

# The Language of Compliance, The Practice of Resistance: A Data-Driven Study of Declarative Ideological Terminology in Slovak Ethnographic Research (1948–1973)

ANDREJ GOGORA



DOI: <https://doi.org/10.31577/SN.2025.2.21> © Ústav etnológie a sociálnej antropológie SAV, v. v. i. © 2025, Andrej Gogora. This is an open access licensed under the Creative Commons

*Andrej Gogora, Institute of Ethnology and Social Anthropology of Slovak Academy of Sciences, Klemensova 19, 813 64, Bratislava; e-mail: andrej.gogora@savba.sk*

The study examines the integration of Marxist-Leninist ideology into Slovak ethnographic fieldwork conducted between 1948 and 1973, focusing on the declarative use of ideologically charged terms. It analyzes descriptive metadata, primarily thematic subject headings, of the *Collection of Research Reports* housed at the Institute of Ethnology and Social Anthropology Slovak Academy of Sciences. Advanced statistical methods, including the chi-squared test, frequency analysis, co-occurrence analysis, and TF-IDF, are employed to identify the use of declarative terms within the documentation of ethnographic research reports from the period. The study aims to quantitatively assess how ethnographers responded to the political pressures of the socialist era and how this response is reflected in their research reports and metadata entries. The findings contribute to an understanding of the complex interplay between ethnography and political ideology in socialist-era Slovakia.

*Keywords:* Slovak ethnography, socialist era, ideological influence, declarative term, descriptive metadata, statistical analysis, digital humanities

*How to cite:* Gogora, A. (2025). The Language of Compliance, The Practice of Resistance: A Data-Driven Study of Declarative Ideological Terminology in Slovak Ethnographic Research (1948–1973). *Slovenský národopis* 73(2), 247–263.

DOI: <https://doi.org/10.31577/SN.2025.2.21>

## Introduction

The article focuses on the issue of the integration of political ideology into ethnography in Slovakia in the post-war period. This discussion pertains exclusively to the Soviet interpretation of Marxism, specifically Marxism-Leninism. This ideology

represents a distinct interpretation and practical application of Marxist theories, adopted in Eastern Bloc countries after World War II. Marxism-Leninism is characterized as a pragmatic and action-oriented ideology that dogmatized classic Marxist tenets. Its key features include, behind the veil of historical materialism and as the politics of the empty signifier (Walker, 1989), an emphasis on the autocratic role of a centralized Communist Party, dictatorial socialist state power, the suppression of any opposition, and comprehensive control over all aspects of life (Arnason, 1982). In this local context, the study specifically addresses the phenomenon of formal declarations of adherence to Marxist-Leninist ideology by scholars, within the context of ethnographic fieldwork and research reports collected at the former Ethnographic Institute Slovak Academy of Sciences (currently the Institute of Ethnology and Social Anthropology Slovak Academy of Sciences – hereinafter IESA SAS). This represents a linguistic strategy employed in academic works, characterized by the declarative and pragmatic utilization of **Marxist-Leninist terminology**,<sup>1</sup> frequently through the inclusion of thematically irrelevant passages or unrelated insertions. Within the context of academic outputs, the obligatory use of Marxist-Leninist terminology served as a pervasive utilization of theoretical frameworks to interpret social, and cultural phenomena, as well as a means of ideological legitimation for the prevailing regime, often necessitating adherence to doctrine and the denunciation of non-Marxist perspectives within academic discourse.

The presented analysis is unique in methodical application of advanced statistical techniques to descriptive metadata of *Collection of Research Report* (characterized in more detail in part 2. Materials and Data Sources), in order to demonstrate the political connotations present in the ethnography of the time. Among the **research questions** are the following:

- What precisely constitutes the declarative use of Marxist-Leninist terminology, and how might it be statistically identified by using a data-driven methodology?
- How did Slovak ethnographers respond, within ethnographic field research, to the ascendant Marxist-Leninist ideology?
- In what manner was a declarative approach to Marxist-Leninist ideology manifested in ethnographic research reports produced?

Based on the research questions provided, the primary aim is to investigate the engagement of Slovak ethnographers with Marxist-Leninist ideology in their field research produced in the years 1948–1973, focusing on the identification and

---

1 Marxist-Leninist terminology denotes the standardized vocabulary, phrases, conceptual frameworks, and discursive style derived from the official state ideology of Marxism-Leninism. Its primary function was to demonstrate alignment with the official ideology of building a socialist society and consolidating state power, thereby legitimizing the ruling Communist Party's authority.

statistical analysis of formal declarative uses of ideologically charged thematic subjects. The structured secondary aims are as follows:

- to operationalize the concept of declarative use of Marxist-Leninist terminology within the context of Slovak ethnographic field research of the specified period,
- to develop a sequence of statistical methods to detect and confirm the instances of declarative use of Marxist-Leninist terminology in the data sample,
- to apply the statistical model, verify its effectiveness in identifying declarative usage, and interpret the declarative usage in relation to the political and academic context of the period.

