Vilnius University Press

Contents lists available at Vilnius University Press

Slavistica Vilnensis
2026, vol. 71(1), pp. 112–126

ISSN 2351-6895 eISSN 2424-6115
DOI: https://doi.org/10.15388/SlavViln.2026.71(1).9

Onomatopoeia in Contemporary Written Ukrainian: A Corpus-Based Study

Milana Hrachova

Pavol Jozef Šafárik University in Košice, Slovakia
E-mail: milana.hrachova@student.upjs.sk
ORCID iD: https://orcid.org/0009-0008-9249-6694
https://ror.org/039965637

Abstract. Despite growing interest in sound symbolism across languages, onomatopoeia in Standard Ukrainian has not been thoroughly investigated [Hatsenko 2003a, 123]. This paper employs a corpus-based approach to investigate lexical onomatopoeia in contemporary written Ukrainian. Within the Ukrainian linguistic tradition, onomatopoeic words are treated as a subclass of interjections [Hatsenko 2003b]. The analysis is based on the *ukTenTen22* corpus accessed via the *Sketch Engine*. A dataset of 45 lexical onomatopoeic expressions was compiled manually from items tagged as interjections. For each item, ten concordance lines were extracted, resulting in a sample of 450 tokens. The selected items were classified into four semantic categories according to the type of sound imitated: human-related sounds, sounds of inanimate nature and plants, animal and bird sounds, and sounds associated with objects set in motion by humans. The analysis identifies tendencies in frequency, genre distribution, and contextual functions. The findings show that lexical onomatopoeia is concentrated mainly in literary texts, especially fiction and children's literature. However, its presence in blogs, online comments, and news articles indicates that its use extends to a broader range of contemporary written discourse. The paper contributes new empirical evidence to the study of Ukrainian onomatopoeia and sound symbolism in Slavic languages.

Keywords: onomatopoeia, Standard Ukrainian, corpus-based analysis, semantic classification, frequency analysis, text genres.

Ономатопея в сучасній писемній українській мові: корпусне дослідження

Анотація. У статті використовується корпусний підхід для дослідження лексичної ономатопеї в сучасній писемній українській мові. В українській мовознавчій традиції ономатопеїчні слова розглядаються як підклас вигуків [Гаценко 2002, 101]. Дослідження ґрунтується на корпусі ukTenTen22, доступному через платформу Sketch Engine. Матеріал дослідження становлять 45 лексичних ономатопеїчних виразів, відібраних вручну з-поміж слів, позначених у корпусі як вигуки. Для кожної одиниці було проаналізовано десять конкордансних рядків, усього 450 токенів. Відібрані одиниці було класифіковано за чотирма семантичними категоріями. На основі аналізу визначено тенденції щодо частотності, жанрового розподілу та контекстуальних функцій ономатопеї. Результати показують, що лексична ономатопея зосереджена переважно в художніх текстах, особливо в дитячій літературі, однак також представлена в блогах, онлайн-коментарях, новинних текстах та інших жанрах сучасного писемного дискурсу.

Received: 2025-10-08. **Accepted:** 2026-05-12.

Ключові слова: ономатопея, стандартна українська мова, корпусний аналіз, семантична класифікація, частотний аналіз, текстові жанри.

Onomatopėja šiuolaikinėje rašytinėje ukrainiečių kalboje: tekstynu paremtas tyrimas

Santrauka. Straipsnyje nagrinėjama leksinė onomatopėja šiuolaikinėje ukrainiečių kalboje taikant tekstynu paremtą tyrimų metodiką. Tyrimas grindžiamas korpusu ukTenTen22, prieinamu per platformą „Sketch Engine“. Rankiniu būdu iš elementų, pažymėtų kaip ištiktukai, buvo sudarytas 45 leksinių onomatopėjinių (ištiktukinių) posakių duomenų rinkinys. Kiekvienam elementui buvo ištraukta po dešimt konkordanso eilučių, gauta 450 vienetų (tokenų) imtis. Atrinkti vienetai suskirstyti į keturias semantines kategorijas: pagal imituojamo garso tipą: su žmogumi susiję garsai, negyvosios gamtos ir augalų garsai, gyvūnų ir paukščių garsai bei garsai, susiję su žmogaus judinamais objektais. Remiantis analize nustatomos dažnumo, žanrų pasiskirstymo ir kontekstinių funkcijų tendencijos. Rezultatai rodo, kad leksinė onomatopėja daugiausia sutelkta meniniuose tekstuose, ypač vaikų literatūroje, tačiau taip pat pasitaiko tinklaraščiuose, internete (komentaruose), kituose šiuolaikinio rašytinio diskurso žanruose. Straipsnyje pateikiama naujų empirinių įrodymų ukrainiečių onomatopėjos ir garsų simbolizmo slavų kalbose tyrimams.

