

3. Looking for Lexical Signatures in *Gomorrah*

Paola Dalla Torre, Paolo Fantozzi,
Maurizio Naldi and Placido Pellegriti

◀ ABSTRACT

Gomorrah is an Italian crime drama TV series that has been appreciated worldwide, being sold in 190 countries, despite its wide use of Neapolitan dialect, hardly understandable without the use of subtitles even for most Italians. Scholars immediately approached the study of this serial phenomenon, analysing it from different points of view, framing it within the broader context of studies on the new Italian television and its serial products. Our approach to *Gomorrah*, taking these elements into account, adds a new perspective that concerns character recognition, an emerging branch of research, to associate dialogues with characters and identify the verbal features of characters. We have then chosen *Gomorrah* as a challenging dataset to perform character recognition. We rely on the transcripts of the series after a pre-processing stage to standardize the lexicon despite the vagaries of dialect and remove stopwords. A machine learning approach, based on a selection of tools, is then employed to identify characters from the lexicon they employ. The problem is approached as a multi-class classification scheme. We compare several representations of texts, including the simple one-hot encoding and more advanced embedding techniques. The results are presented through a confusion matrix, which can also serve to identify similarities in the linguistic profiles of characters.

KEYWORDS

Character recognition; *Gomorrah*; machine learning; text analysis.

Introduction

As early as 2005, Moretti spoke about the need to apply quantitative methods to literary works to give birth to *distant reading* that would take the place of the close reading method used until then. By *distant reading*, he meant a method involving moving away from the text and observing it from a distance. That distance would not be an obstacle to knowledge but rather a specific form of it: “distance makes one see details less, true: but it makes one understand relationships, patterns, forms better”. Distant reading allows us to move from text to pattern or rather to patterns (Moretti 2005: 3).

The book by Moretti represented a seminal work for introducing quantitative methods in the analysis of literature. But the advent of new media and new narrative forms makes a case for the extension and development of those techniques to them. This is particularly true for TV series, which represent a dominant form in the present landscape of fiction works (Hammond 2005).

As a further step in the introduction of quantitative methods in the analysis of TV series, this paper reports some early results of our work on the linguistic corpus of the *Gomorra*, aiming to offer new models of interpretation of television series, following the path of digital humanities and, more specifically, the methods used by Franco Moretti in the literature.

Our approach follows the reflections made by Moretti himself a few years after his seminal work that we have already mentioned. In fact, Moretti himself observed that while it was true that quantitative methods had established themselves in the study of literature, his idea of “abstract models” had not found wide acceptance or yielded the expected results. It was necessary to restart after that “false movement” and integrate new techniques such as text mining, topic modeling, content analysis, and sentiment analysis with twentieth-century literary theories such as hermeneutics or history (Moretti 2022). However, Moretti recognizes that the digital humanities risk losing

sight of form, essence, and social dimension owing to the overwhelming presence of data. In his own words, “Interpretation moves between form and the world, going in search of the historical meaning of works; quantification moves between form and form, attempting to trace the coordinates of a future atlas of literature” (Moretti 2022: 32). A synthesis between those two contrasting movements has to be found as nothing of the sort has yet been fully achieved, and we must keep trying.

In this paper, we apply text-mining techniques to the corpus made of the dialogues in *Gomorra*, integrating media-studies interpretation and linguistic analysis in the footsteps of what linguists have already been conducting on the linguistic corpus of television series for several years (Bednarek 2018). Our final aim is to uncover the underlying association between a character and his lexical signature.

Gomorra

Gomorra has been a huge national and international success. In America, it has even been compared to major crime series such as *The Sopranos* or *The Wire*, and in 2016 the New York Times ranked the series at No. 3 of best series of the year and in 2019 again, the New York Times ranked it at No. 5 among the best international (non-U.S.) productions of the last ten years. In Italy, it has broken all TV rating records and has become a widespread cultural phenomenon, an object of cult, spurring, however, controversies related to the depiction of the figure of the criminal and its glamorous rendering, as is often the case in series devoted to the crime genre.

Born from the idea of Roberto Saviano, to whom we owe the investigative book of departure, the series arrives after the movie directed by Matteo Garrone (2008) and the theatrical show. It was produced by Cattleya, Fandango and Sky from 2014 to 2021 and is structured into five seasons for a total of 58 episodes. The main directors were Stefano Sollima, Francesca Comencini, and Claudio Cupellini, while the main scriptwriters were Leonardo Fasoli, Stefano Bises, and Maddalena Ravagli. Stefano Sollima, the showrunner of the series, is also the showrunner of *Romanzo criminale*, the first serial production of the Sky model, the pay TV that revolutionized the audiovisual environment for the small screen in the 2000s, opening the Third Golden Age of TV in our country as well (Bernardelli 2016). Sky has invested in producing content with uncomfortable topics and high-quality standards. Following the model of great American series, to which audiences are becom-

ing increasingly familiar, Sky has begun to produce series with well-defined genres, characterized by multi-seasonality, with horizontal and vertical storylines, following the serialized series model defined by Mittel. Such series are also careful to interact with their audiences, giving much space to fandom (Napoli and Tirino 2016). Barra and Scaglioni write that the Sky editorial policy is also to tell about “bad guys” (Barra and Scaglioni 2015) on par with the “difficult man” of American quality television (Martin 2013).

The story is well-known and focuses on a Camorra clan (the Savastanos) fighting for control of the area and various illegal trafficking and on the rivalry of the two former friends, Ciro and Genny, son of the boss. The two of them are the real protagonists, present in all seasons, alongside several other characters who rotate in and out of the extended narrative universe of the series. Its characters are divided into very specific hierarchies: the bosses, the sidekicks, women, law enforcement, and the world of finance and business. The topics narrated, in fact, are not only those related to crime, the Camorra, drugs, and power but also those of family, friendship, and love. Public and private interact in a mixture of realistic drama and melodrama. The style is raw, direct, and violent, and for the performers, little-known actors (who became famous later) capable of using dialect were chosen. Opting for crude spoken language also sought strong adherence to reality. Adherence to reality is also visible in the choice of locations: the district of the sails of Scampia as well as the real places of the criminal Naples that Saviano had already described in his books and Garrone had narrated in his film.

Literature Review

Gomorra has received much attention in the TV series criticism context. Scholars have analyzed it right after its appearance, adopting different points of view and different methodological approaches. It has been a global hit and is considered on a par with the great American quality TV series (McCabe and Akass 2007). It is certainly the best result of the new standards of products made by Italian television (Scaglioni and Barra 2013).

