Corpus-based monitoring of dual normativity in orthographic principles⋆

Solomiia Albota[1,†], Zoriana Kunch[2,∗,†], Liliia Kharchuk[3,∗,†], Przemysław Jóźwikiewicz[4,†] and Roman Holoshchuk[5,†]

[1,2,3,5] *Lviv Polytechnic National University, Bandera Str., 12, Lviv, 79013, Ukraine*

[4] *Uniwersytet Wrocławski, Plac Uniwersytecki 1, 50-137 Wroclaw, Poland*

Abstract

This study presents a corpus-based methodology for monitoring dual normativity in Ukrainian orthographic principles, focusing on the lexemes ефір [ephir] and етер [eter]. Using the General Regionally Annotated Corpus of Ukrainian (GRAC) and its subcorpus "Media 2000–2023", the research traces the dynamics of implementing alternative orthographic forms in contemporary media discourse. The methodology allows for quantitative and qualitative analysis of dual norms, providing insights into the gradual establishment of a single normative variant. The historical evolution of ефір and етер reflects broader sociolinguistic processes in Ukraine, balancing the legacy of Soviet standardization with the restoration of pre-Soviet linguistic traditions. The coexistence of these variants highlights tensions between practical usage and cultural-linguistic renewal. This approach not only contributes to corpus-based studies of orthography but also offers material for psycholinguistic research and the analysis of normative language change in the digital media environment. The novelty of the research lies in the development of a replicable approach to tracing the dynamics of alternative orthographic forms in media discourse, which has not been systematically studied before.

Keywords

corpus linguistics, orthography, orthographic norm, variability, double normativity, lemma

1. Introduction

The functioning and development of a national language at any given time rely on a unified orthographic code — the current orthography, whose fundamental principles must be adhered to in everyday speech communication by every speaker. Established orthographic norms contribute to the continuity of verbal traditions, ensure literacy, and unify the national community. The issue of adhering to orthographic norms is addressed in scientific research worldwide. Monitoring orthographic change is particularly relevant for Ukraine today, as orthography reflects broader processes of decolonization and cultural identity restoration. Understanding how dual norms function in practice helps explain not only linguistic habits but also societal shifts after the 2019 reform.

The latest amendments to the current Ukrainian orthography were approved five years ago, making them mandatory across all language domains, particularly in Ukrainian media discourse, following the end of the transition period. Some orthographic rules, which were deliberately distorted due to extralinguistic coercive influences of the Soviet period, now introduce variant forms, i.e., dual normativity (also referred to as the coexistence of parallel orthographic variants or orthographic doublets), which aims to facilitate a gradual transition towards a single possible variant. Researchers rightly emphasize that new media are changing the habits of digital language users [1]. Modern corpus technologies allow for an analysis of how effective and appropriate the

introduction of dual normativity is and to assess the feasibility of achieving exclusive usage of a single form soon. Thus, it is pertinent to investigate the dynamics of the application of one of the two possible normative forms in media discourse. In this study, dual normativity is understood as the official coexistence of two parallel orthographic variants recognized by the codified norm, as opposed to spontaneous variation or non-standard usage.

This work aims to analyze the methodology for monitoring dual normativity of orthographic postulates using corpus linguistics, based on the analysis of the lexemes *ефір* [ephir] / *етер* [eter]. The General Regionally Annotated Corpus of Ukrainian [2] was used as the experimental field for implementing the proposed methodology. Its application creates prerequisites for searching factual data for linguistic analysis and statistical processing of collected materials. In GRAC v. 17, a subcorpus "Media 2000-2023" was created, based on which the applied research was conducted.

2. Related Works

The issue of adhering to orthographic norms is addressed in scientific research worldwide. Among international studies, we focus primarily on those that provide methodological parallels for the Ukrainian case, such as analyses of Hungarian, English, and Spanish orthographic reforms. Recent publications have analyzed the current state of Hungarian orthography from a dual perspective of preservation and modification following the publication of the 12th edition of "Rules of Hungarian Orthography" [3]; examined the correspondence of five generations of an elite English family to understand how the standardization of vernacular writing during this period was reflected in practice [4]; described linguistic ideological debates regarding Spanish orthography in the mid-nineteenth century [5]; and highlighted the patterns and causes of orthographic errors made by Chinese students learning Thai as a foreign language [6], among others. Ukrainian researchers have noted distinctive features in the use of normativity in Business, Informal, and Internet Communication [7], outlined the most controversial innovations of the current orthography and preserved linguistic traditions [8], and established the dynamics of changes in orthographic rules in connection with the choice of one of three different types of communication — traditional official (business), unofficial (handwritten notes), and informal online communication [9]. Separate studies concern new opportunities to utilize orthographic data to create a system of intelligent analysis of Ukrainian literary works to determine the likelihood of text authorship [10]; inform psychological theories about spelling processes based on a large-scale database of orthographic errors [11]; identify trends in adhering to orthographic norms based on the analysis of electronic lexicographic works [12]; and investigate spelling variations in digital written communication beyond the binary paradigm of standard and non-standard [13].

The applied analysis of normativity and adherence to or violation of linguistic regulations remains a relevant and priority area of research, attracting the attention of both Ukrainian and international scholars. A review of recent studies confirms the multifaceted nature of this issue, as evidenced by numerous investigations [14]. In particular, researchers focus on developing general principles for assessing linguistic norm compliance at various levels, as well as analyzing specific instances of norm violations [7].

Notably, considerable attention has been devoted to the study of orthographic norm adherence across different languages, including English [15], Hungarian [16], Spanish [17], Greek [18] and German [19], among others.

The application of corpus linguistics to the study of spelling norm implementation in media discourse has been a key focus of previous research. By employing a targeted methodological framework for corpus-based analysis, it has been possible to examine the dynamics of spelling norm adoption in media discourse during the adaptation period (2019–2023) [20]. Additionally, statistical analyses of specific norms have facilitated the identification of patterns in the usage of key linguistic forms affected by the 2019 orthographic reform, enabling a systematic exploration of their distribution within a compiled text corpus.

Thus, the automated processing of natural language data is gaining increasing importance, while applied linguistics technologies continue to expand their capabilities in the analysis, preservation, and selection of linguistic material. Linguistic corpora have rapidly become an indispensable research resource, as they integrate extensive empirical data with advanced computational techniques. This approach enables objective insights into the functioning of linguistic units, while corpus-based analysis significantly contributes to the formulation of novel scientific generalizations.

