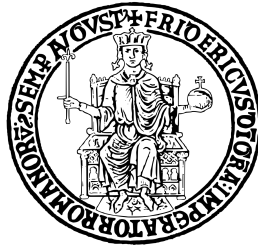


UNIVERSITÀ DEGLI STUDI DI NAPOLI “FEDERICO II”



DIPARTIMENTO DI SCIENZE SOCIALI

DOTTORATO IN

“SCIENZE SOCIALI E STATISTICHE”

XXXII CICLO

**LA VIOLENZA CONTRO LE DONNE IN EUROPA: UNA
STRATEGIA DI *CROSS-LANGUAGE ANALYSIS***

COORDINATORE

Ch.mo Prof. Roberto Serpieri

CANDIDATA

Dott.ssa Alessia Forciniti

TUTOR

Ch.ma Prof.ssa Germana Scepi

CO- TUTOR

Ch.ma Prof.ssa Enrica Morlicchio

A Simona Balbi

Per avermi insegnato che nel “gioco” di potere dei ruoli accademici si possa fare la differenza ... Nell’eterna speranza di essere riuscita a mantenere la nostra promessa, esprimo infinita gratitudine per l’affetto, la stima, l’umiltà ed il rispetto riservatemi.

Ringraziamenti

Un concreto ringraziamento è rivolto al *Prof. Roberto Serpieri*, Coordinatore del Trentaduesimo Ciclo di Dottorato in Scienze Sociali e Statistiche dell'Università degli Studi di Napoli "Federico II", per l'ineccepibile presenza, l'etica del lavoro dimostrata e per non aver mai dimenticato, in nessuna circostanza, la misura del rispetto.

Nei confronti della *Prof.ssa Enrica Morlicchio* esprimo autentica riconoscenza per essere stata la mia *Co-Tutor* e per aver contribuito alla mia crescita emotiva e professionale attraverso i suoi spunti interpretativi.

Grazie alla *Prof.ssa Germana Scepi* per aver accettato di essere la mia *Tutor* dopo la perdita prematura della *Prof.ssa Balbi*.

Un grazie particolare alla *Prof.sse Domenica Fioredistella Iezzi e Lavinia Bifulco* per i preziosi suggerimenti scientifici inoltrati in fase di referaggio.

Autentica gratitudine è rivolta al *Prof. Kenneth Benoit*, per essere stato il mio punto di riferimento presso la *London School of Economics* e per avermi trasmesso il coraggio di osare, la serenità del fallimento e il valore della persistenza.

Con sincero affetto esprimo riconoscenza nei confronti della *Dott.ssa Anna Coccozza*, per la preziosa capacità di saper cogliere l'invisibile, trasformando le percezioni in attribuzioni di significato.

Un grazie speciale è rivolto al mio *papà*, alla mia *mamma* e ai miei *super nonni*, per avermi sempre sostenuta con incommensurabile dedizione. Infinitamente grazie per avermi inculcato valori di lealtà, dignità e rispetto, dotandomi degli strumenti per capire il confine di ogni mia libertà e mostrandomi che è la correttezza e non la convenienza a fare la differenza in questa esistenza. Probabilmente in un mondo di arroganza e prepotenza i miei principi non mi porteranno troppo lontano ma di sicuro non mi condurranno a fare compromessi con la mia coscienza.

A mia *madre* devo un grazie particolare per non aver mai lasciato la mia mano, nonostante le troppe volte in cui la Vita mi ha colpito alla schiena, e per aver ridato luce alla mia esistenza, quando le vicissitudini avevano ormai reso "il mio cuore gonfio e l'anima ignota" (Fabrizio e Morra, 2007).

Ai miei amici di sempre, quelli su cui poter contare al di là delle apparenze, delle logiche e delle distanze. Con autentico affetto a *Giulia, Chiara, Elena, Alice, Cristina*,

Silvia, Eduardo, Claudio, Danilo, Manu e Benedetta, per la loro inestimabile intelligenza emotiva, il costante sostegno e l'originalità di pensiero.

Al mio miglior amico *Francesco*, grazie per la fraterna condivisione dell'esistenza, per i sorrisi più veri, per ogni abbraccio ricevuto al momento giusto, per la capacità di leggere tra le righe anche solo di un sospiro o di uno sguardo, per ogni lacrima condivisa, ogni segreto custodito. Con affetto e gratitudine, perché a dispetto del tempo e dei mutamenti, continuiamo a condividere lo stesso linguaggio dell'anima.

Alla mia preziosa amica *Angela* rivolgo un sincero ringraziamento per essere rimasta sempre al mio fianco con estrema onestà intellettuale e delicatezza dialettica, rendendo più lieve il peso del coraggio di ogni singola scelta.

Nei confronti della mia guida spirituale *Don Michele* esprimo un pensiero di sincera riconoscenza, per avermi insegnato che non occorre cambiare orizzonti ma avere uno sguardo diverso di fronte alla pochezza dello spirito. Grazie per essere riuscito a sedare la mia spasmodica ricerca di giustizia e verità, indicatomi la strada del perdono di fronte alla mediocrità, agli inganni e alle trame ordite dalle frustrazioni altrui.

Un grazie particolare a *Giuseppe*, per avermi concesso, nonostante le ombre e gli errori, l'opportunità di fare i conti con le mie più autentiche fragilità e per la vulnerabilità che ha generato *caos* nel mio perfezionismo innato. Grazie, per aver reso più intensa la consapevole ricerca dell'essenza del vivere.

Voglio ringraziare i miei colleghi del Trentaduesimo Ciclo, poiché è stato un onore poter condividere con ciascuno di loro questo percorso. In particolare, voglio ricordare e ringraziare *Francesca* per avermi trasmesso il coraggio di lottare con amore e dignità per questa Vita, che comunque vada, merita di essere vissuta.

Infine, esprimo profondissima gratitudine nei confronti della *Vita* per avermi concesso il privilegio di sopravvivere alle intemperie materiali ed emotive sopraggiunte, per avermi insegnato ad inciampare e a rialzarmi, malgrado gli innumerevoli "squali" venuti a galla nel mare degli spiacevoli incontri umani.

Indice

Introduzione.....	1
Capitolo primo	
Violenza contro le donne tra identità “sessuata” e dominio maschile	5
1.1. Un genocidio invisibile: definire la Violenza sulle donne.....	6
1.2. Fuori e dentro le mura domestiche.....	12
1.2.1. <i>La violenza nei perimetri nascosti delle pareti di casa.....</i>	<i>13</i>
1.2.2. <i>L’abuso sessuale.....</i>	<i>17</i>
1.2.3. <i>Il mobbing sul posto di lavoro.....</i>	<i>21</i>
1.2.4. <i>L’azione persecutoria dello Stalking.....</i>	<i>23</i>
1.2.5. <i>La cyber-violenza.....</i>	<i>27</i>
1.2.6. <i>Il Vitriolage: l’aggressione con l’acido.....</i>	<i>29</i>
1.2.7. <i>Le mutilazioni genitali.....</i>	<i>31</i>
1.3. La criminalizzazione della violenza attraverso il “Femminicidio”.....	34
1.4. Le conseguenze dei labirinti del male.....	37
1.5. Una breve <i>anamnesi</i> delle possibili costanti che caratterizzano il fenomeno... 	41
1.5.1. <i>Il “genere” come categoria di analisi.....</i>	<i>41</i>
1.5.2. <i>Il dominio maschile.....</i>	<i>44</i>
1.5.3. <i>L’asimmetria simbolica.....</i>	<i>47</i>
1.5.4. <i>La virilità.....</i>	<i>49</i>
1.5.5. <i>L’esibizionismo simbolico: immaginario e rappresentazione mediatica di genere.....</i>	<i>51</i>
1.6. Verso un cambiamento di prospettiva: il Femminismo.....	54
1.7. Le conquiste femminili attraverso l’inquadramento normativo	59
1. 8. Le risposte sul piano sociale.....	66
1. 9. Il fenomeno oggetto di studio attraverso la visione universalistica dei Diritti Umani	69
1. 10. Le statistiche ufficiali sulla violenza contro le donne	71
1. 11. La violenza simbolica attraverso i dati sull’uguaglianza di genere	76
1.11.1. <i>Il settore professionale: tasso occupazionale e gap retributivo.....</i>	<i>79</i>
1.11.2. <i>Cultura ed istruzione.....</i>	<i>83</i>
1.11.3. <i>Il dominio del tempo.....</i>	<i>85</i>
1.11.4. <i>L’acquisizione di potere nell’ambito politico-istituzionale.....</i>	<i>87</i>

1.11.5. <i>La rappresentanza nella leadership aziendale</i>	92
1.11.6. <i>La dimensione della salute come percezione del sé</i>	92
1.12. Cenni conclusivi	95

Capitolo Secondo

L'analisi dei dati testuali come strumento per gli studi <i>Cross-language</i>	99
---	-----------

2.1. Formalizzazione del binomio lingua-cultura e degli studi multilingue in un'era di globalizzazione e plurilinguismo	99
2.2. Dal testo al <i>corpus</i>: i criteri di classificazione	102
2.3. Gli orientamenti teorici delle Scienze Sociali che strutturano l'analisi dei dati testuali: paradigmi positivista ed interpretativista	109
2.4. Dalla teoria alle procedure operative: analisi qualitativa e quantitativa a confronto	111
2.5. L'affermarsi della Statistica Testuale come metodo quantitativo	114
2.5.1. <i>Le fasi operative della Statistica Testuale</i>	116
2.5.2. <i>La codifica e i sistemi di ponderazione dell'unità di analisi</i>	118
2.6. L'analisi multidimensionale	119
2.7. Gli studi <i>cross-language</i> dei dati testuali: principali metodi e tecniche	121
2.8. Cenni conclusivi	128

Capitolo Terzo

Una strategia Statistica per l'Analisi Multilingue Comparativa	131
---	------------

3.1. Il <i>database</i> documentale dell'analisi statistica multilingue	132
3.2. Le scelte strategiche della proposta <i>cross-language</i>	136
3.2.1. <i>Analisi esplorativa monolingue del fenomeno indagato</i>	138
3.2.1.1. <i>Le fasi di pre-processing e feature extraction di ciascun corpus monolingue</i>	138
3.2.1.2. <i>Text summarization dei tweets estratti per ogni lingua</i>	143
3.2.2. <i>Definizione di concetti monolingue</i>	144
3.2.2.1. <i>L'orientamento semantico dei testi ed i principali approcci per il calcolo della polarità</i>	144
3.2.2.2. <i>L'NRC Emotion Lexicon per misurare la polarità dei testi multilingue</i>	146

3.2.2.3. <i>La Concepts Extraction tramite Correspondence e Cluster Analysis</i> ...	148
3.2.2.3.1. <i>L'Analisi delle Corrispondenze</i>	150
3.2.2.3.2. <i>La Cluster Analysis</i>	154
3.2.2.4. <i>La codifica dei dati per la costruzione della matrice di concetti</i>	157
3.2.3. <i>Metodi di Matrix Comparison</i>	160
3.2.4. <i>Il confronto tra matrici di concetti monolingue in uno spazio dimensionale multilingue attraverso l'Analisi di Procuste</i>	163
3.2.4.1. <i>L'Analisi Procustiana Classica</i>	165
3.2.4.2. <i>L'Analisi Procustiana Generalizzata</i>	167
3.3. Cenni conclusivi	173

Capitolo Quarto

I risultati della *Cross-language Analysis* sul fenomeno della Violenza contro le donne...177

4.1. L'estrazione dei dati: tecniche, metodi e parametri di rilevazione	178
4.2. La collezione di <i>tweets</i> multilingue	182
4.3. L'analisi preliminare attraverso lo studio esplorativo multilingue	184
4.3.1. <i>I risultati dell'Italia</i>	185
4.3.2. <i>I risultati del Regno Unito</i>	191
4.3.3. <i>I risultati della Francia</i>	196
4.3.4. <i>I risultati della Spagna</i>	201
4.3.5. <i>I risultati dell'Ucraina</i>	206
4.3.6. <i>Brevi considerazioni sui risultati dello studio esplorativo</i>	211
4.4. La definizione dei concetti per ciascuna lingua	212
4.4.1. <i>Polarità semantica e Concepts Extraction dei tweets italiani</i>	214
4.4.2. <i>Polarità semantica e Concepts Extraction dei tweets inglesi</i>	217
4.4.3. <i>Polarità semantica e Concepts Extraction dei tweets francesi</i>	219
4.4.4. <i>Polarità semantica e Concepts Extraction dei tweets spagnoli</i>	221
4.4.5. <i>Polarità semantica e Concepts Extraction dei tweets ucraini</i>	224
4.4.6. <i>Dalla rilevazione dei concetti alla comparazione cross-language</i>	227
4.5. La GPA per il confronto tra i concetti in unico spazio dimensionale multilingue	231
4.5.1. <i>La Generalized Procrustes Analysis Map: lingue a confronto</i>	231
4.5.2. <i>L'Analisi della Varianza Generalizzata: PANOVA</i>	233
4.5.3. <i>Similarità tra le configurazioni e Consensus Permutation test</i>	236

4.6. I risultati dell'analisi <i>cross-language</i> in un breve confronto con alcune fonti <i>ufficiali e parallele</i>	237
4.7. L'analisi dei concetti e del <i>Consensus tra le lingue per una riflessione sulle Politiche Sociali anti violenza</i>	240
4.8. Cenni conclusivi.....	247
Conclusioni	251
Riferimenti bibliografici	257
<i>Appendice A</i>	287
<i>Appendice B</i>	311

Introduzione

Il tema del presente lavoro di tesi è la violenza contro le donne, universalmente riconosciuto come uno dei più rilevanti fenomeni di violazione dei diritti umani. Le dinamiche trans-storiche dell'interazione tra i *generi* si decompongono e ricompongono caratterizzando e rinnovando una struttura sociale prevalentemente androcentrica. La prospettiva fenomenologica si delinea attraverso la definizione di un'identità *sessuata* ed un costante bilanciamento di potere nei vari domini interazionali.

Malgrado l'esistenza su scala globale di tale fenomeno, riuscire a rilevare il livello dimensionale dello stesso rappresenta un notevole limite, a causa della mancanza di statistiche ufficiali aggiornate e capaci di coprirne l'estensione mondiale.

Pertanto, nel tentativo di studiare il tema oggetto di studio attraverso un'angolazione internazionale, il presente lavoro di tesi propone una strategia statistica per l'analisi di linguaggi e culture differenti che indaga la percezione pubblica europea sul fenomeno della violenza contro le donne. Più propriamente, l'obiettivo è quello di evidenziare eventuali similarità e/o dissimilarità tra i diversi registri linguistici, supponendo l'esistenza di un binomio lingua-cultura. In letteratura esiste infatti l'ipotesi che le strutture grammaticali e semantiche di una lingua inglobino modalità di strutturazione del pensiero collettivo.

La strategia di analisi multilingue comparativa sulla percezione pubblica della violenza contro le donne si presenta nell'ambito della comunicazione digitale e nello specifico nel dominio dei Social Media, intesi come l'insieme delle applicazioni e dei fondamenti ideologici e tecnologici del *Web 2.0*. Nella fattispecie, l'analisi dei dati estratti dai Social Media rappresenta attualmente una delle strategie maggiormente impiegate da diverse aree di studio, al fine di qualificare i fenomeni e mappare l'opinione pubblica. In particolare, nell'ultimo decennio è sempre più rilevante lo studio sul linguaggio della violenza propinato attraverso il *Web*, ponendo una particolare enfasi su forme di linguaggio offensivo e discriminatorio, come ad esempio l'*online hate speech* contro le donne. In modo particolare, lo studio che si presenta in questa tesi si concretizza nell'analisi dell'opinione pubblica su *Twitter*, la cui piattaforma, rispetto ad altri Social Media, alcuni dei quali sicuramente più popolari, si distingue per le sue caratteristiche di apertura ed orizzontalità che gli consentono di tradursi in un valido strumento di ricerca per numerosi temi di

discussione che animano la rete, grazie al potere *influenzativo* della *comunicazione partecipativa* che ne deriva.

Per l'estrazione dei *tweets* risulta necessario ricorrere ad una o più *parole chiave* attinenti al fenomeno, ma il multilinguismo e multiculturalismo esistenti presuppongono la presenza di variabili socio-culturali che influenzano la *notiziabilità* degli argomenti. Quindi, si reputa opportuno affidare la selezione delle *keywords* a fonti ufficiali e specializzate ed in base alla loro disponibilità, il presente studio si realizza analizzando i contenuti estratti da Twitter in cinque diverse lingue europee: italiano, inglese, francese, spagnolo ed ucraino. L'estrazione dei dati avviene in un periodo di monitoraggio compreso tra il primo maggio ed il 31 ottobre 2019.

In senso pratico, al fine di analizzare i dati testuali *cross-language* postati su Twitter si ricorre all'utilizzo di tecniche tipiche di *Text Mining*.

I principali metodi e approcci della letteratura multilingue del dato testuale sono soliti confrontare solo due lingue per volta tramite tecniche di traduzione automatica, l'uso di dizionari bilingue o il ricorso a corpora paralleli. In questo studio invece, con lo scopo di mantenere inalterate le peculiarità proprie di ogni lingua, il confronto si basa su un'unità di analisi alternativa e di ordine concettuale, che riduce la variabilità tra lingue.

La tesi si sviluppa fornendo dapprima uno studio esplorativo sui contenuti testuali di ogni lingua, mappando così la percezione monolingue del fenomeno. In un secondo momento, si individuano le nuove unità di analisi di ordine concettuale, classificando in *cluster* le diverse *features* monolingue, ovvero i termini estratti da ciascun *corpus* di dati in uno spazio vettoriale multidimensionale. Prima di ottenere tali raggruppamenti tuttavia si intende assicurare l'orientamento semantico delle *features*, attraverso il calcolo della polarità semantica. L'assegnazione dei punteggi della polarità si affida ad un *lexicon* multilingue conosciuto come *NRC (National Research Council Canada) Emotion Lexicon* che consente di individuare la direzione semantica di ciascun *tweet* sulla base di un dizionario di parole polarizzate per ciascuna delle lingue considerate. Tale *lexicon* permette al contempo di determinare otto categorie emozionali, che meglio supportano l'interpretazione semantica del testo. A partire da ciascuna collezione di dati monolingue suddivisa in base alla polarità, si propone di adottare in maniera combinata due tecniche di analisi multivariata: l'*Analisi delle Corrispondenze* e la *Cluster Analysis*, che ne consentono la *Concepts Extraction*. Da tale applicazione si ottengono raggruppamenti di classi di concetti solo positivi e raggruppamenti di classi di

concetti solo negativi. Più propriamente, sulle coordinate dell'*Analisi delle Corrispondenze* si realizza una strategia di classificazione mista, che parte dal metodo gerarchico di *Ward* e si consolida con il metodo non gerarchico del *k-means*. Dopo aver identificato i raggruppamenti concettuali delle diverse lingue, è interessante valutare se i concetti individuati nei diversi linguaggi abbiano relazioni simili tra loro e quindi se ci sia una coesistenza dei campi concettuali multilingue. Nello specifico, al fine di confrontare le lingue, si conduce l'*Analisi Procustiana Generalizzata* che consente di comparare simultaneamente in unico spazio fattoriale le cinque configurazioni dei concetti e dare quindi un risultato di *consensus* o accordo tra le lingue. Tale tecnica permette di cogliere similarità e/o dissimilarità tra le lingue ed individuare quali concetti mettano più d'accordo l'opinione pubblica internazionale riguardo al tema della violenza contro le donne. L'analisi dei concetti e del consenso tra le cinque configurazioni linguistiche rappresentano inoltre uno spunto interpretativo di riflessione sull'opinione pubblica internazionale e sulle politiche di sostegno antiviolenza relativamente al fenomeno preso in esame.

Capitolo Primo

Violenza contro le donne tra identità “sessuata” e dominio maschile

L'obiettivo del primo capitolo è presentare il fenomeno della violenza contro le donne non limitandosi a descrivere solo la vittima ma allargando la prospettiva di analisi anche e soprattutto all'autore degli atti violenti, valutando le dinamiche che hanno strutturato una società prettamente fallocratica. L'elaborato si prefigge di delineare alcuni aspetti sociologici inerenti al tema oggetto di studio e di presentare uno scenario quantitativo del fenomeno. In questo contesto, l'intento non è studiare il fenomeno della violenza contro le donne, di cui invece si possono addurre eventuali future ricerche di studio sociologico, ma inquadrare e decostruire l'evoluzione storico-sociale del bilanciamento di potere che nutre l'affermazione identitaria maschile. In particolare, nel tentativo di “denudare” le dimensioni più nascoste della subordinazione della donna che si esaspera attraverso la violenza femicida, si indagano le dimensioni identitarie ed interazionali tra i generi, comprese le forme di interazione che nell'attuale era globale si celano dietro nuovi contesti e domini (cultura e istruzione, salute e percezione del sé, retribuzione, tempo) di latente affermazione di potere. Dalla presentazione di tale analisi si vuole, senza la presunzione dell'esaustività, delineare un profilo sociologico del fenomeno che evidenzia nella complessità i riferimenti storici, socio-culturali e relazionali tra uomo e donna.

Come suggerito da alcuni autori, il fenomeno infatti non può essere letto come una manifestazione fenomenologica autonoma ma è strettamente connesso ad un'analisi dei soggetti che sono coinvolti nel processo interazionale violento.

Si pone al centro della questione la costruzione di un'*identità sessuata* (Bourdieu, 1998) nella quale strutturazione è proprio il sesso a rappresentare il movente sostanziale ad un funzionalismo sociologico, propriamente inteso come globalità di strutture sociali e culturali connesse tra loro. Nel corso del tempo infatti l'*identità sessuata* ha creato e sedimentato una sorta di gerarchizzazione nell'universo simbolico collettivo: il sesso maschile è la figura dominante della società e detiene un'egemonia nei confronti del sesso femminile, la quale supremazia è imposta, riconosciuta e tollerata in un circolo vizioso di persistenza. Pertanto, la costruzione identitaria individuale e sociale del genere maschile si caratterizza per preservare una società di stampo patriarcale, basata sulla determinazione del potere e l'affermazione della

virilità. L'identità del *maschio* si modella e trova conferme nelle dinamiche d'interazione asimmetrica, per mezzo delle quali può affermare il proprio potere. La prevaricazione e la violenza sulle donne si traduce quindi in un tentativo di mantenere ed avvalorare un virilismo storico che appare ormai in declino o in via d'estinzione (Bellassai, 2011).

In definitiva, esiste un profondo legame nella triade *identità, relazione e violenza*. Più specificamente, la supremazia simbolica dell'*identità sessuata* si concretizza mediante le dinamiche relazionali tra uomo e donna, nella cui interazione, solo attraverso la violenza, l'uomo riesce a definire e strutturare la propria *identità di essere e sentirsi maschio*.

1.1. Un genocidio invisibile: definire la Violenza sulle donne

Non esiste una definizione universale della *violenza sulle donne*, tuttavia, nel tentativo di circoscrivere il fenomeno, i diversi domini di interesse condividono a livello globale l'esistenza di una caratteristica comune nel definirla: il *genere*. Il *genere* rappresenta il presupposto causale che definisce una serie di azioni violente che vittimizzano la figura della donna.

In particolare, nel presente lavoro di tesi quando si fa riferimento al tema oggetto di studio si richiamano i principi socialmente riconosciuti e definiti dall'*articolo 3 della Dichiarazione Ufficiale della Convenzione di Istanbul* (2011), conosciuta anche come *Convezione del Consiglio d'Europa sulla prevenzione e lotta contro la violenza nei confronti delle donne e la violenza domestica*. Tale dichiarazione definisce la *violenza sulle donne* nei termini che seguono:

“La violenza di genere è una violenza dei diritti umani, una forma di discriminazione nei confronti delle donne e si intendono tutti gli atti di violazione di genere che determinano o sono suscettibili di provocare danno fisico, sessuale, psicologico o economico o una sofferenza alle donne, comprese le minacce di tali atti, la coercizione o la privazione arbitraria della libertà, sia nella vita pubblica che privata”.

La tematizzazione del fenomeno a livello europeo si delinea dunque attraverso un *file rouge* che sembra inesorabilmente mettere in relazione la determinante logica del *genere* e la sottesa e copiosa disuguaglianza ordinamentale di una

società che vittimizza le donne. Nello specifico, è il concetto di *genere* a rappresentare il fondamento che definisce e motiva la violenza contro le donne. Infatti, tale componente viene ampiamente sottolineata nelle sezioni *c)* e *d)* del suddetto articolo:

c) “con il termine ‘genere’ ci si riferisce a ruoli, comportamenti, attività e attributi socialmente costituiti che una determina società considera appropriati per donne e uomini;

d) “l’espressione ‘violenza contro le donne basata sul genere’ designa qualsiasi violenza diretta contro una donna in quanto tale, o che colpisce le donne in modo sproporzionato”.

Ne risulta quindi che il tema sia strettamente connesso a quell'*identità sessuata* a cui fa riferimento Bourdieu (1998) e che presuppone un bilanciamento di ruoli e poteri da essa dipendenti.

In questo senso, l'identità di ciascun individuo si caratterizza per mezzo di strutture genetiche chiamate *eterosomi* e la cui funzione è legata alla determinazione del sesso di qualsiasi essere umano. Sulla base della struttura anatomica quindi si definisce la costruzione di veri e propri *corpi sociali* attraverso cui si affondano le radici di una topologia sessuale del corpo socializzato, ovvero definito da un insieme di attribuzioni e comportamenti investiti di una significazione sociale. L'attribuzione di senso che viene conferita sia al genere maschile che a quello femminile si caratterizza per un'opposizione di significati, atteggiamenti e ruoli del tutto dicotomici, le cui peculiarità si connotano di una sorta di spessore semantico attraverso la sovradeterminazione di specifiche corrispondenze. La matrice identitaria governata dal *genere* in una prospettiva di visione e divisione è consacrata in un rapporto di causalità che vede inscrivere la legittimazione del dominio maschile. La costruzione arbitraria del biologico che vede il mondo esprimersi attraverso la divisione in *generi relazionali* racchiude in sé il succedersi degli eventi come preordinati ed indipendenti da qualsiasi altra finalità umana, compresi gli atti di violenza materiale e/o morale. È la teoria del genere ad essere quindi parte integrante di una cultura del genere, che legittima le intenzioni della violenza e che nelle definizioni funzionali e strutturali dell'apparato simbolico della società non fanno altro che rilegare in un angolo la parte ritenuta socialmente inferiore. Si tratta pertanto di un genocidio invisibile che nel testo integrale della *Convenzione di Istanbul* consente di far emergere l'atavico giudizio sull'inferiorità femminile come movente strutturale

della violenza. Le differenze tra i *generi* sono messe in risalto proprio nel tentativo di definire il fenomeno e di conseguenza richiamano inevitabilmente un processo storico di *eternizzazione* di modelli che sono incorporati e tramandati nel tempo. La perseveranza di questi modelli va quindi connessa con gli agenti e le istituzioni che da sempre lavorano per assicurare tali permanenze, a partire dal nucleo familiare. Infatti, nel testo preso come riferimento per definire la violenza sulle donne e per chiarire le forme e i domini in cui si realizza, si intende porre l'enfasi proprio sull'ambito domestico, poiché se alle volte i casi di violenza sono agiti da sconosciuti o conoscenti, nella maggior parte dei casi gli autori appartengono all'ambito parentale o del rapporto di coppia (FRA, 2014). La sezione *b*) della *Convezione del Consiglio d'Europa sulla prevenzione e lotta contro la violenza nei confronti delle donne e la violenza domestica* riconosce in particolare la violenza domestica nei termini che seguono:

b) "la 'violenza domestica' designa tutti gli atti di violenza fisica, sessuale, psicologica o economica che si verificano all'interno della famiglia o del nucleo familiare o tra attuali o precedenti coniugi o partner, indipendentemente dal fatto che l'autore di tali atti condivide o abbia condiviso la stessa residenza con la vittima".

L'abuso familiare o *Intimate Partner Violence* identifica spesso relazioni di coppia caratterizzate da comportamenti aggressivi di sopraffazione, controllo e subordinazione fisica, psicologica e/o economica della donna. Le dinamiche di violenza si consumano in nome di uno stereotipo di *genere* e vedono l'uomo nella veste di aggressore, nel tentativo di affermare e legittimare il proprio potere all'interno della relazione.

I casi di violenza domestica si concretizzano principalmente nei rapporti di coppia eterosessuali (Merzagora Betsos, 2006), nei quali si esaspera il concetto di *identità sessuata*, innescando un perverso meccanismo di dominio e possesso. L'egemonia maschile si realizza per lo più a causa di una ribellione femminile (Stark, 2010) alla consacrazione dell'oggettivazione conosciuta e riconosciuta nei corpi e negli *habitus* degli agenti, in uno schema di percezione, pensiero ed azione. Se da una parte, la forza dell'ordine maschile appare contrassegnata da una *normalità* che non ha bisogno di legittimazione, dall'altra parte questo equilibrio viene capovolto e rinnegato. Di conseguenza, il capovolgimento di una tale sedimentata asimmetria relazionale (Kaufman, 1997) e la lotta tra il polo della permanenza e quello del cambiamento

sconvolgono e creano smarrimento in un universo simbolico maschile che utilizza la violenza come strategia strumentale al rinforzo della propria identità, in una società di fondamento patriarcale. Il possesso invece si manifesta attraverso un incessante controllo sia materiale che della vita sociale della donna, che si protrae anche nel momento in cui la relazione coniugale o affettivo-sentimentale si interrompe. Infatti, la persecuzione del controllo talvolta continua a manifestarsi anche quando il rapporto è ormai cessato. In questo ultimo caso, si parla di violenza da parte dell'*ex-partner*.

La Convenzione esplicita inoltre al proprio interno l'esistenza di diverse forme di violenza che devono essere riconosciute nell'ambito della legislazione di ciascun paese firmatario in termini di reato. Nel dettaglio, il testo ritiene opportuno evidenziare e tipizzare una pluralità eterogenea di manifestazioni violente:

- *Violenza psicologica (art. 33)*: a cui afferiscono le denigrazioni, il controllo dei comportamenti, le strategie di isolamento e le intimidazioni. Rientra in questa categoria lo *Stalking (art. 34)*.
- *Violenza fisica (art. 35)*: graduata dalle forme più lievi a quelle più gravi, si concretizza nella minaccia di essere colpite fisicamente, l'essere spinte, afferrate, stratonate, percosse mediante l'utilizzo di oggetti e non, prese a calci, pugni o morsi, il tentativo di strangolamento, di soffocamento, di ustione e minaccia con armi.
- *Violenza sessuale compreso lo stupro (art. 36)*: si considerano in questa categoria le situazioni in cui la donna è costretta ad attuare o subire contro la propria volontà atti sessuali di diverso tipo: molestie (*art. 40*), tentato o avvenuto stupro, procedure degradanti o umilianti attuate per mezzo di pratiche minatorie. Non vengono rilevate le molestie verbali, il pedinamento, gli atti di esibizionismo e le conversazioni dal contenuto osceno.
- *Violenza economica*: rientrano in questa classe le limitazioni economiche, la sottrazione dello stipendio, la mancanza di autonomia di gestione delle finanze e la costrizione ad intraprendere iniziative economiche non volontarie.

La tassonomia delineata ha lo scopo di chiarire nel dettaglio le possibili manifestazioni della violenza, poiché alle volte, alcune forme risultano particolarmente subdole da non essere intercettate. Si tratta di un fenomeno controverso che si caratterizza di aspetti e dimensioni complesse. Infatti, rispetto alla violenza fisica che si realizza con la presenza di ecchimosi o lacerazioni, gli altri tipi di violenza sono più difficili da individuare, a causa della mancanza di quell'aspetto tangibile che invece si presenta nell'abuso fisico.

Non a caso, una delle ultime indagini condotte in Europa nel 2012 e pubblicate nel 2014 dall'*European Union Agency for Fundamental Rights (FRA)* mostra l'opportunità di stimare principalmente i casi di violenza fisica e non di altro tipo, soprattutto se esperita da terzi (il 67%) e quindi da soggetti non appartenenti alla sfera familiare. Si rileva infatti che in Europa il 31% delle donne (la cui età è compresa tra 15 e 74 anni) sia stata vittima di violenza fisica nel 2012 e che tredici milioni di donne che vivono in uno dei Paesi membri dell'Unione Europea (UE) abbia subito violenza fisica nel corso dei 12 mesi precedenti all'indagine del FRA.

La vessazione sessuale rappresenta poi l'aspetto più difficile da indagare, in quanto il livello di effrazione psico-fisica percepita è piuttosto elevato. Nel caso della violenza sessuale, la stima è di 3,7 milioni di donne che vivono nell'UE, delle quali il 5 % è stato costretto ad avere rapporti sessuali, il 6 % ha subito un tentativo di stupro ed un altro 12% è stato obbligato a prendere parte ad attività sessuali anche se non consenziente o per paura delle conseguenze (FRA, 2014). La violenza psicologica invece identifica l'aspetto più insidioso da studiare, in quanto si articola attraverso dinamiche differenti rispetto alla violenza fisica o sessuale. Esiste una linea di confine molto sottile tra ciò che è degno di tolleranza e quello che non lo è. Inoltre, tale soglia muta da individuo ad individuo, anche in base al proprio *background* esperienziale personale. Solitamente, la vessazione emotiva persegue lo scopo di condurre la vittima ad un annichilamento globale e viene eseguita in maniera metodologica con l'intenzionalità di rescindere ogni principio di autonomia, autostima, attitudine alla crescita culturale, economica ed interpersonale. Questa tipologia viene in gran parte agita in ambito parentale o sul posto di lavoro, mentre sono irrilevanti i casi in cui l'autore è un conoscente o uno sconosciuto. Mentre, nel complesso, in Europa, il 43 % delle donne ha subito una forma di violenza psicologica da parte del proprio *partner* (FRA, 2014).

Spesso inclusa nel dominio della violenza psicologica è la subordinazione

economica. Si tratta di un sistematico ed opprimente controllo dei beni economico-finanziari della vittima, attraverso la gestione della contabilità fino alla limitazione o negazione di gestione autonoma del proprio patrimonio personale. La violenza economica rappresenta un aspetto di una certa rilevanza, soprattutto nei casi in cui le donne non riescono ad allontanarsi dal proprio vessatore a causa della dipendenza economica.

Le categorie della violenza possono essere messe in atto singolarmente o simultaneamente, oppure possono essere realizzate ad intervalli ripetuti. La violenza può essere di tipo occasionale o episodica, oppure prolungata nel tempo a seconda della dimensione temporale che viene considerata.

La vittimizzazione può essere sia diretta che indiretta, ovvero si può essere direttamente o indirettamente esposti alla violenza. Nello specifico, l'esposizione alla violenza altrui rappresenta sul piano individuale un tipo di violenza emotiva detta *assistita* che può provocare anche conseguenze gravi, simili a quelle causate in chi è vittima di violenza in prima persona.

Di conseguenza, la violenza può essere categorizzata in due classi di appartenenza nella sua più ampia generalizzazione (che si tratti di violenza psicologica, fisica, sessuale o economica). Si parla pertanto di:

- *violenza diretta*: se il soggetto viene coinvolto in prima persona ed è vittima di una o più categorie di violenza;
- *violenza indiretta o assistita*: è una modalità di esposizione alla violenza di tipo secondario. Il soggetto coinvolto non è la vittima della violenza ma è il testimone oculare e uditivo di aggressioni fisiche o verbali reiterate nel tempo. Nella maggior parte dei casi si concretizza da e/o verso soggetti che costituiscono un riferimento educativo e/o affettivo.

Anche i contesti in cui si consuma la violenza possono essere diversi. Gli spazi che fanno sfondo possono essere sia interni che esterni, quali strade, parcheggi, giardini pubblici, strutture pubbliche, sedi professionali, mezzi di trasporto, dimore private, che siano di proprietà o di ospitalità.

La violenza dunque può essere agita sia fuori che dentro le mura domestiche, riconoscendosi come fenomeno dalla natura endemica (WHO, 1997), senza distinzione geografica o di *status* culturale. Il testo della Convenzione infatti tende proprio a riconoscere:

- a) La violenza che si verifica tra le mura domestiche tra cui:

aggressione fisica e mentale, abuso emotivo e psicologico, stupro e abuso sessuale, incesto, crimini commessi in nome dell'onore, matrimoni forzati, mutilazioni genitali e altre pratiche tradizionali dannose per le donne.

- b) La violenza all'interno della comunità generale in termini di stupri, abusi sessuali, molestie sessuali e intimidazioni sul lavoro, nelle istituzioni o altrove, ed ancora il traffico di donne a fini dello sfruttamento sessuale, economico e del cosiddetto turismo sessuale.
- c) La violenza perpetrata o condonata dallo Stato o dai suoi funzionari.
- d) La violazione dei diritti umani delle donne in situazioni di conflitto armato come presa di ostaggi, sfollamento forzato, stupro sistematico, schiavitù sessuale, gravidanza forzata, sfruttamento sessuale o economico.

La chiarificazione dei contesti rappresenta una maggiore specificazione degli ambiti della violenza anche se la diffusione di alcune forme dipende dal contesto geografico che viene considerato.

Ad esempio, la violenza perpetrata in situazioni di conflitto armato è prevalente in aree geografiche del sud-est asiatico, dell'Europa orientale o dell'Africa meridionale. La violenza della dote (UN Women, 2012) è invece maggiormente riconducibile ad India, Bangladesh, Sri Lanka e Nepal, mentre il lancio dell'acido (AP, 2000; De Castella, 2013; Vig, 2014) è tipico del sud-est asiatico. L'omicidio d'onore (Kirti *et al.*, 2011) è prevalentemente agito in Medio Oriente e in Asia meridionale, mentre le mutilazioni genitali (UNICEF, 2013) sono ascrivibili all'Africa, al Medio Oriente e ad alcune parti dell'Asia, mentre la compravendita delle spose rappresenta una consuetudine che appartiene dell'Africa sub-sahariana e dell'Oceania.

1.2. Fuori e dentro le mura domestiche

Nel tentativo di definire quindi la violenza nelle sue forme più disparate e con l'intenzione di porre l'accento sull'intensificazione di alcune tipologie che afferiscono in prevalenza a specifiche aree territoriali, risulta necessario considerare gli ambiti in cui si realizza la violenza. In particolare, un aspetto per troppo tempo sottostimato e solo da qualche decennio divenuto oggetto di riflessione scientifica è proprio l'analisi del contesto in cui si relazionano la

vittima e l'aguzzino. Talvolta, l'azione violenta è agita da autori sconosciuti o parzialmente conosciuti dalle vittime e altre volte viene messa in atto all'interno delle pareti di casa. Tuttavia, la violenza endofamiliare rappresenta l'aspetto più insidioso da studiare, perché essa racchiude in sé un disagio ed un limite diagnostico del problema, a causa degli elementi affettivo-parentali che caratterizzano la relazione tra la vittima e l'autore della violenza.

1.2.1. La violenza nei perimetri nascosti delle pareti di casa

Solo di recente, la lettura scientifica si è dedicata allo studio di relazioni affettive, siano esse di natura consanguinea o non, che sono patologiche o si caratterizzano di un amore narcisistico, insano, contorto, malato.

La violenza che si consuma nei perimetri nascosti delle pareti di casa è principalmente riconducibile alla sopraffazione del marito sulla moglie, del compagno sulla compagna e della subordinazione tra conviventi legati da un rapporto di intimità o di legge. Tuttavia, la sfera familiare può estendersi anche ad una figura maschile diversa, quale un padre, un figlio, uno zio o un nonno.

Oggi le cronache connotano la violenza domestica spesso fatale come *delitto passionale, amore criminale* o *crimini segreti* (Ponzio, 2004) proprio con lo scopo di esplicitare questa tipologia di violenza che si instaura tra soggetti appartenenti allo stesso nucleo familiare. Infatti, nel 1996, il relatore speciale dell'ONU (Organizzazione delle Nazioni Unite) ha definito la violenza parentale come:

“Violenza che si consuma all'interno della sfera privata, generalmente tra individui legati da un vincolo di intimità, di sangue o legge”.

I soprusi contro il *genere* possono compiersi non solo sulle donne di età adulta ma anche sui minori di sesso femminile, già dal concepimento, attraverso l'aborto selettivo. Infatti, la violenza può originarsi dalla pre-nascita per poi continuare durante tutto il percorso di vita della vittima, in forma sia diretta che indiretta. Quindi, quando il nucleo familiare si trasforma nel palcoscenico della violenza sono inevitabilmente coinvolti anche i minori.

Essenzialmente, se il coinvolgimento del minore è diretto si tratta di situazioni che vanno oltre il dispotico atteggiamento del maltrattamento fisico, dell'aggressione psicologica o dell'indifferenza affettiva. La stragrande maggioranza degli episodi di violenza di genere sui minori evidenzia abusi

principalmente di natura sessuale consanguinea e quindi l'incesto. Il diritto biologico di essere uomo in questo caso scaturisce un principio di dominio anche sul minore che appartiene al nucleo familiare. Se il coinvolgimento è indiretto invece, i minori assistono alla violenza di genere esercitata nei confronti di una figura affettiva di riferimento, che solitamente è la madre. La violenza assistita rappresenta un'importante sezione di ricerca e attraverso l'indagine del FRA (2014) si stima infatti che in Europa, nel 2012, i casi di violenza indiretta siano il 73%. Tuttavia, principalmente frequente tra le mura domestiche è la violenza connessa alle relazioni di coppia, le cui origini conflittuali sono molto antiche. Fin dall'epoca di Tristano ed Isotta infatti va affermandosi il mito occidentale dell'amore assoluto che per alimentarsi necessita di impedimenti mentali e reali. Dalla letteratura, al cinema, alla mitologia, alla cronaca, si rintraccia un'idealizzazione di conflittualità e dominio, dove inevitabilmente deve esistere la supremazia dell'uno sull'altro, identificando la forma più diffusa (Buchili, 2006) delle relazioni di coppia. Il funzionalismo del rapporto uomo-donna si colloca quindi in una prospettiva di passato e presente ed in un *continuum* spazio-temporale che non conosce mutamento e che sembra legare indissolubilmente il trascorso con la contemporaneità. Le modalità relazionali all'interno delle coppie sono state da sempre caratterizzate dal conflitto, dal bilanciamento di dominio e di potere, dal sessismo intrinseco nel DNA maschile che si legittima celandosi dietro effimeri contenuti di natura sentimentale. Talvolta il movente è da rintracciare nei rancori, nelle vendette, nei contrasti sulla gestione del patrimonio o sull'educazione dei figli, in probabili relazioni extra-coniugali ma nella maggior parte dei casi a muovere l'azione violenta è la necessità di affermare il dominio su di un corpo che sfugge via dalle regole del possesso individuale, occupando uno spazio che appare infinito per chi pretende di tenerlo sotto controllo. La cronaca contemporanea offre quasi quotidianamente ricostruzioni di vita drammatiche, dove il cosiddetto *delitto passionale* nasconde solo una crescente ossessione amorosa che mette in luce la dicotomia percettiva di odio e amore che convive nell'universo simbolico maschile della relazione di coppia. Ecco il motivo per cui i maltrattamenti nelle forme più o meno convenzionali trovano ampio spazio tra le mura domestiche e come da manuale il carnefice nel 90% dei casi ha le chiavi di casa o se suona gli viene aperto. Si tratta di una lama invisibile perché da una parte le pareti domestiche rappresentano storicamente l'emblema della sicurezza per ciascun individuo, dall'altra parte è anche vero che "una donna per essere aggredita brutalmente deve proprio entrare

nell'istituzione più sacra: la famiglia" (Dobash & Dobash, 2015). Infatti, i risultati emersi dall'indagine del FRA (2014) hanno indicato che il 34% del campione sottoposto ad esame, nonché un terzo delle donne europee intervistate è stata vittima di più forme di violenza allo stesso tempo da parte del proprio *partner*. Le cause emerse dallo studio sono riconducibili a specifiche aree tematiche partendo dalla possessività e gelosia, per poi passare alla facoltà punitiva di compensare le aspettative maschili sull'operato domestico ed infine nel tentativo di affermare la propria autorità nella coppia. La dinamica comunicativa è quella del perenne conflitto e la violenza verbale talvolta può essere associata alla violenza fisica e/o sessuale. Sono reiterati gli attacchi all'immagine, all'aspetto fisico e alle capacità mentali della donna. Si tratta di un'offensiva alla progettualità personale, di biasimo connesso al ruolo genitoriale e alla mancanza di sostegno morale ed economico. I rapporti si basano essenzialmente su un patologico monitoraggio delle libertà di espressione, interazione sociale e di progresso socio-culturale.

Tali effrazioni psicofisiche vengono inizialmente minimizzate dalle donne ed è solo la reiterata latente aggressività che induce loro a recepirne la reale gravità. Il dispotismo familiare porta le donne a rimanere intrappolate in un rapporto insano in nome della paura, del ricatto dell'eventuale alienazione parentale e talvolta, nei meccanismi più contorti, in nome dell'amore. Una tale contestualizzazione può spiegare il motivo per cui il numero delle denunce di violenza familiare è estremamente ridotto. In Europa, nel 2012, solo nel 31% dei casi, le donne hanno denunciato di aver subito violenza fisica in ambito domestico e nel 39% casi di essere state vittime di violenza sessuale, mentre un altro 32% non ne parla con nessuno (FRA, 2014). Spesso, le donne che subiscono violenza coniugale rimangono legate al proprio *partner* in quanto l'ambivalenza della relazione oscura la razionalità, in una trappola che esorta alla vocazione affettiva. Tuttavia, in situazioni caratterizzate da un crescendo di violenza, l'epilogo è quasi sempre di tipo omicidiario. In particolare, l'esercizio della violenza domestica risulta essere una sorta di *spirale*, un ciclo che si articola in fasi convenzionali, quali la crescita della tensione, l'esplosione della violenza, la fase di latenza e lo scarico della responsabilità (Walker, 1979). Gli studiosi affermano che nella prima fase, la vittima tenta in ogni modo di evitare la situazione conflittuale, reprimendo ogni sorta di paura per sedare l'ira violenta del *partner*. Nello stadio evolutivo sussistono due modalità di reazione: la prima prevede uno stato di inerzia d'innanzi all'impeto della violenza, mentre la seconda consiste nella ricerca di un nascondiglio per evitare la vessazione. La

fase di latenza prevede che passata la fase acuta del maltrattamento, la persona violenta mostri spesso segni di pentimento attraverso tattiche di manipolazione, rendendo così la vittima progressivamente succube del proprio atteggiamento. Nell'ultima fase invece l'aggressore induce la vittima ad assumere la responsabilità del maltrattamento, in modo che questa inizi a tessere un angoscioso dialogo con sé stessa, nel tentativo di convincersi di un fraintendimento dell'accaduto o peggio di esserne responsabile. La spirale della violenza coniugale è il sintomo di un immaginario collettivo nel quale le donne sono inconsapevolmente cresciute, nutrendosi di principi di subordinazione e convincendosi di essere *destinate*, in quanto donne, ad una condizione di mortificazione costante della propria identità, nel corpo e nella psiche. Le donne diventano quindi complici inconsapevoli di un retaggio socio-culturale che alimenta tale accondiscendenza. Le statistiche ufficiali dei paesi dell'UE mostrano una percentuale compresa tra il 30-32% per indicare la violenza agita dal *partner* in in zone geografiche come Finlandia, Danimarca e Lettonia, mentre la diffusione in paesi come Austria, Croazia, Polonia, Slovenia e Spagna è del 13%. Anche il tasso di prevalenza della violenza esperita da terzi è piuttosto considerevole, si parte infatti da un massimo percentuale del 34%-40% in Svezia, Paesi Bassi e Danimarca fino ad arrivare ad un minimo di 10%-11% in Portogallo, Grecia e Polonia (FRA, 2014). Il dato numerico europeo mostra inoltre una correlazione positiva tra il fenomeno della violenza di genere messa in atto dal *partner* e quella generata da un individuo che non appartiene alla sfera intima della vittima.

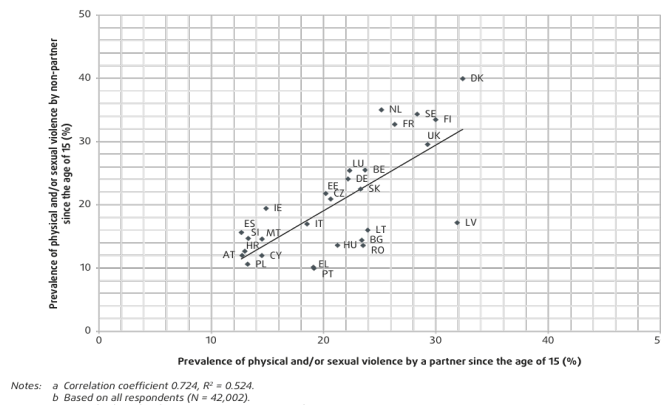


Figura 1.1. Correlazione tra la prevalenza della violenza in cui l'autore è il partner e in cui l'autore è diverso dal partner nell'UE.

Fonte: FRA, 2014 (Database, 2012).

In altre parole, esiste un legame statistico positivo tra la prevalenza della violenza fisica e sessuale agita dal *partner* e la prevalenza della violenza fisica e sessuale esercitata da soggetti diversi dal *partner*.

1.2.2. L'abuso sessuale

La sessualità può essere anch'essa impiegata come forma di potere ginocida (Danna, 2007) sia all'interno che al di fuori delle mura domestiche. Nello specifico, riferirsi all'abuso sessuale indica una serie di varianti, al sesso connesse, che includono una certa trasversalità semantica. Appartiene infatti a tale categoria qualsiasi attività di natura sessuale e riproduttiva che non consente alla donna la libertà di dominio sul proprio corpo. Rientrano tra le violazioni di natura sessuale lo *stupro* in senso stretto, l'aborto forzato, la sterilizzazione, attività riproduttive in assenza di consenso e la prostituzione o lo sfruttamento.

Per eredità culturale, la donna è considerata preda e l'uomo cacciatore (*Ibidem*, 2007), pertanto la violenza sessuale si caratterizza fundamentalmente in una generalizzazione di rapporto sessuale non consenziente, meglio noto come *stupro*. Il termine è stato acquisito dagli antichi Romani che lo utilizzavano per indicare la condivisione intima al di fuori del matrimonio, mentre oggi la definizione copre un'ampia gamma di determinazioni.

Nell'aspetto sessuale il principio di inferiorità della donna che il sistema rituale ratifica e amplifica viene rafforzato. In particolar modo, la disimmetria non fa che irrobustire gli *habitus* che rappresentano la figura maschile come soggetto e quella femminile come oggetto, e se l'uomo è l'agente, la donna è uno strumento che si materializza in un'economia di scambi simbolici (Bourdieu, 1998). Alla radice dell'ordine sociale emerge una rappresentazione della donna che si reifica o che si trasforma in uno strumento capace di contribuire all'accrescimento del capitale simbolico maschile. L'interpretazione puramente "economicistica" attribuisce al corpo della donna una natura strumentale simbolica, come oggetto intercambiabile e circolante, al pari della moneta (Dardigna, 1980). Pertanto, da un punto di vista sessuale, la donna può essere trattata come *merce di scambio* in un segno di comunicazione che si può interpretare in termini di dominio sul corpo. La mercificazione della donna è maggiormente evidentemente nella contemporanea fase storica di globalizzazione, nella quale dominano le forze del cosiddetto capitale privato,

connettendosi a nuove forme di mercato. Nello specifico, il mercato si interfaccia al sesso in un vero e proprio *traffico di donne*, ossia una transizione di giovani ragazze provenienti da località affette da gravi problemi socio-economici, le quali vengono obbligate a prostituirsi mediante attività coercitive. Si tratta di una vera e propria schiavitù sessuale che viene definita *Tratta umana* e solo in Italia è rappresentata da oltre 50 mila donne, un terzo delle quali minorenni, che producono un giro d'affari annuale stimato in 5-7 miliardi di euro. La *Tratta* sessuale è spesso agita da terzi ma alle volte sono gli stessi consanguinei o il proprio *partner* a fare delle vittime una fonte di guadagno.

Altre forme di violenza connesse alla sfera sessuale sono la costrizione alla riproduzione o all'aborto, la sterilizzazione forzata ed il divieto di contraccezione. La coercizione riproduttiva rappresenta un comportamento violento e manipolatore nei confronti del diritto delle donne alla riproduzione e viene prevalentemente perpetrata in ambito domestico. In particolare, nell'*art. 39 della Convenzione di Istanbul (2011)*, la violazione della salute e dei diritti sessuali e riproduttivi sono espressamente proibiti a livello internazionale:

“Violazioni della salute e dei diritti sessuali e riproduttivi delle donne, come sterilizzazioni forzate, aborto forzato, gravidanza forzata, criminalizzazione dell'aborto, la negazione o il tempo dell'assunzione sicura e delle cure post-aborto, la continuazione forzata di gravidanza, l'abuso e il maltrattamento di donne e ragazze che cercano informazioni sulla salute sessuale e riproduttiva, beni e servizi, sono forme di violenza che fanno parte delle circostanze, possono ammontare una tortura o trattamento crudele, inumano o degradante”.

Anche l'interruzione volontaria di gravidanza rientra tra le questioni sulla libertà sessuale delle donne, il quale diritto è stato ampiamente dibattuto nel corso degli anni, basti pensare all'approvazione della legge sull'aborto, la cui giurisdizione di ogni paese europeo delinea dettagli importanti e diversificati. Si tratta di “piccole conquiste” che nel corso del tempo le donne hanno ottenuto, malgrado sussistano ancora molti, troppi casi, di coercizione riproduttiva.

La violenza sessuale è una delle principali forme di abuso in cui le donne sono rlegate ad un'assenza di dignità, intelletto, capacità di discernimento e al contrario sono rapportate ad una mera fisicità. I tempi e le strutture hanno generato e avallato una prospettiva femminile che è ascrivibile alla sola *corporeità sessuata*, tale per cui la donna è solo un oggetto di consumo. Martha Nussbaum (1999) infatti sottolinea proprio la funzione strumentale del corpo

della donna, la negazione dell'autonomia, la sua fungibilità e violabilità. La donna è intesa come una proprietà senza soggettività e nella sfera sessuale, questa prospettiva raggiunge l'apice. La strumentalizzazione si caratterizza di una assoluta reificazione, in un'oggettivazione senza anima, senza voce, senza volto. Una tale *disumanizzazione* della donna si realizza ancor di più nell'ambito della pornografia, che secondo Kimmel (1990) viene creata dagli uomini per compiacere le fantasie sessuali degli uomini. L'autore infatti fa riferimento alla necessità di nutrire un ego maschilista che privilegia la produzione della pornografia e l'utilizzo della donna per il "piacere" privo di intimità emotiva. Nella sessualità trova dunque spazio la corporeità sessuata a cui fa riferimento la Nussbaum che nega alla donna una qualsiasi visione umana. Al contempo però, la violenza sul piano sessuale mette in evidenza un anche nell'universo maschile un distacco di percezione umana, poiché l'uomo prende le distanze dalle proprie emotività e si lascia guidare dall'istinto, e più propriamente dall'istinto animale verso il piacere.

Studi condotti su individui colpevoli di avere compiuto violenza sessuale evidenziano proprio la percezione di ritenere le proprie vittime prive di identità e anzi responsabili dell'accaduto, ignorando completamente le ripercussioni fisiche e psicologiche della donna. Nel rapporto sulla violenza nel mondo stilato dall'Organizzazione mondiale della sanità (OMS) e sviluppato da 160 esperti di settore in tutto il mondo si legge infatti:

"[...] Tra i principali fattori che determinano gli uomini sessualmente violenti si rintraccia un eccessivo senso di mascolinità e misoginia e la preferenza per un legame che non sia emotivo. Le donne sono delle avversarie che devono essere conquistate e sfidate. Un ambiente di crescita violento (la cosiddetta violenza assistita) è senza dubbio un aggravante del fenomeno e talvolta l'atto violento può rappresentare un'attività del tutto ricreativa che non ha un legame intrinseco di rivendicazione o significato [...]".

Pertanto, il distacco dal senso umano o la rivendicazione della mascolinità tendono a sottolineare l'egemonia identitaria che richiama l'identità sessuata tra gli individui, tracciando quello che Kelly e Johnston (2008) definiscono come un bilanciamento di potere tra i due sessi. Un potere che viene esercitato attraverso la sopraffazione di un sesso sull'altro e proprio mediante l'azione sessuale in senso stretto. Talvolta, la sopraffazione sessuale della donna non è esercitata da un singolo individuo ma gli autori della violenza sono due o più di due; in questo caso si parla di *violenza di gruppo* o *stupro da branco*.

Tale tipologia si verifica con maggiore prevalenza nei *campus* universitari ed è un problema piuttosto diffuso soprattutto nelle giovani generazioni. A livello mondiale, è particolarmente frequente negli Stati Uniti d'America (Krebs *et al.*, 2007) mentre è presente ma con una minore incidenza in Canada, Regno Unito e Nuova Zelanda. La violenza sessuale di gruppo ha tuttavia un'eredità culturale molto antica ed è riconducibile ai periodi storico-bellici nei quali i soldati violentavano in gruppo le donne del *luogo conquistato*. Questa tipologia di violenza è passata alla storia con il nome di *stupro di guerra* (Amnesty International, 2006) e mentre in Occidente la pratica è stata abolita e riconosciuta come reato dopo la seconda guerra mondiale, in molti paesi ancora oggi lo stupro di guerra è posto in essere, taciuto e legittimato.

Con lo scopo di valutare i fattori che determinano maggiormente gli episodi di violenza sessuale, i dati dell'*International Crime Victims Survey* propongono le statistiche giudiziarie delle vittime di più di 70 paesi di tutto il mondo.

Dalla relazione riassuntiva si evince che lo *status* culturale ed occupazionale femminile sia spesso correlato all'aggressione sessuale. In altre parole, più il grado culturale di una donna è alto e più risulta elevata la sua vittimizzazione. L'interpretazione di questo dato può essere ricondotta a due motivazioni: da una parte, la necessità di infliggere una pena alla donna per il successo raggiunto, da un'altra parte, l'interpretazione presume che il dato sia più elevato in quanto le donne maggiormente istruite tendono a denunciare di più.

Altri fattori vi sono riconducibili, come l'uso di sostanze stupefacenti o alcoliche, ed in casi più rari la presenza di patologie psichiatriche (Ventimiglia, 1989). Difatti, alcuni studi (*Ibidem*, 1989) ritengono che nella maggior parte dei casi, la violenza sessuale sia direttamente proporzionale alla fragilità identitaria dell'uomo, in un incessante conflitto tra la proiezione del sé e la propria identità cosciente.

Nel considerare il fenomeno tuttavia bisogna tener conto che le statistiche ufficiali soffrono notevolmente di un problema di sottostima, infatti la percentuale di donne propense a denunciare una violenza sessuale è molto bassa. In Europa, nel corso del 2012, solo il 39% delle donne ha denunciato una violenza sessuale inflitta dal *partner* e solo il 30% quella subita da uno sconosciuto (FRA, 2014). I casi di abuso sessuale costituiscono la modalità di violenza in minor misura portata alla luce, in quanto la vittimizzazione spesso passiva della donna che rimane inerme per la paura o lo *shock* viene scambiata per una forma di corresponsabilità nell'esasperato garantismo giudiziario che

vige in Europa. L'abuso di natura sessuale, che si tratti di stupro in senso stretto o violenza in forma di molestia, è di frequente esercitata sul posto di lavoro. In particolare le varie forme di violenza che possono essere esercitate sul luogo di lavoro sono conosciute in letteratura sotto il nome di *mobbing*.

1.2.3. *Il mobbing sul posto di lavoro*

Nell'accezione più comune del termine, il *mobbing* identifica una moltitudine di aggressioni di natura psico-fisica nei confronti di un'unica persona prescelta come vittima, da parte di un singolo individuo o un gruppo di individui. Si tratta di un fenomeno che non può essere circoscritto ad un'univoca definizione ma si caratterizza per una flessibilità di casi a cui può riferirsi.

La nozione di *mobbing* viene introdotta agli inizi degli anni Settanta dall'etologo Konrad Lorenz per descrivere l'atteggiamento aggressivo tra individui della stessa specie e per indicare l'accanimento ripetitivo nei confronti della stessa persona. Nonostante il carattere "aperto" del fenomeno, in cui rientrano una pluralità di azioni mobizzanti e motivazioni che spingono alla prevaricazione della vittima, in questo contesto si approfondisce la persecuzione in nome del *genere* sul posto di lavoro. In altre parole, si mira a contestualizzare il *mobbing* agito in ambito professionale e la cui vittimizzazione è femminile. Infatti, se nel complesso, esso si qualifica come un aspetto comportamentale connesso all'interazione tra i soggetti, nello specifico la casistica giurisprudenziale conferma che le situazioni di *mobbing* sono prevalenti soprattutto nei rapporti di lavoro, di tipo subordinato. Correlato al settore lavorativo verso la fine degli anni Ottanta dallo psicologo svedese Heinz Leyman (1990) identifica proprio una condizione di atteggiamenti molesti volti a produrre danneggiamenti pluri-offensivi in ambito professionale.

Secondo una rilevazione statistica effettuata nel 2016 in Italia dall'Istat (Istituto Nazionale di Statistica), la violenza di genere nel circuito professionale è in netto aumento e si stima che 1 milione 404 donne (8,9%) abbiano subito molestie fisiche o ricatti sessuali sul posto di lavoro. In particolare, le categorie professionali più colpite sono quelle delle impiegate (37,6%) o lavoratrici nel settore del commercio e dei servizi (30,4%). La vittimizzazione sul posto di lavoro riguarda anche donne impegnate in attività professionali scientifiche e tecniche (20%) o nel lavoro domestico (18,2%).

Le manifestazioni mobizzanti nei confronti delle donne possono essere diverse, quali osservazioni sulle caratteristiche corporee, sguardi o gesti allusivi, contatti non desiderati, ricatti ed attenzioni moleste. Tra le pratiche maggiormente messe in atto vi è il ricatto, spesso riconducibile alla richiesta di disponibilità sessuale in cambio di assunzioni, progressioni di carriera o per il mantenimento del posto di lavoro. Si tratta di una tipologia di violenza attuata in modo metodologico e talvolta seguendo logiche strategiche che portano la vittima ad abbandonare il posto di lavoro, pur di far cessare le molestie. L'atteggiamento inopportuno che viene messo in atto da colleghi o superiori è spesso la conseguenza di un rifiuto compromissorio da parte della donna ad una o più richieste poco deontologiche. Nella prassi, esiste una forma di *mobbing* detta di tipo *gerarchico* se esercitata da superiori che in letteratura è nota anche come *bossing*, oppure se messa in atto da colleghi di pari grado è di tipo *orizzontale*.

Dagli inizi degli anni Ottanta, il *mobbing* rappresenta un fenomeno di interesse pluridisciplinare, soprattutto per quel che concerne le scienze sociologiche e psicologiche, le quali nel tentativo di inquadrare il tema, hanno introdotto nelle definizioni il pregiudizio alla personalità o salute del mobizzato (Leymann, 1993), proponendo modelli che descrivono il fenomeno in fasi. Specificamente, Ege (2001) ha illustrato un modello che si origina da una fase zero, nella quale il soggetto *mobber* determina una situazione di conflitto fisiologico al solo scopo di elevarsi sugli altri. Nella prima fase invece, il conflitto è mirato nei confronti di una persona prescelta, dirigendo verso di essa una conflittualità generale o per richiedere prestazioni di natura sessuale in cambio di favoritismi. La seconda fase del modello definisce il generarsi di disagio e fastidio nella vittima, nella cui terza fase inizia a manifestare disturbi psicosomatici. Nella quarta fase, il soggetto mobizzato tende ad allontanarsi dal posto di lavoro, assentandosi ripetutamente per malattia e dando l'impressione di essere una persona poco responsabile, destando così sospetto nei responsabili dell'ufficio amministrativo.

La quinta fase assume un ruolo quasi decisivo, poiché lo stato di salute soprattutto mentale della vittima tende progressivamente ad aggravarsi per poi raggiungere il culmine nell'ultima e sesta fase del modello, nella quale la vittima si esclude dal mondo del lavoro presentando le dimissioni o una richiesta di prepensionamento. Nei casi più gravi invece l'ultima fase si manifesta attraverso gesti estremi, come ad esempio il suicidio.

Il *mobbing* a sfondo sessuale non è tuttavia configurato come reato nella fattispecie concreta in Europa. Nel luglio 2011, il Parlamento Europeo con il

Libro Verde ha introdotto il dibattito in sede comunitaria, senza tuttavia porre l'enfasi sul *mobbing* di genere, il quale riferimento non è stato recepito in diversi ordinamenti giuridici, compreso quello italiano. Inoltre, la risoluzione non è stata seguita da alcuna direttiva europea che obbligasse gli Stati membri a legiferare una legge specifica sul *mobbing*. In giurisprudenza infatti si tende a far confluire la tematica ai testi normativi concernenti la discriminazione di genere o nel caso sussistano elementi probatori, è concesso di avvalersi di indicazioni in materia di lesioni personali gravi e colpose. Il problema del *mobbing* sul posto a carattere sessuale è comunque fortemente sentito in Europa e i *partner* della *Business Europe* hanno siglato un accordo proprio in materia di riconoscimento della violenza sul luogo di lavoro.

1.2.4. L'azione persecutoria dello Stalking

Negli ultimi anni una nuova forma di violenza è stata riconosciuta: lo *Stalking* o *sindrome del molestatore assillante*, la quale definizione nel lessico scientifico rimanda alla determinazione di violenza psicologica. Più propriamente, il termine anglosassone *stalking* mira a coprire l'area semantica dell'intrusione relazionale recidiva nella vita privata di un soggetto da parte di un altro soggetto detto *stalker*. L'etimologia di tale definizione è di interesse criminologico, psichiatrico, psicologico, sociologico e normativo-legislativo, e deriva dal verbo *to stalk*, ossia *fare la posta*, camminare furtivamente e con circospezione. La scelta etimologica ha l'obiettivo di indicare una moltitudine di comportamenti molesti ininterrotti e di intrusioni ripetute nella vita privata della vittima, tra cui la ricerca di un contatto, appostamenti nei pressi del suo domicilio o degli ambienti comunemente frequentati. Nella lingua inglese oltre a *stalking* si utilizzano sinonimi come *obsessional harassment*, *criminal harassment*, *obsessional following*, *obsessional relational intrusion*.

Qualunque sia il termine utilizzato per definirlo, lo *stalking* si differenzia dalla semplice molestia o violenza psicologica per l'intensità, la frequenza e la durata, oltre che per la variegata strategia comportamentale. Non occorre essere in possesso di particolari strumenti, poiché chiunque può diventare o essere uno *stalker*. Sebbene sia stato riconosciuto in tempi relativamente recenti, l'azione persecutoria rappresenta un fenomeno non effettivamente nuovo. Infatti, in un passato non troppo remoto i rapporti tra uomini e donne

erano legittimati dal possesso e dal controllo, di conseguenza le azioni di intrusioni ed effrazione emotiva e relazionale di un uomo nella vita di una donna erano considerati *routine* e non suscitavano particolare scandalo. Solo negli ultimi anni, la gravità del problema ha assunto una certa importanza e nel tentativo di inquadrare il fenomeno nella realtà quotidiana e nei contesti di vita classici e con lo scopo di analizzarne evoluzioni e possibili cause, sono molti gli studiosi e le istituzioni che si focalizzano sul problema, tra cui lo psicologo Ege, interessato alle dinamiche delle condotte persecutorie di *mobbing* e *stalking*, l'*FBI*, i governi di diversi Stati e le associazioni di psicoterapia cognitiva.

Attraverso un'analisi globale della criminogenesi della fenomenologia, l'attore protagonista, quale lo *stalker* è in minima percentuale un estraneo ed il repertorio messo in atto appartiene quasi sempre ad un conoscente della vittima. Può trattarsi di un amico, un collega, un *ex-partner* che agisce spinto dal desiderio di ripristinare lo *status* precedente o per vendicarsi di un torto subito. Più propriamente, si tratta di una relazione patologica in cui lo *stalker* mette in atto un comportamento assillante, intrusivo ed indesiderato di approccio con la vittima, compromettendone la sua salute psichica, fisica e sociale (Caretti e Craparo, 2004).

“Lo stalking è una forma di agguato mentale in cui l'aggressore ripetutamente, inavvertitamente e violentemente irrompe nella vita della vittima con la quale non ha relazioni di sorta. In aggiunta, ogni singolo atto di aggressione non può essere la causa del trauma se non considerati tutti insieme” (Royakkers, 2000).

Nello specifico, per definire lo *stalking* si fa ricorso a sette parametri da Ege (2001) menzionati, di cui il primo è lo spazio d'azione in cui si compie la persecuzione. Talvolta si può parlare di *stalking occupazionale*, che a differenza del *mobbing* che vede nell'ambiente lavorativo solo la causa, si estende anche negli ambienti privati della vita della vittima (De Simone, 2013). Il secondo criterio è quello della frequenza, che l'autore determina in un'azione di persecuzione a settimana, per almeno tre mesi. Tuttavia, questo aspetto non è pienamente condiviso da molti Stati e nei sistemi di ordine giuridico non si tiene realmente conto del lasso di tempo in cui è avvenuta la reiterazione ma si considera la reiterazione stessa che ha eventualmente portato ad alcune specifiche conseguenze nella vita della vittima. Il terzo parametro è strettamente connesso a quello precedente e si riferisce alla durata della persecuzione, mentre il quarto definisce il tipo di azione esperita: violenza

fisica, psicologica, sociale ed altro. Il quinto criterio è il divario che sussiste nella relazione tra la vittima ed il persecutore; infatti la vittima si trova in una posizione di svantaggio rispetto allo *stalker* che conosce ogni sua abitudine e mette in atto una serie di azioni che ne destabilizzano il suo equilibrio psico-fisico. In tal senso, si considera il sesto parametro che mette proprio in evidenza la condotta del persecutore, il quale passa da una prima fase di relazione emotiva conflittuale (legame passato interrotto o immaginario nella sua mente) ad una fase di pedinamenti, tentativi di intrusione a casa, a lavoro, nei luoghi di svago della vittima. Infine, incapace di accettare il rifiuto spesso il persecutore cerca un contatto di scontro con la vittima, il quale conflitto si conclude, nei peggiori dei casi, con l'omicidio della persona perseguitata. L'ultimo aspetto che l'autore considera è il motivo che spinge lo *stalker* a perseguitare la propria vittima: in primo luogo si può indicare la ricerca di un legame affettivo alimentato in via del tutto unilaterale, mentre in seconda battuta si tratta di vendetta o solo per il sadico desiderio di vedere soffrire la propria vittima.

Lo *stalking* affettivo-relazionale che viene agito da parte di *un ex-partner* o di un conoscente infatuato ha spesso lo scopo di maneggiare sentimenti estremi che si polarizzano in amore e odio. Nello specifico, il soggetto-oggetto del proprio interesse assume nella mente del persecutore l'immagine di un qualcosa da possedere. In questo caso, lo *stalker* tenta di imporre la propria presenza nella vita della vittima, perché intende stabilire con lei una qualche relazione. Tuttavia, attraverso il rifiuto della vittima, la figura della donna desiderata sfugge ad ogni aspettativa e tentativo di controllo, così che il sentimento che il persecutore crede di provare si trasforma in rabbia ed odio, in una serie di atteggiamenti violenti che non conoscono limiti spazio-temporali.

Il Centro Presunti Autori–Unità Analisi Psico-Comportamentale dell'Osservatorio Nazionale sullo *Stalking* (CPA) che studia tale fenomeno, ha proposto un *identikit* psico-comportamentale dello *stalker*. In particolare, secondo il CPA oltre il 50% dei persecutori ha vissuto almeno una volta nella vita l'abbandono, la separazione o il lutto di una persona cara che non è riuscito a razionalizzare.

Lo studio condotto dalla Sezione Atti persecutori del Reparto Analisi Criminologiche dei Carabinieri identifica più nello specifico cinque categorie di *stalker*:

- Il *risentito*: caratterizzato da rancori per traumi affettivi pregressi, ritenuti privi di giustificazione (tipicamente un *ex-partner*).

- Il *bisognoso d'affetto*: desideroso di convertire un ordinario rapporto in una relazione sentimentale. Questo tipo di persecutore insiste e fa pressione nella convinzione che prima o poi l'oggetto delle sue attenzioni possa convincersi a considerarlo.
- Il *corteggiatore incompetente*: opera *stalking* in genere di breve durata ed i suoi comportamenti risultano opprimenti ed invadenti principalmente per "ignoranza" delle modalità relazionali. In questo caso, egli arreca un fastidio praticamente preterintenzionale.
- Il *respinto*: è il tipico soggetto rifiutato dalla vittima. Il suo profilo è caratterizzato da un'ambivalenza di scopo, ovvero egli vuole, contemporaneamente, vendicarsi del rifiuto e riprovare ad "allestire" una relazione con la vittima.
- Il *predatore*: quest'ultima categoria identifica un individuo il cui obiettivo è di natura essenzialmente sessuale. Egli trae esaltazione nel rendere noto alla propria vittima di essere oggetto del proprio interesse molesto, soprattutto da un punto di vista sessuale. E raggiunge il proprio intento nel possesso della vittima, solo dopo averle creato pavido e profondo turbamento.

Gli atti persecutori hanno la conseguenza di modificare inevitabilmente le abitudini di vita della vittima. Quest'ultima sentendosi braccata dal proprio *stalker* reagisce limitando la propria esistenza, isolandosi nelle proprie mura domestiche e facendo qualsiasi cosa ritenga necessaria pur di fuggire dal proprio persecutore. La violenza persecutoria dello *stalking* ha quindi la conseguenza di modificare la vita della vittima in tutte le sue declinazioni, soprattutto da un punto delle interazioni sociali e con ripercussioni sulla salute fisica e mentale, a volte permanenti. Nel 2012, il 18% delle donne che vivono in un paese membro dell'UE si dichiara vittima di comportamenti persecutori (FRA, 2014), di cui il 14% ha ricevuto ripetutamente messaggi o telefonate offensive o minacciose da parte della stessa persona, mentre l'8% ha subito pedimenti sul posto di lavoro, al proprio domicilio o in luoghi frequentati per intrattenere le proprie relazioni sociali. Più nello specifico, una donna su dieci, ossia il 9% del campione intervistato è stata vittima di *stalking* da parte di un *ex-partner* (FRA, 2014).

1.2.5. La cyber-violenza

I contesti della violenza stanno diventando sempre più raffinati e particolareggiati negli ultimi decenni, grazie soprattutto all'incremento di nuove tecnologie e alla presenza di una società sempre più digitalizzata. Nello specifico, si fa riferimento a forme di violenza mosse *dall'attivismo da tastiera*, in cui le intimidazioni, le minacce e gli insulti sono trasmessi alla vittima attraverso *il contatto virtuale*, delineando così la cosiddetta *cyber-violenza* (Donne delle Nazioni Unite, 2016). Gli attacchi virtuali sono caratterizzati dalla presenza di un pubblico virtualmente illimitato e da una rapidità di diffusione che può avere conseguenze molto concrete.

Questo nuovo genere di violenza si realizza per mezzo di insulti, minacce, intimidazioni, la diffusione di un linguaggio d'odio verso specifiche categorie, la denigrazione e l'oppressione psicologica a carattere vessatorio. Si tratta di un fenomeno recente e ancora sottovalutato, nel quale le donne vengono colpite in modo sproporzionato, come riportato da uno studio condotto dall'*European Institute of Gender Equality* (EIGE, 2017). In particolare, la *cyber-violenza* si traduce in *hate speech*, *cyber-bullying*, *cyber-stalking* e le donne diventano il bersaglio di minacce di stupro, foto pubblicate senza autorizzazione che le ritraggono nell'intimità, ricatti a sfondo sessuale; si trasformano in vittime di tracciamento *online* nascosto o della divulgazione di informazioni personali e materiale pornografico.

La ricercatrice Bérengère Stassin specializzata in comunicazione e *cyber-molestie* afferma che vi sia il sessismo maschilista al cuore della *cyber-violenza*, dove la natura prettamente sessuale delle intimidazioni e degli insulti non fa che rafforzare le differenze di genere tra uomini e donne. Gli studi infatti dimostrano che ogni 30 secondi una donna è vittima di violenza su un qualche Social Media (Amnesty International, 2018), i cui discorsi sono prevalentemente a carattere discriminatorio (basati su omofobia, transfobia e razzismo). Pertanto, la violenza *online* contro le donne non è altro che un prolungamento della violenza *offline*, amplificata dalla presenza di un pubblico illimitato. La caratteristica più singolare della violenza virtuale è proprio il numero illimitato di individui che possono accedere alle informazioni denigratorie pubblicate a carico di una specifica persona. Infatti, se nella violenza *offline* la molestia, la minaccia o l'umiliazione si sarebbe limitata ad una ristretta cerchia di persone, il mondo *online* rende più problematico il diffondersi delle notizie, in quanto un messaggio provocatorio, una foto o un

video problematico possono trasformarsi in breve tempo in un fenomeno di massa incontrollabile, con conseguenze permanenti per l'identità della vittima. Il fenomeno è sempre più comune nei paesi occidentali ed in particolare nella fascia d'età degli adolescenti (Smith *et al.*, 2008). In un rapporto pubblicato, nel settembre 2015 dalla Commissione della Banda Larga delle Nazioni Unite si afferma che il 75% delle donne abbia riscontrato molestie e minacce su base discriminatoria di genere, durante il proprio transito sul *Web*. Il dibattito pubblico infatti è sempre più fervido in materia, portando gli organi istituzionali a chiedere la collaborazione di Social Media come *Facebook e Twitter* per il monitoraggio delle conversazioni *online*. Le normative e le *policies* si sono dovute adeguare al fenomeno delle *cyber-molestie* definendo in maniera sempre più mirata e dettagliata la disciplina delle comunicazioni sul *Web*. Un'altra caratteristica che i Social Media si prefiggono di controllare è il problema dell'anonimato, che come spiega la studiosa Stassin, incoraggia la violenza digitale proprio perchè favorisce l'attivismo da testiera nell'opportunità di agire di nascosto e rimanere impuniti. I comportamenti persecutori perpetrati attraverso i mezzi informatici colpiscono soprattutto le giovani donne e nell'UE, in una fascia di età compresa tra 18 e 29 anni, il 4% di esse è stata vittima di *stalking on-line*, ovvero oltre un milione e mezzo (FRA, 2014).

Il primo rapporto sulle *cyber-violenze* pubblicato in Francia ha mostrato che spesso la violenza digitale viene esercitata sotto il tetto coniugale. Difatti, 9 donne su 10 vittime di violenza domestica sono state al contempo anche oggetto di *cyber-violenza* da parte del proprio *partner*. Il 34% delle donne intervistate ai fini di tale rapporto ha sperimentato violenza digitale a sfondo sessuale, attraverso la pubblicazione di materiale pornografico, minacce e ricatti. La fascia d'età più colpita riguarda donne al di sotto dei 25 anni ed il 50% di esse è stata costretta ad essere filmata durante i rapporti sessuali. Questa tipologia di violenza rappresenta attualmente una delle forme di violenza domestica più frequenti di questa società post-moderna, identificando il prodotto di una pratica di violenza metodica che ha lo scopo di denigrare ed umiliare la vittima, innescando meccanismi di paura e vergogna. La violenza informatica o digitale ha l'effetto di ridurre l'attività *online* delle donne, rinforzando il loro isolamento da qualsiasi forma di socialità reale o virtuale.

La *cyber-violenza* domestica può manifestarsi anche attraverso il controllo della vita sociale della vittima, violando la sua *privacy* e sicurezza informatica. In particolare, il 21% delle donne intervistate in Francia per il rapporto sulle *cyber-molestie* riferisce un sistematico monitoraggio da parte del proprio

partner, attraverso l'installazione di *spyware* sui propri dispositivi, con lo scopo di accedere ai contatti, monitorare le attività di telefonia e messagistica, gli spostamenti con strumenti di geolocalizzazione e quindi le relazioni della vittima.

1.2.6. Il Vitriolage: l'aggressione con l'acido

Nell'ambito degli studi riguardanti la violenza di genere sempre più conosciuta è l'aggressione con l'acido o *vitriolage*. Fino a qualche decennio fa era una pratica assolutamente rilegata a pochi paesi sotto-sviluppati o in via di sviluppo ma attualmente è diventata rapidamente popolare anche nei paesi occidentali. Il *vitriolage* consiste nell'aggreire la vittima gettandole sul corpo e sul volto una sostanza corrosiva, con l'intento di sfigurarla, mutilarla o ucciderla.

Le sostanze corrosive più comunemente utilizzate dagli autori sono l'acido solforico, nitrico o cloridrico (Welsh, 2006) che sono solitamente usate per colpire il volto della vittima, poichè l'acido ha l'istantaneo potere di deformare i lineamenti e i tratti somatici, bruciando e danneggiando in via irreversibile il tessuto cutaneo e nei casi più estremi, dissolvendo fino alle ossa la zona colpita. L'aggressione con l'acido è senz'altro un'azione di violenza premeditata e preterintenzionale nel suo aspetto più subdolo. Si tratta infatti di una forma di violenza finalizzata a deturpare il corpo della vittima che sopravvive imparando a convivere con la deformazione della propria immagine. L'intenzione dell'autore è di marchiare, stigmatizzare, "ferire e sfigurare l'aspetto fisico della donna per gelosia o vendetta" (Vig, 2014). Le conseguenze dell'aggressione con l'acido possono comprendere cicatrici permanenti sul viso e su tutto il corpo, problemi persistenti alla vista, inclusa la cecità (Bandyopadhyay e Mahmuda, 2003). Pertanto, si tratta di una pratica che genera conseguenze devastanti a livello fisico ma soprattutto sul piano psicologico e relazionale della vittima. Il risultato effettivo di tale forma di violenza è il danneggiamento stabile dell'esistenza della vittima, la cui vita subisce un profondo stravolgimento. Oltre alle sofferenze fisiche di natura medico-chirurgica, la persona aggredita con l'acido deve reinventare la propria identità al di fuori di un'immagine che non riconosce più, perché ha perso i suoi connotati. I danni sul piano psicologico, sociale, professionale ed economico sono pertanto incommensurabili e la vittima deve abituarsi alle condizioni associate a questa

nuova esistenza (Swanson, 2002). La scelta dell'acido come arma da parte dell'aggressore è a scopo principalmente punitivo, nella volontà di annientare l'identità, distruggerne l'aspetto fisico e ledere la psiche della vittima, compromettendone per sempre la qualità dell'esistenza. L'aggressore è quasi sempre una persona che ha avuto un qualche legame con la vittima, di natura sentimentale, amicale, professionale. In particolare, se l'autore ha avuto una relazione di tipo sentimentale con la vittima, questo agisce con l'intento di non permettere all'*ex-partner* di potersi ricostruire una vita all'infuori della sua presenza. Mentre, se l'azione è determinata da un *ex-amico* o collega/subordinato è quasi sempre a scopo punitivo o vendicativo per un torto subito. Nella maggior parte dei casi, la vittima prima di essere aggredita con l'acido è bersaglio di comportamenti ossessivi e distruttivi. Il *vitriolage* rappresenta infatti l'atto finale di una serie di pregressi comportamenti molesti e spesso assillanti nei confronti della vittima. L'aggressore agisce con profonda consapevolezza e predilige questo tipo di violenza sapendo di provocare sfregi e danni permanenti nella vita della persona aggredita. In aggiunta, oltre a ledere l'aspetto fisico della vittima, essi fanno di condannarla ad una vita sociale non più autonoma, le cui relazioni sociali saranno compromesse per sempre. Si tratta della "vendetta perfetta" perché le conseguenze sono irreversibili e la donna deve indossare "i panni della ricostruzione" (Randone, 2018). Attaccare una donna con l'acido rappresenta la deriva della misoginia nel tentativo di cancellare e deformare per sempre il volto della persona che spesso si ritiene di aver amato, in una dinamica di interazione disfunzionale e che si traduce nel possesso e nella necessità di placare la fragilità dell'abbandono o del tradimento (*Ibidem*).

In realtà, il termine retorico *vitriolage* è stato coniato dalla stampa popolare francese nel 1879, per indicare "un'ondata di vetriolo" o attacchi con l'acido definiti come crimini passionali perpetrati prevalentemente da donne contro altre donne. È considerato storicamente uno dei casi più emblematici, l'accidentale e fatale caso del principe Leopoldo Clemente di Sassonia-Coburgo e Gotha, quando nell'ottobre 1915, la sua amante Camilla Rybicka lo ha accidentalmente oppure no, sfigurato con dell'acido¹. La notizia è diventata subito di interesse della stampa, suscitando una certa attenzione ed emulazione da parte del pubblico. Negli anni avvenire infatti la pratica sembra essere

¹ <https://paperspast.natlib.govt.nz/newspapers/TC19160107.2.10>

divenuta di frequente utilizzo in molti paesi in via di sviluppo, come l'Asia meridionale. In effetti, i primi attacchi con l'acido sono stati registrati in Bangladesh, in India e in Cambogia. La casistica mondiale dell'aggressione con l'acido è del 75-80% (De Castella, 2013) attualmente, con una prevalenza proprio nei contesti geografici dell'Asia meridionale, quali Pakistan, Bangladesh, Cambogia, Vietnam, India, Hong Kong, Cina, Sudafrica, Uganda, Etiopia, Iran ed Afghanistan. Tuttavia, la ricerca ha visto un sostanziale aumento della quantità di episodi riconducibili al *vitriolage* anche in altre parti del mondo, compresa l'Europa, malgrado una significativa sottostima degli eventi avvenuti a cavallo degli anni Ottanta e Novanta. In Italia ad esempio, si sono registrati diversi casi di aggressione con l'acido, ad esempio nel 2013 l'avvocata Lucia Annibaldi è stata sfigurata su commissione dell'ex-fidanzato e nel 2017 la ballerina Gessica Notaro ha subito *vitriolage* da parte dell'ex-compagno.

1.2.7. Le mutilazioni genitali

Un altro tipo di violenza sulle donne è quello della mutilazione genitale femminile (*FGM*), portato all'attenzione dell'opinione pubblica mondiale dalle lotte femministe degli anni Settanta. Si tratta di una barbaria nota anche come taglio genitale o circoncisione femminile che identifica una pratica adottata nel tentativo di controllare la sessualità delle donne e rivendicare l'idea di purezza, modestia e bellezza. La mutilazione genitale viene principalmente espletata in Africa, Asia e Medio Oriente. In particolare, nel 2016, l'*UNICEF* ha stimato che circa 200 milioni di donne che vivono in 30 paesi (27 paesi africani, Indonesia, Kurdistan iracheno e Yemen) hanno subito tale pratica.

Nella comunità di riferimento, rifiutare di sottoporsi a tale mutilazione conduce all'esclusione sociale (Cappa *et al.*, 2013).

A livello pratico, l'*FMG* consiste nel taglio o rimozione rituale di alcuni o tutti i genitali femminili esterni, rappresentando un problema di tipo medico-sanitario, poichè identifica una condotta oggettivamente lesiva dell'integrità psico-fisica della donna, e al contempo di riflessione criminologica, perché si sostanzia in un comportamento delittuoso. L'azione di mutilazione genitale è dalla letteratura (Mottin-Sylla, 1991) classificata in due macro-tipologie, ognuna delle

quali corrisponde a tre diversi gradi: i primi due rappresentano l'escissione vera e propria mentre il terzo tipo si identifica con l'infibulazione.

In realtà, le motivazioni variano da un contesto geografico all'altro, includendo una serie di variabili socio-culturali.

Talvolta, il fenomeno dipende da una convenzione sociale indiscussa e reiterata nel tempo e dalla necessità di conformarsi alla propria comunità per non esserne respinti. In molti casi, si tratta di convinzioni sull'accettabilità della sessualità femminile, allo scopo di garantirne la verginità prematrimoniale o la fedeltà coniugale. In altre circostanze invece, la pratica viene eseguita per ideali culturali di purificazione, spesso associate a tradizioni religiose.

Nell'ambito della Terza Giornata Mondiale per l'eliminazione delle *MGF* tenutasi a Roma nel febbraio 2009, è stato invece ribadito che tali pratiche rientrano nell'ambito della violenza sessuale sulle donne, qualificandosi in termini di un rituale disumano che non trova fondamento in nessuna dottrina religiosa o culturale. La natura delittuosa delle *MGF* richiama inevitabilmente la dimensione socio-culturale di una comunità, oltre che l'aspetto sadico prettamente maschilista. Si tratta di uno strumento di controllo e subordinazione della donna, il cui diritto sessuale dipende dalla volontà maschile, legittimando soprattutto il potere economico e sociale attribuito alla relazione matrimoniale, nel cui rapporto si perpetua l'inferiorità della donna rispetto all'uomo.

Sussiste quindi una relazione tra la società, la cultura e la sessualità delle donne che motiva il tema della mutilazione. A tal proposito, si può fare riferimento alla sociologia del corpo che indaga gli influssi sociali esercitati sulla corporeità della donna. Vi sono alcuni studiosi che sostengono l'esistenza di una e vera e propria controversa volontà di stigmatizzazione della donna. La supremazia maschile afferma il proprio potere nel controllo della femminilità attraverso la *MGF*, marchiando la donna proprio nella sua componente femminile più intima. Il termine *mutilare* vuole per l'appunto indicare la modificazione genitale che la cultura maschilista fortemente richiede e tenta di perpetrare come valore imprescindibile di appartenenza alla comunità di riferimento. La donna è quindi vincolata a portare con sé i valori di una struttura sociale di appartenenza che quasi sempre non ha scelto ma subisce per volontà altrui. Le *MGF* rappresentano un legame tra la violenza e la sessualità, poiché l'abuso viene agito *nel* e *sul* corpo. La violazione è quindi esercitata al contempo attraverso il corpo come scenario dell'impulso tra la vita e la morte e a livello di inconscio, compromettendo l'identità personale e quindi il *corpo che si rappresenta* e il *corpo che si è*. Si tratta di una dialettica tra corpo-oggetto che

attraverso la mutilazione del corpo caratterizza l'esistenza umana della donna, in una compiutezza di intergrità interpersonale e affettivo-sessuale.

Tuttavia, di enorme interesse sono anche gli aspetti connessi alla salute psico-fisica della donna (Berg *et al.*, 2014, Reisel & Creighton, 2015). Infatti, sul piano fisico, le *MGF* possono includere infezioni ricorrenti, la trasmissione di alcune malattie come epatite e HIV, problematiche di natura ginecologica e dolore cronico, fino al sanguinolento fatale e alla morte (Abdulcadir *et al.*, 2011). Sul versante psicologico invece l'aspetto più problematico è che la maggior parte delle vittime sono minorenni, il cui trauma e i problemi connessi alle relazioni e ai rapporti sessuali possono essere permanenti.

Non a caso, l'*UNICEF*, l'*ONU* ed altre organizzazioni internazionali governative e non governative si sono impegnate nel corso del tempo per opporsi a tale pratica. L'oggetto del dibattito non è certo privo di critiche, soprattutto da parte degli antropologi che hanno sollevato domande sul relativismo culturale e sull'universalità dei diritti umani (Shell-Duncan, 2008).

Il dibattito è sempre molto fervido a riguardo, in particolar modo perché sebbene questa drastica pratica chirurgia genitale femminile sia di interesse di una specifica popolazione in modo capillare, a seguito del fenomeno migratorio, si è ormai diffusa anche nei paesi occidentali. Pertanto, l'esistenza di un complesso normativo occidentale che predilige gli interessi e la tutela dei membri appartenenti alla società, la pratica delle *MGF* si scontra con i diritti fondamentali riconosciuti dai paesi democratici moderni. In particolare, per quel che concerne il diritto all'integrità fisica, alla salute, alla non discriminazione per questioni di genere, religione, cultura, etnia.

Infatti, nell'*art. 38 della Convenzione del Consiglio d'Europa sulla prevenzione e lotta contro la violenza nei confronti delle donne e la violenza domestica* (2011), il testo afferma:

“Le Parti adottano le misure legislative o di altro tipo necessarie per perseguire penalmente i seguenti atti intenzionali: l'escissione, l'infibulazione o qualsiasi altra mutilazione della totalità o di una parte delle grandi labbra vaginali, delle piccole labbra o asportazione del clitoride; costringere una donna a subire qualsiasi atto indicato al punto a, o fornirle i mezzi a tale fine; indurre, costringere o fornire a una ragazza i mezzi per subire qualsiasi atto enunciato al punto a”.

Il Consiglio europeo considera infatti le *MGF* come una forma di violenza e di brutale manifestazione della disuguaglianza di genere. Nello specifico,

costituisce un abuso contro le donne e contro l'infanzia. Le legislazioni dei paesi membri devono quindi adeguarsi a tali definizioni ed estendere il proprio quadro normativo senza lasciare spazio a reinterpretazioni in materia di *MGF*.

1.3. La criminalizzazione della violenza attraverso il “Femminicidio”

Nell'ambito della violenza contro le donne si sono affermati inoltre due interessanti neologismi volti ad indicare il fenomeno in un'ottica di criminalizzazione, quali *femminicidio* o *femicidio* (Radford & Russell, 1992; Legarde, 1997). Spesso usati come sinonimi, i due termini si collocano da un punto di vista semantico in diversi scenari, ed in particolare in campo letterario e giuridico. Le origini di questi due neologismi si fanno risalire ad un patrimonio culturale femminista degli anni Settanta, la cui introduzione mira a delimitare e comprendere una tipologia criminologica specifica: la violenza sulle donne da parte dell'uomo. Si tratta essenzialmente di differenziare qualsiasi atto criminale compiuto da un uomo verso un altro uomo, sottolineando la vittimizzazione prettamente femminile. La nozione di femminicidio nasce quindi in campo criminologico ma viene utilizzata in altri domini di ricerca per indicare l'omicidio doloso e preterintenzionale di vittime di sesso femminile uccise per motivi di genere (Radford e Russell, 1992), in una società di matrice patriarcale votata alla sudditanza femminile. Tuttavia, al fine di trattare in maniera opportuna il tema oggetto di studio, è bene sottolineare che nel tempo, i neologismi hanno subito un'estensione nella loro definizione. Specificamente, Legarde (1997) ne ha fatto ricorso per identificare la strage di donne avvenuta in Messico tra il 1993 e 2003, periodo nel quale si sono registrati numerosi casi di sequestri e omicidi di donne.

Esiste quindi un'importante ambivalenza semantica (Spinelli, 2008) afferente ai termini femminicidio o femicidio, ed in particolare se si analizzano le definizioni proposte da Radford & Russell (1992) e Legarde (1997). Pertanto, volendone ampliare l'orizzonte semantico, tali termini possono riferirsi al contempo, sia all'uccisione diretta della donna che alla sua mortificazione fisica e morale e di distruzione del proprio ruolo sociale.

In epoca contemporanea infatti, la definizione di femminicidio non è strettamente rilegata alla morte della donna ma si riferisce nella sua accezione più autentica ad ogni forma di violenza esercitata sistematicamente sulle donne,

in nome di una sovrastruttura ideologica di matrice patriarcale, comprendendone l'uccisione da parte di un uomo per motivi di odio, disprezzo, piacere o senso di possesso (*Ibidem*).

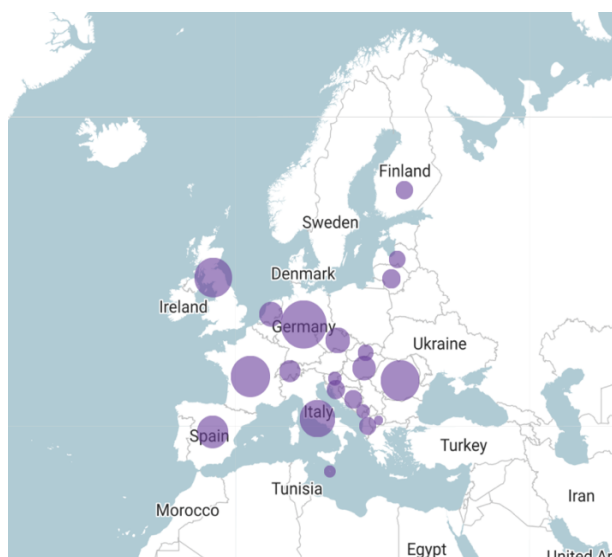
Tale definizione è particolarmente rievocata da parte della comunità scientifica, nell'impellente esigenza di racchiudere in un'unica parola la violenza sulle donne, cercando di circoscrivere una tipologia di crimine che vittimizza la donna per mano maschile, la cui personalità presenta un'inclinazione alla misoginia, all'odio, alla smania di possesso che spesso sfugge al controllo, scatenando una furia omicida che operando in un contesto di autentica follia si trasforma in un delitto.

Tuttavia, nel tentativo di definire meglio quest'osservatorio particolarmente fecondo delle trasformazioni socio-culturali che hanno esasperato le dinamiche interazionali ed identitarie, tanto da sfociare in comportamenti aggressivi di criminalizzazione, sono molti i neologismi che compaiono sulla scena, alcuni dei quali ad esempio *ginocidio* (Danna, 2007). In realtà, Daniela Danna riporta alla ribalta tale vocabolo coniato dal femminismo, per indicare non solo gli omicidi contro le donne ma ogni atto di violenza contro il femminile nella sua accezione più ampia, condannando un disprezzo sociale, una mania di controllo e di confinato spazio di potere e dominio patriarcale, dai quali, i corpi socializzati femminili tentano di fuggire ad ogni costo.

L'escalation femicida che si consuma in questo presente storico non rappresenta statistiche o storie di cronaca, bensì indica una silente distruzione del genere femminile che viene perpetrata in tempi e spazi che fanno da protagonisti ad una conflittualità della quale siamo tutti un po' vittime, un po' carnefici ed un po' spettatori silenziosi. Si tratta effettivamente di un funzionalismo interazionale tra uomo e donna che si colloca in una prospettiva di *continuum* tra passato e presente. Secondo la letteratura, il contesto sociale appare quindi caratterizzato da un vincolo quasi incatenante con la cultura ancestrale. Di conseguenza, l'idea di mettere in discussione l'identità maschile ed il potere che tale sessualità gli conferisce genera una vulnerabilità individuale e collettiva che si esaspera fino a perdere il controllo. In una visione in cui le strutture del passato iniziano a vacillare, l'uomo si sente diseredato del proprio patrimonio genetico e se decade tale egemonia, il funzionalismo si inceppa in un ingranaggio che raggiunge l'apice con l'azione fatale dell'omicidio. I mezzi di comunicazione infatti propinano un meccanismo di violenza maschile piuttosto preoccupante, facendo al contempo echeggiare una

sorta di manipolazione dell'universo simbolico a cui invece la società è abituata e fa riferimento.

Nella maggior parte dei casi, nei media si parla di femminicidio per indicare omicidi di donne e soprattutto per identificare delitti di natura passionale o di rivendicazione maschilista ed in casi più rari, delitti connessi ad una qualche patologia psichiatrica dell'autore. Le cronache sono infatti sature di liste di nomi di donne uccise da autori non identificati, reo confessi o rilevati per mezzo di indagini scientifiche. Nel corso del 2012, in Europa, il numero di donne uccise da parte di un *partner* (uxoricidio), un ex-fidanzato o un membro della famiglia è pari a 43 mila e 600 vittime (FRA, 2014). Nell'ottica in cui si considera il femminicidio nel suo aspetto omicidiario, il maggior numero di casi si registra in Italia, Germania e Regno Unito (Eurostat, 2017; *Figura 1.2*). L'incidenza del fenomeno si può stimare mediante il rapporto tra il numero di vittime e la popolazione femminile di riferimento.



*Figura 1.2. Mappa della prevalenza di Femminicidi nei Paesi membri dell'UE.
Fonte: Eurostat, 2017.*

Si rintraccia altresì una certa prevalenza in Montenegro, Lettonia, Lituania, Repubblica Ceca Ungheria, Bosnia Erzegovina ed infine in Croazia, il cui tasso di mortalità per mano maschile è superiore alla media europea (*Ibidem*).

Esiste tuttavia l'assenza di dati ufficiali e di un repertorio documentale aggiornato a livello mondiale che possa consentire un confronto ed un aggiornamento su scala globale.

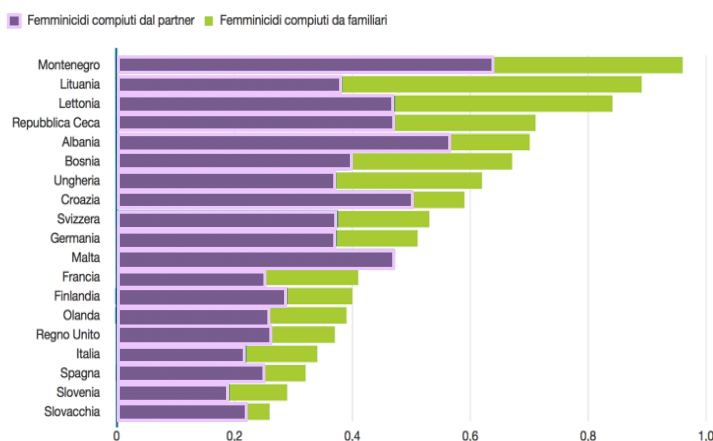


Figura 1.3. Quota percentuale della prevalenza di Femminicidi nei Paesi membri dell'UE. Fonte: Eurostat, 2017.

Nell'ottica in cui si considera il femminicidio nella sua accezione di violenza estrema di tipo omicidiaria, la conseguenza per la vittima è la morte mentre sussistono ripercussioni sul piano comunitario di tipo giuridico, economico e di discussione pubblica. Al contrario, se il femminicidio viene inteso nella sua estensione semantica di violenza contro le donne in tutte le sue forme, vi possono essere conseguenze sia sul piano individuale, per lo più di tipo sanitario, che su quello collettivo.

1.4. Le conseguenze dei labirinti del male

La violenza sulle donne produce inevitabilmente delle conseguenze, sia che si manifesti in forma continua o discontinua e sia che essa venga esercitata in termini di violenza fisica, sessuale, psicologica e/o economica.

Sebbene siano le donne ad essere in maniera diretta le protagoniste degli episodi di violenza, l'impatto può essere notevole anche sulla famiglia o gli eventuali figli della vittima, ed in senso più ampio le ripercussioni possono essere vissute dall'intera collettività.

Sul piano individuale, le conseguenze sono ascrivibili alla salute mentale e fisica della donna, sia nel breve che nel lungo termine, e possono rivelarsi devastanti e talvolta fatali. La violenza si imprime nella memoria relazionale della vittima, compromettendone in maniera indelebile le future esperienze. Una donna che subisce violenza è una persona diversa, il cui universo simbolico tende a mutare, così come le strutture cognitive sul mondo esterno. Le vittime di violenza cambiano opinioni, percezioni e attribuiscono significati diversi alle cose, alle persone, ai ruoli e ai comportamenti.

Nello specifico, l'*Organizzazione Mondiale della Sanità* (OMS) definisce l'esposizione continua o discontinua a microtraumi, reiterata nel tempo, in termini di *sindrome della donna maltrattata* o *disturbo post-traumatico da stress*. La definizione fa riferimento alla proposta avanzata da Herman (1992) per indicare l'entità delle conseguenze scaturite dalla violenza ripetuta e prolungata nel tempo. Tale disturbo clinico racchiude infatti una vasta gamma di sintomi e disfunzioni che compromettono l'intera personalità della vittima.

Tuttavia, molti studiosi sostengono che il *disturbo post-traumatico* non sia abbastanza adatto a fornire un reale quadro sintomatico della violenza subita in uno o più episodi e nelle più svariate e controverse forme. Di conseguenza, al fine di ovviare a possibili lacune semantiche, la comunità scientifica predilige inquadrare le conseguenze dei labirinti della violenza attraverso un assetto patologico denominato *disturbo post-traumatico complesso* (PTSD; Herman, 2005). Il quadro patologico racchiude una micro-ripartizione delle possibili conseguenze sulla salute mentale e fisica delle donne, che nella *Tabella 1.1* vengono riassunte nella versione convenzionalmente indicata dalla letteratura.

Le ripercussioni a livello psicologico sono particolarmente elevate e a dimostrarlo è una ricerca proposta da Romito (2005) che indaga le conseguenze che si sono manifestate in donne che hanno subito violenza nell'ultimo anno. Le vittime studiate dall'autore hanno infatti dimostrato di non avere un buon rapporto con la propria salute e di avvertire malessere, ammettendo di ricorrere all'uso di sostanze che hanno il potere di modificare lo stato psico-fisico (attenzione, percezione, umore, coscienza, comportamento, *etc*), nonostante queste spesso inducano alla dipendenza o all'assuefazione. In molti casi, compaiono depressione, disturbi alimentari o del sonno, fenomeni compulsivo-ossessivi, dissociazione, alterazione della sfera identitaria, somatizzazione, attacchi di panico, autolesionismo e molto frequente è la tendenza al suicidio (OMS, 2002). È bene considerare che esistano reazioni piuttosto singolari che non vengono menzionate dalla letteratura, poiché la violenza può innescare

reazioni diverse e produrre sintomi differenti a seconda della vittima, o se le manifestazioni sono comuni possono mostrarsi con diversa intensità. La condizione postuma varia a seconda del soggetto e delle sue capacità di reazione. Tuttavia, le conseguenze mutano anche in considerazione del contesto e dell'identità dell'autore e talvolta in base all'eventuale procedimento penale e alla condanna del proprio aggressore. Sul piano della salute fisica rientrano conseguenze quali ematomi, lesioni, escoriazioni, danni oculari, sindromi croniche, danni neurologici e in alcuni casi problemi connessi alla sfera ginecologico-riproduttiva, tra cui lesioni gravi o permanenti, malattie sessualmente trasmissibili, gravidanze indesiderate o sterilità. Inoltre, la violenza subita in stato gestazionale può portare ad episodi di aborto o parti prematuri (WHO, 2014), con conseguenze anche successive al parto sia da parte della madre con manifestazioni di depressione *post-partum* che nello sviluppo e temperamento del neonato (McMahon *et al.*, 2013).

Conseguenze fisiche	Conseguenze d'interesse ginecologico/riproduttivo	Conseguenze psicologiche e comportamentali	Conseguenze mortali
Lesioni addominali o toraciche	Malattie infiammatorie pelviche	Somatizzazione: ansia, ipervigilanza,	Mortalità legate a malattie infettive fatali
Contusioni, ematomi, fratture	Disturbi dell'apparato riproduttivo	Alterazione della regolazione affetti ed impulsi: depressione, disturbi del sonno e dell'alimentazione	Suicidio
Ferite da arma da taglio	Isterectomie	Disturbi compulsivo – ossessivi	Omicidio
Ustioni, lacerazioni e abrasioni	Complicanze della gravidanza	Fobie ed attacchi di panico	
Trauma cranico	Aborto spontaneo	Autolesionismo e tendenza suicidaria	
Sindrome da dolore cronico	Gravidanze indesiderate	Episodi di dissociazione: depersonalizzazione	
Fibromialgia	Sterilità	Abuso di fumo, alcool, farmaci e droghe	
Disturbi e sindromi croniche gastrointestinali	Malattie trasmissibili, compreso virus HIV/AIDS	Alterazioni patologiche della sfera identitaria e relazionale: compromissioni professionali e perdita delle relazioni sociali	
Danni oculari	Neoplasie	Alterazione della percezione e cognizione	
Funzione fisica ridotta		Senso di colpa e vergogna	

Tabella 1.1. Le conseguenze della violenza sulle donne.

