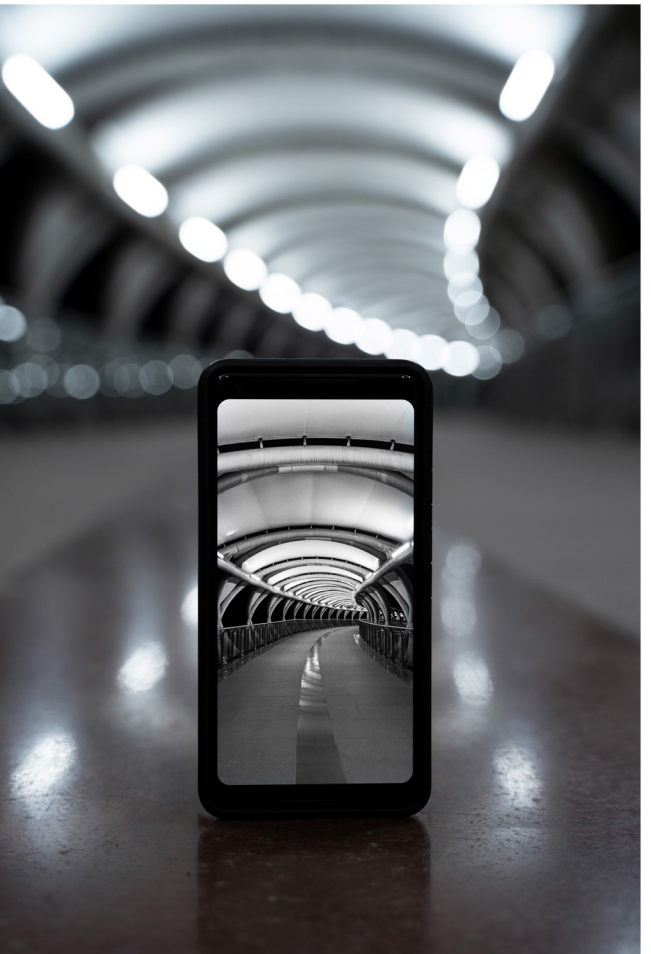
CIRCe
Centro
Interdipartimentale
di Ricerca
sulla Comunicazione

ILEXIA
Rivista di semiotica
Journal of semiotics
33–34

SEMIOTICA E DIGITAL MARKETING

SEMIOTICS AND DIGITAL MARKETING

a cura di
Cinzia Bianchi
Giovanna Cosenza



Lexia. Rivista di semiotica, 33–34
Semiotica e Digital Marketing
ISBN 978-88-255-3542-6
DOI 10.4399/97888255354266
pag. 89–107 (giugno 2020)

Data–driven Semiotics and Semiotics–driven Machine Learning

Leonardo Sanna*

Abstract: Nowadays there is a huge and growing variety of digital data. Despite the obvious relevance for the humanities and the social sciences, these massive quantities of data, usually defined as "big data", are mainly selected and analyzed using computer science and statistics. The paper proposes a theoretical and practical approach to the analysis of large quantities of data within the field of semiotic analysis. The main claim is that semiotics should dialogue with IT and statistics, that are essential to deal with the vastness and continuous variability of data. In particular, machine learning might become really useful from a semiotic perspective. In this work, we use a machine learning technique that is used in Natural Language Processing (NLP), to create a vector space based on probabilities of co–occurrences of words. In a distributional semantics perspective, this space is interpreted as a representation of semantic relations among words. We present then two directions in which we could intend the joint effort of semiotics and machine learning. In the first case, we propose a case study of semiotics–driven machine learning, in which we create a dataset starting from a semiotic analysis. In the second case, we present an example of data–driven semiotics, were the semiotic tools are used on an existing dataset, that was not build with semiotic scopes. The two directions have not to be intended as a dichotomy but instead as a part of a joint effort where semiotics interacts with machine learning and machine learning interacts with qualitative analysis.

Key–Words: Big Data, Digital Methods, Machine Learning, Word Embedding.

1. Background

Digital data and the so–called new media are problematic for the humanities and semiotics is not an exception. The semiotics of the new media has studied signification over the World Wide Web since

* PhD student, University of Modena and Reggio Emilia.

its foundation (Cosenza 2014), and one of the most relevant topics has been interface analysis, especially evaluating the usability of websites and their interactivity, in collaboration with other disciplines such as cognitive ergonomics and usability engineering (Adami 2013; Derboven *et al.* 2012; Nazrul Islam and Tétard 2014).

The literature in the field of applied semiotics is still limited, despite the variety of digital objects that could be studied. Semiotic studies on the social media, for instance, are still scarce, fifteen years after the foundation of the semiotics of the new media. However, it is not just a matter of quantity, it is a matter of the approach that semiotics has been adopting towards social media. Few studies tried to deal with theoretical matters (Kress 2009; Maggi 2014; Mirsarraf *et al.* 2017), while more often there have been applied studies based on text analysis (Peverini 2014; Finocchi 2015; Bonilla and Rosa 2015; Madison 2016). The applied studies that have been carried out on social media are case studies based on small–data corpora such as the works of Marrone (2017) and Peverini (2017).

The aim of this paper is to propose a preliminary outline of a new methodological approach to digital media, in which semiotics could deal with very large corpora. In this perspective, it is essential to make semiotics establish dialogue with other disciplines like computational linguistics and digital methods. In fact, although semiotics has started a reflection on the methodological challenges imposed by a digital scenario (Ferraro and Lorusso eds. 2016), it is not possible to do research on social media and on the web without an interdisciplinary approach.

