironia. Le immagini sono del tutto neutrali, con in un caso una mano che mostra delle monete trovate scavando nel terreno, in un altro caso il disegno molto stilizzato di un uomo che si prostra, e nell'ultimo caso la combinazione delle due. Il meme n. 2131 testimonia invece il caso opposto: l'immagine di Hilary Clinton dalla dentatura distorta, affiancata dalla foto di un cavallo, è accompagnata da un testo neutrale ("TWINS SEPARATED AT BIRTH ?"). Ci sono poi alcuni meme in cui le immagini sono molto esplicite ma il testo è del tutto neutrale, come nel caso del meme FP n. 8480, di cui non riporto l'immagine poiché raffigura una scena pornografica, il cui testo recita "Find yourself someone who will rub your neck after a hard day", oppure il meme n. 10997 la cui immagine contiene dei disegni di una donna nuda, seppure non sessualizzata, con un testo anche in questo caso non misogino.

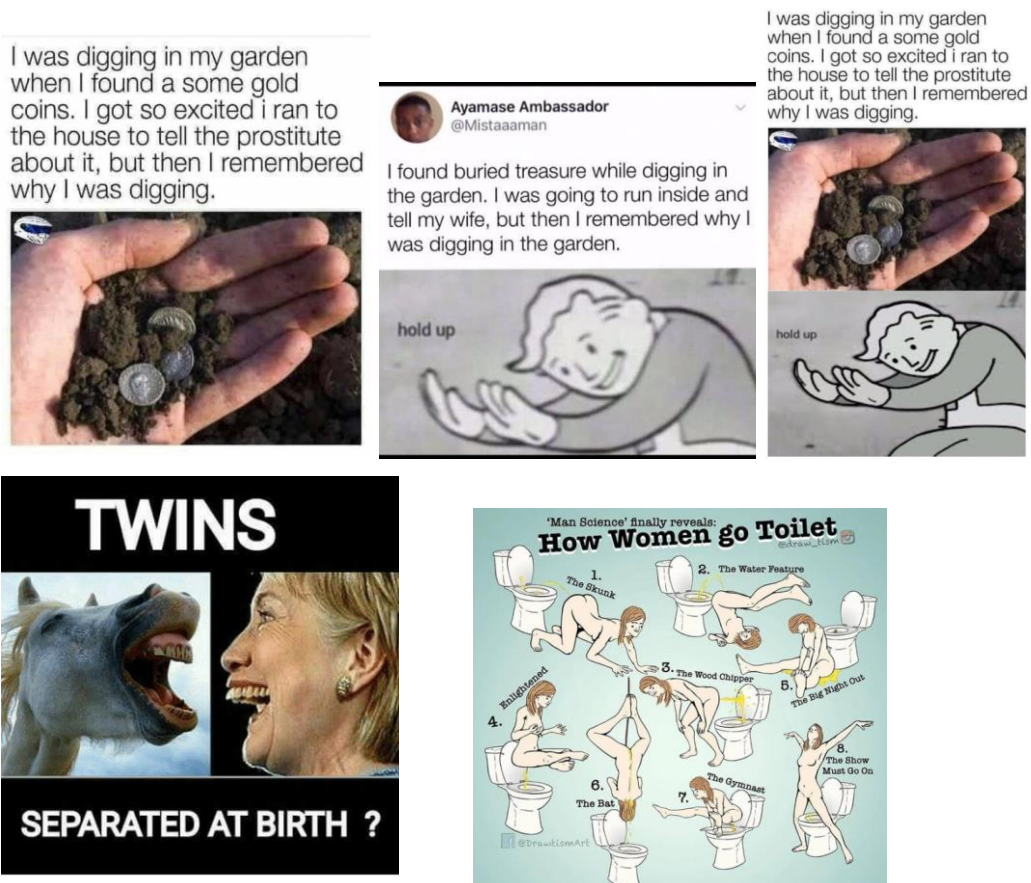


Figura 8. Meme n. 2945, n. 8557, n. 2636, n. 2131, e n. 10997.

6. Conoscenze enciclopediche su personaggi famosi o pratiche culturali.³² Il meme FN n. 11226 recita "Feminist: "All men are pigs" Men who read Animal Farm: You're goddamn right.", con l'immagine del primo piano di un attore dallo sguardo consapevole e sicuro. Senza conoscere il romanzo *La fattoria degli animali* è difficile comprendere il significato del meme. Il meme n. 3468 ha come immagine il disegno di un anziano arrabbiato incorniciato dalla frase "IF YOU HAVE AN OPINION TO SHARE WHY DON'T YOU GO IN THE KITCHEN AND PUT IT IN A BUNT CAKE": se non si è a conoscenza del fatto che quello di andare in cucina nella cultura patriarcale è un invito che si rivolge solo donne per sminuirle e restringere il loro raggio d'azione, il meme non può essere riconosciuto come misogino.

³² Le categorie numero 6 e 2 sono individuate anche da Tao & Kim (2022).

Feminist: "All men are pigs"
Men who read Animal Farm:

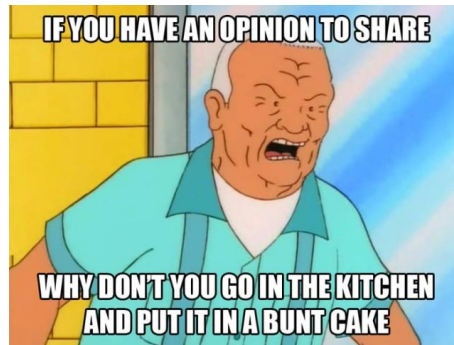


Figura 9. Meme n.11226, e n. 3468.

2.3.4 Individuazione di *identity term*

Per l'estrazione degli *identity term*, che intendo come nomi che hanno un referente umano,³³ ho usato le liste di frequenza per nomi. Il primo tentativo di selezione, etichettatura ed estrazione degli *identity term* ha messo in luce alcune problematiche nella metodologia. La prima criticità è relativa ai dati, ossia le trascrizioni dei testi non normalizzate e non pulite, che non consentono un'analisi del dato ottimale, fra gli altri motivi perché avere due forme di uno stesso lemma divise per numero, oltre che a seconda dei caratteri maiuscoli o minuscoli con cui sono scritte, rende più macchinoso il processo di ricerca della parola. Le categorie che avevo individuato si applicano ai nomi che hanno un referente umano, e sono:

- a. femminili
- b. maschili
- c. non espliciti per genere
- d. offensivi
- e. parentela di sangue
- f. relazione con un uomo
- g. relazione con una donna
- h. nomi propri
- i. nomi di professione.

Fra queste etichette, ho eliminato quella dei nomi non espliciti per genere, e quella dei nomi propri. Ho deciso di ignorare i nomi propri sia perché nei testi ce ne sono pochi, sia perché sono difficilmente individuabili, non avendo abbastanza competenza in inglese, mentre per quanto riguarda i nomi non espliciti per genere è una categoria

³³ Ne parlo meglio nel capitolo 3.

superflua, dato che quelle parole non caratterizzate dal genere grammaticale che mi interessa evidenziare sono comunque già comprese nelle altre categorie, e sarebbero quindi una ripetizione. Ho poi accorpato le categorie *f.* e *g.* La prima idea sulla categoria comprendente i nomi di professione era di individuare il genere dei referenti per ogni nome: nelle barzellette (e buona parte dei meme ha intento umoristico) è comune che siano gli uomini ad essere i referenti sia dei mestieri più prestigiosi, sia di mestieri molto diversificati fra loro, mentre le donne restano limitate a pochi (Chiaro 2018). Volevo quindi verificare se anche all'interno dei meme, che nascono con intento umoristico, ci fosse qualche riscontro di questo tipo. Una ricerca del genere però sarebbe stata troppo dispersiva, oltre che poco utile ai fini del mio lavoro.

L'individuazione dei nomi è stata fatta manualmente, scorrendo le liste di frequenza per nomi. Dato che le liste dei dataset GT+ e GT- sono composte da migliaia di righe, ho deciso di limitarmi a un campione che potesse essere rappresentativo, e ho usato il numero di frequenza assoluta dei nomi, fermandomi a 5 compreso. Al contrario, le liste di FP e FN sono più contenute, e le ho quindi analizzate per intero. Nella fase finale ho poi modificato la scelta del campione.

2.2.5 Misure di impatto

Data la grande quantità di dati, è stato necessario cercare un modo per ordinare gli elementi delle tabelle in maniera funzionale. Poiché l'obiettivo del lavoro è comprendere la ragione degli errori dei modelli, è importante ordinare le liste in modo da evidenziare quali sono le unità linguistiche più problematiche. A questo scopo ho calcolato i rapporti delle frequenze assolute dell'elemento target fra dataset con classificazione corretta e il dataset con classificazione sbagliata, dividendo quindi le frequenze assolute di un elemento in FN per le frequenze assolute dello stesso elemento in GT+, e le frequenze assolute di un elemento in FP per le frequenze assolute dello stesso elemento in GT-. Questo coefficiente dà un valore che va da 0 a 1 alla frequenza dell'errore di quello specifico elemento. Più il coefficiente è simile a 1, più alto è il numero delle volte in cui l'elemento viene classificato in modo sbagliato rispetto al totale delle occorrenze di quell'elemento nel dataset con predizione corrispondente all'annotazione *gold*. Al contrario, più il valore del coefficiente è vicino allo 0, e meno spesso quell'elemento viene sbagliato. Le prime analisi sulle liste di frequenza di type, lemmi, nomi e aggettivi, e quelle di bigrammi, trigrammi e tetragrammi, sono state fatte ordinando il coefficiente

dei rapporti fra FN (la cui *ground truth* è positiva) e GT+, e fra FP (la cui *ground truth* è negativa) e GT-.

Questo tipo di rapporto non tiene però conto dell'impatto che ha l'errore all'interno dei modelli. Facendo un esempio, il bigramma *I see* compare 20 volte nel dataset GT+, e 5 volte nel dataset FN, quindi il coefficiente sarà 0,25; il bigramma *be what* invece compare 12 volte in GT- e 3 volte in FP, e il coefficiente ha un valore anche in questo caso di 0,25, nonostante la rilevanza dell'errore dei due bigrammi è diversa, perché un errore commesso 5 volte su 20 ha un impatto maggiore di un errore commesso 3 volte su 12. Per sopperire a questa mancanza di informazione è stato usato un altro coefficiente empirico, che combina precisione, richiamo, media armonia macro e micro, di cui si parla al paragrafo 4.2.

2.2.6 Analisi della somiglianza fra i testi

Un altro strumento che ho usato, ma che non ha portato a informazioni particolarmente utili, è stato Google Colaboratory, la piattaforma di Google che permette di usare codici Python online. Ho usato tre comandi che servono per restituire gli elementi comuni o esclusivi fra due file. Ho dato come input le liste di frequenza per type, e gli n-grammi, confrontando alcune delle combinazioni possibili fra i quattro dataset.

Tutti gli elementi contenuti nei dataset con classificazione errata sono presenti anche nei corrispettivi dataset GT+ e GT-, per cui non esistono elementi esclusivi di FP ed FN rispetto a GT+ e GT-. Ho quindi estratto gli elementi esclusivi e comuni fra GT+ e GT-, ma questa operazione si è rivelata poco utile per tre motivi principali. Il primo è che per capire le motivazioni per cui i modelli hanno sbagliato il compito di classificazione bisogna porre l'attenzione su FP ed FN rispetto alle GT, piuttosto che sulle GT stesse. Il secondo motivo riguarda le dimensioni di questi output, che contengono migliaia di elementi, ed è quindi difficile riuscire a catturare delle informazioni interessanti da liste così lunghe. Infine, questi comandi hanno restituito soltanto le liste degli elementi comuni ed esclusivi ma non le rispettive frequenze assolute, che sono invece necessarie e per cui Excel si rivela uno strumento più utile. Ho confrontato fra loro anche i due dataset classificati in modo errato dai modelli, ma anche in questo caso Excel si è rivelato uno strumento più idoneo.

2.3 Preparazione definitiva dei corpora

La fase descritta precedentemente è servita a capire su quali punti concentrare l'attenzione e quali strategie abbandonare. Una delle necessità emerse per ottimizzare le analisi sui testi è normalizzare e ripulire almeno in parte i dati.

2.3.1 *Preprocessing* del testo

Le prime perplessità derivate dall'uso del testo grezzo sono emerse nella prima fase di lavoro sullo studio degli *identity term*. Dovendo confrontare la presenza delle parole nei vari dataset, è più informativo confrontare i lemmi, e non due forme della stessa parola distinte per numero, anche perché è improbabile che un modello neurale rappresenti in modo diverso il singolare e il plurale di uno stesso lemma. Anche la *case sensitivity* non è utile per valutare dei dati quantitativi, dato che il modello usato per l'apprendimento del linguaggio non distingue fra caratteri maiuscoli e minuscoli.³⁴ Quindi il testo è stato normalizzato, anche se in modo non invasivo, su tre fronti:

- a. normalizzazione dei caratteri (Sivanaiah, Rajendram & Mirnalinee 2022);
- b. lemmatizzazione e correzione degli errori del lemmatizzatore (Sivanaiah, Rajendram & Mirnalinee 2022; Barnwal, Kumar & Pamula 2022);
- c. uso di un'espressione regolare per eliminare i simboli e parte della punteggiatura (Rao & Rao 2022; Barnwal, Kumar & Pamula 2022; Sivanaiah, Rajendram & Mirnalinee 2022; Frenda 2022).

Questi sono fra i metodi di pulitura del testo più diffusi nei compiti di *natural language processing*. Fra gli altri, il manuale di Jurafsky e Martin (2008) cita fra le tecniche di preparazione del testo il *case folding* e la lemmatizzazione, oltre che l'uso di espressioni regolari per necessità specifiche.

La prima normalizzazione che ho fatto è stata la trasposizione dei testi dei meme in caratteri minuscoli, di seguito ho usato il lemmatizzatore, e infine l'espressione regolare.

³⁴ La conferma sia della non sensibilità al carattere maiuscolo o minuscolo, sia alle forme plurali, si vede nelle liste delle parole simili per una parola target (si veda la lista 1 in appendice), in cui le stesse parole aventi però differenze di carattere maiuscolo o minuscolo hanno sistematicamente la stessa distanza dalla parola target, così come le forme plurali.

2.3.2 Processo di lemmatizzazione

Ho testato due lemmatizzatori, che hanno entrambi commesso alcuni errori. Il primo lemmatizzatore testato è stato quello disponibile in *Spacy*, una libreria per lo svolgimento di alcuni fra i principali compiti di *natural language processing* disponibile per Python.³⁵ Il secondo è invece *Spacy Stanza*, un pacchetto per Python che contiene strumenti per analisi sul linguaggio naturale.³⁶ Per testare la correttezza della lemmatizzazione ho controllato circa cinquanta meme, prendendo nota delle tipologie di errori che ho riscontrato. Gli errori sono stati successivamente corretti con l'editor di testo *TextMax2*.

Spacy non lemmatizza in modo coerente le negazioni verbali abbreviate del tipo *'t*, che in alcuni casi ha riconosciuto e in altri casi no: la sequenza *can't* viene lemmatizzato sotto la forma *cant*, mentre una sequenza come *wasn't* viene analizzata correttamente e divisa sotto i due lemmi *be* e *not*. Cercando altre forme verbali negative non ho trovato altre lemmatizzazioni errate del tipo di *cant*. La forma *Ill* viene lemmatizzata come *ill*, mentre le altre persone, per esempio *you'll*, sono correttamente riconosciute come *you* e *will*. Un altro problema riguarda i pronomi soggetto e gli aggettivi possessivi, che non vengono mai distinti; per esempio, il testo "WHEN YOU WAKE UP ON SATURDAY AND YOUR HOUSE IS CLEAN" viene lemmatizzato "when you wake up on Saturday and you house be clean", senza distinzione fra l'aggettivo possessivo *your* e il pronome soggetto *you*. Si comporta invece correttamente con i pronomi possessivi, che hanno lemmi propri.

Anche *Spacy Stanza* non è privo di problematiche, ma si presentano tutte nel trattamento delle forme verbali contratte dei verbi *to be* e *to have*, e risultano quindi più gestibili da correggere. Dal momento che correggere sistematicamente la lemmatizzazione degli aggettivi possessivi è più complicato, ho scartato il primo lemmatizzatore, *Spacy*.

Per correggere gli errori, ho usato la funzione "trova e sostituisci" di *TextMax2*, e una volta apportate le sostituzioni, ho sempre controllato manualmente che non ci fossero effetti indesiderati di modifica del testo. Di seguito le correzioni apportate:

- a. *I've, Ive, I ve*: lemmatizzati come *I ve*, li ho corretti e sostituiti come *I have*

³⁵ Maggiori informazioni sulle caratteristiche di *Spacy* sono disponibili al link <https://spacy.io/api/edittreelemmatizer>

³⁶ Per maggiori dettagli si può consultare il link <https://stanfordnlp.github.io/stanza/>

- b. You've, youve: lemmatizzati come *you ve*, corretti come *you have*
- c. We've (le forme *weve* o *we ve* non sono attestate): lemmatizzato come *we ve*, corretto *we have*
- d. They've (le altre forme non sono attestate): lemmatizzato *they ve*, corretto con *they have*
- e. Id: compare 2 volte come forma scorretta di *I'd*; è stato lemmatizzato sotto *I d*, e l'ho corretto come *I would*
- f. Youd: lemmatizzato *you d*, corretto con *you would*
- g. Im: lemmatizzato *I m*, corretto *I be*
- h. Youre: lemmatizzato *you re*, corretto con *you be*
- i. Shes: lemmatizzato *she s*, corretto con *she be*
- j. Its: in 66 meme (ma le occorrenze sono di più perché a volte compaiono più volte in uno stesso meme, ma non ho questo numero...) non si tratta del pronome possessivo ma di una disgrafia per *it's*; ho quindi corretto la lemmatizzazione in *its* di questi casi, e l'ho corretta con *it be*
- k. Thats: lemmatizzato come *that s*, sostituito con *that be*. Compare 17 volte
- l. Theyre: compare solo una volta, lemmatizzato come *they re*; corretto come *they be*
- m. U: forma colloquiale per *you*, lemmatizzato come *u* e corretto come *you*.

Non ho controllato la possibile disgrafia *were* per *we're*, perché ci sono 132 occorrenze di questa forma (che è il *past simple* di *to be*), che avrei dovuto controllare singolarmente per verificare se si trattasse del *past simple were* o della disgrafia *we're*, ma soprattutto perché le forme verbali di prima e terza persona plurale sono molto poco attestate nelle trascrizioni dei testi dei meme, e ho quindi ritenuto il dato trascurabile. Controllando un campione di casi, non ho comunque trovato nessuna forma di *present simple*, ma solo il *past simple*.

Le uniche altre grafie non standard che ho controllato sono *havin/havin'* e *bein/bein'*, che però non sono presenti. Altri casi di slang come *gotta*, *gonna* o *watcha* sono riconosciuti correttamente dal lemmatizzatore stesso.

Spacy-Stanza usa due lemmi diversi per le parole che hanno forme diverse a seconda del genere, come *waiter* e *waitress*, o *actor* e *actress*. Questa caratteristica è stata molto utile nel processo di estrazione degli *identity term*, che non è stato quindi inficiato dalla lemmatizzazione.

2.3.3 Eliminazione di alcuni caratteri

I testi dei meme abbondano di simboli e segni diacritici, che disturbano i processi di estrazione di dati da SketchEngine, intervenendo nelle liste di frequenza e negli n-grammi. Un caso limite sono le numerose stringhe simili a *íäóí_íÿíæíœçíœí_íœçíœ*. Ho quindi creato un'espressione regolare che eliminasse tutti i simboli e i segni diacritici che non facessero parte dei grafemi dell'inglese, né dei segni di interpunzione, usando anche in questo caso TextMax2. L'espressione è:

$$[^a-z .,:?!/()0-9][\v]{2,}$$

Questa espressione mantiene nel testo:

- I caratteri alfanumerici dalla *a* alla *z* e da *0* a *9*;
- I segni di interpunzione *.,:?!/*;
- Le parentesi tonde e la barra obliqua, che può essere usata nei testi con significato alternativo o oppositivo fra due o più parole. Non ho incluso invece le parentesi quadre e graffe perché non hanno la stessa funzione della parentesi tonda, ossia di introdurre un inciso, senza che la struttura sintattica venga intaccata;
- Gli *a capo*.

Invece, l'espressione elimina:

- I caratteri alfabetici accentati;
- I caratteri non appartenenti alla codifica di caratteri ISO 8859-1, come *œçșțăîâæ*;
- I simboli matematici, per esempio $\infty \sqrt{f} \Sigma \%$;
- Gli altri simboli, come $\#^{\circ} \text{TM} \text{€} \text{†} \text{©} \text{~} \text{\&}$;
- I segni di interpunzione e le parentesi che non siano *.,:?!/()*.

Ho sostituito tutto quello che non rientrasse nella lista dei caratteri permessi dall'espressione regolare con uno spazio vuoto: così facendo ho mantenuto una separazione nei casi in cui un carattere eliminato fosse inserito senza spaziature fra due parole.

Riguardo ai caratteri alfabetici accentati, in inglese questi non sono previsti, tranne rarissimi casi. Nei meme ci sono soltanto due parole lessicali che contengono caratteri accentati: *fiancé* e la neoformazione *divorcé*. Eliminando la vocale, non viene comunque eliminata la parola, che risulta quindi facilmente ricostruibile.

2.3.4 Varietà lessicale dei dataset

Un metodo comune di studio delle frequenze all'interno di un testo è il modo in cui *type*, le parole tipo, si distribuiscono rispetto ai token, le parole unità. L'indice per misurare la varietà lessicale di un testo è il rapporto tipo/unità, meglio noto come *type/token ratio* (TTR), che si calcola dividendo i *type* per i token, dove più il valore si avvicina a 1 e più il lessico è ricco, più si avvicina allo 0 e più il lessico è povero (Lenci, Montemagni & Pirelli 2020). Secondo i dati calcolati da SketchEngine sui *corpora* delle trascrizioni dei testi grezzi, il *corpus* di testi dei meme misogini GT+ è composto da 92.967 token, e da 16.209 *type*, che comprendono anche le non-parole e quindi i simboli o le stringhe di caratteri non appartenenti all'alfabeto latino. Il *corpus* GT- ha invece 115.778 token e 20.451 *type*. Le due TTR sono uguali: GT+ ha un indice di 0,174, e GT- di 0,176. Possiamo quindi dire che a parità di meme, che sono 5.000 sia per GT+ che per GT-, la lunghezza media dei testi dei meme non misogini è più lunga rispetto a quella dei meme misogini. Nonostante quindi i meme misogini si concentrino su un solo argomento, le donne, la loro varietà lessicale non è minore dei meme che trattano di temi molto più vari: la creatività dei meme misogini non trova riscontro nel numero di *identity term*, che non è maggiore per la GT+ rispetto alla GT-. Potrebbe dipendere allora dalle forme grafiche non standard, e dalle parole appartenenti al registro slang, o simboli, che potrebbero essere maggiori per i meme misogini. Entrambi i risultati sono in contrasto con quanto trovato da Frenda (2022) nel suo studio sui tweet di argomento misogino e offensivo, secondo cui i dataset di tweet non misogini e non offensivi risultano avere una varietà lessicale maggiore,³⁷ e risultano essere meno lunghi di quelli sessisti e offensivi. Per i *corpora* FP e FN i valori del rapporto tipi/unità è rispettivamente di 0,334 e 0,424 (si vedano le tabelle 1 e 2 per il numero dei *type* e dei token dei *corpora*), ma questo valore non è un indice informativo in quanto i due dataset sono numericamente sbilanciati: FP ha il doppio dei token di FN, quindi è possibile che, più un *corpus* è limitato, più è probabile che la varietà lessicale, se calcolata in rapporto alle parole unità, sia maggiore. Il discorso non vale invece per i due dataset di GT, che sono piuttosto bilanciati fra loro.

³⁷ L'autrice spiega la minore varietà lessicale dei *tweet* misogini e sessisti con il fatto che questi trattano di un unico argomento, e con la possibilità che contengano un numero maggiore di insulti, che sono un gruppo di parole limitato, dato che mirano a offendere o urtare il bersaglio.

Con i processi di pulizia il *corpus* GT+ diventa invece di 89.702 token, e 9.785 *type*, mentre il *corpus* GT- di 110.417 token e 12.363 *type*. Quindi, l'espressione regolare ha eliminato nel caso di GT+ 3.262 caratteri, e nel caso di GT- 5.361 caratteri; la differenza fra GT+ e GT- è minima, con un lieve sbilanciamento verso GT- che conta, in proporzione rispetto al numero totale di token, ben pochi caratteri speciali e simboli in più rispetto a GT+. Nonostante l'argomento molto diverso, l'uso di simboli e caratteri speciali rimane quantitativamente quasi uguale. I meme classificati in modo sbagliato come positivi, i FP, hanno invece in proporzione una percentuale più alta, anche in questo caso di poco, rispetto ai FN.

Tabella 1. N. di token e di type nei testi grezzi.

TESTI GREZZI	token	type
GT+	9.2967	16.209
GT-	115.778	20.451
FP	11.992	4.006
FN	5.138	2.182

Tabella 2. N. di token e di type nei testi ripuliti.

TESTI PULITI	token	type
GT+	89.702	9.785
GT-	110.417	12.363
FP	11.385	2.647
FN	4.947	1.534

Capitolo 3: *Identity term*

In questo capitolo mi concentro sugli *identity term*, intesi come nomi che hanno un referente umano. Parlo dei criteri con cui sono stati selezionati (paragrafo 3.1), descrivo le categorie in cui sono stati suddivisi (paragrafo 3.2), e do indicazioni sulla popolazione da cui sono stati estratti e su come ho svolto il processo di estrazione e di annotazione (paragrafo 3.3). Nel paragrafo 3.4 parlo delle ulteriori normalizzazioni che ho operato sul testo per rendere più efficace l'analisi sugli *identity term*, analisi che viene descritta dal paragrafo 3.5 al paragrafo 3.5.7.

3.1 Scelta delle categorie

Per cercare possibili differenze fra contenuti misogini e contenuti non misogini è rilevante individuare quali sono i soggetti che vengono rappresentati in questi contenuti, e come. La rappresentazione di un soggetto viene espressa anche attraverso i nomi che il parlante sceglie per riferirsi al soggetto, e questa scelta non è mai neutrale (Sabatini 1987; Simpson 1993). Per questo ho selezionato una serie di parole che avessero un referente umano. In un'ottica di analisi dell'errore dei modelli, differenziare l'uso di questi *identity term* a seconda del dataset in cui si trovano serve per scoprire se ci sono o meno differenze nella loro distribuzione, in particolare nei gruppi FP e FN: per esempio, è probabile che parole offensive riferite a donne, come *hoe*, siano più spesso contenute in meme misogini, e che quindi le macchine attribuiscono a quelle parole un significato misogino che porti a una tendenza a deviare verso la classificazione misogina se in un meme è presente una certa parola, anche quando il contesto non veicola messaggi di odio verso le donne (Zhou et al. 2022; Muti & Barrón-Cedeño 2020).

Per il processo di annotazione ho usato il metodo *open coding* (Khandkar 2009). Questo è un metodo empirico usato nelle analisi qualitative di contenuti testuali, e consiste nell'etichettatura di concetti, nella definizione e nello sviluppo di categorie, che vengono modificate e raffinate via via che si prosegue con l'annotazione dei dati: nel corso dell'annotazione è probabile che si riconoscano concetti (cioè etichette) nuovi, o che concetti vecchi assumano significati diversi, più ampi o più stretti. Procedendo con la categorizzazione, a seconda delle parole che incontravo, ho eliminato o modificato alcune etichette. La selezione delle categorie si basa sulla letteratura scientifica e

sull'osservazione di alcune parole particolarmente ricorrenti nei testi dei meme: in particolare, ho preso spunto da categorie individuate da alcuni partecipanti alle campagne di valutazione di SemEval, e ho aggiunto delle categorie modellate su alcune delle tattiche di *hate speech* misogino più comuni, e altra letteratura che tratta il sessismo della lingua.

Nell'ambito dei congressi di IberEval, e in particolare nel *task* di detezone e identificazione di tweet misogini di IberEval 2018, il gruppo di ricerca di Frenda e Bilal ha costruito dei lessici che comprendono termini concernenti la sessualità, la volgarità, la femminilità, e il corpo umano (Frenda & Bilal et al. 2018), mentre il gruppo Ex-Lab@UniTo ha identificato alcuni termini su genitali, prostituzione e omosessualità (Pamungkas et al. 2018).

Jennifer Herriman, in un lavoro sul *Brown Corpus*, identifica alcune categorie di aggettivi e nomi che caratterizzano donne e uomini in modo diverso. Fra questi troviamo i nomi di professione, che per gli uomini sono molto più variegati rispetto alle donne, e parole che indicano status coniugali e relazionali, che sono invece più frequenti per le donne (Herriman 1998). Alcune categorie sono modellate su tattiche di *cybersexism*, come il *gender abuse and harassmet* e l'*online threats*, che comportano minacce di morte e intimidazioni, minacce di violenza sessuale, descrizioni di fantasie di violenza sessuale, insulti sessualmente espliciti.³⁸ Questi tipi di contenuti possono essere veicolati con l'uso di parole esplicitamente violente, come *to rape* o *hoe*, ma anche in modi meno evidenti, per esempio con nomi propri che subiscono un ampliamento semantico per significare una categoria di donne con caratteristiche poco apprezzabili, come *Stacy* o *Karen*.

3.2 Descrizione delle etichette

A ogni categoria identificata ho assegnato un'etichetta; una parola può rientrare in una o più categorie e quindi le etichette possono essere combinate fra loro. Le prime etichette sono:

³⁸ Si veda Poland (2016), dove i tipi di *cybersexism* vengono decritti nelle pagine 37-60, oppure Spallaccia (2017) a p. 29.

- Donna: il referente della parola è sempre una donna, o perché la parola a livello grammaticale è solo femminile (per esempio *actress* e *whore*), oppure perché nel dataset di riferimento la parola è sistematicamente associata a una donna
- Uomo: il referente è un uomo
- Generico: il genere del referente non è esplicito, oppure indica entrambi i generi
- Lavoro: la parola è un nome di professione
- Parentela: la parola indica un grado di parentela, che sia però di sangue, come *daughter*, *mother*, *nephew*
- Relazioni sentimentali: parole che indicano relazioni sentimentali, come *wife* e *boyfriend*
- Ambito estetico e sessuale, o sessuale: parole che connotano una persona in base a qualità estetiche e di appetibilità sessuale, come *Chad*, *milf* o *teen*, oppure che caratterizzano un individuo in base ad attributi sessuali, come *virgin*, *hooker* o *stripper*. *Estetica* non è una categoria a parte perché ho trovato soltanto un nome, *bodybuilder*, in cui il significato estetico non contenesse anche sfumature sessuali (al contrario, ci sono termini che sono riconducibili all'ambito strettamente sessuale)
- Offensivo: insulti più o meno espliciti.

3.3 Popolazione e processo di etichettatura

Per la selezione degli *identity term* ho usato le liste di frequenza dei nomi estratte con SketchEngine e visualizzate su Excel, in modo da poter aggiungere le etichette nella colonna a sinistra dell'elenco dei nomi. Il testo dato in input è quello lemmatizzato e ripulito con espressione regolare. Le liste hanno dimensioni molto diverse a seconda dei corpora: quelle dei dataset GT+ e GT- hanno migliaia di righe, e quindi non sarebbe stato possibile scorrerle tutte, mentre le liste di FP e FN possono essere analizzate per intero. Ho deciso quindi di esaminare manualmente tutti i nomi di FP e FN, ma di arrivare alla riga 868 di GT+ e GT-, perché sono le righe totali della lista di FP, la più ristretta fra le quattro. Nel caso in cui in FP e FN ci fossero dei nomi non presenti in GT+ e GT-, li ho sempre cercati manualmente nelle restanti righe delle liste, così da poter fare un confronto almeno per i termini presenti nei dataset classificati in modo sbagliato.

Una volta individuati gli *identity term* a seconda dei gruppi di categorie, li ho rappresentati graficamente, sempre differenziandoli per dataset.

Scorrendo le liste di frequenza dei nomi, ho trovato alcune parole che possono avere un significato ambiguo, e ho quindi controllato nelle trascrizioni complete dei meme tutti i contesti in cui queste compaiono per fare una disambiguazione nei casi in cui fosse necessario. Quando la trascrizione del testo non è stata sufficiente alla disambiguazione, ho usato il meme compreso di immagine. Ho controllato anche quei termini che potrebbero riferirsi a un solo genere. Come si vedrà, non tutte le parole sono usate effettivamente con referenti umani, e altre parole polisemiche vengono usate in alcuni casi per riferirsi a persone, e in altri casi no. In questi casi, ho contato come *identity term* soltanto quei contesti in cui la parola fosse effettivamente riferita a un essere umano, scartando le occorrenze in cui i referenti erano altri.

Alcuni nomi che non hanno un genere grammaticale possono però essere riferiti all'interno dei dataset a un solo genere. Il processo di raggruppamento è stato fatto in base all'uso peculiare di ogni dataset, e non in base alla grammatica. Il caso più esemplare è *dishwasher*, che grammaticalmente ha un referente molto preciso, cioè un oggetto, ma che nel dataset GT+ viene usato come appellativo delle donne: ricade quindi sotto l'etichetta *donne*, ma solo in questo dataset.

Le parole che ho controllato sono:

- Angel: non si riferisce mai a persone reali
- Babe: non indica mai bambini, ma sempre adulti
- Beaver: significa *castoro*, ma può essere anche un termine osceno con cui chiamare i genitali femminili
- Bitch: può essere sia uno dei tanti insulti misogini che si avvalgono della condanna morale contro la libertà sessuale femminile, sia un insulto generico rivolto a entrambi i sessi. Questo doppio significato è perfettamente scandito all'interno dei dataset: in GT+ e FN, quindi i due dataset che per *ground truth* sono misogini, il referente è sempre femminile, mentre in GT- e FP è sia maschile che femminile, e non può essere quindi inserito nel gruppo degli i.t. femminili.
- Bf: è l'abbreviazione tanto di *boyfriend* che di *best friend*
- Bodybuilder: non è mai usato come professione
- Boss: è usato come capo lavorativo, ma anche in scritte su prodotti sullo sfondo delle foto dei meme
- Ceo: ha referenti di entrambi i generi
- Character: non può essere considerato un *identity term*

- Clown: è usato in modo offensivo
- Crush: è usato con tre significati, la persona per cui si ha una cotta, lo scontro di due corpi, e l'innamoramento.
- Cunt: è sia un termine dispregiativo per riferirsi a una donna, che un modo volgare per chiamare i genitali femminili
- Daddy: può essere sia un affettivo di *dad*, sia un affettivo per chiamare il partner maschio
- Dishwasher: è usato anche come appellativo di donna, ma solo in GT+. A livello lessicale, potrebbe essere considerata come la più chiara rappresentazione della cultura oggettificante e della separazione dei ruoli e dei poteri all'interno dell'ordine gerarchico patriarcale nel microcosmo dei meme misogini. Il nome di un oggetto viene usato come sinonimo di metà della popolazione mondiale, e questo termine è scelto in base all'unico ruolo che gli uomini riconoscono alle donne: il lavoro domestico non retribuito.
- Divorcé: neoformazione modellata su *fiancé*, indica una persona divorziata
- Doctor: indica sia il mestiere, che il titolo di studio, ma in alcuni meme si trova anche scritto sulle confezioni di alcuni prodotti da supermercato
- Driver: può essere sia nome di professione, che cognome, che *nomen agentis*, che ricorrere in espressioni come *driver licence*, *driver lessons* e *driver side*
- Editor: non si riferisce mai a persone
- Female: è sia riferito a donne, che ad animali femmine
- Feminist: il genere dei referenti è ben ripartito, solo femminili in GT+, e non espliciti in GT- e FN.
- Fiancé: è attestato molto poco, e ha referenti diversi a seconda dei dataset. È solo femminile in GT+ e FN, e maschile in GT-
- Honey: è usato sia per l'alimento prodotto dalle api, che come appellativo affettivo
- Judge: in alcuni casi all'interno della lista dei nomi è in realtà un verbo
- Karen: non viene mai usato come nome proprio, ma solo come appellativo offensivo per indicare una categoria di donne considerate fastidiose perché ignoranti, razziste, supponenti e arroganti
- King: ben poche volte si riferisce alla carica reale, ma nella maggior parte dei casi fa parte del nome del marchio *Burger King*, oppure soprannomi di personaggi più o meno famosi

- Leader: in alcuni contesti indica un lavoro, in altri no (per esempio in *leader board* o *Christian leader*)
- Lord: può indicare Dio, ma molto più spesso fa parte del titolo della una famosa saga *fantasy* "The Lord of the Rings"
- Maker: nella maggior parte dei casi si trova nell'espressione *sandwich maker*, in altri casi (nelle descrizioni di un prodotto, un prisma, su un sito di vendita online) *rainbow maker*
- Male: ha solo referenti umani
- Mama: si trova solo nella frase usata in senso umoristico *yo mama is so fat/so ugly*
- Master: non è mai riferito a persone
- Model: si trova sia con il significato di *caso esemplare*, come lavoro, e in quest'ultimo caso ha solo referente femminile
- Motherfucker: ha sempre referente umano, tranne in un caso
- Mr: viene usato come appellativo da anteporre a un nome, ma a volte il referente è una persona reale, altre volte invece si tratta di nomi di marchi o altro (vedi per esempio *Mr. T-Rex* e *Mr. Clean*)
- Owner: inteso come *proprietario/a*, ma in un caso appare in un cartellone pubblicitario sullo sfondo dell'immagine
- Partner: in alcuni casi indica il partner in campo amoroso/sexuale, sia animale che umano, e in un solo caso il partner lavorativo
- Player: oltre a indicare una persona che gioca, si trova anche nelle espressioni *single player* (da intendersi come giochi con un solo giocatore), *dvd player*, *game for 2 players*
- Ranger: compare solo in *Power Rangers*
- Rapist: ha sempre un referente maschio
- Scavenger: potrebbe significare sia 'scavatrice' (la macchina) che 'scavatore', ma si trova solo nel nome di una pagina social, *The internet scavengers*
- Speedster: non compare come lavoro di velocista, ma come caratteristica di un supereroe
- Stacy: non ha il significato di nome proprio, ma si riferisce a uno stereotipo di donna, bella e stupida
- Teen: rappresenta sia categorie porno, che il periodo di età, ma fa anche parte del titolo di una serie animata

- Teenager: compare anche, una sola volta, come nome fittizio di un account social
- Veteran: è usato anche senza riferimento all'ambito militare
- Virgin: è sempre riferito a persone, e non è mai usato con il significato di 'incontaminato'

3.4 Normalizzazione delle forme grafiche

Soprattutto nelle parole che indicano relazioni, abbondano i diminutivi e gli affettivi, oltre ad alcune grafie non standard. Una parola come *mother* compare nei dataset sia nella forma *mother*, che *mothers*, *mommy*, *mum* e in tante altre forme ancora. Sebbene ci sia sì una differenza di registro linguistico fra *mother*, *mom* e *mum*, dove la prima parola è più formale, la seconda più colloquiale, e la terza appartiene al registro familiare affettivo, il che si traduce nell'uso di queste parole in contesti differenti, la differenza è comunque abbastanza sottile da poter essere tralasciata in un'analisi quantitativa. Dato che un modello neurale rappresenta in modo molto simile parole come *mother* e *mom*, tutte le forme diverse che avessero però lo stesso referente sono state raggruppate sotto un'unica forma, la più standard, tenendo comunque traccia dei numeri di frequenze delle forme alterate. Di seguito, il primo lemma è la forma sotto cui sono state accorpate le parole che seguono a destra:

- a. Mother → mother, mom, mum, mommy, mummy,³⁹ mamma, mama
- b. Father → father, dad, daddy⁴⁰
- c. Grandmother → grandmother, grandma, granny
- d. Grandfather → grandfather, grandpa
- e. Girlfriend → girlfriend, gf
- f. Boyfriend → boyfriend, bf⁴¹
- g. Girl → girl, gurl, gurl
- h. Transgender → transgender, tran⁴²
- i. Teenager → teenager, teen⁴³
- j. Doctor → doctor, doc

³⁹ Solo nei casi in cui vuol dire *mamma* e non *mummia*.

⁴⁰ Solo nei casi in cui indica il genitore.

⁴¹ Solo quando vuol dire *boyfriend*.

⁴² *Tran* è un errore di Spacy Stanza, che lemmatizza in questo modo la parola *trans*.

⁴³ Solo nei casi in cui indica effettivamente un adolescente.

k. Fiancé → fiancée, fianc⁴⁴

l. Divorced → divorced, divorc.

Nella fase di etichettatura sono emerse anche alcune disgrafie. Le disgrafie, quando notate, non sono state corrette, per perseguire il principio di intervenire meno possibile sul testo, in modo da mantenerlo più naturale possibile. Ma, sempre per lo stesso principio dell'accorpamento di alcuni lemmi, ho accorpato sotto il lemma corretto anche le parole scritte male. Si tratta comunque di una lista piuttosto breve: *vetran* per *veteran*, *gentleman* per *gentleman*, *awoman* per *a woman*, *femenist* per *feminist*, *therapist*⁴⁵ e *apist* per *the rapist*, *transgend* per *transgender*.

3.5 Analisi dei dati

Una volta terminato il processo di etichettatura, corrette le disgrafie e accorpate le varianti di uno stesso lemma, ho creato su Excel delle tabelle e dei grafici. Le tabelle rappresentano le categorie, con l'elenco dei lemmi con l'etichetta pertinente, e i numeri di frequenze assolute dei lemmi corrispondenti a ogni dataset, e si possono vedere in appendice (tabelle dalla 19 alla 25). I grafici, basati sui valori delle tabelle, sono a linee con indicatori, dove l'asse delle x corrisponde al lemma, e l'asse delle y alla frequenza assoluta del lemma. Non ho inserito in uno stesso grafico i dati di più dataset per mantenere una scala, che altrimenti, date le sostanziali differenze dei numeri delle frequenze assolute fra dataset, si perderebbe. L'elenco delle parole all'interno dei vari gruppi di categorie è ordinato in modo decrescente in base alle frequenze del dataset GT+, che è quasi sempre anche quello con i numeri più alti.

La presenza di parole che hanno come referenti delle donne (36) è sensibilmente superiore rispetto a quella degli uomini (28). La categoria *donna* è anche quella più rappresentata, seguita dalla categoria delle professioni che conta 35 termini. Al terzo posto c'è *uomini*, con 27 lemmi, seguiti dai 20 del gruppo relativo a sessualità ed estetica. I termini offensivi e quelli che indicano le relazioni sentimentali contengono 10 lemmi ciascuna, e infine ci sono i 9 termini della categoria *parentela*. Per avere una visione globale su quanto queste categorie siano rappresentate nei singoli dataset, ho creato due grafici (grafici n. 1 e 2) in cui i valori corrispondono alla somma delle

⁴⁴ La forma *fianc*, così come la successiva *divorc*, dipende dalla regex, che ha eliminato le lettere accentate.

⁴⁵ Specifico che non si tratta di *terapista*, ma di un mancato spaziamento fra articolo e nome.

frequenze assolute degli *identity term* che fanno parte della data categoria. Il grafico numero 1 ha i valori di GT+ e GT-, e il grafico numero 2 di FP e FN: in questo modo le colonne sono in scala.

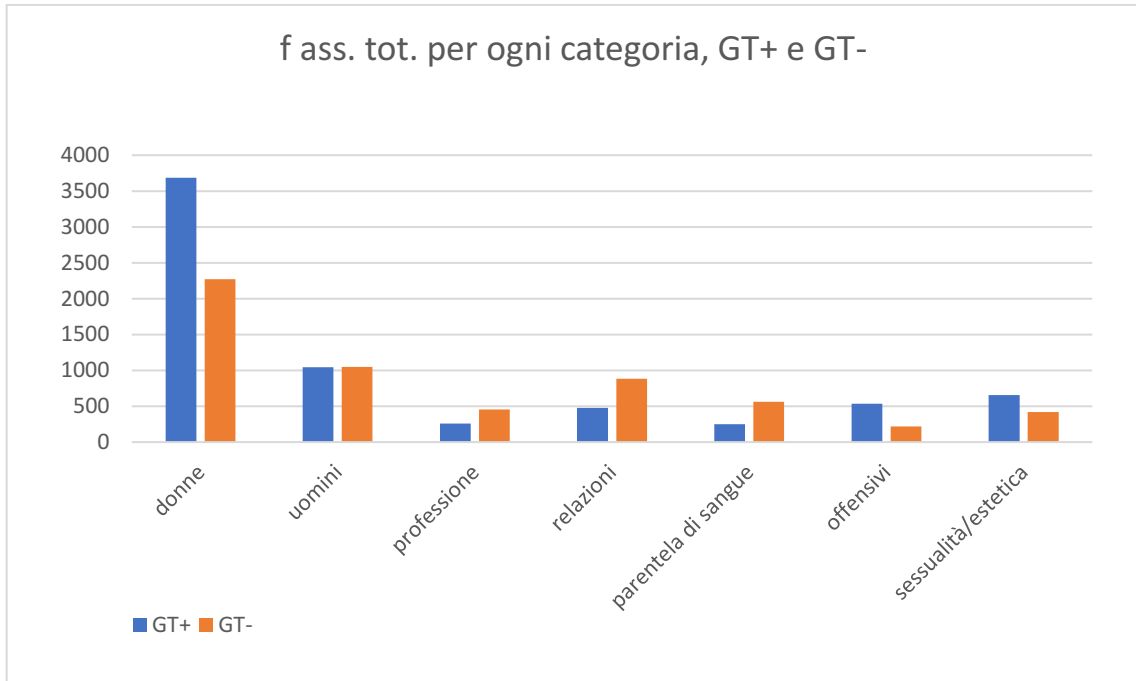


Grafico 1. Frequenze assolute degli identity term divise per categoria, nei dataset GT+ e GT-.

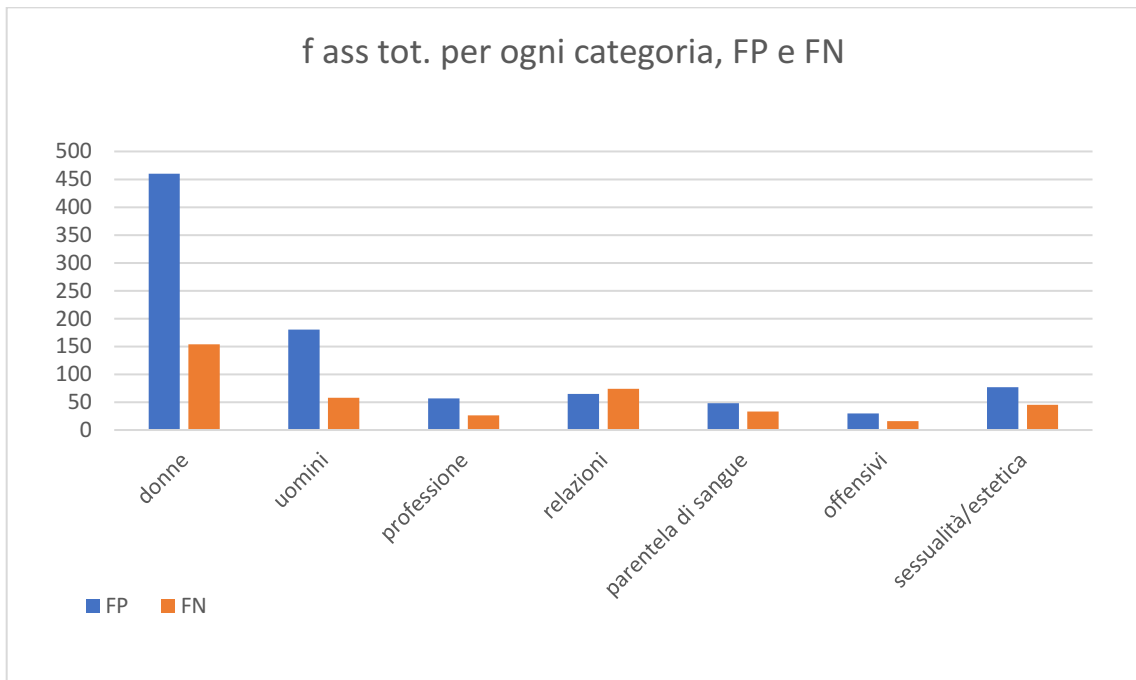


Grafico 2. Frequenze assolute degli identity term divise per categoria, nei dataset GT+ e GT-.

Quantitativamente, le categorie *donne*, *uomini*, *relazioni sentimentali*, *termini offensivi* e *sessualità/estetica* sono più simili fra GT+ e FP, e fra GT- e FN, mentre *professione* e *parentela* sono più simili fra GT+ e FN, e GT- e FP. Quindi, a livello di rappresentazione, guardando il numero di volte in cui certe parole appartenenti a certe categorie semantiche compaiono all'interno dei testi, i dataset classificati in modo errato sono più simili al dataset che corrisponde alla classificazione errata che è stata fatta. Cioè, se le parole per identificare le donne sono più presenti nei meme misogini (GT+) piuttosto che in quelli non misogini (GT-), anche le parole che identificano donne sono contenute di più nei meme che sono stati classificati in modo sbagliato come misogini (FP) piuttosto che nei meme che sono stati classificati in modo sbagliato come non misogini (FN). E questo avviene per cinque categorie su sette. È probabile quindi che i modelli associno la presenza di parole femminili più spesso a un contenuto misogino.

