



Università
Ca' Foscari
Venezia

Corso di Dottorato di ricerca
in Lingue, culture e società moderne e Scienze del linguaggio
Ciclo 34°

Tesi di ricerca

Rappresentazioni di genere dentro e attraverso il linguaggio dell'informazione italiana

SSD: L-LIN 0/01 (GLOTTOLOGIA E LINGUISTICA)

SSD: SPS/08 (SOCIOLOGIA DEI PROCESSI CULTURALI E COMUNICATIVI)

Coordinatore del Dottorato

ch.mo prof. Enric Bou Maqueda

Supervisora

ch.ma prof.ssa Giuliana Giusti

Co-supervisore

ch.mo dott. Gianluca Lebani

Dottoranda

Monia Azzalini

Matricola 956448

Sommario

LISTA DELLE ABBREVIAZIONI	4
INDICE DEGLI ELENCHI	5
INDICE DEI GRAFICI	6
INDICE DELLE FIGURE	7
INDICE DELLE TABELLE	8
INDICE DELLE TABELLE IN APPENDICE	10
RINGRAZIAMENTI.....	11
CAPITOLO 1 INTRODUZIONE.....	13
1.1 LINGUA, GENERE E MEDIA	13
1.2 OBIETTIVO DELLA TESI	16
1.3 DATI E METODOLOGIE DI RICERCA	19
1.4 STRUTTURA DELLA TESI	22
CAPITOLO 2 LINGUA E GENERE	25
2.1 MODELLI TEORICI ED EVIDENZE EMPIRICHE IN AMBITO INTERNAZIONALE	25
2.2 ORIGINE E SVILUPPO DEGLI STUDI IN AMBITO NAZIONALE	35
2.3 IL GENERE FRA GRAMMATICA E CULTURA	53
CAPITOLO 3 MEDIA E GENERE	66
3.1 MODELLI TEORICI ED EVIDENZE EMPIRICHE IN AMBITO INTERNAZIONALE	66
3.2 ORIGINE E SVILUPPO DEGLI STUDI IN AMBITO NAZIONALE	75
3.3 IL GENERE FRA REALTÀ E INFORMAZIONE	97
CAPITOLO 4 COSTRUZIONE E DESCRIZIONE DEI <i>CORPORA</i> ANALIZZATI	118
4.1 DEFINIZIONE DI <i>CORPUS</i>	118
4.2 IL <i>CORPUS</i> TG	120
4.3 IL <i>CORPUS</i> CRS	125
4.4 IL <i>CORPUS</i> ACR	127
CAPITOLO 5 ANALISI DELLE RAPPRESENTAZIONI DI GENERE IN UN <i>CORPUS</i> DI TG ..	130
5.1 DOMANDE DI RICERCA	131
5.2 METODOLOGIA	132
5.3 RISULTATI	146
5.4 DISCUSSIONE DEI RISULTATI	233
5.5 CONCLUSIONI	254
CAPITOLO 6 ANALISI DELL'ACCORDO DI GENERE NEI NOMI DI RUOLO POLITICO IN UN <i>CORPUS</i> DI QUOTIDIANI: UNO STUDIO DEL LEMMA <i>MINISTRA/O</i>	260
6.1 INTRODUZIONE	261
6.2 STATO DELL'ARTE	262
6.3 IPOTESI E DOMANDE DI RICERCA	270
6.4 METODOLOGIA	270
6.5 RISULTATI	274
6.6 DISCUSSIONE DEI RISULTATI	292

6.7	CONCLUSIONI	305
CAPITOLO 7 ANALISI DELLE ASSOCIAZIONI DI GENERE LATENTI NEI NOMI DI PROFESSIONE E CARICA POLITICA IN UN <i>CORPUS</i> DI NOTIZIE.....		
7.1	INTRODUZIONE.....	311
7.2	STATO DELL'ARTE.....	314
7.3	IPOTESI E DOMANDE DI RICERCA.....	329
7.4	METODOLOGIA	330
7.5	RISULTATI	332
7.6	DISCUSSIONE DEI RISULTATI	337
7.7	CONCLUSIONI	339
CAPITOLO 8 CONCLUSIONI		
8.1	RISULTATI PRINCIPALI DELLA TESI.....	343
8.2	LIMITI DELLA TESI E SVILUPPI POSSIBILI	350
BIBLIOGRAFIA		353
APPENDICE: TABELLE INTEGRATIVE (CAPITOLO 5).....		406

Lista delle abbreviazioni

- ACL: Analisi delle Corrispondenze Lessicali
- ADT: Analisi dei Dati Testuali
- CGD: Classificazione Gerarchica Discendente
- DSMs: Distributional Semantic Models (modelli di semantica distribuzionale)
- LSA: Latent Semantic Analysis (analisi semantica latente)
- ST: Segmento di testo (sinonimo di UCE)
- STT: Segmento di testo tipico
- TIC: Tecnologie dell'informazione e della comunicazione
- UCE: Unità di Contesto Elementare (sinonimo di ST)
- UCI: Unità di Contesto Iniziale
- VSMs: Vector Space Models
- WE(s): Word Embedding(s)

Indice degli elenchi

Elenco 1. Selezione di segmenti di testo tipici della classe 1 del <i>sub-corpus</i> P_Fonte	157
Elenco 2. Selezione di segmenti di testo tipici della classe 2 del <i>sub-corpus</i> P_Fonte	160
Elenco 3. Selezione di segmenti di testo tipici della classe 3 del <i>sub-corpus</i> P_Fonte	162
Elenco 4. Selezione di segmenti di testo tipici della classe 1 del <i>sub-corpus</i> PF_Donna	169
Elenco 5. Selezione di segmenti di testo tipici della classe 2 del <i>sub-corpus</i> PF_Donna	173
Elenco 6. Selezione di segmenti di testo tipici della classe 1 del <i>sub-corpus</i> PF_Uomo.....	178
Elenco 7. Selezione di segmenti di testo tipici della classe 2 del <i>sub-corpus</i> PF_Uomo.....	181
Elenco 8. Selezione di segmenti di testo tipici della classe 3 del <i>sub-corpus</i> PF_Uomo.....	184
Elenco 9. Selezione di segmenti di testo tipici della classe 1 del <i>sub-corpus</i> donna	200
Elenco 10. Selezione di segmenti di testo tipici della classe 2 del <i>sub-corpus</i> donna	203
Elenco 11. Selezione di segmenti di testo tipici della classe 3 del <i>sub-corpus</i> donna	207
Elenco 12. Selezione di segmenti di testo tipici della classe 4 del <i>sub-corpus</i> donna	210
Elenco 13. Selezione di segmenti di testo tipici della classe 1 del <i>sub-corpus</i> uomo	216
Elenco 14. Selezione di segmenti di testo tipici della classe 2 del <i>sub-corpus</i> uomo	219
Elenco 15. Selezione di segmenti di testo tipici della classe 3 del <i>sub-corpus</i> uomo	222
Elenco 16. Selezione di segmenti di testo tipici della classe 4 del <i>sub-corpus</i> uomo	226

Indice dei grafici

Grafico 1. <i>Corpus</i> TG: profilo rango/ frequenza	124
Grafico 2. <i>Corpus</i> CRS: profilo rango/frequenza	127
Grafico 3. <i>Sub-corpus</i> P_Fonte: analisi delle somiglianze (cut-off N>100)	147
Grafico 4. <i>Sub-corpus</i> PF_Donna: analisi delle somiglianze (cut-off al rango 30)	152
Grafico 5. <i>Sub-corpus</i> P_Fonte PF_Uomo: analisi delle somiglianze (cut-off al rango 30)	153
Grafico 6. Dendrogramma delle classi lessicali del <i>sub-corpus</i> P_Fonte	155
Grafico 7. Rappresentazione fattoriale delle 3 classi lessicali del <i>sub-corpus</i> P_Fonte	164
Grafico 8. Rappresentazione fattoriale dalle variabili associate alle 3 classi lessicali del <i>sub-corpus</i> P_Fonte	165
Grafico 9. Dendrogramma delle classi lessicali del <i>sub-corpus</i> PF_Donna	167
Grafico 10. Dendrogramma delle classi lessicali del <i>sub-corpus</i> PF_Uomo	176
Grafico 11. Chi-quadrato di associazione delle 3 classi lessicali del <i>sub-corpus</i> P_Fonte con le modalità della variabile *Anno	186
Grafico 12. Chi-quadrato di associazione delle 2 classi lessicali del <i>sub-corpus</i> PF_Donna con le modalità della variabile *Anno	187
Grafico 13. Chi-quadrato di associazione delle 2 classi lessicali del <i>sub-corpus</i> PF_Uomo con le modalità della variabile *Anno	188
Grafico 14. Chi-quadrato di associazione delle 3 classi lessicali del <i>sub-corpus</i> P_Fonte con le modalità della variabile *TG	190
Grafico 15. Chi-quadrato di associazione delle 2 classi lessicali del <i>sub-corpus</i> PF_Donna con le modalità della variabile *TG	191
Grafico 16. Chi-quadrato di associazione delle 2 classi lessicali del <i>sub-corpus</i> PF_Uomo con le modalità della variabile *TG	192
Grafico 17. Dendrogramma delle classi lessicali del <i>sub-corpus</i> donna	196
Grafico 18. Rappresentazione fattoriale delle 4 classi lessicali del <i>sub-corpus</i> donna	212
Grafico 19. Dendrogramma delle classi lessicali del <i>sub-corpus</i> uomo	213
Grafico 20. Rappresentazione fattoriale delle 4 classi lessicali del <i>sub-corpus</i> uomo	227
Grafico 21. Chi-quadrato di associazione delle 4 classi lessicali del <i>sub-corpus</i> donna con le modalità della variabile *Anno	229
Grafico 22. Chi-quadrato di associazione delle 4 classi lessicali del <i>sub-corpus</i> uomo con le modalità della variabile *Anno	231
Grafico 23. Chi-quadrato di associazione delle 4 classi lessicali del <i>sub-corpus</i> donna con le modalità della variabile *TG	231
Grafico 24. Chi-quadrato di associazione delle 4 classi lessicali del <i>sub-corpus</i> uomo per TG	232
Grafico 25. Correlazione fra <i>gender-score</i> dei nomi di professione/carica politica e tasso di occupazione femminile: $r=0,615$, $p<0,05$	334
Grafico 26. Correlazione fra <i>gender-score</i> dei nomi di professione/carica politica distinti fra marcati ($r=-0,339$, $p=0,26$) e non marcati ($r=-0,339$, $p=0,26$) e tasso di occupazione femminile	336

Indice delle figure

Figura 1. Esempio di marcatura del <i>corpus</i>	145
Figura 2. Esempio di una matrice di coricorrenze.....	312
Figura 3. Esempio di rappresentazione vettoriale della parole ricavata dalla matrice in Figura 2 .	312
Figura 4. <i>Bias</i> di genere in nomi di professione in due diversi spazi vettoriali generati da due software diversi (GloVe e word2vec) (Bolukbasi et al. 2016, Fig.1).....	316
Figura 5. Indice di correlazione fra associazione con il genere femminile di nomi di professione in uno spazio vettoriale e percentuale di occupazione delle donne nella relativa professione (Caliskan et al. 2017, 2)	320
Figura 6. Correlazione fra associazione con il genere femminile di nomi di professione in uno spazio vettoriale e percentuale di occupazione delle donne nella relativa professione (Garg et al. 2018, E3636)	324
Figura 7. Correlazione fra associazioni di genere in embeddings di nomi i professione e tasso di occupazione femminile, con un'unica forma nominale, a sinistra, con le due forme femminili e maschili, per i nomi dal genere marcato, a destra (Biasion et al. 2020, 5).....	328

Indice delle tabelle

Tabella 1. Dimensioni del <i>corpus</i> TG.....	124
Tabella 2. Numero di articoli selezionati per la costruzione del <i>corpus</i> di analisi, per quotidiano e per anno.....	125
Tabella 3. Elenco delle variabili con cui è stato marcato il <i>corpus</i> TG.....	145
Tabella 4. Statistiche: dimensioni e caratteristiche del <i>sub-corpus</i> P_Fonte.....	146
Tabella 5. Statistiche: dimensioni dei <i>sub-corpora</i> PF_Donna e PF_Uomo.....	149
Tabella 6. Forme lessicali più frequenti nei <i>sub-corpora</i> PF_Donna e PF_Uomo (occ.).....	149
Tabella 7. Statistiche: risultati della CGD sul <i>sub-corpus</i> P_Fonte.....	154
Tabella 8. Prime 30 forme lessicali significative delle 3 classi del <i>sub-corpus</i> P_Fonte.....	156
Tabella 9. Statistiche: risultati della CGD del <i>sub-corpus</i> PF_Donna.....	166
Tabella 10. Forme lessicali significative della classe 1 del <i>sub-corpus</i> PF_Donna.....	167
Tabella 11. Prime 40 forme lessicali significative della classe 2 del <i>sub-corpus</i> PF_Donna.....	172
Tabella 12. Statistiche: risultati della CGD del <i>sub-corpus</i> PF_Uomo.....	175
Tabella 13. Prime 40 forme lessicali significative della classe 1 del <i>sub-corpus</i> PF_Uomo.....	176
Tabella 14. Prime 40 forme lessicali significative della classe 2 del <i>sub-corpus</i> PF_Uomo.....	179
Tabella 15. Prime 40 forme lessicali significative della classe 3 del <i>sub-corpus</i> PF_Uomo.....	182
Tabella 16. Statistiche: dimensioni e caratteristiche dei <i>sub-corpora</i> donna e uomo.....	193
Tabella 17. Forme lessicali più frequenti nei <i>sub-corpora</i> donna e uomo (occorrenze).....	193
Tabella 18. Statistiche: risultati della CGD del <i>sub-corpus</i> donna.....	195
Tabella 19. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 1 del <i>sub-corpus</i> donna.....	198
Tabella 20. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 2 del <i>sub-corpus</i> donna.....	202
Tabella 21. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 3 del <i>sub-corpus</i> donna.....	206
Tabella 22. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 4 del <i>sub-corpus</i> donna.....	209
Tabella 23. Statistiche: risultati della CGD del <i>sub-corpus</i> uomo.....	212
Tabella 24. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 1 del <i>sub-corpus</i> uomo.....	214
Tabella 25. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 2 del <i>sub-corpus</i> uomo.....	218
Tabella 26. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 3 del <i>sub-corpus</i> uomo.....	221
Tabella 27. Forme lessicali con freq. $N>5$ e $\chi^2>3,84$ ($p<0,05$) della classe 4 del <i>sub-corpus</i> uomo.....	225
Tabella 28. Frequenza dei nomi di cariche politiche nel <i>corpus</i> TG.....	269
Tabella 29. Sistema di classificazione delle espressioni linguistiche per il riferimento a ministre e ministri.....	272
Tabella 30. <i>Corpus</i> TG: frequenza delle quattro forme del lemma <i>ministra/o</i>	274
Tabella 31. <i>Corpus</i> TG: frequenza delle quattro forme di <i>ministra/o</i> per anno.....	275
Tabella 32. <i>Corpus</i> TG: frequenze osservate e attese per le forme femminili e maschili del lemma <i>ministra/o</i> per anno.....	275
Tabella 33. <i>Corpus</i> TG: frequenze osservate e attese per le forme femminili e maschili del lemma <i>ministra/o</i> per anno, solo 2018 e 2019.....	276
Tabella 34. <i>Corpus</i> TG: frequenza delle quattro forme di <i>ministra/o</i> per TG.....	276
Tabella 35. <i>Corpus</i> TG: frequenza delle quattro forme di <i>ministra/o</i> per genere semantico.....	278
Tabella 36. <i>Corpus</i> TG: distribuzione delle frequenze del lemma <i>ministra/o</i> per genere semantico e per anno.....	278
Tabella 37. <i>Corpus</i> TG: distribuzione delle frequenze del lemma <i>ministra/o</i> per genere semantico e per TG.....	278
Tabella 38. <i>Corpus</i> TG: frequenza delle forme femminili e maschili per genere semantico femminile e anno.....	280

Tabella 39. <i>Corpus</i> TG: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per anno	280
Tabella 40. <i>Corpus</i> TG: frequenza delle forme femminili e maschili con genere semantico femminile per TG	280
Tabella 41. <i>Corpus</i> TG: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per TG	281
Tabella 42. <i>Sub-corpus</i> TG 2: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per anno	281
Tabella 43. Tabella di corrispondenza fra il sistema di classificazione delle espressioni linguistiche per il riferimento a ministre e ministri originariamente elaborato e il sistema utilizzato per la presentazione dei risultati.....	282
Tabella 44. <i>Corpus</i> TG: distribuzione delle frequenze del tipo di espressione linguistica per genere semantico e genere grammaticale (solo forme singolari).....	283
Tabella 45. <i>Corpus</i> CRS: frequenza delle quattro forme del lemma <i>ministra/o</i>	284
Tabella 46. <i>Corpus</i> CRS: frequenza delle quattro forme del lemma <i>ministra/o</i> per anno.....	285
Tabella 47. <i>Corpus</i> CRS: frequenze attese e frequenze osservate per anno.....	285
Tabella 48. <i>Corpus</i> CRS: frequenza delle quattro forme di <i>ministra/o</i> per quotidiano.....	285
Tabella 49. <i>Corpus</i> CRS: frequenze attese e frequenze osservate per quotidiano	286
Tabella 50. <i>Corpus</i> CRS: frequenze attese e frequenze osservate per quotidiano	286
Tabella 51. <i>Corpus</i> CRS: frequenza delle quattro forme di <i>ministra/o</i> per genere semantico.....	287
Tabella 52. <i>Corpus</i> CRS: distribuzione delle frequenze del lemma <i>ministra/o</i> per genere semantico e per anno	288
Tabella 53. <i>Corpus</i> CRS: distribuzione delle frequenze del lemma <i>ministra/o</i> per genere semantico e per quotidiano	288
Tabella 54. <i>Corpus</i> CRS: frequenze delle forme femminili e maschili con genere semantico femminile e anno.....	289
Tabella 55. <i>Corpus</i> CRS: frequenza delle forme femminili e maschili con genere semantico femminile per quotidiano.....	289
Tabella 56. <i>Corpus</i> RCS: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per quotidiano	289
Tabella 57. <i>Sub-corpus Il Resto del Carlino</i> : frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per anno.....	290
Tabella 58. <i>Corpus</i> CRS: distribuzione delle frequenze del tipo di espressione linguistica per genere semantico e genere grammaticale	291
Tabella 59. <i>Corpus</i> TG 2018-19: frequenza delle quattro forme del lemma <i>ministra/o</i>	301
Tabella 60. <i>Corpus</i> TG 2018-19: frequenza delle quattro forme di <i>ministra/o</i> per genere semantico	302
Tabella 61. <i>Corpus</i> TG 2018-19: distribuzione delle frequenze del tipo di espressione linguistica per genere semantico e genere grammaticale.....	303
Tabella 62. Elenco dei nomi di professione e carica politica.....	331
Tabella 63. <i>Gender-score</i> dei nomi di professione e carica politica e tasso di occupazione femminile nella professione o carica relativa (in valore decimale)	333
Tabella 64. <i>Gender-score</i> delle forme femminili e maschili dei nomi di professione e carica politica	335

Indice delle tabelle in Appendice

Tab. 1	Forme lessicali significative per la classe 1 (CGD) del <i>sub-corpus</i> P_Fonte.....	407
Tab. 2	Forme lessicali significative per la classe 2 (CGD) del <i>sub-corpus</i> P_Fonte.....	415
Tab. 3	Forme lessicali significative per la classe 3 (CGD) del <i>sub-corpus</i> P_Fonte.....	424
Tab. 4	Forme lessicali significative per la classe 2 (CGD) del <i>sub-corpus</i> PF_Donna	440
Tab. 5	Forme lessicali significative per la classe 1 (CGD) del <i>sub-corpus</i> PF_Uomo	446
Tab. 6	Forme lessicali significative per la classe 2 (CGD) del <i>sub-corpus</i> PF_Uomo	454
Tab. 7	Forme lessicali significative per la classe 3 (CGD) del <i>sub-corpus</i> PF_Uomo	462
Tab. 8	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 1 (CGD) del <i>sub-corpus</i> donna	476
Tab. 9	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 2 (CGD) del <i>sub-corpus</i> donna	478
Tab. 10	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 3 (CGD) del <i>sub-corpus</i> donna.....	480
Tab. 11	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 4 (CGD) del <i>sub-corpus</i> donna.....	482
Tab. 12	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 1 (CGD) del <i>sub-corpus</i> uomo	484
Tab. 13	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 2 (CGD) del <i>sub-corpus</i> uomo	486
Tab. 14	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 3 (CGD) del <i>sub-corpus</i> uomo	489
Tab. 15	Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 4 (CGD) del <i>sub-corpus</i> uomo	491

Ringraziamenti

Molte sono le persone che devo ringraziare per essere riuscita a portare a compimento il mio corso di dottorato e a scrivere questa tesi. Ringrazio anzitutto la mia supervisora, prof.ssa Giuliana Giusti, per il suo rigore professionale e scientifico, nonché per la sua generosità intellettuale e umana, costanti in tutti i tre anni di dottorato. La sua guida è stata fondamentale per introdurmi alla linguistica, in particolare in una prospettiva di genere. Ringrazio poi il mio co-supervisore, dott. Gianluca Lebani, perché senza le sue lezioni di linguistica computazionale, la sua disponibilità e il suo aiuto concreto il risultato di questa tesi non sarebbe stato lo stesso, in particolare, non avrei mai scritto il Capitolo 7. Per la stesura di questo capitolo devo ringraziare anche il dott. Davide Biasion e il dott. Alessandro Fabris, per aver condiviso e discusso con me la versione pre-print dell'articolo scritto con i colleghi dell'Università di Padova (Biasion et al. 2020), che ha ispirato la mia analisi sui *word embeddings* in un *corpus* di quotidiani. Per la stesura del Capitolo 6 ringrazio la prof.ssa Virginia Valian, dell'Hunter College di New York, che nel suo semestre come *visiting scholar* a Ca' Foscari si è confrontata con me, fornendomi un utile sguardo da oltreoceano sulla questione della femminilizzazione dei nomi di professione. Per la stesura del Capitolo 5 ringrazio Milly Buonanno, professoressa emerita presso il Dipartimento di Comunicazione e Ricerca dell'Università La Sapienza di Roma. L'amicizia che ci lega da lungo tempo mi ha permesso di chiederle la lettura di una prima versione della tesi. La ringrazio per averla letta e avermi confortata circa il contributo rilevante del mio lavoro (anche) per la sociologia dei processi culturali e comunicativi, da sempre il settore più produttivo, in Italia, di *gender and media studies*.

Ci sono poi altre amiche del mondo accademico a cui vanno i miei ringraziamenti per avere in vario modo supportato il mio percorso di studi dottorali: prof.ssa Stefania Cavagnoli (Università Tor Vergata di Roma), prof.ssa Karen Ross (Newcastle University, UK), prof.ssa Francesca Santulli (Università Ca' Foscari di Venezia), prof.ssa Arjuna Tuzzi (Università di Padova). Un grazie particolare lo devo

alla prof.ssa Claudia Padovani (Università di Padova), per esserci sempre stata, con preziosi consigli professionali e un indispensabile supporto umano.

Ringrazio poi il Dipartimento di Studi Linguistici e Comparati di Ca' Foscari, per tutte le occasioni di studio offerte, nell'ambito del corso di dottorato in maniera costante, anche durante l'emergenza Covid-19, che è stata un po' per tutti e tutte un momento difficile. In particolare voglio ringraziare per i seminari dedicati alla tematica "lingua e genere", fra cui le stimolanti e vivaci lezioni della prof.ssa Heather Burnett, direttrice di ricerca al CNRS, presso l'Université Diderot di Parigi e ospite di Ca' Foscari come *visiting scholar*.

Ringrazio l'Osservatorio di Pavia, dove mi sono formata come ricercatrice e di cui sono socia da molti anni, per aver acconsentito a rendere *open source* le trascrizioni dei testi di un campione di telegiornali, che ho utilizzato per costruire il *corpus* analizzato nei capitoli 5 e 6. E ringrazio i membri del CDA che hanno appoggiato la mia scelta quando ho deciso di frequentare un corso di dottorato. Fra i colleghi dell'Osservatorio, un grazie va anche a Andrea Caretta e Vittorio Cobianchi, per avermi aiutata a risolvere diversi problemi di ottimizzazione del software IRaMuTeQ, che ho usato in una delle mie analisi.

Infine, ringrazio gli amici e le amiche che ci sono sempre state e quelle che ho incontrato durante il corso di dottorato. Tutte e tutti hanno contribuito ad alleggerire un percorso talora difficoltoso, e mi hanno aiutata a ritrovare fiducia in me stessa, nei momenti di difficoltà.

Un grazie di cuore va alla mia famiglia: ai miei genitori per la cura dedicata ai miei figli, durante le mie assenze, non solo fisiche ma anche mentali, a mia sorella per aver sempre creduto in me, e avermelo sempre ricordato. Infine un grazie speciale va ai mie figli, a cui dedico con tutto il mio affetto questa tesi.

Capitolo 1 Introduzione

Questa tesi analizza le rappresentazioni di genere dentro e attraverso il linguaggio dell'informazione italiana, con un approccio che si propone di integrare l'analisi linguistica con l'analisi mediale.

1.1 Lingua, genere e media

Sia la linguistica sia i *media studies* hanno sviluppato nei rispettivi ambiti disciplinari una riflessione sulle rappresentazioni di genere, condividendo alcuni principi teorici, e talvolta anche pratiche di ricerca, seguendo tuttavia percorsi generalmente indipendenti piuttosto che interdisciplinari.

L'interesse della linguistica per le questioni di genere nasce negli anni Settanta, in ambito anglosassone, su impulso delle teorie femministe, con lo studio pionieristico della linguista statunitense Robin Lakoff, *Language and woman's place* (1973), che pone le basi per un'ampia produzione scientifica che oggi riguarda molte lingue del mondo. In particolare, dal 2001, Marlis Hellinger e Hadumod Bußmann (2001-2003), e poi Marlis Hellinger e Heiko Motschenbacher (2004), pubblicano la serie *Gender across languages. The linguistic representation of women and men*, che arriva a comprendere l'analisi di quarantadue lingue. Lo studio di Lakoff (1973) analizza per la prima volta la lingua (inglese) in una prospettiva di genere, dimostrando che le caratteristiche della lingua parlata dalle donne non sono la manifestazione di differenze biologiche, come ipotizzavano alcuni studi di inizio secolo (Jespersen 1922; Malinowski 1929; Sapir 1929; Haas 1944; Flannery 1946; Lévi-Strauss 1955), bensì il riflesso delle relazioni gerarchiche fra donne e uomini, tipiche di un sistema di potere androcentrico (non solo americano, ma universale), che confina le donne ai margini della società e si manifesta anche nella lingua che parla delle donne.

Le prime riflessioni in materia di media e genere precedono di qualche anno lo studio su lingua e genere di Robin Lakoff (1973), con la pubblicazione, sempre

negli Stati Uniti, di testi come *The feminine mystique* di Betty Friedan (1963) e *Women's estate* di Juliet Mitchell (1966), nei quali vengono denunciate le rappresentazioni stereotipate delle donne, confinate nella sfera privata e commercialmente sfruttate come corpo e oggetto del desiderio sessuale, soprattutto nella pubblicità (Capecchi 2006). Entrambi questi volumi non sviluppano in modo articolato un'analisi della rappresentazione mediale delle donne, per la quale bisogna attendere la pubblicazione di *Heart and home. Images of women in the mass media* a cura di Gaye Tuchman, Arlene Kaplan e James Benét (1978), generalmente considerato il lavoro pionieristico in materia di media e genere. I diversi contributi a questo volume avanzano e dimostrano l'ipotesi che i media operano un simbolico annullamento delle donne, sotto-rappresentandole o rappresentandole in ruoli marginali, poiché non riflettono oggettivamente il mondo reale, ma le idee e i valori dominanti che sono quelli maschili.

In Italia, l'eco delle riflessioni maturate oltreoceano su lingua e genere e media e genere arriva con un ritardo di circa dieci anni, rispettivamente a metà degli anni Ottanta e Settanta del Novecento. Nell'ambito della linguistica, vengono generalmente considerati seminali i lavori di Alma Sabatini (1986, 1987) sul sessismo nella lingua italiana, preceduti di qualche anno da alcuni studi sul parlato delle donne, che hanno conosciuto una minor fortuna e un minor seguito in termini di linea di ricerca, rispetto alla riflessione focalizzata sulla lingua che parla delle donne, che è stata inizialmente sostenuta dalle istituzioni politiche. La prima ricerca sulla lingua italiana analizzata in una prospettiva di genere viene infatti realizzata per volontà del Governo Craxi (1983-1987) che affida ad Alma Sabatini il compito di svolgere una ricerca sul linguaggio della stampa, e, sulla base dei risultati ottenuti, di formulare indicazioni per un uso non sessista dell'italiano. L'obiettivo dichiarato dalla Presidenza del Consiglio Craxi era quello di liberare la società

dai residui pregiudizi nei confronti delle donne [...] non sempre riconoscibili, perché sono spesso nascosti e camuffati sotto forme di apparente valore oggettivo, e sono trasmessi, perpetuati e avvalorati

attraverso la lingua, in modo spesso subdolo e ripetitivo (Sabatini 1987, 23).

Nel 1986 vengono così pubblicate le *Raccomandazioni per un uso non sessista della lingua italiana: per la scuola e per l'editoria scolastica* (Sabatini 1986), e, l'anno successivo, il volume *Il sessismo nella lingua italiana* (Sabatini 1987), dove le *Raccomandazioni* sono anticipate dai risultati della ricerca sulla stampa e dai contributi di Elena Marinucci, Francesco Sabatini e Sergio Lepri. Elena Marinucci, presidente della *Commissione nazionale per la realizzazione della parità tra uomo e donna*, introduce il volume esplicitando l'intenzione istituzionale di offrire «stimoli alla riflessione, con suggerimenti in dimensione aperta e problematica, a chi fa uso della lingua e, usandola, esercita un'azione politica» (Marinucci 1987, 11). Il linguista Francesco Sabatini (1987) concentra il suo contributo sulla relazione fra lingua, pensiero e realtà extralinguistica, spiegando che la lingua è un veicolo storico di idee e valori propri di una società, per questo cambia nel tempo riflettendo i mutamenti sociali, ma i cambiamenti linguistici sono sempre lenti e seguono, piuttosto che precedere, le trasformazioni sociali. Infine, Sergio Lepri, giornalista e direttore dell'ANSA, traccia i limiti dell'azione politica di giornalisti e giornaliste, auspicata da Marinucci, ricordando che sul linguaggio giornalistico «per l'organizzazione del lavoro e gli stessi meccanismi mentali degli operatori, pesano tanto gli elementi imitativi, e anche subliminali, degli ambienti di cui partecipa» (Lepri 1987, 21-22), cioè le classi dirigenti del paese, per cui non basta un impegno consapevole degli organi di stampa, ma occorre anche un impegno degli organi politici e di governo.

Le *Raccomandazioni* del 1986 si rivolgono al mondo dell'istruzione, individuando nel periodo della formazione scolastica una fase cruciale per l'apprendimento linguistico. Il volume del 1987 si concentra invece sul linguaggio dell'informazione, non solo perché propone una ricerca empirica sulla stampa, quotidiana e periodica, ma anche perché attraverso Sergio Lepri chiama in causa direttamente il ruolo dei mezzi d'informazione, e in modo non esente da problematicità. Il direttore dell'ANSA pone infatti l'attenzione su due fattori socio-

culturali rilevanti nella resistenza delle pratiche giornalistiche a un uso inclusivo e paritario dell'italiano: la vicinanza del linguaggio giornalistico alla lingua ufficiale del paese, piuttosto che alla lingua della gente, e una certa subordinazione di giornaliste e giornalisti al sistema del potere. Si tratta di una riflessione importante, nella misura in cui sollecita un dialogo fra la linguistica e i *media studies*, che tuttavia, ad oggi, non è ancora stato intrapreso.

1.2 Obiettivo della tesi

L'obiettivo generale di questa tesi è quello di colmare un vuoto di ricerca costituito dalla mancanza di un'indagine sulle rappresentazioni di genere che integri l'analisi di aspetti lessicali e testuali della lingua con un'analisi delle caratteristiche interne (generi, formati) e/o esterne (cultura, sistema, politiche, organizzazione) dei media, in una prospettiva interdisciplinare, linguistica e mediale, e di genere.

Per rispondere a questo obiettivo generale ho costruito e esplorato tre *corpora* linguistici, optando per una collezione di contenuti dell'informazione quotidiana, televisiva e a stampa. Una scelta guidata da diverse ragioni. In primo luogo, perché le ricerche in materia di media e genere dimostrano che il genere come variabile socio-culturale (*gender*) è sensibile sia al tipo di media, sia al genere mediale (*genre*) (Capecchi 2006, 2018). La rappresentazione delle donne e degli uomini nella pubblicità, per esempio, è diversa, e chiama in causa fattori diversi, dalla rappresentazione delle donne e degli uomini nei telegiornali. Circoscrivere la ricerca a *corpora* di testi informativi prodotti da media tradizionali, come la stampa e la televisione, consente così di delineare un contesto mediale di riferimento ben preciso, con caratteristiche proprie sia in termini di contenuti, sia in termini di organizzazione e processi produttivi, utili a individuare i fattori medialità che si intrecciano con i fattori linguistici nelle rappresentazioni di genere.

In secondo luogo, i mezzi di informazione sono stati individuati, sin dalle prime riflessioni sul sessismo nella lingua italiana, come testimoni dei processi di cambiamento, o di conservazione, nell'uso dell'italiano, e anche come *stakeholder*

nella promozione della parità linguistica e nel contrasto agli stereotipi di genere che attraverso la lingua si trasmettono e si riproducono (Sabatini 1987; Cardinaletti e Giusti 1991; Cavagnoli 2013; Robustelli 2014).

La terza ragione per cui ho deciso di concentrare la mia ricerca su *corpora* di testi informativi è collegata al ruolo dei mezzi di informazione, in particolare di giornaliste e giornalisti, che di recente hanno assunto maggior consapevolezza sul loro potere, e sulla loro responsabilità, nel favorire un linguaggio *gender-fair*, come testimonia la pubblicazione di due linee guida promossa da rappresentanti del mondo dell'informazione: *Donne, grammatica e media* (Robustelli 2014), stampato dall'associazione di giornaliste GiULIA (Giornaliste Unite Libere e Autonome), nata nel 2011 con l'obiettivo, fra gli altri, di «modificare lo squilibrio informativo sulle donne anche utilizzando un linguaggio privo di stereotipi e declinato al femminile»¹, e *Tutt'altro genere d'informazione. Manuale per una corretta rappresentazione delle donne nell'informazione*, con la terza parte focalizzata sull'uso della lingua italiana (Cavagnoli 2015), a cura del Gruppo di lavoro Pari opportunità del Consiglio nazionale dell'Ordine dei giornalisti (CNOG 2015).

Circoscrivere la ricerca all'informazione può agevolare, inoltre, la collocazione dei risultati entro uno spazio di "potere" ben preciso, che è quello del sistema dell'informazione. L'informazione, come il linguaggio, infatti, non è uno strumento di comunicazione neutro, in parte riflette la realtà, e dunque anche le disparità reali fra donne e uomini, ma in parte riflette una cultura giornalistica attiva nella costruzione della realtà sociale, che è sempre un processo interattivo, situato e negoziato con il pubblico (Losito 1998; Grossi e Ruspini 2007), e in quanto tale può contribuire a rinforzare o, viceversa, a sfidare gli stereotipi di genere (Gallagher 2006). Se è vero infatti che possiamo pensare l'informazione come una "finestra sul mondo", vero è anche che dobbiamo essere consapevoli che quello che vediamo dipende dalle caratteristiche della finestra e dalla posizione di chi guarda.

¹ Fonte: <https://giulia.globalist.it/chi-siamo>; ultimo accesso 19 novembre 2021.

The views through a window depends upon whether the window is large or small, has many panes or few, whether the glass is opaque or clear, whether the window faces a street or a backyard. The unfolding scene also depends up one where one stands, far o near, craning one’s neck to the side, or gazing straight ahead, eyes parallel to the wall in which the window is encased (Tuchman 1978, 1).

La scelta di analizzare il linguaggio dell’informazione ha orientato la mia ricerca verso una prospettiva di genere prevalentemente focalizzata sulle rappresentazioni femminili. Le donne sono certamente soltanto una delle due (principali) categorie del genere, ma si tratta di una scelta giustificata – per riprendere le parole della sociologa Milly Buonanno (2015a, 434) che cita Linda Steiner (2012, 210) – dal fatto che «in journalism and mass media, women have been regarded as the intruder, the exception, the problem. So, for better or worse, gender research usually is about women». Una dichiarazione, quest’ultima, che può riferirsi anche alla lingua e alla linguistica: la lingua delle donne è stata originariamente considerata come una deviazione dalla norma (Jespersen 1922) o come manifestazione della subordinazione delle donne al sistema di potere maschile, e la lingua che parla delle donne come espressione di forme di rapporti di potere gerarchici fra donne e uomini (Lakoff 173, Spender 1980, 1985), per cui anche gli studi linguistici sulla relazione fra lingua e genere si sono a lungo concentrati, e ancora oggi si concentrano prevalentemente, sulle donne, come vedremo nel Capitolo 2. In particolare, in Italia, la ricerca linguistica si focalizza su asimmetrie di genere che discriminano le donne, rendendole marginali o invisibili, e su stereotipi che le sminuiscono connotando l’italiano come una lingua sessista, nel senso che spiegherò più avanti (§ 2.2.2), e androcentrica, nel senso brevemente spiegato di seguito.

L’androcentrismo è un fenomeno individuato già all’inizio del Novecento dalla sociologa Charlotte Perkins Gilman (1911), che definisce androcentrica una società caratterizzata da una cultura in cui gli uomini gestiscono tutte le attività e le donne sono posizionate ai margini, e, più tardi da Simone de Beauvoir (1949),

che ne *Le deuxième sexe* argomenta a lungo la sua nota tesi per cui il mondo e la sua rappresentazione sono una costruzione maschile. Solo più tardi ne è stata data una definizione articolata che mi pare ancora valida:

androcentrism is the privileging of male experience and the “otherizing” of female experience, that is, males and male experience are treated as a neutral standard or norm for the culture or the specie as a whole, and females and female experiences are treated as a sex-specific deviation from that allegedly universal standard (Bem 1993, 41).

Uno studio recente (Bailey, LaFrance e Dovidio 2019) sottolinea come l'inglese, così come diverse lingue del mondo, manifestino caratteristiche tali da poter essere considerate forme di androcentrismo linguistico, secondo tre possibili percorsi: « (a) men being more frequently instantiated than women, (b) masculinity being more “ideal” than femininity, and/or (c) masculinity being more common than femininity» (Bailey et al. 2019, 1). In questa tesi utilizzo il termine *androcentrismo*, o *androcentrico*, con riferimento proprio a queste tre caratteristiche rintracciabili sia nella lingua italiana, a livello di uso o di sistema, sia nel linguaggio dei media.

1.3 Dati e metodologie di ricerca

Al fine di raggiungere l'obiettivo generale della tesi, così come dettagliato nel paragrafo precedente, ho deciso di costruire e analizzare tre *corpora* di contenuti dell'informazione: un *corpus* TG che raccoglie testi di telegiornali italiani, e due diversi *corpora* di notizie pubblicate dalla stampa quotidiana nazionale. Il *corpus* TG include la trascrizione integrale dei testi del TG1, del TG2 e del TG5, trasmessi in prima serata nei mesi di gennaio 2018, 2019, 2020. Il primo *corpus* di quotidiani, che ho chiamato *corpus* CRS, raccoglie una selezione di articoli pubblicati dal *Corriere della sera*, *Il Resto del Carlino*, *La Stampa*, nei mesi di gennaio 2018 e 2019. Il secondo *corpus* di notizie quotidiane, che ho chiamato *corpus* ACR, raccoglie tutti gli articoli pubblicati dal notiziario italiano dell'agenzia di stampa ANSA e dalle

edizioni del *Corriere della sera* e *Il Resto del Carlino* nei mesi di gennaio 2018, 2019, 2020.

Dal punto di vista metodologico, ho analizzato i *corpora* attraverso due approcci diversi: uno *corpus-driven* e uno *corpus-based* (Tognini Bonelli 2001).

Nel primo caso, applicato al *corpus* TG, ho esplorato le caratteristiche lessicali e testuali del *corpus*, lasciandomi “guidare” dai risultati ottenuti per descriverlo in termini di rappresentazioni di genere. Non sono partita da qualche ipotesi *a priori* da verificare, ma da una domanda generale: quali sono le rappresentazioni di donne e uomini che emergono dal lessico e dalla struttura testuale del *corpus*? Fra le diverse metodologie *corpus-driven* ho optato per un’analisi semiautomatica computerizzata con pretrattamento manuale dei dati, un tipo di analisi testuale supportata da modelli e metodi della linguistica computazionale, che permette di includere variabili categoriali (per esempio il genere, femminile o maschile, del/la parlante) direttamente nella fase empirica della ricerca. Utilizzando variabili categoriali per descrivere alcune caratteristiche strutturali dei notiziari televisivi, ho potuto così integrare l’analisi linguistica con un’analisi mediale e, al contempo, basare l’analisi mediale su evidenze linguistiche, di tipo sia lessicale che testuale.

Nel secondo caso, *corpus-based*, ho invece utilizzato i *corpora* per verificare due ipotesi di cui anticipo una formulazione sintetica: 1) la persistenza di androcentrismo linguistico nel linguaggio dell’informazione, attestato da Sabatini (1987) e successive ricerche (Burr 1995; Pescia 2010; Cavagnoli 2013; Robustelli 2014; Formato 2016; Castanetto e Ondelli 2020) e 2) la persistenza di stereotipi di genere pertinenti la sfera professionale attestata dai medesimi studi e anche da analisi su *web corpora* (Jacqmain 2005; Nardone 2016, 2018).

Ho verificato la prima ipotesi attraverso lo studio del lemma *ministra/o* nel *corpus* TG e nel *corpus* CRS, che ho analizzato con un approccio quali-quantitativo. Tramite un’analisi quantitativa computerizzata, ho calcolato le frequenze del lemma *ministra/o*, in tutte le sue quattro forme (femminile singolare, femminile

plurale, maschile singolare, maschile plurale), ed estratto le collocazioni da sottoporre successivamente a un'analisi qualitativa manuale, finalizzata a stabilire il genere semantico del/la referente di ciascuna forma linguistica. *Ministra/o* è un nome che, indicando una posizione apicale, storicamente appannaggio maschile, tende a essere declinato al maschile anche quando si riferisce a una donna (Formato 2016; Azzalini 2021a). Per stabilire il genere del/la referente è dunque necessaria un'analisi manuale che si basi sulla conoscenza del contesto extralinguistico di riferimento (per esempio, per sapere che la frase *il ministro della salute Grillo* fa riferimento a una donna, devo sapere che referente di *ministro* è la ministra Giulia Grillo). Per cercare di capire quali possono essere i fattori che influenzano l'androcentrismo del linguaggio giornalistico, ho anche analizzato le strategie discorsive utilizzate per nominare ministre e ministri. E per verificare la possibile incidenza del contesto politico, delle linee editoriali e del tipo di media sul grado di androcentrismo rilevato, ho effettuato analisi comparative, basate su calcoli statistici, per anno, testata giornalistica e *corpus*.

Per quanto riguarda la seconda ipotesi, che ho verificato tramite un'analisi del *corpus* ACR, ho adottato un approccio distribuzionale, estraendo dal *corpus* la rappresentazione vettoriale (*word embedding*) di una selezione di nomi di professioni e di cariche politiche per verificarne l'associazione semantica latente con il genere femminile o maschile, la correlazione o meno fra questo grado di associazione di genere (*gender-score*) con dati di realtà, ovvero l'incidenza femminile nella corrispettiva professione o carica politica, e, infine, il ruolo esercitato dal genere grammaticale sulle associazioni di genere semantiche latenti (Stubbs 1996; Bolukbasi et al. 2016; Caliskan et al. 2017; Biasion et al. 2020; Mc Curdy e Serbetçi 2020). Le associazioni semantiche latenti sono in grado di rilevare eventuali stereotipi che associano il nome di una professione a un genere sociale (donna o uomo); la correlazione con le percentuali di occupazione femminile fa emergere quanto eventuali stereotipi si fondino, o meno, su dati di realtà (il fatto che il lemma *ingegnera/e* sia più associato al genere maschile e il lemma *ostetrica/o* a quello femminile può dipendere anche dal fatto che la prima è una professione

prevalentemente esercitata da uomini, mentre la seconda da donne, e come tali, entrambe, sono percepite ed espresse nell'uso della lingua); infine, l'analisi di quanto il genere grammaticale di un nome incida sulla rappresentazione semantica del nome stesso permette di stabilire quanto la grammatica italiana possa mitigare o meno gli stereotipi di genere.

1.4 Struttura della tesi

La tesi è articolata in otto capitoli, compresa la presente Introduzione (Capitolo 1) e le Conclusioni (Capitolo 8). I primi due capitoli successivi a quello introduttivo (Capitolo 2 e Capitolo 3) offrono la cornice contestuale e teorica entro cui si collocano il Capitolo 5, il Capitolo 6 e il Capitolo 7 che presentano le tre ricerche empiriche, anticipate dalla descrizione dei tre *corpora* dettagliata nel Capitolo 4.

Il Capitolo 2 introduce la letteratura in materia di lingua e genere, illustrando brevemente i principali modelli teorici e le evidenze empiriche della ricerca in ambito internazionale (§ 2.1) e nazionale (§ 2.2), e focalizzandosi, infine, sulla relazione fra genere grammaticale e genere semantico (§ 2.3), che è attualmente il tema dominante nella ricerca su lingua e genere in Italia, su cui sono concentrati il Capitolo 6 e il Capitolo 7 della tesi.

Il Capitolo 3 è dedicato alla letteratura in materia di media e genere, introducendo brevemente i principali modelli teorici e le evidenze empiriche della ricerca in ambito internazionale (§ 3.1) e nazionale (§ 3.2), e concentrandosi, infine (§3.3), sulla relazione fra il potere dell'informazione di sostenere e promuovere l'*empowerment* femminile e i fattori mediali e socio-culturali che intervengono nel limitare questo potenziale, relazione presa in considerazione in tutte e tre le ricerche empiriche, e approfondita in particolare nel Capitolo 5.

Il Capitolo 4 definisce il concetto di *corpus* e presenta i criteri di selezione dei testi raccolti nei tre *corpora* analizzati (*corpus* TG, *corpus* CRS e *corpus* ACR) di cui descrive il processo di costruzione e le caratteristiche, in termini di dimensione e ricchezza di vocabolario.

Il Capitolo 5 presenta la ricerca sulle rappresentazioni di genere nel *corpus* TG, attraverso un'analisi *corpus-driven* che si focalizza sulle donne e sugli uomini, in una prospettiva comparativa, sia come fonti sia come argomento dell'informazione. Poiché la caratteristica delle analisi *corpus-driven* è quella di esplorare la natura di un testo, senza avanzare ipotesi *a priori* (Tognini Bonelli 2001), il capitolo è introdotto direttamente da domande di ricerca generali, formulate partendo dalla cornice teorica e contestuale offerta nel Capitolo 2 e nel Capitolo 3, utili a orientare l'analisi in una prospettiva di genere e a comprendere l'eventuale incidenza di fattori contestuali e mediali sui risultati. Quali sono le rappresentazioni di donne e uomini come soggetto/fonte e come oggetto/argomento di notizia, nel *corpus* TG? Ci sono differenze fra i tre periodi considerati e le tre testate giornalistiche campione?

Il Capitolo 6 verifica l'ipotesi che il linguaggio dell'informazione sia androcentrico per via (anche) di una resistenza a usare forme femminili per nominare donne che ricoprono cariche apicali, come quella di ministra, attraverso uno studio del lemma *ministra/o*, basato sui *corpora* TG e CRS, analizzati anche in una prospettiva comparativa. A partire da tre domande, la ricerca misura la frequenza del lemma *ministra/o* nelle sue quattro forme grammaticali (*ministra*, *ministre*, *ministro*, *ministri*); analizza con quale frequenza il lemma è declinato in modo coerente con il genere semantico di riferimento; infine indaga le espressioni linguistiche ricorrenti nei *corpora* per far riferimento alle ministre e ai ministri citati.

Il Capitolo 7 verifica l'ipotesi che il linguaggio dell'informazione veicoli stereotipi di genere pertinenti la sfera professionale, attraverso un'analisi distribuzionale del *corpus* ACR. Partendo da tre domande di ricerca, lo studio indaga le associazioni di genere nella rappresentazione semantica, più precisamente negli *embeddings* di (una selezione scelta di) nomi di professioni e cariche politiche occorrenti nel *corpus* ACR; verifica poi se c'è una correlazione significativa o meno fra il grado di associazione di genere di un nome e il tasso occupazionale di donne nella professione o carica politica corrispondente al nome, infine analizza il ruolo del genere grammaticale nel confermare associazioni stereotipiche o mitigarle.

Il Capitolo 8 prova a tracciare alcune conclusioni generali, collegando i principali risultati delle tre ricerche empiriche e interpretandoli entro il comune quadro teorico e contestuale di riferimento, così come delineato nei primi capitoli della tesi. Chiude il Capitolo 8 un paragrafo dedicato ai limiti di questa tesi e ai possibili sviluppi di una linea di ricerca che integri l'analisi linguistica con l'analisi mediale in una prospettiva di genere.

Capitolo 2 Lingua e genere

Questo capitolo ha lo scopo di fornire la cornice contestuale e teorica entro cui nascono e si collocano gli studi sulla relazione fra lingua e genere. Il primo paragrafo introduce i principali modelli teorici che si sono sviluppati in ambito internazionale, seguendo la tassonomia proposta da Claudia Bianchi (2006, 6) che, sistematizzando la letteratura in materia e basandosi su precedenti lavori anglosassoni (in particolare Cameron 2003), individua quattro modelli: il «modello del deficit», il «modello del dominio», il «modello della differenza», il «modello dinamico o performativo». Il secondo è dedicato agli studi su lingua e genere in Italia ed è suddiviso in due sotto-paragrafi: il primo pertinente la ricerca sulla lingua parlata dalle donne, che riguarda un ristretto numero di studi, di cui si fornisce una rassegna ragionata; il secondo pertinente la ricerca sulla lingua che parla delle donne, che è stata avviata dai lavori pionieristici di Alma Sabatini ed è diventata la linea di ricerca dominante in Italia. Il terzo paragrafo approfondisce il tema della relazione fra genere grammaticale e genere semantico, che è attualmente il tema prevalente nella ricerca sulla relazione fra la struttura, e l'uso, della lingua italiana e le rappresentazioni di genere, su cui sono concentrati il Capitolo 6 e il Capitolo 7 della tesi, ai quali è rimandata una rassegna approfondita e ragionata degli studi sulla lingua che parla delle donne.

2.1 Modelli teorici ed evidenze empiriche in ambito internazionale

Seguendo la classificazione proposta da Claudia Bianchi (2006), l'evoluzione degli studi su lingua e genere nella letteratura internazionale può essere ricondotta a quattro modelli teorici principali, che sono nati progressivamente nel tempo, seguendo anche l'evolversi del pensiero femminista: il «modello del deficit» (Lakoff 1973), il «modello del dominio» (Spender 1980, 1985), il «modello della differenza», (Tannen 1990, Gray 1992), il «modello dinamico o performativo» (Butler 1990, 1999/2017; Holmes 1997). Si tratta di una tassonomia non esaustiva dell'ormai vasta letteratura in materia, che copre molte lingue del mondo, e riguarda molte

sotto-discipline linguistiche. Tuttavia mi sembra funzionale a ordinare, ancorché astrattamente, l'ampia produzione di studi prodotta nel corso del tempo, individuandone i capisaldi per fornire una cornice contestuale e teorica entro cui collocare la riflessione su lingua e genere, anche in ambito nazionale.

2.1.1 «Talking like a lady» (Lakoff 1973)

Quando Robin Lakoff (1973) pubblica il suo lavoro a metà degli anni Settanta, la relazione fra lingua e genere è una questione già parzialmente indagata in alcune ricerche che all'inizio del Novecento avevano analizzato differenze di comportamenti linguistici basate sul sesso (Jespersen 1922), prevalentemente nell'ambito di un più ampio interesse antropologico (Malinowsky 1929; Sapir 1929; Haas 1944; Flannery 1946; Lévi-Strauss 1955). Basandosi su osservazioni etnografiche, Otto Jespersen (1922) avanza l'ipotesi dell'esistenza di un bilinguismo su base sessuale in alcune popolazioni autoctone dei Caraibi, dell'Africa, dell'America latina, ma anche popolazioni tedesche o scandinave immigrate in Nord America. Inizialmente individua nelle diverse lingue di queste comunità alcuni tratti tipicamente femminili, poi li proietta anche su lingue europee come il francese e l'inglese, trovandone attestazione in brani letterari o in precedenti studi linguistici. Arrivando a profilare una lingua delle donne, il linguista danese ne indica alcune caratteristiche specifiche: presenza di tabù verbali (come per esempio non pronunciare il nome del marito o di uomini membri della famiglia del marito) e conseguente necessità di ricorrere a perifrasi; conservazione della lingua originaria, o del dialetto, in caso di migrazione verso un paese di altra lingua; lessico eufemistico; ampio ricorso ad avverbi e modificatori di quantità; vocabolario ridotto; sintassi povera di subordinate; discorso frammentario e interrotto. Caratteristiche queste ultime che vengono considerate dall'autore come una conseguenza della mancanza di logicità nelle donne (Jespersen 1922, 253).

Tesi simili a questa sono sviluppate negli studi coevi di Bronislaw Malinowsky (1929) e Edward Sapir (1929), e in quelli successivi di Mary Haas (1944), Regina Flannery (1946) e Claude Lévi-Strauss (1955). Partendo dall'osservazione di

comportamenti linguistici tipici di una particolare popolazione, perlopiù minoranze linguistiche, come per esempio i parlanti Koasati in Louisiana (Haas 1944) o i nativi americani Atsina, soprannominati Gros Ventre (Flannery 1946), o popolazioni non occidentali allora ancora poco conosciute, come gli abitanti della Melanesia (Malinowsky 1929) o alcune comunità indigene del Sudamerica (Lévi-Strauss 1955), questi lavori postulano l'esistenza di un bilinguismo su base sessuale, considerando il linguaggio maschile come lingua neutra, e quello femminile come una devianza o anomalia rispetto alla norma (Bodine 1975; Aebischer 1985/1988; Panighel 2014).

Anche le riflessioni di Lakoff (1973) ipotizzano differenze linguistiche fra donne e uomini, ma da una prospettiva che, da un lato, riconduce le differenze osservate nella lingua (inglese) a disuguaglianze sociali fra donne e uomini, e, dall'altro, attribuisce a queste differenze un carattere di universalità, piuttosto che di specificità tipica di qualche popolazione o cultura minoritaria. Scrive infatti la linguista americana: «I do feel that the majority of the claims I make will hold for the majority of speakers of English; that, in fact, much may, *mutatis mutandis*, be universal» (Lakoff 1973, 47).

Lakoff (1973) elenca e argomenta queste differenze in modo poco organizzato, ma chiaro anche alla luce della riflessione accademica che ne è seguita. L'ipotesi da cui parte è che le bambine e i bambini da piccoli apprendono un linguaggio con caratteristiche femminili, essendo esposti prevalentemente alla lingua della madre e delle figure di donna che si occupano del loro accudimento durante l'infanzia. Dai 5 ai 10 anni iniziano a comprendere, se sono bambine, che devono parlare «like a lady» (Lakoff 1973, 48) evitando di usare un linguaggio volgare e diretto, che non si addice alla buona educazione socialmente attesa per una ragazza ed è più adeguato ai ragazzi che crescono nell'aspettativa sociale condivisa di diventare uomini d'affari o politici, in generale uomini di potere. A differenza delle ragazze, che crescono nella prospettiva di avere un ruolo subordinato, o fuori dal sistema. Questo processo di apprendimento della lingua dà origine, secondo Lakoff, a una differenziazione che caratterizza la lingua delle donne come lingua secondaria, di minor potere (vs. la lingua degli uomini, che è la

lingua dominante), in relazione alla loro collocazione sociale. Questo spiega perché, sempre secondo Lakoff (1973), la lingua delle donne non è una lingua parlata soltanto dalle donne, ma anche da alcuni gruppi sociali maschili, come gli accademici, gli hippy e gli omosessuali, che si oppongono al sistema di potere dominante, o vi si collocano ai margini.

I tratti linguistici indicativi di questa differenziazione, basata su una disuguaglianza sociale e politica, sono: la maggior precisione delle donne nel nominare i colori; la tendenza a usare esclamazioni diverse dagli uomini (le donne usano espressioni tipiche del lamento come *oh dear!*, perché a loro è permesso lamentarsi, mentre agli uomini è permesso usare imprecazioni più forti come *shit!*); l'esistenza di vocaboli diversi (con aggettivi che esprimono ammirazione o approvazione per qualcosa, es. *adorable* o *charming*, per le donne, e aggettivi più neutri, es. *great* o *terrific*, per gli uomini); il maggior uso da parte delle donne di *tag-question*, in particolare quelle funzionali ad ammorbidire un'opinione espressa, per evitare conflitti, essere rassicurate, o essere gentili, essendo la *politeness* un tratto socialmente attribuito alle donne.

2.1.2 «Talking about women» (Lakoff 1973)

Per quanto riguarda la lingua che parla delle donne («talking about women», Lakoff 1973, 57), Lakoff individua una serie di asimmetrie linguistiche, prevalentemente di natura semantica, che rendono trasparente la collocazione subordinata delle donne entro il sistema di potere basato sul genere. Per esempio, l'uso di *lady* vs. *gentlemen*: quest'ultimo è raramente usato e non come sinonimo di *man*, ma come titolo, mentre *lady* viene usato come sinonimo di *woman* per varie finalità, come per esempio nobilitare alcune professioni socialmente svalutate, in funzione eufemistica (es. *cleaning lady*), oppure per evitare di usare *woman*, termine che conterrebbe, secondo la linguista statunitense, un tratto semantico che rimanda alla sfera sessuale. L'uso di *mistress* vs. *master*: il primo termine presenta una restrizione rispetto al significato originario di *master-servant* da cui ha origine, indicando una donna nella sua relazione sessuale con un uomo, mentre il secondo

termine, che grammaticalmente è simmetrico al primo, indica un uomo abile a fare qualcosa, risultando così semanticamente asimmetrico rispetto al suo corrispettivo femminile. L'uso di *professional*: in questo caso la parola è unica ma nel cambio di riferimento dal maschile al femminile cambia significato, perché riferito a un uomo significa *professionista*, riferito a una donna significa *prostituta*, riducendo il campo semantico originale alla sola sfera sessuale. E ancora l'uso di *spinster* e *bachelor*, *man* e *wife*, *widow* e *widower*, *Mr.* e *Mrs.* vs. *Miss.* Tutti questi termini sono apparentemente simmetrici, perché grammaticalmente lo sono, ma, se analizzati nel loro uso quotidiano, rilevano asimmetrie semantiche che testimoniano disuguaglianze sociali.

2.1.3 «Man made language» (Spender 1980)

Il «modello del dominio» è quello articolato da Dale Spender in *Man made language* (1980, 1985), dove, attraverso un'argomentazione non sempre lineare, l'autrice afferma che il linguaggio delle donne è subordinato a quello maschile per via di una storica dominanza degli uomini sulle donne. Secondo Spender (1980, 1985), la superiorità maschile è un mito, e su questo mito ha posto le basi il potere patriarcale, anche attraverso il linguaggio che è servito agli uomini per costruire e rafforzare questa credenza, dando sostanza all'ordine patriarcale stesso. Nella struttura societaria patriarcale, il linguaggio maschile rappresenta la norma, quello femminile la deviazione, dando forma a una dominanza maschile che è testimoniata sia nelle modalità di conversazione fra donne e uomini, sia nel modo in cui la lingua rappresenta uomini e donne attraverso le parole: per esempio, la parola *man*, in inglese, è usata per indicare gli uomini, nella loro specificità di genere, ma anche per indicare l'intera specie, mentre il corrispettivo femminile *woman* indica soltanto il genere femminile. Più in generale, Spender (1980, 1985) sostiene che le parole accettabili e comuni hanno più probabilità di esprimere esperienze maschili che femminili, perché attraverso la lingua gli uomini hanno costruito il loro dominio; e individua così nel linguaggio una manifestazione dell'androcentrismo tipico del patriarcato.

Tornando alla tassonomia di Bianchi (2006), il «modello del deficit» (Lakoff 1973) e il «modello del dominio» (Spender 1980, 1985) vengono successivamente criticati: il primo perché basato su intuizioni soggettive dell'autrice, piuttosto che su dati linguistici adeguatamente raccolti, e perché riproduce involontariamente gli stereotipi di genere, attribuendo alle donne una posizione di marginalità e subordinazione nella vita pubblica in generale e nel mondo degli affari e della politica in particolare; il secondo per il suo estremismo e una certa confusione teorica (Bianchi 2006, 89). Tuttavia entrambi questi modelli, avendo posto attenzione non solo sulla lingua, ma anche sulle relazioni sociali fra donne e uomini codificate nella lingua, sono stati cruciali per spingere le ricerche successive verso una maggiore focalizzazione sulle complesse dinamiche di interazione fra lingua, società e genere (Bucholtz 2004) e ancora oggi rappresentano un punto di riferimento importante, in particolare per alcuni studi italiani (cfr. Cavagnoli 2013, 2015, 2019; Cavagnoli e Dragotto 2021; Dragotto 2012, 2016).

2.1.4 «*Different but equal*» (Tannen 1990; Gray 1992, cit. in Cameron 2003)

Il «modello della differenza» è quello che contraddistingue diversi libri divulgativi, pubblicati allo scopo di fornire consigli per il miglioramento della comunicazione, e dunque dei rapporti, fra donne e uomini. Non solo *You just don't understand* di Debora Tannen (1990), citato da Bianchi (2006, 8), ma anche *Men are from Mars, women are from Venus* di John Gray (1992), citato da Cameron (2003). Entrambi questi volumi, divenuti bestseller negli anni Novanta del Novecento², partono dal presupposto che donne e uomini hanno stili comunicativi diversi perché crescono in sotto-culture diverse, con norme sociali differenti, a volte lontane, come possono essere Marte e Venere, i due pianeti citati nel titolo del volume di Gray (1992). Le donne avrebbero uno stile comunicativo cooperativo e non gerarchico, che tende

² Il volume di Tannen (1990) è rimasto per quattro anni nella lista dei best-seller del New York Times; fonte: https://it.scribd.com/book/163656132/You-Just-Don-t-Understand-Women-and-Men-in-Conversation?utm_medium=cpc&utm_source=google_search&utm_campaign=3Q_Google_DSA_NB_All&utm_device=c&gclid=Ci0KCQjw4eaJBhDMARisANhrQACS7s4bi8YTff4_kjIAJDtdMyEA5qOij_nzxPOZnlUDSMpNM-yYwk70aAjqQEALw_wcB (ultimo accesso 19 novembre 2021); il volume di Gray risulta al primo posto dei libri più venduti nella classifica di US Today; fonte: <https://eu.usatoday.com/story/life/books/2013/10/30/book-list-excerpts/3318571/>.

a stabilire relazioni interpersonali e amicali, nonché a esprimere solidarietà. Gli uomini invece avrebbero uno stile comunicativo competitivo e gerarchico, focalizzato sui fatti piuttosto che sulle relazioni e sui sentimenti. In questo modello, le differenze linguistiche non vengono interpretate né come il riflesso di disuguaglianze sociali (Lakoff 1973), né come la codifica del dominio patriarcale (Spender 1980, 1985), bensì come segni di stili conversazionali diversi. E gli stili femminili vengono considerati in taluni casi superiori, nel senso di più performanti, per esempio, per mantenere strette le relazioni, in diversi ambiti della sfera pubblica e privata. Tuttavia – osserva Cameron (2003, 458) – «neither Tannen nor Gray overtly argued for the superiority of women. Rather both took the line that the sexes are 'different but equal' and need to understand and accept one another's differences in order to avoid misunderstandings».

Secondo Cameron (2003, 458), i testi di Tannen (1990) e Gray (1992) rappresentano una svolta, ma soltanto apparente. I primi testi contenenti consigli per una comunicazione più efficiente fra donne e uomini, pur riconoscendo alle donne specifiche abilità linguistiche, chiedevano in sostanza alle donne di «talk like a man» (Cameron 1995, 185-192), dimostrando così la loro adesione a un modello che guardava alle donne come “deficitarie” e agli uomini come “dominanti”. Tannen (1990) e Gray (1992) superano entrambi questi paradigmi, primo perché affermano una differenza nell'uguaglianza, secondo perché nel sottolineare le capacità conversazionali delle donne, riconoscono implicitamente una superiorità femminile, o perlomeno così i loro bestseller sono stati interpretati dal pubblico. Tuttavia, né Tannen (1990) né Gray (1992) superano l'idea stereotipata secondo la quale donne e uomini hanno atteggiamenti linguistici diversi, indipendentemente da altre condizioni (culturali, economico-sociali, generazionali, e così via). Attestano piuttosto un nuovo giudizio di valore che emerge contestualmente all'emergenza di un nuovo ideale linguistico, nell'ambito di un mercato del lavoro che, con lo sviluppo delle nuove tecnologie digitali, valorizza chi eccelle in attività verbali come il *problem-solving* cooperativo, la costruzione del rapporto, l'auto-riflessività emotiva, l'espressione di empatia, e così via (Cameron 200, 458), tutte competenze stereotipicamente associate alle donne, alle quali, si aggiungono competenze

tradizionalmente associate agli uomini come l'autorità, l'intraprendenza e la capacità di leadership. In questo ideale del/la *good communicator* (Cameron 2003, 463) è evidente che donne e uomini sono assunti come parlanti diversi e complementari fra loro, secondo una visione stereotipata della relazione fra lingua e genere.

2.1.5 «Doing gender» (Butler 1990, Holmes 1997)

Il «modello dinamico o performativo» si sviluppa a partire dagli anni Novanta del Novecento, con le riflessioni di Judith Butler (1990, 1999/2017) che reinterpreta il concetto di genere in senso, appunto, dinamico e performativo: il genere non è una proprietà fissa e stabile, ma qualcosa che “facciamo” continuamente, attraverso il linguaggio, il quale ha un ruolo centrale nel costruire, consolidare o modificare le identità di genere. Nel formulare la sua teoria, complessa e mutevole nel tempo, Butler rimanda sia alla teoria degli atti linguistici di Jane Austin (1962), sia alla teoria del discorso di Michel Foucault (1969/1971). Il genere è qualcosa che facciamo attraverso una serie di atti linguistici quotidiani, tuttavia entro i confini di norme sociali condivise. In un passaggio di *Bodies that matter*, Butler (1993) riporta un esempio che chiarisce il concetto di performatività del genere sullo sfondo di questa nozione linguistica:

Consider the medical interpellation which (the recent emergence of the sonogram notwithstanding) shifts an infant from an “it” to a “she” or a “he”, and in that naming, the girl is “girdled”, brought into the domain of language and kinship through the interpellation of gender (Butler 1993, 7).

L'enunciato del medico è performativo del genere nel senso che attraverso di esso il medico avvia un processo di *girling*, che potremmo tradurre con “femminilizzazione”. E «does not end there» (Butler 1993, 7). Questa femminilizzazione è infatti reiterata nel tempo da varie autorità, compresa l'autorità del soggetto, in termini di rinforzo, o anche di contestazione, attraverso atti linguistici quotidiani; sempre e comunque entro i limiti di un discorso che porta

con sé norme e ruoli sociali situati storicamente, secondo la nozione di “discorso” elaborata da Foucault ne *L'archéologie du savoir* (1969/1971).

Per il filosofo francese il discorso è, come per Butler, un atto linguistico non intenzionale, che porta con sé, oltre al suo contenuto, le regole di significazione di un enunciato. Non solo regole sintattiche e semantiche, ma anche contestuali, ovvero regole sociali, situate, nello spazio e nel tempo. Un discorso è sempre un modo di parlare e al contempo di interpretare la realtà, entro possibilità concettuali che costituiscono i confini del pensiero, così come della lingua, che costruisce e insieme esprime quel pensiero entro un dato contesto d'uso e un dato periodo storico.

Uno studio interessante che si basa su questo modello teorico è per esempio quello di Janet Holmes (1997), che dimostra come nell'interazione faccia a faccia, le persone usano il linguaggio per esprimere, creare, sfidare e sovvertire una serie di significati sociali, ma nel farlo attingono a norme sociolinguistiche stabilite. Attraverso le evidenze emerse da un'ampia rassegna di studi linguistici che attestano per le donne maggiori varianti fonologiche e morfologiche, rispetto agli uomini, Holmes (1997) sostiene che una spiegazione esaustiva del fenomeno è data dalla prospettiva che assume il linguaggio come espressione di un'identità di genere, entro i confini delle aspettative sociali di una comunità, detto altrimenti come luogo di produzione e riproduzione del discorso.

From this perspective, women's identity is signaled not so much by the choice of particular linguistic variants which contrast with those preferred by men, but rather by the ways in which women are often required to use language to construct a much wider range of social identities and express a wider range of social roles than men (Holmes 1997, 199).

A titolo esemplificativo, Holmes (1997) riporta i frammenti di due dialoghi, uno spontaneo fra due amiche, raccolto in un *corpus* di parlato neozelandese, e uno fra due uomini, ripreso da una pubblicità televisiva, sempre in inglese della Nuova Zelanda. Attraverso un'analisi dettagliata di questi frammenti, l'autrice

dimostra come le diverse scelte linguistiche di particelle pragmatiche e interattive, così come alcune scelte lessicali, da parte di donne e uomini, rispondono a precise funzioni discorsive e riflettono precisi significati sociali, entro un determinato contesto, perché, in generale, donne e uomini operano scelte linguistiche funzionali a conformarsi con le idee di femminilità e mascolinità dominanti. Per esempio, lo stralcio sul dialogo fra le due amiche riporta la risposta di una delle due, Helen, a una domanda dell'altra, Joan, sulle condizioni di salute del padre. Helen non risponde direttamente alla domanda dell'amica ma inserisce la risposta in una lunga storia che racconta del tentativo di portare la figlia, e un'amica della figlia, in piscina, per convincerla in seguito a far visita al nonno. La storia, costruita con una narrativa classica che usa molte sequenze temporali per ricostruire un evento di per sé molto breve, porta Helen a inserire molti dettagli, in un racconto che le permette di costruire la sua identità di «“good” daughter» e «“good” mother» (Holmes 1997, 207), in senso conservativo rispetto ai canoni della Nuova Zelanda, a cui il suo *status* reale non è conforme, essendo Helen una madre single.

I quattro modelli a cui è possibile ricondurre i principali studi linguistici su lingua e genere che si sono sviluppati in ambito internazionale corrispondono, come anticipato più sopra, a diverse fasi del femminismo, come i modelli a cui possono essere ricondotti i principali studi su media e genere, delineati nel Capitolo 3. In entrambi i casi, si tratta di modelli non necessariamente mutualmente esclusivi, e non necessariamente storicizzati. Vale a dire che, da un lato, le ricerche empiriche possono aderire a più di un modello, per esempio Spender (1980, 1985) riprende e approfondisce diversi fenomeni linguistici attestati da Lakoff (1973), concordando con quest'ultima che sono espressioni di relazioni di genere gerarchiche. Dall'altro, un modello pur essendo nato in un preciso momento storico del femminismo può rappresentare ancora oggi una cornice teorica di riferimento. Questo è particolarmente evidente nella ricerca che si è sviluppata in ambito nazionale. Probabilmente anche per effetto della dominanza della linea di ricerca avviata da Sabatini (1986, 1987), focalizzata sul sessismo nella lingua italiana, la maggior parte delle ricerche linguistiche che indagano la relazione fra lingua

italiana e genere fa riferimento al «modello del deficit» (Lakoff 1973) o al «modello del dominio» (Spender 1980, 1985), o a entrambi i modelli, come risulterà chiaro dalla lettura del paragrafo 2.2.

2.2 Origine e sviluppo degli studi in ambito nazionale

Come premesso nel Capitolo 1, in Italia, l'origine della riflessione sulla relazione fra lingua e genere viene tradizionalmente attribuita ai lavori di Alma Sabatini (1986, 1987), che sono focalizzati sulla lingua che parla delle donne e avviano una linea di ricerca che ha maggior successo rispetto a quella sulla lingua parlata dalle donne. Quest'ultima ha una tradizione più antica, risalendo all'inizio del Novecento (cfr. Tagliavini 1938; Berruto 1977; Attili e Benigni 1977, 1979) ma è prevalentemente concentrata sulla varianza linguistica in ambito dialettologico e, a partire dagli anni Ottanta, viene messa in discussione, per via di una certa debolezza metodologica e teorica, e in larga parte abbandonata a favore di un nuovo approccio che assume la lingua delle donne come costruzione sociale (Berretta 1983). Anche questa nuova linea di ricerca, però, risulta poco produttiva, nel senso che, purtroppo, produce un numero limitato di studi, lasciando parzialmente scoperta la conoscenza sull'italiano parlato dalle donne (e dagli uomini), un aspetto importante della relazione fra lingua e genere, e fortemente correlato a quello del sessismo nell'uso dell'italiano, poiché, come scrive bene Patrizia Violi (1986, 40):

La relazione è circolare: nella lingua si codificano le rappresentazioni collettive del femminile, che le donne stesse riprodurranno nella costruzione della propria immagine, immagine che a sua volta si troverà a interagire con i codici sociali.

Dal punto di vista metodologico, il limite principale della ricerca sul parlato femminile è rappresentato da un'ampia variabilità nelle modalità e nei criteri di raccolta dei dati e dalla generalizzazione di risultati che si riferiscono invece a condizioni di ricerca specifica, attestando in certi casi risultati incongruenti. Per esempio, la maggior conservatività degli usi linguistici femminili è attestata all'inizio del secolo scorso da Tagliavini (1938), che tuttavia esprime quella che Berruto

(1980, 134) definisce una «concezione divergente», affermando che le donne, per via del loro ruolo sociale, da un lato, sono più conservative, perché, lontane dalla vita pubblica e dai contatti esterni, rifuggono o accolgono più tardi le innovazioni che vengono dall'esterno; dall'altro, precorrono gli uomini nelle innovazioni linguistiche spontanee, sempre perché, lontane dalla vita pubblica e dai contatti esterni, sono anche lontane dalla scuola e dalla cultura letteraria, dove si coltiva la tradizione.

Dal punto di vista teorico, l'approccio dialettologico e sociolinguistico è criticato da Patrizia Violi (1986) perché trascura il ruolo del linguaggio nella costruzione della realtà e delle identità di genere, assumendo la lingua semplicemente come riflesso di dinamiche e relazioni sociali, con la conseguente riduzione delle varietà linguistiche a variabili sociali:

ciò che sfugge a questa prospettiva è la capacità di considerare il linguaggio come diretta riproduzione di realtà, come un luogo, ed uno dei più importanti, di costruzione e manipolazione di potere e competenze, dove non soltanto si riflettono le dinamiche sociali, ma soprattutto si creano e producono immagini e rappresentazioni collettive del femminile (Violi 1986, 119).

Una prospettiva che va parzialmente nella direzione auspicata, ma non intrapresa, da Patrizia Violi (1986), la quale si limita a riflessioni teoriche di carattere semiotico, è quella di Monica Berretta (1983) che assume le caratteristiche associate alla lingua delle donne come costruzioni sociali, avviando così la ricerca nazionale verso un modello teorico che indaga i tratti linguistici femminili (e maschili) come stereotipi. Non però, come vedremo, stereotipi latenti, in maniera trasversale e generale al parlato delle donne, in contrasto al parlato degli uomini, bensì stereotipi palesi tipici prevalentemente della classe borghese, soprattutto maschile, e anche degli studi sociolinguistici.

2.2.1 *La lingua delle donne*

Partendo dalla riflessione che i dati linguistici raccolti su scala nazionale e internazionale sono di difficile interpretazione, testimoniando varietà di genere prevalentemente correlate ad altre variabili sociali, quali l'età, la classe sociale, il grado di istruzione, Berretta (1983) rovescia l'approccio delle ricerche che l'hanno preceduta e studia il linguaggio delle donne non come dato oggettivo, bensì come costruzione sociale:

Il tema non è dunque direttamente il linguaggio femminile inteso come dato, quanto le sue regole socialmente condivise: dunque, appunto, la retorica popolare del discorso femminile, o se si vuole la realtà di questo in termini di costruzione sociale (Berretta 1983, 215).

Sperimentando un nuovo approccio metodologico, Berretta (1983) parte dall'ipotesi che le caratteristiche generalmente attribuite al linguaggio parlato dalle donne, e dagli uomini, in particolare le specificità osservate da Lakoff (1973), siano costruzioni sociali, e la verifica per via sperimentale, attraverso un questionario e una prova di attribuzione dell'autorialità di testi scritti da donne e uomini. Per il questionario, seleziona 174 persone, donne e uomini, rappresentative di diverse classi di età e sociali, chiedendo loro di rispondere a 13 domande a risposta chiusa, motivandone la risposta, in uno spazio riservato ai commenti. Il questionario contiene domande generiche, come per esempio «secondo lei c'è differenza fra uomini e donne nel modo di parlare?», o «si può dire che gli uomini, o le donne, parlino "meglio" o "peggio"?» e domande più specifiche, come «usano più parolacce gli uomini o le donne?» (Berretta 1983, 220); domande relative allo stereotipo secondo il quale le donne parlano più degli uomini, parlano di argomenti diversi da quelli di cui parlano gli uomini (prediligendo questioni pertinenti la vita domestica, piuttosto che gli affari, il lavoro, lo sport e il sesso), in modo più gentile e cortese e meno volgare, in modo più corretto (più vicino allo standard), sono più intuitive (capiscono più in fretta leggendo o ascoltando) e apprendono la lingua prima degli uomini, ma, a differenza di questi, costruiscono discorsi meno lineari e meno ordinati (Attili e Benigni 1977, 1979). I risultati del questionario confermano

che, in generale, alle donne sono associate caratteristiche come la loquacità, la precocità, la cortesia, la correttezza formale, mentre agli uomini viene associato l'uso di un linguaggio più volgare, ma anche meglio organizzato (con un capo e una coda, senza interruzioni), con formule più «sostenute» (Berretta 1983, 229), che contrastano con le incertezze ritenute tipiche del linguaggio femminile. Disaggregati per sesso, classe sociale e età, i dati mostrano però differenze su base generazionale e di classe: le persone anziane sono più precise e dirette nei giudizi e più fedeli agli stereotipi tradizionali, in particolare gli uomini di classe medio-alta. Minori sono le differenze per genere: vi è una generale tendenza a privilegiare il proprio gruppo sociale, nelle valutazioni, sia da parte delle donne che da parte degli uomini, e una divergenza netta solo sul carattere della gentilezza femminile, uno stereotipo prevalente fra le donne.

Oltre al questionario, Berretta (1983) riprende e analizza i risultati della tesi di una studente dell'Università di Torino (Stradella 1976 cit. in Berretta 1983, 219) che raccoglie 172 testi descrittivi, assegnando a un gruppo di studenti il compito di attribuire l'autorialità del testo a un uomo o a una donna e motivarne la decisione in un'intervista registrata. Nel 66,4% dei casi l'attribuzione di autorialità del testo (a un uomo o a una donna) è corretta e motivata sulla base di stereotipi sociali. I motivi per cui un testo è attribuito a una donna riguardano caratteristiche come la non-oggettività del discorso, che è basato su impressioni, ricordi, divagazioni, fantasie; l'incertezza semantica o tematica; l'uso di diminutivi come *carino* e *bellino*. Le motivazioni principali di attribuzione di un testo a un uomo sono il carattere oggettivo, concreto, denotante mentalità pratica, senza divagazioni. Guardando alle differenze per genere, generazione e classe sociale, Berretta riporta come principale differenza quella pertinente i tratti di incertezza associati alla scrittura femminile: questa associazione, mediamente pari al 37,5%, raggiunge il valore massimo negli uomini di classe medio-alta (40,7%) e minimo nelle donne di classe bassa (15,4%).

Sintetizzando i risultati delle due indagini, Berretta (1983) ritiene di avere verificato l'ipotesi di un'adesione dei/delle parlanti agli stereotipi sul linguaggio

femminile, non però trasversale a tutte le donne e a tutti gli uomini, bensì specifica degli uomini di classe medio-alta, e di gran parte delle donne dello stesso ceto. Inoltre, afferma, senza tuttavia entrare nel merito della questione, che questa visione è tipica anche dei lavori di sociolinguistica: «gli stereotipi correnti sul linguaggio femminile, quelli che si trovano nelle barzellette come nei lavori di sociolinguistica, sono quelli tipici degli uomini di classe medio-alta» (Berretta 1983, 238).

L'approccio critico nei confronti dei lavori di sociolinguistica emerge in modo più articolato e centrale anche in una ricerca realizzata in Francia da Verena Aebischer nel 1985, tradotta e pubblicata in Italia nel 1988. Benché realizzata oltrelpe, e focalizzata sulla lingua francese, mi sembra importante dedicare qualche riga a questa indagine, perché è funzionale, da un lato, a presentare alcune complessità della ricerca sulla lingua parlata dalle donne, e dall'altro, a comprendere lo spostamento del *focus* della ricerca italiana nella direzione di considerare la lingua delle donne come costruzione sociale. Uno spostamento che produce risultati interessanti ma forse non coglie pienamente l'intuizione originaria di Violi (1986), secondo la quale la lingua parlata dalle donne codifica a livello formale, ovvero di struttura fonologica, morfologica e sintattica, le differenze sessuali che influiscono sulla percezione e la costruzione delle identità di genere, individuali e collettive. Per comprendere questo aspetto sarebbe necessario descrivere la lingua (delle donne e degli uomini), assumendola come oggetto di indagine, poi verificarne il grado di stereotipia, sul modello offerto per esempio da Holmes (1997). Aebischer (1985/1988) invece, come Berretta (1983), si limita a studiare gli stereotipi, senza studiare la lingua.

In linea con la riflessione coeva sviluppata oltremanica da Coates e Cameron (1989), Aebischer (1985/1988) parte dal presupposto che le differenze di linguaggio basate sul genere sono stereotipi prodotti dalla conoscenza: non solo la conoscenza popolare, attestata da canzoni, modi di dire, proverbi e persino opere d'arte figurativa, ma la conoscenza scientifica di discipline come per esempio la sociolinguistica (come affermato anche da Berretta 1983). Secondo la studiosa

francese, queste discipline partono dal presupposto che vi sia una differenza fra donne e uomini e la confermano attestando quella che di fatto, secondo l'autrice, è una rappresentazione sociale. Scrivendo nel 1985, e da psicologa sociale, Aebischer (1985/1988) cita il concetto di rappresentazione sociale così come elaborato in ambito psicologico da Moscovici (1969, 1976), ovvero «una particolare modalità di conoscenza che ha come scopo l'elaborazione dei comportamenti e la comunicazione fra individui» e che ha «una funzione costitutiva della realtà, dell'unica realtà che noi sperimentiamo e nella quale la maggior parte di noi si muove» (Moscovici 1976, 26-27, cit. in Aebischer 1988, 60). Tuttavia Aebischer (1985/1977) non coglie pienamente la complessità di questa definizione concettuale, secondo la quale la costruzione della realtà è sempre mediata da una rappresentazione socialmente condivisa, codificata e decodificata attraverso il pensiero e la lingua, per cui la lingua stessa codifica queste rappresentazioni (Moscovici 1984/1989). Studiare queste rappresentazioni nel giudizio dei parlanti e delle parlanti, senza studiarne le manifestazioni linguistiche, morfologiche, sintattiche e pragmatiche, significa guardare soltanto a un aspetto dell'oggetto indagato.

Aebischer (1985/1977), infatti, a differenza delle colleghe britanniche che, come spiegherò meglio di seguito, dopo aver chiarito che gli stereotipi sul parlato delle donne sono il risultato di generalizzazioni e spiegazioni spesso basate su luoghi comuni, assimilati anche dagli studi linguistici, propongono di «redressing the balance» (Coates e Cameron 1989, 3), e non rinunciano ad assumere la lingua come oggetto di studio, non analizza la lingua parlata, ma gli stereotipi sulla lingua delle donne. In particolare indaga il tratto della chiacchiera, verificandone la condivisione in sottogruppi di donne di età compresa fra i 20 e i 55 anni, e fra un gruppo di studenti universitari. Intervista sessanta donne, in diversa posizione lavorativa e diversamente auto-posizionate rispetto al femminismo (femministe/non femministe), e somministra un questionario a cento studenti (cinquanta maschi e cinquanta femmine). I risultati del questionario attestano la rappresentazione sociale della donna chiacchierona, trasversalmente condivisa da

studenti maschi e femmine. I risultati delle interviste dimostrano, in modo congruente, che questa rappresentazione sociale è ampiamente condivisa dalle donne, anche quando è rifiutata o disprezzata.

In ambito britannico, la presa di coscienza che la ricerca sulla lingua delle donne è essa stessa fondata su stereotipi segna un punto di svolta con la pubblicazione di *Women in their speech community* (Coates e Cameron 1989), che si propone di dare «new perspectives on language and sex», come recita il sottotitolo del volume. Riflettendo sui limiti sia del paradigma quantitativo, focalizzato sulla variazione fonologica e grammaticale, sia del paradigma qualitativo, focalizzato sull'analisi del discorso e l'etnografia del parlato, le autrici propongono dei parametri, ovvero degli standard di ricerca, per evitare di (ri)cadere nell'errore di produrre conoscenze stereotipate. Per quanto riguarda la ricerca quantitativa, la prima raccomandazione è evitare di considerare universali o immutabili i risultati delle ricerche sulle differenze di genere, per quanto possano essere ampiamente diffusi; la seconda è raccogliere dati su parlanti donne e uomini a cui siano applicabili le medesime variabili sociali, piuttosto che partire dalla selezione di un gruppo maschile caratterizzato da specifiche variabili sociali che escludono automaticamente le donne; la terza è fare attenzione a non spiegare i risultati sulla base di stereotipi del senso comune (Coates e Cameron 1989, 11). Per quanto riguarda la ricerca qualitativa, le raccomandazioni sono: spiegare le differenze linguistiche considerando tutti gli aspetti della struttura e delle relazioni di genere nelle diverse sub-culture; incrementare ed estendere a diverse comunità la ricerca empirica per raccogliere informazioni più esaustive e più complete di quelle esistenti; distinguere fra la lingua delle donne parlata in gruppi di sole donne e quella parlata in interazioni miste.

In ambito anglosassone, il progetto di “ristabilire l'equilibrio” diventa produttivo, stimolando una ricerca orientata a indagare oggettivamente le caratteristiche della lingua parlata dalle donne, in una direzione sempre più intersezionale e volta a spiegare le diverse rappresentazioni di genere codificate nella lingua e nelle sue varietà diatopiche, diastratiche e diafasiche (si veda per

esempio Cameron 1998; Holmes e Meyerhoff 2003; Lakoff e Bucholtz 2004; Ehrlich et al. 2017). In Italia, la linea di ricerca privilegiata resta sostanzialmente quella orientata da Berretta (1983) e da Aebischer (1985/1988), la cui traduzione dal francese ne ha forse facilitato la circolazione in Italia, ovvero una linea di ricerca che assume il parlato femminile come stereotipo socialmente condiviso, e lo indaga come tale, senza studiare la lingua come dato oggettivo di realtà. Le uniche due eccezioni sono rappresentate dal lavoro di Anna De Marco (1995) e da una più recente ricerca demoscopica dell'ISTAT (2015). De Marco (1995) indaga l'uso dei diminutivi in parlanti donne e uomini, analizzando il *corpus* LIP (Lessico Italiano Parlato, cfr. De Mauro et al. 1993), poi mette a confronto i dati oggettivi raccolti da questa analisi con gli stereotipi linguistici dei/lle parlanti, che testa tramite dei questionari. I risultati a cui perviene sono: primo, che le donne non usano diminutivi in quantità maggiore rispetto agli uomini, ma in situazioni differenti; secondo, che i parlanti condividono alcuni stereotipi socialmente condivisi sull'uso dei diminutivi. Per esempio, è pregiudizio condiviso che le donne facciano ampio uso di diminutivi quando si rivolgono ai bambini e alle bambine. In realtà, l'analisi del *corpus* LIP dimostra che sia le madri sia i padri ricorrono spesso al cosiddetto *diminutivum puerile*. Le prime però lo fanno in modo costante, mentre i secondi interrompono questo uso in situazioni conflittuali.

Lo studio dell'ISTAT (2015) riporta alcuni dati interessanti sull'uso della lingua italiana, dei dialetti e delle lingue straniere in Italia, in alcuni casi disaggregati per genere. I risultati evidenziano alcune differenze nell'uso del dialetto vs. l'italiano, preferito dalle donne, soprattutto giovani; nella conoscenza delle lingue straniere, prevalente fra le donne, fino ai 44 anni, e fra gli uomini di età superiore ai 45 anni; nelle abilità di ascolto, conversazione, lettura, e scrittura, in cui le donne attestano maggiori competenze rispetto agli uomini. Questa ricerca, tuttavia, avendo un obiettivo puramente descrittivo, non indaga eventuali stereotipi latenti nelle differenze rilevate, né offre elementi sufficienti ad approfondire le statistiche raccolte.

Per il resto, le poche ricerche nazionali focalizzate sulla lingua delle donne partono dall'ipotesi che i tratti tipicamente associati al parlato femminile siano rappresentazioni sociali ampiamente condivise, nel senso di stereotipi lontani dalla lingua effettivamente parlata dalle donne (es. Bazzanella e Fornara 1995; Bazzanella et al. 2006; Messina 2013; Panighel 2014). E anche la riflessione teorica sulla lingua delle donne ne mette in evidenza i caratteri tipicamente stereotipici (Bazzanella 2009; Adamo 2019; Sbisà 2019). Fra le ricerche empiriche, due si focalizzano su contenuti dei media e dimostrano come i prodotti medialti contribuiscono a rafforzare questa rappresentazione della lingua delle donne, veicolandone tratti marcatamente stereotipati. La prima è contenuta nel contributo di Bazzanella et al. (2006) alla collettanea *Linguaggio e genere* (Luraghi e Olita 2006). Questo contributo riporta i risultati di tre ricerche, fra cui un'analisi sul linguaggio di un *corpus* di 100 pubblicità radiofoniche. I risultati attestano la presenza di tratti stereotipicamente associati alla lingua delle donne quali l'uso dei diminutivi, il ricorso ad appellativi come *tesoro* e di esclamazioni come *oh signore* e *oddio*, l'uso di fatismi, focalizzatori e modalizzatori. A questi stereotipi prettamente linguistici, si aggiungono stereotipi narrativi che riguardano i ruoli attribuiti alle donne e agli uomini: le voci femminili hanno un ruolo spesso subalterno, marcato da interruzioni delle voci maschili, a cui è prevalentemente lasciata l'ultima parola e la narrazione fuori campo; gli interventi maschili sono prevalentemente informativi e denotativi, mentre quelli femminili tendono a un maggior coinvolgimento empatico; infine, le donne sono relegate «ad ambiti di acquisto relativi al cibo, alla cura dei bambini e della casa, ai cosmetici» (Bazzanella et al. 2006, 168).

Più di recente, Simona Messina (2013) indaga la lingua parlata dalla donne nella narrazione televisiva italiana, analizzando tre titoli di *fiction* di produzione nazionale trasmesse fra il 1998 e il 2000³, in una prospettiva che, a differenza di Bazzanella et al. (2006), cerca di spiegare i tratti stereotipici individuati su diversi piani linguistici (pianificazione del discorso, lessico, e *idioms*, ovvero proverbi,

³ *Un medico in famiglia* (1998), *Commesse* (1999), *Il bello delle donne* (2000).

metafore, frasi idiomatiche, clichés e frammenti) in relazione a diversi ruoli narrativi interpretati dai personaggi femminili: «la mamma», «la nonna», «la tata», «l'amante buona», «l'amante cattiva», «la donna manager», «la zitella», «la single», «la romantica» (Messina 2013, 271). L'autrice individua questi ruoli narrativi utilizzando come schema di riferimento la mappa sinottica di Eurisko, tradizionalmente utilizzata dall'Auditel per verificare quanto un prodotto televisivo raggiunga un certo *target* di pubblico. La mappa Eurisko distingue, idealmente, il pubblico in sedici gruppi diversi per stili di vita, di cui sette tipicamente femminili, ciascuno rappresentativo di una classe di età, un livello di istruzione, una posizione lavorativa, un livello di reddito, abitudini, consumi, interessi, relazioni: le colleghe, le commesse, le raffinate, le massaie, le appartate 1 («modeste»), le appartate 2 («povere»), le appartate 3 («sole») (Messina 2013, 269). La decisione di profilare i personaggi delle tre *fiction* campione sulla base della mappa Eurisko è motivata dall'autrice a partire dal criterio di verosimiglianza, fondamentale nel decidere il successo di un prodotto mediale. In genere, i personaggi di una serie TV sono profilati così da rappresentare la realtà in maniera verosimile. Fra i tratti che contribuiscono alla creazione della verosimiglianza è compresa la lingua: la lingua parlata dai personaggi femminili per essere verosimile «deve imitare quello che si presume sia il *parlato parlato* femminile» (Messina 2013, 270). Non un parlato unico e indistinto, ma in grado di riprodurre differenze diastratiche, diafasiche e diatopiche che, secondo l'autrice, sono funzionali a una costruzione narrativa di successo.

Sul piano dell'organizzazione del discorso, Messina (2013) osserva alcune caratteristiche comuni a tutti i diversi personaggi femminili, con la sola eccezione della «donna manager». Questi tratti sono l'uso del discorso diretto, di frasi coordinate (stile paratattico), frasi incomplete che tendono a lasciare il discorso interrotto, e caratterizzate da precisazioni tese a riportare molti particolari che a volte rendono il discorso dispersivo, ma che, in realtà, secondo Messina, sono funzionali a mantenere aperto il dialogo. La lingua parlata dalla «donna manager» si caratterizza invece per uno stile ipotattico, informativo, stringato ed essenziale.

Sul piano lessicale, diverse scelte linguistiche risultano funzionali a delineare diversi modelli di donna o caratteristiche specifiche trasversali a diversi modelli. Il «linguaggio infantile» (De Marco 1995) è funzionale a connotare la positività di un personaggio. Il «linguaggio materno», che si contraddistingue per «un'aggettivazione molto vasta», per la presenza ricorrente di reiterazione doppia o multipla di sintagmi aggettivali, nominali e verbali, per l'uso del vocativo affettivo (es. *amore, tesoro*), spesso in funzione fatica, ed è il più diffuso, è tipico dello stereotipo della «mamma», intesa come figura che si prende cura degli altri, anche se non è madre. Allo stereotipo della « donna romantica» è associato un lessico ricco di aggettivi, nomi e verbi che fanno riferimento alla sfera intima e personale. A quello della «donna manager», invece, un lessico scarno, privo di diminutivi, e con poche esclamazioni. Infine, l'«amante cattiva» ha un lessico ricco di parolacce, che contribuisce a connotare il suo stile linguistico come crudo e anche violento, in modo a volte del tutto accessorio alla narrazione.

Sul piano delle espressioni idiomatiche, Messina (2013) ne attesta un uso generalmente ricco, che spiega come un'esigenza delle produzioni seriali rivolte a un pubblico vasto ed eterogeneo (come quello delle tre *fiction* analizzate). Per raggiungere il cosiddetto «grande pubblico», questo genere di produzione ha bisogno di rappresentare una lingua comune di base, facilmente comprensibile a tutti, e di impiegarne le espressioni per caratterizzare i diversi personaggi. A parte le frasi idiomatiche, usate dalle donne soprattutto per descrivere uno stato d'animo, e specialmente negativo, per il resto Messina (2013) non individua specificità femminili.

2.2.2 *Le donne nella lingua*

Se la lingua parlata dalle donne presenta tratti attestati da poche ricerche che riportano risultati diversi e non generalizzabili, poiché basati su dati e metodologie differenti, la lingua che parla delle donne presenta caratteristiche attestate da un più nutrito numero di ricerche, in modo stabile nel corso degli ultimi trent'anni, tanto da poter essere generalmente etichettata come «sessista» (Cavagnoli e

Dragotto 2021). Sebbene il sessismo come scrivono Cavagnoli e Dragotto (2021, 1), «si scrive e riscrive in funzione e come esito degli usi che del termine fa la comunità parlante», e, aggiungo io, la comunità scientifica, mi pare sia necessario assumerne una definizione precisa, quanto meno per orientare la lettura di questa tesi. Fra le varie definizioni di sessismo che ho incontrato nella letteratura su lingua e genere, quella proposta da Cardinaletti e Giusti (1991) mi sembra la più chiara e semplice, e funzionale, fra l'altro, a sottolineare la duplicità, raramente indagata, del sessismo nell'italiano, che si manifesta prevalentemente a livello dell'uso della lingua, ma anche al livello sistemico della grammatica italiana:

dall'inglese *sexism* a sua volta creato in analogia a *racism* ("razzismo"). Con "razzismo" si intende discriminazione secondo la razza, con "sessismo" si intende discriminazione secondo il sesso. Per ciò che riguarda il linguaggio, questo tipo di discriminazione è duplice, apparendo nell'uso della lingua e nel sistema interno alla lingua (Cardinaletti e Giusti 1991, 170).

Riprendendo la classificazione originaria di Alma Sabatini (1987), è possibile riepilogare le forme di sessismo dell'italiano che continuano a essere attestate da ricerche recenti in quattro tipologie principali, qui riepilogate sulla base dei "danni" simbolici che producono.

1. Prevalenza del maschile: contribuisce a connotare l'italiano come una lingua androcentrica⁴, che nasconde le donne, rendendole simbolicamente invisibili.
2. Designazione di donne che ricoprono professioni, cariche o ruoli istituzionali tradizionalmente appannaggio maschile e/o apicali con forme grammaticali maschili o costruzioni lessicali che fanno riferimento agli uomini: rafforzano gli stereotipi che associano la sfera pubblica al genere maschile (vs. la sfera privata, associata al genere femminile) e rafforzano i rapporti di potere fra

⁴ Per precisione, secondo Sabatini «La lingua italiana, come molte altre, è basata su un principio androcentrico: l'uomo è il parametro, attorno a cui ruota e si organizza l'universo linguistico» (Sabatini 1987, 24), per cui non solo la prevalenza del maschile, ma anche tutte le altre forme di sessismo contribuiscono a connotare l'italiano come una lingua androcentrica.

donne e uomini, riconoscendo maggiore autorevolezza e prestigio ai nomi maschili e negando, al contempo, l'*empowerment* femminile in professioni e ruoli politici o di potere apicali, o storicamente appannaggio maschile.

3. Asimmetrie di genere: contribuiscono a rafforzare gli stereotipi sociali e i rapporti gerarchici fra donne e uomini.
4. Forme lessicali alterate e figure retoriche: sminuiscono le donne.

Di seguito presento le diverse forme di sessismo linguistico articolate secondo queste tipologie, riportando per ogni forma un esempio, raccolto nella primavera del 2020, su organi di informazione online⁵.

Anticipo brevemente una spiegazione importante e relativa alla scelta di usare il termine *sovraesteso*, come aggettivo per connotare diversi modi di usare forme maschili con riferimento a gruppi misti di donne e uomini o a qualche referente generica/o di cui non si conosce il genere (tipicamente in espressioni ricorrenti in moduli amministrativi parzialmente pre-compilati, come, per esempio *Il sottoscritto...*). Il termine *sovraesteso* è ripreso da Bazzanella (2010), che utilizza il verbo *sovraestendere* per indicare l'uso di quello che in letteratura è meglio noto come *maschile non marcato* (Sabatini 1987, 39) o *maschile generico* (Marcato e Thüne 2002, 201). Entrambe queste forme mi pare che possano offrire una sponda, non tanto nella letteratura linguistica, quanto nella divulgazione, a un'interpretazione in senso neutro dei termini *generico* e *non marcato*. L'aggettivo *sovraesteso*, che non a caso è utilizzato in opere di divulgazione come *Femminismi singolari* (Gheno 2019, 47), probabilmente proprio per non dare adito a fraintendimenti, mi pare esprima in modo più efficace la pratica di usare una parola maschile estendendone il riferimento ("sopra") alle donne.

⁵ Gli esempi sono stati raccolti per la stesura di un articolo divulgativo che ho scritto per l'associazione UISP (Unione Italiana Sport Per tutti), finalizzato a introdurre il mondo dello sport a un uso paritario e inclusivo della lingua italiana. L'articolo contiene anche una breve guida alla formazione dei femminili, e un elenco esemplificativo di nomi di professioni e ruoli sportivi nelle forme femminili e maschili corrispondenti fra loro (Azzalini 2021b).

1. Prevalenza del maschile

1.1. Uso sovraesteso di *uomo/uomini*, con riferimento alla specie umana, che comprende anche le donne, le quali risultano però oscurate da un nome caratterizzato da ambiguità semantica, potendo anche riferirsi a una persona o più persone esclusivamente di sesso maschile, come esemplificato in (1).

(1) *Coronavirus, negli Usa al via test sull'uomo per un vaccino*⁶.

1.2. Uso sovraesteso del maschile di nomi che si riferiscono genericamente a gruppi sociali notoriamente misti, formati da donne e uomini (maschile generico o non marcato), come esemplificato in (2).

(2) *Coronavirus, in Germania solo cittadini UE e residenti*⁷.

1.3. Uso di nomi con genere grammaticale e semantico maschile, come *fratelli*, *fratellanza*, per indicare relazioni umane generiche, non specificamente maschili, come esemplificato in (3).

(3) *Graziani: «In questo momento siamo tutti fratelli. Playoff sarebbero stuzzicanti»*⁸.

1.4. Assorbimento del femminile nel maschile, nella formazione dei nomi plurali che si riferiscono a gruppi specifici misti di donne e uomini (maschile generico o non marcato), come esemplificato in (4).

(4) *Coronavirus, ecco tutte le donazioni organizzate dai vip. Da Luciana Littizzetto a Giorgia Palmas, è gara di solidarietà*⁹.

1.5. Precedenza del maschile nelle coppie oppositive uomo/donna, come esemplificato in (5)¹⁰.

⁶

Fonte:

<https://www.repubblica.it/salute/2020/03/1.6/news/coronavirus-negli-al-via-test-sull-uomo-per-un-vaccino-251455613/>; ultimo accesso 19 novembre 2021.

⁷ Fonte: https://www.adnkronos.com/fatti/esteri/2020/03/17/coronavirus-germania-solo-cittadini-residenti_VRvgX7T7Mohe5kxW5uWESL.html; ultimo accesso 19 novembre 2021.

⁸ Fonte: <https://www.fiorentina.it/graziani-momento-tutti-fratelli-playo-off-la-serie-stuzzicante/>; ultimo accesso 19 novembre 2021.

⁹ Fonte: https://www.tgcom24.mediaset.it/people/coronavirus-ecco-tutte-le-donazioni-organizzate-dai-vip_15984805-202002a.shtml; ultimo accesso 19 novembre 2021.

¹⁰ La precedenza del maschile nelle coppie oppositive uomo/donna è definita in ambito anglosassone come *male firstness* (Baker 2014, 92-95).

- (5) *Le decisioni sulle date delle altre competizioni UEFA, siano esse club o nazionali per uomini o donne, saranno prese e annunciate a tempo debito*¹¹.
- 1.6. Uso del nome *donne* per indicare una categoria a parte, come esemplificato in (6).
- (6) *gli elettori registrati sono oltre 125.000...neri, ispanici, donne, si sono registrati in massa*¹².
2. Designazione delle donne con forme grammaticali maschili o costruzioni lessicali che fanno riferimento agli uomini.
- 2.1. Uso della forma maschile anche laddove esiste o può essere formata la forma femminile, come esemplificato in (7).
- (7) *Bimbi, ginnastica e Star Wars: Lady Isabelle, medico in famiglia per Matuidi*¹³.
- 2.2. Uso del modificatore *donna*, prima o dopo il nome maschile, e conseguenti possibili discordanze grammaticali, come in (8) e (9).
- (8) *Calcio, follia in campo: arbitro donna 16enne aggredita*¹⁴.
- (9) *Arbitro donna aggredita nel 2014. L'Aia le ritira la tessera*¹⁵.
- 2.3. Uso del suffisso *-essa* originariamente utilizzato per designare le mogli di nobili o titolari di cariche pubbliche, applicato a nomi epiceni/bigenere come *presidente*, che hanno la stessa forma al maschile e al femminile e possono essere usati in modo invariato nel riferimento femminile, come esemplificato in (10).

¹¹ Fonte: <https://it.uefa.com/insideuefa/about-uefa/news/newsid=2641075.html>; ultimo accesso 19 novembre 2021.

¹² Questo esempio è tratto da Sabatini (1987, 49).

¹³ Fonte: https://www.gazzetta.it/Calcio/Serie-A/18-03-2020/bimbi-ginnastica-star-wars-lady-isabelle-medico-famiglia-matuidi-3601801753603_preview.shtml?reason=unauthenticated&origin=http%3A%2F%2Fwww.gazzetta.it%2FCalcio%2FSerie-A%2F18-03-2020%2Fbimbi-ginnastica-star-wars-lady-isabelle-medico-famiglia-matuidi-3601801753603.shtml; ultimo accesso 19 novembre 2021.

¹⁴ Fonte: <https://www.italiasportpress.it/calcio/calcio-follia-in-campo-arbitro-donna-16enne-aggredita-da-allenatore-e-tifosi/>; ultimo accesso 19 novembre 2021.

¹⁵ Fonte: <https://www.gazzetta.it/Calcio/altri-campionati/11-11-2019/arbitro-donna-aggredita-2014-aia-ritira-tessera-3501011579994.shtml>; ultimo accesso 19 novembre 2021.

(10) *Lo Spoleto cambia proprietà: la presidentessa è una principessa araba!*¹⁶.

2.4. Accordo di nomi epiceni/bigenere riferiti a donna con elementi satellite declinati al maschile, come esemplificato in (11).

(11) *La facciata di Palazzo Madama, per decisione del Presidente del Senato Elisabetta Casellati «È una scelta simbolica ma dettata dalla volontà concreta di dare a tutti gli italiani un segnale di fiducia e di vicinanza delle Istituzioni. In questa drammatica fase per il Paese» – dichiara il Presidente Casellati*¹⁷.

3. Asimmetrie di genere

3.1. Uso di prenomi, cognomi, titoli e appellativi diversi per donne e uomini, come in (12).

(12) *Lady Rugani svela: «Sono incinta ma ora ho paura per via del coronavirus». La compagna del difensore della Juventus Daniele Rugani ha svelato a Chi di essere entrata nel quarto mese di gravidanza*¹⁸.

3.2. Uso di aggettivi diversi per donne e uomini, come in (13).

(13) *Ronaldo ha postato su Instagram una foto con la sua bella fidanzata. Insieme hanno festeggiato il compleanno. Ronaldo, il Fenomeno, è stato sicuramente uno dei giocatori più forti della storia*¹⁹.

3.3. Uso di sostantivi uguali ma con connotazione diversa al femminile e al maschile, come esemplificato in (14), dove *il segretario* al maschile è usato per designare la segretaria della CISL Anna Maria Furlan, che ricopre un ruolo apicale nell'organizzazione sindacale, vs. in (15), dove *la segretaria*

¹⁶ Fonte: <https://www.gazzetta.it/Calcio/altri-campionati/02-07-2019/spoleto-cambia-proprietà-presidentessa-principessa-araba-340548019176.shtml>; ultimo accesso 19 novembre 2021.

¹⁷ Fonte: https://www.repubblica.it/politica/2020/03/18/foto/coronavirus_senato_governo_palazzo_chigi_tricolore-251641072/1/; ultimo accesso 19 novembre 2021.

¹⁸ Fonte: <https://www.ilgiornale.it/news/sport/lady-rugani-svela-sono-incinta-ora-ho-paura-coronavirus-1842727.html>; ultimo accesso 19 novembre 2021.

¹⁹ Fonte: <https://www.ilgiornale.it/news/spettacoli/ronaldo-innamorato-celina-locks-ti-auguro-meglio-1836091.html>; ultimo accesso 19 novembre 2021.

al femminile è usato per designare la collaboratrice del segretario di PD Nicola Zingaretti, che svolge una mansione impiegatizia.

(14) *«La nostra priorità oggi è garantire la sicurezza e la salute delle persone in tutti i luoghi di lavoro».* Lo afferma dalle colonne di *Interris.it* il segretario generale della CISL, Anna Maria Furlan²⁰.

(15) *Coronavirus, contagiata la segretaria di Zingaretti [...] La collaboratrice che lavora al Partito Democratico positiva al tampone*²¹.

3.4. Uso di forme di identificazione delle donne per ruolo familiare, vs. identificazione degli uomini per professione, come esemplificato in (16).

(16) *A cui si aggiunge Cristina Parodi, moglie di Giorgio Gori, sindaco di Bergamo*²².

4. Forme lessicali alterate e figure retoriche

4.1. Uso di forme alterate, in particolare diminutivi e vezzeggiativi, per descrivere le donne, come in (17).

(17) *Tra le gigantesse del Grande Nord (altro bacino d'utenza, altro vivaio, altri budget) c'è la puffetta azzurra, che abbraccia le rivali al traguardo con sportività*²³.

4.2. Uso di metafore, metonimie, eufemismi e stilemi stereotipati, come in (18).

(18) *Dorothea, perle blu ai lobi e maschera da combattimento. Solo il sorriso, alla fine, è un po' meno smagliante [...] l'azzurra è la lepre*

²⁰ Fonte: <https://www.interris.it/news/coronavirus-furlan-cisl-su-sicurezza-lavoratori-se-necessario-sospendere-la-produzione/>; ultimo accesso 19 novembre 2021.

²¹ Fonte: https://www.corriere.it/cronache/20_marzo_08/coronavirus-contagiata-segretaria-zingaretti-negativa-giunta-regionale-40f28f9a-612d-11ea-8f33-90c941af0f23.shtml; ultimo accesso 19 novembre 2021.

²² Fonte: https://www.repubblica.it/politica/2021/03/18/news/giornata_nazionale_vittime_covid-292743406/; ultimo accesso 19 novembre 2021.

²³ Fonte: https://www.corriere.it/sport/20_febbraio_23/biathlon-dorothea-wierer-vittoria-argento-mass-start-quarta-medaglia-mondiali-ddf011a2-5636-11ea-b447-d9646dbdb12a.shtml; ultimo accesso 19 novembre 2021.

che corre per non farsi riprendere. [...] È un corpo a corpo tra donne Alfa [...]»²⁴.

Quasi tutte le forme linguistiche che discriminano le donne sono riconducibili a usi linguistici facilmente modificabili in modo da rendere visibili le donne, e al contempo, come vedremo nel paragrafo successivo, rispettare, o addirittura rafforzare la coerenza interna al sistema linguistico italiano. Più problematici risultano invece il maschile generico, o non marcato, utilizzato, al plurale, per indicare gruppi misti di donne e uomini e, al singolare, con riferimento a persone di cui non si conosce il genere specifico oppure a donne che ricoprono un ruolo apicale o storicamente appannaggio maschile; e la declinazione al plurale di tutti gli elementi satellite che si riferiscono a un sintagma nominale composto da più nomi (es. *Le ragazze e i ragazzi sono andati al mare*). Il privilegio accordato al maschile plurale sia in termini di scelta lessicale sia in termini di accordo sintattico rappresenta una norma storicamente acquisita nella lingua italiana a livello di sistema (cfr. Cardinaletti e Giusti 1993; Thornton 2009). L'uso del maschile generico al singolare si basa invece su un fenomeno culturale che riconosce al maschile un maggior prestigio (Giusti 2022), in violazione al sistema della lingua italiana, come vedremo meglio nel paragrafo successivo.

Se per quanto riguarda l'uso del maschile non marcato al singolare è possibile pensare a un cambiamento, quando la referente è chiaramente una donna, utilizzando forme femminili esistenti oppure creandone di nuove attraverso la mozione (Thornton 2004), rafforzando il prestigio delle forme femminili (Giusti 2022), l'uso sovraesteso del maschile sia al singolare, quando il genere semantico referente non è noto, sia al plurale, con referente noto o non noto, rimane problematico. Come osserva Deborah Cameron, riflettendo sull'uso sovraesteso di *men*, che, in inglese, come in italiano, può essere usato per indicare sia gli esseri umani in generale, sia gli uomini come genere specifico (il maschile) della specie umana, in genere il contesto d'uso ci permette di disambiguare il valore di un termine, in senso sovraesteso o specifico: «If I see a banner saying "All men are

²⁴ *Ibidem*.

equal” I know I can join the demonstration, if I see the word MEN on the entrance to a public lavatory, I know I am not supposed to enter» (Cameron 1998, 10). Tuttavia non è privo di problematicità, perché, almeno simbolicamente, oscura la presenza femminile, e, in alcuni casi, come vedremo nello studio del lemma *ministra/o*, attesta un’elevata frequenza (cfr. Capitolo 6).

2.3 Il genere fra grammatica e cultura²⁵

2.3.1 Il genere come categoria grammaticale e semantica

Il genere come categoria grammaticale è una caratteristica morfo-sintattica comune a molte lingue del mondo, anche se non a tutte e non allo stesso modo (Corbett 1991). Ci sono lingue come il francese, lo spagnolo e il tedesco con il genere marcato (*gender-marked*), vale a dire con una struttura tale per cui il genere grammaticale si manifesta, a livello morfologico e/o sintattico, su nomi e pronomi, e regola l’accordo con gli elementi satellite (articoli, aggettivi, participi passati). Ci sono invece lingue, come l’inglese e il danese, con genere naturale (*natural-gender*), che esprimono il genere quasi solo sui pronomi e su pochi nomi. Infine ci sono lingue, come il finlandese (Hellinger e Bußmann 2002), il turco (Hellinger e Bußmann 2001) o l’ungherese (Hellinger e Montschenkacher 2004), prive di genere grammaticale.

Le lingue dal genere marcato possono avere due o più generi grammaticali²⁶, i cui criteri di assegnazione variano a seconda delle lingue (Hockett 1958; Corbett 1991; Hellinger e Bußmann 2001-2003; Hellinger e Montschenkacher 2004). In ogni caso, i generi grammaticali possono considerarsi classi nominali che si riflettono sul comportamento delle parole associate a un nome (Hockett 1958, 231)²⁷. Corbett (1991), prendendo in considerazione le

²⁵ Questo paragrafo riprende parzialmente Azzalini e Giusti (2019).

²⁶ Oltre al genere femminile, maschile e neutro, ci possono essere altri generi tipici di particolari classi di nomi, a seconda del grado di specificità di genere di una lingua; lo Swahili, per esempio, conta 15 diversi generi grammaticali (cfr. Corbett 1991; Hellinger e Bußmann 2003).

²⁷ «Genders are classes of nouns reflected in the behaviour of associated words» (Hockett 1958, 231).

caratteristiche di oltre duecento lingue al mondo, e analizzando l'interrelazione fra il genere grammaticale e le categorie nominali animato/inanimato e umano/non umano, distingue due modalità principali di assegnazione del genere: semantica e formale. Si tratta di una distinzione utile a ordinare l'immensa varietà delle duecento lingue considerate da Corbett (1991) che, per quanto presentino caratteristiche specifiche, possono essere ricondotte, in modo più o meno stretto, all'una o all'altra modalità di assegnazione del genere grammaticale.

I sistemi semantici basano l'attribuzione del genere grammaticale sul significato del nome. Per esempio nella lingua Tamil, i nomi con riferimento umano maschile hanno un genere grammaticale maschile, i nomi con riferimento umano femminile hanno un genere grammaticale femminile, i nomi con riferimento non umano hanno un genere grammaticale neutro. I sistemi formali sono sistemi morfologici, in cui l'assegnazione del genere grammaticale di un nome si basa "anche" sulla morfologia del nome. "Anche" nel senso che non ci sono sistemi morfologici puri, ma solo sistemi morfologici che oltre all'assegnazione del genere su base semantica presentano un'assegnazione del genere su base morfologica. Corbett (1991) riporta l'esempio dei nomi russi *šotlandec* e *šotlandka* (scozzese): il genere maschile del primo è esplicitato dal suffisso *-ec*, indicativo del genere grammaticale maschile; il genere femminile del secondo è esplicitato dal suffisso *-ka*, che in russo è indicativo del genere grammaticale femminile. Tuttavia, in entrambi i casi, la relazione fra il genere grammaticale e il genere semantico non è arbitraria, bensì coerente: il suffisso maschile *-ec* rende trasparente il genere grammaticale maschile di un nome che si riferisce a un uomo, nella fattispecie di origine scozzese, il suffisso femminile *-ka* rende trasparente il genere grammaticale femminile di un nome che si riferisce a una donna (scozzese).

L'italiano è una lingua dal genere marcato a livello formale, come il latino e tutte le altre lingue romanze derivate dal latino. A differenza del latino, da cui deriva, l'italiano non ha però un genere neutro. Tutti i nomi hanno un genere femminile o maschile, che può essere più o meno trasparente a livello morfologico, ed è sempre esplicitato dagli accordi sintattici. Il genere grammaticale dei nomi che

si riferiscono a entità inanimate o astratte non dipende da proprietà semantiche ma morfologiche, ed è arbitrario, tanto che possiamo indicare un medesimo corpo celeste che brilla di luce propria usando la parola femminile *stella* o la parola maschile *astro*, perfettamente intercambiabili fra loro. I nomi che si riferiscono a esseri umani, invece, generalmente hanno una relazione con il genere biologico e sociale del/la referente. Per esempio, *papà* indica un uomo nel suo ruolo di genitore, *mamma* indica una donna nello stesso ruolo; *il fidanzato* può far riferimento a un ragazzo o a un uomo, ma non a una ragazza, o a una donna, per la quale useremo più propriamente *la fidanzata*.

A seconda della classe nominale, tutti i nomi presentano marcatori grammaticali a livello morfologico e/o sintattico. Per quanto riguarda i nomi che si riferiscono agli esseri umani, una prima classe comprende nomi il cui genere grammaticale è espresso a livello di radice e il cui genere semantico è radicato nel significato, vale a dire che è imprescindibile dal significato, come per esempio *donna* (Thornton 2004). La maggior parte di questi nomi sono termini di parentela usati per indicare identità e ruoli femminili e maschili in modo simmetrico, per esempio: *madre* e *padre*, *mamma* e *papà*, *sorella* e *fratello*, *nuora* e *genero*.

Un secondo gruppo comprende le classi dei sostantivi con desinenza flessiva: nomi con desinenza femminile in *-a/-e* (singolare/plurale) e desinenza maschile in *-o/-i*, per esempio *maestr-a/-e* e *maestr-o/-i*; nomi con desinenza femminile in *-a/-e* e maschile in *-e/-i*, per esempio *infermier-a/-e* e *infermier-e/-i*; nomi che al femminile terminano in *-trice/-trici*, per esempio *diret-trice/-trici* e al maschile in *-tore/-tori*, ad es. *diret-tore/-tori*; nomi che al femminile terminano in *-ora/-ore*, ad es. *past-ora/-ore* e al maschile in *-ore/-ori*, ad es. *past-ore/-ori*.

Ci sono poi due classi di nomi per i quali il genere grammaticale non è espresso dalla radice o dalla morfologia di flessione, ma dalla declinazione degli elementi satellite: la classe dei nomi epiceni (detti anche bigenere o ambigenere, perché possono riferirsi sia a donne sia a uomini) con la stessa desinenza per i due generi al singolare e al plurale, ad es. *il/la cantant-e* *i/le cantant-i*; e la classe dei

nomi semi-epiceni (Formato 2019, 48-49) con la stessa desinenza al singolare ma marcatura di genere al plurale, per esempio *il/la regist-a, i regist-i, le regist-e*.

Nei nomi appartenenti alle classi flessive il tratto semantico espresso dalla desinenza è correlato all'identità di genere del/la referente: *maestra* si riferisce a un'insegnante, *maestro* a un insegnante. Nei nomi epiceni o semi-epiceni il tratto semantico di genere è opaco o parzialmente opaco, quindi è impossibile stabilire il genere semantico di parole epicene al singolare e plurale, e semi-epicene, al singolare, estrapolate da una frase; tuttavia, le parole isolate ricorrono probabilmente soltanto nei dizionari, nella maggior parte dei casi le parole che usiamo sono collocate in una frase, composta da diversi elementi grammaticali che, se accordati in modo congruente al genere semantico, ci permettono di stabilire il genere grammaticale e semantico del nome.

Infine, c'è un gruppo ristretto di nomi di genere promiscuo: si tratta di nomi che hanno un genere grammaticale fisso, maschile o femminile, che rimane invariato e indipendente dal genere del/la referente. Sono nomi il cui tratto semantico non specifica il genere sociale del/la referente, che può essere femminile o maschile, a differenza per esempio di parole come *madre* o *padre* la cui caratteristica di genere è ineliminabile dal significato (Thornton 2004), o nomi di professione o ruolo, il cui genere grammaticale ha, o meglio dovrebbe avere, una funzione identitaria (Giusti 2011b). Sono nomi che hanno un significato generico, come per esempio *individuo* e *persona*; nomi di azioni estese metonimicamente all'agente, come *guardia*, *guida*, *sentinella*; nomi di timbri vocali estesi a chi canta, come *contralto* e *soprano*; nomi che hanno origine da una metonimia, come *braccio destro*, *capo*, *cervello*, *mente*, *membro*, *pupilla*, *vittima*.

Fatta eccezione per quest'ultimo gruppo, il cui genere semantico può essere disambiguato soltanto dal contesto d'uso, la lingua italiana ha una struttura morfo-sintattica funzionale, in generale, a rappresentare il genere femminile e maschile²⁸,

²⁸ Non sfugge a chi scrive la recentissima questione relativa alla proposta di introdurre nella lingua italiana una categoria grammaticale di genere non binario, tramite l'uso della *schwa*, per dare visibilità a identità *gender-fluid*, e superare i problemi posti dall'uso del maschile generico o non

in particolare a rendere visibili le donne e nominarle in modo simmetrico rispetto agli uomini, anche nel caso di ruoli storicamente maschili. Tuttavia l'ingresso delle donne italiane nella vita politica e nel mercato del lavoro, e soprattutto la loro ascesa a posizioni apicali, non è stata accompagnata da un uso della lingua paritario. Se è diventato comune e frequente usare la parola *maestra*, simmetrica a *maestro*, non è diventato altrettanto comune usare la parola *ministra* per nominare le ministre in modo simmetrico rispetto ai colleghi maschi (Formato 2016; Azzalini 2021a). Per nominare le donne che occupano posizioni apicali è frequente l'uso di forme maschili (come *ministro* per una ministra) o forme composte in cui il maschile è preceduto o seguito dal modificatore *donna*, come *donna avvocato* o *avvocato donna*, o di nomi derivati dal maschile per suffissazione in *-essa* (Marcato e Thüne 2002). Le forme di femminilizzazione con il suffisso in *-essa* hanno una doppia origine, indicando, nel primo caso, mogli di titolari di una carica, ad esempio la *contessa* per designare la moglie del *conte*; nel secondo, donne esercitanti un'attività, come per esempio *dottoressa*, *professoressa* o *studentessa* (Cortellazzo 1995). Sabatini (1986, 1987) sconsiglia di formare nomi femminili con l'aggiunta del suffisso *-essa*, anche per via di un loro uso talvolta spregiativo e ironico. Tuttavia alcune forme in *-essa*, come quelle appena citate, sono ormai la norma prevalente e hanno perso l'originaria connotazione sessista (cfr. Cortellazzo 1995; Robustelli 2014; Giusti 2015).

I suffissati in *-essa*, così come le forme maschili e le forme composte da un nome maschile e dal modificatore *donna*, deviano però dal sistema grammaticale italiano e allo stesso tempo confermano simbolicamente i rapporti gerarchici tra i generi, perché vengono usati solo per nominare donne in lavori o ruoli storicamente maschili, soprattutto in posizioni di vertice. Nominare una donna insegnante con la parola *maestra* e non nominare una donna membro del governo

marcato, con un simbolo, e un fonema, né maschile né femminile (*a* al singolare e *ɜ* al plurale) che abbia funzione paritaria e inclusiva di tutte le identità di genere (<https://italianoinclusivo.it>). Tuttavia, la questione non verrà qui presa in considerazione sia perché, come specificato nell'introduzione, questa tesi si concentra sulle rappresentazioni di donne e uomini, sia perché l'uso della *schwa* è attualmente ancora sperimentale (<https://www.micromega.net/vera-gheno-intervista-schwa/>).

ministra, ma *ministro*, significa non rendere visibili con le parole le identità emergenti delle donne che ricoprono cariche pubbliche o svolgono lavori prestigiosi, rafforzando indirettamente l'idea che le identità politiche, istituzionali o professionali più autorevoli siano in qualche modo una prerogativa maschile. Inoltre, le oscillazioni tra diverse forme di femminilizzazione provocano identità incerte e diminuiscono, o addirittura ribaltano, la connotazione di *empowerment* che hanno i nomi di ruolo apicale o di prestigio (Giusti 2011b; Cavagnoli 2013). Dire o scrivere *donna* prima o dopo *avvocato*, conservando la declinazione maschile del titolo professionale, implica in qualche modo rafforzare l'idea che la professione legale sia una prerogativa maschile, e il fatto che sia svolta da una donna rappresenti una sorta di eccezione (Azzalini e Giusti 2019). La stessa implicazione si riscontra anche per la resistenza attestata da Sabatini (1987) e, più di recente, da Cavagnoli (2013) e Robustelli (2014), ad accordare al femminile gli elementi satellite di nomi di cariche apicali come *presidente* o *dirigente*.

2.3.2 *La resistenza culturale all'uso dei femminili*

La designazione delle donne che ricoprono certe professioni o ruoli pubblici con forme maschili devianti dal sistema grammaticale italiano va attribuita a una resistenza culturale a riconoscere l'avanzamento delle donne in posizioni di potere tradizionalmente detenute dagli uomini (Cardinaletti e Giusti 1991; Cavagnoli 2013; Giusti 2011b; 2016a). Questa resistenza può essere spiegata dalla concomitanza di una serie di fattori che si rafforzano a vicenda, tra cui la mancanza di metacompetenza linguistica, il rifiuto di alcune donne influenti di utilizzare forme femminili e, infine, la mancanza di una politica linguistica condivisa a livello nazionale (Azzalini e Giusti 2019).

La metacompetenza linguistica è una competenza che va oltre la capacità di parlare una o più lingue, è la conoscenza della natura cognitiva della lingua, del suo ruolo nella costruzione delle identità, comprese quelle di genere, così come del sistema grammaticale, altrimenti appreso senza istruzioni esplicite per via della semplice esposizione ai dati linguistici (Giusti 2011b; Giusti 2016b). Impariamo a

parlare la nostra lingua madre, in modo spontaneo e inconsapevole, fin dalla nascita, attraverso un processo che raggiunge la maturità in condizioni non patologiche entro i 4-5 anni di età e continua per tutta la vita (Jackendoff 1998). Fin dall'infanzia, la lingua gioca un ruolo importante nella costruzione delle nostre identità. Parlando una lingua, ci identifichiamo con una comunità di parlanti, che a sua volta ci identifica come parte (o non parte) di essa. La lingua, con le sue parole, le sue strutture morfologiche e sintattiche, compresa la declinazione e l'accordo di genere, definisce l'ontologia culturale su cui si basa la nostra identità, detto altrimenti riflette valori e idee della società in cui viviamo (Giusti 2011b). In tal senso, l'uso di forme maschili per nominare donne in professioni o ruoli autorevoli è il risultato di un atteggiamento culturale che assorbe le rappresentazioni sociali, inclusi gli stereotipi di genere sedimentati a lungo tempo nella nostra società e al contempo li ritrasmette identificando il genere maschile come il più prestigioso (Giusti 2011b). Non sorprende che alcuni studi dimostrino che le forme femminili sono rifiutate dalle stesse donne perché percepite come meno prestigiose di quelle maschili (Merkel 2016). Questo processo cognitivo è testimoniato ad esempio dall'uso di *segretaria/o*. Come esemplificato nel sotto-paragrafo precedente (§ 2.3.1), in (14) e (15), la forma femminile è comunemente usata per indicare una persona impiegata in compiti amministrativi ordinari, mentre la forma maschile è solitamente usata per indicare una persona in una posizione apicale, anche se questa persona è una donna, come per esempio la segretaria di un'organizzazione sindacale (Nardone 2016, 2018).

Molti esperimenti psicolinguistici e neurolinguistici dimostrano che i nomi con riferimento umano attivano stereotipi di genere, a livello lessicale, morfologico e semantico, e questi stereotipi vincolano l'accordo con gli elementi satellite della frase (Banaji e Hardin 1996; Garnham et al. 2002; Kennison e Trofe 2003; Sturt 2003; Oakill et al. 2005). Per esempio, in inglese il nome *doctor* viene stereotipicamente associato al genere maschile, quindi una frase come «the doctor prepared *herself* for the operation» (Osterhout et al. 1997, 274) viene percepita come strana e genera un'anomalia linguistica che può essere misurata tramite

l'elettroencefalografia, dagli ERPs (Event Related Potential), ovvero differenze del potenziale elettrico cerebrale positive o negative sensibili a stimoli linguistici (Bambini 2017). Gli studi che indagano questi meccanismi mostrano che l'attivazione degli stereotipi è sensibile sia al discorso linguistico (il testo) sia al contesto extralinguistico (la conoscenza del mondo in base alla quale viene interpretato il significato di un nome) e, nelle lingue dal genere marcato, come l'italiano o il francese, l'attivazione degli stereotipi è sensibile anche alla morfologia delle parole (Duffy e Keir 2004; Kreiner et al. 2008). Alcuni esperimenti dimostrano che le forme femminili possono inibire l'attivazione degli stereotipi di genere (Richy e Burnett 2021) e anche mitigare i pregiudizi di genere sulle aspettative di carriera e sulle competenze professionali di donne e uomini (Carreiras et al. 1996; Irmen e Roßberg 2004; Vigliocco et al. 2005; Irmen e Schuman 2011; Esaulova et al. 2013). Questi risultati sono coerenti con i risultati di studi recenti che dimostrano come l'attivazione di stereotipi di genere per i nomi professionali sia inibita in italiano da segnali morfosintattici che esplicitano il genere del riferimento semantico (Cacciari e Padovani 2007; Cacciari et al. 2011; Caffarra et al. 2015; Ronca e Moscati 2019).

2.3.3 *Il ruolo delle politiche linguistiche e dei mezzi di informazione*

In questo contesto, le politiche linguistiche possono giocare un ruolo cruciale. Alcuni paesi hanno infatti adottato con successo politiche linguistiche *gender-fair*. In ambito francofono, il Canada è stato il primo paese ad adottare un documento ufficiale per un uso della lingua francese non sessista, pubblicato nella *Gazette officielle du Québec* nel 1979 (Vachon L'Heureux 1992; Gomez Gane 2017). Il documento comprende indicazioni sulla femminilizzazione degli agentivi, anche attraverso eventuali neologismi da formarsi secondo la morfologia della lingua francese, che come l'italiano è una lingua marcata a livello morfo-sintattico, e successivamente confluisce in una guida dell'*Office de la langue française* (Biron 1991), contribuendo a diffondere ampiamente l'uso di femminili di nomi di professioni e incarichi politici nel Québec (Gomez Gane 2017). Negli anni più recenti, le istituzioni canadesi si sono orientate verso una *rédaction épïcène*, che non va confusa con un uso della lingua ridotto alle forme epicene, bensì come un

uso che segua principi di scelta delle parole e di costruzione del testo funzionali a rappresentare in modo equilibrato donne e uomini (Vachon L'Heureux e Guénette 2006; Vachon L'Heureux 2007). La *rédaction épïcène* è peraltro ripresa da una guida al miglioramento della rappresentazione delle donne nei media dei paesi del Maghreb francofono pubblicata dall'UNESCO e destinata ai professionisti e alle professioniste dell'informazione, che ho coautorato (Azzalini e Malchiodi 2013).

In Francia il primo intervento normativo viene intrapreso nel 1984 da Yvette Roudy, ministra del Governo Fabius, con l'istituzione di una *Commission de terminologie chargée de la féminisation des noms de métiers et des fonctions*, seguita due anni dopo da una *Circulaire* (1986, cit. in Gomez Gane 2017, 75) per la femminilizzazione dei nomi nei documenti ufficiali della pubblica amministrazione e dell'assemblea nazionale francese (Houdebine 1987; Houdebine-Gravaud 1998; Burr 2003; Viennot 2014; Gomez Gane 2017). Queste prime iniziative non hanno successo, per cui una decina d'anni dopo vengono riproposte dal Governo Jospin (giugno 1997-maggio 2002), che emana una *Circulaire* (1998, cit. in Gomez Gane 2017, 75) e un *Rapport* della *Commission générale de terminologie et de néologie* (1998, cit. in Gomez Gane 2017, 75) per introdurre l'uso dei femminili nel linguaggio parlamentare e dell'amministrazione pubblica, e promuove la redazione della *Guide d'aide à la féminisation des noms de métiers, titres, grades et fonctions* (Bequer et al. 1999), un volume completo di un'introduzione a firma del primo ministro Jospin, che conferisce istituzionalità al lavoro, una ricostruzione storica della femminilizzazione dei nomi nella lingua francese nel corso dei secoli, un'ampia sezione dedicata alle regole per la femminilizzazione, completa di una nomenclatura di base, e una parte più di carattere sociolinguistico, che argomenta obiezioni e difficoltà nell'uso di forme femminili. Questa seconda serie di iniziative ottiene un maggior successo rispetto a quella intrapresa dal Governo Fabius (luglio 1984-marzo 1986), contribuendo a diffondere in Francia l'uso dei nomi femminili di professioni e ruoli politici e istituzionali. In particolare Burnett e Bonami (2018) attestano un incremento, ancorché non uniforme, di femminili quali *deputée*, *présidente*, *(la) secrétaire d'état*, *(la) ministre*, *(la) garde des sceaux*, *rapporteuse*,

nei dibattiti dell'assemblea parlamentare francese, a partire dal 1997: per esempio, l'uso dell'espressione *Madame la ministre* da percentuali sotto la soglia del 10% nel periodo 1982-1997, dopo il 1998 (e fino al 2005) raggiunge valori appena sotto la soglia del 100% (Burnett e Bonami 2018, figura 1)²⁹. L'autrice e l'autore di questa ricerca sostengono che, in Francia, le politiche linguistiche del Governo Jospin hanno avuto maggior successo rispetto alle prime iniziative, promosse dal Governo Fabius, per via di un mutato contesto sociale e politico, in particolare per l'emergere di un nuovo «social type or *persona* (Zhang 2005; Podesva 2007; Eckert 2008; among others) for female politicians, which only feminine g-gender can construct» (Burnett e Bonami (2018, 5).

Nel Belgio francofono, l'uso di forme femminili nel linguaggio amministrativo viene introdotto con le linee guida di Lenoble-Pinson (1994), aggiornate nel 2005 (Gomez Gane 2017). In Svizzera, nel cantone francofono di Ginevra, nel 1988 viene adottato un regolamento per l'uso di nomi di professione femminili e nel 2000 è lo stesso governo federale a promuovere la *Guide de formulation non sexiste des textes administratifs et législatifs de la Confédération* (Chancellerie federale 2000). La Confederazione federale svizzera ha linee guida anche per la lingua italiana (Cancelleria federale 1991, 2003, 2012; Cantone Ticino 1995, 2004) e per la lingua tedesca (Bundeskanzlei 2009, 2021). Per quanto riguarda il tedesco, la Germania non ha linee guida promosse a livello governativo, ma sono adottate, a livello nazionale, le linee guida dell'UNESCO (Hellinger e Bierbach 1993) e, a livello amministrativo, un documento del *Ministerium für Justiz Frauen, Jugend und Familie des Landes Schleswig-Holstein* (2000) e il *Bundesstelle für Büroorganisation und Bürotechnik bulletin* (2002), entrambi rivolti all'uso linguistico delle istituzioni di riferimento (Gomez Gane 2017). In Austria, dopo le linee guida coeve all'opera di Sabatini (Wodak et al. 1987), sono state attuate politiche per la loro applicazione ai titoli politici, professionali e accademici. Queste azioni di indirizzo politico-istituzionale hanno portato al consolidamento di un uso inclusivo e paritario della lingua (Doleschal 2009).

²⁹ Il grafico riportato nella Figura 1 di Burnett e Bonami (2018) non riporta le percentuali esatte.

In Italia, l'iniziativa della *Commissione nazionale per la realizzazione della parità tra uomo e donna* che porta alla pubblicazione delle *Raccomandazioni* di Sabatini (1986, 1987) non rappresenta il momento di avvio di una politica linguistica nazionale per un uso non sessismo dell'italiano. Primo perché le *Raccomandazioni* sono indirizzate al mondo dell'istruzione e della scuola, circoscrivendo così due precisi, ancorché fondamentali, ambiti di riferimento, secondo perché non portano allo sviluppo di una politica linguistica nazionale. Tuttavia, negli anni più recenti si sono moltiplicate le iniziative di singole istituzioni finalizzate a introdurre un linguaggio *gender-fair* nella redazione di documenti, comunicazioni interne e/o rivolte al pubblico di riferimento. Si sono dotate di linee guida alcune amministrazioni locali come per esempio la Regione Toscana (Robustelli 2012), la Regione Emilia Romagna (2015) e la Regione Abruzzo (2019), la Provincia di Bolzano (2012), il Comune di Siena (2016) e il Comune di Torino (2017), l'Agenzia delle entrate (2020), e, più di recente, anche diverse università come quella di Palermo (Riggi 2015) e di Torino (Giorcelli et al. 2015), l'Università Ca' Foscari di Venezia (2017) e l'Università degli studi di Padova (2017), l'Università degli studi di Ferrara (2018), l'Università degli studi di Reggio Calabria Mediterranea (2018), l'Università di Bologna (2020) e quella dell'Aquila (Thornton 2020), l'Università per Stranieri di Siena (2020), l'Università degli studi di Milano (2020) e l'Università di Siena (2021), l'Università degli studi di Verona (s.d.) e l'Università della Calabria (s.d.).

A livello nazionale, manca però una politica linguistica unitaria. L'uso di una lingua più inclusiva e paritaria è un tema di scarso interesse politico, confinato a poche iniziative individuali riconducibili prevalentemente ai nomi delle onorevoli Laura Boldrini e Valeria Fedeli. Subito dopo la nomina alla Presidenza della Camera, Laura Boldrini ha chiesto di essere designata come *la presidente*, anche negli atti ufficiali, e chiamata *signora presidente* dai colleghi e dalle colleghe durante le sedute del Parlamento (Villani 2020). A due anni dalla sua nomina ha poi chiesto ai deputati e alle deputate di usare nomi coerenti con l'identità di genere del/la referente nel rivolgersi ai colleghi e alle colleghe (Boldrini 2015; Villani 2020).

Valeria Fedeli da ministra dell'istruzione, dell'università e della ricerca ha promosso linee guida per il suo dicastero dell'istruzione (MIUR 2018), stimolando l'ampia produzione di linee guida universitarie riportata più sopra. Tuttavia nessuna delle due politiche è stata sostenuta dall'azione collegiale del Governo o del Parlamento, dove il 28 febbraio 2017 è stata depositata una proposta di legge dal titolo *Disposizioni in materia di concordanza dei titoli funzionali in base al sesso della persona cui sono attribuiti negli atti delle pubbliche amministrazioni* (Camera dei Deputati 2017), rimasta indiscussa nel corso della 17esima legislatura, che, nel frattempo, si è conclusa. La senatrice Valeria Valente, fra i firmatari di questa proposta di legge, ospite di un evento online organizzato dall'Università Ca' Foscari su *Lingua e Parità di genere in Italia*, in occasione delle celebrazioni dei 60 anni della sentenza n. 33/1960 della Corte costituzionale³⁰, il 9 aprile 2021, ha pubblicamente dichiarato il proprio impegno a ripresentarla. Ma alla data attuale³¹, non risulta ancora nell'agenda dei lavori del Parlamento italiano.

Un'azione mirata, specifica e diffusa su scala nazionale è stata invece di recente intrapresa, come anticipato nell'introduzione, nel mondo del giornalismo, che ha promosso la realizzazione di due linee guida sull'uso inclusivo e paritario della lingua: *Donne, grammatica e media* (Robustelli 2014), stampato dall'associazione di giornaliste GiULIA (Giornaliste Unite Libere e Autonome), e *Tutt'altro genere d'informazione. Manuale per una corretta rappresentazione delle donne nell'informazione*, a cura del Gruppo di lavoro Pari opportunità del Consiglio nazionale dell'Ordine dei giornalisti (CNOG 2015).

Il primo volume contiene, oltre alle linee guida a cura di Cecilia Robustelli, qualche studio di casi specifici per esemplificare asimmetrie, stereotipi e forme di banalizzazione delle donne tipiche del linguaggio dei media, più un'intervista a Sergio Lepri, che rinnova la sua convinzione circa l'importanza del ruolo dei mezzi d'informazione nel promuovere e sostenere una lingua *gender-fair* e, con un po' di cinismo, denuncia l'ignoranza linguistica di colleghe e colleghi che non usano le

³⁰ Fonte: <https://www.unive.it/data/agenda/1/47860>; ultimo accesso 19 novembre 2021.

³¹ 19 novembre 2021.

forme femminili, facendo evidentemente riferimento a una mancanza di metacompetenza linguistica.

Il secondo volume rappresenta un tentativo più evidente di far dialogare la riflessione linguistica con la ricerca sui contenuti dei media. Presenta infatti un'analisi della rappresentazione delle donne nell'informazione quotidiana condotta attraverso una metodologia di analisi del contenuto, tramite scheda strutturata a risposte chiuse (tipicamente usata per l'analisi dei media) che non prende in considerazione, quantomeno dichiaratamente, aspetti linguistici, ma che attesta il carattere androcentrico e stereotipato dell'informazione italiana che sotto-rappresenta le donne e assegna loro ruoli narrativi marginali. Aspetto, quest'ultimo, ripreso dalla linguista Stefania Cavagnoli, a cui è affidata la terza parte del volume, che comprende le regole per un uso dell'italiano adeguato al genere e una riflessione sul ruolo della lingua nel processo di costruzione della realtà e sul ruolo dei media nei processi di trasformazione della lingua:

è la stampa che facilita o rende più difficile la trasformazione dell'uso linguistico [...] è la stampa che "decide", più o meno consapevolmente, che immagine di mondo fornire a chi legge. Spesso tale aspetto non è chiaro nemmeno a chi scrive. Riflettere sull'uso della lingua, in questo caso per una lingua adeguata al genere, significa riflettere sui ruoli di chi agisce all'interno di una determinata società e perfino proporre modifiche e adattamenti dei ruoli. Un'assunzione di responsabilità del potere che è legato a questa professione (Cavagnoli 2015, 126-127).

Capitolo 3 Media e genere

Questo capitolo presenta brevemente l'origine e lo sviluppo della ricerca in materia di media e genere. Il primo paragrafo introduce i principali modelli teorici che sono nati e si sono sviluppati in ambito internazionale, prevalentemente anglosassone, seguendo la classificazione proposta da Saveria Capecchi (2006, 13) che individua tre paradigmi di ricerca corrispondenti anche a diversi orientamenti dei movimenti femministi: «il modello della parità», «il modello della valorizzazione», «il modello post-genere». Il secondo paragrafo si concentra sull'origine e l'evoluzione degli studi in ambito nazionale, focalizzandosi in particolare sulla rappresentazione delle donne nei contenuti dell'informazione e sui più recenti risultati del Global Media Monitoring Project (GMMP), il più ampio e longevo progetto di ricerca sulle donne nell'informazione, a cui l'Italia partecipa dal 1995, e di cui nel 2020 si è svolta l'ultima edizione. Il terzo paragrafo approfondisce il tema della relazione fra il potere dei media di rendere visibili le donne e i loro progressi nella società, e i fattori mediali e socio-culturali che intervengono nel limitare questo potenziale.

3.1 Modelli teorici ed evidenze empiriche in ambito internazionale

Secondo Saveria Capecchi (2006), la riflessione teorica e la conseguente ricerca su media e genere può essere ricondotta a tre modelli principali, nati in ambito prevalentemente anglosassone, seguendo un'evoluzione simile a quello che caratterizza la ricerca su lingua e genere: «il modello della parità fra i sessi», il «modello della valorizzazione femminile», il «modello post-genere». Si tratta di modelli astratti, che non hanno la pretesa di essere esaustivi, ma sono funzionali a una breve rassegna storica, utile a fornire la cornice contestuale e teorica entro cui si collocano gli studi su media e genere, nati nella seconda metà del Novecento, e progressivamente sviluppatasi in coincidenza con le diverse fasi storiche del femminismo, non sostituendosi l'un l'altro ma coesistendo, e talvolta integrandosi, fra loro.

