

An Electronic Ukrainian Dictionary as a Derussification Tool

Vasyl Starko¹, Andriy Rysin²

¹ Ukrainian Catholic University, 2a Kozelnytska Str., 79026 Lviv, Ukraine

² Independent researcher, Cary, NC, USA

E-mail: v.starko@ucu.edu.ua, arysin@gmail.com

Abstract

Due to the policy of Russification in the 20th century, the Ukrainian language underwent an influx of Russianisms, among other forms of interference with its structure. Today, many Ukrainians require guidance regarding non-Russified usage, and a *Large Electronic Dictionary of Ukrainian* (VESUM, vesum.nlp.net.ua) is designed to meet this need. With a register of over 430,000 lemmas, it is the most comprehensive morphological dictionary of Ukrainian. VESUM contains over 9,300 Russianisms, listed alongside their non-Russified equivalents. The decisions on what counts as a Russified item in need of replacement are based on multiple reputable sources, including dictionaries on the r2u.org.ua dictionary portal.

VESUM is the centerpiece of Pravopysnyk, the Ukrainian module of the LanguageTool text checker (check.nlp.net.ua, languagetool.org/uk). The role of VESUM is threefold. First, it supplies single-word Russified items and their replacements. Second, as a machine-readable dictionary, it serves as the source of data for lemmatization and morphological tagging, which are necessary for advanced text checking. Finally, VESUM can also be consulted as a stand-alone online dictionary via a web interface with flexible search options. As part of the Pravopysnyk tool, this electronic dictionary provides users with guidance on derussification when and where such advice is needed.

Keywords: electronic dictionary; Ukrainian; derussification; Russianism; error correction

1. Introduction

George Shevelov (1966) incisively noted that “the main problem faced by Modern Ukrainian is the problem of Russian words and words that come to Ukrainian through Russian.” This dictum is as true today as it was more than half a century ago. During the Soviet period in the 20th century, the Ukrainian language was subjected to intensive Russification, which took various forms, including, notably, interference with its very structure (Shevelov, 1989) and a ban on Ukrainian dictionaries published during the brief period of Ukrainianization in the 1920s. Among other effects were an influx of Russianisms, the impoverishment of vocabulary due to the marginalization or wholesale elimination of lexical items dissimilar to Russian, as evidenced by official Soviet-published dictionaries of Ukrainian, and changes in collocations, phraseology, and syntactic structures (Masenko et al., 2005; Vakulenko, 2018). To a significant extent, these effects persisted after 1991, when Ukraine regained its independence, due to such

factors as linguistic inertia, a lack of awareness, and a shortage of various types of dictionaries. Many Ukrainians, including Russian-speaking citizens, have embraced Ukrainian and strive to improve their proficiency, especially during Russia's ongoing war of aggression against Ukraine, but they need guidance regarding non-Russified usage.

Ukrainian resources that can help Ukrainian speakers achieve this goal include authoritative usage guides, handbooks, and dictionaries in paper format (Horodens'ka 2019; Karavans'kyj, 2001; Karavans'kyj, 2009; Ponomariv, 2011; Ponomariv, 2017; Serbens'ka, 2022; RUS-2011), usage boxes in such linguistic journals as *Ukrajinska Mova*, online sources (KM; JMH), as well as language forums (R2UF) and posts by literary editors on social media. Most of them discuss particular issues one by one, so a user needs to first become aware of a potential problem, formulate their question, and search for answers within the content. However, a much more efficient approach is to alert the user and provide guidance when and where it is needed, i.e., during text production. Moreover, it is crucial to provide a balanced coverage of the spectrum of linguistic opinions, helping users make informed decisions. In what follows, we describe an online tool for Ukrainian based on a large dictionary that operates precisely in this fashion, helping Ukrainian speakers keep their texts free of unnecessary Russianisms.¹

2. VESUM Dictionary: Formats and Features

A *Large Electronic Dictionary of Ukrainian* (VESUM), which is the focus of this paper, was initially launched as a mid-sized spelling dictionary for Linux and has since evolved into the most comprehensive morphological dictionary of Ukrainian. Its approach to Ukrainian morphology rests on reputable academic sources (Kryts'ka et al., 2011; Vyxovanec' & Horodens'ka, 2004), while its content has benefited from dictionary parsing and large-scale processing of corpus data. VESUM is a non-commercial project: its data are available under the CC BY-NC-SA 4.0 license at a GitHub repository (VESUM-GIT).

VESUM is the centerpiece of Pravopysnyk LanguageTool (Pravopysnyk), an advanced spelling, grammar, and style checker for Ukrainian that can be in, among other options, as a browser add-on to provide real-time error identification.