3.5.1 Identity term femminili

È la categoria più rappresentata in tutti i dataset, sia per numero di lemmi che per numero di occorrenze. I lemmi sono in totale 35, di cui 12 sono offensivi. Nei seguenti grafici l'elenco delle parole sull'asse delle x segue lo stesso ordine.

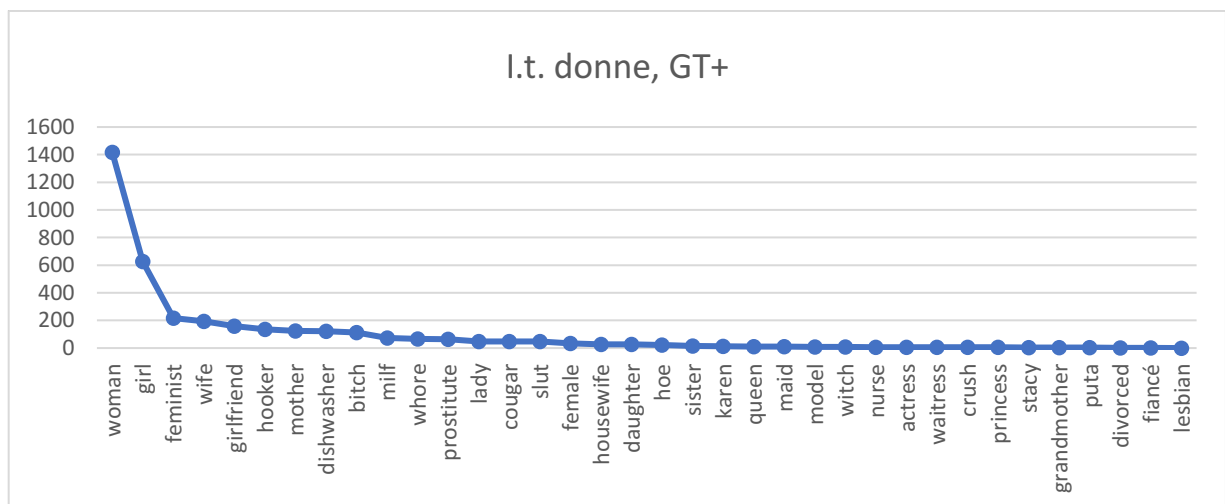


Grafico 3. Identity term femminili del dataset GT+.

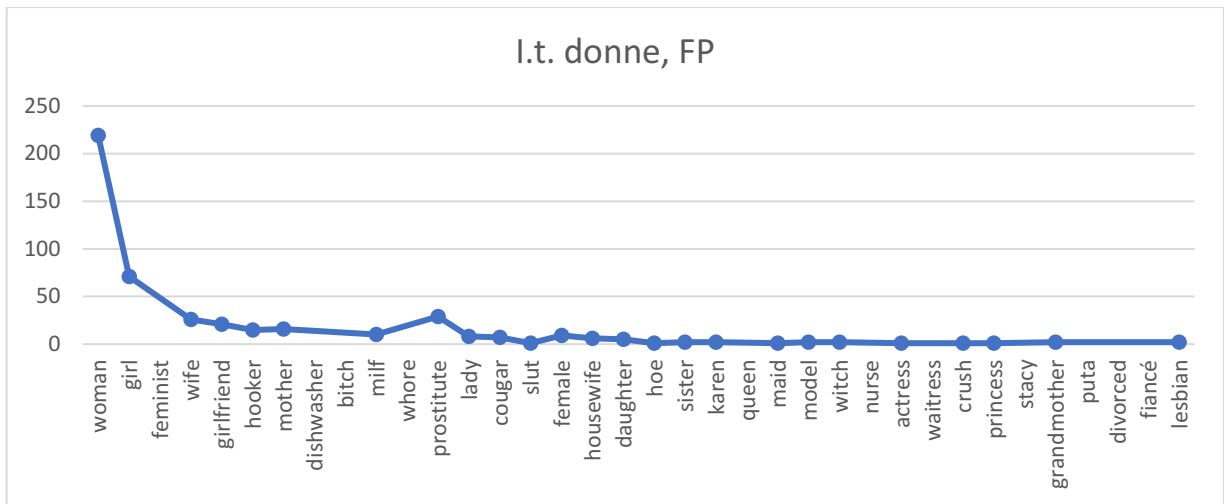


Grafico 4. Identity term femminili del dataset FP.

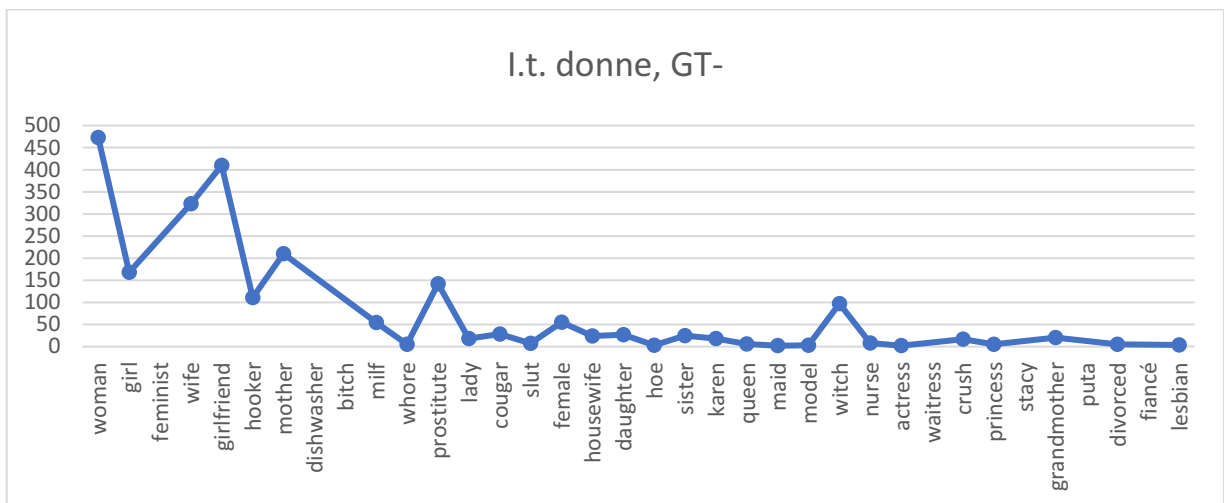


Grafico 5. Identity term femminili del dataset GT-.

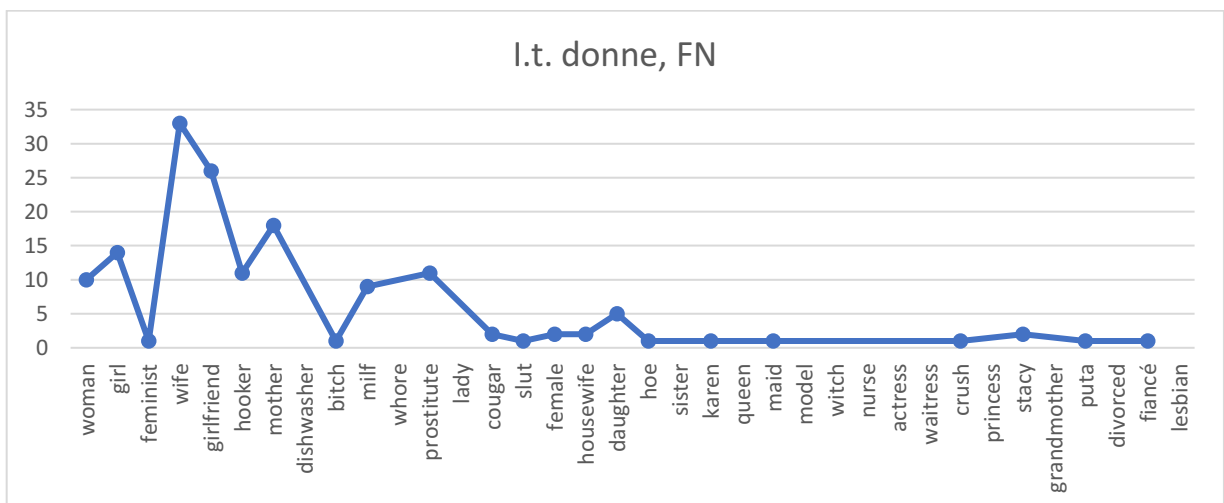


Grafico 6. Identity term femminili del dataset FN.

In GT+ e FP l'andamento degli *identity term* è molto più simile rispetto agli altri due dataset, ma GT- e FN non sono molto simili fra loro. Così come in tanti altri casi, GT+ e FP presentano delle somiglianze, mentre GT- e FN no, e in generale FN è il dataset che si discosta più di tutti. La somiglianza fra GT+ e FP non è tanto nel numero di *identity term* condivisi, che non è molto dissimile fra GT+ e FN, ma nelle frequenze assolute di questi *identity term*: in termini assoluti alcuni lemmi, che è facile pensare siano riconosciuti dai modelli come portatori di contenuti misogini (si pensi a tutti i sinonimi di *prostituta*), sono presenti più spesso in FP che in FN.⁴⁶ Questo fenomeno potrebbe essere riconducibile all'*unintended bias*, ossia la tendenza di un modello a classificare in un certo modo una certa unità perché contiene certe parole. L'*unintended bias* è stato studiato anche appositamente nei compiti di classificazione misogina, in cui si è visto che, soprattutto all'interno dei modelli di elaborazione del linguaggio naturale preaddestrati, errori importanti sono dovuti da un set di *identity term* che sono associati frequentemente a un contenuto misogino, e che portano quindi il modello a interpretare erroneamente il dato nella classificazione (Nozza, Volpetti, Fersini 2019). Questo fenomeno ha portato alla scelta del *modus operandi* sull'analisi dei bigrammi e dei trigrammi, di cui si parla al capitolo 4.⁴⁷

I dataset GT- e FP hanno in comune solo il picco in *prostitute*. Invece nei dataset GT+ e FP l'andamento dei primi sette *identity term* quasi lo stesso, tranne che per *feminist*, *woman*, *girl*, *wife*, *girlfriend*, *hooker* e *mother* sono i termini femminili più presenti sia in GT+ che in FP. Il fatto che le linee dei grafici di GT+ e FP siano più simili rispetto alle linee di GT- e FN potrebbe spiegare perché le macchine hanno classificato come misogini contenuti non misogini molto più spesso che il caso contrario: i meme FP hanno molti *identity term* che possono portare ad un *unintended bias* (Zhou et al. 2022) in comune con i meme effettivamente misogini, e questo potrebbe avere spinto i modelli a riconoscerli come misogini. Al contrario, i contenuti misogini dei meme FN sono molto più sottili, cioè non si avvalgono di lessico esplicito come fanno i meme GT+ e quelli FP. Ma dato che i meme misogini che contengano parole o immagini violente o sessualmente esplicite sono più frequenti di quelli che non ne contengono, capita raramente che i modelli predicano come non misogini contenuti che in realtà lo sono,

⁴⁶ Sarebbe anche interessante, per avvalorare questa ipotesi, vedere se questi i.t. probabilmente soggetti a bias siano per esempio combianti fra di loro all'interno di uno stesso meme.

⁴⁷ In breve, individuare delle parole trigger.

mentre capita più di frequente che facciano una predizione positiva per meme invece non misogini.

In particolare per i primissimi termini dell'elenco, si nota anche una discrepanza delle frequenze assolute fra dataset riconosciuti dai modelli come misogini e riconosciuti invece come non misogini: in GT+ e FP *woman* e *girl* hanno delle frequenze assolute altissime, soprattutto *woman*, che si discostano molto da tutti gli altri termini. In GT+ *woman* compare più del doppio delle volte di *girl*, che a sua volta è attestata quasi il triplo delle volte del termine che occupa la terza posizione. Questa proporzione è più o meno mantenuta anche in FP, dove *woman* compare il triplo di *girl*, e *girl* quasi il triplo della parola successiva. Lo stesso non vale per GT- né per FN. Se è vero che in GT- la parola più frequente di questa categoria è *woman*, la sua frequenza assoluta non si discosta molto da quella di altri termini nelle prime posizioni. In FN invece *woman* e *girl* non sono nemmeno i due termini più frequenti.

3.5.2 Identity term uomini

Anche in questo caso le linee di GT+ e FP sono più simili rispetto a GT- e FN. I primi sette *identity term* seguono lo stesso ordine decrescente, con l'eccezione di *son* e *boyfriend* che in FP hanno un valore rispettivamente di 2 e 3, mentre in GT+ *boyfriend* ricorre meno volte di *son*; la differenza in FP è però talmente bassa, una sola occorrenza, che può essere ignorata. Fino ai primi tre i.t. l'andamento è lo stesso anche nel dataset GT-. Solo FN si differenzia, anche se in minima parte: *husband*, seppure con una sola occorrenza in più, è più presente di *boy*. *Grandfather*, *bouncer* e *cowboy* si attestano soltanto in GT-, ma se *grandfather* compare ben 11 volte, *bouncer* e *cowboy* solo una volta, e questa volta è quella che viene classificate erroneamente e inserita in FP. In quanto alle frequenze assolute, si nota che *man* è molto più rappresentato in GT+ che in GT-, mentre la seconda e la terza parola della lista, *boy* e *husband*, hanno una rappresentazione abbastanza simile.

Come si può vedere, nessuno dei termini esclusivamente maschili è offensivo. Le uniche parole dispregiative che hanno referente solo maschile sono quelle relative all'orientamento sessuale.⁴⁸ L'unica parola che ha una connotazione in parte offensiva

⁴⁸ Al contrario che per le donne, non esistono termini dispregiativi sulla condotta sessuale, presunta o vera che sia, troppo libera o troppo casta. Nell'elenco l'unico termine che ha a che fare con la condotta sessuale è *incel* (che indica uomini che non hanno relazioni sentimentali né sessuali non per scelta, ma

è *Chad*. *Chad* e *Stacy* sono due nomi propri che nella sub-cultura degli *incel* hanno sviluppato un nuovo significato: quelli che sono originariamente dei nomi propri vengono utilizzati come nomi comuni per indicare due figure stereotipiche di uomo e donna che si collocano in cima alla gerarchia sociale, perché rispecchiano perfettamente l'ideale estetico che viene attribuito ai due generi (Gemelli 2021). I *chad* sono la rappresentazione della mascolinità ideale: uomini di bell'aspetto e ricchi, sono gli unici che hanno accesso alle *stacy*, donne belle e desiderabili, ma approfittatrici e superficiali, manipolatrici, egoiste e crudeli (Furl 2022 e Maxwell et al. 2020). I meme GT- numero 921⁴⁹, 8781⁵⁰ e il meme GT+ 1820 (tutti in Figura 11) mostrano la rappresentazione dei *chad*, mentre il meme GT+ n. 4563 (fig. 12) raffigura il modello *stacy* di donna, confermando la descrizione di questi caratteri stereotipici del mondo *incel*. Nel corpus dei meme però rientra anche un'altra sfumatura per le *stacy*, quella espressa nei meme GT+ n. 694 e 6004 (entrambi in Figura 12), dove la madre di una *stacy*, che in quanto madre è una *milf* (*mother I would like to fuck*), è ancora più appetibile della *stacy* stessa. Da questi meme emerge una differenza nella rappresentazione dell'ideale di uomo e donna, dovuta innanzitutto allo spazio dedicato alla descrizione delle caratteristiche degli uni e delle altre. I *chad* sono ampiamente descritti nei meme n. 921 e 8781, mentre non è dato spazio alla descrizione delle *stacy*.

per l'aspetto fisico poco piacevole), che però non ha nessuna connotazione negativa in quanto al comportamento sessuale in senso stretto, ma indica uomini brutti, vergini, con scarso successo sociale, non amichevoli e sgradevoli, termine diventato *mainstream* dopo gli omicidi di massa commessi da due *incel* del 2014 e nel 2018. Si veda per il fenomeno degli *incel* il capitolo 2 di Gemelli (2021).

⁴⁹ Riporto di seguito il testo: "the virgin wholesome meme literally a pixelated dopamine rush for depressed teen be suppose to be subversive repetitive after two post be either make or consume by homosexual or teenage girl I on my way to love you make by people with ridiculously fragile mental state talk like retard be fuck all about encourage positivity and good vibe will actually have an aneurysm at a slightly edgy joke you be do great hun keep it up xoxoxoxoxo needlessly political can not go a day without post about trans right or feminism will inevitably evolve into internal struggle since it be purely emotional have to post advertisement for sweatshirt or mug every other day 6 million na to remain above - destitute just regular meme with a pink filter and heart emoji spamme over they the chad edgy meme require historical knowledge to understand ; only consumable by people with high IQ will eventually say the word be satirically political as a means to humor all audience without alienate son - sjw will never not be relevant since it be more memorable despite be 13 percent be enjoy by nearly everybody . since all joke be in good nature realize the basis of comedy in offense use it accordingly obliterate ever - sensitivity to make a more mentally fortified stock of meme intentionally low quality have create more meme in 2019 than wholesome meme have since 2015 be not addict to le epic fanne cate doggo bork bork pupper meme c be the embodiment of the chad - virgin paradigm since it realize some thing be inherently well than other instead of the virgin you be do great weaty keep it up :)) philosophy"

⁵⁰ Riporto di seguito la parte di testo relativa al *chad*: GREENSBIRO CHAD -Peak Caucasian masculinity -Lives to fuck, fucks to live. Has fucked your girlfriend, sister, mom, and dad -Master of every contact sport SPECIAL: Atomic Cumshot

Inoltre, le *stacy* sono tutte oggettificate, e dove il loro corpo non è esposto in modo sessualizzato, interviene lo sguardo di un uomo per farlo (come nel meme n. 694 in Figura 12). Da queste caratteristiche e dal significato delle figure dei *chad* e delle *stacy* si evince che i *chad* sweatshirts visti in maniera diversa e comunque più positiva rispetto alle *stacy*.⁵¹



Figura 10. Meme n. 921, n. 878, n. 1820.

⁵¹ Questo trova conferma nei *word embedding* (figure 15 e 16), dove i due nomi si trovano in posizioni molto distanti tra loro.



Figura 11. Meme n. 4563, n. 694, n. 6004.

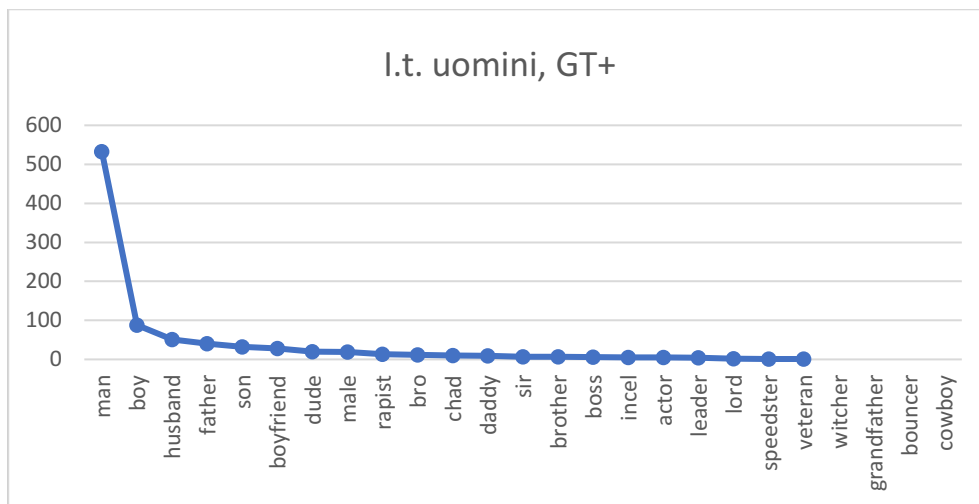


Grafico 7. Identity term maschili nel dataset GT+.

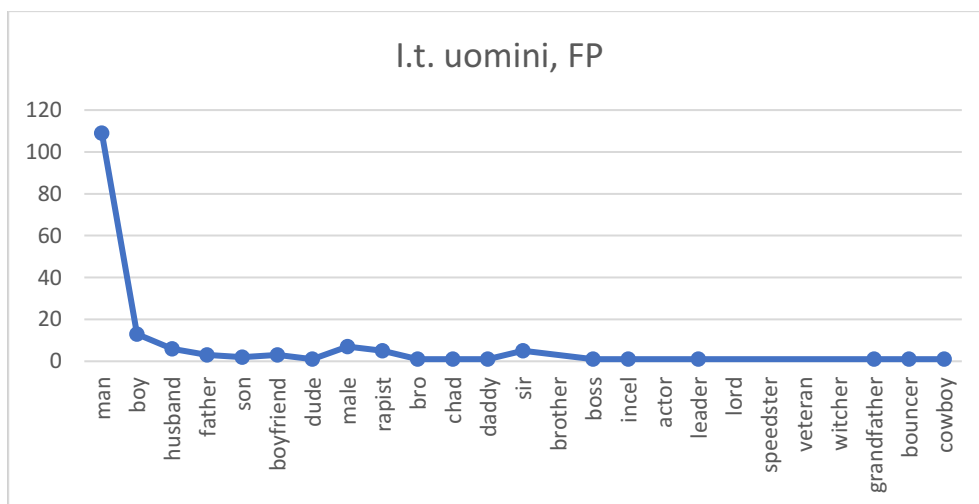


Grafico 8. Identity term maschili nel dataset FP.

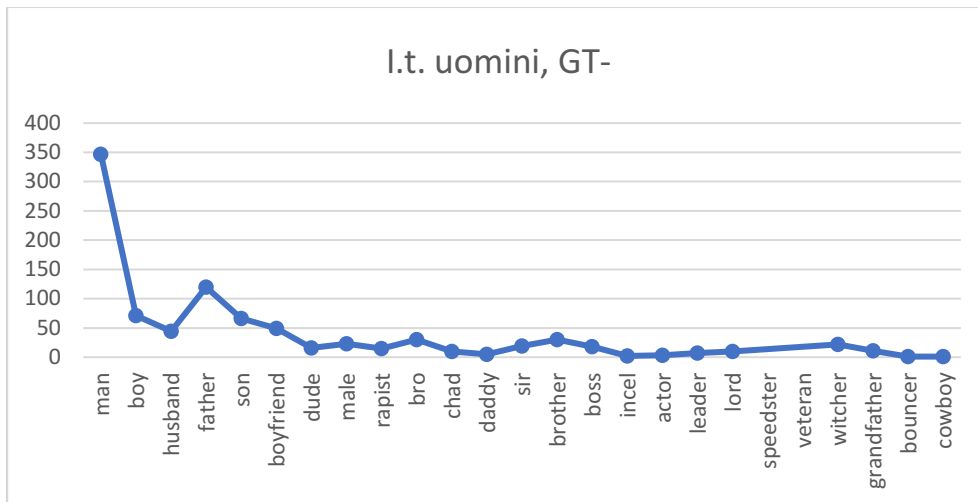


Grafico 9. Identity term maschili nel dataset GT-.

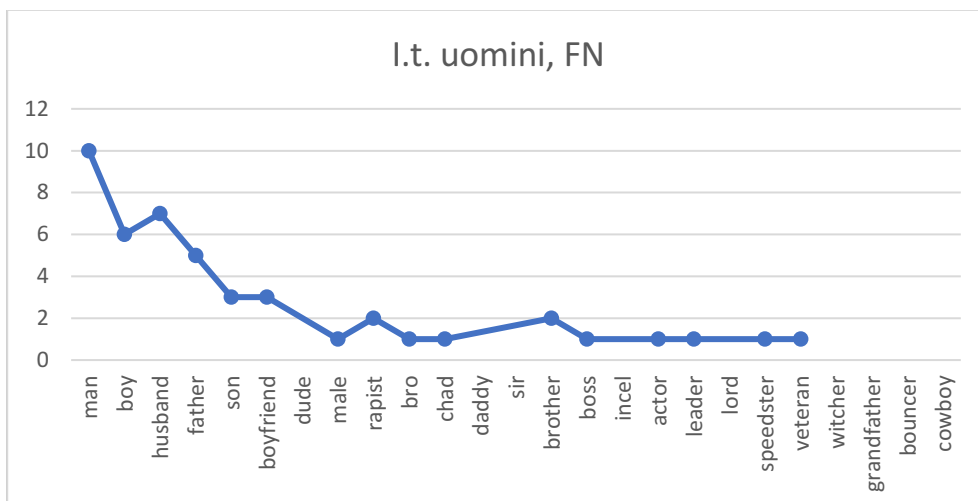


Grafico 10. Identity term maschili nel dataset FN.

3.5.3 Identity term professioni

La professione più rappresentata è quella di prostituta, anche nei meme non misogini, e anzi proprio nei meme non misogini *prostitute* compare 142 volte, contro le 65 dei meme misogini. Questo si può spiegare con due ragioni: innanzitutto *prostitute* è un termine formale, mentre in un messaggio misogino è più probabile che venga usato un termine offensivo per riferirsi a una prostituta; in secondo luogo, è possibile che nei meme misogini venga usata una varietà lessicale più ricca per esprimere il concetto di prostituta. A tale proposito si può vedere la Tabella 3, che raccoglie tutti i sinonimi più o meno formali e offensivi di *prostitute*. Il dataset GT+ ha in realtà pochi termini in più rispetto a GT-, ma la frequenza assoluta totale è decisamente più alta.

Tabella 3. Frequenze assolute dei sinonimi più o meno offensivi di prostitute.

	GT+	GT-	FP	FN
hooker	135	111	15	11
bitch	113	-	-	1
whore	66	5	-	-
prostitute	65	142	29	11
slut	48	7	1	1
hoe	22	3	1	1
puta	3	-	-	1
tot	452	268	46	26

Non sorprende che GT- non abbia *waitress*: la forma non marcata per genere è il maschile *waiter*, e quindi in contesti non reali come quelli creati dai meme si usano le forme non marcate a meno che non ci sia un'esigenza particolare. Nei meme misogini l'esigenza è di creare dell'umorismo che fa leva sugli stereotipi di genere, e quindi la presenza di un cameriere o di una cameriera⁵² è rilevante a seconda dello scherzo che si vuole fare. In tutti i grafici tranne che in GT+ c'è un piccolo picco in corrispondenza di *cop*. FN ha soltanto 13 nomi di questa categoria, contro i 20 di FP.

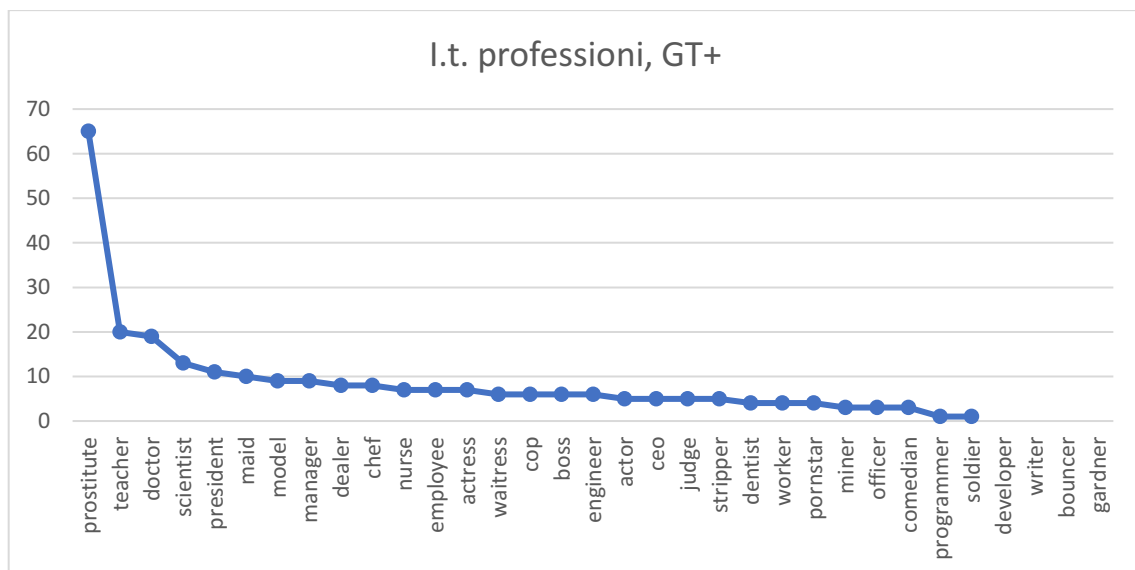


Grafico 11. Identity term delle professioni del dataset GT+.

⁵² In GT+ *waitress* si trova in meme come "I NOTICED MY WAITRESS HAD A BLACK EYE SO I GAVE HER MY ORDER REAL SLOW CAUSE OBVIOUSLY THIS BITCH DONT LISTEN haha punching women", mentre in tutti i casi in cui compare *waiter* questo è funzionale solo a caratterizzare l'ambiente della barzelletta, e il fatto che il cameriere sia un maschio non ha nessuna rilevanza.

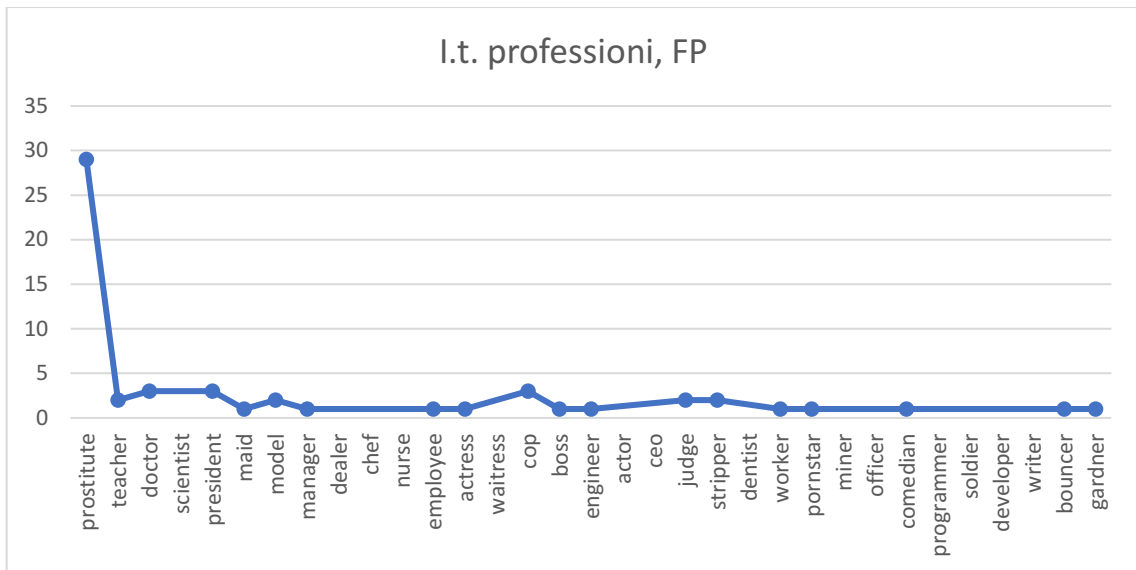


Grafico 12. Identity term delle professioni del dataset FP.

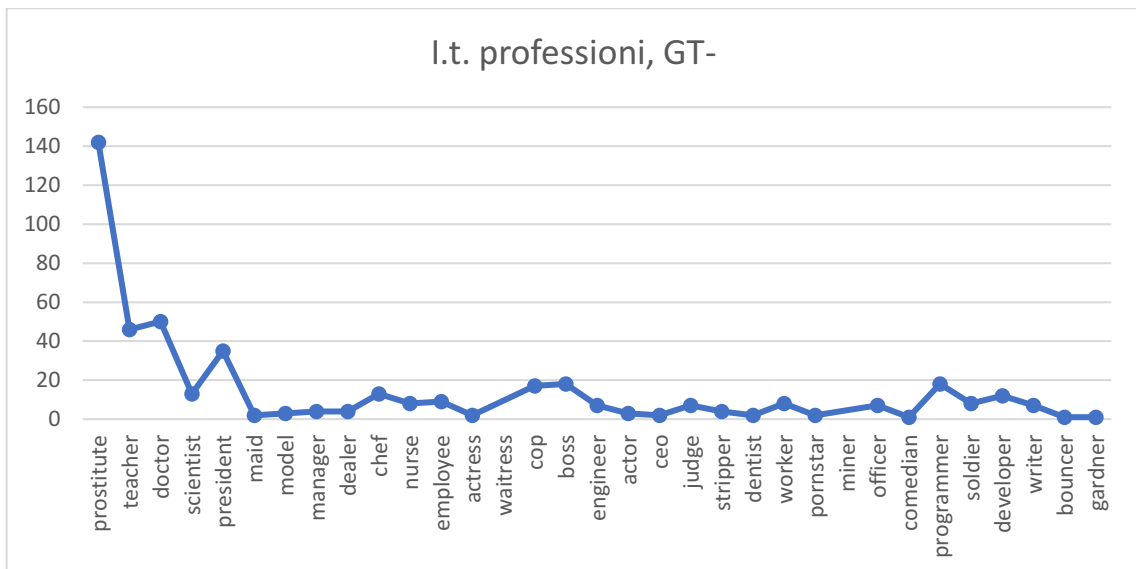


Grafico 13. Identity term delle professioni del dataset GT-.

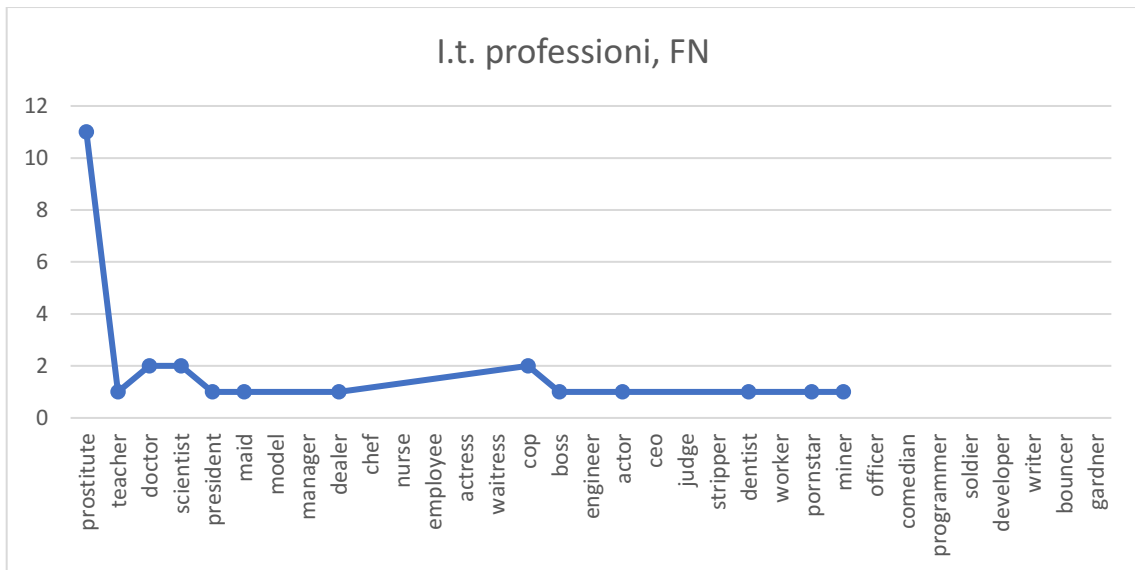


Grafico 14. Identity term delle professioni del dataset FN.

3.5.4 Identity term sessualità ed estetica

Tutti i dataset hanno un picco in *hooker*, e, tranne il dataset GT+, in *prostitute*. 10 termini su 20 sono di genere esclusivamente femminili⁵³ (*hooker, bitch, milf, whore, prostitute, cougar, slut, hoe, stacy, puta, lesbian*); di questi, 7 sono sinonimi più o meno colloquiali di *prostitute* (*hooker, bitch, whore, prostitute, slut, hoe, puta*), uno riguarda l'orientamento sessuale (*lesbian*), tre sono termini oggettificanti che classificano una donna in base alle fantasie sessuali degli uomini (*milf, cougar, stacy* e in parte *teen*).

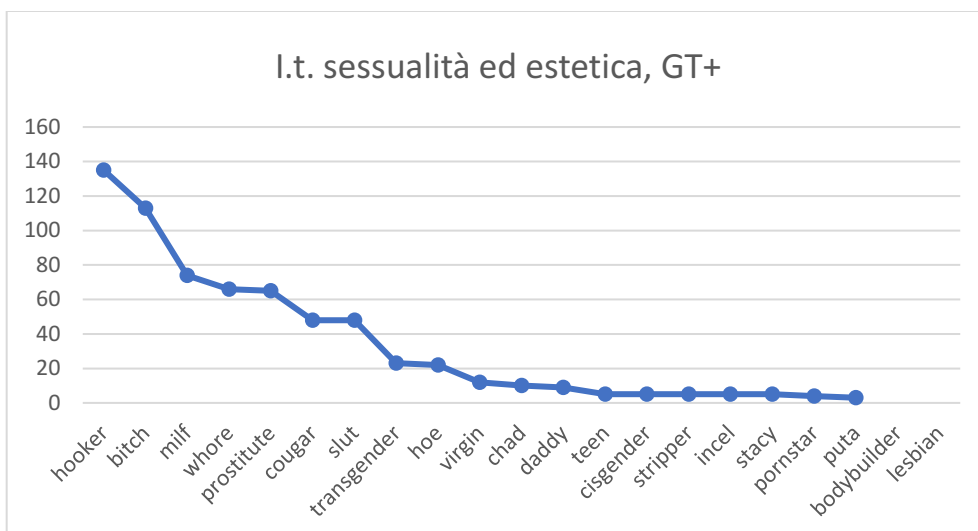


Grafico 15. Identity term nell'ambito della sessualità e dell'estetica nel dataset GT+.

⁵³ Ce ne sono altri che invece sono riferibili a entrambi i sessi, ma nei dataset vengono riferiti soltanto alle donne.

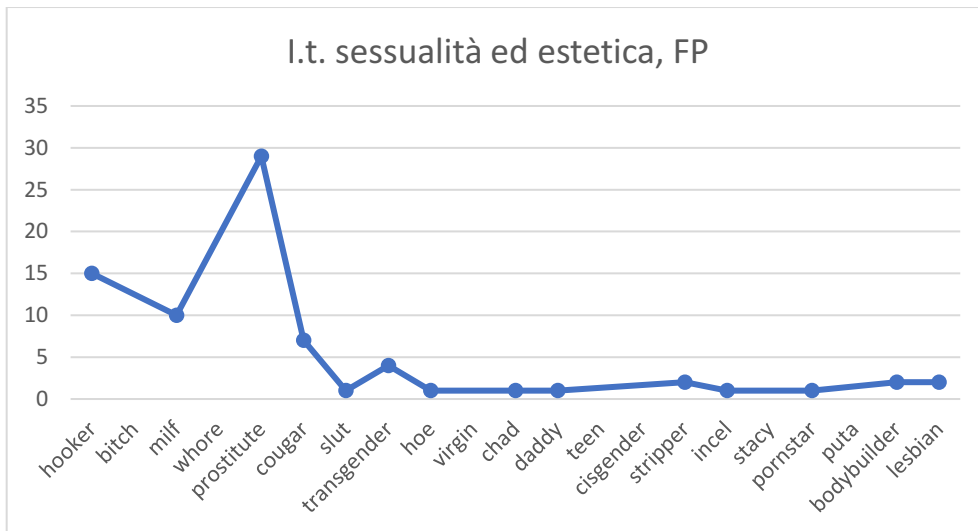


Grafico 16. Identity term nell'ambito della sessualità e dell'estetica nel dataset FP.



Grafico 17. Identity term nell'ambito della sessualità e dell'estetica nel dataset GT-.

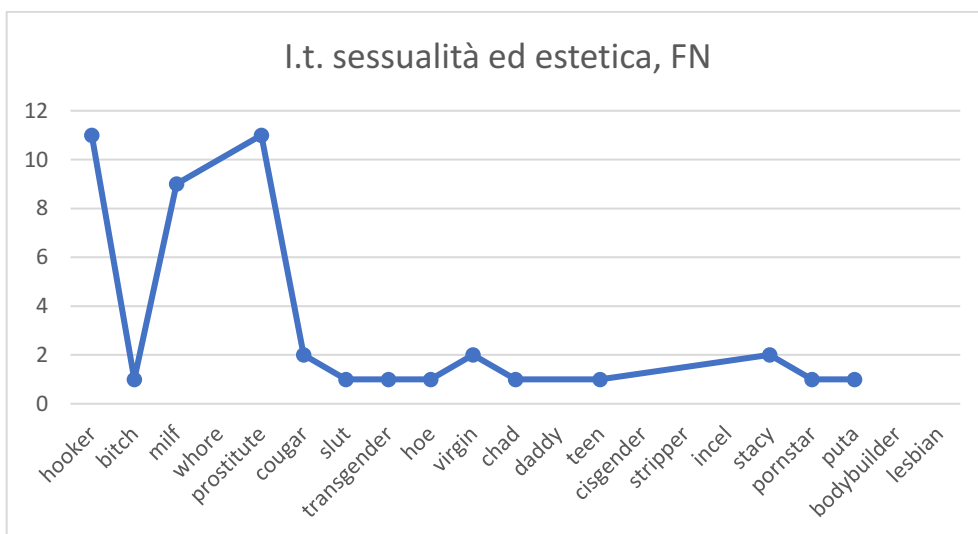


Grafico 18. Identity term nell'ambito della sessualità e dell'estetica nel dataset FN.

3.5.5 Identity term offensivi

Su undici *identity term* offensivi, sette sono esclusivamente attribuibili a donne e riguardano la condotta sessuale, e uno, *bitch*, è usato solo contro le donne in due dataset. In tutti i dataset il nome più usato è *hooker* seguito da *bitch*; nel dataset GT+ questo ordine è interrotto da *dishwasher*, che si colloca quindi al secondo posto. Non ho inserito i due nomi *chad* e *stacy* perché la loro connotazione non è sempre offensiva, ma spesso si concentra solo sulle caratteristiche estetiche e sulle abilità sociali della persona a cui è riferita.

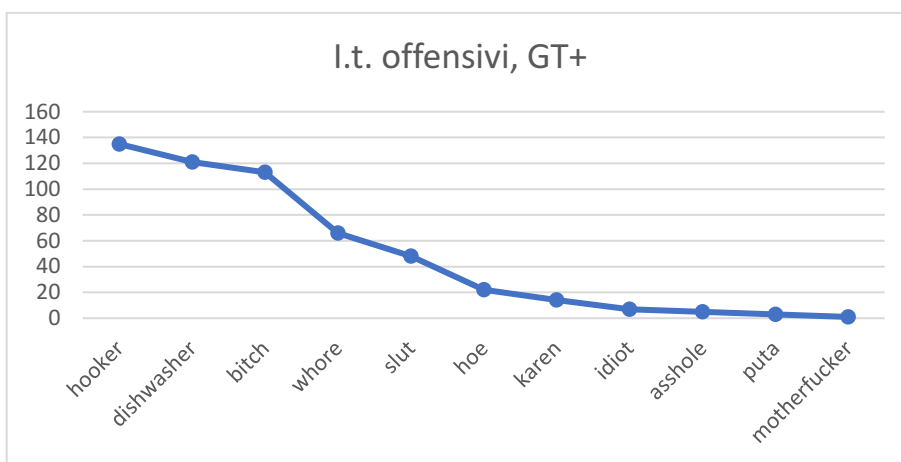


Grafico 19. Identity term offensivi nel dataset GT+.

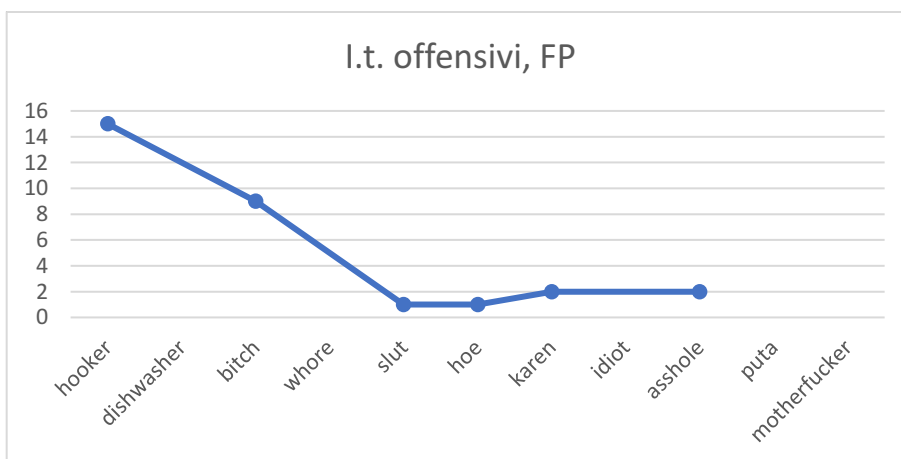


Grafico 20. Identity term offensivi nel dataset FP.

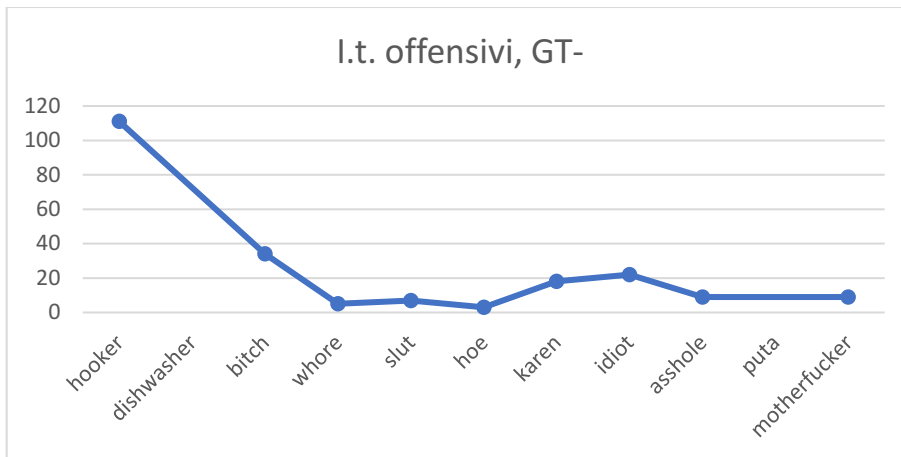


Grafico 21. Identity term offensivi nel dataset GT-.

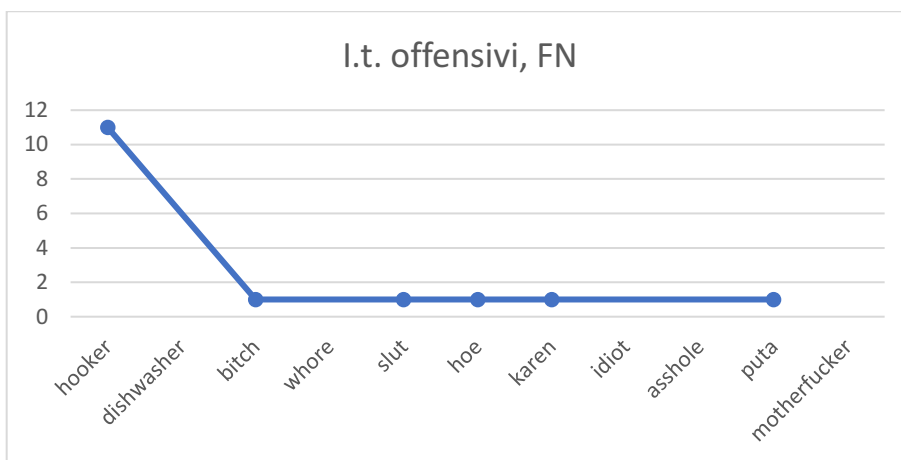


Grafico 22. Identity term offensivi nel dataset FN.

3.5.6 Identity term relazioni

GT- è l'unico dataset in cui la parola più frequente non è *wife* ma *girlfriend*. Sia l'andamento di FP che quello di FN sono molto più simili a GT- che a GT+. La frequenza assoluta delle prime due parole in GT+ è decisamente superiore a GT-; è probabile che il motivo sia sempre attribuibile alla presenza in GT+ di molti termini diversi, spesso offensivi, fra cui *bitch*, che vengono usati per chiamare le donne. Nella sfera relazionale quello più comune nel dataset GT+ è *dishwasher*, che viene usato molto spesso per riferirsi proprio alle mogli e alle fidanzate, oltre che alle donne in generale. Anche con questa disambiguazione le frequenze di GT+ sarebbero comunque più basse. Questa categoria rappresenta l'unico caso in cui le frequenze del dataset FN sono maggiori delle frequenze del dataset FP, il che è tanto più strano dato che FN è un dataset molto meno corposo di FP. Inoltre la suddivisione delle frequenze è netta in tutti e quattro i

dataset: sono soltanto i primi due termini, *wife* e *girlfriend*, che hanno frequenze assolute maggiori, mentre passando alla terza parola, *husband*, si vede che in GT+ e GT- le frequenze scendono rispettivamente a 51 e 44, e in FP e FN rispettivamente a 6 e 7, ottenendo quindi in entrambi i casi valori molto simili. La ragione potrebbe risiedere nel fatto che in contesti misogini si tende a classificare di meno le donne nei loro ruoli sentimentali con gli uomini, e si preferisce invece nominarle secondo altre caratteristiche, come appunto quella di persona che mantiene in ordine la casa, o che ha una condotta sessuale deprecabile. Difatti, quando si passa al successivo *identity term* femminile, *housewife*, la differenza dei valori delle frequenze assolute si dissolve: GT+ ha 27 occorrenze, e GT- 24.

