Language, Gender and Hate Speech
A Multidisciplinary Approach
edited by Giuliana Giusti and Gabriele Iannàccaro

Narrating Gender Violence
A Corpus-Based Study on the Representation of Gender-Based Violence in Italian Media

Lucia Busso
Aston University, Birmingham, UK

Claudia Roberta Combei
Alma Mater Studiorum – Università di Bologna, Italia

Ottavia Tordini
Università di Pisa, Italia

Abstract   Using a corpus-based and cross-modal approach, this study explores mediatic linguistic representations of gendered violence. Specifically, we analyse a journalistic corpus (WItNECS, Women in Italian Newspaper Crime Sections) and a multimedia dataset of the Italian TV program *Amore Criminale* (AC). The corpus is explored via collocational analysis of key terms and topic modelling. AC is investigated with a multimodal analysis of speech and gestures. Finally, we compare similarities and differences of newspaper and television language. Findings from this innovative methodology bring new evidence on the mediatic representation of 'women as victims'.

Keywords   Language and gender. Corpus linguistics. Topic modelling. Multimodal analysis. Media language.

Summary   1 Introduction. – 2 The Study. – 3 Corpus-Based Lexical Analyses. – 4 Inside *WItNECS*. – 4.1 The Definition and Characteristics of the Structural Topic Model. – 4.2 Data Processing for Structural Topic Model. – 4.3 The Exploration of Topics. – 5 Discourse and Gestures Analysis of *Amore Criminale.* – 6 Concluding Remarks.

39

1   Introduction

The social, cultural and political issue of gendered violence has been at the centre of a heated debate in recent years, thanks to global social campaigns (such as *#METOO*, or *ni una menos*) that have drawn more attention to the problem. Nevertheless, quantitative and statistical data on the phenomenon are still scarce. Suffice it to say that according to the European Institute for Gender Equality (EIGE), no European-level statistical analysis has been possible, due to complete absence of useful data (EIGE 2018). In Italy, few (and very recent) statistical data have been analysed (by ISTAT and by the Department for Equal Opportunities). Findings show that family contexts are the highest-risk situations for women: in 2014, 77% of female murder victims were in fact killed at the hands of a partner, a former partner or a family member.[1] Data from the 2018 survey by SDGS (Sustainable Development Goals) on spy crimes reports that in 2014-16 almost 80% of family maltreatment, 74% of persecutory acts and 89% of sexual violence concern a woman.[2] A new attention to the problem is also demonstrated by recent advances in the regulatory field: Law 23/4/2009, art. 612 bis c.p. introduces for the first time the crime of "persecutory acts" (stalking) in the Italian legal system.[3]

Notwithstanding the seriousness of such crimes and the recent global attention to the issue, Italian media and social debate still fails to truly grasp the extent of the problem. The same term "femicide", defined for the first time in Russell and Radford (1992) as the murder of a woman by a man because of her gender, is only first introduced in Italian in 2008 (Spinelli 2008), and is recorded by dictionaries the following year. Furthermore, Italy can be considered (following Formato 2019) a "fruitful epistemological site" (Sunderland 2004, 73), i.e. a place which is particularly interesting for the study of gender and language due to its idiosyncratic characteristics (for an in-depth discussion, see Formato 2019, 3 ff.). These crucial features of Italian sociocultural *milieu* include the cult of femi-

1   EURES 2016, cited in Abis, Orrù 2016.

2   Report SDGS 2018 – Informazioni statistiche per l'agenda 2030 in Italia. `https://www.istat.it/it/files//2018/07/SDGs.pdf`.

3   Law proposal 1455 reinforces the procedural safeguards of victims of sexual and domestic violence has recently been approved (April 2019). To date, there is also a penalty of up to six years and a €15,000 fine for revenge porn.

nine beauty and the ideal of the Mother as the "figure" (Auerbach, [1973]'s terms) of woman, all of which had "a profound sexual asymmetry in Italian society" (Gundle 2007, 266).

We propose an exploratory investigation of the portrayal of gendered crimes in Italian media. Specifically, we present a contrastive analysis of newspaper and television language. The data analysed are a corpus of crime news sections articles and a set of episodes from *Amore Criminale* (Criminal Love), a television show which narrates "the stories of women who through love met death [...] through reconstructions, documentary material and direct testimonies".[4] In other words, we propose an investigation of journalistic and TV language aimed at disentangling attitudes, stereotypes and conceptualizations underlying the contemporary sociocultural attitudes toward women and hate crimes perpetrated against them (Tabbert 2012).

The article is structured as follows. Section 2 outlines the theoretical background, the hypotheses and the data collection of the corpora used for subsequent analyses. Sections 3, 4 and 5 deal with the various types of analyses. Specifically, we first present a corpus-based lexical exploration (§ 3), followed by a description of the computational analysis performed using Structural Topic Modelling (§ 4). Both paragraphs refer to analyses conducted on written language. Section 5 introduces the multimodal and multimedia contrastive analysis performed on television spoken language. We conclude with Section 6, in which we discuss the significance of our findings.

2 The Study

In the last decade, a large body of literature employed corpora in gender linguistics (e.g. Mills, Mullany 2011; Johnson, Ensslin 2013; Baker 2014; Wodak 2015), but to date, very few such studies have addressed the Italian language (cf. Marcato, Thüne 2002; Busso, Combei, Tordini 2019; Formato 2019). Nevertheless, the potential of corpus-based analyses is unquestionable, since they allow large-scale explorations, useful for revealing linguistic and cultural schemes and traits from all sorts of data.

This exploratory study brings a contribution to the field of corpus-based content analysis by describing the language used by Italian media to talk about violence against women. We adopt a multimodal and multimedia approach on texts (i.e. newspaper articles) and audio-video material (i.e. interviews and docu-fiction) in order to unveil patterns of content, style and gestures in the description of gendered

---

4 From the description of the show, our translation.

violence in newspapers and television shows. Unlike Abis and Orrù (2016) and Formato (2019), our study combines corpus-based lexical explorations, Structural Topic Model techniques (Roberts et al. 2014) and metaphor and gesture analysis. We believe that this combined exploratory approach might be beneficial in addressing the complex issue of how language is used to depict gendered violence.