3.1.1 «Heart and home» (Tuchman et al. 1978)

Il «modello della parità fra i sessi» nasce negli anni Sessanta e Settanta del secolo scorso dall'intreccio fra le istanze del movimento femminista liberale statunitense, che, erede dell'*equal rights feminism*, rivendica l'uguaglianza fra donne e uomini, e l'approccio teorico funzionalista dei *media studies*, che attribuisce ai media un ruolo riproduttivo e rinforzante delle relazioni sociali e dei valori culturali consolidati (cfr. Grossi e Ruspini 2007). L'incontro fra queste due prospettive, femminista e mediale, dà l'avvio a una linea di ricerca che analizza, da un lato, la distanza fra l'immagine femminile veicolata dai media e la condizione reale delle donne nella società, e, dall'altro, le disuguaglianze fra donne e uomini nei contenuti così come nell'industria dei media.

Sul versante dei contenuti, a essere messe sotto accusa sono inizialmente la pubblicità, la stampa, in particolare i periodici femminili, la *fiction*, e la televisione più in generale. Alla pubblicità viene addebitata la "coltivazione" di una «mistica della femminilità» (Friedan 1963), ovvero un'ideologia che allontana le donne da prospettive di realizzazione professionale, per spingerle verso l'adesione a ideali di bellezza, maternità e vita casalinga, funzionali al mantenimento dello *status quo* della società, in cui le donne hanno una posizione subordinata rispetto agli uomini. Un'analisi articolata di come, per esempio, a livello simbolico, le promozioni commerciali ritraggano le donne in una posizione regolarmente subalterna rispetto agli uomini è quella offerta dal celebre volume *Gender advertisements* di Erving Goffman (1979). Indagando la prossemica di donne e uomini ritratti in un campione di fotografie e immagini pubblicitarie, il sociologo canadese mostra come le donne siano regolarmente ritratte come soggetti deboli, passivi, dipendenti, bisognose di guida, istruzione e protezione rispetto a uomini forti, attivi, indipendenti, fornitori di guida, istruzione e protezione, secondo aspettative tradizionali di ruoli e relazioni di genere distanti da una realtà sociale in mutamento. Dei periodici femminili si rileva una focalizzazione su tematiche stereotipicamente considerate di interesse femminile, quali la casa, la moda, la bellezza (Ferguson 1978; Phillips 1978; Lopate 1978). La *fiction* è criticata per essere uno spazio simbolico appannaggio maschile,

dove le donne sono numericamente inferiori e hanno ruoli marginali (Lemon 1978). La televisione, nel complesso, è oggetto di una medesima critica aggravata dalla tendenza a proporre contenuti violenti che prediligono le donne nel ruolo di vittima, con ripercussioni negative sulla socializzazione delle giovani generazioni (Gerbner 1978; Gerbner e Connolly 1978). Le ricerche sui diversi generi della programmazione televisiva dimostrano come il grande schermo tenda a confinare le donne nella narrazione della sfera privata degli affetti o del lavoro domestico, tenendole simbolicamente distanti sia dalla sfera pubblica della politica e dell'economia, appannaggio maschile, sia dalla realtà di una società interessata da profondi mutamenti delle identità, dei ruoli e delle relazioni di genere (Gerbner 1978; Lemon 1978; Cantor 1978).

Questa duplice asimmetria è analizzata in modo articolato in *Heart and home. Images of women in the mass media* (Tuchman et al. 1978) considerato il lavoro pionieristico nell'ambito della riflessione su media e genere. Non troppo distintamente da quanto presupposto per la relazione fra lingua e genere da Robin Lakoff (1973), i diversi contributi raccolti in questo volume concordano nel sostenere la tesi delle curatrici, secondo le quali i media non riflettono oggettivamente la realtà, ma le idee e i valori maschili, in quanto dominanti. Sotto-rappresentando le donne e confinandole entro spazi narrativi afferenti alla sfera privata, «heart and home», come recita appunto il titolo del volume, i media operano un «symbolic annihilation» delle donne (Tuchman et al. 1978, 3), escludendole dal mondo degli affari, della politica, e da tutte le questioni di pubblico interesse.

Una radicalizzazione di questa prospettiva è rappresentata dall'approccio di George Gerbner (1978), cui si è già accennato, sviluppato nell'ambito della *cultivation theory*, secondo la quale l'esposizione del pubblico ai contenuti mediali, in particolare televisivi, lungo tutto il processo di crescita dall'infanzia all'età adulta, influenza negativamente la rappresentazione della realtà, nella direzione di un mantenimento degli assetti di potere esistenti, essendo i media, da un lato, prodotti della classe sociale dominante, dall'altro, strumento di mantenimento di

questo dominio. Guardata dalla prospettiva della “teoria della coltivazione”, la mancata rappresentazione dei cambiamenti sociali che interessano identità, ruoli e relazioni di genere è ascrivibile a un problema strutturale che riguarda il sistema dei media e il loro ruolo attivo nel sostenere lo *status quo* culturale, sociale e politico della società.

The logical conclusion of our examination of the “world” constructed by television is that reinforce and maintains conventional beliefs, concepts and behaviors. [...] Women's roles support the traditional stereotypes of women prevalent in society (Gerbner 1978, 55).

Questo orientamento viene progressivamente abbandonato a favore di approcci più complessi e metodologicamente più solidi, attenti sia al mutamento del sistema dei media, sia al mutamento delle questioni di genere. Tuttavia, pone le basi per analizzare la relazione fra media e genere non soltanto a livello di contenuto ma anche a livello sistemico; una prospettiva, quest’ultima, che spinge a indagare, negli anni più recenti, i diversi fattori culturali, sociali e mediali intervenienti nel promuovere o, viceversa, ostacolare una rappresentazione delle identità, dei ruoli e delle relazioni di genere paritaria, inclusiva e aderente alla realtà. Negli anni in cui nascono le prime riflessioni sulla relazione fra media e genere la questione è ancora in qualche modo prematura, poiché il problema principale è garantire alle donne l’accesso e la carriera in professioni e organizzazioni tradizionalmente maschili, e la convinzione condivisa è che un progressivo aumento e avanzamento di carriera delle donne nel settore dei media porterà a un miglioramento dei contenuti (Butler e Paisley 1980). Tuttavia già nel 1990 la studiosa canadese Marjorie Ferguson mette in evidenza la fallacia di questo «mito ottimistico», confermata da ricerche che dimostrano come a un progressivo *empowerment* delle donne nelle organizzazioni dei media non corrisponda una progressiva maggiore e migliore inclusione femminile nei contenuti dei media (cfr. Ferguson 1990; Ross 2001; Steiner 2012). I fattori che favoriscono o sfavoriscono una rappresentazione di genere paritaria e inclusiva sono infatti molto più

complessi, come vedremo nel paragrafo 3.3, e hanno a che fare con diversi aspetti sia socio-culturali sia mediali.

3.1.2 «*Defining women*» (D'Acci 1994)

Il «modello della valorizzazione» nasce, nel corso degli anni Ottanta, su impulso del movimento femminista radicale, che, a differenza del movimento liberale, non ambisce all'uguaglianza, ritenuta una forma di omologazione delle donne agli uomini, bensì al riconoscimento della differenza. O meglio delle differenze. Questo approccio sottolinea infatti come donne e uomini non siano due categorie sociali omogenee al loro interno, distinte semplicemente sulla base del sesso e del genere, ma categorie eterogenee e variabili a seconda di caratteristiche quali il ceto, l'età, l'etnia, il livello di istruzione, la generazione, e queste differenze devono essere riconosciute e valorizzate. Nell'ambito dei *media studies*, la prospettiva femminista della valorizzazione si intreccia con un modello teorico costruzionista e interazionista, secondo il quale i significati dei contenuti mediali sono il risultato di un processo non eterodiretto dai media, ma sempre situato, interattivo e negoziato con il pubblico (cfr. Grossi e Ruspini 2007). Detto altrimenti, le rappresentazioni di genere veicolate dai media non sono il risultato di un mero riflesso dei rapporti di potere fra donne e uomini (Tuchman et al. 1978), e neanche una costruzione univoca e immutabile del dominio maschile (Gerbner 1978), bensì costruzioni sociali condivise, che mutano nello spazio e nel tempo, i cui significati vengono attivamente elaborati dal pubblico e diversamente interpretati a seconda delle caratteristiche dello stesso. Sulla base di questo assunto, a partire dagli anni Ottanta del Novecento, diverse studiose, e studiosi, americane e britanniche, e anche italiane, come vedremo nel paragrafo successivo, rovesciano la prospettiva critica delle colleghe che le hanno precedute e ri-valutano le produzioni mediali focalizzandosi in particolare su quelle che pongono al centro della narrazione l'universo femminile. Orientate/i alla ricerca della differenza, non sempre però trovano che la specificità femminile è valorizzata.

Gli studi sulle pubblicità e sulle riviste femminili, per esempio, colgono sì il valore innovativo offerto da questi spazi narrativi, e il loro ruolo propulsivo nella direzione dell'emancipazione e dell'*empowerment* femminile (cfr. McRobbie 1999; Gauntlett 2002, cit. in Capecchi 2007), ma ne individuano anche l'assoggettamento a esigenze puramente commerciali (cfr. Winship 1987; Douglas 1995, cit. in Capecchi 2007). McRobbie (1999), per esempio, sostiene che alcune riviste femminili, come *Cosmopolitan*³², possono considerarsi femministe, perché propongono l'immagine di una donna sessualmente emancipata, non più aderente al modello della donna-oggetto del desiderio sessuale maschile, tipico delle riviste maschili. Winship (1987) riconosce sì la radicalità del messaggio di una rivista come *Cosmopolitan* che propone modelli diversi dalla tradizionale "donna madre, moglie e casalinga", ma nel contempo ne coglie la contraddizione, riconoscendo che «è evidente come ogni azione della *donna-cosmo* sia in funzione dell'uomo» (Capecchi 2006, 32). Douglas (1995) intuisce una questione divenuta molto attuale di recente e nota con il termine *femvertising*, ovvero «il femminismo che fa vendere» (Capecchi 2018, 67), vale a dire narrazioni, pubblicitarie ma anche di altro genere, che sfruttano le istanze femministe a scopi commerciali, con il risultato di assoggettare le donne alle esigenze del sistema capitalista. Gauntlett (2002), infine, a cui si deve uno sguardo ottimistico sul ruolo della produzione mediale nella ridefinizione di identità, ruoli, e relazioni di genere (cfr. Gauntlett 2008) osserva che i *magazine* femminili non solo propongono immagini di donne forti, libere e indipendenti, ma ritraggono, al contempo, «l'uomo come *uomo-oggetto* e lo fanno con un sorriso consapevole di come gli uomini per secoli hanno trattato le donne; dunque con la finalità, a livello simbolico, di riequilibrare il potere fra i sessi» (Capecchi 2007, 19).

Una rivalutazione più convinta e unanime delle narrazioni femminili è invece quella operata nell'ambito dei cosiddetti *audience studies*, che indagano, per la

³² *Cosmopolitan* è una rivista oggetto di molti studi anglosassoni, probabilmente perché è stato uno dei primi magazine femminili – è nata nel 1964 negli Stati Uniti, ed è stata poi diffusa in Gran Bretagna nel 1972, in Italia nel 1973 – a caratterizzarsi per un target di giovani donne lavoratrici, e a farsi interprete della "liberazione sessuale" (Capecchi 2006, 32).

prima volta, le gratificazioni dell'*audience* femminile nella fruizione di prodotti pensati e realizzati per le donne, come per esempio le *soap opera*. Fra i numerosi lavori che rivalutano il piacere di queste produzioni seriali possiamo ricordare *Watching Dallas. Soap opera and the melodramatic imagination* di Len Ang (1985), che orienta anche alcune importanti ricerche nazionali pubblicate dalla Rai, come quelle di Capecchi (2000, 2004), mostrando come questo genere televisivo, a lungo guardato come un genere minore, rivaluti pubblicamente la sfera privata, mediatizzandola, e permetta anche alle spettatrici di riflettersi in una cultura che riconoscono come propria. In particolare le *soap opera* permettono alle spettatrici, e anche agli spettatori, come dimostra Capecchi (2002, 2004), di interrogare il testo da un punto di vista emozionale.

Infine, aderiscono al «modello della valorizzazione» gli studi focalizzati sulla *fiction* che indagano modelli di donna innovativi e sfidanti gli stereotipi di genere. Fra le ricerche che rientrano in questo filone, rimanendo ancor oggi un punto di riferimento per gli studi nazionali dedicati alla *fiction* italiana (cfr. Buonanno 2014b), possiamo ricordare *Defining women: television and the case of Cagney and Lacey* (D'Acci 1994), che analizza la figura delle protagoniste del poliziesco americano *Cagney and Lacey*: donne emancipate, intelligenti, legate da forti rapporti di amicizia e solidarietà, che si contraddistinguono come “eccezioni” nel panorama (quasi) tutto maschile dei polizieschi anni Ottanta e Novanta del Novecento, rimettendo al centro della narrazione la dimensione del privato valorizzata come specificità femminile.

Sul versante della rappresentanza femminile nelle organizzazioni dei media, il «modello della valorizzazione» non ambisce tanto a un riequilibrio numerico fra professioniste e professionisti, quanto a una messa in discussione dell'industria dei media, la cui cultura si ritiene debba essere rivista in modo tale da fare emergere le competenze e il punto di vista delle donne. Questo cambio di prospettiva viene spinto dalle evidenze emerse da alcune indagini che approfondiscono le condizioni e i ruoli delle professioniste dei media, cresciute, in termini numerici, fra la fine degli anni Ottanta e Novanta del secolo scorso, ma non in termini di potere. In

particolare, si possono citare le ricerche di Liesbet van Zoonen (1994) e Pamela Creedon (1989). Liesbet van Zoonen (1994) evidenzia come i settori della produzione mediale più accessibili alle donne siano anche quelli simbolicamente più vicini agli ambiti stereotipicamente considerati femminili, come la cura, la famiglia, l'evasione. Nell'informazione, per esempio, le donne realizzano servizi di *soft-news*, dedicate a costume, società, cronaca e spettacolo; mentre le notizie di economia e politica (*hard-news*) sono appannaggio dei colleghi maschi. Creedon (1989) indaga la segregazione orizzontale e verticale nel mercato delle professioni giornalistiche, dimostrando come le donne, da un lato, faticano a raggiungere le posizioni di vertice e, dall'altro, a parità di mansioni, hanno salari più ridotti rispetto ai colleghi. Partendo dall'analisi di questa condizione, che evidenzia il fallimento per le donne di adeguarsi al sistema mediale dominante, la studiosa americana riflette sull'opportunità di spostare l'obiettivo dalla parità alla differenza, cercando un modo alternativo di fare giornalismo: che dia più spazio alle donne sia come oggetto di notizia, sia come fonti. Un obiettivo su cui è tornata di recente la ricerca focalizzata sull'informazione, anche a livello nazionale, come vedremo in un paragrafo successivo (§ 3.3).

3.1.3 Oltre il binarismo di genere

Il periodo a cavallo fra la fine del Novecento e il Duemila segna l'affermazione del «modello post-genere», che comprende diverse prospettive femministe accomunate dalla messa in discussione della definizione binaria del genere, basata sulla etero-normatività, e da una riflessione sul potere (ambivalente) delle nuove tecnologie della comunicazione e dell'informazione (TIC). Per dovere di completezza, riporto di seguito una sintesi delle principali teorie afferenti a questo modello anche se è stato sviluppato prevalentemente nell'ambito degli studi femministi e della filosofia, e non dei *media studies*.

Secondo le diverse studiose che possono essere considerate esponenti di questo modello, la nozione di genere, introdotta da Gayle Rubin (1975) per spiegare le differenze fra donne e uomini in termini di costruzioni sociali

storicamente fondate, è un superamento solo parziale del concetto scientifico del sesso, perché è una nozione dualista che pensa la relazione fra donne e uomini in termini di opposizione etero-normata. Le teoriche del «modello post-genere» propongono quindi in generale di ripensare la relazione fra i generi in termini di nuove soggettività, e in particolare, di ripensarle in una relazione alternativa con i media e le TIC.

Teresa De Laurentiis (1987) sostiene che i media sono «gender technologies», tecnologie che partecipano attivamente alla costruzione del genere, favorendo un adeguamento, per le donne, alle rappresentazioni dominanti. Per questo propone di pensare una nuova soggettività, il «soggetto eccentrico», indisciplinato e mobile rispetto ai confini storicamente assegnati al femminile. Non troppo distintamente da quanto sostiene Judith Butler (1990, 1993, 1999) che, rivisitando il tradizionale concetto di genere, orienta non solo la ricerca su lingua e genere, come abbiamo visto nel Capitolo 2 (§ 2.1.5), ma anche quella su genere e media verso un'analisi delle identità *queer* e *gender fluid*.

Donna Haraway (1991) prospetta un superamento del dualismo di genere attraverso la figura del/la «cyborg», né uomo, né donna, né umano, né macchina, una metafora della condizione umana a partire dalla quale la filosofa americana invita le donne a ripensare il rapporto con la tecnologia, in particolare con le nuove TIC che consentono l'uso di identità virtuali de-corporizzate, come spazio dove decostruire il dualismo di genere e sperimentare nuovi modelli identitari che vadano oltre la tradizionale opposizione uomo/donna, mente/corpo, attivo/passivo, e così via.

Finalizzata a rendere visibili identità storicamente minoritarie è la visione della studiosa afroamericana bell hooks³³, pseudonimo di Gloria Jean Watkins, che, in un'ottica intersezionale, suggerisce un «oppositional gaze» permanente che si opponga al «male gaze» (Mulvey 1975) dominante nei contenuti dei media, che non è solo lo sguardo maschile, ma anche la prospettiva «white» della borghesia,

³³ Il minuscolo è una scelta dell'autrice (cfr. Giangrande 2011).

che non tiene conto delle differenze di ceto, razza, istruzione, e così via (bell hooks 1984, 1988, 1992).

Rosi Braidotti (1994), infine, introduce il «soggetto nomade», che, similmente al «soggetto eccentrico» di Teresa De Laurentis, si caratterizza come soggetto di identità complesse, multiple, insubordinate ai modelli di genere convenzionali, senza però pretendere di superare il dualismo di genere, che Braidotti recupera, valorizzandolo, in termini di differenze, non solo fra donna e uomo, ma anche fra donna e donna.

3.2 Origine e sviluppo degli studi in ambito nazionale

In Italia, la riflessione su media e genere segue originariamente e prevalentemente due linee di ricerca, aderenti l'una al «modello della parità fra i sessi», l'altra al «modello della valorizzazione», spesso integrate fra loro. Molti studi che analizzano il grado di prossimità (o distanza) fra la condizione delle donne nella società italiana e la loro rappresentazione mediale non rinunciano infatti a riflettere su come e quanto i media possano contribuire all'*empowerment* femminile. E molti studi finalizzati a rintracciare segnali di innovazione nella rappresentazione femminile, e maschile, nei termini di una valorizzazione della differenza, sottolineano la persistenza di uno scarto fra l'immaginario costruito dai media e la realtà. Quanto al «modello post-genere», non ne farò cenno, in primo luogo perché orienta un esiguo numero di studi nazionali (per esempio, De Maria 2014), in secondo luogo perché esula dall'obiettivo di questa tesi che è quello di indagare le rappresentazioni di genere nel linguaggio dell'informazione da una prospettiva binaria, in particolare focalizzata sulle donne, come anticipato nell'Introduzione (§1.2).

La prima linea di ricerca, in ordine cronologico, nasce nell'ambito della sociologia della comunicazione e degli studi culturali³⁴, ed è prevalentemente riconducibile ai lavori di Milly Buonanno e Saveria Capecchi. Le ricerche di Buonanno si focalizzano, da un lato, sull'immaginario femminile veicolato da diversi tipi di media, generi e sottogeneri mediali, quali la stampa periodica femminile (Buonanno 1967, 1968, 1975, 1978b), la televisione (Buonanno 1982, 1983, 1994), la *fiction* (Buonanno 2008, 2014b, 2017), dall'altro, sulle donne nel giornalismo (Buonanno 1978a, 1988, 1999, 2005, 2006, 2014c, 2015a, 2015b). Saveria Capecchi è la prima in Italia a studiare l'*audience* in una prospettiva di genere, mettendo in evidenza le gratificazioni e gli usi sociali delle *soap opera*, un genere televisivo che, come abbiamo visto (§ 3.1.2), a metà anni Ottanta del Novecento inizia a essere rivalutato dalle studiose dei media che ne mettono in evidenza le caratteristiche testuali e pragmatiche che possono gratificare il pubblico (Capecchi 2000, 2004). Numerose sono poi le ricerche e le riflessioni teoriche di Saveria Capecchi sulla rappresentazione delle donne in televisione (Capecchi e Pallotta 2001), sulla stampa periodica (Capecchi 1995), nella pubblicità (Capecchi 2011), tutte finalizzate a indagare l'influenza biunivoca e circolare fra i contenuti mediali e i cambiamenti sociali, in una prospettiva costruzionista e interazionista, basata sul presupposto che i media «da un lato contribuiscono ad accelerare i cambiamenti sociali stimolando l'emancipazione femminile, dall'altro cristallizzano solo alcuni aspetti della vita delle donne che divengono così nuovi stereotipi, senza dare loro sufficiente voce e dignità» (Capecchi 2002, 116).

Questa prospettiva caratterizza anche le ricerche promosse dalla Rai realizzate nell'ambito del programma di Verifica Qualitativa dei Programmi Trasmessi (VQPT) e pubblicate nell'omonima collana editoriale (Buonanno 1982; Cornero 2001), prevalentemente con l'intento di fornire evidenze empiriche su come e quanto la televisione favorisca (oppure ostacoli) l'*empowerment* femminile,

³⁴ In Italia, l'ambito della sociologia della comunicazione e degli studi culturali è quello in cui prevalentemente sono stati realizzati, e continuano a essere realizzati, gli studi su media e genere, come di recente sottolineato da Milly Buonanno, nel corso del convegno internazionale *Gender and Media Matters. Widening the horizons of the field of study* (Università di Roma La Sapienza, 15-16 ottobre 2021).

sulla base delle quali individuare specifiche azioni di intervento in ambito autoriale e redazionale.

La seconda linea di ricerca è quella sostenuta dalle istituzioni, nazionali o internazionali, nella prospettiva, tipica del «modello della parità fra i sessi», di verificare il grado di inclusione delle donne nei contenuti e nell'organizzazione dei media e di fornire opportune raccomandazioni di intervento. Il primo studio realizzato con questa finalità è promosso dalla *Commissione per la realizzazione della parità tra uomo e donna* istituita durante il Governo Craxi (1983-1987) che, nello stesso periodo in cui incarica Alma Sabatini di studiare il sessismo nella lingua italiana, affida all'antropologa Gioia Di Cristofaro Longo (1986, 1992) la cura di due volumi sull'immagine della donna veicolata dai media. A cura di Gioia Di Cristofaro Longo (1995) è poi anche la pubblicazione dei risultati nazionali della prima edizione del GMMP, il più ampio e longevo progetto di ricerca sulle donne nell'informazione quotidiana, che, in linea con gli obiettivi della *Dichiarazione e Piattaforma d'azione di Pechino* (ONU 1995), sottoscritta dagli stati membri delle Nazioni Unite, dal 1995 fornisce periodicamente dati sulla rappresentazione femminile nelle notizie quotidiane di tutto il mondo ed è diventato un punto di riferimento imprescindibile per lo studio delle relazioni fra donne e informazione quotidiana (Azzalini e Padovani 2016; Djerf-Pierre e Esdröm 2020a, 2020b, 2020c).

Come nel caso della riflessione su lingua e genere, l'interesse delle istituzioni italiane per la relazione fra media e genere è stato discontinuo negli anni. Dopo le ricerche a cavallo fra gli anni Ottanta e Novanta (Di Cristofaro Longo 1986, 1992) è seguito un decennio di sostanziale disinteresse interrotto prima dal Consiglio Nazionale dell'Economia e del Lavoro (CNEL) che, all'inizio del Duemila, commissiona due ricerche all'Osservatorio di Pavia sulla rappresentazione delle donne in televisione (CNEL 2002, 2004), poi dall'Unione europea che finanzia il "libro bianco" *Women and Media in Europe* (CENSIS 2006), realizzato con il patrocinio del Ministero per i Beni e le Attività Culturali. Queste tre ricerche preparano il terreno per il dibattito pubblico, e politico, che caratterizza il decennio successivo, coinvolgendo diverse parti interessate, a partire dalla Rai.

Nel 2009, in concomitanza con il primo scandalo sessuale³⁵ che coinvolge Silvio Berlusconi, fondatore di Mediaset, la principale televisione privata nazionale in competizione diretta con la Rai, e presidente del Consiglio al suo terzo mandato nel periodo 2008-2011, la giornalista e attivista Ida Dominijanni scrive una serie di articoli sul quotidiano *Il Manifesto* ponendo all'attenzione del pubblico le complesse dinamiche di relazione fra sesso, politica e televisione in Italia, nei vent'anni che trascorrono fra l'inizio degli anni Novanta e il 2009 (Guaraldo 2010; Cosenza 2012; Formato 2018). Nello stesso anno esce il documentario di Lorella Zanardo, Marco Melfi e Cesare Cantù, *Il corpo delle donne*³⁶, che denuncia la subalternità dell'immagine femminile alle logiche della competizione televisiva fra Rai e Mediaset. Questo documentario ha il merito di riportare l'attenzione dell'opinione pubblica e del movimento femminista sulla questione della rappresentazione delle donne nei media. Sono infatti questi gli anni in cui nasce *l'Appello donne e media*³⁷, un'azione di lobby presso la *Commissione di vigilanza della Rai*, attraverso la quale la Rai viene portata a siglare un contratto di servizio con il Ministero dello Sviluppo Economico (MISE 2011) in cui, per la prima volta, si impegna a una rappresentazione equilibrata e rispettosa della dignità della donna e a un monitoraggio periodico sulla rappresentazione femminile nella programmazione televisiva, che dal 2012 viene realizzato annualmente (Rai-Osservatorio di Pavia, 2014, 2015, 2018, 2019; ISIMM Ricerche e Università degli studi Roma Tre 2016, 2017; IZI et al. 2020). In questo stesso periodo, si colloca l'avvio di un impegno delle giornaliste, con iniziative come quelle promosse dal Gruppo di lavoro Pari opportunità del Consiglio nazionale dell'Ordine dei giornalisti (CNOG 2015) e dall'associazione GiULiA, che oltre al già menzionato volume su un uso non sessista della lingua (Robustelli 2014), pubblica diversi manuali su come riconoscere e evitare gli stereotipi di genere nei media (GiULiA 2019), in particolare

³⁵ Il riferimento è alla presunta relazione di Silvio Berlusconi con una giovane minorenne (Noemi Letizia) denunciata dall'allora moglie Veronica Lario, in una lettera aperta al quotidiano *la Repubblica*, il 3 maggio 2009. Alla denuncia sono seguite diverse inchieste giornalistiche e anche giudiziarie, che hanno portato alla luce altre relazioni e scambi di favori (sesso in cambio di denaro o carriera politica) fra Berlusconi e giovani ragazze. Per una ricostruzione sintetica si veda: https://it.wikipedia.org/wiki/Silvio_Berlusconi (ultimo accesso 19 novembre 2021).

³⁶ Fonte <https://www.youtube.com/watch?v=nPpIn0b6-x4>; ultimo accesso 19 novembre 2021.

³⁷ Fonte: <https://www.appellodonnemedia.it/storia/>; ultimo accesso 19 novembre 2021.

nella cronaca di casi di violenza contro le donne (GiULiA 2021b) e nell'informazione sportiva (GiULiA 2021a).

Data l'ampia produzione che si è sviluppata nel corso del tempo, in questo paragrafo mi concentrerò sull'informazione, attraverso una rassegna delle principali ricerche sulla rappresentazione delle donne (§ 3.2.1) e sulla condizione e sul ruolo delle giornaliste (§ 3.2.2).

3.2.1 *Le donne nell'informazione*

Il 14 marzo 1985 l'Italia ratifica la *Convenzione sulla eliminazione di ogni forma di discriminazione nei confronti della donna* adottata dalle Nazioni Unite il 18 dicembre del 1979 (ONU 1979). La ratifica impegna l'Italia in un programma di interventi sociali e culturali:

al fine di modificare gli schemi e i modelli di comportamento degli uomini e delle donne e giungere all'eliminazione dei pregiudizi e delle pratiche consuetudinarie o di altro genere, che siano basate sulla convinzione dell'inferiorità o della superiorità dell'uno o l'altro sesso o sull'idea di ruoli stereotipati degli uomini e delle donne (Legge 14 marzo 1985 n. 32, art. 5).

In questo contesto storico, si collocano le iniziative della *Commissione per la realizzazione della parità fra uomo e donna* (istituita nel 1984) sugli usi discriminatori della lingua italiana (Sabatini 1986, 1987, 1993), e sulla rappresentazione delle donne nei media (Cristofaro Longo 1986, 1992):

compito fondamentale della Commissione per la realizzazione della parità uomo donna [...] è quello di sostenere l'accoglimento e l'applicazione delle indicazioni delle organizzazioni internazionali mediante i risultati di ricerche che [...] provino con l'evidenza dei fatti la fondatezza dei presupposti sui quali basare le necessarie "azioni positive" (Marinucci 1986, 10).

Immagine donna (Di Cristofaro Longo 1986) è una ricerca che analizza diversi tipi di media e generi narrativi: la stampa quotidiana, i periodici femminili

«nuovi e di recente mutamento editoriale» (Di Cristofaro Longo 1986, 11), i settimanali di politica, cultura e attualità, le trasmissioni televisive e la pubblicità (nei periodici femminili, nei settimanali di politica, cultura e economia, in TV)³⁸. L'obiettivo della ricerca è quello di verificare quali modelli di donna vengano veicolati dai media italiani e quanto essi siano in linea rispetto al progresso della parità di genere. Attraverso una lettura critica dei materiali di ricerca, analizzati secondo l'approccio della parità fra i sessi, il gruppo di lavoro coordinato da Gioia Di Cristofaro Longo giunge alla conclusione che tutti i diversi modelli di donna rappresentati, anche quelli in apparenza più attuali, confermano una posizione di subalternità femminile. Poiché alcuni fenomeni mediali indicativi di tale subalternità sono attestati anche da studi successivi, sino ai più recenti, e sono rilevanti per il lavoro di ricerca presentato in questa tesi, in questo paragrafo mi soffermo sui risultati che riguardano l'informazione quotidiana della stampa e della televisione.

I risultati relativi ai modelli di donna emergenti nei quotidiani³⁹ sono distinti in 5 *item* tematici: la politica, la cultura, la cronaca, lo spettacolo, lo sport⁴⁰. Le protagoniste dell'informazione politica sono poche, sono rappresentate come eccezioni rispetto ai modelli femminili tradizionali, e, in una sorta di «tentativo di riportare entro canali di normalità una realtà considerata eccezionale» (Di Cristofaro Longo 1986, 21), sono spesso qualificate come madri o mogli di uomini noti, o ritratte mentre sono impiegate in attività ritenute tipicamente femminili (es. fare la maglia). Spesso sono citate solo con il nome proprio, in un modo confidenziale che non viene mai usato per nominare gli uomini, al più designati soltanto attraverso il cognome. Poi ci sono le figure di contorno: le *first lady*, le

³⁸ Il campione copre il mese di novembre 1984.

³⁹ Il campione di quotidiani comprende i seguenti titoli: *Avanti*, *Avvenire*, *Corriere della sera*, *Il Giornale*, *Il Giorno*, *Il Manifesto*, *La Nazione*, *Il Popolo*, *la Repubblica*, *I Resto del Carlino*, *Il Secolo XIX*, *La Stampa*, *Il Tempo*, *L'Unità*, *La Voce Repubblicana*, più un «campione d'appoggio dal 15 al 30 novembre» (Di Cristofaro Longo 1986, 19) comprensivo de *La Gazzetta del Mezzogiorno*, *Il Gazzettino*, *Il Mattino*, *La Notte*, *L'Unione Sarda*.

⁴⁰ Non è chiaro come siano individuati questi 5 *item*, perché manca una presentazione della metodologia di analisi. Quello che si può desumere dal materiale pubblicato è che l'indagine sia stata condotta attraverso una lettura critica dei giornali, con particolare attenzione ad alcune asimmetrie linguistiche messe in evidenza in quello stesso periodo da Sabatini (1986, 1987).

mogli, le figlie, le nuore di-. Rispetto alle tematiche, le politiche fanno notizia in relazione ad argomenti considerati di pertinenza femminile, come il divorzio, l'età pensionabile per le donne, la violenza sessuale (Di Cristofaro Longo 1986, 21). Emergono, positivamente, in particolare nei giornali di partito, le donne come soggetto collettivo: attiviste del movimento femminista italiano, e anche di quello indiano. Nelle pagine culturali, le donne sono recensite come autrici di opere letterarie o di arte figurativa, oppure compaiono come soggetti letterari, o come gruppo sociale interessato da questioni specifiche al centro di iniziative culturali dedicate, per esempio, la violenza contro le donne e il femminismo. Temi, questi ultimi, che vengono trattati però con poca «empatia» (Di Cristofaro Longo 1986, 22) e scarso coinvolgimento delle donne, raramente interpellate come fonti d'informazione.

Nella cronaca si concentra la maggior parte delle notizie che dà spazio alle donne «molto spesso emergenti per il fatto di essere mogli, vedove, ex mogli, sorelle, figlie, madri, nipoti di uomini noti» (Di Cristofaro Longo 1986, 23), oppure vittime di violenza o criminali, sotto processo, o giustiziate, prostitute, suore, nobili, o «madri di drogati» (Di Cristofaro Longo 1986, 23). Quanto alle tematiche, la violenza contro le donne è il tema dominante. Per il resto, soprattutto sui giornali di partito, compaiono notizie sulla famiglia e sul lavoro, in particolare sull'accesso delle donne in magistratura, nell'agenda politica del periodo. Molto spazio alle donne viene dato anche nelle pagine dello spettacolo, dove le attrici sono numerose, spesso ritratte in fotografie, biografie, interviste. Le pagine sportive, infine, danno poco spazio alle protagoniste dello sport e più spazio alle mogli o fidanzate di dirigenti o campioni, prevalentemente del mondo calcistico.

L'analisi sui notiziari televisivi (TG1 e TG2) rileva anzitutto una scarsa attenzione per l'universo femminile: su 29 ore e 19 minuti di trasmissione, equamente ripartiti fra TG1 e TG2, le notizie relative a questioni o figure femminili coprono solo 15 minuti e 16 secondi del TG1 e 14 minuti e 11 secondi del TG2. Le donne compaiono in qualche notizia di politica interna o estera, in poche notizie di cultura mentre sono più numerose nelle notizie di cronaca, dove, nel periodo

considerato (dal 12 novembre al 12 dicembre 1984), trovano spazio soprattutto come madri di persone disabili. Per il resto, solo un'esperta viene interpellata: una sessuologa, per parlare della sessualità «negli handicappati gravi» (Di Cristofaro Longo 1986, 271); e solo una donna fa notizia, in quanto imputata in un processo di omicidio.

Non è azzardato, a questo punto – conclude l'autrice – mettere in evidenza che una delle configurazioni culturali preminenti con cui è vista la donna o quanto meno si è disponibili ad una maggiore evidenziazione, riguarda le donne come madri in situazioni difficili (Di Cristofaro Longo 1986, 271).

La donna nei media (Di Cristofaro Longo 1992), su cui non mi soffermerò, è una pubblicazione che si limita a raccogliere le segnalazioni ricevute dallo *Sportello immagine donna* istituito per segnalare contenuti offensivi o discriminatori nei confronti delle donne e merita di essere citata in quanto prima testimonianza di una consapevolezza condivisa a livello di opinione pubblica sui principali e più evidenti fenomeni di discriminazione mediatica femminile, quale la sessualizzazione e l'oggettivazione del corpo delle donne nelle pubblicità e nei programmi di intrattenimento televisivo.

A distanza di oltre dieci anni dalla ricerca coordinata da Di Cristofaro Longo (1986), nel 1999, la Rai promuove un'analisi della rappresentazione femminile in diversi generi televisivi che compongono i palinsesti di Rai 1, Rai 2 e Rai 3 (Cornero 2001). Si tratta probabilmente della prima ricerca nazionale realizzata su un vasto e articolato campione di osservazione, analizzato con appropriate e diversificate tecniche di analisi della televisione. La ricerca comprende, fra l'altro, un'analisi del contenuto quali-quantitativa dei telegiornali Rai (Siliato 2001) che consente di approfondire e di misurare diversi fenomeni descritti nello studio di Di Cristofaro Longo (1986) e persistenti a distanza di oltre un decennio. Su un campione di analisi composto dalle edizioni di mezzogiorno e prima serata del TG1, TG2 e TG3, trasmesse dal 25 al 31 gennaio 1999, le donne sono citate o intervistate in misura

nettamente inferiore rispetto agli uomini (12% vs. 88%)⁴¹; le notizie in cui le donne hanno maggiore spazio sono quelle di spettacolo e di cronaca, pagina quest'ultima che concentra quasi la metà delle notizie trasmesse (45,6%). Viceversa, le notizie in cui hanno minor spazio sono quelle di politica interna (1%⁴²), che costituisce, dopo la cronaca, l'argomento più trattato dai notiziari analizzati, e quello più esclusivamente appannaggio maschile. Quanto ai ruoli narrativi, le donne sono interpellate prevalentemente come «gente della strada», «gente dello spettacolo» o «vittima», gli uomini invece come «attori politici», «attori istituzionali», «attori sportivi». Gli «esperti o intellettuali» sono tutti uomini, così come gli «sportivi» (Siliato 2001, 99).

Questa forte asimmetria fra ruoli femminili e maschili è confermata anche nell'analisi dei programmi di approfondimento informativo, contenuta nello stesso volume, che non misura l'entità del fenomeno, ma ne mette in evidenza alcuni elementi qualitativi, dimostrando come le donne vengono tendenzialmente svalutate in tutte le loro rappresentazioni (Capecchi e Pallotta, 2001). Gli uomini intervistati o ospiti di questo genere di programmi in qualità di esperti sono molto più frequenti delle donne, coprono molti ambiti del sapere (politica, economia, scienza, e così via), e sono valorizzati sia dai titoli professionali usati per presentarli sia dai giudizi positivi dei conduttori e delle conduttrici dei programmi. Le esperte interpellate sono poche e non godono di altrettanta autorevolezza: spesso sono presentate solo con nome e cognome, senza titolo professionale, a volte le loro dichiarazioni sono riproposte ad esperti maschi, a titolo di validazione. In generale, l'identità professionale delle donne è trascurata a vantaggio di altre identità, come quella della madre, della casalinga, della vittima. E questo indipendentemente dal genere del/la giornalista che le interpellata: «molte conduttrici/giornaliste si adeguano al punto di vista svalorizzante le donne nel fare informazione, omologando i propri atteggiamenti a quelli degli uomini» (Capecchi e Pallotta 2001, 163).

⁴¹ Citazioni 12,4%, tempo di parola nelle interviste 12,0%.

⁴² Citazioni 0,87%, tempo di parola nelle interviste 1,0%.

Un'asimmetria simile è evidenziata nella ricerca CNEL (2004). Anche in questo caso, oggetto dell'analisi non è l'informazione quotidiana ma quella periodica, in particolare dei *talk show* serali, in cui esperte/i e opinioniste/i hanno un ruolo cruciale nello spiegare e interpretare le notizie o i temi di attualità. La ricerca, condotta su un campione di programmi trasmessi da Rai, Mediaset e La7 a cavallo fra il 2003 e il 2004, dimostra che le donne hanno meno spazio degli uomini (24%, su un totale di N=1.964 ospiti), sono più presenti in programmi che alternano l'approfondimento di temi connessi all'attualità economica, politica e sociale con la trattazione di tematiche leggere, di cronaca e costume, e spazi di intrattenimento, e sono meno presenti in programmi che prediligono l'approfondimento politico o temi di interesse generale e di grande attualità. Per quanto riguarda i ruoli narrativi, gli uomini sono invitati e valorizzati più di frequente in quanto competenti e opinionisti, mentre le donne come narratrici di esperienze personali, testimonianze, storie di vita (nel 49% dei casi vs. il 25% dei casi maschili). Le donne tendono anche a esporre più frequentemente la loro dimensione privata, in particolare parlano del loro vissuto biografico più degli uomini (41% vs. 14%), i quali, viceversa, affrontano i temi proposti in trasmissione da una prospettiva generale e teorica, che riguarda l'interesse pubblico, più delle donne (75% vs. 14%). Queste tendenze sono ricondotte a scelte attribuibili alla produzione, ma anche a una maggiore disponibilità delle donne a ricoprire i ruoli loro assegnati: «sono i responsabili dei programmi [...] a invitare le donne più spesso degli uomini in qualità di testimoni piuttosto che di "competenti" o "commentatori", ed esse si attengono alla funzione per la quale sono state invitate» (CNEL 2004, 3).

Risultati simili sono attestati dalla ricerca del CENSIS (2006), sulla rappresentazione delle donne in una settimana campione del mese di marzo 2005, in diversi generi e canali televisivi (Rai, Mediaset e La7). Nei programmi di approfondimento, la figura dell'esperta non è assente, ma è relegata all'approfondimento di tematiche leggere, come l'astrologia (20,7%) o l'artigianato locale (13,8%), mentre argomenti come la politica, le scienze, la guerra, la cultura sono appannaggio maschile. Nei programmi di informazione, come i TG e le

rubriche, la visibilità femminile è concentrata nei servizi di cronaca nera (67,8%), in cui la donna è coinvolta come vittima, e in alcuni casi come carnefice. Infine, la visibilità delle donne impegnate in politica è scarsa, pari al 6,4%.

Segnala qualche cambiamento, ma anche elementi di continuità, la ricerca realizzata dall'Osservatorio di Pavia per il Gruppo di lavoro Pari opportunità del CNOG (2015), che analizza i principali notiziari televisivi nazionali, trasmessi da Rai, Mediaset, La7 e Sky, più 15 giornali quotidiani⁴³; in entrambi i casi circoscrivendo il campione a una settimana ricostruita⁴⁴ nell'arco di due mesi da lunedì 7 febbraio a domenica 23 marzo 2014. Le donne che fanno notizia sono una percentuale ristretta (14%) e ottengono visibilità soprattutto in notizie di cronaca nera, sia in TV (27%), sia sui quotidiani (32%), in quanto vittime o sopravvissute (46% nei quotidiani, 50% nei TG), oppure «gente comune» (54% nei quotidiani, 42% nei TG). Le esperte sono poche (il 17% sui quotidiani, il 19% nei TG) ma sono uscite dall'invisibilità attestata dalla ricerca di Siliato (2001), e ancor prima di Di Cristofaro Longo (1986). E anche l'informazione politica comincia a fare spazio alle donne, che registrano una presenza dell'8% sui quotidiani, dove la pagina politica concentra il 45% delle fonti citate o interpellate, e del 10% nei TG, dove le notizie politiche concentrano il 43% delle fonti.

Quanto la visibilità e la piena inclusione delle donne nell'informazione italiana stia progredendo, o meno, è attestata con dati recenti dal GMMP (Azzalini e Padovani 2021). Come più sopra accennato, il GMMP è il più ampio e longevo progetto di ricerca, e di *advocacy*, sulle donne nei *news media*; l'Italia vi partecipa sin dalla sua prima edizione, che si è svolta nel 1995 ed è stata coordinata per l'Italia da Gioia Di Cristofaro Longo (1995). Questo studio raccoglie ogni 5 anni dati quantitativi e qualitativi sulle donne (e sugli uomini) sia come *newsmaker* e/o fonti

⁴³ *Avvenire, Corriere della sera, Il fatto quotidiano, Il Gazzettino, Il Giornale, Il Mattino, Il Messaggero, Il Sole 24 ore, La Gazzetta del Mezzogiorno, La Gazzetta dello sport, la Repubblica, La Stampa, Metro, QN Il Resto del Carlino, QN La Nazione.*

⁴⁴ Per settimana ricostruita si intende un periodo composto dai 7 giorni della settimana, dal lunedì alla domenica, considerati non in successione l'uno dopo l'altro nell'arco della stessa settimana, ma in successione fra loro nell'arco di 7 settimane per ciascuna delle quali si seleziona un giorno campione a partire dal lunedì, per la prima settimana, fino alla domenica, per l'ultima settimana; si tratta di un metodo utilizzato nell'analisi di contenuti medialti per ottenere un campione *random*.

delle notizie, sia come giornaliste/i che presentano o firmano le notizie, nonché evidenze empiriche sulle caratteristiche delle notizie diffuse a mezzo stampa, radio, TV e, dal 2015 anche Internet e Twitter. Nel 2020 ha visto la partecipazione di 116 gruppi di ricerca distribuiti in altrettanti paesi del mondo. Ogni gruppo di ricerca analizza tradizionalmente un campione di notizie rappresentativo dell'informazione del proprio paese, attraverso una metodologia condivisa, che integra un'analisi del contenuto quali-quantitativa, secondo tecniche di *content analysis*, con un'analisi qualitativa finalizzata a raccogliere esempi di buone e cattive pratiche giornalistiche, incentrate su diversi aspetti di costruzione della notizia, dal linguaggio, alla selezione delle fonti, tramite *case studies* (WACC 2015, 107). Questo approccio è particolarmente funzionale alla raccolta di una grande base di dati comparabili su scala transnazionale e, nel corso del tempo, anche in prospettiva longitudinale, nonché a un approfondimento della qualità dell'informazione utile a formulare raccomandazioni operative per chi fa informazione.

Nel 2020, per l'Italia sono state analizzate N=472 notizie pubblicate o trasmesse il 29 settembre 2020, la giornata prescelta su scala mondiale per il monitoraggio, da un campione rappresentativo del sistema informativo nazionale e comprensivo di otto quotidiani, sei radio, otto TV, otto siti di notizie Internet e otto account Twitter di testate giornalistiche⁴⁵ (Azzalini e Padovani 2021). Il primo risultato emerso riguarda una visibilità delle donne come fonte o oggetto di notizie mediamente pari al 26% (su un totale di N=980 persone): 24% nei media tradizionali, che crescono di tre punti percentuali rispetto al 2015, allineandosi così al dato globale (25%); 28% nei media digitali, dove a distinguersi è Twitter per la crescita di 8 punti percentuali rispetto al 2015 (dal 17% al 25%). Guardati dalla prospettiva del modello della parità dei sessi, che è il *framework* di riferimento del

⁴⁵ Il campione di analisi comprende le seguenti testate giornalistiche: (stampa) *Corriere della sera, la Repubblica, La Stampa, Il Messaggero, Il Gazzettino, QN Il Resto del Carlino, Avvenire, Il Sole 24 ore*; (radio) RTL 102.5, RDS, Radio Deejay, Radio Italia, Radio 105, Radio Rai 1; (TV) Rai 1, Rai 2, Rai 3, Rete 4, Canale 5, Italia 1, La7, TV 2000; (Internet) Corriere.it, repubblica.it, ilmessaggero.it, TGCOM24, Fanpage, ilfattoquotidiano, liberoquotidiano, Leggo; (Twitter) @corriere.it, @fattoquotidiano, @ilpost, @ilsole24ore, @Gazzetta.it, @repubblica.it, @HuffPostItalia, @Fanpage.it (Azzalini e Padovani 2021, 16-18).

GMMP, queste percentuali attestano una sotto-rappresentazione delle donne sia rispetto agli uomini, che costituiscono il restante 74% delle persone nelle notizie, sia rispetto alle donne nel mondo reale, che in Italia, principale, ancorché non esclusivo, contesto di riferimento dell'informazione analizzata, costituiscono il 51,3% della popolazione⁴⁶.

Nei media tradizionali, come nei nuovi media digitali, le donne sono presenti in misura superiore alla media in notizie di «società e giustizia», con valori rispettivamente pari al 33% e al 25%, congruenti con i dati internazionali, e in notizie di «politica e governo» (25% e 34%) o «criminalità e violenza» (31% e 29%), a differenza di quanto registrato per i media a livello globale, dove la presenza femminile in queste macro-aree tematiche è inferiore⁴⁷. In linea con questi risultati, nei media nazionali, le donne politiche attestano una presenza maggiore rispetto a quella osservata su scala globale (22% vs. 18%); così come le donne che fanno notizia come vittime, che, nei media tradizionali sono il 22% vs. il 4% degli uomini e nelle notizie di Internet il 26% vs. il 10% degli uomini⁴⁸ a fronte di dati internazionali inferiori e più bilanciati rispetto al genere (rispettivamente 14% vs. 15% e 12% vs. 6%). Inoltre, i risultati del GMMP Italia, a differenza dei risultati internazionali, attestano anche un'elevata percentuale di donne che fanno notizia come criminali, o sospettate di reato, da un lato, e come professioniste del mondo legale, avvocate, magistrato, e così via, dall'altro: ruoli che insieme a quello della vittima trovano la loro collocazione giornalistica tradizionale nella cronaca nera.

Un altro risultato che segna la distanza fra il quadro internazionale e quello nazionale riguarda la distribuzione di genere nelle diverse funzioni narrative delle persone fonte o oggetto di notizia. In particolare le funzioni dell'esperta/o e del/la portavoce. Quest'ultima registra una presenza femminile del 30% e 31%, a seconda che si considerino i media tradizionali o digitali, molto più elevata rispetto alle

⁴⁶ Dati ISTAT al 1° gennaio 2021; fonte: <https://www.tuttitalia.it/statistiche/popolazione-eta-sesso-stato-civile-2021/> (ultimo accesso 19 novembre 2021).

⁴⁷ Pari al 20% nei media tradizionali e al 21% nei nuovi media digitali, nel primo caso e al 24% e 27%, nel secondo (WACC 2020).

⁴⁸ Per Twitter i dati non sono rilevati.

percentuali registrate nelle edizioni precedenti, e anche a quella osservata a livello internazionale (rispettivamente 22% per i media tradizionali, 25% per Internet e 19% per Twitter). Viceversa, la funzione dell'esperta/o attesta quote femminili del 12% (media tradizionali) e 14% (media digitali), inferiori al 18% e 16% del 2015, e ai dati globali rispettivamente pari al 24% e 25%. Secondo Azzalini e Padovani (2021), il risultato pertinente la (relativamente) elevata quota delle portavoce va ricondotto all'aumento delle donne in due specifici ambiti: la politica e la *leadership* aziendale. Le politiche e le donne di affari rappresentano infatti rispettivamente il 59% e il 12% delle portavoce presenti nei media tradizionali, nell'ambito di un contesto nazionale che, con l'inizio della 18esima legislatura, ha visto raggiungere le parlamentari quota 35% (Senato della Repubblica 2018) e, con la "legge Golfo-Mosca" sulla parità di genere nei consigli di amministrazione delle società quotate adottata nel 2011 e con il Decreto del Presidente della Repubblica del 30 novembre 2012 n. 251 sulla rappresentanza nelle società partecipate pubbliche, ha visto un progressivo aumento della rappresentanza femminile nelle società, dal 7,4% del periodo pre-legge al 36,3% nel 2020 per le società quotate, e dall'11,2% al 28,4% per le partecipate pubbliche, con una svolta nel 2017, quando la soglia del 30% è stata superata per la prima volta. La scarsa visibilità delle esperte è invece ricondotta all'impatto dell'agenda Covid-19, che a livello internazionale, ha favorito la visibilità delle donne in questo ruolo, in particolare espresso dalle professioniste del settore medico-sanitario, mentre in Italia ha avuto un effetto contrario, riducendone sia la visibilità generale (dal 26% all'11%), sia la loro visibilità come esperte (dal 13% al 6%).

Per quanto riguarda le giornaliste, i risultati italiani si distanziano da quelli globali per una minor percentuale di conduttrici di giornali radio e TV (43% vs. 51%), in particolare per una bassa percentuale di conduttrici di giornali radio (21%), ma per una maggiore percentuale di *reporter* di notizie radio, stampa e TV (49% vs. 40%), attestando così un minor divario fra i due ruoli giornalistici per i media tradizionali. Per i nuovi media digitali, invece, i risultati nazionali rilevano un *gender-gap* maggiore, sebbene ridotto rispetto al 2015: le notizie firmate da giornaliste

sono il 30% su Internet e il 33% su Twitter (vs. rispettivamente il 13% e l'8% del 2015)⁴⁹.

Nonostante siano meno numerose, le giornaliste dei nuovi media digitali, nello specifico Internet, l'unico per il quale è stato rilevato il dato, in proporzione, sono più attente alle donne rispetto sia ai loro colleghi maschi, sia alle colleghe dei media tradizionali: il 33% delle persone fonti, o oggetto, delle loro notizie sono donne, mentre le fonti femminili dei servizi firmati dai colleghi (maschi) del web sono solo il 24%, e quelle dei servizi delle colleghe di radio stampa e TV il 22%. Le percentuali relative a Internet sono in linea con quelle registrate a livello globale, pari rispettivamente al 34% e al 25%, e, su scala globale, sono interpretate come indicative di una maggior propensione delle giornaliste a parlare di- o far parlare le donne. Questa tendenza è infatti attestata per il web anche nel 2015, e per i media tradizionali dal 1995. I media tradizionali italiani, invece, non confermano questa tendenza, riportando evidenze instabili di anno in anno. I dati, disponibili dal 2005, mostrano che nel 2005 e nel 2010 l'attenzione delle giornaliste per le donne è superiore a quella dei colleghi maschi, in modo contenuto nel 2005 (15% vs. 12%) maggiore nel 2010 (28% vs. 14%), ma non nel 2015 (21% vs. 22%) e, non nel 2020 (22% vs. 24%).

Una maggiore, ancorché contenuta, attenzione per le questioni di genere nell'informazione digitale rispetto a quella di radio, stampa e TV, nel panorama mediatico italiano, è attestata anche da altri due set di dati. Il primo riguarda le pratiche giornalistiche e il secondo il linguaggio giornalistico. Le pratiche giornalistiche indagano quattro variabili: l'attenzione per l'uguaglianza/disuguaglianza di genere, la trattazione o citazioni di politiche o normative per le pari opportunità e i diritti umani, la centralità narrativa delle donne, come singole o gruppo sociale, nelle notizie, e gli stereotipi di genere, con particolare attenzione

⁴⁹ Il confronto con i dati internazionali non è possibile perché non sono riportati nel rapporto globale del GMMP 2020 (WACC 2020).

agli elementi di sfida e innovazione⁵⁰. Le notizie che mettono in rilievo questioni relative alla parità (o alle disparità) fra donne e uomini e quelle che menzionano politiche o normative per le pari opportunità o i diritti umani sono poche, anche rispetto al dato internazionale (7%), ma proporzionalmente più presenti nei media digitali rispetto ai media tradizionali: 3% vs. 1%. Poche sono le notizie che sfidano gli stereotipi di genere, ma anche in questo caso sono proporzionalmente più frequenti online che nell'informazione di radio, stampa o TV, attestando una percentuale del 4%, in linea con il risultato internazionale, nel primo caso, e una percentuale più bassa, anche rispetto al dato globale (3%), nel secondo caso (1%). Più numerose sono invece le notizie focalizzate sulle donne, che raggiungono l'11% nei media tradizionali, una percentuale quasi doppia rispetto a quella globale (6%), e ben il 21% nei media digitali (vs. il 9% registrato a livello internazionale).

Per quanto riguarda il linguaggio giornalistico, l'uso della lingua italiana nel nominare le fonti è analizzato da tre domande di ricerca che, per la prima volta nel 2020, sono state inserite nella sezione metodologica dedicata a «special questions» che ogni paese ha potuto scegliere in base all'attualità nazionale di specifiche tematiche di genere. Le tre domande di ricerca sono state costruite in modo da verificare se la persona era presentata: 1) con nome e cognome, 2) con titolo professionale o politico, e 3) se, nel caso di risposta affermativa alle prime due domande, il genere grammaticale del titolo era coerente con il genere semantico di riferimento. I risultati evidenziano profili diversi a seconda del tipo di media.

Nei media tradizionali, le donne vengono presentate con nome e cognome meno degli uomini (76% vs. 82%) con alcune differenze tra i diversi media: nella carta stampata le proporzioni sono simili, ma le percentuali sono più alte (84% vs. 90%); in TV le proporzioni sono simili, ma le percentuali sono più basse (61% vs. 73%); per la radio le proporzioni sono invertite, con il 62% delle donne presentate per nome e cognome vs. 52% degli uomini. Nei media digitali, invece le donne

⁵⁰ La metodologia del GMMP per la rilevazione degli stereotipi di genere ha subito diversi cambiamenti nelle varie edizioni della ricerca: inizialmente focalizzata su elementi di rinforzo degli stereotipi, nel corso del tempo è stata modificata per rilevare anche elementi innovativi, di sfida agli stereotipi, su cui è totalmente concentrata la metodologia adottata nell'edizione 2020.

vengono presentate con nome e cognome più degli uomini (80% vs. 70%). In particolare, Twitter anche se nel complesso ha un linguaggio meno attento a identificare la persona con nome e cognome, probabilmente per via delle esigenze di sintesi proprie di questo mezzo, identifica le donne con nome e cognome in una proporzione doppia rispetto agli uomini (69% vs. 31%), mentre Internet ha un profilo molto simile alla carta stampata, con una percentuale di donne nominate per nome e cognome pari all'83% vs. l'89% degli uomini.

Per quanto riguarda l'identificazione professionale, politica o istituzionale, nei media tradizionali le donne sono presentate meno spesso degli uomini con un titolo (54% vs. 72%), ma in proporzioni diverse a seconda del tipo di media: la stampa registra un divario di genere più basso, con il 64% dei soggetti femminili presentati con un titolo professionale o istituzionale vs. il 72% degli uomini; mentre sia la radio che la TV registrano un divario più alto, con, rispettivamente, il 30% e il 39% donne vs. il 73% e il 72% uomini "titolati". I media digitali presentano due profili diversi tra loro: Internet ha un profilo simile a quello della stampa, attestando una frequenza (relativa) d'uso del titolo professionale per gli uomini pari al 54% vs. il 49% dei casi per le donne. Twitter, pur ricorrendo poco al titolo professionale, lo usa più spesso per le donne che per gli uomini: 31% vs. 22%.

Infine, il genere grammaticale risulta coerente con il genere semantico nella maggior parte dei casi sia per i media tradizionali che per i media digitali: con una frequenza del 100% nelle notizie radiofoniche e in quelle postate su Twitter, del 97% nelle notizie pubblicate da Internet e del 77% in quelle pubblicate dalla stampa, e del 75% per quelle TV. Tutti i casi di incoerenza riguardano le donne che, a differenza degli uomini, sono interessate dalla tendenza a essere nominate con la forma maschile, di cui ho trattato nel Capitolo 2 (§ 2.3.2).

3.2.2 L'informazione delle donne

Nell'ambito della ricerca su media e genere, il tema dell'informazione delle donne è da sempre indagato parallelamente, o in modo integrato, al tema della rappresentazione delle donne nell'informazione, nella duplice prospettiva

dell'analisi della condizione lavorativa delle giornaliste, da un lato, e del loro ruolo rispetto a una costruzione della realtà più inclusiva e egualitaria fra donne e uomini (cfr. § 3.1). Questo filone di ricerca a livello internazionale è abbastanza fecondo, fornendo analisi qualitative e quantitative che consentono di misurare e valutare, anche su ampia scala, sia l'avanzamento delle donne nelle professioni dell'informazione (cfr. Byerly 2011), sia il loro ruolo nella produzione di notizie *gender-fair* (cfr. Zoch e van Slyke Turk 1998; Humprecht e Esser 2017); a livello nazionale, invece, la produzione scientifica è più limitata, soprattutto povera di analisi quantitative. I principali studi che consentono una breve retrospettiva storica e una tematizzazione attuale della questione sono quelli realizzati o coordinati da Milly Buonanno, che forniscono analisi del panorama italiano, anche in relazione alle evidenze empiriche e alle riflessioni emergenti in ambito internazionale.

La prima fase di femminilizzazione del giornalismo italiano viene tradizionalmente collocata negli anni Settanta del Novecento. Nel 1978 le giornaliste costituiscono il 10% dei professionisti (Buonanno 2005, 22) e possono già distinguersi in due generazioni, «*le grandi emancipate* (o le pioniere)», entrate nelle redazioni negli anni Cinquanta e Sessanta, e «*le politiche* (o le innovatrici) », il cui ingresso – nella seconda metà degli anni settanta – ha coinciso con l'inizio della fase di accesso allargato per le donne» (Buonanno 2005, 33). Per quanto diverse fra loro, entrambe queste generazioni ricoprono una posizione sostanzialmente marginale nel giornalismo dell'epoca, non solo perché costituiscono una minoranza nelle redazioni, ma anche perché sono escluse da contrattazioni adeguate (spesso svolgono lavoro su base volontaria), non hanno salari congrui e reali opportunità di carriera, sono escluse dai luoghi decisionali e dall'accesso ad ambienti, personaggi e fonti influenti e, infine, sono confinate alla redazione di notizie di cultura, spettacolo e cronaca varia, con scarsa probabilità di fare le inviate all'estero, di intervistare leader, così come di commentare le notizie e avere spazi di opinione. Questa marginalizzazione non si traduce in un moto di ribellione, bensì in reazioni contraddittorie che mescolano sentimenti di soddisfazione per il piacere di fare un

lavoro utile alla democrazia, o semplicemente che piace, a sentimenti di insoddisfazione prevalentemente per la forte competizione interna alle redazioni, e per la marginalizzazione subita in un ambiente maschile e maschilista. Queste reazioni contraddittorie prefigurano tuttavia uno spazio di innovazione della professione e dell'informazione:

proprio perché marginali, infatti, l'ideologia del mestiere su di loro ha agito meno e [...] ha costituito invece le precondizioni efficaci per la presa di distanze da un modello che si avverava ostile o remoto, e il conseguente avvio di processi di elaborazione innovativa (Buonanno 2005, 65).

Già nel 1976 nasce il *Coordinamento delle giornaliste*, che, sollecitato dalla seconda ondata del femminismo, spinge parte delle giornaliste italiane a unirsi per denunciare la propria discriminazione professionale e per rivendicare un'informazione che, nei contenuti, dia più voce e più spazio alle donne (Buonanno 2005).

Gli anni Ottanta si caratterizzano per essere un periodo in cui il giornalismo diventa un ambiente professionale più accogliente per le donne, sotto diversi punti di vista. In primo luogo, la componente femminile cresce, raggiungendo nel 1990 il 19% tra i professionisti e il 36% tra i praticanti (Buonanno 2005, 22). In secondo luogo, le condizioni contrattuali migliorano, con una riduzione del lavoro su base volontaria. In terzo luogo, i meccanismi di selezione, sino ad allora prevalentemente basati su cooptazione, diventano più meritocratici. Infine, le giornaliste, precedentemente collocate nei quotidiani, soprattutto di partito, e nei periodici femminili o femministi, entrano nelle redazioni radiofoniche e televisive. Questo miglioramento delle condizioni favorisce l'ingresso di una nuova generazione di giornaliste, «le *neo-emancipate*» (Buonanno 2006, 84), che introduce un forte elemento di differenziazione rispetto alle generazioni precedenti. Si tratta infatti di una generazione non più motivata da ideali di democratizzazione o realizzazione espressiva personale, ma attratta dal giornalismo in quanto professione prestigiosa. Prevalentemente orientata da questa motivazione, la generazione delle *neo-emancipate*, da un lato, è disposta a

competere nel proprio ambiente di lavoro secondo le regole esistenti, senza alcuna volontà di cambiarle, dall'altro, a fronte della presa di coscienza di una obiettiva difficoltà ad accedere ai luoghi di potere, sposta le proprie traiettorie di carriera verso obiettivi di visibilità. In questo contesto, il duplice interesse per un giornalismo più inclusivo e paritario e un'informazione delle donne "differente", nel duplice senso che favorisca una maggiore visibilità delle donne come oggetto e fonti dell'informazione e che migliori complessivamente la qualità dell'informazione, rimane sullo sfondo come priorità prevalentemente della generazione delle *politiche*.

Gli anni Novanta, e il decennio successivo, si contraddistinguono per l'avvento di una quarta generazione di giornaliste, «le *ultime arrivate*» (Buonanno 2005, 103), le cui caratteristiche, oltre l'etichetta definitoria formulata da Buonanno, devono ancora essere identificate. Originariamente vengono riconosciute come un «arcipelago composito in cui c'è di tutto, nel bene e nel male» (Buonanno 2005, 103); a distanza di trent'anni non sono ancora state profilate, per via di una scarsa e discontinua attenzione della ricerca sulle donne nel giornalismo. Tuttavia, in questa sede, non è tanto rilevante identificarle, quanto annoverarle fra le quattro generazioni che profilano la componente femminile del giornalismo italiano a cavallo del Duemila. Un periodo contraddistinto da un'affermazione delle donne nelle redazioni di tutti i diversi tipi di media, dalla stampa alla radio, dalla televisione alle emergenti testate online. In termini numerici, la percentuale femminile cresce, anche se più lentamente rispetto al decennio precedente: nel 1996 le giornaliste tra i professionisti sono il 26% e tra i praticanti il 46%, nel 2002 sono rispettivamente il 42% e il 48% (Buonanno 2005, 22). Qualitativamente, il loro avanzamento si configura in termini di emersione dall'invisibilità. Sono questi gli anni in cui le giornaliste cominciano a essere visibili alla conduzione dei principali telegiornali nazionali, tradizionalmente visti da un pubblico ampio e generalista, e, dalla guerra nel Golfo in avanti, come inviate al fronte, entrambi ruoli autorevoli e storicamente ricoperti dai colleghi maschi. A questa ampia visibilità non corrisponde però un avanzamento nei luoghi di potere, che rimangono

appannaggio maschile, né una maggiore o migliore rappresentazione delle donne nei contenuti dell'informazione. Una condizione ambivalente efficacemente riassunta da Buonanno (2005) nella locuzione «visibilità senza potere» e che manifesta anche per l'Italia la fallacia del mito ottimistico americano, già sopra citato, secondo il quale il massiccio ingresso delle donne nella professione giornalistica avrebbe automaticamente rotto il "soffitto di cristallo", favorendo direttamente il raggiungimento di posizioni apicali per le donne e indirettamente una migliore e maggiore rappresentazione delle donne nei contenuti dell'informazione.

Queste riflessioni che segnano il passaggio dal 20esimo al 21esimo secolo, lasciano in eredità agli studi attuali un campo di ricerca, da un lato, reso più complesso dai cambiamenti che lo sviluppo delle nuove tecnologie portano sia nel mondo dell'informazione, sia nel mondo del lavoro più in generale, dall'altro, non adeguatamente attrezzato per verificare ipotesi diverse da quelle rivelatisi fallaci nel secolo scorso. In ambito internazionale, alcune ipotesi, su cui torno nel sottoparagrafo successivo (§ 3.3.2), sono verificate su ampie banche dati, analizzate con appropriati strumenti teorici, metodologici, e statistici. A livello nazionale, l'ultimo lavoro che offre un'analisi complessa del giornalismo italiano al femminile è pubblicato in un numero speciale della rivista *Problemi dell'informazione* (Buonanno 2015a), con una serie di contributi che riportano dati di ricerca sulla percentuale di donne nella professione giornalistica, anche ai livelli apicali, sulla loro visibilità nei contenuti dell'informazione, sulla visibilità delle donne, fonti e *newsmaker*, e permettono di riflettere sulla relazione fra i numerosi dati delle analisi empiriche presentate nel volume e i risultati delle (pochissime) ricerche realizzate nell'ultimo quinquennio.

Negli anni più recenti, il giornalismo italiano si contraddistingue per una stabilizzazione della componente femminile attorno alla soglia del 40% e un contestuale accesso delle donne ai luoghi di potere, seppur in proporzioni di una donna ogni tre uomini, alla direzione di qualche testata giornalistica, e di una donna ogni due uomini a capo di una redazione o di un servizio (Azzalini 2015a). Questa

femminilizzazione si riflette nei contenuti dei media, che addirittura sovra-rappresentano le donne alla conduzione dei telegiornali (Azzalini 2015b; Azzalini e Padovani 2021), ma non si riflette nei luoghi in cui il potere esercita più efficacemente il suo esercizio: nelle prime pagine dei quotidiani, dove le firme femminili sono poche (Buonanno 2015b), o alla radio (Perrotta 2015; Azzalini e Padovani 2021), e nei contenuti delle notizie, dove le donne oggetto o fonte di informazione sono una minoranza (Azzalini 2015b; CNOG 2015; Azzalini e Padovani 2021). Questi risultati vengono letti in modo comparativo da un mio contributo al sopracitato numero monografico di *Problemi dell'informazione* (Azzalini 2015b), che torna a riportare al centro delle riflessioni sul ruolo e il potere delle giornaliste nella costruzione di una realtà sociale più paritaria la questione della «consapevolezza del genere», brevemente accennata da Buonanno (2005), come prerequisito indispensabile per orientare le pratiche giornalistiche femminili, ma anche maschili, nella direzione di un'informazione più *gender-fair*. Quella consapevolezza mi sembrava allora, e mi sembra oggi, la caratteristica di un giornalismo, forse di nicchia (Buonanno 2015a, 438), ma innovativo per le attività di informazione, *advocacy* e formazione proposte; precisamente il giornalismo del blog *La 27esima ora* e l'impegno dell'associazione di giornaliste GiULiA.

La 27esima ora è una sezione online, ad accesso libero, del *Corriere della sera*, nato nel 2011 su iniziativa della vicedirettrice (attualmente vicaria) Barbara Stefanelli per dare spazio a tematiche femminili, fra pubblico e privato, tendenzialmente marginali sul quotidiano nazionale. GiULiA è un'associazione di giornaliste nata, nello stesso anno, con il duplice obiettivo di riportare al centro dell'informazione il reale e multiforme universo femminile e di rovesciare le logiche maschili, e maschiliste, delle redazioni italiane. A questa associazione va riconosciuto, fra l'altro, di aver rimesso al centro dell'agenda dell'informazione il tema della lingua che parla delle donne, con la pubblicazione più volte menzionata del volume che riprende e attualizza le riflessioni e le raccomandazioni di Alma Sabatini per un uso non sessista della lingua italiana (Robustelli 2014).

Per «consapevolezza del genere» intendevo, in quel contributo, «una coscienza del *male gaze* che domina l'informazione» (Azzalini 2015b, 474). A distanza di qualche anno, inframezzato da studi di linguistica, questo concetto mi sembra che possa ritenersi corrispondente alla metacompetenza linguistica invocata da Giusti (2016a, 2016 b) e Azzalini e Giusti (2019). Traslata sul piano mediale, e ristretta alle questioni di genere, questa «consapevolezza del genere» si riferisce alla conoscenza di come funzionano i meccanismi dell'informazione e come contribuiscono a coltivare l'immaginario collettivo. In modo più articolato, si tratta di una conoscenza sul modo in cui i media costruiscono la realtà sociale, sui fattori che intervengono nella direzione di promuovere o ostacolare la costruzione di una realtà inclusiva e paritaria e sul ruolo delle *policy*, aziendali e pubbliche, e delle parti interessate. Come i/le parlanti generalmente parlano senza riflettere sul ruolo della lingua nella costruzione della realtà sociale, incluse le identità di genere, né sulle caratteristiche del sistema grammaticale della lingua che parlano, o che scrivono, i giornalisti, e le giornaliste, spesso producono notizie senza riflettere sui processi mentali, socio-culturali e mediali che intervengono in questo processo di produzione dell'informazione e la possono orientare in direzione *gender-fair* o viceversa. A questi argomenti è dedicato il paragrafo successivo.

3.3 Il genere fra realtà e informazione

3.3.1 Il ruolo dei media nelle rappresentazioni di genere

L'idea del GMMP nasce nel febbraio del 1994, durante la *Women Empowering Communication Conference* di Bangkok, in preparazione alla *Quarta Conferenza mondiale sulle donne* convocata dalle Nazioni Unite a Pechino per il mese di settembre del 1995⁵¹ (d'ora innanzi *Conferenza di Pechino*). La prima ricerca nell'ambito del GMMP viene realizzata nel gennaio 1995 e i risultati vengono discussi al *Women's NGO Forum* durante i lavori di Pechino, contribuendo al

⁵¹ Fonte: <https://whomakesthenews.org/gmmp-background/>; ultimo accesso 19 novembre 2021.

confronto che ha portato alla redazione della *Dichiarazione e Piattaforma d'azione di Pechino* (ONU 1995) sottoscritta dai 189 stati membri delle Nazioni Unite⁵².

Dopo le conferenze che, a Città del Messico (1975), Copenaghen (1980) e Nairobi (1985), si erano prefissate come obiettivo il raggiungimento dell'uguaglianza come diritto fondamentale per le donne, la *Conferenza di Pechino* rappresenta un punto di svolta per due ragioni. La prima, e più generale, riguarda la focalizzazione sui diversi fattori culturali, economici, politici e sociali che ostacolano l'avanzamento effettivo delle pari opportunità, anche nei paesi dotati di un sistema legislativo che riconosce formalmente pari diritti a donne e uomini. La seconda, e più particolare, riguarda l'identificazione del settore dei media come uno dei dodici ambiti strategici per l'avanzamento sostanziale dell'uguaglianza di genere. Da Pechino in poi, la riflessione sulla relazione fra media e genere assume infatti una rilevanza internazionale, e contribuisce a orientare l'attenzione, sia della ricerca accademica, sia delle diverse parti interessate (aziende media, associazioni professionali, istituzioni, organizzazioni della società civile), sul potere che i media hanno, in termini di potenziale, di promuovere l'*empowerment* femminile:

the media have a great potential to promote the advancement of women and the equality of women and men by portraying women and men in a non-stereotypical, diverse and balanced manner, and by respecting the dignity and worth of the human person (ONU 1995, 13).

Questa dichiarazione è particolarmente rilevante per due ragioni: in primo luogo, perché richiama il ruolo dei media come *agency*, in secondo luogo, perché riconosce implicitamente che i media hanno un "potere" ambivalente.

Il ruolo dei media come *agency* è una questione complessa ed è fortemente intrecciata al tema del "potere" dei media. Questo tema è al centro della riflessione teorica e della ricerca sui media, sin dagli esordi della cultura di massa, all'inizio del Novecento e, nel corso del tempo, ha sovente polarizzato studiosi e studiosi, fra

⁵² Fonte: <https://www.unwomen.org/en/how-we-work/intergovernmental-support/world-conferences-on-women>; ultimo accesso 19 novembre 2021.

«apocalittici», da un lato, e «integrati», dall'altro, per riprendere una nota diade definitoria coniata da Umberto Eco (1964). In realtà sia la pessimistica posizione dei primi, persuasi del potere omologante dei media, e più in generale della cultura di massa, condizionata dalle esigenze del progresso capitalistico, sia l'ottimistica posizione dei secondi, convinti che i media contribuiscono alla democrazia attenuando barriere geografiche e sociali, «hanno in comune la stessa concezione del potere dei media, considerato da entrambe sostanzialmente illimitato» (Losito 1998, 31). Questa concezione del potere è stata progressivamente (parzialmente) abbandonata a favore di più (o meno) attente analisi degli effetti a breve termine, da un lato, e degli effetti a lungo termine, dall'altro. Nel primo caso, la finalità è quella di comprendere, tramite adeguati approcci teorici e metodologici, gli effetti specifici di circoscritte produzioni mediali su atteggiamenti e comportamenti del pubblico: per esempio, gli effetti di una campagna elettorale sulle intenzioni di voto. Questo filone di ricerca, che risale ai lavori di Lazarsfeld (1948) e Katz e Lazarsfeld (1955), dimostra in modo stabile e coerente nel tempo che i media hanno effetti limitati, ovvero mediati da specifiche condizioni psicologiche e sociali del pubblico. Nel secondo caso, l'obiettivo è più ambizioso, ed è quello di comprendere gli effetti cumulativi derivanti da un'esposizione continua ai contenuti dei media. Questo filone di ricerca, nell'ambito del quale si sono sviluppate teorie come quella della coltivazione (Gerbner 1978), già sopra menzionata, della spirale del silenzio (Noell-Neumann 1973, 1977, 1980/2002) e dell'*agenda setting* (Shaw 1979), si confronta da sempre con difficoltà teoriche e metodologiche dovute alla complessità dell'oggetto, che spesso è stato indagato a partire da ipotesi di ricerca poco convincenti e pratiche di analisi inadeguate (Losito 1998, 37-75).

Secondo Losito (1998), che seguo nel sintetizzare brevemente la questione in oggetto, una teoria funzionale a ricondurre il tema della questione del "potere" dei media entro un solido quadro teorico, sulla base del quale adottare adeguate pratiche di ricerca, è quella delle rappresentazioni sociali, che ho già menzionato, citando la ricerca di Aebischer (1985/1988) nel Capitolo 2 (§ 2.2.1). Questa teoria

consente, secondo il sociologo, di spostare l'analisi da una teoria dei media a una teoria dell'azione sociale, coniugando una prospettiva costruzionista con una prospettiva interazionista. In realtà, la teoria delle rappresentazioni sociali costituisce, come cerco di dimostrare in questa tesi, un *framework* funzionale a un'analisi multidisciplinare a più livelli: tenendo sullo sfondo una teoria dell'azione sociale, secondo la quale i media hanno un ruolo di *agency*, essa consente di analizzare le rappresentazioni di genere dentro e attraverso i media, in considerazione non solo della funzione cognitiva e sociale dei media ma anche della funzione cognitiva e sociale della lingua, nella fattispecie italiana, e delle caratteristiche di entrambi. Quanto al ruolo della lingua nelle rappresentazioni di genere e del "potere", nel senso del potenziale della grammatica italiana di contribuire a rendere visibili le donne in modo simmetrico rispetto agli uomini, e degli ostacoli prevalentemente rappresentati da una resistenza culturale a un uso non sessista, ho ampiamente discusso nel Capitolo 2 (§ 2.3). Qui di seguito mi concentrerò sull'ambivalente "potere" dei media.

Il ruolo dei media come *agency* può essere definito, a livello generale, come ruolo di agenzia di socializzazione: i media, insieme ad altre agenzie, quali la famiglia, la scuola, il gruppo dei pari, e così via, hanno un'influenza sui processi di socializzazione, «ovvero sui processi in virtù dei quali ciascuno di noi apprende prerogative e aspettative di ruolo, norme sociali, mete culturali, modelli di comportamento, valori, competenze comunicative e così via» (Losito 1998, 72). Il rapido sviluppo delle TIC su scala globale ha reso accessibili a donne e uomini informazioni altrimenti irraggiungibili, non solo in termini di conoscenza del mondo, ma anche in termini di identità, ruoli e relazioni sociali, modelli di comportamento e stili di vita, culture, lingue e valori a cui ispirarsi. E queste "informazioni" contribuiscono alla costruzione dell'immaginario collettivo, orientando gli individui. Come scrive Saveria Capecchi (2002, 112):

da un lato l'immaginario si costruisce a partire dagli stimoli, dalle opportunità e dalle risorse presenti nella realtà, dall'altro nella quotidianità i "mondi possibili" prefigurati dall'immaginario risultano indispensabili per

fare scoperte, progettare ed effettuare scelte di vita. A seconda delle esperienze e delle conoscenze personali accumulate fin dall'infanzia, ciascuna e ciascuno di noi sviluppa un proprio orizzonte immaginario, mutevole nel tempo, che lo/la guida nei progetti e nelle scelte future.

E ancora: «i media contribuiscono [...] a creare un immaginario collettivo diffondendo rappresentazioni sociali [...] valori, modelli di comportamento, stili di vita che “segnano” un'epoca, con cui non si può non fare i conti» (Capecchi 2002, 112). Si tratta, come sottolineato da più parti (Losito 1998; Capecchi 2002; Grossi e Ruspini 2007), di una partecipazione, e non di una “direzione” imposta dall'alto, che si caratterizza per essere sempre situata, nello spazio e nel tempo, e negoziata, a livello intrasoggettivo e intersoggettivo, sulla base delle diverse caratteristiche individuali e sociali del “pubblico”.

I media hanno poi anche un ruolo di *agency* più particolare, come attori, anche in questo caso non unici (agiscono insieme ad altre istituzioni) e non univoci (agiscono in varie direzioni), del cambiamento sociale, o viceversa del rafforzamento dello *status quo*, perché «propongono “immagini”, “concezioni”, “rappresentazioni” della realtà che possono influire sui processi in virtù dei quali ciascun membro del pubblico costruisce il proprio sapere sul mondo» (Losito 1998, 145).

Le immagini, concezioni, rappresentazioni della realtà veicolate dai media possono essere considerate come rappresentazioni sociali secondo la nozione originariamente proposta da Moscovici (1969), riportata di seguito:

sistemi cognitivi che hanno una logica ed un linguaggio particolari, una struttura di implicazioni relative sia ai valori sia ai concetti, uno stile di discorso che è loro proprio. Noi non vediamo in esse unicamente “opinioni su”, “immagini di” e “atteggiamento verso”, ma delle “teorie”, delle “scienze” *sui generis* destinate alla scoperta del reale e a mettere in ordine in esso (Moscovici 1969, 10, trad. it. Losito 1998, 146).

Una rappresentazione sociale è un sistema di valori, di nozioni e di pratiche con una duplice vocazione. Innanzi tutto instaurare un ordine che dia agli individui la possibilità di orientarsi nell'ambiente sociale, materiale e di dominarlo. Poi assicurare la comunicazione tra i membri di una comunità offrendo a essi un codice per denominare e classificare in maniera univoca le componenti del mondo, della loro storia individuale e collettiva (Moscovici 1969, 11, trad. it. Losito 1998, 146).

Una rappresentazione sociale è dunque uno schema mentale in cui sono ordinatamente conservati, e progressivamente acquisiti, dati di conoscenza teorica, pratica ed etica, recepiti e archiviati secondo un preciso linguaggio che è al contempo strumento di codifica, decodifica e comunicazione intersoggettiva. La sua dimensione sociale non è accessoria, perché ogni rappresentazione sociale deriva dalla partecipazione delle- e dall'interazione fra- le diverse coscienze individuali, come si comprende dalla descrizione del suo processo di costruzione, che segue due fasi: l'oggettivazione e l'ancoraggio. La prima è la fase di selezione e organizzazione delle informazioni pertinenti l'oggetto della rappresentazione. Il presupposto di Moscovici è che nessuna coscienza, individuale o collettiva, sia *tabula rasa*, ma il risultato di una storia. Per cui la fase di oggettivazione non comporta una ricezione di "informazioni pure", nel senso di dati di una realtà oggettiva, ma una assunzione selettiva di informazioni preesistenti che vengono scomposte, decontestualizzate e ricostruite dando origine a una nuova elaborazione che si presenta come tendenzialmente coerente e sistematica. La seconda è la fase di naturalizzazione e reificazione di questo primo schema figurativo essenziale, tramite il suo ancoraggio a un sistema simbolico, cognitivo e normativo, anch'esso preesistente, con cui lo schema si confronta e interagisce. In questa fase la rappresentazione sociale si presenta come uno strumento di classificazione, gerarchica, della realtà e, al contempo, di orientamento al comportamento sociale.

I media hanno un ruolo cruciale nella costruzione delle rappresentazioni sociali poiché forniscono o riproducono informazioni, le classificano in schemi organizzati sulla base di valori, «cioè, contribuiscono con le altre possibili fonti di

informazione a costituire il “campo” di una rappresentazione sociale» (Losito 1998, 150), e in questo modo possono incidere sugli atteggiamenti mediali. Questo spiega quello che ho definito, più sopra, l’ambivalente potere dei media: nella misura in cui diffondono rappresentazioni di genere stereotipate, sbilanciate, o addirittura degradanti, sedimentano un immaginario, e conseguenti atteggiamenti, che non supportano, o addirittura ostacolano, le pari opportunità, viceversa, nella misura in cui veicolano rappresentazioni di donne e uomini articolate e complesse, bilanciate e rispettose della dignità umana contribuiscono alla costruzione di una realtà democratica, paritaria, inclusiva che può sostenere i cambiamenti sociali nella direzione dell’uguaglianza di genere. Per questo, la *Dichiarazione e Piattaforma d’azione di Pechino* (ONU 1995), presuppone che i media possono promuovere il progresso delle donne e della parità di genere, a condizione però di rappresentarle in modo non stereotipato, bilanciato, dignitoso, e, sulla base di questo presupposto stabilisce due obiettivi strategici:

strategic objective j 1: Increase the participation and access of women to expression and decision-making in and through the media and new technologies of communication (ONU 1995, 100).

strategic objective j 2: Promote a balanced and non-stereotyped portrayal of women in the media (ONU 1995, 101).

3.3.2 *La resistenza dei mezzi d’informazione all’empowerment femminile*

Il processo di formazione delle rappresentazioni sociali è, come abbiamo visto, un processo cognitivo e sociale strutturato in due fasi durante le quali la ritenzione selettiva di informazioni pertinenti un determinato oggetto, per esempio identità, ruoli e relazioni di genere, viene rielaborata, a livello individuale e collettivo, e negoziata, intrasoggettivamente e intersoggettivamente, in un processo situato e mutevole nel corso tempo, di definizione e ridefinizione del significato dell’oggetto così da poterlo comprendere, interpretare e comunicare in modo “ordinato” entro un determinato contesto di riferimento. Questo processo, per quanto attiene la produzione mediale, implica che chi scrive una *fiction* o una notizia, da un lato, è,

più o meno consapevolmente, un soggetto attivo che rielabora le informazioni che possiede e le propone al pubblico, dall'altro lo fa in relazione al patrimonio di competenze che possiede e che condivide con l'ambiente di cui partecipa. Questo ambiente è sia il contesto generale, culturale e sociale, a cui il soggetto appartiene, sia il contesto produttivo in cui opera.

Nell'ambito dell'informazione, una rappresentazione sociale è dunque il risultato di un processo complesso, e dinamico, che riguarda sia la cultura giornalistica, in senso individuale e collettivo, sia la cultura di un paese. Partendo da questo presupposto è possibile individuare alcuni fattori che sono stati indagati per spiegare la resistenza dei mezzi di informazione a una rappresentazione di genere bilanciata e non stereotipata. Accantonata l'ipotesi semplicistica della riflessività dei media e della teoria della coltivazione (Tuchman et al. 1978; Gerbner 1978), per cui la sotto-rappresentazione e la marginalizzazione delle donne nei contenuti dei media è ascrivibile a una più generale resistenza dei media a rappresentare i mutamenti sociali, per via del loro conservatorismo "strutturale" e del loro potere unilaterale, dalla seconda metà degli anni Novanta del secolo scorso, diverse ricerche dimostrano che la rappresentazione delle donne, e degli uomini, nei contenuti dell'informazione dipende da una serie di fattori sociali e mediali che interagiscono in modo complesso e dinamico fra loro.

Il primo fattore, che è stato ampiamente indagato riguarda le pratiche giornalistiche. Il presupposto da cui partono le prime ricerche, probabilmente ancora parzialmente influenzato dall'idea di un potere unilaterale dei media, è che la sotto-rappresentazione e marginalizzazione delle donne nell'informazione sia ascrivibile almeno in parte a una cultura giornalistica presidiata dagli uomini. Zoch e van Slyke Turk (1998), per esempio, indagano, attraverso un'analisi di tre quotidiani distribuiti nel sud degli Stati Uniti fra il 1986 e il 1996, quali sono le componenti della cultura *male-oriented* del giornalismo: i criteri giornalistici di selezione delle fonti prediligono fonti ufficiali o istituzionali, che, nella maggior parte dei casi, sono maschili, e questi criteri sono condivisi a livello redazionale da giornalisti e giornaliste, le quali si differenziano dai colleghi solo per una maggior

tendenza a citare indirettamente fonti femminili (in linea con la tendenza rilevata dal GMMP e più sopra citata). Dunque la selezione delle fonti non è soltanto una questione di giornalisti maschi, ma di cultura maschile condivisa anche dalle giornaliste. Questa ipotesi è confermata da uno studio più recente (Hanitzsch e Hanush 2012) che analizza le concezioni di ruolo di giornaliste e giornalisti in diciotto diversi paesi del mondo, dimostrando una sostanziale similitudine di genere nella percezione e nell'esercizio della professione, su scala transnazionale, anche in contesti socioculturali in cui le donne hanno raggiunto un certo livello di *empowerment*.

Un secondo fattore messo in evidenza più di recente riguarda lo stato di avanzamento delle pari opportunità in un paese, in particolare il livello di partecipazione delle donne alla *governance* politica ed economica. Mediante un'analisi longitudinale su tredici quotidiani americani dal 1983 al 2008, Shor et al. (2015) dimostrano, per esempio, che le quote di donne tra parlamentari e top manager sono predittori dell'uguaglianza di genere nell'informazione più significativi del numero e della posizione delle giornaliste in una redazione, e concludono che la persistenza di differenze di genere nel *coverage* mediatico è più collegato allo stato di avanzamento delle pari opportunità in un paese che alla cultura e alla pratica giornalistica.

Nessuna delle ipotesi brevemente riassunte spiega i risultati di alcune ricerche che dimostrano come il processo di mediatizzazione delle donne in posizioni apicali, in particolare le politiche, tenda a sotto-rappresentarle rispetto alla loro rappresentanza reale nella società, a rappresentarle in modo stereotipato e, talvolta, anche a banalizzarle (Campus 2010; Cameron e Shaw 2016; Ross et al. 2020). Per esempio, l'ultima edizione del GMMP Italia ha rilevato una percentuale di politiche del 22% a fronte di una presenza femminile nel Parlamento italiano pari al 35% (Azzalini e Padovani 2021). Un altro caso esemplare registrato dal GMMP, nella terza edizione (WACC 2005), è quello del Ruanda: a fronte di una componente femminile pari al 51% nell'Assemblea parlamentare, le donne politiche

registravano nel 2005 una visibilità pari soltanto al 13% nell'informazione di radio, stampa e TV del loro paese.

Un'ipotesi avanzata per spiegare questo fenomeno è che vi sia una correlazione positiva fra la rappresentanza mediale delle donne e lo *status* che esse godono come gruppo sociale in una società:

In most societies women are still assumed to have less status than men. Hence their views are regarded as less important. These cultural assumptions link with journalistic practice so as to privilege the male (Portraying Politics Project Partners 2006, 13).

Si tratta di un'affermazione che suggerisce un'ipotesi di ricerca che nella pubblicazione appena citata non viene esplorata, ma è al centro di una recente analisi di Humprecht e Esser (2017) focalizzata sulle notizie politiche di quarantotto quotidiani online, di sei diverse tipologie (popolari, pubblici, privati, corporativi, indipendenti e basati sul web), di sei paesi occidentali (Francia, Germania, Gran Bretagna, Italia, Svizzera e Stati Uniti), per sei settimane dal mese di ottobre del 2012 al mese di febbraio 2013. Humprecht e Esser (2017) provano a indagare la correlazione fra le pratiche giornalistiche, la cultura di genere di un paese e il livello di avanzamento delle donne. Gli autori correlano un indicatore culturale – l'indice di mascolinità nella classificazione di Hofstede (2001)⁵³ – a un indicatore dell'uguaglianza di genere – il Global Gender Gap Index (GGGI) del World Economic Forum (WEF 2015) – ai risultati di un'analisi del contenuto delle quarantotto testate giornalistiche online selezionate. La conclusione a cui giungono è che le *routine* giornalistiche sembrano altamente influenzate sia dal livello di mascolinità di un paese (più il livello è elevato, meno donne presenziano nei media), sia dallo stato

⁵³ Lo psicologo e antropologo olandese Hofstede propone un modello di classificazione culturale secondo un numero limitato di dimensioni, una delle quali riguarda appunto la distanza delle aspettative sociali per donne e uomini, indicando come valori stereotipicamente maschili l'assertività, l'ambizione, il potere e il materialismo, e come valori stereotipicamente femminili l'enfasi data alle relazioni umane. Le culture in posizione alta sulla scala della mascolinità tendono ad avere differenze più rilevanti tra i sessi e tendono a essere più competitive e ambiziose. Quelle con punteggi bassi mostrano meno differenze tra i sessi e danno un valore maggiore allo sviluppo delle relazioni umane (Hofstede 2011).

di avanzamento delle donne in quel paese (più l'uguaglianza di genere è avanzata in un paese più le donne sono incluse nelle notizie politiche online). Per esempio, l'Italia che ha un indice di mascolinità pari a +0,78 (Hoftede 2001, cit. in Humprecht e Esser 2017) e un GGGI pari a -1,57 (WEF 2015, cit. in Humprecht e Esser 2017) attesta una presenza femminile (su un totale di N=276 notizie analizzate) pari al 16%, inferiore a quella del 24% attestata dalla Francia (su un totale di N=279 notizie), che ha un indice di mascolinità inferiore e un GGG superiore, rispettivamente pari a -2,13 e +0,15 (Humprecht e Esser 2017, 445-448). Uno studio più recente (Ross et al. 2020) conferma il ruolo delle pari opportunità nel favorire (o meno) un giornalismo *gender-fair*, in termini non tanto di parità raggiunta quanto di politiche di promozione per l'uguaglianza di genere. Questo studio dimostra infatti l'esistenza di una correlazione fra la rappresentazione delle donne politiche nei contenuti dell'informazione e il grado di impegno generale di un paese per la promozione delle pari opportunità, e anche il grado di bilanciamento fra donne e uomini nelle redazioni, anche ai livelli apicali.

I dati raccolti e analizzati dalle ricerche sopra menzionate mettono in evidenza diversi fattori che ostacolano l'*empowerment* femminile in diversi contesti geo-politici e mediali, enfatizzando talora "cause" pertinenti condizioni sociali, talora "cause" pertinenti condizioni mediali e talora "cause" interagenti. Un tentativo di comprendere l'interazione, e l'influenza reciproca, fra questi, e altri elementi in gioco, è rappresentato dal recente, e ambizioso, progetto di ricerca *Comparing gender and media equality across the globe* (Djerf-Pierre e Esdröm 2020a), che si propone di analizzare qualità, cause e conseguenze delle (dis)uguaglianze di genere dentro e attraverso i mezzi di informazione del mondo, partendo da un'ampia sistematizzazione e analisi statistica di dati raccolti da importanti ricerche transnazionali, riguardanti sia il settore dei media, sia la società nel suo complesso.

Nell'ambito di questo progetto, Monika Djerf-Pierre (2020) analizza le possibili spiegazioni delle variazioni della (dis)uguaglianza di genere nei contenuti mediali partendo da due diversi approcci, quello della modernizzazione e quello del

«gendered media field». L'ipotesi del primo approccio è che il livello di uguaglianza di genere nei media dipenda dalla posizione delle donne in una società e dai processi di sviluppo della stessa (come dimostrano, tra altri, Shor et al. 2015; Humprecht e Esser 2017). L'ipotesi del secondo approccio è che siano le condizioni degli ambienti mediali a influenzare lo *status* delle donne nei media nelle diverse società (come dimostrano, tra altri, Zoch e van Slyke Turk 1998; Hanitzsch e Hanush 2012). L'autrice non assume questi approcci come mutualmente esclusivi fra loro, ma come prospettive in grado di identificare e soppesare diversamente fattori strutturali, culturali e di *agency* pertinenti, nel primo caso, la società, nel secondo caso, i media. Partendo da questi due assunti, Djerf-Pierre (2020) svolge un'innovativa serie di analisi statistiche che misurano la correlazione fra l'indicatore sintetico dell'uguaglianza di genere nei contenuti dell'informazione, GEM-I⁵⁴, messo a punto per 123 paesi al mondo nell'ambito del sopracitato progetto, da un lato, e due diversi set di variabili ascrivibili, l'uno al grado di modernizzazione di un paese, in termini di sviluppo economico, sociale, politico, culturale, di pari opportunità e avanzamento delle donne, l'altro allo *status* dei media di un paese, in termini di libertà di espressione, integrità, autonomia editoriale, livello di influenza, economica, sociale, e così via, e il grado di femminilizzazione del giornalismo, dall'altro. Le analisi svolte a livello comparativo sincronico, fra diversi paesi, e diacronico, fra diversi periodi di tempo per uno stesso paese, dimostrano, in generale, che i paesi in cui le donne hanno una posizione più alta nella società, più donne nel campo del giornalismo e più autonomia per i giornalisti, hanno anche più uguaglianza di genere nelle notizie.

⁵⁴ Il GEM-I (Gender Equality in Media-Index) si basa su 6 indicatori e i valori rilevati per questi indicatori in 5 diverse edizioni del GMMP (1995, 2000, 2005, 2010, 2015). Tre sono le categorie individuate, per ciascuna delle quali sono considerati due indicatori. Per la categoria "presenza" («presence») 1. donne e uomini oggetto o fonte dell'informazione e 2. donne e uomini reporter; per la categoria "argomento" («topics») 3. donne e uomini oggetto o fonte di notizie economiche, 4. donne e uomini oggetto o fonte di notizie di politica e governo; per la categoria "ruoli" («roles») 5. donne e uomini portavoce, 6. donne e uomini esperte/i. Per ogni paese il GEM-I è dato dal divario medio fra donne e uomini nei valori attestati per tutti i 6 indicatori: ovvero la somma del valore dei 6 indicatori relativi alle donne diviso 6 meno la somma dei valori dei 6 indicatori relativi agli uomini diviso 6. L'indice può quindi variare da un minimo di -100 (100% uomini per tutti i 6 indicatori) a +100 (100% donne per tutti i 6 indicatori). Il valore zero (0) indica un equilibrio 50:50 per tutti i 6 indicatori, dunque rappresenta la piena uguaglianza di genere (Djerf-Pierre e Esdröm 2020c, 67-68).

Tutti questi fattori brevemente passati in rassegna e che, in sintesi, riguardano le pratiche giornalistiche, il livello di eguaglianza di donne e uomini e il contesto socio-culturale di un paese, e la loro relazione con le rappresentazioni di genere nei contenuti dei media, sono importanti perché vengono attestati e, al contempo, si riflettono dentro e attraverso il linguaggio dell'informazione analizzato nelle tre ricerche empiriche offerte in questa tesi.

3.3.3 *Il ruolo delle politiche pubbliche e delle parti interessate*

Come abbiamo visto (§ 3.3.1), a livello internazionale la *Dichiarazione e Piattaforma d'azione di Pechino* (ONU 1995) rappresenta un documento programmatico fondamentale che pone al centro dell'agenda globale il settore "donne e media", dopo l'impegno programmatico, più generale, per il contrasto agli stereotipi di genere della *Convenzione sulla eliminazione di ogni forma di discriminazione nei confronti della donna* (ONU 1979) (cfr. § 3.2.1). L'inizio del 21esimo secolo sposta l'attenzione verso le TIC, sempre più diffuse e pervasive su scala globale, e verso il tema della violenza contro le donne. Numerose sono le iniziative del Consiglio d'Europa, che afferma in più di un documento, compresa la *Convenzione sulla prevenzione e la lotta contro violenza nei confronti delle donne e la violenza domestica* (COE 2011), il ruolo strategico dei media e delle TIC nel mettere in discussione gli stereotipi di genere la cui diffusione, rinforzando lo *status quo* della cultura patriarcale, non solo ostacola l'avanzamento delle pari opportunità ma contribuisce ad alimentare un immaginario collettivo che accetta, quando addirittura non legittima, la violenza contro le donne (cfr. COE 2016a, 27).

Il tema della violenza contro le donne è recepito da diverse *policy* e *statement* internazionali sui media, in relazione, da un lato, al tema del *sexist hate speech* (cfr. COE 2016b), dall'altro, al problema delle molestie sessuali e delle violenze contro le giornaliste (cfr. ONU 2018). Contemporaneamente, la questione dell'uguaglianza di genere nei media, in particolare nell'informazione, ha subito un declino nell'agenda globale (Djerf-Pierre e Esdröm 2020, 47). Ne è testimonianza l'Agenda 2030 per lo Sviluppo Sostenibile, di cui l'uguaglianza di genere è un

obiettivo specifico (il numero 5), oltre che un prerequisito per l'avanzamento verso tutti gli altri sedici obiettivi, ma la parola "media" non è mai menzionata. Come notano Monika Djerf-Pierre e Maria Esdröm (2020), il secondo target dell'obiettivo 5 (5 B) può essere indirettamente collegato al settore dei media, prevedendo di «rafforzare l'utilizzo delle tecnologie abilitanti, in particolare le tecnologie dell'informazione e della comunicazione per promuovere l'emancipazione della donna» (ONU 2015, 18), tuttavia l'unico indicatore utilizzato per misurare lo sviluppo di questo obiettivo è la percentuale di persone che usano il telefono cellulare.

A part from the fact that mobile phones also can be a device of control, it is not quite the measure needed to monitor if gender equality is indeed progressing within the news media. In fact, the word "media" is not mentioned in the Agenda 2030 (Djerf-Pierre e Esdröm 2020, 47-48).

Il settore dei media è trascurato anche nei lavori di UN Women (2018) e UN Women (2019), in preparazione di Pechino +25, ovvero il *follow up* della *Dichiarazione e Piattaforma d'azione di Pechino* (ONU 1995), programmato per il 2020. Il primo riduce le aree strategiche per il miglioramento della condizione delle donne nel mondo da dodici a sei, escludendo "donne e media". Il secondo, pur affermando una generale preoccupazione per un arretramento della democrazia e dei diritti delle donne, non menziona né i media, né l'informazione. Il ruolo strategico dei media viene tuttavia ribadito nel rapporto finale di Pechino +25 (ONU 2020), ma, di nuovo, prevalentemente in relazione al tema della violenza contro le donne, e in particolare delle minacce alle giornaliste.

Since legislations and regulations that infringe on publishers' rights to publish freely can be considered a form of censorship or a violation of the freedom of speech, governments are generally reluctant to impose active measures to promote gender equality in the media. A mapping by the Council of Europe in 2020 indicates that only a limited number of European media-related legislation includes gender-equality provisions (Djerf-Pierre e Esdröm 2020b, 51).

Se nell'ultimo decennio, le organizzazioni internazionali, in particolare le Nazioni Unite, hanno parzialmente distolto l'attenzione sull'uguaglianza di genere dentro e attraverso i media, probabilmente anche a causa di una certa inerzia dei governi a legiferare in materia, per via del rischio di essere accusati di limitare la libertà di parola, altre istituzioni hanno invece rafforzato il loro impegno su questo fronte. L'UNESCO, per esempio, dal 2013 promuove la *Global Alliance on Media and Gender* (GAMAG), che attualmente unisce oltre 500 organizzazioni internazionali e singoli/e attivisti/e della società civile e studiosi/e afferenti a diverse università e centri di ricerca, nell'obiettivo comune di promuovere l'uguaglianza di genere dentro e attraverso i media e le TIC. Sul fronte sindacale, la International Federation of Journalists (IFJ), nata nel 1926, e attualmente accoglie più di 600.000 iscritte/i da 187 organizzazioni di 140 paesi del mondo, inclusa l'Italia, con la Federazione Nazionale della Stampa Italiana (FNSI), è da sempre attiva nel sostegno alle giornaliste e nella promozione di un'informazione egualitaria, rispettosa e dignitosa dell'universo femminile. A cura della IFJ è per esempio il decalogo per una corretta informazione sulla violenza contro le donne e la violenza domestica nell'informazione (IFJ s.d.). Sul fronte editoriale, possiamo ricordare tre grandi organizzazioni impegnate sulla parità di genere dei media. La *World Association for News Publishers* (WAN-IFRA), la *International Association for Women in Radio and Television* (IAWRT) e la *European Broadcasting Union* (EBU). La WAN-IFRA ha avviato il progetto *Women in the news* per aumentare la voce delle donne nelle notizie, nel 2020 ha pubblicato il rapporto *Trends in newsrooms 2020. Amplifying women's voices* (WAN-IFRA 2020), che fornisce esempi concreti di redazioni giornalistiche che si sono organizzate per migliorare la rappresentazione di genere nelle notizie, attraverso un miglioramento delle *routine* quotidiane e l'uso di dati per valutazioni interne. L'IAWRT produce rapporti periodici e linee guida finalizzate al miglioramento dei contenuti dell'informazione, ma anche alla sicurezza ai luoghi di lavoro (cfr. IAWRT 2020). L'EBU pubblica annualmente rapporti sulla parità di genere nelle organizzazioni dei media audiovisivi europee; nel 2019 ha pubblicato il rapporto *All things being equal: gender equality guidelines from public service media* (EBU 2019) dichiarando un forte impegno per la parità di genere

condiviso dai *broadcaster* aderenti all'unione, fra i quali c'è anche la Rai, e raccogliendo esempi di buone pratiche.

Per quanto riguarda l'Italia, l'uguaglianza di genere nei media non è materia direttamente normata a livello legislativo. Esistono però due testi normativi che vi fanno indirettamente riferimento. Il primo è il *Testo unico della radiotelevisione* (Decreto Legislativo 31 luglio 2005 n. 177), che, nell'articolo 4, sui «Principi generali del sistema radiotelevisivo a garanzia degli utenti», vieta in generale, e fra altre, le trasmissioni che «inducono ad atteggiamenti di intolleranza basati su differenze di razza, sesso, religione o nazionalità» e, garantisce «la diffusione di trasmissioni pubblicitarie e di televendite [...] che non evocino discriminazioni di razza, sesso e nazionalità» (D.L. 177/2005, 8). Quest'ultima garanzia viene ripresa, sotto forma di divieto, dall'articolo 40, in materia di «Disposizione sulle televendite», vietando «la televendita che vilipenda la dignità umana, comporti discriminazioni di razza, sesso o nazionalità» (D.L. 177/2005, 21). Lo stesso documento, nell'articolo 36 sulle «Trasmissioni transfrontaliere», dispone la sospensione di trasmissioni televisive provenienti da Stati dell'Unione europea e che contengano «incitamento all'odio basato su differenze di razza, sesso, religione o nazionalità» (D.L. 177/2005, 19). Un secondo documento è la Legge 23 novembre 2012 n. 215, che, contestualmente alla regolamentazione della rappresentanza di genere dei soggetti politici, nelle amministrazioni locali e nelle commissioni di concorso nelle pubbliche amministrazioni, introduce anche la cosiddetta *par condicio*⁵⁵ di genere nei media con una modifica della Legge 22 febbraio 2000 n. 28 (nota come legge sulla *par condicio*), che regola il pluralismo delle liste candidate alle elezioni, politiche e amministrative. La L. 215/2012 introduce infatti nell'articolo 1 della L. 28/2000 un comma che prevede che «i mezzi di informazione, nell'ambito delle trasmissioni per la comunicazione politica, sono tenuti al rispetto dei principi di cui all'articolo 51, primo comma, della Costituzione, per la promozione delle pari opportunità tra uomini e donne» (L. 215/2012, 2).

⁵⁵ Per *par condicio* si intende una condizione di accesso ai media paritaria per i diversi soggetti politici; per un approfondimento cfr. Bettinelli 1995.

Sulla base di questi documenti, l’Autorità per le Garanzie nelle Comunicazioni (AGCOM), in quanto organo di garanzia per il settore dei media elettronici e dell’audiovisivo, svolge un’attività di controllo, attraverso un monitoraggio permanente dei contenuti delle principali trasmissioni radiofoniche e televisive nazionali, che raccoglie anche dati sulla presenza delle donne in radio e in televisione, e il recepimento di segnalazioni da parte del pubblico, che può rivolgersi all’autorità per denunciare trasmissioni lesive della dignità femminile, o di altri gruppi sociali. L’AGCOM ha un potere sanzionatorio che esercita però generalmente soltanto in caso di gravi violazioni segnalate dagli/dalle utenti. Più controverso è il tema della rappresentazione bilanciata delle donne nei contenuti dei media, perché non c’è nessuno riferimento normativo che preveda un bilanciamento della rappresentanza femminile, e tanto meno un regolamento che stabilisca delle “quote” di genere. L’unica eccezione è rappresentata dalla sopra menzionata L. 215/2012, che comunque è circoscritta alla comunicazione politica, ed è una normativa stringente soltanto durante le campagne elettorali e per alcuni tipi di programma, come le tribune politiche per le quali è previsto l’obbligo di una partecipazione paritaria di candidate e candidati.

Per quanto riguarda gli operatori dei media, la Rai, come già ricordato più sopra, essendo concessionaria del servizio pubblico radio-televisivo, firma periodicamente un contratto di servizio con il Ministero dello Sviluppo Economico, che, dal 2012, prevede non soltanto un impegno, in termini di principi, a un’offerta equilibrata, rispettosa e non stereotipata rispetto ai genere, ma anche un’attività di monitoraggio che renda conto di quanto questi doveri siano ottemperati. Oltre al contratto di servizio (MISE 2011, 2018) l’azienda è dotata di altri strumenti interni, nessuno dei quali però specificamente rivolti a un uso della lingua italiana paritaria e inclusiva. Il *Codice etico* nella sua più recente formulazione prevede, fra gli altri obiettivi,

la valorizzazione della rappresentazione reale e non stereotipata della molteplicità dei ruoli del mondo femminile nel pieno rispetto della dignità culturale e professionale delle donne, anche al fine di contribuire alla

rimozione degli ostacoli che di fatto limitano le pari opportunità [...] i temi dei diritti civili, della solidarietà, della condizione femminile, delle pari opportunità e dell'integrazione (Rai 2020, 3-4).

Le pari opportunità sono citate come tematica a garanzia di un'offerta di qualità (Rai 2020, 22), e come principio per la selezione e la promozione delle risorse umane (Rai 2020, 12-13). L'insieme di tutte queste *policy* interne non ha portato i contenuti trasmessi dalla Rai a essere pienamente egualitari e rappresentativi del diversificato universo femminile che si muove fuori dallo schermo televisivo. I risultati del monitoraggio televisivo più recente, relativo all'anno 2020, mostrano infatti che persistono alcune asimmetrie di genere, ma l'offerta Rai, valutata attraverso cinque indicatori quali-quantitativi, sintetizzati in un «indice di correttezza della rappresentazione della figura femminile» con valori inclusi fra 0 e 10, raggiunge un valore molto elevato, pari a 8,23 punti (IZI et al. 2021, 17).

Venendo ora alle altre principali aziende media nazionali, molte hanno politiche di inclusione e pari opportunità nell'ambito della selezione e gestione delle risorse umane, ma nessuna politica per l'uguaglianza di genere nei contenuti (EIGE 2013). Un impegno in questo senso è testimoniato invece da alcune iniziative nel mondo dell'associazionismo femminile e/o femminista e del giornalismo.

Per quanto riguarda il primo, il movimento femminista giunge tardivamente a occuparsi di uguaglianza di genere nei media (Dinelli 2011). Come più sopra ricordato (§ 3.2), l'anno 2009 rappresenta un momento di svolta nell'orientare l'attenzione di neonate o vecchie associazioni femministe verso il sessismo nei media, in particolare nei contenuti dell'offerta televisiva. Questo fermento, oltre a portare all'*Appello donne e media*, grazie al quale per la prima volta il contratto di servizio stipulato dalla Rai con il Ministero dello Sviluppo Economico (MISE 2011) include una serie di articoli che stabiliscono in via di principio l'obbligo per l'azienda a una rappresentazione bilanciata, non stereotipata e dignitosa delle donne, sollecita diverse iniziative. All'inizio del 2010 nasce *Pari o Dispare*, un'associazione con presidente onoraria Emma Bonino, che nel 2008, da ministra per il commercio

internazionale, aveva organizzato il convegno dal titolo *Donne, innovazione e crescita. Donne italiane negli stereotipi: vita reale, comunicazione e fiction*, «legando la scarsa presenza delle donne in ruoli di responsabilità al modo in cui sono presentate dal sistema dei media» (Dinelli 2011, 68). L'associazione pone al centro delle sue attività di *advocacy* la questione dei media, promuovendo diverse iniziative, anche di lobby presso le istituzioni, le organizzazioni dei media e le agenzie pubblicitarie, sia online che offline⁵⁶. L'UDI (Unione Donne in Italia), storica associazione femminista, nel 2010 avvia la campagna *Immagini amiche*, basandosi sulla Risoluzione del Parlamento europeo (2008) sull'*Impatto del marketing e della pubblicità sulla parità tra donne e uomini*. Da allora, e fino al 2017, la campagna premia produzioni medial, pubblicitarie e non, ritenute meritevoli per la rappresentazione mediale di modelli identitari, di ruoli e di relazioni di genere innovative, coinvolgendo sia le organizzazioni dei media, in particolare quelle televisive, sia le aziende pubblicitarie. L'iniziativa si interrompe probabilmente a seguito di uno spostamento di interesse dell'UDI, così come di quasi tutte le associazioni femminili e/o femministe italiane, verso il tema della violenza contro le donne. Fra le associazioni che mantengono invece costante un interesse specifico per l'uguaglianza di genere dei media si può annoverare *DonneInQuota*, nata nel 2006, con lo scopo di monitorare e promuovere la rappresentanza politica delle donne e una corretta e bilanciata rappresentazione delle donne sia nei contenuti, sia nell'industria dei media⁵⁷. A questa associazione si deve un'ampia attività di lobby per la modifica della L. 28/2000 al fine di garantire la *par condicio* di genere, nonché una costante attività di segnalazione alle autorità competenti (AGCOM, Dipartimento Pari opportunità, Commissione di vigilanza Rai) di discriminazioni sessiste, sia nei contenuti medial, sia nei processi di nomina di vertici istituzionali quali la Rai o l'AGCOM.

Per quanto riguarda il mondo del giornalismo, oltre alle iniziative promosse dalle giornaliste, menzionate nel sotto-paragrafo precedente (§ 3.2.2), l'impegno

⁵⁶ Fonte: <https://pariodispare.org>; ultimo accesso 19 novembre 2021.

⁵⁷ Fonte: <https://www.donneinquota.org>; ultimo accesso 19 novembre 2021.

per le pari opportunità è costante e continuo nelle principali organizzazioni sindacali, come la FNSI (Federazione Nazionale della Stampa Italiana)⁵⁸, l'USIGRAI (Unione Sindacale dei Giornalisti Rai)⁵⁹ e l'Associazione Stampa Romana⁶⁰ che hanno tutte nei loro organigrammi una Commissione Pari Opportunità, attraverso le quali orientano le loro azioni non solo alla tutela delle pari opportunità, di accesso, carriera e salario, nella professione, ma anche alla garanzia di un'informazione bilanciata rispetto ai generi e non stereotipata. Concretamente, sono attive con campagne di comunicazione, sensibilizzazione e formazione, di recente prevalentemente concentrate sulla violenza contro le donne, declinato, da un lato, come tema della narrazione giornalistica, dall'altro, come problema di molestie, *hate speech* o minacce contro le giornaliste.

L'interesse verso queste tematiche specifiche focalizza anche l'attenzione recente dell'Ordine dei giornalisti, che il 19 novembre 2020 ha approvato una modifica del *Testo unico dei doveri del giornalista* (OdG 2021), inserendo l'articolo 5-bis, intitolato *Rispetto delle differenze di genere*, che prevede fra i doveri di giornalisti e giornaliste, «nei casi di femminicidio, violenza, molestie, discriminazioni e fatti di cronaca, che coinvolgono aspetti legati all'orientamento e all'identità sessuale», l'attenzione a «evitare stereotipi di genere, espressioni e immagini lesive della dignità della persona», l'uso di «un linguaggio rispettoso, corretto e consapevole», che eviti «espressioni, termini e immagini che sminuiscano la gravità del fatto commesso», l'«attenzione a non alimentare la spettacolarizzazione della violenza», e infine, «una narrazione rispettosa anche dei familiari delle persone coinvolte» (OdG 2020, 1). L'introduzione di questo articolo nel *Testo unico dei doveri del giornalista* (OdG 2021), che dal 2016 recepisce tutti i documenti deontologici, divenendo così un punto di riferimento imprescindibile per l'esercizio della professione giornalistica, è certamente un segnale di auto-consapevolezza professionale sul ruolo dell'informazione rispetto alla questione di genere. Tuttavia, l'articolo rimane circoscritto alla narrazione dei casi di violenza. E

⁵⁸ Fonte: <https://www.fnsi.it>; ultimo accesso 19 novembre 2021.

⁵⁹ Fonte: <http://www.usigrai.it>; ultimo accesso 19 novembre 2021.

⁶⁰ Fonte: <https://stamparomana.it>; ultimo accesso 19 novembre 2021.

nessun altro articolo fa riferimento a una più generale e ampia questione di uguaglianza di genere nell'informazione, che, in quanto tale, in Italia rimane nell'alveo di interesse di poche associazioni di categoria, o della società civile, e, circoscritta ai principi generali asseriti nel *Testo unico della radiotelevisione* (D.L. 177/2005).

In questo contesto, sarebbe forse utile che il *Testo unico dei doveri del giornalista* (OdG 2021) contenesse una sezione ampia dedicata in generale alla uguaglianza di genere nell'informazione, e che questa sezione contenesse un riferimento esplicito alla necessità di garantire un'equa rappresentanza e rappresentazione dei generi, anche attraverso un uso paritario e inclusivo della lingua italiana, il cui ruolo, come abbiamo visto, è fondamentale nel promuovere una visione della realtà egualitaria.

Sebbene non esistano studi che dimostrano, su basi statistiche, una correlazione lineare fra una maggiore e migliore rappresentazione delle donne nell'informazione e l'adozione di politiche per l'uguaglianza di genere dentro e attraverso i media, e questo tipo di analisi rappresenti al momento un vuoto di ricerca da colmare (Padovani e Bozzon 2020), diverse studiose evidenziano l'importanza di quadri normativi di riferimento sia a livello internazionale che nazionale (Ross et al. 2017; Padovani 2018), per diverse ragioni. In primo luogo per definire principi e obiettivi condivisi, sulla base dei quali poter valutare i cambiamenti (Gallagher 2011) e garantire una continuità ai progressi nel corso del tempo (Gallagher 2017). In secondo luogo, perché le politiche e le normative pubbliche possono favorire quelle trasformazioni culturali che costituiscono uno dei prerequisiti necessari all'avanzamento dell'uguaglianza di genere (Chaher 2014). Come sottolineano Ross et al. (2017), in relazione al contesto europeo, le relazioni fra politiche, organizzazioni dei media, e contenuti sono complesse, non lineari e non stabili nel corso del tempo, ma «the construction of formal, publicly promoted and integrated equality strategies constitutes the clearest evidence of media organizations' commitment to gender equality» (Padovani 2018, 259).

Capitolo 4 Costruzione e descrizione dei *corpora* analizzati

Questo capitolo è composto da un primo paragrafo in cui viene definito il concetto di *corpus*, seguito da tre paragrafi dedicati alla presentazione dei tre *corpora* analizzati nelle ricerche empiriche (Capitolo 5, Capitolo 6, Capitolo 7): il *corpus* TG, il *corpus* CRS e il *corpus* ACR.

4.1 Definizione di *corpus*

Un *corpus* può essere definito come «una collezione di testi selezionati e organizzati in maniera tale da soddisfare specifici criteri che li rendono funzionali per le analisi linguistiche» (Lenci, Montemagni e Pirrelli 2016, 26). I testi raccolti in un *corpus* possono essere scritti, trascritti dal parlato o multimediali. In quest'ultimo caso si tratta di collezioni di testi di parlato audio o audio-visivo in cui le trascrizioni sono associate a *frame* audio o audio-visivi. Lo sviluppo recente delle TIC ha reso disponibili i cosiddetti *web-corpora* o *corpora* basati sul web: collezioni di testi liberamente accessibili, scaricabili e, in taluni casi, sottoponibili ad analisi linguistiche direttamente online (Davies 2015, 11).

Le caratteristiche dei *corpora* sono principalmente due: 1) si tratta di collezioni di testi esemplificativi del linguaggio naturale, vale a dire testi prodotti allo scopo di comunicare in forma orale o scritta, non per essere inclusi in un *corpus* o in un'analisi linguistica, 2) rappresentativi della varietà linguistica che rappresentano: «a *corpus* is thought to be representative of the language variety it is supposed to represent if the findings based on its contents can be generalized to the said language variety» (Leech 1991, 28).

In termini di rappresentatività, i diversi tipi di *corpora* possono essere distinti in due grandi categorie: *corpora* generali o di riferimento per una lingua, rappresentativi delle diverse varietà di quella lingua, e *corpora* verticali o specialistici, rappresentativi di una specifica varietà linguistica o di un dominio particolare (Lenci et al. 2016, 29). I testi collezionati per costruire i *corpora*

analizzati nelle tre ricerche presentate in questa tesi sono tutti specialistici e, in particolare, rappresentativi del linguaggio dell'informazione italiana, come dettagliato nei paragrafi seguenti.

Altri criteri di classificazione dei *corpora* riguardano la modalità, la cronologia, la lingua, l'integrità e la codifica digitale (Lenci et al. 2016, 27-34). La modalità fa riferimento al tipo di produzione di testi, distinguendo fra modalità orale o scritta. Sulla base di questo criterio, i *corpora* possono essere collezioni di testi scritti, oppure di monologhi o dialoghi orali successivamente trascritti, oppure collezioni miste che raccolgono sia testi scritti che trascritti dal parlato. Come vedremo di seguito, i *corpora* analizzati in questa tesi sono di diversa natura: il *corpus* TG raccoglie testi integralmente trascritti da un campione di telegiornali, raccogliendo sia i testi delle notizie lette in studio, dei servizi e di eventuali collegamenti in diretta, dunque testi orali basati su uno scritto o semiscritto redazionale parlato da giornaliste e giornalisti, sia i testi delle interviste trasmesse nel corso del telegiornali, trascrizione del parlato delle fonti. I *corpora* CRS a ACR, invece, raccolgono solo testi scritti di articoli della stampa quotidiana.

In riferimento alla cronologia, un *corpus* può essere sincronico o diacronico (Lenci et al. 2016, 31), a seconda che comprenda testi prodotti in un periodo di tempo particolare, per esempio un anno, oppure testi di periodi diversi. I *corpora* analizzati in questa tesi coprono un periodo di due o tre anni, tuttavia solo il *corpus* TG e il *corpus* CRS possono considerarsi diacronici, nella misura in cui sono stati organizzati in modo tale da poter essere analizzati sull'asse diacronico, per una comparazione annuale fra il 2018, 2019 e 2020, il primo, e per un'analisi contrastiva, fra 2018 e 2019, il secondo. Il *corpus* ACR, sebbene raccolga testi prodotti in tre anni diversi, di fatto non può considerarsi diacronico perché è stato analizzato, come vedremo, integralmente come rappresentativo di un unico triennio (2018-2020).

Quanto alla lingua, un *corpus* può contenere testi prodotti in una sola lingua o in due o più lingue (*corpora* bilingue e multilingue); questi ultimi, su cui non mi

soffermo, possono poi a loro volta essere distinti in paralleli o comparabili (Lenci et al. 2016, 31). I *corpora* che ho analizzato sono tutti in lingua italiana, ciò non esclude che contengano anche parole di altre lingue, attestandone, nel caso, l'uso diffuso nell'italiano, in particolare nel linguaggio giornalistico.

Il criterio dell'integrità distingue fra raccolte di testi integrali e collezioni di porzioni di testi di lunghezza fissa, che talvolta vengono ritenute più adeguate al bilanciamento di un *corpus*. Testi di lunghezza molto diversa fra loro possono infatti portare a una collezione sbilanciata a favore dei testi più lunghi che condizionano l'intero *corpus*, distorcendo il campione, per questo in alcuni casi si preferisce optare per una selezione di testi della stessa dimensione. I *corpora* TG, CRS e ACR sono tutte collezioni di testi integrali.

Infine, per quanto riguarda la codifica digitale, i *corpora* possono essere annotati o marcati, a vari livelli, con etichette che ne descrivono le caratteristiche linguistiche (morfologiche, sintattiche, semantiche). Come vedremo di seguito i *corpus* TG, CRS e ACR sono stati marcati a diversi livelli, a seconda dell'obiettivo specifico delle indagini a cui sono stati sottoposti.

4.2 Il *corpus* TG

Il *corpus* TG è costituito dai testi integralmente trascritti e annotati dei TG1, TG2 e TG5 trasmessi in prima serata nei mesi di gennaio 2018, 2019 e 2020, rispettivamente da Rai 1, Rai 2 e Canale 5, coprendo 279 edizioni (31 all'anno per ciascuna delle 3 testate) e circa 139,5 ore di trasmesso, se si considera che ogni edizione dura circa 30 minuti. Le trascrizioni integrali sono state fornite dall'Osservatorio di Pavia⁶¹, che dal 2018 utilizza l'API Google Cloud Text-to-

⁶¹ Per una presentazione dell'Osservatorio di Pavia si veda il sito internet: <https://www.osservatorio.it> (ultimo accesso 19 novembre 2021).

Speech⁶² per la trascrizione automatica dell'audio dei principali telegiornali nazionali di fascia *prime time*.

La selezione dei TG e delle edizioni campione si è basata su criteri di *audience*, rappresentatività, bilanciamento e comparabilità. Il TG1 e il TG5 delle ore 20:00 sono i notiziari più visti: nel 2018⁶³ hanno avuto rispettivamente una media di 4.916.212 e 2.775.011 spettatori/trici, per uno *share* medio rispettivamente del 23,6% e 18,6%; nel 2019⁶⁴ un'*audience* di 5.636.000 e 4.523.000, per uno *share* medio rispettivamente del 24,1% e 19,1%; nel 2020⁶⁵ un'*audience* di 6.219.000 e 5.230.000 spettatrici/ori e uno *share* medio del 24,1%, il primo, e del 19,9%, il secondo. Il TG1 è trasmesso dalla Rai, concessionaria del servizio pubblico radio-televisivo in Italia, il TG5 da Mediaset, la seconda *media company* privata nazionale, storica concorrente della Rai, con cui continua a condividere la *leadership* in termini di offerta e ascolto dei notiziari (AGCOM 2020). Il TG2 delle 20:30, che è un notiziario sempre molto seguito – nel 2018⁶⁶ ha registrato una media di 1.749.859 spettatori/trici e uno *share* medio del 7,6%; nel 2019⁶⁷ 1.811.000 spettatori/trici, per uno *share* del 7,2%; nel 2020⁶⁸ un'*audience* di 1.791.000 e *share* del 6,69% – è stato scelto perché nel 2018 era diretto da una giornalista dichiaratamente impegnata a favore di un linguaggio paritario, inclusivo e non stereotipato⁶⁹. La sua

⁶² Fonte: https://cloud.google.com/speech-to-text/?hl=it&utm_source=google&utm_medium=cpc&utm_campaign=emea-it-all-it-dr-bkws-all-all-trial-e-gcp-1010042&utm_content=text-ad-none-any-DEV_c-CRE_170511603295-ADGP_Hybrid%20%7C%20BKWS%20-%20EXA%20%7C%20Ttxt%20~%20AI%20%; ultimo accesso 19 novembre 2021.

⁶³ Fonte: <https://www.affaritaliani.it/blog/prima-serata/ascolti-tv-auditel-tg1-resta-leader-mentana-579773.html>; ultimo accesso 19 novembre 2021.

⁶⁴ Fonte: <https://www.adginforma.it/ecco-la-top-ten-dei-tg-nazionali-nel-2019-in-base-allaudience/>; ultimo accesso 19 novembre 2021.

⁶⁵ Fonte: <https://bubinoblog.altervista.org/analisi-auditel-chi-segue-i-telegiornali-delle-ammiraglie/>; ultimo accesso 19 novembre 2021.

⁶⁶ Fonte: <https://www.affaritaliani.it/blog/prima-serata/ascolti-tv-auditel-tg1-resta-leader-mentana-579773.html>; ultimo accesso 19 novembre 2021.

⁶⁷ Fonte: <https://www.adginforma.it/ecco-la-top-ten-dei-tg-nazionali-nel-2019-in-base-allaudience/>; ultimo accesso 19 novembre 2021.

⁶⁸ Fonte: <https://bubinoblog.altervista.org/analisi-auditel-chi-segue-i-telegiornali-delle-ammiraglie/>; ultimo accesso 19 novembre 2021.

⁶⁹ Cfr. *Ida Colucci, la direttrice che ha cambiato linguaggio al Tg2*, Giulia Giornaliste, 19 agosto 2020, fonte <https://www.globalist.it/attualita/2019/08/19/ida-colucci-la-direttrice-che-ha-cambiato-linguaggio-al-tg2-2045308.html> (ultimo accesso 19 novembre 2021).

inclusione nel *corpus* di analisi consente così di verificare similarità o differenze fra linee editoriali diversamente impegnate sul fronte di un linguaggio giornalistico paritario e non stereotipato.

La scelta di limitare la costruzione del *corpus* ai testi dei TG di gennaio è stata vincolata dalla disponibilità dell'Osservatorio di Pavia a fornire gratuitamente le trascrizioni dei notiziari di un solo mese nel corso dell'anno⁷⁰; ho quindi optato per gennaio, in modo da coprire due diverse legislature (17esima nel gennaio 2018 e 18esima nel gennaio 2019 e 2020), evitando contemporaneamente i giorni della campagna elettorale per le politiche del 2018 (iniziata a metà febbraio 2018), in cui la visibilità televisiva, in particolare femminile (Azzalini 2009), è tradizionalmente monopolizzata dai/dalle *leader* di partito (Mazzoleni 2012), che attualmente sono tutti uomini, con la sola eccezione di Giorgia Meloni. L'inclusione nel campione di analisi di periodi che coprono due legislature diverse è stata motivata dall'esigenza di verificare eventuali correlazioni fra i risultati dell'analisi linguistica e il contesto politico, in particolare in relazione al diverso impegno delle istituzioni, nei due diversi periodi legislativi, nella promozione delle pari opportunità, in generale, e di un linguaggio *gender-fair*, in particolare. Il 23 marzo 2018 si è infatti chiusa la 17esima legislatura caratterizzata dall'impegno della presidente della Camera Laura Boldrini per un uso della lingua non sessista a partire dall'istituzione della *Commissione Jo Cox sui fenomeni d'odio, intolleranza, xenofobia e razzismo*⁷¹, volta a contrastare anche gli stereotipi diffusi a livello linguistico, fino alla promozione dell'uso dei femminili per nominare le cariche politiche (cfr. Capitolo 2, 2.3.3), e si è aperta una legislatura più disimpegnata su questo fronte.

Tutte le trascrizioni fornite dall'Osservatorio di Pavia in singoli file di testo (uno per ogni edizione quotidiana di ciascuna testata giornalistica) sono state controllate e corrette e unificate in un documento elettronico in formato testo (txt). La costruzione del *corpus* TG ha avuto inizio da questo file di testo con la

⁷⁰ È in fase di sviluppo un'interfaccia di consultazione del *corpus* liberamente accessibile online dal sito dell'Osservatorio di Pavia.

⁷¹ Fonte: <https://www.camera.it/leg17/1264m>; ultimo accesso 19 novembre 2021.

normalizzazione e la marcatura, entrambe operazioni eseguite manualmente in fase di pretrattamento dati. La normalizzazione ha consistito nella uniformazione grafica e ortografica del *corpus*, in relazione all'uso di maiuscole e minuscole, all'ortografia di acronimi (con o senza punti), parole composte (con o senza trattino) o espressioni polirematiche. Non è stato necessario alcun controllo della punteggiatura, perché le trascrizioni originali non contenevano punteggiatura e non si è ritenuto necessario introdurla. Per quanto riguarda la marcatura, ho annotato il testo seguendo gli standard previsti da IRaMuTeQ (versione 0.7 alpha 2 2020; Baril e Garnier 2015), il software che ho usato per l'analisi del *corpus* TG e che introduco dettagliatamente nel Capitolo 5 (§ 5.2.3.), a cui rinvio anche le informazioni sui criteri di annotazione.

La dimensione complessiva del *corpus* TG è di N=25.929 testi, N=1.434.733 occorrenze e N=35.447 forme lemmatizzate, avendo optato per la lemmatizzazione automatica del *corpus* più funzionale a un'esplorazione volta prevalentemente a un'analisi testuale. Questa scelta è stata fatta nella consapevolezza che la lemmatizzazione automatica pone diversi problemi, noti a chi fa ricerca, in particolare in termini di controllo efficace del processo e di risultati non sempre corretti, soprattutto nei casi di omografia (Bolasco 2013, 81). Per questo, prima di analizzare il *corpus*, ho apportato qualche modifica al dizionario italiano fornito dalla libreria di R utilizzata da IRaMuTeQ (per esempio la parola *ministra* non era riconosciuta come forma femminile singolare del lemma *ministro*, ma come la forma del verbo *ministrare*, per cui ho modificato il lemma del nome *ministro* in *ministra/o*) e ho ampliato il dizionario delle espressioni regolari (anch'esso fornito da R), aggiungendo alla lista di quelle presenti i nomi di partiti e schieramenti politici, in considerazione del fatto che i TG danno ampio spazio ai soggetti politici, anche collettivi⁷². Nella fase di analisi ho poi prestato attenzione a considerare eventuali errori risultanti dalla lemmatizzazione, in modo da considerarli nell'interpretazione dei risultati. Per esempio, alcuni nomi, come *ragazza/o*, sono presenti nel dizionario non con un'unica entrata, maschile singolare, ma con due

⁷² Fonte: <https://www.agcom.it/pluralismo-politico-sociale-in-televisione>; ultimo accesso 19 novembre 2021.

entrate una femminile, nelle due forme singolare e plurale (*ragazza, ragazze*) e una maschile anch'essa con le due forme singolare e plurale (*ragazzo, ragazzi*).

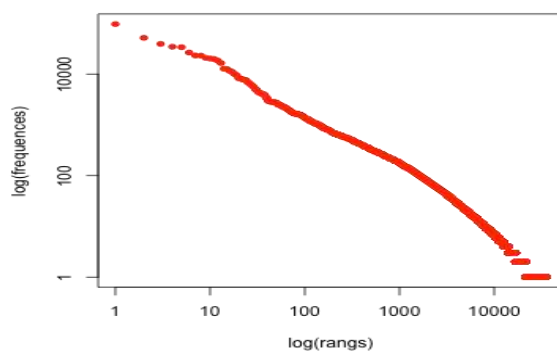
Tornando ora alle statistiche sulla dimensione del *corpus*, la Tabella 1 riporta tutti i dati dettagliati.

Tabella 1. Dimensioni del *corpus* TG

Testi (N)	25.929
Occorrenze (N)	1.434.733
Forme (Lemmi) (N)	35.447
Hapax ⁷³ (N)	14.051
Hapax/occorrenze (%)	0,98
Hapax/forme (%)	39,64
Media occorrenze*testo	55,33
Ratio (occorrenze/forme)	0,025

La curva disegnata dal rapporto fra il rango di frequenza e la frequenza delle parole empiricamente attestata nel *corpus* TG (Grafico 1) mostra un andamento “zipfiano”, ovvero in linea con la legge di Zipf (1949), per cui la frequenza di una parola in un testo è inversamente proporzionale al suo rango. L'output del rapporto fra il rango di frequenza delle parole attestate nel *corpus* e la loro frequenza viene fornito da IRaMuTeQ in un grafico in scala logaritmica – con il rango di frequenza sull'asse delle ascisse e la frequenza delle parole sull'asse delle ordinate – che mostra l'andamento di questo rapporto e consente di verificare se è in linea con l'andamento predetto dalla legge di Zipf.

Grafico 1. *Corpus* TG: profilo rango/ frequenza



⁷³ Gli hapax sono parole con occorrenza N=1.

4.3 Il corpus CRS

Il corpus CRS è costituito dal corpo del testo degli articoli pubblicati dal *Corriere della Sera*, *La Stampa*, *Il Resto del Carlino*, nei mesi di gennaio 2018 e gennaio 2019 e contenenti almeno una delle quattro forme del lemma *ministra/o* (*ministra*, *ministre*, *ministro*, *ministri*): N=3.594 articoli in tutto, distribuiti per anno e testata giornalistica come illustrato nella tabella seguente.

Tabella 2. Numero di articoli selezionati per la costruzione del corpus di analisi, per quotidiano e per anno.

	Gennaio 2018	Gennaio 2019
<i>Corriere della Sera</i>	491	745
<i>Il Resto del Carlino</i>	658	803
<i>La Stampa</i>	371	526
Totale	1.520	2.074

Gli articoli sono stati selezionati da Nexis Unix (Lexis Nexis Academic), una banca dati contenente pubblicazioni a stampa, nazionali e internazionali, scaricabili in file facilmente trasformabili in formato testo, dunque funzionali alla costruzione di un corpus, e accessibile dalla sezione *Risorse bibliografiche* del sito internet dell'Università Ca' Foscari di Venezia⁷⁴.

Le testate giornalistiche sono state scelte, fra le pubblicazioni a stampa quotidiana, sia per la loro reperibilità sulla banca dati Nexis Unix, sia per un confronto longitudinale con lo studio di Formato (2016), che ho parzialmente replicato, avendolo individuato nella rassegna degli studi sulla questione della femminilizzazione dei nomi di professione, come il più innovativo dal punto di vista metodologico.

Il periodo è stato selezionato in modo da coprire due diverse legislature (17esima nel 2018 e 18esima nel 2019) e due diversi governi (Governo Gentiloni nel 2018; Governo Conte I nel 2019), così da poter svolgere una ricerca diacronica e interpretare i risultati anche in relazione al diverso contesto politico nazionale, che, come vedremo, è il contesto di riferimento prevalente delle notizie raccolte

⁷⁴ Fonte: <https://advance.lexis.com/bisacademicresearchhome?crd=16bd77c6-d61e-456e-b1fc-83f33f37aceb&pdmfid=1516831&pdisurlapi=true>; ultimo accesso 19 novembre 2021.

nel *corpus*. Avevo avviato il progetto di ricerca programmando anche l'inclusione del mese di gennaio 2020, in modo da coprire anche il Governo Conte II, e costruire un *corpus* di articoli di quotidiani simmetrico al *corpus* TG, purtroppo Nexis Unix ha cessato di archiviare *La Stampa* il 31 dicembre 2019. Per l'analisi della femminilizzazione dei nomi di professioni, presentata nel Capitolo 7, ho deciso di sostituire il quotidiano di Torino con l'agenzia giornalistica ANSA, che è disponibile su Nexis Unix. Per l'analisi del lemma *ministra/o* ho invece deciso di circoscrivere il *corpus* CRS ai mesi di gennaio 2018 e 2019 perché l'analisi del *corpus* TG ha evidenziato differenze significative fra il 2018 e il 2019 ma non per il 2020.

La scelta di includere solo il mese di gennaio è dettata da ragioni di comparabilità con il *corpus* TG. L'agenda dell'informazione può infatti variare nel corso dell'anno, privilegiando la visibilità di un ministro o una ministra piuttosto che un altro o un'altra, a seconda della contingenza e della notiziabilità degli eventi. È dunque importate che un'analisi comparativa fra testi prodotti da diversi mezzi d'informazione copra lo stesso periodo.

Operativamente, per la costruzione del *corpus* CRS, ho selezionato, attraverso una *query* nella banca dati Nexis Unix, gli articoli contenenti le parole *ministra*, *ministro*, *ministre* e *ministri*, poi ne ho scaricato la versione integrale in singoli file rtf (rich text format) che ho successivamente aggregato. Una volta ottenuto un singolo file l'ho convertito, tramite il software di conversione TextUtil, in formato txt, con codifica UTF8. Tramite uno script in Python 3, ho annotato automaticamente il *corpus* con i metadati disponibili nei file originari (#Title, #Source, #Date, #Length, #Section, #Byline), ho poi separato i metadati dal *corpus*, salvandoli in un file a sé stante, e costruito il *corpus* CRS contenente soltanto i testi degli articoli senza metadati⁷⁵. Questa operazione l'ho poi ripetuta sei volte, costruendo 6 *sub-corpora* uno per ogni testata giornalistica e anno: *Corriere della Sera*, gennaio 2018 (CdS 01/2018) e gennaio 2019 (CdS 01/2019), *Il Resto del*

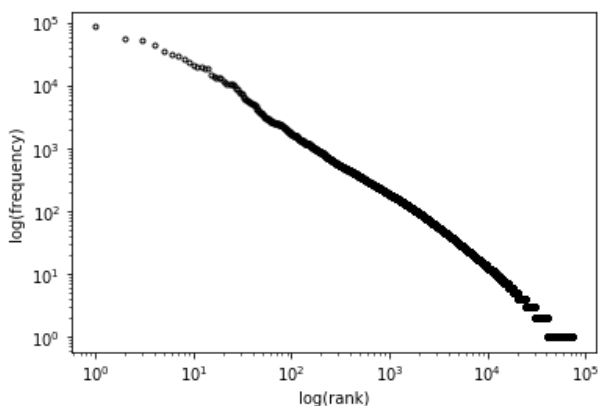
⁷⁵ Per annotare il *corpus* ho usato le seguenti librerie: <https://docs.python.org/3/library/os.html>, <https://docs.python.org/3/library/re.html>; <https://docs.python.org/3/library/glob.html>; <https://docs.python.org/3/library/errno.html>; <https://docs.python.org/3/library/time.html>; <https://docs.python.org/3/library/collections.html>; <https://stanfordnlp.github.io/stanza/>.

Carlino, gennaio 2018 (RdC 01/2018) e gennaio 2019 (RdC 01/2019), *La Stampa*, gennaio 2018 (St 01/2018) e gennaio 2019 (St 01/2919).

Infine, ho analizzato le caratteristiche del *corpus* in termini di dimensione, occorrenze, forme e *ratio*, come dettagliato di seguito. La dimensione complessiva del *corpus* CRS è di N=1.835.938 occorrenze e N=72.565 forme, inclusa la punteggiatura che è stata mantenuta per facilitare l'analisi qualitativa manuale, finalizzata a stabilire il genere semantico dei/lle referenti e a classificare le espressioni linguistiche ricorrenti nei riferimenti alle ministre e ai ministri.

Come risulta evidente dal Grafico 2, la curva disegnata dal rapporto fra il rango di frequenza e la frequenza delle parole empiricamente attestata nel *corpus* CRS mostra un andamento "zipfiano", ovvero in linea con la legge teorica di Zipf (1949).

Grafico 2. *Corpus* CRS: profilo rango/frequenza



4.4 Il *corpus* ACR

Il *corpus* ACR è costituito da tutti gli articoli pubblicati (online) dall'ANSA (notiziario generale in italiano), e dalle edizioni a stampa del *Corriere della Sera* e *Il Resto del Carlino*, nei mesi di gennaio 2018, 2019 e 2020. Le testate giornalistiche sono state scelte, primo, per la loro disponibilità nella banca dati Nexis Unix (Lexis Nexis Academic), secondo, perché sono rappresentative di tre diversi tipi di informazione. L'ANSA, acronimo di Agenzia Nazionale Stampa Associata, è un'agenzia di stampa.

Come molte altre agenzie (es. AdnKronos, AGI), ha un sito internet dove pubblica notizie, ma la sua funzione originaria e prevalente è quella comune a tutte le agenzie di stampa di raccogliere le notizie e fornirle, tramite servizi a pagamento, alle altre testate giornalistiche (radio, TV, stampa, Internet). La maggior parte delle testate giornalistiche non ha infatti una struttura interna tanto grande da poter conoscere gli eventi che accadono in un paese, e nel resto del mondo, in tutti i momenti. Un'agenzia di stampa è invece organizzata, con corrispondenti e inviati in molte località nazionali e internazionale, per coprire il maggior numero di eventi, in tempi rapidi, e fornire una prima sintetica versione dei fatti. Come ben sintetizza Alessandro Meucci (2001), a cui rinvio per approfondire il tema:

L'agenzia di stampa funziona insomma come strumento di "informazione primaria", una sorta di organo di reperimento e trasmissione delle materia prima, la "notizia", a favore dei giornali, un canale di raccolta e distribuzione dell'informazione, che funge da cerniera tra l'avvenimento e i cosiddetti mezzi di comunicazione di massa (Meucci 2001, 5).

Fra le diverse agenzie di stampa ho selezionato l'ANSA perché è la principale agenzia di stampa italiana e anche perché la sua storia è simbolicamente legata al ruolo dei mezzi di informazione nella divulgazione di un uso dell'italiano non sessista. Come ho già scritto nell'introduzione, il volume di Sabatini (1987) contiene infatti un richiamo al ruolo dei mezzi d'informazione nella promozione di un linguaggio paritario e inclusivo, attraverso il contributo di Sergio Lepri, che nei suoi quasi 40 anni di direzione dell'ANSA (1961-1990) si è attivamente impegnato per un uso della lingua italiana paritario e inclusivo, come testimonia una recente intervista (Robustelli 2014, 58-64).

