Emanuel MODOC
Sextil Pușcariu Institute of Linguistics and Literary History, Romanian Academy
Cluj-Napoca, Romania
ermodoc@gmail.com

Daiana GÂRDAN
Faculty of Letters, Babeș-Bolyai University
Cluj-Napoca, Romania
daiana.gardan@lett.ubbcluj.ro

STYLE AT THE SCALE OF THE CANON. A STYLOMETRIC ANALYSIS OF 100 ROMANIAN NOVELS PUBLISHED BETWEEN 1920 AND 1940

Recommended citation: Modoc, Emanuel, and Daiana Gârdan. “Style at the Scale of the Canon. A Stylometric Analysis of 100 Romanian Novels Published between 1920 and 1940”. *Metacritic Journal for Comparative Studies and Theory*, 6.2 (2020): <https://doi.org/10.24193/mjcst.2020.10.03>.

Abstract: The present study proposes an experimental exploration of the Romanian novel written between 1920 and 1940 through the use of stylometry, a method of distant reading employed for the statistical analysis of style. Drawing from the most recent advances in the field of computational stylistics, we select a formal standpoint from which we seek to investigate the relation between the Romanian novelistic canon and minor, tertiary novels published in the same period. In our test cases, we will attempt to establish some of the more promising aspects of stylometric analysis, as well as single out the experiments that yield no relevant result. Because of the relative novelty of the method, the purpose of our investigations is to offer a kind of pilot experiment that can illustrate the benefits of using computational methods on Romanian literary corpora.

Keywords: stylometry, computational literary studies, digital humanities, Romanian literature, Romanian novel, canon

The following is an experiment, an exploratory case study on the merits of computational analysis applied to Romanian literature. The reasons behind taking such precautions *ab initio* concern the state of Romanian literary corpora in the digital era (Olaru 30-7; Coroian-Goldiș et al. 1-8; Pojoga et al., Digital Tools 9-16), but also the novelty of the method (in regard to literary texts written in Romanian) that will be showcased above: *stylometry*, a statistical technique that classifies texts (literary and otherwise) on the basis of the frequency of common words. More of a forensic instrument than one which carefully analyses style at the scale of a sentence, stylometry has shown a fair amount of merit in the last few decades. Most often used in authorship attribution studies, it has a surprisingly long history, dating back to the middle of the nineteenth century, when Augustus de Morgan found that the authenticity of St. Paul's writings might be established by measuring the length of the words used in his Epistles (Kenny 1). Wincenty Lutosławski, a Polish philosopher concerned with dating Platonic texts, coined the term "stylometry" in the same period (Lutosławski 61-81). Since then, researchers expanded the methodology for the exploration of increasingly larger corpora, with new features such as average sentence length, type-token ratio or word classes added to the method's arsenal. Currently, stylometry has evolved into various types of statistical stylistics, with machine-learning supervised methods enabling it to now cover a variety of operations from authorship attribution (Juola 233-334) to comparing the styles of authors or establishing stylistic timelines – a method also known as stylochronometry (Holmes 111-17).

In addition to our investigations into the use of stylometry on a literary corpus written in a language not supported by stylometric instruments, we will also make use of network analysis in order to better visualise our initial results. As is the case of almost any distant reading method, investigations employing network analysis are rather scarce in the field of Romanian literary studies (Modoc, Internaționala 201-17; Modoc, Rețeaua 102-106; Gârdan and Modoc 52-65; Gârdan 87-91; Pojoga et al., Character Network 23-47); even fewer use computational approaches. Therefore, the purpose of our investigations is to offer a kind of pilot experiment, illustrating the benefits of using computational methods on Romanian literary corpora.

Methodology

To perform the computational analyses presented in this study, we used the R package *Stylo*, a stylometric instrument developed by Jan Rybicki, Maciej Eder and Mike Kestemont (Eder et al. 107-121). As we have shown above, stylometry is not at all a new discipline pertaining strictly to digital literary studies, but it has evolved in such a way that it facilitates further inquiries into the study of literature. As the developers of the *Stylo* package themselves note,

[i]nstead of the traditional practice of ‘close reading’ in literary analysis, stylometry does not set out from a single direct reading; instead, it attempts to explore large text collections using computational techniques (and often visualization). Thus, stylometry tries to expand the scope of inquiry in the humanities by scaling up research resources to large text collections in order to find relationships and patterns of similarity and difference invisible to the eye of the human reader (Eder et al. 108).