Reikšminiai žodžiai: onomatopėja, standartinė ukrainiečių kalba, tekstyno analizė, semantinė klasifikacija, dažnis, tekstų žanrai.

---

1. Introduction

Onomatopoeia is commonly defined as the imitation of the sounds of extralinguistic reality [Körtvélyessy & Štekauer 2024, 2] and is considered a non-organic element of a language; therefore, it occupies a peripheral position in the lexicon [de Saussure 1959, 69]. It should be noted that, while the narrow view defines onomatopoeia solely as a direct imitation of sounds (known as primary onomatopoeia), a broader view also includes so-called secondary onomatopoeia, referring to lexical items formed through derivation from primary forms as well as terms that are based on their metaphorical or metonymic shift [Körtvélyessy & Štekauer 2024, 2].

Following Fischer (1999) and Simpson (2004), this study adopts the distinction between lexical and non-lexical onomatopoeia. Lexical onomatopoeia consists of conventionalized sound-imitative words that belong to the established vocabulary of a language. For instance, these are “words like *thud, crack, slurp, and buzz*, whose pronunciation enacts symbolically their referents outside language” [Simpson 2004, 67]. As institutionalized units, lexical onomatopoeia exhibits stable phonological and morphological forms that are listed in dictionaries. In contrast, non-lexical onomatopoeia comprises non-conventionalized clusters of sounds that mimic auditory events without relying on the linguistic structure, for example, *brrrrm brrrrm* [Simpson 2004, 67]. These forms are context-dependent and are not codified in dictionaries.

Given this distinction, the present paper focuses only on lexical onomatopoeia, while non-lexical forms fall outside of its scope.

"Onomatopoeia has not been paid much attention in theoretical linguistics, which may be related to its iconic-symbolic nature" [Körtvélyessy 2020, 1]. However, a growing body of contemporary research on various aspects of this phenomenon demonstrates an increasing interest in this issue. These works concentrate on different aspects such as the exploration of the sound-symbolic nature of onomatopoeia [Akita 2013; Laing 2019; Körtvélyessy 2020], its role in first-language acquisition [Imai & Kita 2014; Laing et al. 2017; Motamedi et al. 2020], its acoustic and perceptual dimensions [Hrytsu 2025], and the status of onomatopoeia in adult speech [see Rydblom 2016 for Japanese and English, and Jahdhami 2020 for Omani Arabic; Gregová 2021 for Slovak]. Despite this progress, there is a notable lack of research exploring the role of onomatopoeia in Ukrainian [Hatsenko 2003, 123].

1.1 Onomatopoeia in Ukrainian

Ukrainian onomatopoeia fits naturally into the broader Slavic linguistic tradition, where onomatopoeia is most commonly treated as a subgroup of interjections [Dvonč et al. 1966; see Oravec, Bajzíková, & Furdík 1984 for Slovak; Rusínová 1996 for Czech; Hatsenko 2003 for Ukrainian; Bańko 2008 for Polish]. Across these languages, onomatopoeia occupies a marginal position yet displays a clear internal structure that aligns with other lexical elements [see Bańko 2008 for Polish; Kleňhová 2010 for Czech; Körtvélyessy 2020 for Slovak], high expressive value [see Zima 1961 for Czech], and considerable derivational potential [Bańko 2008 for Polish; Andrej 2020 for Slovak]. Available evidence suggests that Ukrainian onomatopoeia participates in the same typologically characteristic Slavic tendencies. What distinguishes Ukrainian, however, is the relative lack of detailed phonological description for the standard language [Hrytsu 2025].

"The position of onomatopoeia in languages varies cross-linguistically" [Gregová 2021, 7]. Therefore, it is always important to indicate how such language elements function within each language. As stated above, Ukrainian onomatopoeia is classified as a subclass of interjections [Hatsenko 2002, 1]. Both primary onomatopoeia and secondary onomatopoeia can be found in Ukrainian. For instance, the word *гав (hav*) 'bark' is primary onomatopoeia, while *гавкати (havkaty)* 'to bark' represents secondary onomatopoeia, as it is derived from a primary form *гав (hav)* 'bark'.

Regarding studies on Ukrainian onomatopoeia, it is evident that this topic is rather understudied. In terms of contextual use, onomatopoeia has been described as revealing "the experimental nature of a separate poem, so that it confirms the continuity of language development, particularly in the field

of expressing the nature in sensory images" [Naumenko 2020, 48], which is not surprising. According to Jackson and Sasamoto (2016), onomatopoeia is frequently used by poets and children's writers in other typologically distinct languages. It can also be frequently found in discourse domains that rely on sensory reference, such as romance novels, restaurant reviews, recipe debates, and advertising. However, beyond such literary and stylistic perspectives, onomatopoeia in Ukrainian has not yet been investigated through corpus-based methods, leaving a significant gap in quantitative and functional description. To fill in this gap, the present study employs a corpus-based approach to investigate the Ukrainian onomatopoeia by using the *ukTenTen22* corpus.