Per quanto riguarda il *Corriere della sera* e *Il Resto del Carlino* sono stati selezionati in quanto rappresentativi rispettivamente di un'informazione a copertura nazionale e macro-regionale. Il *Corriere della sera* è il quotidiano che, insieme a *la Repubblica*, detiene da tempo la *leadership* nazionale in termini di

diffusione⁷⁶. La diffusione cartacea e digitale totale è stata per i mesi di gennaio 2018, 2019 e 2020, rispettivamente di 290.881, 279.510, 272.938 copie⁷⁷. *Il Resto del Carlino* è una delle tre testate regionali di QN (*Quotidiano Nazionale*), insieme a *La Nazione* e *Il Giorno*, con cui condivide le pagine nazionali e da cui si differenzia per un'ampia pagina di informazione locale focalizzata sull'Italia centrale, in particolare l'Emilia Romagna. La diffusione cartacea e digitale totale è stata per i mesi di gennaio 2018, 2019 e 2020 rispettivamente di 97.394, 93.315, 84.894 copie⁷⁸.

Per quanto riguarda il periodo di copertura, ho scelto gli stessi tre mesi selezionati per il *corpus* TG, perché l'intenzione originale era di poter svolgere anche un confronto fra *corpus* TG e *corpus* ACR, e anche per anno. Ma il *corpus* TG è risultato troppo piccolo per un'analisi distribuzionale.

Operativamente, ho costruito il *corpus* ACR scaricando tutti i contenuti nella banca dati Nexis Unix (Lexis Nexis Academic) in formato rtf, li ho quindi trasformati in file di solo testo (txt), e uniti fra loro in un solo file.

La dimensione complessiva del *corpus* ACR è di N= 322.467.589 occorrenze e N=442.508 forme, la *ratio* lemmi/occorrenze è di 0,001.

⁷⁶ Fonte: <https://www.adsnotizie.it/index.asp>; ultimo accesso 19 novembre 2021.

⁷⁷ Dati mensili dichiarati e certificati; fonte: https://www.adsnotizie.it/dati_certificati.asp (ultimo accesso 19 novembre 2021).

⁷⁸ *Ibidem*.

Capitolo 5 Analisi delle rappresentazioni di genere in un *corpus* di TG

Questo capitolo presenta un'analisi *corpus-driven* delle rappresentazioni di genere nel *corpus* TG, descritto nel Capitolo 4 (§ 4.2). Poiché la caratteristica delle analisi *corpus-driven* è quella di esplorare la natura di un testo, senza avanzare ipotesi *a priori*, ma lasciando che sia il *corpus* a guidare l'ipotesi (Tognini Bonelli 2001), il capitolo è introdotto direttamente da domande di ricerca molto generali (§ 5.1), utili più che altro a orientare l'analisi del *corpus* in una prospettiva di genere che tenga conto del *framework* teorico delineato nei capitoli 1-3. Quali sono le rappresentazioni di donne e uomini come soggetto/fonte e come oggetto/argomento di notizia, nel *corpus* TG? Ci sono differenze fra i tre periodi considerati e le tre testate giornalistiche campione? Segue un'ampia sezione dedicata alla metodologia (§ 5.2), che include la presentazione del software scelto come supporto per l'analisi computazionale e spiega la modalità di annotazione del *corpus*, che è stato marcato con variabili categoriali in modo da poter essere esplorato con un approccio integrato, linguistico e mediale.

I risultati della ricerca sono presentati nel paragrafo 0, suddiviso in due sotto-paragrafi: il primo si focalizza sulle rappresentazioni di donne e uomini come soggetto/fonte di informazione, esplorando il parlato di donne e uomini fonti delle notizie; il secondo si focalizza sulle rappresentazioni di donne e uomini come oggetto/argomento dell'informazione, analizzando il linguaggio utilizzato per parlare di donne e uomini sia da giornaliste/i sia da fonti interpellate dalla redazione per presentare le notizie. Questi risultati sono discussi distintamente nel paragrafo 5.4, che include anche alcune informazioni extratestuali utili a interpretare i risultati della ricerca.

Il capitolo termina con le conclusioni (§ 5.5), che mettono in evidenza come l'analisi del *corpus* TG rilevi la presenza di asimmetrie di genere sia nell'uso della lingua parlata da donne e uomini, sia nell'uso della lingua che parla di donne e uomini. In entrambi i casi, non si tratta di differenze *tout court*, "oggettive", interpretabili come il semplice riflesso di una realtà extralinguistica, bensì differenze

ascrivibili ai diversi lessici e campi semantici associati in modo stereotipato all'uno o all'altro genere nel discorso dell'informazione, ovvero alle diverse rappresentazioni di donne e uomini come soggetto o oggetto di notizia, in ultima analisi riconducibili sia a una differente interpellazione di donne e uomini, come fonti, sia a una diversa topicalizzazione delle une e degli altri nell'agenda dell'informazione. L'analisi rileva però anche qualche segnale di cambiamento, in particolare emerge chiaramente un'identità femminile soggetto di un discorso pubblico, qual è quello veicolato dai media, che da un lato, si manifesta nel parlato delle donne come propensione a marcare la propria identità di genere, dall'altro, si manifesta nel parlare delle donne come cittadine attive contro la violenza di genere e altre forme di discriminazione e violazione dei diritti umani. Per quanto riguarda il confronto fra i tre periodi e i tre TG considerati, i risultati più interessanti sono due: il primo riguarda la continuità, lungo l'asse diacronico, del discorso mediale in cui le donne sono soggetto attivo di dissenso; il secondo riguarda la linea editoriale del TG2 che si distingue positivamente, da un lato, nel favorire l'emersione delle identità femminili come soggetto di comunicazione politica e, dall'altro, nel non contribuire alla rappresentazione delle donne come vittime (passive) di violenza.

5.1 Domande di ricerca

Come anticipato nell'introduzione, questo capitolo presenta una ricerca sul *corpus* TG finalizzata a rilevare le rappresentazioni di genere latenti nel linguaggio dei notiziari italiani. Poiché la caratteristica propria della ricerca *corpus-driven* è quella di esplorare un *corpus*, senza partire da assunti *a priori* sulla natura dei dati indagati, non ho formulato ipotesi di ricerca *a priori*, ma, partendo dalla cornice teorica e contestuale offerta nel Capitolo 2 e nel Capitolo 3, ho semplicemente formulato una serie di domande utili a orientare l'esplorazione in una prospettiva di genere, e a comprendere l'eventuale incidenza di fattori contestuali e mediali, quali il diverso periodo di riferimento dell'agenda dell'informazione (anno) e le diverse linee editoriali dei notiziari (testata giornalistica), sui risultati.

Data questa premessa, ho articolato le domande come di seguito riportato:

- I. Quali sono le rappresentazioni di genere di donne e uomini, intese/i come soggetto, ovvero fonti dell'informazione nel *corpus* TG?
 - a. Ci sono somiglianze o differenze fra gli anni?
 - b. Ci sono somiglianze o differenze fra le testate giornalistiche?
- II. Quali sono le rappresentazioni di genere di donne e uomini, intese/i come oggetto, ovvero argomenti dell'informazione nel *corpus* TG?
 - a. Ci sono somiglianze o differenze fra gli anni?
 - b. Ci sono somiglianze o differenze fra le testate giornalistiche?

5.2 Metodologia

5.2.1 *L'analisi semiautomatica dei testi*

La metodologia prescelta per l'analisi del *corpus* TG è una specifica tipologia di analisi del contenuto, l'analisi semiautomatica, ovvero parzialmente computerizzata, di testi, scritti o trascritti dal parlato, che, insieme all'analisi automatica, ovvero totalmente computerizzata, ha una consolidata tradizione di ricerca nelle scienze umane e sociali (Bolasco 2013; Pandolfini 2017), per questo è particolarmente adeguata a un'analisi linguistica finalizzata a profilare rappresentazioni sociali, incluse quelle di genere, latenti in un testo mediale.

L'analisi del contenuto può essere definita come

un insieme di metodi che sono orientati al controllo di determinate ipotesi su fatti di comunicazione (emittenti, messaggi, destinatari e loro relazioni) e che a tale scopo utilizzano procedure di scomposizione analitica e di classificazione, normalmente a destinazione statistica, di testi e di altri insiemi simbolici (Rositi 1988, 66).

Storicamente, questo insieme di metodi è stato distinto in due grandi categorie: l'analisi del contenuto quantitativa e l'analisi del contenuto qualitativa (Losito 1996; Tuzzi 2003; Pandolfini 2017). I criteri su cui si basa questa distinzione sono fondamentalmente due: il modo in cui i contenuti vengono scomposti e classificati, e il modo in cui essi vengono analizzati (Pandolfini 2017).

Nell'analisi del contenuto qualitativa, la scomposizione e classificazione dei contenuti avviene tramite processi di individuazione di temi o concetti per deduzione o per induzione. Nel primo caso (deduzione), i contenuti vengono scomposti e classificati per categorie identificate *a priori* sulla base di un'ipotesi di ricerca, di un *framework* teorico di riferimento e di protocolli di ricerca, che vengono sviluppati, e talvolta modificati, in *itinere*, durante l'analisi, la quale si caratterizza per una progressiva revisione delle categorie di scomposizione e classificazione dei contenuti. Nel secondo caso (induzione), queste categorie sono direttamente individuate e definite in fase di analisi, attraverso un processo di progressiva creazione e revisione delle categorie stesse. Indipendentemente dall'adozione di un approccio di tipo deduttivo o induttivo, l'analisi del contenuto qualitativa procede per un'interpretazione argomentativa dei risultati, che vengono confrontati con l'ipotesi di ricerca e il quadro teorico di riferimento, in genere senza il supporto di software o misure statistiche, che sono invece caratteristiche distintive dell'analisi del contenuto quantitativa (Pandolfini 2017).

Nell'analisi del contenuto quantitativa, la scomposizione e classificazione dei contenuti avviene tramite categorie *a priori* che, una volta individuate sulla base delle domande di ricerca e di un *framework* teorico di riferimento, rimangono invariate e sono trattate come variabili categoriali. Il processo di analisi si avvale di misurazioni statistiche che possono essere di vario tipo (per esempio, lessicometriche) e possono riguardare le unità di analisi nella loro interezza (per esempio, una selezione di post di Twitter), oppure elementi costituenti le unità di analisi (per esempio, le parole usate in una selezione di post di Twitter). In quest'ultimo caso, le unità di analisi vengono assunte come unità di contesto, mentre gli elementi analizzati diventano le unità di analisi nel contesto.

L'analisi automatica o semiautomatica dei testi è un tipo di analisi del contenuto che si caratterizza per un superamento della distinzione fra analisi del contenuto qualitativa e quantitativa (Pandolfini 2017). L'analisi automatica prevede l'analisi di testi non precedentemente trattati, ovvero non preliminarmente letti e processati (segmentati e classificati) manualmente. L'analisi viene svolta

automaticamente da un programma informatico che, al più, richiede la configurazione di parametri funzionali all'esecuzione di determinate operazioni. Tuttavia, sia la fase di disegno della ricerca sia la fase di interpretazione dei risultati non possono prescindere da una discussione che consideri aspetti qualitativi quali il *background* teorico e le conoscenze metodologiche e dello stato dell'arte, premesse necessarie per inquadrare il *framework*, teorico e contestuale, della ricerca e formulare specifiche domande di indagine. Valutazioni di ordine qualitativo coinvolgono anche la scelta del software utilizzato, poiché ogni programma ha caratteristiche proprie e si basa su algoritmi e operazioni funzionali che implementano un metodo riferito a un modello di teorico piuttosto che a un altro (Chartier e Meunier 2011; Lahlou 2012; Bolasco 2013).

L'analisi semiautomatica, oltre a includere elementi qualitativi nelle fasi preliminari e finali della ricerca, prevede un'analisi qualitativa anche nella fase empirica della ricerca, che richiede un pretrattamento manuale dei contenuti, mediante lettura, classificazione e segmentazione degli stessi in unità di analisi distinte e marcate per variabili categoriali, le unità di contesto iniziali (UCI). Prendiamo ad esempio una ricerca finalizzata a verificare la presenza di *hate-speech* (linguaggio d'odio) in rete. Dato un campione selezionato di contenuti potenzialmente offensivi o minacciosi sui *social network* (per ricavare i quali occorrerà procedere con *query* per parole chiave), è possibile processare questi contenuti direttamente con un software di analisi testuale, senza pretrattare i contenuti. Il programma restituirà misure statistiche diverse, a seconda delle caratteristiche del software, indicativamente occorrenze e coricorrenze di forme e parole, classi morfologiche e/o sintattiche (le cosiddette POS, Part of Speech), profili lessicali, etc. In questo caso, stiamo eseguendo un'analisi automatica. Per analizzare lo stesso campione di contenuti, possiamo optare per un'analisi semiautomatica, pretrattando i contenuti raccolti, ovvero leggendoli e marcandoli manualmente con variabili di categoria. Se, per esempio, siamo interessati a distinguere i diversi gruppi target di *hate speech* su Twitter, possiamo distinguere e classificare, attraverso una lettura manuale, i post raccolti nel campione di analisi citato più sopra per categorie

sociali a cui sono indirizzati discorsi o parole d'odio: es. donne, comunità religiose, comunità straniera, etc. Questa classificazione richiederà una lettura di un campione di post e una classificazione degli stessi sulla base del loro contenuto.

Il pretrattamento dei testi con l'aggiunta di variabili di categoria ai contenuti, tipico dell'analisi semiautomatica, consente di combinare l'approccio qualitativo con quello quantitativo direttamente nella fase sperimentale della ricerca. Sebbene sia un approccio che implica un dispendio di risorse ed energie piuttosto oneroso, sostanzialmente impraticabile nell'analisi di contenuti di grandi o grandissime dimensioni (*big data*) ha il vantaggio di evitare la perdita di informazioni talvolta difficili da recuperare in altre fasi dell'analisi. Si pensi per esempio a un'analisi testuale dei contenuti di un campione di serie TV. L'informazione sulle caratteristiche dei/le parlanti (per esempio "genere", "fascia di età") sono ricavabili soltanto attraverso una visione integrale delle *fiction* oggetto di analisi, una contestuale segmentazione del testo per dialoghi e una classificazione dei segmenti per caratteristiche dei personaggi. Inoltre le informazioni extratestuali, inserite nel *corpus* tramite variabili categoriali, nell'analisi di un *corpus* di testi prodotti da un certo tipo di media sono indispensabili per tenere in considerazione le caratteristiche strutturali di quel tipo di media.

Per chi fa ricerca può non sempre essere facile scegliere fra un approccio automatico o semiautomatico all'analisi testuale. Un buon punto di partenza per fare una scelta consapevole e funzionale agli obiettivi della ricerca è quello di considerare le caratteristiche del campione di analisi, di valutare vantaggi e svantaggi dei due approcci sulla base delle caratteristiche del *corpus* e degli obiettivi di ricerca. In generale, l'approccio automatico risulta la scelta migliore per l'analisi di *corpora* di grandi o medio-grandi dimensioni, perché il pretrattamento dei dati richiede un dispendio di tempo tanto più oneroso quanto più grande è la dimensione di un *corpus*. Viceversa, l'approccio semiautomatico può essere una buona scelta per l'analisi di piccoli *corpora*, specialmente di tipo specialistico: primo, perché la loro dimensione contenuta ne riduce i tempi di pretrattamento manuale, secondo perché, nel caso di *corpora* mediatici, come il *corpus* TG, il pretrattamento

manuale dei dati permette di inserire informazioni sull'articolazione del testo utili a interpretarne le caratteristiche linguistiche in una prospettiva meta-linguistica che tenga conto delle caratteristiche sia del *medium*, nella fattispecie la televisione, sia del formato, ovvero il telegiornale. Per esempio distinguendo fra testi di titoli, lanci di notizia, notizie lette in studio, notizie confezionate in servizio, e così via.

5.2.2 *La scelta del software IRaMuTeQ*

Per esplorare il *corpus* TG ho optato per un'analisi semiautomatica dei testi, utilizzando per la parte computazionale il software IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires), in particolare la versione 0.7 alpha 2 (2020), basata sul pacchetto R 4.0.3. I motivi che hanno guidato la scelta di questo approccio riguardano, da un lato, la valutazione preliminare della dimensione e delle caratteristiche del *corpus* TG, dall'altro, una considerazione delle metodologie di analisi offerte da IRaMuTeQ e, in particolare, dei loro modelli teorici di riferimento.

Per quanto riguarda le valutazioni preliminari sul *corpus* TG, come abbiamo visto precedentemente, l'analisi semiautomatica dei dati può essere un utile ed efficace strumento di ricerca a condizione che il *corpus* indagato sia di dimensioni piccole, per via del dispendio di tempo richiesto dal pretrattamento manuale dei dati, che deve essere valutato in base ai tempi programmati per una ricerca. Il *corpus* TG è un *corpus* piccolo, sia che si considerino i parametri dimensionali tradizionalmente utilizzati nella linguistica dei *corpora* (Davies 2015, 11), come per esempio il numero di occorrenze che, come abbiamo visto (§ 4.2), sono poco più di un milione (N=1.434.733), e le forme, la cui occorrenza è di N=35.447; sia che si consideri la dimensione del campione di notiziari selezionato originariamente per costruire il *corpus*. Come scritto più sopra (§ 4.2) il *corpus* TG raccoglie le trascrizioni di tutte le edizioni *prime time* dei TG1, TG2, TG5 del mese di gennaio del 2018, 2019, 2020, per una copertura totale 139,5 ore di trasmesso, una quantità di tempo ragionevolmente visionabile, per il pretrattamento manuale dei dati, nel corso di un progetto di ricerca di dottorato distribuito su un periodo di tre anni.

Una volta stabilita la praticabilità di un pretrattamento manuale dei dati, ne ho considerati i vantaggi in termini della possibilità di tenere in considerazione le caratteristiche della collezione originale, che raccoglie trascrizioni di testi che coprono tre testate giornalistiche e tre anni diversi e sono esemplificativi di almeno due tipologie di parlato: il parlato giornalistico e il parlato delle fonti di notizia (persone intervistate in collegamento o in differita o di cui vengono trasmessi messaggi o discorsi ripresi in diretta). Segmentare il *corpus* per UCI (Unità di Classificazione Iniziale) sulla base delle caratteristiche dei/lle parlanti (giornalista o fonte) e del loro genere (donne e uomini), assunte come variabili categoriali nell'analisi multivariata, consente di rispondere alla prima domanda di ricerca, altrimenti destinata a essere inevasa. La segmentazione e analisi del *corpus* (anche) per anno di trasmissione e testata giornalistica rende possibile rispondere alle domande di ricerca focalizzate sulla comparazione per anno e TG. Infine, l'inserimento di variabili categoriali che marcano il *corpus* distinguendo fra loro le varie parti del testo giornalistico (titolo, lancio, servizio, intervista, etc.) consente di avere informazioni sul formato del testo, utili per interpretare i risultati dell'analisi.

Venendo ora alle analisi che IRaMuTeQ è in grado di eseguire, due sono i principali modelli teorici di riferimento. Il primo è il modello della semantica distribuzionale, secondo il quale il significato delle parole risiede nel loro uso (Wittgenstein 1953), nel senso inteso originariamente da Firth che una parola assume un significato proprio in relazione alle parole che l'accompagnano: «you shall know a word by the company it keeps» (Firth 1957, 11). Questo modello è utilizzato anche da Benzécri (1973, 1980) nell'ambito della scuola francese dell'*analyse des données* per sviluppare una metodologia di analisi dei dati testuali di tipo lessicale, focalizzata sulla distribuzione delle parole in un *corpus* o in *corpora* diversi a confronto. Il secondo modello implementato da IRaMuTeQ, con il metodo della "classificazione gerarchica discendente", si propone come una revisione del modello di Benzécri (1973, 1980), nel senso di un'analisi di dati linguistici di tipo testuale focalizzata sulla distribuzione delle parole in segmenti di testo, assunti come Unità di Classificazione Elementari (UCE), e sulla distribuzione e l'inter-

relazione di questi segmenti di testo in un *corpus* la cui struttura “memorizza” le condizioni di produzione del testo. Come sintetizza Reinert:

Cette méthodologie est associée à une orientation particulière des recherches en analyse des données textuelles. Il s'agit, non pas de comparer les distributions statistiques des "mots" dans différents *corpus*, mais d'étudier la structure formelle de leurs cooccurrences dans les "énoncés" d'un *corpus* donné. A ce titre, elle renoue avec l'approche distributionnelle évoquée par Benzécri, mais avec une volonté plus affirmée d'être d'abord une "analyse du discours" : en effet, on se propose de mettre en évidence une dimension d'organisation du texte qui "mémorise" ses conditions de production (Reinert 1993, 9).

Reinert (1990) parte dal presupposto che la semantica di un enunciato si differenzia nettamente dalla semantica di una parola, poiché contiene l'impronta di un soggetto psichico, vale a dire la codifica e decodifica della realtà da parte di un soggetto. Un enunciato semplice è la parte più piccola di un discorso in cui un soggetto esprime qualche cosa sul mondo, un punto di vista, che è la sua rappresentazione del mondo e che lo pone in relazione con un'entità esterna. Un *corpus* può essere considerato, oltre che un insieme di parole, come un insieme di enunciati semplici, che sono rappresentazioni (mentali) elementari della realtà. Studiando somiglianze e differenze del vocabolario fra i diversi testi, o meglio segmenti di testo, che compongono un *corpus* possiamo cogliere somiglianze o differenze fra le forme di relazione con il mondo, forme che, sempre secondo Reinert (1990), ci sono accessibili solo nella forma di rappresentazioni del mondo, che Reinert (1990) chiama “mondi lessicali” o “tipi di mondo”: «les types de mondes référentiels les plus sollicités par un sujet psychique, lors de l'élaboration du *corpus*» (Reinert 1990, 21-22), formulando così una nozione del tutto simile a quella delle rappresentazioni sociali di Moscovici (cfr. Capitolo 3, § 3.3.1).

Partendo da questa ipotesi, il linguista francese mette a punto un metodo di “classificazione gerarchica discendente” (CGD) in grado di suddividere un *corpus*, inteso come un insieme di enunciati, in “classi lessicali” caratterizzate – sul piano

lessicale – da omogeneità interna e diversità esterna, rispetto alle altre classi e – sul piano semantico – da diversi “mondi lessicali” ovvero diverse rappresentazioni della realtà latenti nella struttura di un *corpus*. Laddove il termine “mondo” non va inteso in senso realistico, ma cognitivo: «un monde apparait, au niveau cognitif, à travers un ensemble plus ou moins organisé de signes relatifs à des objets, des actes, des jugements, etc.» (Reinert 1993, 13), e la CGD come un approccio euristico per rilevare non il mondo “reale”, ma le diverse realtà costruite mediante l’uso della lingua, detto altrimenti le rappresentazioni sociali (cfr. più sopra § 3.3.1).

Questa peculiarità della CGD è interessante nell’analisi di un *corpus* TG focalizzata sulle rappresentazioni di donne e uomini. Le diverse testate giornalistiche possono infatti essere annoverate fra quei “soggetti collettivi” di cui scrive Reinert (1993, 12), i cui “mondi lessicali” si caratterizzano come “luoghi comuni” (vs. i luoghi “abitati” dal soggetto individuale) collocati in uno spazio intermedio fra le rappresentazioni individuali (il punto di vista sul mondo del soggetto individuale) e i pre-costrutti culturali (il punto di vista sul mondo condiviso da una comunità) che si possono “imporre” all’enunciatore/trice più di quanto l’enunciatore/trice non li scelga, anche se li ricostruisce conferendo loro una particolare “colorazione”, cioè un senso entro un universo (simbolico) socialmente condiviso. Una considerazione, quest’ultima, che appare particolarmente rilevante se si considera che il *corpus* TG è una raccolta di testi informativi e l’informazione televisiva si contraddistingue, come ampiamente discusso nel Capitolo 3, per una rappresentazione di genere ambivalente fra la resistenza di asimmetrie e stereotipi di genere, da un alto, e la capacità di proporre identità e ruoli di genere innovativi, dall’altro. E in questa ambivalenza giocano un ruolo fondamentale sia la cultura individuale espressa da giornaliste/i e fonti dell’informazione, sia la cultura del paese di cui entrambi sono portatori, sia la cultura giornalistica e/o redazionale condivisa da chi scrive e produce le notizie. E ovviamente l’uso della lingua, attraverso la quale le rappresentazioni sociali si codificano, decodificano e trasmettono (Moscovici 1969).

5.2.3 Caratteristiche di IRaMuTeQ

IRaMuTeQ è un programma informatico per l'analisi testuale, scritto in Python e basato su pacchetti di analisi statistica di R (Baril e Garnier 2015). È distribuito sotto licenza software libero GNU GPL (v2) e può essere scaricato e installato dal sito internet www.IRaMuTeQ.org. Il software è stato sviluppato da Pierre Ratinaud nell'ambito del LERASS (*Laboratoire d'études et de Recherches Appliquées en Science Sociales*) dell'Università di Tolosa 3, originariamente in francese e per l'analisi della lingua francese. Ora è disponibile anche in lingua inglese e supporta l'analisi di undici lingue diverse⁷⁹, incluso l'italiano.

IRaMuTeQ consente di analizzare dati linguistici raccolti sia sotto forma di *corpus* sia sotto forma di matrice, comunque di tipo specialistico, ovvero circoscritti a un determinato genere testuale e/o argomento, per esempio le rappresentazioni sociali nel dibattito dell'assemblea parlamentare francese (Ratinaud e Marchand 2015). In entrambi i casi, i dati testuali raccolti richiedono un pretrattamento manuale: i testi vanno divisi in segmenti iniziali (UCI) sulla base di variabili categoriali scelte dall'analista. Ogni UCI viene individuata da una stringa che precede il testo ed è introdotta da un marcatore composto dalla successione di quattro asterischi (****), seguito dal nome delle variabili (es. var 1, var 2, var n.), a sua volta preceduto da un asterisco, e seguito da un trattino basso più il nome o l'etichetta scelta per nominare la modalità della variabile (es. 1, 2, 3), con un'annotazione del tipo **** *var1_1, *var2_1 *var3_1 e così via. Poiché IRaMuTeQ è stato qui impiegato soltanto per l'analisi di un *corpus*, trascurerò la presentazione delle tecniche di analisi disponibili per l'analisi di matrici, concentrandomi sulle tipologie sviluppate per l'analisi di *corpora*, che sono le seguenti: 1) l'analisi lessicometrica (*statistiques/statistics*), 2) l'analisi delle specificità e delle corrispondenze lessicali (*spécificités/specificities*), 3) la classificazione gerarchica discendente (*classification descendante hiérarchique/ descending hierarchical classification*), 4) l'analisi delle

⁷⁹ Inglese, francese, galiziano, greco, italiano, norvegese, olandese, portoghese, spagnolo, svedese, tedesco.

somiglianze (*similitudes/similarity analysis*) e 5) la “nuvola di parole” (*nuage de mots/ word cloud*). Queste ultime due sono analisi grafiche.

L’analisi lessicometrica – che ho utilizzato per descrivere le caratteristiche del *corpus* TG (Capitolo 4, § 4.2) – identifica e riformatta le UCI in cui il testo è stato segmentato dal/la ricercatore/trice in unità di contesto elementari (UCE), sulla base di un algoritmo originariamente messo a punto per il software francese Alceste (Reinert 1983). Attesta, quindi, la frequenza delle occorrenze, delle forme e degli hapax nel *corpus*, calcola il rapporto fra numero di hapax e numero di occorrenze, fra numero di hapax e numero di forme e, infine, calcola il rapporto fra il rango di frequenza e la frequenza delle parole empiricamente attestata nel *corpus*, verificando la legge di Zipf (1949).

Con IRaMuTeQ la frequenza delle forme può essere calcolata secondo due criteri diversi: lemmatizzando o non lemmatizzando il *corpus*. Come già anticipato, ho scelto di lemmatizzare il *corpus* TG, poiché la lemmatizzazione è più funzionale a un’analisi prevalentemente testuale, come quella che ho condotto. IRaMuTeQ esegue la lemmatizzazione del *corpus* sulla base di due dizionari: il dizionario delle parole e il dizionario delle espressioni. Il dizionario delle parole è organizzato in una tabella in cui a ogni parola corrispondono un lemma e la rispettiva annotazione grammaticale, che, per la lingua italiana, si limita a un livello basilare di POS (Part of Speech), distinguendo aggettivi, articoli determinativi e indeterminativi, avverbi, congiunzioni, locuzioni, nomi, numeri e cifre, preposizioni, pronomi, verbi. Il dizionario delle espressioni contiene una serie di locuzioni, ovvero associazioni di due o più parole che, associate fra loro, formano un’unità grammaticale e lessicale. Sulla base di questi dizionari e della segmentazione delle UCI in UCE, il software francese individua le parti del discorso (POS) di ogni UCE, recuperando la funzione grammaticale di ogni parola nel testo, secondo il contesto, e disambiguandone la funzione in caso di omonimia (es. *sindaca* come indicativo presente, terza persona singolare del verbo *sindacare* e *sindaca* come forma singolare femminile del nome *sindaco*). L’analisi lessicometrica di IRaMuTeQ restituisce l’elenco delle forme/parole nel *corpus* distinte in forme lessicali (nomi, verbi e aggettivi) e forme

grammaticali (articoli, preposizioni, proposizioni, verbi ausiliari)⁸⁰. Per ogni parola o lemma, se si è optato per la lemmatizzazione del testo, è possibile verificare l'annotazione POS, la frequenza e tutte le UCE, ovvero i segmenti di testo, in cui ricorre.

L'analisi delle specificità e delle corrispondenze lessicali è un'analisi di tipo comparativo che permette di analizzare il *corpus* sulla base delle variabili categoriali predefinite e utilizzate per marcare il testo, confrontando fra loro le diverse modalità della variabile, che devono essere almeno tre. Per esempio, se si sceglie di marcare il *corpus* sulla base dell'anno di produzione del testo (es. 2018, 2019, 2020), l'analisi delle specificità consente di confrontare fra loro i testi del 2018, 2019, 2020⁸¹. Nell'analisi delle specificità il *corpus* viene suddiviso in segmenti di testo, sulla base della variabile categoriale prescelta, di cui viene identificato il vocabolario specifico. La specificità del vocabolario dei segmenti di testi correlati a ciascuna modalità della variabile viene misurata attraverso un test statistico che misura la significatività di ogni parola. IRaMuTeQ propone di *default* il test della legge ipergeometrica ma offre in alternativa il test del chi-quadrato (χ^2), che ho preferito, qui come nelle analisi successive, perché è «the most widely used statistical test for nominal data» (Stefanowitsch 2020, 177). I risultati dell'analisi delle specificità vengono utilizzati anche per la costruzione di un piano fattoriale, che rappresenta graficamente la vicinanza e la lontananza di linguaggio tra le diverse modalità delle variabili. La costruzione del piano fattoriale viene effettuata tramite un'analisi delle corrispondenze lessicali (ACL), una tecnica tipicamente usata nell'analisi automatica dei testi e che IRaMuTeQ utilizza anche nella classificazione gerarchica discendente.

L'ACL è un'applicazione dell'analisi delle corrispondenze multiple (ACM) a dati testuali ed è stata adottata dalla scuola francese dell'*analyse des données*,

⁸⁰ IRaMuTeQ utilizza l'etichetta "attive" per definire le forme lessicali e l'etichetta "supplementaire" per definire le forme grammaticali (Baril e Garnier 2015, 8).

⁸¹ Nel caso in cui una variabile presenti soltanto due modalità (per esempio il genere del/la parlante uomo o donna), e non è dunque possibile eseguire l'analisi delle specificità, è possibile svolgere un'analisi contrastiva costruendo due *sub-corpora* e confrontandoli a tutti i livelli di analisi.

originariamente da Benzécri (1992) e successivamente da Reinert (1990). L'ACM è una tecnica di analisi multidimensionale utilizzata nell'analisi del contenuto per fare emergere relazioni latenti tra un alto numero di variabili interdipendenti, la cui relazione viene ricondotta a poche variabili sintetiche, definite "fattori" o "dimensioni latenti". L'applicazione dell'ACM a dati linguistici assume il testo come variabile e le forme lessicali occorrenti nel testo come modalità della variabile "testo" e indaga il variare della variabile "testo" in relazione alle variazioni del "lessico". In pratica, l'ACL si basa sulla costruzione di una tabella con in riga i testi, o meglio i segmenti di testo (ogni riga un segmento) e in colonna le forme lessicali (ogni colonna una forma). A partire da questa tabella, si possono ottenere: 1) le associazioni tra le forme lessicali in relazione al loro profilo di ripartizione tra i testi, 2) le associazioni tra testi in relazione alla loro somiglianza sotto il profilo lessicale, 3) le associazioni privilegiate fra forme lessicali e segmenti che compongono un testo. I segmenti di testo possono essere prescelti sulla base di variabili categoriali con cui il testo è stato pretrattato, come nel caso dell'analisi delle specificità.

La classificazione gerarchica discendente (CGD) è una tecnica di analisi messa a punto da Reinert (1983, 1990), come anticipato nel sotto-paragrafo precedente e originariamente implementata da Alceste, il software francese da cui IRaMuTeQ eredita alcuni modelli e tecniche di analisi. La CGD procede per progressiva suddivisione del *corpus* in "classi lessicali" raggruppate secondo i rispettivi vocabolari costruiti sulla base della frequenza delle parole o lemmi nelle UCE. Il *corpus* viene suddiviso dapprima in due classi e poi successivamente per un numero *n*. di classi che dipende dalle caratteristiche del vocabolario complessivo. Poiché la CGD procede per individuazioni di classi lessicali che allo stesso tempo hanno un vocabolario simile al loro interno e un vocabolario diverso da altre classi, tanto più il vocabolario di un *corpus* è ricco, tanto più elevato sarà il numero di classi ottenute. Come nell'analisi delle specificità, IRaMuTeQ utilizza i risultati della CGD anche per un'ACL che proietta le diverse classi lessicali, ed eventuali variabili categoriali associate, su un piano cartesiano, dove l'asse delle ascisse e l'asse delle ordinate rappresentano i due principali fattori latenti che spiegano la varianza e la

cui natura deve essere interpretata. L'interfaccia consente di recuperare nel *corpus* originale i segmenti di testo associati a ogni classe, permettendo un'analisi più qualitativa dei dati, utile anche a interpretare le dimensioni latenti della proiezione fattoriale.

L'analisi delle somiglianze permette di visualizzare graficamente i rapporti di prossimità fra le parole del *corpus* sulla base di indicatori standard, ovvero misure di distanza e similarità, disponibili nella libreria proxy di R, implementata dal software francese⁸². L'indicatore proposto di default da IRaMuTeQ, per cui si è optato in tutte le analisi effettuate, è quello delle coriccorenze di superficie. Le rappresentazioni grafiche possibili sono più di una, ma hanno tutte in comune l'obiettivo di visualizzare la struttura del *corpus*. Che abbiano una forma simile alle ramificazioni di un albero o di una ragnatela, esse si caratterizzano per rispettare, nella distanza fra parole, le proporzioni reali di vicinanza e lontananza delle stesse nel *corpus*, e, nello spessore dei tratti che le uniscono, la misura della loro relazione effettiva delle parole nel *corpus*, in entrambi i casi sulla base dell'indicatore prescelto.

Infine, la *word cloud* (che non ho utilizzato) è un'altra analisi di tipo grafico che raggruppa e organizza le parole o i lemmi di un *corpus* sulla base della loro frequenza, rappresentandole nella classica forma a nuvola che le visualizza l'una in prossimità all'altra con dimensioni diverse a seconda della loro frequenza, e con al centro le parole più citate.

5.2.4 Marcatatura del corpus TG

Per analizzare il *corpus* TG con IRaMuTeQ, è stato necessario suddividere manualmente l'intero *corpus* in segmenti di testo sulla base di variabili categoriali prescelte e marcare ogni singolo segmento secondo lo standard del software, che, come già scritto, adotta il simbolo dell'asterisco ripetuto quattro volte per

⁸² Fonte: <https://cran.r-project.org/web/packages/proxy/index.html>; ultimo accesso 19 novembre 2021.

individuare l'inizio di ogni UCI e scritto una sola volta per indicare la scrittura di una variabile, il cui nome può essere liberamente scelto (in genere si sceglie una lettera o l'inizio del nome della variabile)⁸³. L'annotazione grammaticale (POS tagging) è stata invece eseguita automaticamente dal software, tramite uno script di R.

Figura 1. Esempio di marcatura del *corpus*

```
**** *A_2019 *G_03-01 *T_TG1 *N_Lancio *P_Giornalista *S_Donna
nuovo scandalo formiche all'ospedale San Giovanni Bosco di Napoli
parenti di un paziente in rianimazione trovano alcuni insetti sul
volto dell'uomo è in corso una nuova bonifica della struttura e il
ministro della salute Grillo dice chi ha sbagliato pagherà urgente
nominare un commissario sentiamo Giovanna Cucè
```

Le variabili categoriali scelte per la marcatura del testo sono le seguenti: Anno (*A), Giorno (*G), Testata giornalistica (*T), tipologia della Notizia (*N), tipologia dei/le Parlanti (*P) distinti in Giornalisti/e e Fonti (persone intervistate o di cui è trasmesso il parlato registrato in esterna), e loro genere sociale, variabile quest'ultima che è stata chiamata Sesso (*S) per evitare una possibile confusione fra la *G della variabile Giorno.

La tabella seguente riporta tutte le variabili categoriali utilizzate per marcare il *corpus* TG, e le relative modalità.

Tabella 3. Elenco delle variabili con cui è stato marcato il *corpus* TG

Variabili (in colonna) e rispettive modalità (in riga)					
Anno (*A)	Giorno (*G)	TG (*T)	Notizia (*N)	Parlante (*P)	Sesso (*S)
2018	1-01	TG1	Introduzione (saluti a inizio TG)	Giornalista	Donna
2019	2-01	TG2	Lancio (annuncio di un servizio)	Fonte	Uomo
2020	3-01	TG5	Notizia da studio (notizia letta in studio)		
	4-01		Servizio (servizio)		
	5-01		Collegamento (collegamento in diretta)		
	6-01		Commento (commento giornalistico)		
	...		Chiusura (saluti a chiusura TG)		
	...		Messaggio (messaggio rilasciato dalla fonte in contesto extra-TG e trasmesso dal notiziario)		
	31-01		Intervista (intervista a/di fonte)		

È importante notare che la variabile *P (Parlante) è fondamentale per distinguere i testi che sono rappresentativi del parlato giornalistico (*P_Giornalista),

⁸³ Operativamente ho proceduto alla marcatura dei testi lavorando sulle trascrizioni delle singole edizioni dei TG, ovvero marcando il testo sulla base della visione e dell'ascolto del notiziario.

nella maggior parte dei casi testi scritti o semi-scritti e condivisi dalla redazione, dal parlato più o meno spontaneo delle fonti (*P_Fonte), persone intervistate o di cui si trasmettono comunicazioni rilasciate in un contesto extra-TG, come per esempio conferenze stampa, dibattiti politici, discorsi pubblici di vario genere. Il parlato giornalistico è ulteriormente articolato nelle modalità della variabile *N (Notizia) Introduzione, Lancio, Notizia da studio, Servizio, Collegamento, Commento e Chiusura; il parlato delle fonti è distinto nelle modalità Intervista, che codifica il parlato di persone intervistate dalla redazione, e Messaggio, che codifica comunicazioni rilasciate in luoghi pubblici, ripresi e ritrasmessi dal TG.

5.3 Risultati

5.3.1 La rappresentazione di donne e uomini come fonte dell'informazione

La prima domanda di ricerca è finalizzata a indagare le rappresentazioni di genere di donne e uomini intesi come soggetto, ovvero parlanti fonti dell'informazione. Per questo è stato costruito un *sub-corpus* P_Fonte, che include soltanto le UCE marcate con la modalità Fonte della variabile Parlante (*P_Fonte)⁸⁴ scorporate dall'intero *corpus* TG, mediante procedura automatica eseguita da IRaMuTeQ. Le dimensioni e caratteristiche di questo *sub-corpus* sono descritte nella tabella seguente.

Tabella 4. Statistiche: dimensioni e caratteristiche del *sub-corpus* P_Fonte

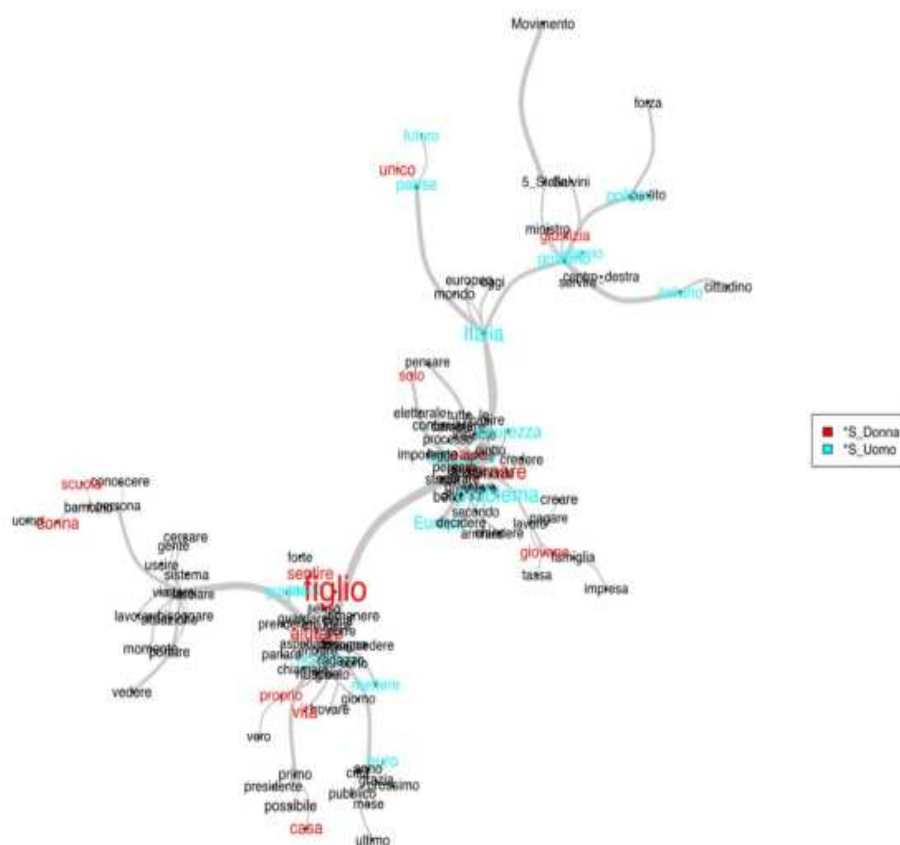
Testi (N)	7.848
Occorrenze (N)	231.875
Forme (Lemmi) (N)	12.149
Hapax (N)	5.407
Hapax/occorrenze (%)	2,33
Hapax/lemmi (%)	44,51
Media occorrenze*testo	29,55
Ratio	0,05

Per un'esplorazione preliminare, è stata eseguita un'analisi delle somiglianze, che calcola le coricorrenze fra parole e ne verifica la significatività

⁸⁴ La variabile Parlante ha due modalità *P_Giornalista e *P_Fonte.

mediante il test del chi-quadrato. Essendo un'analisi di tipo grafico che permette di considerare anche le variabili categoriali (una alla volta), questo tipo di analisi fornisce una prima e generale rappresentazione visuale dei rapporti di prossimità fra le parole (lemmi) e fra le parole e le modalità *S_Donna e *S_Uomo della variabile *Sesso (della fonte) nel *sub-corpus* P-Fonte. Poiché il *corpus* include più di 12.000 parole si è optato per un'analisi delle somiglianze circoscritta alle forme lessicali, più pregnanti dal punto di vista semantico, e con una frequenza assoluta $N > 100$, altrimenti ne sarebbe risultato un grafico illeggibile e di fatto inutilizzabile allo scopo per cui è stato costruito.

Grafico 3. *Sub-corpus* P_Fonte: analisi delle somiglianze (cut-off $N > 100$)



Nel Grafico 3 le parole (lemmi) sono rappresentate ai vertici delle linee che uniscono parole coricorrenti. Più grande è la dimensione delle parole, più le parole sono frequenti nel *corpus*; più spesse sono le linee che uniscono le parole, più le parole sono coricorrenti (Baril e Garnier 2015; Mariola e Ratinaud 2015; Camargo e Justo 2018).

Le parole in rosso sono quelle distintive dei segmenti di testo⁸⁵ con una fonte femminile (UCE marcate *P_Fonte e *S_Donna), le parole in azzurro sono quelle distintive dei segmenti con una fonte maschile (UCE marcate *P_Fonte e *S_Uomo). Prima di analizzare il grafico, vale la pena ricordare, come anticipato nel § 4.2, che il *corpus* TG è stato lemmatizzato, per cui le parole analizzate e riportate nelle tabelle e nei grafici in questo capitolo, sono lemmi, dunque per i verbi la forma infinita, per gli aggettivi la forma maschile singolare e per i nomi dal genere mobile, la forma maschile singolare.

L'addensamento di parole attorno a *figlio* e *tornare* rendono difficoltosa la lettura specialmente delle parole scritte in nero, che tuttavia non sono rilevanti allo scopo per cui è stato costruito il grafico, che illustra invece più chiaramente le parole distintive delle fonti femminili e maschili. Distintive del lessico femminile sono, leggendo il grafico da sinistra verso destra: *scuola* e *donna*, poi *figlio*, che è la parola con la dimensione massima nel grafico ed è collegata a *sentire* (posizionato sopra), *aiutare*, *proprio*, *vita* e *casa* (posizionate sotto), *unico* e *giustizia* (in alto), e infine l'assembramento di parole collegate a *tornare*: *solo* e *paura* (posizionate sopra), *giovane* (posizionata sotto). Distintive del lessico maschile sono, leggendo il grafico dall'alto a sinistra verso il basso a destra: *futuro* e *paese*, sulla sinistra; *politico*, *interno*, *governo*, *italiano*, sulla destra; un po' più in basso, allo snodo delle due ramificazioni precedenti, *Italia* (vicino a *mondo*, *europeo* e *oggi* in nero); poi, scendendo, *sicurezza*, *condizione*, *problema* e *Europa* prossime all'assembramento di molte parole in nero, fra le quali spiccano le due parole in rosso *paura* e *tornare*; infine, *grande* e *punto* (sovrapposte), *riguardare*, *mettere*, *euro*.

Per approfondire le asimmetrie di genere emerse dall'analisi del *sub-corpus* P_Fonte, quest'ultimo è stato suddiviso in due *sub-corpora* per genere della fonte: 1) PF_Donna, che raggruppa segmenti di testo marcati *P_Fonte e *S_Donna e costituisce il *sub-corpus* del parlato delle fonti femminili e 2) PF_Uomo, che raggruppa segmenti di testo marcati *P_Fonte e *S_Uomo e costituisce il *sub-*

⁸⁵ Segmenti di testo (ST) e UCE sono sinonimi.

corpus del parlato delle fonti maschili. Questi due *sub-corpora* sono stati esplorati in modo comparativo. La tabella seguente riporta dimensioni e caratteristiche.

Tabella 5. Statistiche: dimensioni dei *sub-corpora* PF_Donna e PF_Uomo

	<i>Sub-corpus</i> PF_Donna	<i>Sub-corpus</i> PF_Uomo
Testi (N)	1.944	5.904
Occorrenze (N)	51.547	180.328
Lemmi (N)	5.263	10.807
Hapax (N)	2.611	4.818
Hapax/occorrenze (%)	5,07	2,67
Hapax/lemmi (%)	49,61	44,58
Media occorrenze*testo	26,52	30,54
Ratio	0,10	0,06

L'analisi delle statistiche generali dei due *sub-corpora* analizzati in una prospettiva comparativa rileva una prima macro differenza dimensionale: il *sub-corpus* PF_Donna è più piccolo del *sub-corpus* PF_Uomo, registra infatti un numero di testi, di occorrenze e di lemmi pari a circa un terzo rispetto a quello attestato dal *sub-corpus* PF_Uomo.

Il confronto fra le forme lessicali (lemmatizzate) più frequenti (Tabella 6), mostra, a un primo livello di analisi, una "somiglianza" fra i due *sub-corpora*: trascurando i valori delle occorrenze, più elevate per il *sub-corpus* PF_Uomo, per via della maggiore dimensione di quest'ultimo, i due elenchi con parole di rango pari o inferiore a 30 (che coprono circa il 15% del lemmario di entrambi i *corpora*⁸⁶) ne includono 25 uguali, su un totale di 34 del *sub-corpus* PF_Donna e 33 del *sub-corpus* PF_Uomo.

Tabella 6. Forme lessicali più frequenti nei *sub-corpora* PF_Donna e PF_Uomo (occ.⁸⁷)

<i>Sub-corpus</i> PF_Donna				<i>Sub-corpus</i> PF_Uomo			
Rango	Lemma	Occ (N)	Occ. (%)	Rango	Lemma	Occ. (N)	Occ. (%)
1	dire	244	4,64	1	dire	820	7,59
2	volere	185	3,52	2	stare	649	6,01
3	anno	170	3,23	3	anno	506	4,68
4	persona	122	2,32	4	Italia	499	4,62
5	proprio	111	2,11	5	persona	437	4,04
6	figlio	107	2,03	6	governo	391	3,62
7	vedere	105	2,00	7	paese	389	3,60

⁸⁶ 15,4% del *sub-corpus* PF_Donna (occorrenze tot. forme lessicali N= 18.769, occorrenze forme lessicali al cut-off, rango 30, N=3.882); 14,4% del *sub-corpus* PF_Uomo (occorrenze tot. Forma attive N= 70.627, occorrenze forme lessicali al cut-off, rango 30, N=10.150).

⁸⁷ Occorrenza.

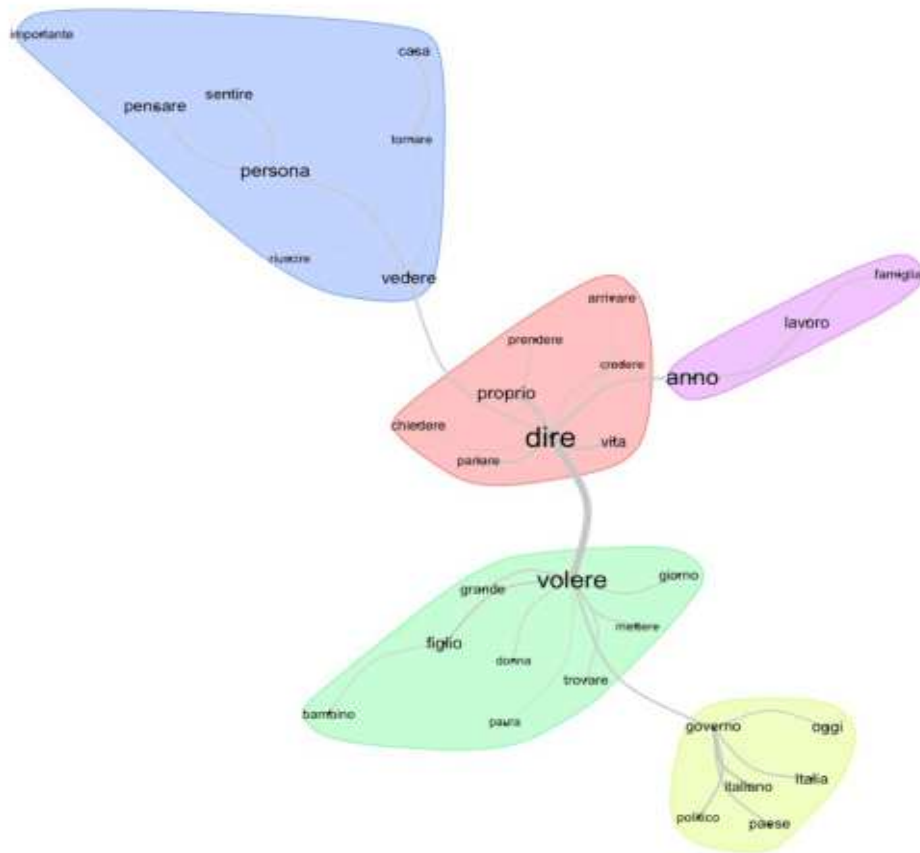
Sub-corpus PF_Donna				Sub-corpus PF_Uomo			
Rango	Lemma	Occ (N)	Occ. (%)	Rango	Lemma	Occ. (N)	Occ. (%)
8	pensare	97	1,84	8	primo	387	3,58
9	vita	94	1,79	9	italiano	373	3,45
10	sentire	86	1,63	10	grande	332	3,07
11	lavoro	85	1,62	11	lavoro	325	3,01
12	oggi	82	1,56	12	politico	313	2,90
13	governo	79	1,50	13	solo	306	2,83
13	Italia	79	1,50	13	proprio	306	2,83
14	paese	76	1,44	14	pensare	305	2,82
15	italiano	75	1,43	15	vedere	301	2,79
16	casa	74	1,41	16	mettere	251	2,32
17	bambino	70	1,33	17	oggi	247	2,29
18	giorno	69	1,31	17	arrivare	247	2,29
19	parlare	68	1,29	18	giorno	234	2,17
19	grande	68	1,29	19	credere	224	2,07
20	trovare	67	1,27	20	prendere	212	1,96
21	chiedere	66	1,25	21	problema	210	1,94
22	prendere	63	1,20	22	portare	205	1,90
23	arrivare	61	1,16	23	trovare	202	1,87
24	famiglia	59	1,12	24	momento	201	1,86
25	politico	58	1,10	25	Europa	193	1,79
26	donna	56	1,06	26	chiedere	192	1,78
27	mettere	55	1,05	27	sentire	180	1,67
27	importante	55	1,05	28	parlare	179	1,66
28	tornare	53	1,01	28	cittadino	179	1,66
29	paura	51	0,97	29	bambino	178	1,65
30	riuscire	50	0,95	30	vita	177	1,64
30	credere	50	4,64				

Scorrendo la tabella, dove sono riportate in grassetto le parole che compaiono in uno dei due elenchi ma non nell'altro, si contano nove parole fra quelle di rango pari o inferiore a 30 nel *sub-corpus* PF_Donna non presenti nell'elenco delle parole di rango pari o inferiore a 30 nel *sub-corpus* PF_Uomo, in ordine decrescente per rango: *volere*, *figlio*, *casa*, *famiglia*, *donna*, *importante*, *tornare*, *paura*, *riuscire*; e otto parole del *sub-corpus* PF_Uomo assenti fra le parole di rango pari o inferiore a 30 del *sub-corpus* PF_Donna: *stare*, *solo*, *primo*, *problema*, *momento*, *Europa*, *cittadino* (in ordine decrescente per rango).

Somiglianze e differenze fra i due *sub-corpora* emergono anche dal confronto fra le due analisi delle somiglianze, che raggruppano parole di rango pari o inferiore a 30 coricorrenti nel testo, rispettivamente nel *sub-corpus* PF_Donna e nel *sub-corpus* PF_Uomo. La rappresentazione grafica dell'analisi delle somiglianze nel *sub-corpus* PF_Donna (Grafico 4) mostra cinque cluster, di colori diversi, collegati fra loro a partire da due snodi centrali attorno a *dire*, in alto, e a *volere*, più in basso. Il cluster rosa, focalizzato su *dire*, può essere interpretato come il

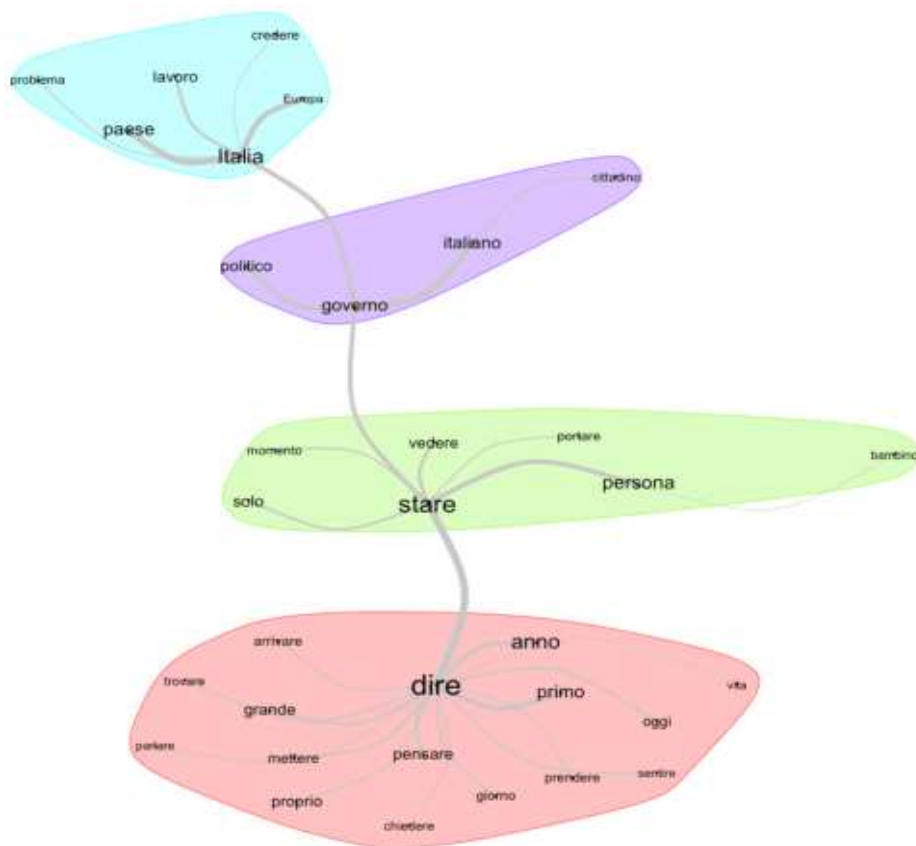
vocabolario minimo della narrazione, in particolare della narrazione del vissuto personale: il verbo dire è infatti collegato al verbo parlare (narrazione), al nome vita (vissuto) e all'aggettivo proprio (personale). Il cluster verde, focalizzato sul verbo volere, contiene le parole donna, che fa riferimento a un'identità di genere, figlio, che fa riferimento un'identità relazionale, specificamente familiare, e la parola bambino, che fa riferimento a un'identità generazionale, non esclusivamente circoscritta alla sfera familiare e domestica, ma riferita in generale al mondo dell'infanzia. In questo stesso cluster si trova anche la parola paura, che evoca un sentimento negativo. Guardando ora ai cluster collocati in una posizione più periferica del grafico, quello con al centro la persona, collegata ai verbi di percezione sentire e vedere, può essere interpretato come il lessico della testimonianza personale. Il cluster color violetto, sulla destra, non ha un nodo centrale, ma contiene due parole, lavoro e famiglia che evocano in modo molto chiaro la classica questione (femminile) della conciliazione fra tempi di vita e tempi di lavoro, che continua a essere semplificata dalla- e nella- etichetta "conciliazione lavoro-famiglia" (cfr. ISTAT 2019). Infine, il cluster giallo, in basso a destra, lontano dal cluster della narrazione del vissuto personale, contiene le parole della politica (governo, politico), in particolare nazionale (Italia, italiano).

Grafico 4. *Sub-corpus* PF_Donna: analisi delle somiglianze (cut-off al rango 30)



La rappresentazione grafica dell'analisi delle somiglianze nel *sub-corpus* PF_Uomo (Grafico 5) mostra quattro raggruppamenti di parole, che partono tutte dallo snodo del verbo *dire*, attorno al quale si delinea un'area simile, ma più estesa, a quella corrispondente nel *sub-corpus* PF_Donna; anche qui ci sono tre parole chiave che rimandano alla dimensione della narrazione del vissuto personale: *parlare, vita, proprio*.

Grafico 5. *Sub-corpus* P_Fonte PF_Uomo: analisi delle somiglianze (cut-off al rango 30)



Leggendo il grafico dal basso verso l'alto, dallo snodo del verbo *dire* si dirama il *cluster* verde chiaro, con al centro il verbo *stare* collegato alla parola *persona* (sulla destra), all'aggettivo *solo* (sulla sinistra) e al verbo *vedere* (in alto). Un raggruppamento simile, di difficile interpretazione, non risulta dall'analisi del *sub-corpus* PF_Donna, dove mancano *stare*, *solo* e *momento* (fra le parole di rango inferiore o uguale a 30), mentre *persona* e *vedere* – come abbiamo visto – sono raggruppate in un *cluster* abbastanza omogeneo, che è interpretato come il vocabolario della testimonianza, e *bambino* rientra in un *cluster* che evoca contemporaneamente la sfera familiare e quella dell'infanzia.

Per quanto riguarda gli altri due *cluster*, quello colorato di viola può considerarsi il raggruppamento delle parole della politica, corrispettivo al *cluster* giallo nel grafico del *sub-corpus* PF_Donna, con il quale condivide le parole *governo*, *politico*, *italiano*. Il *cluster* azzurro fa riferimento al *lavoro*, ma in modo diverso da quanto visto per il *sub-corpus* PF_Donna, dove il binomio lavoro-famiglia richiama

in modo chiaro la questione della conciliazione fra tempi di vita e lavoro. Qui c'è un'esplicita connotazione *problematica*, indicata dalla presenza della parola *problema*, entro un contesto nazionale ed europeo, come indicano le parole coricorrenti *paese, Italia, Europa*. È interessante notare anche la presenza in questo *cluster* del verbo *credere*, tipicamente usato per esprimere un'opinione, piuttosto che per narrare esperienze di vita personale o testimoniare eventi.

Venendo ora alla CGD, come abbiamo visto più sopra, questa analisi suddivide un *corpus* in n. classi lessicali caratterizzate da un vocabolario simile al proprio interno e, al contempo, diverso dalle altre classi. Oltre a ripartire il *corpus* in classi lessicali, la CGD verifica l'associazione fra classi e modalità delle variabili categoriali con cui il *corpus* è stato marcato. Questo consente di attestare eventuali associazioni significative fra le classi del *sub-corpus* P_Fonte e le due modalità *S_Donna/*S_Uomo. La CGD del *sub-corpus* P_Fonte mostra tre classi lessicali che coprono il 95,24% del *corpus* analizzato (N=8.291), come dettagliato nella Tabella 7, che riporta tutte le informazioni sui risultati dell'analisi.

Tabella 7. Statistiche: risultati della CGD sul *sub-corpus* P_Fonte

Testi (N)	7.848
Segmenti di testo (N)	8.705
Forme (N)	19.745
Occorrenze (N)	231.875
Lemmi (N)	12.149
Forme (lemmi) attive (N)	11.499
Forme (lemmi) supplementari (N)	571
Forme (lemmi) attive con frequenza ≥ 5	2.594
Media di forme per segmento	26,64
Classi (N)	3
Segmenti classificati (95,24%)	8.291

La classe 1 (Grafico 6) copre il 49,42 d% delle UCE analizzate (N=4.097) e attesta un'associazione significativa con la modalità della variabile Sesso (del/la parlante) *S_Donna: $\chi^2(1, N=1.995)=160,03, p<0,001$. La classe 2 copre il 13,23% delle UCE (N=1.097) ed è associata all'altra modalità della variabile Sesso, *S_Uomo, con un valore chi-quadrato non alto ma significativo: $\chi^2(1, N=6.296)=9,18, p<0,05$). La classe 3 copre il 37,35% delle UCE (N=3.097) ed è

significativamente associata alla modalità *S_Uomo, con un valore chi-quadrato elevato: $\chi^2(1, N=6.296)= 119,94, p<0,001$.

Grafico 6. Dendrogramma delle classi lessicali del *sub-corpus* P_Fonte



Per comprendere meglio questi risultati si riportano gli elenchi delle forme lessicali (lemmatizzate) più frequenti nei segmenti di testo (ST) per classe, in ordine decrescente per valori del chi-quadrato. Poiché gli elenchi completi sono molto lunghi, qui si è optato per una selezione dei primi 30 lemmi (Tabella 8) che copre una percentuale variabile, a seconda della grandezza delle classi (dal 25% della classe 1, al 22% della classe 3, al 16% della classe 2)⁸⁸, delle occorrenze dei lemmi nei segmenti di testo e include tutte parole il cui χ^2 di associazione alla classe ha una *p-value* < 0,001⁸⁹. Gli elenchi completi, con tutti i lemmi significativi per classe, e le loro occorrenze nei ST che compongono la classe relativa, sono riportati in Appendice (Tab. 1, Tab. 2, Tab. 3).

⁸⁸ Le percentuali sono calcolate sul totale delle frequenze delle forme lessicali nei segmenti di testo che formano le classi (nell'output di IRaMuTeQ eff. s.t.), dunque sono significativamente correlate alla classe con $\chi^2>3,84, p\text{-value}<0,05$. Per la classe 1 $N= 3.337/N=13.080$; per la classe 2 $N=648/N=3.970$; per la classe 3 $N=4.496/N=20.851$.

⁸⁹ Essendo $p<0,001$ per tutte le parole in elenco, il valore è stato omesso dalla tabella, che così risulta più leggibile.

Tabella 8. Prime 30 forme lessicali significative delle 3 classi del *sub-corpus* P_Fonte⁹⁰

Classe 1			Classe 2			Classe 3		
Lemma	χ^2	ST (N)	Lemma	χ^2	ST (N)	Lemma	χ^2	ST (N)
figlio	128,683	176	virus	186,640	34	governo	628,496	388
sentire	104,475	194	zona	186,093	41	Italia	414,620	406
ragazzo	86,175	125	paziente	154,562	27	politico	411,738	303
bello	85,352	122	evento	148,606	25	paese	373,576	343
dire	85,339	575	medico	142,079	28	italiano	263,633	301
succedere	66,147	130	cinese	126,232	28	5_Stelle	250,777	149
vedere	62,858	255	temperatura	122,454	21	Salvini	223,052	147
vita	58,107	178	sintomo	118,299	18	cittadino	217,933	180
mamma	56,822	58	Cina	112,826	25	Europa	206,028	160
bambino	53,791	155	influenza	109,42	19	tassa	199,03	117
casa	46,277	150	infezione	109,195	20	lavoro	183,696	254
ragazza	43,369	45	malattia	100,728	23	Movimento	178,355	105
vivere	41,802	97	area	99,434	22	elettorale	169,758	100
dolore	38,514	48	coronavirus	96,345	18	centro_destra	161,172	95
amore	38,21	40	rischio	94,586	43	legge	159,655	110
film	38,046	37	inverno	89,941	16	europeo	157,954	109
fratello	37,179	39	epidemia	89,941	16	presidente	145,053	105
paura	35,884	95	animale	84,518	21	sicurezza	144,7	138
proprio	35,747	243	freddo	79,648	24	forza	140,499	141
padre	35,487	45	respiratorio	78,809	12	ministro	132,888	96
ricordo	34,868	49	saldo	70,918	12	elezione	132,057	78
morire	32,647	63	produrre	70,851	29	impresa	128,866	90
aspettare	30,242	78	giacca	70,587	13	economico	125,034	84
musico	29,974	32	effetto	69,182	22	maggioranza	123,292	75
piacere	28,945	57	identificare	66,628	15	votare	114,652	72
piangere	28,76	28	Wuhan	65,658	10	Parlamento	112,949	71
moglie	28,311	33	sci	64,176	12	partito	107,099	84
capire	28,299	107	prodotto	63,755	16	Lega	106,467	63
cuore	28,05	56	utilizzare	63,01	29	programma	99,938	71
divertire	27,73	27	computer	59,085	9	riforma	99,439	61

Classe 1: le “testimonianze di vita”

Analizzando l’elenco delle parole più frequenti e significative per classe lessicale (Tabella 8), risulta evidente che la classe 1 contiene un primo gruppo di parole che rimandano all’universo delle emozioni: *sentire, dolore, amore, paura, piacere, piangere, cuore* – riportate nella Tabella 8 – *divertire, emozione, contento, gioia, animo, tranquillo, felice, ridere, emozionare, passione, divertente, sorriso, amare, ansia, spaventare, sentimento, felicità, indifferenza, lamentare, desiderio,*

⁹⁰ χ^2 indica il valore del chi-quadrato di associazione fra il lemma e la relativa classe lessicale, ST (N) indica il numero di occorrenze del lemma nei segmenti di testo della classe lessicale di riferimento.

angoscia, dispiacere, allegro, arrabbiare, abbraccio, abbracciare, odio, sofferenza – non riportate in Tabella 8, ma nell’elenco delle parole significative ($x^2 \geq 3,84$; $p < 0,05$) in Appendice (Tab. 1). Un secondo gruppo di parole fa riferimento a relazioni familiari e affettive: *figlio, mamma, fratello, padre, moglie* (Tabella 8), *sorella, amico, compagno, madre, nipote, genitore, marito, familiare, nonno, coppia, matrimonio* (Appendice, Tab. 1). Infine, un terzo gruppo di parole ha come universo di riferimento l’arte e la dimensione estetica più in generale: *bello, vedere, film, musico* (Tabella 8), *visto, bellezza, guardare, brutto, meraviglioso, raccontare, personaggio, bellissimo, cantare, canzone, racconto, televisione, artista, regista, sguardo, concerto, arte, meraviglia, festival, cinema, foto, danza, fantastico, Celentano, stupendo, splendido, video, Sanremo, ballare, musicale, chitarra, teatro, moda* (Appendice, Tab. 1).

Questa classe è significativamente associata non solo alla modalità *S_Donna, ma anche alle modalità *N_Intervista ($x^2(1, N= 6.142, ST N=3.186)=57,243, p < 0,001$), *N_Collegamento ($x^2(1, N= 65, ST N=55)=32,475, p < 0,001$) e *N_Servizio ($x^2(1, N=82, ST N=57)=13,382, p < 0,05$). L’associazione significativa alle modalità *N_Intervista, *N_Collegamento e *N_Servizio, insieme alla lettura dei segmenti di testo tipici (STT) – di cui si riporta una selezione nell’Elenco 1 – la profila come uno spazio caratterizzato dalla narrazione emozionale espressa per voce di persone intervistate in differita (*N_Intervista), in diretta dallo studio del TG (*N_Collegamento) oppure rievocate in interpretazioni artistiche trasmesse nei servizi (*N_Servizio). In questo mondo lessicale coesistono, nel comune registro delle emozioni, “testimonianze di vita” diverse: di vita privata (es. STT 2, 4, 29, 54, Elenco 1) o artistica (es. STT 20, 25, 123, 323, Elenco 1), raccolte in interviste o collegamenti in diretta (es. STT 20, Elenco 1), e anche rievocate in interpretazioni artistiche trasmesse nel corso dei servizi, come esemplificato dal STT 323(Elenco 1) che contiene uno stralcio della poesia di Primo Levi *Se questo è un uomo*.

Elenco 1. Selezione di segmenti di testo tipici della classe 1 del *sub-corpus* P_Fonte

STT 2 **** *A_2020 *G_29-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 455,48

allora con mia **moglie** durante uno dei colloqui quando **pensavamo** al futuro le **dissi** facciamo una cosa **troviamo** costruiamo una cosa **bella** talmente **bella** che **riesce** a sopraffare il **dolore** avevamo deciso di mantenere 2 **figli** regaliamoci il terzo **figlio** e quando sono **uscito** ci siamo **regalati** **Francesco**

STT 4 **** *A_2020 *G_30-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo

score: 425,10

è stato per me molto **interessante** dopo essere intervenuto in aula avere **sentito** il tributo di tante **persone** anche sconosciute che sono venute a **dirmi** questa è una **storia** che mi riguarda questa è una **storia** che ho **vissuto** in prima **persona** che **vive** mio **figlio**

STT 20 **** *A_2020 *G_10-01 *T_TG5 *N_Collegamento *P_Fonte *S_Uomo

score: 348,81

ciao Cesara beh qui lo studio è molto caldo come potrai **sentire** e comunque devo **dirti** Cesara sono **passati** appena 48 ore ed è **successo** di tutto **stasera** puntatone 8 nuovi vip conquisteranno la **casa** del Grande_ **fratello** sapremo finalmente se

STT 25 **** *A_2020 *G_17-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo

score: 329,51

ho visto tutti i suoi **film** e quest **uomo** è la mia famiglia mentale io sono **figlio** cinematografica di **Fellini** anch io **sogno** tutte_ **le** **notti** anch io **tengo** un diario dei miei **sogni** molto meno **interessante** del suo ha **passato** la **vita** a **sognare** e ad osservare le **persone** a **guardarle** davvero

STT 29 **** *A_2020 *G_21-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna

score: 328,30

mio **padre** mi ha abbandonato alla nascita mia **mamma** era **ragazza** **madre** e da lì **diciamo** siamo andati a **vivere** con i miei **nonni** io e mio **fratello**

STT 54 **** *A_2018 *G_30-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna

score: 291,28

per essere **felice** **cerco** tutti i **giorni** di **stare** con mia **figlia** di **stare** con i miei affetti e **godere** delle **piccole** cose insomma **vivere** la **vita** in modo semplice

STT 123 **** *A_2018 *G_21-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo

score: 243,20

una rossa che può illuminare anche questo **festival** di **Sanremo** che sarà sicuramente un **festival** di grande **successo** ma un premio a Milva **penso** che sia più che dovuto se dovessi **dire** che hai combinato nella **vita** che **diresti** non h

STT 323 **** *A_2019 *G_27-01 *T_TG2 *N_Servizio *P_Fonte *S_Uomo

score 185,52

voi che **vivete** sicuri nelle vostre tiepide **case** voi che **trovate** **tornando** a **sera** il cibo caldo e visi **amici** considerate se questo è un **uomo** che lavora nel fango che non **conosce** pace che lotta per mezzo pane che **muore**

Classe 2: il "discorso scientifico"

La classe 2 raggruppa una serie di parole che fanno prevalentemente riferimento all'informazione sul Covid-19 che nel mese di gennaio 2020 ha cominciato a interessare l'agenda dei notiziari italiani, infatti, come vedremo meglio di seguito, attesta un'associazione significativa con la modalità della variabile Anno *A_2020, con un valore del $\chi^2(1, N= 2.763, ST N=418)=12,99, p<0,05$.

Nella Tabella 8 ci sono parole riferite alla scoperta del coronavirus, come *cinese, Cina, coronavirus, epidemia, animale, Wuhan, identificare*, verbo frequente in molti segmenti di testo tipici di questa classe che riportano discorsi sulla necessità di isolare il virus, come esemplificato dal STT 21 riportato di seguito:

STT 21 **** *A_2020 *G_22-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo

dobbiamo **evitare** che questo nuovo **coronavirus** si **diffonda** quindi è importante come si sta facendo prendere tutti i provvedimenti **necessari** ad **esempio identificare** precocemente **eventuali malati** che **arrivassero** sul nostro **suolo** e isolarli **immediatamente** in modo da **circoscrivere** ogni **focolaio**

Tipica di questa classe è l'occorrenza di parole che fanno riferimento a una più estesa questione medica e di salute – *virus, paziente, medico, temperatura, sintomo, influenza, infezione, malattia, rischio, respiratorio* (Tabella 8), *focolaio, malato, laboratorio, mascherina, tosse, dispositivo, farmaco, prognosi, influenzale, clinico, DNA* (Appendice, Tab. 2) – indicando un prima sovrapposizione dell'informazione specifica sul coronavirus con quella pertinente più in generale ai virus influenzali tipici della stagione invernale e che sono oggetto del discorso anche dei TG del 2018 e 2019, come attestano diversi STT di cui si riporta una selezione nell'Elenco 2.

Una seconda area di sovrapposizione è individuata a partire dalla disambiguazione semantica della parola *temperatura* che ricorre non solo in frammenti di discorso sulla temperatura corporea, dunque sempre riferiti a questioni di salute, ma anche in segmenti di testo che evidenziano un discorso sulle condizioni metereologiche e, più in generale, su questioni ambientali. Indicatori di questo *focus* tematico sono parole come *inverno, freddo, zona, area*, (Tabella 8), *faglia, ghiaccio, riscaldamento, atmosfera, climatico, caldo, allerta, suolo, polvere, nevicata, naturale, plastica, stagione, termico, riciclare, superficie, roccia, nebbia, chimico, istituto, natura, tecnologia, impatto, sismico, ossigeno, neve, energia, pianeta* (Appendice, Tab. 2).

A partire dall'individuazione di questi tre nuclei tematici (coronavirus, influenza, ambiente), la classe 2 può essere definita come la classe del “discorso scientifico”, a cui fanno esplicito riferimento anche due parole chiave come

laboratorio e ricerca (Appendice, Tab. 2). La lettura dei segmenti di testo tipici evidenzia alcune caratteristiche linguistiche che confermano questa interpretazione e, al contempo, consentono di profilare l'identità delle fonti. Gli enunciati della classe 2 sono espressi quasi tutti con forme impersonali (si vedano alcuni esempi riportati nell'Elenco 2) che segnano una marcata distanza dalla narrazione del vissuto personale tipico della classe 1, i cui segmenti di testo tipici attestano una costante autoreferenzialità. Se nella classe 1, l'argomento del discorso è il soggetto parlante, cioè la fonte stessa, che racconta di sé, nella classe 2, il *topic* è una materia, una questione, un evento, di natura scientifica, che non riguarda quasi mai il vissuto personale della fonte. Ci sono alcuni segmenti di testo che testimoniano l'uso di forme verbali personali, ma o al plurale, per esprimere l'appartenenza del/la parlante a un gruppo professionale, come esemplificato nel STT 16, Elenco 2 (*noi infettivologi medici di malattie infettive siamo preoccupati*), o, al singolare, per esprimere un'opinione professionale, come nel STT 19, Elenco 2, o nel più esplicativo STT 182, Elenco 2, o, raramente per raccontare esperienze personali, che si collocano comunque entro l'ambito professionale (es. STT 117, Elenco 2).

Il soggetto tipico degli enunciati della classe 2 è il soggetto di un discorso basato sul sapere e sulla competenza professionale, identificabile nelle fonti esperte tradizionalmente interpellate dai/le giornalisti/e per spiegare gli eventi o i fenomeni di cui parla (Albæk 2011). Infatti questa classe è significativamente associata alla modalità *N_Intervista ($\chi^2(1, N=6.296, ST N=873)=7,23, p<0,05$).

Elenco 2. Selezione di segmenti di testo tipici della classe 2 del *sub-corpus* P_Fonte

STT 1 **** *A_2020 *G_28-01 *T_TG5 *N_Messaggio *P_Fonte *S_Uomo

score: 834,44

sappiamo già molto di questo **coronavirus cinese** anche perché ormai è stato appurato che l'**infezione** si **diffonde** con più **velocità** del previsto e si **trasmette** anche prima che **compaiono** i **sintomi virus cinese** che è stato **immediatamente identificato** fotografato e

STT 5 **** *A_2020 *G_02-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo

score: 621,65

in caso di **influenza** l'**antibiotico** non va mai preso non va fatta la **terapia** fai da te è il **medico** che deve prescrivere se **compare** una superinfezione batterica per **esempio** una bronchite o una broncopolmonite altrimenti **riposo** a letto e un antipiretico se la **temperatura** è troppo **elevata**

STT 7 **** *A_2020 *G_21-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo

score: 589,25

i **sintomi** sono quelli di una **normale** polmonite con **febbre elevata** malessere generale e **tosse** naturalmente da non confondere coi **sintomi influenzali** l'**influenza** è molto **diffusa** in questo **periodo** e nei casi più gravi nei casi critici può **comparire** una **difficoltà respiratoria**

STT 8 **** *A_2019 *G_07-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 563,17

il **freddo** non è di per sé una **condizione** che **causa** l'**influenza** ma può essere una **condizione** che ne favorisce l'insorgenza perché spesso nella **stagione fredda** si sta in **luoghi** chiusi riscaldati e in queste **condizioni** è più facile la **trasmissione** di **virus influenzali** da una persona all'altra

STT 16 **** *A_2018 *G_29-01 *T_TG5 *N_Commento *P_Fonte *S_Uomo
score: 512,67

noi infettivologi **medici** di **malattie** infettive siamo **preoccupati** e vediamo ogni giorno **infezioni** che una volta venivano **trattate** con un solo semplice **antibiotico** e **guarivano** oggi dobbiamo **ricorrere** a 2 3 **antibiotici associazione antibiotici** per la loro guarigione

STT 19 **** *A_2020 *G_31-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 512,67

sotto stretta osservazione in questo **momento** sono innanzitutto i **cinesi** che facevan parte del gruppo in cui erano **presenti** le 2 persone che dove avevano **sviluppato** la **malattia** confermata come **infezione** da **coronavirus** ma io credo che non ci sia assolutamente

STT 23 **** *A_2018 *G_02-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 466,41

i **sintomi** anche quest'anno sono i medesimi degli anni precedenti **tosse** e **febbre** molto **elevata** rinite in molti bambini si accompagnano anche **sintomi** gastrointestinali quali vomito e diarrea è **utile** ricordare che essendo una **malattia** virale per l'**influenza** non

STT 41 **** *A_2019 *G_27-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 404,75

annuale questo è un **gennaio freddo** con **temperatura** al di sotto della **media** finirà così e quindi è un po' in controtendenza come un **ritorno** a quelli che erano gli **inverni** di una volta ma **rientra** sempre in un quadro di **cambiamento climatico**

STT 67 **** *A_2018 *G_06-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 362,86

i **cambiamenti climatici** influiscono soprattutto sul **livello** di aridità e soprattutto estiva e poi anche con degli **eventi** meteorologici **estremi** quindi trombe d'**aria** che creano e minano la stabilità delle **piante** che in molti casi cadono questa è un **area**

STT 117 **** *A_2018 *G_15-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo
score: 264,57

per quanto riguarda i **pazienti** io non truffo mai nessuno perché io **espongo** in **maniera** chiara quali sono i **limiti** e che cosa fa la mia **terapia**

STT 182 **** *A_2020 *G_29-01 *T_TG1 *N_Commento *P_Fonte *S_Uomo
score: 222,65

confidente che il nostro paese ha tutte le competenze per affrontare il problema l'**istituto superiore** di sanità lo **Spallanzani** a Roma il Sacco a **Milano** Elisa Vicenzi qua **vicino** ebbene io penso che queste competenze la dedizione dei **medici** e dei

Classe 3: la “comunicazione politica”

La classe 3 raggruppa parole che fanno riferimento alla vita pubblica, in particolare alla politica, con parole chiave come *governo, politico, 5_Stelle, Salvini, cittadino, elettorale, centro_destra, legge* (Tabella 8), *Partito_Democratico, PD, Forza_Italia, sinistra, parlamentare, governare, cittadinanza, Unione europea, regione, legge, democrazia*, e così via (Appendice, Tab. 3), e all'economia, con parole chiave come *tassa, lavoro* (Tabella 8), *investimento, sviluppo, fiscale, reddito, economia, lavoratore* (Appendice, Tab. 3)

La lettura dei segmenti di testo tipici della classe 3, di cui si riporta una selezione nell'Elenco 3, consentono di riconoscerla come la classe della “comunicazione politica” – significativamente disintermediata, considerando la correlazione con le modalità *N_Messaggio ($\chi^2(1, N=1.875, ST N=902)=119,72, p<0,001$) e *N_Commento ($\chi^2(1, N=125, ST N=64)=10,4 p<0,05$), che marcano frammenti di discorsi pronunciati in un contesto extra-mediale, o comunque extra-TG, e ripresi integralmente o parzialmente dal notiziario, senza una vera e propria mediazione giornalistica, come invece avviene nell'intervista che ha una struttura tipicamente dialogante, fra giornalista e persona intervistata. Le fonti di questo discorso sono identificabili prevalentemente in esponenti di partito o coalizioni politiche, che parlano esplicitamente in nome del gruppo a cui appartengono, esprimendosi in modo autoreferenziale con la seconda persona plurale (es. STT 17, 34, 36, 37, 84, Elenco 3), o a nome proprio, esplicitando la propria identità politica (es. STT 64, 86, 190, Elenco 3).

Elenco 3. Selezione di segmenti di testo tipici della classe 3 del *sub-corpus* P_Fonte

STT 17 **** *A_2018 *G_22-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score 2.060,77

il **Movimento 5_Stelle** si **assume** l'onere di **governare** questo **paese** nel caso in cui non **dovesse** avere la **maggioranza** per **governare** lancerà un **appello** alle altre **forze politiche** per **migliorare** la **qualità** della vita degli **italiani** a partire dai 20 **punti** del nostro **programma**

STT 34 **** *A_2018 *G_29-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 1.927,71

merito esperienza e **rinnovamento** hanno **guidato** la **scelta** dei nostri **candidati** che sono **amministratori** locali **esponenti** del mondo dell' **impresa** del **lavoro** e delle **professioni** e dell

informazione **Forza Italia** ha **scelto** i **migliori** per **vincere** per tornare alla guida del **paese** e **realizzare** così l' **unico progetto politico credibile** per **rilanciare** l' **Italia**

STT 36 **** *A_2020 *G_17-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 1.903,43

la **sinistra** che è **minoranza** nel **paese** vuole una **legge elettorale** che la riporti sempre al **governo** noi invece vogliamo una **legge elettorale** che **garantisca stabilità** e **rispetti** veramente il **voto** degli **italiani**

STT 37 **** *A_2018 *G_14-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 1.887,44

alle **prossime elezioni politiche** il **Movimento 5 Stelle** sarà la **prima forza politica** del **paese** questo **significa** che avremo un gruppo parlamentare che sarà il più **grande** del **Parlamento** della diciottesima **legislatura** chiederemo a **tutte le forze politiche** qualora non **dovessimo** avere la

STT 64 **** *A_2019 *G_12-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 1.788,85

che poi il **governo** valuta e quindi se sulla **TAV** non c'è un **accordo politico** la parola **passa** agli **italiani** io ho ben **chiaro** in testa che **voterei a favore** dello **sviluppo** e della **crescita**

STT 84 **** *A_2020 *G_10-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo
score: 1.704,14

noi abbiamo dato un **contributo** per avvicinare la data delle **elezioni** perché **prima** va a casa questo **governo** di **incapaci** e meglio è per l' **Italia** non per **Salvini** eh meglio è per per l' **Italia** perché stanno soffrendo le aziende **italiane**

STT 86 **** *A_2019 *G_17-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 1.701,93

noi **rappresentiamo** l' **idea liberale** della **politica** i **valori** e le idealità **liberali** che **oggi bisogna difendere** in **Italia** e in **Europa** nel mondo perché io ho **deciso** di **presentarmi** alle **elezioni europee** per portare in **Europa** la mia voce di un **Europa** che **deve** essere **cambiata unita**

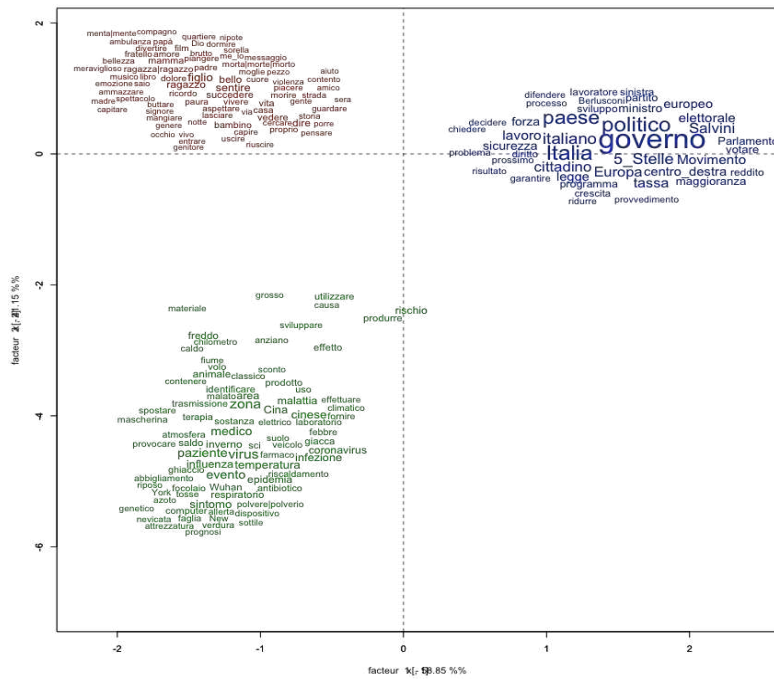
STT 190 **** *A_2018 *G_03-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 1.418,44

su cui **chiederò** un **impegno chiaro** agli **alleati** di **centro_destra** la **prima legge** che il **governo Salvini** **cancella** è la **legge Fornero** su questo non ho dubbio alcuno

IRaMuTeQ utilizza i risultati della CGD anche per un'ACL che implementa l'analisi fattoriale sul *corpus*, assumendo UCE, lemmi e variabili categoriali, come variabili indipendenti, per restituire una proiezione delle classi fattoriali e delle variabili categoriali su un piano cartesiano, dove l'asse delle ascisse e l'asse delle ordinate rappresentano i due principali fattori latenti che spiegano la varianza e la cui natura deve essere interpretata. Il software elabora due grafici distinti, uno con la proiezione delle forme lessicali raggruppate per classe e uno con la proiezione delle variabili categoriali associate alle classi lessicali, ma l'ACL è unica, duplice è soltanto la rappresentazione grafica.

I grafici elaborati nell'ACL della CGD sul *sub-corpus* P_Fonte sono illustrati nel Grafico 7 e nel Grafico 8. In entrambi i grafici, parole e variabili categoriali sono proiettati sul medesimo piano cartesiano, con il fattore 1 che spiega il 58,85% della varianza, sull'asse delle ascisse, e il fattore 2, che spiega il restante 41,15% della varianza, sull'asse delle ordinate. Numerando i quadranti in senso orario, a partire dall'alto a sinistra, il Grafico 7 mostra il raggruppamento delle parole per classi, con la classe 1 posizionata nel primo quadrante (in alto a sinistra), la classe 2 posizionata nel quarto quadrante (in basso a sinistra) e la classe 3 posizionata a destra a cavallo fra il secondo e il terzo quadrante.

Grafico 7. Rappresentazione fattoriale delle 3 classi lessicali del *sub-corpus* P_Fonte



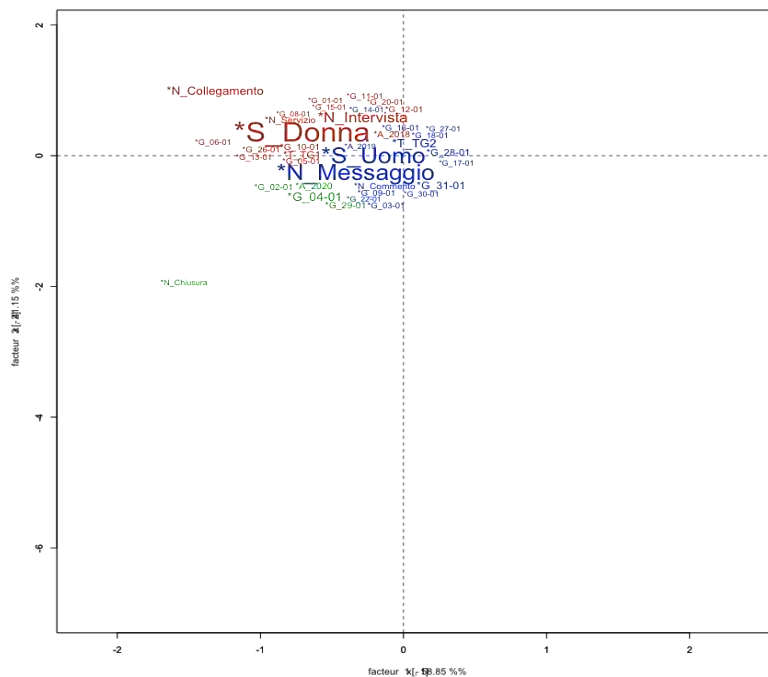
I fattori che sembrano spiegare meglio questa distribuzione sono riconducibili, a una prima analisi, alla polarità cronaca-politica, sull'asse orizzontale, da sinistra a destra, e alla polarità vicino-lontano, sull'asse verticale, dall'alto in basso. Sulla polarità destra dell'asse delle ascisse si posiziona infatti solo la classe 3, che è omogenea per raggruppamento di parole tipiche della politica, mentre sulla polarità sinistra si collocano la classe 1 e 2, entrambe contenenti una parola chiave che indica il riferimento alla dimensione evenemenziale che contraddistingue la

cronaca (vs. la dimensione fenomenologica più tipica della politica), ovvero *succedere* (classe 1) ed *evento* (classe 2).

La polarità vicino-lontano spiega bene la dimensione di lontananza nel senso geografico del termine, a cui fa riferimento la classe 2, totalmente collocata sul versante negativo dell'asse, che contiene il lessico del coronavirus, all'inizio della sua diffusione in Cina; così come spiega il posizionamento centrale della classe 2 che raggruppa parole che fanno riferimento al contesto nazionale sopra l'asse delle ascisse (*Italia, paese, italiano, Salvini*) e la parola *Europa*, sotto l'asse, anche se *europeo* si trova sul versante opposto, ma forse questa collocazione ambivalente dimostra l'ambivalente senso di appartenenza/vicinanza ed estraneità/lontananza della politica italiana rispetto all'Europa, e in particolare all'Unione europea.

Per quanto riguarda la classe 1, la sua collocazione sul versante della vicinanza può essere intesa in senso di prossimità, come riferimento a eventi che riguardano gli ambienti in cui si svolge la vita domestica – come la *casa* (Tabella 8, la *strada*, il *quartiere*, la *scuola*, il *negozio*, il *supermercato* (Appendice, Tab. 1).

Grafico 8. Rappresentazione fattoriale dalle variabili associate alle 3 classi lessicali del *sub-corpus* P_Fonte



Il Grafico 8 illustra chiaramente l'associazione della modalità *S_Donna con la classe 1 e l'associazione della variabile *S_Uomo con la classe 3. Nel contempo evidenzia uno spostamento di tutte le classi sulla polarità sinistra dell'asse orizzontale, indicando una componente cronachistica anche della politica.

Venendo ora alla CGD del *sub-corpus* PF_Donna, l'analisi ha prodotto come risultato due classi lessicali che coprono il 91,80% del *corpus*, classificando N=1.949 segmenti di testo su N=2.123 totali, come illustrato nella tabella seguente che riporta le statistiche generali risultanti dall'analisi.

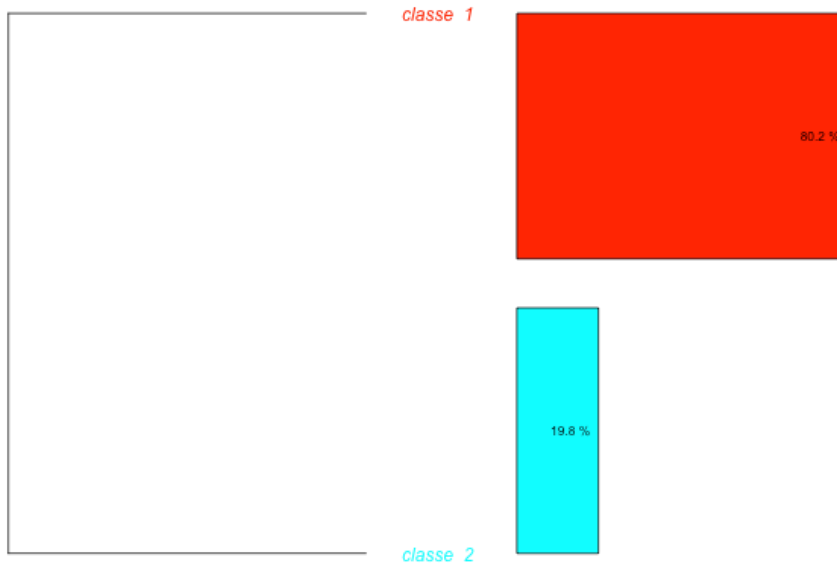
Tabella 9. Statistiche: risultati della CGD del *sub-corpus* PF_Donna

Testi (N)	1.944
Segmenti di testo (N)	2.123
Forme (N)	7.837
Occorrenze (N)	51.547
Lemmi (N)	5.263
Forme lessicali (N)	4.818
Forme supplementari	385
Forme lessicali con frequenza ≥ 3	1.515
Media di forme per segmento	24,28
Classi (N)	2
Segmenti classificati (91,80%)	1.949

Il Grafico 9 mostra la struttura della relazione fra le classi, evidenziando una simmetria strutturale e una asimmetria dimensionale. Le due classi sono infatti in posizione simmetrica fra loro, ma attestano una dimensione diversa: la classe 1 copre l'80,25% dei segmenti di testo (N=1.564), la classe 2 copre il restante 19,75% delle UCE (N=385).

La classe 1 è significativamente associata alle modalità *N_Collegamento ($\chi^2(1, N=40)=7,67, p<0,05$) e *A_2018 ($\chi^2(1, N=718)=6,63p<0,05$). La classe 2 è significativamente associata alle modalità *N_Commento ($\chi^2(1, N=11)=8,45, p<0,05$), *T_TG2 ($\chi^2(1, N=568)=8,15, p<0,05$) e *A_2020 ($\chi^2(1, N=632)=4,35, p<0,05$).

Grafico 9. Dendrogramma delle classi lessicali del *sub-corpus* PF_Donna



Le forme (lemmatizzate) attive significative per la classe 1 sono N=38 riportate integralmente nella tabella seguente, in ordine decrescente per x^2 .

Tabella 10. Forme lessicali significative della classe 1 del *sub-corpus* PF_Donna

ID	Lemma	ST (N)	Occ. (N)	x^2	p-value
0	figlio	89	91	18,56	<0,0001
1	vita	77	80	13,479	0,00024
2	dire	190	213	12,099	0,00050
3	bambino	54	55	11,485	0,00070
4	paura	41	41	10,31	0,00132
5	bello	41	41	10,31	0,00132
6	sentire	70	74	9,99	0,00157
7	anno	132	147	9,147	0,00249
8	scuola	36	36	9,029	0,00265
9	mamma	34	34	8,518	0,00351
10	arrivare	57	60	8,501	0,00354
11	capire	40	41	7,921	0,00488
12	persona	97	107	7,737	0,00541
13	ragazzo	39	40	7,669	0,00561
14	sera	28	28	6,993	0,00818
15	parlare	52	56	5,784	0,01616
16	uomo	23	23	5,729	0,01668
17	terra	23	23	5,729	0,01668
18	situazione	23	23	5,729	0,01668
19	vedere	86	96	5,553	0,01844
20	entrare	22	22	5,477	0,01926
21	guardare	30	31	5,429	0,01980
22	lasciare	21	21	5,226	0,02225
23	messaggio	20	20	4,974	0,02572
24	ricordo	20	20	4,974	0,02572
25	morire	20	20	4,974	0,02572
26	mano	19	19	4,723	0,02575

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
27	dolore	19	19	4,723	0,02575
28	chiamare	27	28	4,693	0,03028
29	pronto	18	18	4,472	0,03444
30	macchina	18	18	4,472	0,03444
31	treno	17	17	4,222	0,03991
32	porre	25	26	4,207	0,04026
33	casa	61	68	3,977	0,04611
34	permettere	16	16	3,971	0,04628
35	marito	16	16	3,971	0,04628
36	difficile	24	25	3,965	0,04645
37	trovare	55	61	3,907	0,04807

Classe 1: le “testimonianze di vita”

L'elenco delle parole significativamente associate alla classe 1 comprende un primo gruppo di nomi che denotano varie identità sociali: dall'identità più generale della *persona*, a quelle familiari (*figlio*, *mamma*, *marito*) a quelle generazionali (*bambino*, *ragazzo*) fino a quella di genere maschile (*uomo*). Manca invece la parola *donna*, che è presente nell'elenco delle parole raggruppate nella classe 1, ma con un valore non significativo: χ^2 (1, N=46, ST N=41)=2,35, p=0,126).

C'è poi un nucleo di parole che fa riferimento alla vita quotidiana: *casa* e *scuola*, luoghi in cui un'ampia parte della popolazione trascorre la propria giornata; *macchina* e *treno*, mezzi di trasporto utilizzati per gli spostamenti, rappresentati anche dai due verbi di movimento *arrivare* e *entrare*. Questa dimensione, che possiamo definire ordinaria, è rappresentata contestualmente alla dimensione straordinaria della vita e della morte: *vita* è la seconda parola nell'elenco, e *morire* è uno dei verbi significativamente associati a questa classe, che si contraddistingue anche per la presenza di parole che evocano sentimenti o situazioni negative, come *paura*, *dolore*, *difficile*.

Scuola e *casa* possono essere inclusi anche in un altro insieme di nomi che indicano una più generale dimensione spazio-temporale, *anno*, *sera* e *situazione*, indicativi di una narrazione evenemenziale, tipica della cronaca, vs. una narrazione fenomenologica, caratteristica dei servizi di approfondimento.

I verbi di percezione *sentire* e *vedere*, come già osservato più sopra, insieme al verbo *guardare* e al nome *ricordo*, sono indicatori di un discorso di testimonianza,

che è quello che sembra connotare questa classe lessicale. Leggendo i segmenti di testo che concorrono a definirla, risulta infatti evidente che si tratta perlopiù di frammenti di interviste a persone interpellate per portare la propria testimonianza, narrare il proprio vissuto personale rispetto a un evento di cui sono state direttamente protagoniste, e che si connota spesso di sentimenti negativi.

In molti casi è possibile identificare la fonte dei segmenti di testo in madri che hanno vissuto il dolore della perdita di un/a figlio/a (es. STT 1, 3, 8, Elenco 4), madri che raccontano della malattia dei propri figli (es. STT 11, Elenco 4) o della violenza subita in ambito domestico (es. STT 14, 50, 122, Elenco 4), oppure ancora madri che chiedono giustizia per la morte, evidentemente in circostanze criminose, dei/lle propri/e figlie/e (es. STT 1, 3, 35, Elenco 4), oppure madri che chiedono protezione, sempre per i/le propri/e figli/e, che percepiscono in una situazione di pericolo (es. STT 43, Elenco 4). Poi ci sono testi che evocano una maternità meno drammatica tuttavia spesso difficoltosa (es. STT 4, 25, 27, 98, 144, Elenco 4), e più raramente appagante (es. STT 25, 144, Elenco 4).

In altri casi è possibile identificare la fonte in figlie che raccontano la gioia di un rapporto positivo con la madre (es. STT 9, Elenco 4) o figlie che raccontano la perdita del padre (es. STT 45, Elenco 4).

Infine ci sono alcuni testi in cui la senatrice a vita Liliana Segre è chiaramente identificabile come fonte che ricorda la sua espulsione dalla scuola all'entrata in vigore delle leggi razziali in Italia (es. STT 2, Elenco 4), evento coperto dai notiziari per celebrare la giornata della memoria che ricorre ogni 27 gennaio, anche attraverso altre fonti che ricordano la storia di Liliana Segre (es. ST 198, Elenco 4). *Ricordo*, peraltro è una parola fra le più significative di questa classe lessicale.

Elenco 4. Selezione di segmenti di testo tipici della classe 1 del *sub-corpus* PF_Donna

STT 1 **** *A_2018 *G_27-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 78,39

una **persona** che va a lavorare ha diritto di avere tutela ha diritto di tornare a **casa** alla **sera** io mia **figlia** non la **vedrò** più non la **sentirò** più però vi supplico indagate ci sarà voglio giustizia continuerò a **dire** finché ho un alito di **vita** voglio giustizia per la memoria di mia **figlia**

STT 2 **** *A_2018 *G_25-01 *T_TG2 *N_Messaggio *P_Fonte *S_Donna

score: 59,79

mi **disse** quella **sera** di fine estate mio **papà** che ero stata espulsa dalla **scuola** beh lo stesso Stato italiano che ha espulso quella **bambina** per la colpa d essere nata e non perché avete fatto qualcosa di male a **scuola** e oggi a 87 **anni** mi **vedo aprire** le porte del senato

STT 3 **** *A_2018 *G_03-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 59,57

tre **ragazzi** che in 3 facevano 64 **anni** ci davo l ergastolo a **vita** perché io mi addormento la **sera** mi ci sveglio la **mattina** sapendo che mio **figlio** sta là l unica consolazione che vado a portà fiori e preghiere

STT 4 **** *A_2020 *G_16-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 57,95

mi auguro per loro insomma che **trovi** strade migliori ecco perché una soluzione c è non è **vita** semplice per nessuno di noi ognuno di noi deve **alzarsi la mattina** e decidere **vivo** o **muoio** o voglio **vivere** perché ho 2 **figli** però voglio **cercare** di farlo con decoro e dignità insomma

STT 8 **** *A_2019 *G_06-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 55,07

una **bambina** nella stessa stanza di nostro **figlio** il 30 **mattina** non c era più e c è stato **detto** che non era **riuscita** a sopravvivere e il **giorno** precedente la morte del nostro piccolo anche il nostro vicino di **letto** non

STT 9 **** *A_2020 *G_19-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna
score: 53,59

la **vita** mi ha regalato intanto dei genitori meravigliosi una **mamma** che io ho adorato ma soprattutto il regalo più **bello** mia **figlia** e che mi ha regalato questi due **nipoti** che son tremendi

STT 11 **** *A_2019 *G_12-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 52,17

mia **figlia** ha 11 **anni** frequenta questo istituto da 2 **anni** io ho visto che c era un regresso nella **bambina** perché anche quando **iniziamo a mangiare** e io mi avvicinavo per dare la posata mia **figlia** faceva con le mani così perché purtroppo non **parla**

STT 14 **** *A_2018 *G_22-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna
score: 46,95

lei sapeva aveva **cercato** di chiudere per la prima **figlia** questi discorsi questa aveva già **detto** alle **ragazze** fate attenzione a non rimanere sole con **papà** quindi lei sapeva e quindi tu non **denunci** alla fine cerchi di lavare i panni sporchi in **casa**

STT 25 **** *A_2018 *G_30-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna
score: 41,96

per essere felice **cerco** tutti i **giorni** di stare con mia **figlia** di stare con i miei affetti e godere delle piccole cose insomma **vivere** la **vita** in modo semplice

STT 27 **** *A_2020 *G_16-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna
score: 41,67

e comunque in qualche **maniera** in 3 **anni** di lavori precari sono rimasta sempre in piedi i miei **figli** sono entusiasti di questa **mamma** forse loro mi hanno ritrovata perché sono i rimproveri che faccio anche a loro **studiate** impegnatevi il senso della responsabilità

STT 35 **** *A_2019 *G_28-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna
score: 39,20

confermati 30 **anni** è quello che mi aspettavo quello che volevo quello che volevo per mia **figlia** per e soprattutto per la **bambina**

STT 43 **** *A_2019 *G_25-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 36,65

ci vogliono le forze dell'ordine la protezione qualcuno ci deve tutelare ci devono proteggere basta i miei **figli** stanno spaventati per colpa di chi fatemi **vedere** se avete il **coraggio** fatevi avanti così li devo crescere nella **paura**

STT 45 **** *A_2020 *G_10-01 *T_TG1 *N_Messaggio *P_Fonte *S_Donna

score: 36,46

sono Daniela Marcone la **figlia** di Francesco Marcone un **uomo** dello Stato ucciso 25 **anni** fa qui a Foggia e perché la sua memoria la sua morte non siano inutili **bisogna** costruire nell'oggi la reazione di questa città

STT 50 **** *A_2018 *G_14-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 35,92

lui mi gridava mi ha puntato la pistola di fronte **dicendomi** era te che dovevo **colpire** poi mentre io gridavo e chiedevo aiuto lui è rientrato dentro e ha **colpito** mio **figlio** al **cuore** per ucciderlo

STT 98 **** *A_2018 *G_29-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna

score: 30,24

sono andata in **ospedale** parecchie volte ho **lasciato** a **casa** mio **figlio**

STT 122 **** *A_2018 *G_10-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna

score: 27,16

soprattutto per una **donna** incinta il **ricordo** è pesante non **dirmelo** poi tanto ricordarmi lui m ha dato delle bastonate sulla schiena sapendo che ero incinta e sono stati molto violenti ma per loro noi eravamo animali non **persone** quando ho

STT 144 **** *A_2020 *G_25-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna

score: 25,78

avevo **paura** di perdere il lavoro e di non essere riconfermata e quindi di ritrovarmi sì con un **bambino** a **casa** felice perché stavo costruendo insomma la mia famiglia però al tempo stesso disoccupata

STT 198 **** *A_2020 *G_27-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna

score: 23,37

la **storia** di Liliana Segre come **bambina** a cui è stato vietato il diritto di andare a **scuola** Anna Frank la **bambina** a cui è stato negato il diritto della libertà i **bambini** del ghetto di Varsavia al quale è stato negato il diritto alla salute

Classe 2: la "comunicazione politica"

La classe 2 attesta un numero di forme (lemmatizzate) attive significative pari a N=340. Essendo un numero elevato, se ne riportano qui le prime 40 per valore di associazione (χ^2) alla classe. Queste selezioni coprono un po' più di un terzo delle occorrenze dei lemmi nei segmenti di testo che formano questa classe⁹¹. L'elenco completo è riportato in Appendice (Tab. 4).

⁹¹ Occorrenze in ST dei 40 lemmi più significativamente associati a questa classe (N=837)/totale occorrenze in ST di lemmi con $\chi^2 > 3,84$, $p < 0,05$ (N=2.387) = 36%.

Tabella 11. Prime 40 forme lessicali significative della classe 2 del *sub-corpus* PF_Donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	governo	64	70	235,31	<0,0001
1	Italia	57	73	162,78	<0,0001
2	5_Stelle	34	34	140,57	<0,0001
3	Movimento	30	30	123,78	<0,0001
4	tassa	28	28	115,4	<0,0001
5	paese	47	70	102,86	<0,0001
6	politico	40	55	100,19	<0,0001
7	prossimo	28	31	98,96	<0,0001
8	Salvini	28	31	98,96	<0,0001
9	centro_destra	24	24	98,71	<0,0001
10	elettorale	26	28	95,77	<0,0001
11	impresa	20	21	76,31	<0,0001
12	cittadino	30	41	75,39	<0,0001
13	campagna	17	17	69,67	<0,0001
14	Fratelli_d	18	19	68,06	<0,0001
15	europeo	19	21	66,98	<0,0001
16	italiano	40	70	64,03	<0,0001
17	Berlusconi	15	15	61,41	<0,0001
18	Europa	19	23	58,01	<0,0001
19	maggioranza	15	16	55,72	<0,0001
20	Forza_Italia	15	16	55,72	<0,0001
21	votare	16	18	54,78	<0,0001
22	legge	19	24	54,11	<0,0001
23	Lega	13	13	53,17	<0,0001
24	Emilia	12	12	49,05	<0,0001
25	riforma	13	14	47,54	<0,0001
26	sviluppo	13	14	47,54	<0,0001
27	tema	14	16	46,71	<0,0001
28	elezione	11	11	44,94	<0,0001
29	Romagna	11	11	44,94	<0,0001
30	coalizione	11	11	44,94	<0,0001
31	governare	11	11	44,94	<0,0001
32	forza	23	37	42,79	<0,0001
33	partito	15	19	42,41	<0,0001
34	prescrizione	10	10	40,83	<0,0001
35	decreto	10	10	40,83	<0,0001
36	accordo	11	12	39,39	<0,0001
37	Parlamento	11	12	39,39	<0,0001
38	processo	15	20	38,91	<0,0001
39	lavoratore	12	14	38,7	<0,0001

Tutte le parole distintive della classe 2 del *sub-corpus* PF_Donna fanno riferimento al mondo della politica, in particolare nazionale, con la parola *Italia* al primo posto in termini di occorrenza assoluta (Tabella 11), anche per via della sua associazione al nome *Fratelli d'Italia* che nel corso del processo di lemmatizzazione

automatica è stato diviso di *Fratelli_d* e *Italia*⁹². Nell'elenco si possono individuare gruppi di parole che fanno riferimento alle diverse tipologie di *issues* tipiche della comunicazione politica (per un'introduzione alla comunicazione politica e alla classificazione fra diverse tipologie di *issues* qui proposta si veda per Mazzoleni 2012). Un primo gruppo di parole fa riferimento a relazioni e dinamiche di partito (le cosiddette *political issues*⁹³): *5_Stelle*, *Movimento*, *centro_destra*, *Fratelli_d*, *maggioranza*, *Forza_Italia*, *Lega*, *forza*, *accordo*. Un secondo gruppo di parole fa riferimento a specifici appuntamenti elettorali (*campaign issues*) incluse le elezioni regionali in Emilia Romagna che si sono svolte il 26 gennaio 2020: *elettorale*, *campagna*, *votare*, *Emilia*, *elezione*, *Romagna*, *coalizione*. Un terzo gruppo di parole si riferisce alle politiche governative o amministrative (*policy issues*), con termini generici come *tema*, *governare* e *decreto*, e termini più specifici indicativi dei diversi settori di intervento. Fra le parole significative della classe 2 troviamo infatti termini afferenti al mondo dell'economia e del lavoro (*tassa*, *impresa*, *sviluppo*, *lavoratore*, della giustizia (*prescrizione*, *processo*), delle riforme istituzionali (*riforma*, *elettorale*), dell'immigrazione (*immigrazione*, *immigrato*, *migrare*, *accoglienza*, *Sea_Watch*, Appendice, Tab. 4), della gestione delle infrastrutture autostradali (*concessione*, *pedaggio*, *Benetton*, *infrastruttura*, *autostrada*, *galleria*, Appendice, Tab. 4). Infine, i nomi di politici come *Salvini*, *Berlusconi* (Tabella 11), *Renzi*, *Meloni*, *Silvio*, *Matteo*, *Bonaccini*, *Macron*, *Di Maio*, *Fornero* (Appendice, Tab. 4) completano il quadro della "comunicazione politica" veicolata dalle fonti femminili nei TG italiani indicando la presenza anche delle *personal issues*, focalizzate sulla vita e l'attività politica, ma anche professionale e privata, dei/le leader politici/che.

Elenco 5. Selezione di segmenti di testo tipici della classe 2 del *sub-corpus* PF_Donna

STT 1 **** *A_2018 *G_04-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna
score: 966,07

sono le questioni imprescindibili che **Fratelli_d Italia** porterà la **prossima settimana** al **tavolo** con **Forza_Italia** e con la **Lega nord** e sono per noi **fondamentali** per **chiudere** qualunque **alleanza** insieme all'impegno da parte di tutte **le forze politiche** a non fare in nessun caso inciuci con le **2 facce** della **sinistra PD** e **Movimento 5_Stelle**

⁹² Per via dei limiti della lemmatizzazione automatica, di cui si è scritto più sopra, il nome del partito Fratelli d'Italia è stato oggetto di un "errore" di cui qui si tiene conto.

⁹³ Per un'introduzione alla comunicazione politica e alla classificazione fra diverse tipologie di *issues* qui proposta si veda per esempio Mazzoleni (2012).

STT 3 **** *A_2019 *G_25-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna

score: 877,46

ci spieghi **Salvini** come pensa di **mandare** le persone della **Sea_Watch** in altri **paesi europei** quando al **Parlamento Europeo Lega** e **5_Stelle** hanno **votato** contro la **riforma** di **Dublino** che **obbliga** l'**Italia** ad **accoglierli** e visto che in **Europa** noi siamo ormai isolati

STT 5 **** *A_2019 *G_08-01 *T_TG5 *N_Intervista *P_Fonte *S_Donna

score: 800,49

il **provvedimento** adottato dal **governo** era **necessario** per difendere **cittadini** famiglie **imprese** e loro risparmi è vergognoso pensare a quanto questi **temi** siano stati strumentalizzati dal **Movimento 5_Stelle** e dalla **Lega** che **oggi** si ritrovano fare esattamente le stesse cose che hanno fatto i **governi** di **sinistra** il **Movimento 5_Stelle** perde la **faccia**

STT 17 **** *A_2020 *G_16-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 671,81

dopo la pronuncia della **consulta** la **maggioranza** **prosegue** verso una **legge elettorale** di tipo **proporzionale** per il **Movimento 5_Stelle** è importante che i **cittadini** si sentano **rappresentati** e con una **soglia** di sbarramento **evitiamo** che **partiti** piccoli possano destabilizzare il parlamento

STT 18 **** *A_2020 *G_13-01 *T_TG5 *N_Messaggio *P_Fonte *S_Donna

score: 671,39

c è tanto **lavoro** da fare col **Movimento 5_Stelle** siamo determinati ad andare avanti pensiamo ad **abbassare** le **tasse** **abbassare** i **pedaggi** delle **autostrade** oltre che ad **ottenere** la **revoca** delle **concessioni** e **avviare** la **commissione** d **inchiesta** sulle **banche** perché i **cittadini** hanno **bisogno** di **chiarezza** e **tutela**

STT 32 **** *A_2020 *G_17-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 608,85

la **sinistra** che è minoranza nel **paese** vuole una **legge elettorale** che la riporti sempre al **governo** noi invece vogliamo una **legge elettorale** che **garantisca** stabilità e rispetti veramente il **voto** degli **italiani**

STT 41 **** *A_2020 *G_24-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 599,07

se anche in queste **elezioni regionali** **segnatamente** qui in **Emilia Romagna** il **centro_destra** dovesse **vincere** gli **italiani** dovessero dire che non vogliono questo **governo** noi **chiederemo elezioni anticipate**

STT 45 **** *A_2018 *G_13-01 *T_TG2 *N_Intervista *P_Fonte *S_Donna

score: 586,79

gli **italiani** il 4 **marzo** avranno 2 scelte la **scelta** del **centro_destra** che è l **unica** tra le 3 **coalizioni** in campo che può **raggiungere** una **soglia** tale da avere un **governo** che è frutto di una **scelta** dei **cittadini** e non frutto magari di un gioco di **palazzo** o l'ingovernabilità

STT 46 **** *A_2019 *G_29-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 586,26

perché **Salvini** ha **agito** secondo quello che gli è stato **chiesto** dai **cittadini italiani** e perché ha **agito** secondo quella che era la **linea politica** del **governo**

STT 49 **** *A_2019 *G_01-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna

score: 579,70

per **Forza_Italia** il 2019 sarà l'anno della **svolta** in **Parlamento** e nelle **piazze** saremo il punto di **riferimento** di famiglie **imprese lavoratori** che non si **riconoscono** nelle **politiche** di questo **governo**

La CGD del *sub-corpus* PF_Uomo mostra tre classi lessicali che coprono il 95,38% del *corpus*, classificando N=6.278 segmenti di testo su N=6.582 totali, come illustrato nella tabella seguente che riporta le statistiche generali risultanti dall'analisi.

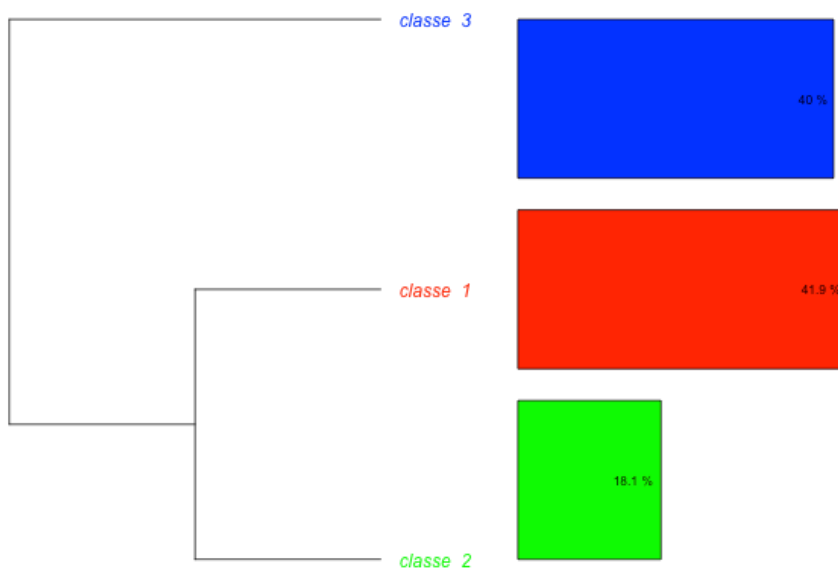
Tabella 12. Statistiche: risultati della CGD del *sub-corpus* PF_Uomo

Testi (N)	5.904
Segmenti di testo (N)	6.582
Forme (N)	17.314
Occorrenze (N)	180.328
Lemmi (N)	10.807
Forme lessicali (N)	10.182
Forme supplementari	549
Forme lessicali con frequenza ≥ 5	2.229
Media di forme per segmento	27,40
Classi (N)	3
Segmenti classificati	6.278

Il Grafico 10 illustra la struttura della relazione fra le classi, evidenziando la posizione gerarchica della classe 3, e la posizione subordinata delle classi 1 e 2 che sono in posizione simmetrica fra loro. In termini dimensionali, la classe maggiore è la 1, che copre il 41,92% dei segmenti di testo (N=2.632), la classe 3 attesta una dimensione dello stesso ordine di grandezza, coprendo il 39,96% delle UCE (N=2.509), infine, la classe 2 attesta una dimensione minore, con una copertura del 18,11% dei testi (N=1.137).

Quanto alle associazioni con le modalità delle variabili categoriali: la classe 1 risulta significativamente correlata *N_Intervista ($\chi^2(1, N=4.539)=56,1, p<0,001$), *N_Collegamento ($\chi^2(1, N=39)=14,38, p<0,05$), *N_Servizio ($\chi^2(1, N=64)=11,24, p<0,05$) e *T_TG1 ($\chi^2(1, N=2.482)=8,41, p<0,05$); la classe 2 è significativamente associata alle modalità *T_TG1 ($\chi^2(1, N=2.482)=38,43, p<0,001$); *A_2020 ($\chi^2(1, N=2.111)=10,04, p<0,05$); *N_Intervista ($\chi^2(1, N=4.539)=6,93, p<0,05$); la classe 3 è significativamente associata alle modalità *N_Messaggio ($\chi^2(1, N=1.520)=114,03, p<0,001$), *T_TG2 ($\chi^2(1, N=1.9732)=30,49, p<0,001$); *T_TG5 ($\chi^2(1, N=1.823)=7,56, p<0,05$) e *A_2019 ($\chi^2(1, N=2.197)=4,43, p<0,05$).

Grafico 10. Dendrogramma delle classi lessicali del *sub-corpus* PF_Uomo



Classe 1: le “testimonianze di vita”

Le forme (lemmatizzate) attive e significative per la classe 1 sono complessivamente N=348, di seguito si riportano le prime 40 per indice di significatività (χ^2), che coprono circa un terzo delle occorrenze nei segmenti di testo che formano la classe⁹⁴.

Tabella 13. Prime 40 forme lessicali significative della classe 1 del *sub-corpus* PF_Uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	dire	397	687	79,73	<0,0001
1	bello	83	97	77,07	<0,0001
2	ragazzo	85	101	75,2	<0,0001
3	sentire	120	162	70,59	<0,0001
4	figlio	85	106	64,84	<0,0001
5	via	74	93	54,95	<0,0001
6	vedere	169	271	48,59	<0,0001
7	vivere	75	100	45,66	<0,0001
8	fratello	30	31	38,5	<0,0001
9	uscire	70	96	38,46	<0,0001
10	film	31	33	36,86	<0,0001
11	succedere	86	126	36,61	<0,0001
12	ragazza	30	32	35,48	<0,0001
13	padre	35	40	34,34	<0,0001
14	moglie	30	33	32,69	<0,0001
15	bambino	98	152	32,53	<0,0001
16	morire	45	57	32,38	<0,0001

⁹⁴ Occorrenze in ST dei 40 lemmi più significativamente associati a questa classe (N=2433)/totale occorrenze in ST di lemmi con $\chi^2 > 3,84$, $p < 0,05$ (N=7.263) = 33,5%.

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
17	buttare	23	23	31,98	<0,0001
18	amore	25	26	31,53	<0,0001
19	storia	62	87	31,19	<0,0001
20	casa	90	139	30,41	<0,0001
21	vita	102	162	30,23	<0,0001
22	mamma	24	25	30,15	<0,0001
23	divertire	20	20	27,79	<0,0001
24	cuore	38	48	27,55	<0,0001
25	paura	58	83	27	<0,0001
26	guardare	59	85	26,74	<0,0001
27	aspettare	56	80	26,23	<0,0001
28	occhio	25	28	25,91	<0,0001
29	strada	63	93	25,84	<0,0001
30	spettacolo	20	21	24,6	<0,0001
31	notte	30	37	23,44	<0,0001
32	quartiere	19	20	23,21	<0,0001
33	musicco	21	23	23,12	<0,0001
34	bellezza	16	16	22,22	<0,0001
35	porre	59	89	22,02	<0,0001
36	dolore	27	33	21,68	<0,0001
37	morta	23	27	20,84	<0,0001
38	piangere	15	15	20,83	<0,0001
39	papà	15	15	20,83	<0,0001

La classe 1 del *sub-corpus* PF_Uomo presenta alcuni tratti comuni con la classe 1 del *sub-corpus* PF_Donna, con la ricorrenza di un primo insieme di nomi che denotano identità basate su relazioni familiari: *figlio* e *mamma* che sono presenti e correlati in modo significativo ai segmenti della classe 1 nel *sub-corpus* PF_Donna; *fratello*⁹⁵, *padre* e *papà*, che sono presenti ma non significativamente correlati alle UCE della classe 1 del *sub-corpus* PF_Donna, e infine *moglie*, che non è presente nella classe 1 del *sub-corpus* PF_Donna, ma, dal punto di vista semantico, è il corrispettivo femminile di *marito*, che invece è significativo per la classe 1 del *sub-corpus* PF_Donna. Un secondo tratto comune alle due classi 1 è la presenza di nomi che fanno riferimento a identità generazionali (*ragazzo*, *ragazza* e *bambino*). Un terzo elemento comune è la presenza della parola *casa*, che qui però non è compresente con la parola *scuola*, ma altri due luoghi pubblici entro cui si svolge la vita quotidiana: la *strada* e il *quartiere* (assenti nella classe 1 del *sub-corpus* PF_Donna). Anche qui, ci sono poi parole che evocano la vita e la morte (*vivere*, *morire*, *morta*), ovvero quella dimensione straordinaria che anche nel *sub-*

⁹⁵ Presente ma non significativo della classe 1 del *sub-corpus* PF_Donna.

corpus PF_Donna risulta contestuale alla narrazione della vita quotidiana. E poi le parole delle emozioni e dei sentimenti: *paura* e *dolore*, frequenti e significative anche nella classe 1 delle fonti femminili, più *amore*, *cuore* e *piangere* (presenti ma non significativamente correlate alla classe 1, PF_Donna), *divertire* (non presente nella classe 1, PF_Donna).

Il verbo *dire*, il più frequente e significativo per questa classe, insieme alla parola *storia*, è indicativo di un'area lessicale che, come la corrispettiva nel *sub-corpus* delle fonti femminili, rappresenta la dimensione della narrazione, specificamente basata sull'esperienza, come indicano i verbi di *sentire*, *vedere*, *guardare*, e *succedere*.

Le parole fin qui elencate, insieme ai segmenti di testo tipici di questa classe, di cui si riporta una selezione nell'Elenco 6, delineano un profilo della classe 1 del *sub-corpus* PF_Uomo molto simile a quello della corrispettiva classe del *sub-corpus* PF_Donna, che è la classe lessicale della testimonianza, dove molte fonti sono identificabili in madri o figlie che raccontano esperienze di vita legate ai propri familiari. Qui ovviamente le fonti non sono identificabili in madri o figlie, o donne soggetto narrante della propria esperienza, ma il loro contraltare maschile: padri o figli, o uomini che raccontano eventi di cui sono testimoni. A questa tipologia di fonte, se ne aggiunge una seconda, emersa nella classe 1 della CGD del *sub-corpus* P_Fonte, ma non nella classe 1 della CGD del *sub-corpus* PF_Donna: le voci del mondo dell'arte e dello spettacolo.

Elenco 6. Selezione di segmenti di testo tipici della classe 1 del *sub-corpus* PF_Uomo

STT 3 **** *A_2019 *G_16-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo

score: 316,39

sì avevo mia **moglie** e mia **figlia** nel sedile posteriore questo **ragazzo** ha bloccato la **strada** con un **secondo sacco** ci ha **fermati** chiedendoci i soldi per il pedaggio se avessi avuto qualcuno alle **spalle** non so cosa sarebbe potuto **succedere** attualmente mia **moglie** ha **paura** di **passare** in quel punto

STT 5 **** *A_2020 *G_15-01 *T_TG5 *N_Messaggio *P_Fonte *S_Uomo

score: 292,10

quando **telefonò** a **casa** e mia **madre disse** c'è Sergio Leone al **telefono** io **sentii** veramente il **cuore** che andava a mille e lui mi **disse** ti voglio **vedere** lo andai a **trovare** sudato e lui mi **disse** dopo un

STT 9 **** *A_2020 *G_08-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo

score: 275,58

è un rapporto molto **bello** molto emozionante tra **padre** e **figlia** e ha il lascito che questa cosa significa forse l'ho *subito* così in quanto **padre** di **figlie** però però devo **dire** che ho **pensato** anche moltissimo al rapporto con mio **padre**

STT 10 **** *A_2018 *G_23-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 269,51

sentendo i boati di tutti siamo **usciti** fuori abbiamo visto la **figlia** che **piangeva** e **gridava** che il **papà** stava **ammazzando** la **mamma**

STT 15 **** *A_2020 *G_20-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 262,02

è molto **bello** quando si passa qui per la **strada** e si **vede** che in questa era bibliocabina per esempio c'è una **mamma** con un **bambino** che sfoglia che sfoglia un **libro** e dietro i **libri** c'è la volontà di un **quartiere** di emanciparsi evidentemente a qualcuno dà **fastidio**

STT 22 **** *A_2020 *G_21-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 241,90

io l'**aspettavo** con Marcello Mastroianni che fumava con tranquillità **contento** come un **ragazzo** come se avesse **preso** lui il quinto Oscar alla carriera Marcello e **arrivò** con Sophia Loren sottobraccio Sophia Loren **teneva** l'Oscar mi **guardò** e mi **disse**

STT 24 **** *A_2020 *G_14-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo
score: 239,29

i **ragazzi** e gli adolescenti sono andati **via** da Facebook soprattutto perché siamo **arrivati** noi **genitori** siamo **arrivati** noi **genitori** mettendo il like ai loro post condividendo magari le loro **foto** e soprattutto pubblicando noi le **foto** dei nostri **figli** e danneggiando quella che loro definiscono ed è un loro diritto una **sorta** di reputazione

STT 31 **** *A_2018 *G_26-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 231,26

sono stato dalla **mattina** dalle 7 e 30 7 e 45 sulla lì sulla banchina dei **treni** per **vedere** la **gente** che **portavano via** se c'era mia **figlia** e mia **figlia** era già **morta** la dentro

Classe 2: il “discorso scientifico”

Le forme (lemmatizzate) attive significative per la classe 2 sono complessivamente N=447, di seguito si riportano le prime 40 per indice di significatività (x^2), che coprono circa un quarto delle occorrenze delle forme lessicali significative nei segmenti di testo che formano la classe⁹⁶.

Tabella 14. Prime 40 forme lessicali significative della classe 2 del *sub-corpus* PF_Uomo

ID	Lemma	ST (N)	Occ.(N)	x^2	p-value
0	zona	35	44	112,76	<0,0001
1	virus	29	34	104,04	<0,0001
2	infezione	20	20	90,72	<0,0001
3	Cina	24	29	82,1	<0,0001
4	situazione	52	100	78,69	<0,0001

⁹⁶ Occorrenze in ST dei primi 40 lemmi (N=801)/totale occorrenze in ST di lemmi con $x^2 > 3,84$, $p < 0,05$ (N=4.341) = 18,34%.

ID	Lemma	ST (N)	Occ.(N)	χ^2	p-value
5	paziente	21	24	78,22	<0,0001
6	medico	22	26	77,86	<0,0001
7	sintomo	17	17	77,08	<0,0001
8	evento	21	25	73,47	<0,0001
9	ricerca	22	27	73,43	<0,0001
10	animale	19	22	69,35	<0,0001
11	materiale	19	22	69,35	<0,0001
12	area	20	24	69,1	<0,0001
13	cinese	23	31	66,07	<0,0001
14	esempio	40	74	65,23	<0,0001
15	malattia	19	23	64,75	<0,0001
16	epidemia	15	16	61,88	<0,0001
17	coronavirus	15	16	61,88	<0,0001
18	temperatura	16	18	60,97	<0,0001
19	contatto	17	20	60,53	<0,0001
20	controllo	30	51	57,47	<0,0001
21	influenza	16	19	56,14	<0,0001
22	identificare	15	18	51,78	<0,0001
23	rischio	37	74	51,35	<0,0001
24	causa	19	27	49,93	<0,0001
25	respiratorio	11	11	49,82	<0,0001
26	sci	11	11	49,82	<0,0001
27	costante	11	11	49,82	<0,0001
28	febbre	12	13	48,36	<0,0001
29	terremoto	15	19	47,56	<0,0001
30	utilizzare	27	49	45,56	<0,0001
31	antibiotico	10	10	45,29	<0,0001
32	scorso	19	29	44,15	<0,0001
33	malato	11	12	43,86	<0,0001
34	produrre	28	53	43,44	<0,0001
35	incendio	12	14	43,24	<0,0001
36	tecnologia	13	16	43,12	<0,0001
37	grosso	20	32	42,73	<0,0001
38	farmaco	9	9	40,75	<0,0001
39	computer	9	9	40,75	<0,0001

La classe 2 non ha nessuna corrispondenza nella CDG del *sub-corpus* PF_Donna, ma ha caratteristiche simili alla classe 2 del *corpus* P_Fonte. Si tratta dell'area del "discorso scientifico", prevalentemente focalizzata sul coronavirus, e più in generale sui problemi di salute stagionali, come l'influenza, ma si estende fino a coprire l'informazione su eventi che interessano la salute e la sicurezza dei cittadini, come incendi e terremoti, o scoperte scientifiche, in particolare nell'ambito dell'intelligenza artificiale, *tecnologia* e *computer* sono infatti fra le parole significativamente associate a questa classe.

Fonti riconoscibili del mondo lessicale circoscritto da questa classe sono i soggetti di un discorso, impersonale, di sapere ed *expertise*, detto altrimenti persone che parlano di un argomento o di un evento in qualità di professioniste o esperte di un settore. Come anticipato, questa classe è infatti correlata significativamente con segmenti di testo marcati *N_Intervista, ovvero frammenti di interviste giornalistiche.

Elenco 7. Selezione di segmenti di testo tipici della classe 2 del *sub-corpus* PF_Uomo

STT 2 **** *A_2020 *G_25-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 469,43

allerta significa **utilizzare** le **normali procedure** igieniche **personali** e e comunque adottare dei **sistemi** che possano **contenere** l' **epidemia** come si sta facendo in **Cina** e adottare **sistemi** che la **controllano** questa **epidemia** per **esempio** negli **aeroporti** il **virus viaggia** con

STT 22 **** *A_2018 *G_04-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 325,44

ha utilizzato una mano pienamente sensorizzata quindi in grado anche di dare informazioni di ritorno al cervello del paziente ed è stato utilizzato in condizioni ecologiche cioè non ambiente normale al di fuori del laboratorio di ricerca

STT 30 **** *A_2019 *G_09-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 305,38

per il mondo scientifico è stato sicuramente ha descritto il primo caso di AIDS in Italia per esempio io lavoravo nella stessa clinica medica dove c'è il professor Aiuti l'ho seguito per anni mentre visitava i pazienti i primi pazienti quando era quel tempo in cui morivano tutt

STT 46 **** *A_2019 *G_03-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 279,42

utilizzare dei **farmaci** contro la **febbre** e gli **antibiotici** mai se non dietro prescrizione **medica** quando **compaiono** delle gravi complicanze

STT 47 **** *A_2018 *G_11-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 277,52

ci possono essere **terremoti importanti** e che da troppo tempo però non **avvengono** quindi quelle sono le **zone** in cui noi ci aspettiamo che ci saranno i prossimi **eventi importanti**

STT 65 **** *A_2018 *G_04-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 254,01

la persona che indossa questa **mano elettronica** fuori dal **laboratorio** di **ricerca** indossa uno zainetto all' **interno** del quale sono stati miniaturizzati un **computer** tutto un **sistema** di stimolatori che **connette direttamente** il suo **cervello** alla **mano** con un andamento in 2 **direzioni** per il movimento in uscita e per le sensibilità in **ingresso**

STT 148 **** *A_2020 *G_07-01 *T_TG5 *N_Messaggio *P_Fonte *S_Uomo
score: 181,27

antincendio e i **piccoli** aeroplani antincendio che vengono **usati** in questo **genere** di **situazioni** **bisogna** organizzare la squadra devono lavorare tutti insieme i Vigili del Fuoco i **volontari** le Forze Armate se **necessario** le comunità **locali** gli enti **locali** deve essere

STT 179 **** *A_2020 *G_07-01 *T_TG5 *N_Messaggio *P_Fonte *S_Uomo

score: 166,94

per quello che **riguarda** il **rischio incendi** boschivi sarebbe un autentica catastrofe ed è quello che sta **accadendo** in Australia dove **fenomeni estremi** che ormai non sono più così **estremi** si stanno **verificando** come si organizza la lotta agli **incendi** boschivi

Classe 3: la “comunicazione politica”

Le forme lessicali significative per la classe 3 sono complessivamente N=687, di seguito si riportano le prime 40 che coprono circa un terzo delle occorrenze delle forme lessicali significative nei segmenti di testo che formano la classe⁹⁷.

Tabella 15. Prime 40 forme lessicali significative della classe 3 del *sub-corpus* PF_Uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	governo	319	333	456,85	<0,0001
1	politico	258	283	323,81	<0,0001
2	Italia	347	438	302,46	<0,0001
3	paese	289	350	280,45	<0,0001
4	italiano	268	327	253,53	<0,0001
5	Europa	144	158	176,91	<0,0001
6	5_Stelle	115	116	172,47	<0,0001
7	Salvini	120	124	170,15	<0,0001
8	cittadino	141	168	139,06	<0,0001
9	tassa	87	89	125,65	<0,0001
10	ministro	83	86	116,2	<0,0001
11	presidente	84	88	114,54	<0,0001
12	elettorale	72	72	109,41	<0,0001
13	Movimento	74	75	109,02	<0,0001
14	forza	115	139	108,37	<0,0001
15	centro_destra	71	71	107,88	<0,0001
16	legge	86	94	105,59	<0,0001
17	lavoro	193	278	105,22	<0,0001
18	elezione	67	67	101,73	<0,0001
19	sicurezza	119	149	101,27	<0,0001
20	economico	75	81	94,73	<0,0001
21	europeo	84	96	91,81	<0,0001
22	diritto	80	90	91,09	<0,0001
23	Parlamento	60	61	87,55	<0,0001
24	maggioranza	58	60	81,18	<0,0001
25	impresa	68	76	78,59	<0,0001
26	programma	61	66	76,51	<0,0001
27	partito	68	77	75,95	<0,0001
28	Lega	50	50	75,71	<0,0001
29	proposta	59	64	73,5	<0,0001
30	fiscale	48	48	72,66	<0,0001
31	Partito_Democratico	48	48	72,66	<0,0001

⁹⁷ Occorrenze in ST dei primi 40 lemmi (N=4.207)/totale occorrenze in ST di lemmi con $\chi^2 > 3,84$, $p < 0,05$ (N=15.554) = 27,05%.

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
32	votare	53	56	70,41	<0,0001
33	PD	48	49	69,23	<0,0001
34	riforma	47	48	67,71	<0,0001
35	sinistra	55	60	67,49	<0,0001
36	investimento	54	59	65,99	<0,0001
37	reddito	42	42	63,52	<0,0001
38	lavoratore	56	63	63,49	<0,0001
39	parlamentare	41	41	61,99	<0,0001

Le parole elencate nella Tabella 15 fanno riferimento in modo chiaro al mondo della politica e profilano questa classe come la classe della “comunicazione politica”, significativamente associata alla modalità *N_Messaggio (che classifica dichiarazioni rilasciate in contesti extra-TG e ritrasmesse dai TG), articolata in modo del tutto simile alla classe 2 del *sub-corpus* S-Donna e alla classe 3 del *sub-corpus* P_Fonte. Un primo gruppo di lemmi circoscrive l’area semantica delle *policy issues*. Parole come *governo, ministro* (Tabella 15), *decreto, decidere, governare, misura, provvedimento, manovra, problema, risolvere, agire, sindaco, scelta, istituzione, scegliere, agenda, adottare, prevedere, decisione, svolgere, approvazione, affrontare, amministrazione, commissario, esecutivo, istituzionale* (Appendice, Tab. 7) fanno infatti riferimento all’azione di governo e delle amministrazioni o istituzioni, in generale. E parole più specifiche contestualizzano i settori di intervento delle *policy* nazionali o locali: dall’economia (*tassa, economico, impresa, fiscale, investimento, reddito*), al lavoro (*lavoro, lavoratore*), dalla sicurezza (*sicurezza*), all’immigrazione (*immigrazione, confine, migrare, accogliere, clandestino, porto, immigrato, sbarcare, scafista, sbarco, Sea Watch, rimpatrio, ricollocamento, profugo*; Appendice, Tab. 7), dalla giustizia (*diritto, Tabella 15; processo, giustizia, processare, prescrizione, magistratura, giudizio*; Appendice, Tab. 7), alle politiche infrastrutturali (*concessione, infrastruttura, autostradale, TAV*; Appendice, Tab. 7), fino alla politica estera ed europea (*Europa, europeo, Tabella 15; Unione europea, euro, Francia, Libia, internazionale, Mediterraneo, Bruxelles, Stati Uniti, Trump, Spagna, occidentale, europeista, Israele, Venezuela, Macron, Malta, Africa, Oriente, Medio, Nazioni Unite, Maduro, America, Olanda*; Appendice, Tab. 7).

Un secondo gruppo di parole descrive l'area semantica delle *political issues*, con i nomi dei protagonisti (*5_Stelle, Salvini, Movimento, forza, centro-destra, partito, Lega, Partito Democratico, PD, sinistra*, Tabella 15; *Forza Italia, Di Maio, Berlusconi, leader, centro-sinistra, Zingaretti, Renzi, Fratelli d'Italia, Più Europa, Italia Viva*, Appendice, Tab. 7), dei luoghi (*Parlamento*, Tabella 15; *Senato*, Appendice, Tab. 7), e delle dinamiche e relazioni fra partiti (*maggioranza, proposta, parlamentare*, Tabella 15; *accordo, discussione, discutere, polemica, coesione, patto, dibattito*, Appendice, Tab. 7).

Un terzo gruppo delinea l'area delle *campaign issues*, con *elettorale, elezione, programma* (Tabella 22), *voto, vincere, campagna, Romagna, Calabria, coalizione, alleanza, vittoria, lista, candidato, propaganda, eleggere, avversario, rappresentare, battere, elettore, candidatura, candidare* (Appendice, Tab. 7).

Infine, i nomi di leader politici attestano la presenza, verificata nella lettura di segmenti di testo tipici, anche di *personal issues*, che riguardano *Salvini* (Tabella 15), *Di Maio, Berlusconi, Zingaretti, Renzi, Grillo* (Appendice, Tab. 7).

Elenco 8. Selezione di segmenti di testo tipici della classe 3 del *sub-corpus* PF_Uomo

STT 1 **** *A_2020 *G_28-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo
score: 1.794,34

dalle **elezioni politiche** si è **votato** in 9 **regioni** e in 8 ha **vinto** il **centro_destra** e **Movimento 5_Stelle** è praticamente scomparso questo **Parlamento** non è più rappresentativo della **volontà** del nostro **popolo** tornare a **votare** per avere un **governo** forte in grado di risanare la nostra **Italia**

STT 3 **** *A_2019 *G_11-01 *T_TG1 *N_Commento *P_Fonte *S_Uomo
score: 1.654,31

la **Lega decida** che cosa vuole fare se vuole **continuare** ad assecondare le **scelte** del **Movimento 5_Stelle** o vuole essere una **forza** di **governo** che **crede** nella **crescita** nello **sviluppo** nella **realizzazione** di grandi **infrastrutture** indispensabile a un **paese** che non può rimanere indietro che deve essere **protagonista** in **Europa**

STT 29 **** *A_2018 *G_19-01 *T_TG5 *N_Intervista *P_Fonte *S_Uomo
score: 1.394,49

allora **oggi** abbiamo **presentato** il **simbolo** al **ministero** dell' **interno** di quella che sarà la **prima forza politica** del **paese** da qui adesso iniziano gli ultimi giorni le ultime settimane di **campagna elettorale** e il 4 **marzo** vogliamo andare al **governo** del **paese** andremo al **governo** del **paese**

STT 44 **** *A_2019 *G_26-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 1.37,79

Salvini sta facendo una **politica** cinica **basata** sulla **bugia** che trattando male gli stranieri gli **italiani** stiano meglio è una **direzione** di marcia pericolosa **chiediamo** al **governo** di **cambiare** rotta e gli **italiani** di non voltarsi dall'altra parte

STT 53 **** *A_2020 *G_08-01 *T_TG5 *N_Messaggio *P_Fonte *S_Uomo
score: 1.291,09

ho **difeso** i **confini** della **sicurezza** l'**onore** nel mio **paese** lo rifarò quando gli **italiani** mi rimanderanno al **governo** se per questo vogliono **processarmi** indagarmi e incarcerarmi facciano pure insieme a me processeranno un **intero popolo** che vuole vivere tranquillo in casa sua

STT 126 **** *A_2019 *G_21-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 1.093,41

evidentemente hanno **bisogno** di nascondere i loro **fallimenti** in **campo economico** e non **solo** in **Europa** noi avremmo **bisogno** di **alleati** e di amici non di nemici in tutto ciò il **ministro** degli **esteri** è scomparso il **PD** chiede con **forza** che venga in **Parlamento** per evitare che il nostro **paese** si rende ridicolo

STT 393 **** *A_2020 *G_05-01 *T_TG5 *N_Messaggio *P_Fonte *S_Uomo
score: 749,46

poi **Trump** l'**unico governo** che non avvisa stranamente è quello **italiano** perché io mi vedo lì che la **segretaria** che gli dà l'agenda guarda chiama Parigi chiama Berlino chiama Roma no Roma no lascia stare Roma però noi sorridiamo

STT 464 **** *A_2020 *G_29-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo
score: 687,60

per quanto **riguarda** il **reddito cittadinanza** noi abbiamo fatto bene a **mantenere** questa **misura** che si sta rivelando **efficace** per **contrastare** la **povertà** che si è **ridotta** allo stesso tempo lo dobbiamo migliorare per migliorare la sua **capacità** di **sostenere** le **politiche** attive del **lavoro**

5.3.2 Comparazione per anno e per TG

La prima domanda di ricerca sul *corpus* TG ("quali sono le rappresentazioni di genere di donne e uomini, intese/i come soggetto, ovvero fonti dell'informazione nel *corpus* TG?"), contiene due quesiti specifici su somiglianze o differenze per anno e TG, che sono state esplorate approfondendo i risultati della CGD del *sub-corpus* P_Fonte e dei due *sub-corpora* PF_Donna e PF_Uomo.

