



TOPIC MODELLING: A QUANTITATIVE RESEARCH TECHNIQUE FOR THE HISTORY OF IDEAS

Anastazja Zakusilo

Regional Centre for Social Policy in Krakow
zakusilo.anastasia@gmail.com

ABSTRACT

The history of ideas emerged as a distinct discipline in the 1940s under the leadership of Arthur Lovejoy, whose foundational principles emphasized meticulous qualitative analysis and the interpretation of selected thinkers' works. However, the increasing volume of authors and publications in recent decades has introduced significant challenges to the field. This article argues that incorporating quantitative methods can help address these challenges. Specifically, it explores the potential of topic modelling as a tool for studying ideas, highlighting its advantages and providing an example of its application through an analysis of sociological articles published in two leading journals: the *American Journal of Sociology* in the United States and *Studia Socjologiczne* in Poland.

KEYWORDS: Methodology of the history of ideas, Sociology of knowledge, Quantitative methods, Topic modelling

1. INTRODUCTION

The history of ideas as an academic discipline emerged in the 1940s with Arthur Lovejoy's article "Reflections on the History of Ideas" (1940), which introduced the field and established the *Journal of the History of Ideas*, dedicated to publishing works in this area. Although studies of intellectual thought were not new, his article helped consolidate the history of ideas as a distinct area of study with its own subject matter, scope, and methodology.

As the father of the discipline, Lovejoy outlined its purpose as an exploration of ideas and the transformation they undergo over time, all while going beyond existing specializations and integrating available knowledge from all

the fields of humanities and social sciences. Lovejoy pointed out that understanding why certain ideas are interesting and important to a person or a group of people should be carried out taking into account both the logical and illogical sides of reasoning (such as psychological factors including personal relationships between authors, and social factors, such as their class position).

Taking into account the foundations laid out by Lovejoy, as well as the way studies in the humanities were conducted in the past, works in the history of ideas primarily focused on the thoughts of a few important and widely recognized authors. These works were studied using historical, biographical, and interpretative approaches. However, recent changes in academic practices over the past few decades, along with technological advancements, have opened the door to new approaches in the study of ideas. Quantitative methods now offer the potential to complement and enrich the qualitative methodologies that have long been dominant in the field.

This article offers a fresh perspective on how quantitative methods – particularly topic modelling – can be applied to the study of ideas, providing an example of its implementation. It begins by exploring key principles and theoretical reflections from influential scholars in the history of ideas, then outlines the potential benefits of using topic modelling in this field.

2. THE FUNDAMENTAL PRINCIPLES OF THE STUDY OF IDEAS AND THEIR CONTEMPORARY SIGNIFICANCE

Lovejoy emphasized the need for the separate and independent field of study, which would act as a bridge between other disciplines that study intellectual thought and the realm of ideas. The purpose of the history of ideas as a new discipline would be to “traverse the customary boundary-lines” (Lovejoy 1940, 7) of existing disciplines, synthesize the knowledge and provide tools for a comprehensive study of ideas present within and across such disciplines like philosophy, sociology, psychology, history, but also art and literature.

Lovejoy proposed several concepts and principles of the study of ideas, as well as some methodological and theoretical insights worth considering when studying intellectual history.

To begin with, he introduced the concept of unit-ideas as a tool for studying ideas. Unit-ideas are fundamental, basic elements that can be combined in various ways to form more complex aggregates. As he pointed out, “most philosophic systems are original or distinctive rather in their patterns than in their components.” (Lovejoy 2001, 3) He observed that, although many philosophical doctrines may seem to have little in common, they can be 'broken up' into the same compounds, or unit-ideas. As Lovejoy (2001, 4) explains, “[t]he seeming novelty of many a system is due solely to the novelty of the applica-

tion or arrangement of the old elements which enter into it.” The concept of unit-ideas is, thus, an alternative way to study relations between humanistic and social doctrines, which, in contrast to the traditional classification by schools of thought, serves to reduce the complexity and multiplicity of ideas. Instead, it provides an opportunity to identify fundamental connections between different concepts, which may initially appear distinct but are based on the same underlying ideas.