Fonte: OMS, 2002

Sul piano comunitario invece, soprattutto nei casi di violenza parentale si possono osservare le prime conseguenze connesse alla violenza indiretta, soprattutto sui minori. L'esposizione ad episodi traumatici verso una figura affettiva di riferimento, come ad esempio la madre, può provocare problemi

emotivi e cognitivi (Evans *et al.*, 2008), innescando inoltre meccanismi relazionali insani che spesso riproducono nel futuro prossimo le stesse dinamiche vissute durante l'infanzia (Guglielmin *et al.*, 2013).

Pertanto, se il supporto sociale avviene durante la prima infanzia o l'adolescenza, gli attori coinvolti nella gestione dei traumi che derivano dalla violenza assistita sono diverse, come ad esempio gli operatori di supporto psicologico o gli assistenti sociali. La vittima invece che subisce in via diretta la violenza rischia di essere sottoposta ad un'ulteriore forma vessatoria da parte degli operatori sanitari, i quali sono spesso distratti, incapaci di gestire la situazione sul piano emotivo e concreto, a causa di una mancanza di conoscenze pertinenti al fenomeno o della mancanza di linee guide specifiche (Romito, 1999; Chang, 2004). In tal senso, l'intervento sul piano medico-sanitario si limita a medicalizzare la donna non riconoscendo lo scenario pericoloso che si cela dietro falsi incidenti o abusi di qualsiasi tipo che non vengono palesemente denunciati dalla vittima. Infatti, la letteratura (Romito, 1999) identifica nel settore medico-sanitario, limiti connessi alla negazione della violenza e alla giustificazione dell'aggressore, affermando complicità pubblica nell'ambito dell'abuso domestico.

Sul versante economico, la violenza influisce poi sul bilancio nazionale. In particolare, costituiscono una "spesa" per lo Stato, sia durante che dopo la cessazione della violenza, i costi sanitari in termini di visite mediche, farmaci, supporto psicologico, degenze e primi soccorsi. Le attività assistenziali di tipo legale, delle forze dell'ordine, dei servizi sociali e dei centri antiviolenza sono inoltre a carico dello Stato in un'ottica di gratuità dei servizi.

A livello collettivo il discorso pubblico assume altresì una certa rilevanza, poiché nel tentativo di educare sulla violenza contro le donne, la comunicazione tende a rafforzare gli stereotipi di genere, evidenziando le differenze tra i sessi. In particolare, nel discorso pubblico emerge il problema della reificazione della donna nell'immaginario collettivo in una sorta di sessualità onnipotente. I meccanismi della comunicazione esprimono modelli che incrementano l'ostilità nei confronti delle donne attraverso la somministrazione di una percezione disumanizzata delle stesse. Si tende a sottolineare una società sessista e maschilista che rappresenta le donne per mezzo di caratteristiche di strumentalità, mancanza di autonomia, fungibilità (Bourdieu, 1998; Nussbaum, 1999). Le donne vengono rilette ad una corporeità pericolosa che ingloba strutture che annullano intelletto, autostima, indipendenza cognitiva, culturale ed economica. Tali meccanismi non fanno che accrescere anche nelle donne

una tale percezione di inferiorità rispetto agli uomini, avvalorando così la *normalità* della violenza.

1.5. Una breve *anamnesi* delle possibili costanti che caratterizzano il fenomeno

Il tema della violenza sulle donne rappresenta un osservatorio delle trasformazioni socio-culturali che nel corso del tempo hanno caratterizzato la società, prestandosi con una certa facilità alla costruzione di strutture basate sulla cultura di genere, sull'identità sessuata degli individui, sul dominio maschile e sulla virilità che hanno determinato un mondo falloocratico. Il fenomeno preso in esame identifica uno strumento archeologico (Bourdieu, 1998) della storia dell'inconscio, che sebbene abbia origini antiche "abita" in ciascuno di noi. Si tenta dunque di inquadrare e trattare le proprietà biologiche e psicologiche inscritte nell'inconscio come elementi costanti, al fine di far emergere gli aspetti naturalizzati in un ambiente di riferimento così sessualmente differenziato e differenziante. Bourdieu (1998) prende spunto da Freud per esprimere la familiarità insita in tali strutture, come una *riscoperta* di una conoscenza al contempo oggettiva e soggettiva, che si riappropria di consapevolezze "a tempo possedute e da sempre perdute" (*Ibidem*). In particolare, Freud, richiamando Platone, le definisce *anamnesi*.

1.5.1. Il "genere" come categoria di analisi

Per rimarcare la costruzione di strutture che determinano l'identità individuale e collettiva basata sul *genere*, risulta necessario dapprima tracciare i confini etimologici del termine, con lo scopo di evitare ogni sorta di ambiguità semantica. Talvolta, alcune terminologie sono talmente incorporate negli schemi sociali che non ci si pone l'esigenza di decostruirli per tentare di individuare un significato univoco all'interno di una moltitudine di possibilità di senso. Infatti, parlare di *genere* significa discutere di una categoria semantica che nel corso dei secoli è stata sottoposta ad attribuzioni di senso piuttosto ambivalenti. L'uso del termine viene impiegato in modo spesso allusivo e traslato, coinvolgendo e contrapponendo allo stesso tempo diversi approcci

teorici.

In ambito sociologico, si definisce di genere una categoria a cui sono imputabili concetti di mascolinità e femminilità, in un mutamento di contesti storico-sociali. L'etimologia è riconducibile al latino *genus*, poi diffusosi in Inghilterra sotto forma di *gender*, con lo scopo di indicare ruoli di genere intesi come modelli di interazione. Più propriamente, il genere rivela l'*identità sessuata* di un individuo e sulla base di questo ne determina i ruoli sociali. È attraverso questa prospettiva che il genere può assumere la definizione di un insieme di norme comportamentali, tramite le quali la società "trasforma la sessualità biologica in prodotti di attività umana" (Reiter, in Rubin, 1975).

In prima istanza, si rischia inevitabilmente di sintetizzare la questione in termini di distinzione genetica, dove sesso e genere sono termini intercambiabili e sinonimi. Sono infatti molti gli studiosi che hanno ben distinto la natura biologica nella sua dimensione genotipica, ovvero maschio genotipicamente *XY* e femmina genotipicamente *XX*, e nel suo aspetto fisiologico, attraverso i cosiddetti *modelli comportamentali* (Money *et al.*, 1955).

uttavia, il vero punto di svolta nell'analisi del genere arriva con gli "*gender studies*" finalizzati a proporre una netta suddivisione tra *sex* (da *sex*) e *gender* (da *gender*), sul piano teorico-concettuale: il primo è inteso come corredo genetico di tratti biologici, fisici e anatomici, mentre il secondo è definito come costruzione culturale, come definizione e rappresentazione di comportamenti associati a quel corredo biologico. È quindi la cultura umana nel suo rinforzo sociale dell'identità che definisce le differenze tra uomo e donna e ne determina i cosiddetti *ruoli di genere*. Di conseguenza, se maschi o femmine si nasce, uomini o donne si diventa (De Beauvoir, 1984).

La diffusione del termine è da attribuire negli anni Settanta ad un'antropologa statunitense conosciuta per il suo attivismo in materia di studi di genere, dal nome Gayle Rubin. Quest'ultima attraverso i suoi scritti approfondisce il filo conduttore che lega proprio la distinzione biologico-anatomica dei sessi alla costruzione sociale dei ruoli di genere. Nel suo saggio *The Traffic in Women* (1975), l'autrice definisce il genere come il prodotto di un processo di adattamento e comportamento attraverso il quale, ogni società trasforma la sessualità biologica in costrutti che organizzano e definiscono i ruoli sociali, sottolineando le differenze tra uomo e donna e consolidando la distinzione di genere. Secondo la Rubin (1975) si tratta di un prodotto sociale e culturale che tenta di "annullare le somiglianze naturali" (*Ibidem*) tra gli individui e che mira a mettere in evidenza rigide divisioni tra le caratteristiche di uomini e donne.

L'appartenenza ad un genere rappresenta una forma di costruzione identitaria che è incorporata ed in continua elaborazione e definizione (West e Zimmerman, 1987). La categorizzazione delle identità e dei ruoli è introiettata nella società, la quale trasforma e plasma le strutture e i contesti attribuendo significati di appartenenza, con lo scopo di legittimarne le differenze. A tal proposito, Raewyn Connell (2006) sostiene che i corpi sociali siano al contempo oggetti e soggetti delle pratiche sociali stesse e che pertanto, le dinamiche di violenza non fanno altro che rappresentare una trasformazione dei corpi sociali che si modificano in relazione ai nuovi contesti ma che al contempo rimangono ancorati ad una visione fallocratica prettamente ancestrale della società. Tale arbitraria divisione delle cose e delle attività connesse alla distinzione di genere esprime il bisogno oggettivo e soggettivo di delineare un sistema di opposizioni. L'opposizione differenziante che ne deriva attribuisce uno spessore semantico che deriva da sovraderminazioni e connotazioni al genere corrispondenti. Questi schemi di pensieri assumono pertanto nella società un ruolo pressoché mitico-rituale, di equivalenza all'ordinamento giuridico di un paese, poiché ne fa confluire i principi di visione che consacrano un ordine stabilito, conosciuto, riconosciuto.

L'archetipo della ricostruzione storica legittima quindi il riconoscimento e la sedimentazione di un patrimonio di concezioni psichiche dell'Uomo, quale strumento in grado di comprendere l'ordine esterno e la realtà interiore. Si tratta di sorta di *DNA psichico* (Jun, 1977) che rappresenta la realizzazione di "strutture profonde capaci di indicare le vie più autentiche dell'esistenza [...] e di riflettere sulla struttura psicologica umana di base [...] e sui significati universali" (*Ibidem*, 1977). L'individuo in qualità di soggetto mutuamente interagente necessita di strutture simboliche volte ad indicare la chiave di lettura della cultura del tempo e del proprio ordinamento sociale. Di conseguenza, "I popoli delle grandi civiltà sono stati inclini ad interpretare figure simboliche [...] intese come supporti e sostegni di ordinamenti culturali e morali e come elementi di coesione [...]" (Campbell, 1958). La tradizione antropologica non ha fatto che difendere l'incorporazione di un'asimmetria inscritta nella natura biologica degli individui, pertanto la forza dell'ordine maschile non deve trovare una giustificazione che ne legittimi l'esistenza, in quanto la visione sociale che incorpora la differenza anatomica si inserisce in un rapporto di causalità che racchiude un'evidenza di dominio maschile.

1.5.2. Il dominio maschile

La determinazione delle differenze di genere indica il prodotto di un lavoro di costruzione sociale che difende la tradizionale antropologica e che organizza il mondo legittimando un rapporto di dominio dell'uomo sulla donna, trovando quindi nella natura biologica una costruzione sociale naturalizzata.

Pertanto, la trattazione teorico-concettuale del dominio maschile non può prescindere dal binomio uomo-donna, in quanto entrambi i protagonisti sono parte fondamentale della relazione conflittuale che esiste tra i generi. Il lavoro di costruzione sociale non si limita ad una *performance* nominativa ma si spinge verso una vera e propria rappresentazione oggettiva che differenzia la donna nella rappresentazione pratica del corpo e del cervello. La cultura quindi incarna una *somatizzazione dei rapporti sociali di dominio* (Bourdieu, 1998), trasponendo sul piano relazionale il corpo socialmente differenziato del genere, oggettivamente e soggettivamente condiviso. È in questo contesto che si iscrive la cultura del patriarcato, dove è l'uomo a dominare la scena, nella sfera pubblica e quindi politico-istituzionale o professionale e nella sfera privata, sottomottendo la donna che fa parte del proprio *entourage* familiare.

In particolare, l'approccio funzionalista vede la società come basata su un sistema organico, tipico delle scienze biologiche, nel quale le singole parti rispondono a funzioni specifiche; ed anche nel nostro apparato socio-culturale ci si aspetta di rintracciare funzionalità di ruoli e poteri connessi agli individui. L'organizzazione sociale ripartisce arbitrariamente i compiti differenziandoli e rafforzando il concetto di genere. Si tratta di un sistema nel quale le donne sono donne e quindi rilegate ad una dimensione prettamente domestica ed uterino-genitoriale, mentre gli uomini sono uomini, e come tali sono destinati ad un'ascesa sociale e professionale. Non basta nascere in uno Stato di diritto perché le facoltà garantite agli esseri umani siano equamente riconosciute ma sono discrezionalmente ripartite e dipendenti dalla natura biologica e dagli attributi e ruoli che ne derivano. Le decisioni spettano pertanto all'uomo che definisce ruoli e condizioni della vita relazionale *con e dell'*altro sesso.

Lo studio combinato che esplora la relazione tra i generi risulta essere particolarmente feconda, se finalizzata ad un'analisi di ruoli sessuali calati nella socialità, soprattutto in periodi di trasformazione che vedono la natura biologica al contempo occultata e legittimata da una forte componente identitaria sociale. Il dominio sembra quindi appartenere all'ordine naturale delle cose, insito nella natura stessa dei corpi; ed è lo stesso Bourdieu (1998) ad attribuire

alla distinzione di genere una caratteristica simbolico-funzionale della società, che si riconosce in una relazione di dominio sessista, come frutto di un processo di costruzione sociale naturalizzato. La distinzione tra i generi ha espletato una caratterizzazione funzionale della comunità, ponendosi come riferimento in grado di modellare e regolamentare intere collettività. Si potrebbe affermare che esista un rapporto di reciproca interdipendenza tra gli aspetti funzionali e strutturali della società passata e quelli della realtà presente. Le condizioni dell'esistenza che si basano sulla distinzione di genere sono profondamente radicate nella storicità del tempo, tale che nella modernità appaiono naturali e senza insinuazione di dubbio, perché è da sempre che esiste il nesso di familiarità tra l'apparenza dell'aspetto biologico e la *biologizzazione* del sociale. Ne consegue quindi una costruzione sociale naturalizzata dei *generi*, in quanto *habitus* sessuati.

I rapporti sociali sono essi stessi dominati da un'influenza puramente simbolica della conoscenza o meglio *mis-conoscenza*, anche e soprattutto in una logica di dominio maschile. "L'ordine delle cose non è un ordine naturale contro il quale non si possa far nulla. È piuttosto una costruzione mentale, una visione del mondo con la quale l'uomo appaga la sua sete di dominio. Una visione esclusiva che le stesse donne, che ne sono vittime, l'hanno integrata nel proprio modo di pensare e nell'accettazione inconscia di inferiorità" (*Ibidem*). La ritualità che si consacra nell'oggettivazione conosciuta e riconosciuta nei corpi, negli *habitus* degli agenti, nelle percezioni e nelle azioni non fa che determinare quello Husserl definisce un'*attitudine naturale* di rapporto con il mondo (Spinicci, 1985). L'ancoraggio alla tradizione passata identifica una relazione di interdipendenza nelle strutture della identità e socialità condivisa. L'apparato simbolico nel quale si muovono le strutture androcentriche sopravvive grazie ad un processo di decodifica da parte degli esseri umani e ciascun soggetto segue, in quanto individuo facente parte di una società, un ordine stabilito di regole e consuetudini, nelle quali sono iscritte le dinamiche del dominio e del potere maschile. Anche la Rubin analizza la storicità del tempo che ha sottomesso la donna in nome della sua natura biologica, attribuendo al processo di naturalizzazione dell'*identità sessuata* il termine di *sex-gender system*, ovvero di sistema nelle sue strutture culturali che ha naturalizzato una suddivisione genetica ed identitaria incorporata nel funzionalismo sociale. Più propriamente, tra la metà degli anni Sessanta e Settanta, una fiorente letteratura ha accelerato incombenti riflessioni innovative a carico della condizione femminile. Sebbene oltreoceano e solo per una minoranza di uomini, la differenza di *genere* ha

fornito interessanti e decisivi spunti interpretativi di riconversione sociale. Infatti, negli anni Novanta, si segna l'esordio di un fenomeno noto come *men's studies*. Questa tipologia di studio vede gli stessi uomini interrogarsi sulla cultura sessista del tempo e attraverso la loro prospettiva, l'emancipazione femminile può rappresentare una forma di riscatto dalla presunzione della mascolinità, solo se tali paradigmi vengono *spogliati* dal simbolismo storico che li contraddistingue. Si tratta di un simbolismo che incatena gli uomini alla detenzione del potere e malgrado molti tentativi di occultamento, l'uomo rimane pur sempre alla ricerca di un'altra identità più fragile da sottomettere e dominare. Il dominio maschile quindi conferisce all'identità del maschio il potere; il potere di scegliere, il potere di comandare, il potere di sottomettere. Alcuni autori, come Corradi (2008) sostengono infatti che l'identità maschile trovi conferma della propria individualità di *maschio* solo attraverso la sopraffazione ed il dominio. L'autore ritiene altresì che sia devastante per un uomo, il timore della prevaricazione che potrebbe derivare se uomini e donne fossero soggetti alla pari. L'uomo si sente minacciato da una donna che non si lascia intimidire dalla sua virilità e che non obbedisce al suo volere di maschio. Pertanto, la frustrazione che deriva dall'ipotesi di perdere la propria autorità e quindi la propria identificazione nel mondo è priva di interrogativi morali e alla costante ricerca di conferme sessiste. In altre parole, la costruzione identitaria maschile si determina per mezzo della formalizzazione del dominio. Senza l'*identità sessuata* che differenzia maschi e femmine, l'uomo non avrebbe modo di esistere e di affermare la propria identità. Se venissero meno i principi identitari individuali e collettivi che si concretizzano attraverso le dinamiche di interazione, non esisterebbe il virilismo che da sempre caratterizza gli uomini. Sono quindi le interazioni e i contesti nei quali essa si formalizza a generare le strutture sociali che caratterizzano il dominio tra i generi. Infatti, come suggerito da Natalie Zemon Devis (1977) la problematizzazione della questione di genere è poco indicativa se la donna viene considerata come un universo isolato dall'uomo.

Il dominio maschile trova particolare spazio d'azione nell'ambito intra-parentale, aderendo ad una cultura patriarcale che tenta di rispondere con assidua prepotenza al principio della *virilità maschile*. Riprendendo le definizioni di Corradi (2008), il patriarcato oggi più che mai si sente messo sotto accusa dalle riformulazioni funzionaliste che mirano alla parità tra i generi, soprattutto tra le mura domestiche. Infatti, è proprio nell'ambito privato che l'idea di dominio si fa spazio, con un aggravante crescendo di severità e di

episodi di violenza. La contaminazione nel bilanciamento di potere e nella determinazione dei ruoli familiari rappresenta un senso di mistificazione e sofisticazione di quelle strutture gelosamente custodite dalla società maschilista. Pertanto, la vulnerabilità e conflittualità identitaria (Corradi, 2009) che ne conseguono, innescano un meccanismo di feroce disappunto che nell'ambito endo-parentale o di coppia può connotarsi come *violent resistance*, che Kelly e Johnston (2008) descrivono come una risposta all'alterazione strutturale della realtà sociale. L'uomo quindi tenta attraverso il dominio violento di arrestare l'affermarsi di una parità tra i generi. Il sottile confine tra il visibile e l'invisibile fornisce una spontanea legittimazione del dominio maschile, che si configura proprio attraverso le dinamiche violente (Jenkins, 1992). La naturalità della prevaricazione maschile conduce il soggetto dominato (la donna) ad una tacita adesione, le strutture cognitive hanno introiettato codici di lettura sociale maschilisti. L'accettazione spesso graduale ed interiorizzata che tendono a subordinare il soggetto dominato, mascherano in realtà la relazione di potere sottostante (Connolly e Healy, 2004).

1.5.3. *L'asimmetria simbolica*

L'effetto del dominio simbolico si esercita per mezzo di schemi percettivi e di modelli di azione che sono costitutivi degli *habitus*, in un rapporto di coscienza profondamente oscuro e paradossale, in quanto la sottomissione femminile diviene al contempo "spontanea ed estorta" (Bourdieu, 1998). La logica del paradosso si può osservare attraverso la forza simbolica esercitata sui corpi e appartenente all'ordine sociale, la quale si basa su disposizioni che appartengono all'universo simbolico depositato sia negli uomini che nelle donne e che agisce come un pulsante di accensione ad elementi e disposizioni che sono inculcati ed incorporati nella quotidianità. Pertanto, nell'ambito della relazione che vede l'asimmetria tra uomo e donna, si può parlare di simbolismo della violenza. Tale simbolismo rappresenta un interessante ponte concettuale e teorico per comprendere il dominio maschile. L'interpretazione del potere e della sottomissione è soprattutto ascrivibile ad una sorta di *significato sociale del sesso* (MacKinnon, 1989), nel quale sono i fattori sociali, politici, economici e culturali a dare luogo ai meccanismi del dominio (da parte dell'uomo) e della subordinazione (da parte della donna). L'invisibilità che

caratterizza la violenza simbolica identifica un silenzioso dispositivo capace di insidiare le strutture cognitive dei soggetti coinvolti nella conflittuale relazione di rivendicazione del potere. Specificamente, la dinamica tenta di stabilire chi è al comando della relazione e chi invece subisce, in una silente abnegazione del soggetto dominato che contribuisce tacitamente alla costruzione delle strutture sociali, *inquadrando i dialoghi dell'inter-soggettività* (Bhambra e Shilliam, 2009) e diventando un'interpretazione naturalizzata. Il linguaggio stesso è una forma di dominio attraverso cui si includono e si escludono i discorsi, formalizzando l'asimmetria tra uomo e donna. La supremazia androcentrica è quindi parte incorporata dell'oggettività del senso comune, che viene intesa come un *consenso pratico e dossico* (Bourdieu, 1998), affermandosi anche nelle pratiche produttive e riproduttive. Quando si parla di forza simbolica, l'uomo non vuole esercitare coercizione fisica o psicologica sulla donna ma fa leva su disposizioni già depositate in modo invisibile ed insidioso mediante un sistema di familiarizzazione, la cui familiarità è contestualmente rilegata all'esperienza prolungata con le strutture del dominio. Il dominatore (l'uomo) agisce attraverso gli strumenti simbolici per avere una "schiava" che spontaneamente accetti e agisca per vocazione alla sua condizione di schiavitù. Il linguaggio simbolico della legittimazione annulla la responsabilità dell'oppressore, poiché la dominazione vive e si nutre di un sistema di strutture preordinate. La vittima diventa tacitamente complice della violenza subita, ricorrendo ad una consuetudine del linguaggio simbolico che la donna ha sempre vissuto. A questo proposito, Bourdieu (2001) propone una similitudine in grado di legittimare la condizione di accettazione conscia ed inconscia del dominio maschile da parte della donna. Egli indaga come la comunicazione giornalistica focalizzi l'argomentazione delle notizie basandosi su un particolare aspetto che determina la cosiddetta *notiziabilità* dei casi. L'uso di un linguaggio sensazionalistico è in grado di generare nell'ambito dei media una teoria detta ipodermica (*Bullet Theory*), secondo la quale i mass-media sono potenti strumenti persuasivi in grado di agire sugli individui che sono soggetti passivi. Allo stesso modo, il linguaggio simbolico agisce sull'ordine sociale e mentale della donna e con la sua complicità ne attiva i meccanismi di subordinazione e passività. Pertanto, il concetto di violenza simbolica non vuole minare o sminuire quello della vessazione fisica o psicologica, poiché talvolta coesistono simultaneamente ma se la violenza simbolica e quella diretta vengono agite separatamente non sono sempre distinguibili e riconoscibili. Bourdieu (1998) suggerisce di superare l'opposizione tra la visione fisica del

mondo sociale (intesa come rapporti di forza tra individui) e la visione *cibernetica* o semiologica (che ritrae i rapporti come relazioni di significato simbolico), poiché egli sottolinea che i rapporti di forza più brutali sono sempre più quelli dati dalle relazioni simboliche (Bourdieu,1998). Un'anima ferita fa più rumore di un corpo moribondo e violenza simbolica è un'interazione asimmetrica governata da una "forma invisibile di violenza" che diventa la modalità di dominio più sicura, in quanto manca della qualità intenzionale e strumentale della forza fisica brutale e coercitiva e "lavora non direttamente sui corpi ma attraverso di loro" (Topper, 2001). Quindi, la violenza simbolica è un circolo vizioso di abuso altamente subdolo per mezzo di dinamiche che non consentono alla vittima di "riconoscere la violenza, in quanto non la si percepisce come tale" (Bourdieu e Wacquant, 1992).

L'asimmetria simbolica è inscritta in un sistema mitico-rituale, tale per cui il principio di divisione è di tipo soggetto/oggetto ed agente/strumento nel dominio degli scambi simbolici. Infatti nel mercato matrimoniale la donna è l'oggetto la cui funzione strumentale è di contribuire alla perpetuazione del capitale simbolico posseduto dal soggetto/agente uomo. La donna appare quindi vocata all'abnegazione e nell'ambito della sfera parentale la violenza simbolica rappresenta l'aspetto più pericoloso del rapporto, in quanto agisce con una serie di strategie che portano la vittima a dipendere dal proprio aggressore, in un'alternanza di armonia e asimmetria. Si tratta di creare *mis-conoscenza* attraverso processi di punizione e ricompensa, in modo che i due processi diventino quasi indistinguibili per la vittima (Bourgois *et al.*, 2004) in una spirale senza via d'uscita. Tuttavia, anche l'uomo rimane vittima del suo stesso dominio, poiché il privilegio di essere uomo richiede imposizione di virtù e costante virilità, in tensione con le caratteristiche tipiche della femminilità. Egli concorre a fare dell'ambita virilità la sua stessa vulnerabilità, nella ricerca di costante approvazione davanti agli altri uomini, attraverso manifestazioni di violenza reale o potenziale come distintivo di appartenenza ai *veri uomini*.

1.5.4. La virilità

La costruzione dell'identità maschile e del suo potere egemonico possiede un forte legame con il concetto di *virilità*. Più propriamente, lo storico Bellasai (2011) indica la virilità come definizione promossa dalle società Occidentali a

partire dal Diciannovesimo Secolo, al fine di affermare una prospettiva politica fondata sulla mascolinità e sulla sua forza dominatrice e violenta. Tale visione nasce in opposizione alla fragilità femminile, che nei processi di modernizzazione e nei primi movimenti femminili si presenta come un attentato finalizzato a ridimensionare l'esistenza della potente forza maschile. La virilità, al pari della nobiltà, identifica una virtù d'onore, inscritta nel corpo per mezzo di un insieme di disposizioni naturali. Si tratta di una forza superiore imprescindibile, di una trascendenza sociale predestinata, che la scrittrice femminista Virginia Woolf mostra come "linee di demarcazione mistica". Tuttavia, questo privilegio ha la sua contropartita, poiché l'uomo deve impegnarsi costantemente a dimostrare in qualsiasi circostanza di essere degno di una tale superiorità, intesa come capacità produttiva, riproduttiva, sessuale, sociale e di predisposizione all'esercizio della violenza. L'esaltazione dei valori maschili, contrapposti a quelli di debolezza, verginità e fedeltà femminili, presenta al contempo un'angosciosa paura del fallimento che porta alla frustrazione e alla vulnerabilità. Pertanto, la virilità si esprime nella relazione disarmonica con la donna e contro la donna, in una sorta di paura contro tutto ciò che è femminile.

La mascolinità e la virilità trovano ampio spazio nella cultura patriarcale, nel quale contesto si legittima e si rafforza lo squilibrio tra i generi e la prevaricazione maschile. L'aderenza alle norme sociali che inglobano una cultura di genere è di fondamentale importanza al fine di raggiungere lo *status* di *vero uomo*, in un processo sociale che riconosce e legittima anche la violenza. Si tratta di una forma di accettazione della violenza maschile sulle donne, che lo scenario sociale conosce e riconosce in un'affermazione di potere e virilità maschile. Quindi, la costruzione della virilità si definisce per mezzo dei rapporti di potere, nei quali si sottomette la donna e si determinano i limiti connessi al rapporto con la stessa. Infatti, le relazioni sono svuotate da ogni interesse o investimento emotivo (Connel, 1996). Inoltre, nella letteratura si fa riferimento alla capacità di produzione che caratterizza la mascolinità, in un'ottica di presenza nell'economia capitalistica (*Ibidem*).

La virilità deve condurre l'uomo a non intraprendere relazioni insane o innaturali ed in particolare, alcuni autori come Flood (2002) indicano alcuni riti di passaggio alla mascolinità virile. In prima istanza rappresenta un esempio di virilità la competizione e la capacità di iniziativa nell'intraprendere delle relazioni. Esiste quindi un legame tra la posizione di potere dell'uomo e la sua adeguatezza ad intrattenere relazioni con l'altro sesso, soprattutto di natura

sessuale. La sessualità non deve affatto rappresentare imbarazzo per un *vero uomo* ma deve anzi essere un elemento di scambio dialettico tra la sua soggettività e la reificazione della donna, in un contesto di profondo distanziamento dalla figura della donna e svalutazione del suo universo femminile. L'attitudine alle pratiche sessiste e pornografiche identificano un ulteriore elemento di appartenenza alla politica del virilismo; così come le pratiche violente e coercitive. Tutto serve ad avvalorare la mascolinità dominante, in opposizione con qualsiasi forma di ambivalenza eterosessuale, all'omosessualità e alla femminilità stessa, delle accezioni negative. Pertanto, lo *status* sessuale rappresenta uno dei punti focali nel riconoscimento della virilità. Nella dicotomia sessuale, all'uomo che presenta un certo numero di esperienze sessuali viene conferita una celebrazione del mondo maschile, mentre in opposizione, tale licenza sessuale conferisce alla donna un'etichetta di denigrazione. Per concludere, il rituale della virilità prevede l'attitudine alla violenza contro le donne nel caso di un eventuale diniego da parte delle stesse e lo spettro dell'omofobia deve essere rinnegato in modo imperante. Tuttavia, nel corso del tempo, la politica del virilismo viene meno, soprattutto con l'avvento della comunicazione di massa, in un primo momento perché conferisce alle donne una qualifica esperta di "custode del focolaio domestico" e successivamente, con le trasformazioni socio-culturali dei vari periodi storici, si è passati ad una stigmatizzazione della donna di successo come carrierista e spietata. In realtà, le sono state attribuite le qualità della *leadership* da sempre connesse all'uomo, le quali però associate allo scenario femminile si sporcano di svilimento e mortificazione. In particolare, la comunicazione ha forgiato un sistema di svilimento femminile, incalzando sui valori del sessismo maschilista, proiettando senza indugi la figura della donna nella sua più autentica mercificazione del corpo. Quindi, se la virilità maschile viene messa a nudi, i contesti socio-culturali si sono spostati verso scenari di autentica misoginia, nel tentativo di mantenere inalterato il potente virilismo ormai in discesa.

1.5.5. L'esibizionismo simbolico: immaginario e rappresentazione mediatica di genere

La mascolinizzazione e femminilizzazione sono spesso connesse ad un esibizionismo di tipo simbolico.

Più propriamente, l'esibizionismo simbolico rappresenta l'essere femminile come *essere-percepito*, in un'ottica di strutturazione simbolica uniformata attraverso elementi stereotipati capaci di *sedimentare il sedimentato*.

La rappresentazione di genere diviene quindi un modo per classificare gli individui e per comprovare la duplice esistenza dei due generi: maschio e femmina (Piccone e Saraceno, 1996), in un'ottica di segno distintivo della realtà sociale. Il genere può essere interpretato come un processo semiotico caratterizzato dall'identità binaria di *significato* e *significante*, nel quale binomio gli individui sono soggetti culturali che vivono, interpretano ed agiscono in un ambiente significante. Pertanto, ai fini dell'elaborazione dei significati e dei significanti su larga scala, un ruolo determinante è conferito all'industria culturale. Essa si impegna ad avvalorare i significati della realtà sociale che sono iscritti nei significanti, ragion per cui, il genere diventa un potente dispositivo semiotico (De Lauretis, 1999) in una realtà di corpi sessuati prodotti e rappresentati attraverso i media. In questo senso, l'immaginario si intende come il prodotto di un repertorio di rappresentazioni simboliche, fantasie e miti, dato dalla "coerenza" in termini di riconoscibilità e di memoria nel tempo. La riconoscibilità non è altro che uno strumento che innesca attributi già impressi nella memoria storica dell'individuo, i quali legittimano lo *status* di inferiorità simbolica della donna.

In particolare, nella genesi dell'*habitus* ogni cosa concorre a fare dell'esperienza femminile del corpo *il limite dell'esperienza universale del corpo -per-gli- altri* (Bourdieu, 1998), e nella rappresentazione mediatica il corpo della donna raggiunge l'apice dell'oggettivazione operata dallo sguardo e dal discorso altrui. La connotazione è data dalla percezione sociale di un corpo nudo o vestito, in una prospettiva dicotomica che si traduce nella rappresentazione della donna nel suo *abito sessuato* (Bobbioni, 2012). L'*habitus* agisce come struttura cognitiva in grado di riconoscere e coordinare le esperienze che derivano dal possedere un abito sessuato o un altro e quindi dall'essere uomo o donna. La visione corporea è dunque sinonimo di identità sessuale, in una reiterata esperienza cognitiva e materiale che può essere reale o mediata. La formazione dell'*habitus* sessuato si origina dall'esperienza ma anche attraverso l'immagine proiettata dalla cultura dominante, come punto di contatto tra il mondo interno e quello esterno. Il mondo interno rappresenta la porzione psichica dell'individuo, nella quale si costituisce la realtà *dell'abito sessuato* e delle *fantasie* in cui la vita del soggetto prende forma. In tal senso, un ruolo fondamentale è giocato dalla coscienza e dal desiderio, le cui istanze

determinano scenari di congiunzione tra la significazione sociale e la realtà materiale. In altri termini, vi è una linea di connessione particolarmente profonda tra il mondo interno ed esterno ed il modo in cui il corpo realizza il suo effetto di significazione. Il soggetto è “coinvolto fisicamente e corporalmente nella produzione di significato, nella rappresentazione e autorappresentazione” (De Lauretis, 1984). Pertanto, il rapporto con il proprio corpo non si riduce alla rappresentazione soggettiva di *self-image o looking-glass self* ma si modella in funzione ad un determinato grado di *self-esteem*, ossia di criterio di percezione oggettiva dell’occhio di chi osserva. Il dominio maschile che collega la donna all’idea di oggetto simbolico finisce con il porre l’individuo femminile in uno stato di permanente insicurezza corporea o alienazione simbolica. La pretesa femminilità non è altro che una forma di compiacenza nei confronti delle aspettative maschili, reali o supposte, in un dominio che vede l’esaltazione ed il nutrimento dell’ego maschile. La comunicazione pubblicitaria attuale si serve di dispositivi comunicativi riecheggiano a contenuti sessuali, assistendo ad una forma di de-umanizzazione e di oggettivazione. Da questo punto di vista, l’individuo è pensato e trattato come merce o strumento sessuale (Volpato, 2011) e l’industria culturale che costituisce la risorsa primaria della comunicazione commerciale, non fa che incentivare tale prospettiva. Sussiste uno stretto legame tra l’esposizione mediatica del corpo femminile e i riferimenti alla sessualità. In altre parole, la donna viene esposta nella sua corporeità in una vetrina delle offerte, pronta a rispondere ad una domanda di mercato finalizzata all’appagamento dell’immaginario maschile. Si può parlare di un’*iper-sessualizzazione*, come punto focale dell’intero sistema mediatico contemporaneo. L’industria culturale presenta un pervasivo e sistemico scambio tra sesso e potere, nel quale il sesso si trasforma in potere, e con lo scopo di compiacere il desiderio maschile vengono propinati corpi gonfiati a dismisura, che si distaccano dalla corporeità come in un *corpo alieno* (Bourdieu, 1998) e che ci rimandano all’ossessiva ricerca della perfezione per mezzo di tecniche di *body-contouring*. Le donne si trasformano in individui “geneticamente modificati in cerca di omologazione, dove il fascino diventa un vero e proprio culto (Wolf, 1991). La rappresentazione mediatica di genere tende a trasformare le donne in un artefatto mediatico nel tentativo di mantenere inalterata la propria giovinezza, mentre il corpo maschile viene accettato nel suo naturale passaggio del tempo e al mutare delle sue caratteristiche. Al contrario, la corporeità femminile diventa oggetto di scrutinio, come una sezione d’indagine a parte.

“[...] Le donne non sono una fauna speciale e non capisco per quale ragione debbano costituire [...] un argomento a parte: come lo sport, la politica, il bollettino meteorologico. I problemi fondamentali degli uomini nascono da questioni economiche, razziali, sociali ma i problemi fondamentali delle donne nascono [...] soprattutto dal fatto di essere donne” (Oriana Fallaci, 1961).

Il “fatto di essere donne” a cui fa riferimento la Fallaci (1961) tenta faticosamente di affermare una presa di coscienza femminile che si propone di spezzare le catene che l’hanno imprigionata nel ruolo di essere *femmina* e quindi un corpo senza cervello. Tale vincolo viene ancora oggi perpetrato dai media, i quali contribuiscono a scolpire nell’immaginario l’oggettivazione della donna, propinando la donna nella sua funzione di cornice o *valletta muta*, in un’ideale di soggetto-oggetto privo di competenza e facilmente manipolabile; come un complemento d’arredo capace solo di catturare l’attenzione dello spettatore maschio. Questo svilimento della dignità femminile e delle sue capacità cognitive è una forma di violenza che offende la dignità umana della donna, imprimendo nell’immaginario collettivo, l’idea che le donne non possano ambire alle medesime posizioni e carriere intraprese dagli uomini

1.6. Verso un cambiamento di prospettiva: il Femminismo

Gli studi sociologici portano a delineare dunque un sistema di relazioni e significati basati sul genere, definendo le caratteristiche e i confini dei ruoli sociali. La consapevolezza della disuguaglianza tra i generi e l’epilogo del mito patriarcale confluiscono in un primo vero movimento politico di critica storica alla famiglia e alla società (Manifesto di Rivolta femminile, 1970), attraverso l’affermarsi del “femminismo”. Non si tratta di un semplice neologismo che si afferma su scala universale ma rappresenta un movimento politico di donne impegnate nell’affermazione della parità politica, sociale ed economica tra i sessi, ritenendo che le donne siano state e siano, in varie misure, discriminate rispetto agli uomini. Tale movimento porta con sé un’elaborazione teorica e concettuale mirata a denunciare, analizzare e approfondire nelle loro implicazioni interdisciplinari, le modalità e le strumentazioni culturali con cui si è perpetuato nei secoli il dominio maschile. Tuttavia, gli approcci teorici sono disparati ed è per questo che molti studiosi parlano di “femminismi”. Nella storia del femminismo intesa come narrazione cronologica degli eventi

riconducibili all'affermazione dell'uguaglianza di genere, tale neologismo si fa risalire al filosofo francese Charles Fourier, nel 1837, il quale attraverso i propri scritti dimostrava di essere un sostenitore della libertà e dell'egualitarismo femminile, anche se il concetto di femminismo non compare mai all'interno della sua produzione creativa ed è solo intorno al 1870, in ambito medico, che esso appare per descrivere lo sviluppo anatomico-funzionale maschile di tipo patologico. Pertanto, in letteratura, si rintracciano i primi riferimenti nell'opera *L'Homme-femme* (1872) di Alexandre Dumas (figlio). Le origini del pensiero femminista come vera e propria azione politica risalgono alla fine del Settecento e all'inizio dell'Ottocento ad opera dell'antesignana Olympe de Gouges, pionera della "*Dichiarazione dei diritti della donna e della cittadina*" (1791). L'anno successivo anche in Inghilterra si affermano idee di egualitarismo e Mary Wollstonecraft scrive "*The Vindication of the Rights of Women*", contrastando la teoria di Rousseau sull'inferiorità "naturale" della donna e rivendicando l'accesso all'istruzione e alla cultura. In realtà, la storia femminista moderna occidentale viene suddivisa in tre periodi temporali (o "ondate"; Weinman, 1968) ognuno dei quali persegue obiettivi di progresso differenti.

La prima ondata femminista si fa risalire tra il XIX secolo e la prima metà del XX secolo, periodo nel quale il *focus* è il ribaltamento delle disuguaglianze tra uomini e donne. Il femminismo di questo periodo storico rivendica il diritto al voto, il diritto al lavoro in condizioni sostenibili ed il diritto all'istruzione. La prima ondata si sviluppa nei Paesi occidentali avanzati, soprattutto Francia, Inghilterra e Stati Uniti. Il primo periodo trova spazio in piena seconda rivoluzione industriale ed evoluzione economico-produttiva, nella quale si assiste ad un processo di urbanizzazione e trasformazione della società agricolo-artigianale e commerciale. In questa prima fase, il "movimento delle donne" si divide in due correnti fondamentali: quella liberale e quella socialista. Nel primo caso, vi afferiscono le donne appartenenti alla classe media che rispetto alle donne che per necessità offrono sul mercato la propria forza-lavoro, rappresentano uno *status* privilegiato perché mantenute dai familiari maschi. Queste tuttavia rivendicano la propria indipendenza, l'accesso all'istruzione, l'autonomia patrimoniale ed il diritto al voto. Una delle figure più significative dell'orientamento liberale è Harriet Hardy Taylor Mill che nel 1851 porta alla luce il concetto di "emancipazione delle donne", pubblicando sulla rivista *The Westminster Review*, a nome del marito John Stuart Mill, una realtà tutta americana che denuncia la mistificazione maschile che insinua la sua

superiorità biologica. La Mill fa riferimento al movimento delle cosiddette “suffragette” prendendo coscienza pubblica del diritto di voto, dei diritti di cittadinanza e della necessità di far cessare l’oppressione domestica. La corrente socialista invece si rivolge alle donne di rango proletario che vivono in una situazione di conflitto con il mondo del lavoro maschilista. Esse rivendicano faticosamente la necessità di accedere alla formazione scolastica, alle libere professioni, tipicamente maschili e chiedono la parità salariale. Il fulcro di ispirazione della corrente socialista è l’opera di Friedrich Engels, *“L’origine della famiglia, della proprietà privata e dello Stato”* (1884), che si fonda sul pensiero marxista e che dà la priorità alla lotta di classi, ponendo in posizione minoritaria la rivendicazione di genere. In un primo momento infatti la questione del suffragio della corrente liberale è stata considerata con scarsa attenzione da parte delle stesse femministe. Solo nel corso del tempo la rappresentanza legislativa ha acquisito importanza nel cambiamento paradigmatico della questione di genere. Il XIX Secolo rappresenta comunque un importante e significativo periodo di mutamento europeo, in quanto le idee femministe si insinuano all’interno di un fervido terreno di movimenti nazionalisti e democratici. Infatti, in Germania, nel 1840, si sviluppa un gruppo di cattolici tedeschi (*“Deutschkatholizismus”*) che iniziano ad inglobare nella loro dottrina e critica la questione delle donne nella società, portando le donne tedesche ad essere, pochi anni, parte attiva di una politica che ha determinato e incentivato la “la primavera dei popoli” e che ha coinvolto l’Europa in moti rivoluzionari borghesi nel 1848 e 1849. Anche in Svizzera, già dal 1845, grazie al contributo di Josephine Zehnder-Stadlin, in Polonia tramite Narcyza Zmichowska e in Giappone con il supporto di Fusae Ichikawa si promuovono i valori del movimento femminista. In Australia invece il forte attivismo di cui Mary Lee è una delle figure più espressive del tempo, fa sì che nel 1894, venga ufficialmente concesso il suffragio universale alle donne. Si tratta del primo caso al mondo in cui le donne acquistano il diritto di voto. Questi sono solo alcuni degli esempi di attivismo femminista, poiché molte figure significative hanno dominato la scena in quegli anni, in diverse località del mondo. Tuttavia, negli anni successivi, le evoluzioni femministe si arrestano a causa dei due conflitti bellici mondiali. Infatti, per quel che concerne il diritto di voto esteso anche alle donne, ad esempio la Francia dovrà attendere il 1945 e l’Italia il 1946. Mentre, in molti paesi del mondo, in quegli anni le donne avevano ottenuto le conquiste più importanti come il diritto di voto, il diritto all’istruzione o il diritto di accedere al mercato del lavoro, le guerre avevano

appannato il femminismo europeo, ma è proprio in questi periodi che alcune scrittrici formulano pensieri che saranno l'*incipit* della seconda ondata di femminismo. In *primis*, Virginia Woolf che pone le basi alla teoria e critica letteraria femminista sulla differenza di genere. L'uguaglianza formale con gli uomini che la scrittrice propone in due saggi "*Una stanza tutta per sé*" ("*A Room of One's Own*", 1929) e "*Tre ghinee*" ("*Three Guineas*", 1938), presentano una contaminazione dei valori etici e politici delle donne a cui sono stati imposti quelli maschili. Secondo la Woolf l'affermazione ed il riconoscimento della parità tra uomo e donna devono rafforzare le differenze e non le somiglianze. In questo stesso periodo, cosiddetto di "riflusso" emerge un'altra figura di spicco che afferma l'identità sessuale intesa come costruzione sociale. Si tratta di Simone De Beauvoir che nell'opera "Il secondo sesso" (1949) sostiene difficile sia per una donna conoscere veramente sé stessa a causa di tutte le identità che le vengono attribuite dalla cultura ufficiale. Una scrittrice che segna il passaggio al secondo periodo femminista dopo la Woolf e la De Beauvoir è la statunitense Betty Friedan. Nel 1957, l'autrice conduce un'indagine sul grado di soddisfazione della vita delle donne. I risultati che ne derivano, pubblicati in "*The Feminine Mystique*" (1963) portano alla luce una cultura dominante che conduce le donne a credere che la felicità risieda nella devozione verso la casa e la famiglia. In realtà è proprio la vita familiare e domestica è proprio il primo ambito di frustrazione e insoddisfazione. Il "femminismo della seconda ondata" conosciuto anche sotto il nome di "femminismo radicale" si sviluppa quindi a partire dal 1968 e perdura sulla scena storica fino al 1980. La letteratura afferma che il "secondo tempo" del femminismo nasce negli Stati Uniti d'America e rapidamente si diffonde prima in Gran Bretagna e poi in tutto resto dell'Europa occidentale. In questa fase storica si rivendica l'uguaglianza e si condannano le azioni spregevoli e violente maschili anche e soprattutto in ambito riproduttivo, ritenendo la cultura il principale deterrente. Mentre l'obiettivo della prima ondata femminista è il raggiungimento della parità nei diritti giuridico-civili tra uomini e donne, il secondo periodo si focalizza sul tema del corpo e sul diritto di disporre in piena libertà. Pertanto, si evidenziano le libertà sessuali, di riproduzione biologica e la contraccezione e l'aborto diventano argomenti collaterali. In questa seconda fase, si lascia inoltre spazio ad un'altra componente importante: il femminismo lesbico e quindi la nascita dei "*Lesbian Studies*". Il tema era, in realtà, già noto nel panorama letterario dal 1971 (di cui la più identificativa "*The Woman Identified Woman*") e poi è stato ampiamente sviluppato in quel periodo dai

gruppi di riferimento (“*Radicalesbians*”) con svariate pubblicazioni (Rich, 1929; Rubin, 1975). Tuttavia, in Francia, il secondo periodo femminista assume connotazioni differenti e la produzione teorica vede nelle differenze di genere non un sinonimo di discriminazione ma un’opportunità di autocoscienza delle specificità femminili (Woolf, 1929; 1938). Con l’avvento degli anni Ottanta, il movimento femminista nella sua azione politica sembra affievolire, restando tuttavia vitale in termini di elaborazione teorica. In particolare, in zona angloamericana si creano dipartimenti dedicati agli studi di genere con un *focus* particolare sull’identità, soggettività, sessualità e corporeità seguendo la teoria decostruzionista francese. In questo periodo storico il ruolo della donna si “accademizza” e diventa parte significativa della ricerca scientifica sul campo. Il corpo umano viene connotato di un dato biologico e di un campo di iscrizioni e codici socio-culturali. Per meglio dire, la tecnologia diventa una modalità attraverso cui costruire il sistema sociale che controlla e condiziona i corpi socializzati. L’aspetto tecnologico consente di uscire dall’ottica binaria di materialismo uomo/donna, poiché non è rintracciabile l’aspetto corporeo. Pertanto, il “*cyborg*” teorizzato da Donna Haraway (1985) non è altro che un individuo ibrido per metà macchina (*cyb-*) e per metà organismo (*-org*) che in qualche modo appartiene tanto alla realtà sociale quanto a quella della finzione e della fantascienza. Il rapporto uomo-donna viene quindi caratterizzato da una mescolanza di tecnologia ed umanità volta ad indicare la mancanza di distinzione tra realtà sociale vissuta e realtà proiettata in termini di finzione, in quanto una categoria definisce e ridefinisce l’altra. Teresa de Lauretis (1990) invece sostiene che nessuna identità sia immutabile ma che ciascuna sia il “simulacro” di un qualcosa che non esiste come dato di fatto presente in natura e parla di “tecnologie di genere”, ossia del prodotto dei discorsi, delle istituzioni e delle narrazioni culturali. Un’altra figurazione è quella ipotizzata dalla Braidotti (1991; 1994) riconducibile all’affermarsi di una filosofia femminista “nomadica”. In altre parole, l’autrice elabora una schematizzazione a tre livelli: la differenza tra uomo e donna è insista nel simbolismo maschile e quindi si parla di una falsa universalità. In secondo luogo, si rifiuta “l’emancipazionismo” poiché si rischia la perdita dei principi etici e politici propri della donna (Woolf, V., 1929; 1938) a favore di un processo di omologazione. Il terzo livello invece è incorporato nella donna stessa, ovvero il concetto di differenziazione transita “*nomadicamente*” tra il conscio e l’incoscio. Dalle prime evoluzioni teoriche e pratiche degli anni novanta ha così inizio il femminismo di “terza ondata”. Il femminismo del terzo periodo

abbandona la battaglia di gruppo per approcciarsi alla necessità di una lotta individuale, sostenendo l'unicità di ogni donna e affermando il cosiddetto femminismo "*sessual-positivo*" che non teme di attaccare il femminismo radicale rivendicando il diritto alla "sessualità differente" (Beasley *et al.*, 2012). Questa ultima ondata si caratterizza per ironia e umorismo. Infatti, i toni così differenti rispetto alla prima e seconda ondata, permettono ad alcuni studiosi, di estendere il femminismo di terza ondata fino agli anni Duemila. Altri ancora invece sostengono la presenza di un quarto periodo del femminismo, iniziato nel 2008 e presente fino ad oggi (Cochrane, 2013b). La cosiddetta quarta ondata è strettamente connessa al concetto di tecnologia (Baumgardner, 2011; Cochrane, 2013a) e si unisce alla politica, psicologia e spiritualità (Diamond, 2009). I temi di maggiore interesse sono le campagne pubblicitarie discriminatorie e misogine proposte sul *Web* e le molestie di natura sessuale esercitate per strada o sul posto di lavoro.

1.7. Le conquiste femminili attraverso l'inquadramento normativo

L'impegno femminile e le lotte femministe che hanno contrassegnato la storia conducendola verso l'emancipazione e la fine del mito patriarcale, al contempo, hanno caratterizzato diversi cambiamenti nell'assetto normativo e nelle politiche di conciliazione. Più specificamente, per molti anni, l'inquadramento normativo non ha prodotto significativi passi in avanti, fino a quando la questione di genere non è diventata una vera e propria emergenza sociale, per poi contraddire questa sua caratteristica ed essere connessa alla peculiarità della persistenza, in un funzionalismo sociale del tutto androcentrico.

Poiché si tratta di un fenomeno assolutamente trasversale, che non conosce alcuna distinzione geografica, di *status*, razza, religione, cultura o *sub-cultura*, la giurisdizione sul tema della violenza contro le donne è principalmente di natura internazionale. Infatti, in questo contesto si mira ad inquadrare l'assetto normativo e politico-legislativo che ha cambiato la società, non soffermandosi su un singolo Stato ma allargando la prospettiva, quanto più possibile, ad una vasta area geografica.

Nel tempo, si sono susseguiti numerosi trattati ed in questa sezione, senza la presunzione dell'eshaustività, si tenta di fornire i principali progressi sul piano giuridico e legislativo che hanno introdotto riconoscimenti e aspetti sanzionatori al fenomeno della disuguaglianza di genere e nello specifico, della

violenza sulle donne. Il primo passo è la *Dichiarazione Universale dei Diritti Umani* (10 dicembre 1948), dalla quale proclamazione sono passati settantuno anni. Il testo redatto dagli Stati membri delle Nazioni Unite, attraverso gli artt. 1, 2 e 3 dichiara il concetto di libertà ed uguaglianza riconosciuto a tutti gli individui senza alcuna discriminazione di genere.

Articolo 1

“Tutti gli esseri umani nascono liberi ed eguali in dignità e diritti”.

Articolo 2

“Ad ogni individuo spettano tutti i diritti e tutte le libertà enunciate nella presente Dichiarazione, senza distinzione alcuna, per ragioni di razza, di colore, di sesso, di lingua, di religione, di opinione politica o di altro genere, di origine nazionale o sociale, di ricchezza, di nascita o di altra condizione. Nessuna distinzione sarà inoltre stabilita sulla base dello statuto politico, giuridico o internazionale del paese o del territorio cui una persona appartiene, sia indipendente, o sottoposto ad amministrazione fiduciaria o non autonomo, o soggetto a qualsiasi limitazione di sovranità”.

Articolo 3

“Ogni individuo ha diritto alla vita, alla libertà ed alla sicurezza della propria persona”.

Successivamente, trentuno anni dopo, i principi della Dichiarazione Universale dei Diritti Umani vengono incorporati nella realizzazione di un Trattato finalizzato a porre rimedio proprio alle questioni di discriminazione di genere: la *Convenzione delle Nazioni Unite sull'eliminazione di tutte le forme di discriminazione contro le donne* (CEDAW, 1979). Adottata dall'Assemblea generale delle Nazioni Unite il 18 dicembre 1979 ed entrata in vigore internazionale dal 3 settembre 1981, questa viene ratificata con tempi diversi nei vari paesi europei, rappresentando comunque il primo vero strumento europeo per il riconoscimento della parità tra i sessi, poiché in esso si afferma l'uguaglianza dei diritti, il principio della dignità e del valore umano. La sezione tratta dal testo integrale che definisce proprio la questione di genere è la seguente:

“Vista la “Dichiarazione Universale dei Diritti Umani”, nel principio d'inammissibilità della discriminazione, secondo cui gli esseri umani nascono

liberi ed uguali in dignità e diritti, sottolineando il principio “senza alcuna distinzione di genere”, l’atto complementare e di rinforzo è dato dalla “Convenzione contro ogni forma di discriminazione di genere”.

Il rafforzamento all’affermazione dell’universalità del diritto contro ogni forma di discriminazione avviene nel 1993, durante la Conferenza mondiale sui Diritti Umani tenutasi a Vienna, la quale introduce una tematica di interesse complementare e di necessità urgente: la questione della violenza di genere. La legge redatta nella risoluzione n. 48/104 del 20 dicembre 1993, conosciuta sotto il nome di *Dichiarazione sull’eliminazione della violenza contro le donne*, richiamando la conclusione del paragrafo 23 in allegato alla risoluzione del Consiglio Economico e Sociale n.15 del 24 maggio 1990 - in cui veniva riconosciuta la trasversale diffusione della violenza contro le donne e l’urgente necessità di eliminarne l’incidenza - ed in riferimento all’allegato n.18 del 30 maggio 1991 - nel quale si proponeva uno schema di strumento giuridico – fornisce una chiara definizione dei diritti applicabili in materia di violenza contro le donne.

“Ai fini della presente Dichiarazione l’espressione “violenza contro le donne” significa ogni atto di violenza fondato sul genere che abbia come risultato, o che possa probabilmente avere come risultato, un danno o una sofferenza fisica, sessuale o psicologica per le donne, incluse le minacce di tali atti, la coercizione o la privazione arbitraria della libertà, che avvenga nella vita pubblica o privata” (Dichiarazione sull’eliminazione della violenza contro le donne”, 1993).

Si riconosce quindi in questa fase non solo la violazione dei Diritti Umani nella disuguaglianza tra i generi ma si dichiara solennemente la violenza contro le donne, definendola come privazione arbitraria delle libertà femminile e delineandone le tipologie di esercizio. Nella solenne dichiarazione si impone inoltre responsabilità allo Stato, chiedendo impegno costante da parte della comunità internazionale a contrastare il fenomeno. L’*articolo 3* dichiara infatti:

“Le donne hanno il diritto ad un uguale godimento e garanzia di tutti i diritti umani e libertà fondamentali in campo politico, economico, sociale, culturale, civile o in ogni altro campo, includendo: il diritto alla vita; il diritto all’uguaglianza; il diritto alla libertà e alla sicurezza della persona; il diritto ad una uguale protezione di fronte alla legge; il diritto di essere libere da tutte le forme di discriminazione; il diritto al più

alto standard raggiungibile di salute fisica e mentale; il diritto a condizioni di lavoro giuste e favorevoli; il diritto a non essere sottoposte a tortura o ad altri trattamenti o punizioni”.

Nel 1995 viene adottata invece la cosiddetta *Beijing Platform for Action* che ha ottenuto il consenso di centoottantanove paesi. La piattaforma identifica le aree che richiedono maggiore intervento in termini di parità di genere e di violenza sulle donne. Il progetto legislativo nel 2005 è stato riesaminato ponendo una lente di ingrandimento sulle diverse aree geografiche. La revisione a cadenza quinquennale, mette i governi nella condizione di mettere in atto misure adeguate a contrastare il fenomeno della violenza sulle donne.

Sei anni dopo, dalla risoluzione del 1993 attraverso l'atto normativo n.54/134 del 17 dicembre 1999, l'*Assemblea Generale* ha designato una ricorrenza annuale al fine di ribadire l'abolizione di ogni forma di violenza di genere. Questa giornata designata il 25 Novembre è stata battezzata come *Giornata internazionale per l'eliminazione della violenza contro le donne*. La data prescelta vuole ricordare il brutale assassinio delle sorelle *Mirabal* avvenuto nel 1960 nella Repubblica Dominicana sotto il regime dittatoriale di *Trujillo*. Annualmente, la *Giornata internazionale per l'eliminazione della violenza contro le donne* segna l'inizio dei *16 Giorni di Attivismo Contro la Violenza di Genere*; ovvero si concretizzano diverse iniziative da parte delle organizzazioni dei diritti umani come il “Centro per la *leadership* globale delle donne”, l'*UNIFEM*, *Women Won't Wait*, *Donne per il cambiamento*, *Soccorso delle donne*.

Dal 1999, Il Consiglio dell'Unione Europea adotta ogni anno un'analisi degli indicatori di riferimento, assicurando un *follow-up* mirato e strutturato.

Tuttavia, nel corso del tempo, si cerca di intensificare gli sforzi, poiché il fenomeno risulta essere particolarmente in attivo, così nel 2006, mediante la risoluzione 61/143, l'*Assemblea Generale* ha riaffermato tutti gli impegni internazionali che i vari Stati membri avevano sottoscritto, incluso l'obbligo di promuovere e proteggere tutti i diritti umani e le libertà fondamentali delle donne con la dovuta diligenza.

Tra gli altri aspetti della violenza contro le donne, il Consiglio di Sicurezza delle Nazioni Unite ha voluto porre l'accento sul tema della violenza in situazioni di conflitto armato per mezzo di due specifiche risoluzioni: n.1325

(2000) e n.1820 (2008). I testi ricordano che “lo stupro e altre forme di violenza sessuale possono costituire un crimine di guerra, un crimine contro l’umanità o un atto costitutivo rispetto al genocidio e sottolinea la necessità di l’esclusione dei reati di violenza sessuale da disposizioni di amnistia nel contesto dei processi di risoluzione dei conflitti [...]”.

A completare il percorso politico-legislativo e normativo con l’intento di contrastare la violenza di genere è la *Convenzione del Consiglio d’Europa sulla prevenzione e la lotta alla violenza contro le donne e la violenza domestica* passata alla storia come la *Convenzione di Istanbul*, approvata dal Comitato dei Ministri del Consiglio d’Europa nell’aprile 2011. Questo atto normativo è forse uno dei più importanti, poiché oltre a definire la violenza nelle sue forme più disparate e a specificarne gli ambiti, include al proprio interno l’abuso parentale, le violenze di natura sessuale e lo *stalking*, definendo una disciplina che dichiara ogni atto di natura psicologica, fisica, sessuale o economica come reato. Tale Convenzione rappresenta il primo strumento internazionale giuridicamente vincolante che crea un quadro normativo completo al fine di proteggere le donne da qualsiasi forma di violenza (*Council Of Europe, 2011*).

Gli Stati membri devono infatti includere tali sezioni nei loro codici penali o in altre forme di legislazioni previste. I reati dichiarati sono essenzialmente la violenza psicologica (articolo 33); gli atti persecutori - *stalking* (art.34); la violenza fisica (art.35), la violenza sessuale, compreso lo stupro (art.36); il matrimonio forzato (art. 37); le mutilazioni genitali femminili (art.38), l’aborto forzato e la sterilizzazione forzata (art.39); le molestie sessuali (articolo 40) e crimini commessi in nome del cosiddetto *delitto d’onore* (art.42).

L’adozione e la ratifica nei diversi Stati membri è piuttosto lunga e variabile da paese a paese. Forse rappresenta proprio questo l’aspetto più problematico, il fatto che molti paesi aderenti non abbiano ancora introdotto e quindi adottato misure e norme che si allinino a tale Trattato. Il *focus* è evidenziare la violenza domestica, la protezione delle vittime e il perseguimento giudiziario degli autori degli atti violenti.

Cionostante, nel corso degli anni, i diversi Stati membri hanno proposto normative relative a casi specifici nel proprio ordinamento giuridico: in Spagna, ad esempio il reato di *stalking* è stato introdotto nel Codice di procedura penale nel 1989 e nel 2004 hanno istituito tribunali ad *hoc* per le violenze che vengono commesse in ambito parentale. In Francia invece attraverso una legge del 2010 vengono stabilite pene più severe, fino all’arresto, per la violenza domestica

riconoscendo non solo la violenza fisica ma anche quella psicologica. In Austria ad esempio, la prima legge sulla violenza parentale è datata 1996 ed è stata più volte modificata nel corso degli anni. L'ordinanza in materia prevede che il soggetto violento debba essere immediatamente allontanato dalla casa comune, disinnescando così il suo potenziale distruttivo. In Germania, dal 2002 vige una legge che allontana gli autori della violenza dal domicilio, proibendo ogni tipo di contatto con la vittima; e solo dal 2008 esiste una normativa in materia di *stalking*. Mentre, in Gran Bretagna lo *stalking* è stato dichiarato reato già nel 1997. Le pene arrivano fino a 6 mesi di detenzione, con un incremento in caso di reato reiterato. Nel caso in cui sussista un'ordinanza restrittiva, se questa viene violata è prevista la reclusione fino a 5 anni.

In Italia, nel 1996 viene introdotto il reato di violenza sessuale attraverso la legge n. 66, nella quale si afferma il principio per cui la coercizione sessuale diventa un crimine contro la persona e non contro la morale pubblica ed il buon costume, come era stato fino ad allora. Accolta dalla comunità internazionale la legge n.66 dello Stato italiano diviene l'anticipazione della risoluzione n.1820 del 19 giugno 2008, attraverso la quale le Nazioni Unite hanno inserito la violenza sessuale fra le armi da guerra. Solo nel 2009, si parla però di “misure urgenti in materia di sicurezza pubblica e di contrasto alla violenza sessuale, estendendo il tema della violenza agli atti persecutori come lo *stalking*”.

Di recente, l'impegno politico ed istituzionale si è contraddistinto per l'assidua necessità di abbattere le disuguaglianze di genere a più livelli. Nei vari paesi europei vi è stata infatti una concreta attivazione di norme che prevedono importanti cambiamenti strutturali. In questo modo, il quadro legislativo di ogni paese tenta di partire dall'affermazione dell'uguaglianza di genere come pietra miliare per contrastare ogni forma di violenza sulle donne.

Al fine di consentire alle donne di conciliare la vita professionale con quella coniugale e genitoriale, nell'aprile 2017, la Commissione europea ha adottato un pacchetto completo di misure politiche e giuridiche. In particolare, la direttiva mira a modernizzare la legislazione dell'UE in materia di congedo familiare e modalità di lavoro praticabile, cercando di creare un maggiore equilibrio tra la vita lavorativa e quella privata delle donne. Tuttavia, a livello europeo si è pensato di proporre anche il congedo di paternità e nel 2016, in Austria, il Parlamento ha proprio adottato un'iniziativa volta a sostenere un rapporto di equa assistenza all'infanzia, concedendo un *bonus di congedo*

parentale chiamato “*Family Time Bonus Act*”. Nel settore privato è entrata in vigore l’opzione del cosiddetto “mese del papà”.

Per quanto riguarda invece il tanto dibattuto problema del *gap* salariale, nell’aprile 2017, la Polonia ha pubblicato un’applicazione gratuita *on-line* che permette a qualsiasi utente di confrontare il proprio salario, nonostante si impegni a fornire retribuzioni eque senza distinzione di genere. Tale *app* misura il *gap* retributivo tra uomini e donne, tenendo ovviamente conto di diverse variabili, come ad esempio le caratteristiche del luogo di lavoro, le peculiarità dei dipendenti e dei datori di lavoro. Può essere utilizzato sia dalle aziende private che dalla pubblica amministrazione.

A questo proposito, nel novembre 2017, la Commissione europea ha adottato una strategia di valutazione per la trasparenza delle retribuzioni (COM, 2017) e nello stesso anno, in Germania è entrata in vigore la legge per promuovere la trasparenza delle strutture salariali (*Entgelttransparenzgesetz*). Il Regno Unito, ha inoltre sancito l’obbligo della rendicontazione sul divario retributivo di genere ed ha imposto alle aziende con oltre duecentocinquanta dipendenti di pubblicare i dati relativi ai *gap* salariali entro aprile 2018.

L’assetto normativo cerca di ridurre inoltre le disparità di genere anche in termini di *leadership* o rappresentanza politica femminile.

In Belgio ad esempio, la rappresentanza paritaria di donne e uomini nelle liste elettorali è prevista per legge dall’anno 2002 e si estende a tutti i livelli di elezione. Anche Bruxelles-Capitale ha applicato il medesimo obbligo per le elezioni municipali del 2018, con l’obbligo di collocare alternativamente donne e uomini nelle liste elettorali. La Francia ha invece introdotto un sistema di candidati bi-nominali per le elezioni dipartimentali, ovvero non si vota più per un solo candidato ma per una squadra di candidati donne e uomini. Ciò garantisce la parità di genere a livello di dipartimento ma stabilisce anche posizioni decisionali condivise nelle politiche territoriali, introducendo il potenziale delle responsabilità condivise e migliorando le possibilità di conciliazione della vita lavorativa per i politici.

La Slovenia ha imposto quote di genere per le elezioni sia locali che europee.

Per quanto riguarda la *leadership* nelle società quotate in borsa invece, le politiche europee risultano particolarmente attive.

Nel luglio 2016 infatti l’Austria ha adottato una legge chiamata “Legge sull’uguaglianza per donne e uomini come amministratori non esecutivi nei consigli di amministrazione”, entrata in vigore il primo gennaio 2018.

Questa normativa prevede che i Consigli d'amministrazione (CdA) delle società quotate in borsa e le aziende con oltre una migliaia di dipendenti possano includere una quota del 30% paritaria tra uomini e donne.

In Portogallo, la legge n. 62/2017 stabilisce le norme per una rappresentanza equilibrata tra donne e uomini negli organi di amministrazione e di vigilanza, sia nel settore pubblico che nelle società quotate in borsa. La legge prevede che dal 2018 le società statali e le società quotate garantiscano una quota minima di donne. Per le società quotate, il minimo delle posizioni previste per il 2018 nei CdA è del 20% ma si prevede un incremento fino ad una quota del 33.3% nell'anno 2020. Mentre, nelle aziende di proprietà statale, almeno il 33.3% delle posizioni nei consigli di amministrazione e negli organi di controllo deve essere svolto da professionisti di sesso femminile.

1.8. Le risposte sul piano sociale

I temi della disuguaglianza di genere e della violenza contro le donne sono stati ormai incorporati nelle strutture sociali, allo stesso modo in cui la cultura patriarcale ha tramandato ed inscritto le differenze di genere, gli attributi e i ruoli sociali che ne derivano. Gli effetti più nascosti del dominio maschile che opera nell'oscurità dei corpi di un ordine sociale naturalizzato sono ormai venuti alla luce e si contrappongono in un continuo bilanciamento tra permanenza e cambiamento. I movimenti femministi, le politiche di conciliazione e le trasformazioni normative hanno contribuito a sottrarre alla storia, un dominio maschile che ha per troppi secoli tentato di assicurare la propria permanenza attraverso Stato, Chiesa, agenti di socializzazione, *etc.* Il patriarcato pubblico ha dunque subito una forte battuta d'arresto, da quando le istituzioni non identificano più la visione ultraconservatrice che fa del patriarcato l'ordine morale pubblico. Di conseguenza, il lavoro critico è riuscito a spezzare anche sul piano sociale quel circolo chiuso di rafforzamento patriarcale generalizzato che da sempre ha caratterizzato la società. I fattori di conversione hanno avuto il potere moltiplicativo e decisivo di modificare nel profondo le strutture cognitive, le tassonomie pratiche, i giudizi etici, estetici portando ad una risposta sociale e collettiva del fenomeno della disparità di genere e soprattutto della violenza contro le donne.

Socialmente condiviso a tal proposito è il cosiddetto fenomeno della *strage di donne*, sensazionalisticamente riproposto dal linguaggio giornalistico. La violenza sulle donne vuole essere una questione di dominio pubblico e non più un fatto privato. Sono infatti innumerevoli le iniziative e le politiche di sostegno che cercano di educare alla *non violenza* e che mirano a definire soprattutto in ambito domestico le dinamiche, spesso contorte, delle relazioni cosiddette tossiche. Le campagne di sensibilizzazione sono finalizzate a far comprendere alle donne l'*a-normalità* della violenza, nella speranza di promuovere l'azione di denuncia e trasmettendo loro il messaggio di non essere sole ma di poter contare su strutture specializzate e dedicate alla protezione e alla tutela delle donne. In tal senso, è ormai nota in Europa l'istituzione di un numero di emergenza (*help line*) dedicato alle donne, il quale servizio si avvale di professionisti in ambito socio-sanitario e legale, in contatto con centri antiviolenza, conosciuti anche come case rifugio o case d'accoglienza. Tali strutture il primo tassello di un mosaico che ha atteso per decenni che il fenomeno della violenza sulle divenisse un problema sociale, la cui matrice ideologica si può ricondurre alle lotte femministe. Si tratta essenzialmente di luoghi in cui vengono ospitate le vittime di violenza, la cui architettura di base è l'assistenza clinica psicologica e fisica attraverso operatori di settore. In Europa è stata istituita la rete *Wave Women Against Violence Europe*, mentre a livello mondiale nel 2008 in Canada è nata la rete *Global Network of Women's*. Tuttavia, oggi, solo quindici su quarantasei Stati europei hanno istituito un numero d'emergenza e mancano oltre quarantasette mila centri antiviolenza. Le politiche di sostegno hanno proposto diverse iniziative di sensibilizzazione e prevenzione, istruendo le donne ma anche le nuove generazioni, cercando di educare alla cultura moltiplicativa dell'uguaglianza di genere. Sono spesso coinvolti individui di sesso maschile, nella speranza che il passaggio dell'informazione che proviene da una persona dello stesso sesso, possa contribuire maggiormente alla riflessione ed al cambiamento. Le attività a sostegno si impegnano a fornire una maggiore preparazione sul tema, a definire protocolli specifici da seguire e a preparare percorsi di formazione per intervenire soprattutto sul piano psicologico con la vittima di violenza. Le istituzioni che necessitano di un tale supporto sono quelle di natura medico-sanitaria, legale, la magistratura e gli operatori delle forze dell'ordine. Sul piano della comunicazione sono state poi realizzate innumerevoli iniziative, soprattutto in termini di pubblicità sociale o *pubblicità progresso* che ha realizzato numerosissime campagne di sensibilizzazione.

Negli ultimi anni, diversi movimenti nati sul *Web* sono diventati virali, come ad esempio, nel 2017, il movimento *#MeToo* è andato affermandosi in modo virale su scala globale. La finalità di partenza consisteva nel denunciare i casi di violenza sessuale o di molestie subite dalle donne ai fini di *far carriera*. Questo movimento ha sottolineato l'importanza degli sforzi della Commissione europea per dare voce e visibilità alle donne che subiscono violenza.

Nello stesso anno, è stata anche lanciata sui social media la campagna "*NON.NO.NEIN campaign — Say NO! Stop violence against women*" che ha aumentato la consapevolezza e finanziato progetti concreti per contrastare la violenza di genere.

Tra le altre iniziative governative e non, a febbraio 2018, ad esempio, il Consiglio della Presidenza maltese ha ospitato una conferenza incentrata sulla violenza di genere nel contesto della discriminazione intersezionale, quale la segregazione di genere in più settori. Mentre, durante il *Mutual Learning Programme in gender equality e The Annual Colloquium on Fundamental Rights*, la Commissione europea si è focalizzata sulla questione della violenza di genere approfittando della presenza dei ministri di circa la metà degli Stati membri, delle organizzazioni internazionali, degli accademici di spicco e di attori facenti parti della società civile, giornalistica e mediatica.

In Italia, la comunicazione sociale è attiva come *pubblicità progresso* e cerca di porre l'accento sulla questione della violenza domestica, focalizzando l'attenzione sui legami amorosi patologici. La rete Rai, nel 2012 ha voluto occuparsi della tematica mediante un ciclo televisivo articolato in quattro diverse storie di donne vittime di violenza. La serie, dal titolo *Mai per amore* regia di Liliana Cavani, Marco Pontecorvo e Margareth von Trotta ha evidenziato gli scenari più comuni del fenomeno, cercando di ricostruire le dinamiche ed i profili dei soggetti maschili archetipo della violenza di genere. La pellicola ha ottenuto complessivamente il 67.76% di *share* con una visibilità di 18 milione 963 mila telespettatori. Altra importante iniziativa Rai è il *format Amore criminale* ideato da Matilde D'Errico, Maurizio Iannelli e Luciano Palmerino. La trasmissione televisiva, dal 2007, in ogni puntata racconta una storia d'amore finita con un omicidio o un tentato omicidio. Le vicende vengono ripercorse attraverso ricostruzioni filmate e materiale documentaristico.

1.9. Il fenomeno oggetto di studio attraverso la visione universalistica dei Diritti Umani

Il fenomeno della violenza sulle donne può essere quindi inquadrato attraverso una prospettiva di analisi che si basa sulla cultura di genere, sulla strutturazione di attributi, ruoli e comportamenti al genere connessi, sul dominio maschile e sulla virilità. Tuttavia, esulando dalle proprietà biologiche, psicologiche e sociali che sono iscritte nell'ambiente di riferimento, si può pensare di trattare il tema oggetto di studio attraverso una visione universalistica dei diritti umani. Infatti, la violenza sulle donne è riconosciuta proprio come una violazione dei diritti umani nella sua effrazione a più livelli, ed in questa sezione della presente tesi si vuole nello specifico fare riferimento alla letteratura che si occupa di definire e trattare i principi di tutela delle capacità umane di Martha Nussbaum (2000). In *Women and Human Development* (Nussbaum, 2000) l'autrice analizza e confronta la condizione delle donne che vivono in India seguendo tre specifici indirizzi di ricerca: uno filosofico-politico, un altro giuridico-comparativo ed infine quello narrativo. La consapevole separatezza che caratterizza le vite delle persone conduce la Nussbaum a determinare gli elementi che qualificano la qualità di una vita degna di essere vissuta, non assumendo come parametro il reddito pro-capite ma ciò che gli individui hanno realmente bisogno di essere o fare in ambiti di vita ritenuti essenziali. La Nussbaum identifica un universalismo dei diritti umani nelle diverse situazioni culturali, dove le *capacità umane* rappresentano un patrimonio di riconoscimenti e potenzialità che appartengono ad ogni individuo; le quali *capacità* indicano un *oggetto di cura* da parte della comunità umana e delle organizzazioni internazionali. La visione dell'autrice identifica le *principali capacità umane* volte a garantire la soglia minima per la dignità della vita umana. La Nussbaum determina dei *fixed provisional points* come frutto di un ampio consenso interculturale. I principi delineati possono essere introdotti nelle garanzie costituzionali dei vari paesi e rappresentano le capacità essenziali per il funzionamento umano. L'elenco di consenso condiviso proposto dall'autrice riguardano diverse aree:

- *La Vita*: intesa nella sua piena dignità ed integrità fisica e morale, la quale non deve essere ridotta in condizioni di indecenza estrema tale per cui si comporti la morte prematura.

- *La salute fisica*: le condizioni di salute garantite includono la possibilità di compiere tutte le funzioni umane come nutrirsi, avere un'abitazione adeguata, rispettare le condizioni igienico-sanitarie necessarie. Rientra in questa categoria anche la facoltà riproduttiva.
- *L'integrità fisica*: che identifica la protezione da eventuali aggressioni inclusi la violenza sessuale, l'abuso sui minori e la violenza domestica. Afferiscono a questa classe di appartenenza le condizioni di libertà di qualsiasi sovranità psico-corporea.
- *Sensi, immaginazione e pensiero*: questa garanzia permette di godere delle condizioni per poter espletare le capacità intellettive e di condividere attivamente nel sociale il proprio pensiero.
- *Emozioni*: vi afferiscono lo sviluppo emotivo e l'uso libero della ragione. Rientrano in questa categoria l'opportunità di poter costruire relazioni emotive stabili senza che eventi traumatici influenzino il modo di rapportarci agli altri.
- *Ragion pratica*: identifica l'uso libero della ragione nella concretizzazione di un pensiero critico e cosciente.
- *Unione*: la libertà ad associarsi e di appartenere a gruppi senza subire discriminazioni di alcun tipo rientrano in questa classe delle capacità umane.
- *Altre specie*: si riconosce in tal senso il valore delle affezioni per il mondo vegetale e quello animale.
- *Gioco*: l'opportunità di vivere esperienze ludico-ricreative.
- *Avere controllo sul proprio ambiente*: si tratta della concreta disponibilità degli spazi politici e materiali compreso il posto di lavoro. Vi rientrano l'opportunità di godere di uguali diritti senza violazioni fisiche o morali in nome di una qualche forma di discriminazione.

La visione dell'autrice esula da ogni tipo di discriminazione o individualismo a cui si possa essere soggetti, poiché riesce con acuta sensibilità a definire le reali ed essenziali esigenze di ogni essere umano. La base su cui prende forma la sua visione universalistica è la dignità della Vita. I principi della Vita si connettono al principio della violenza contro le donne, poiché spesso il fenomeno indagato priva le vittime delle condizioni essenziali per il proprio funzionamento umano. Vi sono parti del Pianeta nei quali esiste una profonda esasperazione delle privazioni materiali e morali e che rendono la vita delle donne indegna di essere vissuta. Il concetto di indegno non si riferisce alle donne in quanto tale ma

aberranti per un essere umano. Si può osservare come l'autrice non abbia di certo trascurato gli aspetti psicologici, emotivi e interazionali. Ha introdotto nelle Capacità Umane Fondamentali aspetti prettamente connessi alle conseguenze della violenza psicologica che possono compromettere definitivamente la percezione di sé, l'autostima, il livello di sicurezza e di serenità emotiva in un contesto sia individuale che collettivo. Se ognuna delle capacità fondamentali delineate dall'autrice venisse rispettata, la violenza contro le donne non esisterebbe, poiché le forme attraverso cui si concretizza la violenza di genere violano a pieno tutti i punti universalistici proposti dalla Nussbaum. Non ci sono dubbi nell'affermare che la violenza contro le donne rispecchi a pieno una violazione dei Diritti Umani.

1.10. Le statistiche ufficiali sulla violenza contro le donne

Con lo scopo di fornire una stima rappresentativa di una delle principali violazioni dei diritti umani e definire quindi lo scenario quantitativo della violenza sulle donne, in questo contesto, si prende come punto di riferimento una delle ultime ricerche condotte in Europa da parte dell'*European Institute of Gender Equality (EIGE)*. Il rapporto pubblicato nel 2017 si basa su dati relativi al 2015 ed è frutto di un attento lavoro di *data integration*; i dati provengono da fonti *Eurostat* (2014), interviste del *FRA* (2014) e atti amministrativi dello stesso ente che realizza il rapporto. L'*EIGE* consente di indagare i vari domini dell'uguaglianza di genere e tra gli ambiti aggiuntivi della ricerca considera per l'appunto il fenomeno della violenza contro le donne in Europa. In realtà, si include nell'analisi dell'uguaglianza di genere il dominio cosiddetto satellite della violenza contro le donne. Tale strutturazione considera la violenza come il corollario delle disuguaglianze che definiscono sul piano funzionale e strutturale le differenze di genere agite nei vari ambiti, quali lavoro, salute, denaro, potere, istruzione e tempo. La prospettiva della violenza apporta un contributo ai settori chiave per il calcolo dell'indice sull'uguaglianza di genere (EIGE, 2017a; 2017b). L'ente di riferimento infatti considera la macro-categoria della violenza sulla base di variabili micro-categoriche. Nello specifico, tale dominio viene calcolato mediante una misura composita costituita da tre sotto-domini: *prevalenza, severità e divulgazione*.

- La *prevalenza* identifica i casi di violenza in Europa e stima la percentuale di donne che hanno subito violenza fisica, sessuale, psicologica o economica, compresa la morte.
- La *severità* misura le conseguenze della violenza, considerando anche percorsi di vittimizzazione multipla.
- La *divulgazione* si basa sull'azione di denuncia vera e propria.

È stata definita una struttura di misurazione a tre livelli che comprende una serie di indicatori, poiché i tre principali *sotto-domini* sono basati, a loro volta, su una serie di ulteriori sotto-indicatori che coprono una gamma ampia delle varie forme di violenza contro le donne ed una serie di fattori contestuali che possono fornire informazioni su alcune delle cause e delle circostanze che circondano la violenza femicida. Gli indicatori identificati per la misura composita sono aggregati allo scopo di ottenere un singolo punteggio medio per ciascun paese, così da evidenziarne l'incidenza. Si tratta di un punteggio unico in grado di monitorare l'estensione delle forme più comuni e ampiamente criminalizzate della violenza sulle donne. Il *global index* si muove in un intervallo che va da 1 a 100, la cui logica interpretativa è 1, quale situazione di violenza scarsa o inesistente e 100, quale condizione di violenza estremamente diffusa.

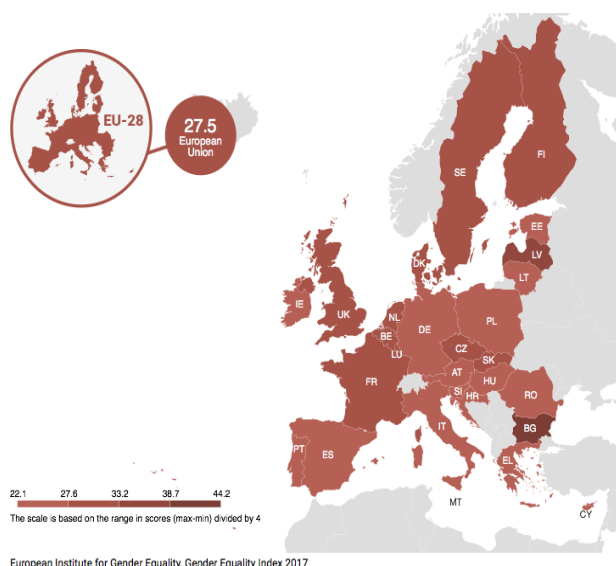


Figura 1.4. Gender Equality Index, Il dominio della violenza sulle donne nei paesi membri dell'UE per il 2015. Fonte: EIGE, 2017a

Secondo le indagini dell'EIGE (2017a) il dominio della violenza nei paesi membri dell'UE mostra un valore di indicatori aggregati pari a 27.5 su 100. Questo punteggio medio risente tuttavia della presenza di eventuali valori anomali, ossia valori particolarmente bassi o alti da parte di uno o più paesi membri. A tal proposito, si osservano infatti i valori medi composti di ogni singolo paese europeo per l'anno 2015. I dati permettono di osservare il valore più alto in Bulgaria (44.2) mentre il valore più basso in Polonia (22.1). Se si considerano le elaborazioni dell'*Eurostat* (2012-2014 in EIGE 2017a) utilizzate nei paragrafi precedenti per indicare i soli femminicidi (intesi in questo caso come omicidi), i dati indicano invece ai primi posti Italia, Germania e Francia. Pertanto, esiste una discrepanza tra i risultati dell'*Eurostat* e quelli della misura composita dell'*EIGE* che permette di osservare valori maggiori in Bulgaria, Lituania e Danimarca. Si sostiene infatti che una tale mancanza di correlazione tra i dati *Eurostat* e quelli *EIGE* derivi da un modo diverso di raccogliere i dati. Per studiare il femminicidio come un caso di omicidio intenzionale bisogna valutare il sesso della vittima e la relazione esistente tra quest'ultima e l'autore. Questo tipo di accertamenti competono agli organi di polizia e secondo un recente rapporto pubblicato da *EIGE*, gli organi di polizia in Danimarca, Grecia, Lussemburgo, Lituania, Malta e Polonia non raccolgono questa tipologia informativa nei casi di omicidio. Mentre, paesi come la Francia ed il Regno Unito rendono maggiormente disponibili i dati, riportando il movente, le armi usate e le circostanze generali dell'atto delittuoso. Proprio per l'insufficienza di tali indicatori quantitativi comparabili il femminicidio è stato escluso dalle analisi *EIGE* successive.

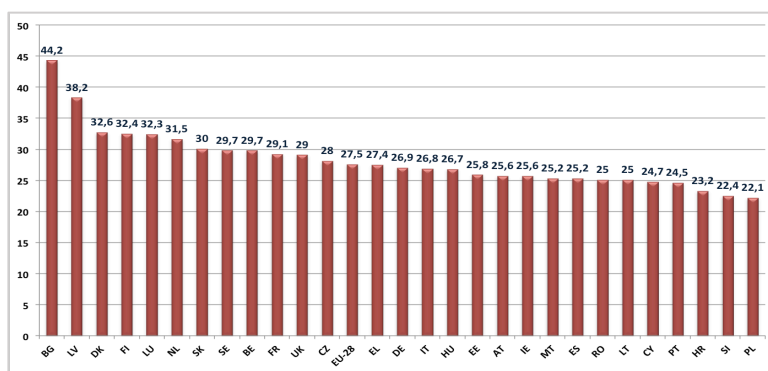


Figura 1.5. Gender Equality Index, punteggio europeo nell'ambito della violenza sulle donne in Europa. Fonte: EIGE, 2017a; 2017b.

Il primo indicatore che concorre a definire la misura di sintesi della violenza è la *prevalenza*, attraverso il quale si ottiene una misura dei casi effettivi. Si tenga conto che questa dimensione ingloba qualsiasi tipo di violenza fisica, sessuale, psicologica, persecutoria, di *mobbing*, di tratta umana, *stalking*, fino all'omicidio. La *prevalenza* mostra un *range* che varia da un valore di 73.6 in Danimarca a 48.9 in Lettonia (Figura 1.6). Comunque considerevole è il valore di *prevalenza* che si evince nei paesi che occupano i primi posti nelle statistiche dell'*Eurostat*: Italia, 61.4; Francia, 66.6 e Germania, 52.9.

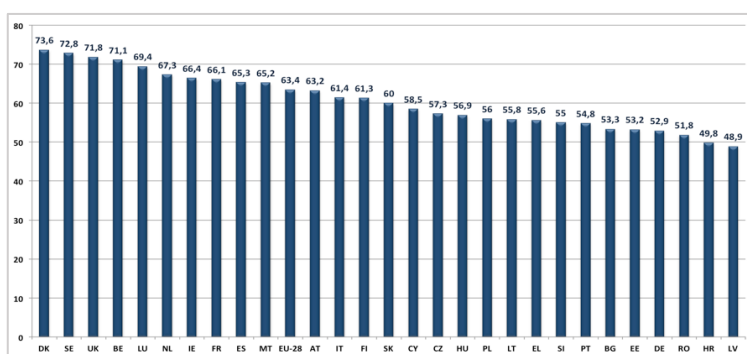


Figura 1.6. Gender Equality Index, punteggio europeo della prevalenza nell'ambito della violenza sulle donne in Europa. Fonte: EIGE, 2017a; 2017b.

La seconda dimensione stima la *severità* (Figura 1.17) e valuta quindi la gravità delle conseguenze per ogni episodio di violenza, includendo i casi di vittimizzazione multipla. Più i valori sono alti e più è grave l'entità.

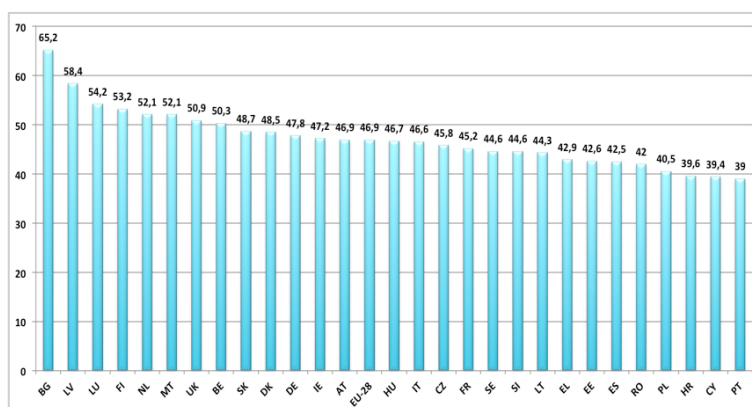


Figura 1.7. Gender Equality Index, punteggio europeo della severità nell'ambito della violenza sulle donne in Europa. Fonte: EIGE, 2017a; 2017b.

Per la *severità*, dati mostrano un intervallo di valori compreso tra 65.2 a 39; il paese con il valore più alto è la Bulgaria mentre una minore incidenza di severità la registra il Portogallo. La Danimarca che nel caso della *prevalenza* si mostra al secondo posto, se si considera invece la gravità degli atti violenti non si posiziona ai primi posti, sebbene registri un valore di *severità* della violenza pari a 48.5.

L'ultima dimensione è la *divulgazione* (Figura 1.18), ossia la percentuale della propensione a denunciare la violenza. Il valore medio europeo è pari a 14.3 e la Bulgaria, anche in questo caso, risulta essere il paese con il valore più alto.

In questo contesto però, un valore di *divulgazione* elevato indica una maggiore propensione a ricorrere alle forze dell'ordine. In tal senso, non sorprende osservare la Bulgaria o la Lituania ai primi posti, considerando l'alto tasso registrato di *prevalenza* e *severità* della violenza.

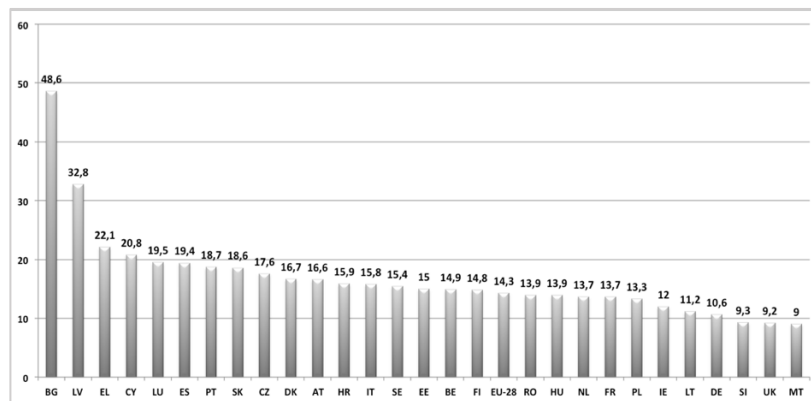


Figura 1.8. Gender Equality Index, punteggio europeo della divulgazione nell'ambito della violenza sulle donne in Europa. Fonte: EIGE, 2017a; 2017b.

Nell'ambito delle statistiche ufficiali tuttavia non si può osservare il fenomeno della violenza contro le donne senza mettere sotto una lente di ingrandimento al contempo, i diversi domini d'interazione che rafforzano il movente sostanziale alla determinazione funzionale e strutturale della società. Infatti, il tema della violenza rappresenta il corollario di un'indagine molto più ampia, finalizzata a studiare la disuguaglianza di genere.

1.11. La violenza simbolica attraverso i dati sull'uguaglianza di genere

Al fine di inquadrare da un punto di vista quantitativo l'asimmetria identitaria e relazionale tra i generi le statistiche ufficiali si configurano in dati che esprimono l'uguaglianza di genere in vari ambiti di riferimento. Nello specifico, si tratta di un bilanciamento di potere tra uomini e donne che non si è né affievolito e né sradicato dalle strutture sociali ma ha solo cambiato volto. L'*identità sessuata*, le strutture e i ruoli che ne derivano, continuano a preservare il passato e in molti contesti hanno preso forma ambito in nuovi domini, come il lavoro, la retribuzione, la formazione, il tempo libero, la salute ed il potere inteso in termini di *leadership*.

A confermare la tesi secondo cui l'uguaglianza tra i generi è ancora lontana vi sono indici deputati a misurare proprio tale disparità. In particolare, nel 2013, l'*EIGE* ha avviato la formalizzazione di un indice sull'uguaglianza di genere denominato *Gender Equality Index*, nel quale si considerano specifici ambiti, quali: lavoro, salario, cultura ed istruzione, tempo libero, *empowerment*, salute e percezione del sé. Rappresentano poi due domini aggiuntivi, quello dell'uguaglianza intersettoriale e quello della violenza, precedentemente trattato.



Figura 1.9. Domini e sotto-domini del Gender Equality Index. Fonte: EIGE, 2017a; 2017b.

Lo studio dell'*EIGE* assegna ad ogni Stato membro dell'UE un punteggio che va da 1 a 100, dove 1 indica un alto tasso di disuguaglianza e 100 una condizione di piena uguaglianza.

Gli otto domini considerati per costituire il *Gender Equality Index* sono a loro volta suddivisi in sotto-ambiti tematici. In particolare pertanto, l'indice è costituito da 31 indicatori, i quali vengono presentati in modo aggregato in un'unica misura di sintesi. L'edizione utilizzata in questo contesto è la medesima di quella impiegata per lo studio del dominio della violenza, al fine di una continuità nella struttura interpretativa dei dati. Il rapporto pubblicato nel 2017 fornisce un andamento evolutivo rispetto a quattro specifici anni: 2005, 2010, 2012 e 2015.

Per l'ultimo anno preso in esame, il *Gender Equality Index* mostra un punteggio medio di 66.2 nei 28 Paesi appartenenti all'UE (*Figura 1.11*).

Tuttavia, se il dato viene confrontato con quello del 2005 si può osservare un aumento di 4.2 punti in un arco temporale di 10 anni. Si tratta di un mutamento alquanto scarso, se si considera il periodo storico un decennio. Il valore riferito al singolo paese (*Figura 1.10*) consente di attribuire una maggiore parità di genere alla Svezia con un punteggio di 82.6, alla Danimarca con valore di 76.8 e alla Finlandia con valore di 73. Al contrario, la maggiore disparità si registra in Grecia il cui dato è pari a 50.

Una considerazione del tutto preliminare sulla misura globale permette di affermare che nessuno dei paesi appartenenti all'UE mostra un livello di disuguaglianza estremo in termini complessivi.

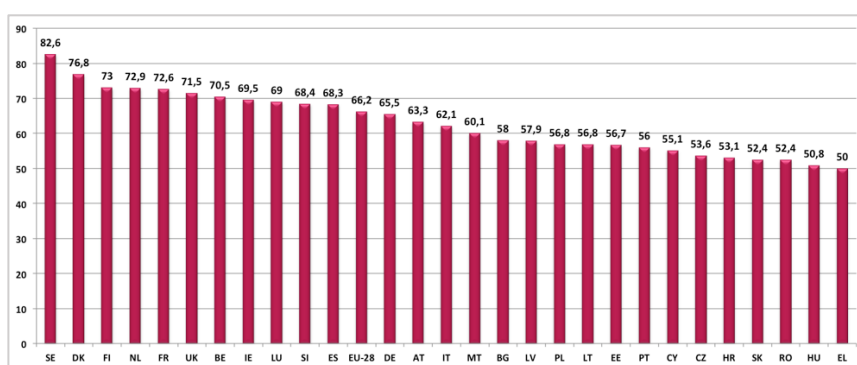


Figura 1.10. Gender Equality Index, punteggio europeo sulla parità di genere in Europa. Fonte: EIGE, 2017a; 2017b

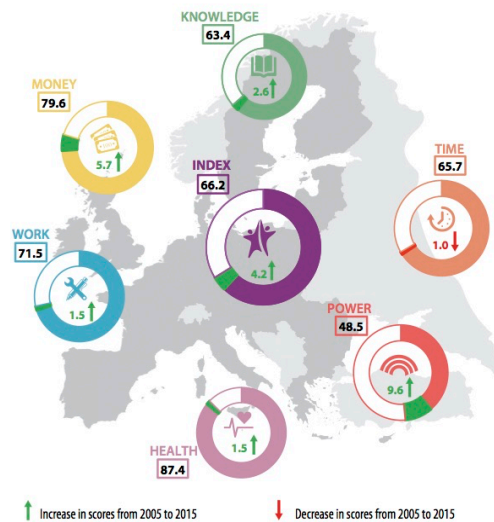


Figura 1.11. Gender Equality Index, punteggio degli Stati membri dell'UE per ciascun dominio. Fonte: EIGE, 2017a; 2017b

Il valore di sintesi se viene scomposto nei vari ambiti può supportare l'interpretazione dei risultati e a seconda della sfera di riferimento considerata può aiutare a valutare in quale dominio la donna stia acquisendo nel tempo un maggiore avanzamento.

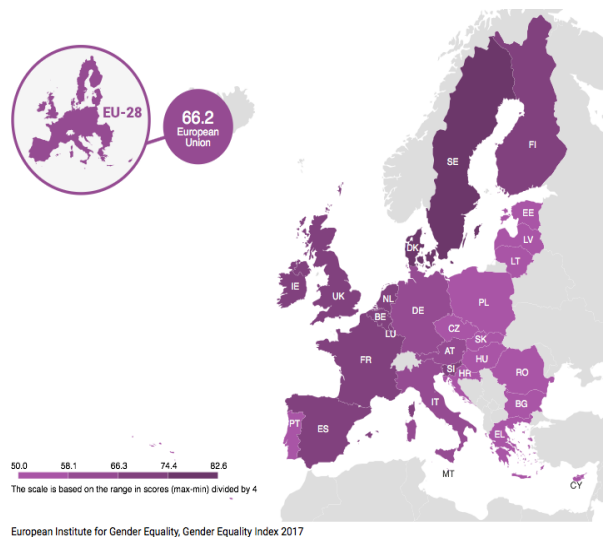


Figura 1.12. Gender Equality Index, punteggio europeo dei domini dal 2005 al 2015. Fonte: EIGE, 2017a; 2017b

È interessante notare come in Europa, dal 2005 al 2015, vi sia stato un

andamento in crescita per tutti i domini indagati, ad eccezione di uno: quello del tempo. Durante il decennio esaminato, le donne hanno subito una decrescita del fattore tempo di circa un punto e questo significa che hanno sicuramente minor tempo a disposizione per sé stesse mentre sempre più coinvolte nella vita familiare e professionale. Nel 2015, l'indice dell'*EIGE* mostra i valori più alti in riferimento al dominio della salute (87.4), della retribuzione (79.6) e del lavoro (71.5). Tuttavia, tali domini identificano al contempo i settori nei quali in questo decennio si sono registrati i più bassi incrementi. Pertanto, salute e lavoro rappresentano i fattori di riferimento più lenti nella scalata verso la parità di genere, con un incremento di appena 1.5 punti. Il dominio che invece sembra crescere con maggiore regolarità è quello dell'*empowerment*, sebbene l'indice proponga un punteggio di 48.5 punti. Quindi, se il valore incrementale è aumentato in maniera notevole ma la stima media è comunque bassa, questo significa che malgrado gli sforzi, la vetta dell'uguaglianza di genere in termini di *empowerment* è ancora piuttosto faticosa.

1.11.1 Il settore professionale: tasso occupazionale e gap retributivo

Se si considera l'ambito professionale si fa riferimento alla misura in cui donne e uomini possono beneficiare di pari opportunità in termini di accesso al mondo occupazionale, nelle medesime condizioni sul mercato del lavoro. Si tiene conto di tre variabili, quali *partecipazione al mercato del lavoro*, *segregazione di genere* e *qualità del lavoro*. L'UE si è prefissata di raggiungere un tasso di occupazione del 75% per uomini e donne entro il 2020 (Commissione Europea, 2010). Guardando all'occupazione femminile negli ultimi 7 anni si sono registrati alcuni incrementi piuttosto minimi, che nell'ultimo anno si discostano da quelli maschili per 11.5 punti di scarto.

I dati più confortanti a livello europeo sono rintracciabili a Malta e a Lussemburgo, con un aumento di rispettivamente 10.2 e 5.9 punti nel decennio tra il 2005-2015. L'occupazione femminile mostra un andamento in crescita costante ma lento.

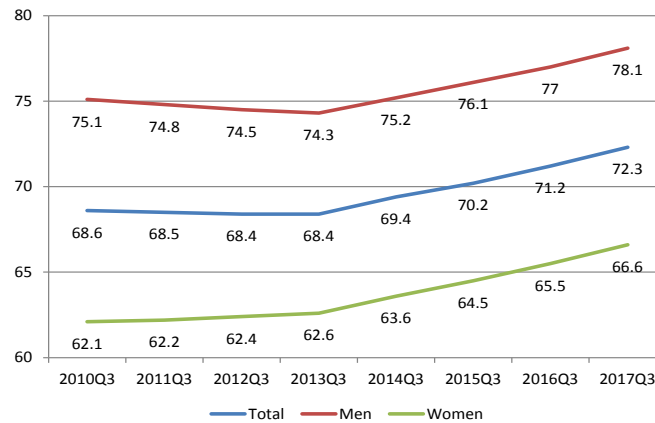


Figura 1.13. Serie Storica del tasso occupazionale per genere (fascia d'età: 20-64 anni). Fonte: Eurostat, 2017.

Il tasso occupazionale dell'Eurostat (2017) indica nel 2017 una percentuale di occupazione femminile del 66.6%. Tuttavia, bisogna tener presente che questo valore occupazionale vede le donne maggiormente impiegate nei settori con livello di retribuzione più basso (OECD, 2017) e con sei ore lavorative in più, in media, rispetto agli uomini (Eurofound, 2016). Le logiche di conformismo al tradizionale modello di divisione sessista non sono pertanto cambiate e le donne subiscono notevoli ritardi anche nell'avanzamento di carriera. In particolare, sul posto di lavoro è luogo comune che le donne siano sottoposte a certi stereotipi e le lavoratrici di sesso femminile devono farsi strada faticosamente per raggiungere gli stessi traguardi conseguiti invece dagli uomini. In ambito professionale, le donne hanno l'obbligo biologico di comprovare la propria le proprie competenze, dimostrando al contempo di saper gestire il carico emotivo, soprattutto in posizioni gerarchiche elevate. In aggiunta, il settore occupazionale femminile deve tener conto dello *status* "invalidante" che deriva dall'aspetto coniugale e/o genitoriale. Infatti, nel 2016, il tasso di occupazione medio di donne senza figli è superiore di 9 punti percentuali rispetto al tasso occupazionale di donne con a carico un figlio di età inferiore a 9 anni (Eurostat, 2017). Tale dato è indicativo di un sistema sessista che preclude alle donne l'opportunità di far carriera, nel tentativo di conciliare la maternità. Sussiste inoltre, un ulteriore elemento strettamente connesso all'occupazione femminile, quale il *gap* o divario salariale che in Europa raggiunge l'11.6 punti (Eurostat, 2017; Figura 1.14).

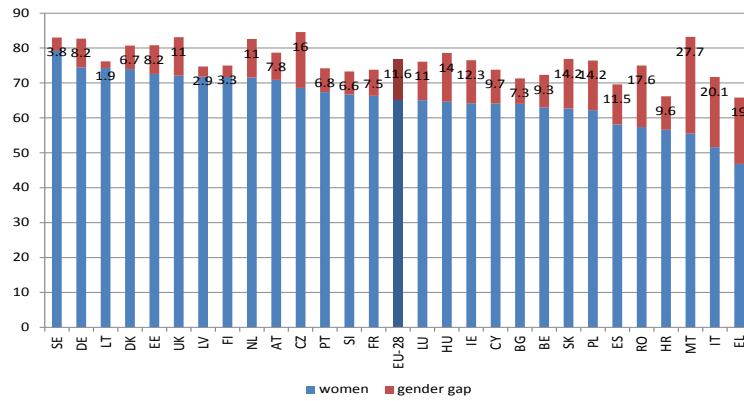


Figura 1.14. Tasso di occupazione femminile e gap occupazione di genere (fascia d'età: 20-64 anni) per il 2016 in Europa. Fonte: Eurostat, 2017.

Il dominio del lavoro come si è detto è costituito da due sotto-domini: in *primis* la partecipazione, la cui dimensione tiene conto di indicatori come il tasso occupazionale di tipo *full-time equivalent* (FTE) e la durata della vita lavorativa; in secondo luogo fa riferimento alla segregazione, intesa come partecipazione delle donne e degli uomini a settori specifici (l'istruzione, la salute umana, attività sociali (EHW)). In questo ambito, infine si tiene conto della qualità del lavoro stesso, in termini di flessibilità oraria e prospettive di carriera, in considerazione di una prospettiva di maternità. Il dominio lavoro registra il primato nei confronti della Svezia (82.6), Finlandia (74.7) ed Estonia (72.1) mentre classifica in modo poco vittorioso l'Italia (62.4) e la Grecia (64.2).