3. Methods and Materials

The research methodology integrates classical general-theoretical methods (analysis, generalization, and explanation) with applied linguistic methods. The study employs the following general scientific and linguistic methods:

1. The methodology is based primarily on corpus-driven procedures, including lemma-based concordance searches, frequency list generation, and diachronic distribution analysis within the "Media 2000-2023" subcorpus.
2. Statistical and corpus-based text analysis methods, widely utilized in modern linguistics, particularly in applied linguistics, were employed. These methods encompass a set of techniques and principles for data collection, systematization, processing, and interpretation, ultimately facilitating scientific and practical conclusions. As targeted methods, they are considered among the most effective research tools in applied linguistics.
3. The structural (descriptive) method, with a comparative analysis component, enabled the systematization, classification, and description of the collected material.
4. Generalization and induction methods allowed for the linguistic-statistical analysis of the obtained results.

As previously noted, lexemes affected by the spelling changes introduced in the 2019 edition of the Ukrainian Orthography were extracted from their usage contexts in media discourse based on the General Regionally Annotated Corpus of the Ukrainian Language (GRAC) [2]. This corpus represents a structured and representative collection of Ukrainian-language texts, accompanied by a software tool that allows for the creation of custom subcorpora, word and grammatical form searches, and the processing, sorting, and balanced sampling of search results to obtain various statistical insights [2]. The extracted data was further subjected to linguistic analysis.

The proposed analytical methodology has the potential to serve as a foundation for similar corpus-based investigations of other orthographic norms, facilitating the examination of the dynamics of standardized spelling adoption. Additionally, this approach may provide valuable material for further psycholinguistic studies employing modern computational technologies.

4. Results

4.1. Title information

Orthographic norms ensure the correct rendering of words in writing. Section 123 of the current orthography states: "The combination th in words of Greek origin is usually rendered by the letter t: *антологія* [anthology], *антропологія* [anthropology], *аптека* [pharmacy] <...>. In words established in the Ukrainian language with *ф* [f], orthographic variability is possible as in: *анафема* [anathema] and *анатема* [anathema], *дифірамб* [dithyramb] and *дитирамб* [dithyramb], *ефір* [ephir] and *етер* [eter]" [21]. Thus, contrary to the Soviet-era norm of exclusively using *ефір* [ephir], the current orthography considers both *ефір* [ephir] and *етер* [eter] as equally correct.

For this study, we extracted a selection of citations with both variants — *ефір* [ephir] and *етер* [eter] — from the "Media 2000-2023" subcorpus of the General Regionally Annotated Corpus of Ukrainian [2]. We searched in the concordance by lemma to identify these words in all grammatical forms. Therefore, the material for monitoring dual normativity included 97425 citations with the lemma *ефір* [ephir] and 2485 usages of the lemma *етер* [eter]. The frequency of use of each lexeme in the subcorpus differs significantly. Based on Figure 1, we observe peaks in the frequency of the lemma *етер* [eter] in specific subcorpus fragments, which may indicate either the temporal usage of the lexeme or the preference for this lexeme by certain publications.
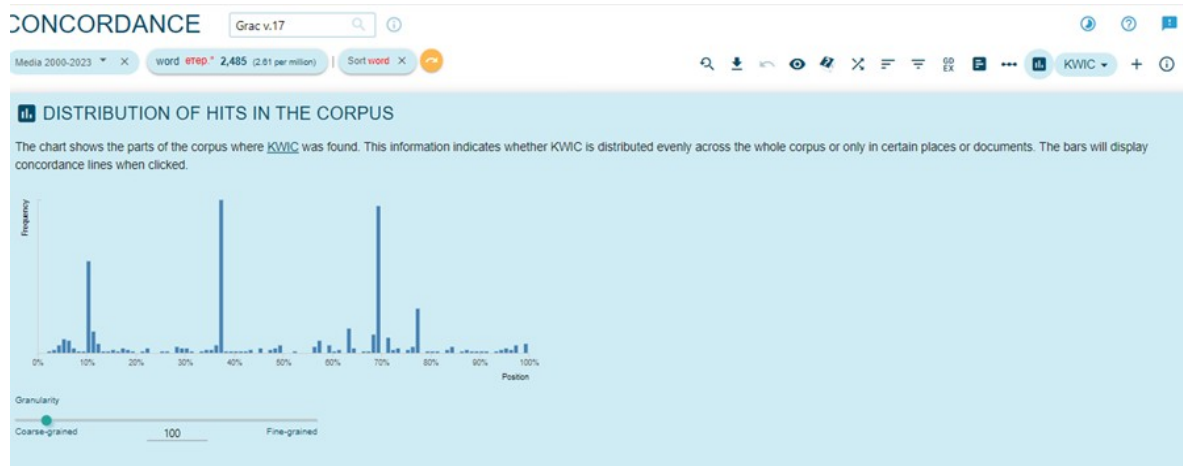

Figure 1: Distribution of lemma *етер* [eter] in the corpus.

The lexeme *ефір* [ephir], which was the sole normative form before the introduction of the new Ukrainian orthography in 2019, clearly demonstrates significantly higher frequency (97.5% of all usages). This is because, within the timeframe from 2000 to 2019, only the lexeme *ефір* [ephir] should have been used exclusively. However, it should be noted that our further analysis does not entirely confirm this. Figure 2 illustrates how this lexeme is distributed within the corpus.

We observe greater homogeneity and an absence of pronounced usage peaks, which provides grounds to consider the lexeme *ефір* [ephir] as consistently used.
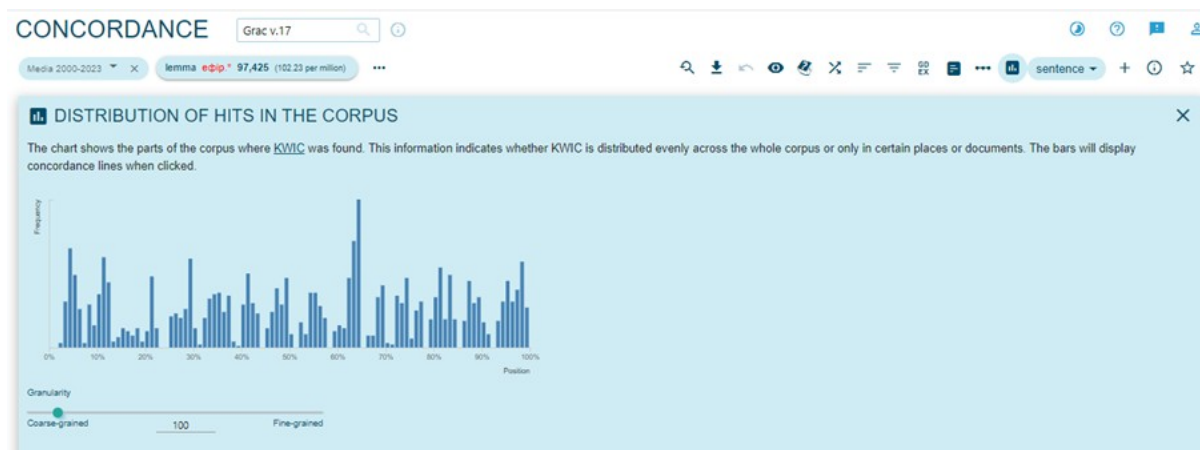

Figure 2: Distribution of lemma *ефір* [ethir] in the corpus.