Our expectations regard both content and style. On the one hand, we expect that newspapers will abound in articles dedicated to femicide perpetrated by the victims' partners; this would act as a proxy for illustrating that this phenomenon has escalated in recent years in Italy. On the other hand, we believe that the audio-video material will be rich in metaphorical content. This will also allow us to briefly compare the two media.

We perform our analyses on a specialized corpus of journalistic language (WItNECS, Women in Italian Newspaper Crime Sections) **[Tab. 1]** and a multimedia and multimodal database for the Italian television series *Amore Criminale* (AC) **[Tab. 2]**. WItNECS was collected manually, choosing as a starting date 13 September 2016, that symbolically matches the suicide of Tiziana Cantone – an Italian victim of revenge porn. The data collection process continued for over 9 months and it consisted in downloading and storing all news regarding violence against women in Italy and the metadata of the articles in individual.txt files. In order to ensure balancedness and representativity to our data, we collected articles from four national newspapers: *Corriere della Sera* (CS), *La Stampa* (LS), *Il Fatto Quotidiano* (IFQ), and *La Repubblica* (R). For the latter, we also included regional editions from Milan (RRM), Florence (RRF), Naples (RRN), and Palermo (RRP), in order to render the data geographically comparable. WItNECS was automatically compiled on Sketch Engine (Kilgarriff et al. 2014).

Table 1 WItNECS statistics

WItNECS	
TOKENS: 271,506	LANGUAGE: Italian
WORDS: 234,178	ENCODING: UTF-8
SENTENCES: 10,556	
	TIMEFRAME: 2016-09-13 to 2017-06-30
DOCUMENTS: 614	

Table 2 AC statistics

AC	
TYPE: Multimodal and multimedia	LANGUAGE: Italian
FORMAT: videos, orthographic transcriptions, gesture and language annotations	ENCODING: UTF-8
CONTENT: docu-fiction, interviews, court records	8 episodes (2015-16; 2016-17)

Our multimedia database (AC) is built around 8 episodes of the TV show *Amore Criminale*.[5] We chose this show not only because it presented real instances of femicide, but also because it was rich in terms of communicative styles and linguistic registers. The AC database consists in 8 episodes aired during the 2015-16 and 2016-17 seasons, as well the orthographic transcription and the multilevel annotation of gestures (see § 5, for a detailed description of the transcription and annotation process). Each episode deals with a different crime and it contains interviews with real people, reconstructions of specific events (i.e. docu-fiction), as well as the presentation of court and legal records.

3 Corpus-Based Lexical Analyses

As mentioned in § 2, we expect crime news articles to convey significant information which reflect communicative strategies, latent ideologies, and general public opinion on the topic of gendered violence. We explore such a hypothesis with a first exploration of the corpus with the online software *SketchEngine* (Kilgarriff et al. 2014), which allows accurate quantitative surveys using fine-grained statistical measures (e.g. association scores, keyness). The corpus was subject only to a mild pre-processing at this stage: numbers, excess white spaces, URLs, hashtags and other non-linguistic symbols were stripped. In fact, the analyses presented are conducted with the *Wordsketch tool*, which allow to select the grammatical relations of interest, therefore excluding stop-words and non-interesting elements at a later stage in the analysis.

---

[5] *Amore Criminale* is an Italian television programme broadcast on Rai Tre. Each episode recounts a love story ended with a murder. The narrations combine various communicative styles: interviews (with the victims' families, friends, and colleagues, with policemen, magistrates and lawyers), docu-fiction (filmed reconstructions) and original documentary material. The presenter introduces and summarizes the story.

For reasons of space, we do not dwell on the detailed lexical analysis (for further details see Busso, Combei, Tordini 2019), but we present two instances of analyses that exemplify well two main trends observed in journalistic language: the overwhelming presence of gendered violence in the domestic environment, and the implicit tendency of mainstream media outlets to exonerate the perpetrator of the crime. Specifically, we analyse the grammatical and collocational behaviour of three relevant terms: "donna", 'woman'; "marito", 'husband'; "moglie", 'wife' and "amore", 'love'.

Figure 1 depicts the Wordsketch of "donna". Wordsketches summarize a word collocational behaviour, using the logDice score, and allows to highlight different grammatical relations between collocates. The dynamic text size relates to frequency. In Figure 1, the relations depicted are modifiers (pink), noun modifiers (green), and nouns occurring with adversative or copulative conjunctions (blue). Some interesting considerations already surface. As we can see from Figure 1, women have collocates which point both to the violence and the abuse: "uccisa", 'killed'; "vittima", 'victim'). Furthermore, women are depicted by their age ("giovane", 'young'; "ragazza", 'girl') or nationality ("straniera", 'foreign'). Interestingly, women are grammatically related to men ("uomo"), children ("bambino"). They are victims and mothers ("madre"). Interestingly, we find no modifier relating to either job descriptions, or personal characteristics not related to external appearance (such as age and nationality).
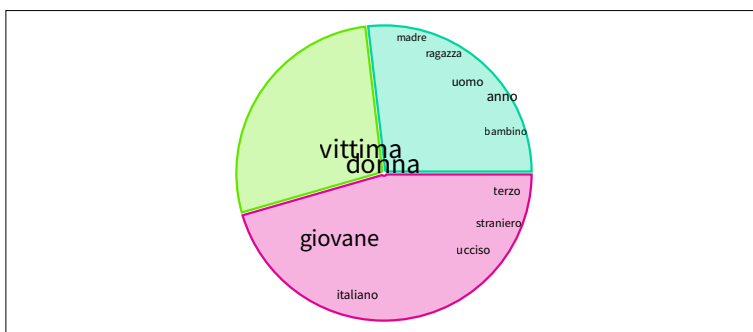