IRaMuTeQ permette di approfondire questi risultati attraverso una procedura che calcola i contributi del chi-quadrato al valore di associazione tra le classi e le diverse modalità di una variabile⁹⁸, anche con segno negativo, consentendo di valutare quanto i segmenti marcati con una modalità della variabile contribuiscano in senso positivo o negativo alla formazione di una classe.

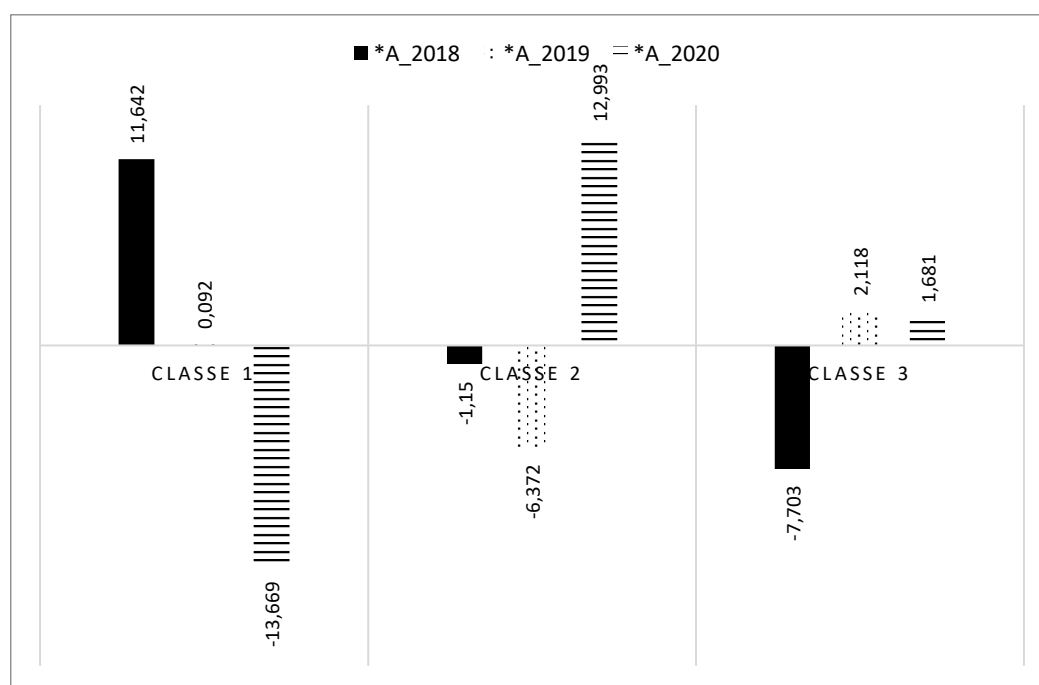
Un chi-quadrato di associazione fra una modalità (es. *A_2018) e una classe lessicale (es. classe 1) di segno positivo dimostra che i segmenti di testo classificati

⁹⁸ Si tratta della stessa procedura che calcola il valore di associazione di un lemma con una classe lessicale.

in quella classe e marcati con quella modalità della variabile contribuiscono alla composizione della classe, in modo più o meno significativo, a seconda che superino o meno il valore soglia $\chi^2 > 3,84$, corrispondente a un valore $p < 0,05$. Un valore χ^2 negativo indica invece che i segmenti di testo classificati in una determinata classe (es. classe 1) e marcati con un determinato valore di una variabile (es. *A_2019) non contribuiscono alla composizione di quella classe, ma contribuiscono, per converso, alla formazione delle altre classi lessicali, in modo più o meno significativo.

Il Grafico 11 illustra i risultati di questa procedura per il *sub-corpus* P_Fonte e la variabile *Anno, evidenziando correlazioni positive e significative fra la modalità *A_2018 e la classe 1 (“testimonianze di vita”), associazioni positive ma non significative fra la modalità *A_2019 e la classe 1, fra la modalità *A_2019 e la classe 3 (“comunicazione politica”) e fra la modalità *A_2020 e la classe 3. Lo stesso grafico mostra correlazioni negative e significative fra la modalità *A_2020 e la classe 1 ($\chi^2 = -13,669$), fra la modalità *A_2019 e la classe 2 ($\chi^2 = -6,372$) e fra la modalità *A_2018 e la classe 3 ($\chi^2 = -7,703$). Infine, si osserva un’associazione negativa ma non significativa fra la modalità *A_2018 e la classe 2 ($\chi^2 = -1,15$).

Grafico 11. Chi-quadrato di associazione delle 3 classi lessicali del *sub-corpus* P_Fonte con le modalità della variabile *Anno



Sintetizzando i principali risultati emersi, i segmenti di testo marcati *A_2018 incidono sulla composizione della classe 1 (“testimonianze di vita”) differenziandola dalle altre classi, in particolare dalla classe 3 (“comunicazione politica”), mentre i segmenti di testo marcati *A_2020 incidono sulla composizione della classe 2 (“discorso scientifico”), differenziandola significativamente dalla classe 1, infine i segmenti di testo *A_2019 incidono negativamente sulla composizione della classe 3, che si caratterizza per essere l’unica non correlata positivamente a qualche modalità della variabile anno.

Il Grafico 12 e il Grafico 13 mostrano i valori del chi-quadrato di associazione delle modalità della variabile *Anno con le tre classi lessicali, rispettivamente, della CGD del *sub-corpus* PF_Donna e della CGD del *sub-corpus* PF_Uomo.

Grafico 12. Chi-quadrato di associazione delle 2 classi lessicali del *sub-corpus* PF_Donna con le modalità della variabile *Anno

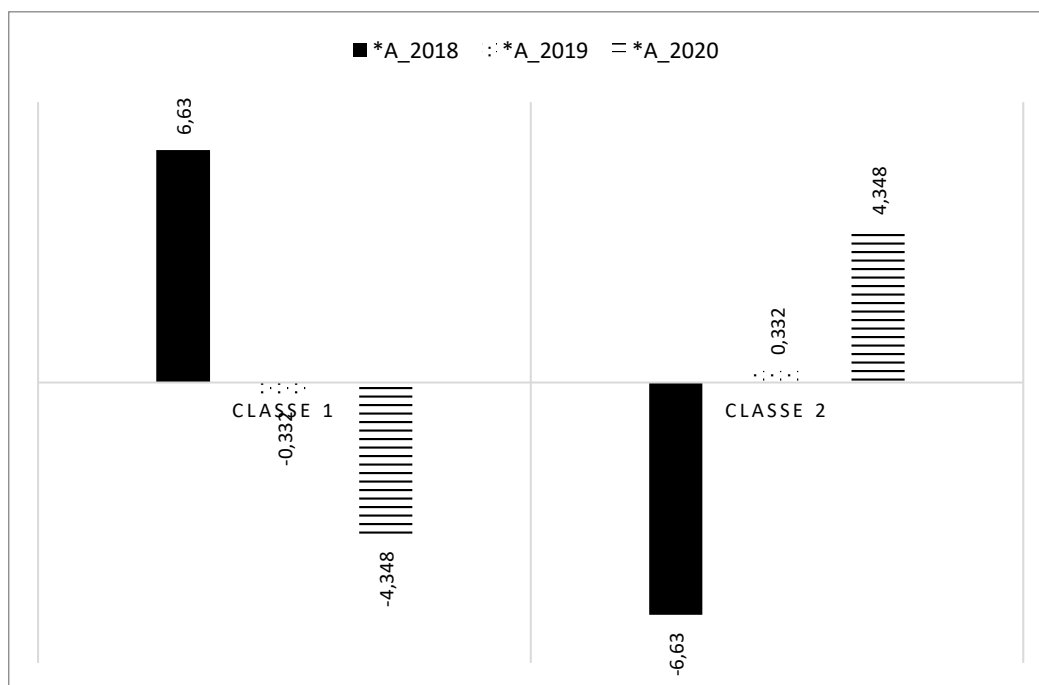
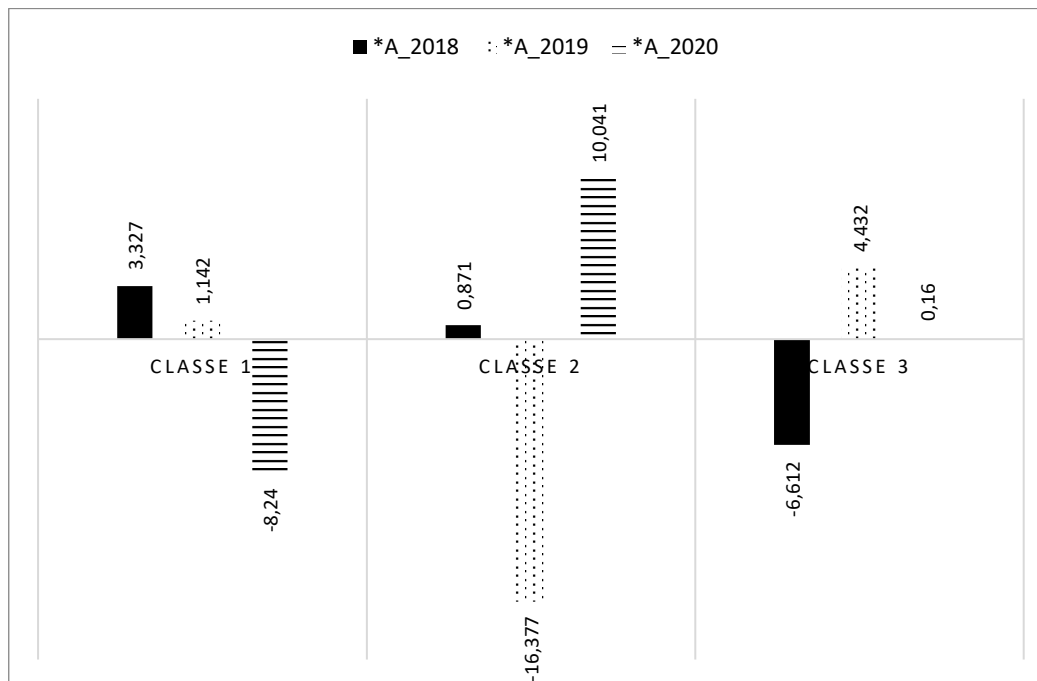


Grafico 13. Chi-quadrato di associazione delle 2 classi lessicali del *sub-corpus* PF_Uomo con le modalità della variabile *Anno



Il primo risultato evidente è una correlazione significativamente positiva fra la classe 1 del *sub-corpus* PF_Donna (“testimonianze di vita”) e la modalità della variabile *Anno_2018, in linea con il risultato riportato nel Grafico 12 per il *sub-corpus* P_Fonte. Mentre il *sub-corpus* PF_Uomo attesta un contributo dei ST marcati *A_2018 positivo ma non significativo alla formazione della classe corrispondente (classe 1).

Il secondo risultato evidente è una correlazione significativamente negativa fra la modalità della variabile *A_2018 e le due classi della “comunicazione politica” nei due *sub-corpora*, che è la classe 2 nel *sub-corpus* PF_Donna e la classe 3 del *sub-corpus* PF_Uomo. Un risultato in linea con quello riportato nel Grafico 13 per il *sub-corpus* P_Fonte, da cui i *sub-corpora* PF_Donna e PF_Uomo sono stati ricavati.

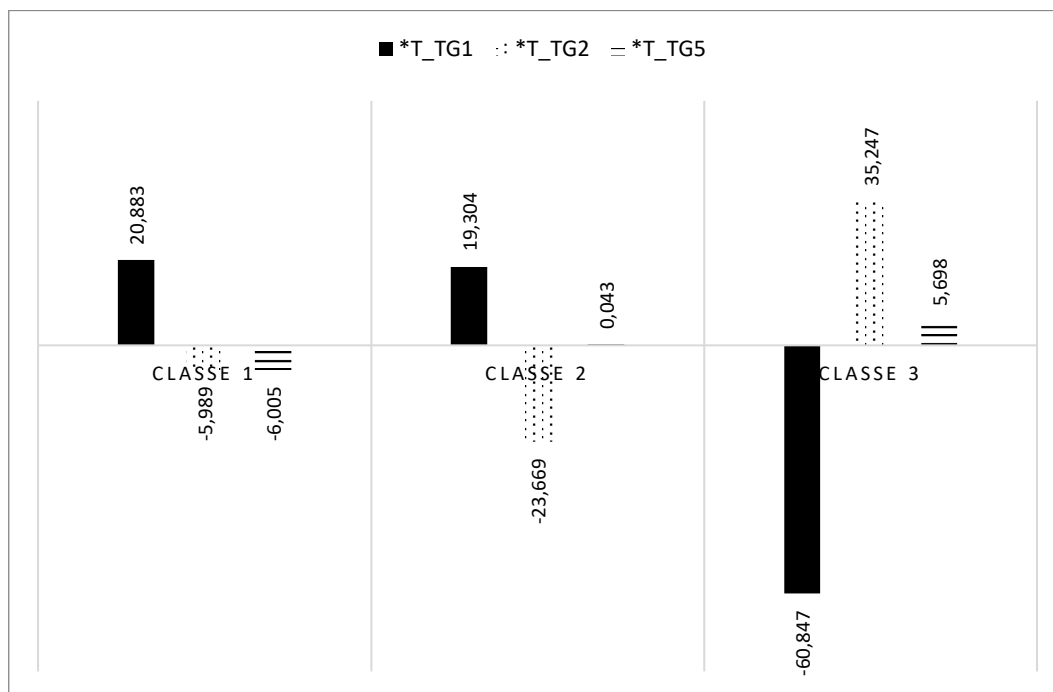
Il terzo risultato riguarda la correlazione significativamente positiva fra la classe 2 (“comunicazione politica”) del *sub-corpora* PF_Donna e la modalità della variabile *A_2020, associazione che non ha corrispondenza negli altri due *corpora*. Nel *sub-corpus* PF_Uomo, sono i segmenti di testo marcati *A_2019 a contribuire

significativamente alla formazione di questa classe, mentre per il *sub-corpus* P_Fonte non si è rilevata alcuna correlazione significativa.

Infine, il chi-quadrato delle associazioni fra le tre modalità della variabile *Anno e la classe del “discorso scientifico”, che profila le classi 2 dei *sub-corpora* P_Fonte e PF_Uomo, attesta ragionevolmente valori di significatività simili per i primi due *corpora*, essendo una classe emergente dalla CGD del *sub-corpus* P_Fonte e del *sub-corpus* PF_Uomo, ma non dalla CGD del *sub-corpus* PF_Donna. Questa classe è significativamente e positivamente associata alla modalità *A_2020 e significativamente e negativamente associata alla modalità *A_2018, sia nel caso del *sub-corpus* P_Fonte, come abbiamo visto sopra, sia nel caso del *sub-corpus* PF_Uomo.

Il Grafico 14 riporta i valori, positivi e negativi, del chi-quadrato delle modalità della variabile *TG, evidenziando, primo, correlazioni significative tra la modalità *T_TG1 e tutte e tre le classi: con valori positivi per le classi 1 ($\chi^2=20,883$) e 2 ($\chi^2=19,304$) e valori negativi per la classe 3 ($\chi^2=-60,847$); secondo, correlazioni significative fra la modalità *T_TG2 e tutte le classi: in senso positivo per la classe 3 ($\chi^2=35,247$), negativo per le classi 1 ($\chi^2=-5,989$) e 2 ($\chi^2=-23,669$); terzo, correlazioni significative tra la modalità *T_TG5 e le classi 1 e 3, in senso negativo, nel primo caso ($\chi^2=-6,005$), positivo, nel secondo ($\chi^2=5,698$).

Grafico 14. Chi-quadrato di associazione delle 3 classi lessicali del *sub-corpus* P_Fonte con le modalità della variabile *TG



Il Grafico 15 e il Grafico 16 mostrano alcune differenze fra il *sub-corpus* PF_Donna e il *sub-corpus* PF_Uomo, e anche in questo caso vanno letti tenendo in considerazione le corrispondenze, sotto il profilo lessicale, fra le classi nei due diversi *sub-corpora* e anche nel *sub-corpus* P_Fonte.

La classe 1 del *sub-corpus* PF_Donna (“testimonianze di vita”) presenta una correlazione significativa in senso negativo con la modalità *T_TG2, mentre non presenta associazioni significative con le modalità *T_TG1 e *T_TG5. La classe corrispondente alle “testimonianze di vita” nel *sub-corpus* PF_Uomo, classe 1, presenta, invece, una correlazione significativa e positiva con la modalità *T_TG1 e negativa, ma non in modo significativo con le modalità *T_TG2 e *T_TG5. In sintesi, questo vuol dire che i segmenti di testo marcati *T_TG2 non hanno contribuito alla formazione di questa classe lessicale, viceversa hanno contribuito alla formazione di un lessico differente che è quello della “comunicazione politica” caratteristico della classe 2. La classe 2 del *sub-corpus* PF_Donna (“comunicazione politica”) attesta, infatti, una correlazione significativamente positiva con la modalità *T_TG2, e una correlazione negativa, ma in modo non significativo, con le modalità *T_TG1 e *T_TG5. La classe 3 del *sub-corpus* PF_Uomo, che dal punto di vista

semantico è la corrispondente della classe 2 del *sub-corpus* PF_Donna, essendo la classe della “comunicazione politica”, presenta invece un profilo del tutto in linea con quello del *sub-corpus* PF_Fonte, attestando correlazioni significative con tutte e tre le modalità della variabile *TG: in senso negativo con la modalità *T_TG1, positivo con le modalità *T_TG2 e *T_TG5.

Infine, la classe 2 del *sub-corpus* PF_Uomo, dove si concentra il “discorso scientifico”, assente nel corrispettivo *sub-corpus* di fonti femminili, evidenzia una correlazione positiva e significativa con la modalità *T_TG1, significativamente negativa con la modalità *T_TG2, in linea con i risultati sul *sub-corpus* P_Fonte, dal quale si differenzia, però, per una correlazione significativamente negativa con la modalità *T_TG5. Considerando che lo scarto fra il *sub-corpus* P_Fonte e il *sub-corpus* PF_Uomo sono i segmenti di testo con una fonte femminile, che sono raggruppati nel *sub-corpus* PF_Donna, senza presentare caratteristiche di omogeneità di vocabolario tali da farne una classe lessicale autonoma, è possibile ipotizzare che questa differenza sia attribuibile a uno scarto di testi, prevalentemente marcati *T_TG5.

Grafico 15. Chi-quadrato di associazione delle 2 classi lessicali del *sub-corpus* PF_Donna con le modalità della variabile *TG

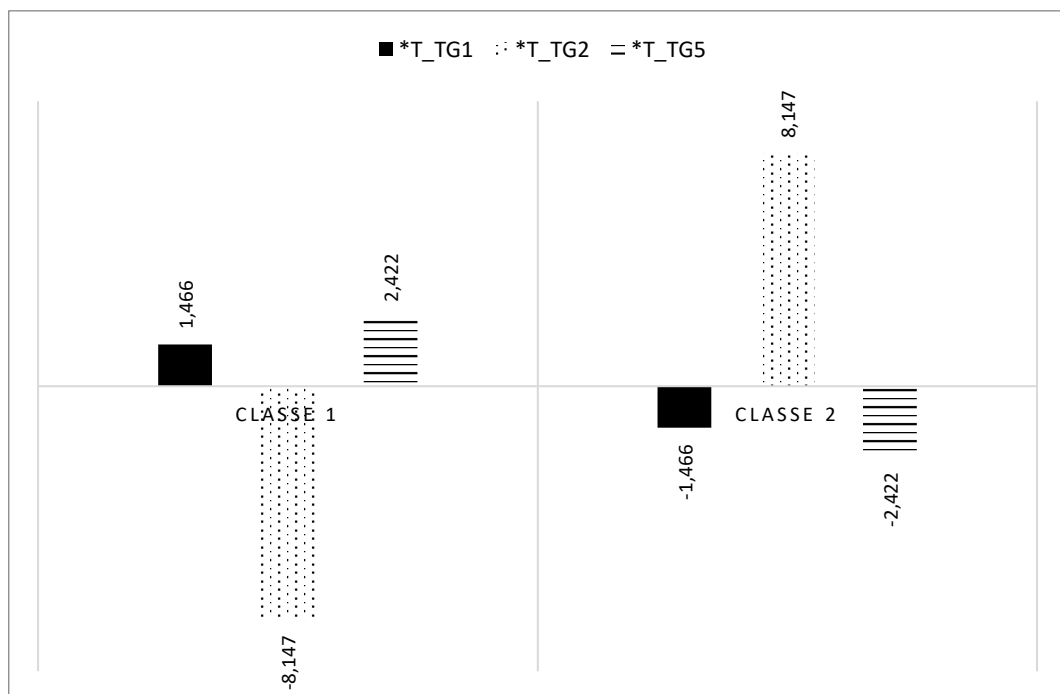
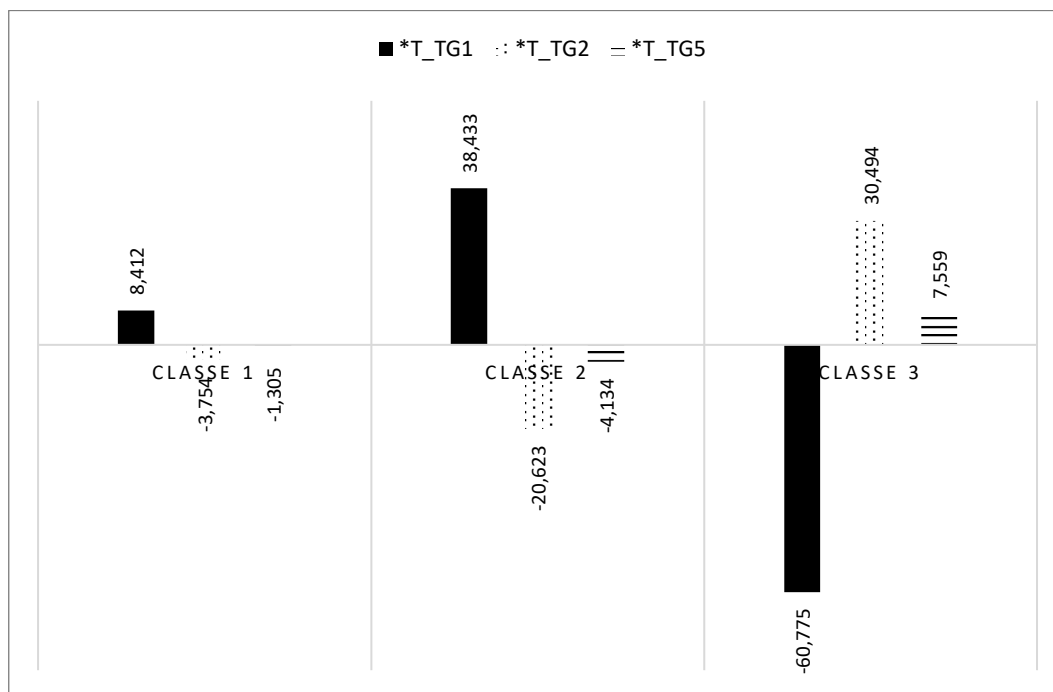


Grafico 16. Chi-quadrato di associazione delle 2 classi lessicali del *sub-corpus* PF_Uomo con le modalità della variabile *TG



5.3.3 La rappresentazione di donne e uomini come argomento dell'informazione

La seconda domanda di ricerca formulata all'inizio di questo capitolo è: "quali sono le rappresentazioni di genere di donne e uomini, intese/i come oggetto, ovvero argomento dell'informazione nel *corpus* TG?". Per rispondere a questa domanda i lemmi *donna* e *uomo* sono stati assunti come *topic* (parole-argomento) del *corpus* TG. Attraverso IRaMuTeQ sono stati costruiti due *sub-corpora* formati da tutti i segmenti di testo contenenti nel primo caso il lemma *donna* (*sub-corpus* donna) e nel secondo caso il lemma *uomo* (*sub-corpus* uomo). I due *sub-corpora* sono poi stati sottoposti, prima, a un'analisi lessicometrica, per profilarne il vocabolario, e, poi, a una classificazione gerarchica discendente (CGD), con l'obiettivo di individuare le rispettive classi lessicali e da queste fare emergere le rappresentazioni di genere latenti, anche attraverso l'analisi fattoriale.

Le dimensioni e caratteristiche dei due *sub-corpora* sono riportate nella tabella seguente che evidenzia profili abbastanza simili. Sebbene nel *sub-corpus* uomo si osservi un più elevato numero di testi, occorrenze, forme e hapax, l'ordine

di grandezza dei due *corpora* è uguale: entrambi i *corpora* sono segmentati in circa un migliaio di testi, attestano circa 40.000 occorrenze e circa 6.000 forme.

Tabella 16. Statistiche: dimensioni e caratteristiche dei *sub-corpora* donna e uomo

	<i>Sub-corpus</i> donna	<i>Sub-corpus</i> uomo
Testi (N)	926	1.002
Occorrenze (N)	41.163	42.255
Forme (Lemmi) (N)	5.948	6.313
Hapax (N)	3.111	3.335
Hapax/occorrenze (%)	7,56	7,89
Hapax/lemmi (%)	52,30	52,83
Media occorrenze*testo	44,45	42,17
Ratio	0,14	0,15

Anche il confronto fra le forme lessicali più frequenti, elencate nella tabella seguente, mostra una similarità fra i due *sub-corpora* che condividono più di un terzo delle parole con rango uguale o inferiore a 30⁹⁹.

Tabella 17. Forme lessicali più frequenti nei *sub-corpora* donna e uomo (occorrenze)

<i>Sub-corpus</i> donna				<i>Sub-corpus</i> uomo			
Rango	Lemma	Occ. (N)	Occ. (%)	Rango	Lemma	Occ. (N)	Occ. (%)
1	donna	1.160	19,50	1	uomo	1.178	18,66
2	anno	225	3,78	2	donna	208	3,29
3	uomo	204	3,43	3	anno	203	3,22
4	primo	136	2,29	4	primo	134	2,12
5	bambino	125	2,10	5	uccidere	81	1,28
6	persona	83	1,40	6	figlio	75	1,19
7	violenza	73	1,23	7	persona	73	1,16
8	Italia	71	1,19	8	morire	69	1,09
9	vita	65	1,09	9	Italia	62	0,98
10	giorno	60	1,01	10	casa	61	0,97
11	paese	59	0,99	11	giorno	59	0,93
11	figlio	59	0,99	12	grande	58	0,92
12	trovare	58	0,98	13	vita	57	0,90
12	nuovo	58	0,98	14	vittima	56	0,89
12	morire	58	0,98	15	secondo	54	0,86
13	arrivare	57	0,96	15	oggi	54	0,86
14	uccidere	56	0,94	16	mondo	52	0,82
14	oggi	56	0,94	17	polizia	49	0,78
15	vittima	54	0,91	17	nuovo	49	0,78
15	grande	54	0,91	18	volere	46	0,73
16	lavoro	53	0,89	19	parlare	45	0,71
17	mondo	52	0,87	19	arrivare	45	0,71
17	casa	52	0,87	20	paese	44	0,70
18	proprio	51	0,86	21	moglie	43	0,68

⁹⁹ I lemmi di rango ≥ 30 coprono il 20% del lemmario del *sub-corpus* donna e il 19% del lemmario del *sub-corpus* uomo.

Sub-corpus donna				Sub-corpus uomo			
Rango	Lemma	Occ. (N)	Occ. (%)	Rango	Lemma	Occ. (N)	Occ. (%)
19	marito	44	0,74	22	strada	41	0,65
20	presidente	43	0,72	22	italiano	41	0,65
20	migrare	43	0,72	23	vedere	40	0,63
20	dare	43	0,72	23	raccontare	40	0,63
21	strada	42	0,71	24	ultimo	39	0,62
21	secondo	42	0,71	24	proprio	39	0,62
22	giovane	41	0,69	24	presidente	39	0,62
23	italiano	39	0,66	25	trovare	38	0,60
23	diritto	39	0,66	26	città	37	0,59
24	vedere	38	0,64	27	bambino	35	0,55
24	accogliere	38	0,64	27	Roma	35	0,55
25	portare	37	0,62	28	forza	34	0,54
26	raccontare	36	0,61	28	chiedere	34	0,54
26	parlare	36	0,61	28	arrestare	34	0,54
27	ultimo	35	0,59	29	storia	33	0,52
27	prendere	35	0,59	29	portare	33	0,52
27	chiedere	35	0,59	29	compagno	33	0,52
28	venire	34	0,57	29	colpire	33	0,52
28	famiglia	34	0,57	30	violenza	32	0,51
28	Di_Maio	34	0,57	30	famiglia	32	0,51
29	compagno	33	0,55	30	corpo	32	0,51
30	madre	32	0,54	30	carabiniere	32	0,51

La comune ricorrenza di nomi di luogo (*casa, Italia, mondo, paese, strada*) e di tempo (*anno, giorno, oggi*), insieme ai verbi *parlare* e *raccontare*, profila entrambi i *sub-corpora* come testi prevalentemente di cronaca, un formato narrativo tipicamente circoscritto ad avvenimenti di vita quotidiana e a precisi contesti territoriali (es. un paese, una città) o argomentativi (es. la politica, lo sport). Le parole *uccidere, violenza* e *vittima*, con *vita* e *morire*, delimitano i contenuti di questa narrazione cronachistica entro i confini di eventi criminosi e/o violenti. Questi confini sono meglio specificati nel *sub-corpus* uomo dove ricorrono con elevate frequenza i nomi *polizia* e *carabiniere* e il verbo *arrestare*. Mentre nel *sub-corpus* donna ricorrono le parole *migrare* e *accogliere* che suggeriscono la presenza di una narrazione anche sui fenomeni migratori. La comune occorrenza di nomi basati su relazioni affettive e familiari quali *compagno, famiglia, figlio* e *marito* e *moglie* che, dal punto di vista semantico-referenziale, sono simmetrici, rispetto ai generi, e sono presenti nei due *sub-corpora* in posizione speculare¹⁰⁰,

¹⁰⁰ *Marito* è nell'elenco delle 30 parole più frequenti nel *sub-corpus* donna (e al rango 46 nel *sub-corpus* uomo), *moglie* è presente nell'elenco delle 30 parole più frequenti nel *sub-corpus* uomo e assente dal *sub-corpus* donna.

contribuiscono a circoscrivere la narrazione cronachistica prevalente e comune ai due *corpora* entro la sfera privata.

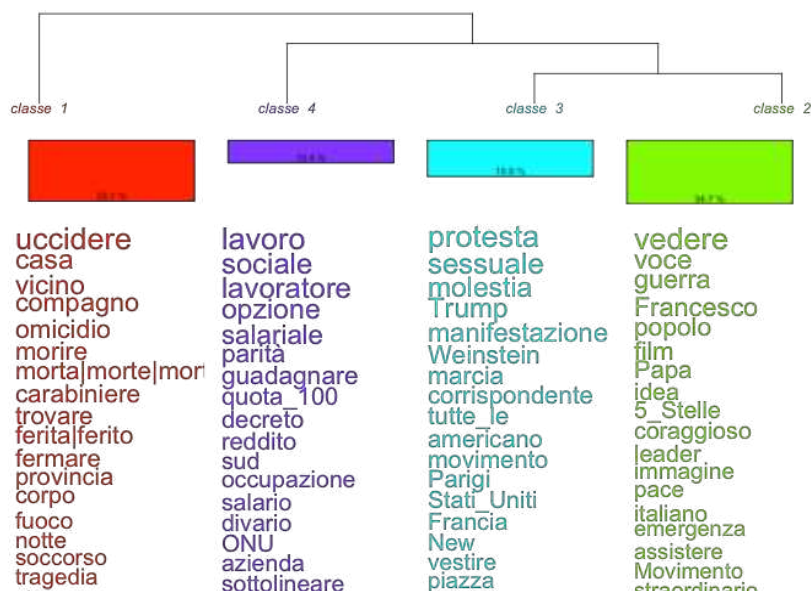
Per approfondire queste prime evidenze, i due *sub-corpora*, come anticipato, sono stati sottoposti a classificazione gerarchica discendente (CGD). L'applicazione della CGD al *sub-corpus* donna ha restituito come risultato quattro classi lessicali che coprono il 76,24% delle UCE in cui il *corpus* è stato segmentato (N=828 su un totale di N=1.086), come dettagliato nella Tabella 18.

Tabella 18. Statistiche: risultati della CGD del *sub-corpus* donna

Testi (N)	926
Segmenti di testo (N)	1.086
Forme (N)	8.111
Occorrenze (N)	41.163
Lemmi (N)	5.948
Forme (lemmi) attive (N)	5.446
Forme (lemmi) supplementari (N)	448
Forme (lemmi) attive con frequenza ≥ 3	1.563
Media di forme per segmento	37,90
Classi (N)	4
Segmenti classificati (76,24%)	828

Il dendrogramma illustrato nel Grafico 17 mostra la relazione gerarchica fra le quattro classi, evidenziando una struttura con la classe 1 al vertice. Leggendo il grafico da sinistra a destra, una prima segmentazione individua in alto a sinistra la classe 1, una seconda segmentazione individua il gruppo delle classi 4, 3 e 2. Una terza segmentazione determina la formazione delle classi 3 e 4, in relazione abbastanza simmetrica fra loro.

Grafico 17. Dendrogramma delle classi lessicali del *sub-corpus* donna



Il Grafico 17 mostra le prime parole più significative delle quattro classi lessicali, delineando sommariamente quattro aree semantiche di cui una, corrispondente alla classe 2, di difficile interpretazione, e tre più chiare che fanno riferimento in modo esplicito alla violenza (classe 1), a proteste e molestie sessuali (classe 3) e al mondo del lavoro (classe 4). Per definire meglio queste aree di seguito si riportano, classe per classe, gli elenchi delle forme lessicali con frequenza superiore a N=5 nei segmenti di testo raggruppati in questa classe (ST N) e indice di correlazione significativa ($\chi^2 > 3,84$, $p < 0,05$)¹⁰¹, più una selezione di segmenti di testo tipici per ogni classe, utili per comprendere l'agenda dell'informazione in cui donne e uomini sono citati.

Classe 1: la "violenza domestica contro le donne"

La classe 1 copre il 33,09% dei segmenti di testo (N=274) ed è significativamente associata alla modalità della variabile Notizia *N_Lancio ($\chi^2(1, N=98)=20,02$, $p < 0,001$) e *P_Giornalista ($\chi^2(1, N=727)=12,11$, $p < 0,05$).

L'associazione significativa di questa classe con le modalità *N_Lancio e *P_Giornalista forniscono informazioni importanti per un'interpretazione di questo

¹⁰¹ Gli elenchi completi sono riportati in Appendice.

“mondo lessicale”, in una prospettiva mediale, che tenga in considerazione la struttura tipica di un telegiornale. La correlazione con la modalità *P_Giornalista indica infatti che questa classe è distintiva del parlato giornalistico (vs. il parlato delle fonti, marcato con la modalità P_Fonte), in particolare del parlato giornalistico dei cosiddetti “lanci” (marcati con la modalità *N_Lancio), ovvero le notizie anticipate da studio a cui seguono le notizie vere e proprie presentate in un servizio giornalistico.

Queste informazioni, insieme all’elenco delle parole distintive e ai segmenti di testo tipici, consentono di identificare questa classe come l’area semantica entro cui si iscrivono le anticipazioni di notizie di cronaca nera, prevalentemente su casi di violenza domestica contro le donne. L’identificazione di questo *topic* specifico deriva da un’analisi dettagliata dei diversi gruppi di parole (sub-aree lessicali) costituenti la classe.

Le parole *uccidere, omicidio, ferita|ferito, picchiata, aggredire, colpire, colpo* si riferiscono in modo molto chiaro alla violenza fisica, un comportamento che mette a rischio la *vita*, fino a limite estremo della *morte*, condizioni entrambe richiamate dai verbi *vivere* e *morire*, dai nomi *vita* e *morta|morte|morto* (unificati in un unico lemma che include *morte*, nel senso della condizione di fine vita, e *morta/o* nel senso di persona che ha cessato di vivere, donna o uomo) e dalle parole *cadavere*, contenuta in STT che si riferiscono a casi di ritrovamento, occultamento o distruzione di cadavere (es. STT 10, Elenco 9), e *corpo* – in espressioni quali *il corpo della vittima* (es. STT 9, Elenco 9). Anche *vittima*, con riferimento a persona che ha subito qualche tipo di violenza, è significativamente associata a questa classe lessicale. Al tema della violenza, in particolare della violenza contro le donne, fa infine riferimento il termine *femminicidio* che definisce atti di violenza basati sulla

concezione condivisa della ‘femmina’ come un nulla sociale [...] non [si tratta del]l’omicidio di una persona di sesso femminile, a cui possono essere riconosciute aggravanti individuali, ma [di] un delitto che trova i suoi profondi motivi in una cultura dura a rinnovarsi e in istituzioni che

ancora la rispecchiano almeno in parte. [...] Qualcosa di più e di diverso (dall'omicidio). Qualcosa che si colloca all'interno di una visione culturale che vede il femminile disprezzato e disprezzabile (Accademia della Crusca 2013) ¹⁰².

Un secondo gruppo di nomi e verbi fa riferimento alle forze dell'ordine, della sicurezza, di soccorsi e alle loro azioni (*carabiniere, fermare, soccorso, vigile, polizia, salvare, arrestare*) e alla giustizia (*interrogatorio, inquirente, carcere, indagare, inchiesta, condannare, indagine, accusare*). Un terzo nucleo di parole fa riferimento alla dimensione domestica (*casa, abitazione, appartamento, porta*). Un quarto insieme fa riferimento a relazioni o identità basate su relazioni di tipo affettivo o familiare: *compagno, relazione, convivente, figlio, marito, moglie, coppia*. Due parole, *marocchino* e *nigeriano* indicano invece un'identità etnica.

Da ultimo, ci sono le parole della narrazione spazio-temporale tipica della *cronaca*, parola anch'essa distintiva di questa classe: aggettivi e nomi di luogo (*vicino, provincia, Napoli, zona, centro, ospedale*) e di tempo (*notte, anno, sera, giorno, scorso*). *Tragedia, urlo, grave, paura*, conferiscono, infine, un registro drammatico a questa narrazione cronachistica.

Tabella 19. Forme lessicali con freq. N>5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 1 del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	uccidere	32	36	52,92	<0,0001
1	casa	36	48	40,42	<0,0001
2	vicino	20	21	37,58	<0,0001
3	compagno	26	32	34,87	<0,0001
4	omicidio	16	16	32,99	<0,0001
5	morire	24	30	30,93	<0,0001
6	morta morte morto	18	20	29,98	<0,0001
7	carabiniere	22	27	29,52	<0,0001
8	trovare	29	40	29,48	<0,0001
9	ferita ferito	14	14	28,79	<0,0001
10	fermare	17	19	27,92	<0,0001
11	provincia	15	16	27,11	<0,0001

¹⁰² Seguendo Giomi e Magaraggia (2017), si è optato per questa definizione di *femminicidio*, perché è chiara e sintetica e, fra l'altro, è fornita dalla Crusca con l'intenzione di spiegare i motivi dell'utilità e della specificità di questo termine (cfr. Accademia della Crusca 2013). Per un dibattito sull'uso del termine *femminicidio*, anche nel discorso dei media, si veda per esempio Bandelli e Porcelli (2016), Giomi (2016), Pavan (2016).

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
12	corpo	17	20	24,94	<0,0001
13	fuoco	11	11	22,54	<0,0001
14	notte	15	18	20,98	<0,0001
15	soccorso	10	10	20,47	<0,0001
16	tragedia	10	10	20,47	<0,0001
17	vita	32	52	20,28	<0,0001
18	ritrovare	11	12	18,87	<0,0001
19	relazione	9	9	18,4	<0,0001
20	vigile	9	9	18,4	<0,0001
21	cronaca	9	9	18,4	<0,0001
22	interrogatorio	8	8	16,33	<0,0001
23	inquirente	8	8	16,33	<0,0001
24	abitazione	8	8	16,33	<0,0001
25	carcere	11	13	15,84	<0,0001
26	urlo	9	10	14,81	0,00011
27	marocchino	9	10	14,81	0,00011
28	polizia	14	19	14,47	0,00014
29	vittima	20	31	14,36	0,00015
30	picchiata	7	7	14,27	0,00015
31	rogo	7	7	14,27	0,00015
32	convivente	7	7	14,27	0,00015
33	aggredire	7	7	14,27	0,00015
34	rimanere	11	14	13,3	0,00026
35	riuscire	11	14	13,3	0,00026
36	figlio	26	45	13,1	0,00029
37	Napoli	8	9	12,79	0,00034
38	anziano	8	9	12,79	0,00034
39	appartamento	8	9	12,79	0,00034
40	marito	16	24	12,58	0,00038
41	cadavere	6	6	12,22	0,00047
42	zona	6	6	12,22	0,00047
43	Valentina	6	6	12,22	0,00047
44	precedente	6	6	12,22	0,00047
45	pista	6	6	12,22	0,00047
46	nigeriano	6	6	12,22	0,00047
47	incendio	6	6	12,22	0,00047
48	Milano	9	11	11,95	0,00054
49	anno	77	175	11,93	0,00055
50	raccontare	19	31	11,57	0,00067
51	ultimo	18	29	11,4	0,00073
52	sera	11	15	11,17	0,00082
53	chiamare	15	23	11,03	0,00089
54	condizione	12	17	11,02	0,0009
55	vivere	12	17	11,02	0,0009
56	indagare	7	8	10,8	0,00101
57	inchiesta	7	8	10,8	0,00101
58	volto	8	10	10,06	0,00151
59	centro	9	12	9,66	0,00188
60	sentire	15	24	9,65	0,00188
61	moglie	14	22	9,52	0,00202
62	salvare	10	14	9,45	0,0021
63	mettere	13	20	9,42	0,00214

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
64	grave	6	7	8,83	0,00296
65	condannare	6	7	8,83	0,00296
66	cadere	6	7	8,83	0,00296
67	pezzo	6	7	8,83	0,00296
68	coppia	6	7	8,83	0,00296
69	indagine	7	9	8,21	0,00417
70	tentare	7	9	8,21	0,00417
71	bimbo	8	11	7,91	0,00491
72	colpire	11	17	7,83	0,00512
73	accusare	9	13	7,79	0,00525
74	portare	15	26	7,34	0,00675
75	arrestare	11	18	6,52	0,01064
76	paura	6	8	6,41	0,01136
77	colpo	6	8	6,41	0,01136
78	ospedale	13	23	5,87	0,01544
79	venire	14	26	5,22	0,0223
80	giorno	19	38	5,14	0,02334
81	proprio	22	46	4,78	0,02886
82	scorso	12	22	4,7	0,03019
83	femminicidio	7	11	4,7	0,0302
84	porta	6	9	4,63	0,03137
85	gestire	6	9	4,63	0,03137

Elenco 9. Selezione di segmenti di testo tipici della classe 1 del *sub-corpus* donna

STT 2 **** *A_2018 *G_24-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 233,94

ha **ucciso** la **moglie** e un **bimbo** di 3 **anni** a **colpi** di **mannaia** a Cremona **arrestato** un cinese di 51 **anni** ai **carabinieri** ha detto era una donna **cattiva** il piccolo era il **figlio** dei **vicini** di **casa** al quale la **coppia** avrebbe dovuto badare Paola Colombo

STT 3 **** *A_2020 *G_30-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 230,43

picchiata per 3 **giorni** fino alla **morte** **vittima** una donna **uccisa** nella **propria abitazione** di Mazara del Vallo in Sicilia **fermato** il **marito** già denunciato per **maltrattamenti** attenzione ai dati un **femminicidio** alla settimana dall **inizio** dell **anno** **Valentina** di **Virgilio**

STT 5 **** *A_2018 *G_06-01 *T_TG1 *N_Servizio *P_Giornalista *S_Donna
score: 224,60

hanno **riservato** alcuna pietà nella sua **vita** questa donna **marocchina** di 46 **anni** aveva sempre scelto l'indipendenza anche il **giorno** della sua **morte** era tornata a **casa** dal **compagno** di **origine albanese** dopo una giornata di lavoro è stata **uccisa**

STT 9 **** *A_2018 *G_10-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 202,37

di nuovo qui da noi purtroppo anche **stasera** con un nuovo **femminicidio** in **provincia** di **Novara** aveva **precedenti penali** il **convivente** della donna **trovata** senza **vita** nella sua **abitazione** adesso l'uomo è sotto **interrogatorio** da ore sul **corpo** della **vittima** sono stati **trovati** diversi lividi e segni di violenza
Giovanna Cucè

STT 10 **** *A_2018 *G_06-01 *T_TG1 *N_Lancio *P_Giornalista *S_Donna
score: 196,82

è svolta nelle **indagini** sulla donna **marocchina uccisa** e fatta a **pezzi** nel **veronese** 2 **albanesi** uno è il **convivente fermati** con l'accusa di **omicidio volontario** e distruzione di **cadavere** **Caterina Proietti**

Classe 2: la “centralità delle donne”

La classe 2 copre il 34,66% dei segmenti di testo (N=287) ed è significativamente associata alle modalità *P_Fonte ($\chi^2(1, N=101)=16,12$, $p<0,001$), *S_Uomo, $\chi^2(1, N=350)=6,83$, $p<0,05$), *N_Intervista ($\chi^2(1, N=71)=7,34$, $p<0,05$), *N_Commento $\chi^2(1, N=75)=5,8$, $p<0,05$, e infine *N_Messaggio ($\chi^2(1, N=23)=4,99$, $p<0,05$).

La correlazione significativa con la modalità *P_Fonte è indicativa di una classe che si distingue dalle altre per raggruppare il parlato delle fonti (vs. il parlato giornalistico), espresso in interviste, commenti o messaggi (rilasciati in luogo pubblico e ritrasmessi dai notiziari) che sono le tre tipologie di parlato delle fonti distinte dalle tre modalità di tipologia della notizia marcate con le variabili *N_Intervista, *N_Commento, *N_Messaggio.

Le parole che la caratterizzano appartengono a diversi discorsi, il cui *trait di union* è poco chiaro. Il primo discorso identificabile, partendo da nomi distintivi di questa classe quali *Francesco, Papa, Bergoglio*, è quello ecclesiastico. I STT contenenti queste parole contengono anche parole quali *guerra, pace, giornata, ricordare, celebrare*, che contestualizzano il discorso ecclesiastico nell'ambito delle celebrazioni della giornata mondiale per la pace, una ricorrenza che la Chiesa cattolica celebra ogni primo gennaio. Una seconda ricorrenza a cui si riferiscono queste ultime parole è la giornata della memoria, che a partire dal 2005, commemora ogni 27 gennaio le vittime della Shoah. Il terzo discorso identificabile in questa classe è quello politico focalizzato sull'ingresso e sulla carriera delle donne nella vita politica. Indicatori di questo discorso sono nomi come *5_Stelle, leader, Mattarella, politico*. Infine, un ultimo gruppo di parole (*film, bello, artista*) fa riferimento a un discorso artistico e culturale, in particolare cinematografico.

La lettura dei STT (di cui si riporta una breve selezione nell'Elenco 10), suggerisce che il *trait d'union* di quest'area semantica sia una visione della centralità delle donne e del loro ruolo nella società, nella storia, nella politica, e, simbolicamente, nelle opere artistiche e cinematografiche. Una visione veicolata

dal discorso mediatico attraverso la soggettività delle fonti, in particolare delle fonti di genere maschile, essendo la modalità della variabile *S_Uomo significativamente correlata a questa classe.

Tabella 20. Forme lessicali con freq. $N > 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 2 del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	vedere	20	27	19,14	<0,0001
1	voce	11	13	14,55	0,00013
2	guerra	9	10	13,69	0,00021
3	Francesco	9	10	13,69	0,00021
4	popolo	7	7	13,31	0,00026
5	film	10	12	12,74	0,00035
6	Papa	11	14	12,12	0,00049
7	idea	6	6	11,39	0,00073
8	5_Stelle	6	6	11,39	0,00073
9	coraggioso	6	6	11,39	0,00073
10	leader	6	6	11,39	0,00073
11	immagine	11	15	10,09	0,00149
12	pace	7	8	9,96	0,0016
13	italiano	15	23	9,75	0,00179
14	consiglio	8	10	9,19	0,00243
15	vero	8	10	9,19	0,00243
16	giornata	9	12	8,75	0,00309
17	dare	17	28	8,69	0,0032
18	passo	10	14	8,5	0,00355
19	momento	12	18	8,32	0,00391
20	accompagnare	6	7	8,12	0,00436
21	difficile	6	7	8,12	0,00436
22	usare	6	7	8,12	0,00436
23	Mattarella	6	7	8,12	0,00436
24	regola	6	7	8,12	0,00436
25	oggetto	6	7	8,12	0,00436
26	esprimere	6	7	8,12	0,00436
27	capace	6	7	8,12	0,00436
28	Bergoglio	6	7	8,12	0,00436
29	diretto	7	9	7,47	0,00627
30	bello	7	9	7,47	0,00627
31	unico	8	11	7,13	0,00756
32	crescere	8	11	7,13	0,00756
33	ricordare	9	13	6,97	0,00829
34	forza	11	17	6,92	0,00853
35	primo	47	103	6,25	0,01242
36	vivo	6	8	5,8	0,01599
37	libero	6	8	5,8	0,01599
38	visita	6	8	5,8	0,01599
39	futuro	6	8	5,8	0,01599
40	candidato	6	8	5,8	0,01599
41	ragazza ragazzo	11	18	5,68	0,01712
42	proteggere	7	10	5,58	0,01815
43	ruolo	7	10	5,58	0,01815

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
44	artista	7	10	5,58	0,01815
45	mano	8	12	5,51	0,01893
46	grande	22	43	5,45	0,01953
47	politico	12	21	4,81	0,02832
48	diventare	11	19	4,63	0,03133
49	celebrare	7	11	4,13	0,04206
50	grazia	6	9	4,12	0,04249
51	chiedere	12	22	3,95	0,04699
52	porre	12	22	3,95	0,04699

Elenco 10. Selezione di segmenti di testo tipici della classe 2 del *sub-corpus* donna

STT 7 **** *A_2020 *G_01-01 *T_TG2 *N_Servizio *P_Giornalista *S_Uomo
score: 59,22

la **giornata** della **pace** che la **Chiesa celebra** ogni **primo** gennaio **diventa** quest anno anche quella delle donne

STT 31 **** *A_2020 *G_15-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 45,32

Papa Francesco nomina la **prima** donna sottosegretario alla segreteria di Stato agirà in nome della diplomazia pontificia è **Francesca** di Giovanni esperta in diritto **internazionale** e umanitario Ignazio Ingrao l ha incontrata

STT 45 **** *A_2018 *G_21-01 *T_TG5 *N_Servizio *P_Giornalista *S_Donna
score: 40,23

arrivo nei cinema **italiani** per la **giornata** della **memoria** intreccia i destini di 3 **personaggi** durante la Seconda Guerra Mondiale un aristocratica russa deportata nei lager per aver **protetto** alcuni bambini ebrei un ufficiale delle SS innamorato della donna che cerca di

STT 87 **** *A_2018 *G_25-01 *T_TG1 *N_Intervista *P_Fonte *S_Donna
score: 30,34

dare all **Italia** un governo di patrioti la **prima** donna alla presidenza del **consiglio** dei ministri e su questo stiamo **lavorando**

STT 96 **** *A_2019 *G_17-01 *T_TG5 *N_Commento *P_Fonte *S_Uomo
score: 28,19

conosco il marito di **Theresa May** l ho visto quello visto qualche volta così isolato all interno dei comuni che osservava che osservava il dibattito e ovviamente poi c è anche un'altra **grande** donna che non è presidente della **repubblica**

STT 160 **** *A_2018 *G_06-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 19,43

lei ha rappresentato un passaggio d epoca ed è stata **capace** di chiudere l epoca in cui le donne erano fatali e aprire l epoca in cui le donne **conquistavano** il **mondo**

Classe 3: il "dissenso delle donne"

La classe 3 copre il 19,81% delle UCE (N=164) ed è significativamente associata alle modalità *N_Collegamento ($\chi^2(1, N=44)=9,5, p<0,05$) e *S_Donna, ($\chi^2(1, N=478)=5,53, p<0,05$).

La correlazione significativa con la modalità della notizia *N_Collegamento è indicativa di un parlato, giornalistico o di fonte esterna, in diretta, da e con lo studio televisivo, in coerenza con quanto emerge dalla lettura dei segmenti di testo e delle parole distintive di questa classe, il cui contesto di riferimento è in modo abbastanza trasparente l'estero, in particolare *Stati Uniti* e *Francia*, paesi dai quali evidentemente sono state trasmesse informazioni in diretta. Entrambi i nomi di questi paesi sono presenti nella Tabella 20, insieme a nomi delle loro capitali, *Parigi* e *Washington* (Appendice, Tab. 10), e a nomi di altre località estere: il generico *America*, i più specifici *Iran*, *Davos*, *Danimarca* e *Grecia* (Appendice, Tab. 10). Il riferimento principale di questa classe lessicale ai contesti statunitense e francese è indicata anche dall'occorrenza significativa degli aggettivi *americano* e *francese*, dal nome dei due presidenti di Stati Uniti e Francia, nel periodo 2018–2020, *Trump* e *Macron*, e dei due artisti *Weinstein* e *Deneuve* (*Catherine*, Appendice, Tab. 10), la cui occorrenza significativa può essere interpretata partendo dalla presenza contestuale delle parole *sessuale* e *molestia*, dalla lettura dei ST tipici e dalla conoscenza di alcune informazioni extra-testuali. Weinstein (Harvey) è il nome del produttore cinematografico statunitense accusato in un'inchiesta del *New York Times* del 6 ottobre 2017 di molestie sessuali ai danni di alcune attrici di Hollywood (nome distintivo di questa classe). Catherine Deneuve è il nome della nota attrice francese che, nel gennaio 2018, è stata al centro di una polemica per aver firmato una lettera aperta pubblicata dal quotidiano *Le Monde* condannando *la caccia alle streghe* (STT 2, Tabella 32), ovvero la ricerca di attrici vittime di molestie sessuali avviata dal movimento *me_too* – altra parola significativamente associata a questa classe (Appendice, Tab. 10) – dopo lo scandalo suscitato dalle accuse di molestie contro Weinstein¹⁰³.

Il lemma più significativo e anche più frequente di questa classe è *protesta*, che insieme a *manifestazione*, *marcia*, *movimento*, *scontro*, *mobilitazione* (Tabella 21), *protestare*, *manifestare*, *corteo*, *rivolta* (Appendice, Tab. 10) e al già citato

¹⁰³ Fonte: https://www.lemonde.fr/idees/article/2018/01/09/nous-defendons-une-liberte-d-importuner-indispensable-a-la-liberte-sexuelle_5239134_3232.html; ultimo accesso 19 novembre 2021.

me_too, identifica questa classe come l'area lessicale del dissenso. Un dissenso che si connota come pubblico (vs. privato), collettivo (vs. soggettivo) e femminista. I nomi di luogo significativamente associati a questa classe indicano infatti spazi pubblici reali, come la *piazza* e la *città*, o virtuali, come *Facebook* e i *social* (Tab. 10, Appendice); i verbi *partecipare* (Tabella 21) e *organizzare* (Tab. 10, Appendice) e il nome *solidarietà* richiamano una dimensione condivisa da un gruppo di persone; infine gli aggettivi *femminista* e *femminile* (Tab 10, Appendice) identificano questo collettivo rispetto al genere femminile, anche in senso militante.

Alcune parole consentono poi di comprendere l'estensione di quest'area lessicale a una dimensione politica (*ministro, governo, politico*) che va oltre la protesta contro le molestie sessuali e la violenza di genere (*violenza, stupro, abuso*) fino a comprendere, in generale, azioni di rivendicazione di diritti (*diritto, libertà*), e in particolare, azioni di denuncia (*difendere, difesa, denunciare*) e, infine, a individuarne il *framing* metaforico della lotta (*battaglia, battere, combattere*). Si noti che fra i ST tipici ci sono frammenti di cronaca su proteste non solo contro le molestie e la violenza di genere, ma anche contro discriminazioni religiose, che riguardano trasversalmente donne e uomini, come quelle denunciate dalle musulmane indiane (es. STT 26, Elenco 11), scese in piazza per contestare la riforma legislativa per la concessione del diritto di cittadinanza a rifugiati e immigrati irregolari afgani, bengalesi e pakistani, ma non di fede islamica¹⁰⁴. O discriminazioni sociali, come quelle denunciate dal movimento dei gilet gialli (es. STT 25, Elenco 11), movimento, quest'ultimo, nato in Francia nel 2018 inizialmente come reazione all'aumento del costo del carburante e progressivamente come protesta contro l'elevato costo della vita e le politiche fiscali del governo gravanti soprattutto sulle lavoratrici e sui lavoratori (Costa 2019; Michelot 2019).

Infine, fra le parole più frequenti e significative, compaiono due nomi propri, *Giovanna* e *Botteri*, significativamente associati a questa classe, e indicano il nome

¹⁰⁴ Fonte: <https://www.ilfattoquotidiano.it/2019/12/13/india-approvata-legge-per-la-cittadinanza-ai-migranti-illegali-ma-non-per-i-musulmani-proteste-nel-nord-est-migliaia-di-arresti/5614551/>; ultimo accesso 19 novembre 2021.

e cognome della giornalista Rai corrispondente dagli Stati Uniti dal 2007 al 2019 e dalla Cina dal 2019 al 2020¹⁰⁵.

Tabella 21. Forme lessicali con freq. N>5 e $\chi^2>3,84$ ($p<0,05$) della classe 3 del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	protesta	20	21	77,18	<0,0001
1	sessuale	19	21	67,75	<0,0001
2	molestia	16	17	60,34	<0,0001
3	Trump	14	14	57,66	<0,0001
4	manifestazione	14	15	52	<0,0001
5	Weinstein	11	11	45,14	<0,0001
6	marcia	13	15	42,99	<0,0001
7	corrispondente	10	10	40,98	<0,0001
8	tutte_le	11	12	39,59	<0,0001
9	americano	15	20	39,31	<0,0001
10	movimento	12	14	38,95	<0,0001
11	Parigi	13	16	38,78	<0,0001
12	Stati_Uniti	13	17	35,08	<0,0001
13	Francia	11	13	34,92	<0,0001
14	New	8	8	32,71	<0,0001
15	vestire	9	10	31,4	<0,0001
16	piazza	16	26	29,43	<0,0001
17	scontro	7	7	28,58	<0,0001
18	giallo	7	7	28,58	<0,0001
19	Deneuve	7	7	28,58	<0,0001
20	diritto	17	29	28,5	<0,0001
21	Macron	8	9	27,34	<0,0001
22	Iran	8	9	27,34	<0,0001
23	motivo	6	6	24,47	<0,0001
24	Golden	6	6	24,47	<0,0001
25	Botteri	6	6	24,47	<0,0001
26	Hollywood	9	12	23,35	<0,0001
27	presidente	18	35	23,01	<0,0001
28	città	12	19	23,01	<0,0001
29	oggi	22	48	21,73	<0,0001
30	scendere	6	7	19,31	<0,0001
31	gilet	6	7	19,31	<0,0001
32	francese	6	7	19,31	<0,0001
33	nero	9	14	17,74	<0,0001
34	Giovanna	7	10	16,06	<0,0001
35	onda	6	8	15,49	<0,0001
36	mobilitazione	6	8	15,49	<0,0001
37	partecipare	6	8	15,49	<0,0001
38	libertà	8	13	14,48	0,00014
39	forte	11	21	14,39	0,00014
40	violenza	25	67	14,07	0,00017
41	continuare	7	11	13,48	0,00024
42	battaglia	6	9	12,58	0,00039

¹⁰⁵ Fonte: https://it.wikipedia.org/wiki/Giovanna_Botteri; ultimo accesso 19 novembre 2021.

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
43	attore	8	14	12,5	0,0004
44	difendere	9	17	12	0,00053
45	difesa	7	12	11,38	0,00074
46	scrivere	6	10	10,3	0,00133
47	abuso	6	10	10,3	0,00133
48	femminile	8	16	9,36	0,00221
49	tema	6	11	8,47	0,00361
50	America	6	11	8,47	0,00361
51	prossimo	7	14	8,17	0,00425
52	milione	8	17	8,12	0,00438
53	denunciare	8	17	8,12	0,00438
54	tempo	10	24	7,44	0,00639
55	ministro	6	13	5,77	0,01628
56	governo	7	17	4,99	0,02549
57	politico	8	21	4,54	0,03316

Elenco 11. Selezione di segmenti di testo tipici della classe 3 del *sub-corpus* donna

STT 1 **** *A_2019 *G_19-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 405,49

e queste invece sono le donne **americane** che sono **scese** in **piazza** per il **terzo** anno consecutivo da **Washington** a **New York** con **slogan** contro le **molestie** e le discriminazioni in alcuni casi ad essere **preso** di mira è stato proprio il **presidente Trump** la **corrispondente Giovanna Botteri**

STT 2 **** *A_2018 *G_09-01 *T_TG2 *N_Lancio *P_Giornalista *S_Uomo
score: 363,38

molestie sessuali in **Francia Catherine Deneuve** e altre 100 donne artiste **scrittrici** accademiche con una **lettera** una **lettera aperta** su **Le Monde** definiscono controproducente la **caccia** alle streghe seguita al caso **Weinstein** da **Parigi** Iman Sabbah

STT 4 **** *A_2018 *G_08-01 *T_TG2 *N_Lancio *P_Giornalista *S_Donna
score: 350,13

negli **Stati Uniti** la **protesta** di **attrici** e **attori vestiti** di **nero** contro le **molestie sessuali** alla **cerimonia** dei **Golden Globe** ha portato alla **ribalta** la donna che potrebbe essere secondo i giornali e i **social** la **prossima** candidata alle presidenziali del 2020 vediamo chi è con a **Maria Lepri**

STT 8 **** *A_2019 *G_06-01 *T_TG2 *N_Titolo *P_Giornalista *S_Donna
score: 311,89

dopo le **proteste** e gli **scontri** di ieri a **Parigi** e in molte città francesi **oggi manifestazione** delle donne contro **Macron** è il primo **corteo femminista** dei **gilet gialli**

STT 26 **** *A_2020 *G_13-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 208,63

la **protesta** dei **musulmani indiani** in **piazza** da settimane contro una legge che li **discrimina** e le donne sono in prima linea la **corrispondente Giovanna Botteri**

Classe 4: l'“occupazione femminile”

La classe 4 copre il 12,44% dei segmenti di testo (N=103) e ha un'associazione, ma non significativa, con la modalità *N_Servizio ($\chi^2(1, N=551)=2,77, p=0,096$).

Dal punto di vista semantico si caratterizza per una chiara omogeneità interna, che la identifica come l'area lessicale della partecipazione delle donne al mercato del lavoro. *Lavoro, lavoratore, occupazione, lavorare* sono parole che fanno esplicito riferimento a questo *focus* tematico che si declina in alcuni discorsi principali, storicamente associati all'occupazione femminile: la parità salariale, le condizioni pensionistiche, il divario nord-sud.

La parola *parità* in 9 STT su 11 ricorre associata all'aggettivo qualificativo *salariale* o al complemento di specificazione (*di*) *salario*, entrambi lemmi distintivi di questa classe, insieme a parole con riferimento allo stesso ambito semantico quali *guadagnare, reddito e stipendio* (Appendice, Tab. 11). *Salariale* e *stipendio*, insieme a *retributivo*, sono anche le collocazioni di *divario*, un antonimo lessicale di *parità* che contribuisce a profilare questo *topic*.

A questioni pertinenti il sistema pensionistico fanno esplicito riferimento le parole *opzione, quota_100 e sociale*, che necessitano di alcune informazioni, anche extra-testuali, per essere collocate entro questo *focus* tematico. Nei ST *opzione* è un collocato di *donna*, e fa riferimento alla cosiddetta *Opzione donna*¹⁰⁶, un trattamento pensionistico erogato alle lavoratrici dipendenti e autonome, a partire dal 2017, con la legge di bilancio che ha introdotto l'APE, acronimo di Anticipo Pensionistico, una misura sperimentale finalizzata ad accompagnare persone in determinate condizioni all'età pensionabile¹⁰⁷. *Quota 100* è la dicitura con cui è stata nominata la misura introdotta dal disegno di legge del 28 gennaio 2019 che prevede, per il triennio 2019-2021, il diritto a conseguire la pensione anticipata, in

¹⁰⁶ Fonte: <https://www.inps.it/nuovoportaleinps/default.aspx?itemdir=50300>; ultimo accesso 19 novembre 2021.

¹⁰⁷ Fonte: <https://www.inps.it/nuovoportaleinps/default.aspx?itemdir=50302>; ultimo accesso 19 novembre 2021.

presenza di un requisito anagrafico pari a 62 anni e di un'anzianità contributiva minima di 38 anni¹⁰⁸. Infine, l'aggettivo *sociale*, in 7 ST su 16 è collocato di *APE* (Appendice, Tab. 11).

Il tema del divario nord-sud è indicato dall'occorrenza di questi due termini nelle parole significativamente associate a questa classe e confermato dalla lettura dei segmenti di testo tipici (es. STT 5, 44, Elenco 12).

Tabella 22. Forme lessicali con freq. N>5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 4 del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	lavoro	27	45	98,82	<0,0001
1	sociale	16	20	85,89	<0,0001
2	lavoratore	12	12	85,71	<0,0001
3	opzione	11	11	78,47	<0,0001
4	salariale	10	10	71,25	<0,0001
5	parità	11	14	57,18	<0,0001
6	guadagnare	8	8	56,86	<0,0001
7	quota_100	7	7	49,69	<0,0001
8	decreto	7	7	49,69	<0,0001
9	reddito	8	9	48,82	<0,0001
10	sud	10	14	45,49	<0,0001
11	occupazione	6	6	42,54	<0,0001
12	salario	6	6	42,54	<0,0001
13	divario	6	6	42,54	<0,0001
14	ONU	6	6	42,54	<0,0001
15	azienda	8	10	42,42	<0,0001
16	sottolineare	7	8	41,78	<0,0001
17	dipendente	7	9	35,66	<0,0001
18	legge	8	14	26,13	<0,0001
19	punto	8	15	23,45	<0,0001
20	giovane	14	38	21,77	<0,0001
21	governo	8	17	19,1	<0,0001
22	euro	8	17	19,1	<0,0001
23	pubblico	7	16	14,68	0,00012
24	lavorare	9	24	14,25	0,00015
25	storia	8	23	10,84	0,00099
26	uomo	32	159	10,67	0,00108
27	paese	10	39	6,55	0,01049
28	mese	6	20	5,8	0,016

¹⁰⁸ Decreto Legge 28 gennaio 2019, n. 4; fonte: <https://www.gazzettaufficiale.it/eli/id/2019/01/28/19G00008/sg>; ultimo accesso 19 novembre 2021.

Elenco 12. Selezione di segmenti di testo tipici della classe 4 del *sub-corpus* donna

STT 1 **** *A_2019 *G_18-01 *T_TG1 *N_Servizio *P_Giornalista *S_Donna

score: 459,06

pagando una somma fissa di 5241 euro per ogni anno di **studio** il **decreto** inoltre proroga di un anno l' **APE sociale** ed **opzione** donna che **consente** alle **lavoratrici** che nel 2018 hanno compiuto 58 anni se dipendenti 59 se autonome

STT 4 **** *A_2018 *G_20-01 *T_TG1 *N_Servizio *P_Giornalista *S_Uomo

score: 411,88

retribuite operano in **settori** a **basso reddito** sono meno **rappresentate** nei **livelli** più alti delle **aziende** ma anche semplicemente perché **ricevono** in **media salari** più **bassi** rispetto ai loro **colleghi** maschi per fare esattamente_lo stesso **lavoro** la stima finale che per ogni **dollaro guadagnato** da un **uomo** una donna **guadagna** in **media** 77 centesimi5

STT 5 **** *A_2018 *G_18-01 *T_TG2 *N_Servizio *P_Giornalista *S_Donna

score: 411,24

l' **occupazione specie** al **sud** e tra le donne per Gentiloni è la **priorità** della prossima legislatura **occorre lavorare** sui **sistemi salariali ridurre** le **povertà** le **disuguaglianze sociali** e sui **conti pubblici** che sono molto in ordine **sottolinea** pone l'accento sul deficit ridotto dal 3 al 2

STT 9 **** *A_2020 *G_29-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo

score: 379,95

penso alla **parità** di **salario** tra **uomo** donna come grande **scelta** di campo di lotta alle **disuguaglianze** quindi sviluppo **crescita lavoro** e **giustizia sociale** ed è **possibile** se da **subito** si apre una nuova stagione di concretezza

STT 12 **** *A_2020 *G_30-01 *T_TG2 *N_Servizio *P_Giornalista *S_Uomo

score: 324,00

una fase 2 nel segno delle **certezze** quelle da dare al **paese** con una **maggioranza** unita e un **governo** solido **Zingaretti** insiste sulle modifiche ai **decreti sicurezza** e rilancia i **temi** chiave dell' **azienda** di **parità salariale uomo** donna **scuola**

STT 44**** *A_2018 *G_31-01 *T_TG1 *N_Servizio *P_Giornalista *S_Uomo

score: 176,20

il **premier** dobbiamo puntare al rafforzamento della **crescita** in **particolare** favorendo le assunzioni di **giovani** e donne facendo diminuire le **disuguaglianze** come quelle geografiche tra **nord** e **sud** impegno garantito sull' Ilva no invece al protezionismo degli avversari **penalizzerebbe** l' **economia**

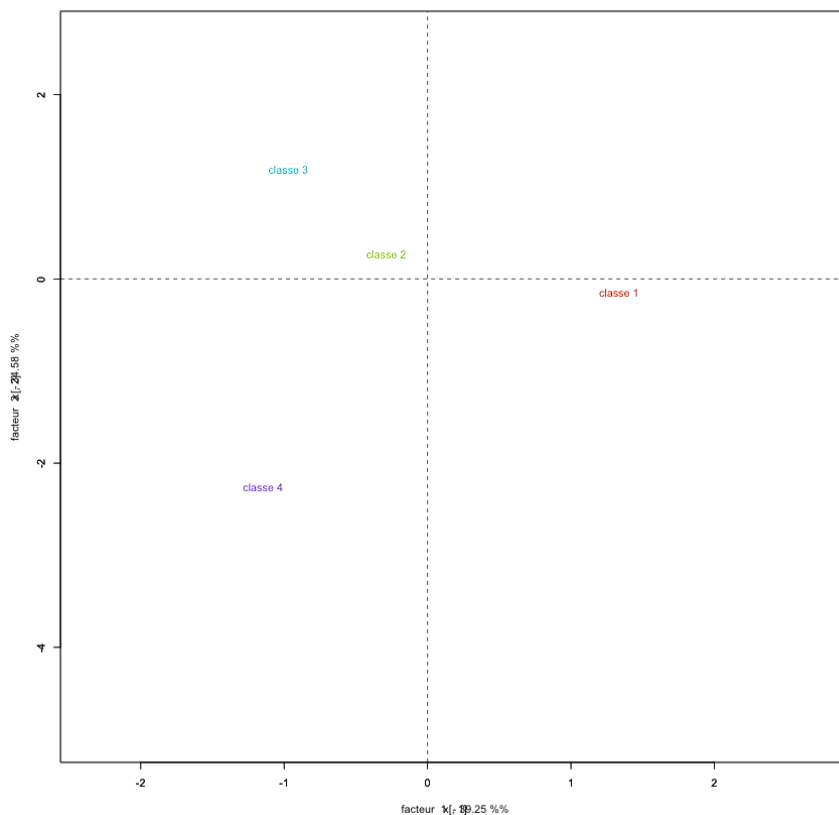
La proiezione dei risultati della CGD su piano fattoriale, illustrata nel grafico seguente, posiziona le quattro classi del *sub-corpus* donna rispetto all'asse delle ascisse, rappresentato dal fattore 1, che spiega il 39,25% della varianza, e all'asse delle ordinate, rappresentato dal fattore 2, che spiega il 34,58%.

Considerando le caratteristiche della classi lessicali, il primo asse, quello orizzontale, può essere interpretato come quello della polarità "pubblico" (sulla sinistra) vs. "privato" (sulla destra): sul versante destro si trovano infatti, a partire dall'alto, la classe 3, che come abbiamo visto si caratterizza per raggruppare le parole del dissenso e della protesta, in particolare contro le molestie sessuali e la

violenza di genere, organizzata e manifestata pubblicamente; poi la classe 2, un'area ibrida che raccoglie diverse posizioni sul ruolo delle donne, nella Chiesa, nella politica o nella società, comunque in ambito pubblico e non nella famiglia o nell'ambito di relazioni intime e private, e infine la classe 4 il cui universo lessicale è quello della "occupazione femminile", come tema al centro di decisioni e misure di politiche pubbliche. Mentre la classe 1, che concentra le parole della cronaca sulla violenza, in particolare la violenza domestica contro le donne, si trova sul versante opposto, che è, appunto, quello della sfera privata.

L'asse verticale potrebbe essere interpretato come l'asse della "soggettività" (in basso) vs. la "oggettività" (in alto). Sopra l'asse delle ascisse sono infatti posizionate la classe 3, più in alto, e la classe 2, più in basso, che si distinguono per rappresentare le donne o come protagoniste di azioni o manifestazioni di protesta e dissenso (classe 3) o come protagoniste di ricorrenze che ne celebrano il ruolo come portatrici di pace o memoria – in momenti simbolici e significativi, come la giornata mondiale per la pace o la giornata della memoria – oppure protagoniste della politica, della storia o dell'arte (classe 2). In ogni caso, le donne sono rappresentate come soggetti attivi. Sotto l'asse delle ascisse sono invece posizionate la classe 1 e la classe 4 che si distinguono seppur diversamente – sono infatti posizionate la prima a destra e la seconda a sinistra dell'asse delle ordinate – per una oggettivazione delle donne, che non sono rappresentate come soggettività attive, ma, nel primo caso, come oggetto di violenza, e, nel secondo caso, come oggetto di politiche o misure per il lavoro specificamente pensate per le donne.

Grafico 18. Rappresentazione fattoriale delle 4 classi lessicali del *sub-corpus* donna



La CGD del *sub-corpus* uomo ha restituito quattro classi lessicali che coprono l'80,20% dei segmenti di testo: N=887 sul totale di N=1.106 in cui il *corpus* è stato suddiviso.

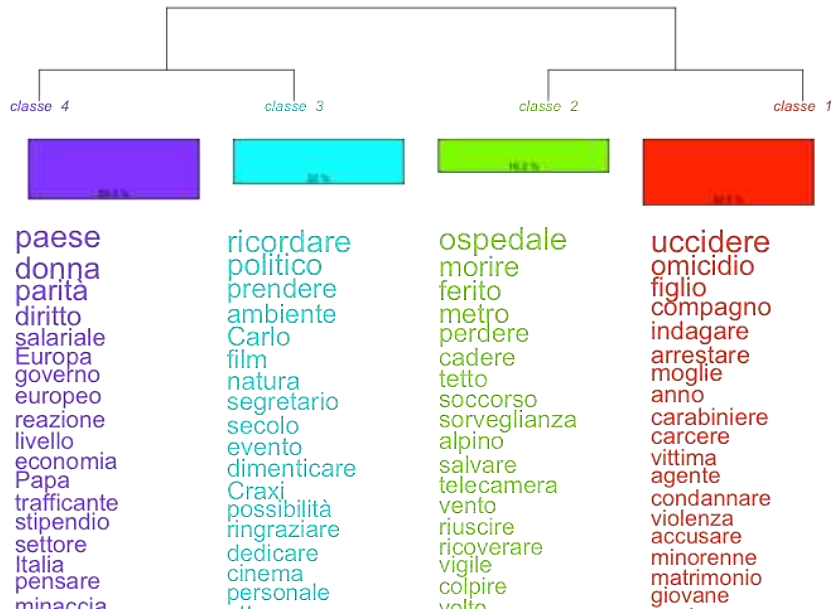
Tabella 23. Statistiche: risultati della CGD del *sub-corpus* uomo

Testi (N)	1.002
Segmenti di testo (N)	1.106
Forme (N)	8.685
Occorrenze (N)	42.255
Lemmi (N)	6.313
Forme (lemmi) attive (N)	5.809
Forme (lemmi) supplementari (N)	451
Forme (lemmi) attive con frequenza ≥ 3	1.675
Media di forme per segmento	38,20
Classi (N)	4
Segmenti classificati (80,20%)	887

Il dendrogramma illustrato nel Grafico 19 mostra la relazione gerarchica fra le quattro classi. Una prima segmentazione, raggruppa, sulla destra, le classi 1 e 2,

sulla sinistra, le classi 3 e 4, segnando la distanza fra loro. Una seconda segmentazione, isola, su entrambi i lati, le due classi, in relazione simmetrica fra loro.

Grafico 19. Dendrogramma delle classi lessicali del *sub-corpus* uomo



Classe 1: la “criminalità e la violenza contro le donne”

La classe 1 copre il 32,47% dei segmenti di testo (N=288), ed è significativamente associata alle modalità *P_Giornalista ($\chi^2(1, N=799)=34,74, p<0,001$) e *N_Lancio ($\chi^2(1, N=111)=22,65, p<0,001$). Queste correlazioni significative indicano che questa classe si distingue per raggruppare segmenti di testo contenenti notizie anticipate in studio dal giornalista o dalla giornalista che conduce il TG, come per la classe 1 del *sub-corpus* donna.

In termini di contenuti tematici, parole come *uccidere, omicidio, vittima, violenza, corpo, ucciso, abusare, rubare, investire, aggredire* rimandano all’ambito della criminalità e della violenza, anche di matrice sessuale – *sessuale* è collocato di *violenza* in 5 ST su 6 – e più in generale di genere – *femminicidio* è infatti un termine distintivo di questa classe (Appendice, Tab. 12). Allo stesso *topic* rinviano, da un lato, i nomi che identificano le forze dell’ordine e i/le rappresentanti della giustizia (*carabiniere, agente, polizia, inquirenti, mobile, investigatore, giudice*), e le loro

azioni (*indagare, arrestare, condannare, accusare, allontanare, denunciare, accusa, fermo, indagine, fermare, arresto*); dall'altro, le parole che identificano le persone reo confesse o sospette di reato (*armato, assassino*) e le loro azioni (*uccidere, confessare, abusare, rubare, investire, aggredire*).

Un altro gruppo di parole fa riferimento a relazioni e/o identità basate su relazioni affettive o familiari: *figlio, compagno, moglie, matrimonio, madre, marito, relazione, fidanzato, coppia, familiare, padre*. Insieme a queste, sono attestate parole con riferimento a identità generazionali, in particolare alla generazione giovani (*minorenne, giovane, ragazzo, bimbo*) e le parole con riferimento a identità etniche (*marocchino, cinese*).

Un gruppo di parole con riferimenti spazio-temporali profila infine la narrazione di questa classe come cronachistica. *Anno, dicembre, scorso, pomeriggio, mese, giorno* indicano un lasso di tempo o un periodo preciso; *carcere, procura, Alessandria, casa, Milano, provincia, Napoli* denotano luoghi specifici. In particolare, l'elenco completo delle parole distintive di questa classe riporta molti nomi di città, come *Novara, Imola, Cremona, Messina, Torino, Ostia* e *Siena* (Appendice, Tab. 12), e aggettivi che indicano specifiche zone geografiche o abitanti di un territorio, come *napoletano, bolognese, alessandrino*, circoscrivendo l'universo di riferimento di questa classe lessicale entro confini di prossimità nazionale.

Tabella 24. Forme lessicali con freq. N>5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 1 del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	uccidere	45	50	79,98	<0,0001
1	omicidio	27	27	57,92	<0,0001
2	figlio	47	65	50,77	<0,0001
3	compagno	27	32	40,79	<0,0001
4	indagare	18	18	38,21	<0,0001
5	arrestare	21	24	34,07	<0,0001
6	moglie	24	30	31,99	<0,0001
7	anno	84	168	29,05	<0,0001
8	carabiniere	21	26	28,5	<0,0001
9	carcere	15	16	27,91	<0,0001
10	vittima	30	47	22,26	<0,0001
11	agente	10	10	21,04	<0,0001
12	condannare	10	10	21,04	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
13	violenza	21	30	19,95	<0,0001
14	accusare	9	9	18,91	<0,0001
15	minorenne	9	9	18,91	<0,0001
16	matrimonio	9	9	18,91	<0,0001
17	giovane	16	21	18,75	<0,0001
18	madre	12	14	18,39	<0,0001
19	corpo	19	27	18,24	<0,0001
20	marocchino	8	8	16,79	<0,0001
21	allontanare	8	8	16,79	<0,0001
22	denunciare	11	13	16,36	<0,0001
23	polizia	24	39	15,72	<0,0001
24	leggere	9	10	15,27	<0,0001
25	procura	7	7	14,67	0,00012
26	fermo	7	7	14,67	0,00012
27	Alessandria	7	7	14,67	0,00012
28	confessare	7	7	14,67	0,00012
29	incidente	11	14	13,79	0,00020
30	marito	11	14	13,79	0,00020
31	tentare	11	14	13,79	0,00020
32	accusa	12	16	13,44	0,00024
33	cellulare	8	9	13,2	0,00027
34	cronaca	8	9	13,2	0,00027
35	casa	26	46	12,8	0,00034
36	trovare	21	35	12,6	0,00038
37	ucciso	6	6	12,56	0,00039
38	relazione	6	6	12,56	0,00039
39	fidanzato	6	6	12,56	0,00039
40	indagine	6	6	12,56	0,00039
41	abusare	6	6	12,56	0,00039
42	dicembre	10	13	11,89	0,00056
43	scorso	16	25	11,66	0,00063
44	Milano	11	15	11,62	0,00065
45	provincia	12	17	11,49	0,00070
46	ragazzo ragazza	14	21	11,47	0,00070
47	ritrovare	13	19	11,45	0,00071
48	rubare	7	8	11,15	0,00084
49	precedente	7	8	11,15	0,00084
50	mobile	7	8	11,15	0,00084
51	clan	7	8	11,15	0,00084
52	coppia	8	10	10,42	0,00124
53	bimbo	8	10	10,42	0,00124
54	armato	6	7	9,12	0,00252
55	pistola	6	7	9,12	0,00252
56	incontro	6	7	9,12	0,00252
57	assassino	6	7	9,12	0,00252
58	cinese	7	9	8,51	0,00352
59	sessuale	7	9	8,51	0,00352
60	investigatore	7	9	8,51	0,00352
61	giudice	7	9	8,51	0,00352
62	pomeriggio	8	11	8,23	0,00411
63	trattare	9	13	8,13	0,00435
64	fermare	13	22	7,29	0,00692

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
65	mese	12	20	7,07	0,00782
66	pericoloso	6	8	6,66	0,00986
67	volontario	6	8	6,66	0,00986
68	servizio	7	10	6,5	0,01080
69	confronto	7	10	6,5	0,01080
70	video	8	12	6,49	0,01085
71	operaio	8	12	6,49	0,01085
72	cercare	14	26	5,58	0,01814
73	arresto	8	13	5,08	0,02413
74	investire	7	11	4,93	0,02632
75	vicenda	6	9	4,85	0,02765
76	squadra	6	9	4,85	0,02765
77	aggredire	6	9	4,85	0,02765
78	familiare	6	9	4,85	0,02765
79	Napoli	10	18	4,47	0,03457
80	secondo	21	45	4,36	0,03683
81	giorno	24	53	4,22	0,03992
82	tornare	9	16	4,2	0,04036
83	passato	9	16	4,2	0,04036
84	padre	8	14	3,95	0,04688

Elenco 13. Selezione di segmenti di testo tipici della classe 1 del *sub-corpus* uomo

STT 1 **** *A_2020 *G_06-01 *T_TG1 *N_Lancio *P_Giornalista *S_Uomo
score: 284,93

Imola investe e uccide un marocchino perché aveva **rubato il cellulare al figlio** l'uomo **operaio** di 44 anni è stato **arrestato** nei **giorni scorsi** tra i due era già scoppiata una rissa **Giacinto Pinto**

STT 2 **** *A_2019 *G_30-01 *T_TG2 *N_Lancio *P_Giornalista *S_Donna
score: 262,21

convalidato il **fermo** dell'uomo che ha **ucciso il figlio** di 7 anni della **compagna** nel **napoletano** la **madre** ha raccontato ha **cercato** di difenderlo il nostro inviato è **Lorenzo Santorelli**

STT 5 **** *A_2018 *G_24-01 *T_TG2 *N_Lancio *P_Giornalista *S_Uomo
score: 239,96

ancora un **femminicidio** è il quinto dall'inizio dell'**anno** come vedete nel nostro contatore a **Cremona** un uomo di nazionalità **cinese** ha **ucciso la moglie** ed un **bimbo** di 3 anni **figlio** di una **coppia** di amici Paola Colombo

STT 13 **** *A_2019 *G_17-01 *T_TG5 *N_Servizio *P_Giornalista *S_Uomo
score: 222,56

una pattuglia di **carabinieri** sta vegliando sulla **casa** dell'ex **compagna** dell'uomo in **fuga** il timore che possa **tornare** qui e **ucciderla** le ricerche non hanno ancora dato esito e non si **arresteranno** c'è un uomo in **fuga armato** ed è **ritenuto pericoloso**

STT 155 **** *A_2020 *G_29-01 *T_TG2 *N_Servizio *P_Giornalista *S_Uomo
score: 75,04

i picchiatori di un **clan** criminale utilizzati per riscuotere **soldi** ma anche per la campagna elettorale l'ex consigliera PDL del Lazio Gina Cetrone è stata **arrestata** dalla **mobile** di Latina insieme al **marito** e a altri 3 uomini del **clan**

Classe 2: gli “incidenti di varia natura”

La classe 2 raggruppa N=144 segmenti di testo che coprono il 16,23% del corpus, ed è significativamente associata alle modalità *P_Giornalista ($\chi^2(1, N=799)=12, p<0,05$), *S_Donna ($\chi^2(1, N=475)=7,39, p<0,05$) e *N_Servizio ($\chi^2(1, N=609)=3,95, p<0,05$). Queste tre correlazioni significative forniscono un'informazione utile a identificare le UCE raggruppate in questa classe come frammenti di testo di servizi firmati da giornaliste.

Dal punto di vista del contenuto tematico, le parole che la distinguono rimandano a incidenti di varia natura, anche letale (*morire*), ovvero avvenimenti inattesi che interrompono il corso regolare di un'azione, per lo più non lieti, disgrazie¹⁰⁹, come suggerito da verbi quali *perdere, cadere, colpire, travolgere, trascinare* e dall'aggettivo *ferito* e dal nome *ferita*. Il filo conduttore che unisce questo gruppo di segmenti di testo è la comune focalizzazione su luoghi e soprattutto su attività di prevenzione o soccorso evocati esplicitamente da termini quali *ospedale, soccorso, sorveglianza, salvare, telecamera, riuscire, ricoverare, trasportare, guardia* (Tabella 25), *soccorritore, ambulanza, medico* (Appendice, Tab. 13). Alcuni gruppi di parole consentono di rilevare diversi ambiti di riferimento extra-testuale e distinguere la natura di questi avvenimenti. *Metro, treno* (Tabella 25) *vagone, metropolitana, passeggero, stazione, convoglio*, (Appendice, Tab. 13) fanno riferimento al settore dei trasporti pubblici identificando il *focus* tematico degli incidenti ferroviari. *Autostrada* e *automobile* (Appendice, Tab. 13) fanno invece riferimento ai viaggi su strada, suggerendo la presenza di un discorso sugli incidenti stradali, confermata dai segmenti di testo tipici (es. STT 10, 13, Tabella 39). *Alpino, vento, neve* (Tabella 25), *raffica, sferzata, montagna, naufragio, mareggiata, maltempo, ghiacciaio, ghiaccio, costiero, aria* (Appendice, Tab. 13) sono tutti nomi che suggeriscono il *focus* tematico su calamità naturali in particolare in località di mare o montagna, come attestato da alcuni segmenti di testo tipici (es. STT 25, 32, Tabella 39). Infine, *fiamma, fuoco* (Tabella 25) e *rogo*

¹⁰⁹ Si veda la definizione del vocabolario Treccani online; fonte <https://www.treccani.it/vocabolario/incidente2/> (ultimo accesso 19 novembre 2021).

(Appendice, Tab. 13) sono nomi specifici di effetti di eventi di natura incendiaria e completano il quadro di questa classe lessicale, che potremmo definire della cronaca su incidenti di varia natura (es. STT 44, 49, Elenco 14), di cui gli uomini, intesi come singoli, sono rappresentati come vittime o sopravvissuti.

Tabella 25. Forme lessicali con freq. $N > 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 2 del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ.(N)	χ^2	p-value
0	ospedale	20	25	76,92	<0,0001
1	morire	28	51	59,49	<0,0001
2	ferito	11	11	57,47	<0,0001
3	metro	17	24	54,07	<0,0001
4	perdere	13	17	46,24	<0,0001
5	cadere	9	10	40,47	<0,0001
6	tetto	10	12	40,27	<0,0001
7	soccorso	11	15	36,58	<0,0001
8	sorveglianza	7	7	36,41	<0,0001
9	alpino	7	7	36,41	<0,0001
10	salvare	12	18	34,36	<0,0001
11	telecamera	11	16	33,04	<0,0001
12	vento	9	12	30,89	<0,0001
13	riuscire	13	22	30,47	<0,0001
14	ricoverare	8	10	30,24	<0,0001
15	vigile	8	10	30,24	<0,0001
16	colpire	15	29	27,77	<0,0001
17	volto	8	11	26,14	<0,0001
18	grave	10	16	25,65	<0,0001
19	condizione	7	9	25,32	<0,0001
20	quarto	6	7	25,05	<0,0001
21	ferita	6	7	25,05	<0,0001
22	trasportare	6	7	25,05	<0,0001
23	neve	6	7	25,05	<0,0001
24	Roma	15	31	24,42	<0,0001
25	volo	6	8	20,5	<0,0001
26	fiamma	6	8	20,5	<0,0001
27	immagine	11	22	18,91	<0,0001
28	treno	7	11	18,4	<0,0001
29	fuoco	8	14	17,5	<0,0001
30	travolgere	6	9	17,01	<0,0001
31	trascinare	6	9	17,01	<0,0001
32	rimanere	10	22	14,16	0,00016
33	sentire	10	23	12,89	0,00033
34	lanciare	6	11	12,02	0,00052
35	porre	8	18	10,75	0,00104
36	notte	8	19	9,56	0,00199
37	punto	7	16	9,07	0,00259
38	persona	15	47	8,97	0,00273
39	polizia	13	39	8,77	0,00306
40	guardia	6	14	7,41	0,00647
41	mettere	9	25	7,39	0,00655

ID	Lemma	ST (N)	Occ.(N)	χ^2	p-value
42	chiamare	8	22	6,72	0.00952
43	momento	8	23	5,97	0.01452
44	terra	6	16	5,42	0.01992
45	cercare	8	26	4,16	0.04136
46	fermare	7	22	4,03	0.04473
47	Napoli	6	18	3,95	0.04686

Elenco 14. Selezione di segmenti di testo tipici della classe 2 del *sub-corpus* uomo

STT 3 **** *A_2018 *G_17-01 *T_TG1 *N_Titolo *P_Giornalista *S_Donna

score: 319,05

I Italia **sferzata** dal **vento raffiche** fino a 200 **orari muore** un uomo in **Calabria** era **salito** sul **tetto** dell **abitazione** danneggiata a **Roma** 2 **feriti** da **alberi caduti** una tegola **colpisce** una studentessa

STT 10 **** *A_2019 *G_03-01 *T_TG1 *N_Servizio *P_Giornalista *S_Donna

score: 262,28

all **ospedale** di Lodi arrivano anche 2 **bambini** di 8 anni stanno bene verranno **dimessi** qualche ora dopo **ricoverato** a Parma in **gravi condizioni** l uomo investito il **branco** di **cinghiali** ha **scavato** un cunicolo nei campi così è **riuscito** a raggiungere l **autostrada** superando la rete di **protezione**

STT 13 **** *A_2018 *G_27-01 *T_TG2 *N_Servizio *P_Giornalista *S_Donna

score: 241,31

gravissime una trentina i **feriti** arrivati negli **ospedali** di Genova oltre 50 uomini a prestare **soccorso** con i mezzi **incastrati** uno sull altro e l **autostrada** chiusa per 8 ore a innescare i tamponamenti pare il gasolio **perso** da un camion

STT 25 **** *A_2018 *G_17-01 *T_TG5 *N_Servizio *P_Giornalista *S_Donna

score: 202,62

ad Acilia un uomo è stato **colpito** da un ramo e **trasportato** in **ospedale** in **codice rosso** in queste **immagini** del Tirreno Livorno **sferzata** da una violenta **mareggiata** che ha costretto i traghetti a **fermarsi** all isola di Capraia problemi per

STT 32 **** *A_2018 *G_17-01 *T_TG5 *N_Servizio *P_Giornalista *S_Donna

score: 175,75

era **salito** sul **tetto** di casa per **verificare** i **danni** del **maltempo** ma è **caduto** nel vuoto davanti allo sguardo del fratello così nelle ore in cui l Italia è **sferzata** dal forte **vento** e da violente **mareggiate** un uomo di

STT 44 **** *A_2020 *G_22-01 *T_TG1 *N_Lancio *P_Giornalista *S_Donna

score: 154,83

si è ustionato nel **disperato tentativo** di **salvarla** dalle **fiamme** ma non ce l ha fatta è il **dramma** di un uomo che ha visto **morire** la figlia quattordicenne nel **rogo** di una villette in lucchesia Andrea Marotta

STT 49 **** *A_2020 *G_12-01 *T_TG5 *N_Servizio *P_Giornalista *S_Uomo

score: 150,18

Galles del **sud** e a Vittoria in **particolare** a Tambaram un uomo è **rimasto gravemente** ustionato mentre **provava** a **proteggere** la sua **abitazione** attaccata dalle **fiamme** un **tentativo disperato** che gli è costato caro ma sono soprattutto i **vigili** del **fuoco**

Classe 3: la “memoria collettiva”

La classe 3 copre il 21,98% del *sub-corpus* uomo, raggruppando N=195 segmenti, ed è significativamente associata alle modalità *P_Fonte ($\chi^2(1, N=88)=28,41, p<0,001$) e *N_Intervista ($\chi^2(1, N=66)=25,92, p<0,001$), questo significa che si distingue dalle altre classi per un raggruppamento di testi del parlato delle fonti, interpellate nel corso di interviste.

Le parole che la caratterizzano rinviano prevalentemente alla dimensione della memoria (*ricordare, dimenticare, conoscere*) e della sua narrazione (*scrivere, parlare*) vs. un'esposizione cronachistica dei fatti. *Storia* è infatti una parola distintiva di questa classe, insieme a *film* e a *libro* (Appendice, Tab. 14), nomi che indicano due formati narrativi diversi, l'uno cinematografico, l'altro letterario, accumulati dall'approfondimento e dalla ricostruzione degli avvenimenti vs. il racconto di eventi circoscritti nello spazio e nel tempo tipico della cronaca.

Diversi sono i nomi propri di persone famose che ricorrono in questo discorso pubblico della memoria che celebra, o semplicemente ricorda, il ruolo politico o sociale di uomini, che hanno in qualche modo segnato la storia recente o passata italiana e non. Il nome più frequente e significativamente associato a questa classe è *Carlo*, collocato di *Ciampi* (Appendice, Tab. 14), con riferimento al decimo presidente della Repubblica italiana. Un altro capo dello Stato presente è *Mattarella*, collocato sia di *Sergio*, dodicesimo presidente della Repubblica italiana, sia di *Piersanti*, il fratello del presidente che fu ucciso da Cosa nostra nel 1980. Altri nomi politici presenti sono quello di *Trump*, presidente degli Stati Uniti nel periodo 2018-2020, e di (Bettino) *Craxi*, ex presidente del consiglio italiano, il cui nome ricorre in frammenti di testo in cui compare anche il nome *Amelio* (Appendice, Tab. 14) con riferimento a Gianni Amelio, regista e autore del film biografico su Bettino Craxi *Hammamet* (Appendice, Tab. 14). C'è poi *Coppi*, con riferimento al ciclista Fausto, e, infine, *Weinstein* il produttore cinematografico americano al centro del già citato scandalo sulle molestie sessuali che ha dato impulso al movimento #MeToo.

Il principale riferimento temporale di questa classe è, come già detto, una storia dai confini imprecisati. L'unico nome di tempo attestato è *secolo*, che letteralmente indica un periodo di cent'anni, ma spesso, essendo la misura massima in uso per contare il tempo, è utilizzato per indicare genericamente un periodo molto lungo o per collocare un evento in un passato remoto¹¹⁰. Quanto ai luoghi, con la sola eccezione di *Livorno* (Appendice, Tab. 14), i nomi di distintivi di questa classe sono nomi propri di località estere: *Hammamet*, *Australia* e *America* (Appendice, Tab. 14). Poi ci sono nomi comuni polisemantici come *ambiente*, *natura* e *mondo* che indicano luoghi, reali o simbolici, imprecisati. *Mondo*, per esempio, che è fra i tre il più frequente in questa classe, può significare, sintetizzando le principali definizioni che ne dà il vocabolario Treccani online: 1) il cosmo, creato, universo, 2) il pianeta Terra, o una sua parte (in espressioni come *il nuovo(o vecchio) mondo*), 3) l'esistenza, la vita (in espressioni come *dare al mondo*, *venire al mondo*, *stare al mondo*, *tornare al mondo*), 4) il regno dei morti (*l'altro mondo*), 5) il genere umano, la gente anche nel senso di una determinata civiltà, cultura o società (*il mondo greco*) o di un determinato ambiente (*il mondo politico*, *il mondo dello spettacolo*), 6) la realtà (fisica o metafisica).

Tabella 26. Forme lessicali con freq. N>5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 3 del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	ricordare	15	19	36,73	<0,0001
1	politico	18	26	34,86	<0,0001
2	prendere	13	18	27,04	<0,0001
3	ambiente	7	7	25,04	<0,0001
4	Carlo	7	7	25,04	<0,0001
5	film	13	19	24,41	<0,0001
6	natura	9	11	23,25	<0,0001
7	segretario	6	6	21,44	<0,0001
8	secolo	6	6	21,44	<0,0001
9	evento	6	6	21,44	<0,0001
10	dimenticare	6	6	21,44	<0,0001
11	Craxi	6	6	21,44	<0,0001
12	personale	6	7	16,71	<0,0001
13	attore	6	7	16,71	<0,0001
14	compiere	6	7	16,71	<0,0001
15	unico	8	12	14,16	0,00016

¹¹⁰ Si veda la definizione di tempo nel vocabolario Treccani online: <https://www.treccani.it/vocabolario/secolo/> (ultimo accesso 19 novembre 2021).

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
16	crescere	6	8	13,23	0,00027
17	impegno	6	8	13,23	0,00027
18	Mattarella	6	8	13,23	0,00027
19	grande	19	43	12,99	0,00031
20	scrivere	9	15	12,86	0,00033
21	vita	21	51	11,62	0,00065
22	volere	18	42	11,2	0,00081
23	vivere	8	14	10,25	0,00136
24	aprire	8	14	10,25	0,00136
25	mondo	16	37	10,17	0,00142
26	storia	14	31	10,06	0,00151
27	comune	9	17	9,68	0,00185
28	presidente	13	29	9,12	0,00252
29	americano	8	15	8,74	0,00310
30	parlare	15	36	8,48	0,00359
31	milione	6	11	6,89	0,00869
32	intero	6	11	6,89	0,00869
33	credere	6	11	6,89	0,00869
34	famiglia	11	26	6,45	0,01109
35	diventare	8	17	6,35	0,01171
36	Stato	8	17	6,35	0,01171
37	capo	6	12	5,57	0,01830
38	conoscere	6	12	5,57	0,01830
39	entrare	8	19	4,58	0,03228
40	tornare	7	16	4,5	0,03388

Elenco 15. Selezione di segmenti di testo tipici della classe 3 del *sub-corpus* uomo

STT 1 **** *A_2020 *G_15-01 *T_TG2 *N_Servizio *P_Giornalista *S_Uomo

score: 177,58

uomo dell' **orizzonte comune** uomo dei tempi **difficili** così il **presidente Mattarella** ricorda **Carlo Azeglio Ciampi** nella sua **Livorno** nel centenario della **nascita** della lezione **politica** del suo predecessore al Quirinale **sottolinea** soprattutto il lascito morale il bene **comune** prima di tutto **Ciampi ricorda Mattarella** si definiva un cittadino al servizio dello **Stato**

STT 3 **** *A_2020 *G_08-01 *T_TG1 *N_Titolo *P_Giornalista *S_Uomo

score: 162,43

Pierfrancesco Favino protagonista di **Hammamet** di **Gianni Amelio** sugli **ultimi** mesi di **Bettino Craxi** racconto l'uomo e la sua **politica** dice il **regista** al Tg1 non sempre **condividendola** il **film** da domani nelle **sale**

STT 5 **** *A_2018 *G_23-01 *T_TG1 *N_Servizio *P_Giornalista *S_Donna

score: 127,50

la sfilata di alta moda di **Armani** anche stavolta si trasforma in un **evento** protagoniste le donne **famose** famosissime tutte coinvolte nel dibattito e **rilanciato** in Francia dalla **presa** di posizione di **Catherine Deneuve** e di 100 tra **intellettuali** e donne dello **spettacolo** poco convinte dalla caccia all'uomo scatenata **Hollywood** dopo il caso **Weinstein**

STT 9 **** *A_2020 *G_29-01 *T_TG5 *N_Servizio *P_Giornalista *S_Uomo

score: 115,44

L'America piange l'uomo che era riuscito ad **unire** anche la **politica americana** e i suoi **ultimi presidenti** da Bush ad Obama fino a **Trump** poi ci sono i fan quelli che amavano l'atleta Kobe e ognuno lo **ricorda** a proprio modo

STT 18 **** *A_2020 *G_29-01 *T_TG5 *N_Lancio *P_Giornalista *S_Donna

score: 93,69

andiamo in Gran_Bretagna perché una **storia** mai **dimenticata torna** ad essere un **vero** grattacapo per la **famiglia reale** un **uomo** che sostiene di essere il figlio **segreto** di **Carlo** e di **Camilla**

STT 24 **** *A_2019 *G_17-01 *T_TG1 *N_Notiziadastudio *P_Giornalista *S_Donna

score: 83,80

il 18 **gennaio** di 100 anni fa **don** Luigi Sturzo diffondeva l **appello** a tutti gli uomini liberi e forti e il programma che fondava il Partito_Popolare_Italiano un **evento** celebrato con una tavola rotonda all istituto Sturzo presenti il **capo** dello **Stato Sergio Mattarella** e la **presidente** del **Senato Elisabetta Alberti Casellati**

STT 28 **** *A_2020 *G_01-01 *T_TG2 *N_Lancio *P_Giornalista *S_Uomo

score: 81,35

e in tutt altro ambito **vogliamo ricordare** un **grande** uomo indimenticabile sportivo **Fausto Coppi** che ci lasciava il 2 **gennaio** del 1960 Laura Gialli

STT 64 **** *A_2020 *G_25-01 *T_TG2 *N_Servizio *P_Giornalista *S_Uomo

score: 57,94

ci sono donne che hanno tenuto in **piedi** d Italia il **mondo** mentre gli uomini si uccidevano fra loro **prendiamo** Chiara Lubich una bella ragazza trentina **nata giusto** 100 anni fa durante i bombardamenti su Trento si ritrova con le amiche

Classe 4: “parità di genere e diritti umani”

La classe 4 copre il 29,31% del *sub-corpus* uomo, con N=260 UCE, ed è significativamente associata alle modalità *N_Messaggio ($\chi^2(1, N=75)=20,61$, $p<0,001$), *P_Fonte ($\chi^2(1, N=88)=18,02$, $p<0,001$) e *N_Intervista ($\chi^2(1, N=66)=7,36$, $p<0,05$). Si tratta dunque di una classe che raggruppa in modo distintivo testi di messaggi o interviste rilasciate da fonti.

Dal punto di vista contenutistico, contiene frammenti di discorsi diversi semanticamente allineati dalla comune focalizzazione sulla diade donne-uomini. *Donna* è infatti la parola più frequente (N=83), e la seconda per chi-quadrato di associazione alla classe, la quale, ricordiamo, è il risultato di una CGD di un *sub-corpus* costruito selezionando dal *corpus* TG i segmenti contenenti il lemma *uomo*.

Il discorso principale e più rilevante è quello sulla parità di salario uomo-donna (*parità, salariale, stipendio, divario, lavoratore*), che si connota come tema politico sia nazionale (*paese, Italia*) sia europeo e internazionale (*Europa, europeo Davos, internazionale, mondo*).

La parola *diritto* è una delle parole più frequenti ed è attestata in diversi segmenti di testo con riferimento non solo alla parità salariale, ma anche ai diritti umani più in generale (es. STT 15, 19, 43, Elenco 16), i quali rappresentano un secondo *focus* tematico.

Un terzo argomento, secondario per frequenza e chi-quadrato di associazione delle parole che lo profilano, è quello pertinente al fenomeno migratorio e ai relativi decreti sicurezza, che fra il 2018 e il 2020 sono stati al centro dell'agenda politica nazionale. La Tabella 27, che riporta le forme con frequenza superiore a $N=5$ e $\chi^2 > 3,484$, $p < 0,05$, attesta infatti l'occorrenza di parole come *trafficante, sicurezza, mare, migrare*, a cui si aggiungono parole con frequenza $N \leq 5$, ma sempre con $p < 0,05$, quali *gommone, decreto, Africa, Tripoli, Libia, Turchia, affondare, Sea_Watch* (Appendice, Tab. 15).

Come la classe 3, e a differenza delle classi 1 e 2, più vicine fra loro nella relazione gerarchica discendente (Grafico 19), la classe 4 si caratterizza per una narrazione più fenomenologica che evenemenziale, raggruppando diversi segmenti di testo che si focalizzano su temi piuttosto che su eventi o azioni specifiche. Le parole che la contraddistinguono sono soprattutto nomi e aggettivi rappresentativi di discorsi tematici, piuttosto che di eventi delimitati entro i confini spazio-temporali tipici della cronaca. Fra i nomi di tempo attestati, ci sono, infatti *oggi*, che rinvia a un tempo presente e non al passato (né quello prossimo, della cronaca, né quello remoto, della storia), *stagione*, che indica un lasso di tempo che ritorna ciclicamente e *tempo*, un nome il cui riferimento semantico non è né un momento, né una durata misurabile, bensì un concetto più generale che può essere brevemente definito, con il vocabolario Treccani, come «l'intuizione e la rappresentazione della modalità secondo la quale i singoli eventi si susseguono e sono in rapporto l'uno con l'altro (per cui essi avvengono prima, dopo, o durante altri eventi)». Quanto ai nomi di luogo, si tratta perlopiù di nomi propri di paesi o aree geografiche estere o internazionali — *Europa, Davos, Stati Uniti* (Tabella 27), *Africa, Cina, Tripoli, Libia, Turchia, Islanda; Mediterraneo, Francia* (Appendice, Tab. 15) — che collocano i diversi discorsi coesistenti in questa classe entro una

dimensione transnazionale, rappresentata anche dalle occorrenze significative dell'aggettivo *internazionale* e del nome *mondo*.

In sintesi, questa classe sembra rappresentare lo spazio dell'agenda tematica delle fonti dei media su una serie di priorità – questioni da affrontare nel tempo presente e/o che si ripresentano ciclicamente – che riguardano l'Italia, vista da una più ampia prospettiva globale e di relazioni internazionali, e che riguardano da vicino la condizione delle donne e degli uomini¹¹¹, intesi come gruppi sociali, come conferma anche una lettura dei STT (cfr. Elenco 16).

Tabella 27. Forme lessicali con freq. N>5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 4 del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	paese	25	32	38,18	<0,0001
1	donna	83	173	36,13	<0,0001
2	parità	13	13	31,82	<0,0001
3	diritto	14	16	26,63	<0,0001
4	salariale	8	8	19,47	<0,0001
5	Europa	9	10	17,98	<0,0001
6	governo	10	12	17,13	<0,0001
7	europeo	10	12	17,13	<0,0001
8	reazione	7	7	17,02	<0,0001
9	livello	7	7	17,02	<0,0001
10	economia	7	7	17,02	<0,0001
11	Papa	7	7	17,02	<0,0001
12	trafficante	7	7	17,02	<0,0001
13	stipendio	7	7	17,02	<0,0001
14	settore	7	7	17,02	<0,0001
15	Italia	30	56	16,98	<0,0001
16	pensare	11	14	16,66	<0,0001
17	minaccia	8	9	15,58	<0,0001
18	azienda	8	9	15,58	<0,0001
19	stagione	6	6	14,57	0,00013
20	popolo	6	6	14,57	0,00013
21	passione	6	6	14,57	0,00013
22	stadio	6	6	14,57	0,00013
23	divario	6	6	14,57	0,00013
24	Davos	6	6	14,57	0,00013
25	guerra	7	8	13,19	0,00028
26	internazionale	8	10	12,54	0,00039
27	euro	8	10	12,54	0,00039
28	forza	12	18	12,37	0,00043

¹¹¹ La forma prevalente del lemma *donna* in questa classe è quella plurale (N=64/83); per il lemma *uomo* IRaMuTeQ non fornisce la statistica per classi lessicali, perché trattandosi di un *sub-corpus* ricavato da un *corpus* più grande (*corpus* TG) selezionando i ST contenenti *uomo*, questo lemma viene escluso dalla CGD del *sub-corpus*.

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
29	sicurezza	9	12	12,25	0,00046
30	mondo	20	37	11,41	0,00073
31	Stati_Uniti	6	7	10,83	0,00099
32	notizia	6	7	10,83	0,00099
33	regione	6	7	10,83	0,00099
34	furto	6	7	10,83	0,00099
35	pagare	7	9	10,31	0,00132
36	annunciare	7	9	10,31	0,00132
37	accordo	7	9	10,31	0,00132
38	candidato	7	9	10,31	0,00132
39	mare	10	15	10,28	0,00134
40	Francesco	8	11	10,13	0,00145
41	presidente	16	29	9,68	0,00186
42	città	14	25	8,84	0,00294
43	Trump	10	16	8,66	0,00325
44	futuro	9	14	8,4	0,00375
45	bastare	6	8	8,13	0,00434
46	tutte_le	6	8	8,13	0,00434
47	voce	6	8	8,13	0,00434
48	lavoratore	6	8	8,13	0,00434
49	bambino	16	31	7,71	0,00549
50	arrivare	18	37	6,97	0,00830
51	vedere	16	32	6,86	0,00882
52	ordine	8	13	6,61	0,01012
53	migrare	8	13	6,61	0,01012
54	oggi	21	46	6,25	0,01240
55	definire	6	9	6,12	0,01334
56	parola	6	9	6,12	0,01334
57	tempo	12	23	5,96	0,01466
58	spiegare	7	12	4,94	0,02617
59	grande	19	43	4,83	0,02804
60	impegnare	6	10	4,6	0,03203
61	spingere	6	10	4,6	0,03203
62	nuovo	14	30	4,51	0,03363

Elenco 16. Selezione di segmenti di testo tipici della classe 4 del *sub-corpus* uomo

STT 1 **** *A_2020 *G_30-01 *T_TG2 *N_Servizio *P_Giornalista *S_Uomo
score: 191,97

una fase 2 nel segno delle certezze quelle da dare al **paese** con una maggioranza unita e un **governo** solido **Zingaretti** insiste sulle modifiche ai **decreti sicurezza** e rilancia i temi **chiave** dell **azienda** di **parità salariale** uomo **donna** scuola

STT 5 **** *A_2018 *G_02-01 *T_TG2 *N_Lancio *P_Giornalista *S_Donna
score: 154,66

ed ora ci **spostiamo** in **Islanda** dove finalmente una **legge** sancisce la **parità salariale** tra uomo e **donna** ed è la prima nazione al **mondo** che a vincere le diseguaglianze di **genere** mentre in **Europa** e anche negli **Stati Uniti** il cammino su

STT 11 **** *A_2020 *G_29-01 *T_TG1 *N_Messaggio *P_Fonte *S_Uomo
score: 139,69

penso alla **parità** di **salario** tra uomo **donna** come **grande** scelta di campo di **lotta** alle disuguaglianze quindi **sviluppo** crescita **lavoro** e giustizia sociale ed è possibile se da **subito** si apre una **nuova stagione** di concretezza

STT 14 **** *A_2020 *G_22-01 *T_TG1 *N_Lancio *P_Giornalista *S_Donna
score: 126,38

Papa Francesco sui migranti basta ostilità uomini e **donne** affrontano viaggi **rischiosi** per **sfuggire** a **guerra** e violenza e ancora **oggi** accusa sono **sfruttati** e trattati come una **minaccia** Ignazio Ingraio

STT 69 **** *A_2019 *G_07-01 *T_TG2 *N_Intervista *P_Fonte *S_Uomo
score: 74,20

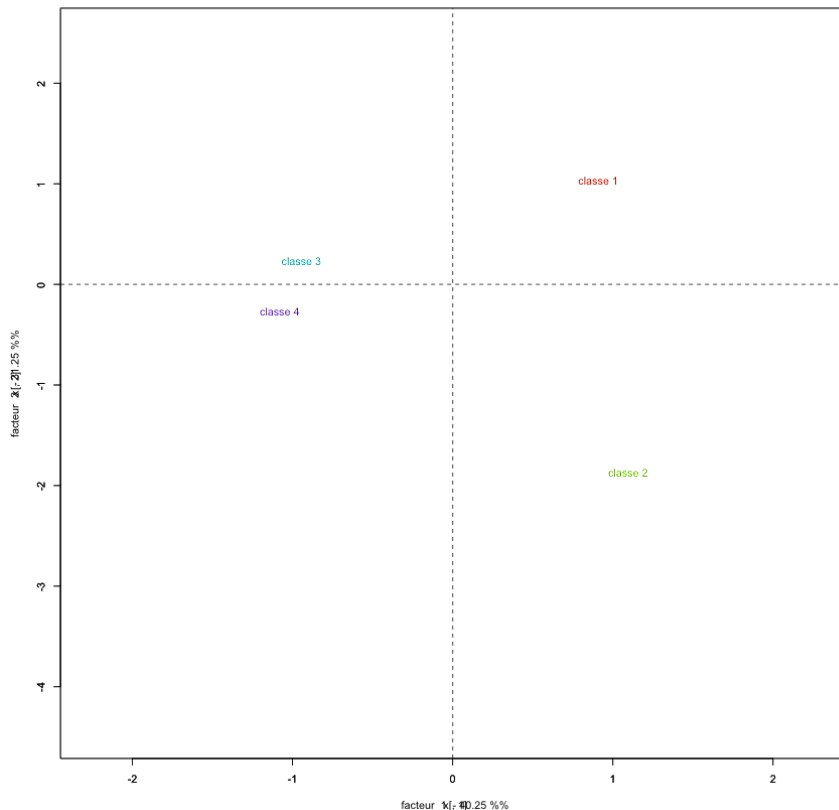
1 15 150 o 1500 fino a che non si fa capire ai **trafficcanti** di esseri **umani** che **l'Italia** non è punto di approdo per questi uomini e queste **donne** usate come merci per poi comprarci armi e droga gli

STT 129 **** *A_2019 *G_19-01 *T_TG1 *N_Intervista *P_Fonte *S_Uomo
score: 52,79

un uomo come me perché partecipa alla marcia perché **penso** che il femminismo non riguardi solo le **donne**

La proiezione dei risultati della CGD su piano fattoriale, illustrata nel grafico seguente, posiziona le quattro classi del *sub-corpus* uomo rispetto all'asse delle ascisse, rappresentato dal fattore 1, che spiega il 40,25% della varianza, e all'asse delle ordinate, rappresentato dal fattore 2, che spiega il 31,25%.

Grafico 20. Rappresentazione fattoriale delle 4 classi lessicali del *sub-corpus* uomo



Tenendo in considerazione le caratteristiche distintive della quattro classi lessicali, l'asse delle ascisse può essere interpretato come la linea su cui le classi di distribuiscono lungo il *continuum* della "tematizzazione" (sulla sinistra) e della "cronaca" (sulla destra). A sinistra, verso la polarità della "tematizzazione" si trovano infatti le classi 3 e 4 contraddistinte rispettivamente per essere, la prima, un luogo simbolico della memoria, detto altrimenti uno spazio narrativo lontano dal racconto cronachistico e vicino alla narrazione storica (vs. la narrazione cronachistica) e, la seconda, uno spazio di tematizzazione sulle questioni di genere portato dentro il discorso mediatico dalle fonti. A destra, invece, ci sono la classe 1 e 2 che si contraddistinguono, rispettivamente, per essere l'una l'area semantica della cronaca su fatti di criminalità e violenza, prevalentemente in ambito domestico, e l'altra lo spazio delle cronaca incidenti di varia natura.

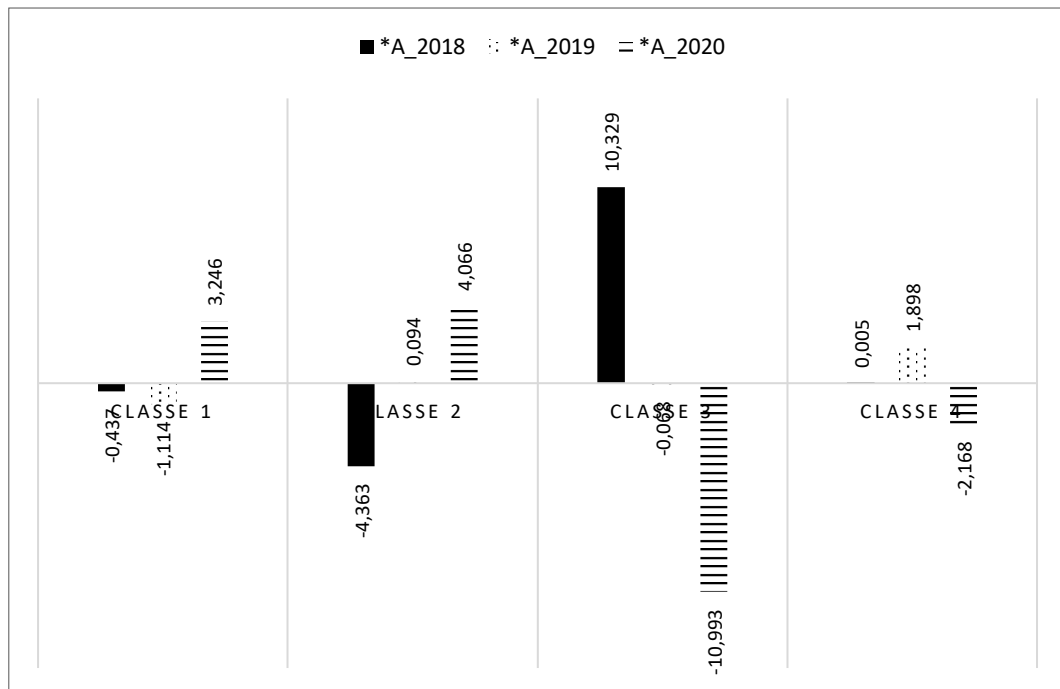
Il primo fattore, che spiega la distribuzione delle classi e del loro vocabolario sull'asse verticale, riguarda strettamente il *framing* mediatico e rende conto del formato dei servizi (tematizzazione/approfondimento vs. cronaca), mentre il secondo fattore, che spiega la distribuzione delle classi e del loro vocabolario lungo l'asse orizzontale, riguarda più da vicino i contenuti e rende conto delle rappresentazioni di genere lungo il *continuum* della soggettività-oggettività, secondo una dimensione già individuata nell'interpretazione della mappa fattoriale della CGD del *sub-corpus* donna. Sopra l'asse delle ascisse si collocano infatti le classi 1 (sulla destra) e la classe 3 (sulla sinistra), che rappresentano rispettivamente il mondo della criminalità e della violenza agita dagli uomini e il mondo della memoria collettiva rappresentata da soggettività maschili. Sotto l'asse delle ascisse si posizionano invece le classi 2 (sulla destra) e 4 (sulla sinistra), caratterizzate, la prima, dalla cronaca di "incidenti di varia natura" di cui gli uomini sono vittime o sopravvissuti, dunque oggetto di eventi ricondotti ad altre soggettività, che possono essere di volta in volta una *raffica di vento* (es. STT 3, Tabella 39) o un *branco di cinghiali* (es. STT 10, Tabella 39); la seconda dalla messa a tema, come oggetto del discorso, della parità di genere e dei diritti umani.

5.3.4 Comparazione per anno e per TG

La seconda domanda di ricerca formulata per l'esplorazione del *corpus* TG ("Quali sono le rappresentazioni di genere di donne e uomini, intese/i come oggetto, ovvero argomenti dell'informazione nel *corpus* TG?") include due quesiti specifici sulle somiglianze o differenze per anno e TG. Come nell'analisi sulle rappresentazioni di genere di donne e uomini fonti di notizia (§ 5.3.2), per rispondere a questa domanda è stato utilizzato l'output di IRaMuTeQ che fornisce i valori positivi e negativi che compongono il chi-quadrato di associazione fra le diverse modalità di una variabile e le n. classi lessicali risultate dalla CGD del *corpus* oggetto di analisi.

Il Grafico 21 rappresenta il chi-quadrato di associazione delle tre modalità della variabile *Anno per ciascuna delle quattro classi lessicali del *sub-corpus* donna.

Grafico 21. Chi-quadrato di associazione delle 4 classi lessicali del *sub-corpus* donna con le modalità della variabile *Anno



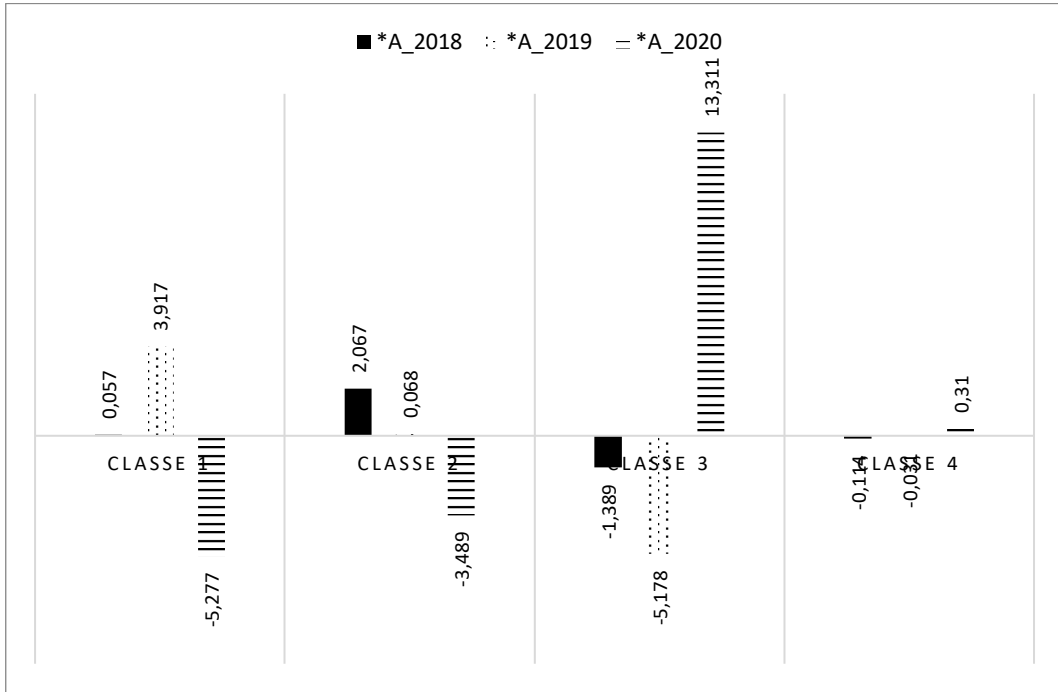
Considerato che il valore soglia per la significatività del chi-quadrato è $\chi^2 > 3,84$ ($p < 0,05$) la classe 1 ("violenza domestica contro le donne") e la classe 4 ("occupazione femminile") non risultano significativamente correlate a nessuna

delle tre modalità della variabile *Anno. Significative sono invece le associazioni fra la modalità *A_2018 e le classi 2 (“centralità delle donne”) e 3 (“dissenso delle donne”) – nel primo caso in senso negativo ($\chi^2=-4,363$), nel secondo caso in senso positivo ($\chi^2=10,329$) – e fra la modalità *A_2020 e le classi 2 e 3 – nel primo caso in senso positivo ($\chi^2=4,066$) e nel secondo caso in senso negativo ($\chi^2=-10,993$). Questo significa che i segmenti di testo marcati *A_2019 incidono in modo equo e trasversale alla formazione di tutte le classi, mentre quelli marcati *A_2018 incidono sensibilmente sulla formazione della classe 3 e alla sua distinzione dalla classe 2, e quelli marcati *A_2020 influiscono positivamente sulla formazione della classe 2 e alla sua distinzione dalla classe 3.

Il Grafico 22 rappresenta il chi-quadrato di associazione delle tre modalità della variabile *Anno per ciascuna delle quattro classi lessicali del *sub-corpus* uomo, evidenziando, in questo caso, nessuna correlazione significativa fra le tre modalità della variabile *Anno e le classi 2 (“incidenti di varia natura”) e 4 (“parità di genere e diritti umani”), e due associazioni significative: una fra la modalità *A_2019 e le classi 1 (“criminalità e violenza contro le donne”) e 3 (“memoria collettiva”) – rispettivamente in senso positivo ($\chi^2=3,917$) e in senso negativo ($\chi^2=-5,178$) – una fra la modalità *A_2020 e le classi 1 e 3 – nel primo caso in senso negativo ($\chi^2=-5,277$), nel secondo caso in senso positivo ($\chi^2=13,311$).

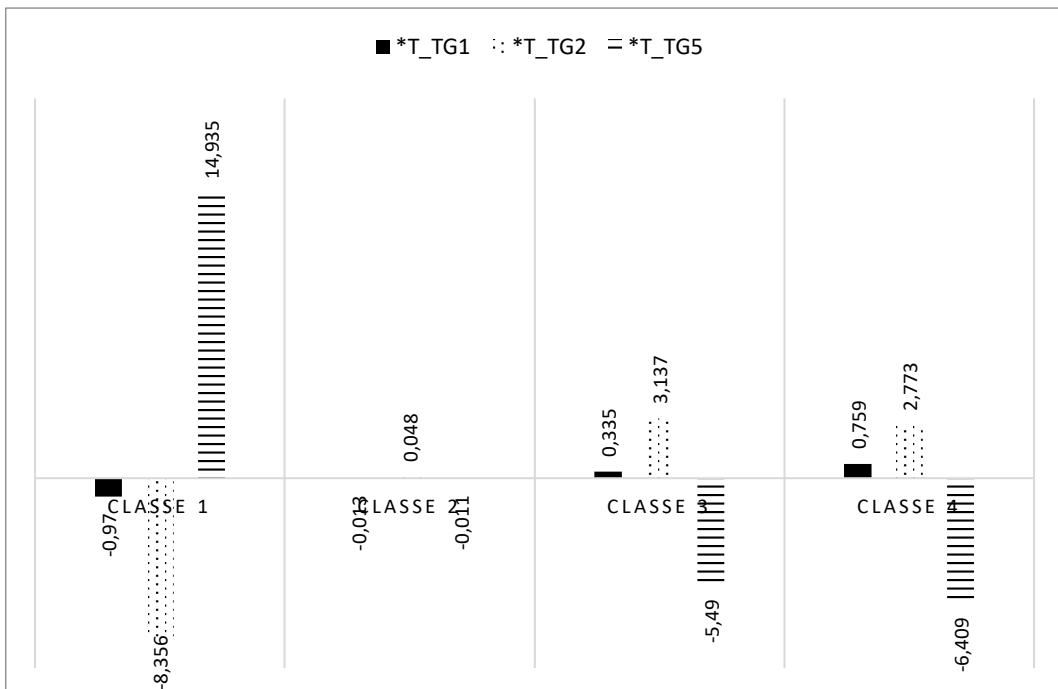
In sintesi, i risultati dimostrano che i segmenti di testo marcati *A_2018 incidono trasversalmente sulla formazione di tutte e quattro le classi lessicali, senza dare un contributo significativo alla composizione dell’una o dell’altra; mentre i segmenti di testo marcati *A_2019 incidono sensibilmente sulla formazione della classe 1 e quelli marcati *A_2020 sulla composizione della classe 3.

Grafico 22. Chi-quadrato di associazione delle 4 classi lessicali del *sub-corpus* uomo con le modalità della variabile *Anno



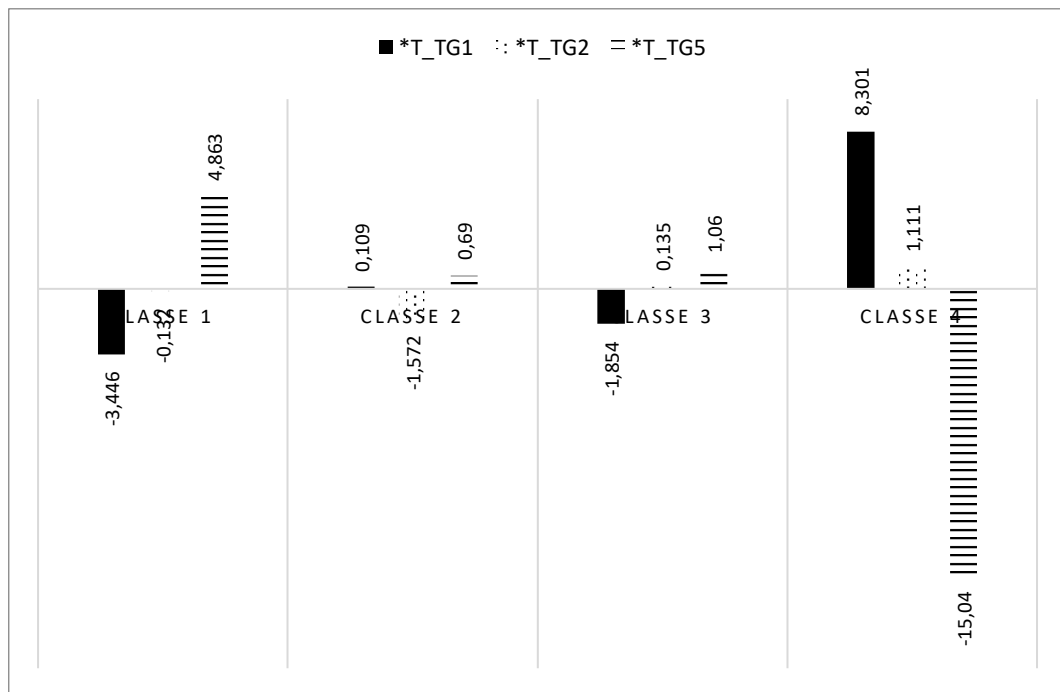
Il Grafico 23 e il Grafico 24 mostrano i valori del chi-quadrato di associazione delle modalità della variabile *TG con le classi lessicali del *sub-corpus* donna e del *sub-corpus* uomo.

Grafico 23. Chi-quadrato di associazione delle 4 classi lessicali del *sub-corpus* donna con le modalità della variabile *TG



Il risultato più evidente riportato dal grafico del sub-corpus donna è la mancanza di correlazioni significative fra la modalità *T_TG1 e tutte e quattro le classi lessicali. Le modalità *T_TG2 risulta invece significativamente correlata in negativo con la classe 1 (“violenza domestica contro le donne”). La modalità *T_TG5 attesta un’associazione significativamente positiva con la classe 1 e sensibilmente negativa con le classi 3 (“dissenso delle donne”) e 4 (“occupazione femminile”). In sintesi, i segmenti di testo del TG1 non influenzano la composizione delle classi; i testi del TG5 incidono positivamente sulla composizione della classe 1 e negativamente sulla formazione delle classi 3 e 4; infine, i segmenti del TG2 indicano negativamente sulla classe 1 (e positivamente, anche se in modo non significativo sulle classi 3 e 4).

Grafico 24. Chi-quadrato di associazione delle 4 classi lessicali del *sub-corpus* uomo per TG



Per quanto riguarda il *sub-corpus* uomo, la modalità *T_TG1 è positivamente e significativamente correlata alla classe 2 (“incidenti di varia natura”), la modalità *T_TG2 non è correlata significativamente a nessuna classe lessicale e la modalità *T_TG5 è correlata positivamente alla classe 1 (“criminalità e violenza contro le donne”) e negativamente alla classe 4 (“parità di genere e diritti umani”). Questo significa che i segmenti di testo del TG2 non incidono

significativamente sulla formazione delle classi lessicali, mentre quelli del TG1 influenzano la formazione della classe 2 e quelli del TG5 la composizione della classe 1 e, in negativo, la composizione della classe 4.

5.4 Discussione dei risultati

5.4.1 Donne e uomini come fonte dell'informazione

La rappresentazione di genere delle donne e degli uomini come fonti/soggetto dell'informazione è stata analizzata attraverso: 1) un'analisi lessicografica del *sub-corpus* P_Fonte che, includendo le variabili *S_Donna e *S_Uomo, ha evidenziato alcune differenze e somiglianze generali fra il lessico delle fonti femminili e quello delle fonti maschili; 2) un'analisi lessicografica dei due *sub-corpora* PF_Donna e PF_Uomo, ricavati dal P_Fonte per distinzione dei segmenti di testo con fonti rispettivamente di sesso femminile e maschile, che ha approfondito somiglianze e differenze emerse al primo livello di analisi; 3) una CGD del *sub-corpus* P_Fonte che ha individuato, sul piano testuale, le aree lessicali del *corpus* e le loro correlazioni con le variabili categoriali inserite nel *corpus*; 4) due CGD separate per i due *sub-corpora* PF_Donna e PF_Uomo che ha approfondito le prime correlazioni emerse dal *corpus* P_Fonte con le due modalità della variabile Sesso, *S_Donna e *S_Uomo, consentendo così di interpretare le classi lessicali individuate dalle CGD dei due *corpora* in un'ottica comparativa, anche attraverso l'analisi fattoriale; 5) un'analisi comparativa per anno e per TG, utile a verificare differenze diacroniche e editoriali, che ha evidenziato l'impatto dell'agenda dell'informazione e delle linee editoriali sulle diverse rappresentazioni di genere emerse. I risultati di queste analisi sono di seguito interpretati in un'ottica comparativa e integrata, in modo da tener conto delle principali evidenze emerse in modo dialettico e complessivo.

L'analisi delle somiglianze del *sub-corpus* P_Fonte, basata sulla correlazione fra le parole (lemmatizzate) e le due modalità della variabile *S_Donna/S_*Uomo, restituisce principalmente due risultati rilevanti: 1) *donna* è fra i lemmi più frequenti (freq. N>100) del *corpus* ed è significativamente associato ai segmenti

con fonti femminili, viceversa, *uomo*, anch'esso fra i lemmi più frequenti nel *corpus*, non è significativamente associato né a fonti femminili né a fonti maschili; 2) le parole più frequenti e significativamente associate alla modalità *S_Donna fanno riferimento alla sfera della maternità (*figlio*), a quella dell'educazione (*scuola*) e a quella domestica (*casa*), mentre le parole correlate alla modalità *S_Uomo fanno riferimento in generale alla sfera pubblica (*mondo, Europa, europeo, italiano*) e in particolare alla politica (*politico, governo*).

Il primo risultato può essere interpretato come l'attestazione di una presenza e rilevanza, in termini di significatività, di un discorso autoreferenziale per le fonti femminili, che non ha corrispondenza per le fonti maschili. Ciò non vuol dire che gli uomini non parlino di sé, ma che l'uso della parola *uomo* (e *uomini*) non è distintivo del loro parlato. L'analisi lessicografica del *sub-corpus* PF_Uomo attesta infatti N=94 occorrenze del lemma *uomo*, che è un valore superiore a quello attestato da *donna* nel *sub-corpus* PF_Donna (N=56), ma *uomo* si posiziona al rango 71 del *sub-corpus* PF_Uomo mentre *donna* al rango 26 del *sub-corpus* PF_Donna, anche per effetto della diversa dimensione dei due *sub-corpora*: come già scritto, quello maschile è più grande di quello femminile di circa tre volte (si veda la Tabella 5). Ciò non sminuisce tuttavia la rilevanza dell'asimmetria attestata dall'analisi delle somiglianze nel *sub-corpus* P_Fonte, che può essere interpretata come la testimonianza di una propensione delle donne a marcare la propria identità di genere (la parola *donna* usata come *marker* identitario) e/o a parlare di altre donne o delle donne in generale come gruppo sociale (la parola *donna* usata come *topic-marker*), come conferma l'analisi delle rappresentazioni di donne e uomini come oggetto/argomento dell'informazione. In entrambi i casi, questo risultato dimostra la presenza nel discorso mediatico di un'esplicita autoconsapevolezza dell'identità di genere femminile, a cui non corrisponde una altrettanto esplicita autoconsapevolezza maschile.

Il secondo risultato attesta una rappresentazione generalmente stereotipata rispetto al genere: le parole distintive del lessico femminile rinviano infatti a domini tradizionalmente appannaggio delle donne (la maternità,

l'educazione, la vita domestica) mentre quelle distintive del lessico maschile rinviano alla sfera pubblica, e in particolare alla politica, storicamente una prerogativa di pensiero e di azione maschile.

Lo scorporo del *sub-corpus* PF_Fonte nei due *sub-corpora* PF_Donna e PF_Uomo ha prodotto, come sopra accennato, due *corpora* dalla dimensione molto diversa, attestando un minor spazio di parola delle fonti femminili rispetto a quelli maschili. Questa evidenza non può essere correlata a caratteristiche "generaliste" del parlato femminile, peraltro secondo un diffuso luogo comune più prolifico di quello maschile (si veda più sopra, § 0), ma va riportata ai risultati delle ricerche sulla presenza delle donne nei mezzi di informazione, italiani e non, che ne attestano una costante sotto-rappresentazione, sia rispetto agli uomini, sia rispetto all'incidenza femminile sulla popolazione complessiva (cfr. Capitolo 3).

L'analisi comparativa delle somiglianze nei due *sub-corpora* PF_Donna e PF_Uomo attesta sia elementi di similarità, sia elementi di differenza. Elementi di somiglianza sono evidenti nei rispettivi vocabolari, con la maggior parte dei lemmi più frequenti condivisi (N=25, con soglia al rango 30) e nella dimensione di senso indicata dalle coricorrenze lessicali. Il Grafico 4 e il Grafico 5 mostrano infatti una similarità in tre raggruppamenti di parole che attestano la presenza di una narrazione del vissuto personale, la presenza di contenuti politici e la presenza di un tema lavoro.

Elementi di distinzione emergono da un'analisi più attenta alla semantica. Fra i lemmi distintivi del *sub-corpus* PF_Donna si trova la parola *donna* (N=30 occorrenze al plurale e N=26 al singolare), mentre la parola *uomo* non è fra i lemmi più frequenti del *sub-corpus* PF_Uomo¹¹², risultato che conferma quanto emerso nell'analisi delle somiglianze del *sub-corpus* P_Fonte, indicando che le fonti femminili sono più autoreferenziali rispetto a quelle maschili. Fra i lemmi distintivi del *sub-corpus* PF_Donna ci sono poi tre parole che evocano la sfera privata e domestica: *figlio*, *casa*, *famiglia*, mentre nel *sub-corpus* PF_Uomo non ci sono

¹¹² Il lemma *uomo* nel *sub-corpus* PF_Uomo ha una frequenza N=94 (66 singolare, 22 plurale), con rango 71.

parole che rinviano alla medesima sfera; ci sono, invece, due parole che fanno riferimento alla sfera pubblica e politica: *cittadino* e *Europa*, assenti dall'elenco dei lemmi con rango pari o inferiore a 30 nel *sub-corpus* S_Donna. C'è poi la parola *paura*, al rango 29 del *sub-corpus* PF_Donna, assente dal *sub-corpus* PF_Uomo, dove non compaiono parole associate alla dimensione emotiva, una dimensione che già a un primo livello di analisi del *corpus* P_Fonte emerge come significativamente associata al genere femminile (S_Donna). Infine, trascurando il verbo *dire*, che è il più frequente in entrambi i *sub-corpora*, il secondo verbo più frequente nel *sub-corpus* PF_Donna è *volere*, un verbo che, come precisa il vocabolario Treccani «assume varie sfumature di significato e soprattutto di tono sia dal contenuto della frase, sia, nel parlare, dal modo di pronunciarla»¹¹³ ma il cui significato principale è «tendere con decisione, o anche soltanto con il desiderio, a fare o conseguire qualche cosa, o essere risoluto (e comandare e disporre) che altri la faccia»¹¹⁴, è un verbo dunque che indica una tensione verso un obiettivo, al contrario di *stare* che può avere diversi significati – come «rimanere immobile, fermo [...] restare [...] essere, trovarsi, permanere in un dato luogo, o in una determinata condizione»¹¹⁵ – ma tutti ascrivibili all'area semantica della permanenza, della staticità, della collocazione.

La CGD del *sub-corpus* P_Fonte individua, come abbiamo visto nel paragrafo precedente, tre classi lessicali che attestano associazioni significative diverse con il genere della fonte. La classe 1, denominata “testimonianze di vita”, è quella di maggiore dimensione — copre quasi la metà dei segmenti di testo classificati (49,42%) — è significativamente associata alle fonti femminili (con un valore chi-quadrato elevato, $\chi^2=160,03$) e, sul piano fattoriale, si colloca nel 1° quadrante delimitato dalle polarità “vicino” (in alto, sull'asse verticale) e “cronaca” (a sinistra, sull'asse orizzontale). La classe 2, del “discorso scientifico”, è quella di minore dimensione (13,3%), è significativamente associata alle fonti maschili (con un valore chi-quadrato significativo ma non elevato, $\chi^2=12,19$) e, sul piano fattoriale, si

¹¹³ Fonte: <https://www.treccani.it/vocabolario/volere2/>; ultimo accesso 19 novembre 2021.

¹¹⁴ *Ibidem*.

¹¹⁵ Fonte: <https://www.treccani.it/vocabolario/stare/>; ultimo accesso 19 novembre 2021.

colloca in opposizione simmetrica con la classe 1 sul 4° quadrante, definito dalle polarità “lontano” (in basso, sull’asse delle ordinate) e “politica” (a destra, sull’asse delle ascisse). La classe 3, infine, che coincide con l’area semantica della “comunicazione politica”, ha una dimensione estesa (37,35%), è significativamente associata al genere maschile (con un valore chi-quadrato elevato: $\chi^2=119,94$) e, nella mappa fattoriale, si colloca principalmente nel 2° quadrante, delimitato dalle polarità “vicino” (in alto, sull’asse delle y) e “politica” (a destra, sull’asse delle x), con alcune parole sovrapposte all’asse delle ascisse e anche sotto l’asse delle ascisse, verso la polarità “lontano”.