1.2 Aim of the Current Study

This study is based on the *ukTenTen22* corpus, a large Ukrainian web corpus compiled within the *TenTen* family of comparable web-based corpora available through Sketch Engine [Kilgarriff et al. 2014]. The *TenTen* corpora follow a unified methodology for building large web corpora [Jakubíček et al. 2013] and rely on efficient web-crawling procedures designed specifically for linguistic analysis [Suchomel & Pomikálek 2012; Suchomel 2020]. The *ukTenTen22* corpus contains over 7.5 billion words, collected from Ukrainian websites during major web-crawling periods in 2014, 2020, and 2023, and supplemented with *Wikipedia* data from 2020 and 2022 [Kilgarriff et al. 2014].

Given its size and composition, *ukTenTen22* contains a wide range of online text types, including news articles, blogs, forums, encyclopedic entries, reviews, and other forms of user-generated content. These genres differ in their stylistic, grammatical, and semantic properties, which directly affect the likelihood of encountering onomatopoeia. More expressive and narrative-oriented registers, such as blogs, personal posts, reviews, entertainment media, and discussion forums, may frequently include onomatopoeia because they rely on sensory description and evaluation and allow for affective expression [Naumenko 2020, 48]. In contrast, formal genres, such as news, reporting, encyclopedic writing, and administrative texts, typically avoid expressive vocabulary and therefore contain little or no onomatopoeia.

Although literary genres such as fiction, poetry, and children's literature are known to be rich in onomatopoeia [Junaid et al. 2023] and are usually considered underrepresented in web-derived corpora, the *ukTenTen22* data used in this study contained a notably large proportion of such materials. This indicates that online publication practices may increase the visibility of literary texts in contemporary web corpora.

This paper investigates the use of lexical onomatopoeia in contemporary written Ukrainian through corpus analysis, since there is, as already men-

tioned, a gap in research dedicated to this topic. This study focuses on the following research questions:

1. Which lexical onomatopoeic words appear most frequently in contemporary written Ukrainian?
2. How are they distributed across different written genres?
3. What stylistic and functional roles do they perform in context?

The next section outlines the methodology, followed by the results of the analysis and the conclusion. In summary, this paper aims to provide a novel corpus-based investigation of Ukrainian onomatopoeia.

2. Data and Methodology

In order to answer the three research questions mentioned above (see Section 1.2), the *ukTenTen22* corpus was analyzed through several sequential steps. Since Ukrainian onomatopoeia is classified as a subclass of interjections (Section 1.1), the corpus does not provide a separate tag for onomatopoeic words. Moreover, unlike some other Slavic corpora, interjections in *ukTenTen22* are not presented as an independently searchable part-of-speech category in the standard interface.

For comparison, the Slovak National Corpus treats interjections as a distinct grammatical category, which facilitates the identification of onomatopoeic expressions. “Since Slovak onomatopoeic words are treated as institutionalized interjections, the chance to find them in a representative dictionary is high. They can be identified in the Slovak National Corpus, but there is no separate list or dictionary of Slovak onomatopoeic words” [Körtvélyessy 2020, 15]. Due to the fact that the Ukrainian corpus does not provide interjections as a separate searchable category, several methodological steps were required to manually isolate and analyze the selected lexical onomatopoeic expressions in the research material.

First, a frequency wordlist grouped by morphological tags was consulted to identify the position of interjections in the corpus. In this list, interjections appear at position 180, with approximately 5.4 million tokens. After locating this category, all items marked with the interjection tag ([tag==”I”]) were retrieved for further analysis. However, automatic tagging proved partially unreliable, as many items labelled as interjections were semantically or functionally unrelated to this class. For instance, the list contained items such as *гей* (*hei*, ‘hey’) and *ну* (*nu*, ‘well’). Although frequent in everyday discourse, these items do not function as sound-imitative expressions, and were therefore excluded from the analysis.

To ensure accuracy, a list of approximately 30,000 interjections was manually reviewed. From this list, 45 lexical onomatopoeic expressions (see Appendix) were identified based on the definition presented in the Introduction.

Therefore, rather than being a comprehensive list of all Ukrainian onomatopoeia, the resulting list should be interpreted as a manually assembled working dataset.

Each onomatopoeic expression was then examined separately, including the frequency counts restricted by the filter ([tag=="I"]). In addition, ten concordance lines were examined to determine the contexts in which individual lexical sound imitations were used. The analysis revealed that some forms appeared in both literal and extended uses. The form *гав* (*hav*, 'bark'), for example, appeared both in its literal onomatopoeic function and in extended, non-imitative uses. In educational materials, it is usually presented as a typical onomatopoeic form, as in the concordance line:

(1) *Звуконаслідування – зображальний прийом імітації реальних звуків, що існують у навколишньому житті ("ку-ку", "гав", "тпру", "цок-цок" тощо).*
'Onomatopoeia is an expressive technique that imitates real sounds occurring in everyday life ("cuckoo", "bark", "tpru", "clip-clop", etc.).'