The declaration of adherence to Marxist-Leninist ideology by scholars of that period is clearly evident (see the following section). Research on this issue has relied on qualitative methods, such as literature reviews, archival research, and interviews with contemporaries (Kiliánová, Popelková, 2010: 410). The article contributes by reinforcing this knowledge through quantitative statistical data analysis of research collection metadata. The **purpose** is not to reveal the general occurrence of Marxist-Leninist declarative expression, but to verify whether it can be statistically captured using this methodology within the sample data. Within the domain of documentation, it validates the application of advanced statistical methods to metadata entries and tests their innovative application for further ethnological analyses.

## 1. Historical Background

This section summarizes how Slovak ethnographers responded to the rise of Marxist-Leninist ideology after the communist seizure of power in Czechoslovakia. In this period, Slovak ethnography faced considerable obstacles stemming from the political climate. The ideological shift affected the organization of scholarly endeavours, particularly within the IESA SAS and related institutions (Beňušková, Jágerová, Nádaská, 2013: 38–52). The establishment of the Communist regime in 1948 resulted in the political imposition of Marxism-Leninism as the sole acceptable theoretical and methodological framework. This imposition manifested through several mechanisms, including institutional restructuring, labour law uncertainty, a reorientation of research priorities, and modifications to scholarly methodology and terminology (Kiliánová, Zajonc, 2016: 127–130).

Political screenings and purges substantially diminished both institutional and individual independence and autonomy. These actions resulted in dismissals, compelling some scholars to depart from the Academy. Research activities were integrated into a centralized plan, culminating in the establishment of state-sponsored programs for science and technology. State research priorities within ethnology were

directed toward topics such as the formation of socialist society, the transformations within folk culture resulting from collectivization, industrialization, and modernization. The overarching objective was to identify elements conducive to the development of a new socialist culture (Kiliánová, 2017).

Methodological advancement was hampered by a pervasive climate of apprehension and political pressure. Over time, Marxist-Leninist ethnography became the consolidated dominant perspective, thereby suppressing alternative perspectives. It is arguable that ideological conformity became as significant as scholarly expertise. However, a detailed qualitative analysis of research activities at IESA SAS revealed that the introduction of Marxism-Leninism during this period did not induce a radical discontinuity in research activity: “The application of Marxist principles often remained only at a declarative or formal level” (Kiliánová, Zajonc, 2016: 54). Instead, ethnographers primarily focused on manifestations of traditional folk culture, while demonstrating only a perfunctory interest in politically mandated topics during fieldwork (Zajonc, 2010: 436).

By superficially aligning with ideology, researchers could potentially maintain some degree of autonomy in their actual research focus and avoid potential censorship or repercussions: “The implementation of political objectives depended on the ability of the relevant personnel to consistently negotiate compromises between the directives of political authorities and the mission of scholarship” (Podolák, 2003: 23). The motivations behind this resistance were not solely political or ideological; often it was a matter of researchers’ unfamiliarity or lack of experience with the theory and methodology of Marxist-Leninist ethnography, which led to the emergence of various forms of approaches (Kiliánová, Zajonc, 2016: 54).

Outside the field of ethnography, the various forms of censorship (Harrington, 2020), self-censorship (Wagner, 2022), post-publication censorship and the issues of academic freedom in scholarly writing are a frequent subject of research. For the Central European region, the work of Oates-Indruchová is crucial. Through oral history interviews and archival research, she reveals how ideological control permeated the publishing process, leading to subtle yet pervasive pressures on authors and institutions. Within this framework, she also underscores the authors’ strategies of self-censorship and linguistic camouflage (Oates-Indruchová, 2020, 2021).

In summary, Slovak ethnographers during this period exhibited a complex and nuanced response to the imposition of Marxist-Leninist ideology (Zachar Podolinská, Popelková, 2023: 32–33). They adapted their research to conform to political expectations while simultaneously striving to uphold scholarly standards and preserve established ethnographic traditions. The imperative to reconcile ideological conformity with the pursuit of knowledge fundamentally shaped the trajectory of ethnographic research in Slovakia during this era.

## 2. Materials and Data Sources

This section details the demanding and intricate process of sample data preparation and selection. The data source for conducting analysis represents the *Collection of Research Reports* (CRR) stored at the Information and Documentation Department of IESA SAS. Since 1946, the research staff have been aggregating this primarily text-based collection, which comprises over 1,500 documents, totalling approx. 158,000 items in various formats (A4 papers, A5 archival cards, notebooks, photographs, drawings, maps, sheet music). The research reports were systematically collected through rigorous ethnographic fieldwork and archival research (thematically and locally focused) conducted by the institute's personnel, the resulting work being considered a 'commissioned piece of work' (utilizing methods of either direct observation or the collection and synthesis of various secondary written and visual sources). The CRR is thematically very diverse and contains records of most phenomena of traditional culture in Slovakia (see Table 1.). After completing the fieldwork, the employees submitted the draft of report to the department and it was processed by documentalists. When indexing research reports, the documentalists followed the metadata records that an author filled in on the standardized title page (including thematic subjects) and the established hierarchical classification system (Žatko, 1953); if necessary, they consulted directly with the author on specific cases. Subsequently, the CRR's reports were and remain freely available for on-site examination at the IESA SAS for the needs of the ethnographic research community, students, but also for the public (along with the corresponding archival registers, including metadata for localities and thematic subjects).