The next stage of monitoring, enabled by corpus methodology, allows us to identify that the lemma *етер* [eter] is part of derived lexemes: common nouns such as *етерний* [eternyi],

етеричний [eterychnyi], *етеранський* [eteranskyi], as well as proper names such as *Етерія* [Eteriia], *Етерінгтон* [Eterington], *Етерович* [Eterovych], etc. This is because our selection includes not only the common noun *етер* [eter], but also several derived words. By creating a frequency list of word forms, we find that, for example, the proper name *Етерінгтон* [Eterington] appears 139 times in various grammatical forms in our study materials, and the proper name *Етерович* [Eterovych] appears 51 times in various cases, and so on (see Table 1). Overall, as shown in Table 1, we observe 65 variants of the lemma *етер* [eter], indicating its high productivity, even though it was not normative until 2019.

Table 1
Frequency of usage of variant forms of the lemma *етер [eter]*

Word	Frequency	Frequency per million
етері	1,715	0.96
етер	175	0.10
етеру	114	0.06
Етерінгтон	101	0.06
Етері	45	0.03
Етерович	35	0.02
ЕТЕРОМ	33	0.02
етери	26	0.01
етерах	25	0.01
Етер	20	0.01
Етерінгтона	16	<0.01
Етерідж	16	<0.01
Етерія	14	<0.01
Етеровича	12	<0.01
етерів	10	<0.01
Етерінгтоном	10	<0.01
Етерна	9	<0.01
Етертон	7	<0.01
ЕТЕРІ	7	<0.01
ЕТЕР	7	<0.01

Conducting similar research on the frequency of the lemma *ефір* [ephir], we first determine the nature of the lexemes derived from the word *ефір* [ephir] in this subcorpus. Creating a frequency list by word forms, we find that all variants of the lemma *ефір* [ephir] include grammatical forms of this word in different cases of singular and plural, derived common nouns such as *ефір* [ephir] *оманія* [etheromania], *ефір* [ephir] *ність* (etherness), and others, as well as grammatical forms of the adjective *ефір* [ephir] *ний* [etheric], and compound adjectives such as *ефір* [ephir] *олійний* [ether-oil], *ефір* [ephir] *но-цифровий* [ether-digital], *ефір* [ephir] *но-рожевий* [ether-pink], *ефір* [ephir] *но-прозорий* [ether-transparent], *ефір* [ephir] *но-проводовий* [ether-conductive], *ефір* [ephir] *но-кабельний* [ether-cable], *ефір* [ephir] *но-дерев'яний* [ether-wooden], as shown in Table 2. However, it is notable that no proper names derived from this word have been found. This observation confirms the appropriateness of adopting the normative use of the variant *етер* [eter], as the Ukrainian language, after the forced removal of *ефір* [ephir] from the lexical system, has not demonstrated the ability to form proper names from the lemma *ефір* [ephir].

Table 2

Frequency of usage of variant forms of the lemma *ефір* [ephir]

Word	Frequency	Frequency per million
ефірі	73,186	41.08
ефір	8,561	4.81
ефіру	6,711	3.77
ефіри	1,317	0.74
ефірах	1,267	0.71
Ефір	1,097	0.62
ефірного	1,094	0.61
ефірів	1,073	0.60
ефірні	451	0.25
ефірний	383	0.21
ефірних	372	0.21
ефірне	342	0.19
ефіром	318	0.18
ефірною	233	0.13
ефірними	118	0.07
ефірну	113	0.06
ефірна	89	0.05
ефірним	85	0.05
ефірами	83	0.05
ефірному	75	0.04

4.2. Monitoring the chronology of usages

To study the dynamics of implementing dual normativity over time, we create a frequency list based on the chronology of usage. We find that only 79.4% of the instances of the lemma етер [eter] occur in the period from 2019 to 2022, i.e., after the norm was introduced. Thus, more than 20% of the citations pertain to the period when the current spelling did not include this lemma as a normative variant. Therefore, we can conclude that contemporary media were guided by the recommendations of the spelling commission that worked in the 1990s under the "leadership of Professor V. Nimchuk, who developed the project "Ukrainian Spelling. The Project of the Latest Edition" [22]. From 2004 onwards, the lemma етер [eter] shows a stable frequency of usage in the media, especially leading up to its official adoption in 2018, when the spelling commission made the corresponding decision (117 instances of this lemma were found). The chronological analysis of the usage of the lemma *ефір* [ephir] shows significant frequency of its usage in the period from 2019 to 2022: a total of 39,169 citations, which constitutes 40.2% of all recorded instances in this subcorpus (as seen in Table 3), even though during this period both *ефір* [ephir] and *етер* [eter] were normative. This indicates a certain inertia in the Ukrainian media space regarding the adoption of a significant language change and confirms the presence of a stable linguistic habit. Detailed information is presented in Table 3.