Lovejoy also emphasized the importance of both synchronic and diachronic analysis. He argued that it is crucial to explore both dimensions of intellectual history: the synchronic dimension, which looks at how ideas are conceptualized by thinkers within the same time period, and the diachronic dimension, which traces the evolution of ideas over time. As he put it,

It is in the persistent dynamic factors, the ideas that produce effects in the history of thought, that he [a historian of ideas] is especially interested. (...) And motives and reasons partly identical may contribute to the production of very diverse conclusions, and the same substantive conclusions may, at different periods or in different minds, be generated by entirely distinct logical or other motives. (Lovejoy 2001, 5)

As a third point, Lovejoy accentuated the importance of interdisciplinary approach. Since ideas tend to appear in different areas of intellectual thought, the goal of the historian of ideas should be to cross the boundaries of an existing discipline; those which are not usually crossed by their representatives – philosophers, historians, sociologists or anthropologists. Lovejoy outlines the risks associated with compartmentalisation of historical research:

In so far as the endeavor to trace such relations [between events] stops at the boundaries of one or another of these divisions, there is always a high probability that some of the most significant – that is, the most illuminating and explanatory – relations will be missed. It has even sometimes happened that a conception of major historic influence and importance has long gone unrecognized, because its various manifestations, the parts which make up the whole story, are so widely dispersed among different fields of historical study, that no specialist in any one of these fields became distinctly aware of it at all. (Lovejoy 1940, 4)

Finally, the father of the history of ideas highlighted that to study ideas in all their complexity, contextual knowledge of social and cultural milieu, as well as an author’s biography, is indispensable. It is therefore important to study not only author’s ideas, but also the broader context (his social position, nationality, crucial life experiences, historical age and dominant ideas of the time etc.). For example, when it comes to understanding of art, “[s]uch

knowledge adds what may be called a new dimension to a work of art, the dramatic dimension” (Lovejoy 1940, 12-13). He also provides an example which emphasizes the degree of its importance:

How meager would be the aesthetic content of the *Divine Comedy* as a whole, or of most of its parts, to a modern reader – especially a non-Catholic reader – wholly ignorant of medieval ideas and feelings and pieties, or incapable, while reading it, of making these in some degree his own, by an effort of the imagination. (Lovejoy 1940, 12)

This fragment demonstrates that Quentin Skinner's criticism of Lovejoy for conceiving a decontextualized history of ideas is, at best, unfair (Mouton 2023). However, it is true that Skinner delved deeper into the matter. He argued that understanding a text requires not only situating it within its social context but also analysing its linguistic and rhetorical dimensions. This involves uncovering the author's intentions and examining the illocutionary force of their statements – that is, the intended purpose or effect of their utterances, as conveyed through language (Skinner 1969). This method, known as ‘contextualism,’ emphasizes interpreting texts historically, taking into account the language, conventions, and purposes in which they were created.

In particular, Skinner warns against the imposition of contemporary meanings and interpretations or, as he called them, ‘anachronistic mythologies,’ onto historical texts. He underlined the need for an analysis that respects the historical specificity of the ideas under study. For example, in discussing the notion of the government of consent in Locke's writings, Skinner notes that there is a danger of misinterpreting the idea as intended by the author if the researcher applies modern framework to understand it:

When *we* speak of government by consent we usually have in mind a theory concerned with the best organization of government. It is thus natural, or rather fatally easy, to turn with this conceptualization in mind to Locke's text, and duly to find some such theory rather bunglingly set out there. There is decisive evidence, however, that when Locke spoke of government by consent this simply did not happen to be what *he* had in mind at all. It is now clear that Locke's concern with the concept of consent arises solely in connection with the *origin* of legitimate societies. This is hardly what we should regard as an argument for consent, but it happens to be Locke's argument, and the only result of failing to start from this point is to misdescribe the whole theory, and so to accuse Locke of having bungled an account which he was not, in fact, trying to write. (Skinner 1969, 28)

Yet another scholar who provided important insights into the scope and methodology of the history of ideas was Jacques Le Goff, a French historian

who aimed at understanding the cultural structures of past societies, focusing particularly on the Middle Ages. He was particularly interested in examining how long-term social, economic, and cultural forces shaped historical developments, moving beyond the traditional focus on specific events or individuals. This approach aligned closely with the principles of the Annales School of historiography, where Le Goff was a prominent figure, which emphasized the study of long-term historical structures over individual events (Le Goff 1996, xiii).

The key object of study by Le Goff were mentalities, that is collective ways of thinking. He studied interactions between mentalities and social structures, examining how broad cultural patterns and social practices influenced and were influenced by ideas over extended periods of time. The effects of his research can be found in such books as *History and Memory* (1996), *Time, Work and Culture in the Middle Ages* (1980), and *The Medieval Imagination* (1992).