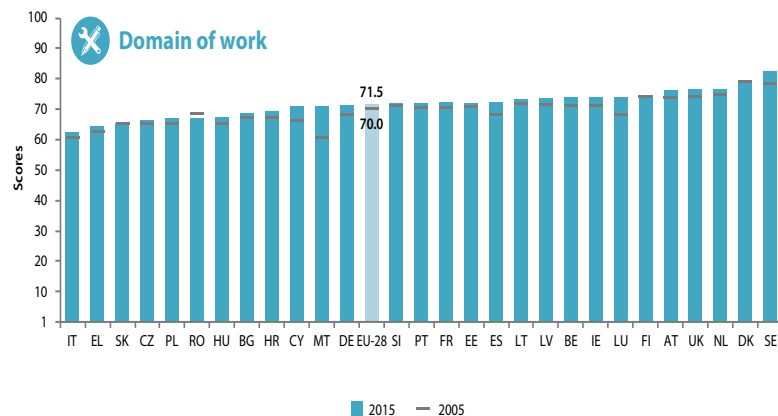


Figura 1.15. Gender Equality Index, Dominio del lavoro ed indicatori di sotto-dominio per il 2015 in Europa. Fonte: 2017a; 2017b

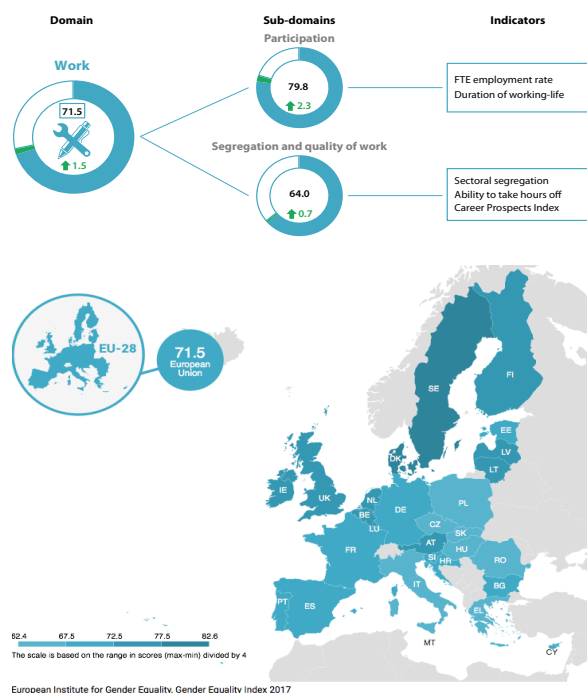


Figura 1.16. Gender Equality Index, Dominio del lavoro nei paesi membri dell'UE per il 2015. Fonte: 2017a; 2017b

Al settore del lavoro corrisponde anche il dominio della retribuzione. Tale dimensione esamina le disuguaglianze di genere in termini di retribuzione e reddito mensile. Nonostante le donne siano abilmente in grado di ottenere le medesime qualifiche ottenute dagli uomini, la loro retribuzione è nettamente inferiore. I risultati del *Gender Equality Index* (EIGE, 2017a; 2017b) mostrano un valore europeo di 79.6 e quindi notevolmente predisposto ad una condizione di parità tra i sessi. L'incremento decennale è pari a 5.7 punti e il reddito mensile presenta un incremento di 12.1 mentre la disponibilità economico-finanziaria mostra una decrescita di 3 punti. I progressi sono marginali se si considera l'estensione del decennio in cui lo studio è stato condotto. Rispetto al 2005 la Grecia propone addirittura una decrescita di -1.2 punti, dimostrando che la disparità nella retribuzione tra uomini e donne, dal 2005 al 2015 è aumentata.

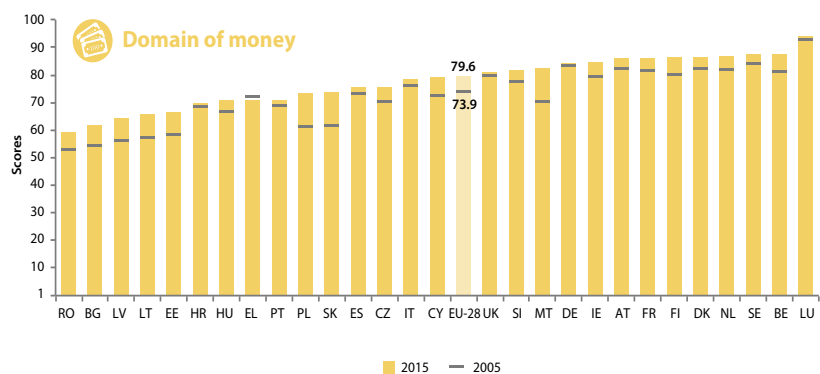


Figura 1.17. Gender Equality Index, Dominio della retribuzione ed indicatori di sotto-dominio per il 2015 in Europa. Fonte: EIGE, 2017a; 2017b.

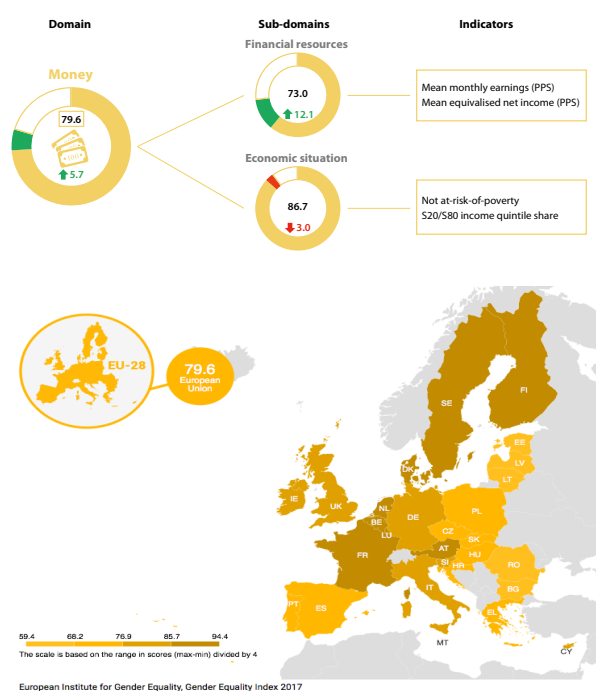


Figura 1.18. Gender Equality Index, Dominio della retribuzione nei paesi membri dell'UE per il 2015. Fonte: EIGE, 2017a; 2017b.

1.11.2. Cultura ed istruzione

Un aspetto determinante e strettamente connesso a quello della retribuzione è il fattore istruzione. Si pensa che la formazione sia uno dei fattori più influenti per il raggiungimento della parità di genere, di inclusione sociale e di eliminazione della povertà (EIGE, 2017a). Il decennio dal 2005 al 2015 vede secondo il

Gender Equality Index (EIGE, 2017a; 2017b) un leggero aumento in termini di istruzione e acquisizione di competenze da parte delle donne: si registra un incremento di 2.6 punti. Il dominio dell'istruzione mostra un punteggio europeo pari a 63.4. La maggiore crescita si può osservare a Cipro (+15.1), in Grecia (8.4), in Lussemburgo (+ 7.4) e in Italia (+7.3). Il *Gender Equity Index* indica una percentuale del 24% in tema di istruzione terziaria nei paesi dell'UE, sia per quanto riguarda il genere maschile che femminile, mentre si assiste ad un'intersezione di genere ed età. Infatti, nelle generazioni precedenti si dimostra che il completamento degli studi è in maggior numero realizzato da individui di sesso maschile. Dal 2015 in poi, si assiste ad un'inversione di tendenza, ovvero sono le donne di età compresa tra 30 e 34 anni che hanno superato di 3 punti il 40% del valore *target* previsto dalla Commissione Europa per il 2020 per entrambi i sessi. L'incremento percentuale delle donne nell'istruzione terziaria si riferisce prevalentemente a settori come l'educazione e la formazione, la salute, le discipline umanistiche ed artistiche.

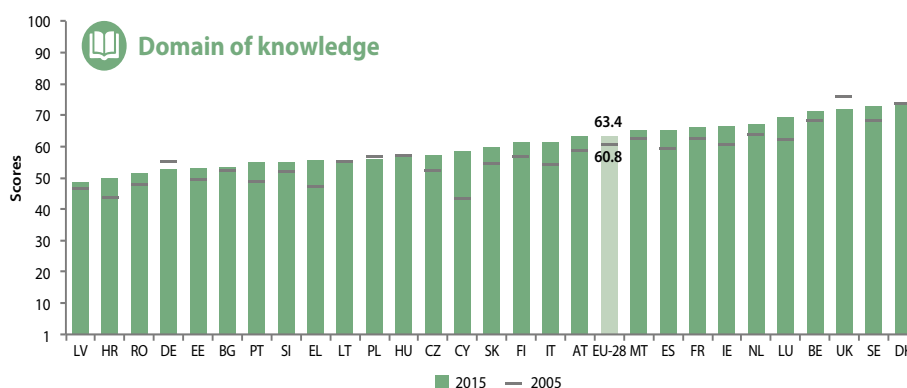


Figura 1.19. Gender Equality Index, Dominio dell'istruzione ed indicatori di sotto-dominio per il 2015 in Europa. Fonte: EIGE, 2017a; 2017b

Le statistiche dimostrano anche nel 2016 risultati formativi decisamente migliori da parte delle donne rispetto agli uomini. Il 44% delle donne di età compresa tra 30 e 34 anni completa il ciclo di studi nell'istruzione universitaria, mentre la percentuale di uomini che nel 2016 porta a termine il percorso scolastico è pari al 34% (*Eurostat, 2017; Figura 1.21*).

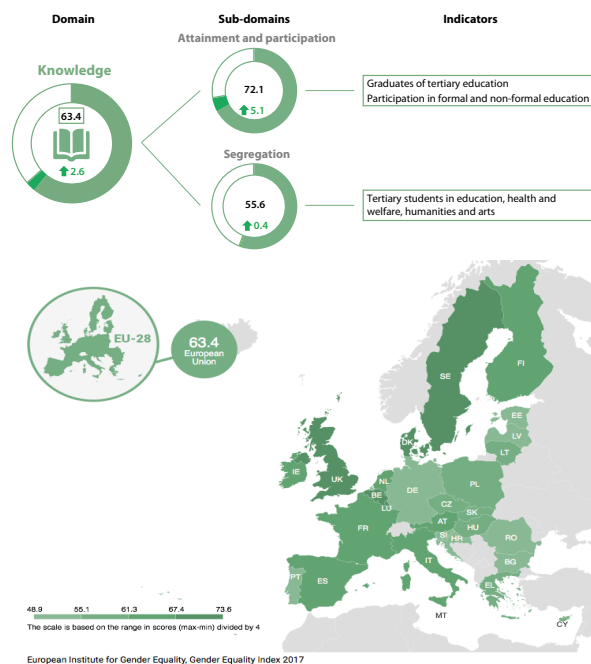


Figura 1.20. Gender Equality Index, Dominio dell'istruzione nei paesi membri dell'UE per il 2015. Fonte: EIGE, 2017a; 2017b

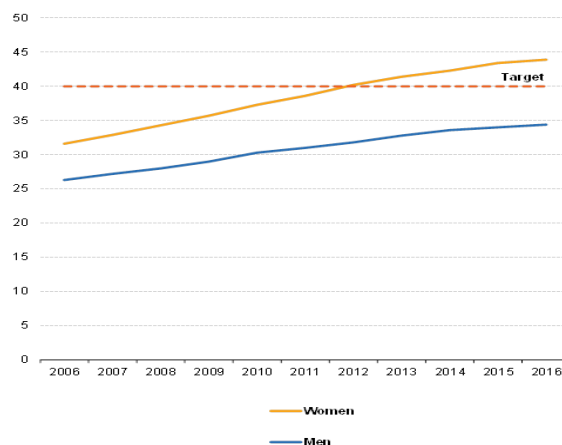


Figura 1.21. Andamento del livello di istruzione e formazione per genere (fascia d'età: 30-34 anni). Fonte: Eurostat, 2017

1.11.3. Il dominio del tempo

Dai risultati presentati attraverso il *Gender Equality Index* (EIGE, 2017a; 2017b) il dominio maggiormente critico è quello del tempo. Questa dimensione

riguarda la disuguaglianza di genere in termini di tempo a disposizione per assistenza, lavoro domestico e attività sociali. In questo ambito si comprendono anche attività lavorative di assistenzialismo non retribuito. Tali attività sono strutturate seguendo uno schema di sessismo maschilista alquanto stereotipato che crede che l'assistenzialismo rientri tra i lavori di esclusiva mansione femminile. È l'unico dominio del rapporto *EIGE* (2017a; 2017b) che propone un decremento nel decennio 2005-2015. Infatti, negli Stati membri dell'UE si distribuisce in un *range* che varia da 42.7 (in Bulgaria) a 90.1 (in Svezia). Nell'intervallo temporale della ricerca, si registra una crescita solo in otto Stati membri ed in particolare in Lettonia (+ 6,7), Repubblica Ceca e Spagna (entrambe con + 6). La decrescita si assiste invece in dodici paesi ma il maggior calo si osserva in Slovacchia (- 9.1), Belgio (- 9) e Bulgaria (- 8.3).

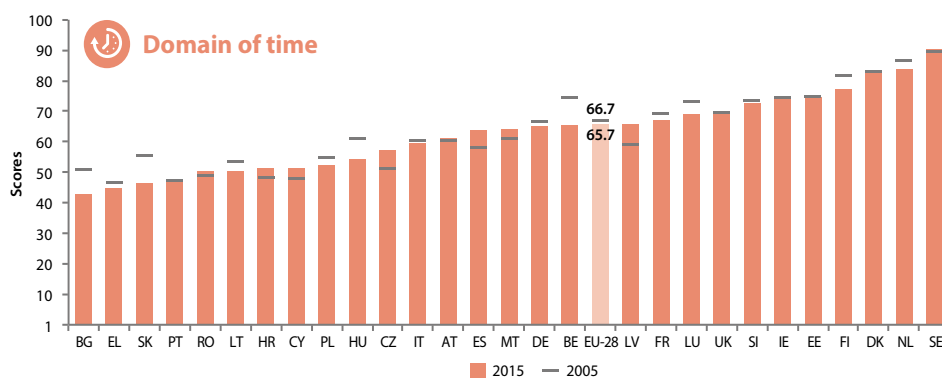


Figura 1.22. Gender Equality Index, Dominio del tempo ed indicatori di sotto-dominio per il 2015 in Europa. Fonte: EIGE, 2017a; 2017b

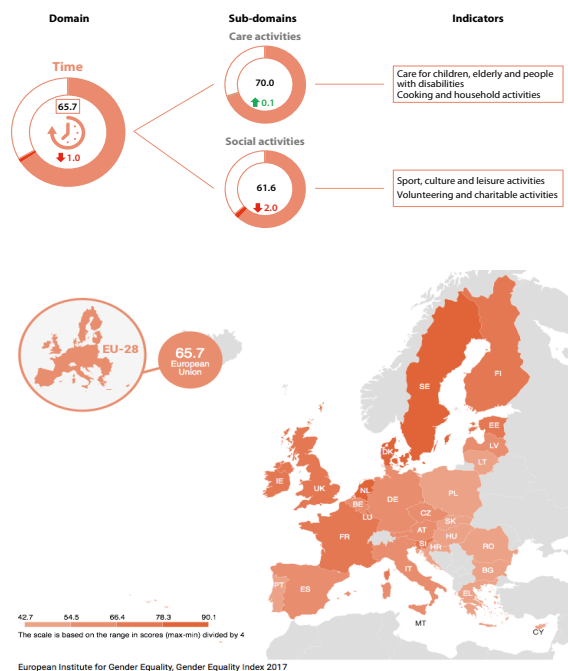


Figura 1.23. Gender Equality Index, Dominio del tempo nei paesi membri dell'UE per il 2015. Fonte: 2017a; 2017b

Il sotto-dominio del tempo si articola prevalentemente in attività ludico-ricreative che comprendono manifestazioni sportive, spesso combinate con attività di volontariato e filantropia. Tale settore mostra una diminuzione media europea di 2 punti nel decennio 2005-2015.

1.11.4. L'acquisizione di potere nell'ambito politico-istituzionale

Un'ulteriore dimensione è il cosiddetto dominio di potere o *empowerment*, il quale è deputato a stimare l'uguaglianza di genere nelle posizioni decisionali dei settori politico-istituzionale, economico e sociale. La parità in termini di partecipazione nelle posizioni di potere identifica la tutela dei diritti fondamentali degli esseri umani ed è alcontempo sinonimo di un governo efficace nelle sue politiche. Tale ambito rileva i maggiori progressi, mostrando una crescita di 9.6 punti nella media europea, nel decennio 2005-2015. In particolare, il progresso dell'*empowerment* si osserva in Italia (+ 29,2), Francia (+ 24,6) e Slovenia (+ 24,1); mentre la decrescita si registra in Repubblica Ceca (- 7,0), Slovacchia (- 3,8), Finlandia (- 3,1), Lituania (- 0,7) e Malta (- 0,4). La dimensione del potere si concretizza principalmente attraverso il potere

politico (esaminato attraverso la rappresentanza di donne nei ruoli politico-istituzionali) ed il potere economico (misurato dalla proporzione di donne e uomini nei consigli di amministrazione delle più grandi società registrate a livello nazionale e quotate in Borsa). Nel 2015, detengono il primato del sotto-dominio del potere politico la Svezia, Finlandia e Francia, mentre il più basso livello si registra in Ungheria, Cipro e Slovacchia.

Per quanto riguarda il potere economico dell'*empowerment*, il più alto livello di parità di genere è presente in Francia, Slovenia e Svezia, mentre il più basso valore si evince nella Repubblica Ceca, in Grecia ed in Slovacchia.

L'aumento del numero di donne che ricoprono incarichi parlamentari nazionali si caratterizza per un lento incremento: dal 21% nel 2005, al 28% nel 2015. La medesima tendenza si coglie nell'ambito delle istituzioni regionali e locali, con una crescita della rappresentanza femminile che va dal 26% nel 2010 al 28% nel 2015. La cosiddetta *quota femminile* dei seggi parlamentari adottata in Italia è più che raddoppiata dal 2005 (12%) al 2015 (30%). In Slovenia, nel decennio considerato, si passa dal 12% al 27% ed in Francia dal 14% al 26%. Tuttavia, la proporzione europea è frenata da paesi come la Bulgaria, dove si assiste ad una decrescita che va dal 25% nel 2005, al 21% nel 2015. La maggior parte dei progressi si osservano nei contesti in cui le *quote rosa* sono in vigore già da più tempo. “Nei recenti risultati elettorali (compreso il 2017) solo quattro Stati membri (ES, FR, PT, SI) sono vicini al raggiungimento di un obiettivo di quota” (EIGE, 2017a).

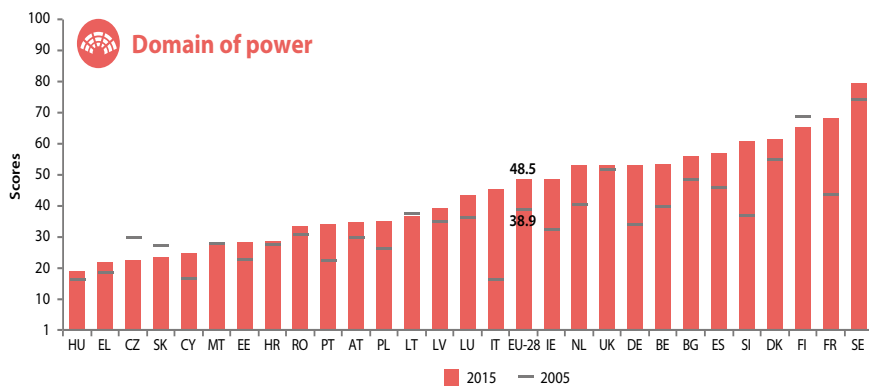


Figura 1.24. Gender Equality Index, Dominio del potere ed indicatori di sotto-dominio per il 2015 in Europa. Fonte: EIGE, 2017a; 2017b

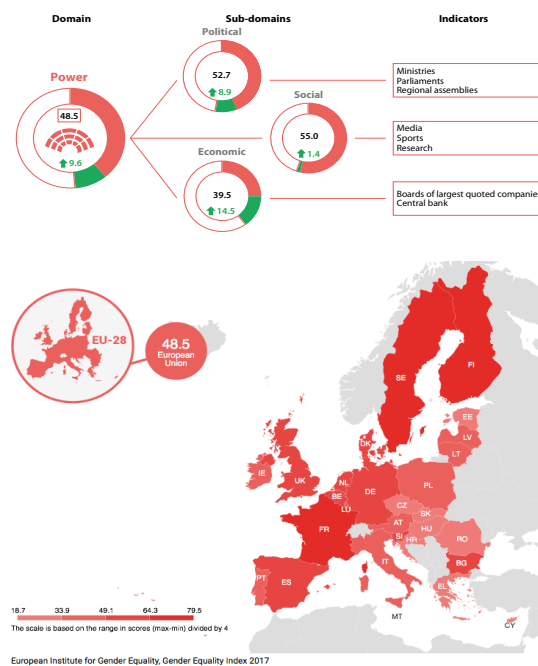


Figura 1.25. Gender Equality Index, Dominio del potere nei paesi membri dell'UE per il 2015. Fonte: EIGE, 2017a; 2017b

Se si guarda alla storia delle donne e agli incarichi politici della Presidenza del Parlamento europeo, si possono ascrivere solo due figure femminili durante tutto il periodo ciclo di vita del Parlamento stesso: Simone Veil (1979-1982) e Nicole Fontaine (1999-2002). Nell'ambito dei singoli contesti nazionali invece, negli ultimi anni, i governi hanno indicato un segnale di progresso con ad esempio Angela Merkel in Germania (mandato dal 2005 ad oggi), Beata Szydło in Polonia (2015–2017) e Theresa May in Regno Unito (dal 2016 al 2019). Se di considera invece la Presidenza di Stato, si possono menzionare in Croazia Kolinda Grabar-Kitarović (dal 2005), in Lituania Dalia Grybauskaitė (dal 2009) e nella Repubblica di Malta, Marie Louise Coleiro Preca (dal 2014). Per la Presidenza di Stato vi sono inoltre due donne eredi monarchiche, quali Elisabetta II in Regno Unito e Margherita II in Danimarca.

Quindi, le posizioni di *leadership* rimangono dominate dagli uomini, facendo sì che le donne rimangano sistematicamente fuori dalle posizioni di vertice. Se secondo il *Gender Equality Index* il valore di sintesi che esprime la massima parità tra i sessi è 100, un punteggio di 48.5 attribuito al dominio dell'*empowerment* permette di constatare quanto lontana sia l'uguaglianza tra i generi e quanto ancora lunga sia la strada per favorire l'inclusione femminile. Si tratta di una segregazione di tipo verticale radicata e riscontrata soprattutto nelle organizzazioni di ricerca, nelle emittenti pubbliche e nelle principali

federazioni sportive. Il *gap* di genere è comune anche nel panorama dei media dell'UE, sebbene quasi due terzi dei laureati in giornalismo siano donne. Esse infatti rappresentano solo un terzo delle iscrizioni al CdA e solo il 22% di loro raggiunge la presidenza di commissione in tutta l'UE (EIGE, 2017a).

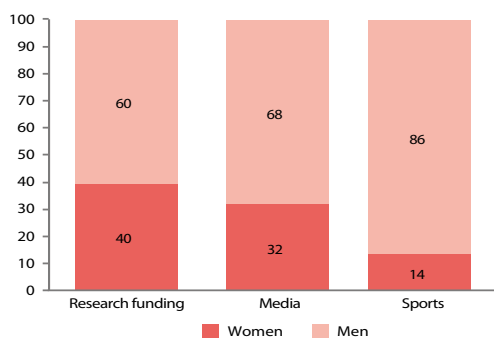


Figura 1.26. Percentuale di potere di uomini e donne nell'indagine dei settori: Ricerca, Media e Sport. Fonti: EIGE, Gender Statistics Database, WMID (Women and Men in Decision-Making), 2017a.

La problematica dell'*empowerment* non è solo una questione europea ma è un problema su scala mondiale. Negli ultimi vent'anni, solo il 23.6% di tutti i membri del parlamento del mondo sono donne. Mentre, alla fine del 2016 l'Europa è riuscita ad ottenere un picco del 37.3% nell'ambito della rappresentanza presso il Parlamento Europeo, il quale valore percentuale nel 2017 ha subito un calo di 1.2 punti. In tutti gli Stati membri dell'UE la maggioranza parlamentare è quindi di sesso maschile, ad eccezione della Finlandia che presenta valori di rappresentanza femminile uguali al 61.5%.

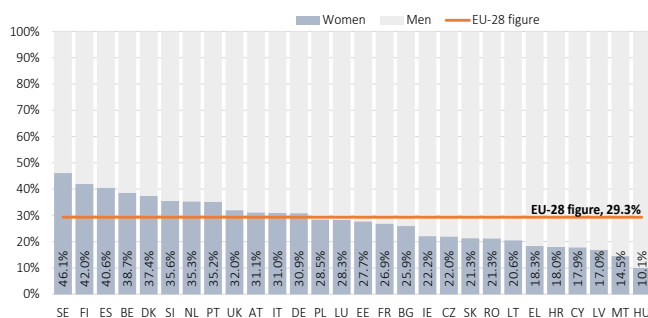


Figura 1.27. Quota di uomini e donne in rappresentanza presso il Parlamento. Fonte: EIGE, Gender Statistic Database, 2017b.

La segregazione di genere è altresì ascrivibile alla nomina ed assegnazione del Ministero di riferimento, quando le donne vengono elette. Infatti, se agli uomini vengono assegnati alti profili come Ministero di Grazia e Giustizia, Ministero della Difesa, Ministero degli Interni o degli Affari Esteri, alle donne vengono attribuiti ad esempio il Ministero dell'Istruzione e della Ricerca, quello della Salute o altri afferenti a discipline socio-culturali (EIGE, *Gender Statistic Database*, 2017). I dati parlamentari e governativi sono esplicitativi nell'ultimo trimestre del 2017 e mostrano un calo di 0.4 punti in Parlamento. Mentre al Governo la quota di donne *senior minister* è passata dal massimo storico del 28.5% nel febbraio 2017, a 27.7% nel novembre dello stesso anno.

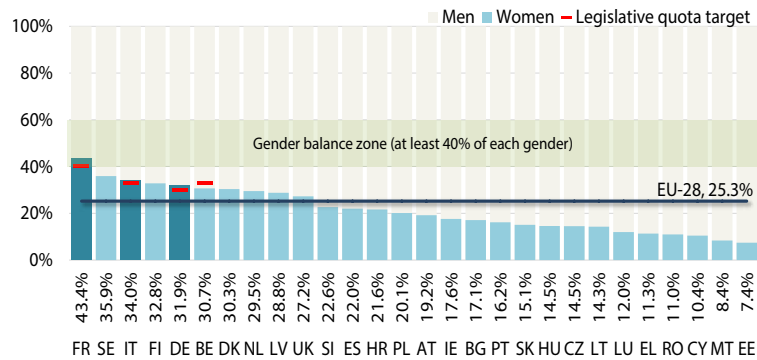


Figura 1.28. Andamento storico di uomini e donne in rappresentanza presso Parlamento e Governo. Fonte: EIGE, *Gender Statistic Database*, 2017b.

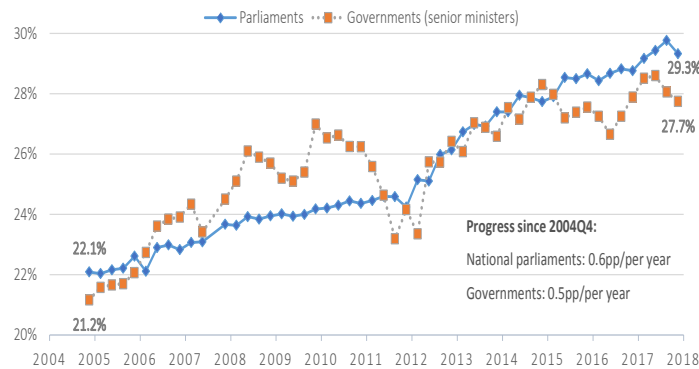


Figura 1.29. Quota di rappresentanza di uomini e donne nella leadership di aziende quotate in Borsa. Fonte: EIGE, *Gender Statistic Database*, 2017b.

Migliora invece l'andamento delle donne in termini di rappresentanza presso le assemblee regionali di tutti i paesi membri dell'UE, sebbene la *trend* sia molto lento e mostra una rappresentanza del 33.3% nel 2017. Anche la proporzione di donne che dirigono assemblee regionali ha raggiunto un nuovo massimo storico del 21,5% (EIGE, *Gender Statistic Database*, 2017b).

1.11.5. La rappresentanza nella leadership aziendale

Oltre alla rappresentanza nel contesto politico-istituzionale, non bisogna sottovalutare l'ambito della *leadership* aziendale. I dati dell'EIGE (2017) indicano che solo il 25.3% delle donne è membro nel CdA delle società quotate in Borsa. Nei paesi dell'UE, la Francia detiene il primato, con un valore pari a 43.4%, probabilmente grazie ad una misura legislativa del 2011 che ha obbligato le aziende a raggiungere un obiettivo del 40% entro gennaio 2017 (EIGE, 2017a). Nonostante si registri tale progresso all'interno dei CdA delle società quotate in Borsa, solo poche donne raggiungono i vertici, in termini di Presidenza o Direttore Esecutivo (*Chief Executive Officer*). Nel 2017, si osserva un andamento positivo: il 7.1% delle donne raggiunge la carica di presidenti ed il 5.5% è invece Amministratore Delegato.

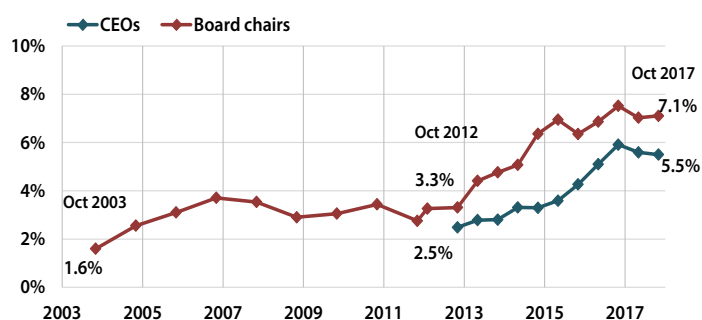


Figura 1.30. Andamento di rappresentanza di uomini e donne in posizioni di vertice di aziende quotate in Borsa. Fonte: EIGE, *Gender Statistic Database*, 2017b.

1.11.6. La dimensione della salute come percezione del sé

L'ambito della salute riveste altresì un ruolo decisivo, poiché essa è

strettamente connessa a fattori biologici, comportamentali, ambientali, economici, sociali, culturali e politici. L'eterogeneità del punteggio calcolato attraverso il *Gender Equality Index* mostra quanto differenziata sia la dimensione salutistica in Europa. Infatti, nel decennio 2005-2015 l'incremento è di 1.5 punti ed il valore medio è pari a 87.4.

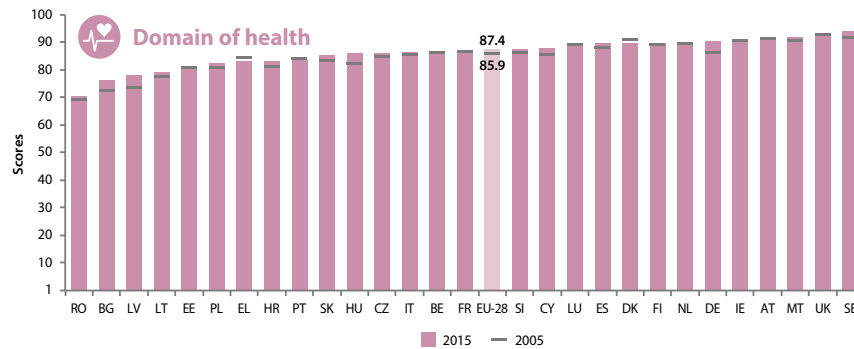


Figura 1.31. *Gender Equality Index, Dominio della salute ed indicatori di sotto dominio per il 2015 in Europa. Fonte: EIGE, 2017a; 2017b.*

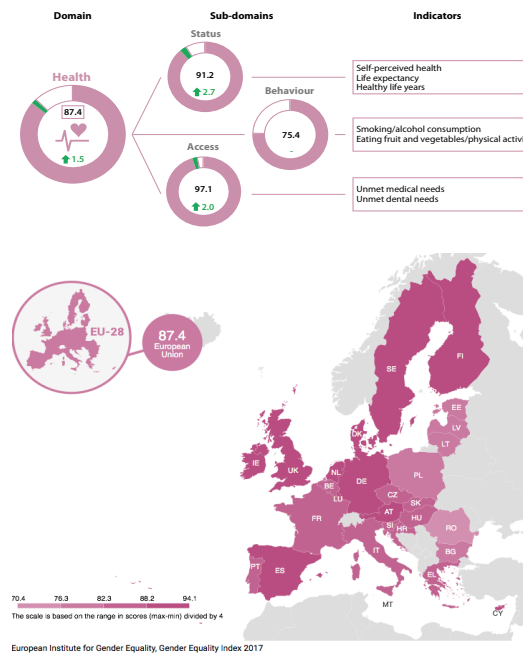


Figura 1.32. *Gender Equality Index, Dominio della salute nei paesi membri dell'UE per il 2015. Fonte: EIGE, 2017a; 2017b.*

L'auto-percezione delle proprie condizioni salutistiche si interseca in fattori sociali come livello di istruzione o come il tasso di attività sportiva. Pertanto, gli indicatori dell'uguaglianza di genere in Europa aiutano a determinare meglio quale sia il livello effettivo di trasformazione socio-culturale, politico-legislativa, economico-sanitaria ed educativo-professionale.

L'obiettivo di indagare la disuguaglianza di genere nei vari ambiti risulta essere un compito particolarmente facilitato rispetto all'intento di rilevare l'aspetto dimensionale della violenza fisica, psicologica, sessuale, economica e/o mortale nei confronti delle donne. Infatti, le statistiche ufficiali sull'uguaglianza di genere risultano aggiornate sia a livello di contesto nazionale che a livello internazionale. Sussistono misure di sintesi, indicatori costruiti ad *hoc* che consentono di realizzare un'attenta e puntuale analisi dei domini in cui la violenza simbolica continua a perpetuarsi, quali per l'appunto l'accesso all'istruzione; alla carriera, anche in posizioni di *leadership*; di ingresso al potere, in termini di gerarchia politico-istituzionale; di tempo libero o di salute e percezione del sé. Le fonti documentali permettono quindi di misurare i diversi contesti di analisi e di confrontare i dati a livello *cross-country*.

Al contrario, l'accurata ricerca di risorse documentali e statistiche ufficiali sulla violenza contro le donne in senso stretto può rivelarsi particolarmente complicata, a causa della mancanza di dati nazionali ed internazionali resi pubblici allo scopo di inquadrare la portata dimensionale del fenomeno. A tal proposito infatti, non è nota l'esistenza di alcuna organizzazione governativa o non governativa a cui sia stato affidato il compito di monitorare la violenza contro le donne nel mondo e di aggiornare in tempo reale le casistiche. La maggior parte dei *database* sono incompleti o non permettono di valutare il *trend* nel tempo o per area geografica. Subentrano questioni connesse alla all'accuratezza, accessibilità, tempestività e confrontabilità dei dati. Si assiste altresì a problemi legati alla praticabilità dei dati stessi, la cui compatibilità è caratterizzata anche dalle normative vigenti in materia di tutela della *privacy*. Pertanto, si avverte un certo limite nell'esigenza di osservare l'aspetto dimensionale e quantitativo del fenomeno della violenza contro le donne, soprattutto in un'ottica di comparabilità *cross-country*.

Nel tentativo di delineare uno scenario quantitativo, anche se poco aggiornato, le fonti che si possono interrogare sono vari, come *report* resi disponibili da enti sanitari o giudiziari, resoconti pubblicati da istituti statistici, rapporti stilati da enti di ricerca sociale, reportistica che deriva dai centri antiviolenza ma

nessuno di questi ci consente realmente di confrontare i dati e tracciare un andamento dei vari paesi nel corso tempo. Tuttavia, ciascuna di queste fonti, qualora fossero disponibili per ogni singolo paese, possiede un impianto metodologico e di campionamento differente. Alcuni studi si basano infatti su campioni di donne che vengono intervistate (FRA, 2014; Eurostat, 2017), mentre altre indagini cercano di allineare le notizie della stampa con gli atti processuali (ad esempio EURES-ANSA, 2014), tenendo conto di una sottostima dei crimini che non vengono denunciati.

1.12. Cenni conclusivi

In questo primo capitolo, attraverso le principali teorie emerse dagli studi della letteratura sul tema della violenza contro le donne, si è tentato di delineare un profilo sociologico del fenomeno, indagando le dinamiche dell'interazione tra uomo e donna. Più propriamente, con lo scopo di definire le potenziali ragioni sottese a spiegare la violenza maschile sulle donne, la chiave interpretativa si è focalizzata sul processo di mutua interazione tra i generi, che nel corso della storia ha riprodotto con arte copiosa una cultura androcentrica che ai giorni nostri ha raggiunto l'apice del paradosso. L'evoluzione contemporanea dei modelli di riferimento e i mutamenti socio-culturali dovrebbero aver trasformato l'intero funzionalismo sociologico, propriamente inteso come globalità di strutture sociali e culturali connesse tra loro, che invece negli anni non hanno fatto altro che contribuire a preservare ed avvalorare le condizioni essenziali per l'esistenza di un paradigma sociale fondamentalmente maschilista. La manifestazione fenomenologica in questo contesto viene osservata non isolando la vittima dal carnefice, ma indagando il processo che mette proprio in relazione la dinamica di interazione e prevaricazione tra gli individui, considerati come *soggetti sessuati*. Il tema oggetto di studio rappresenta in realtà un privilegiato osservatorio delle trasformazioni della socialità contemporanea, che seppure non afferma un concreto cambiamento rispetto al passato, riesce a portare alla luce il riconoscimento dell'*a-normalità* del fenomeno stesso e quindi a rappresentare un'importante presa di coscienza. L'effetto del dominio simbolico che deriva da una cultura ancestrale sedimentata nel tempo e tramandata di generazione in generazione è talmente incorporata nelle percezioni, nelle azioni e negli *habitus* degli agenti, che appartiene all'ordine naturale delle cose, ed è l'elemento essenziale di cui si

nutre la frustrazione della modernità contemporanea, che necessita della violenza per affermare alla propria identità, la propria esistenza. Lo smarrimento costante tra l'incessante esigenza di dimostrare di possedere potere e la frustrazione di non essere all'altezza della virilità tanto osannata dalla cultura simbolica sociale, scaturisce nella criminalizzazione genocida del più debole: la donna. La vittima è la figura che da sempre ha rappresentato l'esistenza di un'impari importanza, che è stata designata di una diversità non solo anatomica o genetica ma di diritto e dovere verso gli altri e verso il mondo. La sopraffazione dunque identifica la disperata esigenza da parte dell'identità maschile di dimostrare la propria esistenza, la propria forza e di compensare l'umiliazione, la disillusione e l'insoddisfazione che proviene dall'esterno. Le dinamiche oggi più che mai sono controverse e si celano dietro scenari spesso subdoli, come i rapporti affettivo-sentimentali o familiari. La violenza esperita sia dentro che fuori le mura domestiche si manifesta nei vari domini d'interazione, in un processo circolare che va dal collettivo, all'individuale e all'identitario, dove in quest'ultima istanza, dopo aver appagato e rafforzato la propria identità si ripropone nella collettività, sentendosi parte integrante di un sistema che funziona.

Malgrado la violenza sulle donne sia un fenomeno su scala globale, senza distinzione di longitudine o latitudine, riuscire a rilevare il livello dimensionale dello stesso rappresenta un notevole limite, a causa della mancanza di statistiche ufficiali aggiornate e capaci di coprire l'estensione mondiale. Pertanto, al fine di determinare un'angolazione internazionale del fenomeno e proporre un eventuale confronto socio-culturale tra diversi paesi, una strategia di analisi può concretizzarsi attraverso lo studio della percezione pubblica.

In questo contesto, l'opinione pubblica può essere intesa come l'insieme delle attitudini di pensiero e convinzioni di una popolazione. In particolare, in considerazione del fatto che esiste il limite umano di non poter avere conoscenza diretta di tutto ciò che accade al di fuori di un ambito di vita associativa circoscritto e limitato e poiché non può esserci esatta corrispondenza tra il mondo e le immagini che gli individui si fanno del mondo, sussiste una necessità di rappresentazione. In sostanza, "il bisogno di rappresentazione dello mondo determina 'pseudo-ambienti', attraverso cui si agisce nell'ambiente reale", così che lo pseudo-ambiente che si viene a creare tra i fatti e l'individuo non è reale ma si tratta di una sorta di ambiente invisibile, costituito dalle immagini che l'individuo riceve da diverse fonti, tra le quali i mezzi di comunicazione (Lippmann, 2015). Più propriamente, in

un'ottica in cui le dinamiche di opinione sono profondamente influenzate dai processi di comunicazione collettiva, può risultare utile indagare l'opinione pubblica internazionale che viene espressa proprio attraverso l'uso del linguaggio naturale. Lo studio dei diversi linguaggi o registri linguistici identifica una specifica area di ricerca delle Scienze Statistiche che verrà presentata nel secondo capitolo e che mira ad un'analisi *cross-language* capace di identificare attraverso la lingua, le similarità o dissimilarità tra le differenti culture sul tema della violenza contro le donne, in accordo con l'ipotesi che sussista un legame tra lingua e strutturazione collettiva di pensiero.

Capitolo Secondo

L'analisi dei dati testuali come strumento per gli studi *Cross-language*

La finalità del secondo capitolo esplicita l'interesse da parte della comunità scientifica per lo studio del *linguaggio naturale*, dato il suo diretto legame con la strutturazione cognitiva e culturale di una comunità. In particolare, lo studio del fenomeno linguistico, in un'era di globalizzazione e plurilinguismo, si formalizza attraverso il dominio delle analisi *cross-language* e quindi dei testi multilingue. In questo contesto, l'analisi del testo viene dettagliatamente presentata attraverso il passaggio dalla teoria alle fasi operative, e più propriamente tramite la definizione dei paradigmi *positivista* ed *interpretativista* delle Scienze Sociali che determinano l'affermarsi di approcci *qualitativi* e *quantitativi*. L'orientamento quantitativo indica il *focus* metodologico del presente lavoro di tesi e pertanto in questo capitolo, si propone un approfondimento delle definizioni e delle tecniche di analisi proprie della *Statistica Testuale*. Nell'ambito dello studio dei dati testuali, si presentano inoltre alcuni metodi fattoriali e di classificazione maggiormente utilizzati, ponendo particolare enfasi sulle tecniche adottate per i dati multilingue. Al fine di esplorare le soluzioni strategiche proposte dalla letteratura *cross-language*, si realizza una rassegna dei principali metodi e tecniche presentati dalla comunità scientifica afferente.

2.1. Formalizzazione del binomio lingua-cultura e degli studi multilingue in un'era di globalizzazione e plurilinguismo

I fattori che caratterizzano un paese sono molteplici ed uno degli elementi maggiormente identificativi è la dimensione linguistica o meglio le espressioni del *linguaggio naturale* che costituiscono una determinata lingua.

La facoltà esclusiva degli esseri umani di comunicare per mezzo di segni convenzionali si concretizza proprio attraverso il *linguaggio naturale*, che la filosofia del linguaggio, la linguistica e la neuropsicologia definisce come una *lingua* di lenta e naturale evoluzione nel tempo (Olivieri e Margiarotta, 2014).

Il fenomeno linguistico rappresenta un oggetto di indagine interdisciplinare, che vede l'interazione di diversi domini scientifici, come la linguistica, la psicologia, la sociologia, la statistica, l'informatica.

La scienza che per prima si è occupata di studiare le caratteristiche strutturali della *lingua* è nota come linguistica (Genetti, 2014). Originariamente la disciplina era definita “prescrittiva”, poiché finalizzata allo studio delle regolarità ricorrenti di una *lingua* senza tener conto di possibili eccezioni. Successivamente, la scienza è divenuta “descrittiva”, perseguendo l'obiettivo di organizzare e strutturare come un sistema gli elementi che la costituiscono. Nella linguistica moderna, denominata strutturale, il cui fondatore è il semiologo svizzero Ferdinand de Saussure (1916), essa si configura come l'insieme delle “abitudini” linguistiche di una comunità, grazie al quale ciascun individuo è in grado di comprendere e farsi comprendere. La parola si serve della lingua e consiste nella realtà concreta di ogni singolo atto linguistico in un rapporto di interdipendenza (Magli, 2008). Quindi, la lingua è un sistema di comunicazione, parlato o segnato, che concretizza la facoltà umana del linguaggio (Graffi e Scalise, 2002); consentendo una mutua comprensione tra i membri che appartengono alla stessa comunità.

Lo studio di una lingua si realizza fundamentalmente attraverso l'analisi di un *testo*, sia esso frutto di un'opera scritta o di trascrizione orale. Infatti, a partire dagli anni sessanta è il testo a costituire il segno linguistico primario e non più la singola parola o frase isolata da un contesto discorsivo (Magli, 2008). La cosiddetta *svolta testuale* della semiotica avviene quando de Saussure (1916) connota il testo di due componenti: il *significato* ed il *significante*. Il *significato* è il senso della narrazione; ovvero il contenuto che deriva dall'osservazione di un oggetto reale chiamato *referente*. Il *significante* è l'idioma, nonché l'insieme delle espressioni naturali della lingua di un paese. Louis Hjelmsley (1943, citato in Magli, 2008) meglio definisce la corrispondenza tra *significato* e *significante*: egli chiama piano dell'espressione il *significante* e piano del contenuto il *significato*. Nei casi di traduzione, in linea di massima, muta il *significante* ma non il *significato* (*Ibidem*). Infatti, ad esempio, un romanzo è una storia che può essere scritta in italiano e poi tradotta in francese, inglese, tedesco, spagnolo e così via. Cambia l'idioma ma il senso della storia e della narrazione resta lo stesso.

Secondo un linguista ed antropologo statunitense di origine tedesca Edward Sapir ed il suo allievo Benjamin Lee Whorf, le strutture grammaticali e semantiche di una lingua inglobano modalità di strutturazione del pensiero

(Sapir e Worf in Carassai e Crucianelli, 2017). La teoria nota come *Sapir-Whorf Hypothesis (SWH)* o *Ipotesi della relatività linguistica* afferma che lo sviluppo cognitivo di ciascun essere umano sia influenzato dalla lingua che parla (*Ibidem*). In una concettualizzazione più ampia e per certi versi estremizzata, il modo di esprimersi (scritto o orale che sia) determina il modo di pensare di un individuo. La formalizzazione dell'*Ipotesi della relatività linguistica* deriva da una sperimentazione condotta dal maestro di Sapir, il tedesco Franz Boas (*Ibidem*); il quale esplora le categorie lessicali e gli stili di vita di alcuni gruppi di nativi americani appartenenti a differenti famiglie linguistiche. Lo studio condotto da Boas mostra come al mutare degli stili di vita, mutino anche le categorie strutturali dei diversi luoghi; definendo così un interessante binomio tra la cultura di un popolo e la lingua parlata. Sulla base della concettualizzazione di Boas, gli approfondimenti condotti da Sapir consentono di non limitarci a considerare una connessione tra lingua e comportamento ma di spingerci verso una teorizzazione di interazione sistematica più ampia: si considera l'esistenza di un rapporto di influenza e interdipendenza reciproca tra lingua e comunicazione, tra comunicazione e cultura. Lo stesso de Saussure (Magli, 2008) ipotizza un rapporto tra lingua e cultura, nel momento in cui definisce la lingua come “un prodotto sociale della facoltà del linguaggio e come un insieme di convenzioni” (*Ibidem*). Il “corpo sociale” a cui fa riferimento de Saussure è il patrimonio delle consuetudini e del linguaggio specifico di un popolo che ne permette una strutturazione identitaria e di appartenenza.

Quindi, la naturalità di una lingua si concretizza in un determinato contesto storico, geografico e sociale. In particolare, in un'era di profonda *globalizzazione* economica e *culturale* esiste un dirompente pluralismo linguistico. L'attuale globalizzazione su scala mondiale e l'intensificazione dei reciproci scambi da parte delle economie nazionali ha generato un'interdipendenza di natura sociale, culturale, politica e tecnologica.

Consolidatasi durante il Ventunesimo secolo, la globalizzazione ha generato la diffusione delle reti telematiche ed il rafforzamento di Internet. Più propriamente, l'avvento del *Web 2.0* ed il moltiplicarsi delle piattaforme di *social network* ha consentito una rapida circolazione delle informazioni. I contenuti trasmessi e condivisi da una parte all'altra del mondo, attraverso il *Web*, possono essere espressi in qualsiasi registro linguistico. Di conseguenza, la globalizzazione plurisettoriale teorizza e concretizza una forte “apertura” di confini, da cui emerge la necessità di non “chiudersi” di fronte alla disponibilità

di risorse di testo multilingue. La letteratura suggerisce di definire il multilinguismo dei testi come *cross-language analysis* (Givòn, 1983; Vinokourov *et al.*, 2003; Enweiji *et al.*, 2017; Ratna *et al.*, 2018). Non esiste una definizione formale ma sussiste una nomenclatura intercambiabile e diversificata proposta dai diversi contributi di studio, come ad esempio *analisi multilingue (multilingual analysis)*; Bader e Chew, 2008; Banea *et al.*, 2008; Zhang e Rettinger, 2013; Balahur e Turchi, 2014; Beinborn e Choenni, 2019) o *analisi cross-lingual* (Wang *et al.*, 2012; Zhang *et al.*, 2013; Zhang e Rettinger, 2013; Wang e Manning, 2014; Xiao e Guo, 2014).

Qualunque sia la terminologia adottata, l'obiettivo consiste nel cercare di *allineare* il significato di più lingue in una rappresentazione comune, così da superare barriere linguistiche, spesso sfida dell'umanità (Beinborn e Choenni, 2019). I processi di globalizzazione (siano essi riferiti a fenomeni di politica internazionale, istruzione, politica sociale, diritto internazionale, occupazione, giustizia o sicurezza) e la compresenza del plurilinguismo rappresentano un "ponte" per la stretta interazione esistente tra contesti e fenomeni (de Saussure in Magli 2008; Sapir e Worf in Carassai e Crucianelli, 2017). Gli studi *cross-language* tentano di rispondere all'esigenza di agevolare l'incontro tra le diverse lingue. In altre parole, mirano a studiare i diversi *significanti* per giungere ai diversi o comuni *significati*, rappresentando così uno strumento di analisi *cross-cultural*. Attraverso la dimensione linguistica si possono quindi esplorare gli aspetti socio-culturali che caratterizzano i paesi in cui una lingua è quella ufficiale. Le diverse culture possono essere interpretate come "costellazioni di valori ridondanti che abitano i discorsi e che approdano a comunità diverse, talvolta intrecciandole" (Basso Fossali, 2008). Quindi, presupposta la coesistenza del rapporto lingua-cultura, in un contesto multilingue, l'analisi dei *significati* e dei *significanti* tenta di studiare in un'unica rappresentazione le diverse strutture linguistiche, tentando per mezzo di esse, di identificare strutture di senso comuni.

2.2. Dal testo al *corpus*: i criteri di classificazione

La lingua può essere indagata da differenti punti di vista:

✓ *fonologico*, in relazione ai suoni;

- ✓ *morfologico*, struttura interna delle parole;
- ✓ *sintattico*, struttura delle frasi;
- ✓ *semantico*, analisi del significato;
- ✓ *pragmatico*, analisi delle proprietà degli atti comunicativi;
- ✓ *lessicologico*, analisi e strutturazione del lessico.

L' 'oggetto di analisi di una qualsiasi rappresentazione del linguaggio - segnica o orale – diviene il testo; come suggerito dalla linguistica moderna (de Saussure in Magli, 2008). Dal latino *textus*, come participio passato del verbo *texere* raffigura l'ordine insito nella trama dei fili di una tela (Giuliano e La Rocca, 2008). Su questo presupposto etimologico, il testo viene inteso come il complesso linguistico di un discorso (*Ibidem*). In particolare, lo studio del linguaggio naturale non deriva dall'analisi di un singolo testo ma da una raccolta di più testi. Una collezione di testi o documenti omogenei tra loro secondo un preciso criterio di valutazione (Bolasco, 1999) viene detta *corpus*. Si determina quindi il passaggio dal testo al *corpus* come un oggetto di indagine. Un *corpus* può derivare da diverse fonti: "indagini sul campo, domande aperte o interviste; analisi di frammenti di un discorso o di testi corti, quali *abstract*, bibliografie, manifesti, messaggi, *etc...*" (Bolasco, 2005). La nozione di *corpus*² è nota già dai tempi del *Corpus Iuris Civilis* (528-534 d.C.) come raccolta del materiale normativo del diritto civile ai tempi dell'Impero di Giustiniano. Si può ricondurre, inoltre, al concetto primitivo di *corpus* anche la collezione, in lingua latina, della documentazione epigrafica, suddivisa per territori, degli usi e consuetudini dell'Impero Romano; la cui raccolta di 180.000 iscrizioni è conosciuta come *Corpus Inscriptionum Latinarum*. testi che costituiscono un *corpus* vengono selezionati ed organizzati seguendo precisi criteri (Lenci *et al.*, 2005; Morini, 2007). I parametri maggiormente rilevanti sono:

- *Generalità*: la selezione dei testi avviene in base al tipo di linguaggio utilizzato; ovvero se specialistico o generalista. Tramite il criterio della generalità è possibile identificare due tipologie di sotto-*corpus*:

² Plurale *corpora*

- a. *Corpus specialistico*: vi fanno parte insiemi di testi orientati allo studio di una particolare tipologia linguistica (*sub-language*; Lenci *et al.*, 2005) o quelli circoscritti ad un determinato ambito applicativo; come quello giornalistico, giuridico, medico, *etc.* I corpora specialistici hanno un livello di generalità minimo.
 - b. *Corpus generale o di riferimento*: vi appartengono quei testi che possiedono un'elevata generalità, poiché sono caratterizzati da una certa trasversalità del linguaggio. In altre parole, questo tipo di *corpus* consente di studiare il linguaggio proponendosi come una risorsa capace di rappresentare gli aspetti più caratteristici di una determinata lingua.
- *Modalità*: è il parametro che determina il modo attraverso il quale i testi vengono prodotti e si traduce nella seguente sotto-classificazione (*Ibidem*):
 - a. *Corpus scritto*: se è costituito solo da una collezione di testi scritti.
 - b. *Corpus parlato*: se deriva da una trascrizione di notiziari radiotelevisivi, conversazioni telefoniche, comunicazioni face-to-face, interviste, *etc...*
 - c. *Corpus misto*: rappresentato sia da testi scritti sia da trascrizioni di parlato in proporzioni variabili (*Ibidem*); ne è un esempio il *British National Corpus (BNC)*; <https://www.english-corpora.org/bnc/>).
 - d. *Speech database corpus*: questo tipo di *corpus* è costituito da campioni di linguaggio parlato in forma di segnale acustico, come TIMIT ad esempio (<https://catalog.ldc.upenn.edu/LDC93S1>).
 - e. *Corpus multimediale*: si tratta di una collezione di testo audio-video; ossia si hanno allo stesso tempo testi scritti, video e parlato in forma di segnato acustico (Lenci *et al.*, 2005).
 - *Cronologia*: rappresenta un criterio la cui selezione è dipendente dall'asse temporale. Nello specifico, si hanno:
 - a. *Corpora sincronici*: a cui afferiscono quei testi che sono riconducibili al medesimo *slot* temporale;
 - b. *Corpora diacronici*: testi che appartengono a periodi storici diversi; magari con lo scopo di descrivere il mutamento linguistico nel tempo.
 - *Lingua*: in questa categoria è possibile distinguere due tipologie di *corpus* (Osimo, 2001; Laviosa, 2002):

- a. *Corpus monolingue*: se la collezione di testi oggetto di analisi è espressa in una sola lingua e rappresenta il caso più frequente nello studio di un *corpus*.
- b. *Corpora multilingue*: rappresentano una collezione di testi i cui contenuti sono espressi in lingue diverse. A sua volta, la categoria dei corpora multilingue può essere scomposta in due sotto-tipologie (Osimo, 2001; Laviosa, 2002, Lenci *et al.*, 2005):
 - ✓ *Corpora paralleli o parallel corpora*: quando un testo viene tradotto in una o più lingue (Laviosa, 2002). Un *corpus* è parallelo ad un altro, attraverso un esatto allineamento delle frasi. Ordinariamente, il “parallelismo linguistico” riguarda due lingue alla volta; ecco perché nella maggior parte dei casi, lo studio multilingue si focalizza su *parallel corpora* bilingue.
 - ✓ *Corpora comparabili*: i testi non vengono sottoposti ad operazioni di traduzione ma ogni *corpus* conserva la propria struttura monolingue originale. Questa tipologia di *corpus* permette di confrontare due o più lingue rispetto ad uno stesso dominio tematico o ad uno stesso argomento. I corpora comparabili possiedono un maggiore grado di naturalezza, in quanto non sono sottoposti ad eventuali artificiosità che possono derivare dal processo di traduzione.
- *Codifica*: il criterio della codifica avviene attraverso delle etichette (o codici), le quali forniscono informazioni aggiuntive sul testo (Lenci *et al.*, 2005). Un particolare tipo di *corpus* codificato è il cosiddetto *corpus* annotato. In tal caso, le informazioni di codifica riguardano la struttura linguistica del testo a differenti livelli di rappresentazione, come la morfologia, la sintassi, la semantica, *etc...*
- *Dimensione*: è determinata dal conteggio delle parole contenute nel *corpus* stesso (*Ibidem*). Questa definizione è pertinente ad un testo di lingua scritta, mentre se il *corpus* oggetto di studio è parlato, il criterio della dimensione dipende dalla durata della registrazione vocale (*Ibidem*). I sub-corpora che derivano da questa categoria sono:

- a. *Corpora chiusi*: sono caratterizzati da uno standard tradizionale secondo il quale la quantità dei testi e delle parole è prestabilita. I corpora chiusi si formalizzano in una sorta di “scatto fotografico” della lingua, poiché data la staticità nel tempo non sono sottoposti a cambiamenti (*Ibidem*).
 - b. *Corpora aperti*: si tratta di una categoria venuta a battesimo grazie a John Sinclair (1991), il quale trasforma il concetto di standard tradizionale in quello di *corpus* di monitoraggio (*monitor corpus*). Un sistema si dice aperto, perché permette di introdurre nuovi testi nel *corpus* - rispettando sempre gli stessi criteri che hanno permesso di selezionare la collezione iniziale - con lo scopo di monitorare nel tempo la lingua. È ampiamente utilizzato per la sua dinamicità come risorsa per i lessicografi, al fine di aggiornare sistematicamente il vocabolario di una lingua (Lenci *et al.*, 2005).
- *Evoluzione*: questo criterio è strettamente connesso a quello della dimensione. Più specificamente, l’evoluzione in termini di capacità di analisi di un *corpus* da parte della scienza è proporzionalmente collegata alla quantità di testo oggetto di indagine. A seconda dei progressi che hanno caratterizzato la scienza del dato testuale è possibile profilare tre specifici domini connessi all’evoluzione (*Ibidem*):
 - a. *Corpus di prima generazione*: tra gli anni Sessanta e Settanta, dove i testi erano costituiti da milioni di parole;
 - b. *Corpus di seconda generazione*: tra gli anni Ottanta e Novanta in cui i testi analizzati contenevano decine di milioni di parole e poi dagli anni Duemila, in cui vi è stato un salto dimensionale a centinaia di milioni di parole;
 - c. *Corpus di ultima generazione*: identificano i tempi attuali ed indicano testi costituiti da miliardi di parole. In particolare, i cosiddetti *web corpora*.
- *Rappresentatività*: in termini statistici, le frasi che costituiscono un linguaggio formano una popolazione infinita, mentre un *corpus* identifica una porzione limitata e finita di un linguaggio (McEnery e Wilson, 1996). Pertanto, all’interno di una popolazione linguistica, un *corpus* è una raccolta di testi selezionati e organizzati secondo espliciti criteri, allo scopo di essere

usata come un campione rappresentativo del linguaggio o di una sua particolare varietà (*Ibidem*).

- *Bilanciamento*: è una condizione essenziale per garantire la rappresentatività di un *corpus*. Si presuppone la creazione di un profilo accurato della popolazione linguistica da studiare e da cui estrarre il campione di testo. Il bilanciamento prevede una mappatura dei confini spazio - temporali e tipologici dei corpora (Lenci *et al.*, 2005).

Tra i parametri di classificazione fino ad ora esaminati, il criterio della lingua ci consente a sua volta di organizzare i corpora attraverso un'ulteriore classificazione.

Poiché è indubbia la condizione per cui esiste una relazione imprescindibile tra fattori linguistici e fattori culturali (de Saussure in Magli 2008; Sapir e Worf in Carassai e Crucianelli, 2017),

si possono organizzare i corpora tenendo conto solo di questi due elementi: lingua cultura.

Vi possono essere testi che possiedono una struttura linguistica simile ma un inquadramento culturale differente, oppure testi le cui lingue presentano una simile struttura linguistica ed al contempo un *background* culturale affine (Krupa, 1969).

La classificazione dei corpora in base alla lingua e alla cultura si formalizza quindi in quattro diverse categorie:

- *Corpora monolingue monoculturali*: i testi provengono da paesi diversi ma il loro contenuto è espresso nella stessa lingua. Allo stesso tempo, i paesi presentano una certa similarità culturale. È il caso del Regno Unito e dell'Irlanda. Da un punto di vista linguistico, l'inglese è la lingua ufficiale del Regno Unito ma allo stesso tempo è la seconda lingua ufficiale dell'Irlanda (dopo il Gaelico irlandese). Per quanto riguarda il tessuto culturale, l'Irlanda è profondamente influenzata dal Regno Unito; poiché storica è la pluri-secolare egemonia britannica iniziata a partire dalle invasioni anglo-normanne del Decimo Secolo, con l'ufficiale annessione al Regno Unito nel 1801. Solo nel 1922, l'Irlanda si è conquistata la propria indipendenza anche se, tuttora, 26 delle 33 contee appartengono all'Irlanda; mentre le altre 6 sono ancora appartenenti al Regno Unito sotto il nome di Irlanda del Nord.

- *Corpora monolingue multiculturali*: in questo caso, i testi permettono di dare voce a più di una cultura anche nell'ambito di una stessa lingua. I corpora possono provenire da aree geografiche diverse e presentare aspetti socio-culturali differenti, nonostante la struttura linguistica presenti una forte similarità. Ascrivibile a questa categoria è il caso dell'inglese britannico e di quello nord-americano; le cui lingue prima della digitalizzazione sono state caratterizzate da sviluppi linguistici indipendenti dovuti ad una connotazione culturale completamente diversa. L'inglese britannico e quello americano presentano una struttura linguistica non identica ma affine e mostrano un inquadramento culturale estremamente diverso, poiché sono geograficamente distanti e storicamente difforni tra loro.
- *Corpora multilingue monoculturali*: a tal proposito, i testi fanno ricorso a lingue diverse ma con una certa prossimità culturale. Un esempio che mostra una struttura linguistica completamente difforme ma un inquadramento culturale simile è il caso dell'ungherese e dello slovacco. L'ungherese è una lingua facente parte del ceppo ugro-finnico mentre lo slovacco è una lingua slava. Tuttavia, il *background* culturale è simile, in quanto le due regioni sono state amministrate dallo stesso governo centrale, ovvero il Regno d'Ungheria o Impero austriaco o austroungarico (Logos Group, 2014). Un caso piuttosto particolare è dato dall'italiano e dal francese, le quali lingue malgrado siano diverse esprimono una prossimità linguistica strutturale (ceppo indo-europeo) e allo stesso tempo una certa prossimità culturale.
- *Corpora multilingue multiculturali*: sono testi di lingue diverse che esprimono al contempo culture differenti. Varia la struttura linguistica così come il tessuto culturale di cui è intrisa la comunità di riferimento; come ad esempio l'italiano e l'eschimese, l'inglese ed il cinese.

Definire i criteri per la classificazione di un *corpus* è strettamente connesso alla determinazione degli obiettivi di ricerca e all'individuazione dei metodi e delle tecniche, che consentono di perseguire le finalità desiderate.

Una collezione di testi può rappresentare oggetto di indagine per studiosi di varie discipline, i quali percorrono strategie differenti in base al dominio scientifico a cui appartengono.

Nello specifico, l'interesse per lo studio dei dati testuali proviene in particolare modo da tutte quelle aree scientifiche che afferiscono alle scienze sociali per lo studio dei comportamenti umani e delle teorie strutturaliste della società.

2.3. Gli orientamenti teorici delle Scienze Sociali che strutturano l'analisi dei dati testuali: paradigma positivista ed interpretativista

Nelle scienze sociali, l'osservazione di un qualsiasi fenomeno si fonda su un paradigma, come una guida per la scelta dei fatti rilevanti da studiare, la formulazione di ipotesi e la determinazione dei metodi e delle tecniche di ricerca.

Di fondamentale importanza è il cosiddetto "cambiamento di paradigma" (Kuhn, 1962), tramite il quale si struttura la scienza non solo come l'analisi specialistica di un fenomeno ma come uno strumento per comprendere le strutture della scienza stessa (Corbetta, 2014).

Infatti, attraverso il cambiamento paradigmatico si intravede l'organizzazione degli orientamenti teorici e metodologici su cui si basano le scienze sociali. I paradigmi fondamentali sono due: uno *positivista* e l'altro *interpretativista*.

Gli elementi cardine su cui i due paradigmi si fondano sono: la questione *ontologica*, ovvero il modo di concepire e sviluppare la scienza per studiare i fenomeni; l'aspetto *epistemologico* in termini di conoscibilità della realtà sociale ed infine l'approccio *metodologico*, nonché gli "strumenti" attraverso i quali si può conoscere la realtà o una sua angolazione.

I due paradigmi vanno a concretizzare due diversi orientamenti di ricerca. In particolare, è la diversa prospettiva *epistemologica* ad influenzare notevolmente l'approccio *metodologico* dell'analisi dei dati nelle scienze sociali e quindi anche il modo di studiare il testo.

Il paradigma *positivista* si afferma durante la metà del Diciannovesimo secolo e si interroga sulla realtà sociale "utilizzando gli apparati concettuali, le tecniche di osservazione, i metodi di misurazione e gli strumenti dell'analisi matematica" (*Ibidem*). L'analisi del testo è quindi sottoposta ai procedimenti di

inferenza delle scienze naturali. Poiché la sua *ontologia* vede la realtà caratterizzata da una materialità oggettiva (realismo ingenuo) e l'*epistemologia* si basa su una mancanza di reciproca influenza tra lo studioso ed il fenomeno analizzato, la conoscenza assume connotazione di “legge naturale” e l'aspetto *metodologico* è quindi sperimentale e manipolativo. L'analisi che deriva dal paradigma positivista, il cui massimo esponente è Durkheim, è prevalentemente induttiva e basata su tecniche assolutamente *quantitative*. Nel tempo, si afferma anche un sotto-paradigma di quello *positivista*, quale il *neopositivismo* che fondamentalmente si origina per rispondere alle critiche avanzate in merito al *positivismo* e metodologicamente si estende ad una visione di apertura verso i metodi *qualitativi* (*Ibidem*).

L'assunto teorico su cui si fonda il paradigma *interpretativista* afferma invece che la realtà non possa essere semplicemente “osservata” ma necessiti di “essere interpretata”. Quindi, il testo diventa oggetto di interpretazione da parte del ricercatore; attraverso l'esplorazione approfondita delle narrazioni di senso veicolate dai testi. L'*interpretativismo* si presenta come paradigma opposto a quello del *positivismo* e vede in Weber il suo massimo esponente. In termini *ontologici*, sostiene che non esista una sola realtà oggettiva ma ogni singolo individuo ne produce una personale e riconoscibile (costruttivismo e relativismo). Difatti, l'analisi di un testo è totalmente dipendente dalla capacità interpretativa dell'analista/ricercatore. Lo scopo è la ricerca di significato, piuttosto che una sperimentazione in cerca di legge (*Ibidem*). L'aspetto *metodologico* si caratterizza per tecniche di analisi *qualitative* e soggettive.

I due paradigmi esprimono chiaramente due approcci differenti all'analisi del linguaggio naturale. Il *positivismo* si basa su uno studio oggettivo e quantitativo, scomponendo il testo in unità elementari ed osservando le strutture linguistiche per mezzo di operazioni statistico-matematiche. Al contrario, l'*interpretativismo* mira ad un'interpretazione più che ad un'osservazione delle parole presenti in un testo. Un *corpus* viene studiato attraverso le parole presenti nel testo, con lo scopo di determinarne i contesti. “Il testo è il luogo in cui, nel ‘circolo dell'interpretazione’ si ricompona la dialettica tra spiegare e comprendere” (Ricoeur, 1987).

Direttamente dalla concezione paradigmatica *positivista* ed *interpretativista* deriva una contrapposizione tra *analisi qualitativa* e *quantitativa* dei dati.

Si tratta di due orientamenti di ricerca che identificano una storica e dogmatica opposizione di

metodi e tecniche, che nel tentativo di analizzare il fenomeno linguistico tendono a confondersi e completarsi (Giuliano e La Rocca, 2008).

Trobia (2012) definisce l'opposizione terminologica come "etichette diabolicamente utilizzate con eccessiva disinvoltura" e che possono risultare pericolose nell'ambito delle scienze sociali (Silverman, 2000 citato in Trobia, 2012). Per alcuni studiosi, si tratta di un contrasto che nel "vuoto definitorio su cui si regge la contrapposizione - data la pluralità di linee di convergenza e divergenza - si riduce ad uno stereotipo" (*Ibidem*). Il criterio di mutua esclusività che predilige un approccio piuttosto che un altro non soddisfa più la scienza e in alcuni ambiti richiede una compresenza quali-quantitativa. In particolare, nell'ambito dei dati testuali, l'orientamento quali-quantitativo consente di indagare il "campo micro-sociale e macro-sociale" in cui si produce il testo, afferrando la totalità dei significati per scendere nel particolare e controllare le ipotesi attraverso classificazioni e confronti (Giuliano e La Rocca, 2008). In questo modo, diventa "molto più sfumato il confine tra 'parole che contano' e 'conteggio di parole'" (*Ibidem*). Nei due approcci *qualitativo* e *quantitativo*, ciò che fondamentale cambia è il rapporto che soggiace la teoria e i metodi di ricerca. Per l'analisi dei dati testuali, il paradigma *positivista* si basa su tecniche e metodi *quantitativi* mentre il paradigma *interpretativista* su metodi e tecniche *qualitative*.

2.4. Dalla teoria alle procedure operative: analisi qualitativa e quantitativa a confronto

L'etimologia del termine *qualitativo* deriva dal latino *qualitatem* e rimanda alla nozione di determinazione logica; ovvero indica i metodi e le tecniche finalizzate a "determinare la 'natura' di un certo oggetto di studio" (Trobia, 2012). Il termine *qualitativo* implica un assetto teorico ed operativo che nell'ambito dell'analisi testuale tenta di "privilegiare le parole rispetto ai numeri, assumendo il punto di vista classico dell'ermeneutica" (Giuliano e La Rocca, 2008), che in filosofia è la pura interpretazione dei testi scritti.

Lo studio *qualitativo* per eccellenza dei dati testuali è di fatto proprio l'analisi ermeneutica, che si concretizza nell'analisi diretta di un testo, nel tentativo di

oggettivarlo attraverso la comprensione della complessità dei significati (polisemia) nascosti nella discorsività.

Le origini dell'ermeneutica si riconducono all'antica Grecia ed in particolare, la sua etimologia viene fatta risalire al dio Hermes, quale messaggero dell'Olimpo che incarna uno "strumento" di interpretazione e spiegazione del pensiero divino sugli uomini. Non a caso, la pratica ermeneutica viene consolidata per l'analisi dei Testi Sacri in termini di esegesi; nonché come interpretazione dell'intento autorale. Oggi, l'approccio ermeneutico identifica la teoria generale delle regole interpretative di qualsiasi tipo di testo: documenti storici, interviste, questionari, biografie, *etc.* Si basa su due principi fondamentali: la *coesione*, in termini di regole e collegamenti formali all'interno di un testo e la *coerenza*, come criterio per identificare possibili legami semantici. Il principio della coesione è determinante al fine di rintracciare nella successione delle frasi, il senso narrativo ed una sorta di continuità o discontinuità tematica. Si definiscono innanzitutto le dimensioni di testo da analizzare, determinando così l'unità di analisi ermeneutica che può essere un solo segmento di testo, la selezione di alcuni paragrafi, il testo nella sua interezza o solo frammenti che contengono alcune parole caratteristiche. La scelta dipende dal ricercatore e dagli obiettivi conoscitivi. In un'analisi ermeneutica "classica" si individua il tipo di narrazione e l'utilizzo delle parole, attraverso l'osservazione degli elementi lessicali e gli elementi retorici. I primi identificano il modo in cui la discorsività del testo propone l'uso dei sostantivi, dei pronomi (se al singolare o al plurale), dei tempi verbali, *etc.* Gli elementi retorici consentono invece di individuare temi o argomentazioni ricorrenti. Spesso, all'interno di un testo si determinano varie "categorie concettuali" che vengono analizzate singolarmente, oppure si rintraccia una categoria principale e si decide di "far ruotare attorno ad essa l'interpretazione del testo" (La Rocca, 2007). L'ermeneutica si prefigge inoltre di fornire un'interpretazione dotata di spirito critico, che smascheri i meccanismi latenti ideologici presenti nel testo; non attraverso l'oggettivismo ma mediante una critica razionale. Nascono spontanee le problematiche connesse proprio all'interpretazione critica. Lo stesso testo, se letto da persone diverse, può produrre differenti interpretazioni, poiché si tratta di una "attualizzazione dei significati di testo" (Giuliano e La Rocca, 2008). L'oggettività viene inevitabilmente compromessa da una dirompente soggettività; chiamata anzi intersoggettività, che attraverso la competenza dei codici e dei riferimenti contestuali organizza la sua interpretazione (*Ibidem*).

L'approccio qualitativo del testo presenta nei metodi e negli strumenti tre tratti distintivi: un'assenza di matrici di dati, una ispezionabilità della base empirica (poiché su base testuale e non etnografica) ed un carattere informale delle procedure (Ricolfi, 1997; La Rocca, 2007).

Il termine *quantitativo*, dal latino *quantum* fa invece riferimento alla "determinazione numerica" (Odifreddi, 2004 citato da Trobia 2012). Di conseguenza, quando si pensa ad un approccio quantitativo si mira a determinare in quale proporzione elementi e proprietà siano presenti in un oggetto di studio. Nell'ambito dell'analisi testuale lo studio quantitativo si prefigge di sottoporre le parole al dominio dei numeri (Giuliano e La Rocca, 2008). "Anche quando i dati testuali assumono una codifica numerica e le parole vengono sottoposte a conteggio, le relazioni tentano di misurarne un significato" (*Ibidem*). In termini di operativizzazione rispetto alla teoria, la prospettiva quantitativa si caratterizza per almeno tre caratteristiche (Ricolfi, 1997). La prima è l'impiego di una matrice di dati, che nel caso dell'analisi testuale è la *tabella lessicale*. La seconda caratteristica è la presenza di definizioni operative circa le modalità della matrice, che per lo studio del testo vede *termini* e *documenti* (testi). L'ultimo elemento è il ricorso alle tecniche statistico-matematiche per l'analisi dei dati, che nella fattispecie si basa sulla stima della frequenza delle parole che costituiscono un testo e delle possibili combinazioni terminologiche.

Fino alla metà del Ventesimo Secolo era assolutamente impensabile contemplare un'analisi quantitativa dei dati testuali, poiché si trattava di un campo di ricerca tradizionalmente ed esclusivamente riservato a linguisti, sociologi, psicologi ed esperti di comunicazione.

Nell'ambito dello studio quantitativo del testo si afferma uno specifico dominio di ricerca e di analisi come quello della *Statistica Testuale*.

Tra gli approcci qualitativo e quantitativo, vale la pena menzionare una particolare tipologia di studio: l'*analisi del contenuto* o *content analysis*; ampiamente dibattuta dalla scienza, nel tentativo di collocarla in uno dei due orientamenti. Essa identifica una serie ampia ed eterogena di tecniche appartenenti al dominio delle scienze sociali, orientate a stabilire "inferenze valide e replicabili a partire dai dati, per arrivare al relativo contesto" (Krippendorff, 1983). L'utilizzo pionieristico dell'analisi del contenuto risale ad una raccolta di novanta inni religiosi, i cosiddetti *Canti di Sion*. Un'opera commissionata dal clero della Chiesa svedese del Diciottesimo secolo, per

evidenziare i contenuti eterodossi della raccolta in oggetto, rispetto ai canti ufficiali proposti dalla Chiesa (Tuzzi, 2003; La Rocca, 2007). Rappresenta uno degli approcci più tradizionali per l'analisi del testo, attraverso cui si rilevano le dimensioni di senso che sono presenti in un dato testo (Bolasco, 1997). Il modello della *content analysis* mira a classificare le espressioni del testo (parole, frase, proposizioni) in opportune categorie, chiamate "categorie di contenuto", che sono definite in rapporto a specifici obiettivi conoscitivi. La *content analysis*, per quel che riguarda il campo della ricerca empirica delle scienze sociali, ha potuto esplorare un ventaglio di definizioni e di approcci piuttosto variegato, con mutamenti significativi sia dal punto di vista *epistemologico* che *metodologico*. In particolare, viene caratterizzata da un'identità puramente quantitativa dal momento che Laswell *et al.* (1949) estendono la definizione di *content analysis* ad una metodologia basata sulla "semantica quantitativa"; che esplicitamente richiama le tecniche statistico-matematiche. Tuttavia, nel corso del tempo, la trattazione puramente quantitativa dell'analisi del contenuto viene spostata verso metodi e tecniche proprie dell'analisi qualitativa (Ricolfi, 1997; Berg, 1998). La chiave di rilettura attribuisce all'analisi del contenuto un taglio maggiormente qualitativo, che deriva dall'intercettare "le dimensioni fondamentali di senso" a cui fa riferimento Bolasco (1997) e che conferiscono, attraverso l'oggetto di studio puramente qualitativo, un *habitus* tipicamente qualitativo anche alla tecnica.

2.5. L'affermarsi della Statistica Testuale come metodo quantitativo

L'approccio prettamente quantitativo all'analisi del testo si identifica attraverso la cosiddetta *Statistica Testuale* (Lebart e Salem 1994), nata a partire dagli anni Sessanta-Settanta e fortemente condizionata dall'evoluzione dell'Informatica. Il connubio disciplinare ha reso possibile un tipo di studio noto come analisi automatica dei testi (Lebart e Salem, 1988; *Ibidem*). Precursori dell'approccio quantitativo moderno sono Zipf (1935) e Yule (1944) che si sono focalizzati sul compito di codificare le regolarità presenti nella lingua e Benzécri (1963, 1973, 1981, 1982), il quale formalizza proprio sui dati linguistici le sue prime sperimentazioni. L'analisi testuale quantitativa si caratterizza di un progressivo mutamento in termini di finalità conoscitive della "lingua": si parte da una logica linguistica ad una lessicale, poi testuale fino ad approdare all'analisi lessico-testuale (Bolasco, 2005). Di pari passo, si assiste ad un progressivo

cambiamento di tecniche, strumenti, definizioni dell'unità di analisi e vere e proprie "filieri per l'elaborazione dei dati testuali" (*Ibidem*).

Dapprima, i fondamenti della statistica per il trattamento del linguaggio naturale si applicavano per lo studio di testi letterari, quali opere di grandi autori che richiedevano complesse operazioni di preparazione del testo. Da allora, la statistica testuale ha subito una notevole evoluzione sia in termini di metodi e tecniche, sia in termini di ambiti di applicazione. Un'analisi di dati testuali capace di produrre valore in qualsiasi settore, rientra in quel dominio di ricerca oggi noto come *Text Mining*. L'analisi del testo, malgrado il supporto della statistica e delle scienze informatiche rimane un fenomeno complesso, poiché il testo rappresenta un oggetto di studio controverso a causa delle proprietà tipiche del linguaggio naturale e delle ambiguità di senso in esse presenti.

Riuscire ad analizzare un testo da un punto di vista quantitativo significa considerare un oggetto di analisi che non è tipico delle scienze quantitative. Il testo non rappresenta una risorsa strutturata o espressa in formato numerico, che è la consuetudine delle applicazioni statistico-matematiche. Pertanto, il compito principale della statistica testuale è di *strutturare il non strutturato*. In particolare, l'analisi testuale quantitativa si prefigge di studiare il *significante* per arrivare al *significato*, muovendosi dalla sintattica e grammatica verso la semantica (Bolasco, 2004).

Trasponendo le entità oggetto di analisi in termini puramente statistici, il collettivo che viene analizzato è rappresentato da un *corpus* testuale che ne costituisce il *database* documentale. L'organizzazione di un *corpus* avviene secondo un criterio di omogeneità così come definito da Bolasco (1999) nella sezione 2.2 di questo capitolo. Tra i diversi livelli, cui è possibile studiare e descrivere le strutture linguistiche, la morfologia, la sintassi e la lessicologia rientrano tra le discipline più interessanti, in una visione di trattamento puramente statistico del fenomeno linguistico. Lo studio statistico dei dati testuali prevede una fase di preparazione del testo, una di organizzazione dei dati e l'analisi statistica in senso stretto.

La fase di preparazione mira a trasformare l'informazione testuale in "dato" testuale, attraverso gli strumenti tipici della linguistica, come dizionari o lessici di frequenza che consentono lo studio delle categorie sintattiche e grammaticali. L'organizzazione dei dati testuali avviene attraverso le procedure di codifica, che permettono di strutturare i dati in matrici, con lo scopo di consentire il passaggio alla fase di analisi statistica vera e propria.

2.5.1. Le fasi operative della Statistica Testuale

Essenzialmente l'analisi automatica dei testi si concretizza in tre fasi operative (Bolasco, 2005), di cui le ultime due rappresentano un approccio integrato:

1. *pre-elaborazione del testo;*
2. *analisi lessicale;*
3. *analisi testuale.*

Una delle principali difficoltà del trattamento ed analisi di un testo è riuscire a strutturare ed organizzare i dati prima che vengano elaborati. Il processo di organizzazione e strutturazione seleziona le parole che rappresentano l'asse sintagmatico di un testo, attraverso una fase preliminare di elaborazione nota come *pre-processing* o *pre-trattamento* del testo. Non esiste né una definizione univoca del concetto di *pre-trattamento* di un testo, né una successione prestabilita di operazioni sequenziali da compiere. Si tratta di scelte che mutano a seconda del tipo di testo e degli obiettivi di ricerca perseguiti. Nel pre-trattamento si delineano delle sotto-fasi legate alle leggi del linguaggio naturale, quali il *parsing*, la *normalizzazione*, la *lemmatizzazione*, il *tagging grammaticale*, il *tagging semantico*, *etc.* Malgrado un certo livello di soggettività presente in questa fase, è bene considerare che il pre-trattamento risulta essere molto importante per l'analisi di un testo, tale da condizionare e compromettere i risultati dell'analisi stessa.

Convenzionalmente, l'unità di analisi di un testo è la *parola* (Bolasco, 2005). Riguardo alla scelta dell'unità di analisi, in letteratura, sussiste un vero e proprio dibattito nell'ambito dello studio quantitativo dei testi (Bolasco, 1999). Una volta acquisito un *corpus*, esso si considera come una successione di *forme grafiche* (Lebart e Salem, 1988; Lebart *et al.*, 1998); ovvero una sequenza di caratteri alfabetici delimitata da due separatori. L'analisi delle forme grafiche è indipendente dalla lingua, in quanto privilegia i segni (*significanti*) per arrivare al contenuto o senso (*significato*).

Il primo passo che consente di scansionare il testo ed ottenere la lista delle forme grafiche che lo costituiscono, al di là di un'informazione del tutto semantica, è detta *parsing*.

L'operazione di *normalizzazione* agisce sull'insieme dei caratteri non separatori, al fine di eliminare possibili "duplicazioni" del dato testuale; come ad esempio le stesse forme grafiche che si presentano con lettera sia maiuscola

che minuscola. L'operazione di normalizzazione, inoltre, consente di uniformare il testo eliminando le forme grafiche che mostrano una maggiore variabilità, come le date, le sigle o i nomi propri.

Attraverso la *lessicalizzazione* è possibile individuare, all'interno del *corpus*, le unità minime di senso con un'estrazione di *segmenti ripetuti* e fissando a priori una soglia di frequenza. I segmenti possono derivare da una ripetizione di elementi grammaticali o da caratteristiche che identificano unità di senso indipendenti. In questa fase, è il ricercatore che può decidere quali segmenti considerare.

Con il *tagging grammaticale*, si può attribuire ad ogni termine la categoria grammaticale a cui afferisce (pronome, aggettivo, verbo, sostantivo, *etc...*). Questa fase di pre-trattamento consente di identificare le parti del discorso, comunemente chiamate *Part Of Speech (POS)*; le quali sono ascrivibili a due classi di appartenenza: una lessicale e l'altra funzionale. La prima raggruppa le categorie lessicali come sostantivi, aggettivi e verbi; mentre la seconda categoria comprende articoli, congiunzioni, pronomi, avverbi, preposizioni ed hanno all'interno di una grammatica un ruolo strumentale alla costruzione delle frasi. Le forme strumentali sono anche dette "parole vuote", poiché non sono portatrici di significato e non aiutano certo ad arrivare al contenuto di un testo. Per tale motivo, nella fase di pre-trattamento vengono filtrate, riducendo così la quantità di rumore presente nei dati. Sono le forme principali o "parole piene" a determinare le parti sostanziali del contenuto di un testo.

Il *tagging semantico* determina un processo attraverso il quale si etichettano le parole che costituiscono il vocabolario del *corpus* con informazioni di natura semantica.

Definita la *POS* di un testo, si può passare ad un'operazione di *lemmatizzazione*, in cui si richiama l'analisi dei processi morfologici e si riconduce ogni forma flessa al *lemma* di appartenenza. Un *lemma* identifica la forma usuale con cui le parole si presentano sui dizionari attraverso le determinate voci lessicali; come ad esempio il singolare per i sostantivi o l'infinito per i verbi.

Terminato il *pre-trattamento*, la fase successiva consiste nell'organizzazione dei dati e coincide con la fase lessicometrica, nella quale il dato statistico rilevato è il numero di volte in cui un'unità lessicale si presenta nel *corpus*. Questa fase fornisce una rappresentazione paradigmatica del testo e si compone di due momenti principali: la *costruzione del vocabolario (o lessico di frequenza)* e la realizzazione della *matrice lessicale* del *corpus*.

2.5.2. La codifica e i sistemi di ponderazione dell'unità di analisi

La collezione di testi presa in esame si sottopone ad una codifica che associa a ciascun testo una rappresentazione compatta di contenuto. Si tratta, quindi, di introdurre un adeguato sistema di pesi che tenga conto della diversa importanza delle parole.

Nel trattamento automatico dei testi si ricorre generalmente al cosiddetto *bag of words*, ovvero una trasformazione dei documenti in vettori, in uno spazio multidimensionale (Bolasco, 2005). Un generico vettore/documento d_j è rappresentato come:

$$d_j = [w_{1j}, \dots, w_{ij}, \dots, w_{pj}]$$

dove w_{ij} indica l'importanza della parola *i-esima* nel testo o documento *j-esimo*, in termini di contenuto informativo ($i=1, \dots, p$ e $j=1, \dots, n$) nel caso in cui la base testuale sia costituita da p parole differenti presenti in n testi.

Contestualmente allo *step* di codifica, è opportuno determinare anche il sistema di ponderazione dell'unità d'analisi. La letteratura a seconda dei casi e delle finalità propone diverse alternative:

- *Booleano*, in cui il peso w_{ij} assume valore pari ad 1 se il termine i è presente nel testo o documento j . In caso contrario assume valore uguale a 0.
- *Frequentista*, quando il peso w_{ij} è uguale a n_{ij} , nonché la frequenza del termine i nel documento j .
- *Frequentista normalizzato*, dove il peso w_{ij} assume valore pari a $n_{ij}/\max n_j$; con $\max n_j$ uguale alla frequenza del termine che si presenta più volte nel documento j ;
- *Term Frequency/Inverse Document Frequency (TF/IDF)*, il cui sistema di ponderazione, originariamente presentato da Salton e Buckley nel 1988, viene principalmente impiegato nell'ambito dell'*Information Retrieval (IR)*.

Tra gli schemi più usati quello *frequentista* assume un ruolo rilevante (Bolasco, 2005). Attraverso questo sistema si crea un *vocabolario del corpus*, ovvero le forme grafiche diventano un elenco di parole a cui viene associato il valore di occorrenza; una vera e propria frequenza assoluta in termini statistici, quindi il "conteggio" delle parole rappresenta la distribuzione statistica dei termini.

Il numero di termini presenti nella raccolta di testi presa in esame definisce l'ampiezza del vocabolario di un corpus, denominata V :

$$V = V_1 + \dots + V_k \dots + V_{fmax}$$

dove V_1 identifica il numero di termini che si presentano una sola volta all'interno della collezione di testi ed è detto *hapax*, V_k è il numero di parole che si presentano k volte ed infine V_{fmax} rappresenta la frequenza della parola con il maggior numero di occorrenze nel vocabolario.

Le dimensioni che un vocabolario può raggiungere sono notevoli e non sempre facilmente maneggiabili; ecco perché le procedure di *pre-trattamento* mirano anche a ridurre la variabilità presente in un *corpus*.

Dalla realizzazione del *vocabolario* si può determinare la matrice *lessicale del corpus*.

La codifica permette di considerare i testi come vettori, i quali vengono giustapposti nella costruzione di una *tabella lessicale*, in cui si avranno in riga i testi ed in colonna i termini (nel caso di una tabella lessicale *termini x documenti*). Nella rispettiva cella si avrà il valore di occorrenza di ogni termine per ogni testo (o documento). Dall'analisi paradigmatica del testo ottenuta attraverso lo *step lessicale*, è possibile adesso realizzare la rappresentazione sintagmatica, mediante una fase di analisi *testuale* in senso stretto.

L'ambiguità di senso che è insita nel linguaggio naturale può risolversi attraverso i metodi e le tecniche dell'analisi multidimensionale; la quale produce visualizzazioni grafiche capaci di definire strutture di senso sulla base della vicinanza o distanza, somiglianza o contrapposizione delle forme grafiche.

2.6. L'analisi multidimensionale

I fondamenti su cui si basa la statistica testuale rappresentano il modello di riferimento per l'analisi di una raccolta omogenea di testi. Le fasi che concretizzano l'analisi statistica dei dati testuali sono caratterizzate da una certa discrezionalità conferita al ricercatore; ovvero è l'analista ad identificare quali siano le operazioni di *pre-processamento* più opportune da compiere, gli schemi di ponderazione da adottare e quali, gli strumenti più adatti all'interpretazione dei risultati. L'aspetto decisionale connotato di una certa soggettività nel processo di analisi automatica dei testi si trasforma non solo in

base agli obiettivi perseguiti ma anche in relazione all'oggetto di indagine. Un *corpus* come oggetto di studio, in base ai criteri che lo classificano e diversificano, può determinare una serie di scelte tecniche e metodologiche ed una successione di complessità del tutto rilevanti ai fini dell'analisi.

Per la visualizzazione dei dati testuali si fa principalmente ricorso a tecniche di riduzione dimensionale, dando particolare rilevanza all'individuazione di sottospazi di proiezione ottimale delle relazioni esistenti tra le unità linguistiche e i documenti (*metodi fattoriali*) o determinando classi di equivalenza mediante il supporto di strutture ad albero e grafi (*metodi di classificazione*).

Le tecniche di riduzione dimensionale maggiormente impiegate per i dati testuali sono l'*analisi delle corrispondenze* (Benzécri, 1973), il *singular value decomposition* (SVD; Eckart e Young, 1935), il *latent semantic index* (Dumais *et al.*, 1988) la *cluster analysis* (Tryon, 1939), il *multidimensional scaling* (Torgerson, 1952).

I metodi fattoriali hanno lo scopo di ridurre lo spazio dimensionale costituito dai termini del vocabolario, attraverso la creazione di nuove variabili sintetiche (Bolasco *et al.*, 2004) che sono capaci di determinare i principali assi semantici su cui si sviluppano i contenuti del testo.

Una delle tecniche fattoriali utilizzate per l'analisi dei dati testuali è l'*analisi delle corrispondenze* (AC; Benzécri, 1973), che a partire dalla matrice lessicale restituisce la relazione esistente tra i termini, tra i documenti e, indirettamente, tra termini e documenti,

sia da un punto di vista geometrico che algebrico. Se l'AC è finalizzata a studiare solo le relazioni tra i profili lessicali, in base alla vicinanza o distanza delle unità linguistiche sul piano, è possibile ricostruire raggruppamenti tematici in grado di rappresentare il senso globale di un testo.

Il *singular value decomposition* (SVD; Eckart e Young, 1935), ovvero la *decomposizione in valori singolari* è una particolare tecnica di fattorizzazione di una matrice basata sull'uso di autovalori e autovettori. Essa si propone di risolvere il problema della ricerca di un sottospazio di riferimento, individuando una struttura di associazione tra le variabili osservate. L'SVD trova spazio tra le tecniche di analisi multidimensionale dei dati, soprattutto nelle strategie di *Information Retrieval* (IR).

Sempre connesso al dominio dell'IR, si individua il *latent semantic index* (LSI; Dumais *et al.*, 1988). Questo metodo individua i documenti rilevanti a seguito di una *query* (richiesta da parte di un utente), determinando uno spazio concettuale definito "sintetico" all'interno di una rappresentazione vettoriale di

documenti. L' *LSI* stima la frequenza con la quale determinati termini vengono usati insieme, restituendo così le associazioni semantiche latenti tra le parole e i documenti.

Il *multidimensional scaling* (Torgerson, 1952) è una tecnica di analisi multidimensionale che, se viene utilizzata nell'ambito dell'analisi dei dati testuali, serve a reperire informazioni semantiche dai contesti. Attraverso la posizione che le unità lessicali o i documenti assumono in un dato spazio dimensionale, è possibile determinare, sulla base della vicinanza o distanza, un principio di similarità o dissimilarità.

Tra i metodi di classificazione multidimensionale, la *cluster analysis* (Tryon, 1939) è un metodo statistico che consente di individuare, all'interno di un ampio collettivo, categorie o gruppi (*cluster*) di unità statistiche, secondo un principio di omogeneità e similarità interna; con lo scopo di descriverne alcune caratteristiche. Nell'ambito dei dati testuali, ad esempio, raggruppa classi di parole o di testi caratterizzati da una forte omogeneità *intra-cluster*, concepita in termini di distanza in un dato spazio multidimensionale (Bolasco, 1999; Vichi e Optiz, 1999).

2.7. Gli studi *cross-language* dei dati testuali: principali metodi e tecniche

Per l'analisi dei dati testuali si fa dunque riferimento alle definizioni, ai criteri e alle fasi di elaborazione che caratterizzano la statistica testuale. Si ricorre a metodi e tecniche di analisi multidimensionale, prediligendo uno strumento di analisi piuttosto che un altro, in base al tipo di testo oggetto di studio ed in relazione agli obiettivi perseguiti. Si analizzano fonti testuali eterogenee, come opere letterarie, contenuti multimediali, corpora monolingue o multilingue; ed in base al *corpus* si decide di ricorrere ad un metodo o alla combinazione di più metodi allo stesso tempo. In particolare, se lo studio riguarda risorse di testo multilingue (*cross-language analysis*), scegliere una strategia di analisi diventa particolarmente ambizioso, soprattutto in un'era di digitalizzazione della comunicazione e di globalizzazione culturale. Per essere considerato multilingue, un sistema deve essere in grado di gestire testi espressi in più lingue, allo stesso tempo ed allo stesso modo; producendo buoni risultati per ciascuna delle lingue coinvolte nello studio (Litvak e Vanetik, 2019).

In letteratura si annoverano differenti tecniche di elaborazione per un'analisi *cross-language*:

- **Topic Modeling**: nell'ambito dell'apprendimento automatico e dell'elaborazione del linguaggio naturale, la definizione di *Topic Modeling* riconducibile al critico e storico Franco Moretti (Ciotti, 2017) indica un modello statistico finalizzato ad individuare gli "argomenti" (o *topics*) su cui si strutturano i testi. Più specificamente, gli argomenti sono costituiti da una distribuzione di parole maggiormente frequenti che definiscono la semantica racchiusa in un testo.

Un esempio di *Topic Modeling* usato per l'analisi dei dati testuali è la *Latent Dirichlet Allocation*, (*LDA*; Blei *et al.*, 2003); il cui modello statistico individua un *set* di argomenti latenti, per ognuno dei quali associa una sequenza di parole in termini di distribuzione multinomiale. L'*LDA* in un primo momento determina l'argomento o gli argomenti centrali di un *corpus* e successivamente, definisce la distribuzione di parole che ruotano attorno al *topic* identificato.

Chen *et al.* (2017) usano l'*LDA* per mappare l'opinione pubblica *online* e rilevare gli argomenti denominati "caldi" in un *corpus* di lingua inglese-cinese. La strategia proposta si concretizza con l'applicazione dell'*LDA* su un *corpus* misto inglese-cinese, al fine di rilevare un insieme di argomenti co-occorrenti. Successivamente, attraverso la similarità del coseno sulla distribuzione dei *topic* per ogni testo, gli autori stimano la distanza tra due argomenti. Pruss *et al.* (2019) intendono individuare, per mezzo dell'*LDA*, gli argomenti chiave che ruotano attorno ad un *set* di dati in lingua inglese, spagnola e portoghese. Lo studio consiste nell'analisi dello *speech online*, su Twitter, della discussione relativa allo scoppio, nel 2015, in America meridionale e settentrionale, del virus *Zika*, associato a problemi teratogeni e sindromi neurologiche. L'applicazione dell'*LDA* su corpora multilingue non paralleli, come dimostrato in questo studio, evidenzia una riduzione della sua efficacia richiamando la necessità di effettuare un lavoro di traduzione, seppure parziale. La ridotta efficacia dell'*LDA* in un contesto multilingue dipende da una mancanza di sovrapposizione tra le parole delle diverse lingue, in quanto ogni *topic* avrà una distribuzione monolingue specifica. Per ovviare a questo problema, Pruss *et al.* (2019) impiegano un modello di *Topic Modeling*

multilingue, chiamato *Polylingual Topic Model (PLTM; Mimno et al., 2009)*. Infatti, Mimno *et al.* (2009) propongono per primi il modello *PLTM*, la cui applicazione è riservata a corpora paralleli o comparabili. La *PLTM* è un'estensione dell'*LDA* che tuttavia si caratterizza per alcuni tratti distintivi: mentre l'*LDA* presuppone che ogni documento abbia una propria distribuzione di argomenti e parole ad essi connessi, il modello *PLTM* ritiene che i testi condividano la stessa distribuzione specifica sugli argomenti e si assume che ciascun *topic* sia costituito da un insieme di distribuzioni discrete.

Un'altra tecnica di *Topic Modeling* è la *Latent Semantic Analysis (LSA; Dumais et al., 1988)*, usata nel *Natural Language Processing (NLP)* per definire il contenuto di un documento. Si ritiene che la vicinanza di un testo ad un particolare argomento derivi dal presupposto che ogni parola assuma un determinato concetto.

Bader e Chew (2008) ricorrono all'*LSA* per studiare cinque lingue diverse, quali arabo, inglese, cinese, russo e spagnolo. Per ridurre la dimensionalità dei dati, gli autori adottano tecniche di decomposizione matriciale come il metodo degli autovalori o *eigenvalue decomposition (EVD; Mcwhirter et al., 2007)*. Tale metodo arricchisce l'*LSA* e consente di incorporare informazioni sugli allineamenti dei termini nelle diverse lingue, utilizzando i sistemi della *Statistical Machine Translation (SMT)*. L'allineamento serve a migliorare l'analisi semantica e a trasformare i vettori in matrici *termine per concetto*. Ratna *et al.* (2018) utilizzano l'*LSA* per verificare casi di plagio nelle pubblicazioni scientifiche. Più propriamente, implementano un algoritmo che allinea documenti di *test* in lingua indonesiana e documenti di riferimento in lingua inglese, progettato dal Dipartimento di Ingegneria Elettronica dell'Università dell'Indonesia. Attraverso l'*LSA* si intendono analizzare le relazioni esistenti tra un *set* di documenti e un insieme di termini, generando così un *set* di concetti, connessi a documenti e termini.

- **Reti neurali:** o *Artificial Neural Networks* (Hebb, 1949) rappresentano modelli matematici che, ispirati al funzionamento biologico della corteccia celebrale umana, si configurano attraverso neuroni artificiali. Nello specifico, in ambito multilingue, Lee e Yang (2003) applicano un approccio di rete definito *Self-Organizing Maps* (Kohonen *et al.*, 2000).

Le *Self-Organizing Maps (SOM)* identificano una soluzione di reti neurali per l'apprendimento non supervisionato. Esse consentono di determinare, in uno spazio euclideo, la similitudine o differenza tra corpora multilingue inglese-cinesi. Inoltre, ascrivibile all'applicazione delle *SOM* è il lavoro di ricerca presentato da Al-Marghilani *et al.* (2008), con lo scopo di raggruppare i termini che appartengono ad una stessa lingua partendo da un *corpus* bilingue inglese-arabo. L'arabo, rispetto all'inglese, è una lingua altamente derivata e flessa, che richiede un'adeguata analisi morfologica per un'efficace estrazione e processamento del testo. Quindi, l'applicazione delle *SOM* si traduce in un tentativo di sviluppare uno strumento utile nell'analisi di un *corpus* bilingue combinato. A conoscenza degli autori non vi è alcuna letteratura significativa riguardo al *mining SOM* per dati testuali *cross-language* di lingua araba (Al-Marghilani *et al.*,2008).

Tra i modelli di rete neurale presenti in letteratura, per l'analisi dei dati testuali si ricorre spesso alle reti neurali ricorrenti (*RNNs*; Elman, 1990) e nello specifico a quelli con lunga memoria a breve termine o *Long Short-Term Memory (LSTM*; Hochreiter e Schmidhuber, 1997). Le *RNNs* si identificano come una forma di apprendimento non supervisionato, che nell'ambito dell'*NLP* hanno una funzione predittiva (*predictive learning*) ampiamente utile nel campo del *deep learning* (*Ibidem*), dell'*handwriting* e *speech recognition*. Li *et al.* (2018) ricorrono all'*LSTM* per il *knowledge transfer* tra una lingua d'origine ed una di destinazione, con l'obiettivo di calcolare la similarità semantica bilingue sia per testi di breve che di lunga estensione.

Sempre restando nel campo applicativo delle reti neurali, Plahl *et al.* (2011) ne fanno ricorso per indagare le *features* probabilistiche di corpora multilingue, valutando il livello di affinità tra una lingua di *training* ed una di *testing*. Gli autori stimano il *word error rate (WER)* e la quantità di dati utilizzati durante la fase di addestramento. Ogniquale volta che l'ammontare dei dati per la fase di *training neurale* è uguale, si ha una relazione molto simile nei risultati tra la lingua di *training* e quella di *testing*; mentre se aumentano i dati di addestramento, la relazione diminuisce.

- **Word embedding:** identifica un insieme di tecniche che memorizzano le informazioni sia semantiche che sintattiche del vocabolario di un

corpus, le quali vengono mappate partendo da una collezione di testi non annotata (Turian *et al.*, 2010; Mikolov *et al.*, 2013). Secondo l'ipotesi della semantica distribuzionale, si costruisce uno spazio vettoriale, in cui i vettori delle parole che presentano una vicinanza nello spazio, mostrano di occorrere negli stessi contesti linguistici, riconoscendosi come semanticamente più simili. L'utilizzo del *word embedding* diventa particolarmente interessante nel tentativo di proiettare lingue diverse nello stesso spazio vettoriale, così da permettere il trasferimento di quanto appreso da una lingua origine ad una di destinazione.

Mikolov *et al.* (2013) sono stati i primi a notare che gli spazi di *word embedding* mostrano strutture simili in tutte le lingue, anche nei casi di studio bilingue in cui non è presente l'inglese, notoriamente diffuso per gli studi dei dati testuali. In questo contributo si utilizza un vocabolario parallelo inglese-vietnamita come punti di riferimento per l'apprendimento di una mappatura. Tra gli approcci relativi al *word embedding* lo spazio vettoriale multilingue può anche essere mappato attraverso metodi di trasformazione ortogonale (Xing *et al.*, 2015; Artetxe *et al.*, 2016; 2017; Zhang *et al.*, 2016; Smith *et al.*, 2017). Nella fattispecie, Smith *et al.* (2017) confrontano due lingue, quali inglese ed italiano attraverso una trasformazione di rotazione tra spazi vettoriali basata sull'*SVD*. La proposta introduce un nuovo "*inverted softmax*" per identificare coppie di traduzione che tengano conto anche di espressioni rare sia in lingua inglese che in lingua italiana. Nella tecnica non si richiede un dizionario di addestramento ma si costruisce, tramite l'*SVD*, un *corpus* bilingue di testo allineato su vettori di frasi che sommano e normalizzano vettori di parole. Artetxe *et al.* (2017) evidenziano l'indisponibilità di molti corpora bilingue finalizzati all'apprendimento *word embeddings*. Al fine di ridurre la necessità di risorse bilingue, gli autori realizzano un approccio di autoapprendimento; ossia conducono uno studio basato sulla somiglianza strutturale degli spazi di *embedding*. Questa strategia funziona con tecniche *dictionary-based* e ne dimostrano la validità empirica con un dizionario bilingue molto limitato. Hermann e Blunsom (2014) utilizzano la tecnica del *bilingual word embedding* per l'apprendimento delle rappresentazioni semantiche ed estendono l'ipotesi distribuzionale a dati multilingue, in uno spazio condiviso

chiamato *joint-space embeddings*. Il modello presentato si serve di dati paralleli per allineare gli *embeddings* delle frasi semanticamente equivalenti, seppur mantenendo una distanza sufficiente tra quelle frasi che si mostrano dissimili tra loro. L'approccio viene, inoltre, esteso anche a livello di documento e non solo di singole frasi. Il *word embedding* multilingue può essere impiegato anche per la traduzione automatica (Zou *et al.*, 2013), il collegamento tra le entità multilingue (Tsai e Roth, 2016), il *tagging* di parti del discorso (Gouws e Søgaard, 2015; Zhang *et al.*, 2016) e la classificazione dei documenti (Klementiev *et al.*, 2012). Conneau *et al.* (2018) sperimentano il *word embedding* per la costruzione di un dizionario bilingue attraverso un allineamento delle parole ed incorporano gli spazi vettoriali in modo non supervisionato. Il modello sperimenta coppie di lingue con scarse risorse linguistiche ed una limitata quantità di dati paralleli (Mihalcea *et al.*, 2007; Balahur, e Turchi, 2014; Barnes *et al.*, 2018), aprendo così la strada ai sistemi di traduzione automatica non supervisionata per dati *cross-language* (Lample *et al.*, 2018; Artetxe *et al.*, 2018). Tuttavia, Søgaard *et al.* (2018) identificano dei limiti in relazione ad un modello basato sulla traduzione automatica non supervisionata, poiché è poco efficace per un linguaggio ricco di risorse che morfologicamente non dipende da un'annotazione del *corpus*. Essi suggeriscono infatti che la supervisione delinea, invece, una robustezza statistica quasi perfetta al bilinguismo. Al contrario, altri autori (Barone, 2016; Zhang *et al.*, 2017; Lample *et al.*, 2018) sostengono che gli *embeddings* multilingue senza supervisione, in assenza di dizionari e traduzioni, siano efficaci per l'induzione dei lessici bilingui e per i modelli di *machine translation*.

- ***Machine translation***: Il supporto dei sistemi di traduzione automatica (Lûdskanov, 1967) o *machine translation (MT)* identificano un potente alleato per le pratiche di traduzione; ovvero sono capaci di trasferire informazioni tra le lingue anche dove un linguaggio è piuttosto ricco di risorse mentre l'altro ne è povero (Mihalcea *et al.*, 2007). Uno studio orientato ad indagare la soggettività di una lingua sfruttando le risorse e gli strumenti disponibili di un'altra lingua è quello di Mihalcea *et al.* (2007), che utilizza l'inglese come "ponte" per mezzo dei sistemi di traduzione automatica. Più specificamente, l'obiettivo consiste nello sviluppo di un classificatore statistico capace di trasferire la soggettività

dall'inglese al rumeno: in un primo momento attraverso un lessico di soggettività tradotto dall'inglese e successivamente, propone l'utilizzo di un *corpus* parallelo con lo scopo di generare un *training set* di dati nella lingua di destinazione. Lo studio della soggettività attraverso i sistemi di traduzione automatica viene affrontato anche da Banea *et al.* (2008) che sfruttano le risorse disponibili per l'inglese per l'analisi del rumeno e dello spagnolo. La strategia si fonda sull'impiego di corpora annotati e viene applicata dapprima al rumeno, poiché rappresenta una lingua che dispone di poche risorse linguistiche e in seguito allo spagnolo, per convalidarne la tecnica. Gouws *et al.* (2015) introducono un modello di traduzione automatica *BilBOWA* (*Bilingual Bag-of-Words without Alignments*) semplice ed efficiente dal punto di vista computazionale, che consente l'apprendimento di dati bilingue riducibili a grandi *set* monolingue senza richiedere *parallel corpora*.