Digital methods are an extensive and multidisciplinary framework to approach digital–born data (Rogers 2013). During the last decade, this discipline has been dealing with a variety of case studies mainly in the area of social sciences (Rogers 2010; Rieder 2012; Borra *et al.* 2014; Rieder *et al.* 2015; Helmond *et al.* 2017; Rogers 2017; Nieborg and Helmond 2019). Even though digital methods are mainly developed by social scientists, their epistemology is broad enough to work, with the necessary adjustments, in every discipline within the area of the humanities. Hence, we propose to use digital methods as the main discipline to take into consideration for methodologies of social media analysis.

On the other hand, computational linguistics is necessary to have the know–how on Natural Language Processing (henceforth NLP). Computational linguistics has a quite long tradition of social media analysis, using a set of different tools all structured on a corpus–based

approach (Zappavigna 2011; Bamman *et al.* 2014; Aditya *et al.* 2015; Mewari *et al.* 2015; Apoorva and Pradeep 2017). Recently there have been some attempts of including NLP in a semiotic analysis (Chartier *et al.* 2019) and also some research works about the interaction between semiotics and quantitative methodologies (Compagno 2018).

In the following two sections, we present two different case studies in which NLP and digital methods have been used in combination with semiotics; in the discussion, after a brief overview on the findings of the two studies, we focus on the methodological aspects. As a matter of fact, both cases are explorative, the aim of presenting them is mainly to provide an insight on the development of new semiotic approaches to digital data, highlighting also their limitations. Finally, we conclude the paper proposing a first interdisciplinary research protocol to approach big data in semiotics.

2. Machine Learning: Word and Document Embedding

In this work, we present a set of machine learning techniques called *word embedding*, used in NLP to create a semantic model of the language that can be explored and visualized. The main idea is that every word is mapped to a vector and then represented as a geometrical n–dimensional space; words with similar meanings are represented close to each other, while words that have less semantic relations are represented more distant in the space. The basic concept on which word embeddings are founded is what is called the distributional hypothesis (Harris 1954). This hypothesis basically affirms that similar words usually occur in the same contexts.

> If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution. (Harris 1954, p. 156)

There are several ways of doing word embedding: word2vec (Mikolov *et al.* 2013a), fastText (Bojanowski *et al.* 2017), Star Space (Wu *et al.* 2018) and RAND–WALK (Arora *et al.* 2015). In this paper, we will use the renowned version of word2vec, that has been largely used and documented in other studies (Ouyang *et al.* 2015; Ma *et al.* 2015,

Jang *et al.* 2019; Li *et al.* 2019). There are two versions of word2vec (Mikolov *et al.* 2013b), meaning that the training could be performed in two different ways: the first is the Continuous–Bag–Of–Words model (CBOW) and the second the skip–gram.

A continuous–bag–of–words is basically a bag–of–words in which context words are concatenated to predict the target word; the missing word is usually put at the center of the context window; thus, to predict the focus word Wt, the input will be Wt–2, Wt–1, Wt+1, Wt+2, so in this case the two following and the two preceding words.

The skip–gram model works in the opposite direction, predicting the context starting from a target word. So, we have to imagine inputting Wt and having its context as output. In a nutshell, the C–BOW computes the probability of a word to occur given its context, while the Skip–Gram computes the probability of a context to occur given a word.

The fact that word embedding is based on probabilities, rather than on simple co–occurrences, is really the added value of this tool. Counting co–occurrences was the most used method before word embeddings. One of the most famous count–based algorithms is for sure Latent Semantic Analysis (Landauer and Dumais 1997). Basically, Latent Semantic Analysis works creating a matrix in which one column represents words and the other represents contexts. An empirical evaluation performed by Baroni, Dinu and Kruszewski (2014) showed that predictive models outperformed their ancestors based on co–occurrences.

In a few words, the real innovation of word embedding is its capacity to capture semantic relations even between words that never co–occur in the same context. In fact, the main reason to use word embedding is that the word space model is not a representation of co–occurrences, but instead a representation of probabilities of co–occurrences. That means that word embedding can capture the semantic relation even *in absentia*, capturing also relations among words that never co–occur in our corpus. In more semiotic terms, we might say that word embeddings allow us to explore the semantic field of a word, highlighting the contexts that the term would evoke.

Besides word embedding, it is important to consider an algorithm that works on entire texts, called doc2vec or Paragraph Vector. According to what Le and Mikolov wrote in their paper (2014), paragraph vectors are created from word vectors:

> In our Paragraph Vector framework, every paragraph is mapped to a unique vector, represented by a column in matrix D and every word is also mapped to a unique vector, represented by a column in matrix W. The paragraph vector and word vectors are averaged or concatenated to predict the next word in a context. (Le and Mikolov 2014)

Basically, word vectors are somehow condensed in a single big vector representing the semantics of the whole document, creating a document embedding. With doc2vec we can easily explore text similarity in our corpus.

3. Semiotics–driven Machine Learning: Building a Dataset

By semiotics–driven machine learning we mean an analysis in which the dataset used for machine learning is built following the results of a semiotic analysis. In this case, the object of the study was the post–truth[1] issue, in particular the case of vaccines. The concept of post–truth is challenging for semiotics, since in traditional textual semiotics we do not intend "truth" as a direct correspondence with some reality outside the text (Lorusso and Violi 2004; Lorusso 2018): truth and veridiction are seen rather as the product of textual strategies whose aim is to gain the trust of the reader (Greimas *et al.* 1989). In our case study, we decided to investigate post–truth starting from the assumption that the circulation of alternative and competing truths is a mass–mediatic process, hence we decided to investigate its evolution starting from an analysis of the journalistic discourse in British newspapers.