Figure 1 Wordsketch of the term "donna"

We also present the Wordsketch difference of 'husband' and 'wife', as they represent the most frequent terms to denote perpetrator and victim after the generic terms 'woman' ("donna", freq.: 2037, freq. pmw: 5976.27; "moglie", freq.: 358, freq. pmw: 1050,3) and 'man' ("uomo", freq.: 1017, freq. pmw: 2983.7; "marito", freq.: 341, freq. pmw: 1000). This datum corroborates our hypothesis: not only women are depicted as mothers, or in relation to a man, but also a prevalence of fam-

ily-related environments seems to emerge. Figure 2 represents the Wordsketch difference between the two terms, set on the grammatical relations of "verbs with 'moglie/marito' as objects".
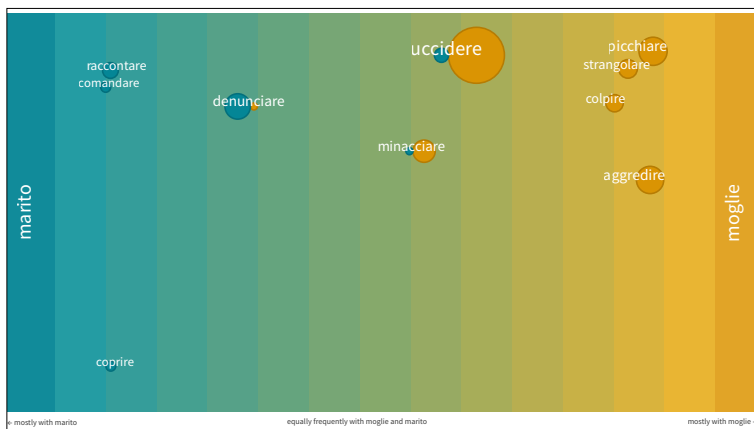

Figure 2 Wordsketch difference of verbs with "moglie/marito" as objects

As we can see, wives tend to appear more as direct objects. The verbs displayed are all actional, confirming the agentive role of the husband as a perpetrator of different kinds of violence ("uccidere", 'to kill'; "aggredire", 'to assault').

Notwithstanding the clear agentivity of the actions perpetrated by men and husbands, we can make some noteworthy observations by analysing the adjectival collocates of the term "amore", which in Italian is often used as a metonymy for relationship. Since most episodes of violence occur between partners, in fact, we considered it an interesting linguistic cue to explore how newspaper address the issue of violent relationships, a widely spread sociological phenomenon which often result in abuse, stalking or murder. The lexical collocates reveal that violent relationships are frequently not considered as a criminal, abusive situation, but rather 'justified' in various ways. The love is metaphorically defined as "malato" (sick). The metaphor, poetic as it may be, is however extremely dangerous: by equating a violent relationship to an illness, the perpetrator of the violence appears not to be held responsible for his actions, as violence is merely a symptom of an uncontrollable disease.

We also find that relationships are "sfortunato" (unlucky) or "eterno" (eternal). Both these terms are again extremely problematic, as they tend to put the focus on external elements (luck) or to reinforce the belief that despite everything love (especially marriage) should be eternal.

Love is also defined as "sbagliato" (wrong) or "criminale" (criminal). Although these terms are more appropriate to describe abuse

and violence, yet again we find a shift of focus from the perpetrator to the general relation of love, which disguise the responsibility of the agent.

Figure 3 Wordsketch of the term "amore"

In sum, the main findings of the exploration of the lexicon of WIt-NECS can be summarized as follows: gender violence mostly occurs in the domestic environment; for this reason – perhaps – women are defined as "in relation to" a man. Moreover, female victims are qualified with adjectives relating to their age or nationality, rather than with personal traits or professional titles. Although wives are also found to appear principally as direct objects, husbands (and men)'s agentivity is found to be disguised with various (implicit) strategies (for a more detailed discussion; see the parallel study Busso, Combei, Tordini 2019).

This first corpus-driven exploration of our data allows for a preliminary understanding of collocational behaviour and general corpus structure. For a more in-depth analysis, we employ computational techniques for a better understanding of hidden patterns inside our corpus.

4   Inside *WItNECS*

In the last decade, an exponential interest in text mining has been reported worldwide (Sinoara, Antunes, Rezende 2017). This might be due to the enormous availability of texts on the web – often unstructured – and to the ongoing communication shift from traditional media to online content. For these reasons, various machine-assisted approaches for the analysis of large corpora have been proposed so far (see Watanabe, Zhou 2020 for a review). In an attempt to bring a contribution to content analysis, our study employs the Structural Topic Model technique (Roberts et al. 2014) to inductively identi-

fy semantic themes and issues on which Italian journalists focus in their crime articles regarding violence against women. This allows us to verify one of our expectations, namely that most articles will report femicide perpetrated by the victims' partners. Unlike Busso, Combei, Tordini (2019), this paper also explores the variation in issue attention across the newspapers considered. This enables us to capture communicative intentions and topic patterns for each newspaper. In § 4.1 the Structural Topic Model technique is presented; in § 4.2 we describe how pre-processing was applied to our data; finally, in § 4.3 we present our model and comment the topics.

4.1 The Definition and Characteristics of the Structural Topic Model

The Structural Topic Model (STM) is a generative model of word counts with document-level covariate information (Roberts et al. 2014; Roberts, Stewart, Airoldi 2016). Covariates are measurable variables that have a statistical relationship with the dependent (response) variable; hence, a covariate may predict or explain a dependent variable.

STM is part of the unsupervised learning methods that use modelling assumptions, text properties, and covariates to estimate topics within a corpus of often unstructured material and to organize it according to word co-occurrences. In the STM, and in similar frameworks such as Latent Dirichlet Allocation (Blei, Ng, Jordan 2003), a topic is a mixture over words where each word has a probability of belonging to a topic, while a document is a mixture over topics. One of the major novelties of the STM is the fact that the prior distribution of topics may be varied as a function of covariates (Roberts et al. 2014; Roberts, Stewart, Airoldi 2016). Covariates may enhance the interpretability of topics. Moreover, incorporating covariates in topic models improves inferences and enables the exploration of relationships among variables in regression-like schemes. This feature allows researchers to explain topical prevalence,[6] topical content or both, as a function of the relevant variables.