Stylo automatically loads and processes a corpus of electronic text files from a specified location and performs a variety of statistical operations. The core function of *Stylo* is the production of an MFW (most-frequent-word) list taken from the entire corpus. It will then place the frequencies of the MFWs of each individual text in a table of frequencies. For certain languages, it is able to “cull” words (automatic deletion of personal pronouns, for instance) in order to make the analysis more manageable. Finally, it will produce a wordlist used for the actual analysis, and then compare the results for individual texts by performing distance calculations (measuring similarities or differences). By using various measures of distance (such as Burrow’s Delta, Manhattan distance, Euclidean distance, etc.), *Stylo* attributes each text a score in relation to every other text in the corpus, in the form of “weight”, which is in turn used to visualise, in different abstract forms, the relation between texts or authors. *Stylo* can also perform other various statistical procedures, such as cluster analysis, multidimensional scaling, or PCA (Principal Component Analysis). For visual aid, it will also produce graphical representations of distances in the form of a dendrogram or, in other cases, consensus trees.

The standard measure for stylometry, Burrow’s Delta, established by John F. Burrows in 2002, is still one of the most stable metrics for establishing the relative stylistic difference between two or more texts (Burrows 267-287) and it is featured in

the *Stylo* package. Since it relies on what is known in computing as “stop words” (function words, for the more linguistically inclined), this ground-breaking measure can easily be transferred to other languages without risking major inaccuracies. Furthermore, stop words have been established as relevant for computational stylistics for their sheer statistical relevance. As Andrew Piper and Mark Algee-Hewitt note, “stop words are usually semantically poor and yet stylistically rich” (Piper and Algee-Hewitt 158). Moreover, as Burrows himself has shown, and many others have subsequently confirmed through a number of test cases, “Delta has a high probability of correctly indicating the author – that is, if other texts by the author are included in the comparison” (Jannidis and Lauer 32). Although it could be argued that stop words do not make style simply because they are statistically dominant¹, an opposite argument, stemming from psycholinguistics, is that these seemingly innocuous words – pronouns, articles, prepositions, auxiliary verbs, conjunctions – do contribute to individual style (see Pennebaker).

As a method of distant reading, stylometry, in all its various forms, has two main features: the conversion of textual information (in electronic form) into numbers and the processing of those numbers using statistics. In this respect, the *Stylo* package can perform both tasks; however, given that, from a certain point onward, a literary corpus may reach large numbers of texts, there are other instruments that can help with data visualisation. Such an instrument is *Gephi* (Bastian et al. 361-2), a software able to render, through its statistical algorithms, the stylometric networks that will be showcased below.

A few observations regarding our option for network visualisation are in order. First, *Gephi* is a software that creates descriptive, but ultimately arbitrary, networks. As Dennis Tenen points out: “[d]espite the apparent quantification, network analysis is more of an art than a science” (Tenen 260). The numbers fed into *Gephi*’s statistical algorithms are taken from *Stylo*, which strives for exhaustiveness in the sense that it requires large corpora of literary texts in order to yield the best results,

¹ Here Nan Z. Da argues, in a seemingly “computational case against computational literary studies”, that “[t]he hitch of using textual pattern mining for forensic stylometry is that even if you apply pattern recognition techniques that reduce noise and nonlinear interactions between data, the stylistic differences that can be captured for literature tend to be driven by stop words – if, but, and, the, of. Why is that the case? [...] In reality, stylistic differences boiling down to stop words is not surprising at all. To locate a statistical difference of occurrence means having enough things to compare in the first place. If the word cake only occurs once in one text and four times in another, there’s no way to really compare them, statistically. By the numbers, stop words are the words that texts have most in common with one another, which is why their differentiated patterns of use will yield the readiest statistical differences and why they have to be removed for text mining” (Da 622-23).

so the weights established in analysing one hundred novels may vary when adding another fifty to the corpus. We then coax Gephi into rendering the results provided by *Stylo* into visual representations of a stylistic network that is able to exhibit, at the very least, the relation between a set number of factors, such as authors, texts, genre, or narrative perspective. This network is in turn generated by using different layouts, which are governed by various mathematical principles. Through Gephi's interfaces, we can fine-tune the properties of each layout in order to create a network that best suits our needs. For the purpose of our experiment, we used Force Atlas 2, a physics-based layout that treats nodes as particles that repel each other. It is the preferred layout of many authors who attempt to render literary networks as a "field", thus extending Bourdieu's metaphor to expose its implicit mechanism. However, we have to be very careful when interpreting these networks, because, as Tenen puts it,

[t]he visualization belies a mathematical model, a metaphor, which should not be confused for the thing it is meant to represent. The graphic is subject to the usual pitfalls of interpretation: it is at once overdetermined and necessarily reductive; it leads to an excess of signification (Tenen 262).