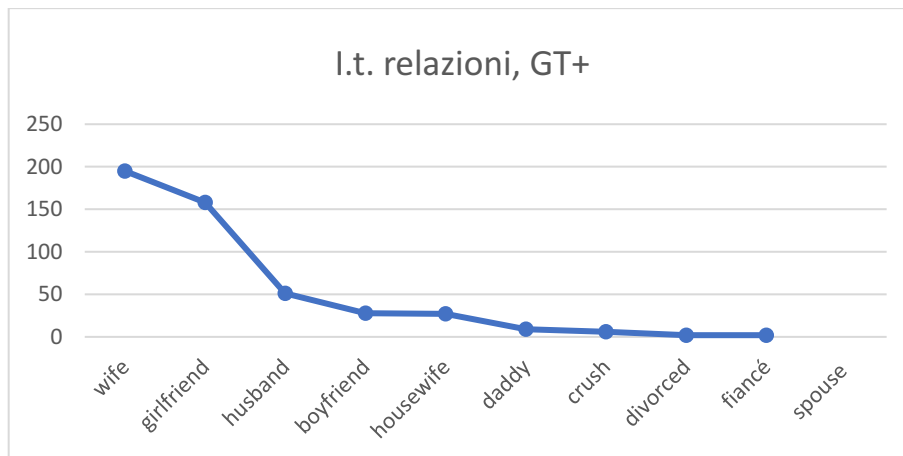


Grafico 23. Identity term delle relazioni sentimentali nel dataset GT+.

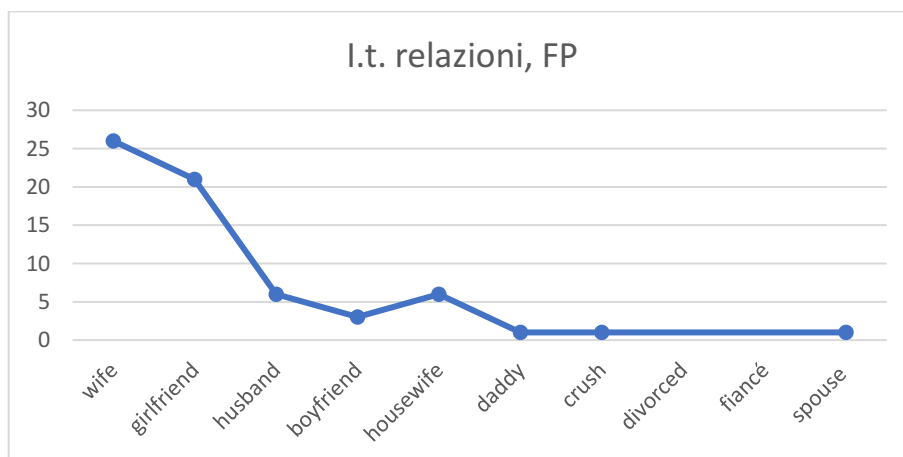


Grafico 24. Identity term delle relazioni sentimentali nel dataset FP.

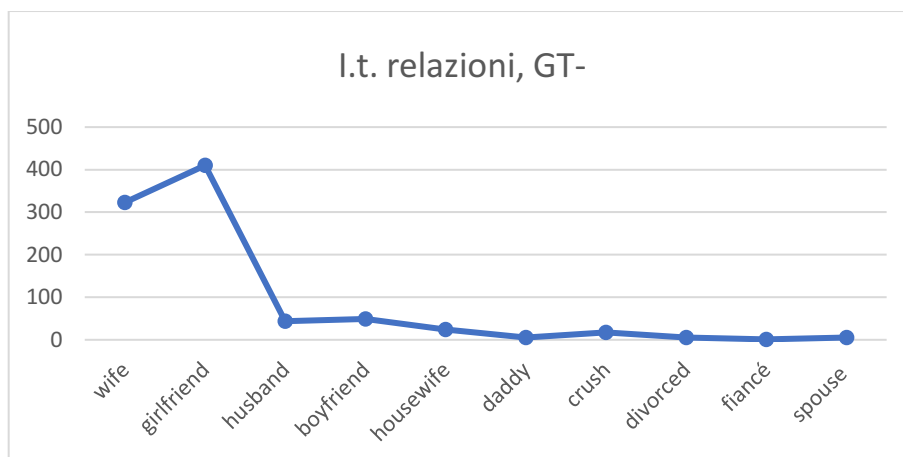


Grafico 25. Identity term delle relazioni sentimentali nel dataset GT-.

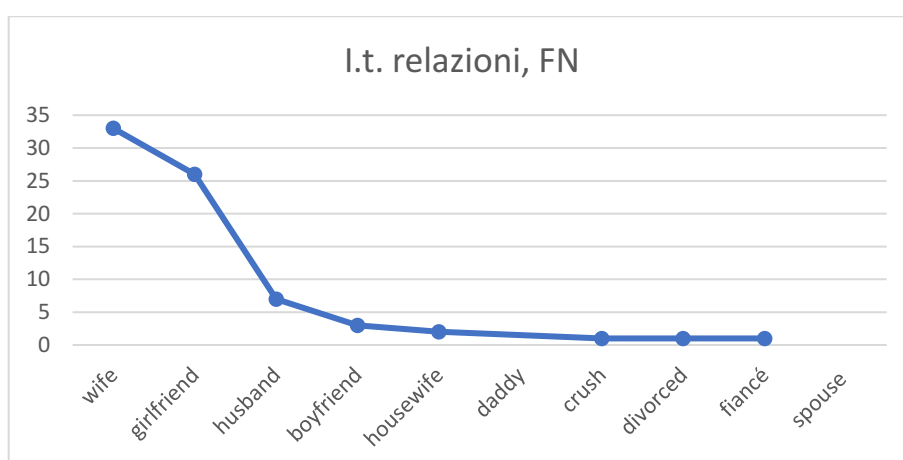


Grafico 26. Identity term delle relazioni sentimentali nel dataset FN.

3.5.7 Identity term di parentela

Il dataset misogino ha delle frequenze minori rispetto alla GT+, cosa che accade solo per due gruppi, le relazioni sentimentali e la parentela di sangue. Anche nella sfera della parentela ci si può aspettare che meme dal contenuto misogino abbiano un lessico più ristretto, perché è più comune che contenuti misogini siano indirizzati verso donne generiche o donne disprezzate dallo scrivente, piuttosto che dalle donne appartenenti al proprio nucleo familiare. Chiaramente i termini di parentela non necessariamente sono usati per indicare una propria parente, ma possono indicare la parente di qualcun altro, così come non per forza se c'è una parente nel contesto di un meme misogino, allora la misoginia è indirizzata verso questa, ma proprio per questo la parentela di sangue può assumere un ruolo più marginale, in un universo in cui l'unico argomento è l'odio contro le donne. Questa differenza si può apprezzare meglio se si prende la lista

di nomi di parentela non accorpati (in Tabella 4), dove si vede che su quattordici termini che indicano parentela, molti dei quali sono varianti affettive, il dataset GT+ ne contiene 9, e il dataset GT- ne contiene 12, ma la differenza maggiore è data dalla frequenza in cui questi termini compaiono, che in GT- è il doppio rispetto a GT+. Inoltre si riscontra la stessa cosa anche nei dataset classificati in modo scorretto: in FP gli *identity term* familiari sono sia più numerosi che più frequenti di quelli in FN.

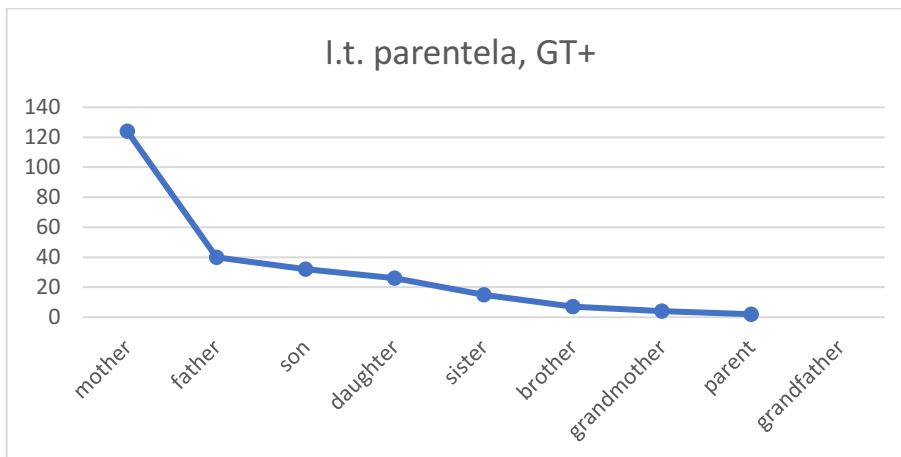


Grafico 27. Identity term di parentela nel dataset GT+.

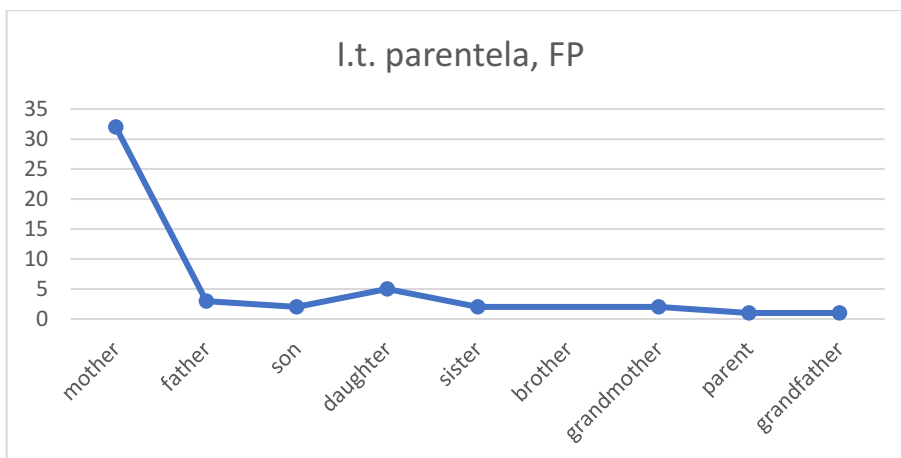


Grafico 28. Identity term di parentela nel dataset FP.

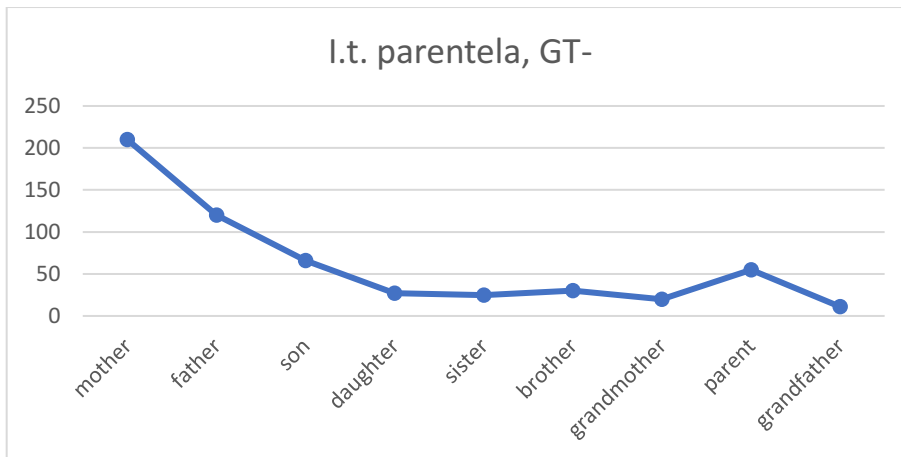


Grafico 29. Identity term di parentela nel dataset GT-.

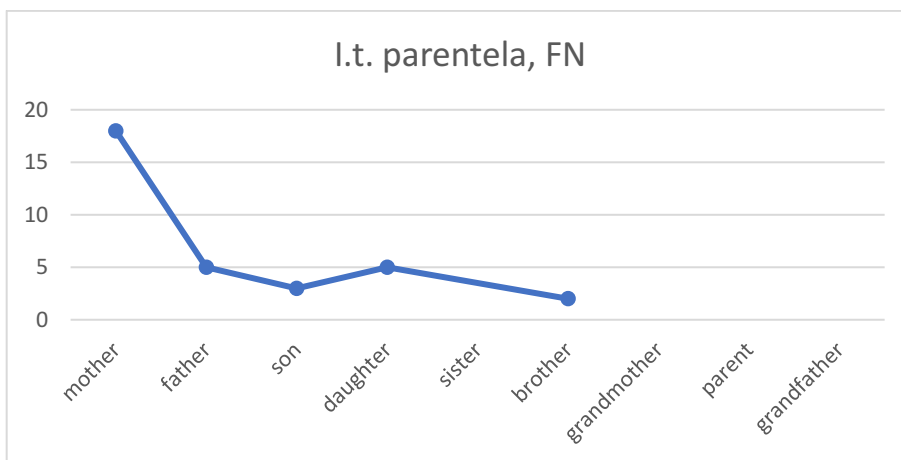


Grafico 30. Identity term di parentela nel dataset FN.

Tabella 4. Frequenze assolute nei quattro dataset di tutti gli identity term di parentela, che sono stati poi accorpati sotto parole seme.

I.t. di parentela non accorpati	GT+	FN	GT-	FP
mom	65	9	149	7
mother	38	2	27	6
daddy	14	1	20	1
mama	13	-	10	1
father	10	2	23	-
mommy	6	-	12	1
grandmother	4	-	-	-
mum	2	-	24	1
momma	2	-	-	-
grandma	-	-	17	1
grandpa	-	-	6	-
grandfather	-	-	5	1
granny	-	-	3	1
mummy	-	-	3	-

Tot. frequenze	154	14	299	20
-----------------------	------------	-----------	------------	-----------

Capitolo 4: Analisi su bigrammi e trigrammi

Descrivo in questo capitolo le analisi svolte sui bigrammi e sui trigrammi, aprendo l'argomento con le ragioni che mi hanno spinto a scegliere queste unità come oggetto di analisi (paragrafo 4.1). Proseguo poi con le problematiche, e la soluzione, relative alla ricerca di una misura che potesse ordinare gli n-grammi in base all'impatto che questi avessero per il modello (paragrafo 4.2). Nei paragrafi 4.3 e 4.3 descrivo, motivo e analizzo i processi di annotazione semantica svolti su bigrammi e trigrammi. Nell'ultimo paragrafo (4.4.1) introduco una nuova analisi, basata sull'individuazione dei soggetti parlanti dei meme.

4.1 Gli n-grammi

Gli n-grammi possono essere sia di caratteri che di parola. Nel mio lavoro uso soltanto quelli di parola, per cui ogni volta che si parlerà di n-grammi si intenderanno solo gli n-grammi di parola. Per n-gramma si intende una sequenza più o meno lunga di parole direttamente contigue all'interno di un testo, che può andare da una parola sola, gli unigrammi, fino a un numero che di solito si limita a quattro o cinque. Per esempio, SketchEngine prevede un massimo di sei parole per la ricerca di n-grammi. L'uso degli n-grammi di parola è largamente usato nei compiti di elaborazione del linguaggio naturale (per citare alcuni lavori, HaCohen-Kerner et al. 2010; Pamungkas et al. 2018; Frenda 2022), ed è uno strumento utile nella linguistica dei corpora, poiché isolare una sequenza ristretta di elementi permette di catturare il contesto frasale di un testo a diversi livelli di granularità. Isolare e comparare n-grammi fra *corpora* può servire per individuare dei pattern linguistici ricorrenti: è proprio questa la motivazione che mi ha spinto ad usarli.

Avendo a disposizione quattro *corpora* di testi di meme, suddivisi a seconda della correttezza o meno della predizione dei modelli, grazie alla funzione *n-grams* di SketchEngine ho estratto i bigrammi e i trigrammi. La funzione *n-grams* del software si basa sui token (quindi le forme verbali negative come *don't* e *haven't* sono considerate dei bigrammi), e richiede che la sequenza di parole compaia almeno due volte per

essere rappresentata.⁵⁴ Non ho usato sequenze composte da più di tre parole perché i dataset, soprattutto quelli con predizione errata, sono molto limitati, e i risultati prodotti non sarebbero sufficienti per un'analisi di questo tipo.⁵⁵ Mettendo a confronto le frequenze assolute dei bigrammi e dei trigrammi nei quattro dataset si può vedere come le stesse sequenze di due o tre parole si distribuiscono fra di essi. Gli n-grammi sono anche interessanti perché possono raccogliere anche parole funzionali, restituendo combinazioni di parole contenute e funzionali che più difficilmente si riescono a notare senza l'estrazione automatica di gruppi di parole.

Per l'analisi mi sono concentrata sugli n-grammi presenti anche nei dataset FP e FN, e ho individuato delle categorie ricorrenti. Una parte importante del lavoro è stata la ricerca di una misura che desse un valore all'impatto di queste unità nei modelli, misura di cui si parla nel paragrafo successivo.

4.2 Misure di valutazione dell'impatto

Dare un valore all'impatto che gli n-grammi hanno per i modelli è un'operazione fondamentale per valutare la rilevanza dell'errore: grazie al valore dell'impatto le liste degli n-grammi possono essere ordinate proprio in base al valore che il loro impatto assume. Un valore di questo tipo restituisce un'informazione per così dire interna al metodo di classificazione dei modelli stessi. A questo scopo, per l'annotazione semantica degli n-grammi è stato importante cercare una formula che tenesse conto di tutti gli aspetti di cui invece non tiene conto il coefficiente che mette in relazione le unità di FP o di FN con quelle rispettivamente di TN e TP, di cui si è parlato nel paragrafo 2.2.5. Sono state usate quindi le micro e macro *F1-measure*, e l'impatto (Ting 2011).

Il primo metodo usato è stata la misura F1. Le matrici di confusione si usano per valutare le prestazioni dei modelli di classificazione automatica, e le metriche più comuni di cui si servono sono precisione, richiamo, e macro-F (Ting 2011). La precisione è il rapporto fra veri positivi, o veri negativi, e tutti i positivi, o negativi, a livello di annotazione. Il richiamo è invece il rapporto fra veri positivi, o veri negativi, e tutti i positivi, o negativi, a livello di predizione, mentre la macro-F è una media armonica di precisione e richiamo. Più il valore della macro-F si avvicina ad 1 e meno impatto ha

⁵⁴ Il software offre anche la possibilità di scegliere la frequenza minima per cui un n-gramma sia restituito, ma il limite minimo è di due.

⁵⁵ I tetragrammi di FN sono 91, mentre in FP addirittura non esiste nessun tetragramma.

l'elemento, mentre più il valore si avvicina allo 0 e più l'elemento è impattante. Questa misura non ha però risolto i problemi relativi alla considerazione di tutti gli aspetti di classificazione: guardando la Tabella 5 si può vedere che il valore della macro-F negativa è lo stesso per bigrammi che dovrebbero avere invece un impatto differente. Il bigramma *woman s* viene classificato correttamente come negativo solo 7 volte su 21, venendo quindi sbagliato ben 14 volte, e la sua macro-F ha lo stesso valore del bigramma *girl in* che viene classificato in modo errato 3 volte su 6. Di seguito riporto le formule delle tre misure solo per la classe positiva: per la classe negativa basta sostituire TP con TN, e FP con FN; per la macro F invece basta sostituire precisione e richiamo positivi con i valori di quelli negativi.

$$P(+)=\frac{TP}{TP+FP}$$

Formula 1, Precisione positiva.

$$R(+)=\frac{TP}{TP+FN}$$

Formula 2, Richiamo positivo.

$$macro\ F(+)=2\times\frac{P(+)\times R(+)}{P(+)+R(+)}$$

Formula 3, Macro F positiva.

Tabella 5. Valori di precisione, richiamo, e macro F positivi e negativi di alcuni bigrammi.

BIGRAMMI	GT+	TP	FN	GT-	TN	FP	P+	R+	F+	P-	R-	F-
woman s	113	113	0	21	7	14	0,88976378	1	0,941666667	1	0,333333333	0,5
for woman	35	35	0	6	2	4	0,897435897	1	0,945945946	1	0,333333333	0,5
girl in	21	18	3	6	3	3	0,857142857	0,857142857	0,857142857	0,5	0,5	0,5
she in	16	16	0	3	1	2	0,888888889	1	0,941176471	1	0,333333333	0,5
girl when	15	15	0	3	1	2	0,882352941	1	0,9375	1	0,333333333	0,5
a milf	10	8	2	7	3	4	0,666666667	0,8	0,727272727	0,6	0,428571429	0,5
woman on	12	12	0	3	1	2	0,857142857	1	0,923076923	1	0,333333333	0,5

Per risolvere il problema ho utilizzato la misura micro-F, o F1 pesata, che aggiunge un aspetto in più al valore che assegna: la somma delle frequenze assolute di tutti

gli elementi della GT della classe corrispondente.⁵⁶ La micro-F riesce a distinguere questi casi assegnando coefficienti diversi ad elementi che hanno uno stesso valore di macro-F (si veda la Tabella 6). Il coefficiente è direttamente proporzionale alla correttezza della classificazione, e quindi inversamente proporzionale alla portata dell'impatto, come quello della macro-F. Riporto la formula della micro F positiva (per quella negativa, anche in questo caso basta invertire i dataset):

$$micro\ F(+)=\frac{TP+FN}{\sum_{k=m}^n f(k)(+)}$$

Formula 4. Micro F positiva.

Tabella 6. Macro e micro F positive e negative, per gli stessi bigrammi della tabella 5.

BIGRAMMI	GT+	TP	FN	GT-	TN	FP	F+	F-	micro F1+	micro F1-
woman s	113	113	0	21	7	14	0,94166666667	0,5	0,00246481048	0,00020798669
for woman	35	35	0	6	2	4	0,94594594595	0,5	0,00076690621	0,00005942477
girl in	21	18	3	6	3	3	0,85714285714	0,5	0,00041694656	0,00005942477
she in	16	16	0	3	1	2	0,94117647059	0,5	0,00034881804	0,00002971238
girl when	15	15	0	3	1	2	0,93750000000	0,5	0,00032573950	0,00002971238
a milf	10	8	2	7	3	4	0,72727272727	0,5	0,00016846326	0,00006932890
woman on	12	12	0	3	1	2	0,92307692308	0,5	0,00025658250	0,00002971238

Anche la micro-F ha però presentato un problema. Il coefficiente pari a 0 viene assegnato in due casi opposti: sia quando un elemento non compare in una delle due classi di GT (si veda ad esempio la Tabella 7, dove il coefficiente della micro-F+ del bigramma *to finanzi ally*, bigramma che non è attestato nella classe positiva, è 0), sia nei casi in cui la classificazione di un elemento venga sbagliato tutte le volte, come si vede per il bigramma *to dress* della Tabella 7, il cui coefficiente di micro-F- è 0, appiattendosi così l'impatto. Il primo caso, ossia un coefficiente pari a 0 se l'elemento non esiste, avviene anche per i coefficienti di precisione, richiamo, e macro-F. È problematico quando lo 0 viene assegnato anche per gli elementi classificati erroneamente tutte le volte che ricorrono, poiché l'impatto di un elemento che viene

⁵⁶ Più nello specifico, la micro-F per esempio positiva si calcola moltiplicando la macro-F+ per (frequenza assoluta dell'elemento diviso la somma delle frequenze assolute della GT+).

sbagliato 2 volte su 2 non può essere uguale a quello di un elemento sbagliato 6 volte su 6.

Tabella 7. Esempi della diversa capacità descrittiva di micro F ed impatto.

BIGRAMMI	GT+	TP	FN	GT-	TN	FP	micro F1+	micro F1-	Impatto+	Impatto-
to dress	4	4	0	2	0	2	0,00007412	0	0,00001853	0,00003962
to financially	0	0	0	5	5	0	0	0,00009904	not present	0

La misura risolutiva, che è stata usata per ordinare le liste degli n-grammi, è l'impatto, che disambigua i due casi, e che, se un elemento viene classificato erroneamente ogni volta che compare, assegna valori diversi in rapporto alla frequenza assoluta dell'elemento sbagliato in tutte le sue occorrenze. Si tratta semplicemente di aggiungere una funzione logica in Excel, chiedendo al foglio di calcolo di restituire "not present" nel caso in cui la frequenza assoluta di un n-gramma della *ground truth* sia pari a 0, e mantenendo il resto della formula uguale a quella della micro F, ma sottraendo il valore della macro F a 1, di modo che il valore finale dell'impatto sia invertito: un valore pari a 1 indica il massimo impatto, mentre un valore pari a 0 indica un impatto inesistente, perché l'elemento non esiste (si veda la tabella 7).

4.3 Annotazione semantica di bigrammi e trigrammi

Per cercare di capire se ci sono degli elementi lessicali che possono portare i modelli a un *unintended bias* ho svolto un processo di etichettatura su bigrammi e trigrammi dei dataset che sono stati predetti in modo errato. Anche in questo caso, come per l'etichettatura degli *identity term*, ho seguito il metodo *open coding* ma ho selezionato categorie diverse ed etichettato parole appartenenti a tutte le classi grammaticali, e non solo a quella dei nomi. In modo empirico, scorrendo le liste, ho quindi individuato parole potenzialmente *trigger*, cioè che possono portare i modelli a sbagliare la classe di predizione per via della loro maggiore frequenza nei meme appartenenti a una certa classe piuttosto che alla classe opposta (Nozza, Volpetti, Fersini 2019). Questo meccanismo di deviazione si riscontra in particolare per la classe misogina, che si caratterizza per un contenuto specifico, cioè l'odio verso le donne, e che quindi si serve più frequentemente di un certo tipo di linguaggio per esprimere tali contenuti. Al contrario, la classe non misogina non è uniforme dal punto di vista del contenuto, ma

comprende qualsiasi argomento che non sia l'odio verso le donne. Il processo di annotazione di bigrammi e trigrammi è perciò utile in particolare per approfondire la deviazione verso la classificazione misogina dei modelli, che è interessante dal momento che i meme FP sono 545, contro i 236 meme FN.

È possibile quindi che alcune parole semanticamente significative, come per esempio *hoe*, aumentino la probabilità che la predizione venga sbagliata. Per controllare questa ipotesi ho cercato di rilevare in che modo queste parole si distribuivano all'interno dei dataset classificati in modo sbagliato, FP e FN, al fine di verificare se davvero i bigrammi e i trigrammi sbagliati più spesso contenessero più parole *trigger* di quelli predetti male meno di frequente. A questo scopo disporre gli n-grammi a seconda della rilevanza dell'errore di classificazione è fondamentale, e per questo ho ordinato le liste dei bigrammi e dei trigrammi secondo il valore dell'impatto, che dà un valore all'impatto che questi hanno all'interno dei modelli.

Le parole *trigger* sono state divise in categorie semantiche, alcune più granulari, come la categoria *donna* e quella *sex working*, altre meno, come le categorie *sessualità* e *parti del corpo* e *parolacce*. Per ridurre gli errori, l'etichettatura sia dei bigrammi che dei trigrammi è stata ripetuta per quattro volte su ognuno dei due *corpora*. Le categorie sono:

- a. Donna (F): parole che hanno come referente una donna. Possono essere nomi, come *girlfriend*, aggettivi, come *female*, o pronomi, come *she* o *her*, e anche i nomi propri. Ho escluso tutte le parole sinonimo di *prostituta*, che sono state inserite nella categoria successiva;
- b. Sex working (W): tutti i nomi come *prostitute*, *hooker*, *bitch* e sinonimi, che descrivono la donna per la sua condotta sessuale o per il mestiere svolto in ambito sessuale;
- c. Uomo (M): parole che hanno come referente un uomo. Come per la categoria *donna*, possono essere nomi (*husband*), aggettivi (*male*), pronomi (*him*) e nomi propri;
- d. Sessualità e parti del corpo (S): ne fanno parte le parole che riguardano la sfera sessuale (come *blowjob*, *rape* e *gender*, ma anche verbi che indicano l'atto sessuale come *to get lay* e *to have sex*); che indicano l'orientamento sessuale, per esempio *lesbian* e *straight*; che indicano parti del corpo, come *body* e *nipple*;
- e. Oscenità e parole dal significato fortemente negativo (P): ne fanno parte parolacce come *shit* e *stupid*; parole che indicano crimini della sfera sessuale, come *to rape* o

sexual assault; verbi di violenza come *to torture* e *to kill*. Ho inserito in questa categoria *fuck* perché è più comunemente usato come parolaccia piuttosto che come verbo indicante l'atto sessuale (Frenda et al. 2018; Frenda 2022);

- f. Discorso di genere (G): ci sono alcune parole che di per sé indicano realtà della vita quotidiana, ma che sono molto usate nella classe di meme misogini per veicolare stereotipi di genere sui compiti di una donna. Fra queste, *kitchen*, *house*, *cook* o *sandwich*.⁵⁷ Fanno parte della categoria anche termini strettamente connessi ai discorsi di genere, come *feminist*,⁵⁸ *feminism* o *equal right*;
- g. Altro (A): parole che non rientrano in un unico campo semantico, ma che possono essere di interesse per il modello, come per esempio i nomi delle pagine o dei siti internet di creazione dei meme, quali *mematic* (che fa parte della sequenza *made with mematic*), o *imgflip.com*.

4.4 Analisi dei dati

I bigrammi e i trigrammi sono stati estratti con SketchEngine, divisi quindi per dataset: GT+, FP, GT- e FN. Le trascrizioni dei testi dei meme date in input al software sono quelle lemmatizzate, quindi anche normalizzate in carattere minuscolo, e ripulite con espressione regolare (cfr. paragrafo 2.3). Gli n-grammi devono avere una frequenza assoluta maggiore o uguale a 2 per essere tali. I bigrammi del dataset GT+ sono 8.397, quelli del dataset GT- sono 9.701, quelli di FP sono 963, e i bigrammi in FN 367, mentre i trigrammi del dataset GT+ sono 5.869, quelli del dataset GT- 6.463, e quelli dataset FP e FN sono rispettivamente 331 e 104 (si veda la Tabella 8).

Tabella 8. Numeri totali dei bigrammi e dei trigrammi nei quattro dataset.

	GT+	GT-	FP	FN
Bigrammi	8.397	9.701	963	367
Trigrammi	5.869	6.463	331	104

Ho eseguito l'etichettatura solo sui dataset con predizione sbagliata sia perché le due GT sono troppo ampie per essere annotate a mano, sia per concentrarmi sulle

⁵⁷ Ho già citato la considerazione sull'umorismo sessista fatta da Citron (2009) in nota 31.

⁵⁸ Non si pone il problema di come considerare *feminist* in FN, dove viene sempre riferito alle donne, perché non compare.

predizioni errate. Le liste di bi- e trigrammi sono state ordinate in modo crescente secondo la misura dell'impatto, dove più il numero si avvicina allo 0 e minore è l'impatto, e più il numero si avvicina a 1 e maggiore è l'impatto. Una prima considerazione da fare è che fra gli n-grammi più impattanti, molti sono formati da parole grammaticali, o contengono un verbo o un nome dal significato estremamente generico (si veda in appendice, dalla Tabella 26 alla 29): come è prevedibile, in un *corpus* di una certa lunghezza fra le parole più frequenti si trovano proprio quelle dal significato più generico. Questo si può vedere chiaramente osservando i primi 20 bi- e trigrammi più impattanti, si vede che molti sono in comune. In particolare per i bigrammi si nota che dei primi 20 con maggiore impatto per la classe FN, i primi 18 sono presenti anche nei FP, e dei primi 20 bigrammi FP ben 18 (ma non in ordine) sono presenti anche nei primi 20 bigrammi FN (si vedano i grafici 31 e 32). Per i trigrammi invece i due dataset, sempre in un campione dei primi 20 più impattanti, ne condividono 12 su 20 (grafici 33 e 34).

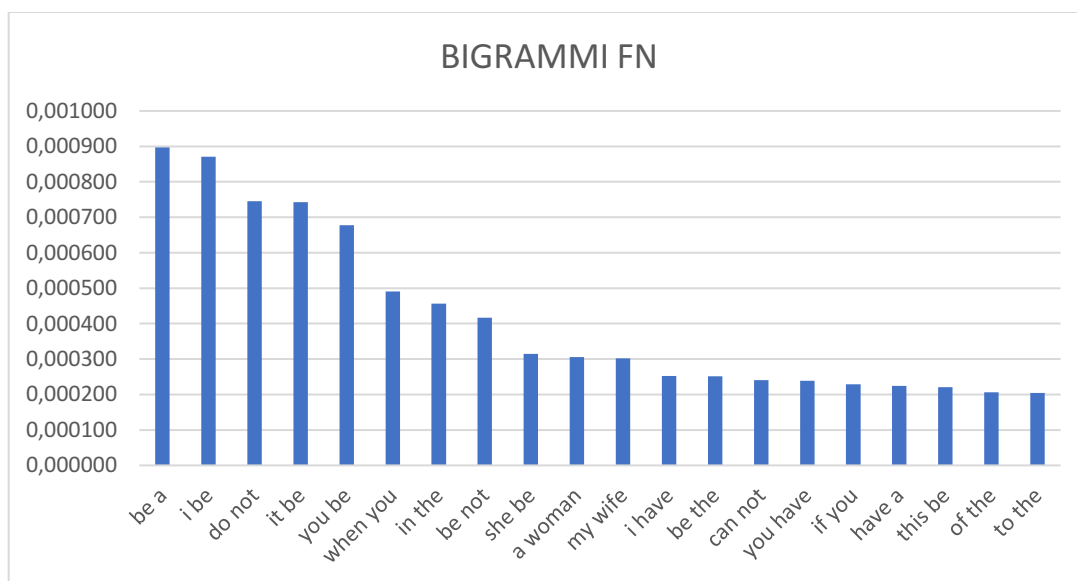


Grafico 31, I primi 20 bigrammi in ordine di impatto (sull'asse delle y) del dataset FN.

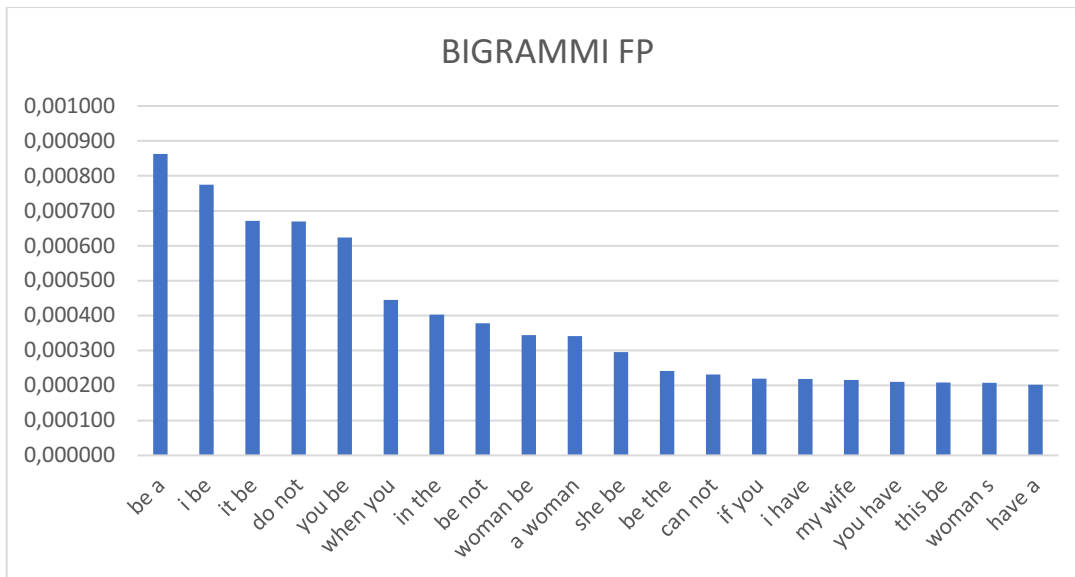


Grafico 32, I primi 20 bigrammi in ordine di impatto (sull'asse delle y) del dataset FP.

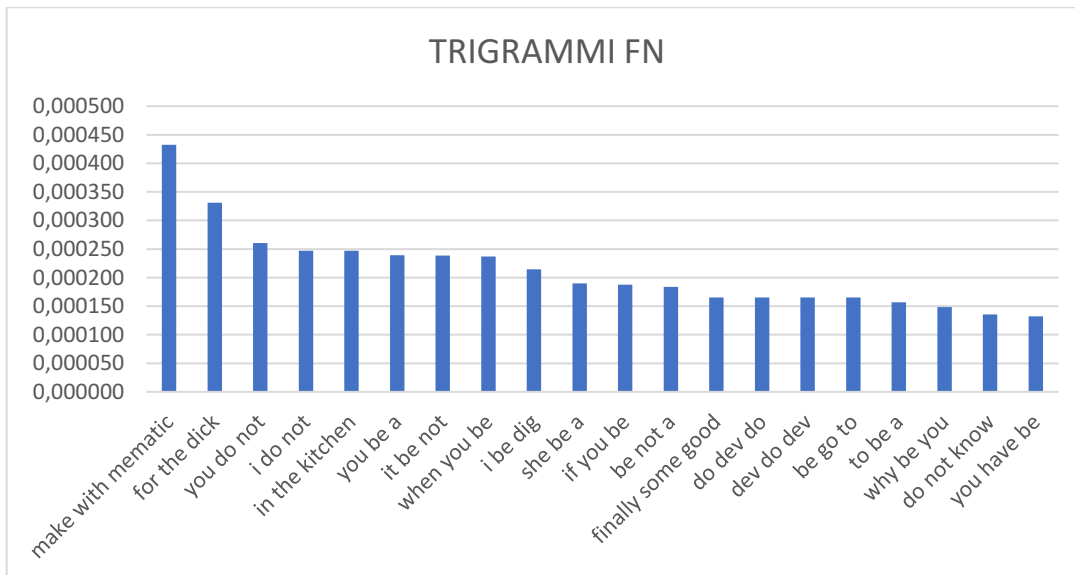


Grafico 33, I primi 20 trigrammi in ordine di impatto (sull'asse delle y) del dataset FN.

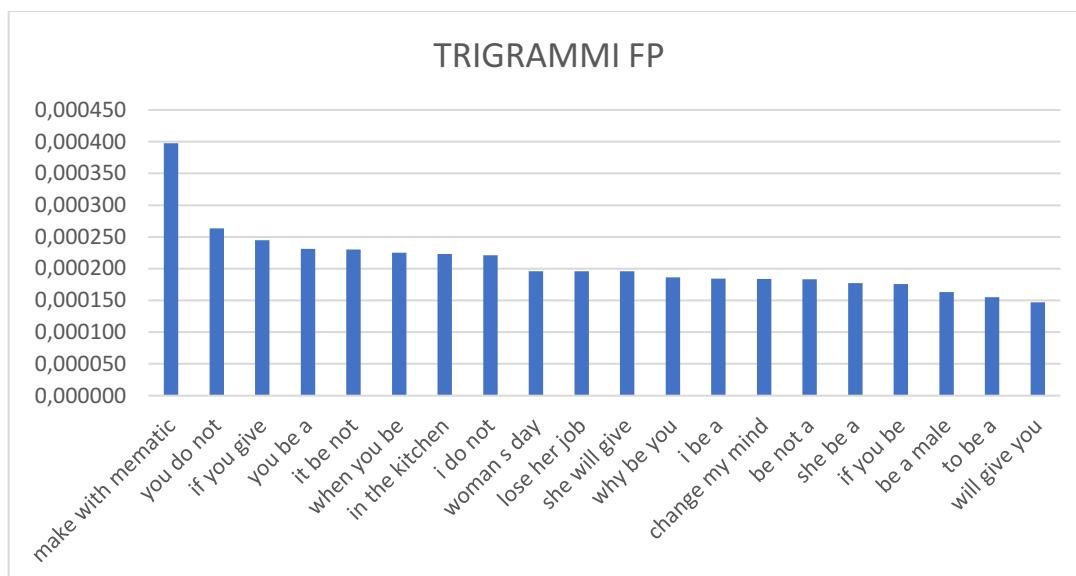


Grafico 34. I primi 20 trigrammi in ordine di impatto (sull'asse delle y) del dataset FP.

Per guardare come le etichette si distribuiscono nelle liste di n-grammi a seconda dell'impatto che questi hanno sul modello, le liste sono state divise in quartili. Le tabelle 9, 10, 11 e 12 mostrano i numeri delle etichette dei bigrammi e dei trigrammi, divise per quartili, a cui si aggiunge un'ulteriore distinzione: bigrammi e trigrammi sono composti da due e tre parole, e quindi possono avere più di una etichetta per unità. Per entrambi, la riga 1° delle tabelle si riferisce alle etichette assegnate alle prime parole delle unità, la riga 2° delle tabelle si riferisce alle etichette corrispondenti alle seconde (e ultime nel caso dei bigrammi) parole delle unità, e, solo nei trigrammi, la riga 3° corrisponde alle etichette date alla terza e ultima parola.

Tabella 9. Distribuzione delle etichette nei bigrammi del dataset FP, con suddivisione in quartili (Q).

BIGRAMMI FP		F	W	M	G	S	P	A
Q1	1°	27	1	12	6	5	4	1
	2°	-	1	-	-	1	-	-
Q2	1°	48	2	16	6	9	4	-
	2°	3	-	4	1	-	-	-
Q3	1°	41	2	23	7	10	6	4
	2°	4	-	1	-	1	1	-
Q4	1°	49	6	12	8	6	2	3
	2°	2	-	-	2	1	-	-

Tabella 10. Distribuzione delle etichette nei bigrammi del dataset FN, con suddivisione in quartili (Q).

BIGRAMMI FN		F	W	M	G	S	P	A
Q1	1°	6	1	4	-	1	1	1
	2°	-	-	-	-	-	-	-

Q2	1°	7	1	-	1	-	3	1
	2°	1	-	-	-	-	-	-
Q3	1°	13	-	3	-	1	3	3
	2°	-	1	-	-	-	-	-
Q4	1°	6	3	1	2	1	-	1
	2°	-	-	-	-	-	-	-

Tabella 11. Distribuzione delle etichette nei trigrammi del dataset FP, con suddivisione in quartili (Q).

TRIGRAMMI FP		F	W	M	G	S	P	A
Q1	1°	11	3	7	-	1	1	2
	2°	1	-	1	-	1	-	-
	3°	-	-	-	-	-	-	-
Q2	1°	21	2	7	6	2	1	1
	2°	2	-	-	1	1	-	-
	3°	-	-	-	-	-	-	-
Q3	1°	17	6	12	3	5	1	-
	2°	3	1	-	1	1	-	-
	3°	-	-	-	-	-	-	-
Q4	1°	11	1	7	5	4	2	1
	2°	1	1	-	1	2	1	-
	3°	-	-	-	1	-	-	-

Tabella 12. Distribuzione delle etichette nei trigrammi del dataset FN, con suddivisione in quartili (Q).

TRIGRAMMI FN		F	W	M	G	S	P	A
Q1	1°	1	1	1	1	1	2	-
	2°	1	1	-	-	-	-	-
	3°	-	-	-	-	-	-	-
Q2	1°	-	2	1	2	-	-	-
	2°	-	-	-	-	-	-	-
	3°	-	-	-	-	-	-	-
Q3	1°	2	-	1	-	-	-	2
	2°	-	-	-	-	-	-	-
	3°	-	-	-	-	-	-	-
Q4	1°	1	-	-	1	1	-	1
	2°	-	-	-	-	-	-	-
	3°	-	-	-	-	-	-	-

Guardando i grafici 31 e 32, si nota innanzitutto che bigrammi e trigrammi del dataset FP hanno più etichette rispetto a bigrammi e trigrammi FN, dato che non stupisce, dati i numeri maggiori di FP rispetto a FN, ma che conferma comunque una presenza più consistente di parole che possono essere problematiche. Inoltre, le etichette nei FP

compaiono più spesso in combinazione all'interno di uno stesso n-gramma, rispetto a quanto accade nei FN. Queste inclinazioni sembrano confermare la teoria del bias non intenzionale.

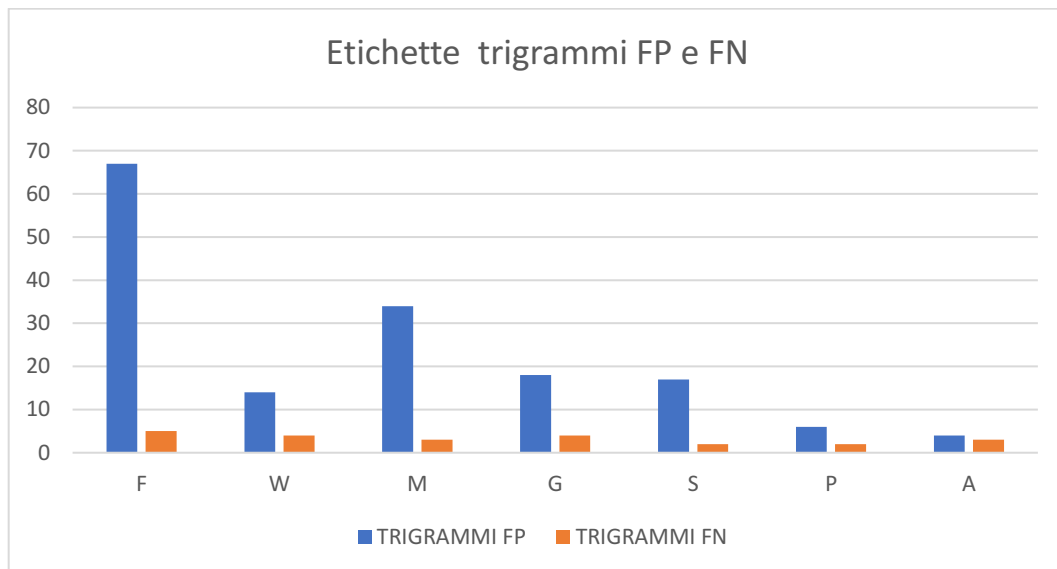


Grafico 35. Somma di tutte le etichette nei trigrammi appartenenti a una categoria, con distinzione fra FP e FN.

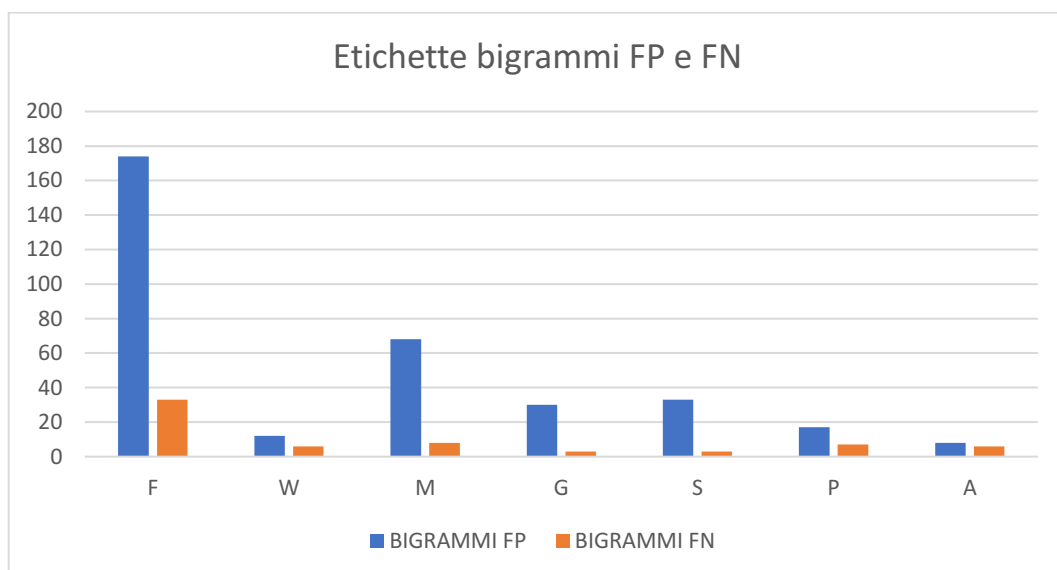


Grafico 36. Somma di tutte le etichette nei bigrammi appartenenti a una categoria, con distinzione fra FP e FN.

Guardando la distribuzione delle etichette all'interno dei quartili, si notano alcune tendenze: i bigrammi e i trigrammi FP, predetti come misogini ma che in realtà non lo sono, hanno sistematicamente più etichette negli ultimi due quartili, rispetto ai primi due (si vedano le tabelle 9 e 11), quindi i bigrammi e i trigrammi che hanno un impatto

maggiormente presentano effettivamente più parole *trigger* in tutte le categorie, rispetto alla prima metà di popolazione, avente meno impatto. L'unica eccezione è rappresentata dalla categoria che indica parole femminili (F) dei trigrammi, che conta 35 parole nei primi due quartili, contro le 32 degli ultimi due. Al contrario, la tendenza ad avere più etichette negli n-grammi più impattanti non si verifica negli n-grammi FN, in particolare nei trigrammi, dove anzi la tendenza è di avere meno etichette proprio in quelli con maggiore impatto, tranne che nelle categorie di donne (F) e altro (A), e in parte anche per le oscenità (P) dove la cifra rimane uguale. Bisogna però tenere a mente che si tratta di numeri molto bassi, ed è quindi difficile giungere a conclusioni. Nei bigrammi FN invece non c'è una tendenza dominante. Questa tendenza si nota in modo più accentuato suddividendo le liste in coppie di quartili (tabelle 13, 14, 15 e 16) in base al valore della mediana.

Tabella 13. Etichette dei trigrammi FP accorpate tra primo e secondo quartile (Q1 e Q2), e terzo e quarto (Q3 e Q4).

TRIGRAMMI FP	F	W	M	G	S	P	A
Q1 + Q2	35	5	15	7	5	2	3
Q3 + Q4	32	9	19	11	12	4	1

Tabella 14. Etichette dei trigrammi FN accorpate tra primo e secondo quartile (Q1 e Q2), e terzo e quarto (Q3 e Q4).

