

**Content Analysis and Digital Data:  
Methodological Solutions between Tradition and  
Innovation. The Case of Digital Campaigns as  
Meeting Ground of Expertises and Research  
Perspectives**

*Maria Paola Faggiano, Raffaella Gallo*

**How to cite**

Faggiano, M. P., Gallo, R. (2021). Content Analysis and Digital Data: Methodological Solutions between Tradition and Innovation. The Case of Digital Campaigns as Meeting Ground of Expertises and Research Perspectives. [Italian Sociological Review, 11 (4S), 293-318]  
Retrieved from <http://dx.doi.org/10.13136/isr.v11i4S.436>

[DOI: 10.13136/isr.v11i4S.436]

**1. Author information**

*Maria Paola Faggiano*

Communication and Social Research Department (CoRis), Sapienza  
University of Rome, Rome, Italy

*Raffaella Gallo*

Statistical Sciences Department (DSS), Sapienza University of Rome,  
Italy

**2. Author e-mail address**

*Maria Paola Faggiano*

E-mail: mariapaola.faggiano@uniroma1.it

*Raffaella Gallo*

E-mail: raffaella.gallo@uniroma1.it

**3. Article accepted for publication**

Date: January 2021

**Additional information about**

**Italian Sociological Review**

**can be found at:**

**About ISR-Editorial Board-Manuscript submission**



*Content Analysis and Digital Data: Methodological Solutions between Tradition and Innovation. The Case of Digital Campaigns as Meeting Ground of Expertises and Research Perspectives<sup>1</sup>*

Maria Paola Faggiano\*, Raffaella Gallo\*\*

Corresponding author:  
Maria Paola Faggiano  
E-mail: mariapaola.faggiano@uniroma1.it

Corresponding author:  
Raffaella Gallo  
E-mail: raffaella.gallo@uniroma1.it

**Abstract**

In our Platform Society, Content Analysis represents a bridge between the past and the future: thanks to technological advancements, this technique can perform its typical functions more efficiently, but it can also penetrate new fields.

The authors present the case of digital campaigns, by way of a strategy that enhances both the text-based material in a strict sense, and the entire multimedia system interrelated with the units under analysis.

The research examples – selected from the analysis of Facebook posts published by the main Italian political forces during the digital campaign for the 2018 General Elections – have the purpose of explicating how specific methodological choices and investigation styles can be “reused” in other study contexts.

---

<sup>1</sup> Maria Paola Faggiano: responsible for paragraphs 1, 2, 3.2. and 4. Raffaella Gallo: Responsible for paragraphs 3 and 3.1.

\* Communication and Social Research Department (CoRis), Sapienza University of Rome, Rome, Italy.

\*\* Statistical Sciences Department (DSS), Sapienza University of Rome, Italy.

This paper proposes the integrated use of *CA* as a survey, statistical textual analysis and qualitative *CA*, further to the idea that the combined application of these can limit the disadvantages connected to the exclusive use of specific instruments, increasing overall advantages.

Finally, *CA* is presented as the springboard for other research initiatives: from the investigation of the points of view/the social profiles of the political actors associated with the posting activity, to the political message analysis in the other (on/offline) contexts in which it appears; from the analysis of communication strategies (interpreted on the basis of the media channels), to the study of social reactions-interactions the message generates.

Keywords: content analysis, digital campaigns, integrated analysis approach.

## **1. Digital Society and Digital Data: introductory notes on the environment**

As inhabitants of the Digital Society we are aware, in varying degrees of consciousness and breadth, of the profound influence of new digital technologies on our everyday lives (van Dijck et al. 2019[2018]). Most of the activities we carry out on a daily basis involve the use of electronic devices connected to the Internet; the use of Social Network Sites, blogs and online newspapers, the use of e-commerce and of a wide range of services on the Net represent some of the many habits rooted in broad and transversal segments of the population, as well as in specific and interesting targets. The Net has become the largest vehicle for information on how social actors make their choices and consumption; they express opinions, preferences and orientations; they communicate and interact with others. Such collective actions and social exchanges, systematically transformed into digital traces, often take the form of Big Data, by virtue of their vast proportions and complex forms of interconnection. Beyond the breadth and structural characteristics of the data that settle on the Net – which, moreover, often have a “textual” nature – they constitute a veritable mine of knowledge for Social Sciences, which are ready to seize the opportunities and the challenges of the digital revolution to understand the social transformations taking place, intercept collective needs and forms of inequality, and identify solutions to urgent social problems.

The vast range of data that settles on the Net, of great interest for Social Sciences, includes the extensive administrative databases available online in relation to important sectors such as health, work, education and safety; the multiple forms of online communication (one need only think of the variegated world of Social Network Sites); the digital traces connected to the uses of digital devices, which are very often geo-referenced (Martire, Pitrone, 2016; Conte,

2016). This scenario largely constitutes a joint work opportunity for scholars from different disciplinary sectors (from sociology to psychology, from marketing to information technology, from statistics to jurisprudence) and with diversified, while simultaneously complementary, skills; it represents the framework within which scientific research can take further steps forward in terms of breadth, relevance and scope of undertaken studies, complexity of the analyses carried out, and robustness of the observation models. The conduct of investigations characterized by a wider space-time perspective than in the past – with the consequent attainment of new spaces of knowledge, the acquisition of new frontiers of analysis, the probing of new themes, as well as, contemporaneously, the need to deal with new challenges, limits and difficulties on the theoretical and methodological front – constitute a valuable innovation compared to the traditional way of conducting research. For these reasons, as mentioned, the opportunity to develop a “more open attitude” in the scientific field transversally concerns all disciplinary sectors, called to converse, merge skills and approaches, and refer to transversal methodological frameworks.

Today more than ever – looking both at the ever wider boundaries of analyzable empirical material, and at the means available to carry out elaborations and reach results – social scientists have the possibility of carrying out sophisticated statistical and statistical-textual analyses of the automatic and semi-automatic kind, of studying the attitudes of vast universes, extensively reconstructing a climate of opinion, systematically assuming a longitudinal research perspective (one may think of phenomena such as the productivity of a scientific community, youth subcultures, e-learning, electoral behaviour, the values and lifestyles of heterogeneous social groups, leadership, the political agenda, the representation of ethnic and cultural minorities). Furthermore, they can do all this with advanced and innovative equipment that, in a certain sense, breaks with the past. On the other hand, while positively acknowledging the innovative lunges of digital research, the comparison between results obtained in an automated way (Sentiment Analysis, Network Analysis applied to Social Networks, etc.) and qualitative analyses appears particularly relevant, aimed at investigating the empirical material downloadable from the Net in an interpretative sense (Di Giammaria, 2016). In other words, the detection and processing of large masses of data sometimes implies the adoption of research strategies based on highly automated algorithmic options that disciplines which are traditionally distant from these approaches, such as sociology or political science, have recently received and made their own. On the other hand, the possible perspectives of integrating Big Data and Small Data, as that of Quality and Quantity, translate into the maximum exploitation of the plurality of available research and analysis techniques, in the hypothesis that it is difficult to describe or even explain a phenomenon exclusively through Big Data, or

through mere quantifications obtained automatically (Bail, 2014). Pursuing a scientific perspective of maximum integration (in relation to the phases of data collection, archiving and processing) it is desirable, on the one hand, to gradually adapt the scope and complexity of scientific research to social complexity, through the constant acquisition of skills and the willingness to change on the part of its stakeholders. On the other hand, it is advisable for new technologies and current frontiers of data analysis to continue to interact with traditional research tools and strategies (Di Giammaria, Faggiano, 2016; Faggiano, 2016), a precious basis on which innovative research practices can be successfully grafted.