In the last couple of years, the STM has been greatly employed in social and political science research (Curry, Fix 2019; Mourtgos, Adams 2019; Pinto 2019; Combei, Giannetti 2020), but to the best of our knowledge, in linguistics it is still unused (Brookes, McEnery 2019; Combei 2019; Busso, Combei, Tordini 2019).

---

6 We use the topical prevalence feature in this work.

4.2 Data Processing for Structural Topic Model

STM is based on the 'bag of words' approach; this means that at the document level (i.e. the single newspaper article, in our case), every word is treated as a unique feature, and then its co-occurrences are computed. At the corpus level (i.e. all newspaper articles, in our case), words are systematized into a term document matrix, with each document representing a matrix column, and each word representing a matrix row.

After having built the term document matrix and in order for STM to be applied to the newspaper articles, several pre-processing operations were necessary. This helped reducing the 'noise' in our data. After the text tokenization (i.e. identifying and separating each word), we defined an ad hoc 'stop words' list for Italian, consisting of 1300 lexically empty or uninformative words (i.e. prepositions, conjunctions, auxiliary verbs, determiners etc.). Then, we cleaned and normalized the text with the pre-processing functions available on the *quanteda* (Benoit et al. 2018), *tm* (Feinerer, Hornik, Meyer 2008), and *qdapRegex* (Rinker 2017) R packages (R Core Team 2018). Hence, numbers, punctuation and one-character sequences were deleted. Additionally, we converted all letters to lowercase. Finally, following similar research (Banks et al. 2018; Mourtgos, Adams 2019), we opted for a frequency threshold of 1; this means that a word that appeared only once in one document was dropped.

The corpus generated with the "textProcessor" and "prepDocuments" functions in STM consisted of 77,087 single word forms and 17,840 lemmas. By applying the "searchK" function, we were able to perform several tests, such as held-out likelihood, residual analysis and semantic coherence, in order to establish the correct number of topics. According to these three tests, the 20-topics model had high levels of held-out likelihood and of semantic coherence, as well as low residuals. STM also gave us the possibility to set the type of initialization, therefore we used the spectral initialization, since its stability and consistence had already been proven (Roberts, Stewart, Tingley 2016; Combei 2019). We chose the highest probability as a word profile. Previous studies have shown that the stemming function available for STM does not perform well on Italian (Combei, Giannetti 2020); as a result, we used word forms instead of stemmed words in the model.

To use the name of the newspaper as a covariate for the topical prevalence, the information was coded in the file names during the data collection and further extracted and processed with R (e.g. from "CS_2016_12_14.txt" we obtained the newspaper name "*Corriere della Sera*" and the date "2016-12-14").

4.3 The Exploration of Topics

The topics that emerged from the STM model on WItNECS are shown in Figure 4. Each topic was assigned a label after the examination of their most semantically coherent words and after the analysis of the most exemplar documents, namely newspaper articles that had the highest proportion of words associated with the specific topic.

The results of our empirical observation display interesting societal, linguistic, and journalistic insights. Firstly, based on the topic proportions and the word probabilities of the STM model, it is clear that the most prevalent topic in our corpus regarded domestic violence and uxoricides (topics: 1, 4, 6, 10, 14, 16, 18), with an overall expected proportion of 38%. Words such as "marito", "coppia", "relazione", "famiglia", "fidanzato", "moglie" have high values of probabilities.[7] This finding fully confirms our first expectation, indirectly suggesting that this phenomenon has seen a rapid growth in recent years in Italy.

News describing cases of sexual harassment were second in terms of their presence in newspapers, as suggested by the high proportion of topics (i.e. 21%) related to this issue (topics: 5, 7, 9, 13, 20). On a different note, topics 8 and 17 (expected topic proportion: 13%) suggest that various articles also delineated crime scenes, giving deictic information on the crime setting and its participants, through words from the semantic field of crime (e.g. "corpo", "cadavere", "fuoco", "coltello" etc.).[8] Next, topics 3 and 19 (expected topic proportion: 12%) concerned news on arrests, accusations and sentences for crimes of violence against women; most terms in these topics belong to the semantic field of law (e.g. "gip", "custodia cautelare", "corte", "pena", "condanna" etc.).[9] Additionally, topics 2 and 12 regarded initiatives of helping victims and preventing femicides (estimated topic proportion: 7%). Topics 11 and 15 (estimated topic proportion: 9%) revealed the attention newspapers dedicated to new media and revenge porn (e.g. "video hard", "social", "facebook", "diffusione", "rete" etc.).[10]

As one of the aims of the STM model was to reveal communicative intentions and topic patterns for each newspaper, we also computed the estimated effects for the covariate 'newspaper'. The gamma matrix and the "findThoughts" function of the STM package allowed us, on the one hand, to examine how topics were distributed in each

---

[7] Translation into English: 'husband', 'couple', 'relation', 'family', 'fiancé', 'wife'.

[8] Translation into English: 'body', 'corpse', 'fire', 'knife'.