As a machine-readable morphological dictionary, VESUM is utilized in combination with the TagText tagger for Ukrainian from the NLP-UK toolkit (NLP-UK), achieving over 99% accuracy on texts written in Modern Ukrainian. Since 2017, it has served as the basis for a revamped search engine for the Ukrainian-language Wikipedia. It has also been successfully applied in multiple other projects, such as the lemmatization and

¹ In what follows, *Russianism* and *Russified* refer to linguistic units that are predominantly regarded by Ukrainian linguists, lexicographers, and editors as such that are best replaced with existing Ukrainian equivalents. This does not include numerous words borrowed from or through Russian that have become established in Ukrainian and are unobjectionable.

morphological tagging of several Ukrainian corpora (BRUK, UFLTC, GRAC) (Starko & Rysin, 2023), the construction of word vectors in the lang-uk project (LANG-UK), and the creation of the Ukrainian Vocabulary Profile (PULS). The General Regionally Annotated Corpus of Ukrainian (Shvedova et al., 2017–2025), with its detailed regional annotation and high variety of textual sources (Shvedova, 2020), has been especially valuable in the development of VESUM over the years as a source of a wide range of vocabulary, from dialectal and archaic to modern colloquialisms and slang.

In its current version, VESUM contains over 430,000 lemmas, from which more than 6.5 million word forms are generated. Its complete tagset is available online (VESUM-GIT), while the details on the inflection classes and the generation of word forms from lemmas are described separately (Starko & Rysin, 2020; Starko & Rysin, 2022).

In addition to strictly morphological tags, VESUM supplies stylistic tags, some of which are utilized during text analysis and error correction. These include, among others, :subst (substandard), :arch (archaic or, in some cases, dialectal), :slang, :vulg (vulgar), :alt (alternative spelling), :var (variant form), and :bad (erroneous or objectionable lemma). The :bad tag is of interest for us here, as it is used to mark, *inter alia*, single-word Russianisms in texts.

VESUM can also be consulted as a stand-alone online dictionary via a web interface with flexible search options. Fig. 1 below shows search lemma-only results (without full paradigms) for the query beginning with *slovn*yk ‘dictionary.’ The retrieved words include such derivatives as *slovn*yk^{kar} and *slovn*yk^{ka} ‘dictionary maker’ (masculine and feminine, respectively), *slovn*yk^{karstvo} ‘lexicography,’ and others.

☐ Серед усіх словоформ
 ☐ Вивід таблицею

☒ Показувати лише лему

Знайдено 9 статей
 [Шукати «словник*» на інших ресурсах:](#)


словник noun:inanim:m:v_naz
словникар noun:anim:m:v_naz
словникарка noun:anim:f:v_naz
словникарство noun:inanim:n:v_naz
словникарський adj:m:v_naz
словниковий adj:m:v_naz
словниково-граматичний adj:m:v_naz
словниково-довідковий adj:m:v_naz
словниково-технічний adj:m:v_naz

Figure 1: Lemma-only search results in the web version of VESUM.

Alternatively, users can search for a particular inflected form and/or see complete paradigms in table format, which is especially useful for educational purposes. Fig. 2 shows the search query for *dijem*, a substandard first-person plural form of the verb

dijaty ‘to act,’ which is retrieved in table format, where it appears highlighted next to the correct form *dijemo* and is supplied with a note indicating its substandard status.

дієм ☒ Серед усіх словоформ
☒ Вивід таблицею ☐ Показувати лише лему

Знайдено 1 статтю [Шукати «дієм» на інших ресурсах:](#)   

діяти - дієслово *недок.*
 Узгодження відмінків: о.в.

Інф.		
	діяти	діять
Однина	Множина	
Наказ.		
1-ша		діймо
2-га	дій	дійте
Тепер.		
1-ша	дію	діємо дієм (просторічна форма)
2-га	дієш	дієте
3-тя	діє	діють

Figure 2: Partial results for the query *dijem* in table format in the web version of VESUM.

If a user looks up a Russified word, its lemma and all inflected forms in the paradigm will be marked :bad in the list format, as shown in Fig. 3 for the query *існуjučyі* ‘existing,’ while in the table format, only the lemma will have the descriptive marker *neprav.* ‘incorrect.’

існуючий ☐ Серед усіх словоформ
☐ Вивід таблицею ☐ Показувати лише лему

Знайдено 1 статтю [Шукати «існуючий» на інших ресурсах:](#)   

існуючий adj:m:v_naz:adjp:actv:imperf:bad
 існуючого adj:m:v_rod:adjp:actv:imperf:bad
 існуючому adj:m:v_dav:adjp:actv:imperf:bad
 існуючого adj:m:v_zna:ranim:adjp:actv:imperf:bad
 існуючий adj:m:v_zna:rinanim:adjp:actv:imperf:bad
 існуючим adj:m:v_oru:adjp:actv:imperf:bad
 існуючим adj:m:v_mis:adjp:actv:imperf:bad
 існуючому adj:m:v_mis:adjp:actv:imperf:bad
 існуючий adj:m:v_kly:adjp:actv:imperf:bad
 існуюча adj:f:v_naz:adjp:actv:imperf:bad
 існуючої adj:f:v_rod:adjp:actv:imperf:bad

Figure 3: Partial results for the query *існуjučyі* in list format in the web version of VESUM.

Thus, VESUM functions in machine-readable format for NLP purposes, in web format for human lookup, and as a key component of an advanced spellchecker.