## **2. Digital Society and Content Analysis: aims and perspectives**

The acquisition and analysis of the data offered by the Net are not exempt from numerous problems on the methodological front, which warrant inquiry: intrusiveness, social desirability, ethics of analysis, validity, fidelity, and statistical representativeness are just a few, particularly relevant, aspects (Stefanizzi, 2016; Martire, Pitrone, 2016; Richterich, 2018). Herein, where the focus is represented by Content Analysis, it will not be possible to question the entire pool of issues. The key word in this work is “integration”, in the use of techniques, in the adoption of approaches and perspectives and in the design of research procedures. The main idea, the common thread of this entire paper, is maximizing benefits and results during investigations centered around CA.

As mentioned, the digital turnaround and continuous technological innovation have generated considerable effects (in terms of data collection, data analysis and reporting techniques) on social research in general, which negatively affect and, in essence, benefit CA. The Net, reflecting specifically on CA as well, offers new opportunities and launches just as many challenges, making the (new) data flows particularly fast, wide, substantial and interconnected.

In this Platform Society we are witnessing a proliferation of content, a transformation and expansion of the space/platform for observation of the social world (one need only think of the new relational and communicative dynamics and the “fusion” between the online and offline spheres). In this sense, the objects we can look at, the spaces in which we can insert ourselves as observers of the social world, the methodological styles that we can adopt have increased (one may think of the development of dredging systems and mathematical models/algorithms for analysis, innovative archiving and sharing techniques, the application of sophisticated analysis models and techniques, the

push towards automation, the possibilities open to the longitudinal and transnational analysis of large data archives).

We are faced with a real and downright expansion of the meaning of the term “text”, which determines an exponential growth in the potential of CA: the range of textual corpora to which we can refer is increasingly vast; the breadth of the empirical bases that can be analyzed for scientific purposes can also reach enormous proportions; the multimedia kit in which the elementary semantic units are structured is complex and articulated. The “text” to which we can aspire, as researchers of the digital age, is a complex plot that not only includes words in the strict sense, but also, for example, images and videos, and represents an interconnected and multi-platform “system of meaning”.

CA represents a real bridge between the past and the future (Krippendorff, 1980; Tiplado, 2014), suggesting a creative and profitable combination between tradition and innovation: technological advancement allows technique to perform its typical functions more efficiently (one may think of the “reinvigoration” of techniques such as in-depth interviews and focus groups as a result of digitization, which breaks down space-time boundaries and speeds up many steps, from transcription to analysis), while also allowing it to enter new fields. Digital Data thus reveal, in terms of CA as well, new methodological horizons and suggest how the future lies in the integration and intersection of sensibilities on the technical-practical and substantive levels, in methodological pluralism and interdisciplinarity. Old and new merge fruitfully, by virtue of technological advancement, affecting both the objects and contexts of study, as well as the techniques of detection and analysis. What we can talk about is a real revolution for CA, which affects, in our opinion in a decidedly positive sense, both the form and the “substance” of the data, and expands the context of use and the spread of such technique.

Faced with the increased potential of CA in Digital Society in a general sense, in this paper our attention is focused on data coming from Social Media, in particular from Facebook, considered particularly interesting for their structure and complexity, because of the “social worlds” they interconnect. With the aim of showing how a collection of traditional analysis techniques, possibly used in combination, can adapt to the new “digitized” research spaces, even in a direction that is not merely exploratory, a case study on Italian digital campaigns is presented. Consistently with the above, for the empirical part, the paper focuses on the analysis of Facebook posts published by the main political forces in Italy during the digital campaign for the 2018 General Elections, adopting a strategy that summarizes different types of Content Analysis, capable of enhancing both the textual material in the strictest sense, as well as the entire multimedia kit interconnected with the posts under analysis (photos, videos, links, etc.). The examples offered, far from representing a research report, have

the sole purpose of explaining the methodological choices made and the style of research – “exportable” elsewhere – adopted. Constantly underlining how the prospective investigation practices can be applied to other fields of study (the empirical case presented has the function of exemplifying research horizons and “transferable” practices in other – and numerous – contexts of empirical research), the analysis of Italian digital campaigns is therefore intended as a meeting ground between different knowledge and research perspectives.

Trying to make an evaluation – efficiently combining quantitative analysis and qualitative interpretation of the text, as well as responding to needs of empirical control, robustness and deepening/exploration of data – the use of Content Analysis as a survey (managed via a semi-standardized form, similar to a questionnaire – Losito, 1993, 2009; Tipaldo, 2013; Faggiano, 2016), statistical textual analysis and qualitative Content Analysis are critically discussed. The proposed perspective calls for integrated use of different Content Analysis techniques, oscillating between quality and quantity, the use of automatisms such as ‘paper and pencil’ operations, the simultaneous adoption of an exploratory and interpretative approach. The paper proposes, compatibly with available resources, the integrated use of Content Analysis as a survey, statistical textual analysis and qualitative Content Analysis, according to the idea that the combined application of them can limit the disadvantages resulting from exclusive use of specific instruments, and increase the overall advantages from a theoretical point of view, and in terms of robustness and data control (Hashemi, Babaii, 2013; Pashakhanlou, 2017).

Finally, the analysis of the content is presented as the starting point for other research initiatives: from the investigation of perspectives and social profiles of the political actors with whom the posting activity is associated (using more or less structured interviews, individual or group, as a tool) to the analysis of the political message in other contexts – on/offline – in which it appears; from the analysis of communication strategies based on the media channel used, to the study of social reactions and interactions the message generates.

Essentially, and without further revelation, looking at the potential of CA applied to empirical material from Social Media, the axes of discourse on which, through the presented case study and the conclusions drawn, the areas of focus in this paper are: adoption of a mixed methods approach, the propensity towards the articulation of intricate and complex research designs (“coherent” in terms of the observed social complexity), the preference for research procedures in which methodological pluralism, interdisciplinarity and pooling of skills are valorized.



### 3. Content Analysis and Political Communication: the empirical study of Digital Campaign for the 2018 Italian General Election

Since 2013, the Internet has played a progressively central role in the political communication strategy of the Italian parties (Morcellini et al., 2013). As underlined by Cepernich (2015), the current relationship between politics and the web could be defined as “networked politics”. The contemporary scenario, which moves within the so-called hybrid media system (Chadwick, 2013), characterized by the integration between traditional media logic and new media logic, defines social networking sites as the main hubs of every hybrid campaign (Wells et al., 2016). In fact, Social Networks Sites have become the fundamental tools to publish political messages aimed at reaching specific target of users, encouraging a microtargeting strategy and spreading personalized content. In the apical phase of hybrid digital campaigns that we are living, constant communication/information/interaction between political forces/leaders and audiences is combined (Tipaldo, 2014).