Table 3
Relative frequency of usage of the lemmas *ефір* [ephir] / *етер* [eter] by chronology

Year	*Ефір* [ephir], %	Year	*Етер* [eter], %
2020	11,2	2022	30,8
2019	10,6	2021	22,3
2022	9,8	2020	19
2018	9,2	2019	7,3
2016	8,8	2018	4,7
2021	8,6	2014	2,9
2015	6,9	2015	2,7
2014	5,9	2011	2,7
2017	5,7	2016	1,4
2012	4,1	2017	1,3

4.3. Detecting the influence of subjective factors on the choice of normative variants

Recognizing that corpus mechanisms allow us to observe which media outlets prefer using one lexeme over another is crucial. Our frequency analysis reveals that the lexeme *етер* [eter] appears 1,299 times, which is more than half of all identified instances, in the online media outlet "Hromadske TV". The remaining sources in the corpus show significantly lower frequencies for this lexeme: "NV" magazine — 278 instances, online media "LB.ua" — 115 instances, online media "UNIAN.NET" — 102 instances, online media "European Pravda" — 72 instances, and "Svoboda"

newspaper — 72 instances. Regarding the lexeme *ефір* [ephir], our observations show that "Hromadske TV" exhibits a significantly lower frequency of this lexeme compared to other sources: its usage of *ефір* [ephir] is only 16.7% compared to the online media outlet "UNIAN.NET", which has the highest frequency of its usage. Preference for the lexeme *ефір* [ephir] is shown by sources such as "NV" magazine, "LB.ua", and "Zaxid.net". Thus, the study of the frequency of use of the lexemes *ефір* [ephir] / *етер* [eter] across different publications, presented in Figure 3, suggests that the use of one of the normative variants is to some extent influenced by the linguistic consciousness and preferences of the authors and editors of these publications. In other words, when faced with dual norms, subjective factors come into play.
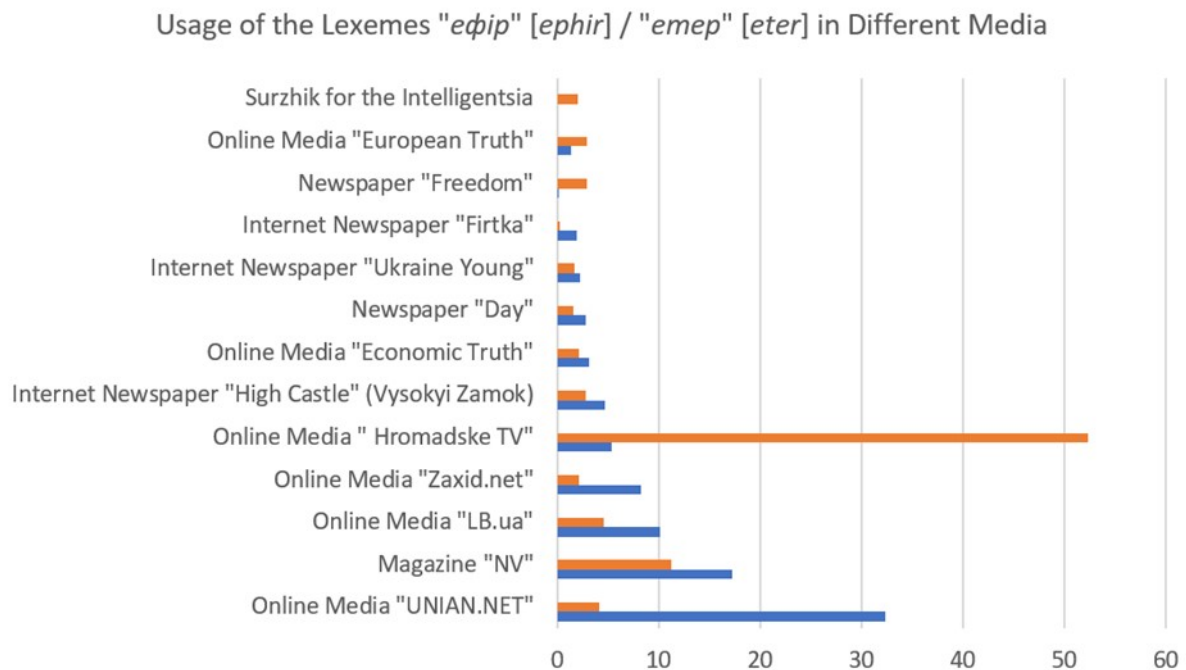

Figure 3: Frequency of usage of the lexemes *ефір* [ephir] / *етер* [eter] in different media.

When comparing the frequency of usage of the lexeme *ефір* [ephir] with the names of publications over specific periods, it is evident that the online media outlet "UNIAN NET" and the magazine "HB" rank highest in frequency during the period 2019–2022. This confirms our previous hypothesis about the preference for one of the variants of dual normativity. More detailed information can be traced in Table 4.

Table 4
Chronology of the frequency of usage of the lexeme *ефір* [ephir] in the publications of the "Media 2000–2023" corpus

№	Year	Title	Frequency
1	2022	Online Media "UNIAN.NET"	5,219
2	2019	Magazine "NV"	3,299
3	2018	Magazine "NV"	3,255
4	2020	Magazine "NV"	3,180
5	2021	Online Media "UNIAN.NET"	3,063

6	2020	Online Media "UNIAN.NET"	2,836
7	2019	Online Media "UNIAN.NET"	2,262
8	2021	Magazine "NV"	1,957
9	2018	Online Media "UNIAN.NET"	1,946
10	2011	Online Media "UNIAN.NET"	1,864
11	2016	Magazine "NV"	1,716
12	2014	Online Media "UNIAN.NET"	1,669
13	2016	Online Media "UNIAN.NET"	1,664
14	2020	Online Media "LB.ua"	1,612
15	2017	Magazine "NV"	1,590
16	2015	Online Media "UNIAN.NET"	1,545
17	2019	Online Media "Hromadske TV"	1,442
18	2012	Online Media "UNIAN.NET"	1,399
19	2016	Online Media "Hromadske TV"	1,373