Le Goff's main focus is on widespread ways of thinking, and his approach represents a perspective alternative to the study of the intellectual activity of a narrow elite. His works represent the trend of moving toward the study of popular ideas, shared by large groups of people. This approach was, to some extent, shared by Lovejoy, who stated that:

the study of the history of ideas (...) is especially concerned with the manifestations of specific unit-ideas in the collective thought of large groups of persons, not merely in the doctrines or opinions of a small number of profound thinkers or eminent writers. (...) It is, in short, most interested in ideas which attain a wide diffusion, which become a part of the stock of many minds. (Lovejoy 2001, 19)

Precisely to measure the greater or lesser diffusion of certain terms and concepts, as well as their variation over time in both absolute and relative frequency, Riccardo Campa has applied scientometric tools to research in the history of ideas. While not abandoning qualitative analysis, he has quantified not only the statistical presence of certain ideas in scientific literature but also the co-presence and theoretical relationships between concepts (Campa 2019, 2020, 2021). Moreover, Campa (2014) has raised the issue of expertise or disciplinary competence in this field of investigation. The inherent interdisciplinarity of the history of ideas directly impacts this challenge. Tracing the trajectory of an idea across various disciplines requires considerable erudition and a different kind of expertise than that of a hyperspecialized expert. The specific qualifications and competence of the historian of ideas have been examined in detail by multiple authors in two volumes edited by Michel Kowalewicz (2013, 2014).

To summarize the argument above, several key principles have been highlighted by the most prominent historians of ideas. These include interdisciplinarity, contextualism, the integration of synchronic and diachronic approaches, and a focus on ideas that became widespread and had a tangible historical impact. When seeking to implement these principles in contemporary research within the field of the history of ideas, it is crucial to consider the unique characteristics of modern academic work. The past few decades have witnessed profound changes in how intellectual activities are conducted, particularly within academia, as well as in the publication and dissemination of research outputs. These developments pose new challenges for the history of ideas, especially in managing and processing the ever-growing volume of academic publications. The following section of this article will propose a solution for addressing these challenges while adhering to the aforementioned principles for studying ideas.

3. A QUANTITATIVE RESEARCH TECHNIQUE FOR THE HISTORY OF IDEAS

There exist a range of methods and techniques of quantitative text analysis. The following section presents one of these methods, focusing specifically on its utility in the field of the history of ideas, namely topic modelling. This method of quantitative text research is presented along with the benefits it provides for the history of ideas, as exemplified by research on diversity and dynamics of topics present in two important sociological journals based in two different countries: the United States of America and Poland.

3.1. Methodology of the research

The journals taken into consideration in the mentioned study are the *American Journal of Sociology* (USA) and *Studia Socjologiczne*, which is translated into English as *Sociological Studies* (Poland). Both journals are considered among the most important sociological journals in their respective countries.

The units of analysis were the abstracts of 1703 articles published over the course of 30 years (1993-2022). Among them, 988 were published in the *American Journal of Sociology* and 714 in *Studia Socjologiczne*. Apart from abstracts, other data characterizing articles were collected, such as the year of publication, gender of authors, their country and institution of affiliation and number of authors for each article¹.

¹ The abstracts and other data on the articles were collected using the program ParseHub for web scraping.

The topic analysis was conducted on abstracts in English in order to ensure the consistency and comparability of results. Almost all abstracts were available in English. Some abstracts that were only available in Polish (5% of all abstracts from *Studia Socjologiczne*) have been translated into English before the analysis. Additionally, the study of topics was complemented by the analysis of topic distributions between gender groups and the two journals.

Over the course of the analysis, three different models were created:

1. a complete model based on all the texts in the database (1703 articles);
2. a partial model based on texts published only in the *American Journal of Sociology* (988 articles);
3. a partial model based on texts published only in *Studia Socjologiczne* (714 articles).

Topic modelling was conducted in the program R using the *stm* package² developed by Margaret E. Roberts, Brandon E. Stewart and Dustin Tingley (2019a, 2019b). This package offers the possibility to perform the analysis using a generative probabilistic LDA (Latent Dirichlet Allocation) algorithm, which can be used as a means of reducing the dimensionality in a set of documents³. This algorithm assumes that individual documents are represented with a certain probability by a number of topics, and topics are represented with a certain probability by a number of words (Blei, Ng, Jordan 2003).

The results of the analysis include:

1. The θ /*theta* parameter, that is the distribution of topics within each document. This parameter indicates the extent to which specific documents contain different topics.
2. The β /*beta* parameter, that is the distribution of words within topics. This parameter indicates which words are associated with each topic, i.e. how topics are represented by specific words.

Thus, for example, if for a given text the topic 'race' has a value of $\theta=0.25$, then according to the algorithm 25% of its content is related to this topic. The θ values help understand which topics dominate a given text or a set of texts and allow for the identification of the main themes in a series of documents (the examples of topic assignment are presented in Table 1).

² Other packages supporting data processing were used: *tidyverse* (Wickham et al. 2019) and *quanteda* (Benoit et al. 2018).