In view of these changes, that have affected not only the field of political communication but many other areas as well, and in the presence of a heterogeneous empirical material of enormous proportions and in continuous production, social researchers have new methodological goals and opportunities. Scholars need to adopt and refine research tools through which, on the one hand, this new political communication speed and complexity can be grasped; on the other hand, the observed phenomena evolutions can be analyzed from a longitudinal study point of view.

In this regard, we believe that Content Analysis is a particularly suitable methodology for this purpose, with respect to both Small Data and Big Data analysis. In our opinion, this is even more true if it is applied in a *mixed methods approach* that combines its different methodological variations by integrating the different techniques in which it is articulated.

The empirical studies characterized by the application of Content Analysis show the tendency to use only one technique at a time. In particular, if we consider the field of political communication, especially if this is analyzed on Social Network Sites (known for the overabundance of data and the availability of big text corpora), we note the systematic use of automated and semi-automated text analysis techniques (only by way of example, see: West et al., 2006; Klüver, 2009; Antenore, Faggiano, 2013; Bimber, 2014; Medzihorsky et al., 2014; Kropf, Grasset, 2017; Nobile, 2018a, b).

Concerning the *exclusive or concomitant use* of data collection and analysis techniques and tools, the *mixed methods approach* has not gained much attention

with reference to the Content Analysis sphere<sup>1</sup>, while it registers a strong diffusion spread compared to other types of data collection and data analysis (for example, within research projects that mix questionnaires, focus groups and/or in-depth interviews and allow researchers to reach qualitative and quantitative results).

The data that we will show in the following pages are taken from a wider study on the electoral campaign for the 2018 Italian General Elections (Cavallaro et al., 2018; Valbruzzi, 2018; Lombardo et al., 2019a). The empirical basis refers to the political message that leaders and parties convey to their electorate and consists, more operatively, of the posting activity on Facebook of the main political forces (the League, the Democratic Party and the Five Star MoVement) during the four weeks preceding the electoral appointment. In particular, the observation activity concerns the weeks of opening (5-11 February 2018) and closing (26 February-3 March 2018) of the electoral campaign, given their strategic value on the decision-making process in view of the vote. The data collection was carried out on the public Facebook page of a targeted selection of political parties-movements (the ‘pivots’ of Italian politics): for the center-right wing the League was chosen, for the center-left wing the Democratic Party and for the ‘Third political prong’ the Five Star MoVement. The total number of posts referring to the first week of the campaign amounts to 838 records and the fourth week to 973.

Data collection includes information on user activity in response to content published by the three political parties considered in the selected time unit (number of likes, comments and shares recorded on these posts). During an exploratory phase of analysis, the main descriptive statistics were calculated on these data (average, min, max, standard deviation)<sup>2</sup>.

On the quantitative front, the Content Analysis techniques selected for comparative purposes are *Content Analysis as a survey* and *statistical-textual analysis* (semi-automatic – Roberts, 2000). *Qualitative analysis* showed its importance in all phases of the text interpretation: here, the latter – which is essential both in the pre-treatment phase of the data, and during the entire process of quantitative analysis to interpret and contextualize the results correctly – is used mainly for further investigation and empirical control. On the other hand, we do not provide specific guidance on highly automated techniques (for example, sentiment analysis, etc.), more suitable to big corpora connected with audiences<sup>3</sup>

---

<sup>1</sup> In the literature there are only a few research examples that propose an integrated use of Content Analysis techniques. Just to give some examples among the most recent works, we can recall: Frame et al., (2016); Di Giammaria, Faggiano (2017); Cernison, (2019); D’Ambrosio et al. (2019); Lombardo et al. (2019b).

<sup>2</sup> For further information on this analysis, see Gallo, Sallusti (2019).

postings and comments<sup>3</sup>. In an attempt to carry out an evaluation, the instruments and the application methods will be considered analytically with respect to each technique, reconstructing the limits and benefits system.

### ***3.1 The statistical-textual analysis of an online political campaign: an assessment***

The first methodological reconstruction of the empirical case, aimed at highlighting the positive and negative peculiarities of the adopted instrumentation, concerns the lexicometric analysis applied to electoral campaign data downloaded from Facebook. In particular, the statistical-textual analysis focused on all the collected posts in their strictly textual component.

The essential features of a study concerning the posts that leaders and parties publish on their official pages during the election campaign are the following:

1. although it is desirable to carry out comparative analyses in time and space, the lexicometric analysis is characterized by distinctly exploratory and descriptive goals, given the continuous mutability of the observed content (Grimmer, Brandon, 2013).
2. The analysis focuses on the strictly textual component of the considered units, ignoring the extra-textual elements (gifs, images, videos). If, thanks to special tools, the downloading of posts takes place automatically and easily, allowing a textual corpus to be obtained as an immediate result, it is instead far from being quick and simple to proceed to the ‘translation into words’ of the multimedia system that normally combines with the written message (e.g. slogans reproduced in image format, interviews with public figures, links to newspaper articles, extracts from television broadcasts, live streaming). For these reasons, the analyses one can more frequently find in the empirical literature exclusively focus on the posts’ text.
3. If centering the analysis only on the textual component of the posts represents a renunciation in some ways, on the other hand this choice allows one to work on the whole universe of the units produced by an issuer within a given space-time frame.

---

<sup>3</sup> We are convinced that the fully automated analysis is not particularly suitable for the study of political messages, which are in constant evolution (also due to the technological progress and the hybrid nature of the media system) and vary significantly depending on: the specific political force, the precise historical moment, the country/geographical sub-context. In this scenario, in which a plurality of innovative languages emerges, political vocabulary seems to be anything but static.

4. The preparation of the corpus to be subjected to analysis (cleaning, lemmatization, disambiguation, etc.) is particularly onerous, requires great attention and a lot of time (Nobile, 2018a, b).

Once a textual corpus has been imported, to which 'reading keys' have previously been associated (in our example the variables 'electoral campaign week' and 'political party'), whatever analysis software is used, it automatically produces descriptive statistics about the size and quality/heterogeneity of the text (lexical richness indices). The evaluation of these statistical parameters directs the subsequent analyses, providing indications on the analysis level that can be carried out, on the controls to be developed, or on the comparability among the textual sub-corpora under observation.

The 'key words', a set of words and linguistic expressions that describe in a synthetic form the overall meaning of the text, perform a particularly relevant function; they register a high frequency in the text and an appreciable statistical association with other terms. A focused study of the key-words makes it possible to identify the structural elements of a corpus (subjects, distinctive features, expressive modalities, actions, etc.), allowing to isolate a series of elements such as actors (parties, leaders), communication contexts (platforms and SNS grammar such as 'invitation to share', 'Facebook', 'Instagram'), contexts of action (electoral campaign, media and local events, news), territorial references (Italy, regions, cities), topics (scientific research, employment, education)<sup>4</sup>.

