

Dr Maciej Maryl

Institute of Literary Research of the Polish Academy of Sciences

maciej.maryl@ibl.waw.pl

**Operationalising the Change:
The dispersion of Polish literary life (1989–2002)**

1. Introduction

Let us picture the history of literature as an unknown city one is trying to explore. An individual may contemplate some details and visit places which have been recommended by others. However, one needs to climb higher to perceive the regularities of the landmarks and the connections between them, to see the hidden patterns of the city's life. While climbing higher and higher we lose many of the details and subjective impressions of particular buildings, yet, we get a fuller picture and discover new relationships which are invisible from ground level. However, you can never climb high enough to get the full picture – or can you? This is precisely what this project aims to explore – the possibilities and limitations of a bird's eye view of events, works, and processes during an important period in Polish literary history, and to compare it with intersubjective claims made by various critics and scholars.

Why is the intersubjective view not enough? All syntheses in literary history, as well as in other branches of humanistic inquiry, are, by necessity, based on certain approximations due to the obvious cognitive limitations of mortal, selective, and forgetful human beings. How about expanding the capacities of individual researchers by providing them with the opportunity to read all the books at once, to learn all the facts (including the minor and seemingly unimportant ones), and combine everything into a meaningful theory. If we consider digital research tools as being not only more effective typewriters but – in McLuhanian terms – extensions of our senses and capacities, we could treat such a 'view from above' as a great opportunity to generate new insights which could support our interpretative practices. Moreover, apart from cognitive limitations, we may go beyond the set of works considered to be canonical and also explore those objects which went unnoticed by contemporaries or the next generations of critics. According to the laws of 'historical selection' proposed by Robert Escarpit, 80% of literary production is forgotten within a year, and 99% within twenty years (1966, 34). By looking 'from above,' we may now account for those forgotten works and try to present a more accurate picture of literary processes during a given period.

The digital revolution in humanities brought about not only increased storage capacity and accessibility to literary resources, but more importantly the possibility of reusing these resources for large-scale quantitative research. In doing so, we may take advantage of existing resources that have been diligently collected and annotated by various institutions over the years with the intention of facilitating the research process: bibliographies, catalogues, chronicles, calendars, fact sheets, reports, etc. There is a growing body of new, computational methods and approaches to the study of literature which have marked a distinction between traditional, close-reading and detail-oriented approaches on the one hand, and new, synthetic studies that apply statistical methods to larger corpora, on the other. In order to highlight the qualitative and empirical angle of this approach, it has been dubbed ‘distant reading’ (Moretti 2009), ‘algorithmic criticism’ (Ramsay 2011), or ‘macroanalysis’ (Jockers 2013). The main premise behind this research corresponds to the metaphor I used in the introduction: when looking ‘from above’ we may lose some granularity concerning individual works, but instead, we can survey larger-scale processes, sequences, and cycles. Among these digital ‘bird’s-eye view’ methods, we may distinguish between the scholarship that analyses the features of the actual texts and the metadata-driven research.

The former approach concentrates on quantifying textual features. For instance, stylometric research focuses on the most frequently used words in a given corpus of texts (Eder, Rybicki, Kestemont 2011). Such research may entail authorship attribution (e.g. Eder 2015), genre analysis (e.g. Schöch and Pielström 2014), or large-scale analyses of national literary output (e.g. Jannidis & Lauer 2014 for German literature, or Rybicki 2014 for Polish novels). The stylometric analysis is also combined with functional or grammatical linguistic categories for the study of prose (Allison et al. 2011) or other literary genres (Hoover, Culpeper, O’Halloran 2014).

The second strand of research employs metadata, that is, data about texts such as title, genre, author, publication date, etc. Although such research often draws on book history, current scholarship proves its validity in literary studies. Franco Moretti visualised and interpreted the genre life-cycle using the example of metadata from the nineteenth-century British novels (Moretti 2005, 15–19). Matthew Jockers used bibliographical data to explore the history, topics, and geographical distribution of Irish American fiction (2013, 35–62). While Katherine Bode (2014) used various bibliographical datasets to ‘recalibrate’ the history of Australian literature and challenge some established arguments, including views on the relationships between local and foreign publishing markets and gender distribution among authors and their coverage in press and scholarship. Through a visualisation of data about poet–journal affiliation, Long and

So (2013) pictured the differences between literary networks in the U.S., Japan, and China. It should also be noted that the division between the linguistic and the metadata approaches is purely conceptual and they could be incorporated into a single research design, as in the case of Moretti's 'Reflections on 7,000 titles' (in Moretti 2013).

This paper explores the possibility of quantitative research into the transitions of literary history using the example of Polish literary life 1989–2002. This period is by no means under-researched; on the contrary, the abundance of scholarly and critical writing makes it even more interesting from the data-driven research perspective, since it allows the critical claims against existing data to be validated.

This exploration will be carried out on the basis of data obtained from the Polish Literary Bibliography (PBL),¹ a comprehensive database of Polish literary and cultural life. It indexes not only literary books but all instances of literary life and reception: reviews, journal articles, newspaper mentions, dramas, adaptations of literary works (stage, film, radio, and television), and literary prizes. So, not only do we have access to rich data about literature but also to the relationships between the objects (e.g. A is a review of B; C is an adaptation of D). The content is also annotated according to its subject. This resource has been in development since 1954 and already covers Polish literary culture from 1944 to 2002. Unfortunately, most of the data are still available only in print, with the approximately 500,000 digital records only covering the years 1988 to 2003. An ongoing development project (Maryl & Weislik 2016) aims to convert the PBL's Oracle database into a new, more accessible format with some tools for researchers. The project is also engaged in the retrodigitisation of earlier volumes, but given the vast scope of the task and difficulties with parsing the complicated PBL records, it is hard to foresee when a full database will be available.

Against this backdrop, the following discussion focuses on demonstrating how the PBL's digital 'view from above' can be used to test selected qualitative hypotheses concerning Polish literature during the transition period (1989–2002), as a prelude to a general methodology for the data-driven study of literary history using bibliographical datasets. The discussion will be supplemented with a reflection on how documentation methodologies and practices affect the quantitative analysis.

¹ <https://pbl.ibl.waw.pl/>

2. The literary transition as a challenge to literary theory

As Anna Nasiłowska (1997) observes, the periodisation of twentieth-century Polish literature is almost never based solely on artistic criteria, and has been heavily influenced by external events – both political and economic. This is one of the reasons why she chooses 1976 as the starting point for her book on contemporary Polish literary history – the year of the massive workers’ strikes which were violently broken by the authorities, events that served as a decisive impulse for the establishment of the first organized oppositional bodies and, in turn, underground publishing. Nasiłowska uses the term ‘divided literature’ to describe Polish literature between 1976 and 1989, highlighting the remarkable dualism of literary life after the birth of *drugi obieg*, i.e. the *second circulation*, which was independent of both the authorities and censorship. On the one hand, there was the official circulation comprising state-controlled publishing and dissemination institutions, while on the other, a variety of *samizdat* books and journals illegally printed and distributed by various dissident bodies. The year 1976 marks the beginning of organised underground printing in Poland, but Polish literature was already being published by émigré circles from the end of World War II, mostly in France and the UK but also in smaller centres in Germany, USA, Canada, and Sweden. These endeavours targeted local Polish communities, but their output was also smuggled into Poland and, after the establishment of illegal presses, reprinted (and *vice versa* – Polish *samizdat* were often republished abroad). Moreover, there was also the *third circulation* (*trzeci obieg*), that is, counterculture publications, which were also published illegally but which remained critical of the political affinities and goals of the underground presses.

Defining the literary transition poses a challenge for literary theory because of the sheer fuzziness of the concept. First of all, it is difficult to set up a precise date for the political breakthrough – was it 4 June 1989, the day of the first partially free parliamentary election; or 24 August, when the first non-communist government was appointed; or maybe even more pertinent from the cultural point of view, 11 April 1990, the day when censorship was officially abolished? As Przemysław Czapliński notes, ‘Transition, whether we want it or not, lasts. It is not a rupture but a movement past a critical point [*przesilenie*], not an event but a process’² (2002, 6). For this reason literary critics discuss their observations of literary change in a broader timeframe, which makes those elusive processes visible: Nasiłowska focuses on a twenty year period (1976–1996), which she calls a *transitory period* (*okres przejściowy*) (1997); Czapliński writes about the *traces of the end* in the literature of the 1980s and *traces of the*

² All translations from the Polish sources are by the author.

beginning in the 1990s (1997); and Janusz Sławiński writes about the *disappearance of the centre (zanik centrali)* in 1994, retrospectively naming the processes which gradually took place during the decade 1983–1993. There always has to be a point of reference. Although this problem receives proper attention a bit later in this section, it should be noted here that the demarcation of the transition period's end point poses similar challenges.

Secondly, in thinking about the literary transition of 1989, there is an enthymematic assumption that political events bring about cultural change. Many literary critics, as well as audiences, anticipated a transformed, brand new literature (cf. Czapliński & Śliwiński 1999, 209–213; Czapliński 2002, 11–24). These expectations are neatly captured as 'Appetite for Change' in the title of Jerzy Jarzębski's book (1997), as well as in the book itself. Czapliński and Śliwiński note that the need for a literary breakthrough was intensified by the ongoing debates about the transformation of society, and they conclude that,

'...the question, whether a political change [przełom] triggers a literary one, is and is not unreasonable at the same time. It is unreasonable, because – luckily – literature is not an addition to political life. It is not, because literature is a part of reality which underwent an enormous change.' (1999, 213)

So, perhaps, political change should be perceived as a sort of catalyst for cultural transition, because 'it was not only a beginning of serious and extensive transformations of the infrastructure of the literary life but it also ... made the novelty legitimate, because it built up the expectation of novelty.' (Czapliński 1999, 113)

In his book, which is exclusively dedicated to the transition, Czapliński notes that a literary breakthrough takes place in the three spheres of poetics, ideas, and institutions (1999, 5). It affects poetics when certain genres and conventions lose their function as carriers of significant content, and new, different forms replace the accepted ones (*ibid.*). The field of ideas changes thanks to new modes of using literature (*ibid.*) Finally, the institutions are transformed when the means of literary production and distribution are changed, together with the relationships between the partners of literary communication: writers, patrons, publishers, and educators (*ibid.*). Given the vagueness of the plane of ideas (and, to some extent, the poetics), the change is mostly visible in the institutional field, Czapliński concludes (*ibid.*, 5). In this paper, I try to corroborate the critical discourse concerning this period, and map the evolution of literary institutions from the centralised model to the commercialised literary market of late capitalism.