TRIGRAMMI FN	F	W	M	G	S	P	A
Q1 + Q2	2	4	2	3	1	2	0
Q3 + Q4	3	0	1	1	1	0	3

Tabella 15. Etichette dei bigrammi FP accorpate tra primo e secondo quartile (Q1 e Q2), e terzo e quarto (Q3 e Q4).

BIGRAMMI FP	F	W	M	G	S	P	A
Q1 + Q2	78	4	32	13	15	8	1
Q3 + Q4	96	8	36	17	18	9	7

Tabella 16. Etichette dei bigrammi FN accorpate tra primo e secondo quartile (Q1 e Q2), e terzo e quarto (Q3 e Q4).

BIGRAMMI FN	F	W	M	G	S	P	A
Q1 + Q2	14	2	4	1	1	4	2
Q3 + Q4	19	4	4	2	2	3	4

Da questi dati si può concludere che sicuramente i meme FP hanno molte più parole potenzialmente problematiche per l'ambito della misoginia, ma le parole oscene e dal

significato fortemente negativo (P) e i nomi delle pagine di creazione dei meme (A) non sono termini che, quantomeno presi singolarmente e non all'interno di un contesto frasale, hanno influenza nella classificazione. Le parole di genere femminile (F) invece costituiscono un discorso a parte: sia nei meme FN, che a livello di annotazione sono misogini, che in quelli FP, che a livello di annotazione non sono misogini, rappresentano la categoria più presente, e quindi la loro difficoltà di classificazione dipende strettamente sia dalla carica offensiva del singolo lemma (*prostitute*, che di per sé è una parola neutra, abbiamo visto nei grafici 4, 5, 6 e 7 che è presente in modo simile sia nei meme misogini che in quelli non misogini, mentre così non è per parole più offensive come *hoe* o *slut*), sia dal contesto, che per i lemmi neutrali (come *girlfriend* o *woman*) diventa ancora più rilevante.

Sui bigrammi e sui trigrammi FP e FN, ordinati in modo decrescente secondo la misura dell'impatto, ho calcolato anche i valori di asimmetria, un indice statistico applicabile alle curve di frequenza (Aron, Coups, Aron 2018). I valori dell'asimmetria, sono visibili nella Tabella 17.⁵⁹ In tutti i casi è un'asimmetria positiva: i valori più alti si concentrano nella parte iniziale della popolazione. Sia i bigrammi che i trigrammi del dataset FP hanno un'asimmetria maggiore di quelli FN, e i trigrammi di entrambi i dataset hanno invece un'asimmetria minore dei rispettivi bigrammi. Il valore dell'asimmetria è direttamente proporzionale al numero di n-grammi che hanno poco impatto per il modello, ma questo dato va relazionato con il campione preso in considerazione: sia i bigrammi che i trigrammi FP sono circa il triplo di quelli FN, e l'indice risulta quindi diluito in un campione molto ampio. Più che affermare che gli n-grammi FP hanno in media un impatto minore sul modello, si può invece dire che, rispetto a FN, gli n-grammi, e in particolare i bigrammi, di FP che hanno valori di impatto molto bassi sono di più.

Tabella 17. Valori dell'asimmetria in bigrammi e trigrammi FP e FN.

	FP	FN
Bigrammi	7,3	5
Trigrammi	2,7	2,5

⁵⁹ Sono valori calcolati su Excel non normalizzati; i valori standard dell'asimmetria ricoprono un intervallo che va da -1 a 1.

4.4.1 Referenti del trigramma più comune

All'interno dei meme si fa largo uso dei pronomi soggetto di prima e seconda persona singolare con una stessa funzione: raccontare un'esperienza personale o un fatto accaduto perché i fruitori del meme si possano identificare in quell'esperienza, che diventa quindi condivisa, creando un senso identitario. Hanno questo scopo per esempio i meme n. 615, 964 e 4581 (Figura 13), che fanno parte di un modello di meme molto usato, che inizia con *when you*, proseguendo con il racconto di una situazione, di solito breve, che molte persone hanno vissuto, o che si immagina o si prevede che possa accadere, e la narrazione viene accompagnata da un'immagine che fornisce più dettagli su ciò che è accaduto, o sulla reazione del soggetto alla situazione descritta. Oppure, si fa spesso uso di una sorta di copione di sceneggiatura, come nel meme n. 11227 (in Figura 14), dove due o più persone sono gli attori della scena, e i loro ruoli vengono introdotti da nomi o pronomi personali (*me, Trump, my girlfriend, feminists*) seguiti dai due punti, cui segue l'enunciato o il contesto del meme. Comunissimi, ma non appartenenti a un modello prototipico, sono i meme in cui si usa la prima persona singolare per raccontare la propria esperienza, o un pensiero, o un avvenimento vissuto in prima persona, con il medesimo intento di rivolgersi a quelle persone che hanno condiviso lo stesso vissuto o che pensano la stessa cosa.



Figura 12. Meme n.651, n. 964, n. 4581.

trump: men and women
will get \$1200

me : *identify myself
as both* get \$2400.



Figura 13. Meme n. 11227.

Secondo lo studio di Nissenbaum e Shifman (2018) condotto su 4.000 meme, di cui 1.000 in lingua inglese, i gruppi dominanti nella vita offline sono gli stessi che dominano quella online, e che quindi creano e dominano i contenuti della cultura digitale: i maggiori autori di meme sono maschi giovani, bianchi e di classe media. Dei 1.000 meme scritti in inglese, l'81% rappresenta uomini, e il 19% donne.⁶⁰

Individuare il genere dei referenti di un pronome personale può dare indicazioni non necessariamente su chi ha creato il meme, ma su chi invece è coinvolto nella realtà espressa dal meme. La ricerca sul genere del referente del trigramma può essere interessante per verificare se ci sono differenze sostanziali a seconda della classe dei meme, e che possono di conseguenza fornire ulteriori indizi sull'origine degli errori della classificazione del modello. Ho quindi scelto di selezionare il primo trigramma in quanto con frequenza assoluta più alta che contenesse un pronome, soggetto o possessivo, o un aggettivo possessivo. Il trigramma più comune in assoluto in entrambe le GT, con un totale di 235 occorrenze, è *I do not*, che contiene un pronome personale soggetto di prima persona singolare; questo trigramma è interessante perché il suo referente è un attore parlante, attivo, all'interno del contesto del meme. Chi sono quindi i soggetti parlanti che stanno dietro a questo trigramma, e come si distribuiscono fra meme misogini e meme non misogini?

⁶⁰ Se per i meme non misogini questa percentuale può essere buona, è facile pensare che per i meme misogini la percentuale maschile sia ancora più alta. L'identità collettiva più condivisa, e per questo legittimata, è quella maschile.

Tabella 18. Il genere dei referenti del trigramma "I do not" nei quattro dataset.

	GT+	TP	FN	GT-	TN	FP
Donna	48	46	2	20	19	1
Uomo	46	44	2	54	53	1
Generico	8	8	-	53	50	3
Animale	-	-	-	4	4	-

Controllando il genere dei referenti dei trigrammi *I do not* nei dataset GT+, GT-, FP e FN, si vede una netta distinzione fra meme classificati positivamente e negativamente come misogini (cfr. Tabella 13): mentre il genere dei soggetti nei meme non misogini è più frequentemente maschile, il genere dei soggetti nei meme misogini si distribuisce equamente fra donne e uomini. Il risultato è perfettamente in linea con la tipologia di meme che queste due classi rappresentano: i meme misogini per essere tali devono avere due ingredienti, le donne e qualcuno che le odi, che nei meme sono spesso, se non sempre, gli uomini. In tutti i meme della categoria positiva che ho guardato, non mi risulta esserci nessun caso in cui l'odio verso una o più donne parta da altre donne. I due generi canonici, donna e uomo, sono quindi gli ingredienti non solo necessari ma anche sufficienti per esprimere questi contenuti, e ciò spiega la quasi assenza di casi in cui il genere del soggetto non è comprensibile in modo esplicito, categoria che invece nei meme non misogini è largamente presente. Nei meme misogini le donne sono soggetti parlanti tanto quanto gli uomini sia perché sono il principale e unico argomento di questa classe di meme, sia per una sorta di comodità comunicativa: per esprimere un sentimento negativo e nocivo nei confronti di qualcuno spesso ci si oppone e si denigra ciò che questa persona dice, fa, o pensa, e, in un mezzo comunicativo che deve essere veloce e immediato, e che ha poco spazio per le parole, come sono i meme, per far dire qualcosa a una persona si fa largo uso di piccoli dialoghi o di citazioni. Far pronunciare qualcosa direttamente a una persona, usando la forma della citazione o del dialogo, è anche un metodo efficace per legittimare il proprio pensiero, e delegittimare quello degli altri. Come si vede dal meme n. 2985 (fig. 15), si possono mettere in bocca a qualcuno parole mai dette, per fare passare il messaggio che quella categoria, in questo caso le donne, è portatrice di un concetto condannabile.



Figura 14. Meme n. 2985.



Figura 15. Meme n. 5984, n. 4597, e n. 6152.

Fra i trigrammi *I do not* non misogini, quelli in cui il genere del soggetto non è ricavabile né da altri elementi testuali né dall'immagine sono il 40%. Buona parte di questi hanno come immagine quella di un maschio, che può essere un uomo, spesso un attore, o un personaggio maschile di un cartone animato, come nei meme in Figura 16. Se il trigramma non fa parte di una citazione diretta o di un dialogo i cui attori sono quelli delle immagini, allora, anche quando l'immagine rappresenta un uomo non ho considerato il soggetto come genere maschile. Questo perché i meme costruiti in questo modo vogliono narrare esperienze collettive, e non di un singolo soggetto, quindi nonostante è probabile che chi ha creato il meme sia un uomo, e che pensando all'esperienza o al pensiero da raccontare abbia pensato a un uomo e non a una donna, non è comunque possibile assegnare a priori il genere maschile.

Capitolo 5: *Word embedding*

Nell'ultimo capitolo mi concentro sugli *embedding* di parola e sulle rappresentazioni da questi derivate, i plot. Dopo avere spiegato cosa sono e come sono stati estratti i dati (paragrafo 5.1), analizzo le liste di similarità ricavate dai vettori (paragrafo 5.2), e nella sezione seguente (paragrafi 5.3, 5.3.1 e 5.3.2) continuo con la descrizione delle due diverse annotazioni che ho svolto su un campione di parole in vista dei plot, che analizzo nell'ultimo paragrafo (5.4).

5.1 Gli *embedding* di parola

I *word embedding* sono rappresentazioni di parole o frasi che, memorizzando le informazioni semantiche e sintattiche in un corpus, costituiscono uno spazio vettoriale in cui i vettori delle parole sono più vicini se le parole occorrono negli stessi contesti linguistici, cioè se sono riconosciute come semanticamente più simil, seguendo l'ipotesi della semantica distribuzionale (Lenci 2008). Nello specifico, con *embedding* si intende il vettore usato per la mappatura della parola (Jurafsky & Martin 2008). La semantica vettoriale si occupa quindi di trasformare l'ipotesi della semantica distribuzionale imparando i significati delle parole in base alla loro distribuzione all'interno del testo (Jurafsky & Martin 2008).

Il modello di codifica del linguaggio usato nella *challenge*, USE, è stato addestrato su Wikipedia, articoli internet, siti di domanda e risposta, e forum di discussione, ed è a partire da questi testi che ha ricavato le sue conoscenze linguistiche. Il modello, che non è stato sviluppato con lo scopo di riconoscere *hate speech*, è stato successivamente perfezionato con una tecnica di *fine tuning* in cui gli è stato insegnato a riconoscere contenuti misogini e d'odio. Una premessa da fare quando si parla di *word embedding* è che non sappiamo quali siano le caratteristiche su cui il modello basa la similarità tra parole, ma possiamo solo fare delle inferenze a riguardo. Inoltre, il fatto che due parole siano vicine non vuol dire necessariamente che si trovino spesso all'interno di uno stesso meme, ma solo che i contesti linguistici in cui ricorrono sono stati ritenuti per qualche ragione affini. Questa similarità può inoltre essere derivata sia dai testi somministrati al modello nella fase di *training*, che dai testi dei meme. Come si vedrà nei paragrafi successivi, in molti casi la vicinanza semantica

che il modello attribuisce ad alcune parole è molto accurata e di facile comprensione,⁶¹ mentre in altri casi le ragioni della similarità sono meno chiare.

5.2 Analisi delle liste di similarità

La prima analisi che ho svolto sui *word embedding* si è basata su liste di similarità. Le reti neurali comprendono il linguaggio in base a vettori, per cui più due parole sono vicine fra loro, e più sono semanticamente connesse, nel senso che compaiono in contesti simili. È possibile quindi ricavare a partire da questi vettori delle liste di parole semanticamente vicine a una parola target, liste che, per ricercare alcuni tipi di informazioni, sono di più facile lettura rispetto a un grafico. Per questo ho selezionato alcuni *identity term* in ognuna delle categorie descritte al paragrafo 3.2, in modo da avere parole di tutti i generi grammaticali in ognuno dei campi semantici delle categorie. L'unica eccezione riguarda le categorie degli *identity term* femminili e di quelli maschili, che ho accorpato in una nuova categoria per così dire neutrale (che chiamo *generici*), selezionando le parole più usate per i lemmi femminili (*woman* e *girl*), per i lemmi maschili (*man* e *boy*), e ho aggiunto una parola non compresa negli *identity term*, ossia *guy*, che può essere considerata una parola neutrale per genere, poiché all'interno dei meme viene usata sia per riferirsi alle donne che agli uomini, e ho ritenuto per questo interessante verificare se ci fossero differenze di rappresentazione fra parole esplicite per genere e una parola che invece non lo è. Non ho selezionato tutti gli *identity term* perché la lunghezza di queste liste ne rende l'analisi piuttosto lunga da svolgere. Gli *identity term* che ho selezionato sono:

- Generici:
 - Woman
 - Man
 - Girl
 - Boy
 - Guy
- Sessualità e/o estetica:
 - Prostitute
 - Hooker

⁶¹ Per fare un esempio, tutte le varianti di *mother* sono nella stessa zona del diagramma (Figura 15).

- Hoe
 - Slut
 - Whore
 - Bitch
 - Cougar
 - Milf
- Termini offensivi:
 - Chad
 - Stacy
 - Karen
 - Dishwasher
 - Idiot
 - Asshole
- Relazioni sentimentali:
 - Wife
 - Husband
 - Housewife
 - Girlfriend
 - Boyfriend
- Parentela:
 - Mother
 - Father
 - Parent
 - Daughter
 - Son

Per avere queste liste è stato scelto come vocabolario l'insieme delle parole di tutti e 10.000 i meme, poiché il vocabolario usato dal modello USE nella fase di *training* non è accessibile. La similarità fra parole è stata calcolata tramite la *cosine similarity* tra gli *embedding* dei termini target e quelli del vocabolario selezionato.⁶² Il valore di *threshold*

⁶² Cioè il *corpus* dei testi grezzi dei 10.000 meme.

scelto è di 0.90, e più la misura di similarità si avvicina a 1, maggiore è la somiglianza.⁶³ Le liste di similarità si possono vedere in appendice (lista 1).

Ciò che si nota da queste liste di similarità è in alcuni casi molto interessante. Innanzitutto, nelle coppie in cui il genere è identificato su base lessicale (come *woman* e *man*, o *mother* e *father*), la parola di genere femminile ha sistematicamente più parole che appaiono in contesti simili rispetto a quella maschile, tranne la coppia *wife* – *husband*, che ha un numero di sinonimi identico. Questo può dipendere dal fatto che le donne, all'interno dei meme, vengono associate a più concetti diversi, o descritte in modi più vari rispetto agli uomini, il che è connesso anche con il numero notevolmente più alto di frequenze assolute che alcune nomi femminili hanno rispetto alla loro controparte maschile.⁶⁴

La differenza più lampante si ha con la coppia *woman* – *man*. *Woman* ha quasi il triplo di parole simili rispetto a *man* (101 contro 37), e queste sono anche più genderizzate: si trovano parole come *patriarchal*, *misogynist*, *sexist*, *genderless*. In più, sotto *woman* si trova anche *wife*, ma sotto *man* non c'è nessun *husband*. Le parole correlate a *man* invece sono estremamente neutrali: *guy*, *dude* e sue varianti slang, *him*, *woman* e infine *gentleman*. Fra *boy* e *girl*, l'unica differenza interessante sembra essere la presenza di *child* e *kid* sotto *boy*, che non ha corrispettivi in *girl*. Inoltre, *man* è meno simile a *boy* di quanto *woman* non lo sia per *girl*. *Guy* invece non mostra particolarità, se non che la parola più vicina, dopo *guy* stessa, è una parola femminile, *girl*, e non una maschile, come mi sarei aspettata.

I lemmi di ambito sessuale e/o estetico più volgari (*hoe*, *slut*, *whore*, *bitch*) condividono una peculiarità con alcuni dei termini offensivi (*idiot* e *asshole*): presentano sia alcune parole che sembrano non avere senso, come *Ã,299*, *PokÃfmoy*, *FOX13*, *INA*, *LER*, *Ãœ*, e anche un numero (62460, sotto *asshole*), sia molte parole gergali appartenenti allo slang,⁶⁵ che invece non si trovano sotto tutti gli altri lemmi, sia molti termini offensivi. Ciò è facilmente spiegabile con la volgarità e l'offensività dei lemmi stessi, che quindi si inseriscono in contesti altrettanto bassi e offensivi. Invece sotto

⁶³ Non è possibile però avere un valore pari a 1 perché le approssimazioni dei valori fatte durante il passaggio dagli *embedding* impediscono una corrispondenza perfetta anche quando si tratta della stessa parola.

⁶⁴ Per menzionare alcuni numeri, il lemma *woman* ha 1891 occorrenze, contro le 880 di *man*; *girl* ha 796 occorrenze, mentre *boy* ne ha 159; *wife* ne ha 518 e *husband* 95.

⁶⁵ Ne riporto solo alcune: *Biatch*, *FUCKA*, *twat*, *THOT* (acronimo per *that ho over there*), *Niggaz*, *nigga*, *skank*.

prostitute e *hooker* non si trovano termini simili. Come si poteva sospettare dall'analisi sui lemmi, in cui *hooker* è risultata essere molto presente nei meme non misogini nonostante sia un termine offensivo, il modello non sembra riconoscere la differenza diafasica tra *prostitute* e *hooker*: non solo la loro rappresentazione nello spazio vettoriale è molto vicina, ma ognuna delle due costituisce la parola più vicina per l'altra, dopo la parola target stessa. Inoltre *prostitute* e *hooker* condividono alcuni termini come *hostesses* e *waitress*; in più, *prostitute* ha anche *maid* e *babysitter* fra le parole simili, che *hooker* non ha. È anche interessante notare che i lemmi della categoria della sessualità, che sono tutti femminili, non si trovano (almeno nell'intervallo di distanza considerato) in contesti molto simili a quelli di parole non offensive che si riferiscono in qualche modo a una donna, come *woman* o *girlfriend*; gli unici casi che si avvicinano a questa situazione sono casi di parole scritte male o appartenenti allo slang, come *awomen* (sotto *hoe*), *wifegf* (sotto *hoe*, *slut* e *bitch*), *girlfiend* (sotto *slut*, *whore*, ma anche sotto *girlfriend* e *boyfriend*) e *gurlz* (sotto *hoe*, *slut*, ma anche sotto *girl*).⁶⁶

I tre nomi propri usati per raffigurare i tipi umani, *chad*, *stacy* e *karen*, hanno fra le parole correlate soltanto altri nomi propri. Questo fa pensare che il modello non ne abbia compreso il vero significato. Passando a *dishwasher*, anche in questo caso non c'è nessun riferimento a persone umane, nemmeno con aggettivi, tranne che il nome di professione *plumbers*. Nonostante le apparenti carenze di questa rappresentazione, la classificazione di *dishwasher* viene sbagliato soltanto due volte dal modello.⁶⁷

I lemmi che si riferiscono a relazioni e parentela non hanno rappresentazioni particolarmente significative, se non che fra le parole collegate ad *housewife* appaiono alcuni termini relativi all'ambito lavorativo, come *maid* e *waitress* – la cui similarità con *housewife* è però facilmente comprensibile – *workplace*, *apprentice*, *businesswoman* e anche *bachelor* (che si trova anche per *wife*).

5.3 Selezione delle parole per i *plot*

Gli *embedding* di parola possono essere rappresentati su di un plot a due dimensioni: poiché è impossibile rappresentare le 512 dimensioni dei vettori del modello USE, usando un algoritmo euristico è possibile ridurle a 2. L'algoritmo usato per le

⁶⁶ Non trascrivo fedelmente i caratteri minuscoli e maiuscoli, poiché si trovano in diverse versioni.

⁶⁷ In meme FP, quindi non misogini a livello di annotazione.

rappresentazioni delle parole è t-SNE (*distributed stochastic neighbor embedding*) (Laurens & Hinton 2008), molto efficace per la comprensione di dati fortemente dimensionali e per la loro trasposizione in spazi con poche dimensioni. L'algoritmo seleziona automaticamente delle caratteristiche creando una rappresentazione significativa delle parole, e limitando la perdita di informazioni derivata dall'appiattimento delle dimensioni del vettore. Il valore di perplessità usato è pari a 5; maggiore è il valore della perplessità e maggiore è il valore della varianza tra i punti rappresentati, che tenderanno a raggrupparsi in strutture dense. Al contrario, una perplessità bassa comporta meno raggruppamenti fra token, rendendo la rappresentazione più sparsa. La posizione che una parola assume all'interno del plot è una media di tutte le posizioni della parola stessa.

L'idea iniziale era quella di estrarre sia delle rappresentazioni di parola che rappresentazioni di frase. Le rappresentazioni di frase avrebbero permesso di scegliere una serie di caratteristiche all'interno del diagramma, come la suddivisione in classe sia a livello di annotazione che a livello di predizione. Il codice necessario all'estrazione delle rappresentazioni interne di frase non è stato reso disponibile, e non è stato quindi possibile estrarre gli *embedding* di frase. Con la rappresentazione interna degli *embedding* di parola non è invece possibile ottenere una suddivisione per classe di annotazione e/o di predizione, perché non si può suddividere il dato di input in base alle quattro classi, per ciascuna delle quali è necessario fornire un *corpus*, i testi dei meme. Ho quindi scelto una serie di parole modellate in parte sugli *identity term* e in parte sulle parole etichettate nell'annotazione di bigrammi e trigrammi, etichettando queste parole in due modi differenti: sia suddividendole in base a criteri semantici (paragrafo 5.3.1), sia riproducendo, per quanto possibile, le classi misogino – non misogino a livello di annotazione e di predizione (paragrafo 5.3.2), tenendo conto del fatto che ad una parola non può essere attribuita più di un'etichetta. Le parole collezionate sono le stesse per entrambi i tipi di annotazione, e comprendono anche le loro versioni ortograficamente errate o non standard, poiché la rappresentazione dell'*embedding* non si basa sui testi normalizzati ma sui testi grezzi; in questo modo si può vedere come vengono trattate dal modello quelle parole che con molta probabilità sono state viste poche volte anche durante la fase di *training* perché appunto scritte in modo non convenzionale. Sempre per la stessa motivazione ho deciso anche di aggiungere alla lista le parole che nelle fasi di annotazione precedenti sono state accorpate sotto una parola seme.

5.3.1 Criteri di raggruppamento delle parole in base alla semantica

Le parole selezionate sono state suddivise in gruppi semantici, che in parte coincidono e in parte no alle categorie degli *identity term* e dei bigrammi e trigrammi. Elenco le parole suddivise per categorie:

- Donne:
 - Woman
 - Girl
 - Gurl
 - Gurlz
 - Girrllll
 - Lady
 - Female
 - Queen
 - Princess
 - Witch
 - Awoman
- Uomini:
 - Man
 - Boy
 - Dude
 - Bro
 - Male
 - Rapist
 - Therapist
 - Apist
 - Sir
 - Lord
 - Boss
- Leader
- Cowboy
- Witcher
- Gentleman
- gentleman
- Professioni:
 - Teacher
 - Writer
 - Doctor
 - Doc
 - Dentist
 - Scientist
 - President
 - Manager
 - Dealer⁶⁸
 - Chef
 - Cop
 - Officer
 - Soldier
 - Engineer
 - Programmer
 - Developer
 - Ceo
 - Judge

⁶⁸ Lo inserisco fra le professioni perché, nei 10.000 meme, in 11 casi su 12 occorrenze totali viene usato con il significato di "spacciatore/spacciatrice".

- Stripper
 - Pornstar
 - Worker
 - Miner
 - Employee
 - Comedian
 - Gardner
- Professioni donne:
 - Actress
 - Model
 - Supermodel
 - Waitress
 - Maid
 - Nurse
 - Professioni uomini:
 - Actor
 - Bouncer
 - Speedster
 - Veteran
 - Vetran
 - Sessualità/estetica:
 - Milf⁶⁹
 - Cougar
 - Transgender
 - Trans
 - Transgend
 - Tran
 - Cisgender
 - Lesbian
 - Virgin
- Incel
 - Chad
 - Stacy
 - Teenager
 - Teen
 - Bodybuilder
- Prostituzione:
 - Bitch
 - Bitchhhhh
 - Hoe
 - Hooker
 - Puta
 - Prostitute
 - Slut
 - Whore
 - Offensivi:
 - Dishwasher
 - Karen
 - Asshole
 - Idiot
 - Motherfucker
 - Parentela:
 - Mother
 - Mom
 - Mum
 - Mama
 - Momma
 - Mommy
 - Mummy
 - Father

⁶⁹ Non ho aggiunto le altre scritte come "M.I.L.F." e simili perché, grazie alla lista di

similitudine della parola, si vede che vengono ricondotte a *milf*.

- Dad
 - Parent
 - Daughter
 - Son
 - Sister
 - Brother
 - Grandmother
 - Grandma
 - Granny
 - Grandfather
 - Grandpa
- Girlfrand
 - Gf
 - Boyfriend
 - Bf
 - Crush
 - Daddy
- Generici:
 - Feminist
 - Femenist
 - Guy
 - Discordi relativi al genere:
 - Home
 - House
 - Kitchen
 - Cook
 - Sandwich
 - Feminism
 - Equal
 - Right
 - Manspreading
 - Sexism
 - Prostitution
- Relazioni:
 - Wife
 - Housewife
 - Husband
 - Spouse
 - Divorced
 - Divorcée
 - Fiancee
 - Fiancé
 - Fiancée
 - Girlfriend

Il plot si può vedere in Figura 17.

5.3.2 Criteri di raggruppamento delle parole in base alla annotazione e alla predizione del modello

Poiché non è stato possibile estrarre dei *plot* con suddivisione a seconda dell'annotazione e della predizione, ho cercato di riprodurre, per quanto possibile, questa suddivisione con delle categorie. Ho usato le medesime parole elencate nel paragrafo precedente, ma raggruppate in modo differente. Per ottenere una

rappresentazione il più chiara possibile, che si traduce nell'avere meno categorie possibili, mi sono limitata a otto combinazioni di annotazione e predizione, che elenco di seguito:

- Altro: le parole che sono presenti in tutti e 4 i dataset (GT+, GT-, FP, FN);
- FN: le parole GT+ che però vengono sistematicamente classificate male come FN;
- FP: le parole GT- che però vengono sistematicamente classificate male come FP;
- GT: le parole la cui classificazione è sempre corretta (quindi quelle che si trovano solo in GT+ e GT-);
- MAI FN: le parole che non vengono mai classificate come FN (e possono comparire in due contesti: a) GT+, FP, GT-; b) GT-, FP);
- MAI FP: le parole che non vengono mai classificate come FP (e possono comparire in due contesti: a) GT-, FN, GT+; b) GT+, FN);
- TN: parole che compaiono solo nel dataset TN;
- TP: parole che compaiono solo nel dataset TP.

Riporto le parole secondo questa suddivisione:

- Altro:
 - Veteran
 - Hoe
 - Hooker
 - Bitch
 - Slut
 - Puta
 - Prostitute
 - Maid
 - Model
 - Milf
 - Cougar
 - Virgin
 - Chad
 - Teen
 - Stacy
 - Incel
 - Transgender
 - Home
 - House
 - Kitchen
 - Sandwich
 - Feminist
 - Guy
 - Feminism
 - Wife
 - Husband
 - Spouse
 - Girlfriend
 - Boyfriend
 - Housewife
 - Mother

- Mom
- Gf
- Dad
- Daddy
- Crush
- Grandma
- Karen
- Son
- Daughter
- Sister
- Granny
- Grandfather
- Teacher
- Doctor
- President
- Bodybuilder
- Lesbian
- Trans
- Pornstar
- Miner
- Man
- Boy
- Male
- Rapist
- Bro
- Boss
- Leader
- Cop
- Woman
- Girl
- Female
- Witch

- GT:

- Lord
- Gentleman
- Chef
- Queen
- Ceo
- Officer
- Programmer
- Soldier
- Cisgender
- Nurse
- Motherfucker
- Bf
- Fiancé
- Divorcée
- Divorced

- TP:

- Gurlz
- Waitress
- Supermodel
- Bitchhhhh
- Momma
- Grandmother
- Femenist
- Fiancee
- Fiancée

- TN:

- Witcher
- Grrllll
- Developer
- Writer
- Tran

- Mummy
- Grandpa
- Girlfrand
- FP:
 - Cowboy
 - Gentleman
 - Gardner
 - Bouncer
- FN:
 - Apist
 - Speedster
 - Vetran
- Mai FN:
 - Gurl
 - Lady
 - Princess
 - Awoman
 - Dude
 - Sir
 - Engineer
 - Judge
 - Stripper
 - Doc
 - Manager
 - Employee
 - Worker
- Comedian
- Incel
- Transgender
- Actress
- Asshole
- Mama
- Dishwasher
- Parent
- Mum
- Mommy
- Prostitution
- Manspreading
- Equal
- Right
- Sexism
- Cook
- Mai FP:
 - Teenager
 - Dealer
 - Scientist
 - Father
 - Idiot
 - Brother
 - Whore
 - Actor
 - Dentist

Il plot è riportato in Figura 18.

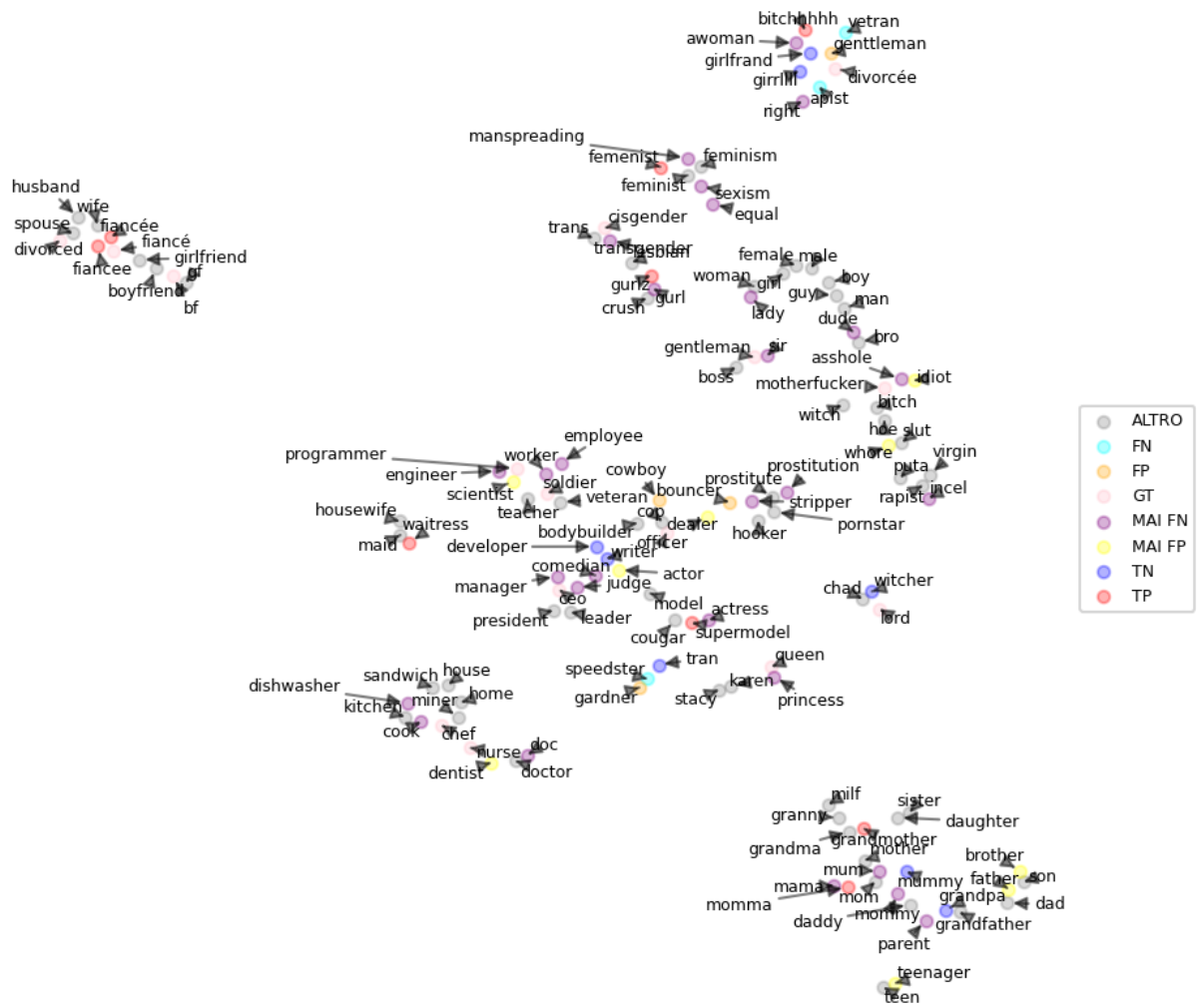


Figura 17. Plot delle parole con categorizzazione su base di annotazione e predizione dei modelli. Segnalo un errore: il neologismo *divorcée* dovrebbe essere segnato in giallo, poiché appartiene alle classi FN e TP.

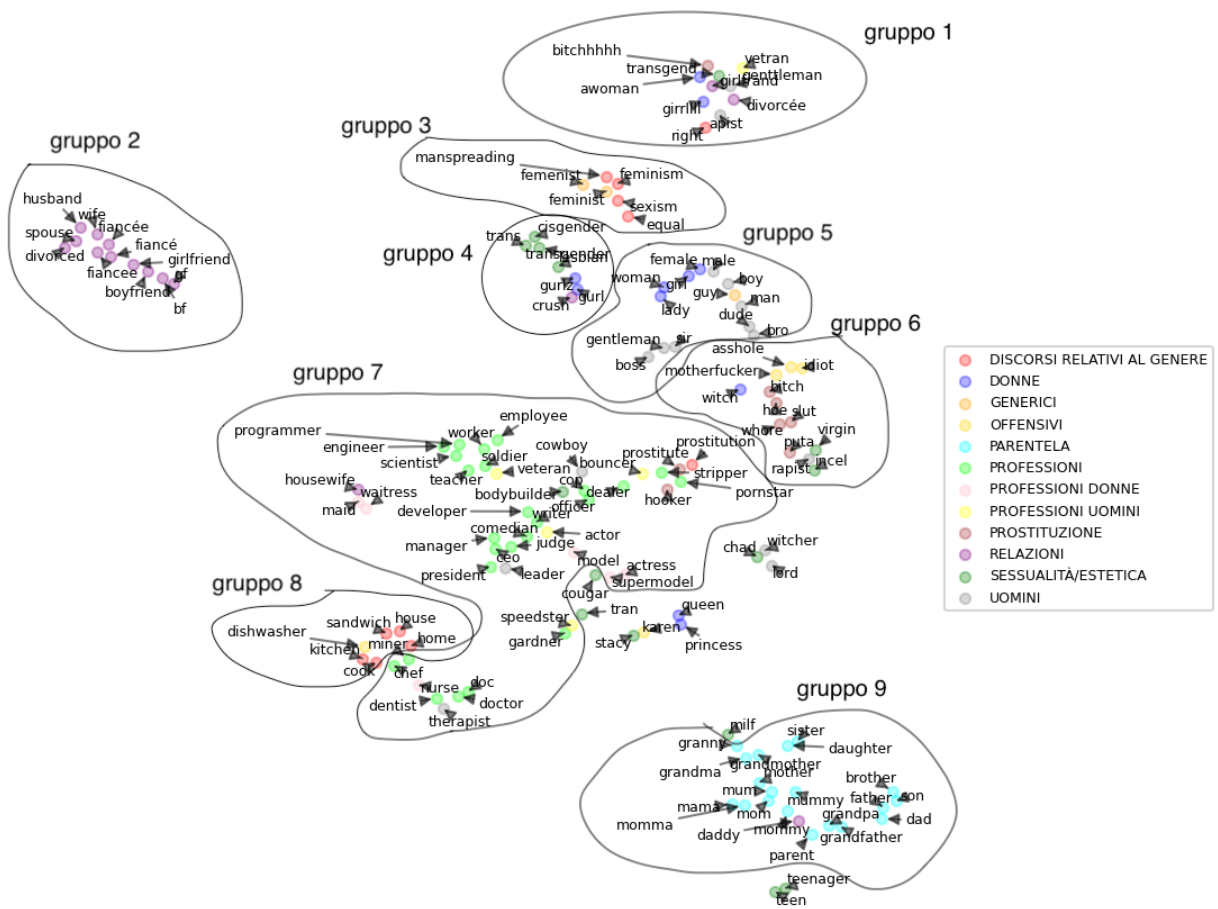


Figura 18. Plot con suddivisione per macro-gruppi semantici.

La disposizione delle parole nei plot in Figura 17 e Figura 18 è uguale, ma cambia l'etichettatura. In generale, si tratta di rappresentazioni molto parziali, per varie ragioni. Innanzitutto, come già detto, la posizione di una parola è la traduzione di un algoritmo, che appiattisce in due dimensioni qualcosa che in origine ne ha 512. Poi, solo un campione estremamente ristretto di parole è stato rappresentato,⁷⁰ e vengono tralasciate le restanti centinaia di migliaia del vocabolario dei testi. È quindi molto difficile trarre delle conclusioni o cogliere dei collegamenti a partire dai plot, tanto più che le logiche interne ai modelli sono a noi sconosciute.

La maggior parte degli errori di classificazione sono stati commessi verso la classe misogina, e difatti nel plot in Figura 18 si può vedere che le parole che sono sempre state classificate come FP e mai FN sono molto diffuse. La classe più rappresentata rimane comunque quella dei termini sbagliati sia come FP che come FN (ALTRO). A livello semantico, le parole si distribuiscono in modo abbastanza conforme.

⁷⁰ Sono in totale 140 parole.

In Figura 19 ho segnato i gruppi semantici che corrispondono in parte all'annotazione semantica da me fatta, ma in modo meno granulare: non c'è distinzione in quanto al genere delle parole. Si possono vedere i macro-gruppi di relazioni sentimentali (gruppo 2), di parentela (gruppo 9), le professioni (gruppo 7), parole sui discorsi di genere (gruppo 3), sull'identità di genere e l'orientamento sessuale (gruppo 4), i termini neutrali e quelli offensivi (gruppi 5 e 6), e le parole che hanno a che fare con la casa (gruppo 8) che io ho individuato come potenzialmente problematiche perché legate a stereotipi di genere. Inoltre, tutte le parole scritte male o con grafie non standard si concentrano in un'unica zona, ad eccezione di *tran*,⁷¹ che compare in una posizione molto distante, ma si trova vicino ad altre tre parole che ricorrono in totale una volta ciascuna (*gardner*, *speedster* e *supermodel*); di *fiancee*, che però si trova vicino sia al maschile *fiancé* che al femminile *fiancée*, fra i termini delle relazioni sentimentali; e di *femenist*, anche questo avvicinato alla versione corretta *feminist*. Negli ultimi due casi evidentemente i contesti sono semanticamente molto simili a quelli delle grafie corrette. Questa disposizione conferma che il modello ha difficoltà evidenti nella gestione degli errori di battitura, e difatti USE non è implementato a livello di carattere in modo da fare fronte a stringhe dove un solo carattere è sbagliato. Si vede anche una notevole disomogeneità a livello di classificazione (Figura 18) rispetto a tutti gli altri gruppi, dove si può invece individuare una tendenza nella direzione delle predizioni.

Alcune parole appartenenti alla varietà slang sono inserite molto bene per quanto riguarda le relazioni semantiche: *dude* e *bro* sono nel gruppo dei nomi neutrali, vicino a *man*, *guy* e *boy*; *gurlz* e *gurl* sono vicine tra loro, e abbastanza vicine a *woman* e *girl*; *incel* è vicino a *virgin*; le abbreviazioni *bf* e *gf* sono affianco alle forme estese *boyfriend* e *girlfriend*. Di contro, *crush* è molto lontano dal gruppo dove si trovano le altre parole che indicano relazioni sentimentali,⁷² e *chad* e *stacy* sono abbastanza lontane fra loro. Tra *chad*, *witcher* e *lord*, che sono vicini tra loro e separati da altre parole, non ho trovato particolari correlazioni, se non il fatto che nelle immagini ci siano spesso uomini

⁷¹ Compare due volte in GT-, di cui una con significato di *trans*.

⁷² Preciso che come verbo appare solo 5 volte, mentre come nome 25, di cui però due non sono riconosciute da SketchEngine come nomi (si tratta del meme n. 5202, il cui testo è "CRUSH ME ME CRUSH", che rappresenta una scenetta in cui il nome e il pronome soggetto sono apposti a degli animali, a simboleggiare che gli atteggiamenti che questi animali assumono rappresentano i sentimenti dei due protagonisti ideali: *crush* è quindi da intendersi come un *identity term*). Di queste 25 occorrenze di *crush* come nome, ben 17 indicano la persona per cui si prova infatuazione, mentre il significato delle altre si divide fra l'infatuazione stessa, e la botta o lo scontro che si può fare per esempio con un veicolo.

muoscolosi, o primi piani di uomini con sfondo scuro e/o ambientazioni e vestiti dal sapore un po' medievale. *Chad*, *stacy* e *karen* sono comunque lontani dalle parole offensive, come è emerso dalle liste di similarità, in cui i vicini erano solo altri nomi propri.

Tra i termini che indicano relazioni di parentela spicca *milf*, che, nonostante con le madri abbia poco a che fare, benchè sia un acronimo contenente *mother*, evidentemente compare in contesti più simili a quelli di *mother* che di altri termini collegati alla sessualità. *Teen* e *teenager*, nonostante non indichino parentele, sono collegati con l'ambiente familiare: *teenager* si usa spesso proprio in contesti familiari, mentre *teen*, quando indica categorie porno, è spesso in combinazione con altri termini familiari. *Brother*, *father* e *teenager* sono le uniche tre parole di questo gruppo che non vengono mai classificate come FP. *Daddy* è inserito in questa zona; d'altronde 34 volte su 48 ha proprio il significato di *padre*.

I nomi della categoria *prostituzione* si concentrano nella stessa zona dove sono presenti gli insulti, tranne che per *hooker* e *prostitute*, di cui ho già parlato. La categoria *sessualità/estetica* è suddivisa in due zone principali, in accordo con identità di genere e orientamento sessuale da una parte, e *virgin* e *incel* dall'altra. Dalla posizione di *cougar* si può intuire che la caratteristica semantica principale catturata dai modelli è l'aspetto estetico: si trova infatti vicina a *model*, *supermodel* e *actress*; anche per *actress* sembra quindi che una piacevole estetica sia più importante di altre caratteristiche.

Il plot suddiviso per classificazione conferma che le parole collegate ai discorsi relativi al genere tendono ad essere considerati misogini anche quando non lo sono: 6 su 11 vengono classificati solo come FP, e guardando le frequenze assolute nei quattro dataset si vede che 27 volte sono presenti in FN, contro le 103 di FP. La categoria si divide fra parole strettamente collegate a discorsi sul genere come *feminism*, *manspreading* e *sexism*, e parole che invece rientrano nella sfera domestica, tra cui c'è anche *dishwasher*, che, come si è visto dalla lista di similarità, non sembra essere riconosciuto come un *identity term*. Le uniche parole che si discostano dai due gruppi sono il nome *prostitution*, che si trova comprensibilmente vicino a *prostitute*, e l'aggettivo *right* che è molto vicino alle parole scritte con errori di battitura. Di tutte le parole appartenenti a questa categoria semantica dei discorsi relativi al genere, soltanto una

parola su 11, *femenist*, non ha classificazioni sbagliate, configurandosi così come il gruppo che ha più parole classificate in modo scorretto di tutti.

Dal plot è emerso che il modello, a meno che il contesto non sia estremamente chiaro, non è capace di gestire le disgrafie, e che le parole appartenenti alla varietà gergale sono problematiche. Per il primo punto esistono soluzioni di codice interne all'architettura delle reti neurali, mentre il secondo punto è strettamente connesso alla scarsità dei dati: è difficile che un modello neurale individui correttamente le relazioni semantiche di una parola che non è stata mai vista prima, e questo si scontra con la grande creatività della lingua dei meme, che fa invece largo uso di parole non convenzionali e sempre nuove. La categorizzazione per predizione (Figura 18) ha confermato che le parole più frequenti sono inserite in due gruppi di predizione: o vengono predette sia come falsi positivi che come falsi negativi, oppure soltanto come falsi positivi. Al di sopra delle 30 occorrenze sono solo tre le eccezioni:⁷³ *father*, *brother* che sono parole maschili e della più stretta cerchia familiare, e *whore*, nome dispregiativo che per sua natura non può che essere usato in contesti misogini.

⁷³ Le eccezioni sono comunque poche anche abbassando la soglia a 25 occorrenze.

Conclusione

In questa tesi ho cercato di comprendere le ragioni dell'errore nel rilevamento di contenuti misogini nei meme in modelli di *deep learning*, usando metodi comuni della linguistica dei *corpora*, come l'analisi delle frequenze delle parole nel secondo capitolo e quella degli n-grammi presentata nel terzo capitolo, ampliando queste analisi con processi di annotazione semantica. Mi sono avvalsa anche di alcuni metodi statistici di base, e di alcune metriche nate appositamente per analizzare gli errori nel *machine learning*, quali la misura micro-F.

Il primo fattore da evidenziare riguarda l'importanza dei processi di pulitura e di normalizzazione del testo, che sono risultati fondamentali per tutti i processi di analisi svolti in questo lavoro. Anche per gli *embedding* di parola, grazie alle liste di similarità, si è visto che le forme plurali di un lemma, e i caratteri maiuscoli e minuscolo non sono elementi di disturbo. A tutti i livelli di analisi dell'errore è emerso poi un aspetto, il bias non intenzionale: ci sono una serie di *identity term* femminili e di parole relative a discorsi di genere che spingono i modelli ad andare verso una classificazione misogina del meme, anche quando il messaggio d'odio non è presente, perché statisticamente si trovano più spesso in contesti che veicolano odio verso le donne. Con gli *identity term* ciò è emerso con più chiarezza dal confronto tra le frequenze assolute dei nomi femminili (dal grafico 3 al grafico 6) e di quelli maschili, dove nel primo caso la GT+ ma soprattutto i FP hanno un numero sensibilmente più alto rispetto ai meme negativi e ai falsi negativi, mentre la differenza non è così ampia con i nomi maschili. Passando ai bigrammi e ai trigrammi, il bias non intenzionale è emerso soprattutto con le parole relative ai discorsi di genere, ma in generale si è osservato che tutte categorie individuate hanno decisamente più unità nel dataset dei FP. Bisognerebbe quindi limitare il peso che alcune parole hanno sulle macchine, e privilegiare invece l'intero contesto. Si è visto anche che i modelli hanno qualche difficoltà a far dialogare i contenuti espressi dalle due modalità di comunicazione che il meme combina, l'immagine e il testo, sia nei casi in cui non sono coerenti tra di loro sia quando una delle due fa uso di elementi impliciti o ambigui, e l'implicitezza del contenuto è un altro dei problemi emersi, come per esempio nel caso dei meme FN n. 2945, 8557 e 2636 (tutti in Figura 9), che sono la stessa variante di una stessa barzelletta, che combina neutralità dell'immagine ad implicitezza del contenuto, o dei meme FP n. 8480 e n. 10997 (Figura 9), in cui rispettivamente una scena pornografica e dei disegni di nudo femminile sono accompagnati da testi del tutto neutrali.

È stato poi messo in luce, in particolar modo con le rappresentazioni bidimensionali degli *embedding* di parola, che la scarsità dei dati, in cui è compresa anche la fase di *training*

del modello di riconoscimento del linguaggio, costituisce un ulteriore problema. Soprattutto dall'analisi sugli *identity term*, nel capitolo 3, si è visto che ci sono effettivamente delle somiglianze non trascurabili tra i meme predetti correttamente come misogini e quelli predetti come misogini ma in modo errato; si è visto anche che la stessa somiglianza non è così evidente fra i meme predetti sia correttamente sia non correttamente come non misogini (TP e FP), confermando il bias non intenzionale verso la classificazione misogina, tendenza già evidente dai numeri sbilanciati dei dataset FP e FN. Disponendo di un dataset più grande, il modello potrebbe imparare a riconoscere anche le differenze più sottili, poiché avrebbe più contesti dai quali trarre relazioni semantiche tra parole, che coprirebbero quindi un maggior numero di possibilità attraverso le quali un certo contenuto può essere espresso. Tutte queste considerazioni sono però da riportare sia alla parzialità del lavoro, che è stato svolto principalmente su uno solo dei due elementi che compongono il meme, cioè il testo, mentre l'immagine è rimasta sullo sfondo, sia ai criteri di annotazione che non sono coerenti e che non possono quindi portare ad avere dei dati conformi a ciò che è misogino e ciò che non lo è.

