

Studying the dynamics of the implementation of orthographic norms in media discourse using corpus linguistics

Zoriana Kunch^{1,†}, Liliia Kharchuk^{1,*,†}, Olena Shevchenko^{2,*,†} and Svitlana Shekhavtsova^{3,*,†}

¹ Lviv Polytechnic National University, Bandery Str. 12, 79013 Lviv, Ukraine

² Odesa Polytechnic National University, 1 Shevchenko Ave., 65037, Odessa, Ukraine

³ Matej Bel University, Národná 12, 974 01 Banská Bystrica, Slovakia

Abstract

The research aims to propose an algorithm for studying the dynamics of the implementation of orthographic norms in the Ukrainian media space during the transitional five-year period following the introduction of changes to the current Ukrainian orthography (2019-2023) using corpus linguistics tools. The General Regionally Annotated Corpus of Ukrainian (GRAC) was used as the experimental field for the study. Through corpus linguistics tools, the method of exhaustive sampling from the sub-corpus "Ukrainian Media Discourse 2000-2023" and statistical data analysis revealed an overall positive trend in the implementation of the new Ukrainian orthographic rules. It was found that the choice of normative form is significantly influenced by subjective factors (the publication's editorial policy, the inclination of certain authors towards the newly introduced norms, their linguistic habits, etc.). In particular, among the media represented in the sub-corpus, the clearly defined media demonstrate consistent progress in adoption before the adoption of new variant technologies forms. The proposed analysis method can serve as a basis for conducting similar observations using corpus methods concerning other orthographic norms to identify the dynamics of establishing the correct variant. This may also provide material for further psycholinguistic research using modern computer technologies

Keywords

Corpus linguistics, text corpus, linguistic-statistical text analysis, orthographic norm, variability

1. Introduction

The issue of codifying Ukrainian orthography is a pressing concern for both scholars and all speakers of the language. External interference and prolonged assimilation have unfortunately led to the destabilization of Ukrainian orthographic norms. The assimilation policies directed towards the Ukrainian language included the introduction of orthographic rules that, by

SCIA-2024: 3rd International Workshop on Social Communication and Information Activity in Digital Humanities, October 31, 2024, Lviv, Ukraine

* Corresponding author.

† These authors contributed equally.

✉ zoriana.y.kunch@lpnu.ua (Z. Kunch); liliia.v.kharchuk@lpnu.ua (L. Kharchuk); shevchenko.o.v@op.edu.ua (O. Shevchenko); svitlana.shekhavtsova@umb.sk (S. Shekhavtsova)

 0000-0002-8924-7274 (Z. Kunch); 0000-0003-0063-1956 (L. Kharchuk); 0000-0003-2313-5939 (O. Shevchenko); 0000-0001-8196-8977 (S. Shekhavtsova)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

interfering with the natural development of the language, effectively carried out a form of linguicide – the deliberate and targeted destruction of a language, aimed at ethnocide (the destruction of a people as a cultural-historical community). This deliberate erosion of national specificity turned the Ukrainian language, as a self-sufficient system, into a victim of a totalitarian society, alienating its speakers from their native language and, consequently, from themselves. Notably, in 1933, a Ukrainian orthography was approved to bring Ukrainians closer to Russian: native Ukrainian features of word formation, grammar, and phonetics were abolished, and tens of thousands of lexical units were borrowed from Russian. Thus, the colonial Russified norm, which permeated all linguistic levels, required reform, and finally, on May 22, 2019, the Cabinet of Ministers of Ukraine approved a new version of Ukrainian orthography, which came into effect on May 30, 2019.

The current version of the Ukrainian orthography revives certain features from the 1928 orthography, which are part of the Ukrainian orthographic tradition and whose restoration has a strong scientific foundation. At the same time, the orthographic commission recognized that the linguistic practices of Ukrainians in the second half of the 20th century and the early 21st century have also become part of the Ukrainian orthographic tradition [1]. Therefore, the implementation of the new version of the Orthography included a transitional period until May 2024, during which each institution independently decided how to apply the new Ukrainian orthographic rules.

Therefore, it is essential to investigate the current trends in the implementation of orthographic norms that were revised in 2019 and analyze the dynamics of adopting these normative forms, particularly in the media space. The media can be considered markers that reflect the current state of the language system. The use of corpus technologies allows for the analysis of a large volume of linguistic material and the effective application of linguistic-statistical analysis to both normative and non-normative variants of language units.

This article proposes, for the first time, a methodology for analyzing the dynamics of orthographic norm implementation in the media space during the adaptation period. The study aims to identify patterns in the usage of the main normative forms that were revised in the 2019 orthography edition using a text corpus, based on the proposed methodology. Achieving this goal involves the following tasks:

1. Select, using the General Regionally Annotated Corpus of the Ukrainian Language (GRAC), excerpts where specific changes from the 2019 Orthography have been implemented or not implemented.
2. Propose a methodology for targeted processing of research materials using corpus mechanisms.
3. Analyze the dynamics of the implementation of orthographic norms in the media discourse within the adaptation period (2019-2023) and conduct a statistical analysis of specific norms.
4. Identify the potential for studying the impact of subjective factors on the introduction of changes to Ukrainian orthography.

The object of the research is the lexemes whose spelling was altered according to the 2019 edition of the Ukrainian orthography, identified in the context of their usage in the media space based on the General Regionally Annotated Corpus of the Ukrainian Language (GRAC) [2].

The research is based on a card index created from the exhaustive selection of material for analysis from the "Ukrainian Media Discourse 2000-2023" corpus.