In this section, we review the most relevant essays in Media Studies concerning the series and the most recent results in the quantitative analysis of TV series. The first line of research we analyze concerns the transmediality of *Gomorra* and its huge success both as a narrative ecosystem (De Pascalis 2018) and as a global brand (Benvenuti 2017). The series, in fact, is only

the third remediation, as Grusin would say (Bolter and Grusin 1999), of the original text consisting of Roberto Saviano's novel (already a collation of previous reports). A play was derived from this novel, followed by the film directed by Matteo Garrone and, finally, the series. This movement across media may be dubbed "media colonization" (Guerra et al. 2017). However, it does not result from a reasoned expansion but rather through random and irregular choices, according to a weak transmediality, where projects accumulate only by a logic of simple reference to the source novel.

This research line also includes studies concerning fandom and the public's creativity in the resemantization of the series with the making of memes and videos. *Gomorra* is an expanded universe also because of the audience: as described in (Napoli and Tirino 2016: 108), a symbolic pulp space has been built with the contributions of the public, the platform, and viral videos, where elements borrowed from gangster and spaghetti western movies mix with the cult expressions of *Gomorra*. Videos mainly adopt a parody and quasi-mythological tone with inter-media links and often resort to memetic techniques (Chen 2012, Dawkins 2006) with remix culture elements (Manovich 2007).

Another line of research has paid attention to the narrative structure and characters, e.g., the analysis of female characters in a story where male characters dominate (Milkowska-Samul 2017) or the analysis of main characters as anti-heroes (Martin 2017, Russo 2017, Bernardelli 2016). As to the latter issue, the process of viewers' identification with a morally repugnant protagonist has been described by Russo with the following words:

Ciro Di Marzio, Genny and the other primary characters of *Gomorra* are quintessentially modern inasmuch as they function in the narrative as heroes with whom we align and to whom we temporarily pledge our allegiance, only to eventually be disappointed and repulsed by them. William Brown called on Stanley Cavell's notion of revulsion to suggest an 'ethical mode of engagement' with explicit violence in extreme films (Russo 2017: 17).

Many analyses have focused on the problem of realism in the series, relating it to the original text. The series has been accused of too much realism and excessive violence. Instead, most studies observe that the series' structure is not so realistic but is often closer to other narrative types. For example, Bernardelli highlighted how the series moves between realistic content and melodramatic fictional representation (Bernardelli 2017). Applying the

meaning that Mittel assigns to melodrama, defined as a style of excess and exaggeration of feelings, actions, and words, aimed at narrating feelings and creating pathos, Bernardelli states that contemporary TV series are dominated by anti-heroes who perform deplorable actions and with whom it would be impossible to connect if there were not this very melodramatic mechanism. *Gomorrah*, therefore, has a realistic setting for its raw and direct depiction, bringing to life almost chronicle-like scenes from the lives of the bosses, but, at the same time, it takes on the melodramatic style to make its narrative acceptable to the viewer. In his own words:

The rawness of the events described, the realism, is constantly balanced by strongly theatrical, melodramatic, and often emphatic acting. It is like saying that if you have the viewer bear the weight of reality, you must then emphasize the fictionality of what you represent through the melodrama exhibited by acting: a realistic substance corresponds to a melodramatic form (Bernardelli 2017: 9).

Pescatore and Rocchi (2020) also mentioned a particular “realism formula” for *Gomorrah*, which is closer to the structure of a narrative genre than a documentary one. They analyzed the lexical corpus of Saviano’s book and the lexical corpus of the synopses of the episodes to extract latent topics, that is, lexical/semantic regularities, and find the structure of the series to be related to the gangster genre and its stereotypes. The realism of *Gomorrah* is only conventional because all the stereotypes associated with the gangster genre, such as the mafia crime family, the relationship between generations, and the mechanism of drug dealing, clearly emerge from the study (Pescatore and Rocchi 2020: 105).

The issue of *Gomorrah*’s realism is also investigated by Variano (2019), who defines the series’ language as “furbesco seriale”: a combination of realistic dialectal slang and a reinvention by the scriptwriters, aimed at making the dialogue comprehensible (Variano 2019). The language of *Gomorrah* moves between standard Italian and dialectal Italian.

As with all the ingredients in the series, the language has also conformed to the choice of realism. The Gomorrese is the Neapolitan spoken in Scampia and Secondigliano, softened by some simplifying interventions (made in dubbing in some cases) to make the language as understandable as possible without taking away authenticity to the dialogues (Variano 2019: 287).

Therefore, the series ends up exhibiting a language that is realistic but reinvented by screenwriters at the same time.

The resulting mixture of languages is difficult to translate. Some studies have focused on the difficulties of translating dialogues in English for subtitles (Fruttaldo 2018, Raffi 2017). However, Fruttaldo has pointed out how *Gomorra*'s dialogues can be translated in subtitles and how cross-cultural remediation is possible, stating that

the (re)translation process, when it comes to works that are representative of an extremely local society, may in a way impoverish the source context references in the target language environment; this was the case for *Gomorra*, where a lower culture-specificity can be highlighted in the translation process (Fruttaldo 2018: 156).

If we move to the quantitative analysis of TV series, despite the seminal work by Moretti on fiction works, dating nearly a decade (Moretti 2005), the studies are relatively recent. Most works have concentrated on applying social network analysis techniques (the resulting network is aptly named character network) based on dialogues or other forms of interaction among the characters to describe the dynamics of the character ecosystem and its power structure. Labatut and Bost have surveyed the tools employed for that purpose in (Labatut and Bost 2019). The tools do not differ significantly from what can be employed for novels and other works of literary fiction. Both Edwards et al. and Vala et al. (the latter with specific reference to literary works) have highlighted the difficulties of the task in (Edwards et al. 2018, Vala et al. 2015). Specific examples of the application of social network analysis to TV series are the analysis of the *Game of Thrones* performed by Beveridge and Shan (2016), and the analysis of the narration of *Breaking Bad*, again *Game of Thrones*, and *House of Cards* (Bost et al. 2018). Chao et al. (2018) have employed a two-mode network to analyse the dynamics of character activities and plots in movies. Dialogues have been examined in (Fronzetti Colladon and Naldi 2018) to analyse the dominance of characters in the plot (or, vice versa, the choral features of the narration) regarding the *Big Bang Theory* series. A more extensive analysis, including the concentration analysis of video presence and locations in addition to dialogues has been carried out by Naldi and Dalla Torre (2022) for *Breaking Bad*. Also, Fronzetti Colladon and Naldi (2019) have analysed the relationship between the text and the series' success, again in the *Big Bang Theory*. Social networks have also been employed to validate literary theories, e.g.,

about the relationship between the number of characters and the settings (Jayannavar et al. 2015). Waumans et al. carried out an even more ambitious plan of obtaining the signature of a novel’s story through the topological analysis of its character network (Waumans et al. 2015).