The case of vaccines arose in 1998 when Andrew Wakefield published a controversial study about the MMR vaccine, claiming that he found out some correlation between MMR vaccine and autism. In 2004 a journalistic inquiry conducted by the "Sunday Times" found out that the study was fraudulent and in 2010 a tribunal proved the accusation to be true; Wakefield got expelled from the UK medical register and "The Lancet" officially retracted his paper.

1. According to Oxford Dictionary, post–truth is relating to or denoting circumstances in which "objective facts are less influential in shaping public opinion than appeals to emotion and personal belief".

In our analysis, we decided to take into consideration the English context, in the period of the controversy (hence from 1998 to 2010), taking into account four different newspapers, analyzing how the story evolved in the press. Our initial corpus was built with 34 articles from "The Independent", 20 from "The Telegraph", 20 from the "Daily Mail" and 20 from "The Guardian". These articles were selected trying to keep a balance between the newspapers, selecting at least 2 articles per year with goal of inquiring the evolution of the narrative[2].

The analysis highlighted that two newspapers were in favor of vaccines ("The Independent" and "The Guardian") and two against ("The Telegraph" and the "Daily Mail"). However, the point of view of the newspapers has sometimes been ambiguous; in fact, in the period between 1998 and 2004 even the newspapers that were clearly in favor of vaccines had sometimes a critical view towards health institutions, insinuating a series of doubt (more or less sneakily) towards vaccines safety.

This aspect appeared in texts as an isotopy that we called "the matter of trust" or simply "mistrust". Hence, despite differences in enunciation strategies, we noticed that each newspaper, no matter of its axiology, presented this isotopy in some way (see Figure 1).

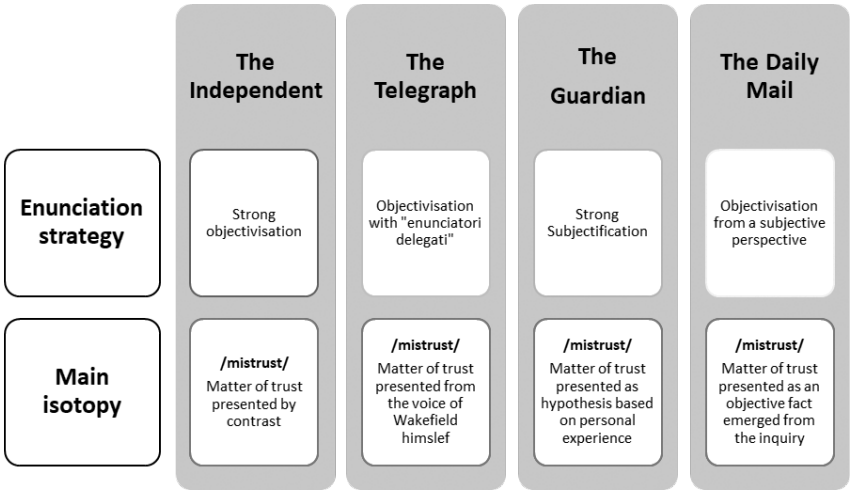


Figure 1. "Matter of trust" regarding vaccines safety.

2. In the "The Independent" archive we retrieved more articles about vaccines. Hence, we decided to include more articles to represent the proportions among our newspapers archives.

In the end the analysis suggested that texts involved in post–truth processes are structurally very similar. In particular, they are similar in their semantic structures, as they express similar isotopies. However, the similarity summarized in Figure 1 is the product of a qualitative investigation on a small corpus. Moreover, it appeared that the isotopy of mistrust somehow evolved diachronically; in fact, despite being sometimes present in its original form (as distrust towards institutions), in "The Guardian" and in "The Independent" this isotopy was also used to discredit the anti–vaccine counterpart.

Therefore, we decided to test our hypothesis, using a digital methods perspective to enlarge our corpus. We scraped the search engine Bing using BootCat, a computational tool that queries the search engine starting from a set of defined seed–words. These words are then combined in each possible combination, creating what is called a *tuple*. Each tuple can retrieve up to 50 URLs.

We used the results of our semiotic analysis to build our tuples, selecting four topics that emerged in the first stage of analysis: (a) The MMR controversy and Andrew Wakefield, (b) Mercury and chemicals in vaccines, (c) Flu vaccine, (d) Cervical cancer vaccine. Each tuple was built using words related to each of the topics, collecting the isotopies within our initial corpus. While building our queries we also tried to use ideological keywords (Rogers 2018) that were more likely to orient our search results in favor or against vaccines.

Afterwards, we manually identified and labelled texts in favor (henceforth ProVaxx) and against vaccines (henceforth AntiVaxx). We considered AntiVaxx all those texts which mentioned explicitly that vaccines were dangerous and also all texts which had the isotopy of mistrust against official science or institutions. On the other hand, we considered ProVaxx all texts which simply explained vaccines effects according to official science, those that clearly expressed their support for vaccinations and also debunking texts. In the end, we had 516 texts for AntiVaxx and 486 texts for ProVaxx, mainly coming from blogs and webzines, as shown in Figure 2. In total our BootCat corpus had 1002 documents and 1349252 words (word tokens), of which 29541 were unique words (word types).