However, concordance searches also revealed figurative, idiomatic uses, such as *"І тут щось не ловіть гав"* ('And here, don't let your guard down'), where *гав (hav)* forms part of a fixed phrase and no longer denotes an imitated sound (i.e., a dog's bark), but, instead, functions as an idiomatic element rather than as onomatopoeia. Such contexts were excluded from the analysis, as the study focuses exclusively on lexical onomatopoeia in its imitative function.

To examine potential patterns within the dataset, the identified items were classified into semantic categories. Based on Hatsenko's (2003) framework, the onomatopoeic expressions were assigned to four groups:

1. Onomatopoeia related to humans and their actions (*ахчхи* (*akhchkhy*) 'achoo', *сьорб* (*s'orb*) 'slurp', *ням-ням* (*niam-niam*) 'yum-yum').
2. Onomatopoeic words that imitate the sounds of inanimate nature (*шурх-шурх* (*shurkh-shurkh*) 'rustle', *вжух* (*vzhukh*) 'whoosh', *плюсь* (*plius'*) 'splash').
3. Onomatopoeic words that convey the sounds of animals and birds (*гав* (*hav*) 'bark', *няв* (*niav)* 'meow', *мур-мур* (*mur-mur*) 'purr').
4. Sound imitations associated with the actions of objects set in motion by humans (*гуп* (*hup*) 'thud', *грюк* (*hriuk*) 'slam', *бах/бац* (*bakh/bats*) 'bang').

In total, the final dataset consisted of 450 tokens, representing individual corpus occurrences of 45 distinct lexical onomatopoeic types (see Appendix). These 45 types were identified manually and served as the basis for the semantic and contextual analysis. The 450-token sample enabled an initial evaluation of how lexical onomatopoeia functions in contemporary written

Ukrainian. However, several limitations must be acknowledged. First, some inconsistency in corpus tagging may have affected the results. Second, this paper focuses only on 45 onomatopoeic types; a much wider selection would provide a more comprehensive picture and allow for broader generalization. Despite these limitations, the dataset offers a solid foundation for demonstrating the main functional tendencies of lexical onomatopoeia in contemporary written Ukrainian.

3. Research Results

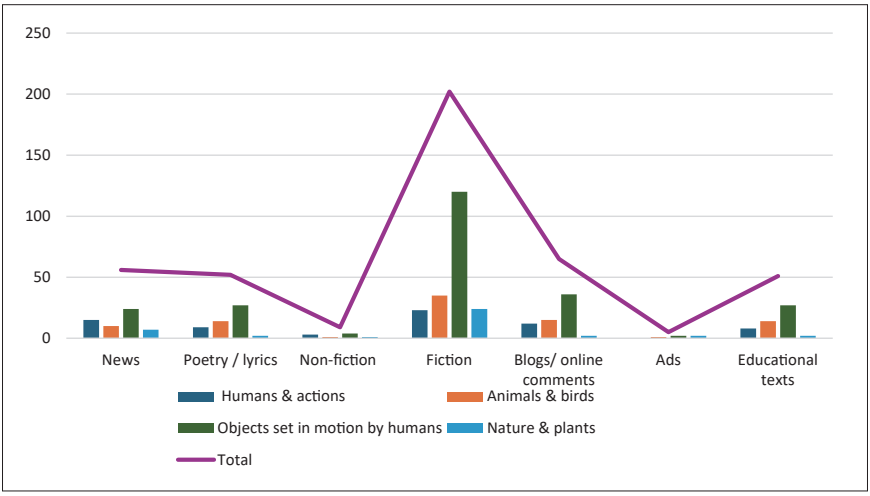
As explained in detail above (see Section 2), the analysis of the *ukTenTen22* corpus produced a set of onomatopoeic expressions that were manually extracted from the interjection category and classified into four semantic groups. Table 1 and Graph 1 below summarize the distribution of the 45 types of lexical onomatopoeic expressions across the four semantic categories, showing for each group the total frequency, the average frequency per item, and the tendency (a qualitative description of the genres in which these forms most commonly appear).

Graph 1 complements this list by plotting the distribution of genres across these categories in the *ukTenTen22* corpus. Bars represent category frequencies within each genre, and the total line shows the overall concentration of onomatopoeia across the text types. The analysis reveals that the distribution of onomatopoeic words in the *ukTenTen22* dataset demonstrates both contextual (see Section 3.1) and functional (see Section 3.2) differences across the semantic categories.