For the purpose of statistical analysis, it is not possible to use the CRR's research reports files.<sup>2</sup> Instead, respective metadata (descriptive and administrative) were employed. These were obtained through time-consuming manual transcription of the accession book and the card-based subject catalogue containing semantic keywords. Finally, not all documents and respective metadata are included, but a selection was performed according to the following criteria:

**a. Method of acquisition.** Documents produced through fieldwork, archival research, exploratory research, and questionnaire surveys were accepted; documents obtained by purchase, personal papers such as bequests and undefined documents were not accepted.

**b. Year of research implementation.** The author's original optimistic scope of the analysis encompassed documents dating from 1948 to 1989. However, preliminary statistical analysis revealed a substantial decrease in the number of thematic subjects

---

<sup>2</sup> Only rough post-processed scan images of each document are available so far, not electronically converted texts suitable for natural language processing (NLP).

Code	Category	Count	Proportion_all	Proportion_doc
I.	Employment of the population, production, transport, trade	10,771	0.2691	0.0757
II.	Food (in general)	2,065	0.0516	0.0717
III.	Clothing	5,442	0.136	0.0542
IV.	Construction (settlement, types of housing, furniture)	4,622	0.1155	0.0728
V.	Hygiene and cosmetics	241	0.006	0.0556
VI.	Social organizations, groups, social and family relationships, customs	2,616	0.0654	0.238
VII.	Customs, social practices, rituals, festive events	9,126	0.228	0.079
VIII.	Knowledge of the people, religion	1,764	0.0441	0.1316
IX.	Folk fine arts	332	0.0083	0.2491
X.	Folk theatre, narrative folklore	928	0.0232	0.2395
XI.	Singing, music, dance	1,005	0.0251	0.1733
XII.	Auxiliary disciplines	777	0.0194	0.416
XIII.	History of ethnography and folk culture	333	0.0083	0.296

*Table 1. Number of thematic categories for all CRR's documents. The "Proportion\_all" column represents the relative frequency of each code. It shows how common each thematic category is compared to the total number of categories (0.2691 for category I. means that 26.91% of all subjects fall under that category.). The "Proportion\_doc" column provides insights into the average relative importance of each category within individual documents. It suggests that when a category appears in a document, it tends to represent a larger portion of the document's content (0.416 for category XII. as a highest overall value means that when the category is assigned to a document, it tends to dominate the document's thematic content). Source: Author.*

assigned to individual documents from 1973 onward (approx. 2–5 subjects per document). Notably, the number of documents themselves did not decrease significantly during this period (see Table 2., Table 3.). This indicates that the data after 1973 are not representative. Consequently, the data sample was limited to documents originating from research conducted between 1948 and 1973.<sup>3</sup> This circumstance constrained the research aims, requiring them to be adjusted retroactively to conform to the available representative data. As will be demonstrated, this data limitation did not ultimately compromise the achieved results in a meaningful way. Therefore, the temporal selection criterion is as follows: documents conducted between 1948 and 1973; sustained multi-year research projects, which were disaggregated into individual years during metadata preparation, were also accepted.

<sup>3</sup> Several factors could have contributed to this decline. Firstly, the *Ethnographic Atlas of Slovakia* project commenced extensive fieldwork in 1971, potentially diverting resources and personnel. Secondly, around the 1970s, the task of assigning thematic subjects began to be undertaken by documentarians, rather than the authors of research reports, and without clearly defined, standardized criteria. Finally, limitations in the capacity of the documentation department may have played a role.

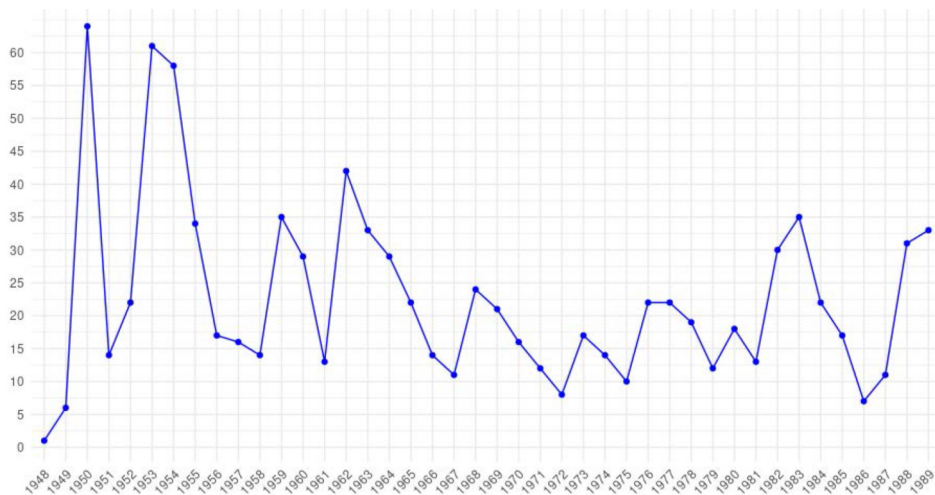


Table 2. The total number of CRR's documents per year. Source: Author.