'Mapping' is not a random word here since critics have often used spatial metaphors to describe the processes of transition. The concept of the *disappearing centre* was coined by

Sławiński, an eminent Polish structuralist, in a tiny (3-pages long) and highly influential essay (1994). Although he focuses solely on poetry, his remarks were applied, with much success, to literary life in the broadest sense. Sławiński's point of departure is a vision of literature in which individual works are parts of the 'greater Whole' (*większa Całość*); in which authoritative (*miarodajni*) poets (i.e. those 'read by everyone, whether they like them or not') serve as orientation points for the assessment of other works (1994, 15). In other words, they constitute a canon. Sławiński observes the collapse of this system and concludes:

'This Whole, which had supplied common frames of literary communication and had organised it around a certain centre, has been replaced by a polycentric system: multiplicity of – social and spiritual – microworlds, in which poets and readers find each other. In each of them, there are local tastes, local evaluation measures, local circulations, local relationships, authorities and hierarchies, local lucidities and unlucidities. One writes mostly for fellows ... and expects new poems – from fellows.' (ibid.16)

This spatial metaphor is eventually elaborated on by Bałajewski (1995), who claims that literature is in a *state of dispersion* (*stan rozproszenia*) by demonstrating how top-down mechanisms of literary prestige are being substituted by bottom-up canon formation: locally acclaimed writers who write about their local experience gain national recognition. 'Locality,' in his understanding, goes beyond geography, and covers like-minded, small initiatives in different parts of the country (1995, 19). Moreover, as Bałajewski observes elsewhere, the attention of literary critics of the first half of the decade is focused on the novelty and thus dubbed by him *debutocentrism* (*debiutocentrism*) (2012, X) What is important in this diagnosis is the feeling of temporariness: '[the] Situation is still quite fluid. New initiatives appear all the time, transforming a – seemingly – ready cultural-institutional-literary system. We are in the state of dispersion' (1995, 17). This flexible relationship between the centre and peripheries is later elaborated on by Czapliński (2002), who describes this process through the metaphor of *moveable margins* (*ruchome marginesy*). It is a vision of literary culture in which different phenomena coexist and either move to the centre or are pushed to the margins, depending on their importance for the greater community.

These transformations are, if not welcome, at least accepted by the critics, who – perhaps sharing the *appetite for change* – embrace this new configuration of cultural production, acknowledging its potential. Yet quite quickly this optimism yields to disappointment, mostly because the decentralisation processes meet the market economy. Underfunded local initiatives find themselves on the margins, whereas popular and commercial literature occupies the centre.

‘Thanks to the domination of the market ... the centre has become established, and the margins fixed’ (Czapliński 2002, 7).

Somewhere in the second half of the 1990s this trend is reversed, marking the beginning of a slow normalisation; understood as a process in which literature is marginalised by the mass media, which in turn is treated as the natural order of things by the public (Dunin 2006). Critics start challenging the dominant discourse, which establishes new, commercially-generated hierarchies. They call for literature which challenges the reader and confronts them with otherness (Czapliński 2002, 21–36). These processes coincide with the ‘return of the old masters’ (Bağlajewski 2012), that is, a revaluation of the great writers who are again featured in the mainstream and receive literary prizes. In the end, critics claim, we are witnessing the *return of the centre (powrót centrali)* (Czapliński 2007), but the centre has changed. It is not the official, state-controlled system, but the dominant media discourse (Dunin 2006) and free market processes. Hence, the state of dispersion is considered to be a temporary stage between two centralised systems (Bağlajewski 2012). As Czapliński claims (2007), the only difference between these two is that the former was politically-driven, while the latter is established on commercial principles.

Let us translate these claims into a set of hypotheses which are verifiable using the available data. The main hypothesis that I am going to test in this paper is that, according to previous scholarship, we witness a post-1989 dispersion of the centralised system of literary life, and then, around the mid-1990s, its subsequent re-centralisation according to market principles. Given the access we have to the data at hand, described in more detail in the next section, I will aim to corroborate four hypotheses (or sub-hypotheses) which have been derived from the main one. These are as follows:

1. The institutions of literary life undergo dispersion in the early 1990s, followed by the subsequent return of a more centralised market.
2. Literary reception follows the same pattern.
3. Literary reception in the ‘dispersed’ phase is *debutocentric* (i.e. focused on debutants), which changes in the re-centralisation phase with more attention being given to the ‘old masters,’ that is, eminent writers who earned their position before 1989.
4. Processes of dispersion and re-centralisation also have a geographical edge, which should be visible as the rise, and eventual decline, of smaller publishing centres.

3. Operationalisation

In order to validate these critical claims, one needs to first translate the elusive metaphors into variables or indicators. As Franco Moretti puts it in his pamphlet on character-space: ‘Operationalizing means building a bridge from concepts to measurement and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts’ (Moretti 2013b, 3). But in our case: from the concepts of literary history, through some form of quantification, to bibliographical data.

This task looks like an equation with two unknowns: on the one hand there are metaphoric claims, on the other bibliographical material prepared by humans, full of possible errors and biases. However, in a spirit most recently advocated by Kathrine Bode, we should use ‘bibliographical consciousness’ (Bode 2017, 94), and recognise the shortcomings of both the data and the practices of collecting; or rather, as she puts it, constructing literary data: ‘grounding data-rich literary history in scholarly editions of literary systems emphasizes that constructing literary data is just as much an interpretive and critical activity as its analysis’ (ibid. 101). In other words, only by acknowledging the construction of our data – its shortcomings and advantages – are we able to account for the phenomena which are indeed graspable by this dataset (and such an account is given later in this section).

However, how should we approach the other unknown, namely, the claims made by literary critics of the transition period? Firstly, there is the question as to whether these generalisations are of a descriptive or rather prescriptive nature, as critics are, *volens nolens*, actors of the literary life they investigate.

Unlike critics, who form their judgments on the basis of vast cultural material (although accessed unevenly and unsystematically), empirical approaches are faced with the scarcity and fragmentation of the data at hand. Unlike scholars in the life sciences, who collect data relevant to their hypotheses, we need to tune our research scope to the available material. Hence operationalisation becomes a tricky activity which entails ‘building a bridge from concepts to measurement’ with only that many bricks we can gather.

The most important thing to be noted about the PBL data is that it is a construct guided by certain methodological principles which have evolved over time. Luckily, the period I am working with does not witness any major methodological changes, however, there are important reservations to be taken into account. First of all, although comprehensive and detailed, the PBL database does not contain every piece of literature published in Poland. For instance, it does not account for some regional literary journals. It may also favour certain eminent writers by collecting all of their output, even that of a clearly non-literary nature. Second, literary documentation (the bibliography, in our case), unlike literary scholarship, has to work within

the sharp definition of literature. Hermeneuts can dedicate entire volumes to the discussion of literariness, whereas literary documentation has to operate within well-defined borders. The consequences are grave – if something is considered literary, it belongs in the database (cf. Umerle 2017). Third, only certain data fields (creator, reviewer, publisher, journal, particular work) have controlled vocabularies, while some entities have duplicate entries. I have tried to minimise this influence during the data-cleaning process, however, there may be human mistakes I have overlooked. Moreover, an open ‘comment’ field for each data entry contains lots of unstructured information (e.g. if a critical work discusses two authors, the one less featured will be mentioned only in annotations, thus not available for quantitative analyses). Ongoing modernisation of the PBL will minimise these problems in the future.

With all these reservations in mind, we may treat the PBL database as a broad, representative sample of Polish literary life, which is carefully annotated and ready for analysis. It is safe to assume that although the absolute results obtained from this data may not be comparable with other resources (e.g. the actual number of books published, or the creators active in a given year), yet, thanks to consistent methodology, it allows for relative comparisons to be made over time, that is, for looking at certain trends in literary production.

The initial dataset was downloaded from the PBL database. The database is relational, so the materials which were requested consisted of all published books and articles, including secondary literature. Although the PBL covers Polish literature from 1944 onwards, the online database only contains the data for 1988–2003, with some earlier works if they were referenced in the modern material. At the time of the study, the year 2003 was still not finished, and the year 1988 was not complete, as it served as a test year for the first online database’s setup in 2000.³ Given these constraints, I have limited the material to the period 1989–2002. Furthermore, as the collection methodology was changed after the data for the year 1989 was entered into the database, most notably through excluding some regional journals; I have decided to exclude from the sample all journals which appear only in 1989.

Limiting the transformation period to 1989–2002 is an arbitrary move, dictated merely by the availability of the research data for comparative analysis. However, based on the research claims discussed above, one should be able to trace the processes of dispersion and subsequent re-centralisation during this period. The ideal scenario would require a dataset spanning approximately five more years in either directions. Hopefully, future data will allow for a re-examination of the findings presented below.

³ Certain efforts are being made towards digitising printed volumes and hence expanding the range of the online bibliography.