Questi risultati indicano la prevalenza di due mondi lessicali, particolarmente ampi e coesi al loro interno, e polarizzati rispetto al genere delle fonti: quello delle “testimonianze di vita”, ampio e circoscritto alla cronaca di eventi di prossimità temporale e spaziale (“vicino”), che si contraddistingue per essere “specialmente” femminile, e quello della “comunicazione politica”, che ha una dimensione inferiore, in termini di segmenti di testo classificati, ma un’estensione maggiore in termini di ampiezza semantica sovrapponendosi all’asse delle ascisse, e che si distingue per essere “specialmente” maschile. Questa polarizzazione assiologica delle fonti femminili vs. quelle maschili non solo ricalca la tradizionale collocazione di genere delle donne nella sfera privata, alla quale si può ricondurre una qualsiasi testimonianza di vita, ancorché resa pubblica dall’esposizione mediale, e degli uomini nella sfera pubblica, qual è per definizione la politica, ma appare in linea con tutti i risultati sui ruoli delle fonti nei contenuti dei mezzi d’informazione. Se prendiamo, a esempio, i risultati italiani dell’ultima edizione del GMMP (WACC 2020), che la più recente ricerca sulla rappresentanza e rappresentazione di genere nei mezzi di informazione, questi dimostrano che, nonostante la visibilità femminile nelle notizie sia aumentata dal 24% al 26%, le politiche, attestando una presenza del 22%, continuano a essere sotto-rappresentate sia rispetto agli uomini, che rappresentano il restante 78%, sia

rispetto alla loro rappresentanza reale che, a titolo indicativo può essere attestata dalla percentuale di donne nel Parlamento italiano¹¹⁶, pari al 35%.

Questa congruenza di risultati, fra quanto emerge dall'analisi del *sub-corpus* P_Fonte e i dati del GMMP, nonché da precedenti ricerche sulle donne nei contenuti nell'informazione, induce a interpretare i primi come indicatori non tanto, o almeno non solo, di una diversità di linguaggio fra donne e uomini, quanto di una loro diversa collocazione dentro il discorso dei media: le prime come soggetto di una narrazione privata, i secondi come soggetto di un discorso pubblico. In entrambi i casi non si tratta di una esclusività, ma di una prevalenza, come confermano i risultati delle analisi dei *sub-corpora* PF_Donna e PF_Uomo, in particolare la CGD, interpretati in un'ottica comparativa.

La CGD del *sub-corpus* PF_Donna ha restituito due classi lessicali di cui la prima ha caratteristiche del tutto simili alla classe 1 del *sub-corpus* P_Fonte, e anche alla classe 1 del *sub-corpus* PF_Uomo, e si configura come l'area lessicale delle "testimonianze di vita", caratterizzata da una narrazione basata sull'esperienza personale, con una forte connotazione emotiva e sentimentale. La seconda classe lessicale emersa dalla CGD del *sub-corpus* PF_Donna ha caratteristiche affini a quelle della classe 3 del *sub-corpus* P_Fonte, e alla classe 3 del PF_Uomo, corrispondenti alla "comunicazione politica". Tuttavia, la classe 1 nel *sub-corpus* PF_Donna ha una grande dimensione, coprendo l'80,25% dei segmenti di testo classificati vs. il 41,92% dei segmenti di testo classificati in quest'area lessicale nel *sub-corpus* PF_Uomo. L'area semantica della "comunicazione politica", invece, nel *sub-corpus* PF_Donna copre soltanto il 19,75% dei segmenti di testo vs. il 39,93% nel *sub-corpus* PF_Uomo. Infine, la classe del "discorso scientifico", che copre il 18,11% del *sub-corpus* PF_Uomo, risulta assente dal *sub-corpus* PF_Donna.

Donne e uomini sono dunque soggetti non esclusivi, ma prevalenti, le une di un discorso di testimonianza, che ha a che fare con l'esperienza personale e la

¹¹⁶ Il riferimento al Parlamento italiano è puramente indicativo perché le politiche, e i politici, visibili nell'informazione italiana non sono necessariamente rappresentanti del Parlamento italiano, possono essere politiche, e politici, locali, o di paesi esteri.

vita privata, e gli altri di un discorso politico, che ha a che fare con questioni pubbliche. Mentre l'esclusività maschile riguarda il "discorso scientifico". Un risultato quest'ultimo che può essere interpretato sia sulla base della storica marginalizzazione delle donne, come studiose e professioniste, dai diversi ambiti del sapere scientifico, sia sulla base di evidenze empiriche che attestano la scarsa visibilità mediatica delle scienziate. Per esempio, secondo una recente ricerca del King's College of London (Jones 2020) che ha analizzato 146.867 articoli sul coronavirus pubblicati fra il 1° marzo e il 31 luglio 2020, in quindici fonti dell'informazione di Regno Unito, Australia e Stati Uniti, solo il 5% fra gli esperti di STEM (Science, Technology, Mathematics and Engineering) citati da questi articoli sono donne.

5.4.2 *L'impatto dell'agenda e della linea editoriale*

Il *corpus* TG, da cui sono ricavati i *sub-corpora* P_Fonte, PF_Donna e PF_Uomo, copre tre anni rappresentati ciascuno da un mese (gennaio), per questo i risultati emersi dalla comparazione per anno non possono essere interpretati come segnali esaustivi di elementi di stabilità o instabilità. Tuttavia, il confronto incrociato per anno e per classi lessicali delle CGD dei tre diversi *sub-corpora* consente di avanzare qualche ipotesi interpretativa sull'impatto dell'agenda dei tre mesi considerati rispetto alla rappresentazione di genere di donne e uomini come soggetti dell'informazione. Per questo, è importante tenere presente le corrispondenze (e non-corrispondenze) fra le classi lessicali nei tre *sub-corpora*. La classe delle "testimonianze di vita" è attestata in tutti i tre i *sub-corpora*, come classe 1. La classe della "comunicazione politica" risulta anch'essa in tutti i tre i *sub-corpora*, come classe 2 nel *sub-corpus* PF_Donna e come classe 3 nei *sub-corpora* PF_Uomo e P_Fonte. La classe del "discorso scientifico" è attestata soltanto per questi ultimi due, come classe 2.

Il primo risultato emergente dalla comparazione per anno evidenzia una correlazione positiva fra la modalità *A_2018 e tutte le classi 1 dei tre *corpora*, ma significativa soltanto per i *corpora* P_Fonte e PF_Donna. Questo significa che i

segmenti di testo datati 2018 contribuiscono alla formazione di questa classe nel *corpus* P_Fonte, soprattutto per l'apporto delle fonti femminili. Detto altrimenti l'anno 2018 è caratterizzato da un mondo lessicale che rappresenta soggetti narranti un vissuto personale, fra i quali si distinguono le soggettività femminili. Il secondo risultato rilevante riguarda la formazione della classe della "comunicazione politica" che risulta significativamente distinta dalla classe 1 per il mancato apporto di segmenti di testo dell'anno 2018, in tutti e tre i *sub-corpora*. Questo sembra indicare che il 2018 sia caratterizzato da un'agenda prevalentemente cronachistica (come abbiamo visto più sopra e come conferma anche l'analisi sulle rappresentazioni di genere di donne e uomini oggetto dell'informazione), che abbassa la soglia d'ingresso delle fonti testimoni di esperienza alzando al contempo la soglia d'ingresso delle fonti della "comunicazione politica", in modo trasversale per donne e uomini. Il terzo risultato riguarda sempre la formazione della classe della "comunicazione politica" sulla quale incidono sensibilmente, per il *sub-corpora* PF_Donna, i segmenti di testo del 2020, per il *sub-corpora* PF_Uomo quelli del 2019. Questo indica che non ci sono fattori stabili nel favorire soggettività femminili o maschili nella costruzione della "comunicazione politica", ma piuttosto fattori imputabili all'agenda, dunque variabili di anno in anno. Questa ipotesi è confermata anche dal risultato che riguarda la classe del "discorso scientifico", alla cui formazione concorrono significativamente segmenti di testo dell'anno 2020, che, come abbiamo visto, sono prevalentemente, anche se non esclusivamente, focalizzati sull'emergenza coronavirus.

La comparazione dei risultati per TG restituisce tre risultati rilevanti che evidenziano le specificità dei diversi telegiornali. Il primo riguarda il TG1: nel complesso i segmenti di testo di questo notiziario contribuiscono in modo significativo alla formazione della classe 1, ma questo per effetto dei segmenti di testo di fonti maschili, piuttosto che femminili. La modalità *T_TG1 è infatti significativamente e positivamente associata alla classe 1 dei *sub-corpora* P_Fonte e PF_Uomo ma non alla classe 1 del *sub-corpora* PF_Donna. Il secondo risultato

riguarda il TG2: nel complesso i segmenti di testo di questo notiziario contribuiscono in modo significativo alla formazione della classe della “comunicazione politica”, sia nel *sub-corpus* PF_Donna, sia nel *sub-corpus* PF_Uomo; al contempo non contribuiscono, o contribuiscono negativamente, alla formazione della classe 1, in modo significativo per il *sub-corpus* PF_Donna. Infine, il TG5 si distingue per contribuire in modo significativamente positivo alla formazione della classe della “comunicazione politica” nel *sub-corpus* PF_Uomo.

In sintesi i risultati sulla comparazione per TG evidenziano diverse linee editoriali, fra cui merita di essere notata quella del TG2 che testimonia una rappresentazione delle fonti femminili come soggetto della “comunicazione politica”, in una direzione in linea con il progresso delle pari opportunità e, al contempo, si distingue per non contribuire alla rappresentazione delle fonti femminili come soggetto di “testimonianze di vita”, una narrazione, che, per quanto importante, tende a ricollocare le donne entro lo spazio della sfera privata e dunque a rappresentarle in modo stereotipato.

5.4.3 *Donne e uomini come argomento dell'informazione*

La rappresentazione di genere delle donne e degli uomini come oggetto/argomenti dell'informazione è stata analizzata attraverso: 1) un'analisi lessicale dei *sub-corpora* donna e uomo che ha evidenziato similarità dimensionali e di vocabolario fra i due *sub-corpora*; 2) due CGD separate per i due *sub-corpora* che, attraverso anche l'analisi fattoriale incrociata con le variabili categoriali, hanno evidenziato alcune similarità e alcune differenze di rappresentazione delle donne e degli uomini; 3) un'analisi compartiva per anno e TG, che ha evidenziato rappresentazioni ricorrenti o non ricorrenti sul piano diacronico e rappresentazioni trasversali o non trasversali fra le diverse testate giornalistiche.

L'analisi dei *sub-corpora* donna e uomo ha restituito tre evidenze di similarità, a livello di dimensione, ricchezza di vocabolario e profilo lessicale. In termini dimensionali, le statistiche dei due *sub-corpora* registrano misure molto simili fra loro: sebbene il *sub-corpus* donna sia segmentato in un numero di testi

inferiore rispetto al *sub-corpus* uomo (N=926 vs. N=1.002) e registri un numero di occorrenze (N=41.163 vs. N=42.255), forme (N=5.948 vs. N=6.313) e hapax (N=3.111 vs. N=3.335) inferiore, i due *corpora* hanno lo stesso ordine di grandezza. Inoltre, i due *sub-corpora* hanno una ricchezza di vocabolario del tutto simile, come risulta dalla *ratio* forme/occorrenze, pari a 0,14 per il *sub-corpus* donna e a 0,15 per il *sub-corpus* uomo. Questo dimostra che i lemmi *donna* e *uomo* sono al centro di un discorso mediale equilibrato nell'assumere le donne e gli uomini come argomento del discorso, a differenza di quanto emerge nell'analisi delle rappresentazioni di genere di donne e uomini come soggetto/fonti dell'informazione. Anche se l'equilibrio di genere attestato riguarda segmenti di testo correlati a due lemmi che si caratterizzano per una identica trasparenza del genere grammaticale a livello morfosintattico ma per una diversa coerenza fra genere grammaticale e genere semantico. Come ho già scritto, *donna* si riferisce infatti a persone dal genere biologico e sociale femminile, mentre *uomo* viene usato anche con riferimento all'umanità in generale, con una sovra-estensione del significato anche alla componente femminile.

Per quanto riguarda i profili lessicali i due *sub-corpora*, analizzati comparativamente, dimostrano caratteristiche molto simili fra loro tanto da profilare un lessico dominante che colloca la rappresentazione di entrambi i generi come argomento/oggetto dell'informazione prevalentemente dentro la cronaca. In particolare la cronaca nera incentrata, da un lato, su fatti di criminalità e violenza e, dall'altro, su relazioni affettive e familiari. La prevalenza di un discorso focalizzato sulla cronaca di fatti di criminalità e violenza può essere ricondotta a una tendenza tipica dei notiziari italiani a dare ampio spazio alla cronaca. Secondo il X *Rapporto sulla sicurezza in Italia e in Europa* (Osservatorio Europeo sulla Sicurezza 2017¹¹⁷), che riporta i risultati di un'attività di monitoraggio sull'agenda dei notiziari italiani, a confronto con quelli di Francia, Germania, Inghilterra e Spagna, la criminalità rappresenta in modo costante la seconda o terza voce dell'agenda tematica dei TG

¹¹⁷ Questo è l'ultimo rapporto dell'Osservatorio Europeo sulla sicurezza che contiene dati sull'informazione dei TG italiani ed esteri.

italiani, che per alcuni anni registrano il triplo di notizie di questo tipo rispetto, per esempio, ai notiziari britannici. Inoltre, a seconda dei periodi, i TG italiani declinano questa tematica collegandola a qualche tema sociale, come per esempio l'immigrazione (2007-2008) o le relazioni affettive (2011-2012), ma approdano, nel secondo semestre del 2016, a una semplice «elencazione delle manifestazioni della “banalità del male”» (Osservatorio Europeo sulla Sicurezza 2017, 21), vale a dire diversi fatti di violenza presentati come tali senza una contestualizzazione di tipo politico o sociale. Forse perché, come spiega il sociologo Pierre Bourdieu, nel suo volume *Sulla televisione* (1997), l'informazione quotidiana ha sempre bisogno di fatti di “straordinaria-ordinarietà” per competere all'interno di un sistema mediale che non risponde solo a logiche di servizio informativo, ma anche di competizione sul mercato dell'Auditel. Secondo Bourdieu (1997), gli organi d'informazione quotidiana necessitano di presentare come straordinari eventi che, accadendo ogni giorno, sono di fatto ordinari, per alzare l'attenzione del pubblico e guadagnare dati di ascolto maggiori rispetto ai propri *competitor*. Questo si traduce, sul piano dei contenuti, nella ricerca di una “messa in scena” che renda straordinari eventi ordinari. Per questo la televisione ricorre alla spettacolarizzazione e alla drammatizzazione non solo attraverso le immagini ma anche e soprattutto attraverso le parole.

Con quelle comuni non si può sperare di “épater les bourgeois”, di far colpo sui ceti medi, e neppure sul “popolo”. Per questo occorrono parole straordinarie. In effetti, paradossalmente il mondo dell'immagine è dominato dalle parole. [...] Nominare, si sa, vale a far vedere, a creare, a portare all'esistenza. E le parole possono avere effetti catastrofici (Bourdieu 1997, 20).

La prevalenza di un lessico familiare può essere compresa in relazione all'attualità sociale, e mediatica, del tema della violenza domestica (Oddone 2015; Giomi e Magaraggia 2017; IZI et al. 2020; Buonanno e Faccioli 2020), che risulta infatti dominante nel *sub-corpus* donna e presente anche nel *sub-corpus* uomo, secondo quanto attestato dai risultati delle CGD e dell'analisi fattoriale.

La CGD del *sub-corpus* donna e la la CGD del *sub-corpus* uomo hanno restituito quattro classi lessicali, con struttura, dimensioni, profili e rappresentazioni fattoriali parzialmente differenti fra loro. L'interpretazione dei risultati tiene conto in maniera dialettica di questi diversi livelli di analisi. La struttura della CGD del *sub-corpus* donna mostra una gerarchia con al vertice la classe 1 ("la violenza contro le donne") seguita dalla classe 4 ("l'occupazione femminile") e poi dalle classi 3 ("il dissenso delle donne") e 2 ("la centralità delle donne") che, pur avendo dimensioni differenti, sono le uniche in relazione simmetrica fra loro. Le relazioni fra classi nella CGD del *sub-corpus* uomo non disegnano invece una simile struttura gerarchica, bensì una bipartizione del *corpus* in quattro classi poste in relazione duale e simmetrica fra loro, con la classe 1 ("criminalità e violenza contro le donne") e la classe 2 ("incidenti di varia natura"), da una parte, e la classe 3 ("memoria collettiva") e la classe 4 ("parità di genere e diritti umani"), dall'altra. Questa differenza strutturale delle due CGD è il primo risultato dell'analisi comparativa fra i due *sub-corpora* e indica una diversa rappresentazione di genere, a un primo e più generale livello. La rappresentazione dominante per le donne come oggetto dell'informazione rinvia a un "mondo" di violenza. Un mondo che delimita anche la rappresentazione maschile in modo rilevante, ma non dominante.

La classe 1 del *sub-corpus* donna si può infatti considerare parzialmente corrispondente alla classe 1 del *sub-corpus* uomo. Entrambe, oltre a coincidere con l'area lessicale della violenza, hanno dimensioni simili, coprendo il 33,09% delle UCE nel primo caso e il 32,47% nel secondo caso. Tuttavia non si può trascurare che la classe 1 del *sub-corpus* uomo abbia un'estensione semantica maggiore, includendo non soltanto il lessico della violenza di genere, ma anche il lessico della criminalità più in generale, con la ricorrenza di parole che fanno riferimento per esempio a furti o ad attività della criminalità organizzata. Un risultato che riflette, in parte, una realtà storica e sociale, ovvero il predominio maschile nelle attività criminali (Fadda 2012), in parte una tendenza mediale abbastanza recente, che è

l'emersione della violenza contro le donne nel discorso dei media (si veda il Capitolo Capitolo 3).

Il secondo risultato attestato dalla comparazione delle CGD dei due *sub-corpora* è una parziale corrispondenza della classe 4, che nel *sub-corpora* donna coincide con l'area lessicale dell'"occupazione femminile" e nel *sub-corpora* uomo all'area della "parità di genere e dei diritti umani". A livello dimensionale, la classe 4 del *sub-corpora* donna è la più piccola individuata alla CGD, con una copertura del 12,44% dei segmenti di testo; mentre la classe 4 del *sub-corpora* uomo è più grande, la seconda per dimensioni dopo la classe 1, con una copertura del 29,31% dei segmenti di testo. A questa differenza di grandezza corrisponde una parziale condivisione di contenuti: la classe 4 del *sub-corpora* donna è specificatamente focalizzata sull'"occupazione femminile", in particolare sul tema della parità salariale e su alcune misure legislative specifiche quali *Quota 100*, *Opzione donna*, *APE*; la classe 4 del *sub-corpora* uomo ha una maggiore estensione semantica, rinviando anch'essa alla parità salariale come discorso centrale, ma entro un contesto più ampio di diritti, incluso il diritto alla sicurezza che per effetto di un apparentamento semantico sfocia nel tema delle migrazioni. In questa comparazione sui contenuti è possibile osservare la centralità del discorso sulla parità salariale sia nel *sub-corpora* donna, sia nel *sub-corpora* uomo a testimonianza del fatto che un argomento specificamente femminile – il *gender pay gap* interessa infatti il genere femminile, storicamente sottopagato anche a parità di ruoli e mansioni professionali (WEF 2020) – trova uno spazio di rappresentazione anche nell'universo delle rappresentazioni maschili. E questo sembrerebbe accadere per effetto di un contributo delle fonti giornalistiche. Se infatti consideriamo le correlazioni significative fra le classi 4 dei due *sub-corpora* e le modalità della variabili categoriali, la classe 4 nel *sub-corpora* donna è significativamente associata alla modalità *N_Servizio, mentre la classe 4 nel *sub-corpora* uomo è significativamente correlata alla modalità *P_Fonte. Nel primo caso, dunque, sono i segmenti di testo di servizi giornalistici a incidere sensibilmente sulla formazione della classe, mentre nel secondo caso sono i segmenti di testo del parlato di fonti

giornalistiche, che hanno accesso all'informazione tramite interviste (la modalità *N_Intervista è significativamente correlata alla classe 4 del *sub-corpus* 4) o messaggi espressi in luoghi pubblici e ritrasmessi dai notiziari (*N_Messaggio è l'altra modalità della variabile significativamente associata alla classe 4 del *sub-corpus* uomo). Detto altrimenti, il tema della parità salariale, con una concentrazione semantica specifica e circoscritta all'"occupazione femminile", nel *sub-corpus* donna emerge come tema dell'agenda (tele)giornalistica; lo stesso tema, con un'estensione semantica generica e ampliata ai diritti umani, nel *sub-corpus* uomo si profila come tema dell'agenda delle fonti. Questo spiega anche la diversa collocazione delle due classi all'interno della rispettiva CGD. La classe 4 nella CGD del *sub-corpus* donna è la prima individuata per separazione dalla classe 1, che – come la classe 1 del *sub-corpus* uomo – è significativamente correlata alla modalità *N_Lancio, dunque generata per maggiore incidenza dell'agenda (tele)giornalistica rispetto all'agenda delle fonti. I lanci dei servizi sono infatti tipicamente preparati in redazione, e presentati da chi conduce il telegiornale. La classe 4 nella CGD del *sub-corpus* uomo, invece, si trova in posizione diametralmente opposta alla classe 1, e in relazione di prossimità con la classe della "memoria collettiva" (classe 3), sulla cui formazione incidono sensibilmente i segmenti di testo marcati *P_Fonte.

Il terzo risultato attestato dalla comparazione delle CGD dei due *sub-corpora* è l'assenza di qualche tipo corrispondenza fra le restanti classi lessicali (2 e 3), che hanno caratteristiche molto diverse fra loro. La classe 2 del *sub-corpus* donna è l'area semanticamente più estesa (34,66%) e concentra una serie di discorsi il cui *trait d'union* è la centralità riconosciuta al ruolo delle donne in diversi ambiti pubblici. Le correlazioni significative con le modalità *P_Fonte, *S_Uomo, *N_Intervista, *N_Commento attestano l'incidenza di segmenti di testo ascrivibili al parlato di fonti maschili – interpellate tramite interviste o spazi di commento – sulla formazione di questa classe che, per questo, risulta la più distante dalla classe 1, dove si concentra il discorso (tele)giornalistico della cronaca. Un posizionamento, quest'ultimo, rilevante perché, da un lato, segna simbolicamente la distanza fra un

mondo di violenza, dove le donne sono oggetto di un discorso giornalistico focalizzato sul loro ruolo di vittime nell'ambito della sfera domestica ("la violenza domestica contro le donne"), e un mondo di protagonismo, dove le donne sono oggetto di un discorso pubblico, veicolato dai notiziari tramite le fonti, focalizzato sul ruolo delle donne nella vita pubblica; dall'altro, testimonia la compresenza nel discorso mediale di due rappresentazioni femminili agli antipodi l'una dall'altra.

La classe 2 del *sub-corpus* uomo è la più piccola individuata dalla CGD (16,23%) e si contraddistingue per l'area lessicale della narrazione cronachistica di "incidenti di varia natura", sulla cui formazione incidono sensibilmente i segmenti di testo di servizi firmati da giornaliste (la classe 2 è infatti correlata significativamente con le modalità *P_Giornalista, *S_Donna e *N_Servizio). A livello gerarchico, si posiziona in relazione di prossimità simmetrica con la classe 1, testimoniando una relazione di vicinanza semantica con l'area della criminalità e della violenza contro le donne, evidentemente in ragione della natura cronachistica di entrambe.

La classe 3 del *sub-corpus* donna è l'area lessicale del "dissenso delle donne", che occupa circa un quinto del *corpus*, ed è significativamente correlata alle modalità *N_Collegamento e *S_Donna, attestando l'incidenza sulla sua formazione di segmenti di testo di discorsi trasmessi dai notiziari tramite collegamenti in diretta in particolari con fonti femminili, fra i quali si distingue la giornalista Giovanna Botteri (corrispondente per la Rai dagli Stati Uniti dal 2007 al 2019 e dalla Cina dal 2019 al 2020) il cui nome e cognome sono associati in modo significativo a questa classe¹¹⁸. Quest'area rappresenta un discorso internazionale in cui le donne sono soggetti di cittadinanza attiva, o, per meglio dire, militanza espressa in diversi movimenti di protesta: dal #MeToo, notoriamente nato per reazione allo scandalo Weinstein negli Stati Uniti, alla presa di posizione di un collettivo di cento donne francesi, fra cui l'attrice Catherine Deneuve, contro il radicalismo generato dal #MeToo accusato di confondere forme di corteggiamento

¹¹⁸ Fonte: https://it.wikipedia.org/wiki/Giovanna_Botteri; ultimo accesso 19 novembre 2021.

e molestie sessuali, condannare alla gogna mediatica e giudiziaria uomini colpevoli di aver flirtato con colleghe, e ridurre le donne a soggetto debole che va protetto¹¹⁹, dalla protesta delle donne indiane contro una legge che discrimina i/le richiedenti cittadinanza di origine afghana, bengalese e pakistana, su base religiosa, fino alla protesta delle donne aderenti al movimento francese dei gilet gialli. In quest'area lessicale le donne sono dunque rappresentate come protagoniste della vita pubblica, e in particolare di movimenti collettivi di protesta contro discriminazioni basate non solo sul genere, ma anche sulla religione o la classe sociale. A incidere su questa rappresentazione sono soprattutto, come abbiamo visto, fonti femminili, in collegamento con lo studio del telegiornale, dunque fonti giornalistiche, perché la modalità *N_Collegamento classifica soltanto segmenti di fonti interne (alla redazione del telegiornale o di altri programmi di rete). Si tratta di un risultato interessante che testimonia una correlazione positiva fra il giornalismo femminile e una rappresentazione delle donne come cittadine attive nella vita pubblica, piuttosto che come osservatrici passive (Ross e Carter 2011).

La classe 3 del *sub-corpus* uomo rinvia a tutt'altro universo semantico, lontano dall'attualità del "dissenso femminile": "la memoria collettiva". Questa classe copre circa un quinto del *corpus* (21,98%) ed è strutturalmente prossima alla classe 4, della "parità di genere e diritti umani", con la quali condivide la caratteristica di essere uno spazio di tematizzazione ascrivibile alle fonti. Le modalità *P_Fonte, *N_Messaggio e *N_Intervista sono infatti significativamente correlate a questa classe. In termini contenutistici, si contraddistingue per una

¹¹⁹ A titolo esemplificativo, si riporta un pezzo saliente della lettera pubblicata da *Le Monde*: «Questa giustizia rapida ha già avuto le sue vittime, uomini puniti nell'esercizio della loro professione, costretti a dimettersi, ecc., quando la loro unica colpa era quella di aver toccato un ginocchio, cercato di rubare un bacio, parlato di cose "intime" durante una cena professionale o aver inviato messaggi con connotazioni sessuali a una donna per la quale l'attrazione non era reciproca. Questa febbre di mandare i "maiali" al macello, lungi dall'aiutare le donne a responsabilizzarsi, serve in realtà gli interessi dei nemici della libertà sessuale, degli estremisti religiosi, dei peggiori reazionari e di coloro che credono, in nome di una concezione sostanziale del bene e della morale vittoriana che ne deriva, che le donne siano esseri "separati", bambini con il volto adulto, che chiedono di essere protetti» (la traduzione è mia); fonte: https://www.lemonde.fr/idees/article/2018/01/09/nous-defendons-une-liberte-d-importuner-indispensable-a-la-liberte-sexuelle_5239134_3232.html (ultimo accesso 19 novembre 2021).

sorta di celebrazione di persone famose con un ruolo riconosciuto nella vita pubblica.

L'interpretazione dei profili lessicali delle classi e del loro grado di corrispondenza/non corrispondenza sul piano sintagmatico, nei due *sub-corpora*, ha un valore euristico di per sé, ma anche ai fini della comparazione delle mappe fattoriali dei due *corpora*. Il primo fattore che spiega la distribuzione delle classi sul piano cartesiano del *sub-corpus* donna è la polarità pubblico-privato (Grafico 18), mentre il primo fattore che spiega la distribuzione delle classi sul piano cartesiano del *sub-corpus* uomo è la polarità tematizzazione-cronaca (Grafico 20). Questa differenza indica una rappresentazione stereotipata anzitutto rispetto al genere maschile, che continua a essere incorniciato, e simbolicamente confinato, entro lo spazio della vita pubblica. L'assenza di un fattore latente che spieghi la distribuzione delle classi lessicali su una polarità pubblico-privato, nel *sub-corpus* uomo, è infatti riconducibile a una incongruenza di questo fattore rispetto al profilo delle classi e alla loro relazione gerarchica: la classe della "criminalità e violenza contro le donne", per via dell'estensione semantica alla criminalità comune (non agita contro le donne e non agita in ambito domestico), ha come riferimento uno spazio più ampio rispetto alla sfera domestica o privata, inoltre è posizionata sullo stesso versante della classe degli "incidenti di varia natura", che, secondo quanto attestato dal profilo lessicale e dai STT, rinvia a contesti sia pubblici sia privati, dunque non può essere spiegata in termini di pubblico-privato, ma più propriamente in termini di tematizzazione-cronaca.

La classe 1 nel *sub-corpus* donna, invece, fa riferimento in modo abbastanza chiaro a una dimensione prettamente privata, non solo per via del riferimento specifico alla violenza contro le donne, in particolare subita nell'ambito di relazioni intime, ma soprattutto per via del suo posizionamento rispetto alle altre classi. Tutte le altre classi del *sub-corpus* donna si collocano infatti sul versante del pubblico, lungo l'asse delle ascisse, in coerenza con il riferimento ad ambiti che sono pubblici o meglio rappresentati come tali (il lavoro femminile, come tema politico; il dissenso delle donne manifestato nelle piazze e nelle città di diversi paesi

del mondo; il ruolo delle donne in diversi ambiti della vita politica, sociale o artistico-culturale). Questo posizionamento isolato della classe 1 sul versante privato dell'asse pubblico-privato, da un lato, conferma la presenza di una narrazione della violenza contro le donne stereotipata, nel senso di ricondotta a una questione privata, familiare, piuttosto che alla matrice culturale e sociale, dunque pubblica, da cui trae origine e forza di perpetuazione; dall'altra, attesta – per distanziamento – un discorso mediale “altro”, non stereotipato, che celebra le donne per il loro ruolo in diversi ambiti della vita pubblica (classe 2), che le rappresenta come soggetti attivi nelle manifestazioni di protesta (classe 3) e, infine, che le mette al centro di politiche attive volte a superare disparità in ambito lavorativo (classe 4).

Il secondo fattore latente che spiega la distribuzione delle classi sul piano cartesiano è stato interpretato per entrambi i *corpora* in termini di polarizzazione sull'asse soggettività-oggettività. Sia nel *sub-corpus* donna sia nel *sub-corpus* uomo è possibile osservare una distribuzione delle classi secondo una rappresentazione che privilegia la dimensione soggettiva, in alcuni contesti discorsivi, e la dimensione oggettiva, in altri. Nel *sub-corpus* donna, le classi lessicali collocate sul versante della soggettività sono la 2 e la 3, rispettivamente coincidenti con l'area della “centralità del ruolo delle donne” e con l'area del “dissenso delle donne”, che non hanno alcun corrispettivo nel *sub-corpus* uomo, dove sullo stesso versante sono collocate la classe 1, della “criminalità e della violenza contro le donne” che ha una parzialmente corrispondenza nella classe 1 del *sub-corpus* donna, e la classe 3, della “memoria collettiva”, che non ha alcuna corrispondenza con il *sub-corpus* donna. Questa diversa distribuzione delle classi nei quadranti superiori del piano cartesiano dei due *sub-corpora* testimonia come la soggettività femminile sia rappresentata sia in contesti di attualità, com'è l'area coperta dalla classe 3, sia in contesti misti, dove l'attualità si mescola con la storia nel celebrare il ruolo pubblico delle donne; viceversa, la soggettività maschile è rappresentata o nel tradizionale contesto della criminalità o nel contesto storico della memoria.

Le classi posizionate sul versante dell'oggettività, nel *sub-corpora* donna, sono la 1 e la 4, coincidenti con "la violenza contro le donne" e "l'occupazione femminile", caratterizzate tutte e due da un profilo specifico presente anche nelle classi 1 e 4 del *sub-corpora* uomo, entro un contesto però più esteso. Anche la classe 4 del *sub-corpora* uomo è posizionate sul versante dell'oggettività, testimoniando la presenza di un'area in cui la parità di genere e i diritti umani sono l'oggetto di un discorso in modo del tutto simile alla "occupazione femminile" nel *sub-corpora* donna. La classe 1, invece, nel *sub-corpora* uomo, è collocata sul versante della soggettività, rinviando a un mondo di criminalità e violenza in cui gli uomini agiscono piuttosto che subire la violenza. Un risultato quest'ultimo che almeno in parte riflette una realtà sociale in cui la devianza dalla norma è ancora appannaggio maschile.

5.4.4 *L'impatto dell'agenda e della linea editoriale*

La comparazione dei risultati per anno e per TG sulle rappresentazioni di genere di donne e uomini come argomento/oggetto dell'informazione consente di avanzare ipotesi sull'impatto dell'agenda mediale entro i limiti della composizione dei due *sub-corpora* donna e uomo che sono ricavati da testi di notiziari che coprono di fatto soltanto un mese per anno. Per questo, come è stato fatto nel paragrafo sulle rappresentazioni delle fonti, è importante tenere presente la composizione delle classi ed eventuali corrispondenze di classe fra i due *sub-corpora*.

Il primo risultato emergente dal confronto delle diverse correlazioni fra le classi lessicali dei due *corpora* e le modalità della variabile *Anno è l'assenza di associazioni significative per alcune classi lessicali: la classe 1 del *sub-corpora* donna ("la violenza domestica contro le donne"), la classe 2 del *sub-corpora* uomo ("gli incidenti di varia natura") e le classi 4 di entrambi i *sub-corpora* ("occupazione femminile" e "parità di genere e diritti umani"). L'assenza di correlazioni indica che sulla formazione di queste classi non incidono i segmenti di testo di un anno o di un altro, testimoniando dunque che queste aree semantiche presentano una stabilità lungo l'asse diacronico. Detto altrimenti "la violenza domestica contro le donne",

“gli incidenti di varia natura”, “l’occupazione femminile” e “la parità di genere e i diritti umani” sono temi costanti nell’agenda dei telegiornali nei tre periodi considerati.

Le restanti aree semantiche attestano invece qualche associazione più o meno significativa con alcune modalità della variabile *Anno indicando variazioni periodiche. La correlazione della classe 1 del *sub-corpus* uomo (“la criminalità e la violenza contro le donne”), positiva con la modalità *A_2019 e negativa con la modalità *A_2020, dimostra quanto questa area lessicale, focalizzata sulle soggettività maschili che agiscono la violenza, sia un discorso mediale sensibile al variare del contesto, a differenza della “violenza domestica contro le donne” (classe 1 del *sub-corpus* donna) focalizzata sulle vittime di violenza. Anche le associazioni osservate per la classe 3 del *sub-corpus* uomo (“la memoria collettiva”) riflettono una instabilità pertinente la tematizzazione della memoria collettiva, su cui incidono sensibilmente segmenti di testo dell’anno 2020. Nel *sub-corpus* donna, le correlazioni osservate per la classe 2 (“centralità del ruolo delle donne”) attestano una maggiore incidenza alla formazione di questa classe dei segmenti di testo del 2020, viceversa le correlazioni osservate per la classe 3 (“dissenso delle donne”) attestano una maggiore incidenza alla formazione di questa classe dei segmenti di testo del 2018. Due risultati che possono essere spiegati, data la prossimità semantica delle classi, anche come testimonianza di bilanciamenti periodici di un discorso mediale che rimane costantemente focalizzato sulle soggettività femminili, con un’attenzione che nel 2018 è prevalentemente incentrata su manifestazioni di protesta e nel 2020 sul ruolo delle donne nella società.

Per quanto riguarda la comparazione per TG, anche in questo caso si osserva l’assenza di associazioni significative per alcune classi: la classe 2 del *sub-corpus* donna e le classi 2 e 3 del *sub-corpus* uomo. Un risultato che indica, per quanto riguarda il *sub-corpus* donna, una trasversalità della rappresentazione della “centralità del ruolo delle donne” (classe 2), rispetto a tutte e tre le testate giornalistiche e, per quanto riguarda il *sub-corpus* uomo, una trasversalità del

discorso cronachistico sugli “incidenti di varia natura”(classe 2) e del discorso sulla “memoria collettiva” (classe 3).

Le classi per le quali si osservano invece correlazioni significative, positive o negative, con le diverse modalità della variabile *TG, nel *sub-corpus* donna, sono la 1, la 3 e la 4 e nel *sub-corpus* uomo, sono la 1 e la 4. Nel *sub-corpus* donna, la modalità *T_TG5 è positivamente correlata con la classe 1 e negativamente con le classi 3 e 4, attestando quanto l’informazione del TG5 contribuisca sensibilmente alla rappresentazione della “violenza domestica contro le donne” (classe 1) e alla distinzione di quest’area lessicale da quelle del “dissenso delle donne” (classe 3) e della “occupazione femminile” (classe 4). Sempre nel *sub-corpus* donna, la modalità *T_TG2 è invece negativamente correlata alla formazione della classe 1, testimoniando come i segmenti di testo del TG2, a differenza di quelli del TG5, non contribuiscano alla rappresentazione della “violenza domestica contro le donne”.

Nel *sub-corpus* uomo, la modalità *T_TG5 è positivamente correlata con la classe 1 e negativamente con la classe 4, attestando quanto l’informazione del TG5 contribuisca sensibilmente alla rappresentazione della “criminalità e violenza contro le donne” (classe 1) e alla distinzione di quest’area lessicale dall’area della “parità di genere e diritti umani” (classe 4). Classe, quest’ultima, che è invece positivamente correlata alla modalità *T_TG1, attestando una incidenza dei segmenti di testo del notiziario di Raiuno sulla formazione di quest’area semantica.

In sintesi, i risultati sulle correlazioni significative più rilevanti sono tre. Il primo riguarda l’incidenza positiva dell’informazione del TG5 sulla classe 1 del *sub-corpus* uomo, la “criminalità e violenza contro le donne” parzialmente corrispondente alla classe 1 del *sub-corpus* donna, la “violenza contro le donne”, e la contestuale non incidenza sulla classe 4, “parità di genere e diritti umani”, che ha una parziale corrispondenza con la classe “occupazione femminile” del *sub-corpus* donna. Il secondo risultato riguarda la non incidenza dei segmenti di testo del TG2 sulla rappresentazione della “violenza domestica contro le donne” (*sub-corpus*

donna). Il terzo risultato è invece il contributo distintivo del TG1 alla tematizzazione della “parità di genere e diritti umani” nel *sub-corpus* uomo.

5.5 Conclusioni

L’analisi delle rappresentazioni di genere di donne e uomini come soggetto/fonte e come oggetto/argomento dell’informazione televisiva, attraverso l’esplorazione lessicale e testuale del *corpus* TG, ha restituito un quadro complesso, dove sono riconoscibili asimmetrie e stereotipi di genere evidenziati in precedenti ricerche, ma anche segnali di attualità e innovazione.

Le donne come fonte di informazione sono anzitutto interessate da un discorso identitario autoreferenziale che non ha corrispondenza per le fonti maschili. Attraverso la parola *donna*, le fonti femminili esplicitano la propria identità di genere, a livello individuale o collettivo, come singole o appartenenti al gruppo sociale delle donne, o l’identità di genere di altre donne. Viceversa, la parola *uomo* non ha questo valore identitario, o per lo meno non viene usato dagli uomini per marcare la propria appartenenza al genere maschile. Questo risultato è confermato ai vari livelli dell’analisi sulle fonti e attesta due fenomeni: primo, come l’informazione, più o meno consapevolmente, veicola l’asimmetria semantica che interessa la coppia di nomi *donna-uomo* in termini di androcentrismo linguistico (cfr. Capitolo 1, § 1.2), laddove il primo indica un genere della specie umana (quello femminile), mentre il secondo indica l’altro genere (maschile) ma anche l’intera specie; secondo, come l’informazione sia uno spazio in cui le donne esprimono esplicitamente la propria identità di genere, forse anche su una sollecitazione in tal senso da parte dei giornalisti e delle giornaliste, come dimostrato anche da precedenti ricerche (cfr. CNEL 2004).

Il secondo risultato attesta una polarizzazione di genere molto tradizionale, quella fra la sfera pubblica, associata agli uomini, e la sfera privata, associata alle donne: le parole distintive del lessico femminile rinviano prevalentemente a questioni come la maternità, l’educazione, la vita domestica, mentre quello

maschile a questioni pubbliche, in particolare di politica e governo. Questo risultato conferma come l'informazione televisiva rafforzi uno degli stereotipi più datati sulla lingua parlata dalle donne, quello che riguarda gli argomenti di cui le donne parlano da una posizione ai margini del sistema del potere: argomenti «out of power», per usare le parole di Lakoff (1973, 53), argomenti che riguardano le esperienze tradizionalmente femminili come la maternità (Spender 1980, 1985), argomenti che riguardano la vita domestica (Berretta 1983). Con la differenza che, se negli anni Settanta del Novecento, le donne, fatta qualche rara eccezione, erano realmente «out of power», lo stesso non si può dire oggi, con la presenza di numerose donne in tutti gli ambiti della sfera pubblica. Il fatto che esse non trovino un adeguato spazio di espressione nei mezzi di informazione dipende forse dal fatto che sono soprattutto “out of the power of the news media”, ovvero fuori dal potere di ottenere una rappresentazione bilanciata e non stereotipata, come dimostrano almeno altri due risultati dell'analisi del *corpus* TG.

La diversa dimensione dei due *sub-corpora* PF_Donna e PF_Uomo, l'uno circa un terzo rispetto all'altro, conferma infatti la “regola del terzo”, secondo la quale la visibilità delle donne nei contenuti dell'informazione, dopo un periodo di crescita, si è fermata a una soglia di attenzione di una donna ogni due uomini nella maggior parte dei paesi del mondo, inclusa l'Italia (Ross e Azzalini 2017, 41; ISIMM Ricerche e Università degli studi Roma Tre 2017, 27; IZI et al. 2020, 19; Djerf-Pierre e Esdröm 2020b, 33).

Inoltre, le CGD dei *sub-corpora* P_Fonte, e poi PF_Donna e PF_Uomo a confronto, hanno messo in evidenza, primo, come le donne siano escluse dal discorso scientifico, un campo semantico presente solo nel *sub-corpora* uomo, e, secondo, come i due campi semantici condivisi, quello della “testimonianza di vita” e della “comunicazione politica”, non siano pienamente condivisi. Nel *sub-corpora* P_Fonte, le fonti femminili sono infatti significativamente associate alla classe delle “testimonianze di vita”, mentre gli uomini alla “comunicazione politica”. Entrambe queste classi sono attestate separatamente sia per il *sub-corpora* PF_Donna che il per *sub-corpora* PF_Uomo ma con proporzioni diverse. Nel *sub-corpora* PF_Donna, le

“testimonianze di vita” coprono l’80,25% dei 1.949 segmenti di testo classificati dalla CGD vs. il 41,92% dei 6.278 segmenti di testo classificati dalla CGD del *sub-corpus* PF_Uomo. La “comunicazione politica”, invece, copre il 19,75% dei segmenti di testo del *sub-corpus* PF_Donna vs. il 39,96% del *sub-corpus* PF_Uomo. In sintesi, questi dati indicano come le “testimonianze di vita” sono testimoniate tanto dalle fonti femminili, quanto da quelle maschili, mentre la “comunicazione politica” è prevalentemente appannaggio maschile. A maggior ragione se consideriamo che le fonti maschili sono più produttive di quelle femminili, evidentemente per via di uno spazio maggiore riservato alle prime. Come già osservato nella discussione dei risultati, queste differenze non sembrano tanto ascrivibili a specificità linguistiche, quanto ai ruoli, e soprattutto ai diversi spazi offerti per esercitare questi ruoli, assegnati alle donne e agli uomini.

Gaetano Berruto, in uno dei primi contributi della sociolinguistica italiana sulla questione delle differenze di lingua, fra uomini e donne, scrive:

le differenze certamente riconoscibili nella propensione a usare e nell’uso effettivo di termini e campi semantici specifici sono da ricondurre in massima parte non a fatti “linguistici”, bensì alle diverse sfere di argomenti e settori di esperienze che uomini e donne sono soliti trattare preferenzialmente in dipendenza della divisione sociale dei rispettivi ruoli. Si tratta cioè di diverse sfere di riferimento esplicitate linguisticamente, e non di diversità di “trattamento” linguistico delle stesse sfere (Berruto 1980, 146).

Questa affermazione coglie in modo molto pertinente i risultati emersi dall’esplorazione del *corpus* TG. Le donne come fonte dell’informazione sono prevalentemente associate a un certo vocabolario e a certi campi semantici, perché sono prevalentemente interpellate come testimoni della vita quotidiana, una sfera quest’ultima tradizionalmente, e stereotipicamente, associata al genere femminile. Fatta eccezione per l’uso della parola *donna*, il lessico di uomini e donne è infatti molto simile in corrispondenza dei medesimi campi semantici, ma le variazioni

vanno nella direzione di ampliare il campo semantico femminile verso la sfera privata e quello maschile verso la sfera pubblica.

Per quanto riguarda l'impatto dell'agenda dell'informazione, in termini di notizie che la compongono nei tre periodi coperti dal *corpus* TG, i risultati attestati dalla comparazione per anno evidenziano tutti una instabilità delle classi lessicali di tutti i *sub-corpora* analizzati, indicativa di come sia le fonti femminili sia le fonti maschili siano portatrici di discorsi "instabili", che possono variare in relazione ai diversi eventi che superano la soglia di notiziabilità. Mentre la comparazione per TG mette in evidenza diverse linee editoriali, fra cui si distingue la linea "innovativa" del TG2, che privilegia le fonti femminili come soggetto della "comunicazione politica" piuttosto che come soggetto di "testimonianza di vita". Considerato che il TG2 è stato inserito nel *corpus* TG per verificare l'eventuale differenza apportata da una linea editoriale dichiaratamente impegnata sul fronte delle pari opportunità di genere, come quella di Ida Colucci direttrice del TG2 nel gennaio 2018, è ragionevole ipotizzare che questo risultato sia dovuto alla sua direzione, anche se non è stato possibile confrontare il TG2 del 2018, con le edizioni del 2019 e 2020 per verificare puntualmente questa ipotesi (una comparazione simile avrebbe richiesto una disaggregazione dei *sub-corpora* analizzati in ulteriori *sub-corpora*, che sarebbero diventati troppo piccoli).

Venendo ora alla rappresentazione di donne e uomini come argomento/oggetto dell'informazione, il primo risultato rilevante riguarda un equilibrio di genere in termini spazio di attenzione per donne e uomini come *topic* dell'informazione. A questo equilibrio quantitativo corrisponde una similarità lessicale, tanto da poter profilare un lessico dominante che rappresenta entrambi i generi come oggetto di un'informazione prevalentemente cronachistica, incentrata, da un lato, su eventi di criminalità e violenza e, dall'altro, su relazioni affettive e familiari fra i protagonisti. Le differenze di genere emergono, come secondo risultato, da un'analisi più approfondita (CGD e analisi fattoriale) che porta in evidenza stereotipi e asimmetrie tradizionali, ma anche qualche elemento di innovazione.

Si collocano sul primo versante, la rappresentazione dominante delle donne nel campo semantico della violenza contro le donne, contraddistinto dal fattore latente della “oggettività”, indicativo di una rappresentazione vittimistica delle donne, vs. la rappresentazione non dominante degli uomini nel campo semantico più esteso della criminalità, di cui la violenza di genere è soltanto una componente, ed è peraltro caratterizzato dal fattore latente della “soggettività”, indicativo di una rappresentazione attiva degli uomini in questa sfera dell’agire sociale, non solo come criminali, ma anche come rappresentanti delle forze dell’ordine, di soccorso e così via. Un’asimmetria simile interessa anche le questioni di genere tematizzate in modo stabile nei tre periodi considerati, ovvero quelle riconoscibili entro il campo semantico della “occupazione femminile”, che tematizza le donne come oggetto di un discorso specifico, e quelle riconoscibili entro il campo semantico delle “pari opportunità e diritti”, che topicalizza gli uomini in modo più esteso, come soggetti di un discorso di parità e diritti più ampio, confermando l’uso sovraesteso del lemma *uomo*. Su questo versante è possibile collocare anche il risultato pertinente la “centralità delle donne”, un’area semantica che pone al centro del discorso mediatico il ruolo delle donne nella società, nella storia, nella politica, e nell’arte, ma che è significativamente associata alla modalità della variabile *S_Uomo, ciò vuol dire che la rilevanza dei ruoli femminili è mediatizzata da parlanti di genere maschile, giornalisti o uomini fonti di notizia. Questo risultato conferma la tendenza dell’informazione a legittimare la valorizzazione delle donne attraverso la voce, il pensiero, la testimonianza degli uomini (cfr. Capecchi 2001; CNEL 2004),

Rappresenta invece un elemento di innovazione, la rappresentazione delle donne come voce e soggetto del “dissenso delle donne”, che, come abbiamo visto, è un campo semantico che fa riferimento non solo alla ribellione pubblicamente manifestata dalle donne contro le molestie sessuali, ma anche contro altre forme di violenza e discriminazione, per esempio su base religiosa, politica, sociale. La portata innovativa di questo risultato è duplice: da un lato, evidenzia l’emersione di uno spazio di attenzione, ancorché piccolo, per una questione specificamente femminile e femminista a cui, in genere, l’informazione *mainstream* presta poca

attenzione (cfr. Azzalini e Padovani 2015, 2016, 2021), dall'altro, mostra come questo spazio si configuri come una rappresentazione delle donne quali cittadine, soggetti attivi nella sfera pubblica, piuttosto che osservatrici passive (Ross e Carter 2011). Infine, è interessante notare come questo "campo semantico" sia positivamente correlato con fonti giornalistiche femminili, testimoniando come le giornaliste possano avere un ruolo nel configurare nuovi spazi identitari nei contenuti dell'informazione.

Questa ipotesi è indirettamente (e parzialmente) confermata anche dal risultato che riguarda la positiva non incidenza dei segmenti di testo del TG2, che nel 2018 era diretto da una giornalista, sulla formazione della classe lessicale della "violenza domestica contro le donne", nel *sub-corpus* donna. Quest'area semantica, come abbiamo visto, si contraddistingue per una rappresentazione delle donne come vittime vs. una rappresentazione delle stesse come cittadine attive, quindi si configura come problematica, nella misura in cui diventa dominante (33,09%) rispetto ad altre aree che contribuiscono a una costruzione di identità femminili attive, come per esempio "il dissenso delle donne" (19,81%). Purtroppo la conferma è parziale perché una conferma completa potrebbe derivare soltanto da una disaggregazione del *sub-corpus* donna per anno e una successiva analisi comparativa dei *sub-corpora* risultanti dalla disaggregazione, la quale tuttavia avrebbe ridotto i *sub-corpora* a una dimensione troppo piccola, inadeguata a un'analisi con IRaMuTeQ.

Capitolo 6 Analisi dell'accordo di genere nei nomi di ruolo politico in un *corpus* di quotidiani: uno studio del lemma *ministra/o*

Questo capitolo presenta un'analisi dell'accordo di genere nei nomi di ruolo politico *corpus-based*, attraverso uno studio del lemma *ministra/o* nel *corpus* TG e nel *corpus* CRS, descritti nel Capitolo 4 (§ 4.1 e § 4.3). Il capitolo è introdotto da una breve contestualizzazione del dibattito attualmente in atto in Italia sul sessismo nella lingua italiana, che è prevalentemente focalizzato sulla questione dell'uso dei femminili (§ 6.1). Il paragrafo successivo (§ 6.2) è dedicato allo stato dell'arte sulle ricerche che, a partire dagli studi pioneristici di Alma Sabatini (1986, 1987), rilevano una resistenza a declinare al femminile nomi di professioni e/o ruoli ricoperti da donne, attraverso l'analisi del linguaggio giornalistico assunto come indicatore di usi linguistici consolidati, così come di cambiamenti linguistici.

Partendo da una rassegna degli studi basati su *corpora* giornalistici o del web, nel paragrafo 6.3 è avanzata l'ipotesi che il linguaggio dei media sia caratterizzato da un androcentrismo linguistico, per via anche, ma non solo, di una resistenza all'uso delle forme femminili. Sulla base di questa ipotesi, nello stesso paragrafo vengono formulate le seguenti domande di ricerca: con quale frequenza il lemma *ministra/o* ricorre nelle sue quattro forme (*ministra, ministre, ministro, ministri*)? Con quale frequenza il lemma è declinato in modo coerente con il genere semantico di riferimento? Quali espressioni linguistiche vengono usate nel riferimento a ministre e ministri? Tutte e tre le domande di ricerca sono applicate al *corpus* CRS, al *corpus* TG integrale, e anche al *corpus* TG ridotto al 2018 e 2019, per una comparazione puntuale con il *corpus* CRS. Le prime domande di ricerca includono entrambe due domande specifiche pertinenti eventuali differenze fra i periodi e le testate giornalistiche considerate, funzionali a verificare se il rinnovato interesse di alcune istituzioni politiche, in particolare l'impegno di Laura Boldrini durante la Presidenza alla Camera nel corso della 17esima legislatura, per le questioni della parità linguistica, e la recente pubblicazione di linee guida per un uso *gender-fair* dell'italiano destinate al mondo dell'informazione, hanno in

qualche modo favorito un minore androcentrismo linguistico nell'informazione italiana. Dopo aver introdotto le ipotesi e le domande di ricerca, il paragrafo 6.4 espone la metodologia di analisi utilizzata, che replica parzialmente lo studio di Formato (2016), individuato come il più innovativo fra gli studi recenti in materia. Nei paragrafi successivi (§ 6.5 e § 6.6) vengono presentati e discussi i risultati. Il capitolo si conclude con il paragrafo 6.7 che mette in evidenza come nei *corpora* analizzati si rilevi un indebolimento dell'androcentrismo linguistico attestato dagli studi precedenti, in particolare nell'analisi di Formato (2016), qui parzialmente replicata, e avanza l'ipotesi che diverse linee editoriali possano svolgere un ruolo importante nel promuovere un linguaggio di genere paritario, specie se sostenute da un clima politico favorevole.

6.1 Introduzione

Nonostante l'italiano sia una lingua dal genere marcato a livello morfo-sintattico e si caratterizzi nella maggior parte dei casi per una flessione coerente con l'identità di genere del/la referente, la concordanza di genere nei nomi di ruolo apicale riferiti a donne è una questione ancora aperta nell'agenda culturale del nostro paese (Giusti 2016a; Gomez Gane 2017; Adamo et al. 2019; Azzalini e Giusti 2019; Somma e Maestri 2020). Introdotta nelle prime riflessioni sul rapporto fra genere e lingua dai lavori Alma Sabatini (1986, 1987), promossi dalla *Commissione per la realizzazione della parità tra uomo e donna* e dalla Presidenza del Consiglio dei ministri, con l'obiettivo di rimuovere gli stereotipi «perpetuati e avvalorati attraverso la lingua, in modo spesso subdolo e ripetitivo» (Sabatini 1987, 23), la questione è rimasta a lungo circoscritta all'ambito accademico e al pensiero femminista. Solo di recente è rientrata nell'agenda nazionale, con un rinnovato impegno, degli organismi politici e degli organi d'informazione, per un uso della lingua più paritario e inclusivo, come auspicato dal volume di Sabatini, attraverso il contributo di Sergio Lepri (1987).

Come anticipato nel Capitolo 2 (§ 2.3.3), l'8 marzo 2015, l'allora presidente della Camera Laura Boldrini ha inviato una lettera a deputate e deputati esortandoli

a richiamare, nei loro interventi, le cariche e i ruoli istituzionali «correttamente, secondo il genere proprio della persona cui essi si riferiscono»¹²⁰. In quello stesso anno, l'Ordine dei giornalisti italiani ha pubblicato il volume *Tutt'altro genere d'informazione. Manuale per una corretta rappresentazione delle donne nell'informazione* (CNOG 2015), contenente un capitolo dedicato alla riflessione sulla lingua completo di raccomandazioni per un uso paritario e inclusivo dell'italiano e indicazioni per declinare correttamente al femminile nomi di professioni, ruoli e cariche pubbliche (Cavagnoli 2015). L'anno prima, l'associazione di giornaliste GiULiA ha stampato un opuscolo eloquentemente intitolato *Donne, grammatica e media* (Robustelli 2014) destinato a divulgare fra giornaliste e giornalisti una maggior consapevolezza sul sessismo palese e latente in numerosi usi linguistici attestati nei contenuti dei media, fra cui la designazione di professioniste e politiche con nomi maschili. Usare un nome di professione, ruolo o carica pubblica declinato al maschile quando si fa riferimento a una donna implica infatti, da un lato, nascondere la sua identità di genere, dall'altro, trasmettere l'idea che una carica, un ruolo o una professione perde di prestigio o autorevolezza se nominata al femminile (Azzalini e Giusti 2019; Fusco 2020).

Partendo dallo stato dell'arte, a cui è dedicato il secondo paragrafo, questo capitolo della tesi si propone di analizzare il grado di androcentrismo linguistico dell'informazione italiana, attraverso uno studio sulla concordanza di genere del lemma *ministra/o* nel *corpus* TG e nel *corpus* CRS, e verificare se l'uso dei femminili è favorito o meno dall'«impegno consapevole degli organismi politici e di governo» e dall'«impegno convinto dei responsabili delle testate giornalistiche» (Lepri 1987, 22).

6.2 Stato dell'arte

Dato il ruolo fondamentale dei media nel «captare e ridiffondere i segnali dell'attualità» (Masini 2016, 31) e allo stesso tempo testimoniare usi consolidati,

¹²⁰ Fonte: <https://twitter.com/lauraboldrini/status/573436739773153280>; ultimo accesso 19 novembre 2021.

originariamente intuito da Alma Sabatini (1987), molti studi sulla concordanza tra genere grammaticale e genere semantico in Italia si concentrano su *corpora* di giornali quotidiani (Burr 1995; Pescia 2010; Cavagnoli 2013; Robustelli 2014; Formato 2016; Castenetto e Ondelli 2020), e più di recente su *web corpora* (Jacqmain 2005; Nardone 2016, 2018).

Trascurando gli studi con un approccio qualitativo, che si limitano a osservare diverse strategie linguistiche ricorrenti nei quotidiani per nominare le donne in professioni o ruoli di prestigio (Cavagnoli 2013; Robustelli 2014), Burr (1995) è il primo lavoro che adotta un approccio quantitativo su un piccolo *corpus* di giornali (4 edizioni quotidiane per il *Corriere della sera*, 2 per *Il Mattino*, 2 per *la Repubblica* e 2 per *La Stampa*, per un totale di 9 edizioni). Attraverso un'analisi lessicometrica computazionale, l'autrice calcola la frequenza di tutte le forme femminili e maschili degli agentivi occorrenti nel *corpus* e ne verifica la congruenza con il genere semantico del/la referente, attestando un livello di congruenza nei riferimenti alle donne pari all'85,9%, riferimenti che però nel complesso sono solo il 5,75%. Nel 46,01% delle occorrenze, infatti, il referente è un uomo, e nel 48,24% non è determinabile, perché il genere grammaticale è maschile e il genere semantico non è ulteriormente specificato da elementi satellite e, data la tendenza a usare forme maschili per designare le donne, non può essere attribuito con certezza a un uomo. Sommando tutte le occorrenze delle forme maschili, con riferimento chiaramente maschile o riferimento indeterminabile rispetto al genere, la percentuale degli agentivi maschili attestata da Burr (1995) è pari al 94,25%, un valore indicativo di un linguaggio della stampa quotidiana fortemente androcentrico.

Pescia (2010) svolge uno studio comparativo sull'uso di forme femminili e maschili di nomi di professione e cariche politiche fra l'italiano parlato in Svizzera e quello parlato in Italia, analizzando due *corpora*, uno costituito da testi rappresentativi di tre diversi registri dell'italiano svizzero e l'altro da testi rappresentativi dell'italiano standard. Il primo *corpus* include quotidiani in lingua italiana distribuiti nel Canton Ticino della Svizzera (189 edizioni totali), contenenti

tre tipi di articoli: 1) articoli redazionali/di cronaca locale in italiano ticinese pubblicati da tre quotidiani locali (*La Regione, Corriere del Ticino e Giornale del Popolo*); 2) articoli dell'ATS (Agenzia Telegrafica Svizzera), in italiano elvetico, e 3) articoli dell'ANSA (Agenzia Nazionale Stampa Associata), in italiano standard, tutti pubblicati dal 1° novembre 2008 al 30 novembre 2009. Il secondo *corpus* contiene articoli pubblicati nello stesso periodo da *la Repubblica* e, di nuovo, dall'ANSA (inclusa in entrambi i *corpora* per una doppia comparazione). Le conclusioni a cui perviene l'autrice è che nel primo *corpus* prevalgono le forme femminili, in particolare nel *sub-corpus* dell'ATS, in italiano elvetico, mentre i *sub-corpora* dei tre quotidiani locali e dell'ANSA sono interessati, come il *sub-corpus* de *la Repubblica* da maggiori oscillazioni fra forme femminili e maschili. La maggiore produttività delle forme femminili riscontrata nei tre *sub-corpora* rappresentativi dell'italiano svizzero viene attribuita dall'autrice in parte all'intervento delle istituzioni locali, che nel 1995 hanno introdotto un manuale con le *Tecniche per la redazione degli atti ufficiali* (Cantone Ticino 1995, 2004) contenente indicazioni precise per la femminilizzazione delle cariche ufficiali, resa contestualmente obbligatoria dal Consiglio di Stato ticinese¹²¹, e in parte all'influenza della cultura tedesca che, sia in Austria, sia Germania ha adottato sin dagli anni Ottanta politiche linguistiche per un linguaggio di genere paritario (cfr. Capitolo 2, §2.3.3). Influenza quest'ultima particolarmente evidente per gli articoli dell'ATS, con sede a Berna, quindi nella Svizzera tedesca, meno evidente invece negli articoli redazionali dei tre quotidiani in italiano ticinese interessati da oscillazioni indicative di una residuale tendenza all'uso dei maschili nel riferimento alle donne, come i *corpora* in italiano standard dell'ANSA e de *la Repubblica*, dove nomi di cariche politiche, come per esempio *assessore, ministro, sindaco e sotto-segretario*, e di cariche giudiziarie o professioni storicamente appannaggio maschile, come *avvocato, ingegnere e architetto* ricorrono quasi solo al maschile, mentre *consigliera, governatrice, deputata e senatrice* ricorrono prevalentemente al femminile. Pescia (201) osserva anche che

¹²¹ Successiva allo studio di Pescia (2010) è la pubblicazione del volume *Pari trattamento linguistico. Guida al pari trattamento linguistico di donna e uomo nei testi ufficiali della Confederazione Cancelleria federale* (2012).

nei *corpora* in italiano i nomi di cariche o professioni sono più facilmente declinati al femminile nel riferimento a persone non famose o straniere, forse in funzione di rendere esplicita l'identità di genere della referente poco nota o straniera. Purtroppo i risultati non sono riportati con precisione, ma solo con grafici che indicano le frequenze sull'asse delle ordinate, in un intervallo di 10 punti percentuali. Per quanto riguarda il lemma *ministra/o*, da una lettura dei grafici si possono ricavare i seguenti valori: l'uso di *ministra* nel riferimento alle ministre ha una frequenza relativa del 15% circa, nel *sub-corpus* di *la Repubblica* (vs. il 75% circa di *ministro*), di circa il 18% in quello dell'ANSA (vs. circa il 72% di *ministro*) vs. il 100% nel *sub-corpus* dell'ATS (vs. lo 0% di *ministro*) e poco più dell'80% nel *sub-corpus* di articoli redazionali/ di cronaca dei tre quotidiani ticinesi (vs. il 20% ca. di *ministro*).

Jacqmain (2005) è il primo studio che interroga il web per ricavare la frequenza d'uso di varie modalità ricorrenti per designare le ministre, di seguito elencate con i rispettivi dati di frequenza: *ministra* N=9.130.00, *ministressa* N=64, *signora ministro* N=56, *donna ministro* N=729, *ministro donna* N=357. L'autrice specifica di aver svolto la ricerca a metà novembre del 2005, ma non indaga la data originaria della fonte, così come non verifica i casi in cui le ministre sono designate con una forma maschile, indagine che avrebbe richiesto un'analisi manuale, e, infine, non considera il fatto che *ministra* è una parola presente anche nel vocabolario spagnolo e portoghese, con lo stesso significato che in italiano.

Analisi più articolate e approfondite su *web corpora* sono invece quelle offerte da Nardone (2016, 2018). Nardone (2016) analizza la frequenza d'uso e le collocazioni nel *corpus* itWac (Baroni et al. 2006; Baroni e Kilgarriff 2006), che raccoglie i contenuti divulgati via internet da diverse fonti non solo informative, comunque tutte in lingua italiana, di tre gruppi di nomi femminili. Il primo gruppo comprende i nomi individuati da Robustelli (2014) come quelli con una maggiore resistenza all'uso: *collaboratrice*, *direttrice*, *segretaria*, *architetta*, *chirurga*, *funzionaria*, *ingegnera*, *ispettora*, *magistrata*, *medica*, *notaia*, *procuratrice*, *rettora* e *revisora*. Il secondo gruppo comprende i nomi che, sulla base di una preliminare rassegna bibliografica svolta dall'autrice, risultano i nomi al centro del dibattito sul

sessismo linguistico: *collaboratrice* e *segretaria*. Infine, il terzo cluster include nomi che presentano due forme possibili di femminilizzazione: *dottora* e *dottoressa*, *professora* e *professoressa*, *avvocata* e *avvocatessa*, *sindaca* e *sindachessa*. I risultati della ricerca attestano una forte resistenza all'uso dei femminili appartenenti al primo gruppo di nomi, con una minor frequenza d'uso delle forme femminili rispetto a quelle maschili, e diverse asimmetrie semantiche per alcuni nomi appartenenti al secondo e al terzo gruppo, precisamente *segretaria*, *direttrice*, *collaboratrice*, *dottoressa* e *professoressa*, e i loro corrispettivi maschili. In generale, le forme maschili sono più frequenti di quelle femminili e hanno un campo semantico più esteso. In particolare, i diversi nomi e le loro collocazioni dimostrano specifiche asimmetrie. Per esempio, il nome *segretaria* è spesso collocato di *ufficio*, *redazione*, *impiegata*, *personale*, *lavoro*, che fanno riferimento a un ruolo subalterno e non di prestigio; *segretario*, invece è prevalentemente collocato di parole che fanno riferimento all'ambito sindacale o istituzionale (*CGIL*, *generale*, *confederale*, *nazionale*, *provinciale*, *regionale*, *CISL*, *Amnesty*, *Stato*, *Associazione*). *Direttrice* è collocato di parole come *biblioteca*, *centro*, *museo*, *didattica*, *scuola*, *generale*, *rivista*, *istituto*, *carcere*, *nazionale*, *dipartimento* che rinviano prevalentemente all'ambito culturale e scolastico, mentre *direttore* è presente in collocazioni che rinviano a più ambiti e incarichi professionali: *generale*, *dipartimento*, *istituto*, *responsabile*, *artistico*, *ufficio*, *lavori*, *centro*, *amministrativo*, *tecnico*, *ex*, *regionale*, *incarico*, *agenzia*, *sanitario*, *nuovo*, *nominato*, *orchestra*, *nomina*, *rivista*, *editoriale*, *vice*.

Nardone (2018) fa un'analisi contrastiva tra italiano e tedesco, indagando una lista di nomi professionali nei due *corpora* itWac (Baroni et al. 2006; Baroni e Kilgarriff 2006) e deWac (Ivanova et al. 2008; Kilgarriff et al. 2010; Faaß e Eckart 2013), e in due *corpora* di annunci di lavoro online pubblicati da aziende italiane e tedesche, costruiti manualmente dall'autrice. I risultati mostrano che in italiano alcuni femminili, come *segretaria*, sono caratterizzati da asimmetrie semantiche rispetto ai loro omologhi maschili: mentre questi ultimi indicano ruoli lavorativi di prestigio (es. *segretario di organizzazione sindacale*), i primi designano posizioni più

subordinate (es. *segretaria d'ufficio*). Anche in tedesco alcuni nomi sono interessati da una asimmetria ma di tipo diverso rispetto all'italiano: sono usati al maschile per indicare ruoli o professioni in senso generico, e mai una donna specifica, mentre sono usati al femminile con riferimento esclusivo alle donne. Nardone (2018) conclude che le differenze tra il tedesco e l'italiano sono, probabilmente, ascrivibili a un maggiore impatto delle politiche linguistiche tedesche sull'uso della lingua, meno resistente dell'italiano alla femminilizzazione di nomi di professioni di prestigio.

Formato (2016) analizza un ampio *corpus* di quotidiani (25.446 articoli e 12.249.139 parole) concentrandosi sulle due forme singolari del lemma *ministra/o*, al femminile e al maschile, adottando un approccio diacronico innovativo, con risultati interessanti. L'autrice esamina le forme ricorrenti per nominare i ministri e le ministre di tre governi italiani dal 2012 al 2014 (Governo Monti, Governo Letta, Governo Renzi), analizzando un *corpus* contenenti articoli del *Corriere della sera*, *Il Resto del Carlino*, *La Stampa*. Focalizzandosi sulla congruenza o meno del genere grammaticale con il genere semantico femminile, Formato (2016) attesta che l'uso di *ministro* e *ministra* per riferirsi a una ministra del governo italiano aumenta e diminuisce in modo instabile, nel corso dei tre anni. La forma maschile è la più frequente – con percentuali pari al 91,38% per l'anno 2012, all'88,94% per il 2013 e all'88,72% nel 2014 – anche se non in tutti i giornali: la frequenza di *ministro* è significativamente più alta di quella di *ministra* nei *sub-corpora* del *Corriere della Sera* e *Il Resto del Carlino* ma non nel *sub-corpora* de *La Stampa*. Ciò indica, conclude l'autrice, che l'uso del maschile per i referenti femminili è ancora la norma linguistica più seguita; tuttavia, si sta affermando anche l'uso della forma femminile (Formato 2016, 397). Attraverso una classificazione delle diverse espressioni linguistiche utilizzate nel *corpus* per menzionare le ministre, nominando o meno il loro nome e cognome e/o il ministero che ricoprono, Formato (2016) riporta anche una preferenza, seppur non costante negli anni e tra i giornali, a nominare le ministre utilizzando la forma maschile seguita dal loro nome e cognome. Ciò dimostra, secondo l'autrice, che i giornali conferiscono maggiore ufficialità alla

forma maschile, tendenza confermata anche da un approfondimento sugli aggettivi utilizzati per descrivere ministre e ministri: *ministra* è più frequentemente accompagnato da aggettivi che descrivono caratteristiche personali, come per esempio caratteristiche fisiche o provenienza geografica, mentre *ministro* è più frequentemente accompagnato da attributi professionali, come *buono* e *competente*.

Un recente studio condotto da Giorgia Castenetto e Stefano Ondelli (2020) indaga l'accettabilità dei titoli professionali femminili negli articoli di giornali italiani, attraverso un sondaggio a persone di madrelingua italiana. I risultati mostrano un aumento dell'accettabilità delle forme femminili per professioni e ruoli istituzionali, tra cui il termine *ministra*, che viene approvato da circa 5/6 degli intervistati. Secondo gli autori, ciò è probabilmente dovuto al loro frequente uso nei media. Per esempio, riportano un aumento significativo dell'uso di *ministra* da una media annuale di N=111 occorrenze nel periodo 2000-2010 a N=1.225,25 nel periodo 2015-2018, nell'archivio online del quotidiano *la Repubblica*. Questo lavoro è interessante, in primo luogo perché dimostra l'accettabilità della forma *ministra*, così come di *sindaca* e *avvocata* e, in secondo luogo, perché presuppone, ancorché senza testarlo empiricamente, il ruolo cruciale dei mass media.

Essendo focalizzato sulla sola accettabilità delle forme femminili, lo studio di Castenetto e Ondelli (2020) non approfondisce il ruolo dei media, nel promuovere o meno un linguaggio paritario e inclusivo. Un aspetto toccato invece da Formato (2016) che esamina similarità e differenze fra tre diverse testate giornalistiche. Lo studio di Formato si conclude però con l'anno 2014, poco prima della pubblicazione di quelle che possiamo considerare le linee guida specificamente destinate a professionisti e professioniste dell'informazione, contenute nei volumi *Donne, grammatica e media* (Robustelli 2014) e *Tutt'altro genere d'informazione* (CNOG 2015).

L'ipotesi e le domande di ricerca formulate nel paragrafo 6.3 perseguono l'obiettivo generale di continuare la ricerca sull'uso delle forme femminili per

nominare le donne in posizione apicale, e un obiettivo più specifico che è quello di cercare di capire se e in che misura i mezzi di informazione, a cui sono rivolte specifiche linee guida (Sabatini 1987; Robustelli 2014; CNOG 2015), sostenuti anche da un rinnovato interesse delle istituzioni politiche, in particolare nel corso della 17esima legislatura, possano effettivamente svolgere un ruolo nella diffusione di un linguaggio *gender-fair*.

Ho circoscritto la ricerca al lemma *ministra/o* per tre ragioni: innanzitutto perché la classe politica è, fra le categorie professionali, quella più visibile nei notiziari italiani e il lemma *ministra/o* è risultato il più frequente, fra i paradigmi nominali di ruolo politico apicale, nel *corpus* TG, da cui ha preso avvio la mia analisi (Tabella 28); secondariamente, perché questo lemma è l'unico indagato trasversalmente in tutti gli studi empirici sull'uso delle forme femminili di nomi di professioni e ruoli; infine, perché esso è l'oggetto di indagine di Formato (2016), che, a mio parere, è la più interessante ed innovativa analisi dell'accordo fra genere grammaticale e genere semantico dei nomi di professione e ruolo.

Tabella 28. Frequenza dei nomi di cariche politiche nel *corpus* TG

Titolo	Occorrenza (N)
Ministra/o (e/i)	1.401
Sindaca/o (e/i)	573
Parlamentare (i)	330
Governatrice/tore (trici/tori)	152
Deputata/o (e/i)	128
Senatrice/tore (trici/tori)	120
Sottosegretaria/o (e/i)	45
Assessora/e (e/i)	39
Viceministra/o (e/i)	12
Vicesindaca/o (e/i)	11
Eurodeputata/o (e/i)	11
Europarlamentare (i)	9
Vicegovernatrice/tore	1

6.3 Ipotesi e domande di ricerca

L'ipotesi generale da cui parte l'analisi sull'uso del lemma *ministra/o* nei *corpora* TG e CRS è che esso si caratterizzi per una forma di androcentrismo linguistico. Sulla base di questa ipotesi, ho formulato un set di domande specifiche applicate ai due *corpora* e anche al *corpus* TG ridotto al 2018-2019.

- I. Con quale frequenza il lemma *ministra/o* ricorre nelle quattro forme *ministra*, *ministro*, *ministre*, *ministri*?
 - a. Ci sono differenze significative fra gli anni?
 - b. Ci sono differenze significative fra le testate giornalistiche?
- II. Con quale frequenza il lemma *ministra/o* è declinato in coerenza con il genere semantico?
 - a. Ci sono differenze significative fra gli anni?
 - b. Ci sono differenze significative fra le testate giornalistiche?
- III. Quali espressioni linguistiche vengono utilizzate nel riferimento a ministre e ministri?
- IV. Ci sono differenze o somiglianze fra il *corpus* TG e il *corpus* CRS in termini di frequenza e concordanza tra genere grammaticale e genere semantico del lemma *ministra/o* ed espressioni linguistiche utilizzate nel riferimento a singole/i donne e uomini?

6.4 Metodologia

Al fine di rispondere alle domande di ricerca così formulate, ho adottato un approccio misto fra analisi quantitativa computerizzata e analisi manuale qualitativa. Tramite l'analisi quantitativa ho calcolato le frequenze del lemma *ministra/o*, in tutte le sue quattro forme, femminile singolare, femminile plurale, maschile singolare, maschile plurale, ed effettuato tutti i test statistici utili a valutare la significatività o meno delle comparazioni per anno, testata giornalistica e *corpus*. Per la misurazione delle frequenze nel *corpus* TG ho utilizzato il software IRaMuTeQ, che include fra le varie funzioni operative l'analisi lessicometrica e che avevo già utilizzato per esplorare, anche se con finalità diverse, il *corpus* TG. Per il

corpus CRS ho invece utilizzato Python per descriverne le caratteristiche e AntConc, un software *open source*¹²², per l'analisi delle frequenze. Essendo un *concordancer*, AntConc è uno strumento più agile rispetto a IRaMuTeQ per analizzare occorrenze e collocazioni delle parole in stringhe di testo. Per i test statistici, ho utilizzato in entrambi i casi il test del chi-quadrato, che, come scritto più sopra, è il test statistico più utilizzato per i dati nominali (Capitolo 5, § 5.2.3, Stefanowitsch 2020, 177).

Dopo aver calcolato le frequenze delle quattro forme di *ministra/o*, nei due *corpora*, ho estratto le collocazioni per verificare manualmente, tramite un'analisi qualitativa, sia l'accordo fra genere grammaticale e genere semantico, sia le espressioni utilizzate per nominare le ministre e i ministri. Data la tendenza a nominare le ministre con forme maschili, per stabilire il genere semantico del/la referente è stato necessario analizzare le collocazioni di *ministra/o* valutando sia elementi testuali, di natura sintattica o semantica, sia elementi extra-testuali pertinenti il contesto di riferimento. Per esempio, per decidere che il genere semantico de *il ministro Grillo* è femminile è stato necessario raccogliere informazioni sulla composizione del Governo, o meglio dei diversi governi, a seconda del periodo di riferimento. Questo tipo di analisi manuale, si è resa necessaria anche per indagare le espressioni linguistiche utilizzate per citare le ministre e i ministri. Seguendo il modello sviluppato da Formato, ho messo a punto un sistema di riclassificazione delle collocazioni del lemma *ministra/o* per tipologia di espressione linguistica utilizzata. La Tabella 29 contiene la descrizione delle tipologia di espressione linguistica pre-codificata, per ordine di codice assegnato (Id), più un esempio o due per ciascuna categoria, estratto dal *corpus* CRS. Lo schema di classificazione comprende tipologie indicative dell'espressione linguistica utilizzata per citare ministre e ministri, distinguendo fra utilizzo del titolo generico (*ministra, ministro, ministre, ministri* e basta) oppure titolo specifico, ovvero completo del nome del dicastero di afferenza (es. *ministra della salute*), la presenza o meno di nome e cognome e la presenza o meno di elementi testuali (es.

¹²² Fonte: <https://www.laurenceanthony.net/software.html>.

elementi satellite come articoli o aggettivi) e/o contestuali (es. nome del partito), utili a determinare l'identità di genere della persona di riferimento.

Tabella 29. Sistema di classificazione delle espressioni linguistiche per il riferimento a ministre e ministri

Id	Tipo di espressione linguistica	Esempio di collocazione
0	Non pertinente (titolo usato con significato diverso da quello indagato, quindi non pertinente l'analisi)	Id 2298 famiglia protestante. Anche mio padre era un ministro della Chiesa. In quanto cristiano durante il
1	Titolo generico (senza nessuna specificazione o altra informazione che consenta di definire il genere del referente)	Id 257 «meri auspici politici» che «non legittimavano il ministro a disattendere le Convenzioni internazionali Id 130 “Abbiamo chiesto all'Avvocatura – ha spiegato la ministra – di darci le linee attuative. Appena arriverà
2	Titolo generico più cognome	Id 11 disparità di trattamento» ha però risposto la ministra Bongiorno: «Anpal Servizi non è una pubblica Id 323 prima di sottoporla all'Italia”. Soddisfatto il ministro Alfano. “Ho molto apprezzato le parole del
3	Titolo generico più nome e cognome	Id 134 La mossa a sorpresa con cui la ministra Elisabetta Trenta ha annunciato il ritiro delle Id 3117 fonti della Difesa hanno informato che il ministro Elisabetta Trenta avrebbe dato disposizioni per i Id 327 Con il senno di poi, forse il ministro Alfonso Bonafede non lancerebbe dalla sua pagina
4	Titolo generico in funzione cataforica o anaforica di cognome	Id 202 dall'alleanza la lista di Lorenzin. La ministra minaccia di candidarsi alla presidenza della Regi Id 4602 tra Pd e Calenda. SUL PUNTO il ministro è netto: “Mi colloco nel centrosinistra, e
5	Titolo generico in funzione cataforica o anaforica di nome e cognome	Id 32 di +Europa, Emma Bonino (foto). L'ex ministra dei Radicali italiani ha commentato il sondaggio Id 3112 Danilo Toninelli. La delegazione che incontrerà il ministro ed i competenti funzionari ministeriali, sarà
6	Titolo generico in stringa di testo con informazioni testuali utili a determinare identità e genere referente	Id 1225 essere stata all'epoca dei fatti un ministro del governo di Enrico Letta: «Abbiamo vinto Id 2991 ha suscitato anche reazioni poco urbane: «Un ministro di sinistra e gay che guida la
7	Titolo generico in stringa di testo con informazioni contestuali utili a determinare identità e genere referente	Id 3078 già sindaco di Reggio per due legislature, ministro e ora capogruppo del Pd alla Camera, Id 4709 politico pentastellato, non a caso, riunisce i ministri Cinque Stelle per fare un punto della
9	Titolo specifico (indicazione del ministero ricoperto)	Id 8 in più stanziati dall'ultima manovra La ministra (ama farsi chiamare così) dell'Istruzione corre Id 1006 tedesco. Un tempo, era il posto di ministro degli Esteri, quello considerato più importante d Id 1292 TORNANDO a Leonardo, fin da quando il ministro del Lavoro prospettò il salvataggio pubblico, l'
10	Titolo specifico più cognome	Id 149 la lettera suscita la replica stizzita della ministra francese degli affari europei, Loiseau: “Salvini Id 290 Esteri, Enzo Moavero Milanese, ha incontrato il ministro agli Affari europei francese Loiseau. Loiseau Id 981 il pd Andrea Romano chiede se il ministro degli Esteri Moavero condivide, mentre Matteo Renzi
11	Titolo specifico più nome e cognome	Id 36 governo. C'è chi, come la influente ministra del Lavoro Amber Rudd, una moderata, minaccia Id 359 riflettere sicuramente”. Nel luglio 2008, l'allora ministro all'istruzione Mariastella Gelmini fece sua la

Id	Tipo di espressione linguistica	Esempio di collocazione
		Id 293 il taglio del nastro ufficiale affidato al ministro ai Beni culturali Dario Franceschini. Insieme a
12	Titolo specifico in funzione cataforica o anaforica di cognome	Id 541 democrazie del mondo», nelle parole del primo ministro britannico. May vuole dimostrare di avere ancora Id 1296 optassimo per nuove elezioni», ha detto il ministro del Lavoro uscente. Nahles ha annunciato anche
13	Titolo specifico in funzione cataforica o anaforica di nome e cognome	Id 878 ? Non certo Boris Johnson. Perché l'ex ministro degli Esteri, capofila degli euroscettici e aspira Id 4104 ha sollevato Nunzia De Girolamo. L'ex ministro, poi capogruppo Ncd, tornata in FI a
14	Titolo specifico in stringa di testo con informazioni testuali utili a determinare identità e genere referente	Id 366 che è stata in predicato di diventare ministro alla Farnesina, renziana ma non della prima Id 1621 donna premier del Paese. È stata anche ministro dell'Educazione e degli Affari ecclesiastici oltre
15	Titolo specifico in stringa di testo con informazioni contestuali utili a determinare identità e genere referente	Id 845 60 anni, membro del Partito socialista francese, ministro degli Affari europei dal 1997 al 2002, ministro Id 886 ufficiale del Pd – dice infatti l'ex ministro degli Esteri del governo di Enrico Letta - Id 1171 sabato da Matteo Renzi. Presente anche il ministro dei Trasporti, che domani mattina è a

Dal punto di vista operativo, dopo aver raccolto i dati di frequenza delle quattro forme di *ministra/o* ho esportato le collocazioni in due matrici di Excel, una per ogni *corpus*, con le collocazioni in riga e i metadati (anno e testata giornalistica di riferimento) e le informazioni qualitative (genere e numero della forma linguistica, genere semantico, espressione linguistica) in colonna, in modo da assumere sia i metadati sia le informazioni qualitative come variabili categoriali ed eseguire su entrambi un'analisi di tipo quantitativo. Ho quindi effettuato l'elaborazione statistica dei dati, calcolando le frequenze delle forme linguistiche per genere semantico di riferimento, anche per anno e testata giornalistica, e le frequenze delle diverse tipologie di espressione linguistica.

Per il confronto fra i due *corpora*, ho costruito un *corpus* TG 2018-2019, scorporando dal *corpus* TG i testi delle notizie del gennaio 2020, in modo da fare una comparazione fra due *corpora* di testi prodotti nello stesso periodo. Inoltre ho optato per un confronto qualitativo, piuttosto che statistico, basato sulla discussione dei dati di frequenza relativa. Il *corpus* TG (sia integrale, sia 2018-2019) è infatti un po' più piccolo del *corpus* CRS e anche il numero di occorrenze del lemma *ministra/o* nel primo è inferiore rispetto al secondo. Questa diversa dimensione ha invitato a una certa cautela nella comparazione fra i due *corpora*. Infine, poiché parte dei risultati relativi ai testi dei TG del gennaio 2018 e 2019 sono

già presentati nel paragrafo 6.5, sotto-paragrafi 6.5.1, 6.5.2 e 6.5.3, ho evitato di dedicare un paragrafo specifico ai risultati del *corpus* TG 2018-2019, riportando dati integrativi per la comparazione direttamente nel paragrafo dove discuto i risultati (§ 0), in un sotto-paragrafo dedicato (§ 6.6.3).

6.5 Risultati¹²³

6.5.1 Frequenza delle quattro forme del lemma *ministra/o* nel *corpus* TG

L'analisi lessicometrica del *corpus* TG ha attestato una frequenza del lemma *ministra/o* pari a N=1.400 occorrenze totali, con una netta prevalenza delle forme maschili che attestano N=1.336 occorrenze, pari a una frequenza relativa del 95,43%; vs. le forme femminili che ne attestano N=64, pari a una frequenza relativa del 4,57%. In particolare, risulta elevata la frequenza della forma maschile singolare, con N=1.114 occorrenze (79,57%).

Tabella 30. *Corpus* TG: frequenza delle quattro forme del lemma *ministra/o*

Forma linguistica	Occorrenza (N)	Occorrenza (%)
ministra	63	4,50
ministre	1	0,07
<i>Tot. femminili</i>	64	4,57
ministro	1.114	79,57
ministri	222	15,86
<i>Tot. maschili</i>	1.336	95,43
Totale	1.400	100,00

La comparazione dei risultati per anno, riportata nella Tabella 31, attesta una frequenza delle forme femminili sostanzialmente stabile in termini assoluti, variabile invece in termini di frequenza relativa sul totale delle occorrenze: con una percentuale di femminili sul totale delle occorrenze pari al 7,61% nel gennaio 2018, vs. il 3,23% del 2019 e il 4,36% del 2020. Questo per via di una variazione in termini assoluti delle forme maschili, più numerose nel 2019 e nel 2020, rispetto al 2018.

¹²³ I risultati presentati in questo paragrafo sono già stati parzialmente pubblicati in un articolo della rivista *Problemi dell'informazione* (Azzalini 2021a), dove ho presentato e discusso i dati pertinenti al *corpus* TG limitatamente ai periodi di gennaio 2018 e gennaio 2019.

Tabella 31. *Corpus* TG: frequenza delle quattro forme di *ministra/o* per anno

Forma linguistica	Gennaio 2018		Gennaio 2019		Gennaio 2020	
	N	%	N	%	N	%
ministra	21	7,61	20	3,23	22	4,36
ministre			1	0,16		0,00
<i>Tot. femminili</i>	21	7,61	21	3,39	22	4,36
ministro	212	76,81	503	81,26	399	79,01
ministri	43	15,58	95	15,35	84	16,63
<i>Tot. maschili</i>	255	92,39	598	96,61	483	95,64
Totale	276	100,00	619	100,00	505	100,00

Per verificare se queste differenze sono significative o meno, ho calcolato il chi-quadrato sulle frequenze femminili e maschili osservate nel gennaio 2018, 2019, 2020, assumendo l'anno come variabile indipendente. Il test del chi-quadrato ha restituito un valore $\chi^2(1, N=1.400)=7,89$, $p<0,05$, indicativo di una differenza significativa. La tabella di corrispondenza fra le frequenze osservate e le frequenze attese (Tabella 32) mostra che nel 2018 la frequenza delle forme femminili (singolari e plurali aggregate) è superiore alle attese (+8,38), nel 2019 è inferiore (-7,3) e nel 2020 è sostanzialmente stabile (+1,09).

Tabella 32. *Corpus* TG: frequenze osservate e attese per le forme femminili e maschili del lemma *ministra/o* per anno

Periodo	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Gennaio 2018	21	12,62	255	263,38
Gennaio 2019	21	28,30	598	590,70
Gennaio 2020	22	23,09	483	481,91

Al fine di avere dati utili a una comparazione puntuale fra il *corpus* TG e il *corpus* CRS, che copre solo il 2018 e 2019 ed è sviluppata nel §6.6.3, ho confrontato a due a due le frequenze osservate nei tre periodi: 2018 con 2019, 2019 con 2020, e 2018-2020, calcolando anche in questo caso il chi-quadrato. Il test statistico ha rilevato una varianza significativa solo fra il gennaio 2018 e il gennaio 2019 ($\chi^2(1, N=895)=6,67$, $p<0,05$). La Tabella 33 riporta frequenze osservate e frequenze attese, evidenziando che la frequenza delle forme femminili osservata per il 2018 è superiore a quella attesa, viceversa quella osservata per il 2019 è inferiore al valore atteso.

Tabella 33. *Corpus* TG: frequenze osservate e attese per le forme femminili e maschili del lemma *ministra/o* per anno, solo 2018 e 2019

Periodo	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Gennaio 2018	21	12,95	255	263,05
Gennaio 2019	21	29,05	598	589,95

La comparazione dei dati per testata giornalistica, pur rilevando qualche differenza, sia in termini di frequenze assolute che in termini di frequenze relative, non risulta statisticamente significativa ($\chi^2(1, N=1.400)=2,35, p=0,3$).

Tabella 34. *Corpus* TG: frequenza delle quattro forme di *ministra/o* per TG

Forma linguistica	TG1		TG2		TG5	
	N	%	N	%	N	%
ministra	24	4,7	26	5,3	13	3,3
ministre	1	0,2	0	0	0	0
<i>Tot. femminili</i>	25	4,9	26	5,3	13	3,3
ministro	400	77,7	392	80,7	322	80,7
ministri	90	17,5	68	14	64	16
<i>Tot. maschili</i>	490	95,1	460	94,7	386	96,7
Totale	515	100,0%	486	100,0%	399	100,0%

Avendo scelto di includere il TG2 nel *corpus* TG per verificare se una linea editoriale sensibile alle questioni di genere, anche linguistiche, può incidere sulla frequenza d'uso delle forme femminili (cfr. Capitolo 4, § 4.1), ho costruito una tabella di contingenza per anno e occorrenza delle forme femminili/maschili nel *corpus* del TG2. Sottoposta al test del chi-quadrato, la tabella ha restituito un valore $\chi^2(1, N=486)=3,19, p=0,2$, indicativo di una differenza non significativa.

In sintesi, l'analisi delle occorrenze del lemma *ministra/o* nel *corpus* TG attesta una netta prevalenza di forme maschili rispetto alle forme femminili (95,43% vs. 4,57%), e in particolare della forma maschile singolare (79,57%); una significativa maggior frequenza di forme femminili nel 2018 (7,61%) rispetto al 2019 (3,39%) e 2020 (4,36%); frequenze di forme femminili e maschili diverse ma non significative fra le diverse testate giornalistiche. Un approfondimento sui dati del TG2, selezionato per verificare la possibile incidenza di una linea editoriale sensibile alle questioni di genere sul linguaggio giornalistico (ascrivibile al TG2 del 2018, diretto da una giornalista dichiaratamente impegnata per le pari

opportunità), attesta che non ci sono differenze significative fra la frequenza di forme femminili/maschili nel TG2 del 2018, rispetto al TG2 degli altri anni.

6.5.2 *Accordo fra genere grammaticale e genere semantico nel corpus TG*

Per rispondere alla seconda domanda di ricerca, ho costruito una tabella di contingenza, incrociando le frequenze delle quattro forme di *ministra/o* e le frequenze del genere semantico classificato manualmente, dopo una lettura analitica delle collocazioni del lemma nei frammenti di testo estrapolati dal *corpus* TG tramite IRaMuTeQ. La Tabella 35 illustra i risultati ottenuti, riportando in riga la forma linguistica e in colonna il genere semantico, inclusi i casi di riferimento non determinabile (n.d.), vale a dire casi in cui le informazioni contenute nelle collocazioni esaminate erano insufficienti per stabilire il genere semantico, pur essendo evidente, dalla costruzione sintattica, il riferimento specifico a una persona particolare, e non un riferimento generico al ruolo di *ministra/o*, che ho invece classificato come riferito al genere semantico “Donna e/o Uomo”. La terza colonna della Tabella 35, con l’intestazione “Donna e/o Uomo”, riporta infatti le frequenze sia della forma singolare usata in modo generico, con riferimento alla carica in generale e non a una persona specifica, , dunque potenzialmente donna o uomo, in frasi del tipo *serve un’Europa incisiva e più autorevole ma soprattutto un ministro degli esteri a tempo pieno* (TG2, 13 gennaio 2020), sia della forma plurale riferita a coppie, in frasi del tipo *i ministri Alfano e Pinotti intanto hanno illustrato al Parlamento la missione italiana in Niger* (TG1, 15 gennaio 2018), o gruppi misti comprensivi di ministre e ministri, come in frasi il cui sintagma nominale è per esempio *il consiglio dei ministri*.

Tabella 35. *Corpus* TG: frequenza delle quattro forme di *ministra/o* per genere semantico

Forma linguistica	Referente Donna		Referente Uomo		Referente Donna e/o Uomo		Referenti n.d. ¹²⁴		Totale	
	N	%	N	%	N	%	N	%	N	%
ministra	63	37,72							63	4,50
ministre	1	0,60							1	0,07
<i>Tot. Femminili</i>	64	38,32							64	4,57
ministro	103	61,68	933	97,09	3	1,88	78	66,96	1.114	79,57
ministri			28	2,91	157	98,13	37	33,04	222	15,86
<i>Tot. Maschili</i>	103	61,68	961	100,00	160	100,00	112	100,00	1.336	95,43
Totale	167	100,00	961	100,00	160	100,00	112	100,00	1.400	100,00

I risultati riportati nella Tabella 35 evidenziano un primo dato generale (nella riga del Totale) che riguarda un numero di riferimenti agli uomini superiori ai riferimenti alle donne, indicativa di una maggiore visibilità dei ministri, rispetto alle ministre, nel discorso mediatico. Al netto dei riferimenti ambigenere o non determinabili, i riferimenti alle ministre nel complesso, cioè indipendentemente dal genere grammaticale della forma linguistica utilizzata per designarle, sono infatti N=167 vs. N=961 attestati per gli uomini. In termini percentuali, i riferimenti alle ministre costituiscono il 14,8% vs. l'85,2% dei riferimenti ai ministri, con qualche differenza fra anno e TG, riportata nelle tabelle seguenti.

Tabella 36. *Corpus* TG: distribuzione delle frequenze del lemma *ministra/o* per genere semantico e per anno

Periodo	Referente Donna		Referente Uomo		Totale	
	N	%	N	%	N	%
Gennaio 2018	38	16,74	189	83,26	227	100,00
Gennaio 2019	67	13,32	423	86,68	503	100,00
Gennaio 2020	62	15,58	336	84,42	398	100,00

Tabella 37. *Corpus* TG: distribuzione delle frequenze del lemma *ministra/o* per genere semantico e per TG

TG	Referente Donna		Referente Uomo		Totale	
	N	%	N	%	N	%
TG1	57	14,00	350	14,00	407	100,00
TG2	52	13,27	340	13,27	392	100,00
TG5	58	17,63	271	17,63	329	100,00

Tornando all'accordo fra genere grammaticale e genere semantico, le forme femminili sono attestate con riferimento coerente al genere semantico

¹²⁴ Non determinabile.

femminile con frequenza N=63 al singolare e N=1 al plurale. Per quanto riguarda le forme maschili: la forma singolare *ministro* è utilizzata per nominare sia singole ministre (N=103), sia singoli ministri (N=933) e anche, sebbene raramente, in senso generico con riferimento alla carica in quanto tale con referente che può essere sia donna sia uomo (N=3). È inoltre attestata una frequenza N=78 con riferimento non determinabile, per insufficienza di informazioni nelle collocazioni analizzate. La forma maschile plurale *ministri* attesta una frequenza nulla con riferimento a coppie o gruppi di sole donne, pari a N=28 per indicare gruppi o coppie di soli uomini, pari a N=157, con riferimento a gruppi misti di ministre e ministri e, infine, pari a N=37 con riferimento a casi non determinabili. Nel caso di riferimento a gruppi o coppie di soli uomini, si tratta di un uso di *ministri* specifico e perfettamente congruente con il genere semantico di riferimento, e incide per il 12,61% sulla frequenza totale delle forme maschili al netto delle occorrenze di forme con genere semantico non determinabile (N=185). Nel caso di riferimento a gruppi o coppie miste di ministre e ministri, si tratta di un uso del maschile plurale sovraesteso anche alle ministre, dunque non perfettamente congruente con il genere semantico, e incide per l'84,86% sul totale al netto delle occorrenze con riferimento n.d. (N=185).

Trascurando i casi indeterminabili e le rare forme maschili usate in senso generico (al singolare) o in riferimento a coppie o gruppi misti di ministre e ministri (al plurale), che, concordando con Robustelli (2012) possono considerarsi inclusive, ancorché non senza una ripercussione di tipo socio-culturale, su cui tornerò nel paragrafo successivo, il genere grammaticale è sempre congruente con il genere semantico nella designazione dei ministri (100%), mentre, nel caso delle ministre, lo è nel 38,32% delle occorrenze vs. il 61,68% (sul totale di N=167 occorrenze).

Per una comparazione fra i tre periodi analizzati, funzionale a verificare l'eventuale incidenza del contesto sulle frequenze di accordo e disaccordo fra genere grammaticale e genere semantico femminile – che è l'unico soggetto a variazioni – ho assunto il genere grammaticale come variabile categoriale e l'ho incrociata con la variabile anno, ottenendo così la distribuzione di frequenza delle

forme femminili e maschili, con riferimento a una donna o più donne, per anno (Tabella 36).

Tabella 38. *Corpus* TG: frequenza delle forme femminili e maschili per genere semantico femminile e anno

Periodo	Forme femminili		Forme maschili		Totale	
	N	%	N	%	N	%
Gennaio 2018	21	80,00	17	20,00	38	100,00
Gennaio 2019	21	33,33	46	66,67	67	100,00
Gennaio 2020	22	60,0	40	40,00	62	100,00

Avendo ottenuto una distribuzione delle frequenze variabile, ne ho verificato la significatività, attraverso un test del chi-quadrato. Il test ha riportato un valore significativo: $\chi^2(1, N=167)=6,21, p<0,5$. Ho quindi calcolato le frequenze attese in modo da compararle con le frequenze osservate.

La Tabella 39 evidenzia che nel 2018 la frequenza osservata delle forme femminili è superiore alla frequenza attesa (+6,44), mentre nel 2019 e nel 2020 risulta inferiore (-4,68 e -1,76).

Tabella 39. *Corpus* TG: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per anno

Periodo	Forme femminili		Forme maschili	
	F. osservata (N)	F. attesa (N)	F. osservata (N)	F. attesa (N)
Gennaio 2018	21	14,56	17	23,44
Gennaio 2019	21	25,68	46	41,32
Gennaio 2020	22	23,76	40	38,24

Anche il confronto fra TG, analizzato con la stessa tecnica, rileva una significatività nelle differenze riportate nella Tabella 37 ($\chi^2(1, N=167)=9,95, p<0,05$).

Tabella 40. *Corpus* TG: frequenza delle forme femminili e maschili con genere semantico femminile per TG

TG	Forme femminili		Forme maschili		Totale	
	N	%	N	%	N	%
TG1	25	43,86	32	56,14	57	100,00
TG2	26	50	26	50,00	52	100,00
TG5	13	22,41	45	77,59	58	100,00

Come riportato nella Tabella 41, la frequenza delle forme femminili osservata è superiore a quella attesa per il TG1 (+3,16) e per il TG2 (+6,07), inferiore per il TG5 (-9,23).

Tabella 41. *Corpus* TG: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per TG

TG	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
TG1	25	32	21,84	35,16
TG2	26	26	19,93	32,07
TG5	13	45	22,23	35,77

Poiché approfondire i risultati sul TG2, ho costruito una tabella (Tabella 42) con le frequenze osservate e le frequenze attese nel *sub-corpus* relativo, per anno. I risultati, benché su frequenze che, disaggregate per anno, risultano molto basse, evidenziano un uso della forma femminile (solo singolare, poiché per il TG2 non è stata osservata nessuna occorrenza della forma plurale) significativamente superiore nel 2018, rispetto sia al 2019 sia al 2020 ($\chi^2(1, N=52)=7,13, p>0,5$).

Tabella 42. *Sub-corpus* TG 2: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per anno

Periodo	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Gennaio 2018	8	5,00	2	5,00
Gennaio 2019	9	13,50	18	13,50
Gennaio 2020	9	7,50	6	7,50

In sintesi, l'analisi dell'accordo fra genere grammaticale e genere semantico evidenzia una coerenza totale fra genere grammaticale e genere semantico nel caso di riferimento ai ministri, e una coerenza complessivamente pari al 38,32% con riferimento alle ministre, che nel 61,68% dei casi sono nominate con forme maschili. Le comparazioni per anno e TG evidenziano differenze significative con una maggior frequenza di forme femminili, rispetto a quelle maschili, utilizzate per nominare ministre nel gennaio del 2018 (55,26%) rispetto al gennaio 2019 (31,34%) e 2020 (35,48%) e nel TG2 (43,86%) rispetto al TG1 (43,86) e al TG5 /22,41%). L'approfondimento sul TG2 attesta una frequenza di riferimenti alle ministre con forme femminili significativamente superiore nel 2018 (80%), rispetto al 2019 (33,33%) e 2020 (60%).

6.5.3 *Espressioni linguistiche utilizzate nel riferimento a ministre e ministri nel corpus TG*

Per rispondere alla terza domanda di ricerca ho elaborato, come spiegato nella metodologia, un sistema di classificazione delle collocazioni contenente 15 diverse espressioni linguistiche ricorrenti nel riferimento a ministre e ministri. In fase di elaborazione dei dati, alcune modalità sono risultate poco frequenti, ho quindi proceduto a una aggregazione di alcune categorie coerenti fra loro, mettendo a punto un sistema di classificazione semplificato. La Tabella 43 riporta nelle prime due colonne codice e denominazione delle categorie originali e nelle ultime due colonne codice e denominazione delle categorie utilizzate per presentare i risultati, escludendo la categoria “non pertinente” (Id=0) utilizzata in fase preliminare per escludere le occorrenze del lemma *ministra/o* con significato diverso da membro del Governo (es. *ministro della Chiesa*).

Tabella 43. Tabella di corrispondenza fra il sistema di classificazione delle espressioni linguistiche per il riferimento a ministre e ministri originariamente elaborato e il sistema utilizzato per la presentazione dei risultati

Id	Tipo di espressione linguistica analitica/originaria	Tipo di espressione linguistica sintetica/derivato
1	Titolo generico senza nessuna specificazione o altra informazione che consenta di definire il genere del referente)	Titolo generico
2	Titolo generico più cognome	Titolo generico più cognome
3	Titolo generico più nome e cognome	Titolo generico più nome e cognome
4	Titolo generico in funzione cataforica o anaforica di cognome	Titolo generico in collocazione con elementi utili a determinare il genere semantico
5	Titolo generico in funzione cataforica o anaforica di nome e cognome	
6	Titolo generico in stringa di testo con informazioni testuali utili a determinare identità e genere referente	
7	Titolo generico in stringa di testo con informazioni contestuali utili a determinare identità e genere referente	
9	Titolo specifico (indicazione del ministero ricoperto)	Titolo specifico (indicazione del ministero ricoperto)
10	Titolo specifico più cognome	Titolo specifico più cognome
11	Titolo specifico più nome e cognome	Titolo specifico più nome e cognome
12	Titolo specifico in funzione cataforica o anaforica di cognome	Titolo specifico in collocazione con elementi utili a determinare il genere semantico
13	Titolo specifico in funzione cataforica o anaforica di nome e cognome	

Id	Tipo di espressione linguistica analitica/originaria	Tipo di espressione linguistica sintetica/derivato
14	Titolo specifico in stringa di testo con informazioni testuali utili a determinare identità e genere referente	
15	Titolo specifico in stringa di testo con informazioni contestuali utili a determinare identità e genere referente	

I dati riportati nella Tabella 44 evidenziano che l'espressione più ricorrente per designare le singole ministre è il titolo specifico al maschile più nome e cognome (N=31, 18,67%), a seguire il titolo specifico al maschile più cognome (N=29, 17,47%) – che è il modo più frequente di designare anche i ministri (N=288, 30,97%) – e poi il titolo specifico al femminile più cognome (N=25, 15,06%). Meno frequenti sono i casi di titolo generico al femminile o al maschile più cognome, che è invece la terza espressione linguistica preferita per nominare i ministri.

Tabella 44. *Corpus* TG: distribuzione delle frequenze del tipo di espressione linguistica per genere semantico e genere grammaticale (solo forme singolari)

Genere semantico	Donna						Uomo	
<i>Genere grammaticale</i>	<i>F.</i>		<i>M.</i>		<i>Tot. (F. più M.)</i>		<i>Tot. M.</i>	
Espressione linguistica	N	%	N	%	N	%	N	%
Titolo generico	10	6,02			10	6,02		
Titolo generico più cognome	18	10,84	13	7,83	31	18,67	243	26,05
Titolo generico più nome e cognome			3	1,81	3	1,81	7	0,75
Titolo generico in collocazione con informazioni utili a determinare il genere semantico			7	4,22	7	4,22	99	10,61
Titolo specifico	1	0,60			1	0,60		
Titolo specifico più cognome	25	15,06	29	17,47	54	32,53	288	30,87
Titolo specifico più nome e cognome	7	4,22	31	18,67	38	22,89	98	10,50
Titolo specifico in collocazione con elementi utili a determinare il genere semantico ¹²⁵	2	1,20	20	12,05	22	13,25	198	21,22
Totale	63	37,95	103	62,05	166	100,0	933	100,0

In generale si può notare che fra le strategie che contengono un riferimento nominale, nel caso di riferimento alle ministre, sono più frequenti quelle contengono anche il nome, mentre, nel caso del riferimento ai ministri, prevalgono

¹²⁵ Le occorrenze di questo tipo di espressione attestate nel riferimento femminile sono casi di riprese anaforiche o cataforiche in cui nome e/o cognome non sono la testa del sintagma nominale.

le strategie che contengono solo il cognome senza il nome. Se sommiamo le occorrenze del titolo specifico più nome e cognome a quelle del titolo generico più nome e cognome (le due strategie complete di nome e cognome) otteniamo infatti una percentuale del 24,7% per le ministre e del 11,25% per i ministri; e se sommiamo le occorrenze del titolo specifico più cognome a quelle del titolo generico più cognome (le due strategie che comprendono solo il cognome) otteniamo una percentuale del 51,20% per le ministre vs. il 59,62% per i ministri.

6.5.4 Frequenza delle quattro forme del lemma *ministra/o* nel corpus CRS

La frequenza assoluta del lemma *ministra/o* nel corpus CRS è di N= 5.324 occorrenze. Nella fase di classificazione delle quattro forme (femminile singolare/plurale e maschile singolare/plurale), per genere semantico ed espressione linguistica, sono stati riscontrati 35 casi non pertinenti in quanto riferiti a *i ministri del culto* o a usi metaforici quali *ministro della difesa* con riferimento a difensori nel gioco del calcio. Esclusi questi 35 casi, le frequenze ottenute per ciascuna delle quattro forme grammaticali sono le seguenti:

Tabella 45. Corpus CRS: frequenza delle quattro forme del lemma *ministra/o*

Forma linguistica	Occorrenza (N)	Occorrenza (%)
ministra	251	4,75
ministre	3	0,06
<i>Tot. femminili</i>	254	4,80
ministro	4.340	82,06
ministri	695	13,14
<i>Tot. maschili</i>	5.035	95,20
Totale	5.289	100,00

Come risulta evidente dalla Tabella 45, le forme femminili sono molto meno frequenti di quelle maschili, attestando una frequenza relativa del 4,80% vs. il 95,2% delle forme maschili. In particolare; è molto bassa la frequenza della forma femminile plurale (solo N=3 occorrenze, pari allo 0,06%), viceversa è elevata la frequenza della forma maschile singolare (82,06%).

Il confronto fra i due anni considerati (Tabella 47) dimostra che le flessioni femminili, in generale, senza considerare il genere del/la referente, in proporzione, sono più frequenti nel 2018 rispetto al 2019: 6,53% vs. 3,66%.

Tabella 46. *Corpus* CRS: frequenza delle quattro forme del lemma *ministra/o* per anno

Forma linguistica	Gennaio 2018		Gennaio 2019	
	N	%	N	%
ministra	131	6,53	120	3,66
ministre	2	0,10	1	0,03
<i>Tot. femminili</i>	<i>133</i>	<i>6,63</i>	<i>121</i>	<i>3,69</i>
ministro	1.581	78,77	2.759	84,06
ministri	293	14,60	402	12,25
<i>Tot. maschili</i>	<i>1.874</i>	<i>93,37</i>	<i>3.161</i>	<i>96,31</i>
Totale	2.007	100,00	3.282	100,00

Sottoposta al test del chi-quadrato, la differenza fra i due diversi periodi considerati risulta significativa con un valore χ^2 (1, N=5.289)=22,91, $p<0,001$. La tabella di corrispondenza fra le frequenze osservate e le frequenze attese (Tabella 47) mostra che nel 2018 la frequenza delle forme femminili (singolari e plurali aggregate) è superiore alle attese (+34,62), mentre nel 2019 è inferiore (-36,62).

Tabella 47. *Corpus* CRS: frequenze attese e frequenze osservate per anno

Periodo	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Gennaio 2018	133	98,38	1.874	1.910,62
Gennaio 2019	121	157,62	3.161	3.124,39

Anche la comparazione le diverse testate giornalistiche rileva differenze (Tabella 48), che sottoposte al test del chi-quadrato risultano significative (χ^2 (1, N=5.289)=20,44, $p<0,001$).

Tabella 48. *Corpus* CRS: frequenza delle quattro forme di *ministra/o* per quotidiano

Forma linguistica	Corriere della sera		Il Resto del Carlino		La Stampa	
	N	%	N	%	N	%
ministra	125	6,36	84	4,31	42	3,06
ministre	1	0,05	1	0,05	1	0,07
<i>Tot. femminili</i>	<i>126</i>	<i>6,42</i>	<i>85</i>	<i>4,36</i>	<i>43</i>	<i>3,13</i>
ministro	1.579	80,40%	1.644	84,26	1.117	81,30
ministri	259	13,19	222	11,38	214	15,57
<i>Tot. maschili</i>	<i>1.838</i>	<i>93,58</i>	<i>1.866</i>	<i>95,64</i>	<i>1.374</i>	<i>98,87</i>
Totale	1.964	100,00	1.951	100,00	81,30	100,00

Il confronto fra le frequenze osservate e le frequenze attese delle forme femminili e maschili mostra che la frequenza di quelle femminili è superiore alle attese per il *Corriere della Sera* (+31,68), mentre è inferiore per *Il Resto del Carlino* (-8,70) e per *La Stampa* (-22,99).

Tabella 49. *Corpus* CRS: frequenze attese e frequenze osservate per quotidiano

Quotidiano	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Corriere della Sera	126	94,32	1.838	1.869,68
Il Resto del Carlino	85	93,70	1.866	1.857,30
La Stampa	43	65,99	1.331	1.308,01

Poiché *Il Resto del Carlino* nei due periodi considerati aveva due direttori diversi, e le linee editoriali, in genere stabilite dalla direzione di un organo di informazione, possono essere un fattore incidente sull'uso del linguaggio anche in termini di resistenza o meno all'uso di termini femminili per designare donne in posizioni apicali – come testimoniato dal caso del TG2 più sopra analizzato – ho costruito una tabella di contingenza, con in riga i sei *sub-corpora* e in colonna le forme femminili/maschili, assumendo il genere grammaticale come variabile indipendente. Il test del chi-quadrato ha restituito un $\chi^2(1, N=252)=9,11, p<0,05$ indicativo di una differenza significativa. Ho quindi costruito una tabella riportando le frequenze delle forme femminili e maschili osservate e attese per i sei *sub-corpora* (Tabella 50).

Tabella 50. *Corpus* CRS: frequenze attese e frequenze osservate per quotidiano

	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
CdS 2018	47	32,85	637	651,15
RdC 2018	66	40,77	783	808,23
St 2018	20	22,76	454	451,24
CdS 2019	79	61,47	1.201	1.218,53
RdC 2019	19	52,92	1.083	1.049,08
St 2019	23	43,22	877	856,78

I risultati dimostrano che il *Corriere della Sera* ha una frequenza delle forme femminili superiore alle attese sia nel gennaio 2018 (CdS 2018), sia nel gennaio 2019 (CdS 2019). Mentre *Il Resto del Carlino* attesta una frequenza delle forme femminili superiore alle attese nel 2018, inferiore nel 2019. *La Stampa*, infine,

registra per entrambi gli anni frequenze inferiori alle attese, in misura maggiore nel 2019 rispetto al 2018.

In sintesi, l'analisi delle occorrenze del lemma *ministra/o* nel *corpus* CRS attesta: 1) una diversa frequenza di forme femminili e maschili, con una netta prevalenza di quelle maschili rispetto a quelle femminili (4,80% vs. 95,29%), 2) una significativa maggior frequenza di forme femminili nel 2018 (6,63%) rispetto al 2019 (3,69%) e 3) una significativa maggior frequenza delle forme femminili nel *Corriere della sera* (6,42%) rispetto a *Il Resto del Carlino* (4,36%) e *La Stampa* (3,13%). Un approfondimento sui dati de *Il Resto del Carlino*, che è l'unico quotidiano fra i tre ad avere avuto due diversi direttori nei due periodi considerati, dimostra una frequenza di forme femminili superiori alle attese per il 2018, rispetto al 2019.

6.5.5 Accordo fra genere grammaticale e genere semantico nel corpus CRS

Per analizzare l'accordo fra genere grammaticale e genere semantico nel *corpus* CRS, ho elaborato una tabella di contingenza fra le frequenze delle quattro forme di *ministra/o* e quelle del genere semantico classificato manualmente. La Tabella 51 illustra i risultati ottenuti, riportando in riga la forma linguistica e in colonna il genere semantico, inclusi i casi di riferimento non determinabile (n.d.) e i casi con riferimento semantico ambigenere ("Donna e/o Uomo"), corrispondenti, al singolare a un uso generico del maschile, e al plurale a un uso sovraesteso a coppie o gruppi di ministre e ministri.

Tabella 51. *Corpus* CRS: frequenza delle quattro forme di *ministra/o* per genere semantico

Forma linguistica	Referente Donna		Referente Uomo		Referente Donna e/o Uomo		Referenti n.d.		Totale	
	N	%	N	%	N	%	N	%	N	%
ministra	251	43,43							251	4,75
ministre	3	0,52							3	0,06
<i>Tot. Femminili</i>	254	43,94							254	4,80
ministro	324	56,06	2.830	97,69	36	8,16	1.150	83,76	4.340	82,06
ministri			67	2,31	405	91,84	223	16,24	695	13,14
<i>Tot. Maschili</i>	324	56,06	2.897	100,0	441	100,0	1.373	100,0	5.035	95,20
Totale	578	100,0	2.897	100,0	441	100,0	1.373	100,0	5.289	100,0

Al netto dei riferimenti ambigenere o non determinabili, i riferimenti complessivamente osservati per le ministre, indipendentemente dalla forma linguistica femminile o maschile, sono N=578 vs. N=2.897 attestati per i ministri. In termini percentuali 16,63% vs. 83,37%, con valori leggermente variabili per anno e testata giornalistica, come riportato nelle tabelle seguenti.

Tabella 52. *Corpus* CRS: distribuzione delle frequenze del lemma *ministra/o* per genere semantico e per anno

Periodo	Referente Donna		Referente Uomo		Totale	
	N	%	N	%	N	%
Gennaio 2018	299	21,65	1.082	78,35	1.381	100,00
Gennaio 2019	279	13,32	1.815	86,68	2.094	100,00

Tabella 53. *Corpus* CRS: distribuzione delle frequenze del lemma *ministra/o* per genere semantico e per quotidiano

Quotidiano	Referente Donna		Referente Uomo		Totale	
	N	%	N	%	N	%
Corriere della sera	230	18,24	1031	81,76	1261	100,00
Il Resto del Carlino	252	18,33	1123	81,67	1375	100,00
La Stampa	96	11,44	743	88,56	839	100,00

Tornando all'accordo fra genere grammaticale e genere semantico, tutti i riferimenti ai ministri (N=2.897) sono congruenti con l'identità maschile. Su N=578 riferimenti all'identità di genere femminile, N=254 sono espressi con forme femminili, N=251 al singolare e N=3 al plurale, mentre N=324 sono espressi con una forma maschile, solo al singolare. Non sono stati attestati riferimenti a coppie o gruppi di ministre con la forma maschile plurale.

La Tabella 51 riporta anche i dati a riferimenti ambigenere, espressi solo con una forma maschile: la forma singolare *ministro* con valore generico, riferito alla carica senza alcun preciso riferimento a una persona ricoprente la stessa, ha una frequenza assoluta di N=36 occorrenze, che incide sul totale della frequenza delle forme maschili per lo 0,83%; la forma plurale *ministri* con valore sovraesteso anche alle donne ha una frequenza maggiore, N=405 occorrenze, pari al 58,27% sul totale (N=695). Trascurando i riferimenti ambigenere e i casi indeterminabili, il genere grammaticale è sempre coerente con il genere semantico maschile, mentre è

coerente con il genere semantico femminile nel 43,94% delle occorrenze vs. il 56,06% di occorrenze di forme femminili per designare le ministre.

Il confronto per anno attesta differenze minime e non significative (χ^2 (1, N=578)=0,03, p=0,85)

Tabella 54. *Corpus* CRS: frequenze delle forme femminili e maschili con genere semantico femminile e anno

Periodo	Forme femminili		Forme maschili		Totale	
	N	%	N	%	N	%
Gennaio 2018	133	44,48	166	55,52	299	100,00
Gennaio 2019	121	43,37	158	56,63	279	100,00

La comparazione per quotidiano rileva invece variazioni significative fra le diverse testate giornalistiche (χ^2 (1, N=578)=21,83, p<0,001).

Tabella 55. *Corpus* CRS: frequenza delle forme femminili e maschili con genere semantico femminile per quotidiano

Quotidiano	Forme femminili		Forme maschili		Totale	
	N	%	N	%	N	%
Corriere della sera	126	54,78	104	45,22	254	100,0
Il Resto del Carlino	85	33,73	167	66,27	324	100,0
La Stampa	43	44,79	53	55,21	578	100,0

La Tabella 56 riporta le frequenze attestate e attese delle forme femminili e maschili con riferimento semantico femminile per testata giornalistica.

Tabella 56. *Corpus* RCS: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per quotidiano

Quotidiano	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Corriere della sera	126	101,07	104	128,93
Il Resto del Carlino	85	110,74	167	141,26
La Stampa	43	42,19	53	53,81

Il *Corriere della Sera* attesta una frequenza di accordi coerenti con il genere semantico femminile superiore alle attese (+24,93), *Il Resto del Carlino* inferiore alle attese (-25,74), *La Stampa* infine riporta frequenze osservate e attese sostanzialmente simili (+0,81). Poiché *Il Resto del Carlino* è l'unico, fra i tre quotidiani considerati, ad aver due direttori diversi nei due periodi, al fine di verificare se il cambio di direzione ha inciso o meno sull'uso delle forme femminili

ho costruito una tabella di contingenza (periodo per forma femminile/maschile), sottoponendola al test del chi-quadrato. Il test statistico ha restituito una $p\text{-value} < 0,05$, indicativa di differenze significative per anno, nel *corpus* di articoli.

I dati nella Tabella 57 riportano frequenze osservate e frequenze attese delle forme femminili e maschili con riferimento femminile nel sub-*corpus* de *Il Resto del Carlino*, evidenziano una frequenza di forme femminili osservata superiore a quella attesa per il gennaio 2018 (+11,36).

Tabella 57. *Sub-corpus Il Resto del Carlino*: frequenze osservate e attese per le forme femminili e maschili con genere semantico femminile per anno

Periodo	Forme femminili		Forme maschili	
	F. osservate (N)	F. attese (N)	F. osservate (N)	F. attese (N)
Gennaio 2018	66	54,64	96	107,36
Gennaio 2019	19	30,36	71	59,64

In sintesi, l'analisi dell'accordo fra genere grammaticale e genere semantico evidenzia una coerenza totale fra genere grammaticale e genere semantico nel caso di riferimento ai ministri, parziale nel caso di riferimento alle ministre per designare le quali nel 43,94% è usata una forma femminile, nel 56,06% una forma maschile. Le comparazioni fra anno e testata giornalistica rilevano differenze minime e non significative per anno, e variazioni più marcate e significative per quotidiano: il *Corriere della sera* attesta una maggior frequenza d'uso di forme femminili (54,78%) rispetto a *Il Resto del Carlino* (33,73%) e *La Stampa* (44,79%). Un approfondimento su *Il Resto del Carlino* che, come il TG2, ha due direttori diversi nei due periodi considerati attesta un significativa maggior frequenza di forme femminili per nominare le ministre nel 2018, rispetto al 2019.

6.5.6 *Espressioni linguistiche utilizzate nel riferimento a ministre e ministri nel corpus CRS*

Come nell'analisi del *corpus* TG, per la presentazione di risultati sulle espressioni linguistiche utilizzate nel riferimento a ministre e ministri, ho semplificato il sistema di classificazione originariamente elaborato per classificare le espressioni linguistiche utilizzate nel riferimento a ministre e ministri, poiché alcune modalità

previste inizialmente sono risultate poco frequenti. Non ho creato un nuovo schema specifico per il *corpus* CRS, ma ho utilizzato lo stesso messo a punto per l'analisi del *corpus* TG, sia per consentire la comparazione puntuale fra i due *corpora*, sia perché le espressioni linguistiche che conveniva aggregare fra loro, in quanto poco frequenti, erano le stesse nei due *corpora*. I risultati riportati nella Tabella 58 evidenziano che il modo più frequente per nominare le ministre è il titolo specifico al maschile più nome e cognome (N=137, 23,83%), come per i ministri (N=1.211, 42,79%), seguito dal titolo declinato al femminile più nome e cognome (N=113, 19,65%).

Tabella 58. *Corpus* CRS: distribuzione delle frequenze del tipo di espressione linguistica per genere semantico e genere grammaticale

Genere semantico	Donna						Uomo	
	F.		M.		Tot. (F. più M.)		Tot. M.	
Genere grammaticale	N	%	N	%	N	%	N	%
Espressione linguistica	N	%	N	%	N	%	N	%
Titolo generico	32	5,57			32	5,57		
Titolo generico più cognome	35	6,09	63	10,96	98	17,04	563	19,89
Titolo generico più nome e cognome	32	5,57	55	9,57	87	15,13	415	14,66
Titolo generico in collocazione con informazioni utili a determinare il genere semantico			8	1,39	24	4,17	65	2,30
Titolo specifico	31	5,39			31	5,39		
Titolo specifico più cognome	8	6,09	16	2,78	24	4,17	217	7,67
Titolo specifico più nome e cognome	113	19,65	137	23,83	250	43,48	1.211	42,79
Titolo specifico in collocazione con informazioni utili a determinare il genere semantico			45	7,83	45	7,83	359	12,69
Totale	251	43,65	324	56,35	575	100,0	2.830	100,0

In generale, si può notare che fra le strategie che contengono un riferimento nominale prevalgono quelle che contengono nome e cognome, rispetto a quelle che contengono solo il cognome, in misura abbastanza simile per ministre e ministri. Se sommiamo le occorrenze del titolo specifico più nome e cognome a quelle del titolo generico più nome e cognome (le due strategie complete di nome e cognome) otteniamo infatti una percentuale del 21,21% per le ministre e del 27,56% per i ministri; e se sommiamo le occorrenze del titolo specifico più cognome a quelle del titolo generico più cognome (le due strategie

che comprendono solo il cognome) otteniamo una percentuale del 58,61% per le ministre vs. il 57,45% per i ministri.

6.6 Discussione dei risultati

6.6.1 Comportamento delle quattro forme di ministra/o nel corpus TG

Il primo risultato sulla frequenza d'uso sulle quattro forme del lemma *ministra/o* nel *corpus* TG evidenzia un'elevata prevalenza delle forme maschili rispetto a quelle femminili (95,43% vs. 4,57%), indicativa di un androcentrismo linguistico caratterizzato soprattutto dall'alta frequenza di forme maschili singolari (79,57%). La comparazione dei dati per anno attesta una frequenza di forme femminili sostanzialmente invariata in termini assoluti, ma non in termini relativi, per via del contestuale aumento delle forme maschili, nel 2019 e 2020 rispetto al 2018. Il calcolo delle occorrenze delle forme femminili ponderate sul totale delle occorrenze, restituisce infatti una percentuale di forme femminili sensibilmente più elevata nel gennaio 2018 rispetto al gennaio 2019 (7,61% vs. 3,23%; $\chi^2(1, N=1.400)=7,89, p<0,05$).

La comparazione della distribuzione delle frequenze fra forme femminili e maschili per TG risulta statisticamente non significativa, indicando una non incidenza delle diverse linee editoriali dei notiziari sulla frequenza d'uso complessiva delle forme femminili e maschili. L'approfondimento sul TG2, finalizzato a verificare l'ipotesi di un'incidenza della linea editoriale di Ida Colucci, direttrice del notiziario nel gennaio 2018 e dichiaratamente sensibile alle questioni della parità di genere, sull'uso delle forme femminili, non ha prodotto risultati significativi. In sintesi, l'ipotesi che una linea editoriale sensibile alle questioni di genere incida sull'incremento di uso delle forme femminili in questo caso non è confermata.

Confrontando questo primo risultato con il contesto politico di riferimento, ovvero la composizione dei tre diversi Governi in carica nei tre diversi periodi

considerati, che sono il contesto extra-linguistico di riferimento prevalente (85,48%¹²⁶), ancorché non esclusivo, è evidente che la sproporzione fra forme femminili e forme maschili attestata nel *corpus* TG non può considerarsi solamente il riflesso di una prevalenza maschile nella formazione, o meglio nelle formazioni governative. Si tratta infatti di una disparità di proporzioni maggiore rispetto al *gender-gap* nei Governi italiani in carica nei mesi di gennaio 2018, 2019 e 2020. Nel gennaio 2018, le ministre al Governo erano 5 su 19 (26,32%), nel gennaio 2019, 5 su 20 (25%), e nel gennaio 2020, 8 su 23 (34,78%). Se calcoliamo la media della presenza femminile nei Governi italiani nei tre periodi, otteniamo una percentuale del 29,03%, ben più elevata rispetto al 4,57% di frequenza d'uso delle forme femminili del lemma *ministra/o*. Inoltre, non c'è corrispondenza fra il periodo di maggior presenza delle donne nel Governo italiano (gennaio 2020) e il periodo di maggiore visibilità linguistica nel *corpus* TG (gennaio 2018).

Per comprendere meglio questo risultato bisogna considerare altri due fattori: la sotto-rappresentazione mediatica delle ministre rispetto ai ministri, e la tendenza a designare le ministre prevalentemente, anche se non esclusivamente, con il titolo declinato al maschile. Entrambi questi fattori emergono dall'analisi dei risultati sull'accordo fra genere grammaticale e genere semantico.

La percentuale di riferimenti alle ministre, indipendentemente dalla forma linguistica utilizzata, nel *corpus* TG, è pari al 14,8%, un valore inferiore alla percentuale dei riferimenti ai ministri (il restante 85,2%), e anche alla presenza femminile nei tre Governi di riferimento (mediamente pari al 29,3%). L'androcentrismo linguistico del *corpus* TG è dunque anzitutto la conseguenza di un'informazione in cui le ministre fanno meno notizia dei ministri, sotto-rappresentandole sia rispetto ai colleghi sia rispetto alla realtà, ovvero la loro effettiva incidenza nelle varie compagini governative. Ma è anche il risultato di una tendenza a nominare le ministre utilizzando prevalentemente forme maschili. I dati riportati nel paragrafo precedente attestano infatti che le ministre vengono

¹²⁶ Totale N=1.288.

designate con il titolo declinato al femminile nel 38,32% dei casi, per il restante 61,68% di casi la loro identità di genere rimane celata sotto una designazione maschile.

Venendo ora alla nomina di coppie o gruppi di genere eterogeneo, il *corpus* TG attesta N=157 occorrenze della forma maschile *ministri* sovra-estesa anche alle ministre. Questo uso può essere considerato in qualche modo inclusivo e non discriminante, perché in genere il contesto permette di distinguere fra riferimento a un gruppo ambigenere o solo maschile, come discusso nel Capitolo 2 (§ 2.2.2). Tuttavia non è privo di problematicità, perché, almeno simbolicamente, oscura la presenza femminile e, come attestato dai risultati presentati nel paragrafo precedente, ha una frequenza relativa elevata: su un totale di N=222 forme maschili plurali il 70,72% ha come riferimento una coppia o un gruppo misto, il cui genere rimane ambiguo, quanto meno a livello linguistico.

L'alternativa più efficace per rendere visibile la presenza delle donne in gruppi misti è sicuramente l'uso di forme sdoppiate estese, come per es. *la ministra e il ministro, le ministre e i ministri*. D'altro canto come argomenta Robustelli (2012), l'uso delle forme sdoppiate, è una pratica difficile da utilizzare per via dell'appesantimento di testi e delle difficoltà nel governare gli accordi con articoli, aggettivi, participi passati e pronomi. Nel caso specifico di *ministra/o* un'alternativa praticabile potrebbe essere utilizzare il termine *governo*, per lo meno nei casi in cui il riferimento è al consiglio dei ministri. Per quanto i notiziari siano sempre corredati da immagini che possono visualizzare i gruppi di persone, e mostrare anche volti femminili, è pur vero che la pratica attestata di usare il maschile in riferimento a gruppi di uomini e donne contribuisce a mantenere l'invisibilità linguistica delle donne, che già faticano a essere equamente rappresentate in uno spazio mediatico che, come testimoniato anche dai risultati di questa analisi, è ancora occupato prevalentemente da uomini.

La comparazione per anno, circoscritta al genere semantico femminile, che è l'unico a presentare variazioni d'accordo, rileva che nel gennaio 2018 la frequenza

delle forme femminili osservate è sensibilmente superiore a quelle attese, viceversa nel gennaio 2019 e 2020 le frequenze osservate sono inferiori a quelle attese. Questo risultato sebbene circoscritto a un numero di frequenze basso, che richiede una certa cautela interpretativa, può essere spiegato alla luce del mutato contesto politico-istituzionale, che, con l'inizio della 18esima legislatura (23 marzo 2018) ha perso una figura come la presidente della Camera Laura Boldrini, politica molto attiva nel contrasto alle discriminazioni di genere, incluse quelle linguistiche e questo potrebbe avere influito sul linguaggio giornalistico. È un'ipotesi che può essere avanzata partendo dalle suggestioni offerte dal direttore dell'ANSA Sergio Lepri, nel suo contributo al volume di Alma Sabatini del 1987. Suggestioni che possono essere così sintetizzate: giornaliste e giornalisti appartengono alle classi dirigenti del paese e difficilmente possono innovare il loro linguaggio con forme non gradite agli ambienti a cui appartengono, specie di fronte a richieste specifiche come, per esempio, quelle di una presidente che chiede di essere chiamata *il presidente* o una senatrice che si presenta come *senatore* e pretende di essere chiamata così dai giornalisti (Lepri 1987, 22). Accogliendo le suggestioni di Lepri è ragionevole ipotizzare che la presa di posizione pubblica della presidente Laura Boldrini a favore dell'uso delle forme femminili abbia trasmesso l'idea a giornaliste e giornalisti che queste forme sono gradite alla classe dirigente, o per lo meno, a una parte di essa.

Anche il confronto per TG, sempre circoscritto all'accordo fra genere grammaticale e genere semantico femminile, rileva differenze significative, attestando una frequenza di forme femminili osservate superiori alle attese per il TG1 e per il TG2, che risulta il notiziario più virtuoso da questo punto di vista. L'approfondimento specifico sul TG2, inserito nel *corpus* TG per verificare l'ipotesi dell'incidenza di una direzione editoriale sensibile alle questioni di genere sull'uso della forma femminile per nominare le donne, o per lo meno le ministre, essendo stato diretto da una giornalista dichiaratamente sensibile alle questioni di genere, nel gennaio 2018, ma non nel 2019 e nel 2020, ha evidenziato un uso della forma femminile (solo singolare, poiché per il TG2 non è stata osservata nessuna

occorrenza della forma plurale) significativamente superiore nel 2018, rispetto sia al 2019 sia al 2020. Risultato che non incide sul complessivo androcentrismo del TG2, che in termini di frequenza di forme femminili e maschili, indipendentemente dal genere semantico di riferimento, non si distingue significativamente dagli altri TG, come abbiamo visto nei risultati commentati nel paragrafo precedente. Questo può dipendere dal fatto che l'androcentrismo linguistico non è solo il risultato di una tendenza all'uso di forme maschili per designare le ministre, ma anche l'effetto di una loro generale sotto-rappresentazione mediatica. Per verificare se sia questo il caso del TG2, ho calcolato l'incidenza dei riferimenti alle ministre, indipendentemente dalla forma linguistica utilizzata, sul totale dei riferimenti a ministre e ministri, al netto dei casi di riferimento ambigenere o non determinabile. Questo calcolo ci dà una misura della visibilità mediatica delle ministre (rispetto ai ministri), attraverso le frequenze del genere semantico femminile (rispetto al maschile). I risultati in effetti attestano una visibilità mediatica delle ministre nel TG2 pari al 13,27%, valore inferiore al 14% attestato dal TG1 e al 17,63% attestato dal TG5. Dunque, il TG2 attesta un elevato androcentrismo linguistico per via di una minore attenzione generale alle ministre (vs. i ministri) rispetto agli altri notiziari, piuttosto che per una tendenza a nominarle al maschile, tendenza sensibilmente inferiore rispetto agli altri TG.

Per quanto riguarda le espressioni linguistiche, il modo più frequente per designare le ministre è il titolo specifico al maschile più nome e cognome, in linea con quanto riportato da Formato (2016, 393): una strategia discorsiva che potrebbe essere considerata, come suggerisce Formato (2016), connotata da maggiore ufficialità entro un discorso linguistico che continua a conferire maggiore autorevolezza alle forme maschili, che in effetti sono la maggior parte. Tuttavia in questo caso bisogna prendere in considerazione anche un'altra ipotesi, non avanzata da Formato: la citazione del nome proprio permette di disambiguare il riferimento, esplicitando l'identità femminile della referente, senza rinunciare, al contempo, alla declinazione maschile del titolo. In effetti, nei casi in cui il titolo è declinato al femminile, le strategie linguistiche più frequenti sono il titolo generico

più cognome oppure il titolo specifico più cognome, ovvero strategie che non includono la citazione del nome proprio, esplicitando l'identità di genere delle referenti attraverso la declinazione femminile del titolo. Contribuisce a rafforzare questa ipotesi, la contestuale attestazione di una più bassa frequenza di strategie che includono il nome proprio per i ministri rispetto alle ministre (11,25% vs. 24,7%).

6.6.2 *Comportamento delle quattro forme di ministra/o nel corpus CRS*

Il primo risultato dell'analisi sulla frequenza delle quattro forme del lemma *ministra/o* nel *corpus* CRS è un'elevata sproporzione fra le forme maschili e quelle femminili: 95,20% vs. il 4,80%, una *ratio* di 1 forma femminile ogni circa 20 maschili, che attesta un androcentrismo linguistico, a delineare il quale concorre soprattutto l'elevata frequenza di forme maschili singolari (82,06%). I risultati dell'analisi comparativa fra il 2018 e il 2019 non dimostrano differenze significative per anno. Differenze significative sono invece evidenti per testata giornalistica: il *Corriere della Sera* attesta una frequenza di forme femminili superiore alle attese, sia nel 2018 sia nel 2019; *Il Resto del Carlino* attesta frequenze femminili superiori alle attese per il 2018, ma non per il 2019; *La Stampa* attesta frequenze inferiori alle attese per entrambi gli anni, minime nel 2018, più marcate nel 2019. L'approfondimento sul *corpus* de *Il Resto del Carlino*, che è l'unico quotidiano ad avere due direttori diversi nei due periodi considerati, ancorché né l'uno né l'altro dichiaratamente impegnati sul fronte delle pari opportunità (perlomeno non ne ho trovata attestazione pubblica), evidenzia differenze significative per anno, attestando una frequenza di forme femminili significativamente superiore a quella attesa nel 2018 rispetto al 2019, a differenza del *Corriere della sera* e *La Stampa* che attestano, in modo stabile nei due anni, l'uno frequenze femminili superiori, l'altro frequenze femminili inferiori alle attese.

Come nel caso del *corpus* TG, questi risultati non possono essere interpretati soltanto come il riflesso di una disparità realmente esistente nel contesto di riferimento, che è il contesto extra-linguistico prevalentemente nel

corpus CRS (83%¹²⁷) ed è rappresentato dai Governi Gentiloni e Conte I, in carica rispettivamente, nel gennaio 2018 e nel gennaio 2019. La presenza delle donne nei Governi italiani nei due periodi considerati è infatti uguale e pari a 5, con una differenza minima in termini di incidenza sull'intera compagine governativa (25,64% vs. 26,3%)

Come per il *corpus* TG, i risultati sull'accordo fra genere grammaticale e genere semantico evidenziano due fattori che giocano un ruolo cruciale nel connotare il linguaggio del *corpus* CRS come androcentrico: da un lato, la sotto-rappresentazione mediatica delle ministre, rispetto ai colleghi maschi, dall'altro il loro parziale occultamento linguistico. La sotto-rappresentazione mediatica delle ministre, rispetto sia i ministri sia alla loro incidenza nella composizione dei due Governi di riferimento, è attestata da una frequenza di riferimenti alle ministre complessivamente, cioè indipendentemente dalla forma linguistica femminile o maschile, pari soltanto al 16,63% del totale. Il loro oscuramento linguistico è indicato dalla prevalenza delle forme maschili nel loro riferimento. I risultati sull'accordo fra genere grammaticale e genere semantico attestano, infatti, come prevedibile, un accordo di genere grammaticale sempre coerente con il genere semantico nel caso di referenti uomini. Mentre, nel caso di referenti donne, indicano una prevalenza della forma maschile nel 56,06% dei casi vs. il 43,94% di forme femminili. L'androcentrismo linguistico del *corpus* CRS è dunque il risultato di almeno tre fattori concomitanti: 1) la scarsa presenza delle donne nelle posizioni apicali della politica nazionale, ovvero nei Governi italiani degli anni più recenti; 2) la loro sotto-rappresentazione mediatica; 3) la tendenza a nominarle ancora preferibilmente con forme maschili.

Venendo ora all'uso del maschile generico, le frequenze complessivamente osservate nel *corpus* CRS sono N=36, con una scarsa incidenza sul totale delle forme maschili singolari occorrenti: 0,83%. Mentre l'uso del maschile sovraesteso è più frequente, attestando N=405 occorrenze pari al 58,27% della frequenza totale delle

¹²⁷ Totale N=3.891.

occorrenze di *ministri* (N=695). Una frequenza relativamente elevata che pone i medesimi problemi di nascondimento parziale delle ministre, aggravato dal fatto che, la carta stampata non può supplire come la televisione con un efficace uso delle immagini per esplicitare le identità di genere. Anche se spesso gli articoli sono corredati da fotografie, non sempre le fotografie ritraggono tutte le persone citati.

I confronti per anno e per testata giornalistica, limitatamente alle forme con genere semantico femminile, che sono le uniche a presentare oscillazioni fra, attestano differenze significative per testata giornalistica ma non per anno. Quest'ultimo risultato è particolarmente interessante se confrontato con il risultato relativo alla significativa maggior frequenza delle forme femminili nel 2018 rispetto al 2019. Alla significativa maggior frequenza d'uso delle forme femminili nel 2018 non corrisponde una significativa maggior frequenza di accordi coerenti con il genere semantico femminile, nello stesso anno. Questo perché le ministre nel 2018 sono state nel complesso più citate, costituendo il 21,65% dei riferimenti a singoli ministri/e vs. il 13,32% del 2019 (Tabella 52), ma non significativamente più citate con forme coerenti rispetto al loro genere. Se infatti si calcola la frequenza relativa delle forme femminili sul totale delle ministre citate, la percentuale che si ottiene è simile per i due anni, pari al 44,48% nel 2018 e al 43,37% nel 2019 (cfr. Tabella 54).

Quanto alle differenze fra quotidiani, i risultati evidenziano che il *Corriere della Sera* non solo è la testata giornalistica per cui si osserva una significativa maggior frequenza d'uso delle forme femminili, in generale, ma anche una maggior frequenza delle forme femminili rispetto a quelle maschili nel riferimento al genere semantico femminile. *Il Resto del Carlino*, che è l'unico quotidiano ad aver cambiato direzione nei due periodi considerati, è anche l'unico che attesta una variazione di frequenza di forme femminili accordate in modo coerente con il genere semantico femminile nei due anni: in linea con i risultati sulla distribuzione delle forme femminili e maschili, indipendentemente dal genere semantico di riferimento, superiore alle attese nel 2018, rispetto al 2019. Infine *La Stampa* attesta una congruenza fra genere grammaticale femminile e genere semantico femminile in

linea con le attese, a fronte di una significativa maggior frequenza di forme maschili rispetto a quelle femminili, ascrivibile a una maggior presenza dei ministri nel discorso giornalistico (88,56% vs. 11,44%).

In linea con quanto dimostrato da Formato (2016, 393) e quanto riportato nell'analisi del *corpus* TG, il modo più frequente per citare le ministre è il titolo specifico (completo del nome del ministero ricoperto) declinato al maschile più nome e cognome (23,83%). A differenza di quanto attestato da Formato (2016), il grado di prevalenza di questa strategia linguistica non sembra però imputabile soltanto al maggior prestigio attribuito al titolo declinato al maschile e, neanche soltanto, a differenza di quanto rilevato per il *corpus* TG, alla funzione di rendere visibile l'identità di genere senza rinunciare alla declinazione maschile del titolo, perché per il *corpus* CRS è stata attestata un'elevata frequenza anche di titoli declinati al femminile più nome e cognome (19,65% vs. 4,11% attestato per il *corpus* TG. Quello che sembra emergere in questo caso è una preferenza della stampa a citare sia le ministre sia i ministri con nome e cognome, oscillando poi nella scelta del titolo, generico o specifico, e nella declinazione al femminile e al maschile per le donna: l'incidenza complessiva dei riferimenti alle ministre con nome e cognome e titolo specifico, con forme femminili o maschili, sul totale dei riferimenti è pari al 43,48% una percentuale del tutto in linea a quella maschile pari al 42,79% (Tabella 58). Inoltre, l'analisi delle strategie linguistiche utilizzate per nominare le ministre nel *corpus* CRS evidenzia una netta prevalenza di quelle comprensive di nome e cognome sia per le ministre (58,61%), sia per i ministri (57,45%).

6.6.3 *Comportamento delle quattro forme di ministra/o nei corpora TG e CRS a confronto*

Il *corpus* TG e il *corpus* CRS sono entrambi collezioni di testi giornalistici e hanno dimensioni simili fra loro, attestando N=1.434.733 occorrenze il primo, N=1.835.938 il secondo. Tuttavia si riferiscono a due periodi parzialmente diversi, poiché il *corpus* TG comprende non solo testi di notizie del gennaio 2018 e gennaio

2019, ma anche del 2020. Per questo circoscriverò, come anticipato nel paragrafo 6.4, la comparazione al biennio 2018-2019, che è comunque risultato il periodo più interessato da differenze significative nell'analisi del *corpus* TG. Inoltre, la selezione delle quattro forme del lemma *ministra/o*, a partire dalla quale ho costruito la base dati su cui svolgere l'analisi quantitativa, ha restituito due campioni di dimensioni diverse, attestando N=1.400 occorrenze totali per il *corpus* TG nel suo complesso e N=5.289 per il *corpus* CRS. Lo scorporo dei testi datati gennaio 2020 dal *corpus* TG, ne ha ridotto ulteriormente la base dati, riducendo a N=895 le occorrenze totali delle quattro forme del lemma *ministra/o*, che risultano così meno di un quinto rispetto alle occorrenze attestate per il *corpus* CRS. Per questo ho optato per un confronto prevalentemente qualitativo, focalizzato sulle evidenze di similarità e differenze più interessanti e supportato dai risultati in valori percentuali, in modo da ponderare le frequenze assolute osservate nei due campioni con le diverse dimensioni degli stessi.