The multi-words, i.e. complex linguistic expressions that contribute to the reconstruction of topics and specific aspects of the investigation, add information in the same direction, also providing interesting data on the communicative style (in our empirical case, the one adopted by each political force). Here are some particularly significant expressions (each of them is, as is well known, associated with a specific frequency): 'vote League' (vota\_Lega), 'start chatting' (inizia\_a\_chattare), 'use the web' (usa\_la\_Rete), 'the Captain's page' (la\_pagina\_del\_Capitano), 'Italians first' (prima\_gli\_italiani), 'citizenship income' (reddito\_di\_cittadinanza), 'asylum seekers' (richiedenti\_asilo), 'DP program' (programma\_PD), 'common sense revolution' (rivoluzione\_del\_buon\_senso), 'European Union', 'illegal immigration', 'fake news', 'flat tax'. The expressions given as an example show the consistent

---

<sup>4</sup> It is important to clarify that some of the words/expressions herein cited as examples are the result of a conspicuous synthesis work aimed at bringing words back to broader semantic categories. For example, 'link' mode includes all the specific references to websites reported in the analyzed posts or even 'invitation to share' ('condivisione\_social') summarizes all the expressions that represent an invitation to spread the political message through social channels.

heterogeneity of the textual material. On the one hand, there are slogans and rhetorical expressions with a strong suggestive power, on the other, the call for online mobilization, voting instructions, reference to specific social categories and political intervention fields.

The appropriate use of the selected keys, which makes it possible to divide the corpus into several sub-corpora based on the modalities of the variables chosen with strategic function, allows for a deep study of the 'specificities'. In our example, the League is characterized by the strong presence of linguistic elements connected with: personalization of politics and leadership, online and offline mobilization of voters, enhancement of national identity and traditional values. The 5SM, on the other hand, is characterized by the following and recurrent indications: activation of the common citizen in support of the movement (online mobilization and fundraising); discrediting political opponents and traditional political parties; presentation of the candidates and of the future government team; radical change projects; contrast of social inequalities through the measurement of citizenship income. Finally, the Democratic Party is characterized by the use of a language which emphasizes the party role rather than the leader figure; the achievements (as a party of the outgoing government) and the political program in a logic of continuity; the use of a technical language supported by data, especially in relation to the Country's economic conditions; the reference to values, topics and actors typically associated with left-wing political orientation.

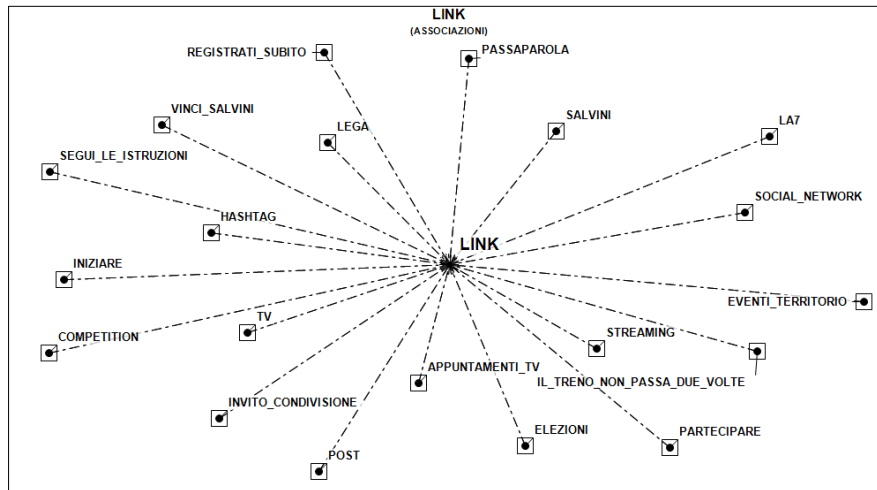
An operation that is as simple as it is essential, which connects the quantitative analysis with the necessary qualitative in-depth analysis, consists in the contexts of use study of the lemmas and of the key expressions that are particularly relevant in the research. Another crucial operation consists in tracing the polysemic verbal expressions that require clarification. This analysis allows for the production of useful classifications and detailed comparisons among two or more sub-corpora.

Another interesting output consists in the production of significant statistical association between linguistic lemmas/expressions (previously processed), in which the single selected lemma connects, with greater or lesser significance, with the others words (for example graph 1). The smaller the distance between two lemmas, the stronger their bond.

Being aware that a rich and large textual corpus can be analyzed with many other tools besides those briefly outlined, it is worth referring, in closing, to the multidimensional analysis techniques applicable to textual data. In greater detail, the analysis of lexical correspondences allows for the identification of the conceptual axes around which the empirical basis rotates (graph 2). Known for its power of synthesis based on the extracted factors, it is able to provide an

overall view of the data, which are semantically interpreted based on the known statistical coefficients associated with each lemma.

GRAPH 1. Radar: associations with 'link' lemma in the sub-corpus of League.



A graphic example is offered, in which two factors have emerged, that explain a large amount of the common inertia, named in the following way:

- First factor (which explains 42.56% of the common inertia) – *Communication strategy of the anti-system forces*. In particular, the first factor contains the indications coming from the two most visible political movements in the Italian political scene at the time of the 2018 General Elections. On the positive semi-axis, we observe the *salient elements of the 5SM campaign*, based on the size of the political change/cleavage. On the negative semi-axis we observe instead *the League's communication style*, completely focused on the dense interaction via social network on part of its leader Matteo Salvini with his own electorate, on crime-related news, anti-immigration politics, the media and territorial agenda.
- Second factor – *Communication strategy of traditional parties* (33.33% of the inertia explained). This conveys the political message of traditional parties. Here, since the authors have exclusively selected the empirical base referring to the main parties within the Italian three-way structure, only the connotative elements of the positive semi-axis emerge, coming under the *political campaign of the DP*. In this case, references to the work of the previous five-year government, to future projects in continuity with the past and to teamwork are evident.



analysis is co-essential with respect to the lexicometric analysis and, even if you work with big corpora, you cannot miss appropriate controls and in-depth analysis at the small data level (Popping, 2012; Grimmer, Brandon, 2013). It is obvious that only an accurate knowledge of the overall text – to safeguard against gross interpretative mistakes – can guide all subsequent processing, including the manipulation and synthesis operations, quality controls, targeted analyses.

- The intensity and quality of manipulative interventions for the corpus preparation, including the delicate process of explaining the previous passages, is established and managed in different ways by the single research group (there are no fixed thresholds or criteria in this sense). All this involves a strong weight of subjectivity and substantial limits on the results comparability, although it is possible to maintain a certain control of the specific final dictionary created (even considering its future re-use). The textual material at its original stage is comparable to a large skein: it needs to be both untangled and reduced in terms of complexity for the purpose of an adequate text reading (Grimmer, Brandon, 2013). It is evident that every time analytical verbal forms are transferred to more synthetic categories a conspicuous loss of information is generated (in terms of reduced lexical richness and lower data articulation); however, it allows statistical associations of sociological interest to emerge. The most complex task is to find the right balance between the need for synthesis and analytical data reporting.

Furthermore, the reflections expressed in the introduction are valid. In particular, if, on the one hand, the lexicometric analysis is suitable for elaborations involving even very large textual data universes, on the other, it neglects all textual and meta-textual elements, which are increasingly relevant and abundant in social networks sites, not present in the merely textual part of the post.