Because Gephi's topography is non-spatial (there is no "up" or "down", "left" or "right"), and thus subject to positional randomness, the same data processed with the same layout can generate different positions within the network. This is why, in order to produce the least biased results, we have produced a single network for all our tests.

In many ways, our study is an exploration of the possibilities that *Stylo* offers to those who work with collections of Romanian texts. Being among the first attempts at using this analytical tool on a language it does not yet support, our investigation should be taken with a grain of salt. If our experiment does not provide any dramatic insights into computational analyses of style, it will, at the very least, grant a preliminary evaluation of the method.

Corpus preparation

For reasons pertaining to a host of factors, from the novelty of the method in the field of Romanian literary studies to *Stylo*'s lack of language support for Romanian, choosing a literary corpus that would yield both promising *and* verifiable results

(especially through alternative or traditional means) was one of the main challenges of our study. We had to tackle not only methodological limitations, but also our working categories of what we do analyse. We started out with the preliminary notions set out by Stanford Literary Lab in their eleventh pamphlet (Algee-Hewitt et al. 2): the published, the archive, the corpus. We had a comprehensive list of *the published*, thanks to the *Chronological Dictionary of the Romanian Novel from its Origins to 1989* (Istrate et al.); we had access to *the archive* courtesy of the many national libraries that preserved the novels; we also had a part of the corpus provided by the project ASTRA Data Mining. The Digital Museum of the Romanian Novel 1900-1932 (Baghiu et al.) formalized and ready for analysis. However, we were missing a governing principle for our inquiry. What do we want to compare, and how can we be sure that our investigation will be fruitful? We needed an anchor point, something to ensure that our results are verifiable at least when confronted with our original assumptions: *the canon*. More specifically, *canons*, because we had to take into consideration not only what our critical tradition deems canonical (canonization as a result of choices made by literary critics, historians, publishers, or the public), but also what a canon is from a distant reading perspective. Thus, our canonical selection includes both a representation of the Romanian critical canon and a historically relevant sample of the Romanian field of production.

Scaled down to our own – relatively small – literature, our canonical corpus consists of 50 novels published between 1920 and 1940. It is the core canonical period of our modern novel. We added another 50 novels. A small part of *the corpus*, an even smaller part of *the archive*, but altogether necessary for the “background noise” that would serve our stylistic network. Furthermore, and with very few exceptions to this self-imposed rule, we selected minor novels following the dominant novelistic genres dictated by the canon. Therefore, we selected 17 subgenres spanning these 100 novels: adventure, biographies (and autobiographies), *Bildungsroman*, sensation, *hajduk*, science fiction, erotic, sentimental, family novels, *livres à clef*, parables, historic, psychological, rural, social, war. In order to avoid further expansion of our corpus (this is, after all, merely an experiment), we did not take into account canon-to-corpus ratios in selecting the *number* of novels belonging to each subgenre. After appropriately tagging the corpus with genre metadata, we added an additional tag, this time relating to narrative perspective. Not wanting to muddy the waters with narration *styles* (such as free indirect speech), since it would

require a great deal of close reading, we limited ourselves to only tagging the point of view of the narrator (first, second or third person POV).

It should also be made clear that our corpus is not exactly balanced, nor is it representative of the *entire* literary production of the period it covered. Firstly, it represents a mere 13% of the novelistic production, a far cry from an exhaustive literary corpus. Secondly, it strives to achieve results that can be reached with similar research methods, be them traditional (close readings) or digital/quantitative. A similar endeavour could be attempted using, for instance, topic modelling, an altogether different type of computational analysis that requires a great deal more corpus preparation and management. Lastly, the corpus is merely a test run, an exercise in providing a firmer foundation for later uses. If, however, we can confirm or refute traditional knowledge based on previously established hermeneutical inquiries, then we can start developing further investigations that challenge clichés or prejudices related to the Romanian novelistic canon, as set by the critical establishment.