There are three key datasets I use in this study:

- a) **Books.** Literary books either published in Poland (incl. translations of foreign literature), or by Polish authors abroad. Each entry includes information about the author, publisher, place, and year of publication. In some cases, book genre was available or was able to be inferred from metadata or the title (e.g. play, children's literature, music, nonfiction, short stories, novel, reportage, art, poem, memoir).
- b) **Articles.** Literary articles published in Polish literary journals; including poems, fiction, letters, essays, nonfiction, and adaptations. The metadata covers information about the author, journal, year, and place of publication (in some cases).
- c) **Reception.** Various kinds of secondary literature, i.e. reviews, interviews, summaries, letters, polemics, etc. Each reception piece is connected to one literary author (creator). Some reviewers were also authors, but in this study, I have omitted this relationship. Certain reception pieces noted in the bibliography are about non-literary books and authors (e.g. a review of a literary study).

I prepared tables for all the datasets, which will allow for further computation. This process entailed data cleaning (esp. removing typos and mistakes) and enriching the data with additional information (e.g. author's gender, and latitude and longitude of cities). Certain variables have a low reliability due to changing descriptive conventions and human errors, for example, the number of pages, or the journal issue numbers.⁴

Altogether, the dataset I extracted from the PBL for this study indexes 33,142 literary books and 138,925 literary articles, which were together by 21,865 authors. These were published by 8,233 publishers and 958 journals. There are 93,874 reception texts, including pieces concerning 9,286 authors and 22,335 books.

4. Studies of dispersion

The main hypothesis, that is, the movement from dispersion towards centralisation and back, as described by critics; could be captured on many levels. The most obvious one, stemming from hypothesis one (H1), is simple bibliometric statistics, and could reveal the structure and evolution of literary life over the transformation period. I commence with some statistics for the literary production of that period, then combine those findings with the reception data (H2), and subsequently, with a comparison of the reception of debutants and 'old masters' (H3). Next,

⁴ I am very thankful to Cezary Rosiński from the PBL team, who extracted the data and prepared the list of debutants referred to in the next section.

I analyse the dispersion using the example of geography (H4); and finally use network analysis to obtain centrality measures in order to address the main hypothesis.

4.1. Literary production

Let us start with the sheer numbers concerning the literary production of that period. The dispersion of literary production (H1), operationalised using the available dataset, would entail an increase in the number of smaller publishers, journals, or publishing centres disseminating literary content over time. Recentralisation would mean the reverse, that is, a smaller number of big players overtaking literary communication. Let us see how this looks through some of the available measures.

To put the discussion into context, we should start with the general statement that in Polish literature, the period in question could be characterised regarding growth. Figure 1 combines the yearly numbers of publications (books and journal pieces), the active literary journals (in this example only those which published literary works, not reception pieces), as well as a count of publishers and authors (i.e. those who published at least one book or journal piece). The trend lines show a steady growth of the first two groups and quite a stable situation for the last one. This tendency is not so evident if we take into consideration economic factors (the diminishing of state-subsidised publishing) as well as competition from other media and computers during the 1990s.

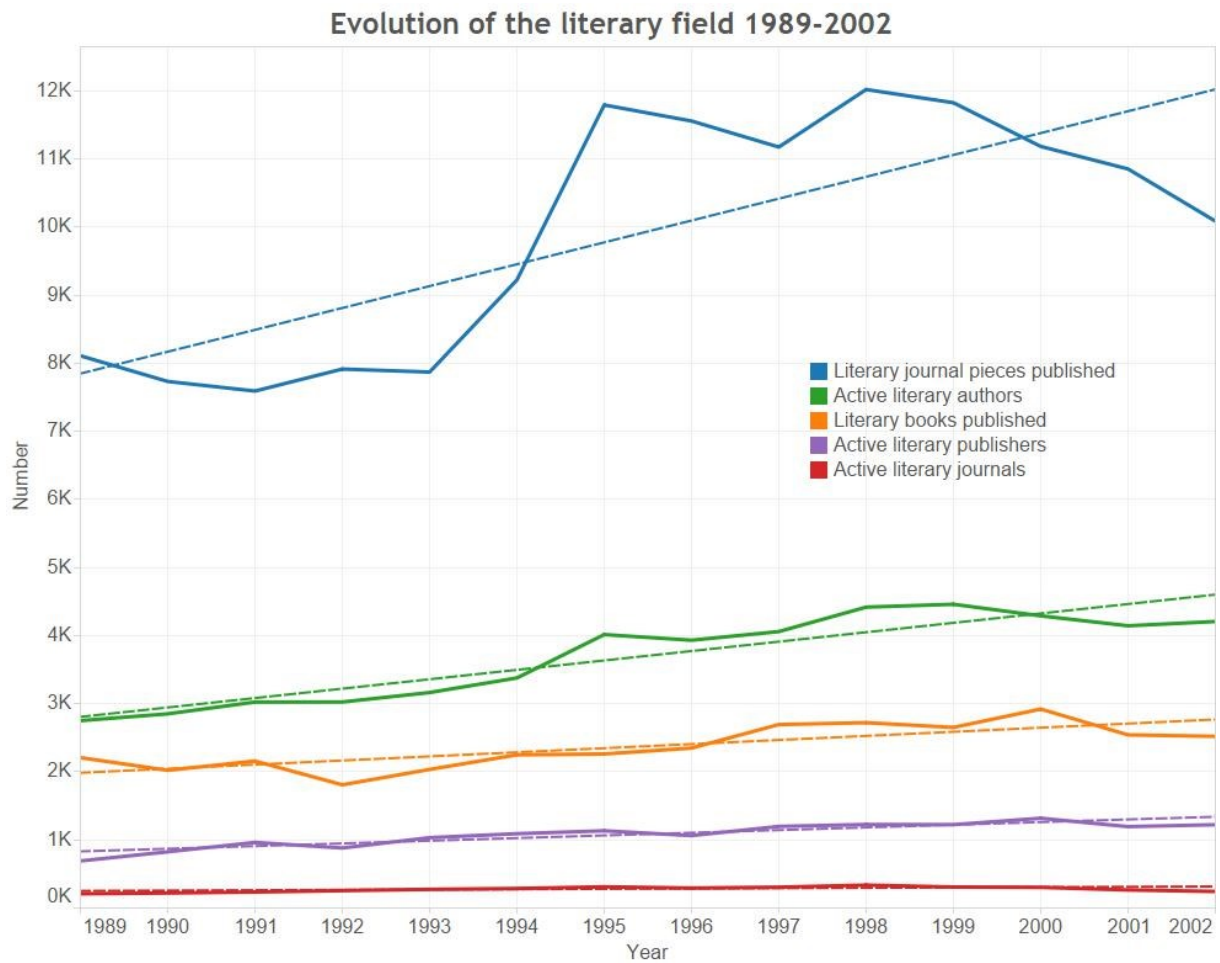


Figure 1. Evolution of the literary field.

In order to add some granularity to these frequencies, I divided the publishers and journals into big and small. If the critical assumptions are correct, we should see many small, dispersed presses in the early 1990s; and a subsequent market re-concentration towards the turn of the decade. In order to differentiate the publishers and journals, I took their annual publication numbers and applied the ninetieth percentile as the threshold for qualifying a journal or publisher as being big or small. In other words, big players made up 10% of the most prolific journals and publishers each year. It should be noted that the quality of being ‘big’ or ‘small’ in this study is always relative to a single year and the publishing numbers of other entities. For instance, a publisher or journal with the same number of publications over time may be qualified as big in one year, and small in another, depending on the output of other publishers.

The area graphs in Figure 2 show an interesting discrepancy between the actual number of journals and publishers, and their average output, that is, the average number of publications for active entities in all four categories. The production seems to be dominated by large players, which is especially true for big journals after 1995, and big publishers after around 1998.

Interestingly, this growth of big publishing companies coincides with a noticeable drop in the number of active players in the early 2000s.