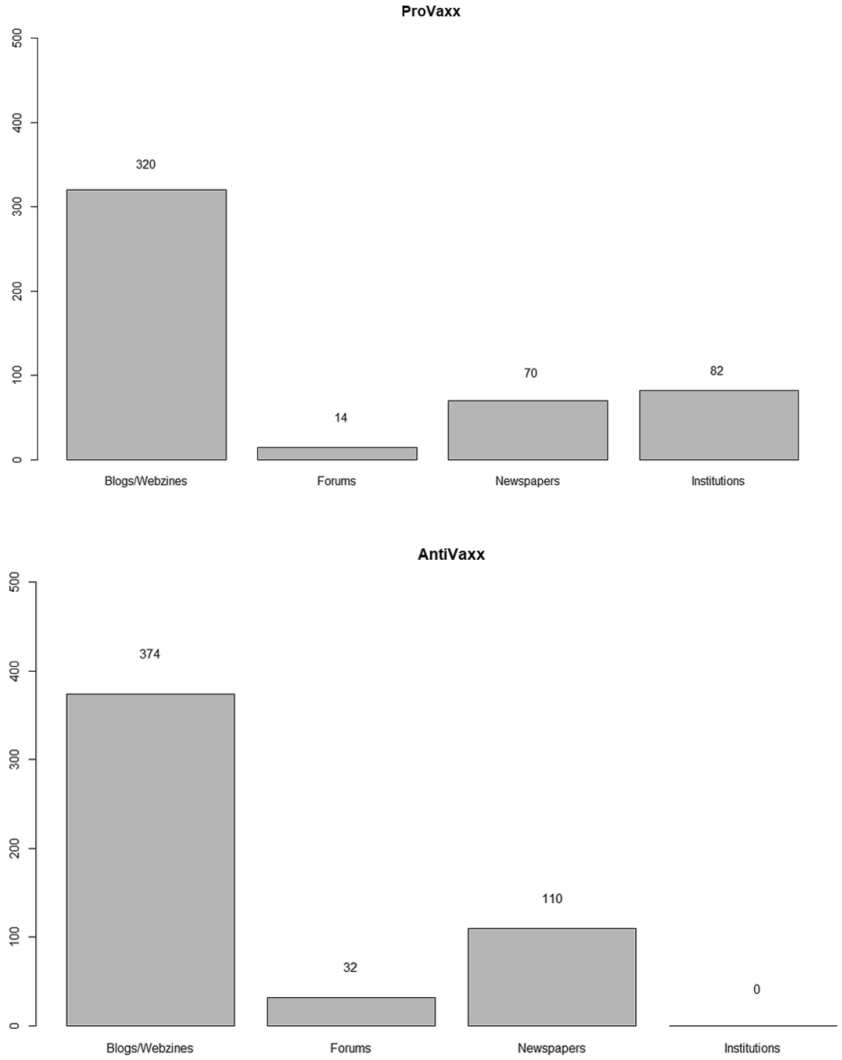


Figure 2. Text genres within the corpus built with BootCat.

The BootCat corpus was then processed with doc2vec, to visualize the degree of semantic similarity between the texts of the two factions. As shown in Figure 3, vectors are all mixed together, meaning that there is a semantic affinity between ProVaxx and AntiVaxx. The only

exception is a very small cluster[3] on the left, that might also be an artifact of our model[4].

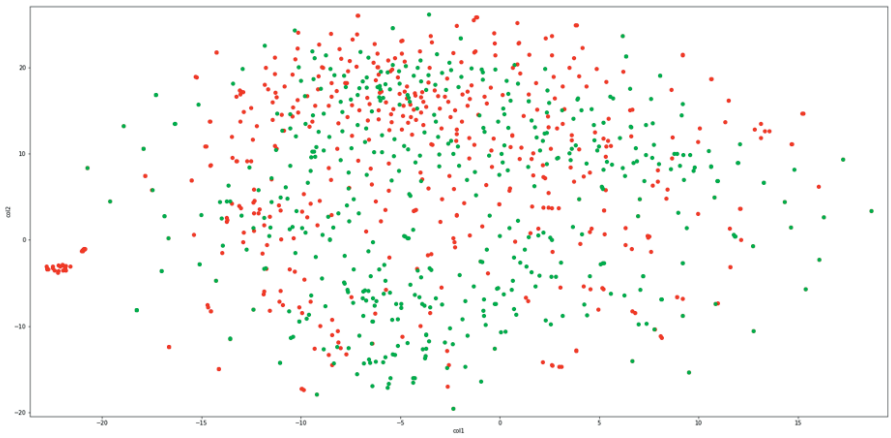


Figure 3. Doc2Vec showing that among AntiVaxx texts (in red) and ProVaxx texts (in green) there is no clear polarization.

4. Data–driven Semiotics: Analyzing the Data

Another example of semiotics working together with digital methods and NLP could be a case of what we call data–driven semiotics, namely when the semiotic elaboration follows the creation of the dataset.

In this paper, we present the case of the dataset built by Facebook Tracking Exposed, which is the first project of the tracking exposed manifesto. The main interest of the Tracking Exposed[5] group is to look for evidence of user profiling and algorithmic personalization. During the Italian elections in 2018, they studied algorithmic personalization of Facebook, highlighting that different users were unevenly

3. In data science, a cluster is a group of objects that are similar to each other compared to the rest of the dataset.

4. The small cluster has only articles from the "Daily Mail". This might be either for their high semantic coherence, compared to the rest of the corpus, or simply because those vectors have not been trained well. This might be due to words and contexts that appeared very few times in our corpus, e.g. advertisement banners.

5. See https://tracking.exposed/.

exposed to political communication during the electoral campaign (Hargreaves *et al.* 2018). The phenomena at stake in their project are also known as filter bubbles (Pariser 2011).

In a nutshell, they built six experimental Facebook profiles (henceforth bots); each profile followed 30 different political sources, covering the whole Italian political spectrum. However, each of these profiles interacted (liking posts) just with a specific political party, simulating an interest towards a political ideology. With the help of a tool built by the researchers, they collected evidence of what each bot was seeing on its personal newsfeed. According to their findings we have 6 different Facebook newsfeeds and each one is coherent with the simulated political ideology of the experimental profile.