Table 1. Distribution of onomatopoeia by category

Category	**Examples**	**Total frequency**	**Average frequency**	**Tendency**
Humans and their actions	ахчхи (*akhchkhý*) ‘achoo’, ням-ням (*niam-niam*) ‘yum-yum’	4246	606	Mainly fiction, but also in news, blogs, teacher handbooks
Inanimate nature and plants	бульк (*bul'k)* ‘plop’, вжух (*vzhukh)* ‘whoosh’	745	186	Almost exclusively literary context, except for *whoosh* (newspapers)
Animals and birds	гав (*hav*) ‘bark’, ку-ку (ku-ku) ‘cuckoo’	7065	785	Mostly fiction, some presence in blogs, online comments, news, and educational texts

Category	Examples	Total frequency	Average frequency	Tendency
Objects (set in motion by humans)	бабах (*babakh)* 'bang', цінь (*tsin')* 'ting'	9578	400	Most diverse category: many forms appear in literary works, educational texts and others are used across blogs, online comments. Explosion-related words occur predominantly in the news



Graph 1. Distribution of onomatopoeia across text genres

3.1 Contextual Differences

The distribution of onomatopoeic words in the *ukTenTen22* corpus reveals clear contextual differences across the semantic categories. Although many onomatopoeic expressions are associated with literary writing, others occur more widely across different types of texts, including news, blogs, online communication, educational texts, and other forms of written discourse. The following paragraphs describe these patterns in more detail, organized by semantic category.

The category "Objects set in motion by humans" is the most frequent, with a total frequency of 9578 tokens and an average of approximately 400 occurrences per item. Onomatopoeia referring to animals and birds shows the highest average frequency per lexical item, with an average of 785 occurrences and a total frequency of 7065 occurrences. Human-related sound imitations

occupy an intermediate position, with a total frequency of 4246 and an average frequency of 606 occurrences per item. The least frequent category, "Inanimate nature and plants", has a total frequency of 745 and an average frequency of 186 occurrences per item.

The category "Objects set in motion by humans" includes words such as *бабах (babakh*) 'bang' and *цінь (tsin')* 'ting', which appear in a broad range of genres, including news, blogs, fiction, poetry/lyrics, and educational materials. Unlike some other categories, these expressions are not used solely for expressive or stylistic purposes. They often appear in the news discourse, where object-related onomatopoeia is frequently used to foreground dynamic events, including impact- and explosion-related phenomena, often in connection with the current political events. This distribution explains why the category has the highest total frequency in the corpus: its forms are not limited to literary stylization but can also function in the informational and public discourse.

The "Animals and birds" category includes words such as *гав (hav*) 'bark' and *мур-мур* (*mur-mur*) 'purr'. These forms are well-established and occur most often in fictional texts, particularly in children's literature and fairytales. They can also be found in song lyrics and poetry; however, these occurrences are still limited in number. Some occurrences are evident in educational texts and online blogs. Outside of fiction, their use is typically linked to contexts in which animals are explicitly involved, for example, in descriptions of animal behavior or in discussions of human-animal interaction. The high average frequency of this group suggests that a smaller number of strongly conventionalized forms accounts for a substantial proportion of the data.

The category "Humans and their actions" includes words such as *ахчхи (akhchkhý*) 'achoo' and *ням-ням (niam-niam)* 'yum-yum', which are used mainly in fiction. However, some onomatopoeic expressions may appear in news texts and online discourse, but these remain relatively rare and tend to occur in passages that allow for a more informal tone.

By contrast, onomatopoeia from the last semantic category, the sounds of inanimate nature and plants, remains marginal. Nearly all of these expressions are used in literary writing. Only a few occurrences can be found in non-literary writing, most notably, in weather forecasts, and, even there, their use is highly limited.

3.2 Functional Differences

Beyond contextual distribution, the data reveal functional differences among the semantic categories. Object-related onomatopoeia frequently serves an informational or attention-directing function. In the news discourse in particular, such forms are used to foreground dynamic events. This category, therefore, demonstrates that onomatopoeia in written Ukrainian is not restricted to play-

ful or poetic usage but may also participate in the representation of socially significant events.

Animal-related forms, while widely recognized, function predominantly as referential labels for well-established sounds and as symbolic cultural markers. Their frequent occurrence in children's literature and some educational materials confirms their conventional status.

Onomatopoeia related to humans and their actions functions primarily as an expressive device, enhancing descriptions of bodily movements, reactions, or emotional states, particularly in narrative and descriptive contexts. By contrast, sound imitations of inanimate nature and plants serve primarily aesthetic and poetic functions, contributing to the sensory atmosphere of literary texts rather than to the transmission of concrete information.

Ukrainian onomatopoeia does not constitute a functionally uniform class, as demonstrated by the distinction between the categories. Rather, the type of sound being imitated and the genre in which the form appears have a significant impact on its function in written discourse.