2. Related Works

The issue of adhering to linguistic norms across different styles of literary language is a subject of scholarly research in several contemporary studies. Given that our research is dedicated to the dynamics of implementing the orthographic norm in the new edition of the Ukrainian orthography of 2019, it is worth noting that prominent Ukrainian linguists have been actively working in this area, including V. Nimchuk [3], O. Ponomariv [4], I. Vykhovanets [5], L. Masenko [6], I. Farion [7], [8], and many other scholars. A fundamental work that examines the evolution of the main linguistic norms, including orthographic ones, and identifies the extralinguistic reasons for the distortion of the Ukrainian linguistic norm throughout the 20th century is the monograph "Language Norm: Destruction, Search, Restoration" [8]. Recent studies by young, applied linguistics scholars are also noteworthy. For example, in A. Bordovska's research, the issue of orthographic and lexical variability in journalism during the World War II period is addressed [9]. The researcher conducted a study of the variability of the linguistic norm in newspapers of different political orientations during World War II using the General Regionally Annotated Corpus of the Ukrainian Language (GRAC) and traced the influence of ideology on the formation of the linguistic norm. As A. Bordovska aptly notes, "The task of linguists is to develop a system of rules that will not cause societal rejection due to habit with established norms, while also returning speakers to their native sources" [9].

The topic of applied analysis of normativity and adherence to or violation of linguistic laws remains consistently relevant and a priority, thus attracting the attention of both Ukrainian and foreign researchers. An analysis of recent studies confirms that scholars are focused on various aspects of this subject, as evidenced by some studies [10], [11], [12]. Specifically, researchers address the problem of developing general principles for studying adherence to linguistic norms at different levels and analyzing specific instances of such violations [13], [14], [15]; issues of adherence to orthographic norms in different languages: English [16], [17], [18], Hungarian [19], Thai [20], Spanish [21], [22], Greek [23], [24], Scottish [25], German [26], among others.

However, the methodology for conducting a scientific analysis of the dynamics of adherence to orthographic norms in Ukrainian literary language, which changed according to the new edition of the Ukrainian Orthography in 2019, particularly in the media space, has not yet been studied.

It is worth noting that the issue of automatic processing of natural language information is becoming increasingly significant, and applied linguistics technologies are progressively offering their technical capabilities for processing, storing, and selecting linguistic material. Linguistic corpora have quickly evolved into invaluable research resources because they are based on the use of large volumes of empirical material and the involvement of advanced computer technologies. These tools enable researchers to obtain objective conclusions about the functioning of linguistic units, and corpus data analysis contributes to the formulation of qualitatively new conclusions on the subject matter under investigation.

3. Methods and Materials

To conduct a linguistic-statistical analysis and determine the dynamics of the implementation of spelling norms in the media space during the adaptation period (2019-2023), we used data from the annotated GRAC linguistic corpus. This corpus was chosen as our primary data source due to the large amount of information it contains and the ability to create a sub-corpus of journalistic works from the required time period. The choice of the media space for analysis can be explained by the fact that "the linguistic features of newspaper genres unite characteristics that make the language of newspapers a means of mass communication: addressing a mass audience, orientation towards diverse linguistic preferences, and publicistic nature. Genre features influence the choice of vocabulary, syntactic constructions, and expressive means for headlines" [27].

The methodology of our research combines general theoretical methods with those of applied linguistics:

- Statistical and corpus methods. This involves a set of techniques and principles for the collection, systematization, processing, and interpretation of statistical data to draw scientific and practical conclusions.
- Content analysis method. A qualitative-quantitative method of text processing that includes both the quantitative characterization of content and the interpretation of results. This method involves preparing the material and selecting texts based on the date or place of writing.
- Structural (descriptive) method. The core of this method is comparative analysis, which allows us to systematize, classify, and describe the collected material base.
- Methods of generalization and induction. These enabled the linguistic-statistical analysis of the obtained results.

As mentioned earlier, we identified lexemes that underwent orthographic changes according to the 2019 edition of the Ukrainian orthography by examining their usage in the media space, utilizing the General Regionally Annotated Corpus of the Ukrainian Language (GRAC). This corpus is a representative and structured collection of Ukrainian texts, accompanied by a program that allows users to create their sub-corpora, search for words, grammatical forms, and their combinations, and process search results. The tool enables sorting, creating balanced samples, obtaining various statistical information [2], and conducting linguistic analysis.

4. Results and Discussions

4.1. Main Changes in the 2019 Edition of the Ukrainian Orthography

The main changes in the 2019 edition of the Ukrainian orthography, compared to the 1993 orthography, which we have included in our research, can be observed in Table 1.

Table 1

Comparative Characteristics of Orthographic Changes Used as the Basis for the Research

Norm Essence	1993 Orthography (was)	2019 Orthography (became mandatory norm)
The spelling of the word "священник" [sviashchennyk] (priest)	The word "свящеНик" [sviashchenyk] (priest) was an exception and was spelled without the doubling of 'н'.	The word "свящеННик" [sviashchennyk] (priest) is now spelled with the doubling of 'н', as in other words where identical consonants meet at the boundary of the root and suffix.
The spelling of the word "госпіс" [hospis] (hospice)	The word "Хоспіс" [khospis] (hospice) was spelled with 'х' at the beginning of the word.	The word is now spelled with 'г' at the beginning: "Госпіс" [hospis] (hospice).
The spelling of the word "проект" [proiekt] (project) "проекція" [proiektiia] (projection) and derivatives	The words "проект" [proekt] (project), "проекція" [proektsiia] (projection) were spelled without the letter 'й' despite the pronunciation.	The words "проєкт" [proiekt] (project), "проєкція" [proiektiia] (projection) and their derivatives are now pronounced and spelled with the letter 'й'.