Based on the data provided, line graph, bar graph and pie chart representing lexemes *ефір [ephir]* / *етер* usage trend, comparison and distribution have been provided in Figure 4, Figure 5, Figure 6. The image contains three different visual representations of data regarding "Usage" over time for two categories: *ефір [ephir]* (blue) and *етер* (red). The first visual sketch *Usage Trend Over Time* predisposes that the x-axis represents years (from 2005 to 2022). The y-axis represents frequency (%) of usage. The blue line (*ефір*) shows a gradual increase over time with some fluctuations. The red line (*етер*) remains low and inconsistent until around 2017, after which it rapidly increases, surpassing *ефір* [ephir] after 2020. This suggests that *етер* experienced a sudden surge in popularity after 2017, while *ефір* grew steadily (see Figure 5). The second visual sketch *Comparison of Usage by Year* leads to that the bar chart shows usage frequency (%) by year for both categories: before 2017, *ефір* dominated in most years. From 2017 onward, *етер* started increasing significantly, surpassing *ефір* around 2020–2021. The gap widened sharply in 2022, where *етер* usage surged past *ефір* (see Figure 6). The third visual sketch *Usage Distribution in 2022 (Bottom Pie Chart)* shows that red (*етер*) occupies 75.9% of total usage. Blue (*ефір*) occupies 24.1%. This confirms that in 2022, *етер* significantly outpaced *ефір*, taking over three-quarters of the total usage. So, *етер* saw a major increase after 2017, with an exponential surge from 2020 to 2022. *Ефір* remained stable but did not grow at the same pace as *етер*. By 2022, *етер* usage was three times higher than *ефір*, indicating a major shift in preference (see Figure 6).
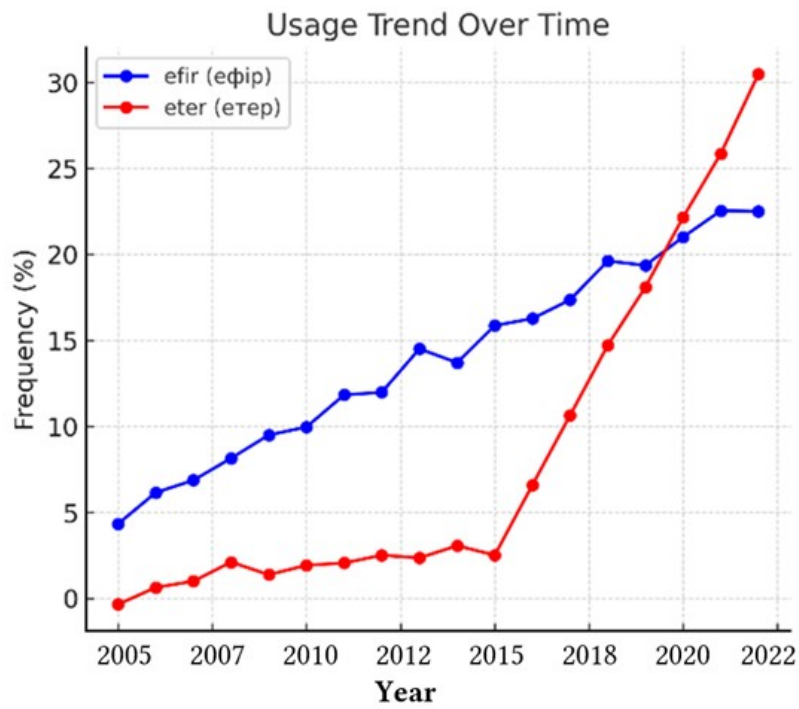
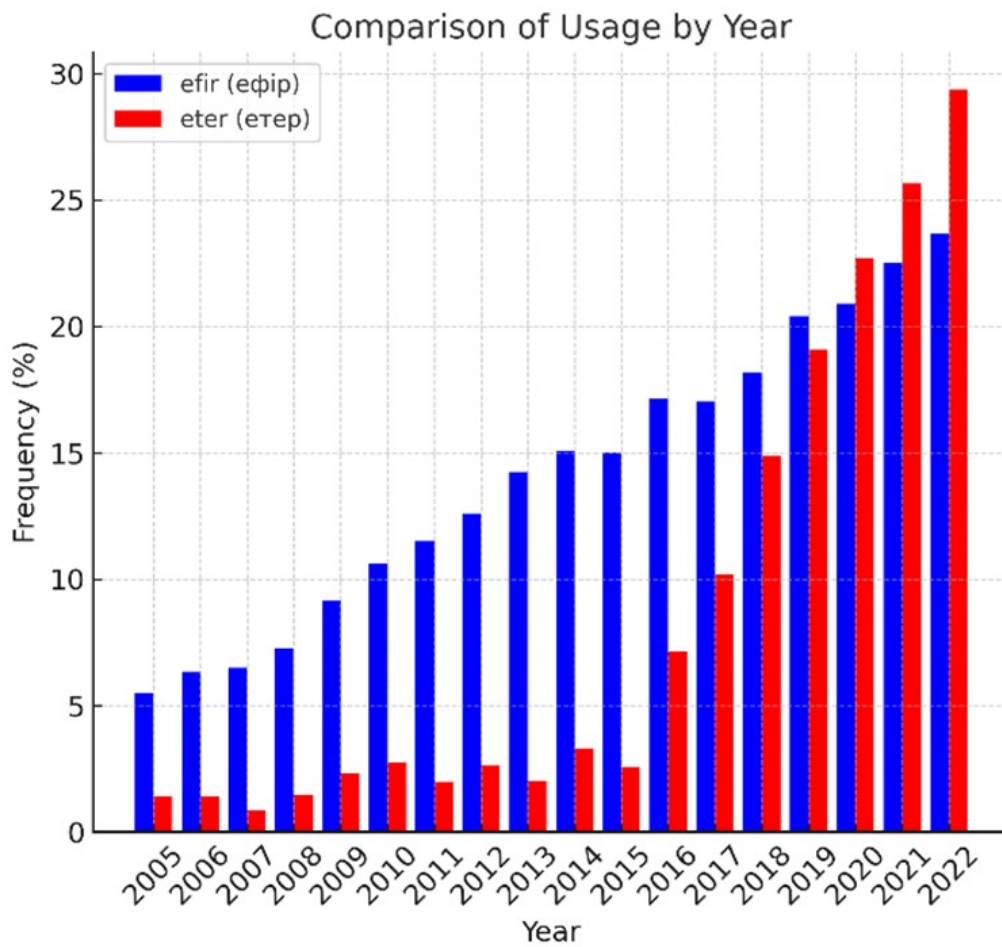**Figure 4:** Diachronic trends in the usage of the lexeme *ефір* [ephir] / *етер* [eter].

Figure 5: Comparative analysis of the frequency of the lexeme *ефір* [ephir] / *етер* [eter].
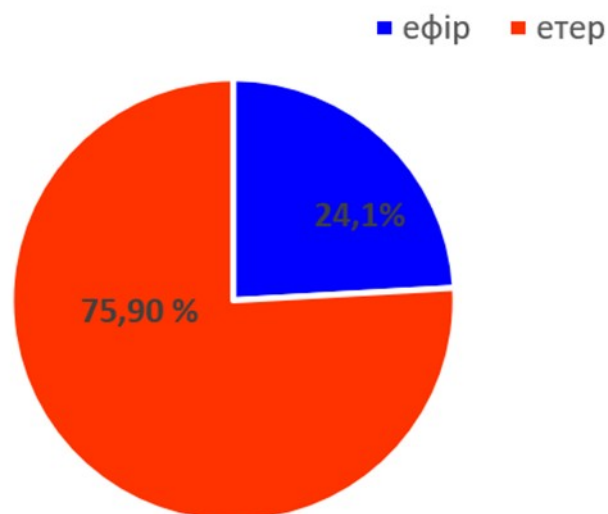
Figure 6: Proportional distribution of the lexeme *ефір* [ephir] / *eme*р [eter] across functional categories.