As a final observation, it is worth noting that, since our preferred metric for computational stylistics is based on Classic Delta, and since this measure relies on z-scores – i.e. normalized word frequencies – and thus dependent on the number of texts analysed and on a balance between these texts, we have reduced the number of maximum novels per author to five. The only exception here is Liviu Rebreanu, who features with eight novels in the corpus. This is because Rebreanu is so stylistically cohesive (from a computational standpoint) as opposed to Hortensia Papadat-Bengescu or Mihail Sadoveanu, that having him in his own separate community stabilizes the network. As we will see below, far from having to do with “stylistic fingerprints” or other forensic categories, the clustered communities in our network are indicative only of the relation between a text and other texts in the context of a given corpus. In this respect, authorial footprints have little to do with the way in which our computational analysis expresses the network.

Application

With these general parameters in mind, let us commence our experiment. We will look at the selected one hundred Romanian novels published between 1920 and 1940 through two different types of visualisation: cluster analysis consensus trees (Fig. 1) and cluster analysis networks (Fig. 3, 4, 5). While the networks are based on

preliminary dendrograms generated by *Stylo*, they are so densely packed (see Fig. 2) that it is impossible to discern any meaningful results. The dendrogram (and, by extension, the network) contains information gathered during a single run of the Stylo program (set at 500 MFWs), while the consensus tree is produced by running four consecutive runs (starting at 200, 300, 400, and 500 MFWs). When looking at the consensus tree shown below, one point is rather obvious: without any meaningfully separate branches, one cannot point toward anything relevant about the stylistic similarities or differences between the analysed texts. However, upon closer inspection, we can see that the Delta measurement is quite good at indicating authorship in the corpus. Conversely, dendrograms show the similarities between two entities quite well. Liviu Rebreanu's novels are clustered very closely, so are Mihail Sadoveanu's or Anton Holban's. The x-axis of the dendrogram featuring numerical values represents the degree of affinity between two or more texts. The closer the novel to the value 0, the more similar it is to its neighbours. Usually, novels with values between 0 and 0.5 are considered to be authored by the same person. Camil Petrescu, Anton Holban, Hortensia Papadat-Bengescu, E. Lovinescu, and of course Liviu Rebreanu and Mihail Sadoveanu feature in this category. A notable exception is G. Călinescu, who clusters, with two of his novels, in different positions: *Enigma Otiliei* clusters closely with E. Lovinescu's *Mite*, *Bălăuca*, and *Lulu*, while *Cartea nunții* features alongside novels written by Ionel Teodoreanu and Gib Mihăescu.

Fig. 1. Romanian novels published between 1920 and 1940. Consensus tree, Classic Delta, 200-500 MFW, consensus 0.5

Fig. 2. Romanian novels published between 1920 and 1940. Dendrogram, Classic Delta, 500 MFW

The more promising visualisations, however, are the ones generated by *Gephi*. As can be seen in **Fig. 3**, the stylistic *network* of the Romanian novel published between 1920 and 1940 is much more accessible for further interpretations. In order to cluster the authors and texts, we applied a modularity statistic, which assesses the number of distinct groupings within a network and parses the nodes of the network into distinct groups. The purpose of a modularity statistic is to group nodes based on the strength of their relationships, thus creating „communities” (Cherven 189) that facilitate the visualisation of the relations between groups of authors and their respective novels. In **Fig. 3** we have also ranked the authors by their modularity class, so we can observe some fairly interesting communities shaping up. In this regard, we can easily identify some strong authorial communities, as well as “weaker” ones that group together in a less cohesive manner. Of the stronger communities, Mihail Sadoveanu, Liviu Rebreanu, Hortensia Papadat-Bengescu and Anton Holban are the most individualized authors in the network, while Camil Petrescu, Ion Minulescu, Mihail Drumeș, Mihail Sebastian, Mircea Eliade, and Victor Eftimiu, most of which are representative for the Romanian novelistic canon, create a multi-author stylistic community. The same is true for authors such as Garabet Ibrăileanu, Gib Mihăescu, Cezar Petrescu and Mateiu Caragiale. However, what we call the “white noise” of the network has its own coherence, in the sense that the canonical novels are isolated enough for them to make a distinct community, separate from the more critically acclaimed authors. In this respect, prodigious tertiary authors such as N.R. Niger, V. Demetrius, Damian Stănoiu, Felix Aderca, and Ion Agârbiceanu seem to cover large swathes of network real estate.