Our analysis of *Gomorra* is lexically based and uses subtitles as its source, observing that dialectal expressions are often not amenable to consistent lemmatization, and there are significant differences between scripts and the actual dialogues. We avoid the language dilemma by excerpting dialogues from subtitles and then performing lemmatization. Since subtitles exploit a softened form of the dialect that mimics the Italian language, we do not have to consider pure dialect expressions at this stage. We use an approach that Franco Moretti has called “Quantitative Formalism”: “take an aesthetic form and disassemble it down to its basic elements: transforming – or rather, let’s say, reducing – a novel to its paragraphs, or a drama to a sequence of linguistic exchanges” (Moretti 2022). In our case, we disassemble by reducing dialogues to their words, adopting the bag-of-words approach. Our final aim is to uncover the lexical signature of each character in *Gomorra*. As long as the lexicon is not distorted (e.g., flattened) in the spoken language-to-subtitles process (i.e., words that are different in the spoken dialogues stay different in their subtitle version), using subtitles does not impact lexical signatures.

Our main research question is then: can we identify a *Gomorra* character from what he/she says? Can we find a lexical signature in *Gomorra*? To do this, we apply a perspective that concerns character identification, an emerging branch of research, to associate dialogues with characters and identify the lexical features of characters.

Methodology and Results

We have applied several quantitative techniques to help us build a profile of each character and identify them based on their lexicon only. In this section, we describe the features we have analysed so far and the methods we have employed. The results of this section are obtained based on the dialogues of the whole series as transcribed in the subtitles.

The features we are interested most in concern both the lexicon employed and its use by the characters. Our underlying research hypothesis is that each character bears a unique lexicon signature, i.e., a personal set of

words that he/she employs most. If this hypothesis holds true, assigning a text extracted from the dialogues to a specific character could be possible.

In order to progress towards that ultimate goal, we have so far analysed the following features:

- Word presence;
- Lexicon similarity and variety;
- Character presence;
- Character identification.

Word Presence

We first examine the lexicon employed in *Gomorra*. A lexicon is made of words. The first thing we wish to observe is the set of words employed in the series, their overall quantity and frequency.

We start with observing the overall number of words uttered during each episode. In Figure 1, we show the time series of the number of words per minute. Using the density of words removes the bias due to the length of episodes. In fact, episodes differ widely in length, ranging from 35 to 55 minutes. The average number of words per minute is 47 over the whole series. If we take a closer look at the season, that average is 46 minutes for Seasons 1 and 2, then lowers to 43 and 41 minutes, respectively, in Seasons 3 and 4, to grow to 45 minutes in the last season. Is that average figure low or high? It should be compared with that of other series in the same genre. Of course, as we can see from the graph, we also have variations from epi-

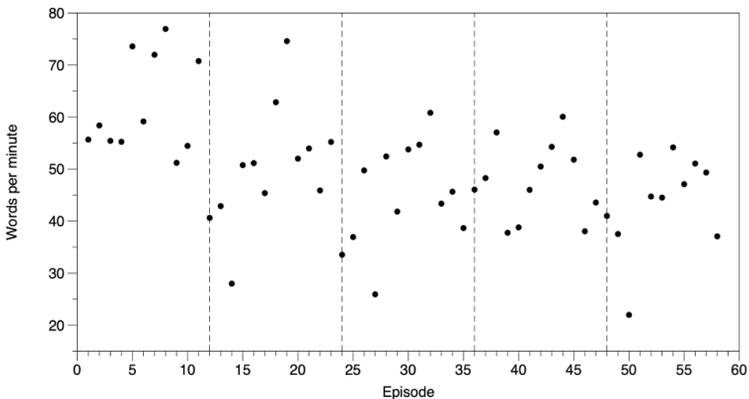
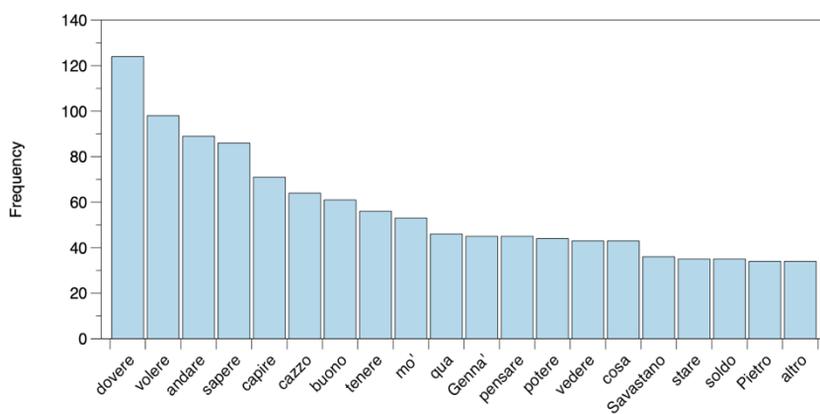
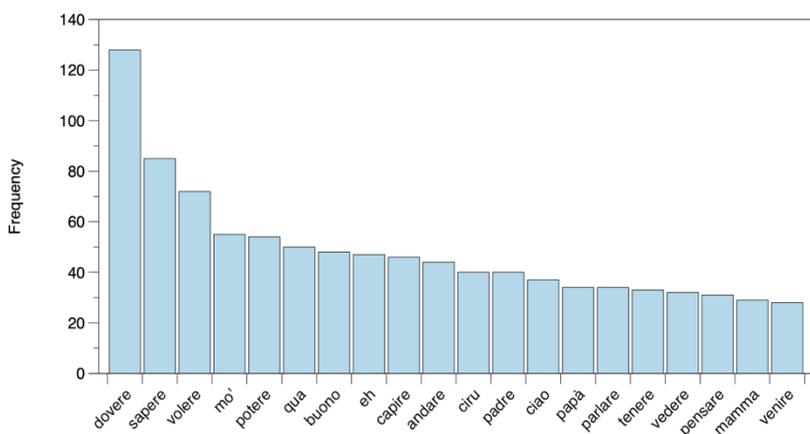


FIGURE 1
Overall density of words throughout the series.