3.3 Pragmatic and Typological Extensions

Although the use of onomatopoeia remains most prevalent in fiction, especially in children's literature and educational texts, as confirmed by previous studies [Jackson & Sasamoto 2016; Naumenko 2020], the data also show a wider range of contexts. The fact that onomatopoeic words can be found in blogs, online comments, and news articles indicates that their use is not limited to literary stylization but extends to informal, public, and media-based written discourse (see Graph 1).

From a semiotic point of view, these results show that Ukrainian onomatopoeia serves as a symbolic marker, with its significance and scope varying by semantic category. Sound imitations of inanimate nature and plants primarily fall within the aesthetic sphere, while imitations of animal and bird sounds serve as culturally rooted reference points. Onomatopoeia of human-related sounds primarily serves as an expressive nuance, whereas onomatopoeia of object-related sounds attracts attention in written speech. The difference between the average and total frequency indicated an unequal distribution, with certain extremely common forms dominating each category while others remained on the periphery. This imbalance suggests that, while the overall inventory is broad, the corpus data are dominated by a limited set of highly conventionalized items.

Over time, certain onomatopoeic expressions have also developed additional symbolic meanings. For example, *ку-ку* (*ku-ku*) 'cuckoo' not only imitates the sound made by the bird but also is used when someone wants to draw attention:

(2) “**Кукý, кукý**! А де ти? Тут, татусю? Візьми на руці, поцілуй Марусю!” [Леся Українка]
‘**Cuckoo, cuckoo**! Where are you? Here, daddy? Hold me in your arms, kiss Marusya!’ [Lesya Ukrainka]

This demonstrates how onomatopoeia may evolve into a culturally significant expression that goes beyond its original imitative function, which can also be observed in the corpus. This pragmatic evolution is observable in several Slavic languages, and it positions Ukrainian within the same broader typological pattern. For instance, in Czech, Blatná (1996) notes that while onomatopoeia is semantically neutral in its primary, imitative use, it frequently acquires expressive or figurative meanings (e.g., “[…] ty hrozně kuňkáš, vůbec ti nerozumím” (‘you’re croaking/whining so much, I can’t understand you at all’)) and is especially common in child-directed speech (e.g., *ťapy ťap* ‘little steps’).

4 Conclusion

The corpus evidence demonstrates that Ukrainian onomatopoeia participates in a wide range of communicative and stylistic functions. While some categories are predominantly literary, others (especially object- and human-related forms) extend across everyday written communication, including news media and digital discourse. The presence of figurative extensions further demonstrates the semiotic adaptability of onomatopoeia.

The findings indicate that Ukrainian onomatopoeia is a dynamic and multifunctional component of the lexicon. Its use, distribution, and pragmatic functions require further linguistic investigation, as it is not merely a peripheral element of children’s literature. The study also illustrates that, particularly in terms of the genre distribution and contextual function, a corpus-based approach can uncover patterns that are less apparent in purely theoretical or dictionary-based descriptions.

In summary, this study relies only on written data from the *ukTenTen22* corpus processed by *Sketch Engine*. As a result, spontaneous spoken onomatopoeia is not included in the current analysis (see Data and Methodology). Furthermore, the data were manually identified and categorized because onomatopoeic words are not tagged as a distinct category in the corpus. Despite these limitations, the research provides valuable insights into the frequency of occurrence, genre distribution, and functions of onomatopoeia in contemporary written Ukrainian.

Data Sources

UKTENTEN22 CORPUS, 2022. Ukrainian web corpus (TenTen family). Brno: Lexical Computing / Sketch Engine. https://www.sketchengine.eu/uktenten-ukrainian-corpus/

References

AKITA, K., 2013. Constraints on the semantic extension of onomatopoeia, *The Public Journal of Semiotics*, 5(1). 21–37. https://doi.org/10.37693/pjos.2013.5.9646

ANDREJ, Ľ., 2020. Derivational networks of onomatopoeia, *Zeszyty Naukowe Uniwersytetu Rzeszowskiego. Seria Filologiczna: Studia Anglica Resoviensia*, 17. 47–60. https://doi.org/10.15584/sar.2020.17.4

BAŃKO, M., 2008. *Współczesny polski onomatopeikon: ikoniczność w języku.* Warszawa: PWN.

BLATNÁ, R., 1996. Zvukomalba a pragmatika. In NEBESKÁ, I., MACUROVÁ, A. (eds.). *Jazyk a jeho užívání: sborník k životnímu jubileu profesora Oldřicha Uličného.* Praha: FFUK, 93–102.

de SAUSSURE, F., 1959. *Course in general linguistics.* Transl. W. Baskin; ed. A. Sechehaye and C. Bally. New York: Philosophical Library.