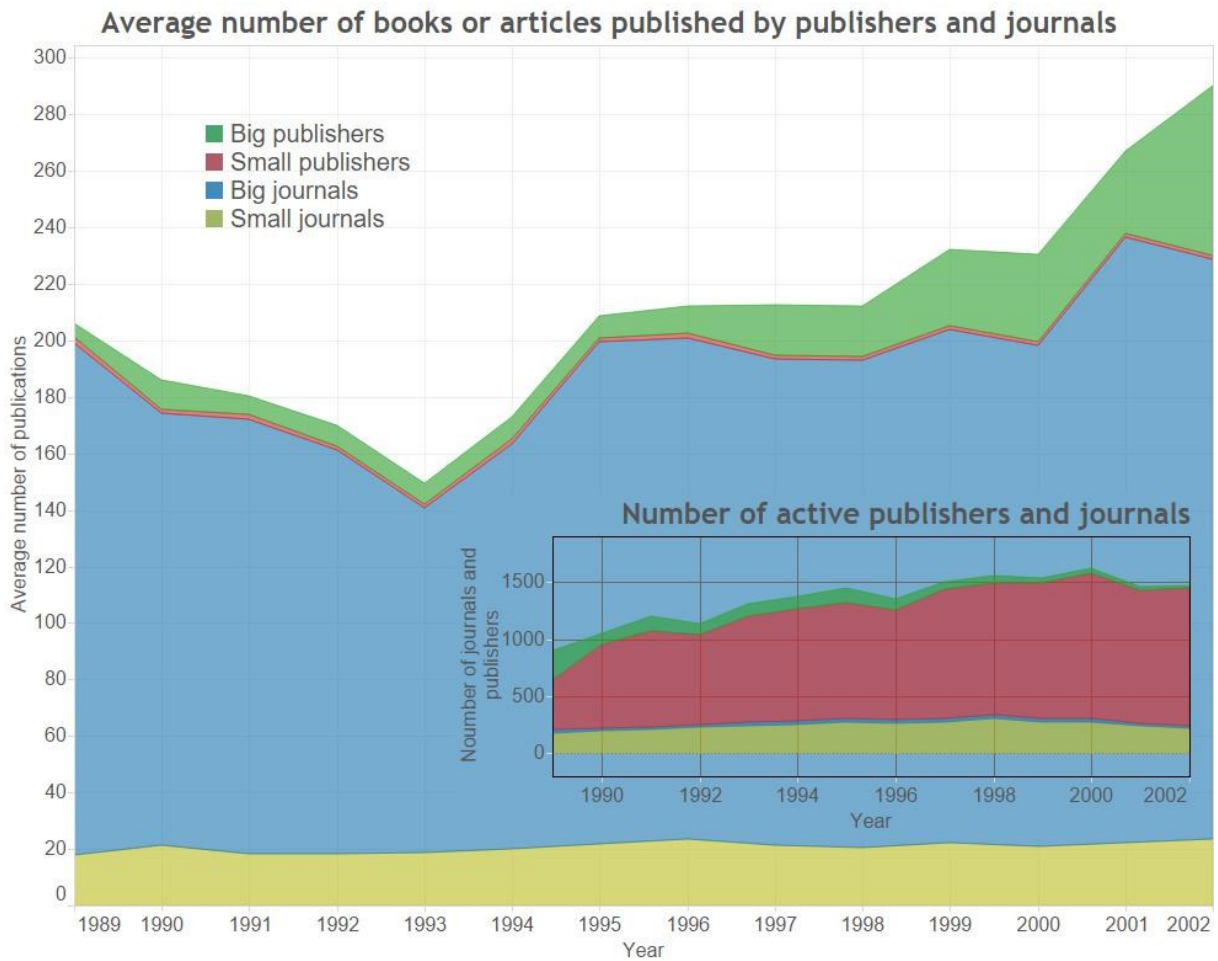


Figure 2. Average number of books or articles published by publishers and journals.

The overrepresentation of small literary players may be partially caused by the data collection methodology, which does not register non-literary books. For instance, if a big publisher in the field of history releases one volume which is pertinent to literary studies (e.g. a collection of letters by a literary figure), only this book will be counted, and the publisher would appear in the PBL as being small.

4.2. Critical reception

Sheer production statistics could be misleading if we fail to take into consideration the reception practices over time (H2). If we want to analyse the tendency for centralising literary taste, we need to compare the number of books with their reception; namely, how many times a particular book is reviewed, discussed, or mentioned. The data in Figure 2 seem to at least partially corroborate the claims of literary critics about the eventual recentralisation of literary life by a few of the bigger players. This trend is even more visible if we take into consideration the

critical attention which different publications attracted. Figure 3 plots the number of book reviews by big and small publishers.



Figure 3. Number of reviews dedicated to books from big and small publishers

The critical attention devoted to books mirrors the processes described by the critics. Books by big publishers which attracted most of the critical attention clearly dominated Poland’s centralised market in 1989. Yet they are swiftly outpaced by small publishers after 1991. However, this trend quickly reverses around 1995, and within two years big publishers regain their dominance in the struggle for critical attention. This trend is perhaps even more strikingly visible when compared with the actual publishing output.

In Figure 4, the books by big (blue) and small (red) publishers are contrasted with the ratio of critical attention (the green line represents the number of book reviews of publications by small publishers, divided by reviews of the output from big publishers; a score above 1 means a higher interest in small publishers, whereas, a value below 1 indicates that critical attention shifted towards the big players). It is worth noting that 1989 was the only year when big publishers published more books than the small publishers (55%). During the entire period, on average, small publishers published 57% of all books (median = 58%); and in 2001 and 2002

they had record numbers with 58% and 62% of all books published, respectively. This pattern is compatible with the critical narrative about the recentralisation and emergence of the mainstream in the late 1990s.

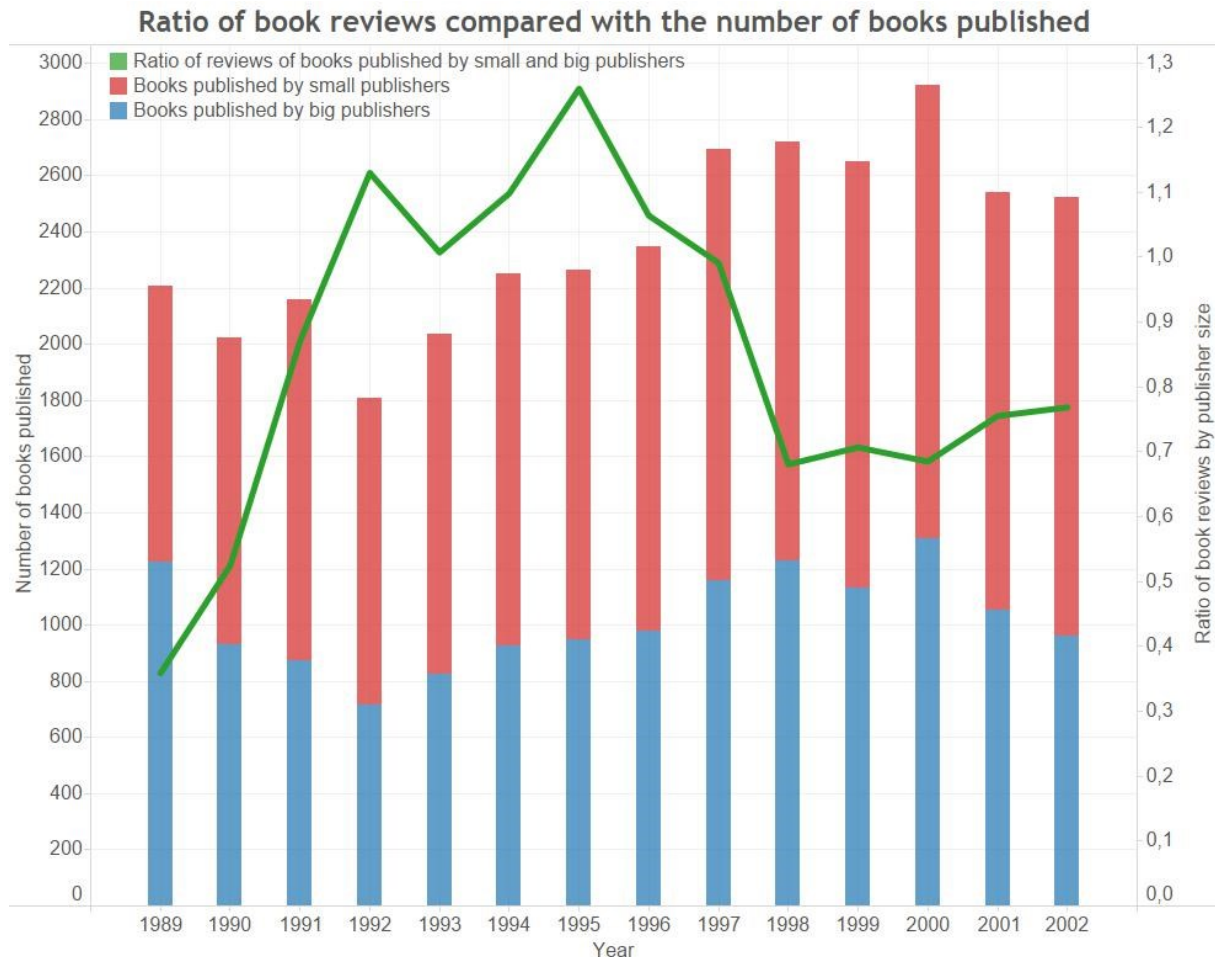


Figure 4. Ratio of book reviews of publications from small and big publishers compared to their output.

In order to better understand these relationships, one can apply some measurements related to the economy.⁵ In this case, the Gini coefficient, which is usually employed to measure inequality in societies, is particularly useful. We may consider literary life to be a system in which attention is given to authors in three ways: by publishing their books, by publishing their shorter pieces in journals, and by writing about the author or their work. Measuring the distribution of this attention allows the level of the writers' equality to be judged according to these variables, and for further inferences about the processes of literary life to be made.

⁵ I am grateful to Mark Algee-Hewitt and Ryan Heuser for pointing me in this direction.