[9] Translation into English: 'magistrate in charge of preliminary investigations', 'pre-trial detention', 'court', 'punishment', 'sentence',

[10] Translation into English: 'pornographic video', 'social media', 'Facebook', 'sharing', 'network'.
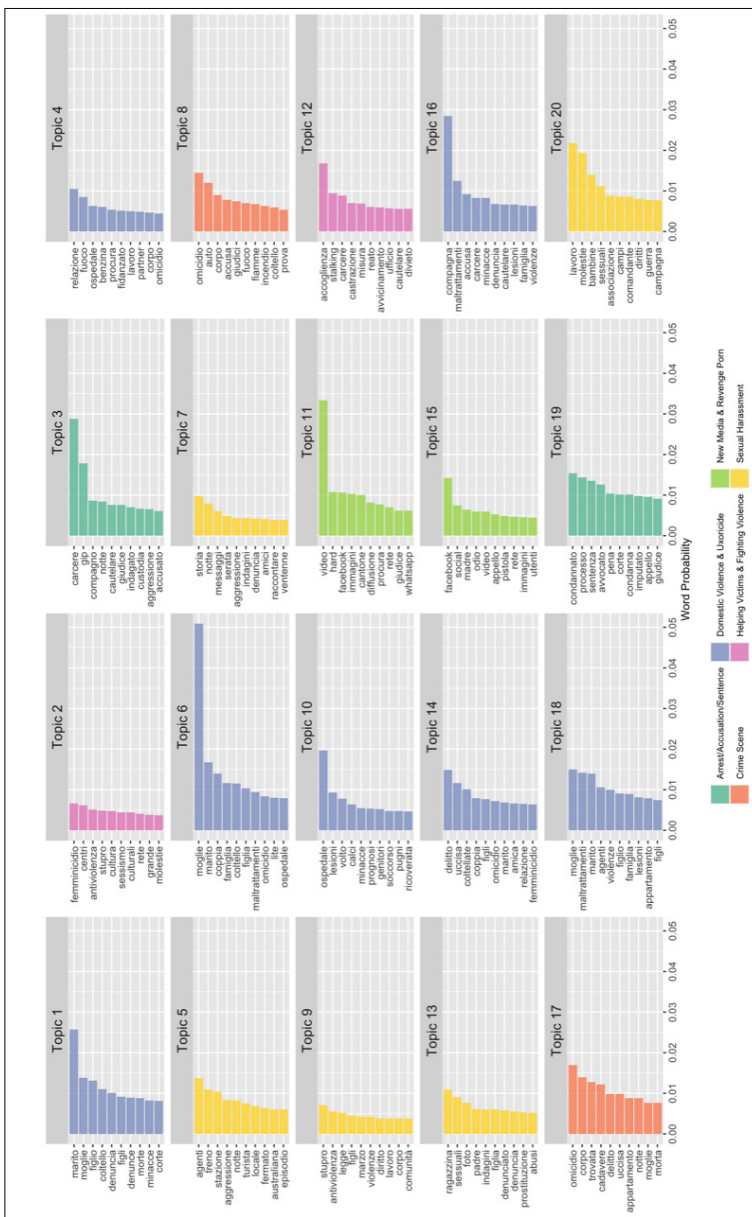
Figure 4 Topics in WItNECS

newspaper, and on the other hand, to assess the goodness of our model. Regarding the topic distribution, several tendencies, although not particularly strong, were revealed. We will mention only results that reached statistical significance (p-values < 0.05 in the regression-like scheme). Firstly, initiatives of helping victims and preventing femicides were reported often by IFQ and RRN (e.g. an IFQ article on 5 March 2017). Crimes involving new media and revenge porn were more frequent in CS and RRN (e.g. a CS article on 14 January 2017). Next, topics related to domestic violence had high expected proportions in LS, RRM, and RRP (e.g. a LS article on 11 April 2017). Topics regarding arrests and sentences for crimes against women were more frequently reported in LS (e.g. a LS article on 11 March 2017). Also, sexual harassment crimes were reported more often in IFQ (e.g. an IFQ article on 13 March 2017). Finally, the national and regional editions of *Repubblica* dedicated more space than other newspapers to the description of crime scenes (e.g. a R article on 1st May 2017).

5   Discourse and Gestures Analysis of *Amore Criminale*

Along with lexical and computational investigations, we also carried out a multimodal analysis on the audiovisual material extracted from the AC database (see § 2), with the purpose to offer a more thorough representation of 'women as victims' in Italian crime news. The novelty of this approach consists in the fact that this popular TV show has never been object of a linguistic research, albeit it can offer crucial insights into the narration and perception of everyday gender violence. To explore the AC content, we employed the software ELAN, a tool for multimodal transcription and processing developed by the Max Planck Institute for Psycholinguistics[11] (for further details, cf. Sloetjes 2014). This method was chosen to provide a coherent methodological basis for the annotation and examination of audio and video resources and the interpretation of the multimedia content. In fact, ELAN allows the researcher to create and manipulate multiple annotations on different 'tiers', namely multiple levels of analysis in which both linguistic and metalinguistic information can be segmented and labelled. As methodological framework for the transcription and the analysis of multimodal data, we adopted the method developed by Magno Caldognetto et al. (2004). According to this system, which is called "partitura", communicative acts can be processed either separately or in their mutual interaction (see Busso, Combei, Tordini 2019).

For the purposes of this work, we followed the norms for orthographic transcription and coding originally provided by McNeill

---

11   Available at `https://tla.mpi.nl/tools/tla-tools/elan`.

(1992), and subsequently adapted in more recent studies, for instance Poggi (2007) and Kong et al. (2015) (see Wagner, Malisz, Kopp 2014 for an overview). This classification method was employed to label either the gestures of the participants (presenter, actors, victims/relatives in interviews), and the content of their oral productions (for example, M-speech act). Specifically, we employed the following tags:

- *I-gesture* for iconic gestures that shape the form of an object and/or try to reproduce a dynamic movement (i.e., the act of miming a real scene);
- *M-gesture* for metaphoric gestures, which convey and abstract idea to conceptualize a concrete experience;[12]
- *D-gesture* for deictic gestures, used by the speaker to indicate a specific element in the physical communicative context (i.e., pointing finger);
- *E-gesture* for emblematic gestures, i.e. with 'standard'/shared properties (for instance, gestures which are not language-specific, such as the thumb-index gesture for "OK").

In total, we analysed about 16 hours of the TV show, and tagged 116 observations. The content and percentages resulting from these classifications are shown in Figure 5.