The surge in media mentions from 2016-2022 aligns with the rise of *emep* [eter] over *eфір* [ephir]. Online journalism and digital media likely played a major role in accelerating this change. 2022 remains a peak year, indicating that *emep* [eter] has solidified as the preferred term.

5. Discussions

5.1. Historical Evolution of Orthographic Norms

Orthographic norms in Ukrainian are far more than mere spelling conventions — they are evidence to the nation's complex history of cultural survival and political disruption. The lexemes *eфір* [ephir] and *emep*, both translating to ether or airwaves in English, embody this dynamic. *Eфір* [ephir], influenced by the Russian "*эфир*", became involved during the Soviet period, while *emep* recollects pre-Soviet Ukrainian traditions. This chapter explores the historical forces — imperial domination, nationalist revival, and modern reforms — that have shaped the trajectory of these terms, reflecting broader struggles over linguistic identity.

Before the Soviet era, Ukrainian orthography was shaped by a jumble of regional influences and imperial pressures. In Western Ukraine, under Austro-Hungarian governance, *emep* [eter] was prevalent, reflecting ties to West Slavic languages like Polish (eter). Literary giants such as Ivan Franko and Lesya Ukrainka employed *emep* [eter] in their works, embedding it in the cultural fabric of Galicia. This form aligned with the phonetic and lexical preferences of local dialects, which resisted the Russification pressures felt in the east (Lesya Ukrainka and Ivan Franko's Works (Collected Letters and Published Works) [23].

Eastern Ukraine, under Russian imperial rule, experienced a different trajectory. Here, *eфір* [ephir] emerged as a linguistic borrowing, particularly in urban centers like Kyiv and Kharkiv, where Russian administrative and cultural influence was strong. The late 19th and early 20th centuries saw attempts to standardize Ukrainian orthography, notably through the efforts of the Ukrainian Scientific Society in Kyiv. However, these initiatives were fragmented until the Soviet period imposed a more uniform approach (Soviet Decrees on Language and Nationality (Soviet Archives) [24].

The 1928 Kharkiv orthography, ratifies during a brief "breath" of Soviet Ukrainization, marked a significant milestone. Led the way by Mykola Skrypnyk, a key advocate for Ukrainian culture, this reform embraced *emep* [eter] as part of a broader push to distinguish Ukrainian from Russian. The policy reflected a nationalist impulse to reclaim linguistic autonomy, but its lifespan was ended by the Stalinist repressions of the 1930s. (Skrypnyk, Mykola. The Ukrainian National Movement and Language Reform in the Soviet Era) [25].

The 1933 Soviet orthographic reform fundamentally altered Ukrainian spelling norms. As part of Stalin's Russification campaign, the reform replaced *emep* [eter] with *eфip* [ephir], aligning Ukrainian orthography with Russian conventions. This shift was justified as a means of fostering "linguistic unity" across the Soviet Union, but it provoked resistance from Ukrainian intellectuals. Linguists like Olena Kurylo and Ivan Ohienko decried the change as a deliberate erosion of Ukrainian identity, arguing that *eфip* [ephir] violated the language's phonetic heritage (Kurylo, Olena. Linguistic Resistance in Soviet Ukraine: Language as a Marker of Identity; Ohienko, Ivan. Ukrainian Linguistic Nationalism in the 20th Century) [24] [26].

Throughout the Soviet era, *eфip* [ephir] dominated official domains – textbooks, newspapers, and state broadcasting — while *emep* [eter] survived in émigré publications and rural speech. The suppression of dissent ensured compliance, but underground literary circles preserved *emep* [eter] as a symbol of resistance. By the late Soviet period, *eфip* [ephir] had become so normalized that its Russian origins were often overlooked by younger generations (Soviet Decrees on Language and Nationality (Soviet Archives) [27].

Ukraine's independence in 1991 ignited a reevaluation of Soviet-imposed norms. The 1993 orthographic guidelines, developed by the Ukrainian Language Institute, aimed to reverse Russification but stopped short of fully endorsing *emep* [eter]. Instead, *eфip* [ephir] was retained due to its entrenched usage, sparking debate among linguists. Yuriy Shevelov, a leading figure in Ukrainian linguistics, criticized this decision in a 1994 essay, arguing that *eфip* [ephir] perpetuated a "colonial legacy". Shevelov advocated for *emep* [eter] as a marker of cultural authenticity, a view echoed by nationalist writers and educators (Shevelov, Yuriy. Language and Nationalism in Ukraine: From the Soviet Era to Independence) [26].

Despite this critique, *emep* [eter] began to reemerge in the post-independence period, particularly in literary and academic circles. Publications like "Krytyka" and "Suchasnist" adopted *emep* [eter] to signal alignment with Ukraine's European aspirations, contrasting it with the Soviet connotations of *eфip* [ephir] ("Suchasnist" and "Krytyka" Magazines. Literary and Cultural Discussions on Language Revival) [28]. This gradual revival laid the groundwork for later reforms.

The 2019 orthographic reform, approved by Ukraine's Cabinet of Ministers on May 22, marked a turning point. Led by Pavlo Hrytsenko of the National Academy of Sciences, the reform reinstated *emep* [eter] as a legitimate variant alongside *eфip* [ephir]. This decision followed years of public debate and coincided with the post-Maidan push for de-Russification after 2014. The reform was framed as a restoration of Ukraine's linguistic heritage, with *emep* [eter] celebrated as a symbol of cultural resilience (Ukrainian Language Institute. The Orthographic Guidelines of 1993 and 2019) [28].

Media responses varied: progressive outlets like "Ukrainska Pravda" hailed the reform as a step toward linguistic sovereignty, while some eastern publications worried about practical challenges, such as reader unfamiliarity with *emep* [eter]. The reform's implementation was bolstered by the 2019 Law on Ensuring the Functioning of the Ukrainian Language as the State Language, which mandated Ukrainian in public life, reinforcing the shift toward *emep* in official discourse ("Ukrainska Pravda", Articles on the 2019 Orthographic Reform). The diachronic corpus data confirm these historical tendencies: while *eфip* [ephir] remained dominant in the Soviet and post-Soviet periods, the gradual revival of *emep* [eter] since the 2000s reflects not only cultural memory but also measurable frequency shifts in contemporary media usage.

5.2. Regional Variations in Usage

Ukraine's linguistic landscape is a mosaic shaped by centuries of geopolitical variability. Western Ukraine, with its Polish and Austro-Hungarian heritage, contrasts sharply with Eastern Ukraine, where Russian influence has left a deep imprint. Central Ukraine, anchored by Kyiv, bridges these extremes, blending national standards with regional diversity. This chapter examines how these regional identities influence the use of *ефір* [ephir] and *етер* [eter], revealing the geographic dimensions of orthographic variation in Ukrainian media.