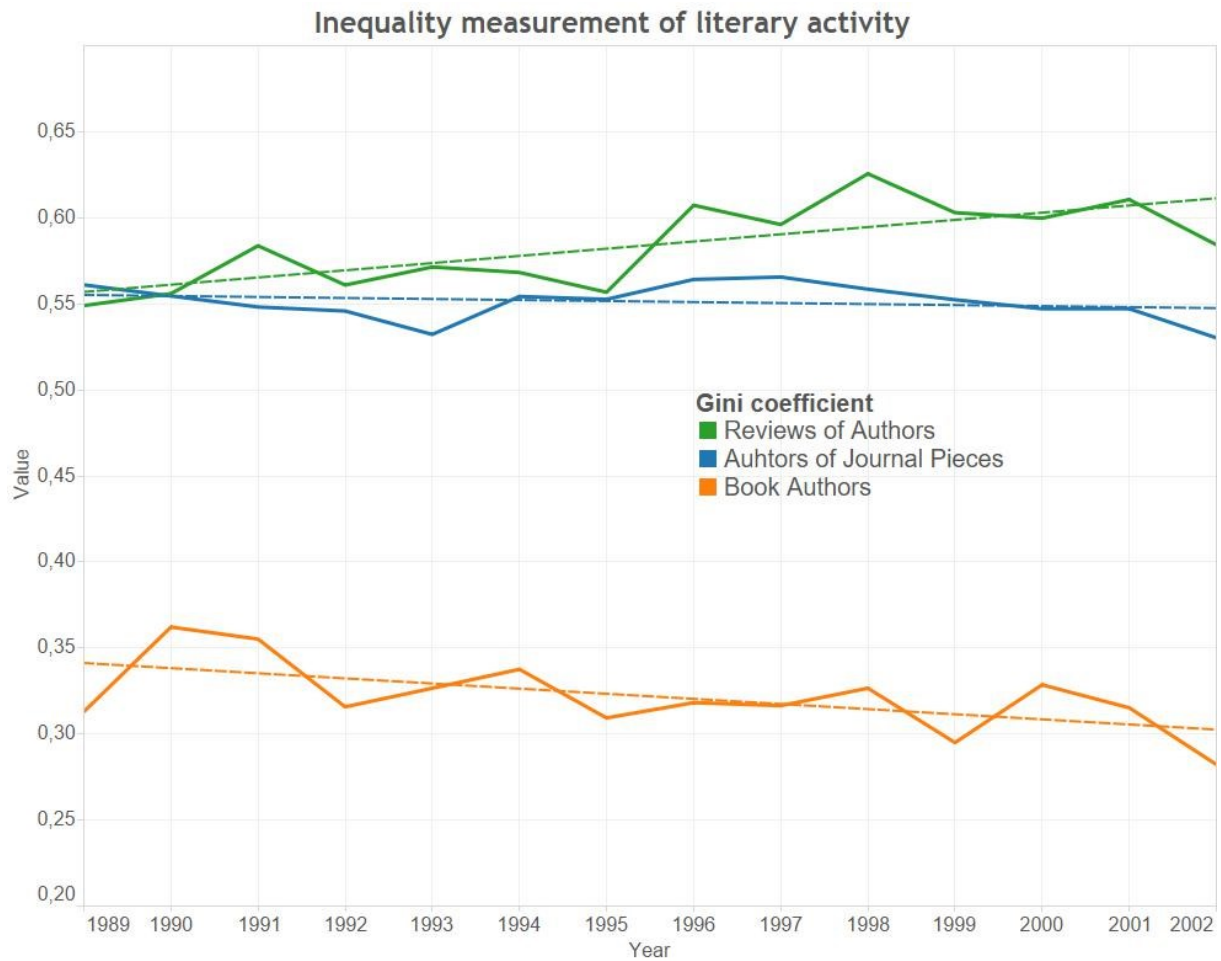


Figure 5. Gini coefficient of literary attention

The Gini coefficient is ‘based on the Lorenz curve, a cumulative frequency curve that compares the distribution of a specific variable (for example, income) with the uniform distribution that represents equality’ (Haughton & Khandker 2010, 104). The Gini coefficient has a value of between 0 and 1, which indicates the distance between an actual distribution (as represented by a Lorenz curve) and a perfectly equal one. Hence, the value 0 means perfect equality (all publish the same number of books), and 1 stands for the opposite, absolute inequality (one author publishes all the books).

The data used for computation was prepared for three variables (books published, journal pieces published, personal reception pieces about the author) for every year between 1989 and 2002. Authors not present in a given year were omitted. The Gini coefficient was computed in the *ineq* package for R (Zeileis 2014). The results are presented in Figure 5.

First, it should be noted that it is hardly surprising that literary culture seems to be unequal, yet what strikes us in Figure 5 is the discrepancy between the distribution of the published books and the other variables. How to interpret this finding? If we look at the World Bank statistics for economic distribution in various countries, we will notice that even the most unequal societies rarely score higher than 0.5 on the Gini index. Through this comparison we may conclude that the distribution of critical attention and journal pieces is highly unequal, whereas the authors who publish books are represented more equally.

Moreover, the data on publishing show a slight downward trend, whereas the inequality of critical attention is on the rise and seems to follow the same pattern that is visible in the previous figures (i.e. a sharp rise in the second half of the 1990s and a drop in the last year). This is the only variable in this figure which follows the critics' claims, and seems to point to some form of re-centralisation processes in the field of secondary literature. Let us examine this point more closely through the PBL data-based analysis of another metaphor which describes the transformation processes of Polish literary life for this period – the return of the 'old masters.'

4.3. *Debutants and masters*

In order to assess hypothesis 3, (i.e. the claim about the *debutocentrism* of the early 1990s and the subsequent 'return of the old masters'), using the PBL dataset, the notions of 'debutants and masters' were operationalised as follows: 'debutants' were those authors who had not published anything before 1989, whereas 'masters' were those authors who were highly popular before 1989 and also featured in criticism (e.g. Czesław Miłosz, Wisława Szymborska, Tadeusz Różewicz, Zbigniew Herbert, Stanisław Lem). The third, and largest, group of writers (including historical figures) was left out. This operationalisation resulted in two groups consisting of 11 masters and 3,907 debutants. The reception for both groups is plotted in Figure 6. The y-axis represents the percentage of critical texts dedicated to authors for both groups, relative to all articles about authors published in a given year. Yet despite the similarities in the scores, we must not forget that the groups being compared differ exceptionally in size.

We may start with the observation that the trends in both cases point to an increasing interest in both groups by reviewers. Furthermore, although the steep rise in writings dedicated to debutants in the early 1990s may be an indicator of *debutocentrism*, we hardly see any decline in the second part of the decade. This could be a side effect of the operationalisation, which defines debutants as all authors who published for the first time in 1989 or later. Hence the number of debutants increases over time, which in turn may cause the gradual rise of critical attention. Yet if we concentrated on the trends in the first half of the 1990s, we can observe a

decline in interest in acclaimed writers, who are overtaken by debutants by 1995. Suddenly, this trend almost reverses between 1996 and 1998 and afterwards.

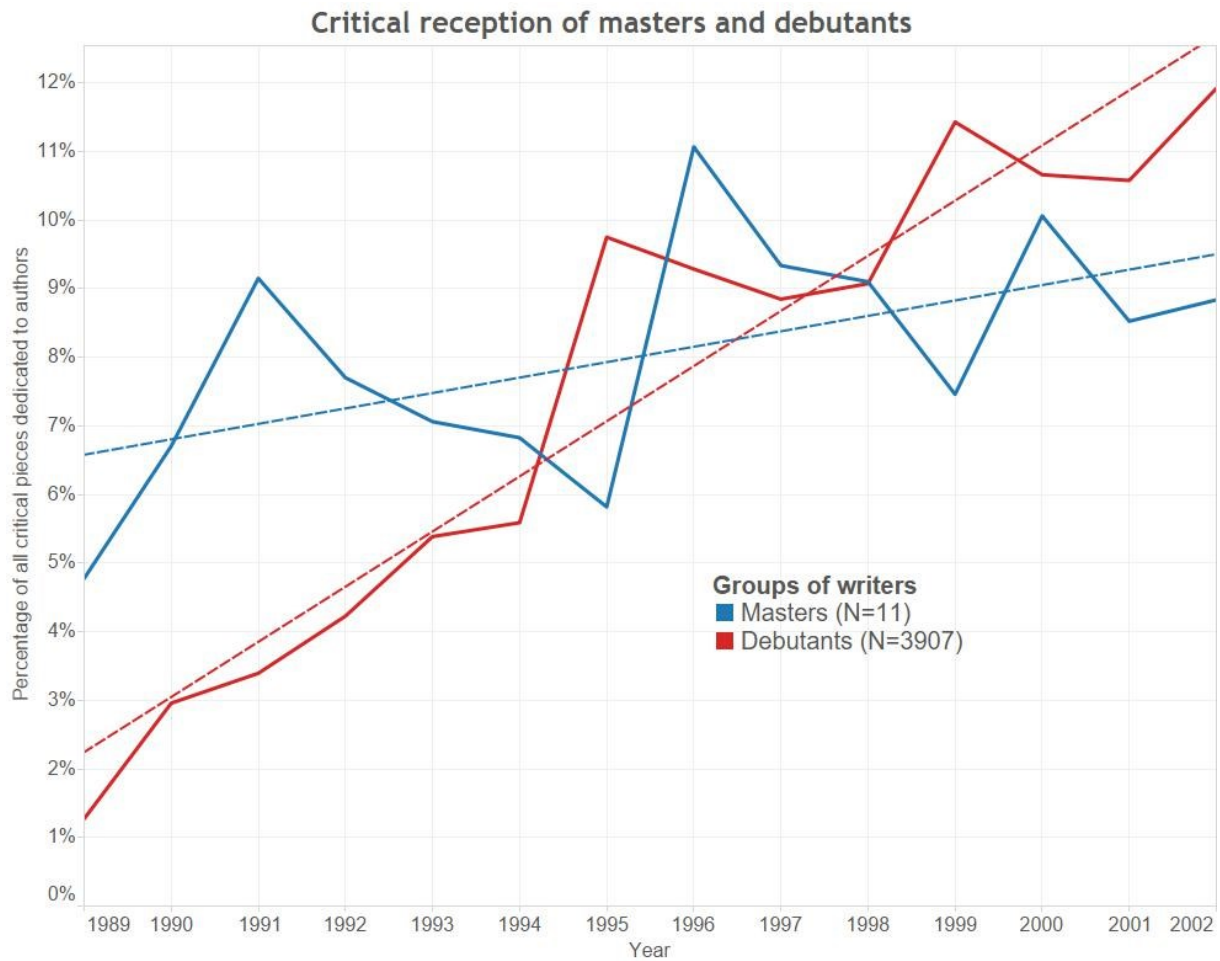


Figure 6. Critical reception of masters and debutants

A closer scrutiny of the data allows us to link those irregular spikes on the masters' line with particular events: such as Wisława Szymborska being awarded the Nobel prize (1996); the deaths of the eminent writers, Zbigniew Herbert (1998) and Gustaw Herling Grudziński (2000); as well as the ninetieth birthday of another Polish Nobel-prize winner, Czesław Miłosz (2001). In all cases, the abundance of writings about these authors would dramatically increase the overall reception of the masters. Moreover, such events were not only limited to living authors. For instance, the two-hundredth birthday of Adam Mickiewicz (1998), an eminent poet of Polish romanticism, generated over one thousand articles about him. And it seems that the Nobel prize for Szymborska may have been a turning point in (or at least an influence on) Polish literary history in respect to it renewing interest in established writers. It should be highlighted, however, that this pattern of interest seems to be incidental and related solely to big events concerning particular writers. Szymborska is a good example here – in her Nobel Prize year there were 306 critical pieces about her, almost twice as many as during the preceding seven

years combined (127). A year later, there were 180 pieces about her, and in 1999 she was featured in roughly the same number of articles as in the year before winning the prize (18). This interest seems to be quite short-lived and seasonal, yet it may explain the inequality patterns in critical attention of the late 1990s when individuals were the subject of a disproportionately high degree of attention.