- **Annotazione di corpora:** rappresenta una tecnica di *labeling* morfologico, sintattico e semantico di un *corpus*, attraverso l'aggiunta di informazioni o etichette (*tagging* o *mark-up*) ad una o più porzioni di testo (Chiari, 2004). L'annotazione maggiormente utilizzata è la morfo-sintattica (POS), che associa ad ogni parola la corrispondente categoria grammaticale. La procedura può essere condotta manualmente, con sistemi automatici o semi-automatici (*Ibidem*).

Ehrmann *et al.* (2011) sviluppano un sistema detto *multilingual named entity recognition (NER)* per affrontare il problema della comparabilità automatica tra le lingue, che si risolve per mezzo dell'annotazione dei corpora. Wang *et al.* (2014) utilizzano l'annotazione dei corpora per l'apprendimento multilingue, sviluppando un modello basato su annotazioni *bitext* su *set* di dati cinese-inglese e tedesco-inglese. Padò e Lapata (2009) evidenziano invece la differenza tra l'annotazione manuale e quella automatica, attraverso lo studio semantico di un *corpus* parallelo inglese-tedesco. Tufis (2007) attraverso il *tagging* di una parte del discorso contrassegna le parole presenti nel *corpus*, come corrispondenti a specifici frammenti o porzioni di testo. L'annotazione determina un allineamento di frasi, capace di generare corpora paralleli. Una volta identificati gli equivalenti di traduzione e quindi i *parallel corpora*, è ragionevole aspettarsi che le parole della traduzione bilingue

condividano almeno un significato concettuale memorizzato in un “inventario” di senso interlinguistico.

2.8. Cenni conclusivi

A conclusione del secondo capitolo, lo studio specifico sul legame tra l'uso di un registro linguistico e la parallela costruzione culturale ha conferito importanza all'analisi del linguaggio naturale. In particolare, ha assunto rilevanza lo studio dell'analisi testuale in un contesto multilingue. Dalla letteratura passata in rassegna è emersa una diversificazione delle scelte strategiche, in termini tecnico-metodologici, evidenziando nell'ambito comunità scientifica un dominio di ricerca estremamente competitivo e in attesa di costanti implementazioni. Nello specifico, le analisi *cross-language* rappresentano un campo di indagine caratterizzato da evidenti limiti. Allo stato attuale infatti si individua una profonda carenza di risorse sintattico-morfologiche in grado di coprire la notevole vastità linguistica. Quindi, le finalità conoscitive di un ricercatore sono spesso influenzate dalla disponibilità degli strumenti linguistici e questo comporta che molti contesti e fenomeni rimangano inesplorati, a causa della carenza o mancanza strumentale. Essenzialmente, dalla letteratura è merso che la maggior parte degli studi multilingue sono basati su corpora paralleli, ossia testi i cui contenuti sono espressi in lingue diverse ma di equivalente traduzione. L'opportunità di ottenere corpora paralleli non sempre rappresenta una condizione possibile per tutte le varie combinazioni linguistiche e una parte di produzione scientifica, seppure meno corposa, evidenzia l'opportunità di indagare corpora comparabili, i quali conservano la propria struttura monolingue ma afferiscono ad uno stesso dominio tematico.

Dal momento che la maggior parte degli studi *cross-language* si basa su corpora paralleli, la letteratura ha evidenziato il ricorso ad operazioni di traduzione, condotte sia manualmente che automaticamente attraverso sistemi di *machine translation*. Nei processi di traduzione, la lingua inglese è spesso usata come “ponte” per lo studio delle altre lingue, la quale tuttavia tende a “manipolare” la naturalezza del linguaggio, producendo una conseguente artificiosità anche dal punto di vista semantico. Inoltre, vi sono lingue talmente povere di risorse che non riuscirebbero con nessun sistema di traduzione

manuale o automatico, a coprire la vastità linguistica di un determinato linguaggio naturale. Considerate tali complessità, gli studiosi delle analisi multilingue avanzano prevalentemente proposte metodologiche per casi studio che indagano al massimo due lingue alla volta, ecco perché in letteratura l'espressione *cross-language* è spesso usata come sinonimo di analisi bilingue. Questo aspetto risulta essere particolarmente limitante, poiché un'analisi bilingue può essere utile per verificare ad esempio la corrispondenza linguistica di opere letterarie tradotte ma se l'interesse di studio si focalizza sull'analisi dei contenuti multimediali, risulta inadatta una prospettiva di ricerca basata sui soli processi di traduzione.

I principali metodi e tecniche presentate dalla letteratura *cross-language* tendono a classificare le distribuzioni lessicali multilingue in raggruppamenti monolingue, talvolta di natura tematica. In sintesi, dallo stato dell'arte sugli studi *cross-language* emerge una prevalente produzione scientifica interessata a testi bilingue paralleli di traduzione automatica. Pertanto, con l'obiettivo di analizzare lo *speech online* multilingue relativo al fenomeno della violenza contro le donne, senza ricorrere a sistemi di *machine translation*, o all'inglese come *testing e/o parallel corpora*, nel capitolo successivo si vuole presentare un'alternativa strategica che si concretizzi nell'analisi sincronica di corpora *cross-language*. La proposta mira a presentare sullo stesso piano e allo stesso tempo il passaggio dalla sintattica monolingue alla semantica multilingue, valutando le dirette e conseguenti implicazioni socio-culturali.

Capitolo Terzo

Una strategia Statistica per l'Analisi Multilingue Comparativa

Il terzo capitolo delinea da un punto di vista teorico e con il supporto della letteratura afferente, una strategia di analisi *cross-language* comparativa per l'analisi dei linguaggi e delle culture differenti sul tema della violenza contro le donne. L'obiettivo finale è quello di evidenziare eventuali similarità e/o dissimilarità tra i registri linguistici, supponendo l'esistenza del binomio lingua-cultura. Affinché si possa cogliere la percezione del fenomeno tra diversi contesti socio-culturali con l'analisi dei differenti linguaggi si è ritenuto opportuno approcciarsi al dominio della comunicazione digitale. Infatti, l'analisi dei dati estratti dai Social Media rappresenta attualmente una delle strategie maggiormente impiegate da diverse aree di studio, al fine di qualificare i fenomeni e mappare l'opinione pubblica. Nell'ultimo decennio è sempre più rilevante lo studio sul linguaggio della violenza propinato attraverso il *Web*, con una particolare enfasi su forme di linguaggio offensivo e discriminatorio, come ad esempio *l'online hate speech contro le donne*. Più propriamente lo studio si concretizza nell'analisi dell'opinione pubblica su *Twitter*, la cui piattaforma, rispetto ad altri Social Media, alcuni dei quali sicuramente più popolari, si distingue per le sue caratteristiche di apertura ed orizzontalità che gli consentono di tradursi in un valido strumento di ricerca per numerosi temi di discussione che animano la rete, grazie al potere *influenzativo* della *comunicazione partecipativa*.

Tenendo conto della struttura dei dati e delle peculiarità proprie delle diverse lingue, in questo contesto si espongono gli aspetti metodologici per il trattamento ed il confronto dei dati testuali multilingue. Il primo passo della proposta strategica risiede nelle operazioni di *pre-processing* e *feature extraction* di ciascun *corpus* monolingue, ai fini di un'analisi esplorativa preliminare che consente di mappare il fenomeno per ogni lingua. Con lo scopo di tutelare le caratteristiche originali di ciascun registro linguistico, in questo contesto si propone di determinare un'unità di analisi alternativa alla *forma grafica* e considerata più opportuna per il successivo confronto multilingue, ovvero la definizione di raggruppamenti concettuali. La determinazione dei concetti introduce quindi la seconda fase della strategia *cross-language*. Nello specifico, la *Concepts Extraction* si realizza mediante la combinazione di due

tecniche di analisi multidimensionale: l'*Analisi delle Corrispondenze* e la *Cluster Analysis*, attraverso cui si individuano gruppi di concetti monolingue classificando in *cluster* le diverse *features* (termini estratti in uno spazio vettoriale multidimensionale). Tuttavia, per ottenere tali raggruppamenti si è inteso inizialmente assicurare l'orientamento semantico delle *features*, attraverso il calcolo della polarità semantica dell'*NRC (National Research Council Canada) Emotion Lexicon*. In questo modo, le classi di concetto che si ottengono identificano solo *features* positive e solo *features* negative. Determinato lo *spazio concettuale* delle diverse lingue in base all'appartenenza delle *features* a ciascun gruppo di concetto, si costruisce la matrice dei concetti positivi e negativi per ciascuna lingua. Tuttavia, per indagare come differiscono le relazioni tra i concetti delle diverse lingue, la strategia conclusiva consiste nell'applicazione dell'*Analisi Procustiana Generalizzata*, per confrontare simultaneamente le matrici *concetti x concetti* delle diverse lingue in un unico spazio dimensionale multilingue, analizzandone così l'accordo e valutando similarità o dissimilarità linguistico-concettuali e quindi socio-culturali.

3.1. Il database documentale dell'analisi statistica multilingue

La strategia di analisi multilingue comparativa di questa tesi viene presentata nell'ambito della comunicazione digitale e più propriamente nel dominio dei Social Media, intesi come l'insieme delle applicazioni e dei fondamenti ideologici e tecnologici del *Web 2.0* (Kaplan e Haenlein, 2010). La crescita esponenziale di piattaforme *online* attraverso cui condividere notizie ed opinioni influenza notevolmente il modo di percepire i diversi aspetti della vita di ogni giorno (Perrin, 2015), come una rivoluzione trasformativa della società (Mondal *et al.*, 2017). In particolare, la diffusione *online* di notizie ed opinioni permette di creare un *pensiero comune* su uno specifico tema che Ceron *et al.* (2014) definiscono come un "cervello collettivo", denotando anche e soprattutto una cultura dell'odio verso specifiche categorie di persone. Nell'ultimo decennio infatti è diventato sempre più rilevante lo studio sul linguaggio della violenza propinato attraverso il *Web*, con una particolare enfasi su forme di linguaggio offensivo e discriminatorio (Malmasi e Zampieri, 2018), come per l'appunto l'*online hate speech* contro le donne. La proliferazione dei discorsi d'odio *online* proposti su *blog*, *micro-blog* e *social network* è

l'espressione più autentica dei conflitti e delle tensioni che si instaurano all'interno della società e che mirano ad attaccare categorie di persone basandosi su razza, religione, etnia, orientamento sessuale e genere (Johnson *et al.*, 2019). Sono infatti numerosi gli studi che di recente si sono occupati di indagare attraverso il *Web* il linguaggio della violenza, molti dei quali finalizzati all'analisi del *cyberbullismo* (Xu *et al.*, 2012), dell'incitamento all'odio (Burnap e Williams, 2015) e della rilevazione del razzismo (Tulkens *et al.*, 2016). Quindi, con lo scopo di qualificare i fenomeni (Fish *et al.*, 2011) e di mappare l'opinione pubblica (Murphy *et al.*, 2014), l'analisi dei dati estratti dai Social Media rappresenta attualmente una delle strategie maggiormente impiegate da diverse aree di studio. Per quanto riguarda l'analisi della percezione pubblica, un ruolo chiave è svolto dalla piattaforma Twitter (De Stefano e Santelli, 2019) lanciata da Dorsey *et al.* nel 2006, che rappresenta il servizio di *micro-blogging* tra i più popolari e diffusi al mondo. Gli utenti di Twitter possono postare e condividere brevi messaggi di testo chiamati *tweets* con una lunghezza massima di 280 caratteri, i quali possono includere anche immagini, video, *link*, *URLs*. A seconda delle impostazioni sulla *privacy*, i contenuti sono visibili a tutti o solo a determinati contatti appartenenti alla propria rete. Possono essere etichettati mediante l'uso di uno o più *hashtag* (#) seguiti da una o più combinazioni di parole, generando così un collegamento ipertestuale che permette di categorizzare i messaggi recenti che possiedono lo stesso *hashtag tematico*. L'aspetto cruciale che caratterizza la piattaforma nello studio dell'opinione pubblica è il flusso di circolazione delle informazioni che si sposta da un utente all'altro attraverso un'interazione condivisa (Onorati e Diaz, 2016). Rispetto ad altri Social Media, alcuni dei quali sicuramente più popolari, Twitter si distingue per le sue caratteristiche di apertura ed orizzontalità (Ceron *et al.*, 2014), che gli consentono di tradursi in un valido strumento di ricerca per numerosi temi di discussione che animano la rete. Difatti, i messaggi condivisi su Twitter vengono automaticamente mostrati nella pagina del profilo utente ed inoltrati a tutti coloro che si sono registrati per riceverli, ovvero i cosiddetti *followers*. I contenuti postati possono essere visualizzati anche senza l'autorizzazione dell'utente, ad eccezione di specifiche impostazioni sulla *privacy*. I *tweets* possono inoltre essere pubblicati sul profilo di altri utenti e rilanciati integralmente da un altro *user*, inviandoli sistematicamente a tutti i propri *followers*. In sintesi, Twitter consente, in tempo reale, una modalità di diffusione delle notizie come *giornalismo partecipativo* o come condivisione di idee, inclusa la disseminazione di una cultura basata

sull'odio. La piattaforma rappresenta infatti il Social Media più utilizzato per indagare il linguaggio dell'odio e della violenza, soprattutto per quel che concerne la violenza contro le donne, in quanto solo in Italia si stima, nel 2017, la produzione di oltre 1 miliardo di *tweets* sessisti rilevati su un campione di oltre 2 miliardi di messaggi (Vox- Osservatorio sui diritti della Presidenza della Camera dei Deputati, 2017). Un'analisi dello scenario digitale mondiale indica nel 2019 la presenza di 330 milioni di utenti Twitter attivi (Statista, 2019; Clement, 2019: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>), con una produzione di 500 milioni di *tweets* al giorno. Solo nel 2018 si contavano almeno 6000 *tweets* al secondo (Social Media Statistics 2018 for Business: blog.statusbrew.com/social-media-statistics-2018-for-business). L'80% degli utenti attivi utilizza la piattaforma da dispositivo mobile, con una propensione all'informazione giornalistica, alla condivisione di notizie e all'acquisto di prodotti (Twitter Internal Data, 2019: <https://business.twitter.com>). Gli *users* sono particolarmente interessati al mercato della compravendita e all'*advertising*: non a caso il 53% degli utenti attivi è pioniere di novità (Kantar, 2017 in Twitter Internal Data, 2019). L'*audience* della piattaforma presenta una profilatura utente prettamente maschile, in una fascia d'età compresa tra 13 e 50 anni e più, ovvero il 34% sono donne mentre il 66% sono uomini. Il 37% degli utenti ha un'età compresa tra 18 e 29 anni, mentre il 25% ha tra i 30 e 49 anni. L'80% degli utenti appartiene alla categoria dei *millennials* benestanti, i quali tendono ad influenzare percezioni, usi, costumi e tendenze d'acquisto. Il livello di istruzione è medio-alto, prevalentemente di tipo universitario. L'aspetto economico può definirsi agiato, poichè il reddito annuo è pari a \$ 50 000 e oltre, per il 56% dei fruitori (*Ibidem*).

Si tratta di individui che vivono perlopiù nelle zone urbane delle città e che prediligono Twitter rispetto ad altri Social Media soprattutto per le informazioni e le notizie riguardanti la cronaca e la politica³ (Walton, 2019: <https://www.investopedia.com/articles/markets/100215/twitter-vs-facebook-vs-instagram-who-target-audience.asp>).

³ Si registra un incremento di utenti Twitter durante la candidatura ed elezione del Presidente degli Stati Uniti d'America Donald Trump, il quale ha più di 60 milioni di *followers* (seguaci) che seguono con interesse i contenuti da lui postati (*tweets*; Walton, 2019).

Sulla base di tale profilatura utenti, scegliere di condurre un'indagine utilizzando i contenuti di Twitter presenta alcuni evidenti svantaggi, tra cui la mancanza della rappresentatività dei dati. Infatti, non si possono estendere le caratteristiche degli utenti all'intera popolazione ma vanno circoscritte a specifiche categorie socio-demografiche e socio-economiche, tenendo conto che non tutti gli utenti esprimono la propria opinione su tutti gli argomenti e non tutti i membri di una popolazione possiedono un *account* di Twitter. Pertanto, l'interesse nei confronti dei soli *users-generated content* produce una sottostima della totalità della popolazione, in quanto ad esempio alcune categorie di età o di livello di istruzione sono escluse (Wei e Hindman, 2011). Un altro aspetto rilevante è la parziale accessibilità dei dati, in quanto i processi di aggregazione o i regolamenti vigenti in accordo con le normative sulla tutela della *privacy* limitano l'opportunità di ottenere in maniera diretta le informazioni complete. Ai dati provenienti dai Social Media così come Twitter si può altresì imputare una ridotta attendibilità, perché spesso gli *account* sono falsi o duplicati.

Ciononostante, se si considera l'opportunità di estrarre dati che in altro modo non sarebbe possibile ottenere e se si tiene conto del potere *influenzativo* della *comunicazione partecipativa* dei Social Media, si può ritenere ragionevole utilizzare comunque tali informazioni, tenendo presente eventuali implicazioni e distorsioni durante l'interpretazione dei risultati. Tenuto quindi conto delle caratteristiche di Twitter e considerata la sua capacità di generare un flusso di comunicazione di ampia condivisione e diffusione, diversamente da altri Social Media, si reputa che la piattaforma possa contenere un elevato potenziale per l'analisi del linguaggio sulla violenza e l'incitamento all'odio contro le donne, in grado di svelare importanti scenari della cultura collettiva sul tema della violenza contro le donne. Conseguentemente inserirsi nell'ambito della comunicazione sui Social Media per indagare il tema oggetto di studio consente di ottenere in *real-time*, una moltitudine di commenti ed opinioni pubblicate *online* da utenti che vivono in diversi contesti e si esprimono tramite registri linguistici differenti. Lo studio della percezione pubblica internazionale sul tema della violenza contro le donne attraverso Twitter permette di cogliere gli aspetti caratteristici delle culture e dei linguaggi appartenenti a qualsiasi contesto geografico di interesse, anche quelli più difficili da raggiungere in altro modo; magari attraverso strumenti di indagine più tradizionali e diretti.

Pertanto, tenuto conto al contempo dei limiti connessi all'interpretazione dei risultati, il *database* documentale che si sottopone ad indagine nel presente

lavoro di ricerca è costituito dall'estrazione di più *set* di dati monolingue comparabili⁴ tra loro, quindi riferiti allo stesso oggetto di studio ma i cui registri linguistici sono diversi.

I dati estratti da Twitter sono di tipo multimediale e presentano una serie di informazioni strutturate, semi-strutturate e destrutturate, la cui natura diversificata permette di condurre diverse strategie di analisi. La prima dimensione che si ottiene dall'estrazione dei dati da Twitter è l'informazione testuale (*text*), ossia il corpo del messaggio, mentre un vettore logico TRUE o FALSE viene indicato per determinare se il *tweet* è tra i preferiti dagli utenti (*favorited*), seguito dal conteggio del numero di preferenze (*favoriteCount*). Viene mostrato il nome dell'utente che eventualmente riceve il messaggio (*replyToSN*) ed il relativo *id* identificativo (*replyToUID*), la data di creazione (*created*) del *tweet* (nel formato *aaaa-mm-gg* e *hh-mm-ss*) e le coordinate geografiche di latitudine e longitudine. Tra le altre dimensioni considerate vi sono: l'*username* (*screenName*) dell'utente che ha postato il *tweet*, il dispositivo con cui è stato creato (*statusSource*) lo *status*, se è stato ripubblicato da terzi (*retweeted*) ed il corrispondente numero di volte (*retweetCount*).

La sezione testuale del *database* rappresenta la collezione di *tweets* o *corpus* su cui è possibile condurre l'analisi testuale e portare quindi alla luce la percezione pubblica dei singoli contesti monolingue per poi confrontarli successivamente in uno spazio *cross-language*. Tuttavia, poiché il processo di *Text Mining* è da Hearst (1999) definito come l'estrazione di informazione e conoscenza interessante e non banale da un'informazione non strutturata, in questo contesto, si richiede un accurato lavoro di preparazione del testo prima di sottoporlo alle procedure e ad operazioni tipiche della statistica testuale. In particolare, nell'ambito dell'analisi testuale di corpora multilingue risulta indispensabile un'attenta procedura di pre-elaborazione dei *tweets* che tenga conto della natura dei dati e delle peculiarità proprie di ogni lingua.

3.2. Le scelte strategiche della proposta *cross-language*

Prendendo in considerazione le diversità linguistiche e gli aspetti annessi alle procedure di analisi testuale, la proposta strategica *multilingue* si concretizza

⁴ La comparabilità dei corpora si rimanda al capitolo secondo: § 2.2.

dapprima attraverso un'analisi esplorativa in uno spazio monolingue. In questa prima fase lo scopo è di descrivere il vocabolario di ciascun *corpus*, per evidenziare gli aspetti più peculiari di ogni lingua in termini di frequenze.

Ciascun *corpus* monolingue viene suddiviso in tre sotto-corpora, il primo contenente *tweets* e *retweets* così da ottenere la percezione pubblica globalmente diffusa sul tema della violenza contro le donne; il secondo è costituito dai soli *hashtag* al fine di esplorare le parole o gli argomenti utilizzati come etichette di contenuto; ed infine l'ultima analisi si focalizza sullo studio dei soli *retweets*, per comprendere quali parole fungono da *pensiero collettivo* maggiormente *condiviso* e rilanciato. Per passare da uno spazio di analisi monolingue ad uno multilingue, rispondendo all'obiettivo di poter confrontare i diversi corpora monolingue in un unico spazio dimensionale, la strategia mira a ridurre la variabilità linguistica attraverso la definizione di un'unità di analisi di ordine concettuale. Più propriamente i termini estratti da ciascun *corpus* monolingue vengono classificati in gruppi di concetti. In un primo momento il *corpus* di ciascuna lingua viene suddiviso in base al calcolo della polarità semantica per assicurare che i raggruppamenti concettuali racchiudano al proprio interno parole con lo stesso orientamento semantico. Successivamente, i concetti di ogni lingua si costruiscono combinando due tecniche di analisi multivariata: l'*Analisi delle Corrispondenze* e la *Cluster Analysis*, tipicamente impiegate per la riduzione dimensionale dei dati testuali.

L'applicazione congiunta di queste due tecniche presenta il vantaggio di cogliere l'accezione interna ai vari corpora riscontrando possibili ed eventuali *modelli di senso* del contenuto di ogni testo (Bolasco, 2005), attraverso la posizione attribuita dalle coordinate fattoriali ai vari termini. Mantenendo poi inalterata la posizione datagli dalle coordinate fattoriali, si procede con l'operazione di classificazione. Una volta identificati i gruppi concettuali per categoria semantica di ogni lingua, la fase finale dell'analisi si concretizza nella comparazione simultanea di tali unità concettuali mediante l'applicazione di una tecnica di studio multivariata nota come *Analisi di Procuste*. Si predilige la *performance* di tale tecnica poiché da un punto di vista metrico i dati originari possono essere comparati in modo simultaneo e senza ricorrere ad operazioni di trasformazioni di algebra lineare, restituendo altresì una configurazione media del consenso o accordo tra le lingue prese in esame.

In sintesi, dunque, la proposta di analisi *cross-language* consiste in tre diversi e consequenziali *step*:

1. Analisi esplorativa monolingue del fenomeno;
2. Definizione di concetti monolingue;
3. Confronto tra concetti multilingue in un'unica rappresentazione comparativa.

3.2.1. *Analisi esplorativa monolingue del fenomeno indagato*

L'*Exploratory Data Analysis (EDA)* si riferisce ai metodi descrittivi utilizzati per rappresentare *set* di dati multivariati. Generalmente, nelle analisi statistiche lo studio esplorativo rappresenta l'obiettivo principale o lo studio preliminare che anticipa un determinato modello di analisi (Husson *et al.*, 2010).

Nel caso di fattispecie, l'analisi esplorativa viene condotta su ciascuna collezione di *tweets* monolingue per mappare il fenomeno da ogni prospettiva linguistica, con lo scopo di cogliere gli aspetti più peculiari delineati dallo studio del linguaggio di ciascuna lingua sul tema preso in esame.

L'analisi automatica di ciascun *corpus* monolingue segue le fasi delineate da Bolasco (2005), che prevedono fasi di *pre-trattamento* e *feature extraction*.

3.2.1.1. *Le fasi di pre-processing e feature extraction di ciascun corpus monolingue*

La prima fase della strategia si concretizza nello studio esplorativo e preliminare di ciascuna collezione di *tweets* monolingue, in modo separato rispetto all'intero *database* multilingue. La raccolta dei *tweets* di ogni lingua costituisce un diverso *corpus* su cui realizzare l'analisi del testo. Tuttavia, prima della fase di analisi vera e propria, la pre-elaborazione dei testi richiede scelte che possono influenzare l'accuratezza, la validità e i risultati (Crone *et al.*, 2006; Günther e Quandt, 2016). Conseguentemente, la preparazione di ciascun *corpus* monolingue che precede la fase di *feature extraction* si realizza attraverso specifiche operazioni di *pre-processing*.

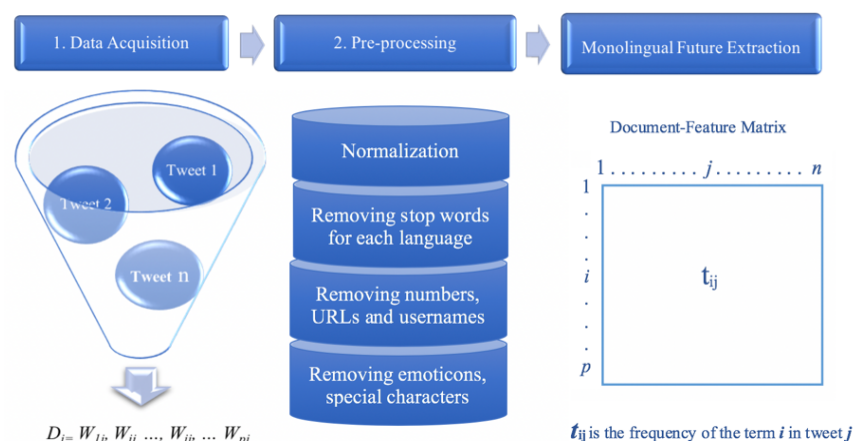


Figura 3.1. Il processo dell'analisi lessico-testuale di un corpus
Fonte: Elaborazione propria

Il principale problema dei dati estratti dal *Web* è la presenza di elementi che creano rumore all'informazione sottostante, come *link* e *URLs*, simboli quali l'*RT* dei *retweets*, il cancelletto (#) tipico dell'*hashtag* o la chiocciola (@) che accompagna il nome dell'utente che ha postato o *ritwittato* un messaggio. Nel corpo del testo possono essere incluse le estensioni di dominio dei siti come ".it", ".com", ".org", ".uk", *etc*, ed ancora altri segni o caratteri speciali (*, /, |, *etc*). Sempre più frequente nella comunicazione digitale è la presenza di simboli pittografici simili ad *emoticon* e conosciute come *emoji*. Divenute popolari in Giappone alla fine degli anni Novanta ed introdotte dalla società di comunicazione giapponese NTT DoCoMo (Negishi, 2014), le *emoji* vengono largamente utilizzate per esprimere uno stato d'animo o per rafforzare il contenuto di un messaggio. Infatti, attualmente Twitter supporta 1100 differenti tipologie di *emoji* (Twitter Symbols, 2019: <https://it.piliapp.com/twitter-symbols/>). Quindi, in un'ottica di trattamento automatico dei testi questi aspetti devono essere tenuti in considerazione. L'analisi automatica o semi-automatica avviene attraverso l'utilizzo di *software*⁵ che permettono il pre-trattamento e l'elaborazione del testo per consentirne l'esplorazione, la descrizione e l'analisi.

⁵ Uno degli strumenti attualmente più popolari (Welbers et al., 2017) è l'ambiente statistico *R* (R Core Team, 2019), il quale rappresenta lo strumento di analisi utilizzato per il presente lavoro di tesi.

Fondamentalmente, gli algoritmi computazionali operano sul testo un procedimento di *tokenizzazione* attraverso il quale la raccolta di documenti viene suddivisa in un insieme di unità lessicali. Le unità minimali di senso sono le parole, poiché rappresentano le componenti semanticamente più significative (Wilbers *et al.*, 2017) e più propriamente intese nella sua forma grafica, ossia come una stringa di caratteri delimitata da due separatori. Pertanto, la *tokenizzazione* che definisce l'occorrenza (*token*) o apparizione delle parole nel testo identifica l'aspetto cruciale dello studio, in quanto i documenti nella loro interezza non possono essere sottoposti a procedure di analisi.

Lo *step* di *pre-processing* inizia con la *normalizzazione* attraverso cui le parole vengono per l'appunto *normalizzate*, ovvero l'intero *tweet* viene convertito in minuscolo, uniformando così spazi, apostrofi, accenti e caratteri ed evitando che l'algoritmo possa considerare come elementi differenti due parole identiche che si presentano con l'iniziale sia maiuscola che minuscola. Poiché l'intento è quello di minimizzare l'ampiezza del vocabolario di ogni *corpus* monolingue, si ricorre alla *lemmatizzazione*. In realtà, in letteratura esistono due diverse soluzioni finalizzate a ridurre la variabilità del testo ed entrambe richiamano l'analisi morfologica: lo *stemming* e la *lemmatizzazione*. Lo *stemming* utilizza uno specifico dizionario per sostituire i termini nella forma morfologica radicale, tralasciando la desinenza ed eliminando il problema del conteggio multiplo delle parole, mentre la *lemmatizzazione* che si decide di adottare in questo contesto riduce ogni forma flessa al lemma di appartenenza. Durante il *pre-trattamento* si procede inoltre con l'eliminazione di: punteggiatura, caratteri numerici, simboli come # dell'*hashtag*, @ del nominativo utenti, *rt* del *retweet* e trattandosi di contenuti multimediali si rimuovono gli *URLs*, *link* eventuali caratteri speciali, segni come *, /, |, \, \$, % e le *emoji*. Come noto, non tutte le parole di un testo sono portatrici di significato e si distinguono le parole piene come sostantivi, aggettivi e verbi che hanno una connotazione semantica dalle parole vuote o *stop words*, che sorreggono invece la costruzione delle frasi. Dal momento che queste ultime non contribuiscono in maniera significativa all'interpretazione del testo vengono rimosse. Nello specifico, la lista delle parole vuote è di consueto predefinita nei diversi *software* o nelle *libraries* degli strumenti di analisi deputati al *Text Mining*; tuttavia, tale elenco può talvolta essere implementato o importato dall'esterno.

Considerando che il presente lavoro di tesi si sviluppa in un contesto *cross-language* con il supporto dello strumento *R*, bisogna tener presente che le liste

di *stop words* contenute nei diversi pacchetti non sono disponibili per tutte le lingue e al fine di sopperire a questa condizione, si ricorre ad una *library* aggiuntiva denominata per l'appunto *Multilingual Stopword Lists* (<https://cran.r-project.org/web/packages/stopwords/stopwords.pdf>) che contiene elenchi di parole vuote per ben 57 diverse lingue in base al codice ISO 639-1 (International Information Centre for Terminology, 2011). Rappresenta un'eccezione di ricorso alla lista delle *stop words* ISO 639-1 il pre-trattamento del *corpus* inglese, per il quale si ritiene invece più opportuno utilizzare l'elenco di parole vuote SMART (*Mechanical Analysis and Retrieval of Text*, Lewis *et al.*, 2004). Tale lista rappresenta un esempio di implementazione delle *stop words* inglesi, infatti Lewis *et al.*, nel 2014 hanno arricchito con 571 *stop words* la precedente lista costituita da 175 parole vuote sviluppata negli anni Sessanta dalla Cornell University.

Nella fase di rimozione delle parole si reputa inoltre ragionevole eliminare le *keywords* utilizzate per l'estrazione dei dati da Twitter, poiché è sensato aspettarsi che tali parole risultino tra i termini più frequenti nella collezione di *tweets* durante la fase di *features extraction*. Questo comporterebbe una distorsione del contenuto informativo effettivo.

La fase di *features extraction* quindi si basa sull'estrazione di informazione in base alle parole che costituiscono la collezione di *tweets* presa in esame.

Per la codifica del testo, il trattamento automatico segue il modello *bag of words*⁶ che trasforma i *tweets* in vettori in uno spazio multidimensionale (Salton, 1975), considerando le forme grafiche indipendentemente dalla loro posizione all'interno del testo e a prescindere dal contesto semantico. Il modello restituisce il vocabolario di ogni lingua, determinando il valore di occorrenza di ciascuna parola o *feature* (ossia termine estratto in uno spazio vettoriale multidimensionale) che è presente nel *corpus*. Malgrado questo approccio sembri penalizzare considerevolmente il contenuto di un documento, diverse ricerche empiriche (Grimmer e Stewart, 2013) dimostrano invece che il conteggio delle parole più frequenti all'interno di un *corpus* è sufficiente per rispondere ad innumerevoli interrogativi di ricerca. In particolare, per individuare l'importanza dei diversi termini nei diversi *tweets*, si adotta uno schema di ponderazione di tipo frequentista⁷, il quale sistema di pesi permette

⁶ La trattazione del modello *bag of words* si rimanda al capitolo secondo: § 2.5.2.

⁷ La codifica e i sistemi di ponderazione dell'unità di analisi si rinviano al capitolo secondo: § 2.5.2.

di ottenere una *tokenlist*, nonché un elenco di termini per ogni lingua a cui viene associato il valore di occorrenza come frequenza assoluta in termini statistici. Per la sua determinazione, in questo caso, si considerano al contempo sia singole parole in termini di *unigrammi* e sia coppie di *tokens* adiacenti, in termini di *n-grams*, ed in particolare *bigrammi* e *trigrammi*.

Durante la fase di estrazione delle *features* i dati vengono organizzati in una struttura matriciale *testi x forme* o *forme x testi* chiamata *document-feature matrix*. Si tratta di una tabella lessicale (**T**) *document-term matrix* o della sua trasposta *term-document matrix* che registra il valore di occorrenza di ciascun termine monolingue in ciascun *tweet*.

Ogniqualvolta si conduce un'analisi statistica ed in particolare uno studio multidimensionale dei dati testuali, in genere l'obiettivo si concretizza nell'opportunità di presentare l'intero vocabolario di un *corpus*, con una copertura del testo del 100%. In pratica però tale finalità è impossibile e l'intento del ricercatore si trasforma nella possibilità di minimizzare il numero delle variabili in analisi, massimizzando la copertura del testo. Questo equivale a definire un'opportuna *soglia* di frequenza, al di sotto della quale si assume che le *features* non considerate non siano rilevanti. In letteratura non esistono regole generali per definire soglie di frequenze valide (Bolasco, 1999), però la determinazione di una soglia di occorrenza aiuta ad identificare i termini (verbi, avverbi, aggettivi e sostantivi) che appartengono ad una fascia di frequenza definita come bassa da Bolasco (1999)⁸, la quale consente di estrarre un maggior numero di parole principali.

Nella presente tesi, trattandosi di corpora medio-grandi, dove il vocabolario di ciascun *corpus* monolingue supera le 200 000/ 300 000 occorrenze, è praticamente indispensabile adottare un livello di *soglia* di frequenza. La soglia prescelta per ogni caso linguistico si prefigge in ogni caso di assicurare la copertura di ciascun *corpus*.

Si tiene conto altresì di specifiche misure lessicometriche, nonché di una serie di indici statistici calcolati sul vocabolario, come ad esempio il valore percentuale che si ottiene dal rapporto tra *type* e *token* (*Type/Token Ratio, TTR*),

⁸Ad ogni soglia di frequenza prescelta, è importante tenere sotto controllo il tasso di copertura del testo, ovvero la percentuale delle occorrenze dovute al vocabolario di ampiezza (*V*) sul totale delle occorrenze del *corpus* (*N*). Il livello di soglia accettabile è posizionato nei primi decili delle basse frequenze, quindi 10-20% delle parole appartenenti a tale fascia di soglia. Gli approfondimenti si rinviano a Bolasco (1999), pp.202-205.

ovvero tra il valore di occorrenza dovuto al numero di *type* o ampiezza del vocabolario, sul totale delle occorrenze (*token*) del *corpus*. Un'altra misura è invece il numero di *hapax* presenti nell'intero vocabolario.

3.2.1.2. *Text summarization dei tweets estratti per ogni lingua*

L'analisi di carattere esplorativo viene condotta su ciascuna matrice lessicale monolingue, al fine esplorare, descrivere e analizzare ogni collezione di *tweets* monolingue. In particolare, tale studio si prefigge di presentare una sintesi del contenuto (o *Text summarization*) di ciascun *corpus*. Una delle visualizzazioni grafiche più comuni è il *word cloud* (Deleuze e Guattari, 1992) che nel linguaggio del *visual design* indica una nuvola di parole come *elenco ponderato*. Tale rappresentazione contribuisce a dare rilevanza ai termini maggiormente frequenti. La nuvola di parole restituisce una immediata intellegibilità grafica del contenuto, poiché mostra le parole presenti nel *corpus* filtrate per soglia di frequenza, con una maggiore dimensione per tutti quei termini che occorrono un maggior numero di volte. Vale dire che la dimensione del carattere è proporzionale al valore di frequenza. In questo modo, si può da subito osservare quali siano i termini più usati nello *speech online* di ciascuna lingua. La *Text summarization* permette quindi di cogliere le principali *features* in uno spazio monolingue, attribuendo al valore di frequenza delle parole un ruolo di decisiva centralità.

Tuttavia, con lo scopo di proporre uno studio comparativo dei linguaggi e delle culture in uno spazio dimensionale *cross-language*, riducendo il più possibile la variabilità linguistica monolingue, una soluzione efficace può consistere nel raggruppare le *features* di ogni lingua in gruppi tematici o concettuali, mantenendo così intatte le peculiarità proprie di ogni registro linguistico.

La strategia di classificazione delle *features* deve però tenere conto che i contenuti sono stati estratti dal *Web* e sono riferiti alla percezione degli utenti su un fenomeno assai controverso come quello della violenza contro le donne. Le parole utilizzate dagli utenti possono quindi rivelare scenari di utilizzo contestualmente diversi ed il loro uso può essere influenzato da orientamenti *emozionali* che consentono di associare ad ogni parola una valenza positiva o negativa. Pertanto, prima di individuare le categorie concettuali è opportuno dapprima determinare se i termini estratti siano stati utilizzati con accezione negativa o positiva.

Al fine di rispondere a tale esigenza entra in gioco la seconda fase della proposta multilingue di questo lavoro di ricerca, che si concretizza per l'appunto nella determinazione di concetti monolingue raggruppando le parole attraverso tecniche di analisi statistica multivariata ma solo dopo averne identificato e assicurato l'orientamento semantico.

3.2.2. Definizione di concetti monolingue

La seconda fase dell'analisi consiste quindi nel determinare i concetti per ciascuna lingua. Questo obiettivo viene realizzato attraverso due sotto-fasi:

1. Nella prima, ciascuna collezione di *tweets* monolingue viene classificata in base al *calcolo della polarità semantica*, così da ottenere due diversi *dataset* testuali: uno solo di *tweets* positivi e l'altro solo di *tweets* negativi.
2. Nella seconda sotto-fase invece, da ogni *dataset* testuale sia positivo che negativo vengono estratte le rispettive *features*, le quali vengono raggruppate in classi di concetti positivi e negativi mediante la combinazione di due tecniche di analisi multidimensionale: l'*Analisi delle Corrispondenze* e la *Cluster Analysis*.

3.2.2.1. L'orientamento semantico dei testi ed i principali approcci per il calcolo della polarità

La determinazione dell'orientamento semantico dei testi rientra tra le operazioni di classificazione dei documenti. Più propriamente, i testi ed in questo caso i *tweets* vengono classificati in base alla *polarità* delle parole.

La *polarità* delle parole rappresenta la direzione con cui la parola si discosta dalla norma per il suo gruppo semantico o *campo lezico* (Lehrer, 1974 citato in Hatzivassiloglou e McKeown, 1997), denotando una valutazione positiva o negativa circa un argomento (Hatzivassiloglou e McKeown, 1997; Löbner, 2000; Esuli e Sebastiani, 2005; Giannakidou, 2008).

La letteratura propone tre diversi approcci per l'orientamento semantico dei testi (ad esempio Esuli e Sebastiani, 2005):

- *Soggettività/Oggettività (SO-orientation)*: determina se un documento possiede una natura fattuale, ovvero se il testo è in grado di descrivere una situazione o un evento senza esprimere un'opinione soggettiva (Yu e Hatzivassiloglou, 2003; Pang e Lee, 2004).
- *Positività/Negatività (PN-orientation)*: identifica se un testo soggettivo esprime un'opinione positiva o negativa (Turney, 2002; Pang e Lee, 2004).
- *Intensità dell'orientamento Positivo/Negativo (PN-strength)*: indica il grado di positività e negatività con un'intensità espressa a più livelli. L'opinione di un testo può essere debolmente positiva, lievemente positiva o fortemente positiva. In maniera analoga si classifica l'intensità dell'orientamento negativo (Wilson *et al.*, 2004; Nielsen, 2011).

Generalmente, lo studio dell'orientamento semantico di un testo avviene a tre diversi livelli: *documento, frase, entità-aspetto* (Liu, 2012).

- *A livello di documento (document-level)*: si propone di classificare un intero testo o documento determinando se esso esprima un'opinione positiva o negativa (Pang *et al.*, 2002; Turney, 2002).
- *A livello di frase (sentence-level)*: ogni documento viene frammentato in frasi e si misura la polarità di ciascuna proposizione, che può essere sintetizzata più in generale a livello di documento (Tan *et al.*, 2011).
- *A livello di entità e aspetto (entity and aspect level)*: partendo dall'assunto che ciascuna opinione miri ad un obiettivo di comunicazione positiva o negativa, questo livello studia l'orientamento semantico dell'opinione stessa, non fermandosi agli elementi basilari del linguaggio (frasi, paragrafi, documenti, *etc.*). Si tratta di un livello di analisi particolarmente dettagliato e in precedenza chiamato *livello di funzionalità* (Hu e Liu, 2004).

L'orientamento semantico dei testi si basa essenzialmente sul *calcolo della polarità*. Il calcolo della polarità consiste nell'attribuire un punteggio positivo o negativo ad un testo o una frase, sulla base delle parole presenti in un *dizionario* o *lexicon*. Un *lessico* non è altro che un elenco di parole polarizzate, ovvero una

lista di termini a cui viene assegnato un determinato *score* capace di contestualizzare la positività o negatività di un documento o di una frase.

In letteratura esistono diversi dizionari di polarità, alcuni creati manualmente (come ad esempio Elarnaoty *et al.*, 2012) ed altri in maniera automatica o semi-automatica (come il caso di Abdul-Mageed e Diab, 2012). Alcune risorse sono riferite a specifici domini, come ad esempio l'ambito finanziario (Loughran-McDonald *financial dictionary* di Loughran e McDonald, 2011) o quello della psicologia sociale (Harvard IV *dictionary* in General Inquirer: <http://www.wjh.harvard.edu/~inquirer/homecat.htm>). Altri dizionari invece sono sviluppati in base al campo d'applicazione e agli interrogativi della ricerca (ad esempio Pröllochs *et al.*, 2015). Oltre al dominio di applicazione, si diversifica anche il punteggio che la polarità può assumere. Nello specifico, nell'approccio basato sull'orientamento *PN* (*PN-orientation*), la direzione semantica viene calcolata assegnando ai termini un punteggio di polarità positiva + 1, negativa - 1 o neutrale 0 (Hu e Liu, 2004). L'approccio che studia l'*intensità dell'orientamento semantico Positivo/Negativo* (*PN-strength*) invece non definisce la polarità solo in termini di segno (+ positivo e - negativo) ma considera il grado di positività e negatività, con un'*intensità* espressa a più livelli (Nielsen, 2011). In questo caso, i lessici considerano una scala discreta o continua per la polarità: dal punteggio più negativo a quello più positivo (Wilson *et al.*, 2004).

Mohammad e Turney (2010; 2013a) propongono peraltro un dizionario noto come *NRC (National Research Council Canada) Word-Emotion Association Lexicon* che calcola la polarità e mette in relazione le parole con le emozioni. In particolar modo, tale dizionario conosciuto anche come *EmoLex* o *NRC Emotion Lexicon* (Mohammad e Turney, 2013a) oltre ad attribuire la polarità alle parole, le assegna in otto diverse categorie di emozioni: rabbia, paura, anticipazione, fiducia, sorpresa, tristezza, gioia o disgusto (Plutchik, 1991).

3.2.2.2. *L'NRC Emotion Lexicon per misurare la polarità dei testi multilingue*

La maggior parte delle risorse disponibili per il calcolo della polarità semantica sono sviluppate per lo studio della lingua inglese (Hu e Liu, 2004; Nielsen, 2011; Loughran e McDonald, 2011; Pröllochs *et al.*, 2015; Bose, 2016; Rinker,

2016; Schmidt, 2017). Esiste invece un numero piuttosto esiguo di dizionari adatti a testi multilingue.

Tra i *lexicon cross-language* vi è *LIWC (Linguistic Inquiry and Word Count; Pennebaker et al., 2007)* che misura l'*intensità dell'orientamento semantico Positivo/Negativo (PN-strength)* per testi multilingue. Per l'utilizzo dei lessici *LIWC* è prevista una licenza che ne consente l'uso attraverso specifiche condizioni economiche.

Un dizionario di libero accesso è invece il sopracitato *NRC Emotion Lexicon* (Mohammad e Turney, 2013a; 2013b), creato con l'ausilio di *Amazon Mechanical Turk*⁹ mediante operazioni di annotazione manuale e per mezzo di *crowdsourcing*. Le procedure di annotazione del lessico determinano la polarità positiva o negativa e stabiliscono l'associazione delle parole alle otto categorie emozionali (Plutchik, 1991) per testi di varie lingue. La costruzione di questo lessico deriva da tre annotatori diversi che tengono conto di tutte le possibili ambiguità semantiche delle parole. Infatti, i termini sono divisi in due categorie: parole con più di un significato e parole con un solo significato. Ad esempio, il dizionario inglese dell'*NRC* è costituito da 14 182 *unigrammi* e all'incirca 25 000 *word-sense*, ovvero parole con più di un significato. Il lessico *NRC* offre dizionari di polarità semantica ed *emotions detection* per all'incirca cento diverse lingue, spesso affidandosi a sistemi di traduzione automatica (Mohammad e Turney, 2013a).

Sono diversi i *packages* di *R* che sfruttano l'*NRC Emotion Lexicon*, tra cui ad esempio *tidytext* (De Queiroz et al., 2017), *syuzhet* (Jockers, 2015) e *quanteda* (Benoit et al., 2018; <https://github.com/kbenoit/quanteda.dictionaries>).

Considerando l'opportunità di impiegare tale *lexicon* per lo studio di più lingue ed in virtù della sua accessibilità attraverso le diverse *libraries* di *R*, l'*NRC Emotion Lexicon* rappresenta una valida soluzione per coprire la varietà linguistica richiesta dal presente contesto di ricerca.

Specificatamente, in questo caso, il calcolo della polarità avviene a livello di documento o *tweet* per ciascuna collezione di dati monolingue. Attraverso la determinazione dell'orientamento semantico si ritiene valido organizzare i testi positivi come un *dataset* positivo e i testi negativi come un *dataset* negativo. La raccolta di dati positivi e negativi identificano due *sotto-corpora* per ciascuna

⁹ Mechanical Turk: <http://www.mturk.com/mturk/welcome>

lingua, i quali vengono processati con le tipiche operazioni di *Text Mining* consentendone l'estrazione delle *features*: da una parte solo *features* positive e dall'altra solo *features* negative. Dall'elaborazione dei dati, le *features* vengono organizzate mediante una tabella lessicale positiva (\mathbf{T}_p), $n_p \times p_p$ dove n_p sono i *tweets* positivi e p_p sono le *features* positive e una tabella lessicale negativa (\mathbf{T}_n), $n_n \times p_n$, dove n_n sono i *tweets* negativi e p_n sono le *features* negative. Il calcolo della polarità semantica dei *tweets* e l'estrazione di *features* positive e negative per ciascuna collezione di dati monolingue rappresenta la fase preliminare alla determinazione vera e propria dei concetti per ogni lingua, la quale procedura cosiddetta di *Concepts Extraction* viene realizzata attraverso tecniche di analisi multidimensionale adatte allo studio dei dati testuali.

3.2.2.3. *La Concepts Extraction tramite Correspondence e Cluster Analysis*

Determinare i concetti monolingue significa quindi raggruppare in classi di concetto le *features* positive e negative di ciascuna collezione di *tweets* monolingue. A tale scopo però non ci si può limitare a tecniche di analisi che si basano principalmente sul valore di frequenza delle *features*. Risulta invece opportuno basarsi su metodi finalizzati ad esplorare le relazioni o le similarità tra le unità lessicali di ciascuna lingua. Più propriamente, si tratta di ricorrere a metodi considerati particolarmente induttivi, perchè capaci di indagare schemi interpretativi da un punto di vista matematico che non sono identificabili a priori (Boumans e Trilling, 2016),

La letteratura concernente l'analisi dei dati testuali annovera diverse tecniche e metodi che permettono di categorizzare gruppi di parole in temi o *topic* differenti. Tra le tecniche più frequentemente utilizzate vi sono quelle di *Topic Modeling*, come la *Latent Dirichlet Allocation (LDA)*; Blei *et al.*, 2003; ad esempio Anoop *et al.*, 2011; Yin e Wang, 2014; Tong e Zhang, 2016) su base probabilistica o i modelli di classificazione come il *Random Forest* (Breiman, 2001; ad esempio Bouaziz *et al.*, 2014; Elagamy, 2017; Islam, *et al.*, 2019) che impiega esclusivamente alberi decisionali. Altri approcci sono invece di tipo *network-based* (ad esempio Syyadi e Raschid, 2013; Misuraca *et al.*, 2019), i quali raggruppano le *features* in gruppi o comunità di concetti attraverso i principi teorici e pratici dell'analisi delle reti sociali (Wasserman e Faust, 1994). Tra le altre proposte, vi è anche la *Cluster Analysis* (Tryon, 1939) nota

in letteratura per raggruppare *set* di dati di varia natura. Tuttavia, basare la rilevazione dei concetti su strategie esclusivamente classificatorie non sempre permette di individuare fattori latenti che possono avere invece la loro importanza (Lebart *et al.*, 1995), soprattutto in un ambito così delicato come quello corrente che mira a definire gruppi concettuali e dove particolare rilevanza possiede l'ambiguità semantica racchiusa nel linguaggio.

A tal proposito, infatti, la letteratura sottolinea l'opportunità di combinare metodi fattoriali con metodi di classificazione, così che l'organizzazione spaziale delle classi sugli assi fattoriali possa evidenziare eventuali tratti latenti inattesi. In particolar modo, nell'ambito dei dati testuali, combinare metodi fattoriali con metodi di classificazione può rivelarsi utile, in quanto si possono evincere le relazioni di similarità esistenti tra i profili lessicali e sulla base di tali similarità si possono classificare gruppi di parole che individuano domini terminologici o concettuali.

Pertanto, se da un'analisi del testo di tipo paradigmatico le parole possono essere listate seguendo un qualche ordine, l'ambiguità semantica può essere invece risolta mediante il ricorso a tecniche di analisi multidimensionale. In particolare, nell'ambito dei dati testuali, uno dei metodi fattoriali prevalentemente impiegati con l'obiettivo di cogliere l'accezione interna al testo, sulla base della vicinanza e quindi somiglianza delle *features* sul piano fattoriale, è l'*Analisi delle Corrispondenze*. Tale tecnica consente di rilevare sul piano eventuali *modelli di senso* del contenuto di ogni testo (Bolasco, 2005).

Sulla base di tali considerazioni, nel presente contesto di ricerca, la scelta metodologica per la determinazione dei concetti monolingue ha quindi lo scopo di voler rilevare potenziali tratti latenti in ogni *corpus*, limitando così il più possibile l'ambiguità semantica e restituendo la globalità di senso racchiusa in ogni testo, sulla base della vicinanza delle *features* sul piano fattoriale. Preservando la posizione assegnata a ciascuna *feature* dalle coordinate fattoriali, in questo lavoro si intende quindi procedere alla loro classificazione attraverso la combinazione di queste ultime due tecniche di analisi multivariata, quali per l'appunto l'*Analisi delle Corrispondenze* e la *Cluster Analysis*.

Il vantaggio di combinare queste due tecniche di analisi multivariata è che la classificazione delle parole viene realizzata a partire dalla posizione datagli dalle coordinate rispetto agli assi fattoriali considerati e quindi, il calcolo della distanza tra le *features* è considerevolmente ridotto, proprio perché si riferisce allo spazio dimensionale limitato ai primi fattori.

3.2.2.3.1. L'Analisi delle Corrispondenze

L'Analisi delle Corrispondenze (AC, Benzécri, 1973) rappresenta uno dei metodi più efficaci per il trattamento multidimensionale di dati qualitativi (Gherghi e Lauro, 2008).

Si tratta strumento particolarmente utile per lo studio di una distribuzione congiunta, in quanto consente di descrivere, sia da un punto di vista geometrico che algebrico, la relazione esistente tra le modalità di due o più variabili di un *set* di dati organizzati in matrice. Il modello mira alla ricerca di sottospazi ottimali su cui rappresentare simultaneamente la relazione esistente tra le righe e le colonne della tabella dei dati.

In uno schema tipico, si consideri $\mathbf{N}(p,q)$ come una tabella di contingenza in cui si incrociano due variabili qualitative x e y , rispettivamente con p e q modalità e con n_{ij} come valore generico che esprime la frequenza con cui si presentano simultaneamente la i -esima modalità di x e la j -esima modalità di y . Partendo dai caratteri originari, lo scopo dell'AC è di calcolare una serie di fattori, ognuno dei quali indica un tratto latente del tipo di associazione interna ai dati, facendo riferimento a procedure di decomposizione matriciale.

La visualizzazione grafica dei fattori individua la struttura dell'associazione della tabella dei dati e quindi l'interdipendenza tra le modalità di riga e colonna. Le modalità dei due caratteri vengono rappresentate in uno spazio di riduzione dimensionale di tipo cartesiano.

La rappresentazione in forma grafica identifica la caratteristica propriamente induttiva attribuita alla tecnica, in quanto permette di cogliere gli aspetti non rilevabili dalla sola lettura dei dati (Boumans e Trilling, 2016).

Si supponga che \mathbf{F} (con $f_{ij}=n_{ij}/n_{i.}$) sia la matrice delle frequenze relative e che i vettori delle frequenze marginali di riga \mathbf{r} e di colonna \mathbf{c} siano calcolati rispettivamente nel seguente modo: $f_{i.}=n_{i.}/n$ e $f_{.j}=n_{.j}/n$. Si consideri inoltre di poter ottenere le matrici diagonali delle distribuzioni di riga $\mathbf{D}_p \equiv \text{diag}(\mathbf{r})$ e di colonna $\mathbf{D}_q \equiv \text{diag}(\mathbf{c})$. Calcolando il prodotto della matrice delle frequenze \mathbf{F} con \mathbf{D}_p^{-1} e \mathbf{D}_q^{-1} si ottengono le matrici delle distribuzioni condizionate, rispettivamente $\tilde{\mathbf{R}}$ per i profili di riga e $\tilde{\mathbf{C}}$ per i profili di colonna:

$$\tilde{\mathbf{R}} \equiv \mathbf{D}_p^{-1} \mathbf{F} \equiv \begin{bmatrix} \tilde{r}'_1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{r}'_p \end{bmatrix} \quad \tilde{\mathbf{C}} \equiv \mathbf{F} \mathbf{D}_q^{-1} \equiv \begin{bmatrix} \tilde{c}'_1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{c}'_q \end{bmatrix} \quad [3.1]$$

L'interpretazione geometrica dei profili di riga e di colonna è data da una nuvola di punti proiettati in uno spazio dimensionale. In particolare, la nube dei profili di riga $\tilde{\mathbf{r}}'_1, \dots, \tilde{\mathbf{r}}'_p$ è rappresentata in uno spazio $\tilde{\mathbf{R}}^{q-1}$ con pesi¹⁰ \mathbf{D}_p e metrica¹¹ \mathbf{D}_q^{-1} . Allo stesso modo, la nube dei profili di colonna $\tilde{\mathbf{c}}'_1, \dots, \tilde{\mathbf{c}}'_q$ è rappresentata in uno spazio $\tilde{\mathbf{R}}^{p-1}$ con pesi \mathbf{D}_q e metrica \mathbf{D}_p^{-1} .

La ricerca di sottospazi ottimali, in termini di minimi quadrati, su cui rappresentare simultaneamente la relazione esistente tra i profili di riga e quelli di colonna ha origine dalla decomposizione in valori singolari generalizzata della tabella delle frequenze \mathbf{F} (Greenacre, 1984). I vettori singolari di destra e di sinistra determinano gli assi principali degli spazi dimensionali su cui proiettare le nuvole di punti dei profili di riga e colonna (*Ibidem*). Si determina che:

$$\mathbf{F} = \mathbf{U} \mathbf{D}_\mu \mathbf{V}'$$

$$\text{con } \mathbf{U}' \mathbf{D}_p^{-1} \mathbf{U} = \mathbf{V}' \mathbf{D}_q^{-1} \mathbf{V} = \mathbf{I} \quad [3.2]$$

\mathbf{D}_μ rappresenta la tabella della decomposizione in valori singolari, mentre $\mathbf{D}_p^{-1} \mathbf{U}$ e $\mathbf{D}_q^{-1} \mathbf{V}$ indicano le colonne delle matrici che identificano gli assi principali degli spazi dei profili di riga e colonna. Le coordinate di ciascun punto di riga e colonna proiettate sull'asse α -esimo sono le coordinate principali. Il loro valore si ottiene dalle seguenti operazioni:

$$\psi_{\alpha i} = \sqrt{\lambda_\alpha} f_{i.}^{-1/2} u_{\alpha i} \quad \psi_{\alpha j} = \sqrt{\lambda_\alpha} f_{.j}^{-1/2} v_{\alpha j} \quad [3.3]$$

Individuati quindi gli assi principali su cui proiettare i punti, per una corretta interpretazione grafica si determinano dei coefficienti che tengano conto della prossimità tra i punti e i piani principali e che al contempo, considerino l'importanza di ciascun punto nella determinazione degli assi (Gherghi e Lauro, 2008). La qualità della rappresentazione è data dalla bontà dell'approssimazione, che si misura in termini di variabilità del fenomeno. Tale variabilità può essere spiegata dai sottospazi identificati dall'analisi fattoriale, mediante la misura del rapporto tra il quadrato dell' α -esimo valore singolare e

¹⁰ Il sistema di pesi è definito dalle masse dei punti nello spazio dimensionale.

¹¹ La metrica adottata è quella del *Chi-quadrato*.

la somma dei quadrati di tutti i valori singolari ottenuti dalla decomposizione matriciale. L'AC mira ad indagare la struttura di associazione tra le variabili e quindi prima di procedere alla riduzione dello spazio dimensionale di riga e di colonna risulta stima il grado di interdipendenza tra i caratteri presi in esame. Il test di indipendenza adottato è la statistica del *chi-quadrato* (χ^2) che misura la distanza tra i punti delle modalità di x e y . Il contributo assoluto alla variabilità degli assi di ciascun punto di riga e colonna è dato da:

$$ca_{\alpha}(i) = f_i \cdot \psi^2 \alpha_i \qquad ca_{\alpha}(j) = f_j \cdot \psi^2 \alpha_j \qquad [3.4]$$

A differenza di altri approcci fattoriali, l'aspetto peculiare dell'AC è dato dalla rappresentazione ed interpretazione simultanea dei profili di riga e di colonna, restituendo così una simmetria complessiva che deriva dalla sovrapposizione di due analisi asimmetriche (*Ibidem*). La rappresentazione è possibile grazie alla proprietà baricentrica dell'AC che consente di misurare la posizione di ciascun punto dei profili riga come media ponderata dei punti-vertici definiti dai profili colonna, e i cui pesi sono uguali agli elementi del profilo riga stesso. In maniera analoga, la posizione di ciascun punto dei profili colonna si misura attraverso la media ponderata dei punti-vertici definiti dai profili riga, i cui pesi sono pari agli elementi del profilo colonna stesso. Di conseguenza, la rappresentazione è definita β -baricentrica (*Ibidem*); ovvero ogni riga risulta essere baricentro delle colonne ed ogni colonna è baricentro delle righe.

Nel caso dell'analisi dei dati testuali, l'AC è detta *Analisi delle Corrispondenze Lessicali* (ACL; Rositi, 1988 citato in Losito, 2002) e la matrice di contingenza è rappresentata da una tabella lessicale (\mathbf{T}) di dimensioni $p \times n$, il cui elemento generico di frequenza t_{ij} indica, come noto, il numero di occorrenze della parola i -esima nel documento j -esimo.

Se rapportato allo studio di fattispecie, lo schema tipico dell'ACL permette di rappresentare i profili di riga e di colonna della matrice \mathbf{T} e quindi indagare la relazione tra le parole nello spazio definito dai *tweets* e la relazione tra i *tweets* nello spazio definito dalle occorrenze nel *corpus*. In altre parole, l'ACL restituisce la relazione esistente tra i termini, tra i documenti e, indirettamente, tra termini e documenti.

Spesso, le dimensioni di una matrice lessicale (\mathbf{T}) sono elevate a causa di un'eccessiva *text sparsity* che genera la conseguente presenza di celle nulle, riducendo così il contenuto informativo e i tratti latenti dei testi. In tal caso, si

considera una *tabella lessicale aggregata* (Lebart *et al.*, 1998) generata da una matrice \mathbf{Q} , di dimensioni $n \times q$, di codifica disgiuntiva completa, dove q sono le modalità della variabile qualitativa relativa ai testi. Questa rappresentazione matriciale \mathbf{T}_q è data da $\mathbf{T} * \mathbf{Q}$, e presenta in riga le p parole ed in colonna le q classi di documenti.

L'obiettivo delle tecniche fattoriali risiede nel ridurre il numero di variabili che caratterizzano un dato fenomeno, producendo delle variabili *sintetiche* in grado di ricostruire gli assi semantici su cui si modella la variabilità del contenuto di un testo. Tanto più le parole sono vicine sul piano fattoriale, tanto più *probabilmente* sono utilizzate congiuntamente nei testi analizzati (Bolasco, 2004). In generale, nella mappa fattoriale dell'*ACL* la dispersione dei punti attorno all'origine degli assi principali è sinonimo di una forte associazione interna alla tabella lessicale (\mathbf{T}). La vicinanza tra le parole indica una similarità nel modo in cui sono utilizzate nel *corpus*, e l'adiacenza sul piano delle modalità appartenenti ad una partizione di classe di documenti è espressione di un vocabolario comune. La logica interpretativa della rappresentazione β -baricentrica ci consente di valutare la posizione delle parole rispetto all'intera nuvola di punti sul piano, in quanto il contributo di un punto rispetto alla determinazione dell'asse principale è letto in termini assoluti.

Nella rappresentazione fattoriale, la metrica del χ^2 tende ad assegnare un maggior peso alle parole che si presentano poche volte nella collezione di documenti, evidenziando così una forma di linguaggio peculiare ma portando anche ad una probabile distorsione dei risultati.

In base alla vicinanza o distanza delle unità linguistiche sul piano, le relazioni tra i profili lessicali possono essere interpretate, in base a quanto definito da Borrelli *et al.* (2018), come costellazioni semantiche, in grado di ricostruire raggruppamenti tematici che identificano il senso globale di un testo. In sintesi, attraverso l'*AC* è possibile determinare categorie tematiche o concettuali che individuano termini che si riferiscono ad uno stesso tema o argomento.

Tuttavia, l'*AC* studia le relazioni esistenti tra le modalità di due variabili qualitative rappresentandole in un sottospazio ottimale di riduzione dimensionale, dove talvolta la complessità delle relazioni è di difficile intelligibilità grafica (Ghergi e Lauro, 2008). La finalità di partizionamento può essere quindi raggiunta con le tecniche di classificazione automatica come per l'appunto la *Cluster Analysis*. Quindi, con lo scopo di determinare i principali concetti monolingue e per una complementarità informativa dei dati, in questo

contesto, si classificano le *top features* di ciascuna lingua combinando l'*AC* e la *Cluster Analysis*.

3.2.2.3.2. *La Cluster Analysis*

La *Cluster Analysis* è una tecnica di analisi multidimensionale che ha lo scopo di raggruppare gli oggetti di un *set* di dati multivariati basandosi su misure di similarità/dissimilarità (o mutua distanza) tra le entità.

Normalmente i metodi di classificazione mirano a raggruppare le modalità di riga e di colonna di una matrice di contingenza, come quella lessicale, per mezzo di algoritmi formalizzati in base a criteri di ottimizzazione. In tal senso, la letteratura della *Cluster Analysis* definisce come miglior raggruppamento possibile quello che minimizza la prossimità tra gli elementi di ogni classe individuata (Bolasco, 1999). Il concetto di raggruppamento presuppone dunque l'esistenza di un criterio di riferimento. Nel caso di fattispecie, la *Cluster Analysis* si prefigge di generare gruppi (*cluster*) in cui le parole che vi fanno parte sono simili tra loro e non appartengono ad altri raggruppamenti. In questo modo si determina il criterio di omogeneità *intra-cluster* ed eterogeneità *tra cluster*, garantendo valori di devianza minima *within-group* e devianza massima *between-group* (Jain *et al.*, 1999).

I metodi di classificazione operano attraverso le procedure gerarchiche o non gerarchiche: le prime creano un insieme di partizioni ordinate gerarchicamente, nelle quali ogni gruppo ad un qualunque livello fa parte di un *cluster* più ampio ad un livello successivo (Barbaranelli, 2007). Se sono di tipo agglomerativo si originano da una partizione iniziale formata da k gruppi, ognuno dei quali è costituito da un solo elemento e per mezzo di successive fusioni dei raggruppamenti meno distanti tra loro, si giunge ad una condizione in cui un solo *cluster* contiene tutti gli elementi. Se le procedure sono di tipo divisivo invece, l'intero *set* di dati iniziali è raggruppato in un'unica grande partizione e mediante le fasi successive, si realizza una classificazione gerarchica che divide il *cluster* iniziale in gruppi via via sempre più omogenei al loro interno (Barbaranelli, 2007; Gherghi e Lauro, 2008).

Le procedure non gerarchiche invece ripartiscono gli oggetti iniziali in k gruppi stabiliti a priori dal ricercatore, fino al raggiungimento delle proprietà ottimali definite da un criterio di partenza. I metodi gerarchici danno quindi luogo ad una gerarchia mentre le procedure non gerarchiche realizzano una partizione.

Gli algoritmi di *clustering* sono strettamente dipendenti dal criterio di raggruppamento, che a sua volta comporta la definizione di una metrica per il calcolo della distanza¹². Tali scelte derivano dalla natura dei dati e dagli obiettivi di ricerca perseguiti. Spesso, con lo scopo di circoscrivere gli effetti della classificazione automatica che possono derivare da tali scelte preliminari, si adottano *strategie di classificazione miste* (Lebart *et al.*, 1995; Bolasco, 1999), come nel presente lavoro di tesi. Nello specifico, in questo contesto di ricerca si intende ricorrere ad un algoritmo di *clustering* misto proposto da Lebart *et al.*, (1995) ed implementato in diversi *software* per la classificazione dei dati, il quale parte da un metodo di *clustering* gerarchico per poi consolidare la classificazione attraverso un metodo di *clustering* non gerarchico. Specificamente, questo algoritmo in un primo momento classifica le parole partendo da un raggruppamento gerarchico basato sul metodo di *Ward* (1963), che in letteratura è menzionato tra i più efficaci metodi di aggregazione (Bolasco, 1999) finalizzati alla minimizzazione della varianza *within*. Il metodo di *Ward* si modella sul teorema Huygens che consente di scomporre l'*inerzia totale* (o varianza totale) in due quantità: *inerzia inter-classe* (varianza *between*) e *inerzia intra-classe* (varianza *within*):

$$\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^{I_q} (x_{iqk} - \bar{x}_k)^2 = \sum_{k=1}^K \sum_{q=1}^Q I_q (\bar{x}_{qk} - \bar{x}_k)^2 + \sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^{I_q} (x_{iqk} - \bar{x}_{qk})^2 \quad [3.5]$$

dove x_{iqk} è il valore di una variabile k per le *features* i del *cluster* q , \bar{x}_{qk} è la media della variabile k per il *cluster* q , \bar{x}_k identifica la media complessiva della variabile k mentre I_q è il numero delle parole presenti nel cluster q . La finalità perseguita dal metodo di *Ward* è di minimizzare l'*inerzia intra-classe* ad ogni *step* dell'esecuzione dell'algoritmo, caratterizzando così di omogeneità interna ciascun raggruppamento di *features*. In letteratura, tale metodo è considerato particolarmente adatto a classificare gli oggetti individuati dalle coordinate fattoriali in uno spazio euclideo¹³ multidimensionale, come le *features* delle diverse lingue nel caso di fattispecie.

¹² Approfondimenti sui criteri di raggruppamento e metodi per il calcolo della distanza si rimandano a: Leti (1979); Bolasco (1999); Gherghi e Lauro (2008); Khachumov (2012).

¹³ Approfondimenti sulla distanza Euclidea si rinviano a Arabie (1991); Solomentsev, (2002).

La classificazione gerarchica che ne deriva viene poi messa in discussione e “riprocessata” assumendo il numero di *cluster* identificati come una scelta ragionata di nuclei iniziali. In particolare, basandosi sulla proposta di Lebart *et al.*, (1995), l’algoritmo riclassifica le parole attraverso una procedura di *clustering* non gerarchica detta di *ottimizzazione* o *consolidamento*, con lo scopo di migliorare le classi ottenute dalla classificazione gerarchica. L’eventuale miglioramento è misurato dall’aumento del rapporto tra *inerzia inter-classe* e *inerzia totale*. Alle volte il risultato può rimanere immutato e ciò vuol dire che la classificazione gerarchica ha già ripartito in modo ottimale le classi. Per la soluzione di partizionamento consolidato finale, in questo contesto si ricorre al metodo delle *k-medie* (MacQueen, 1967), considerato uno dei metodi più comunemente utilizzati tra quelli non gerarchici. L’adozione di tale tecnica partizionale massimizza l’omogeneità *intra-classe* e l’eterogeneità *inter-classe*. Ogni *cluster* è rappresentato dal suo *centroide* che corrisponde alla media dei dati (punti) appartenenti a ciascun gruppo. Il processo di classificazione è iterativo fino a convergenza ottimale dell’algoritmo e le parole vengono classificate in *k* gruppi distinti, di dimensione maggiore o uguale a 2, nel rispetto del principio di coesione interna e separazione esterna (Jain *et al.*, 1999).

$$D_{w(C_k)} = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad [3.6]$$

dove x_i è un punto dei dati appartenenti al *cluster* C_k e μ_k è la media dei punti assegnati al raggruppamento C_k . Ciascuna osservazione x_i viene assegnata ad un dato *cluster*, tale per cui la somma dei quadrati (*SS*) delle distanze di ogni osservazione rispetto al *centroide* del gruppo è minima. La devianza totale *inter-cluster* si può formalizzare nel seguente modo:

$$D_{Tw} = \sum_{k=1}^k D_{w(C_k)} = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad [3.7]$$

L’algoritmo *k-means* misura attraverso la [3.7] la compattezza, nonché la bontà della minimizzazione della sommatoria totale della devianza *within-group*.

Al fine di caratterizzare ciascun *cluster* di parole ottenuto, si utilizza la *statistica test* (t_k) secondo una legge ipergeometrica (Lebart *et al.*, 1995)

significativa con $\alpha = 0.05$. Questo *test* viene utilizzato come strumento descrittivo per ordinare e selezionare le parole caratteristiche di ciascun raggruppamento ma non ne fornisce un'interpretazione probabilistica. Il suo valore (noto come *value-test*) misura in termini di scarti normali standardizzati la *significatività* della differenza tra la percentuale di presenza in una modalità della classe e la percentuale della modalità nel *corpus*.

In definitiva, dall'applicazione della *Cluster Analysis* sulle coordinate fattoriali ottenute dall'*AC* ne deriva che ogni gruppo di parole positive e ogni gruppo di parole negative possa essere interpretato in un vero e proprio *spazio di concetto* (*concept space*; Chew, 2009) multidimensionale, dove per *concetto* si intende un'unità di pensiero costituita per astrazione sulla base di proprietà comuni ad un insieme di oggetti o parole (Norma Internazionale, Terminologia-Vocabolario, 1995). Ad ogni *cluster* di *features* viene inoltre attribuita un'etichetta concettuale finalizzata ad esprimere la semantica racchiusa in ogni raggruppamento, la cui assegnazione si basa altresì sui valori ottenuti dalla *statistica-test*.

Identificati quindi i concetti delle diverse lingue e determinate le parole appartenenti ad ogni ogni raggruppamento, le etichette concettuali e le *features* possono essere organizzate in una rappresentazione matriciale.

3.2.2.4. La codifica dei dati per la costruzione della matrice di concetti

Al termine della fase di *Concepts Extraction* per ogni lingua è possibile dunque costruire una tabella *termini x concetti*, dove in riga vengono poste le diverse parole ed in colonna vengono indicati i diversi concetti. Tale configurazione matriciale può definirsi una *matrice dei concetti* (**C**), come tabella bi-dimensionale $p \times m$, dove p indicano le parole ed m i diversi concetti. La tabella **C** è una matrice disgiuntiva completa in cui l'elemento generico c_{ik} registra elementi dicotomici 1 o 0: 0 quando la parola i -esima non appartiene al k -esimo concetto, 1 se la parola i -esima appartiene invece al k -esimo concetto. La *matrice dei concetti* (**C**) viene costruita per ogni lingua, sia per i soli concetti positivi (**C_p**) e sia per i soli concetti negativi (**C_n**).

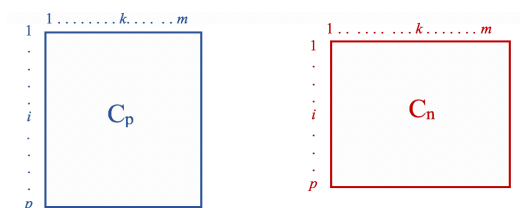


Figura 3.2. Struttura dei dati per la costruzione della matrice dei concetti
 Fonte: Elaborazione propria

Mediante trasformazioni di algebra matriciale, dalla matrice *termini x concetti* positiva o negativa si può ricavare la matrice *documenti x concetti* (Misuraca *et al.*, 2019) o *document-concept matrix*, la quale configurazione indica in quali *tweets* sono contenuti i concetti individuati, sulla base della proporzione dei termini afferenti ad ogni concetto. La *document-concept matrix* si ottiene moltiplicando la trasposta della matrice lessicale positiva (T_p) o negativa (T_n) con la matrice dei concetti positiva (C_p) o negativa (C_n). Il prodotto ottenuto da tale operazione viene poi moltiplicato con la matrice diagonale ricavata dalla distribuzione marginale di colonna della matrice dei concetti positiva e negativa. In sintesi: $T_p^* \equiv (T_p^T C_p) D_{Gp}^{-1}$ e $T_n^* \equiv (T_n^T C_n) D_{Gn}^{-1}$. In questo modo si ottiene una matrice (E) di dimensioni $n \times m$, dove n sono i *tweets* ed m sono i concetti.

Tuttavia, potrebbe risultare interessante individuare quali siano le associazioni tra i concetti che più accomunano le lingue, ovvero rilevare se esiste una corrispondenza tra le lingue in base a come si relazionano i concetti in ciascun vocabolario monolingue.

In altre parole, si tratta di determinare se i concetti, che possono talvolta essere simili tra corpora comparabili, siano utilizzati anche nello stesso contesto e se sussiste quindi una co-esistenza o co-occorrenza tra i concetti.

In tal senso si possono indagare le relazioni tra i concetti al fine di definire veri e propri *campi concettuali* (Norma Internazionale, Terminologia-Vocabolario, ISO 1087, 1995), nonché gruppi di concetti basati su nessi tematici.

Il campo concettuale di ogni registro linguistico è il punto di partenza che stabilisce in linguistica il *sistema concettuale*, il quale sistema può mutare in base al contesto extralinguistico, ovvero in base a dimensioni socio-culturali che possono influenzare la struttura della comunicazione e quindi l'associazione tra i concetti.

Ad esempio, se nel vocabolario di una lingua A un concetto y viene associato ad un altro concetto z non è detto che in un diverso vocabolario di una lingua B, lo stesso concetto y sia associato allo stesso z concetto.

L'esistenza di una corrispondenza tra le relazioni concettuali di due o più lingue identifica il grado di coincidenza linguistica (Norma Internazionale, Terminologia -Vocabolario, ISO 1087, 1995).

	LANGUAGE A			LANGUAGE B	
CONCEPTS	y	z	CONCEPTS	y	z
y	1	1	y	1	0
z	1	1	z	0	1

Figura 3.3. Esempio di concordanza o non concordanza tra concetti

Fonte: Elaborazione propria

Constatare la corrispondenza tra le associazioni dei concetti quindi può significare effettuare uno studio comparato delle diverse lingue, per individuare un'eventuale inclusione o assenza di coincidenza. Per studiare la relazione tra i concetti e dunque la eventuale coincidenza tra le lingue, si ritiene opportuno considerare il passaggio dalla matrice *documenti x concetti* alla matrice *concetti x concetti*, che meglio indaga questo tipo di relazione.

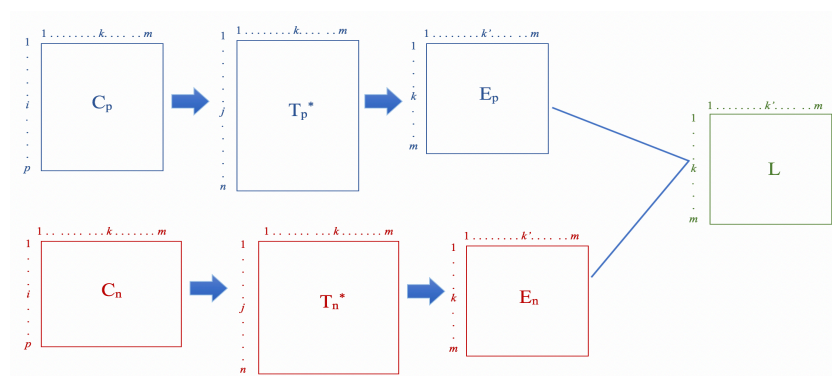


Figura 3.4. Struttura dei dati per la costruzione della matrice concetti x concetti

Fonte: Elaborazione propria

Per la costruzione della matrice *concetti x concetti* positiva E_p e negativa E_n di ciascuna lingua si fa ricorso, come nel caso precedente, ad operazioni tipiche dell'algebra matriciale. Più propriamente, la matrice *concetti x concetti* è una matrice $m \times m$ che si ottiene dalla moltiplicazione della *document-concept matrix* positiva o negativa con la sua trasposta, ossia: $T_p^* T_p^{*T}$ e $T_n^* T_n^{*T}$. Questa rappresentazione permette di indagare le combinazioni di concetti che si presentano più frequentemente, definendo la relazione concettuale in un'ottica di comprensione associativa.

Infine, per ottenere una configurazione matriciale che tenga conto allo stesso tempo sia delle relazioni tra i concetti positivi che delle relazioni tra i concetti negativi, può risultare utile creare un'unica matrice dei concetti (L) unificata, alle cui etichette viene aggiunto un elemento identificativo positivo (+) o negativo (-) allo scopo di conoscerne l'orientamento semantico.

A livello pratico, si costruiscono due blocchi matriciali distinti dove da una parte sono presenti gli incroci tra concetti solo positivi e dall'altra, gli incroci tra concetti solo negativi. Nella stessa matrice, si realizzano poi altri due blocchi distinti, nel cui incrocio tra i concetti positivi e quelli negativi si esprimono solo relazioni nulle. Quest'ultima configurazione matriciale identifica la tabella del *sistema concettuale* positivo e negativo di ogni lingua sul tema della violenza contro le donne.

Per confrontare le relazioni esistenti tra i concetti estratti dal vocabolario di ciascuna lingua, si può procedere ad uno studio comparativo che si basa sul confronto delle matrici *concetti x concetti* unificate (o *matrici del sistema concettuale*) monolingue.

Infatti, l'ultima fase della proposta *cross-language* consiste proprio nel confrontare le matrici *concetti x concetti* unificate di ciascuna lingua, in un comune spazio concettuale multilingue.

3.2.3. Metodi di Matrix Comparison

In letteratura, la nozione di confronto tra matrici copre una serie di tecniche statistiche correlate che mirano ad indagare il grado di somiglianza tra due o più matrici di dati. Si tratta di un dominio di ricerca comunemente utilizzato in biologia, ecologia (Irschick *et al.*, 1997; Damme *et al.*, 1998 citato in Peres-Neto e Jackson, 2001; Poe, 2014), tassonomia numerica, psicologia, geografia

ed antropologia (Sokal e Rohlf, 1962; Pielou, 1984; Legendre & Legendre, 1998).

Tecniche di analisi multidimensionale come l'*Analisi in Componenti Principali* (ACP; Pearson, 1901) o l'*Analisi delle Corrispondenze* (AC; Benzécri, 1973) *Semplici* o *Multiple* indagano *patterns*, associazioni o variabilità interne ai dati per un *unico blocco di dati*, mentre l'analisi comparativa tra matrici identifica metodi di analisi cosiddette *multi-blocco* (o *multi-block*). Tale denominazione è conosciuta in letteratura per indicare *set* di dati multivariati suddivisi in *blocchi* di variabili su uno stesso insieme di osservazioni, con lo scopo di analizzare le relazioni esistenti tra questi blocchi di variabili (Jöreskog, 1971 in De Roover *et al.*, 2012; Westerhuis *et al.*, 1997; Westerhuis *et al.*, 1998; Westerhuis e Smilde, 2001; De Roover *et al.*, 2012; Bougeard e Cardinal, 2014; Bougeard *et al.*, 2018). In particolare, lo studio esplorativo comparativo e sintetico di più matrici si concretizza in un insieme di metodi statistici noti come *Analisi a più vie* (*Multi-way data analysis* (Law *et al.*, 1984), il cui ambito comprende tecniche che mutano a seconda dei diversi contesti di ricerca. Per il confronto tra matrici è frequente il ricorso all'*Analisi Multifattoriale* (*Multiple Factor Analysis*; *MFA*; Escofier e Pagès, 1990) che studia dati *multi-blocco* di varie grandezze sugli stessi oggetti. Ne sono un esempio applicativo i contributi di Nestrud e Lawless (2011), Abdi *et al.* (2013), Kemp *et al.* (2018). L'*MFA* presuppone che i valori di tutte le matrici siano standardizzati e sottoposti a trasformazione lineare come la decomposizione in valori singolari generalizzata. Tale trasformazione permette di ottenere un'unica grande matrice su cui applicare l'ACP, al fine di ottenere una *configurazione di compromesso* o di *consensus*. Per indagare il consenso tra i dati, i punti di ciascun blocco rispetto alla *configurazione di compromesso* possono essere graficamente proiettati.

Altri metodi di analisi *multi-way* mirano ad evidenziare l'esistenza di similarità tra fattori o dimensioni di più *set* di variabili organizzate in più matrici. Ad esempio nell'*Analisi a tre vie* (o *Three-way data analysis*; Bolasco, 1986) la similarità si rintraccia attraverso la matrice delle correlazioni, limitatamente alle prime due componenti principali che spiegano la variabilità totale attraverso gli autovalori. È il caso dello studio presentato da Fraire (2000), in cui l'analisi dell'*interstruttura* si traduce nella distanza rispetto alla matrice media maggiormente rappresentativa, che è un compromesso di sintesi di tutte le matrici prese in esame. Altri studi (ad esempio Fraire, 2007) invece calcolano la matrice delle distanze euclidee fra matrici di correlazione, con lo scopo di

osservare similarità o dissimilarità tra i dati. Ulteriori contributi come ad esempio i lavori presentati da Garcia (2012) e Bookstein & Mitteroecker (2014) fanno ricorso al criterio della covarianza (Tucker, 1958), le cui soluzioni esprimono un legame tra le matrici in termini di inerzia.

Un altro metodo simile alla *MFA* che studia dati *multi-blocco* di varie grandezze sugli stessi oggetti è l'*Analisi di Procuste* (Hurley e Cattell, 1962) che nella versione *Generalizzata* (Gower, 1975) ha l'intento di individuare una configurazione media di consenso, adottando però un approccio statistico differente rispetto a quelli precedentemente menzionati (Host *et al.*, 1993; Perrin *et al.*, 2008; Tomic *et al.*, 2015; Næs *et al.*, 2018), in cui il *set* di dati può essere rappresentato da assi di ordinazione o dati originari (dati grezzi; Buttigieg e Ramette, 2014). Nell'approccio procustiano, le matrici dei dati non vengono trasformate in matrici di distanze o di correlazioni ma si considerano direttamente i dati originari ed il loro possibile ordinamento *traslato*, *ruotato* o *scalato*, al fine di evidenziare il loro massimo accordo in termini di minimi quadrati (Host *et al.*, 1993, Tomic *et al.*, 2015). Lo dimostrano i contributi proposti ad esempio da Gower (1975), Næs *et al.* (2010), Alcalde *et al.* (2014), Paulos, *et al.* (2015), e quello di Tomic *et al.* (2015) che si basa proprio sul confronto tra l'*AMF* e l'*Analisi Procustiana Generalizzata*. Al termine delle *trasformazioni procustiane* (*translation*, *rotation*, *scaling*), la metrica di similarità o dissimilarità tra le matrici e gli oggetti è rappresentata dalla somma dei residui al quadrato. Al contrario di altre tecniche sopramenzionate, per determinare la configurazione di consenso nell'*Analisi Procustiana Generalizzata* si considerano tutte le dimensioni che hanno contribuito a definire il compromesso medio e si ottengono i valori di variazione residua in termini di minimi quadrati per ciascuna dimensione (Host *et al.*, 1993). Rispetto alle classiche procedure di trasformazione, uno dei vantaggi dell'*Analisi Procustiana* è il ricorso alla decomposizione spettrale delle matrici, ricavata dal prodotto delle matrici originarie che contengono i valori delle coordinate dei punti da mettere in relazione. Questa condizione rende possibile il calcolo diretto dei parametri di trasformazione, senza dover studiare misure di prossimità sottostanti. Inoltre, tale metodo permette di schivare eventuali problemi connessi alla linearizzazione algebrica di combinazioni non lineari, mediante il ricorso a procedure di iterazione che permettono una notevole contrazione dei tempi di calcolo e laboriosità analitica. Da un punto di vista geometrico, l'*Analisi Procustiana Generalizzata* restituisce una mappa fattoriale su cui visualizzare i dati, dove ciascun oggetto è posizionato nello

spazio dimensionale in base al proprio centroide. Quindi è facilmente visibile, in modo separato, la concordanza delle osservazioni per ciascuna dimensione, agevolando l'interpretazione della struttura dei dati.

Per indagare la reale congruenza tra matrici di dati multivariati, gli approcci basati sulla correlazione sono soliti utilizzare una *statistica test* cosiddetta di Mantel (1967) per validare il confronto tra due o più *set* di matrici (ad esempio i contributi di Douglas e Endler, 1982; Jackson e Harvey 1989; Thorpe *et al.*, 1996; Douglas *et al.*, 1999; Dutilleul *et al.*, 2000). Nello specifico, si tratta di determinare una matrice di distanze a coppie che esprime la differenza tra ogni duo di oggetti in uno spazio multivariato. La *statistica test* di Mantel permette di calcolare la correlazione tra le due distanze matriciali tenendo conto di possibili implicazioni sull'ordinamento dei dati. Come supplemento o in alternativa si può utilizzare direttamente l'analisi di *Procuste* che contrariamente al *test* di Mantel individua quanta varianza presente in una matrice sia attribuibile alla varianza presente in un'altra matrice (Peres-Neto e Jackson, 2001; Schneider e Borlund, 2007). L'ispezione visiva *dell'Analisi di Procuste* è un vantaggio aggiuntivo, poiché sono mappati i punti residui di ciascuna matrice, consentendo l'identificazione dei singoli oggetti che mostrano concordanza (relativamente) insolita (Jackson, 1995 citato in Buttigieg e Ramette, 2014).

3.2.4. Il confronto tra matrici di concetti monolingue in uno spazio dimensionale multilingue attraverso l'Analisi di Procuste

In base alle caratteristiche che il metodo procustiano presenta (§ 3.2.3), si ritiene opportuno adottare proprio l'*Analisi di Procuste* per l'ultima fase della strategia *cross-language*. La *performance* di tale tecnica per la comparazione tra le lingue risulta da prediligere soprattutto da un punto di vista metrico, poiché consente di operare sui dati originari delle matrici dei concetti senza la necessità di sottoporre i dati a trasformazioni di algebra lineare. Inoltre, la prospettiva computazionale è agevolata dalla procedura iterativa che riduce tempo e complessità analitica. Infine, i risultati che ne derivano sia da un punto di vista geometrico che algebrico consentono di individuare eventuali similarità o dissimilarità tra le lingue, permettendo altresì di osservare i singoli concetti

che contribuiscono ad un'eventuale concordanza o discordanza tra le configurazioni matriciali di ogni lingua.

La denominazione dell'*Analisi Procustiana* è da ricondurre alla mitologia greca e si riferisce a *Procrustes* (dal greco “Προκρούστης”), il soprannome attribuito ad un brigante dell'antica Grecia il cui nome era Damaste. L'uomo straziava le proprie vittime distendendole su un letto di ferro (Diodorus Siculus, 1935), sul quale verificava che la statura dei malcapitati corrispondesse esattamente alla dimensione del giaciglio. Se così non era, con lo scopo di eguagliare gli arti delle vittime all'incudine, *Procruste* “stirava”, amputava e torturava le loro membra a martellate, fino a condurre le proprie vittime alla morte (Plutarch, 1914). È su questo presupposto etimologico che inizialmente l'*Analisi Procustiana* si riferisce alla *shape analysis* (Dryden e Mardia, 2016) e mira al confronto tra le forme per esplorarne la struttura di variabilità. Più propriamente, il termine *Analisi Procustiana* vuole indicare un insieme di metodi e tecniche matematiche per operare trasformazioni fra *set* di dati o matrici di coordinate, appartenenti ad un generico spazio p -dimensionale, al fine di ottenere la loro massima concordanza (Hurley e Cattel, 1962; Sneath, 1967; Gower, 1975).

Il metodo scientifico è finalizzato al *matching* tra configurazioni geometriche, per mezzo di trasformazioni che tendono a minimizzare il più possibile la distanza tra una forma e l'altra. Il confronto tra le forme viene eseguito in uno spazio euclideo, utilizzando il metodo dei minimi quadrati. L'*Analisi Procustiana* è una tecnica utilizzata in diversi domini di ricerca:

- in campo biometrico per il riconoscimento e la classificazione genetica e molecolare;
- in morfometria, per indagare eventuali deformazioni o variazioni in organi e tessuti come informazione patologica;
- in informatica, per l'*image matching*;
- in statistica, per gli studi multifattoriali.

Il suo modello può essere applicato a tutte le situazioni reali in cui si necessita di confrontare in maniera diretta diverse configurazioni di uno stesso oggetto di studio. Essenzialmente, l'analisi di *Procruste* si basa su trasformazioni ortogonali (Schönemann, 1966; Crosilla, 1983) che mirano a determinare i criteri della trasformazione che meglio adattano la sovrapposizione

(*superimposition*) della prima configurazione sulla seconda, nel rispetto del principio dei minimi quadrati e della condizione di ortogonalità¹⁴. Le tecniche procustiane ortogonali si considerano come modelli matematici che operano trasformazioni tra coppie di matrici, indicate rispettivamente come *matrice origine* o *sorgente* e *matrice destinazione* o *obiettivo*.

L'*Analisi Procustiana Ortogonale* che analizza solo due configurazioni di dati è detta *Analisi Procustiana Classica*, mentre, se nel processo di adattamento intervengono più di due configurazioni matriciali l'*Analisi Procustiana* è detta *Generalizzata* (Gower, 1975).

Nei paragrafi successivi, la trattazione teorica dell'*Analisi Procustiana Classica* si limiterà ai punti essenziali della metodologia, in quanto è l'*Analisi Procustiana Generalizzata* a rappresentare il *core* di interesse del lavoro di tesi. Infatti, nell'ultima fase della strategia *cross-language* proposta, il confronto tra le lingue si realizza proprio attraverso l'applicazione dell'*Analisi Procustiana Generalizzata* per confrontare le matrici *concetti x concetti* delle diverse lingue.

3.2.4.1. L'*Analisi Procustiana Classica*

L'origine terminologica dell'*Analisi Procustiana Classica* è riconducibile ad Hurley e Cattell (1962), tuttavia alcuni autori come Goodall (1991) attribuiscono a psicometristi come Mosier (1939) e morfometristi come Boas (1905) la sua reale formalizzazione tecnica.

L'*Analisi Procustiana Classica* risponde all'obiettivo di indagare il massimo accordo (*maximal agreement*) sempre e solo tra due *set* di dati ed esegue le trasformazioni di similitudine al fine di minimizzare la distanza Euclidea quadratica tra le due configurazioni. La trasformazione di similitudine di una matrice è data da un insieme di operazioni: *traslazione*, *rotazione* (e *reflection*) e *trasformazione scalare* (o *scaling*). Tali interventi di trasformazione mirano a minimizzare al massimo la distanza tra due configurazioni, consentendone il confronto nello stesso spazio dimensionale.

Più è piccola la distanza e più le due configurazioni possono dirsi simili tra loro.

¹⁴ In algebra lineare, il concetto di trasformazione lineare indica il movimento di un oggetto o di una figura geometrica in uno spazio metrico euclideo che preserva le distanze.

Si considerino $m=2$ configurazioni matriciali *concetti x concetti* ($m \times m$), rispettivamente \mathbf{X}_1 e \mathbf{X}_2 di m punti in uno spazio p -dimensionale. Ciascuna matrice $\mathbf{X}_{k'}$ ($k'=1, 2, \dots, m$) è costituita da m righe e $m_{k'}$ colonne, dove la k -esima riga esprime le coordinate di un punto $P_k^{(k')}$ ($k'=1, 2, \dots, m; k=1, 2, \dots, m$). Con l'obiettivo di voler sovrapporre e quindi adattare la seconda matrice (\mathbf{X}_2) sulla prima (\mathbf{X}_1), Schönemann & Carroll (1970) e dopo Gower (1971) eseguono le trasformazioni ortogonali affinché la somma delle differenze al quadrato tra gli elementi di \mathbf{X}_2 e gli elementi di \mathbf{X}_1 sia la più piccola possibile. A tale scopo si utilizza il criterio dei minimi quadrati per misurare il grado di adattamento tra le due configurazioni di concetti, ovvero:

$$m_{12^2} = \sum_{k=1}^m \Delta^2 \left(P_k^{(1)}; P_k^{(2)} \right) \quad [3.8]$$

dove $\Delta \left(P_k^{(1)}; P_k^{(2)} \right)$ rappresenta la distanza Euclidea tra coppie di punti appartenenti a \mathbf{X}_2 e \mathbf{X}_1 .

Le coordinate delle due matrici vengono centrate rispetto all'origine e rappresentate come matrici di deviazione o scarti al quadrato \mathbf{X}'_1 e \mathbf{X}'_2 . Tale operazione rappresenta la *traslazione* delle configurazioni matriciali.

Dopo aver sovrapposto i centroidi (*centroid size*, Bookstein, 1991) delle due matrici, per un migliore adattamento si ricorre ad un movimento di *rotazione* (e *reflection*) della seconda matrice sulla prima, così che $\mathbf{X}_2^* = \mathbf{X}'_2 \mathbf{H}$, dove \mathbf{H} è la matrice di rotazione. In questo modo, la somma delle deviazioni al quadrato è minima rispetto alla matrice di riferimento \mathbf{X}_1 .

La matrice di rotazione \mathbf{H} che meglio performa la sovrapposizione di $\mathbf{X}'_2 \mathbf{H}$ su \mathbf{X}_1 è data da:

$$\mathbf{H} = \mathbf{V}' \mathbf{U} \quad \text{dove } \mathbf{X}'_1 \mathbf{X}_2 = \mathbf{U}' \mathbf{\Gamma} \mathbf{V} \quad [3.9]$$

come decomposizione in valori singolari di Eckart-Young (1935) e con \mathbf{U} e \mathbf{V} ortogonali e $\mathbf{\Gamma}$ diagonale.

La stima dei minimi quadrati attraverso il *fattore di scala* si ottiene adattando \mathbf{X}_2 su \mathbf{X}_1 :

$$\rho = \text{tr}(\mathbf{X}_2 \mathbf{H} \mathbf{X}_1') / \text{tr}(\mathbf{X}_2 \mathbf{X}_2') \quad [3.10]$$

Una precisazione è doverosa nell'ambito della sovrapposizione di una configurazione sull'altra: qualora l'ordine delle matrici venisse invertito, i parametri di somiglianza non restano uguali (Gower, 1975). Se si sovrappone la matrice \mathbf{X}_2 sulla matrice \mathbf{X}_1 si ha una trasformazione algebrica e geometrica che differisce dall'adattamento della matrice \mathbf{X}_1 sulla matrice \mathbf{X}_2 ¹⁵.

3.2.4.2. L'Analisi Procustiana Generalizzata

Nel caso in cui il confronto riguarda più di due matrici di dati e quindi $m > 2$, i principi dell'*Analisi Procustiana* sono generalizzati, determinando così l'affermarsi dell'*Analisi Procustiana Generalizzata* o *Generalized Procrustes Analysis* (GPA; Gower, 1975).

La sua formulazione si deve a Gower (1975) sulla base di una precedente formalizzazione sulla trasformazione ortogonale per più di due matrici avanzata da Kristof e Wingersky (1971). Tale tecnica permette di *traslare, ruotare e scalare* simultaneamente un numero ipoteticamente infinito di *set* di dati, restituendo un'unica configurazione media.

Essa trova ampio spazio nel campo delle analisi sensoriali (Williams e Langron, 1984; Arnold e Williams, 1986; Pagès, 2014) e viene infatti adottata nelle analisi descrittive del *Free Choice Profiling* (FCP; Williams e Langron, 1984; Arnold e Williams, 1986; Jack e Piggott, 1992), nelle quali un *panel* di valutatori esprime un giudizio su un *set* di attributi relativi ad uno o più prodotti e tipicamente la GPA è usata con lo scopo di confrontare le differenze nei punteggi attribuiti agli *items*, restituendo così una configurazione di consenso delle valutazioni.

Si consideri ad esempio il caso di $m=3$ matrici delle relazioni tra i concetti, $m \times m$, dove ciascuna configurazione $\mathbf{X}_{k'}$ ($k'=1, 2, \dots, m$) presenta il k -esimo concetto di riga come le coordinate del punto $P_k^{(k')}$ ($k'=1, 2, \dots, m; k=1, 2, \dots, m$) in uno spazio Euclideo. Piuttosto che considerare la distanza euclidea tra

¹⁵ Una dimostrazione algebrica e geometrica di tale principio è presente in Dryden, I.L., Mardia, K.V. (2016). *Statistical shape analysis, Wiley series in probability and statistics, (1^a ed.1998), Capitolo 7, pp.129-130.*

tutte le possibili coppie delle m configurazioni ($m_{12}, m_{13}, m_{23}, \dots, etc$ [3.8]), la *GPA* consente di *traslare*, *ruotare* e *scalare* singolarmente e contemporaneamente tutte le configurazioni nello stesso spazio dimensionale, con l'obiettivo di fornire una corrispondenza in termini di minimi quadrati (Gower, 1975; Goodall, 1991; Borg e Groenen, 1997; Dryden e Mardia, 2016). La comparabilità è detta *ottimale* ed il miglior adattamento conforme ai minimi quadrati restituisce una configurazione media complessiva dei dati che è conosciuta come *consensus*. Il *consensus* identifica la rappresentazione intermedia tra le configurazioni date e i criteri delle varie trasformazioni. L'intento di minimizzare la distanza Euclidea quadratica per la *GPA* si concretizza attraverso un criterio della bontà di adattamento che Gower (1975) formalizza nel seguente modo:

$$\sum_{k'=1}^m \sum_{k=1}^m \Delta^2 (P_k^{(k')}; G_k) \quad [3.11]$$

dove G_k rappresenta il cosiddetto "centroide" dell'insieme dei punti $P_k^{(k')}$ ($k'=1, 2, \dots, m$) che costituiscono e caratterizzano le diverse matrici.

Da un punto di vista geometrico, il metodo permette di rappresentare la pluralità delle matrici m (dove $m > 2$) in uno spazio dimensionale comune, minimizzando la somma dei quadrati tra gli mm punti e la rappresentazione media o centroide di ogni concetto (G_k).

La distanza tra ciascun punto $P_k^{(k')}$ ed il concetto centroide G_k è chiamata *residuo*, mentre la *somma dei residui al quadrato* viene indicata come S_r e si calcola attraverso la [3.11].

Il modello di rappresentazione geometrica della *GPA* viene presentato a seguire in un esempio grafico che mostra $m=3$ matrici bidimensionali di $m=4$ concetti, dove $p_1= p_2= p_3= p_4=2$ sono i punti delle mm configurazioni che si riferiscono allo stesso k -esimo concetto.

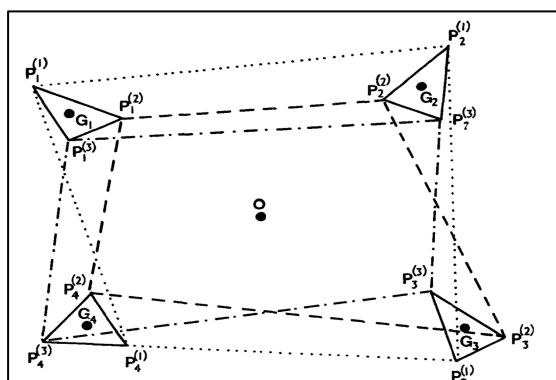


Figura 3.5. Rappresentazione geometrica della GPA.

Fonte: Gower (1975).

La visualizzazione grafica (Figura 3.5) mostra come ciascun concetto origini un gruppo triangolare di vertici, in cui è possibile identificare i centroidi G_k ($k=1,2,3,4$) dei diversi punti.

Il centroide dell'intero sistema è invece in riferimento a 0 ed il criterio della bontà di adattamento [3.11] minimizza la somma dei quadrati tra ciascun cluster di punti $P_{(k)}^{(k')}$ ($k'=1,2, \dots, m$) appartenenti a m ed i loro centroidi G_k , sommati a tutti gli m gruppi.

Come nel caso dell'Analisi Procustiana Classica, anche per la GPA le trasformazioni procustiane si originano dall'operazione di traslazione, la quale implica che le m configurazioni $X_{k'}$ ($k'=1,2, \dots, m$) vengano centrate, come matrici di deviazione o scarti al quadrato, $X'_{k'}$.

Le matrici centrate $X'_{k'}$ permettono di ottenere un baricentro che senza perdita di generalità per ogni matrice viene convenzionalmente scelto come origine, affinché le configurazioni vengano sovrapposte in base al baricentro.

Poiché i dati possono presentare diverse grandezze in letteratura è consigliato riscalarci ciascuna matrice $X'_{k'}$ ¹⁶ attraverso la formula seguente:

$$\sum_{k'=1}^m \lambda^{17} \text{tr}(X_{k'} X'_{k'}) = \sum_{k'=1}^m \text{tr}(X_{k'} X'_{k'}) = m \quad [3.12]$$

¹⁶ Per tornare alle unità originali, i risultati devono essere moltiplicati per il fattore di ridimensionamento.

¹⁷ λ indica il fattore di ridimensionamento.

Lasciando inalterata la posizione del baricentro si procede al movimento di *rotazione*. Le *rotazioni* iniziali si calcolano come nel caso dell'*Analisi Procustiana Classica*, ovvero si ruota \mathbf{X}_2 su \mathbf{X}_1 e si ottiene \mathbf{Y} come matrice dei centroidi delle m coppie di vertici $P_{(k')}^{(1)}, P_{(k')}^{(2)}$ ($k' = 1, 2, \dots, m$). In questo modo si assegnano le nuove posizioni di \mathbf{X}_1 e \mathbf{X}_2 . Il movimento di rotazione viene eseguito per ciascuna matrice $\mathbf{X}_{k'}$ dei concetti al fine di adattarla al corrente valore di \mathbf{Y} , il quale viene ricalcolato dopo ogni rotazione. Tale operazione consente così di ottenere i valori iniziali di ciascuna matrice *concetti* \times *concetti* $\mathbf{X}_{k'}$ e delle loro medie \mathbf{Y} . Quindi, la matrice \mathbf{Y} non è altro che la tabella delle medie iniziali di $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{k'}$.

A questo punto è possibile valutare la somma dei residui al quadrato di partenza, mediante:

$$S_r = m (1 - tr(\mathbf{Y}\mathbf{Y}')) \quad [3.13]$$

La formula mantiene fissa la matrice \mathbf{Y} e ruota nuovamente ciascuna matrice $\mathbf{X}_{k'}$, dopo che questa, a seguito della prima rotazione, ha acquisito la nuova posizione. Lo scopo è di adattare ogni matrice $\mathbf{X}_{k'}$ alla matrice delle medie \mathbf{Y} . Infatti, dopo aver ruotato tutte le m configurazioni, si ricalcola la matrice \mathbf{Y}^* che registra i nuovi centroidi, nonché le nuove posizioni acquisite dalle matrici $\mathbf{X}_{k'}$ dopo la successiva rotazione. In questo caso, la nuova somma dei residui al quadrato è pari a:

$$S_r^* = S_r - m tr(\mathbf{Y}^*\mathbf{Y}^{*'} - \mathbf{Y}\mathbf{Y}') \quad [3.14]$$

Quando la matrice delle medie iniziali (\mathbf{Y}) viene sostituita con la matrice dei nuovi centroidi (\mathbf{Y}^*), la somma dei residui al quadrato si riduce ulteriormente. Secondo Kristof & Wingersky (1971) se dopo la procedura di *rotazione*, il valore di S_r non ha subito un cambiamento considerevole rispetto ad una specifica soglia prefissata, il processo di *rotazione* risulta terminato. In caso contrario, si procede con l'iterazione fino a convergenza globale.

Tuttavia, secondo Gower (1975) per dirsi efficacemente convergenti le trasformazioni procustiane devono essere complete e quindi bisogna includere nelle operazioni il *fattore di scala* $\rho_{k'}$.