4.2. Selection of Examples of Spelling of Certain Modified Lexemes

For this study, we extracted a selection of quotes from the "Ukrainian Media Discourse 2000-2023" sub-corpus of the General Regionally Annotated Corpus of the Ukrainian Language [2]. The selection includes quotes featuring both variants (old and new norms) of the most commonly used lexemes, such as "священик" [sviashchenyk] (priest) and "священник" [sviashchennyk] (priest); "проект" [proekt] (project) and "проєкт" [proiekt] (project); "проекція" [proektsiia] (projection) and "проекція" [proiektiia] (projection); "хоспіс" [khospis] (hospice) and "госпіс" [hospis] (hospice). Since the aim of this research is to propose a methodology for identifying patterns in the use of key normative forms that were changed in the 2019 spelling edition using a text corpus, we limit our analysis to just three spelling norms. This allows for a more comprehensive and detailed analysis of each change and enables us to more thoroughly characterize the patterns associated with the implementation of the new norm. To identify these words in all grammatical forms, we searched each lexeme in concordance using the following query formula: [lc="lexeme" | lemma_lc="lexeme"]. Thus, the material for observing the dynamics of the implementation of the mandatory orthographic norm consisted of:

- 24,364 quotes with the word "священик" [sviashchenyk] (priest) following the old orthographic rules and 5,267 quotes with the word "священник" [sviashchennyk] (priest) spelled according to the new orthography (see Table 2);
- 1,157 quotes with the lemma "проекція" [proektsiia] (projection) and 221 quotes using the lemma "проекція" [proiektiia] (projection) (see Table 3);
- 96,324 quotes with the lemma "проєкт" [proiekt] (project) and 540,430 instances of the lemma "проект" [proekt] (project) (see Table 4);

- 1,407 quotes with the lemma "хоспіс" [khospis] (hospice) and 81 quotes with the word "госпіс" [hospis] (hospice) (see Table 5).

Table 2

Frequency of Usage of Lexemes "священик" [sviashchenyk] (priest) / "священник" [sviashchennyk] (priest) in the Analyzed Sub-corpus

Parameters	Hits	Per million
[lc="священик" lemma_lc="священик"]	24 364	25.85
[lc="священник" lemma_lc="священник"]	5 267	5.59

Table 3

Frequency of Usage of Lexemes "проєкція" [proektsiia] (projection) / "проекція" [proiektssiia] (projection) in the Analyzed Sub-corpus

Parameters	Hits	Per million
[lc="проєкція" lemma_lc="проєкція"]	1 157	1.23
[lc="проекція" lemma_lc="проекція"]	221	0.23

Table 4

Frequency of Usage of Lexeme "проект" [proekt] (project) / "прое́кт" [proiekt] (project) in the Analyzed Sub-corpus

Parameters	Hits	Per million
[lc="проект" lemma_lc="проект"]	96 324	102.21
[lc="прое́кт" lemma_lc="прое́кт"]	540 430	573.43

Table 5

Frequency of Usage of Lexemes "хоспіс" [khospis] (hospice) / "госпіс" [hospis] (hospice) in the Analyzed Sub-corpus

Parameters	Hits	Per million
[lc="хоспіс" lemma_lc="хоспіс"]	1 407	1.49
[lc="госпіс" lemma_lc="госпіс"]	81	0,09

As can be seen from the statistical data presented above, only the lexeme "прое́кт" [proiekt] (project) shows a significant advantage over the outdated variant "проект" [proiekt] (project) with 540,430 instances of usage compared to 96,324 instances of "проект" [proiekt] (project). The weakest dynamic in usage is observed for the lexeme "госпіс" [hospis] (hospice) which, according to the recommendations of the Orthographic Commission, should be used instead of the older variant "хоспіс" [khospis] (hospice). There were only 81 instances of "госпіс" [hospis] (hospice) compared to 1,407 instances of the Russified lexeme "хоспіс" [khospis] (hospice). Since our sample covers the period from 2000, when the new normative variants had not yet

been officially approved, we will narrow our search to the chronological boundaries defined in the research objectives and proceed to the second stage of our analysis.

At this stage, we create queries regarding the use of the analyzed lexemes to observe patterns and trends in their usage within a chronological context. By setting the temporal boundaries for the usage of the specified lexemes (2019-2023) according to the query formula [lc="requested lexeme" | lemma_lc="requested lexeme"] within <doc (date="2019" | date="2020" | date="2021" | date="2022" | date="2023") />, we determine the frequency of usage of the aforementioned words specifically during the transitional period of the new orthographic rules. At this stage of observation, we operate with the following statistical data:

- 4,857 citations using the word "священик" [sviashchenyk] (priest) according to the old orthographic rules and 4,102 citations with the word "священник" [sviashchennyk] (priest) according to the new edition (54.2% vs. 45.8%) (see Table 6);
- 332 citations with the lemma "проєкція" [proektsiia] (projection) and 206 citations with the lemma "проєкція" [proiektsiia] (projection) (61.7% vs. 38.3%) (see Table 7);
- 108,399 citations with the lemma "проект" [proekt] (project) and 94,903 citations with the lemma "проект" [proiekt] (project) (53.3% vs. 46.7%) (see Table 8);
- 348 citations with the lemma "хоспіс" [khospis] (hospice) and 28 citations with the word "госпіс" [hospis] (hospice) (92.6% vs. 7.4%) (see Table 9).

This analysis allows us to examine the adoption of new orthographic norms and their prevalence compared to previous standards during the transitional period.

Table 6

Frequency of the Use of the Lexemes "священик" [sviashchenyk] (priest) / "священник" [sviashchennyk] (priest) from 2019 to 2023

Parameters	Hits	Per million
[lc="священик" lemma_lc="священик"] within <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	4 857	5.15
[lc="священник" lemma_lc="священник"] within <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	4 102	4.35

Table 7

Frequency of the Use of the Lexemes "проєкція" [proektsiia] (projection) / "проєкція" [proiektsiia] (projection) from 2019 to 2023

Parameters	Hits	Per million
[lc="проєкція" lemma_lc="проєкція"] within <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	332	0.35

[lc="проєкція" lemma_lc="проєкція"] within <doc (date="2020" date="2021" date="2022" date="2023" date="2019") />	206	0.22
--	-----	------

Table 8

Frequency of the Use of the Lexemes "проєкт" [proekt] (project) / "проєкт" [proiekt] (project) from 2019 to 2023