4.4. Geography

H4 states that the processes of dispersion also have a geographical dimension. If the critics are right, we should see the rise and subsequent decline of smaller publishing centres throughout the transformation period. In order to map this process, I compiled information about the locations of the publishers and journals featured in the PBL dataset. I excluded publishing locations outside Poland in order to minimise the number of random locations. I computed the number of books and articles published each year in each of the 607 Polish cities. In order to locate larger centres, I used the same procedure I used in the case of publishers and journals, that is, setting a high publishing level at above the ninetieth percentile of all the city publications for a given year. Importantly, these levels were computed separately for books and for journal pieces; hence, a given city could be considered to be a large centre regarding book publishing, but simultaneously classified as a small centre in journal pieces published (and *vice versa*). The classification of a given city may also differ over time.

Figure 7 contains two representations of these data. The top graph shows the number of works published in cities, where the blue colour signifies a high level of literary production in a city, and red represents a low level of publishing. Numbers are given for both books (squares) and journal pieces (ovals). In terms of publishing, we see a steady rise for small publishing centres, who reach a similar level to that of large publishers in 2002. Regarding journals, we witness growth for both types of cities. The bottom graph uses the same features but arranges them according to the actual number of cities falling under a given category. Here again, journals behave differently as we see a small rise and subsequent decline in the number of smaller locations who host literary journals. Concerning books, both levels seem to rise, and there are more locations with both smaller and larger outputs.

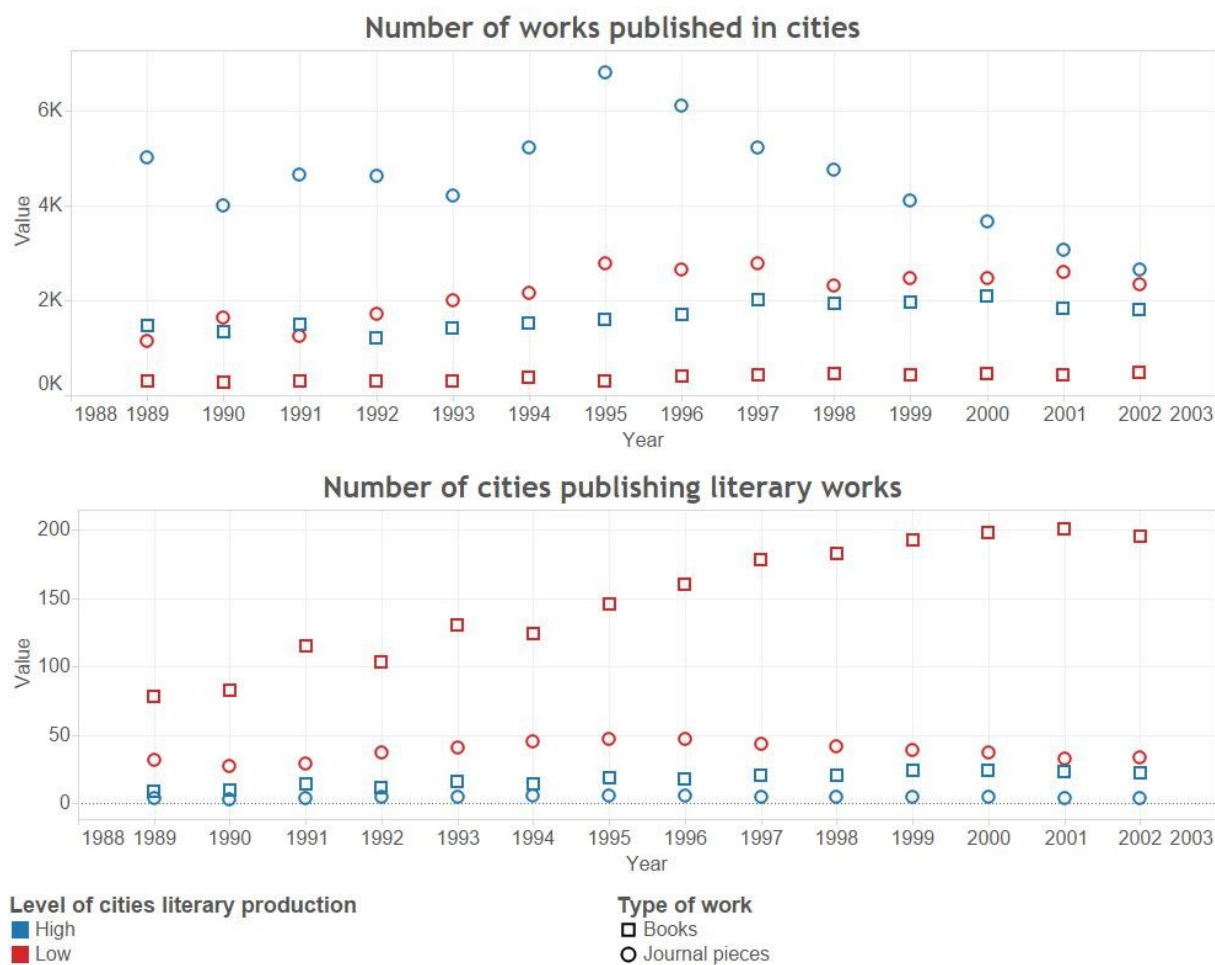


Figure 7. Number of publications in Polish cities

In order to shed more light on these numbers, let us once again turn to the economy as a reference field. In economic studies the Herfindahl-Hirschman Index (HHI) is usually used to describe market concentration, and can therefore be used to answer the very question we are dealing with here. The HHI is the sum of the squares of all the shares in the market, and I computed it for the PBL dataset for R, using the *DescTools* package (Signorelli et al. 2008). As in the case of the Gini coefficient, I generated a table containing the number of books and journal pieces published in a given city each year. Cities without any publications in a given year were omitted. The results are presented in Figure 8. The HHI usually has values from 0 to 1,000, with the highest value representing a monopoly. As it is the sum of the squares of all entities' shares in the market, we should not be surprised to find that the distribution of literary production among a few hundred cities is highly deconcentrated, as many cities have a meagre publishing output (cf. Figure 7). We should, therefore, not concentrate on the absolute values, but rather on the trends and changes over time.