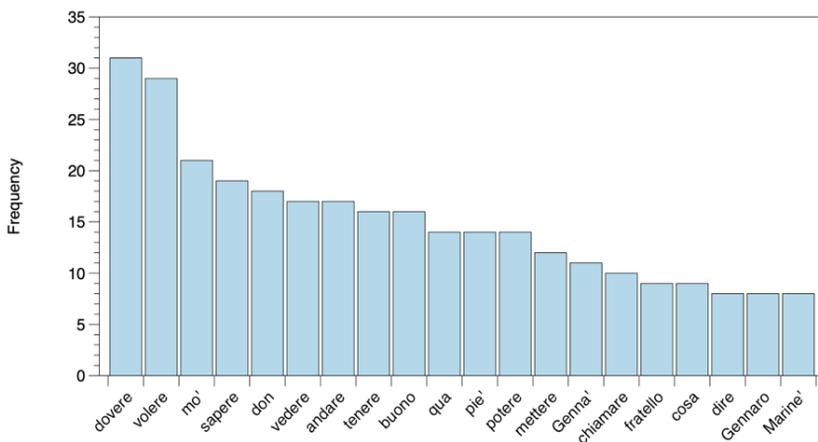
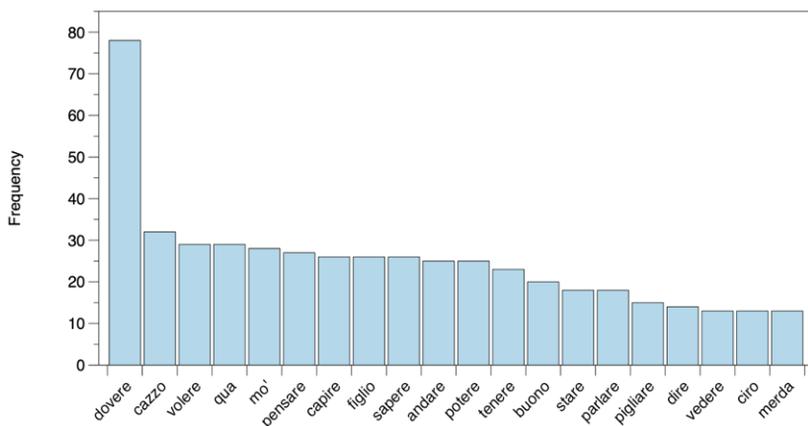
nessuno but also *qualcosa*, or even *interessarsi*, *preoccuparsi di qualcuno* or also *braccare* (Variano 2019: 296).

On closer inspection, two verbs emerge. On the one hand, we see the verb *dovere* (*to must*): as if the whole universe of *Gomorra* were dominated by unwritten rules that everyone must obey. On the other hand, we observe the verb *volere* (*to want*), which corresponds to the protagonists' aspirations, and their quest for power, money, and supremacy. The other strong element that emerges is the topic of family: *padre* (*father*), *mamma* (*mother*), *figlio* (*son*), *famiglia* (*family*) appear as dominant in the cloud.



FIGURES 3, 4
Lexicon employed by Genny (Gennaro Savastano) and Lexicon employed by Ciro.

Gomorra runs on this triad: duty, will, and family. And it does according to a mode that follows genre conventions, especially for gangsters, with its stereotyped elements and character typecasting. So, the first lexical signature we can identify in *Gomorra* is a stereotyped narrative, as other scholars have already said (Pescatore and Rocchi 2020).



FIGURES 5, 6
Lexicon employed by Pietro and Lexicon employed by Patrizia.

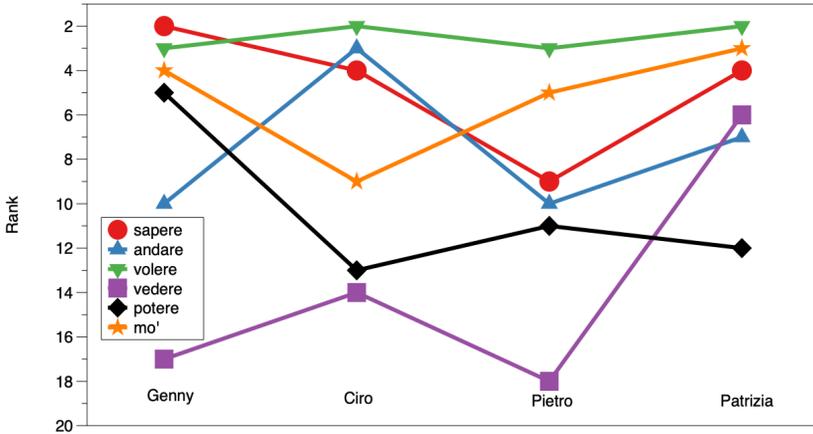


FIGURE 7
Word ranking comparison.

Lexicon Similarity and Variety

We can now dig deeper and examine the specific lexicon employed by the main characters in a comparative fashion. In Figures 3 through 6, we show the words mostly employed by four major characters; the old boss Pietro Savastano, his son Genny (Gennaro Savastano), the once-aid and then rival Ciro, and Patrizia, who starts her criminal career as a liaison officer for Pietro to later become the sole boss of Secondigliano.

As already observed for the whole series, the word *dovere* (to must) ranks first for all those characters. In Figure 7, we have reported the ranks of some words for the four characters. We see that *volere* (to want) ranks second for Ciro and Patrizia, and third for Genny and Pietro. Also, a common interjection (*mo*) ranks quite high for all the characters. If we compare the most used words by Ciro and Genny, the two bosses, we can spot many similarities. The vocabulary is almost the same: they use the same phraseology, as if they share a linguistic code that is that of the bosses, of those in charge. If we compare Figures 3 and 5, we see that Pietro and Genny (father and son) share 14 words in their Top 20.