Parameters	Hits	Per million
[lc="проєкт" lemma_lc="проєкт"] within <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	108	115.02
[lc="проєкт" lemma_lc="проєкт"] within <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	94 903	100.7

Table 9

Frequency of the Use of the Lexemes "хоспіс" [khospis] (hospice) / "роспіс" [hospis] (hospice) from 2019 to 2023

Parameters	Hits	Per million
[lc="хоспіс" lemma_lc="хоспіс"] within <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	348	0.37
[lc="роспіс" lemma_lc="роспіс"] протягом <doc (date="2019" date="2020" date="2021" date="2022" date="2023") />	28	0,03

Thus, we observe that the frequency of use for each lexeme in the sub-corpus differs significantly both overall and from 2019 to 2023, during the transition period to the new orthographic norm. The best dynamics is again seen with the lemma "проєкт" [proiekt] (project): 46.7% versus 53.3%, while the old spelling "хоспіс" [khospis] (hospice) significantly prevails over the new normative form "роспіс" [hospis] (hospice): 92.6% versus 7.4%.

In the next phase of the chronological analysis, we examine the frequency of use of each variant of the studied lexemes in different years of the research period to trace the presence or absence of a trend toward an increase in the usage of variants that adhere to the 2019 orthographic recommendations. Table 10 provides a summary analysis of the frequency of use of the analyzed lexemes over the specified period.

Table 10

Frequency of Use of Analyzed Lexemes from 2019 to 2023

Year	Parameters	Hits	Percent
2019	[lemma_lc="священик"] within <doc (date="2019") />	2 528	77,5
	[lemma_lc="священник"] within <doc (date="2019") />	734	22,5
	[lemma_lc="проєкція"] within <doc (date="2019") />	117	84,2
	[lemma_lc="проєкція"] within <doc (date="2019") />	22	15,8
	[lemma_lc="проєкт"] within <doc (date="2019") />	47 451	80,7

	[lemma_lc="проект"] within <doc (date="2019") />	11 329	19,3
	[lemma_lc="хоспіс"] within <doc (date="2019") />	113	97,4
	[lemma_lc="госпіс"] within <doc (date="2019") />	3	2,6
2020	[lemma_lc="священик"] within <doc (date="2020") />	1 207	41,1
	[lemma_lc="священник"] within <doc (date="2020") />	1 728	58,9
	[lemma_lc="проекція"] within <doc (date="2020") />	117	66,1
	[lemma_lc="проекція"] within <doc (date="2020") />	60	33,9
	[lemma_lc="проект"] within <doc (date="2020") />	28 802	43,8
	[lemma_lc="проект"] within <doc (date="2020") />	36 974	56,2
	[lemma_lc="хоспіс"] within <doc (date="2020") />	99	90
	[lemma_lc="госпіс"] within <doc (date="2020") />	11	10
2021	[lemma_lc="священик"] within <doc (date="2021") />	639	47,1
	[lemma_lc="священник"] within <doc (date="2021") />	719	52,9
	[lemma_lc="проекція"] within <doc (date="2021") />	54	46,2
	[lemma_lc="проекція"] within <doc (date="2021") />	63	53,8
	[lemma_lc="проект"] within <doc (date="2021") />	21 325	40,5
	[lemma_lc="проект"] within <doc (date="2021") />	31 349	59,5
	[lemma_lc="хоспіс"] within <doc (date="2021") />	89	90,8
	[lemma_lc="госпіс"] within <doc (date="2021") />	9	9,2
2022	[lemma_lc="священик"] within <doc (date="2022") />	483	34,4
	[lemma_lc="священник"] within <doc (date="2022") />	921	65,6
	[lemma_lc="проекція"] within <doc (date="2022") />	44	41,9
	[lemma_lc="проекція"] within <doc (date="2022") />	61	58,1
	[lemma_lc="проект"] within <doc (date="2022") />	10 821	41,5
	[lemma_lc="проект"] within <doc (date="2022") />	15 251	58,5
	[lemma_lc="хоспіс"] within <doc (date="2022") />	47	90,4
	[lemma_lc="госпіс"] within <doc (date="2022") />	5	9,6
2023	[lemma_lc="священик"] within <doc (date="2023") />	27	64,3
	[lemma_lc="священник"] within <doc (date="2023") />	15	35,7
	[lemma_lc="проекція"] within <doc (date="2023") />	0	0
	[lemma_lc="проекція"] within <doc (date="2023") />	0	0
	[lemma_lc="проект"] within <doc (date="2023") />	27	73
	[lemma_lc="проект"] within <doc (date="2023") />	10	27
	[lemma_lc="хоспіс"] within <doc (date="2023") />	0	0
	[lemma_lc="госпіс"] within <doc (date="2023") />	0	0

Based on Figure 1, we observe the dynamics of the frequency of use of the analyzed lemmas in different segments of the sub-corpus, with an indication of the time frames for the use of the lexeme (2019-2023 — the years of adaptation to the new orthographic norm) in relative terms.