Nello specifico, se si considera $\mathbf{X}_{k'} (k'=1,2, \dots, m)$ con media \mathbf{Y} , le stime dei minimi quadrati dei coefficienti di scala $\rho_{k'}$ che meglio si adattano ad \mathbf{Y} sono:

$$X_{k'}' = \rho_{k'} \mathbf{X}_{k'} \mathbf{H}_{k'} \quad [3.15]$$

attraverso cui si ruota il fattore di scala $\rho_{k'}$ per ogni $k'=1,2, \dots, m$, adattandolo a \mathbf{Y} . Mentre, $\mathbf{X}_{k'} \mathbf{H}_{k'}$ può essere sostituita da una matrice $\mathbf{X}_{k'}^*$ che senza perdita di generalità esprime le posizioni ruotate correnti interpretate come configurazioni iniziali.

Se $\mathbf{X}_{k'}^*$ possiede già un coefficiente di scala $\rho_{k'}$, il nuovo fattore a scalare è $\rho_{k'}^*$ e si misura con la formula che segue:

$$\rho_{k'}^* X_{k'}^{**} = \frac{\rho_{k'}}{\rho_{k'}} (\rho_{k'} \mathbf{X}_{k'}^*) \quad [3.16]$$

Si ottengono così i nuovi valori $\rho_{k'}^* X_{k'}^{**}$ per ogni configurazione $\rho_{k'} \mathbf{X}_{k'}^*$.

A questo punto, si calcola la nuova media \mathbf{Y}^{**} che è pari a:

$$\mathbf{Y}^{**} = \sum \rho_{k'}^* \frac{X_{k'}^{**}}{m} \quad [3.17]$$

Mentre la nuova somma dei residui al quadrato è:

$$S_r^{**} = S_r^* - m (Y^{**} Y^{**} - Y^* Y^{*'}) = S_r^* - m \text{tr}(Y^{**} Y^{**'} - Y Y')$$
 [3.18]

Se si considera $S_r - S_r^{**} <$ ad una soglia di tolleranza prestabilita, il processo di iterazione può dirsi concluso e si ottiene la rappresentazione media ottimale del *consensus*.

In particolare, nel presente lavoro di tesi, l'algoritmo che esegue le trasformazioni ortogonali delle matrici *concetti x concetti* di ciascuna lingua consente un numero massimo di iterazioni possibili, adottando un criterio d'arresto che interrompe il processo quando i valori successivi di S_r differiscono di meno 0.0001.

Da un punto di vista geometrico il *consensus* si osserva attraverso una mappa fattoriale, nella quale ogni concetto identifica un centroide attorno al quale si originano le proiezioni delle diverse lingue.

La vicinanza o distanza tra le proiezioni delle lingue esprimono accordo o disaccordo in merito a ciascun concetto preso in esame. Più propriamente, lo spazio dimensionale multilingue del *consensus* agevola l'intelligibilità grafica dei risultati, evidenziando quale sia o quali siano i concetti che mettono più d'accordo o meno d'accordo l'opinione pubblica sul fenomeno della violenza contro le donne. L'accordo ed il disaccordo tra le lingue possono inoltre essere utilizzati per cogliere la somiglianza tra lingue, la cui similarità è traducibile anche in termini di affinità socio-culturale¹⁸.

Da un punto di vista algebrico invece il consenso tra gli mm punti si può indagare attraverso la proporzione di variabilità dei dati. Infatti, Gower (1975) suggerisce di interpretare i risultati combinando la visualizzazione fattoriale procustiana (come in *Figura 3.5*) con l'*Analisi della Varianza (ANOVA)*, che nel caso della *GPA* è conosciuta come *Generalized Procrustes Analysis of Variance* o meglio *PANOVA*¹⁹ (Commandeur, 1991).

Come noto la *Varianza* può essere divisa in due componenti: *Varianza Between* e *Varianza Within*. La *Between-group* rappresenta il contributo dei termini di traslazione, mentre la *Within-group* a sua volta si divide in *consensus* ed elementi residui.

Essa è prevalentemente utile per identificare l'importanza dei concetti che vanno a costituire la somma dei residui al quadrato totale (S_r). Nello specifico, (S_r) misura i valori residui per ciascun gruppo di consenso [3.19₍₁₎] e per ciascun punto delle configurazioni [3.19₍₂₎]:

$$\sum_{k=1}^m \Delta^2 = (P_k^{(k')}, G_k) (k' = 1, 2, \dots, m)_{(1)} ; \quad [3.19]$$

$$\sum_{k'=1}^m \Delta^2 = (P_k^{(k')}, G_k) (k = 1, 2, \dots, m)_{(2)}$$

Quindi, con l'*Analisi della Varianza* si può stabilire quali concetti esprimano un maggiore o minore consenso (accordo) attraverso i valori residui [3.19₍₂₎]: più i

¹⁸ A tal proposito si rimanda al capitolo precedente, § 2.1: Sapir e Worf in Carassai e Crucianelli (2017).

¹⁹ Per la sua interpretazione riferirsi a Gower (1975); Dijksterhuis e Punter (1990) e Commandeur (1991).

valori residui sono alti e più esiste un basso consenso, mentre più i valori residui sono bassi e più sussiste un alto consenso.

Questo significa che se i valori residui sono elevati ed il *consensus* di un concetto è basso, vi è una maggiore differenza tra le lingue e quindi un minore accordo nel modo in cui i concetti si relazionano in ciascun vocabolario monolingue. Al contrario, se i valori residui sono bassi ed il *consensus* è alto, significa che un dato concetto esprime un buono o massimo accordo tra le lingue. Oltre a studiare i singoli concetti che contribuiscono a determinare la somma dei residui al quadrato totale e quindi il singolo consenso, è altresì possibile attraverso la [3.19₍₁₎] indagare nel complesso quali siano le lingue più simili tra loro, in base a tutti i concetti. Più precisamente, la [3.19₍₁₎] aiuta proprio ad identificare quali lingue presentino un maggiore consenso generale.

Per completare il processo di analisi procustiana, in letteratura si predilige eseguire la *performance* del *permutation test* (Walkeling *et al.*, 1992), come strumento di *significatività* del *consensus*. In particolare, la statistica usata per valutare il livello di concordanza è la proporzione di varianza spiegata dal *consensus* che viene identificata come R_c . Il test consente di determinare se i valori osservati di R_c siano significativamente superiori al 95% di quelli ottenuti con la permutazione dei dati, se confrontati con una distribuzione randomizzata (King e Arents, 1991).

Pertanto, nella fase conclusiva della strategia *cross-language* proposta nel corrente lavoro di tesi si intende eseguire proprio il *permutation test* della *GPA*, al fine di verificare che il *consensus* ottenuto tra le lingue sia statisticamente significativo.

3.3. Cenni conclusivi

Al fine di proporre una strategia di analisi multilingue comparativa sul tema della violenza contro le donne, il terzo capitolo ha evidenziato l'efficacia della comunicazione partecipativa che caratterizza il dominio dei Social Media ed in particolare per lo studio della percezione pubblica sul fenomeno oggetto di studio. Nel contempo, la trattazione del capitolo ha esposto nel dettaglio le fasi di pre-elaborazione e trattamento automatico dei dati testuali. Gli aspetti connessi alle scelte del ricercatore su come processare i dati rivestono un ruolo chiave per l'efficacia dei risultati, quindi in questo capitolo si è voluta

sottolineare l'importanza di operare scelte di *pre-processing* e *feature extraction* che siano appropriate, sebbene dipendano da considerazioni del tutto individuali. Uno degli aspetti maggiormente rilevanti quando si studia l'opinione pubblica è anche il problema dell'ambiguità semantica, poiché le parole possono essere utilizzate in maniera differente in riferimento ad una molteplicità di scenari emotivi diversi. Risulta quindi importante, al fine di cogliere il senso globale della comunicazione, poter risalire dapprima al loro orientamento semantico. Le implicazioni che ne derivano si amplificano se si considera la comunicazione digitale calata in un contesto multilingue, poiché si aggiungono una serie di fattori strettamente connessi alle peculiarità di ciascuna lingua. Solo dopo aver affrontato con scrupolosa meticolosità questi aspetti, in tale contesto, è stata delineata da un punto di vista teorico e con il supporto della letteratura afferente, la strategia di analisi *cross-language* che si intende adottare su dati reali nel prossimo capitolo. Nello specifico, la proposta è stata sviluppata in tre fasi principali. Il primo passo consiste proprio nella definizione delle operazioni di *pre-processing* e *feature extraction* di ciascun *corpus* monolingue, con lo scopo di mappare la percezione del fenomeno per ogni lingua. Tale analisi può indicare, in via preliminare, una sintesi del contenuto di ciascun *corpus*. I risultati dell'analisi del testo che vengono proposti in questa prima fase di *Text summarization* considerano come unità minimale di senso la parola o *forma grafica*.

Nella seconda fase della strategia invece, al fine di assicurare la non *artificialità* delle caratteristiche di ciascuna lingua, si è proposto di individuare un'unità di analisi alternativa alla *forma grafica* e considerata più opportuna per il successivo confronto multilingue, perché di ordine superiore: i cosiddetti concetti. Specificamente, la determinazione dei concetti per ogni lingua si concretizza attraverso l'adozione consequenziale di due tecniche di analisi multidimensionale: l'*AC* la *Cluster Analysis*. Applicando la *Cluster Analysis* nel sottospazio dimensionale dell'*AC* si determinano gruppi di concetti monolingue che categorizzano in *cluster* di concetti le diverse *features*. Tuttavia, si è inteso dapprima assicurare l'orientamento semantico delle *features* attraverso il calcolo della polarità, così da ottenere gruppi di *features* con la stessa direzione semantica. Poiché si opera in un contesto multilingue, è stato necessario individuare un *lexicon* adatto per più lingue, nello specifico per le sue caratteristiche e versatilità d'utilizzo su *software* e *packages* è stato introdotto l'*NRC Emotion Lexicon*. Il calcolo della polarità che si intende operare su ciascuna collezione di *tweets* è a livello di documento, in modo da

ottenere da una parte un *dataset* positivo e dall'altro un *dataset* negativo, per ciascuna lingua. Le collezioni di *tweets* positive e quelle negative vengono processate separatamente, così da ottenere da una parte solo *features* positive e dall'altra solo *features* negative. Dall'estrazione di *features* positive e negative si genera una matrice lessicale monolingue positiva e una matrice lessicale monolingue negativa, sulle quali attraverso la combinazione dell'*AC* e della *Cluster Analysis* si ottengono classi di concetto positive e classi di concetto negative. Determinato quindi lo *spazio concettuale* delle diverse lingue, si codificano i dati in modo da costruire un'unica matrice dei concetti in base all'appartenenza delle *features* ai diversi *tweets*.

Attraverso operazioni di algebra matriciale si definisce poi la costruzione di una matrice *concetti x concetti* che indaga le relazioni tra i diversi concetti.

Sulla base di un'ampia valutazione dei metodi statistici esistenti per il confronto tra più *set* di dati e a fronte delle caratteristiche teoriche e pratiche, si ritiene opportuno adottare per l'ultima fase della strategia, ovvero il confronto delle diverse matrici *concetti x concetti* monolingue, l'*Analisi procustiana*. Nello specifico, quella *generalizzata (GPA)* rappresenta lo strumento cruciale per il confronto tra più lingue in unico spazio dimensionale multilingue, permettendo di conservare al massimo l'originalità dei dati di partenza.

La tecnica confronta simultaneamente l'accordo o il disaccordo tra le diverse matrici attraverso una configurazione media multilingue. Per l'interpretazione della rappresentazione media o di *consensus* da un punto di vista geometrico si ricorre alla mappa fattoriale, mentre da un punto di vista algebrico si consulta l'*Analisi della Varianza (PANOVA)*.

La mappa fattoriale determina per ogni concetto le proiezioni di trasformazione procustiana delle diverse lingue permettendo di osservare il massimo accordo o disaccordo tra le lingue per ogni concetto. L'interpretazione algebrica della *Varianza* invece supporta la definizione di un eventuale consenso tra le lingue, consentendo quindi di individuare le configurazioni più simili tra le lingue. La similarità o dissimilarità può essere letta sia in termini linguistico-concettuali che socio-culturali.

In definitiva, le tre fasi combinate della strategia *cross-language* proposta in questo capitolo consentono di raggiungere la finalità di un confronto tra più lingue senza ricorrere a dizionari bilingue, operazioni di traduzione manuale o di *machine translation* e senza la realizzazione di corpora paralleli.

Il prossimo capitolo illustrerà i risultati dell'applicazione della strategia multilingue su una collezione di dati reali, al fine di dimostrare l'efficacia della

proposta. Nello specifico, lo studio che verrà presentato confronterà i linguaggi e le culture, analizzando la percezione pubblica dei contenuti testuali estratti da Twitter in cinque diverse lingue europee.

Capitolo Quarto

I risultati della *Cross-language Analysis* sul fenomeno della violenza contro le donne

Il quarto capitolo ha lo scopo di presentare i risultati dell'applicazione della strategia multilingue delineata nel capitolo precedente. Tuttavia, in una fase preliminare del medesimo capitolo risulta necessario introdurre i metodi, le tecniche e i parametri di rilevazione dei dati. In particolare, nel caso di fattispecie, per indagare il tema della violenza sulle donne, l'estrazione dei *tweets* si concretizza attraverso il ricorso ad una o più *parole chiave* attinenti al fenomeno. Il multilinguismo e multiculturalismo esistenti presuppongono la presenza di variabili socio-culturali che influenzano la *notiziabilità* degli argomenti, pertanto, si reputa opportuno affidare la selezione delle *keywords* a fonti ufficiali e specializzate. In base alla disponibilità di tali fonti, nel presente studio è stata estratta ed analizzata la percezione pubblica dei contenuti estratti da Twitter in cinque diverse lingue europee: italiano, inglese, francese, spagnolo ed ucraino. L'utilizzo delle tecniche di *Text Mining* consente di esplorare il vocabolario di ciascun *corpus* monolingue rispetto alla collezione di *tweets cross-language*, al fine di mappare la percezione sul tema preso in esame. Per mantenere inalterate le peculiarità delle lingue originali e per limitare la variabilità tra le lingue, si considera un'unità di analisi di ordine concettuale. Con lo scopo di determinare tali concetti, per ogni collezione di dati monolingue si determina l'orientamento semantico attraverso l'*NRC Emotion Lexicon*, così da estrarre solo *features* positive e solo *features* negative. In questo modo, su ciascuna collezione di *tweets* divisa per polarità semantica, la combinazione di due tecniche di analisi multivariata, *Analisi delle Corrispondenze* e *Cluster Analysis*, consente la *Concepts Extraction*. Si tratta di ottenere raggruppamenti di classi concettuali costituiti da *features* positive e raggruppamenti di classi concettuali costituiti da *features* negative. Al fine di indagare la relazione tra i concetti delle cinque diverse lingue, si codificano i dati della classificazione in una matrice *concetti x concetti*. Infine, per studiare la similarità o dissimilarità tra le associazioni concettuali delle diverse lingue si applica l'*Analisi Procustiana Generalizzata* (GPA) che ha lo scopo di confrontare simultaneamente in unico piano fattoriale *cross-language* le cinque

configurazioni dei concetti, determinando i concetti di maggiore *consensus* e l'accordo e/o disaccordo tra le lingue.

4.1. L'estrazione dei dati: tecniche, metodi e parametri di rilevazione

Nel corso degli ultimi anni, molte piattaforme di Social Media hanno consentito lo sviluppo di un sistema interconnesso di applicazioni per l'accesso ai dati, con lo scopo di estrarre ed analizzare²⁰ una parte della quantità virtualmente infinita di dati generati dagli utenti sul *Web*. Tuttavia, dopo lo scandalo *Cambridge Analytica* (Tirino, 2019) lo scenario è decisamente mutato e sono state apportate restrizioni che hanno modificato metodi e tecniche di rilevazione dei dati. Attualmente le procedure di *data extraction* vengono strutturate nel rispetto del *GDPR* (*General Data Protection Regulation*)²¹ che promulgato nel 2018 ha abrogato precedenti articoli o codici incompatibili in materia di protezione dei dati personali, modificando altresì le *policies* dei contratti di licenza dei Social Media fino a limitare o interrompere la procedura di estrazione dei dati. Malgrado i limiti imposti dal *GDPR*, Twitter è una tra le piattaforme a consentire attualmente la rilevazione dei dati e ad esprimere un allineamento tra i principi di tutela della *privacy* e la propria *policy*²², in conformità con l'articolo 5, paragrafo 1, lettera b). Nello specifico, la fase di *data extraction* avviene in maniera autorizzata attraverso l'accesso ad un'interfaccia di programmazione delle applicazioni conosciuta con l'acronimo di *API* (*Application Programming Interface*). Nel linguaggio informatico, l'*API* indica un insieme di procedure che consente ad entità esterne come *software* o *packages* di linguaggi di programmazione²³ di eseguire operazioni su una piattaforma di cui sono sconosciute le informazioni implementative. Infatti,

²⁰ L'estrazione e l'analisi dei dati provenienti dai Social Media è finalizzata a rispondere ad interrogativi perlopiù di natura scientifica o commerciale (Di Donato, 2013).

²¹ Unione Europea (2016). Regolamento del Parlamento Europeo e del Consiglio d'Europa in materia di trattamento dei dati personali: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN>.

²² Sezione 1.2. del regolamento sulla *privacy* di Twitter: <https://twitter.com/it/privacy>.

²³ <https://docs.microsoft.com/en-us/dotnet/visual-basic/programming-guide/com-interop/walkthrough-calling-windows-apis>.

previa autorizzazione, Twitter permette di accedere solo a parti del servizio di *micro-blogging* tramite *API*. L'autorizzazione consiste nella richiesta di sviluppo di un'*app* al servizio *developer*²⁴ di Twitter che viene esaminata al fine di verificare eventuali difformità con la *policy*. La richiesta accompagnata da un'esplicitazione degli scopi d'uso, dell'oggetto di indagine e dei metodi di estrazione e analisi dei dati può essere rigettata o approvata. In caso di esito affermativo, come nel caso di fattispecie, Twitter consente all'applicazione di *chiamare* durante l'estrazione dati un determinato *endpoint*, ossia un indirizzo riferito a un tipo specifico di informazione che per impostazione predefinita corrisponde solo ad informazioni rese pubbliche sui profili utenti. Per accedere ai dati tramite l'*endopint API* si necessita di un'autenticazione *OAuth*, come generico protocollo di verifica dell'identità coinvolta nel trasferimento delle informazioni, e di un'autenticazione *token bearer*, che è uno schema di autenticazione *HTTP* come *token* di sicurezza. Tali credenziali rilasciate dal servizio di *developer* richiamano l'*API* per l'accesso ai dati di Twitter, attraverso il supporto di metodi di estrazione che si basano su tecniche di *Web scraping*. In pratica, tramite le *libraries* di alcuni *software* è possibile richiamare l'*API* e accedere a funzioni o dati inaccessibili mediante altri programmi o servizi *Web*. In questa tesi, per l'estrazione dei dati si fa riferimento ai *packages* di *R* che supportano l'*API* di Twitter ed in particolare a *TwitterR* (Gentry, 2009) sostenuto da altre *libraries* come *RJSONIO* (Temple Lang e Wallace, 2010), *ROAuth* (Gentry e Temple Lang, 2011), *streamR* (Barbera, 2013), *RCurl* (Temple Lang e CRAN team, 2004). Per mezzo di tali pacchetti si autentica la propria utenza e si richiama l'*API* di Twitter che in risposta fornisce solo gli elementi degli *account* resi pubblici e in accordo con i parametri della richiesta. I dati rilevati vengono sistematicamente salvati durante il periodo di monitoraggio prestabilito, realizzando così il *data storage* su cui condurre l'analisi (*Figura 4.1*).

²⁴ <https://developer.twitter.com>

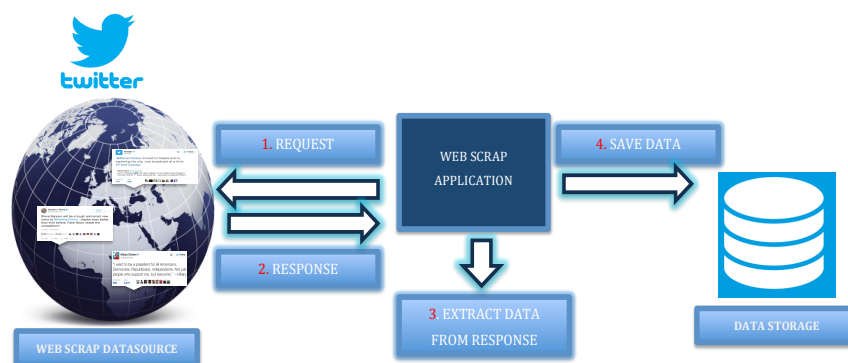


Figura 4.1. Struttura per l'estrazione dei dati da Twitter

Fonte: Elaborazione propria

I parametri della richiesta sono a discrezione del ricercatore e prevedono l'estrazione delle informazioni pubblicate o da uno specifico utente o su uno specifico tema, attraverso il ricorso ad una o più *parole chiave*.

Poiché la finalità del presente lavoro consiste nell'analisi dei linguaggi multilingue *online* sul fenomeno della violenza contro le donne, per estrarre i contenuti da Twitter si utilizzano specifiche *keywords* monolingue pertinenti al fenomeno, così da acquisire il *database* per l'indagine.

Tuttavia, il multilinguismo esistente presuppone al contempo un importante multiculturalismo, tale per cui non solo muta la lingua da un contesto geografico all'altro ma cambiamo le parole chiave, perché quello che può essere *notiziabile* in un contesto territoriale potrebbe non esserlo in un altro contesto, a causa di variabili socio-culturali non palesemente note. Per tale motivo, si considera opportuno affidare la selezione delle *keywords* per la rilevazione dei dati a fonti specializzate e diversificate che conoscono il tema della violenza contro le donne, sono inserite nel tessuto socio-culturale del luogo e sono madrelingua. In realtà, è bene sottolineare che la scelta delle lingue è strettamente legata alla disponibilità degli enti o delle figure professionali consultate. La laboriosa attività di ricerca delle fonti circoscritta al contesto europeo è durata diversi mesi ed ha prodotto cinque diversi risultati:

- **Italia:** Consiglio Nazionale delle Ricerche (CNR) di Roma, Istituto di Ricerche sulla Popolazione e le Politiche Sociali (<https://www.irpps.cnr.it>);

- **Regno Unito:** Membro del Chartered Institute of Linguists (<https://www.ciol.org.uk>);
- **Francia:** Docente di Sociologia;
- **Spagna:** Esperto di Sociolinguistica e Traduzione specializzata;
- **Ucraina:** Membro dell'Istituto per la Psicologia Politica e Sociale dell'Accademia Nazionale per le Scienze dell'Educazione (<http://ispp.org.ua>) e co-fondatore dell'*European Society for Traumatic Stress Studies (ESTSS)*; (<https://estss.org>).

I referenti di ciascuna lingua hanno fornito una lista di dieci diverse *keywords* (Tabella 4.1) monolingue. Nello specifico, al fine di non ottenere dati esclusivamente basati sul linguaggio dell'odio, i referenti di ogni lingua hanno ritenuto appropriato indicare cinque *keywords* che contrastano il fenomeno (ad esempio: *empowerment, inclusione sociale, co-educazione, etc...*) e cinque *keywords* che esprimono invece proprio il tema della violenza (ad esempio: *sessismo, violenza domestica, femminicidio, etc...*).

	KEYWORDS INGLESI	TRADUZIONE ITALIANA	KEYWORDS FRANCESI	TRADUZIONE ITALIANA
	rape	stupro	fémicide	femminicidio
	sexism	sessismo	meurtre conjugal	morte coniugale
	domestic violence	violenza domestica	violence sexuelles	violenza sessuale
	sexual abuse	abuso sessuale	violence physiques	violenza fisica
	cunt	genitale femminile	crime passionnel	crimine passionale
KEYWORDS ITALIANE	power balance	equilibrio di potere	brisons le silence	rompiamo il silenzio
violenza di genere	feminism	femminismo	pouvoir	potere
uomini maltrattati	against rape	contro lo stupro	reconstruction	ricostruzione
colpevolizzazione vittima	culture of equality	cultura dell'uguaglianza	élimination de la violence	eliminazione della violenza
violenza psicologica	social inclusion	inclusione sociale	journée internationale	giornata internazionale
femminicidio	KEYWORDS SPAGNOLE	TRADUZIONE ITALIANA	KEYWORDS UCRAINE	TRADUZIONE ITALIANA
centro antiviolenza	estereotipos de género	stereotipo di genere	життєва цінність	valore della vita
leadership femminile	crimen sexual	crimine sessuale	сімейні	famiglia
empowerment	violencia machista	violenza maschilista	знати	conoscenza
inclusione sociale	cultura de violación	cultura dello stupro	заткнись	stai zitta
casa rifugio	brecha de género	divario di genere	влада чоловікові	potere al marito
	coeducación	co-educazione	соціальні дослідження	ricerca sociale
	techo de cristal	soffitto di cristallo	змінити суспільство	cambiare la società
	reformular la cultura	riformulare la cultura	політична довіра	fiducia politica
	perspectiva de género	prospettiva di genere	гендерна рівність	uguaglianza di genere
	no cosificación	nessuna reificazione	лідерство жінок	leadership femminile

Tabella 4.1. Lista delle keywords monolingue per l'estrazione dei dati

Oltre alle parole chiave e alla specifica della lingua, tra gli altri parametri si utilizza il criterio di *geo-codifica* che circoscrive l'area geografica da cui estrarre i dati. In particolare, si intendono utilizzare le coordinate di latitudine e

longitudine delle capitali dei paesi coinvolti nell'indagine dei linguaggi e delle culture di questa ricerca.

PAESI E CAPITALI	COORDINATE DI LATITUDINE E LONGITUDINE
Italia: Roma	'41.9027835,12.4963655,100mi'
Regno Unito: Londra	'51.5073509,-0.1277583,100mi'
Francia: Parigi	'48.856614, 2.3522219,100mi'
Spagna: Madrid	'40.4167047,-3.7035825,100mi'
Ucraina: Kiev	'50.4500644, 30.5241037,100mi'

Tabella 4.2. Geo-codifica dei paesi per l'estrazione dei dati

Il periodo di monitoraggio è comune a tutte le lingue ed è compreso dal primo maggio al 31 ottobre 2019. La richiesta di estrazione dati per la costruzione del *database* multilingue viene rinnovata ogni giorno per tutta la *timeline*.

Esistono tuttavia dei limiti²⁵ connessi alla fase di *data extraction* che dipendono dal servizio di sviluppo dell'applicazione. Il primo condizionamento deriva dall'opportunità di poter eseguire al massimo 100 000 richieste di *chiamata* all'*API endpoint* di Twitter in 24 ore. Il secondo svantaggio risiede nella restituzione dei soli *tweets* più recentemente pubblicati in accordo con i parametri della richiesta. Infine, il sistema consente di rilevare un massimo di 3200 *tweets* al giorno per ogni *keyword*, inclusi i *retweets*, indipendentemente dal codice utilizzato nel comando di esecuzione.

4.2. La collezione di *tweets* multilingue

Attraverso l'utilizzo dei metodi, delle tecniche e dei parametri di rilevazione sopramenzionati si ottiene un *database* pari a 4 milioni 62 mila e 725 *tweets* sul tema della violenza contro le donne.

²⁵

<https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user-timeline>

PAESE	N. TWEETS	PERCENTUALE
ITALIA	308 664	7.6
REGNO UNITO	973 478	24.0
FRANCIA	1 055 619	26.0
SPAGNA	1 621 746	39.9
UCRAINA	103 218	2.5
TOTALE	4 062 725	100

Tabella 4.3. La collezione dei dati multilingue

I risultati del monitoraggio iniziato a maggio e terminato ad ottobre 2019 mostrano che quasi il 40% dei *tweets* sono stati raccolti in Spagna, la quale collezione di dati rappresenta la maggior parte dell'intero *dataset* multilingue. Infatti, solo in Spagna sono stati estratti 1 milione 621 mila 746 *tweets*.

Una raccolta di dati così numerosa in confronto alle altre lingue può sicuramente essere motivata da quanto sia elevata la popolarità del Social Media in ciascuno dei paesi studiati. Infatti, il numero di utenti attivi su Twitter in Spagna è maggiore che negli altri paesi europei, ovvero il 49%. La piattaforma è al quinto posto tra i Social Media più utilizzati e durante il periodo di monitoraggio la media di utilizzo è stata del 4.07% (Global Stats 2019, <https://gs.statcounter.com/social-media-stats/all/spain>).

Si considera altresì interessante la quantità di *tweets* ottenuta dalla Francia che è pari al 26% del *database* globale, nonché 1 milione 55 mila e 619 *tweets*. Twitter è al settimo posto tra i Social Media più utilizzati in Francia con una percentuale d'uso del 28%. Gli utenti attivi sono 5.56 milioni con una predominanza anche in questo caso maschile, pari al 63% (We are Social, 2019: <https://wearesocial.com/global-digital-report-2019>). Nel periodo di rilevazione dei dati la piattaforma ha registrato un utilizzo medio del 5.7% (Global Stats 2019, <https://gs.statcounter.com/social-media-stats/all/france>).

Al contrario della Spagna e della Francia, l'Ucraina è il paese che ha restituito il minor numero di *tweets*, solo 103 mila e 218 dati. In realtà, una quantità di *tweets* così esigua può dipendere da un uso limitato di Twitter, la cui utenza attiva è di 576 mila persone, con una percentuale maschile del 73% (We are Social, 2019: <https://www.slideshare.net/DataReportal/digital-2019-ukraine-january-2019-v01>). Tuttavia, nel periodo di estrazione dei dati, la media d'utilizzo della piattaforma è stata dell'8.56% (Global Stats 2019, <https://gs.statcounter.com/social-media-stats/all/ukraine>).

La diffusione del Social Media utilizzato per la rilevazione dei dati ha sicuramente influito sulla quantità di *tweets* estratti per ogni lingua.

È proprio per questo motivo che si è ritenuto ragionevole dapprima condurre uno studio esplorativo monolingue prima di passare al confronto multilingue.

4.3. L'analisi preliminare attraverso lo studio esplorativo monolingue

Lo studio esplorativo monolingue possiede quindi lo scopo di indagare i dati estratti da Twitter per ciascuna lingua al fine di descrivere ed analizzare i tratti maggiormente caratteristici racchiusi nel linguaggio e nella cultura di ciascun contesto esaminato.

Specificamente, in un primo momento si osserva l'andamento dei *tweets* nel periodo di monitoraggio, per valutare ed individuare all'interno di ciascuna distribuzione di dati monolingue i potenziali eventi che possano aver determinato un incremento dello *speech online* sul tema della violenza contro le donne.

In un secondo momento, si procede all'applicazione delle procedure tipiche del *Text Mining* sui dati testuali, mediante le operazioni di *pre-processing* e *feature extraction* che ne consentono la *data visualization* o sintesi del contenuto.

Tali procedure nel presente contesto sono state condotte attraverso un *package* di R denominato *quanteda* (Benoit *et al.*, 2018), ovvero *quantitative analysis of textual data*, il quale pacchetto²⁶ consente di elaborare testi in linguaggio naturale, grazie ad un *workflow* e *toolkit* di *data analysis* e *data visualization* che supera le *performance* di molti altri *software* costosi e non *open-source* (Welbers *et al.*, 2017).

Per la fase di *pre-processing*²⁷, ciascun *corpus* monolingue è stato sottoposto a procedure di *normalizzazione*, *lemmatizzazione*, rimozione di numeri, punteggiatura, *stop words* per ogni lingua, caratteri speciali, *link*, *URLs*, *emoji* ed altri simboli. Tuttavia, per quanto riguarda la *lemmatizzazione* è doveroso puntualizzare che su alcuni specifici termini tale operazione non è stata eseguita (ad esempio per la lingua francese *jeune fille* o per lo spagnolo *joven chica* o

²⁶ *Quanteda package* viene sviluppato nel 2013 da Kenneth Benoit con finanziamenti dell'European Research Council (ERC-2011-StG 283794-QUANTESS): <https://cran.r-project.org/web/packages/quanteda/quanteda.pdf>

²⁷ I dettagli relativi alle operazioni di *pre-processing* si rimandano dal capito terzo: § 3.2.1.1.

per l'italiano *bambina*), con la finalità di sottolineare aspetti legati alla cultura di genere o al linguaggio della violenza.

Per la fase di *feature extraction* invece, una volta definita la soglia di frequenza, in grado di coprire in maniera appropriata il testo, sono state visualizzate le prime 100/ 200 parole, in termini di *top features*. La rappresentazione di sintesi del contenuto (o *Text summarization*) di ciascuna raccolta di dati monolingue (*tweets* e *retweets*, solo *hashtag* e solo *retweets*) ha restituito una mappatura della percezione pubblica sul tema della violenza contro le donne per ogni contesto linguistico.

4.3.1. I risultati dell'Italia

La collezione dei dati in lingua italiana è costituita da 308 mila e 664 *tweets*. In particolare, il grafico della serie storica (*Figura 4.2*) mostra l'esatto andamento della comunicazione *online* sul tema, permettendo di osservare un *trend* con continue variazioni nel tempo: si ha una scarsa produzione di *tweets* durante maggio e giugno, mentre si evidenzia un *outlier* additivo nel mese di luglio. A settembre ed ottobre i dati mostrano un aumento di *tweets* sul fenomeno della violenza contro le donne ed in particolare il 9 settembre si osserva un picco particolarmente alto. In quella data infatti, in Italia si registrano 10 mila 168 *tweets* quando la stampa parla di un nuovo caso di femminicidio: il delitto di Elisa Pomarelli²⁸.

²⁸ Per approfondimenti sul caso di cronaca:

https://www.repubblica.it/protagonisti/Elisa_Pomarelli?refresh_ce

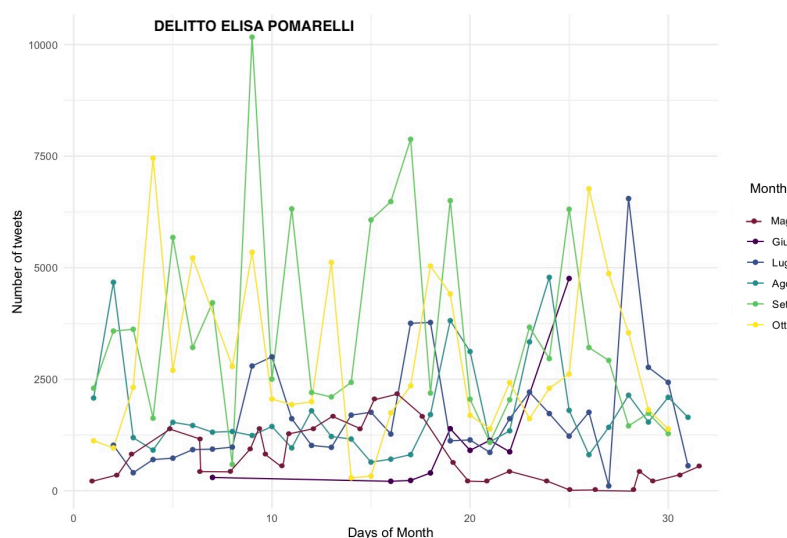


Figura 4.2. Serie storica dei tweets italiani

Con lo scopo invece di analizzare e descrivere il vocabolario italiano esplorando potenziali scenari sul linguaggio dell'odio e della violenza *online* si realizza l'analisi del testo sul *corpus* italiano.

Le fasi essenziali per ottenere una sintesi del contenuto di ciascuna lingua sono partite innanzitutto dalla creazione di tre sotto-corpora: il primo contiene *tweets* e *retweets*, il secondo solo gli *hashtag* ed infine il terzo i soli *retweets*.

Tuttavia, prima di procedere con l'analisi, per assicurare la fattibilità del processamento statistico dei dati, su ciascuno dei tre sotto-corpora sono stati calcolati due indicatori lessicometrici, quali il *Type/Token ratio (TTR)* ed il numero di *hapax*²⁹. Solo dopo che gli indicatori indicano di poter procedere con l'analisi, sono state effettuate le procedure di *pre-processing* e *feature extraction* su ognuno dei tre *sub-corpora*.

In particolare, per il pre-trattamento di tutti e tre i sotto-corpora italiani, dopo la normalizzazione, la rimozione della punteggiatura, dei caratteri numerici, dei simboli e delle *keywords* utilizzate per il *download*, sono state rimosse 657 *stop words* (dalla *Multilingual Stopwords List ISO-639-1*).

²⁹ *Hapax*: forma grafica che si presenta una sola volta nel *corpus*: § 2.5.2.

Mentre, per la fase di estrazione delle *features* è stata scelta un'appropriata soglia di frequenza per ciascuna collezione di dati, al di sotto della quale soglia, le parole tralasciate possono considerarsi poco rilevanti³⁰.

La prima raccolta di dati che viene sottoposta ad indagine è comprensiva di tutti i *tweets* e *retweets* del *corpus* italiano. Gli indicatori lessicometrici che permettono di descrivere da un punto di vista quantitativo il testo indicano che il *Type/Token ratio* del *corpus* italiano è del 2% (TRR= 0.02), mostrando chiaramente che il numero di parole distinte non supera il 20% delle occorrenze, pertanto, è sufficientemente esteso per coglierne la ricchezza lessicale. Il numero di *hapax* presenti è del $\approx 38\%$, quindi inferiore al 50% e trattabile da un punto di vista statistico. Il limite tra le medie e le basse frequenze (partendo dal basso) suggerisce di considerare le forme grafiche con frequenza maggiore e uguale a 37, al cui rango corrisponde una copertura del testo del 61%.

L'approccio di *Text summarization* adottato è il *word cloud* la cui rappresentazione grafica ha permesso di visualizzare le dimensioni delle *features* in maniera proporzionale al valore di occorrenza. Si ricorda, che nel calcolo di frequenza sono stati considerati anche i *token* adiacenti in termini di *n-grams*.

La rappresentazione grafica della *Figura 4.3*. visualizza le occorrenze che derivano dall'analisi dei *tweets* e *retweets* italiani. In particolare, il *word cloud* ha permesso di osservare la predominanza dei termini *maschilismo* e *dimenticare*, seguiti dalle *features* *numero agghiacciante*, *sessuale* e *abuso domestico*. Se si pone poi l'attenzione sui quei termini che presentano un valore di occorrenza inferiore, risulta interessante osservare la presenza delle *features*: *matteo salvini*, *stupratore*, *vergogna*, *donna*, *amore*, *conseguenza*, *polizia*, *legge*, *sangue* e *retorica*.

Pertanto, si evince una sintesi del contenuto dei *tweets* e *retweets* italiani essenzialmente basata sull'esistenza di una società *maschilista* e sul bisogno di *dimenticare*, probabilmente, tale prospettiva esistenziale.

Il *corpus* tende inoltre a sottolineare l'*escalation* della violenza attraverso parole che mirano ad evidenziare i numeri sempre più alti delle vittime. Infatti, l'opinione pubblica studiata ha prevalentemente indicato la presenza di un

³⁰ Nota 7 del § 3.2.1.2.

In sintesi, lo studio esplorativo dei dati estratti in lingua italiana ha rilevato un primo *corpus* comprensivo di *tweets* e *retweets* essenzialmente basato sul sessismo maschilista, sull'*esclation* numerica della vittimizzazione femminile, sulla violenza sessuale e sull'abuso intrafamiliare.

Al contrario, la collezione di *hashtag* risulta incentrata sul tema della violenza in tutte le sue molteplici forme, inclusa quella di genere ma non escludendo altri tipi di abuso come la violenza sui minori o la sopraffazione politico-giornalistica. Infine, la collezione dei *retweets* ha confermato l'analisi del primo *corpus* comprensivo di *tweets* e *retweets* portando alla luce l'importanza del maschilismo nel dibattito pubblico, tuttavia dando primaria importanza alla sessualità. La sua trama argomentativa si sposta verso la criminalizzazione di tipo omicidiaria e poi verso la politica attraverso la figura di Matteo Salvini.

4.3.2. I risultati del Regno Unito

Il *dataset* inglese è costituito da 973 mila e 478 *tweets*. La serie storica (Figura 4.7) indica nel mese di maggio una scarsa produzione di *tweets*, mentre registra continui aumenti e diminuzioni durante i primi dieci giorni di agosto, settembre ed ottobre. L'aspetto più interessante si evince nel mese di giugno, periodo che presenta costanti valori bassi ad eccezione di un unico incremento anomalo. Infatti, il 22 giugno si registrano 32 mila 854 *tweets*, il picco più alto nell'intera distribuzione di dati. L'evento che potrebbe aver influito è l'inchiesta sullo sfruttamento sessuale minorile dell'area periferica del West Yorkshire³².

³²Ad esempio: <https://www.bbc.com/news/uk-england-leeds-48705621>

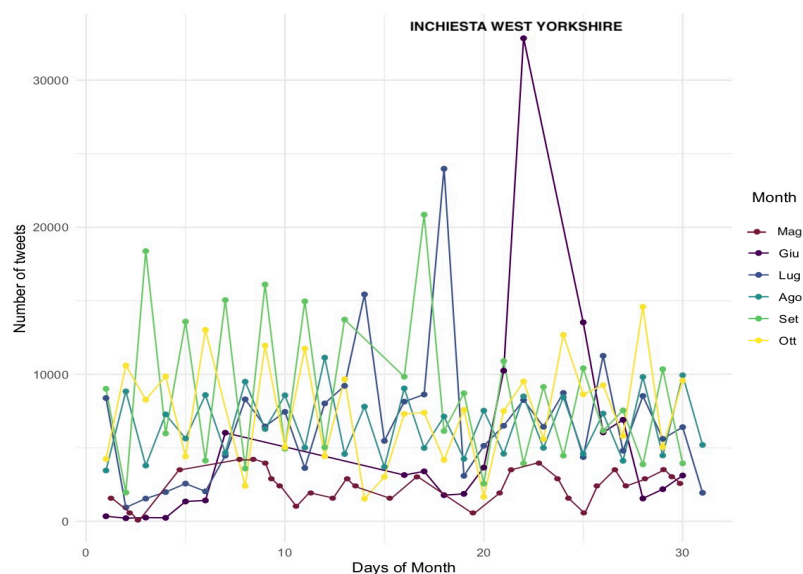


Figura 4.7. Serie storica dei tweets inglesi

I tre *sub-corpora* inglesi, ovvero quello dei *tweets* e *retweets*, quello degli *hashtag* e dei soli *retweets* sono stati sottoposti a pre-trattamento, ovvero è stata eseguita la normalizzazione, rimozione della punteggiatura, dei caratteri numerici, dei simboli, delle *keywords* utilizzate per la *data extraction* e sono state rimosse 571 *stop words* presenti nell'elenco *SMART*³³(2004).

La prima raccolta di dati inglesi, comprensiva di *tweets* e *retweets* mostra che l'estensione del *corpus* è al di sotto del 20% e nello specifico è pari a 2.4% (TTR=0.024), mentre la presenza di *hapax* è del 49%. Il confine tra le medie e le basse frequenze indica di considerare le forme grafiche partendo dal valore di frequenza di 85, la cui copertura del testo è del $\approx 62\%$.

³³ Si rinviano gli approfondimenti al capitolo terzo: § 3.2.1.1.

L'analisi degli *hashtag* invece ha mostrato un numero di parole distinte pari al 19% (TRR= 0.19) ed un numero di *hapax* del 48.9%. Il limite tra le medie e le basse frequenze indica di considerare gli *hashtag* che si presentano nel *corpus* con frequenza maggiore e uguale a 18, la cui copertura del testo è del 71%.

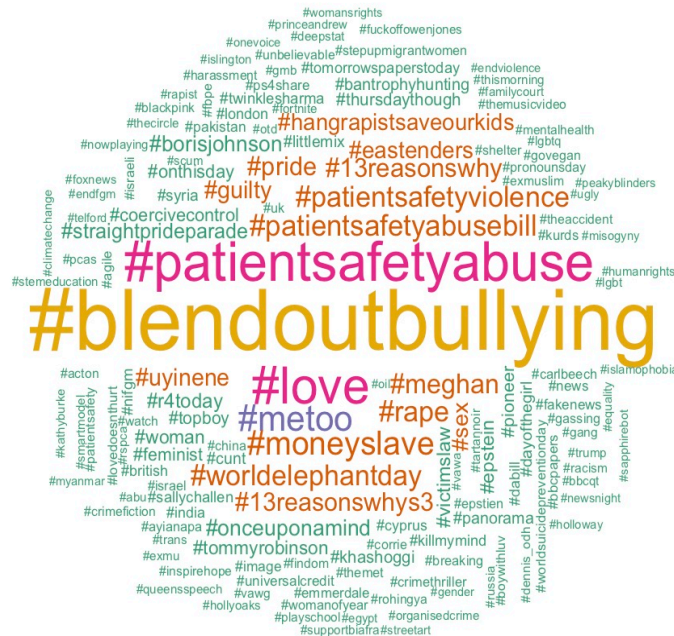


Figura 4.9. Word cloud degli hashtag inglesi

L'*hashtag* maggiormente utilizzato è *#blendoutbullying* chiaramente riferito al tema del bullismo e linguaggio d'odio *online*. In particolare, esso identifica una campagna virale che propone di evidenziare sul proprio viso la cosa peggiore che è stata indirizzata a ciascun individuo attraverso il *Web*. L'offesa scelta da ciascuna "vittima" viene scritta con il *make up* sul proprio viso e poi diffusa tramite la rete, con lo scopo di evidenziare l'importanza e il peso delle parole. L'analisi degli *hashtag* mette inoltre in evidenza l'abuso che avviene nei luoghi sanitari attraverso *#patientsafetyabuse*, *#patientsafetyviolen*, *#patientsafetyabusebill* e sottolinea la presenza dell'amore in un contesto di studio che parla di violenza. Infatti, proprio attraverso *#love* esprime quanto spesso l'abuso sia connesso alla sfera affettiva.

Tra le altre etichette di contenuto prevalentemente utilizzate in lingua inglese vi è *#metoo*, il quale rappresenta un movimento femminista contro le molestie sessuali e la violenza sulle donne che si è diffuso in modo virale, a livello

giustizia (*justice system*), al governo (*government*), alla legislazione (*legislation*), al sesso (*sex*) e all'incesto (*incest*). Si possono identificare *features* che si riferiscono alla vita (*life*), al razzismo (*racism*), alla cultura (*literacy*), al contrasto (*eliminate*) e *stop* ed ancora, che invece sottolineano l'efferezza della violenza che si traduce in terrorismo (*terrorist*), in omicidio (*killling*) e in violenza carnale (*rapist*).

In sintesi, l'analisi esplorativa dei dati estratti in lingua inglese ha evidenziato un primo *corpus* di *tweets* e *retweets* incentrato sul maltrattamento, sulla famiglia come nucleo di violenza e sull'importanza offensiva del fenomeno. In secondo luogo, lo *speech* che emerge si basa sul potere pubblico, interpretabile in senso plurivalente, ossia da una parte in grado di incentivare attraverso il linguaggio dell'odio il fenomeno e dall'altro di contrastarlo attraverso le proprie azioni. Il *corpus* parla ancora di violenza sessuale, sfruttamento, giustizia, forze dell'ordine, femminismo. La raccolta degli *hashtag* invece evidenzia il bullismo e violenza in tutte le forme, facendo emergere anche tipi di abuso non particolarmente noti e rilanciati dalla comunicazione pubblica, come la violenza nei luoghi ospedalieri-ambulatoriali. Di particolare interesse sono inoltre gli *hashtag*: *#love* e *#metoo* finalizzati proprio a sottolineare la violenza di genere. Da una parte si sottolinea la trappola dell'amore dietro cui si celano gli abusi e le sopraffazioni e dell'altra riecheggia il movimento internazionale e virale del *#metoo* che accusa e condanna ogni tipo di molestia di natura sessuale contro le donne. Infine, lo studio dei *retweets* si è mostrato in linea con i risultati del *corpus* comprensivo di *tweets* e *retweets*, evidenziando i temi del maltrattamento, della violenza sessuale e del potere pubblico. Tuttavia, in questo ultimo *corpus* emergono aspetti legati alla religione islamica e alla vita che si trasforma in una prigione di violenza.

4.3.3. I risultati della Francia

La raccolta dei dati in lingua francese è composta da 1 milione 556 mila e 619 *tweets*.

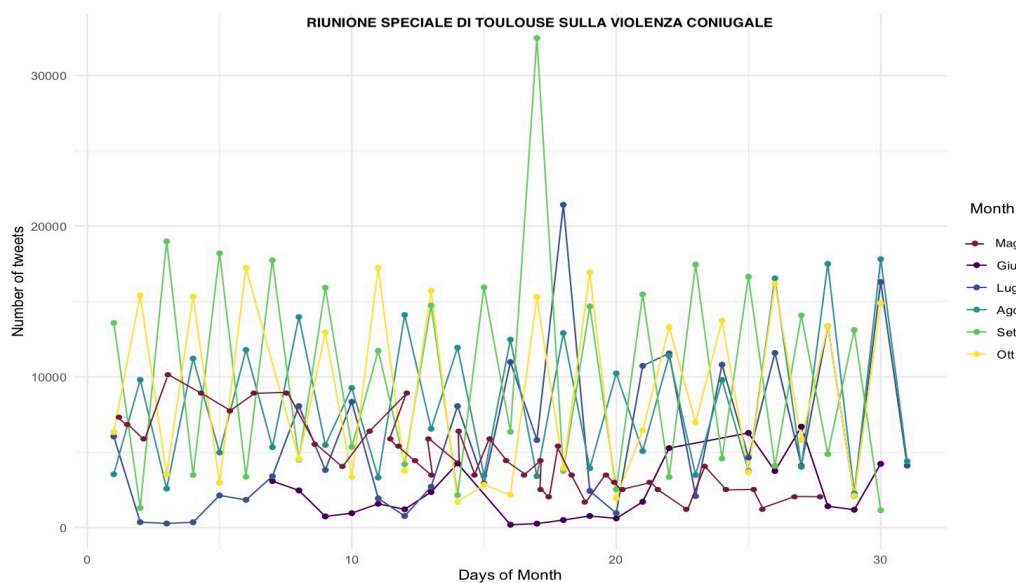


Figura 4.11. Serie storica dei tweets francesi

L'andamento dello *speech online* è piuttosto simile durante i mesi di luglio, agosto, settembre ed ottobre (Figura 4.11) e a seconda dei giorni si registrano crescite e decrescite nel numero dei *tweets*. Settembre mostra il maggiore incremento ed in particolare il 17 settembre si osservano 32 mila 486 *tweets*. L'evento che potrebbe aver influito è la riunione speciale del Comitato locale dell'Alta Garonna tenutasi sul tema della violenza coniugale proprio lo scorso 17 settembre a Toulouse³⁵.

I tre *sub-corpora* francesi, quello dei *tweets* e *retweets*, quello degli *hashtag* e dei soli *retweets* sono stati pre-processati attraverso normalizzazione, rimozione della punteggiatura, dei caratteri numerici, dei simboli, delle *keywords* utilizzate per l'estrazione dei dati e sono state rimosse 689 *stop words* (*Multilingual Stopwords List ISO-639-1*).

Nel caso di fattispecie, la ricchezza lessicale del primo *corpus* è pari al 2% (TTR=0.02) mentre la presenza degli *hapax* è del 43%. Il limite tra le medie e le basse frequenze mostra una soglia di frequenza pari a 101, con una copertura del testo del 50.65%.

35

http://www.haute-garonne.gouv.fr/content/download/31297/208752/file/20190916_CP_CLAV.pdf

familiare e sessuale, la discriminazione ed il maschilismo sono comunque alquanto emergenti.

Anche la collezione degli *hashtag* risulta incentrata sulla violenza di genere, sottolineando per l'appunto la solidarietà femminile, intesa anche come movimento francese nato per contrastare la violenza sulle donne. Inoltre, la traduzione dall'inglese *#metoo* in *#balancetonporc* ha rappresentato la condivisione virale ed internazionale di vicinanza al tema. Gli *hashtag* quindi tendono ad evidenziare l'unione delle donne e delle organizzazioni governative e non, per l'eliminazione della vittimizzazione femminile.

Infine, la collezione dei *retweets* ha confermato l'analisi del primo *corpus* comprensivo di *tweets* e *retweets*, portando alla luce proprio l'aspetto legislativo-normativo e sottolineando l'importanza delle iniziative governative e della giustizia, ponendo altresì l'enfasi sulle molestie e sugli autori della violenza, che rientrano nella maggior parte dei casi negli ex-congiunti delle vittime.

4.3.4. I risultati della Spagna

Il *dataset* spagnolo è costituito da 1 milione 621 mila 746 *tweets*.

L'andamento dei *tweets* (Figura 4.15) mostra durante i primi giorni di maggio e giugno l'assenza delle *keywords* indicate per l'estrazione dei dati, mentre nei giorni successivi si possono osservare graduali aumenti.

Nella distribuzione dei dati si individuano due particolari picchi: il primo è il 18 luglio con 38 mila 140 *tweets*, e il fenomeno che potrebbe aver contribuito ad aggiungere tale *speech* è la ricorrenza della commemorazione dell'avvento del regime Franchista che nel 1936 ha generato una violenta repressione delle donne³⁶. Il secondo incremento si registra il 17 settembre con 43 mila 300 *tweets*, in concomitanza ad un uxoricidio commesso a Madrid in rivendicazione della supremazia genere³⁷.

³⁶<https://kaosenlared.net/la-violencia-extrema-contras-las-mujeres-como-venganza-machista-del-franquismo-desde-el-18-de-julio/>

³⁷<https://theobjective.com/further/violencia-machista-asesinadas-2019/>

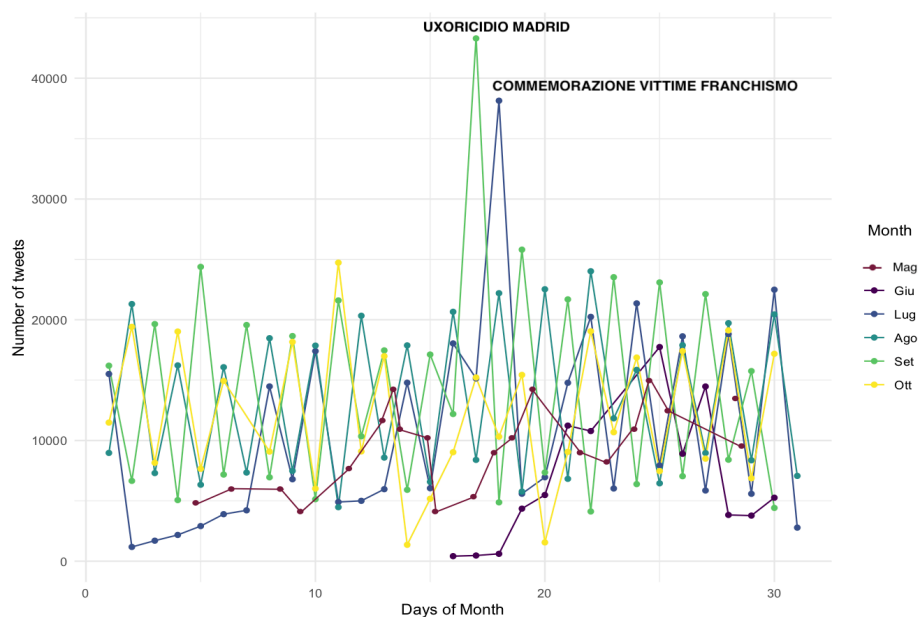


Figura 4.15. Serie storica dei tweets spagnoli

I *sub-corpora* spagnoli, dei *tweets* e *retweets*, poi degli *hashtag* e quello dei soli *retweets* sono stati sottoposti a pre-trattamento, attraverso la normalizzazione, rimozione della punteggiatura, dei caratteri numerici, dei simboli, delle *keywords* utilizzate per la *data extraction* e sono state rimosse 732 *stop words* (*Multilingual Stopwords List ISO-639-1*).

La prima collezione di dati comprende sia i *tweets* che i *retweets* evidenziando attraverso gli indicatori lessicometrici un'estensione del *corpus* pari all'1% (TTR=0.01), mentre la presenza degli *hapax* è del 39%. Il confine tra le medie e le basse frequenze suggerisce di considerare le forme grafiche partendo dal valore di frequenza di 26, la cui copertura del testo è del 50.18%.

La rappresentazione del contenuto attraverso il *word cloud* (Figura 4.16) indica tra i termini più frequentemente utilizzati *patriarcado*, *empoderamiento* e *libertad*. Pertanto, se da una parte la percezione pubblica spagnola tende a sottolineare l'esistenza di una cultura patriarcale, dall'altra parte vuole esprimere la rivincita delle donne e la loro inclusione sociale attraverso la libertà e l'*empowerment*. La trama argomentativa di una società patriarcale si rafforza attraverso *features* come *discriminación*, *retrógrada* e *misoginia* che sottolineano per l'appunto la presenza di una cultura basata sulla discriminazione di genere e sul sessismo, particolarmente insensibile al progresso ed ostinata a rimanere ancorata alla tradizione e al passato.

(*igualdad*), le misure di conciliazione (*medidas conciliación*), il governo (*gobierno*), la giustizia (*justicia*), il femminismo (*feminismo*), la voce (*vox*) e la legge (*ley*).

In sintesi, l'analisi esplorativa dei dati estratti in lingua spagnola ha evidenziato un primo *corpus* comprensivo di *tweets* e *retweets* incentrato da una parte sull'autoderminazione femminile attraverso l'*empowerment* e la libertà, e dall'altra parte basato sulla condanna di una cultura patriarcale, discriminatoria e misogina. La raccolta degli *hashtag* evidenzia con un maggior numero di rilanci il tema del femminicidio vero e proprio, per poi confermare gli stessi temi emersi dai *tweets* e *retweets*. Infine, l'ultimo *corpus*, ossia quello dei *retweets* non rappresenta altro che un rafforzamento degli aspetti che sono emersi attraverso lo studio della prima collezione di dati, nella quale si contrapponevano l'*empowerment* e la presenza di una società retrograda e maschilista.

4.3.5. I risultati dell'Ucraina

La raccolta dei dati in lingua ucraina è composta da 103 mila e 218 *tweets*. L'andamento dei dati (*Figura 4.19*) mostra nel mese di maggio una scarsa produzione di *tweets* che tende sempre più a contrarsi fino all'estrazione nulla di contenuti. Durante la fine di giugno si può invece osservare un *outlier* additivo, per poi assistere ad una nuova decrescita.

Nell'intera distribuzione si registrano due particolari picchi. Il primo con 1939 *tweets* il 2 ottobre, in riferimento alla *Giornata Internazionale della Non Violenza*, che sottolinea una delle forme di violenza più diffuse in Ucraina: il conflitto domestico, viene riconosciuto legalmente come crimine³⁹ solo nel 2019. Il secondo evento si osserva il 4 ottobre, con 2 mila e 38 *tweets*, ovvero il giorno successivo alla divulgazione pubblica dell'Agenda del Governo Ucraino

³⁹ Ad esempio: <http://turadm.gov.ua/protydiia-domashnomu-nasyilstvu/item/5938-2-zhovtnia-mizhnarodnyi-den-nenasyilstva>

2025⁴⁰, il cui programma ignora il tema della violenza domestica, del sessismo e della disuguaglianza di genere.

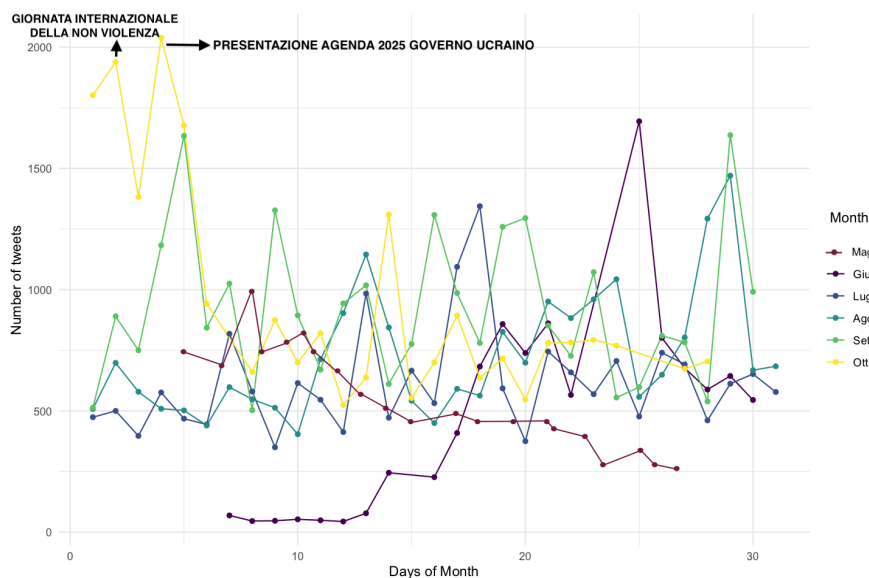


Figura 4.19. Serie storica dei tweets ucraini

I tre *sub-corpora* ucraini, nonché quello dei *tweets* e *retweets*, quello degli *hashtag* e dei soli *retweets* sono stati sottoposti a pre-trattamento mediante la normalizzazione, rimozione della punteggiatura, dei caratteri numerici, dei simboli, delle *keywords* utilizzate per la rilevazione dei dati e sono state rimosse 28 *stop words* (*Multilingual Stopwords List ISO-639-1*).

L'analisi del primo *corpus*, ossia comprensivo di *tweets* e *retweets* ha consentito di individuare una ricchezza lessicale del 5% (TTR=0.05), mentre la presenza degli *hapax* è del 47%. La soglia di frequenza suggerita è di 5, coprendo così il 58% del *corpus*.

Attraverso la visualizzazione grafica del *word cloud* (Figura 4.20) le parole più frequenti che emergono per rappresentare il fenomeno sono *il tuo posto* (вашому місці), *denunciare* (дeнoнcyвaти) e *potere* (пoтужнiсть), le quali *features* esprimono in modo chiaro la subordinazione della donna sottomessa al

⁴⁰ Ad esempio: <https://povaha.org.ua/ukrajintsi-ukrajineti-i-lyudy-uryad-zapropnuvav-prohramu-do-2025-roku/>

La raccolta degli *hashtag* invece evidenzia la battaglia per i diritti umani, definendo quindi la violenza di genere come una una violazione dei diritti contro l'umanità e sottolineando il tema delle disabilità e del linguaggio d'odio *online*. L'analisi dei *retweets* infine si focalizza sulla conoscenza pubblica e sulla denuncia del fenomeno, evidenziando inoltre la prospettiva di cambiamento e di interesse circa l'ordinamento giuridico.

4.3.6. Brevi considerazioni sui risultati dello studio esplorativo

Lo studio esplorativo monolingue operato sul *corpus* dei *tweets* e *retweets*, degli *hashtag* e dei soli *retweets* ha permesso di inquadrare lo *speech online* sul tema della violenza contro le donne per ciascuna lingua, e di individuare possibili elementi di connessione tra le lingue.

Se si osservano i corpora comprensivi di *tweets* e *retweets* ad esempio, uno degli aspetti che sembra mettere in maggiore relazione le diverse lingue è la *violenza sessuale*, il quale elemento risulta presentarsi con una considerevole frequenza nei discorsi monolingue analizzati.

L'*abuso domestico* e la *violenza sessuale* in particolare identificano un tratto di similarità emerso dallo studio della lingua italiana ed inglese. Il tema della famiglia si rileva tuttavia anche dallo studio della lingua francese, seppure con una minore predominanza nel *corpus* rispetto ad altri termini. Difatti, il *corpus* francese è particolarmente focalizzato sulla *legislazione* e *giurisdizione*, trovando punti di connessione con la Spagna, che nella propria collezione di dati sottolinea la *difesa* e la *giustizia* e con l'Ucraina, che nei propri *tweets* e *retweets* evidenzia il *sistema giudiziario*.

Il confronto tra gli *hashtag* invece non mostra punti di accordo tra le lingue, quanto piuttosto presenta diversificazioni strettamente connesse al linguaggio monolingue. L'unica eccezione è *#metoo* per l'inglese e *#balancetonporc* per il francese, le quali etichette fanno riferimento allo stesso movimento femminista ma nei registri linguistici propri di ogni lingua.

Le collezioni dei *retweets* monolingue in parte confermano i risultati dei corpora comprensivi dei *tweets* e *retweets* di ogni lingua, e da un'altra parte mostrano un punto d'accordo tra le cinque le lingue studiate. In particolare, l'elemento che accomuna l'italiano, l'inglese, il francese, lo spagnolo e l'ucraino è l'aspetto legislativo e della giustizia. Benché sia evidente nei *retweets* francesi ed ucraini, tale aspetto emerge anche nei dati spagnoli però

con frequenza minore. Inoltre, via via con occorrenza sempre più inferiore, tale risultato si può osservare anche nello *speech* inglese ed italiano.

Complessivamente, lo studio esplorativo realizzato in questo contesto ha permesso di individuare il vocabolario di ciascuna lingua e di ricostruire attraverso gli *hashtag* e i *retweets* gli scenari tematici che contraddistinguono in maggior misura il fenomeno della violenza sulle donne per ogni caso linguistico.

Tuttavia, per esplorare un eventuale accordo o disaccordo tra le lingue è opportuno ricorrere a tecniche e metodi di analisi maggiormente induttivi e più adatti a cogliere la semantica racchiusa nei testi di ogni lingua.

Più propriamente, il problema della diversità linguistica in questo contesto si risolve classificando in raggruppamenti tematici le diverse *features* monolingue, così da interpretare tali gruppi in un vero e proprio *spazio di concetto* (Chew, 2009), per poi basare il confronto multilingue sui concetti che emergono per ogni lingua.

4.4. La definizione dei concetti per ciascuna lingua

La definizione dei concetti monolingue di questa tesi si realizza quindi classificando le *features* di ciascun *corpus* monolingue comprensivo di *tweets* e *retweets* in gruppi di concetto.

Dapprima però, si considera l'esigenza di determinare l'orientamento semantico delle *features*, in modo da assicurare che i termini classificati abbiano la stessa direzione semantica: ovvero solo positiva o solo negativa. In questo modo, i gruppi di concetto delle *features* positive determinano raggruppamenti concettuali positivi, mentre i gruppi di concetto delle *features* negative determinano raggruppamenti concettuali negativi.

A tale scopo, è necessario sottoporre il *corpus* di ciascuna lingua a calcolo della polarità semantica e trattandosi di corpora multilingue si fa ricorso all'*NRC Emotion Lexicon* (§ 3.2.2.2).

Tuttavia, prima di procedere al calcolo della polarità dei testi è bene tenere presente che l'analisi esplorativa precedentemente condotta ha permesso di individuare il vocabolario di ogni lingua in via preliminare. Pertanto, tale informazione può essere utilizzata per verificare la presenza o assenza di tali

parole all'interno del dizionario dell'*NRC Lexicon*, ai fini di una corretta classificazione del testo. In altre parole, attraverso i risultati dell'analisi esplorativa è possibile individuare se i termini che costituiscono ciascun vocabolario monolingue sono effettivamente contenuti nel *lexicon* utilizzato per il calcolo della polarità semantica. Questo aspetto risulta cruciale per la determinazione dell'orientamento semantico, poiché se i dizionari polarizzati delle diverse lingue non contengono al proprio interno la maggior parte delle parole presenti nel vocabolario di ogni lingua, la classificazione che l'algoritmo restituirà sarà impropria.

Difatti, da un confronto tra il vocabolario di ciascuna lingua ed il corrispondente *NRC Lexicon* risulta l'assenza di alcuni termini che si riferiscono a fenomeni di natura sociale, come ad esempio: *razzismo*, *omofobia*, *misoginia*, che risultano essere presenti nei corpora indagati. Una tale constatazione ci pone quindi di fronte all'esigenza di personalizzare il dizionario dell'*NRC*, con lo scopo di adattarlo alla natura sociale del tema indagato. Prende così forma la seconda fase dell'analisi *cross-language* presentata in questa tesi, la quale parte proprio dal calcolo della polarità semantica dei diversi corpora monolingue prima di sottoporli a qualsiasi pre-trattamento. Mediante il *package* di *R syuzhet* (Jockers, 2015) ha luogo la personalizzazione del dizionario ed il calcolo della polarità semantica che avviene a livello di documento, o meglio a livello di *tweet*, dove la polarità delle parole contenute nel *lexicon* conferisce a ciascun *tweet* un orientamento semantico positivo o negativo. Una volta terminato l'orientamento semantico di ciascun *tweet*, si ottengono due diversi *dataset* testuali: uno che raggruppa solo i *tweets* positivi e l'altro solo i *tweets* negativi. Ciascun *set* di dati positivi monolingue è stato sottoposto a pre-trattamento attraverso la normalizzazione, rimozione della punteggiatura, dei caratteri numerici, dei simboli, delle *keywords* utilizzate per la rilevazione dei dati e delle *stop words* relative ad ogni lingua. Successivamente si è proceduto con la *features extraction* tenendo conto delle opportune soglie di frequenza e degli indicatori lessicometrici, così da estrarre le cosiddette *features* positive. In modo analogo, ogni *set* di dati negativi monolingue è stato separatamente pre-trattato e sono state estratte le corrispondenti *features* negative. Infine, per determinare i gruppi concettuali positivi e negativi di ciascuna lingua, le *features* positive e le *features* negative sono state analizzate attraverso la combinazione dell'*AC* e della *Cluster Analysis* (§ 3.2.2.3). Nello specifico, sulle coordinate dell'*AC* è stata adottata

una strategia di classificazione mista, che parte dal metodo gerarchico di *Ward* e si consolida con il metodo non gerarchico del *k-means*.

L'interpretazione dei *cluster* e l'attribuzione dell'etichetta concettuale di ogni gruppo è determinata dal *value-test*, che in questo caso non fornisce un'interpretazione probabilistica delle classi ma supporta l'individuazione di una gerarchia tra le *features* di ogni gruppo, in termini di *significatività*.

Per *significatività* si intende un valore di probabilità non inferiore al 95% e in valori assoluti non inferiori a 2.

Per economia del testo e per evitare di trattare in maniera particolarmente estesa gli esiti della *Cluster Analysis* nei paragrafi successivi, in considerazione del fatto che il *focus* del presente lavoro di ricerca è il confronto *cross-language*, i risultati della *statistica test* (solo quelli non inferiori a 2) si rinviano all'Appendice B.

4.4.1. Polarità semantica e Concepts Extraction dei tweets italiani

La raccolta dei 308 mila 664 *tweets* italiani è costituita da 237 mila 826 testi negativi e 70 mila 838 testi positivi. L'interpretazione della polarità può essere affiancata dalla distribuzione delle otto categorie emozionali (Plutchik, 1991) che possono aiutare a motivare la prevalenza dell'orientamento semantico positivo o negativo. Infatti, nel caso del *corpus* italiano (*Figura 4.23*) la predominanza dei *tweets* negativi è determinata da contenuti testuali che esprimono *paura* (16.4% dei casi). Tuttavia, la distribuzione delle emozioni non presenta una netta contrapposizione tra categorie positive e negative, in quanto vi è una mescolanza di emozioni, che passano per l'appunto dalla *paura* all'*anticipazione* (15.3%), dall'*anticipazione* alla *tristezza* (14.8%) e poi alla *fiducia* (14.7%).

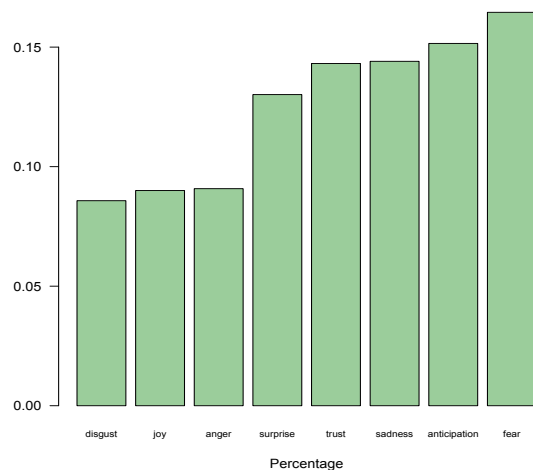


Figura 4.23. Distribuzione delle emozioni dei tweets italiani

Dalla collezione dei *tweets* positivi con TTR del 6% e il 27.5% di *hapax* sono state estratte le *features* cosiddette positive, adottando una soglia di frequenza uguale a 15. Le *features* sono state poi classificate attraverso la *Cluster Analysis* nel sottospazio dimensionale prodotto dall'*AC*. La classificazione delle prime 200 *features* positive può essere visualizzata attraverso la *Tabella 4.4*, che mostra 8 raggruppamenti concettuali positivi italiani.

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	N. FEATURES	%
1	PROGRESSO STRATEGICO	bilanciare_potere , radicale, progresso, equilibrio, pianificazione , sfatare , intuizione, sviluppare	8	4,0
2	RIDUZIONE DELLA DISPARITA	sviluppo , integrità, fiducia, affermazione, mutamento, coscienza, legittimare, reazione, raggiungere, morale, regolamento, includere, attuale, femminista, leghista, decisione, battaglia, paritario, sindaco, solidarietà, uguaglianza, valore, media, chiesa, poliziotto, luciana_lamorgese, età, legge, risucire, cambiare, uguaglianza, parità	32	16,0
3	EDUCARE ALL'UGUAGLIANZA DI GENERE	franco_gabrielli, giovane, risposta, spirito, carabinieri, tentativo, giornalista, apertura, dovere, effettivo, figlio, posizione, partecipazione, giorgia_meloni, solenne, fisica, insistere, sicurezza, strategia, differenza, bisogno, determinazione, passare, sociale, trasformazione_politica, questione, giudicare, capire, impegno, bruno_vespa, origine, repubblica, arrivare, messaggi, ms5, richiamare, attenzione, web, assistenza, pd, protezione, relazione, iniziativa, carrefour, salario, importante, risalire, soluzione, libertà, istruire, inculcare	59	29,5
4	FIDUCIA NELLE ISTITUZIONI	fiducia, cooperazione, forze, ordine, dialogo, matrimonio, sentimentale istituzione, legale, tolleranza, comunitaria	11	5,5
5	ATTIVISMO	attivista, azione, campagna, pubblicitaria, mondiale, agire, manifestazione, nessuna_justificazione, esprimere, accoglienza, premier_conte, ragazza, rapporto, tema, parola, attenzione, italia, normale, approvare, marito, confine, buona, aiutare, visione, definire, italiano, psicologica, padre, matteo_salvini, sollevare, famiglia, carriera, onore, migliore, corpo, chiedere, educazione, recuperare, foto, ascoltare, storia, risorsa, inasprimento, obbligo, eliminazione, penale, pena	41	20,5
6	LEGGE	inasprimento, obbligo, eliminazione, penale, pena	5	2,5
7	COOPERAZIONE	sostegno, collaborazione, polizia, altruismo, solidarietà, adattamento, contesto, cambiare, permanente, donna, madre, bambino, televisione, proposta, gerarchia, giustizia, parlare, vicinanza, codice, scarpe_rosse, fermare, femmina, garanzia, pazienza, verbale, obiettivo, progredire	27	13,5
8	COMUNICAZIONE SOCIALE	comunicazione, stampa, istruzione, vincere, costruire, antiviolenza, riposizionare, coraggio, denunciare, movimento, emancipazione, centrale, fondamento, esigenza, sistema, genere, questione	17	8,5

Tabella 4.4. I Concepts Cluster positivi dell'Italia

Il TTR del *set* di dati negativo è del 17% con una presenza di *hapax* del 49%. Per la fase di *feature extraction* si è adottata una soglia di occorrenza pari a 6. Dalla classificazione delle prime 200 *features* estratte dal *dataset* negativo invece sono stati ricavati 9 diversi *cluster* di concetti (Tabella 4.5).

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	N. FEATURES	%
1	GARANTISMO GIURIDICO	elisa_pomarelli, garantismo, condanna, criminale, indifferenza, assassino, isolamento, ragazzo, vessare, istinto, reato, follia, lite, minore, aggressione, omicidio, fallimento, indagare, accusa, arma, affermazione, gelosia, offesa, ennesima, balordo, stupratore, beppe_grillo, pentirsi, media, gratuito, impossibile, tribunale, giudice, sistema, corrotto, pena, vergognoso, illegale,	38	19,0
2	MALTRATTAMENTO	subire, fragilità, assistere, bibbiano, bambina, guerra, sottomissione, privazione, oppressione, immorale, capire, colpire, odio, ipocrisia, sudditanza, negare, terrorismo, ancestrale, immigrazione, peggiore, estremo, urgenza, maschilismo, roberto_saviano	24	12,0
3	INEFFICACIA DEL SISTEMA GIURIDICO	bimba, nessuna_riduzione, ingiustizia, indulgenza, ergastolo, pedofilia, instabile, norma, legislazione, prostituzione, esemplare, premeditazione, coltello, attentato, orrore, frustrazione, galera, castrazione, ferire, pugnì, clandestino, possesso, strangolare, scontare	24	12,0
4	STRUMENTALIZZAZIONE DELLA COMUNICAZIONE	ragazza, segni, strumentalizzazione, giornalista, rituale, tradizione, discutere, giornale, errore, nascondere, infamia, grave, ledere, retroscena, famiglia, menzogna, padrone, cronaca, rabbia, repubblica, meschino, strage, punire, mentire, esercitare	25	12,5
5	PSICOPATOLOGIE	psicopatologia, onorare, psicosi, assassinio, efficacia, convalida, sottomettere, malattia, mentale, recupero, ricovero, ospedale, curare	13	6,5
6	BRUTALITA'	truce, stupro, carnefice, vergognarsi, massacro, attenzione, belva, politica, inganno, pestaggio, livido, insano, branco, storia, mostruoso, sanguinoso	15	7,5
7	AMORE CRIMINALE	partner, amore, timore, falso_incidente, passionale, problema, italiano, piangere, delitto, spietato, colpa, patetico, mostro, mura, calci, psicologica, sfruttamento, sessuale, razzismo, dannoso, spose_bambine, retorica, fisica, convivente, bugia	25	12,5
8	DISTRUZIONE DELL'IDENTITA'	identità, abuso, minaccia, impunito, obbligo_coniugale, denigrare, penale, iperbole, morte, botte, ostilità, moglie, insulto, denuncia, distrutto, disuguaglianza, carcere, rapporto, padre, fenomeno, gesto, camera	17	8,5
9	ABUSO DOMESTICO	domestico, martirio, indegno, violenza, donna, ratifica, battaglia, percosse, vittima, arresto, contrastare, italia, rapimento, malato, paura, casa, silenzio, vergogna, segregazione	19	9,5

Tabella 4.5. I Concepts Cluster negativi dell'Italia

I risultati della classificazione consentono di far emergere un contesto italiano che sul versante positivo è per lo più focalizzato sui temi dell'*educazione* all'*uguaglianza di genere*, mediante iniziative di partecipazione sociale e di collaborazione politico-istituzionale, sottolineando l'esigenza di una trasformazione attiva. Quest'ultima si definisce attraverso un'assidua attenzione sul problema, manifestazioni e campagne per non giustificare nessun tipo di violenza. Sul versante negativo invece la numerosità dei raggruppamenti è meno diversificata e attraverso tali *cluster* si evince in maniera piuttosto netta l'esigenza di combattere il garantismo nazionale e l'inefficacia del sistema

giuridico che tende a non assicurare una giusta condanna ai reati di violenza sulle donne.

4.4.2. Polarità semantica e Concepts Extraction dei tweets inglesi

La collezione dei 973 mila 478 *tweets* di lingua inglese indica un orientamento semantico prevalentemente negativo, con 644 mila 625 testi negativi e 328 mila 853 testi positivi (Figura 4.24). Le categorie emozionali indicano in maggior misura *paura e timore* (20.2%), *tristezza* (17.3%), *rabbia* (15.5%) e *disgusto* (15%).

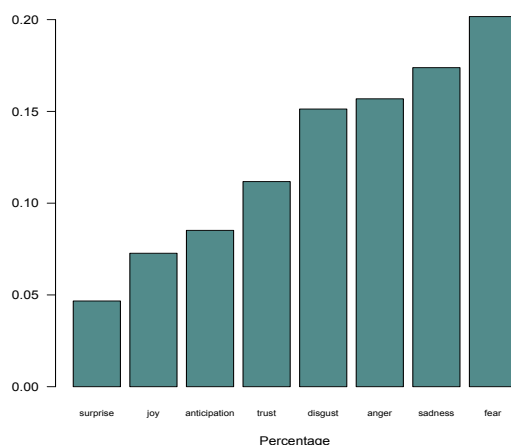


Figura 4.24. Distribuzione della polarità semantica dei tweets inglesi

Il *corpus* positivo ha presentato il TTR pari al 2% ed un conteggio di *hapax* del 31%. La soglia di frequenza utilizzata per l'estrazione delle *features* è di 21. Attraverso l'applicazione congiunta dell'*AC* e *Cluster Analysis*, le prime 200 *features* positive inglesi sono state raggruppate in 10 *cluster* concettuali positivi (Tabella 4.6). Mentre, dal *corpus* dei dati negativi con TTR uguale al 4% ed *hapax* uguali al 46%, le *features* sono state estratte con una soglia di frequenza pari a 3. Dalla classificazione delle prime 200 *features* rilevate dal *dataset* negativo, si sono ottenuti 8 gruppi concettuali negativi (Tabella 4.7).

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	N. FEATURES	%
1	GARANTISMO GIURIDICO	elisa_pomarelli, garantismo, condanna, criminale, indifferenza, assassino, isolamento, ragazzo, vessare, istinto, reato, follia, lite, minore, aggressione, omicidio, fallimento, indagare, accusa, arma, affermazione, gelosia, offesa, ennesima, balordo, stupratore, beppe_grillo, pentirsi, media, gratuito, impossibile, tribunale, giudice, sistema, corrotto, pena, vergognoso, illegale,	38	19,0
2	MALTRATTAMENTO	subire, fragilità, assistere, bibbiano, bambina, guerra, sottomissione, privazione, oppressione, immorale, capire, colpire, odio, ipocrisia, sudditanza, negare, terrorismo, ancestrale, immigrazione, peggiore, estremo, urgenza, maschilismo, roberto_saviano	24	12,0
3	INEFFICACIA DEL SISTEMA GIURIDICO	bimba, nessuna_riduzione, ingiustizia, indulgenza, ergastolo, pedofilia, instabile, norma, legislazione, prostituzione, esemplare, premeditazione, coltello, attentato, orrore, frustrazione, galera, castrazione, ferire, pugni, clandestino, possesso, strangolare, scontare	24	12,0
4	STRUMENTALIZZAZIONE DELLA COMUNICAZIONE	ragazza, segni, strumentalizzazione, giornalista, rituale, tradizione, discutere, giornale, errore, nascondere, infamia, grave, ledere, retroscena, famiglia, menzogna, padrone, cronaca, rabbia, repubblica, meschino, strage, punire, mentire, esercitare	25	12,5
5	PSICOPATOLOGIE	psicopatologia, onorare, psicosi, assassino, efficacia, convalida, sottomettere, malattia, mentale, recupero, ricovero, ospedale, curare	13	6,5
6	BRUTALITA'	truce, stupro, carnefice, vergognarsi, massacro, attenzione, belva, politica, inganno, pestaggio, livido, insano, branco, storia, mostruoso, sanguinoso	15	7,5
7	AMORE CRIMINALE	partner, amore, timore, falso_incidente, passionale, problema, italiano, piangere, delitto, spietato, colpa, patetico, mostro, mura, calci, psicologica, sfruttamento, sessuale, razzismo, dannoso, spose_bambine, retorica, fisica, convivente, bugia	25	12,5
8	DISTRUZIONE DELL'IDENTITA'	identità, abuso, minaccia, impunito, obbligo_coniugale, denigrare, penale, iperbole, morte, botte, ostilità, moglie, insulto, denuncia, distrutto, disuguaglianza, carcere, rapporto, padre, fenomeno, gesto, camera	17	8,5
9	ABUSO DOMESTICO	domestico, martirio, indegno, violenza, donna, ratifica, battaglia, percosse, vittima, arresto, contrastare, italia, rapimento, malato, paura, casa, silenzio, vergogna, segregazione	19	9,5

Tabella 4.6. I Concepts Cluster positivi del Regno Unito

La classificazione consente di individuare un contesto inglese che da un punto di vista concettuale positivo si basa sulla *comunicazione progresso*, che in altre parole si traduce nella divulgazione della difesa, della civilizzazione, della *leadership*, del *controllo*, dell'*autonomia* e delle *risorse*. Come nel caso dell'Italia, anche il Regno Unito presenta una propensione positiva nei confronti dell'attivismo sociale, mediante l'idea di *comunità* e attraverso *messaggi* di inclusione e sicurezza pubblica. Al contrario, il polo negativo predilige opinioni che si basano sull'oppressione familiare e quindi sul tema della violenza domestica che oscilla tra la retorica e la vittimizzazione con conseguenze annesse.

CLUSTER	CONCETTI	FEATURES	N. FEATURES	%
1	DOMINIO MASCHILE	patriarchal_culture, domination, psychological, disobedience, gang, offensive, threat suicide, trauma, kill, murder, crime, bill, authority, dark, leave, depany, lying, consequence, prison, economic_restriction, accuse, campaign, breaking_news, refuse, aversion, antifeminism, bout, institution, femicide	30	15,0
2	SUPREMAZIA MASCHILE	abnegation, interfering, masculine, delay, disequilibrium	5	2,5
3	OPPRESSIONE FAMILIARE	domestic, sick, knife, mortify, rhetoric, survivor, dad, shut, master, fragility, damage, pretend, depression, beating, social_role, victim, trap, bitch, house, conflict, act, oppression, livid, blame, mask, allegation, message, exploitation, tension, beast, insane, husband, violation, psychopatic, obsession, abortion, condemn, blood, abuse, show, report, emergency, hell, menace, assault, kidnapping, extort, killer, oppressor, suffer, daughter, torture	51	25,5
4	MALTRATTAMENTO E VIOLENZA	monster, martyr, indifference, story, disturb, question, journalism, extreme, pornography, disparity, word, anger, sad, bullying, pain, incest, violence, media burder, evidence, labour, remember, stop, abandonment, public, government, court, law, impulse, precaution, harassment, mortality, homophobia, fear, disgust, dignity, horror, destriy	38	19,0
5	GARANTISMO SOCIALE	guarantee, innocence, segregation, life released_assassin, sentence, judge, effect, discussion, inability, politics, family, disgust, young, girl, case, growth, radical, inhuman, disgrace, silence, death, coercion, sacrifice, global, person, ruthless, penal, contradiction, denunciation, risk, inadequate, fight	33	16,5
6	AMORE OSTILE	hostility, passion, weak, wrong, fallacy, female, love, scream, bad, street, cowardice, talk, respect, men, issue, wife, activism, shame, distruction, slavery, aggression, arrogant, contrast, bury	24	12,0
7	SFRUTTAMENTO DEL CORPO	commodify, body, sex, gender, hate, sexual, harassment, black, transsexual, discrimination, common	11	5,5
8	MISOGINIA	racism, muslim, genital_mutilation, misogyny, woman, gap, islam, brutality	8	4,0

Tabella 4.7. I Concepts Cluster negativi del Regno Unito

4.4.3. Polarità semantica e Concepts Extraction dei tweets francesi

Il dataset dei dati in lingua francese composto da 1 milione 55 mila e 619 tweets mostra un orientamento semantico prevalentemente negativo (Figura 4.25), come l'Italia e il Regno Unito. Più propriamente, il corpus della Francia rileva 866 mila 122 testi negativi e 189 mila 497 testi positivi. Per quanto riguarda le categorie emozionali, i tweets francesi esprimono la predominanza di emozioni quali rabbia (28%), paura (22.6%) e tristezza (19.02%).

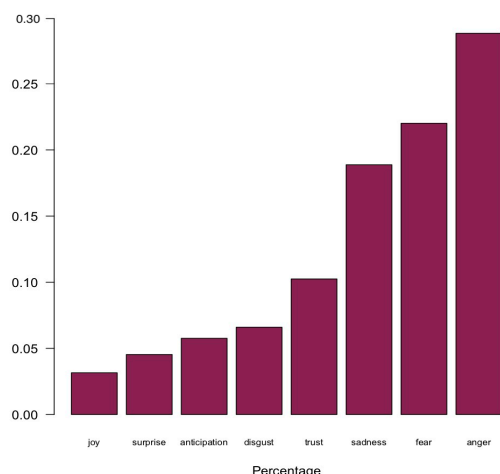


Figura 4.25. Distribuzione delle categorie emozionali dei tweets francesi

La collezione di dati positivi francesi ha indicato TTR pari al 2% e presenza di hapax del 42.3%. La soglia di frequenza adottata per rilevare le features è pari a 2. La Cluster Analysis applicata nel sottospazio dimensionale dell'AC ha permesso di rilevare 8 classi di concetti positivi (Tabella 4.7), classificando le prime 200 features.

CLUSTER	CONCETTI	FEATURES	N. FEATURES	%
1	EDUCARE	légalité, économique, civiliser, instruire, normalité, solidarité, génération, développement, orientation, indépendance, confrontation, mari, perspective, conjoint, inviolable, emploi, réaction, gagner, défense, caroline_de_haas, actif, réglage, député, augmenter, forme, reportage, humanité, racheter, principe, pas_de_compromis, européen, entité, expérience, école, délivrance	35	17,5
2	PROGRESSO ISTITUZIONALE	responsable, juridictionnel, normatif, énergiquement, marlene_schiappa, représenter, moral, réforme, rue, moyen, institution, résultat, activisme, message, constitution, dialogue, pays, officiel, aide, confiance, police, série, annonce, projection, témoin, sincérité, avantageux, culture, cible, couple, communication, positivement, media, ambitieux, genre, éthique, espoir, divulgation, phénomène, promulgation, protection, ministre	42	21,0
3	COMUNICAZIONE SOCIALE	équipe, lucidement, publicité, réconcilier, television, tournant, campagne, expérience, tolérance, occupation, reconnaître, emplacement, efficacement, courage, coexistence, identité, équilibre, vidéo, égalité, imprimer, connaissance, intégrité, secteur, divulgation, ensemble, liberté, mariage, information, sensibiliser, service, discuter, mots, conscience, bébé, valeur, famille, vie, témoignage, responsabilisation	39	19,5
4	IMPEGNO COLLETTIVO	mouvement, réhabiliter, engagement, différence, emmanuel_macron, public, travail, global, accepter, collectif, code, le_parisien, système, exposer, conjugal, ordre, dignité, volonté, féminisme, accueil, plan, conseil des ministres, exceptionnel, intérêt, convocation, réunion, solution, personnes, augmentation, encourager, renforcer	32	16,0
5	GIUSTIZIA	prudent, français, déclaration, juge, innocent, justice, honnêteté, restaurer, vérité, femme, paris, enquête, presse, france, intention, affirmation, exécutif, approbation, décret	19	9,5
6	SICUREZZA	sécurité, juridique, privé, planifier, précaution, décret, respecter, droit, important, relations, prévention, rôle, assistance, approbation, décider, anti-rapprochement	16	8,0
7	PROGRESSO STRATEGICO	protéger, programme, recherche, populaire, stratégie, progrès, nations_unies, soutien, europe, proposition, application, association, compter	13	6,5
8	COOPERAZIONE INTERNAZIONALE	repositionnement, coopération, international, national	4	2,0

Tabella 4.7. I Concepts Cluster positivi della Francia

Il *dataset* negativo francese invece ha mostrato un TTR dell'1.6% e la presenza di *hapax* è del 33.5%. La soglia di occorrenza considerata per estrarre le *features* è pari a 32 e dall'elaborazione di classificazione dei dati negativi, è stato possibile ricavare 10 *cluster* (Tabella 4.8).

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	N.FEATURES	%
1	SUPREMAZIA	extrémisme, suprématie, sujétion, déclaration	4	2,1
2	PSICOPATOLOGIA	traumatisme, psychose, enfer, pathologie, survivre, monstre, féroce, fou, psychologie, santé, destruction	11	4,8
3	MORTE	poignardée, humiliation, aggravant, responsabilité, folie, malade, enclose, meurtrière, cadavre, corruption, torture, accusation, mépris, indécent, abuser, éviter, meurtre, illégale, cimetière, prison, mentalité, maniaque, inhumain, bourreau, urgence, impitoyable, pédophilie, jalousie, subir, répugnant, acquittement, terrorisme, immoralité, risque, peur, auteur, assassin, aboutissement, gravité	37	21,4
4	DOMINIO E DISPARITA'	disparité, obliger, chômage, domaine, esclavage, privation, contrôle, imposer, mesure	9	2,8
5	MALTRATTAMENTO	physique, psychologique, mensonge, vengeance, mari, calomnie, sang agression, raison, raciste, injustice, peur, condamné, tromper, série, tyran, conjoint, prostituée, délire, pornographie, choc, déséquilibré, dégoût, cruel, dénoncer, web, victime, strangulation, mutilation, coercition, intime, pression, agresseur, erreur, agressif, instrument, illégitime	37	17,2
6	MISOGINIA	plainte, peine, vulgaire, violence, incapable, chienne, piège, discuter, suicide, évident, virilità, répulsion, conflit, personnalité, décision, conjugal, insulte, viol, supériorité, tergiversent, problème, système, prévalence	23	13,8
7	VIOLENZA FAMILIARE	famille, terreur, haine, contraste, colère, exploitation, homme, cage, profiter, dénigrer, trouble, misogynie, tourment, détérioration, manque, cauchemar, abandon, maison, crime, tromperie, amour, mordere, mariage_forcé, menace, consommer, mauvais, conséquence	26	13,8
8	CRIMINE PASSIONALE	passionnel, contusion, morte, dommage, fatalité, couteau, implorer, dévaster, horreur, tragique, arme, chantage, danger, corps, femme, sérieux, épouse, brûler, proie, domestique, acte, situationnelle	22	10,3
9	DISTRUZIONE DELL'IDENTITA'	relations, esclavage, genre, oppression, annihiler, détruire, identité, exploitation, solitude, mâle, isolation, ségrégation, indigne, violation, sexe, rabaisser, blessants, isolement, social	19	8,3
10	DISCRIMINAZIONE	raillerie, harcèlement, pouvoir, travail, protagonisme, intimidation, cynisme, discrimination, racisme, homophobie, question, ouvert	12	6

Tabella 4.8. I Concepts Cluster negativi della Francia

I risultati della classificazione mostrano un contesto francese prevalentemente fondato sul *progresso istituzionale* e sull'*educazione*, da un punto di vista concettuale positivo. Mentre, sul versante negativo l'opinione pubblica è principalmente focalizzata sulla morte, sulla follia omicida e sulla tortura, e si sottolineano le tipologie attraverso cui si può esperire il maltrattamento.

4.4.4. Polarità semantica e Concepts Extraction dei tweets spagnoli

Il *set* dei dati spagnoli formato da 1 milione 621 mila e 746 *tweets* esprime invece una maggioranza di contenuti positivi (Figura 4.26), al contrario dei risultati ottenuti per l'Italia, il Regno Unito e la Francia. Il *corpus* è costituito

da 838 mila e 45 testi positivi e 783 mila 701 testi negativi. Nello specifico, i contenuti testuali della Spagna esprimono prevalentemente *fiducia* (21%). Da quest'ultima emozione vi è il passaggio ad un versante emozionale prettamente negativo, costituito da *tristezza* (15.8%), *paura* (15.5%), *rabbia* (14.2%) e *disgusto* (13.7%).

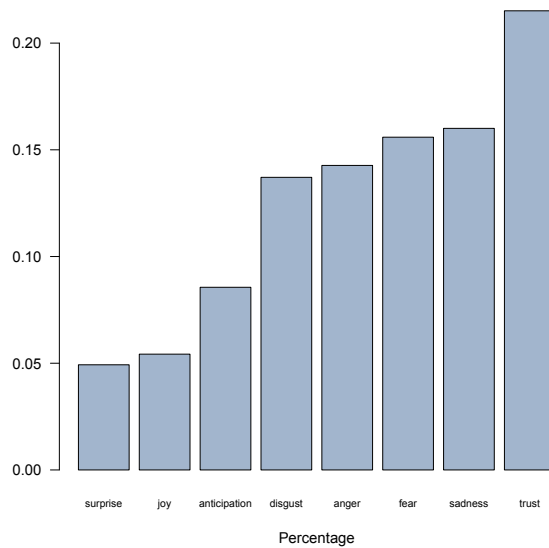


Figura 4.26. Distribuzione delle categorie emozionali dei tweets spagnoli

Il *corpus* positivo spagnolo ha mostrato TTR uguale a 1% e presenza di *hapax* del 16%. La soglia di frequenza indicata per rilevare le *features* è pari a 2. Dall'applicazione della *Cluster Analysis* sulle coordinate fattoriali dell'*AC* ottenute sul *set* di dati positivi spagnoli sono stati ricavati 9 gruppi concettuali positivi (Tabella 4.9), classificando le prime 200 *features*. Mentre, la collezione di dati negativa ha mostrato TTR uguale al 4.7% ed un contaggio di *hapax* del 48%. Il valore di occorrenza utilizzata come soglia è di 10. Dall'elaborazione classificatoria del *dataset* negativo si sono ottenuti 10 raggruppamenti concettuali (Tabella 4.10).

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	N. FEATURES	%
1	RIDURRE LA DISPARITA'	lealtà, onu, ética, empoderamiento, pareja, fortalecer, liderazgo, confianza, idealista, igualdad, programa, sentimiento, moral, innovar, principio, religión, eficiente, económico, lista, interés, recuperar, normativas, hablar, oficial, coraje, obtener, hospitalidad, mensaje, cambiar, poner, fin, eliminación, trabajar, inclusión, existencia, escuchar, contexto, creer, visibilizar, motivar, expresión, incesante, formación, tratamiento, significativos, congreso, normativa, aprobar, creación, responsabilidad, históricamente, innovación, salario, acceptable, derecho, conciliación, institución, educativo, flexible, proceso, interseccionalidad	61	30,5
2	PROGRESSO SOCIALE	europa, legalmente, recuperación, progreso, experiencia, aumentar, nacer, madrid, futuro, unión, compromiso, seguridad, perspectiva, necesidad, padre, humano, españa, estrategia, canción, éxito, género, medios, pública, nación, condición, sensibilización, pericia, información, diferente, persistente, privilegio, protección, estima, mejorar, fronteras, momento	35	17,5
3	EDUCARE ALLA CULTURA DELL'UGUAGLIANZA	feminismo, educación, equilibrio, igualdad, identidad, video, cultura, role, ley, social, público, papel, respeto, policías, comprensión, campaña, vuelco, hombre, realidad, instrucción, pensar, mentalidad, acompañar, conquista, imagen, historia, integral, promover	28	14,0
4	LEGGE	legal, tramitar, código, judicial, correcto, forma, favor, tema, tanacidad, revolución, estabilidad, uniformar, rebelarse, indispensable, honor, familia, sano, propuesta, legalidad, prevalecientes, respetar	21	10,5
5	GIUSTIZIA	justicia, proteger, reforzamiento, dios, internacional, comunidad, resultado, institución, sociedad, consciente, único, idea, global, definir, confirmar, victoria, reducción, estable, referente	19	9,5
6	ATTIVISMO	acción, seguir, asegurar, importancia, activo, positivo, concreto, modelo, solución, ayuda, vida, ventaja, desplazamiento, mensaje, colectivo, efecto	16	8,0
7	AZIONE E COOPERAZIONE	mundial, cambio, grupo, apoyo, esperanza, acción, atención, precaución, intervención, reforzar, asistencia, nacimiento	12	6,0
8	LIBERTA'	mundo, libertad, nacional, rosa, capacidad, participar, vida, niña	8	4,0

Tabella 4.9. I Concepts Cluster positivi della Spagna

La classificazione delle *features* spagnole ha evidenziato sul piano positivo concetti le cui parole in maggior misura afferiscono alla *riduzione della disparità di genere* attraverso l'*empowerment*, la *fiducia*, l'*evoluzione e l'innovazione etica, economica e normativa*. La trasformazione si sviluppa inoltre attraverso il *progresso sociale* della pianificazione del *futuro* che coinvolge l'*Europa*. La prospettiva sociale negativa invece è basata sulla *misoginia*, sul *maschilismo*, il *patriarcato* e sugli *stereotipi di genere*. Un raggruppamento concettuale a parte si focalizza solo sul sesso e sullo *sfruttamento del corpo*, indicando una cultura che tende a sminuire la donna.

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	N. FEATURES	%
1	MISOGINIA	misoginia, cosificación, machista, patriarcado, salarial, feminicidio, promover_desigualdad, acoso, importancia, maltrato, propagación, estereotipo, psoe, pardo_devera, aborto, brecha, taller, femenino, agresión, pacto, marido, custodia, hostil, agustín_laje, europa, disposición, masculino, distinción, subordinación	29	14,5
2	MANIPOLAZIONE MEDIATICA	mauricio_macri, mastro, feminismo, desprecio, falsas, izquierda, capitalismo, matar, rasgón, cabeza, atención, situación, llama, seguro, policías, cólera, sufrir, orden contraste, cárcel, tema, explotación, access, subversión, medios, ciberacoso, reticencias	27	13,5
3	SESSO	desafío, retirar, sexo, redundancia, paula_fraga, porno, abolicionista, desnudo, enseñar, reproductiva	10	5,0
4	SFRUTTAMENTO DEL CORPO	cuero, lidere, estúpidas, campaña, cultura, juicio, jaula, techo, lgbt, enemigo, medida, público, silencio, polémico, scapar, problema, educación, manos, saca, crimen, españa, titular, defensores, lucha, revolución, policía, periodista, seguir	28	14,0
5	VIOLENZA FAMILIARE	familia, casa, violencia, historia, abuso, sexual, voluntad, destrucción, signo, obsesión, perspectiva, género, verdugo, pelea, pareja, privación, opresión, desfavorable, terror, demolición, comunidad, privado, obstáculo	23	11,5
6	COMPORAMENTO PSICOPATOLOGICO	prostituta, falta, latido, nacional, atroces, odio, delito, refutar, enfermedad, sangre, prisión, social, ideología, fiera, importante, patológico, madrid, niño, mental, derechos, grito, afectar, negativamente, salud, conflicto	25	12,5
7	BRUTALITA'	psicosi, indigno, torturas, animal, piojoso, indiferencia, cadáver, abandono, matanza, trágico, comprobar, incorrecto, condenado, reprimir, exclusión, patología, clase, régimen, sociedad, trampa, intimidación, agravar	22	11,0
8	VIOLENZA DOMESTICA	doméstica, violación, anormales, monstruo, extremo, corrupción, invisible, irracional, clima, locura, asesinato, culpa, conciencia, vergonzoso, dolor, especie, degradante	17	8,5
9	DISCRIMINAZIONE	discriminando, burlarse, homofobia, racismo, mujer, machismo, implacable, hombre, critica, descuido, ineficaz, galopante, implícitamente	13	6,5
10	FALLACIA DEL SISTEMA GIURIDICO	penal, código, regulación, jurídico, sistema, falacia	6	3,0

Tabella 4.10. I Concepts Cluster negativi della Spagna

4.4.5. Polarità semantica e Concepts Extraction dei tweets ucraini

La collezione dei *tweets* di lingua ucraina costituita da 103 mila 218 *tweets* evidenzia una polarità semantica prevalentemente positiva (Figura 4.27), con 65 mila 779 testi positivi e 37 mila 439 testi negativi, come nel caso della Spagna. La distribuzione delle emozioni pone al primo posto la *fiducia* (39.9%) e *gioia* (15.4%) per poi trasferirsi sul polo negativo del *disgusto* (11.1%), della *paura* (10%), *tristezza* (9.1%) e *rabbia* (9%).

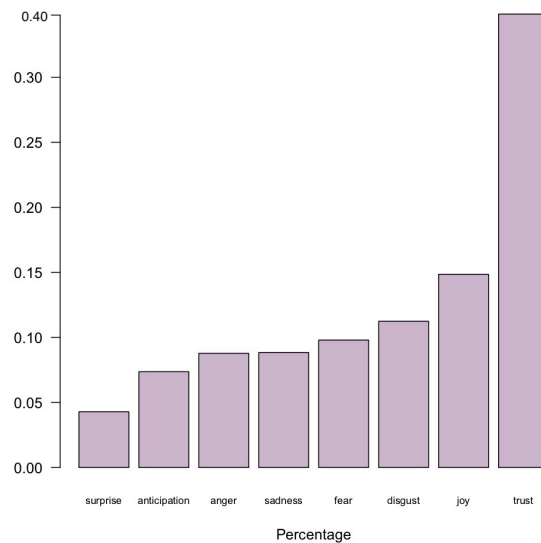


Figura 4.27. Distribuzione delle categorie emozionali dei tweets ucraini

Il corpus positivo ucraino ha rilevato un TTR del 15% e la presenza di *hapax* è del 49.9%. Per estrarre le *features* si è considerata una soglia di frequenza pari a 2. Attraverso la combinazione dell'*AC* e della *Cluster Analysis* sul dataset positivo ucraino la classificazione delle prime 200 *features* ha consentito di estrarre 9 raggruppamenti concettuali positivi (Tabella 4.11). Il corpus negativo invece ha indicato il TTR uguale all'11.2% ed il conteggio degli *hapax* è del 46.3%. La soglia di occorrenza utilizzata per l'estrazione delle *features* è pari a 2. Dall'elaborazione di classificazione del dataset negativo è stato possibile rilevare 7 gruppi concettuali negativi, categorizzando le prime 200 *features* (Tabella 4.12).