³ In addition, the package provides the possibility to select the algorithm used for parameter estimation in the initial training phase. For the analysis of data in subsets containing articles from each journal separately, the method of moments was used. This method allowed for a more unambiguous classification of documents than Gibbs sampling. Analysis on merged data containing all articles was performed using the LDA algorithm, which is less likely to provide unambiguous classification, but proved more effective in terms of assigning topics to texts for more diverse data.

Table 1. Four examples of topic attribution based on article title and abstract (complete model)

Title and abstract	Topic assignment
Making Markets on the Margins: Housing Finance Agencies and the Racial Politics of Credit Expansion Widespread reliance on credit increasingly defines realities of economic citizenship in American society. This article theorizes the racial politics of credit expansion. It examines the federal initiative in the 1960s and '70s to broaden financial access for poor renters in communities of color, which unintentionally sparked the rise of new state-level credit agencies. Drawing on historical evidence, much of it never used before, the author's findings reveal the contentious politics at the heart of this policy shift. Doing so highlights the constitutive whiteness of credit and also illuminates how the project of expanding credit to marginalized groups tests the categorical seams of markets in the public imagination: such initiatives fuel racial contestation around taken-for-granted market rules, which draws governing officials toward increasingly speculative and convoluted financial instruments as a means of rule-bending subversion. Ultimately, this article sheds much-needed light on, and encourages further research into, the racial stratification of the state's market-making power. (Robinson 2020)	Topic 1 economy, business $\beta=0.79$
	Topic 2 race $\beta=0.1$
Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined According to recent research, ethnoracial diversity negatively affects trust and social capital. This article challenges the current conception and measurement of "diversity" and invites scholars to rethink "social capital" in complex societies. It reproduces the analysis of Putnam and shows that the association between diversity and self-reported trust is a compositional artifact attributable to residential sorting: nonwhites report lower trust and are overrepresented in heterogeneous communities. The association between diversity and trust is better explained by differences between communities and their residents in terms of race/ethnicity, residential stability, and economic conditions; these classic indicators of inequality, not diversity, strongly and consistently predict self-reported trust. Diversity indexes also obscure the distinction between in-group and out-group contact. For whites, heterogeneity means more out-group neighbors; for nonwhites, heterogeneity means more in-group neighbors. Therefore, separate analyses were conducted by ethnoracial groups. Only for whites does living among out-group members – not in diverse communities per se – negatively predict trust. (Abascal, Baldassarri 2015)	Topic 1 race $\beta=0.66$
	Topic 2 social networks $\beta=0.24$
Social Cohesion in Decentralized Ukraine: From Old Practices to New Order The article presents a systematic study of social cohesion phenomenon at the level of amalgamated hromadas as a key local entity of decentralization reform in Ukraine. Building on the analysis of the 26 semi-structured interviews conducted in amalgamated hromadas of two border regions of Ukraine – Kharkiv and Chernivtsi, the author has identified social cohesion components, their interconnection as well as posi-	Topic 1 political system, transformation $\beta=0.76$
	Topic 2 regional identity $\beta=0.2$

Title and abstract	Topic assignment
<p>tive and negative factors of social cohesion strengthening at community level. Relying on Chan's empirical model and perceived perspective of social cohesion, hromada amalgamation is conceptualized as a transformation process of avoiding 'old practices' to form 'new order'. In the process, the establishing of democratic tools, local activist growth, reducing gaps between center and periphery, formation of common sociocultural space are emphasized. Strengthening social cohesion components at the hromada level are stated to become a sure basis for 'a giant leap' of Ukraine's democratisation in the nearest future. (Deineko 2021)</p>	
<p>The Right-Wing Attitudes of Young Poles. The Context of the 2015 Parliamentary Elections</p> <p>In the article, we analyze political preferences of the young Polish voters who associate themselves with the right side of the political scene. We are interested in the views, attitudes and assessments of socio-political situation before and after parliamentary elections in 2015 that are shared by the young Poles declaring sympathy to the political right or voting for the right-wing parties. What exactly are the right-wing attitudes of the young Polish generation and how do they relate to the definitions of the political right assumed in political science (i.e. as conservative moral views and liberal economic views)? In the article, we propose a theoretical conceptualization of the political right, and then confront it with its popular understanding in Poland. In particular, we focus on the right-wing attitudes of young people, analyze it by reference to the existing data and to the results of our own research carried out in 2016, a year after the last Polish presidential and parliamentary elections. We analyze the electoral behaviours and motivations of the surveyed group, its views on the economy, social issues, the state, collective identity, attitudes to the actions of the political opposition and the independence of the Constitutional Tribunal. The results of the analyses demonstrate that many young voters who voted for the right had formerly been undecided voters seeking political identification. Above all, however, the findings of our research provide the characteristic of the right-wing attitudes of young people, pointing at their liquid, incoherent and exploratory character, possible political turnovers and significance. (Szafraniec, Grygień 2019)</p>	<p>Topic 1</p> <p>national and ethnic identity, minorities $\beta=0.58$</p> <p>Topic 2</p> <p>politics, voting $\beta=0.39$</p>