Il calcolo delle frequenze delle quattro forme del lemma *ministra/o* sul *corpus* TG 2018-19 attesta valori percentuali molto simili a quelli ottenuti per il *corpus* TG integrale e a quella ottenuta per il *corpus* CRS.

Tabella 59. *Corpus* TG 2018-19: frequenza delle quattro forme del lemma *ministra/o*

Forma linguistica	Occorrenza (N)	Occorrenza (%)
ministra	1	0,11
ministre	41	4,58
<i>Tot. femminili</i>	42	4,69
ministro	138	15,42
ministri	715	79,89
<i>Tot. maschili</i>	853	95,31
Totale	895	100,00

La frequenza delle forme femminili nel *corpus* TG 2018-2019 è pari al 4,69% vs. il 4,57% del *corpus* TG e il 4,80% del *corpus* CRS. Un risultato indicativo di un androcentrismo linguistico che accomuna l'informazione dei notiziari con l'informazione della stampa quotidiana. Questo androcentrismo linguistico può di nuovo essere interpretato come il risultato dei tre già citati fattori congiunti: 1) la disparità di genere nel contesto extra-linguistico di riferimento; 2) la sotto-

rappresentazione mediatica delle ministre rispetto ai ministri; 3) l'oscuramento linguistico delle ministre designate con titoli declinati al maschile. Il contesto di riferimento prevalente del *corpus* TG 2018-19 è infatti, come per il *corpus* TG intero e per il *corpus* CRS, quello nazionale (87,79%), caratterizzato da una prevalenza di uomini fra i membri di entrambi i governi in carica nel gennaio 2018 e nel gennaio 2019, mediamente pari al 74,36%, un valore tuttavia inferiore di circa 20 punti percentuali alla frequenza delle forme maschili osservata nei *corpora* (95,31% nel *corpus* TG 2018-19 e 95,20% nel *corpus* CRS).

L'analisi dell'accordo fra genere grammaticale e genere semantico porta alla luce come la disparità di genere fra le più alte cariche politiche nazionali, principale contesto di riferimento extra-linguistico, sia amplificata dal linguaggio empiricamente osservato nei *corpora*, da un lato per effetto di una minor visibilità mediatica delle ministre, dall'altro per effetto di una tendenza a nominare le singole ministre con forme maschili e i gruppi ambigenere esclusivamente con la forma plurale *ministri*.

Tabella 60. *Corpus* TG 2018-19: frequenza delle quattro forme di *ministra/o* per genere semantico

Forma linguistica	Referente Donna		Referente Uomo		Referente Donna e/o Uomo		Referenti n.d.		Totale	
	N	%	N	%	N	%	N	%	N	%
ministra	41	39,05							41	4,58
ministre	1	0,95							1	0,11
<i>Tot. Femminili</i>	42	40,00							42	4,69
ministro	63	60,00	608	97,28					715	79,89
ministri			17	2,72	97	100,0	44	64,71	138	15,42
<i>Tot. Maschili</i>		60,00	635	100,0	97	100,0	24	35,29	853	95,31
Totale	105	100,0	635	100,0	97	100,0	68	100,0	895	100,0

Calcolando la percentuale dei riferimenti alle ministre, indipendentemente dalla forma femminile o maschile utilizzata nel riferimento, sul totale dei riferimenti con genere semantico specifico, donna o uomo, otteniamo infatti una visibilità delle ministre complessivamente pari a 14,38% (su totale N=730), un valore inferiore alla percentuale media delle ministre italiane nei Governi in carica nel gennaio 2018 e nel gennaio 2019, pari al 25,64%.

A questo primo effetto di riduzione della visibilità delle ministre, si aggiunge la tendenza a nominare le singole ministre con il maschile *ministro* (60%) e a nominare coppie o gruppi di ministre e ministri con il maschile plurale *ministri* (100%). Entrambi questi valori percentuali sono in linea con quelli ottenuti nell'analisi del *corpus* CRS, che, a fronte di una visibilità mediatica femminile pari al 16,63%, attesta una frequenza d'uso di *ministro*, nel riferimento alle singole ministre, pari al 56,06% e una frequenza d'uso di *ministro*, nel riferimento a gruppi ambigenere, pari al 100%. A differenza del *corpus* CRS, il *corpus* TG 2018-2019 non attesta alcuna frequenza del maschile singolare con valore generico, per indicare la carica ministeriale senza riferimento specifico a una persona che lo ricopre. Una forma, tuttavia, molto poco ricorrente anche nel *corpus* CRS: N=36 occorrenze, per un'incidenza pari allo 0,83%.

Per quanto riguarda le espressioni linguistiche ricorrenti nel riferimento alle ministre e ai ministri, i risultati sul *corpus* TG 2018-19 attestano una distribuzione di frequenze, per i diversi tipi di espressione, per genere grammaticale e genere semantico, abbastanza in linea con quelle osservate per il *corpus* TG integrale.

Tabella 61. *Corpus* TG 2018-19: distribuzione delle frequenze del tipo di espressione linguistica per genere semantico e genere grammaticale

Genere semantico	Donna						Uomo	
	F.		M.		Tot. (F. più M.)		Tot. (M.)	
Genere grammaticale	N	%	N	%	N	%	N	%
Espressione linguistica	N	%	N	%	N	%	N	%
Titolo generico	9	8,65			9	8,65		
Titolo generico più cognome	14	13,46	8	7,69	22	21,15	178	29,28
Titolo generico più nome e cognome			1	0,96	1	0,96	7	1,15
Titolo generico in collocazione con informazioni utili a determinare il genere semantico			3	2,88	3	2,88	68	11,18
Titolo specifico	1	0,96			1	0,96		
Titolo specifico più cognome	12	11,54	17	16,35	29	27,88	163	26,81
Titolo specifico più nome e cognome	4	3,85	24	23,08	28	26,92	57	9,38
Titolo specifico in collocazione con elementi utili a determinare il genere semantico ¹²⁸	1	0,96	10	9,62	11	10,58	135	22,20
Totale	41	39,42	63	60,58	104	100,00	608	100,00

¹²⁸ Le occorrenze di questo tipo di espressione attestate nel riferimento femminile sono casi di riprese anaforiche o cataforiche in cui nome e/o cognome non sono la testa del sintagma nominale.

Sia nel corpus TG integrale, sia nel il *corpus* TG 2018-19, le due espressioni più ricorrenti con riferimento a singole ministre sono il titolo specifico declinato al maschile più nome e cognome (23,08%), e, a seguire, il titolo specifico, sempre declinato al maschile, seguito dal cognome (16,35%). A differenza del *corpus* TG, il *corpus* TG 2018-19 attesta come terza strategia linguistica prevalente nel riferimento alle ministre non il titolo specifico, bensì quello generico, declinato al femminile, più cognome (13,46%). Per quanto riguarda il riferimento ai ministri, tutti e due i *corpora* attestano la prevalenza del titolo generico o specifico seguito dal cognome, ma con frequenze relative leggermente diverse: nel *corpus* TG la categoria più frequente è il titolo specifico più cognome (30,87%), e, a seguire, il titolo generico più cognome (26,05%), mentre nel *corpus* TG 2018-19 la tipologia più frequente è il titolo generico più cognome (29,28%), e poi il titolo specifico più cognome (26,81%).

Venendo al confronto fra il *corpus* TG 2018-2019 e il *corpus* CRS, i due *corpora* attestano strategie linguistiche parzialmente simili nel riferimento alle ministre e ai ministri. In entrambi i *corpora*, l'espressione linguistica prevalente nel riferimento alle ministre è il titolo specifico declinato al maschile seguito da nome e cognome. Questa strategia linguistica può essere ricondotta, come più sopra ricordato, a un discorso giornalistico che assume il maschile come forma di prestigio, specie nella citazione specifica del titolo, che, riportando il nome del ministero di afferenza, riconduce la figura ministeriale entro il suo preciso ambito istituzionale.

Nel caso dei TG, il titolo specifico più nome e cognome sembra però essere funzionale anche a rendere trasparente l'identità di genere delle ministre, senza rinunciare alla declinazione maschile del titolo, poiché il nome proprio permette di disambiguare il genere semantico. Un'indicazione in tal senso deriva dalla bassa frequenza di questa espressione linguistica sia nel caso di riferimento alle ministre con titolo declinato al femminile (3,85%), sia nel caso di riferimento ai ministri (9,38%). In questi due casi, in cui il genere grammaticale di *ministra/o* è coerente con il genere semantico, e dunque il genere del/la referente è esplicitato dalla

forma grammaticale del titolo, le espressioni linguistiche più frequenti non includono infatti il nome proprio.

Nel caso del *corpus* CRS, invece, l'elevata frequenza dei riferimenti alle ministre con titolo specifico declinato al maschile più nome e cognome non sembra ascrivibile alla medesima funzione di esplicitare l'identità di genere, poiché il titolo specifico più nome e cognome è anche la modalità più frequente nel riferimento ai ministri (42,79%) ed è molto frequente anche nel caso di riferimento alle ministre con titolo declinato al femminile (19,65%), un contesto in cui la citazione del nome proprio è superflua a esplicitare il genere della referente, già reso trasparente dalla declinazione del titolo al femminile.

6.7 Conclusioni

L'analisi sull'accordo di genere, grammaticale e semantico, del lemma *ministra/o* nel *corpus* TG (integrale e ridotto al 2018-2019) e nel *corpus* CRS ha attestato come primo risultato generale un elevato androcentrismo linguistico. Se confrontiamo le frequenze delle forme femminili con le forme maschili, senza considerare il genere semantico dei riferimenti, le prime risultano nettamente inferiori con una frequenza relativa che, arrotondata all'unità, è pari al 5% in tutti tre i *corpora*: 4,57% nel *corpus* TG, 4,69% nel *corpus* TG 2018-19, 4,80% nel *corpus* CRS. Questo androcentrismo linguistico sembra essere riconducibile a tre fattori che interagiscono fra loro in modo parzialmente diverso nell'informazione televisiva e in quella a stampa, per anno e per testata giornalistica. Il primo fattore riguarda la prevalenza degli uomini in tutti i tre governi italiani in carica nei tre diversi periodi analizzati, che rappresentano il principale contesto extra-linguistico di riferimento dei *corpora* analizzati. Il secondo fattore concerne la sotto-rappresentazione mediatica delle ministre, che, indipendentemente dal genere grammaticale della forma linguistica ricorrente per designarle, sono complessivamente meno presenti dei loro colleghi nel discorso politico attestato dall'analisi delle collocazioni del lemma *ministra/o* nei *corpora*. Il terzo fattore è riconducibile a una tendenza a designare le ministre con il titolo declinato al maschile: nel 61,68% dei casi nel

corpus TG integrale, nel 60,00% dei casi nel *corpus* TG 2018-19, e nel 56,06% dei casi nel *corpus* CSR. Una tendenza, quest'ultima, che è indicativa di una resistenza alla femminilizzazione dei nomi di cariche politiche apicali, qual è un membro governativo, tuttavia molto attenuata rispetto a quanto rilevato nello studio di Formato (2016), che attesta frequenze d'uso del maschile per nominare le ministre pari al 91,38% per l'anno 2012, all'88,94% per il 2013 e all'88,72% nel 2014, su un *corpus* di quotidiani direttamente comparabile al *corpus* CRS, poiché costituito dalle stesse testate giornalistiche: *Corriere della sera*, *Il Resto del Carlino* e *La Stampa*.

La comparazione per anno sulla frequenza delle forme femminili e maschili, indipendentemente dal genere semantico, ha rilevato differenze significative, fra il gennaio 2018 e il gennaio 2019 (e anche il gennaio 2020, per il *corpus* TG), attestando una frequenza di forme femminili superiore alla frequenza attesa per il gennaio 2018. Questa differenza non sembra potersi considerare come il riflesso di un cambiamento nel contesto extra-linguistico di riferimento prevalente, ovvero le diverse compagini governative in carica periodi considerati, poiché il numero delle ministre in carica nel gennaio 2018 e 2019 era lo stesso (N=5), con una percentuale di incidenza sulla composizione totale del governo superiore nel 2018 rispetto al 2019, per via di un ministro in più nel Governo Conte I, rispetto al Governo Gentiloni, ma di solo 1,32 punti percentuali (26,32% vs. 25%). Inoltre, se consideriamo anche il Governo Conte II, in carica nel mese di gennaio 2020, coperto dal *corpus* TG, esso includeva più ministre rispetto ai precedenti, sia in termini assoluti (N=8) che percentuali (34,78%), eppure la frequenza delle forme femminili del lemma *ministra/o* attestata nel *sub-corpus* TG 2020 è sostanzialmente in linea con quelle attese, +1,09 vs. +8,38 del 2018.

I risultati sull'analisi dell'accordo fra genere grammaticale e genere semantico dimostrano che il minor androcentrismo linguistico attestato per il 2018 è contestuale ad altre due evidenze: la minore sotto-rappresentazione mediatica delle ministre, rispetto ai ministri, nel 2018 rispetto al 2019 (e anche al 2020 per il *corpus* TG), e la minor tendenza a nominarle al maschile. Il primo fattore è più evidente nel *corpus* CRS, che attesta una percentuale di riferimenti alle ministre

pari al 21,65% nel gennaio 2018 vs. il 13,32% nel gennaio 2019 e, allo stesso tempo, evidenzia differenze minime e non significative in termini di accordi di genere con il genere semantico femminile, fra i due periodi. Il secondo fattore invece è più evidente nel *corpus* TG, che attesta un significativo maggiore numero di accordi fra il genere grammaticale femminile e il genere semantico femminile, per il gennaio 2018, rispetto al gennaio 2019 (e anche al gennaio 2020 per il *corpus* TG). Detto altrimenti, il linguaggio dell'informazione televisiva risulta meno androcentrico nel gennaio 2018 rispetto al 2019 (e anche al 2020) per effetto di una maggiore tendenza a designare le ministre con il titolo al femminile; il linguaggio della stampa risulta meno androcentrico nel 2018 rispetto al 2019 per via anche di una maggior presenza delle ministre nel discorso giornalistico dei quotidiani, i quali, nei due periodi, risultano comunque un po' più attenti a declinare al femminile il lemma *ministra/o* nel riferimento alle ministre rispetto ai TG (*corpus* CRS 43,94% vs. 40% del *corpus* TG 2018-19).

Questo risultato si riflette anche sui dati ottenuti dall'analisi specificamente dedicata alle due testate giornalistiche interessate da un cambio di direzione editoriale, ovvero il TG2 e *Il Resto del Carlino*. Il TG2 attesta una maggior frequenza di forme femminili nel 2018 rispetto al 2019, a fronte di un minor numero di riferimenti alle ministre, contestuale a una significativa maggior frequenza d'uso di forme femminili per citare le pur meno numerose ministre, rispetto ai ministri. Questo risultato è indicativo di un minor androcentrismo linguistico nel *sub-corpus* TG2 2018, rispetto al *sub-corpus* TG2 2019 (e anche al *sub-corpus* TG2 2020) ascrivibile a una tendenza a rendere visibili le ministre attraverso l'uso delle forme femminili. Una tendenza che, a monte, può forse essere spiegata con scelte editoriali dettate da una direttrice dichiaratamente attenta all'uguaglianza di genere. Per quanto riguarda *Il Resto del Carlino* il minor androcentrismo linguistico attestato per il gennaio 2018, rispetto al gennaio 2019, è contestuale sia a una maggior numero di riferimenti alle ministre sia a una maggiore frequenza di forme femminili. Sebbene non abbia trovato informazioni attestanti un impegno per le pari opportunità del direttore del quotidiano *Il Resto del Carlino* nel gennaio 2018,

i risultati evidenziano che il cambio di direzione coincide con una diversa attenzione sia alle ministre come oggetto del discorso giornalistico, sia al linguaggio di genere. Un risultato quest'ultimo che potrebbe essere spiegato anche dal fatto che nel 2018, periodo per il quale si osserva in tutti i *sub-corpora* analizzati una maggior frequenza di forme femminili per designare le ministre, era ancora in vigore la 17esima legislatura caratterizzata da un impegno della presidente della Camera Laura Boldrini a favore di un uso della lingua italiana paritario e inclusivo, rispetto al genere. Le elezioni politiche del 4 marzo 2018 hanno avviato una legislatura che è rimasta sprovvista di una "paladina" del linguaggio *gender-fair* e questo può aver inciso nel senso di un disimpegno di alcune testate giornalistiche, in particolare proprio quelle come il TG2 e *Il Resto del Carlino* che nel gennaio 2018 risultavano più impegnate in tal senso.

Infine, per quanto riguarda le espressioni linguistiche, le conclusioni che si possono avanzare sono parzialmente diverse per i *corpora* analizzati. Per quanto riguarda il *corpus* TG (e anche il *corpus* TG 2018-19), la prevalenza del titolo specifico declinato al maschile seguito da nome e cognome sembra essere ascrivibile a una maggiore autorevolezza associata alla forma maschile, come attestato da Formato (2016), ma anche alla funzione di rendere esplicita l'identità di genere del/la referente, opaco nella declinazione del titolo al maschile, attraverso la citazione del nome proprio. Tant'è vero che, in generale, quando il titolo è declinato al femminile, nella maggior parte dei casi (51,20%) è collocato in espressioni linguistiche che non includono il nome proprio, superfluo a identificare l'identità di genere, resa già trasparente dalla forma femminile; mentre quando il titolo è riferito ai ministri include meno frequentemente anche il nome proprio (11,25% vs. 59,62%).

Per quanto riguarda il linguaggio della carta stampata i risultati attestano un'elevata frequenza di collocazioni che comprendono il titolo specifico, cioè completo del nome del dicastero di appartenenza, più nome e cognome, a prescindere dal genere semantico del/la referente e dal genere grammaticale della forma linguistica utilizzata nel riferimento alle ministre, che è maschile nel 23,83% delle

occorrenze e femminile nel 19,85%. Questo risultato sembrerebbe indicare una maggiore ufficialità del discorso giornalistico della carta stampata che predilige, in tutti i casi, designare i membri del governo specificandone il dicastero, e in particolare nome e cognome. A differenza di quanto attestato da Formato (2016) non sembra che entro questo discorso ufficiale il titolo declinato al femminile abbia meno prestigio del titolo declinato al maschile nel riferimento alle ministre. È questo senz'altro un risultato interessante che testimonia un cambiamento in atto nel linguaggio dei mezzi d'informazione italiana, meno resistenti all'uso delle forme femminili rispetto a qualche anno fa, forse anche per merito della crescente metacompetenza linguistica di giornaliste e giornalisti che, dal 2014 dispongono di almeno due pubblicazioni dedicate all'argomento, una delle quali promossa dall'Ordine dei Giornalisti (CNOG 2015), l'altra dall'associazione GiULiA (Robustelli 2014), che conta oltre 1.000 socie, ha una pagina Facebook con oltre 16.000 *follower*¹²⁹ e promuove periodicamente corsi di formazione in materia.

¹²⁹ Fonte: https://www.facebook.com/GiuliaGiornaliste/about/?ref=page_internal; ultima consultazione 19 novembre 2021.

Capitolo 7 Analisi delle associazioni di genere latenti nei nomi di professione e carica politica in un *corpus* di notizie

Questo capitolo studia le associazioni semantiche di genere negli *embeddings* di una selezione scelta di nomi di professione e carica politica estratti dal *corpus* ACR, presentato nel Capitolo 4 (§ 4.4). Il paragrafo introduttivo (§ 7.1) presenta l'ipotesi della semantica distribuzionale su cui si basano i Distributional Semantic Models (DSMs) in grado di rappresentare numericamente il significato implicito delle parole, così come di apprendere tali rappresentazioni da dati linguistici: i *word embeddings* (WEs) sono infatti parole rappresentate sotto forma di vettori costruiti, usando una rete neurale, in uno spazio implicito (Lenci 2018, 160). In questo stesso paragrafo sono brevemente descritti i due principali DSMs attualmente utilizzati per rappresentare (e apprendere) empiricamente il significato delle parole di un *corpus* testuale. Il paragrafo successivo (§ 7.2) ricostruisce lo stato dell'arte degli studi che attestano associazioni (o *bias*) di genere in *word embeddings* estratti da *big-corpora*. Questa linea di ricerca si è sviluppata prevalentemente su *corpora* in lingua inglese, nell'ambito di studi internazionali, ma associazioni di genere in *embeddings* di parole della lingua italiana sono stati recentemente indagati da ricerche multilingue (Grave et al. 2018; Lewis e Lupyan 2020) e anche da uno studio nazionale, che stima, fra l'altro, gli stereotipi latenti in nomi di professione e attesta il ruolo del genere grammaticale nel mitigare questi stereotipi (Biasion et al. 2020).

Partendo dallo stato dell'arte, nel paragrafo 7.3 viene introdotta l'ipotesi che gli *embeddings* di una selezione di nomi di professioni e cariche politiche attestino la persistenza di stereotipi di genere. Nello stesso paragrafo vengono introdotte tre domande di ricerca: qual è l'associazione semantica di genere nei *word embeddings* di una selezione di nomi di professioni e cariche politiche estratti dal *corpus* ACR? C'è una correlazione fra questa associazione e la realtà sociale, ovvero la partecipazione femminile nei diversi tipi di professione considerata? Quale ruolo ha il genere grammaticale dei nomi nel confermare associazioni stereotipiche o mitigarle? Il paragrafo successivo (§ 7.4) spiega la metodologia di analisi, basata

sull'approccio di Bolukbasi et al. (2016) per misurare le associazioni di genere dei nomi indagati, e sull'approccio di Biasion et al. (2020) per verificarne la correlazione con la partecipazione effettiva delle donne nella professione o funzione politica corrispondente. Nei paragrafi successivi (§ 7.5 e § 7.6) vengono illustrati e discussi i risultati che dimostrano che i nomi di professione e carica politica analizzati hanno diversi gradi di associazione con il genere femminile o maschile e che questa associazione è correlata con l'incidenza femminile nella professione o carica politica corrispondente, dunque riflette una conoscenza della realtà, ma riflette anche pregiudizi sociali come dimostrano due nomi: *ingegnera/e*, sul versante dell'associazione con il genere maschile, e *ostetrica/o*, sul versante dell'associazione con il genere femminile. Nonostante la forma femminile, nel primo caso, e quella maschile, nel secondo, risultino più associate al genere semantico rispettivamente femminile e maschile, in modo coerente con la relazione di congruenza fra genere grammaticale e genere semantico, che caratterizza la maggior parte dei nomi con riferimento umano nel sistema linguistico italiano, *ingegnera/e* rimane un nome fortemente associato al genere maschile e *ostetrica/o* a quello femminile nel linguaggio giornalistico italiano, o perlomeno nel *corpus* indagato. Infine, l'analisi dimostra che il genere grammaticale può mitigare gli stereotipi di genere. Il capitolo si conclude con il paragrafo 7.7 che propone alcune riflessioni sul ruolo della stampa nella prevenzione e nel contrasto degli stereotipi di genere attraverso l'uso di un linguaggio *gender-fair*, che nel caso della lingua italiana può iniziare con un linguaggio attento alla corretta declinazione di genere dei nomi di professione .

7.1 Introduzione

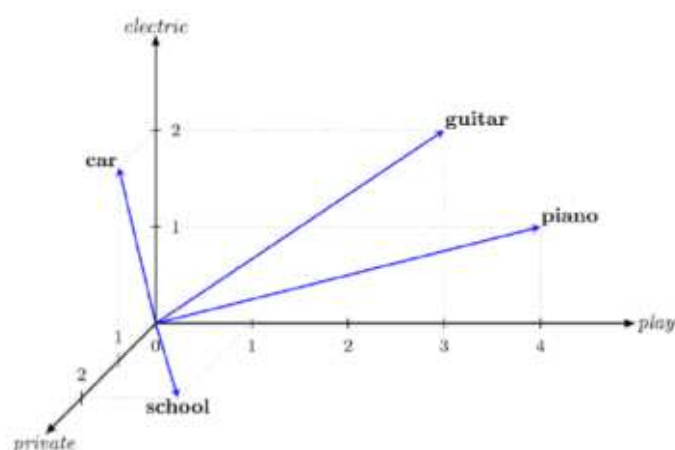
I DSMs si basano sull'ipotesi distribuzionale intuita da Wittgenstein (1953) e Firth (1957) e formulata in modo più articolato da Harris (1954), che non si limita a osservare come il significato di una parola dipenda dall'uso o, detto altrimenti, dalle parole che l'accompagnano, ma afferma che «difference of meaning correlates with difference of distribution» (Harris 1954, 156). Sulla base di questo presupposto, i DSMs rappresentano similarità e differenze di significato fra le

parole in termini di distanza vettoriale in uno spazio geometrico. Partendo da una matrice di dati con parole in riga e parole (o testi) in colonna, trasformano (tramite programmi informatici) le parole in vettori, ovvero elementi geometrici definiti da tre valori (origine, dimensione, direzione) assegnati partendo dai numeri di coricorrenza fra parole (o fra parole e testi). I rapporti di similarità e differenza fra le parole vengono misurati in termini di distanza fra vettori. Per esempio partendo dalla matrice di coricorrenze rappresentata nella Figura 2 si ottiene una rappresentazione vettoriale come quella riportata nella Figura 3.

Figura 2. Esempio di una matrice di coricorrenze¹³⁰

	play	electric	classical	private	high	...	the	new
guitar	3	2	1	0	0	...	0	1
piano	4	1	1	0	0	...	4	0
car	0	2	0	1	0	...	4	2
school	1	0	0	2	2	...	1	1

Figura 3. Esempio di rappresentazione vettoriale della parole ricavata dalla matrice in Figura 2¹³¹



¹³⁰ La tabella è ripresa dai materiali didattici del corso di Linguistica Computazionale di Gianluca Leboni presso il Dipartimento di Studi Linguistici e Culturali Comparati, Università Ca' Foscari di Venezia, aa 2019-2020.

¹³¹ *Ibidem.*

I DSMs sviluppati nel corso del tempo sono numerosi e vengono impiegati per varie funzioni, come ad esempio la decisione automatica per la selezione di CV aderenti a determinate posizioni lavorative (De-Arteaga et al. 2019). Qui mi limito a presentare la Latent Semantica Analysis (LSA), uno dei principali «count models» ovvero, seguendo la distinzione proposta da Baroni et al. (2014), modelli che imparano la rappresentazione di una parola target registrando e contando le sue coricorrenze in contesti linguistici, e word2vec¹³² (Mikolov et al. 2013), uno dei principali «prediction models», ovvero modelli a rete neurale che creano direttamente rappresentazioni distributive implicite imparando a prevedere in modo ottimale i contesti di una parola target.

La LSA analizza la relazione tra parole e testi, ipotizzando che parole vicine nel significato si presentino in simili segmenti di testo. Operativamente, nella LSA, i dati linguistici sono organizzati in una matrice con i testi in colonna e le parole in riga, e le celle compilate con valore n. per ogni corrispondenza fra testo e parola, ovvero la frequenza delle coricorrenze. Il numero di righe viene poi successivamente ridotto, tramite un processo di fattorizzazione chiamato “decomposizione ai valori singolari”, detta anche SVD, dall’inglese *singular value decomposition*, che crea un nuovo spazio di dimensionalità ridotta in modo da evidenziare certe relazioni.

Word2vec (Mikolov et al. 2013) è uno dei principali modelli a rete neurale attualmente in uso, in grado di creare *word embeddings*. Operativamente esso consiste in un pacchetto software con due architetture di apprendimento alternative: CBOW (Continuous Bag Of Words) e Skip-gram. CBOW si basa sull’occorrenza delle parole nel contesto, Skip-gram si basa invece sulla coricorrenza fra parole.

¹³² Word2vec è una open source, disponibile online all’indirizzo: code.google.com/p/word2vec/; ultimo accesso 19 novembre 2021.

Word2vec, così come altri modelli a reti neurale, come per esempio GloVe (Pennington et al. 2014), hanno avuto un grande successo e sono attualmente ampiamente utilizzati nel campo dell'Intelligenza Artificiale. Questo spiega perché sono contestualmente diventati oggetto di studio della linguistica computazionale e dell'ingegneria informatica che, da un lato, ne testano l'efficacia e l'applicazione anche per lingue diverse dall'inglese (al-Rfou et al. 2013; Bojanowski et al. 2017; Grave et al. 2018; Lewis e Lopyan 2020), incluso l'italiano (Berardi et al. 2015; Lomonaco 2015; Marelli 2017; Tripodi e Li Pira 2017), dall'altro, ne indagano i problemi in termini di stereotipi e *bias* cognitivi latenti (Bolukbasi et al. 2016; Caliksan et al. 2017; Garg et al. 2018; Biasion et al. 2020), non solo per riflettere su come e quanto una lingua codifichi pregiudizi sociali, ma anche per migliorare le performance degli algoritmi stessi (Bolukbasi et al. 2016; Claiksan et al. 2017; Manzini et al. 2019), onde evitare che riproducano rappresentazioni stereotipate, provocando danni simbolici ed effettivi che possono ricadere su alcuni gruppi sociali (Barocas et al. 2017, 2019¹³³; Crawford 2017). Lo studio di De-Arteaga et al. (2019) dimostra, per esempio, che i processi di classificazione automatica dei profili professionali discriminano le donne codificando informazioni linguistiche sul genere dei/delle candidati/e per una posizione lavorativa, sulla base di marcatori quali i pronomi e i nomi propri.

7.2 Stato dell'arte

Il primo studio che affronta il problema degli stereotipi nei WEs è quello di Bolukbasi et al. (2016). Partendo dal presupposto che gli algoritmi di Machine Learning sono ottimizzati per modellizzare proprietà statistiche di dati linguistici che possono codificare anche *bias* cognitivi e pregiudizi sociali, Bolukbasi e colleghi si focalizzano sugli stereotipi di genere, in particolare quelli codificati nei nomi di professione: elaborano un metodo per quantificare gli stereotipi di genere in WEs

¹³³ «When decision-making systems in criminal justice, health care, etc. are discriminatory, they create *allocative harms*, which are caused when a system withholds certain groups an opportunity or a resource. In contrast, the other examples—stereotype perpetuation and cultural denigration—are examples of *representational harms*, which occur when systems reinforce the subordination of some groups along the lines of identity—race, class, gender, etc.» Barocas et al. (2017, 32).

pre-addestrati, e poi sviluppano un algoritmo di *de-biasing*, in grado di mitigare gli stereotipi di genere, senza che l'algoritmo perda di efficacia in termini di *performance*. In particolare, studiano gli stereotipi in applicazioni utilizzate per funzioni di recupero basate su compiti di analogia del tipo $A: B :: C: D$, per recuperare l'elemento sconosciuto D , a partire dai dati conosciuti A , B e C ¹³⁴, avendo preliminarmente verificato che word2vec (Mikolov et al. 2013) può restituire come risultato di un simile compito risposte (stereotipate) del tipo: «*father : doctor :: mother : nurse*», oppure «*man : computer programmer :: woman : homemaker*» (Bolukbasi et al. 2016, 41).

Per quantificare gli stereotipi di genere, analizzano un elenco di 215 nomi di professioni, nessuno dei quali marcati rispetto al genere (escludono per esempio nomi come *waitress* e *waiter*) e ne misurano la distribuzione lungo la «gender-direction *he-she*», ottenuta dalla differenza fra \vec{he} e \vec{she} .

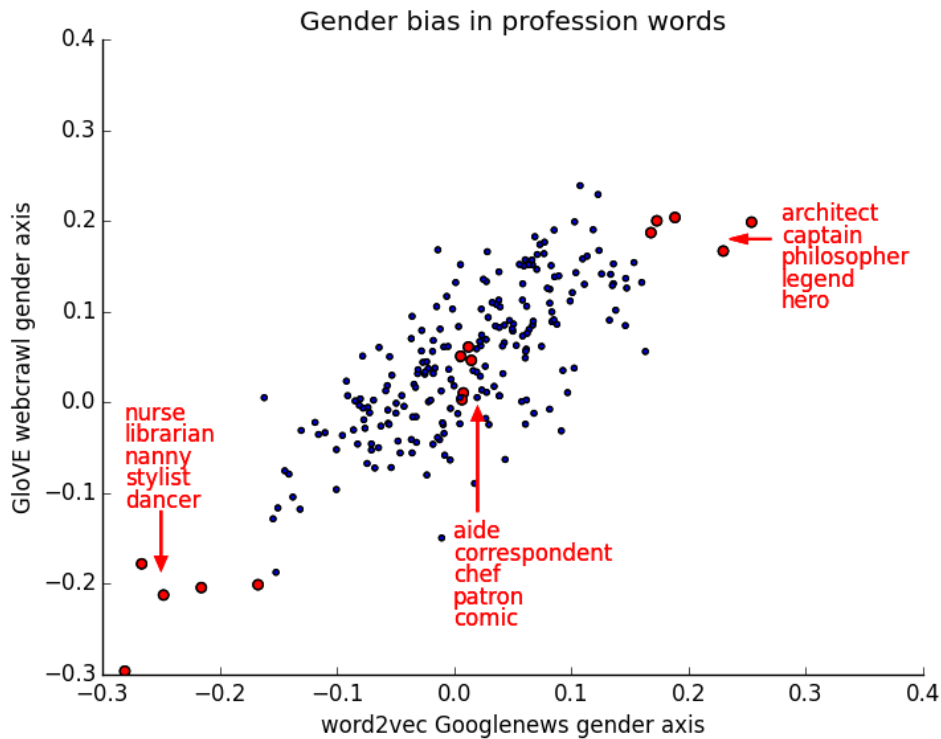
A simple approach to explore how gender stereotypes manifest in embeddings is to quantify which words are closer to *he* versus *she* in the embedding space (Bolukbasi et al. 2016, 42).

Per ogni nome di professione, ne calcolano la proiezione sulla direzione di genere. Rappresentano questa proiezione su un grafico con i WEs ricavati da word2vec (addestrato su un *corpus* di Google news) sull'asse delle ascisse e i WEs estratti da GloVe (addestrati su un set di dati scaricati dal web) sull'asse delle ordinate, ottenendo come risultato una rappresentazione coerente fra i due WEs in termini di associazione stereotipata, al femminile e al maschile, dei 215 nomi di professione analizzati. Bolukbasi et al. (2016) non riportano l'elenco completo dei nomi, ma solo i più stereotipati al femminile: *nurse, librarian, nanny, stylist, dancer*; al maschile: *architect, captain, philosopher, legend, hero*; e quelli che si collocano

¹³⁴ Word2vec sfrutta la potenzialità dei WEs di rappresentare regolarità linguistiche per eseguire compiti di recupero. Per esempio, partendo dal fatto che i WEs rilevano una medesima relazione di genere (semantico) fra \vec{king} e \vec{queen} e fra \vec{man} e \vec{woman} , \vec{queen} può essere recuperato dalle rappresentazioni delle altre parole attraverso una semplice aritmetica vettoriale "sottraendo" a \vec{king} , \vec{man} e aggiungendo \vec{woman} .

in una posizione intermedia rispetto alla direzione di genere: *aide*, *correspondant*, *chef*, *patron*, *comic* (Figura 4).

Figura 4. *Bias* di genere in nomi di professione in due diversi spazi vettoriali generati da due software diversi (GloVe e word2vec) (Bolukbasi et al. 2016, Fig.1)¹³⁵



Il lavoro di Caliskan et al. (2017) si propone dichiaratamente come complementare allo studio di Bolukbasi et al. (2016), sviluppando un metodo che va oltre la “formulazione algebrica” delle associazioni di genere, cercandone i nessi sia con le rappresentazioni mentali, attestate e attestabili tramite test psicologici, sia con i dati di realtà. Per quanto riguarda il primo obiettivo, lo studio di Caliskan et al. (2017) elabora un metodo basato sull’Implicit Association Test (IAT), utilizzato in psicologia sociale per misurare la forza di associazione implicita fra diversi tipi di rappresentazioni mentali. In particolare, riprende gli studi di Greenwald et al.

¹³⁵ Riporto di seguito la didascalia completa in Bolukbasi et al. (2016, 42): «Comparison of gender bias of profession words across two embeddings: word2vec trained on Googlenews and GloVe trained web-crawl texts. The x and y axes show projections onto the *he-she* direction in the two embeddings. Each dot is one of 249 common profession words. Words closest to *he*, closest to *she*, and in between the two are colored in red and shown in the plot».

(1998), Nosek et al. (2002a, 2002b), Bertrand e Mullainathan (2004), Monteith e Pettit (2011).

Lo studio seminale di Greenwald et al. (1998) dimostra, attraverso un test di associazione implicita (IAT), che il tempo di reazione in compiti (sperimentali) di associazione fra due concetti target (per esempio nomi di fiori vs. nomi di insetti) e un attributo che esprime gradevolezza vs. sgradevolezza (es. *beautiful-ugly*), variano in relazione a tre fattori: 1) differenze valutative quasi-universali¹³⁶, 2) differenze individuali attese in associazioni valutative, 3) differenze valutative consapevolmente negate. Se un'associazione è già presente come rappresentazione mentale in un soggetto, i tempi di reazione sono inferiori e variano sulla base dei tre fattori sopra elencati. Il primo fattore è reso evidente, per esempio, da un minor tempo di reazione nell'associazione del nome di un fiore a un attributo di gradevolezza rispetto al tempo di reazione nell'associazione del nome di un insetto allo stesso gruppo di aggettivi. Il secondo fattore è dimostrato dalla più rapida associazione di aggettivi che esprimono gradevolezza a nomi propri di persona giapponesi per soggetti (partecipanti all'esperimento) giapponesi vs. soggetti coreani, che associano in modo sensibilmente più rapido attributi positivi a nomi propri di persona coreani. Il terzo fattore è dimostrato dalla differenza dei tempi di associazione fra nomi tipicamente usati da americani di origine europea («white») e aggettivi positivi vs. nomi tipicamente usati da americani di origine africana («black») e attributi negativi per soggetti partecipanti all'esperimento americani «white» che si sono descritti come senza pregiudizi.

Lo studio di Nosek et al. (2002a) rileva, attraverso nove test di associazione implicita, pregiudizi razziali, generazionali, politici e di genere, dimostrando, rispetto a quest'ultima variabile, che la scienza e la matematica (vs. le arti), così come la carriera (vs. la famiglia) sono implicitamente associate più agli uomini che alle donne. L'associazione implicita fra matematica (vs. le arti) e genere maschile (vs. femminile) viene confermato anche da Nosek et al. (2002b) con un

¹³⁶ «Near-universal» (Greenwald et al. 1998, 1464).

esperimento su studenti universitari che considera cinque indicatori: 1) la preferenza per la matematica, 2) l'identificazione di sé con la matematica, 3) la credenza che la matematica sia più maschile che femminile, 4) l'identificazione di sé con il genere femminile o maschile, 5) la performance in test matematici.

Marianne Bertrand e Sendhil Mullainathan (2004) misurano l'impatto di pregiudizi razziali codificati dai nomi propri di persona sulla selezione delle risorse umane. Le autrici creano *curricula vitae* fittizi assegnandoli casualmente a persone dal nome tipicamente euro-americano o afro-americano. Invia questi profili professionali in risposta ad annunci di lavoro pubblicati su due quotidiani, di Boston e Chicago, ottenendo come risultato che i CV assegnati a persone con un nome tipicamente euro-americano ricevono il 50% in più di risposte positive, ovvero risposte accompagnate dall'invito a un colloquio di lavoro. Monhith e Pettit (2011) indagano gli stereotipi relativi alla depressione, rilevando atteggiamenti impliciti generalmente più negativi verso questo tipo di patologia, rispetto a malattie fisiche.

Nel loro primo esperimento Caliskan et al. (2017) verificano con successo una congruenza fra le associazioni mentali attestate dai diversi IAT sopra citati e i WEs pre-addestrati con GloVe sul *web corpus* Common Crawl (Pennington et al. 2014). Operativamente sviluppano un modello (WEAT, Word-Embedding Association Test) che usa il valore della distanza vettoriale tra nomi e attributi, precisamente il coseno, come unità di misura delle associazioni al posto dei millisecondi, con cui vengono misurati i tempi di reazione nell'IAT, e poi comparano la prima misurazione con la seconda. Introducendo qualche modifica in alcune liste originarie di nomi (utilizzate negli IAT con cui confrontano i loro dati), per via della loro bassa frequenza nel *corpus* di analisi, Caliskan et al. (2017) verificano l'associazione con attributi di gradevolezza/sgradevolezza delle seguenti categorie: 1) fiori vs. insetti, 2) armi vs. strumenti musicali, 3) nomi propri di persona euro-americani vs. afro-americani (Greenwald et al. 1998, Bertrand e Mullainathan 2004), 4) giovani e anziani (Nosek et al. 2002a), più 5) l'associazione implicita di nomi maschili vs. femminili con termini che fanno riferimento alla sfera della carriera professionale vs. attributi che fanno riferimento alla famiglia (Nosek et al.

2002a), 6) l'associazione di nomi maschili vs. femminili con termini che fanno riferimento alla scienza vs. le arti (Nosek et al. 2002a), 7) l'associazione di nomi maschili vs. femminili con termini che fanno riferimento alla matematica vs. le arti (Nosek et al. 2002a, 2000b), 8) l'associazione di nomi di patologie mentali vs. patologie fisiche con attributi che qualificano la patologia come temporanea vs. permanente (Montheith e Pettit 2011).

Le conclusioni a cui giungono Caliskan et al. (2017) nel loro primo esperimento è che i WEs attestano diversi *bias*, che possono riguardare fiori o insetti, armi o strumenti musicali, ma anche alcuni gruppi sociali. Per quanto riguarda i nomi femminili e maschili, gli autori considerano l'ipotesi che i WEs possono rappresentare non solo stereotipi ma anche un *gender-gap* reale, codificando una conoscenza della realtà. Verificano questa ipotesi confrontando gli stereotipi di genere codificati nei nomi propri di persona androgini (es. *Kelly, Tracy, Jamie, Jackie*, usati negli Stati Uniti sia per le donne che per gli uomini) e nei nomi di 50 professioni¹³⁷ con la corrispondente situazione reale, ovvero l'uso effettivo, registrato all'anagrafe, dei nomi androgeni per donne e uomini e il tasso di occupazione femminile nelle 50 professioni corrispondenti ai nomi selezionati. Per confrontare gli stereotipi di genere con l'incidenza femminile effettiva in queste due categorie sviluppano il Word-Embedding Factual Association Test (WEFAT) e lo testano con due esperimenti: WEFAT 1, per nomi propri di persona; WEFAT 2, per i nomi di professione. Per calcolare l'associazione semantica confrontano i due gruppi di parole target (professioni e nomi propri) con due gruppi di attributi di genere descrittivi, come *male/female* e *woman/man*. Assumono come ipotesi nulla che non vi sia differenza tra i due gruppi di parole target in termini di somiglianza relativa con i due insiemi di parole attributo. Tramite un test di permutazione calcolano la probabilità che una permutazione casuale degli attributi produca la

¹³⁷ *technician, accountant, supervisor, engineer, worker, educator, clerk, counselor, inspector, mechanic, manager, therapist, administrator, salesperson, receptionist, librarian, advisor, pharmacist, janitor, psychologist, physician, carpenter, nurse, investigator, bartender, specialist, electrician, officer, pathologist, teacher, lawyer, planner, practitioner, plumber, instructor, surgeon, veterinarian, paramedic, examiner, chemist, machinist, appraiser, nutritionist, architect, hairdresser, baker, programmer, paralegal, hygienist, scientist.*

differenza osservata (o maggiore) nelle medie del campione. Come dati reali di confronto, utilizzano i dati riportati dal censimento statunitense del 1990. I risultati sono riportati in due grafici che mostrano come al crescere della percentuale di donne in una professione (WEFAT 2), o effettivamente registrate all'anagrafe statunitense con un nome androgino (WEFAT 1), corrisponda una crescente associazione semantica con il genere femminile (Figura 5).

Figura 5. Indice di correlazione fra associazione con il genere femminile di nomi di professione in uno spazio vettoriale e percentuale di occupazione delle donne nella relativa professione (Caliskan et al. 2017, 2)

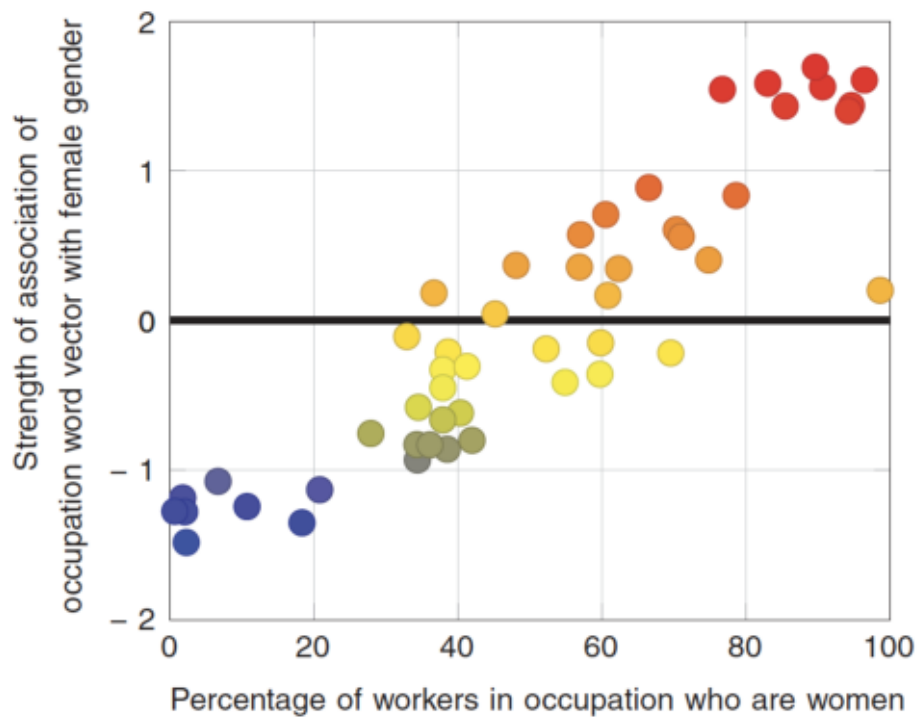


Fig. 1. Occupation-gender association. Pearson's correlation coefficient $\rho = 0.90$ with $P < 10^{-18}$.

Caliskan et al. (2017) concludono osservando che i loro modelli sono utili a rilevare pregiudizi sociali, ma anche altre conoscenze della realtà, come le preferenze d'uso di un nome androgino per una donna o un uomo e la distribuzione di genere nelle professioni.

Uno sviluppo della ricerca in tal senso è proposta dagli studi di Garg et al. (2018) e Wevers (2019). Quest'ultimo esamina *bias* di genere in un *corpus* di sei

quotidiani olandesi pubblicati fra il 1950 e il 1990, ponendosi due obiettivi: primo, misurare *bias* di genere pertinenti diversi gruppi di parole che rappresentano la vita sociale, gli stati psicologici e il mondo del lavoro e delle professioni, lungo le quattro decadi comprese nel periodo selezionato, caratterizzato da una rapida affermazione dell'uguaglianza fra uomini e donne nei Paesi Bassi, indagando contestualmente se e quali differenze ci sono fra giornali di diverso orientamento ideologico; secondo, verificare, dal punto di vista metodologico, se i WEs sono un approccio efficace nel cogliere *bias* di genere variabili a seconda dei diversi orientamenti dell'informazione e del periodo storico. Operativamente, Wevers (2019) seleziona tre data-set di parole: 1) 12.500 nomi di titoli professionali contenuti nella banca dati HISCO (Historical International Classification of Occupations) (Zijdeman et al. 2013), 2) il vocabolario olandese delle emozioni di Cornetto (Combinatoric and Relational Network for Language Technology, Vossen et al. 2007 cit. in Wevers), 3) la traduzione olandese delle parole che misurano stati psicologici e cognitivi delle seguenti categorie del LIWC (Linguistic Inquiry and Word Count): processi emozionali e affettivi, processi cognitivi, processi sensoriali e percettivi, processi sociali, occupazione, attività del tempo libero, denaro e questioni finanziarie, questioni metafisiche, stati fisici (Pennebaker et al. 2001; Bloothoof 2010; Tausczik e Pennebaker 2010; Boot et al. 2017, cit. in Wevers cit. in Wevers 2019). Di tutte queste parole costruisce una rappresentazione vettoriale, utilizzando l'implementazione di word2vec di Gensim¹³⁸, optando per l'algoritmo CBOW.

Per misurare i *bias* di genere, Wevers (2019) costruisce due vettori che rappresentano le due dimensioni di genere (femminile e maschile) modellate come i centroidi delle parole che esprimono il genere a livello lessicale (es. *uomo*¹³⁹) e dei più popolari nomi propri di donne e uomini in Olanda, basandosi sulla banca dati *Nederlandse Voornamenbank* (Bloothoof 2010, cit. in Wevers 2019). Calcola quindi

¹³⁸ Gensim è una libreria open source, disponibile online all'indirizzo: <https://radimrehurek.com/gensim/> (ultimo accesso 169 novembre 2021).

¹³⁹ La traduzione è mia dall'inglese utilizzato da Wevers (2019) che riporta a titolo esemplificativo tre termini, tutti in inglese (*man, his, father*), senza citare la versione originale olandese.

la distanza di tutte le parole target dei tre data set da questi due vettori e usa la distanza delle parole target dall'uno e dall'altro centroide come misura di *bias* di genere: maggiore è la distanza della parola target dal vettore di genere maschile (o femminile), minore è l'associazione di quella parola al genere maschile (o femminile). Infine, per determinare variazioni significative nel corso del tempo, calcola il coefficiente di regressione lineare fra la variabile tempo (misurata in decenni) e la variabile *bias* (misurata come distanza di una parola dal genere femminile e maschile).

I risultati a cui perviene Wevers (2019) è che nel complesso il *corpus* testimonia una certa stabilità, mostrando deboli tendenze, in parte livellate da cambiamenti in direzione opposta dei singoli quotidiani. Fra queste tendenze, le categorie che si muovono da una associazione con il genere maschile a inizio anni Cinquanta del Novecento verso una associazione con il genere femminile negli anni Novanta sono: TV, musica, questioni metafisiche e sessualità, e anche i titoli professionali. Mentre le categorie che subiscono un cambiamento nella direzione inversa sono: il denaro, la cura di sé, le emozioni negative e il mondo del lavoro, distinto dai titoli professionali¹⁴⁰. Nel caso dei titoli professionali rileva un leggero cambiamento nella direzione di una progressiva maggiore associazione con le donne; mentre per il mondo del lavoro, un leggero cambiamento nella direzione di una maggiore associazione con il genere maschile. «This suggests – conclude l'autore – that job titles might be more closely related to women, while the notion of working gravitates toward men» (Wevers 2019, 4).

Lo studio di Garg et al. (2018) si propone di verificare i cambiamenti in termini di stereotipi verso le donne e verso le minoranze etniche, negli Stati Uniti, in un periodo di tempo che copre quasi 100 anni, dal 1910 al 1990, confrontando WEs pre-addestrati sui *corpora* Google news (word2vec, Mikolov et al. 2013) e

¹⁴⁰ Wevers (2019) utilizza due diversi data-set per misurare gli stereotipi di genere in ambito lavorativo: la banca dati HISCO (Historical International Classification of Occupations) che comprende 12.500 titoli professionali, più la traduzione olandese del data set LIWC (Linguistic Inquiry and Word Count) 2001 che contiene un vocabolario di parole raggruppate per diverse categorie, fra cui l'Occupazione.

Google books nel COHA (*Corpus of Historical American English*, Davies 2010, cit. in Garg et al. 2018), e WEs estratti dal *New York Times Annotated Corpus* con GloVe (Pennington et al. 2014) con dati ufficiali sull'occupazione di donne e uomini, e, per quanto riguarda le minoranze etniche, dati sugli spostamenti demografici interni agli Stati Uniti. I risultati generali a cui giunge lo studio è che i WEs "catturano" i cambiamenti sociali, come per esempio il movimento delle donne negli anni Sessanta del secolo scorso, che segna un passaggio nell'associazione di aggettivi che descrivono competenze e abilità quali *intelligent, logical, thoughtful* alle donne, precedentemente associati prevalentemente agli uomini.

Per quanto riguarda gli stereotipi di genere nei nomi di professione, Garg e colleghi calcolano le associazioni di genere in termini di distanza media (vettoriale) fra nomi che rappresentano le donne, cioè nomi il cui tratto di genere è marcato alla radice, e i nomi di professione. A fini comparativi, calcolano la stessa distanza per gli uomini. Assumono quindi come «natural metric for the embeddings bias» (Garg et al. 2018, E3636) la distanza media ottenuta da questo calcolo per le donne meno la distanza media per gli uomini. In tal modo, se un nome di professione ha un valore negativo vuol dire che è più associato agli uomini, se ha un valore positivo è più associato alle donne. Utilizzano quindi una regressione lineare per misurare l'associazione fra questa misurazione di *embedding bias*¹⁴¹, calcolata per tutti i nomi di professione considerati, e la percentuale di occupazione femminile reale in ogni professione, e anche con i risultati di una serie di questionari sugli stereotipi di genere condotti fra il 1977 e il 1990. I risultati a cui pervengono è che, in generale, c'è una correlazione lineare fra *women-bias* e percentuale femminile nelle diverse professioni: il primo valore è più elevato, quando la percentuale femminile è inferiore a quella maschile. Nello specifico, lo studio riporta i risultati sulla correlazione fra *embedding bias* nel *corpus* Google news e dati occupazionali pertinenti il 2015. La figura utilizzata per illustrare i risultati (si veda più sotto, Figura

¹⁴¹ Gli autori dell'articolo usano in modo intercambiabile i termini *embedding bias* e *women bias*, evidentemente perché l'indicatore è ottenuto per sottrazione di \overrightarrow{man} da \overrightarrow{woman} , dunque restituisce valori positivi quando c'è un'associazione stereotipata con il genere femminile, viceversa, restituisce un valore negativo.

6) riporta solo alcuni nomi di professione: *carpenter*, *engineer* e *mechanic* con valori di *embedding bias* negativi (dunque stereotipati come maschili), in corrispondenza di percentuali di occupazione femminile basse, che vanno da pochi punti sopra lo 0% per *carpenter*, a valori più alti, ma sempre sotto il 50%, per *engineer* e *mechanic*; nomi di professione come *dancer*, *librarian* e *nurse* con valori di *embedding bias* positivi (dunque stereotipati come femminili) in corrispondenza di percentuali di occupazione femminile elevata, sopra il 50%. Infine, lo stesso grafico riporta il nome *secretary* con un valore di *women-bias* inferiore a 0, molto distante dalla linea di correlazione lineare, indicativo di una sorta di “anomalia”, ovvero un’associazione con il genere maschile, in corrispondenza di una elevata percentuale di occupazione femminile in questa professione.

Figura 6. Correlazione fra associazione con il genere femminile di nomi di professione in uno spazio vettoriale e percentuale di occupazione delle donne nella relativa professione (Garg et al. 2018, E3636)

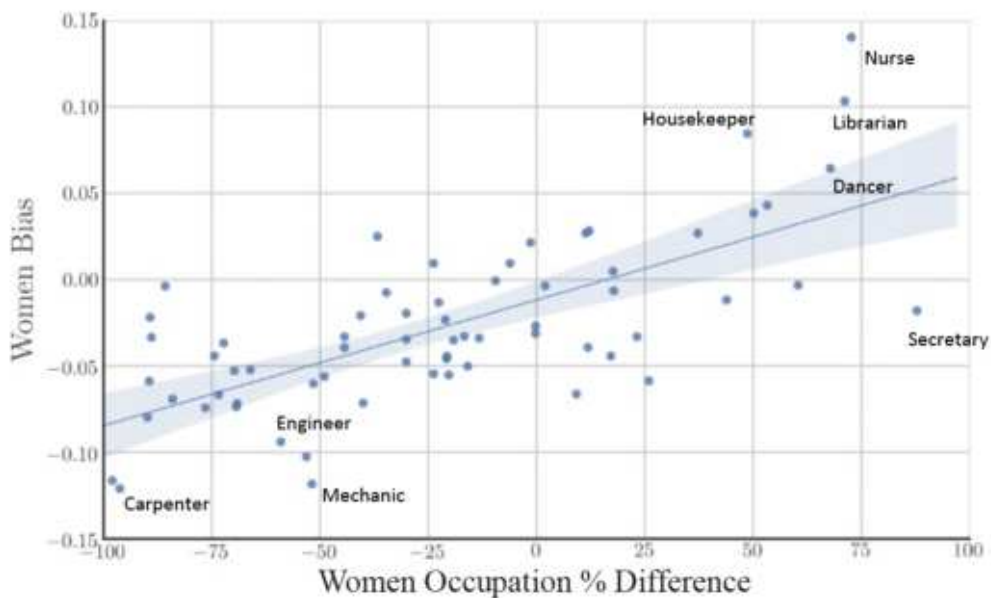


Fig. 1. Women’s occupation relative percentage vs. embedding bias in Google News vectors. More positive indicates more associated with women on both axes. $P < 10^{-10}$, $r^2 = 0.499$. The shaded region is the 95% bootstrapped confidence interval of the regression line. In this single embedding, then, the association in the embedding effectively captures the percentage of women in an occupation.

Gli autori non commentano quest’ultimo risultato, che secondo me è interessante. Il nome *secretary* è infatti caratterizzato da un’ambiguità

semantica¹⁴², potendo indicare sia persone che si occupano di mansioni di ufficio (prima accezione), sia persone che ricoprono cariche pubbliche a livello apicale (seconda accezione). E questa ambiguità semantica è a sua volta interessata da un'asimmetria di genere. In inglese, il termine nella sua prima accezione è stereotipicamente associato al genere femminile, nella seconda al genere maschile, come osservato da Hellinger e Bußman (2001) e attestato da diversi studi di neuro e psicolinguistica (Banaji e Hardin 1996; 2001; Duffy e Keir 2004; Oakill et al. 2005). Tuttavia, considerando che il data set di partenza è Google news, può darsi che la parola *secretary* sia più frequente nella seconda accezione, con riferimento a leader/posizioni apicali di partiti, organizzazioni internazionali, sindacati o associazioni, attestando così una maggiore associazione con il genere maschile. Google news raccoglie infatti notizie dal mondo, e in genere fa più notizia chi ha un ruolo pubblico, di una persona comune (van Zoonen 1994); inoltre, come dichiarano gli autori, il *corpus* Google news, a cui si riferiscono i WEs utilizzati, raccoglie notizie anche precedenti al 2015. Mentre i dati occupazionali di *secretary* fanno riferimento molto probabilmente soltanto all'impiego in lavori di segreteria e, certamente, soltanto all'anno 2015.

Quanto i WEs codifichino questa ambiguità semantica, restituendo una "anomala" rappresentazione di *secretary* è attestato, indirettamente, anche da uno studio di Lewis e Lupyan (2020), che analizza gli stereotipi di genere in WEs di 25 lingue attraverso un *corpus* di sottotitoli (Lison e Tiedemann 2016; van Paridon e Thompson 2019, cit. in Garg et al. 2018) e un *corpus* Wikipedia (Bojanowski et al. 2017). Gli autori riportano che *secretary* ha una più forte associazione di genere (femminile) nei WEs italiani rispetto ai WEs inglesi. Focalizzandosi sulla lingua italiana, attribuiscono questa associazione al fatto che l'italiano, come lo spagnolo,

¹⁴² Queste le definizioni di *secretary* dell'Oxford dictionary online: 1. a person who works in an office, working for another person, dealing with mail and phone calls, keeping records, arranging meetings with people, etc.; 2. an official of a club, society, etc. who deals with mail, keeping records and making business arrangements; (in the UK) the head of an important government department; 4. (in the US) the head of a government department, chosen by the president; (in the UK) an assistant of a government minister, an ambassador, etc.; fonte: <https://www.oxfordlearnersdictionaries.com/definition/english/secretary?q=secretary> (ultimo accesso 19 novembre 2021).

e a differenza dell'inglese, è una lingua dal genere marcato, per cui molti nomi di professione hanno il tratto di genere espresso morfologicamente e questo influenza le rappresentazioni semantiche, come dimostrato da McCurdy e Serbetçi (2020), in un'analisi comparativa fra spagnolo e tedesco, lingue dal genere marcato, e inglese, lingua dal genere naturale. Non indagano però le ragioni per cui l'inglese *secretary* ha una minore associazione con il genere femminile. E arrivano a dimostrare che le lingue che hanno una maggiore marcatezza di genere, come l'italiano e tutte le lingue romanze, hanno parlanti con associazioni di genere implicite più forti, contrariamente, in verità, a quanto dimostrano molti studi di neuro e psicolinguistica citati nel Capitolo 2.

Lo studio di Lewis e Lupyan (2020) è comunque interessante perché verifica che l'ipotesi di Caliskan et al. (2017) è attestabile anche per lingue diverse dall'inglese americano, e indaga il ruolo del genere grammaticale in *embeddings* di nomi di professione in diverse lingue.

Uno studio che analizza in maniera specifica il ruolo del genere grammaticale negli stereotipi di genere sui nomi di professione, attraverso WEs italiani, è quello di un gruppo di ricercatori dell'Università di Padova (Biasion et al. 2020) che ha replicato il WEAT di Caliskan et al. (2017) su due associazioni di genere implicite, uomo-carriera vs. donna-famiglia e uomo-scienze vs. donna-arti, e l'esperimento di Bolukbasi et al. (2016) sui nomi di professione, utilizzando WEs pre-addestrati in una versione migliorativa di word2vec, fastText¹⁴³ su due *corpora*: wiki, basato su Wikipedia 2016 (Bojanowski et al. 2017), e wiki cc, basato su dati di Common Crawl, maggio 2017, e Wikipedia, 11 settembre 2017 (Grave et al. 2018). Lo studio di Biasion et al. (2020) dimostra che il genere grammaticale, in italiano, mitiga gli stereotipi di genere, sia a livello di associazioni implicite uomo-carriera vs. donna-famiglia e uomo-scienze vs. donna-arti, sia a livello di nomi di professione. Per quanto riguarda questi ultimi, lo studio misura l'associazione di genere di 30 nomi di professione di cui 15 epiceni/bigenere e 15 con genere marcato. Basandosi

¹⁴³FastText è una libreria open source, disponibile online all'indirizzo: <https://fasttext.cc> (ultimo accesso 16 novembre 2021).

su studi precedenti, che osservano un'influenza del genere grammaticale nei WEs su nomi inanimati – i nomi con il genere grammaticale uguale risultano più vicini fra loro (Gonen et al. 2019) – e su associazioni di genere cognitive, come per esempio carriera-uomo vs. famiglia-donna (McCurdy e Serbetçi 2020), Biasion e colleghi sviluppano un metodo per cercare di comprendere il ruolo del genere grammaticale in WEs italiani: svolgono due diverse misurazioni dell'associazione di genere dei nomi di professione. Per i 15 nomi epiceni/bigenere¹⁴⁴ calcolano in entrambe le misurazioni un solo *gender-score*. Per i 15 nomi flessivi, nella prima misurazione calcolano un *gender-score* medio (fra la forma femminile e maschile), nella seconda misurazione calcolano due *gender-score* per le due diverse forme. Riportano quindi i risultati in due rappresentazioni grafiche che mostrano una correlazione lineare inversa fra il *gender-score* e la percentuale di occupazione femminile nella professione corrispondente a ciascun nome.

Dal primo grafico (Figura 7., grafico a sinistra) risulta evidente che c'è una correlazione lineare inversa fra *gender-score* e percentuale di donne nelle professioni: al crescere della percentuale di donne in una occupazione corrisponde una diminuzione del *gender-score*. Detto altrimenti, i nomi di professione che hanno una più forte associazione con il genere semantico maschile (es. *elettricista*) sono nomi di professione in cui la percentuale di donne è bassa, viceversa, i nomi di professione che hanno una più forte associazione con il genere semantico femminile (es. *badante*) sono nomi di professione in cui la percentuale di donne è alta. Un altro risultato evidente da questo primo grafico è che i nomi marcati rispetto al genere, anche "degenderizzati" (cioè sottoposti a una misurazione del *gender-score* media, fra forma femminile e maschile) hanno un *gender-score* più basso, vale a dire sono meno stereotipati al maschile. Fatta eccezione per *calzolaio*, attestano tutti un *gender-score* inferiore a 0, indipendentemente dalla percentuale di donne che svolge quella professione. Questo risultato indica un'influenza del genere grammaticale sulle associazioni di genere, come conferma il secondo

¹⁴⁴ Gli autori considerano bigenere tre nomi che in realtà appartengono a classi nominali flessive e quindi hanno la doppia forma, maschile e femminile: *architetto (architetta)*, *ingegnere (ingegnera)*, *medico (medica)*, *notaio (notaia)*.

grafico, su cui Biasion e colleghi non si soffermano per un'analisi dettagliata, dato che va oltre il loro scopo di indagare a quale livello, in generale, una lingua può codificare stereotipi di genere e quanto i WEs possono essere strumenti efficaci per rilevarli. Tuttavia la loro analisi evidenzia bene alcuni fenomeni tipici della relazione fra lingua e genere, su alcuni dei quali mi soffermo brevemente, perché sono stati particolarmente utili nell'impostare il mio studio.

Figura 7. Correlazione fra associazioni di genere in embeddings di nomi i professione e tasso di occupazione femminile, con un'unica forma nominale, a sinistra, con le due forme femminili e maschili, per i nomi dal genere marcato, a destra (Biasion et al. 2020, 5)

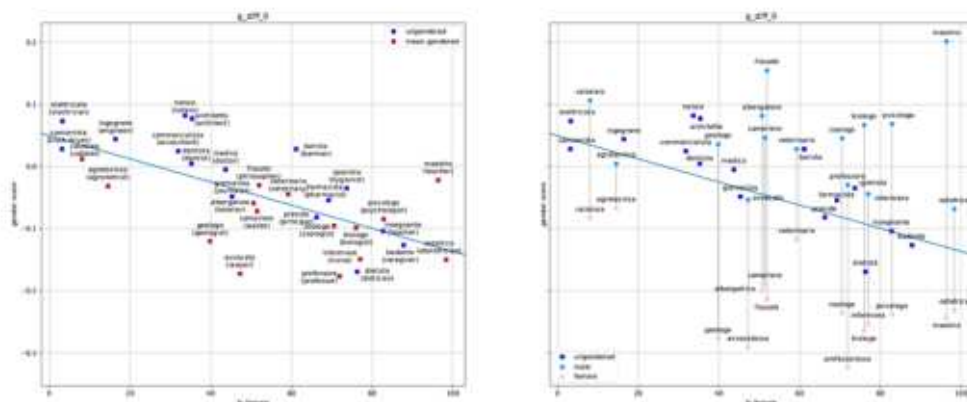


Figure 1: (Left) Gender score of occupations g_{diff} , on the y axis, vs percentage of women in that occupation, represented on the x axis. (Right) Same variables, including both feminine and masculine versions of gendered jobs. Translations, omitted for readability, can be found in the left pane.

Dal secondo grafico (Figura 7. a destra) risulta evidente che, nella maggior parte dei casi, i nomi di professione appartenenti alle classi flessive ottengono *gender-score* abbastanza simmetrici per la forma femminile – collocata sul versante negativo, ovvero dell'associazione con il genere semantico femminile – e la corrispettiva forma maschile – collocata sul versante positivo, ovvero dell'associazione con il genere semantico maschile. Fanno eccezione: *avvocatessa* (l'unica delle due possibili forme femminili considerata dallo studio) e *avvocato*, *professoressa* e *professore*, *ostetrica* e *ostetrico*. In questo caso sia la forma femminile sia la forma maschile hanno un *gender-score* basso, cioè sono semanticamente associate al genere femminile. Se nel caso di *ostetrica/o* questo potrebbe dipendere dal fatto che gli ostetrici (uomini) sono un'esigua minoranza

nella professione, negli altri due casi potrebbe più ragionevolmente dipendere dal fatto che le forme maschili, *avvocato* e *professore* non sono usate solo con riferimento maschile, ma anche femminile, questo ne spiegherebbe la loro associazione semantica con il genere femminile.

Un altro caso interessante riguarda la forma maschile maestro che si colloca in posizione simmetrica con la forma *maestra*: entrambi hanno un *gender-score* di circa 0,2¹⁴⁵ (positivo la prima, negativa la seconda), in corrispondenza di una professione con un'occupazione femminile prossima al 100%. Detto altrimenti, la forma maschile attesta un'associazione semantica con il genere maschile, tanto quanto la forma femminile attesta un'associazione semantica con il genere femminile, a fronte di una realtà sociale in cui i maestri di scuola sono una esigua minoranza. Questo risultato potrebbe dipendere dal fatto che *maestra/o* è una parola interessata contestualmente da un'ambiguità semantica e un'asimmetria di genere, come *segretaria/o*: la forma femminile è infatti usata in modo simmetrico alla forma maschile per indicare chi svolge la professione di insegnante di scuola dell'infanzia o elementare, mentre la forma maschile è usata per designare ruoli di prestigio, anche nel caso in cui siano ricoperti da donne, come riportato dal vocabolario Treccani online: «chi eccelle in un'arte, in una scienza, in una disciplina, o in singole forme d'arte e manifestazioni di cultura, così da poter essere considerato una guida, un caposcuola», e «nella pratica musicale, titolo attribuito, fin dal primo medioevo, all'istruttore e direttore di un insieme di cantori o di strumentisti»¹⁴⁶.

7.3 Ipotesi e domande di ricerca

L'ipotesi da cui sono partita, sulla base di quanto attestato sia dalla ricerca su lingua e genere, in particolare sui nomi di professione e carica politica, sia dallo stato dell'arte appena descritto, è che i WEs del *corpus* ACR attestino la persistenza di

¹⁴⁵ I dati non sono riportati in modo specifico né nell'articolo, né nel grafico, che riporta il *gender-score* in intervalli di valore di 0,1, in un intervallo complessivo che va da -0,35 a +0,25.

¹⁴⁶ Fonte: <https://www.treccani.it/vocabolario/maestro1/>, ultimo accesso 19 novembre 2021.

stereotipi di genere pertinenti la sfera professionale femminile (e maschile). Per verificare questa ipotesi ho formulato le seguenti domande:

- I. qual è l'associazione semantica di genere negli *embeddings* di una selezione di nomi di professioni e cariche politiche estratti dal *corpus* ACR?
- II. C'è una correlazione fra questa associazione e la realtà sociale, ovvero la partecipazione femminile nei diversi tipi di professione considerata?
- III. Quale ruolo ha il genere grammaticale dei nomi nel confermare associazioni stereotipiche o mitigarle?

7.4 Metodologia

Per rispondere alle domande sopra formulate, ho anzitutto selezionato una lista di nomi di professione e carica politica. Per i nomi di professione, sono partita dalla lista di Biasion et al. (2020), scartando le professioni per cui non sono riuscita a trovare dati recenti sulla percentuale di occupate e occupati in quella professione¹⁴⁷, e anche *professoressa/professore* termine interessato da un'asimmetria semantica a seconda dell'ambito di riferimento, come attestato da Nardone (2018, 235), che scrive: «si può notare una sorta di opposizione binaria tra il mondo scolastico associato al sostantivo *professoressa* e il mondo universitario associato al sostantivo *professore*». Data questa asimmetria, sarebbe stato necessario disambiguare a priori, manualmente, il riferimento specifico, un lavoro troppo oneroso su un *corpus* di così ampie dimensioni. Per quanto riguarda i nomi di cariche politiche, ho considerato i principali ruoli istituzionali nazionali, *ministra/o, senatrice/tore, deputata/o*, più la carica locale *sindaca/o*.

La tabella seguente riporta l'elenco dei 26 nomi considerati, specificandone la categoria grammaticale in termini di marcatezza (nome dal genere marcato) o meno (nome epiceno/bigenere), riportando anche la percentuale di donne presenti in quella professione/carica politica.

¹⁴⁷ *agrotecnica/o, barista, biologa/o, cameriera/o, dietista, filosofa/o, geologa/o, professore/essa, zoologa/o*.

Tabella 62. Elenco dei nomi di professione e carica politica

Id	Nome di professione/carica	Categoria grammaticale	Occupazione femminile
1	Albergatrice/tore	marcato	30%
2	Architetta/o	marcato	41%
3	Avvocata essa/o	marcato	48%
4	Badante	epiceno/bigenere	89%
5	Calzolaia/o	marcato	10%
6	Camionista	epiceno/bigenere	6%
7	Commercialista	epiceno/bigenere	33%
8	Dentista	epiceno/bigenere	27%
9	Deputata/o	marcato	32%
10	Elettricista	epiceno/bigenere	4%
11	Farmacista	epiceno/bigenere	69%
12	Giornalista	epiceno/bigenere	42%
13	Igienista	epiceno/bigenere	76%
14	Infermiera/e	marcato	77%
15	Ingegnera/e	marcato	28%
16	Insegnante	epiceno/bigenere	82%
17	Maestra/o	marcato	97%
18	Medica/o	marcato	45%
19	Ministra/o	marcato	29%
20	Notaia/o	marcato	37%
21	Ostetrica/o	marcato	99%
22	Preside	epiceno/bigenere	66%
23	Psicologa/o	marcato	83%
24	Senatrice/tore	marcato	34%
25	Sindaca/o	marcato	14%
26	Veterinaria/o	marcato	47%

Una volta selezionati i nomi, ho creato i loro *embeddings* dal *corpus* ACR, applicando il modello Skip-gram con *negative sampling* (Mikolov et al. 2013), scegliendo un intervallo simmetrico di ± 5 parole (a sinistra e a destra della parola target) e fissando a $N=5$ l'occorrenza minima dei nomi, e a 100 la dimensione vettoriale. Questo modello, è stato applicato sulla lingua italiana con successo in diversi studi: Berardi et al. (2015), Lomonaco (2015), Bojanowski et al. (2017), Marelli (2017), Tripodi e Li Pira (2017). Per una rassegna completa si veda Lenci et al. (2021).

Per rispondere alla prima domanda di ricerca, ho calcolato il *gender-score* dei nomi selezionati, seguendo l'approccio di Bolukbasi et al. (2016), che quantifica l'associazione di genere di una lista di titoli professionali in termini di *gender-*

direction rispetto alla differenza fra \vec{he} e \vec{she} , calcolandone la distanza in termini di coseno fra vettori. Seguendo Biasion et al. (2020) ho scelto di assumere come corrispondenti dei pronomi personali inglesi *he* e *she* gli italiani *lui* e *lei* e calcolato la *gender-direction* come differenza fra \vec{lui} e \vec{lei} , e il *gender-score* come distanza (coseno) della parola target (il nome di professione o carica politica) dal valore della *gender-direction*. In questo modo, un *gender-score* elevato indica una più forte associazione del nome al genere semantico maschile, mentre un *gender-score* basso un'associazione del nome al genere semantico femminile.

Per rispondere alla seconda e alla terza domanda di ricerca, ho calcolato il coefficiente di correlazione lineare fra la proporzione di donne nelle professioni e il *gender-score* in due condizioni diverse: 1) considerando tutti i nomi nel loro complesso, senza distinguere la forma femminile e maschile per i nomi dal genere marcato (per cui ho calcolato un *gender-score* medio fra la forma femminile e maschile); 2) distinguendo la forma femminile e maschile per i nomi dal genere marcato. Il primo coefficiente consente di verificare se c'è una corrispondenza o meno fra il *gender-score* semantico dei nomi di professione e il reale stato occupazionale delle donne (e degli uomini) nella corrispettiva posizione professionale o politica. Il secondo consente di verificare il ruolo del genere grammaticale.

7.5 Risultati

I risultati nella Tabella 63 riportano il *gender-score* dei nomi di professioni e cariche politiche in ordine dal più alto, vale a dire il più fortemente associato al genere maschile, al più basso, il più fortemente associato al genere femminile. Quattro sono i nomi, in coda alla lista in tabella, per cui il calcolo del *gender-score* non ha restituito alcun valore (NaN¹⁴⁸). Fatta eccezione per questi ultimi termini, il calcolo del *gender-score* attesta come primo risultato diversi gradi di associazione con il

¹⁴⁸ Not a Number.

genere maschile o femminile variabili entro un valore massimo di 0,147, per *ingegnera/e*, e un valore minimo di -0,119, per *infermiera/e*, a seconda dei nomi.

Tabella 63. *Gender-score* dei nomi di professione e carica politica e tasso di occupazione femminile nella professione o carica relativa (in valore decimale) ¹⁴⁹

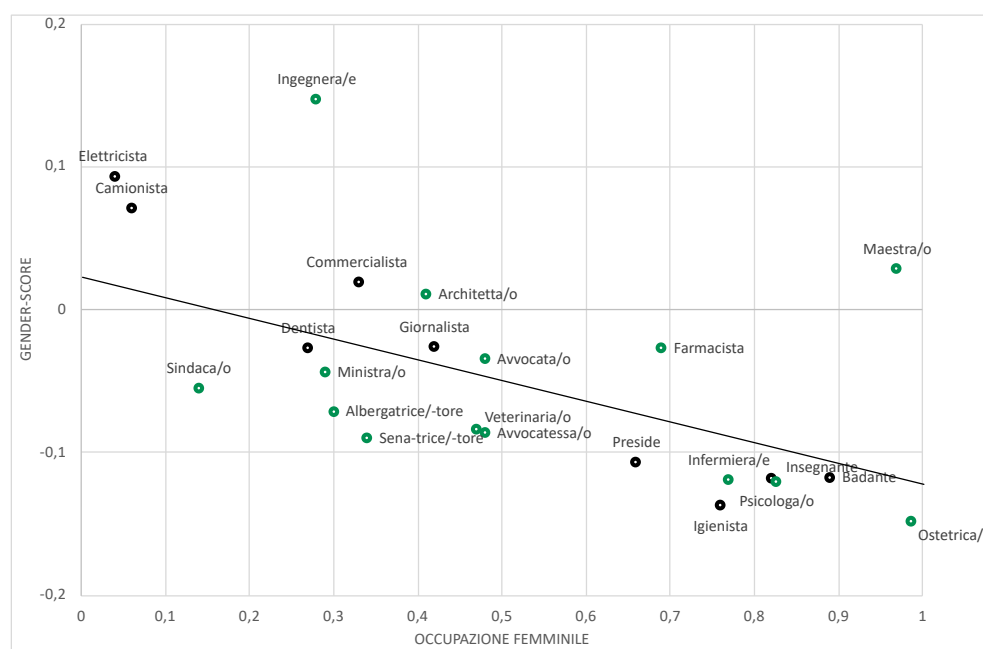
Id	Nome	Categoria nome	Gender-score	Occupazione femminile
5	Calzolaia/o	marcato	NaN	0,10
9	Deputata/o	marcato	NaN	0,32
18	Medica/o	marcato	NaN	0,45
20	Notaia/o	marcato	NaN	0,37
21	Ostetrica/o	marcato	-0,148	0,99
13	Igienista	epiceno/bigenere	-0,138	0,76
23	Psicologa/o	marcato	-0,120	0,83
14	Infermiera/e	marcato	-0,119	0,77
4	Badante	epiceno/bigenere	-0,118	0,89
16	Insegnante	epiceno/bigenere	-0,118	0,82
22	Preside	epiceno/bigenere	-0,107	0,66
24	Senatrice/tore	marcato	-0,090	0,34
3(b)	Avvocatessa/o	marcato	-0,087	0,48
26	Veterinaria/o	marcato	-0,083	0,47
1	Albergatrice/tore	marcato	-0,071	0,30
25	Sindaca/o	marcato	-0,055	0,14
19	Ministra/o	marcato	-0,044	0,29
3(a)	Avvocata/o	marcato	-0,034	0,48
8	Dentista	epiceno/bigenere	-0,027	0,27
11	Farmacista	epiceno/bigenere	-0,027	0,69
12	Giornalista	epiceno/bigenere	-0,026	0,42
2	Architetta/o	marcato	0,010	0,41
7	Commercialista	epiceno/bigenere	0,019	0,33
17	Maestre/o	marcato	0,029	0,97
6	Camionista	epiceno/bigenere	0,071	0,06
10	Elettricista	epiceno/bigenere	0,094	0,04
15	Ingegnera/e	marcato	0,147	0,28

Una prima lettura della tabella mostra come i nomi più associati al genere maschile corrispondono prevalentemente a professioni in cui le donne sono meno degli uomini, fatta eccezione per *maestra/o*; viceversa i nomi più associati al genere femminile corrispondono prevalentemente a professioni in cui le donne sono più degli uomini, fatta eccezione per *senatrice/tore*. Questa lettura suggerisce che vi possa essere una correlazione fra il *gender-score* e l'incidenza femminile in una

¹⁴⁹ Per poter confrontare le due serie di dati, le percentuali di occupazione femminile sono state convertite in numeri decimali.

professione. Per verificare questa ipotesi, formulata nella seconda domanda di ricerca, ho calcolato il coefficiente di correlazione lineare fra le due variabili. Il risultato ottenuto è un coefficiente $r=-0,615$ ($p<0,05$) indicativo di una correlazione inversa di media entità (Ratner 2009): al crescere della percentuale di donne in una professione corrisponde tendenzialmente una diminuzione del *gender-score*, ovvero una riduzione dell'associazione del nome di quella posizione professionale o politica al genere maschile, e una corrispettiva maggiore associazione al genere femminile. Come mostra in modo abbastanza chiaro il Grafico 25, la maggior parte dei nomi si distribuisce attorno alla retta di regressione che spiega la correlazione fra le due variabili indagate. Si distanziano "visibilmente" dalla retta due nomi: *ingegnera/e* e *maestra/o*: attestando, il primo, un *gender-score* più elevato di altri nomi corrispondenti a professioni con una percentuale di donne simile (attorno al 30%), come *commercialista*, *dentista*, *ministra/o*, *albergatrice/tore*, *senatrice/tore*, o più alta, come *camionista* e *elettricista*; il secondo, un *gender-score* più elevato rispetto a nomi che, come *maestra/o*, corrispondono a professioni con un'elevata percentuale di donne, per esempio *ostetrica/o* e *badante*.

Grafico 25. Correlazione fra *gender-score* dei nomi di professione/carica politica e tasso di occupazione femminile: $r=0,615$, $p<0,05$



Per verificare il ruolo del genere grammaticale nella correlazione fra il *gender-score* dei nomi di posizioni professionali o politiche e la percentuale delle

donne nella professione o carica politica corrispondente (terza domanda di ricerca), per i nomi marcati ho anzitutto calcolato il *gender-score* delle due forme, femminile e maschile.

Tabella 64. *Gender-score* delle forme femminili e maschili dei nomi di professione e carica politica

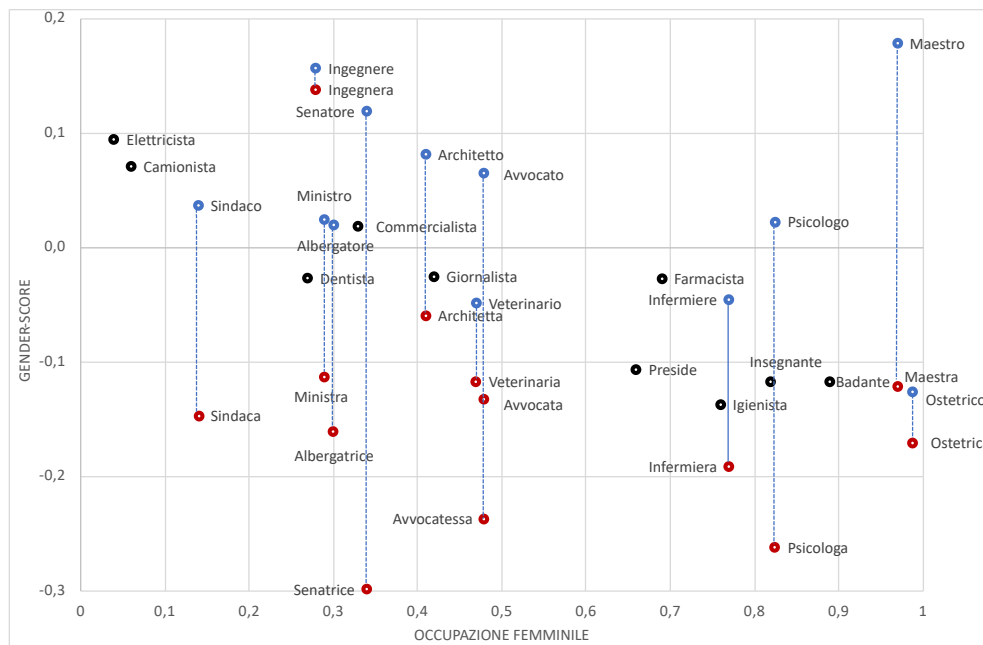
Id	Nome	Gender-score f. femminile	Gender-score f. maschile
1	Albergatrice/tore	-0,162	0,020
2	Architetta/o	-0,060	0,081
3(a)	Avvocata/o	-0,133	0,065
3(b)	Avvocatessa/o	-0,238	0,065
14	Infermiera/e	-0,192	-0,046
15	Ingegnera/e	0,138	0,157
17	Maestra/o	-0,121	0,178
19	Ministra/o	-0,113	0,024
21	Ostetrica/o	-0,171	-0,126
23	Psicologa/o	-0,263	0,023
24	Sena-trice/tore	-0,299	0,119
25	Sindaca/o	-0,147	0,037
26	Veterinaria/o	-0,118	-0,049

I risultati, riportati nella Tabella 64, evidenziano come tutte le forme femminili dei nomi dal genere marcato abbiano un *gender-score* più basso rispetto alle corrispondenti forme maschili, anche se in misura diversa a seconda del nome: per esempio, *ingegnera* ha un *gender-score* più basso rispetto a *ingegnere*, ma con un valore comunque alto, indicativo di una forte associazione con il genere maschile. Sul versante, opposto, *ostetrico* ha un *gender-score* più alto rispetto a *ostetrica*, ma con un valore comunque basso, indicativo di una forte associazione con il genere femminile. Diverso invece è il comportamento di *senatrice/tore* o *maestra/o* le cui forme maschili hanno, in entrambi i casi, un *gender-score* alto, indicativo di una associazione con il genere maschile, e, simmetricamente, le forme femminili un *gender-score* basso, indicativo di una associazione con il genere femminile. Anche se nel complesso, sembrerebbe che la forma femminile *senatrice* contribuisca ad abbassare il *gender-score* medio della coppia *senatrice-senatore*, mentre *maestra* non contribuisca ad abbassare il *gender-score* medio di *maestra-maestro*.

Interessante è la diversa associazione di genere delle due forme femminili corrispondenti al maschile *avvocato*: *avvocata* e *avvocatessa*. La prima, raccomandata da Sabatini (1987), ma anche successivamente (Cavagnoli 2013), in particolare dalle linee guida destinate ai giornalisti e alle giornaliste (Robustelli 2014; CNOG 2015) ha un *gender-score* più alto, rispetto alla seconda che è più associata al genere semantico femminile.

Dopo aver calcolato il *gender-score* separatamente per le forme femminili e maschili dei nomi dal genere marcato, ho anche calcolato separatamente il coefficiente di correlazione per i nomi epiceni/bigenere e per i nomi dal genere marcato. Il primo (nomi epiceni/bigenere) è risultato pari a un valore $r=-0,927$ ($p<0,001$) indicativo di una forte correlazione inversa; il secondo è risultato pari a un valore $r=-0,339$ ($p=0,26$) indicativo di una correlazione, sempre inversa, ma di debole entità e non significativa.

Grafico 26. Correlazione fra *gender-score* dei nomi di professione/carica politici distinti fra marcati ($r=-0,339$, $p=0,26$) e non marcati ($r=-0,339$, $p=0,26$) e tasso di occupazione femminile



7.6 Discussione dei risultati

I risultati presentati nel paragrafo precedente dimostrano, in generale, che i diversi nomi di professioni e cariche politiche selezionati occorrenti nel *corpus* ACR sono associati in diversa misura con il genere femminile o maschile. Nella maggior parte dei casi, i nomi corrispondenti ad ambiti professionali in cui le donne sono percentualmente inferiori rispetto agli uomini, come per esempio *ingegnera/e*, *elettricista*, *camionista*, sono più associati al genere maschile, e nomi corrispondenti ad ambiti professionali in cui le donne sono percentualmente superiori rispetto agli uomini, come per esempio *infermiera/e*, *ostetrica/o* e *igienista*, sono più associati al genere femminile.

In effetti, il coefficiente di correlazione fra il *gender-score* dei nomi e la percentuale di donne nella professione o carica corrispondente indica una correlazione inversa di media entità fra le due variabili: al crescere della percentuale di donne in una professione corrisponde, tendenzialmente, una diminuzione dell'associazione semantica con il genere maschile (e una crescita dell'associazione con il genere femminile). Le associazioni di genere latenti nel *corpus* ACR sembrano riflettere, dunque, almeno in parte, una conoscenza della realtà (Caliskan et al. 2017), nella fattispecie la diversa distribuzione di donne e uomini nelle diverse posizioni professionali o politiche considerate.

La conoscenza della realtà non sembra però l'unica informazione codificata dal *corpus* ACR. Se consideriamo la posizione dei nomi (Grafico 25) rispetto alla retta di regressione, alcuni risultano tanto distanti da poter essere considerati *outlier*, valori anomali, che non risultano spiegati dalla correlazione. Due in particolare sono i nomi che possono considerarsi *outlier*: *ingegnera/e* e *maestra/o*. Il primo ha un *gender-score* particolarmente elevato (0,147), più elevato di nomi come *elettricista* (0,094) o *camionista* (0,071) corrispondenti a professioni in cui le donne sono una percentuale molto bassa, rispettivamente 4% e 6% vs. il 28% delle ingegnere. Il secondo ha un *gender-score* elevato (0,029), indicativo di un'associazione prevalente con il genere maschile, nonostante faccia riferimento a

una professione in cui la percentuale di donne è molto alta (97%). Le professioni che hanno percentuali di occupazione femminile simili, come ostetrica/o (98,7%) o badante (89%), hanno *gender-score* molto più bassi (rispettivamente -0,149 e -0,118). Essendo entrambi nomi dal genere marcato, la loro “anomalia” non può essere ricondotta al genere grammaticale, anche se come discuterò di seguito, i risultati attestano che il genere grammaticale ha un ruolo nel mitigare le associazioni con entrambi i generi semantici (femminile e maschile), bensì a stereotipi codificati nella lingua, in coerenza con quanto dimostrato dagli studi sopra citati (cfr. § 7.2).

Per quanto riguarda *maestra/o* una possibile interpretazione è quella già argomentata più sopra nella presentazione dello studio di Biasion et al. (2020), che attesta, senza argomentarla, un medesimo risultato. Il termine *maestra/o* è caratterizzato da un’ambiguità semantica interessata contestualmente da un’asimmetria di genere: l’accezione del termine che riferimento agli/lle insegnanti di scuola è caratterizzata da un uso simmetrico delle forme femminile e maschile, mentre l’accezione del termine che fa riferimento a «chi eccelle in un’arte, in una scienza, in una disciplina, o in singole forme d’arte e manifestazioni di cultura, così da poter essere considerato una guida, un caposcuola», e «nella pratica musicale, titolo attribuito, fin dal primo medioevo, all’istruttore e direttore di un insieme di cantori o di strumentisti»¹⁵⁰, per riprendere la definizione del vocabolario Treccani, è spesso usata al maschile anche per designare una donna. Probabilmente il *corpus* ACR codifica questa informazione, e questo spiega la posizione di *outlier* di *maestra/o*, in corrispondenza di una percentuale femminile che riguarda solo la professione a cui si riferisce la prima accezione del termine. Per quanto riguarda *ingegnera/e* l’unica interpretazione che mi sembra possibile avanzare è che il termine sia interessato da un’associazione semantica con il genere maschile così forte da resistere all’effetto di mitigazione del tratto di genere che si esprime in modo trasparente sulla morfologia del nome.

¹⁵⁰ Fonte: <https://www.treccani.it/vocabolario/maestro1/>; ultimo accesso 19 novembre 2021.

Venendo ora all'ultima domanda di ricerca focalizzata proprio sul ruolo della grammatica italiana, in particolare del suo potere di rendere visibili le donne attraverso il tratto di genere marcato a livello morfologico, fatta eccezione per *ingegnera/e*, nel complesso questa indagine dimostra che la correlazione fra le associazioni di genere (maschili o femminili che siano) codificate da (una selezione) di nomi di professioni e cariche politiche e la percentuale di donne nella professione e carica corrispondente è forte nel caso dei nomi epiceni/bigenere, mentre è debole per i nomi dal genere marcato. Analizzando il Grafico 25 risulta evidente come le forme femminili e maschili si collochino (quasi) tutte in modo simmetrico rispettivamente sul versante dell'associazione con il genere femminile o maschile indicato dal *gender-score*, tranne le già citate coppie *ingegnera-ingegnere*, *maestro-maestra*, e, sul versante opposto, le coppie *veterinaria-veterinario* e *ostetrica-ostetrico*, le cui forme maschili rimangono comunque fortemente associate al genere semantico femminile. Il fatto che in entrambi questi ultimi casi siamo in presenza di termini che possono avere anche una funzione aggettivale suggerisce che questo risultato possa essere ricondotto a una codifica che sovrappone i due valori di entrambi i termini (aggettivo e sostantivo), restituendo un risultato "sporco".

7.7 Conclusioni

L'analisi delle associazioni di genere (femminile e maschile) nel *corpus* ACR, tramite un approccio distribuzionale, ha evidenziato che i nomi di professioni o cariche politiche codificano informazioni sul genere, in modo del tutto simile a quanto attestato da Biasion et al. (2020). Queste informazioni riflettono in parte una conoscenza della realtà, ovvero la diversa ripartizione di donne e uomini nelle posizioni professionali o politiche corrispondenti ai nomi selezionati; in parte riflettono però anche stereotipi di genere, ovvero associazioni di genere sedimentate nel corso del tempo che possono arrecare danni simbolici o allocativi (Barocas et al. 2017, 2019; Crawford 2017), alle donne così come agli uomini. Se assumiamo l'ipotesi per cui tutti gli stereotipi originano da un "nocciolo di verità",

come sostengono Prothro e Melikina (1953), la correlazione fra associazioni di genere e dati di realtà si può considerare un risultato atteso. Ma gli stereotipi si distanziano dalla realtà nella misura in cui irrigidiscono posizioni sociali acquisite nel corso del tempo, rendendo difficile, specialmente per quanto riguarda i rapporti di genere, mutamenti nella direzione di una maggiore uguaglianza sostanziale fra donne e uomini. Di questo “irrigidimento”, che in termini linguistici può considerarsi un “sedimento”, sembrano essere testimoni in particolare due nomi: sul versante maschile, *ingegnera/e*, e sul versante femminile *ostetrica/o*. Sebbene la forma femminile *ingegnera* attesti una maggiore associazione con il genere semantico femminile rispetto alla corrispondente forma maschile, in modo coerente con la relazione che intercorre fra genere grammaticale e genere semantico, quella dell'*ingegnera/e* rimane una professione codificata nel linguaggio giornalistico come una professione maschile. Lo stesso, ma sul versante opposto, si può dire per la professione dell'*ostetrica/a*: sebbene la forma maschile attesti una maggiore associazione con il genere maschile, rispetto alla corrispondente forma femminile, *ostetrica/o* rimane un nome più associato al genere femminile.

Come approfondito nel Capitolo 2 (§ 2.3), la struttura della grammatica italiana è funzionale a disambiguare il genere semantico dei nomi con riferimento umano a livello morfo-sintattico. L'approccio distribuzionale adottato in questo studio dimostra, in modo coerente con quanto dimostrato da Biasion et al. (2020), che questa caratteristica della grammatica italiana mitiga le associazioni codificate nel linguaggio, e basate su pregiudizi sociali, in molti casi, associando in modo congruente la forma femminile al genere semantico femminile e la forma maschile al genere semantico maschile. Questa evidenza è in linea con quanto attestato da diversi esperimenti di neuro e psicolinguistica che dimostrano, da un lato, che i nomi con riferimento umano attivano stereotipi di genere (Banaji e Hardin 1996; Osterhout et al. 1997; Garnham et al. 2002; Kennison e Trofe 2003; Sturt 2003; Oakill et al. 2005), anche nella lingua italiana (Cacciari e Padovani 2007) e, dall'altro, che, nelle lingue dal genere marcato l'uso delle forme femminili, le meno diffuse nell'uso dell'italiano, data la tendenza a usare forme maschili anche per nominare

le donne, possono inibire l'attivazione degli stereotipi di genere (Duffy e Keir 2004; Kreiner et al. 2008; Richy e Burnett 2021) e, sia nei bambini che negli adolescenti, mitigare i pregiudizi di genere sulle aspettative di carriera e sulle competenze professionali (Carreiras et al. 1996; Irmen e Roßberg 2004; Vigliocco et al. 2005; Irmen e Schuman 2011; Esaulova et al. 2013). Guardati da questa prospettiva i risultati di questo studio confermano come il linguaggio giornalistico non si discosti dal linguaggio di un *corpus* meno specialistico, come quello di Wikipedia analizzato da Biasion et al. (2020), codificando stereotipi di genere, che tuttavia possono essere messi in discussione dalla scelta di usare la lingua italiana in modo congruente rispetto a un sistema grammaticale che prevede la declinazione dei nomi di posizione professionale o carica politica in modo coerente rispetto al genere semantico del/la referente.

Capitolo 8 Conclusioni

Obiettivo di questa tesi è stato analizzare le rappresentazioni di genere dentro e attraverso il linguaggio dell'informazione italiana, con un approccio in grado di integrare l'analisi linguistica e l'analisi mediale nella comune prospettiva di genere. Al fine di raggiungere questo obiettivo ho costruito tre *corpora* linguistici raccogliendo testi esemplificativi del linguaggio dell'informazione quotidiana, televisiva (*corpus* TG) e a stampa (*corpus* CRS e *corpus* ACR), in modo da circoscrivere l'analisi a un genere specifico, di cui tenere in considerazione le peculiarità, sia linguistiche che mediali.

Le tre ricerche attraverso le quali ho analizzato questi *corpora* presentano caratteristiche diverse fra loro, sotto diversi punti di vista, come è emerso nei capitoli 5-7. Nel primo paragrafo di questo ultimo capitolo (§ 8.1), proverò a tracciare alcune conclusioni generali che colleghino fra loro le diverse prospettive ed evidenze emerse dalle tre analisi svolte e fra queste e il quadro di riferimento teorico e contestuale delineato nei capitoli introduttivi 1-3. Le tre ricerche si differenziano infatti prevalentemente per i diversi usi linguistici indagati, l'approccio metodologico e le tecniche di analisi impiegate, ma possono essere tutte ricondotte al medesimo *framework* teorico delle rappresentazioni sociali (cfr. Capitolo 2, § 2.2 e 2.3; Capitolo 3, § 3.3) e allo stesso contesto nazionale del periodo compreso fra il 2018 e il 2020.

Nel secondo paragrafo (§ 8.2), individuo alcuni limiti di questa tesi, a partire dai quali delinea alcuni sviluppi possibili per approfondire, e ampliare, una linea di ricerca che indaghi le rappresentazioni di genere nei contenuti dei media, in particolare nell'informazione, con un approccio interdisciplinare fra linguistica e *media studies*.

8.1 Risultati principali della tesi

Tutte le tre ricerche presentate rilevano rappresentazioni di genere in mutamento, attestando asimmetrie e stereotipi osservate in precedenti ricerche, ma anche interessanti segnali di attualità e innovazione, correlati, laddove è stato verificato, con diverse linee editoriali e/o diversi contesti politico-istituzionali, più o meno impegnati sul fronte delle pari opportunità di genere, anche linguistiche.

La prima ricerca (Capitolo 5) ha esplorato il *corpus* TG seguendo un approccio *corpus-driven* evidenziando come le donne, in quanto fonti di informazione, dunque parlanti, sono anzitutto il soggetto di un discorso identitario autoreferenziale che non ha un corrispettivo maschile. Questo risultato mostra come il linguaggio veicolato dall'informazione rifletta un'asimmetria semantica tipica dell'italiano per cui il nome *donna* ha un'accezione ristretta al genere femminile, e in questa accezione viene usata dalle parlanti, mentre il corrispettivo *uomo* ha un'accezione più estesa, riferendosi non solo al genere maschile, ma anche alla specie umana, e, probabilmente per questo, non viene usato dagli uomini per marcare la propria identità di genere, o semplicemente auto-rappresentarsi. Questa asimmetria fra fonti maschili e fonti femminili è anche indicativa di come l'informazione rappresenti le donne come un genere specifico entro una costruzione sociale della realtà in cui gli uomini, invece, non hanno tratti di genere peculiari ma sono rappresentativi della società *tout court*. In questo senso, l'androcentrismo testimoniato dal linguaggio dell'informazione italiana assume i caratteri di una rappresentazione pubblica della società, qual è quella veicolata dai media, che coincide ancora ampiamente con l'universo maschile. Come confermano i risultati sui profili lessicali e sui campi semantici delle fonti femminili e maschili a confronto.

Il lessico delle donne rinvia infatti a una dimensione prevalentemente privata mentre quello maschile prevalentemente a una dimensione pubblica. Questa diversità lessicale non è riconducibile a una differenza *tout court* fra il parlato delle donne e quello degli uomini, bensì a una diversa rappresentazione

mediale. Entrambi i parlari contribuiscono alla formazione di due campi semantici dominanti, quello delle “testimonianze di vita”, in cui si concentra il mondo del privato, che è nettamente prevalentemente nel discorso veicolato dalle fonti femminili (con una copertura dell’80,25% dei segmenti di testo di fonti femminili vs. il 41,92% dei ST di fonti maschili); e quello della “comunicazione politica”, in cui si concentra la sfera pubblica, che è, non esclusivamente, ma prevalentemente maschile (questa classe copre il 29,96% dei ST di fonti maschili vs. il 19,75% dei ST di fonti femminili).

Questo risultato conferma come l’informazione contribuisca a una rappresentazione stereotipata, che relega le donne alla trattazione di argomenti «out of power», per dirlo con le parole di Robin Lakoff (1973, 53), nonostante il fatto che, rispetto agli anni Settanta, le donne in Italia abbiano certamente fatto progressi in termini di maggiore presenza nella vita pubblica, sia politica, dove con la 18esima legislatura hanno raggiunto una rappresentanza parlamentare complessivamente, cioè a camere unite, pari al 35% (Senato della Repubblica 2018), sia lavorativa, dove a fine 2020 hanno raggiunto un tasso di attività del 54,1% e di occupazione del 48,6%¹⁵¹. Se, entro il quadro teorico delle rappresentazioni sociali, ipotizziamo che sia la lingua, sia i media attestino, e al contempo contribuiscano a, una costruzione della realtà sociale pubblicamente condivisa, interagendo fra loro, questo risultato mostra come tre fra i più seguiti telegiornali nazionali partecipino attivamente a una costruzione stereotipata del parlato delle donne (come già notato da Berretta 1983). E contestualmente testimonia una collocazione simbolica delle donne “out of the power of the news media”, ovvero fuori dal potere di ottenere una rappresentazione bilanciata e non stereotipata, come confermano anche i risultati sulla rappresentazione di genere in termini di spazio di parola. Il *sub-corpus* che raggruppa i testi prodotti dalle fonti femminili ha infatti una dimensione pari a circa un terzo rispetto al *sub-corpus* dei testi delle fonti maschili, confermando la “regola del terzo” rilevata da molti studi nazionali e

¹⁵¹ Dati ISTAT (serie storiche) al 31 dicembre 2020; fonte: <https://www.istat.it/it/archivio/253019> (ultimo accesso 19 novembre 2021).

internazionali che attestano una sorta di “tetto di attenzione” a un terzo dello spazio o della voce assegnata alle donne nelle notizie della maggior parte dei paesi del mondo (Djerf-Pierre e Esdröm 2020b, 33).

Le analisi comparative per anno e per TG dimostrano tuttavia che una diversa linea editoriale può fare la differenza. In particolare, il TG2 si distingue per contribuire a una maggiore rappresentazione delle fonti femminili come soggetto di “comunicazione politica” e per non contribuire alla rappresentazione delle fonti femminili come soggetto di “testimonianze di vita”. Poiché il TG2 nel 2018 è stato diretto da una giornalista, dichiaratamente attenta alla parità di genere, questo risultato può ragionevolmente essere attribuito a un orientamento della redazione, ragionevolmente lasciato in eredità dalla direttrice, verso un impegno per una maggiore e migliore rappresentazione di genere. Come abbiamo visto nel Capitolo 3, le giornaliste possono infatti fare una informazione differente, ma a condizioni variamente correlate con la loro posizione di potere entro gli organi di informazione in cui operano, alla loro “consapevolezza del genere” e anche al contesto del paese.

L’analisi della rappresentazione delle donne, e degli uomini, come oggetto/argomento delle notizie ha evidenziato una rappresentazione in buona parte coerente con quella sino a qui sintetizzata. Nel *corpus* TG donne e uomini sono topicalizzati in modo equilibrato dal punto di vista dello spazio di attenzione che ottengono e anche sotto il profilo lessicale, per via della prevalenza di un’ampia tematizzazione cronachistica caratterizzata da un lessico dominante e comune ai due generi. Questa tematizzazione tuttavia, analizzata sotto il profilo sintagmatico, evidenzia diverse asimmetrie di genere. Le donne sono prevalentemente rappresentate come vittime, dirette o indirette, di violenza contro le donne e domestica, mentre gli uomini come soggetti attivi di un più ampio ambito criminoso e violento, dove emergono non solo come criminali, in modo riflessivo rispetto a un

ambiente criminoso nazionale appannaggio maschile¹⁵², ma anche come rappresentanti delle forze dell'ordine e dei soccorsi.

Una dissimmetria, sempre sbilanciata verso uno specifico universo femminile e un sovraesteso mondo maschile, emerge anche in relazione alle tematiche di genere stabili lungo l'asse diacronico del periodo considerato (2018-2020), vale a dire l'occupazione femminile e le pari opportunità e i diritti. Il primo è emerso come campo semantico che tematizza le donne in modo specifico e le include come oggetto, piuttosto che come soggetto del discorso. Il secondo invece è emerso come campo semantico che topicalizza gli uomini in modo più esteso, come soggetto, piuttosto che come oggetto di un discorso che riguarda più in generale le pari opportunità e i diritti umani. Su questo versante è possibile collocare anche il risultato pertinente la "centralità delle donne", un'area semantica che pone al centro del discorso mediale il ruolo delle donne nella società, nella storia, nella politica, e nell'arte, ma che è significativamente associata alle fonti maschili. Questo risultato conferma una (certa) tendenza dell'informazione a legittimare la valorizzazione delle donne attraverso la parola degli uomini (cfr. Capecchi 2001; CNEL 2004).

La complessa rappresentazione di genere emersa dall'analisi del *corpus* TG ha tuttavia fatto emergere anche un segnale di cambiamento, ovvero la rappresentazione delle donne come voce e soggetto del "dissenso delle donne", una classe lessicale che fa riferimento non solo alla ribellione pubblicamente manifestata dalle donne contro le molestie sessuali, ma anche contro altre forme di violenza e discriminazione, per esempio su base religiosa, politica, sociale. Benché si tratti di uno spazio tematico limitato e instabile nei tre periodi considerati, è indicativo dell'emersione di una questione di genere nell'informazione *mainstream*, qual è quella dei TG nazionali, storicamente poco attenta alle rivendicazioni delle donne, e in particolare poco attenta a una

¹⁵² Al 31 dicembre 2020 la percentuale maschile fra i carcerati, dunque persone per cui sono accertati reati punibili con il carcere, è pari al 96,4%. Fonte dati: mia elaborazione su dati ISTAT [online] https://dati.istat.it/Index.aspx?DataSetCode=DCCV_DETENUTI# (ultimo accesso 19 novembre 2021).

rappresentazione delle donne come cittadine attive nella sfera pubblica (Ross e Carter 2011). Infine non è irrilevante il fatto che questo campo semantico sia positivamente correlato con fonti giornalistiche femminili, perché è un'altra attestazione di come le giornaliste possano avere un ruolo attivo in una maggiore e migliore rappresentazione delle donne.

La seconda ricerca (Capitolo 6) si è focalizzata su una questione specifica che riguarda l'uso di sostantivi femminili per nominare donne in posizioni apicali, e verifica l'ipotesi che il linguaggio dell'informazione sia androcentrico per via anche di una resistenza a usare forme femminili per nominare donne che ricoprono cariche apicali, come quella di ministra, attraverso uno studio sul lemma *ministra/o*. Ho verificato questa ipotesi con un'analisi *corpus-based* sul *corpus* CRS e sul *corpus* TG, integrale e ridotto al 2018-2019 per una comparazione puntuale con il *corpus* CRS, che include solo articoli pubblicati nel gennaio 2018 e 2019, per ragioni spiegate nel Capitolo 4 (§ 4.3).

Le analisi che ho condotto per verificare questa ipotesi dimostrano, coerentemente con quanto emerso nella prima ricerca, che l'androcentrismo dell'informazione politica è ascrivibile in parte a una maggiore citazione dei politici, nella fattispecie dei ministri, evidentemente per via della loro posizione di maggioranza nelle diverse compagini governative italiane che costituiscono il principale, ancorché non esclusivo, contesto di riferimento nei periodi considerati (Governo Gentiloni, Governo Conte I e Conte II), in parte all'uso di forme maschili, specialmente nel riferimento al singolare, per nominare le ministre.

La comparazione per anno (limitatamente al biennio 2018-2019) e per *corpus* ha dimostrato una correlazione significativa fra il (gennaio) 2018 e un minor androcentrismo linguistico, che nel caso dell'informazione a stampa è contestuale con una minore sotto-rappresentazione delle ministre (rispetto ai ministri), mentre nel caso dell'informazione televisiva è contestuale con una minore tendenza a nominarle al maschile. In entrambi i casi, il minore androcentrismo non sembra esser riconducibile alla rappresentazione di una diversa realtà politica, dato che il

numero delle ministre nel Governo Gentiloni (gennaio 2018) e nel Governo Conte I (gennaio 2019) è lo stesso (N=5), bensì a un diverso contesto politico-istituzionale in termini di attenzione alla parità di genere. Nel gennaio 2018 era in vigore la 17esima legislatura caratterizzata da un impegno della presidente della Camera Laura Boldrini a favore di un uso della lingua italiana paritario e inclusivo, rispetto al genere. Le elezioni politiche del 4 marzo 2018 hanno avviato una legislatura che è rimasta sprovvista di figure istituzionali impegnate nella causa del linguaggio *gender-fair* e questo può aver inciso nel senso di un disimpegno del mondo dell'informazione. Questo disimpegno risulta particolarmente evidente nell'analisi dei dati pertinenti le due testate giornalistiche che fra il gennaio 2018 e il gennaio 2019 hanno cambiato direzione, il TG2 e *Il Resto del Carlino*. Il TG2 2018 attesta una significativa maggior frequenza di forme femminili nel 2018 rispetto al 2019, a fronte di un minor numero di riferimenti alle ministre. *Il Resto del Carlino* attesta un minor androcentrismo linguistico nel gennaio 2018, rispetto al gennaio 2019, contestuale sia a una maggior numero di riferimenti alle ministre (con forme maschili e femminili) sia a una maggiore frequenza di forme femminili. Questo risultato è particolarmente interessante perché sembrerebbe dimostrare, coerentemente con quanto emerso anche dalla prima ricerca (Capitolo 5), che una diversa linea editoriale può fare la differenza, ma per quanto riguarda l'uso dei femminili, che è un tema ancora divisivo nel nostro paese, probabilmente non basta l'«impegno convinto dei responsabili delle testate giornalistiche», per riprendere le parole di Sergio Lepri (10987, 22), bensì anche l'«impegno consapevole degli organismi politici e di governo».

Quanto l'impegno di Laura Boldrini, e poche altre politiche che, specialmente nel corso della 17esima legislatura, hanno lavorato per un linguaggio *gender-fair* (per esempio Valeria Fedeli che ha promosso le linee guida del MIUR, 2018), abbia contribuito a erodere una resistenza culturale all'uso dei femminili non è infatti trascurabile, perché i risultati basati sul *corpus* CRS, costruito per un confronto diretto con il *corpus* di Formato (2016) e una verifica dello stato di avanzamento dell'uso dei femminili su un lungo periodo, dimostra che qualche

progresso è stato fatto: le percentuali attestate da Formato (2016) oscillano infatti fra l'8,14% nel 2012, il 10,82% nel 2013 e il 10,00% nel 2014, valori inferiori rispetto al 43,43%¹⁵³ attestato nella mia ricerca per il biennio 2018-2019.

Nella terza ricerca (Capitolo 7) ho studiato le associazioni semantiche di genere latenti negli *embeddings* di una selezione scelta di nomi di professione e carica politica estratti dal *corpus* ACR, verificando l'ipotesi che i WEs attestino la persistenza di stereotipi di genere pertinenti la sfera professionale femminile (e maschile) nel linguaggio della stampa quotidiana. L'analisi delle associazioni di genere (femminile e maschile) nel *corpus* ha evidenziato che i nomi di professioni o cariche politiche codificano informazioni sul genere. Queste informazioni riflettono in parte una conoscenza della realtà, ovvero la diversa ripartizione di donne e uomini nelle posizioni professionali o politiche corrispondenti ai nomi selezionati; in parte riflettono però anche stereotipi di genere, ovvero associazioni di genere pregiudizievoli e sedimentate nella lingua nel corso del tempo. Di questo "sedimento" sembrano essere testimoni in particolare due nomi: *ingegnera/e* e *ostetrica/o*, il primo fortemente associato al genere maschile e il secondo al genere femminile. In entrambi i casi, così come in tutti i casi dei nomi analizzati e con genere mobile, tuttavia le forme femminili (e quelle maschili) mitigano le associazioni di genere stereotipate, dimostrando come la struttura *gender-marked* della grammatica italiana possa contribuire a una riduzione dei *bias* di genere, in linea con quanto attestato anche da diversi esperimenti di neuro e psicolinguistica che dimostrano, da un lato, che i nomi con riferimento umano attivano stereotipi di genere e, dall'altro, che, nelle lingue dal genere marcato, incluso l'italiano, l'uso delle forme femminili possono inibire l'attivazione degli stereotipi di genere, e, sia nei bambini che negli adolescenti, mitigare i pregiudizi di genere sulle aspettative di carriera e sulle competenze professionali.

¹⁵³ Il dato riportato è quello relativo soltanto alla forma singolare, poiché Formato analizza solo il singolare.

8.2 Limiti della tesi e sviluppi possibili

Il limite principale di questa tesi è rappresentato dalla dimensione e dalla copertura, in termini di periodo di tempo, dei *corpora* di analisi. La disponibilità dell'Osservatorio di Pavia di fornire gratuitamente solo i testi trascritti dei notiziari di un mese all'anno, a partire dal gennaio 2018, e di mettere a disposizione i file video relativi, per poter rivedere i testi, correggerne eventuali errori e marcarli, inserendo nel *corpus* TG alcune variabili categoriali utili a un'analisi linguistica integrata con un'analisi mediale, ha orientato la scelta di circoscrivere la costruzione anche degli altri due *corpora* al solo mese di gennaio (2018, 2019, 2020). Devo tuttavia riconoscere che, anche a fronte di una più ampia disponibilità dell'istituto di ricerca pavese, non credo sarei riuscita a costruire un *corpus* TG più grande, poiché l'operazione richiede un impiego di risorse molto oneroso, sia in termini di tempo che di attenzione. L'investimento in termini di tempo si è peraltro dimostrato particolarmente gravoso anche per la costruzione dei *corpora* CRS e ACR. Sebbene non sia stato necessario annotarli manualmente, l'operazione di *download* dei singoli articoli ha richiesto diversi giorni di lavoro, mentre più agili sono state l'aggregazione dei singoli file, eseguita automaticamente con una procedura informatica, e la successiva eliminazione dei metadati, eseguita tramite un algoritmo di Python. Questo limite non ha ridotto il valore euristico dei risultati delle analisi sui *corpora* TG e CRS, le cui dimensioni sono risultate adeguate a tutte le analisi statistiche, anche comparative, effettuate; ha invece precluso la possibilità di ricavare WEs anche dai *corpora* TG e CRS, perché troppo piccoli, e di svolgere analisi comparative per anno e testata giornalistica sul *corpus* ACR. Per questo tipo di analisi sarebbe stato infatti necessario disaggregare lo stesso in *sub-corpora* (per anno e testata giornalistica), che sarebbero però risultati di dimensioni troppo ridotte per un efficace funzionamento degli algoritmi disponibili per estrarre WEs.

Valutando in modo obiettivo i fattori che hanno contribuito a determinare questo limite, credo che uno sviluppo futuro in termini di costruzione di *corpora* possa ragionevolmente prefigurarsi non come una raccolta di testi informativi pertinenti altri mesi degli stessi anni, ma eventualmente come una raccolta di

notizie diffuse, magari sempre nel mese di gennaio, nel corso dell'anno 2021 e successivi. Questo consentirebbe di continuare tutte e tre le ricerche in una prospettiva longitudinale e, al contempo, di disporre di un *corpus* TG abbastanza grande da poter essere analizzato tramite approccio WEs, che è efficace soltanto per *corpora* di dimensione grande o medio-grande. Quanto al *corpus* CRS, per integrarlo con testi di articoli pubblicati dopo il 2019, si potrebbe provare a contattare direttamente il quotidiano *La Stampa* e verificare la disponibilità dell'editore a fornire gli articoli in un formato facilmente trattabile con un approccio linguistico computazionale.

Un secondo limite di questa tesi, sempre correlato al campione di analisi, riguarda la circoscrizione a un campione esemplificativo del linguaggio dell'informazione quotidiana della televisione e della stampa. Questa scelta, oltre a essere orientata da una preferenza personale, è stata dettata da due ragioni "pratiche". La prima ragione riguarda la relativamente facile reperibilità dei testi per la costruzione dei *corpora*, o meglio la difficoltà di reperire testi esemplificativi del linguaggio di altri tipi di media, come per esempio la radio, o i quotidiani online. Tuttavia, se un'estensione nella direzione di costruire un campione rappresentativo anche del linguaggio radiofonico può richiedere un programma di ricerca per un (forse troppo) lungo periodo, l'estensione di analisi su *corpora* esemplificativi dell'informazione online può essere programmata entro tempi di ricerca più brevi, perché non richiede la trascrizione di testi audio.

Un terzo limite di questa tesi è quello di non essere riuscita a prendere in considerazione direttamente tutti i diversi fattori che interagiscono, o potrebbero interagire, con le rappresentazioni di genere nel linguaggio dell'informazione. Avendo individuato, nel corso della rassegna bibliografica su lingua e genere e media e genere, diversi fattori intervenienti, mi sono limitata a considerare i principali, basandomi su dati *open source* che ho trovato on line, per esempio i dati sul tasso di occupazione femminile nelle professioni relative ai nomi analizzati nella terza ricerca.

A conclusione della tesi rimane in me la curiosità di indagare direttamente, attraverso un adeguato approccio metodologico, quale sia la metacompetenza linguistica dei giornalisti e delle giornaliste italiane e quale sia la metacompetenza mediale delle stesse. E quale sia, in entrambi i casi, la consapevolezza di quanto la lingua italiana e l'informazione, che ne è veicolo, possano contribuire a una rappresentazione di genere più paritaria e inclusiva, a supporto dell'avanzamento della parità di genere nei contenuti dei media. Ipotesizzo infatti che il livello di queste due metacompetenze sia correlato a rappresentazioni di genere generalmente più equilibrate e meno stereotipate. Un'intervista a un campione rappresentativo delle diverse professioni giornalistiche esercitate nei diversi tipi di organi informativi italiani potrebbe rappresentare, in tal senso, una proficua futura linea di ricerca.

Bibliografia

- Accademia della Crusca. 2013. *Femminicidio: i perché di una parola*. Firenze: Accademia della Crusca. [online] <https://accademiadellacrusca.it/it/consulenza/femminicidio-i-perché-di-una-parola/803>.
- Adamo, Sergia, Zanfabro Giulia e Elisabetta Tigani Sava, a cura di. 2019. *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere*. Trieste: EUT.
- Adamo, Sergia. 2019. "Non esiste solo il maschile. Alcune riflessioni su teorie e pratiche per un uso non discriminatorio del linguaggio da un punto di vista di genere". In *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere*, a cura di Sergia Adamo, Giulia Zanfabro e Elisabetta Tigani Sava. 9-16. Trieste: EUT.
- Aebischer, Verena. 1985. *Les femmes et le langage. Représentations sociales d'une différence*. Parigi: Presses Universitaires de France. Trad. it. 1988, *Il linguaggio delle donne. Rappresentazioni sociali di una differenza*. Roma: Armando Editore.
- AGCOM (Autorità per le Garanzie nelle Comunicazioni). 2020. *Relazione annuale 2020 sull'attività svolta e sui programmi di lavoro*. [online] https://www.agcom.it/documentazione/documento?p_p_auth=fLw7zRht&p_p_id=101_INSTANCE_FnOw5IVOIXoE&p_p_lifecycle=0&p_p_col_id=column-1&p_p_col_count=1&_101_INSTANCE_FnOw5IVOIXoE_struts_action=%2Fasset_publisher%2Fview_content&_101_INSTANCE_FnOw5IVOIXoE_asse.
- Agenzia delle entrate. 2020. *Linee guida per un linguaggio rispettoso delle differenze di genere*. [online] <https://www.agenziaentrate.gov.it/portale/documents/20143/1742359/Li>

[nee guida linguaggio genere 2020.pdf/0327598d-9607-4929-ceae-a3760b081ab4](https://www.researchgate.net/publication/327598496/figure/fig/1/figure-pdf/327598496-9607-4929-ceae-a3760b081ab4).

al-Rfou, Rami, Perozzi, Bryan e Steven Skiena. 2013. "Polyglot: distributed word representations for multilingual NLP". In *Proceedings of the 17th conference on Computational Natural Language Learning (CoNLL 2013)*. Sofia, Bulgaria. 183-192.

Albæk, Erik. 2011. "The interaction between experts and journalists in news journalism". *Journalism* 12 (3): 335-348.

Ang, Ien. 1985. *Watching Dallas. Soap Opera and the Melodramatic Imagination*. Londra: Methuen.

Attili, Grazia e Laura Benigni. 1977. "Retorica naturale e linguaggio femminile". In *Psicologia e retorica*, a cura di Giuseppe Mosconi e Valentina D'Urso. 85-91. Bologna: Il Mulino.

Attili, Grazia e Laura Benigni. 1979. "Interazione sociale, ruolo sessuale e comportamento verbale: lo stile retorico naturale del linguaggio femminile nell'interazione faccia a faccia". In *Retorica e scienze del linguaggio. Atti del X convegno internazionale di studi*, a cura di Federico Albano Leoni e Maria Rosaria Pigliasco. 261-280. Roma: Bulzoni.

Austin, Jane. 1962. *How to do things with words*. Oxford: Oxford University Press.

Azzalini, Monia e Claudia Padovani. 2010. *Who makes the news? Global media monitoring project 2015. National report. Italy* [online] https://whomakesthenews.org/wp-content/uploads/who-makes-the-news/imported/reports_2010/national/Italy.pdf.

Azzalini, Monia e Claudia Padovani. 2015. *Who makes the news? Global media monitoring project 2015. National report. Italy* [online] https://whomakesthenews.org/wp-content/uploads/who-makes-the-news/imported/reports_2015/national/Italy.pdf.

- Azzalini, Monia e Claudia Padovani. 2016. "Informazione ed uguaglianza di genere". *Il Mulino* 2: 276-284.
- Azzalini, Monia e Claudia Padovani. 2021. *Who makes the news? Global media monitoring project 2020. National report. Italy* [online] <https://whomakesthenews.org/wp-content/uploads/2021/07/Italy-Report-GMMPrev2.pdf>.
- Azzalini, Monia e Giuliana Giusti. 2019. "Lingua e genere fra grammatica e cultura". *Economia della cultura* 34 (4): 537-546.
- Azzalini, Monia e Manuela Malchiodi. 2013. *Femmes et presse écrite au Maghreb. Amélioration de la représentation des femmes dans les médias au Maghreb*. Rabat, Marocco: UNESCO.
- Azzalini, Monia. 2006. "Le quote della TV nella campagna elettorale 2006". *Inchiesta* 153: 24-25.
- Azzalini, Monia. 2009. "Europee 2009: la visibilità femminile nei programmi Rai durante la campagna elettorale". *Comunicazione politica* 3: 483-487.
- Azzalini, Monia. 2015a. "Donne nel giornalismo italiano". *Problemi dell'informazione* 3: 465-482.
- Azzalini, Monia. 2015b. "New media: quale rappresentazione femminile? I risultati nazionali del Global Media Monitoring Project 2015". *Problemi dell'informazione* 3: 669-674.
- Azzalini, Monia. 2016a. "Discriminazioni di genere nell'informazione. Una sfida ancora aperta". *Aggiornamenti sociali* 8/9: 580-590.
- Azzalini, Monia. 2016b. "Media e genere: l'immaginario svelato dai numeri". In *Siamo le parole che usiamo*, a cura di Veronica Bacci Bonivento, Nadia Cairo, Julia Di Candia, Alisa del Re, Bruna Mura e Lorenza Perini. 56-58. Padova: Padova University Press.

- Azzalini, Monia. 2021a. "Lingua e genere nell'informazione televisiva. Un caso di studio su *ministra e ministro*". *Problemi dell'informazione* 2: 213-235.
- Azzalini, Monia. 2021b. "Parole inclusive e paritarie per raccontare le donne nel mondo dello sport". *Uispres. Numero speciale Donne, diritti e parità di genere* 29. [online] http://www.uisp.it/nazionale/files/principale/2021/Monia%20Azzalini_Parole%20inclusive_2021.pdf.
- Bailey, April H. , LaFrance, Marianne e John F. Dovidio. 2019. "Is Man the Measure of All Things? A Social Cognitive Account of Androcentrism". *Personality and Social Psychology Review* 23(4) :307-321.
- Baker, Paul. 2006. *Using corpora in discourse analysis*. Londra: Routledge.
- Baker, Paul. 2014. *Using corpora to analyse gender*. Londra: Bloomsbury.
- Bambini, Valentina. 2017. *Il cervello pragmatico*. Roma: Carocci.
- Banaji, Mahazarin R. e Curtis D. Hardin. 1996. "Automatic stereotyping". *Psychological science* 7 (3): 136-141.
- Bandelli, Daniela e Giorgio Porcelli. 2016. "'Femminicidio' in Italy: A critique of feminist gender discourse and constructivist reading of the human identity". *Current sociology* 64 (7): 1071-1089.
- Baril, Elodie e Bénédicte Garnier. 2015. *IRaMuteQ 0.7 alpha 2 Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*. [online] http://www.iramuteq.org/documentation/fichiers/Pas%20a%20Pas%20IRAMUTEQ_0.7alpha2.pdf.
- Barocas, Solon, Crawford, Kate, Shapiro, Aaron e Hanna Wallach. 2017. "The problem with bias: allocative versus representational harms in machine learning". In *9th Annual Conference of the Special Interest Group for Computing, Information and Society*.

- Barocas, Solon, Hardt, Moritza e Arvind Naryanan. 2019. *Fairness and machine learning. Limitations and opportunities*. [online] <https://fairmlbook.org/pdf/fairmlbook.pdf>.
- Baron, Deborah. 1986. *Language and gender*. New Haven: Yale University Press.
- Baroni, Marco e Adam Kilgarriff. 2006. "Large linguistically-processed web corpora for multiple languages". In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*. 87–90.
- Baroni, Marco, Bernardini, Silvia, Ferraresi, Adriano e Eros Zanchetta. 2006. "The WaCky wide web: a collection of very large linguistically processed web-crawled corpora". *Language resources and evaluation* 43 (3): 209-226.
- Baroni, Marco, Dinu, Georgiana e Germán Kruszewski. 2014. "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors". In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, USA*. 238-247.
- Bazzanella, Carla e Orsola Fornara. 1995. "Segnali discorsivi e linguaggio femminile: evidenze da un corpus". In *Donna e linguaggio. Atti del convegno internazionale di studi. Sappada/Plodn (Belluno)*, a cura di Gianna Marcato. 73-84. Padova: CLEUP.
- Bazzanella, Carla, Fornara Orsola e Manuela Manera. 2006. "Indicatori linguistici e stereotipi femminili". In *Linguaggio e genere*, a cura di Silvia Luraghi e Anna Olita. 155-169. Roma: Carocci.
- Bazzanella, Carla. 2009. "Stereotipi e categorizzazioni del femminile/maschile". In *Mi fai male... Atti del convegno. Venezia, Auditorium Santa Margherita 8-19-20 novembre 2008*, a cura di Giuliana Giusti e Susanna Regazzoni. 99-114. Venezia: Cafoscarina.

- Bazzanella, Carla. 2010. *Genere e lingua*. In Enciclopedia dell'italiano. Treccani. [online] [https://www.treccani.it/enciclopedia/genere-e-lingua_\(Enciclopedia-dell'Italiano\)/](https://www.treccani.it/enciclopedia/genere-e-lingua_(Enciclopedia-dell'Italiano)/).
- Bem, Sandra Lipsitz. 1993. *The lenses of gender. Transforming the debate on sexual inequality*. New Haven: Yale University Press.
- Benzécri, Jean-Paul. 1973. *L'analyse des données*. Paris-Montréal: Dunod.
- Benzécri, Jean-Paul. 1980. *Pratique de l'analyse des données. Analyse des correspondances & classification. Exposé élémentaire*. Parigi-Montréal: Dunod.
- Benzécri, Jean-Paul. 1992. *Correspondence analysis handbook*. New-York-Basel-Hong Kong: Marcel Dekker.
- Bequer, Anne-Marie, Cholewka, Nicole, Coutier, Martine e Marie-Josèphe Mathieu. 1999. *Femme, j'écris ton nom... Guide d'aide à la féminisation des nomes de métiers, titres, grades et fonctions*. Parigi: Centre national de la recherche scientifique Institut national de la langue française.
- Berardi, Giacomo, Esuli, Andrea e Diego Marchegiani. 2015. "Word embeddings go to Italy: a comparison of models and training datasets". In *Proceedings of the 6th Italian information retrieval workshop. Cagliari, Italia*, a cura di Paolo Boldi, Reffaele Perego e Fabrizio Sebastiani. 1404.
- Berretta, Monica. 1983. "Per una retorica popolare del linguaggio femminile, ovvero la lingua delle donne come costruzione sociale". In *Comunicare nella vita quotidiana*, a cura di Franca Orletti. 215-240. Bologna: Il Mulino.
- Berruto, Gaetano. 1977. "Dialettologia e sociolinguistica in Italia". In *Gli aspetti sociolinguistici dell'Italia contemporanea*, a cura di Raffaele Simone e Giulianella Ruggiero. 75-86. Roma: Bulzoni.
- Berruto, Gaetano. 1980. *La variabilità sociale della lingua*. Torino: Loescher.

- Bertrand, Marianne e Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination". *American economic review* 94 (4): 991-1013.
- Bettinelli, Ernesto. 1995. *Par condicio: regole, opinioni, fatti*. Torino: Einaudi.
- Bianchi, Claudia. 2006. "La parola". In *Donna m'apparve*, a cura di Nicla Vassallo. 83-102. Torino: Codice Edizioni.
- Biasion, Davide, Fabris, Alessandro, Silvello, Gianmaria e Gian Antonio Susto. 2020. "Gender bias in italian word embeddings". In *Proceedings of the seventh Italian Conference on Computational Linguistics. CLiC-it 2020. Bologna, Italia*, a cura di Johanna Monti, Felice Dell'Orletta e Fabio Tamburini. [online] <http://ceur-ws.org/Vol-2769/>.
- Biron, Monique. 1991. *Au féminin. Guide de féminisation des titres de fonction et des textes*. Québec: Les publications du Québec. [online] <https://numerique.banq.qc.ca/patrimoine/details/52327/46231>.
- Bloothoof, Gerrit. 2010. *Nederlandse Voornamenbank*. <https://www.meertens.knaw.nl/nvb/veelgestelde vragen>.
- Bodine, Ann. 1975. "Androcentrism in prescriptive grammar: singular 'they', sex-indefinite 'he', and 'he or she'". *Language in society* 4:129-146.
- Bojanowski, Piotr, Grave, Eduard, Joulin, Armand e Tomas Mikolov. 2017. "Enriching word vectors with subword information". *Transactions of the association for computational linguistics* 5: 135–146.
- Bolasco, Sergio. 2013. *L'analisi automatica dei testi. Fare ricerca con il text mining*. Roma: Carocci.
- Boldrini, Laura. 2015. #NonSiamoCosì. Twitter. <https://twitter.com/lauraboldrini/status/573436739773153280>. 5 marzo.

- Bolukbasi, Tolga, Chang, Kai-Wei, Zou, James, Saligrama, Venkatesh e Asam Kalai. 2016. "Quantifying and reducing stereotypes in word embeddings." In *Proceedings of the 2016 ICML workshop on #Data4Good: machine learning in social good applications*. New York, USA, a cura di Kush R. Varshney. 41-45.
- Boot, Peter, Zijlstra, Hanna e Rinie Geenen. 2017. "The Dutch translation of the Linguistic Inquiry and Word Count (LIWC) 2007 dictionary". *Dutch Journal of Applied Linguistics* 6 (1): 65–76.
- Bourdieu, Pierre. 1997. *Sulla televisione*. Milano: La Feltrinelli.
- Braidotti, Rosi. 1994. *Nomadic subjects. Embodiment and sexual difference in contemporary feminist theory*. New York: Columbia University Press. Trad. it. 1995. *Soggetto nomade. Femminismo e crisi della modernità*. Roma: Donzelli.
- Bucholtz, Mary. 2004. "Editor's Introduction". In Robin T. Lakoff e Mary Bucholtz, *Language and women's place. Text and commentaries. Revised and expanded edition*. 3-14. New York: Oxford University Press.
- Bundeskanzlei. 2009. *Geschlechtergerechte Sprache. Leitfaden zum geschlechtergerechten Formulieren im Deutschen*. Prima edizione. Berna.
- Bundeskanzlei. 2021. *Geschlechtergerechte Sprache. Leitfaden zum geschlechtergerechten Formulieren im Deutschen*. Seconda edizione. Berna.
- Bundesstelle für Büroorganisation und Bürotechnik bulletin. 2000. *Sprachliche Gleichbehandlung und Beispiel*. Koln.
- Buonanno, Milly e Franca Faccioli, a cura di. 2020. *Genere e media: non solo immagini. Soggetti, politiche, rappresentazioni*. Milano: Franco Angeli.
- Buonanno, Milly, a cura di. 2014a. *Il prisma dei generi. Immagini di donne in tv*. Milano: Franco Angeli.

- Buonanno, Milly, a cura di. 2017. *Television antiheroines. Women behaving badly in crime and prison drama*. Bristol: Intellect.
- Buonanno, Milly. 1967. "La demistificazione della femminilità". *Tempi moderni* 29: 186-188.
- Buonanno, Milly. 1968. "La stampa femminile". *Tempi moderni* 32: 151-155.
- Buonanno, Milly. 1975. *Naturale come sei. Indagine sulla stampa femminile in Italia*. Firenze: Guaraldi.
- Buonanno, Milly. 1978a. "L'informazione prodotta dalla donne". *Problemi dell'informazione* (2) 1: 465-472
- Buonanno, Milly. 1978b. *La donna nella stampa. Giornaliste, lettrici, modelli di femminilità*. Roma: Editori Riuniti.
- Buonanno, Milly. 1982. *L'immagine inattesa. La donna nei programmi televisivi tra realtà e immaginario*. Roma: Rai-Eri VQPT 34.
- Buonanno, Milly. 1983. *Cultura di massa e identità femminili. L'immagine della donna in televisione*. Torino: ERI.
- Buonanno, Milly. 1988. *L'élite senza sapere. Uomini e donne nel giornalismo italiano*. Napoli: Liguori.
- Buonanno, Milly. 1993. "Costruire il punto di viste delle donne sull'informazione". *Problemi dell'informazione*. 3: 271-281.
- Buonanno, Milly. 1994. *Narrami o diva. Studi sull'immaginario televisivo*. Napoli: Liguori.
- Buonanno, Milly. 1999. "Occhio di donna e visione etica del giornalismo". In Milly Buonanno, *Faction. Soggetti mobili e generi ibridi nel giornalismo italiano degli anni novanta*. 29-41. Napoli: Liguori.

- Buonanno, Milly. 2005. *Visibilità senza potere. Le sorti progressive ma non magnifiche delle giornaliste in Italia*. Napoli: Liguori.
- Buonanno, Milly. 2006. "Donne e informazione: accesso, potere e cambiamento". *Inchiesta* (36) 153: 57-62
- Buonanno, Milly. 2008. "Le ragazze con la pistola. La femminilizzazione del poliziesco televisivo italiano". In *Gender e media*, a cura di Annalisa Tota. 63-82. Roma: Meltemi.
- Buonanno, Milly. 2014b. "Donne al comando fra action e melodramma. Il caso di Squadra antimafia". In *Il prisma dei generi. Immagini di donne in tv*, a cura di Milly Buonanno. 49-77. Milano: Franco Angeli.
- Buonanno, Milly. 2014c. "Forward but not too fast. Donne e informazione: a che punto siamo?". *Problemi dell'informazione* 39 (1): 75-84.
- Buonanno, Milly. 2015a. "Introduzione". *Problemi dell'informazione* 15 (3): 433-440.
- Buonanno, Milly. 2015b. "Al fronte, ma non sulla *front page*. Giornaliste in prima pagina". *Problemi dell'informazione* 15 (3): 483-502.
- Burnett, Heather e Olivier Bonami. 2018. "Linguistic prescription, ideological structure, and the actualization of linguistic changes: grammatical gender in French parliamentary debates". *Language in society* 49(1): 65-93.
- Burr, Elisabeth. 1995. "Agentivi e sessi in un corpus di giornali italiani". In *Donna e linguaggio. Atti del convegno internazionale di studi. Sappada/Plodn (Belluno)*, a cura di Gianna Marcato. 349-365. Padova: CLEUP.
- Burr, Elisabeth. 2003. "Gender and language politics in France". In *Gender across languages. The linguistic representation of women and men*, a cura di Marlis Hellinger e Hadumod Bußmann. Vol. 3. 119–39. Amsterdam-Philadelphia: John Benjamins.

- Busso, Lucia e Gianmarco Vignozzi. 2017. "A corpus-assisted analysis". In *Proceedings of the fourth Italian Conference on Computational Linguistics CLiC-it 2017. Roma, Italia*, a cura di Roberto Basili, Malvina Nissim e Giorgio Satta. 71-76.
- Butler, Judith. 1990. *Gender trouble. Feminism and the subversion of identity*. New York: Routledge.
- Butler, Judith. 1993. *Bodies that matter. On the discursive limits of sex*. New York-Londra: Routledge.
- Butler, Judith. 1999. *Gender trouble. Feminism and the subversion of identity*. Seconda edizione. New York: Routledge. Trad. it. 2017. *Questioni di genere. Identità, sesso e desiderio*. Bari-Roma: Laterza.
- Butler, Matilda e William J. Paisley. 1980. *Women and the mass-media*. New York: Human Science Press.
- Byerly, Carolyn M. 2011. *Global report on the status of women in the news media*. Washington DC: International Women's media foundation IWVF. [online] <https://www.iwmf.org/wp-content/uploads/2018/06/IWMF-Global-Report.pdf>.
- Cacciari, Cristina e Roberto Padovani 2007. "Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns". *Applied psycholinguistics* 28 (2): 277-293.
- Cacciari, Cristina, Corradini, Paola, Padovani, Roberto e Manuel Carreiras. 2011. "Pronoun resolution in Italian: the role of grammatical gender and context". *Journal of cognitive psychology* 23 (4): 416-434.
- Caffarra, Sendy, Siyanova-Chanturia, Anna, Pesciarelli, Francesca, Vespignani, Francesco e Cristina Cacciari. 2015. "Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian". *Psychophysiology* 52: 1010-1030.

- Caliskan, Aylin, Brysono, Joanna J. e Arvind Narayanan. 2017. "Semantics derived automatically from language *corpora* contain human-like biases". *Science* 356: 183-186.
- Camargo, Brigido V. e Ana M. Justo. 2013. "IRAMUTEQ: um software gratuito par análise de dados textuais". *Temas em psicologia* 21 (2): 513-518.
- Camargo, Brigido V. e Ana M. Justo. 2018. *Tutorial para uso di software IRaMuTeQ*, . 2018. [online] <http://www.iramuteq.org/documentation/fichiers/tutoriel-portugais-22-11-2018>.
- Camera dei Deputati. 2017. *Proposta di legge 4335. Disposizioni in materia di concordanza dei titoli funzionali in base al sesso della persona cui sono attribuiti negli atti delle pubbliche amministrazioni*. [online] <http://documenti.camera.it/dati/leg17/lavori/stampati/pdf/17PDL0050530.pdf>.
- Camera dei deputati. Servizio Studi (2021). *La partecipazione delle donne alla vita politica e istituzionale*. [online] <https://documenti.camera.it/Leg18/Dossier/Pdf/AC0340.Pdf>
- Cameron, Deborah e Sylvia Shaw. 2016. *Gender power and political speech. Women and language in the 2015 UK general elections*. Londra: Macmillan.
- Cameron, Deborah. 1995. "The new Pygmalion: verbal hygiene for women". In *Verbal Hygiene*. 167-211. Londra e New York: Routledge.
- Cameron, Deborah. 1998. *The feminist critique of language*. New York e Londra: Routledge.
- Cameron, Deborah. 2003. "Gender and Language Ideologies". In *The handbook of language and gender*, a cura di Janet Holmes e Miriam Meyerhoff. 447-467. Oxford: Blackwell.

- Cameron, Deborah. 2009. "Sex/gender, language and the new biologism". *Applied linguistics* 31 (2): 173-192.
- Campus, Donatella. 2010. "Gli stereotipi di genere e le donne in politica: un connubio apparentemente inestricabile". In *L'immagine della donna leader*, a cura di Donatella Campus. 29-56. Bologna: Bonomia University Press.
- Cancelleria federale. 1991. *Parità tra donna e uomo nel linguaggio normativo e amministrativo. Rapporto di un gruppo interdipartimentale della Confederazione*. Berna.
- Cancelleria federale. 2003. *Istruzioni per la redazione dei testi ufficiali in italiano*. Berna.
- Cancelleria federale. 2012. *Pari trattamento linguistico. Guida al pari trattamento linguistico di donna e uomo nei testi ufficiali della Confederazione*. Berna.
- Cantone Ticino. 1995. *Tecniche per la redazione di atti ufficiali*. Prima edizione. Bellinzona.
- Cantone Ticino. 2004. *Tecniche per la redazione di atti ufficiali*. Seconda edizione. Bellinzona.
- Cantor, Muriel S. 1978. "Where are the women in public broadcasting?". In *Heart and home. Images of women in the mass media*, a cura di Gaye Tuchman, Arlene Kaplan e James Benét. 78-90. New York: Oxford University Press.
- Capecchi, Saveria e Clelia Pallotta. 2001. "La rappresentazione di genere nei programmi di approfondimento e attualità". In *Una, nessuna... a quando centomila? La rappresentazione della donna in televisione*, a cura di Loredana Cornero. 105-177. Roma: Rai-Eri VQPT 184.
- Capecchi, Saveria. 1995. "Immagini di uomini e donne nella stampa periodica". *Problemi dell'informazione* 1: 93-115.