In the dataset, we have the total amount of 79345 Facebook posts (called "sources dataset") made by the political sources and also all the posts that have been effectively selected for each bot (called "impressions dataset", 193809 posts in total). The sources dataset was explored with a semiotic analysis to inquiry semantic similarity among posts. In fact, since algorithmic personalization operates by homophily, it should be questioned whether semantic similarity is an influential criterion in the algorithm choices. We used doc2vec to inquire semantic similarity within the sources dataset, mapping the different political ideologies within our semantic space. Then, we compared the semantic affinity of each ideology with their impressions (what each bot saw on its newsfeed).

While the right–wing profile (Figure 4) seems to have visualized posts that are quite coherent with its alleged ideology, in the case of the Five–Stars movement (Figure 5) there is a huge part of the cluster that has not been shown to the Five–Stars bot[6].

6. We also remark that that part of the cluster was not shown to any of the profiles.

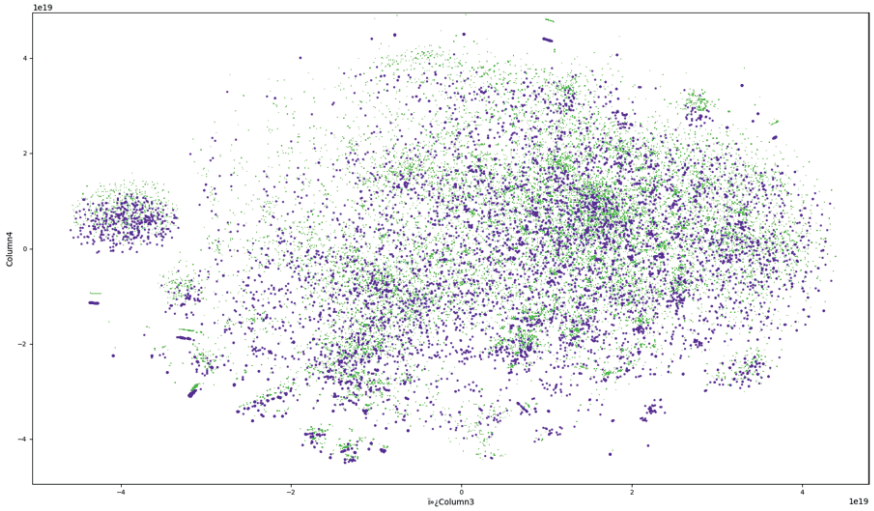


Figure 4. Right–wing filter bubble. In blue the posts seen by the profile, in green the actual posts made by right–wing sources.

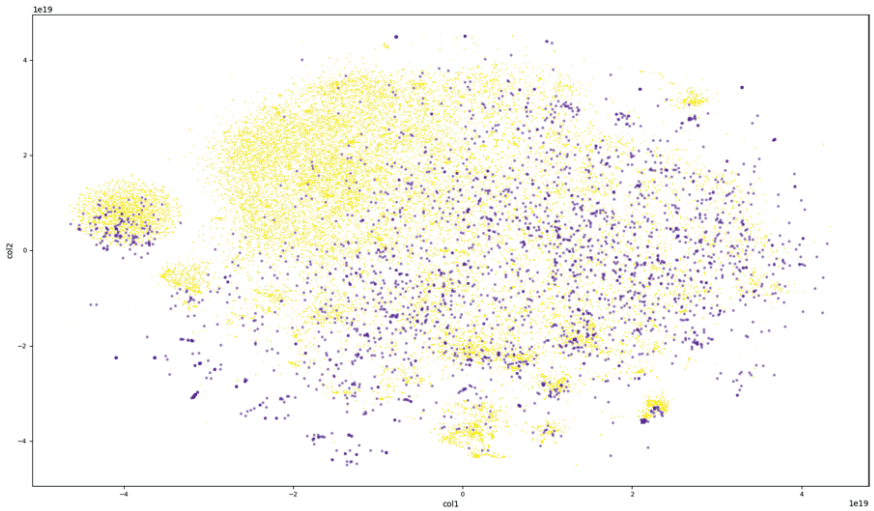


Figure 5. Five Star Movement filter bubble. In purple the posts seen by the profile, in yellow the actual posts made by Five Stars Movement sources.

At this point, a semiotician has to go back to the text to understand the machine learning output. A qualitative investigation showed then that the omitted part of the Five–Star Movement cluster was a set of "supportive" posts towards the party and the candidates, while in the case of the right–wing the content was mainly ideological and almost never simply supportive towards the main candidates. The role of the semiotic analysis is to further explore the dataset. In a data–driven perspective, the analysis starts from word vectors, exploring the semantic framing of words and expression as in Figure 6.