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	TRADUZIONE	N. FEATURES	%
1	VALORE DELL'IDENTITA'	цінності, пріоритети, ідентичність, можливий, досвід, колективний, разом, подумав, гроші, довіру, пожавите, мужність, істотний, догляд	valori, priorità, identità, possibile, esperienza, collettivo, insieme, pensiero, denaro, fiducia, rivivere, coraggio, essenziale, cura	14	7,0
2	PROGRESSO	Павло Клімкін, свідомість, влада, Україна, закон, необхідність, принцип, європейський, необхідний, повага, час, Порошенко, трансформація, проект, підтвердження, зміна, щире закінчення, негайно, сьогодні, без вини, стан, прогрес, роботу	Pavlo Klimkin, coscienza, potere, Ucraina, legge, necessità, principio, europeo, necessario, rispettare, tempo, Poroshenko, trasformazione, progetto, conferma, cambiamento, finale sincero, subito, oggi, senza colpa, condizione, progresso, posto di lavoro	23	11,5
3	RIDUZIONE DELLA DISUGUAGLIANZA	потреба, свобода, знання, профілактика обізнаність, новини, прогрес, воля революція, світ, визначення, очікування, здатність, вірність, включеність, президент, додати, захист, новий, втручатися, відновити, простір, завзятість, законодавчий, глобальний, valere_la_pena, визначеність, незалежність, київ, економічний, соціальний, рівність, легітимний, пояснити, підтримка, психологічний, розгортається, сам, воля, добро, справа, національний, тиша, чинити опір, пробувати, психологічної, корисний, замовлення	bisogno, libertà, sapere, prevenzione, consapevolezza, notizia, progresso, volontà, rivoluzione, mondo, definizione, aspettativa, abilità, fedeltà, inclusione, presidente, aggiungere, protezione, nuovo, interferire, recuperare, spazio, tenacia, legislativo, globale, valere_la_pena, certezza, indipendenza, kiev, economico, sociale, uguaglianza, legittimo, spiegare, supporto, psicologico, svolgimento, se stesso, volontà, buono, causa, nazionale, silenzio, combattere, tentare, psicologico, utile, costume	48	24,0
4	COOPERAZIONE PER LA TRASFORMAZIONE	виживши, управління, захист, толерантність, цілісність, співтовариство, чітко, ефективно, вчити, Російська Федерація, підність, культура, виховувати, запитувати, корисно, жест, чудово, відкликатись, вирішити, бачити, головне, голос, допомога, релігія, надія, потужність, стабільний, початок, чоловік, Зеленьський, пристосуватися, слуга, гордий, обирати, реакція, місце, прояв, автономія, Росія, країна, конвенція, в.о., сумісні, спілкування, українська, нейтральна, пам'ять, міністерство соціальної політики, подія, команда, російська, рятувальна, міжнародна, виявити, напрямок, сили, експерт, відповідь, прибуток, марш, пропозиція, дзвінок, підтримка, сервіси соціальні, координують, доступними, момент, анонімно	sopravvissuto, gestione, difesa, tolleranza, integrità, insieme, chiaramente, efficace, insegnare, Federazione Russa, dignità, cultura, educare, domanda, utile, gesto, grande, ritirare, decidere, vedere, principale, voce, aiuto, religione, speranza, potere, stabile, inizio, marito, Zelensky, adattare, servitore, orgoglioso, scegliere, reazione, posto, manifestazione, autonomia, Russia, paese, convenzione, recitazione, compiacente, comunicazione, ucraino, neutro, memoria, Ministero delle politiche sociali, evento, squadra, russo, salvataggio, internazionale, rivelare, direzione, forze, esperto, risposta, profitto, marcia, offerta, chiamare, sostegno, servizi sociali, coordinare, disponibile, tempo, anonimo	68	34,0
5	ATTIVISMO	прапор, вшанування пам'яті, сім'я, дипломат, власність, юрист, комітет, інституційна діяльність, лінія, червоний, ОБСЕ, сфера, позитивно, сторінка, центр боротьби з насильством, невинний, зобов'язання	bandiera, commemorazione, famiglia, diplomatico, proprietà, avvocato, comitato, attività istituzionale, linea, rosso, OSCE, ambito, positivamente, pagina, centro antiviolenza, incessante, impegno	17	8,5
6	SICUREZZA	розслідування, збереження, безпека, дівчино, захищати, безпечний, громадськості	inchiesta, salvare, sicurezza, ragazza, proteggere, garantire, pubblico	7	3,5
7	FIDUCIA NELLE ISTITUZIONI	рівний, інститут, засоби, триумф, спільність, автономія, програма, честолюбство, дію	uguale, istituzione, mezzi, trionfo, comunità, autonomia, programma, ambizione, effetto	9	4,5
8	COMUNICAZIONE	шанувати, спілкуватися, говорити, повідомлення, життя, підхід, орієнтація	onore, comunicare, parlare, messaggio, vita, approccio, orientamento	7	3,5
9	LEGGE	затверджують, правові, радикальні, незворотні зміни, енергія, юрисдикція, легальність	approvare, cambiamento giuridico, radicale, irreversibile, potere, giurisdizione, legalità	7	3,5

Tabella 4.11. I Concepts Cluster positivi dell'Ucraina

La categorizzazione delle *features* ha consentito di osservare, sul versante positivo, un quadro socio-culturale ucraino prevalentemente interessato alla *cooperazione* per la trasformazione. Per meglio dire, le parole che appartengono a questo gruppo si riferiscono alla *tolleranza*, alla *difesa*, all'*efficacia*, all'*autonomia*, alla *comunicazione*, al *Ministero per le Politiche Sociali*. Sul piano negativo, invece, la prevalenza di *features* afferisce al raggruppamento concettuale della *brutalità*, espressa attraverso la *deturpazione*, l'*oscenità*, la *subordinazione*, la *violenza minorile*, il *suicidio*, la *disperazione*, la *molestia*, il *sangue*.

CLUSTER	ETICHETTE CONCETTUALI	FEATURES	TRADUZIONE	N. FEATURES	%
1	AUTORITA' MASCHILE	дискримінація, патріархат, позбавлення, погіршення, авторитарність, підбурювання, агітація, гроші, розчарування, проблема байдужість, падіння, гнітюче погано, шахрайство, кислота, Україна, хворі, погода, фатальність, дурний, переляк, цна, обкладинка, сім'я, нижче, недобросовісний, мстити, заборона, незбалансованість, вище, будинок, садизм	discriminazione, patriarcato, privazione, deterioramento, autoritario, incitamento, agitazione, soldi, frustrazione, problema indifferenza, caduta, deprimente cattivo, frode, acido, Ucraina, malato, tempo, fatalità, stupido, spavento, prezzo, copertura, famiglia, inferiorità, ingiusto, vendicare, vietare, disequilibrio, superiore, casa, sadismo	33	16,5
2	SESSO	огидний, нелегітимний, секс, жорстокий, сексуальне_насильство, нецивілізований, зґвалтування, війна, солдат, почуття провини, необґрунтований, зловживання, безпорадний, тортури	disgustoso, illegittimo, sesso, brutale, violenza sessuale, incivile, stupro, guerra, soldato, senso_di_colpa, irragionevole, zlovживання, indifeso, tortura	14	7,0
3	BRUTALITA'	звірство, криза, прикриття, зневіра, роки, переслідування, крайній, покарання, вогонь, порушувати, ганьба, маніяк, необхідне, непристойність, зброя, підлеглий, тюрма, вирок, колоти, необачність, тварина, гнів, воля, раб, темний_радж, зґвалтування, питання, невдача, особистість кїв, терор, страшний, запуганий, небезпека, закон, світ, насильство, кордон, проблема, самоґубство, скаржитися, бруд, знищити, збегтеження, проти, цькування_on_mіноґи, Порошенко, вбити, пістолет, кров, людина, зневажливий, загроза, україна, повернутися, побачити, важко, розшук, гідота, свавілля, сила, вульгарний, глузування, бідність, виснаження, несправедливість, країна, Російська федерація, траур, українська, пам'ять, вбивця, мотузка, відчай, незаконний, відсутність, розгортання, сука, травма, смерть, ліквідувати, удар_ножем	brutalità, crisi, coprire, deturpazione, anni, persecuzione, estremo, punire, fuoco, violare, vergogna, maniaco, richiesto, oscenità, arma, subordinato, prigionie, pena, accoltellare, imprudenza, animale, rabbia, volontà, schiavo, rabbia_oscura, stupro, domanda, sfortuna, identità, kiev, terrore, terribile, confuso, pericolo, legge, mondo, violenza, confine, problema, suicidio, lamentarsi, sporcizia, distruggere, imbarazzo, contro, molestia_su_minori, Poroshenko, uccidere, pistola, sangue, uomo, spregevole, minaccia, Ucraina, tornare, vedere, difficile, ricercato, abominio, tirannia, forza, volgare, beffa, povertà, esaurimento, ingiustizia, paese, Federazione Russa, lutto, ucraino, memoria, assassino, corda, disperazione, illegale, assenza, svolgimento, punire, puttana, ferita, morte, eliminare, pugnolare	83	41,5
4	PSICOPATOLOGIA	відхилення, психічний, хвороба, психопат, збочення, вибух, звільнення, гнівний, божевільно, переслідування, поганий, марення	deviazione, mentale, malattia, psicopatico, perversito, esplosione, rilasciare, rabbioso, manicomio, persecuzione, cattivo, delirio	11	5,5
5	INEFFICACIA ISTITUZIONALE	федерація, усамітнення, інституціональність, нелегальний, непридатний, необхідність, померти, затримка, постійні_пошкодження, вбивство, уряд, недовіру, нездатний, невідповідний, дисфункціональних, загальний	federazione, solitudine, istituzionale, illegale, inadatto, esigenza, morire, ritardo, danno_permanente, uccisione, governo, diffidenza, incapace, inadeguato, disfunzionale, globale	15	7,5
6	MISOGINIA	різанина, мізоґінія, автономія, сварка, скраб, ненависть, приниження, зарозумілість, потужність, зневажати, причина, повія	strage, misoginia, autonomia, battibecco, macchia, odio, umiliazione, arroganza, potenza, disprezzare, causa, prostituta	12	6,0
7	STRUMENTALIZZAZIONE MEDIATICA	Телеканал, невідчужуваність, відликання, втеча, тиша, ЗМІ, безчестя, обман, окупація, влада, конфлікт, пастка, рос_види, ОБСЕ, не помічаю, травма_голови, пряма, червона, помилк, адміністративне правопорушення, преса, Телешоу, фантастика, звинувачувати, слабкість, маніпулювати, стріляти, біль, озлоблений, приховати, труп	canale televisivo, inalienabilità, ritirare, scappare, omertà, media, disonore, ingannare, occupazione, potere, conflitto, trappola, russo venire_fuori, OSCE, trascurare, trauma_cranico, diritto, rosso, errore, reato amministrativo, stampa, spettacolo televisivo, finzione, accusare, debolezza, manipolare, sparare, dolore, amareggiato, coprire, cadavere	32	16,0

Tabella 4.12. I Concepts Cluster negativi dell'Ucraina

4.4.6. Dalla rilevazione dei concetti alla comparazione Cross-language

La rilevazione dei concetti dai diversi corpora multilingue comparabili ha portato alla luce un aspetto alquanto interessante: malgrado le *features* estratte da ciascuna collezione di *tweets* positiva e negativa monolingue siano diverse da lingua a lingua, attraverso la loro classificazione sulle coordinate fattoriali esse racchiudono molti significati comuni. Per meglio dire, quando le parole delle diverse lingue vengono categorizzate in raggruppamenti, queste nel

complesso presentano una semantica concettuale comune, nonostante le differenze linguistiche esistenti. Quindi, malgrado ogni *cluster* presenti al proprio interno *features* diverse, così come diverso è il numero delle parole contenute in ciascun raggruppamento, nella maggior parte dei casi, la procedura di classificazione ha restituito concetti semanticamente equivalenti tra le lingue. Questo aspetto tende ad indicare che malgrado il multilinguismo e multiculturalismo dei diversi vocabolari monolingue, il fenomeno della violenza contro le donne può essere interpretato attraverso un universo concettuale simile.

Tale spazio concettuale comune tra le lingue può essere ulteriormente scomposto e studiato. In particolar modo, si possono individuare quali sono i raggruppamenti concettuali che più accomunano le lingue e si può determinare quali lingue siano eventualmente più simili tra loro. A tal fine, l'insieme dei concetti rilevati per le diverse lingue identificano uno spazio concettuale multilingue su cui operare il confronto. I *cluster* concettuali individuati attraverso la classificazione delle varie *features* monolingue diventano così le dimensioni di riferimento per la comparazione tra le lingue.

Per la costruzione dello spazio dimensionale multilingue, si ritiene opportuno tuttavia non considerare quei raggruppamenti che tendono a rafforzare uno stesso concetto e che quindi non fanno altro che ripetere un'etichetta semantica già individuata (come ad esempio *abuso domestico* e *violenza familiare* oppure *dominio* e *supremazia*). La scelta consiste quindi nel considerare solo un'etichetta e racchiudere le *features* del *cluster* duplicati in un unico raggruppamento concettuale, al solo scopo di evitare ripetizioni concettuali ridondanti. Pertanto, lo spazio concettuale *cross-language* di questo studio si basa essenzialmente su *12 raggruppamenti concettuali positivi* e *15 raggruppamenti concettuali negativi*.

Nella *Tabella 4.13* viene presentata la sintesi dei domini concettuali che sono stati esplorati attraverso l'analisi comparativa multilingue proposta in questa ultima fase della tesi.

CONCETTI POSITIVI	CONCETTI NEGATIVI
Riduzione della disparità	Discriminazione
Fiducia Istituzionale	Abuso domestico
Comunicazione Sociale	Amore criminale
Identità	Dominio maschile
Sicurezza	Misoginia
Libertà	Maltrattamento
Cooperazione	Distruzione dell'identità
Progresso strategico	Brutalità
Legge	Morte
Attivismo	Garantismo nazionale
Educazione all'uguaglianza di genere	Sfruttamento del corpo
Giustizia	Fallacia del sistema giuridico
	Sesso
	Psicopatologia
	Strumentalizzazione mediatica

Tabella 4.13. Sintesi dei concetti multilingue rilevati

La chiave di lettura e di confronto unificata tra le lingue si basa fondamentalmente su una percezione pubblica che vede la società impegnata in una continua lotta per l'equilibrio di potere e per la parità tra uomini e donne. La riduzione di tale disparità viene infatti interpretata come uno degli aspetti propositivi di questa analisi, nel tentativo di rovesciare l'ordine naturale delle cose. Sempre sul versante positivo dell'opinione pubblica, si osserva la fiducia e l'affidamento nelle istituzioni, quali organizzazioni e forze dell'ordine, in un clima bisognoso di cambiamenti e alla ricerca di tutela, protezione, affermazione della propria sicurezza, libertà ed identità. Sono questi gli elementi concettuali che emergono dall'analisi complessiva dei corpora positivi multilingue. Dalla determinazione delle unità di analisi di ordine concettuale si è rilevato inoltre di particolare interesse l'attivismo, inteso come un insieme di azioni e collaborazioni volte a movimentare le coscienze e a produrre risultati sul piano pratico. Si delineano infatti operazioni concrete che coinvolgono la legge, richiamano la giustizia e si fondano sulla cooperazione nazionale ed internazionale attraverso piani di sviluppo e strategie.

Sul versante negativo invece i concetti delle diverse lingue si sviluppano basandosi sulla discriminazione, sulla sudditanza delle donne rispetto al dominio maschile e sul maltrattamento, spostandosi lungo un asse interpretativo di vittimizzazione familiare e coniugale. I temi delineati si focalizzano altresì sulla sessualità e sullo sfruttamento del corpo, sulla misoginia e sulla criminalizzazione della violenza attraverso la brutalità e la morte. Le lingue

indagate ripropongono anche il tema delle psicopatologie e dei disturbi mentali connessi all'efferatezza delle dinamiche violente.

I corpora negativi hanno consentito inoltre di rilevare un'insofferenza per l'inefficacia del sistema giuridico e la presenza di un garantismo che non assicura tutela o giustizia alle vittime. Infine, si richiama il tema della manipolazione ed esasperazione dei mezzi di comunicazione.

Al termine della fase di *Concepts Extraction* di ciascuna lingua e più propriamente dopo aver identificato lo spazio multilingue per il confronto, si è passati alla codifica dei dati attraverso la costruzione della tabella *termini x concetti*, sia per i concetti positivi che per i concetti negativi di ogni lingua.

Nella rappresentazione matriciale è stata codificata in forma disgiuntiva completa, l'appartenenza delle diverse *features* monolingue ad ogni *k*-esimo concetto positivo o negativo multilingue (*Tabella 4.13*).

Tenendo poi conto della proporzione di termini afferenti ad ogni concetto, attraverso operazioni proprie dell'algebra matriciale, dalla tabella *termini x concetti* si è ricavata la *document-concept matrix* (Misuraca *et al.*, 2019; § 3.2.2.4) positiva e negativa di ciascuna lingua, che esprime l'appartenenza dei diversi concetti ai diversi *tweet*.

Tuttavia, per evidenziare se i concetti delle diverse lingue hanno relazioni simili tra loro e quindi se sono associati nello stesso modo all'interno dei vocabolari monolingue, dalla *document-concept matrix* si è ottenuta la matrice *concetti x concetti* positiva e negativa di ciascuna lingua (§ 3.2.2.4). In questo modo, moltiplicando la *document-concept matrix* positiva o negativa di ciascuna lingua con la sua trasposta, si sono ottenute le relazioni tra i diversi concetti positivi e negativi di ogni lingua. Le co-esistenze tra i concetti permettono così di indagare la relazione tra i campi concettuali delle diverse lingue (Norma Internazionale, Terminologia-Vocabolario, ISO 1087, 1995), evidenziando quindi un'eventuale inclusione o assenza di coincidenza.

Per ultimo, al fine di per ottenere una configurazione matriciale che tenga conto allo stesso tempo sia delle relazioni tra i concetti positivi che delle relazioni tra i concetti negativi, è stata creata un'unica matrice dei concetti unificata, alle cui etichette è stato aggiunto un elemento identificativo positivo (+) o negativo (-) allo scopo di non perdere l'orientamento semantico. A livello pratico, sono stati costruiti due blocchi matriciali distinti dove da una parte sono presenti gli incroci tra concetti solo positivi e dall'altra sono presenti gli incroci tra concetti solo negativi. Nella stessa matrice, sono stati poi generati altri due blocchi

distinti, nel cui incrocio tra i concetti positivi e quelli negativi sono state espresse solo relazioni nulle. Quest'ultima configurazione matriciale rappresenta la tabella del *sistema concettuale* positivo e negativo di ogni lingua sul tema della violenza contro le donne.

Per confrontare le relazioni esistenti tra i concetti estratti dal vocabolario di ciascuna lingua e quindi per individuare un'eventuale similarità tra le lingue si è passati all'ultima fase della proposta *cross-language*, che si realizza confrontando simultaneamente le diverse matrici *concetti x concetti* unificate di ciascuna lingua, in un comune spazio dimensionale multilingue.

4.5. La GPA per il confronto tra i concetti in unico spazio dimensionale multilingue

Per l'analisi comparativa dei concetti si è adottata la tecnica statistica dell'approccio procustiano generalizzato o GPA (§ 3.2.4.2), il cui metodo ha permesso di confrontare i dati originari delle diverse matrici in maniera simultanea e sullo stesso piano dimensionale. Nel caso di fattispecie, la GPA è stata applicata alle $m=5$ configurazioni matriciali, dove ciascuna tabella $m \times m$ nella k -esima riga indica le coordinate dello stesso punto $P_k^{(k')}$ ($k'=1, \dots, 27$; $k=1, \dots, 27$). Le $m=5$ matrici sono state allo stesso tempo *traslate*, *ruotate* e *scalate*. Attraverso il processo di trasformazione ortogonale si è ottenuta la configurazione media (o *consensus*) delle lingue indagate per ciascun concetto. Il massimo accordo tra le lingue è stato studiato da un punto di vista geometrico mediante la mappa fattoriale e da un punto di vista algebrico con l'*Analisi della Varianza*.

4.5.1. La Generalized Procrustes Analysis Map: lingue a confronto

La visualizzazione grafica consiste per l'appunto in una mappa fattoriale, sulla quale ogni concetto identifica un centroide G_k ($k=1, \dots, 27$) attorno a cui si originano cinque segmenti che rappresentano le lingue. La direzione che ogni segmento assume dipende dai risultati dell'adattamento procustiano, mentre la sua distanza o vicinanza e la sua estensione sul piano (rispetto alle altre configurazioni linguistiche) esprime il suo accordo o disaccordo.

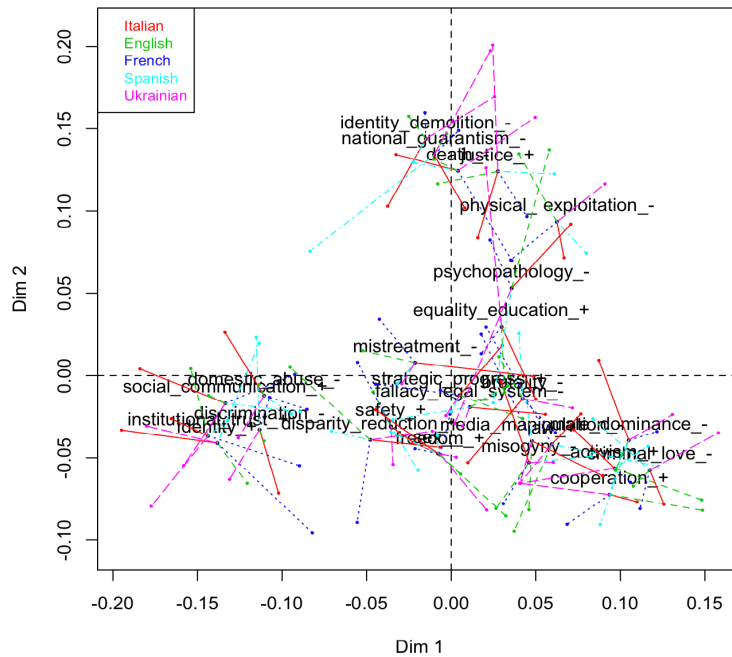


Figura 4.28. Mappa GPA del confronto cross-language

La Mappa GPA del confronto *cross-language* tra i concetti (Figura 4.28) evidenzia da subito alcune sovrapposizioni sulla trama fattoriale, le quali non permettono un'intelligibilità grafica pienamente nitida. Pertanto, non è possibile stabilire geometricamente quale concetto o quali concetti mostrino un maggiore consenso tra le configurazioni. Tuttavia, si può osservare ad esempio il concetto *discrimination*, il quale rispetto alla sua rappresentazione media sembra mostrare distanze più contenute tra i segmenti e quindi ipotizzare un maggiore *consensus* tra le lingue. Tra gli altri punti sul piano, è possibile altresì indicare il concetto *activism* che evidenzia un tangibile disaccordo da parte dell'Italia rispetto al Regno Unito, la Francia, la Spagna e l'Ucraina. Il consenso o accordo tra le lingue rispetto a ciascun concetto o centroide si interpreta in termini di similarità, dove per similarità si intende la somiglianza del sistema concettuale monolingue utilizzato per parlare della violenza contro le donne. Si può dunque affermare che i diversi sistemi concettuali monolingue siano maggiormente simili tra loro nell'argomentare il tema della *discriminazione*. Al contrario, il sistema concettuale italiano differisce da quello inglese, francese, spagnolo ed ucraino per quanto riguarda l'*attivismo*. Un'ulteriore

interpretazione si può riferire al concetto della *morte* che vede lo spagnolo come la lingua che maggiormente si distanzia dalle altre.

Anche la *riduzione della disparità di genere* e la *libertà* sono concetti attraverso le cui relazioni si può individuare il minor *consensus* da parte della lingua inglese. Diversamente, la lingua francese mostra una certa distanza geometrica dalle altre configurazioni, in relazione al concetto *dell'identità*, mentre l'Ucraina è particolarmente in disaccordo per quel che concerne le *psicopatologie* e l'*educazione alla parità di genere*.

4.5.2. L'Analisi della Varianza Generalizzata: PANOVA

Per analizzare con precisione quali siano i concetti che esprimono una maggiore similarità tra i sistemi concettuali delle lingue è bene corredare lo studio geometrico con l'*Analisi della Varianza (PANOVA)*, come suggerito dallo stesso Gower (1975). I risultati di PANOVA stabiliscono il maggiore o minore *consensus* in termini di analisi dei residui.

Più i valori residui sono alti e più esiste un basso consenso tra le lingue, mentre più i residui si presentano bassi e più le relazioni tra i concetti sono simili nelle diverse lingue. Questo significa che il concetto *misoginia* che presenta il valore residuo in assoluto più basso, ovvero dello ≈ 0.52 (*Tabella 4.14*) indica il massimo accordo tra le lingue. Quindi, i diversi sistemi concettuali monolingue sulla violenza contro le donne utilizzano concetti simili tra loro per parlare della *misoginia* e in un'ottica *cross-language* si può affermare che in assoluto, il massimo consenso tra i vocabolari monolingue riguarda proprio la tendenza alla repulsione e all'avversione verso le donne.

Concepts	SS_{fit}	SS_{residual}	SS_{total}
disparity_reduction_+	2.081912	0.7109582	2.792870
institutional_trust_+	3.822182	0.7832996	4.605482
social_communication_+	3.007899	0.6792613	3.687160
identity_+	5.732384	0.7307586	6.463143
safety_+	2.191701	0.7291259	2.920826
freedom_+	1.894324	0.7347733	2.629097
cooperation_+	3.047156	1.0880977	4.135254
strategic_progress_+	2.648261	0.6527114	3.300972
law_+	2.286506	1.0129619	3.299468
activism_+	2.702436	1.1443190	3.846756
equality_education_+	3.537350	0.6629391	4.200289
justice_+	3.143484	0.9534504	4.096934
discrimination_-	3.916751	0.6078980	4.524649
domestic_abuse_-	2.965226	0.7144446	3.679670
criminal_love_-	3.787444	0.6948122	4.482257
male_dominance_-	2.832905	0.8224504	3.655355
misogyny_-	2.924233	0.5225146	3.446748
mistreatment_-	2.621175	0.6081878	3.229363
identity_demolition_-	2.787092	0.6514189	3.438511
brutality_-	2.818503	0.7838995	3.602402
death_-	3.089959	0.6682248	3.758184
national_guarantism_-	2.940094	0.6158925	3.555986
physical_exploitation_-	2.767089	0.8002814	3.567370
fallacy_legal_system_-	2.396004	0.6118880	3.007891
sex_-	2.393278	0.6241496	3.017427
psychopathology_-	2.873599	1.0295387	3.903137
media_manipulation_-	2.100010	1.0527884	3.152798
sum	79.308954	20.6910457	100.000000

Tabella 4.14. PANOVA: Analisi della varianza dei concetti

Osservando i residui via via più alti, dopo *misoginia*, si vince con valore pari a ≈ 0.607 il concetto *discriminazione*, che mette in evidenza una serie di attributi, comportamenti e azioni private e sociali che impongono un trattamento non paritario tra l'uomo e la donna. Quindi, i campi concettuali che si riferiscono alla *discriminazione* indicano un considerevole accordo tra le lingue.

Dall'analisi dei residui, emerge inoltre il consenso riferito al *maltrattamento* (≈ 0.608), alla *fallacia del sistema giuridico* (≈ 0.611) e al *garantismo nazionale* (≈ 0.615), i quali campi concettuali affini tra le lingue indicano chiaramente l'insoddisfazione popolare riguardo agli strumenti giuridici e legislativi che dovrebbero contrastare il fenomeno, sottolineando l'esistenza di un garantismo nazionale che rende inefficace l'intero sistema di azione. Il tema del *sex* (≈ 0.62) è altresì particolarmente dominante, il cui accordo esprime proprio una semantica comune nel definire la violenza sulle donne attraverso la mera fisicità che si concretizza con l'abuso e lo sfruttamento sessuale.

Il *focus* sul maggiore disaccordo invece si esprime attraverso il concetto *attivismo*, il cui residuo è di ≈ 1.14 . Pertanto, i campi concettuali monolingue connessi all'*attivismo* sono diversi tra loro e i vocabolari delle diverse lingue si presentano in modo differente per quel che concerne le iniziative, la pianificazione e le strategie dell'azione sociale. Difatti, a seguire, tra i valori residui più alti vi è proprio il concetto di *cooperazione* (≈ 1.08) che si lega perfettamente al disaccordo dell'*attivismo*.

Presenta inoltre valori residui alti il concetto *strumentalizzazione mediatica* (≈ 1.05), il cui scarso *consensus* permette di asserire che i concetti connessi alla manipolazione dei Social Media siano utilizzati in modo diverso da lingua a lingua e dunque non riferiti agli stessi temi.

Dall'analisi complessiva della *Varianza* si può sicuramente affermare l'esistenza di un considerevole consenso tra le lingue per quel che concerne i campi concettuali negativi. Infatti, la propensione alla semantica negativa accomuna in maggior misura i paesi indagati sul fenomeno della violenza contro le donne. Al contrario, i concetti positivi rappresentano il maggiore disaccordo e restituiscono quindi sistemi concettuali monolingue particolarmente diversi.

La sezione della tabella PANOVA finora analizzata ha permesso di studiare i concetti per i quali esiste un minore o maggiore accordo rispetto alle cinque lingue studiate. Tuttavia, se si intende indagare il massimo accordo o disaccordo tra le configurazioni linguistiche, in base al contributo che ciascuna lingua apporta alla rappresentazione media di ogni concetto, si osservano i valori residui delle configurazioni (*Tabella 4.15*). Dalla *Tabella 4.15* si evince che il valore residuo più basso è riferito all'italiano, che risulta essere la lingua maggiormente in accordo con tutte le altre sul piano concettuale. In opposto, la lingua ucraina presenta il valore residuo più alto, indicando quindi il più basso consenso con le altre lingue.

<i>Configurations</i>	<i>SS_{fit}</i>	<i>SS_{residual}</i>	<i>SS_{total}</i>
Italian	0	3.727155	20.70124
English	0	3.753493	20.65631
French	0	4.389883	19.57065
Spanish	0	4.199919	19.89473
Ukrainian	0	4.620595	19.17707
sum	0	20.691046	100.00000

Tabella 4.15. PANOVA: Analisi dei residui tra le configurazioni linguistiche

I valori residui tra le configurazioni linguistiche permettono inoltre di individuare una maggiore similarità tra l'italiano e l'inglese, la cui somiglianza del vocabolario era effettivamente emersa attraverso l'analisi esplorativa dei risultati (§ 4.3) i cui numerosi punti di connessione tra i due corpora riguardavano la violenza sessuale e l'abuso domestico.

4.5.3. Similarità tra le configurazioni e Consensus Permutation test

La similarità tra le configurazioni è inoltre visibile attraverso la matrice degli indici di similarità procustiana tra le configurazioni parziali (*Tabella 4.16*). In questo caso, più l'indice è alto e più esprime similarità tra le configurazioni linguistiche.

<i>Configurations</i>	<i>Italian</i>	<i>English</i>	<i>French</i>	<i>Spanish</i>	<i>Ukrainian</i>
<i>Italian</i>	1.0000000	0.7951419	0.7571396	0.7649120	0.7549434
<i>English</i>	0.7951419	1.0000000	0.7804084	0.7638709	0.7294187
<i>French</i>	0.7571396	0.7804084	1.0000000	0.7265819	0.7089476
<i>Spanish</i>	0.7649120	0.7638709	0.7265819	1.0000000	0.7511768
<i>Ukrainian</i>	0.7549434	0.7294187	0.7089476	0.7511768	1.0000000

Tabella 4.16. Indici di similarità procustiana tra le configurazioni parziali

Tale rappresentazione consente di constatare la similarità tra coppie di lingue e di proporre quindi la maggiore somiglianza per ciascuna configurazione linguistica. La *Tabella 4.16* indica infatti una maggiore somiglianza tra l'italiano e l'inglese, questo significa che i campi concettuali e quindi le relazioni tra i concetti di queste due lingue sono più simili tra loro rispetto a quelle delle altre lingue. Pertanto, le relazioni tra i concetti della lingua italiana sono più simili all'inglese e meno simili all'ucraino. Analoga prospettiva, si osserva per il sistema concettuale inglese che presenta una maggiore affinità con l'italiano e una minore similarità con l'ucraino. Il sistema concettuale francese invece risulta essere più simile a quello inglese e meno a quello ucraino, mentre quello spagnolo si presenta più affine a quello italiano e meno a quello ucraino. Per concludere, gli indici procustiani rilevano che le relazioni tra i concetti della lingua ucraina siano più concordi con il sistema concettuale italiano e meno concordi con il sistema concettuale francese.

La rappresentazione di similarità procustiana tra ogni duo di configurazione linguistiche ha confermato gli esiti di PANOVA riguardo all'analisi dei residui

delle lingue, dalla cui rappresentazione era emersa una maggiore similarità del vocabolario italiano ed una minore concordanza con il vocabolario della lingua ucraina.

In definitiva, l'accordo complessivo tra le cinque lingue indagate ha permesso di esprimere una maggiore similarità concettuale tra l'italiano e l'inglese, indicando attraverso lo studio del linguaggio, la presenza di un tessuto socio-culturale particolarmente simile sul tema della violenza contro le donne. Al contrario, si può affermare l'esistenza di una maggiore differenza nella percezione pubblica e nel modo di argomentare tale tema da parte dell'Ucraina. L'analisi procustiana generalizzata condotta sulle relazioni concettuali dei diversi vocabolari monolingue conferma alcune deduzioni che potevano essere tratte dallo studio esplorativo monolingue, che vedeva per l'appunto particolarmente affini la lingua italiana ed inglese mentre mostrava la lingua ucraina come quella più distante dall'analisi degli altri vocabolari.

Per completare il processo della comparazione *cross-language* tramite la *GPA* è stata eseguita la *performance* del *permutation test* (Walkeling *et al.*, 1992), con lo scopo di valutare la significatività del *consensus* tra le configurazioni. La proporzione di varianza spiegata dal consenso è stata espressa attraverso la statistica dell' R_c con $\alpha = 0.05$. Nel caso di fattispecie, mediante 999 permutazioni, il *test di significatività* del *consensus* ha restituito la varianza spiegata uguale al 79.31% e il *p-value* < 0.000 , il che indica che il *test* del *consensus* è statisticamente significativo.

4.6. I risultati dell'analisi *cross-language* in un breve confronto con alcune fonti ufficiali e parallele

I risultati della *GPA* confrontano l'opinione pubblica sul tema della violenza contro le donne in cinque diversi contesti linguistici e di conseguenza differenti contesti socio-culturali.

Risulta chiaramente difficile poter confrontare gli esiti ottenuti dal lavoro di ricerca con i risultati di eventuali fonti ufficiali e parallele che abbiano indagato nello stesso periodo di tempo e negli stessi contesti linguistici il tema oggetto di studio.

La percezione pubblica rimane comunque un aspetto di particolare interesse da parte delle fonti più diverse e ai fini di innumerevoli prospettive di ricerca. Ad

esempio, uno studio che ha coinvolto l'Unione Europea (UE) in materia di violenza di genere è l'indagine dell'Eurobarometro (428) condotta nel 2014. In particolare, tale ricerca ha studiato il livello di percezione della violenza contro le donne in svariati ambiti, rivelando l'esistenza di consapevolezza negli intervistati riguardo alle dinamiche del fenomeno e indicando l'esigenza di un maggior impegno da parte dell'UE. Gli esiti dell'Eurobarometro confermano i risultati ottenuti dal confronto procustiano dei diversi linguaggi, portando alla luce un'ampia conoscenza del fenomeno, seppure per certi versi stereotipata, e indicando l'insofferenza di un'inefficacia giuridica e di garantismo nazionale. Nonostante la ricerca dell'Eurobarometro sia del 2014 mentre i dati dell'analisi presentata in questa tesi sono del 2019, nell'opinione pubblica permane questo *status* di vuoto definitorio e d'azione che si riferisce alle attività di contrasto della violenza. Specificamente, l'*opinion survey* condotta dall'UE ha evidenziato che la violenza contro le donne è strettamente connessa alla disuguaglianza di genere, dimostrando come nella definizione dei concetti monolingue, la riduzione della disparità e l'educare alla parità identifichino le intenzioni pubbliche di voler ridurre la disuguaglianza, nel tentativo di limitare al contempo la violenza. Tra gli aspetti che emergono vi è la violenza sessuale, considerata la tipologia di abuso maggiormente esperita o denunciata dalle donne, secondo quanto riporta l'indagine dell'Eurobarometro e anche un'altra indagine proposta sempre nel contesto europeo da parte del FRA (2014). L'aspetto sessuale rappresenta quindi un punto di connessione estremamente forte tra le indagini di *survey* condotte in Europa e lo studio dei *tweets* presentato in questa tesi. Tra l'altro, la violenza sessuale è uno dei principali concetti che si presenta all'interno di ciascun *corpus* di *tweets* esaminato, per poi connotarsi di un significativo valore residuo in termini di *consensus* tra le lingue. Anche la violenza intrafamiliare rappresenta sia nell'indagine dell'Eurobarometro che del FRA una delle forme di abuso di cui si ha più conoscenza pubblica. Allo stesso modo, dall'analisi dei *tweets*, l'abuso domestico è risultato essere uno degli aspetti più dibattuti nei corpora delle diverse lingue.

Altri studi tuttavia si sono focalizzati sull'analisi dei contenuti estratti da Twitter, al fine di monitorare l'insorgenza del linguaggio d'odio *online* contro le donne. Quindi, confrontando i risultati emersi dall'analisi *cross-language* proposta in questa tesi con i risultati determinati da altre applicazioni metodologiche, si può tentare di determinare se esistano ulteriori punti di connessione con l'opinione pubblica specificatamente estratta dal *Web*. Il tema

della misoginia ad esempio, che risulta essere uno dei concetti che mette maggiormente d'accordo l'opinione pubblica studiata trova un allineamento con il contributo di ricerca presentato da Fulper *et al.* (2014). Gli autori infatti hanno esplorato negli Stati Uniti d'America l'*online hate speech* contro le donne a sfondo sessuale, indagando un campione di *tweets* estratti da gennaio a febbraio 2012. Il contributo di ricerca ha confrontato i contenuti del Social Media con le statistiche di stupro fornite dall'FBI, trovando una significativa associazione tra i tassi pro-capite di stupro e l'aumento del linguaggio misogino. Pertanto, lo studio di Fulper *et al.* e la proposta multilingue presentata in questo contesto mettono chiaramente in evidenza che nonostante i contesti esplorati siano diversi, così come i metodi adottati, il linguaggio utilizzato ha riferimenti chiaramente misogini e si focalizza sull'aspetto prettamente sessuale del rapporto tra uomo e donna. Il grado di conoscenza che emerge in tutti e cinque i contesti esplorati nella tesi può trovare conferme in un'altra analisi sulla violenza di genere che coinvolge però altre località geografiche, come gli Stati Uniti d'America, la Nigeria, le Filippine, l'India ed il Sud Africa. La ricerca di Purohit *et al.* (2015) infatti si è prefissata di analizzare i linguaggi della violenza selezionando tre temi principali, quali la violenza fisica, la violenza sessuale e le pratiche socio-culturali dannose. Gli esiti della ricerca hanno dimostrato che nonostante il livello di alfabetizzazione sulla violenza possa variare dal 99% negli Stati Uniti fino al 61.3% in Nigeria, esiste una dipendenza semantica che sorpassa le definizioni di stupro attraverso le culture. Questa ricerca in un certo qual modo conferma l'idea che nonostante si possano indagare registri linguistici differenti e tessuti socio-culturali diversi, il tema della violenza sulle donne è ormai divenuto un fenomeno su scala globale e alcuni argomenti che lo caratterizzano come la disuguaglianza di genere, l'abuso familiare o sessuale rappresentano elementi costanti ed imprescindibili di qualsiasi area di studio. Le opinioni estratte dai Social Media o l'analisi della percezione pubblica indagata per via diretta portano alla luce un'imprescindibile risultanza: le persone che si esprimano per mezzo di un questionario o davanti alla sezione predefinita di un Social Media possiedono idee contaminate da un'eccesso informativo che ha plasmato a livello globale la percezione pubblica. Vi possono essere sottili differenze nel dare più importanza ad un argomento o ad un altro ma fondamentalmente il tema della violenza sulle donne si muove, in qualsiasi cultura, lungo un *continuum* che considera come punti costanti la *misoginia*, la *disparità di genere*, la *violenza*

familiare e l'abuso sessuale, lamentando forme di *inefficacia risolutiva* da parte delle istituzioni.

4.7. L'Analisi dei concetti e del *Consensus* tra le lingue per una riflessione sulle Politiche Sociali antiviolenza

Lo studio *cross-language* proposto in questa tesi ha risposto all'interrogativo di indagare, attraverso l'analisi dei linguaggi, la percezione pubblica del fenomeno della violenza contro le donne, con l'obiettivo di evidenziare eventuali similarità e/o dissimilarità tra i paesi presi in esame.

Tale analisi è stata possibile grazie all'utilizzo di tecniche tipiche del *Text Mining* sui contenuti testuali estratti da Twitter durante sei mesi di monitoraggio, ovvero dal primo maggio al 31 ottobre 2019. Dunque, l'interpretazione dei risultati si limita a rappresentare la porzione di opinione pubblica che è stata raggiunta e studiata in questo contesto. Più propriamente, l'indagine ha riguardato l'applicazione di una strategia di analisi statistica su 4 milioni 62 mila e 725 *tweets* multilingue. I risultati hanno evidenziato che le radici strutturali della violenza maschile sulle donne appaiono come una questione molto dibattuta. Infatti, dall'analisi sono complessivamente emerse alcune dimensioni concettuali comuni per tutte le lingue. Principalmente, i concetti rilevati racchiudono una semantica sia positiva che negativa riferita al tema oggetto di studio. Sul versante positivo sono stati individuati concetti connessi alla giustizia, libertà, sicurezza, progresso strategico, attivismo, comunicazione sociale, cooperazione *etc.* Al contrario, sul versante negativo, sono stati evidenziati concetti basati sulla discriminazione, misoginia, abuso domestico, dominio maschile, maltrattamento, distruzione dell'identità, fallacia del sistema giuridico, garantismo nazionale, sfruttamento del corpo, *etc.*

Il modo in cui si relazionano i concetti delle cinque lingue studiate, nonché come l'opinione pubblica viene argomentata nei *tweets*, ha evidenziato una sostanziale omogeneità di rappresentazioni stereotipate, legate ai temi della *misoginia*, *discriminazione*, *abuso familiare* o *coniugale* e del *dominio maschile*. È risultata altresì convenzionale nelle argomentazioni il tema della *giustizia* e della *riduzione della disparità di genere*, dell'*educazione alla parità tra i sessi*, della *comunicazione dei mass media*.

In particolare, a livello globale, l'analisi statistica ha consentito di evidenziare similarità e/o dissimilarità tra i diversi registri linguistici, valutando in termini

numerici il livello effettivo di *consenso* tra le lingue. I risultati hanno quindi permesso di riscontrare un maggiore *consensus* interlinguistico in merito ai concetti della *misoginia*, della *discriminazione*, del *garantismo*, dell'*inefficacia giuridica* e della *sessualità*.

Tale consenso rappresenta i temi di massimo accordo nello studio dei registri linguistici di Italia, Regno Unito, Francia e Spagna ed al contempo anche dell'Ucraina, benché territorio dell'Europa dell'Est caratterizzato da un contesto sociale particolarmente diverso. Infatti, in Ucraina sussistono violenze e abusi strettamente connessi al genere soprattutto nell'ambito dei conflitti armati⁴¹ e permangono vuoti legislativi volti a definire e punire i crimini di guerra a sfondo sessuale. Rispetto alle altre lingue e culture indagate, lo studio della lingua ucraina ha consentito di identificare il contesto più peculiare, proprio perché la cultura della violenza sulle donne si è affermata in maniera molto più lenta. Pertanto, osservare l'esistenza di una somiglianza con gli altri paesi indica che nonostante le consapevolezze sociali, culturali e legislative abbiano tardato ad affermarsi e malgrado le lacune definitorie, la questione di genere non rappresenti più un tabù di cui non si possa parlare. Questo spiegherebbe anche perché in Ucraina l'anno scorso è stato riconosciuto come reato l'abuso familiare⁴².

Nello specifico, attraverso le tecniche statistiche messe a punto in questa indagine si è potuta mettere in evidenza una trama argomentativa della violenza sulle donne che in Italia è basata sul maltrattamento, la brutalità, la distruzione identitaria femminile, l'abuso domestico e l'amore violento. Nel contesto inglese si ripropongono gli stessi temi, sottolineando tuttavia concetti di dominio maschile, misoginia e sfruttamento del corpo. L'analisi dell'opinione pubblica francese invece ha restituito, oltre ai temi della violenza familiare e della supremazia maschile, un profilo culturale che tende ad evidenziare la criminalizzazione della violenza, il cui culmine è raggiunto attraverso la morte. Prevalentemente affini sono gli esiti emersi dall'analisi della collezione di dati in lingua sia spagnola che ucraina, ovvero concezioni culturali misogine e patriarcali e la violenza sessuale ne hanno caratterizzato i vocabolari.

⁴¹ <https://www.internazionale.it/reportage/marie-alix-detrie/2019/02/05/ucraina-violenze-sessuali>

⁴² <http://turadm.gov.ua/protydiia-domashnomu-nasyilstvu/item/5938-2-zhovtnia-mizhnarodnyiden-nenasyilstva>

Lo studio comparativo che ha confrontato le relazioni tra i concetti delle diverse lingue in unico dimensionale multilingue, attraverso i risultati del *consensus* ha permesso di far emergere inoltre il disappunto circa l'esecutivo della giurisdizione. In altre parole, sono emerse coesistenze concettuali comuni ed omogenee nel modo di argomentare il tema della *fallacia giuridica* e del *garantismo*. Sono infatti frequenti i riferimenti all'*ingiustizia*, *l'inasprimento delle pene*, all'*indulgenza*, al *rilascio degli assassini*, all'*ergastolo*, *etc.* Tale risultato ha portato alla luce un consenso pubblico che può essere interpretato come mancanza di fiducia nelle istituzioni e nel lavoro di tutela e difesa delle donne da parte degli apparati dello Stato. L'esito può essere prevalentemente giustificato nel contesto ucraino dove sussistono forti critiche circa la depenalizzazione dei crimini contro le donne, tuttavia, lo studio comparativo *cross-language* ha permesso di avanzare comunque delle riflessioni su un clima culturale globale che richiama le Politiche Sociali attraverso i poteri dello Stato. Dai risultati dell'analisi statistica, il consenso pubblico sembra essere propenso a concepire come fallimentari le politiche nazionali ed internazionali, in quanto l'impegno istituzionale e l'adeguatezza normativo-legislativa dell'azione di contrasto appaiono palesemente sottovalutati.

In realtà, invece, il più ragguardevole impegno delle Politiche Sociali è proprio rivolto alla strutturazione di un sistema di apparati adeguati al contrasto, che si definiscono per mezzo di strumenti sempre più consapevoli delle necessità di tutela e difesa della vittima e sempre più preparati alle strategie di punizione e rieducazione dell'aggressore. Esiste quindi un'inversione di tendenza in merito al consenso internazionale che emerge dai risultati dello studio, poichè sebbene gli interventi delle Politiche Sociali fino ad ora attuati indichino un impegno attivo nella disciplina sanzionatoria che ha cercato di inasprire le pene sul tema della violenza contro le donne, tale progresso non è stato collettivamente né accolto, né percepito.

Infatti, un primo passo importante di valenza giuridica per la tutela dei diritti delle donne è la Convenzione CEDAW⁴³ (*Convention on the Elimination of All Forms of Discrimination against Women*) adottata dall'Assemblea delle Nazioni Unite il 18 Dicembre 1979 per monitorare la condizione delle donne e promuoverne i diritti. Mentre, nel 1995 con l'adozione della *Piattaforma di Azione di Pechino* sono state individuate 12 aree in cui i governi sono stati

⁴³ <http://www.un.org/womenwatch/daw/cedaw/text/econvention.htm>

esortati a migliorare la condizione femminile. Tra i contributi di massima importanza rientra poi la *Convenzione di Istanbul* (2011) che ha chiarito definizioni e contesti della violenza sulle donne e della violenza domestica, delineando quali aspetti considerare come reato e indicandone le misure punitive. La *Convenzione di Istanbul* ha rappresentato il primo vero atto di consapevolezza internazionale e del chiaro intento di contrastare tale fenomeno, sebbene sul piano locale non tutti i paesi europei abbiano ancora pienamente accolto tale trattato. Ad esempio, tra i paesi coinvolti nella presente indagine, l'Italia (settembre 2013), la Spagna (aprile 2014) e la Francia (luglio 2014) lo hanno firmato e ratificato, al contrario il Regno Unito (giugno 2014) e l'Ucraina (novembre 2011)⁴⁴ hanno aderito all'accordo ma non l'hanno ancora ratificato. Proprio in attuazione della *Convenzione di Istanbul* ad esempio in Italia è stata approvata la legge sul Femminicidio (*L.119/2013*) che dal punto di vista penale ha inserito un inasprimento delle pene e delle misure cautelari, ponendo l'enfasi sull'aggravante dell'uxoricidio. La legge ha inoltre introdotto il reato di *stalking* e in flagranza di reato l'opportunità di misure restrittive.

Pertanto, se si confrontano alcuni dei provvedimenti delle politiche di sostegno attuate con i risultati dell'analisi statistica proposta in questo contesto, ne deriva una legittimazione culturale della violenza maschile sulle donne che ricalca una resilienza dei sistemi giuridici e legislativi sia locali che internazionali. La percezione collettiva dunque non richiama ulteriori operazioni di educazione alla *cultura della non violenza* e alla *parità di genere* ma rivendica maggiori e concrete azioni punitive. Dai risultati emersi si può quindi avanzare un'ipotesi di una mancanza di allineamento tra gli interventi delle Politiche Sociali e la percezione pubblica europea dell'impegno delle politiche stesse. Sono questi elementi che infatti mostrano la più rilevante somiglianza tra i paesi.

In definitiva, l'analisi statistica ha permesso di identificare una cultura sociale consapevole delle dinamiche fenomenologiche della violenza ma poco cosciente degli strumenti di contrasto.

Un ulteriore elemento problematico che si evince dall'osservazione analitica dei risultati è l'aspetto risolutivo del problema, che nella visione pubblica si concretizza solo attraverso un piano punitivo.

Tuttavia, è interessante osservare che i fattori di trasformazione e lotta sono molteplici e non possono essere semplificati alla stregua di un semplicistico atto

⁴⁴ <https://www.coe.int/it/web/conventions/full-list/-/conventions/treaty/210/signatures>

punitivo, poiché le radici della conversione sociale e culturale hanno una maggiore importanza dell'azione punitiva in sé; eppure questo elemento sembra sfuggire completamente dalla *voce* pubblica che è stata studiata in questo contesto. Le fasi delineate dalla strategia hanno quindi permesso di mettere in evidenza la prevalenza di un accordo interlinguistico particolarmente focalizzato su concetti che si connotano di un orientamento semantico negativo. Mentre, il livello di minore accordo tra le lingue e quindi di maggiore dissimilarità nel modo di argomentare il tema della violenza contro le donne, è riferito ai concetti della *cooperazione* e l'*attivismo*, i quali si caratterizzano per un orientamento positivo. Questi elementi di minore accordo tra le lingue indicano in maniera evidente un pericoloso radicamento di sospetto e scetticismo nelle politiche di sostegno. Quindi, questa opposta polarità percettiva nell'opinione pubblica che riguarda lo sviluppo di un progresso strategico partecipato e le azioni di trasformazione socio-culturale vengono argomentate in maniera più differente tra le lingue, evidenziando una maggiore diversificazione di linguaggio.

Pertanto, se i problemi del *garantismo* e dell'*inefficacia delle politiche* in ambito normativo possono essere trattati a livello interculturale, i temi della *cooperazione* e dell'*attivismo* comprendono Politiche Sociali mirate ai contesti locali.

In altre parole, attraverso l'analisi statistica dei dati si può avanzare l'idea che le Politiche Sociali svolgano un ruolo cruciale nel modello percettivo della violenza contro le donne, tuttavia sono gli attori di contrasto a livello nazionale che rappresentano i nodi fondamentali per definire la conversione del modello socio-culturale di ogni lingua. È chiaro che il quadro regolativo ha bisogno di una cooperazione che sia al contempo frammentata e omologata, a seconda degli ambiti specifici e delle peculiarità locali. La pertinenza delle operazioni deve essere adattata ad ogni singolo contesto, affinché le iniziative possano essere accolte e percepite senza resistenze di tipo ambientale o sociale, per un cambiamento in grado di equilibrare un *funzionalismo che non funziona più*.

In ambito di *attivismo* e *cooperazione*, le principali iniziative⁴⁵ delle Politiche di sostegno fino ad oggi realizzate in Europa a livello omogeneo si sono focalizzate sulla prevenzione, attraverso la sensibilizzazione dei settori privati e dei *mass media*, proponendo una cultura moltiplicativa dell'uguaglianza di

⁴⁵ Ad esempio: <https://www.unicef-irc.org/publications/pdf/digest6i.pdf>

genere. Il *fil rouge* delle pari opportunità si è concretizzato con lo scopo di ristabilire nell'immaginario collettivo, un'identità femminile paritaria e non proiettata alla mercificazione del corpo, dove fondamentalmente le agenzie educative e di socializzazione hanno adottato strategie a livello micro e macro per educare le nuove generazioni. Un ulteriore aspetto particolarmente dibattuto a livello europeo risulta la promozione partecipativa nel mercato del lavoro, tema accompagnato anche dall'equa distribuzione salariale. Sul piano della comunicazione sono state realizzate poi innumerevoli campagne sull'insidiosa trappola della violenza e delle relazioni tossiche, incentivando a denunciare il fenomeno e ponendo l'accento sull'esistenza di servizi sviluppati per accogliere e sostenere le vittime di violenza, in particolare quelle della violenza domestica. Le iniziative delle politiche di sostegno si sono altresì impegnate nella realizzazione di percorsi di formazione specializzati per l'assistenza medico-sanitaria, legale e degli operatori delle forze dell'ordine, così come la presa in carico a protezione e supporto delle vittime, nel tentativo di realizzare il cosiddetto *empowerment* a più livelli (*Ibidem*).

A livello locale, identificano il perno delle Politiche Sociali, la valorizzazione dell'autodeterminazione e la realizzazione di servizi specializzati che assicurino la gratuità dell'assistenza ed il principio di anonimato e riservatezza, elementi che necessitano di una importante frammentazione in base al contesto geografico.

Di conseguenza, le politiche concernenti lo sviluppo dei servizi e delle strutture specializzate a difesa delle donne identificano il punto focale delle iniziative che non possono essere considerate attraverso una prospettiva omologata e globale. Ad esempio, in Ucraina, questi aspetti devono essere rafforzati ed occorre un ingente potenziamento delle politiche a sostegno della causa e un attento monitoraggio delle forze pubbliche che rappresentano uno dei principali tasselli mancanti per la reazione attiva al contrasto della violenza⁴⁶. Infatti, tra i paesi indagati in questo lavoro di ricerca, l'Ucraina rappresenta il contesto più complesso in cui monitorare le politiche di sostegno, a causa di una tardiva affermazione sociale del fenomeno che determina un altrettanto inadeguato sviluppo di strumenti e ambienti adatti alla protezione e tutela delle donne. Il posticipato riconoscimento ha portato le strutture e i sistemi a risentire di una

⁴⁶ <https://www.amnesty.it/rapporti-annuali/rapporto-annuale-2017-2018/europa/ucraina/>

serie di aspetti e limiti a livello pratico che non incentivano l'azione di denuncia e la mobilità della donna da un nucleo familiare violento. Quindi, l'implementazione di una rete di politiche che cooperano in maniera attiva in questa direzione si presume possa rivelarsi particolarmente significativa in contesti meno avanzati, come per l'appunto quello ucraino. Al contrario, in domini socio-culturali post-moderni e avanzati, come Italia, Regno Unito, Francia e Spagna, questi modelli di sostegno e supporto sono stati ormai *assorbiti* e non ritenuti efficacemente adeguati.

In sintesi, dall'analisi statistica dei concetti e del *consensus* tra le lingue si può concludere che il tessuto socio-culturale europeo analizzato in questo contesto sia intriso di stereotipi di genere e di risentimento nei confronti dell'assetto legislativo-normativo.

Tuttavia, in base ai risultati dell'analisi statistica, l'aspetto che potrebbe indurre ad una sorta di perplessità e riflessione non è la persistenza radicalizzata del cliché di genere di cui si nutre la percezione pubblica o al contempo il livore espresso nei confronti dell'inefficacia del sistema giuridico, quanto piuttosto il credere che solo l'atto punitivo possa convertire l'attuale società in uno scenario che abbia davvero imparato a rispettare le donne.

I risultati della ricerca possono essere interpretati attraverso un'opinione pubblica che è alla costante ricerca di un colpevole, spesso rintracciabile nella fallacia delle istituzioni, non interrogandosi su questioni molto più profonde e complesse, legate invece al funzionalismo di una società che reitera in svariate forme la violenza sulle donne.

Le risposte che possono risultare mancanti all'interno della moltitudine così vasta di *tweets* analizzata sono in primo luogo una scarsa consapevolezza della necessità di un potenziamento dei servizi volti a fornire una prima ed immediata protezione per le donne, nonché l'istituzione di percorsi lavorativi mirati ad un supporto psicologico per donne che non riescono ad uscire da relazioni tossiche e violente. Particolare attenzione viene poi data alla violenza maschile, ai contesti familiari e educativi che ne fanno da incubatore.

Tra gli aspetti altresì sottovalutati dall'opinione pubblica e quindi non emersi dall'indagine statistica può annoverarsi un richiamo ai diritti umani. Nella fattispecie, si fa riferimento al diritto alla vita, al rispetto e alla tutela delle integrità ed identità personali, indipendentemente dalla sessualità degli individui. I processi di educazione, comunicazione di massa ed istituzione legislativa hanno continuato a propinare modelli che inglobano sempre e comunque, anche involontariamente, principi sessisti che rivendicano

inevitabilmente la questione di genere. Sussiste pertanto anche nelle politiche un vincolo connesso al genere e non all'universalità del diritto (Nussbaum, 2000). È questo, probabilmente l'aspetto mancante e forse più problematico che non emerge dall'analisi statistica dei linguaggi multilingue. L'assenza della dimensione dell'universalità dei diritti umani rappresenta forse il punto focale su cui le politiche sociali possono intervenire, perché la violenza in tutte le sue forme anche quelle più subdole e nascoste rende qualsiasi vita indegna di essere vissuta. Quindi, le iniziative di sostegno devono marcare un territorio nuovo, che non si basa sull'istruire riguardo alla parità di genere ma che mira alla definizione di modelli che ormai vacillano in un'epoca di costante frustrazione personale e alla spasmodica ricerca di colmare i vuoti della coscienza individuale. Si tratta essenzialmente di modelli che si focalizzano sul rispetto della vita umana.

4.8. Cenni conclusivi

Il capitolo conclusivo sui risultati della strategia di analisi *cross-language* ha permesso di valutare l'efficacia e l'utilità della metodologia statistica utilizzata per lo studio dei linguaggi e delle culture differenti sul tema della violenza contro le donne. In particolare, essa ha consentito di evidenziare eventuali similarità e/o dissimilarità tra i registri linguistici, supponendo l'esistenza di un binomio lingua-cultura. Lo studio ha focalizzato l'attenzione sulla tutela delle peculiarità linguistiche dei diversi corpora monolingue, rinunciando ad ogni sorta di traduzione, per tentare di rendere il più coerente possibile il confronto tra le lingue, senza alterarne la dimensione socio-culturale.

L'interpretazione dei risultati monolingue, ha mostrato che il quadro socio-culturale italiano e inglese sia particolarmente affine, in particolare, per quel che riguarda la violenza parentale e l'abuso sessuale. Gli esiti della Francia invece hanno messo in evidenza una tendenza sociale prevalentemente interessata all'assetto legislativo-normativo. L'analisi esplorativa dei dati spagnoli da una parte ha evidenziato l'autoderminazione femminile attraverso l'*empowerment* e la libertà, e dall'altra parte ha sottolineato la presenza di cultura patriarcale, discriminatoria e misogina. L'indagine della lingua ucraina ha permesso invece di rilevare il ruolo simbolico della donna, subordinata al

potere autoritario dell'uomo, l'esigenza di un sistema giudiziario e di un lavoro preventivo per il relativo contrasto.

Tuttavia, l'intento ultimo della ricerca non è stata l'analisi monolingue, quanto il confronto multilingue. Da un punto di vista statistico-testuale, il problema della comparazione e quindi della diversità linguistica è stato risolto attraverso la determinazione di un'unità di analisi di ordine concettuale: sono stati individuati gruppi di concetti monolingue, classificando in *cluster* le diverse *features*. Tuttavia, per ottenere i raggruppamenti concettuali si è voluto assicurare l'orientamento semantico delle parole, attraverso il calcolo della polarità semantica. Attraverso l'utilizzo combinato di due tecniche di analisi multivariata, quali l'*Analisi delle Corrispondenze* e la *Cluster Analysis*, ha preso forma la *Concepts Extraction*, da cui è stato possibile rilevare una chiave di lettura comune tra le lingue, cogliendo un dominio semantico simile. Più propriamente, nonostante le *features* estratte da ciascuna collezione di *tweets* positiva e negativa siano diverse da lingua a lingua, attraverso la loro classificazione sulle coordinate fattoriali si determinano gruppi concettuali che racchiudono spesso e volentieri una semantica comune. Di conseguenza, il fenomeno della violenza contro le donne può essere interpretato attraverso una percezione pubblica piuttosto simile. Ciononostante, per stabilire se i concetti individuati nei diversi linguaggi avessero relazioni simili tra loro, si è ritenuto interessante effettuare la comparazione tra i campi concettuali (sia positivi sia negativi) e studiarne quindi le relazioni. Tale proposito si è concretizzato con l'applicazione della *GPA* che ha consentito di confrontare simultaneamente in unico piano fattoriale le cinque configurazioni dei concetti. Complessivamente, le dimensioni multilingue emerse e confrontate si riferiscono principalmente a 12 raggruppamenti concettuali positivi e 15 raggruppamenti concettuali negativi. Sul versante positivo sono stati individuati concetti connessi alla giustizia, libertà, sicurezza, progresso strategico, attivismo, comunicazione sociale, *etc.* Al contrario, sul versante negativo, sono stati evidenziati concetti basati sulla discriminazione, misoginia, abuso domestico, dominio maschile, maltrattamento, distruzione dell'identità, *etc.*

I risultati della *GPA* attraverso la rappresentazione fattoriale e l'*Analisi della Varianza (PANOVA)* hanno permesso di cogliere le similarità o le dissimilarità tra le lingue e di individuare quindi quali concetti mettano più d'accordo l'opinione pubblica internazionale riguardo al tema preso in esame, suggerendo importanti riflessioni.

In definitiva, la proiezione fattoriale dei concetti in un comune spazio multidimensionale *cross-language* e l'*Analisi della Varianza* hanno evidenziato un maggiore accordo o *consensus* sui concetti della *misoginia*, della *discriminazione*, del *garantismo* e dell'*inefficacia giuridica*, mentre hanno espresso un minore accordo in relazione all'*attivismo*, alla *cooperazione* e alla *strumentalizzazione dei Social Media*. La propensione alla semantica negativa accomuna in maggior misura le lingue indagate sul fenomeno della violenza contro le donne. Al contrario, i concetti positivi restituiscono il maggiore disaccordo. È altresì emerso che l'italiano e l'inglese sono più simili tra loro nelle relazioni tra i concetti mentre la lingua ucraina è risultata la più dissimile nel confronto multilingue. La proposta di analisi *cross-language* adottata si è quindi rivelata particolarmente utile a definire un contesto europeo poco diversificato in termini di opinione pubblica, suggerendo una chiara e comune consapevolezza del problema ed evidenziando una diversa percezione sulla cooperazione e l'attivismo operati per la trasformazione socio-culturale.

Conclusioni

Lo studio multilingue comparativo proposto in questa tesi ha avuto l'obiettivo di confrontare in un unico e comune spazio dimensionale la percezione europea sul tema della violenza contro le donne. L'intento si è concretizzato comparando linguaggi e culture diverse, individuandone similarità o dissimilarità.

La proposta metodologica della presente ricerca ha definito una procedura di analisi *cross-language* in grado di studiare il fenomeno di interesse attraverso la combinazione di tecniche di analisi statistica. Più propriamente, è stata realizzata una strategia di analisi che mediante l'utilizzo di tecniche di *Text Mining* su dati reali ha consentito di dimostrare la sua efficacia strategica, indagando l'opinione pubblica di cinque paesi europei: Italia, Regno Unito, Francia, Spagna ed Ucraina. Nella fattispecie, si è inteso monitorare in un periodo di tempo che varia dal primo maggio al 31 Ottobre 2019, lo *speech online* realizzato sulla piattaforma Twitter in relazione al tema della violenza contro le donne. Il *database* documentale multilingue che è stato sottoposto ad analisi è costituito da 4 milioni 62 mila e 725 *tweets*. Tuttavia, la maggior parte dell'informazione estratta è in lingua spagnola, $\approx 40\%$, mentre solo il 2.5% della collezione dei *tweets cross-language* appartiene al registro linguistico ucraino. Una tale differenza nella quantità del dato estratto ha probabilmente influenzato l'emergere dell'effettiva percezione pubblica sul tema preso in esame in una prospettiva di analisi comparativa. Ciononostante, i risultati emersi sono riusciti a restituire una mappatura monolingue sul fenomeno, consentendone altresì il confronto multilingue. Una tale distorsione dipende probabilmente dalla popolarità e dall'utilizzo attivo del Social Media impiegato per l'estrazione dei dati testuali *online*. Malgrado tali criticità, di cui molte sono note in letteratura proprio sull'utilizzo dei dati provenienti dai Social Media, il dominio della comunicazione digitale ed in particolare Twitter hanno rappresentato, in questo contesto, la soluzione più adatta ai fini dell'analisi dei linguaggi multilingue. In particolare, tale piattaforma, grazie al potere *influenzativo* della *comunicazione partecipativa* si rivela infatti uno strumento di ricerca valido per numerosi temi di discussione che animano la rete, e rispetto ad altri Social Media si caratterizza per apertura ed orizzontalità. Inoltre, l'analisi della comunicazione sul *Web* permette di cogliere l'uso frequente di un linguaggio offensivo e discriminatorio verso specifiche categorie, come ad esempio l'*online hate speech* contro le donne.

Nella fattispecie, l'analisi statistica *cross-language* adottata in questo contesto per studiare l'opinione pubblica di Twitter sul tema della violenza contro le donne è risultata idonea ad indagare le diverse lingue e culture, producendo risultati rilevanti. Sono infatti emerse somiglianze e dissomiglianze tra i diversi registri linguistici considerando il binomio lingua-cultura, che mette in relazione le strutture grammaticali e semantiche di una lingua con la determinazione del pensiero collettivo.

I principali metodi ed approcci della letteratura *cross-language* del dato testuale solitamente basano il confronto solo fra due lingue per volta e attraverso tecniche di *machine traslation* o l'uso di dizionari bilingue. La presente tesi invece ha perseguito l'intento di mantenere inalterate le peculiarità proprie di ogni lingua, e senza ricorrere ad alcuna sorta di traduzione, dimostrando l'efficacia di un confronto che si basa sulla definizione di un'unità di analisi di ordine concettuale. In questo modo la variabilità tra le lingue che deriva dalla presenza di *features* monolingue differenti è stata notevolmente ridotta. Sono stati infatti individuati gruppi di concetti per ogni lingua, attraverso la classificazione delle singole *features* in *cluster*. Nello specifico, per definire tali classi, dapprima si è inteso determinare l'orientamento semantico delle parole finalizzato a raggruppare solo *features* che possiedono la stessa direzione semantica, così da contenere l'ambiguità che deriva dai diversi scenari emotivi degli utenti. L'assegnazione dei punteggi per il calcolo della polarità si è ottenuta utilizzando il *lexicon NRC (National Research Council Canada)* che copre la vastità linguistica richiesta in questo contesto, il quale è stato personalizzato sulla base dei risultati ottenuti dallo esplorativo preliminare dei cinque corpora monolingue. Tale miglioria ha avuto lo scopo di adattare il *lexicon* alla natura del fenomeno indagato, ossia ha permesso di allineare il dizionario delle parole polarizzate ai temi emersi dall'analisi esplorativa, consentendo quindi una più peculiare classificazione dei *tweets*.

Ogni collezione di dati monolingue è stata suddivisa in base alla polarità e attraverso l'utilizzo combinato di due tecniche di analisi multivariata: *Analisi delle Corrispondenze* e *Cluster Analysis*, si è realizzata la fase di *Concepts Extraction*. Ne è conseguita la rilevazione di classi di concetti solo positivi e classi di concetti solo negativi.

L'applicazione congiunta di queste due tecniche ha presentato il vantaggio di cogliere l'accezione interna ai vari corpora riscontrando possibili ed eventuali *modelli di senso* del contenuto di ogni testo, attraverso la posizione attribuita dalle coordinate fattoriali ai vari termini. Mantenendo poi inalterata la posizione

datagli dalle coordinate fattoriali, è seguita l'operazione di classificazione mista che partendo dal metodo gerarchico di *Ward*, si è consolidata con il metodo non gerarchico del *k-means*.

Dalla *Concepts Extraction* delle cinque lingue si è potuta riscontrare la presenza di raggruppamenti concettuali simili, ovvero la chiave di lettura ha evidenziato un dominio semantico simile a livello internazionale sul tema della violenza contro le donne. Successivamente, si è ritenuto interessante indagare se i concetti individuati nei diversi linguaggi avessero relazioni simili tra loro, ovvero se vi fosse una coesistenza tra i campi concettuali delle diverse lingue. L'intento in questo caso è stato quello di comprendere se i concetti siano stati concepiti allo egual modo in termini di nessi tematici.

Allo scopo di confrontare tali relazioni tra i concetti è stato quindi utilizzata l'*Analisi Procustiana Generalizzata (GPA)*. Tale metodo, tipicamente impiegato per confrontare i punteggi dei giudizi nelle analisi sensoriali, in questo contesto invece ha permesso di comparare simultaneamente in unico piano fattoriale le cinque configurazioni di concetti e dare così un risultato di *consensus* o accordo tra le lingue. La *performance* di tale tecnica per la comparazione tra le lingue è da prediligere soprattutto da un punto di vista metrico, in quanto consente di operare sui dati originari delle matrici senza la necessità di sottoporre i dati a trasformazioni di algebra lineare. Inoltre, la prospettiva computazionale è facilitata dalla procedura iterativa che riduce tempo e complessità analitica.

I risultati emersi dall'applicazione della *GPA* che si concretizzano in una rappresentazione fattoriale e nell'*Analisi della Varianza (PANOVA)* hanno permesso di osservare le similarità e/o dissimilarità tra le lingue, identificando quali concetti hanno rappresentato il maggiore consenso nell'opinione pubblica internazionale riguardo al tema preso in esame.

Nello specifico, la proiezione fattoriale dei concetti in un comune spazio multidimensionale *cross-language* e l'*Analisi della Varianza* hanno evidenziato un maggiore accordo sui concetti della *misoginia*, della *discriminazione*, del *garantismo* e dell'*inefficacia giuridica*, mentre hanno espresso un minore consenso in relazione all'*attivismo*, alla *cooperazione* e alla *strumentalizzazione dei Social Media*. Ne consegue quindi, che nonostante l'omogeneità emersa dai raggruppamenti concettuali simili, sussiste una percezione pubblica che si diversifica nell'argomentare i temi dell'attivismo, della manipolazione mediatica, della cooperazione e partecipazione pubblica. La propensione alla semantica negativa accomuna in maggior misura le lingue indagate sul

fenomeno della violenza contro le donne, suggerendo una comune conoscenza del problema e dei rischi.

La condivisione delle concezioni patriarcali e misogine e la visione partecipativa della violenza sessuale e dell'abuso domestico rappresentano la chiave di lettura di una società europea che è culturalmente cosciente dell'assoggettamento della donna nella sua dimensione pubblica e privata. Tale aspetto non sorprende nell'analisi di una società democratica consolidata e post-moderna occidentale, mentre restituisce un'interessante prospettiva sul grado di avanzamento socio-culturale dell'Ucraina, dove sussistono gravi deficienze nelle politiche di contrasto, tutela e protezione delle vittime. I risultati dell'analisi statistica dei concetti e del *consensus* hanno condotto ad importanti riflessioni sul tema delle Politiche Sociali, poiché il tessuto socio-culturale europeo analizzato in questo contesto appare intriso di stereotipi di genere e di risentimento nei confronti dell'assetto legislativo-normativo. Si richiamano infatti gli apparati dello Stato, facendo emergere il proprio disappunto sull'esecutivo della giurisdizione, intesa come unica strategia per contrastare il fenomeno.

In altre parole, lo studio della percezione pubblica *cross-language* ha evidenziato un maggiore accordo nel modo di argomentare i temi del garantismo e dell'inefficacia giuridica mentre ha mostrato maggiore dissimilarità in riferimento ai concetti della cooperazione e l'attivismo. Questi elementi di minore accordo tra le lingue indicano in maniera evidente un pericoloso radicamento di sospetto e scetticismo nelle politiche di sostegno. Pertanto, se i problemi del garantismo e della fallacia giuridica possono essere trattati attraverso politiche di sostegno omologate a livello interculturale, i temi della cooperazione e dell'attivismo comprendono attività di Politiche sociali locali e frammentate.

I risultati della ricerca possono essere interpretati attraverso un'opinione pubblica che colpevolizza le istituzioni, non interrogandosi su questioni più profonde e complesse, legate invece al funzionalismo di una società che reitera in svariate forme la violenza sulle donne. Quindi, in base ai risultati emersi dalla strategia di analisi statistica proposta in questo elaborato, le risposte che possono risultare mancanti nell'opinione pubblica sono la scarsa consapevolezza delle necessità di potenziamento dei servizi per la tutela delle donne e per i processi di autodeterminazione che favoriscono l'*empowerment* e l'allontanamento da relazioni tossiche e violente.

Tra gli aspetti altresì sottovalutati dall'opinione pubblica si può cogliere la mancanza di un chiaro riferimento ai diritti umani e nella fattispecie, il diritto alla vita nel rispetto e tutela delle integrità ed identità personali, indipendentemente dalla sessualità degli individui. L'assenza della dimensione dell'universalità dei diritti umani rappresenta forse il punto focale su cui le politiche sociali possono intervenire, esplorando un territorio nuovo, che non si basa sull'istruire riguardo alla parità di genere ma che mira alla definizione di modelli vacillanti in un'epoca di costante frustrazione personale e alla ricerca di colmare i vuoti della coscienza individuale, portando solo così ad una radicale conversione socio-culturale.

Riferimenti bibliografici

- Abdi, H., Williams, L.J., Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets, *WIREs Computational Statistics*, Wiley Periodicals, Inc.
- Abdul-Mageed, M., Diab, M. (2012). Toward building a large-scale Arabic sentiment lexicon, in *Proceedings of the 6th International Global WordNet Conference*.
- Abdulcadir, J., et al. (2011). Cura delle donne con mutilazione- taglio genitale femminile, *Swiss Medical Weekly*.
- Accademia della Crusca (2013). Femminicidio: i perché di una parola. Disponibile all'indirizzo: <http://www.accademiadellacrusca.it/it/lingua-italiana/consulenza-linguistica/domande-risposte/femminicidio-perch-parola> (Ultima consultazione: 12 Maggio 2018).
- Al-Marghilani, A., Zedan, H., Ayesh, A. (2008). Text Mining Based on the Self-Organizing Map Method for Arabic-English Documents, In *Proceedings of the Nineteenth Midwest Artificial Intelligence and Cognitive Science Conference (MAICS 2008)*.
- Alcalde, M.J. et al. (2014). Generalized procrustes analysis (GPA) as a tool to discriminate among sheep breeds, *Archiv Tierzucht* 57: 28, 1-10.
- Amnesty International (2006). Danni collaterali. La violenza contro le donne nei conflitti armati, EGA Editore.
- Anderberg, M.R. (1973). Cluster Analysis for Applications, *Academic Press, New York*.
- Angelari, M. (1997). Gli statuti del crimine di odio: uno strumento promettente per combattere la violenza contro le donne, In Maschke, Karen J. Pornografia, lavoro sessuale e incitamento all'odio, *New York, Taylor e Francis*.
- Anoop, V.S., Asharaf, S., Deepak, P. (2011). Topic Modeling for Unsupervised Concept Extraction and Document Ranking, *The International Symposium on Intelligent Systems Technologies and Applications, ISTA 2017, pp.123-135*.
- AP (2000). Bangladesh combats an acid onslaught against women, *CNN*.
- Apel, K.O., Gadamer, H.G., Habermas, J. (1992). Ermeneutica e critica dell'ideologia (2^a edizione). Queriniana Editore.
- Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist ?, *Psychometrika*, 56, 567-587.
- Arimatsu, L., Chinkin, C. (2017). International Women's Day 2017: Moving forward in a time of uncertainty and upheaval. Disponibile all'indirizzo: <http://blogs.lse.ac.uk/wps/2017/03/07/international-womens-day-2017/> (Ultima consultazione: 15 Maggio 2018).
- Arnold, G.M., Williams, A.A. (1986). The Use of Generalized Procrustes Techniques in Sensory Analysis. In: Piggot, J.R. (ed.), *Statistical Procedures in Food Research*, pp. 233-253. Elsevier, London.

- Artetxe, M., Labaka, G., Agirre, E. (2016). Learning principled bilingual mappings of word embeddings while preserving monolingual invariance, *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2289–2294.
- Artetxe, M., Labaka, G., Agirre, E. (2017). Learning bilingual word embeddings with (almost) no bilingual data, *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 451–462.
- Artetxe, M., Labaka, G., Agirre, E. (2018). A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings, *In Proceedings of ACL 2018*.
- Assemblea Generale delle Nazioni Unite (1948). Universal Declaration of Human Rights. Disponibile all'indirizzo: un.org/en (Ultima consultazione: aprile 2018).
- Bader, B.W., Chew, P.A. (2008). Enhancing Multilingual Latent Semantic Analysis with Term Alignment Information, *In COLING 2008, 22nd International Conference on Computational Linguistics, Proceedings of the Conference, Manchester, UK*.
- Bahl, T., Syed, M. H. (2003). *Encyclopaedia of Muslim world*. New Delhi, Anmol Publications.
- Balahur, A., Turchi, M. (2014). Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis, *Computer Speech & Language*, 28 (1):56–75.
- Balbi, S., Misuraca, M., (2005). Pesi e Metriche nell'Analisi dei Dati Testuali, *Quaderni di Statistica, Liguori*, 7, 55–68.
- Bandyopadhyay, M., Mahmuda, R. K. (2003). Loss of face: violence against women in South Asia in Lenore Manderson, Linda Rae Bennett (eds) *Violence Against Women in Asian Societies*, Routledge Curzon.
- Banea, C., Mihalcea, R., Wiebe, J., Hassan, S. (2008). Multilingual subjectivity analysis using machine translation, *In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 127–135.
- Barbaranelli, C. (2007). *Analisi dei dati. Tecniche Multivariate per la ricerca psicologica e sociale*, Seconda Edizione, Strumenti e Metodi per le Scienze Sociali, Edizioni Universitarie di Lettere Economia e Diritto.
- Barbera, P. (2013). streamR package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/streamR/streamR.pdf>.
- Barnes, J. et al. (2018). Bilingual Sentiment Embeddings: Joint Projection of Sentiment Across Language, *In Conference: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Vol. 1: Long Papers, Melbourne*.
- Barone, A.V.M. (2016). Towards cross-lingual distributed representations without parallel text trained with adversarial autoencoders, *Proceedings of the 1st Workshop on Representation Learning for NLP*, pp. 121–126.

- Basso Fossali, P. (2008). *Vissuti di significazione. Temi per una semiotica viva*. Edizioni ETS.
- Baumgardner, J. (2011). *F 'em!: Goo Goo, Gaga, and Some Thoughts on Balls*, Seal Press, pp. 250.
- Beasley, C., Brook, H., Holmes, M. (2012). *Heterosexuality in Theory and Practice*, New-York, Taylor and Francis.
- Beh, E.J., Lombardo, R. (2014). *Correspondence Analysis: Theory, Practice and New Strategies*, John Wiley & Sons, Chicester.
- Beinborn, L., Choenni, R. (2019). *Semantic Drift in Multilingual Representations*, *In Computation and Language (cs.CL)*, *arXiv:1904.10820v1*.
- Bellassai, S. (2011). *L'invenzione della virilità. Politica e immaginario maschile nell'Italia contemporanea*. Carocci Editore.
- Benoit *et al.*, (2018). *quanteda: An R package for the quantitative analysis of textual data*, *Journal of Open Source Software*, 3(30), 774.
- Benzécri J. P. (1963). *Cours de linguistique mathématique*, Rennes: Université de Rennes, Rennes.
- Benzécri J. P. (1973). *L'Analyse des données (2 tomes)*, Dunod. Paris.
- Benzécri J. P. *et al.* (1981), *Pratique de l'analyse des données - Linguistique et lexicologie*, Dunod, Paris.
- Benzécri J.P. (1982). *Histoire et préhistoire de l'analyse des données*, Dunod, Paris.
- Berelson, B. (1952). *Content Analysis in Communication Research*, *The Free Press of Glencoe, New York*.
- Berg, B. L. (1998). *Qualitative research methods for the social sciences*, Boston, Allyn and Bacon.
- Berg, R. C. *et al.* (2014). *Effects of female genital cutting on physical health outcomes: a systematic review and meta-analysis*, *BMJ Open*. 4 (11).
- Bhambra, G., Shilliam, R (2009). *Silencing Human Rights: Critical Engagements with a Contested Project*, London: Palgrave MacMillan.
- Bianchi, C. (2003). *Pragmatica del linguaggio*. Laterza, Bari.
- Bianco, F. (2002). *Introduzione all'ermeneutica*, Laterza, Roma.
- Blei, D.M., Ng, A.Y., Jordan, M.I (2003). *Latent Dirichlet Allocation*, *Journal of Machine Learning Research*. 3 (4-5): pp. 993-1022.
- Boas, F. (1905). *The horizontal plane of the skull and the general problem of the comparison of variable forms*. *Science*. 21, 862-863.
- Bobbioni, M., P. (2012). *L'abito sessuato*, disponibile all'indirizzo: <http://www.psychiatryonline.it/node/2367>

- Bolasco S. (1990). Sur différentes stratégies dans une analyse des formes textuelles: une expérimentation à partir de données d'enquête, *In Bécue M., Lebart L., Rajadell N. editors, Actes des Premières journées JADT, UPC, Barcelone, (1992), 69-88.*
- Bolasco, S. (1986). Per una teoria sulla costruzione e analisi delle matrici a tre modi, *in Atti della XXXIII Riunione Scientifiche SIS, Bari 28-30 Aprile, Vol. 1, pp.183-95.*
- Bolasco, S. (1997). L'analisi informatica dei testi, *in Ricolfi, L. (a cura di) La ricerca qualitativa. Nis, Roma.*
- Bolasco, S. (1999). Analisi multidimensionale dei dati. Metodi, strategie e criteri di applicazione, Carocci Editore, Roma (2^a ed., 2004).
- Bolasco, S. (2004). L'analisi statistica dei dati testuali: intrecci problematici e prospettive, *in E. Aureli Cutillo e S. Bolasco (a cura di), Applicazioni di analisi statistica dei dati testuali, Casa Editrice La Sapienza, Roma, pp. 9-26.*
- Bolasco, S. (2005). Statistica testuale e text mining: alcuni paradigmi applicativi, *Quaderni di Statistica, Vol.7.*
- Bolasco, S., Bisceglia, B., Baiocchi, F. (2004). Estrazione automatica di informazione dai testi, *Mondo Digitale, vol. 3, n. 1, pp. 27-43.*
- Bookstein, F. L. (1991). Morphometric tools for landmark data. *Cambridge: New York Press.*
- Bookstein; F.L., Mitteroecker, P. (2014). Comparing Covariance Matrices by Relative Eigenanalysis, with Applications to Organismal Biology, *Evol Biol 41:336-350, Springer Science+Business Media New York, 2013.*
- Borg, I., Groenen, P. (1997). Modern Multidimensional Scaling. Theory and Applications, *Springer-Verlag, New York.*
- Borrelli, D., Serpieri, S., Taglietti, D., Trezza, D. (2018). The Promises of Sociological Degrees: A Lexical Correspondence Analysis of Masters Syllabi, *in Lauro, C., Amaturò, E., Grassia, M.G., Aragona, B., Marino, M., Data Science and Social Research. Epistemology, Methods, Technology and Applications, Studies in Classification, Data Analysis and, Knowledge Organization, Springer.*
- Bose, S. (2016). RSentiment package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/RSentiment/RSentiment.pdf>.
- Bott S., Morrison A., Ellsberg M. (2005). *Preventing and responding to gender-based violence in middle and low-income countries a global review and analysis*, World Bank.
- Bouaziz, A., Dartigues-Pallez, C., da Costa Pereira, C., Precioso, F., Lloret, P. (2014). Short Text Classification Using Semantic Random Forest, *International Conference on Data Warehousing and Knowledge Discovery, pp. 288-299.*
- Bougeard, S., Abdi, H., Saporta, G., Niang, N. (2018). Clusterwise analysis for multiblock component methods, *Advances in Data Analysis and Classification 2018, Vol. 12, Issue 2, pp. 285-313.*

- Bougeard, S., Cardinal, M. (2014). Multiblock modeling for complex preference study. Application to European preferences for smoked salmon, *Food Qual Prefer* 32:56–64.
- Boumans, J. W., Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars, *Digital Journalism*, 4(1), 8–23.
- Bourdieu, P. (1977). *Outline of a Theory of Practice*, Cambridge University Press.
- Bourdieu, P. (1980). *Le Sens pratique*, Edizione Minuit, pp. 246-247, 358, 361-409, 441-461.
- Bourdieu, P. (1989). Social Space and Symbolic Power. *Sociological Theory*, Vol. 7, No 1: 14- 25.
- Bourdieu, P. (1994). Social Space and Symbolic Power, in Pierre Bourdieu. *Other Words*, pp. 122-139, Cambridge: Polity Press.
- Bourdieu, P. (1998). *La domination masculine*. Édition du Seuil.
- Bourdieu, P. (2001). *Television, European Review*, Vol. 9, No. 3: 245–256.
- Bourdieu, P. (2002). *Language and Symbolic Power*, Cambridge: Polity Press.
- Bourdieu, P., Passeron, J. (1977). *Reproduction in Education, Society and Culture*, London: Sage.
- Bourdieu, P., Wacquant, L. (1992). *An Invitation to Reflexive Sociology*, Chicago, University of Chicago Press.
- Bourgois, P. (2004). The Everyday Violence of Gang Rape, in Scheper-Hughes, N. and Bourgois, P. (2004) editors, *Violence in War and Peace: An Anthology*, Malden, MA, Blackwell, pp.343-348.
- Bourgois, P., Prince, P., Moss, A. (2004). The Everyday Violence of Hepatitis C among Young Women Who Inject Drugs, in *San Francisco, Human Organization*, Vol. 63, No. 3: 253- 263.
- Braidotti, R. (1991). *Patterns of Dissonance*, Traduzione italiana, *Dissonanze. Le donne e la filosofia contemporanea*, La Tartaruga Editore, 1994.
- Braidotti, R. (1994). *The Nomadic Subject*, Traduzione italiana, *Soggetto nomade. Femminismo e crisi della modernità*, Donzelli, Roma 1995.
- Brownmiller, S. (1935). *Against Our Will: Men, Women, and Rape*, New York, Ballantine Books, 1975.
- Brzezińska, J. (2018). *Singular Value Decomposition Approaches in A Correspondence Analysis with The Use of R*, *Folia Oeconomica Stetinensia* 18(2):178-189.
- Buchili, E. (2006). *L'amore fatale*. Baldini Castoldi Dalai Editori.
- Burgess, R., Campbell, C. (2016). Creating social policy to support women's agency in coercive settings: A case study from Uganda, *Global public health*, 11(1-2), 48.

- Burnap, P., Williams, M.L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making, *Policy & Internet* 7(2):223–242.
- Buttigieg, P.L., Ramette, A. (2014). A Guide to Statistical Analysis in Microbial Ecology: a community-focused, living review of multivariate data analyses, *FEMS Microbiol Ecol.* 90: 543–550.
- Campbell, J. (1958). *L'eroe dai mille volti*. Feltrinelli Editore.
- Cantarella, E. (1996). *Passato prossimo. Donne romane da Tacita a Sulpicia*, Feltrinelli, Milano.
- Capecchi, S. (2016). *Identità di genere e media*, Carocci Editore, Roma.
- Cappa, C. et al. (2013). *Mutilazione - taglio genitale femminile: una panoramica statistica e l'esplorazione delle dinamiche del cambiamento*, UNICEF.
- Caretti, V., Craparo, C. (2004). Prefazione all'edizione italiana del libro di W.R. Cupach e B.H. Spitzberg, *The dark side of relationship pursuit. From attraction to obsession and stalking*, Lawrence Erlbaum Associates, Inc., New Jersey-London.
- Carr, L. (1884). The Social and Political Position of Women Among the Huron-Iroquois Tribes, *Report of the Peabody Museum of American Archaeology*, p. 223.
- CEDAW (1979). Convention on the Elimination of All Forms of Discrimination against Women. Disponibile all'indirizzo: <https://www.un.org/womenwatch/daw/cedaw/text/econvention.htm> (Ultima consultazione: aprile 2018).
- Centers for Disease Control and Prevention (2006). *Intimate partner violence*, National Center for Injury Prevention and Control.
- Ceron, A., Curini, L., Iacus, S.M. (2014). *Social media e Sentiment analysis. Milano: Springer per l'Innovazione*.
- Chang, J. C., et al. (2004). Health care interventions for intimate partner violence: What women want, *Women's Health Issues*, 15, 21-30.
- Chen, X., Luo, L., Wang, H., Wang, W., Gao, Y. (2017). Analysis and Research on Cross Language Topic Discovery in Chinese and English, *Advanced Engineering Sciences*, 49(2):100-106.
- Chew, P.A. (2009), *Multilingual Text Mining*, Encyclopedia of Data Warehousing and Mining, Second Edition.
- Chiari, I. (2004). *Informatica e lingue naturali. Teorie e applicazioni computazionali per la ricerca sulle lingue*, Aracne.
- Chinkin, C. (2014). *Research impact: making a difference, Making the world better, safer place for women*, Collection Law, in Library of London School of Economics.
- Ciotti, F. (2017). What's in a Topic Model? Critica teorica di un metodo computazionale per l'analisi del testo, *Testo e Senso, Digital Humanities*, n. 18.

- Cipriani, R., Bolasco, S. (1995). *Ricerca qualitativa e computer. Teoria, metodi e applicazioni*, Franco Angeli, Milano.
- Cirone, J., Li, W., Page, T., Phipps, A., Waye, F. (2017). *Report It. Stop It: how can reports of sexual harassment and sexual violence change the culture*, disponibile all'indirizzo: <http://www.lse.ac.uk/website-archive/newsAndMedia/videoAndAudio/channels/publicLecturesAndEvents/player.aspx?id=3906>
- Cixous, H. (1975). *Le Rire de la Méduse et autres ironies*, Galilée Édition, 2010.
- Clement, J. (2019). Twitter: number of monthly active users 2010-2019, in Statista, Disponibile all'indirizzo: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- Cochrane, K. (2013a). *The fourth wave of feminism: meet the rebel women*, The Guardian.
- Cochrane, K. (2013b). *All the rebel women: the rise of the fourth wave of feminism*, London, Guardian Books.
- Comitato dei ministri del Consiglio d'Europa (2002). *Raccomandazioni del Comitato dei Ministri agli Stati Membri a protezione delle donne contro la violenza*.
- Commandeur, J.J.F. (1991). *Matching configurations*. DSWO press, Leiden University.
- Commissione Europea (2015). *Strategic engagement for gender equality 2016-2019*, Pubblicazione dell'Ufficio dell'Unione Europea, disponibile all'indirizzo: https://ec.europa.eu/anti-trafficking/sites/antitrafficking/files/strategic_engagement_for_gender_equality_en.pdf
- Commissione Europea, COM (2010), 2020 definitivo. *Comunicazione della Commissione Europea 2020, Una strategia per una crescita intelligente, sostenibile e inclusiva*; disponibile: <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52010DC2020>.
- Conneau, A., Lample, G., Ranzato, M., Denoyer, L., Jégou, H. (2018). Word Translation Without Parallel Data, in *Proceedings of ICLR 2018*.
- Connell, R. W. (1996). *Maschilità. Identità e trasformazioni del maschio occidentale* Feltrinelli.
- Connell, R. W. (2006). *Questioni di genere. Edizioni, Il Mulino ISBN 88-15-10538-7*.
- Connolly, P., Healy, J. (2004). Symbolic Violence, Locality and Social Class: the educational and Career Aspirations of 10-12 year old boys in Belfast. *Pedagogy, Culture and Society, Vol.12, No.1: 15-34*.
- Consiglio d'Europa (2011). *Convenzione del Consiglio d'Europa sulla prevenzione e lotta contro la violenza nei confronti delle donne e la violenza domestica*. Disponibile all'indirizzo: https://www.gazzettaufficiale.it/do/atto/serie_generale/caricaPdf?cdimg=13A057890000010110002&dgu=2013-07-02&art.dataPubblicazioneGazzetta=2013-07-02&art.codiceRedazionale=13A05789&art.num=1&art.tiposerie=SG (Ultima consultazione: Maggio 2018).

- Corbetta, P. (2014). *Metodologia e tecniche della ricerca sociale* (2^a ed), Il Mulino. Bologna.
- Corradi, C. (2008). *I modelli sociali della violenza contro le donne. Rileggere la violenza nella modernità*, Franco Angeli Editore.
- Council of Europe (2011). *Council of Europe Convention on preventing and combating violence against women and domestic violence*.
- Cresti E., Panunzi, A. (2013). *Introduzione ai corpora dell'italiano*, Il Mulino, Bologna.
- Crone, S. F., Lessmann, S., Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing, *European Journal of Operational Research*, 173(3), 781–800.
- Crosilla, F. (1983). A criterion matrix for the second order design of control networks. *Bulletin Geodesique*. 57.
- Curci, P., Galeazzi, G.M., Secchi, C. (2003). *La sindrome delle molestie assillanti (stalking)*. Torino, Bollati Boringhieri Casa Editrice.
- Dandini, S. (2014). *Ferite a morte, E se le vittime potessero parlare?*, Brossura Edizioni.
- Danna, D. (2007). *Ginocidio. La violenza contro le donne nell'era globale*. Elèuthera Editrice.
- De Beauvoir, S. (1989). *The Second Sex*, Versione Inglese, *Vintage Book*, p.105.
- De Beauvoir, S. (1984). *Il secondo sesso*, Il Saggiatore.
- De Castella, T. (2013). *How many acid attacks are there?*, BBC News.
- De Lauretis, T. (1987). *Technologies Of Gender. Essays on Theory, Film, and Fiction*, Indiana University Press, Bloomington (Traduzione italiana parziale, 1996).
- De Lauretis, T. (1991). Queer Theory: Lesbian and Gay Sexualities, in *Differences, A Journal of Feminist Cultural Studies*, n. 2, vol. 3.
- De Lauretis, T. (1996). *Sui generis. Scritti di teoria femminista*, Feltrinelli, Milano.
- De Lauretis, T. (1984). *Alice doesn't: feminism, semiotics, cinema*. Indiana University Press.
- De Lauretis, T. (1990). *Soggetti Eccentrici*, Feltrinelli Editore, 1999.
- De Lauretis, T. (1994). *The Practice of Love: Lesbian Sexuality and Perverse Desire*, Indiana University Press, Bloomington.
- De Queiroz, G., Hvitfeldt, E., Keyes, O., Misra, K., Mastny, T., Erickson, J., Robinson, D., Silge, J. (2017). Tidytext package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/tidytext/tidytext.pdf>.
- De Roover, K., Ceulemans, C., Timmerman, M. (2012). Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data, *Psychol Methods* 17:100–119.

- De Roover, K., Ceulemans, E., Timmerman, M.E. (2012). How to perform multiblock component analysis in practice, *Behav Res* 44:41–56.
- De Saussure, F. (1916). *Cours de linguistique générale*, Lausanne-Paris, Payot.
- De Simone, G. (2013). *Il delitto di atti persecutori*, Aracne, Roma.
- De Soete, G., Carroll, J. D. (1994). K-means clustering in a low-dimensional euclidean space, in E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, & B. Burtschy (Eds.), *New approaches in classification and data analysis* (pp. 212–219). Berlin: Springer.
- De Stefano, D., Santelli, F. (2019). Combining Sentiment Analysis and Social Network Analysis to Explore Twitter Opinion Spreading, In *Proceedings 28th International Conference on Computer Communication and Networks (ICCCN)*.
- Deerwester, S, Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R. (1990). Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science*. 41 (6): 391–40.
- Deleuze, G., Guattari, F. (1992). *Tausend Plateaus. Kapitalismus und Schizophrenie*.
- Di Donato, L. (2013). Guida all'estrazione di dati dai Social Network. Disponibile all'indirizzo: <http://www.dima.unige.it/~riccomag/CorsoISTAT/Twitter/article-130719135843-phpapp01.pdf>.
- Diamond, D. (2009). *The fourth wave of feminism: psychoanalytic perspectives*, pp. 213–223.
- Dijksterhuis, G., Punter, P. (1990). Interpreting generalized procrustes analysis, Analysis of Variance tables, *Food Quality and Preference*, 2, 255–265.
- Diodorus Siculus (1935). *Library of History, Books III-IV*. Traduzione a cura di Oldfather, C.H., *Loeb Classical Library, Vol. 303 e 340*. Cambridge, MA, Harvard University Press, London, William Heinemann Ltd., 1935.
- Dobash, R., E., Dobash, R., P. (2015). *When Men Murer Women, Interpersonal Violence*, Oxford Press.
- Donati, P. (1991). *Teoria relazionale della società*, Franco Angeli Editore.
- Donne delle Nazioni Unite (2016). *Comunicato stampa sul rapporto sulla violenza informatica*.
- Douglas, M.A., Endler, J.A. (1982). Quantitative matrix comparisons in ecological and evolutionary investigations, *Journal of Theoretical Biology, Volume 99, Issue 4, 21*, pp. 777-795.
- Douglas, M.E., Minckley, W.L., De Marais, B.D. (1999). Did vicariance mold phenotypes of western North American fishes? Evidence from Gila river Cyprinids, *Evolution* 53: 238–246.
- Dryden, I. L. (2018). Shapes package. R. *Foundation for Statistical Computing, Vienna, Austria. Contributed package. Version 1.2.4*. URL <http://www.R-project.org>.

- Dryden, I.L., Mardia, K.V. (2016). Statistical shape analysis. Wiley series in probability and statistics, (1st ed. 1998).
- Dumais *et al.*, 1988. Using latent semantic analysis to improve access to textual information, *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 281-285.
- Dumas, A. (1872). *L'Homme-femme : réponse à Henri d'Ideville*, Michel Levy frères, Paris.
- Dustin, M. (2016). Culture or masculinity? Understanding gender-based violence in the UK, *Journal of Poverty and Social Justice*, 24(1): 51-62, Copyright The Policy Press, disponibile presso: Library of London School of Economics.
- Dutilleul, P., Stockwell, J.D., Frigon, D., Legendre, P. (2000) The Mantel test versus Pearson's correlation analysis: assessment of the differences for biological and environmental studies, *J Agric Biol Environ Stud* 5:131–150.
- Eckart, V., Young, G. (1935). The approximation of one matrix by another of lower rank, *Psychometrika*, 1935, 1, 211-218.
- Eckart, V., Young, G. (1935). The approximation of one matrix by another of lower rank. *Psychometrika*, 1935, 1, 211-218.
- Ehrmann, M., Turchi, M., Steinberger, R. (2011). Building a multilingual named entity-annotated corpus using annotation projection, *In RANLP*, pp. 118–124.
- EIGE, European Institute for Gender Equality (2017a). *Gender Equality Index 2017, Measuring gender equality in the European Union 2005-2015, Main Findings*, European Union Publishing.
- EIGE, European Institute for Gender Equality (2017b). *Statistical Database*.
- EIGE, European Institute for Gender Equality (2018). *Gender Equality Index 2017, Measuring gender equality in the European Union 2005-2015, Main Findings*, European Union Publishing.
- Elagamy, M.N. (2017). Doctor of Philosophy Thesis: Stock Market Random Forest-Text Mining (SMRF-TM) Approach to Analyse Critical Indicators of Stock Market Movements, Staffordshire University.
- Elarnaoty, M., Rahman, S., A., Fahmy, A. (2012). A machine learning approach for opinion holder extraction Arabic language, *in CoRR*.
- Ellsberg, M., Heise, L. (2002). Bearing witness: ethics in domestic violence research, *Lancet*, Vol. 359: 1599-1604.
- Elman, J.L. (1990). Finding Structure in Time, *in Cognitive Science*, vol. 14, n° 2.
- Engels, F. (2005). *L' origine della famiglia, della proprietà privata e dello Stato* (1884), Traduzione a cura di Della Terza, D., Riuniti Editore.
- Enweiji, M.Z., Lehinevych, T., Glybovets, A. (2017). Cross-Language Text Classification with Convolutional Neural Networks from scratch, *Eureka: Physics and Engineering*, n.2.

- Escofier, B., Pagès, J. (1990). *Analyses factorielles simples et multiples: objectifs, méthodes, interprétation*, Dunod, Paris.
- Esiodo (2008). *Opere e giorni*. Garzanti Editore.
- Esuli, A., Sebastiani, F. (2005). Determining the Semantic Orientation of Terms through Gloss Classification, *In CIKM'05: Proceedings 14th ACM International Conference on Information and Knowledge Management*, Bremen, Germany.
- EURES-ANSA, Ricerche Economiche e Sociali (2013). *L'omicidio volontario in Italia*, Rapporto EURES 2013. In collaborazione con l'Agenzia Ansa.
- Eurofound (2016). *Sixth European Working Conditions Survey*, disponibile all'indirizzo: <https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys/sixth-european-working-conditions-survey-2015>.
- European Commission (2015). Eurobarometer Special Surveys. Disponibile all'indirizzo: https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_428_en.pdf
- European Commission (2017). *Report on equality between women and men in EU*, European Union Publishing.
- European Commission (2018). *EU guidelines on violence against women and girls and combating all forms of discrimination against them*, European Union Publishing.
- European Commission (2018). *Report on equality between women and men in EU*, European Union Publishing.
- Eurostat. (2017) Sustainable Development in the European Union — Overview of progress towards the SDGs in an EU context, available at: http://ec.europa.eu/eurostat/documents/4_031_688/8_461_538/KS-01-17-796-EN-N.pdf/f9c4e3f9-57eb-4f02-ab7a-42a7ebcf0748 (Ultima consultazione: aprile 2018).
- Evans, S. E., Davies C., DiLillo D. (2008). Exposure to domestic violence: A meta-analysis of child and adolescent outcomes, *Aggression and Violent Behavior Volume 13, Issue 2, pp. 131–140*
- Fabris, G. (1992). *La pubblicità. Teoria e prassi*, Franco Angeli Editore.
- Faga, A. et al. (2000). Sulphuric acid burned women in Bangladesh: A social and medical problem, *Burns 26 (8):701–709*.
- Fallaci, O. (2009). *Il sesso inutile. Viaggio intorno alla donna*. Milano: Bureau Biblioteca Univ. Rizzoli.
- Faruqui, M. Dyer, C. (2014). Improving vector space word representations using multilingual correlation, *In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 462–471, Gothenburg, Sweden. Association for Computational Linguistics.
- Faust, K. (2005). Using Correspondence Analysis for Joint Displays of Affiliation Network in Models and Methods in Social Network Analysis, eds. Peter J. Carrington, John Scott and Stanley Wasserman.

- Feci, F., Schettini., L. (2017). *La violenza contro le donne nella storia. Contesti, linguaggi, politiche del diritto (secoli XV-XXI). Storia delle donne e di genere.* Viella Editore.
- Ferraris, M. (1988). *Storia dell'ermeneutica*, Bompiani, Milano.
- Fish *et al.* (2011). Birds of the Internet: Towards a field guide to the organization and governance of the participation, *Journal of Cultural Economy*.
- Flood, M. (2002). Pathways to manhood: The social and sexual ordering of young men's lives, *Health Education Australia, Vol. 2, No. 2: pp-24-30*.
- FRA, European Union Agency for fundamental Rights (2014). *Report summary: Violence against women: an EU-wide survey. Results at a glance.* Publications Office of the European Union.
- Fraire M. (2007). Analisi Multi-way delle principali caratteristiche d'ingresso, percorso ed esito degli studenti delle facoltà di Sociologia italiane, in Fasanella A., L'impatto della riforma universitaria 3+2 nella formazione sociologica, Studi, ricerche e percorsi di Sociologia, Franco Angeli.
- Fraire, M. (2000). Analisi dei Dati a Tre-vie delle Risposte a Domande Aperte e Indicatori Empirici, *Proceedings in 5^{es} Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2000)*.
- Friedan, B. (1963). *The Feminine Mystique*, W. W. Norton and Co. Editori.
- Fulper, R., Ciampaglia, G. L., Ferrara, E., Ahn, Y., Flammini, A., Menczer, F., Lewis, B. & Rowe, K. (2014). Misogynistic Language on Twitter and Sexual Violence, *ChASM Workshop at ACM WebSci-2014*.
- Garcia, C. (2012). A simple procedure for the comparison of covariance matrices, *BMC Evolutionary Biology, volume 12: 222*.
- Garofano, L., Diaz, R. (2013). *I labirinti del male*, Modena, Edizione Infinito.
- Genetti, C. (2014). *How Languages Work: An Introduction to Language and Linguistics*, Cambridge, *Cambridge University Press*.
- Gentry, J. (2009). TwitteR package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/twitteR/twitteR.pdf>.
- Gentry, J., Temple Lang, D. (2011). ROAuth package: Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/ROAuth/ROAuth.pdf>.
- Gherghi M., Lauro, C. (2008). *Appunti di Analisi dei Dati Multidimensionali. Metodologia ed esempi*, RCE Multimedia.
- Giannakidou, A. (2008). Negative and positive polarity items: licensing, compositionality and variation, in K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning (2, pp. 1660–1712)*. Berlin: Mouton de Gruyter.
- Gimbutas, M. (1990). *Il linguaggio della dea. Mito e culto della dea madre nell'Europa neolitica*. Longanesi Editore.

- Giuliano L., La Rocca, G. (2008). L'analisi automatica e semi-automatica dei dati testuali, LED Edizioni Universitarie, Milano.
- Givón, T. (1983). Topic Continuity in Discourse: A quantitative cross-language study. *A cura di John Benjamins, in Typological Studies in Language, 3*.
- Goke-Pariola, A. (1993). The Role of Language in the Struggle for Power and Legitimacy in Africa, *African Studies, 31*, The Edwin Mellen Press.
- Golub, G.H., Van Loan, C.F. (1983). Matrix computation. *John Hopkins University Press*.
- Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal Royal Stat. Soc. Part B 53, 2*.
- Gouws, S., Bengio, Y., Corrado, G. (2015). BilBOWA: Fast bilingual distributed representations without word alignments, *Proceedings of The 32nd International Conference on Machine Learning*, pp. 748–756.
- Gouws, S., Søgaard, A. (2015). Simple task-specific bilingual word embeddings, *In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1386–1390.
- Gower, J. C. (1971). Statistical methods of comparing different multivariate analyses of the same data. *In F. R. Hodson, D. G. Kendall, and P. Tautu (eds.) Mathematics in the Archaeological and Historical Sciences*, pp. 138-149. *Edinburgh Univ. Press: Edinburgh*.
- Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*. 40(1).
- Graffi, G., Scalise, S. (2002). Le lingue e il linguaggio. Introduzione alla linguistica, Bologna, Il Mulino.
- Graves, A. (2013). Generating sequences with recurrent neural networks, *Neural and Evolutionary Computing (cs.NE)*.
- Greenacre, M. J. (1984). Theory and applications of correspondence analysis, *London: Academic Press*.
- Grimmer, J., Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts, *Political Analysis, 21(3)*, 267–297.
- Guglielmin, M. S. *et al.* (2013). The mother who cannot provide liberation: family atom analysis of women victims of domestic violence, *in Testoni I., Zamperini A., Moita G., Bacuță M. D. (a cura di) La camera blu. Rivista di studi di genere, volume 218 monografico: Against gender-based violence: from Italian debate to intercultural dialogue, 10*.
- Günther, E., Quandt, T. (2016). Word counts and topic models: Automated text analysis methods for digital journalism research, *Digital Journalism, 4(1)*, 75–88.
- Haraway, J., D. (1985). *Simians, Cyborgs and Women: The Reinvention of Nature*, New York, Routledge, Taylor and Francis Group, 1991.