Results show that 48.7% of communicative acts[13] in the AC database have a metaphorical content. If we look at the correspondences with the audiovisual material, we note that metaphors are specifically employed to describe the behaviour of the subjects – both men and women – involved in the events. These metaphors are drawn from various semantic domains, for example: nature/animals/plants ("era un parassita", 'he was a parasite'; "scatta la caccia all'assassino", 'the killer hunt starts'; "Francesca era un leone […] una combattente", 'Francesca was a lion […] a fighter'; "il fuoco della gelosia", 'the jealousy fire'), perception ("lento sprofondare nel mondo oscuro del ragazzo che amava", 'slowly sinking into the dark world of the man she loved'; "refrattario ad ogni luce di umanità", 'impervious to any light of humanity'), war ("eravamo tutti nel suo mirino", 'we were all in his gunsight'; "assediata da un uomo", 'besieged by a man'), or technology and games ("piano piano […] ho cominciato a vederla come un avversario", 'Little by little, I was starting to see her as a rival'; "era come se avessi combattuto in un videogame", 'it was as if I fought in a videogame'; "per me lei era solo una pedina", 'she was just a pawn to me').

---

12   As argued by Chui: "without concomitant linguistic representations, metaphorical thoughts cannot be interpreted without perceiving the manual configurations" (2011, 448).

13   For both iconic and metaphoric acts, we summed the total of *speech acts* and *gestures* (for instance, 42.6% *M-speech* + 6.1% *M-gestures*).
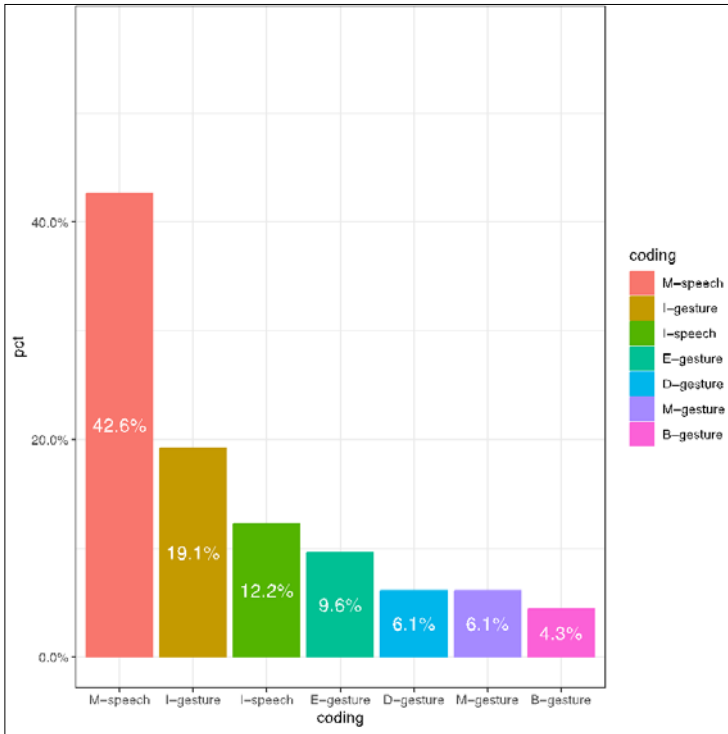
Figure 5 Percentages resulting from the gesture-speech tagging

As can be inferred from some of the examples presented here, some episodes of AC (season 2016-17) include interviews to men who had previously perpetrated violence on women (i.e. "uomini maltrattanti", 'abusing men'), and who are currently being treated in specialized centres. Specifically, they are asked to report (anonymously) their experience as serial aggressors and to describe the escalation of brutality against their partners. Interestingly, we note that they adopt recurrent schemes that might suggest a displacement of violence outside their own moral dimension. That is, these men would undergo *raptus* (such as "rabbia cieca", 'blind rage') that temporarily take over they consciousness, making them unable to control their behaviour and to act responsibly. Moreover, most of them report to have faced past traumas as victims of violence, which left a permanent trail on their personality. From the interviews, we deduce that they are deeply aware of what they have inflicted on their victims. In fact, they have decided to undergo psychological treatments in specialized rehabilitation centres. Nonetheless, the overall impression is that men are still – unconsciously – looking for some external justification of their viciousness.

Overall, gestures revealed to be the most significant strategy to convey the *authentic* content of the AC show, namely the characterization at socio-psychological level of both victims and "uomini maltrattanti". It is also worth noting the meaningful presence of *I-gestures* (19.1%), combined with *I-speech acts* (12.2%). Iconic pragmatic strategies are employed either by victims (if present) and by the presenter/interviewer to describe or mime acts of violence, (for instance: "avevo un occhio così", 'I had an eye like that'; "mi acchiappa per i capelli e mi tira fuori", 'he grabs me by my hair and pulls me out'). In some cases (6.1%), *I-gestures* are supported by *D-gestures* ("ero tutta piena di lividi […] qua […] dappertutto", 'I was all full of bruises […] here […] everywhere').

In light of these observations, we can infer that the representations of gendered violence in AC and WItNECS exhibit visible differences, yet also share some commonalities. First, it is evident that the nature of the TV show allows to provide a socio-psychological portrait of the actors, which on the contrary is absent in newspapers. Also, the spectacularization of the events is widespread in AC, as many details of the narration are stressed through adjectives. Besides, when describing both aggressors and victims, metaphoric speech is the most pervasive among the various communicative strategies. At the same time, the analyses conducted on both media show that the majority of crimes are committed by someone within the victim's domestic circle. Furthermore, the aggressor is often represented as someone who may have had no control of his own actions. In particular, equating "violent love" as consequence of a disease (for instance, "raptus") (see § 3), seems to suggest that – since he was sick – the culprit is to some extent less guilty.