### ***3.2 Content Analysis as a survey applied to a digital campaign: advantages and disadvantages***

The second methodological report, which the authors propose to support this point of view on Content Analysis techniques, concerns Content Analysis as a survey<sup>5</sup>, applied to the same electoral campaign data, extracted from Facebook. For this second analysis about 50% of the posts of each political

---

<sup>5</sup> As is known, a Content Analysis form, comparable from a technical point of view to a semi-structured questionnaire, is appropriately tested before its extensive use. In the specific case, testing was performed on a heterogeneous set of cases per party.



party were randomly extracted. In this second case, the selected units of analysis were not only observed in their textual aspect, but also in relation to their multimedia dimension (external links connected to contents produced on other platforms/mainstream media; video: live streaming, spots, amateur footage; pictures and photos).

The features of such a study are schematically the following:

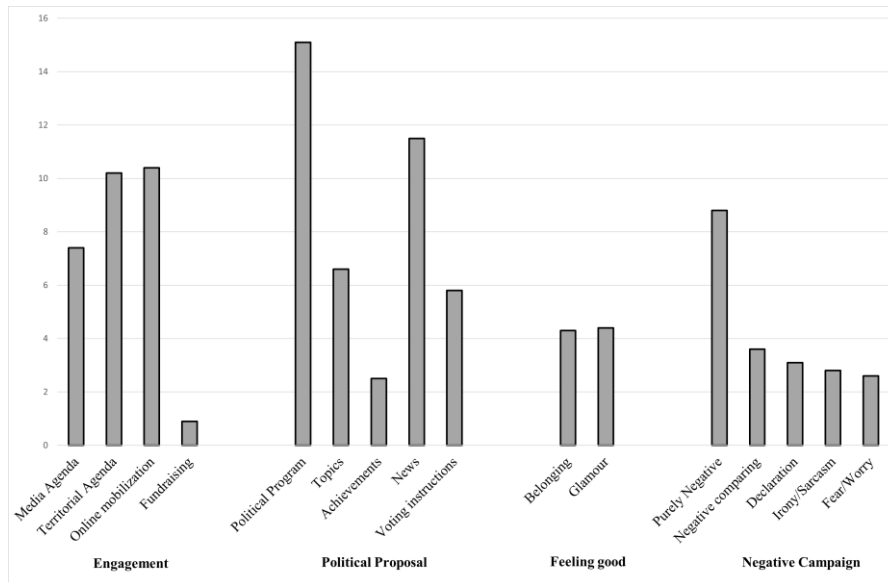
1. The Content Analysis as a survey aims to *enhance the interpretative-evaluative component of the analysis*, designing research instrumentation that, ‘upstream’, sift the text in a targeted way – classifying it, breaking it down, selecting it based on pre-established goals -, and not opting for an analysis based on the statistical association among words. For example, in our research example, the authors aim to *reconstruct* the communication strategies in use, the tones, the social representations, the roles of the actors involved, the media universe in which the electoral campaign unfolded. The *subjective component* that inevitably intervenes during the Content Analysis as a survey, must be kept under control by carrying out the classification work constantly in a group setting, and solving the critical aspects of meaning attribution through appropriate discussions and negotiations.
2. This technique allows one to *analyze the posts in their entirety* (text and visual/audio-visual equipment) taking into account the richness of the extra-textual elements that, especially nowadays, characterize online communication (political communication in our example);
3. If the chance to analyze the posts in their entirety makes it possible to identify elements that are not immediately visible in the purely textual part of a post, this choice involves a renouncement of the exhaustiveness of data collection and analysis; without forgetting that social research must also deal with the ‘time’ and the need to promptly produce scientific results. In fact, since the analysis is particularly onerous, complex and extensive, most of the time (unless the statistical universe is very limited) it is advisable to sample the empirical basis in order to obtain analytical and quality results a timely fashion.
4. It allows one to *benefit from the advantages of quantitative analysis* in the narrow sense (at the end of the inquiry one gets a data matrix ‘cases by variables’) *and of qualitative analysis* (by appropriately recording original extracts and impressions coming from the empirical basis). If, on the one hand, the information obtained through the quantitative analysis is valued, simplified and made comparable through appropriate descriptive statistics and graphical-tabular representations (on a mono-, bi- and multivariate level), on the other hand, the study of images/audiovisual content lends itself to qualitative in-depth analysis (in our example, political jargon and

communicative styles have been reconstructed and deepened through the punctual annotation of adjectives, verbs, key-expressions, etc.).

5. The Content Analysis form is designed with an 'ideal' data collection in mind which yields particularly rich results, complete and articulated, but, most of the time, several variables contained in the data matrix are only partially full (if not almost empty), due to an imperfect correspondence between the designed instrument and the available empirical basis. All this has an impact on the analyses that can actually be performed on the quantitative front. Indeed, the analysis of data – sophisticated, in varying degrees, based on the nature of the variables and the size of the sample – can exceed the bivariate level as long as an excessive number of missing values has not been recorded on a large set of variables. In the exemplified case, although the total number of registered records is high, some zooms on the data could not exceed the bivariate level, due to the limited number of valid cases on different variables.

Below the authors present some typical examples of analysis with the relative graphic-tabular equipment obtained by performing both quantitative and qualitative analyses (descriptive statistics coefficients, tables, graphs, synoptic schemes).

GRAPH 3. Function of the posts published on Facebook on the official pages of the League, the 5SM and the DP during the electoral campaign (%).



The first example graphically represents the data on the overall communication strategy used in the electoral campaign by the selected parties, distinguishing between conceptual macro- and micro-categories (graph 3).

Moreover, the double entry table shows the political parties, classifying the posts between those that contain a reference to a specific territorial context and those which do not contain it.

TABLE 1. Political parties (League, 5SM, DP) based on the territorial reference contained in the posts.

Reference to territorial context	Political Party/Movement			Total
	Lega	Movimento 5 Stelle	Partito Democratico	
No	51.8	49.6	85.2	54.0
Yes	48.2	50.4	14.8	46.0
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

p=.000

As already mentioned, with the aim of analyzing the communication style and political jargon, the entire system of adjectives, verbs, key-expressions found in each Facebook post was annotated, classified and represented in elaborate descriptive tables<sup>6</sup>.

Enhancing the variables without/with a few missing values recorded in the data matrix, we applied the technique of multiple correspondence analysis (MCA – graph 4) after their simplification (mostly dichotomization); through this technique we identified two particularly significant factors (which summarize respectively 21% and 14% of the common inertia).