3. Types of Russianisms in the VESUM Dictionary

VESUM contains an extensive inventory (over 9,300 items) of Russianisms listed together with their replacements. All of these Russianisms are single-word items; each lemma and the corresponding inflected forms carry the :bad tag (as illustrated above), allowing for their automatic identification and marking in texts. Thus, the dictionary serves as a source of suggestions for single-word items. Any Russified multiword expressions or constructions are handled separately by Pravopysnyk LanguageTool (see section 4 below), while VESUM provides lemmas, part-of-speech information, and morphological features for text processing.

The largest group of Russianisms (nearly 3,500 items) consists of imperfective active participles, e.g., *blokujučyj* ‘blocking,’ *akumuljujučyj* ‘accumulating,’ *vertajučyj* ‘returning,’ *bahatoobicjajučyj* ‘promising, hopeful,’ *bidnijučyj* ‘becoming poor or poorer,’ *kryjučyj* ‘covering,’ etc. These forms occur more than 730,000 times (410 pm) in the GRAC corpus (version 17a). To convey the same meaning, Ukrainian employs different linguistic devices, depending on the word in question, its meaning in context, syntactic structure, and the overall style of the text. The first replacement option provided in VESUM is most often the subordinate construction consisting of the conjunction *ščo* and the corresponding verb in the third-person singular, e.g., *ščo blokuje* ‘that blocks.’ Another frequently used option offered to the user includes a deverbal adjective with the suffix *-l'n-*, e.g., *blokuval'nyj* ‘blocking,’ which can denote both the purpose (one that is designed/intended to block) and the actual use (one that blocks). In cases of substantivization, nominal counterparts are also suggested. For example, the following words can replace *atakujučyj* ‘attacking’ in different contexts: the adjectives *atakuval'nyi* and *napadnyj* ‘attacking’ or the noun *napadnyk* ‘attacker, assailant; forward (in sports).’

The suggested Ukrainian equivalents are taken from dictionaries and corpora. In some cases, entire synonym sets are provided. For example, *vbyvajučyj* ‘killing’ can be replaced with *ubyvčyj*, *dušohubnyj*, *smertonosnyj*, *smertel'nyj*, or *smertovbyvčyj*, all referring to killing, murder, and causing death. Paraphrases are also suggested, e.g., *oxoplenyj ahoniyeju* ‘grasped by death’s agony,’ *v ahoniji* ‘in death’s agony,’ and *pry smerti* ‘in one’s death throes’ are among the replacements of *ahonizujučyj* ‘in death’s agony.’

In a number of cases, a general recommendation and explanation, rather than specific alternatives, is provided to the user via a pop-up message in Pravopysnyk LanguageTool along the following lines (English equivalents are added here for clarity):

“Active participles are not characteristic of Ukrainian. They can be replaced with proper Ukrainian words in different ways: *ščo* + verb (*robljačyj* ‘working’ — *ščo robyt* ‘that works’), a verbal root + suffixes *-l'n-*, *-lyv-*, etc. (*zbyrajučyj* ‘gathering’ — *zbyral'nyj*, *obtjažujučyj* ‘burdening’ — *obtjažlyvyj*),

a noun (*zavidujučyj* ‘chief, head’ — *zaviduvač*), an adjective with a fitting meaning (*dijuča model* ‘working model’ — *roboča model*), by changing the construction (*z nastupajučym Novym rokom* — *z nastannjam Novoho roku*), etc.”

Other significant groups of Russianisms include deverbal nouns ending in *-ka* that designate processes (e.g., *aranžyrovka* ‘arrangement,’ *verbovka* ‘recruitment,’ *vidpravka* ‘sending (off), dispatching; shipping,’ *holodovka* ‘hunger strike,’ *dyktovka* ‘dictation,’ and *dovodka* ‘finishing (work)’). These are best replaced with the corresponding nouns ending in *-nnja*, where the suffix reflects the procedural component of meaning, which is important in Ukrainian: *aranžuvannja*, *verbuuvannja*, *vidrjadžannja* (also *posylannja*, *vyrušennja*, etc.), *holoduvannja*, *dyktuvannja*, and *dovedennja* (*vykinčennja*). Occasionally, nouns with a different morphological structure are also matched, e.g., *dyktant* ‘dictation’ and *dovedenist* (*vykinčenist*) ‘finished state.’ This group accounts for over 239,000 occurrences (165 pm) in GRAC.

The initial component *dvox-* ‘two-’ and *trjox-* ‘two’ in adjectives and nouns, e.g., *dvoxdennyi* ‘two-day’ and *trjoxrička* ‘three-year period,’ is best replaced with *dvo-* and *try-* (*dvodennyj*, *tryrička*). These forms occur over 15,000 times (8.6 pm) in the GRAC corpus. While this replacement pattern has been part of VESUM for over eight years, it was codified in the new official Ukrainian orthography in 2019 (UP).

Nomina actionis with the suffixes *-ščyk-* and *-ščyc-* (e.g., *velohonščyk* ‘male bicycle racer’ and *muzejščycja* ‘female museum employee’) structurally copy the Russian counterparts. Instead, the corresponding Ukrainian nouns with the suffix *-n-* (*veloperehonnyk*, *muzejnycja*) are given preference. This group of Russianisms is attested in GRAC with over 13,000 occurrences (7.6 pm).