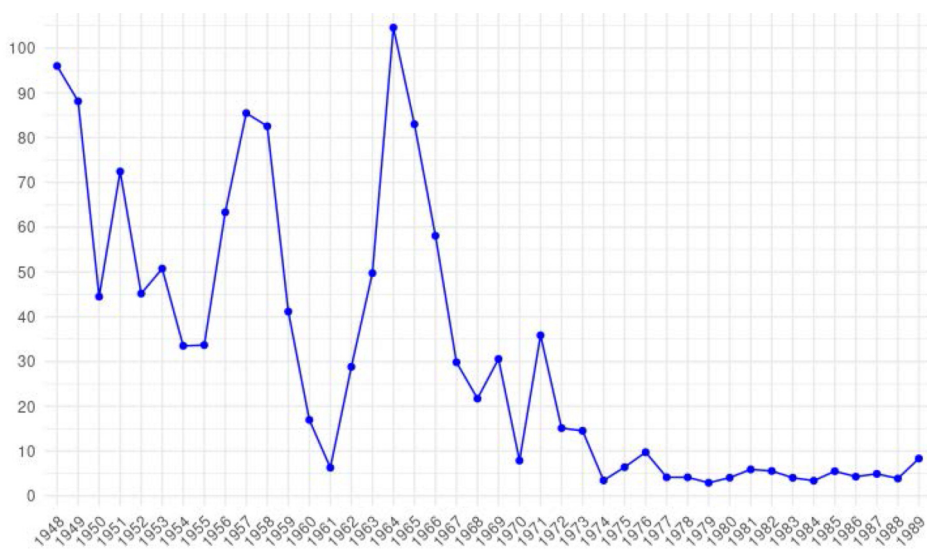


Table 3. Average number of CRR's thematic subjects per document per year. Source: Author.

Based on the strict application of these two criteria, the sample was reduced to **633 unique documents**, with 930 documents filtered out from the initial total of 1,563. The segmentation of certain documents into multiple sub-components was retained (e.g., AT1219a, AT1219b, AT1219c), as these generally contain distinct thematic subjects. Following this filtration process and the removal of duplicates, **39,751 unique thematic subject entries** remained, along with the documents' persistent identifier, and years of the research execution.<sup>4</sup>

4 At the time of the study's preparation, the FAIR principles were not applied to these data, owing to procedural and authorial reasons. The metadata are currently being prepared for online publication

Finally, it is necessary to address the two issues of the **representativeness**. Firstly, grounds for scepticism exist, given that the limited period may be not sufficient to detect the declarativeness. The initial decades following the imposition of ideological pressures are crucial for identifying how the research environment responded to them. Precisely, the tension between ethnologists' accommodation of the political imperative to study the nascent socialist society and the pressing need to conduct neglected research on traditional folk culture can be observed during this 25-year period. That's when declarative strategies emerged, aiming to maintain academic positions and preserve the integrity of the research itself.

Secondly, it appears questionable whether declarativeness can be identified within the metadata of research reports. In scholarly publications, this often manifested itself as obligatory Marxist-Leninist formulations. In the case of assigning subject topics to research reports, this will not be as explicit or clearly identifiable. However, the prevailing practice was such that fieldwork was rigorously and collectively organized according to ideological frameworks. The research requirements, including data collection, were politically driven and mandated (Kiliánová, Zajonc, 2016; Prandová, 1982). These factors, combined with the public accessibility of research reports and registers upon request, compelled researchers and documentalists to formally declare thematic content aligned with ideological imperatives, using ideologically charged subject headings. Ultimately, this was confirmed by the occurrences of ideological terms.

### 3. Methodology

The utilized methodology innovatively applies data-driven techniques to humanities research by statistically analyzing metadata (thematic subject headings) from ethnographic research reports. The study exemplifies the increasing role of quantitative methods in humanities, demonstrating how statistical analysis can complement traditional qualitative approaches (Porsdam, 2013) to provide new insights into historical and cultural phenomena (Flanders, Jannidis, *Eds.*, 2019; Gold, Klein, *Eds.*, 2019; Tilton, Mimno, Johnson, *Eds.*, 2024). By analyzing a large volume of metadata, the study offers a “distant reading” perspective, uncovering large-scale trends that might be missed in close readings of individual data. The methodology aligns with digital humanities' emphasis on leveraging computational tools to analyze and interpret textual cultural data (Lin, 2012). This section details the data preparation process; the operationalization of the concept of declarativeness; and an explanation of selected methodological tools employed in advanced statistical analysis.

---

on the IESA SAS institutional repository. Subsequently, these data, along with the research reports themselves (which are presently in the post-processing phase), will be (F)indable, (A)ccessible, (I)nteroperable, and (R)eusable.