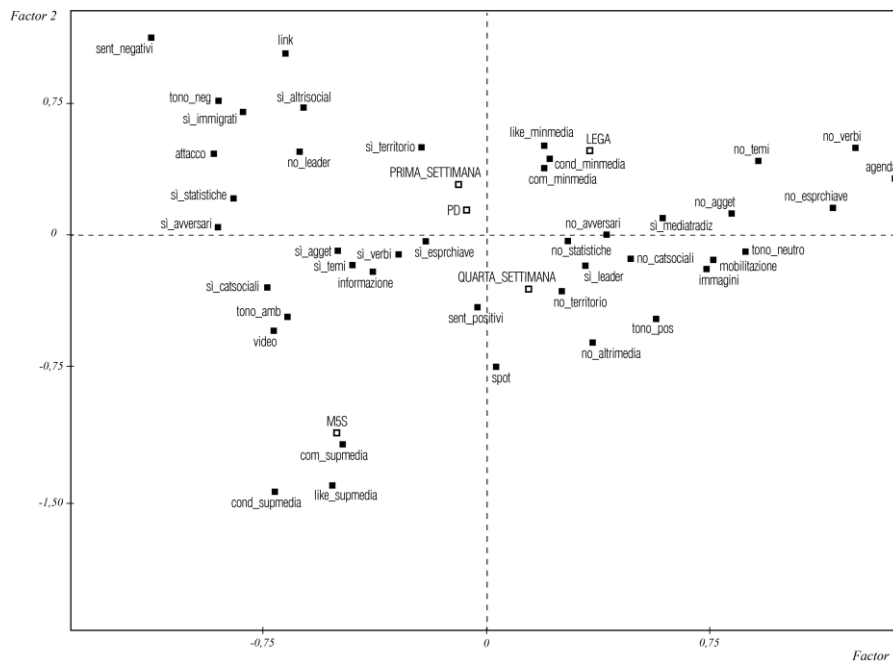
The first refers to League (the negative semi-axis summarizes the sentiment of hostility against immigrant-enemies and the positive one reflects the so-called nativism-primatism typical of right-wing populist movements)<sup>7</sup>. The second factor represents on the negative semi-axis the 5SM and its political communication centered on the use of the web and digital platforms (*values of*

<sup>6</sup> For further information on this topic and on the data analysis, see Lombardo et. al., 2019a.

<sup>7</sup> The graph shows the short strings associated with the modes of the synthesized categorical variables. For example, on the negative semi-axis you can see modes like: *negative feelings* ('sent\_negativi'), *the indication* (with a negative connotation) *of immigrants* ('si\_immigrati'), *negative communication strategy* ('attacco'), *numbers and statistics use* ('si\_statistiche'); on the positive semi-axis there are modes like: *traditional media* ('si\_mediadiz'), *online mobilization* ('mobilitazione'), *Salvini leader* ('si\_leader'), *italian people rights* ('no\_catsociali'), etc.

comments, shares and likes on SNS higher than average – com\_supmedia; cond\_supmedia; like\_supmedia); on the positive one the DP (in a position rather close to the origin of the axes), less inclined to use slogan and particularly sensitive to certain topics and social categories (*sensitivity towards the weaker social categories; posts function: information; topics variety*, etc. – ‘si\_catsociali’; ‘informazione’; ‘si\_temi’).

GRAPH 4. Projection of active and illustrative variables-modes on factor axes - Multiple Correspondences Analysis (MCA).



Also for the Content Analysis as a survey it is advantageous to draw up an initial assessment of its usefulness and robustness within the sociological analysis of textual corpora downloadable from Social Network Sites, in particular in the field of political communication. Below are the main schematic observations:

- This instrumentation allows to quantify the most recurrent phenomenologies and to highlight the main statistical trends starting from a given text, on the basis of a careful extrapolation, counting, interpretation/classification/referral to general categories; on the other hand, at the same time, it allows us to investigate less visible and minority aspects, judged as relevant by the research team. For obvious reasons, the

more detailed/ambitious the planned Content Analysis form is (given an optimal knowledge of the analysis context and of the text), the more achievable the goal of fruitfully combining quantification with deep qualitative study appears.

- The interpretative and evaluative purpose that characterizes the Content Analysis as a survey leads to combining two levels of results production: on the one hand, the research team reaches relevant outputs on a *manifest* plan (e.g. producing statistics and percentages after punctual counting operations); on a *latent* level, it gets to conclusions that respond to a logic of abstraction, contextualization, comparison, modeling. The interpretation, during the data collection and analysis, implies, as mentioned above, the enhancement of subjectivity; on the other hand, to guarantee scientific results, the cognitive-interpretative skill required of each analyst must be constantly monitored and translated, within an assiduous group work, into inter-subjective knowledge. A well-tested and promising Content Analysis form is a tool that can be reused within a given field of study, even in a perspective of longitudinal and comparative analysis.
- Content Analysis as a survey, carried out through a full and detailed form (aimed at enhancing, as previously mentioned, also all the extra-textual elements), can be considered a self-sufficient text observation technique, even when, for reasons of time and availability of resources, sampling is necessary<sup>8</sup>. Nevertheless, the authors consider the combination of the Content Analysis as a survey with statistical-textual analysis techniques desirable, both to get the data quality, and to widen and strengthen the results.

Conducting Content Analysis as a survey – contrary to what happens for the statistical-textual analysis which ‘starts’ from the raw text – implies the ex-ante construction of a data collection instrument, deeply bound to the analysis goals. The concrete fallout of this methodological choice is that the questions addressed to the text enhance a portion, however broad, and not the totality of the contents. In fact, a larger or smaller portion/segment of the original textual/extra-textual material will never be subjected to analysis. This outcome is evidently connected to the choices set upstream, which can be defended as much as they are the result of shared negotiation and the development of wide and lasting research programs. The loss of information also in this case is inevitable, but with it the research team does the math in the investigation start-up phase, assessing ‘costs’ and ‘benefits’ during the research planning step.

---

<sup>8</sup> It is assumed that the sampling is random and the specific extraction strategy is based on the peculiarities of the empirical study and its interpretation keys.

#### 4. Findings

In order to draw a conclusive evaluation, it appears useful to make first-order reflections based on the empirical case presented: what *useful methodological legacy* does it leave us? Which *virtuous scheme of work* does it represent? This while always bearing in mind that we are talking about *style* and *heritage* in the belief that the concatenation of research steps presented can effectively be transposed into other research opportunities, with reference to 1. *other research topics*; 2. *other contexts* (virtual or otherwise) *of observation and collection of text-based material*; 3. *a dataset much larger than even the one considered for illustrative purposes* (to the point of crossing over into the panorama of Big textual corpora).

We aimed to highlight the usefulness of adopting a mixed approach intended to maximize the advantages of the combined use of analysis techniques and tools, that is, to reduce the disadvantages arising from exclusive use of a technique. The choice fell on this strategy because: from a methodological point of view, it guarantees greater quality and robustness of results; from a substantive point of view, it allows for a greater variety and depth of the results themselves (Faggiano, 2016; Chakrabarti, Frye, 2017). As can be seen, differentiated outputs are obtained based on the adopted technique, constantly considering their complementarity and correspondence as advantageous if the collected data and the performed analyzes are high quality. The combination of techniques and tools is, in our opinion, always very fruitful, although each research design must necessarily deal with the resources available to it; therefore, what we have presented through a selected series of examples is an ideal process of analysis, capable of enhancing every textual and meta-textual element of the empirical basis. In other words, the techniques illustrated in the previous pages – Content analysis as a survey, Statistical-textual analysis, Qualitative analysis of textual and extra-textual material – are considered co-essential in an analysis design aimed at maximizing the value of a rich and varied initial corpus (Hashemi, Babaii, 2013; Pashakhanlou, 2017). However, in the actual research practice one can rarely find opportunities for joint use of all available techniques, also because – in every concrete investigative situation – data collection and analysis designs are structured based on the screening of the actual available resources (human, temporal, economic). Nevertheless, full awareness of the limits and virtues of each instrument in use can shorten the distance between an *ideal* and a *practically feasible* research design, which tends towards maximum effectiveness of the research results. For each perspective, the large amount of work required to complete a satisfactory path of analysis was highlighted. In addition, for each technique – when an exclusive and non-combined use was established – an evaluation was made in terms of benefits and limitations (distinguishing the effects on data collection from those

connected to the analyses' potential).