- Hartigan, J.A., Wong, M.A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics, Journal of the Royal Statistical Society, Vol. 28, No. 1 (1979), pp. 100-108.*
- Harvard IV Dictionary in General Inquirer (2019). Disponibile all'indirizzo: <http://www.wjh.harvard.edu/~inquirer/homecat.htm> (Ultima consultazione: 20 Novembre 2019).
- Hatzivassiloglou, V., McKeown, R. K. (1997). Predicting the semantic orientation of adjectives, *In Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics, EACL '97, pages 174–181, Stroudsburg, PA, USA.*
- Hearst, M. (1999). Untangling Text Data Mining, *in Proceedings of the 37th Annual Meeting of the Association of Computational Linguistics, Kaufmann Publishers, San Francisco, 3-10.*
- Hebb, D.O. (ed. 2002). *The Organization of Behavior: A Neuropsychological Theory, Taylor & Francis, edizione originale 1949.*
- Herman, J. (2005). *Guarire dal trauma. Affrontare le conseguenze della violenza, dall'abuso domestico al terrorismo, Magi Edizioni.*
- Hermann, K.M., Blunsom, P. (2014). Multilingual models for compositional distributed semantics, *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Vol. 1: Long Papers, Association for Computational Linguistics, Baltimore, Maryland, pp. 58– 68.*
- Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory, *Neural Computation. 9 (8): 1735–1780.*
- Hosseini, K. (2007). *A Thousand Splendid Suns, Bloomsbury or Andre Brink (2012) Philida, Harvill Secker.*
- Host, G.E., Ramm, C.W., Padley, E.A., Pregitzer, K.S., Hart, J.B., Cleland, D.T. (1993). Field Sampling and Data Analysis Methods for DEvelopment of Ecological Land Classifications: An Application on tyhe Manistee National Forest, *General Technical Report NC-162, United States Department of Agriculture, Forest Service.*
- Hu, M., Liu, B. (2004). Mining and Summarizing Customer Reviews, *in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), USA.*
- Humm, M. (1990). Wave (definition), in Maggie Humm (a cura di), *The dictionary of feminist theory, Columbus, Ohio State University Press, p. 251.*
- Hurley, J.R., Cattell, R., B. (1962). The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Computers in Behavioral Science. Journal of the Society for General System Research, Vol.7, Issue 2, pp. 258-262.*
- Husson, F., Josse, J., Pagès, J. (2010). Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data?, *Technical Report- Agrocampus, Department of Applied Mathematics.*

- IDEA (2017). *Regional Organisations, Gender Equality, and the Political Empowerment of Women*, disponibile: <https://www.idea.int/publications/catalogue/regional-organisations-gender-equality-and-political-empowerment-women>.
- International Information Centre for Terminology (2011). ISO 639 Codes. Disponibile all'indirizzo: http://www.infoterm.info/standardization/iso_639_1_2002.php (Ultima consultazione: 17 Novembre 2019).
- Irigaray, L. (1974). *Speculum de l'autre femme (1974), Traduzione italiana. Speculum. L'altra donna*, Feltrinelli Editore, 1975.
- Irschick, D.J., Vitt, L.J., Zani, P.A., (1997). A Comparison of Evolutionary Radiations in Mainland and Caribbean Anolis Lizards Losos, J.B., *Ecology Vol. 78, No. 7 (Oct., 1997)*, pp. 2191-2203.
- Islam, M.Z., et al. (2019). A Semantics Aware Random Forest for Text Classification, in *CIKM '19 Proceedings of the 28th ACM International Conference on Information and Knowledge Management* pp. 1061-1070.
- Istat (2016). La violenza contro le donne. Disponibile all'indirizzo: <https://www.istat.it/it/archivio/194779> (Ultima consultazione: Maggio 2018).
- Jack, F.R., Piggott, J.R. (1992). Free choice profiling in consumer research, *Food Quality and Preference, Volume 3, Issue 3, 1991–1992*, pp. 129-134.
- Jackson, D.A., Harvey, K.M. (1989). Are probability estimates from the permutation model of Mantel's test stable? *Can. J. Zool*, 67: 766-769.
- Jacobs, Renée, E. (1991). Iroquois Great Law of Peace and the United States Constitution: How the Founding Fathers Ignored the Clan Mothers, in *American Indian Law Review*, vol. 16, no. 2, pp. 497-531.
- Jain, A.K., Murty, M.N., Flynn, P.J. (1999). Data clustering: a review, in *ACM Computing Surveys*, 31(3), 264-323.
- Jenkins, R. (1992). Pierre Bourdieu, *London: Routledge*.
- Jockers, M. (2017). Syuzhet package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf>.
- Johnson, M., P. (1995). Patriarchal Terrorism and Common Couple Violence: Two Forms of Violence against Women, in *Journal of Marriage and Family*, Vol. 57, No. 2, pp. 283-294.
- Johnson, N. F., Leahy, R., Johnson Restrepo, N., Velasquez, N., Zheng, M., Manrique, P., Devkota, P., Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology, *Nature*, 573 (7773): 261–265.
- Jung, C. (1977). Gli archetipi dell'inconscio collettivo, *Bollati Boringhieri Editore*.
- Kabeer, N. (2014). Violence against women as 'Relational' vulnerability: engendering the sustainable human development agenda. *Occasional Paper, United Nations Development Programme*, New York, USA.

- Kaplan, A. M., Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media, *Business Horizons*, 53(1), 59–68.
- Kaufman, M. (1997). *Working with men and boys to challenge sexism and end men's violence* presented at UNESCO experts group meeting Male roles and Masculinities in the perspective of a Culture of Peace, Oslo, Norway, 24-28 Settembre 1997.
- Kazi-Aoual, F., Hitier, S., Sabatier, R., Lebreton, J.-D. (1995). Refined approximations to permutations tests for multivariate inference. *Computational Statistics and Data Analysis*, 20, 643–656.
- Kelly, J. B. & Johnson, M. P. (2008). Differentiation among types of intimate partner violence: Research update and implications for interventions. *Family Court Review* 46.3: 476-499.
- Kemp, S.E., Hort, J., Hollowood, T. (2018). *Descriptive Analysis in Sensory Evaluation*, John Wiley & Sons Ltd.
- Khachumov, M. V. (2012). Distances, metrics and cluster analysis, *Scientific and Technical Information Processing, Volume 39, Issue 6*, pp 310–316.
- Kimmel, M. (1990) After fifteen years: the impact of the sociology of masculinity on the masculinity of sociology. In J. Hearn & D. Morgan (Eds) *Men, Masculinities and Social Theory (Critical Studies on Men and Masculinities)* Unwin Hyman.
- King, B.M., Arents, P. (1991). A statistical test of consensus obtained from Generalized Procrustes Analysis, *J. Sen. Stud.* 6, 37-48.
- Kirti, A., Kumar, P., Yadav, R. (2011). The face of honour based crimes: global concerns and solutions, *International Journal of Criminal Justice Sciences. South Asian Society of Criminology and Victimology*, pp. 343–357.
- Klementiev, A. Titov, I, Bhattarai, B. (2012). Inducing crosslingual distributed representations of words, *In Proceedings of COLING 2012. The COLING 2012 Organizing Committee, Mumbai, India*, pp. 1459–1474.
- Kohonen, T., Huang, T.S., Schroeder, M.R. (ed. 2000). *Self-Organizing Maps, Series in Information Sciences*, Springer, Verlag.
- Krebs, C.P., et al. (2007). *Campus sexual assault (CSA): final report*, RTI International.
- Krippendorff, K. (1983). *Analisi del contenuto. Introduzione metodologica*, Trad. italiana, Eri, Torino (1^a ed. 1980).
- Kristeva, J. (1969). *Semeiotiké. Recherche pour une sémanalyse*, Seuil, Paris.
- Kristof, W., Wingersky, B. (1971). Generalization of the orthogonal Procrustes rotation procedure to more than two matrices. *Proceedings of the 79-th Annual Conv. of the American Psychological Ass.*, 6.
- Krupa, V. (1969). Some remarks on the translation process, *in Asian and African Studies*, n. 4, Bratislava 1968.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

- La Rocca, G. (2007). Software per l'analisi qualitativa dei testi. *Contributi su aree tematiche differenti*, vol.5, n.3.
- Lacan, J. (1949). Lo stadio dello specchio come formatore della funzione dell'Io, *XVI Congresso internazionale di psicoanalisi*, Zurigo, luglio 1949.
- Lagarde, M. (1997). Identidades de género y derechos humanos. La construcción de las humanas, VII curso de verano, *Educación, democracia y nueva ciudadanía*, Universidad Autónoma de Aguascalientes, dal sito della Cátedra UNESCO de Derechos Humanos de la UNAM.
- Lam, C. (2016). Correspondence analysis: A statistical technique ripe for technical and professional communication researchers, *IEEE Transactions on Visualization and Computer Graphics*. 53, pp 299-310.
- Lample, G., Denoyer, L., Ranzato, M.A. (2018). Unsupervised machine translation using monolingual corpora only, *In Proceedings of ICLR (Conference Papers)*.
- Laswell, H.D. et al. (1949). Language of Politics: Studies in Quantitative Semantics, George Stuart, New York (Traduzione italiano: Il linguaggio della politica: studi di semantica quantitativa, ERI, Torino, 1979).
- Laviosa, S. (2002). Corpus-based Translation Studies: Theory, Findings, Applications. Approaches of Translation Studies.
- Law, H.G., Snyder, C.W., Hattie, J.A., Mc Donald, R.P. (eds.) (1984). Multimode Data Analysis, Praeger, New York.
- Lebart L., Salem A. (1988). Analyse statistique des données textuelles, Dunod, Paris.
- Lebart L., Salem A. (1994). *Statistique textuelle*, Dunod, Paris.
- Lebart L., Salem A., Berry L. (1998). Exploring textual data, *Kluwer Academic Publishers, Dordrecht (The Netherlands)*.
- Lebart, L., Morineau, A., Piron, M. (1995). *Statistique exploratoire multidimensionnelle*, Dundod, Paris.
- Lebart, L., Salem, A., Berry, L. (1998). Exploring Textual Data, *Kluwer Academic Publishers, Dordrecht, The Netherlands*.
- Lee, C.H., Yang, H.C. (2003). A Multilingual Text Mining Approach Based on Self-Organizing Maps, *In Applied Intelligence 18(3):295-310*.
- Legendre, P., & Legendre, L. (1998). Numerical ecology (2[^] ed.), *Amsterdam: Elsevier*.
- Lenci, A., Montemagni, S., Pirrelli, V. (2005). Testo e computer. Carocci Editore, Roma.
- Leti G. (1979), Distanze ed indici statistici, Ed. La Goliardica Roma.
- Lévi-Strauss, C. (1947). *Les structures elementaires de la parenté*, Presses Universitaires de France, Paris.
- Lewis, D., et al. (2004). Rcv1: A new benchmark collection for text categorization research, *Journal of machine learning research 5: 361-397*.

- Leymann, H. (1990). *Psychological reactions to violence in working life: bank robberies*, Umea University, medical dissertation.
- Li, Y.J., Zhao, X., Xu, W., Yan, Y. (2018). Cross-Lingual Multi-Task Neural Architecture for Spoken Language Understanding, *Interspeech 2018, 19th Annual Conference of the International Speech Communication Association, Hyderabad, India*.
- Lippman, W. (2015). *Public Opinion*, Wilder Publications (1[^] ed. 1922).
- Lis, A., Sambin, M. (1977) *Analisi dei Cluster*, Cleup.
- Litvak, M., Vanetik, N. (2019). *Multilingual Text Analysis: Challenges, Models, and Approaches*, World Scientific Editor.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies*, Series Editor Graeme Hirst, University of Toronto, Morgan & Claypool Publishers.
- Liu, B. (2015). *Sentiment Analysis: mining opinions, sentiments, and emotions*, Cambridge, England: Cambridge University Press.
- Liu, B., Hu, M., Cheng, J. (2005). Opinion Observer: Analyzing and Comparing Opinions on the Web, in *Proceedings of the 14th International World Wide Web conference (WWW-2005), May 10-14, 2005, Chiba, Japan*.
- Löbner, S. (2000). Polarity in Natural Language: Predication, Quantification and Negation in Particular and Characterizing Sentences, *Linguistics and Philosophy*, 23, 213–308.
- Logos Group (2014). Disponibile all'indirizzo: http://courses.logos.it/IT/1_24.html (Ultima consultazione: 10 luglio 2019).
- Losito, G. (2002). *L'analisi del contenuto nella ricerca sociale* (3[^] edizione), Franco Angeli Editore.
- Loughran, T., McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance*, 66:1, 35-65.
- Lûdskanov, A. (2008). *Un approccio semiotico alla traduzione. Dalla prospettiva informatica alla scienza traduttiva*, a cura di Bruno Osimo, Milano, Hoepli, versione originale 1967.
- MacKinnon, C. (1989). *Toward a Feminist Theory of the State*, Cambridge: Harvard University Press.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol.1, N.14, pp. 281-297*.
- Magli, P. (2008). *Semiotica. Teoria, metodo, analisi* (3[^] Edizione). Marsilio Editori.
- Malmasi, S., Zampieri, M. (2018). Challenges in discriminating profanity from hate speech, *Journal of Experimental & Theoretical Artificial Intelligence*, 30(2), pp. 187-202.

- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach, *Cancer Res.*
- Mazzaglia, S. (2010). Il “danno invisibile” nella violenza assistita dai minori tra aspetti penali, civili e psicologici, Edizioni Universitarie Romane.
- McEnery, T., Wilson, A. (1996). *Corpus Linguistics*, Edinburgh University Press.
- McMahon, R., (2016). Histories of Interpersonal Violence in Europe and North America, 1700-Present, *The Oxford Handbook of the History of Crime and Criminal Justice*, a cura di Paul Knepper, Ania Johansen, OUP, Oxford, pp. 111-131.
- McMahon, S., Huang, C.C., Boxer, P. & Postmus, J. L. (2013). The impact of emotional and physical violence during pregnancy on maternal and child health at one year post-partum, *Children and Youth Services Review* 33, pp. 2103 – 2111.
- McQuarrie, M., Walker, T., E. (2015). *Democratizing Inequalities, Dilemmas of the New Public Participation*, Caroline W. Lee Edition.
- McRobbie, A. (2004). Notes on “What Not To Wear” and Post-Feminist Symbolic Violence, in Lisa Adkins and Beverley Skeggs (eds) *Feminism after Bourdieu*, pp. 99-109.
- McWhirter, J., Baxter, P., Cooper, T., Redif, S. (2007). An EVD algorithm for para-Hermitian polynomial matrices, *IEEE Trans. Signal Process.* 55(5), 2158–2169.
- Merzagora Betsos, I. (2006). *Criminologia della violenza e dell’omicidio, dei reati sessuali, dei fenomeni di dipendenza*. Padova: Cedam.
- Mihalcea, R., Banea, C., Wiebe, J. (2007). Learning multilingual subjective language via cross-lingual projections, *In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 976–983.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *arXiv:1310.4546 [cs.CL]*.
- Mikolov, T., Karafiát, M., Burget, L., Černocký, J., Khudanpur, S. (2010). Recurrent neural network-based language model, *in Eleventh Annual Conference of the International Speech Communication Association*.
- Mimno, D., Wallach, H.M., Naradowsky, J., Smith, D.A., McCallum, A. (2009). Polylingual Topic Models, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 880–889.
- Misuraca, M., Scepi, G., Spano, M. (2019). A network-based concept extraction for managing customer requests in a social media care context, *International Journal of Information Management*.
- Mohammad, S. (2017). Word Affect Intensities, *Computation and Language (cs.CL)*, [arXiv:1704.08798 \[cs.CL\]](https://arxiv.org/abs/1704.08798).
- Mohammad, S., Turney, P. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, *in Proceedings of the NAACL-*

HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, LA, California.

- Mohammad, S., Turney, P. (2013a). Crowdsourcing a word-emotion association lexicon, in *CoRR*, *abs/1308.6297*.
- Mohammad, S., Turney, P. (2013b). NRC Emotion LEXicon, in NRC Publications Archive. Disponibile all'indirizzo: <https://nrc-publications.canada.ca/eng/view/fulltext/?id=0b6a5b58-a656-49d3-ab3e-252050a7a88c>.
- Mondal, M., Silva, L.A., Benevenuto, F. (2017). A Measurement Study of Hate Speech in Social Media, *HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media, July 2017*, pp. 85–94.
- Money, J., Hampson J.,G., Hampson, J. L. (1955). Hermaphroditism: recommendations concerning assignment of sex, change of sex, and psychologic management, *Bullettin of the Johns Hopkins Hospital*, n.97, pp.284-300, Johns Hopkins Press.
- Morini, M. (2007). *La traduzione. Teorie. Strumenti. Pratiche*, Sironi Editore.
- Morris, C., W. (1938). Foundations of the Theory of Signs, International Encyclopedia of Unified Sciences, *University of Chicago Press*.
- Mosier, C. I. (1939). Determining a simple structure when loadings for certain tests are known. *Psychometrika*, 4, 149-162.
- Murphy, J. *et al.* (2014). Social Media in Public Opinion Research: Executive Summary of the Aapor Task Force on Emerging Technologies in Public Opinion Research, in *Public Opinion Quarterly*, Volume 78, Issue 4, Winter 2014, pp. 788–794.
- Næs, T., Brockhoff, P. B., and Tomic, O. (2010). *Statistics for Sensory and Consumer Science*, Wiley, Chichester.
- Næs, T., Varela, P., Berget, I. (2018). Individual differences in Sensory and Consumer Science. Experimentation, Analysis and Interpretation, in *Woodhead Publishing Series in Food Science, Technology and Nutrition*.
- Negishi, M. (2014). Meet Shigetaka Kurita, the Father of Emoji, *The Wall Street Journal*, 26 marzo 2014. Disponibile all'indirizzo: <https://blogs.wsj.com/japanrealtime/2014/03/26/meet-shigetaka-kurita-the-father-of-emoji/> (Ultima Consultazione: 15 Novembre 2019).
- Nestrud, M.A., Lawless, H.T. (2011). Recovery of subsampled dimensions and configurations derived from napping data by MFA and MDS, *Attention, Perception, & Psychophysics*, Volume 73, Issue 4, pp 1266–1278.
- Nielsen, A.F. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, in *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings : 93-98. 2011*.

- Norma Internazionale, Terminologia-Vocabolario (1995). Appendice I, ISO 1087. Traduzione italiana a cura di Pucci, C.R. e Fondazione UGo Bordoni. Disponibile all'indirizzo: http://www.term-minator.it/corso/doc/mod3_normaISO.pdf.
- Nussbaum, M. (2011). *Creating Capabilities: The Human Development Approach*, Harvard University Press.
- OECD (2017). *The Pursuit of Gender Equality. An Uphill Battle*, disponibile all'indirizzo: <http://www.oecd.org/publications/the-pursuit-of-gender-equality-9789264281318-en.htm>
- Offen, K., M. (2000). *European Feminism 1700-1950: a political History*, Stanford University press.
- Olivieri, D., Margiarotta, U. (2014). Il grafo della formazione. L'albero generativo della conoscenza pedagogica. In Margiarotta Umberto, *Ontologie Pedagogiche, Pensa Multimedia*.
- Omicore Agency (2019). Twitter by the Numbers: Stats, Demographics & Fun Facts. Disponibile all'indirizzo: <https://www.omicoreagency.com/twitter-statistics/>.
- OMS (2002). Conseguenze della violenza da parte del partner sulla salute, *World report on violence and health*, p.138.
- Onorati, T., Diaz, P. (2016). Giving meaning to tweets in emergency situations: a semantic approach for filtering and visualizing social data, *SpringerPlus 5 (1)*, p. 1782.
- ONU (1993). *Conferenza Mondiale sui diritti umani*, Dichiarazione di Vienna e Programma d'azione.
- Osimo, B. (2001). Traduzione e nuove tecnologie: informatica e internet per traduttori: guida pratica con glossario, Hoepli Editore.
- Package 'quanteda' (2019). Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/quanteda/quanteda.pdf> (Ultima consultazione: 17 Novembre 2019).
- Package 'stopwords' (2019). Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/stopwords/stopwords.pdf> (Ultima consultazione: 17 Novembre 2019).
- Padò, S., Lapata, M. (2009). Cross-lingual Annotation Projection for Semantic Roles, *Journal of Artificial Intelligence Research 36*:307-340.
- Pagès, J. (2014). Multiple Factor Analysis by Example Using R, *Chapman & Hall/CRC The R Series London*.
- Pallotti, G. (1999). Relatività linguistica e traduzione, In *Versus, Quaderni di studi semiotici*, 82, pp. 109-138.
- Pang, B., Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, In *Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics*, pp. 271-278, Barcelona, ES.

- Pang, B., Lee, L., Vaithyanathan, S (2002). Thumbs up?: sentiment classification using machine learning techniques, *In Proceedings of conference on empirical methods in natural Language Processing (emnLP-2002)*.
- Paulos, K. *et al.* (2015). Sensory Characterization and Consumer Preference Mapping of Fresh Sausages Manufactured with Goat and Sheep Meat, *Journal of Food Science Vol. 80, Nr. 7, Institute of Food Technologists*.
- Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J. (2007). The development and psychometric properties of LIWC2007, [Software manual], Austin, TX. Disponibile all'indirizzo: <http://liwc.wpengine.com>.
- Peres-Neto P. R., Jackson D.A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test, *Oecologia, 129, Springer-Verlag, pp.169–178*.
- Perrin, A. (2015). Social Media Usage: 2005-2015. Washington, D.C.: Pew Internet & American Life Project; Smith, T., 2009, The Social Media Revolution, *International Journal of Market Research, 51 (4), 559–561*.
- Perrin, L., Symoneaux, R., Maitre, I., Asselin, C., Jourjon, F., Pagès, J. (2008). Comparison of three sensory methods for use with the napping procedure: Case of ten wines from loire valley, *Food Quality and Preference, 19:1–11*.
- Piccone, S., Saraceno, C. (1996). *Genere. La costruzione sociale del femminile e del maschile*, Il Mulino Editore.
- Pielou, E. C. (1984). The interpretation of ecological data: A primer on classification and ordination, *New York: Wiley*.
- Plahl, C., Schlüter, R., Ney, H. (2011). Cross-lingual portability of Chinese and English neural network features for French and German LVCSR, *In 2011 IEEE Workshop on Automatic Speech Recognition & Understanding*.
- Plutarch (1914). The Parallel Lives. *Loeb Classical Library edition, Vol.1*.
- Plutchik, R. (1991). The Emotions, *University Press of America*.
- Poe, S. (2014). Comparison of Natural and Nonnative Two-Species Communities of Anolis Lizards, *The American Naturalist, Vol. 184, No. 1 (July 2014), pp. 132-140*.
- Ponzio, G. (2004). *Crimini segreti: maltrattamento e violenza alle donne nella relazione di coppia*. Baldini e Castoldi Editori.
- Pröllochs, N., Feuerriegel, S., Neumann, D. (2015). Generating Domain-Specific Dictionaries Using Bayesian Learning, *23rd European Conference on Information Systems (ECIS)*.
- Pruss, D. *et al.* (2019). Zika discourse in the Americas: A multilingual topic analysis of Twitter, *PLoS ONE 14(5): e0216922*.
- Purohit, H. *et al.* (2015). Gender-Based Violence in 140 Characters or Fewer: A #BigData Case Study of Twitter, *Social and Information Networks (cs.SI)*.

- Qannari, E.M., MacFie, H.J.H, Courcoux, P. (1999). Performance indices and isotropic scaling factors in sensory profiling. *Food Quality and Preference*, 10, 17–21.
- R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Disponibile all'indirizzo: <https://www.R-project.org/>.
- Radford, J., Russell, D. (1992). “*Femicide: The Politics of Woman Killing*”, Twayne Publishers, p.379.
- Randone, V. (2018). *Ex/forse ex: Gli amori affamati*, Editore indipendente.
- Ratna, A.A.P *et al.* (2018). Cross-Language Automatic Plagiarism Detector Using Latent Semantic Analysis and Self-Organizing Map, in *Proceedings AIVR 2018, International Conference on Artificial Intelligence and Virtual Reality 2018*, pp. 83-87.
- Reisel, D., Creighton, S. M. (2015). Long term health consequences of Female Genital Mutilation (FGM), *Maturitas*. 80 (1): 48–51.
- Reiter, R. (1975). *Toward an Anthropology of Women*, Monthly Review Press.
- Representation of the People Act (1918). *Parliament of the United Kingdom*.
- Rich, A. (1929). Compulsory Heterosexuality and Lesbian Existence, in C. R. Stimpson, E. S. Pearson (edd.), *Women, Sex, and Sexuality*, Chicago University Press, 1980.
- Ricoeur, P. (1987). Logica ermeneutica? In *Aut Aut*, 217-218, pp.64-100 (ed.1981).
- Ricolfi, L. (1997). *La ricerca qualitativa*, Carocci, Roma.
- Rinker, T. (2016). Sentimentr package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf>.
- Romito, P. (1999). Dalla padella alla brace. Donne maltrattate, violenza privata e complicità pubbliche, *POLIS XIII*, pp. 235-254.
- Romito, P. (2011). *La violenza di genere su donne e minori. Un'introduzione*, Milano, Franco Angeli Editore.
- Royakkers, L. (2000). The Dutch Approach to Stalking Laws, *Berkeley Journal of Criminal Law*.
- Rubin, S., G. (1975). The Traffic in Women: Notes on the *Political Economy of Sex*, in M. Rayna Reiter, R.R. (ed.), *Towards an Anthropology of Women*, New York, *Monthly Review Press*, pp. 157-210.
- Russell D. E. H. (1974). *The Politics of Rape: The Victim's Perspective*, New York, *Stein and Day*.
- Salton G., Buckley C. (1988). Term weighting approaches in automatic text retrieval, *Information Processing and Management*, 24(5), 513-523.
- Salton, G., Wong, A., Yang, C. S. (1975). A vector space model for automatic indexing, *Communications of the ACM*, 18, 613–620.

- Sapir, E., Whorf, B.L. (2017). *Linguaggio e relatività*. A cura di Marco Carassai ed Enrico Crucianelli, Castelveccchi, Roma.
- Saussure, F. (1916). *Corso di linguistica generale*, Traduzione Italiana, Editori Laterza, 2017.
- Sayyadi, H., Raschid, L. (2013). A graph analytical approach for topic detection, *ACM Transactions on Internet Technology*, 13, 1–23.
- Scheper-Hughes, N. (1992). *Death Without Weeping: The Violence of Everyday Life in Brazil*, Berkeley, University of California Press.
- Scheper-Hughes, N. (1996). Small Wars and Invisible Genocides, *Social Science and Medicine*, Vol. 43, No. 5: 889-900.
- Schmitd, D. (2017). Meanr package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/meanr/meanr.pdf>.
- Schneider, J.W., Borlund, P. (2007). Matrix comparison, Part 2: Measuring the resemblance between proximity measures or ordination results by use of the mantel and procrustes statistics, *Journal of the American Society for Information Science and Technology*.
- Schönenemann, P.H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1), pp. 1-10.
- Schönenemann, P. H., Carroll., R. M., (1970). Fitting one matrix to another under choice of a central dilation and a rigid motion, *Psychometrika*, 35(2).
- Sen, A. (2000). *Development as Freedom*, Oxford: Oxford paperbacks.
- Sen, P., (2003). Successes and Challenges: Understanding the Global Movement to end Violence against Women, in *Global Civil Society yearbook*, Anheier H, Glasius M, Kaldor M (eds).
- Shaw, E., F., (1979). Agenda Setting and Mass Communication Theory, in *Gazette (international Journal for Mass Communication Studies)*.
- Shell-Duncan, B. (2008). Dalla salute ai diritti umani: taglio genitale femminile e politica di intervento, *Antropologo americano*.
- Shilling, C. (1993). *The Body and Social Theory*. London: Sage .
- Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *J.R Statistical Society*, 40(2):234-238.
- Smith, P.K., et al. (2008). Cyberbullismo: la sua natura e il suo impatto negli alunni delle scuole secondarie, *Journal of Child Psychology and Psychiatry*, Wiley 49 (4): 376-385.
- Smith, S.L., Turban, D.H.P., Hamblin, S., Hammerla, N.Y. (2017). Offline Bilingual Word Vectors, Orthogonal Transformations and the inverted softmax, *In Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)*.
- Sneath, P.H.A. (1967). Trend surface analysis of transformation grids. *Journal of Zoology*, London, 151:65-122. pp.32, 125,303, 306.

- Soavi, G. (2009). La violenza assistita e il male che resta, in *L'invisibile violenza contro i bambini*, Scuola della non violenza.
- Sobrero, A. (1993). Pragmatica, in Alberto Sobrero (a cura di), *Introduzione all'italiano contemporaneo. Le strutture*, ed. Laterza, Roma-Bari.
- Social Media Statistics 2018 for Business (2019). Disponibile all'indirizzo: blog.statusbrew.com/social-media-statistics-2018-for-business. (Ultima consultazione: 15 Novembre 2019).
- Søgaard, A., Ruder, S., Vulić, I. (2018). On the Limitations of Unsupervised Bilingual Dictionary Induction, *In Proceedings of ACL 2018*.
- Sokal, R., & Rohlf, F. (1962). The comparison of dendograms by objective methods, *Taxon*, 11, 33-40.
- Solomentsev, E.D. (2002). Euclidean space, in *Encyclopaedia of Mathematics*, Springer e European Mathematical Society.
- Spencer, MC. (1981). *Charles Fourier. Twayne's World Author's Series 578*, Boston.
- Spina, S. (2001). *Fare i conti con le parole. Introduzione alla linguistica dei corpora*. Perugia,
- Spinicci, P. (1985). *I pensieri dell'esperienza: interpretazione di Esperienza e giudizio di Edmund Husserl*, La Nuova Italia.
- Stark, E. (2010). Do violent acts equal abuse? Resolving the gender parity/asymmetry dilemma. *Sex Roles*, 62: 201–211 DOI 10.1007/s11199-009-9717-2
- Statista (2019). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019, Internet, Social Media & User-Generated Content. Disponibile all'indirizzo: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- Swanson, J. (2002). Acid attacks: Bangladesh's efforts to stop the violence, in *Harvard Health Policy Review*, vol. 3, n. 1, pp. 1-4.
- Tamale, S. (2011). *African Sexualities, A reader*. Cape Town: Pambazuka Press.
- Tan, L.W., Na, J.C., Theng, Y.L., Chang, K. (2011). Sentence-level sentiment polarity classification using a linguistic approach, In C. Xing, F. Crestani, & A. Rauber (Eds.), *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation* (pp. 77–87), Berlin Heidelberg: Springer-Verlag.
- Taylor, Mill, A. (1851). *Enfranchisement of Women*, *Westminster Review*.
- Temple Lang, CRAN team (2004). RCurl package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/RCurl/RCurl.pdf>.
- Temple Lang, D., Wallace, J. (2010). RJSONIO package. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/RJSONIO/RJSONIO.pdf>.

- Thorpe, R.S.M., Black, H., Malhotra, A. (1996) Matrix correspondence of the DNA phylogeny of the Tenerife Lacertid elucidated both historical causes and morphological adaptation, *Syst Biol* 45:335–343.
- Thurstone, L. L. (1927). Psychophysical analysis. *Amer, J. Psychol.*, 38, 368–389.
- Tipaldo, G. (2007). L'analisi del contenuto nella ricerca sociale. Spunti per una riflessione multidisciplinare, *Quaderni di Ricerca del Dipartimento di Scienze Sociali dell'Università di Torino*.
- Tipaldo, G. (2014). L'analisi del contenuto e i mass media. Oggetti, metodi e strumenti, Bologna, Il Mulino.
- Tirino, N. (2019). Cambridge Analytica. Il potere segreto, la gestione del consenso e la fine della propaganda, Lecce, Libellula Edizioni.
- Tomic, O., Berget, I., Næs, T. (2015). A comparison of generalised procrustes analysis and multiple factor analysis for projective mapping data, *Food Quality and Preference*, Volume 43, July 2015, pp. 34-46.
- Tong, Z., Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling, *Conference: The Sixth International Conference on Computer Science, Engineering and Information Technology*, pp.201, 210.
- Topper, K. (2001). Not so trifling nuances: Pierre Bourdieu, Symbolic violence and the perversions of democracy, *Constellations*, Vol. 8, No. 1: 30–56.
- Torgerson, W.S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, Vol.17, pp. 401–419.
- Trobia, A. (2012). La ricerca sociale quali-quantitativa. Collana di Sociologia, Franco Angeli, Milano.
- Tryon R.C. (1939). Cluster analysis; correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality, *Ann Arbor, Mich, Edwards Brother, Inc., lithoprinters and Publishers*.
- Tsai, C.T., Roth, D. (2016). Cross-lingual wikification using multilingual embeddings, *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, San Diego, California*, pp. 589–598.
- Tufis, D.I. (2007). Exploiting Aligned Parallel Corpora in Multilingual Studies and Applications, *in International Workshop on Intercultural Collaboration, IWIC 2007: Intercultural Collaboration*, pp. 103-117.
- Tulkens, S., Hilte, L., Lodewyckx, E., Verhoeven, B., Daelemans, W. (2016). A Dictionary-based Approach to Racism Detection in Dutch Social Media, *In Proceedings of the Workshop Text Analytics for Cybersecurity and Online Safety (TA-COS), Portoroz, Slovenia*.
- Turian, J., Lev, R., Bengio, Y. (2010). Word representations: a simple and general method for semi-supervised learning, *In Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics*.

- Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *In Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417–424.
- Tuzzi, A. (2003). *L'analisi del contenuto. Introduzione ai metodi e alle tecniche di ricerca*, Carocci, Roma.
- Twitter Internal Data (2019). The power of Twitter. Disponibile all'indirizzo: <https://business.twitter.com>.
- Twitter Symbols (2019). Disponibile all'indirizzo: <https://it.piliapp.com/twitter-symbols/> (Ultima consultazione: 15 Novembre 2019).
- Ulivieri, S., (2014). *Corpi violati. Condizionamenti educativi e violenze di genere*, Franco Angeli Editore.
- UN Women (2012). *Confrontando la violenza legata alla danza in India: le donne al centro della giustizia*.
- UNICEF (2013). *Female Genital Mutilation/Cutting: A statistical overview and exploration of the dynamics of change*.
- UNICEF, Innocenti Research Centre (2000). *Violence against women and girls*, Innocenti Digest n. 6.
- Unione Europea (2016). Regolamento 2016/679 del Parlamento Europeo e del Consiglio. *Official Journal of European Union, Legislation, Volume 59*. Disponibile all'indirizzo: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN> (Ultima consultazione: 02 Dicembre 2019).
- Ventimiglia, C. (1989). *La differenza negata. Ricerca sulla violenza sessuale in Italia*. Franco Angeli Editore.
- Vichi, M., Kiers, H. A. L. (2001). Factorial k-means analysis for two-way data, *Computational Statistics and Data Analysis*, 37, 49–64.
- Vichi, M., Opitz, O. (ed.1999). *Classification and Data Analysis. Theory and Application*, Springer Verlag, Heidelberg.
- Vig, K. (2014). Corrosive poisons: vitriolage. In Vig, Krishan. *Textbook of forensic medicine and toxicology: principles and practice (5th ed.)*. India: Elsevier. p. 462.
- Vinokourov, A., Shawe-Taylor, J., John, Cristianini, N. (2003). Inferring a Semantic Representation of Text via Cross-Language Correlation Analysis. *Advances in neural information processing systems*.
- Volpato, C. (2011). *Deumanizzazione. Come si legittima la violenza*, Editori Laterza.
- Wakeling, I.N., Raats, M.M., MacFie, H.J.H. (1992). A new significance test for consensus in Generalized Procrustes Analysis, *Journal of Sensory Studies* 7:91-96.
- Walker, L. (1979). *The Battered Woman*. New York: Harper and Row.
- Walker, R. (1992). Becoming the Third Wave, *Liberty Media for Women*, p. 39-41.

- Walton, J. (2019). Twitter vs. Facebook vs. Instagram: What's the Difference? Disponibile all'indirizzo: <https://www.investopedia.com/articles/markets/100215/twitter-vs-facebook-vs-instagram-who-target-audience.asp>.
- Wang, M., Manning, C.D. (2014). Cross-lingual projected expectation regularization for weakly supervised learning, *Transactions of the Association for Computational Linguistics (TACL)*, 2(5):55–66.
- Wang, Z., Liu, E., Sakaji, H., Ito, T. *et al.* (2012). Estimation of Cross-Lingual News Similarities Using Text-Mining Methods, *In Journal of Risk and Financial Management*.
- Ward, J.H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58, 236–244.
- Wasserman, S., Faust, K. (1994). Social network analysis. *Cambridge University Press*.
- We are Social. (2019). Global Digital Report 2019. Disponibile all'indirizzo: <https://wearesocial.com/global-digital-report-2019>, (Ultima consultazione 15 Novembre 2019).
- Weber, R. (1995). Fondamenti di analisi del contenuto, *trad. it. a cura di G. Badami e S. Costantino, Palermo, Sigma Edizioni (edizione originale 1990)*.
- Wei, L., Hindman, D. B. (2011). Does the Digital Divide Matter More? Comparing the Effects of New Media and Old Media Use on the Education-Based Knowledge Gap. *Mass Communication and Society*, 14(2), 216-235.
- Weinman, Lear, M. (1968). *The Second Feminist Wave*. *New York Times*.
- Welbers, K. Van Attevelde, W., Benoit, K. (2017). Text Analysis in R, *Communication Methods and Measures Journal, Volume 11, 2017, Issue 4, pp. 245-265*.
- Welsh, J. (2006). It was like burning in hell: A comprehensive exploration of acid attack violence, *Carolina Papers on International Health. Center for Global Initiatives, University of North Carolina*.
- West, C. & Zimmerman, D. H. (1987). Doing gender. *Gender and Society, Vol. 1, No. 2, pp - 125-151*.
- Westerhuis, J., Coenegracht, P. (1997). Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares, *J Chemom 11:379–392*.
- Westerhuis, J., Kourti, T., MacGregor, J. (1998). Analysis of multiblock and hierarchical PCA and PLS model, *J Chemom 12:301–321*.
- Westerhuis, J., Smilde, A. (2001) Deflation in multiblock PLS, *J Chemom 15:485–493*.
- WHO, World Health Organization, (1997). *Violence Against Women: A priority health issue*.

- Williams, A.A., Langron, S. P. (1984). The Use of Free-Choice Profiling for the Evaluation of Commercial Ports. *Journal of the Science of Food and Agriculture* 35(5): 558-568.
- Wilson, T., Wiebe, J., Hwa., R. (2004). Just how mad are you? Finding strong and weak opinion clauses, in *Proceedings of AAAI-04, 21st Conference of the American Association for Artificial Intelligence, San Jose, US*.
- Wolf, M. (2006). *Teorie delle comunicazioni di massa*, Bompiani, Milano.
- Wolf, N. (1991). *Il Mito Della Bellezza*, Mondadori .
- Wollstonecraft, M. (1792). *A Vindication of the Rights of Woman*, London, Penguin, 1992 .
- Woolf, V. (1929). *Una stanza tutta per sé*. Traduzione italiana di Maura Del Serra, Newton Compton Editori, 1993.
- Woolf, V.(1938). *Three Guineas. Room of One's Own and Three Guineas*, Penguin Book, London-New York, 1992, pp.230-231.
- World Economic Forum (2017). *The Global Gender Gap Report 2017*, disponibile: <https://www.weforum.org/reports/the-global-gender-gap-report-2017>.
- Xiao, M., Guo, Y. (2014). Distributed word representation learning for cross-lingual dependency parsing, In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, Ann Arbor, Michigan, pp. 119–129*.
- Xing, C., Wang, D., Liu, C., Lin, Y. (2015). Normalized word embedding and orthogonal transform for bilingual word translation, In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp- 1006–1011, Denver, Colorado. Association for Computational Linguistics.
- Xu, J.M., Jun, K.S., Zhu, X., Bellmore, A. (2012). Learning from bullying traces in social media, In *Proceedings of the 2012 Conference of the North American chapter of the asso- ciation for computational linguistics: Human language technologies. Association for Computational Linguistics, pp.656–666*.
- Yan, R., Cao, X.B., Li, K. (2009). Dynamic assembly classification algorithm for short text, *Acta Electronica Sinica*, 37, 1019–1024.
- Yu, H., Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences, In *Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language Processing, pp. 129–136*.
- Yule, G. U. (1944). *A statistical study of vocabulary*, Cambridge University Press, Cambridge.
- Zemond Davis, N. (1977). La storia delle donne in transizione: il caso europeo, *Donna e ricerca storica*, 1977, n. 3., pp. 7-33. Nuova DWF. *Donna Woman Femme, Quaderni di studi internazionali sulla donna*, Roma, Coines Edizioni.

- Zhang, L., Rettinger, A. (2013). Semantic Annotation, Analysis and Comparison: A Multilingual and Cross-lingual Text Analytics Toolkit, *In Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 13–16, Gothenburg, Sweden.
- Zhang, L., Rettinger, A., Färber, M., Tadić, M. (2013). A comparative evaluation of cross-lingual text annotation techniques, *International Conference of the Cross-Language Evaluation Forum for European Languages*, pp.124-135, Springer, Berlin, Heidelberg.
- Zhang, M., Liu, Y., Luan, H., Sun, M. (2017). Adversarial training for unsupervised bilingual lexicon induction, *In Proceedings of ACL*, pp. 1959–1970.
- Zhang, Y., Gaddy, D., Barzilay, R., Jaakkola, T. (2016). Ten pairs to tag – multilingual pos tagging via coarse mapping between embeddings, *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*.
- Zipf, G. K. (1935). The psychobiology of language - An introduction to dynamic philology, *Houghton-Mifflin, Boston*, (traduzione francese *La psychobiologie du langage*, Paris, RETZ-CEPL, 1974).
- Zou, W.Y., Socher, R., Cer, D., Manning, C.D. (2013). Bilingual word embeddings for phrase-based machine translation, *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Seattle, Washington, USA*, pp. 1393–1398.

Appendice A

Tabella 1. Punteggio di sintesi della Parità di genere, Gender Equality Index, nei Paesi membri dell'EU, anno 2015.

(Riferimento: Figure 1.2 e 1.3)

time	cod_geo	geo	value	domain
2015	AT	Austria	63.3	Overall Gender Equality Index
2015	BE	Belgium	70.5	Overall Gender Equality Index
2015	BG	Bulgaria	58	Overall Gender Equality Index
2015	CY	Cyprus	55.1	Overall Gender Equality Index
2015	CZ	Czech Republic	53.6	Overall Gender Equality Index
2015	DE	Germany	65.5	Overall Gender Equality Index
2015	DK	Denmark	76,8	Overall Gender Equality Index
2015	EE	Estonia	56.7	Overall Gender Equality Index
2015	EL	Greece	50	Overall Gender Equality Index
2015	ES	Spain	68.3	Overall Gender Equality Index
2015	EU28	European Union (28 countries)	66.2	Overall Gender Equality Index
2015	FI	Finland	73.0	Overall Gender Equality Index
2015	FR	France	72.6	Overall Gender Equality Index
2015	HR	Croatia	53.1	Overall Gender Equality Index
2015	HU	Hungary	50.8	Overall Gender Equality Index
2015	IE	Ireland	69.5	Overall Gender Equality Index
2015	IT	Italy	62.1	Overall Gender Equality Index
2015	LT	Lithuania	56.8	Overall Gender Equality Index
2015	LU	Luxembourg	69.0	Overall Gender Equality Index
2015	LV	Latvia	57.9	Overall Gender Equality Index
2015	MT	Malta	60.1	Overall Gender Equality Index
2015	NL	Netherlands	72.9	Overall Gender Equality Index
2015	PL	Poland	56.8	Overall Gender Equality Index
2015	PT	Portugal	56.0	Overall Gender Equality Index
2015	RO	Romania	52.4	Overall Gender Equality Index
2015	SE	Sweden	82.6	Overall Gender Equality Index
2015	SI	Slovenia	68.4	Overall Gender Equality Index
2015	SK	Slovakia	52.4	Overall Gender Equality Index
2015	UK	United Kingdom	71.5	Overall Gender Equality Index

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 2. Punteggio di sintesi della Parità di genere, Gender Equality Index, nei Paesi membri dell'EU, anno 2005.

(Riferimento: Figura 4)

time	cod_geo	geo	value	domain
2005	AT	Austria	59.5	Overall Gender Equality Index
2005	CY	Cyprus	45.9	Overall Gender Equality Index
2005	CZ	Czech Republic	53.6	Overall Gender Equality Index
2005	DE	Germany	60.0	Overall Gender Equality Index
2005	DK	Denmark	74.6	Overall Gender Equality Index
2005	EE	Estonia	52.2	Overall Gender Equality Index
2005	EL	Greece	46.8	Overall Gender Equality Index
2005	ES	Spain	62.2	Overall Gender Equality Index
2005	EU28	European Union (28 countries)	62.0	Overall Gender Equality Index
2005	FI	Finland	72.0	Overall Gender Equality Index
2005	FR	France	65.2	Overall Gender Equality Index
2005	BE	Belgium	66.0	Overall Gender Equality Index
2005	BG	Bulgaria	56.0	Overall Gender Equality Index
2005	HR	Croatia	50.3	Overall Gender Equality Index
2005	HU	Hungary	49.5	Overall Gender Equality Index
2005	IE	Ireland	61.9	Overall Gender Equality Index
2005	IT	Italy	49.2	Overall Gender Equality Index
2005	LT	Lithuania	55.8	Overall Gender Equality Index
2005	LU	Luxembourg	64.4	Overall Gender Equality Index
2005	LV	Latvia	53.4	Overall Gender Equality Index
2005	MT	Malta	56.0	Overall Gender Equality Index
2005	NL	Netherlands	67.8	Overall Gender Equality Index
2005	PL	Poland	52.4	Overall Gender Equality Index
2005	PT	Portugal	49.9	Overall Gender Equality Index
2005	RO	Romania	49.9	Overall Gender Equality Index
2005	SE	Sweden	78.8	Overall Gender Equality Index
2005	SI	Slovenia	60.8	Overall Gender Equality Index
2005	SK	Slovakia	52.5	Overall Gender Equality Index
2005	UK	United Kingdom	71.2	Overall Gender Equality Index

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 3. Andamento occupazionale per genere (fascia d'età: 20-64 anni), in Europa, dall'anno 2010 all'anno 2017. (Riferimento: Figura 1.5)

European Union		2010Q3	2011Q3	2012Q3
Total		68.6	68.6	68.5
Males		75.2	75.0	74.7
Females		60.1	61.2	62.1
2013Q3	2014Q3	2015Q3	2016Q3	2017Q3
68.5	69.3	70.1	71.2	72.3
74.4	75.1	76.0	77.0	78.1
62.9	62.3	62.2	62.3	62.5

Fonte: Eurostat, LFS, 2017.

Tabella 4. Gap occupazionale di genere (fascia d'età: 20-64 anni) per il 2016, nei Paesi membri, nell'anno 2016. (Riferimento: Figura 1.6)

cod_geo	2016
European Union	11.6
AT	7.8
BE	9.3
BG	7.3
CY	9.7
CZ	16.0
DE	8.2
DK	6.7
EE	8.2
EL	19.0
ES	11.5
FI	3.3
FR	7.5
HR	9.6
HU	14.0
IE	12.1
IT	20.1
LT	1.9
LU	11.0
LV	2.9
MT	27.7
NL	11.0

PL	14.2
PT	6.8
RO	17.6
SE	3.8
SI	6.6
SK	14.2
UK	3.8

Fonte: Eurostat, LFS, 2017.

Tabella 5. Punteggio Gender Equality Index, dominio del lavoro, nei Paesi membri dell'EU, anno 2015.

(Riferimento: Figure 1.7 e 1.8)

time	cod_geo	geo	value	domain
2015	AT	Austria	76.1	Work (Domain score)
2015	BE	Belgium	73.8	Work (Domain score)
2015	BG	Bulgaria	68.6	Work (Domain score)
2015	CY	Cyprus	70.7	Work (Domain score)
2015	CZ	Czech Republic	66.1	Work (Domain score)
2015	DE	Germany	71.4	Work (Domain score)
2015	DK	Denmark	79.2	Work (Domain score)
2015	EE	Estonia	72.1	Work (Domain score)
2015	EL	Greece	64.2	Work (Domain score)
2015	ES	Spain	72.4	Work (Domain score)
2015	EU28	European Union (28 countries)	71.5	Work (Domain score)
2015	FI	Finland	74.7	Work (Domain score)
2015	FR	France	72.1	Work (Domain score)
2015	HR	Croatia	69.4	Work (Domain score)
2015	HU	Hungary	67.2	Work (Domain score)
2015	IE	Ireland	73.9	Work (Domain score)
2015	IT	Italy	62.4	Work (Domain score)
2015	LT	Lithuania	73.2	Work (Domain score)
2015	LU	Luxembourg	74.0	Work (Domain score)
2015	LV	Latvia	73.6	Work (Domain score)
2015	MT	Malta	71.0	Work (Domain score)
2015	NL	Netherlands	76.7	Work (Domain score)
2015	PL	Poland	66.8	Work (Domain score)
2015	PT	Portugal	72.0	Work (Domain score)
2015	RO	Romania	67.1	Work (Domain score)
2015	SE	Sweden	82.6	Work (Domain score)

2015	SI	Slovenia	71.8	Work (Domain score)
2015	SK	Slovakia	65.5	Work (Domain score)
2015	UK	United Kingdom	76.6	Work (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 6. Punteggio Gender Equality Index, dominio del lavoro, nei Paesi membri dell'EU, anno 2005.

(Riferimento: Figura 1.8)

time	cod_geo	geo	value	domain
2005	AT	Austria	73.7	Work (Domain score)
2005	CY	Cyprus	66.3	Work (Domain score)
2005	CZ	Czech Republic	65.3	Work (Domain score)
2005	DE	Germany	68.1	Work (Domain score)
2005	DK	Denmark	78.9	Work (Domain score)
2005	EE	Estonia	71.0	Work (Domain score)
2005	EL	Greece	62.5	Work (Domain score)
2005	ES	Spain	68.1	Work (Domain score)
2005	EU28	European Union (28 countries)	70.0	Work (Domain score)
2005	FI	Finland	74.2	Work (Domain score)
2005	FR	France	70,5	Work (Domain score)
2005	BE	Belgium	71.0	Work (Domain score)
2005	BG	Bulgaria	67.3	Work (Domain score)
2005	HR	Croatia	67.5	Work (Domain score)
2005	HU	Hungary	65.4	Work (Domain score)
2005	IE	Ireland	71.1	Work (Domain score)
2005	IT	Italy	60.8	Work (Domain score)
2005	LT	Lithuania	71.9	Work (Domain score)
2005	LU	Luxembourg	68.1	Work (Domain score)
2005	LV	Latvia	71.7	Work (Domain score)
2005	MT	Malta	60.8	Work (Domain score)
2005	NL	Netherlands	74.8	Work (Domain score)
2005	PL	Poland	65.2	Work (Domain score)
2005	PT	Portugal	70.6	Work (Domain score)
2005	RO	Romania	68.6	Work (Domain score)
2005	SE	Sweden	78.7	Work (Domain score)
2005	SI	Slovenia	71.2	Work (Domain score)
2005	SK	Slovakia	65.3	Work (Domain score)
2005	UK	United Kingdom	74.2	Work (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database

Tabella 7. Punteggio Gender Equality Index, dominio del denaro, nei Paesi membri dell'EU, anno 2015.
(Riferimento: Figure 9 e 10)

time	cod_geo	geo	value	domain
2015	AT	Austria	85.9	Money (Domain score)
2015	BE	Belgium	87.5	Money (Domain score)
2015	BG	Bulgaria	61.9	Money (Domain score)
2015	CY	Cyprus	79.2	Money (Domain score)
2015	CZ	Czech Republic	75.9	Money (Domain score)
2015	DE	Germany	84.2	Money (Domain score)
2015	DK	Denmark	86.6	Money (Domain score)
2015	EE	Estonia	66.7	Money (Domain score)
2015	EL	Greece	70.7	Money (Domain score)
2015	ES	Spain	75.9	Money (Domain score)
2015	EU28	European Union (28 countries)	79.6	Money (Domain score)
2015	FI	Finland	86.4	Money (Domain score)
2015	FR	France	86.1	Money (Domain score)
2015	HR	Croatia	69.9	Money (Domain score)
2015	HU	Hungary	70.7	Money (Domain score)
2015	IE	Ireland	84.7	Money (Domain score)
2015	IT	Italy	78.6	Money (Domain score)
2015	LT	Lithuania	65.6	Money (Domain score)
2015	LU	Luxembourg	94.4	Money (Domain score)
2015	LV	Latvia	64.3	Money (Domain score)
2015	MT	Malta	82.4	Money (Domain score)
2015	NL	Netherlands	86.8	Money (Domain score)
2015	PL	Poland	73.3	Money (Domain score)
2015	PT	Portugal	70.9	Money (Domain score)
2015	RO	Romania	59.4	Money (Domain score)
2015	SE	Sweden	87.5	Money (Domain score)
2015	SI	Slovenia	81.6	Money (Domain score)
2015	SK	Slovakia	74.0	Money (Domain score)
2015	UK	United Kingdom	81.2	Money (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 8. Punteggio Gender Equality Index, dominio del denaro, nei Paesi membri dell'EU, anno 2005.

(Riferimento: Figura 1.10)

time	cod_geo	geo	value	domain
2005	AT	Austria	82.5	Money (Domain score)
2005	CY	Cyprus	72.6	Money (Domain score)
2005	CZ	Czech Republic	70.2	Money (Domain score)
2005	DE	Germany	83.3	Money (Domain score)
2005	DK	Denmark	82.7	Money (Domain score)
2005	EE	Estonia	58.4	Money (Domain score)
2005	EL	Greece	71.9	Money (Domain score)
2005	ES	Spain	73.6	Money (Domain score)
2005	EU28	European Union (28 countries)	73.9	Money (Domain score)
2005	FI	Finland	80.1	Money (Domain score)
2005	FR	France	81.6	Money (Domain score)
2005	BE	Belgium	81.3	Money (Domain score)
2005	BG	Bulgaria	54.3	Money (Domain score)
2005	HR	Croatia	68.6	Money (Domain score)
2005	HU	Hungary	66.5	Money (Domain score)
2005	IE	Ireland	79.5	Money (Domain score)
2005	IT	Italy	76,2	Money (Domain score)
2005	LT	Lithuania	57.0	Money (Domain score)
2005	LU	Luxembourg	93.1	Money (Domain score)
2005	LV	Latvia	56.3	Money (Domain score)
2005	MT	Malta	70.3	Money (Domain score)
2005	NL	Netherlands	82.2	Money (Domain score)
2005	PL	Poland	61.4	Money (Domain score)
2005	PT	Portugal	68.8	Money (Domain score)
2005	RO	Romania	53.2	Money (Domain score)
2005	SE	Sweden	84.1	Money (Domain score)
2005	SI	Slovenia	77.7	Money (Domain score)
2005	SK	Slovakia	61.5	Money (Domain score)
2005	UK	United Kingdom	79.7	Money (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 9. Punteggio Gender Equality Index, dominio dell'istruzione, nei Paesi membri dell'EU, anno 2015.

(Riferimento: Figure 11 e 12)

time	cod_geo	geo	value	domain
2015	AT	Austria	63.2	Knowledge (Domain score)
2015	BE	Belgium	71.1	Knowledge (Domain score)
2015	BG	Bulgaria	53.3	Knowledge (Domain score)
2015	CY	Cyprus	58.5	Knowledge (Domain score)
2015	CZ	Czech Republic	57.3	Knowledge (Domain score)
2015	DE	Germany	52.9	Knowledge (Domain score)
2015	DK	Denmark	73.6	Knowledge (Domain score)
2015	EE	Estonia	53.2	Knowledge (Domain score)
2015	EL	Greece	55.6	Knowledge (Domain score)
2015	ES	Spain	65.3	Knowledge (Domain score)
2015	EU28	European Union (28 countries)	63.4	Knowledge (Domain score)
2015	FI	Finland	61.3	Knowledge (Domain score)
2015	FR	France	66.1	Knowledge (Domain score)
2015	HR	Croatia	49.8	Knowledge (Domain score)
2015	HU	Hungary	56.9	Knowledge (Domain score)
2015	IE	Ireland	66.4	Knowledge (Domain score)
2015	IT	Italy	61.4	Knowledge (Domain score)
2015	LT	Lithuania	55.8	Knowledge (Domain score)
2015	LU	Luxembourg	69.4	Knowledge (Domain score)
2015	LV	Latvia	48.9	Knowledge (Domain score)
2015	MT	Malta	65.2	Knowledge (Domain score)
2015	NL	Netherlands	67.3	Knowledge (Domain score)
2015	PL	Poland	56.0	Knowledge (Domain score)
2015	PT	Portugal	54.8	Knowledge (Domain score)
2015	RO	Romania	51.8	Knowledge (Domain score)
2015	SE	Sweden	72.8	Knowledge (Domain score)
2015	SI	Slovenia	55.0	Knowledge (Domain score)
2015	SK	Slovakia	60.0	Knowledge (Domain score)
2015	UK	United Kingdom	71.8	Knowledge (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 10. Punteggio Gender Equality Index, dominio dell'istruzione, nei Paesi membri dell'EU, anno 2005.

(Riferimento: Figura 1.12)

time	cod_geo	geo	value	domain
2005	AT	Austria	58.9	Knowledge (Domain score)
2005	CZ	Czech Republic	52.2	Knowledge (Domain score)
2005	DE	Germany	55.3	Knowledge (Domain score)
2005	DK	Denmark	73.7	Knowledge (Domain score)
2005	EE	Estonia	49.5	Knowledge (Domain score)
2005	EL	Greece	47.2	Knowledge (Domain score)
2005	ES	Spain	59.3	Knowledge (Domain score)
2005	EU28	European Union (28 countries)	60.8	Knowledge (Domain score)
2005	FI	Finland	56.6	Knowledge (Domain score)
2005	FR	France	62.3	Knowledge (Domain score)
2005	BE	Belgium	68.1	Knowledge (Domain score)
2005	BG	Bulgaria	52.5	Knowledge (Domain score)
2005	CY	Cyprus	43.4	Knowledge (Domain score)
2005	HR	Croatia	43.6	Knowledge (Domain score)
2005	HU	Hungary	56.9	Knowledge (Domain score)
2005	IE	Ireland	60.8	Knowledge (Domain score)
2005	IT	Italy	54.1	Knowledge (Domain score)
2005	LT	Lithuania	55.1	Knowledge (Domain score)
2005	LU	Luxembourg	62.0	Knowledge (Domain score)
2005	LV	Latvia	46.6	Knowledge (Domain score)
2005	MT	Malta	62.4	Knowledge (Domain score)
2005	NL	Netherlands	63.9	Knowledge (Domain score)
2005	PL	Poland	56.7	Knowledge (Domain score)
2005	PT	Portugal	48.6	Knowledge (Domain score)
2005	RO	Romania	47.9	Knowledge (Domain score)
2005	SE	Sweden	68.1	Knowledge (Domain score)
2005	SI	Slovenia	52.1	Knowledge (Domain score)
2005	SK	Slovakia	54.5	Knowledge (Domain score)
2005	UK	United Kingdom	75.8	Knowledge (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 11. Andamento del livello di istruzione terziaria, per genere (fascia di età: 30-34 anni), in Europa, dall'anno 2006 all'anno 2016. (Riferimento: Figura 1.13)

European Union	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Males	26.3	27.2	28.0	29.0	30.3	31.0	31.8	32.8	33.6	34.0	34.4
Females	31.6	32.9	34.3	35.7	37.3	38.6	40.2	41.4	42.3	43.4	43.9

Fonte: Eurostat, 2017

Tabella 12. Punteggio Gender Equality Index, dominio del tempo, nei Paesi membri dell'EU, anno 2015. (Riferimento: Figure 14 e 15)

time	cod_geo	geo	value	domain
2015	AT	Austria	61.2	Time (Domain score)
2015	BE	Belgium	65.3	Time (Domain score)
2015	BG	Bulgaria	42.7	Time (Domain score)
2015	CY	Cyprus	51.3	Time (Domain score)
2015	CZ	Czech Republic	57.3	Time (Domain score)
2015	DE	Germany	65.0	Time (Domain score)
2015	DK	Denmark	83.1	Time (Domain score)
2015	EE	Estonia	74.7	Time (Domain score)
2015	EL	Greece	44.7	Time (Domain score)
2015	ES	Spain	64.0	Time (Domain score)
2015	EU28	European Union (28 countries)	65.7	Time (Domain score)
2015	FI	Finland	77.4	Time (Domain score)
2015	FR	France	67.3	Time (Domain score)
2015	HR	Croatia	51.0	Time (Domain score)
2015	HU	Hungary	54.3	Time (Domain score)
2015	IE	Ireland	74.2	Time (Domain score)
2015	IT	Italy	59.3	Time (Domain score)
2015	LT	Lithuania	50.6	Time (Domain score)
2015	LU	Luxembourg	69.1	Time (Domain score)
2015	LV	Latvia	65.8	Time (Domain score)
2015	MT	Malta	64.2	Time (Domain score)
2015	NL	Netherlands	83.9	Time (Domain score)
2015	PL	Poland	52.5	Time (Domain score)
2015	PT	Portugal	47.5	Time (Domain score)
2015	RO	Romania	50.3	Time (Domain score)

2015	SE	Sweden	90.1	Time (Domain score)
2015	SI	Slovenia	72.9	Time (Domain score)
2015	SK	Slovakia	46.3	Time (Domain score)
2015	UK	United Kingdom	69.9	Time (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 13. Punteggio Gender Equality Index, dominio del tempo, nei Paesi membri dell'EU, anno 2005.
(Riferimento: Figura 1.15)

time	cod_geo	geo	value	domain
2005	AT	Austria	60.2	Time (Domain score)
2005	CY	Cyprus	47.7	Time (Domain score)
2005	CZ	Czech Republic	51.2	Time (Domain score)
2005	DE	Germany	66.6	Time (Domain score)
2005	DK	Denmark	82.7	Time (Domain score)
2005	EE	Estonia	74.6	Time (Domain score)
2005	EL	Greece	46.2	Time (Domain score)
2005	ES	Spain	58.0	Time (Domain score)
2005	EU28	European Union (28 countries)	66.7	Time (Domain score)
2005	FI	Finland	81.6	Time (Domain score)
2005	FR	France	69.1	Time (Domain score)
2005	BE	Belgium	74.3	Time (Domain score)
2005	BG	Bulgaria	50.9	Time (Domain score)
2005	HR	Croatia	48.3	Time (Domain score)
2005	HU	Hungary	61.1	Time (Domain score)
2005	IE	Ireland	74.2	Time (Domain score)
2005	IT	Italy	60.1	Time (Domain score)
2005	LT	Lithuania	53.5	Time (Domain score)
2005	LU	Luxembourg	73.2	Time (Domain score)
2005	LV	Latvia	59.1	Time (Domain score)
2005	MT	Malta	60.8	Time (Domain score)
2005	NL	Netherlands	86.4	Time (Domain score)
2005	PL	Poland	54.6	Time (Domain score)
2005	PT	Portugal	47.3	Time (Domain score)
2005	RO	Romania	48.9	Time (Domain score)
2005	SE	Sweden	89.6	Time (Domain score)
2005	SI	Slovenia	73.4	Time (Domain score)
2005	SK	Slovakia	55.3	Time (Domain score)

2005	UK	United Kingdom	69.4	Time (Domain score)
------	----	----------------	------	---------------------

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 14. Punteggio Gender Equality Index, dominio del potere, nei Paesi membri dell'EU, anno 2015.
(Riferimento: Figure 1.16 e 1.17)

time	cod_geo	geo	value	domain
2015	AT	Austria	34.9	Power (Domain score)
2015	BE	Belgium	53.4	Power (Domain score)
2015	BG	Bulgaria	56.0	Power (Domain score)
2015	CY	Cyprus	24.7	Power (Domain score)
2015	CZ	Czech Republic	22.6	Power (Domain score)
2015	DE	Germany	53.0	Power (Domain score)
2015	DK	Denmark	61.5	Power (Domain score)
2015	EE	Estonia	28.2	Power (Domain score)
2015	EL	Greece	21.7	Power (Domain score)
2015	ES	Spain	57.0	Power (Domain score)
2015	EU28	European Union (28 countries)	48.5	Power (Domain score)
2015	FI	Finland	65.3	Power (Domain score)
2015	FR	France	68.2	Power (Domain score)
2015	HR	Croatia	28.5	Power (Domain score)
2015	HU	Hungary	18.7	Power (Domain score)
2015	IE	Ireland	48.6	Power (Domain score)
2015	IT	Italy	45.3	Power (Domain score)
2015	LT	Lithuania	36.6	Power (Domain score)
2015	LU	Luxembourg	43.5	Power (Domain score)
2015	LV	Latvia	39.0	Power (Domain score)
2015	MT	Malta	27.4	Power (Domain score)
2015	NL	Netherlands	52.9	Power (Domain score)
2015	PL	Poland	35.1	Power (Domain score)
2015	PT	Portugal	33.9	Power (Domain score)
2015	RO	Romania	33.2	Power (Domain score)
2015	SE	Sweden	79.5	Power (Domain score)
2015	SI	Slovenia	60.6	Power (Domain score)
2015	SK	Slovakia	23.1	Power (Domain score)
2015	UK	United Kingdom	53.0	Power (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 15. Punteggio Gender Equality Index, dominio del potere, nei Paesi membri dell'EU, anno 2005.

(Riferimento: Figura 1.17)

time	cod_geo	geo	value	domain
2005	AT	Austria	29.5	Power (Domain score)
2005	CY	Cyprus	16.4	Power (Domain score)
2005	CZ	Czech Republic	29.6	Power (Domain score)
2005	DE	Germany	34.0	Power (Domain score)
2005	DK	Denmark	54.7	Power (Domain score)
2005	EE	Estonia	22.5	Power (Domain score)
2005	EL	Greece	18.2	Power (Domain score)
2005	ES	Spain	45,9	Power (Domain score)
2005	EU28	European Union (28 countries)	38.9	Power (Domain score)
2005	FI	Finland	68.4	Power (Domain score)
2005	FR	France	43.6	Power (Domain score)
2005	BE	Belgium	39.8	Power (Domain score)
2005	BG	Bulgaria	48.4	Power (Domain score)
2005	HR	Croatia	27.4	Power (Domain score)
2005	HU	Hungary	16.3	Power (Domain score)
2005	IE	Ireland	32.1	Power (Domain score)
2005	IT	Italy	16.1	Power (Domain score)
2005	LT	Lithuania	37.3	Power (Domain score)
2005	LU	Luxembourg	36.2	Power (Domain score)
2005	LV	Latvia	34.8	Power (Domain score)
2005	MT	Malta	27.8	Power (Domain score)
2005	NL	Netherlands	40.3	Power (Domain score)
2005	PL	Poland	26,3	Power (Domain score)
2005	PT	Portugal	22.2	Power (Domain score)
2005	RO	Romania	30.7	Power (Domain score)
2005	SE	Sweden	74.1	Power (Domain score)
2005	SI	Slovenia	36.5	Power (Domain score)
2005	SK	Slovakia	26.9	Power (Domain score)
2005	UK	United Kingdom	51.4	Power (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 16. Punteggio Gender Equality Index, Decision Making Section, Percentuale di potere di uomini e donne nell'indagine dei settori: Ricerca, Media e Sport. (Riferimento: Figura 1.18)

Settore Ricerca		Settore Media			
European Union	2016	2014	2015	2016	Average
Males	65.7	70	67.6	65.7	68
Females	34.3	30	32.4	34.3	32

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 17. Punteggio Gender Equality Index, Decision Making Section, Proporzione di donne al Parlamento (%), nei Paesi membri dell'EU, anno 2017. (Riferimento: Figura 1.19)

time	cod_geo	geo	value
2017	AT	Austria	31.1
2017	CZ	Czech Republic	22.0
2017	DE	Germany	30.9
2017	DK	Denmark	37.4
2017	EE	Estonia	27.7
2017	EL	Greece	18.3
2017	ES	Spain	40.6
2017	EU28	European Union (28 countries)	29.3
2017	FI	Finland	42.0
2017	FR	France	26.9
2017	BE	Belgium	38.7
2017	BG	Bulgaria	25.9
2017	CY	Cyprus	17.9
2017	HR	Croatia	18.0
2017	HU	Hungary	10.1
2017	IE	Ireland	22.2
2017	IT	Italy	31.0
2017	LT	Lithuania	20.6
2017	LU	Luxembourg	28.3
2017	LV	Latvia	17.0
2017	MT	Malta	14.5
2017	NL	Netherlands	35.3
2017	PL	Poland	28.5
2017	PT	Portugal	35.2
2017	RO	Romania	21.3

2017	SE	Sweden	46.1
2017	SI	Slovenia	35.6
2017	SK	Slovakia	21.3
2017	UK	United Kingdom	32.0

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 18. Punteggio Gender Equality Index, Decision Making Section, Proporzione di donne al Governo (%), nei Paesi membri dell'EU, anno 2017. (Riferimento: Figura 1.20)

time	cod_geo	geo	value
2017	AT	Austria	21.4
2017	CZ	Czech Republic	11.8
2017	DE	Germany	46.2
2017	DK	Denmark	40.9
2017	EE	Estonia	26.7
2017	EL	Greece	17.4
2017	ES	Spain	35.7
2017	EU28	European Union (28 countries)	27.7
2017	FI	Finland	35.3
2017	FR	France	50.0
2017	BE	Belgium	21.4
2017	BG	Bulgaria	23.8
2017	CY	Cyprus	8.3
2017	HR	Croatia	28.6
2017	HU	Hungary	0.0
2017	IE	Ireland	26.7
2017	IT	Italy	27.8
2017	LT	Lithuania	20.0
2017	LU	Luxembourg	20.0
2017	LV	Latvia	21.4
2017	MT	Malta	13.3
2017	NL	Netherlands	37.5
2017	PL	Poland	27.3
2017	PT	Portugal	16.7
2017	RO	Romania	21.4
2017	SE	Sweden	52.2
2017	SI	Slovenia	47.1
2017	SK	Slovakia	20.0
2017	UK	United Kingdom	26.1

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 19. Punteggio Gender Equality Index, Decision Making Section, Proporzione di donne al Parlamento e al Governo, nei Paesi membri dell'EU, dall'anno 2004 all'anno 2017. (Riferimento: Figura 1.21)

EU28		2004	2005	2006	2007	2008
PARLIAMENTS		22.1	22.6	22.8	23.7	23.9
GOVERNMENTS		21.2	22.1	23.9	24.5	25.7
2009	2010	2011	2012	2013	2014	2015
24.2	24.4	24.2	26.1	27.4	27.7	28.7
27	26.2	24.2	26.4	26.6	28.3	27.6

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 20. Punteggio Gender Equality Index, Decision Making Section, Proporzione di donne a capo di aziende quotate in borsa (%), nei Paesi membri dell'EU, anno 2017. (Riferimento: Figura 1.22)

time	cod_geo	geo	value
2017	AT	Austria	19.2
2017	CZ	Czech Republic	14.5
2017	DE	Germany	31.9
2017	DK	Denmark	30.3
2017	EE	Estonia	7.4
2017	EL	Greece	11.3
2017	ES	Spain	22.0
2017	EU28	European Union (28 countries)	25.3
2017	FI	Finland	32.0
2017	FR	France	43.4
2017	BE	Belgium	30.7
2017	BG	Bulgaria	17.1
2017	CY	Cyprus	10.4
2017	HR	Croatia	21.6
2017	HU	Hungary	14.5
2017	IE	Ireland	17.6
2017	IT	Italy	34.0
2017	LT	Lithuania	14.3
2017	LU	Luxembourg	12.0
2017	LV	Latvia	28.8
2017	MT	Malta	8.4
2017	NL	Netherlands	29.5
2017	PL	Poland	20.1
2017	PT	Portugal	16.2

2017	RO	Romania	11.0
2017	SE	Sweden	35.9
2017	SI	Slovenia	22.6
2017	SK	Slovakia	15.1
2017	UK	United Kingdom	27.2

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 21. Punteggio Gender Equality Index, Decision Making Section, Proporzione di donne nella leadership di aziende quotate in borsa (%), in Europa, dall'anno 2003 all'anno 2017. (Riferimento:

Figura 1.23)

EU28	2003	2004	2005	2006	2007	2008	2009
CEO	NA	NA	NA	NA	NA	NA	NA
BOARD CHAIRS	1.6	2.6	3.1	3.7	3.5	2.9	3.1
2010	2011	2012	2013	2014	2015	2016	2017
NA	NA	2.5	2.8	3.3	4.3	5.9	5.5
3.4	2.8	3.3	4.8	6.4	6.4	7.5	7.1

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 22. Punteggio Gender Equality Index, dominio della salute, nei Paesi membri dell'EU, anno 2015.

(Riferimento: Figure 1.24 e 1.25)

time	cod_geo	geo	value	domain
2015	AT	Austria	91.7	Health (Domain score)
2015	BE	Belgium	86.3	Health (Domain score)
2015	BG	Bulgaria	76.4	Health (Domain score)
2015	CY	Cyprus	88.2	Health (Domain score)
2015	CZ	Czech Republic	86.0	Health (Domain score)
2015	DE	Germany	90.5	Health (Domain score)
2015	DK	Denmark	89.6	Health (Domain score)
2015	EE	Estonia	81.5	Health (Domain score)
2015	EL	Greece	83.1	Health (Domain score)
2015	ES	Spain	89.6	Health (Domain score)
2015	EU28	European Union (28 countries)	87.4	Health (Domain score)
2015	FI	Finland	89.7	Health (Domain score)
2015	FR	France	87.1	Health (Domain score)
2015	HR	Croatia	83.3	Health (Domain score)
2015	HU	Hungary	86.0	Health (Domain score)
2015	IE	Ireland	90.6	Health (Domain score)
2015	IT	Italy	86.3	Health (Domain score)

2015	LT	Lithuania	79.1	Health (Domain score)
2015	LU	Luxembourg	89.0	Health (Domain score)
2015	LV	Latvia	78.4	Health (Domain score)
2015	MT	Malta	91.8	Health (Domain score)
2015	NL	Netherlands	89.9	Health (Domain score)
2015	PL	Poland	82.2	Health (Domain score)
2015	PT	Portugal	83.6	Health (Domain score)
2015	RO	Romania	70.4	Health (Domain score)
2015	SE	Sweden	94.1	Health (Domain score)
2015	SI	Slovenia	87.7	Health (Domain score)
2015	SK	Slovakia	85.3	Health (Domain score)
2015	UK	United Kingdom	93.1	Health (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 23. Punteggio Gender Equality Index, dominio della salute, nei Paesi membri dell'EU, anno 2005.

(Riferimento: Figura 1.25)

time	cod_geo	geo	value	domain
2005	AT	Austria	91.4	Health (Domain score)
2005	CZ	Czech Republic	84.6	Health (Domain score)
2005	DE	Germany	86.6	Health (Domain score)
2005	DK	Denmark	91.1	Health (Domain score)
2005	EE	Estonia	81.0	Health (Domain score)
2005	EL	Greece	84.6	Health (Domain score)
2005	ES	Spain	88.1	Health (Domain score)
2005	EU28	European Union (28 countries)	85.9	Health (Domain score)
2005	FI	Finland	89.2	Health (Domain score)
2005	FR	France	86.9	Health (Domain score)
2005	BE	Belgium	86.3	Health (Domain score)
2005	BG	Bulgaria	72.6	Health (Domain score)
2005	CY	Cyprus	85.8	Health (Domain score)
2005	HR	Croatia	81.4	Health (Domain score)
2005	HU	Hungary	82.4	Health (Domain score)
2005	IE	Ireland	90.4	Health (Domain score)
2005	IT	Italy	85.8	Health (Domain score)
2005	LT	Lithuania	77.6	Health (Domain score)
2005	LU	Luxembourg	89.2	Health (Domain score)
2005	LV	Latvia	73.8	Health (Domain score)
2005	MT	Malta	90.7	Health (Domain score)

2005	NL	Netherlands	89.7	Health (Domain score)
2005	PL	Poland	80.6	Health (Domain score)
2005	PT	Portugal	83.8	Health (Domain score)
2005	RO	Romania	69.5	Health (Domain score)
2005	SE	Sweden	91.7	Health (Domain score)
2005	SI	Slovenia	86.3	Health (Domain score)
2005	SK	Slovakia	83.5	Health (Domain score)
2005	UK	United Kingdom	93.1	Health (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 24. Quota percentuale della prevalenza di Femminicidi commessi dal partner della vittima, nei Paesi membri dell'UE, dall'anno 2008 all'anno 2015. (Riferimento: Figure 1.27 e 1.28)

GEO	2008	2009	2010	2011	2012	2013	2014	2015
CZ	0.32	0.28	0.49	0.37	0.30	0.39	0.65	0.45
DE	NA	NA	NA	0.39	0.27	0.35	0.41	0.34
ES	NA	NA	NA	NA	NA	0.23	0.24	0.25
FR	NA	NA	NA	NA	NA	NA	NA	0.25
HR	NA	NA	NA	NA	0.32	0.36	0.36	0.50
IT	0.19	0.26	0.16	0.23	0.24	0.25	0.26	0.22
LV	NA	NA	NA	NA	0.18	0.27	0.18	0.47
LT	1.04	1.11	1.24	0.97	0.68	1.12	0.94	0.38
HU	NA	0.59	0.46	0.36	0.54	0.52	0.46	0.37
MT	NA	NA	NA	NA	0.95	0.00	0.47	0.47
NL	0.20	0.20	0.36	0.30	0.25	0.32	0.19	0.26
SI	0.29	0.19	0.19	0.39	0.19	0.10	0.38	0.19
SK	NA	NA	NA	0.33	0.11	NA	0.00	0.22
FI	0.92	0.88	0.62	0.77	0.58	0.51	0.54	0.29
UKC-L	0.37	0.34	0.35	0.31	0.27	0.30	0.28	NA
UKM	0.56	0.30	0.44	0.33	0.18	0.29	0.22	NA
IS	NA	NA	0.00	0.00	0.63	0.00	1.23	0.00
LI	NA	NA	NA	NA	NA	NA	NA	0.00
CH	NA	0.28	0.38	0.40	0.25	0.34	0.32	0.36
ME	NA	NA	NA	NA	NA	NA	NA	0.64
MK	NA	NA	NA	NA	0.19	NA	NA	NA
AL	NA	NA	NA	NA	NA	NA	NA	NA
BA	NA	NA	NA	NA	0.15	NA	NA	NA

Fonte: Eurostat, 2017

Tabella 25. Quota percentuale della prevalenza di Femminicidi commessi da un familiare della vittima, nei Paesi membri dell'UE, dall'anno 2008 all'anno 2015. (Riferimento: Figure 1.27 e 1.28)

GEO	2008	2009	2010	2011	2012	2013	2014	2015
CZ	0.21	0.13	0.19	0.19	0.28	0.24	0.34	0.26
DE	NA	NA	NA	0.17	0.17	0.19	0.18	0.17
ES	NA	NA	NA	NA	NA	0.08	0.05	0.07
FR	NA	NA	NA	NA	NA	NA	NA	0.16
HR	NA	NA	NA	NA	0.23	0.23	0.27	0.09
IT	0.17	0.17	0.20	0.16	0.12	0.16	0.11	0.12
LV	NA	NA	NA	NA	0.27	0.18	1.38	0.37
LT	0.52	0.41	0.65	0.43	0.37	0.25	0.25	0.51
HU	NA	0.13	0.11	0.10	0.19	0.31	0.25	0.25
MT	NA	NA	NA	NA	0.00	0.00	0.00	0.00
NL	0.14	0.22	0.12	0.12	0.06	0.09	0.06	0.13
SI	0.10	0.29	0.10	0.00	0.29	0.10	0.00	0.10
SK	NA	NA	NA	0.00	0.07	NA	0.04	0.04
FI	0.11	0.11	0.15	0.26	0.51	0.11	0.14	0.11
UKC-L	0.11	0.15	0.14	0.11	0.15	0.18	0.11	NA
UKM	0.11	0.15	0.15	0.11	0.07	0.04	0.15	NA
IS	NA	NA	0.00	0.63	0.00	0.00	0.00	0.00
LI	NA	NA	NA	NA	NA	NA	NA	0.00
CH	NA	0.13	0.10	0.13	0.10	0.10	0.10	0.17
ME	NA	NA	NA	NA	NA	NA	NA	0.32
MK	NA	NA	NA	NA	0.19	NA	NA	NA
AL	NA	NA	NA	NA	NA	NA	NA	NA
BA	NA	NA	NA	NA	0.46	NA	NA	NA

Fonte: Eurostat, 2017

Tabella 26. Punteggio Gender Equality Index, dominio della violenza in generale, nei Paesi membri dell'EU, anno 2012. (Riferimento: Figure 1.29 e 1.30)

time	cod_geo	geo	value	domain
2012	AT	Austria	25.6	Violence (Domain score)
2012	BE	Belgium	29.7	Violence (Domain score)
2012	BG	Bulgaria	44.2	Violence (Domain score)
2012	CY	Cyprus	24.7	Violence (Domain score)
2012	CZ	Czech Republic	28.0	Violence (Domain score)
2012	DE	Germany	26.9	Violence (Domain score)

2012	DK	Denmark	32.6	Violence (Domain score)
2012	EE	Estonia	25.8	Violence (Domain score)
2012	EL	Greece	27.4	Violence (Domain score)
2012	ES	Spain	25.2	Violence (Domain score)
2012	EU28	European Union (28 countries)	27.5	Violence (Domain score)
2012	FI	Finland	32.4	Violence (Domain score)
2012	FR	France	29.1	Violence (Domain score)
2012	HR	Croatia	23.2	Violence (Domain score)
2012	HU	Hungary	26.7	Violence (Domain score)
2012	IE	Ireland	25.6	Violence (Domain score)
2012	IT	Italy	26.8	Violence (Domain score)
2012	LT	Lithuania	25.0	Violence (Domain score)
2012	LU	Luxembourg	32.3	Violence (Domain score)
2012	LV	Latvia	38.2	Violence (Domain score)
2012	MT	Malta	25.2	Violence (Domain score)
2012	NL	Netherlands	31.5	Violence (Domain score)
2012	PL	Poland	22.1	Violence (Domain score)
2012	PT	Portugal	24.5	Violence (Domain score)
2012	RO	Romania	25.0	Violence (Domain score)
2012	SE	Sweden	29.7	Violence (Domain score)
2012	SI	Slovenia	22.4	Violence (Domain score)
2012	SK	Slovakia	30.0	Violence (Domain score)
2012	UK	United Kingdom	29.0	Violence (Domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 27. Punteggio Gender Equality Index, dominio della violenza: sotto-dominio della prevalenza, nei Paesi membri dell'EU, anno 2012. (Riferimento: Figura 1.31)

time	cod_geo	geo	value	domain
2012	AT	Austria	13.2	Prevalence (Sub-domain score)
2012	BE	Belgium	24.0	Prevalence (Sub-domain score)
2012	BG	Bulgaria	18.7	Prevalence (Sub-domain score)
2012	CY	Cyprus	14.1	Prevalence (Sub-domain score)
2012	CZ	Czech Republic	20.7	Prevalence (Sub-domain score)
2012	DE	Germany	22.4	Prevalence (Sub-domain score)
2012	DK	Denmark	32.4	Prevalence (Sub-domain score)
2012	EE	Estonia	19.9	Prevalence (Sub-domain score)
2012	EL	Greece	17.1	Prevalence (Sub-domain score)
2012	ES	Spain	13.7	Prevalence (Sub-domain score)
2012	EU28	European Union (28 countries)	21.2	Prevalence (Sub-domain score)
2012	FI	Finland	29.2	Prevalence (Sub-domain score)

2012	FR	France	28.3	Prevalence (Sub-domain score)
2012	HR	Croatia	14.0	Prevalence (Sub-domain score)
2012	HU	Hungary	19.4	Prevalence (Sub-domain score)
2012	IE	Ireland	17.7	Prevalence (Sub-domain score)
2012	IT	Italy	17.9	Prevalence (Sub-domain score)
2012	LT	Lithuania	19.3	Prevalence (Sub-domain score)
2012	LU	Luxembourg	23.1	Prevalence (Sub-domain score)
2012	LV	Latvia	23.3	Prevalence (Sub-domain score)
2012	MT	Malta	14.5	Prevalence (Sub-domain score)
2012	NL	Netherlands	28.8	Prevalence (Sub-domain score)
2012	PL	Poland	12.4	Prevalence (Sub-domain score)
2012	PT	Portugal	15.7	Prevalence (Sub-domain score)
2012	RO	Romania	19.1	Prevalence (Sub-domain score)
2012	SE	Sweden	29.2	Prevalence (Sub-domain score)
2012	SI	Slovenia	13.4	Prevalence (Sub-domain score)
2012	SK	Slovakia	22.7	Prevalence (Sub-domain score)
2012	UK	United Kingdom	26.9	Prevalence (Sub-domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 27. Punteggio Gender Equality Index, dominio della violenza: sotto-dominio della severità, nei Paesi membri dell'EU, anno 2012. (Riferimento: Figura 1.32)

time	cod_geo	geo	value	domain
2012	AT	Austria	46.9	Severity (Sub-domain score)
2012	BE	Belgium	50.3	Severity (Sub-domain score)
2012	BG	Bulgaria	65.2	Severity (Sub-domain score)
2012	CY	Cyprus	39.4	Severity (Sub-domain score)
2012	CZ	Czech Republic	45.8	Severity (Sub-domain score)
2012	DE	Germany	47.8	Severity (Sub-domain score)
2012	DK	Denmark	48.5	Severity (Sub-domain score)
2012	EE	Estonia	42.6	Severity (Sub-domain score)
2012	EL	Greece	42.9	Severity (Sub-domain score)
2012	ES	Spain	42.5	Severity (Sub-domain score)
2012	EU28	European Union (28 countries)	46.9	Severity (Sub-domain score)
2012	FI	Finland	53.2	Severity (Sub-domain score)
2012	FR	France	45.2	Severity (Sub-domain score)
2012	HR	Croatia	39.6	Severity (Sub-domain score)
2012	HU	Hungary	46.7	Severity (Sub-domain score)
2012	IE	Ireland	47.2	Severity (Sub-domain score)
2012	IT	Italy	46.6	Severity (Sub-domain score)
2012	LT	Lithuania	44.3	Severity (Sub-domain score)

2012	LU	Luxembourg	54.2	Severity (Sub-domain score)
2012	LV	Latvia	58.4	Severity (Sub-domain score)
2012	MT	Malta	52.1	Severity (Sub-domain score)
2012	NL	Netherlands	52.1	Severity (Sub-domain score)
2012	PL	Poland	40.5	Severity (Sub-domain score)
2012	PT	Portugal	39.0	Severity (Sub-domain score)
2012	RO	Romania	42.0	Severity (Sub-domain score)
2012	SE	Sweden	44.6	Severity (Sub-domain score)
2012	SI	Slovenia	44.6	Severity (Sub-domain score)
2012	SK	Slovakia	48.7	Severity (Sub-domain score)
2012	UK	United Kingdom	50.9	Severity (Sub-domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Tabella 28. Punteggio Gender Equality Index, dominio della violenza: sotto-dominio della divulgazione, nei Paesi membri dell'EU, anno 2012. (Riferimento: Figura 1.33)

time	cod_geo	geo	value	domain
2012	AT	Austria	16.6	Disclosure (Sub-domain score)
2012	BE	Belgium	14.9	Disclosure (Sub-domain score)
2012	BG	Bulgaria	48.6	Disclosure (Sub-domain score)
2012	CY	Cyprus	20.8	Disclosure (Sub-domain score)
2012	CZ	Czech Republic	17.6	Disclosure (Sub-domain score)
2012	DE	Germany	10.6	Disclosure (Sub-domain score)
2012	DK	Denmark	16.7	Disclosure (Sub-domain score)
2012	EE	Estonia	15.0	Disclosure (Sub-domain score)
2012	EL	Greece	22.1	Disclosure (Sub-domain score)
2012	ES	Spain	19.4	Disclosure (Sub-domain score)
2012	EU28	European Union (28 countries)	14.3	Disclosure (Sub-domain score)
2012	FI	Finland	14.8	Disclosure (Sub-domain score)
2012	FR	France	13.7	Disclosure (Sub-domain score)
2012	HR	Croatia	15.9	Disclosure (Sub-domain score)
2012	HU	Hungary	13.9	Disclosure (Sub-domain score)
2012	IE	Ireland	12.0	Disclosure (Sub-domain score)
2012	IT	Italy	15.8	Disclosure (Sub-domain score)
2012	LT	Lithuania	11.2	Disclosure (Sub-domain score)
2012	LU	Luxembourg	19.5	Disclosure (Sub-domain score)
2012	LV	Latvia	32.8	Disclosure (Sub-domain score)
2012	MT	Malta	9.0	Disclosure (Sub-domain score)
2012	NL	Netherlands	13.7	Disclosure (Sub-domain score)

2012	PL	Poland	13.3	Disclosure (Sub-domain score)
2012	PT	Portugal	18.7	Disclosure (Sub-domain score)
2012	RO	Romania	13.9	Disclosure (Sub-domain score)
2012	SE	Sweden	15.4	Disclosure (Sub-domain score)
2012	SI	Slovenia	9.3	Disclosure (Sub-domain score)
2012	SK	Slovakia	18.6	Disclosure (Sub-domain score)
2012	UK	United Kingdom	9.2	Disclosure (Sub-domain score)

Fonte: EIGE, 2017b, Gender Equality Index, Statistics Database.

Appendice B

CLUSTER 1: PROGRESSO STRATEGICO						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
bilanciare_potere	1.0002942	0.18999012	34	644	0.000	7.634721
sfatare	0.5589879	0.07375393	19	250	0.000	6.631754
progresso	0.9708738	0.24191288	33	820	0.000	6.499652
radicale	0.8826125	0.20533093	30	696	0.000	6.441342
pianificazione	0.6178288	0.10709070	21	363	0.000	6.226683
equilibrio	0.7649309	0.18467983	26	626	0.000	5.859075
intuizione	0.7589625	0.182567878	21	525	0.000	4.856986
sviluppare	0.6958695	0.20154872	18	618	0.000	3.258962
CLUSTER 2: RIDUZIONE DELLA DISPARITA'						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
sindaco	1.0700016	0.77117106	474	2614	0.000	7.331249
solidarietà	1.2235039	0.90776334	542	3077	0.000	7.177249
femminista	0.5936929	0.39296093	263	1332	0.000	6.786404
includere	0.8081447	0.57675571	358	1955	0.000	6.542097
morale	0.6975327	0.48795598	309	1654	0.000	6.416991
leghista	0.7043048	0.50831207	312	1723	0.000	5.906636
chiesa	0.7968577	0.58973640	353	1999	0.000	5.820983
legittimare	0.6727014	0.48677592	298	1650	0.000	5.726139
attuale	0.9097271	0.69092679	403	2342	0.000	5.707530
mutamento	0.1106120	0.04749753	49	161	0.000	5.669467
coscienza	0.6365832	0.45786438	282	1552	0.000	5.668217
regolamento	0.7426804	0.55167938	329	1870	0.000	5.550073
reazione	0.6388406	0.46317466	283	1570	0.000	5.545776
uguaglianza	0.5101695	0.35549393	226	1205	0.000	5.533594
battaglia	0.5891781	0.43042792	261	1459	0.000	5.201342
valore	0.9548748	0.74874987	423	2538	0.000	5.190110
media	1.0587146	0.84699010	469	2871	0.000	5.030474
affermazione	0.5124269	0.37083475	227	1257	0.000	4.986176
sviluppo	0.6975327	0.54341894	309	1842	0.000	4.539433
poliziotto	0.7562247	0.60094700	335	2037	0.000	4.363619
raggiungere	0.1918779	0.12586661	85	424	0.000	3.954810
paritario	0.3070047	0.22952220	136	778	0.000	3.465886
luciana_lamorgese	0.2641143	0.19553808	117	661	0.000	3.342567
integrità	0.4108445	0.32342204	182	1095	0.000	3.341355

legge	0.4153593	0.32982756	184	1118	0.000	3.225296
fiducia	0.3860132	0.32363223	171	1097	0.000	2.384939
età	0.3724689	0.31212662	165	1058	0.000	2.347756
decisione	1.0496851	0.95083563	465	3223	0.000	2.247943
battaglia	1.0677442	0.96912661	473	3285	0.000	2.222201
...
parità	1.0678958	0.698596	685	2589	0.000	2.00255

CLUSTER 3: EDUCARE ALL'UGUAGLIANZA DI GENERE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
educare	1.04520325	0.91867892	1607	3114	0.000	6.999077
crescita	1.24813008	1.12725503	1919	3821	0.000	6.044128
posizione	1.17723577	1.06294160	1810	3603	0.000	5.882583
sociale	1.14861789	1.03609517	1766	3512	0.000	5.864818
tentativo	0.73235772	0.64460933	1126	2185	0.000	5.780417
trasformazione_politica	1.16422764	1.05291107	1790	3569	0.000	5.755974
condanna_sicura	0.88715447	0.79793712	1364	2681	0.000	5.728991
solenne	1.40682927	1.28803859	2163	4366	0.000	5.562306
giorgia_meloni	1.34699187	1.23139557	2071	4174	0.000	5.533679
repubblica	1.35024390	1.23611582	2076	4190	0.000	5.453020
partecipazione	1.21105691	1.10424380	1862	3743	0.000	5.394652
dovere	1.12845528	1.02635965	1735	3479	0.000	5.345395
apertura	0.59707317	0.52837314	918	1791	0.000	4.991009
capire	1.05951220	0.96942162	1629	3286	0.000	4.850453
bisogno	1.05560976	0.96824156	1623	3282	0.000	4.706399
origine	0.99967480	0.91985898	1537	3118	0.000	4.408833
protezione	1.09528455	1.01514906	1684	3441	0.000	4.216279
messaggi	0.75707317	0.69926679	1164	2342	0.000	4.206907
pd	1.04390244	0.96617645	1605	3275	0.000	4.190201
spirito	0.86504065	0.79595238	1330	2698	0.000	4.097459
giornalista	0.63284553	0.57469060	973	1948	0.000	4.050088
fisica	1.01593496	0.94346024	1562	3198	0.000	3.952485
passare	0.89495935	0.82751907	1376	2805	0.000	3.923226
relazione	1.14666667	1.07179207	1763	3633	0.000	3.834627
risposta	0.63934959	0.58472114	983	1982	0.000	3.771019
carabiniere	1.03869919	0.96912661	1597	3285	0.000	3.743775
decreto	0.97365854	0.90776334	1497	3077	0.000	3.661654
determinazione	0.84747967	0.78651188	1303	2666	0.000	3.635660
arrivare	0.80845528	0.74933990	1243	2540	0.000	3.610217

richiamare	0.91186992	0.85235248	1402	2882	0.000	3.536670
pena	1.12585366	1.05822135	1731	3587	0.000	3.484430
insistere	1.01398374	0.95835631	1559	3223	0.000	3.429029
soluzione	1.06016260	0.99685808	1630	3379	0.000	3.358247
conoscenza	0.90536585	0.84699010	1392	2871	0.000	3.355283
risalire	0.94829268	0.89389760	1458	3030	0.000	3.043333
differenza	0.71284553	0.66585046	1096	2257	0.000	3.039590
assistenza	0.87284553	0.82397888	1342	2793	0.000	2.844859
strategia	0.61138211	0.57233048	940	1940	0.000	2.719456
salario	0.91577236	0.86793622	1408	2942	0.000	2.713884
pubblicità	0.54439024	0.50831207	837	1723	0.000	2.663117
giudicare	0.37918699	0.34959362	583	1185	0.000	2.626188
canali	1.03154472	0.98240231	1586	3330	0.000	2.622803
sicurezza	0.75772358	0.71688817	1165	2430	0.000	2.544172
web	0.69333333	0.65552491	1066	2222	0.000	2.461031
importante	0.87349593	0.83135427	1343	2818	0.000	2.440436
bruno_vespa	0.69268293	0.65670497	1065	2226	0.000	2.338927
franco_gabrielli	0.78699187	0.74874987	1210	2538	0.000	2.330781
libertà	0.89300813	0.85230038	1373	2889	0.000	2.328023
impegno	0.63869919	0.60537224	982	2052	0.000	2.254387
cultura	0.88585366	0.84758013	1362	2873	0.000	2.193882
...
inculcare	0.7825698	0.6524587	1245	2185	0.000	2.00584

CLUSTER 4: FIDUCIA NELLE ISTITUZIONI

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
fiducia	0.6450992	0.15871845	67	538	0.000	9.551504
istituzione	1.1361448	0.53220834	118	1804	0.000	7.461052
tolleranza	0.6450992	0.22716210	67	770	0.000	7.422656
dialogo	0.9435779	0.47262689	98	1444	0.000	7.084113
sentimentale	0.6258425	0.22952222	65	778	0.000	7.006461
legale	1.1842865	0.60035697	123	2035	0.000	6.899975
matrimonio	0.9339495	0.43750830	97	1483	0.000	6.753268
cooperazione	0.8761795	0.39974629	91	1355	0.000	6.741509
comunitaria	0.1444252	0.05280781	15	179	0.000	3.319007
...
forze	0.478578	0.052689	14	1526	0.000	2.05286

CLUSTER 5: ATTIVISMO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
----------	----------	--------	-------------------	-------------------	---------	--------

agire	1.05057782	0.88740725	910	3008	0.000	5.809906
definire	0.86355188	0.71924830	748	2438	0.000	5.691380
educazione	1.01940683	0.86380600	883	2928	0.000	5.615976
matteo_salvini	0.93512971	0.78916702	810	2675	0.000	5.505915
italia	1.09213914	0.93401974	946	3166	0.000	5.495946
manifestazione	0.88317806	0.74815984	765	2536	0.000	5.230423
normale	0.83815329	0.70774269	726	2399	0.000	5.190627
rapporto	1.01478890	0.87118139	879	2953	0.000	5.167272
confine	0.72039622	0.60271709	624	2043	0.000	5.065732
chiedere	0.84854362	0.72160843	735	2446	0.000	5.007215
visione	0.76657546	0.64726447	664	2194	0.000	4.962444
marito	0.77696579	0.65729500	673	2228	0.000	4.940627
recuperare	0.92012145	0.79347091	797	2679	0.000	4.898991
storia	0.97207310	0.84499760	842	2849	0.000	4.821377
parola	0.86586084	0.74255454	750	2517	0.000	4.799327
tema	0.81737263	0.70243240	708	2381	0.000	4.598467
esprimere	0.66959905	0.56702019	580	1922	0.000	4.554952
padre	0.57493160	0.48117062	498	1631	0.000	4.509825
aiutare	0.64535495	0.54813919	559	1858	0.000	4.390392
nessuna_justificazione	1.02748820	0.90481318	890	3067	0.000	4.341479
campagna	0.99054480	0.87472158	858	2965	0.000	4.168544
risorsa	1.08405777	0.96588143	939	3274	0.000	4.053706
premier_conte	0.69961556	0.60625728	606	2055	0.000	4.018111
accoglienza	0.80813678	0.71334799	700	2418	0.000	3.770919
italiano	0.62226532	0.54468782	539	1832	0.000	3.725552
mondiale	0.60148466	0.52158777	521	1768	0.000	3.702788
corpo	0.58647641	0.5137718	508	1730	0.000	3.565400
sollevare	0.57954952	0.50949213	502	1727	0.000	3.286981
attenzione	0.47333726	0.41567713	410	1409	0.000	2.988861
migliore	0.62688325	0.56495508	543	1915	0.000	2.764982
onore	0.66844457	0.61682522	579	2070	0.000	2.483385
famiglia	1.19373348	1.11692948	1034	3786	0.000	2.459396
carriera	0.83699881	0.78267668	725	2653	0.000	2.068891
...
foto	0.475845	0.589658	625	1854	0.000	2.067859

CLUSTER 6: LEGGE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
eliminazione	1.3975904	0.09853525	29	334	0.000	10.159500

obbligo	1.6385542	0.24250291	34	822	0.000	8.503138
inasprimento	1.6385542	0.27111944	34	919	0.000	8.111580
penale	1.635258	0.1524824	25	985	0.000	7.25690
pena	1.256985	0.2568975	21	825	0.000	6.25896

CLUSTER 7: COOPERAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
obiettivo	1.09155368	0.60979747	351	2067	0.000	10.603712
altruismo	0.85520587	0.44989896	275	1525	0.000	10.265278
stampa	0.84898619	0.46317466	273	1570	0.000	9.690490
parlare	1.00758801	0.58531117	324	1984	0.000	9.545771
scarpe_rosse	0.82099764	0.45784001	264	1528	0.000	9.434220
cambiare	0.65928598	0.33808810	212	1146	0.000	9.330208
garanzia	0.83343699	0.46346968	268	1571	0.000	9.318967
polizia	0.65928598	0.34743114	212	1155	0.000	9.229390
televisione	0.80544844	0.44783385	259	1518	0.000	9.161713
proposta	1.05423560	0.64667444	339	2192	0.000	8.848112
giustizia	0.92984202	0.55403950	299	1878	0.000	8.764814
vicinanza	0.61263839	0.32540233	197	1103	0.000	8.552743
verbale	0.60330887	0.32569734	194	1104	0.000	8.284564
contesto	0.54733176	0.28675527	176	972	0.000	8.240702
fermare	0.93295186	0.59593173	300	2020	0.000	7.666900
permanente	0.49135465	0.26315401	158	892	0.000	7.556820
madre	0.50379400	0.27583969	162	935	0.000	7.401615
bambino	0.89563379	0.57941085	288	1964	0.000	7.308759
pazienza	0.50379400	0.28970543	162	982	0.000	6.838560
femmina	0.47580545	0.26905433	153	912	0.000	6.831378
codice	0.37629058	0.19972564	121	677	0.000	6.689557
donna	0.66861550	0.41833228	215	1418	0.000	6.758023
gerarchia	0.37629058	0.19972564	121	677	0.000	6.689557
...
sostegno	0.3158758	0.254879	125	569	0.000	6.36589

CLUSTER 8: COMUNICAZIONE SOCIALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
comunicazione	0.76420952	0.19972564	48	677	0.000	7.692613
genere	0.71644643	0.17789447	45	603	0.000	7.680344
antiviolenza	0.62092024	0.13629726	39	462	0.000	7.659547
emancipazione	0.66868333	0.15724337	42	533	0.000	7.654856
movimento	0.71644643	0.18113965	45	614	0.000	7.597214

riposizionare	0.65276230	0.15606331	41	529	0.000	7.489666
costruire	0.60499920	0.16166861	38	548	0.000	6.734021
istruire	0.93934087	0.13474314	59	1155	0.000	6.729553
coraggio	0.47763095	0.10738572	30	364	0.000	6.611840
fondamento	0.52539405	0.14337763	33	486	0.000	6.181259
sistema	0.39802579	0.08732465	25	296	0.000	6.103305
permanente	0.76420952	0.28675527	48	972	0.000	5.880330
vincere	0.44578889	0.21271390	28	407	0.000	5.730013
centro	0.28657857	0.05841311	18	198	0.000	5.357614
...
centrale	0.8578895	0.0878586	25	258	0.000	4.236585

Tabella 1. Caratterizzazione dei cluster dei concetti positivi italiani

CLUSTER 1: GARANTISMO GIURIDICO						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
garantismo	1.25699237	1.05508137	1346	5277	0.000	7.128.053
condanna	1.03566459	0.87073704	1109	4355	0.000	6.401.608
reato	1.24111654	1.06747762	1329	5339	0.000	6.110.966
gelosia	1.17761321	1.01069477	1261	5055	0.000	6.033.321
vessare	1.07301949	0.91572345	1149	4580	0.000	5.965.160
assassino	1.04873881	0.89492973	1123	4476	0.000	5.899.279
media	0.88624499	0.75837097	949	3793	0.000	5.323.233
indagato	0.75923833	0.64260593	813	3214	0.000	5.262.757
omicidio	1.10757277	0.96830757	1186	4843	0.000	5.150.923
lite	0.93387249	0.81095509	1000	4056	0.000	4.957.873
arma	0.65931398	0.55763160	706	2789	0.000	4.920.425
accusa	0.87784014	0.76097019	940	3806	0.000	4.863.238
fallimento	1.00764842	0.88233354	1079	4413	0.000	4.852.815
minore	1.16640674	1.03588716	1249	5181	0.000	4.677.045
folia	1.04126783	0.91852261	1115	4594	0.000	4.663.493
isolamento	0.82927877	0.72418130	888	3622	0.000	4.484.618
reato	0.99550807	0.88353317	1066	4419	0.000	4.338.867
ragazzo	0.97776450	0.86753800	1047	4339	0.000	4.309.577
criminale	0.85916269	0.75637158	920	3783	0.000	4.296.458
elisa_pomarelli	0.46973786	0.40047906	503	2003	0.000	3.948.533
aggressione	0.62009133	0.54763461	664	2739	0.000	3.551.316
offesa	0.61262035	0.54363582	656	2719	0.000	3.394.571
ennesima	0.61822359	0.55023383	662	2752	0.000	3.326.547

beppe_grillo	0.72655280	0.65300279	778	3266	0.000	3.310.374
balordo	0.70694147	0.63660774	757	3184	0.000	3.205.851
stupratore	0.57059609	0.50924621	611	2547	0.000	3.118.655
affermazione	0.21665842	0.17974572	232	899	0.000	3.112.335
indifferenza	0.73215603	0.68279380	784	3415	0.000	2.177.293
pentirsi	0.35767316	0.32450200	383	1623	0.000	2.103.548
...
illegale	0.325896	0.5548721	256	1259	0.000	2.012587

CLUSTER 2: MALTRATTAMENTO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
subire	1.1824746	0.9715066	994	4859	0.000	6.644642
terrorismo	0.7815753	0.6296099	657	3149	0.000	5.913753
bibbiano	0.9112430	0.7581710	766	3792	0.000	5.456215
fragilità	0.4532423	0.3498943	381	1750	0.000	5.347861
sottomissione	0.8339182	0.6943903	71	3473	0.000	5.194147
guerra	0.8815027	0.7473743	741	3738	0.000	4.825441
assistere	1.0647030	0.9171230	895	4587	0.000	4.808991
bambina	1.0325835	0.8901312	868	4452	0.000	4.711047
ancestrale	0.6864063	0.5730269	577	2866	0.000	4.641392
negare	1.0908745	0.9477138	917	4740	0.000	4.594665
colpire	0.9778613	0.8501433	822	4252	0.000	4.324416
odio	0.9968951	0.8705371	838	4354	0.000	4.230770
peggiore	0.5912373	0.5034480	497	2518	0.000	3.838564
ipocrisia	0.9695340	0.8589406	815	4296	0.000	3.732935
oppressione	0.6721310	0.5824241	565	2913	0.000	3.657355
immigrazione	0.6530972	0.5698279	549	2850	0.000	3.433816
sudditanza	0.7887130	0.7005884	663	3504	0.000	3.288553
privazione	0.8589001	0.7741662	722	3872	0.000	3.14436
urgenza	0.7292323	0.6568016	613	3285	0.000	2.792956
maschilismo	0.3652110	0.3149049	37	1575	0.000	2.772412
estremo	0.6947336	0.6246114	584	3124	0.000	2.771035
roberto_saviano	0.3568837	0.3097065	30	1549	0.000	2.622324
negare	0.5733931	0.5150445	482	2576	0.000	2.534108
immorale	0.5162917	0.4662592	434	2332	0.000	2.281860
...
indulgenza	0.254875	0.254789	425	2358	0.000	2.258755