DVONČ, L., HORÁK, G., MIKO, F., MISTRÍK, J., ORAVEC, J., RUŽIČKA, J., URBANČOK, M., 1966. *Morfológia slovenského jazyka.* Bratislava: Vydavateľstvo Slovenskej akadémie vied.

FISCHER, A., 1999. What, if anything, is phonological iconicity? In NÄNNY, M., FISCHER, O. (eds.). *Form miming meaning: Iconicity in Language and Literature.* Amsterdam: John Benjamins, 123–134. https://doi.org/10.1075/ill.1

GREGOVÁ, R., 2021. On the position of onomatopoeia in adult language: evidence from Slovak, *Lingua Posnaniensis*, 63(1). 7–19. https://doi.org/10.2478/linpo-2021-0001

HATSENKO, I. O., 2002. Sintaksychna funktsiia zvukoimidatyvnykh sliv ukrainskoi movy [ГАЦЕНКО, І. О. Синтаксична функція звуконаслідувальних слів української мови], *Siversianskyi litopys*, 6. 97–100.

HATSENKO, I. O., 2003. Klasyfikatsiia zvukoimidatyvnykh sliv ukrainskoi movy [ГАЦЕНКО, І. О. Класифікація звуконаслідувальних слів української мови за їх семантичними ознаками], *Current Issues of Ukrainian Linguistics : Theory and Practice*, 7. 122–127.

HATSENKO, I. O., 2003. *Typolohichni osoblyvosti zvukoimidatyvnykh sliv: na materiali ukrainskoi, rosiiskoi ta angliiskoi mov.* PhD thesis. [ГАЦЕНКО, І. О. *Типологічні особливості звуконаслідувальних слів (на матеріалі української, російської та англійської мов): автореф. дис. ... канд. філол. наук*]. Kyiv: Taras Shevchenko National University of Kyiv.

HRYTSU, D., 2025. *An acoustic and perceptual analysis of onomatopoeia in the sample of Indo-European languages: a cross-linguistic analysis of English, Slovak, Ukrainian, and Spanish.* PhD thesis. Pavol Jozef Šafárik University in Košice. https://opac.crzp.sk/?fn=detailBiblioForm&sid=F3AF90715DF24108A63D17E80399

IMAI, M., KITA, S., 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651). Article 20130298. https://doi.org/10.1098/rstb.2013.0298

JACKSON, R., SASAMOTO, R., 2016. Onomatopoeia – showing-word or saying-word? Relevance theory, lexis, and the communication of impressions, *Lingua*, 175–176. 36–53. https://doi.org/10.1016/j.lingua.2015.11.003

JAHDHAMI, S., 2020. Onomatopoeia in Omani Arabic, *Journal of Advances in Linguistics*, 11. 63–74. https://doi.org/10.24297/jal.v11i.8898

JAKUBÍČEK, M., KILGARRIFF, A., KOVÁŘ, V., RYCHLÝ, P., SUCHOMEL, V., 2013. The TenTen corpus family. In *Proceedings of the 7th International Corpus Linguistics Conference CL*, 125–127. https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf

JUNAID, S., MUZZAMMIL, A., MUJIZAT, A., ANDINI, C., 2023. Onomatopoeia variation among cultures: an exploration in selected children's story books, *ELS Journal on Interdisciplinary Studies in Humanities*, 6(4). 658–664. https://journal.unhas.ac.id/index.php/jish/article/view/31437

KILGARRIFF, A., BAISA, V., BUŠTA, J., JAKUBÍČEK, M., KOVÁŘ, V., MICHELFEIT, J., RYCHLÝ, P., SUCHOMEL, V., 2014. The Sketch Engine: ten years on, *Lexicography*, 1(1). 7–36. https://doi.org/10.1007/s40607-014-0009-9

KLEŇHOVÁ, E., 2010. *Interjekce v českém jazykovém systému.* Diplomová práce. Praha: FFUK.

KÖRTVÉLYESSY, L., 2020. Onomatopoeia – a unique species? *Studia Linguistica*, 74(2). 506–551. https://doi.org/10.1111/stul.12133

KÖRTVÉLYESSY, L., ŠTEKAUER, P. (eds.), 2024. *Onomatopoeia in the world's languages: A comparative handbook.* Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783111053226

LAING, C., VIHMAN, M., KEREN-PORTNOY, T., 2017. How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech, *Journal of Child Language*, 44(5). 1117–1139. https://doi.org/10.1017/S0305000916000428

LAING, C., 2019. A role for onomatopoeia in early language: evidence from phonological development, *Language and Cognition*, 11(2). 173–187. https://doi.org/10.1017/langcog.2018.23

MOTAMEDI, Y., MURGIANO, M., PERNISS, P., WONNACOTT, E., MARSHALL, C., GOLDIN-MEADOW, S., VIGLIOCCO, G., 2020. Linking language to sensory experience: onomatopoeia in early language development, *Developmental Science*, 24(3). Article e13066. https://doi.org/10.1111/desc.13066

NAUMENKO, N. V., 2020. Onomatopoeias of contemporary Ukrainian poetry in the aspect of artistic synthesis, *International Humanitarian University Herald. Philology*, 1(45). 46–49. https://doi.org/10.32841/2409-1154.2020.45-1.12

ORAVEC, J., BAJZÍKOVÁ, E., FURDÍK, J., 1984. *Súčasný spisovný slovenský jazyk. Morfológia.* Bratislava: SPN.