Adjectives ending in *-vydnyj* (e.g., *hrybovydnyj* ‘mushroom-like’ and *derevovydnyj* ‘dendriform’), which occur over 4,000 times (2.33 pm) in GRAC, are structural calques, copying the Russian formant *-vidnyj* (*gribovidnyj* and *drevovidnyj*). Their standard Ukrainian equivalents are words ending in *-podibnyj* ‘-like,’ while other suffixes to express similarity are also sometimes used, e.g., *derevuvatyj*, *derevystyj* ‘dendriform, tree-like.’

In addition to the groups described above, the top 10,000 words by frequency in Modern Ukrainian (based on the 2001-2023 subcorpus of GRAC, version 17a) include a handful of Russianisms, which are presented in Table 1 below along with their Russian counterparts (illustrating the similarity between the two), as well as one of the possible English equivalents and one of the Ukrainian alternatives supplied by VESUM.

Russianism	Russian counterpart	English equivalent	Replacement
policejskyj	policejskij	policeperson	policijant
postavka	postavka	delivery	postačannja
pryjom	prijom	reception	pryjnattja
sxože	sxože	likely	vydno
da	da	yes	tak
zarubižnyj	zarubežnyj	foreign	zakordonnyj
budučy	buduči	(while) being	buvšy
zadijaty	zadejat'	involve	zalučyty
vyručka	vyručka	proceeds	vytorh
peredvybornyj	predvybornyj	pre-election	peredvyborčyj
nevidjemnyj	neotjemlemyj	inseparable	neviddilnyj
zaključnyj	zaključitel'nyj	final	zaveršal'nyj
vuz (abbr.)	vuz (abbr.)	higher-education institution	vyš
dostovirnyj	dostovernij	credible	virohidnyj
optovyj	optovyj	wholesale	hurtovyj
mynuloričnyj	prošlogodnij	last-year	torišnij
misceznaxodžennia	mestonaxoždenije	location	misceperebuvannia
predstojatel'	predstojatel'	primate	nastojatel'
protyriččja	protivorečije	contradiction	superečnist'
blahopoluččja	blagopolučije	well-being	dobrobut
jemnist'	ěmkost'	container	mistkist'
vytoky	istoki	sources	džerela
oxarakteryzuvaty	oxarakterizirovat'	characterize	sxarakteryzuvaty
vlast'	vlast'	power	vlada
postavljaty	postavljat'	to supply	postačaty
specnaz	specnaz	riot police	speczahony
zavidomo	zavedomo	knowingly	svidomo

Table 1: Most frequent Russianisms in GRAC with replacements from VESUM.

The same frequency list extracted from GRAC also contains lemmas for which VESUM recommends using a more suitable Ukrainian alternative. While these lexemes also arose under the influence of Russian, they are more established and/or generally less objectionable in Ukrainian, as we have established through analysis of literature and corpus data. Thus, they are marked blue (as stylistic issues) in texts by Pravopysnyk LanguageTool. Here are these lemmas with replacements: *skladova* ‘component’ – *skladnyk*, *pidnimaty* ‘to lift’ – *pidijmaty*, *obstanovka* ‘furniture; situation’ – *obstava*, *obstavyyny*, *spivvitčyznyk* ‘compatriot’ – *zemljak*, *krymčanyn* ‘Crimea resident’ – *krymec*, *vybačytysja* ‘to apologize’ – *pereprošuvaty*, *ruslo* ‘riverbed’ – *ričyšče*, *podali* ‘as far as possible’ – *jakomoha dali*, and *kaznačejstvo* ‘treasury’ – *skarbnycja*.

VESUM provides replacement alternatives for these and many more Russianisms identified in Ukrainian texts during the processing of GRAC texts and other collections. Moreover, it labels as substandard (:subst) those morphological forms that arose under the influence of Russian and fall outside Standard Ukrainian. The most frequent of these is the first-person plural *-m* ending (instead of the standard *-mo*) in the future tense for perfective verbs (e.g., *zanesem* ‘will take’) and in the present tense for imperfective verbs (e.g., *robym* ‘are doing’). Such substandard forms occur more than 29,000 times (16.58 pm) in the GRAC corpus.

4. Pravopysnyk LanguageTool

The overall approach to suggesting replacements for Russianisms in both VESUM and Pravopysnyk LanguageTool can be described as moderate purism along the lines formulated by Shevelov (1966). He called for a kind of filter to be implemented to weed out only those Russianisms that are unnecessary. In our case, this filter is an electronic tool (an advanced spellchecker), which draws on three types of sources: dictionaries, corpus data, and linguistic advice on usage. Let us consider how each type contributes to the decision making regarding Russianisms and how the results are implemented in practice.