CLUSTER 3: INEFFICACIA DEL SISTEMA GIURIDICO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
----------	----------	--------	-------------------	-------------------	---------	--------

nessuna_riduzione	0.5320578	0.34909457	297	1746	0.000	7.264982
esemplare	0.6950789	0.48305412	388	2416	0.000	7.231653
premeditazione	0.8688486	0.63660774	485	3184	0.000	6.965318
ergastolo	0.7309077	0.52424168	408	2622	0.000	6.801058
legislazione	0.6753731	0.48405382	377	2421	0.000	6.548341
ingiustizia	0.5607209	0.39308129	313	1966	0.000	6.338981
nessuna_riduzione	0.3367908	0.21513503	188	1076	0.000	6.111836
pedofilia	0.7900253	0.60661680	441	3034	0.000	5.671096
frustrazione	0.3851597	0.26272066	215	1314	0.000	5.626736
indulgenza	0.6771645	0.51204536	378	2561	0.000	5.538951
prostituzione	0.3332079	0.22193298	186	1110	0.000	5.538599
coltello	0.7882338	0.61141535	440	3058	0.000	5.453785
attentato	0.6252127	0.46945822	349	2348	0.000	5.448193
orrore	0.7864424	0.62381161	439	3120	0.000	4.981897
galera	0.7219505	0.56962797	403	2849	0.000	4.876317
norma	0.7882338	0.62920998	440	3147	0.000	4.854870
clandestino	0.4836889	0.36349023	270	1818	0.000	4.770255
instabile	0.4191971	0.30810695	234	1541	0.000	4.768616
castrazione	0.3690367	0.27231776	206	1362	0.000	4.414228
scontare	0.3278336	0.23892784	183	1195	0.000	4.320217
ferire	0.4460687	0.34589554	249	1730	0.000	4.090412
pugni	0.4514430	0.36349023	252	1818	0.000	3.520844
bimba	0.4066570	0.32790097	227	1640	0.000	3.316621
strangolare	0.5016033	0.41967326	280	2099	0.000	3.069368
possesso	0.3762025	0.30790701	210	1540	0.000	2.970683
...
scontare	0.356987	0.365897	158	1547	0.000	2.56989

CLUSTER 4: STRUMENTALIZZAZIONE DELLA COMUNICAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
strumentalizzazione	0.5320578	0.34909457	297	1746	0.000	7.264982
segni	0.6950789	0.48305412	388	2416	0.000	7.231653
ragazza	0.8688486	0.63660774	485	3184	0.000	6.965318
giornalista	0.7309077	0.52424168	408	2622	0.000	6.801058
rituale	0.6753731	0.48405382	377	2421	0.000	6.548341
tradizione	0.5607209	0.39308129	313	1966	0.000	6.338981
infamia	0.3367908	0.21513503	188	1076	0.000	6.111836
nascondere	0.7900253	0.60661680	441	3034	0.000	5.671096
errore	0.3851597	0.26272066	215	1314	0.000	5.626736

giornale	0.6771645	0.51204536	378	2561	0.000	5.538951
grave	0.3332079	0.22193298	186	1110	0.000	5.538599
retroscena	0.7882338	0.61141535	440	3058	0.000	5.453785
famiglia	0.6252127	0.46945822	349	2348	0.000	5.448193
cronaca	0.7846509	0.62101245	438	3106	0.000	5.022394
strage	0.7864424	0.62381161	439	3120	0.000	4.981897
repubblica	0.7219505	0.56962797	403	2849	0.000	4.876317
punire	0.7882338	0.62920998	440	3147	0.000	4.854870
clandestino	0.4836889	0.36349023	270	1818	0.000	4.770255
rabbia	0.4191971	0.30810695	234	1541	0.000	4.768616
menzogna	0.3690367	0.27231776	206	1362	0.000	4.414228
...
meschino	0.258965	0.478575	458	1589	0.000	3.45879

CLUSTER 5: PSICOPATOLOGIE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
sottomettere	0.8211353	0.17414741	46	871	0.000	8.426818
onorare	1.0531953	0.28871281	59	1444	0.000	8.238520
psicosi	0.4462692	0.08597404	25	430	0.000	6.482176
assassinio	0.5712246	0.14715556	32	736	0.000	6.271565
convalida	0.7318815	0.24212688	41	1211	0.000	5.991783
efficacia	0.4819707	0.18194505	27	910	0.000	4.328802
...
mentale	0.485795	0.1258745	24	1458	0.000	3.25896

CLUSTER 6: BRUTALITA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
vergogna	1.1127000	0.8693375	498	4348	0.000	5.577173
massacro	0.9384217	0.7199826	420	3601	0.000	5.478435
branco	0.8825632	0.6727968	395	3365	0.000	5.434506
carnefice	0.8646885	0.6647992	387	3325	0.000	5.215568
belva	0.9205470	0.7155839	412	3579	0.000	5.165672
politica	0.8915006	0.6937905	399	3470	0.000	5.601378
truce	1.0210921	0.8193526	457	4098	0.000	4.777906
livido	0.7373313	0.5896219	330	2949	0.000	4.110665
attenzione	0.6747699	0.5340387	32	2671	0.000	4.106199
stupro	0.9585307	0.7921608	429	3962	0.000	4.201929
...
inganno	0.985965	0.764585	458	2548	0.000	3.25698

CLUSTER 7: AMORE CRIMINALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
partner	0.8853525	0.6544024	553	3273	0.000	7.300846
amore	0.9445894	0.7145842	590	3574	0.000	6.984822
passionale	0.6724196	0.4828542	420	2415	0.000	6.935314
psicologica	0.8309185	0.6192130	519	3097	0.000	6.885677
mura	0.8085045	0.6004187	55	3003	0.000	6.868396
falso_incidente	0.6692176	0.4938509	418	2470	0.000	6.369326
sfruttamento	0.7556715	0.5716274	472	2859	0.000	6.240129
fisica	0.7044396	0.5278406	440	2640	0.000	6.219840
delitto	0.8453275	0.6516032	528	3259	0.000	6.173225
problema	0.6067786	0.4454655	379	2228	0.000	6.160632
razzismo	0.7796865	0.5958201	487	2980	0.000	6.116402
dannoso	0.6420006	0.4770559	41	2386	0.000	6.999.207
patetico	0.6756216	0.5128451	422	2565	0.000	5.825541
timore	0.9157714	0.7255809	572	3629	0.000	5.769084
retorica	0.5971726	0.4478647	373	2240	0.000	5.703758
spietato	0.6516066	0.4988493	47	2495	0.000	5.548552
colpa	0.6131826	0.4666591	383	2334	0.000	5.495855
mostro	0.7604745	0.6086162	475	3044	0.000	5.300.230
sessuale	0.8309185	0.6727968	519	3365	0.000	4.992556
calci	0.5491427	0.4232722	343	2117	0.000	4.962397
italiano	0.5907686	0.4608608	369	2305	0.000	4.918850
spose_bambine	0.5235267	0.4034781	327	2018	0.000	4.845879
piangere	0.5667537	0.4448657	354	2225	0.000	4.699991
...
convivente	0.548799	0.458785	358	2547	0.000	3.25985

CLUSTER 8: DISTRUZIONE DELL'IDENTITA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
identità	0.7131214	0.38868262	280	1944	0.000	9.712420
denigrare	0.7105746	0.39328123	279	1967	0.000	9.468182
minaccia	0.6596373	0.36329029	259	1817	0.000	9.189511
botte	0.6519967	0.36628938	256	1832	0.000	8.853445
moglie	0.6036064	0.33129995	237	1657	0.000	8.833074
padre	0.7767930	0.46086082	305	2305	0.000	8.819011
obbligo_coniugale	0.6978403	0.40347815	274	2018	0.000	8.737731
ostilità	0.7385901	0.43746788	290	2188	0.000	8.622596
abuso	0.5730440	0.31650442	225	1583	0.000	8.520994
insulto	0.6163407	0.35329331	242	1767	0.000	8.326441

impunito	0.5730440	0.32150291	225	1608	0.000	8.312419
morte	0.6952934	0.41527459	273	2077	0.000	8.238884
fenomeno	0.7589650	0.46665907	298	2334	0.000	8.157404
carcere	0.6952934	0.42327217	273	2117	0.000	7.953478
distretto	0.6137938	0.37188769	241	1860	0.000	7.534420
penale	0.5221068	0.30430810	205	1522	0.000	7.442911
rapporto	0.6596373	0.41187561	259	2060	0.000	7.373073
gesto	0.4507946	0.25872187	177	1294	0.000	7.093827
camera	0.4278729	0.24332652	168	1217	0.000	7.014995
denuncia	0.4431540	0.25492301	174	1275	0.000	7.005589
indegno	0.4406072	0.25852193	173	1293	0.000	6.751934
paura	0.3896699	0.22053340	153	1103	0.000	6.743886
iperbole	0.3820293	0.21693449	150	1085	0.000	6.640071
camera	0.5781377	0.36808884	227	1841	0.000	6.628447
...
distretto	0.587485	0.3598698	148	1584	0.000	4.23569

CLUSTER 9: ABUSO DOMESTICO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
domestico	0.6226745	0.22053340	82	1103	0.000	8.153993
martirio	0.5619257	0.18854306	74	943	0.000	8.088128
indegno	0.6682360	0.25492301	88	1275	0.000	7.921615
violenza	0.5011770	0.16734946	66	837	0.000	7.661220
malato	0.4556155	0.14535610	60	727	0.000	7.559751
ratifica	0.5998937	0.24332652	79	1217	0.000	7.078578
vittima	0.6226745	0.25852193	82	1293	0.000	7.048936
battaglia	0.6226745	0.25872187	82	1294	0.000	7.043450
italia	0.3948667	0.12876111	52	644	0.000	6.913491
percosse	0.4404283	0.15675266	58	784	0.000	6.811510
paura	0.7062040	0.32150291	93	1608	0.000	6.809034
malato	0.5239578	0.21593479	69	1080	0.000	6.505046
donna	0.3113372	0.10116945	41	506	0.000	6.141201
ratifica	0.4632091	0.19414137	61	971	0.000	6.006561
denuncia	0.6302681	0.31650442	83	1583	0.000	5.699496
minaccia	0.6910168	0.36329029	91	1817	0.000	5.609777
arresto	0.2202141	0.06737965	29	337	0.000	5.379408
moglie	0.6302681	0.33129995	83	1657	0.000	5.354070
rapporto	0.7062040	0.41187561	93	2060	0.000	4.823730
morte	0.6986104	0.41527459	92	2077	0.000	4.639701

abuso	0.6606424	0.38868262	87	1944	0.000	4.592068
ostilità	0.6606424	0.39328123	87	1967	0.000	4.497283
insulto	0.6074873	0.35329331	80	1767	0.000	4.488514
denigrare	0.7137976	0.43746788	94	2188	0.000	4.433462
rapimento	0.3493052	0.17094837	46	855	0.000	4.358228
botte	0.6074873	0.36628938	80	1832	0.000	4.209346
contrastare	0.1139039	0.03159046	15	158	0.000	4.091627
...
casa	0.589685	0.021458	58	147	0.000	3.25896

Tabella 2. Caratterizzazione dei cluster dei concetti negativi italiani

CLUSTER 1: RIDURRE LA DISPARITA' DI GENERE						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
legitimate	0.598335878	0.29238127	192	1982	0.000	9.144170
rehabilitation	0.666895198	0.34357012	214	2329	0.000	9.009081
welfare	1.081367447	0.67902673	347	4603	0.000	8.292484
communication	0.479915236	0.23425906	154	1588	0.000	8.191391
present	0.766617844	0.44137475	246	2992	0.000	8.168217
triumph	0.747919848	0.42750803	240	2898	0.000	8.164422
dignity	0.694942192	0.39343131	223	2667	0.000	7.991269
task	0.448751909	0.22791577	144	1545	0.000	7.518720
composure	0.835177164	0.53121340	268	3601	0.000	7.087533
advocacy	0.782199508	0.49861185	251	3380	0.000	6.824947
journalism	0.529776559	0.30964091	170	2099	0.000	6.604093
organization	0.289818941	0.15917225	93	1079	0.000	5.382866
direction	0.330331266	0.19560927	106	1326	0.000	5.079311
...
change	0.4785965	0.1478562	145	1258	0.000	4.235696

CLUSTER 2: COMUNICAZIONE PROGRESSO						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
divulgation	0.916827179	0.73346099	1685	4972	0.000	10.532297
manifestation	1.267778830	1.06050906	2330	7189	0.000	9.964456
discourse	1.233499831	1.03513591	2267	7017	0.000	9.653861
forgiveness	1.253631941	1.06257431	2304	7203	0.000	9.187711
police	1.291175606	1.10299433	2373	7477	0.000	8.890857
practical	1.203029611	1.02333444	2211	6937	0.000	8.806166
tolerance	1.071354728	0.90723754	1969	6150	0.000	8.531728

resource	1.185618056	1.01315568	2179	6868	0.000	8.497142
literacy	1.250911386	1.07703111	2299	7301	0.000	8.318347
picture	0.947297400	0.80102437	1741	5430	0.000	8.085478
improvement	1.333616271	1.15919880	2451	7858	0.000	8.054193
wellness	1.056119617	0.90266448	1941	6119	0.000	8.004265
house	1.148618502	0.98940524	2111	6707	0.000	7.942588
target	1.053943173	0.90399214	1937	6128	0.000	7.818356
domestic	1.317837050	1.15123281	2422	7804	0.000	7.723132
model	1.254720164	1.09709359	2306	7437	0.000	7.483086
ethics	1.036531618	0.89543608	1905	6070	0.000	7.396082
leadership	1.222073499	1.07009775	2246	7254	0.000	7.304926
control	1.013134842	0.88850272	1862	6023	0.000	6.567031
europe	1.069178283	0.94278945	1965	6391	0.000	6.470637
attention	1.008237842	0.88983038	1853	6032	0.000	6.237892
democracy	0.964164844	0.85118059	1772	5770	0.000	6.084224
legal	0.902680291	0.79718889	1659	5404	0.000	5.867684
motivation	0.731829410	0.63772161	1345	4323	0.000	5.838079
advertising	0.617021971	0.53180347	1134	3605	0.000	5.778011
civilization	0.853710293	0.75485114	1569	5117	0.000	5.649525
therapy	0.572404862	0.49890689	1052	3382	0.000	5.147857
resource	0.584919417	0.51085587	1075	3463	0.000	5.127961
resource	0.710609078	0.62975562	1306	4269	0.000	5.054000
money	0.535405308	0.47471389	984	3218	0.000	4.362129
wear	0.785152297	0.71811908	1443	4868	0.000	3.936145
...
equity	0.8754895	0.6352895	1558	4785	0.000	3.2569

CLUSTER 3: PARTECIPAZIONE SOCIALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
without_tie	0.6329114	0.0857813	24	581	0.000	7.360939
celebration	0.9757384	0.21626183	37	146	0.000	7.333696
emancipation	0.8175105	0.16404035	31	112	0.000	7.055981
partecipation	0.4483122	0.09337908	17	633	0.000	5.082172
...
support	0.789585	0.1254788	22	587	0.000	6.25896

CLUSTER 4: ATTIVISMO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
activism	1.17618068	1.07747366	1766	7304	0.000	4.144019
attempt	1.11157732	1.02038408	1669	6917	0.000	3.933126

community	0.89778684	0.81621875	1348	5533	0.000	3.923441
contribution	1.04497592	0.95872143	1569	6499	0.000	3.835771
news	1.11357536	1.02643233	1672	6958	0.000	3.748783
school	1.13821788	1.05224803	1709	7133	0.000	3.654275
confirmation	1.10225313	1.01787627	1655	6900	0.000	3.645277
message	0.87114628	0.79718889	1308	5404	0.000	3.600793
end	0.97238040	0.89617367	1460	6075	0.000	3.504525
safety	1.00301704	0.92774259	1506	6289	0.000	3.404103
group	0.93508362	0.86534235	1404	5866	0.000	3.263831
membership	1.09559299	1.02008904	1645	6915	0.000	3.260396
inclusion	0.92775746	0.85870402	1393	5821	0.000	3.243936
security	1.09559299	1.02053160	1645	6918	0.000	3.240689
administration	0.82186124	0.76089939	1234	5158	0.000	3.039300
regulation	0.95439802	0.88983038	1433	6032	0.000	2.981926
wife	0.68399635	0.62975562	127	4269	0.000	2.966888
coherent	0.90045089	0.83982168	1352	5693	0.000	2.880793
child	1.22146963	1.15123281	1834	7804	0.000	2.860251
online	0.95506404	0.89307579	1434	6054	0.000	2.858141
court	0.79721873	0.74201705	1197	5030	0.000	2.787116
history	1.22746375	1.15919880	1843	7858	0.000	2.770840
attraction	0.63204726	0.58431998	949	3961	0.000	2.709056
individual	0.57676810	0.53209851	866	3607	0.000	2.654604
heart	0.93108753	0.88319206	1398	5987	0.000	2.221459
ideal	1.15486823	1.10299433	1734	7477	0.000	2.158046
favour	0.67667020	0.63772161	116	4323	0.000	2.118878
...
message	0.5248965	0.6985696	454	4253	0.000	2.032569

CLUSTER 5: LEGGE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
law	1.01795885	0.90576236	1285	6140	0.000	4.581952
fundamental	1.02429634	0.91181061	1293	6181	0.000	4.578946
right	0.91180595	0.80928539	1151	5486	0.000	4.424526
board	1.00449169	0.89941907	1268	6097	0.000	4.308685
affection	0.94586994	0.84616497	1194	5736	0.000	4.212944
respect	0.91180595	0.81430101	1151	5520	0.000	4.197901
donald_trump	1.03459476	0.93231565	1306	6320	0.000	4.123250
society	0.97518082	0.87596366	1231	5938	0.000	4.123163
support	0.97359644	0.87507855	1229	5932	0.000	4.096390

future	0.91655906	0.82300459	1157	5579	0.000	4.008927
idealist	0.93398715	0.8470679	1179	5699	0.000	3.956449
face	1.09321651	0.99515845	1380	6746	0.000	3.831975
sense	0.84922326	0.76635757	1072	5195	0.000	3.679516
project	0.82308113	0.74186953	1039	5029	0.000	3.663784
dream	0.91339032	0.82787270	1153	5612	0.000	3.657120
home	0.83892485	0.75780151	1059	5137	0.000	3.622431
efficacy	0.41114447	0.35787939	519	2426	0.000	3.434200
armony	1.04093224	0.95724625	1314	6489	0.000	3.336359
revolution	1.05756815	0.97435837	1335	6605	0.000	3.289121
british	0.70108450	0.63447621	885	4301	0.000	3.246521
country	0.70900636	0.64406490	895	4366	0.000	3.142935
mobilization	1.05756815	0.97834136	1335	6632	0.000	3.126412
date	0.85952168	0.79114064	1085	5363	0.000	2.994041
pride	0.50145366	0.45243863	633	3067	0.000	2.825920
...
protect	0.1258798	0.4785962	558	4257	0.000	2.589633

CLUSTER 6: LIBERTA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
freedom	0.9126071	0.7290354	781	4942	0.000	6.520874
independent	0.7688802	0.6043825	658	4097	0.000	6.396217
public_interest	0.9172811	0.7375915	785	5000	0.000	6.352020
change	0.8670351	0.6933360	742	4700	0.000	6.325948
boris_johnson	0.7723857	0.6116109	661	4146	0.000	6.221105
media	0.9535049	0.7737335	816	5245	0.000	6.214327
maternity	0.9254607	0.7558838	792	5124	0.000	5.935048
body	0.7829023	0.6282804	670	4259	0.000	5.914804
inviolable	0.8109466	0.6546862	694	4438	0.000	5.861903
space	0.7104547	0.5649951	608	3830	0.000	5.856740
no_gap	0.7992615	0.6465727	684	4383	0.000	5.764878
enthusiasm	0.9523364	0.7858300	815	5327	0.000	5.724771
understand	0.9780437	0.8094329	837	5487	0.000	5.715372
heroism	0.7092862	0.5738462	607	3890	0.000	5.424055
quietly	0.7349934	0.6034974	629	4091	0.000	5.146715
care	0.9231237	0.7759463	790	5260	0.000	5.142416
london	0.8833943	0.7398043	756	5015	0.000	5.095894
marriage	0.7034436	0.5781242	602	3919	0.000	5.010647
word	0.9359773	0.7911406	801	5363	0.000	4.978832

...
benefit	0.8958696	0.785986	457	4578	0.000	3.25698

CLUSTER 7: FIDUCIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
trust	0.77988839	0.48400754	443	3281	0.000	9.811578
talk	0.79397215	0.49757922	451	3373	0.000	9.710502
white	0.83622344	0.53445880	475	3623	0.000	9.573518
reaction	0.88375614	0.57812422	502	3919	0.000	9.362053
positive_vision	0.83446297	0.53947442	474	3657	0.000	9.332815
self_respect	0.73939757	0.46453513	420	3149	0.000	9.320599
show	0.91368414	0.60349736	519	4091	0.000	9.316392
life	0.70946957	0.44358753	403	3007	0.000	9.217145
energy	0.67426016	0.41865693	383	2838	0.000	9.107917
labour	0.67954157	0.42411511	386	2875	0.000	9.051071
power	0.69890675	0.43945701	397	2979	0.000	9.044291
family	0.92072602	0.61942934	523	4199	0.000	8.960057
justice	0.92776790	0.64613015	527	4380	0.000	8.247915
idea	0.92776790	0.65468621	527	4438	0.000	7.962044
...
strategy	0.9856985	0.6253896	558	4458	0.000	6.32589

CLUSTER 8: TRASFORMAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
trasformation	0.76040273	0.32380267	216	2195	0.000	11.321738
critical_thinking	0.78856580	0.34224246	224	2320	0.000	11.302535
humanity	0.78504541	0.34696304	223	2352	0.000	11.057802
mindness	0.65831162	0.26951593	187	1827	0.000	10.956714
work	0.68999507	0.28943090	196	1962	0.000	10.948774
woman	0.80264733	0.36348509	228	2464	0.000	10.883663
evolution	0.64423009	0.26302513	183	1783	0.000	10.867437
stop	0.61606703	0.24930593	175	1690	0.000	10.718115
rights	0.57030205	0.23189877	162	1572	0.000	10.260169
girl	0.67943392	0.30241251	193	2050	0.000	10.200695
job	0.70759699	0.32557289	201	2207	0.000	10.028867
feel	0.76392311	0.36658297	217	2485	0.000	9.914914
...
witness	0.5289662	0.3656889	212	2545	0.000	8.754855

CLUSTER 9: EDUCARE ALL'UGUAGLIANZA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
----------	----------	--------	-------------------	-------------------	---------	--------

read	0.80661424	0.19000357	80	1288	0.000	10.558534
education	0.76628352	0.17303897	76	1173	0.000	10.534207
identity	0.72595281	0.15710699	72	1065	0.000	10.488888
person	0.62512603	0.12450544	62	844	0.000	10.151434
choice	0.85702763	0.23189877	85	1572	0.000	10.014760
equality	0.54446461	0.10208266	54	692	0.000	9.751008
go_out	0.60496068	0.12686574	60	860	0.000	9.724412
young	0.69570478	0.17053115	69	1156	0.000	9.574099
world	0.85702763	0.24930593	85	1690	0.000	9.547673
woman	0.85702763	0.26951593	85	1827	0.000	9.035291
mindness	0.81669691	0.26302513	81	1783	0.000	8.662456
reply	0.51421658	0.11948982	51	810	0.000	8.470730
gender	0.36297641	0.06579316	36	446	0.000	8.066271
purpose	0.36297641	0.06756338	36	458	0.000	7.964974
...
image	0.3256896	0.0365896	25	475	0.000	6.23569

CLUSTER 10: SVILUPPARE LA CULTURA DI GENERE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
gender	0.8857396	0.6579316	10	446	0.000	5.689624
development	0.7971656	0.5266403	9	357	0.000	5.561696
success	0.6200177	0.3304410	7	224	0.000	5.150054
culture	0.8857396	0.1020826	10	692	0.000	4.950667
Phenomenon	0.6895696	0.1245875	12	586	0.000	3.253689

Tabella 3. Caratterizzazione dei cluster dei concetti positivi inglesi

CLUSTER 1: DOMINIO MASCHILE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
domination	1.01677199	0.79356522	1498	6109	0.000	10.388.621
patriarchal_culture	1.39551616	1.13481516	2056	8736	0.000	10.221.353
psychological	1.26315932	1.01998267	1861	7852	0.000	10.040.599
authority	0.36584786	0.24486339	539	1885	0.000	9.894337
murder	0.87219760	0.67951214	1285	5231	0.000	9.681517
offensive	1.19596278	0.98049277	1762	7548	0.000	9.088694
economic_restriction	0.66585669	0.51077074	981	3932	0.000	8.958445
depeny	0.97944057	0.78992800	1443	6081	0.000	8.876537
dark	1.18849649	0.98166188	1751	7557	0.000	8.727279

crime	1.09143482	0.89410860	1608	6883	0.000	8.710333
threat	0.95839923	0.78486186	1412	6042	0.000	8.169413
bill	1.11722743	0.93346860	1646	7186	0.000	7.959768
trauma	0.92785534	0.76355809	1367	5878	0.000	7.844983
disobedience	0.99708815	0.82811889	1469	6375	0.000	7.759437
consequence	0.98351309	0.81590820	1449	6281	0.000	7.752411
lying	1.03984959	0.86773870	1532	6680	0.000	7.727746
suicide	0.91495904	0.75797235	1348	5835	0.000	7.529192
kill	1.12944498	0.95555177	1664	7356	0.000	7.457175
prison	0.11199424	0.06443090	165	496	0.000	7.400089
accuse	0.34073400	0.25148834	502	1936	0.000	7.290360
campaign	1.02966829	0.87150583	1517	6709	0.000	7.098272
leave	0.60612642	0.48764836	893	3754	0.000	7.043418
refuse	0.88305765	0.74147492	1301	5708	0.000	6.875676
breaking_news	0.62309525	0.50635411	918	3898	0.000	6.820126
aversion	1.02355952	0.88696404	1508	6828	0.000	6.094042
bout	0.41743309	0.33345587	615	2567	0.000	6.021210
...
institution	0.4578562	0.256896	457	2547	0.000	6.02250

CLUSTER 2: SUPREMAZIA MASCHILE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
masculine	0.7593838	0.15756992	35	1213	0.000	7.380955
abnegation	0.9546539	0.25824319	44	1988	0.000	7.127989
interfering	0.4339336	0.07534258	20	580	0.000	6.040510
delay	0.3905402	0.10755830	18	828	0.000	4.449601
disequilibrium	0.5345629	0.02145785	21	1254	0.000	3.025689

CLUSTER 3: OPPRESSIONE FAMILIARE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
domestic	0.97137109	0.88241751	2636	6793	0.000	6.103069
sick	0.96363254	0.88085870	2615	6781	0.000	5.685781
mortify	1.22600592	1.13481516	3327	8736	0.000	5.531864
victim	1.08118466	0.99582108	2934	7666	0.000	5.521389
damage	1.08855470	1.00452445	2954	7733	0.000	5.412364
pretend	0.94815546	0.87150583	2573	6709	0.000	5.294263
oppression	1.11361283	1.03154386	322	7941	0.000	5.218239
fragility	1.02738338	0.94879692	2788	7304	0.000	5.206347
depression	1.03364791	0.95555177	2805	7356	0.000	5.156028
beating	0.85529298	0.78486186	2321	6042	0.000	5.122386

dad	1.10255777	1.02349000	2992	7879	0.000	5.047333
social_role	1.05760054	0.98166188	2870	7557	0.000	4.948263
mask	0.96621206	0.89410860	2622	6883	0.000	4.919172
trap	1.00637877	0.93346860	2731	7186	0.000	4.870134
house	0.56491346	0.51077074	1533	3932	0.000	4.865829
survivor	1.01780233	0.94489989	2762	7274	0.000	4.840661
livid	0.96694906	0.89787573	2624	6912	0.000	4.703327
blame	1.02554087	0.95451257	2783	7348	0.000	4.693220
shut	0.90061871	0.83461394	2444	6425	0.000	4.658908
master	0.85676699	0.79356522	2325	6109	0.000	4.573344
bitch	0.28006147	0.24486339	760	1885	0.000	4.544181
insane	1.00416776	0.93749553	2725	7217	0.000	4.445112
exploitation	1.08855470	1.01998267	2954	7852	0.000	4.386315
psychopatic	0.94926097	0.88540523	2576	6816	0.000	4.378837
abortion	0.88035111	0.81954542	2389	6309	0.000	4.331361
rhetoric	0.25316083	0.22563804	687	1737	0.000	3.701269
knife	0.63971935	0.59546620	1736	4584	0.000	3.690103
violation	0.98537416	0.93035098	2674	7162	0.000	3.683434
allegation	0.93083587	0.87774107	2526	6757	0.000	3.657592
act	0.93636340	0.88696404	2541	6828	0.000	3.385762
message	0.80738773	0.76225908	2191	5868	0.000	3.332501
emergency	0.93967992	0.89371890	2550	6880	0.000	3.138420
abuse	1.02332986	0.97555653	2777	7510	0.000	3.124609
blood	0.81107275	0.76927374	2201	5922	0.000	3.072758
condemn	0.73147633	0.69237234	1985	5330	0.000	3.027615
show	0.96510655	0.92242702	2619	7101	0.000	2.869286
report	0.87629759	0.83890067	2378	6458	0.000	2.634076
hell	0.86303152	0.82617038	2342	6360	0.000	2.615919
assault	0.88219362	0.84604523	2394	6513	0.000	2.535393
beast	0.07443739	0.06443090	202	496	0.000	2.484375
conflict	0.43888580	0.41438420	1191	3190	0.000	2.442001
tension	0.71784176	0.69029393	1948	5314	0.000	2.134719
obsession	0.71599925	0.68860521	1943	5301	0.000	2.125331
...
husband	0.6365896	0.587845	1589	5478	0.000	2.02569

CLUSTER 4: MALTRATTAMENTO E VIOLENZA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
monster	0.85110122	0.75459492	1453	5809	0.000	5.136209

martyr	0.89737582	0.79993037	1532	6158	0.000	5.041638
violence	0.94892221	0.84929275	1620	6538	0.000	5.006543
evidence	0.87570291	0.78369275	1495	6033	0.000	4.810656
incest	0.62031396	0.54493471	159	4195	0.000	4.707333
abandonment	0.77788191	0.69328165	1328	5337	0.000	4.697245
disparity	0.83001406	0.74355334	1417	5724	0.000	4.639749
pain	0.95770853	0.86604998	1635	6667	0.000	4.566277
bullying	0.88800375	0.80070978	1516	6164	0.000	4.519141
question	0.87218838	0.78745988	1489	6062	0.000	4.423088
burder	0.58165417	0.51323886	993	3951	0.000	4.402595
impulse	0.82005623	0.73913670	1400	5690	0.000	4.357604
media	0.98406748	0.89631692	1680	6900	0.000	4.301003
anger	0.85168697	0.77109235	1454	5936	0.000	4.252144
public	0.76030928	0.68639690	1298	5284	0.000	4.128872
word	0.86750234	0.78862899	1481	6071	0.000	4.116915
sad	0.90323336	0.82279295	1542	6334	0.000	4.112583
disturb	0.80164011	0.91788048	1710	7066	0.000	4.059739
stop	0.78491097	0.71198739	1340	5481	0.000	4.002381
precaution	0.80306935	0.73017353	1371	5621	0.000	3.952192
government	0.79194002	0.72043096	1352	5546	0.000	3.902934
court	0.79662605	0.72822502	1360	5606	0.000	3.715005
story	0.94130740	0.86994701	1607	6697	0.000	3.553725
pornography	0.86398782	0.80096958	1475	6166	0.000	3.269232
labour	0.87746017	0.81512879	1498	6275	0.000	3.206271
extreme	0.67596064	0.62144640	1154	4784	0.000	3.202647
remember	0.76148079	0.70627175	1300	5437	0.000	3.047146
journalism	0.76148079	0.70757076	1300	5447	0.000	2.973077
law	0.72574977	0.67431610	1239	5191	0.000	2.904469
indifference	0.89503280	0.83890067	1528	6458	0.000	2.848338
...
homophobia	0.5896585	0.2568956	1254	5896	0.000	2.35689

CLUSTER 5: GARANTISMO SOCIALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
released_assassin	0.83077496	0.6204072	893	4776	0.000	9.066467
sentence	0.77123453	0.5707850	829	4394	0.000	8.991574
effect	0.77030421	0.5715644	828	4400	0.000	8.912098
politics	0.75727975	0.5618218	814	4325	0.000	8.839689
death	0.76286166	0.5668880	820	4364	0.000	8.825740

case	0.72285794	0.5361014	777	4127	0.000	8.644044
sacrifice	0.72750954	0.5401284	782	4158	0.000	8.642175
disgrace	0.76751326	0.5748119	825	4425	0.000	8.628402
coercion	0.72471858	0.5389593	779	4149	0.000	8.578597
ruthless	0.68657550	0.5089521	738	3918	0.000	8.435109
family	0.72378826	0.5420769	778	4173	0.000	8.376017
global	0.81961113	0.6310591	881	4858	0.000	8.092641
young	0.84379942	0.6549609	907	5042	0.000	7.966170
girl	0.86519676	0.6756151	930	5201	0.000	7.882183
person	0.64192018	0.4808935	690	3702	0.000	7.876387
growth	0.75076751	0.5771502	807	4443	0.000	7.786610
life	0.61866220	0.4629672	665	3564	0.000	7.759072
radical	0.60935901	0.4563422	655	3513	0.000	7.681021
inhuman	0.62052284	0.4686828	667	3608	0.000	7.530788
radical	0.50981487	0.3747644	548	2885	0.000	7.454608
disgust	0.85217230	0.6743161	916	5191	0.000	7.415190
...
inability	0.2577856	0.478578	581	5147	0.000	6.589658

CLUSTER 6: AMORE OSTILE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
love	0.70555033	0.36748994	330	2829	0.000	11.133.282
hostility	0.72479261	0.38125944	339	2935	0.000	11.128.909
fallacy	0.79534764	0.43932519	372	3382	0.000	10.845556
cowardice	0.73120670	0.39502895	342	3041	0.000	10.752133
men	0.63071923	0.32397310	295	2494	0.000	10.726324
female	0.73120670	0.39736717	342	3059	0.000	10.657393
bad	0.63285727	0.32774023	296	2523	0.000	10.624066
street	0.67347986	0.35670815	315	2746	0.000	10.619270
wrong	0.80603780	0.45491331	377	3502	0.000	10.553357
issue	0.59864876	0.30630656	280	2358	0.000	10.503435
respect	0.60720089	0.31708835	284	2441	0.000	10.283880
weak	0.81458993	0.46868282	381	3608	0.000	10.278948
activism	0.56444026	0.28773072	264	2215	0.000	10.248065
talk	0.53878389	0.27435092	252	2112	0.000	10.025344
scream	0.59437270	0.32111528	278	2472	0.000	9.684662
...
shame	0.58785965	0.5698569	458	2547	0.000	5.258966

CLUSTER 7: SFRUTTAMENTO DEL CORPO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
commodify	0.72834870	0.22966497	128	1768	0.000	11.153169
body	0.75110959	0.24564280	132	1891	0.000	11.021037
sex	0.66575623	0.20212596	117	1556	0.000	10.955544
sexual	0.63161489	0.18653784	111	1436	0.000	10.877505
harassment	0.60885399	0.17770457	107	1368	0.000	10.765281
hate	0.77956071	0.27435092	137	2112	0.000	10.593087
gender	0.72265847	0.24785111	127	1908	0.000	10.411941
transsexual	0.78525094	0.28773072	138	2215	0.000	10.275534
black	0.54057130	0.15782972	95	1215	0.000	10.135150
discrimination	0.78525094	0.30630656	138	2358	0.000	9.715365
common	0.47857854	0.25689965	258	5478	0.000	8.236588

CLUSTER 8: MISOGINIA						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
muslim	0.63227112	0.09235961	25	711	0.000	7.289548
homophobie	0.70814365	0.12353585	28	951	0.000	7.157569
racism	0.55639858	0.07599209	22	585	0.000	7.012796
woman	0.70814365	0.13951368	28	1074	0.000	6.756133
genital_mutilation	0.45523520	0.05715644	18	440	0.000	6.532325
gap	0.70814365	0.15782972	28	1215	0.000	6.337196
islam	0.47857785	0.14785478	25	1245	0.000	6.25896
brutality	0.65891235	0.25894367	26	1234	0.000	6.32569

Tabella 4. Caratterizzazione dei cluster dei concetti negativi inglesi

CLUSTER 1: EDUCARE						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
civiliser	0.415343119	0.28875834	676	2117	0.000	10.308.500
instruire	0.997806546	0.79521073	1624	5830	0.000	10.141.636
orientation	1.050646055	0.84445105	1710	6191	0.000	10.029.387
économique	0.764329645	0.59170226	1244	4338	0.000	9.970054
indépendance	1.055561358	0.85086184	1718	6238	0.000	9.922763
légalité	0.828228586	0.64967216	1348	4763	0.000	9.860776
normalité	1.102256739	0.89532817	1794	6564	0.000	9.789969
confrontation	1.018696584	0.82617348	1658	6057	0.000	9.475901
perspective	0.964628249	0.77856996	1570	5708	0.000	9.425273
député	0.914246392	0.73737722	1488	5406	0.000	9.203200

humanité	1.209778996	1.01004039	1969	7405	0.000	8.931372
inviolable	0.826999760	0.66822253	1346	4899	0.000	8.676126
mari	1.072150507	0.89164538	1745	6537	0.000	8.580006
reaction	0.327482075	0.23651722	533	1734	0.000	8.214191
conjoint	1.196876325	1.01535998	1948	7444	0.000	8.109960
gagner	0.947424688	0.78907274	1542	5785	0.000	7.996820
actif	1.184588067	1.00690319	1928	7382	0.000	7.973317
emploi	1.245414944	1.06364550	2027	7798	0.000	7.943175
augmenter	0.940666146	0.78402595	1531	5748	0.000	7.936155
développement	1.100413500	0.93433851	1791	6850	0.000	7.731618
caroline_de_haas	0.897657244	0.75033520	1461	5501	0.000	7.630302
génération	0.895814005	0.74883480	1458	5490	0.000	7.620127
défense	1.045116339	0.89164538	1701	6537	0.000	7.315544
solidarité	0.820855631	0.69400209	1336	5088	0.000	6.837087
forme	1.240499641	1.08710626	2019	7970	0.000	6.649985
européen	0.597209337	0.49281241	972	3613	0.000	6.647984
reportage	1.179058351	1.03022756	1919	7553	0.000	6.623288
principe	0.803652070	0.68349931	1308	5011	0.000	6.529068
réglage	0.350215352	0.27811916	570	2039	0.000	6.069911
pas_de_compromis	1.122532364	0.98944402	1827	7254	0.000	6.047364
racheter	0.395067493	0.31958469	643	2343	0.000	5.943862
entité	0.992276830	0.87173101	1615	6391	0.000	5.828547
...
conjoint	0.35894655	0.3978462	1236	5289	0.000	5.2356

CLUSTER 2: PROGRESSO ISTITUZIONALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
énergiquement	1.150823932	1.03172795	2790	7564	0.000	7.019412
représenter	1.102563574	0.98944402	2673	7254	0.000	6.806643
responsable	0.817951203	0.72223685	1983	5295	0.000	6.721153
institution	1.182172541	1.06951069	2866	7841	0.000	6.527281
officiel	1.155361231	1.04414033	2801	7655	0.000	6.520068
confiance	1.165673273	1.05668911	2826	7747	0.000	6.352648
juridictionnel	1.095963867	0.99108082	2657	7266	0.000	6.308937
constitution	1.130612329	1.02449876	2741	7511	0.000	6.280220
marlène_schiappa	1.163610865	1.05750751	2821	7753	0.000	6.183339
résultat	1.070390001	0.96925685	2595	7106	0.000	6.151054
police	0.878173531	0.78907274	2129	5785	0.000	5.995448
normatif	0.985418772	0.89164538	2389	6537	0.000	5.942898

pays	1.086889269	0.99230842	2635	7275	0.000	5.689036
dialogue	0.925608926	0.84090466	2244	6165	0.000	5.526922
annonce	1.022954606	0.93433851	2480	6850	0.000	5.490965
moral	1.122775177	1.03022756	2722	7553	0.000	5.466369
moyen	1.013055046	0.92574532	2456	6787	0.000	5.434874
culture	1.095963867	1.00663039	2657	7380	0.000	5.337416
positivement	1.092664013	1.00690319	2649	7382	0.000	5.124196
aide	1.127312476	1.04100314	2733	7632	0.000	5.073614
série	1.150411450	1.06364550	2789	7798	0.000	5.047020
sincérité	0.985418772	0.90651295	2389	6646	0.000	4.964649
media	0.975931693	0.89751057	2366	6580	0.000	4.958414
message	0.868686452	0.79521073	2106	5830	0.000	4.930320
réforme	1.170623054	1.08710626	2838	7970	0.000	4.807210
humanité	0.966032132	0.89260018	2342	6544	0.000	4.656631
témoin	1.087301751	1.01004039	2636	7405	0.000	4.611167
rue	0.721430487	0.66344854	1749	4864	0.000	4.255408
projection	0.890135500	0.82617348	2158	6057	0.000	4.214559
ambitieux	0.766390991	0.70777847	1858	5189	0.000	4.167306
éthique	1.008105265	0.94074930	2444	6897	0.000	4.164017
cible	1.092664013	1.02258917	2649	7497	0.000	4.158279
communication	0.928083816	0.86422902	2250	6336	0.000	4.115691
propagation	0.960257389	0.89532817	2328	6564	0.000	4.112891
couple	0.991605997	0.92615452	2404	6790	0.000	4.077697
divulgation	0.703693774	0.64967216	1706	4763	0.000	4.006627
genre	1.086476788	1.01931557	2634	7473	0.000	3.992060
espoir	0.836100398	0.77856996	2027	5708	0.000	3.903783
...
protection	0.985966655	0.77485662	1254	5478	0.000	3.256895

CLUSTER 3: COMUNICAZIONE SOCIALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
publicité	0.9275709	0.8181259	1441	5998	0.000	5.299933
divulgation	0.8979608	0.7949379	1395	5828	0.000	5.621070
television	0.9842165	0.8765050	1529	6426	0.000	5.465708
imprimer	0.9745610	0.8679118	1514	6363	0.000	5.211402
vidéo	0.9809980	0.8762322	1524	6424	0.000	4.910768
tournant	0.9076163	0.8091235	1410	5932	0.000	4.800812
tolérance	0.9996653	0.8973742	1553	6579	0.000	4.741078
équipe	0.8696380	0.7772060	1351	5698	0.000	4.596831

campagne	0.9449508	0.8493614	1468	6227	0.000	4.552746
identité	0.9320768	0.8374947	1448	6140	0.000	4.535973
reconnaître	1.0234822	0.9247905	1590	6780	0.000	4.509757
équilibre	0.8252227	0.7377864	1282	5409	0.000	4.461618
efficacement	0.8496833	0.7616564	1320	5584	0.000	4.422867
intégrité	0.8438900	0.7586556	1311	5562	0.000	4.292059
lucidement	0.9449508	0.8591822	1468	6299	0.000	4.662330
courage	0.8020495	0.7238736	1246	5307	0.000	4.300190
occupation	1.0485864	0.9602545	1629	7040	0.000	3.967107
expérience	0.8548329	0.7754328	1328	5685	0.000	3.958582
emplacement	0.8490396	0.7755692	1319	5686	0.000	3.664783
coexistence	0.6231011	0.5614215	968	4116	0.000	3.603659
égalité	0.8368093	0.7657484	1300	5614	0.000	3.567413
connaissance	0.5484319	0.4915848	852	3604	0.000	3.544615
secteur	0.8348782	0.7661576	1297	5617	0.000	3.449778
réconcilier	0.8194294	0.7523812	1273	5516	0.000	3.396158
liberté	0.8786498	0.8113059	1365	5948	0.000	3.288224
mariage	1.0009527	0.9302465	1555	6820	0.000	3.229242
ensemble	0.8934549	0.8287651	1388	6076	0.000	3.126711
...
bébé	0.478568	0.254785	14785	5258	0.000	3.256849

CLUSTER 4: IMPEGNO COLLETTIVO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
systeme	0.8721270	0.6552646	746	4804	0.000	7.999877
réhabiliter	0.8861559	0.6698593	758	4911	0.000	7.899557
engagement	0.8417312	0.6315310	720	4630	0.000	7.895239
collectif	0.8686198	0.6626301	743	4858	0.000	7.573634
travail	0.7797704	0.5895199	667	4322	0.000	7.401587
emmanuel_macron	0.7224859	0.5420527	618	3974	0.000	7.309177
global	0.8417312	0.6461258	720	4737	0.000	7.288609
public	0.7295003	0.5517371	624	4045	0.000	7.146512
mouvement	0.6885829	0.5169552	589	3790	0.000	7.118136
accepter	0.8499147	0.6582653	727	4826	0.000	7.840872
dignité	0.8826486	0.6882733	755	5046	0.000	7.359000
conjugal	0.8487456	0.6634485	726	4864	0.000	6.832513
volonté	0.8171807	0.6367142	699	4668	0.000	6.787920
code	0.9411022	0.7523812	805	5516	0.000	6.560356
exposer	0.7470364	0.5829727	639	4274	0.000	6.447289

le_parisien	0.9434403	0.7661576	807	5617	0.000	6.121122
ordre	0.8557600	0.6877277	732	5042	0.000	6.110585
différence	0.9387641	0.7657484	803	5614	0.000	5.979017
...
personnes	0.8578568	0.5478547	841	5247	0.000	5.235695

CLUSTER 5: GIUSTIZIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
justice	0.78039379	0.43634290	413	3199	0.000	11.299727
innocent	0.79928952	0.45121048	423	3308	0.000	11.262529
juge	0.73504403	0.40660775	389	2981	0.000	11.150066
enquête	0.79551037	0.45639367	421	3346	0.000	10.941206
vérité	0.81062696	0.47112485	429	3454	0.000	10.806701
honnêteté	0.87298288	0.52418436	462	3843	0.000	10.590069
paris	0.67079853	0.37291700	355	2734	0.000	10.564883
femme	0.65001323	0.36091382	344	2646	0.000	10.418799
intention	0.63678621	0.35736743	337	2620	0.000	10.138096
restaurer	0.74449189	0.44002570	394	3226	0.000	10.053871
affirmation	0.58387816	0.32558628	309	2387	0.000	9.803460
presse	0.55175541	0.30730871	292	2253	0.000	9.545528
déclaration	0.57820944	0.32954187	306	2416	0.000	9.417225
français	0.82763312	0.52773076	438	3869	0.000	9.167606
prudent	0.89565776	0.58229067	474	4269	0.000	9.153571
france	0.53852840	0.30553551	285	2240	0.000	9.153469
...
exécutif	0.5487568	0.12547852	478	5252	0.000	8.256895

CLUSTER 6: SICUREZZA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
sécurité	0.69877807	0.24551961	187	1800	0.000	12.534.869
prévention	0.76604013	0.28575754	205	2095	0.000	12.472.176
précaution	0.66141026	0.22833324	177	1674	0.000	12.370.298
respecter	0.78472404	0.30294392	210	2221	0.000	12.239.239
décret	0.62030567	0.21073766	166	1545	0.000	12.132.817
juridique	0.82209185	0.32954187	220	2416	0.000	12.093.217
planifier	0.57172751	0.19232369	153	1410	0.000	11.736.303
important	0.83703897	0.35736743	224	2620	0.000	11.456.427
relations	0.63525279	0.24088202	170	1766	0.000	11.183.374
approbation	0.73988267	0.30553551	198	2240	0.000	11.137.398
privé	0.58293786	0.21223806	156	1556	0.000	11.100.446

rôle	0.60909533	0.22778764	163	1670	0.000	11.085.408
droit	0.75482979	0.32558628	202	2387	0.000	10.757674
assistance	0.46336086	0.18564010	124	1361	0.000	9.059364
...
anti-rapprochement	0.85478547	0.23225698	125	2458	0.000	9.589652

CLUSTER 7: PROGRESSO STRATEGICO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
progrès	0.81049219	0.14308337	55	1049	0.000	10.071901
stratégie	0.83996463	0.15904215	57	1166	0.000	9923909
programme	0.70733864	0.11362102	48	833	0.000	9803759
soutien	0.85470085	0.17854732	58	1309	0.000	9532908
nations_unies	0.66312997	0.10993822	45	806	0.000	9358318
protéger	0.54524020	0.07733868	37	567	0.000	9046439
recherche	0.53050398	0.08484066	36	622	0.000	8487750
populaire	0.44208665	0.05919751	30	434	0.000	8317687
...
proposition	0.85478547	0.23225698	125	2458	0.000	9.589652

CLUSTER 8: COOPERAZIONE INTERNAZIONALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
international	1.0204082	0.4228393	6	310	0.000	5.028030
coopération	0.8503401	0.2632516	5	193	0.000	4.852307
ripositionnement	0.5102041	0.1159398	3	85	0.000	3.899117
national	0.54524020	0.07733868	37	567	0.000	9046439

Tabella 5. Caratterizzazione dei cluster dei concetti positivi francesi

CLUSTER 1: SUPREMAZIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
suprématie	1.2690355	0.2102751	50	1239	0.000	9.855747
sujétion	1.2944162	0.2294528	51	1352	0.000	9.658900
extrémisme	1.1167513	0.2289436	44	1349	0.000	8.353043
declaration	1.85478547	0.23225698	125	2458	0.000	9.589652

CLUSTER 2: PSICOPATOLOGIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
psychose	0.8244455	0.4791015	126	2823	0.000	5.662251
pathologie	0.7851862	0.4522867	120	2665	0.000	5.606234
féroce	0.7001243	0.3925475	107	2313	0.000	5.526857
enfer	0.5103710	0.2647532	78	1560	0.000	5.281683

monstre	0.5888896	0.3221164	90	1898	0.000	5.259453
traumatisme	0.4253092	0.2072203	65	1221	0.000	5.229080
survivre	0.7720997	0.4757072	118	2803	0.000	4.925649
...
fou	0.1167513	0.2289436	44	1349	0.000	2.353043

CLUSTER 3: MORTE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
meurtre	1.40273016	1.1773031	8245	96937	0.000	10.259401
meurtrière	1.17740344	0.9890908	1206	45828	0.000	9.339798
poignardée	0.98858535	0.8193772	2173	34828	0.000	9.199674
cadavre	1.29206337	1.1014412	1226	56490	0.000	8.976647
cimetière	0.90016600	0.7447032	4157	84388	0.000	8.859903
torture	1.29947918	1.1156971	5227	86574	0.000	8.604938
inhumain	1.38732808	1.2046270	7243	27098	0.000	8.244820
enclose	1.08442051	0.9295213	4190	15477	0.000	7.934919
corruption	1.28635889	1.1173942	9225	56584	0.000	7.912804
humiliation	1.28008397	1.1150183	0224	46570	0.000	7.740057
malade	0.76953355	0.6433842	2134	93791	0.000	7.735252
bourreau	1.13119720	0.9790777	1198	35769	0.000	7.601297
prison	0.48031671	0.3849104	384	22268	0.000	7.519428
indécent	1.08042738	0.9386858	7189	45531	0.000	7.233660
mentalité	0.85053708	0.7267136	0149	14282	0.000	7.161199
pédophilie	1.31944484	1.1642352	4231	36860	0.000	7.132064
folie	1.21391207	1.0678379	2212	86292	0.000	7.002520
accusation	1.06274351	0.9268059	2186	35461	0.000	6.983270
mépris	1.26182965	1.1140000	1221	26564	0.000	6.942411
illégal	1.22132789	1.0812452	9214	16371	0.000	6.677469
impitoyable	0.72104552	0.6175877	6126	43639	0.000	6.485763
responsabilité	1.21847565	1.0854881	3213	66396	0.000	6.330082
abuser	1.26182965	1.1285953	8221	26650	0.000	6.223243
acquittement	1.12092914	0.9967279	2196	55873	0.000	6.164775
terrorisme	0.92184300	0.8100429	7161	64773	0.000	6.141506
éviter	0.57900411	0.4936968	4101	52909	0.000	5.972057
aggravant	0.85681200	0.7545466	3150	24446	0.000	5.818984
répugnant	1.12321093	1.0057227	4196	95926	0.000	5.808803
maniaque	0.73530670	0.6416870	9128	93781	0.000	5.767008
jalousie	0.63547841	0.5498720	4111	43240	0.000	5.688069
immoralité	0.91385674	0.8242989	1160	24857	0.000	4.886531

urgence	0.72218641	0.6449116	5126	63800	0.000	4.755867
subir	0.94466090	0.8567142	1165	65048	0.000	4.709897
...
peur	0.5895895	0.5287545	1254	2547	0.000	3.552562

CLUSTER 4: DOMINIO E DISPARITA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
domaine	1.0826567	0.4544930	104	2678	0.000	7.814536
disparité	0.7495315	0.2752754	72	1622	0.000	7.349052
obliger	1.0410160	0.4619604	100	2722	0.000	7.225083
chômage	1.0826567	0.4385399	104	2584	0.000	8.096761
...
privation	2.254785	0.254785	142	5478	0.000	6.2548525

CLUSTER 5: MALTRATTAMENTO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
agression	1.09233261	1.00114048	1504	5899	0.000	3.831395
physique	0.99646299	0.90949514	1372	5359	0.000	3.829345
peur	1.04221895	0.95616637	1435	5634	0.000	3.698358
psychologique	1.09814289	1.01217186	1512	5964	0.000	3.594123
délire	1.17367653	1.08548813	1616	6396	0.000	3.563188
cruel	1.17876052	1.09567095	1623	6456	0.000	3.342943
tyran	1.03059839	0.95497838	1419	5627	0.000	3.254165
sang	1.12428915	1.04560544	1548	6161	0.000	3.239302
raciste	0.97394816	0.90372487	1341	5325	0.000	3.105169
choc	1.02043040	0.94869898	1405	5590	0.000	3.097444
dégoût	1.01316755	0.94445614	1395	5565	0.000	2.973994
prostituée	0.84031172	0.77932481	1157	4592	0.000	2.900131
déséquilibré	1.03495610	0.96940403	1425	5712	0.000	2.801816
condamné	0.73572668	0.68089093	1013	4012	0.000	2.786137
calomnie	0.96232760	0.90083974	1325	5308	0.000	2.724322
mari	0.59482740	0.54749605	819	3226	0.000	2.676031
pornographie	1.22088505	1.15320385	1681	6795	0.000	2.657723
injustice	1.05601836	0.99452165	1454	5860	0.000	2.596249
vengeance	0.87081569	0.81530409	1199	4804	0.000	2.582899
mensonge	0.85628999	0.80444242	1179	4740	0.000	2.428539
série	1.04875551	0.99553993	1444	5866	0.000	2.245839
raison	1.08289090	1.03016150	1491	6070	0.000	2.188420
...
web	1.55854785	0.58968595	1478	5241	0.000	2.0124511

CLUSTER 6: MISOGINIA						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
virilité	1.0325291	0.8954089	1057	5276	0.000	5.014064
répulsion	1.0618345	0.9274848	1087	5465	0.000	4.832243
violence	1.0100615	0.8808135	1034	5190	0.000	4.767578
chienne	1.0696493	0.9390253	1095	5533	0.000	4.672318
viol	1.0120152	0.8872627	1036	5228	0.000	4.587953
conflit	1.0354596	0.9108528	1060	5367	0.000	4.525412
discuter	1.0716030	0.9464927	1097	5577	0.000	4.460632
conjugal	0.9055387	0.7962962	927	4692	0.000	4.238682
insulte	0.8996776	0.7917139	921	4665	0.000	4.201300
évident	1.0139689	0.9016883	1038	5313	0.000	4.102789
piège	1.1009085	0.9845085	1127	5801	0.000	4.075752
tergiversent	0.8391130	0.7418181	859	4371	0.000	3.911268
violence	1.1243528	1.0133599	1151	5971	0.000	3.834636
personnalité	0.9377747	0.8411006	960	4956	0.000	3.658554
suicide	0.9397284	0.8429674	962	4967	0.000	3.657898
plainte	1.0843020	0.9841691	1110	5799	0.000	3.511804
vulgaire	0.9534043	0.8714793	976	5135	0.000	3.051838
supériorité	1.0120152	0.8872627	1036	5228	0.000	4.587953
peine	0.8752564	0.7976539	896	4700	0.000	3.018510
incapable	0.8264140	0.7514918	846	4428	0.000	3.000315
peine	1.1184917	1.0313495	1145	6077	0.000	2.990735
...
problème	1.245758	1.0254785	1452	2541	0.000	2.12452

CLUSTER 7: VIOLENZA FAMILIARE						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
amour	0.9245952	0.6961652	772	4102	0.000	8.213067
crime	0.8347705	0.6233580	697	3673	0.000	8.018682
homme	0.9305835	0.7065177	777	4163	0.000	8.006802
maison	0.9569321	0.7314656	799	4310	0.000	7.927297
contraste	0.8551308	0.6440631	714	3795	0.000	7.886664
haine	0.8132126	0.6096112	679	3592	0.000	7.812224
détérioration	0.8263869	0.6211517	690	3660	0.000	7.805114
dénigrer	0.8443518	0.6381231	705	3760	0.000	7.744889
cage	0.8970490	0.6847943	749	4035	0.000	7.708518
profiter	0.9796876	0.7582803	818	4468	0.000	7.660503
cauchemar	0.9293858	0.7163611	776	4221	0.000	7.576142

trouble	0.7976430	0.6045198	666	3562	0.000	7.452191
abandon	0.9868736	0.7706694	824	4541	0.000	7.429871
tromperie	1.0108269	0.8017270	844	4724	0.000	7.061769
manque	0.7329692	0.5595457	612	3297	0.000	6.959650
tourment	0.9533391	0.7535283	796	4440	0.000	6.954555
exploiter	0.6910511	0.5291670	577	3118	0.000	6.681127
colère	0.6850628	0.5242453	572	3089	0.000	6.667265
mordre	0.6886557	0.5284881	575	3114	0.000	6.616432
misogynie	0.9880713	0.7976539	825	4700	0.000	6.462243
...
famille	0.9876543	0.3657891	369	3800	0.000	6.213752

CLUSTER 8: CRIMINE PASSIONALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
passionnel	0.83850850	0.45568099	320	2685	0.000	10.339189
contusion	0.77300003	0.41019775	295	2417	0.000	10.278528
corps	0.76775935	0.40951890	293	2413	0.000	10.166897
mort	0.73631528	0.39051097	281	2301	0.000	10.037479
tragique	0.77300003	0.42139885	295	2483	0.000	9.876164
couteau	0.90663732	0.52424528	346	3089	0.000	9.740650
dévaster	0.72845426	0.39712980	278	2340	0.000	9.584000
fatalité	0.90139664	0.52848812	344	3114	0.000	9.483990
chantage	0.69701019	0.37829160	266	2229	0.000	9.436422
femme	0.66294579	0.35554997	253	2095	0.000	9.365236
horreur	0.71535257	0.39848751	273	2348	0.000	9.184898
dommage	0.88829495	0.52916698	339	3118	0.000	9.152761
danger	0.64984409	0.35215570	248	2075	0.000	9.129296
imploser	0.63412206	0.34282145	242	2020	0.000	9.049220
arme	0.62626104	0.33756033	239	1989	0.000	9.032632
...
mort	0.25478552	0.3589658	542	1458	0.000	6.589625

CLUSTER 9: DISTRUZIONE DELL'IDENTITA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
identità	0.77641381	0.27510573	148	1621	0.000	11.032425
exploitation	0.73969153	0.25677666	141	1513	0.000	10.949946
détruire	0.77641381	0.28206399	148	1662	0.000	10.801193
solitude	0.69772322	0.23997502	133	1414	0.000	10.712217
genre	0.86559647	0.33756033	165	1989	0.000	10.709890
mâle	0.65050887	0.21757282	124	1282	0.000	10.573290

annihilent	0.79215193	0.30192048	151	1779	0.000	10.458753
oppression	0.69772322	0.24947898	133	1470	0.000	10.373264
esclavage	0.87608855	0.35554997	167	2095	0.000	10.373058
ségrégation	0.59804847	0.19567977	114	1153	0.000	10.308729
isolation	0.54558808	0.18057526	104	1064	0.000	9.754289
relations	0.80789004	0.34282145	154	2020	0.000	9.520295
...
rabaisser	0.5874852	0.4785478	154	2014	0.000	9.859865

CLUSTER 10: DISCRIMINAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
discrimination	0.7638889	0.12253321	33	722	0.000	8.096237
travail	0.8101852	0.13967429	35	823	0.000	8.081578
harcèlement	0.6944444	0.10488300	30	618	0.000	7.911162
raillerie	0.8333333	0.16003992	36	943	0.000	7.794111
pouvoir	0.6018519	0.08180195	26	482	0.000	7.672425
protagonisme	0.8333333	0.18057526	36	1064	0.000	7.330042
intimidation	0.4861111	0.06262432	21	369	0.000	7.021036
cynisme	0.8333333	0.19567977	36	1153	0.000	7.013877
...
social0	0.245875	0.2457854	25	1547	0.000	7.005478

Tabella 6. Caratterizzazione dei cluster dei concetti negativi francesi

CLUSTER 1: RIDURRE LA DISPARITA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
empoderamiento	1.43680374	1.17641479	2239	5953	0.000	11.245.251
idealista	1.59017403	1.31731581	2478	6666	0.000	11.155.754
liderazgo	1.25327276	1.01654253	1953	5144	0.000	10.978.841
pareja	1.50161713	1.24202368	2340	6285	0.000	10.924.040
onu	1.52792751	1.27344480	2381	6444	0.000	10.584.506
ética	1.39188357	1.15250312	2169	5832	0.000	10.452.484
lealtad	1.23787155	1.01930917	1929	5158	0.000	10.134.803
moral	1.47273987	1.24064036	2295	6278	0.000	9.786695
principio	1.44001232	1.21396205	2244	6143	0.000	9.634751
económico	1.49584168	1.26573773	2331	6405	0.000	9.610767
fortalecer	1.52536064	1.29735648	2377	6565	0.000	9.411960
igualdads	1.43167000	1.21277634	2231	6137	0.000	9.337776
confianza	1.45669696	1.24202368	2270	6285	0.000	9.055563

interés	1.22632065	1.03729233	1911	5249	0.000	8.708358
recuperar	1.21926177	1.03571139	1900	5241	0.000	8.465144
eficiente	0.92792238	0.77050920	1446	3899	0.000	8.386492
normativas	1.41498537	1.22048341	2205	6176	0.000	8.283098
religión	1.47530674	1.28036140	2299	6479	0.000	8.112459
oficial	1.05241542	0.89283421	1640	4518	0.000	7.918156
hospitalidad	1.30396838	1.12799859	2032	5708	0.000	7.792110
innovar	1.26867396	1.10349407	1977	5584	0.000	7.396949
coraje	0.96064993	0.81991348	1497	4149	0.000	7.286412
lista	1.38803327	1.22087865	2163	6178	0.000	7.127937
programa	1.38418297	1.22245958	2157	6186	0.000	6.894025
religión	1.36814005	1.21317158	2132	6139	0.000	6.632862
sentimiento	1.38546640	1.23313091	2159	6240	0.000	6.469792
obtener	1.07936752	0.94678368	1682	4791	0.000	6.406727
...
aprobar	1.2012454	0.4578545	1254	4785	0.000	6.25689

CLUSTER 2: PROGRESSO SOCIALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
progreso	1.34619608	1.22680716	1655	6208	0.000	4.313141
estrategia	1.23150506	1.11851297	1514	5660	0.000	4.270407
unión	1.16968578	1.05982068	1438	5363	0.000	4.262850
compromiso	1.30471209	1.19242178	1604	6034	0.000	4.114626
nacer	1.27136222	1.16159351	1563	5878	0.000	4.074187
madrid	1.10949333	1.00745214	1364	5098	0.000	4.059809
seguridad	1.32911444	1.22127388	1634	6180	0.000	3.907069
perspectiva	1.26973540	1.16475538	1561	5894	0.000	3.892420
legalmente	1.02164488	0.93057908	1256	4709	0.000	3.768152
futuro	1.18839424	1.09064895	1461	5519	0.000	3.743153
experiencia	1.22418435	1.12760336	1505	5706	0.000	3.639485
necesidad	1.18432719	1.08966087	1456	5514	0.000	3.627488
unión	1.13552249	1.04420893	1396	5284	0.000	3.572843
españa	1.22662459	1.13214855	1508	5729	0.000	3.553625
recuperación	1.34050220	1.24202368	1648	6285	0.000	3.540573
medios	1.24858670	1.15388644	1535	5839	0.000	3.529274
padre	1.13958955	1.04954459	1401	5311	0.000	3.514752
europa	0.94030373	0.85923929	1156	4348	0.000	3.489170
género	1.31528644	1.22245958	1617	6186	0.000	3.364112
aumentar	1.20791612	1.12483672	1485	5692	0.000	3.136560

éxito	1.20303565	1.12266293	1479	5681	0.000	3.037578
canción	1.31691327	1.23313091	1619	6240	0.000	3.024665
...
estima	0.4578524	1.254785	1548	6256	0.000	3.20145

CLUSTER 3: EDUCARE ALLA CULTURA DELL'UGUAGLIANZA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
educación	1.1313849	0.9997451	1051	5059	0.000	4.378497
igualdad	1.1486086	1.0171354	1067	5147	0.000	4.336994
identidad	1.1604500	1.0283996	1078	5204	0.000	4.332853
cultura	1.1216965	0.9922356	1042	5021	0.000	4.322483
role	1.1249260	0.9959903	1045	5040	0.000	4.297354
ley	1.1561440	1.0272139	1074	5198	0.000	4.233971
social	0.8601109	0.7550951	799	3821	0.000	4.073991
respeto	1.1626029	1.0416399	1080	5271	0.000	3.948474
realidad	1.0129716	0.9172700	941	4563	0.000	3.895510
feminismo	1.0269659	0.9155602	954	4633	0.000	3.872624
público	1.1496851	1.0335376	1068	5230	0.000	3.870700
comprensión	1.0474191	0.9374957	973	4744	0.000	3.778310
equilibrio	1.0592605	0.9491551	984	4803	0.000	3.762044
campaña	1.0398837	0.9361124	966	4737	0.000	3.571309
video	1.0689488	0.9649645	993	4883	0.000	3.526565
papel	1.1679854	1.0643659	1085	5386	0.000	3.351841
vuelco	1.0032833	0.9127935	932	4619	0.000	3.155820
hombre	1.1938210	1.1428045	1109	5588	0.000	2.847702
policías	0.9365413	0.8576584	870	4340	0.000	2.837722
...
integral	1.2458752	0.854785	871	2458	0.000	2.58462

CLUSTER 4: LEGGE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
judicial	1.0462423	0.8189254	655	4144	0.000	6.487838
código	1.0590208	0.8333514	663	4217	0.000	6.390306
uniformar	0.9967255	0.7790067	624	3942	0.000	6.367602
tema	0.9759604	0.7618140	611	3855	0.000	6.331326
estabilidad	0.9504033	0.7402738	595	3746	0.000	6.299210
indispensablesir	0.9456114	0.7386928	592	3738	0.000	6.211991
forma	1.0430477	0.8311777	653	4206	0.000	6.017641
rebelarse	0.8913026	0.6987742	558	3536	0.000	5.943464
tanacidad	1.0142960	0.8122064	635	4110	0.000	5.809356

favor	1.0733967	0.8671440	672	4388	0.000	5.747724
revolución	1.0494369	0.8483704	657	4293	0.000	5.664454
correcto	1.0462423	0.8576584	655	4340	0.000	5.294157
honor	0.6405239	0.4960190	41	2510	0.000	5.273893
familia	0.7954636	0.6333629	498	3205	0.000	5.263578
sano	0.7874770	0.6278296	493	3177	0.000	5.207295
tramitar	1.1005511	0.9127935	689	4619	0.000	5.119587
...
legalidad	0.4578555	0.987545	547	4578	0.000	5.24578

CLUSTER 5:GIUSITZIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
justicia	0.86702606	0.52981153	386	2681	0.000	9.463639
proteger	0.90745732	0.56478977	404	2858	0.000	9.345348
reforzamiento	0.86028751	0.52823060	383	2673	0.000	9.340036
definir	0.92991914	0.58297054	414	2950	0.000	9.325999
institución	0.83108715	0.50906173	370	2576	0.000	9.221343
idea	0.86702606	0.54245903	386	2745	0.000	9.037752
internacional	0.76819407	0.46736452	342	2365	0.000	8.976461
único	0.90071878	0.57467062	401	2908	0.000	8.851704
confirmar	0.80413297	0.49997135	358	2530	0.000	8.809212
global	0.85804133	0.54226141	382	2744	0.000	8.808995
sociedad	0.76594789	0.47171210	341	2387	0.000	8.756969
victoria	0.73225517	0.45254323	326	2290	0.000	8.502145
consciente	0.93665768	0.62032018	417	3139	0.000	8.319317
comunidad	0.94339623	0.63336291	420	3205	0.000	8.087631
resultado	0.73899371	0.47151448	329	2386	0.000	8.007934
dios	0.66486972	0.41519360	296	2101	0.000	7.932808
...
referente	0.658956	0.4578542	458	2155	0.000	7.8548

CLUSTER 6:ATTIVISMO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
activo	0.8308575	0.34286573	165	1735	0.000	10.192245
ayuda	0.8610705	0.36183697	171	1831	0.000	10.189216
positivo	0.7956090	0.32211593	158	1630	0.000	10.161244
importancia	0.7553250	0.30116851	150	1524	0.000	10.045466
modelo	0.9063900	0.39641997	180	2006	0.000	10.026889
solución	0.8358930	0.35788463	166	1811	0.000	9.845054
vida	0.7100055	0.28002348	182	1417	0.000	9.838123

concreto	0.9214966	0.41519360	183	2101	0.000	9.785746
ventaja	0.6596505	0.25986653	184	1315	0.000	9.489323
asegurar	0.9315676	0.43456007	185	2199	0.000	9.454706
seguir	0.9366031	0.45254323	186	2290	0.000	9.086065
...
acciòn	0.6719834	0.36985274	187	1456	0.000	9.348627

CLUSTER 7: AZIONE E COOPERAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
apoyo	0.8507223	0.19781475	53	1001	0.000	8.621564
grupo	0.8025682	0.17726257	50	897	0.000	8.611128
acción	0.8828250	0.21816931	55	1104	0.000	8.481977
atención	0.7383628	0.15493183	46	784	0.000	8.481652
cambio	0.6581059	0.13200824	41	668	0.000	8.186543
precaución	0.8988764	0.23951196	56	1212	0.000	8.175333
esperanza	0.8988764	0.25986653	56	1315	0.000	7.750784
...
mundial	0.3578914	0.45612897	56	1313	0.000	8.346791

CLUSTER 8: LIBERTÀ

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
libertad	0.7832898	0.06857315	9	347	0.000	5.133672
mundo	0.8703220	0.09129912	10	462	0.000	5.118341
rosa	0.8703220	0.11145606	10	564	0.000	4.766442
nacional	0.6092254	0.04485909	7	227	0.000	4.731959
niña	0.8988764	0.25986653	6	115	0.000	4.750784
vida	0.7988764	0.25986653	7	1315	0.000	2.750784
participar	0.5788764	0.45986653	6	1215	0.000	2.450784
capacidad	0.6988764	0.58986653	8	1315	0.000	2.350784

Tabella 7. Caratterizzazione dei cluster dei concetti positivi spagnoli

CLUSTER 1: MISOGINIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
misoginia	1.04647595	0.85141678	2865	12205	0.000	12.018.721
cosificación	1.11879785	0.91678160	3063	13142	0.000	12.011.059
machista	1.11952837	0.91908366	3065	13175	0.000	11.905.634
patriarcado	0.74549997	0.58444652	2041	8378	0.000	11.900.674
feminicidio	1.11295366	0.91440977	3047	13108	0.000	11.823.674
propagación	1.06839168	0.88176224	2925	12640	0.000	11.322.171

esteriotipo	1.00921922	0.83237239	2763	11932	0.000	11.038.097
salarial	0.97415405	0.80091078	2667	11481	0.000	11.016.957
promover_desigualdad	1.01725498	0.84311538	2785	12086	0.000	10.806.995
acoso	1.01835077	0.84876592	2788	12167	0.000	10.496.950
maltrato	1.08884636	0.91392145	2981	13101	0.000	10.447.764
importancia	0.87845538	0.72382635	2405	10376	0.000	10.341.306
psoe	1.03661387	0.86815911	2838	12445	0.000	10.317.320
brecha	0.96684881	0.80600324	2647	11554	0.000	10.214.195
femenino	1.08446321	0.91371217	2969	13098	0.000	10.204.449
taller	0.86640173	0.71971053	2372	10317	0.000	9.848125
disposición	0.51319327	0.40390849	1405	5790	0.000	9.705436
pacto	0.89233534	0.74552160	2443	10687	0.000	9.692779
agresión	0.91753843	0.76993749	2512	11037	0.000	9.595722
custodia	0.53766583	0.43118453	1472	6181	0.000	9.176016
hostil	0.72285372	0.61297823	1979	8787	0.000	8.007253
marido	0.76668517	0.65462475	2099	9384	0.000	7.911402
aborto	0.94237625	0.82337340	2580	11803	0.000	7.520791
pardo_devera	0.92813103	0.81137473	2541	11631	0.000	7.433062
europa	0.73819473	0.63585940	2021	9115	0.000	7.337378
agustín_laje	0.88247326	0.77237908	2416	11072	0.000	7.182793
...
masculino	0.8703220	0.09129912	2510	46214	0.000	5.118341

CLUSTER 2: MANIPOLAZIONE MEDIATICA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
medios	1.029823606	0.91371217	2894	13098	0.000	7.098030
disprezio	0.933388846	0.82400123	2623	11812	0.000	7.033306
falsas	0.950469541	0.84088307	2671	12054	0.000	6.977536
mauricio_macri	0.978581519	0.86920550	2750	12460	0.000	6.854194
subversión	0.875029802	0.77237908	2459	11072	0.000	6.814618
matar	1.025197585	0.91392145	2881	13101	0.000	6.805400
situación	0.970041172	0.86215977	2726	12359	0.000	6.788270
atención	0.966838541	0.86041578	2717	12334	0.000	6.704182
sufrir	1.023774193	0.91440977	2877	13108	0.000	6.688199
feminismo	0.796387433	0.70171253	2238	10059	0.000	6.589342
cólera	0.911326280	0.81137473	2561	11631	0.000	6.482180
capitalismo	1.021283258	0.91678160	2870	13142	0.000	6.386311
tema	0.939082411	0.83983668	2639	12039	0.000	6.330870
contraste	0.920578324	0.82337340	2587	11803	0.000	6.261733

orden	0.860795889	0.76805398	2419	11010	0.000	6.181550
izquierda	0.864710215	0.77203028	2430	11067	0.000	6.162102
llama	0.902430085	0.81004930	2536	11612	0.000	6.001586
access	0.756176629	0.67304131	2125	9648	0.000	5.913649
cárcel	0.841580107	0.75459036	2365	10817	0.000	5.852288
mastro	0.943352585	0.85141678	2651	12205	0.000	5.831014
exploitation	0.723794477	0.64485840	2034	9244	0.000	5.735701
policías	0.721659390	0.64743951	2028	9281	0.000	5.386470
rasgón	0.774324868	0.70017782	2176	10037	0.000	5.181135
seguro	0.929118672	0.84876592	2611	12167	0.000	5.110932
cabeza	0.845494433	0.77000725	2376	11038	0.000	5.036455
...
ciberacoso	0.422785755	0.84575551	2547	14572	0.000	5.235698

CLUSTER 3: SESSO

FEATURES	intern %	glob %	Frecuencia Interna	Frecuencia Globale	p.value	v.test
sexo	0.7411492	0.4198835	336	6019	0.000	9.711526
paula_fraga	0.5382155	0.2800851	244	4015	0.000	9.386970
abolicionista	0.7212970	0.4209996	327	6035	0.000	9.115379
porno	0.8536451	0.5434976	387	7791	0.000	8.427028
desafío	0.6661520	0.4039085	32	5790	0.000	8.181366
enseñar	0.2536671	0.1161499	115	1665	0.000	7.547766
desnudo	0.6793868	0.4380907	38	6280	0.000	7.310348
retirar	0.3132238	0.1827006	142	2619	0.000	5.987240
redundancia	0.5315981	0.3572393	241	5121	0.000	5.884368
...
reproductiva	0.685478	0.325698	125	5478	0.000	4.23563

CLUSTER 4: SFRUTTAMENTO DEL CORPO

FEATURES	intern %	glob %	Frecuencia Interna	Frecuencia Globale	p.value	v.test
cuerpo	0.86836202	0.78563341	2435	11262	0.000	5.453793
medida	0.78099090	0.70443316	2190	10098	0.000	5.32496
público	0.79953497	0.72361707	2242	10373	0.000	5.212625
lgbt	0.69718594	0.62769752	1955	8998	0.000	5.115944
campaña	0.82378492	0.74810271	2310	10724	0.000	5.113515
estúpidas	0.81807905	0.7430126	2294	10651	0.000	5.089287
cultura	0.88191346	0.8068435	2473	11566	0.000	4.890230
silencio	0.79061955	0.72061740	2217	10330	0.000	4.819919
polémico	0.83519666	0.763388	2342	10943	0.000	4.807362
manos	0.71180723	0.64799758	1996	9289	0.000	4.629842

scapar	0.83163049	0.76268248	2332	10933	0.000	4.619025
problema	0.58913103	0.53184773	1652	7624	0.000	4.578990
españa	0.83412681	0.76596119	2339	10980	0.000	4.557542
crimen	0.86336939	0.79553929	2421	11404	0.000	4.452568
educación	0.78455706	0.72061740	2200	10330	0.000	4.405947
lídere	0.59875969	0.54384639	1679	7796	0.000	4.344022
revolución	0.82307168	0.75856666	2308	10874	0.000	4.335088
titular	0.78384383	0.72166380	2198	10345	0.000	4.282622
lucha	0.81486950	0.75145118	2285	10772	0.000	4.282189
defensores	0.88084361	0.81779262	2470	11723	0.000	4.085872
enemigo	0.79596880	0.73666213	2232	10560	0.000	4.045461
jaula	0.82556800	0.76568215	2315	10976	0.000	4.008575
periodista	0.69647270	0.64297489	1953	9217	0.000	3.901930
policía	0.77599826	0.72047788	2176	10328	0.000	3.830094
techo	0.86622232	0.81102593	2429	11626	0.000	3.594312
seguir	0.77136224	0.72033836	2163	10326	0.000	3.522027
saca	0.55703552	0.51510541	1562	7384	0.000	3.412781
juicio	0.78954970	0.73966179	2214	10603	0.000	3.399815

CLUSTER 5: VIOLENZA FAMILIARE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
familia	0.82182591	0.7114091	1617	10198	0.000	6.136669
casa	0.77608421	0.6689952	1527	9590	0.000	6.132087
opresión	0.76642762	0.6613217	1508	9480	0.000	6.053765
violencia	0.75982049	0.6558107	1495	9401	0.000	6.015712
género	0.74813094	0.6466721	1472	9270	0.000	5.910181
abuso	0.72017768	0.6217679	1417	8913	0.000	5.844142
signo	0.72830954	0.6309065	1433	9044	0.000	5.745254
privación	0.81623525	0.7129438	1606	10220	0.000	5.741241
desfavorable	0.73694964	0.6401845	1450	9177	0.000	5.668602
destrucción	0.77100179	0.6728320	1517	9645	0.000	5.614424
pareja	0.69069970	0.6015376	1359	8623	0.000	5.388363
verdugo	0.65715578	0.5703551	1293	8176	0.000	5.383292
sexual	0.82436711	0.7268958	1622	10420	0.000	5.372578
demolición	0.61954594	0.5365216	1219	7691	0.000	5.305795
pelea	0.63276021	0.5491481	1245	7872	0.000	5.283642
terror	0.68104311	0.5944919	1340	8522	0.000	5.262689
perspectiva	0.71102934	0.6259535	1399	8973	0.000	5.048070
voluntad	0.81166108	0.7206872	1597	10331	0.000	5.040139

comunidad	0.69730683	0.6132573	1372	8791	0.000	5.037342
obsesión	0.72576833	0.6414402	1428	9195	0.000	4.946011
historia	0.83605666	0.7453821	1645	10685	0.000	4.943113
...
obstáculo	0.69069970	0.6015376	1359	8623	0.000	3.388363

CLUSTER 6: COMPORTAMIENTO PSICOPATOLOGICO

FEATURES	intern %	glob %	Frecuencia Interna	Frecuencia Globale	p.value	v.test
patológico	0.6800622	0.5192910	112	7444	0.000	9.169728
enfermedad	0.6961072	0.5337312	1128	7651	0.000	9.141756
mental	0.7263459	0.5607282	1177	8038	0.000	9.108185
odio	0.7096838	0.5467065	1150	7837	0.000	9.073336
ideología	0.6689541	0.5111989	1084	7328	0.000	9.069077
refutar	0.7460937	0.5808190	129	8326	0.000	8.943207
atrocés	0.7454765	0.5814469	128	8335	0.000	8.873590
prisión	0.6313100	0.4823184	1023	6914	0.000	8.815552
delito	0.6467379	0.4957820	1048	7107	0.000	8.815090
social	0.6498235	0.5029672	1053	7210	0.000	8.526855
sangre	0.5800894	0.4449969	940	6379	0.000	8.323890
madrid	0.7214090	0.5705643	1169	8179	0.000	8.253926
fiera	0.5825578	0.4483454	944	6427	0.000	8.242927
derechos	0.6053911	0.4710173	981	6752	0.000	8.066296
nacional	0.7423910	0.5935153	123	8508	0.000	8.001129
grito	0.5936659	0.4618090	962	6620	0.000	7.992854
latido	0.7257288	0.5815166	1176	8336	0.000	7.831888
importante	0.7201748	0.5807493	1167	8325	0.000	7.583858
falta	0.7615216	0.6259535	1234	8973	0.000	7.124873
prostituta	0.7460937	0.6132573	129	8791	0.000	7.052465
niño	0.5091210	0.4057920	825	5817	0.000	6.701772
...
salud	0.55703552	0.51510541	1562	6384	0.000	3.412781

CLUSTER 7: BRUTALITA'

FEATURES	intern %	glob %	Frecuencia Interna	Frecuencia Globale	p.value	v.test
trágico	0.64251096	0.40593153	696	5819	0.000	11.805170
animal	0.61573967	0.38528266	667	5523	0.000	11.782351
torturas	0.61020078	0.38130636	661	5466	0.000	11.760223
trampa	0.59543042	0.37007505	645	5305	0.000	11.740802
cadáver	0.57881375	0.36044822	627	5167	0.000	11.530292
matanza	0.57327487	0.35633240	621	5108	0.000	11.517010

exclusión	0.55665820	0.34586845	603	4958	0.000	11.356461
régimen	0.63143319	0.40579201	684	5817	0.000	11.293624
sociedad	0.60466190	0.38493386	655	5518	0.000	11.271613
indiferencia	0.67112855	0.43990449	727	6306	0.000	11.155467
abandono	0.53173321	0.32961445	576	4725	0.000	11.148319
comprobar	0.59635357	0.38228300	646	5480	0.000	11.031948
incorrecto	0.68497577	0.45908839	742	6581	0.000	10.709198
indigno	0.50773136	0.31817386	550	4561	0.000	10.660020
piojoso	0.50496192	0.31622059	547	4533	0.000	10.645391
patología	0.60466190	0.39602565	655	5677	0.000	10.603064
reprimir	0.70343873	0.48469019	762	6948	0.000	10.141483
condenado	0.47265174	0.30115250	512	4317	0.000	9.939056
psicosi	0.48188322	0.31043054	522	4450	0.000	9.805779
clase	0.67759058	0.47101730	734	6752	0.000	9.727897
...
intimidación	0.36785412	0.36914782	666	5492	0.000	9.271613

CLUSTER 8: VIOLENZA DOMESTICA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
doméstica	0.62248025	0.27806205	386	3986	0.000	14.400202
vergonzoso	0.56603774	0.24297293	351	3483	0.000	14.331371
anormales	0.57893888	0.25608775	359	3671	0.000	14.035070
culpa	0.56281245	0.24743755	349	3547	0.000	13.929470
dolor	0.53055959	0.22720725	329	3257	0.000	13.906596
locura	0.62570553	0.28929336	388	4147	0.000	13.884008
clima	0.65150782	0.30729135	404	4405	0.000	13.839712
asesinato	0.59345267	0.27038848	368	3876	0.000	13.749209
monstruo	0.50153201	0.21199964	311	3039	0.000	13.702737
irracional	0.65795839	0.31622059	408	4533	0.000	13.594917
violación	0.49024351	0.20760478	304	2976	0.000	13.521293
especie	0.46444122	0.19504804	288	2796	0.000	13.271623
invisible	0.64344461	0.31817386	399	4561	0.000	12.970473
conciencia	0.46927915	0.20760478	291	2976	0.000	12.626043
degradante	0.40638607	0.16888816	252	2421	0.000	12.542225
...
extremo	0.34896171	0.21735846	525	1242	0.000	11.684357

CLUSTER 9: DISCRIMINAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
discriminando	0.57394488	0.14440252	116	2070	0.000	12.237966

machismo	0.52446687	0.12298630	106	1763	0.000	12.183856
homofobia	0.49478007	0.11154571	100	1599	0.000	12.093843
burlarse	0.60363169	0.16428403	122	2355	0.000	11.956235
racismo	0.48983227	0.11580105	99	1660	0.000	11.717066
mujer	0.61352729	0.17523629	124	2512	0.000	11.679234
hombre	0.46509327	0.10645326	94	1526	0.000	11.624574
implacable	0.53931028	0.14203069	109	2036	0.000	11.533127
critica	0.42056306	0.09438483	85	1353	0.000	11.169760
descuido	0.61847509	0.19504804	125	2796	0.000	10.938910
ineficaz	0.37603285	0.08287449	76	1188	0.000	10.660941
...
galopante	0.56603774	0.24297293	51	1483	0.000	11.331371

CLUSTER 10: FALLACIA DEL SISTEMA GIURIDICO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
falacia	0.5550930	0.06020260	20	863	0.000	7.240428
penal	0.4995837	0.04813417	18	690	0.000	7.132906
código	0.5828476	0.07178270	21	1029	0.000	7.098260
regulacion	0.5828476	0.08287449	21	1188	0.000	6.719825
jurídico	0.4163197	0.03613551	15	518	0.000	6.709748
sistema	0.5828476	0.09438483	21	1353	0.000	6.366852

Tabella 8. Caratterizzazione dei cluster dei concetti negativi spagnoli

CLUSTER 1: VALORE DELL'IDENTITA'

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
ідентичність	0.86573734	0.20484773	59	830	0.000	9.050500
цінності	0.86573734	0.20484773	59	830	0.000	9.050500
пріоритети	0.82171680	0.18757142	56	760	0.000	8.994622
можливий	0.74834923	0.16017612	51	649	0.000	8.883023
досвід	0.66030814	0.13204041	45	535	0.000	8.634700
колективний	0.92443140	0.25914472	63	1050	0.000	8.430698
разом	0.55759354	0.10810037	38	438	0.000	8.047008
подумав	0.89508437	0.26358720	61	1068	0.000	8.022004
гроші	0.93910492	0.31097367	64	1260	0.000	7.543634
довіру	0.92443140	0.31492254	63	1276	0.000	7.318174
пожавите	0.44020543	0.08169229	30	331	0.000	7.287172
...
сам	0.52847489	0.04785695	25	1254	0.000	6.2358966

CLUSTER 2: PROGRESSO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
зміна	0.814674939	0.42647817	302	1728	0.000	10.807938
проект	0.849743728	0.45387347	315	1839	0.000	10.723477
влада	0.954950094	0.53655298	354	2174	0.000	10.521958
Україна	0.930671702	0.52174471	345	2114	0.000	10.423269
закон	0.855138926	0.46794133	317	1896	0.000	10.373384
необхідність	0.714863771	0.36872592	265	1494	0.000	10.330896
принцип	0.882114918	0.48990693	327	1985	0.000	10.297279
європейський	0.906393310	0.50792366	336	2058	0.000	10.291791
необхідний	0.690585379	0.35811333	256	1451	0.000	10.076667
повага	0.647423793	0.33244566	240	1347	0.000	9.888079
час	0.604262207	0.31097367	224	1260	0.000	9.520190
Порошенко	0.930671702	0.55456971	345	2247	0.000	9.395815
трансформація	0.604262207	0.31492254	224	1276	0.000	9.353376
свідомість	0.995414081	0.61750486	369	2502	0.000	9.011274
підтвердження	0.941462099	0.57850975	349	2344	0.000	8.924364
Павло_Клімкін	0.507148638	0.25914472	188	1050	0.000	8.799920
щире_закінчення	0.509846237	0.26358720	189	1068	0.000	8.681821
негайно	1.033180469	0.66118925	383	2679	0.000	8.618683
сьогодн	0.774210952	0.47781351	287	1936	0.000	8.015772
без_вини	0.404639871	0.20484773	150	830	0.000	7.951031
...
стан	0.2457859511	0.35698566	154	874	0.000	7.365896

CLUSTER 3: RIDUZIONE DELLA DISUGUAGLIANZA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
незалежність	1.06231643	0.95587382	1501	3873	0.000	5.040366
рівність	0.94907817	0.84900748	1341	3440	0.000	5.022207
економічний,	0.94270852	0.84382458	1332	3419	0.000	4.977761
свобода	0.98234191	0.88133887	1388	3571	0.000	4.977172
обізнаність	0.91652217	0.81914413	1295	3319	0.000	4.973835
новини	0.91227574	0.81963774	1289	3321	0.000	4.731305
прогрес	0.95403234	0.86011368	1348	3485	0.000	4.684845
воля	0.94624721	0.85320315	1337	3457	0.000	4.659704
революція	1.12955165	1.02991517	1596	4173	0.000	4.550471
світ	1.09416469	0.99634976	1546	4037	0.000	4.540423
визначенн	0.98234191	0.88997702	1388	3606	0.000	4.531416
очікування	0.83796313	0.75349413	1184	3053	0.000	4.496704

здатність	0.83159347	0.74855804	1175	3033	0.000	4.434876
вірність	0.83017800	0.74781763	1173	3030	0.000	4.401070
включеність	0.82239287	0.74238793	1162	3008	0.000	4.290862
президент	0.79549878	0.71820109	1124	2910	0.000	4.213823
дати	0.84220956	0.76410673	1190	3096	0.000	4.130449
захист	0.73321774	0.66118925	1036	2679	0.000	4.089585
новий	0.74100287	0.66859339	1047	2709	0.000	4.088809
втручатися	0.77355887	0.69969075	1093	2835	0.000	4.079160
відновити	0.80116069	0.72634564	1132	2943	0.000	4.056397
простір	1.13096713	1.04200859	1598	4222	0.000	4.040922
завзятість	0.96252521	0.88084526	1360	3569	0.000	4.028870
законодавчий	0.85919530	0.78261707	1214	3171	0.000	4.002849
глобальний	0.68013730	0.61750486	961	2502	0.000	3.678284
варті_зусиль	0.36377791	0.31887141	514	1292	0.000	3.647797
визначеність	0.84999469	0.78088943	1201	3164	0.000	3.616815
потреба	1.05240808	0.97537138	1487	3952	0.000	3.615264
київ	1.12035104	1.04151498	1583	4220	0.000	3.582652
профілактика	0.67022895	0.61034753	947	2473	0.000	3.537101
соціальний	1.18617078	1.10617776	1676	4482	0.000	3.529567
свобода	0.74949574	0.68735053	1059	2785	0.000	3.462904
легітимний	0.92855374	0.85986687	1312	3484	0.000	3.429130
пояснити	0.63130330	0.57850975	892	2344	0.000	3.201741
підтримка	1.17130826	1.10025446	1655	4458	0.000	3.143570
психологічний	1.04320747	0.97882664	1474	3966	0.000	3.016329
розгортається	1.16706182	1.10469694	1649	4476	0.000	2.753499
сам	1.09274921	1.03731931	1544	4203	0.000	2.523590
воля	1.17838565	1.12172645	1665	4545	0.000	2.482497
добро	0.59450087	0.55456971	840	2247	0.000	2.471510
справа	0.51452635	0.47781351	727	1936	0.000	2.444544
національний	1.14229095	1.08964186	1614	4415	0.000	2.339486

CLUSTER 4: COOPERAZIONE PER LA TRASFORMAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
співтовариство	0.67901425	0.58467986	1322	2369	0.000	7.554117
автономія	1.19315439	1.07952288	2323	4374	0.000	6.713547
захист	1.36316476	1.24512870	2654	5045	0.000	6.499282
толерантність	1.34107882	1.22612475	2611	4968	0.000	63.77447
цілісність	1.20085878	1.09704600	2338	4445	0.000	6.083328
виживши	1.19058625	1.08840784	2318	4410	0.000	6.010761

чітко	1.19264076	1.09136949	2322	4422	0.000	5.949253
ефективно	1.22089022	1.12222006	2377	4547	0.000	5.716699
вчити	1.31437024	1.21304411	2559	4915	0.000	5.649501
Російська_Федерація	1.29279793	1.19280614	2517	4833	0.000	5.621434
гідність	0.67541886	0.60541144	1315	2453	0.000	.5502455
культура	1.34261970	1.24488189	2614	5044	0.000	5.379671
виховувати	1.10224249	1.01461329	2146	4111	0.000	5.334801
корисно	0.51670827	0.45806915	1006	1856	0.000	5.291268
жест	1.08991546	1.00597514	2122	4076	0.000	5.131231
відкликатись	0.33026185	0.28555281	643	1157	0.000	5.099511
вирішити	0.90090090	0.82531424	1754	3344	0.000	5.094947
бачити	1.16798669	1.08199092	2274	4384	0.000	5.071177
головне	0.83721121	0.76805560	1630	3112	0.000	4.829166
допомога	0.69236854	0.63058549	1348	2555	0.000	4.756098
релігія	0.99489455	0.92329563	1937	3741	0.000	4.564276
надія	0.50489486	0.45609471	983	1848	0.000	4.408557
потужність	0.83721121	0.77718737	1630	3149	0.000	4.164636
стабільний	0.47202276	0.42746539	919	1732	0.000	4.155159
початок	1.23989440	1.16787889	2414	4732	0.000	4.087176
чоловік	0.55831202	0.51063851	1087	2069	0.000	4.070746
Зеленський	0.79355296	0.73646462	1545	2984	0.000	4.067269
приспосуватися	1.21113131	1.14048359	2358	4621	0.000	4.056574
слуга	0.54701223	0.50002591	1065	2026	0.000	4.053888
гордий	0.58142521	0.53334452	1132	2161	0.000	4.017799
обирати	1.05396160	0.99092006	2052	4015	0.000	3.878758
реакція	0.99489455	0.93563585	1937	3791	0.000	3.750166
місце	0.15357433	0.13129999	299	532	0.000	3.722546
прояв	1.13870998	1.07754844	2217	4366	0.000	3.609872
управління	0.42939176	0.39315957	836	1593	0.000	3.517770
Росія	1.04471632	0.98795841	2034	4003	0.000	3.495829
країна	0.07653035	0.06244154	149	253	0.000	3.392968
конвенція	1.17877284	1.12246686	2295	4548	0.000	3.255779
в.о.	0.40833308	0.37588325	795	1523	0.000	3.219175
сумісні	0.98359477	0.93316781	1915	3781	0.000	3.193128
спілкування	0.91476882	0.86677740	1781	3512	0.000	3.151252
українська	0.27016754	0.24483006	526	992	0.000	3.105597
нейтральна	0.80228461	0.75818342	1562	3072	0.000	3.093066
нейтральна	0.30098514	0.27444660	586	1112	0.000	3.074121

пам'ять	1.08529282	1.03805972	2113	4206	0.000	2.836411
нейтральна	0.46175023	0.43116746	899	1747	0.000	2.832244
міністерство_соціальної_політики	0.67952787	0.64243211	1323	2603	0.000	2.821679
подія	0.57217993	0.53828061	1114	2181	0.000	2.813611
команда	1.20445417	1.15775990	2345	4691	0.000	2.656673
російська	1.03547105	0.99264769	2016	4022	0.000	2.627733
рятувальна	0.15151982	0.13549567	295	549	0.000	2.623173
міжнародна	0.89422376	0.85616481	1741	3469	0.000	2.511010
виявити	0.28711722	0.26580845	559	1077	0.000	2.502021
напрямок	0.13354289	0.11920657	260	483	0.000	2.497120
сили	0.36005218	0.33664134	701	1364	0.000	2.446166
експерт	0.27581744	0.25593627	537	1037	0.000	2.376813
відповідь	0.47305002	0.44720975	921	1812	0.000	2.346401
прибуток	1.19469527	1.15381103	2326	4675	0.000	2.328309
марш	1.14795525	1.10790539	2235	4489	0.000	2.326708
пропозиція	0.36929746	0.34725393	719	1407	0.000	2.266445
дзвінок	0.61840632	0.58986275	1204	2390	0.000	2.260658
жест	0.67233710	0.64366613	1309	2608	0.000	2.174499

CLUSTER 5: ATTIVISMO

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
ОБСЕ	0.56877713	0.13549567	76	549	0.000	10.396611
інституційна_діяльність	0.98039216	0.44720975	131	1812	0.000	8.120752
прапор	0.53884149	0.17794604	72	721	0.000	8.119742
вшанування_пам'яті	0.44155067	0.13129999	59	532	0.000	7.933104
дипломат	0.97290825	0.45806915	130	1856	0.000	7.796009
сім'я	0.40413112	0.11920657	54	483	0.000	7.627412
комітет	0.47148630	0.16066973	63	651	0.000	7.401116
власність	0.65858404	0.28925487	88	1172	0.000	6.921669
юрист	0.72593923	0.33664134	97	1364	0.000	6.840882
лінія	0.44155067	0.16239736	59	658	0.000	6.714017
червоний	0.46400239	0.17918007	62	726	0.000	6.589971
сфера	0.62864841	0.28555281	84	1157	0.000	6.512119
позитивно	0.38916330	0.14043176	52	569	0.000	6.400615
сторінка	0.56129322	0.27444660	75	1112	0.000	5.627243
...
центр_боротьби_з_насильством	0.58968966	0.21457899	74	1145	0.000	5.23698

CLUSTER 6: SICUREZZA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
----------	----------	--------	-------------------	-------------------	---------	--------

безпека	1.6553800	0.1244060	44	488	0.000	12.316498
збереження	2.1068473	0.24433645	56	990	0.000	12.024450
розслідування	1.5425132	0.11278966	41	457	0.000	11.866601
дівчино	1.6930023	0.17843965	45	723	0.000	11.117532
...
громадськості	1.2365899	0.1475896	40	748	0.000	10.235698

CLUSTER 7: FIDUCIA NELLE ISTITUZIONI

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
інститут	1.3262599	0.09427932	45	382	0.000	12.574952
автономія	1.6209844	0.16486541	55	668	0.000	12.481077
засоби	1.1199528	0.06244154	38	253	0.000	12.318829
тріумф	1.6504568	0.20213288	56	819	0.000	11.807430
спільність	1.3852048	0.19522236	47	791	0.000	10.240656
рівний	0.9136457	0.11846616	31	480	0.000	8.559442
...
дію	1.368956	0.125478	40	587	0.000	8.235688

CLUSTER 8: COMUNICAZIONE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
спілкуватися	0.5384450	0.05133534	25	208	0.000	8.577712
говорити	0.6245962	0.07453496	29	302	0.000	8.563651
шанувати	1.0338143	0.25593627	48	1037	0.000	7.905071
...
орієнтація	1.2568963	0.254789	41	1025	0.000	6.235698

CLUSTER 9: LEGGE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
правові_радикальні	3.52281825	0.05133534	44	208	0.000	17.161167
радикальні	3.52281825	0.07280733	44	295	0.000	16.165668
затверджують	3.76301041	0.10958120	47	444	0.000	15.679950
зміни	3.04243395	0.07453496	38	302	0.000	14.550171
...
легальність	3.25989655	0.04785651	45	201	0.000	14.235698

Tabella 9. Caratterizzazione dei cluster dei concetti positivi ucraini

CLUSTER 1: AUTORITA' MASCHILE						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
патріархат	1.137186730	0.49305984	157	687	0.000	9.894536
дискримінація	1.339997103	0.63085823	185	879	0.000	9.782318
позбавлення	1.318267420	0.63444673	182	884	0.000	9.441424
погіршення	1.209619006	0.56195903	167	783	0.000	9.437000
авторитарність	0.905403448	0.37535706	125	523	0.000	9.243488
підбурювання	1.361726785	0.68181492	188	950	0.000	9.120147
агітація	0.840214400	0.34234286	116	477	0.000	9.058784
гроші	0.832971172	0.33803666	115	471	0.000	9.055547
розчарування	0.818484717	0.33157736	113	462	0.000	8.991132
проблема	1.419672606	0.75860881	196	1057	0.000	8.503785
байдужість	0.637404027	0.24904187	88	347	0.000	8.206853
падіння	1.448645516	0.80023541	200	1115	0.000	8.168483
гнітюче	0.796755034	0.35095526	110	489	0.000	8.118572
погано	1.535564247	0.87774700	212	1223	0.000	7.962412
шахрайство	1.086484137	0.55765283	150	777	0.000	7.863020
кислота	0.564971751	0.21818077	78	304	0.000	7.805853
Україна	0.521512386	0.20526218	72	286	0.000	7.358578
хворі	1.542807475	0.93085679	213	1297	0.000	7.258110
погода	0.434593655	0.15932938	60	222	0.000	7.167382
фатальність	1.542807475	0.99544978	213	1387	0.000	6.346776
дурний	0.311458786	0.11124349	43	155	0.000	6.187835
переляк	1.419672606	0.96602409	196	1346	0.000	5.379903
ціна	0.260756193	0.10119569	36	141	0.000	5.232124
...
садизм	0.25874555	0.10254785	25	145	0.000	5.2365896

CLUSTER 2: SESSO						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
секс	5.7450628	0.09832489	32	137	0.000	2.4315254
жорстокий	5.5655296	0.09760719	31	136	0.000	2.4025618
огидний	5.0269300	0.10478419	28	146	0.000	2.2913034
нелегітимний	4.4883303	0.09904259	25	138	0.000	2.2060995
...
Безпорадний	4.54785125	0.0258745	21	145	0.000	2.1025478

CLUSTER 3: BRUTALITA'						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
зневіра	1.37764033	1.31985014	1519	1839	0.000	3.725684

переслідування	1.46561341	1.40884493	1616	1963	0.000	3.536791
тварина	0.85433653	0.81243630	942	1132	0.000	3.436204
гнів	1.39215135	1.34281654	1535	1871	0.000	3.136888
темний_радж	0.50607196	0.47727044	558	665	0.000	3.072668
зневажливий	1.50914648	1.45980163	1664	2034	0.000	3.006419
криза	1.35496685	1.30980235	1494	1825	0.000	2.900846
прикриття	0.59857973	0.56913603	660	793	0.000	2.865434
роки	1.66604693	1.61697791	1837	2253	0.000	2.837687
крайній	1.24432029	1.20214736	1372	1675	0.000	2.824279
покарання	0.83529081	0.80310621	921	1119	0.000	2.627787
вогонь	1.42570890	1.38444314	1572	1929	0.000	2.569841
порушувати	1.09830312	1.06219587	1211	1480	0.000	2.563539
ганьба	1.36494318	1.32487404	1505	1846	0.000	2.549630
маніяк	1.77578654	1.73037450	1958	2411	0.000	2.532661
необхідне,	0.72645813	0.69760432	801	972	0.000	2.522997
непристойність	1.30689980	1.26817575	1441	1767	0.000	2.516919
зброя	1.44566075	1.40525643	1594	1958	0.000	2.495771
підлеглий	1.53000680	1.48850962	1687	2074	0.000	2.491370
тюрма	0.23761820	0.22176927	262	309	0.000	2.452495
вирок	1.01758555	0.98611968	1122	1374	0.000	2.309999
колоты	1.53091320	1.49281582	1688	2080	0.000	2.278661
необачність	0.92145010	0.89210099	1016	1243	0.000	2.262461
воля	0.31833559	0.30143396	351	420	0.000	2.230147
раб	0.98856350	0.95956479	1090	1337	0.000	2.152771
з'валтування	1.10827945	1.07798527	1222	1502	0.000	2.122584

CLUSTER 4: PSICOPATOLOGIA

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
відхилення	1.92414947	0.26554897	69	370	0.000	12.823026
психічний	2.1193534	0.43277305	76	603	0.000	11.267259
хвороба,	2.00780814	0.40191195	72	560	0.000	11.074169
психопат	1.89626325	0.42846685	68	597	0.000	10.086208
збочення	2.20301171	0.59569093	79	830	0.000	9.779997
вибух	0.78081428	0.21100378	28	294	0.000	5.757572
звільнення	1.84049080	0.96458869	66	1344	0.000	4.802163
...
марення	1.23589663	0.96589652	55	1245	0.000	4.235698

CLUSTER 5: INEFFICACIA ISTITUZIONALE

FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
----------	----------	--------	-------------------	-------------------	---------	--------

Інституціональність	1.02782392	0.33947206	99	473	0.000	9.842693
Федерація	0.64368771	0.17798958	62	248	0.000	8.844148
Усамітнення	1.24584718	0.53109794	120	740	0.000	8.597945
...
уряд	1.2547856	0.25478552	124	147	0.000	8.2569856
CLUSTER 6: MISOGINIA						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
мізогінія	3.5958904	0.1779896	21	248	0.000	9.328976
різанина	3.0821918	0.1162674	18	162	0.000	9.166986
автономія	2.0547945	0.1248798	12	174	0.000	6.641835
сварка	2.9109589	0.3394721	17	473	0.000	6.531743
скраб	1.8835616	0.1995206	11	278	0.000	5.382713
...
повія	1.25478565	0.2014578	10	478	0.000	5.2356985
CLUSTER 7: STRUMENTALIZZAZIONE MEDIATICA						
FEATURES	intern %	glob %	Frequenza Interna	Frequenza Globale	p.value	v.test
Телеканал	4.7356828	0.10693729	43	149	0.000	15.968157
невідчужваність	4.6255507	0.10621959	42	148	0.000	15.726971
відкликання	4.5154185	0.16507098	41	230	0.000	14.122914
втеча	3.7444934	0.9904259	34	138	0.000	13.736342
тиша	3.7444934	0.17440108	34	243	0.000	12.153262
ЗМІ	4.4052863	0.47727044	40	665	0.000	10.418354
Безчестя	2.2026432	0.31076406	20	433	0.000	6.621652
обман	2.8634361	0.62081043	26	865	0.000	6.227377
окупація	2.5330396	1.10310477	23	1537	0.000	3.468841
влада	1.3215859	0.52105014	12	726	0.000	2.710736
конфлікт	0.5506608	0.12344439	5	172	0.000	2.534295
пастка	0.5506608	0.12487979	5	174	0.000	2.517448
російська	1.3215859	0.59569093	12	830	0.000	2.353510
вийди	1.2114537	0.55119353	11	768	0.000	2.219077
ОБСЕ	1.2114537	0.58636083	11	817	0.000	2.560228
...
труп	1.25478523	0.41254785	10	478	0.000	2.365896

Tabella 10. Caratterizzazione dei cluster dei concetti negativi ucraini