Metadata were cleaned, structured, and prepared for analysis as follows:

- The CRR subject catalogue (80,846 entries) was cleaned by means of form correction, deduplication, removal of superfluous columns, table restructuring, and UTF-8 encoding.
- Selected metadata (year, acquisition method, document type) from the accession book were integrated into corresponding catalogue entries. Missing data were manually populated.
- The table was filtered out based on predefined criteria. Multi-year research periods (e.g., 1950–1951) were disaggregated into individual years (e.g., 1950, 1951), with corresponding subjects assigned. Generic codes (e.g., III.B.2.d) were expanded into specific subjects (e.g., clothing, women’s clothing, festive clothing, skirts). Metadata were consolidated per document into single rows with the following columns: document ID, year, subjects.
- These operations utilized LibreOffice Calc, OpenRefine, and semi-automated processing via the Gemini 2.0 generative AI model.<sup>5</sup>

The concept of declarativeness must be defined using categorizable indicators to facilitate analysis and achieve the research aim. A trivial definition posits that it constitutes a formal act that outwardly affirms something (formal compliance), yet is substantively empty and uncorrelated (superficial alignment). In the case of ethnographic research, it means the strategic employment of declarative expressions for the purpose of formally demonstrating political adherence to Marxism-Leninism (as defined in the Introduction). In addition, it is essential to expand the declarativeness to include the substitutive implementation of something that lacks an intrinsic connection to the ideology. For example, a researcher studying rural communities might prominently use the subject “collectivization” to signal adherence to socialist ideology, even if the primary research interest lies in traditional agricultural grain processing. To improve clarity, here is the dichotomy of **declarative expression (DE)** and **meaningful expression (ME)**:

- DE tends to be broad in scope, often representing common themes. In contrast, ME is more specific or unique, often representing niche topics or specialized aspects of the content.
- DE often exhibits high frequencies and co-occurrences with other DE, indicating widespread use and association with general themes. Conversely, ME is relatively rare, with low overall frequency, yet it appears prominently, indicating its importance for the content.

---

5 The Gemini 2.0 generative AI model was used for formal data operations such as formatting correction, table structure modification, and deduplication, not for processing metadata content. The outputs of these operations were manually reviewed.

- DE may be employed in a formal or standardized manner, as part of a controlled vocabulary or classification. ME, on the other hand, co-occurs with other related ME, forming clusters of specialized topics.

From this, the **operationalization of DE** is derived: DE is likely to appear frequently, may co-occur with other DE, and exhibits weak co-occurrence with other expressions within a document. Consequently, a methodical procedure was established to examine the probable distribution of DE. Initially, subjects that are evidently ideologically charged were manually selected from the CRR's subject catalogue. Themes were selected that related to the working class, industrialization, socialist production, collectivization, socialist culture, folk traditions reinterpreted within a socialist narrative, or internationalism. Their presence was verified within the filtered sample data.<sup>6</sup> This list was then utilized in the sequence of the following statistical analyses:

a.) **Chi-squared test:** This test determines whether there is a statistically significant association between documents and the presence of ideologically charged subjects. A significant chi-squared test reveals a strong association between documents and declarative subjects, highlighting their prevalent use. It utilizes a contingency table with documents as rows and selected subjects as columns, from which it is calculated whether the distribution of ideological subjects across documents is random or not. If the p-value is low (typically < 0.05), it suggests that the occurrences are not random and that an association exists.

b.) **Frequency analysis:** This analysis calculates the frequency of each ideologically charged subject across all documents and determines the percentage of documents that contain each subject. This provides an overview of how common each subject is. High frequencies of ideologically charged subjects would confirm their strategic deployment for ideological compliance.

c.) **Co-occurrence Analysis:** This analysis examines how often pairs or groups of selected ideological subjects appear together in the same documents, thus revealing connections between subjects. Co-occurrence patterns can uncover relationships between ideologically charged subjects (as well as meaningful subjects), further supporting their strategic use.

---

<sup>6</sup> The terms in this list consist of a limited number of initial characters followed by an asterisk (\*), which represents any subsequent string of characters. This approach is used because CRR thematic subjects are typically not formulated in their lemmatized form within the catalogue. Consequently, this method allows for retrieval of subjects in various morphological forms. Sample data consists of these terms: "collect.\*", "capit.\*", "socialis.\*", "work.\*", "industr.\*", "manufact.\*", "class.\*", "polit.\*", "UAC.\*", "mining.\*", "cooperat.\*", "associat.\*", "organ.\*".

d.) **TF-IDF (Term Frequency-Inverse Document Frequency):** Although traditionally employed in text analysis, TF-IDF can be adapted to metadata entries (Arnold, Tilton, 2024: 97–99). It can determine the importance or distinctiveness of a subject within a specific document relative to a larger collection of documents. Subjects with high TF-IDF scores are likely to be more meaningful within those documents. Conversely, low TF-IDF scores would suggest that the subjects are probably declarative, as they are not particularly meaningful in specific documents.

For the execution of the analyses, the R programming language was selected, as it is commonly and effectively used for statistical computing and graphics. It was implemented using RStudio software (version 2024.12.0), which serves as an integrated development environment.