In a similar way, for each topic there is a probability distribution of words (β), which allows us to identify which words are most characteristic of specific topics on the basis of the documents analysed. That is, if the word 'racial' has a value of $\beta=0.02$ for the topic 'race', this means that 2% of that topic is related to this word. Due to the large number of words, the highest values of β amount to just a few per cent.⁴

⁴ The probability distributions of both θ and β add up to unity, which is the condition for the correctness of the probability distribution in probabilistic models.

3.2. *The benefits of quantitative analysis: exemplification*

The following section outlines three key benefits of applying quantitative methods, such as statistical analysis and topic modelling, to the study of ideas. These benefits include the ability to broaden research to cover large datasets spanning diverse temporal and social contexts, the opportunity to examine additional important factors (such as the characteristics of authors and texts), and the capacity to distil and represent the core essence of ideas and their compounds.

1. Amplification of research to include more extensive temporal and geographical range

One of the most manifest advantages of quantitative research techniques that involve automatised data collection, transformation, and processing, is the potential to analyse large sets of data.

Implementing quantitative techniques into the study of the history of ideas make it possible to expand the research to include many authors and extensive historical periods instead of focusing in detail on the thought of few specific authors. This in turn offers new possibilities to conduct synchronic and diachronic analyses, and combine these approaches.

I would like to outline several new opportunities related to the use of topic modelling in the study of ideas. First of all, this technique can serve as a method to identify the main ideas discussed in a set of texts, as well as classify them based on a selected characteristic (e.g. source, like in the presented example; yet other criteria can be used, like author's gender, discipline and school of thought). The classification may then be used to identify main ideas discussed by authors belonging to each of the distinguished groups. As Table 2 shows, some topics appear predominantly in one of the two journals, while others are discussed in both of them⁵.

For example, topics such as race, gender and related inequalities, as well as spatial segregation are often discussed by authors publishing in the *American Journal of Sociology*, while they almost do not appear in texts published in *Studia Socjologiczne*. This may be due to specific social problems which are relevant in the USA (but not so much in Poland) that encourage American authors to address these topics. Issues of race, racial violence, police brutality against people of certain races, or spatial segregation of neighbourhoods re-

⁵ The Duncan dissimilarity index can be used to measure the degree of topic diversification between journals. Here, the index indicates the percentage of articles that would need to be reclassified under different topics to equalize the topic distribution across both journals. Here, $D=0.47$, indicating that nearly half of the Polish (or American) texts would need to be reassigned to different topics for the thematic structures of the two journals to become identical.

lated to race do not occur in Poland, and therefore are not of interest to Polish sociologists, which much more often discuss ethnic and regional, rather than racial identities. Other topics present predominantly in Polish sociological debate are the public sphere and public discourses, as well as politics, specifically democracy and elections. One of the possible hypotheses explaining this phenomenon may be the need to develop new organizational and social foundations for the new political system following Poland's political and economic transformation after the dissolution of the Polish People's Republic in 1990s.

Table 2. Distribution of topics within two analysed journals and the difference in each topic's share between them (complete model)

Topic	<i>American Journal of Sociology</i>	<i>Studia Socjologiczne</i>	Difference (StS – AJS)
race	9.53	0.23	-9.29
organizations	8.88	0.61	-8.27
gender, occupational inequalities	7.5	1.29	-6.21
housing and neighbourhoods	5.38	1.61	-3.78
state, policy, welfare	5.4	1.86	-3.53
economy, business	5.52	2.16	-3.36
migration	4.19	1.1	-3.08
social capital and networks	6.14	3.2	-2.94
school	4.69	1.89	-2.8
mobility, education, occupations	6.46	4.57	-1.89
labour market	4.3	3.09	-1.21
political and economic systems	3.61	3.08	-0.53
law, rights	3.67	3.91	0.24
cultural movements and practices	3.58	4.28	0.7
global society	3.46	4.34	0.87
class culture	2.51	4.12	1.61
family	3.83	5.44	1.61
power, institutions	1.96	4.37	2.41
national and ethnic identity, minorities	2.23	5.14	2.91
public discourse, media	1.96	5.19	3.23
voting, politics	2.36	8.2	5.84
regional communities and identity	0.76	8.27	7.52
culture, ethics	0.65	8.94	8.29
sociology and social theory	1.43	13.11	11.68
Sum (%)	100	100	
Number of articles	988	714	

Meaning of abbreviations: AJS – *American Journal of Sociology*, StS – *Studia Socjologiczne*.