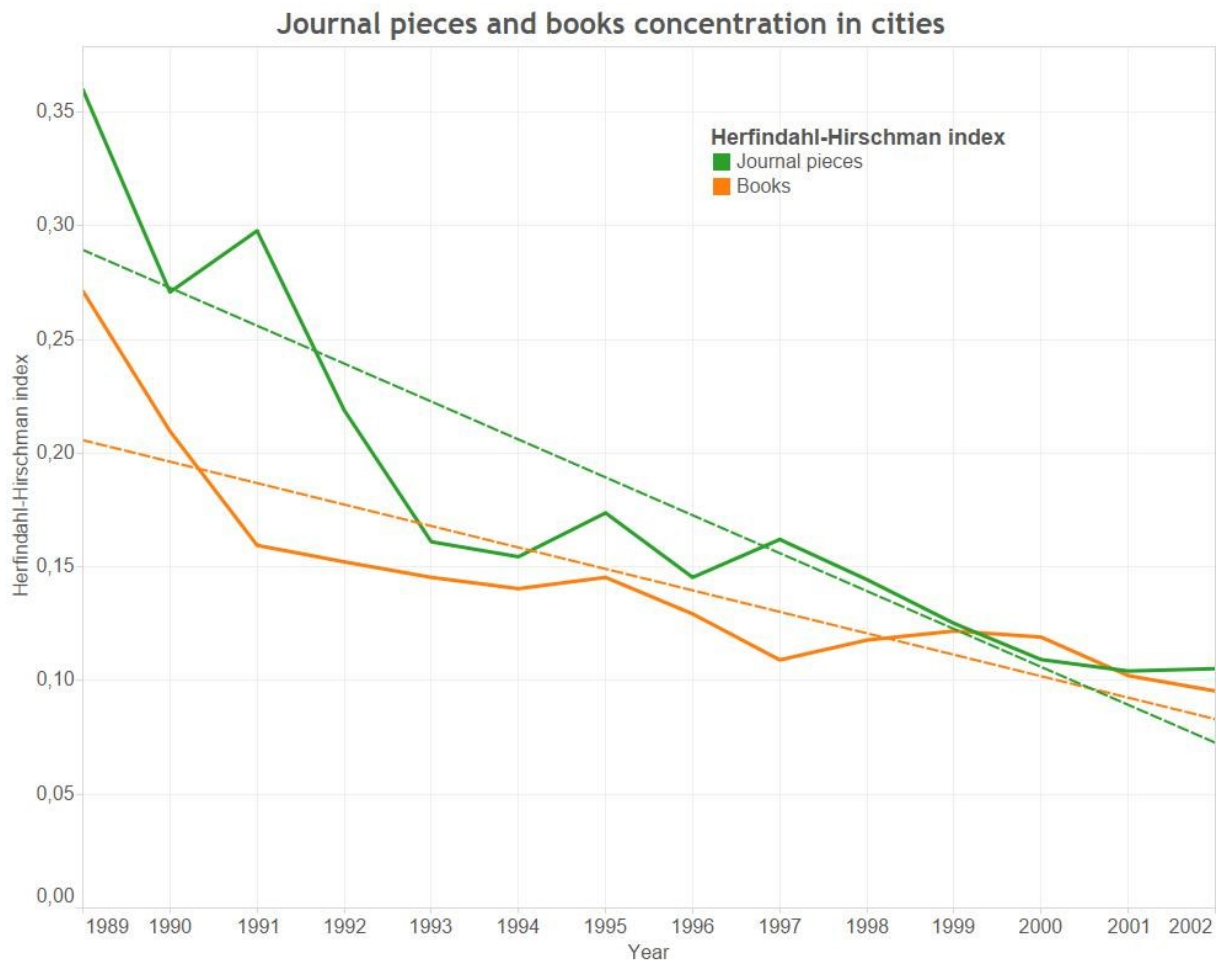


Figure 8. Journal pieces and books concentration in cities

The HHI plot in Figure 8 shows a clear trend towards deconcentration in terms of both the number of books and the number of journal pieces published in cities. It seems that almost every year new actors emerge and take their small share of the market. Although we can hardly see any trend towards re-centralisation, it should be noted that in both cases we can distinguish between a radical deconcentration during the first five years, and a somewhat steady downward movement in the last five years. In between (i.e. around 1995) we even witness a small movement towards concentration.

4.5. Network analysis: a combined view

So far I have used various measurements with single variables to account for the processes in question; namely, the dispersion and recentralisation of literary production, reception, and attention to particular groups of writers, as well as the geographical context of those processes. In the concluding study, I borrow tools from network analysis to look at literary life as a network of relationships and interdependencies in order to test the main hypothesis of dispersion and recentralisation on the model of literary life. On the basis of relationships such as, being published by the same publisher, being covered in a journal, and receiving the same prize, one

can build a network of literary life and explore its shape, which could be centralised or dispersed (cf. Long and So 2012; So and Long 2013).

The network of literary life that emerges from the PBL dataset is complex because it features different types of actors and relationships between them. I decided to work with the author-centred network and link them with other entities: publishers (published by), journals (published in, or reviewed in), and reviewers (reviewed by). One can, of course, oppose to using reviews twice – first, to mark the author’s relationship with a reviewer, and then with a journal – but it creates an additional layer of connections (a reviewer usually publishes in other journals, and a journal usually publishes many reviewers). I did not link reviewers to journals (so as to keep the network author-centred), and I did not use cities (in order not to introduce another entity class). The network contains a total of 339,287 unweighted connections (including duplicate entries). For each year I prepared a list of relationships between the authors on one side and the remaining entities on the other. These tables were then transformed into lists of nodes and edges using *Gephi* (Bastian, Heyman and Jacomy 2009). Each node (i.e. the relationship between two entities) was assigned a weight which indicates the number of times they were connected to in a given year.

This network contains many different kinds of entities but I will simplify its interpretation and will treat it as a two-mode undirected network of writers and literary institutions. In order to deepen the understanding of this network, I also use a one-mode ‘projection’ of this network, which was prepared with the use of bibliographic coupling in *Sci2* (Sci Team 2009; cf. Graham 2012). A ‘projection’ is a translation of a two-mode network into the relationships among entities of one kind (Cf. Opsahl, Agneessens, and Skvoretz 2010). For instance, authors A and B, published novels with publisher X, while authors B and C published their poems with publisher Y. In a two-mode network there are no connections between authors, but only between authors and publishers (A-X, B-X, B-Y, C-Y). In a one-mode network, relationship with publishers are translated into connections between authors on the basis of co-occurrence. So, B is connected to both A and C, but A and C are not connected directly because they did not publish together. Transforming a network into a one-mode projection is often essential in order to proceed with network statistics as most measures are designed for this kind of network (cf. Graham, Milligan, and Weingart 2015, 195–234).

The network concentration was measured using the average degree statistics. A node’s degree is the sum of its connections, while the average degree is the mean of all the degrees in the network. The concept of degrees differs in one and two-mode networks. If a writer published with only two publishers, both of whom published the books of ten other authors, her two-mode

degree would be 2 (2 publishers), while her one-mode degree would amount to 20 (2 x 10 authors). And so the latter is larger than the former. Figure 9 shows the average degree for one-mode and two-mode networks – the former was computed using *Gephi*, and the latter using the *tnet* package for R (Opsahl 2009).

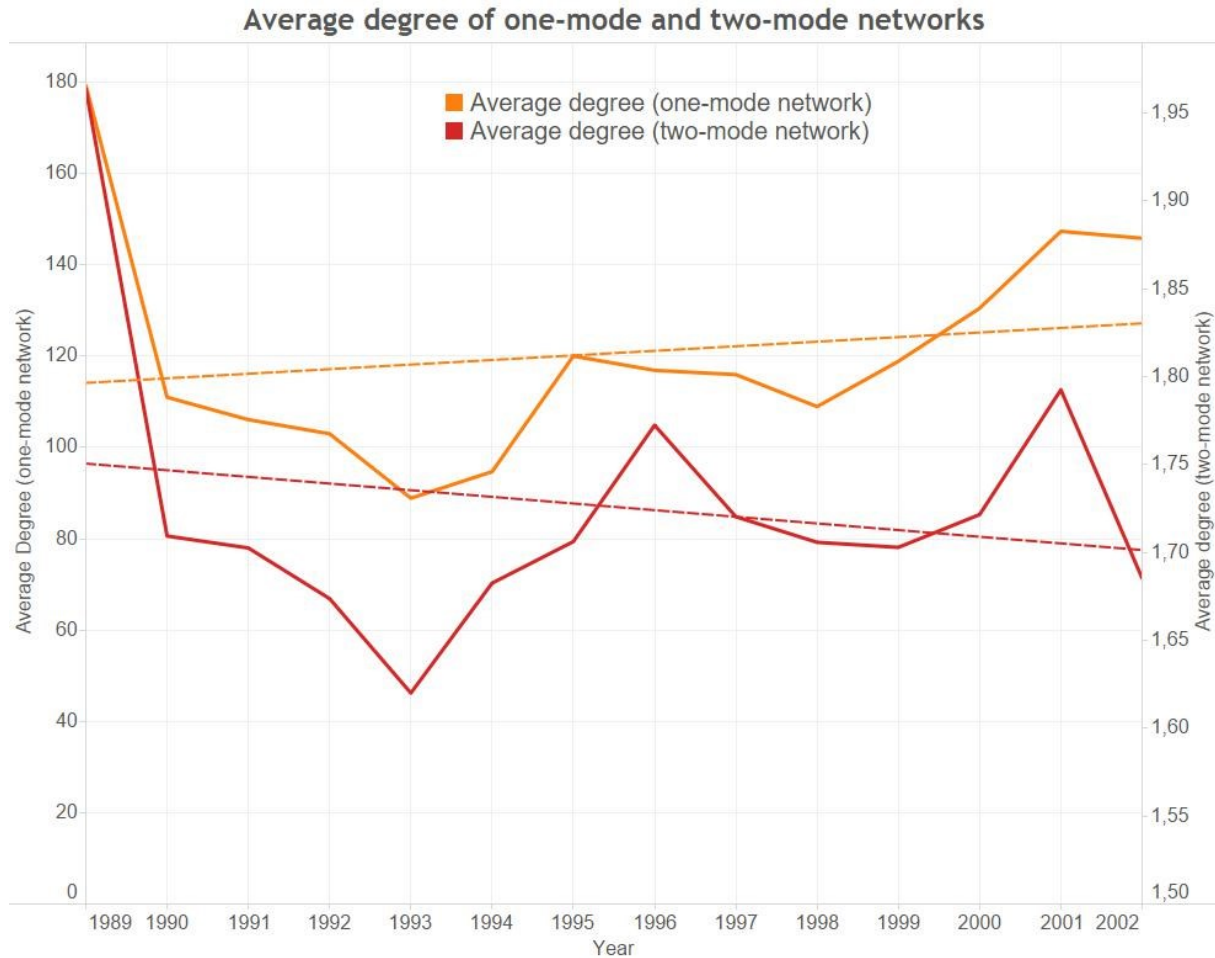


Figure 9. Average degree of one and two-mode networks

For reasons described above, both networks operate on different scales, but in the graph they were adjusted by the starting point, to allow for visual comparison. The one-mode network of relationships between writers (orange) seems a perfect fit for the dispersion–recentralisation narrative, because after 1994 writers, on average, seem to increase their connections. Quite interestingly, the two-mode network (red), which focuses more on institutional relationships, tells a similar, yet not precisely the same narrative. It seems as if it was more prone to extremes (1993, 1996, 2001), as, generally, the fewer number of connections makes individual writers more influential in the overall score. The drop in 1993 may signify the peak of decentralisation, whereas spikes in 1996 and 2001 correspond with Szyborska and Miłosz respectively becoming the centres of literary attention. The sudden drop at the end corresponds with the general growth of book publishing, also visible on previous graphs. So, although the big players

may actually publish more books, they remain obscured on the two-node network due to the wave of new publishers. The one-mode network, in contrast, rises, probably because large publishers produce more books in general, which allows new connections to be established through co-occurrence.

5. Conclusions

Probably the most obvious conclusion of this study is that there is no single clear way to operationalise the claims made by the critics for the purposes of a data-set based validation. Once we start to disentangle the complex nature of literary life, it appears that certain measures correspond with the critical predictions and some do not. Interestingly the patterns which seem to corroborate the critical claims almost always contained the reception variable: patterns of critical attention given to big and small publishers, the reception of debutants and old masters, the Gini coefficient, and the network analysis. The only time the decentralisation pattern was visible without the reception component was in the study of literary production concentration using the Herfindahl-Hirschman Index.