The similarity between the characters' lexicons can be shown quantitatively for the whole set of characters if we compute the Jaccard Index between any two characters. The Jaccard Index is the ratio of the number of common words between the two characters and the overall number of words

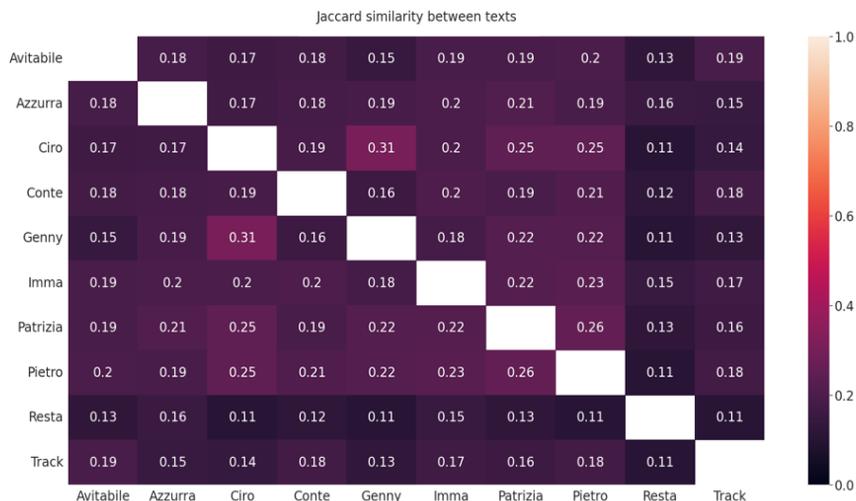


FIGURE 8
Jaccard Index.

they employ and measures the similarity between the lexicons employed by two characters and takes values in the $[0,1]$ range. If that index is close to 0, the two characters use nearly disjoint sets of words. On the opposite end of the range, a Jaccard Index close to 1 shows a nearly identical vocabulary. In Figure 8, we show the Jaccard Index in a matrix form to compare the most important characters. We see that it takes the highest value (0.31) for the Genny-Ciro pair (i.e., they roughly share 30% of their vocabulary), two emerging bosses belonging to the same generation. Slightly lower values show up between Pietro and Ciro (0.25), and Pietro and his son (0.22). On the other end of the spectrum, we find Resta, whose lexicon exhibits very low similarity values with all the other characters (0.11 with Pietro, Ciro, and Genny). Resta is a white collar, who, though joining Genny in a business endeavour, comes from a very different background, being a successful manager. It is to be noted that the Jaccard Index does not account for the frequency of the words, but just for their presence, in a binary fashion. For that reason, we observe a Jaccard Index of 0.22 between Genny and Pietro, while they share 14 out of their Top 20 words (i.e., a commonality ratio of 70%). Probably, they do not share many of their least frequent words.

This linguistic bosses' code, this bosses' signature, frequently uses verbs like must and want and, also, many words concerning the family (father,

son, mother) or proper names of characters linked to the Savastano's family or otherwise to bosses. So, we have exactly the three corners of the triangle delimiting the lexical morphology of *Gomorra*. We can also note another feature: the use of interjections, trivialisms, slang and colloquial phraseology that make up the low and lower-middle register of the characters. In Figure 7, we have noted the widespread presence of the interjection *mo'*. We also observe bad language in the male characters' lexicon. We observe *cazzo* (*dick*) and *merda* (*shit*) in ranks 2 and 20, respectively, for Pietro, with *cazzo* appearing in rank 6 for Ciro. So the stereotyped sign is combined with the cadence of "furbesco seriale", which makes *Gomorra*'s lexicon highly expressive and endowed with great representative force (Variano 2019: 296).

We can conclude that in *Gomorra* we have a bosses' signature: a code of words that all the bosses share and use, a code that marks the general linguistic nature of the series. As noted by Variano (2019: 295), the lingo of *Gomorra* is not the spoken language in Naples or its outskirts but rather a language used for a purpose in certain situations.

After examining the similarity of the lexica employed by different characters, which may serve as a discriminating feature to identify each character by his/her lexicon, we now consider lexicon variety. We are back to a characteristic of each character, though we can, of course, compare how the different characters perform. By variety we mean the richness of their vocabulary, how many words they employ and how often they keep em-

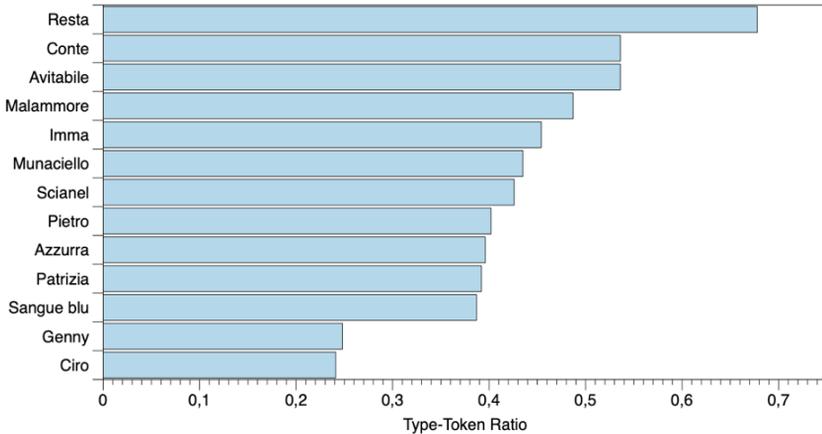


FIGURE 9
Type-Token Ratio.

ploying the same words repeatedly. We measure variety by a single indicator: the Type-Token Ratio (TTR), i.e., the ratio between the number of unique words and the overall number of tokens. Values closer to 1 imply a rich language, while values closer to 0 show a poor vocabulary. In Figure 9, we report the TTR for the major characters. We observe the poor language used by *Ciro* and *Genny*, who exhibit a TTR of around 0.25, while *Resta*, a white-collar character, reaches the highest value here. *Conte* and *Avitabile* are the two runner-ups, respectively, who play the role of a somewhat mannered boss and a borderline boss who runs a hotel aside from his criminal activities. Additionally, even other bosses rank better than *Ciro* and *Genny*, whose common feature is to be two young bosses.