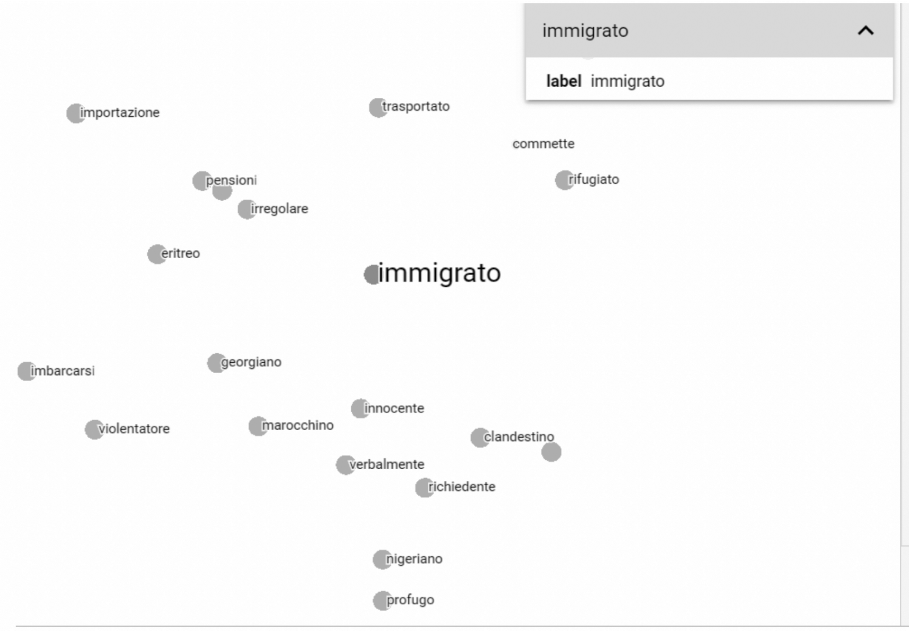


Figure 6. The semantic frame of the word "immigrato" (immigrant), showing the 20 most similar words according to Word2Vec. This semantic frame is the "general" frame, meaning that it emerges from the whole corpus.

We argue that the Word2Vec visualization (Figure 6) is nothing more than a computational operationalization of what has been theorized by Violi (1997); words have the power to select their context, therefore meaning is not constrained by context but instead words select their contexts. In this case the contexts selected are ascribable mainly to a right–wing ideology, meaning that the right–wing ideology, on this theme, was dominant in the dataset.

5. Discussion on the Findings

The main advantage of a semiotics–driven approach is that we can verify a hypothesis on a larger dataset. In applied semiotics, this might be really helpful, as it allows to generalize assumptions and to test possible biases due to researcher interpretation. On the other hand, the main limitation of this approach is the size of the corpus. In fact, although it was large enough for machine learning, it is quite small to have meaningful results in a quantitative perspective. However, this is the compromise of using machine learning within semiotic analysis; indeed, in the semiotic tradition the corpus is not built with statistical criteria[7].

In the end the hypothesis of similarity between the two factions seemed founded, as it appeared that the semantic structures were effectively similar. Hence, it is reasonable to think that post–truth processes might imply always semantic similarity because of opposed veridiction strategies; however, it also emerged that similarity might be also due to contradictory discursive patterns, e.g. "vaccines are dangerous" and "vaccines are NOT dangerous". In fact, distributional semantic models detect any type of semantic affinity among lexemes, including complementarity. Thus, it was not possible to determine if document embedding captured the same type of semantic similarity that we observed in our semiotic investigation. More general assumptions on post–truth would need larger corpora built with statistical criteria so that we can investigate both veridiction and complementary terms.

Conversely, working in a data–driven perspective gave us the advantage of working on a larger corpus and also to start our analysis without making a preliminary semiotic analysis. This has the benefit of working directly on evidence that emerges from data, which is normally not possible in semiotics. The research question in this case was to understand if semantic similarity was influential for algorithmic personalization.

Specifically, we found that:

1. the filter bubble is not always influenced by semantic similarity;

7. According to Greimas and Courtés (1979, Eng. trans. p. 64), "the representativity of the corpus can be obtained either by statistical sampling or by saturation of the model. In the latter case, the model, constructed on the basis of a segment chosen intuitively, is later on applied for confirmation, complement, or rejection to other segments until all the data are used".

2. the right–wing sources influenced the semantic framing of ideological words within the sources dataset, as shown in the example in Figure 6. This might be the reason of the different algorithmic treatment, but at this stage of analysis it is not possible to make more solid statements.

Even though the dominance of right–wing within the sources is not an algorithmic product, it is relevant to show that the right–wing and the Five–Stars Movement have been treated differently, as in the case of Five–Stars Movement a relevant part of their content was never shown to the bot. Nonetheless, further research is necessary to study the composition of the filter bubbles trying to understand: (a) how diverse the bubbles are and (b) if their composition is coherent with the composition of the sources dataset. The main limitation of this approach is that it allows only a preliminary exploration, leading to more semiotic–oriented research questions. Semantic similarity alone is not properly a semiotic research question, although its reasons are for sure interesting for the discipline. Starting from these findings it should be necessary to do a qualitative investigation on algorithmic personalization and on the reasons of semantic similarity within the clusters.

Regarding other technical limits of the approaches we presented, we must use machine learning always keeping a critical perspective. For instance, word embedding has some difficulties in representing words that occur few times in the corpus (Sergienya and Schütze 2015: Faruqui *et al.* 2016; Pilehvar and Collier 2017) and might be a significant issue, depending on our case study. It is also important to remember that any computational tool should be used critically, understanding that the output is heavily influenced also by the quality of our dataset. In computer science, the motto in this case is "garbage in, garbage out", meaning that a poor dataset would outcome poor results.

6. Conclusions

We presented two new approaches to digital data in a semiotic analysis. As we hope to have shown, these two approaches have both strengths and limitations. Specifically, the data–driven method allows us to start the

analysis on robust findings, while the semiotics–driven method allows us to explore in deep the signification structures of our corpus. On the other hand, the huge amount of data creates also a great complexity, that entails a multidisciplinary approach. Thus, a purely quantitative methodology cannot capture the details of the object of study, at least not with the granularity required in semiotics. Conversely, keeping an idiographic approach prevents us from generalizing our conclusions, regardless of the enlargement of our corpus.

Following the tradition of digital methods, we should then create a new methodological know–how that is specifically meant to effectively study social media and related topics, since digital media and digital spaces are by now a part of our reality. In this perspective, it is necessary to overcome the dichotomy between virtual and real (Rogers 2009), combining qualitative and quantitative methods to enhance the semiotic approach to digital media.