In questo lavoro ho guardato principalmente al lessico, ma indagare altri livelli della lingua, e in particolare la sintassi, potrebbe essere molto interessante: è possibile che le strutture più profonde della lingua, come per esempio le dipendenze sintattiche, portino in luce configurazioni ricorrenti nei modi in cui i messaggi di odio verso le donne vengono comunicati. SketchEngine stesso offre altri strumenti di indagine del *corpus* più vicini alla sintassi, come la funzione *word sketch* che estrae i collocati, ossia quelle combinazioni di parole che hanno un grado di associazione forte; SketchEngine restituisce anche, con questa funzione, il comportamento sintattico delle parole, che, per l'inglese, prevede fino a 27 relazioni grammaticali, e offre la possibilità di estrarre collocati formati anche da più di due parole. Per indagare invece sempre il livello lessicale, un'altra possibilità di SketchEngine è quella di estrarre le parole chiave, ossia quelle parole che, in realzione ad un altro *corpus* di testi, si presentano con una frequenza relativa maggiore, e che quindi caratterizzano il *corpus* di focus rispetto al *corpus* di riferimento (Kilgarriff et al. 2004; Rychlý 2008; Kilgarriff et al. 2014). Tornando alla sintassi, si potrebbero creare, quantomeno per il testo, dei *corpora* gold, annotati da esperti di misoginia,⁷⁴ magari con un'annotazione morfo-sintattica oltre che semantica. Ad oggi non pare che esistano delle catene di annotazione automatica in grado di gestire testi rumorosi come quelli dei meme, che, per l'organizzazione del contenuto testuale

⁷⁴ Come è stato fatto per i 1.000 meme del dataset usato come *test* per la *challenge* di MAMI.

spesso implicito grazie ai rimandi intrinseci all'immagine, si differenziano almeno in parte dai testi dei *social network* e della messaggistica istantanea. Sempre per fronteggiare la multimodalità dei meme, porterebbe sicuramente a risultati innanzitutto meno parziali, ma anche più precisi, combinare le analisi linguistiche svolte sui testi ad analisi svolte invece sull'immagine.

Appendice

1. Tabelle delle frequenze assolute degli *identity term*

Tabella 19. Frequenze assolute degli identity term femminili nei quattro dataset.

I.T. DONNE	GT+	FN	GT-	FP
woman	1418	10	473	219
girl	628	14	168	71
feminist	218	1	#N/D	#N/D
wife	195	33	323	26
girlfriend	158	26	410	21
hooker	135	11	111	15
mother	124	18	210	16
dishwasher	121	#N/D	#N/D	#N/D
bitch	113	1	#N/D	#N/D
milf	74	9	54	10
whore	66	#N/D	5	#N/D
prostitute	65	11	142	29
lady	48	#N/D	18	8
cougar	48	2	28	7
slut	48	1	7	1
female	35	2	55	9
housewife	27	2	24	6
daughter	26	5	27	5
hoe	22	1	3	1
sister	15	#N/D	25	2
karen	14	1	18	2
queen	12	#N/D	6	#N/D
maid	10	1	2	1
model	9	#N/D	3	2
witch	8	#N/D	97	2
nurse	7	#N/D	8	#N/D
actress	7	#N/D	2	1
waitress	6	#N/D	#N/D	#N/D
crush	6	1	17	1
princess	6	#N/D	5	1
stacy	5	2	#N/D	#N/D
grandmother	4	#N/D	20	2
puta	3	1	#N/D	#N/D
divorced	2	#N/D	5	#N/D
fiancé	1	1	#N/D	#N/D
lesbian	0	#N/D	4	2

Tabella 20. Frequenze assolute degli identity term maschili nei quattro dataset.

I.T. UOMINI	GT+	FN	GT-	FP
man	533	10	347	109
boy	88	6	71	13
husband	51	7	44	6
father	40	5	120	3
son	32	3	66	2
boyfriend	28	3	49	3
dude	20	#N/D	16	1
male	19	1	23	7
rapist	13	2	15	5
bro	12	1	30	1
chad	10	1	10	1
daddy	9	#N/D	5	1
sir	7	#N/D	19	5
brother	7	2	30	#N/D
boss	6	1	18	1
incel	5	#N/D	2	1
actor	5	1	3	#N/D
leader	4	1	7	1
lord	2	#N/D	10	#N/D
speedster	1	1	#N/D	#N/D
veteran	1	1	#N/D	#N/D
witcher	#N/D	#N/D	22	#N/D
grandfather	#N/D	#N/D	11	1
bouncer	#N/D	#N/D	1	1
cowboy	#N/D	#N/D	1	1

Tabella 21. Frequenze assolute degli identity term di professione nei dataset.

I.T. PROFESSIONI	GT+	FN	GT-	FP
prostitute	65	11	142	29
teacher	20	1	46	2
doctor	19	2	50	3
scientist	13	2	13	#N/D
president	11	1	35	3
maid	10	1	2	1
model	9	#N/D	3	2
manager	9	#N/D	4	1
dealer	8	1	4	#N/D
chef	8	#N/D	13	#N/D
nurse	7	#N/D	8	#N/D
employee	7	#N/D	9	1
actress	7	#N/D	2	1
waitress	6	#N/D	#N/D	#N/D
cop	6	2	17	3

boss	6	1	18	1
engineer	6	#N/D	7	1
actor	5	1	3	#N/D
ceo	5	#N/D	2	#N/D
judge	5	#N/D	7	2
stripper	5	#N/D	4	2
dentist	4	1	2	#N/D
worker	4	#N/D	8	1
pornstar	4	1	2	1
miner	3	1	#N/D	#N/D
officer	3	#N/D	7	#N/D
comedian	3	#N/D	1	1
programmer	1	#N/D	18	#N/D
soldier	1	#N/D	8	#N/D
developer	#N/D	#N/D	12	#N/D
writer	#N/D	#N/D	7	#N/D
bouncer	#N/D	#N/D	1	1
gardner	#N/D	#N/D	1	1

Tabella 22. Frequenze assolute degli identity term relativi alla sessualità e/o all'estetica.

I.T. SESSUALITÀ/ ESTETICA	GT+	FN	GT-	FP
hooker	135	11	111	15
bitch	113	1	#N/D	#N/D
milf	74	9	54	10
whore	66	#N/D	5	#N/D
prostitute	65	11	142	29
cougar	48	2	28	7
slut	48	1	7	1
transgender	23	1	22	4
hoe	22	1	3	1
virgin	12	2	13	#N/D
chad	10	1	10	1
daddy	9	#N/D	5	1
teen	5	1	3	#N/D
cisgender	5	#N/D	1	#N/D
stripper	5	#N/D	4	2
incel	5	#N/D	2	1
stacy	5	2	#N/D	#N/D
pornstar	4	1	2	1
puta	3	1	#N/D	#N/D
bodybuilder	#N/D	#N/D	4	2
lesbian	#N/D	#N/D	4	2

Tabella 23. Frequenze assolute degli identity term offensivi.

I.T. OFFENSIVI	GT+	FN	GT-	FP
hooker	135	11	111	15
dishwasher	121	#N/D	#N/D	#N/D
bitch	113	1	34	9
whore	66	#N/D	5	#N/D
slut	48	1	7	1
hoe	22	1	3	1
karen	14	1	18	2
idiot	7	#N/D	22	#N/D
asshole	5	#N/D	9	2
puta	3	1	#N/D	#N/D
motherfucker	1	#N/D	9	#N/D

Tabella 24. Frequenze assolute degli identity term delle relazioni sentimentali.

I.T. RELAZIONI	GT+	FN	GT-	FP
wife	195	33	323	26
girlfriend	158	26	410	21
husband	51	7	44	6
boyfriend	28	3	49	3
housewife	27	2	24	6
daddy	9	#N/D	5	1
crush	6	1	17	1
divorced	2	1	5	#N/D
fiancé	2	1	1	#N/D
spouse	#N/D	#N/D	5	1

Tabella 25. Frequenze assolute degli identity term di parentela.

I.T. PARENTELA	GT+	FN	GT-	FP
mother	124	18	210	32
father	40	5	120	3
son	32	3	66	2
daughter	26	5	27	5
sister	15	#N/D	25	2
brother	7	2	30	#N/D
grandmother	4	#N/D	20	2
parent	2	#N/D	55	1
grandfather	#N/D	#N/D	11	1

2. Tabelle con annotazione semantica di bigrammi e trigrammi FP e FN

Tabella 26. Etichettatura sui bigrammi FP, ordinati in modo decrescente a seconda della misura dell'impatto.

Etichetta riferita alla seconda parola del bigramma	Etichetta riferita alla prima parola del bigramma	BIGRAMMI FP	impatto -
		be a	0,0008623268335545630
		i be	0,0007748248846698920
		it be	0,0006709653726530360
		do not	0,0006692980869486170
		you be	0,0006237610655286500
		when you	0,0004453856367333910
		in the	0,0004032080571182070
		be not	0,0003783763623654480
	F	woman be	0,0003440381319354960
	F	a woman	0,0003413114893980120
	F	she be	0,0002958032908116100
		be the	0,0002412880252949360
		can not	0,0002310963209465700
		if you	0,0002194834404969940
		i have	0,0002190814770940780
	F	my wife	0,0002160900663396500
		you have	0,0002106418380712990
		this be	0,0002083993608536040
	F	woman s	0,0002079866888519130
		have a	0,0002019413119959830
	M	a man	0,0001828454407489350
		to the	0,0001814296021590650
		there be	0,0001798299269355370
		of the	0,0001796047099893920
	F	when she	0,0001716715527031670
		that be	0,0001702762541616020
		in a	0,0001643351615620060
		be you	0,0001620675497547380
	A	with mematic	0,0001607298493440640
		make with	0,0001604976646573990
		who be	0,0001586583618149130
		do you	0,0001572182844864860
		you do	0,0001565682516592680
	F	a girl	0,0001534086215645320
	F	woman in	0,0001530017021439360
		to be	0,0001529058364157760
		for the	0,0001527383601436890
		they be	0,0001479016454058050
	M	he be	0,0001464889635767630
	W	a prostitute	0,0001446863922448090
		as a	0,0001434218035241650
		look like	0,0001398805244248030
		when the	0,0001355301731867110
		when a	0,0001307344901354880
		out of	0,0001275937216037050
		the first	0,0001245090382242750
	G	a feminist	0,0001232513711715040
	F	this woman	0,0001232513711715040

		to get	0,0001230196956617830
		on the	0,0001225252953472250
		i do	0,0001202477663278290
	M	man be	0,0001188495364868080
	G	the kitchen	0,0001188495364868080
	G	equal right	0,0001188495364868080
G	M	male feminist	0,0001188495364868080
		when i	0,0001176956574917900
		go to	0,0001174261887444510
		not a	0,0001170756628079000
	W	the prostitute	0,0001155481604732850
	F	of woman	0,0001134472848283160
		need to	0,0001127905405090490
		you can	0,0001121686453825200
		get a	0,0001120229623571510
		in my	0,0001112633958599900
		and you	0,0001111994513796110
		with a	0,0001110815929255780
		try to	0,0001095540420129790
		see a	0,0001084737833014510
		have to	0,0001074346092536110
		like a	0,0001058717135370990
		will not	0,0001055004943451740
	F	woman have	0,0001039933444259570
		a a	0,0001039933444259570
	S	a bikini	0,0001039933444259570
		you will	0,0001035431567877490
		of my	0,0001016476298900330
		why be	0,0001011943517188400
		be just	0,0001006133642216360
		want to	0,0001002064719398580
		in your	0,0000990412804056731
	F	woman without	0,0000990412804056731
	F	her job	0,0000990412804056731
	G	reverse prostitution	0,0000990412804056731
		i will	0,0000990412804056731
		find a	0,0000982327801574635
	F	she will	0,0000978760888714887
		when your	0,0000977550300107942
		so you	0,0000969258355814743
		that you	0,0000949252791420608
	M	man s	0,0000924385283786282
		for a	0,0000909562779235773
		what the	0,0000905984171579764
		you give	0,0000891371523651058
		and i	0,0000878198545469024
		i like	0,0000867790266411612
		have be	0,0000858357763515833
	F	woman to	0,0000848925260620055
	F	that woman	0,0000848925260620055
	F	girl s	0,0000848925260620055
		make a	0,0000841850883448222
	F	she say	0,0000840350257987529
		but you	0,0000826046423809017
		be to	0,0000824451739593170
		what i	0,0000822804483370207
		i think	0,0000820131304411888
		of a	0,0000819651975771088

		instead of	0,0000819651975771088
	F	a girlfriend	0,0000817090563346804
		with i	0,0000816340250616458
		be in	0,0000805759569402087
		cheat on	0,0000805759569402087
		what be	0,0000803813290248940
		be go	0,0000802112098100264
		be it	0,0000796544765815839
		i can	0,0000792330243245385
	F	the wife	0,0000792330243245385
		a strong	0,0000792330243245385
	S	get lay	0,0000792330243245385
	F	lose her	0,0000792330243245385
	M	man hire	0,0000792330243245385
	F	strong woman	0,0000792330243245385
	S	s nipple	0,0000792330243245385
	W	prostitute but	0,0000792330243245385
		you because	0,0000792330243245385
F	F	woman woman	0,0000792330243245385
	W	a hooker	0,0000792330243245384
		when they	0,0000778429712662133
	G	the house	0,0000773687178698434
	M	a male	0,0000772521987164250
		what you	0,0000770321069821901
		i would	0,0000763622625736495
		do what	0,0000756315232188776
	M	man who	0,0000756315232188776
		should be	0,0000752713731083116
		change my	0,0000742809603042548
		with you	0,0000739508227029026
		my mind	0,0000731381762995740
		how to	0,0000724331752220595
		on a	0,0000720300221132168
	F	girl be	0,0000720300221132168
	W	prostitute i	0,0000720300221132168
		and then	0,0000715653122931315
		what do	0,0000715653122931315
		will never	0,0000713097218920846
		i want	0,0000712370310440804
	F	a female	0,0000706882864071863
	F	woman who	0,0000705669122890420
	F	your wife	0,0000701641841910070
		do it	0,0000701257801493042
		be your	0,0000700153209996540
		we be	0,0000697880280474412
	F	a milf	0,0000693288962839712
		talk about	0,0000688076263870992
		from the	0,0000685132622100421
	F	your girl	0,0000679140208496044
		give you	0,0000675281457311407
		call a	0,0000671758249708043
		i see	0,0000663995063001414
	S	a dick	0,0000660275202704487
	F	she a	0,0000660275202704487
		s day	0,0000660275202704487
	F	woman can	0,0000660275202704487
		their life	0,0000660275202704487
		we have	0,0000645183198071242

	F	wife be	0,0000639959042621273
		you for	0,0000639959042621273
		when it	0,0000637729220173115
		be like	0,0000633002966071041
		what if	0,0000631388162586166
		here be	0,0000627261442569263
		i the	0,0000627261442569263
		make you	0,0000619008002535457
	F	if she	0,0000613112688225595
	M	man in	0,0000613112688225595
		you want	0,0000611226187646440
		you get	0,0000606891250145402
		you know	0,0000605459902857322
		year old	0,0000604493332131178
		to i	0,0000603389954471484
		with the	0,0000594247682434039
		i get	0,0000594247682434039
		why i	0,0000594247682434039
		say that	0,0000594247682434039
		be be	0,0000594247682434038
	F	for woman	0,0000594247682434038
	M	think man	0,0000594247682434038
		will give	0,0000594247682434038
		like when	0,0000594247682434038
	F	girl in	0,0000594247682434038
	F	girl on	0,0000594247682434038
	F	woman do	0,0000594247682434038
	F	treat woman	0,0000594247682434038
G	G	feminist feminist	0,0000594247682434038
	A	imgflip.com this	0,0000594247682434038
		not spell	0,0000594247682434038
F	A	imgflip.com girl	0,0000594247682434038
	F	where woman	0,0000594247682434038
	P	be toxic	0,0000594247682434038
		a alamy	0,0000594247682434038
		alamy a	0,0000594247682434038
		must go	0,0000594247682434038
	F	give her	0,0000594247682434038
	F	prostitute from	0,0000594247682434038
S	S	bikini body	0,0000594247682434038
	F	bindelj bindelj	0,0000594247682434038
	F	bindelj dec	0,0000594247682434038
	F	julie bindel	0,0000594247682434038
		york post	0,0000594247682434038
		lay by	0,0000594247682434038
		be my	0,0000580427968889061
		one of	0,0000570237675062967
	W	the hooker	0,0000570237675062967
		more than	0,0000570237675062967
		not the	0,0000570237675062967
		to you	0,0000558694402288412
	G	a sandwich	0,0000554631170271769
	M	to man	0,0000554631170271769
	F	say no	0,0000554631170271769
	G	right to	0,0000554631170271769
	F	give she	0,0000554631170271769
		make i	0,0000552788541799105
		think you	0,0000552788541799105

		i find	0,0000551801419403036
		be no	0,0000547887934159043
		let us	0,0000538645560101029
		i love	0,0000535797090719216
		be also	0,0000528220162163590
		without a	0,0000528220162163590
	F	that she	0,0000528220162163590
	F	wife i	0,0000528220162163590
		to a	0,0000527362661900336
	S	have sex	0,0000526548579371933
		the same	0,0000525774698449869
		i a	0,0000524336190382975
	F	because she	0,0000524336190382975
		not have	0,0000522415544996957
	P	be fuck	0,0000521269896871964
	M	guy who	0,0000516737115160034
		be so	0,0000516369068781715
		go out	0,0000513547379881268
	F	say she	0,0000512282484856930
		not leave	0,0000511180802093796
		want a	0,0000509355156372033
		by the	0,0000509355156372033
		must be	0,0000509355156372033
		you go	0,0000505312655130985
		to see	0,0000504916331479901
		be too	0,0000501990051371220
		be an	0,0000501809154055411
		and a	0,0000501474837497079
		and the	0,0000500770518905088
		not know	0,0000500419100997086
	M	real man	0,0000495206402028366
		time for	0,0000495206402028366
		be still	0,0000488231663971629
	F	but she	0,0000484201815316624
		you need	0,0000484201815316624
		ask for	0,0000483689974074217
		after a	0,0000482508801976356
		but then	0,0000478130319199801
		get out	0,0000476865424175463
	F	woman like	0,0000475398145947231
	G	feminism be	0,0000475398145947231
	F	be woman	0,0000475398145947231
	M	all man	0,0000475398145947231
	F	woman that	0,0000475398145947231
	F	this girl	0,0000475398145947231
	M	than man	0,0000475398145947231
	F	all woman	0,0000475398145947231
	F	she like	0,0000475398145947231
	F	so she	0,0000475398145947231
		fight for	0,0000475398145947231
		be literally	0,0000475398145947231
	F	when girl	0,0000475398145947231
	F	she that	0,0000475398145947231
		us see	0,0000475398145947231
		the secret	0,0000475398145947231
	S	dick be	0,0000475398145947231
		modern solution	0,0000475398145947231
		modern problem	0,0000475398145947231

		be tall	0,0000475398145947231
		require modern	0,0000475398145947231
	F	and woman	0,0000475398145947231
	M	with his	0,0000469142907184767
		you must	0,0000468195143735909
		be only	0,0000457113601872338
	M	a guy	0,0000452760138997363
		like to	0,0000448987137839052
		the other	0,0000445685761825528
		talk to	0,0000442772782990068
		from a	0,0000442772782990068
		they say	0,0000440183468469658
		can i	0,0000440183468469658
		like you	0,0000437866713372449
		how do	0,0000437866713372449
		you a	0,0000433895133205806
		to eat	0,0000433895133205806
		have the	0,0000432180132679301
		would be	0,0000429178881757917
		be on	0,0000427858331352508
		but it	0,0000427858331352508
		be this	0,0000427858331352508
		we will	0,0000426639361747515
	S	sex with	0,0000425510686187336
	M	man do	0,0000424462630310027
	M	man want	0,0000424462630310027
		realize that	0,0000424462630310027
	F	on woman	0,0000424462630310027
	M	how man	0,0000424462630310027
		can we	0,0000424462630310027
		someone be	0,0000424462630310027
		you what	0,0000424462630310027
	M	man when	0,0000424462630310027
	F	with woman	0,0000424462630310027
		problem require	0,0000424462630310027
		the only	0,0000424462630310027
		be that	0,0000423486854148395
		need a	0,0000421724161727382
		tell you	0,0000420925441724111
		no one	0,0000420175128993765
		you see	0,0000418803128572561
		to have	0,0000418803128572561
		if i	0,0000417015917497570
		the world	0,0000414591406349329
	F	your girlfriend	0,0000411701008745151
		but i	0,0000409825987885543
	A	imgflip.com i	0,0000396165121622693
		to my	0,0000396165121622693
	P	the fuck	0,0000396165121622693
		for my	0,0000396165121622692
		you think	0,0000396165121622692
		like it	0,0000396165121622692
	P	fuck your	0,0000396165121622692
		and your	0,0000396165121622692
		to your	0,0000396165121622692
		do this	0,0000396165121622692
	P	fuck a	0,0000396165121622692
		and be	0,0000396165121622692

	F	wife say	0,0000396165121622692
	F	on her	0,0000396165121622692
	F	to she	0,0000396165121622692
		be ready	0,0000396165121622692
		i wish	0,0000396165121622692
		you you	0,0000396165121622692
		the middle	0,0000396165121622692
		my business	0,0000396165121622692
		someone who	0,0000396165121622692
		job for	0,0000396165121622692
		sit down	0,0000396165121622692
	M	for man	0,0000396165121622692
	G	be equal	0,0000396165121622692
	G	equal to	0,0000396165121622692
	P	to rape	0,0000396165121622692
	F	woman make	0,0000396165121622692
	F	respect woman	0,0000396165121622692
	F	understand woman	0,0000396165121622692
	S	be cute	0,0000396165121622692
	F	woman a	0,0000396165121622692
	G	right for	0,0000396165121622692
	F	woman as	0,0000396165121622692
	M	man the	0,0000396165121622692
	F	woman never	0,0000396165121622692
	P	sexual assault	0,0000396165121622692
	F	woman want	0,0000396165121622692
	F	woman love	0,0000396165121622692
		money back	0,0000396165121622692
		let they	0,0000396165121622692
		be pig	0,0000396165121622692
	F	think woman	0,0000396165121622692
		only a	0,0000396165121622692
	G	about feminism	0,0000396165121622692
		s chair	0,0000396165121622692
	F	she see	0,0000396165121622692
	F	some woman	0,0000396165121622692
		good driver	0,0000396165121622692
		to dress	0,0000396165121622692
		be oppress	0,0000396165121622692
	A	mematic what	0,0000396165121622692
		safe space	0,0000396165121622692
		to share	0,0000396165121622692
	M	be male	0,0000396165121622692
F	F	woman she	0,0000396165121622692
		be sometimes	0,0000396165121622692
	G	manspreade chair	0,0000396165121622692
	S	before sex	0,0000396165121622692
		first game	0,0000396165121622692
	M	boy s	0,0000396165121622692
		game where	0,0000396165121622692
	G	the manspreade	0,0000396165121622692
	F	she must	0,0000396165121622692
		s march	0,0000396165121622692
		award for	0,0000396165121622692
	S	a barbie	0,0000396165121622692
		the snack	0,0000396165121622692
	F	strong female	0,0000396165121622692
	W	hooker you	0,0000396165121622692

		you order	0,0000396165121622692
	A	image istock	0,0000396165121622692
		world look	0,0000396165121622692
	F	like woman	0,0000396165121622692
		well it	0,0000396165121622692
	P	and kill	0,0000396165121622692
	M	my man	0,0000396165121622692
	G	feminist will	0,0000396165121622692
		no to	0,0000396165121622692
		bleach hair	0,0000396165121622692
F	S	anime girl	0,0000396165121622692
		from date	0,0000396165121622692
		bring balance	0,0000396165121622692
	M	or man	0,0000396165121622692
	G	promote sexism	0,0000396165121622692
	F	random woman	0,0000396165121622692
M	P	rape his	0,0000396165121622692
		can never	0,0000396165121622692
	F	she or	0,0000396165121622692
		sjw sjw	0,0000396165121622692
		hey how	0,0000396165121622692
		balance to	0,0000396165121622692
	W	hire prostitute	0,0000396165121622692
F	S	straight woman	0,0000396165121622692
		create chair	0,0000396165121622692
		their clothing	0,0000396165121622692
	G	their right	0,0000396165121622692
	F	daughter turn	0,0000396165121622692
		day you	0,0000396165121622692
	P	torture and	0,0000396165121622692
	F	you her	0,0000396165121622692
	F	lady in	0,0000396165121622692
F	F	woman girl	0,0000396165121622692
		be replace	0,0000396165121622692
		wong ho	0,0000396165121622692
		would ask	0,0000396165121622692
		during yoga	0,0000396165121622692
		looter in	0,0000396165121622692
	M	love man	0,0000396165121622692
		empower some	0,0000396165121622692
		a yacht	0,0000396165121622692
		zero zero	0,0000396165121622692
S	M	male nipple	0,0000396165121622692
		be unable	0,0000396165121622692
	M	man describe	0,0000396165121622692
		maximum distance	0,0000396165121622692
		bear without	0,0000396165121622692
	F	female team	0,0000396165121622692
		my people	0,0000396165121622692
		non alcoholic	0,0000396165121622692
	S	blowjob at	0,0000396165121622692
		for march	0,0000396165121622692
P	P	fuck stupid	0,0000396165121622692
	S	gender assume	0,0000396165121622692
	A	around imgflip.com	0,0000396165121622692
		people around	0,0000396165121622692
		by multiple	0,0000396165121622692
	F	pregnant girl	0,0000396165121622692

		give to	0,0000396165121622692
		pretty sure	0,0000396165121622692
	M	problem sir	0,0000396165121622692
		pumpkin pie	0,0000396165121622692
	F	rabbit girlfriend	0,0000396165121622692
		age range	0,0000396165121622692
		role model	0,0000396165121622692
		s dodgeball	0,0000396165121622692
		s pocket	0,0000396165121622692
		sale like	0,0000396165121622692
		at what	0,0000396165121622692
	M	sanni gbe	0,0000396165121622692
	F	scare she	0,0000396165121622692
		have equal	0,0000396165121622692
	S	sex class	0,0000396165121622692
	M	shahmir sanni	0,0000396165121622692
		have win	0,0000396165121622692
	F	she bang	0,0000396165121622692
	M	sir shahmir	0,0000396165121622692
		sjws have	0,0000396165121622692
		claim it	0,0000396165121622692
		clean shave	0,0000396165121622692
		speak the	0,0000396165121622692
		spend their	0,0000396165121622692
		all must	0,0000396165121622692
	F	sure she	0,0000396165121622692
		team at	0,0000396165121622692
		the backlash	0,0000396165121622692
		the cage	0,0000396165121622692
		could this	0,0000396165121622692
		the sentence	0,0000396165121622692
		the type	0,0000396165121622692
		their time	0,0000396165121622692
		cut of	0,0000396165121622692
		they pose	0,0000396165121622692
		be fight	0,0000396165121622692
		in darkness	0,0000396165121622692
	M	dave j	0,0000396165121622692
		defend a	0,0000396165121622692
		islamic revolution	0,0000396165121622692
	F	trixie mattel	0,0000396165121622692
	F	describe woman	0,0000396165121622692
		trust fund	0,0000396165121622692
		uk news	0,0000396165121622692
	M	underpay man	0,0000396165121622692
		it underpay	0,0000396165121622692
		j hogan	0,0000396165121622692
	F	different woman	0,0000396165121622692
	S	be lesbian	0,0000396165121622692
	F	woman more	0,0000396165121622692
		watch what	0,0000396165121622692
		distance age	0,0000396165121622692
		diy electronic	0,0000396165121622692
		what cost	0,0000396165121622692
	F	what girl	0,0000396165121622692
	M	what guy	0,0000396165121622692
		who lose	0,0000396165121622692
M	F	his wife	0,0000396165121622692

		that the	0,0000396165121622692
		live in	0,0000396165121622692
		of you	0,0000396165121622692
		star war	0,0000396165121622692
		of they	0,0000396165121622692
		be all	0,0000396165121622692
		pick up	0,0000396165121622692
	F	she do	0,0000396165121622692
		be actually	0,0000396165121622692
		to tell	0,0000396165121622692
		so that	0,0000396165121622692
		and get	0,0000396165121622692
	F	she have	0,0000378157616094388
		if your	0,0000378157616094388
	F	what woman	0,0000378157616094388
		that a	0,0000378157616094388
	F	to woman	0,0000378157616094388
	M	what man	0,0000378157616094388
		a problem	0,0000378157616094388
		on tinder	0,0000378157616094388
		identify as	0,0000378157616094388
		with they	0,0000378157616094388
		i wanna	0,0000378157616094388
	F	the female	0,0000365690881497870
	F	a cougar	0,0000365690881497870
		ask you	0,0000365690881497870
		they see	0,0000365690881497870
		meet a	0,0000365690881497870
		but a	0,0000365690881497870
	S	non binary	0,0000365690881497870
		none of	0,0000365690881497870
	F	pregnant woman	0,0000365690881497870
		the truth	0,0000365690881497870
		like my	0,0000356548609460423
		a real	0,0000356548609460423
	F	with her	0,0000356548609460423
		picture of	0,0000356548609460423
	F	female character	0,0000356548609460423
		it in	0,0000356548609460423
		what a	0,0000349557460255317
		be where	0,0000349557460255317
		in their	0,0000349557460255317
		be watch	0,0000349557460255317
	M	he can	0,0000349557460255317
	F	she to	0,0000344038131935496
		the big	0,0000344038131935496
		not worry	0,0000344038131935496
		the word	0,0000344038131935496
		but that	0,0000344038131935496
		be sure	0,0000344038131935496
		be what	0,0000339570104248022
		it do	0,0000339570104248022
		but your	0,0000339570104248022
		hire a	0,0000339570104248022
	M	and his	0,0000339570104248022
		new york	0,0000339570104248022
		because it	0,0000335879124854022
		they can	0,0000335879124854022

	M	man and	0,0000335879124854022
		will make	0,0000335879124854022
		not need	0,0000332778702163062
		come to	0,0000332778702163062
		be really	0,0000332778702163062
	F	the woman	0,0000330137601352243
	F	what she	0,0000330137601352243
		you would	0,0000330137601352243
		the way	0,0000330137601352243
		be tell	0,0000330137601352243
		or a	0,0000330137601352243
		because they	0,0000325877761334795
		find the	0,0000325877761334795
		how be	0,0000325877761334795
		due to	0,0000325877761334795
		walk in	0,0000325877761334795
		at a	0,0000324135099509476
		you in	0,0000322591599035621
	F	your mom	0,0000322591599035621
		the most	0,0000321214963477859
		they have	0,0000319979521310636
		the good	0,0000318864610086558
		to do	0,0000318864610086558
		do the	0,0000317853411534486
		be get	0,0000316089192784063
		have no	0,0000312231833143309
		will be	0,0000312231833143309
		not get	0,0000309678369719146
		i i	0,0000309678369719146
		wait for	0,0000309008794865699
		at the	0,0000303357208515278
	F	she in	0,0000297123841217019
	F	girl when	0,0000297123841217019
	F	woman on	0,0000297123841217019
	F	she for	0,0000297123841217019
	F	white woman	0,0000297123841217019
	G	the feminist	0,0000297123841217019
	F	for her	0,0000297123841217019
		dress like	0,0000297123841217019
		the harpoon	0,0000297123841217019
		hit a	0,0000297123841217019
	F	how woman	0,0000297123841217019
	F	girl and	0,0000297123841217019
		if this	0,0000297123841217019
		do in	0,0000297123841217019
	F	she would	0,0000297123841217019
	F	her first	0,0000297123841217019
	M	by man	0,0000297123841217019
	F	i woman	0,0000297123841217019
	F	do she	0,0000297123841217019
G	F	woman right	0,0000297123841217019
		touch a	0,0000297123841217019
		make more	0,0000297123841217019
	S	the eye	0,0000297123841217019
	F	she back	0,0000297123841217019
	S	pussy but	0,0000297123841217019
F	F	she she	0,0000297123841217019
		capable of	0,0000297123841217019

		to impress	0,0000297123841217019
M	F	woman man	0,0000297123841217019
	M	boy in	0,0000297123841217019
		they with	0,0000297123841217019
		people will	0,0000297123841217019
		worried about	0,0000297123841217019
	F	she but	0,0000297123841217019
		in sport	0,0000297123841217019
		see who	0,0000297123841217019
		be give	0,0000297123841217019
		be cold	0,0000297123841217019
		turn up	0,0000297123841217019
		if someone	0,0000297123841217019
	F	wife when	0,0000297123841217019
	F	be female	0,0000297123841217019
		to bring	0,0000297123841217019
		it get	0,0000297123841217019
		you over	0,0000297123841217019
	M	man on	0,0000297123841217019
		ever have	0,0000297123841217019
		farm remember	0,0000297123841217019
		but at	0,0000297123841217019
		pepperidge farm	0,0000297123841217019
		put on	0,0000297123841217019
		here we	0,0000297123841217019
F	M	his daughter	0,0000297123841217019
		drink it	0,0000297123841217019
		also the	0,0000297123841217019
		could stop	0,0000297123841217019
		visible confusion	0,0000297123841217019
		yes no	0,0000297123841217019
		your privilege	0,0000297123841217019
		n a	0,0000297123841217019
		never touch	0,0000297123841217019
	M	of male	0,0000297123841217019
		of meat	0,0000297123841217019
		on not	0,0000297123841217019
	W	prostitute say	0,0000297123841217019
		a role	0,0000297123841217019
	W	cheap prostitute	0,0000297123841217019
		ten year	0,0000297123841217019
		a haircut	0,0000297123841217019
	F	her room	0,0000297123841217019
		stuff like	0,0000297123841217019
F	G	cook her	0,0000297123841217019
	S	the gender	0,0000297123841217019
		the pumpkin	0,0000297123841217019
		there will	0,0000297123841217019
	S	a trans	0,0000297123841217019
		be quite	0,0000297123841217019
	F	girl who	0,0000264110081081795
	F	michelle obama	0,0000264110081081795
	F	she ask	0,0000264110081081795
		all your	0,0000264110081081795
		a black	0,0000264110081081795
	F	woman when	0,0000264110081081795
S	F	her pussy	0,0000264110081081795
	S	be ugly	0,0000264110081081795

		really want	0,0000264110081081795
		you really	0,0000264110081081795
		who can	0,0000264110081081795
	M	a boy	0,0000264110081081795
	S	a sex	0,0000264110081081795
	F	about her	0,0000264110081081795
		the side	0,0000264110081081795
		these be	0,0000264110081081795
M	F	her boyfriend	0,0000264110081081795
		i pay	0,0000264110081081795
	F	woman go	0,0000264110081081795
		who think	0,0000264110081081795
		not change	0,0000264110081081795
	M	your husband	0,0000264110081081795
		your family	0,0000264110081081795
	M	man with	0,0000264110081081795
	M	man you	0,0000264110081081795
		they and	0,0000264110081081795
		become an	0,0000264110081081795
	P	shit out	0,0000264110081081795
	F	another girl	0,0000264110081081795
		who make	0,0000264110081081795
		president of	0,0000264110081081795
		one will	0,0000264110081081795
		remember when	0,0000264110081081795
		as long	0,0000264110081081795
	M	man really	0,0000264110081081795
		so if	0,0000264110081081795
		i by	0,0000264110081081795
		be both	0,0000264110081081795
		in real	0,0000264110081081795
		leave it	0,0000264110081081795
	G	feminism and	0,0000264110081081795
		pay to	0,0000264110081081795
		a pregnant	0,0000264110081081795
M	S	straight man	0,0000264110081081795
	S	not gay	0,0000264110081081795
		carry on	0,0000264110081081795
		change be	0,0000264110081081795
		the wood	0,0000264110081081795
		unable to	0,0000264110081081795
		to cover	0,0000264110081081795
		invite you	0,0000264110081081795
		it feel	0,0000264110081081795
	F	woman stand	0,0000264110081081795
		you ok	0,0000264110081081795
	F	girl with	0,0000247603201014183
	F	she make	0,0000247603201014183
	S	your dick	0,0000247603201014183
		hard to	0,0000247603201014183
	F	she start	0,0000247603201014183
	F	woman think	0,0000247603201014183
	P	kill a	0,0000247603201014183
		no matter	0,0000247603201014183
		real life	0,0000247603201014183
		getty image	0,0000247603201014183
		put it	0,0000247603201014183
		up for	0,0000247603201014183

		on facebook	0,0000247603201014183
		matter what	0,0000247603201014183
		before after	0,0000247603201014183
		not handle	0,0000247603201014183
		your mind	0,0000247603201014183
		that will	0,0000247603201014183
		wish i	0,0000247603201014183
		as they	0,0000247603201014183
		not wanna	0,0000247603201014183
		over there	0,0000247603201014183
		be none	0,0000247603201014183
		base on	0,0000247603201014183
	F	have her	0,0000247603201014183
		to they	0,0000247603201014183
	P	fuck the	0,0000247603201014183
		the president	0,0000247603201014183
		could have	0,0000247603201014183
		date site	0,0000247603201014183
	F	all female	0,0000247603201014183
		not live	0,0000247603201014183
		of culture	0,0000247603201014183
		or be	0,0000247603201014183
	F	she just	0,0000247603201014183
	F	eat her	0,0000247603201014183
	F	my mother	0,0000247603201014183
	F	her cat	0,0000247603201014183
		war fan	0,0000247603201014183
	S	two gender	0,0000247603201014183
		type of	0,0000237699072973615
		not find	0,0000237699072973615
		it when	0,0000237699072973615
	F	off her	0,0000237699072973615
		of be	0,0000237699072973615
		we use	0,0000237699072973615
		not matter	0,0000237699072973615
		your phone	0,0000237699072973615
		one who	0,0000237699072973615
		keep the	0,0000237699072973615
	F	her clothe	0,0000237699072973615
	F	take she	0,0000237699072973615
	S	a beautiful	0,0000237699072973615
		over to	0,0000237699072973615
		i or	0,0000237699072973615
		be leave	0,0000237699072973615
	G	cook and	0,0000237699072973615
		pay you	0,0000237699072973615
	G	of cook	0,0000237699072973615
		a tiger	0,0000237699072973615
		kind of	0,0000231096320946571
		be one	0,0000231096320946571
		in front	0,0000231096320946571
		front of	0,0000231096320946571
	M	guy be	0,0000231096320946571
	F	ask she	0,0000231096320946571
		for iphone	0,0000231096320946571
		do a	0,0000231096320946571
		twitter for	0,0000231096320946571
		show i	0,0000231096320946571

	P	a fucking	0,0000231096320946571
		not give	0,0000231096320946571
		into your	0,0000231096320946571
		of their	0,0000231096320946571
	F	for she	0,0000231096320946571
		what it	0,0000231096320946571
		hope you	0,0000231096320946571
		see it	0,0000231096320946571
		room and	0,0000231096320946571
		have any	0,0000231096320946571
	M	his own	0,0000231096320946571
		just the	0,0000231096320946571
	F	and her	0,0000231096320946571
		than the	0,0000231096320946571
	F	on she	0,0000231096320946571
	S	gender be	0,0000231096320946571
		deal with	0,0000231096320946571
	G	home and	0,0000231096320946571
		my gear	0,0000231096320946571
	W	prostitute to	0,0000231096320946571
		my money	0,0000231096320946571
		a clean	0,0000231096320946571
		a bad	0,0000226380069498681
		how much	0,0000226380069498681
		if we	0,0000226380069498681
		the internet	0,0000226380069498681
		be time	0,0000226380069498681
		up a	0,0000226380069498681
		out to	0,0000226380069498681
		life be	0,0000226380069498681
		never be	0,0000226380069498681
		to come	0,0000226380069498681
	M	he will	0,0000226380069498681
		people to	0,0000226380069498681
		away from	0,0000226380069498681
		what your	0,0000226380069498681
	P	you fuck	0,0000226380069498681
	M	white man	0,0000226380069498681
		really do	0,0000226380069498681
		be funny	0,0000226380069498681
		the sun	0,0000226380069498681
	M	guy with	0,0000226380069498681
		be hard	0,0000226380069498681
	F	amber hear	0,0000226380069498681
W	M	male prostitute	0,0000226380069498681
		be always	0,0000222842880912764
		know how	0,0000222842880912764
		but they	0,0000222842880912764
		the left	0,0000222842880912764
		to understand	0,0000222842880912764
		a day	0,0000222842880912764
		that make	0,0000222842880912764
		get you	0,0000222842880912764
		and this	0,0000222842880912764
		you may	0,0000222842880912764
		be ok	0,0000222842880912764
		that it	0,0000222842880912764
	G	to cook	0,0000222842880912764

		you ever	0,0000222842880912764
		that can	0,0000222842880912764
		the water	0,0000222842880912764
		the difference	0,0000220091734234829
		difference between	0,0000220091734234829
	F	woman and	0,0000220091734234829
		you make	0,0000220091734234829
		out the	0,0000220091734234829
		be how	0,0000220091734234829
		use the	0,0000220091734234829
		ready for	0,0000220091734234829
		be with	0,0000220091734234829
		to talk	0,0000220091734234829
		time in	0,0000220091734234829
	M	but he	0,0000220091734234829
		laugh at	0,0000220091734234829
	M	to his	0,0000220091734234829
		before you	0,0000220091734234829
	M	male and	0,0000220091734234829
		never have	0,0000220091734234829
		take off	0,0000220091734234829
		give a	0,0000220091734234829
		social distancing	0,0000220091734234829
		they will	0,0000217890816892481
	G	the right	0,0000217890816892481
		this picture	0,0000217890816892481
		you look	0,0000217890816892481
		a joke	0,0000217890816892481
	M	the man	0,0000217890816892481
		not think	0,0000217890816892481
	F	female friend	0,0000217890816892481
		i look	0,0000217890816892481
		not i	0,0000217890816892481
	M	trump be	0,0000217890816892481
		a white	0,0000217890816892481
		b b	0,0000217890816892481
		only one	0,0000216090066339650
		i hope	0,0000216090066339650
		not believe	0,0000216090066339650
		who say	0,0000216090066339650
		of all	0,0000216090066339650
		return to	0,0000216090066339650
		pay for	0,0000214589440878958
	M	when he	0,0000214589440878958
		you mean	0,0000214589440878958
		not care	0,0000214589440878958
		be bear	0,0000214589440878958
		meanwhile in	0,0000214589440878958
	F	a witch	0,0000214589440878958
		to stop	0,0000213319680873758
		a big	0,0000213319680873758
		think that	0,0000213319680873758
		time i	0,0000213319680873758
		be probably	0,0000213319680873758
		a baby	0,0000213319680873758
		first time	0,0000212231315155014
		be call	0,0000212231315155014
		not just	0,0000212231315155014

		the real	0,0000212231315155014
		will get	0,0000212231315155014
		on your	0,0000211288064865436
		become a	0,0000211288064865436
		have you	0,0000211288064865436
		you hear	0,0000211288064865436
		for your	0,0000210462720862055
		say you	0,0000210462720862055
		i at	0,0000210462720862055
		x81 x81	0,0000210462720862055
		a lot	0,0000209734476153190
		well than	0,0000209734476153190
	A	imgflip.com when	0,0000209734476153190
		not like	0,0000209734476153190
		know you	0,0000209734476153190
		to find	0,0000209087147523088
		a nice	0,0000209087147523088
		you like	0,0000208507958748785
		at my	0,0000208507958748785
		for you	0,0000208507958748785
		get the	0,0000207986688851913
		like this	0,0000207515063707125
		not want	0,0000207515063707125
		we do	0,0000207515063707125
		so much	0,0000207515063707125
		where be	0,0000206694846064013
		i make	0,0000206694846064013
	F	and she	0,0000206336000845152
		because of	0,0000206336000845152
		and it	0,0000206005863243800
		of your	0,0000205701120842552
		know what	0,0000205418951952507
		now i	0,0000205418951952507
		to call	0,0000205418951952507
		i need	0,0000205156937983180
		a good	0,0000204912993942772
		you to	0,0000203908518482268
		be do	0,0000203908518482268
		i in	0,0000203908518482268
		think i	0,0000203584854167217
		so i	0,0000203436143535977
		to cheat	0,0000203436143535977
	G	at home	0,0000203034624831630
		all the	0,0000202297083381801
		tell i	0,0000201277440824434

Tabella 27. Etichettatura dei bigrammi FN, ordinati in modo decrescente a seconda della misura dell'impatto.