Another important approach to the study of ideas is a diachronic analysis, which involves the change in their presence and spread over time. The change in problems which arouse interest among authors who represent various fields of social science, humanities and art may serve as a signal of ongoing and forthcoming social change. To trace such evolution in sociological thought in the USA, the analysed articles were divided into three groups, corresponding to three 10-year periods between 1993 and 2022. Table 3 presents the distribution of topics' share within each of those periods, showing changes in research interests among the authors.

Table 3. Share of topics in articles published in the *American Journal of Sociology* by publication period and its change over time (complete model)

Topics	The share of the topic in the articles			Change (Period 3- Period 1)
	Period 1 (1993-2002)	Period 2 (2003-2012)	Period 3 (2013-2022)	
race	7.05	9.65	11.77	4.72
school	3.61	3.79	6.64	3.03
state, policy, welfare	2.97	7.13	5.94	2.97
housing and neighbourhoods	3.69	6.18	6.18	2.48
national and ethnic identity, minorities	1.28	2.46	2.91	1.63
power, institutions	1.42	1.51	2.95	1.52
labour market	3.98	3.7	5.23	1.25
gender, occupational inequalities	7.14	7.29	8.06	0.92
voting, politics	2.31	2.21	2.58	0.27
economy, business	6	4.76	5.85	-0.16
migration	4.25	4.37	3.94	-0.31
regional communities and identity	1.12	0.49	0.68	-0.43
organizations	8.82	9.49	8.3	-0.53
law, rights	3.54	4.43	3.01	-0.54
culture, ethics	0.86	0.83	0.27	-0.59
class culture	3.48	1.76	2.36	-1.12
mobility, education, occupations	7.26	6.25	5.91	-1.35
sociology and social theory	2.18	1.33	0.81	-1.37
family	4.73	3.57	3.24	-1.5
global society	4.46	3.29	2.69	-1.76
political and economic systems	4.64	3.43	2.8	-1.84
public discourse, media	3.15	1.71	1.08	-2.08
cultural movements and practices	4.99	3.22	2.59	-2.4
social capital and networks	7.04	7.16	4.23	-2.82
Sum (%)	100	100	100	
Number of articles	317	341	330	

As shown in the table, several problems related to inequalities and power have become more prominent, including race and gender inequalities, as well as spatial segregation in areas like poorer and wealthier neighbourhoods, education, and the labour market. While these topics already had a significant share in sociological discourse in the USA from 1993 to 2022, being the main topic of 25% of articles, they gained even more attention, emerging as the primary focus in 38% of articles published in the *American Journal of Sociology* between 2013 and 2022. On the other hand, topics related to public discourse, political systems, cultural practices and movements, as well as social networks and social capital, have seen a decline in importance, with their share decreasing from 20% in 1993-2002 to 11% in 2013-2022.

The above examples show that one benefit provided by quantitative topic modelling is a more general analysis of texts, which allows one to determine the most important ideas pondered by authors living in different times or coming from different social, cultural and disciplinary backgrounds.

2. *A new approach to the study of the compounds of ideas*

Topic modelling provides the opportunity to explore the content of ideas and the meaning attributed to them by the authors. What is important is that the technique may prove useful even if the idea in question is not the main focus of the analysed works. Therefore, it offers a glimpse into latent, not fully conscious or intended associations.

One of the results of topic modelling is the distribution of words within topics which indicates what terms most often appear close to each other in the text, thus indicating the relation of proximity between them. Table 4 presents two examples of such decomposition of ideas into words they are associated with in the analysed journals.

For example, if we consider the idea of a woman in the analysed sociological works in Poland and in the USA, the first thing that comes to our attention is that the context in which we find references to women is richer in the *American Journal of Sociology*. In this journal they appear in such contexts as gender differences on the labour market, family life, marriage and parenting, as well as issues (also related to mental health) associated with the role of women in the household, such as housework and care. In the Polish context the notion of women mostly appears in contexts related to family life, children or social change affecting family as an institution. Thus, the same idea of a woman is presented differently in Polish and American sociological discourse, with the latter bringing forward more social contexts of their life.