Pronouncements on lexical usage are abundant in Ukrainian linguistics. From the variety of sources available, we rely on the most reputable publications (see the list of handbooks above) and note cases of agreement. Regarding dictionaries, we have developed and maintain the Russian-Ukrainian dictionary portal (R2U), which features a collection of hand-picked lexicographic works, including those that were banned during Soviet times. Even though they reflect the state of Ukrainian more than a century ago and need to be viewed critically, these dictionaries serve as important guides to us. The gem of the r2u collection is the academic Russian-Ukrainian Dictionary edited by Ahatanhel Kryms'kyj and Serhij Jefremov (RUS-1924). It was banned during Soviet times, and its fourth and last volume was destroyed. It was only in the 2000s that the dictionary was brought back into circulation (in electronic format), thanks to the efforts of the r2u team, which used the scan produced by Valentyn Kul'kov and distributed by Viktor Kubaičuk and Olga Kočerga, theoretical physicists from Kyiv (Starko 2017). The dictionary provides the last thorough, academic snapshot of Ukrainian before the Russification campaign was launched in the 1930s and is extremely valuable for the development of Ukrainian today. Among contemporary dictionaries, the four-volume academic Russian-Ukrainian dictionary (RUS-2011) has proven to be most useful in our experience. Other dictionaries, primarily monolingual, have also been consulted. The third key source of information about word usage is provided by the GRAC corpus, which encompasses more than 150,000 texts written by approximately 35,000 authors over the period from 1816 to the present. Crucially, it is the only Ukrainian corpus that contains a diverse range of older texts, allowing for the tracing of lexical usage dynamics.

When these sources are in agreement, Russianisms are assigned the :bad tag in VESUM, highlighted in yellow in the text, and supplied with a message offering possible replacements (see Fig. 4).

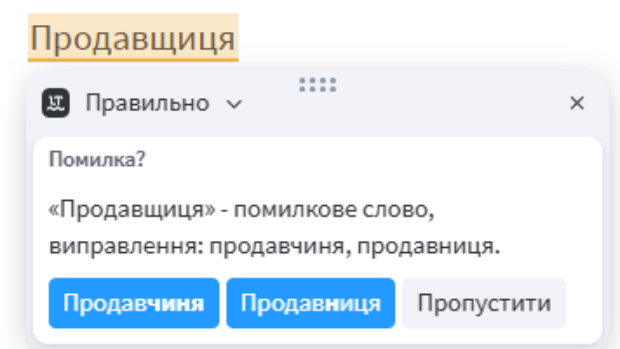


Figure 4: Pop-up message for *prodavščyčja* in Pravopysnyk LanguageTool.

In this case, the Russianism *prodavščyčja* ‘saleswoman’ (cf., the Russian *prodavščica*) is marked yellow, and replacements are suggested in the pop-up message. The user is given the choice of two Ukrainian alternatives, *prodavčynja* and *prodavnycja*. The first option suggested by Pravopysnyk LanguageTool is usually the best fit in terms of meaning and usage.

In cases of disagreement between our sources or when there are reasonable doubts that the use of a word or phrase that looks like a borrowing from Russian is indeed unwarranted, we prefer to err on the side of caution: a different color code (blue) is used to highlight the item in text, signaling that it is a stylistic issue. The corresponding pop-up message contains explanatory information, sometimes accompanied by a link to a detailed discussion, helping users make an informed decision. A good case in point is the expression *na protjazi* ‘in the course of’ which has been generally stigmatized in Ukrainian. When found in a text, the Pravopysnyk tool will highlight it blue and provide the following message (Fig. 5): “Some linguists advise against using *na protjazi* to refer to a time period.” The user is offered two alternatives, *uprodovž* and *protjahom*, both meaning ‘during,’ or they can choose to deactivate this kind of correction in the entire text.

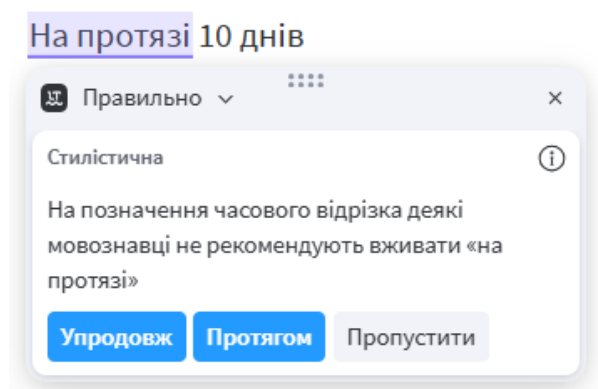


Figure 5: Pop-up message for the phrase *na protjazi* in Pravopysnyk LanguageTool.

The information button in the top right corner takes the user to a separate page with a detailed explanation. A summary of our findings from three types of sources (usage guides, dictionaries, and corpus data) is provided there, along with references and links. In this case, the user is informed that “data from the dictionaries published in the 1920s suggest that their compilers did not consider the phrase *na protjazi* a Russianism. However, the word *protjahom* has since then become established in usage as the dominant designation of a time period. The words *vprodovž/uprodovž* are also used in this case, albeit not as often, while *na protjazi* occurs only rarely. A number of specialists and many speakers now believe *na protjazi* to be incorrect.” Having received ample information about this particular expression and its alternatives, users can now decide for themselves which option best suits them. This way, our tools also perform an educational function.