RELATIVE FREQUENCY OF USE OF ANALYZED LEXEMES FROM 2019 TO 2023

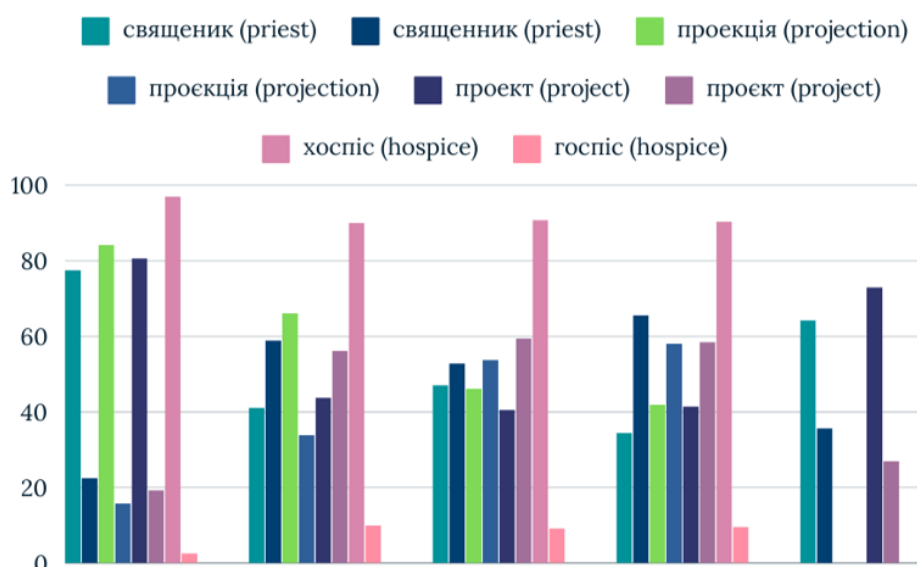


Figure 1: Relative Frequency of Use of Analyzed Lexemes from 2019 to 2023

Based on the comprehensive analysis of the chronology of the use of outdated and new orthographic variants of the words under analysis, we can conclude that, overall, there is a positive trend in the introduction of the analyzed orthographic changes into the media space. However, different lexemes show varying degrees of implementation of the corresponding norm: from 2.6% (госпіс [hospis] (hospice), 2019) to 65.6% (священник [sviashchennyk] (priest), 2022). We assume a significant influence on the choice of spelling variant from subjective factors, such as the editorial policy of publications, spelling habits, spelling experience of individual journalists, etc. To investigate this hypothesis, our methodology includes the next stage of analysis.

4.3. Identifying the Influence of Subjective Factors on the Choice of Norm Variant

Since an important aspect of implementing orthographic changes during the adaptation period, as we have noted, was the subjective factor, it is crucial to observe which specific media outlets more consistently adopted the new spelling rules. Corpus technologies allow us to identify the frequency of use of specific individual spellings in particular media texts.

Tracing the frequency of use of the normative lexeme “священник” [sviashchennyk] (priest) in some media, we notice that in the transition period (2019-2023) this lexeme in more than half of the identified cases is used online-media “LB.ua” (93 uses in 2019, 550 times in 2020, 95 uses in 2021, and 202 uses in 2022). Other publications presented in the corpus demonstrate a significantly lower frequency of use of this lexeme (see Table 11 and Figure 2).

Table 11

The chronology of the lexeme frequency “священник” [sviashchennyk] (priest) in the publications of the sub-corpus "Ukrainian Media Discourse 2000-2023"

№	Year	Title	Frequency
1	2020	LB.ua	550
2	2022	LB.ua	202
3	2022	UNIAN.NET	199
4	2020	High Castle	191
5	2020	HB	176
6	2020	UNIAN.NET	153
7	2020	Public television	145
8	2019	Reporter	121
9	2022	Public television	118
10	2021	UNIAN.NET	117
11	2020	Young Ukraine	115
12	2021	LB.ua	95
13	2019	HB	93
14	2021	Public television	81
15	2019	High castle	77

FREQUENCY OF USE OF THE LEXEME "СВЯЩЕННИК" (PRIEST) IN DIFFERENT MEDIA

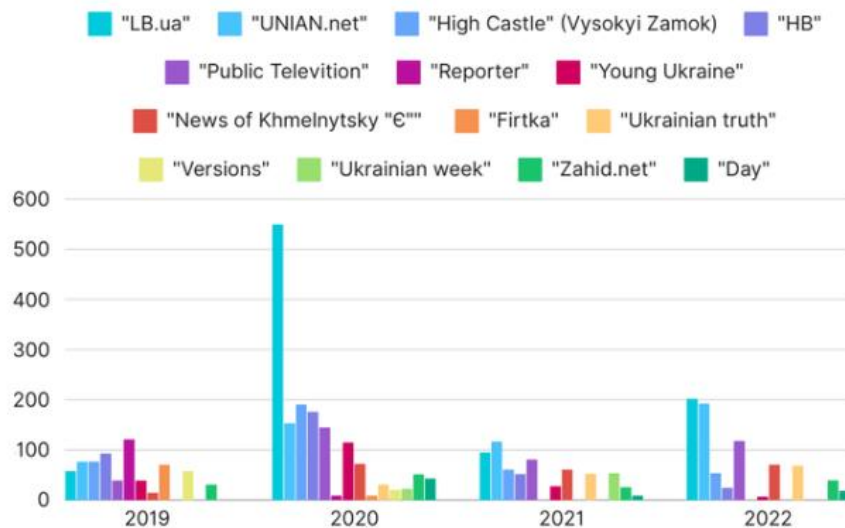


Figure 2: Frequency of Use of the Lexeme "священник" [sviashchennyk] (priest) in Different Media

Using a similar model, we observe the use of other analyzed lexemes, whose spelling has changed with the introduction of a new edition Ukrainian spelling of 2019. The spelling dynamics norm implementation in individual mass media are summarized in Table 12 and Table 13 below and schematically depicted in Figure 3, Figure 4, Figure 5.