Character Presence

We now examine the presence of characters. Since our analysis is based on the lexicon, we assess their presence by dialogues. In particular, we consider both the number of lines and the number of tokens. In Figure 10, *Genny* and *Ciro* dominate the scene, exhibiting figures roughly three times as large as their runner-ups under both indicators. Their heavy presence seems to

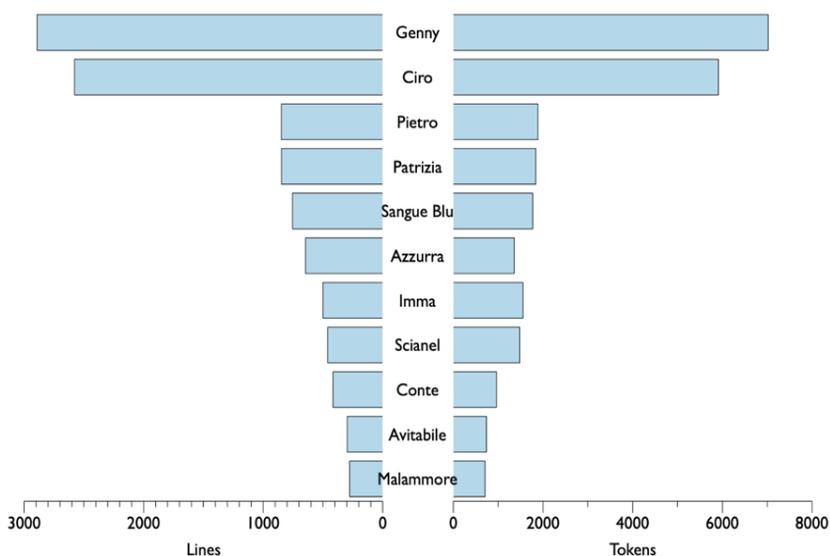


FIGURE 10
Characters' presence.

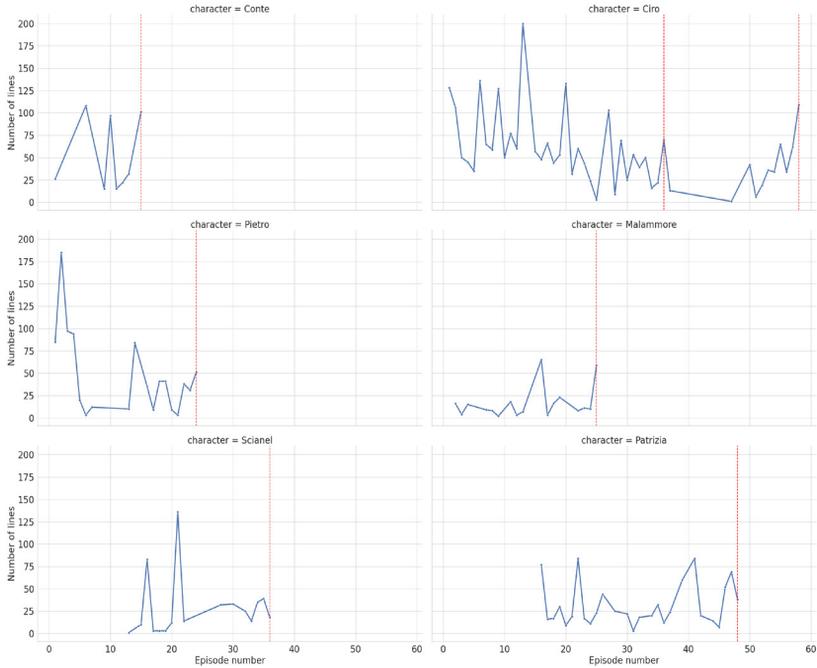


FIGURE 11
 Characters' presence across the series and the death factor.

show that though Don Pietro ranks higher in the criminal hierarchy for most of the series (being respectively Genny's father and Ciro's boss), the series is mainly centred on the two younger bosses.

We can also track the characters' presence throughout the series. In Figure 11, we see that some characters have an intermittent presence (see, e.g., both Conte and Scianel), and even the most prominent characters have a presence that is far from uniform (see the ups and down in Ciro's and Pietro's presence). On the same graph, we also notice what we might call "the death factor". For characters about to die, we observe a growing presence in the dialogues right before their death. In Figure 11, we have marked the episode where the character dies with a red label. We observe that death happens right at a local or global peak of presence. We observe a local peak for Pietro, Ciro (his first death before his apparent resurrection), Scianel, and Patrizia. Instead, we observe a global peak for Conte, Malammo'. It is

a narrative strategy that naturally creates a rising tension designed to capture the viewer's attention, a lexical hook, reminding us of the discursive hooks of narrative strategies in contemporary TV series scripts mentioned by Kristin Thompson (Thompson 2012).

Character Identification

Previously, we have seen that the lexica of some characters exhibit a degree of similarity. This observation significantly reduces the hope of identifying characters based on their lexicon. Here we investigate a character identification approach based on cosine similarity.

Cosine similarity is a measure of similarity between texts. We consider a space whose dimension is the overall number of unique words obtained after merging the two texts. Each text is treated as a term-frequency vector. The cosine similarity of the two texts is the cosine of the angle between the two vectors.

For character identification, we subdivide the overall text spoken by each character into two parts, which we call the train and test portions, with 80% and 20% proportions, respectively. We then examine the test portion of the vectors of character X, compute its cosine similarity with the training vector of all the characters, and attribute the test vector to the character with the highest cosine similarity. Character identification is correct if that test vector

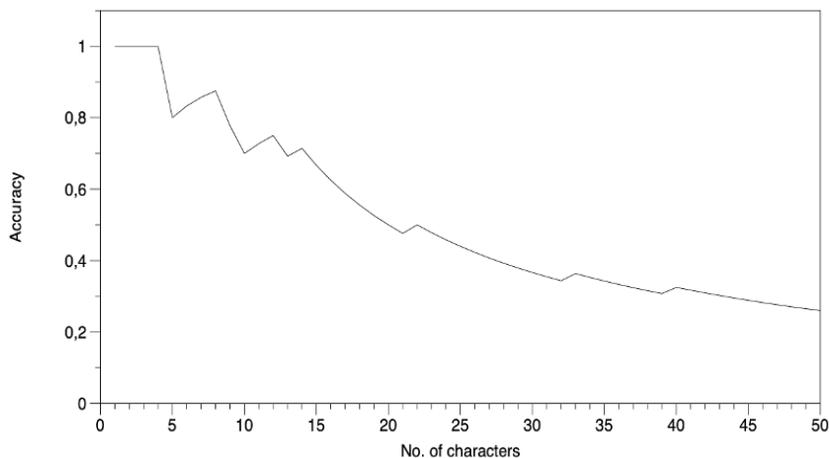


FIGURE 12
Characters' identification accuracy.

happens to be the test vector of character X. We evaluate the performance of this identification criterion by its accuracy, computed as the fraction of characters correctly identified. We expect the identification task to become more difficult as the number of potential characters to be identified grows. For that reason, we evaluate the algorithm's accuracy as we progressively increase the number of characters. We start with the most present characters (Genny) and then add one character at a time in order of decreasing presence (see Figure 10). When we have just two characters to identify, we may identify none, one, or two (i.e., with an accuracy of 0, 50%, or 100%).