- Capecchi, Saveria. 2000. *Ridendo e sognando (con le soap). Il pubblico di Un posto al sole e di Beautiful*. Roma: Rai-Eri VQPT 176.
- Capecchi, Saveria. 2002. "Media e immaginari femminili". In *Tra i generi. Rileggendo le differenze di genere di generazione di orientamento sessuale*, a cura di Carmen Leccardi. 111-132. Milano: Guerini e Associati.
- Capecchi, Saveria. 2004. *L'audience "attiva". Effetti e usi sociali dei media*. Roma: Carocci.
- Capecchi, Saveria. 2006. *Identità di genere e media*. Roma: Carocci.
- Capecchi, Saveria. 2007. "Che donna sei? Modelli femminili proposti dai media (e dibattiti in corso)". In *Ofelia e Parsifal. Modelli e differenze di genere nel mondo dei media*, a cura di Giorgio Grossi e Elisabetta Ruspini. 1-34. Milano: edizioni libreria Cortina.
- Capecchi, Saveria. 2011. "Il corpo erotizzato delle donne negli spot pubblicitari e nelle riviste di moda femminile". *Polis* 3: 393-417.
- Capecchi, Saveria. 2018. *La comunicazione di genere. Prospettive teoriche e buone pratiche*. Roma: Carocci.
- Cardinaletti, Anna e Giuliana Giusti. 1991. "Il sessismo nella lingua italiana. Riflessione sui lavori di Alma Sabatini". *Rassegna italiana di linguistica applicata* 91 (2): 169-189.
- Carreiras, Manuel, Garnham, Alan, Oakill, Jane e Kate Cain. 1996. "The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish". *Quarterly journal of experimental psychology* 49: 639-663.
- Castenetto, Giorgia e Stefano Ondelli. 2020. "The acceptability of feminine job titles in Italian newspaper articles". In *Language, gender and hate speech*. A

- multidisciplinary approach*, a cura di Giuliana Giusti e Gabriele Iannacaro. 75-89. Venezia: Edizioni Ca' Foscari.
- Cavagnoli, Stefania e Francesca Dragotto. 2021. *Sessismo*. Milano: Mondadori.
- Cavagnoli, Stefania. 2013. *Linguaggio giuridico e lingua di genere: una simbiosi possibile*. Alessandria: Edizioni dell'Orso.
- Cavagnoli, Stefania. 2015. "Riflessioni sulla lingua". In CNOG. *Tutt'altro genere d'informazione. Manuale per una corretta rappresentazione delle donne nell'informazione*. 119-136. [online] <http://old.odg.it/content/tuttaltro-genere-dinformazione>.
- Cavagnoli, Stefania. 2019. "La grammatica della lingua di genere: dissimmetrie semantiche e stereotipi". In *Stereotipi. Donne nei media*, a cura di GiULiA giornaliste. Milano: Ledizioni.
- CENSIS. 2006. *Women and media in Europe. Libro bianco*. Roma: Fondazione Adkins Chiti, Fondazione Censis e Fondazione Risorsa Donna. [Sintesi online]: <https://parita.regione.emilia-romagna.it/documentazione/documentazione-temi/documentazione-stereotipi-di-genere/donne-e-media-in-europa>.
- Chaher, Sandra, a cura di. 2014. *Public policies on communication and gender in Latin America: the path ahead of us*. Buenos Aires: Asociación Civil Comunicación para la Igualdad, Defensoría del Público.
- Chancellerie federale. 2000. *Guide de formulation non sexiste des textes administratifs et législatifs de la Confédération*. [online] <https://www.bk.admin.ch/bk/fr/home/documentation/langues/aides-redaction-et-traduction/guide-de-formulation-non-sexiste.html>
- Chartier, Jean-François e Jean G. Meunier. 2011. "Text mining methods for social representation analysis". *Papers on social representations* 20: 37.1-37.47.

- CNEL (Consiglio Nazionale dell'Economia e del Lavoro). 2002. *Donne, lavoro e TV. La rappresentazione femminile nei programmi televisivi*. Roma: CNEL.
- CNEL (Consiglio Nazionale dell'Economia e del Lavoro). 2004. *Donne, lavoro e TV. L'immagine della donna nei programmi d'informazione*. Roma: CNEL. [Sintesi online]: <http://www.tempiespazi.it/segnala/?action=item&id=20050509092451530&template=tempisegbio&folderid=139>.
- CNOG (Consiglio Nazionale dell'Ordine dei Giornalisti). 2015. *Tutt'altro genere d'informazione. Manuale per una corretta rappresentazione delle donne nell'informazione*. [online] <http://old.odg.it/content/tuttaltro-genere-dinformazione>.
- Coates, Jennifer e Deborah Cameron, a cura di. 1989. *Women in their speech communities*. Londra: Longman.
- COE (Consiglio d'Europa). 1990. *Recommendation No. R (90) 4 of the Committee of Ministers to member states on the elimination of sexism from language*. Adopted by the Committee of Ministers on 21 February 1990 at the 434th meeting of the Ministers' Deputies. [online] <https://rm.coe.int/1680505480>.
- COE (Consiglio d'Europa). 1994. *Instruction No. 33 of 1 June 1994 concerning the use of non sexist language at the Council of Europe*. [online] <https://rm.coe.int/1680781cf4>.
- COE (Consiglio d'Europa). 2007. *Recommendation CM/Rec(2007)17 of the Committee of Ministers to member states on gender equality standards and mechanisms*. Adopted by the Committee of Ministers on 21 November 2007 at the 1011th meeting of the Ministers' Deputies. [online] <https://www.coe.int/en/web/genderequality/standards-and-mechanisms>.

- COE (Consiglio d'Europa). 2011. *Convenzione sulla prevenzione e la lotta contro violenza nei confronti delle donne e la violenza domestica*. [online] <https://www.coe.int/it/web/conventions/full-list?module=treaty-detail&treatynum=210>.
- COE (Consiglio d'Europa). 2016a. *Encouraging the participation of the private sector and the media in the prevention of violence against women and domestic violence: article 17 of the Istanbul Convention* [online] <https://edoc.coe.int/en/violence-against-women/6804-encouraging-the-participation-of-the-private-sector-and-the-media-in-the-prevention-of-violence-against-women-and-domestic-violence-article-17-of-the-istanbul-convention.html>.
- COE (Consiglio d'Europa). 2016b. *Combating sexist hate speech*. [online] <https://edoc.coe.int/en/gender-equality/6995-combating-sexist-hate-speech.html>.
- Comune di Siena. 2016. *Guida operativa per l'uso di un linguaggio amministrativo non sessista negli atti e nei documenti del Comune di Siena*. [online] <https://www.comune.siena.it/Il-Comune/Servizi/Pari-Opportunita-e-Politiche-di-Genere/Guida-operativa-per-l-uso-non-sessista-del-linguaggio-amministrativo>.
- Comune di Torino. 2017. *Linee guida per un utilizzo non discriminatorio del linguaggio in base al genere nell'attività amministrativa*. [online] <http://www.comune.torino.it/ediliziaprivata/normativa/pdf/all1-DelGC201703517.pdf>.
- Corbett, Greville G. 1991. *Gender*. Cambridge-New York-Port Chester-Melbourne-Sindney: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press
- Corbett, Greville G. e Sebastian Fedden. 2016. "Canonical gender". *Journal of Linguistics* 52: 495-531.

- Corbett, Greville G., a cura di. 2013. *The expression of gender*. De Grutier.
- Cornero, Loredana, a cura di. 2001. *Una, nessuna... a quando centomila? La rappresentazione della donna in televisione*. Roma: Rai-Eri VQPT 184.
- Cortellazzo, Manlio. 1995. "Perché non si vuole la presidentessa?" In *Donna e Linguaggio. Atti del Convegno internazionale di studi. Sappada/Plodn (Belluno)*, a cura di Gianna Marcato. 49-52. Padova: CLEUP.
- Cosenza, Giovanna. 2012. *SpotPolitik. Perché la «casta» non sa comunicare*. Milano: Laterza.
- Costa, Giacomo. 2019. "Gilet gialli: in ascolto di un conflitto sociale inedito". *Aggiornamenti sociali* 2 (70): 93-100.
- Crawford, Kate. 2017. *The trouble with bias*. NIPS 2017 Keynote. Youtube. 10 dicembre. https://www.youtube.com/watch?v=fMym_BKWQzk. Ultimo accesso 10 giugno 2021.
- Creedon, Pamela J. 1989. *Women in mass communication. Challenging gender values*. Newbury Park: Sage.
- Criado Perez, Caroline. 2020. *Invisibili. Come il nostro mondo ignora le donne in ogni campo. Dati alla mano*. Torino: Einaudi.
- D'Acci, Julie. 1994. *Defining women: television and the case of Cagney and Lacey*. Chapel Hill-Londra: The University of North Caroline Press.
- Davies, Mark. 2015. "Corpora: an introduction". In *The Cambridge handbook of English corpus linguistics*, a cura di Douglas Biber e Randi Reppen. 11-31. Cambridge: Cambridge University Press.
- De Beauvoir, Simone. 1949. *Le deuxième sexe*. Parigi: Gallimard.
- De Laurentiis, Teresa. 1987. *Technologies of gender: essays on theory, film and fiction*. Bloomington: Indiana University Press.

- De Marco Anna. 1995. "L'influenza del sesso nell'uso dei diminutivi in italiano". In *Donna e linguaggio. Atti del convegno internazionale di studi. Sappada/Plodn (Belluno)*, a cura di Gianna Marcato. 87-98. Padova: CLEUP.
- De Maria, Cristina. 2014. "Abiti di genere. Other world e inner world tra azione, immaginazione e Phantasie". *Rivista italiana di filosofia del linguaggio*. Numero speciale SFL: 258-269.
- De Mauro, Tullio, Mancini, Federico, Vedovelli Massimo e Miriam Voghera. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etas libri.
- De-Arteaga, Maria, Romanov, Alexey, Wallach, Hanna, Chayes, Jennifer, Borgs, Christian Alexandra, Sahin Geyik, Chouldechova, Kenthapadi, Krishnaram e Adam Tauman Kalai. 2019. "Bias in bios: a case study of semantic representation bias in a high-stakes setting". *Proceedings of the conference on Fairness, Accountability, and Transparency. FAT*19. Atlanta GA, USA*. 120-128.
- Decreto del Presidente della Repubblica 30 novembre 2012, n. 251. Regolamento concernente la parità di accesso agli organi di amministrazione e di controllo nelle società, costituite in Italia, controllate da pubbliche amministrazioni, ai sensi dell'articolo 2359, commi primo e secondo, del codice civile, non quotate in mercati regolamentati, in attuazione dell'articolo 3, comma 2, della legge 12 luglio 2011, n. 120. *Gazzetta ufficiale della Repubblica Italiana*. 154 (23): 25-26. Roma, 28 gennaio 2013. [online] <https://www.gazzettaufficiale.it/eli/gu/2013/01/28/23/sg/pdf>.
- Decreto Legislativo 31 luglio 2005, n. 177. Testo unico della radiotelevisione. *Gazzetta ufficiale della Repubblica italiana*. Supplemento ordinario al Numero 208 del 7 settembre 2005. [online] <https://www.gazzettaufficiale.it/eli/gu/2005/09/07/208/so/150/sg/pdf>.

- Di Cristofaro Longo, Gioia. 1986. *Immagine donna. Modelli di donna emergenti nei mezzi di comunicazione di massa*. Roma: Istituto Poligrafico e Zecca dello Stato.
- Di Cristofaro Longo, Gioia. 1992. *La donna nei media. Denunce, analisi, ricerche: modelli culturali emergenti*. Roma: Istituto Poligrafico e Zecca dello Stato.
- Di Cristofaro Longo, Gioia. 1995. *La disparità virtuale. Donne e mass media*. Roma: Armando editore.
- Dinelli, Serena. 2011. "Donne, media, comunicazione trent'anni dopo: presenze e linee di lavoro nel movimento delle donne in Italia". *La camera blu. Rivista di studi di genere*. 7: 59-84 [online] <http://www.serena.unina.it/index.php/camerablu/article/view/1366>.
- Djerf-Pierre, Monika e Maria Esdröm, a cura di. 2020a. *Comparing gender and media equality across the globe*. Gothenburg: Nordicom.
- Djerf-Pierre, Monika e Maria Esdröm. 2020b. "Introduction". In *Comparing gender and media equality across the globe*, a cura di Djerf-Pierre, Monika e Maria Esdröm. 11-56. Gothenburg: Nordicom.
- Djerf-Pierre, Monika e Maria Esdröm. 2020c. "The GEM Index: constructing a unitary measure of gender equality in the news". In *Comparing gender and media equality across the globe*, a cura di Djerf-Pierre, Monika e Maria Esdröm. 59-98. Gothenburg: Nordicom.
- Djerf-Pierre, Monika. 2020. "Explaining gender equality in news content". In *Comparing gender and media equality across the globe*, a cura di Djerf-Pierre, Monika e Maria Esdröm. 147-189. Gothenburg: Nordicom.
- Doleschal, Ursula .2009. "Linee guida e uguaglianza linguistica". In *Mi fai male... Atti del convegno. Venezia, Auditorium Santa Margherita 8-19-20 novembre 2008*, a cura di Giuliana Giusti e Susanna Regazzoni. 135-147. Venezia: Cafoscarina.

- Douglas, Biere e Randi Reppen, a cura di. 2015. *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press.
- Douglas, Susan J. 1995. *Where the girls are: growing un female with the mass media*. Londra: Penguin.
- Dragotto, Francesca, a cura di. 2012. *Grammatica e sessismo*. Vol. 1. Roma: Universitalia.
- Dragotto, Francesca, a cura di. 2016. *Grammatica e sessismo*. Vol. 2 Roma: Universitalia.
- Duffy, Susan A. e Jessica A. Keir. 2004. "Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge". *Memory & cognition* 32 (4): 551-559.
- EBU (European Broadcasting Union). 2019. *All things being equal: gender equality guidelines from public service media*. [online] https://www.ebu.ch/publications/strategic/login_only/all-things-being-equal-gender-equality-guidelines-from-public-service-media.
- Ehrlich, Susan, Meyerhoff, Miriam e Janet Holmes, a cura di. 2017. *The handbook of language, gender, and sexuality*. Seconda edizione. Hoboken-New York: Wiley Balckwell.
- EIGE (European Insitute for Gender Equality). 2013. *Review of the implementation if the Beijing platform for action in the EU member states: women and the media. Advancing gender equality in decision-making in media organizations*. [online] <https://op.europa.eu/en/publication-detail/-/publication/0406d3e9-84ac-48f8-b939-7f549ba77650>.
- Esaulova, Yulia, Reali, Chiara e Lisa von Stockhausen. 2013. "Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution". *Language, cognition and neuroscience* 29 (7): 781-803.

- Faaß, Gertrud e Kerstin Eckart. 2013. "Sdewac a corpus of parasable sentences from the web". In *Language processing and knowledge in the web*. 61-68. Berlino-Heidelberg: Springer.
- Fabre, Cécile e Alessandro Lenci. 2015. "Distributional semantics today". In *ATALA, Revue de l'association pour le traitement automatique des langues* 56 (2): 7-22.
- Fadda, Maria Laura. 2012. "Differenza di genere e criminalità. Alcuni cenni in ordine ad un approccio storico, sociologico e criminologico". *Diritto penale contemporaneo*. 20 settembre. [online] <https://archivioldpc.dirittopenaleuomo.org/d/1717-differenza-di-genere-e-criminalita-alcuni-cenni-in-ordine-ad-un-approccio-storico-sociologico-e-cri>.
- Ferguson, Marjorie. 1978. "Imagery and ideology: the cover photographs of traditional women's magazines". In *Heart and home. Images of women in the mass media*, a cura di Gaye Tuchman, Arlene Kaplan e James Benét. 97-115. New York: Oxford University Press.
- Ferguson, Marjorie. 1990. "Images of power and feminist fallacy". *Critical studies in mass communication* 6: 20-25.
- Firth, Rupert J. 1935. "The technique of semantics." *Transactions of the philological society* 34 (1): 36-73.
- Firth, Rupert J. 1957. "A synopsis of linguistic theory, 1930-1955". In *Studies in linguistic analysis*. 1-32. Oxford: Blackwell.
- Flannery, Regina. 1946. "Men's and women's speech in Gros Ventre". *International Journal of American Linguistics* 12: 133-135.
- Formato, Federica. 2016. "Linguistic markers of sexism in the Italian media: a case study of *ministra* and *ministro*". *Corpora* 11 (3): 371-399.

- Formato, Federica. 2018. *Gender, discourse and ideology in Italian*. Palgrave Mcmillan.
- Foucault, Michel. 1969. *L'archéologie du savoir*. Parigi: Gallimard. Trad. it. 1971. *L'archeologia del sapere*. Milano: Rizzoli.
- Friedan, Betty. 1963. *The feminine mystique*. New York: Dell. Trad. it. 1964. *La mistica della femminilità*. Milano: Edizioni di Comunità.
- Fusco, Fabiana. 2020. «L'abitudine fa la sindaca e l'avvocata». Il genere femminile nella lingua italiana, anche a partire da Alma Sabatini". In *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere*, a cura di Sergia Adamo, Giulia Zanfabro e Elisabetta Tigani Sava. 37-58. Trieste: EUT.
- Gallagher, Margaret. 2001. *Gender setting. New agendas for media monitoring and advocacy*. New York e Londra: Zed Books.
- Gallagher, Margaret. 2006. "Sintesi dei risultati GMMP 2005 a livello mondiale". *Inchiesta* 153: 10-12.
- Gallagher, Margaret. 2011. "Gender and communication policy: struggling for space." In *The handbook of global media and communication policy*, a cura di Robin Mansell e Marc Raboy. 449–466. Oxford: Wiley-Blackwell.
- Gallagher, Margaret. 2017. "Gender and media: a critical analysis after 20 years of the BPfA." *ComPol: Comunicazione politica* 2: 191–208.
- Garg, Nikil, Schiebinger, Londa, Jurafsky, Dan e James Zou. 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes". *PNAS Latest articles*. 115 (16): E3635-E3644.
- Garnham, Alan, Oakill, Jane e David Reynolds. 2002. "Are inferences from stereotyped role names to characters' gender made elaboratively?". *Memory & cognition* 30 (3): 439-446.

- Gauntlett, David. 2002. *Media, gender and identity. An introduction*. Prima edizione. Londra e New York: Routledge
- Gauntlett, David. 2008. *Media, gender and identity. An introduction*. Second edition. Londra-New York: Routledge.
- Gerbner, George e Kathleen Connolly. 1978. "Television as new religion". *New catholic world* 221 (1322): 52-56.
- Gerbner, George. 1978. "The dynamics of cultural resistance". In *Heart and home. Images of women in the mass media*, a cura di Gaye Tuchman, Arlene Kaplan e James Benét. 46-50. New York: Oxford University Press.
- Gheno, Vera. 2019. *Femminili singolari. Il femminismo è nelle parole*. Firenze: Effequ.
- Giangrande, Rossella. 2001. "bell hooks". *Lingue e linguaggi* 5: 111-124.
- Gilman, Charlotte Perkins. 1911. *The man-made world, or, Our androcentric culture*. New York, NY: Charlotte Company.
- Giomi, Elisa e Sveva Magaraggia. 2017. *Relazioni brutali. Genere e violenza nella cultura mediale*. Bologna: Il Mulino.
- Giomi, Elisa. 2016. "Comment on Bandelli and Porcelli/1. Against Moral Panic, in Defence of Data". *Sociologica* 2: 0-0.
- Giorcelli, Silvia, Spanò, Maria, Raus, Rachele, Abouyaala, Miriam, Catrano, Igor e Viviana Patti. 2015. *Un approccio di genere al linguaggio amministrativo*. Università degli studi di Torino. [online] https://www.unito.it/sites/default/files/linee_guida_approccio_genere.pdf.
- GiULiA giornaliste, a cura di. 2019. *Stereotipi. Donne nei media*. Milano: Ledizioni.
- GiULiA giornaliste, a cura di. 2021a. *Donne sport e media. Idee guida per una diversa informazione*. Roma: All Around.

- GiULiA giornaliste, a cura di. 2021b. *Stop violenza: le parole per dirlo*. Roma: All Around.
- Giusti, Giuliana e Gabriele Iannaccaro, a cura di. 2020. *Language, gender and hate speech. A multidisciplinary approach*. Venezia: Edizioni Ca' Foscari.
- Giusti, Giuliana e Susanna Regazzoni, a cura di. 2009. *Mi fai male... Atti del convegno. Venezia, Auditorium Santa Margherita 8-19-20 novembre 2008*. Venezia: Cafoscarina.
- Giusti, Giuliana, a cura di. 2011a. *Nominare per esistere. Atti del primo convegno Lingua e identità di Genere. Venezia, Auditorium S. Margherita, 19 settembre 2011*. Venezia: Cafoscarina.
- Giusti, Giuliana. 2011b. "Riferimento al genere e costruzione di identità". In *Nominare per esistere. Atti del primo convegno Lingua e identità di genere. Venezia, Auditorium S. Margherita, 19 settembre 2011*, a cura di Giuliana Giusti. 9-24. Venezia: Cafoscarina.
- Giusti, Giuliana. 2015. "Ruoli e nomi di ruolo in classe: una prospettiva di genere". In *Identità e discorsi. Studi offerti a Franca Orletti*, a cura di Laura Mariottini. 39-54. Roma: TrePress.
- Giusti, Giuliana. 2016a. "La frontiera della lingua: una questione ancora irrisolta". In *Cinquant'anni non sono bastati*, a cura di Anna Maria Isastia e Rosanna Oliva. 239-245. Trieste: Scienza Express, Trieste.
- Giusti, Giuliana. 2016b. "Metacompetenza linguistica e costruzione dell'identità di genere". In *Siamo le parole che usiamo*, a cura di Veronica Bacci Bonivento, Nadia Cairo, Julia Di Candia, Alisa del Re, Bruna Mura e Lorenza Perini. 21-25. Padova: Padova University Press.
- Giusti, Giuliana. 2018. "Lingua italiana e parità di genere, ricerca e formazione linguistica a Ca' Foscari". In *Le lingue occidentali nei 150 anni di storia di Ca' Foscari*, a cura di Anna Cardinaletti, Laura Cerasi e Patrizio Rigobon. Venezia:

Edizioni Ca' Foscari. [online] <http://doi.org/10.30687/978-88-6969-262-8/017>.

Giusti, Giuliana. 2022. "Inclusività della lingua italiana, nella lingua italiana: come e perché. Fondamenti teorici e proposte operative". *DEP Rivista telematica di studi sulla memoria femminile* 48: 1-19. [online] https://www.unive.it/pag/fileadmin/user_upload/dipartimenti/DSLCC/documenti/DEP/numeri/n48/05_Giusti.pdf.

Goffman, Irving. 1979. *Gender advertisements*. New York: Haper Colophon Books.

Gomez Gane, Yorick, a cura di. 2017. «*Quasi una rivoluzione*». *I femminili di professione e cariche in Italia e all'estero*. Firenze: Accademia della Crusca.

Gonen, Hila, Kementchedjhieva, Yova, Goldberg, Yoav. 2019. "How does grammatical gender affect noun representations". In *Proceedings of the international Conference on Language Resources and Evaluation (LREC 2018, Miyasaki, Giappone)*.

Grave, Edouard, Bojanowski, Piotr, Gupta, Prakhar, Joulin, Armand e Tomas Mikolov. 2018. "Learning word vectors for 157 languages". In *Proceedings of the eleventh international conference on Language Resources and Evaluation (LREC 2018), Miyasaki, Giappone*. 3483-3487.

Gray, John. 1992. *Men are from Mars, women are from Venus*. New York: HarperCollins.

Greenwald, Anthony G., McGhee, Debbie E. e Jordan L. K. Schwartz. 1998. "Measuring individual differences in implicit cognition: the implicit association test". *Journal of personality and social psychology* 74 (6): 1464-1480.

Grossi, Giorgio e Elisabetta Ruspini. 2007. "Introduzione". In *Ofelia e Parsifal. Modelli e differenze di genere nel mondo dei media*, a cura di Giorgio Grossi e Elisabetta Ruspini. XI-XLIII. Milano: edizioni libreria Cortina.

- Guaraldo, Olivia. “(In)significante padrone. Media, sesso e potere nell’Italia contemporanea”. In *Filosofia di Berlusconi*, a cura di Carlo Chiurco. 97-128. Verona: Ombre Corte.
- Guérin-Pace, France e Elodie Baril. 2018. “Les outils de la statistique textuelle pour analyser les *corpus* de données d’enquêtes de la statistique publique”. In *Proceedings of the 14th international conference on statistical analysis of textual data (JADT 2018)*. 359-366. Roma: Universitalia.
- Haas, Mary R. 1944. “Men’s and women’s speech in Koasati”. *Language* 30: 60-69.
- Hamilton, William L., Leskovec Jure e Dan Jurafsky. 2016. “Diachronic word embeddings reveal statistical laws of semantic change”. In *Proceedings of the 54th annual meeting of the association for computational linguistics. Berlino, Germania*, a cura di Katrin Erk e Noha Smith. Vol. 1, Long Papers. 1489–1501.
- Hanitzsch, Thomas e Folker Hanush. 2012. “Does gender determine journalists’ professional views? A reassessment based on cross-national evidence”. *European journal of communication* 27 (3): 257-277.
- Haraway, Donna. 1991. “A cyborg manifesto: science, technology, and socialist feminism in the late twentieth century”. In *Simians, cyborgs and women: The reinvention of nature*. 149-181 New York: Routledge. Trad. it. 1995 *Manifesto cyborg*. Milano: Feltrinelli.
- Harris, Zellig Sabbetai. 1954. “Distributional structure”. *WORD* 10 (2-3): 146-162.
- Hellinger, Marlis e Christine Bierbach. 1993. *Eine Sprache für beide Geschlechter*. Bonn: Deutsche UNESCO Kommission.
- Hellinger, Marlis e Hadumod Bußmann, a cura di. 2001-2003. *Gender across languages. The linguistic representation of women and men*. Vol. 1-3. Amsterdam-Philadelphia: John Benjamins.

- Hellingher, Marlis e Heiko Montschenbacher, a cura di. 2004. *Gender across languages. The linguistic representation of women and men*. Vol. 4. Amsterdam-Philadelphia: John Benjamins.
- Hockett, Charles. 1958. *A course in modern linguistics*. New York: MacMillan.
- Hofstede, Geert. 2001. *Culture's consequences: comparing values, behaviors, Institutions and organizations across nations*. Seconda edizione. Thousand Oaks: SAGE.
- Holmes, Janet e Miriam Meyerhoff, a cura di. 2003. *The handbook of language and gender*. Bleckweel Publishing Ltd.
- Holmes, Janet. 1997. "Women Language and Identity". *Journal of sociolinguistics* 1/2: 195-223.
- hooks, bell. 1984. *Feminist theory: from margin to center*. Boston: South and Press.
- hooks, bell. 1988. *Elogio del margine. Razza, sesso e mercato culturale*. Milano: Feltrinelli.
- hooks, bell. 1992. *Black looks: race and representation*. Boston: South and Press.
- Houdebine-Gravaud, Anne-Marie. 1998. *La féminisation des noms de métiers*. Parigi: L'Harmattan.
- Houdebine, Anne-Marie. 1987. "Le français au féminin". *La Linguistique* 23(1):13-34.
- Humprecht, Edda e Frank Esser. 2017. "A glass ceiling in the online age? Explaining the underrepresentation of women in online politica news". *European journal of communication* 32 (5): 439-456.
- IAWRT (Association for Women in Radio and Television). 2020. *Safety hanbook for women journalists*. Seconda edizione. [online]

<https://www.iawrt.org/sites/default/files/field/pdf/2020/11/SAFETY%20HANDBOOK%202nd%20EDITION%20%28ENGLISH%20VERSION%29.pdf>.

IFJ (International Federation of Journalists). s.d. *Guidelines for reporting violence against women*. [online] https://www.ifj.org/fileadmin/user_upload/IFJ_Guidelines_for_Reporting_on_Violence_Against_Women_EN.pdf.

Irmen, Lisa e Eva Schuman. 2011. "Processing grammatical gender of role nouns: evidence from eye movements". *Journal of cognitive psychology* (23) 8: 998-1014.

Irmen, Lisa e Nadaja Roßberg. 2004. "Gender markedness of language. The impact of grammatical and nonlinguistic information on the mental representation of person information". *Journal of language and social psychology* 23 (3): 272-307.

ISIMM Ricerche e Università degli studi Roma Tre. 2016. *Monitoraggio sulla rappresentazione femminile nella programmazione della Rai. Anno 2016*. [online] https://www.rai.it/dl/doc/1586513103808_2016%20Monitoraggio%20della%20figura%20femminile%20nei%20programmi%20Rai.pdf.

ISIMM Ricerche e Università degli studi Roma Tre. 2017. *Monitoraggio sulla rappresentazione femminile nella programmazione della Rai. Anno 2017*. [online] https://www.rai.it/dl/doc/1586512843025_2017%20Analisi%20dei%20contenuti%20-%20Monitoraggio%20della%20Figura%20Femminile%20nei%20programmi%20Rai.pdf.

ISTAT. 2015. *Lingua italiana, dialetti, lingue straniere*. [online] <https://www.istat.it/it/archivio/207961>.

- ISTAT. 2019. *Conciliazione tra lavoro e famiglia*. [online]
<https://www.istat.it/it/archivio/235619>.
- ISTAT. 2021. *Bes 2020. Il benessere equo e sostenibile in Italia*. [online]
https://www.istat.it/it/files//2021/03/BES_2020.pdf.
- Ivanova, Kremena, Heid, Ulrich, Schulte im Walde, Sabine e Adam Kilgarrif. 2008.
 “Evaluating a German sketch grammar. A case study on noun phrase case”.
 In *Proceedings of the Sixth International Conference on Language Resources
 and Evaluation (LREC 08), Marrakech, Marocco*.
- IZI, ISIMM Ricerche e InfoJuice. 2020. *Monitoraggio rappresentazione della figura
 femminile, pluralismo di temi, soggetti e linguaggi e contributo alla
 creazione di coesione sociale nella programmazione Rai 2020*. [online]
https://www.rai.it/dl/doc/1623168190198_Monitoraggio%20Figura%20femminile%20Pluralismo%20sociale%20e%20Coesione%20sociale%20Rai%202020_Analisi%20dei%20contenuti.pdf.
- Jackendoff, Ray. 1998. *Linguaggio e natura umana*. Bologna: Il Mulino.
- Jacqmain, Monique. 2005. “Ministro, ministra, signora ministro: quali appellativi
 per le donne ‘in carriera?’”. *Studi di grammatica italiana* 24: 211-217.
- Jespersen, Otto. 1922. *Language: its nature, development and origins*. New York:
 McMillan.
- Jones, Laura. 2020. *Women' representation and voice in media coverage of the
 coronavirus crisis*. Londra: The Global Institute for women's leadership,
 King's college.
- Katz, Elihu e Paul F. Lazarsfeld. 1955. *Personal influence. The role played by people
 in the flow of communication*. Glencoe: The Free Press. Trad. It. 1968.
L'influenza personale nelle comunicazioni di massa. Torino: ERI.

- Kennison, Sheila e Jessie L. Trofe. 2003. "Comprehending Pronouns: A Role for Word-Specific Gender Stereotype Information". *Journal of psycholinguistic research* 32 (3): 355-378.
- Kilgarriff, Adam, Reddy, Siva, Pomikálek, Jan e PVS Avinesh. 2010. "A corpus factory for many language". In *Workshop on web services and processing pipelines. LREC (Language Resources Evaluation Conference) 2010. Malta*.
- Kreiner, Hamutal, Sturt, Patrick e Simon Garrod. 2008. "Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements". *Journal of memory and language* 58: 239-261.
- Lahlou, Saadi. 2012. "Text mining methods: an answer to Chartier and Meunier". *Papers on social representations* 20 (38): 1-7.
- Lakoff, Robin T. e Mary Bucholtz. 2004. *Language and women's place. Text and commentaries. Revised and expanded edition*. New York: Oxford University Press.
- Lakoff, Robin. 1973. "Language and woman's place". *Language in society* (2)1: 45-79.
- Lakoff, Robin. 1975. *Language and woman's place*. New York: Harper & Row.
- Lazersfeld, Paul F. 1948. "Communication research and the social psychologist". In *Current trends in social psychology*, a cura di Wayne Dennis. 218-273. Pittsburg: University of Pittsburg Press.
- Leccardi, Carmen, a cura di. 2002. *Trai generi. Rileggendo le differenze di genere di generazione di orientamento sessuale*. Milano: Guerini.
- Leech, Geoffrey 1991. "The state of art in corpus linguistics". In *English corpus linguistics: studies in honour of Jan Svartvit*, a cura di Karim Aijmer e Bengt Altenberg. 8-29. Londra: Longman

- Legge 14 marzo 1985 n. 32. Ratifica ed esecuzione della Convenzione sull'eliminazione di ogni forma di discriminazione nei confronti della donna, adottata a New York il 18 dicembre 1979. *Gazzetta ufficiale della Repubblica italiana*. Serie generale. Supplemento ordinario. 89. Roma, 10 giugno 1985. [online] <https://www.gazzettaufficiale.it/eli/id/1985/04/15/085U0132/sg>.
- Legge 22 febbraio 2000 n. 28. Disposizioni per la parità di accesso ai mezzi di informazione durante le campagne elettorali e referendarie e per la comunicazione politica. *Gazzetta ufficiale della Repubblica italiana*. Serie generale. 43: 5-12. Roma, 22 febbraio 2000. [online] <https://www.gazzettaufficiale.it/eli/gu/2000/02/22/43/sg/pdf>.
- Legge 23 novembre 2012 n. 215. Disposizioni per promuovere il riequilibrio delle rappresentanze di genere nei consigli e nelle giunte degli enti locali e nei consigli regionali. Disposizioni in materia di pari opportunità nella composizione delle commissioni di concorso nelle pubbliche amministrazioni. *Gazzetta ufficiale della Repubblica italiana*. Serie generale. 288: 1-2. Roma, 11 dicembre 2012. [online]. <https://www.gazzettaufficiale.it/eli/gu/2012/12/11/288/sg/pdf>.
- Lemon, Judith. 1978. "Dominant or dominated? Women on prime-time television. In *Heart and home. Images of women in the mass media*, a cura di Gaye Tuchman, Arlene Kaplan e James Benét. 51-68. New York: Oxford University Press.
- Lenci, Alessandro, Montemagni, Simonetta e Vito Pirrelli. 2016. *Testo e computer. Elementi di linguistica computazionale*. Roma: Carocci.
- Lenci, Alessandro, Sahlgren, Magnus, Jeuniaux, Patrick, Gyllensten, Aamru Cuba e Martina Miliani. 2021. "A comprehensive comparative evaluation and analysis of Distributional Semantic Models". [online] arXiv preprint arXiv:2105.09825.

- Lenci, Alessandro. 2018. "Distributional models of word meaning". *Annual review of linguistics* 4: 151-171.
- Lenoble-Pinson, Michéle, a cura di. 1994. Prima edizione. *Mettre au féminin. Guide de féminisation des noms de métier, fonction, grade ou titre*. Bruxelles: Ministère de la Communauté française de Belgique. Conseil supérieur de la langue française et service de la langue française. [2005. Seconda edizione].
- Lepri, Sergio. 1987. "Ammissione di colpa e chiamata di correo". In Alma Sabatini, *Il sessismo nella lingua italiana*. 21-23. Roma: Istituto poligrafico e Zecca dello Stato.
- Lévi-Strauss, Claude. 1955. *Tristes tropiques*. Paris: Plon. Trad. it. 1960. *Tristi tropici*. Milano: Il Saggiatore.
- Lewis, Molly e Gary Lupyan. 2020. "Gender stereotypes are reflected in the distributional structure of 25 languages". *Nature human behaviour* 4: 1021-1028.
- Lison, Pierre e Jörg Tiedemann. 2016. "OpenSubtitles2016: extracting large parallel corpora from movie and TV subtitles". In *Proceedings of tenth international conference on Language Resources and Evaluation (LREC) 2016*. Portorož, Slovenia. 923-929.
- Livolsi, Marino. 2013. "Premessa. Watching Dallas e watching le soap". In Ian Ang, *Watching Dallas. Cultura di massa e imperialismo culturale*, a cura di Marino Livolsi. Roma: Armando editore.
- Lomonaco, Vincenzo. 2015. "Word2vec on the Italian language: first experiments". Slideshare. 19 febbraio. [online] <https://www.slideshare.net/VincenzoLomonaco/word2vec-on-the-italian-language-first-experiments>.

- Lopate, Carol. 1978. "Jackie!". In *Heart and home. Images of women in the mass media*, a cura di Gaye Tuchman, Arlene Kaplan e James Benét. 130-140. New York: Oxford University Press.
- Losito, Gianni. 1996. *L'analisi del contenuto nella ricerca sociale*. Seconda edizione riveduta e corretta. Milano: Franco Angeli.
- Losito, Gianni. 1998. *Il potere dei media*. Seconda edizione. Roma: Carocci.
- Loubère, Lucie e Pierre Ratinaud. 2014. *Documentation IRaMuTeQ 0.6 alpha 3 version 0.1*. [online] http://www.iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf.
- Luraghi, Silvia e Anna Olita. 2006. *Linguaggio e genere*. Roma: Carocci.
- Malinowski, Bronislaw. 1929. *The sexual life of savages in North-Western Melanesia. An ethnographic account of courtship, marriage, and family life among the natives of the Trobriand Islands, British New Guinea*. Londra: Routledge.
- Manzini, Thomas, Lim, Yao Chong, Tsvetkov, Yulia e Alan W. Black. 2019. "Blak is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings". *Proceedings of NAACL-HLT. Minneapolis, Minnesota, USA*. 615–621.
- Marcato, Gianna e Eva M. Thüne. 2002. "Gender and female visibility in Italian". In *Gender across languages. The linguistic representation of women and men*, a cura di Marlis Hellinger e Hadumod Bußmann. Vol. 2. 188-217. Amsterdam-Philadelphia: John Benjamins.
- Marcato, Gianna, a cura di. 1995. *Donna e linguaggio. Atti del convegno internazionale di studi. Sappada/Plodn (Belluno)*. Padova: CLEUP.

- Marchand, Pascal e Pierre Ratinaud. 2014. "Analyse lexicométrique des tweets sur le #mariagepourtous". *Colloque "Comprendre les mondes sociaux 2014"*, organisé par Le Labex Structuration des Mondes Sociaux (SMS) de l'Université Toulouse Jean-Jaures-campus Mirail. Toulouse, Centre de congrès Dominique Baudis, 7-9 avril 2014.
- Marelli, Marco. 2017. "Word-embeddings Italian semantic spaces: a semantic model for psycholinguistic research". *PSIHOLOGIJA* 50 (4): 503-520.
- Marinucci, Elena. 1986. "Presentazione". In Gioia Di Cristofaro Longo, *Immagine donna. Modelli di donna emergenti nei mezzi di comunicazione di massa*. 10. Roma: Istituto poligrafico e Zecca dello Stato.
- Marinucci, Elena. 1987. "Presentazione". In Alma Sabatini, *Il sessismo nella lingua italiana*. 11. Roma: Istituto poligrafico e Zecca dello Stato.
- Masini, Andrea. 2016. "L'italiano contemporaneo e la lingua dei media". In *La lingua italiana e i mass media*, a cura di Ilaria Bonomi e Silvia Morgana. Nuova edizione. 17-40. Roma: Carocci.
- Mazzoleni, Gianpietro. 2012. *La comunicazione politica*. Terza edizione. Bologna: Il Mulino.
- McCurdy, Katherine e Ogus Serbetçi. 2020. *Grammatical gender associations outweigh topical gender*. [online] <https://arxiv.org/pdf/2005.08864.pdf>.
- McRobbie, Angela. 1999. *In the culture society, art, fashion and popular music*. Londra: Routledge.
- Mediawatch. 1995. *Global Media Monitoring Project. Women's Participation in the News*. Toronto: National Watch on Images of Women in the Media Inc.
- Merkel, Elisa. "Le due facce del linguaggio". In *Siamo le parole che usiamo*, a cura di Veronica Bacci Bonivento, Nadia Cario, Julia Di Campo, Alisa Del Re, Bruna Mura e Lorenza. 47-51. Padova: Padova University Press.

- Messina, Simona. 2013. "Che carino! Alcune considerazioni sulla «lingua delle donne» nella *fiction* tra stereotipo e realtà". *Quaderns d'Italia* 18: 265-283.
- Meucci, Alessandro. 2001. "Agenzie di stampa e quotidiani. Una notizia dall'ANSA ai giornali". In *Working papers del Dipartimento di Scienze Storiche, Giuridiche, Politiche e Sociali dell'Università degli Studi di Siena*. Working paper 42. [online] <https://www.dispi.unisi.it/sites/st06/files/allegatiparagrafo/22-05-2013/wp42.pdf>.
- Michelot, Vincent. 2019. "Gilet gialli, dalla rivolta fiscale alla crisi della rappresentatività". *Il Mulino* 1: 74-82.
- Mikolov, Tomas, Chen, Kai, Corrado, Greg e Jeffrey Dean. 2013. "Efficient estimation of word representations in vector space". *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA*.
- Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg e Jeffrey Dean. 2013. "Distributed representations of words and phrases and their compositionality". [online] <https://arxiv.org/abs/1301.3781>.
- Ministerium für Justiz Frauen, Jugend und Familie des Landes Schleswig-Holstein. 2002. *Meher frauen in die sprache. Leitfaden zur geschlechtergerechten formulierunt*. Kiel, Pirwitz Druck.
- Ministero dello Sviluppo Economico. 2011. Contratto Nazionale di Servizio tra il Ministero dello sviluppo economico e la Rai- Radiotelevisione italiana S.p.A. -2018-2022. Gazzetta ufficiale della Repubblica Italiana. Serie generale. 147: 71-102. [online] <https://www.mise.gov.it/images/stories/documenti/Contratto-di-servizio-Rai-2018-2022.pdf>.

- Ministero dello Sviluppo Economico. 2018. Approvazione del Contratto Nazionale di Servizio tra il Ministero dello sviluppo economico e la Rai-Radiotelevisione italiana S.p.A. -2018-2022. Gazzetta ufficiale della Repubblica Italiana. Serie generale. 55: 25-36. [online] <https://www.mise.gov.it/images/stories/documenti/Contratto-di-servizio-Rai-2018-2022.pdf>.
- Mitchell, Juliet. 1966. "Women's estate". *New left review* 40: 11-37. Trad. it. 1972. *La condizione della donna*. Torino: Einaudi.
- MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca). 2018. *Linee guida per l'uso del genere nel linguaggio amministrativo del MIUR*. [online] <https://www.miur.gov.it/-/linee-guida-per-l-uso-del-genere-nel-linguaggio-amministrativo-del-miur>.
- Molfino, Francesca. 2006. *Donne, politica e stereotipi. Perché l'ovvio non cambia?* Milano: Baldini Castoldi Dalai.
- Monteith, Lindsey L. e Jeremy Pettit. 2011. "Implicit and explicit stigmatizing attitudes and stereotypes about depression". *Journal of social and clinical psychology* 30 (5): 484-505.
- Moreno, Mariola e Pierre Retinaud. 2015. *Manual uso de IRaMuTeQ. Version 0.7 alpha 2*. [online] <http://www.iramuteq.org/documentation/fichiers/guia-iramuteq>.
- Moscovici, Serge. 1969. "Introduction". In Claudine Herzlich, *Santé et maladie. Analyse d'une représentation sociale*. 8-15. Parigi: Mouton.
- Moscovici, Serge. 1976. *Social influences and social change*. London: Academic presse.
- Moscovici, Serge. 1984. "The phenomenon of social representations". In *Social representations*, a cura di Robert M. Farr e Serge Moscovici. 3-69. Cambridge: Cambridge University Press. Trad. it. 1989. "Il fenomeno delle

- rappresentazioni sociali". In *Rappresentazioni sociali*, a cura di Robert M. Farr e Serge Moscovici. 23-94. Bologna: Il Mulino.
- Mulvey, Laura. 1975. "Visual pleasure and narrative cinema". *Screen*. 16 (2): 6-18.
- Nardone, Chiara. 2016. "Asimmetrie semantiche di genere: un'analisi sull'italiano del corpus itWaC". *gender/sexuality/italy* 3. [online] <https://www.gendersexualityitaly.com/wp-content/uploads/2016/12/1.-Nardone.pdf>.
- Nardone, Chiara. 2018. *Lingua, genere e lavoro in Italia e in Germania: un'analisi comparativa su annunci di lavoro, sui corpora itWac e deWac e sulla stampa*. Tesi di Dottorato. Bologna-Düsseldorf: Alma Mater Studiorum Università di Bologna – Heinrich-Heine Universität Düsseldorf. [online] <http://amsdottorato.unibo.it/8470/#>.
- Noell-Neumann, Elisabeth. 1973. "Return to the concept of powerful mass media". *Studies of broadcasting* 9: 66-112.
- Noell-Neumann, Elisabeth. 1977. "Turbulence in the climate of opinion: methodological application of the spiral of silence". *Public opinion quarterly* 41: 143-158.
- Noell-Neumann, Elisabeth. 1980. *Die schweigespirale: oeffentliche meinung*. Monaco: Piper Verlag. Trad. it. 2002. *La spirale del silenzio*. Roma: Meltemi.
- Nosek, Brian A., Banaji, Mahzarin R. e Anthony G. Greenwald. 2002a. "Harvesting implicit group attitudes and beliefs from a demonstration web site." *Group dynamics: theory, research, and practice* 6 (1): 101–115.
- Nosek, Brian A., Banaji, Mahzarin R. e Anthony G. Greenwald. 2002b. "Math = Male, Me = Female, Therefore Math ≠ Me". *Journal of personality and social psychology* 83(1): 44–59.

- Oakill, Jane, Garnham, Alan e David Reynolds. 2005. "Immediate activation of stereotypical gender information". *Memory & cognition* 33(6): 972-983.
- Oddone, Cristina. 2015. "Dal silenzio al rumore: come parlare di violenza maschile contro le donne? Evoluzioni del discorso pubblico prodotto dai media dagli anni Settanta a oggi". *Egeria. Rivista dell'Istituto superiore di scienze religiose* 8-9: 11-27.
- OdG (Ordine dei Giornalisti). 2020. *Modifiche al Testo unico dei doveri del giornalista*. Approvate dal Consiglio Nazionale il 19 novembre 2020. [online] https://www.odg.it/wp-content/uploads/2020/11/Modifiche-al-TU-giornalisti_19_11_20-.pdf.
- OdG (Ordine dei Giornalisti). 2021. *Testo unico dei doveri del giornalista. In vigore dal 1° gennaio 2021*. [online] <https://www.odg.it/testo-unico-dei-doveri-del-giornalista/24288>.
- ONU (Organizzazione delle Nazioni Unite). 1979. *Convenzione sulla eliminazione di ogni forma di discriminazione nei confronti della donna*. [online] <https://www.ohchr.org/Documents/ProfessionalInterest/cedaw.pdf>.
- ONU (Organizzazione delle Nazioni Unite). 1995. *Dichiarazione e Piattaforma d'azione di Pechino*. [online] <https://www.un.org/womenwatch/daw/beijing/pdf/BDPfA%20E.pdf>.
- ONU (Organizzazione delle Nazioni Unite). 2015. *Risoluzione adottata dall'Assemblea Generale il 25 settembre 2015. Trasformare il nostro mondo: l'Agenda 2030 per lo sviluppo sostenibile*. [online] <https://unric.org/it/wp-content/uploads/sites/3/2019/11/Agenda-2030-Onu-italia.pdf>.
- ONU (Organizzazione delle Nazioni Unite). 2018. *Agreed conclusions: CSW62. Commission on the Status of sixty-second session of the Commission on the Status of Women*. [online] <https://www.unwomen.org/en/csw/csw62-2018/session-outcomes>.

- ONU (Organizzazione delle Nazioni Unite). 2020. *Review and appraisal of the implementation of the Beijing declaration and platform for action and the outcomes of the twenty-third special session of the general assembly* [online] <https://undocs.org/E/CN.6/2020/3>.
- Orletti, Franca. 2001. *Identità di genere nella lingua, nella cultura, nella società*. Roma: Armando Editore.
- Osservatorio Europeo sulla Sicurezza. 2017. *Rapporto sulla sicurezza in Italia e in Europa*. [online] https://www.osservatorio.it/wp-content/uploads/2017/02/Fondazione-Unipolis-X-Rapporto-sulla-sicurezza-e-insicurezza-sociale-2017_light.pdf.
- Osterhout, Lee, Bersick, Michael e Judoth McLaughlin. 1997. "Brain potentials reflect violations of gender stereotypes". *Memory & cognition* 25 (3): 273-285.
- Padovani, Claudia e Rossella Bozzon. 2020. "Media gender-equality regimes. Exploring media organisations' policy adoption across nations". In *Comparing gender and media equality across the globe*, a cura di Djerf-Pierre, Monika e Maria Esdröm. 99-144. Gothenburg: Nordicom.
- Padovani, Claudia. 2018. "Gendering media policy research and communication governance". *Javnost-The Public* 25 (1/2): 256-264.
- Pandolfini, Valeria. 2017. *Il sociologo e l'algoritmo. L'analisi dei dati testuali al tempo di Internet*. Milano: Franco Angeli.
- Panighel, Marta. 2014. "La questione della 'lingua al femminile'. Aspetti, temi, stereotipi sociali (con una ricerca sul campo)". *Rivista italiana di dialettologia* 37: 161-204.
- Parlamento europeo. 2008. *Risoluzione del Parlamento europeo del 3 settembre 2008 sull'impatto del marketing e della pubblicità sulla parità tra donne e*

uomini (2008/2038(INI)). [online]
http://www.unionedonneinitalia.org/stage/Risoluzione_PA_2008.pdf.

Pavan, Elena. 2016. "Comment on Bandelli and Porcelli/2 The Politics of Knowledge Production in the Field of Violence Against Women". *Sociologica* 2: 0-0.

Pearson, Karl. 1895. "Note on regression and inheritance in the case of two parents". *Proceedings of the Royal Society of London*. 58: 240-242.

Pearsons, Talcott. 1954. *Essays in sociological theory*. Glencoe: The Free Press.

Pennebaker, James W., Francis, Martha E. e Roger J. Booth. 2001. *Linguistic inquiry and word count: Liwc 2001*. Mahway: Lawrence Erlbaum Associates.

Pennington, Jeffrey, Socher, Richard e Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation". In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar*. 1532-1543.

Perrotta, Marta. 2015. "Corpi senza voce o voci senza copro? Donne al microfono nell'informazione radiofonica". *Problemi dell'informazione* 3: 503-526.

Pescia, Lorenza. 2010. "Il maschile e il femminile nella stampa scritta del Canton Ticino (Svizzera) e dell'Italia". In *Che genere di lingua?*, a cura di Maria S. Sapegno. 57-74. Roma: Carocci.

Phillips, Barbara. 1978. "Is Ms. Just another member of the *Family Circle*?". In *Heart and home. Images of women in the mass media*, a cura di Gaye Tuchman, Arlene Kaplan e James Benét. 116-129. New York: Oxford University Press.

Portraying Politics Project Partners. 2006. *Portraying politics. A toolkit on gender and television*. [online] www.coe.int/t/dg4/cultureheritage/mars/source/resources/references/others/53%20-%20Portraying%20Politics%20-%202006%20COMP.pdf.

- Prothro, Terry E. e Levon H. Melikian. 1953. "Studies in stereotypes: V. familiarity and the kernel of truth hypothesis". *The journal of social psychology* 41: 3-10
- Provincia di Bolzano. 2012. *Direttive per il rispetto del genere nei testi dell'Amministrazione provinciale*. [online] http://www.provincia.bz.it/politica-diritto-relazioni-estere/diritto/downloads/Richtlinien_geschl_Spr_ita_HP.pdf.
- Pusch, Luise F. 1984. *Das Deutsche als Männersprache*. Frankfurt am Main: Suhrkamp.
- Rai-Osservatorio di Pavia. 2014. *Monitoraggio sulla rappresentazione femminile nella programmazione televisiva in TV. Anno 2014*. [online] <https://www.osservatorio.it/wp-content/uploads/MonitorDonne2014.pdf>.
- Rai-Osservatorio di Pavia. 2015. *Monitoraggio sulla rappresentazione femminile nella programmazione televisiva in TV. Anno 2015*. [online] https://www.rai.it/dl/docs/MONITOR_DONNE_ANNO_2015.pdf.
- Rai-Osservatorio di Pavia. 2018. *Monitoraggio sulla rappresentazione femminile nella programmazione della Rai. Anno 2018*. [online] https://www.rai.it/dl/doc/1586509200289_2018%20Monitoraggio%20sulla%20rappresentazione%20della%20figura%20femminile.pdf.
- Rai-Osservatorio di Pavia. 2019. *Monitoraggio sulla rappresentazione femminile nella programmazione della Rai. Anno 2019*. [online] https://www.rai.it/dl/doc/1593072471021_Monitoraggio_della_Figura_Femminile_2019_Analisi_dei_contenuti.pdf.
- Rai. 2020. *Codice etico*. [online] https://www.rai.it/dl/doc/1586535485463_Codice%20Etico%20%20Marzo%202020.pdf.
- Ratinaud, Pierre e Pascal Marchand. 2015. "Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les

débats à l'Assemblée nationale (1998-2014)". *Mots. Les langages du politique* 108: 57-77.

Ratinaud, Pierre. 2009. *IRaMuTeQ: Interface de R pur les Analyse Multidimensionnelles des Textes et des Questionnaires*. [online]

Ratner, Bruce. 2009. "The correlation coefficient: its values range between +1/-1, or do they?". *Journal of targeting, measurement and analysis for marketing* 17 (2): 139-142.

Regione Abruzzo. 2019. *Guida all'utilizzo corretto del genere nel linguaggio amministrativo*. [online] https://www.regione.abruzzo.it/system/files/cug/documenti/LINGUAGGIO_DI_GENERE.pdf.

Regione Emilia Romagna. 2015. *Uno sguardo nuovo alla comunicazione pubblica. Linee guida in ottica di genere*. [online] <https://parita.regione.emilia-romagna.it/documentazione/documentazione-temi/media-e-comunicazione/linee-guida-in-ottica-di-genere-della-regione-emilia-romagna-uno-sguardo-nuovo-nella-comunicazione-pubblica>.

Reinert, Max. 1983. "Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte". *Les cahiers de l'analyse des données* 8 (2): 187-198.

Reinert, Max. 1986. "Un logiciel d'analyse lexicale". *Les cahiers de l'analyse des données* 11 (4): 471-484.

Reinert, Max. 1990. "Une méthode de classification des énoncés des *corpus* présentée à l'aide d'une application". *Les cahiers de l'analyse des données* 15 (1): 21-36.

Reinert, Max. 1993. "Les 'mondes lexicaux' et leur 'logique' à travers l'analyse statistique d'un *corpus* de récits de Cauchemars". *Langage & société* 66: 5-39.

- Richy, Célia e Heather Burnett. 2021. "Démêler les effets des stéréotypes et le genre grammatical dans le biais masculin : une approche expérimentale". *GLAD! Revue sur le langage, le genre, les sexualités*. 10. [online] <http://journals.openedition.org/glad/2839>.
- Riggi, Riccardo. 2015. *Manuale di stile. Scrivi bene e parla chiaro*. Università degli studi di Palermo. [online] <https://www.unipa.it/amministrazione/rettorato/coordinamentodeiservizi/direttorato/settorerelazioneconimediawebesociale/u.o.ufficiorelazioniconipubblicourp/.content/Manuale-di-Stile.pdf>.
- Robustelli, Cecilia. 2012. *Linee guida per l'uso del genere nel linguaggio amministrativo*. Progetto "Genere e Linguaggio, Parole e immagini della Comunicazione", svolto in collaborazione con l'Accademia della Crusca e il Comune di Firenze. [online] https://www.uniss.it/sites/default/files/documentazione/c.robustelli_line_e_guida_uso_del_genere_nel_linguaggio_amministrativo.pdf.
- Robustelli, Cecilia. 2014. *Donne, grammatica e media*. Roma: GiULiA Giornaliste.
- Ronca Daniela e Vincenzo Moscati. 2019. "The Interaction of Morphological Gender With Stereotypical Information: An Eye Tracking Study on Gender Inferences". *International journal of linguistics* 11 (4): 111-125.
- Rositi, Franco. 1988. "Analisi del contenuto". In *La ricerca nell'industria culturale*, a cura di Franco Rositi e Marino Livolsi. 59-94. Roma: La Nuova Italia Scientifica.
- Ross, Karen e Cynthia Carter. 2011. "Women and news: a long and winding road, Media". *Culture & Society* 33 (8): 1148-1165.
- Ross, Karen e Monia Azzalini. 2017. "The WIME study: context, methods and summaries", In *Gender equality and the media*, a cura di Karen Ross e Claudia Padovani. 31-43. New York: Routledge.

- Ross, Karen, Marloes, Janen e Tobias Bürger. 2020. "The media world versus the real world of women and political representation. Questioning differences and struggling for answers". In *Comparing gender and media equality across the globe*, a cura di Djerf-Pierre, Monika e Maria Esdröm. 233-257. Gothenburg: Nordicom.
- Ross, Karen, Padovani, Claudia e Margareth Gallagher. 2017. "Conclusion". In *Gender equality and the media*, a cura di Karen Ross e Claudia Padovani. 233-247. New York: Routledge.
- Ross, Karen. 2001. "Women at work: journalism as an en-gendered practice". *Journalism studies* 2 (4): 531-544.
- Rubin, Gayle. 1975. "The traffic in women: notes on the 'political economy' of sex. In *The second wave. A reader in feminist theory*, a cura di Linda Nicholson. 27-72. New York-Londra: Routledge.
- Ruspini, Elisabetta. 2003. *Le identità di genere*. Roma: Carocci.
- Sabatini, Alma. 1986. *Raccomandazioni per un uso non sessista della lingua italiana: per la scuola e l'editoria scolastica*. Roma: Istituto poligrafico e Zecca dello Stato.
- Sabatini, Alma. 1987. *Il sessismo nella lingua italiana*. Roma: Istituto poligrafico e Zecca dello Stato.
- Sabatini, Alma. 1993 *Il sessismo nella lingua italiana*. Ristampa. Roma: Istituto poligrafico e Zecca dello Stato.
- Sabatini, Francesco. 1987. "Più che una prefazione". In Alma Sabatini, *Il sessismo nella lingua italiana*. 13-19. Roma: Istituto poligrafico e Zecca dello Stato.
- Salviati, Maria Elisabeth. 2017. *Manual do aplicativo Iramuteq (versão 0.7 Alpha 2 e R versão 3.2.3)*. [online]

<http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>.

Sapegno, Maria S., a cura di. 2010. *Che genere di lingua?*. Roma: Carocci.

Sapir, Edward. 1929. "Male and female forms of speech in Yoga". In *Selected writings of Edward Sapir*, a cura di David G. Mandelbaum. 1949. Berkeley: University of California Press.

Sbisà, Marina. 2019. "Il genere tra stereotipi e impliciti". In *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere*, a cura di Sergia Adamo, Giulia Zanfabro e Elisabetta Tigani Sava. 17-26. Trieste: EUT

Senato della Repubblica. Ufficio Valutazione Impatto. 2018. *Parità vo cercando. 1948-2018. Settant'anni di elezioni in Italia: a che punto siamo con il potere delle donne?* [online]
https://www.senato.it/application/xmanager/projects/leg17/attachments/documento/files/000/028/741/Focus_8_marzo.pdf.

Shaw, Edward F. 1979. "Agenda setting and mass communication theory". *Gazette* 2. 96-105.

Shor, Eran, van de Rijt, Arnout, Miltsov, Alex, Kulkarni, Vivek e Steven Skiena. 2015. "A paper ceiling: explaining the persistent underrepresentation of women in printed news". *American sociological review* 80 (5): 960-984.

Siliato, Francesco. 2001. "Gli uomini, le donne ed I telegiornali". In *Una, nessuna... a quando centomila? La rappresentazione della donna in televisione*, a cura di Loredana Cornero. 87-104. Roma: Rai-Eri VQPT 184.

Somma, Annalisa e Gabriele Maestri, a cura di. 2020. *Il sessismo nella lingua italiana*. Blonk editore.

- Souza MAR, Wall ML, Thuler ACMC, Lowen IMV e Aida M. Peres. 2018. "The use of IRAMUTEQ software for data analysis in qualitative research". *Revista de escola de enfermagem USP*. 52: e03353.
- Spender, Dale. 1980. *Man made language*. Prima edizione. Londra: Routledge & Kegan Paul.
- Spender, Dale. 1985. *Man made language*. Seconda edizione. Londra: Routledge & Kegan Paul.
- Spender, Dale. 1998. *Man made language*. Ristampa . Londra: Pandora Press.
- Stefanowitsch, Anatol. 2022. *Corpus linguistics. A guide to the methodology*. Berlin: Language Science Press.
- Steiner, Linda. 2012. "Failed theories: explaining gender difference in journalism". *Review of communication* 2 (3): 201-223.
- Stradella, L. 1976. *La variabile sesso nell'uso descrittivo della lingua parlata: un'inchiesta preliminare*. Tesi di laurea in Lettere e Filosofia. Università degli studi di Torino.
- Stubbs, Michael. 1996. *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell.
- Sturt, Patrick. 2003. "The time-course of the application of the binding constraints in reference resolution". *Journal of memory and language* 48, 542-562.
- Tagliavini, Carlo. 1938. "Modificazioni del linguaggio nella parlata delle donne". In *AAVV, Scritti in onore di Alfredo Trombetti*. 87-142. Milano: Hoepli.
- Tannen, Deborah. 1990. *You just don't understand: women and men in conversation*. New York: William Marrow.

- Tausczik, Yla R. e James W. Pennebaker. 2010. "The psychological meaning of words: Liwc and computerized text analysis methods". *Journal of language and social psychology* 29(1):24–54.
- Thornton, Anna M. 2004. "Mozione". In *La formazione delle parole in italiano*, a cura di Maria Grossmann e Franz Rainer. 218-227. Tübingen: Niemeyer.
- Thornton, Anna M. 2009. "Constraining gender assignment rules". *Language Sciences* 31 (1): 185-208.
- Thornton, Anna M. 2020. *Per un uso della lingua italiana rispettoso dei generi*. L'Aquila: Università degli studi dell'Aquila. [online] <https://www.univaq.it/include/utilities/blob.php?item=file&table=allegato&id=4925>
- Tognini Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam-Philadelphia: John Benjamins.
- Tripodi, Rocco e Stefano Li Pira. 2017. "Analysis of Italian. Word embeddings". In *Proceedings of the fourth Italian Conference on Computational Linguistics CLiC-it 2017, Roma, Italia*, a cura di Roberto Basili, Malvina Nissim e Giorgio Satta. 321-326.
- Trömel-Plötz, Senta. 1982. *Frauensprache: Sprache der Veränderung*. Frankfurt am Main: Fisher Taschenbuch Verlag.
- Tuchman, Gaye, Daniels, Arlene Kaplan e James Benét, a cura di. 1978. *Heart and home. Images of women in the mass media*. New York: Oxford University Press.
- Tuchman, Gaye. 1978. *Making news: a study in the construction of reality*. New York: The free press.
- Tuzzi, Arjuna. 2003. *L'analisi del contenuto. Introduzione ai metodi e alle tecniche di ricerca*. Roma: Carocci.

UN Women. 2018. *CSW64 preparations*. [online]
<https://www.unwomen.org/en/csw/csw64-2020/preparations#regional-review-processes>.

UN Women. 2019. *Expert group meeting: Beijing +25: Current context, emerging issues and prospects for gender equality and women's rights*. [online]
<https://www.unwomen.org/en/csw/csw64-2020/preparations/expert-group-meeting>.

Università Ca' Foscari di Venezia. 2017. *Linee guida per il linguaggio di genere*. [online]
https://www.unive.it/pag/fileadmin/user_upload/dipartimenti/DSLCC/documenti/sostenibilita/Linee_guida_linguaggio_genere.pdf.

Università degli studi di Ferrara. 2018. *Prontuario dell'Università degli Studi di Ferrara per l'uso del genere nel linguaggio amministrativo e per la redazione di documenti accessibili*. [online]
<http://www.unife.it/it/ateneo/comunicazione/unife-e-il-linguaggio-di-genere/prontuario-linguaggio-di-genere-e-accessibilita.pdf>.

Università degli studi di Milano. 2020. *Linee guida per l'adozione della parità di genere nei testi amministrativi e nella comunicazione istituzionale dell'Università degli Studi di Milano*. [online]
https://www.unimi.it/sites/default/files/regolamenti/Linee%20guida%20linguaggio%20di%20genere_2020_Università%20degli%20Studi%20di%20Milano.pdf.

Università degli studi di Padova. 2017. *Generi e linguaggi. Linee guida per un linguaggio amministrativo e istituzionale attento alle differenze di genere*. [online]
<https://www.unipd.it/sites/unipd.it/files/2017/Generi%20e%20linguaggi.pdf>.

- Università degli studi di Reggio Calabria Mediterranea. 2018. *Indicazioni per un uso del linguaggio rispettoso delle differenze*. [online] http://geecco.unirc.it/images/Linee_guida_uso_linguaggio_Unirc.pdf.
- Università degli studi di Verona. s.d. *Linee guida per il linguaggio di genere*. [online] <https://docs.univr.it/documenti/Documento/allegati/allegati044384.pdf>.
- Università della Calabria. s.d. *Linee guida per l'uso del genere nel linguaggio amministrativo e nella comunicazione istituzionale*. [online] https://www.unical.it/portale/strutture/dipartimenti_240/fisica/parioppourtunita/Linee%20guida%20Linguaggio%20di%20genere_15%20marzo%2021.pdf.
- Università di Bologna. 2020. *Linee guida per la visibilità del genere nella comunicazione istituzionale dell'Università di Bologna*. [online] <https://www.unibo.it/it/ateneo/chi-siamo/linee-guida-per-la-visibilita-del-genere-nella-comunicazione-istituzionale-universita-di-bologna>.
- Università di Siena. 2021. *Genere e linguaggi. Linee guida per un linguaggio amministrativo e istituzionale inclusivo*. [online] https://www.unisi.it/sites/default/files/allegatiparagrafo/LINEE%20GUIDA_Linguaggi%20e%20Generi_0.pdf.
- Università per Stranieri di Siena. 2020. *Linguaggio amministrativo e differenze di genere. Linee guida*. [online] <https://www.unistrasi.it/public/articoli/318/Files/DR%20396.%2020%20Linee%20guida%20linguaggio%20rispettoso%20del%20genere.pdf>.
- Vachon L'Heureux, Pierrette e Luoise Guénette. 2006. *Guide de rédaction épiciène*. Québec: Office québécois de la langue française.
- Vachon L'Heureux, Pierrette. 1992. "Quinze ans de féminisation au Québec : de 1976 à 1991". *Des femmes de la francophonie* 5 (1): 139-142.

- Vachon L'Heureux, Pierrette. 2007. "Au Québec, la rédaction épiscopale devient une proposition officielle". *Nouvelle questions féministe* 26 (3):70-80.
- Van Paridon, Jeroen e Bill Thompson. 2019. "Subs2vec: word embeddings from subtitles in 55 languages". *PsyArXiv preprints*. 14 ottobre. [online] <https://psyarxiv.com/fcrmy/>.
- Van Zoonen, Liesbet. 1994. *Feminist Media Studies*. Londra: Sage.
- Viennot, Éliane. 2014. *Non, le masculin ne l'emporte pas sur le féminin!* Donnemarie-Dontilly: Éditions lxe.
- Vigliocco, Gabriella, Vinson, David P., Paganelli, Federica e Katharina Dworzynski. 2005. "Grammatical gender effects on cognition: implications for language learning and language use". *Journal of experimental psychology: General* 134 (4): 501-520.
- Villani, Paola. 2020. *Il femminile come "genere del disprezzo". Il caso di presidentia: parola d'odio e fake news*. Accademia della Crusca. [online] <https://accademiadellacrusca.it/it/contenuti/il-femminile-come-genere-del-disprezzo-il-caso-di-presidentia-parola-d-odio-e-fake-news/8109>.
- Violi, Patrizia. 1986. *L'infinito singolare: considerazioni sulla differenza sessuale nel linguaggio*. Verona: Essedue.
- Vossen, Piek, Hofmann, Katja, De Rijke, Maarten, Tjong, Erik, Sang Kim e Koen Deschacht. 2007. *The Cornetto database: architecture and user-scenarios*. [online] https://www.researchgate.net/publication/229004970_The_Cornetto_database_Architecture_and_user-scenarios.
- WACC (World Association for Christian Communication). 2005. *Who makes the news? Global media monitoring project 2005*. [online] https://www.mediamonitoringafrica.org/images/uploads/gmmp-report-en-2005_small.pdf.

- WACC (World Association for Christian Communication). 2010. *Who makes the news? Global media monitoring project 2010*. [online] <https://whomakesthenews.org/gmmp-2010-final-reports/>.
- WACC (World Association for Christian Communication). 2015. *Who makes the news? Global media monitoring project 2015*. [online] <https://whomakesthenews.org/gmmp-2015-reports/>.
- WACC (World Association for Christian Communication). 2020. *Who makes the news? Global media monitoring project 2020*. [online] <https://whomakesthenews.org/gmmp-2020-final-reports/>.
- WAN-IFRA. 2020. *Trends in newsrooms 2020. Amplifying women's voices*. [online] https://womeninnews.org/ckfinder/userfiles/files/TIN2020%232_final_low_res.pdf.
- WEF (World Economic Forum). 2015. *Global gender gap report*. [online] <https://reliefweb.int/sites/reliefweb.int/files/resources/cover.pdf>.
- WEF (World Economic Forum). 2020. *Global gender gap report*. [online] https://www3.weforum.org/docs/WEF_GGGR_2020.pdf.
- Wevers, Melvin. 2019. "Using word embeddings to examine gender bias in Dutch newspapers, 1950-1990". *Proceedings of the 1st international workshop on computational approaches to historical language change*. 92–97.
- Winship, Janice. 1987. *Inside Women's Magazines*. Londra: Pandora.
- Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Basil Blackwell.
- Wodak, Ruth, Feistritz, Gert, Moosmüller, Silvia e Ursula Doleschal. 1987. *Sprachliche Gleichbehandlung von Frau und Mann. Linguistische Empfehlungen zur sprachlichen Gleichbehandlung von Frau und Mann im öffentlichen Bereich*. Vienna: Bundesministerium für Soziales.

Zijdeman, Richard, Mandemakers, Kees Muurling, Sanne, Maas, Ineke, van de Putte, bart, Lambert, Paul, van Leeuwen, Marco, van popple, rans e Andrew Miles. 2013. *HSN standardized, HISCO-coded and classified occupational titles*.

Zipf, George K. 1949. *Human behavior and the principle of least-effort*. Cambridge: Addison-Wesley Press.

Zoch, Lynn M. e Judy van Slyke Turk. 1998. "Women making news: gender as variable in source selection and use". *Journalism & mass communication quaterly* 75 (4): 762-775.

Appendice: tabelle integrative (Capitolo 5)

Tab. 1 Forme lessicali significative per la classe 1 (CGD) del *sub-corpus* P_Fonte

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	figlio	176	197	128,68	<0,0001
1	sentire	194	236	104,47	<0,0001
2	ragazzo	125	142	86,17	<0,0001
3	bello	122	138	85,35	<0,0001
4	dire	575	899	85,34	<0,0001
5	succedere	130	160	66,15	<0,0001
6	vedere	255	366	62,86	<0,0001
7	vita	178	242	58,11	<0,0001
8	mamma	58	59	56,82	<0,0001
9	bambino	155	208	53,79	<0,0001
10	casa	150	206	46,28	<0,0001
11	ragazza	45	46	43,37	<0,0001
12	vivere	97	124	41,8	<0,0001
13	dolore	48	52	38,51	<0,0001
14	amore	40	41	38,21	<0,0001
15	film	37	37	38,05	<0,0001
16	fratello	39	40	37,18	<0,0001
17	paura	95	125	35,88	<0,0001
18	proprio	243	377	35,75	<0,0001
19	padre	45	49	35,49	<0,0001
20	ricordo	49	55	34,87	<0,0001
21	morire	63	77	32,65	<0,0001
22	aspettare	78	102	30,24	<0,0001
23	musico	32	33	29,97	<0,0001
24	piacere	57	70	28,94	<0,0001
25	piangere	28	28	28,76	<0,0001
26	moglie	33	35	28,31	<0,0001
27	capire	107	151	28,3	<0,0001
28	cuore	56	69	28,05	<0,0001
29	divertire	27	27	27,73	<0,0001
30	uscire	94	130	72,31	<0,0001
31	lasciare	76	101	75,25	<0,0001
32	storia	84	114	73,68	<0,0001
33	morte	39	44	88,64	<0,0001
34	via	88	121	72,73	<0,0001
35	cercare	97	136	71,32	<0,0001
36	papà	24	24	100	<0,0001
37	bellezza	24	24	100	<0,0001
38	guardare	82	113	72,57	<0,0001
39	me_lo	29	31	93,55	<0,0001
40	sai	26	27	96,3	<0,0001
41	gente	97	139	69,78	<0,0001
42	messaggio	37	43	86,05	<0,0001
43	strada	86	121	71,07	<0,0001
44	libro	24	25	96	<0,0001
45	sorella	21	21	100	<0,0001
46	mangiare	33	38	86,84	<0,0001
47	porre	82	116	70,69	<0,0001
48	dormire	26	28	92,86	<0,0001
49	pensare	228	374	60,96	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
50	emozione	28	31	20,83	<0,0001
51	brutto	23	24	20,75	<0,0001
52	riuscire	104	154	20,61	<0,0001
53	quartiere	22	23	19,73	<0,0001
54	notte	41	51	19,7	<0,0001
55	amico	46	59	19,38	<0,0001
56	ammazzare	24	26	19,2	<0,0001
57	vivo	40	50	18,82	<0,0001
58	spettacolo	21	22	18,7	<0,0001
59	buttare	28	32	18,64	<0,0001
60	contento	30	35	18,53	<0,0001
61	signore	32	38	18,49	<0,0001
62	sera	46	60	17,96	<0,0001
63	entrare	75	108	17,56	<0,0001
64	ambulanza	17	17	17,44	<0,0001
65	meraviglioso	17	17	17,44	<0,0001
66	compagno	17	17	17,44	<0,0001
67	madre	22	24	17,19	<0,0001
68	violenza	50	67	17,18	<0,0001
69	genere	45	59	17,15	<0,0001
70	occhio	32	39	16,7	<0,0001
71	pezzo	28	33	16,64	<0,0001
72	Dio	16	16	16,41	<0,0001
73	nipote	16	16	16,41	<0,0001
74	mente	16	16	16,41	<0,0001
75	aiuto	37	47	16,24	<0,0001
76	incredibile	21	23	16,19	<0,0001
77	capitare	21	23	16,19	<0,0001
78	raccontare	42	55	16,09	<0,0001
79	genitore	42	55	16,09	<0,0001
80	macchina	53	73	15,84	<0,0001
81	tirare	29	35	15,73	<0,0001
82	uccidere	27	32	15,71	<0,0001
83	personaggio	18	19	15,65	<0,0001
84	isola	18	19	15,65	<0,0001
85	passare	49	67	15,2	<0,0001
86	maestro	20	22	15,19	<0,0001
88	pensiero	22	25	14,94	0,00011
89	Battisti	17	18	14,63	0,00013
90	gioia	14	14	14,36	0,00015
91	Marco	14	14	14,36	0,00015
92	cielo	14	14	14,36	0,00015
93	dimenticare	25	30	13,86	0,00019
94	cantare	16	17	13,62	0,00022
95	tantissimo	16	17	13,62	0,00022
96	secondo	98	153	13,36	0,00025
97	calcio	13	13	13,33	0,00026
98	anima	13	13	13,33	0,00026
99	urlare	13	13	13,33	0,00026
100	tranquillo	30	38	13,32	0,00026
101	scusa	18	20	13,21	0,00027
102	treno	28	35	13,15	0,00028

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
103	mattina	43	59	13,09	0,00029
104	felice	26	32	13,02	0,0003
105	sorpresa	20	23	13	0,00031
106	arrivare	164	273	12,83	0,00034
107	domanda	31	40	12,68	0,00036
108	albergo	15	16	12,61	0,00038
109	canzone	12	12	12,3	0,00045
110	Fellini	12	12	12,3	0,00045
111	racconto	12	12	12,3	0,00045
112	ridere	12	12	12,3	0,00045
113	fumo	17	19	12,23	0,00047
114	cattivo	17	19	12,23	0,00047
115	certo	97	153	12,19	0,00047
116	prova	25	31	12,14	0,00049
117	marito	21	25	12	0,00053
118	telefono	21	25	12	0,00053
119	ultimo	70	106	11,87	0,00057
120	scuola	74	113	11,84	0,00058
121	anno	330	587	11,7	0,00062
122	attimo	14	15	11,6	0,00066
123	corpo	31	41	11,31	0,00077
124	Elena	11	11	11,28	0,00078
125	emozionare	11	11	11,28	0,00078
126	tantissime	11	11	11,28	0,00078
127	ergastolo	11	11	11,28	0,00078
128	bimbo	11	11	11,28	0,00078
129	imparare	16	18	11,24	0,00079
130	sparare	16	18	11,24	0,00079
131	sacco	16	18	11,24	0,00079
132	persona	278	490	11,16	0,00083
133	bravo	22	27	11,14	0,00084
134	sorta	18	21	11,1	0,00086
135	sperare	68	104	10,75	0,00104
136	strano	13	14	10,59	0,00113
137	saltare	13	14	10,59	0,00113
138	abituare	13	14	10,59	0,00113
139	bastare	54	80	10,57	0,00114
140	familiare	23	29	10,41	0,00125
141	negozio	15	17	10,27	0,00135
142	Luca	15	17	10,27	0,00135
143	televisione	15	17	10,27	0,00135
144	passione	15	17	10,27	0,00135
145	muro	15	17	10,27	0,00135
146	scontare	15	17	10,27	0,00135
147	gesto	21	26	10,26	0,00136
148	sogno	28	37	10,25	0,00136
149	artista	10	10	10,25	0,00136
150	rifiutare	10	10	10,25	0,00136
151	regista	10	10	10,25	0,00136
152	mo	10	10	10,25	0,00136
153	gridare	10	10	10,25	0,00136
154	vendetta	10	10	10,25	0,00136

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
155	divertente	10	10	10,25	0,00136
156	Natale	10	10	10,25	0,00136
157	diventare	75	117	10,24	0,00137
158	festa	19	23	10,17	0,00143
159	bellissima	19	23	10,17	0,00143
160	denunciare	19	23	10,17	0,00143
161	difficile	54	81	9,74	0,0018
162	donna	84	134	9,6	0,00194
163	festeggiare	12	13	9,58	0,00196
164	tratto	12	13	9,58	0,00196
165	sguardo	12	13	9,58	0,00196
166	arresto	12	13	9,58	0,00196
167	gamba	14	16	9,3	0,00228
168	stazione	14	16	9,3	0,00228
169	abitare	14	16	9,3	0,00228
170	roba	18	22	9,27	0,00233
171	ce_la	18	22	9,27	0,00233
172	silenzio	16	19	9,22	0,00238
173	fatica	16	19	9,22	0,00238
174	binario	16	19	9,22	0,00238
175	Francesco	9	9	9,22	0,00238
176	camion	9	9	9,22	0,00238
177	concerto	9	9	9,22	0,00238
178	botto	9	9	9,22	0,00238
179	tifoso	9	9	9,22	0,00238
180	supermercato	9	9	9,22	0,00238
181	sorriso	9	9	9,22	0,00238
182	incominciare	9	9	9,22	0,00238
183	improvviso	9	9	9,22	0,00238
184	cadavere	9	9	9,22	0,00238
185	minuto	25	33	9,2	0,00242
186	chiamare	70	110	9,02	0,00267
187	amare	28	38	9	0,0027
188	fortuna	23	30	8,95	0,00278
189	arte	26	35	8,7	0,00318
190	Papa	11	12	8,58	0,00339
191	insegnare	11	12	8,58	0,00339
192	caro	11	12	8,58	0,00339
193	stamattina	19	24	8,52	0,0035
194	marcia	13	15	8,34	0,00387
195	caffè	13	15	8,34	0,00387
196	colpa	13	15	8,34	0,00387
197	ansia	13	15	8,34	0,00387
198	accompagnare	13	15	8,34	0,00387
199	spaventare	13	15	8,34	0,00387
200	sentimento	15	18	8,3	0,00395
201	violento	15	18	8,3	0,00395
202	nonno	15	18	8,3	0,00395
203	iniziare	46	69	8,28	0,00399
204	filo	8	8	8,2	0,00419
205	vabbé	8	8	8,2	0,00419
206	meraviglia	8	8	8,2	0,00419

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
207	impressione	8	8	8,2	0,00419
208	felicità	8	8	8,2	0,00419
209	citare	8	8	8,2	0,00419
210	Arturo	8	8	8,2	0,00419
211	conoscere	59	92	8,06	0,00452
212	piccolo	59	92	8,06	0,00452
213	cominciare	52	80	7,85	0,00508
214	festival	10	11	7,59	0,00587
215	perfetto	10	11	7,59	0,00587
216	giornale	10	11	7,59	0,00587
217	significato	10	11	7,59	0,00587
218	interessante	10	11	7,59	0,00587
219	bruciare	10	11	7,59	0,00587
220	Maria	10	11	7,59	0,00587
221	distuggere	16	20	7,5	0,00616
222	pochissimo	12	14	7,39	0,00655
223	cinema	14	17	7,39	0,00654
224	cane	14	17	7,39	0,00654
225	indifferenza	12	14	7,39	0,00655
226	foto	12	14	7,39	0,00655
227	ideale	12	14	7,39	0,00655
228	spalla	21	28	7,36	0,00667
229	uomo	64	102	7,34	0,00673
230	forte	60	95	7,26	0,00704
231	accadere	40	60	7,2	0,0073
232	augurio	7	7	7,17	0,0074
233	coppia	7	7	7,17	0,0074
234	De	7	7	7,17	0,0074
235	odore	7	7	7,17	0,0074
236	danza	7	7	7,17	0,0074
237	pistola	7	7	7,17	0,0074
238	suonare	7	7	7,17	0,0074
239	pianto	7	7	7,17	0,0074
240	pesare	7	7	7,17	0,0074
241	omertà	7	7	7,17	0,0074
242	meccanico	7	7	7,17	0,0074
243	maltese	7	7	7,17	0,0074
244	lamentare	7	7	7,17	0,0074
245	hotel	7	7	7,17	0,0074
246	giubbotto	7	7	7,17	0,0074
247	fretta	7	7	7,17	0,0074
248	fortunato	7	7	7,17	0,0074
249	desiderio	7	7	7,17	0,0074
250	deragliare	7	7	7,17	0,0074
251	centesimo	7	7	7,17	0,0074
252	angoscia	7	7	7,17	0,0074
253	Federico	7	7	7,17	0,0074
254	Daniele	7	7	7,17	0,0074
255	aprire	66	106	7,09	0,00774
256	luce	19	25	7,09	0,00775
257	dispiacere	19	25	7,09	0,00775
258	giorno	153	267	6,87	0,00877

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
259	ospite	9	10	6,6	0,01021
260	vetro	9	10	6,6	0,01021
261	sciare	9	10	6,6	0,01021
262	lingua	9	10	6,6	0,01021
263	fantastico	9	10	6,6	0,01021
264	bagno	9	10	6,6	0,01021
265	adulto	9	10	6,6	0,01021
266	abbattere	9	10	6,6	0,01021
267	poliziotto	9	10	6,6	0,01021
268	montagna	9	10	6,6	0,01021
269	incontrare	20	27	6,59	0,01025
270	bomba	13	16	6,5	0,01078
271	volto	13	16	6,5	0,01078
272	veloce	13	16	6,5	0,01078
273	collega	26	37	6,47	0,01099
274	rubare	11	13	6,45	0,01106
275	ricominciare	11	13	6,45	0,01106
276	ferita	11	13	6,45	0,01106
277	professore	11	13	6,45	0,01106
278	parecchio	11	13	6,45	0,01106
279	mostrare	11	13	6,45	0,01106
280	provare	37	56	6,26	0,01236
281	colpire	27	39	6,16	0,0131
282	Angela	6	6	6,15	0,01316
283	infanzia	6	6	6,15	0,01316
284	tremendo	6	6	6,15	0,01316
285	finestrino	6	6	6,15	0,01316
286	Andrea	6	6	6,15	0,01316
287	stanco	6	6	6,15	0,01316
288	scoop	6	6	6,15	0,01316
289	lettura	6	6	6,15	0,01316
290	lager	6	6	6,15	0,01316
291	interpretare	6	6	6,15	0,01316
292	coltello	6	6	6,15	0,01316
293	Sergio	6	6	6,15	0,01316
294	Celentano	6	6	6,15	0,01316
295	stupendo	6	6	6,15	0,01316
296	splendido	6	6	6,15	0,01316
297	scoperta	6	6	6,15	0,01316
298	rumore	6	6	6,15	0,01316
299	pugno	6	6	6,15	0,01316
300	petardo	6	6	6,15	0,01316
301	pago	6	6	6,15	0,01316
302	nonno	6	6	6,15	0,01316
303	fortissimo	6	6	6,15	0,01316
304	dolcezza	6	6	6,15	0,01316
305	casino	6	6	6,15	0,01316
306	boato	6	6	6,15	0,01316
307	bicchiere	6	6	6,15	0,01316
308	bastone	6	6	6,15	0,01316
309	animo	6	6	6,15	0,01316
310	altrui	6	6	6,15	0,01316

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
311	realità	50	79	6,14	0,01318
312	scappare	24	34	6,12	0,01334
313	voce	24	34	6,12	0,01334
314	trovare	138	241	6,11	0,01341
315	colpo	16	21	6,04	0,01399
316	tornare	72	119	5,94	0,0148
317	stanza	19	26	5,84	0,01564
318	alzare	25	36	5,8	0,01599
319	stare	382	711	5,79	0,01616
320	chiudere	53	85	5,75	0,01647
321	terra	44	69	5,73	0,01663
322	parlare	123	214	5,71	0,01685
323	testa	36	55	5,7	0,01698
324	vendere	12	15	5,62	0,01772
325	gravissimo	12	15	5,62	0,01772
326	collezione	12	15	5,62	0,01772
327	assassino	12	15	5,62	0,01772
328	albero	12	15	5,62	0,01772
329	Luigi	12	15	5,62	0,01772
330	scatola	8	9	5,62	0,01779
331	tardo	8	9	5,62	0,01779
332	bar	8	9	5,62	0,01779
333	riparare	8	9	5,62	0,01779
334	insulto	8	9	5,62	0,01779
335	accendere	8	9	5,62	0,01779
336	puro	8	9	5,62	0,01779
337	preghiera	8	9	5,62	0,01779
338	pochino	8	9	5,62	0,01779
339	matrimonio	8	9	5,62	0,01779
340	fiore	8	9	5,62	0,01779
341	famoso	8	9	5,62	0,01779
342	decennio	8	9	5,62	0,01779
343	cammino	8	9	5,62	0,01779
344	Brasile	8	9	5,62	0,01779
345	povero	17	23	5,54	0,01861
346	sopravvivere	10	12	5,53	0,01868
347	sala sale	10	12	5,53	0,01868
348	fiamma	10	12	5,53	0,01868
349	ebreo	10	12	5,53	0,01868
350	video	10	12	5,53	0,01868
351	fame	10	12	5,52	0,01868
352	duro	26	38	5,45	0,01883
353	giro	46	73	5,45	0,01958
354	passato	46	73	5,45	0,01958
355	ritornare	23	33	5,45	0,01954
356	ricevere	23	33	5,26	0,01954
357	memoria	27	40	5,25	0,02183
358	dedicare	15	20	5,25	0,02194
359	epoca	15	20	5,25	0,02194
360	riprendere	15	20	5,18	0,02194
361	rimanere	96	165	5,12	0,02289
362	edizione	5	5	5,12	0,02363

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
363	Sanremo	5	5	5,12	0,02363
364	dottore	5	5	5,12	0,02363
365	mille	5	5	5,12	0,02363
366	curiosità	5	5	5,12	0,02363
367	Facebook	5	5	5,12	0,02363
368	trasferta	5	5	5,12	0,02363
369	rock	5	5	5,12	0,02363
370	perdonare	5	5	5,12	0,02363
371	pentirsi	5	5	5,12	0,02363
372	offendere	5	5	5,12	0,02363
373	nube	5	5	5,12	0,02363
374	inferno	5	5	5,12	0,02363
375	incinta	5	5	5,12	0,02363
376	guerriero	5	5	5,12	0,02363
377	donare	5	5	5,12	0,02363
378	battezzare	5	5	5,12	0,02363
379	ballare	5	5	5,12	0,02363
380	allegro	5	5	5,12	0,02363
381	Claudio	5	5	4,99	0,02363
382	chiesa	13	17	4,99	0,02551
383	galera	13	17	4,8	0,02551
384	gioco	22	32	4,8	0,02838
385	piede	22	32	4,8	0,02838
386	svegliare	16	22	4,77	0,02852
387	eccetera	11	14	4,77	0,02897
388	regalare	11	14	4,77	0,02897
389	valanga	11	14	4,77	0,02897
390	parente	11	14	4,65	0,02897
391	stile	7	8	4,65	0,03111
392	offesa offeso	7	8	4,65	0,03111
393	dov	7	8	4,65	0,03111
394	spegnere	7	8	4,65	0,03111
395	rifugio	7	8	4,65	0,03111
396	portafoglio	7	8	4,65	0,03111
397	educare	7	8	4,65	0,03111
398	discarica	7	8	4,65	0,03111
399	telefonare	7	8	4,65	0,03111
400	sasso	7	8	5,53	0,03111
401	ragazzini	7	8	4,65	0,03111
402	musicale	7	8	4,65	0,03111
403	han	7	8	4,65	0,03111
404	corridoio	7	8	4,65	0,03111
405	chitarra	7	8	4,65	0,03111
406	arrabbiare	7	8	4,65	0,03111
407	abbraccio	7	8	4,65	0,03111
408	cucina	9	11	4,63	0,03147
409	lontano	9	11	4,63	0,03147
410	vietare	9	11	4,63	0,03147
411	legno	9	11	4,63	0,03147
412	ricostruzione	9	11	4,63	0,03147
413	frase	9	11	4,63	0,03147
414	Parigi	9	11	4,63	0,03147

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
415	conto	23	34	4,54	0,03311
416	parola	49	80	4,53	0,03337
417	nascere	44	71	4,52	0,03355
418	fermo	14	19	4,49	0,03414
419	abbracciare	14	19	4,49	0,03414
420	correre	17	24	4,42	0,03557
421	arrivo	17	24	4,42	0,03557
422	ricordare	41	66	4,3	0,03817
423	odio	24	36	4,3	0,038
424	immaginare	24	36	4,3	0,038
425	sofferenza	12	16	4,2	0,04047
426	generazione	18	26	4,1	0,04295
427	Segre	4	4	4,1	0,04296
428	Liliana	4	4	4,1	0,04296
429	pecora	4	4	4,1	0,04296
430	monumento	4	4	4,1	0,04296
431	esperienza	25	38	4,09	0,04301
432	viaggio	15	21	4,08	0,04335
433	rientrare	15	21	4,08	0,04335
434	rabbia	15	21	4,08	0,04335
435	orgoglioso	15	21	4,08	0,04335
436	vittima	44	72	3,97	0,04618
437	stasera	10	13	3,94	0,0471
438	teatro	10	13	3,94	0,0471
439	cellulare	10	13	3,94	0,0471
440	negare	10	13	3,94	0,0471
441	moda	10	13	3,94	0,0471
442	omicidio	10	13	3,94	0,0471
443	tedesco	10	13	3,94	0,0471
444	scena	10	13	3,94	0,0471
445	tenere	56	94	3,93	0,04754

Tab. 2 Forme lessicali significative per la classe 2 (CGD) del *sub-corpus* P_Fonte

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	virus	34	39	186,64	<0,0001
1	zona	41	54	186,09	<0,0001
2	paziente	27	30	154,56	<0,0001
3	evento	25	27	148,61	<0,0001
4	medico	28	34	142,08	<0,0001
5	cinese	28	37	126,23	<0,0001
6	temperatura	21	23	122,45	<0,0001
7	sintomo	18	18	118,3	<0,0001
8	Cina	25	33	112,83	<0,0001
9	influenza	19	21	109,42	<0,0001
10	infezione	20	23	109,2	<0,0001
11	malattia	23	31	100,73	<0,0001
12	area	22	29	99,43	<0,0001
13	coronavirus	18	21	96,35	<0,0001
14	rischio	43	90	94,59	<0,0001
15	epidemia	16	18	89,94	<0,0001
16	inverno	16	18	89,94	<0,0001
17	animale	21	30	84,52	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
18	freddo	24	39	79,65	<0,0001
19	respiratorio	12	12	78,81	<0,0001
20	saldo	12	13	70,92	<0,0001
21	produrre	29	57	70,85	<0,0001
22	giacca	13	15	70,59	<0,0001
23	effetto	22	37	69,18	<0,0001
24	identificare	15	20	66,63	<0,0001
25	Wuhan	10	10	65,66	<0,0001
26	sci	12	14	64,18	<0,0001
27	prodotto	16	23	63,75	<0,0001
28	utilizzare	29	61	63,01	<0,0001
29	computer	9	9	59,09	<0,0001
30	sostanza	12	15	58,35	<0,0001
31	trasmissione	12	15	58,35	<0,0001
32	volo	15	22	58,02	<0,0001
33	antibiotico	10	11	57,89	<0,0001
34	ghiaccio	10	11	57,89	<0,0001
35	febbre	11	13	57,79	<0,0001
36	terapia	11	13	57,79	<0,0001
37	uso	14	20	56,28	<0,0001
38	fornire	12	16	53,28	<0,0001
39	focolaio	8	8	52,51	<0,0001
40	faglia	8	8	52,51	<0,0001
41	malato	11	14	52,15	<0,0001
42	spostare	11	14	52,15	<0,0001
43	veicolo	10	12	51,44	<0,0001
44	laboratorio	10	12	51,44	<0,0001
45	mascherina	10	12	51,44	<0,0001
46	riscaldamento	9	10	51,4	<0,0001
47	provocare	9	10	51,4	<0,0001
48	atmosfera	9	10	51,4	<0,0001
49	anziano	15	25	47,77	<0,0001
50	classico	11	15	47,28	<0,0001
51	climatico	11	15	47,28	<0,0001
52	caldo	13	20	46,8	<0,0001
53	verdura	7	7	45,94	<0,0001
54	allerta	7	7	45,94	<0,0001
55	York	7	7	45,94	<0,0001
56	New	7	7	45,94	<0,0001
57	tosse	7	7	45,94	<0,0001
58	dispositivo	7	7	45,94	<0,0001
59	abbigliamento	7	7	45,94	<0,0001
60	elettrico	9	11	45,13	<0,0001
61	polvere	8	9	44,92	<0,0001
62	suolo	8	9	44,92	<0,0001
63	farmaco	8	9	44,92	<0,0001
64	chilometro	13	21	43,45	<0,0001
65	causa	18	36	42,58	<0,0001
66	grosso	18	36	42,58	<0,0001
67	sviluppare	15	27	42,27	<0,0001
68	sconto	10	14	41,37	<0,0001
69	contenere	10	14	41,37	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
70	materiale	14	25	39,95	<0,0001
71	effettuare	9	12	39,94	<0,0001
72	azoto	6	6	39,38	<0,0001
73	sottile	6	6	39,38	<0,0001
74	genetico	6	6	39,38	<0,0001
75	attrezzatura	6	6	39,38	<0,0001
76	riposo	6	6	39,38	<0,0001
77	prognosi	6	6	39,38	<0,0001
78	nevicata	6	6	39,38	<0,0001
79	influenzale	6	6	39,38	<0,0001
80	fortunatamente	6	6	39,38	<0,0001
81	cronico	6	6	39,38	<0,0001
82	circoscrivere	6	6	39,38	<0,0001
83	operatore	11	17	39,32	<0,0001
84	naturale	11	17	39,32	<0,0001
85	eventuale	11	17	39,32	<0,0001
86	clinico	8	10	38,88	<0,0001
88	sportivo	7	8	38,47	<0,0001
89	sciatore	7	8	38,47	<0,0001
90	abito	7	8	38,47	<0,0001
91	comparire	7	8	38,47	<0,0001
92	Spallanzani	7	8	38,47	<0,0001
93	DNA	7	8	38,47	<0,0001
94	ricerca	17	35	38,24	<0,0001
95	fonte	12	20	38,2	<0,0001
96	forma	16	32	37,83	<0,0001
97	metro	15	29	37,56	<0,0001
98	diffondere	11	18	36,02	<0,0001
99	nord	11	18	36,02	<0,0001
100	falso	9	13	35,57	<0,0001
101	indicare	9	13	35,57	<0,0001
102	vento	8	11	33,96	<0,0001
103	curare	10	16	33,9	<0,0001
104	tecnica tecnico	10	16	33,9	<0,0001
105	plastica	11	19	33,09	<0,0001
106	stagione	13	25	32,83	<0,0001
107	termico	5	5	32,81	<0,0001
108	interazione	5	5	32,81	<0,0001
109	riciclare	7	9	32,7	<0,0001
110	superficie	7	9	32,7	<0,0001
111	eccessivo	7	9	32,7	<0,0001
112	importazione	6	7	32,06	<0,0001
113	impiegare	6	7	32,06	<0,0001
114	prontamente	6	7	32,06	<0,0001
115	profondità	6	7	32,06	<0,0001
116	monitorare	6	7	32,06	<0,0001
117	documentare	6	7	32,06	<0,0001
118	acqua	21	54	31,17	<0,0001
119	elevato	10	17	30,84	<0,0001
120	trasmettere	13	26	30,71	<0,0001
121	grado	22	59	29,96	<0,0001
122	trasferire	8	12	29,89	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
123	preoccupare	18	44	29,52	<0,0001
124	sequestro	9	15	28,63	<0,0001
125	pista	15	34	28,37	<0,0001
126	distanza	12	24	28,34	<0,0001
127	procedura	7	10	28,1	<0,0001
128	chilo	7	10	28,1	<0,0001
129	fase	17	42	27,29	<0,0001
130	profilo	6	8	26,61	<0,0001
131	precauzione	6	8	26,61	<0,0001
132	mobilità	6	8	26,61	<0,0001
133	pregiudizio	6	8	26,61	<0,0001
134	considerazione	6	8	26,61	<0,0001
135	roccia	4	4	26,24	<0,0001
136	velluto	4	4	26,24	<0,0001
137	fibra	4	4	26,24	<0,0001
138	terremoto	11	22	25,98	<0,0001
139	avvenire	18	47	25,87	<0,0001
140	scientifico	9	16	25,84	<0,0001
141	tessuto	10	19	25,75	<0,0001
142	nebbia	5	6	25,7	<0,0001
143	cavo	5	6	25,7	<0,0001
144	lago	5	6	25,7	<0,0001
145	ricorrere	5	6	25,7	<0,0001
146	pulizia	5	6	25,7	<0,0001
147	morbido	5	6	25,7	<0,0001
148	lentamente	5	6	25,7	<0,0001
149	guarire	5	6	25,7	<0,0001
150	formaggio	5	6	25,7	<0,0001
151	circolazione	5	6	25,7	<0,0001
152	chimico	5	6	25,7	<0,0001
153	cappotto	5	6	25,7	<0,0001
154	bottiglia	5	6	25,7	<0,0001
155	alcol	5	6	25,7	<0,0001
156	accontentare	5	6	25,7	<0,0001
157	B	5	6	25,7	<0,0001
158	turista	7	11	24,38	<0,0001
159	istituto	7	11	24,38	<0,0001
160	limite	7	11	24,38	<0,0001
161	circolare	7	11	24,38	<0,0001
162	medio	15	37	24,14	<0,0001
163	periodo	15	37	24,14	<0,0001
164	natura	11	23	24,04	<0,0001
165	esistere	23	70	23,69	<0,0001
166	mezza mezzo	8	14	23,55	<0,0001
167	breve	8	14	23,55	<0,0001
168	caratteristica	8	14	23,55	<0,0001
169	motore	9	17	23,4	<0,0001
170	tecnologia	9	17	23,4	<0,0001
171	esempio	28	93	23,33	<0,0001
172	informazione	15	38	22,9	<0,0001
173	fede	6	9	22,41	<0,0001
174	basso	14	35	21,94	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
175	scorso	14	35	21,94	<0,0001
176	numero	24	77	21,78	<0,0001
177	impatto	10	21	21,69	<0,0001
178	visitare	7	12	21,29	<0,0001
179	formare	7	12	21,29	<0,0001
180	normalmente	7	12	21,29	<0,0001
181	inserire	7	12	21,29	<0,0001
182	utilizzo	9	18	21,24	<0,0001
183	escludere	9	18	21,24	<0,0001
184	sicuramente	35	131	21,09	<0,0001
185	tumore	8	15	21,05	<0,0001
186	accesso	8	15	21,05	<0,0001
187	informare	8	15	21,05	<0,0001
188	ritorno	11	25	20,68	<0,0001
189	proteina	5	7	20,67	<0,0001
190	rapidamente	5	7	20,67	<0,0001
191	organico	5	7	20,67	<0,0001
192	edificio	5	7	20,67	<0,0001
193	televisivo	5	7	20,67	<0,0001
194	sismico	5	7	20,67	<0,0001
195	rallentare	5	7	20,67	<0,0001
196	medicina	5	7	20,67	<0,0001
197	lavare	5	7	20,67	<0,0001
198	accumulare	5	7	20,67	<0,0001
199	situazione	33	123	20,11	<0,0001
200	robot	4	5	19,43	<0,0001
201	gelato	4	5	19,43	<0,0001
202	ossigeno	4	5	19,43	<0,0001
203	luna	4	5	19,43	<0,0001
204	trattare	19	58	19,4	<0,0001
205	apertura	6	10	19,08	<0,0001
206	volontario	6	10	19,08	<0,0001
207	ingresso	6	10	19,08	<0,0001
208	esporre	6	10	19,08	<0,0001
209	adeguato	6	10	19,08	<0,0001
210	incendio	8	16	18,88	<0,0001
211	eccezionale	7	13	18,71	<0,0001
212	riservare	7	13	18,71	<0,0001
213	globale	7	13	18,71	<0,0001
214	serie	15	42	18,59	<0,0001
215	corso	13	34	18,59	<0,0001
216	aria	14	38	18,54	<0,0001
217	particolare	23	78	18,12	<0,0001
218	pesante	9	20	17,62	<0,0001
219	neve	14	39	17,54	<0,0001
220	energia	13	35	17,5	<0,0001
221	capo	13	35	17,5	<0,0001
222	fenomeno	13	35	17,5	<0,0001
223	campione	8	17	16,98	<0,0001
224	stupefacente	5	8	16,93	<0,0001
225	pianeta	5	8	16,93	<0,0001
226	ambito	5	8	16,93	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
227	test	5	8	16,93	<0,0001
228	psicosi	5	8	16,93	<0,0001
229	notare	5	8	16,93	<0,0001
230	fragile	5	8	16,93	<0,0001
231	velocità	7	14	16,51	<0,0001
232	soggetto	12	32	16,48	<0,0001
233	aeroporto	6	11	16,38	<0,0001
234	specie	9	21	16,1	<0,0001
235	reale	11	29	15,46	<0,0001
236	estremo	8	18	15,31	<0,0001
237	attualmente	8	18	15,31	<0,0001
238	partire	17	55	15,07	0,0001
239	emettere	4	6	14,93	0,00011
240	contemporaneo	4	6	14,93	0,00011
241	schiena	4	6	14,93	0,00011
242	notevole	4	6	14,93	0,00011
243	fondamentalmente	4	6	14,93	0,00011
244	estendere	4	6	14,93	0,00011
245	confine	4	6	14,93	0,00011
246	comando	4	6	14,93	0,00011
247	ciclo	4	6	14,93	0,00011
248	agricolo	4	6	14,93	0,00011
249	accertamento	4	6	14,93	0,00011
250	unità	7	15	14,63	0,00013
251	comunicazione	7	15	14,63	0,00013
252	vaccino	11	30	14,4	0,00014
253	cura	11	30	14,4	0,00014
254	elicottero	6	12	14,15	0,00016
255	scarpa	6	12	14,15	0,00016
256	quantità	6	12	14,15	0,00016
257	inquinamento	6	12	14,15	0,00016
258	estate	6	12	14,15	0,00016
259	diffusione	6	12	14,15	0,00016
260	costante	6	12	14,15	0,00016
261	sciogliere	5	9	14,06	0,00017
262	condurre	5	9	14,06	0,00017
263	pianta	5	9	14,06	0,00017
264	turismo	5	9	14,06	0,00017
265	intenso	5	9	14,06	0,00017
266	antico	5	9	14,06	0,00017
267	ottenere	12	35	13,57	0,00022
268	accorgersi	7	16	13,01	0,00031
269	verificare	13	40	13	0,00031
270	intervenire	14	45	12,6	0,00038
271	disastro	8	20	12,51	0,0004
272	Milano	9	24	12,35	0,00044
273	funzione	9	24	12,35	0,00044
274	zero	6	13	12,29	0,00045
275	principalmente	6	13	12,29	0,00045
276	origine	6	13	12,29	0,00045
277	occupare	14	46	11,92	0,00055
278	eroe	5	10	11,79	0,00059

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
279	costruzione	5	10	11,79	0,00059
280	verde	5	10	11,79	0,00059
281	online	5	10	11,79	0,00059
282	inquinare	5	10	11,79	0,00059
283	completo	5	10	11,79	0,00059
284	cervello	5	10	11,79	0,00059
285	caduta	5	10	11,79	0,00059
286	esatto	4	7	11,77	0,0006
287	vino	4	7	11,77	0,0006
288	sperimentare	4	7	11,77	0,0006
289	indossare	4	7	11,77	0,0006
290	bicicletta	4	7	11,77	0,0006
291	sospetto	4	7	11,77	0,0006
292	primavera	4	7	11,77	0,0006
293	inferiore	4	7	11,77	0,0006
294	compatto	4	7	11,77	0,0006
295	luogo	15	51	11,7	0,00062
296	cambiamento	11	34	10,87	0,00097
297	preciso	6	14	10,72	0,00105
298	individuare	6	14	10,72	0,00105
299	presente	14	48	10,68	0,00108
300	disponibile	10	30	10,6	0,00113
301	effettivamente	10	30	10,6	0,00113
302	leggero	9	26	10,39	0,00126
303	pulire	7	18	10,34	0,00129
304	specifico	7	18	10,34	0,00129
305	preoccupazione	7	18	10,34	0,00129
306	sanitario	11	35	10,14	0,00145
307	momento	45	221	10,06	0,00151
308	attività	17	64	9,98	0,00157
309	risparmiare	5	11	9,96	0,00159
310	mostra	5	11	9,96	0,00159
311	eventualmente	5	11	9,96	0,00159
312	deficit	5	11	9,96	0,00159
313	interno	28	123	9,88	0,00166
314	scoprire	9	27	9,53	0,00201
315	direttamente	9	27	9,53	0,00201
316	filtro	3	5	9,53	0,00201
317	residente	3	5	9,53	0,00201
318	polo	3	5	9,53	0,00201
319	parallelo	3	5	9,53	0,00201
320	creatività	3	5	9,53	0,00201
321	mano	16	60	9,5	0,00205
322	studiare	11	36	9,45	0,0021
323	connettere	4	8	9,43	0,00213
324	visitatore	4	8	9,43	0,00213
325	verifica	4	8	9,43	0,00213
326	scienziato	4	8	9,43	0,00213
327	ripristinare	4	8	9,43	0,00213
328	causare	4	8	9,43	0,00213
329	altezza	4	8	9,43	0,00213
330	immediatamente	6	15	9,38	0,00219

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
331	sottoporre	6	15	9,38	0,00219
332	sostenibile	6	15	9,38	0,00219
333	sito	6	15	9,38	0,00219
334	passaggio	6	15	9,38	0,00219
335	necessità	6	15	9,38	0,00219
336	fisico	6	15	9,38	0,00219
337	impianto	7	19	9,25	0,00235
338	città	26	115	8,93	0,0028
339	livello	21	88	8,76	0,00308
340	maniera	18	72	8,76	0,00307
341	alto	19	78	8,49	0,00356
342	colore	5	12	8,46	0,00362
343	ospedale	10	33	8,41	0,00372
344	principale	7	20	8,28	0,00401
345	camminare	6	16	8,22	0,00413
346	ufficio	6	16	8,22	0,00413
347	paio	6	16	8,22	0,00413
348	mare	13	48	8,07	0,0045
349	gennaio	8	25	7,69	0,00554
350	importantissimo	4	9	7,65	0,00569
351	bocca	4	9	7,65	0,00569
352	percentuale	4	9	7,65	0,00569
353	quota	4	9	7,65	0,00569
354	autorità	7	21	7,41	0,00648
355	cliente	5	13	7,22	0,00721
356	arco	5	13	7,22	0,00721
357	sufficiente	5	13	7,22	0,00721
358	pericoloso	10	35	7,2	0,00727
359	normale	12	45	7,11	0,00764
360	mani mano	12	45	7,11	0,00764
361	maglia	3	6	7,07	0,00783
362	emissione	3	6	7,07	0,00783
363	solare	3	6	7,07	0,00783
364	risvegliare	3	6	7,07	0,00783
365	ristorante	3	6	7,07	0,00783
366	risentire	3	6	7,07	0,00783
367	resistenza	3	6	7,07	0,00783
368	rame ramo	3	6	7,07	0,00783
369	essenzialmente	3	6	7,07	0,00783
370	elementare	3	6	7,07	0,00783
371	drammaticamente	3	6	7,07	0,00783
372	diversa	3	6	7,07	0,00783
373	aggressivo	3	6	7,07	0,00783
374	settimana	21	94	6,87	0,00875
375	contatto	9	31	6,77	0,00928
376	pronto	14	57	6,42	0,01129
377	tipo	20	90	6,41	0,01136
378	cibo	6	18	6,35	0,01174
379	sud	6	18	6,35	0,01174
380	ottimo	6	18	6,35	0,01174
381	aereo	8	27	6,34	0,01177
382	conseguenza	8	27	6,34	0,01177

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
383	ufficiale	4	10	6,25	0,01242
384	sicuramente_la	4	10	6,25	0,01242
385	evoluzione	4	10	6,25	0,01242
386	drammatico	4	10	6,25	0,01242
387	altissimo	4	10	6,25	0,01242
388	incidente	9	32	6,21	0,01272
389	indicazione	5	14	6,17	0,01295
390	analisi	5	14	6,17	0,01295
391	proteggere	8	28	5,76	0,0164
392	particolarmente	8	28	5,76	0,0164
393	mesa mese	11	43	5,74	0,01655
394	massimo	10	38	5,69	0,01703
395	controllare	9	33	5,69	0,01706
396	stima	3	7	5,36	0,02065
397	aggressore	3	7	5,36	0,02065
398	pratico	3	7	5,36	0,02065
399	frequentare	3	7	5,36	0,02065
400	trascurare	3	7	5,36	0,02065
401	record	3	7	5,36	0,02065
402	preso	3	7	5,36	0,02065
403	massa	3	7	5,36	0,02065
404	indico	3	7	5,36	0,02065
405	immediato	3	7	5,36	0,02065
406	giovanissimi	3	7	5,36	0,02065
407	divieto	3	7	5,36	0,02065
408	superare	7	24	5,32	0,02103
409	sport	5	15	5,29	0,02145
410	totale	8	29	5,22	0,02228
411	inizio	9	34	5,21	0,02242
412	social	4	11	5,13	0,02346
413	panico	4	11	5,13	0,02346
414	matematico	4	11	5,13	0,02346
415	stadio	4	11	5,13	0,02346
416	meccanismo	4	11	5,13	0,02346
417	generare	4	11	5,13	0,02346
418	acquistare	4	11	5,13	0,02346
419	coprire	6	20	4,91	0,02669
420	macchina	16	73	4,84	0,02779
421	possibilità	17	79	4,77	0,02892
422	rete	7	25	4,76	0,02905
423	interrogare	2	4	4,71	0,02994
424	lungo	13	57	4,58	0,03226
425	consentire	13	57	4,58	0,03226
426	collegare	5	16	4,53	0,03323
427	ambiente	10	41	4,47	0,03451
428	procura	6	21	4,32	0,03776
429	diminuire	6	21	4,32	0,03776
430	vecchio	8	31	4,29	0,03842
431	studio	8	31	4,29	0,03842
432	recuperare	7	26	4,26	0,03904
433	pizza	4	12	4,23	0,03971
434	perdita	4	12	4,23	0,03971

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
435	denuncia	4	12	4,23	0,03971
436	ecologico	4	12	4,23	0,03971
437	avanzare	4	12	4,23	0,03971
438	slittare	3	8	4,11	0,04267
439	scorta	3	8	4,11	0,04267
440	desiderare	3	8	4,11	0,04267
441	cravatta	3	8	4,11	0,04267
442	abitante	3	8	4,11	0,04267
443	attendere	5	17	3,88	0,04872
444	terreno	5	17	3,88	0,04872
445	episodio	5	17	3,88	0,04872

Tab. 3 Forme lessicali significative per la classe 3 (CGD) del *sub-corpus* P_Fonte

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	governo	388	403	628,5	<0,0001
1	Italia	406	510	414,62	<0,0001
2	politico	303	338	411,74	<0,0001
3	paese	343	419	373,58	<0,0001
4	italiano	301	397	263,63	<0,0001
5	5_Stelle	149	150	250,78	<0,0001
6	Salvini	147	155	223,05	<0,0001
7	cittadino	180	209	217,93	<0,0001
8	Europa	160	181	206,03	<0,0001
9	tassa	117	117	199,03	<0,0001
10	lavoro	254	356	183,7	<0,0001
11	Movimento	105	105	178,35	<0,0001
12	elettorale	100	100	169,76	<0,0001
13	centro_destra	95	95	161,17	<0,0001
14	legge	110	118	159,65	<0,0001
15	europeo	109	117	157,95	<0,0001
16	presidente	105	115	145,05	<0,0001
17	sicurezza	138	169	144,7	<0,0001
18	forza	141	176	140,5	<0,0001
19	ministro	96	105	132,89	<0,0001
20	elezione	78	78	132,06	<0,0001
21	impresa	90	97	128,87	<0,0001
22	economico	84	89	125,03	<0,0001
23	maggioranza	75	76	123,29	<0,0001
24	votare	72	74	114,65	<0,0001
25	Parlamento	71	73	112,95	<0,0001
26	partito	84	95	107,1	<0,0001
27	Lega	63	63	106,47	<0,0001
28	programma	71	77	99,94	<0,0001
29	riforma	61	62	99,44	<0,0001
30	proposta	71	78	96,93	<0,0001
31	investimento	67	73	93,23	<0,0001
32	Partito_Democratico	54	54	91,16	<0,0001
33	sviluppo	56	57	90,94	<0,0001
34	fiscale	53	53	89,46	<0,0001
35	PD	55	56	89,24	<0,0001
36	reddito	52	52	87,76	<0,0001
37	Forza_Italia	52	53	84,15	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
38	economia	56	59	84,14	<0,0001
39	immigrazione	53	55	82,39	<0,0001
40	sinistra	62	69	81,96	<0,0001
41	lavoratore	67	77	81,91	<0,0001
42	accordo	55	59	79,26	<0,0001
43	diritto	91	119	78,94	<0,0001
44	voto	54	58	77,58	<0,0001
45	decreto	52	55	77,39	<0,0001
46	campagna	55	60	76,19	<0,0001
47	parlamentare	45	45	75,88	<0,0001
48	governare	47	48	75,67	<0,0001
49	crescita	61	70	74,79	<0,0001
50	soluzione	48	51	70,66	<0,0001
51	cittadinanza	46	48	70,56	<0,0001
52	prossimo	86	115	69,81	<0,0001
53	Unione_Europea	41	41	69,1	<0,0001
54	garantire	50	55	67,86	<0,0001
55	regione	51	57	66,63	<0,0001
56	tema	59	70	66,45	<0,0001
57	sociale	69	88	64,07	<0,0001
58	processo	71	92	63,04	<0,0001
59	Emilia	41	43	62,13	<0,0001
60	popolo	53	62	61,84	<0,0001
61	problema	126	197	61,04	<0,0001
62	impegno	50	58	59,57	<0,0001
63	Legge	43	47	59,2	<0,0001
64	milione	58	72	57,93	<0,0001
65	decidere	75	102	57,75	<0,0001
66	crisi	40	43	57,24	<0,0001
67	concreto	40	43	57,24	<0,0001
68	democrazia	38	40	57,08	<0,0001
69	Berlusconi	38	40	57,08	<0,0001
70	Romagna	36	37	57,07	<0,0001
71	Di_Maio	36	37	57,07	<0,0001
72	difendere	50	60	54,6	<0,0001
73	risultato	51	62	53,83	<0,0001
74	coalizione	34	35	53,69	<0,0001
75	provvedimento	34	35	53,69	<0,0001
76	discussione	34	35	53,69	<0,0001
77	ridurre	46	54	53,14	<0,0001
78	chiedere	142	237	53,07	<0,0001
79	vincere	56	71	52,76	<0,0001
80	internazionale	60	78	52,69	<0,0001
81	convincere	44	51	52,48	<0,0001
82	regola	47	56	52,26	<0,0001
83	responsabilità	61	80	52,23	<0,0001
84	sindaco	41	47	50,26	<0,0001
85	misura	44	52	49,95	<0,0001
86	dimostrare	47	57	49,89	<0,0001
88	chiaro	65	89	48,94	<0,0001
89	questione	40	46	48,64	<0,0001
90	manovra	31	32	48,63	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
91	credere	147	253	48,01	<0,0001
92	nazionale	47	58	47,62	<0,0001
93	candidato	34	37	47,24	<0,0001
94	discutere	28	28	47,12	<0,0001
95	Fratelli_d	28	28	47,12	<0,0001
96	Calabria	28	28	47,12	<0,0001
97	fondamentale	39	45	47,02	<0,0001
98	prescrizione	32	34	47,01	<0,0001
99	euro	63	87	46,18	<0,0001
100	oggi	173	312	45,36	<0,0001
101	miliardo	29	30	45,27	<0,0001
102	Conte	29	30	45,27	<0,0001
103	rispettare	45	56	44,56	<0,0001
104	pagare	71	103	44,44	<0,0001
105	risorsa	32	35	43,92	<0,0001
106	alleanza	26	26	43,74	<0,0001
107	concessione	28	29	43,58	<0,0001
108	pensione	35	41	40,59	<0,0001
109	appello	38	46	40,48	<0,0001
110	banca	28	30	40,32	<0,0001
111	dovere	414	878	40,29	<0,0001
112	presentare	42	53	40	<0,0001
113	sistema	64	93	39,79	<0,0001
114	continuare	91	146	39,61	<0,0001
115	cambiare	83	130	39,61	<0,0001
116	migrare	31	35	39,4	<0,0001
117	approvare	23	23	38,68	<0,0001
118	infrastruttura	23	23	38,68	<0,0001
119	Renzi	27	29	38,65	<0,0001
120	immigrato	27	29	38,65	<0,0001
121	assumere	36	44	37,37	<0,0001
122	contratto	26	28	36,99	<0,0001
123	processare	22	22	36,99	<0,0001
124	TAV	22	22	36,99	<0,0001
125	impegnare	31	36	36,73	<0,0001
126	posto	37	46	36,69	<0,0001
127	risolvere	45	60	36,6	<0,0001
128	repubblica	35	43	35,83	<0,0001
129	lista	35	43	35,83	<0,0001
130	Libia	25	27	35,32	<0,0001
131	tavolo	26	29	34,02	<0,0001
132	istituzione	39	51	33,55	<0,0001
133	soldo	51	73	33,26	<0,0001
134	bisogno	67	104	32,98	<0,0001
135	merito	25	28	32,38	<0,0001
136	obiettivo	48	68	32,36	<0,0001
137	Stato	49	70	32,15	<0,0001
138	tutelare	28	33	31,94	<0,0001
139	imporre	19	19	31,94	<0,0001
140	Fornero	19	19	31,94	<0,0001
141	ruolo	31	38	31,91	<0,0001
142	Trump	21	22	31,82	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
143	ragione	39	52	31,69	<0,0001
144	confine o	26	30	31,29	<0,0001
145	campo	48	69	30,85	<0,0001
146	bilancio	24	27	30,74	<0,0001
147	emergenza	33	42	30,65	<0,0001
148	trafficante	22	24	30,34	<0,0001
149	Senato	22	24	30,34	<0,0001
150	alleato	18	18	30,25	<0,0001
151	leader	20	21	30,14	<0,0001
152	democratico	20	21	30,14	<0,0001
153	polemica	25	29	29,68	<0,0001
154	pubblico	71	115	29,63	<0,0001
155	servire	70	113	29,61	<0,0001
156	priorità	28	34	29,54	<0,0001
157	investire	23	26	29,11	<0,0001
158	formazione	21	23	28,69	<0,0001
159	assicurare	21	23	28,69	<0,0001
160	promessa promesso	17	17	28,57	<0,0001
161	occupazione	17	17	28,57	<0,0001
162	popolare	17	17	28,57	<0,0001
163	complesso	17	17	28,57	<0,0001
164	Zingaretti	17	17	28,57	<0,0001
165	propaganda	19	20	28,47	<0,0001
166	atto	41	58	27,74	<0,0001
167	stipendio	22	25	27,49	<0,0001
168	consiglio	36	49	27,47	<0,0001
169	agire	28	35	27,32	<0,0001
170	stabilità	20	22	27,04	<0,0001
171	abolire	16	16	26,89	<0,0001
172	collaborazione	16	16	26,89	<0,0001
173	centro_sinistra	16	16	26,89	<0,0001
174	Bruxelles	16	16	26,89	<0,0001
175	inaccettabile	18	19	26,8	<0,0001
176	fiducia	33	44	26,79	<0,0001
177	decisione	39	55	26,64	<0,0001
178	Stati_Uniti	23	27	26,48	<0,0001
179	rischiare	30	39	26,22	<0,0001
180	restare	35	48	26,09	<0,0001
181	corte	19	21	25,39	<0,0001
182	eleggere	15	15	25,2	<0,0001
183	liberale	15	15	25,2	<0,0001
184	coesione	15	15	25,2	<0,0001
185	benessere	17	18	25,13	<0,0001
186	mandare	39	56	25,12	<0,0001
187	vittoria	25	31	24,92	<0,0001
188	rilanciare	20	23	24,25	<0,0001
189	rappresentare	48	74	24,15	<0,0001
190	missione	23	28	24,08	<0,0001
191	adottare	23	28	24,08	<0,0001
192	progetto	31	42	23,97	<0,0001
193	base	31	42	23,97	<0,0001
194	favore	18	20	23,75	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
195	clandestino	18	20	23,75	<0,0001
196	futuro	78	136	23,63	<0,0001
197	operativo	14	14	23,52	<0,0001
198	elettore	14	14	23,52	<0,0001
199	credibilità	14	14	23,52	<0,0001
200	credibile	14	14	23,52	<0,0001
201	Europeo	14	14	23,52	<0,0001
202	industriale	16	17	23,46	<0,0001
203	avversario	16	17	23,46	<0,0001
204	difesa	24	30	23,4	<0,0001
205	prospettiva	24	30	23,4	<0,0001
206	lotta	33	46	23,37	<0,0001
207	unico	75	130	23,35	<0,0001
208	scelta	42	63	23,31	<0,0001
209	piazza	29	39	22,93	<0,0001
210	alternativo	25	32	22,82	<0,0001
211	creare	73	127	22,33	<0,0001
212	guidare	17	19	22,11	<0,0001
213	applicare	17	19	22,11	<0,0001
214	riguardare	56	92	21,98	<0,0001
215	incompetenza	13	13	21,84	<0,0001
216	recessione	13	13	21,84	<0,0001
217	opposizione	13	13	21,84	<0,0001
218	terrorismo	13	13	21,84	<0,0001
219	serietà	13	13	21,84	<0,0001
220	ribadire	13	13	21,84	<0,0001
221	revoca	13	13	21,84	<0,0001
222	incapace	13	13	21,84	<0,0001
223	collegio	13	13	21,84	<0,0001
224	aumentare	48	76	21,83	<0,0001
225	debito	15	16	21,79	<0,0001
226	costituzionale	15	16	21,79	<0,0001
227	referendum	20	24	21,75	<0,0001
228	efficace	20	24	21,75	<0,0001
229	funzionare	28	38	21,53	<0,0001
230	prevedere	38	57	21,07	<0,0001
231	regionale	18	21	21,04	<0,0001
232	reato	21	26	21,01	<0,0001
233	Mediterraneo	21	26	21,01	<0,0001
234	volontà	30	42	20,95	<0,0001
235	raggiungere	32	46	20,51	<0,0001
236	sfida	26	35	20,49	<0,0001
237	garanzia	16	18	20,47	<0,0001
238	valutare	19	23	20,19	<0,0001
239	battere	19	23	20,19	<0,0001
240	sbarco	12	12	20,15	<0,0001
241	obbligo	12	12	20,15	<0,0001
242	costituzione	12	12	20,15	<0,0001
243	proporzionale	12	12	20,15	<0,0001
244	salario	14	15	20,12	<0,0001
245	contrastare	14	15	20,12	<0,0001
246	unire	23	30	19,89	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
247	bloccare	37	56	19,87	<0,0001
248	maggiore	40	62	19,7	<0,0001
249	giustizia	69	122	19,51	<0,0001
250	aggiungere	20	25	19,49	<0,0001
251	droga	17	20	19,45	<0,0001
252	busta	17	20	19,45	<0,0001
253	cancellare	25	34	19,09	<0,0001
254	significare	52	87	18,88	<0,0001
255	tradire	15	17	18,85	<0,0001
256	imprenditore	15	17	18,85	<0,0001
257	evidente	18	22	18,64	<0,0001
258	evitare	45	73	18,57	<0,0001
259	valore	40	63	18,53	<0,0001
260	scafista	11	11	18,47	<0,0001
261	agenda	11	11	18,47	<0,0001
262	Jobs_act	11	11	18,47	<0,0001
263	attivare	11	11	18,47	<0,0001
264	Italia_Viva	11	11	18,47	<0,0001
265	pensionato	11	11	18,47	<0,0001
266	finanziario	11	11	18,47	<0,0001
267	burocrazia	11	11	18,47	<0,0001
268	moderato	13	14	18,46	<0,0001
269	proseguire	13	14	18,46	<0,0001
270	gestione	13	14	18,46	<0,0001
271	chiacchiera	13	14	18,46	<0,0001
272	penale	13	14	18,46	<0,0001
273	Venezuela	13	14	18,46	<0,0001
274	Lombardia	19	24	17,98	<0,0001
275	tutela	19	24	17,98	<0,0001
276	mercato	24	33	17,72	<0,0001
277	rimettere	24	33	17,72	<0,0001
278	contributo	14	16	17,23	<0,0001
279	istituzionale	14	16	17,23	<0,0001
280	movimento	26	37	17,21	<0,0001
281	rafforzare	17	21	17,1	<0,0001
282	blocco	17	21	17,1	<0,0001
283	sostegno	21	28	17,02	<0,0001
284	caos	12	13	16,8	<0,0001
285	pedaggio	12	13	16,8	<0,0001
286	largo	12	13	16,8	<0,0001
287	consenso	12	13	16,8	<0,0001
288	legislatura	10	10	16,79	<0,0001
289	autostradale	10	10	16,79	<0,0001
290	Grillo	10	10	16,79	<0,0001
291	sottoscrivere	10	10	16,79	<0,0001
292	parlamento	10	10	16,79	<0,0001
293	cessare	10	10	16,79	<0,0001
294	legittimo	10	10	16,79	<0,0001
295	approvazione	10	10	16,79	<0,0001
296	Sea_Watch	10	10	16,79	<0,0001
297	Diciotti	10	10	16,79	<0,0001
298	Camera	10	10	16,79	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
299	intero	30	45	16,61	<0,0001
300	sostenere	31	47	16,53	<0,0001
301	Francia	23	32	16,36	<0,0001
302	controllo	35	55	16,34	<0,0001
303	comune	36	57	16,33	<0,0001
304	taglio	15	18	16,3	<0,0001
305	riduzione	15	18	16,3	<0,0001
306	obbligatorio	15	18	16,3	<0,0001
307	contare	15	18	16,3	<0,0001
308	azione	24	34	16,11	<0,0001
309	costo	19	25	16	<0,0001
310	norma	19	25	16	<0,0001
311	ministero	19	25	16	<0,0001
312	Iran	19	25	16	<0,0001
313	militare	19	25	16	<0,0001
314	destra	25	36	15,91	<0,0001
315	passo	45	76	15,66	<0,0001
316	interno	67	123	15,63	<0,0001
317	mafioso	13	15	15,62	<0,0001
318	certezza	13	15	15,62	<0,0001
319	aula	13	15	15,62	<0,0001
320	estero	27	40	15,61	<0,0001
321	linea	27	40	15,61	<0,0001
322	accoglienza	20	27	15,61	<0,0001
323	solidarietà	20	27	15,61	<0,0001
324	organizzazione	16	20	15,58	<0,0001
325	dato	28	42	15,5	<0,0001
326	crescere	37	60	15,27	<0,0001
327	sede	11	12	15,15	<0,0001
328	Silvio	11	12	15,15	<0,0001
329	vertice	11	12	15,15	<0,0001
330	segretario	11	12	15,15	<0,0001
331	flat_tax	11	12	15,15	<0,0001
332	firmare	11	12	15,15	<0,0001
333	burocratico	11	12	15,15	<0,0001
334	cellula	9	9	15,11	0,0001
335	Più_Europa	9	9	15,11	0,0001
336	rilevante	9	9	15,11	0,0001
337	patto	9	9	15,11	0,0001
338	irregolare	9	9	15,11	0,0001
339	dimissione	9	9	15,11	0,0001
340	occidentale	9	9	15,11	0,0001
341	irrealizzabile	9	9	15,11	0,0001
342	europeista	9	9	15,11	0,0001
343	estremismo	9	9	15,11	0,0001
344	abolizione	9	9	15,11	0,0001
345	Spagna	9	9	15,11	0,0001
346	tagliare	17	22	15,02	0,0001
347	aumento	23	33	14,81	0,00011
348	libico	14	17	14,74	0,00012
349	giudizio	14	17	14,74	0,00012
350	dichiarare	14	17	14,74	0,00012

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
351	flusso	14	17	14,74	0,00012
352	svolgere	24	35	14,64	0,00013
353	cantiere	18	24	14,58	0,00013
354	salute	39	65	14,36	0,00015
355	confronto	38	63	14,31	0,00015
356	libertà	27	41	14,3	0,00015
357	arma arme armo	19	26	14,22	0,00016
358	necessario	34	55	14,16	0,00016
359	realizzare	30	47	14,16	0,00016
360	eliminare	15	19	14,08	0,00017
361	concentrare	15	19	14,08	0,00017
362	monitoraggio	12	14	14,02	0,00018
363	commissario	12	14	14,02	0,00018
364	innovazione	12	14	14,02	0,00018
365	occorrere	20	28	13,94	0,00018
366	centro	42	72	13,66	0,00021
367	regime	10	11	13,5	0,00023
368	testo	10	11	13,5	0,00023
369	evasione	10	11	13,5	0,00023
370	consapevolezza	10	11	13,5	0,00023
371	ridare	10	11	13,5	0,00023
372	prevalere	10	11	13,5	0,00023
373	Stati	10	11	13,5	0,00023
374	profugo	8	8	13,43	0,00024
375	inciucio	8	8	13,43	0,00024
376	confusione	8	8	13,43	0,00024
377	bonus	8	8	13,43	0,00024
378	versare	8	8	13,43	0,00024
379	stabilizzare	8	8	13,43	0,00024
380	sollecitare	8	8	13,43	0,00024
381	rimpatrio	8	8	13,43	0,00024
382	parlamentarie	8	8	13,43	0,00024
383	pagano	8	8	13,43	0,00024
384	news	8	8	13,43	0,00024
385	fake	8	8	13,43	0,00024
386	esecutivo	8	8	13,43	0,00024
387	cuneo	8	8	13,43	0,00024
388	candidare	8	8	13,43	0,00024
389	agricoltura	8	8	13,43	0,00024
390	affidare	8	8	13,43	0,00024
391	Israele	8	8	13,43	0,00024
392	ISTAT	8	8	13,43	0,00024
393	strategia	13	16	13,2	0,00027
394	Macron	13	16	13,2	0,00027
395	amministrazione	13	16	13,2	0,00027
396	scontro	13	16	13,2	0,00027
397	magistratura	13	16	13,2	0,00027
398	povertà	17	23	13,17	0,00028
399	dividere	17	23	13,17	0,00028
400	grave	32	52	13,08	0,00029
401	iniziativa	18	25	12,86	0,00033
402	livello	49	88	12,77	0,00035

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
403	Africa	14	18	12,6	0,00038
404	promettere	14	18	12,6	0,00038
405	dibattito	14	18	12,6	0,00038
406	contrario	14	18	12,6	0,00038
407	smettere	14	18	12,6	0,00038
408	confrontare	14	18	12,6	0,00038
409	marzo	20	29	12,43	0,00042
410	introdurre	11	13	12,43	0,00042
411	Palazzo	11	13	12,43	0,00042
412	Meloni	11	13	12,43	0,00042
413	elemento	21	31	12,28	0,00045
414	umano	44	78	12,22	0,00047
415	costruire	34	57	12,19	0,00047
416	annunciare	15	20	12,14	0,00049
417	favorire	15	20	12,14	0,00049
418	positivo	24	37	12,02	0,00052
419	strumento	24	37	12,02	0,00052
420	ritenere	29	47	11,98	0,00053
421	fatto	41	72	11,91	0,00055
422	unione	9	10	11,86	0,00057
423	razziale	9	10	11,86	0,00057
424	Lazio	9	10	11,86	0,00057
425	unito	9	10	11,86	0,00057
426	revocare	9	10	11,86	0,00057
427	iscrivere	9	10	11,86	0,00057
428	ampio	9	10	11,86	0,00057
429	Craxi	9	10	11,86	0,00057
430	Chigi	9	10	11,86	0,00057
431	punto	78	154	11,85	0,00057
432	Matteo	16	22	11,8	0,00059
433	aliquota	7	7	11,75	0,0006
434	penalizzare	7	7	11,75	0,0006
435	statuto	7	7	11,75	0,0006
436	relativo	7	7	11,75	0,0006
437	competente	7	7	11,75	0,0006
438	Gentiloni	7	7	11,75	0,0006
439	sbloccare	7	7	11,75	0,0006
440	rinnovare	7	7	11,75	0,0006
441	redistribuzione	7	7	11,75	0,0006
442	populismo	7	7	11,75	0,0006
443	garante	7	7	11,75	0,0006
444	fallimento	7	7	11,75	0,0006
445	fallimentare	7	7	11,75	0,0006
446	espulsione	7	7	11,75	0,0006
447	decisivo	7	7	11,75	0,0006
448	concretezza	7	7	11,75	0,0006
449	candidatura	7	7	11,75	0,0006
450	bandiera	7	7	11,75	0,0006
451	assenza	7	7	11,75	0,0006
452	amministratore	7	7	11,75	0,0006
453	alleare	7	7	11,75	0,0006
454	affermazione	7	7	11,75	0,0006

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
455	abbassamento	7	7	11,75	0,0006
456	Raggi	7	7	11,75	0,0006
457	Costituzione	7	7	11,75	0,0006
458	onore	12	15	11,68	0,00063
459	servizio	36	62	11,45	0,00071
460	privato	19	28	11,17	0,00083
461	ripetere	19	28	11,17	0,00083
462	trasformare	20	30	11,06	0,00088
463	combattere	28	46	10,93	0,00094
464	capacità	27	44	10,9	0,00096
465	serio	23	36	10,88	0,00097
466	qualità	26	42	10,87	0,00097
467	motivo	24	38	10,86	0,00098
468	sanità	10	12	10,86	0,00098
469	partecipazione	10	12	10,86	0,00098
470	membro	10	12	10,86	0,00098
471	stabilire	10	12	10,86	0,00098
472	sconfiggere	10	12	10,86	0,00098
473	scambio	10	12	10,86	0,00098
474	chiarezza	10	12	10,86	0,00098
475	giudice	14	19	10,74	0,00104
476	lavorare	89	182	10,6	0,00112
477	nazione	15	21	10,45	0,00122
478	carico	16	23	10,23	0,00138
479	sconfitta sconfitto	8	9	10,23	0,00138
480	scienza	8	9	10,23	0,00138
481	strutturale	8	9	10,23	0,00138
482	spostamento	8	9	10,23	0,00138
483	signor	8	9	10,23	0,00138
484	sovrano	8	9	10,23	0,00138
485	soglia	8	9	10,23	0,00138
486	rivendicare	8	9	10,23	0,00138
487	rapido	8	9	10,23	0,00138
488	industria	8	9	10,23	0,00138
489	incremento	8	9	10,23	0,00138
490	incertezza	8	9	10,23	0,00138
491	incentivo	8	9	10,23	0,00138
492	incapacità	8	9	10,23	0,00138
493	fattore	8	9	10,23	0,00138
494	dimostrazione	8	9	10,23	0,00138
495	comunale	8	9	10,23	0,00138
496	coerente	8	9	10,23	0,00138
497	acquisire	8	9	10,23	0,00138
498	territorio	45	83	10,19	0,00141
499	ONG	11	14	10,18	0,00141
500	partenza	11	14	10,18	0,00141
501	correggere	11	14	10,18	0,00141
502	minaccia	11	14	10,18	0,00141
503	giungere	11	14	10,18	0,00141
504	CGIL	6	6	10,07	0,0015
505	disperdere	6	6	10,07	0,0015
506	criticare	6	6	10,07	0,0015

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
507	consumatore	6	6	10,07	0,0015
508	Liberi_e_Uguali	6	6	10,07	0,0015
509	Bonafede	6	6	10,07	0,0015
510	viadotto	6	6	10,07	0,0015
511	sorveglianza	6	6	10,07	0,0015
512	sgravio	6	6	10,07	0,0015
513	rinnovamento	6	6	10,07	0,0015
514	ricollocaimento	6	6	10,07	0,0015
515	rappresentativo	6	6	10,07	0,0015
516	populista	6	6	10,07	0,0015
517	poltrona poltrone	6	6	10,07	0,0015
518	organizzativo	6	6	10,07	0,0015
519	occidente	6	6	10,07	0,0015
520	migratore migratorio	6	6	10,07	0,0015
521	istituire	6	6	10,07	0,0015
522	individuazione	6	6	10,07	0,0015
523	illegale	6	6	10,07	0,0015
524	garantista	6	6	10,07	0,0015
525	force	6	6	10,07	0,0015
526	estremista	6	6	10,07	0,0015
527	divisione	6	6	10,07	0,0015
528	coinvolgimento	6	6	10,07	0,0015
529	centrista	6	6	10,07	0,0015
530	bugia	6	6	10,07	0,0015
531	azzerare	6	6	10,07	0,0015
532	IVA	6	6	10,07	0,0015
533	IRPEF	6	6	10,07	0,0015
534	Gregoretti	6	6	10,07	0,0015
535	Fontana	6	6	10,07	0,0015
536	Bonino	6	6	10,07	0,0015
537	Bonaccini	6	6	10,07	0,0015
538	Berlino	6	6	10,07	0,0015
539	Benetton	6	6	10,07	0,0015
540	Banca	6	6	10,07	0,0015
541	mantenere	28	47	9,97	0,00158
542	ambiente	25	41	9,83	0,00172
543	augurare	24	39	9,79	0,00175
544	termine	23	37	9,77	0,00176
545	offrire	22	35	9,77	0,00177
546	definitivo	12	16	9,71	0,00183
547	minoranza	9	11	9,31	0,00228
548	invitare	9	11	9,31	0,00228
549	modificare	9	11	9,31	0,00228
550	insicurezza	9	11	9,31	0,00228
551	strategico	9	11	9,31	0,00228
552	fondare	9	11	9,31	0,00228
553	autorizzazione	9	11	9,31	0,00228
554	affermare	9	11	9,31	0,00228
555	litigare	15	22	8,96	0,00276
556	traffico	25	42	8,87	0,0029
557	errore	25	42	8,87	0,0029
558	rivolgere	17	26	8,76	0,00308

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
559	confermare	18	28	8,71	0,00316
560	dichiarazione	10	13	8,71	0,00316
561	principio	10	13	8,71	0,00316
562	indeterminato	10	13	8,71	0,00316
563	visione	10	13	8,71	0,00316
564	spendere	10	13	8,71	0,00316
565	principe principio	10	13	8,71	0,00316
566	prevenzione	10	13	8,71	0,00316
567	minimo	21	34	8,69	0,00319
568	civile	21	34	8,69	0,00319
569	proporre	20	32	8,68	0,00321
570	instabilità	7	8	8,61	0,00335
571	contraddizione	7	8	8,61	0,00335
572	stampa	7	8	8,61	0,00335
573	regolamento	7	8	8,61	0,00335
574	organismo	7	8	8,61	0,00335
575	manutenzione	7	8	8,61	0,00335
576	indebolire	7	8	8,61	0,00335
577	firma	7	8	8,61	0,00335
578	autonomia	7	8	8,61	0,00335
579	vigore	7	8	8,61	0,00335
580	vaccinazione	7	8	8,61	0,00335
581	soldato	7	8	8,61	0,00335
582	significativo	7	8	8,61	0,00335
583	senatorio	7	8	8,61	0,00335
584	sanzione	7	8	8,61	0,00335
585	riunire	7	8	8,61	0,00335
586	riunione	7	8	8,61	0,00335
587	riaprire	7	8	8,61	0,00335
588	realizzazione	7	8	8,61	0,00335
589	quota_100	7	8	8,61	0,00335
590	presidenza	7	8	8,61	0,00335
591	francamente	7	8	8,61	0,00335
592	finanza	7	8	8,61	0,00335
593	favorevole	7	8	8,61	0,00335
594	esercitare	7	8	8,61	0,00335
595	costi_benefici	7	8	8,61	0,00335
596	Veneto	7	8	8,61	0,00335
597	fronte	36	66	8,4	0,00374
598	oppressione	5	5	8,39	0,00377
599	stampare	5	5	8,39	0,00377
600	sondaggio	5	5	8,39	0,00377
601	raddoppiare	5	5	8,39	0,00377
602	prestigio	5	5	8,39	0,00377
603	divergenza	5	5	8,39	0,00377
604	diciottenne	5	5	8,39	0,00377
605	Nicola	5	5	8,39	0,00377
606	Carige	5	5	8,39	0,00377
607	disporre	11	15	8,31	0,00393
608	emergere	11	15	8,31	0,00393
609	basare	11	15	8,31	0,00393
610	condividere	28	49	8,25	0,00407

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
611	mese	51	100	8,06	0,00453
612	pressione	12	17	8,04	0,00457
613	bisognare	87	184	7,93	0,00486
614	popolazione	13	19	7,85	0,00506
615	ripartire	13	19	7,85	0,00506
616	materia	13	19	7,85	0,00506
617	affrontare	31	56	7,81	0,00519
618	spreco	8	10	7,78	0,00527
619	integrazione	8	10	7,78	0,00527
620	cogliere	8	10	7,78	0,00527
621	avviare	8	10	7,78	0,00527
622	altrettanto	8	10	7,78	0,00527
623	posizione	22	37	7,76	0,00533
624	precedente	14	21	7,73	0,00542
625	sbarcare	14	21	7,73	0,00542
626	responsabile	14	21	7,73	0,00542
627	protagonista	14	21	7,73	0,00542
628	scenario	14	21	7,73	0,00542
629	sanitario	21	35	7,7	0,00551
630	commissione	15	23	7,65	0,00567
631	coinvolgere	15	23	7,65	0,00567
632	abbassare	18	29	7,6	0,00584
633	intendere	18	29	7,6	0,00584
634	differenza	18	29	7,6	0,00584
635	famiglia	100	216	7,58	0,0059
636	salvare	28	50	7,47	0,00625
637	Roma	38	72	7,38	0,00658
638	modello	9	12	7,28	0,00697
639	pelle	9	12	7,28	0,00697
640	opporre	9	12	7,28	0,00697
641	complice	9	12	7,28	0,00697
642	possibilità	41	79	7,21	0,00724
643	porto	31	57	7,12	0,00764
644	implicare	6	7	7	0,00814
645	improprio	6	7	7	0,00814
646	illusione	6	7	7	0,00814
647	disponibilità	6	7	7	0,00814
648	chiarire	6	7	7	0,00814
649	trattativa	6	7	7	0,00814
650	task	6	7	7	0,00814
651	stragrande	6	7	7	0,00814
652	premiare	6	7	7	0,00814
653	opinione	6	7	7	0,00814
654	nido	6	7	7	0,00814
655	navale	6	7	7	0,00814
656	maggiormente	6	7	7	0,00814
657	imputato	6	7	7	0,00814
658	esito	6	7	7	0,00814
659	diseguaglianze	6	7	7	0,00814
660	complessivo	6	7	7	0,00814
661	cinismo	6	7	7	0,00814
662	calabrese	6	7	7	0,00814