Amid the overall aspiration to free Ukrainian from Russianisms, there is a trend among some Ukrainian speakers to reject all and any lexical items that mirror their Russian counterparts. While this distancing is understandable, there are a number of Ukrainian words that resemble Russian ones but still have their place in Ukrainian, even when they are not part of the standard repertoire. Taken together, they may be termed “false Russianisms” or “pseudo-Russianisms,” and it is crucial to distinguish two subgroups among them. One includes standard Ukrainian words, such as *rysuvaty* ‘to draw,’ *temnota* ‘darkness,’ and *lyšnij* ‘extra,’ which are left unmarked by Pravopysnyk LanguageTool. The other group comprises words that are either dialectal or archaic/outdated from the standpoint of Modern Ukrainian. In trying to assess the status of such items, we have found stylistic labels in existing Ukrainian dictionaries to be inconsistent and insufficiently reliable. In contrast, data from the GRAC corpus has proven to be invaluable for this task, particularly thanks to its coverage of older and regional texts, enabling us to provide corpus-based labeling of vocabulary items. For example, the following “pseudo-Russianisms” are labeled :arch in VESUM and marked blue in texts: *dovh* ‘debt,’ *vydity* ‘to see,’ *vremja* ‘time,’ *nihde* ‘nowhere,’ *zavyst* ‘envy,’ *ložyty* ‘to lay down,’ *nahyj* ‘naked,’ *ohon* ‘fire,’ *pervyj* ‘first,’ *poslidnij* ‘last,’ and many others.

VESUM is a continually expanded and improved dictionary. Over time, we have made numerous updates in the dictionary files, reflecting a deeper and more nuanced understanding of the language dynamics. As larger amounts of textual data become available for analysis, primarily in the GRAC corpus, we anticipate that VESUM will undergo further modifications. This kind of continual, cyclical lexicographic process distinguishes VESUM from other Ukrainian dictionaries, which are largely static by nature.

As more Russian speakers produce texts in Ukrainian, the number of various Russified forms occurring in texts may increase at a high rate. The VESUM dictionary and Pravopysnyk LanguageTool explicitly identify the most frequent words and forms, while the rest are handled automatically: such items are marked pink as unrecognized words

in a text; a pop-up message notifies the user that “a potential orthographic error has been found,” and the tool suggests corrections based on the Levenshtein distance and a frequency list of some 150,000 lemmas. While this approach is fairly standard in spellchecking, its utility directly depends on the vocabulary size, as a small vocabulary leads to many valid words not being recognized in text. The large and constantly growing size of VESUM’s register ensures that such cases are reduced to a minimum.

Pravopysnyk LanguageTool follows the principle “suggest corrections only if certain,” which is applied to all kinds of issues, including Russianisms. The benefit lies in avoiding many false positives, which can overburden the user, while the downside is that some cases that require correction may be overlooked.

The vast majority of items listed as replacements for Russianisms in VESUM are part of the established lexical stock of the Ukrainian language, well-attested in corpus data and dictionaries.

Some words borrowed from Russian were adopted in Ukrainian and have become standard designations for the respective concepts, even though they are not attested in corpus data before a certain point, such as the late 19th or early 20th century, and are not found in dictionaries published before the 1930s. In cases like this, if there are no better alternatives, we treat these items as standard Ukrainian words. For example, some extreme purists view the word *urok* ‘lesson’ as an undesirable Russianism primarily based on its near absence in the dictionaries before the 1930s and argue for other Ukrainian words instead. However, *urok* has been the standard term for a lesson in the Ukrainian school system for a long time, while its suggested alternatives have fallen out of use in this sense. Therefore, there are no reasonable grounds for suggesting any corrections for this lexical item.

Pravopysnyk LanguageTool can be accessed online at two primary web addresses (Pravopysnyk) and on other websites that utilize its engine for spellchecking. At the same time, LanguageTool can be installed as a web add-on, office plugin, or desktop application. No data is stored and no statistics are collected on LT’s servers, so it is impossible to gauge which items, including Russianisms, appear most frequently in user texts. Instead, we utilize frequencies and other data from GRAC and our internal news corpus in the development of VESUM and Pravopysnyk LanguageTool.

While leveraging VESUM for lemmas and morphological data, which are critical for text processing and error detection, Pravopysnyk LanguageTool handles multiword expressions and constructions on its own, including those that have arisen under the influence of Russian. For example, it employs fairly complex logic to verify agreement in texts, which detects, among other things, common Russified constructions, such as *po+NOUN.DAT.PL* instead of *po+NOUN.LOC.PL* (*po mistam* instead of *po mistax*). This confusion occurs in texts because the plural locative ending in Russian nouns coincides with the plural dative ending in Ukrainian nouns. Another frequent construction involves a numeral such as *dva* ‘two,’ *try* ‘three,’ or *čotyry* ‘four’ and a

noun: it must have the plural nominative ending in Ukrainian (e.g., *dva dyvany* ‘two sofas’), whereas Russian requires the singular genitive case (*dva divana*). Pravopysnyk LanguageTool is also helpful in detecting cases when the superlative degree of adjectives is formed incorrectly, following the Russian pattern of using the word *samyj* and an adjective in the positive degree, e.g., *samyj cikavyj* instead of the correct *najcikavišyj*.