6   Concluding Remarks

To conclude, we presented an exploratory cross-modal study of newspaper and TV language concerning the sociological phenomenon of gendered violence, through quantitative and qualitative analyses of an ad hoc compiled newspaper corpus and multimodal database. The journalistic corpus was firstly explored to reveal the collocational behaviour of several relevant terms. This first step was essential to a preliminary understanding of the corpus and of some of its patterns. It was found that women tend to be described as 'in-relation-to' their family members, based on their age and nationality, and never with professional titles. Furthermore, the collocations of the term "amore" (love) seemed to suggest an implicit exoneration of the culprit in favour of an 'external' causality: violence is a sickness, an unlucky circumstance, a mistake. To refine our preliminary analysis, we used the computational STM technique for in-depth quantitative explora-

tions. The STM allowed us to display and organize the content of the over 600 crime news articles of WItNECS. Besides revealing cues on the stylistic simplicity and objectivity, the major finding of this analysis, which confirms our expectations, is that most crimes are committed by someone in the victim's domestic circle (e.g. husband, exboyfriend etc.). Most importantly, it is to note that such quantitative analysis – that fully confirms data provided by the EIGE, ISTAT and the Department for Equal Opportunities – has never been so far performed. Regarding the distribution of topics in the newspapers under consideration, very few patterns were traceable: news depicting crime scenes were more recurrent in the national and regional editions of *Repubblica*; crimes regarding the distribution of revenge porn through new media appeared more often in CS and in RRN; finally, the issue of domestic violence was more frequent in LS, RRM, and RRP.

The multimodal analysis on the AC database allowed us to directly compare the two different mediatic languages and communicative strategies. In particular, it shed light on a large variety of communicative patterns, which highly depend on the focus of the interview/presentation. First, it is to note that metaphors (in both discourse and gestures) are frequently used to sketch the socio-psychological portrait of the actors, i.e. "uomini maltrattanti" and victims. Undoubtedly, this type of strategy contributes to emphasizing the spectacularization of the events presented in the TV show, in contrast with the descriptive approach that traditionally characterizes newspapers. Second, iconic speech and gestures are also frequently adopted by both victims and interviewees to report, testify and/or mime episodes of violence – either physical or psychological.

In conclusion, our data confirm the results of the surveys conducted by ISTAT and by the Department for Equal Opportunities (§ 1): women are harassed or killed mostly by men they know well. Also, the contrastive analysis revealed a striking similarity between AC and WItNECS. In both corpora, the gravity of violence seems to be somehow attenuated by external circumstances, or by internal factors, i.e. a temporary illness affecting the aggressor (*raptus*). Moreover, we also find a crucial difference between the representation of gender violence in WItNECS and AC. While – for its idiosyncratic nature – the TV show also reports on psychological violence and domestic abuse, crime news is principally concerned with femicide.

Bibliography

Abis, S.; Orrù, P. (2016). "Il femminicidio nella stampa italiana. Un'indagine linguistica". *Gender/sexuality/Italy*, 3, 18-33.

Auerbach, E. (1973). *Figura. ‖ 1938. Auerbach. Scenes from the Drama of European Literature*. Gloucester (MA): Peter Smith.

Baker, P. (2014). *Using Corpora to Analyse Gender*. London: Bloomsbury.

Banks, G.C. et al. (2018). "A Review of Best Practice Recommendations for Text Analysis in R (and a User-Friendly App)". *Journal of Business and Psychology*, 33(4), 445-59. https://doi.org/10.1007/s10869-017-9528-3.

Benoit, K. et al. (2018). "quanteda. An R Package for the Quantitative Analysis of Textual Data". *Journal of Open Source Software*, 3(30), 774. https://doi.org/10.21105/joss.00774.

Blei, D.M.; Ng, A.Y.; Jordan, M.I. (2003). "Latent Dirichlet Allocation". *Journal of Machine Learning Research*, 3(3), 993-1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Brookes, G.; McEnery, T. (2019). "The Utility of Topic Modelling for Discourse Studies. A Critical Evaluation". *Discourse Studies*, 21(1), 3-21. https://doi.org/10.1177/1461445618814032.

Busso, L.; Combei, C.R.; Tordini, O. (2019). "La rappresentazione lessicale della violenza di genere. 'Donne come vittime' nei media italiani". Aldinucci, B. et al. (a cura di), *Parola. Una nozione unica per una ricerca multidisciplinare*. Siena: Edizioni dell'Università per Stranieri di Siena, 261-79.

Busso, L.; Vignozzi, G. (2017). "Gender Stereotypes in Film Language. A Corpus-Assisted Analysis". Basili, R.; Nissim, M.; Satta, G. (eds), *Proceedings of CLiC-it 2017*. Rome: CEUR-WS, 71-6.

Chui, K. (2011). "Conceptual Metaphors in Gesture". *Cognitive Linguistics*, 22(3), 437-58. https://doi.org/10.1515/cogl.2011.017.

Combei, C.R. (2019). "Is '*manovra*' really '*del popolo*'? Linguistic Insights into Twitter Reactions to the Annual Italian Budget Law". Bernardi, R. et al. (eds), *Proceedings of the Sixth Italian Conference on Computational Linguistics. CLiC-it 2019* (Bari, 13-15 November 2019). Aachen: CEUR-WS, 1-8.

Combei, C.R.; Giannetti, D. (2020). "The Immigration Issue on Twitter Political Communication. Italy 2018-2019". *Comunicazione Politica*, 21(2), 231-64. https://doi.org/10.3270/97905.

Curry, T.A.; Fix, M.P. (2019). "May It Please the Twitterverse. The Use of Twitter by State High Court Judges". *Journal of Information Technology & Politics*, 16(4), 379-93. https://doi.org/10.1080/19331681.2019.1657048.

EIGE, European Institute for Gender Equality (2018). *Improving Police and Justice Data on Intimate Partner Violence Against Women in the European Union*.

EURES (2016). *III Rapporto su caratteristiche, dinamiche e profili di rischio del femminicidio in Italia*. http://www.quotidianosanita.it/allegati/allegato4803121.pdf.