Table 4. Topics and words that constitute them in the *American Journal of Sociology* and *Studia Socjologiczne* (comparison of partial models)

<i>American Journal of Sociology</i>			<i>Studia Socjologiczne</i>		
Women					
gender	family	work, health and family	family		
gender	family	work	family		
women	children	health	social		
men	social	time	policy		
job	families	effect	solidarity		
labor	effects	effects	families		
employment	parents	mental	life		
women's	life	social	women		
differences	marriage	care	relationships		
social	parental	labor	public		
career	status	wives	changes		
female	age	housework	children		
characteristics	intergenerational	marital	context		
Politics					
political organizations	political movements	organizations and law	voting and political participation	politics, democracy and Internet	social structure, class movements
political	movement	organizational	electoral	political	social
corporate	social	law	voting	internet	class
interests	rights	legal	elections	social	movement
state	movements	theory	theory	democracy	structure
organizations	protest	organizations	models	system	human
groups	state	change	participation	public	capital
professional	mobilization	institutional	social	democratic	religion
organizational	political	effects	model	party	activity
power	action	political	political	legitimacy	well
institutional	collective	state	voter	society	cultural
firms	policy	social	turnout	attitudes	relations
market	organizations	sentencing	local	part	movements

Only the words with the highest β parameter are presented for each topic.

The second area that can be considered is politics, which both journals explore in various contexts. However, some differences can be traced. In the *American Journal of Sociology*, politics is viewed through institutional, economic, and behavioural lenses, such as in relation to organizations (corporate interests, professional groups), political activity (movements, protests), and the law, including the regulatory role of the state. In contrast, *Studia Socjo-*

logiczne discusses politics as an institutional complex focused on elections and democratic political systems, while exploring its ideological dimension (political attitudes and legitimacy), along with topics such as social structure, class movements, social capital, religion, and online political activity.

3. Exploration of the influence of different factors other than time and space (e.g. authors' gender)

Other than the expansion of the analysis on larger sets of texts coming from different époques, cultural and disciplinary backgrounds, quantitative approach makes it possible to statistically explore the interactions between ideas and different characteristics of texts or authors. It allows to determine such interconnectivities more broadly, which is almost impossible in qualitative studies.

For example, the topics and ideas that authors focus on may vary depending on their gender. In the articles published in the *American Journal of Sociology*, some topics appeared to be gender-neutral, while others were more commonly addressed by authors of a specific gender. As shown in Table 5, female authors tended to focus more on issues related to women, particularly in the contexts of family, care, and health – subjects that may be influenced by personal experience.

Table 5. Gender of authors and the topics they discuss in the *American Journal of Sociology* (%)

Topics	Gender of author(s)			Difference		
	F	(FM)	M	F-(FM)	F-M	(FM)-M
work, health, and family	9.49	5.12	5.26	4.37	4.23	-0.14
organizations and law	9.79	5.54	7.09	4.26	2.71	-1.55
political movements	10.81	7.49	5.90	3.32	4.90	1.58
family	7.08	7.87	5.09	-0.80	1.99	2.79
gender	9.24	10.84	4.75	-1.61	4.49	6.10
social networks	3.45	9.08	11.01	-5.63	-7.55	-1.93
labour market	1.58	5.02	5.24	-3.44	-3.66	-0.22
race	5.71	5.29	7.79	0.42	-2.08	-2.50
economic status	5.83	5.38	7.14	0.45	-1.31	-1.76
class culture	5.17	3.83	6.84	1.33	-1.68	-3.01

Meaning of abbreviations: F – female author(s); M – male author(s); FM – mixed gender composition of authors' team (both women and men).

On the other hand, male authors were more likely to engage with topics such as class culture, economic status, and the labour market, as well as social networks. Interestingly, these latter two areas of study were often tackled by

male or mixed-gender teams of authors, which could suggest that these works were more likely to be collective efforts arising from larger studies, such as surveys or experiments. The predominance of male authors in these areas may also be partly explained by the fact that male researchers tend to rely more on quantitative methods, which are common in such studies (Thelwall et al. 2019).

Apart from the application illustrated by the example, the characteristics of texts and authors can be leveraged in various ways at different stages of analysis to gain deeper insights into their content and structure. For instance, categorical variables (with two or few categories) could be used similarly to journals in the presented analysis, dividing the set of analysed texts into subsets before estimating models to examine topics and their components within each group. Another possibility is to use continuous variables (with many quantitative categories, such as different measures of spread and influence, including the number of citations or views) to compare their means or medians across different topics or groups, such as journals, gender groups, or nationalities. These variables can also be combined, allowing researchers to visualize relationships between them on scatterplots. Expanding on this, future studies could integrate advanced statistical techniques or machine learning models to identify hidden patterns and relationships, further enriching the analysis of texts and authors. Such approaches could open new avenues for understanding the dynamics of ideas and their dissemination across various contexts.

4. CONCLUSION

The presented quantitative technique proposed for application to the history of ideas constitutes a response to modern challenges posed by the way modern academic system functions (specifically concerning an increasing number of scholars and publications). This approach allows for a shift from the study of one or a few chosen authors to many authors, and may be used to identify major and minor trends in intellectual thought.