The analytically-performed assessments allow us to confirm, in our findings, the usefulness of an integration perspective, and to guarantee the most appropriate balancing of the analysis, appropriately mixing a series of operational choices with broad repercussions on the quality and solidity of the results. Through a targeted series of research examples drawn from the world of political communication on Social Network Sites, we have not provided a system of alternative options but, rather, outlined an analysis process in which possibilities are not discarded, but opportunities are combined on several levels: whether or not to sample analysis units according to the set goals and the specific research phase; big and small data in the selection of relevant and significant contents; exploratory-descriptive and interpretative-explanatory objectives; automatisms and 'paper and pencil' operations; enhancement of textual and extra-textual elements of a given corpus; quality and quantity in the result production.

Although the predilection for research designs that enhance the concomitant use of multiple Content Analysis techniques – adopting different observation and analysis perspectives, as well as constantly oscillating between quality and quantity – is the main indication of this paper, numerous further reflections can be made explicit at the end of this *methodological evaluation process*, in which the key words are *integration* and *combination*.

A first possible consideration refers to the vast range of *software* available today for the *analysis of textual* corpora, which, in our opinion and wherever possible, would be worth using jointly within the same research design, oscillating once again between quality and quantity, combining multiple approaches rather than proceeding in a unidirectional way for a single opportunity. One may think, for the qualitative aspect, of *packages* such as N-Vivo (also suitable in many cases for broad empirical bases), connected to the tradition of *the constructivist Grounded Theory*, or, in strictly quantitative terms, to software such as T2K, aimed at automated text analysis and sentiment analysis. On the one hand, more analytically, computer-assisted reading and interrogation of the textual base in the direction of hypothesis formulation, construction of emerging theories and an accurate exploration of the corpus are featured. In this sense, the essential units in which the analysis is structured are the *sensitizing concepts* (themes, key concepts, semantic categories assigned by the researcher to single portions of the text when reading and interpreting the corpus through the software). On the other hand, among the potential of the analysis, the following stand out: the extraction of key terms and significant semantic entities on the basis of an automated language learning process, connected – after studying and polishing the text – to specific algorithms; the construction of multidimensional maps of the relationships between extracted

information; the creation of glossaries and the computational inspection of the textual basis; the extraction, again in automated form, of information connected with widespread tones and attitudes. Between these two analysis scenarios, in some ways opposite, yet in our opinion complementary, in the ever-growing *contemporary toolbox* there are many types of software and as many possibilities for processing textual material, which can be associated with the *different traditions* and *vocations of research and analysis*, precisely, interpreted with a view to *maximum joint enhancement*, and a *constant sensitivity* towards a fitting *balance between automation/use of Big Data* and *depth/use of Small Data* (Di Giammaria, Faggiano, 2016).

These considerations and those expressed in the previous paragraphs, starting from the introduction, lead us to draw attention to another interesting point: CA, in addition to associating itself fruitfully with a *Mixed Methods* approach (Amaturo, Punziano, 2016), is often inserted in the context of articulated and complex research designs, “coherent”, as it were, with the observed social complexity, wherein consequent effective research processes make room for methodological pluralism, interdisciplinarity (the field of Humanities increasingly collaborates and dialogues with disciplines considered – until some time ago – light years away and, simultaneously, *being contaminated by* and *itself contaminating knowledge*) and the sharing of differentiated skills (Corchia, 2019). In this sense, still trying to “export” what has been exemplified to other areas, it seems useful to recall the broader study program in which the reported case study is inserted. CA, in its various forms, applied to the 2018 digital campaign of the main Italian political parties, as well as connecting to a broader time perspective (the Electoral Sociology Observatory of Communication and Social Research Department of the Sapienza University has a decades-long tradition of studies and deals with digital campaigns, conducted on Facebook and Twitter, since 2013 in a comparative perspective), is associated with an elaborate system of scientific works that contemplates at its core: the analysis of the political offer over time (electoral programs, speeches by politicians, analysis of political party/movement websites – in this case, too, CA plays a leading role); the study of voters’ attitudes and voting intentions (through *survey research*, in the version of the *web survey* as well); the analysis of electoral flows and the geopolitical mapping of electoral results; the analysis of the political message on digital platforms other than Facebook and in the broader media landscape (TV, radio, etc.). In the previous pages we have presented analysis results obtained with different techniques; as one may recall, the posts published by the parties during the election campaign on Facebook represented the units of analysis. One may think, having reached this point – still adopting a broader perspective, which allows one to “go beyond” concrete examples and draw useful methodological ideas for other research opportunities



– of the possibility of further examining the points of view of political subjects involved in posting activities (and/or their entourage) through survey techniques such as in-depth interviews or focus groups (also in their online capacity) and/or the opportunity to reconstruct the social profiles of a generation of political actors, also by making comparisons that go beyond national borders, through a web survey or an adequate “interrogation” of the Net; think, moreover, of the possibility of studying the political message in other media contexts, not necessarily “social” ones (from the news, to television broadcasts, to the press, etc.) with the same temporal unit; or, also, an in-depth study of the communication strategies in use according to the channel being analyzed/the selected observation context (traditional media, Social Network Sites, local events); or the analysis of the reactions and social interactions generated by the political message (one need only think of the likes or comments on politicians’ posts by ordinary people); or, finally, that of the vast universe of links and hashtags shared on the Internet in relation to a specific thematic/problematic area. CA lends itself to meeting various of the study needs listed above and will produce its best results whenever quality and quantity can meet, as well as when analysts from different disciplinary sectors (IT, statisticians, political scientists, communication experts, sociologists, etc.) will decide to form a united and collaborative team. As anticipated, such an articulated system of opportunities, which substantially marks a rebirth of CA today, fits into the wider panorama of Digital Society, fruitfully interconnecting many of its objects, contexts and tools.

## References

- Amaturo, E., Punziano, G. (2016), *I «Mixed Methods» nella ricerca sociale*, Roma, Carocci.
- Antenore, M., Faggiano, M.P. (2013), ‘RT is (not) endorsement. Chiacchiere, commenti e racconti dalla campagna elettorale’, in Morcellini, M. et al. (eds), *Talk&Tweet. La campagna elettorale 2013 tra TV e Twitter*, Santarcangelo di Romagna, Maggioli, 209-238.
- Bail, C.A. (2014), *The cultural environment: measuring culture with big data*, *Theory and Society*, 43, 465-482.
- Bimber, B. (2014), Digital media in the Obama Campaigns of 2008 and 2012: Adaptation to the personalized political communication environment, *Journal of Information Technology, Politics*, 11, 2, 130-150.
- Cavallaro, M., Diamanti, G., Pregliasco, L. (2018), *Una nuova Italia. Dalla comunicazione ai risultati, un’analisi delle elezioni del 4 marzo*, Roma, Castelvecchi Editore.