Etichette riferite alla seconda parola del bigramma	Etichette riferite alla prima parola del bigramma	BIGRAMMI FN	impatto +
		be a	0,0008966820295679320
		i be	0,0008710118826927810
		do not	0,0007454106743786920
		it be	0,0007429902075953970

		you be	0,0006772297697651790
		when you	0,0004910703944777740
		in the	0,0004561739730383430
		be not	0,0004166906088865920
	F	she be	0,0003142101605100240
	F	a woman	0,0003056721467838090
	F	my wife	0,0003020832761205920
		i have	0,0002525256703408700
		be the	0,0002517271760176410
		can not	0,0002406450832477570
		you have	0,0002389013334751480
		if you	0,0002285991165137800
		have a	0,0002239157459096770
		this be	0,0002205005853288260
		of the	0,0002065429725201330
		to the	0,0002048476584019260
		for the	0,0001948945612436190
		that be	0,0001945192458234880
		do you	0,0001871626786500200
		there be	0,0001840291774128570
		make with	0,0001820585375091380
	A	with mematic	0,0001816760615900020
	M	he be	0,0001812811136300230
		be you	0,0001778971995089300
		in a	0,0001755564468838120
		to be	0,0001696109629500490
		you do	0,0001644759610844390
		who be	0,0001590759973018480
		they be	0,0001550185933220840
		out of	0,0001522946571466530
	F	woman in	0,0001502815519225590
	W	a prostitute	0,0001459312964721690
		look like	0,0001433943200406800
		when the	0,0001428030032570070
		when i	0,0001416044925336410
		go to	0,0001411040525685100
	F	a girl	0,0001403754922503500
	G	the kitchen	0,0001389821871163510
		i do	0,0001368922294153530
		to get	0,0001353247611396050
		the first	0,0001331912626531700
		on the	0,0001328506200376890
		try to	0,0001264127386691360
		with a	0,0001246105497624710
		you can	0,0001241993246676270
		as a	0,0001239939120351760
		and you	0,0001235397218812010
		in my	0,0001234601215449380
		need to	0,0001225623506637870
		get a	0,0001223043246623890
		not a	0,0001204003196081210
		when a	0,0001193660826284580
		when your	0,0001183362825084870
		dev do	0,0001158184892636260
		in your	0,0001158184892636260
		i will	0,0001158184892636260
		want to	0,0001146830138786880
		be just	0,0001139504491142130

		why be	0,0001126013090063030
		like a	0,0001120211617467860
	S	the dick	0,0001111857496930810
		of my	0,0001103033231082150
		will not	0,0001099294135383570
		have be	0,0001097227793023830
		and i	0,0001089039525911710
		you will	0,0001083463286659730
		see a	0,0001061669484916570
	G	the house	0,0001058911901838870
		what the	0,0001010779542664370
		for a	0,0001010128267185850
	F	my girlfriend	0,0001006422734290820
	W	the prostitute	0,0000965154077196884
		that you	0,0000950930753953983
		find a	0,0000950930753953983
	W	a hooker	0,0000926547914109009
		i can	0,0000926547914109009
		what be	0,0000915647350413609
		be go	0,0000910843712174958
		i like	0,0000907041642233030
		be in	0,0000901506078592550
		be dig	0,0000900810472050425
		instead of	0,0000892231324697564
		what i	0,0000890911455874047
		cheat on	0,0000886263222191225
		but you	0,0000879032636462392
		make a	0,0000876464243076089
		i see	0,0000876464243076089
		of a	0,0000875073029991842
		i think	0,0000872045095632008
		with i	0,0000855274997639084
		from the	0,0000826026206446238
		we be	0,0000802011904148119
		do it	0,0000801820310286641
	F	your wife	0,0000798979723036029
	F	a female	0,0000798256664463146
		be your	0,0000793479862614629
		and then	0,0000792713215404374
		will never	0,0000788817818768480
		what do	0,0000786648352820271
		how to	0,0000786161866516735
		should be	0,0000774428107314993
		on a	0,0000773311143698673
		i want	0,0000772123261757507
		remember why	0,0000772123261757507
	F	a girlfriend	0,0000772123261757507
		with you	0,0000756680796522357
		i find	0,0000738342869055616
		be my	0,0000721638279257978
	F	girl in	0,0000694910935581757
		be be	0,0000694910935581757
		say that	0,0000694910935581756
		why i	0,0000694910935581756
		have an	0,0000694910935581756
		dig in	0,0000694910935581756
		my birthday	0,0000694910935581756
		finally some	0,0000694910935581756

		do dev	0,0000694910935581756
		i get	0,0000694910935581756
		with the	0,0000694910935581756
		you know	0,0000678750216149623
		you want	0,0000677961888372446
		you get	0,0000677092706464275
		make you	0,0000672494453788797
	F	the wife	0,0000667659526343256
		year old	0,0000663324074873495
	F	your girl	0,0000663324074873495
	F	wife be	0,0000656304772493881
		to i	0,0000654033821724006
		what if	0,0000654033821724006
		but then	0,0000643436051464589
		here be	0,0000631737214165233
		you for	0,0000631737214165233
	F	a milf	0,0000631737214165233
		get out	0,0000623638019111833
		i the	0,0000617698609406006
		after a	0,0000609570996124348
		ask for	0,0000606668277095184
		with his	0,0000602256144170855
		x x	0,0000595637944784362
		be still	0,0000593216652325890
		you need	0,0000593216652325890
		and a	0,0000573779671581267
		and the	0,0000571571765197116
		not know	0,0000570449275477560
		be an	0,0000569277320109348
	F	say she	0,0000568932929716057
		you go	0,0000566223725288838
		be too	0,0000565625180124685
		want a	0,0000564968240310371
		to see	0,0000563441299120343
		go out	0,0000560412044823997
	F	because she	0,0000559123741272677
	F	that she	0,0000559123741272677
	M	guy who	0,0000557644577935977
		some good	0,0000555928748465405
		i remember	0,0000555928748465405
	M	male tear	0,0000555928748465405
	F	wife i	0,0000551516615541076
		by the	0,0000545028184770005
	P	be fuck	0,0000540486283230255
W	P	dead hooker	0,0000540486283230255
		not leave	0,0000514748841171671
		why do	0,0000486437654907229
		for my	0,0000463273957054504
		you think	0,0000463273957054504
F	M	his wife	0,0000463273957054504
		to my	0,0000463273957054504
		to tell	0,0000463273957054504
		and your	0,0000463273957054504
		do this	0,0000463273957054504
		and be	0,0000463273957054504
		live in	0,0000463273957054504
		so that	0,0000463273957054504
	P	fuck a	0,0000463273957054504

		to your	0,0000463273957054504
		and get	0,0000463273957054504
		be actually	0,0000463273957054504
		that the	0,0000463273957054504
		like it	0,0000463273957054504
		of you	0,0000463273957054504
	P	fuck your	0,0000463273957054504
		sit down	0,0000463273957054504
		star war	0,0000463273957054504
	F	wife say	0,0000463273957054504
		hand with	0,0000463273957054504
		go outside	0,0000463273957054504
	P	you idiot	0,0000463273957054504
		good news	0,0000463273957054504
		i choose	0,0000463273957054504
		the garden	0,0000463273957054504
		your age	0,0000463273957054504
		mematic excuse	0,0000463273957054504
	F	girlfriend want	0,0000463273957054504
		sell my	0,0000463273957054504
		cheat the	0,0000463273957054504
		complete saga	0,0000463273957054504
		to order	0,0000463273957054504
	S	an orgasm	0,0000463273957054504
		lego star	0,0000463273957054504
	F	maker she	0,0000463273957054504
		marriage i	0,0000463273957054504
		excited i	0,0000463273957054504
		mint flavor	0,0000463273957054504
		miss my	0,0000463273957054504
	F	mom during	0,0000463273957054504
		nuclear arm	0,0000463273957054504
		rainbow maker	0,0000463273957054504
		response time	0,0000463273957054504
		shake hand	0,0000463273957054504
		some gold	0,0000463273957054504
	M	tag he	0,0000463273957054504
		a some	0,0000463273957054504
		cover your	0,0000463273957054504
	S	crazy ex	0,0000463273957054504
		think drug	0,0000463273957054504
	A	imgflip.com new	0,0000463273957054504
		drink because	0,0000463273957054504
	A	imgflip.com i	0,0000463273957054504
	P	the fuck	0,0000463273957054504
		pick up	0,0000463273957054504
		of they	0,0000463273957054504
	F	she do	0,0000463273957054504
		be all	0,0000463273957054504
	M	say he	0,0000442216049915663
		your room	0,0000442216049915663
		finally get	0,0000427637498819542
M	F	her husband	0,0000427637498819542
		be suppose	0,0000408771138577504
		leave you	0,0000408771138577504
		be at	0,0000402316857442069
		think about	0,0000392775746198384
		suppose to	0,0000392775746198384

		be try	0,0000392775746198384
		and that	0,0000389150123925784
		then i	0,0000389150123925784
		with my	0,0000386061630878753
	F	girlfriend be	0,0000386061630878753
		not go	0,0000379042328499140
	S	a virgin	0,0000379042328499140
		you just	0,0000377237365030096
		be never	0,0000377237365030096
		find out	0,0000368728251533177
	G	house to	0,0000347455467790878
		in bed	0,0000347455467790878
	A	imgflip.com you	0,0000347455467790878
	S	my virginity	0,0000347455467790878
		watch a	0,0000347455467790878
		year in	0,0000347455467790878
	S	the relationship	0,0000347455467790878
		explain to	0,0000347455467790878
		to run	0,0000347455467790878
		i run	0,0000347455467790878
		about what	0,0000347455467790878
	F	out she	0,0000347455467790878
		and finally	0,0000347455467790878
		so excited	0,0000347455467790878
		i hire	0,0000347455467790878
		the west	0,0000347455467790878
		be direct	0,0000347455467790878
		wanna see	0,0000347455467790878
		totally worth	0,0000347455467790878
		up o	0,0000347455467790878
		my garden	0,0000347455467790878
		garden when	0,0000347455467790878
		by memri	0,0000347455467790878
		please tag	0,0000347455467790878
	W	prostitute about	0,0000347455467790878
		gold coin	0,0000347455467790878
		a slow	0,0000347455467790878
		the rainbow	0,0000347455467790878
		and my	0,0000308849304703003
		i when	0,0000308849304703003
		friend s	0,0000308849304703003
		your good	0,0000308849304703003
	A	memeful.com i	0,0000308849304703003
		and one	0,0000308849304703003
		i miss	0,0000308849304703003
		because my	0,0000308849304703003
		get so	0,0000308849304703003
		loyal to	0,0000308849304703003
		memri tv	0,0000308849304703003
		of fun	0,0000308849304703003
		run to	0,0000308849304703003
		meet i	0,0000308849304703003
		and tell	0,0000289546223159065
		to sleep	0,0000289546223159065
		be red	0,0000289546223159065
		i take	0,0000289546223159065
		the entire	0,0000289546223159065
	W	prostitute and	0,0000289546223159065

		my car	0,0000289546223159065
		turn out	0,0000289546223159065
	M	he would	0,0000289546223159065
		will take	0,0000289546223159065
	S	stay single	0,0000289546223159065
		my area	0,0000289546223159065
		the moon	0,0000289546223159065
		will cheat	0,0000289546223159065
		i try	0,0000277964374232702
		clean your	0,0000277964374232702
		good friend	0,0000277964374232702
		hold up	0,0000277964374232702
		for some	0,0000277964374232702
	F	my daughter	0,0000277964374232702
		take my	0,0000277964374232702
		my phone	0,0000277964374232702
		it but	0,0000277964374232702
		tell the	0,0000277964374232702
		the work	0,0000277964374232702
	P	a dead	0,0000277964374232702
		never leave	0,0000277964374232702
		marry a	0,0000277964374232702
	F	female fox	0,0000277964374232702
		my friend	0,0000270243141615128
		it to	0,0000270243141615128
		leave i	0,0000270243141615128
		be here	0,0000270243141615128
		want i	0,0000270243141615128
		it mean	0,0000270243141615128
		as you	0,0000270243141615128
		worth it	0,0000270243141615128
	M	and he	0,0000264727975459717
		come in	0,0000264727975459717
		be use	0,0000264727975459717
		about it	0,0000264727975459717
		have never	0,0000260591600843158
	M	of his	0,0000260591600843158
		see this	0,0000260591600843158
		i what	0,0000260591600843158
		if they	0,0000260591600843158
		rest of	0,0000260591600843158
		the rest	0,0000260591600843158
		video game	0,0000257374420585835
	F	wife and	0,0000257374420585835
		be think	0,0000257374420585835
		my face	0,0000257374420585835
		why you	0,0000257374420585835
	F	make she	0,0000257374420585835
		to speak	0,0000257374420585835
		about to	0,0000254800676379977
	F	my gf	0,0000254800676379977
		should not	0,0000254800676379977
		come back	0,0000254800676379977
		start to	0,0000254800676379977
		to say	0,0000254800676379977
	F	girl i	0,0000254800676379977
		to play	0,0000254800676379977
		the floor	0,0000252694885666093

	M	he a	0,0000252694885666093
		i my	0,0000250940060071190
		with your	0,0000249455207644733
		you should	0,0000249455207644733
		of this	0,0000248182476993484
		a few	0,0000247079443762402
		be about	0,0000246114289685205
		excuse i	0,0000246114289685205
		not even	0,0000245262683146502
		because i	0,0000243828398449739
		look for	0,0000243828398449739
		i to	0,0000242667310838074
		a new	0,0000240902457668342
		i say	0,0000240902457668342
		to make	0,0000236269718097797

Tabella 28. Etichettatura dei trigrammi FP ordinati in modo decrescente a seconda dell'impatto.

Etichette riferite alla terza parola del trigramma	Etichetta riferita alla seconda parola del trigramma	Etichetta riferita alla prima parola del trigramma	TRIGRAMMI FP	impatto -
		A	make with mematic	0,0003975642192203
			you do not	0,0002633512983831
			if you give	0,0002449779519843
			you be a	0,0002309792118709
			it be not	0,0002300662505592
			when you be	0,0002253081894162
		G	in the kitchen	0,0002230144114616
			i do not	0,0002213121573775
		F	woman s day	0,0001959823615875
		F	lose her job	0,0001959823615875
		F	she will give	0,0001959823615875
			why be you	0,0001861832435081
			i be a	0,0001841202712808
		G	change my mind	0,0001837334639882
			be not a	0,0001834486059045
		F	she be a	0,0001773978272990
			if you be	0,0001758816065528
		M	be a male	0,0001633186346562
			to be a	0,0001551527029234
			will give you	0,0001469867711906
		G	i a sandwich	0,0001469867711906
		F	woman do not	0,0001469867711906
			can not spell	0,0001469867711906
			a a a	0,0001469867711906
		M	man in the	0,0001469867711906
			alamy a a	0,0001469867711906
		F	a woman without	0,0001469867711906
		F	her job for	0,0001469867711906
	S	S	a bikini body	0,0001469867711906
		W	a prostitute but	0,0001469867711906
		F	bindel bindelj dec	0,0001469867711906
			new york post	0,0001469867711906
		F	julie bindel bindelj	0,0001469867711906
		S	get lay by	0,0001469867711906
		F	you give she	0,0001469867711906

			be go to	0,0001469867711906
			there be a	0,0001381926908630
			it be a	0,0001322103232931
			you can not	0,0001313649887452
			when it be	0,0001302414428271
			do not know	0,0001238968952564
			make i a	0,0001175894169525
		G	be a feminist	0,0001175894169525
			what the world	0,0001175894169525
			let us see	0,0001175894169525
			problem require modern	0,0001175894169525
			require modern solution	0,0001175894169525
			modern problem require	0,0001175894169525
			in their life	0,0001175894169525
			i be the	0,0001158077591199
		M	be a man	0,0001110566715662
			there be no	0,0001095195550048
			need to be	0,0001049905508504
			in the middle	0,0001049905508504
			do not have	0,0001039300402358
			i be not	0,0001016204837861
			if someone be	0,0000979911807937
			i can not	0,0000979911807937
			out of the	0,0000979911807937
			when you have	0,0000979911807937
			give you a	0,0000979911807937
			think you do	0,0000979911807937
			of my business	0,0000979911807937
			none of my	0,0000979911807937
		W	hire a prostitute	0,0000979911807937
		G	be equal to	0,0000979911807937
	G	M	equal to man	0,0000979911807937
		F	she in the	0,0000979911807937
		F	when she say	0,0000979911807937
		F	be a girl	0,0000979911807937
		F	a woman to	0,0000979911807937
		G	equal right for	0,0000979911807937
		M	more than man	0,0000979911807937
		M	man be pig	0,0000979911807937
		F	that woman be	0,0000979911807937
		A	with mematic what	0,0000979911807937
			you look like	0,0000979911807937
			dress like a	0,0000979911807937
		P	you be fuck	0,0000979911807937
		F	see a girl	0,0000979911807937
			the first game	0,0000979911807937
		F	this woman be	0,0000979911807937
			be only a	0,0000979911807937
		F	woman s march	0,0000979911807937
			look like when	0,0000979911807937
			the world look	0,0000979911807937
			to dress like	0,0000979911807937
		G	the manspreade chair	0,0000979911807937
			well it be	0,0000979911807937
			first game where	0,0000979911807937
			talk to you	0,0000979911807937
			world look like	0,0000979911807937
		F	game where woman	0,0000979911807937

		F	girl in a	0,0000979911807937
		W	prostitute from date	0,0000979911807937
		F	daughter turn up	0,0000979911807937
			bring balance to	0,0000979911807937
		G	talk about feminism	0,0000979911807937
			to come to	0,0000979911807937
		W	prostitute to come	0,0000979911807937
			say no to	0,0000979911807937
		F	she must be	0,0000979911807937
			so if you	0,0000979911807937
			hey how be	0,0000979911807937
		W	hire prostitute from	0,0000979911807937
	F	M	his daughter turn	0,0000979911807937
			here be where	0,0000979911807937
			tell you what	0,0000979911807937
		M	man hire a	0,0000979911807937
	W	M	man hire prostitute	0,0000979911807937
			a a alamy	0,0000979911807937
	F	M	and his daughter	0,0000979911807937
			from date site	0,0000979911807937
	S	F	woman s nipple	0,0000979911807937
		M	a man hire	0,0000979911807937
		W	prostitute but you	0,0000979911807937
		W	prostitute say that	0,0000979911807937
	G	S	blowjob at home	0,0000979911807937
		G	have equal right	0,0000979911807937
			but at what	0,0000979911807937
			at what cost	0,0000979911807937
	F	F	she that she	0,0000979911807937
			all must go	0,0000979911807937
			speak the truth	0,0000979911807937
			spend their time	0,0000979911807937
		S	call a trans	0,0000979911807937
			balance to the	0,0000979911807937
		M	a problem sir	0,0000979911807937
		S	i be lesbian	0,0000979911807937
			i can never	0,0000979911807937
			the type of	0,0000979911807937
			be a role	0,0000979911807937
		S	i get lay	0,0000979911807937
			i have win	0,0000979911807937
			to bring balance	0,0000979911807937
		S	a sex class	0,0000979911807937
			i would ask	0,0000979911807937
			be bear without	0,0000979911807937
		M	up a man	0,0000979911807937
		F	could this woman	0,0000979911807937
			in a yacht	0,0000979911807937
			wait for march	0,0000979911807937
			in the cage	0,0000979911807937
		M	dave j hogan	0,0000979911807937
		F	a strong female	0,0000979911807937
		W	prostitute i would	0,0000979911807937
			invite you over	0,0000979911807937
			when you order	0,0000979911807937
		F	scare she that	0,0000979911807937
		M	shahmir sannigbe	0,0000979911807937
			distance age range	0,0000979911807937

		F	who lose her	0,0000979911807937
			it in darkness	0,0000979911807937
			why i like	0,0000979911807937
		M	sir shahmir sannu	0,0000979911807937
		M	without a man	0,0000979911807937
		F	woman be one	0,0000979911807937
		F	kind of woman	0,0000979911807937
		F	woman in sport	0,0000979911807937
		P	and kill a	0,0000979911807937
		F	woman without a	0,0000979911807937
			lay by multiple	0,0000979911807937
			leave it in	0,0000979911807937
		M	be male and	0,0000979911807937
			you be to	0,0000979911807937
			like to dress	0,0000979911807937
			the pumpkin pie	0,0000979911807937
		F	you give her	0,0000979911807937
		F	the wife i	0,0000979911807937
		G	their right to	0,0000979911807937
		F	empower some woman	0,0000979911807937
	W	M	male prostitute i	0,0000979911807937
	F	M	man describe woman	0,0000979911807937
	S	M	man s nipple	0,0000979911807937
		M	man when it	0,0000979911807937
			maximum distance age	0,0000979911807937
	P	P	torture and kill	0,0000979911807937
			turn up a	0,0000979911807937
			us see who	0,0000979911807937
			use the word	0,0000979911807937
		F	female team at	0,0000979911807937
		S	want a bikini	0,0000979911807937
G	G	G	feminist feminist feminist	0,0000979911807937
			for the backlash	0,0000979911807937
			not leave it	0,0000979911807937
			be unable to	0,0000979911807937
			of a black	0,0000979911807937
		P	fuck the pumpkin	0,0000979911807937
			of my people	0,0000979911807937
			get a tiger	0,0000979911807937
		S	on a bikini	0,0000979911807937
			as a strong	0,0000979911807937
			you over to	0,0000979911807937
			over to eat	0,0000979911807937
			a alamy a	0,0000979911807937
			what do you	0,0000979911807937
			you do what	0,0000935370362122
			you will never	0,0000904533976557
		F	my wife be	0,0000904533976557
			when they see	0,0000904533976557
			but that be	0,0000904533976557
		F	when she be	0,0000881920627144
		F	be a woman	0,0000881920627144
			if i be	0,0000881920627144
			it do not	0,0000864628065827
			do not worry	0,0000850976043735
			be the first	0,0000839924406803
			do not need	0,0000830794793686
			but i be	0,0000810961496224

			but it be	0,0000794523087517
		S	have sex with	0,0000786208311019
		F	a woman be	0,0000734933855953
		F	a woman s	0,0000734933855953
		F	when a woman	0,0000734933855953
			when they be	0,0000734933855953
		F	a girl who	0,0000734933855953
		M	man do not	0,0000734933855953
		M	all man be	0,0000734933855953
			they can not	0,0000734933855953
			that you do	0,0000734933855953
		F	when your girl	0,0000734933855953
			how be you	0,0000734933855953
			be time for	0,0000734933855953
		F	be a milf	0,0000734933855953
		S	as a sex	0,0000734933855953
		F	cheat on she	0,0000734933855953
			be one of	0,0000734933855953
			i wish i	0,0000734933855953
		W	the prostitute i	0,0000734933855953
			one of my	0,0000734933855953
		M	a man who	0,0000734933855953
			pepperidge farm remember	0,0000734933855953
	S	F	her pussy but	0,0000734933855953
		F	to eat her	0,0000734933855953
			it be cold	0,0000734933855953
		F	woman like you	0,0000734933855953
			be a white	0,0000734933855953
			can not live	0,0000734933855953
			carry on not	0,0000734933855953
			on not a	0,0000734933855953
			put on a	0,0000734933855953
			it be hard	0,0000734933855953
		F	she be also	0,0000734933855953
			can we have	0,0000734933855953
		M	man with a	0,0000734933855953
	F	G	cook her cat	0,0000734933855953
		G	instead of cook	0,0000734933855953
			never touch a	0,0000734933855953
	F	G	of cook her	0,0000734933855953
		F	woman be not	0,0000653274538625
			this be what	0,0000653274538625
		F	what woman think	0,0000653274538625
		F	because she be	0,0000653274538625
			i at my	0,0000653274538625
	M	F	woman think man	0,0000653274538625
		M	think man want	0,0000653274538625
		M	a real man	0,0000653274538625
		F	a woman in	0,0000653274538625
			no matter what	0,0000653274538625
		P	shit out of	0,0000653274538625
		M	what man really	0,0000653274538625
			no one will	0,0000653274538625
		M	man really want	0,0000653274538625
			in real life	0,0000653274538625
			be in my	0,0000653274538625
			do what you	0,0000653274538625
		M	he be get	0,0000653274538625

		W	a prostitute to	0,0000653274538625
			they see a	0,0000653274538625
	S	F	eat her pussy	0,0000653274538625
			to talk about	0,0000653274538625
			be hard to	0,0000653274538625
		M	he will not	0,0000653274538625
			it be the	0,0000612444879961
			know how to	0,0000612444879961
			if you can	0,0000612444879961
			i like to	0,0000612444879961
		W	when the hooker	0,0000612444879961
		F	if she be	0,0000612444879961
			what it be	0,0000612444879961
			be get a	0,0000612444879961
			do not wanna	0,0000612444879961
		F	give she a	0,0000612444879961
			i be sure	0,0000612444879961
			be none of	0,0000612444879961
			that be none	0,0000612444879961
			it be ok	0,0000612444879961
			you have any	0,0000612444879961
			star war fan	0,0000612444879961
			not a problem	0,0000612444879961
		A	twitter for iphone	0,0000587947084762
			because it be	0,0000587947084762
			do not give	0,0000587947084762
			do not matter	0,0000587947084762
			you see a	0,0000587947084762
		F	take off her	0,0000587947084762
			be call a	0,0000587947084762
			i in my	0,0000587947084762
			in front of	0,0000571615221297
			this be how	0,0000571615221297
			you be tell	0,0000571615221297
			it be time	0,0000559949604536
			this be a	0,0000559949604536
			you will not	0,0000559949604536
		F	she do not	0,0000551200391965
			be the only	0,0000551200391965
			know you be	0,0000551200391965
	F	M	man and woman	0,0000551200391965
		S	to have sex	0,0000551200391965
			the difference between	0,0000544395448854
			the first time	0,0000544395448854
			for the first	0,0000544395448854
			i be your	0,0000544395448854
			to get a	0,0000538951494366
			can not get	0,0000538951494366
			do not think	0,0000538951494366
		A	imgflip.com when you	0,0000538951494366
			but you be	0,0000538951494366
			i have to	0,0000534497349784
			do not care	0,0000534497349784
			how do you	0,0000534497349784
			you be in	0,0000534497349784
			i need a	0,0000534497349784
			what be the	0,0000527644819659
			and it be	0,0000527644819659

		W	be a prostitute	0,0000524952754252
			you be the	0,0000524952754252
			do not want	0,0000515743056809
			i want to	0,0000514453699167
			i be just	0,0000512226626876
			and i be	0,0000510370733301

Tabella 29. Etichette dei trigrammi FN ordinate in modo decrescente a seconda dell'impatto.

Etichette riferite alla terza parola	Etichette riferite alla seconda parola	Etichette riferite alla prima parola	TRIGRAMMI FN	impatto +
		A	make with mematic	0,00043220464890125500
		S	for the dick	0,00033090668431502300
			you do not	0,00026043581635904600
			i do not	0,00024696937902535900
		G	in the kitchen	0,00024682383830055000
			you be a	0,00023920965131206500
			it be not	0,00023851066207121800
			when you be	0,00023706747533016600
			i be dig	0,00021447655464862600
		F	she be a	0,00019005922381170600
			if you be	0,00018751378777851300
			be not a	0,00018353567463374200
			finally some good	0,00016545334215751200
			do dev do	0,00016545334215751200
			dev do dev	0,00016545334215751200
			be go to	0,00016545334215751200
			to be a	0,00015674527151764200
			why be you	0,00014848376860289500
			do not know	0,00013561749357173100
			you have be	0,00013236267372600900
			but then i	0,00013236267372600900
			then i remember	0,00013236267372600900
			i remember why	0,00013236267372600900
			be dig in	0,00013236267372600900
			remember why i	0,00013236267372600900
			when i find	0,00011818095868393700
			what do you	0,00011030222810500800
			out of the	0,00011030222810500800
			get out of	0,00011030222810500800
			why i be	0,00011030222810500800
			out of my	0,00011030222810500800
			for my birthday	0,00011030222810500800
		A	with mematic excuse	0,00011030222810500800
		F	girlfriend want to	0,00011030222810500800
			in the garden	0,00011030222810500800
			be not leave	0,00011030222810500800
		A	mematic excuse i	0,00011030222810500800
			lego star war	0,00011030222810500800
		M	please tag he	0,00011030222810500800
			so excited i	0,00011030222810500800
			some good news	0,00011030222810500800
			i miss my	0,00011030222810500800
			a some gold	0,00011030222810500800
		F	rainbow maker she	0,00011030222810500800
			shake hand with	0,00011030222810500800

			will never leave	0,00011030222810500800
			some gold coin	0,00011030222810500800
			the rainbow maker	0,00011030222810500800
			drink because my	0,00011030222810500800
			excited i run	0,00011030222810500800
			find a some	0,00011030222810500800
			never leave you	0,00011030222810500800
			when you have	0,00011030222810500800
			i can not	0,00011030222810500800
			be suppose to	0,00009732549538677150
			be try to	0,00009454476694714950
			you need to	0,00009265387160820650
			it but then	0,00008272667107875580
			want i to	0,00008272667107875580
			your good friend	0,00008272667107875580
		G	the house to	0,00008272667107875580
			you should not	0,00008272667107875580
			excuse i what	0,00008272667107875580
			totally worth it	0,00008272667107875580
			pick up o	0,00008272667107875580
		G	to the house	0,00008272667107875580
			i will take	0,00008272667107875580
		M	say he would	0,00008272667107875580
		W	prostitute about it	0,00008272667107875580
			my garden when	0,00008272667107875580
			i get so	0,00008272667107875580
			in the west	0,00008272667107875580
			run to the	0,00008272667107875580
			by memri tv	0,00008272667107875580
		G	house to tell	0,00008272667107875580
			x x x	0,00008272667107875580
		W	the prostitute about	0,00008272667107875580
			i run to	0,00008272667107875580
			in my garden	0,00008272667107875580
			about it but	0,00008272667107875580
			dig in my	0,00008272667107875580
			dig in the	0,00008272667107875580
			garden when i	0,00008272667107875580
			get so excited	0,00008272667107875580
		P	what the fuck	0,00007353481873667180
			you be my	0,00007353481873667180
			i what the	0,00007353481873667180
			i see a	0,00007353481873667180
		W	tell the prostitute	0,00007353481873667180
	F	M	with his wife	0,00007353481873667180
			why do you	0,00006893889256562990
			and i have	0,00006893889256562990
			but you have	0,00006893889256562990
		F	my wife i	0,00006893889256562990
			in my area	0,00006893889256562990
			to tell the	0,00006893889256562990
	W	P	a dead hooker	0,00006618133686300460
			do you think	0,00006434296639458790
			i find a	0,00006434296639458790
			the rest of	0,00006204500330906680
		S	be a virgin	0,00006127901561389310
			because i be	0,00005974704022354590
			i have be	0,00005974704022354590

		i be go	0,00005839529723206290
--	--	---------	------------------------

3. Liste di similarità

Riporto di seguito la lista 1, le liste di similarità.

woman [0.7039182113350886, 0.8727775483580702, 0.7144477323721208, 0.7039182113350886, 0.7384694574643361, 0.7184680156719669, 0.868033807577512, 0.7054353836686575, 0.7039182113350886, 0.7184680156719669, 0.722765168608347, 0.7808448239209353, 0.7993685938602284, 0.7183762916918626, 0.7115128853395083, 0.7934696909560405, 0.7032510949524735, 0.7642751224630179, 0.7934696909560405, 0.729757009250365, 0.8905466734231605, 0.7141399438625695, 0.7122586136396153, 0.7934696909560405, 0.868033807577512, 0.7184680156719669, 0.7860129878677752, 0.709245160740714, 0.7122586136396153, 0.7032510949524735, 0.9999999999999747, 0.7959752989724832, 0.725659898285399, 0.7210054668066792, 0.7654194735683546, 0.7384694574643361, 0.729757009250365, 0.7808448239209353, 0.7785426382051884, 0.9999999999999747, 0.8727775483580702, 0.868033807577512, 0.722765168608347, 0.7388747842277752, 0.725659898285399, 0.7100078011653157, 0.7642751224630179, 0.7144477323721208, 0.7183762916918626, 0.7292381075820495, 0.8727775483580702, 0.7292381075820495, 0.7115128853395083, 0.9999999999999747, 0.7100078011653157, 0.7993685938602284, 0.7813725707074948, 0.7032510949524735, 0.7054353836686575, 0.7934696909560405, 0.7642751224630179, 0.7223488745108093, 0.8905466734231605, 0.7122586109561146, 0.7210054668066792, 0.7949570205785226, 0.7047003291260836, 0.7388747842277752, 0.7119288552711103, 0.736037662111926, 0.8905466734231605, 0.7808448239209353, 0.7949570205785226, 0.722765168608347, 0.737983556512224, 0.7860129878677752, 0.9332010155124377, 0.7949570205785226, 0.722765168608347, 0.7119288552711103, 0.7183762916918626, 0.7054353836686575, 0.70737536620208, 0.7881734736271491, 0.729757009250365, 0.7144477323721208, 0.7100078011653157, 0.7754630166338643, 0.7141399438625695, 0.729757009250365, 0.7860129878677752, 0.7141399438625695, 0.7230267181869716, 0.7388747842277752, 0.725659898285399, 0.7813725707074948, 0.7813725707074948, 0.701991808075612, 0.7184680156719669, 0.9332010155124377, 0.7119288552711103] ['Girls', 'GIRL', 'MAN', 'GIRLS', 'womyn', 'wife', 'chick', 'Her', 'girls', 'Wife', 'SHE', 'Females', 'Gendered', 'Sexist', 'HERS', 'women', 'MANS', 'Guy', 'WoMeN', 'men', 'LADY', 'male', 'MISOGYNIST', 'Women', 'Chick', 'W.I.F.E.', 'genders', 'Lorena', 'misogynist', 'Mans', 'WOMAN', 'WOMANLY', 'Ladies', 'PATRIARCHAL', 'Allfemale', 'WOMYN', 'MEN', 'females', 'Gal', 'Woman', 'Girl', 'CHICK', 'she', 'FEMININE', 'LADIES', 'Femininity', 'guy', 'Man', 'sexist', 'Madam', 'girl', 'madam', 'hers', 'woman', 'femininity', 'gendered', 'Gender', 'mans', 'her', 'WOMEN', 'GUY', 'unladylike', 'lady', 'Misogynist', 'patriarchal', 'female', 'LADYLIKE', 'feminine', 'shes', 'SEXISTS', 'Lady', 'FEMALES', 'Female', 'SHe', 'Unisex', 'GENDERS', 'womans', 'FEMALE', 'She', 'SHES', 'SEXIST', 'HER', 'chauvinistic', 'SEXES', 'Men', 'man', 'FEMININITY', 'womankind', 'Male', 'mEn', 'Genders', 'MALE', 'misogynistic', 'Feminine', 'ladies', 'GENDER', 'gender', 'genderless', 'WIFE', 'WOMANS', 'Shes']

man [0.735814330623712, 0.7453222412078582, 0.9999999999999764, 0.7046213140112492, 0.71029320652772, 0.7046213140112492, 0.7204489021240599, 0.7453222412078582, 0.7503586919842749, 0.8666523844729125, 0.8144919400221611, 0.7234060828701444, 0.803298656045745, 0.759731321899796, 0.8666523844729125, 0.7144477494389287, 0.803298656045745, 0.7234060828701444, 0.735814330623712, 0.7144477494389287, 0.7204489021240599, 0.8144919400221611, 0.9999999999999764, 0.7144477494389287, 0.7204489021240599, 0.8666523844729125, 0.8144919400221611, 0.735814330623712, 0.7503586919842749, 0.7453222412078582, 0.7234060828701444, 0.9999999999999764, 0.803298656045745, 0.7234060828701444, 0.7055311351942428, 0.759731321899796, 0.7503586919842749] ['Dudes', 'boy', 'MAN', 'gentleman', 'fellas', 'GENTLEMAN', 'him', 'Boy', 'BRO', 'MANS', 'Guy', 'men', 'DUDE', 'FELLA', 'Mans', 'WOMAN', 'Dude', 'MEN', 'dudes', 'Woman', 'HIM', 'guy', 'Man', 'woman', 'Him', 'mans', 'GUY', 'DUDES', 'Bro', 'BOY', 'Men', 'man', 'dude', 'mEn', 'DOOD', 'fella', 'bro']

girl [0.7894255142101294, 0.9999999999999785, 0.7003967396670108, 0.7947972910966268, 0.7894255142101294, 0.9420753307453593, 0.7313788200768265, 0.7894255142101294, 0.7413620621147134, 0.7794003902209543, 0.7947972910966268, 0.7670434360330058, 0.8259542925860566, 0.844016096113302, 0.7382657304264573, 0.9420753307453593, 0.7963377305605156, 0.8727775473705608, 0.7407472409861915, 0.7203323622818598, 0.7003967396670108, 0.7225302938857773, 0.7794003902209543, 0.8339760775232277, 0.7203323622818598, 0.8727775473705608, 0.9999999999999785, 0.7003967396670108, 0.9420753307453593, 0.7413620621147134, 0.7622420268344792, 0.8259542925860566, 0.9999999999999785, 0.8727775473705608, 0.7670434360330058, 0.7942081229080157, 0.8259542925860566, 0.7203323622818598, 0.844016096113302, 0.8236890617071679, 0.7622420268344792, 0.7134800215916566, 0.844016096113302, 0.7794003902209543, 0.7947972910966268, 0.8236890617071679, 0.7413620621147134, 0.7687259894550664, 0.7963377305605156, 0.8074559268958492, 0.8236890617071679, 0.7413620621147134, 0.7063103003030001, 0.7134800215916566, 0.7052439930040865, 0.7619458773277588, 0.7382657304264573, 0.7963377305605156, 0.7382657304264573, 0.7622420268344792, 0.7942081229080157, 0.7942081229080157, 0.7430835863211392, 0.8074559268958492, 0.7134800215916566] ['Girls', 'GIRL', 'Boys', 'boy', 'GIRLS', 'chick', 'GURLZ', 'girls', 'SHE', 'Females', 'Boy', 'Gendered', 'Guy', 'LADY', 'male', 'Chick', 'genders', 'WOMAN', 'WOMANLY', 'BABE', 'BOYS', 'Allfemale', 'females', 'Gal', 'babe', 'Woman', 'Girl', 'boys', 'CHICK', 'she', 'FEMININE', 'guy', 'girl', 'woman', 'gendered', 'Gender', 'GUY', 'Babe', 'lady', 'female', 'feminine', 'shes', 'Lady', 'FEMALES', 'BOY', 'Female', 'SHe', 'Unisex', 'GENDERS', 'womans', 'FEMALE', 'She', 'Androgynous', 'SHES', 'LESBO', 'SEXES', 'Male', 'Genders', 'MALE', 'Feminine', 'GENDER', 'gender', 'genderless', 'WOMANS', 'Shes']

boy [0.7947972996636973, 0.8193999284351702, 0.9999999999999758, 0.7453222492231049, 0.7467518734576168, 0.7185649323288653, 0.9999999999999758, 0.7420320360282557, 0.7155147992856912, 0.810744045613209, 0.7196232433926754, 0.7045322369365841, 0.7185649323288653, 0.7168823685164278, 0.7420320360282557, 0.7155147992856912, 0.7471014410220164, 0.8193999284351702, 0.7467518734576168, 0.7155147992856912, 0.7054328585491906, 0.7947972996636973, 0.8193999284351702, 0.7185649323288653, 0.810744045613209, 0.7453222492231049, 0.7947972996636973, 0.7272505544766645, 0.7045322369365841, 0.7420320360282557, 0.810744045613209, 0.7471014410220164, 0.9999999999999758, 0.7118988955209155, 0.7168823685164278, 0.7467518734576168, 0.7453222492231049, 0.7196232433926754, 0.7045322369365841, 0.7168823685164278, 0.7196232433926754, 0.7659643041283372, 0.7272505544766645, 0.7272505544766645, 0.7471014410220164] ['GIRL', 'Boys', 'boy', 'MAN', 'Son', 'chick', 'Boy', 'MANS', 'DAUGHTER', 'Guy', 'male', 'kid', 'Chick', 'genders', 'Mans', 'daughter', 'Child', 'BOYS', 'SON', 'Daughter', 'Gal', 'Girl', 'boys', 'CHICK', 'guy', 'Man', 'girl', 'Gender', 'Kid', 'mans', 'GUY', 'child', 'BOY', 'Unisex', 'GENDERS', 'son', 'man', 'Male', 'KID', 'Genders', 'MALE', 'childs', 'GENDER', 'gender', 'CHILD']

guy [0.7478656288830325, 0.8259542931300285, 0.7126469165294099, 0.8107440480053151, 0.8144919414130403, 0.729678873438796, 0.8003047756081065, 0.7099662042408305, 0.729678873438796, 0.7642140507902273, 0.8107440480053151, 0.7070640029747018, 0.7444977004851125, 0.9999999999999745, 0.733559194225539, 0.7454874939122762, 0.7467160104389314, 0.8003047756081065, 0.731907461738083, 0.7482268343722104, 0.7444977004851125, 0.7642751147885514, 0.7126469165294099, 0.733559194225539, 0.7478656288830325, 0.7261874903962663, 0.7195880118799978, 0.7642751147885514, 0.8259542931300285, 0.7642140507902273, 0.7126469165294099, 0.8003047756081065, 0.9999999999999745, 0.8144919414130403, 0.8259542931300285, 0.7642751147885514, 0.7070640029747018, 0.7642140507902273, 0.7151713669459365, 0.7444977004851125, 0.9999999999999745, 0.7478656288830325, 0.7454874939122762, 0.7111463767818068, 0.7454874939122762, 0.8107440480053151, 0.7111463767818068, 0.7423534689311891, 0.731907461738083, 0.7255096633225533, 0.7111463767818068, 0.7352125694467475, 0.733559194225539, 0.8144919414130403, 0.7467160104389314, 0.7195880118799978, 0.733559194225539, 0.731907461738083, 0.7467160104389314, 0.7151713669459365, 0.7151713669459365, 0.7195880118799978, 0.7482268343722104, 0.7255096633225533] ['Dudes', 'GIRL', 'Boys', 'boy', 'MAN', 'gentleman', 'chick', 'fellas', 'GENTLEMAN', 'him', 'Boy', 'Gendered', 'MANS', 'Guy', 'men', 'LADY', 'male', 'Chick', 'genders', 'FELLA', 'Mans', 'WOMAN', 'BOYS', 'MEN', 'dudes', 'Gal', 'MALES', 'Woman', 'Girl', 'HIM', 'boys', 'CHICK', 'guy', 'Man', 'girl', 'woman', 'gendered', 'Him', 'Gender', 'mans', 'GUY', 'DUDES', 'lady', 'female', 'Lady', 'BOY', 'Female', 'Unisex', 'GENDERS',

'womans', 'FEMALE', 'SEXES', 'Men', 'man', 'Male', 'males', 'mEn', 'Genders', 'MALE', 'GENDER', 'gender', 'Males', 'fella', 'WOMANS']

prostitute [0.7844458810835525, 0.7516263496419426, 0.8723168147939049, 0.7539103557810464, 0.7609207996931164, 0.7030925363510194, 0.7412396935379891, 0.7926108809134134, 0.8962002368294406, 0.8680241621062696, 0.7030925363510194, 0.7180566443217978, 0.7412396935379891, 0.8723168147939049, 0.8962002368294406, 0.7587398718234657, 0.7180566443217978, 0.8680241621062696, 0.7079985899334663, 0.8766125404841785, 0.7516263496419426, 0.764320785080423, 0.7002450863344165, 0.7002450863344165, 0.7030925299657723, 0.8629506651471883, 0.7733196740443078, 0.8723168147939049, 0.999999999999722, 0.7202257127902311, 0.999999999999722, 0.7202257127902311, 0.8962002368294406, 0.7539103557810464, 0.8064551717548761, 0.7180566443217978, 0.7412396935379891, 0.8064551717548761, 0.7926108809134134, 0.70309675812432, 0.8629506651471883, 0.7002450863344165, 0.7202257127902311, 0.7030925363510194, 0.8766125404841785, 0.70309675812432, 0.7383314222748246, 0.8479014141425626, 0.7079985899334663, 0.800848302940818, 0.8766125404841785, 0.7926108809134134, 0.999999999999722] ['stripping', 'STRIPPERS', 'PROSTITUTES', 'Waitress', 'businesswoman', 'BLOWJOB', 'Whores', 'Hookers', 'Prostitution', 'ESCORT', 'BlowJob', 'Trafficking', 'WHORES', 'prostitutes', 'PROSTITUTION', 'CAMGIRL', 'trafficking', 'escort', 'MAIDS', 'HOOKER', 'strippers', 'Hostesses', 'Whore', 'WHORE', 'Blowjob', 'STRIPPER', 'golddigger', 'Prostitutes', 'prostitute', 'maid', 'PROSTITUTE', 'MAID', 'prostitution', 'WAITRESS', 'pornstar', 'TRAFFICKING', 'whores', 'Pornstar', 'hookers', 'babysitter', 'stripper', 'whore', 'Maid', 'blowjob', 'hooker', 'Babysitter', 'pimps', 'BROTHEL', 'maids', 'Prostituted', 'Hooker', 'HOOKERS', 'Prostitute']

hooker [0.7026968917060307, 0.7293624966184222, 0.7675959628882326, 0.7298899636466927, 0.8046629000322589, 0.7540433536336402, 0.8151966569483353, 0.7675959628882326, 0.7540433536336402, 0.8151966569483353, 0.999999999999767, 0.7293624966184222, 0.7161829247199518, 0.8076452401481713, 0.7601303176844749, 0.7124076930355547, 0.7675959628882326, 0.8766125377015983, 0.8766125377015983, 0.7540433536336402, 0.7298899636466927, 0.7331221044663594, 0.7289957319698729, 0.7331221044663594, 0.8046629000322589, 0.8076452401481713, 0.999999999999767, 0.7021930126452718, 0.7266015064444202, 0.756801622501203, 0.7124076930355547, 0.999999999999767, 0.8046629000322589, 0.8766125377015983] ['stripping', 'STRIPPERS', 'PROSTITUTES', 'Waitress', 'Hookers', 'Prostitution', 'ESCORT', 'prostitutes', 'PROSTITUTION', 'escort', 'HOOKER', 'strippers', 'Hostesses', 'STRIPPER', 'golddigger', 'Bj', 'Prostitutes', 'prostitute', 'PROSTITUTE', 'prostitution', 'WAITRESS', 'pornstar', 'SHAG', 'Pornstar', 'hookers', 'stripper', 'hooker', 'banging', 'pimps', 'BROTHEL', 'BJ', 'Hooker', 'HOOKERS', 'Prostitute']

hoe [0.7026114359534861, 0.7542482728961566, 0.7082719939383634, 0.716852784168698, 0.7041741734009069, 0.7294128226398908, 0.7083396332310121, 0.7033923986391951, 0.7354674043076317, 0.7096292117837866, 0.8132990461901738, 0.8169095854664636, 0.7770011281851895, 0.7143252144833545, 0.7410715671533807, 0.7696656820776224, 0.7833611818776027, 0.7207662271955279, 0.7059340761987936, 0.7250157213493458, 0.7729669646873372, 0.7615861967754421, 0.7189932039671927, 0.7317974763680757, 0.7055137067029988, 0.781952538644677, 0.8169095854664636, 0.999999999999979, 0.7770011281851895, 0.7685681078385029, 0.7993522311073833, 0.7026114359534861, 0.8177578956674899, 0.7003924158326909, 0.7831659424658369, 0.75841616939049, 0.7111246957528858, 0.7436764938570228, 0.7162407823739497, 0.720292829043246, 0.781952538644677, 0.999999999999979, 0.7357079558212225, 0.7249030276005612, 0.7615861967754421, 0.7833611818776027, 0.7602366610442093, 0.7168916360489901, 0.8730255677663472, 0.7106247593809171, 0.720292829043246, 0.7496518077720772, 0.7041741734009069, 0.7143977875852661, 0.7642721624779906, 0.8351520867038643, 0.8730255677663472, 0.7201289630092882, 0.8132990461901738, 0.7198500363915707, 0.7305596131902095, 0.735392119312294, 0.7305596131902095, 0.7109301869390017, 0.7354674043076317, 0.7348358406311044, 0.8096434733847873, 0.7993522311073833, 0.735392119312294, 0.7312536094654767, 0.7354674043076317, 0.7065234220037748, 0.7249030276005612, 0.7921737907296318, 0.7150405514240099, 0.7104524631215894, 0.7344099838140704, 0.7567447278499744, 0.7194583027675846, 0.7143977875852661, 0.7729669646873372, 0.7229931172514028, 0.8096434733847873, 0.7140691910531137,

0.7041741734009069, 0.7051758803211785, 0.7238884787934234, 0.7160182265676451, 0.75841616939049, 0.7143252144833545, 0.7436764938570228, 0.8730255677663472, 0.75841616939049, 0.7707707766001959, 0.7402542539073647, 0.7065234220037748, 0.7297898129475207, 0.7046252353696034, 0.7395864776601283, 0.72451528056218, 0.7348358406311044, 0.7833611818776027, 0.7554926247788755, 0.7720572278393963, 0.7440019687445983, 0.7081855258464641, 0.725651592410588, 0.7111246957528858, 0.7435184856873243, 0.7283673184852619, 0.7315808708548465, 0.7305596131902095, 0.7993522311073833, 0.7143977875852661, 0.7192237875748523, 0.7038598750718674, 0.9999999999999979, 0.7395864776601283, 0.8351520867038643, 0.7185968344796984, 0.7177207451054566, 0.7514191722085806, 0.7519345133331254, 0.7232893203182307, 0.7290477934259689, 0.735392119312294, 0.7014971391790569, 0.7831659424658369, 0.7615861967754421, 0.7357394998756742, 0.7207717048217905, 0.7109301869390017, 0.8096434733847873, 0.7026853381809242] ['ary', 'THUR', 'NIGGER', 'INA', 'Imma', 'packin', 'JAS', 'beez_', 'Ima', 'sayin', 'pimp', 'skank', 'HIPPIITY', 'ell', 'MUTHAFUCKAS', 'fuckboi', 'BITCH', 'FING', 'GURLZ', 'tryna', 'DIGGER', 'SLUT', 'motherfucking', 'Ä,299', 'DONK', 'NIGGA', 'SKANK', 'HOE', 'Hippity', 'SHAWTY', 'Simp', 'ARY', 'bich', 'DRO', 'FUCKBOY', 'NIGGAS', 'Awomen', 'fuckin', 'PUH', 'cha', 'nigga', 'Hoe', 'SWAG', 'RACK', 'slut', 'Bitch', 'MUTHAFUCKA', 'Grils', 'thot', 'Yoy', 'Cha', 'biddies', 'imma', 'yall', 'GEY', 'HOES', 'Thot', 'BWORD', 'PIMP', 'FUCKA', 'Whore', 'BITCHES', 'WHORE', 'BAE', 'IMA', 'WIFEGF', 'HO', 'simp', 'bitches', 'yout', 'ima', 'wit', 'rack', 'Niggaz', 'dere', 'matta', 'diz', 'finna', 'chea', 'YALL', 'digger', 'SHAG', 'Ho', 'Äœ', 'IMMA', 'Mong', 'Boi', 'LER', 'Niggas', 'ELL', 'FUCKIN', 'THOT', 'niggas', 'BTCH', 'coochie', 'WIT', 'THHHHILL', 'shitface', 'FUKIN', 'Herpina', 'wifegf', 'bitch', 'mothafuckin', 'Biatch', 'LELE', 'SIK', 'asf', 'awomen', 'mofo', 'Thug', 'LESBO', 'whore', 'SIMP', 'Yall', 'harlot', 'nword', 'hoe', 'Fukin', 'hoes', 'MAK', 'sluttiest', 'DOOD', 'MOTHAFUCKA', 'pimps', 'lyf', 'Bitches', 'youyou', 'fuckboy', 'Slut', 'fuckboys', 'Wack', 'bae', 'ho', 'BAK']

slut [0.7283085309499133, 0.7085268063439552, 0.8976763101239622, 0.7155552221806569, 0.7264306039537396, 0.7026290695827797, 0.7327391730510548, 0.7404701532653994, 0.7129010882190521, 0.749965484898353, 0.7301277738771412, 0.7161398688786977, 0.7303966280728906, 0.9999999999999762, 0.7085268063439552, 0.8976763101239622, 0.7615862020335025, 0.7065343098899712, 0.7585912648637758, 0.7770570054264833, 0.7582577518465169, 0.7037634974673057, 0.7615862020335025, 0.7037634974673057, 0.7231817835601391, 0.9999999999999762, 0.7129010882190521, 0.7433061768128262, 0.7212842574139222, 0.7132108836702036, 0.7347219972825745, 0.7653025507703008, 0.7212842574139222, 0.7139119050226186, 0.9146137080988653, 0.8695753732096336, 0.7021690387251573, 0.9146137080988653, 0.7164827882559102, 0.72114011002933, 0.8695753720999159, 0.7305666538594808, 0.7018754356774686, 0.7256260247580176, 0.7063026788829582, 0.7550248332228737, 0.7585912648637758, 0.7726009657289326, 0.7085268063439552, 0.8170604910952508, 0.7550248332228737, 0.7327391730510548, 0.8170604910952508, 0.7638690287091748, 0.7550248332228737, 0.741008115906121, 0.7037634959977405, 0.8170604910952508, 0.7428405778489007, 0.7471260012685332, 0.7256260247580176, 0.7379681049624051, 0.7212842574139222, 0.7188134339777986, 0.7327391730510548, 0.7231817835601391, 0.716982102041001, 0.7045275757346777, 0.7673717701450351, 0.7444862431035936, 0.7271223779475372, 0.7305666538594808, 0.7129010882190521, 0.7423727612973126, 0.7202597413262782, 0.7019904325848314, 0.830455160469389, 0.9146137080988653, 0.701314307325534, 0.7585912648637758, 0.7037029216613457, 0.7322097966298723, 0.7037029216613457, 0.7202597413262782, 0.8695753720999159, 0.821258081441083, 0.7066640954605787, 0.7615862020335025, 0.7264306039537396, 0.7471260012685332, 0.7202597413262782, 0.9401652193033956, 0.7597005232654672, 0.7231817835601391, 0.7582577518465169, 0.9999999999999762, 0.7090121100031166, 0.7256260247580176, 0.7021690387251573] ['Wannabe', 'ASSHOLE', 'skank', 'TRASHY', 'LOSER', 'sexualising', 'douchebag', 'fuckboi', 'BITCH', 'girlfiend', 'Naughty', 'GURLZ', 'Fetishist', 'SLUT', 'asshole', 'SKANK', 'HOE', 'PRICK', 'Simp', 'bich', 'FUCKBOY', 'MISOGYNIST', 'Hoe', 'misogynist', 'GAY', 'slut', 'Bitch', 'Douche', 'thot', 'BTCHES', 'biddies', 'GEY', 'Thot', 'Sucker', 'Whore', 'Slutty', 'homosexual', 'WHORE', 'jerk', 'degrading', 'slutty', 'WIFEGF', 'androgynne', 'HO', 'demeaning', 'SEXUAL', 'simp', 'FUCKABLE', 'Asshole', 'SLUTS', 'Sexual', 'DOUCHEBAG', 'sluts', 'COCKBLOCKER', 'sexual', 'unladylike', 'Misogynist', 'Sluts', 'beingron.com', 'SEXUALLY', 'Ho', 'sexs', 'THOT', 'deviants', 'Douchebag', 'Gay', 'BTCH', 'coochie', 'TRAMP', 'shitface', 'Herpina', 'wifegf', 'bitch', 'Biatch', 'horny', 'BITCHBOY', 'LESBO', 'whore', 'twat', 'SIMP', 'lover', 'weirdo', 'LOVER', 'Horny', 'SLUTTY', 'harlot',