Analyzing regional usage requires examining media sources from Western, Central, and Eastern Ukraine, considering factors like geographic origin, editorial stance, and audience preferences. While specific data is not provided here, the discussion draws on well-documented sociolinguistic trends to illustrate how historical and cultural contexts shape orthographic choices.

Western Ukraine has historically favored *етер* [eter], a preference rooted in its resistance to Russian linguistic dominance. Under Austro-Hungarian rule, Galician intellectuals like Yurii Fedkovych used *етер* [eter] in poetry and prose, embedding it in the region's literary tradition. This form persisted in local dialects and was reinforced by the area's cultural ties to Central Europe. After the 2019 reform, Western media outlets — such as Lviv-based Zaxid.net — increasingly adopted *етер* [eter], framing it as a rejection of Soviet norms and a return to authentic Ukrainian roots.

This preference is not merely linguistic but ideological, reflecting Western Ukraine's pro-European orientation and post-Maidan nationalism. Local dialects, which retain older Ukrainian forms, further support *етер* [eter], making it a natural choice for both spoken and written communication.

Central Ukraine, particularly Kyiv, exhibits a more balanced approach to *ефір* [ephir] and *етер* [eter]. As the nation's political and cultural center, the region mediates between Western and Eastern influences. Kyiv-based media outlets like Hromadske use *ефір* [ephir] in formal reporting, reflecting its long-standing acceptance, but increasingly employ *етер* [eter] in content tied to cultural or national themes. This duality mirrors Central Ukraine's role as a unifying force within the country.

Audience demographics also shape usage: younger, urban readers in Kyiv are more receptive to *етер* [eter], associating it with modernity and reform, while older generations favor *ефір* [ephir] due to familiarity. The presence of Surzhyk — a Ukrainian-Russian hybrid — further complicates orthographic choices, blending regional and bilingual influences.

In Eastern Ukraine, *ефір* [ephir] remains predominant, a legacy of the region's Russification under Soviet rule. Cities like Kharkiv and Dnipro, with significant Russian-speaking populations, continue to favor *ефір* [ephir] in media and everyday speech. This preference reflects not only linguistic habit but also cultural proximity to Russia, particularly in industrial areas with historical ties to Soviet institutions.

Since the 2019 reform, however, some Eastern outlets have begun incorporating *етер* [eter], especially in content targeting pro-Ukrainian audiences. This shift is tentative, constrained by reader expectations and the region's bilingual environment, where Russian often dominates informal communication.

Regional orthographic preferences are deeply tied to Ukraine's linguistic diversity. Western Ukrainian dialects, by contrast, preserve pre-Soviet forms like *етер* [eter], resisting external influence. Central Ukraine's mix of standard Ukrainian, Surzhyk, and regional idioms creates a fluid orthographic landscape, where both lexemes coexist depending on context.

Ukrainian media outlets often tailor orthography to reflect regional identities. In Western Ukraine, *етер* [eter] signals cultural autonomy and anti-Russian sentiment, appealing to nationalist readers. In Eastern Ukraine, *ефір* [ephir] aligns with the region's historical and linguistic ties to Russia, maintaining continuity with Soviet-era norms. Central Ukraine's media, aiming for national appeal, navigate this divide by using both forms strategically — *ефір* [ephir] for broad accessibility, *етер* [eter] for cultural resonance.

Regional variations in *ефір* [ephir] and *етер* [eter] highlight the interplay of geography, history, and identity in shaping Ukrainian orthography. Western Ukraine's embrace of *етер* [eter] underscores its rejection of Russification, while Eastern Ukraine's preference for *ефір* [ephir] reflects the enduring impact of Soviet influence. Central Ukraine's mixed usage illustrates its role as a linguistic bridge, balancing tradition with reform. These patterns reveal that orthographic norms are not uniform but are deeply embedded in Ukraine's regional diversity. Taken together, the historical trajectory of orthographic reforms and the current regional preferences demonstrate that language change in Ukraine is never purely linguistic. Rather, it reflects overlapping layers of political history, cultural memory, and local identity.

6. Conclusions

The proposed methodology allows for the detection of the dynamics of the implementation of spelling rules with the introduction of the dual norm *ефір* [ephir] / *етер* [eter] as opposed to the non-alternative *ефір* [ephir] in the previous edition of the spelling. The research can serve as a basis for similar observations using corpus methods for other doublet variants of norms to identify the dynamics of establishing one variant with subsequent removal of dual normativity. It may also provide material for further psycholinguistic studies using modern computer technologies. In addition to its theoretical contribution, the proposed methodology can be applied in lexicography (dictionary compilation of variant forms), educational practice (teaching orthography in schools and universities), and language policy (monitoring the effectiveness of orthographic reforms).

The historical journey of *ефір* [ephir] and *етер* [eter] reflects Ukraine's broader struggle for self-definition. *Ефір* [ephir] embodies the legacy of Soviet standardization, while *етер* [eter] represents a reclaimed Ukrainian identity rooted in pre-Soviet traditions. Their coexistence today underscores a tension between practicality and cultural renewal, with orthographic choices serving as proxies for larger debates about history, power, and national belonging.

A comprehensive exploration of the historical and regional factors driving the dual normativity of *ефір* [ephir] and *етер* [eter] has been considered. By tracing their evolution through pivotal historical moments and mapping their usage across Ukraine's diverse regions, the linguistic analysis concerning regional linguistic variations provided a detailed foundation for understanding how orthographic choices in Ukrainian media reflect broader socio-political currents. The interplay of history, geography, and identity underscores the richness of Ukraine's linguistic heritage and its ongoing evolution.

Overall, the study contributes to a deeper understanding of the mechanisms and dynamics of dual normativity in Ukrainian orthography, offering both a replicable methodological framework for corpus-based monitoring of language reforms and a practical foundation for lexicographic, educational, and language policy initiatives aimed at promoting consistent and culturally grounded standardization.

Declaration on Generative AI

The authors have not employed any Generative AI tools in the creation of the manuscript text.

References

[1] S. Bralic, Deviations from the language norm — the Italian language in the digital age, Folia linguistica et litteraria, 2021, 35, pp. 235–257. doi: 10.31902/fll.35.2021.12.
[2] M. Shvedova, R. von Waldenfels, S. Yaryhin, A. Rysin, V. Starko, T. Nikolajenko et al. (2017–2023): GRAC: General Regionally Annotated Corpus of Ukrainian. Electronic resource: Kyiv, Lviv, Jena. [Online]. URL: www.uacorpus.org.