Hence, we propose a preliminary methodological framework to approach digital objects within the semiotic analysis:

1. build an exploratory corpus for traditional semiotic analysis;
2. starting from the findings of step (1), build a larger dataset, scraping the web and/or social media using digital methods and computational techniques;
3. build a computational dataset (big–data corpus) for a semiotics–driven machine learning;
4. starting from the findings of step (3), carry out a data–driven semiotic analysis.

This methodological proposal is a combination of qualitative and quantitative techniques. In stage (1), the composition of the corpus follows the usual semiotic methodology, having in mind that we have to output something useful for stage two. With "useful" we mean something that can be used for data collection and machine learning, such as keywords or topics. However, in stage (2) it would be necessary to build the dataset with statistical criteria, so that we can dialogue better with other disciplines and we can generalize our findings of the qualitative stage.

There are still a number of questions that have not been discussed in this work. We conclude suggesting that further research on these

themes might follow two main axes: the epistemological issues of the quali–quantitative compound and the use of computational tools in experimental studies. As we are aware of the difficulties brought by these subjects, we suggest keeping a broad interdisciplinary perspective, dialoguing with statisticians, computer scientists and social sciences experts.

Bibliographic References

Adami E. (2013) *A Social Semiotic Multimodal Analysis Framework for Website Interactivity*, National Center for Research Methods, https://doi.org/10.1177/1470357214565583.

Aditya S. *et al.* (2015) *Recognizing Social Constructs from Textual Conversation*, "Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies": 1293–1298.

Apoorva T. and Pradeep N. (2017) "Aspect Based Sentiment Analysis with Text Compression", *International Journal of Computer Sciences and Engineering*, 5, 8: 63–66.

Arora S., Li Y., Liang Y., Ma T., and Risteski A. (2015) *A Latent Variable Model Approach to Word Embeddings*, arXiv preprint arXiv:1502.03520.

Bamman D., Eisenstein J., and Schnoebelen T. (2014) *Gender Identity and Lexical Variation in Social Media*, "Journal of Sociolinguistics", 18, 2: 135–160.

Baroni M., Dinu G., and Kruszewski G. (2014) *Don't Count, Predict! A Systematic Comparison of Context–counting vs. Context–Predicting Semantic Vectors*, "Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1)": 238–47.

Bojanowski P., Grave E., Joulin A., and Mikolov T. (2017) *Enriching Word Vectors with Subword Information*, "Transactions of the Association for Computational Linguistics", 5: 135–146.

Bonilla Y. and Rosa J. (2015) "#Ferguson: Digital Protest, Hashtag Ethnography, and the Racial Politics of Social Media in the United States", *American Ethnologist*, 42, 1: 4–17.

Borra E. *et al.* (2014) *Contropedia — The Analysis and Visualization of Controversies in Wikipedia Articles*, "OpenSym 2014 Proceedings".

Chartier J.F. *et al.* (2019) *A Data–driven Computational Semiotics: The Semantic Vector Space of Magritte's Artworks*, "Semiotica", 230: 19–69.

Compagno D. (ed.) (2018) *Quantitative Semiotic Analysis*, Springer, Basel.

Cosenza G. (2014) *Introduzione alla semiotica dei nuovi media*, Laterza, Rome–Bari.

Derboven J., De Roeck D., and Verstraete M. (2012) *Semiotic Analysis of Multi–touch Interface Design: The MuTable Case Study*, "International Journal of Human–Computer Studies", 70, 10: 714–728.

Faruqui M. *et al.* (2016) *Problems with Evaluation of Word Embeddings Using Word Similarity Tasks*, arXiv preprint arXiv:1605.02276.

Ferraro G. and Lorusso A.M. (eds.) (2016) *Nuove forme d'interazione: dal web al mobile*, Libellula Edizioni, Tricase (LE).

Finocchi G. (ed.) (2015) *Carte Semiotiche. Strategia dell'ironia del web*, La Casa Usher, Florence.

Greimas A.J. and Courtés J. (1979) *Sémiotique. Dictionnaire raisonné de la théorie du langage*, Hachette, Paris (Eng. trans. *Semiotics and Language: An Analytical Dictionary*, Indiana University Press, Bloomington, 1982).

Greimas A.J., Collins F., and Perron P. (1989) "The Veridiction Contract", *New Literary History*, 20, 3: 651–660.

Hargreaves E. *et al.* (2018) *Biases in the Facebook News Feed: A Case Study on the Italian Elections*, "IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE", 2018.

Harris Z.S. (1954) *Distributional Structure*, "Word", 10, 2–3: 146–162.

Helmond A., Nieborg D.B., and van der Vlist F.N. (2017) *The Political Economy of Social Data: A Historical Analysis of Platform–Industry Partnerships*, "Proceedings of the 8th International Conference on Social Media & Society", Article n. 38: 1–5.

Jang B., Inhwan K., and Jong W.K. (2019) *Word2vec Convolutional Neural Networks for Classification of News Articles and Tweets*, "PLS ONE", 14, 8.

Kress G. (2009) *Multimodality: A Social Semiotic Approach to Contemporary Communication*, Routledge, London.

Landauer T. K. and Dumais S.T. (1997) *A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge*, "Psychological Review", 104: 211–140.

Le Q. and Mikolov T. (2014) *Distributed Representations of Sentences and Documents*, "Proceedings of the 31st International Conference on Machine Learning" ICML–14, 32: 1188–1196.

Li B., Drozd A., Guo Y., Liu T., Matsuoka S., and Du X. (2019) *Scaling Word2vec on Big Corpus*, "Data Science and Engineering", 4, 2: 157–175.

Lorusso A.M. (2018) *Postverità*, Laterza, Rome–Bari.