#### 4. Results

By applying these methods to the sample data and a list of ideologically charged subjects, the following findings were obtained.

a.) **Chi-squared test:** Pearson's Chi-squared test was performed on a contingency table, which represents the cross-tabulation of document IDs and selected subjects. The result is:

$$X\text{-squared} = 5749.1, df = 4590, p\text{-value} < 2.2e-16$$

*X-squared* measures the discrepancy between the observed frequencies in the contingency table and the expected frequencies under the null hypothesis (that there is no association between documents and subjects). A larger value indicates a greater difference between observed and expected frequencies. *Df* is the degrees of freedom. It is calculated based on the dimensions of the contingency table and it indicates that the table has a large number of rows and columns. *P-value* “2.2e-16” is a scientific notation, representing an extremely small number, very close to zero ( $2.2 \times 10^{-16}$ ). The small *p-value* indicates that the results are statistically significant.

The chi-squared test result strongly suggests that the occurrences of ideological subjects are not random and that there is an association between documents and subjects. In other words, there is strong evidence to reject the null hypothesis of no association, but it doesn't directly tell whether it is truly declarative. Further analysis is needed to determine if the subjects are used in a declarative way.

b.) **Frequency analysis:** The percentage results offer insight into the relative frequencies of ideological subjects within the total number of subject occurrences. The findings indicate that “class structure” constitutes 22.42% of all identified ideological subjects, with “Unified Agricultural Cooperative (UAC)” (19.73%) and

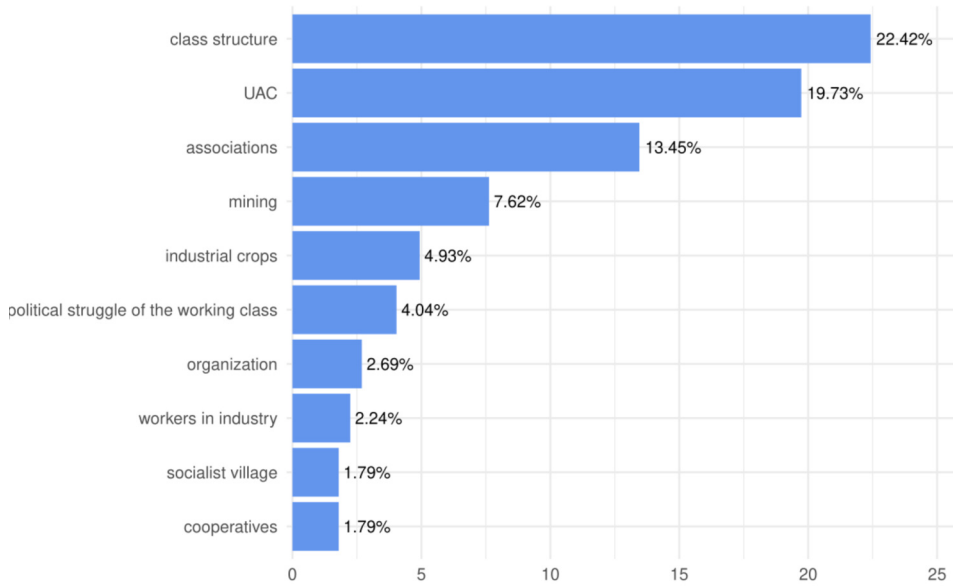


Table 4. Top ten ideological subjects by percentage. Source: Author.

“associations” (13.45%) also featuring prominently. The subject “the political struggle of the working class” (4.04%) appears with moderate frequency. The elevated percentages associated with “class structure”, “UAC”, and “associations” support the hypothesis that they might possess a more declarative character. They likely function as broad categories applicable to a wide range of documents, suggesting a potentially formal or standardized application of these terms.

c.) **Co-occurrence Analysis:** The results indicate that certain ideological subjects exhibit high co-occurrence counts with one another. The subject “class structure” shows high co-occurrence with “UAC” (22), “socialist village” (4), and “class differentiations” (1). “UAC” also demonstrates high co-occurrence with “associations” (13) and “socialist village” (4). These co-occurrences suggest that these subjects tend to appear together within the same documents, implying potential relationships or shared thematic elements. The co-occurrence matrix reinforces the previously identified clusters of politics-related and class-related subjects. The matrix is characterized by considerable sparsity, with numerous zero values. This sparsity indicates that many subject pairs do not co-occur, a finding that is unsurprising given the large number of subjects and the potential specificity of some. In the context of the preceding findings, the high co-occurrence rates between subjects such as “class structure”, “UAC”, “associations” further bolster the notion that these subjects may function in a more declarative capacity.

id	Subject	tf_idf
AT0674	class structure	0.142448975500464
AT0327	class structure	0.249285707125813
AT0447	class structure	0.332380942834417
AT0474	class structure	0.332380942834417
AT0506d	class structure	0.332380942834417
AT0038	class structure	0.498571414251625
AT0241	class structure	0.498571414251625
AT0276	class structure	0.498571414251625
AT0280	class structure	0.498571414251625
AT0372	class structure	0.498571414251625

Table 5. The ten lowest TF-IDF scores for the subject “class structure” in particular documents. Source: Author.

d.) **TF-IDF:** The results indicate that the ideological subject “class structure” exhibits relatively low TF-IDF scores (0.498 and below) in the majority of documents, likely due to its overall high frequency. This suggests that subjects with low TF-IDF scores, particularly those that are frequent, tend to be declarative rather than distinctive. Conversely, subjects with high TF-IDF scores are more likely to be semantically significant, as they differentiate those documents from others within the dataset. Among the ideological subjects, “miners’ artistic expression”, for example, has a high TF-IDF score (4.63) in document AT0008. These elevated scores indicate that such subjects are relatively rare across the entire dataset but appear prominently within specific documents.