[3] L. A. Krisztina, Panorama of hungarian orthography after the publication of the 12th edition [Körkép a magyar helyesírásról — áttekintés a 12. kiadás után], 2018, Magyar Nyelvor, 142(2), pp. 136–149.

[4] C. Shammas, Acquiring written communication skills as the vernacular standardizes: A case study of an English family's letters, 1560–1700, 2019, Huntington Library Quarterly, 82 (3), pp. 429–482, doi:10.1353/hlq.2019.0022.

[5] L. Villa, Official orthographies, spelling debates and nation-building projects after the fall of the spanish empire, Written Language and Literacy, 2015, 18(2), pp. 228–247, doi:10.1075/wll.18.2.03vil.

[6] P. Hou, Spelling errors in thai made by Chinese students learning Thai as a foreign language, Manusya, 2019, 22 (3), pp. 358–374, doi:10.1163/26659077-02203005I.

[7] Z. Kunch, O. Mykytyuk, L. Kharchuk, M. Hnatyuk, Lexical Norms in Business, Informal and Internet Communication, Studies in Systems, Decision and Control, 2023, 466, pp. 99–120.

[8] K. S. Bondarchuk, O. A. Chumachenko, Positive Changes and Inaccuracies in the Ukrainian 2019 Spelling, Alfred Nobel University Journal of Philology, 2022, 2 (24), pp. 153–164, doi:10.32342/2523-4463-2022-2-24-13.

[9] Z. Kunch, A. Serednytska, I. Vasylyshyn, H. Horodylovska, I. Farion, The Peculiarities of Spelling Rules in Formal, Informal Handwriting and Internet Communication, CEUR Workshop Proceedings, 2022, 3171, pp. 91–103.

[10] R. Romanchuk, V. Vysotska, V. Andrunyk, Analysis System Project for Ukrainian-language Artistic Works to Determine the Text Authorship Attribution Probability, International Scientific and Technical Conference on Computer Sciences and Information Technologies English family's letters 1560–1700, 2019, Huntington Library Quarterly, 82 (3), pp. 429–482.

[11] A. Rey, J.-L. Manguin, C. Olivier, S. Pacton, P. Courrieu, Spelling performance on the web and in the lab, PLoS ONE, 2019, 14(12). doi:10.1371/journal.pone.0226647.

[12] Z. Kunch, O. Lytvyn, I. Mentynska, Modern Ukrainian Electronic Dictionaries: the Problem of Implementing Spelling Changes, CEUR Workshop Proceedings, 2023, 3396, pp. 32–47.

[13] F. Busch, Enregistered spellings in interaction Social indexicality in digital written communication, So zeitschrift fur sprachwissenschaft, 2021, 40 (3), pp. 297–323. doi:10.1515/zfs-2021-2033.

[14] K. Ashwini, M. P. Brundha, V. B. Preejitha, A Survey on Increasing Spelling Errors due to Increase Use of Digital Technology Among Students. Bioscience biotechnology research communications, 2020, 13 (7), 247-250. doi:10.21786/bbrc/13.7/41.

[15] M. Modiano, Identity and standards for English as a European Union lingua franca. World Englishes, 2024, 43 (2), pp. 210–227. doi:10.1111/weng.12646.

[16] L. Krisztina, A panorama of hungarian orthography after the publication of the 12th edition [Körkép a magyar helyesírásról — áttekintés a 12. kiadás után]. Magyar Nyelvor, 2018, 142(2), pp. 136–149.

[17] N. I. Franch, The languages of Spain in the parliamentary discourse of the Spanish transition: an analysis of arguments on linguistic officialdom (Part two) [Les llengües D'Espanya en el discurs parlamentari de la transició Espanyola: una anàlisi dels arguments sobre l'oficialitat lingüística (segona part)]. 2024. Revista de Llengua i Dret, (81), pp. 170–188. doi: 10.58992/rld.i81.2024.4187.

[18] D. Rafiyenko, I. A. Seržant, Ideology and identity in grammar: A diachronic-quantitative approach to language standardization processes in Ancient Greek. In: Müller, Johannes. (Ed.), Connectivity Matters! Social, Environmental and Cultural Connectivity in Past Societies, 2022, 187–210. Sidestone Press.

[19] C. Fandrych, F. Wallner, Functional and stylistic features of spoken advanced learner language: Methodological and conceptual considerations based on GeWiss. Zeitschrift Fur Germanistische Linguistik, 2022, 50(1), 202–239. doi:10.1515/zgl-2022-2053.

[20] Z. Kunch, L. Kharchuk, S. Shekhavtsova, O. Shevchenko, Studying the dynamics of the implementation of orthographic norms in media discourse using corpus linguistics. CEUR

Workshop Proceedings. 2024. Vol. 3851: Proceedings of the 3rd International workshop on social communication and information activity in digital humanities SCIA 2024, Lviv, Ukraine, October 31, 2024.

[21] Ukrainskyi pravopys [Ukrainian spelling]. Naukova Dumka, Kyiv, 2019.

[22] Ukrainskyi pravopys. Proekt novoi redaktsii [Ukrainian spelling. The project for the latest edition], ed. V. Nimchuk. Naukova Dumka, Kyiv, 1999.

[23] A. Liubarets, The Politics of Memory in Ukraine in 2014 Removal of the Soviet Cultural Legacy and Euromaidan Commemorations. Kyiv-Mohyla Humanities Journal. 2016 (3), 197-214. URL: https://ekmair.ukma.edu.ua/items/d54d5143-8b16-44bc-8a31-e9baf5dc819d.

[24] A. Nedashkivska, Language ideological encounters over the new 2019 Ukrainian orthography. URL: https://www.ideopol.org/wp-content/uploads/2023/09/5.-Nedashkivska.pdf.

[25] T. Kuzio, Post-Soviet Ukrainian Historiography in Ukraine. Internationale Schulbuchforschung, 2001, 23(1), 27–42. URL: http://www.jstor.org/stable/43057221.

[26] National Academy of Sciences of Ukraine. The 2019 Ukrainian Orthographic Reform: A New Era for Language and Identity.

[27] O. Kivilyuk, Ukrainian Linguistic Identity: From Soviet Legacy to European Aspirations.

[28] A. Pavlenko, Language and Identity in Post-Soviet Ukraine: The Quest for Linguistic Sovereignty.