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
663	mandato	10	14	6,96	0,00834
664	orgoglio	10	14	6,96	0,00834
665	crimine	10	14	6,96	0,00834
666	scegliere	38	73	6,8	0,0091
667	copertura	11	16	6,75	0,00935
668	vita vite	11	16	6,75	0,00935
669	operare	11	16	6,75	0,00935
670	culturale	11	16	6,75	0,00935
671	riconoscere	20	34	6,72	0,0095
672	imposta	4	4	6,71	0,00957
673	Tunisia	4	4	6,71	0,00957
674	truffa	4	4	6,71	0,00957
675	risparmiatore	4	4	6,71	0,00957
676	miliardario	4	4	6,71	0,00957
677	Sardine	4	4	6,71	0,00957
678	web	4	4	6,71	0,00957
679	squadra	19	32	6,66	0,00987
680	Malta	12	18	6,62	0,01006
681	capitale	12	18	6,62	0,01006
682	procedere	12	18	6,62	0,01006
683	vaccino	18	30	6,6	0,0102
684	attivo	13	20	6,55	0,01049
685	produzione	16	26	6,52	0,01067
686	singolo	16	26	6,52	0,01067
687	africano	7	9	6,29	0,01213
688	Sophia	7	9	6,29	0,01213
689	prevenire	7	9	6,29	0,01213
690	moneta	7	9	6,29	0,01213
691	comunista	7	9	6,29	0,01213
692	opzione	7	9	6,29	0,01213
693	frattura	7	9	6,29	0,01213
694	applicazione	7	9	6,29	0,01213
695	Maduro	7	9	6,29	0,01213
696	professione	7	9	6,29	0,01213
697	primario	7	9	6,29	0,01213
698	netto	7	9	6,29	0,01213
699	equilibrio	7	9	6,29	0,01213
700	consumo	7	9	6,29	0,01213
701	attenzione	34	65	6,26	0,01234
702	idea	45	90	6,22	0,01264
703	generale	29	54	6,21	0,0127
704	libero	23	41	6,19	0,01287
705	permettere	47	95	6,03	0,01404
706	gestire	21	37	5,98	0,01447
707	assoluto	21	37	5,98	0,01447
708	determinazione	8	11	5,89	0,01522
709	risparmio	8	11	5,89	0,01522
710	ennesimo	8	11	5,89	0,01522
711	conferma	8	11	5,89	0,01522
712	stabile	8	11	5,89	0,01522
713	iraniano	8	11	5,89	0,01522
714	inchiesta	8	11	5,89	0,01522

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
715	auspicare	8	11	5,89	0,01522
716	risposta	35	68	5,84	0,01567
717	consentire	30	57	5,72	0,01672
718	vero	65	138	5,7	0,01698
719	dipendere	9	13	5,65	0,01741
720	manifestare	9	13	5,65	0,01741
721	limitare	9	13	5,65	0,01741
722	conflitto	9	13	5,65	0,01741
723	importante	94	208	5,6	0,01794
724	determinare	16	27	5,55	0,01843
725	importanza	10	15	5,52	0,01882
726	tecnico	15	25	5,5	0,01906
727	protezione	14	23	5,45	0,01956
728	rivoluzione	11	17	5,45	0,01961
729	francese	11	17	5,45	0,01961
730	giudiziario	11	17	5,45	0,01961
731	centrale	11	17	5,45	0,01961
732	storico	13	21	5,42	0,01987
733	diminuire	13	21	5,42	0,01987
734	settore	12	19	5,42	0,01992
735	legalità	12	19	5,42	0,01992
736	debole	12	19	5,42	0,01992
737	personalmente	12	19	5,42	0,01992
738	razzista	5	6	5,42	0,01985
739	punire	5	6	5,42	0,01985
740	premier	5	6	5,42	0,01985
741	islamico	5	6	5,42	0,01985
742	giustificare	5	6	5,42	0,01985
743	diplomatico	5	6	5,42	0,01985
744	umanamente	5	6	5,42	0,01985
745	scomparsa scomparso	5	6	5,42	0,01985
746	procedimento	5	6	5,42	0,01985
747	perseguire	5	6	5,42	0,01985
748	ordinanza	5	6	5,42	0,01985
749	moltiplicare	5	6	5,42	0,01985
750	istanza	5	6	5,42	0,01985
751	intelligenza	5	6	5,42	0,01985
752	incidere	5	6	5,42	0,01985
753	grillini	5	6	5,42	0,01985
754	gratuito	5	6	5,42	0,01985
755	formativo	5	6	5,42	0,01985
756	fermezza	5	6	5,42	0,01985
757	esponente	5	6	5,42	0,01985
758	esaminare	5	6	5,42	0,01985
759	convergenza	5	6	5,42	0,01985
760	coerenza	5	6	5,42	0,01985
761	certificare	5	6	5,42	0,01985
762	Livorno	5	6	5,42	0,01985
763	rendere	41	83	5,2	0,02262
764	riscrivere	3	3	5,03	0,02486
765	cambiamento	19	34	5,01	0,02522
766	parere	19	34	5,01	0,02522

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
767	tutte_le	79	174	4,92	0,02654
768	Emma	6	8	4,85	0,02764
769	vantaggio	6	8	4,85	0,02764
770	taglia	6	8	4,85	0,02764
771	rinnovabile	6	8	4,85	0,02764
772	offerta	6	8	4,85	0,02764
773	apprezzare	6	8	4,85	0,02764
774	trasparente	6	8	4,85	0,02764
775	rifugiato	6	8	4,85	0,02764
776	rappresentante	6	8	4,85	0,02764
777	ragionare	6	8	4,85	0,02764
778	manetta manette	6	8	4,85	0,02764
779	disumano	6	8	4,85	0,02764
780	addestramento	6	8	4,85	0,02764
781	accogliere	17	30	4,8	0,02847
782	partecipare	17	30	4,8	0,02847
783	bianco	16	28	4,7	0,03013
784	utile	16	28	4,7	0,03013
785	grande	145	338	4,63	0,0314
786	tribunale	15	26	4,61	0,03177
787	universale	7	10	4,56	0,03272
788	trasporto	7	10	4,56	0,03272
789	razza	7	10	4,56	0,03272
790	apparire	7	10	4,56	0,03272
791	regolare	7	10	4,56	0,03272
792	tranquillamente	7	10	4,56	0,03272
793	tradizionale	7	10	4,56	0,03272
794	magistrato	7	10	4,56	0,03272
795	invidia	7	10	4,56	0,03272
796	continente	7	10	4,56	0,03272
797	avviso	7	10	4,56	0,03272
798	riferimento	13	22	4,45	0,03481
799	peso	8	12	4,41	0,03567
800	pagina	8	12	4,41	0,03567
801	ambientale	8	12	4,41	0,03567
802	direzione	12	20	4,39	0,03606
803	impedire	12	20	4,39	0,03606
804	costituire	12	20	4,39	0,03606
805	dipendente	11	18	4,35	0,03698
806	sud	11	18	4,35	0,03698
807	Torino	11	18	4,35	0,03698
808	cambio	11	18	4,35	0,03698
809	colpevole	9	14	4,35	0,03707
810	attacco	9	14	4,35	0,03707
811	complicare	9	14	4,35	0,03707
812	Mattarella	9	14	4,35	0,03707
813	analisi	9	14	4,35	0,03707
814	contribuire	10	16	4,33	0,0374
815	chiusura	10	16	4,33	0,0374
816	trivella	4	5	3,89	0,04862
817	sintesi	4	5	3,89	0,04862
818	pagamento	4	5	3,89	0,04862

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
819	derivare	4	5	3,89	0,04862
820	abbattimento	4	5	3,89	0,04862

Tab. 4 Forme lessicali significative per la classe 2 (CGD) del *sub-corpus* PF_Donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	governo	64	70	235,31	<0,0001
1	Italia	57	73	162,78	<0,0001
2	5_Stelle	34	34	140,57	<0,0001
3	Movimento	30	30	123,78	<0,0001
4	tassa	28	28	115,4	<0,0001
5	paese	47	70	102,86	<0,0001
6	politico	40	55	100,19	<0,0001
7	prossimo	28	31	98,96	<0,0001
8	Salvini	28	31	98,96	<0,0001
9	centro_destra	24	24	98,71	<0,0001
10	elettorale	26	28	95,77	<0,0001
11	impresa	20	21	76,31	<0,0001
12	cittadino	30	41	75,39	<0,0001
13	campagna	17	17	69,67	<0,0001
14	Fratelli_d	18	19	68,06	<0,0001
15	europeo	19	21	66,98	<0,0001
16	italiano	40	70	64,03	<0,0001
17	Berlusconi	15	15	61,41	<0,0001
18	Europa	19	23	58,01	<0,0001
19	maggioranza	15	16	55,72	<0,0001
20	Forza_Italia	15	16	55,72	<0,0001
21	votare	16	18	54,78	<0,0001
22	legge	19	24	54,11	<0,0001
23	Lega	13	13	53,17	<0,0001
24	Emilia	12	12	49,05	<0,0001
25	riforma	13	14	47,54	<0,0001
26	sviluppo	13	14	47,54	<0,0001
27	tema	14	16	46,71	<0,0001
28	elezione	11	11	44,94	<0,0001
29	Romagna	11	11	44,94	<0,0001
30	coalizione	11	11	44,94	<0,0001
31	governare	11	11	44,94	<0,0001
32	forza	23	37	42,79	<0,0001
33	partito	15	19	42,41	<0,0001
34	prescrizione	10	10	40,83	<0,0001
35	decreto	10	10	40,83	<0,0001
36	accordo	11	12	39,39	<0,0001
37	Parlamento	11	12	39,39	<0,0001
38	processo	15	20	38,91	<0,0001
39	lavoratore	12	14	38,7	<0,0001
40	proposta	12	14	38,7	<0,0001
41	pagare	16	23	36,43	<0,0001
42	ministro	14	19	35,2	<0,0001
43	obiettivo	13	17	34,8	<0,0001
44	crescita	13	17	34,8	<0,0001
45	restare	12	15	34,61	<0,0001
46	unico	24	44	34,38	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
47	servire	18	29	33,25	<0,0001
48	sindaco	8	8	32,63	<0,0001
49	immigrazione	8	8	32,63	<0,0001
50	Renzi	8	8	32,63	<0,0001
51	misura	8	8	32,63	<0,0001
52	concessione	8	8	32,63	<0,0001
53	reddito	9	10	31,29	<0,0001
54	assumere	9	10	31,29	<0,0001
55	programma	10	12	30,79	<0,0001
56	ruolo	10	12	30,79	<0,0001
57	consentire	10	12	30,79	<0,0001
58	voto	11	14	30,78	<0,0001
59	priorità	7	7	28,54	<0,0001
60	immigrato	7	7	28,54	<0,0001
61	dividere	7	7	28,54	<0,0001
62	blocco	7	7	28,54	<0,0001
63	alleanza	7	7	28,54	<0,0001
64	PD	7	7	28,54	<0,0001
65	Meloni	7	7	28,54	<0,0001
66	Calabria	7	7	28,54	<0,0001
67	interesse	8	9	27,26	<0,0001
68	candidato	8	9	27,26	<0,0001
69	lavoro	34	80	27,23	<0,0001
70	inciucio	6	6	24,45	<0,0001
71	accoglienza	6	6	24,45	<0,0001
72	riduzione	6	6	24,45	<0,0001
73	revoca	6	6	24,45	<0,0001
74	referendum	6	6	24,45	<0,0001
75	merito	6	6	24,45	<0,0001
76	credibile	6	6	24,45	<0,0001
77	chiacchiera	6	6	24,45	<0,0001
78	Silvio	6	6	24,45	<0,0001
79	Partito_Democratico	6	6	24,45	<0,0001
80	banca	7	8	23,26	<0,0001
81	economia	9	12	23,25	<0,0001
82	decisione	8	10	23,02	<0,0001
83	abbassare	5	5	20,36	<0,0001
84	proseguire	5	5	20,36	<0,0001
85	penale	5	5	20,36	<0,0001
86	pedaggio	5	5	20,36	<0,0001
88	manovra	5	5	20,36	<0,0001
89	legislatura	5	5	20,36	<0,0001
90	inchiesta	5	5	20,36	<0,0001
91	fiscale	5	5	20,36	<0,0001
92	fallimentare	5	5	20,36	<0,0001
93	approvare	5	5	20,36	<0,0001
94	analisi	5	5	20,36	<0,0001
95	Matteo	5	5	20,36	<0,0001
96	ISTAT	5	5	20,36	<0,0001
97	ridurre	8	11	19,58	<0,0001
98	concreto	8	11	19,58	<0,0001
99	pensione	6	7	19,28	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
100	busta	6	7	19,28	<0,0001
101	annunciare	6	7	19,28	<0,0001
102	TAV	7	9	19,2	<0,0001
103	rischio	10	16	18,6	<0,0001
104	vincere	11	19	17,61	<0,0001
105	presentare	8	12	16,76	<0,0001
106	provvedimento	4	4	16,28	<0,0001
107	unità	4	4	16,28	<0,0001
108	premio	4	4	16,28	<0,0001
109	irregolare	4	4	16,28	<0,0001
110	insicurezza	4	4	16,28	<0,0001
111	trasformare	4	4	16,28	<0,0001
112	stipendio	4	4	16,28	<0,0001
113	serietà	4	4	16,28	<0,0001
114	segnare	4	4	16,28	<0,0001
115	rilanciare	4	4	16,28	<0,0001
116	regolare	4	4	16,28	<0,0001
117	regionale	4	4	16,28	<0,0001
118	reato	4	4	16,28	<0,0001
119	proporzionale	4	4	16,28	<0,0001
120	piegare	4	4	16,28	<0,0001
121	parlamentare	4	4	16,28	<0,0001
122	opposizione	4	4	16,28	<0,0001
123	navale	4	4	16,28	<0,0001
124	argo	4	4	16,28	<0,0001
125	iscrivere	4	4	16,28	<0,0001
126	guidare	4	4	16,28	<0,0001
127	elettore	4	4	16,28	<0,0001
128	costi_benefici	4	4	16,28	<0,0001
129	controllo	4	4	16,28	<0,0001
130	centralità	4	4	16,28	<0,0001
131	caos	4	4	16,28	<0,0001
132	abbassamento	4	4	16,28	<0,0001
133	Jobs_act	4	4	16,28	<0,0001
134	Italia_Viva	4	4	16,28	<0,0001
135	Europeo	4	4	16,28	<0,0001
136	Benetton	4	4	16,28	<0,0001
137	mare	7	10	16,01	<0,0001
138	confronto	7	10	16,01	<0,0001
139	evitare	7	10	16,01	<0,0001
140	volontà	6	8	15,47	<0,0001
141	contratto	6	8	15,47	<0,0001
142	tagliare	6	8	15,47	<0,0001
143	costa	5	6	15,35	<0,0001
144	traffico	5	6	15,35	<0,0001
145	obbligatorio	5	6	15,35	<0,0001
146	capitale	5	6	15,35	<0,0001
147	bilancio	5	6	15,35	<0,0001
148	Senato	5	6	15,35	<0,0001
149	bloccare	8	13	14,41	0,00014
150	lotta	8	13	14,41	0,00014
151	passo	12	24	14,02	0,00018

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
152	aumentare	9	16	13,56	0,00023
153	responsabilità	9	16	13,56	0,00023
154	garantire	7	11	13,44	0,00024
155	sinistra	6	9	12,55	0,00039
156	risultato	6	9	12,55	0,00039
157	consiglio	6	9	12,55	0,00039
158	norma	6	9	12,55	0,00039
159	Roma	8	14	12,44	0,00042
160	investimento	8	14	12,44	0,00042
161	propaganda	3	3	12,21	0,00047
162	tavolo	3	3	12,21	0,00047
163	soglia	3	3	12,21	0,00047
164	segnatamente	3	3	12,21	0,00047
165	scandaloso	3	3	12,21	0,00047
166	riferimento	3	3	12,21	0,00047
167	ridare	3	3	12,21	0,00047
168	redistribuzione	3	3	12,21	0,00047
169	professione	3	3	12,21	0,00047
170	premiare	3	3	12,21	0,00047
171	popolare	3	3	12,21	0,00047
172	opporre	3	3	12,21	0,00047
173	nord	3	3	12,21	0,00047
174	migrare	3	3	12,21	0,00047
175	incapace	3	3	12,21	0,00047
176	governabilità	3	3	12,21	0,00047
177	gestione	3	3	12,21	0,00047
178	franco	3	3	12,21	0,00047
179	disabile	3	3	12,21	0,00047
180	definire	3	3	12,21	0,00047
181	criterio	3	3	12,21	0,00047
182	credibilità	3	3	12,21	0,00047
183	costituzionale	3	3	12,21	0,00047
184	consulta	3	3	12,21	0,00047
185	coerente	3	3	12,21	0,00047
186	burocrazia	3	3	12,21	0,00047
187	bonus	3	3	12,21	0,00047
188	anticipato	3	3	12,21	0,00047
189	abolizione	3	3	12,21	0,00047
190	Sea_Watch	3	3	12,21	0,00047
191	Palazzo	3	3	12,21	0,00047
192	Macron	3	3	12,21	0,00047
193	Dublino	3	3	12,21	0,00047
194	Bonaccini	3	3	12,21	0,00047
195	differenza	5	7	11,83	0,00058
196	bisogno	12	26	11,59	0,00066
197	ritardo	4	5	11,48	0,0007
198	primario	4	5	11,48	0,0007
199	partecipazione	4	5	11,48	0,0007
200	medio	4	5	11,48	0,0007
201	efficace	4	5	11,48	0,0007
202	discussione	4	5	11,48	0,0007
203	collegio	4	5	11,48	0,0007

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
204	autorizzazione	4	5	11,48	0,0007
205	Libia	4	5	11,48	0,0007
206	sostenere	7	12	11,34	0,00075
207	faccia	9	18	10,49	0,0012
208	soluzione	6	10	10,27	0,00135
209	reale	6	10	10,27	0,00135
210	milione	5	8	9,26	0,00234
211	causa	5	8	9,26	0,00234
212	solidarietà	5	8	9,26	0,00234
213	rinunciare	5	8	9,26	0,00234
214	interessare	5	8	9,26	0,00234
215	cittadinanza	5	8	9,26	0,00234
216	dimostrare	6	11	8,45	0,00365
217	meritare	4	6	8,36	0,00384
218	svolgere	4	6	8,36	0,00384
219	processare	4	6	8,36	0,00384
220	povertà	4	6	8,36	0,00384
221	giudice	4	6	8,36	0,00384
222	vero	13	33	8,17	0,00426
223	opera	7	14	8,14	0,00433
224	sicurezza	9	20	8,13	0,00436
225	scienza	3	4	7,72	0,00546
226	palazzo	3	4	7,72	0,00546
227	stabilire	3	4	7,72	0,00546
228	seriamente	3	4	7,72	0,00546
229	scontro	3	4	7,72	0,00546
230	pressione	3	4	7,72	0,00546
231	partenza	3	4	7,72	0,00546
232	nemico	3	4	7,72	0,00546
233	manco	3	4	7,72	0,00546
234	litigare	3	4	7,72	0,00546
235	infrastruttura	3	4	7,72	0,00546
236	funzionare	3	4	7,72	0,00546
237	francese	3	4	7,72	0,00546
238	diciottenne	3	4	7,72	0,00546
239	democrazia	3	4	7,72	0,00546
240	democratico	3	4	7,72	0,00546
241	crisi	3	4	7,72	0,00546
242	contraddizione	3	4	7,72	0,00546
243	comprendere	3	4	7,72	0,00546
244	chiarezza	3	4	7,72	0,00546
245	base	3	4	7,72	0,00546
246	autostrada	3	4	7,72	0,00546
247	assassino	3	4	7,72	0,00546
248	Di_Maio	3	4	7,72	0,00546
249	presidente	11	27	7,61	0,00581
250	continuare	15	41	7,49	0,00622
251	commissione	5	9	7,31	0,00685
252	riconoscere	5	9	7,31	0,00685
253	lista	5	9	7,31	0,00685
254	rappresentare	8	18	6,99	0,00821
255	condividere	6	12	6,97	0,00829

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
256	settimana	11	28	6,84	0,00892
257	tutela	4	7	6,2	0,0128
258	progetto	4	7	6,2	0,0128
259	accogliere	4	7	6,2	0,0128
260	piazza	4	7	6,2	0,0128
261	fondamentale	4	7	6,2	0,0128
262	nuovo	14	40	5,99	0,01439
263	sistema	7	16	5,86	0,01548
264	porto	7	16	5,86	0,01548
265	nazionale	5	10	5,8	0,01601
266	maggiore	5	10	5,8	0,01601
267	internazionale	6	13	5,75	0,01644
268	fatto	6	13	5,75	0,01644
269	scelta	9	23	5,51	0,01887
270	oggi	23	77	5,18	0,0229
271	universale	3	5	5,12	0,02362
272	agire	3	5	5,12	0,02362
273	fiducia	3	5	5,12	0,02362
274	sostegno	3	5	5,12	0,02362
275	sanitario	3	5	5,12	0,02362
276	romano	3	5	5,12	0,02362
277	raggiungere	3	5	5,12	0,02362
278	impedire	3	5	5,12	0,02362
279	eliminare	3	5	5,12	0,02362
280	discutere	3	5	5,12	0,02362
281	comune	3	5	5,12	0,02362
282	decidere	7	17	4,97	0,02586
283	necessario	6	14	4,75	0,02932
284	esistere	6	14	4,75	0,02932
285	terzo	4	8	4,64	0,0313
286	posto	4	8	4,64	0,0313
287	ricerca	4	8	4,64	0,0313
288	popolo	4	8	4,64	0,0313
289	economico	4	8	4,64	0,0313
290	creare	8	21	4,51	0,03379
291	polvere	2	3	4,17	0,0411
292	impatto	2	3	4,17	0,0411
293	utile	2	3	4,17	0,0411
294	strategico	2	3	4,17	0,0411
295	strategia	2	3	4,17	0,0411
296	spettare	2	3	4,17	0,0411
297	solido	2	3	4,17	0,0411
298	settore	2	3	4,17	0,0411
299	schiaffo	2	3	4,17	0,0411
300	scenario	2	3	4,17	0,0411
301	ripartire	2	3	4,17	0,0411
302	revocare	2	3	4,17	0,0411
303	record	2	3	4,17	0,0411
304	rapido	2	3	4,17	0,0411
305	presidenza	2	3	4,17	0,0411
306	plastica	2	3	4,17	0,0411
307	pagano	2	3	4,17	0,0411

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
308	news	2	3	4,17	0,0411
309	minacciare	2	3	4,17	0,0411
310	luglio	2	3	4,17	0,0411
311	leader	2	3	4,17	0,0411
312	iniziativa	2	3	4,17	0,0411
313	inclusione	2	3	4,17	0,0411
314	incentivo	2	3	4,17	0,0411
315	incapacità	2	3	4,17	0,0411
316	guida	2	3	4,17	0,0411
317	galleria	2	3	4,17	0,0411
318	funzione	2	3	4,17	0,0411
319	firmare	2	3	4,17	0,0411
320	fake	2	3	4,17	0,0411
321	estremo	2	3	4,17	0,0411
322	decretare	2	3	4,17	0,0411
323	crimine	2	3	4,17	0,0411
324	criminalità	2	3	4,17	0,0411
325	corrispondere	2	3	4,17	0,0411
326	constatare	2	3	4,17	0,0411
327	consenso	2	3	4,17	0,0411
328	concentrare	2	3	4,17	0,0411
329	comporre	2	3	4,17	0,0411
330	cognome	2	3	4,17	0,0411
331	clima	2	3	4,17	0,0411
332	burocratico	2	3	4,17	0,0411
333	avviare	2	3	4,17	0,0411
334	affidare	2	3	4,17	0,0411
335	acquisto	2	3	4,17	0,0411
336	Morandi	2	3	4,17	0,0411
337	Iran	2	3	4,17	0,0411
338	Fornero	2	3	4,17	0,0411
339	soldo	6	15	3,91	0,04802

Tab. 5 Forme lessicali significative per la classe 1 (CGD) del *sub-corpus* PF_Uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	dire	397	687	79,73	<0,0001
1	bello	83	97	77,07	<0,0001
2	ragazzo	85	101	75,2	<0,0001
3	sentire	120	162	70,59	<0,0001
4	figlio	85	106	64,84	<0,0001
5	via	74	93	54,95	<0,0001
6	vedere	169	271	48,59	<0,0001
7	vivere	75	100	45,66	<0,0001
8	fratello	30	31	38,5	<0,0001
9	uscire	70	96	38,46	<0,0001
10	film	31	33	36,86	<0,0001
11	succedere	86	126	36,61	<0,0001
12	ragazza	30	32	35,48	<0,0001
13	padre	35	40	34,34	<0,0001
14	moglie	30	33	32,69	<0,0001
15	bambino	98	152	32,53	<0,0001
16	morire	45	57	32,38	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
17	buttare	23	23	31,98	<0,0001
18	amore	25	26	31,53	<0,0001
19	storia	62	87	31,19	<0,0001
20	casa	90	139	30,41	<0,0001
21	vita	102	162	30,23	<0,0001
22	mamma	24	25	30,15	<0,0001
23	divertire	20	20	27,79	<0,0001
24	cuore	38	48	27,55	<0,0001
25	paura	58	83	27	<0,0001
26	guardare	59	85	26,74	<0,0001
27	aspettare	56	80	26,23	<0,0001
28	occhio	25	28	25,91	<0,0001
29	strada	63	93	25,84	<0,0001
30	spettacolo	20	21	24,6	<0,0001
31	notte	30	37	23,44	<0,0001
32	quartiere	19	20	23,21	<0,0001
33	musicista	21	23	23,12	<0,0001
34	bellezza	16	16	22,22	<0,0001
35	porre	59	89	22,02	<0,0001
36	dolore	27	33	21,68	<0,0001
37	morta	23	27	20,84	<0,0001
38	piangere	15	15	20,83	<0,0001
39	papà	15	15	20,83	<0,0001
40	sacco	15	15	20,83	<0,0001
41	riuscire	69	109	20,82	<0,0001
42	capire	70	111	20,74	<0,0001
43	violenza	36	49	20,18	<0,0001
44	vivo	30	39	19,74	<0,0001
45	pensare	156	286	19,6	<0,0001
46	cercare	62	97	19,57	<0,0001
47	incredibile	16	17	19,07	<0,0001
48	emozione	18	20	19,05	<0,0001
49	piacere	33	45	18,36	<0,0001
50	pezzo	24	30	17,95	<0,0001
51	ricordo	27	35	17,93	<0,0001
52	fortuna	19	22	17,91	<0,0001
53	dormire	17	19	17,7	<0,0001
54	pensiero	17	19	17,7	<0,0001
55	libro	15	16	17,7	<0,0001
56	saio	15	16	17,7	<0,0001
57	brutto	15	16	17,7	<0,0001
58	arte	22	27	17,43	<0,0001
59	proprio	148	274	17,2	<0,0001
60	signore	20	24	16,97	<0,0001
62	tranquillo	26	34	16,76	<0,0001
63	meraviglioso	12	12	16,65	<0,0001
64	gioia	12	12	16,65	<0,0001
65	calcio	12	12	16,65	<0,0001
66	bellissima	12	12	16,65	<0,0001
67	bravo	18	21	16,59	<0,0001
68	albergo	14	15	16,32	<0,0001
69	bellissimo	14	15	16,32	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
70	arrivare	119	216	15,93	<0,0001
71	mangiare	19	23	15,69	<0,0001
72	Fellini	11	11	15,26	<0,0001
73	anima	11	11	15,26	<0,0001
74	saltare	11	11	15,26	<0,0001
75	silenzio	15	17	15,02	0,0001
76	gesto	15	17	15,02	0,0001
77	sorpresa	13	14	14,95	0,00011
78	sorella	13	14	14,95	0,00011
79	Luca	13	14	14,95	0,00011
80	raccontare	28	39	14,38	0,00014
81	sogno	21	27	14,32	0,00015
82	entrare	53	86	13,9	0,00019
83	coda	10	10	13,87	0,00019
84	cielo	10	10	13,87	0,00019
85	luce	14	16	13,68	0,00021
86	capitare	14	16	13,68	0,00021
87	personaggio	12	13	13,58	0,00022
88	accorgersi	12	13	13,58	0,00022
89	trovare	99	179	13,55	0,00023
90	secondo	71	122	13,53	0,00023
91	genitore	27	38	13,32	0,00026
92	amico	33	49	13,11	0,00029
93	certo	68	117	12,84	0,00033
94	maestro	15	18	12,71	0,00036
95	ammazzare	15	18	12,71	0,00036
96	macchina	36	55	12,62	0,00038
97	canzone	9	9	12,49	0,00041
98	fantastico	9	9	12,49	0,00041
99	cucina	9	9	12,49	0,00041
100	urlare	9	9	12,49	0,00041
50	pezzo	24	30	17,95	<0,0001
51	ricordo	27	35	17,93	<0,0001
52	fortuna	19	22	17,91	<0,0001
53	dormire	17	19	17,7	<0,0001
54	pensiero	17	19	17,7	<0,0001
55	libro	15	16	17,7	<0,0001
56	saio	15	16	17,7	<0,0001
57	brutto	15	16	17,7	<0,0001
58	arte	22	27	17,43	<0,0001
59	proprio	148	274	17,2	<0,0001
60	signore	20	24	16,97	<0,0001
62	tranquillo	26	34	16,76	<0,0001
63	meraviglioso	12	12	16,65	<0,0001
64	gioia	12	12	16,65	<0,0001
65	calcio	12	12	16,65	<0,0001
66	bellissima	12	12	16,65	<0,0001
67	bravo	18	21	16,59	<0,0001
68	albergo	14	15	16,32	<0,0001
69	bellissimo	14	15	16,32	<0,0001
70	arrivare	119	216	15,93	<0,0001
71	mangiare	19	23	15,69	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
72	Fellini	11	11	15,26	<0,0001
73	anima	11	11	15,26	<0,0001
74	saltare	11	11	15,26	<0,0001
75	silenzio	15	17	15,02	0,0001
76	gesto	15	17	15,02	0,0001
77	sorpresa	13	14	14,95	0,00011
78	sorella	13	14	14,95	0,00011
79	Luca	13	14	14,95	0,00011
80	raccontare	28	39	14,38	0,00014
81	sogno	21	27	14,32	0,00015
82	entrare	53	86	13,9	0,00019
83	coda	10	10	13,87	0,00019
84	cielo	10	10	13,87	0,00019
85	luce	14	16	13,68	0,00021
86	capitare	14	16	13,68	0,00021
87	personaggio	12	13	13,58	0,00022
88	accorgersi	12	13	13,58	0,00022
89	trovare	99	179	13,55	0,00023
90	secondo	71	122	13,53	0,00023
91	genitore	27	38	13,32	0,00026
92	amico	33	49	13,11	0,00029
93	certo	68	117	12,84	0,00033
94	maestro	15	18	12,71	0,00036
95	ammazzare	15	18	12,71	0,00036
96	macchina	36	55	12,62	0,00038
97	canzone	9	9	12,49	0,00041
98	fantastico	9	9	12,49	0,00041
99	cucina	9	9	12,49	0,00041
100	urlare	9	9	12,49	0,00041
101	tifoso	9	9	12,49	0,00041
102	puro	9	9	12,49	0,00041
103	cinema	13	15	12,36	0,00043
104	compagno	11	12	12,22	0,00047
105	negozio	11	12	12,22	0,00047
106	me_lo	11	12	12,22	0,00047
107	guardo	11	12	12,22	0,00047
108	cominciare	42	67	11,99	0,00053
109	aiuto	23	32	11,85	0,00057
110	dimenticare	17	22	11,33	0,00076
111	festival	8	8	11,1	0,00086
112	perfetto	8	8	11,1	0,00086
113	Marco	8	8	11,1	0,00086
114	bar	8	8	11,1	0,00086
115	mo	8	8	11,1	0,00086
116	camion	8	8	11,1	0,00086
117	supermercato	8	8	11,1	0,00086
118	soddisfazione	8	8	11,1	0,00086
119	significato	8	8	11,1	0,00086
120	scarpa	8	8	11,1	0,00086
121	saldo	8	8	11,1	0,00086
122	regista	8	8	11,1	0,00086
123	insegnare	8	8	11,1	0,00086

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
124	gridare	8	8	11,1	0,00086
125	pochissimo	10	11	10,86	0,00098
126	pizza	10	11	10,86	0,00098
127	madre	10	11	10,86	0,00098
128	arresto	10	11	10,86	0,00098
129	spalla	18	24	10,83	0,001
130	domanda	22	31	10,79	0,00101
131	porta	36	57	10,65	0,00109
132	duro	23	33	10,51	0,00118
133	minuto	19	26	10,41	0,00125
134	lasciare	47	79	10,14	0,00144
135	corpo	20	28	10,06	0,00151
136	iniziare	34	54	9,9	0,00165
137	cattivo	11	13	9,75	0,00179
138	ricominciare	11	13	9,75	0,00179
139	odore	7	7	9,71	0,00183
140	danza	7	7	9,71	0,00183
141	botto	7	7	9,71	0,00183
142	suonare	7	7	9,71	0,00183
143	sorriso	7	7	9,71	0,00183
144	schiaffo	7	7	9,71	0,00183
145	montagna	7	7	9,71	0,00183
146	menta	7	7	9,71	0,00183
147	incominciare	7	7	9,71	0,00183
148	deragliare	7	7	9,71	0,00183
149	citare	7	7	9,71	0,00183
150	Alpi	7	7	9,71	0,00183
151	tirare	17	23	9,7	0,00184
152	contento	17	23	9,7	0,00184
153	chiudere	39	64	9,6	0,00194
154	treno	14	18	9,53	0,00202
155	sorta	14	18	9,53	0,00202
156	eccetera	9	10	9,51	0,00204
157	legno	9	10	9,51	0,00204
158	caffè	9	10	9,51	0,00204
159	tentare	9	10	9,51	0,00204
160	scoppiare	9	10	9,51	0,00204
161	muro	9	10	9,51	0,00204
162	bruciare	9	10	9,51	0,00204
163	mattina	29	45	9,44	0,00212
164	neve	23	34	9,29	0,0023
165	gente	63	113	9,04	0,00264
166	cercare	15	20	9,02	0,00267
167	stasera	12	15	8,95	0,00277
168	pulire	12	15	8,95	0,00277
169	stamattina	12	15	8,95	0,00277
170	buono	38	63	8,84	0,00294
171	provare	26	40	8,8	0,003
172	festeggiare	10	12	8,47	0,00361
173	violento	10	12	8,47	0,00361
174	parente	10	12	8,47	0,00361
175	festa	13	17	8,36	0,00384

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
176	uccidere	13	17	8,36	0,00384
177	colpo	13	17	8,36	0,00384
178	ricordare	31	50	8,34	0,00387
179	coppia	6	6	8,32	0,00392
180	racconto	6	6	8,32	0,00392
181	finestrino	6	6	8,32	0,00392
182	Andrea	6	6	8,32	0,00392
183	sognare	6	6	8,32	0,00392
184	ridere	6	6	8,32	0,00392
185	piatto	6	6	8,32	0,00392
186	concerto	6	6	8,32	0,00392
187	vabbè	6	6	8,32	0,00392
188	splendido	6	6	8,32	0,00392
189	sguardo	6	6	8,32	0,00392
190	musicale	6	6	8,32	0,00392
191	impressione	6	6	8,32	0,00392
192	ideale	6	6	8,32	0,00392
193	fretta	6	6	8,32	0,00392
194	felicità	6	6	8,32	0,00392
195	ergastolo	6	6	8,32	0,00392
196	educare	6	6	8,32	0,00392
197	corridoio	6	6	8,32	0,00392
198	cadavere	6	6	8,32	0,00392
199	bimbo	6	6	8,32	0,00392
200	bastone	6	6	8,32	0,00392
201	arma	6	6	8,32	0,00392
202	angoscia	6	6	8,32	0,00392
203	Daniele	6	6	8,32	0,00392
204	sparare	8	9	8,16	0,00427
205	turista	8	9	8,16	0,00427
206	finestra	8	9	8,16	0,00427
207	ricostruzione	8	9	8,16	0,00427
208	pietra	8	9	8,16	0,00427
209	accompagnare	8	9	8,16	0,00427
210	sconto	8	9	8,16	0,00427
211	passione	8	9	8,16	0,00427
212	imparare	8	9	8,16	0,00427
213	familiare	14	19	7,9	0,00495
214	anno	211	437	7,8	0,00521
215	stazione	11	14	7,74	0,0054
216	carabiniere	15	21	7,53	0,00605
217	viaggio	12	16	7,21	0,00725
218	ce_la	12	16	7,21	0,00725
219	abbracciare	12	16	7,21	0,00725
220	chiesa	12	16	7,21	0,00725
221	attore	12	16	7,21	0,00725
222	cantare	9	11	7,2	0,00727
223	video	9	11	7,2	0,00727
224	professore	9	11	7,2	0,00727
225	freddo	17	25	7,01	0,0081
226	puntata	5	5	6,93	0,00846
227	artista	5	5	6,93	0,00846

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
228	trasferta	5	5	6,93	0,00846
229	interessante	5	5	6,93	0,00846
230	fastidio	5	5	6,93	0,00846
231	coltello	5	5	6,93	0,00846
232	battezzare	5	5	6,93	0,00846
233	Celentano	5	5	6,93	0,00846
234	Papa	7	8	6,83	0,00894
235	sopravvivere	7	8	6,83	0,00894
236	avventura	7	8	6,83	0,00894
237	accendere	7	8	6,83	0,00894
238	telefonare	7	8	6,83	0,00894
239	tantissimo	7	8	6,83	0,00894
240	scena	7	8	6,83	0,00894
241	preghiera	7	8	6,83	0,00894
242	mostrare	7	8	6,83	0,00894
243	han	7	8	6,83	0,00894
244	frase	7	8	6,83	0,00894
245	caro	7	8	6,83	0,00894
246	passare	30	50	6,76	0,0093
247	bastare	31	52	6,74	0,00942
248	ritornare	19	29	6,66	0,00985
249	Luigi	10	13	6,55	0,01046
251	epoca	14	20	6,5	0,01081
252	senso	47	85	6,33	0,01189
253	salire	11	15	6,09	0,01357
254	fumo	11	15	6,09	0,01357
255	partita	16	24	6,06	0,01384
256	moda	8	10	5,96	0,0146
257	strano	8	10	5,96	0,0146
258	poliziotto	8	10	5,96	0,0146
259	nonno	8	10	5,96	0,0146
260	attimo	8	10	5,96	0,0146
261	collega	17	26	5,9	0,01512
262	gioco	18	28	5,78	0,01624
263	svegliare	12	17	5,75	0,01646
264	talento	4	4	5,54	0,01853
265	Vincenzo	4	4	5,54	0,01853
266	monumento	4	4	5,54	0,01853
267	tornare	39	70	5,53	0,0187
268	vetro	6	7	5,52	0,0188
269	foto	6	7	5,52	0,0188
270	sinceramente	6	7	5,52	0,0188
271	sciare	6	7	5,52	0,0188
272	sala	6	7	5,52	0,0188
273	ferro	6	7	5,52	0,0188
274	fantasia	6	7	5,52	0,0188
275	divertente	6	7	5,52	0,0188
276	bagno	6	7	5,52	0,0188
277	arrabbiare	6	7	5,52	0,0188
278	luogo	23	38	5,43	0,01975
279	galleria	9	12	5,4	0,02011
280	vestire	9	12	5,4	0,02011

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
281	abitare	9	12	5,4	0,02011
282	carcere	14	21	5,3	0,02135
283	scuola	42	77	5,1	0,02392
284	sperare	42	77	5,1	0,02392
285	scappare	16	25	5,02	0,02499
286	tenere	38	69	4,95	0,02603
287	occasione	17	27	4,93	0,02639
288	pregare	7	9	4,76	0,02915
289	identità	7	9	4,76	0,02915
290	cifra	7	9	4,76	0,02915
291	spiacere	7	9	4,76	0,02915
292	Parigi	7	9	4,76	0,02915
293	comprendere	11	16	4,74	0,02945
294	fermo	11	16	4,74	0,02945
295	specie	11	16	4,74	0,02945
296	rivedere	12	18	4,54	0,03313
297	impossibile	12	18	4,54	0,03313
298	negare	8	11	4,29	0,03824
299	assassino	8	11	4,29	0,03824
300	albero	8	11	4,29	0,03824
301	modalità	8	11	4,29	0,03824
302	finire	33	60	4,25	0,03915
303	ospite	5	6	4,23	0,03972
304	magico	5	6	4,23	0,03972
305	triste	5	6	4,23	0,03972
306	spendere	5	6	4,23	0,03972
307	spegnere	5	6	4,23	0,03972
308	riparare	5	6	4,23	0,03972
309	guaio	5	6	4,23	0,03972
310	fotografia	5	6	4,23	0,03972
311	discarica	5	6	4,23	0,03972
312	abbattere	5	6	4,23	0,03972
313	Cesare	5	6	4,23	0,03972
314	senno	5	6	4,23	0,03972
315	sasso	5	6	4,23	0,03972
316	sala	5	6	4,23	0,03972
317	piovere	5	6	4,23	0,03972
318	matrimonio	5	6	4,23	0,03972
319	interpretare	5	6	4,23	0,03972
320	indifferenza	5	6	4,23	0,03972
321	furgone	5	6	4,23	0,03972
322	educazione	5	6	4,23	0,03972
323	compagnia	5	6	4,23	0,03972
324	chitarra	5	6	4,23	0,03972
325	cancello	5	6	4,23	0,03972
326	battuto	5	6	4,23	0,03972
327	ahimé	5	6	4,23	0,03972
328	abbraccio	5	6	4,23	0,03972
329	Natale	5	6	4,23	0,03972
330	vittima	32	58	4,22	0,03994
331	chiamare	43	81	4,2	0,04044
332	notizia	15	24	4,19	0,04068

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
333	bambù	3	3	4,16	0,04144
334	Sanremo	3	3	4,16	0,04144
335	portafoglio	3	3	4,16	0,04144
336	stile	3	3	4,16	0,04144
337	faccia	29	52	4,13	0,04217
338	memoria	17	28	4,08	0,04343
339	immaginare	17	28	4,08	0,04343
340	girare	17	28	4,08	0,04343
341	resto	17	28	4,08	0,04343
342	mondo	67	133	3,99	0,04586
343	sport	9	13	3,99	0,04578
344	telefono	9	13	3,99	0,04578
345	fatica	9	13	3,99	0,04578
346	paio	9	13	3,99	0,04578
347	abituare	9	13	3,99	0,04578

Tab. 6 Forme lessicali significative per la classe 2 (CGD) del *sub-corpus* PF_Uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	zona	35	44	112,76	<0,0001
1	virus	29	34	104,04	<0,0001
2	infezione	20	20	90,72	<0,0001
3	Cina	24	29	82,1	<0,0001
4	situazione	52	100	78,69	<0,0001
5	paziente	21	24	78,22	<0,0001
6	medico	22	26	77,86	<0,0001
7	sintomo	17	17	77,08	<0,0001
8	evento	21	25	73,47	<0,0001
9	ricerca	22	27	73,43	<0,0001
10	animale	19	22	69,35	<0,0001
11	materiale	19	22	69,35	<0,0001
12	area	20	24	69,1	<0,0001
13	cinese	23	31	66,07	<0,0001
14	esempio	40	74	65,23	<0,0001
15	malattia	19	23	64,75	<0,0001
16	epidemia	15	16	61,88	<0,0001
17	coronavirus	15	16	61,88	<0,0001
18	temperatura	16	18	60,97	<0,0001
19	contatto	17	20	60,53	<0,0001
20	controllo	30	51	57,47	<0,0001
21	influenza	16	19	56,14	<0,0001
22	identificare	15	18	51,78	<0,0001
23	rischio	37	74	51,35	<0,0001
24	causa	19	27	49,93	<0,0001
25	respiratorio	11	11	49,82	<0,0001
26	sci	11	11	49,82	<0,0001
27	costante	11	11	49,82	<0,0001
28	febbre	12	13	48,36	<0,0001
29	terremoto	15	19	47,56	<0,0001
30	utilizzare	27	49	45,56	<0,0001
31	antibiotico	10	10	45,29	<0,0001
32	scorso	19	29	44,15	<0,0001
33	malato	11	12	43,86	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
34	produrre	28	53	43,44	<0,0001
35	incendio	12	14	43,24	<0,0001
36	tecnologia	13	16	43,12	<0,0001
37	grosso	20	32	42,73	<0,0001
38	farmaco	9	9	40,75	<0,0001
39	computer	9	9	40,75	<0,0001
40	terapia	10	11	39,38	<0,0001
41	soggetto	19	31	39,17	<0,0001
42	popolazione	13	17	39,15	<0,0001
43	uso	13	17	39,15	<0,0001
44	eventuale	13	17	39,15	<0,0001
45	trasmissione	11	13	38,85	<0,0001
46	operatore	11	13	38,85	<0,0001
47	climatico	11	13	38,85	<0,0001
48	vaccinare	12	15	38,83	<0,0001
49	nord	12	15	38,83	<0,0001
50	pista	16	24	38,3	<0,0001
51	struttura	15	22	37,32	<0,0001
52	trasmettere	14	20	36,43	<0,0001
53	focolaio	8	8	36,22	<0,0001
54	atmosfera	8	8	36,22	<0,0001
55	settimana	31	67	36,2	<0,0001
56	tipo	34	77	35,66	<0,0001
57	informazione	16	25	35,64	<0,0001
58	tessuto	12	16	35	<0,0001
59	sicuramente	41	101	34,99	<0,0001
60	avvenire	24	47	34,67	<0,0001
61	sequestro	10	12	34,49	<0,0001
62	periodo	17	28	34,42	<0,0001
63	pericoloso	17	29	32,24	<0,0001
64	faglia	7	7	31,69	<0,0001
65	tosse	7	7	31,69	<0,0001
66	riscaldamento	7	7	31,69	<0,0001
67	immunitario	7	7	31,69	<0,0001
68	caduta	7	7	31,69	<0,0001
69	Wuhan	7	7	31,69	<0,0001
70	prodotto	12	17	31,65	<0,0001
71	ospedale	14	22	30,85	<0,0001
72	numero	26	56	30,55	<0,0001
73	laboratorio	8	9	30,44	<0,0001
74	clinico	8	9	30,44	<0,0001
75	ghiaccio	8	9	30,44	<0,0001
76	classico	8	9	30,44	<0,0001
77	volontario	8	9	30,44	<0,0001
78	riservare	8	9	30,44	<0,0001
79	effettuare	8	9	30,44	<0,0001
80	fornire	10	13	30,38	<0,0001
81	fenomeno	17	30	30,21	<0,0001
82	mezza	9	11	30,16	<0,0001
83	sostanza	9	11	30,16	<0,0001
84	indicare	9	11	30,16	<0,0001
85	caldo	11	16	27,74	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
86	rete	13	21	27,25	<0,0001
88	proteggere	13	21	27,25	<0,0001
89	distanza	13	21	27,25	<0,0001
90	anziano	13	21	27,25	<0,0001
91	recuperare	13	21	27,25	<0,0001
92	proteina	6	6	27,16	<0,0001
93	sensorio	6	6	27,16	<0,0001
94	polvere	6	6	27,16	<0,0001
95	tumore	6	6	27,16	<0,0001
96	ASL	6	6	27,16	<0,0001
97	prognosi	6	6	27,16	<0,0001
98	lentamente	6	6	27,16	<0,0001
99	dispositivo	6	6	27,16	<0,0001
100	coordinamento	6	6	27,16	<0,0001
101	velocità	10	14	26,9	<0,0001
102	evitare	27	63	26,28	<0,0001
103	metro	14	24	26,28	<0,0001
104	contenere	9	12	26,24	<0,0001
105	disagio	12	19	26,08	<0,0001
106	sciatore	7	8	26,01	<0,0001
107	provocare	7	8	26,01	<0,0001
108	cervello	7	8	26,01	<0,0001
109	veicolo	8	10	25,87	<0,0001
110	veloce	8	10	25,87	<0,0001
111	limite	8	10	25,87	<0,0001
112	fisico	8	10	25,87	<0,0001
113	possibilità	27	64	25,27	<0,0001
114	stagione	13	22	25	<0,0001
115	impianto	11	17	24,95	<0,0001
116	elevato	10	15	23,9	<0,0001
117	episodio	10	15	23,9	<0,0001
118	effetto	15	28	23,85	<0,0001
119	presente	19	40	23,44	<0,0001
120	verificare	17	34	23,44	<0,0001
121	salute	23	53	23,04	<0,0001
122	fiume	9	13	22,95	<0,0001
123	sottile	5	5	22,63	<0,0001
124	robot	5	5	22,63	<0,0001
125	termico	5	5	22,63	<0,0001
126	allerta	5	5	22,63	<0,0001
127	disposizione	14	26	22,48	<0,0001
128	provincia	8	11	22,16	<0,0001
129	interno	37	104	21,75	<0,0001
130	cura	12	21	21,64	<0,0001
131	eccessivo	7	9	21,64	<0,0001
132	mascherina	7	9	21,64	<0,0001
133	apertura	7	9	21,64	<0,0001
134	altissimo	7	9	21,64	<0,0001
135	importazione	6	7	21,6	<0,0001
136	impiegare	6	7	21,6	<0,0001
137	formare	6	7	21,6	<0,0001
138	prontamente	6	7	21,6	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
139	profondità	6	7	21,6	<0,0001
140	fragile	6	7	21,6	<0,0001
141	comparire	6	7	21,6	<0,0001
142	diffondere	10	16	21,31	<0,0001
143	territorio	26	65	21,22	<0,0001
144	produzione	13	24	21,12	<0,0001
145	acqua	18	39	20,81	<0,0001
146	occupare	17	36	20,69	<0,0001
147	attività	23	56	20,09	<0,0001
148	massimo	16	34	19,32	<0,0001
149	preoccupare	16	34	19,32	<0,0001
150	basso	14	28	19,28	<0,0001
151	termine	15	31	19,26	<0,0001
152	attenzione	23	57	19,18	<0,0001
153	perdita	8	12	19,11	<0,0001
154	spostare	8	12	19,11	<0,0001
155	caratteristica	8	12	19,11	<0,0001
156	personalmente	10	17	19,05	<0,0001
157	viaggiare	7	10	18,18	<0,0001
158	Ostia	7	10	18,18	<0,0001
159	elettronico	7	10	18,18	<0,0001
160	roccia	4	4	18,1	<0,0001
161	velluto	4	4	18,1	<0,0001
162	interrogare	4	4	18,1	<0,0001
163	fase	16	35	18,08	<0,0001
164	dialogo	9	15	17,79	<0,0001
165	superficie	6	8	17,48	<0,0001
166	suolo	6	8	17,48	<0,0001
167	precauzione	6	8	17,48	<0,0001
168	arco	6	8	17,48	<0,0001
169	telecamera	6	8	17,48	<0,0001
170	pregiudizio	6	8	17,48	<0,0001
171	associazione	6	8	17,48	<0,0001
172	Spallanzani	6	8	17,48	<0,0001
173	Spada	6	8	17,48	<0,0001
174	nebbia	5	6	17,23	<0,0001
175	sciogliere	5	6	17,23	<0,0001
176	connettere	5	6	17,23	<0,0001
177	ricorrere	5	6	17,23	<0,0001
178	procedura	5	6	17,23	<0,0001
179	primavera	5	6	17,23	<0,0001
180	notevole	5	6	17,23	<0,0001
181	comando	5	6	17,23	<0,0001
182	bottiglia	5	6	17,23	<0,0001
183	allarmismo	5	6	17,23	<0,0001
184	gestire	14	30	16,57	<0,0001
185	sanitario	14	30	16,57	<0,0001
186	valanga	8	13	16,57	<0,0001
187	sito	8	13	16,57	<0,0001
188	specifico	8	13	16,57	<0,0001
189	sviluppare	12	24	16,52	<0,0001
190	curare	7	11	15,4	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
191	elettrico	7	11	15,4	<0,0001
192	scientifico	7	11	15,4	<0,0001
193	principalmente	7	11	15,4	<0,0001
194	diffusione	7	11	15,4	<0,0001
195	chilometro	10	19	15,31	<0,0001
196	controllare	13	28	15,21	<0,0001
197	forma	11	22	15,14	<0,0001
198	trattare	20	52	14,64	0,00012
199	livello	27	78	14,51	0,00013
200	tecnica	8	14	14,41	0,00014
201	richiesta	8	14	14,41	0,00014
202	elicottero	6	9	14,33	0,00015
203	applicazione	6	9	14,33	0,00015
204	stretto	6	9	14,33	0,00015
205	incidente	14	32	14,26	0,00015
206	piccolo	22	60	14,06	0,00017
207	motore	9	17	13,94	0,00018
208	fonte	9	17	13,94	0,00018
209	ulteriore	11	23	13,74	0,0002
210	ricevere	11	23	13,74	0,0002
211	speciale	5	7	13,43	0,00024
212	stupefacente	5	7	13,43	0,00024
213	rapidamente	5	7	13,43	0,00024
214	organico	5	7	13,43	0,00024
215	visitatore	5	7	13,43	0,00024
216	ripristinare	5	7	13,43	0,00024
217	rallentare	5	7	13,43	0,00024
218	orario	5	7	13,43	0,00024
219	completo	5	7	13,43	0,00024
220	circolare	5	7	13,43	0,00024
221	capannone	5	7	13,43	0,00024
222	antico	5	7	13,43	0,00024
223	ansia	5	7	13,43	0,00024
224	accumulare	5	7	13,43	0,00024
225	lungo	17	43	13,4	0,00025
226	camminare	7	12	13,12	0,00029
227	utilizzo	7	12	13,12	0,00029
228	naturale	7	12	13,12	0,00029
229	Roma	21	58	12,93	0,00032
230	corso	13	30	12,93	0,00032
231	cavo	4	5	12,92	0,00032
232	settembre	4	5	12,92	0,00032
233	psicologico	4	5	12,92	0,00032
234	dinamico	4	5	12,92	0,00032
235	abuso	4	5	12,92	0,00032
236	abbattimento	4	5	12,92	0,00032
237	condizione	29	89	12,75	0,00035
238	energia	12	27	12,68	0,00036
239	ambulanza	8	15	12,58	0,00039
240	estremo	8	15	12,58	0,00039
241	gennaio	9	18	12,38	0,00043
242	usare	19	52	12,01	0,00053

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
243	colore	6	10	11,85	0,00057
244	attacco	6	10	11,85	0,00057
245	inquinare	6	10	11,85	0,00057
246	informare	6	10	11,85	0,00057
247	esporre	6	10	11,85	0,00057
248	consentire	17	45	11,82	0,00058
249	vaccino	11	25	11,34	0,00075
250	indagine	11	25	11,34	0,00075
251	sufficiente	7	13	11,22	0,00081
252	globale	7	13	11,22	0,00081
253	maniera	21	61	11,06	0,00088
254	autorità	9	19	11	0,00091
255	indagare	8	16	11	0,00091
256	escludere	8	16	11	0,00091
257	attualmente	8	16	11	0,00091
258	presenza	13	32	10,99	0,00091
259	ufficiale	5	8	10,64	0,0011
260	sportivo	5	8	10,64	0,0011
261	sano	5	8	10,64	0,0011
262	professionale	5	8	10,64	0,0011
263	precedenza	5	8	10,64	0,0011
264	origine	5	8	10,64	0,0011
265	drammatico	5	8	10,64	0,0011
266	dito	5	8	10,64	0,0011
267	considerazione	5	8	10,64	0,0011
268	avanzare	5	8	10,64	0,0011
269	immagine	10	23	10,02	0,00155
270	totale	10	23	10,02	0,00155
271	intervenire	16	44	9,95	0,0016
272	alto	23	71	9,88	0,00167
273	unità	6	11	9,86	0,00168
274	fiamma	6	11	9,86	0,00168
275	prevenzione	6	11	9,86	0,00168
276	normalmente	6	11	9,86	0,00168
277	reale	9	20	9,78	0,00176
278	cambiamento	12	30	9,74	0,0018
279	procura	8	17	9,63	0,00191
280	preciso	7	14	9,62	0,00192
281	procedere	7	14	9,62	0,00192
282	comportare	7	14	9,62	0,00192
283	rilievo	4	6	9,55	0,002
284	pianta	4	6	9,55	0,002
285	offerta	4	6	9,55	0,002
286	lago	4	6	9,55	0,002
287	verdura	4	6	9,55	0,002
288	sismico	4	6	9,55	0,002
289	mostra	4	6	9,55	0,002
290	merce	4	6	9,55	0,002
291	inferiore	4	6	9,55	0,002
292	indico	4	6	9,55	0,002
293	importantissimo	4	6	9,55	0,002
294	effettivo	4	6	9,55	0,002

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
295	edificio	4	6	9,55	0,002
296	documentare	4	6	9,55	0,002
297	confine	4	6	9,55	0,002
298	comandare	4	6	9,55	0,002
299	circoscrivere	4	6	9,55	0,002
300	ciclo	4	6	9,55	0,002
301	agricolo	4	6	9,55	0,002
302	mano	15	41	9,5	0,00205
303	particolare	23	72	9,4	0,00217
304	importante	43	158	9,06	0,00261
305	danno	18	53	9,06	0,00261
306	oggetto	10	24	9,01	0,00267
307	muovere	10	24	9,01	0,00267
308	intervento	15	42	8,83	0,00295
309	filtro	3	4	8,73	0,00312
310	fibra	3	4	8,73	0,00312
311	spazio	9	21	8,7	0,00318
312	aereo	9	21	8,7	0,00318
313	terra	16	46	8,68	0,0032
314	dato	13	35	8,6	0,00336
315	traffico	13	35	8,6	0,00336
316	riciclare	5	9	8,52	0,00351
317	turismo	5	9	8,52	0,00351
318	preparazione	5	9	8,52	0,00351
319	clan	5	9	8,52	0,00351
320	chilo	5	9	8,52	0,00351
321	aria	11	28	8,5	0,00354
322	criminale	8	18	8,44	0,00366
323	sottoporre	7	15	8,27	0,00403
324	complesso	7	15	8,27	0,00403
325	collegare	7	15	8,27	0,00403
326	noto	6	12	8,24	0,00408
327	quantità	6	12	8,24	0,00408
328	etico	6	12	8,24	0,00408
329	denunciare	6	12	8,24	0,00408
330	accesso	6	12	8,24	0,00408
331	medio	12	32	8,15	0,00429
332	studiare	10	25	8,11	0,0044
333	momento	47	181	7,76	0,00535
334	Milano	9	22	7,74	0,0054
335	particolarmente	9	22	7,74	0,0054
336	esatto	4	7	7,2	0,00729
337	condurre	4	7	7,2	0,00729
338	cellula	4	7	7,2	0,00729
339	biglietto	4	7	7,2	0,00729
340	segnalazione	4	7	7,2	0,00729
341	quotidianamente	4	7	7,2	0,00729
342	ambito	4	7	7,2	0,00729
343	ritardo	4	7	7,2	0,00729
344	notare	4	7	7,2	0,00729
345	massa	4	7	7,2	0,00729
346	ingresso	4	7	7,2	0,00729

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
347	funzionamento	4	7	7,2	0,00729
348	evoluzione	4	7	7,2	0,00729
349	divieto	4	7	7,2	0,00729
350	avviare	4	7	7,2	0,00729
351	adeguato	4	7	7,2	0,00729
352	genere	14	41	7,16	0,00747
353	plastica	7	16	7,11	0,00766
354	pesante	7	16	7,11	0,00766
355	indicazione	6	13	6,91	0,00858
356	atteggiamento	6	13	6,91	0,00858
357	sforzo	6	13	6,91	0,00858
358	ritorno	9	23	6,88	0,00873
359	cellulare	5	10	6,87	0,00877
360	sensazione	5	10	6,87	0,00877
361	trasferire	5	10	6,87	0,00877
362	scaricare	5	10	6,87	0,00877
363	inquinamento	5	10	6,87	0,00877
364	delicato	5	10	6,87	0,00877
365	binario	5	10	6,87	0,00877
366	ultimo	25	87	6,72	0,00955
367	strumento	11	31	6,34	0,0118
368	macchina	17	55	6,13	0,0133
369	esistere	17	55	6,13	0,0133
370	volo	7	17	6,11	0,0134
371	bisognare	38	147	6,08	0,01367
372	servizio	15	47	6,08	0,01363
373	implicare	3	5	5,92	0,01496
374	scopo	3	5	5,92	0,01496
375	emettere	3	5	5,92	0,01496
376	ordine	10	28	5,88	0,01534
377	cibo	6	14	5,79	0,01608
378	solamente	6	14	5,79	0,01608
379	funzione	8	21	5,67	0,01721
380	conseguenza	8	21	5,67	0,01721
381	rosso	5	11	5,56	0,01842
382	lontano	5	11	5,56	0,01842
383	omicidio	5	11	5,56	0,01842
384	ndrangheta	5	11	5,56	0,01842
385	meccanismo	5	11	5,56	0,01842
386	legare	5	11	5,56	0,01842
387	eventualmente	5	11	5,56	0,01842
388	contesto	5	11	5,56	0,01842
389	eroe	4	8	5,49	0,01909
390	estremamente	4	8	5,49	0,01909
391	vietare	4	8	5,49	0,01909
392	planeta	4	8	5,49	0,01909
393	istituto	4	8	5,49	0,01909
394	fede	4	8	5,49	0,01909
395	verifica	4	8	5,49	0,01909
396	vento	4	8	5,49	0,01909
397	provenire	4	8	5,49	0,01909
398	desiderare	4	8	5,49	0,01909

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
399	concittadino	4	8	5,49	0,01909
400	scenario	7	18	5,25	0,02188
401	impatto	7	18	5,25	0,02188
402	attivo	7	18	5,25	0,02188
403	dimensione	8	22	4,96	0,02594
404	Napoli	6	15	4,86	0,02752
405	ottimo	6	15	4,86	0,02752
406	operazione	6	15	4,86	0,02752
407	inverno	6	15	4,86	0,02752
408	preoccupazione	6	15	4,86	0,02752
409	prevedere	15	50	4,8	0,02839
410	disponibile	9	26	4,8	0,02853
411	neve	11	34	4,68	0,03059
412	mani	11	34	4,68	0,03059
413	complicare	5	12	4,5	0,03392
414	Dio	5	12	4,5	0,03392
415	<i>subire</i>	5	12	4,5	0,03392
416	prezzo	5	12	4,5	0,03392
417	lingua	4	9	4,21	0,04008
418	parco	4	9	4,21	0,04008
419	panico	4	9	4,21	0,04008
420	vario	4	9	4,21	0,04008
421	immediatamente	4	9	4,21	0,04008
422	frontiera	4	9	4,21	0,04008
423	estrarre	4	9	4,21	0,04008
424	sabbia	3	6	4,12	0,04242
425	trasferimento	3	6	4,12	0,04242
426	slittare	3	6	4,12	0,04242
427	bocca	3	6	4,12	0,04242
428	assistenza	3	6	4,12	0,04242
429	vigore	3	6	4,12	0,04242
430	scienziato	3	6	4,12	0,04242
431	rassicurare	3	6	4,12	0,04242
432	psicosi	3	6	4,12	0,04242
433	introduzione	3	6	4,12	0,04242
434	intelligente	3	6	4,12	0,04242
435	drammaticamente	3	6	4,12	0,04242
436	criminalità	3	6	4,12	0,04242
437	comodo	3	6	4,12	0,04242
438	DNA	3	6	4,12	0,04242
439	Chiesa	3	6	4,12	0,04242
440	rientrare	6	16	4,07	0,04375
441	leggero	6	16	4,07	0,04375
442	favorire	6	16	4,07	0,04375
443	diminuire	6	16	4,07	0,04375
444	concentrare	6	16	4,07	0,04375
445	comportamento	6	16	4,07	0,04375
446	direttamente	7	20	3,86	0,04948

Tab. 7 Forme lessicali significative per la classe 3 (CGD) del *sub-corpus* PF_Uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
0	governo	319	333	456,85	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	politico	258	283	323,81	<0,0001
2	Italia	347	438	302,46	<0,0001
3	paese	289	350	280,45	<0,0001
4	italiano	268	327	253,53	<0,0001
5	Europa	144	158	176,91	<0,0001
6	5_Stelle	115	116	172,47	<0,0001
7	Salvini	120	124	170,15	<0,0001
8	cittadino	141	168	139,06	<0,0001
9	tassa	87	89	125,65	<0,0001
10	ministro	83	86	116,2	<0,0001
11	presidente	84	88	114,54	<0,0001
12	elettorale	72	72	109,41	<0,0001
13	Movimento	74	75	109,02	<0,0001
14	forza	115	139	108,37	<0,0001
15	centro_destra	71	71	107,88	<0,0001
16	legge	86	94	105,59	<0,0001
17	lavoro	193	278	105,22	<0,0001
18	elezione	67	67	101,73	<0,0001
19	sicurezza	119	149	101,27	<0,0001
20	economico	75	81	94,73	<0,0001
21	europeo	84	96	91,81	<0,0001
22	diritto	80	90	91,09	<0,0001
23	Parlamento	60	61	87,55	<0,0001
24	maggioranza	58	60	81,18	<0,0001
25	impresa	68	76	78,59	<0,0001
26	programma	61	66	76,51	<0,0001
27	partito	68	77	75,95	<0,0001
28	Lega	50	50	75,71	<0,0001
29	proposta	59	64	73,5	<0,0001
30	fiscale	48	48	72,66	<0,0001
31	Partito_Democratico	48	48	72,66	<0,0001
32	votare	53	56	70,41	<0,0001
33	PD	48	49	69,23	<0,0001
34	riforma	47	48	67,71	<0,0001
35	sinistra	55	60	67,49	<0,0001
36	investimento	54	59	65,99	<0,0001
37	reddito	42	42	63,52	<0,0001
38	lavoratore	56	63	63,49	<0,0001
39	parlamentare	41	41	61,99	<0,0001
40	chiedere	118	171	61,79	<0,0001
41	immigrazione	45	47	61,41	<0,0001
42	accordo	45	47	61,41	<0,0001
43	sociale	65	79	59,7	<0,0001
44	Unione_Europea	38	38	57,43	<0,0001
45	Forza_Italia	37	37	55,91	<0,0001
46	cittadinanza	39	40	55,54	<0,0001
47	popolo	48	54	54,34	<0,0001
48	difendere	44	48	53,89	<0,0001
49	decreto	42	45	53,8	<0,0001
50	crescita	47	53	52,87	<0,0001
51	economia	43	47	52,4	<0,0001
52	voto	41	44	52,3	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
53	decidere	66	85	50,99	<0,0001
54	governare	36	37	50,99	<0,0001
55	questione	36	37	50,99	<0,0001
56	legge	38	40	50,82	<0,0001
57	sviluppo	40	43	50,8	<0,0001
58	vincere	46	53	48,85	<0,0001
59	regione	42	47	48,16	<0,0001
60	euro	62	80	47,58	<0,0001
61	cambiare	77	107	46,45	<0,0001
62	chiaro	58	74	46,05	<0,0001
63	responsabilità	52	64	45,93	<0,0001
64	convincere	42	48	45,55	<0,0001
65	discussione	30	30	45,28	<0,0001
66	Di_Maio	32	33	44,93	<0,0001
67	misura	39	44	43,75	<0,0001
68	soluzione	37	41	43,48	<0,0001
69	campagna	38	43	42,29	<0,0001
70	miliardo	28	28	42,25	<0,0001
71	provvedimento	30	31	41,9	<0,0001
72	nazionale	41	48	41,65	<0,0001
73	democrazia	33	36	40,34	<0,0001
74	risultato	43	53	37,76	<0,0001
75	Conte	25	25	37,7	<0,0001
76	Emilia	29	31	37,28	<0,0001
77	processo	54	72	37,26	<0,0001
78	prossimo	60	83	36,63	<0,0001
79	coalizione	24	24	36,19	<0,0001
80	manovra	26	27	35,86	<0,0001
81	pagare	58	80	35,75	<0,0001
82	tema	43	54	35,72	<0,0001
83	giustizia	60	84	35,13	<0,0001
84	dimostrare	38	46	35,12	<0,0001
85	Stato	44	56	35,1	<0,0001
86	discutere	23	23	34,68	<0,0001
88	Romagna	25	26	34,36	<0,0001
89	interesse	37	45	33,73	<0,0001
90	confine	24	25	32,85	<0,0001
91	Berlusconi	24	25	32,85	<0,0001
92	crisi	33	39	32,61	<0,0001
93	impegno	36	44	32,35	<0,0001
94	problema	109	181	31,87	<0,0001
95	concessione	21	21	31,65	<0,0001
96	Calabria	21	21	31,65	<0,0001
97	ridurre	35	43	30,97	<0,0001
98	pubblico	63	93	30,36	<0,0001
99	migrare	28	32	30,29	<0,0001
100	concreto	28	32	30,29	<0,0001
101	regola	40	52	29,85	<0,0001
102	ruolo	24	26	29,81	<0,0001
103	fiducia	32	39	28,97	<0,0001
104	garantire	35	44	28,93	<0,0001
105	credere	119	205	28,88	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
106	Francia	27	31	28,84	<0,0001
107	cancellare	27	31	28,84	<0,0001
108	alleanza	19	19	28,63	<0,0001
109	infrastruttura	19	19	28,63	<0,0001
110	repubblica	30	36	28,38	<0,0001
111	banca	21	22	28,33	<0,0001
112	Libia	21	22	28,33	<0,0001
113	tutelare	21	22	28,33	<0,0001
114	risolvere	40	53	28,09	<0,0001
115	posto	31	38	27,59	<0,0001
116	fondamentale	31	38	27,59	<0,0001
117	continuare	68	105	27,37	<0,0001
118	campo	42	57	27,26	<0,0001
119	approvare	18	18	27,12	<0,0001
120	vittoria	24	27	27,05	<0,0001
121	oggi	131	233	26,66	<0,0001
122	internazionale	46	65	25,98	<0,0001
123	agire	25	29	25,97	<0,0001
124	imporre	17	17	25,61	<0,0001
125	lista	28	34	25,6	<0,0001
126	polemica	23	26	25,59	<0,0001
127	impegnare	23	26	25,59	<0,0001
128	sindaco	31	39	25,55	<0,0001
129	contratto	19	20	25,33	<0,0001
130	raggiungere	32	41	24,95	<0,0001
131	progetto	29	36	24,86	<0,0001
132	rispettare	36	48	24,75	<0,0001
133	candidato	24	28	24,53	<0,0001
134	Mediterraneo	22	25	24,14	<0,0001
135	abolire	16	16	24,1	<0,0001
136	propaganda	16	16	24,1	<0,0001
137	processare	16	16	24,1	<0,0001
138	alleato	16	16	24,1	<0,0001
139	Bruxelles	16	16	24,1	<0,0001
140	ragione	34	45	23,93	<0,0001
141	trafficante	20	22	23,88	<0,0001
142	scelta	31	40	23,64	<0,0001
143	soldo	41	58	23,03	<0,0001
144	appello	29	37	22,89	<0,0001
145	Stati Uniti	21	24	22,69	<0,0001
146	promessa	15	15	22,59	<0,0001
147	liberale	15	15	22,59	<0,0001
148	coesione	15	15	22,59	<0,0001
149	bisogno	51	77	22,42	<0,0001
150	stabilità	19	21	22,41	<0,0001
151	bilancio	19	21	22,41	<0,0001
152	applicare	17	18	22,33	<0,0001
153	Senato	17	18	22,33	<0,0001
154	leader	17	18	22,33	<0,0001
155	clandestino	17	18	22,33	<0,0001
156	corte	17	18	22,33	<0,0001
157	alternativo	24	29	22,24	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
158	prospettiva	24	29	22,24	<0,0001
159	pensione	27	34	22,17	<0,0001
160	ONG	14	14	21,08	<0,0001
161	eleggere	14	14	21,08	<0,0001
162	occupazione	14	14	21,08	<0,0001
163	contributo	14	14	21,08	<0,0001
164	centro_sinistra	14	14	21,08	<0,0001
165	popolare	14	14	21,08	<0,0001
166	flusso	14	14	21,08	<0,0001
167	Zingaretti	14	14	21,08	<0,0001
168	destra	26	33	20,84	<0,0001
169	democratico	16	17	20,83	<0,0001
170	significare	47	71	20,6	<0,0001
171	mandare	30	40	20,59	<0,0001
172	sfida	21	25	20,29	<0,0001
173	investire	21	25	20,29	<0,0001
174	presentare	31	42	20,19	<0,0001
175	risorsa	24	30	20,14	<0,0001
176	obiettivo	36	51	20,1	<0,0001
177	Renzi	19	22	19,81	<0,0001
178	merito	19	22	19,81	<0,0001
179	incompetenza	13	13	19,57	<0,0001
180	salario	13	13	19,57	<0,0001
181	piazza	25	32	19,52	<0,0001
182	Trump	17	19	19,47	<0,0001
183	favore	17	19	19,47	<0,0001
184	debito	15	16	19,34	<0,0001
185	Fornero	15	16	19,34	<0,0001
186	porto	30	41	18,97	<0,0001
187	consiglio	30	41	18,97	<0,0001
188	prescrizione	20	24	18,89	<0,0001
189	stipendio	18	21	18,38	<0,0001
190	immigrato	18	21	18,38	<0,0001
191	crescere	32	45	18,33	<0,0001
192	tavolo	21	26	18,12	<0,0001
193	unire	21	26	18,12	<0,0001
194	costituzione	12	12	18,06	<0,0001
195	sbarcare	16	18	18,01	<0,0001
196	inaccettabile	16	18	18,01	<0,0001
197	benessere	16	18	18,01	<0,0001
198	avversario	14	15	17,85	<0,0001
199	libertà	25	33	17,71	<0,0001
200	assicurare	19	23	17,5	<0,0001
201	istituzione	30	42	17,45	<0,0001
202	intero	30	42	17,45	<0,0001
203	serio	26	35	17,28	<0,0001
204	scegliere	31	44	17,17	<0,0001
205	rafforzare	17	20	16,96	<0,0001
206	droga	15	17	16,55	<0,0001
207	scafista	11	11	16,55	<0,0001
208	contare	15	17	16,55	<0,0001
209	recessione	11	11	16,55	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
210	agenda	11	11	16,55	<0,0001
211	sbarco	11	11	16,55	<0,0001
212	ribadire	11	11	16,55	<0,0001
213	prevalere	11	11	16,55	<0,0001
214	pensionato	11	11	16,55	<0,0001
215	giungere	11	11	16,55	<0,0001
216	credibilità	11	11	16,55	<0,0001
217	complice	11	11	16,55	<0,0001
218	onore	13	14	16,36	<0,0001
219	patrimonio	13	14	16,36	<0,0001
220	aula	13	14	16,36	<0,0001
221	rappresentare	37	56	16,05	<0,0001
222	rischiare	25	34	16,05	<0,0001
223	assumere	25	34	16,05	<0,0001
224	futuro	67	116	15,6	<0,0001
225	rilanciare	16	19	15,55	<0,0001
226	arma	19	24	15,43	<0,0001
227	linea	23	31	15,21	<0,0001
228	imprenditore	14	16	15,11	0,0001
229	garanzia	14	16	15,11	0,0001
230	battere	14	16	15,11	0,0001
231	autostradale	10	10	15,05	0,0001
232	Grillo	10	10	15,05	0,0001
233	sottoscrivere	10	10	15,05	0,0001
234	vertice	10	10	15,05	0,0001
235	segretario	10	10	15,05	0,0001
236	incapace	10	10	15,05	0,0001
237	finanziario	10	10	15,05	0,0001
238	elettore	10	10	15,05	0,0001
239	Europeo	10	10	15,05	0,0001
240	Diciotti	10	10	15,05	0,0001
241	Camera	10	10	15,05	0,0001
242	valore	36	55	15,03	0,0001
243	libico	12	13	14,88	0,00011
244	taglio	12	13	14,88	0,00011
245	TAV	12	13	14,88	0,00011
246	costituzionale	12	13	14,88	0,00011
247	contrastare	12	13	14,88	0,00011
248	busta	12	13	14,88	0,00011
249	adottare	20	26	14,86	0,00011
250	restare	24	33		
251	lotta	24	33	14,84	0,00011
252	Lombardia	17	21	14,75	0,00012
253	estero	25	35	14,52	0,00013
254	elemento	21	28	14,39	0,00014
255	difesa	21	28	14,39	0,00014
256	prevedere	33	50	14,24	0,00016
257	referendum	15	18	14,15	0,00016
258	maggiore	34	52	14,12	0,00017
259	passo	34	52	14,12	0,00017
260	salvare	28	41	13,8	0,0002
261	guidare	13	15	13,67	0,00021

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
262	magistratura	13	15	13,67	0,00021
263	occorrere	19	25	13,58	0,00022
264	patto	9	9	13,54	0,00023
265	parlamento	9	9	13,54	0,00023
266	opposizione	9	9	13,54	0,00023
267	cessare	9	9	13,54	0,00023
268	serietà	9	9	13,54	0,00023
269	legittimo	9	9	13,54	0,00023
270	dimostrazione	9	9	13,54	0,00023
271	Spagna	9	9	13,54	0,00023
272	Fratelli_d	9	9	13,54	0,00023
273	decisione	30	45	13,47	0,00024
274	evidente	16	20	13,4	0,00025
275	tradire	11	12	13,39	0,00025
276	visione	11	12	13,39	0,00025
277	definitivo	11	12	13,39	0,00025
278	missione	20	27	13,15	0,00028
279	priorità	20	27	13,15	0,00028
280	costo	17	22	12,81	0,00034
281	rimettere	21	29	12,79	0,00034
282	svolgere	21	29	12,79	0,00034
283	giudizio	14	17	12,77	0,00035
284	dibattito	14	17	12,77	0,00035
285	tutela	14	17	12,77	0,00035
286	relazione	12	14	12,24	0,00046
287	operativo	12	14	12,24	0,00046
288	disporre	12	14	12,24	0,00046
289	costruire	31	48	12,22	0,00047
290	funzionare	23	33	12,22	0,00047
291	efficace	15	19	12,07	0,00051
292	carico	15	19	12,07	0,00051
293	unione	8	8	12,03	0,00052
294	razziale	8	8	12,03	0,00052
295	chiacchiera	8	8	12,03	0,00052
296	senatorio	8	8	12,03	0,00052
297	ridare	8	8	12,03	0,00052
298	proporzionale	8	8	12,03	0,00052
299	occidentale	8	8	12,03	0,00052
300	europeista	8	8	12,03	0,00052
301	cuneo	8	8	12,03	0,00052
302	credibile	8	8	12,03	0,00052
303	collegio	8	8	12,03	0,00052
304	burocrazia	8	8	12,03	0,00052
305	approvazione	8	8	12,03	0,00052
306	Israele	8	8	12,03	0,00052
307	dichiarare	10	11	11,92	0,00055
308	sacrificio	10	11	11,92	0,00055
309	obbligo	10	11	11,92	0,00055
310	gestione	10	11	11,92	0,00055
311	evasione	10	11	11,92	0,00055
312	flat_tax	10	11	11,92	0,00055
313	crimine	10	11	11,92	0,00055

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
314	Venezuela	10	11	11,92	0,00055
315	libero	20	28	11,6	0,00065
316	ritenere	27	41	11,53	0,00068
317	nave	16	21	11,52	0,00068
318	tribunale	16	21	11,52	0,00068
319	affrontare	28	43	11,42	0,00072
320	costa	13	16	11,4	0,00073
321	amministrazione	13	16	11,4	0,00073
322	mercato	21	30	11,33	0,00076
323	riguardare	46	79	11,12	0,00085
324	movimento	22	32	11,11	0,00085
325	positivo	22	32	11,11	0,00085
326	sostegno	17	23	11,09	0,00086
327	militare	17	23	11,09	0,00086
328	creare	59	106	11,07	0,00087
329	emergenza	23	34	10,92	0,00095
330	commissario	11	13	10,82	0,001
331	Macron	11	13	10,82	0,001
332	confrontare	11	13	10,82	0,001
333	precedente	14	18	10,76	0,00103
334	aggiungere	14	18	10,76	0,00103
335	ripetere	18	25	10,74	0,00105
336	privato	18	25	10,74	0,00105
337	condividere	25	38	10,63	0,00111
338	aliquota	7	7	10,53	0,00117
339	apparire	7	7	10,53	0,00117
340	Jobs_act	7	7	10,53	0,00117
341	autonomia	7	7	10,53	0,00117
342	Più_Europa	7	7	10,53	0,00117
343	Italia_Viva	7	7	10,53	0,00117
344	sottrarre	7	7	10,53	0,00117
345	soldato	7	7	10,53	0,00117
346	rivendicare	7	7	10,53	0,00117
347	revocare	7	7	10,53	0,00117
348	revoca	7	7	10,53	0,00117
349	irrealizzabile	7	7	10,53	0,00117
350	imputato	7	7	10,53	0,00117
351	espulsione	7	7	10,53	0,00117
352	esito	7	7	10,53	0,00117
353	esecutivo	7	7	10,53	0,00117
354	entusiasmo	7	7	10,53	0,00117
355	decisivo	7	7	10,53	0,00117
356	calabrese	7	7	10,53	0,00117
357	alleare	7	7	10,53	0,00117
358	Sea_Watch	7	7	10,53	0,00117
359	Costituzione	7	7	10,53	0,00117
360	attivare	9	10	10,45	0,00122
361	terrorismo	9	10	10,45	0,00122
362	confronto	32	52	10,17	0,00142
363	comune	32	52	10,17	0,00142
364	Malta	12	15	10,04	0,00152
365	Vita	12	15	10,04	0,00152

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
366	iniziativa	16	22	9,88	0,00167
367	differenza	16	22	9,88	0,00167
368	bianco	16	22	9,88	0,00167
369	sostenere	23	35	9,73	0,00181
370	mese	46	81	9,68	0,00186
371	cantiere	17	24	9,57	0,00198
372	rivoluzione	13	17	9,47	0,00209
373	Africa	13	17	9,47	0,00209
374	regionale	13	17	9,47	0,00209
375	contrario	13	17	9,47	0,00209
376	materia	10	12	9,42	0,00214
377	esigenza	10	12	9,42	0,00214
378	umano	39	67	9,39	0,00217
379	trasformare	18	26	9,32	0,00226
380	percorso	19	28	9,12	0,00252
381	offrire	19	28	9,12	0,00252
382	servire	47	84	9,07	0,00259
383	dignità	14	19	9,03	0,00265
384	CGIL	6	6	9,02	0,00266
385	penalizzare	6	6	9,02	0,00266
386	statuto	6	6	9,02	0,00266
387	indebolire	6	6	9,02	0,00266
388	guadagnare	6	6	9,02	0,00266
389	criticare	6	6	9,02	0,00266
390	Silvio	6	6	9,02	0,00266
391	Gentiloni	6	6	9,02	0,00266
392	terroristico	6	6	9,02	0,00266
393	stabilizzare	6	6	9,02	0,00266
394	sovrano	6	6	9,02	0,00266
395	sollecitare	6	6	9,02	0,00266
396	sgravio	6	6	9,02	0,00266
397	rimpatro	6	6	9,02	0,00266
398	profugo	6	6	9,02	0,00266
399	populismo	6	6	9,02	0,00266
400	poltrona	6	6	9,02	0,00266
401	organizzativo	6	6	9,02	0,00266
402	occidente	6	6	9,02	0,00266
403	inedito	6	6	9,02	0,00266
404	incremento	6	6	9,02	0,00266
405	garante	6	6	9,02	0,00266
406	collaborare	6	6	9,02	0,00266
407	candidatura	6	6	9,02	0,00266
408	candidare	6	6	9,02	0,00266
409	azzerare	6	6	9,02	0,00266
410	assenza	6	6	9,02	0,00266
411	amministratore	6	6	9,02	0,00266
412	affermazione	6	6	9,02	0,00266
413	abolizione	6	6	9,02	0,00266
414	Veneto	6	6	9,02	0,00266
415	Toninelli	6	6	9,02	0,00266
416	IVA	6	6	9,02	0,00266
417	IRPEF	6	6	9,02	0,00266

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
418	Banca	6	6	9,02	0,00266
419	sconfitta	8	9	8,99	0,00271
420	perseguire	8	9	8,99	0,00271
421	precario	8	9	8,99	0,00271
422	consapevolezza	8	9	8,99	0,00271
423	unito	8	9	8,99	0,00271
424	penale	8	9	8,99	0,00271
425	largo	8	9	8,99	0,00271
426	ampio	8	9	8,99	0,00271
427	Stati	8	9	8,99	0,00271
428	posizione	20	30	8,96	0,00276
429	ambiente	21	32	8,83	0,00296
430	volontà	22	34	8,72	0,00314
431	orgoglio	11	14	8,72	0,00315
432	nazione	11	14	8,72	0,00315
433	istituzionale	11	14	8,72	0,00315
434	eliminare	11	14	8,72	0,00315
435	commissione	11	14	8,72	0,00315
436	blocco	11	14	8,72	0,00315
437	base	24	38	8,57	0,00341
438	realizzare	24	38	8,57	0,00341
439	punto	68	130	8,43	0,00369
440	grave	29	48	8,43	0,00368
441	marzo	16	23	8,43	0,00369
442	formazione	12	16	8,21	0,00417
443	promettere	12	16	8,21	0,00417
444	operare	12	16	8,21	0,00417
445	industriale	12	16	8,21	0,00417
446	sistema	43	77	8,19	0,0042
447	principio	9	11	8,05	0,00456
448	indeterminato	9	11	8,05	0,00456
449	membro	9	11	8,05	0,00456
450	introdurre	9	11	8,05	0,00456
451	operare	12	16	8,21	0,00417
452	industriale	12	16	8,21	0,00417
453	sistema	43	77	8,19	0,0042
454	principio	9	11	8,05	0,00456
455	indeterminato	9	11	8,05	0,00456
456	membro	9	11	8,05	0,00456
457	introdurre	9	11	8,05	0,00456
458	sconfiggere	9	11	8,05	0,00456
459	fondare	9	11	8,05	0,00456
460	affermare	9	11	8,05	0,00456
461	accogliere	13	18	7,83	0,00514
462	litigare	13	18	7,83	0,00514
463	comunità	23	37	7,64	0,00569
464	capacità	24	39	7,61	0,00579
465	atto	29	49	7,6	0,00582
466	instabilità	7	8	7,54	0,00602
467	africano	7	8	7,54	0,00602
468	moneta	7	8	7,54	0,00602
469	comunista	7	8	7,54	0,00602

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
470	testo	7	8	7,54	0,00602
471	strutturale	7	8	7,54	0,00602
472	razza	7	8	7,54	0,00602
473	firma	7	8	7,54	0,00602
474	confusione	7	8	7,54	0,00602
475	carattere	7	8	7,54	0,00602
476	versare	7	8	7,54	0,00602
477	strategico	7	8	7,54	0,00602
478	riaprire	7	8	7,54	0,00602
479	pelle	7	8	7,54	0,00602
480	pedaggio	7	8	7,54	0,00602
481	netto	7	8	7,54	0,00602
482	francamente	7	8	7,54	0,00602
483	favorevole	7	8	7,54	0,00602
484	crollare	7	8	7,54	0,00602
485	comunale	7	8	7,54	0,00602
486	chiarezza	7	8	7,54	0,00602
487	agricoltura	7	8	7,54	0,00602
488	Unite	7	8	7,54	0,00602
489	Oriente	7	8	7,54	0,00602
490	Nazioni	7	8	7,54	0,00602
491	Medio	7	8	7,54	0,00602
492	Mattarella	7	8	7,54	0,00602
493	legislatura	5	5	7,52	0,00611
494	disperdere	5	5	7,52	0,00611
495	relativo	5	5	7,52	0,00611
496	islamico	5	5	7,52	0,00611
497	ennesimo	5	5	7,52	0,00611
498	bonus	5	5	7,52	0,00611
499	Maduro	5	5	7,52	0,00611
500	Carige	5	5	7,52	0,00611
501	strategia	10	13	7,42	0,00646
502	colpevole	10	13	7,42	0,00646
503	moderato	10	13	7,42	0,00646
504	collaborazione	10	13	7,42	0,00646
505	sostenibile	10	13	7,42	0,00646
506	minaccia	10	13	7,42	0,00646
507	valutare	15	22	7,33	0,00679
508	singolo	15	22	7,33	0,00679
509	famiglia	81	162	6,98	0,00824
510	certezza	11	15	6,98	0,00825
511	responsabile	11	15	6,98	0,00825
512	opera	18	28	6,93	0,00845
513	parola	30	52	6,87	0,00877
514	qualità	21	34	6,77	0,00926
515	risparmio	8	10	6,69	0,00969
516	partenza	8	10	6,69	0,00969
517	consenso	8	10	6,69	0,00969
518	povertà	12	17	6,66	0,00984
519	storico	12	17	6,66	0,00984
520	preferire	12	17	6,66	0,00984
521	Matteo	12	17	6,66	0,00984

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
522	confermare	13	19	6,43	0,0112
523	aumentare	33	59	6,33	0,01187
524	idea	41	76	6,27	0,01228
525	accoglienza	14	21	6,26	0,01233
526	modello	9	12	6,15	0,01313
527	dichiarazione	9	12	6,15	0,01313
528	comunicare	9	12	6,15	0,01313
529	riduzione	9	12	6,15	0,01313
530	obbligatorio	9	12	6,15	0,01313
531	università	9	12	6,15	0,01313
532	basare	9	12	6,15	0,01313
533	Emma	6	7	6,11	0,01341
534	prevenire	6	7	6,11	0,01341
535	regolamento	6	7	6,11	0,01341
536	partecipazione	6	7	6,11	0,01341
537	improprio	6	7	6,11	0,01341
538	depositare	6	7	6,11	0,01341
539	chiarire	6	7	6,11	0,01341
540	sospendere	6	7	6,11	0,01341
541	significativo	6	7	6,11	0,01341
542	riunione	6	7	6,11	0,01341
543	rifugiato	6	7	6,11	0,01341
544	realizzazione	6	7	6,11	0,01341
545	mandato	6	7	6,11	0,01341
546	estremismo	6	7	6,11	0,01341
547	complessivo	6	7	6,11	0,01341
548	appuntamento	6	7	6,11	0,01341
549	anticipo	6	7	6,11	0,01341
550	rivolgere	16	25	6,04	0,01395
551	oppressione	4	4	6,01	0,0142
552	Rai	4	4	6,01	0,0142
553	imposta	4	4	6,01	0,0142
554	terrorista	4	4	6,01	0,0142
555	miliardario	4	4	6,01	0,0142
556	essenziale	4	4	6,01	0,0142
557	contraddizione	4	4	6,01	0,0142
558	aumento	17	27	5,98	0,01448
559	bloccare	25	43	5,96	0,01462
560	primo	159	344	5,94	0,01482
561	straordinario	24	41	5,93	0,01486
562	motivo	19	31	5,91	0,01509
563	squadra	19	31	5,91	0,01509
564	tagliare	10	14	5,79	0,01612
565	smettere	10	14	5,79	0,01612
566	culturale	10	14	5,79	0,01612
567	norma	11	16	5,54	0,01859
568	legalità	11	16	5,54	0,01859
569	Torino	11	16	5,54	0,01859
570	dividere	11	16	5,54	0,01859
571	Sophia	7	9	5,37	0,02047
572	dimissione	7	9	5,37	0,02047
573	tradizionale	7	9	5,37	0,02047

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
574	stabile	7	9	5,37	0,02047
575	integrazione	7	9	5,37	0,02047
576	incertezza	7	9	5,37	0,02047
577	firmare	7	9	5,37	0,02047
578	fattore	7	9	5,37	0,02047
579	conflitto	7	9	5,37	0,02047
580	burocratico	7	9	5,37	0,02047
581	appoggiare	7	9	5,37	0,02047
582	acquisire	7	9	5,37	0,02047
583	compito	12	18	5,36	0,02055
584	costituire	12	18	5,36	0,02055
585	generale	25	44	5,25	0,022
586	protagonista	13	20	5,24	0,02205
587	solo	134	289	5,17	0,02291
588	reato	14	22	5,16	0,02316
589	Iran	14	22	5,16	0,02316
590	ministero	15	24	5,1	0,02393
591	intendere	15	24	5,1	0,02393
592	proporre	16	26	5,06	0,02442
593	assoluto	19	32	5,05	0,02461
594	sede	8	11	4,93	0,02639
595	pagina	8	11	4,93	0,02639
596	invitare	8	11	4,93	0,02639
597	manifestare	8	11	4,93	0,02639
598	limitare	8	11	4,93	0,02639
599	importanza	8	11	4,93	0,02639
600	auspicare	8	11	4,93	0,02639
601	centro	30	55	4,92	0,0266
602	giovane	35	66	4,75	0,02937
603	razzista	5	6	4,71	0,03002
604	punire	5	6	4,71	0,03002
605	pilastro	5	6	4,71	0,03002
606	illusione	5	6	4,71	0,03002
607	disponibilità	5	6	4,71	0,03002
608	ambizione	5	6	4,71	0,03002
609	stragrande	5	6	4,71	0,03002
610	selezionare	5	6	4,71	0,03002
611	revisione	5	6	4,71	0,03002
612	pubblicare	5	6	4,71	0,03002
613	procedimento	5	6	4,71	0,03002
614	perseguire	5	6	4,71	0,03002
615	parlamentarie	5	6	4,71	0,03002
616	opinione	5	6	4,71	0,03002
617	nido	5	6	4,71	0,03002
618	luglio	5	6	4,71	0,03002
619	iscrivere	5	6	4,71	0,03002
620	intelligenza	5	6	4,71	0,03002
621	incarico	5	6	4,71	0,03002
622	incapacità	5	6	4,71	0,03002
623	grillini	5	6	4,71	0,03002
624	finanza	5	6	4,71	0,03002
625	fermezza	5	6	4,71	0,03002

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
626	febbraio	5	6	4,71	0,03002
627	fascia	5	6	4,71	0,03002
628	fallimento	5	6	4,71	0,03002
629	espellere	5	6	4,71	0,03002
630	diseguaglianze	5	6	4,71	0,03002
631	convergenza	5	6	4,71	0,03002
632	consegnare	5	6	4,71	0,03002
633	coerente	5	6	4,71	0,03002
634	cinismo	5	6	4,71	0,03002
635	bugia	5	6	4,71	0,03002
636	brasiliano	5	6	4,71	0,03002
637	boom	5	6	4,71	0,03002
638	Livorno	5	6	4,71	0,03002
639	Gregoretti	5	6	4,71	0,03002
640	permettere	40	77	4,67	0,03077
641	francese	9	13	4,65	0,03104
642	America	9	13	4,65	0,03104
643	principe	9	13	4,65	0,03104
644	contribuire	9	13	4,65	0,03104
645	cambio	9	13	4,65	0,03104
646	unico	44	86	4,56	0,03278
647	olandese	3	3	4,51	0,03372
648	Olanda	3	3	4,51	0,03372
649	organizzazione	10	15	4,47	0,03453
650	chiusura	10	15	4,47	0,03453
651	interno	52	104	4,44	0,03513
652	fatto	31	58	4,44	0,03519
653	combattere	21	37	4,37	0,03649
654	bisognare	71	147	4,36	0,03683
655	direzione	11	17	4,35	0,03703
656	traffico	20	35	4,33	0,03747
657	richiedere	12	19	4,27	0,03873
658	solidarietà	12	19	4,27	0,03873
659	augurare	18	31	4,25	0,03916
660	partecipare	16	27	4,21	0,04024
661	vicenda	15	25	4,2	0,04044
662	utile	15	25	4,2	0,04044
663	tragedia	14	23	4,2	0,04031
664	vantaggio	6	8	4,1	0,04293
665	modificare	6	8	4,1	0,04293
666	dimettere	6	8	4,1	0,04293
667	stampa	6	8	4,1	0,04293
668	ricco	6	8	4,1	0,04293
669	verde	6	8	4,1	0,04293
670	testimonianza	6	8	4,1	0,04293
671	stabilire	6	8	4,1	0,04293
672	spreco	6	8	4,1	0,04293
673	sottolineare	6	8	4,1	0,04293
674	sopportare	6	8	4,1	0,04293
675	signor	6	8	4,1	0,04293
676	sanzione	6	8	4,1	0,04293
677	ricchezza	6	8	4,1	0,04293

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
678	mediatico	6	8	4,1	0,04293
679	manetta	6	8	4,1	0,04293
680	magistrato	6	8	4,1	0,04293
681	industria	6	8	4,1	0,04293
682	frattura	6	8	4,1	0,04293
683	esercitare	6	8	4,1	0,04293
684	disumano	6	8	4,1	0,04293
685	cogliere	6	8	4,1	0,04293
686	Chigi	6	8	4,1	0,04293

Tab. 8 Forme lessicali con freq. N \leq 5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 1 (CGD) del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	acqua	5	5	10,17	0,00142
2	frequentare	5	5	10,17	0,00142
3	guardia	5	5	10,17	0,00142
4	pomeriggio	5	5	10,17	0,00142
5	locale	5	5	10,17	0,00142
6	trasportare	5	5	10,17	0,00142
7	tendopoli	5	5	10,17	0,00142
8	massacrare	5	5	10,17	0,00142
9	maltrattamento	5	5	10,17	0,00142
10	mafia	5	5	10,17	0,00142
11	cane	5	5	10,17	0,00142
12	soccorritore	4	4	8,13	0,00436
13	fatica	4	4	8,13	0,00436
14	drammatico	4	4	8,13	0,00436
15	costiero	4	4	8,13	0,00436
16	albero	4	4	8,13	0,00436
17	stasera	4	4	8,13	0,00436
18	gelo	4	4	8,13	0,00436
19	esplosione	4	4	8,13	0,00436
20	culmine	4	4	8,13	0,00436
21	stretto	4	4	8,13	0,00436
22	shock	4	4	8,13	0,00436
23	riposo	4	4	8,13	0,00436
24	pistola	4	4	8,13	0,00436
25	passante	4	4	8,13	0,00436
26	parco	4	4	8,13	0,00436
27	immediato	4	4	8,13	0,00436
28	fiamma	4	4	8,13	0,00436
29	conto	4	4	8,13	0,00436
30	bracciante	4	4	8,13	0,00436
31	botta botte	4	4	8,13	0,00436
32	Virgilio	4	4	8,13	0,00436
33	Novara	4	4	8,13	0,00436
34	metro	5	6	6,89	0,00866
35	ucciso	5	6	6,89	0,00866
36	dimenticare	5	6	6,89	0,00866
37	mangiare	5	6	6,89	0,00866
38	ambulanza	3	3	6,09	0,01361
39	riservare	3	3	6,09	0,01361

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
40	procura	3	3	6,09	0,01361
41	genitore	3	3	6,09	0,01361
42	ammazzare	3	3	6,09	0,01361
43	traccia	3	3	6,09	0,01361
44	terremoto	3	3	6,09	0,01361
45	fermata fermato	3	3	6,09	0,01361
46	amare	3	3	6,09	0,01361
47	aiuto	3	3	6,09	0,01361
48	trauma	3	3	6,09	0,01361
49	sopravvissuto	3	3	6,09	0,01361
50	soccorso	3	3	6,09	0,01361
51	reparto	3	3	6,09	0,01361
52	impossibile	3	3	6,09	0,01361
53	gelosia	3	3	6,09	0,01361
54	dormire	3	3	6,09	0,01361
55	Sri	3	3	6,09	0,01361
56	Lanka	3	3	6,09	0,01361
57	Barge	3	3	6,09	0,01361
58	veronese	3	3	6,09	0,01361
59	valigia	3	3	6,09	0,01361
60	trentenne	3	3	6,09	0,01361
61	trasferire	3	3	6,09	0,01361
62	ricostruire	3	3	6,09	0,01361
63	questura	3	3	6,09	0,01361
64	presunto	3	3	6,09	0,01361
65	penale	3	3	6,09	0,01361
66	parcheggio	3	3	6,09	0,01361
67	mannaia	3	3	6,09	0,01361
68	legato	3	3	6,09	0,01361
69	ingiuria	3	3	6,09	0,01361
70	guinzaglio	3	3	6,09	0,01361
71	giardino	3	3	6,09	0,01361
72	furia	3	3	6,09	0,01361
73	emorragia	3	3	6,09	0,01361
74	commerciale	3	3	6,09	0,01361
75	casertano	3	3	6,09	0,01361
76	barbaramente	3	3	6,09	0,01361
77	bar	3	3	6,09	0,01361
78	baby_sitter	3	3	6,09	0,01361
79	albanese	3	3	6,09	0,01361
80	agrume	3	3	6,09	0,01361
81	agriturismo	3	3	6,09	0,01361
82	accorgersi	3	3	6,09	0,01361
83	abbracciare	3	3	6,09	0,01361
84	Sozzago	3	3	6,09	0,01361
85	San_Ferdinando	3	3	6,09	0,01361
86	Rosarno	3	3	6,09	0,01361
87	Riccardo	3	3	6,09	0,01361
88	Piccato	3	3	6,09	0,01361
89	Perla	3	3	6,09	0,01361
90	Novarese	3	3	6,09	0,01361
91	Malizia	3	3	6,09	0,01361

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
92	Mali	3	3	6,09	0,01361
93	Lucherini	3	3	6,09	0,01361
94	Gerace	3	3	6,09	0,01361
95	Cuneo	3	3	6,09	0,01361
96	Croce	3	3	6,09	0,01361
97	Caterina	3	3	6,09	0,01361
98	travolgere	4	5	5	0,02535
99	ricoverare	4	5	5	0,02535
100	telecamera	4	5	5	0,02535
101	maceria	4	5	5	0,02535
102	inizio	4	5	5	0,02535
103	violentare	4	5	5	0,02535
104	straniero	4	5	5	0,02535
105	sala sale	4	5	5	0,02535
106	racconto	4	5	5	0,02535
107	dolore	4	5	5	0,02535
108	scomparsa scomparso	4	5	5	0,02535
109	interrogare	4	5	5	0,02535
110	fortuna	4	5	5	0,02535
111	abitare	4	5	5	0,02535
112	Di	4	5	5	0,02535
113	mare	5	7	4,69	0,0304
114	mattina mattino	5	7	4,69	0,0304
115	giugno	5	7	4,69	0,0304
116	agente	5	7	4,69	0,0304
117	struttura	5	7	4,69	0,0304
118	sabato	5	7	4,69	0,0304
119	Torino	5	7	4,69	0,0304

Tab. 9 Forme lessicali con freq. N \leq 5 e $\chi^2 > 3,84$ ($p < 0,05$) della classe 2 (CGD) del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	emergenza	5	5	9,48	0,00207
2	assistere	5	5	9,48	0,00207
3	Movimento	5	5	9,48	0,00207
4	straordinario	5	5	9,48	0,00207
5	speciale	5	5	9,48	0,00207
6	May	5	5	9,48	0,00207
7	internazionale	4	4	7,58	0,00591
8	mandare	4	4	7,58	0,00591
9	superiore	4	4	7,58	0,00591
10	tradizionale	4	4	7,58	0,00591
11	repubblica	4	4	7,58	0,00591
12	moderno	4	4	7,58	0,00591
13	spaventare	4	4	7,58	0,00591
14	sfogare	4	4	7,58	0,00591
15	sfida	4	4	7,58	0,00591
16	reagire	4	4	7,58	0,00591
17	interpretare	4	4	7,58	0,00591
18	imparare	4	4	7,58	0,00591
19	generazione	4	4	7,58	0,00591
20	forma	4	4	7,58	0,00591

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
21	conquistare	4	4	7,58	0,00591
22	bionica	4	4	7,58	0,00591
23	assoluto	4	4	7,58	0,00591
24	arrivo	4	4	7,58	0,00591
25	Sanremo	4	4	7,58	0,00591
26	Chiesa	4	4	7,58	0,00591
27	Di_Maio	5	6	6,32	0,01192
28	umano	5	6	6,32	0,01192
29	divieto	5	6	6,32	0,01192
30	amore	5	6	6,32	0,01192
31	scegliere	5	6	6,32	0,01192
32	difficoltv†	5	6	6,32	0,01192
33	Dio	5	6	6,32	0,01192
34	serie	5	6	6,32	0,01192
35	esempio	5	6	6,32	0,01192
36	dedicare	5	6	6,32	0,01192
37	aspettare	5	6	6,32	0,01192
38	tribunale	5	6	6,32	0,01192
39	strattonare	5	6	6,32	0,01192
40	contatto	5	6	6,32	0,01192
41	tedesco	3	3	5,68	0,0172
42	decisione	3	3	5,68	0,0172
43	Francesca	3	3	5,68	0,0172
44	tradizione	3	3	5,68	0,0172
45	musicale	3	3	5,68	0,0172
46	coscienza	3	3	5,68	0,0172
47	afghano	3	3	5,68	0,0172
48	serata	3	3	5,68	0,0172
49	personaggio	3	3	5,68	0,0172
50	palco	3	3	5,68	0,0172
51	museo	3	3	5,68	0,0172
52	lingua	3	3	5,68	0,0172
53	impiantare	3	3	5,68	0,0172
54	edizione	3	3	5,68	0,0172
55	culto	3	3	5,68	0,0172
56	come_le	3	3	5,68	0,0172
57	collezione	3	3	5,68	0,0172
58	capello	3	3	5,68	0,0172
59	Theresa	3	3	5,68	0,0172
60	Mapuche	3	3	5,68	0,0172
61	Kelly	3	3	5,68	0,0172
62	Junior	3	3	5,68	0,0172
63	comunitv†	4	5	4,57	0,03261
64	guida	4	5	4,57	0,03261
65	aggiungere	4	5	4,57	0,03261
66	silenzio	4	5	4,57	0,03261
67	cultura	4	5	4,57	0,03261
68	comprare	4	5	4,57	0,03261
69	trasmettere	4	5	4,57	0,03261
70	terrorismo	4	5	4,57	0,03261
71	stringere	4	5	4,57	0,03261
72	personale	4	5	4,57	0,03261

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
73	maschile	4	5	4,57	0,03261
74	fondare	4	5	4,57	0,03261
75	esercito	4	5	4,57	0,03261
76	Roma	5	7	4,21	0,04009
77	riprendere	5	7	4,21	0,04009
78	reale	5	7	4,21	0,04009
79	musicista	5	7	4,21	0,04009
80	vincere	5	7	4,21	0,04009
81	vecchio	5	7	4,21	0,04009

Tab. 10 Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 3 (CGD) del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	democratico	5	5	20,37	<0,0001
2	stupro	5	5	20,37	<0,0001
3	slogan	5	5	20,37	<0,0001
4	me_too	5	5	20,37	<0,0001
5	festeggiare	5	5	20,37	<0,0001
6	femminismo	5	5	20,37	<0,0001
7	Catherine	5	5	20,37	<0,0001
8	tutela	4	4	16,27	<0,0001
9	solidariet�	4	4	16,27	<0,0001
10	scrittore	4	4	16,27	<0,0001
11	produttore	4	4	16,27	<0,0001
12	pacifico	4	4	16,27	<0,0001
13	Globe	4	4	16,27	<0,0001
14	scoprire	5	6	15,36	<0,0001
15	lettera	5	6	15,36	<0,0001
16	famoso	5	6	15,36	<0,0001
17	eleggere	5	6	15,36	<0,0001
18	cerimonia	5	6	15,36	<0,0001
19	aperto	5	6	15,36	<0,0001
20	dichiarare	3	3	12,19	0,00048
21	up	3	3	12,19	0,00048
22	time	3	3	12,19	0,00048
23	musulmano	3	3	12,19	0,00048
24	manifestare	3	3	12,19	0,00048
25	corteo	3	3	12,19	0,00048
26	bandiera	3	3	12,19	0,00048
27	Monde	3	3	12,19	0,00048
28	Danimarca	3	3	12,19	0,00048
29	migliaio	5	7	11,84	0,00057
30	organizzare	5	7	11,84	0,00057
31	velo	5	7	11,84	0,00057
32	Washington	5	7	11,84	0,00057
33	domenica	4	5	11,47	0,0007
34	scandalo	4	5	11,47	0,0007
35	protestare	4	5	11,47	0,0007
36	intellettuale	4	5	11,47	0,0007
37	rivelare	5	8	9,27	0,00232
38	appello	4	6	8,36	0,00384
39	regista	4	6	8,36	0,00384

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
40	voto	4	6	8,36	0,00384
41	moda	4	6	8,36	0,00384
42	combattere	4	6	8,36	0,00384
43	bisogno	4	6	8,36	0,00384
44	Facebook	4	6	8,36	0,00384
45	partita	3	4	7,71	0,00549
46	immigrazione	3	4	7,71	0,00549
47	servire	3	4	7,71	0,00549
48	indossare	3	4	7,71	0,00549
49	giornalista	3	4	7,71	0,00549
50	documento	3	4	7,71	0,00549
51	critica	3	4	7,71	0,00549
52	mondiale	3	4	7,71	0,00549
53	dibattito	3	4	7,71	0,00549
54	sogno	3	4	7,71	0,00549
55	rivolta	3	4	7,71	0,00549
56	presidenza	3	4	7,71	0,00549
57	parlamento	3	4	7,71	0,00549
58	occidentale	3	4	7,71	0,00549
59	bianco	3	4	7,71	0,00549
60	battere	3	4	7,71	0,00549
61	Davos	3	4	7,71	0,00549
62	nascondere	4	7	6,2	0,0128
63	importante	4	7	6,2	0,0128
64	abito	4	7	6,2	0,0128
65	rete	4	7	6,2	0,0128
66	isolare	4	7	6,2	0,0128
67	femminista	4	7	6,2	0,0128
68	risultato	3	5	5,12	0,0237
69	vertice	3	5	5,12	0,0237
70	distanza	3	5	5,12	0,0237
71	passare	3	5	5,12	0,0237
72	scadere	3	5	5,12	0,0237
73	contare	3	5	5,12	0,0237
74	colore	3	5	5,12	0,0237
75	decina	4	8	4,64	0,0313
76	spingere	4	8	4,64	0,0313
77	numero	4	8	4,64	0,0313
78	social	4	8	4,64	0,0313
79	cinema	4	8	4,64	0,0313
80	danno	4	8	4,64	0,0313
81	accogliere	2	3	4,16	0,04132
82	disporre	2	3	4,16	0,04132
83	Wuhan	2	3	4,16	0,04132
84	ondata	2	3	4,16	0,04132
85	interesse	2	3	4,16	0,04132
86	esplodere	2	3	4,16	0,04132
87	Maria	2	3	4,16	0,04132
88	vita	2	3	4,16	0,04132
89	vestito	2	3	4,16	0,04132
90	innocente	2	3	4,16	0,04132
91	caccia	2	3	4,16	0,04132

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
92	D	2	3	4,16	0,04132
93	tetto	2	3	4,16	0,04132
94	tabVrt	2	3	4,16	0,04132
95	sfilata	2	3	4,16	0,04132
96	schiaivitVrt	2	3	4,16	0,04132
97	quotidiano	2	3	4,16	0,04132
98	passione	2	3	4,16	0,04132
99	legislatura	2	3	4,16	0,04132
100	intervista	2	3	4,16	0,04132
101	iniziativa	2	3	4,16	0,04132
102	indiano	2	3	4,16	0,04132
103	ideale	2	3	4,16	0,04132
104	forum	2	3	4,16	0,04132
105	formare	2	3	4,16	0,04132
106	est	2	3	4,16	0,04132
107	enorme	2	3	4,16	0,04132
108	elezione	2	3	4,16	0,04132
109	divorzio	2	3	4,16	0,04132
110	discriminare	2	3	4,16	0,04132
111	deputato	2	3	4,16	0,04132
112	corruzione	2	3	4,16	0,04132
113	contemporaneo	2	3	4,16	0,04132
114	cinematografico	2	3	4,16	0,04132
115	Repubblica	2	3	4,16	0,04132
116	Le	2	3	4,16	0,04132
117	Grecia	2	3	4,16	0,04132

Tab. 11 Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 4 (CGD) del *sub-corpus* donna

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	cosiddetto	5	5	35,41	<0,0001
2	furto	5	5	35,41	<0,0001
3	crescita	5	5	35,41	<0,0001
4	APE	5	5	35,41	<0,0001
5	aumento	4	4	28,29	<0,0001
6	penalizzare	4	4	28,29	<0,0001
7	stabile	4	4	28,29	<0,0001
8	requisito	4	4	28,29	<0,0001
9	organizzazione	4	4	28,29	<0,0001
10	disuguaglianza	4	4	28,29	<0,0001
11	contratto	4	4	28,29	<0,0001
12	autonomo	4	4	28,29	<0,0001
13	consentire	5	6	27,89	<0,0001
14	disabile	3	3	21,19	<0,0001
15	prioritV†	3	3	21,19	<0,0001
16	vulnerabilitV†	3	3	21,19	<0,0001
17	prorogare	3	3	21,19	<0,0001
18	fenomeno	3	3	21,19	<0,0001
19	diseguaglianze	3	3	21,19	<0,0001
20	comparto	3	3	21,19	<0,0001
21	canale	3	3	21,19	<0,0001
22	Zingaretti	3	3	21,19	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
23	settore	4	5	21,08	<0,0001
24	prevedere	4	5	21,08	<0,0001
25	tasso	4	5	21,08	<0,0001
26	stipendio	4	5	21,08	<0,0001
27	PD	4	5	21,08	<0,0001
28	novembre	4	5	21,08	<0,0001
29	differenza	4	5	21,08	<0,0001
30	collega	4	5	21,08	<0,0001
31	cittadinanza	4	5	21,08	<0,0001
32	regione	5	8	18,59	<0,0001
33	scuola	5	8	18,59	<0,0001
34	medio	5	8	18,59	<0,0001
35	pensione	4	6	16,32	<0,0001
36	rischio	5	9	15,53	<0,0001
37	familiare	5	9	15,53	<0,0001
38	significare	3	4	14,44	0,00014
39	riguardare	3	4	14,44	0,00014
40	maternitv†	3	4	14,44	0,00014
41	massimo	3	4	14,44	0,00014
42	dollaro	3	4	14,44	0,00014
43	controllare	3	4	14,44	0,00014
44	Santelli	3	4	14,44	0,00014
45	mesa mese	4	7	12,95	0,00031
46	studio	4	7	12,95	0,00031
47	sicurezza	5	11	11,16	0,00083
48	sistema	3	5	10,45	0,00122
49	record	3	5	10,45	0,00122
50	possibile	3	5	10,45	0,00122
51	basso	3	5	10,45	0,00122
52	CGIL	3	5	10,45	0,00122
53	povertv†	3	5	10,45	0,00122
54	livello	3	5	10,45	0,00122
55	gravidanza	3	5	10,45	0,00122
56	esistere	3	5	10,45	0,00122
57	vista	3	5	10,45	0,00122
58	rappresentare	3	5	10,45	0,00122
59	dettaglio	3	5	10,45	0,00122
60	consenso	3	5	10,45	0,00122
61	minore	2	3	8,13	0,00435
62	carriera	2	3	8,13	0,00435
63	manco	2	3	8,13	0,00435
64	maggioranza	2	3	8,13	0,00435
65	interessare	2	3	8,13	0,00435
66	termine	2	3	8,13	0,00435
67	tendenza	2	3	8,13	0,00435
68	specie	2	3	8,13	0,00435
69	richiedere	2	3	8,13	0,00435
70	integrazione	2	3	8,13	0,00435
71	grosso	2	3	8,13	0,00435
72	disparitv†	2	3	8,13	0,00435
73	disoccupare	2	3	8,13	0,00435
74	disabilitv†	2	3	8,13	0,00435

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
75	conte conto	2	3	8,13	0,00435
76	certezza	2	3	8,13	0,00435
77	approvare	2	3	8,13	0,00435
78	Jole	2	3	8,13	0,00435
79	Camusso	2	3	8,13	0,00435
80	crisi	3	6	7,83	0,00514
81	Calabria	3	6	7,83	0,00514
82	pubblicare	3	6	7,83	0,00514
83	privato	3	6	7,83	0,00514
84	posto	3	6	7,83	0,00514
85	mercato	3	6	7,83	0,00514
86	giustizia	3	6	7,83	0,00514
87	uscire	5	14	7,08	0,00778
88	confermare	3	7	6	0,01433
89	lista	3	7	6	0,01433
90	dato	3	7	6	0,01433
91	tema	4	11	5,86	0,0155
92	ricevere	2	4	5,21	0,0225
93	permettere	2	4	5,21	0,0225
94	iniziare	2	4	5,21	0,0225
95	maggiore	2	4	5,21	0,0225
96	stabilire	2	4	5,21	0,0225
97	incontro	2	4	5,21	0,0225
98	annunciare	2	4	5,21	0,0225
99	umanitario	2	4	5,21	0,0225
100	sposare	2	4	5,21	0,0225
101	obiettivo	2	4	5,21	0,0225
102	economia	2	4	5,21	0,0225
103	chiaro	2	4	5,21	0,0225
104	assumere	2	4	5,21	0,0225
105	accordo	2	4	5,21	0,0225
106	sconto	2	4	5,21	0,0225
107	ridurre	2	4	5,21	0,0225
108	greco	2	4	5,21	0,0225
109	contributo	2	4	5,21	0,0225
110	chiusura	2	4	5,21	0,0225
111	assegno	2	4	5,21	0,0225
112	amministrazione	2	4	5,21	0,0225
113	Vicenza	2	4	5,21	0,0225
114	Europa	3	8	4,66	0,03091
115	problema	3	8	4,66	0,03091
116	nord	3	8	4,66	0,03091
117	particolare	3	8	4,66	0,03091
118	svolgere	4	13	4,07	0,04355
119	definire	4	13	4,07	0,04355

Tab. 12 Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 1 (CGD) del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	frequentare	5	5	10,46	0,00122
2	procuratore	5	5	10,46	0,00122

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
3	piazza	5	5	10,46	0,00122
4	aggressione	5	5	10,46	0,00122
5	scontare	5	5	10,46	0,00122
6	massacrare	5	5	10,46	0,00122
7	femminicidio	5	5	10,46	0,00122
8	ennesimo	5	5	10,46	0,00122
9	approfittare	5	5	10,46	0,00122
10	Novara	5	5	10,46	0,00122
11	episodio	4	4	8,36	0,00384
12	ipotesi	4	4	8,36	0,00384
13	discoteca	4	4	8,36	0,00384
14	condanna	4	4	8,36	0,00384
15	Imola	4	4	8,36	0,00384
16	pena	4	4	8,36	0,00384
17	ordinanza	4	4	8,36	0,00384
18	insegnante	4	4	8,36	0,00384
19	estorsione	4	4	8,36	0,00384
20	verbale	4	4	8,36	0,00384
21	sequestrare	4	4	8,36	0,00384
22	scontro	4	4	8,36	0,00384
23	proprietario	4	4	8,36	0,00384
24	professore	4	4	8,36	0,00384
25	penitenziario	4	4	8,36	0,00384
26	organizzare	4	4	8,36	0,00384
27	online	4	4	8,36	0,00384
28	omicida	4	4	8,36	0,00384
29	napoletano	4	4	8,36	0,00384
30	ergastolo	4	4	8,36	0,00384
31	eccolo	4	4	8,36	0,00384
32	confessione	4	4	8,36	0,00384
33	botta	4	4	8,36	0,00384
34	autopsia	4	4	8,36	0,00384
35	Lorenzo	4	4	8,36	0,00384
36	Gip	4	4	8,36	0,00384
37	Cremona	4	4	8,36	0,00384
38	coltello	5	6	7,13	0,00758
39	residente	5	6	7,13	0,00758
40	delitto	5	6	7,13	0,00758
41	capodanno	5	6	7,13	0,00758
42	Pinto	5	6	7,13	0,00758
43	Giacinto	5	6	7,13	0,00758
44	Spada	3	3	6,26	0,01234
45	presunto	3	3	6,26	0,01234
46	Sara	3	3	6,26	0,01234
47	Messina	3	3	6,26	0,01234
48	usura	3	3	6,26	0,01234
49	rabbia	3	3	6,26	0,01234
50	postare	3	3	6,26	0,01234
51	minuto	3	3	6,26	0,01234
52	litigio	3	3	6,26	0,01234
53	coltellata	3	3	6,26	0,01234
54	appartenere	3	3	6,26	0,01234

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
55	Torino	3	3	6,26	0,01234
56	Ostia	3	3	6,26	0,01234
57	tatuaggio	3	3	6,26	0,01234
58	sedere	3	3	6,26	0,01234
59	ripetizione	3	3	6,26	0,01234
60	probatorio	3	3	6,26	0,01234
61	perizia	3	3	6,26	0,01234
62	nozze	3	3	6,26	0,01234
63	movente	3	3	6,26	0,01234
64	monossido	3	3	6,26	0,01234
65	maltrattamento	3	3	6,26	0,01234
66	liceo	3	3	6,26	0,01234
67	esalazione	3	3	6,26	0,01234
68	dirigente	3	3	6,26	0,01234
69	convivente	3	3	6,26	0,01234
70	confesso	3	3	6,26	0,01234
71	camera	3	3	6,26	0,01234
72	braccialetto	3	3	6,26	0,01234
73	bolognese	3	3	6,26	0,01234
74	aula	3	3	6,26	0,01234
75	alessandrino	3	3	6,26	0,01234
76	aggravare	3	3	6,26	0,01234
77	Siena	3	3	6,26	0,01234
78	Santorelli	3	3	6,26	0,01234
79	Romano	3	3	6,26	0,01234
80	Natale	3	3	6,26	0,01234
81	Magliana	3	3	6,26	0,01234
82	Lucherini	3	3	6,26	0,01234
83	Digos	3	3	6,26	0,01234
84	DNA	3	3	6,26	0,01234
85	escludere	4	5	5,18	0,02283
86	risultare	4	5	5,18	0,02283
87	quartiere	4	5	5,18	0,02283
88	interrogatorio	4	5	5,18	0,02283
89	documento	4	5	5,18	0,02283
90	genitore	4	5	5,18	0,02283
91	esame	4	5	5,18	0,02283
92	La	4	5	5,18	0,02283
93	picchiare	5	7	4,88	0,0271
94	origine	5	7	4,88	0,0271
95	suicidio	5	7	4,88	0,0271
96	sparire	5	7	4,88	0,0271
97	soldo	5	7	4,88	0,0271

Tab. 13 Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 2 (CGD) del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	testimone	5	5	25,94	<0,0001
2	soccorritore	5	5	25,94	<0,0001
3	raffica	5	5	25,94	<0,0001
4	gravemente	4	4	20,73	<0,0001

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
5	vagone	4	4	20,73	<0,0001
6	signore	4	4	20,73	<0,0001
7	scivolare	4	4	20,73	<0,0001
8	scavare	4	4	20,73	<0,0001
9	metropolitano	4	4	20,73	<0,0001
10	autostrada	4	4	20,73	<0,0001
11	D	4	4	20,73	<0,0001
12	scendere	5	6	20	<0,0001
13	ambulanza	5	6	20	<0,0001
14	tentativo	5	6	20	<0,0001
15	chiudere	5	7	15,81	<0,0001
16	spinta	5	7	15,81	<0,0001
17	paziente	3	3	15,53	<0,0001
18	proseguire	3	3	15,53	<0,0001
19	proiettile	3	3	15,53	<0,0001
20	sferzata	3	3	15,53	<0,0001
21	sciatore	3	3	15,53	<0,0001
22	petardo	3	3	15,53	<0,0001
23	peruviano	3	3	15,53	<0,0001
24	naufragio	3	3	15,53	<0,0001
25	montagna	3	3	15,53	<0,0001
26	mareggiata	3	3	15,53	<0,0001
27	maltempo	3	3	15,53	<0,0001
28	attraversare	3	3	15,53	<0,0001
29	attimo	3	3	15,53	<0,0001
30	agosto	3	3	15,53	<0,0001
31	San_Camillo	3	3	15,53	<0,0001
32	Crotone	3	3	15,53	<0,0001
33	Giovanna	4	5	15,03	0,0001
34	passaggero	4	5	15,03	0,0001
35	puntare	4	5	15,03	0,0001
36	medico	5	8	12,71	0,00036
37	riportare	5	8	12,71	0,00036
38	protezione	5	8	12,71	0,00036
39	elicottero	5	8	12,71	0,00036
40	albero	5	8	12,71	0,00036
41	area	4	6	11,3	0,00077
42	precipitare	4	6	11,3	0,00077
43	shock	4	6	11,3	0,00077
44	sequestro	4	6	11,3	0,00077
45	cane	5	9	10,34	0,0013
46	accadere	5	9	10,34	0,0013
47	disperato	5	9	10,34	0,0013
48	urlò	3	4	10,2	0,0014
49	test	3	4	10,2	0,0014
50	ghiacciaio	3	4	10,2	0,0014
51	telefonata	3	4	10,2	0,0014
52	stamattina	3	4	10,2	0,0014
53	istante	3	4	10,2	0,0014
54	ghiaccio	3	4	10,2	0,0014
55	codice	3	4	10,2	0,0014
56	branco	3	4	10,2	0,0014

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
57	Francesca	3	4	10,2	0,0014
58	Cucv®	3	4	10,2	0,0014
59	rosso	4	7	8,68	0,00321
60	costiero	4	7	8,68	0,00321
61	civile	4	7	8,68	0,00321
62	servire	4	7	8,68	0,00321
63	partita	4	7	8,68	0,00321
64	stazione	3	5	7,08	0,00778
65	salire	3	5	7,08	0,00778
66	identificare	3	5	7,08	0,00778
67	resto	3	5	7,08	0,00778
68	caduta	3	5	7,08	0,00778
69	aria	3	5	7,08	0,00778
70	parente	3	5	7,08	0,00778
71	orario	3	5	7,08	0,00778
72	cominciare	4	8	6,77	0,00928
73	testa	4	8	6,77	0,00928
74	rogo	4	8	6,77	0,00928
75	danno	4	8	6,77	0,00928
76	avvicinare	4	8	6,77	0,00928
77	hotel	2	3	5,63	0,01765
78	poliziotto	2	3	5,63	0,01765
79	bilancio	2	3	5,63	0,01765
80	viaggiare	2	3	5,63	0,01765
81	girare	2	3	5,63	0,01765
82	voce	2	3	5,63	0,01765
83	senso	2	3	5,63	0,01765
84	liberare	2	3	5,63	0,01765
85	giungere	2	3	5,63	0,01765
86	automobile	2	3	5,63	0,01765
87	velivolo	2	3	5,63	0,01765
88	veicolo	2	3	5,63	0,01765
89	valanga	2	3	5,63	0,01765
90	urgente	2	3	5,63	0,01765
91	spezzare	2	3	5,63	0,01765
92	senzatetto	2	3	5,63	0,01765
93	segnalare	2	3	5,63	0,01765
94	scoperchiare	2	3	5,63	0,01765
95	sala	2	3	5,63	0,01765
96	invadere	2	3	5,63	0,01765
97	fotogramma	2	3	5,63	0,01765
98	filmato	2	3	5,63	0,01765
99	dormire	2	3	5,63	0,01765
100	dinamico	2	3	5,63	0,01765
101	dimora	2	3	5,63	0,01765
102	convoglio	2	3	5,63	0,01765
103	caduto	2	3	5,63	0,01765
104	bar	2	3	5,63	0,01765
105	afferrare	2	3	5,63	0,01765
106	Gabriella	2	3	5,63	0,01765
107	Capparelli	2	3	5,63	0,01765
108	Aosta	2	3	5,63	0,01765

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
109	interrogare	4	9	5,32	0,02107
110	fuggire	3	6	5,06	0,02441
111	dimettere	3	6	5,06	0,02441
112	centinaio	3	6	5,06	0,02441
113	proteggere	3	6	5,06	0,02441
114	canale	3	6	5,06	0,02441
115	migrare	5	13	4,79	0,02857
116	recuperare	5	13	4,79	0,02857
117	sud	4	10	4,2	0,0404
118	spingere	4	10	4,2	0,0404

Tab. 14 Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 3 (CGD) del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	possibilità	5	5	17,84	<0,0001
2	ringraziare	5	5	17,84	<0,0001
3	dedicare	5	5	17,84	<0,0001
4	cinema	5	5	17,84	<0,0001
5	complesso	4	4	14,26	0,00015
6	scena	4	4	14,26	0,00015
7	regista	4	4	14,26	0,00015
8	progetto	4	4	14,26	0,00015
9	luna	4	4	14,26	0,00015
10	incredibile	4	4	14,26	0,00015
11	amore	4	4	14,26	0,00015
12	Coppi	4	4	14,26	0,00015
13	titolo	4	4	14,26	0,00015
14	ricordo	4	4	14,26	0,00015
15	realizzare	4	4	14,26	0,00015
16	musicista	4	4	14,26	0,00015
17	discorso	4	4	14,26	0,00015
18	contribuire	4	4	14,26	0,00015
19	cantare	4	4	14,26	0,00015
20	africano	4	4	14,26	0,00015
21	Weinstein	4	4	14,26	0,00015
22	Steve	4	4	14,26	0,00015
23	Hammamet	4	4	14,26	0,00015
24	Amelio	4	4	14,26	0,00015
25	simbolo	5	6	13,26	0,00027
26	libro	5	6	13,26	0,00027
27	Armani	3	3	10,68	0,00108
28	re	3	3	10,68	0,00108
29	elegante	3	3	10,68	0,00108
30	continente	3	3	10,68	0,00108
31	Piersanti	3	3	10,68	0,00108
32	Ciampi	3	3	10,68	0,00108
33	Charles	3	3	10,68	0,00108
34	stratega	3	3	10,68	0,00108
35	ingegnere	3	3	10,68	0,00108
36	etico	3	3	10,68	0,00108
37	Elisabetta	3	3	10,68	0,00108
38	Deneuve	3	3	10,68	0,00108

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
39	Catherine	3	3	10,68	0,00108
40	Camilla	3	3	10,68	0,00108
41	Bettino	3	3	10,68	0,00108
42	5_Stelle	3	3	10,68	0,00108
43	importante	5	7	10,06	0,00151
44	immaginare	5	7	10,06	0,00151
45	storico	4	5	9,87	0,00168
46	consentire	4	5	9,87	0,00168
47	segreto	4	5	9,87	0,00168
48	sala	4	5	9,87	0,00168
49	piangere	4	5	9,87	0,00168
50	migliore	4	5	9,87	0,00168
51	dimostrare	4	5	9,87	0,00168
52	coraggioso	4	5	9,87	0,00168
53	orizzonte	4	5	9,87	0,00168
54	esempio	4	5	9,87	0,00168
55	capace	5	8	7,73	0,00544
56	ritorno	5	8	7,73	0,00544
57	giusto	5	8	7,73	0,00544
58	diretto	4	6	7,03	0,008
59	seguire	4	6	7,03	0,008
61	ospitare	4	6	7,03	0,008
62	nascere	4	6	7,03	0,008
63	difficile	4	6	7,03	0,008
64	attivitv†	4	6	7,03	0,008
65	intellettuale	4	6	7,03	0,008
66	bianco	4	6	7,03	0,008
67	cultura	3	4	6,58	0,01028
68	Australia	3	4	6,58	0,01028
69	sogno	3	4	6,58	0,01028
70	riserva	3	4	6,58	0,01028
71	eccezionale	3	4	6,58	0,01028
72	vincere	3	4	6,58	0,01028
73	thriller	3	4	6,58	0,01028
74	sognare	3	4	6,58	0,01028
75	sicuramente	3	4	6,58	0,01028
76	ereditv†	3	4	6,58	0,01028
77	commuovere	3	4	6,58	0,01028
78	cambiamento	3	4	6,58	0,01028
79	amicizia	3	4	6,58	0,01028
80	Livorno	3	4	6,58	0,01028
81	Casa_Bianca	3	4	6,58	0,01028
82	numero	5	9	5,97	0,01451
83	piede	5	9	5,97	0,01451
84	America	4	7	5,09	0,02413
85	aspettare	4	7	5,09	0,02413
86	apparire	4	7	5,09	0,02413
87	sicuro	4	7	5,09	0,02413
88	unire	4	7	5,09	0,02413
89	mafia	4	7	5,09	0,02413
90	dividere	4	7	5,09	0,02413
91	braccio	4	7	5,09	0,02413

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
92	virus	3	5	4,24	0,03955
93	naturale	3	5	4,24	0,03955
94	cancellare	3	5	4,24	0,03955
95	tecnologia	3	5	4,24	0,03955
96	sport	3	5	4,24	0,03955
97	personaggio	3	5	4,24	0,03955
98	nazione	3	5	4,24	0,03955
99	effetto	3	5	4,24	0,03955
100	chiaro	3	5	4,24	0,03955
101	rilanciare	3	5	4,24	0,03955
102	condividere	3	5	4,24	0,03955
103	Sergio	3	5	4,24	0,03955

Tab. 15 Forme lessicali con freq. $N \leq 5$ e $\chi^2 > 3,84$ ($p < 0,05$) della classe 4 (CGD) del *sub-corpus* uomo

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
1	guadagnare	5	5	12,13	0,00049
2	distanza	5	5	12,13	0,00049
3	Salvini	5	5	12,13	0,00049
4	ONU	5	5	12,13	0,00049
5	sviluppo	5	5	12,13	0,00049
6	salario	5	5	12,13	0,00049
7	popolare	5	5	12,13	0,00049
8	differenza	5	5	12,13	0,00049
9	rispondere	4	4	9,69	0,00185
10	unità	4	4	9,69	0,00185
11	avanzare	4	4	9,69	0,00185
12	porto	4	4	9,69	0,00185
13	intelligente	4	4	9,69	0,00185
14	grandissimo	4	4	9,69	0,00185
15	energia	4	4	9,69	0,00185
16	ambientale	4	4	9,69	0,00185
17	team	4	4	9,69	0,00185
18	spostare	4	4	9,69	0,00185
19	spirito	4	4	9,69	0,00185
20	recente	4	4	9,69	0,00185
21	occidentale	4	4	9,69	0,00185
22	manifesto	4	4	9,69	0,00185
23	gommone	4	4	9,69	0,00185
24	decreto	4	4	9,69	0,00185
25	bisogno	4	4	9,69	0,00185
26	Zingaretti	4	4	9,69	0,00185
27	World_Economic_Forum	4	4	9,69	0,00185
28	Vaticano	4	4	9,69	0,00185
29	PD	4	4	9,69	0,00185
30	situazione	5	6	8,51	0,00353
31	impresa	5	6	8,51	0,00353
32	buono	5	6	8,51	0,00353
33	Africa	5	6	8,51	0,00353
34	miliardo	5	6	8,51	0,00353
35	Cina	3	3	7,26	0,00705

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
36	Tripoli	3	3	7,26	0,00705
37	Libia	3	3	7,26	0,00705
38	Turchia	3	3	7,26	0,00705
39	vista	3	3	7,26	0,00705
40	tavolo	3	3	7,26	0,00705
41	mondiale	3	3	7,26	0,00705
42	ebreo	3	3	7,26	0,00705
43	sostenibile	3	3	7,26	0,00705
44	robot	3	3	7,26	0,00705
45	libico	3	3	7,26	0,00705
46	fondamentale	3	3	7,26	0,00705
47	export	3	3	7,26	0,00705
48	esercito	3	3	7,26	0,00705
49	contratto	3	3	7,26	0,00705
50	vertice	3	3	7,26	0,00705
51	sbarco	3	3	7,26	0,00705
52	rischioso	3	3	7,26	0,00705
53	nucleare	3	3	7,26	0,00705
54	motore	3	3	7,26	0,00705
55	maschio	3	3	7,26	0,00705
56	fronteggiare	3	3	7,26	0,00705
57	fenomeno	3	3	7,26	0,00705
58	dollaro	3	3	7,26	0,00705
59	agenda	3	3	7,26	0,00705
60	affondare	3	3	7,26	0,00705
61	Sea_Watch	3	3	7,26	0,00705
62	Sant	3	3	7,26	0,00705
63	Islanda	3	3	7,26	0,00705
64	Costa	3	3	7,26	0,00705
65	pronto	4	5	6,24	0,01252
66	vestire	4	5	6,24	0,01252
67	combattere	4	5	6,24	0,01252
68	genere	4	5	6,24	0,01252
69	esprimere	4	5	6,24	0,01252
70	esistere	4	5	6,24	0,01252
71	presenza	4	5	6,24	0,01252
72	polemica	4	5	6,24	0,01252
73	conta	4	5	6,24	0,01252
74	campagna	4	5	6,24	0,01252
75	battaglia	4	5	6,24	0,01252
76	agire	4	5	6,24	0,01252
77	volontv†	4	5	6,24	0,01252
78	sorta	4	5	6,24	0,01252
79	prevedere	4	5	6,24	0,01252
80	leggenda	4	5	6,24	0,01252
81	giocare	4	5	6,24	0,01252
82	cosiddetto	4	5	6,24	0,01252
83	sostenere	5	7	6,04	0,01398
84	consigliere	5	7	6,04	0,01398
85	nave	5	7	6,04	0,01398
86	medio	5	7	6,04	0,01398
87	guardare	5	7	6,04	0,01398

ID	Lemma	ST (N)	Occ. (N)	χ^2	p-value
88	ricco	5	8	4,29	0,0383
89	decisione	5	8	4,29	0,0383
90	scegliere	5	8	4,29	0,0383
91	TV	5	8	4,29	0,0383
92	climatico	5	8	4,29	0,0383
93	crisi	4	6	4,07	0,0437
94	umano	4	6	4,07	0,0437
95	economico	4	6	4,07	0,0437
96	attenzione	4	6	4,07	0,0437
97	partire	4	6	4,07	0,0437
98	lotta	4	6	4,07	0,0437
99	creare	4	6	4,07	0,0437
100	consegnare	4	6	4,07	0,0437
101	Donald	4	6	4,07	0,0437
102	Mediterraneo	4	6	4,07	0,0437
103	Francia	3	4	4,05	0,04422
104	ponte	3	4	4,05	0,04422
105	firmare	3	4	4,05	0,04422
106	scimmia	3	4	4,05	0,04422
107	ripetere	3	4	4,05	0,04422
108	rifiutare	3	4	4,05	0,04422
109	missione	3	4	4,05	0,04422
110	concludere	3	4	4,05	0,04422
111	Giovanni	3	4	4,05	0,04422
112	soldato	3	4	4,05	0,04422
113	sfuggire	3	4	4,05	0,04422
114	preparare	3	4	4,05	0,04422
115	pietra	3	4	4,05	0,04422
116	operare	3	4	4,05	0,04422
117	lista	3	4	4,05	0,04422
118	imporre	3	4	4,05	0,04422
119	copertina	3	4	4,05	0,04422
120	cibo	3	4	4,05	0,04422



Università
Ca' Foscari
Venezia

DEPOSITO ELETTRONICO DELLA TESI DI DOTTORATO

DICHIARAZIONE SOSTITUTIVA DELL'ATTO DI NOTORIETA'
(Art. 47 D.P.R. 445 del 28/12/2000 e relative modifiche)

Io sottoscritto Monia Azzalini
nat. a Sondrio (prov. SO) il 25/09/1970
residente a Pavia (PV) in Via Federico Faruffini n. 6
Matricola (se posseduta) 956448 Autore della tesi di dottorato dal titolo:
Rappresentazioni di genere dentro e attraverso il linguaggio dell'informazione italiana
.....
.....
Dottorato di ricerca in Lingue, culture e società moderne e Scienze del linguaggio
(in cotutela con)
Ciclo 34°
Anno di conseguimento del titolo 2022

DICHIARO

di essere a conoscenza:

- 1) del fatto che in caso di dichiarazioni mendaci, oltre alle sanzioni previste dal codice penale e dalle Leggi speciali per l'ipotesi di falsità in atti ed uso di atti falsi, decado fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
- 2) dell'obbligo per l'Università di provvedere, per via telematica, al deposito di legge delle tesi di dottorato presso le Biblioteche Nazionali Centrali di Roma e di Firenze al fine di assicurarne la conservazione e la consultabilità da parte di terzi;
- 3) che l'Università si riserva i diritti di riproduzione per scopi didattici, con citazione della fonte;
- 4) del fatto che il testo integrale della tesi di dottorato di cui alla presente dichiarazione viene archiviato e reso consultabile via internet attraverso l'Archivio Istituzionale ad Accesso Aperto dell'Università Ca' Foscari, oltre che attraverso i cataloghi delle Biblioteche Nazionali Centrali di Roma e Firenze;
- 5) del fatto che, ai sensi e per gli effetti di cui al D.Lgs. n. 196/2003, i dati personali raccolti saranno trattati, anche con strumenti informatici, esclusivamente nell'ambito del procedimento per il quale la presentazione viene resa;
- 6) del fatto che la copia della tesi in formato elettronico depositato nell'Archivio Istituzionale ad Accesso Aperto è del tutto corrispondente alla tesi in formato cartaceo, controfirmata dal tutor, consegnata presso la segreteria didattica del dipartimento di riferimento del corso di dottorato ai fini del deposito presso l'Archivio di Ateneo, e che di conseguenza va esclusa qualsiasi responsabilità dell'Ateneo stesso per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
- 7) del fatto che la copia consegnata in formato cartaceo, controfirmata dal tutor, depositata nell'Archivio di Ateneo, è l'unica alla quale farà riferimento l'Università per rilasciare, a richiesta, la dichiarazione di conformità di eventuali copie.

Data 12 marzo 2022

Firma Monia Azzalini

Mod. TD-Lib-09-a 1

AUTORIZZO

- l'Università a riprodurre ai fini dell'immissione in rete e a comunicare al pubblico tramite servizio on line entro l'Archivio Istituzionale ad Accesso Aperto il testo integrale della tesi depositata;
- l'Università a consentire:
 - la riproduzione a fini personali e di ricerca, escludendo ogni utilizzo di carattere commerciale;
 - la citazione purché completa di tutti i dati bibliografici (nome e cognome dell'autore, titolo della tesi, relatore e correlatore, l'università, l'anno accademico e il numero delle pagine citate).

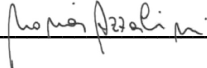
DICHIARO

- 1) che il contenuto e l'organizzazione della tesi è opera originale da me realizzata e non infrange in alcun modo il diritto d'autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta, né compromette in alcun modo i diritti di terzi relativi alla sicurezza dei dati personali;
- 2) che la tesi di dottorato non è il risultato di attività rientranti nella normativa sulla proprietà industriale, non è stata prodotta nell'ambito di progetti finanziati da soggetti pubblici o privati con vincoli alla divulgazione dei risultati, non è oggetto di eventuali registrazione di tipo brevettuale o di tutela;
- 3) che pertanto l'Università è in ogni caso esente da responsabilità di qualsivoglia natura civile, amministrativa o penale e sarà tenuta indenne a qualsiasi richiesta o rivendicazione da parte di terzi.

A tal fine:

- dichiaro di aver autoarchiviato la copia integrale della tesi in formato elettronico nell'Archivio Istituzionale ad Accesso Aperto dell'Università Ca' Foscari;
- consegno la copia integrale della tesi in formato cartaceo presso la segreteria didattica del dipartimento di riferimento del corso di dottorato ai fini del deposito presso l'Archivio di Ateneo.

Data 12 marzo 2022

Firma 

La presente dichiarazione è sottoscritta dall'interessato in presenza del dipendente addetto, ovvero sottoscritta e inviata, unitamente a copia fotostatica non autenticata di un documento di identità del dichiarante, all'ufficio competente via fax, ovvero tramite un incaricato, oppure a mezzo posta

Firma del dipendente addetto

Ai sensi dell'art. 13 del D.Lgs. n. 196/03 si informa che il titolare del trattamento dei dati forniti è l'Università Ca' Foscari - Venezia.

I dati sono acquisiti e trattati esclusivamente per l'espletamento delle finalità istituzionali d'Ateneo; l'eventuale rifiuto di fornire i propri dati personali potrebbe comportare il mancato espletamento degli adempimenti necessari e delle procedure amministrative di gestione delle carriere studenti. Sono comunque riconosciuti i diritti di cui all'art. 7 D. Lgs. n. 196/03.

Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente: Monia Azzalini

matricola: 956448

Dottorato: Lingue, culture e società moderne e Scienze del linguaggio

Ciclo: 34°

Titolo della tesi : Rappresentazioni di genere dentro e attraverso il linguaggio dell'informazione italiana

Abstract:

Questa tesi analizza le rappresentazioni di genere nel linguaggio dell'informazione italiana, con un approccio interdisciplinare che integra l'analisi linguistica con l'analisi mediale. Dopo una rassegna bibliografica ragionata in materia di "lingua e genere" e "media e genere", la tesi presenta tre analisi di *corpora*. La prima, *corpus-driven*, esplora la rappresentazione di donne e uomini in un *corpus* di notiziari televisivi, attraverso un'analisi testuale computerizzata. La seconda, *corpus-based*, studia, attraverso un'analisi lessicometrica integrata con un'analisi manuale, l'accordo fra genere grammaticale e genere semantico del lemma *ministra/o* in un *corpus* di notizie di quotidiani e nel *corpus* di notiziari televisivi sopra citato. La terza analizza le associazioni di genere latenti nei *word embeddings* di una selezione di nomi di professione estratti da un *corpus* di notizie quotidiane. Nel complesso i risultati dimostrano che il linguaggio dell'informazione italiana veicola ancora attivamente rappresentazioni di genere stereotipate, attestando però anche alcuni segnali di innovazione rispetto al passato.

This thesis focuses on gender representations in the Italian news language, using an interdisciplinary approach that integrates linguistic analysis with media analysis. It first provided an annotated bibliographic review on 'language and gender' and 'media and gender', and then it presents three *corpus* analyses. The first analysis is *corpus-driven* and explores the representations of women and men in a *corpus* of Italian television news, using a computer-assisted textual approach. Then a *corpus-based* lexicometric analysis is presented to investigate the use of the grammatical gender of the lemma *ministra/o* in a *corpus* that expands the above-mentioned *corpus* of Italian television news with a collection of Italian newspapers. The last analysis focuses on the gender associations in the *word embeddings* of a selection of nouns of professions extracted from a *corpus* of daily news. Overall, the results show that the Italian news language still actively spreads stereotypical gender representations, yet some results also indicate signs of innovation compared to the past.

Firma dello studente