'rapey', 'hoe', 'loser', 'sexually', 'HORNY', 'sluttiest', 'pervert', 'gay', 'fuckboy', 'Slut', 'Homophobe', 'ho', 'HOMOSEXUAL']

whore [0.8452945050288219, 0.7003966192245281, 0.7239816817770661, 0.7152272134319158, 0.7622113024721993, 0.9146137067439432, 0.8452945050288219, 0.730559602187727, 0.7116381448396083, 0.7235335031664663, 0.7500120841506738, 0.7622113024721993, 0.730559602187727, 0.9146137067439432, 0.7239816817770661, 0.7003966192245281, 0.7562153426478041, 0.9999999999999756, 0.7596284828324066, 0.9999999999999756, 0.7143138065861666, 0.7596284840874142, 0.7002450878754787, 0.7119375176306688, 0.7235335031664663, 0.7211832876967192, 0.7002450878754787, 0.7105441189016586, 0.7105441189016586, 0.7105441189016586, 0.7119375176306688, 0.7622113024721993, 0.7175045987800482, 0.7117034381576186, 0.7013796993744308, 0.7239816817770661, 0.7391895237339696, 0.7303051435038885, 0.9999999999999756, 0.7235335031664663, 0.7596284840874142, 0.7883091326010523, 0.730559602187727, 0.8418908788350632, 0.7110656481466124, 0.9146137067439432, 0.7003966192245281, 0.7002450878754787, 0.7119375176306688] ['skank', 'Cunt', 'BITCH', 'girlfiend', 'Whores', 'SLUT', 'SKANK', 'HOE', 'PRICK', 'Simp', 'bich', 'WHORES', 'Hoe', 'slut', 'Bitch', 'cunt', 'Sucker', 'Whore', 'Slutty', 'WHORE', 'jerk', 'slutty', 'prostitute', 'HO', 'simp', 'FUCKABLE', 'PROSTITUTE', 'SLUTS', 'sluts', 'Sluts', 'Ho', 'whores', 'BTCH', 'TRAMP', 'shitface', 'bitch', 'Biatch', 'LESBO', 'whore', 'SIMP', 'SLUTTY', 'harlot', 'hoe', 'sluttiest', 'pervert', 'Slut', 'CUNT', 'Prostitute', 'ho']

bitch [0.7090494110888299, 0.7233585416877678, 0.7244298739366253, 0.7026641964149706, 0.7643816016340494, 0.8055262915706465, 0.7351521443947872, 0.7024167193061435, 0.7366050441530765, 0.7871307089412907, 0.7228523893969034, 0.9999999999999809, 0.700126695253649, 0.7739294252791618, 0.725222842065573, 0.7420361662596775, 0.7129010737540237, 0.70022832328944, 0.7643816016340494, 0.7859916790816184, 0.7091651010629948, 0.7223145509477915, 0.7170364233260961, 0.7294799678696438, 0.7162621955294783, 0.8055262915706465, 0.7833611769190119, 0.7503588154777924, 0.7351521443947872, 0.8135377920588749, 0.8253298402230329, 0.7101170504927804, 0.7308276357882293, 0.700126981275043, 0.7026641964149706, 0.7294799678696438, 0.7833611769190119, 0.7836009197348497, 0.7129010737540237, 0.70022832328944, 0.9999999999999809, 0.8621427670329574, 0.7057943619309055, 0.700585803128137, 0.7366050441530765, 0.7194941833874064, 0.700126981275043, 0.7071006913453506, 0.7057943619309055, 0.7071006913453506, 0.7194941833874064, 0.7050050296909596, 0.8746839313501447, 0.7239816873920485, 0.7080784625851311, 0.8282902819257077, 0.7239816873920485, 0.7343024568551172, 0.7244751291974993, 0.7859897859698162, 0.8135377920588749, 0.8282902819257077, 0.7643816016340494, 0.7607839608810106, 0.7011360234966293, 0.7101493422474348, 0.7070605296841743, 0.7013118220558953, 0.7024167193061435, 0.7228523893969034, 0.7057943619309055, 0.7074801417696143, 0.7859897859698162, 0.700126695253649, 0.7308276357882293, 0.7180830010133823, 0.7194941833874064, 0.7101170504927804, 0.7024167193061435, 0.7011360234966293, 0.815558059954488, 0.9744380802756533, 0.7335594146924572, 0.7430153246956306, 0.7244751291974993, 0.7101493404137693, 0.9999999999999809, 0.7667960635578434, 0.9527582092726056, 0.750444273460672, 0.7067075252286801, 0.7839932373603142, 0.7101170504927804, 0.7239816873920485, 0.7072697210197847, 0.8135377920588749, 0.7101493422474348, 0.7074801417696143, 0.7228523893969034, 0.7180830010133823, 0.7009874784737773, 0.7365033879435656, 0.7074801417696143, 0.7086285450116578, 0.7833611769190119, 0.7430153246956306, 0.700126695253649, 0.815558059954488, 0.7272143621326133, 0.7315392608183815, 0.7022706345798979, 0.8450390073613445, 0.7836009197348497, 0.8282902819257077, 0.710481714977317, 0.7129010737540237, 0.7272023668500429, 0.7878150123056206, 0.7366050441530765, 0.7859897859698162, 0.7170364233260961] ['STFU', 'THUR', 'sayin', 'heck', 'ASSHOLE', 'skank', 'HIPPIITY', 'douchebag', 'Cunt', 'MUTHAFUCKAS', 'Shut', 'BITCH', 'dick', 'FING', 'FUC', 'motherfuckers', 'SLUT', 'MAMAS', 'asshole', 'motherfucking', 'JUSTA', 'SWEAR', 'NAG', 'NIGGA', 'fricken', 'SKANK', 'HOE', 'PRICK', 'Hippity', 'Simp', 'bich', 'FCK', 'fuckin', 'cha', 'HECK', 'nigga', 'Hoe', 'BITCHY', 'slut', 'mamas', 'Bitch', 'MUTHAFUCKA', 'BABE', 'ISHE', 'cunt', 'thot', 'Cha', 'BASTARD', 'babe', 'bastard', 'Thot', 'Sucker', 'FUCKA', 'Whore', 'DUMBASS', 'BITCHES', 'WHORE', 'jerk', 'WIFEGF', 'HO', 'simp', 'bitches', 'Asshole', 'GOSH', 'complains', 'Mean', 'Niggaz', 'dere', 'DOUCHEBAG', 'SHUT', 'Babe', 'DEAR', 'Ho', 'DICK', 'FUCKIN', 'SHT', 'THOT', 'Fck', 'Douchebag', 'COMPLAINS', 'motherfucker', 'BTCH', 'shitface', 'FUKIN',

'wifegf', 'MEAN', 'bitch', 'mothafuckin', 'Biatch', 'fcking', 'hooo', 'mofo', 'fck', 'whore', 'twat', 'SIMP', 'mean', 'Dear', 'shut', 'sht', 'whoop', 'MOMMA', 'dear', 'WHOUSE', 'hoe', 'Fukin', 'Dick', 'MOTHERFUCKER', 'MOOOOM', 'SHUTUP', 'DOOD', 'MOTHAFUCKA', 'bitchy', 'Bitches', 'youyou', 'Slut', 'FU', 'fucker', 'CUNT', 'ho', 'nag']

cougar [0.8060778826489496, 0.999999999999752, 0.999999999999752, 0.7305499310774868, 0.7305499310774868, 0.7305499310774868, 0.8060778826489496, 0.8060778826489496, 0.999999999999752] ['cougars', 'COUGAR', 'cougar', 'Cowgirl', 'COWGIRL', 'cowgirl', 'COUGARS', 'Cougars', 'Cougar']

milf [0.7171042840910113, 0.7299360350679454, 0.999999999999791, 0.7097831171536257, 0.7094377525155586, 0.705708185919895, 0.7171042840910113, 0.999999999999791, 0.7094377525155586, 0.7040866322116035, 0.7018671673050891, 0.7097831171536257, 0.7006374494989092, 0.7005142062145366, 0.999999999999791, 0.8729139194140094, 0.7264982212820741, 0.702846525242538, 0.7040866322116035, 0.7006374494989092, 0.7006374494989092, 0.7005142062145366, 0.7094377525155586, 0.7264982212820741, 0.7301186172122399, 0.8729139194140094, 0.7240982447364706, 0.7536157289278965, 0.8729139194140094, 0.7018671673050891, 0.7168848592839174, 0.7168848592839174, 0.999999999999791, 0.7041972610250287, 0.7301186172122399, 0.7018671673050891, 0.7080785114351824, 0.7199578353028957, 0.7416014830265459, 0.7041972610250287, 0.7459009159991992, 0.7240982447364706, 0.7299360350679454] ['STEPMOM', 'Babes', 'MILF', 'skank', 'DADDY', 'sexualising', 'stepmom', 'M.I.L.F', 'daddy', 'MAMAS', 'Butterface', 'SKANK', 'DAUGHTER', 'daddys', 'milf', 'MILFS', 'granddaughter', 'Lorena', 'mamas', 'daughter', 'Daughter', 'DADDYS', 'Daddy', 'Granddaughter', 'granny', 'Milfs', 'HOTNESS', 'FUCKABLE', 'milfs', 'BUTTERFACE', 'Brazzers', 'BRAZZERS', 'Milf', 'babysitter', 'GRANNY', 'butterface', 'MOMMA', 'auntie', 'MOOOOM', 'Babysitter', 'sluttiest', 'hotness', 'babes']

chad offensivi [0.7818764045234416, 0.7021553576823734, 0.7236445015592643, 0.7039068070387873, 0.7418927818455675, 0.7517099933188023, 0.7115480568375285, 0.7860959951102342, 0.722214898668154, 0.8963594596568911, 0.7418927818455675, 0.7282780655329395, 0.728512341571707, 0.7236445015592643, 0.7551493335268822, 0.7039068070387873, 0.7423685286607042, 0.7756090727913583, 0.7433001425464267, 0.999999999999741, 0.7188763345353324, 0.7517099933188023, 0.7860959951102342, 0.7334301927584099, 0.722214898668154, 0.7396767594908962, 0.728512341571707, 0.7418927818455675, 0.7084890823380547, 0.7243307279490384, 0.7188763345353324, 0.7396767594908962, 0.7860959951102342, 0.7762244958649358, 0.722214898668154, 0.7467412241272685, 0.999999999999741, 0.7334301927584099, 0.7567483420543656, 0.7115480568375285, 0.7517099933188023, 0.7129395228292987, 0.7069208981548865, 0.7311541607917211, 0.999999999999741, 0.751222389122002, 0.727707758737002, 0.7243307279490384, 0.7129395228292987, 0.7567483420543656, 0.7021553701782975, 0.7398721980622021, 0.7321118640058114, 0.7218365268674598, 0.7754011340180864, 0.7334301927584099, 0.7567483420543656, 0.727707758737002, 0.728512341571707, 0.7243307279490384, 0.7069208981548865] ['Corey', 'KEVIN', 'Jake', 'JOE', 'jeff', 'Cyrus', 'Dave', 'Justin', 'PATrick', 'chads', 'Jeff', 'DANIEL', 'Mike', 'JAKE', 'SEAN', 'Joe', 'Erin', 'KYLE', 'Brent', 'chad', 'Steve', 'cyrus', 'JUSTIN', 'CHRIS', 'PATRICK', 'joey', 'mike', 'JEFF', 'Dan', 'sam', 'STEVE', 'Joey', 'justin', 'Andrew', 'Patrick', 'EVAN', 'Chad', 'chris', 'Zach', 'DAVE', 'CYRUS', 'JOHN', 'ADAM', 'William', 'CHAD', 'Kyles', 'Brad', 'Sam', 'John', 'ZACH', 'Kevin', 'Jamie', 'BRIAN', 'Nick', 'Ryan', 'Chris', 'zach', 'BRAD', 'MIKE', 'SAM', 'Adam']

stacy [0.7760785717081011, 0.7541653327826375, 0.7967104220911375, 0.7219440797073986, 0.999999999999702, 0.7450438067692847, 0.7864515287735565, 0.7639692680539961, 0.7677504725678777, 0.7090430161344756, 0.7154098435367989, 0.7480582773981244, 0.7755108543746421, 0.7563149858551899, 0.7038886609791768, 0.7315379587555504, 0.7984304624670494, 0.7772813499571274, 0.7267290492412195, 0.7707487049930898, 0.7530659088692544, 0.7479984299818481, 0.744095094733162, 0.7213447024837315, 0.7680444088403934, 0.7333564914777235, 0.7405801667963213, 0.7678871728275941, 0.706594371495584, 0.7772813499571274, 0.7149077254645313, 0.7992596938153196,

0.7019842577783262, 0.7221007221232066, 0.7633733610544805, 0.7136331705398186, 0.7253211978359577, 0.729971987237355, 0.758956175233563, 0.708969884241688, 0.7633733610544805, 0.7678871728275941, 0.8481581113469961, 0.7090430161344756, 0.7710758204454438, 0.8193180468270006, 0.7728735330158958, 0.7523456353946634, 0.7337241812905004, 0.8125247398182665, 0.7063824516468251, 0.7011559670238636, 0.7403075853145938, 0.7616381247029389, 0.7267290492412195, 0.7235876722238723, 0.7330745447413529, 0.7710758204454438, 0.7992596938153196, 0.7443921128353408, 0.7391128330259225, 0.7405801667963213, 0.7438058594911547, 0.8005399526456539, 0.7286586855712416, 0.7033245812388537, 0.7635288160444316, 0.8167165507308063, 0.7453258072752393, 0.744095094733162, 0.7677504725678777, 0.7864515287735565, 0.999999999999702, 0.7431921810450127, 0.7438058594911547, 0.7541653327826375, 0.7406714457219327, 0.7967104220911375, 0.7133575528810814, 0.7039810009927323, 0.765515923958494, 0.7680444088403934, 0.7776576264151995, 0.7443800071721435, 0.7038886609791768, 0.7664600915375828, 0.7743383342249517, 0.7743383342249517, 0.7182939450391324, 0.7386224424583132, 0.7700132064463462, 0.7743383342249517, 0.7182939450391324, 0.7563149858551899, 0.765515923958494, 0.7000488472482811, 0.7687455781879562, 0.7349576744695069, 0.7006064912951515, 0.7531908478721321, 0.7453258072752393, 0.8126174751631495, 0.708969884241688, 0.7662906020063978, 0.8286174576364165, 0.7862653235333383, 0.7359701104961036, 0.7286586855712416, 0.7204497866658688, 0.7231277152433704, 0.8271490921330018, 0.7149077254645313, 0.7609215022612039, 0.7070681641594352, 0.7358649058868536, 0.7615311813611817, 0.7206643730576503, 0.7216447650348548, 0.7202952993644187, 0.783415565255488, 0.7723505886060543, 0.7168593939811023] ['Jenn', 'Rebecca', 'Megan', 'CHRISTINA', 'STACY', 'Veronica', 'DIANE', 'TAMI', 'Susan', 'MONICA', 'CASEY', 'Fay', 'becca', 'BECKY', 'AMBER', 'LILA', 'Lauren', 'JENNIFER', 'Barbara', 'MEREDITH', 'Linda', 'Margaret', 'Nancy', 'Sabrina', 'Carole', 'Jocelyn', 'MICHELLE', 'MARY', 'Kayleigh', 'Jennifer', 'Elizabeth', 'JESSICA', 'Robyn', 'BRITTNEY', 'Amy', 'JANIS', 'Lorena', 'lizzie', 'Agnes', 'natalie', 'AMY', 'Mary', 'STACYS', 'Monica', 'FIONA', 'Patricia', 'Kate', 'VALERIE', 'jenny', 'Erin', 'Nancys', 'Rita', 'Leah', 'Ann', 'BARBARA', 'ELLA', 'DIANNE', 'Fiona', 'Jessica', 'Caroline', 'Suzanne', 'Michelle', 'SARAH', 'KELLY', 'Taylor', 'KALEY', 'Amanda', 'SAMANTHA', 'Jen', 'NANCY', 'SUSAN', 'Diane', 'Stacy', 'Julie', 'Sarah', 'REBECCA', 'nell', 'MEGAN', 'Barbra', 'Zoey', 'JANE', 'CAROLE', 'Kristy', 'Abby', 'Amber', 'KATHLEEN', 'Karen', 'karen', 'Yvonne', 'Cindy', 'KAREN', 'Tina', 'Becky', 'Jane', 'nicky', 'Candace', 'HALI', 'Rhianna', 'GINA', 'JEN', 'Melissa', 'Natalie', 'Letty', 'HAILEY', 'Katherine', 'Julianne', 'TAYLOR', 'BETTY', 'Jeffrey', 'Lacey', 'ELIZABETH', 'Kendall', 'Lori', 'JEFFERY', 'Katie', 'Tessa', 'Ellen', 'Riley', 'TRACY', 'VICKI', 'HEIDI']

karen [0.8657748004773197, 0.7453351341069008, 0.9040324605279073, 0.751670796095198, 0.8676009443318168, 0.8499135947044469, 0.7237491258609361, 0.725376969605873, 0.7285328411730558, 0.7713809646659535, 0.7308663182953796, 0.7743383321770485, 0.7837053958777226, 0.9221083642212833, 0.7231808290343011, 0.7376983035078151, 0.8450312105665077, 0.8723622983995808, 0.7507413009219133, 0.7028775447494193, 0.7058379612377795, 0.7069182858213889, 0.8098676614099112, 0.7218372366899417, 0.7964054887994914, 0.7465381588646032, 0.8375693613151758, 0.7206777970901332, 0.8428039519882055, 0.8391488032138339, 0.7058379612377795, 0.7770566318127541, 0.8420974323956756, 0.7009811905793248, 0.8778511984342434, 0.7165540883209185, 0.7067379324577909, 0.8223803619524617, 0.7206411709131194, 0.7166538487746155, 0.7235030217946908, 0.8879169808683844, 0.8351488411327372, 0.9239740182739906, 0.8759693677253402, 0.7987719034238541, 0.7051510400496737, 0.8101117546436145, 0.7726013516045369, 0.7177175144064907, 0.8600141257091769, 0.9020917124000106, 0.7069182858213889, 0.800308038777334, 0.7561009121546349, 0.8161464744488808, 0.7537924925808784, 0.7707070327508907, 0.8223803619524617, 0.7013211237869508, 0.7871913414813445, 0.7991733526128444, 0.9035390759320973, 0.8020808198790189, 0.8170577659646809, 0.8509947654514067, 0.9007400145805406, 0.7198267166353747, 0.7332187471248548, 0.8347325869329258, 0.8426226425796555, 0.777891549841633, 0.9007400145805406, 0.8151337654574146, 0.7231808290343011, 0.7537924925808784, 0.7144286521023506, 0.8098676614099112, 0.7904348931632951, 0.8785521556250719, 0.9081547046548102, 0.7892113798217182, 0.7363206024804735, 0.8588025327036511, 0.89933768227053, 0.7545080886838097, 0.7260739588270109, 0.7309817581235623, 0.8215695353606529, 0.7666459794592817, 0.7454750441698025, 0.8546678778709788, 0.7664536665136291, 0.8376743766488159, 0.8326018752091522, 0.8879169808683844,

0.7125880738001762, 0.8223024022843488, 0.7287042566024389, 0.8585559605485807,
0.7410689316949763, 0.7713809646659535, 0.7012782796639837, 0.834696837902086,
0.8509947654514067, 0.7244572750395628, 0.8229704436853235, 0.8785521556250719,
0.8252626674769508, 0.7353194620858436, 0.7664536665136291, 0.9035390759320973,
0.8433942190266206, 0.9084452771383716, 0.8161464744488808, 0.7217823540971975,
0.8068342201584967, 0.7404654676833258, 0.7678252323512804, 0.7112647389947785,
0.7152506100949514, 0.7334101614925537, 0.7735863745104521, 0.7985435107711707,
0.7639769912988015, 0.7024597924609377, 0.7058850442801743, 0.8371401257508331,
0.8480949015744459, 0.7755901733571766, 0.8751104999705825, 0.7317325736639213,
0.7013211237869508, 0.7069182858213889, 0.8825052469022198, 0.7030934158168345,
0.7975648643387783, 0.7063394481989649, 0.8758005883615961, 0.7987719034238541,
0.8723622983995808, 0.7146255340743879, 0.7112647389947785, 0.9221083642212833,
0.7856467089522705, 0.7743383321770485, 0.7139710293413722, 0.7825634466374198,
0.7084496488209991, 0.8786844698343765, 0.8068342201584967, 0.9040324605279073,
0.7614253508806015, 0.803552768927631, 0.8676009443318168, 0.8184023375110393,
0.7143468047681516, 0.7000239852938537, 0.7641725962345267, 0.8570980244893723,
0.7807333289182169, 0.725376969605873, 0.7455008457544745, 0.8420499976844236,
0.9020917124000106, 0.883363046906678, 0.7631224899997988, 0.759116544091599,
0.7931435495242003, 0.7543218439579292, 0.7770566318127541, 0.7009811905793248,
0.7012782796639837, 0.9087607931510366, 0.9999999999999754, 0.9999999999999754,
0.7140508454719192, 0.7030934158168345, 0.8377328423230004, 0.7208389961990498,
0.7360494034274383, 0.862343831078295, 0.7028292436293291, 0.9999999999999754,
0.8390943964232311, 0.720624434349656, 0.7588128833845497, 0.8391488032138339,
0.8420499976844236, 0.7423700922798301, 0.7831454917573035, 0.9001637288082609,
0.735045290809792, 0.7030199811431919, 0.7584983106724716, 0.7261126662710531,
0.7074032691147641, 0.7549603939373521, 0.8189288371972279, 0.7011203254771153,
0.8758005883615961, 0.9302962735807729, 0.7166538487746155, 0.777891549841633,
0.7549604290060965, 0.7561009121546349, 0.8728211747780041, 0.708071352245432,
0.7507413009219133, 0.8293972784939498, 0.7825634466374198, 0.7631224899997988,
0.8306934794817161, 0.7848553388343236, 0.8458500956163293, 0.8497249482529337,
0.7473044855914509, 0.7236673822028798, 0.7991733526128444, 0.705940303941278,
0.788474156294891, 0.7078370887660033, 0.7013211237869508, 0.7629726388489795,
0.7208389961990498, 0.7823414357213205, 0.8859559602713174, 0.7034280362975326,
0.7109941338350534, 0.8344914499990164, 0.8064710160072033, 0.8014108024540829,
0.7117694112721114, 0.8008454855799784, 0.927164328559422, 0.7983080112176714] ['Jenn',
'Olga', 'Rebecca', 'Mira', 'Megan', 'CHRISTINA', 'INA', 'EMMA', 'lass', 'PAM', 'Janeane', 'STACY',
'Veronica', 'DIANE', 'JO', 'Paige', 'TAMI', 'Susan', 'ruby', 'Maria', 'TIMOTHY', 'jeff', 'MONICA', 'CASEY',
'Elle', 'NANA', 'Fay', 'RINA', 'becca', 'BECKY', 'Timothy', 'AMBER', 'LILA', 'Dave', 'Lauren', 'LORRAINE',
'THORA', 'JENNIFER', 'ken', 'Lena', 'SASHA', 'Barbara', 'MEREDITH', 'Linda', 'Margaret', 'Nancy',
'Margery', 'Laila', 'Rosie', 'Madeleine', 'Sabrina', 'Carole', 'Jeff', 'Jocelyn', 'ALEXA', 'MICHELLE', 'MARY',
'Kayleigh', 'Jennifer', 'Mike', 'Lois', 'Elizabeth', 'JESSICA', 'Robyn', 'BRITTNEY', 'DONNA', 'Amy',
'Hellen', 'Lorena', 'lizzie', 'Agnes', 'natalie', 'AMY', 'BEA', 'Jo', 'Mary', 'STACYS', 'Monica', 'Greta',
'FIONA', 'Patricia', 'CHRISTY', 'ANDI', 'Kate', 'VALERIE', 'ESTHER', 'jenny', 'BILLIE', 'Erin', 'meg',
'Nancys', 'Rita', 'KYLIE', 'Leah', 'Ann', 'BARBARA', 'BRITTANY', 'ELLA', 'David', 'DIANNE', 'cate', 'Pam',
'STELLA', 'Mae', 'Donna', 'Lillith', 'EFFIE', 'Fiona', 'Sharon', '37863', 'Kylie', 'Jessica', 'Caroline',
'Suzanne', 'Michelle', 'Bitsy', 'SARAH', 'Sophie', 'Radhika', 'joey', 'MEL', 'GIOVANNA', 'KELLY', 'CAROL',
'DANI', 'DINGLING', 'Christiane', 'Joan', 'KALEY', 'CLEA', 'Amanda', 'VIV', 'mike', 'JEFF', 'SAMANTHA',
'Katy', 'Alice', 'VICTA', 'Jen', 'NANCY', 'SUSAN', 'Vanessa', 'Joey', 'Diane', 'Lise', 'Stacy',
'ÄfÄ¼bertragbar', 'KARENS', 'Junko', 'Julie', 'Sarah', 'REBECCA', 'Mari', 'nell', 'MEGAN', 'Barbra',
'Raven', 'PLOWMAN', 'Anita', 'NIA', 'Zoey', 'Emma', 'Jackie', 'JANE', 'CAROLE', 'Kristy', 'SERENA',
'Mima', 'Abby', 'JANU', 'Amber', 'DAVE', 'Stella', 'KATHLEEN', 'Karen', 'karen', 'trang', 'KATY', 'Yvonne',
'Winnie', 'Shirley', 'Cindy', 'Sharmila', 'KAREN', 'Tina', 'Skylar', 'grace', 'Becky', 'Jane', 'nicky', 'RIKO',
'Candace', 'Denises', 'Pearl', 'A00A', 'Rhianna', 'Lala', 'lucy', 'GINA', 'Dora', 'JEN', 'Melissa', 'LENA',
'Natalie', 'Lucy', 'Alexa', 'Letty', 'Terry', 'Ruby', 'HAILEY', 'Karens', 'Serena', 'Katherine', 'Julianne',
'PATTI', 'BETTY', 'Lacey', 'PATTY', 'ELIZABETH', 'REE', 'DREA', 'Roxanne', 'MIKE', 'Lori', 'WINNIE',
'Teresa', 'Katie', 'fred', 'Lyle', 'Tessa', 'Peggy', 'Ellen', 'Riley', 'TRACY', 'VICKI', 'HEIDI']

dishwasher [0.7483044899395241, 0.7922827614177599, 0.7394883690224285,
0.9999999999999742, 0.7260130570830599, 0.911224656343697, 0.8862124592049418,

0.7062039023728306, 0.7483044899395241, 0.7311683461853096, 0.7260130570830599, 0.7606226244291109, 0.7606226244291109, 0.8473303173295447, 0.7190536556339632, 0.7350140216150314, 0.778537066609295, 0.7138344818028683, 0.9999999999999742, 0.911224656343697, 0.778537066609295, 0.7380819088242276, 0.778537066609295, 0.7453965919449339, 0.7031556931266509, 0.7031556931266509, 0.7311683461853096, 0.7162767110109005, 0.7394883690224285, 0.7162767110109005, 0.7006412849810463, 0.7062039023728306, 0.911224656343697, 0.7453965919449339, 0.8862124592049418, 0.7138344818028683, 0.7162767110109005, 0.7138344818028683, 0.7242672477812984, 0.7684713666508316, 0.7606226244291109, 0.7538957024312753, 0.7896759290423555, 0.7453965919449339, 0.9999999999999742, 0.7394353348058125, 0.8473303173295447, 0.7315023660953542, 0.7483044899395241, 0.8473303173295447] ['Dryer', 'Plumbers', 'APPLIANCES', 'DISHWASHER', 'SINK', 'dishwashing', 'dishwashers', 'nonstick', 'DRYER', 'STOVE', 'sink', 'dishes', 'Dishes', 'WASHER', 'housekeeping', 'vacuumed', 'KITCHEN', 'CLEANING', 'Dishwasher', 'DISHWASHING', 'kitchen', 'cleaners', 'Kitchen', 'washing', 'LAUNDRY', 'laundry', 'stove', 'HOUSEWORK', 'appliances', 'Housework', 'TUPPERWARE', 'Nonstick', 'Dishwashing', 'Washing', 'DISHWASHERS', 'cleaning', 'housework', 'Cleaning', 'vacuuming', 'Spatula', 'DISHES', 'kitchens', 'appliance', 'WASHING', 'dishwasher', 'vacuums', 'washer', 'Plunger', 'dryer', 'Washer']

idiot [0.7938466101735703, 0.9999999999999838, 0.7623786338481204, 0.7150462688305701, 0.7529811181630761, 0.7476111687123439, 0.8115445939010283, 0.7220098084771714, 0.7507032291155666, 0.8019829280407518, 0.8503730001752827, 0.7195759884270634, 0.9999999999999838, 0.8767156499438133, 0.7041286835055587, 0.7039510439752632, 0.7132826150861942, 0.7195759884270634, 0.7479598987399223, 0.7150462688305701, 0.7079797571630628, 0.8115445939010283, 0.7081879599708254, 0.7036669313204911, 0.7225324420066008, 0.7007907772235253, 0.8160323473130681, 0.7193100083538038, 0.7059314178454396, 0.7041286835055587, 0.7729813781225948, 0.813952174643223, 0.8452272674984074, 0.7007907772235253, 0.7479598987399223, 0.7060518838526866, 0.7079797571630628, 0.7430699538291732, 0.8767156499438133, 0.7507032291155666, 0.838417309034255, 0.7328304221555721, 0.7081879599708254, 0.751844684797812, 0.8975358415383575, 0.7361770193094488, 0.7229920668094955, 0.97763480458979, 0.819079557290517, 0.7162213191269129, 0.7036669313204911, 0.7188428895199326, 0.7873897651242823, 0.8452272674984074, 0.7079797571630628, 0.7073211475309104, 0.8524157185119501, 0.7193100083538038, 0.7187100465109209, 0.8355059830959105, 0.8115445939010283, 0.7232983583392657, 0.706542129310848, 0.8829808376062461, 0.7851004992137283, 0.8503730001752827, 0.7115342842032175, 0.7264920833866337, 0.7036669313204911, 0.8711646092155545, 0.7034794934582234, 0.7851004992137283, 0.7162213191269129, 0.9999999999999838, 0.7521614161889978, 0.7273579926216062, 0.7039510439752632, 0.8355059830959105, 0.7476111687123439, 0.8503730001752827, 0.7873897651242823, 0.7361770193094488, 0.7521614161889978, 0.7873897651242823, 0.705059232073285, 0.8829808376062461, 0.7521614161889978, 0.7512123678073577, 0.7251245522453871, 0.7059314178454396, 0.7877723931218095, 0.7193100083538038, 0.9273044185432487, 0.7851004992137283, 0.8711646092155545, 0.8975358415383575, 0.7012014019274248, 0.9273044185432487, 0.7162213191269129, 0.7039510439752632, 0.8019829280407518, 0.7479598987399223, 0.8767156499438133, 0.7507032291155666, 0.7095929988991081, 0.7429987831201551, 0.7007907772235253, 0.751844684797812, 0.7344592844146971, 0.751844684797812, 0.7132826150861942, 0.8975358415383575, 0.7031506370545028, 0.8829808376062461, 0.819007196257272] ['STFU', 'idiot', 'Wannabe', 'OWNED', 'aufog', 'COWARD', 'ASSHOLE', 'GTFO', 'pathetic', 'LOSER', 'douchebag', 'hypocrite', 'IDIOT', 'retarded', 'lolz', 'dick', 'SCUM', 'Hypocrite', 'SILLY', 'owned', 'nonsense', 'asshole', 'TROLL', 'WHATEVER', 'FACEPALMED', 'intelligent', 'PRICK', 'Simp', 'Dummies', 'Lolz', 'PWNERD', 'DUMBEST', 'fool', 'Intelligent', 'silly', 'lmao', 'Nonsense', 'misguided', 'RETARDED', 'PATHETIC', 'Douche', 'BTCHES', 'troll', 'derp', 'DUMB', 'stupidity', 'childish', 'DUMBASS', 'DURR', 'NOOB', 'Whatever', 'jerk', 'IDIOTS', 'FOOL', 'ONSENSE', 'rude', 'foolish', 'simp', 'bigot', 'IDIOCY', 'Asshole', 'AREST', 'HURR', 'Stupid', 'fail', 'DOUCHEBAG', 'DUMBER', 'immature', 'whatever', 'dummy', 'MISINFORMED', 'Fail', 'Noob', 'Idiot', 'Fools', 'Mong', 'DICK', 'idiocy', 'coward', 'Douchebag', 'Idiots', 'STUPIDITY', 'fools', 'idiots', 'kiddo', 'STUPID', 'FOOLS', 'ignorantly', 'INSULTING', 'DUMMIES', 'twat', 'SIMP', 'Retard', 'FAIL', 'DUMMY', 'Dumb', 'roflmao', 'retard', 'noob', 'Dick', 'loser', 'Silly', 'Retarded', 'Pathetic', 'asspie', 'SHUTUP', 'INTELLIGENT', 'DERP', 'UNINFORMED', 'Derp', 'scum', 'dumb', 'INTELLEAGENT', 'stupid', 'IGNORANT']

asshole [0.8004224385779843, 0.7115844650123635, 0.7392936597424111, 0.7098877933556984, 0.811544595827216, 0.7049587347293753, 0.7379999315483311, 0.7518252134359571, 0.9999999999999809, 0.717847165618476, 0.7304709886430918, 0.7005762739559169, 0.7005245915401521, 0.7855683976836523, 0.8204197642124936, 0.9143113381642858, 0.7140234319453647, 0.7136855149499561, 0.8004224385779843, 0.7206872845757777, 0.811544595827216, 0.7062818350546229, 0.7643816236911846, 0.8829001435090187, 0.7420946442630256, 0.7727943672839823, 0.7011552645937786, 0.7140234319453647, 0.7078303221379849, 0.7079459908618173, 0.7379999315483311, 0.7085268001770683, 0.9999999999999809, 0.7606080828348114, 0.7297608841170353, 0.7563884989290146, 0.7022554859635856, 0.7050288173268298, 0.7524849029833607, 0.7005762739559169, 0.9177199958137755, 0.7002100837953595, 0.807903455314366, 0.7835056696603746, 0.7038116422300005, 0.7161515395675266, 0.7365059526112905, 0.7159843964414385, 0.7509023080211346, 0.7294720316349044, 0.7307135748040505, 0.7104601152197467, 0.7252070700032508, 0.7085268001770683, 0.7651854170827168, 0.7643816236911846, 0.7198689970188472, 0.7010570125045126, 0.7173653288427321, 0.7062818350546229, 0.7294720316349044, 0.7206872845757777, 0.7005245915401521, 0.9115851642018298, 0.7071719879720693, 0.708392375126023, 0.820790347983659, 0.7173210979036713, 0.7078008446933148, 0.7466988996935102, 0.7466988996935102, 0.7292451768301556, 0.7432754114108922, 0.788889464621253, 0.7212124939312694, 0.7067514203666238, 0.7812444444167106, 0.7044860452160489, 0.7132323076443549, 0.8415729474684907, 0.7305871711287547, 0.7106014397901828, 0.7089246614597166, 0.8921828092195708, 0.7097243249024079, 0.7045457792293968, 0.7307135748040505, 0.8078817715293378, 0.7046777745607802, 0.7011552255235851, 0.70657044916967, 0.7138373597603596, 0.7182440332113278, 0.7442405791823884, 0.7835056696603746, 0.703002479325882, 0.7067514311447409, 0.9999999999999809, 0.7011552230161021, 0.7338121034011311, 0.7106518713231869, 0.7036652222876916, 0.7219352521128322, 0.7102481301925724, 0.7185039939611718, 0.8204197703738627, 0.7045553472604791, 0.7261737462210002, 0.9143113381642858, 0.7509064218774707, 0.7006282008766596, 0.7306243855876474, 0.7212975288314887, 0.7100690576179449, 0.7171825065716171, 0.7045553472604791, 0.7067514203666238, 0.70657044916967, 0.811544595827216, 0.7233103552087212, 0.7067514203666238, 0.8829001435090187, 0.7501141074988782, 0.9143113381642858, 0.7546684168432497, 0.7622711338334811, 0.7098407583137065, 0.7809690041964663, 0.7328329592955225, 0.7261737569267491, 0.7643816236911846, 0.7963264367781012, 0.7071719879720693, 0.7031346087058046, 0.770255221796235, 0.7185039939611718, 0.7012471001632451, 0.7391766850930088, 0.7509023080211346, 0.7458692182630866, 0.794096091465196, 0.7835056696603746, 0.7659198273330867, 0.7261737462210002, 0.7100690576179449, 0.7181977734315292, 0.7501141074988782, 0.7045553472604791, 0.7509023080211346, 0.7196145783844465, 0.7306243855876474, 0.716122212669205, 0.7292451768301556, 0.7659198273330867, 0.7026615372012228, 0.7297608841170353, 0.7100690576179449, 0.7171825065716171, 0.7071719879720693, 0.703042225362926, 0.8829001435090187, 0.7855683976836523, 0.7546684168432497, 0.7062818350546229, 0.7005245915401521, 0.7196145783844465, 0.7171825065716171, 0.8204197642124936, 0.7220630559410839, 0.7284961735879226, 0.734251027664651, 0.7025303244078536, 0.7252070700032508, 0.7081547037519305, 0.7025182078820914, 0.7085268001770683, 0.7294720316349044, 0.7290578983621131, 0.7727943672839823, 0.8113492086899002, 0.7128236581557652, 0.7292451768301556, 0.7049769372816628, 0.7206872845757777, 0.7185039939611718, 0.70657044916967] ['Jerks', 'hurtful', 'STFU', 'FUCKERS', 'idiot', 'Wannabe', 'OWNED', 'aufog', 'ASSHOLE', 'GTFO', 'UNAPPRECIATIVE', 'skank', 'pathetic', 'LOSER', 'assholes', 'douchebag', 'hypocrite', 'petty', 'jerks', 'Cunt', 'IDIOT', 'retarded', 'BITCH', 'dick', 'FING', 'SCUM', 'screw', 'Hypocrite', 'FUC', 'stink', 'owned', 'SLUT', 'asshole', 'DOUCHEBAGS', 'Typical', 'cruel', 'PITY', 'fricken', 'POOPY', 'SKANK', 'PRICK', 'penisy', 'Inconsiderate', 'Simp', 'Bostock', 'bich', 'INSULTED', 'inocent', 'INSULT', 'SUCK', 'fool', 'jussayin', 'BITCHY', 'slut', 'Scumbag', 'Bitch', 'Chelsey', 'WEARHOU', 'poo', 'RETARDED', 'suck', 'cunt', 'PATHETIC', 'Douche', 'Shit', 'Ppxleyes.com', 'BTCHES', 'vindictive', 'PokÅfmoy', 'BASTARD', 'bastard', 'DUMB', 'Sucker', 'selfrighteous', 'spiteful', 'eViL', 'FUCKA', 'NICENESS', 'pisses', 'DUMBASS', 'DURR', 'vile', 'damned', 'jerk', 'MEANNESS', 'UNGRATEFUL', 'FOOL', 'rude', 'montags', 'Screw', 'HO', 'demeaning', 'HATEFUL', 'despicable', 'simp', 'FOX13', 'Evil', 'Asshole', 'SCREW', 'AREST', 'disrespectul', 'HURR', 'GOSH', 'Juvenile', 'Stupid', 'ASSHOLES', 'fail', 'Mean', 'DOUCHEBAG', 'immature', 'stinks', 'dummy', 'MOUTHY', 'DEAR', 'DICKS', 'Fail', 'evil', 'Ho', 'Idiot', '62460', 'EVIL', 'DICK', 'SHT', 'Douchebag', 'motherfucker', 'BTCH', 'Jeez', 'shitface', 'Wlove', 'MEAN', 'bitch', 'Biatch', 'shit', 'fcking', 'BITCHBOY', 'STUPID', 'BCH', 'mofo', 'insult', 'INSULTING', 'twat',

'SIMP', 'Retard', 'mean', 'Dear', 'CONSIDERATE', 'sht', 'FAIL', 'Insult', 'crap', 'DUMMY', 'nofucksgiven', 'Dumb', 'retard', 'unnecessary', 'TYPICAL', 'dear', 'dicks', 'SHIT', 'Sucky', 'Dick', 'loser', 'MOTHERFUCKER', 'Retarded', 'Pathetic', 'CRAP', 'Dicks', 'Assholes', 'asspie', 'SHUTUP', 'wifebeating', 'polite', 'bitchy', 'youyou', 'KINDEST', 'Slut', 'Suck', 'FU', 'scum', 'fucker', 'rmeme', 'dumb', 'whiny', 'CUNT', 'stupid', 'ho']

wife [0.711810734319952, 0.7234647045421287, 0.9999999999999702, 0.7198221361116212, 0.829424151406686, 0.7234647045421287, 0.7461894445965165, 0.7543599640083092, 0.829424151406686, 0.9999999999999702, 0.9523504639241169, 0.7465661555760856, 0.8921214728407451, 0.7326643867735451, 0.7178512908706187, 0.7543599640083092, 0.8773541459353161, 0.7356391218768007, 0.9999999999999702, 0.7466703781450066, 0.7431213800589755, 0.7184680160533518, 0.8581066965324429, 0.9523504639241169, 0.7431213800589755, 0.8921214728407451, 0.7178512908706187, 0.7198221361116212, 0.7070053968507586, 0.7356391218768007, 0.7198221361116212, 0.7178512908706187, 0.7234647045421287, 0.7184680160533518, 0.7225183536007532, 0.7137972207633698, 0.8773334692188388, 0.7830468678497436, 0.7225183536007532, 0.8026137155804567, 0.7213796578806431, 0.7116936904164041, 0.8176735290851429, 0.8444367584248893, 0.7465661555760856, 0.8366420852178453, 0.7326643867735451, 0.7184680160533518, 0.7356391218768007, 0.7431213800589755, 0.7466703781450066, 0.8921214728407451, 0.8444367584248893, 0.7225183536007532, 0.8762357811699788, 0.7392520828663534, 0.8773334692188388, 0.8581066965324429, 0.7908088182839577, 0.8997116019052318, 0.7461894445965165, 0.8444367584248893, 0.7351556811945084, 0.9523504639241169, 0.829424151406686, 0.7466703781450066, 0.7461894445965165, 0.8773541459353161, 0.7543599640083092, 0.8366420852178453, 0.8581066965324429, 0.9999999999999702, 0.7351556811945084] ['HOMEMAKER', 'divorce', 'wife', 'MOTHERS', 'girlfriend', 'DIVORCE', 'DIVORCED', 'MOTHER', 'Girlfriend', 'Wife', 'Wifes', 'bride', 'Husband', 'HERS', 'DAUGHTER', 'Mother', 'SPOUSE', 'Housewife', 'W.I.F.E.', 'MARRIES', 'girlfriends', 'WOMAN', 'Married', 'WIFES', 'Girlfriends', 'HUSBAND', 'daughter', 'Mothers', 'Bachelor', 'housewife', 'mothers', 'Daughter', 'Divorce', 'Woman', 'boyfriend', 'inlaws', 'WIVES', 'Sisterinlaw', 'BOYFRIEND', 'exwife', 'REMARRIED', 'girlfriend', 'SPOUSAL', 'Husbands', 'BRIDE', 'SPOUSES', 'hers', 'woman', 'HOUSEWIFE', 'GIRLFRIENDS', 'marries', 'husband', 'HUSBANDS', 'Boyfriend', 'WIFED', 'SONINLAW', 'wives', 'married', 'widowed', 'FIANCEE', 'divorced', 'husbands', 'womans', 'wives', 'GIRLFRIEND', 'Marries', 'Divorced', 'spouse', 'mother', 'spouses', 'MARRIED', 'WIFE', 'WOMANS']

husband relazioni [0.7545271498155636, 0.7065540143212166, 0.8921214671544195, 0.7284289739259531, 0.7088054101257736, 0.7439258739054844, 0.7204548826430407, 0.7284289739259531, 0.7668550674571684, 0.8921214671544195, 0.869869278067202, 0.9999999999999736, 0.7180028847923617, 0.710256916972644, 0.7065540143212166, 0.8905465168117572, 0.7148170503013396, 0.8921214671544195, 0.7368934017810591, 0.8437789630435151, 0.7448069996245565, 0.869869278067202, 0.9999999999999736, 0.710256916972644, 0.7324682502190043, 0.7148170503013396, 0.710256916972644, 0.7324682502190043, 0.8035341968527585, 0.7225936325815048, 0.7667698730097724, 0.7204548826430407, 0.789963520260045, 0.8035341968527585, 0.7338554622070312, 0.7180028847923617, 0.7379757192967903, 0.7109731329432953, 0.8242702321677132, 0.9114540997940981, 0.8276937628258687, 0.7324682502190043, 0.7072257799513003, 0.7148170503013396, 0.7368934017810591, 0.9999999999999736, 0.7448069996245565, 0.9114540997940981, 0.8035341968527585, 0.7088054083692723, 0.878566172845189, 0.7109731329432953, 0.7781436567459139, 0.7667698730097724, 0.8437789630435151, 0.7204548826430407, 0.7109731329432953, 0.7625305191456329, 0.7180028847923617, 0.8929240640400291, 0.7439258739054844, 0.9114540997940981, 0.869869278067202, 0.7284289739259531, 0.7368934017810591, 0.7439258739054844, 0.8905465168117572, 0.8276937628258687, 0.8437789630435151, 0.7072257799513003, 0.8921214671544195, 0.716548868427843, 0.7637009024177338] ['HOMEMAKER', 'ENGAGED', 'wife', 'girlfriend', 'partner', 'DIVORCED', 'dads', 'Girlfriend', 'stepson', 'Wife', 'Wifes', 'Husband', 'boyfriends', 'DAUGHTER', 'engaged', 'SPOUSE', 'Housewife', 'W.I.F.E.', 'MARRIES', 'Married', 'stepdad', 'WIFES', 'HUSBAND', 'daughter', 'FATHER', 'housewife', 'Daughter', 'father', 'boyfriend', 'inlaws', 'WIVES', 'Dads', 'Sisterinlaw', 'BOYFRIEND', 'exwife', 'Boyfriends', 'REMARRIED', 'fathers', 'SPOUSAL', 'Husbands', 'SPOUSES', 'Father', 'MRS', 'HOUSEWIFE', 'marries', 'husband', 'STEPDAD', 'HUSBANDS', 'Boyfriend', 'PARTNER', 'WIFED', 'FATHERS', 'SONINLAW', 'wives', 'married', 'DADS', 'Fathers', 'widowed', 'BOYFRIENDS',