The proposed technique of topic modelling, and quantitative methods in general, cannot substitute a more profound study of ideas with the use of qualitative, interpretative and understanding approach, but rather can serve to add new value to the traditional technique by expanding the area of the research. There is still a need for a deeper understanding of an investigated topic by the researcher, which is essential on the stage of planning and design of the research, as well as for understanding the obtained results.

What is more, technical challenges and dilemmas should be taken into account when using a quantitative approach to text analysis. Some of the problems are the need of linguistic unification of texts before the analysis, the mul-

tiplicity of word forms (especially in synthetic languages), as well as the inevitable simplification of the material and the loss of context, which affects the accuracy of the results. In addition, the outcome of the research depends a lot on the initial parameters and the multiple decisions taken by the researcher during the process (e.g. concerning such questions as what time periods to adapt, how to select the material, how many topics to choose, which method to use, whether to unify the language or different subsets or work on original material). This being said, quantitative study of ideas requires craftsmanship and skill, making indispensable also theoretical and qualitative in-depth research.

BIBLIOGRAPHY

- Abascal, M., Baldassarri, D. 2015. "Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined." *American Journal of Sociology* 121 (3): 722-782.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A. 2018. "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software* 3 (30).
- Blei, D. M., Ng, A. Y., Jordan, M. I. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993-1022.
- Campa, R. 2014. "Il problema della competenza disciplinare nella storia delle idee." In *Formen der Ideengeschichte*, edited by Michel H. Kowalewicz. Münster: mentis Verlag.
- Campa, R. 2019. "Technological Unemployment and Universal Basic Income: A Scientometric Analysis." *Sociologies in Dialogue. Journal of the Brazilian Sociological Society* 5 (1): 57-85.
- Campa, R. 2020. "Fourth Industrial Revolution and Emotional Intelligence: A Conceptual and Scientometric Analysis." *Changing Societies & Personalities* 4 (1): 8-30.
- Campa, R. 2021. "The Sociology of Global Warming: A Scientometric Look." *Studia Humana* 10 (1): 1-17.
- Deineko, O. 2021. "Social Cohesion in Decentralized Ukraine: From Old Practices to New Order." *Studia Socjologiczne* 1 (240): 117-38.
- Kowalewicz, M. H., ed. 2013. *Spiel. Facetten seiner Ideengeschichte*. Münster: mentis Verlag.
- Kowalewicz, M. H., ed. 2014. *Formen der Ideengeschichte*. Münster: mentis Verlag.
- Le Goff, J. 1980. *Time, Work and Culture in the Middle Ages*. Chicago and London: University of Chicago Press.
- Le Goff, J. 1992. *The Medieval Imagination*. Chicago and London: University of Chicago Press.
- Le Goff, J. 1996. *History and Memory*. New York: Columbia University Press.

- Lovejoy, A. O. 1940. "Reflections on the History of Ideas." *Journal of the History of Ideas* 1 (1): 3-23.
- Lovejoy, A. O. 2001. *The Great Chain of Being. A Study of the History of an Idea*. Cambridge, Massachusetts, and London: Harvard University Press.
- Mouton, N. 2023. "An apologia for Arthur Lovejoy's long-range approach to the history of ideas." *History and Theory* 62: 272-95. <https://doi.org/10.1111/hith.12298>.
- Roberts, M. E., Stewart, B. M., Tingley, D. 2019a. "stm: R Package for Structural Topic Models." *Journal of Statistical Software* 91 (2): 1-40. <https://www.jstatsoft.org/index.php/jss/article/view/v091i02/1319>.
- Roberts, M. E., Stewart, B. M., Tingley, D. 2019b. "stm: R Package for Structural Topic Models." R package version 1.3.4. DOI: 10.32614/CRAN.package.stm.
- Robinson, J. N. 2020. "Making Markets on the Margins: Housing Finance Agencies and the Racial Politics of Credit Expansion." *American Journal of Sociology* 125 (4): 974-1029.
- Skinner, Q. 1969. "Meaning and Understanding in the History of Ideas." *History and Theory* 8 (1): 3-53.
- Szafraniec, K., Grygień, J. 2019. „Prawicowość młodych Polaków. Kontekst wyborów parlamentarnych z 2015 roku.” *Studia Socjologiczne* 2 (233): 2545-2770.
- Thelwall, M., Bailey, C., Tobin, C., Bradshaw, N. 2019. "Gender differences in research areas, methods and topics: Can people and thing orientations explain the results?" *Journal of Informetrics* 13 (1): 149-169.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43).