The performance we get is shown in Figure 12. The accuracy is still larger than 80% when we have no more than eight characters and goes below 50% when we have more than 20 characters.

Conclusions

We started with a research question asking if we could identify a lexical signature of characters in *Gomorra*. A positive response would allow us to identify speakers based on their dialogues. We have shown early results in a research effort that is still in progress.

Our analysis of dialogues through simple similarity indices has shown that characters show both features common to the role of characters (e.g., the frequent use of certain words in bosses' dialogues) and features that can be exploited to identify individual characters (e.g., the type-to-token ratio to introduce the lexical variety of characters). Our first attempt shows that character identification is possible with a limited number of characters, and identification performances degrade as we increase the range of characters.

Our distant reading has, however, highlighted a lexical pattern that may be named bosses' code. Bosses appear to share a standard set of words that they frequently employ. That lexical corpus is typical of the gangster genre followed by the *Gomorra* series but also mimics the criminal reality that inspired *Gomorra* scriptwriters. The lexicon of Genny and Ciro is then both a real-life and a genre lingo, which we may dub Gomorrese.

Also, the distinguishing feature of complex tv series, which has been adopted as a reference to compare *Gomorra* with, is to narrate stories framed within a genre but set within a powerfully realistic context. The first series employing this mixture, at least within the gangster genre, was *The Sopranos*, created by David Chase and to which *Gomorra* has been

compared. *The Sopranos* ushered in the third golden age of American television and started the HBO model, followed by many producers-broadcasters (e.g., Sky): the story of the depressed gangster Tony Soprano moves within classic themes and stylistic elements of the American mafia film, but at the same time it also refers to the great American realist tradition.

Gomorra moves along the same lines. It is based on an investigative journalism book that is considered to belong to the new Italian epic. The term has been coined by the group of authors acting under the collective name of Wu Ming and refers to a trend in contemporary Italian fiction to deal with current Italian history by adopting the narrative conventions of a pop genre (Wu Ming 2009: 5-61).

Another distinguishing feature emerging from the analysis of dialogues is what we could call the death factor, i.e., the sudden increase of a character's relevance right before his/her death. *Gomorra*'s authors give dying characters the dead man walking stylistic feature: characters talk more when they are about to be silent forever.

On the other hand, the linguistic analysis confirms what the general audience may recognise: the Genny-Ciro pair's centrality. Though the story is rather coral, those characters represent two poles grabbing most of the video and dialogue presence. They also share the lexicon of older bosses. A distinguishing feature of that lexicon is that of being somewhat limited, using few words that often include the verbs *dovere* and *volere* (to want and to must). Those verbs embody the paradigm of the whole series: a perennial fight between obligations and wishes, which will remain unsolved for the two main characters.

This early quantitative analysis of dialogues in *Gomorra* has shown that we can uncover a set of patterns and models to build a layered interpretation of the series, following Moretti's suggestion to keep together the analysis of details and the drawing of a wider morphology.

BIBLIOGRAPHY

- Barra, Luca and Massimo Scaglioni (2015). "Saints, Cops and Camorristi. Editorial Policies and Production Models of Italian Tv Fiction." *SERIES – International Journal of TV Serial Narratives* 1(1): 65-76. <https://doi.org/10.6092/issn.2421-454X/5115>.
- Bednarek, Monika (2017). "The Role of Dialogue in Fiction." In *Pragmatics of Fiction* edited by Miriam H. Locher and Andreas Jucker, 129-158. Berlin-Boston: De Gruyter Mouton.
- Bednarek, Monika (2018). *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge: Cambridge University Press.
- Benvenuti, Giuliana (2017). *Il brand Gomorra. Dal romanzo alla serie TV*. Bologna: Il Mulino.
- Bernardelli, Andrea (2017). "Melodramma e realismo in Gomorra. La serie." *Ocula* 18: 1-14. <https://doi.org/10.12977/ocula85>.
- Beveridge, Andrew and Jie Shan (2016). "Network of Thrones." *Math Horizons* 23(4): 18-22. <https://doi.org/10.4169/mathhorizons.23.4.18>.
- Bolter, Jay David and Richard Grusin (1999). *Remediation: Understanding New Media*. Cambridge: MIT Press.
- Bost, Xavier, Vincent Labatut, Serigne Gueye and Georges Linarès (2016). "Narrative Smoothing: Dynamic Conversational Network for the Analysis of TV Series Plots." In *DyNo: 2nd International Workshop on Dynamics in Networks, in conjunction with the 2016 IEEE/ACM International Conference ASONAM*. <https://doi.org/10.1109/ASONAM.2016.7752379>.
- Chao, Dingding, Taro Kanno, Kazuo Furuta and Chen Lin (2018). "Representing Stories as Interdependent Dynamics of Character Activities and Plots: A Two-Mode Network Relational Event Model." *Digital Scholarship in the Humanities* 34(3): 471-81. <https://doi.org/10.1093/lcf/fqy062>.
- De Pascalis, Ilaria (2018). "An Italian Ecosystem: *Gomorra*." In *Reading Contemporary Serial Television Universes*, edited by Paola Brembilla and Ilaria De Pascalis, 113-27. New York-London: Routledge.
- Dusi, Nicola (2019). "Adapting, Translating and Reworking *Gomorra*." *Adaptation* 12(3): 222-39. <https://doi.org/10.1093/adaptation/apz007>.
- Edwards, Michelle, Lewis Mitchell, Jonathan Tuke and Matthew Roughan (2018). "The One Comparing Narrative Social Network Extraction Techniques." *arXiv*:1811.01467. <https://doi.org/10.1109/ASONAM49781.2020.9381346>.
- Erbaggio, Pierluigi (2015). "#*GomorraLaSerie*: Converging Audience and Enhanced Authors on Twenty-First-century Italian Screens." *Modern Italy* 20(4): 335-49. <https://doi.org/10.1080/13532944.2015.1096241>.