'FIANCEE', 'divorced', 'husbands', 'wives', 'GIRLFRIEND', 'Marries', 'Divorced', 'spouse', 'spouses', 'MARRIED', 'Mrs', 'WIFE', 'boyfriendgirlfriend', 'stepfather']

housewife [0.7632070104639275, 0.9000538495115378, 0.735639140903112, 0.7784919913661618, 0.7632070104639275, 0.7954501823896523, 0.735639140903112, 0.7522185421977066, 0.7177225875371508, 0.714817052742852, 0.7495430121840837, 0.7120786613641381, 0.9999999999999725, 0.735639140903112, 0.7503450836886114, 0.7439377988179334, 0.7522185421977066, 0.714817052742852, 0.7884578803094651, 0.7583273652218989, 0.772944407739217, 0.9999999999999725, 0.7027094084306571, 0.7095686951352502, 0.7027094084306571, 0.7029890310573225, 0.7206494936866537, 0.7177225875371508, 0.7075477684687614, 0.7027094084306571, 0.856389688682043, 0.7049454488564371, 0.856389688682043, 0.9999999999999725, 0.7784919913661618, 0.714817052742852, 0.7206494936866537, 0.7207763663988962, 0.7099749730058277, 0.7754330101103021, 0.7144491498219326, 0.7206494936866537, 0.7522185421977066, 0.7075477684687614, 0.7599640978551083, 0.856389688682043, 0.7599640978551083, 0.7049454488564371, 0.7075477684687614, 0.735639140903112, 0.7583273652218989] ['STEPMOM', 'HOMEMAKER', 'wife', 'Waitress', 'stepmom', 'businesswoman', 'Wife', 'Wives', 'bride', 'Husband', 'supervising', 'nursing', 'Housewife', 'W.I.F.E.', 'housekeeping', 'Lorena', 'WIFES', 'HUSBAND', 'stayathome', 'MAIDS', 'Bachelor', 'housewife', 'KITCHEN', 'Apprentice', 'kitchen', 'SPOUSAL', 'Husbands', 'BRIDE', 'CHEF', 'Kitchen', 'maid', 'SPOUSES', 'MAID', 'HOUSEWIFE', 'WAITRESS', 'husband', 'HUSBANDS', 'WIFED', 'WORKPLACE', 'widowed', 'FIANCEE', 'husbands', 'wives', 'chef', 'babysitter', 'Maid', 'Babysitter', 'spouses', 'Chef', 'WIFE', 'maids']

girlfriend [0.7529701309659291, 0.8294241531730182, 0.7137395046102304, 0.9999999999999741, 0.7121719448771529, 0.701291659006963, 0.8497820313958924, 0.8023320574407993, 0.9999999999999741, 0.8294241531730182, 0.7902472051225226, 0.7103964599656433, 0.7284289878879464, 0.8491365188074236, 0.7338031126346711, 0.7084262529365626, 0.7072897981952877, 0.7114575567503523, 0.8294241531730182, 0.9139957404535017, 0.7200131612735514, 0.7200131612735514, 0.7439764012603145, 0.7902472051225226, 0.7025680747182481, 0.9139957404535017, 0.7284289878879464, 0.7084262529365626, 0.7137395046102304, 0.8545458891813603, 0.8881752114939719, 0.7072897981952877, 0.7209796622317215, 0.7089760379785628, 0.8881752114939719, 0.7686240743380507, 0.8491365188074236, 0.8821740902839117, 0.7215984692138364, 0.7025680747182481, 0.7050825004884393, 0.7338031126346711, 0.7072897981952877, 0.9139957404535017, 0.7284289878879464, 0.7400006005482468, 0.7103964599656433, 0.8881752114939719, 0.7121719367362082, 0.8545458891813603, 0.8017561012163441, 0.7529701309659291, 0.7209796622317215, 0.7439764012603145, 0.7050825004884393, 0.8491365188074236, 0.7200131612735514, 0.8567863112531593, 0.7902472051225226, 0.9999999999999741, 0.7400006005482468, 0.7529701309659291, 0.8497820313958924, 0.7114575567503523, 0.7400006005482468, 0.7439764012603145, 0.8294241531730182, 0.822844913157119] ['RELATIONSHIP', 'wife', 'gf', 'girlfriend', 'partner', 'Ménage', 'exgirlfriend', 'girlfiend', 'Girlfriend', 'Wife', 'Wives', 'RELATIONSHIPS', 'Husband', 'boyfriends', 'HERS', 'breakup', 'Herself', 'SPOUSE', 'W.I.F.E.', 'girlfriends', 'Bf', 'BF', 'Married', 'WIFES', 'COUPLES', 'Girlfriends', 'HUSBAND', 'BREAKUP', 'GF', 'GFS', 'boyfriend', 'HERSELF', 'WIVES', 'Sisterinlaw', 'BOYFRIEND', 'exwife', 'Boyfriends', 'girlfiend', 'SPOUSAL', 'Couples', 'Whipped', 'hers', 'herself', 'GIRLFRIENDS', 'husband', 'Single', 'relationships', 'Boyfriend', 'PARTNER', 'GFs', 'WIFED', 'relationship', 'wives', 'married', 'WHIPPED', 'BOYFRIENDS', 'bf', 'FIANCEE', 'wives', 'GIRLFRIEND', 'SINGLE', 'Relationship', 'EXGIRLFRIEND', 'spouse', 'single', 'MARRIED', 'WIFE', 'boyfriendgirlfriend']

boyfriend [0.7233637683442151, 0.752791444284532, 0.7225183640353565, 0.888175213752568, 0.7432257221644417, 0.714493953371228, 0.7679443270793412, 0.7843679606265266, 0.888175213752568, 0.7225183640353565, 0.8035342085869761, 0.9345113151480839, 0.7225183640353565, 0.7984132373379567, 0.7984622129418755, 0.7984622129418755, 0.7137678849040567, 0.7293851169050461, 0.7984132373379567, 0.8035342085869761, 0.750140724808319, 0.7044056753032963, 0.9999999999999696, 0.9999999999999696, 0.9345113151480839, 0.82263984693986, 0.7065827123489969, 0.7293851169050461, 0.7338827074267892, 0.7984132373379567, 0.8035342085869761, 0.7484166518252371, 0.7338827074267892, 0.9999999999999696, 0.7432257255302976, 0.750140724808319,

0.7877392424234726, 0.752791444284532, 0.7010517690487371, 0.7137678849040567, 0.9345113151480839, 0.7984622129418755, 0.8173317549206455, 0.7338827074267892, 0.888175213752568, 0.7233637683442151, 0.7484166518252371, 0.752791444284532, 0.7679443270793412, 0.7484166518252371, 0.7137678849040567, 0.7225183640353565, 0.8777812224140964] ['bestfriend', 'RELATIONSHIP', 'wife', 'girlfriend', 'partner', 'Menage', 'exgirlfriend', 'girlfiend', 'Girlfriend', 'Wife', 'Husband', 'boyfriends', 'W.I.F.E.', 'girlfriends', 'Bf', 'BF', 'Married', 'COUPLES', 'Girlfriends', 'HUSBAND', 'GFS', 'FREIND', 'boyfriend', 'BOYFRIEND', 'Boyfriends', 'girfriend', 'SPOUSAL', 'Couples', 'Husbands', 'GIRLFRIENDS', 'husband', 'Single', 'HUSBANDS', 'Boyfriend', 'PARTNER', 'GFs', 'WIFED', 'relationship', 'sexs', 'married', 'BOYFRIENDS', 'bf', 'FIANCEE', 'husbands', 'GIRLFRIEND', 'BESTFRIEND', 'SINGLE', 'Relationship', 'EXGIRLFRIEND', 'single', 'MARRIED', 'WIFE', 'boyfriendgirlfriend']

mother [0.8012904435317071, 0.832846458263756, 0.7543599591017797, 0.8734871647341412, 0.7787588062075013, 0.8096365013735608, 0.8176637837393945, 0.832846458263756, 0.771220611302798, 0.7219720616397678, 0.7327276814059933, 0.7094351328073771, 0.8116818945732137, 0.7010474239195434, 0.8176637837393945, 0.7894992027829948, 0.999999999999745, 0.7560733013620335, 0.7543599591017797, 0.76823301006622, 0.7050720651881153, 0.8104285998121853, 0.873033741488509, 0.7601747849267264, 0.999999999999745, 0.8421912032000514, 0.9219210417498012, 0.8159347205024414, 0.7543599591017797, 0.7049689923193081, 0.7146598910300845, 0.8936974771759263, 0.8033503948616036, 0.8104285998121853, 0.7050720651881153, 0.873033741488509, 0.8637090189085759, 0.8290869052709772, 0.7135322023095654, 0.8492862779221599, 0.8734871647341412, 0.8096365013735608, 0.7787588062075013, 0.8734871647341412, 0.873033741488509, 0.8492862779221599, 0.8290869052709772, 0.9219210417498012, 0.8936974771759263, 0.76823301006622, 0.7655497080812681, 0.7369207859900089, 0.7894992027829948, 0.7638170738738945, 0.7459597996389837, 0.8637090189085759, 0.9219210417498012, 0.7601747849267264, 0.8176637837393945, 0.8159347205024414, 0.7817271789640224, 0.7472353167884082, 0.769301355717815, 0.717959734888235, 0.8492862779221599, 0.8051146805107303, 0.7459597996389837, 0.7585839943414365, 0.7094351328073771, 0.8208537427871394, 0.8051146805107303, 0.7408688741992503, 0.8033503948616036, 0.7135322023095654, 0.769301355717815, 0.7248659462597917, 0.8936974771759263, 0.8051146805107303, 0.7817271789640224, 0.8290869052709772, 0.8637090189085759, 0.7894992027829948, 0.769301355717815, 0.7186707159634058, 0.8116818945732137, 0.7459597996389837, 0.7817271789640224, 0.7050720651881153, 0.7249806204294069, 0.7327276814059933, 0.7472353167884082, 0.7249806204294069, 0.8421912032000514, 0.8096365013735608, 0.7219720616397678, 0.8709903232708641, 0.771220611302798, 0.8496478929276874, 0.771220611302798, 0.7219720616397678, 0.7968198145955927, 0.999999999999745, 0.76823301006622, 0.7244971744472266, 0.7585839943414365, 0.8421912032000514, 0.7369207859900089, 0.7327276814059933, 0.8208537427871394, 0.7724622004827971, 0.7787588062075013, 0.7543599591017797, 0.7249806204294069, 0.8049095558356628, 0.7135322023095654] ['parent', 'STEPMOM', 'wife', 'MOTHERS', 'daughters', 'Son', 'DADDY', 'stepmom', 'SISTER', 'GRANDFATHER', 'nephew', 'grandmas', 'Grandmother', 'NANA', 'daddy', 'dads', 'MOTHER', 'stepson', 'Wife', 'Dad', 'Wifes', 'MAMAS', 'DAUGHTER', 'daddys', 'Mother', 'mama', 'Mom', 'granddaughter', 'W.I.F.E.', 'Lorena', 'EENY', 'Mum', 'stepdad', 'mamas', 'WIFES', 'daughter', 'MOMS', 'AUNT', 'Child', 'FATHER', 'Mothers', 'SON', 'DAUGHTERS', 'mothers', 'Daughter', 'father', 'aunt', 'mom', 'mum', 'DAD', 'STEPFATHERS', 'Sis', 'Dads', 'Sisterinlaw', 'grandma', 'Moms', 'MOM', 'DADDYS', 'Daddy', 'Granddaughter', 'uncle', 'granny', 'fathers', 'grandfathers', 'Father', 'mommy', 'Grandma', 'MRS', 'Grandmas', 'mums', 'Mommy', 'grandson', 'STEPDAD', 'child', 'FATHERS', 'SONINLAW', 'MUM', 'MOMMY', 'Uncle', 'Aunt', 'moms', 'DADS', 'Fathers', 'widowed', 'grandmother', 'GRANDMA', 'UNCLE', 'wives', 'sons', 'Nephew', 'GRANNY', 'Sons', 'MAMA', 'son', 'grandfather', 'MOMMA', 'sister', 'auntie', 'Sister', 'Grandfather', 'MOOOOM', 'mother', 'dad', 'childs', 'Mrs', 'Mama', 'sis', 'NEPHEW', 'MUMS', 'aunts', 'Daughters', 'WIFE', 'SONS', 'stepfather', 'CHILD']

father [0.7698088039772302, 0.7056979159833143, 0.7368426236079686, 0.7125921962594325, 0.9116304129160753, 0.7690616755689864, 0.7056979159833143, 0.7947569250569015, 0.8175007575585365, 0.7690616755689864, 0.8126464424194629, 0.8492862765069459, 0.7800438470726365, 0.8574717552559961, 0.7324682532817, 0.8003804974081377, 0.8492862765069459, 0.7810019467846703, 0.7089063489417721, 0.742417006729146,

0.8329572556154048, 0.7324682532817, 0.8003804974081377, 0.7217579195733914,
0.741287059795982, 0.7036654884463258, 0.999999999999974, 0.7368426236079686,
0.9116304129160753, 0.7125921962594325, 0.7368426236079686, 0.8003804974081377,
0.999999999999974, 0.741287059795982, 0.7810019467846703, 0.742417006729146,
0.8574717552559961, 0.7441626805930497, 0.8126464424194629, 0.7217579195733914,
0.7810019467846703, 0.7690616755689864, 0.7089063489417721, 0.8516415270207238,
0.81296095099054, 0.718156880505008, 0.7317958178557864, 0.999999999999974,
0.7317958178557864, 0.7324682532817, 0.8063549589884691, 0.8329572556154048,
0.7036654884463258, 0.81296095099054, 0.742417006729146, 0.8516415270207238,
0.741287059795982, 0.7217579195733914, 0.8126464424194629, 0.81296095099054,
0.7317958178557864, 0.8516415270207238, 0.7669561659570825, 0.8175007575585365,
0.7669561659570825, 0.9116304129160753, 0.7947569250569015, 0.7947569250569015,
0.8492862765069459, 0.8574717552559961, 0.8175007575585365, 0.7125921962594325,
0.7669561659570825, 0.8444660827296759, 0.7036654884463258] ['parent', 'STEPMOM',
'MOTHERS', 'daughters', 'Son', 'DADDY', 'stepmom', 'GRANDFATHER', 'nephew', 'daddy', 'dads',
'MOTHER', 'stepson', 'Dad', 'Husband', 'DAUGHTER', 'Mother', 'Mom', 'granddaughter', 'Mum',
'stepdad', 'HUSBAND', 'daughter', 'MOMS', 'AUNT', 'Child', 'FATHER', 'Mothers', 'SON', 'DAUGHTERS',
'mothers', 'Daughter', 'father', 'aunt', 'mom', 'mum', 'DAD', 'STEPFATHERS', 'Dads', 'Moms', 'MOM',
'Daddy', 'Granddaughter', 'uncle', 'fathers', 'grandfathers', 'brother', 'Father', 'BROTHER', 'husband',
'grandson', 'STEPPAD', 'child', 'FATHERS', 'MUM', 'Uncle', 'Aunt', 'moms', 'DADS', 'Fathers', 'Brother',
'UNCLE', 'sons', 'Nephew', 'Sons', 'son', 'grandfather', 'Grandfather', 'mother', 'dad', 'NEPHEW',
'Daughters', 'SONS', 'stepfather', 'CHILD']

daughter [0.7608738118956022, 0.802974308857674, 0.7155147961546491, 0.7006374480153639,
0.7178512890825146, 0.7853654450234999, 0.9028576797478006, 0.8429466770005554,
0.797433742028358, 0.802974308857674, 0.8516180287205685, 0.8475700404060218,
0.7076405914581654, 0.7006374480153639, 0.7195707300983152, 0.797433742028358,
0.7256059622245477, 0.7301093616046194, 0.8730337376074058, 0.7636980978076677,
0.7662899688246293, 0.8246261396338931, 0.7178512890825146, 0.7102569171064212,
0.7472615347013742, 0.7155147961546491, 0.7204575868161809, 0.9999999999999758,
0.768314160898462, 0.8730337376074058, 0.7006374480153639, 0.7664856909252341,
0.711415887760676, 0.7692193006352888, 0.9003500393002857, 0.7178512890825146,
0.7798633154589736, 0.7665760777894945, 0.7479895926784389, 0.7472615347013742,
0.7102569171064212, 0.9999999999999758, 0.7459303944013058, 0.7926934841496215,
0.7530060333871859, 0.8003804991629353, 0.7853654450234999, 0.8429466770005554,
0.9028576797478006, 0.7853654450234999, 0.9999999999999758, 0.8003804991629353,
0.7926934841496215, 0.7692193006352888, 0.7798633154589736, 0.8006522951087324,
0.7792899551496101, 0.7301093616046194, 0.7933194618733426, 0.7459303944013058,
0.7692193006352888, 0.768314160898462, 0.797433742028358, 0.9003500393002857,
0.7452560866449143, 0.7472796959995507, 0.754252931123564, 0.746121326022599,
0.8003804991629353, 0.737405488346566, 0.7204575868161809, 0.7250322907018802,
0.7344632303532157, 0.737405488346566, 0.7102569171064212, 0.8227250591856322,
0.7479895926784389, 0.7018612568155931, 0.7530060333871859, 0.7472796959995507,
0.7707203581248657, 0.7006374480153639, 0.7798633154589736, 0.7260185585490518,
0.737405488346566, 0.7155147961546491, 0.7452560866449143, 0.7926934841496215,
0.7459303944013058, 0.7301093616046194, 0.7472796959995507, 0.7189033190000563,
0.754252931123564, 0.7195707300983152, 0.7452560866449143, 0.7980898298923648,
0.7662899688246293, 0.8475700404060218, 0.7980898298923648, 0.7664856909252341,
0.8429466770005554, 0.7636980978076677, 0.7662899688246293, 0.7736588677835818,
0.8516180287205685, 0.8188461635728885, 0.8516180287205685, 0.746121326022599,
0.7500734851034032, 0.8730337376074058, 0.7076405914581654, 0.7804174660889583,
0.7250322907018802, 0.7664856909252341, 0.7792899551496101, 0.8475700404060218,
0.7344632303532157, 0.7488041387314884, 0.9028576797478006, 0.7178512890825146,
0.7980898298923648, 0.8040507042092037, 0.7530060333871859, 0.7260185585490518] ['parent',
'STEPMOM', 'boy', 'MILF', 'wife', 'MOTHERS', 'daughters', 'Son', 'DADDY', 'stepmom', 'SISTER',
'nephew', 'BAZZINGA', 'M.I.L.F', 'Grandmother', 'daddy', 'siblings', 'dads', 'MOTHER', 'sibling', 'Sisters',
'stepson', 'Wife', 'Husband', 'MAMAS', 'Boy', 'HERS', 'DAUGHTER', 'daddys', 'Mother', 'milf', 'mama',
'mummys', 'Mom', 'granddaughter', 'W.I.F.E.', 'Mum', 'youngest', 'stepdad', 'mamas', 'HUSBAND',
'daughter', 'MOMS', 'AUNT', 'Child', 'FATHER', 'Mothers', 'SON', 'DAUGHTERS', 'mothers', 'Daughter',

'father', 'aunt', 'mom', 'mum', 'STEPFATHERS', 'Sis', 'Dads', 'Sisterinlaw', 'Moms', 'MOM', 'DADDYS', 'Daddy', 'Granddaughter', 'uncle', 'fathers', 'stepbrother', 'adopted', 'Father', 'mommy', 'hers', 'MRS', 'mums', 'Mommy', 'husband', 'grandson', 'STEPPAD', 'newborn', 'child', 'FATHERS', 'SONINLAW', 'Milf', 'MUM', 'cousin', 'MOMMY', 'BOY', 'Uncle', 'Aunt', 'moms', 'DADS', 'Fathers', 'widowed', 'STEPBROTHER', 'grandmother', 'UNCLE', 'sons', 'SISTERS', 'Nephew', 'Sons', 'MAMA', 'son', 'SIBLING', 'sisters', 'MOMMA', 'sister', 'auntie', 'Sister', 'ADOPTED', 'MOOOOM', 'mother', '4YEAROLD', 'childs', 'Mrs', 'Mama', 'sis', 'NEPHEW', 'MUMS', 'aunts', 'Daughters', 'WIFE', 'SONS', 'stepfather', 'CHILD', 'COUSIN']

son [0.7467518799059971, 0.7567096934112802, 0.9999999999999714, 0.7830748889050924, 0.7223193652845012, 0.7305929893219485, 0.8637292915122383, 0.7830748889050924, 0.7770412493010239, 0.8096364978997641, 0.7909088653969543, 0.8078279311141123, 0.7467518799059971, 0.8429466703682442, 0.715366065752433, 0.8096364978997641, 0.7350311383722963, 0.7457033628224046, 0.7265927971647681, 0.7322411945215271, 0.7808392509025464, 0.8429466703682442, 0.7090378228168628, 0.7307562876098324, 0.91163040850286, 0.9999999999999714, 0.7567096934112802, 0.8429466703682442, 0.91163040850286, 0.7090378228168628, 0.7457033628224046, 0.7322411945215271, 0.8078279311141123, 0.731062433839749, 0.7770412493010239, 0.7457033628224046, 0.715366065752433, 0.7830748889050924, 0.7265927971647681, 0.8198525860276226, 0.7570137611510881, 0.7674239522663802, 0.7331282454712796, 0.91163040850286, 0.7674239522663802, 0.8230287495912907, 0.7808392509025464, 0.7307562876098324, 0.7570137611510881, 0.7322411945215271, 0.7220917306693238, 0.7467518799059971, 0.8198525860276226, 0.7090378228168628, 0.7770412493010239, 0.7570137611510881, 0.7331282454712796, 0.7674239522663802, 0.8198525860276226, 0.8274570476716365, 0.8637292915122383, 0.8274570476716365, 0.7350311383722963, 0.9999999999999714, 0.7305929893219485, 0.7194689269156139, 0.7223193652845012, 0.7223193652845012, 0.7305929893219485, 0.7120754421654281, 0.8096364978997641, 0.8078279311141123, 0.7282846999145399, 0.7350311383722963, 0.8637292915122383, 0.7567096934112802, 0.8274570476716365, 0.7996824117613999, 0.7307562876098324, 0.7220917306693238] ['boy', 'daughters', 'Son', 'DADDY', 'SISTER', 'GRANDFATHER', 'nephew', 'daddy', 'dads', 'MOTHER', 'stepson', 'Dad', 'Boy', 'DAUGHTER', 'daddys', 'Mother', 'mama', 'Mom', 'granddaughter', 'Mum', 'stepdad', 'daughter', 'AUNT', 'Child', 'FATHER', 'SON', 'DAUGHTERS', 'Daughter', 'father', 'aunt', 'mom', 'mum', 'DAD', 'STEPFATHERS', 'Dads', 'MOM', 'DADDYS', 'Daddy', 'Granddaughter', 'uncle', 'fathers', 'brother', 'stepbrother', 'Father', 'BROTHER', 'grandson', 'STEPPAD', 'child', 'FATHERS', 'MUM', 'cousin', 'BOY', 'Uncle', 'Aunt', 'DADS', 'Fathers', 'STEPBROTHER', 'Brother', 'UNCLE', 'sons', 'Nephew', 'Sons', 'MAMA', 'son', 'grandfather', 'MOMMA', 'sister', 'Sister', 'Grandfather', 'MOOOOM', 'mother', 'dad', 'childs', 'Mama', 'NEPHEW', 'Daughters', 'SONS', 'stepfather', 'CHILD', 'COUSIN']

parent [0.9999999999999312, 0.7728684715240043, 0.7146096229136755, 0.7086339100957072, 0.7418643487909674, 0.8019666851097031, 0.7279378929156084, 0.7910148826420531, 0.7728684715240043, 0.7012516870313825, 0.7278659060066068, 0.7459898783586096, 0.7273383244290206, 0.7910148826420531, 0.7960636275390629, 0.8012904352992452, 0.7351942590329666, 0.79244897313876, 0.7244802155617898, 0.7552786362130843, 0.7305610783456402, 0.7608738317575886, 0.7879170137662974, 0.8012904352992452, 0.7178292268648654, 0.7674968366212722, 0.7943968175555295, 0.7388310902817936, 0.7705777275454497, 0.7451700846092062, 0.7245502621747686, 0.8231676226873422, 0.7552786362130843, 0.7608738317575886, 0.7952352971252371, 0.7595563150621156, 0.7084625197582639, 0.7690948084093878, 0.8664828251501779, 0.7698087938438426, 0.8019666851097031, 0.741836978103313, 0.7279378929156084, 0.741836978103313, 0.7858188896774987, 0.8019666851097031, 0.7608738317575886, 0.7120153482162142, 0.7422391543384274, 0.7698087938438426, 0.7084625197582639, 0.7943968175555295, 0.7451700846092062, 0.7244802155617898, 0.8501563800279339, 0.7960636275390629, 0.7445805675546721, 0.8664828251501779, 0.7952352971252371, 0.7943968175555295, 0.7879170137662974, 0.7910148826420531, 0.7388310902817936, 0.7277387858350217, 0.7782024987312485, 0.7238437500084836, 0.7157547624554715, 0.7023761035923848, 0.7176737668415364, 0.7698087938438426, 0.7901660747103786, 0.7674968366212722, 0.7106443124701183, 0.7901660747103786, 0.7086339100957072, 0.7089013797860754, 0.8231676226873422, 0.7690948084093878, 0.7782024987312485, 0.7338992386326555, 0.7451700846092062, 0.8664828251501779, 0.7318893448029009, 0.7901660747103786,

0.7168247080385189, 0.7277387858350217, 0.7495155673784013, 0.7084625197582639,
0.7086339100957072, 0.7952352971252371, 0.7960636275390629, 0.7782024987312485,
0.7157547624554715, 0.7318893448029009, 0.7277387858350217, 0.7168247080385189,
0.7236536845033905, 0.7278659060066068, 0.7176737668415364, 0.7351942590329666,
0.7012516870313825, 0.7477111699534724, 0.7253458478248174, 0.7184550716233304,
0.7674968366212722, 0.7023761035923848, 0.7273383244290206, 0.7012516870313825,
0.7402174717958797, 0.8012904352992452, 0.7236536845033905, 0.7459898783586096,
0.7244802155617898, 0.7176737668415364, 0.7602925698082081, 0.7278659060066068,
0.7106443124701183, 0.7486243435213918, 0.7279378929156084, 0.7495155673784013,
0.7318893448029009, 0.848393745920269, 0.7690948084093878] ['parent', 'STEPMOM', 'families',
'Adult', 'grandparents', 'MOTHERS', 'daughters', 'DADDY', 'stepmom', 'GRANDFATHER', 'nephew',
'BAZZINGA', 'TEENAGER', 'daddy', 'dads', 'MOTHER', 'sibling', 'stepson', 'Dad', 'MAMAS', 'supervised',
'DAUGHTER', 'daddys', 'Mother', 'supervising', 'kid', 'Mom', 'granddaughter', 'EENY', 'Mum', 'youngest',
'stepdad', 'mamas', 'daughter', 'MOMS', 'HOUSEHOLD', 'AUNT', 'Child', 'Parents', 'FATHER', 'Mothers',
'12yo', 'DAUGHTERS', '12YO', 'misbehaved', 'mothers', 'Daughter', 'intergenerational', 'parenting',
'father', 'aunt', 'mom', 'mum', 'DAD', 'STEPFATHERS', 'Dads', 'Sisterinlaw', 'PARENTS', 'Moms', 'MOM',
'DADDYS', 'Daddy', 'Granddaughter', 'uncle', 'fathers', 'grandfathers', 'stepbrother', 'adopted', 'TEEN',
'Father', 'mommy', 'Kid', 'mums', 'Mommy', 'adult', 'grandson', 'STEPMOM', 'child', 'FATHERS',
'SONINLAW', 'MUM', 'parents', 'kids', 'MOMMY', 'RESPONSIBLE', 'Uncle', 'childrens', 'Aunt', 'ADULT',
'moms', 'DADS', 'Fathers', 'STEPBROTHER', 'Kids', 'UNCLE', 'responsible', 'babysitter', 'Nephew',
'Teen', 'SIBLING', 'grandfather', 'childless', 'MOMMA', 'auntie', 'KID', 'ADOPTED', 'teenager',
'Grandfather', 'MOOOOM', 'mother', 'Babysitter', '4YEAROLD', 'dad', 'teen', 'childs', 'NEPHEW', 'MUMS',
'aunts', 'Daughters', 'Childrens', 'KIDS', 'stepfather', 'CHILD']

Bibliografia

- Ahluwalia, Resham, Himani Soni, Edward Callow, Anderson Nascimento, Martine De Cock. 2018. Detecting Hate Speech Against Women in English Tweets. In *Proceedings of Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018)*. Torino, CEUR.org.
- Anzovino, Maria, Elisabetta Fersini, Paolo Rosso. 2018. Automatic Identification and Classification of Misogynistic Language on Twitter. In *Proceedings of the 23rd International Conference on Applications of Natural Language to Information Systems (NLDB 2018) (Lecture Notes in Computer Science)*, Max Silberstein, Faten Atigui, Elena Kornysheva, Elisabeth Métais, Farid Meziane (Eds.), Vol. 10859. Springer, 57–64.
- Aron, Arthur, Elliot J. Coups, Elaine Aron. 2018. *Fondamenti di statistica. Introduzione alla ricerca in psicologia*. Pearson Italia, Milano, Torino.
- Bach, Kent, Robert M. Harnish. 1980. *Linguistic Communication and Speech Acts*. Massachusetts, MA: The MIT Press.
- Barnwal, Shubham, Ritesh Kumar, Rajendra Pamula. 2022. IIT DHANBAD CODECHAMPS at semeval-2022 task 5: MAMI - multimedia automatic misogyny identification. In *The 16th International Workshop on Semantic Evaluation*, pp. 733-735.
- Blackmore, Susan, Lee Alan Dugatkin, Robert Boyd, Peter J. Richerson, Henry Plotkin. 2000. "THE POWER OF MEMES." *Scientific American* 283, no. 4 (2000): 64–73. <http://www.jstor.org/stable/26058899>.
- Breazzano, Claudia, Edoardo Rubino, Danilo Croce, Roberto Basili. 2020. UNITOR @ DANKMEMES: Combining Convolutional Models and Transformer-based architectures for accurate MEME management. In Valerio Basile, Danilo Croce, Maria Di Maro, Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.
- Busso, Lucia, Ottavia Tordini, and Claudia Roberta Combei. 2020. "The Mediatization of Femicide: A Corpus-Based Study on the Representation of Gendered Violence in Italian Media". *L'Analisi Linguistica E Letteraria* 28 (3):20. <https://www.analisilinguisticaeletteraria.eu/index.php/ojs/article/view/2>.
- Capozzi, Arthur T. E., Mirko Lai, Valerio Basile, Fabio Poletto, Manuela Sanguinetti, Cristina Bosco, Viviana Patti, Giancarlo Ruffo, Cataldo Musto, Marco Polignano, Giovanni Semeraro, Marco Stranisci. 2019. Computational linguistics against hate: Hate speech

- detection and visualization on social media in the "Contro L'Odio" project. In *Proceedings of the Sixth Italian Conference on Computational Linguistics, CLiC-it 2019*. Cer, Daniel. Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. *Universal sentence encoder*. arXiv preprint arXiv:1803.11175.
- Chiaro, Delia. 2018. *The Language of Jokes in the Digital Age: #like, #share, #lol*. London New York: Routledge.
- Cignarella, Alessandra Teresa, Simona Frenda, Valerio Basile, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2018. Overview of the evalita 2018 task on irony detection in italian tweets (ironita). In Tommaso Caselli, Nicole Novielli, Viviana Patti, and Paolo Rosso, editors, *Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'18)*, Torino. CEUR.org.
- Citron, Danielle. 2009. Law's Expressive Value in Combating Cyber Gender Harassment. *Michigan Law Review* 108, no. 3: 373-415.
- Coates, Jennifer. 1986. *Women, men and language*. Londra, Longman.
- Crenshaw, Kimberle 1989. *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*. The University of Chicago Legal Forum 140:139-167.
- Criado-Perez, Caroline. 2020. *Invisibili. Come il nostro mondo ignora le donne in ogni campo. Dati alla mano*. Einaudi, Torino.
- Dancygier, Barbara, Lieven Vandelanotte. 2017. Internet memes as multimodal constructions. *Cognitive Linguistics*, volume 28, numero 3, pp. 565-598.
- Davison, Patrick. 2012. "The language of internet memes." *The social media reader*: 120-134.
- de Beauvoir, Simone. 1969. *Il secondo sesso*. ilSaggiatore, Milano.
- DeCook, Julia R. 2018. Memes and symbolic violence: #proudboys and the use of memes for propaganda and the construction of collective identity. *Learning, Media and Technology*, 43:4, 485-504, DOI: 10.1080/17439884.2018.1544149
- Doyle, Jude Ellison Sady. 2021. *Il mostruoso femminile. Il patriarcato e la paura delle donne*. Edizioni Tlon.
- Emerson, Guy, Natalie Schluter, Gabriel Stanovsky, Ritesh Kumar, Alexis Palmer, Nathan Schneider, Siddharth Singh, Shyam Ratan. 2022. *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics, Seattle, United States, edition.
- Fersini, Elisabetta Francesca Gasparini, Giulia Rizzi, Aurora Saibene, Berta Chulvi, Paolo

- Rosso, Alyssa Lees, and Jeffrey Sorensen. 2022. SemEval-2022 Task 5: Multimedia Automatic Misogyny Identification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 533–549, Seattle, United States. Association for Computational Linguistics.
- Fersini, Elisabetta, Debora Nozza. Paolo Rosso. 2020. "AMI @ EVALITA2020: Automatic Misogyny Identification." International Workshop on Evaluation of Natural Language and Speech Tools for Italian.
- Fersini, Elisabetta, Francesca Gasparini, Silvia Corchs. 2019. "Detecting Sexist MEME On The Web: A Study on Textual and Visual Cues," in 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, United Kingdom, pp. 226-231.
- Fersini, Elisabetta, Luca Rosato, Antonio Candelieri, Francesco Archetti, Enza Messina. 2021. *Deep Learning Representations in Automatic Misogyny Identification: What Do We Gain and What Do We Miss?.* In CLiC-it.
- Fersini, Elisabetta, Mary E. Anzovino, Paolo Rosso. 2018. Overview of the task on automatic misogyny identification at IberEval. In Proc. of IberEval 2018, volume 2150 of CEUR-WS, pages 214–228.
- Fersini, Elisabetta, Vincenzina Messina, Francesco Antonio Archetti. 2012. Emotional states in judicial courtrooms: An experimental investigation. *SPEECH COMMUNICATION*, 54(1), 11-12 [10.1016/j.specom.2011.06.001].
- Fiorino, Vinzia. 2006. Una storia di genere maschile: riflessioni su un approccio storiografico. *Contemporanea* 9.2 : 381-390.
- Fiorucci, Stefano. 2020. SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.
- Frazer, Rayan, Bronwyn Carlson. 2017. Indigenous memes and the invention of a people. *Social Media + Society*, 3(4). doi:10.1177/2056305117738993
- Frenda, Simona, Ghanem Bilal, et al. 2018. *Exploration of misogyny in Spanish and English tweets*, in: *Third Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2018*, volume 2150, pp. 260–267. Ceur Workshop Proceedings.

- Frenda, Simona. 2022. *Sarcasm and Implicitness in Abusive Language Detection: A Multilingual Perspective*. Tesi di dottorato. Universitat Politècnica de València.
- Furl, Katherine. 2022. *Denigrating Women, Venerating "Chad": Ingroup and Outgroup Evaluations among Male Supremacists on Reddit*. *Social Psychology Quarterly*, volume 85, (3), 279–299. <https://doi.org/10.1177/01902725221090907>
- García-Díaz, José Antonio, Mar Cánovas-García, Ricardo Colomo-Palacios, and Rafael Valencia- García. 2021. Detecting Misogyny in Spanish Tweets. An Approach based on Linguistics Features and Word embeddings. *Future Generation Computer Systems*, 114:506–518.
- Gasparini, Francesca, Ilaria Erba, Elisabetta Fersini, Silvia Corchs. 2018. Multimodal Classification of Sexist Advertisements. In *Proceedings of the 15th International Joint Conference on e-Business and Telecommunications - ICETE*, ISBN 978-989-758-319-3; ISSN 2184-2825, pages 399-406. DOI: 10.5220/0006859405650572
- Gemelli, Sara. 2021. *Bei faccini e brutti veri: un'analisi sociolinguistica e critica del discorso in un forum incel italiano*. Tesi magistrale, Università degli studi di Pavia, relatrice Chiara Zanchi, correlatrice Ilaria Fiorentini.
- Gianini Belotti, Elena. 2014. *Dalla parte delle bambine: l'influenza dei condizionamenti sociali nella formazione del ruolo femminile nei primi anni di vita*. 29 ed. Milano: Feltrinelli.
- Grundlingh, Lezandra. 2017. *Memes as speech acts*. *Social Semiotics*, volume 28, numero 2, pp. 147-168. Londra: Routledge.
- Gualco, Barbara, Regina Rensi, Giovanna Fossa. 2017. "Violenza assistita e comportamenti devianti dei giovani in Italia: i risultati dell'International Self-report Delinquency Study-3." *Rassegna Italiana di Criminologia*: 104-114.
- Guest, Ella, Bertie Vidgen, Alexandros Mittos, Nishanth Sastry, Gareth Tyson, Helen Margetts. 2021. An expert annotated dataset for the detection of online misogyny. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1336–1350, Online. Association for Computational Linguistics.
- HaCohen-Kerner, Yaakov, Hananya Beck, Elchai Yehudai, Mordechai Rosenstein, Dror Mughaz. 2010. Cuisine: classification using stylistic feature sets and/or name-based feature sets. *J. Assoc. Inf. Sci. Technol.* 61(8), 1644–1657.
- Hakimov, Sherzod, Gullal Singh Cheema, Ralph Ewerth. 2022. *TIB-VA at semeval-2022 task 5: A multimodal architecture for the detection and classification of misogynous memes*. In *The 16th International Workshop on Semantic Evaluation*, pp. 756-760.

- Herriman, Jennifer. 1998. Description of Woman and Man in Present-Day English. In *Moderna Sprak*, vol. XCII/2, pp. 136-142.
- Jurafsky, Daniel, James H. Martin. 2008. *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2. edition. Upper Saddle River (NJ): Prentice Hall.
- Karras, Ruth Mazo. 2003. *From boys to men: Formations of masculinity in late medieval Europe*. University of Pennsylvania Press.
- Khandkar, Shahedul Huq. 2009. *Open coding*. University of Calgary. <http://pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/opencoding.pdf>
- Kiela, Douwe, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, Davide Testuggine. 2020. *The hateful memes challenge: Detecting hate speech in multimodal memes*. arXiv preprint arXiv:2005.04790.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž, David Tugwell. 2004. The sketch engine. *Proceedings of the 11th EURALEX International Congress*: pp. 105-116.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography*, 1: 7-36.
- Lenci, Alessandro, Simonetta Montemagni, Vito Pirrelli. 2020. *Testo e computer: elementi di linguistica computazionale*. Roma: Carocci Editore.
- Lenci, Alessandro. 2008. Distributional semantics in linguistic and cognitive research. In *Italian journal of linguistics* 20.1: 1-31.
- Lippe, Phillip, Nithin Holla, Shantanu Chandra, Santhosh Rajamanickam, Georgios Antoniou, Ekaterina Shutova, Helen Yannakoudakis. 2020. *A multi-modal framework for the detection of hateful memes*. *CoRR*, abs/2012.12871.
- Lipperini, Loredana. 2007. *Ancora dalla parte delle bambine*. Milano: Feltrinelli.
- Maeve, Duggan, Nicole B. Ellison, Cliff Lampe, Amanda Lenhart, Mary Madden. 2015. "Online Harassment." Report. Pew Research Center, 2015.
- Maltz, Daniel, Ruth A. Borker. 1982. A Cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and social identity*. New York: Cambridge University Press.
- Marchetti, Sabrina. 2022. Bello, BG (2020), Intersezionalità. Teorie e pratiche tra diritto e società, Milano, FrancoAngeli. *AG About Gender-International Journal of Gender Studies* 11.21.
- Maxwell, December, Sarah R. Robinson, Jessica Roberts Williams, Craig Keaton. 2020. "A Short Story of a Lonely Guy": A Qualitative Thematic Analysis of Involuntary Celibacy

Using Reddit. Sexuality & Culture (2020): 1-23.

- Meluzzi, Chiara, Erica Pinelli, Elena Valvason, Chiara Zanchi. 2021. Responsibility attribution in gender-based domestic violence: A study bridging corpus-assisted discourse analysis and readers' perception. *Journal of Pragmatics*, 185(1):73–92. DOI: 10.1016/j.pragma.2021.07.023
- Milani, Martina, Giulia Giorgi, Ilir Rama, Guido Anselmi, Gianluca E. Lebani. 2020. DANKMEMES@EVALITA2020: The memeing of life: memes, multimodality and politics. In Valerio Basile, Danilo Croce, Maria Di Maro, Lucia C. Passaro, editors, *Proceedings of the 17th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Torino: Accademia University Press. doi:10.4000/books.accademia.7330 .
- Milner, Rayan M. 2012. *The World Made Meme: Discourse and Identity in Participatory Media*. PhD diss., University of Kansas. Accessed March 13 2015. https://kuscholarworks.ku.edu/bitstream/handle/1808/10256/Milner_ku_0099D_12255_DATA_1.pdf?sequence=1.
- Miltner, Kate M. 2014. There's no place for lulz on lolcats: The role of genre, gender and group identity in the interpretation and enjoyment of an internet meme. *First Monday*, 19(8), <http://firstmonday.org/ojs/index.php/fm/article/view/5391/4103>.
- Minnema, Gosse, Sara Gemelli, Chiara Zanchi, Viviana Patti, Tommaso Caselli, Malvina Nissim. 2021. Frame Semantics for Social NLP in Italian: Analyzing Responsibility Framing in Femicide News Reports. In *CLiC-it*.
- Muti, Arianna, Alberto Barròn-Cedeño. 2020. *UniBO@AMI: A Multi-Class Approach to Misogyny and Aggressiveness Identification on Twitter Posts Using AIBERTo*. In *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Work-shop (EVALITA 2020)*, Bologna, Italy. CEUR.org
- Nencioni, Giovanni. 1976. Parlato-parlato, parlato-scritto, parlato-recitato, «Strumenti critici» 10, pp. 1-56 (poi in Id., *Di scritto e parlato, Discorsi linguistici*, Bologna, Zanichelli, 1983, pp. 126-179).
- Nissenbaum, Asaf, Limor Shifman. 2018. Meme templates as expressive repertoires in a globalizing world: A cross-linguistic study. *Journal of Computer-Mediated Communication* 23.5: 294-310.
- Nobata, Chikashi, Joel Tetreault, Achint Thomas, Yashar Mehdad, and Yi Chang. 2016. Abusive language detection in online user content. In *Proc. of the 25th International Conference on World Wide Web*, pages 145–153.

- Nozza, Debora, Claudia Volpetti, Elisabetta Fersini. 2019. Unintended bias in misogyny detection. In *Proceedings - 2019 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2019* (pp.149-155). New York, NY, USA: Association for Computing Machinery, Inc [10.1145/3350546.3352512].
- Ostanina-Olszewska, Julia, Aleksandra Majdzińska-Koczorowicz. 2019. *A Cognitive Linguistics approach to internet memes on selected Polish internet sites*. *Cognitive Studies*, (19): http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.ojs-doi-10_11649_cs_1939
- Pamungkas, Endang Wahyu, Alessandra Teresa Cignarella, Valerio Basile, Viviana Patti, et al. 2018. 14-exlab@unito for ami at ibereval2018: Exploiting lexical knowledge for detecting misogyny in english and spanish tweets. In *3rd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2018*, volume 2150, pages 234–241. Ceur Workshop Proceedings.
- Pamungkas, Endang Wahyu, Valerio Basile, and Viviana Patti. 2020. Misogyny detection in twitter: a multilingual and cross-domain study. *Information Processing & Management*, 57(6):102360.
- Phillips, Whitney, Rayan M. Milner. 2017. Decoding memes: Barthes' punctum, feminist standpoint theory, and the political significance of #YesAllWomen. In S. Harrington (Ed.), *Entertainment Values* (pp. 195–211). London, England: Palgrave Macmillan.
- Poland, Bailey. 2016. *Haters: Harassment, Abuse, and Violence Online*. University of Nebraska Press. <https://doi.org/10.2307/j.ctt1fq9wdp>
- Prada, Massimo. 2015. *L'italiano in rete. Usi e generi della comunicazione mediata tecnicamente*, FrancoAngeli, Milano.
- Prada, Massimo. 2016. "Lingue e Internet", in *La lingua italiana e i mass media*, a cura di Ilaria Bonomi e Silvia Morgana, Carrocci editore, Roma.
- Pranesh, Raj Ratn, Ambesh Shekhar. 2020. *Memesem: a multi-modal framework for sentimental analysis of meme via transfer learning*. In *Thirty-seventh International Conference on Machine Learning (ICML)*, 4th Lifelong Learning Workshop.
- Rao, Ailneni Rakshitha, Arjun Rao. 2022. ASRtrans at semeval-2022 task 5: Transformer - based models for meme classification. In *The 16th International Workshop on Semantic Evaluation*, pp. 597-604.
- Razavi, Amir H., Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive language detection using multi-level classification. In *Canadian Conference on Artificial Intelligence*, pages 16–27. Springer.

- Rinaldi, Cirus. 2021. La violencia contra las mujeres como prueba de masculinidad. Reflexiones socio-criminológicas. *La Aljaba*. 2021; Vol.25 no.1
- Ruiz, Vicki, Ellen Carol DuBois. 2008. *Unequal Sisters: an Inclusive Reader in U.S. Women's History*. 4. ed. / edited by Vicki L. Ruiz with Ellen Carol DuBois. New York London: Routledge.
- Rychlý, Pavel. 2008. A Lexicographer-Friendly Association Score. *Proc. 2nd Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN*, 2: 6-9.
- Sabatini, Alma. 1987. *Il sessismo nella lingua italiana*, Presidenza del Consiglio dei ministri – Direzione generale delle informazioni della editoria e della proprietà letteraria artistica e scientifica, Roma.
- Schettini, Laura. 2012. Un sesso che non è un sesso: medicina, ermafroditismo e intersessualità in Italia tra Otto e Novecento. *Un sesso che non è un sesso: medicina, ermafroditismo e intersessualità in Italia tra Otto e Novecento*: 19-40.
- Scott, Joan Wallach. 1999. Gender: A Useful Category of Historical Analysis, in *Gender and the Politics of History*, revised edition, New York: Columbia University Press: 28–50.
- Setpal, Jinen, Gabriele Sarti. 2020. Dankmemesteam @ dankmemes: Archimede: A new model architecture for meme detection. In Valerio Basile, Danilo Croce, Maria Di Maro, Lucia C. Passaro, a cura di, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.
- Sharma, Chhavi, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, Bjorn Gambäck. 2020. SemEval-2020 Task 8: Memotion Analysis – the Visuo-Lingual Metaphor! *arXiv Preprint arXiv:2008.03781*. <http://arxiv.org/abs/2008.03781>.
- Sharma, Mayukh, Ilanthenral Kandasamy, Vasantha W B. 2022. R2d2 at semeval-2022 task 5: Attention is only as good as its values! a multimodal system for identifying misogynist memes. In *The 16th International Workshop on Semantic Evaluation*, pp. 761-770.
- Shifman, Limor. 2013. Memes in a Digital World: Reconciling with a Conceptual Troublemaker. *Journal of Computer-Mediated Communication*, 18(3):362– 377.
- Simonyan, Karen, Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Simpson, Paul. 1993. *Language, Ideology and Point of View*. Londra: Routledge.
- Sivanaiah, Rajalakshmi, Angel Deborah S, Sakaya Milton Rajendram, Mirnalinee T T. 2022. TechSSN at semeval-2022 task 5: Multimedia automatic misogyny identification using

- deep learning models. In *The 16th International Workshop on Semantic Evaluation*, pp. 571-574.
- Spallaccia, Beatrice. 2017. *Misogynistic hate speech on social networks: a critical discourse analysis*, tesi di dottorato, Università degli Studi di Bologna.
- Spina, Stefania. 2001. *Fare i conti con le parole. Introduzione alla linguistica dei corpora*. Guerra Edizioni, Perugia.
- Tao, Chen, Jung jae Kim. 2022. taochen at semeval- 2022 task 5: Multimodal multitask learning and ensemble learning. In *The 16th International Workshop on Semantic Evaluation*.
- UN Women. "Cyber Violence against Women and Girls. A World-Wide Wake-Up Call." Report. United Nation Women, 2015.
- Van Aken, Betty, Julian Risch, Ralf Krestel, Alexander Löser. 2018. Challenges for toxic comment classification: An in-depth error analysis. *arXiv preprint arXiv:1809.07572*.
- Van der Maaten, Laurens, Geoffrey Hinton. 2008. "Visualizing data using t-SNE." *Journal of machine learning research* 9.11.
- Vlad, George-Alexandru, George-Eduard Zaharia, Dumitru-Clementin Cercel, Mihai Dascalu. 2020. Upb @ dankmemes: Italian memes analysis: Employing visual models and graph convolutional networks for meme identification and hate speech detection. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*. Final Workshop (EVALITA 2020), Online. CEUR.org.
- Vox. "Misoginia: la rete si scatena contro le donne." Web. 2016. *voxdiritti.it*. 3 January 2017. <<http://www.voxdiritti.it/misoginia-la-rete-si-scatena-contro-le-donne/>>.
- Zampieri, Marcos, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, Ritesh Kumar. 2019. SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, 75–86.
- Zhang, Ziqi, Lei Luo. 2018. Hate speech detection: A solved problem? The challenging case of long tail on Twitter. *arXiv preprint arXiv:1803.03662*.
- Zhi, Jin Mei, Zhou Mengyuan, Mengfei Yuan, Dou Hu, Xiyang Du, Lianxin Jiang, Yang Mo, XiaoFeng Shi. 2022. PAIC at semeval-2022 task 5: Multimodal misogynous detection in MEMES with multi-task learning and multi-model fusion. In *The 16th International Workshop on Semantic Evaluation*.
- Zhou, Ziming, Han Zhao, Jingjing Dong, Ning Ding, Xiaolong Liu, Kangli Zhang. 2022. DD-TIG

at semeval-2022 task 5: Investigating the relationships between multimodal and unimodal information in misogynous memes detection and classification. In *The 16th International Workshop on Semantic Evaluation*.

SITOGRAFIA

Statistiche sul genere degli utenti dei social network:
<https://www.statista.com/statistics/471345/us-adults-who-use-social-networks-gender/>

Statistiche dell'ONU sui femminicidi: <https://www.unwomen.org/sites/default/files/2022-11/Gender-related-killings-of-women-and-girls-improving-data-to-improve-responses-to-femicide-feminicide-en.pdf>

Gender gap report del 2022: https://www3.weforum.org/docs/WEF_GGGR_2022.pdf

Sito del Working to Halt Online Abuse (WHOA)
<http://www.haltabuse.org/resources/stats/index.shtml>

Sito dell'OCR di Google usato per la challenge MAMI:
<https://cloud.google.com/vision/docs/ocr?hl=it>

Link github dei dati della challenge MAMI: <https://github.com/MIND-Lab/SemEval2022-Task-5-Multimedia-Automatic-Misogyny-Identification-MAMI->

Link github per consultare le liste di frequenza usate per il presente lavoro:
<https://github.com/MIND-Lab/Misogynous-N-Gram-Lists>