The claims of the critics are based on the various materials they have at hand: literary texts, critical discourse, scholarly articles etc. Some of those claims are grounded in empirical data (e.g. Czaplński and Śliwiński [1999] use a number of journals having been shut down in the early 1990s, new literary prizes, number of debuts, etc.). Their interpretation of macro processes is then an individual interpretation of the events. However, the sharper the diagnosis, the more questions it raises: What phenomena are taken into account? Which ones were left out? Were all texts treated equally? However, similar questions may be directed towards the bibliographical material in the PBL, which, as the methodological section of this paper pointed out, is far from being an ‘objective representation,’ and is quite a comprehensive sample of literary material, gathered according to a specific methodology and by no means free from errors.

Furthermore, critics do not work in a vacuum – they engage in literary debates and take stands on issues raised by other critics, and they are not mere observers but active participants in the events in question. Hence, it is hard to draw a clear border between descriptive and prescriptive activity. It is quite telling that the critics quoted in this volume authored 1.6 % of the reviews in the PBL dataset...

Quantitative approaches, on the other hand, allow us to go beyond the critical writings and take a look at some different dimensions of literary life, especially at those which usually

escape critical scrutiny, as the ‘long tail’ of literary production. The obvious question that arises is which construct of literary life is the ‘right’ one: quantitative analysis or critical synthesis? Are we talking about the same literary culture? If literary processes are visible most clearly in the secondary literature, which is itself a particular interpretation of literary facts, the attempted reconstruction of critical diagnoses may be a reconstruction of a specific aspect of literary life, namely, the actual critical debates.

The primary goal of this study was to corroborate certain claims concerning the processes of literary life, and the results show that a particular type of generalisation is possible: the dispersion–recentralisation narrative is visible, especially when literary reception is considered, or in the exploration of the literary-world as a network. This is why the central, yet problematic variable in this study is time. As we all know too well, the fact that a certain literary piece was published in a given year and not the previous or the following one is often a matter of chance. One can blur the time boundaries by using a moving average but this, in turn, would soften the impact of individual literary events, as the case of Szymborska’s Nobel Prize demonstrates. This study marks a point of departure for the further work with the PBL data, which – freed from the constraints imposed by the corroborative goal of the present study – would look at further avenues of macroanalytical research into literary using bibliographic data.

Acknowledgments

The research presented in this study was made possible by the National Programme for the Development of Humanities (Research grant: The Polish Literary Bibliography – a knowledge lab on contemporary Polish culture, NR 0061/NPRH3/H11/82/2014), the Polish-American Fulbright Commission (Fulbright Senior Award: Literary transition as viewed from above. Quantitative history of Polish literature [1989–2000]. Research visit to Stanford University), and COST Association (COST Action CA16213: New Exploratory Phase in Research on East European Cultures of Dissent). The author is greatly indebted to colleagues from the Stanford Literary Lab (Mark Algee-Hewitt, Ryan Heuser, Erik Fredner, J.D. Porter, and Hannah Walser) for their insightful feedback on this project, as well as to the participants of the DFG conference at Villa Vigoni. Special thanks to Cezary Rosiński for his help with the data dump. All mistakes remain my own.

Bibliography

- Allison, Sarah, Ryan Heuser, Matthew L. Jockers, Franco Moretti, and Michael Witmore. *Quantitative Formalism: An Experiment*. Pamphlets of the Stanford Literary Lab, pamphlet 1. Stanford, California, 2011.
- Bagłajewski, Arkadiusz: Stan rozproszenia. In: *Odra* . 1(1996), 51-55.
- Bagłajewski, Arkadiusz. *Mapy Dwudziestolecia 1989-2009: Linie Ciągłości*. Lublin 2012.
- Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media, 2009.
- Bode, Katherine: *Reading by numbers. Recalibrating the literary field*. London 2014.
- Bode, Katherine: The Equivalence of “Close” And “Distant” Reading. Or, toward a New Object for Data-Rich Literary History. In: *Modern Language Quarterly*, 78/1(2017), 77–106.
- Czapliński, Przemysław, and Piotr Śliwiński: *Literatura Polska 1976-1998: Przewodnik po prozie i poezji*. Kraków 1999.
- Czapliński, Przemysław: *Kalendarium Życia Literackiego 1976-2000: Wydarzenia, Dyskusje, Bilanse*. Kraków 2003.
- Czapliński, Przemysław: *Powrót centrali: Literatura w nowej rzeczywistości*. Kraków 2007.
- Czapliński, Przemysław: *Ruchome Marginesy: Szkice O Literaturze Lat 90*. Kraków 2002.
- Czapliński, Przemysław: *Ślady Przełomu*. Kraków 1997.
- Eder, Maciej: Style-markers in authorship attribution: a cross-language study of authorial fingerprint. In: *Studies in Polish Linguistics* 6 (2011), 99-114.
- Eder, Maciej, Kestemont, Mike and Rybicki, Jan: Stylometry with R: a suite of tools. In: *Digital Humanities 2013: Conference Abstracts*, University of Nebraska-Lincoln, NE, 2011, 487-489.
- Graham, Shawn, Converting 2-mode networks to 1-mode networks (2012), <https://electricarchaeology.ca/2012/02/08/converting-2-mode-networks-to-1-mode-networks/>.
- Gawliński, Stanisław: Wstęp In: Stanisław Gawliński and D. Siwor (Eds.): *Dwie dekady nowej (?) literatury 1989-2009*, Kraków 2011.
- Graham Shawn, Ian Milligan, and Scott Weingart: *Exploring Big Historical Data: The Historian's Macroscope*. London, 2015.
- Haughton, Jonathan, and Shahidur R. Khandker. 2009. ‘Handbook on Poverty and Inequality’. The World Bank.

- <http://documents.worldbank.org/curated/en/488081468157174849/Handbook-on-poverty-and-inequality>.
- Hoover, David L, Jonathan Culpeper, and Kieran O'Halloran: *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*, New York 2014.
- Jannidis, Fotis and Lauer, Gerhard: Burrows's Delta and Its Use in German Literary History. In: M. Erlin and L. Tatlock (Eds.) *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, New York 2014, 29-54.
- Jarzębski, Jerzy: *Apetyt na przemianę. Notatki o prozie współczesnej*. Kraków 1997.
- Jockers, Matthew L.: *Macroanalysis. Digital Methods & Literary History*. Chicago 2013.
- Jockers, Matthew L.: *Text Analysis with R for Students of Literature*. Cham 2014.
- Klejnocki, Jarosław, and Jerzy Sosnowski: *Chwilowe Zawieszenie Broni: O Twórczości Tzw. Pokolenia "brulionu" (1986-1996)*. Warszawa 1996.
- Long, Hoyt, and Richard So: Network Science and Literary History. In: *Leonardo*. 46/3 (2013).
- Maryl, Maciej, Wciślik, Piotr: Remediations of Polish Literary Bibliography: Towards a Lossless and Sustainable Retro-Conversion Model for Bibliographical Data. In: *Digital humanities 2016: Conference Abstracts.*, Kraków 2016, 621-623.
- Maryl, Maciej, Eder, Maciej: Topic Patterns in an Academic Literary Journal: The Case Of *Teksty Drugie*. In: *Digital humanities 2017: Conference Abstracts*. Montreal 2017.
- Moretti, Franco: *Literature, Measured*. Pamphlets of the Stanford Literary Lab, pamphlet 12 Stanford, California 2016.
- Moretti, Franco: *Distant Reading*. London 2013a.
- Moretti, Franco: "Operationalizing": or, the function of measurement in modern literary theory. Pamphlets of the Stanford Literary Lab, pamphlet 6 Stanford, California 2013b.
- Nasiłowska, Anna. *Literatura Okresu Przejściowego 1975-1996*. Warszawa 2006.
- Opsahl, Tore: *Structure and Evolution of Weighted Networks*. University of London (Queen Mary College), London 2009, pp. 104-122. Unpublished thesis available at <http://toreopsahl.com/publications/thesis/>
- Opsahl, Tore, Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. In: *Social Networks* 32/3 2010, 245-251.
- Rybicki, Jan and Eder, Maciej: Deeper Delta across genres and languages: do we really need the most frequent words? In: *Literary and Linguistic Computing* 26/3 (2011), 315-21.
- Rybicki, Jan: Pierwszy rzut oka na stylometryczną mapę literatury polskiej. In: *Teksty Drugie*. 2 (2014), 106-128.

- Sci2 Team. Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, 2009, <https://sci2.cns.iu.edu>.
- Signorell, Andri et. al., *DescTools: Tools for Descriptive Statistics*, R package version 0.99.24, 2018, <https://cran.r-project.org/package=DescTools>.
- So, Richard and Long, Hoyt: Network analysis and the sociology of Modernism. In: *Boundary* 2, 40/2 (2013), 147–82.
- Schöch, Christof and Pielström, Streffen. Für eine computergestützte literarische Gattungsstilistik. In: *1. Jahrestagung der Digital Humanities im deutschsprachigen Raum (DHd)*. 2014.
- Sławiński, Janusz: Zanik centrali. In: *Kresy* 18 (1994), 14-16. (1994).
- Umerle, Tomasz: Fluid Borders of ‘Literature’: Literary Theory versus the Practice of Literary Documentation, In: *(e)motion Cultural Literacy in Europe: Second Biennial Conference Book of Abstracts*, 2017, <http://cleurope.eu/wp-content/uploads/2017/02/CLE-2017-Book-of-Abstracts-final.pdf>
- Zeileis, Achim, *ineq: Measuring Inequality, Concentration, and Poverty*, R package version 0.2-13, 2014. <https://CRAN.R-project.org/package=ineq>.