While it is largely accepted that genre is an important aspect of style and can be a fairly good classifying feature, novelistic subgenres do not determine authorial style. As shown above (Fig. 4), authors that dabble in different subgenres and are prolific enough will showcase some interesting results. Mihail Sadoveanu is featured in our network with three historical novels, a rural novel, and a parabola. Stylistically, the historical novels do group together in the small Sadoveanu community. Liviu Rebreanu, on the other hand, is more cronologically cohesive, with earlier psychological novels such as *Pădurea spânzuraților* and *Ciuleandra* slightly separate from his later novel, *Amândoi*, published in 1940. Somewhat similarly, Hortensia Papadat-Bengescu, who in the same series (the *Hallipa Saga*) experiments with different subgenres (psychological, social, or family). Judging by this subgenre

distribution, it is worth noting that, as a classifying feature, the novelistic subgenre could yield interesting results when analysing singular authors with a large (and diverse) enough literary corpus. In larger scale analyses, however, the results are not as reliable due to the very stable authorship attribution that Stylo generates.

Fig. 3. Romanian novels published between 1920 and 1940. Network analysis.
Force Atlas 2. Modularity Class ranking

Fig. 4. Romanian novels published between 1920 and 1940. Network analysis.
Genre distribution

Another aspect we wanted to point out in our analysis concerns narrative perspective, mainly because the Romanian novel undergoes, in this period of concurrent modernization and consolidation, significant changes in terms of authorial perspective. As shown in **Fig. 5** below, novels written from a first-person point of view are statistically dominant after 1930, and that is no accident, seeing that the same period is regarded as the peak of Romanian novelistic history – both from a quantitative and a qualitative standpoint (Terian 63) –, with 1933 being dubbed “the golden year” (Cornea 380-2). It is small wonder, then, that the most flourishing period of the Romanian novel features most of the subjective novels of the period. Other first-person novels such as Max Blecher’s *Întâmplări în realitatea imediată* [Adventures in Immediate Reality], Felix Aderca’s *Omul descompus* [The Decomposed Man] or Virgiliu Serdaru’s *Apocalipsul* [The Apocalypse] are more or less entropic occurrences in the network.

Fig. 5. Romanian novels published between 1920 and 1940. Network analysis. Narrative perspective distribution (blue for third person point of view, red for first person point of view)

Our observations do not exclude, however, the existence of psychological novels written in the third person, which employ narrative styles such as indirect free speech to a great extent, but their positions within the network do show that, in terms of computational stylistics, this narrative style cannot be differentiated through statistical distance measures. **Fig. 5** also highlights the positions of these novels within the stylistic network. To a certain degree, the network is historically coherent. Apart from the highly individualized communities of Rebreanu, Sadoveanu, and Blecher, it can be easily observed that the network also displays, to some extent, a chronology of the Romanian novel.

Conclusions

In many ways, our experiment challenged us to make inferences beyond our original assumptions. In some cases, authors marked as stylistically special by traditional literary historiography do end up in an isolated position within the network, while, in other cases (Camil Petrescu, Mircea Eliade or G. Călinescu), they are a great deal more closely linked to the rest of the authors. We believe that these results have the potential to reopen some debates concerning personal style using formal markers such as stop-words. Nevertheless, stylometry is only one of many methods of distant reading. It does show some promise for research in Romanian literary history, given how well it discerns authors and narrative points of view, notwithstanding its lower accuracy in the case of novelistic subgenres. It opens up new avenues for the inclusion of minor authors in the analysis of literary style, as well as challenging pre-existent hypotheses regarding the place of literary canon within a broader spectrum. In our test cases, stylometry proved very capable of attributing authorship, while also helping contextualize the most well individualized authors in a network. Last, but not least, it enabled a type of flexible reading by establishing differences and similarities between the styles of authors (such as Mihail Sadoveanu and Liviu Rebreanu) and those of their immediate contemporaries.