- Cernison, M. (2019), 'The Campaign for Water on Facebook', in Cernison, M. (ed), *Social Media Activism. Water as a Common Good*, Amsterdam, Amsterdam University Press, 165-202.
- Cepernich, C. (2015), *Web politics». Un bilancio*, Storia del pensiero politico, 3, 475-492.
- Chadwick, A. (2013), *The hybrid media system: Politics and Power*, New York, Oxford University Press.
- Chakrabarti, P., Frye, M. (2017), *A mixed-methods framework for analyzing text data Integrating computational techniques with qualitative methods in demography*, Demographic Research, 37, 1351-1382.
- Conte, R. (2016), *Big Data: un'opportunità per le scienze sociali?*, Sociologia e Ricerca Sociale, 109, 11-27.
- Corchia, L. (2019), *Political communication in Social Networks Election campaigns and Digital Data Analysis: A Bibliographic Review*, Rivista Trimestrale di Scienza dell'Amministrazione 2/2019, 1-50.
- D'Ambrosio, G., De Marchis, E., Gallo, R., (2019), 'Two Presidents in Comparison: Explore the Obama and Trump Communication Approaches by Means of Content Analysis Tools', in Ruggiero, C., Arslan, H., Gianturco, G. (eds). *Contemporary Approaches in Social Sciences Researches*, Bialystok, Poland: E-BWN, 363-370.
- Di Giammaria, L. (2016), 'Il M5S dal blog ai social network. Analisi etnografica di temi e conversazioni on line', in Faggiano M.P., Nobile S. (eds), *La politica partecipata. Il M5S di Roma e il suo blog*, Roma, Aracne.
- Di Giammaria, L., Faggiano, M.P. (2016), *Tra big e small data. La partecipazione politica degli utenti Twitter*, Sociologia e Ricerca Sociale, 109, 143-158.
- Faggiano, M.P. (2016), *L'analisi del contenuto di oggi e di ieri. Testi e contesti on e offline*, Milano, FrancoAngeli, 1-116.
- Di Giammaria, L., Faggiano, M.P. (2017), *Big Text Corpora & Mixed Methods - The Roman Five Star Movement Blog*, Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, Sage Journal, 133, 46-64.
- Frame, A., Mercier, A., Branchotte, G., Thimm, C. (2016), *Tweets from the Campaign Trail. Researching Candidates' Use of Twitter During the European Parliamentary Elections*, Bern, Peter Lang AG.
- Gallo, R., Sallusti, S. (2019), 'The Statistics of 2018 Italian Digital Campaign: Political Parties on Facebook and the Engaged Audience', in Ruggiero, C., Arslan, H., Gianturco, G. (eds), *Contemporary Approaches in Social Sciences Researches*, Bialystok, Poland: E-BWN, 353-362.
- Grimmer, J., Brandon, M.S. (2013), *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts*, Political Analysis, 21 (3), 267-97.

- Hashemi, M.R., Babaii, E. (2013), *Mixed Methods Research: Toward New Research Designs in Applied Linguistics*, *The Modern Language Journal*, Vol. 97, No. 4, 828-852.
- Klüver, H. (2009), *Measuring Interest Group Influence Using Quantitative Text Analysis*, *European Union Politics*, 10 (4), 535-49.
- Krippendorff, K. (1980), *Content Analysis. An Introduction to Its Methodology*, New York, Sage.
- Kropf, M.E., Grasset, E. (2017), 'Gendered Linguistics: a Large-scale Text Analysis of U.S. Senate Candidate Debates', in Browning R.X. (ed), *Advances in Research Using the C-SPAN Archives*. West Lafayette, Purdue University Press, 83-101.
- Lombardo, C., Novelli, E., Ruggiero, C. (2019a), *La società nelle urne. Strategie comunicative, attori e risultati delle elezioni politiche 2018*, Milano, FrancoAngeli.
- Lombardo, C., Faggiano, M.P., Gallo, R. (2019b), *Le campagne digitali del Movimento Cinque Stelle 2018-2019. Dalla squadra al leader?*, *DEMOCRAZIA E DIRITTO*, 1/2019, 7-49.
- Losito, G. (1993), *L'analisi del contenuto nella ricerca sociale*, Milano, FrancoAngeli.
- Losito, G. (2009), *La ricerca sociale sui media. Oggetti, metodi, tecniche*, Roma, Carocci Editore.
- Martire, F., Pitrone, M.C. (2016), *Lo studio dell'opinione pubblica al tempo dei big data. Una sfida per la ricerca sociale*, *Sociologia e Ricerca Sociale*, 109, 102-115.
- Medzihorsky, J., Littvay, L., Jenne, E.K. (2014), *Has the Tea Party Era Radicalized the Republican Party? Evidence from Text Analysis of the 2008 and 2012 Republican Primary Debates*, *Political Science and Politics*, Vol. 47, n. 4, 806-812.
- Morcellini, M., Antenore, M., Ruggiero, C., (2013), *Talk&Tweet. La campagna elettorale 2013 tra Tv e Twitter*, Sant'Arcangelo di Romagna, Maggioli.
- Nobile, S. (2018a), *Testi musicali e contesti sociali. Un'indagine empirica*, *Sociologia e Ricerca Sociale*, n. 117, anno XXXIX, 45-66.
- Nobile, S. (2018b), 'Analisi testuale, rumore semantico e peculiarità morfosintattiche: problemi e strategie di pretrattamento di corpora speciali', in Fioredistella Iezzi, D., Celardo, L., Misuraca, M. (eds), *Jadt' 18. Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*, Roma, Universitalia, 578-585.
- Pashakhanlou, A.H. (2017), *Fully integrated content analysis in International Relations*, *International Relations*, Vol. 31, Issue 4, 447-465.
- Popping, R. (2012), *Qualitative Decisions in Quantitative Text Analysis Research*, *Sociological Methodology*, Vol. 42, 88-90.
- Richerich, A. (2018), *The Big Data Agenda. Data Ethics and Critical Data Studies*, Westminster, University of Westminster Press.
- Roberts, C.W. (2000), *A Conceptual Framework for Quantitative Text Analysis*, *Quality and Quantity*, Vol. 34, 3, 259-274.

- Stefanizzi, S. (2016), *Small, open, big: i dati e la conoscenza scientifica*, *Sociologia e Ricerca Sociale*, 109, 117-126.
- Tipaldo, G. (2013), 'A Method for Analysing Television Quality', in Santangeli A., Tipaldo G. (eds), *Handbook of tv quality assessment a socio-semiotic approach for prix Italia jurors*, Exeter, Short Run Press Ltd, 16-43.
- Tipaldo, G. (2014), *L'analisi del contenuto e i mass media. Oggetti, metodi e strumenti*, Bologna, Il Mulino.
- Valbruzzi, M. (2018), *Analisi della campagna elettorale 2018*, Istituto Cattaneo, [www.cattaneo.org](http://www.cattaneo.org).
- van Dijck, J., Poell, T., de Waal, M. (2018), *The Platform Society: Public Values in a Connective World*, New York, Oxford University Press, trad. it. Boccia Artieri G., Marinelli A. (eds) (2019), *Platform Society. Valori pubblici e società connessa*, Milano, Guerini e Associati.
- Wells, C., Shah, D.V., Pevehouse, J.C.W., Yang, J.H. (2016), *How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning*, *Political Communication*, 33, 669-676.
- West, D.M., Maisel, L.S., Clifton, B.M. (2006), *The Impact of Campaign Reform on Political Discourse*, *Political Science Quarterly*, Vol. 120, n. 4, 637-651.