## 5. Discussion: the case of “class structure”

In summary, to confirm the presence of DE, it was necessary for some of the ideologically charged subjects to exhibit the following characteristics: low p-values, high frequency, high co-occurrence rates, low TF-IDF scores. Based on the combined results, the subject “class structure” in particular exhibits a high frequency, suggesting broad applicability and potentially standardized usage. Other high-frequency subjects (“UAC”, “associations”) demonstrate frequent co-occurrence with “class structure”, reinforcing their declarative nature. Finally, “class structure” possesses low TF-IDF

scores, indicating that it is not distinctive to individual documents. It is highly probable that this ideological subject demonstrates attributes associated with DE.

In the case of such analyses, it is recommended that by synthesizing quantitative findings with expert qualitative insights, a more nuanced comprehension of the issue can be attained. To corroborate the statistical findings, an **archival survey** was conducted on documents indexed with the term “class structure”. This involved a manual review of the contents of 38 extensive research reports (approx. total of 5,451 units). The majority of these documents were formatted as archival cards with a pre-printed layout, including a clearly designated field for subject heading. Surprisingly, the laborious examination of all units demonstrated that the term “class structure” was absent as a subject in any of the research reports items, notwithstanding its assignment to the documents in the subject catalogue. This finding substantially strengthens the hypothesis that the term is used declaratively.

Furthermore, an examination of the sets of subjects associated with individual documents confirmed that the term “class structure” lacks any semantic connection to the documents’ content, which is predominantly monothematic (e.g., viticulture, furriery, shepherding, folk art). In the subject catalogue, it appears as a supplementary and unrelated term, situated among subjects that are thematically coherent with each other but indifferent to it. While some reports contain reflection on socialist society, these are nowhere interpreted through the lens of Marxist-Leninist ideology. This stands in contrast to the strong ideological connotation of the term “class structure” which, in the given context, represents a fundamental concept of dialectical materialism. This further reinforces the assertion that, while ethnographic fieldwork of the period declared obedient adherence to the ideology, it was, in practice, primarily focused on the systematic and meaningful collection of data on traditional folk culture in Slovakia.

It is important to note that, in the case of this type of sample data, these analyses primarily reflect the methods of metadata preparation. Consequently, the findings inherently pertain to the ethnographic documentation practices of the time. The CRR case repeatedly underscores the crucial and complex relationship between the ethnographer in fieldwork and the documentalist in archive, alongside the necessity for precise documentation of curatorial working procedures.<sup>7</sup> Therefore, it is not entirely clear whether the term “class structure” was intentionally used as a DE, as the rationale for its assignment to the research reports in the subject catalogue is, unfortunately, irretrievable.

---

7 Within the documentation department’s resources, there is no original manual or handbook that clearly and systematically prescribes the practice of documenting and indexing research collection materials.

## Conclusion

This study investigated the presence of declarative uses of Marxist-Leninist terminology in Slovak ethnographic field research conducted between 1948 and 1973. Key findings confirm the presence of ideologically charged subjects in the sample of CRR documents, particularly “class structure”. These subjects exhibit high frequency, co-occurrence, and low TF-IDF scores, indicating their broad, potentially standardized, and declarative nature. The study contributes to understanding the relationship between ethnographers and the political ideology during the socialist era in Slovakia. Methodologically, the research innovates by applying advanced statistical techniques to metadata of the CRR in order to uncover broad trends in ethnographic data, such as the prevalence of ideologically charged terms over time. This approach allows for quantitative analysis of documentation practices and the identification of declarative expressions. The study’s methodology offers valuable insights for future ethnological research based on data-driven approaches, such as extending the analysis to other periods, thematic clusters or regions. Digital humanities’ statistical methods can provide a systematic way to analyze large collections of ethnographic materials, enabling researchers to identify patterns in thematic focus across different periods or regions. Furthermore, these methods can facilitate comparative studies, allowing scholars to statistically examine similarities and differences between various ethnographic traditions or the impact of external factors. By employing techniques like network analysis or geospatial analysis, it can also reveal connections between ethnographers, research sites, or cultural phenomena, offering new perspectives on the social and intellectual networks within the field of ethnography.