Lorusso A.M. and Violi P. (2004) *Semiotica del testo giornalistico*, Laterza, Rome–Bari.

Madison N. (2016) *Digital Bisexuality: The Semiotics of Online Sexual Identity*, "AoIR Selected Papers of Internet Research", 5.

Maggi R. (2014) "Toward a Semiotics of Digital Places", in A. Resmini (ed.) *Reframing Information Architecture*, Springer, Basel: 85–102.

Marrone G. (2017) *Social media e comunione fàtica: verso una tipologia delle pratiche in rete*, "Versus", 125, 2: 249–272.

Mewari R., Singh A., and Srivastava A. (2015) *Opinion Mining Techniques on Social Media Data*, "International Journal of Computer Applications", 118, 6: 39–44.

Mikolov T., Sutskever I., Chen K., Corrado G.S., and Dean J. (2013a) *Distributed Representations of Words and Phrases and Their Compositionality*, "Advances in Neural Information Processing Systems", 26: 3111–3119.

Mikolov T., Chen K., Corrado G.S., and Dean J. (2013b) *Efficient Estimation of Word Representations in Vector Space*, arXiv preprint arXiv:1301.3781.

Mirsarraf M., Shairi H., and Ahmadpanah A. (2017) *Social Semiotic Aspects of Instagram Social Network*, "2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA)": 460–465.

Nazrul Islam M., and Tétard F. (2014) *Exploring the Impact of Interface Signs' Interpretation Accuracy, Design, and Evaluation on Web Usability: A Semiotics Perspective*, "Journal of Systems and Information Technology", 16, 4: 250–276.

Nieborg D.B. and Helmond A. (2019) *The Political Economy of Facebook's Platformization in the Mobile Ecosystem: Facebook Messenger as a Platform Instance*, "Media, Culture & Society", 41, 2: 196–218.

Ouyang X., Zhou P., Li C.H., and Liu L. (2015) *Sentiment Analysis Using Convolutional Neural Network*, "2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing": 2359–2364.

Pariser E. (2011) *The Filter Bubble: What the Internet Is Hiding from You*, Viking, London.

Peverini P. (2014) "Reputazione e influenza nei social media. Una prospettiva sociosemiotica", in I. Pezzini and L. Spaziante (eds.) *Corpi mediali. Semiotica e contemporaneità*, Edizioni ETS, Pisa.

——— (2017) *Daily Life in the Instagram Age. A Socio–semiotic Perspective*, "Versus", 125: 285–302.

Pilehvar M.T. and Collier N. (2017) *Inducing Embeddings for Rare and Unseen Words by Leveraging Lexical Resources*, "Association for Computational Linguistics", 2: 388–393.

Rieder B. (2012) *The Refraction Chamber: Twitter as Sphere and Network*, "First Monday", 17, 11.

Rieder B. *et al.* (2015) *Data Critique and Analytical Opportunities for Very Large Facebook Pages: Lessons Learned from Exploring "We are all Khaled Said"*, "Big Data & Society" 2, 2 https://journals.sagepub.com/doi/10.1177/2053951715614980.

Rogers R. (2009) *The End of the Virtual: Digital Methods*, Text prepared for the Inaugural Speech, Chair, New Media & Digital Culture, University of Amsterdam, Amsterdam University Press.

——— (2010) *Internet Research: The Question of Method — A Keynote Address from the YouTube and the 2008 Election Cycle in the United States Conference*, "Journal of Information Technology & Politics", 7, 2–3: 241–260.

——— (2013) *Digital methods*, The MIT Press, Cambridge (MA).

——— (2017) *Doing Web History with the Internet Archive: Screencast Documentaries*, "Internet Histories", 1, 1–2: 160–172.

——— (2018) *Otherwise Engaged: Social Media from Vanity Metrics to Critical Analytics*, "International Journal of Communication", 12: 450–72.

Sergienya I. and Schütze H. (2015) *Learning Better Embeddings for Rare Words Using Distributional Representations*, "Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing": 280–285.

Violi P. (1997) *Significato ed esperienza,* Bompiani, Milan.

Wu L.Y., Fisch A., Chopra S., Adams K., Bordes A., and Weston J. (2018) *Starspace: Embed all the things!*, "Thirty–Second AAAI Conference on Artificial Intelligence (AAAI 18)": 5569–5577.

Zappavigna M. (2011) *Ambient Affiliation: A Linguistic Perspective on Twitter*, "New Media & Society", 13, 5: 788–806.

LEXIA 33-34

Semiotica e Digital Marketing

Semiotics and Digital Marketing

Oggi la semiotica può dialogare con il Digital Marketing in molti ambiti. Fra questi, ci sono gli studi e le attività di Search Engine Optimization (SEO), il Digital Storytelling, il mondo della Web Usability e dell'Interface Design, l'analisi della User Experience (UX), le ricerche sulla viralità online, il videogame design, gli studi sulla gamification, le strategie di new branding nella comunicazione d'impresa, l'analisi qualitativa di grandi masse di dati. Questo numero di Lexia mostra, in modo rigoroso e con linguaggio accessibile anche ai non addetti ai lavori, quali teorie, concetti e metodologie della semiotica possano essere applicati in modo proficuo a ciascuno di questi ambiti.

Contributi di / Contributions by Paula Almiron–Chamadoira, Cinzia Bianchi, Giovanna Cosenza, Giuliana Laurita, Gabriele Marino, Gustavo Niklander Ribera, Paolo Odoardi, Ruggero Ragonese, Mauro Salvador, Leonardo Sanna, Elsa Soro, Mattia Thibault.

In copertina
Foto di Abhishek Saini da www.pexels.com.

ISBN 978-88-255-3542-6

ISSN 1720-5298

euro 35,00