Feinerer, I.; Hornik, K.; Meyer, D. (2008). "Text Mining Infrastructure in R". *Journal of Statistical Software*, 25(5), 1-54. https://doi.org/10.18637/jss.v025.i05.

Formato, F. (2019). *Gender, Discourse and Ideology in Italian*. Cham: Palgrave Macmillan.

Gundle, S. (2007). *Bellissima. Feminine Beauty and the Idea of Italy*. Lexington (KY): University Press of Kentucky.

Jamet, D. (2010). "What do Internet Metaphors Reveal about the Perception of the Internet?". *Metaphorik.de*, 18(2), 17-32.

Johnson, S.; Ensslin, A. (2013). "'But Her Language Skills Shifted the Family Dynamics Dramatically'. Language, Gender and the Construction of Publics in two British Newspapers". *Gender and Language*, 1(1), 229-54. `https://doi.org/10.1558/genl.v1i2.229`.

Kilgarriff, A. et al. (2014). "The Sketch Engine. Ten Years on". *Lexicography*, 1(1), 7-36. `https://doi.org/10.1007/s40607-014-0009-9`.

Kong, A.P. et al. (2015). "A Coding System with Independent Annotations of Gesture Forms and Functions during Verbal Communication. Development of a Database of Speech and GEsture (DoSaGE)". *Journal of Nonverbal Behaviour*, 39(1), 93-111. `https://doi.org/10.1007/s10919-014-0200-6`.

Magno Caldognetto, E. et al. (2004). "La 'partitura'. Un sistema di annotazione multilivello di interazioni multimodali basata su ANVIL". De Dominicis, A. et al. (a cura di), *Atti delle XIV Giornate di Studio del Gruppo di Fonetica Sperimentale (GFS)*. Roma: Esagrafica, 253-8.

Marcato, G.; Thüne, E.M. (2002). "Gender and Female Visibility in Italian". Hellinger, M.; Bussman, H. (eds), *Gender Across Languages. The Linguistic Representation of Women and Men*. Amsterdam: John Benjamins, 188-217.

McNeill, D. (1992). *Hand and Mind. What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Mills, S.; Mullany, L. (2011). *Language, Gender and Feminism. Theory, Methodology and Practice*. London: Routledge.

Mourtgos, S.M.; Adams, I.T. (2019). "The Rhetoric of De-Policing. Evaluating Open-Ended Survey Responses from Police Officers with Machine Learning-Based Structural Topic Modeling". *Journal of Criminal Justice*, 64(1), 61-73. `https://doi.org/10.1016/j.jcrimjus.2019.101627`.

Phillips, M. (1985). *Aspects of Text Structure. An Investigation of the Lexical Organisation of Text*. Amsterdam: North-Holland.

Pinto, L. (2019). "Structural Topic Model per le scienze sociali e politiche". *POLIS*, 33, 163-74. `https://doi.org/10.1424/92923`.

Poggi, I. (2007). *Mind, Hands, Face and Body. A Goal and Belief View of Multimodal Communication*. Berlin: Weidler.

R Core Team (2018). *R. A Language and Environment for Statistical Computing*. Wien: R Foundation for Statistical Computing. `https://www.r-project.org`.

Rinker, T.W. (2017). *qdapRegex. Regular Expression Removal, Extraction, and Replacement Tools*. 0.7.2. Buffalo, New York: University of Buffalo.

Roberts, M.E. et al. (2014). "Structural Topic Models for Open-Ended Survey Responses". *American Journal of Political Science*, 58(4), 1064-82. `https://doi.org/10.1111/ajps.12103`.

Roberts, M.E.; Stewart, B.M.; Airoldi, E.M. (2016). "A Model of Text for Experimentation in the Social Sciences". *Journal of the American Statistical Association*, 111(515), 988-1003. `https://doi.org/10.1080/01621459.2016.1141684`.

Roberts, M.E.; Stewart, B.M.; Tingley, D. (2016). "Navigating the Local Modes of Big Data: The Case of Topic Models". Alvarez, R.M. (ed.), *Computational Social Science. Discovery and Prediction* (*Analytical Methods for Social Research*). Cambridge: Cambridge University Press, 51-97. `https://doi.org/10.1017/CBO9781316257340.004`.

Russell, D.; Radford, J. (1992). *Femicide. The Politics of Woman Killing*. New York: Twayne.

Sinoara, R.A.; Antunes, J.; Rezende, S.O. (2017). "Text Mining and Semantics. A Systematic Mapping Study". *Journal of the Brazilian Computer Society*, 23(1), 1-20. `https://doi.org/10.1186/s13173-017-0058-7`.

Sloetjes, H. (2014). "ELAN. Multimedia Annotation Application". Durand, J. et al. (eds), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press, 305-20. `https://doi.org/10.1093/oxfordhb/9780199571932.013.019`.

Spinelli, B. (2008). *Femminicidio. Dalla denuncia sociale al riconoscimento giuridico internazionale*. Milano: FrancoAngeli.

Sunderland, J. (2004). *Gendered Discourses*. London: Palgrave Macmillan.

Tabbert, U. (2012). "Crime through a Corpus. The Linguistic Construction of Offenders in the British Press". Gregoriou, C. (ed.), *Constructing Crime*. London: Palgrave Macmillan, 130-44.

Wagner, P.; Malisz, Z.; Kopp, S. (2014). "Gesture and Speech in Interaction. An Overview". *Speech Communication*, 57, 209-32. `https://doi.org/10.1016/j.specom.2013.09.008`.

Watanabe, K.; Zhou, Y. (2020). "Theory-Driven Analysis of Large Corpora. Semi-supervised Topic Classification of the UN Speeches". *Social Science Computer Review*, 2020(1), 1-21. `https://doi.org/10.1177/0894439320907027`.

Wodak, R. (2015). "Gender and Language. Cultural Concerns". Smelser, N.J.; Baltes, P.B. (eds), *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam: Elsevier, 698-703.