Pravopysnyk LanguageTool identifies grammatical constructions, including case government, that mirror those in Russian but for which established Ukrainian constructions exist. For example, the phrase *ity za hrybamy* ‘to go pick up mushrooms’ refers to the goal of the action, in which case the preposition *po* is required in Ukrainian (*ity po hryby*). On the other hand, Ukrainian speakers may incorrectly use the preposition *po*, mirroring Russian, where a different structure is required in Ukrainian, e.g., *po subotax* ‘every Saturday’ instead of *u subotu* or *ščosuboty*; *po svjatax* ‘on holidays’ instead of *na svjata*; *po cini* ‘at the price (of)’ instead of *za cinoju*; *po imeni* ‘by one’s name’ instead of *na imja*, etc. The verb *navčyty* ‘to teach’ is illustrative of an issue with case government: its Russian counterpart *naučit’* requires a noun in the dative case (*naučit’ remeslu* ‘to teach a craft’), whereas the genitive case needs to be used in Ukrainian (*navčyty remesla*). A similar issue of case misalignment is presented by the verb *probačyty*, which takes the indirect object in the dative case, and its Russian counterpart *prostit’*, which requires the genitive case. Many speakers, especially those whose native language is Russian or who grew up in a Russian-language environment, find it challenging to keep track of these fine distinctions and require the kind of on-the-spot guidance provided by Pravopysnyk.

Moreover, this tool successfully handles a number of multiword expressions. Table 2 lists Russified phrases along with their Russian points of origin, English equivalents, and Ukrainian replacements (one or two of the options provided to the user).

Russianism	Russian expression	English equivalent	Replacement
za raxunok	za sčet	at the expense of	koštom
v peršyj raz	v pervyj raz	for the first time	vperše
robyty vyhliad	delat’ vid	to pretend	udavaty
bil’š za vse	bolee vsego	more than anything	ponad use
odyn na odyn	odin na odin	tête-à-tête	sam na sam
hirše nikudy	xuže nekuda	could not be worse	dali nema kudy
ne po sobi	ne po sebe	ill at ease	nijakovo
ne v sylax	ne v silax	unable to	nesyla, nespromoha
jak by tam ne bulo	kak by tam ni bylo	in any case	xaj tam jak
vse rivno	vsë ravno	all the same	vse odno
lamaty holovu	lomat’ golovu	to rack one’s brains	sušyty holovu

Table 2: Russified multiword expressions processed by Pravopysnyk LanguageTool.

The text checker is sufficiently sophisticated to identify some “creeping Russianisms” (Shevelov 1966), which have sneaked into Ukrainian unnoticed, subtly affecting usage. One example is the verb *prysvojuvaty* (Rus. *prisvaivat'*), which means ‘to appropriate’ in both languages and ‘to lend, confer’ only in Russian. This latter use has crept into Ukrainian, leading to such expressions as *prysvojity komus' zvannja* ‘to confer a title on smb.,’ which Pravopysnyk LanguageTool identifies in various constructions with intervening words and suggests replacing with *nadavaty komus' zvannja*.

5. Conclusions

Resulting from decades of Russification, Russianisms present a key challenge in Ukrainian. Speakers with varying degrees of Ukrainian proficiency require careful, tailored guidance regarding different types of Russianisms.

VESUM is the largest morphological dictionary of Ukrainian, designed to perform a variety of practical tasks. As a machine-readable dictionary, it serves as the source of data for lemmatization and morphological tagging, which are necessary for advanced text checking. In this role, it has been successfully used in a number of Ukrainian NLP projects. Notably, the dictionary has been developed in close synergy with GRAC, a large reference corpus of Ukrainian. VESUM can also be consulted as a stand-alone online dictionary via a web interface with flexible search options. VESUM’s register includes more than 9,300 single-word Russianisms with suggested Ukrainian replacements. This information, along with rich morphological data, is fed into Pravopysnyk LanguageTool, the Ukrainian-language module of the widely used text checker, where it is utilized for text processing, error identification, and correction.

Pravopysnyk LanguageTool leverages VESUM data for dealing with single-word Russianisms. Furthermore, it handles Russified multiword expressions and constructions in a flexible way. Together, VESUM and Pravopysnyk LanguageTool serve to provide Ukrainian users with guidance on derussification when and where such advice is needed.

6. References

- BRUK: *Ukrainian Brown Corpus*. Accessed at: https://github.com/brown-uk/dict_uk. (7 July 2025)
- Horodens'ka, K. (2019). *Ukrajins'ke slovo u vymirax s'ohodennja*. 2nd ed. Kyiv: KMM.
- JMH: Antonenko-Davydovyč, B. *Jak my hovorymo*. Accessed at: <http://yak-my-hovorymo.wikidot.com>. (7 July 2025)
- Karavans'kyj, S. (2001). *Pošuk ukrajins'koho slova, abo borot'ba za nacionalne “Ja.”* Kyiv: Akademia.
- Karavans'kyj, S. (2009). *Sekrety ukrajins'koji movy*. 2nd ed. L'viv: BaK.
- KM: *Kultura movy na ščoden'*. Accessed at: <http://kulturamovy.univ.kiev.ua>. (7 July