Table 12

The chronology of the lexeme frequency "проєкція" [proiektiia] (projection) in the editions of the sub-corpus "Ukrainian Media Discourse 2000-2023"

№	Year	Title	Frequency
1	2022	Ukraine Truth	20
2	2021	Ukrainian Week	12
3	2022	HB	10
4	2021	LB.ua	9
5	2020	UNIAN.NET	9
6	2020	Hmarochos	8
7	2020	HB	8
8	2022	Ukrainian Week	7
9	2021	HB	7
10	2020	Young Ukraine	6
11	2020	Mirror of week	6
12	2022	Military	5
13	2021	Hmarochos	5
14	2021	Ukraine Truth	5
15	2021	Public TV	5

FREQUENCY OF USE OF THE LEXEME "ПРОЕКЦІЯ" (PROJECTION) IN DIFFERENT MEDIA

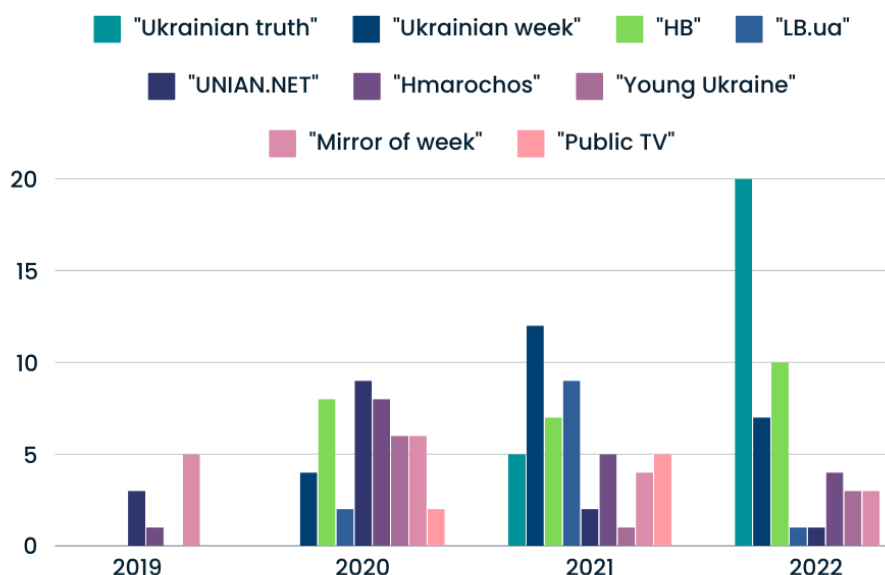


Figure 3: Frequency of the Lexeme "проєкція" [proieksiia] (projection) in Different Media

As you can see, observing the position on writing the word "проєкція" [proieksiia] (projection) in accordance with the new edition of the Ukrainian spelling in 2019. prefer such publications as "Ukrainian truth", "Ukrainian Week", "NV", "LB.ua", and "UNIAN.NET". If we talk about the frequency of using the widely used the project's lexeme, then we can trace the extremely high frequency of its use, compared to the use of other words whose spelling has changed in 2019. Our observations showed that the lexeme "проєкт" [proiekt] (project) during the adaptation years of the introduction of the mandatory spelling norm with a high frequency of use – 9014 times (2020) – used in online media "LB.ua". Note that the high frequency of the analyzed lexeme usage in this media has been observed throughout the adaptation period – 1812 uses in 2019, 3397 – 2020 and 5570 – 2021. Other publications presented in the corpus also demonstrate a rather high this lexeme frequency: the NV magazine – 5191 (2020), 4492 (2021) and 1463 the lexeme usages in 2022; "Hmarochos" – 2884 (2020) 1469 (2022), 1111 uses in 2019; "Hmarochos" online media – 3488 (2021), 2884 (2020), 1469 (2022). Our observations regarding the frequency of the lexeme project are presented in Table 13 and Figure 4.

Table 13

The chronology of the lexeme frequency "проєкт" [proiekt] (project) in the editions of the sub-corpus "Ukrainian Media Discourse 2000-2023"

№	Year	Title	Frequency
1	2020	LB.ua	9014
2	2021	LB.ua	5570

3	2020	HB	5191
4	2021	HB	4492
5	2021	Hmarochos	3488
6	2022	LB.ua	3397
7	2020	Hmarochos	2884
8	2020	Economical truth	2512
9	2021	Economical truth	2456
10	2021	Zhytomyr. info	2288
11	2020	Zhytomyr. info	2288
12	2020	Public TV	1940
13	2019	LB.ua	1812
14	2021	Public TV	1784
15	2021	European truth	1767

FREQUENCY OF USE OF THE LEXEME "ПРОЄКТ" (PROJECT) IN DIFFERENT MEDIA

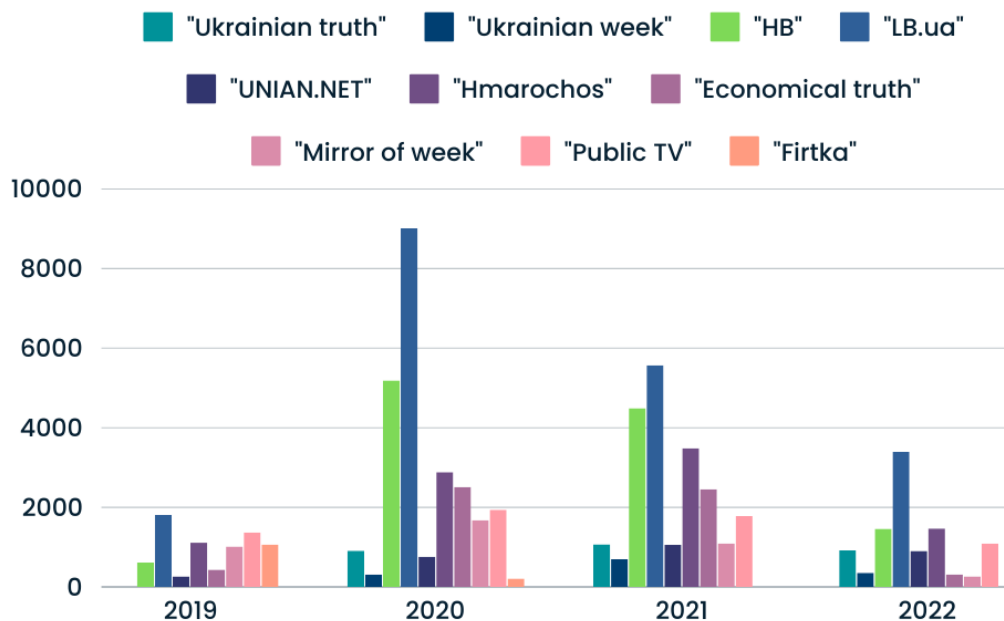


Figure 4: Frequency of the Lexeme "проект" [proiekt] (project) in Different Media