Acknowledgement: Emanuel Modoc's work was supported by a grant given by Babeş-Bolyai University, GTC-UBB, "Rețeaua romanului românesc 1920–1940. O abordare computațională". Project number GTC 31372/2020.

References

- Baghiu, Ștefan et al. *Muzeul Digital al Romanului Românesc: 1901-1932* [*The Digital Museum of the Romanian Novel: 1901-1932*]. Complexul Național Muzeal ASTRA, 2020: <https://revistatransilvania.ro/mdrr1901-1932>.
- Bastian, Mathieu et al. "Gephi: An open source software for exploring and manipulating networks". *Proceedings of the Third International ICWSM Conference*, 2009: 361-362.
- Burrows, John F. "Delta: A Measure for Stylistic Difference and a Guide to Likely Authorship". *Literary & Linguistic Computing*, vol. 17, no. 3, 2002: 267-87.
- Cherven, Ken. *Mastering Gephi Network Visualization*. Packt Publishing, 2015.
- Cornea, Paul. *Aproapele și departele*. Cartea Românească, 1990.

- Coroian-Goldiș, Andreea et al. "The Archives of the Romanian Novel and Digitization Possibilities". *Transilvania*, nr. 9, 2019: 1-8.
- Da, Nan Z. "The Computational Case against Computational Literary Studies". *Critical Inquiry* 45.1, 2019: 601-639.
- Eder, Maciej et al. "Stylometry with R: A package for computational text analysis", *R Journal*, 8(1), 2016, journal.r-project.org/archive/2016/RJ-2016-007/index.html): 107-121.
- Gârdan, Daiana, and Emanuel Modoc. "Mapping Literature through Quantitative Instruments. The Case of Current Romanian Literary Studies." *Interlitteraria*, 25.1, 2020: 52-65. doi.org/10.12697/IL.2020.25.1.6.
- Gârdan, Daiana. "Privind de departe modernitatea: Cazul 'Sburătorul'". *Vatra*, no. 8-9, 2020: 87-91.
- Holmes, David L. "The Evolution of Stylometry in Humanities Scholarship," *Literary & Linguistic Computing* 13, no. 3 (September 1998): 111-117.
- Istrate, Ion et al. *Dicționarul cronologic al romanului românesc de la origini până în 1989* [Chronological Dictionary of the Romanian Novel from its Origins to 1989]. Editura Academiei Române, 2004.
- Jannidis, Fotis, and Gerhard Lauer. "Burrow's Delta and Its Use in German Literary History". *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. Edited by Matt Erlin and Lynne Tatlock. Boydell & Brewer, 2014: 29-54.
- Juola, Patrick. "Authorship Attribution". *Foundations and Trends in Information Retrieval*, vol. 1, no. 3, 2006: 233-334.
- Lutosławski, Wincenty. "Principes de stylométrie". *Revue des études grecques*, no. 41, 1890: 61-81.
- Modoc, Emanuel. "Rețeaua conceptuală a modernității românești. O abordare computațională". *Vatra*, no. 8-9, 2020: 102-6.
- Modoc, Emanuel. *Internaționala periferiilor. Rețeaua avangardelor din Europa Centrală și de Est*. Muzeul Literaturii Române, 2020.
- Olaru, Ovio. "What is Digital Humanities and What's It Doing in Romanian Departments?". *Transilvania*, no. 5-6, 2019: 30-7.
- Pennebaker, James. *The Secret Life of Pronouns: What Our Words Say about Us*. Bloomsbury, 2011.

- Piper, Andrew, and Mark Algee-Hewitt. "The Werther Effect I: Goethe, Objecthood, and the Handling of Knowledge". *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*. Edited by Matt Erlin and Lynne Tatlock. Boydell & Brewer, 2014: 155-184.
- Pojoga, Vlad et al. "Digital Tools for the Analysis of the Romanian Novel". *Transilvania*, no. 10, 2019: 9-16.
- Pojoga, Vlad et al. "The Character Network in Liviu Rebreanu's Ion: A Quantitative Analysis of Dialogue". *Metacritic Journal for Comparative Studies and Theory*, 6.2, 2020: 23-47: <https://doi.org/10.24193/mjcst.2020.10.02>.
- Terian, Andrei. "Big Numbers. A Quantitative Analysis of the Development of the Novel in Romania". *Transylvanian Review*, XXVIII, supplement no. 1, 2019: 55-71.