- Fronzetti Colladon, Andrea and Maurizio Naldi (2018). "Concentration Indices for Dialogue Dominance Phenomena in TV Series: The Case of *The Big Bang Theory*." In *International Conference on the Statistical Analysis of Textual Data*, 55-64. Cham: Springer.
- Fronzetti Colladon, Andrea and Maurizio Naldi (2019). "Predicting the Performance of TV Series through Textual and Network analysis: The Case of *Big Bang Theory*." *PloS ONE* 14(11), e0225306. <https://doi.org/10.1371/journal.pone.0225306>.
- Fruzzardo, Antonio (2018). "(Re) Translating Culture-Bound Elements in *Gomorra-The Series*: A Corpus-Based Investigation into Relocated Identities." *Status Quaestionis* 15: 135-60. <https://doi.org/10.13133/2239-1983/14574>.
- Guerra, Michele, Sara Martin and Stefania Rimini (eds) (2017). "Un altro mondo in cambio. *Gomorra* fra teatro, cinema e televisione." *Arabeschi*, 9.
- Guerra, Michele, Sara Martin and Stefania Rimini (eds) (2018). *Universo Gomorra. Da libro a film, da film a serie*. Milano-Udine: Mimesis.
- Hammond, Michael and Lucy Mazdon (eds) (2005). *The Contemporary Television Series*. Edinburgh: Edinburgh University Press.
- Heimerl, Florian, Steffen Lohmann, Simon Lange and Thomas Ertl (2014). "Word Cloud Explorer: Text Analytics Based on Word Clouds." In *2014 47th Hawaii International Conference on System Sciences*, 1833-42. <https://doi.org/10.1109/HICSS.2014.231>.
- Holdaway, Dom (2014). "Osservazioni sulla retorica di *Gomorra*." In *Un nuovo cinema politico*, edited by Luciana Arcangeli, William Hope and Silvana Serra, 198-210. Leicester: Troubadour.
- Jayannavar, Prashant Arun, Apoorv Agarwal, Melody Ju and Owen Rambow, (2015). "Validating Literary Theories Using Automatic Social Network Extraction." In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 32-41.
- Jenkins, Harry (2006). *Convergence Culture: Where Old and New Media Collide*. New York-London: New York University Press.
- Labatut, Vincent and Xavier Bost (2019). "Extraction and Analysis of Fictional Character Networks: A Survey." *ACM Computing Surveys (CSUR)* 52(5): 1-40.
- Martin, Brett (2013) *Difficult men: Behind the scenes of a creative revolution: From The Sopranos and The Wire to Mad Men and Breaking Bad*. New York: Penguin.
- Martin, Sara (2017). "Il fascino sfacciato dei boss." In *Arabeschi* 9. *Un altro mondo in cambio. Gomorra fra teatro, cinema e televisione*, edited by Michele Guerra, Sara Martin and Stefania Rimini, 9: 238-41.
- McCabe, Janet and Kim Akass (eds) (2007). *Quality TV: Contemporary American Television and Beyond*. London: IB Tauris.
- Milkowska-Samul, Kamila (2017). "Per una nuova immagine delle donne in TV italiana. Alcune riflessioni in base alla fiction *Gomorra*." *International Journal for Linguistic, Literary and Cultural Research* 2(3): 193-203.
- Mittel, Jason (2015). *Complex tv: The Poetics of Contemporary Television Storytelling*. New York: New York University Press.
- Moretti, Franco (2005). *La letteratura vista da lontano*. Torino: Einaudi.
- Moretti, Franco (2022). *Falso movimento. La svolta quantitativa nello studio della letteratura*. Milano: Nottetempo.

- Naldi Maurizio and Paola Dalla Torre (2022). “Uncovering the Narrative Structure of *Breaking Bad* through a Multi-Dimensional Quantitative Analysis.” *SERIES – International Journal of TV Serial Narratives* 8(1): 17-32. <https://doi.org/10.6092/issn.2421-454X/14503>.
- Napoli, Antonella and Mario Tirino (2015). “*Gomorra* Remixed: Transmedia storytelling tra politiche di engagement, mainstream e produttività del fandom.” *SERIES – International Journal of TV Serial Narratives* 1(2): 191-204. <https://doi.org/10.6092/issn.2421-454X/5904>.
- Napoli, Antonella and Mario Tirino (2016). “Senza pensieri. *Gomorra – La serie*: dal contesto produttivo alle audience della Rete, fenomenologia di un processo culturale, transmediale.” *Mediascapes Journal* 7: 102-14.
- Pescatore, Guglielmo and Marta Rocchi (2020). “*Gomorra*: La serie, la formula del realismo.” *L'avventura. International Journal of Italian Film and Media Landscape* 6(Fascicolo Speciale): 87-108.
- Raffi, Francesca (2017). “*Gomorra – The Series* Flies to the UK: How Is *Gomorra*'s World Rendered in English Subtitles?” In *Transnational Subject. Linguistic Encounters. Vol. 2. Selected Papers from XXVII AIA Conference*, edited by Liliana Landolfi, Eleonora Federici and Flavia Cavaliere, 179-93. Napoli: Liguori.
- Renga, Dana (2015). “Making Men in *Gomorra: La serie*.” *L'avventura. International Journal of Italian Film and Media Landscape* 1: 105-20.
- Renga, Dana (ed.) (2016). “*Gomorra: La Serie*: Beyond Realism.” *The Italianist* 36(2): 287-354.
- Russo, Paolo (2017). “Storylining Engagement with Repulsive Antiheroes. Towards a Cognitive Poetics of TV Serial Drama Narrative: The Case of *Gomorra – The Series*.” *Journal of Screenwriting* 8 (1): 5-21.
- Scaglioni, Massimo and Luca Barra (eds) (2013). *Tutta un'altra fiction. La serialità pay in Italia e nel mondo*. Roma: Carocci.
- Thompson, Kristin (2012). *Storytelling. Forme del racconto tra cinema e televisione*. Soveria Manelli: Rubbettino.
- Vala, Hardik, Davide Jurgens, Andrew Piper and Derek Andrew (2015). “Mr. Bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the Difficulty of Detecting Characters in Literary Texts.” In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 769-74.
- Variano, Angelo (2019). “Il furbesco seriale della fiction: La lingua di *Gomorra: La Serie*.” *Lingue e Linguaggi* 30: 285-305. <https://doi.org/10.1285/122390359V30P285>.
- Waumans, Michael C., Thiabaut Nicodème and Hugues Bersini (2015). “Topology Analysis of Social Networks Extracted from Literature.” *PLoS ONE* 10(6), e0126470. <https://doi.org/10.1371/journal.pone.0126470>.
- Wu Ming (2009). *New Italian Epic. Letteratura, sguardo obliquo, ritorno al futuro*. Torino: Einaudi.

LOOKING FOR LEXICAL SIGNATURES IN GOMORRAH



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.a8ccd1c0>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.