Because the lexeme "госпіс" [hospis] (hospice) is not widely used because it has synonymous correspondences in the Ukrainian language (лікарня [likarnia] (hospital), шпиталь [shpytal] (hospital), госпіталь [hospital] (hospital), клініка [klinika] (clinic)), the use of this word demonstrates a much lower frequency than, for example, the word project. Despite this, the corpus analysis tools showed that in the media discourse, they consistently adhere to the innovations regarding the spelling of this lexeme. In particular, such media as "Zaxid.net",

"European Truth", "Mirror of Week", "Reporter", "Public Television", "High Castle", and "NV" since 2019, although with low frequency, but this word is used in accordance with the new spelling. The frequency of the "rocnic" [hospis] (hospice) lexeme in various media can be seen in Figure 5.

FREQUENCY OF USE OF THE LEXEME "ГОСПИС" (HOSPICE) IN DIFFERENT MEDIA

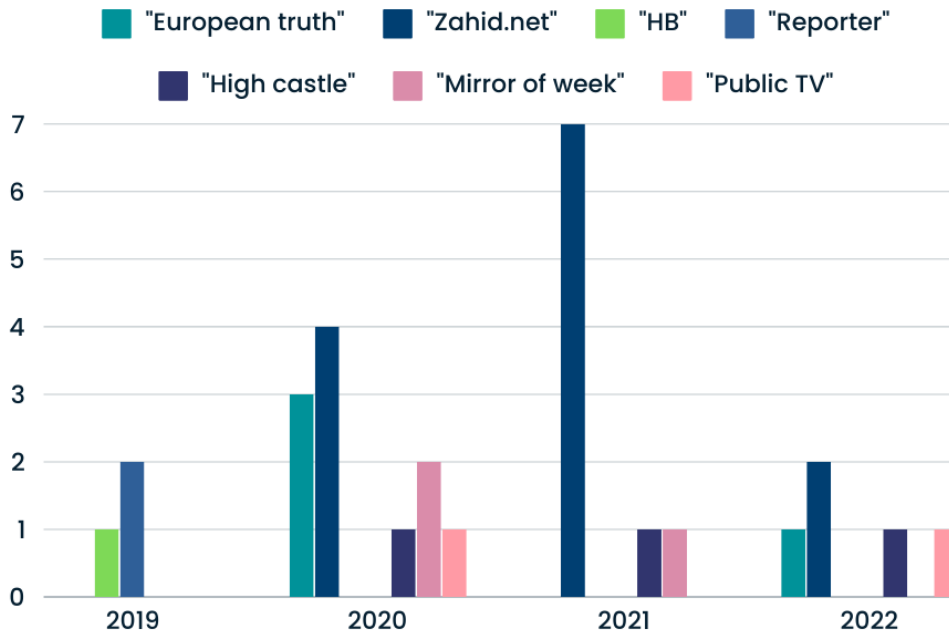


Figure 5: Frequency of the Lexeme "rocnic" [hospis] (hospice) in Different Media

Therefore, the study of the frequency of the analyzed lexemes (priest, projection, project, hospice) in various editions, which is presented above, gives reason to assert that the use of the mandatory norm in accordance with the edition of the Ukrainian spelling 2019 depends on the linguistic consciousness and linguistic the authors preferences of publications and their editors, that is, on subjective factors.

5. Conclusion

The proposed research algorithm utilizing corpus linguistics to analyze the dynamics of the introducing spelling norms in media discourse makes it possible to comprehensively analyze with the help of software products various aspects of the implementation of spelling postulates introduced by the new edition of "Ukrainian Spelling" during the transitional five-year period (2019-2023). Through corpus technologies, the study involved continuous material selection from the sub-corpus "Ukrainian Media Discourse 2000-2023" and conducted a statistical analysis of the data, revealing a clear positive trend in the implementation and practical application of

new Ukrainian spelling rules in the media. Based on a comprehensive analysis of the chronology of using outdated and new spelling variants of the words taken for analysis, we can state that the analyzed lexemes reveal a different degree of implementation of new spelling postulates. In particular, with the help of case technologies, it was found that the lexeme "проєкт" [proiekt] (project) demonstrates superiority over the outdated "проєкт" [proekt] (project) version. The dynamics of using the "роєчнїк" [hospis] (hospice) lexeme turned out to be the lowest, since this word belongs to the underused and has a number of synonymous correspondences. The analyzed material allowed to reveal that a subjective factor (editorial policy of the publication, adherence to the introduced norm of individual authors, their language habits, etc.) has a significant influence on the choice of the normative form. From among the media presented in the sub-corpus, it is possible to clearly identify publications that, in a gradual but clear progression, introduce new spelling rules into use. The analysis of the individual variants frequency in media texts made it possible to single out progressive, in our opinion, mass media that do not neglect the introduction of new spelling rules and each time increase the use of variants that meet the recommendations of the new Ukrainian spelling of 2019. Such media include LB.ua, UNIAN.NET, HB, "Economical Truth", "Public TV", "Hmarochos", "Ukrainian Truth", "Ukrainian Week" and "Young Ukraine".

The proposed method of analysis can become the basis for conducting similar observations using corpus methods relative to other orthographic norms to identify the dynamics of the becoming permanent of the correct variant. It can also serve as material for further psycholinguistic research using modern computer technology.