### Acknowledgements

*The article was supported from the public funds of the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic and the Slovak Academy of Sciences, Grant No. VEGA 2/0149/25 (2025-2027) Digital thematic research collections and the ethnographic research process (principal investigator: A. Gogora).*

*The author gratefully acknowledges the former colleague Tomáš Kubisa, and all others who contributed to the transcription of the CRR subject catalogue and the scanning of research reports. The author also expresses sincere appreciation for the reviewers’ insightful comments and valuable suggestions, which significantly enhanced the quality of the manuscript.*

## REFERENCES

- Arnason, J. P. (1982). Perspectives and Problems of Critical Marxism in Eastern Europe (Part Two). *Thesis Eleven*, 5-6(1), 215–245. <https://doi.org/10.1177/072551368200500115>
- Arnold, T., Tilton, L. (2024). *Humanities Data in R. Exploring Networks, Geospatial Data, Images, and Text*. Cham: Springer. <https://doi.org/10.1007/978-3-031-62566-4>
- Beňušková, Z., Jágerová, M., Nádaská, K. (2013). *Dejiny slovenskej etnológie v 20. storočí*. Nitra: Univerzita Konštantína Filozofa, Filozofická fakulta.
- Flanders, J., Jannidis, F. (Eds.) (2019). *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*. Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781315552941>
- Gold, M. K., Klein, L. F. (Eds.) (2019). *Debates in the Digital Humanities 2019*. University of Minnesota Press. Available at: <https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2019>
- Harrington, A. (2020). Censorship: The Challenge of Writing in Oppressive Regimes. In: T. Boes, R. Braun, E. Spiers (Eds.), *World Authorship* (pp. 46–59). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198819653.013.4>
- Kiliánová, G. (2017). Ethnology in Slovakia in crucial historical periods (after 1968 and 1989): From a historical to a social sciences discipline? *Národopisná revue*, 27(5), 65–76.
- Kiliánová, G., Popelková, K. (2010). Zavádzanie marxistickej etnografie v národopise na Slovensku: zmena vedeckého myslenia? *Slovenský národopis*, 58(4), 410–424.
- Kiliánová, G., Zajonc, J. (2016). *70 rokov Ústavu etnológie Slovenskej akadémie vied: kontinuity a diskontinuity bádania a jednej inštitúcie*. Bratislava: VEDA, vydavateľstvo Slovenskej akadémie vied. <https://doi.org/10.31577/2016.9788022415125>
- Lin, Y. (2012). Transdisciplinarity and Digital Humanities: Lessons Learned from Developing Text-Mining Tools for Textual Analysis. In: D. M. Berry (Ed.), *Understanding Digital Humanities* (pp. 295–314). Palgrave Macmillan.
- Oates-Indruchová, L. (2020). *Censorship in Czech and Hungarian Academic Publishing, 1969–89: Snakes and Ladders* (1<sup>st</sup> ed.). Bloomsbury Publishing Plc. <https://doi.org/10.5040/9781350106673>
- Oates-Indruchová, L. (2021). The Limits of Thought: Academic Censorship during Normalization and its Legacy. *Securitas imperii*, 39, 17–26.
- Podolák, J. (2003). Etnológia na Slovensku v 20. storočí. Etapy jej vývoja. In: O. Elschek (Ed.), *Etnologické a folkloristické inštitúcie na Slovensku. Ethnologia Actualis Slovaca*, 3, 9–58.
- Porsdam, H. (2013). Digital Humanities: On Finding the Proper Balance between Qualitative and Quantitative Ways of Doing Research in the Humanities. *Digital Humanities Quarterly*, 7(3). <http://www.digitalhumanities.org/dhq/vol/7/3/000167/000167.html>
- Prandová, E. (1982). Dokumentačné fondy Národopisného ústavu SAV 1946–1977. Krajské osvetové stredisko v Bratislave.
- Tilton, L., Mimno, D., Johnson, J. M. (Eds.) (2024). *Computational Humanities*. University of Minnesota Press.
- Wagner, I. (2022). Auto-censura (self-censorship) – an ordinary and integral part of research life(?). *Zoon Politikon*, 13, 130–174. <https://doi.org/10.4467/2543408XZOP.22.005.16308>
- Walker, R. (1989). Marxism–Leninism as Discourse: The Politics of the Empty Signifier and the Double Bind. *British Journal of Political Science*, 19(2), 161–189.
- Zachar Podolinská, T., Popelková, K. (2023). *Story of Cohesion: On the History of an*

*Academic Institution*. Bratislava: VEDA, Publishing House of the SAS, Institute of Ethnology and Social Anthropology SAS. <https://doi.org/10.31577/2023.9788022420228>  
Zajonc, J. (2010). Výskum v okrese Modrý Kameň v roku 1950 ako príklad prenikania marxizmu do slovenského národopisu. *Slovenský národopis*, 58(4), 425–442.  
Žatko, R. (1953). *Vecný systém triedenia materiálu*. Národopisný ústav SAV.

## ABOUT THE AUTHOR

ANDREJ GOGORA (ORCID: 0000-0001-5169-5330) – is a senior researcher at the Institute of Ethnology and Social Anthropology SAS and Head of the Research Collections Department. He completed his doctoral studies in systematic philosophy at the Institute of Philosophy SAS. He is engaged in basic research on the methodological foundations of digital humanities and digital curation. In applied research, he is actively involved in building ethnological digital research collections. He is the coordinator on behalf of the IESA SAS as a cooperating partner of the European research infrastructure DARIAH-EU.