- 2025)
- Kryts'ka, V., Nedozym, T., Orlova, L., Puzdyrjeva, T., Romaniuk, Ju. (2011). *Hramatyčnyi slovnyk ukrajinskoji literaturnoji movy. Slovozmına*. Kyiv: Dmytro Burago Publishing House.
- LANG-UK: Lang-uk. Accessed at: <https://lang.org.ua/en/>. (7 July 2025)
- Masenko, L., Kubajčuk, V. & Dems'ka-Kul'čycka, O. (eds.) (2005). *Ukrajins'ka mova u XX storičči: istorija linhvocydu: Dokumenty i materialy*. Kyiv: Kyjevo-Mohyljans'ka akademija.
- NLP-UK: NLP-UK toolkit for Ukrainian. Accessed at: https://github.com/brown-uk/nlp_uk. (7 July 2025)
- Ponomariv, O. (2011). *Kultura slova: Movnostylistyčni porady*. 4th ed. Kyiv: Lybid'.
- Ponomariv, O. (2017). *Ukrajins'ke slovo dlja vsix i dlja kožnoho*. 2nd ed. Kyiv: Lybid'.
- Pravopysnyk: Pravopysnyk LanguageTool. Available at: languagetool.org/uk,check.nlp.net.ua. (7 July 2025)
- PULS: *Ukrainian Vocabulary Profile*. (2025). Lviv. Accessed at: <https://puls.peremova.org>. (7 July 2025)
- R2U: Russian-Ukrainian dictionary portal. Accessed at: <https://r2u.org.ua>. (7 July 2025)
- R2UF: R2U Forums. Accessed at: <https://r2u.org.ua/forum>. (7 July 2025)
- RUS-1924: *Rosijs'ko-ukrajins'kyj slovnyk*. (1924-1933). Vols. 1–3, Kyiv, Xarkiv: DVU, URE.
- RUS-2011: *Rosijs'ko-ukrajins'kyj slovnyk*. (2011-2014). Vols. 1–4, Kyiv: Znannja.
- Serbens'ka, O. (2022). *Antysuržyk*. Lviv: Apriori.
- Shevelov, G. (1966). Puryzm v ukrajinskij movi. *Ukrajinski visti*, 23, pp. 2–3. Available at: <https://zbruc.eu/node/104522>.
- Shevelov, G. Y. (1989). *The Ukrainian Language in the First Half of the Twentieth Century (1900-1941): Its State and Status*. Cambridge, Mass.: Harvard University Press.
- Shvedova, M. (2020). The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorporus.org): Architecture and Functionality. In *Computational Linguistics and Intelligent Systems. Proc. 4th Intl. Conf. COLINS 2020*, pp. 489–506.
- Shvedova, M., Waldenfels, R. von, Yarygin, S., Rysin, A., Starko, V., Nikolajenko, T. et al. (2017–2025). *GRAC: General Regionally Annotated Corpus of Ukrainian*. Kyiv, Lviv, Jena. Accessed at: <http://uacorporus.org>. (7 July 2025)
- Starko, V. (2017). Kompjuterni linhvistyčni proekty hurtu r2u: stan ta zastosuvannja. *Ukrajins'ka mova*, 3, pp. 86-100.
- Starko, V. & Rysin, A. (2020). Velykyj elektronnyj slovnyk ukrajinskoji movy (VESUM) jak zasib NLP dlja ukrajinskoji movy. In *Halaktyka Slova. Halyni Makarivni Hnatiuk*. Kyiv: Dmytro Burago Publishing House, Kyiv, pp. 135–141.
- Starko, V. & Rysin, A. (2022). VESUM: A Large Morphological Dictionary of Ukrainian As a Dynamic Tool. *Computational Linguistics and Intelligent Systems. Proc. 6th Int. Conf. COLINS 2022*, Gliwice, Poland, pp. 71–80.

- Starko, V. & Rysin, A. (2023). Creating a POS Gold Standard Corpus of Modern Ukrainian. *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*. ACL, pp. 91–95.
- UFLTC: *The UFL Textbook Corpus*. Accessed at: <http://corpus-puls.peremova.org>. (7 July 2025)
- UP: *Ukrajins'kyj pravopys*. Accessed at: <https://www.inmo.org.ua/pravopys-2019.html> (17 July 2025)
- Vakulenko, S. (ed.) (2018). *Ukrajins'ka mova: unormuvannja, rozunormuvannja, perevnormuvannja (1920-2015)*. Xarkiv: Xarkivs'ke istoryko-filolohične tovarystvo.
- VESUM: Rysin, A. & Starko, V. *Large Electronic Dictionary of Ukrainian (VESUM)*. (2005-2025). Web version 6.7.1-SNAPSHOT. Accessed at: <https://vesum.nlp.net.ua>. (7 October 2025)
- VESUM-GIT: Rysin, A. & Starko, V. *Large Electronic Dictionary of Ukrainian (VESUM)*. (2005-2025). Version 6.7.1-SNAPSHOT. Accessed at: https://github.com/brown-uk/dict_uk. (7 October 2025)
- Vyxovanec', I. & Horodens'ka, K. (2004). *Teoretyčna morfolohija ukrajins'koji movy*. Kyiv: Pul'sary.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