References

- [1] Ukrainian Orthography 2019. URL: https://uk.wikipedia.org/wiki/%D0%A3%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D1%81%D1%8C%D0%BA%D0%B8%D0%B9_%D0%BF%D1%80%D0%B0%D0%B2%D0%BE%D0%BF%D0%B8%D1%81_2019_%D1%80%D0%BE%D0%BA%D1%83 [in Ukrainian].
- [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. URL: www.uacorpus.org.
- [3] V. Nimchuk, Pro hrafiku ta pravopys yak elementy etnichnoi kultury: istoriia g [On Graphical and Orthographic Elements as Aspects of Ethnic Culture: The History of Phonology]. *Linguistics*, 1990, 6 [in Ukrainian].
- [4] O. Ponomaryov, Pravopys: khto proty? [Spelling: Who Is Against?]. 2001, *Svoboda*, 17, 21 [in Ukrainian].
- [5] I. Vyhovanets, Nenaukovi prystrasti navkolo ukrainskoho pravopysu [Non-Scientific Passions Surrounding Ukrainian Orthography]. *Ukrainian Language*, 2004. 2, pp. 3–24 [in Ukrainian].
- [6] *Ukrainska mova u XX storichchi: istoriia linhvotsydu: dokumenty i materialy [Ukrainian Language in the 20th Century: History of Linguocide: Documents and Materials]* / Edited by: L. Masenko et al. Kyiv: Publishing House "Kyiv-Mohyla Academy", 2005. 399 p. [in Ukrainian].
- [7] I. Faryon, Pravopys — korset movy? Ukrainskyi pravopys yak kulturno-politychnyi vybir [Orthography — A Corset for the Language? Ukrainian Orthography as a Cultural-Political Choice]. Monastery of St. Ivan Lavra; Svichado, 2004 [in Ukrainian].

- [8] I. Farion, *Movna norma: znyshchennia, poshuk, vidnova* [Language norm: destruction, search, restoration]. Ivano-Frankivs'k: Misto NV, 2010 [in Ukrainian].
- [9] A. Bordovska, *Pravopysna ta leksychna variantnist u publitsystytsi periodu druhoi svitovoi viiny (na materiiali HRAKu)* [The Spelling and Lexical Variation in Newspapers and Magazines Published During the Second World War (Based on GRAC)]. *Linguistic and Conceptual Views of the World*, 2024, 75 (1), pp. 36–59. doi: 10.17721/2520-6397.2024.1.03. URL: <https://ouci.dntb.gov.ua/en/works/4kJWGVN7/> [in Ukrainian].
- [10] 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.
- [11] P. Milin, N. Bermel, P. Blevins, *Cognitive approaches to uniformity and variability in morphology*. *Cognitive Linguistics*, 2024, 35 (2), pp. 167–176, doi: 10.1515/cog-2024-0027.
- [12] A. Curzan, M. Robin, Queen, Kristin VanEyck, and Rachel Elizabeth Weissler, *Language Standardization & Linguistic Subordination*. *Dædalus*. SUMMER 2023. URL: https://www.amacad.org/sites/default/files/publication/downloads/Daedalus_Su23_02_Curzan-Queen-VanEyck-Weissler_1.pdf.
- [13] Z. Kunch, O. Mykytyuk, L. Kharchuk, M. Hnatyuk, *Lexical Norms in Business, Informal and Internet Communication*, 2023. *Studies in Systems, Decision and Control*, 466, pp. 99–120.
- [14] Z. Kunch, A. Serednytska, I. Vasylyshyn, H. Horodylovska, I. Farion, *The Peculiarities of Spelling Rules in Formal, Informal Handwriting and Internet Communication*, 2022, *CEUR Workshop Proceedings*, 3171, pp. 91–103.
- [15] O. Lytvyn, L. Kharchuk, O. Mykytyuk, I. Mentynska, *Practice of compliance of punctuation rules in Internet communication*. *CEUR Workshop Proceedings*. 2022. Vol. 3171: *Computational Linguistics and Intelligent Systems 2022 : Proceedings of the 6th International conference on computational linguistics and intelligent systems (COLINS 2022)*. Vol. 1: Main conference, Gliwice, Poland, May 12-13, 2022, pp. 177–191.
- [16] C. Shammas, *Acquiring written communication skills as the vernacular standardizes: A case study of an English family's letters 1560–1700*. *Huntington Library Quarterly*, 2019, 82 (3), pp. 429-482. doi:10.1353/hlq.2019.0022.
- [17] 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.
- [18] E. Suhrob, K. Vasila, *Parts Of Speech And Sentence Structure In English Grammar*. *Galaxy International Interdisciplinary Research Journal*, 2022, 10(7), 156–160.
- [19] L. Krisztina, *A panorama of hungarian orthography after the publication of the 12th edition* [Körkép a magyar helyesírásról – áttekintés a 12. kiadás után]. *Magyar Nyelv*, 2018, 142(2), pp. 136–149.
- [20] P. Hou, *Spelling errors in thai made by Chinese students learning Thai as a foreign language*, 2019, *Manusya*, 22 (3), pp. 358–374. doi:10.1163/26659077-02203005I.
- [21] J. Del Valle, (Ed.). *A Political History of Spanish: The Making of a Language*. Cambridge: Cambridge University Press 2013. doi: 10.1017/CBO9780511794339.
- [22] 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.
- [23] J. L. García Ramón, Grammatical und lexical structures on change in Postclassical Greek: local dialects and supradialectal tendencies. In: D. Rafiyenko and I. A. Seržant, eds. *Contemporary Approaches to Postclassical Greek*. Trends in Linguistics series. Berlin, New York: De Gruyter, 2020, 303–336.
- [24] 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.
- [25] L. Hall-Lew, R. Friskney, J. M. Scobbie, Accommodation or political identity: Scottish members of the UK Parliament. *Language Variation and Change*, 2017, 29(3), 341–363. doi:10.1017/S0954394517000175.
- [26] 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.
- [27] S. Albota, Z. Kunch, L. Kharchuk, M. Hnatyuk, Corpus research of semantic transformations of deterministic lexicon using the example of the lexeme virus. *CEUR Workshop Proceedings*, 2024, 3722, pp. 425–456.