RYDBLOM, O., 2016. *Universals in usage of and attitudes to onomatopoeia.* Bachelor's thesis. Lund University. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8888824&fileOId=8888825

RUSÍNOVÁ, Z., 1996. *Příruční mluvnice češtiny.* 2. vyd. Praha: Nakladatelství Lidové noviny.

SIMPSON, P., 2004. *Stylistics: A resource book for students.* London: Routledge.

SUCHOMEL, V., 2020. *Better web corpora for corpus linguistics and NLP.* Brno: Masaryk University. https://is.muni.cz/th/u4rmz/

SUCHOMEL, V., POMIKÁLEK, J., 2012. Efficient web crawling for large text corpora. In *Proceedings of the Seventh Web as Corpus Workshop (WAC7)*, 39–43. https://www.sketchengine.eu/wp-content/uploads/Efficient_Web_2012.pdf

ZIMA, J., 1961. *Expresivita slova v současné češtině: studie lexikologická a stylistická.* Praha: Nakladatelství Československé akademie věd.

Appendix. List of Analyzed Lexical Onomatopoeia

1. Onomatopoeia Related to Humans and Their Actions

Ukrainian Onomatopoeia	**Meaning**	**English Equivalent**
ахчхи (akhchkhý)	sneeze sound	achoo
ням-ням (niam-niam)	eating/chewing sound	yum-yum
цьом (ts′om)	kissing sound	mwah / kiss
тьфу-тьфу (tfu-tfu)	spitting / dismissive sound	ptoo / spit
туп (tup)	thudding step	thump
гам (ham)	low humming / murmuring sound	hum
блим (blym)	brief flash / blink sound	blink / flash

2. Onomatopoeic Words that Imitate the Sounds of Inanimate Nature

Ukrainian Onomatopoeia	**Meaning**	**English Equivalent**
бульк (bul′k)	bubbling / soft plop	plop / blub
вжух (vzhukh)	fast movement through air	whoosh
беркиць (berkyts′)	heavy falling sound	plop / thud
шелесь (sheles′)	light rustling / swish	swish / rustle

3. Onomatopoeic Words that Convey the Sounds of Animals and Birds

Ukrainian Onomatopoeia	**Meaning**	**English Equivalent**
гав (hav)	dog bark	bark / woof
ку-ку (ku-ku)	cuckoo bird sound	cuckoo
ку-ку-рі-ку (ku-ku-ri-ku)	rooster crow	cock-a-doodle-doo
кум-кум (kum-kum)	frog call	ribbit / croak
кря (kria)	duck sound	quack
мур-мур (mur-mur)	cat purring	purr
няв (niav)	cat meowing	meow
хрю (khriu)	pig grunt	oink
цвірінь (tsvirin′)	bird chirping	tweet / chirp

4. Sound Imitations of Objects Set in Motion by Humans

Ukrainian Onomatopoeia	**Meaning**	**English Equivalent**
бабах (babakh)	loud bang	bang
бам-бам (bam-bam)	repeated banging	bam-bam
бац (bats)	sudden hit / bam	bam / smack
бринь (bryn′)	light metallic sound	plink / ding
гуп (hup)	dull thud	thud
грюк (hriuk)	heavy slam	slam
гульк (hulk)	pop / sudden appearance	pop
клац (klats)	sharp click	click
лусь (lus′)	snap / crack	snap / crack
тік-так (tik-tak)	rhythm of a clock	tick-tock
торох (torokh)	clatter	clatter
тук-тук (tuk-tuk)	knocking	knock-knock
ту-ту-ту (tu-tu-tu)	horn sound	toot-toot-toot
хрум (khrum)	crunch	crunch
хрусь (khrus′)	crunch / snap	crunch / snap
хрясь (khrias′)	crash / smack	crash
цок-цок (tsok-tsok)	hoof-like clicking	clip-clop / click-click
цінь (tsin′)	light metallic ting	ting
чух-чух (chukh-chukh)	sound of a moving train	choo-choo
хвіть (khvit′)	swish / slap	swish / slap
туп-туц (tup-tuts)	rhythmic thumping	thump-thump
брязь (briaz′)	sharp clang	clang
дзень (dzen′)	clang / ding	clang
дзінь (dzin′)	ding	ding
трах (trakh)	explosive bang	bang
тріськ (tris′)	sharp crack	crack